



The 18<sup>th</sup> National Conference on Computing  
and Information Technology

May 19-20 2022

# Conference Proceedings



**KMUTNB**



Faculty of Information Technology and Digital Innovation  
King Mongkut's University of Technology North Bangkok  
Kanchanaburi Rajabhat University  
Council of IT Deans of Thailand

**[www.nccit.net](http://www.nccit.net)**

## TABLE OF CONTENTS

Message from KMUTNB President.....	ii
Message from KRU President.....	iv
Message from General Chair .....	vi
Conference Organizers.....	x
Conference Organization Committee .....	xii
Steering Committee .....	xiii
Technical Program Committee .....	xiv
Technical Program Contents.....	xiv
Regular Papers .....	1
Author Index .....	529

## สารจากอธิการบดี มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ในยุคเศรษฐกิจดิจิทัลที่ต้องอาศัยพื้นฐานจากระบบคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ในการสนับสนุนการจัดการด้านต่าง ๆ เช่น ด้านการจัดการเรียนการสอน ด้านธุรกิจ ด้านเศรษฐกิจ ด้านการเงิน ด้านการทหาร ด้านการแพทย์ ด้านคมนาคม ด้านอุตสาหกรรม ด้านการเกษตร และอื่น ๆ อีกมากมาย ประเทศไทยได้เห็นความสำคัญของคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ดังได้มีการลงทุนทรัพยากร เพื่อการวางโครงสร้างและปัจจัยพื้นฐานด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ เพื่อให้การศึกษาของ ประชาชนในชาติพร้อมพัฒนานวัตกรรมการเรียนการสอน ที่มุ่งเน้นการนำเทคโนโลยีสารสนเทศมาใช้เป็น ประโยชน์ได้อย่างเป็นรูปธรรมในการจัดการศึกษาทุกระดับ ในด้านการประยุกต์ใช้งานภาคอุตสาหกรรม เทคโนโลยีสารสนเทศจะช่วยสนับสนุนให้การบริการงานและการตัดสินใจต่าง ๆ เป็นไปอย่างรวดเร็วเพิ่ม ผลผลิตของอุตสาหกรรม การขนส่งที่รวดเร็ว และธุรกิจได้กำไรทางการค้าที่สูง

เป็นเรื่องที่น่ายินดีที่เครือข่ายหน่วยงานวิชาการด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศได้ ร่วมกันจัดงานประชุมวิชาการด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศอย่างต่อเนื่องทุก ๆ ปี เป็นเวที วิชาการสำหรับอาจารย์ นักวิจัย และนักศึกษา

ท้ายที่สุดนี้ ขอขอบพระคุณกรรมการเครือข่ายหน่วยงานวิชาการ คณะกรรมการจัดงานทุกภาคส่วน โดยเฉพาะคณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ที่ได้ร่วมจัดงานประชุมทางวิชาการอย่างต่อเนื่อง ขออวยพรให้การประชุมทางวิชาการครั้งนี้ ประสบความสำเร็จลุล่วงตามวัตถุประสงค์ที่ตั้งไว้ทุกประการ



(ศาสตราจารย์ ดร.สุชาติ เชิญฉิน)

อธิการบดี มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

## **MESSAGE FROM KMUTNB PRESIDENT**

Base on Government's policy to move forward Thailand's Economy, Digital is a key mechanism in economy and society to enhance efficiency of business competition especially 10 new aimed industries (New S-Curve) which are mechanism in innovation economy to increase potential of Thailand's business competition in the future. The Government has developed a strategy to improve an efficient basis of digital structure to completely cover the expected services. To move forward the economy by digital technology, create a thorough and digitalized society, reform government units to be digitalized, prepare human resources for digitalized society and economy, and digital literacy.

The National Conference on Computing and Information Technology (NCCIT) responses the Government's policy in digital strategy to support and encourage publication of Computer Innovation, Information Technology, and Digital Innovation. NCCIT is a conference for professors, researchers, and students who prefer to present and publish their papers, exchange knowledge or find cooperation. It is a potential developing encouragement to enhance Computer and Information Technology researches for applicable usage and move forward to globalization.

Finally, I would like to thank everyone including the NCCIT collaborative universities, all co-organizers, especially the staff members of the Faculty of Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok to organize and manage this conference. I wish the conference achieves its purposes.



(Professor Dr.-Ing. habil. Suchart Siengchin)  
President, King Mongkut's University of Technology North Bangkok

## สารจากอธิการบดี มหาวิทยาลัยราชภัฏกาญจนบุรี

หนังสือเล่มนี้เป็นเอกสารรวมเล่มรายงานผลวิจัยจากการประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 18 (NCCIT2022) จัดขึ้น ณ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ระหว่างวันที่ 19-20 พฤษภาคม พ.ศ. 2565 ซึ่งจัดรวมกับการประชุมวิชาการระดับนานาชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศครั้งที่ 18 (The 18<sup>th</sup> International Conference on Computing and Information Technology: IC<sup>2</sup>IT2022) เป็นการจัดงานร่วมกันของมหาวิทยาลัยเครือข่ายทั้งในประเทศและต่างประเทศ

สำหรับปีนี้ NCCIT2022 มีบทความที่ผ่านการพิจารณาจากผู้ทรงคุณวุฒิพิจารณาจำนวน 3 ท่าน และได้รับการตอบรับจากผู้ทรงคุณวุฒิอย่างน้อย 2 ใน 3 ท่าน ผลงานวิจัยที่ได้รับการตอบรับมีจำนวน 86 บทความจากสถาบันทั่วประเทศ ที่นำเสนอปากเปล่าและนำบรรจุในเอกสารรวมเล่มรายงานผลวิจัยงานประชุมวิชาการ NCCIT2022

คณะกรรมการจัดงานประชุม NCCIT2022 ขอขอบคุณนักวิจัยทุกท่านที่ส่งบทความวิจัยเพื่อพิจารณา ทั้งผู้ที่ได้รับการตอบรับและผู้ที่ไม่ได้รับการตอบรับในการนำเสนอบทความวิจัย ขอขอบคุณคณะกรรมการพิจารณาบทความที่เสียสละเวลาอันมีค่าเพื่อพิจารณาบทความอย่างเข้มข้นได้มาตรฐาน ทั้งนี้ เพื่อให้ได้ผลงานวิจัยที่มีคุณภาพในการลงพิมพ์ในเอกสารรวมเล่มรายงานผลวิจัย และการนำเสนอในวันประชุมวิชาการ NCCIT2022

ขอขอบคุณมหาวิทยาลัยเครือข่ายในประเทศทั้ง 8 แห่ง ที่ให้การสนับสนุนเป็นอย่างดีทั้งด้านการเงินและบุคลากร ในการร่วมเป็นกรรมการเครือข่าย กรรมการดำเนินงาน และกรรมการพิจารณาบทความ ได้แก่ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ มหาวิทยาลัยราชภัฏกาญจนบุรี มหาวิทยาลัยราชภัฏนครปฐม มหาวิทยาลัยราชภัฏวชิรวิทยาดงขี้เหล็ก มหาวิทยาลัยราชภัฏธนบุรี สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ และคณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏเพชรบุรี

หวังเป็นอย่างยิ่งว่า NCCIT จะ ได้ให้โอกาสที่ดีสำหรับคณาจารย์ นักวิจัย นักวิชาการ และนักศึกษา ได้นำเสนอผลงานเป็นการพัฒนาตนเองไปสู่ความเป็นนักวิจัยที่เข้มแข็งยิ่ง ๆ ขึ้นไป



(ผู้ช่วยศาสตราจารย์ ดร.พนนิษฐ์ สุขชวานา)  
อธิการบดี มหาวิทยาลัยราชภัฏกาญจนบุรี

## MESSAGE FROM KRU PRESIDENT

This volume contains the papers of the 18<sup>th</sup> National Conference on Computing and Information Technology (NCCIT2022) held at Faculty of Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok, on May 19<sup>th</sup> – 20<sup>th</sup>, 2022. Traditionally, it is organized in conjunction with The 18<sup>th</sup> International Conference on Computing and Information Technology (IC2IT2022), which is supported by many universities both in Thailand and overseas.

This year, the program committee of NCCIT2022 Each submission was reviewed by at least 2, mostly 3 members of the program committee to avoid contradictory results. On these judgments, the committee decided to accept 86 papers for oral presentation and inclusion in the conference proceedings of NCCIT2022.

I would like to also thank the 8 co-operative universities for supporting both in finance and personnel, which are Faculty of Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok (KMUTNB), Kanchanaburi Rajabhat University (KRU), NakhonPathom Rajabhat University (NPRU), Rajamangala University of Technology Thanyaburi (RMUTT), Rajamangala University of Technology Krung Thep (RMUTK), Faculty of Information Science Mahasarakham University (MSU), Faculty of Applied Statistics, National Institute of Development Administration (NIDA), and Faculty of Information Technology, Phetchaburi Rajabhat University (PRU).

Hopefully, NCCIT will provide good opportunities for faculty, researchers, scholars and students to present their work to develop themselves to become qualified researchers.



(Assistant Professor Dr. Photjane Sukchaona)  
President, Kanchanaburi Rajabhat University

## สารจากประธานจัดงานประชุมวิชาการ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

งานประชุมวิชาการระดับชาติและระดับนานาชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ จัดขึ้นอย่างต่อเนื่องเป็นประจำทุกปี ตั้งแต่ พ.ศ. 2548 ในครั้งนี้เป็นการจัดงาน ครั้งที่ 18 เป็นความร่วมมือของหน่วยงานเครือข่ายวิชาการจัดงานประชุมวิชาการ NCCIT และ IC<sup>2</sup>IT โดยมีวัตถุประสงค์เพื่อเป็นเวทีสำหรับการนำเสนอผลงานวิจัยด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ของอาจารย์ นักศึกษา นักวิจัย จากสถาบันต่าง ๆ ทั้งในประเทศและต่างประเทศ เพื่อให้ได้พบปะกันทั้งภาครัฐและภาคธุรกิจ เป็นการเผยแพร่ผลงานวิจัยด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ จากทฤษฎีงานประยุกต์ เพื่อขับเคลื่อนสังคม

เทคโนโลยีสารสนเทศกับสถานการณ์ในปัจจุบันของการแพร่ระบาดของเชื้อไวรัสโคโรนา หรือ COVID-19 เกิดขึ้นทั่วโลกในช่วงระยะเวลา 2 ปีนี้ ส่งผลกระทบต่อการใช้ชีวิตของมนุษย์ในปัจจุบันเป็นอย่างมาก อย่างไรก็ตาม วิกฤติในสถานการณ์เช่นนี้เราไม่สามารถปฏิเสธความสำคัญของเทคโนโลยีดิจิทัล ในฐานะเครื่องมือในการช่วยลดการแพร่ระบาดของ COVID-19 ได้เลย ซึ่งวิกฤตครั้งนี้ทำให้มีการนำนวัตกรรม และเทคโนโลยีเข้ามาเป็นกำลังเสริม เพื่อเป็นการลดการแพร่ระบาดของโรคประยุกต์ใช้เทคโนโลยีให้เหมาะสมกับสังคมปัจจุบัน และอำนวยความสะดวกในการดำเนินชีวิตได้อย่างปลอดภัย การสร้างนวัตกรรมใหม่จากเทคโนโลยีดิจิทัลก็เป็นอีกหนึ่งพลังขับเคลื่อนสำคัญที่จะช่วยให้มนุษยชาติบริหารวิกฤติและผ่านพ้นจุดเปลี่ยนสำคัญของศตวรรษนี้ไปได้อย่างรวดเร็วและมีประสิทธิภาพ

ในงานประชุมวิชาการครั้งนี้ต้องขอขอบคุณทุกท่านสำหรับการนำเสนอผลงานการวิจัยที่มีคุณภาพสูง ทั้งนี้มีบทความจำนวนหนึ่งที่ไม่ได้รับการนำเสนอผลงาน เนื่องจากความเข้มข้นของการพิจารณาจากคณะกรรมการที่ต้องรักษาคุณภาพและมาตรฐานงานวิจัย ขอขอบพระคุณคณะกรรมการพิจารณาบทความทุกท่านที่เสียสละเวลาอันมีค่าเพื่อให้ได้ผลงานวิจัยที่มีคุณภาพนำเสนอในงานประชุมวิชาการในครั้งนี้ด้วย

ขอขอบคุณมหาวิทยาลัยเครือข่ายทั้ง 8 แห่ง ที่ให้การสนับสนุนเป็นอย่างดีทั้งด้านคณาจารย์ บุคลากรในการร่วมเป็นคณะกรรมการเครือข่าย คณะกรรมการดำเนินงาน และคณะกรรมการพิจารณาบทความ อีกทั้งด้านงบประมาณ ขอขอบพระคุณ ศาสตราจารย์ ดร.สุชาติ เชี่ยงฉิน อธิการบดี มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ สำหรับการสนับสนุนการประชุมวิชาการ NCCIT และ IC<sup>2</sup>IT อย่างต่อเนื่อง ขอขอบคุณ ผู้ช่วยศาสตราจารย์ ดร.ศักดิ์ชาย ตั้งวรรณวิทย์ อาจารย์ ดร.วัชรวิกรม จิตต์สกุล นางสาวพรพิมล ฝ่ายเทศ และทีมงานจากคณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ที่ได้ประสานงานกับนักวิจัยและผู้ที่เกี่ยวข้องเพื่อรวบรวม ผลงานการวิจัยสำหรับงานประชุมทางวิชาการให้เสร็จสิ้นภายในระยะเวลาที่เหมาะสม

ทำนนี้หวังเป็นอย่างยิ่งว่างานประชุมวิชาการ NCCIT และ IC²IT จะเป็นเวทีในการนำเสนอผลงานวิจัยและนำไปสู่การพัฒนาตนเองเพื่อให้นักวิจัยที่ดีและมีคุณภาพสูงยิ่ง ๆ ขึ้นไป NCCIT และ IC²IT จะเป็นแพลตฟอร์มสำหรับการแลกเปลี่ยนความรู้ในด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ สร้างแรงบันดาลใจให้กับนักวิจัยเพื่อคิดค้นนวัตกรรมและพัฒนาองค์ความรู้ใหม่ ๆ เพื่อการพัฒนาชาติอย่างยั่งยืนตลอดไป



(ผู้ช่วยศาสตราจารย์ ดร.สุนันทา สดสี)  
คณบดีคณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล  
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ  
ประธานเครือข่ายงานประชุมวิชาการ NCCIT



## MESSAGE FROM GENERAL CHAIR

The National Conference on Computing and Information Technology (NCCIT) and the International Conference on Computing and Information Technology (IC<sup>2</sup>IT) are organized annually in conjunction with an academic network comprising eight universities. Since their inception in 2005, NCCIT & IC<sup>2</sup>IT have continually contributed to the state-of-the-art through knowledge dissemination and discussion of Computing and Information Technologies and their impact on our lives. This, the 18<sup>th</sup> gathering, once again provides a unique platform both for showcasing innovative research in computer and information technology, and for international and domestic academics and students to unite with and discuss their findings with delegates from government and business. Research contributions at this year's Conference consider computer and Information Technology theory and practice for social mobilization.

For the past 2 years, the Coronavirus (COVID-19) pandemic has impacted dramatically on people's way of life. However, it is apparent the application and adaptation of both new and existing digital technologies have played a significant role in reducing spread of the disease, while making our lives safer and more convenient. It is therefore vital mankind continues to push back the frontiers of digital innovation and infrastructure and to foster a culture of global cooperation to ensure our readiness to meet future challenges in the decades to come.

A debt of gratitude is owed to our eight university partners in Thailand for their financial support and for the contribution of all staffs in serving as network committees, executive director committees, and technical program committees.

Huge thanks are extended to Professor Dr.-Ing. habil. Suchart Siengchin, President of King Mongkut's University of Technology North Bangkok (KMUTNB), for his unwavering support for NCCIT & IC<sup>2</sup>IT and for providing us with the necessary resources.

A very special thank you goes to Assistant Professor Dr. Sakchai Tangwannawit, Dr. Watchareewan Jitsakul, Miss Pornpimon Faythet, and all academic staffs from the Faculty of Information Technology and Digital Innovation (ITD), KMUTNB for collaborating, collating and timely production of our quality conference proceedings. Thank you also to NCCIT & IC<sup>2</sup>IT network universities for providing personnel and budget support.

This conference could not happen without innovative and progressive contributions from the research community at large. Special thanks are due to all our technical committee members for their diligent consideration of all submissions and for maintaining and

perpetuating the high standards for which NCCIT & IC<sup>2</sup>IT is justifiably renowned. Regrettably, as a consequence of our rigorous peer review process, we were unable to accept some articles for presentation.

Finally, looking to the future, NCCIT & IC<sup>2</sup>IT will continue to serve as a community platform for research, while raising the bar for high quality academic investigation. It will provide an enduring forum for knowledge exchange in computer and information technology. And it will inspire researchers to develop innovative products, to exchange ideas and to build up a research network geared to sustainable development.

A handwritten signature in blue ink, reading "Sunantha Sodsee". The signature is written in a cursive style with a large initial 'S'.

Assistant Professor Dr. Dr.-Ing. Sunantha Sodsee  
Dean, Faculty of Information Technology and Digital Innovation  
King Mongkut's University of Technology North Bangkok  
NCCIT2022 General Chair

## CONFERENCE ORGANIZERS



King Mongkut's University of Technology North Bangkok, Thailand  
Faculty of Information Technology and Digital Innovation

Fern Universität in Hagen, Germany



Oklahoma State University, USA

Chemnitz University, Germany



Edith Cowan University, Australia

Hanoi National University of Education, Vietnam



Gesellschaft für Informatik (GI), Germany

Kanchanaburi Rajabhat University, Thailand



Phetchaburi Rajabhat University, Thailand  
Faculty of Information Technology

## CONFERENCE ORGANIZERS



Nakhon Pathom Rajabhat University, Thailand

Rajamangala University of Technology Thanyaburi, Thailand  
Office of Academic Resources and Information Technology



Rajamangala University of Technology Krungthep, Thailand  
Center of Academic Resources and Information Technology

Maharakham University, Thailand  
Faculty of Informatics



National Institute of Development Administration, Thailand  
Information Technology Management Program,  
Graduate School of Applied Statistics

Council of IT Deans of Thailand (CITT), Thailand



## **CONFERENCE ORGANIZATION COMMITTEE**

General Chair: Asst. Prof. Dr. Dr.-Ing. Sunantha Sodsee, KMUTNB  
Secretary: Asst. Prof. Dr. Sakchai Tangwannawit, KMUTNB  
Assistant Secretary: Dr. Watchareewan Jitsakul, KMUTNB  
Assistant Secretary: Asst. Prof. Sutat Gammanee

## **STEERING COMMITTEE**

Asst. Prof. Dr. Photjanee Sukchovna, KRU

Asst. Prof. Dr. Pramote Luenam, NIDA

Asst. Prof. Coosak Kamonkhantithorn, RMUTK

Asst. Prof. Sasitorn Kaewman, MSU

Asst. Prof. Suwat Tachaphetpiboo, PBRU

Asst. Prof. Pongpith Tuenpusa, RMUTT

Dr. Wirat Pinkaew, NPRU

## **TECHNICAL PROGRAM COMMITTEE**

Akkarat Boonyapalanant, KMUTNB	Panana Tangwannawit, PCRU
Anusorn Bunteong, UBU	Panida Songram, MSU
Chakadkit Thaenchaikun, PSU	Paranya Palawisut, NPRU
Chatchawin Namman, UBU	Pattaraporn Warintarawej, PSU
Choochart Haruechaiyasak, NECTEC	Phayung Meesad, KMUTNB
Gangarn Somprasertsri, MSU	Pichate Kunakornvong, RMUTT
Jantima Polpinij, MSU	Pilapan Phonarin, RMUTK
Jaree Thongkam, MSU	Piyawad Kasabai, MSU
Jarut Busarathid, PBRU	Pongpisit Wuttidittachotti, KMUTNB
Jeerasak Numpradit, KMUTNB	Pramote Kuacharoen, NIDA
Jirawat Thaenthong, PSU	Pudsadee Boonrawd, KMUTNB
Kanchana Viriyapant, KMUTNB	Rungtip Charoensak, MSU
Kanjana Laosen, PSU	Sakchai Tangwannawit, KMUTNB
Komsan Kanjanasit, PSU	Sarawut Plaengsorn, PBRU
Kriengsak Treeprapin, UBU	Sompong Valuvanathorn, UBU
Kwankamon Dittakan, PSU	Sucha Smanchat, KMUTNB
Laddawan Champa, KRU	Suchart Khummanee, MSU
Maleerat Maliyaem, KMUTNB	Sukuma Uamcharoen, PBRU
Monvorath Phongpaibul, TU	Sunantha Sodsee, KMUTNB
Nadh Ditcharoen, UBU	Supawadee Hiranpongsin, UBU
Nalinpat Porrawatpreyakorn, KMUTNB	Supoj Hengpraproh, NPRU
Narong Phothi, KMUTNB	Sutat Gammanee, KRU
Narong Punkong, KRU	Sutep Tongngam, NIDA
Nathaporn Utakrit, KMUTNB	Suwat Tachaphetpiboon, PBRU
Nattapong Tongtep, PSU	Suwich Tirakoat, MSU
Nattavee Utakrit, KMUTNB	Tanapon Jensuttiwetchakul, KMUTNB
Nilubon Tongchai, KRU	Thanawat Yochanang, KMUTNB
Nongluk Promthong, RMUTT	Thitirat Siriborvornratanakul, NIDA
Olarik Surinta, MSU	Tongpool Heeptaisong, KMUTNB

## **TECHNICAL PROGRAM COMMITTEE**

Uraivan Inyaem, RMUT

Watchareewan Jitsakul, KMUTNB

Vatinee Nuijian, KMUTNB

Weena Janratchakool, RMUTT

Wanida Saetang, PSU

Wongkot Sriurai, UBU



## **TECHNICAL PROGRAM CONTENTS**

## Program Schedule

<b>Thursday 19<sup>th</sup> May 2022</b>	
08:00 – 09:00	Starting Virtual Conference System
09:00 – 09:15	Report by <b>Assistant Professor Dr. Dr.-Ing. Sunantha Sodsee</b> , <i>Dean, Faculty of Information Technology and Digital Innovation</i> <i>King Mongkut's University of Technology North Bangkok</i>
09.15-09.30	Welcome Speech by <b>Assistant Professor Dr. PhotJanee Sukchaona</b> , <i>President of Kanchanaburi Rajabhat University</i>
09:30 – 09:45	Opening Ceremony by <b>Professor. Dr.-Ing. habil. Suchart Siengchin</b> , <i>President of King Mongkut's University of Technology North Bangkok</i>
09:45 – 11:00	Invited Keynote Speech by: <b>Assoc. Prof. Dr. Guidong Zhang</b> , <ul style="list-style-type: none"> <li>• School of Automation, Guangdong University of Technology, China.</li> </ul>
11:00 – 12:00	Invited Keynote Speech by: <b>Prof. Dr. David Abramson</b> , <ul style="list-style-type: none"> <li>• Director, Research Computing Centre, The University of Queensland, Australia, Professor of Computer Science, FACM, FIEEE, FTSE, FACE.</li> </ul>
12:00 – 13:00	<i>Lunch</i>
13:00 – 17:00	<b>Paper Presentation</b>

<b>Thursday 19<sup>th</sup> May 2022</b>		
<b>ROOM: I (Sensor, Networking, Signal Processing, Image Processing and Related fields)</b>		<b>Page</b>
13:00 – 13:20 NCCIT2022-081	<b>Cyber Threat Analysis and Monitoring System with Random Forest and Gradient Boosted Tree</b> <i>Watcharakrit Lertthamphaiboon and Pongsarun Boonyopakorn</i>	1
13:20 – 13:40 NCCIT2022-082	<b>Load Balancing by Neighbour Node Discovery for RPL Routing Protocol in Wireless Sensor Networks</b> <i>Suttitee Jathuphornpaserd and Suphakit Awiphan</i>	7
13:40 – 14:00 NCCIT2022-030	<b>Extracting Key Features of Metaphase G-Band Chromosome Image for Counting Enhancement for Initial Genetic Diseases Diagnosis</b> <i>Bussayamas Pakkeetong, Anurak Yutthanawa and Pichet Wayalun</i>	13
14:00 – 14:20 NCCIT2022-100	<b>Convolutional Neural Network for Image Classification of Crown Flowers</b> <i>Kewalin Khamnipat and On-Uma Pramote</i>	19
14:20 – 14:40 NCCIT2022-023	<b>Contactless Water Dispenser Based on Internet of Things</b> <i>Wuttachai Chumyen and Thepparit Banditwattanawong</i>	25
14:40 – 15:00 NCCIT2022-035	<b>Indoor Mobile Performance Improvement by Inter-Band Carrier Aggregation</b> <i>Pakorn Kaewkhampaeng and Suwat Pattaramalai</i>	31
15:00 – 15:20	<i>Break</i>	
15:20 – 15:40 NCCIT2022-085	<b>The Assessment of an Organization's Cyber Security Awareness by Using a Phishing Simulation</b> <i>Preecha Harnyuenyongsakul and Pongsarun Boonyopakorn</i>	37
15:40 – 16:00 NCCIT2022-019	<b>The Impact of Perceived Security Risk and Privacy Risk on the Acceptance of IoT Technologies</b> <i>Maruchai Wattanarungson and Pramote Luenam</i>	43
16:00 – 16:20 NCCIT2022-067	<b>Generating an Air Force Intrusion Detection Dataset to Compare Between Standardized Datasets</b> <i>Sappanyou Chukaew, Prasong Praneetpolgrang and Payap Sirinam</i>	49
16:20 – 16:40 NCCIT2022-109	<b>Continuous Integration and Continuous Deployment on Cloud Computing Using Gitlab</b> <i>Thanthicha Nimmuan and Pudsadee Boonrawd</i>	55
	<i>End Session</i>	

<b>Thursday 19<sup>th</sup> May 2022</b>		
<b>ROOM: II (Applications/ Information Systems, Management Information System and Related Fields)</b>		<b>Page</b>
13:00 – 13:20 NCCIT2022-010	<b>Diagnosis System of Orchid Disease using Deep Learning on Mobile Application</b> <i>Nattavadee Hongboonmee and Kawinthip Thumchon</i>	61
13:20 – 13:40 NCCIT2022-089	<b>Project Management Information System According to the Term of Reference by Using the Theory of Task Priorities for Information Consulting Service</b> <i>Taechathat Sirimai and Nattavee Utakrit</i>	67
13:40 – 14:00 NCCIT2022-006	<b>Wireless Network Management for Green University Case Study Air Control System and Semi-Smart Bin System of Walailak University</b> <i>Satit Khwanchum</i>	73
14:00 – 14:20 NCCIT2022-018	<b>Simultaneous Multi-facial Recognition Software-as-a-Service for Class Attendance Checking</b> <i>Arisa Vaichiaranai and Thepparit Banditwattanawong</i>	80
14:20 – 14:40 NCCIT2022-026	<b>An Analysis of Game-Based Learning Materials in Java Programming Language to Enhance Learning Achievement by Using User Experience</b> <i>Phattaradon Ekthammabordee and Apichaya Nimkoompai</i>	87
14:40 – 15:00 NCCIT2022-027	<b>An Efficiency Tuning for Registration and Educational Statistic Application with MongoDB Replication</b> <i>Wiroj Tamboonlertchai, Nutsurang Dumlakor and Chumpol Mokrat</i>	93
15:00 – 15:20	<i>Break</i>	
15:20 – 15:40 NCCIT2022-091	<b>A Developing of IT Service Management System</b> <i>Wuttinun Chaipromnimit and Nattavee Utakrit</i>	99
15:40 – 16:00 NCCIT2022-107	<b>The Development of Asset Trading Platform Using Non-Fungible Token</b> <i>Jidapa Dumrongsombut and Pudsadee Boonrawd</i>	105
16:00 – 16:20 NCCIT2022-012	<b>Indoor Positioning System by Analyzing Wireless of Network Using Machine Learning</b> <i>Nattaphon Chaiwong and Atikhom Siri</i>	111
	<i>End Session</i>	

<b>Thursday 19<sup>th</sup> May 2022</b>		
<b>ROOM: III (Data Mining, Text Mining and Related Fields)</b>		<b>Page</b>
13:00 – 13:20 NCCIT2022-001	<b>Analysis of Image-Based Data to Track Rainfall and Forecast Flood Risk Areas</b> <i>Kanyarat Charoenpru and Prattana Deeprasertkul</i>	117
13:20 – 13:40 NCCIT2022-055	<b>Imbalanced Data Problem Solving in Predicting Accidental Deaths</b> <i>Thaninthorn Whasphutthisit and Watchareewan Jitsakul</i>	122
13:40 – 14:00 NCCIT2022-072	<b>Forecasting the Fraud Risk of Electronic Card Transactions</b> <i>Sirepong Padit and Watchareewan Jitsakul</i>	128
14:00 – 14:20 NCCIT2022-074	<b>The Development of Business Intelligence for Analyzing Data from the Fault Reporting System Using Data Mining Techniques</b> <i>Jirarat Srichok and Montean Rattanasiriwongwut</i>	134
14:20 – 14:40 NCCIT2022-022	<b>Thai Language Level Analysis System Using Text Mining Technique</b> <i>Sarin Songkarin, Apisit Kraiyaso and Warunya Wunnasri</i>	140
14:40 – 15:00 NCCIT2022-092	<b>Comparison of Time Series Analysis Techniques for Forecasting Facebook Ad Values</b> <i>Sujanya Tubtong and Watchareewan Jitsakul</i>	146
15:00 – 15:20	<i>Break</i>	
15:20 – 15:40 NCCIT2022-098	<b>Sales Data Analysis of Lighting Products Using Clustering Technique</b> <i>Suwapat Janmaitree and Sakchai Tangwannawit</i>	152
15:40 – 16:00 NCCIT2022-108	<b>Road Accidents Analytics System Using Business Intelligence</b> <i>Thanisa Sungsaeng and Pudsadee Boonrawd</i>	158
16:00 – 16:20 NCCIT2022-052	<b>Enhancement Data Analytics for Medical Diagnosis</b> <i>Suthatip Parittapongsachai and Phayung Meesad</i>	164
16:20 – 16:40 NCCIT2022-060	<b>Short and Long-Term Stock Price Forecasting with Deep Learning</b> <i>Sutasorn Yumak and Phayung Meesad</i>	171
	<i>End Session</i>	

<b>Thursday 19<sup>th</sup> May 2022</b>		
<b>ROOM: IV (Artificial Intelligent, Text Mining, Natural Language Processing and Related Fields)</b>		<b>Page</b>
13:00 – 13:20 NCCIT2022-002	<b>A Comparison of Machine Learning Models for COVID-19 Patients Screening</b> <i>Siwakorn Banluesapy and Waraporn Jirapanthong</i>	178
13:20 – 13:40 NCCIT2022-013	<b>Thai Sentiment Analysis about Education in Thailand on Twitter</b> <i>Wipada Silarach, Phithayarat Phochompu and Thanaphon Tangchoopong</i>	184
13:40 – 14:00 NCCIT2022-093	<b>Forecasting Home Loan Refinance Out Using Machine Learning</b> <i>Pongpan Kaeduang and Kanchana Viriyapant</i>	190
14:00 – 14:20 NCCIT2022-066	<b>Integrating Cyber Dataset for Cyber Intrusion Prediction Model</b> <i>Somboon Udnan, Prasong Praneetaphonkrang and Payap Sirinam</i>	196
14:20 – 14:40 NCCIT2022-068	<b>Face Detection Systems for Effective Facial Recognition with Artificial Intelligence Using Closed Circuit Television Camera</b> <i>Apisit Seelsat, Prasong Praneetpolgrang and Pahron Sanguanbhokai</i>	202
14:40 – 15:00 NCCIT2022-050	<b>Comparison of Deep Learning Models for Morphological Image Identification of Wild Mushrooms</b> <i>Wacharaphol Ketwongsa and Urachart Kokaew</i>	208
15:00 – 15:20	<i>Break</i>	
15:20 – 15:40 NCCIT2022-088	<b>Sentiment Analysis of Twitter Comments About COVID-19 Using Text Mining Techniques</b> <i>Natarda deearom and Tanapon Jentsuttiwetchakul</i>	214
15:40 – 16:00 NCCIT2022-090	<b>Classification of Information Problems from Text</b> <i>Thanakorn Liwveha and Nattavee Utakrit</i>	220
16:00 – 16:20 NCCIT2022-029	<b>Information Extraction with Knowledge Graph for Hotel Business in Thailand</b> <i>Tantikorn Poonyakariyakorn and Thanachart Ritbumroong</i>	226
16:20 – 16:40 NCCIT2022-073	<b>Sentiment Classification of Electric Vehicle Using Text Mining</b> <i>Damrongpol Suwanchatree and Watchareewan Jitsakul</i>	235
	<i>End Session</i>	

<b>Thursday 19<sup>th</sup> May 2022</b>		
<b>ROOM: V (Factor Analysis and Related Fields)</b>		<b>Page</b>
13:00 – 13:20 NCCIT2022-045	<b>Factors Influencing Customer Engagement in Using Facebook Messenger ChatBot Services</b> <i>Supawan Pongsupawan and Pramote Luenam</i>	241
13:20 – 13:40 NCCIT2022-033	<b>Factors Influencing the Intention to Use Robotic Process Automation Technology</b> <i>Chiraporn Thepthong and Tanapon Jentsuttiwetchakul</i>	248
13:40 – 14:00 NCCIT2022-034	<b>Factors Influencing Patient Tracking Application Acceptance</b> <i>Nawakarn Ninprawit and Tanapon Jentsuttiwetchakul</i>	254
14:00 – 14:20 NCCIT2022-043	<b>Parasocial Factors and Their Impact to the Intention to Follow Influencers on TikTok</b> <i>Jamnapa Yodho and Pramote Luenam</i>	260
14:20 – 14:40 NCCIT2022-061	<b>An Analysis of Factors Influencing Cryptocurrency Investments</b> <i>Anongwanneeya Suayroop and Montean Rattanasiriwongwut</i>	266
14:40 – 15:00 NCCIT2022-070	<b>Classification Factors Affecting Customer's Responsiveness Toward Service Quality Behavior During Covid-19 Pandemic</b> <i>Thanaporn Tepkila and Tanapon Jentsuttiwetchakul</i>	272
15:00 – 15:20	<i>Break</i>	
15:20 – 15:40 NCCIT2022-087	<b>A Study of Factors Affecting Consumers' Purchases on the E-Marketplace</b> <i>Sasima Tanapatpiboon and Tanapon Jentsuttiwetchakul</i>	278
15:40 – 16:00 NCCIT2022-042	<b>Stress Factors and Their Impact on Teaching Efficiency at Fundamental Education Level During the COVID-19 Pandemic</b> <i>Wirote Cherdawat and Pramote Luenam</i>	284
16:00 – 16:20 NCCIT2022-096	<b>An Factor Analysis of Components Influencing the Acceptance of e-Government (G2E) Information Technology Among Civil Servants Under the Ministry of Justice</b> <i>Krittamuk Lexseng and Montean Rattanasiriwongwut</i>	290
16:20 – 16:40 NCCIT2022-025	<b>An Analysis of the Dark Pattern of User Experience Online Shopping Application in Thailand</b> <i>Pumarin Tiangpanich and Apichaya Nimkoompai</i>	296
	<i>End Session</i>	

## **Program Schedule**

<b>Friday 20<sup>th</sup> May 2022</b>	
08:00 – 09:00	Starting Virtual Conference System
09:00 – 10.20	<b>Paper Presentation</b>
10.20 – 10.40	<i>Break</i>
10.40 – 13.00	<b>Paper Presentation</b>



<b>Friday 20<sup>th</sup> May 2022</b>		
<b>ROOM: I (Data Mining, Text Mining, Natural Language Processing and Related Fields)</b>		<b>Page</b>
09:00 – 09:20 NCCIT2022-095	<b>Predicting Longan Yield in Northern of Thailand Using Time Series Data</b> <i>Thanchanok Rangkaew, Eakasit Pacharawongsakda, Kantarakorn Jitharn, Sakolpuck Jearawattanasawadi and Chutimon Pakdeesirote</i>	302
09:20 – 09:40 NCCIT2022-051	<b>Business Intelligence System for Customer Consent of Objectives for Data Protection Officer</b> <i>Passorn Sukmon and Sakchai Tangwannawit</i>	308
09:40 – 10:00 NCCIT2022-015	<b>Decision Support System for Opening or Closing Branches Using Network Analysis</b> <i>Kornthip Whanchaaim, Tanasanee Phienthrakul and Konglit Hunchangsinh</i>	315
10:00 – 10:20 NCCIT2022-020	<b>Neural Network Training for Prediction of Side Effects After COVID-19 Vaccination Using Optimization Algorithms Inspired by Nature</b> <i>Supaporn Deankasai, Neungrutai Deemeewong, Phumin Duankhan and Khamron Sunat</i>	321
10:20 – 10:40	<i>Break</i>	
10:40 – 11:00 NCCIT2022-069	<b>The Development of Business Intelligence to Predict the Behaviors of the Working-Age Population Resulting in Chronic Non-Communicable Diseases</b> <i>Suphaluck Charoenthum and Montean Rattanasiriwongwut</i>	327
11:00 – 11:20 NCCIT2022-083	<b>Network Intrusions Analysis with Neural Network and K-Means</b> <i>Saksit Jantaraplin and Pongsarun Boonyopakorn</i>	334
11:20 – 11:40 NCCIT2022-080	<b>The Development of Decision Support System for Diabetes Risk Prediction Using Data Mining Technology</b> <i>Nattapong Laehte and Montean Rattanasiriwongwut</i>	340
11:40 – 12:00 NCCIT2022-084	<b>The Development of ChatBot Line to Provide Information about Coronavirus</b> <i>Kan Khemnakh and Nattavee Utakrit</i>	346
	<i>End Session</i>	

<b>Friday 20<sup>th</sup> May 2022</b>		
<b>ROOM: II (Data Mining and Related Fields)</b>		<b>Page</b>
09:00 – 09:20 NCCIT2022-102	<b>A Comparison of Prediction Models for Estimates the Cost of Treatment of Patients Undergoing Surgery</b> <i>Siriporn Mongkolchaiwanich, Wuntana Tientong and Sakchai Tangwannawit</i>	352
09:20 – 09:40 NCCIT2022-037	<b>Predictive Models for Auto Loan Approval using the Decision Tree Algorithm</b> <i>Supachanun Wanapu, Naris Mingmora, Janejeera Aksornpim, Ketsuda Chainok, Awissara Dokauboi and Thati Phiphatprueksa</i>	358
09:40 – 10:00 NCCIT2022-049	<b>Forecasting Model for Agricultural Export Using Data Mining Techniques</b> <i>Suphaporn Jamrurai and Sakchai Tangwannawit</i>	364
10:00 – 10:20 NCCIT2022-054	<b>Forecasting the Number Registration of the Land Officer Using Time Series Techniques</b> <i>Chatchadawan Chanprasert and Watchareewan Jitsakul</i>	370
10:20 – 10:40	<i>Break</i>	
10:40 – 11:00 NCCIT2022-062	<b>Prediction mai Growth Stock in the Securities Exchange of Thailand</b> <i>Sudthida Changpan and Tanapon Jensuttiwetchakul</i>	376
11:00 – 11:20 NCCIT2022-106	<b>Car Insurance Recommendation System Using Rules-Based Personalization with K-Means Clustering</b> <i>Peemai Keelakitpaibool and Pudsadee Boonrawd</i>	382
11:20 – 11:40 NCCIT2022-028	<b>Assessing the Risk of Polycystic Ovary Syndrome</b> <i>Kamonchat Deela, Pawitra Sariwong and Thanaphon Tangchoopong</i>	388
11:40 – 12:00 NCCIT2022-075	<b>Sales Forecasting Using Time Series Techniques</b> <i>Adithep Panwanichayakij and Watchareewan Jitsakul</i>	394
	<i>End Session</i>	

<b>Friday 20<sup>th</sup> May 2022</b>		
<b>ROOM: III (Applications, Information System, Mining and Related Fields)</b>		<b>Page</b>
09:00 – 09:20 NCCIT2022-053	<b>Application Attendance System Base 2FA and Location</b> <i>Natkamon Siripanurak and Tanapon Jentsuttiwetchakul</i>	400
09:20 – 09:40 NCCIT2022-078	<b>The Development of Automation System for Support Process of Purchasing and Accounting Payable</b> <i>Kontorn Inprasith and Sakchai Tangwannawit</i>	406
09:40 – 10:00 NCCIT2022-007	<b>Automatic Quiz Creation and Verification for the Volume and Surface Area of Three-dimensional Objects</b> <i>Rattanakorn Saeng-Aun and Phongthep Raxpakawong</i>	412
10:00 – 10:20 NCCIT2022-008	<b>An Incident Tracking System: A Case Study of Civil Court</b> <i>Non Chitchamnong and Waraporn Jirapanthong</i>	418
10:20 – 10:40	<i>Break</i>	
10:40 – 11:00 NCCIT2022-011	<b>Performance Testing of Converting Tools from PDF File to CSV Text File</b> <i>Phutares Suwannarat and Ponrudee Netisopakul</i>	424
11:00 – 11:20 NCCIT2022-063	<b>Text Matching Using BLEU Score for Voice-based Food Ordering System</b> <i>Kittamet Montheankul, Vanitchaporn Masomboon, Supasuda Sudprasert and Sopon Phumeechanya</i>	430
11:20 – 11:40 NCCIT2022-056	<b>Development of Model for Forecasting Business Credit Approvals Using Data Mining Techniques</b> <i>Karsiya Klintharvorn and Montean Rattanasiriwongwut</i>	436
	<i>End Session</i>	

<b>Friday 20<sup>th</sup> May 2022</b>		
<b>ROOM: IV (Data Mining, Text Mining, Natural Language Processing and Related Fields)</b>		<b>Page</b>
09:00 – 09:20 NCCIT2022-038	<b>Classification of Weapons Using Convolutional Neural Networks</b> <i>Eakarach Nintra, Surapan Airphaiboon and Somchat Jiriwibhakorn</i>	442
09:20 – 09:40 NCCIT2022-040	<b>A Study and Development of Bitmap Intersection Lookup Algorithm for Packet Classification</b> <i>Worachet Nimcharoen, Supakkavee Supho and Akharin Khunkitti</i>	448
09:40 – 10:00 NCCIT2022-057	<b>Business Intelligence System for Price Forecasting of Agricultural Product</b> <i>Phonpan Charoensin and Sakchai Tangwannawit</i>	454
10:00 – 10:20 NCCIT2022-065	<b>SAPPY: Emotion-Based Assistant Chatbot</b> <i>Natthaphat Kongthong, Vacharavat Kesbuntherng and Supannada Chotipant</i>	460
10:20 – 10:40	<i>Break</i>	
10:40 – 11:00 NCCIT2022-094	<b>Forecasting Sugar Cane Yield in Thailand with Machine Learning</b> <i>Teerawat Sinlaparatsamee and Kanchana Viriyapant</i>	467
11:00 – 11:20 NCCIT2022-077	<b>Business Intelligence System of Predicts Online Purchase Behavior for Support Marketing Plan</b> <i>Panatda Yeemasa and Sakchai Tangwannawit</i>	473
11:20 – 11:40 NCCIT2022-097	<b>Sentiment Analysis for Service Chatbot</b> <i>Thiti Tongumpun and Kanchana Viriyapant</i>	479
	<i>End Session</i>	

<b>Friday 20<sup>th</sup> May 2022</b>		
<b>ROOM: V (Applications, Information System, Management Information System and Related Fields)</b>		<b>Page</b>
09:00 – 09:20 NCCIT2022-086	<b>Development of Learning Media Topic: Taxonomy Animal Kingdoms</b> <i>Nathapoom Boonruangroj and Nattavee Utakrit</i>	485
09:20 – 09:40 NCCIT2022-024	<b>User Experience Design of Augmented Reality Teaching Materials for Basic Level Japanese Subjects to Increase Learning Achievement</b> <i>Niratchaporn Ounnimit and Apichaya Nimkoompai</i>	491
09:40 – 10:00 NCCIT2022-101	<b>Threatening Virus Game Development Using Virtual Reality</b> <i>Phuriphat Theerawut, Sutasinee Thovuttikul and Watcharee Jumpamule</i>	497
10:00 – 10:20 NCCIT2022-039	<b>Undergrad: IT KMITL Student Life Simulation Game</b> <i>Sunisa Chanbang, Suphakit Nachom and Supannada Chotipant</i>	503
10:20 – 10:40	<i>Break</i>	
10:40 – 11:00 NCCIT2022-076	<b>Applied an Artificial Intelligence with Chatbot's on Line Office Account</b> <i>Theerawat Patcham and Montean Rattanasiriwongwut</i>	509
11:00 – 11:20 NCCIT2022-017	<b>Data Privacy Management System on Cloud according to Personal Data Protection Act</b> <i>Kantapon Seshajoti and Thepparit Banditwattanawong</i>	516
11:20 – 11:40 NCCIT2022-079	<b>The Development of Chatbot Prototype on an Application for Problems Recommender of Female's Wellness</b> <i>Uraiwan Inyaem, Weena Janratchakool, Nattawadee Sukkasem and Suwimon Tuanthong</i>	523
	<i>End Session</i>	

## **REGULAR PAPERS**

# ระบบวิเคราะห์และตรวจสอบภัยคุกคามทางไซเบอร์ด้วย Random forest และ Gradient Boosted Tree Cyber Threat Analysis and Monitoring System with Random Forest and Gradient Boosted Tree

วัชรกฤษฎ์ เลิศธรรมไพบูลย์ (Watcharakrit Lertthamphaiboon)<sup>1</sup>

และพงศ์ศรีณย์ บุญโยปกรณ์ (Pongsarun Boonyopakorn)<sup>2</sup>

ภาควิชาการบริหารเครือข่ายดิจิทัลและความมั่นคงปลอดภัยสารสนเทศ

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307031858069@email.kmutnb.ac.th, <sup>2</sup>pongsarunb@kmutnb.ac.th

## บทคัดย่อ

บทความนี้มีวัตถุประสงค์เพื่อศึกษา สร้าง และเปรียบเทียบแบบจำลองวิเคราะห์ความเสี่ยงของภัยคุกคามทางไซเบอร์ เพื่อหาความสัมพันธ์ระหว่างแพ็คเกจของการสื่อสารข้อมูลเพื่อวิเคราะห์ความผิดปกติในระบบเครือข่ายโดยแยกแยะชนิดของการโจมตีประเภทต่าง ๆ โดยมีชุดข้อมูล KDDCUP-99 จำนวน 4,898,431 records และ UNSW-NB15 จำนวน 2,218,761 records นำมาใช้เป็นชุดฝึกข้อมูลเพื่อสร้างและเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองที่ใช้อัลกอริทึมการเรียนรู้แบบมีผู้สอน ทั้ง 2 อัลกอริทึม ได้แก่ Random Forest และ Gradient Boosted Tree จากการพิจารณาอัลกอริทึมพบว่าสำหรับชุดข้อมูล KDDCUP-99 อัลกอริทึม Gradient Boosted Tree สามารถวิเคราะห์ความเสี่ยงและชนิดของภัยคุกคามทางไซเบอร์ได้แม่นยำถึง 96.56% ในขณะที่ Random Forest มีความแม่นยำ 95.56% สำหรับชุดข้อมูล UNSW-NB15 อัลกอริทึม Gradient Boosted Tree สามารถวิเคราะห์ความเสี่ยงและชนิดของภัยคุกคามทางไซเบอร์ได้แม่นยำถึง 82.57% ในขณะที่ Random Forest มีความแม่นยำเพียง 78.46% ผู้วิจัยได้นำเสนอแบบจำลองอัลกอริทึมแบบผสมระหว่าง Gradient Boosted Tree และ Random Forest พบว่าจากชุดข้อมูล KDDCUP-99 สามารถวิเคราะห์ความเสี่ยงและชนิดของภัยคุกคามทางไซเบอร์ได้แม่นยำถึง 99.89% และชุดข้อมูล UNSW-NB15 ได้แม่นยำถึง

88.64% ผู้พัฒนาจึงได้เลือกอัลกอริทึมแบบผสมระหว่าง Gradient Boosted Tree และ Random Forest ในการพัฒนาเว็บแอปพลิเคชันสำหรับวิเคราะห์ความเสี่ยงและชนิดของภัยคุกคามทางไซเบอร์  
คำสำคัญ: การทำนายความเสี่ยง การโจมตีทางไซเบอร์ การเรียนรู้ของเครื่อง

## Abstract

The objective of this paper is to study, create, and compare disease risk analysis models in the Cyber Threat Attack to find the correlation between the packet data and Cyber Threat To determine an algorithm that is appropriate for each attack type, 4,898,431 records of data packet KDDCUP-99 and 2,218,761 records of data packet UNSW-NB15 were used as a training data set to generate models to be used for comparing performance of the three machine learning algorithms: Random Forest, Gradient Boosted Tree and Hybrid between Random Forest and Gradient Boosted Tree. For KDDCUP99 dataset, the experiment showed that Gradient Boosted Tree algorithm was able to analyze disease risk with accuracy of 96.56% and Random Forest could achieve 96.56%, while Hybrid algorithm could achieve 99.89%. For UNSW-NB15 Gradient Boosted Tree algorithm was able to analyze disease risk of

82.57% and Random Forest could achieve 78.46%, while Hybrid algorithm accurate to 88.64%. Finally, Hybrid algorithm was used to develop web applications for analyzing Cyber Threat and attack type.

**Keyword:** Predicting Cyber Risks, Cyber Attacks, Machine Learning

## 1. บทนำ

ในปัจจุบันความก้าวหน้าทางเทคโนโลยีเข้ามามีอิทธิพลต่อพฤติกรรมของผู้คนในปัจจุบันมากยิ่งขึ้นทั้งยังสร้างความสะดวกสบายในด้านต่าง ๆ เช่น การค้นหาข้อมูล การติดต่อสื่อสาร ไปจนถึงการพัฒนาธุรกิจผ่านช่องทางต่างๆ ในทางกลับกันก็เป็นช่องทางหนึ่งที่เปิดโอกาสให้ผู้ที่ไม่หวังดีเข้ามาเพื่อทำการขโมยข้อมูลที่มีความสำคัญหรือหาประโยชน์จากเราโดยไม่ได้รับอนุญาตได้เฉกเช่นเดียวกัน โดยมีผลกระทบทั้งทางตรงและทางอ้อมต่อผู้ใช้งานได้ เช่นผลกระทบทางตรงจากการถูกภัยคุกคามทางไซเบอร์โจมตีจนไม่สามารถให้บริการต่อได้ (DENIAL OF SERVICE ATTACK) ส่งผลเสียต่อองค์กรโดยตรงก่อให้เกิดความเสียหายต่อรายได้ของทางบริษัทและจากภัยคุกคามข้างต้นยังส่งผลทางอ้อมคือการเสื่อมเสียชื่อเสียง ความเชื่อมั่นต่อองค์กร ซึ่งอาจส่งผลกระทบต่อการดำเนินงานในภายภาคหน้านั้นเอง

จากที่กล่าวมามีข้อมูลที่น่าสนใจจาก “แคสเปอร์สกี” เปิดเผยในปี 2563 มีการโจมตีผ่านเว็บไซต์ในประเทศไทยจำนวน 20 ล้านครั้ง โดยมัลแวร์นิยมแฝงตัวผ่านเว็บไซต์ฟิชชิ่ง โฆษณาออนไลน์ รวมไปถึงการดาวน์โหลดโปรแกรม ไฟล์แนบต่าง ๆ จากอีเมลแคสเปอร์สกี ซิเคียวริตี้เน็ตเวิร์ค ระบุว่าปี 2563 สามารถสกัดกั้นภัยไซเบอร์ผ่านเว็บที่คุกคามคอมพิวเตอร์ของผู้ใช้ในประเทศไทยจำนวนทั้งสิ้น 20,598,223 รายการ

จากข้างต้นผู้วิจัยมีความสนใจใช้เทคโนโลยีทางด้าน Machine Learning ในการเรียนรู้ความสัมพันธ์ลักษณะของการโจมตีทางไซเบอร์ในรูปแบบต่าง ๆ โดยพัฒนาระบบการตรวจสอบจับและตรวจสอบความผิดปกติของเครือข่ายหรือระบบที่ถูกใช้งานอยู่

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 Machine Learning

การเรียนรู้ของเครื่อง (Machine Learning) [1] เป็นสาขาหนึ่งในด้านการสร้างและพัฒนาปัญญาประดิษฐ์ (Artificial Intelligence) มีจุดประสงค์เพื่อการสร้างและพัฒนาทำให้ระบบคอมพิวเตอร์สามารถประมวลผลจากข้อมูลต่าง ๆ คาดการณ์ ทำนายค่าและตัดสินใจแก้ปัญหาได้ด้วยตัวเองผ่านชุดข้อมูลฝึกฝนที่ได้ใส่เข้าไปเพื่อเป็นข้อมูลในการเรียนรู้

โดยในปี 1959 Arthur Samuel (อาเธอร์ ซามูเอล) [2] นักวิทยาศาสตร์คอมพิวเตอร์ชาวอเมริกันผู้เชี่ยวชาญด้านเกมส์คอมพิวเตอร์ ปัญญาประดิษฐ์ และการเรียนรู้ของเครื่องได้ทำการบินนิยามคำว่า “Machine Learning” ขึ้นมาเป็นครั้งแรกโดยได้ให้นิยาม Machine Learning ว่าเป็นศาสตร์หนึ่งที่มีมุ่งเน้นศึกษาการเพื่อให้คอมพิวเตอร์มีความสามารถที่จะเรียนรู้ได้โดยตัวเองโดยไม่ต้องเขียนโปรแกรมกำกับไว้อย่างชัดเจนในทุก ๆ การกระทำ

การเรียนรู้ของเครื่องถือว่ามี ความเกี่ยวข้องทางสถิติศาสตร์เป็นอย่างมาก เนื่องจากทั้งสองสาขามีการศึกษาการวิเคราะห์ข้อมูลเพื่อทำนายและคาดการณ์ผลลัพธ์ นอกจากนี้ยังนำความรู้ทางสาขาการเพิ่มประสิทธิภาพทางคณิตศาสตร์ (Mathematical Optimization) มาประยุกต์ใช้

#### 2.1.1 องค์ประกอบของการเรียนรู้ของเครื่อง

##### 2.1.1.1 ชุดข้อมูลฝึกฝน (Training Set)

เป็นชุดข้อมูลสำหรับฝึกฝนแบบจำลองโดยให้ข้อมูลสำหรับเรียนรู้แก่แบบจำลองเพื่อให้แบบจำลองสามารถคาดการณ์ผลลัพธ์ออกมาเป็นไปตามชุดข้อมูลต้นฉบับได้มากที่สุด

##### 2.1.1.2 ชุดข้อมูลทดสอบ (Test Set)

ชุดข้อมูลสำหรับทดสอบความถูกต้องของแบบจำลองสร้างขึ้น

##### 2.1.1.3 Label, Non-Label หรือ Class

ผลลัพธ์ของข้อมูลฝึกฝนที่กำหนดขึ้น Machine Learning จะต้องทำการกำหนด Label เพื่อวัตถุประสงค์ในการทำนาย คาดการณ์ผลลัพธ์ที่ต้องการโดย Label จะใช้กับการเรียนรู้ของเครื่องแบบมีผู้ช่วยสอน (Supervised)



และ non-Label ใช้กับการเรียนรู้แบบไม่มีผู้ช่วยสอน (Unsupervised) โดยอาศัยการแบ่งกลุ่ม

#### 2.1.1.4 Feature

คุณลักษณะเด่นที่ปรากฏอยู่ภายในข้อมูลที่มีผลต่อการวิเคราะห์ผลลัพธ์ เช่น ข้อมูลการโจมตีแบบ Denial of Service จะมีการรับส่งข้อมูลจำนวนมากในระยะเวลาสั้น ๆ ทำให้การให้บริการต้องหยุดชะงัก

#### 2.1.1.5 Machine Learning Algorithm / Machine Learning Model

การนำชุดข้อมูลฝึกฝน (Training Set) ภายในประกอบไปด้วย Class ซึ่งมีหรือไม่มี Label ก็ได้ โดยต้องมีการเลือกคุณลักษณะเด่น (Feature) ของข้อมูล จากนั้นเอาข้อมูลเหล่านั้นมาคำนวณผ่านหลักคณิตศาสตร์ และสถิติ โดยขั้นตอนนี้เราเรียกว่าการฝึกฝนแบบจำลอง (Train Model) เมื่อฝึกฝนเสร็จจะได้โมเดลสำหรับการทำนาย

### 2.2 Random Forest

เป็นอัลกอริทึมที่นำโครงสร้างต้นไม้ Decision Tree ในหลายๆแบบจำลองมาพิจารณาด้วยกัน [3] โดยแต่ละแบบจำลองจะใช้ชุดข้อมูลทดสอบเดียวกันจากการสุ่มเลือกเพียงคุณลักษณะมาพิจารณาเท่านั้น จากนั้นจึงนำผลลัพธ์ที่ได้จากทุกแบบจำลองมาตัดสินใจเป็นผลลัพธ์สุดท้าย ซึ่งโดยทั่วไปแต่ละแบบจำลองจะฝึกฝนแยกกันอย่างอิสระและได้ผลลัพธ์สุดท้ายจากการนำผลลัพธ์จากทุก ๆ แบบจำลองมาหาค่าเฉลี่ยหรือการคำนวณผลลัพธ์ที่มีความแม่นยำมากที่สุด อัลกอริทึมนี้เป็นอัลกอริทึมการเรียนรู้แบบมีผู้สอนที่ถือว่ามีประสิทธิภาพสูงเนื่องจากใช้หลายแบบจำลองในการตัดสินใจจึงได้ความแม่นยำที่สูงกว่า Decision Tree และนำไปใช้งานได้ทั้งกรณีการจำแนกประเภทและการถดถอยขั้นตอนอัลกอริทึมของ Random Forests สามารถทำได้ดังนี้ (กำหนด Decision tree จำนวน  $N$  แบบจำลอง)

2.2.1 สุ่มเลือกคุณลักษณะมาจำนวน  $K$  จากทั้งหมด  $M$  โดยที่  $K < M$

2.2.2 สร้างแบบจำลอง Decision tree จากคุณลักษณะที่สุ่มเลือก

2.2.3 ทำกระบวนการตามข้อ 2.2.1 และ 2.2.2 จนได้จำนวนแบบจำลอง Decision tree ตามที่กำหนด

### 2.3 Gradient Boosted Tree

วิธี Gradient Boosted Trees (GBT) คือวิธีที่มีพื้นฐาน[4] มาจาก Decision tree ซึ่งเป็นการปรับปรุงประสิทธิภาพของแบบจำลองใหม่ให้มีค่าความแม่นยำสูงขึ้น โดยการสุ่มสร้าง Decision tree หลายร้อยแบบจำลอง และประเมินผลแต่ละแบบจำลองจนกว่าจะได้ Decision tree ที่สมบูรณ์ และใช้วิธีการ Optimize Parameter เพื่อหาค่าพารามิเตอร์ที่ดีที่สุดของแต่ละแบบจำลอง ด้วยวิธีการ Evolutionary Optimization โดยมีพื้นฐานมาจาก Genetic Algorithms เป็นวิธีการหาคำตอบจากการเลียนแบบพันธุกรรมและกระบวนการวิวัฒนาการ (Evolution) ของสิ่งมีชีวิตโดยมีหลักการในการหาปัจจัยที่ทำให้คำตอบดีขึ้นและทำการไขว้คำตอบเพื่อนำไปสู่ชุดปัจจัยที่ทำให้คำตอบดีขึ้นกว่าเดิม อีกทั้งทำการประเมินประสิทธิภาพแบบจำลองโดยการทำ K-fold Cross-Validation ซึ่งก็คือการแบ่งข้อมูลเป็น  $k$  ส่วน โดยในแต่ละส่วนของข้อมูลจะถูกแบ่งเป็น Training set และ Test set และเมื่อทดสอบจนครบ  $k$  ครั้งแล้วจะมีการนำเอาค่า Accuracy ของแต่ละครั้งมาเฉลี่ยเพื่อให้ได้ค่า Accuracy สุดท้าย และหากผลลัพธ์มีค่า Accuracy สูงก็หมายความว่า แบบจำลองนั้นมีประสิทธิภาพหรือความแม่นยำสูงนั่นเองโดยทั่วไปค่า  $K$  ที่ใช้ในการทดลองจะใช้เป็น 10 เนื่องจากเป็นการแบ่งข้อมูลเป็น Training set เป็นร้อยละ 90 และ Test set เป็นร้อยละ 10 และทำการสลับการสร้างแบบจำลองเพื่อแน่ใจว่าข้อมูลทุกส่วนถูกนำมาสร้างและทดสอบด้วยความน่าจะเป็นเท่า ๆ กัน

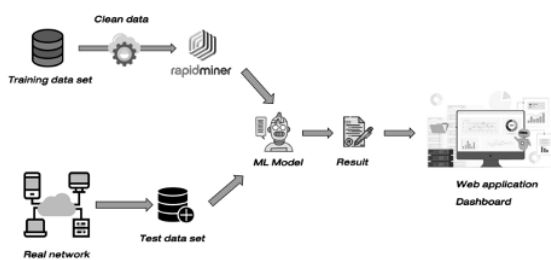
### 2.4 Hybrid Model

ผู้วิจัยได้นำอัลกอริทึม Random forest ร่วมกันกับ Gradient Boosted Tree เพื่อสร้างเป็น โมเดลแบบผสม โดยเป็นการทำงานแบบคู่ขนานเพื่อรวบรวมข้อดีของทั้งสองอัลกอริทึมในการคำนวณและทำนายผลโดยจากข้อดีของ Random forest ในด้านการคำนวณจากหลาย ๆ Tree และเลือกมาหนึ่งเพื่อผลลัพธ์ที่แม่นยำที่สุดหลังจากนั้นจะนำไปควมรวมกับข้อดีของ Gradient Boosted Tree โดยการ Optimize Parameter เพื่อหาค่าพารามิเตอร์ที่ดีที่สุดของแต่ละแบบจำลองอีกครั้ง จึงได้เป็นโมเดลที่มีความแม่นยำสูงกว่าการใช้งานโมเดลใดโมเดลหนึ่งแบบเดียว

### 3. วิธีการดำเนินงาน

ผู้ทำการวิจัยเริ่มต้นจากการศึกษาชนิดและประเภทของภัยคุกคามทางไซเบอร์ ความรุนแรง ผลกระทบและโอกาสในการเกิด รวมถึงการตรวจสอบลักษณะการทำงานของ Machine Learning อัลกอริทึมประเภทต่าง ๆ ว่ามีข้อดีข้อเสียและปัจจัยในการเพิ่มประสิทธิภาพของแบบจำลองอย่างไร เพื่อนำข้อมูลจากสองส่วนมาสรุป วิเคราะห์และออกแบบระบบวิเคราะห์ความเสี่ยงและการโจมตีประเภทต่าง ๆ ได้อย่างแม่นยำ

1) ทำการกำหนดขอบเขตในการวิจัยจากข้อมูลที่รวบรวมมาจากข้างต้น กำหนดขอบเขตในการวิจัยพัฒนาระบบตรวจสอบภัยคุกคามทางไซเบอร์ โดยแบ่งเป็นชุดข้อมูลจาก 2 แหล่งข้อมูล จากนั้นนำผลความแม่นยำมาวิเคราะห์เพื่อเลือก Machine Learning อัลกอริทึมที่มีความแม่นยำสูงที่สุดรวมไปถึงการผสมผสาน Machine Learning อัลกอริทึม เพื่อผลลัพธ์ที่ดีที่สุด จากนั้นจึงนำผลทำนายแสดงผ่านหน้าเว็บไซต์เพื่อแสดงผลให้ง่ายแก่การทำความเข้าใจของผู้ใช้งาน นอกจากนี้ผู้วิจัยได้กำหนดกระบวนการวิจัยดังแสดงในแผนภาพเพื่อให้การวิจัยเป็นไปอย่างมีประสิทธิภาพ สอดคล้องกับขอบเขตและวัตถุประสงค์



ภาพที่ 1: แสดงผลกระบวนการวิจัยเพื่อสร้างโมเดล

2) ใช้งานโปรแกรม RapidMiner Studio สำหรับขั้นตอนการตรวจสอบและประเมินประสิทธิภาพ โดยเริ่มจากการนำข้อมูลชุดฝึกฝนเข้าสู่โปรแกรมโดยกำหนดคุณลักษณะที่ต้องการทำนายเป็น Label ในที่นี้ระบุเป็นในส่วนของคุณลักษณะการโจมตีทางไซเบอร์โดยแบ่งชุดข้อมูลเป็น 2 ส่วนในส่วนแรกนำไปฝึกสอนเพื่อให้แบบจำลองเกิดการเรียนรู้ ส่วนที่สองแยกไว้เพื่อทำการทดสอบความแม่นยำของแบบจำลองอัตราส่วน 70:30

กล่าวคือชุดข้อมูลฝึกฝนร้อยละ 70 และข้อมูลชุดทดสอบประสิทธิภาพร้อยละ 30

3) ทำการทดสอบจำครบบทั้งสามโมเดล ประกอบไปด้วย Gradient Boosted Trees, Random Forest และ Hybrid อัลกอริทึม

4) ประเมินประสิทธิภาพ จะเป็นการประเมินประสิทธิภาพความแม่นยำในการตรวจสอบชนิดของภัยคุกคามทางไซเบอร์ในแต่ละอัลกอริทึม ดังที่กล่าวมา

5) ทำการพัฒนาเว็บแอปพลิเคชันด้วยภาษา PHP และ Python เพื่อทำการแสดงผลให้งานต่อการทำงานเข้าใจและพัฒนาต่อไป

แสดงผลพีเจอร์ที่นำมาคำนวณจากชุดข้อมูล KDDCUP99 ดังตารางที่ 1 และข้อมูลเพิ่มเติมของชุดข้อมูล UNSW-NB15 ดังตารางที่ 2 โดยทั้งสองชุดมีค่าพารามิเตอร์ที่ใช้งานแตกต่างกัน

ตารางที่ 1: แสดงผลพีเจอร์ที่สำหรับชุดข้อมูล KDDCUP99

Feature	Usage
duration	✓
protocol type	✓
service	✓
flag	✓
Src_bytes	✓
Dst_bytes	✓
Land	✓
Wrong	✓
Urgent	✓
hot	✓
num failed_login	✓
logged in	✓
num_compromised	✓
root_shell	✓
su_attempted	✓
num_root	✓
num_file_creations	✓
num_shells	✓
num_access_files	✓
num_outbound_cmds	✓
is_host_login	✓
is_guest_login	✓
count	✓
srv_count	✓
serror_rate	✓
srv_serror_rate	✓
rerror_rate	✓
srv_rerror_rate	✓
same_srv_rate	✓
diff_srv_rate	✓
srv_diff_host_rate	✓

Feature	Usage
dst_host_count	✓
dst_host_srv_count	✓
dst_host_same_srv_rate	✓
dst_host_diff_srv_rate	✓
dst_host_same_src_port_rate	✓
dst_host_srv_diff_host_rate	✓
dst_host_serror_rate	✓
dst_host_srv_serror_rate	✓
dst_host_rerror_rate	✓
dst_host_srv_rerror_rate	✓
attack_type	✓

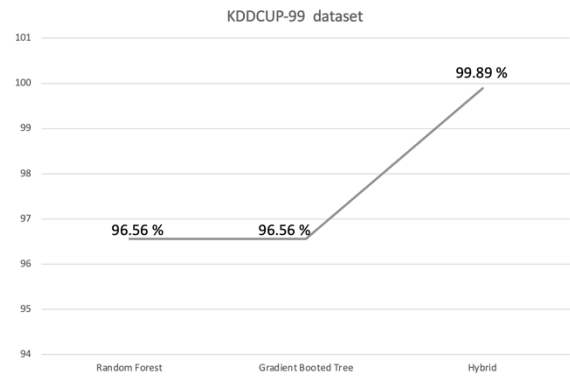
ตารางที่ 2: แสดงผลฟีเจอร์เพิ่มเติมสำหรับชุดข้อมูล UNSW-NB15

Feature	Usage
attack_cat	✓
dur	✓
proto	✓
service	✓
state	✓
spkts	✓
dpkts	✓
rate	✓
sttl	✓
dttl	✓
sload	✓
dload	✓
response_body_len	✓
ct_srv_src	✓
ct_state_ttl	✓
ct_src_ltm	✓
ct_src_dport_ltm	✓
ct_dst_dport_ltm	✓
Is_ftp_login	✓
Is_sm_ips_ports	✓

#### 4. ผลการดำเนินงาน

##### 4.1 ผลการทดสอบสำหรับชุดข้อมูล KDDCUP99

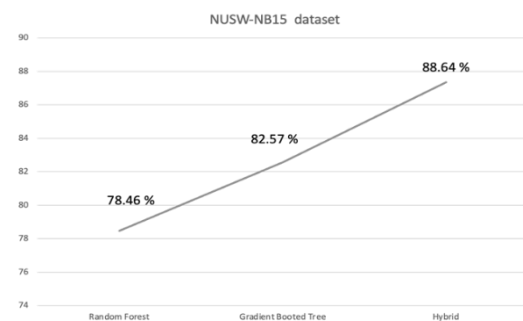
จากการทดสอบพบว่าค่าความแม่นยำในการตรวจสอบภัยคุกคามทางไซเบอร์ด้วยอัลกอริทึมทั้งสามประกอบด้วย Gradient Boosted Trees, Random Forest และ Hybrid กับชุดข้อมูลฝึกฝน KDDCUP99 อัลกอริทึมแบบ Hybrid ให้ความแม่นยำสูงที่สุดที่ 99.89% ตามมาด้วย Random Forest และ Gradient Boosted Trees ที่ 96.56% ดังภาพที่ 2



ภาพที่ 2: แสดงผลความแม่นยำชุดข้อมูลฝึกฝน KDDCUP99

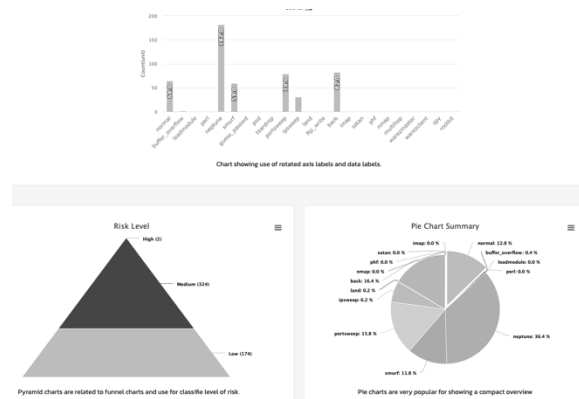
##### 4.2 ผลการทดสอบสำหรับชุดข้อมูล UNSW-NB15

จากการทดสอบพบว่าค่าความแม่นยำในการตรวจสอบภัยคุกคามทางไซเบอร์ด้วยอัลกอริทึมทั้งสามประกอบด้วย Gradient Boosted Trees, Random Forest และ Hybrid กับชุดข้อมูลฝึกฝน UNSW-NB15 อัลกอริทึมแบบ Hybrid ให้ความแม่นยำสูงที่สุดที่ 88.64% ตามมาด้วย Gradient Boosted Trees 82.57 และ Random Forest ที่ 78.46% ดังภาพที่ 3



ภาพที่ 3: แสดงผลความแม่นยำชุดข้อมูลฝึกฝน UNSW- NB15

จากผลการทดสอบผู้วิจัยได้นำผลลัพธ์ที่ได้รับหลังจากประมวลผลแล้วมาแสดงในรูปแบบของเว็บแอปพลิเคชันแสดงดังภาพที่ 4



ภาพที่ 4: แสดงผลการทดสอบผ่านเว็บแอปพลิเคชัน

## 5. สรุป

งานวิจัยนี้ผู้วิจัยได้ทำการตรวจสอบการโจมตีทางด้านไซเบอร์ชนิดต่าง ๆ ผ่านชุดข้อมูลฝึกฝนประกอบด้วย KDDCUP99 และ UNSW-NB15 ด้วยอัลกอริทึมสามชนิดประกอบด้วย Random Forest, Gradient Boosted Tree และแบบ Hybrid

ผลการทดสอบพบว่าการทำงานของ Gradient Boosted Tree อัลกอริทึมมีความแม่นยำสำหรับชุดข้อมูลทั้งสองสูงกว่า Random Forest อัลกอริทึมและการทำงานผสานระหว่างสองอัลกอริทึม Random Forest และ Gradient Boosted Tree กล่าวคือการทำงานของ Hybrid อัลกอริทึมพบว่าจากการนำข้อดีของทั้งสองอัลกอริทึมมาควรรวมกันพบว่ามีความแม่นยำในการตรวจจับภัยคุกคามทางไซเบอร์ได้สูงที่สุดจากโมเดลที่กล่าวมาทั้งหมด

### เอกสารอ้างอิง

- [1] Mitchell, T. *Machine Learning*. McGraw Hill. 1997.
- [2] อาเธอร์ ลี ซามูเอล (อังกฤษ: Arthur Lee Samuel) Available Online at [https://th.wikipedia.org/wiki/อาเธอร์\\_ซามูเอล](https://th.wikipedia.org/wiki/อาเธอร์_ซามูเอล), สืบค้นเมื่อ 18 ธันวาคม 2564
- [3] Afroz Chakure (2019). *Random Forest Regression* Available Online at <https://medium.com/swlh/random-forest-and-itsimplementation-71824ced454f>, สืบค้นเมื่อ 18 ธันวาคม 2564
- [4] Sakulrat Khunsoongnoen, Dr.Sotarathammaboosadee, and Somkiat Wattanasirichaigoon “Using Predicting Analytics to Determine Discharge Status and Mortality in Sepsis and Septic Shock Patients based on Surgery and Medical Procedures” *KKU Research* Vol. 19, NO. 2, April-June 2019
- [5] W. Hu, W. Hu, and S. Maybank, “AdaBoost-based algorithm for network intrusion detection,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, Vol. 38, No. 2, pp. 577–583, Apr. 2008.
- [6] M. Alazab, S. Venkatraman, P. Watters, and M. Alazab, “Zero-day mal-ware detection based on supervised learning algorithms of API call signatures,” in *Proc. 9th Australas. Data Mining Conf., Ballarat, VIC, Australia*, Vol. 121, Dec. pp. 171–182, 2011.
- [7] Hongyu Liu and Bo Lang “Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey”, Beihang University, *Beijing 100191*, 17 October 2019.
- [8] Cheng Wang , Senior Member, IEEE, Bo Yang , Jipeng Cui, and Chaodong Wang, *IEEE Transactions On Computational Social Systems*, Vol. 6, No. 4, August 2019.
- [9]. N.Sirikulviriyaya, S.Sinthupinyo. Integration of Rules from a Random Forest, *2011 International Conference on Information and Electronics Engineering IPCSIT*, Vol.6. 2011.
- [10] M. M. Waldrop, “How to hack the hackers: The human side of cybercrime,” *Nature*, Vol. 533, No. 7602, pp. 164, 2016.
- [11] รศ.ดร. ไกรศักดิ์ เกสร, “โครงข่ายประสาทเทียมอัจฉริยะ ศิลปะแห่งวิทยาการข้อมูลสำหรับผู้เริ่มต้นโดยใช้ Python และ Azure ML.” ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร. หน้าที่ 263-272, มกราคม, 2564.
- [12] Samson Ho, Saleh Al Jufout, Khalil Dajani, and Mohammad Mozumdar “A Novel Intrusion Detection Model for Detecting Known and Innovative Cyberattacks Using Convolutional Neural Network” *Electrical Engineering Department, California State University Digital Object Identifier*, 10.1109/OJCS.2021.3050917 Vol. 2, pp. 14-25, 2021.

# การสกัดคุณลักษณะสำคัญของภาพโครโมโซมชนิด G-Band ในระยะเมทาเฟสสำหรับการเพิ่มประสิทธิภาพในการนับจำนวนเพื่อการวินิจฉัยโรคทางพันธุกรรมเบื้องต้น

## Extracting Key Features of Metaphase G-Band Chromosome Image for Counting Enhancement for Initial Genetic Diseases Diagnosis

บุษยามาส ปากคิบทอง (Bussayamas Pakkeetong)<sup>1</sup> อนูรักษ์ ยูทรานาวา (Anurak Yutthanawa)<sup>2</sup>

และพิเชษ วะยะลุน (Pichet Wayalun)<sup>3</sup>

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

<sup>1</sup>59160580@go.buu.ac.th, <sup>2</sup>yutt.anurak@gmail.com, <sup>3</sup>pichet.wa@go.buu.ac.th

### บทคัดย่อ

การสกัดกระดูกภาพเป็นหนึ่งในกระบวนการวิเคราะห์และการนับจำนวนโครโมโซม ซึ่งเป็นเทคนิคการประมวลผลภาพโดยปกติกระดูกภาพของโครโมโซมเดี่ยวจะมีจำนวนจุดปลาย 2 จุด ซึ่งการแตกแขนงของกระดูกภาพทำให้มีจำนวนจุดปลายมากกว่า 2 จุดและส่งผลทำให้เกิดการนับจำนวนโครโมโซมที่ผิดพลาด ดังนั้นงานวิจัยนี้จึงได้นำเสนอวิธีการ Skeleton Intersection Filling (SIF) และ Skeleton Intersection Point Detection (SIPD) เพื่อเพิ่มประสิทธิภาพและหาจุดสำคัญจากกระดูกภาพโครโมโซมเพื่อนำมาหาจำนวนจุดปลายที่เหมาะสม วิธีการที่นำเสนอนี้ช่วยในการนับจำนวนโครโมโซมและให้ผล 88% สำหรับโครโมโซมที่ซับซ้อน

**คำสำคัญ:** เทคโนโลยีการประมวลผลภาพ โครโมโซม

กระดูกภาพ จุดสำคัญ การนับจำนวน

### Abstract

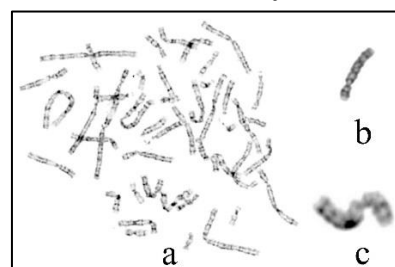
Skeleton image extraction is one of analysis and counting of chromosomes, which are image processing techniques. Normally, a single chromosome skeleton has two endpoints. However, the endpoints number of a skeleton branched are greater than two points. As a result, the number of chromosomes counted is incorrect. Therefore, this paper proposed the Skeleton Intersection Filling (SIF) and Skeleton Intersection Point Detection (SIPD), which help to enhance skeleton image and detection

interesting point. The proposed methods help to counted chromosome numbers and yield 88% for complicated chromosomes.

**Keywords:** Image Processing, Chromosome, Count, Skeleton, Interesting Point.

### 1. บทนำ

การวิเคราะห์ภาพโครโมโซมด้วยการนับจำนวนโครโมโซมเป็นกระบวนการสำคัญอย่างหนึ่งที่ช่วยให้แพทย์สามารถจำแนกโรคทางพันธุกรรมเบื้องต้นเช่น ดาวน์ซินโดรม และเป็นข้อบ่งชี้ที่สำคัญวินิจฉัยโรคทางพันธุกรรมต่าง ๆ แต่ปัจจุบันมีแพทย์ผู้เชี่ยวชาญในการอ่านผลที่ไม่เพียงพอต่อความต้องการของผู้ป่วย การใช้เทคโนโลยีการประมวลผลภาพด้วยเทคนิคการสกัดคุณลักษณะของเส้นกระดูกภาพจึงมีความสำคัญที่จะเข้ามาช่วยในการอ่านผลแต่เทคโนโลยีที่มีอยู่ในปัจจุบัน [1] [2] การประมวลผลรูปโครโมโซมเดี่ยวที่มีความผิดปกติตัวอย่างดังภาพที่ 1 จะพบปัญหาในการประมวลผลเส้นกระดูกภาพ



**ภาพที่ 1:** ตัวอย่าง (a) ภาพสไลด์โครโมโซม (b) ภาพโครโมโซมเดี่ยว (c) ภาพโครโมโซมเดี่ยวที่มีความผิดปกติ

การวิจัยนี้จึงได้นำเสนอวิธีการปรับปรุงกระดุกภาพโครโมโซมเดี่ยวและการสกัดคุณลักษณะสำคัญของกระดุกภาพเพื่อนำมาหาจุดปลายที่เหมาะสมแก่การนับจำนวนโครโมโซม เพื่อให้ได้ผลลัพธ์ที่น่าพึงพอใจมากยิ่งขึ้น

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 วิธีการนับจำนวนของโครโมโซม

การนับจำนวนโครโมโซมจากจำนวนจุดปลาย ด้วยการหาจำนวนจุดปลายจากกระดุกภาพ โดยสามารถอธิบายดังสมการที่ 1 [1]

$$A = \left[ \frac{E}{2} \right] \quad (1)$$

โดยที่  $A$  คือ ผลลัพธ์ของการนับโครโมโซมจากจุดปลาย

$E$  คือ จำนวนจุดปลาย

### 2.2 เครื่องมือในการวัดประสิทธิภาพ

1. Accuracy คือเครื่องมือในการวัดความถูกต้อง ณ ที่นี้จะใช้วัดความถูกต้องของการหาจุดสำคัญสามารถอธิบาย ดังสมการที่ 2 [3]

$$Accuracy = \frac{A}{B} \times 100 \quad (2)$$

โดยที่  $Accuracy$  คือ ค่าความถูกต้อง

$A$  คือ จำนวนภาพที่หาจุดสำคัญได้ถูกต้อง

$B$  คือ จำนวนภาพทั้งหมด

2. Sensitivity คือเครื่องมือการประเมินประสิทธิภาพของการนับจำนวนโครโมโซม อธิบายได้ดังสมการที่ 3 [1]

$$SE = \frac{TP}{TP + FN} \times 100 \quad (3)$$

โดยที่  $SE$  คือ สัดส่วนของการนับจำนวนโครโมโซมเดี่ยว

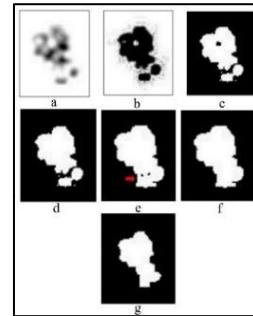
$TP$  คือ จำนวนโครโมโซมที่ระบุเป็นโครโมโซมเดี่ยว

$FN$  คือ จำนวนโครโมโซมที่ระบุว่าเป็นโครโมโซมเดี่ยว

โครโมโซมเดี่ยว

### 2.3 งานวิจัยที่เกี่ยวข้อง

1. P. Wayalun และ K. Kuloba (2018) ได้นำเสนอวิธีการเตรียมรูปภาพของกล้องจุลทรรศน์แบบใช้แสง เพื่อวิเคราะห์ความผิดปกติทางพันธุกรรมด้วยเทคนิคใหม่ คือ Determine the Number of Complicated Chromosome (DNCC) โดยประกอบไปด้วยขั้นตอนต่าง ๆ ดังภาพที่ 2



ภาพที่ 2: การปรับปรุงคุณภาพของรูปภาพ (a) รูปภาพ

ต้นฉบับ (b) Gamma Correction (c) Otsu (d) Flood-filled

(e) Dilation (f) Flood-filled (g) Erosion

จากภาพที่ 2 จะเห็นว่าวิธีการ DNCC จะมีกระบวนการ Gamma Correction เพื่อเพิ่มความเข้มของรูปภาพ Otsu Threshold ใช้ในการแยกวัตถุออกจากพื้นหลัง Flood-fill เพื่อเติมส่วนที่ขาดหาย Dilation เพื่อขยายวัตถุให้มีขนาดใหญ่ขึ้น Flood-fill อีกครั้งเพื่อให้ปลายโครโมโซมเชื่อมติดกับตัวโครโมโซม และ Erosion เพื่อให้พื้นที่ของวัตถุมีขนาดเล็กลง โดยจะให้ผลของกลุ่มโครโมโซมเดี่ยว 100% กลุ่มโครโมโซมทับซ้อน 100% และกลุ่มโครโมโซมที่มีรูปแบบซับซ้อน 79.12% [1]

2. P. Wayalun และ K. Kuloba (2019) ได้เสนอการเพิ่มประสิทธิภาพการปรับภาพให้เหมาะสมสำหรับการหาจำนวนโครโมโซมชนิด G-band ที่ซับซ้อนแบบอัตโนมัติ โดยใช้เทคนิคที่เรียกว่า Adaptive Complicated Chromosome Image Enhancement (ACCIE) ซึ่งจะทำให้การพัฒนาต่อจากเทคนิค HAFCO โดยใช้วิธีการกำหนดค่าเกณฑ์ ซึ่งพัฒนาวิธีการหาค่าเกณฑ์ที่ขึ้นอยู่กับวิธีการ Otsu และวิธีการปรับความเข้มแบบพลวัตที่ได้พัฒนาวิธีการปรับความเข้มแบบพลวัตตาม Gamma Correction ให้ค่าประสิทธิภาพสูงสุดที่ 86.81% [2]

### 3. วิธีการดำเนินงานวิจัย

#### 3.1 ขั้นตอนการดำเนินงานวิจัย

งานวิจัยนี้ ได้ใช้ขั้นตอนวิธีการดำเนินงานวิจัยดังนี้

##### 1. การปรับปรุงคุณภาพของรูปภาพ

การปรับปรุงคุณภาพของรูปภาพก่อนการประมวลผล หากระดุกภาพด้วยวิธีการ DNCC [1]

##### 2. การพัฒนาวิธีการเพิ่มประสิทธิภาพของเส้นกระดูกภาพโครโมโซม

การพัฒนาประสิทธิภาพของเส้นกระดูกภาพคือการเติมพิกเซลที่หายไปจากบริเวณจุด Intersection Point ให้มีความสมบูรณ์ซึ่งการวิจัยนี้ได้พัฒนาวิธีการที่เรียกว่า Skeleton Intersection Filling (SIF) ดังสมการที่ 4

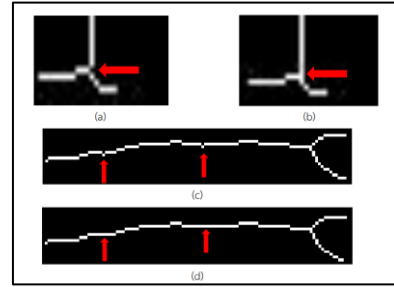
$$SIF \begin{cases} Fw(i,j) & \text{if } img_{i,j} = 255 \text{ and } img_{i,j+1} = 0 \\ Dw(i,j) & \text{if } img_{i,j} = 255 \text{ and } img_{i+1,j} = 0 \end{cases} \quad | \quad i \in \{0, \dots, y\}, y \in \{0, \dots, x\}$$

$$Dw(i,j) \begin{cases} img_{i+1,j} = 255 & \text{if } (img_{i+1,j+1} \text{ or } img_{i+1,j+1} = 255) \\ & \text{and } img_{i+2,j} = 255 \\ img_{i+1,j-1} = 0 & \text{if } img_{i,j-2} \text{ and } img_{i+1,j-2} \\ & \text{and } img_{i+2,j-2} \neq 255 \\ img_{i+1,j+1} = 0 & \text{if } img_{i,j+2} \text{ and } img_{i+1,j+2} \\ & \text{and } img_{i+2,j+2} \neq 255 \\ img_{i,j} = 0 & \text{if } img_{i-1,j} \text{ and } img_{i-1,j-1} \text{ and } img_{i-1,j+1} \\ & \text{and } img_{i,j-1} \text{ and } img_{i,j+1} \neq 255 \\ img_{i+2,j} = 0 & \text{if } img_{i+3,j} \text{ and } img_{i-1,j-1} \text{ and } img_{i+3,j+1} \\ & \text{and } img_{i+2,j-1} \text{ and } img_{i+2,j+1} \neq 255 \end{cases} \quad (4)$$

$$Fw(i,j) \begin{cases} img_{i,j+1} = 255 & \text{if } (img_{i-1,j+1} \text{ or } img_{i+1,j+1} = 255) \\ & \text{and } img_{i,j+2} = 255 \\ img_{i-1,j+1} = 0 & \text{if } img_{i-2,j} \text{ and } img_{i-2,j+1} \\ & \text{and } img_{i-2,j+2} \neq 255 \\ img_{i+1,j+1} = 0 & \text{if } img_{i+2,j} \text{ and } img_{i+2,j+1} \\ & \text{and } img_{i+2,j+2} \neq 255 \\ img_{i,j+2} = 0 & \text{if } img_{i,j+3} \text{ and } img_{i-1,j+3} \text{ and } img_{i+1,j+3} \\ & \text{and } img_{i-1,j+2} \text{ and } img_{i+2,j+2} \neq 255 \\ img_{i,j-1} = 0 & \text{if } img_{i,j-1} \text{ and } img_{i-1,j-1} \text{ and } img_{i+1,j-1} \\ & \text{and } img_{i-1,j} \text{ and } img_{i+1,j} \neq 255 \end{cases}$$

โดยที่  $img$  คือ เมทริกซ์ของภาพโครโมโซมระดับเทา  $i$  คือ ค่าตำแหน่งพิกเซลหลักในพิกเซลของ  $img$   $j$  คือ ค่าตำแหน่งพิกเซลแถวในพิกเซลของ  $img$   $Fw$  คือ ฟังก์ชันตรวจสอบเพื่อนบ้านแบบ Forward  $Dw$  คือ ฟังก์ชันตรวจสอบเพื่อนบ้านแบบ Downward

จากสมการที่ 4 จะประกอบด้วยการตรวจสอบเพื่อนบ้านแบบ Forward (Fw) ที่ตรวจสอบจากพิกเซลปัจจุบันไปทางขวาและ Downward (Dw) ที่ตรวจสอบจากพิกเซลปัจจุบันไปข้างล่าง ซึ่งแต่ละแบบจะมีเงื่อนไขที่จะทำการตรวจเช็คและเปลี่ยนแปลงค่าในพิกเซลเพื่อนบ้านให้มีความเหมาะสม ซึ่งทำให้พิกเซลที่ขาดหายไปบริเวณจุด Intersection Point ถูกเติมและกระดูกภาพบริเวณจุด Intersection Point มีความสมบูรณ์ดังตัวอย่างดังภาพที่ 3



ภาพที่ 3: (a) ภาพต้นฉบับ (b) ภาพที่ถูกปรับปรุงด้วย SIF (c) ภาพที่ถูกปรับปรุงด้วย SIF (d) ภาพที่ถูกปรับปรุงด้วย SIF

##### 3. การหาจำนวนจุดปลายของกระดูกภาพ

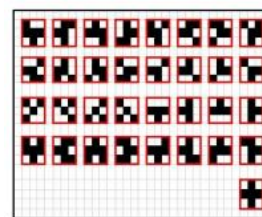
การหาจำนวนจุดปลายของกระดูกภาพด้วยวิธีการ Filter2D [5] ซึ่งทำให้ได้จำนวนจุดปลายและตำแหน่งจุดปลายของกระดูกภาพ โดยจะใช้ Kernel ดังภาพที่ 4

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 10 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

ภาพที่ 4: Kernel ที่ใช้ในการหาจุดปลายของกระดูกภาพ

##### 4. การพัฒนาวิธีการหาจุดสำคัญของกระดูกภาพของโครโมโซม

การพัฒนาวิธีการหาจุดสำคัญของกระดูกภาพของโครโมโซมด้วย Skeleton Intersection Point Detection (SIPD) ที่การวิจัยครั้งนี้ได้พัฒนาบนพื้นฐานของ Hit-or-Miss Transform [6] โดย SIPD ได้พัฒนา Intersection Kernel จำนวน 33 รูปแบบ ดังภาพที่ 5



ภาพที่ 5: Intersection Kernel ของ SIPD ที่ใช้ในการหา Intersection Point

จากภาพที่ 5 จะประกอบไปด้วย Kernel ต่าง ๆ จำนวน 33 รูปแบบในการหาจุด Intersection Point บนรูปภาพโครโมโซมระดับเทา ด้วยการหารูปแบบของกลุ่มพิกเซลที่ตรงกับ kernel ที่ได้สร้างไว้เพื่อกำหนดจุด Intersection Point ที่เป็นคุณลักษณะสำคัญ

5. การวิเคราะห์เพื่อหาจำนวนจุดปลายที่เหมาะสมของกระดูกภาพ

การวิเคราะห์เพื่อหาจำนวนจุดปลายที่เหมาะสม จะนำระยะห่างระหว่างจุดปลายและจุด Intersection Point มาเทียบกับระยะห่างระหว่างจุดปลายและจุด Intersection Point ที่มีค่ามากที่สุดหารด้วยสองหรือในกรณีที่มีจำนวนจุด Intersection Point 2 จุด จะเทียบกับค่าระยะห่างระหว่างจุด Intersection Point ด้วยกัน ซึ่งจะทำให้ได้จำนวนจุดปลายใหม่ที่เกิดจากการวิเคราะห์เพื่อนำไปนับจำนวนโครโมโซม

### 3.2 ออกแบบการทดลอง

การวิจัยนี้ได้แบ่งการทดลองออกเป็น 3 กลุ่มดังนี้

1. กลุ่มภาพโครโมโซมเดี่ยวที่มีความเข้มของภาพสูง
2. กลุ่มภาพโครโมโซมเดี่ยวที่มีความสว่างสูง
3. กลุ่มภาพโครโมโซมเดี่ยวที่มีความผิดปกติ

โดยแสดงตัวอย่างของภาพโครโมโซมตามตารางที่ 1

ตารางที่ 1: ตัวอย่างของภาพโครโมโซมตามประเภทการทดลอง

ประเภทการทดลอง	โครโมโซมเดี่ยวที่มีความเข้มของภาพสูง	โครโมโซมเดี่ยวที่มีความสว่างของภาพสูง	โครโมโซมเดี่ยวที่มีความผิดปกติ
ภาพโครโมโซมเดี่ยวชนิดเซลล์สูงน้ำคร่ำ			
ภาพโครโมโซมเดี่ยวชนิดเซลล์เม็ดเลือดขาวลิ้มโฟไซต์			

### 3.3 ที่มาของข้อมูล

ข้อมูลรูปภาพโครโมโซมเดี่ยวที่ใช้ในการทำวิจัยได้รับมาจาก Bioweb, the University of Wisconsin System [7]

โดยการวิจัยนี้ได้รับการสนับสนุนทุนวิจัยจากคณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

## 4. ผลการทดลอง

1. การทดลองการนับจำนวนกับกลุ่มภาพโครโมโซมเดี่ยวที่มีความเข้มของภาพสูง

ภาพที่นำมาทดลองสามารถแบ่งออกเป็นภาพโครโมโซมเดี่ยวชนิดเซลล์สูงน้ำคร่ำที่มีค่าเฉลี่ย Histogram น้อยกว่า 211.89368 จำนวน 50 ภาพ ซึ่งเป็นภาพระดับสีเทาที่มีประเภทของภาพคือ .jpg โดยมีขนาด 37x32 Pixel ขึ้นไปและภาพโครโมโซมเดี่ยวชนิดเซลล์เม็ดเลือดขาวลิ้มโฟไซต์ [4] ที่มีค่าเฉลี่ย Histogram น้อยกว่า 224.9581 จำนวน 50 ภาพ ซึ่งเป็นภาพระดับสีเทาที่มีประเภทของภาพคือ .jpg โดยมีขนาด 27x26 Pixel ขึ้นไป ซึ่งให้ค่าความถูกต้องในการหาจุดสำคัญของกระดูกภาพดังตารางที่ 2

ตารางที่ 2: ค่าความถูกต้องในการหาจุดสำคัญของกลุ่มภาพโครโมโซมเดี่ยวที่มีความเข้มของภาพสูงตามสมการที่ 2

Chromosome Type	จำนวนรูปภาพโครโมโซมที่นับได้ถูกต้อง	จำนวนรูปภาพโครโมโซมทั้งหมด	Accuracy (%)
ภาพโครโมโซมเดี่ยวชนิดเซลล์สูงน้ำคร่ำ	49	50	98
ภาพโครโมโซมเดี่ยวชนิดเซลล์เม็ดเลือดขาวลิ้มโฟไซต์	50	50	100

สามารถแสดงค่าประสิทธิภาพในการนับจำนวน ดังตารางที่ 3

ตารางที่ 3: ค่าประสิทธิภาพในการนับจำนวนของกลุ่มภาพโครโมโซมเดี่ยวที่มีความเข้มของภาพสูงตามสมการที่ 3

Chromosome Type	TP (จำนวน)	FN (จำนวน)	SE (%)
ภาพโครโมโซมเดี่ยวชนิดเซลล์สูงน้ำคร่ำ	43	7	86



ภาพโครโมโซมเดี่ยว ชนิดเซลล์เม็ดเลือด ขาวลิมโฟไซต์	50	0	100
---------------------------------------------------------	----	---	-----

2. การทดลองการนับจำนวนกับกลุ่มภาพโครโมโซมเดี่ยวที่มีความสว่างของภาพสูง

ภาพที่นำมาทดลองสามารถแบ่งออกเป็นภาพโครโมโซมเดี่ยวชนิดเซลล์ถุงน้ำคร่ำที่มีค่าเฉลี่ย Histogram มากกว่า 211.89368 จำนวน 50 ภาพ ซึ่งเป็นภาพระดับสีเทาที่มีประเภทของภาพคือ .jpg โดยมีขนาด 45x46 Pixel ขึ้นไป และภาพโครโมโซมเดี่ยวชนิดเซลล์เม็ดเลือดขาวลิมโฟไซต์ที่มีค่าเฉลี่ย Histogram มากกว่า 224.9581 จำนวน 50 ภาพ ซึ่งเป็นภาพระดับสีเทาที่มีประเภทของภาพคือ .jpg โดยมีขนาด 36x45 Pixel ขึ้นไป

ซึ่งให้ค่าความถูกต้องในการหาจุดสำคัญของกระดูกภาพ ดังตารางที่ 4

**ตารางที่ 4:** ค่าความถูกต้องในการหาจุดสำคัญของกลุ่มภาพโครโมโซมเดี่ยวที่มีความสว่างของภาพสูงตามสมการที่ 2

Chromosome Type	จำนวนรูปภาพโครโมโซมที่นับได้ถูกต้อง	จำนวนรูปภาพโครโมโซมทั้งหมด	Accuracy (%)
ภาพโครโมโซมเดี่ยวชนิดเซลล์ถุงน้ำคร่ำ	49	50	98
ภาพโครโมโซมเดี่ยวชนิดเซลล์เม็ดเลือดขาวลิมโฟไซต์	50	50	100

สามารถแสดงค่าประสิทธิภาพในการนับจำนวน ดังตารางที่ 5

**ตารางที่ 5:** ค่าประสิทธิภาพในการนับจำนวนของกลุ่มภาพโครโมโซมเดี่ยวที่มีความสว่างของภาพสูงตามสมการที่ 3

Chromosome Type	TP (จำนวน)	FN (จำนวน)	SE (%)
ภาพโครโมโซมเดี่ยวชนิดเซลล์ถุงน้ำคร่ำ	50	0	100
ภาพโครโมโซมเดี่ยวชนิดเซลล์เม็ดเลือดขาวลิมโฟไซต์	50	0	100

3. การทดลองการนับจำนวนกับกลุ่มภาพโครโมโซมเดี่ยวที่มีความมืดปกติ

ภาพที่นำมาทดลองสามารถแบ่งออกเป็นภาพโครโมโซมเดี่ยวชนิดเซลล์ถุงน้ำคร่ำที่มีแท่งโครโมโซมที่คดงอ และแท่งโครโมโซมที่หักพับจำนวน 50 ภาพ ซึ่งเป็นภาพระดับสีเทาที่มีประเภทของภาพคือ .jpg โดยมีขนาด 52x38 Pixel ขึ้นไป และภาพโครโมโซมเดี่ยวชนิดเซลล์เม็ดเลือดขาวลิมโฟไซต์ที่มีแท่งโครโมโซมที่คดงอและแท่งโครโมโซมที่หักพับจำนวน 50 ภาพ ซึ่งเป็นภาพระดับสีเทาที่มีประเภทของภาพคือ .jpg โดยมีขนาด 38x28 Pixel ขึ้นไป

ซึ่งให้ค่าความถูกต้องในการหาจุดสำคัญของกระดูกภาพ ดังตารางที่ 6

**ตารางที่ 6:** ค่าความถูกต้องในการหาจุดสำคัญของกลุ่มภาพโครโมโซมเดี่ยวที่มีความมืดปกติตามสมการที่ 2

Chromosome Type	จำนวนรูปภาพโครโมโซมที่นับได้ถูกต้อง	จำนวนรูปภาพโครโมโซมทั้งหมด	Accuracy (%)
ภาพโครโมโซมเดี่ยวชนิดเซลล์ถุงน้ำคร่ำ	47	50	94
ภาพโครโมโซมเดี่ยวชนิดเซลล์เม็ดเลือดขาวลิมโฟไซต์	50	50	100

สามารถแสดงค่าประสิทธิภาพในการนับจำนวน ดังตารางที่ 7

**ตารางที่ 7:** ค่าประสิทธิภาพในการนับจำนวนของกลุ่มภาพโครโมโซมเดี่ยวที่มีความมืดปกติตามสมการที่ 3

Chromosome Type	TP (จำนวน)	FN (จำนวน)	SE (%)
ภาพโครโมโซมเดี่ยวชนิดเซลล์ถุงน้ำคร่ำ	43	7	86
ภาพโครโมโซมเดี่ยวชนิดเซลล์เม็ดเลือดขาวลิมโฟไซต์	45	5	90

จากตารางที่ 2 ตารางที่ 4 และตารางที่ 6 แสดงถึงการหาจุดสำคัญผิดพลาดนั้นเกิดจากการเติมพิกเซลด้วยเทคนิค

SIF ทำให้การหาจุดคาดเคลื่อนจากจุดที่แท้จริง 1-2 พิกเซล ซึ่งส่งผลกระทบต่อการวิเคราะห์เพียงเล็กน้อย ดังนั้นจุด Intersection Point ที่หาได้คาดเคลื่อนสามารถนำไปใช้วิเคราะห์เพื่อหาจำนวนจุดปลายที่เหมาะสมได้

จากตารางที่ 3 แสดงถึงโครโมโซมเดี่ยวชนิดเซลล์ดุงน้ำคร่ำที่มีความเข้มของภาพสูงให้ค่าประสิทธิภาพที่ 86% เนื่องจากบริเวณขอบของแท่งโครโมโซมมีความคมชัดมาก ซึ่งทำให้ขอบของโครโมโซมชนิดเซลล์ดุงน้ำคร่ำที่มีความหยาบและเว้าแหว่งกว่าภาพโครโมโซมชนิดเซลล์เม็ดเลือดขาวลิมโฟไซต์มีความเด่นชัดทำให้เกิดการแตกแขนงของกระดูกภาพที่มากขึ้น

จากตารางที่ 5 แสดงถึงโครโมโซมเดี่ยวทั้งสองชนิดให้ประสิทธิภาพในการนับ 100% เนื่องจากบริเวณขอบของแท่งโครโมโซมมีความคมชัดไม่มาก ซึ่งทำให้มีความหยาบและเว้าแหว่งน้อยจึงเกิดการแตกแขนงของกระดูกภาพที่น้อย

จากตารางที่ 7 แสดงถึงประสิทธิภาพในการนับจำนวนโครโมโซมของภาพโครโมโซมเดี่ยวชนิดเซลล์ดุงน้ำคร่ำและภาพโครโมโซมเดี่ยว ชนิดเซลล์เม็ดเลือดขาวลิมโฟไซต์เท่ากับ 86% และ 90% ตามลำดับ เนื่องจากภาพโครโมโซมที่มีความคดงอและหักพับ จะทำให้เกิดการแตกแขนงของกระดูกภาพที่มาก

## 5. สรุป

การวิจัยนี้ได้นำเสนอวิธีการ SIF และ SIPD เพื่อเพิ่มประสิทธิภาพและหาจุดตัดจากกระดูกภาพโครโมโซมเดี่ยวเพื่อนำมาหาจำนวนจุดปลายที่เหมาะสมแก่การนับจำนวนโครโมโซม จากการทดลองพบว่าประสิทธิภาพในการนับจำนวนโครโมโซมเมื่อผ่านวิธีการที่นำเสนอ โดยโครโมโซมเดี่ยวที่มีความเข้มของภาพสูงมีประสิทธิภาพถึง 93% และโครโมโซมเดี่ยวที่มีความสว่างของภาพสูงมีประสิทธิภาพถึง 100% นอกจากนี้พบว่าประสิทธิภาพในการนับจำนวนโครโมโซมที่มีความผิดปกติมีประสิทธิภาพถึง 88% ซึ่งมากกว่าวิธีการ DNCC [1] และ ACCIE [2] ดังตารางที่ 8

**ตารางที่ 8:** เปรียบเทียบค่าความถูกต้องระหว่าง DNCC ACCIE และ SIF + SIPD ในภาพโครโมโซมเดี่ยวที่มีความผิดปกติ

Chromosome Type	DNCC (%)	ACCIE (%)	SIF + SIPD (%)
ภาพโครโมโซมเดี่ยวที่มีความผิดปกติ	79.12	86.81	88.0

ดังนั้นวิธีการ SIF และ SIPD จึงสามารถช่วยเพิ่มประสิทธิภาพในการนับจำนวนภาพโครโมโซมชนิดความเข้มของภาพสูง ความสว่างของภาพสูง และ โครโมโซมที่มีความผิดปกติ

## เอกสารอ้างอิง

- [1] K. Kubola, and P. Wayalun, "Automatic Determination of The G-band Chromosomes Number based on Geometric Features," *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1-5, 2018.
- [2] P. Wayalun, and K. Kubola, "Adaptive Image Enhancement for Automatic Complicated G-band Chromosome Number Determination," *2019 12th Biomedical Engineering International Conference (BMEiCON)*, pp. 1-5, 2019.
- [3] K. Kubola and P. Wayalun, "Amniotic Fluid Bag Chromosome Geometric Features Detection" *2020 – 5th International Conference on Information Technology (InCIT)*, Chonburi, Thailand, pp. 294-194, 2020.
- [4] Marisa44638, "โครโมโซมและการถ่ายทอดลักษณะพันธุกรรม", <https://bit.ly/3bUJb3d>, 2015.
- [5] Kang&atul, *Spatial Filtering*, Available Online at <https://theailearner.com/tag/cv2-filter2d/>, 2019.
- [6] R. Fisher, S. Perkins, A. Walker, and E. Wolfart, *Hit-and-Miss Transform*, Available Online at <https://homepages.inf.ed.ac.uk/rbf/HIPR2/hitmiss.htm>,
- [7] L. Phelps, *Human karyotypes for teaching*, Available Online at <http://worms.zoology.wisc.edu/zooweb/Phelps/karyotype.html>.

# การปรับสมดุลโหลดด้วยการค้นพบโหนดใกล้เคียง สำหรับเกณฑ์วิธีจัดสรรเส้นทางอาร์พีแอลในเครือข่ายเซนเซอร์ไร้สาย

## Load Balancing by Neighbour Node Discovery for RPL Routing Protocol in Wireless Sensor Networks

สุทธิธีร์ จาตุพรประเสริฐ (Suttitee Jathuphornsaserd)<sup>1</sup> และศุภกิจ อวัพันธ์ (Suphakit Awiphan)<sup>2</sup>

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

suttitee\_j@cmu.ac.th, suphakit.a@cmu.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการจัดการจัดเส้นทางของโปรโตคอลอาร์พีแอลของเครือข่ายเซนเซอร์ไร้สาย และพัฒนาเทคนิคการปรับสมดุลโหลดของโหนดในเครือข่ายเพื่อปรับปรุงประสิทธิภาพของการขนส่งข้อมูล โดยทำการเปรียบเทียบประสิทธิภาพของฟังก์ชันวัตถุประสงค์ที่นำเสนอกับวิธีโอเอฟซีโร และเอ็มอาร์เอสโอเอฟ ซึ่งเป็นฟังก์ชันวัตถุประสงค์ที่ถูกจัดเป็นมาตรฐาน จากการทดลองพบว่าในเครือข่ายที่มีโหนด 50, 100 และ 200 โหนดนั้นฟังก์ชันวัตถุประสงค์ที่นำเสนอ มีอัตราส่วนการจัดส่งแพ็กเก็ตที่สูงกว่าวิธีการอื่นที่นำมาเปรียบเทียบ เนื่องจากโหนดแม่ในเครือข่ายมีจำนวนของโหนดลูกที่เหมาะสม จึงช่วยลดความแออัดของทราฟฟิกที่ทำให้เกิดการส่งข้อมูลซ้ำได้

**คำสำคัญ:** อาร์พีแอล เครือข่ายเซนเซอร์ไร้สาย การปรับสมดุล

### Abstract

*This research aims to study the RPL protocol of wireless sensor networks and propose the load balancing technique of the nodes in the network to improve the data delivery performance. The proposed objective function was compared with other two methods which are OF0 and MRHOF. The results showed that in the networks with 50, 100 and 200 nodes, the proposed approach provided better packet delivery ratio than the comparisons. This is because the parent node has the appropriate number of*

*children nodes. Accordingly, the traffic congestion which might lead to unnecessary retransmission could be reduced.*

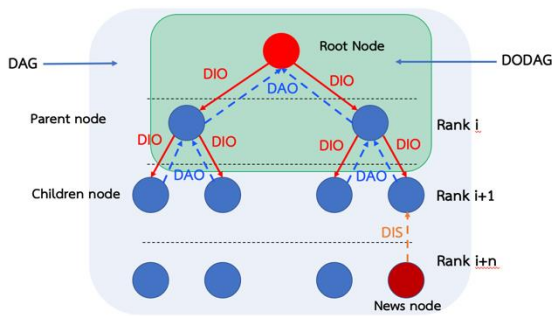
**Keyword:** RPL; Wireless Sensor Network; Load Balancing.

### 1. บทนำ

อาร์พีแอล (RPL: IPv6 Routing Protocol for Low Power and Lossy Network) เป็นโปรโตคอลการจัดสรรเส้นทางที่ถูกพัฒนาขึ้นเพื่อใช้งานกับเครือข่ายเซนเซอร์ไร้สาย (Wireless Sensor Network) ซึ่งจัดเป็นเครือข่ายที่มีการใช้พลังงานต่ำ และมีโอกาสเกิดการสูญหายของข้อมูลสูง หรือ แอลแอลเอ็น (LLN: Low-Power and Lossy Networks) [1] ด้วยลักษณะดังกล่าวทำให้อาร์พีแอลสามารถประยุกต์ใช้งานในเครือข่ายที่มีขนาดเล็ก ซึ่งอาจมีอุปกรณ์อินเทอร์เน็ตของสรรพสิ่ง หรือ ไอโอที (Internet of Things: IoT) ในเครือข่ายร่วมด้วย

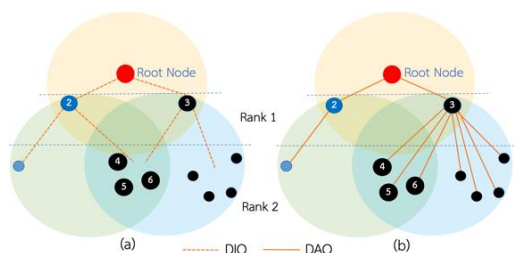
การแลกเปลี่ยนข้อมูลบนเครือข่ายมีทั้งแบบผู้ส่งรายเดียว ไปยังผู้รับรายเดียว (One-to-One) และแบบผู้ส่งรายเดียว ไปยังผู้รับหลายราย (One-to-Many) [2] ลักษณะโครงสร้างของเครือข่ายที่ใช้โปรโตคอลอาร์พีแอลจะเป็นแบบกราฟอวัฏจักรระบุทิศทาง หรือ ดีเอจี (Directed Acyclic Graph: DAG) ซึ่งแต่ละโหนดจะสามารถสร้างโครงข่ายย่อยในลักษณะของกราฟอวัฏจักรระบุทิศทางเชิงปลายทาง หรือ ดีโอดีเอจี (Destination Oriented Directed Acyclic Graph: DODAG) [3] ที่มีโหนดเริ่มต้น (Root Node)

เพียงโหนดเดียว โดยในการสร้างเส้นทางส่งข้อมูลนั้น โหนดเริ่มต้นจะส่งข้อความที่เรียกว่า ดีไอโอ (DIO: DODAG Information Object) ซึ่งเป็นข้อความที่ช่วยให้เกิดการแบ่งปันข้อมูลของกราฟไปยังโหนดอื่น ๆ โดยที่จะมีการส่งข้อความดังกล่าวเป็นระยะเพื่อปรับปรุงข้อมูลของโหนดที่อยู่บนโทโพโลยีของเครือข่าย [4,5] นอกจากนี้ยังมีข้อมูลของการจัดลำดับ (Rank) ซึ่งเป็นค่าที่ระบุตำแหน่งเชิงสัมพันธ์ของโหนดภายในกราฟ เพื่อป้องกันการเกิดเส้นทางภายในกราฟที่มีลักษณะเป็นวงกลม ภาพที่ 1 แสดงกราฟตัวอย่างของเครือข่ายที่ถูกสร้างด้วยโปรโตคอลอาร์พีแอล



ภาพที่ 1: กราฟตัวอย่างของเครือข่ายที่ถูกสร้างด้วยโปรโตคอลอาร์พีแอล

อย่างไรก็ตาม จากการศึกษาพบว่า โปรโตคอลอาร์พีแอลยังคงต้องการการปรับปรุงเพิ่มเติมในแง่ของการปรับสมดุลโหลด (Load Balancing) การใช้พลังงาน (Power Consumption) และการรับรู้บริบท (Context Aware) เนื่องจากฟังก์ชันวัตถุประสงค์ที่เป็นมาตรฐานเดิมของโปรโตคอลอาร์พีแอล มักทำให้เกิดการกระจายของโหลดที่ไม่สมดุลในเครือข่าย ซึ่งนำไปสู่ความแออัดของทราฟฟิก (Traffic Congestion) ของการรับส่งข้อมูลที่โหนดใดโหนดหนึ่ง



ภาพที่ 2: ปัญหาการกระจายโหลดแบบไม่สมดุลในเครือข่ายเซนเซอร์ไร้สาย

จากภาพที่ 2 (a) โหนดแม่ทำการส่งข้อความ DIO ไปยังโหนดลูก และโหนดลูกตอบรับ DAO เข้าร่วมเครือข่ายดังภาพที่ 2 (b) ทำให้จำนวนโหนดลูก (Child Node) ต่อโหนดแม่ (Parent Node) และโหนดเริ่มต้น (Root Node) มีความไม่สมดุล โดยปัญหานี้จะมีผลกระทบที่สูงขึ้นหากโหนดมีโหลดเกินกว่าที่จะสามารถรองรับได้เนื่องจากส่งผลให้เกิดความแออัดของทราฟฟิกการรับส่งข้อมูล และเมื่อเกิดการสูญหายของข้อมูล โหนดที่เกี่ยวข้องจำเป็นต้องส่งข้อมูลซ้ำอีกครั้ง ทำให้เกิดการใช้งานทรัพยากรของเครือข่ายโดยไม่จำเป็น ซึ่งสามารถส่งผลกระทบต่อการใช้พลังงาน และในบางสถานการณ์ที่ต้องการการขนส่งข้อมูลที่มีความล่าช้าต่ำ ระดับความลึกของกราฟที่มากจากการที่กราฟไม่สมดุล สามารถส่งผลกระทบต่อความล่าช้าในการขนส่งข้อมูลจากปลายถึงปลาย (End-to-End Delay) ได้

จากปัญหาดังกล่าว ผู้วิจัยจึงมีแนวคิดที่จะปรับปรุงวิธีการปรับสมดุลโหลด โดยใช้เทคนิคการสำรวจโหนดใกล้เคียง (Neighbor Node Discovery) ร่วมกับการนับจำนวนของการส่งที่คาดไว้ หรืออีทีเอ็กซ์ (Expected Transmission Count : ETX) เพื่อช่วยปรับสมดุลโหลดของโหนด และเพิ่มประสิทธิภาพของการสื่อสารข้อมูลบนเครือข่ายเซนเซอร์ไร้สาย โดยมีการเปรียบเทียบประสิทธิภาพของฟังก์ชันวัตถุประสงค์ที่นำเสนอกับฟังก์ชันวัตถุประสงค์แบบเดิม ซึ่งได้แก่ โอเอฟซีโร (OFO: Objective Function Zero) และเอ็มอาร์เอชโอเอฟ (MRHOF: Minimum Rank with Hysteresis Objective Function) ด้วยโปรแกรมจำลองเครือข่าย

## 2. งานวิจัยที่เกี่ยวข้อง

### 2.1 การจัดสรรเส้นทางโดยปรับสมดุลการใช้พลังงาน

Chen และ Nasser [6] ทำการกำหนดเส้นทางการปรับสมดุลพลังงานในเครือข่ายเซนเซอร์ไร้สาย โดยที่โหนดแม่จะเลือกเส้นทางที่จะส่งแพ็กเก็ตข้อมูลเป็นระยะ โดยพิจารณาจากระดับพลังงานของเส้นทางต่าง ๆ ที่รักษาไว้ โดยใช้เทคนิคโปรโตคอลการกำหนดเส้นทางแบบหลายเส้นทาง (Multipath Routing Protocol) และทำการวัด

ประสิทธิภาพของอัตราส่วนการจัดส่ง (Delivery Ratio) และพลังงานคงเหลือเฉลี่ยของโหนด อย่างไรก็ตาม การกำหนดเส้นทางด้วยวิธีดังกล่าว จำเป็นต้องมีการคำนวณเส้นทางที่เหมาะสมที่สุดในแบบรวมศูนย์ เนื่องจากโทโพโลยีอาจมีการเปลี่ยนแปลงที่ไม่เสถียร

## 2.2 การแก้ไขความแออัดของการขนส่งข้อมูลบนเครือข่ายเซนเซอร์ไร้สาย

Pradeska และคณะ [7] ต้องการแก้ไขปัญหาความแออัดของการขนส่งข้อมูลบนเครือข่ายเซนเซอร์ไร้สาย และปรับปรุงประสิทธิภาพเครือข่ายเมื่อเกิดความแออัด โดยทำการทดลองเพื่อประเมินและวิเคราะห์สภาวะของเครือข่ายที่มีความแออัด ซึ่งแพ็กเก็ตส่วนใหญ่จะสูญหายเนื่องจากบัฟเฟอร์ล้น ดังนั้น จึงเป็นสิ่งสำคัญที่จะต้องคำนึงถึงการครอบครองบัฟเฟอร์ โดยใช้เทคนิค (BO: Buffer Occupancy) และอีทีเอ็กซ์ (ETX) ในฟังก์ชันวัตถุประสงค์ การใช้บัฟเฟอร์จึงถูกใช้เป็นตัวบ่งชี้เพื่อให้ทราบถึงความน่าจะเป็นของการเกิดบัฟเฟอร์ล้น และทำการประเมินประสิทธิภาพโดยนำมาเปรียบเทียบกับฟังก์ชันวัตถุประสงค์สามประการ ได้แก่ โอเอฟโอ (OFO) อีทีเอ็กซ์-โอเอฟ (ETX-OF) และ ENERGY-OF

## 2.3 ฟังก์ชันวัตถุประสงค์

ฟังก์ชันวัตถุประสงค์เป็นกลไกการเลือกเส้นทางโดยที่โหนดลูกจะเลือกโหนดแม่ตามวิธีการที่ระบุไว้ โดยโอเอฟซีโร (OF0: Objective Function Zero) เป็นฟังก์ชันวัตถุประสงค์ที่ใช้ค่าต่ำสุดของฮอป (Hop) เพื่อเข้าถึงโหนดเริ่มต้น ซึ่งแต่ละโหนดจะคำนวณอันดับตามค่าการข้ามไปยังโหนดเริ่มต้น โดยจำนวนของฮอปที่น้อยลงจะได้รับลิงก์ที่มีลำดับความสำคัญสูงกว่า สำหรับโอเอฟซีโรนั้น ค่าอันดับบนโหนดย่อยจะสูงกว่าโหนดหลักเสมอ ขณะที่เอ็มอาร์เอชโอเอฟ (MRHOF: Minimum Rank with Hysteresis Objective Function) เป็นฟังก์ชันวัตถุประสงค์ที่ใช้การนับจำนวนของการส่งที่คาดหวัง ที่ได้รับ โดยไม่มีข้อผิดพลาดที่ปลายทาง หรืออีทีเอ็กซ์ (ETX) ในการเลือกโหนดหลัก [8,9] การคำนวณ (ETX) แสดงในสมการที่ (1)

$$ETX = \frac{1}{Df \times Dr} \quad (1)$$

ค่า  $Df$  ที่ได้จากการคำนวณความน่าจะเป็นของแพ็กเก็ตที่ได้รับความเสียหายจากโหนดข้างเคียง ในขณะที่ค่า  $Dr$  ได้จากการคำนวณความน่าจะเป็นของแพ็กเก็ตที่ตอบรับ (ACK) ที่ผู้รับค่า (ETX) สามารถเป็นค่าที่ไม่ต่อเนื่อง หรือไม่ใช่จำนวนเต็ม [10]

พารามิเตอร์เชิงปริมาณ เช่น ความน่าเชื่อถือของลิงก์ (Link Reliability) และ ปริมาณของพลังงานที่เหลือของโหนด (Remaining Node Energy) สามารถใช้เป็นเมตริกส์ซึ่งเป็นชนิดของพารามิเตอร์ในฟังก์ชันวัตถุประสงค์หรือเงื่อนไขที่ต้องพิจารณาเมื่อเลือกเส้นทางในกรณีที่ใช้ความน่าเชื่อถือของลิงก์เป็นเงื่อนไข เส้นทางที่ถูกเลือกจะเป็นเส้นทางที่มีการรับประกันค่าความน่าเชื่อถือ และเส้นทางที่มีระดับ (Rank) น้อยที่สุดจะถูกพิจารณาเป็นเส้นทางที่ดีที่สุดโดยข้อมูลดังกล่าวจะมีอยู่ในข้อความดีไอโอ (DIO) เมื่อสร้างดีไอโอดีเอจ (DODAG)

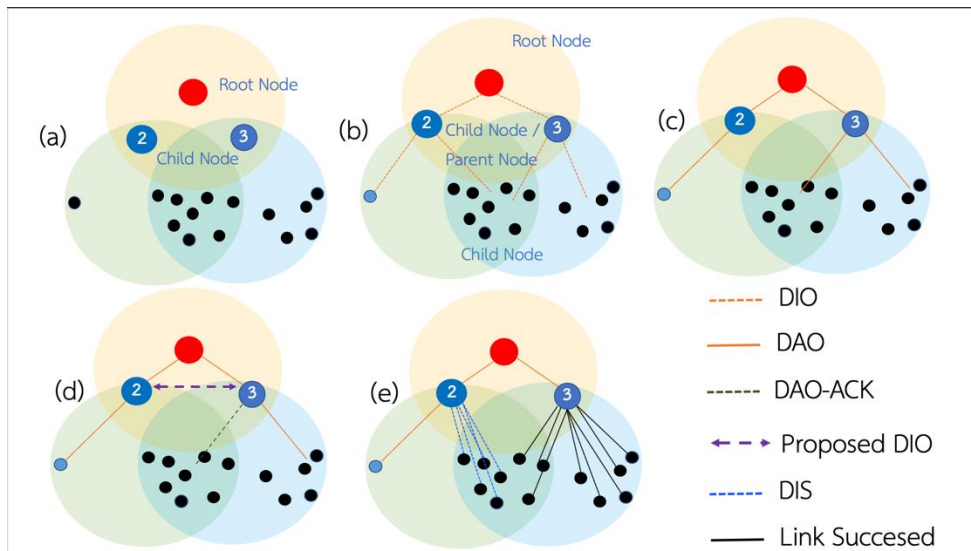
## 3. วิธีการดำเนินการวิจัย

### 3.1 กรอบแนวคิดการวิจัย

จากการศึกษางานวิจัยที่กล่าวมาข้างต้น ผู้วิจัยจึงมีแนวคิดในการเสนอการปรับปรุงวิธีการปรับสมดุลโหลด โดยใช้เทคนิคการสำรวจโหนดใกล้เคียง ร่วมกับการนับจำนวนของการส่งที่คาดไว้ เพื่อช่วยปรับสมดุลโหลดของโหนดแม่กับโหนดลูก

โดยทั่วไป อัตราส่วนของแพ็กเก็ตที่ส่งจะเพิ่มขึ้นตามการใช้งานในเครือข่ายขนาดใหญ่ โหนดลูกสามารถเลือกโหนดแม่ที่น่าเชื่อถือเพื่อกำหนดเส้นทางแพ็กเก็ตข้อมูลไปยังโหนดหลัก อย่างไรก็ตามหากโหนดลูกส่งข้อมูลผ่านเส้นทางที่มีความแออัด อาจทำให้เกิดการสูญหายของข้อมูล ผู้วิจัยจึงพยายามกระจายโหลดของโหนดแม่เพื่อหลีกเลี่ยงความแออัดของการขนส่งข้อมูล ซึ่งมีส่วนทำให้การส่งต่อแพ็กเก็ตสามารถทำได้อย่างมีประสิทธิภาพ โดยมีขั้นตอนวิธี แสดงในภาพที่ 3 ดังนี้

- ภาพที่ 3 (a) แสดงตัวอย่างโครงสร้างดีเอจ (DAG) ของโปรโตคอลอาร์พีแอล
- ภาพที่ 3 (b) โหนดเริ่มต้นและโหนดแม่จะทำการส่งข้อความดีไอโอ (DIO: DODAG)



ภาพที่ 3: ขั้นตอนวิธีการค้นพบโหนดใกล้เคียงสำหรับการปรับสมดุลโหลดในเครือข่ายเซนเซอร์ไร้สาย

Information Object) ไปยังโหนดลูกที่อยู่ในพื้นที่ทั้งหมด

- ภาพที่ 3 (c) โหนดลูกทำการตอบข้อความดีเอโอ (DAO: Destination Advertisement Object) กลับมายังโหนดแม่ ซึ่งจากภาพที่ 3 (c) โหนดลูกเลือกที่จะเข้าร่วมกับโหนด 3 ทำให้เกิดความไม่สมดุลของโหลดในเครือข่าย
- ภาพที่ 3 (d) ในกรณีที่สามารับโหนดลูกใหม่ได้โหนด 3 ทำการส่งข้อความดีเอโอ - เอชเค (DAO-ACK) ไปให้โหนดลูกเพื่อตอบรับให้เข้าร่วมเครือข่าย แต่หากไม่สามารถรับโหนดลูกใหม่ได้ โหนด 3 จะส่งข้อความ Proposed DIO ไปยังโหนดใกล้เคียงเพื่อสอบถามความสมดุลโหลดของโหนดใกล้เคียง โดยใช้ร่วมกับค่า ETX เพื่อตรวจสอบความแออัดข้อมูลของโหนดข้างเคียง โดยเลือกค่า ETX ที่มีเส้นทางที่มีอันดับต่ำสุด
- ภาพที่ 3 (e) โหนดลูกที่ไม่ได้รับการตอบรับจากโหนดแม่จะส่งข้อความดีไอเอส (DIS DODAG Information Solicitation) ไปยังโหนดใกล้เคียงเพื่อขอเข้าร่วมเครือข่ายกับโหนด 2 ต่อไป

ตารางที่ 1: ขั้นตอนวิธีการค้นพบโหนดใกล้เคียงสำหรับการปรับสมดุลโหลดในเครือข่ายเซนเซอร์ไร้สาย

1. Initialize
2. Parent node sends DIO message to all children nodes
3. Child node sends DAO message to prospective parent
4. **if** parent sends DAO-ACK message  
child node joins prospective parent  
**else**  
parent node sends Proposed DIO to neighbor parent and  $ETX \leq Min$   
child node sends DIS to neighbor parent  
**end**
5. Repeat step 4

### 3.2 การตั้งค่าการทดลอง

ข้อมูลที่น่ามาวิจัยเป็นข้อมูลที่ได้จากการจำลองด้วยโปรแกรมจำลองเครือข่ายทีเอสซีเอช (TSCH Simulator) โดยทำการทดลองใช้โหนดลูก (Children Nodes) จำนวน 50, 100, 200 และ 400 โหนด และใช้วิธีการสร้างเครือข่ายแบบสุ่มตำแหน่ง (Random Location) โดยมีโหนดเริ่มต้น (Root Node) จำนวน 1 โหนด [10] โดยมีการทดลองมากกว่า 5 ครั้งในแต่ละสถานการณ์และหาค่าเฉลี่ยที่ได้

ตารางที่ 2: การกำหนดค่าในโปรแกรมจำลอง TSCH

Parameter	Value
Number of node	50, 100, 200, 400
Area	100m x 100m
Routing Protocol	RPL
Positioning Num Degree	5
Data rate	1 Packet = 1 Byte
Simulation Duration	60 seconds
Positioning Layout	Mesh

### 3.3 วิธีการวัดประสิทธิภาพ

ผู้วิจัยได้ทำการทดสอบผลกระทบของอัตราส่วนการจัดส่งแพ็กเก็ต (PDR: Packet Delivery Ratio) โดยการวัดค่าและคำนวณดังสมการที่ (2)

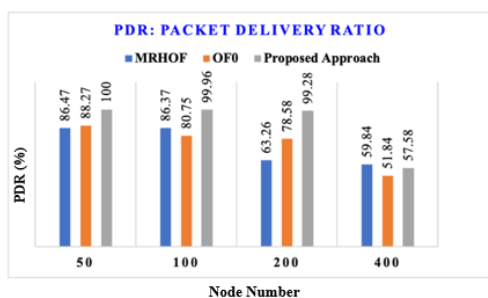
$$PDR = \frac{TPR}{TPS} \quad (2)$$

โดยที่ TPR คือข้อมูลที่ได้รับทั้งหมดจากปลายทาง (Total Packets Received) และ TPS คือข้อมูลที่ส่งไปทั้งหมดจากต้นทาง (Total Packet Sent)

## 4. ผลการทดลอง

### 4.1 อัตราส่วนการจัดส่งแพ็กเก็ต (PDR: Packet Delivery Ratio)

ผู้วิจัยทำการเปรียบเทียบผลกระทบของอัตราส่วนการจัดส่งแพ็กเก็ตระหว่างฟังก์ชันวัตถุประสงค์ที่นำเสนอกับไอเอฟซีโร และเอ็มอาร์เอชไอเอฟ ได้ผลการทดลองดังภาพที่ 4



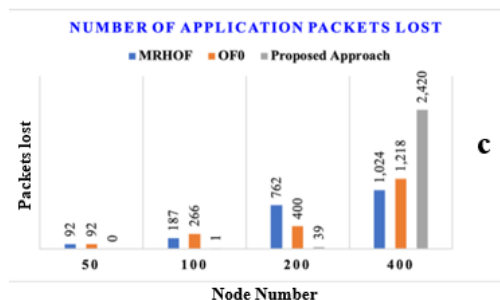
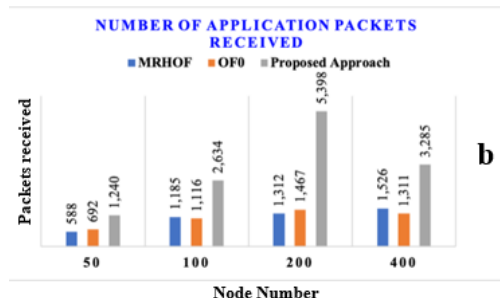
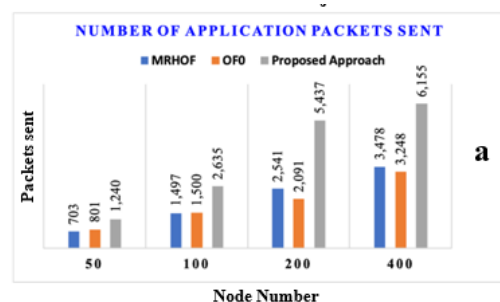
ภาพที่ 4: การเปรียบเทียบอัตราส่วนการจัดส่งแพ็กเก็ต

ผลการทดลองครั้งที่ 1 ถึงครั้งที่ 3 ด้วยจำนวนโหนดที่ 50, 100 และ 200 โหนด แสดงให้เห็นว่า ในเครือข่ายมีจำนวนไม่สูงเกิน 200 โหนด วิธีการที่นำเสนอสามารถ

ช่วยเพิ่มอัตราส่วนการจัดส่งแพ็กเก็ตได้อย่างมีประสิทธิภาพ เนื่องจากโหนดแม่ในพื้นที่ใกล้เคียงกันสามารถพิจารณาการรับโหนดลูกได้อย่างเหมาะสมมากขึ้นจากการแลกเปลี่ยนข้อความระหว่างโหนดเป็นระยะ

อย่างไรก็ตาม หากพิจารณาผลการทดลองครั้งที่ 4 ซึ่งมีจำนวนของโหนดที่ 400 โหนดนั้น พบว่า วิธีการที่นำเสนอมีอัตราการจัดส่งแพ็กเก็ตที่ใกล้เคียงกับวิธีที่นำมาเปรียบเทียบ ทั้งนี้ผู้วิจัยคาดว่าเกิดจากปริมาณของข้อความควบคุม (Control Message) ที่ใช้ในการแลกเปลี่ยนข้อมูลระหว่างโหนดของวิธีที่นำมาเสนอมีค่าสูงขึ้นจนเริ่มกระทบกับความแออัดของทราฟฟิกข้อมูล

### 4.2 จำนวนของแอปพลิเคชันแพ็กเก็ตที่รับส่งและสูญหาย



ภาพที่ 5: การเปรียบเทียบจำนวนของแอปพลิเคชันแพ็กเก็ตที่รับส่งและสูญหาย

จากภาพที่ 5 (a) จะเห็นได้ว่าจำนวนของแอปพลิเคชันแพ็กเก็ต (Application Packet) ที่ถูกส่งของวิธีที่นำเสนอ มีจำนวนมากกว่าแบบอื่น ซึ่งส่งผลมายังจำนวน

ของแอปพลิเคชันแพ็กเก็ตที่ถูกปรับ ตามภาพที่ 5 (b) ที่แสดงให้เห็นว่าจำนวนของแอปพลิเคชันแพ็กเก็ตที่ได้รับทั้งหมดของวิธีที่นำเสนอ มีจำนวนที่มากกว่าอีกสองแบบ

สำหรับจำนวนของแอปพลิเคชันแพ็กเก็ตที่สูญหาย จากภาพ 5 (c) จะเห็นได้ว่า วิธีที่นำเสนอมีจำนวนของแอปพลิเคชันแพ็กเก็ตที่สูญหาย ที่น้อยมากที่โหนด 50, 100 และ 200 แต่ที่โหนด 400 โหนดแบบเอ็มอาร์เอชไอเอฟจะทำได้ดีกว่า

จากผลการทดลองในภาพที่ 5 นี้แสดงให้เห็นว่าวิธีการที่นำเสนอสามารถช่วยให้การส่งแอปพลิเคชันแพ็กเก็ตมีความสำเร็จมากขึ้น เนื่องจากการปรับสมดุลของโหลดส่งผลให้ความแออัดของกราฟฟิกข้อมูลบนเครือข่ายมีน้อยลง เมื่อเปรียบเทียบกับวิธีการอื่น

แต่ในขณะเดียวกันในการทดลองที่จำนวน 400 โหนด ปริมาณของข้อความควบคุม (Control Message) ของวิธีที่นำเสนอ มีค่าสูงกว่าวิธีอื่นมาก จนทำให้กระทบกับอัตราส่วนของการส่งแพ็กเก็ตข้อมูลบนเครือข่ายเซนเซอร์ไร้สาย

## 5. สรุปผลการวิจัย

ในงานวิจัยนี้ผู้วิจัยได้ทำศึกษาแนวทางการปรับสมดุลโหลดในเครือข่ายเซนเซอร์ไร้สายที่ใช้โปรโตคอลอาร์พีแอล โดยการใช้การแลกเปลี่ยนข้อความระหว่างโหนดแม่ที่อยู่ในพื้นที่ใกล้เคียงกัน เพื่อให้แต่ละโหนดมีข้อมูลที่สามารถตัดสินใจตอบรับโหนดลูกที่ร้องขอการเชื่อมต่อได้อย่างมีประสิทธิภาพมากขึ้น

ผู้วิจัยได้ทำการเปรียบเทียบอัตราส่วนการจัดส่งแพ็กเก็ตและจำนวนของแอปพลิเคชันแพ็กเก็ตที่ถูกปรับส่งบนเครือข่าย โดยเปรียบเทียบวิธีการที่นำเสนอกับ ไอเอฟซีโร และเอ็มอาร์เอชไอเอฟซึ่งเป็นฟังก์ชันวัตถุประสงค์ที่เป็นมาตรฐานของโปรโตคอลอาร์พีแอล ด้วยการใช้เครื่องมือโปรแกรมจำลองเครือข่ายที่เอสซีเอส ผลการทดลองแสดงให้เห็นว่าวิธีการที่นำเสนอ มีอัตราส่วนการจัดส่งแพ็กเก็ตสำเร็จสูงกว่าวิธีการอื่นที่นำมาเปรียบเทียบ เนื่องจากการปรับสมดุลโหลดภายในเครือข่าย ช่วยลดโอกาสที่โหนดใดโหนดหนึ่งจะประสบกับความแออัดของกราฟฟิกข้อมูล ส่งผลให้การรับส่งแพ็กเก็ตบนเครือข่ายมีประสิทธิภาพมากขึ้น

อย่างไรก็ตามปริมาณของข้อความควบคุม (Control Message) บนเครือข่ายยังเป็นประเด็นที่ผู้วิจัยวางแผนจะศึกษาในอนาคต เนื่องจากด้วยวิธีการที่นำเสนอในงานวิจัยนี้ ปริมาณของข้อความควบคุมมีแนวโน้มที่จะสูงขึ้น จนกระทบกับอัตราส่วนของการส่งแพ็กเก็ตข้อมูลบนเครือข่ายเซนเซอร์ไร้สาย และผู้วิจัยวางแผนศึกษาเกี่ยวกับการใช้พลังงานของโหนดในเครือข่าย เมื่อมีการรับส่งข้อความควบคุมที่เพิ่มขึ้น เพื่อเพิ่มประสิทธิภาพของการใช้งานในทุกส่วน

## เอกสารอ้างอิง

- [1] Gaddour and A. Koubaa, "RPL in a nutshell: A survey", *Computer Networks*, Vol. 56, pp. 3163–3178, 2012.
- [2] N. Accettura, L. A. Grieco, G. Boggia, and P. Camarda, "Performance Analysis of the RPL Routing Protocol", *IEEE ICM*, pp. 767-772, 2011.
- [3] D. Wang, et al. "RPL Based Routing for Advanced Metering Infrastructure in Smart Grid", *IEEE ICC Workshops*, pp. 1-6, 2010.
- [4] B. Pavkovic, et al., "Multipath opportunistic RPL routing over IEEE 802.15.4," *ACM MSWiM*, pp. 179-186, 2010.
- [5] P. Kulkarni, et al., "A Self-Organising Mesh Networking Solution Based on Enhanced RPL for Smart Metering Communications," *IEEE WoWMoM*, pp. 1-6, 2011.
- [6] Y. Chen and N. Nasser, "Energy-Balancing Multipath Routing Protocol for Wireless Sensor Networks," *QSHINE*, pp. 21-24, 2016.
- [7] N. Pradeska, et al., "Performance analysis of objective function MRHOF and OF0 in routing protocol RPL IPV6 over low power wireless personal area networks (6LoWPAN)", *ICITEE*, pp. 1-6, 2016.
- [8] JP. Vasseur, et al., "Routing Metrics Used for Path Calculation in Low-Power and Lossy Networks", *Internet Engineering Task Force (IETF)*, 2012.
- [9] O. Gnawali, "The Minimum Rank with Hysteresis Objective Function", *Internet Engineering Task Force (IETF)*, 2012.
- [10] Z. Magubane, et al., "Performance Evaluation of ETXPC-RPL Routing Algorithm in IoT Network", *IMITEC*, pp. 1-8, 2020.



# โครงข่ายประสาทเทียมแบบคอนโวลูชันสำหรับการจำแนกภาพดอกกรัก

## Convolutional Neural Network for Image Classification of Crown Flowers

เกวลิน ขำนิพัฒน์ (Kewalin Khamnepat)<sup>1</sup> และอรอุมา พรีมาโมต (On-Uma Pramote)<sup>1</sup>

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพิบูลสงคราม

kewalin.k@psru.ac.th, onbee@psru.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ 1) เพื่อสร้างโมเดลจำแนกภาพด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชันสำหรับการจำแนกภาพดอกกรักดี ดอกกรักเสีย และ 2) เพื่อทดสอบประสิทธิภาพของโมเดลที่พัฒนา วิธีการประกอบด้วย 3 ส่วน ได้แก่ 1) การได้มาซึ่งภาพ 2) การสร้างโมเดลจำแนกดอกกรัก และ 3) การนำโมเดลไปใช้กับชุดข้อมูลภาพดอกกรักอื่น ๆ ชุดข้อมูลสำหรับสร้างโมเดลได้จากกล้องดิจิทัลมิลเลอร์เลส โดยใช้ดอกกรักสีขาวจำนวน 800 ดอก 2 กลุ่ม ได้แก่ กลุ่มดอกดี 400 ดอก และกลุ่มดอกเสีย 400 ดอก ใช้เป็นข้อมูลสำหรับฝึกสอน โมเดล 640 ดอก ข้อมูลสำหรับทดสอบโมเดล 160 ดอก โมเดลได้ค่าความถูกต้องร้อยละ 93 สร้างชุดข้อมูลภาพจากกล้องโทรศัพท์มือถือ และชุดข้อมูลภาพจากเซนเซอร์รับภาพร่วมกับเลนส์คุณภาพสูงของรายเบอร์รี่พาย โดยใช้โครงสร้างและค่าพารามิเตอร์ของโมเดลเดียวกัน และทำการเปรียบเทียบค่าความถูกต้องกับโมเดลที่สร้างจากเว็บทีโอเบิล แมชชีน พบว่าได้ค่าความถูกต้องร้อยละ 98

**คำสำคัญ:** การจำแนกประเภทภาพ โครงข่ายประสาทเทียมแบบคอนโวลูชัน ดอกกรัก

### Abstract

This research aims 1) to create an image classification model with a convolutional neural network for grading Crown Flowers and 2) to test the efficiency of the model. It has 3 steps 1) Image Acquisition 2) Create Modeling and 3) Using the model to another crown flowers data set. We create the model from the white Crown Flowers dataset from a mirrorless digital camera and macro lens.

The dataset has 2 classes i.e., Good and Bad each class has 400 flowers total are 800. We divided to train data are 640 flowers and test data are 160 flowers. The evaluation of the model's accuracy is 93%. Then, we use the CNN model with the Crown flowers dataset from Smartphone, Raspberry Pi, and High Camera. We compare the model to Teachable Machine. The result of accuracy is 98%.

**Keyword:** Image Classification, CNN, Crown Flowers.

### 1. บทนำ

การใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ในงานจำแนกประเภทรูปภาพ (Image Classification) เป็นเทคโนโลยีที่ได้รับความนิยมมากในปัจจุบัน เพราะภาพถ่ายเป็นข้อมูลที่เข้าใจได้ง่ายในการสื่อสารและอุปกรณ์สำหรับบันทึกภาพนั้นกลายเป็นส่วนหนึ่งในฟังก์ชันของโทรศัพท์มือถือซึ่งเป็นเครื่องมือที่จำเป็นของมนุษย์ ณ ขณะนี้ ส่งผลให้นำติดตัวได้ตลอด การจำแนกประเภทรูปภาพถูกนำไปใช้งานด้านต่าง ๆ ได้แก่ ด้านการศึกษา [1] ด้านธุรกิจ [2] ด้านกีฬา [3] ด้านความปลอดภัย [4] ด้านสุขภาพ [5] ด้านที่ส่งผลต่อประเทศไทยซึ่งรัฐบาลให้ความสำคัญและพยายามผลักดันให้เกิดกับประชากรหมู่มาก คือ ด้านเกษตรกรรม โดยมีงานวิจัยที่นำเทคโนโลยีการจำแนกประเภทรูปภาพไปประยุกต์ใช้ได้แก่ การจำแนกประเภทโรคเชื้อราของใบยางพารา [6] การวิเคราะห์คุณภาพของมะขามหวาน [7] การวิเคราะห์สายพันธุ์ผลไม้ [8] การวิเคราะห์ชนิดของแมลงศัตรูพืช [9] เป็นต้น จะเห็นได้ว่าเทคโนโลยีการจำแนกประเภทรูปภาพสามารถนำมาใช้กับงานได้หลากหลาย ผู้วิจัยจึงมีแนวคิดในการนำเทคโนโลยีนี้มาใช้

กับดอกกรัก โดยใช้จำแนกดอกดี ดอกเสีย ซึ่งการจำแนกจะส่งผลให้สามารถขายดอกกรักได้ราคาสูงขึ้น ซึ่งดอกกรักเป็นดอกไม้ที่นิยมใช้ในประเทศไทยมานานในกิจกรรมทางศาสนา หรืองานพิธีต่าง ๆ ดังนั้นจึงเป็นดอกไม้ที่ขายได้ตลอดทั้งปี ด้วยเหตุนี้เกษตรกรจึงหันมาเก็บดอกกรักขายเป็นอาชีพเสริม การซื้อขายแบ่งออกเป็น 2 แบบ ได้แก่ การขายเหมารวม และการขายคัดดอก ราคาขายของการขายคัดดอกจะสูงกว่าแต่เกษตรกรไม่นิยมเพราะต้องใช้ทักษะและดอกกรักมีขนาดเล็ก ดังนั้นงานวิจัยนี้จะเป็นประโยชน์ต่อเกษตรกรในส่วนของการเพิ่มรายได้ให้ดีขึ้น ช่วยส่งเสริมการมีอาชีพที่ 2 ของเกษตรกรได้

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 โครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงข่ายประสาทเทียมแบบคอนโวลูชัน แบ่งเป็น 3 ส่วน คือ 1) ส่วนรับข้อมูล (Input) ทำหน้าที่นำเข้ารูปภาพ 2) ส่วนการเรียนรู้คุณลักษณะเด่น (Feature Learning) ทำหน้าที่ สกัดคุณลักษณะเด่นของภาพ และ 3) ส่วนการจำแนก (Classification) ทำหน้าที่จำแนกวัตถุในภาพไปยังคลาสซึ่งพิจารณาจากคุณลักษณะเด่น

โครงข่ายประสาทเทียมแบบคอนโวลูชัน มีชั้น (Layer) สำคัญ ดังนี้ ชั้นคอนโวลูชัน (Convolution Layer) ทำหน้าที่คำนวณชั้นผลลัพธ์ด้วยวิธีการคอนโวลูชัน (Convolutional) ชั้นพูลลิ่ง (Pooling Layer) ทำหน้าที่เลือกค่าตัวแทนค่าตอบของชั้นผลลัพธ์ และชั้นเชื่อมต่อสมบูรณ์ (Fully Connected Layer) ทำหน้าที่แปลงเมตริกซ์ของตัวแทนชั้นค่าตอบให้เป็นเวกเตอร์แนวตั้ง และทำการเรียนรู้เพื่อจำแนกวัตถุ ผลลัพธ์ที่ได้คือคือคลาสของวัตถุ [10]

### 2.2 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่ใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันในการจำแนกประเภทรูปภาพดอกไม้มีดังตารางที่ 1

ตารางที่ 1: สรุปงานวิจัยจำแนกประเภทภาพดอกไม้

งานวิจัย	ชุดข้อมูลที่ใช้ (Dataset)	ค่าความถูกต้อง (Accuracy)
[11]	Oxford-102 flowers	98.6
[12]	flower images from the internet	78.0
[13]	Daisy, Dandelion, Sunflower, Rose and Tulip	95.0
[14]	Orchid	92.2

จากงานวิจัยข้างต้น ชุดข้อมูลที่ใช้เป็นดอกไม้ในต่างประเทศหรือกล้วยไม้ในไทยซึ่งยังไม่มีดอกกรักบรรจุเป็นชุดข้อมูล และการจำแนกดอกกรักจะพิจารณาจากส่วนของฐานดอกซึ่งแตกต่างจากภาพดอกไม้ที่อยู่ในชุดข้อมูล

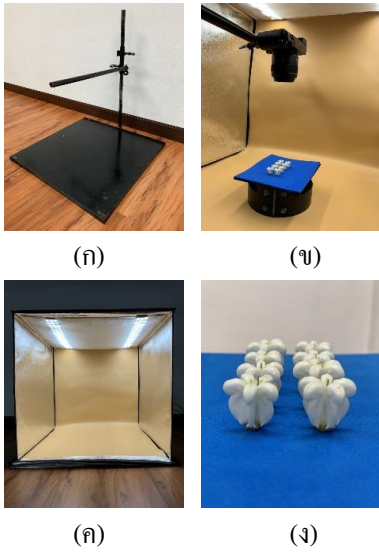
## 3. วิธีการดำเนินงานวิจัย

งานวิจัยนี้ประกอบด้วยขั้นตอนหลัก 3 ขั้นตอน 1) การได้มาซึ่งภาพ 2) การสร้างโมเดลจำแนกดอกกรัก และ 3) การนำโมเดลไปใช้กับชุดข้อมูลภาพดอกกรักอื่น ๆ

### 3.1 การได้มาซึ่งภาพ (Image Acquisitions)

การได้มาซึ่งภาพดอกกรักประกอบด้วย 2 ส่วน ได้แก่ การถ่ายภาพ (Image Capture) และ การเตรียมภาพ (Pre-process) โดยมีรายละเอียดดังนี้

3.1.1 การถ่ายภาพใช้กล้องดิจิทัลคอมมิวนิตี เลส ซีรี่ส์ หือพานาโซนิค รุ่นดีเอ็มซีจีเอ็กซ์ 85 (Panasonic DMC-GX85) และเลนส์มาโคร 45 มิลลิเมตร กำหนดค่าความไวแสง 200 ค่ารับแสง 2.8 ค่าความเร็วชัตเตอร์ 1/400 และโหมดกล้องแบบปรับตั้งค่าเอง (Manual) ภาพที่ 1 (ก) ฐานถ่ายภาพเป็นโครงเหล็กสำหรับติดตั้งกล้องดิจิทัลสามารถปรับระยะความสูงของกล้องกับวัตถุที่ต้องการถ่ายได้ ระยะห่างระหว่างกล้องและดอกกรักเท่ากับ 28 เซนติเมตร ภาพที่ 1 (ข) มุมถ่ายภาพดอกกรักเป็นมุมมองภาพจากบนลงล่าง (Top View) ภาพที่ 1 (ค) สตูดิโอ ลูกบาศก์ ขนาด 60 เซนติเมตร ใช้กำหนดสภาพแวดล้อมของแสง และภาพที่ 1 (ง) ดอกกรักถูกตรึงด้วยเข็มหมุด หนายฐานดอกกรักขึ้น เพราะใช้จำแนกดอกดี-ดอกเสียได้ชัดเจนกว่าบริเวณอื่น และใช้กระดาษสีน้ำเงินรองเป็นพื้นหลัง เพื่อให้เห็นดอกกรักได้ชัดขึ้น

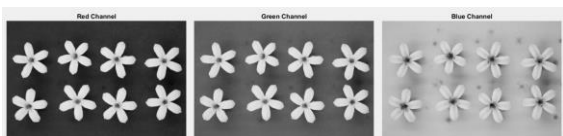


ภาพที่ 1: อุปกรณ์และการเก็บบันทึกภาพดอกกรัก

3.1.2 การเตรียมข้อมูลภาพเพื่อให้ได้ภาพที่เหมาะสมต่อการเรียนรู้ของโมเดล ภาพที่ได้รับจากกล้องเป็นไฟล์เจพีจี (JPG) กว้าง 4592 พิกเซล สูง 3064 พิกเซล และความละเอียดภาพ 180 ดิพีไอ (dpi) ซึ่งมีขนาดใหญ่และมีดอกกรัก 8 ดอก การเรียนรู้ของโมเดลเป็นการเรียนรู้แบบ 1 ภาพ 1 ดอก จำเป็นต้องตัดแบ่งภาพดอกกรักออกจากกันและปรับขนาดภาพให้มีความกว้าง ความสูงเท่ากันทุกภาพ โดยมี 2 ขั้นตอนดังนี้

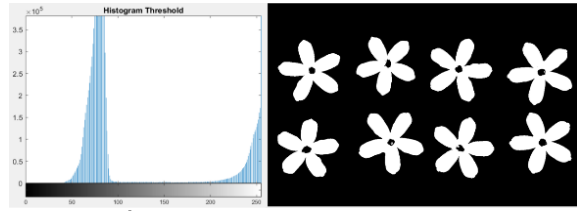
ขั้นตอนที่ 1 การตีกรอบดอกกรัก เป็นการสร้างพื้นที่สี่เหลี่ยมมุมฉากล้อมรอบดอกกรัก ในขั้นตอนนี้จะทำให้ทราบความกว้างและความสูงของดอกกรัก โดยมีรายละเอียด คือ

1) เลือกช่องสีที่แยกระหว่างสีดอกกรักและสีพื้นหลังออกจากกันได้มากที่สุด งานวิจัยนี้ใช้ช่องสีแดง (Red Channel) ดังภาพที่ 2



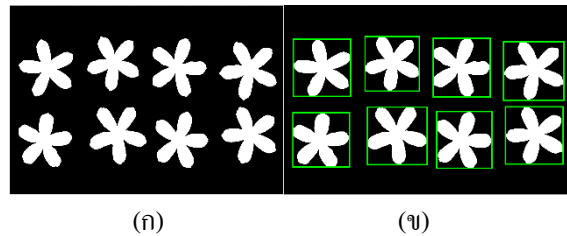
ภาพที่ 2: ช่องสีทั้ง 3 สี แดง เขียว และน้ำเงิน

2) ใช้เทคนิคฮิสโตแกรมกับภาพช่องสีแดง เพื่อหาค่าที่เหมาะสมสำหรับใช้แบ่งสีดอกกรักกับสีพื้นหลัง ดังภาพที่ 3



ภาพที่ 3: กราฟฮิสโตแกรมและผลลัพธ์จากการแบ่ง

3) ทำให้ภาพดอกกรักสมบูรณ์ เพื่อให้ได้ภาพดอกกรักเต็มดอกด้วยเทคนิคเติมหลุม (Fill holes) ดังภาพที่ 4 (ก) และ 4) ตีกรอบดอกกรักแต่ละดอกด้วยเทคนิคการหากรอบวัตถุในภาพ เพื่อให้ได้ค่าความกว้างและความสูงแต่ละดอก ดังภาพที่ 4 (ข)



ภาพที่ 4: (ก) ดอกกรักที่ไม่มีรู (ข) กรอบรอบดอกกรัก

ขั้นตอนที่ 2 หลังจากตีกรอบดอกกรักทุกดอกจะหาค่าความกว้างและค่าความสูงได้ ซึ่งแต่ละดอกจะมีค่าไม่เท่ากัน จำเป็นที่ต้องตีกรอบใหม่เพื่อให้ได้ขนาดภาพดอกกรักที่เท่ากัน โดยมีขั้นตอนคำนวณหาค่ากรอบใหม่ตามอัลกอริทึมดังภาพที่ 5

**Algorithm 1:** Cropping crown flowers by new border

```

Input: Crown flower images (Image)
Output: New Crown flower images (New_image)
// Load image and get high, width
1: for n=1 to 1000
2:    $Img_n \leftarrow \text{imread}(Image_n)$ 
3:    $[x_n, y_n, width_n, high_n] \leftarrow \text{Regionprops}(\text{BoundingBox}[Img_n])$ 
4: end for



// Compute new border
6:  $new\_high \leftarrow \text{Max}([All[high_n])$ 
7:  $new\_width \leftarrow \text{Max}([All[width_n])$ 

// Cropping border of crown flowers
9: for n=1 to 1000
10:   $[x_n, y_n, width_n, high_n] \leftarrow \text{Regionprops}(\text{BoundingBox}[Img_n])$ 
11:   $Dis\_width_n = (new\_width - width_n)/2$ 
12:   $Dis\_high_n = (new\_high - high_n)/2$ 
13:   $Pos\_x_n = \text{round}(x_n) - Dis\_width_n$ 
14:   $Pos\_y_n = \text{round}(y_n) - Dis\_high_n$ 
15:   $new\_img_n \leftarrow \text{imcrop}(Img_n[Pos\_x_n, Pos\_y_n, Dis\_width_n, Dis\_high_n])$ 
16: end for
    
```

ภาพที่ 5: อัลกอริทึมตัดภาพโดยคำนวณหาพื้นที่ที่กรอบใหม่

ภาพผลลัพธ์ที่ได้มีความกว้างและความสูง 1050 พิกเซล และมีความละเอียด 96 ดีพีไอ ทุกภาพ ภาพดอกกรักดี มีจำนวน 400 ภาพ คลาส 1 และภาพดอกกรักเสียมีจำนวน 400 ภาพ คลาส 0 ได้ชุดข้อมูลภาพดอกกรักรวม 800 ภาพ

ตารางที่ 2: คุณลักษณะของดอกกรัก

ประเภทดอก	คุณลักษณะ	ภาพ	คลาส
ดอกดี	กลีบดอกสีขาวล้วน มี 5 กลีบ ขนาดแต่ละกลีบมีความยาวใกล้เคียงกัน ฐานกลีบไม่ชัดเจน		1
ดอกเสีย	กลีบดอกมีรา หรือจำนวนกลีบ น้อย/มากกว่า 5 กลีบ หรือ มีบางกลีบที่ขนาดสั้น หรือ ฐานกลีบชัดเจน		0

3.2 การพัฒนาโมเดลจำแนก (Modeling)

ข้อมูลภาพที่ได้จะถูกนำมาใช้พัฒนาโมเดลจำแนกโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน ซึ่งมีขั้นตอนดังนี้ 1) การสร้างโมเดลด้วยชุดข้อมูลฝึกสอน และการปรับพารามิเตอร์ 2) การทดสอบ โมเดลด้วยชุดข้อมูลทดสอบ ข้อมูลภาพจะถูกแบ่งใช้เป็นชุดข้อมูลฝึกสอนจำนวน 640 ภาพ (ร้อยละ 80) และชุดข้อมูลทดสอบจำนวน 160 ภาพ (ร้อยละ 20) ซึ่งชุดข้อมูลทดสอบจะถูกนำเข้าโมเดลด้วยเทคนิค 10 fold Cross-Validation

3.3 การวัดประสิทธิภาพโมเดล

การวัดประสิทธิภาพของโมเดลในงานวิจัยนี้ใช้ตารางคอนฟิวชันเมตริกซ์ (Confusion Metric) ดังภาพที่ 6 เพื่อหาค่าความถูกต้อง (Accuracy) สมการ (1) [15]

**Predicted**

		Positive	Negative
		True Positive (TP)	False Negative (FN)
<b>Actual</b>	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

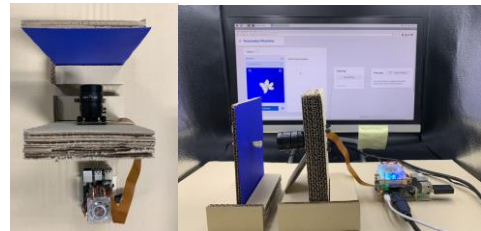
ภาพที่ 6: รายละเอียดของคอนฟิวชัน เมตริกซ์

$$Accuracy = (TP + TN) / (FP + FN) \quad (1)$$

3.4 ชุดข้อมูลดอกกรักอื่น (Others Data Set of Crown)

ในงานวิจัยนี้ได้ทำการเก็บภาพดอกกรักเพื่อสร้างชุดข้อมูลสำหรับทดสอบความทนทานของโมเดลที่สร้างขึ้น โดยใช้อุปกรณ์เก็บภาพ 2 ลักษณะดังนี้

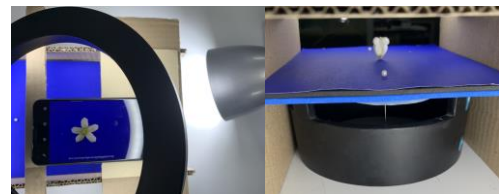
1) กล้องคุณภาพสูงที่ใช้กับราสเบอร์รี่พาย ความละเอียด 12 ล้านพิกเซล พร้อมเลนส์ทางยาวโฟกัส 6 มิลลิเมตร ระยะห่างระหว่างดอกกรักกับกล้องเท่ากับ 10 ซม. โดยถ่ายแบบควบคุมสภาพแวดล้อมด้วยสตูดิโอถ่ายภาพดังภาพที่ 7



ภาพที่ 7: การถ่ายภาพดอกกรักด้วยกล้องจากราสเบอร์รี่พาย

ภาพที่ได้มีความกว้างและความสูง 224 พิกเซล ความละเอียด 96 ดีพีไอ ถ่ายดอกกรักดี จำนวน 400 ภาพ และดอกกรักเสีย จำนวน 400 ภาพ รวม 800 ภาพ

2) กล้องของโทรศัพท์มือถือ ยี่ห้อ Realme X3 Super Zoom ด้วยโหมดถ่ายภาพแบบมาโคร ระยะห่างระหว่างดอกกรักกับกล้องเท่ากับ 5 ซม. โดยกำหนดสภาพแวดล้อมในการถ่ายด้วยไฟวงแหวนแอลอีดีสีขาวและหลอดไฟแสงขาว 4500 ลูเมน 1 หลอด ดังภาพที่ 8



ภาพที่ 8: การถ่ายภาพดอกกรักด้วยกล้องมือถือ

ภาพที่ได้มีความกว้างและความสูง 626 พิกเซล ความละเอียด 96 ดีพีไอ ถ่ายดอกกรักดี จำนวน 400 ภาพ และดอกกรักเสีย จำนวน 400 ภาพ รวม 800 ภาพ

#### 4. ผลการดำเนินการวิจัย

งานวิจัยนี้แบ่งการทดลองเป็น 3 ส่วนคือ 1) การหาพารามิเตอร์ทำให้โมเดลมีความถูกต้องในการจำแนกสูงที่สุด 2) การใช้ค่าพารามิเตอร์สร้างโมเดลจำแนกภาพดอกกุ๊ก โดยใช้ชุดข้อมูลอื่น 3) การเปรียบเทียบค่าความถูกต้องของโมเดลที่พัฒนาขึ้นกับโมเดลจาก ทีชเอเบิล แมชชีน (Teachable Machine) ซึ่งเป็นเว็บไซต์ให้บริการสร้างโมเดลจำแนกรูปภาพฟรี

##### 4.1 การหาพารามิเตอร์ (Finetune Parameter)

จำนวนเลเยอร์และพารามิเตอร์ที่ใช้ใน โมเดล CNN สำหรับการจำแนกภาพดอกกุ๊ก ดังภาพที่ 9 กำหนด epochs เท่ากับ 100, batch size เท่ากับ 32 โมเดลให้ค่าความถูกต้อง ร้อยละ 93

```

Model: "sequential_13"
-----
Layer (type)                Output Shape              Param #
-----
conv2d_31 (Conv2D)          (None, 222, 222, 8)      224
max_pooling2d_31 (MaxPoolin (None, 111, 111, 8)      0
g2D)
conv2d_32 (Conv2D)          (None, 107, 107, 16)     3216
max_pooling2d_32 (MaxPoolin (None, 53, 53, 16)      0
g2D)
dropout_32 (Dropout)        (None, 53, 53, 16)      0
conv2d_33 (Conv2D)          (None, 47, 47, 32)      25120
max_pooling2d_33 (MaxPoolin (None, 23, 23, 32)      0
g2D)
dropout_33 (Dropout)        (None, 23, 23, 32)      0
flatten_13 (Flatten)        (None, 16928)            0
dropout_34 (Dropout)        (None, 16928)            0
dense_26 (Dense)            (None, 64)               1083456
dense_27 (Dense)            (None, 1)                65
-----
Total params: 1,112,081
Trainable params: 1,112,081
Non-trainable params: 0
    
```

ภาพที่ 9: โครงสร้างของโมเดล CNN ที่ใช้จำแนกดอกกุ๊ก

##### 4.2 ประสิทธิภาพโมเดล CNN ในชุดข้อมูลอื่น

ในการใช้พารามิเตอร์และชั้น โครงสร้างเลเยอร์ของ โมเดล CNN ในงานวิจัยนี้กับชุดข้อมูลดอกกุ๊กจากกล้อง ราชเบอร์รี่พาย และชุดข้อมูลดอกกุ๊กที่จากกล้อง โทรศัพท์มือถือ ให้ค่าความถูกต้องร้อยละ 98

##### 4.3 การเปรียบเทียบกับโมเดลทีชเอเบิล แมชชีน

ในงานวิจัยนี้นำชุดข้อมูลภาพไปสร้าง โมเดลการเรียนรู้ จา ก เ รื บ ไ ช ต์ <https://teachablemachine.withgoogle.com/train/image> และกำหนดค่าพารามิเตอร์เช่นเดียวกับโมเดล CNN ภาพทั้ง 3 ชุดข้อมูลจะถูกปรับขนาดความกว้างและความสูงเป็น 224 พิกเซล ตามข้อกำหนดของการสร้างโมเดลบนเว็บไซต์

ตารางที่ 3: เปรียบเทียบค่าความถูกต้องของชุดข้อมูลภาพ

ชุดข้อมูลภาพ	ค่าความถูกต้อง (ร้อยละ)	
	CNN	Teachable Machine
ภาพจากกล้องดิจิทัล	93	90
ภาพจากกล้อง โทรศัพท์มือถือ	98	98
ภาพจากกล้อง ราชเบอร์รี่พาย	98	98

#### 5. สรุป

โมเดล CNN ที่พัฒนาในงานวิจัยนี้ สามารถใช้จำแนกภาพดอกกุ๊กดี ดอกกุ๊กเสียได้ โดยใช้ชุดข้อมูลในการทดลอง 3 ชุด ได้แก่ ชุดข้อมูลจากกล้องดิจิทัลมิลเลอร์เลส และเลนส์แบบมาโคร กล้องโทรศัพท์มือถือ โหมคเลนส์มาโคร และเซนเซอร์รับภาพของราชเบอร์รี่พายร่วมกับเลนส์คุณภาพสูง ซึ่งให้ค่าความละเอียดร้อยละ 93 98 และ 98 ทางผู้วิจัยวิเคราะห์สาเหตุที่ชุดข้อมูลภาพจากกล้องดิจิทัล ให้ค่าความถูกต้องต่ำกว่าเพราะภาพต้องถูกย่อขนาดลงมากกว่าชุดข้อมูลภาพจากกล้อง โทรศัพท์และชุดข้อมูลภาพจากเซนเซอร์รับภาพของราชเบอร์รี่พาย ส่งผลให้รายละเอียดถูกลดทอนไป แต่ก็มีความถูกต้องเกินร้อยละ 80 ทั้ง 3 ชุดข้อมูลแสดงให้เห็นว่าโมเดล CNN นี้สามารถใช้ได้กับชุดข้อมูลภาพดอกกุ๊กที่ได้จากหลายอุปกรณ์ และการเปรียบเทียบกับโมเดลที่สร้างจากเว็บไซต์ ทีชเอเบิล แมชชีนนั้นค่าความถูกต้องมีความแตกต่างกัน้อย ดังนั้นโมเดล CNN นี้สามารถนำไปใช้จำแนกประเภทภาพดอกกุ๊กดี-เสียได้อย่างน่าเชื่อถือ

## เอกสารอ้างอิง

- [1] จิราณี นวีรัตน์ อรวรรณ เขาวลิต และ ชามาพัชร วราวุฒิโรจน์, “แอปพลิเคชันการเรียนรู้คำศัพท์ภาษาอังกฤษสำหรับเด็ก,” *Proceeding of the 8th ASEAN Undergraduate Conference in Computing: AUCC*, กรุงเทพมหานคร, หน้าที่ 103-109, 2563.
- [2] ธัญชนก มะหังสา ณิชูโชติ พรหมฤทธิ์ และ สัจจาภรณ์ ไวจรรยา, “ระบบคัดเลือกรูปภาพสวยของสินค้าแบรนด์ BoxJourney โดยใช้เทคนิคการเรียนรู้เชิงลึก,” *Proceeding of the 8th ASEAN Undergraduate Conference in Computing: AUCC*, กรุงเทพมหานคร, หน้าที่ 27-33, 2563.
- [3] อรณิชา เพ็ชรรัตน์ รัชดาพร คณาวงษ์ และ จิตคารง ปรีชาสุข, “การตรวจจับ BIB ด้วยการวิเคราะห์รูปภาพโดยโครงข่ายประสาทแบบคอนโวลูชัน,” *Proceeding of the 8th ASEAN Undergraduate Conference in Computing: AUCC*, กรุงเทพมหานคร, หน้าที่ 263-269, 2563.
- [4] ปรมศร์ ทองชมพู และ สายัญญ์ สายยศ, “การตรวจจับและจำแนกยานพาหนะ ด้วยเทคนิคการเรียนรู้เชิงลึกจากกล้องวีดิโอรักษาความปลอดภัย กรณีศึกษาศูนย์ประชุมและแสดงสินค้านานาชาติขอนแก่น,” *Proceeding of the 16th National Conference on Computing and Information Technology: NCCIT*, กรุงเทพมหานคร, หน้า 86-91, 2563.
- [5] ณัฐนนท์ ลำสมุทร และสมปอง เวฬุวนาธร, “การจำแนกภาพเอกสารเรย์คอมพิวเตอร์โรคหลอดเลือดสมองด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน,” *Proceeding of the 17th National Conference on Computing and Information Technology: NCCIT*, กรุงเทพมหานคร, หน้า 43-48, 2564.
- [6] จิวรีย์ อาดา จิรพิริ แก้วประชา จิตคารง ปรีชาสุข และ รัชดาพร คณาวงษ์, “การจำแนกประเภทโรคเชื้อราของใบยางพารา ด้วยการวิเคราะห์จากภาพถ่าย,” *Proceeding of the 8th ASEAN Undergraduate Conference in Computing: AUCC*, กรุงเทพมหานคร, หน้าที่ 148-154, 2563.
- [7] นพรัตน์ คาทิพย์ และ เทวิน ธนะวงษ์, “ระบบวิเคราะห์คุณภาพของมะขามหวานด้วยเทคนิคการเรียนรู้เชิงลึก,” *Proceeding of the 8th ASEAN Undergraduate Conference in Computing: AUCC*, กรุงเทพมหานคร, หน้าที่ 505-511, 2563.
- [8] นันทนา แดงชีว และ เทวิน ธนะวงษ์, “ระบบวิเคราะห์ชนิดสายพันธุ์ผลไม้ด้วยเทคนิคการประมวลผลภาพ,” *Proceeding of the 8th ASEAN Undergraduate Conference in Computing: AUCC*, กรุงเทพมหานคร, หน้าที่ 512-517, 2563.
- [9] นนทวัฒน์ นกเพ็ชร และ เทวิน ธนะวงษ์, “ระบบวิเคราะห์ชนิดของแมลงศัตรูพืชโดยใช้เทคนิคการเรียนรู้เชิงลึก,” *Proceeding of the 9th ASEAN Undergraduate Conference in Computing: AUCC*, กรุงเทพมหานคร, หน้าที่ 1291-1297, 2564.
- [10] Pimpa Cheewaprakobkit. “Improving the Performance of an Image Classification with Convolutional Neural Network Model by Using Image Augmentations Technique” , *TNI Journal of Engineering and Technology*, Vol.7, No.1, pp 59-64. January - June 2019.
- [11] N. Alipour, O. Tarkhaneh, M. Awrangjeb and H. Tian, “Flower Image Classification Using Deep Convolutional Neural Network,” *2021 7th International Conference on Web Research (ICWR)*, pp. 1-4, 2021.
- [12] F. Hu, F. Yao and C. Pu, “Learning Salient Features for Flower Classification Using Convolutional Neural Network,” *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS)*, pp. 476-479, 2020.
- [13] S. Giraddi, S. Seeri, P. S. Hiremath and J. G.N, “Flower Classification using Deep Learning models,” *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, pp. 130-133, 2020.
- [14] รสริน ตรีธนะสมบัติ และ เทวิน ธนะวงษ์, “ระบบวิเคราะห์พันธุ์กล้วยไม้ด้วยเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน,” *Proceeding of the 8th ASEAN Undergraduate Conference in Computing: AUCC*, กรุงเทพมหานคร, หน้าที่ 1337-1343, 2564.
- [15] ไพสิฐ วัฒนภักทราภรณ์ ณิชูโชติ พรหมฤทธิ์ และ สัจจาภรณ์ ไวจรรยา, “การเพิ่มประสิทธิภาพการตรวจจับป้ายสำหรับผู้พิการทางสายตาโดยใช้เทคนิคการขยายภาพ,” *Proceeding of the 8th ASEAN Undergraduate Conference in Computing: AUCC*, กรุงเทพมหานคร, หน้าที่ 173-179, 2563.

# ตู้กดน้ำลดสัมผัสด้วยอินเทอร์เน็ตประสานสรรพสิ่ง

## Contactless Water Dispenser Based on Internet of Things

วุฒิชัย ชุ่มเย็น (Wuttachai Chumyen)<sup>1</sup> และเทพฤทธิ์ บัณฑิตวัฒนาวงศ์ (Thepparit Banditwattanawong)<sup>2</sup>

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

<sup>1</sup>wuttachai.c@ku.th, <sup>2</sup>fscitr@ku.ac.th

### บทคัดย่อ

เนื่องด้วยการแพร่ระบาดของไวรัสโควิด 19 ผู้วิจัยจึงเสนอระบบตู้กดน้ำที่ใช้อินเทอร์เน็ตประสานสรรพสิ่งเพื่อลดการสัมผัสวัตถุร่วมกันกับผู้ขายและผู้ซื้อรายอื่นการใช้งานระบบค่อนข้างสะดวกโดยการนำแก้วน้ำมาไว้ในระยะทำงานของตัวรับรู้อัลตราโซนิกของผู้และสั่งซื้อผ่านโปรแกรมประยุกต์เคลื่อนที่ ระบบสามารถตรวจสอบปริมาณน้ำคงเหลือ อุณหภูมิและความชื้นภายในตู้จากไฟฟ้าลัดวงจรและน้ำรั่ว ผู้ขายสามารถกำหนดช่วงเวลาเปิดให้บริการของระบบ ตรวจสอบสถานะการทำงานของแต่ละตู้ กำหนดราคาขาย และดูแลประวัติการจำหน่ายน้ำ ผลการทดลองใช้งานระบบโดยผู้ใช้งาน 20 คนพบว่ามีความพึงพอใจในภาพรวมเท่ากับ 4.59

**คำสำคัญ:** อินเทอร์เน็ตประสานสรรพสิ่ง โปรแกรมประยุกต์เคลื่อนที่ โปรแกรมประยุกต์บนเว็บ ตู้กดน้ำ โควิด19

### Abstract

Due to the spread of COVID-19, we propose a water dispenser system using Internet of Things to reduce the exposure of the same objects by sellers and other buyers. It is easy to use the system by bringing a glass of water within the working range of the dispenser's ultrasonic sensor and ordering via a mobile application. The system can check the remaining water volume and working temperature and humidity inside the cabinet that might be caused by short circuit and water leakage. The seller can

set the operating hours of the system, check the working status of each dispenser, set selling prices, and view water selling history. Evaluation results by 20 users showed that an overall satisfaction score was 4.59.

**Keywords:** Internet of Things, mobile application, web application, water dispenser, Covid-19.

### 1. บทนำ

ปัจจุบันคนไทยจำนวนมากนิยมใช้บริการเครื่องดื่มจากตู้กดน้ำ เช่น ตู้กดน้ำในร้านสะดวกซื้อ ตู้กดน้ำตามงานนิทรรศการและงานแสดงสินค้าบริการต่างๆ ตู้กดน้ำเหล่านี้ต้องมีพนักงานประจำที่คอยดูแลการขายเครื่องดื่ม ทำให้มีค่าใช้จ่ายสำหรับการจ้างพนักงานและมีโอกาสสูงที่ผู้ขายและผู้ซื้อต้องสัมผัสด้วยใบเดียวกันอันนำไปสู่การแพร่เชื้อโควิด 19 ได้ ผลกระทบจากปัญหาเหล่านี้สามารถบรรเทาได้ด้วยการใช้ตู้กดน้ำลดสัมผัสด้วยอินเทอร์เน็ตประสานสรรพสิ่งที่เหมาะสมในบทความนี้

ในการทำงานของตู้กดน้ำ ผู้ซื้อเพียงแค่นำภาชนะส่วนตัวสำหรับใส่น้ำมาวางไว้ในตำแหน่งจ่ายน้ำ ตู้กดน้ำก็จะสามารถตรวจสอบระยะของภาชนะ จากนั้นผู้ใช้สามารถเลือกปริมาณน้ำที่ต้องการซื้อและชำระเงินผ่านโปรแกรมประยุกต์เคลื่อนที่ (mobile application) นอกจากนี้ตู้กดน้ำยังสามารถวัดปริมาณน้ำคงเหลือ วัดความชื้นและอุณหภูมิภายในตู้เพื่อให้ผู้บริหารระบบตรวจสอบการทำงานภายในตู้ว่าเกิดปัญหาน้ำรั่วและไฟฟ้าลัดวงจรหรือไม่และยังสามารถดูรายงานสถิติการใช้ตู้กดน้ำได้

## 2. เทคโนโลยีและระบบที่เกี่ยวข้อง

### 2.1 เทคโนโลยีที่เกี่ยวข้อง

อินเทอร์เน็ตประสานสรรพสิ่งมีวัตถุประสงค์เพื่อ การสร้างโครงข่ายเพื่อเชื่อมโยงอุปกรณ์ที่มีมาตรฐาน แตกต่างกันที่พัฒนาโดยผู้ผลิตที่มีเทคโนโลยีแตกต่างกัน ให้สามารถ สื่อสารกันได้ในรูปแบบที่หลากหลายมากยิ่งขึ้น [1] เทคโนโลยีอินเทอร์เน็ตประสานสรรพสิ่งที่ใช้ ในการพัฒนาตู้กดน้ำลดสัมผัสได้แก่

2.1.1 อะคูโน (Arduino) บอร์ดอะคูโน ยูโน ตัว ประมวลผล 16 เมกะเฮิร์ต หน่วยความจำข้อมูล 2 กิโลไบต์ เอสแรม (SRAM) 32 กิโลไบต์เฟรช อินพุต/ เอาท์พุต 14 พอร์ต [2]

2.1.2 ราสเบอร์รี่พาย (Raspberry Pi) พอร์ตยูเอสบี (USB 3.0) จำนวน 2 พอร์ตยูเอสบี (USB 2.0) จำนวน 2 พอร์ตไวไฟ (WiFi 802.11b/g/n/ac) ใช้กับอะแดปเตอร์ 5 โวลต์ 3 แอมป์ ชั่วต่อแบบยูเอสบีซี (USB-C) [3]

2.1.2 จอแอลซีดี (LCD) 480x320 พิกเซล 16 บิต

2.1.3 อาร์ทีซี (RTC) จุดต่อแบบไอดีซี (IDC) ตัวผู้ และตัวเมีย ให้ข้อมูลเวลาวัน วันที่ เดือน ปี ถึง ปี ค.ศ. 2099 [4]

2.1.4 อี เอ็ม ป ร อ ม (EEPROM) ความจุ 32 กิโลไบต์ เชื่อมต่อผ่านระบบบัสไอทูซี (I2C)

2.1.5 ปุ่มลมม่านหุ้มเส้นผ่าศูนย์กลาง 9.2 มิลลิเมตร ขนาด 85x40x32 มิลลิเมตร (ยาว \* กว้าง \* สูง) [5]

2.1.6 รีเลย์ (Relay) 5 โวลต์ใช้สำหรับขับปั้มน้ำ ขนาดเล็กซื้อผ่านอินเทอร์เน็ต ราคา 125 บาท

2.1.7 สเตรนเกจ (Strain gauge) ขนาด 1.75 x 1.5 นิ้วรองรับน้ำหนักการกดได้ 100 กรัมถึง 10 กิโลกรัม

2.1.8 เซเวนเซกเมน (7 segment) แบบ 4 อักขร

2.1.9 ปุ่มกด (Button) รูปร่างวงกลมสีแดง

2.1.10 ตัวรับรู้ (sensor) อุณหภูมิความชื้นดีเอชแอล (DHT) วัดความชื้นตั้งแต่ 0 – 100 เปอร์เซ็นต์อาร์เอช (RH) วัดอุณหภูมิ -40 ถึง 80 องศาเซลเซียส [6]

2.1.11 ตัวรับรู้อัลตราโซนิก (Ultrasonic Sensors) ระยะสูงสุดที่สามารถวัดได้ 4 เมตร และระยะต่ำสุดที่สามารถวัดได้ 0.02 เมตร

2.1.12 การเชื่อมต่อระหว่างบอร์ดของอะคูโนและ ราสเบอร์รี่พายใช้พอร์ตซีเรียลยูเอสบี ยูเอสบี (Serial UART (USB))

2.1.13 กูเกิลไฟเบอร์เบส (Google Firebase) เป็น ฐานข้อมูลแบบโนเอสคิวแอล (No SQL) ซึ่งเก็บข้อมูล อยู่ในรูปเจสัน (JSON) โดยโปรแกรมประยุกต์มือถือจะ ทำการเชื่อมต่อเพื่อบันทึกค่าลงบนกูเกิลไฟเบอร์เบสซึ่งเป็นระบบคลาวด์ไฟสโตนให้ผู้กดน้ำประมวลคำสั่งซื้อ เมื่อมีข้อมูลการสั่งซื้อเพิ่มเข้ามา

2.1.14 แซมมาริน (Xamarin) ใช้พัฒนาโปรแกรม ประยุกต์เคลื่อนที่ (Mobile Application) บนระบบ ปฏิบัติการแอนดรอยด์ (Android) และ ไอโอเอส (IOS) ด้วยภาษาซีชาร์ป (C#)

### 2.2 ระบบและงานวิจัยที่เกี่ยวข้อง

ตู้กดกาแฟแบบบริการตนเอง [7] มีลักษณะของ เครื่องกดกาแฟคล้าย ๆ กับ เครื่องกดกาแฟอัตโนมัติที่ พนักงานใช้ทั่วไป เมื่อเลือกเครื่องดื่มที่ต้องการเรียบร้อยแล้วก็เข้าสู่ขั้นตอนในการชำระเงินผ่านธนาคารหรือ กระเป๋าอิเล็กทรอนิกส์ อย่างไรก็ตามทั้งการเลือกเมนู เครื่องดื่มและการชำระเงินต้องอาศัยหน้าจอสัมผัสซึ่งมีความเสี่ยงด้านสุขอนามัย

ตู้จำหน่ายขนมและเครื่องดื่มอัตโนมัติ [8] ผู้ซื้อ สามารถเลือกสินค้าผ่านหน้าจอสัมผัสที่เครื่องและ สามารถสแกนจ่ายเงินที่หน้าตู้ผ่านคิวอาร์โค้ด อย่างไรก็ตามผู้ซื้อไม่สามารถสั่งซื้อสินค้าผ่านโปรแกรมประยุกต์เคลื่อนที่ได้และยังต้องมีการสัมผัสหน้าจอร่วมกันกับผู้ซื้อรายอื่นทำให้เกิดความเสี่ยง

ตู้กดน้ำหยอดเหรียญระบบอาร์โอ (RO) ขนาด 200 ลิตร [9] เป็นตู้กดน้ำที่พบตามชุมชนหรือตามหอพัก นักศึกษา การใช้ตู้กดน้ำต้องทำการหยอดเหรียญเพื่อให้ได้ปริมาณน้ำที่ต้องการ ระบบไม่มีการทำงานผ่าน อุปกรณ์ที่ช่วยลดการสัมผัส เช่น โปรแกรมประยุกต์เคลื่อนที่

ตู้กดน้ำในร้านสะดวก [10] ผู้มีน้ำอยู่ด้านบนของผู้ ให้ถูกค้ำมองเห็นได้ เมื่อต้องการกดน้ำจะต้องเลือกหยิบ แก้วน้ำตามปริมาณที่ต้องการเพื่อกดน้ำ จากนั้นทำการ ชำระเงินที่จุดชำระเงิน จึงยังมีความเสี่ยงจากแก้วน้ำที่วาง



ให้ลูกค้าเลือกจากปณเป็อนเชื้อโรคที่แพร่กระจาย ณ บริเวณจุดจำหน่าย การออกแบบอุปกรณ์จ่ายอาหาร ยา และระบบอัตโนมัติ [11] ช่วยเหลือผู้ที่ได้รับผลกระทบ จากภัยพิบัติทางธรรมชาติโดยใช้วิธีการจัดกลุ่มเคมีนแบบคลุมเครือ (fuzzy K-means clustering) เพื่อจัดหมวดหมู่และวินิจฉัยเอนทีตี้ในรูปแบบอัจฉริยะเพื่อการกระจายอาหาร ยา และน้ำที่เหมาะสมที่สุด หากมีวัตถุประสงค์ผ่านตัวรับรู้

ตรวจจับได้ว่าเป็นมนุษย์ มีการใช้เทคนิคนี้เพื่อกำจัด การจ่ายอาหาร เนื่องจากแต่ละคนต้องให้อาหารของตัวเอง เป็นเวลา 12 ชั่วโมง ในขั้นต้น บุคคลจะกดปุ่มเพื่อให้อาหารหรือยา ระบบจะจ่ายรายการตามลำดับหลังจากวิเคราะห์วิดีโอที่ถ่ายโดยกล้องจะถูกแปลงเป็นภาพและ นำเฟรมที่เข้าซ้ยนออก จากนั้นการตรวจจับใบหน้า มนุษย์จากวัตถุพื้นหลังจึงทำได้โดยใช้เทคนิคการตรวจจับสีผิวตามค่า fuzzy K-means clustering

การประหยัดพลังงานของระบบจ่ายน้ำที่ใช้ อินเทอร์เน็ตประสานสรรพสิ่ง [12] อาศัยโมดูลการ สื่อสารของเครื่องจ่ายน้ำเชิงพาณิชย์และเกตเวย์เพื่อ รวบรวมอุณหภูมิของถังน้ำร้อน น้ำอุ่น และน้ำเย็น และ ใช้แคลมป์มิเตอร์เชิงพาณิชย์เพื่อประเมินการใช้พลังงาน ของตู้กดน้ำเพื่อเรียนรู้การทำงานของตู้กดน้ำและ พฤติกรรมของผู้ใช้เพื่อนำมาปรับพารามิเตอร์ในการเพิ่ม ประสิทธิภาพการใช้พลังงานของเครื่องจ่ายน้ำ

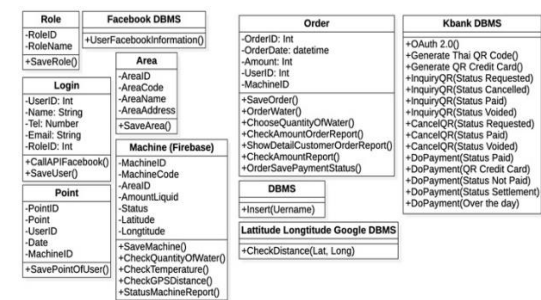
โปรแกรมประยุกต์อินเทอร์เน็ตประสานสรรพสิ่ง สำหรับเครื่องจ่ายน้ำในการตรวจจับระดับน้ำอัตโนมัติ [13] ใช้ตัวรับรู้ระดับน้ำซึ่งเมื่อลดลงจนถึงระดับการเติม ราวเบอร์รี่พาย 3 เพื่อส่งข้อมูลตำแหน่งของตู้ไปยัง คลาวด์เพื่อติดต่อกับโปรแกรมประยุกต์เคลื่อนที่ที่ติดตั้ง ในสมาร์ตโฟนของบุคคลที่รับผิดชอบในการเติมน้ำ

### 3. การวิเคราะห์ห้ออกแบบระบบ

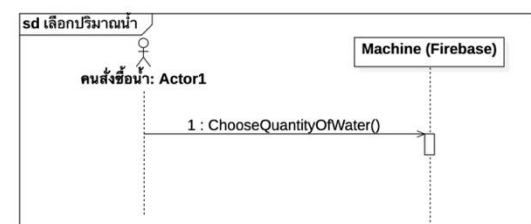


ภาพที่ 1: แผนภาพยูสเคสตู้กดน้ำลดส้ผลิตด้วยอินเทอร์เน็ตประสาน สรรพสิ่ง

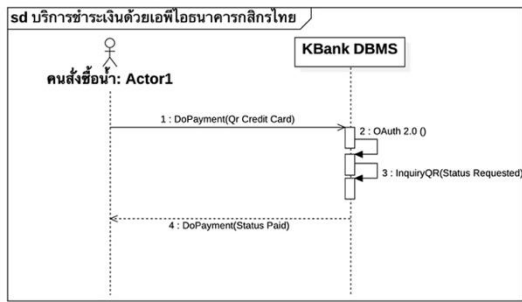
เมื่อลงบันทึกเข้าสู่ระบบด้วยบัญชีผู้ใช้เฟซบุ๊กหรือ บัญชีส่วนตัวของระบบ โดยตรงที่เคลงทะเบียน เรียบร้อยแล้ว ระบบจะทำการรับคำสั่งจากการกดสั่งซ้ยน น้ำและค้นหาตู้กดน้ำด้วยจีพีเอส จากนั้นทำการชำระเงิน ผ่านเอพีไอธนาคารกสิกรไทย (KBank API) หลังจากนั้นระบบจะทำการตรวจสอบปริมาณน้ำ ตรวจสอบ อุณหภูมิความชื้นภายในตู้ ตรวจสอบระยะภาชนะบรรจุ น้ำ เมื่อพร้อมแล้วจึงทำการจ่ายน้ำ และในส่วน ของเจ้าหน้าที่ดูแลระบบจะสามารถดูรายงานจำนวนการ สั่งซ้ยนน้ำได้ โดยการค้นหาจากวันที่ทำการสั่งซ้ยน หรือจากรายชื่อคนสั่งซ้ยนน้ำ สามารถดูเพิ่ม ลบ แก้ไข และรายงานรายละเอียดคนสั่งซ้ยน จำนวนของการสั่งซ้ยน และยอดจำนวนของการสั่งซ้ยนได้



ภาพที่ 2: แผนภาพคลาสของตู้กดน้ำ



ภาพที่ 3: แผนภาพซีควเอนสเลือกปริมาณน้ำ



ภาพที่ 4: แผนภาพซีควเอนสบริการชำระเงินด้วยเอทีไอธนาคารกสิกรไทย

## 4. ผลการสร้างระบบ

### 4.1 ต้นแบบตู้กดน้ำลดสัมผัสด้วยอินเทอร์เน็ตประสานสรรพสิ่ง



ภาพที่ 5: ต้นแบบตู้กดน้ำลดสัมผัสด้วยอินเทอร์เน็ตประสานสรรพสิ่ง

ภาพที่ 5 แสดงต้นแบบตู้กดน้ำลดสัมผัสด้วยอินเทอร์เน็ตประสานสรรพสิ่งที่ประกอบไปด้วยปุ่มเปิดปิดตู้กดน้ำ โดยที่เมื่อกดปุ่มเปิดปิดสถานะของหลอดแอลอีดีจะถูกเก็บไว้ในอาร์ทีซี/อีเอ็ปรอม ทำให้เปิดไฟแล้วค่าหลอดแอลอีดีไม่หายไปจากหน่วยความจำในตัวเครื่องตู้กดน้ำ และตัวรับรู้อุณหภูมิความชื้นภายในตู้ จะทำการตรวจสอบว่าตู้เหมาะสมกับการใช้งานหรือไม่ ถ้าไม่เหมาะสมแก่การใช้งานจะทำการแจ้งเตือนที่หน้าจอแอลซีดี และที่หน้าจอแอลซีดีจะทำการแสดงคิวของรายการที่มีคนสั่งซื้อเข้ามาตามลำดับ สมมติขาย ก สั่งซื้อน้ำดื่มเป็นลำดับแรกจำนวน 10 แก้ว ระบบจะแสดงคิวว่าจะจ่ายน้ำขาย ก เป็นคนแรก เมื่อตู้กดน้ำทำการจ่ายน้ำไปจนครบปริมาณ 1 แก้วระบบจะแจ้งเตือนให้นาย ก ทำการยกแก้วใหม่มาได้ และเมื่อตู้กดน้ำจะจ่ายน้ำขาย ก ตู้กดน้ำจะทำการใช้เครื่องชั่งเพื่อรู้ปริมาณน้ำคงเหลือของผู้ว่าเพียงพอต่อการจ่ายน้ำหรือไม่ และใช้ตัวรับรู้อัตราโซนิคตรวจจบบรรยากาศของแก้วน้ำ เมื่อแก้ว

น้ำอยู่ในระยะที่สามารถจ่ายน้ำได้ ตู้กดน้ำก็จะทำการจ่ายน้ำโดยผู้ใช้ป้อนเงินในการคั่นน้ำจากถังเก็บน้ำจ่ายให้แก่ลูกค้าต่อไป ส่วนสำหรับการสั่งซื้อน้ำผ่านตู้กดน้ำนั้น จะต้องทำการสั่งซื้อผ่านโปรแกรมประยุกต์เคลื่อนที่เท่านั้น

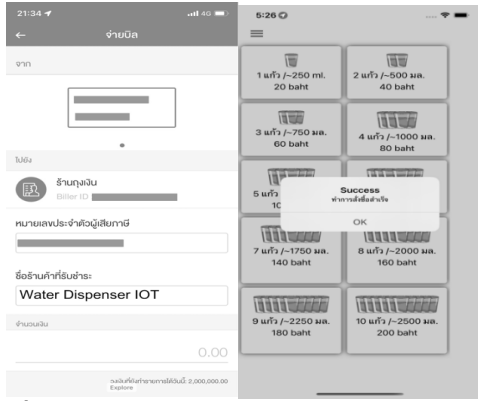
### 4.2 โปรแกรมประยุกต์

โปรแกรมประยุกต์เคลื่อนที่บนระบบปฏิบัติการไอโอเอสและแอนดรอยด์จะเริ่มต้นให้กรอกชื่อผู้ใช้และรหัสผ่านถ้าได้ทำการสมัครสมาชิกจะสามารถทำการกรอกชื่อผู้ใช้และรหัสผ่านแล้ว แต่ถ้าไม่อยากจะสมัครสมาชิกก็สามารถลงบันทึกผ่านระบบของเฟซบุ๊กได้ เมื่อลงบันทึกเข้าสำเร็จจะปรากฏหน้าจอตามภาพที่ 6 (ซ้าย) ที่ประกอบไปด้วยรายการของเมนูการสั่งซื้อ (Ordering) ประวัติการสั่งซื้อ (History) และการลงบันทึกออกจากระบบ (Log out) และถ้ากดเมนูการสั่งซื้อจะปรากฏเมนูสำหรับเลือกปริมาณน้ำดื่มที่ต้องการตามภาพที่ 6 (ขวา)



ภาพที่ 6: หน้าจอแรกหลังลงบันทึกเข้าระบบ (ซ้าย) และหน้าจอแสดงรายการปริมาณน้ำดื่มที่ต้องการ (ขวา)

เมื่อผู้ใช้ทำการกดเลือกปริมาณของน้ำดื่มที่ต้องการระบบจะทำการตรวจสอบตำแหน่งของตู้กดน้ำที่อยู่ใกล้ผู้ซื้อที่สุดและเข้าสู่ขั้นตอนการชำระเงินผ่านเว็บเอทีไอของธนาคารตามภาพที่ 7 (ซ้าย) ส่วนภาพที่ 7 (ขวา) เมื่อทำการชำระเงินสำเร็จ ระบบจะทำการแจ้งเตือนว่าทำการสั่งซื้อสำเร็จ ส่วนไบโเสิร์ฟระบบของธนาคารจะทำการบันทึกลงในคลังรูปภาพของผู้ใช้โทรศัพท์เคลื่อนที่



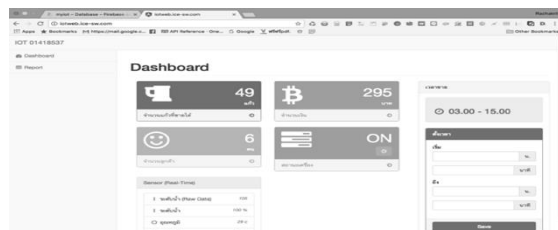
ภาพที่ 7: หน้าจอแสดงการชำระเงินจากผู้ใช้งาน (ชาย) และหน้าจอแสดงการชำระเงินสำเร็จ (ขวา)



ภาพที่ 8: หน้าจอประวัติการใช้บริการของผู้ซื้อ

ส่วนภาพที่ 8 แสดงหน้าจอประวัติการใช้บริการที่ดูผ่านเมนูประวัติการสั่งซื้อ ซึ่งแสดงปริมาณน้ำที่สั่งซื้อ และวันเวลาที่ทำการสั่งซื้อตามวันเวลาล่าสุดตามลำดับ

หน้าเว็บผู้ดูแลระบบแสดงดังภาพที่ 9 ซึ่งประกอบด้วยแผงหน้าปัด (Dashboard) ที่แสดงรายละเอียดของจำนวนแก้วที่ขายได้ จำนวนเงิน จำนวนลูกค้าที่ใช้บริการ และสถานะเครื่องของแต่ละตู้ โดยสามารถกำหนดเวลาในการขายหรือเปิดปิดตู้ได้ และมีตารางแสดงข้อมูลจากตัวรับรู้ต่าง ๆ ได้แก่ ระดับน้ำ อุณหภูมิ ความชื้น



ภาพที่ 9: หน้าจอแผงหน้าปัดของผู้ดูแลระบบ

### 5. การประเมินผล

ผู้วิจัยทำการวัดความพึงพอใจของผู้ใช้ระบบจำนวน 20 คน แบ่งเป็นผู้ซื้อ 18 คน และผู้ดูแลระบบ 2 คน มีรายละเอียดของผลการประเมินดังในตารางที่ 1 โดยมีคะแนนเต็มสำหรับเกณฑ์แต่ละข้อเท่ากับ 5 พบว่าค่าความพึงพอใจเฉลี่ยคิดเป็น 4.59 และส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับ 0.26 ซึ่งโดยรวมถือว่าอยู่ในระดับที่ดี

ตารางที่ 1: ผลการวัดความพึงพอใจของผู้ใช้ใน โปรแกรมประยุกต์

เกณฑ์การประเมิน	$\bar{X}$	S.D.
1) ความถูกต้องและน่าเชื่อถือ ของระบบ (Accuracy)	4.8	0.34
2) ความง่ายในการใช้ระบบ (Ease of Use)	4.9	0.33
3) ความรวดเร็วในการทำงาน ของระบบ (Timeliness)	4.8	0.31
4) ฟังก์ชันการทำงานของระบบ (Functionality)	4.9	0.31
5) ประโยชน์จากการใช้ระบบ (Usefulness)	4.3	0.25
6) ความสมบูรณ์ของโปรแกรมประยุกต์ (Completeness)	4.5	0.24
7) ความทันสมัยของโปรแกรมประยุกต์ (Currency)	4.6	0.27
8) ความเข้าใจได้ของระบบ (Understandability)	4.0	0.30
9) ความสมัครใจในการใช้ระบบ (Voluntariness)	4.5	0.02
ค่าเฉลี่ย	4.59	0.26

ผู้วิจัยได้ทำการวัดประสิทธิภาพของผู้คนี่ลดสัมผัสด้วยอินเทอร์เน็ตประสานสรรพสิ่งจากความเร็วตอบสนองหลังกดซื้อจากโปรแกรมประยุกต์เคลื่อนที่ โดยเมื่อทำการกดจะบันทึกค่าลงบนภูิกิลไฟเซอร์เบส หลังจากนั้นผู้คนี่ก็จะทำการอ่านค่าในฐานข้อมูลบนภูิกิลไฟเซอร์เบสเสร็จแล้วนำมาประมวลผลเพื่อแสดงกิวรายการสำหรับทำการจ่ายน้ำ จากระยะเวลาดังกล่าว นำมาสู่การวัดประสิทธิภาพในตารางที่ 2 ข้อที่ 1 และสำหรับการวัดประสิทธิภาพจากระยะเวลาเดิมน้ำเต็ม 1 แก้วนั้น จะทำการวัดค่าเมื่อผู้คนี่ได้รับคำสั่ง จนได้สั่งการให้ทำการปั้มน้ำจากถัง ผ่านตัวควบคุมเปิดปิดน้ำไหล

ลงสู่แก้ว โดยการคำนวณประมาณเทียบกับระยะเวลาที่  
จ่ายน้ำเต็ม 1 แก้วเหมือนในตารางที่ 2 ข้อที่ 2

**ตารางที่ 2:** ผลการวัดประสิทธิภาพของตู้กดน้ำลดสัมผัสด้วย  
อินเทอร์เน็ตประสานสรรพสิ่ง

เกณฑ์การประเมิน	ครั้ง	ครั้ง	ครั้ง	เฉลี่ย
	ที่ 1	ที่ 2	ที่ 3	
1) ความเร็วการตอบสนอง หลังกดซื้อ	3s	5s	4s	4s
2) ระยะเวลาเติมน้ำเต็ม 1 แก้ว	6s	7s	8s	7s

### 6. สรุปผลการศึกษา

ระบบของตู้กดน้ำลดสัมผัสด้วยอินเทอร์เน็ตประสาน  
สรรพสิ่งช่วยให้ผู้ใช้สั่งซื้อน้ำผ่านโปรแกรมประยุกต์  
เคลื่อนที่ได้ตามปริมาณที่ต้องการได้โดยปราศจากการ  
สัมผัสวัตถุเดียวกันร่วมกับผู้อื่น ผลประเมินระบบโดย  
ผู้ใช้พบว่าตู้กดน้ำสามารถจำหน่ายน้ำได้อย่างมี  
ประสิทธิภาพ เป็นการใช้อินเทอร์เน็ตประสานสรรพสิ่ง  
มาช่วยลดความเสี่ยงจากการสัมผัสเชื้อโรคโควิด 19

### เอกสารอ้างอิง

[1] Thailand World Economic Outlook Database, Apr 2012, International Monetary Fund. “เทคโนโลยี Internet of Things และนโยบาย Thailand 4.0,” <http://www.nbt.go.th/getattachment/Services/quarter2560/ปี-2560/ไตรมาส-3-ปี-2560/เอกสารแนบ.pdf.aspx>, สำนักงานคณะกรรมการกฤษฎีกากระจายเสียงกิจการโทรทัศน์ และกิจการโทรคมนาคมแห่งชาติ, 2560.

[2] เทพร้านค้า. *Arduino คืออะไร*. จาก <https://www.ec-bot.com/article/1/arduino-คืออะไร-ร้านค้าออนไลน์>.

[3] Shopee. *Raspberry Pi 4 Model B Ram4 - ราสเบอร์รี่พาย 4 โมเดล B แรม 4*. จาก <https://shopee.co.th/Raspberry-Pi-4-Model-B-Ram4-ราสเบอร์รี่พาย-4-โมเดล-B-แรม-4-i.270502312.7042582666>.

[4] อินโนเวตีฟ เอ็กเพอริเมนต์. *ZX-RTC มินิบอร์ดฐานเวลานาฬิกาจริง*. จาก <https://inex.co.th/home/product/zx-rtc/> บริษัท อินโนเวตีฟ เอ็กเพอริเมนต์ จำกัด.

[5] Shopee. *ของพร้อมส่งโครตฮิตปั๊มไดอะแฟรม 12V ปั๊มไดอะแฟรม ปั๊มน้ำ ปั๊มลม ขนาด 12VDC คุณน้ำ-ปั๊มน้ำ คุณลม-ปั๊มลม Diaphragm Pum*. จาก <https://shopee.co.th/> ของพร้อมส่ง โครตฮิตปั๊มไดอะแฟรม - 12V-ปั๊มไดอะแฟรม-ปั๊มน้ำ-ปั๊มลม-ขนาด-12VDC-คุณน้ำ-ปั๊มน้ำ-คุณลม - ปั๊มลม - Diaphragm-Pum-i.516323909.12918459578 Shopee. 2022.

[6] Shopee. *DHT22/AM2302แทนที่ SHT11 SHT15ความชื้น อุณหภูมิและความชื้นเซนเซอร์*. จาก <https://thai.alibaba.com/product-detail/dht22-am2302-replace-sht11-sht15-humidity-temperature-and-humidity-sensor-60568635366.html>.

[7] Beank. *รีวิวเครื่องกดกาแฟ Starbucks self serve เครื่องแรกในไทยคอกาแฟต้องลอง*. <https://food.trueid.net/detail/2G71MvpaQL2e>, True Digital Group Company Limited. 2020.

[8] สมาชิกหมายเลข 3258261. *[Review] ง่ายๆ ลองใช้ตู้สะดวกซื้อ bluemarkt สะดวกไปหมด ถูกใจคนรุ่นใหม่พกเงินน้อย*. <https://pantip.com/topic/37818036>, สมาชิกหมายเลข 3258261, 2017.

[9] เทพร้านค้า. *ตู้น้ำหยอดเหรียญ ระบบ RO ขนาด 200 ลิตร ระบบกรอง 7 ชั้น ตอน*. <https://www.deetisud.com/product/130/ตู้น้ำหยอดเหรียญ-ระบบ-ro-ขนาด-200-ลิตร-ระบบกรอง-7-ชั้นตอน>, 2564.

[10] Investory. *ธุรกิจเครื่องกดกาแฟใน 7-11 รายได้ดีแค่ไหน*. <https://www.thebangkokinsight.com/news/business/branding/184474/>, The Bangkok Insight, 2018.

[11] N. Nilabh, D. Dhruvajyoti, D. Dipak, and B. Subhabrata. “Design of Food, Medicine & Water Dispensing Automation Device.” *Proceedings of the Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, IEEE Xplore Part Number:CFP20OSV-ART, 2020.

[12] C. Wen-Zhi, C. Ray-Guang, and C. Shuo-Yan. “Power-saving for IoT-enabled Water Dispenser System.” *National Taiwan University of Science and Technology (Taiwan Tech)*, Taipei, Taiwan, Republic of China, IEEE, 2019.

[13] M. Albin Libi, and S. Lipsa. “IoT Applications in Automated Water Level Detections.” *International Conference on Intelligent Engineering and Management (ICIEM)*, IEEE, 2020.

# การเพิ่มประสิทธิภาพการใช้สัญญาณมือถือภายในอาคาร โดยใช้การรวมช่องสัญญาณต่างความถี่ Indoor Mobile Performance Improvement by Inter-Band Carrier Aggregation

ปกรณ์ แก้วคำแพง (Pakorn Kaewkhampaeng)<sup>1</sup> และสุวัฒน์ ภัทรมาลัย (Suwat Pattaramalai)<sup>2</sup>

ภาควิชาวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี  
<sup>1</sup>pakorn.ene@mail.kmutt.ac.th, <sup>2</sup>suwat.pat@kmutt.ac.th

## บทคัดย่อ

ในงานวิจัยนี้ได้ทดลองวัดผลการเพิ่มประสิทธิภาพการใช้สัญญาณมือถือภายในอาคาร โดยใช้การรวมความถี่ Carrier aggregation (CA) ต่างความถี่กัน (Inter-Band) และต่างสถานีกัน (Inter-site) ซึ่งจะวัดผลเปรียบเทียบกับ 2 สถานการณ์คือไม่รวมความถี่ (Non-CA) 900MHz. และรวม 3 ความถี่ (1800 2100 และ 900MHz) โดยเก็บผลด้วยโปรแกรม GENEX PHU Smart และวิเคราะห์ผลด้วยโปรแกรม Genex Assistant มี 4 ตัวชี้วัดได้แก่ ความเร็วในการรับ-ส่งข้อมูล (Speed Test) ทราฟฟิค (Throughput) ความแรงสัญญาณ (RSRP) และคุณภาพสัญญาณ (SINR) ผลสรุปพบว่าการรวมความถี่นั้นความเร็วในการรับ-ส่งข้อมูล มีค่าความเร็วดาวน์โหลดเพิ่มขึ้น 409.90% ความเร็วอัพโหลดเพิ่มขึ้น 297.81% และปิงดีขึ้น 45.16% ทราฟฟิคเพิ่มขึ้น 249.69% แต่ค่า RSRP และค่า SINR ลดลง 7.68% 14.69% ตามลำดับ ซึ่งจากการรวม 3 ความถี่ช่วยทำให้ขนาดช่องสัญญาณเพิ่มขึ้นเป็น 3.5 เท่า ส่งผลให้ประสิทธิภาพสัญญาณในอาคารได้ดียิ่งขึ้น

**คำสำคัญ:** การรวมความถี่ ความเร็วเน็ต ทราฟฟิค ความแรงสัญญาณ คุณภาพสัญญาณ

## Abstract

In this study, an experiment was conducted to measure the effectiveness of mobile phone signal usage in the building by utilizing carrier aggregation (CA). Experiment settings include using different types of CA, such as different frequencies (Inter-Band) and different stations (Inter-site), as well as two

scenarios, one using a single frequency and another using CA. The GENEX PHU Smart program was used to collect the data, which was then analyzed by the Genex Assistant program. Data transmission speed (speed Test), throughput, signal strength (RSRP), and signal quality (SINR) were the four metrics used to evaluate experiment. The results show that the effective of using CA is higher than using single frequency as download speed increased by 409.90%, upload speed increased by 297.81%, Ping improved by 45.16%, throughput increased by 249.69%, RSRP decreased by 7.68% and SINR decreased by 14.69%, resulting in 3.5 times more bandwidth and better indoor signal performance.

**Keywords:** Carrier aggregation, Speed Test, Throughput, RSRP, SINR

## 1. บทนำ

ระบบเครือข่ายมือถือมีบทบาทสำคัญต่อผู้ใช้งานในปัจจุบันเป็นอย่างมากเช่น ความต้องการรับ-ส่งข้อมูลที่รวดเร็ว การรับส่งข้อมูลขนาดใหญ่ ซึ่งเทคโนโลยีรวมความถี่ (Carrier aggregation) [1] เพื่อเพิ่มขนาดช่องสัญญาณทำให้สามารถรองรับการใช้งานที่เพิ่มขึ้นและรับส่งข้อมูลได้เร็วขึ้น ซึ่งเทคโนโลยี 5G มีการทดลองรวมความถี่ได้และ LTE-A (Long Term Evolution Advance) ก็สามารถรวมความถี่ได้เช่นกัน และเมื่อปี 2563 ประเทศไทยได้มีการประมูลคลื่นความถี่ที่รองรับ 5G และในปัจจุบันการรวมความถี่ของ 5G ยังไม่เป็น

แพร่หลายเนื่องด้วยความปลอดภัยของสัญญาณ อุปกรณ์มือถือที่รองรับ รวมไปถึงผู้ใช้ส่วนใหญ่ยังใช้งานบน 4G ที่มีสัญญาณความปลอดภัยมากกว่า อย่างไรก็ตามผู้ใช้บริการเครือข่ายมือถือแต่ละเครือข่ายในประเทศไทยมีการถือครองความถี่แต่ละความถี่ที่มีขนาดช่องสัญญาณที่จำกัดไม่สามารถรองรับการใช้งานอย่างเพียงพอได้ โดยเฉพาะในอาคารที่มีความหนาแน่นของผู้ใช้งานทำให้เกิดการใช้งานรับ-ส่งข้อมูลที่ช้า การรวมความถี่เป็นหนึ่งในวิธีแก้ปัญหาแต่ยังไม่เคยมีการวัดผลในบริบทของประเทศไทย

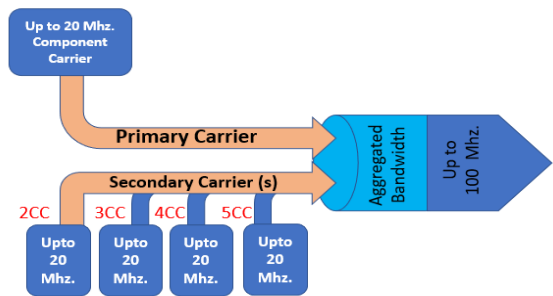
จากงานวิจัยที่เกี่ยวข้องพบว่า Sangmin Lee และคณะ [2] ได้นำเสนอประสิทธิภาพของการรวม 2 ความถี่ระหว่าง 2600 และ 850MHz เก็บผลด้วยวิธีการขับรถ (Drive Test) ซึ่งคล้ายกับ Aleksi Tanner [3] ได้นำเสนอประสิทธิภาพของการรวมความถี่ 3 คลื่นความถี่และ 2 ความถี่บน Network ของผู้ใช้บริการที่ฟินแลนด์โดยใช้โปรแกรม Anite's Nemo วัดผลทดสอบและ Lea Fadlan และคณะ [4] ได้นำเสนอประสิทธิภาพของการรวม 2 ความถี่ระหว่าง 1800 และ 2100MHz ได้จำลองการใช้งานโดยใช้โปรแกรม Atoll จำลองผลและวิเคราะห์ซึ่งคล้ายกับ Subuh Pramono และคณะ [5] จำลองผลของเมืองซูราการ์ตา ประเทศอินโดนีเซียโดยใช้โปรแกรมเดียวกันในการจำลองผลขึ้นมาและวิเคราะห์ผลพบว่าประสิทธิภาพโดยรวมเพิ่มขึ้นเล็กน้อยและ Pongsakorn Pantham และคณะ [6] ได้นำเสนอประสิทธิภาพของการรวม 2 ความถี่ระหว่าง 2300 และ 2100MHz แต่ใช้โปรแกรมจำลองผลชื่อ ns-3 และวัดประสิทธิภาพของทรูพุทเช่นเดียวกัน

ในงานวิจัยนี้วัดผลจากการเพิ่มประสิทธิภาพการใช้สัญญาณมือถือภายในอาคาร โดยใช้การรวมความถี่ต่างความถี่กันและต่างสถานีกัน ซึ่งจะทำการรวมช่องสัญญาณของเทคโนโลยี 4G ภายในอาคารมีสถานีฐานความถี่ 900MHz มีแบนด์วิธ 10MHz กับเสาสถานีนอกอาคารที่มีความถี่ 1800 และ 2100MHz มีแบนด์วิธ 15 และ 10MHz กลายเป็น 3 ความถี่ที่มีแบนด์วิธรวม 35MHz ซึ่งจะวัดผลในอาคาร โดยมี 4 ตัวชี้วัดได้แก่ ค่าความเร็วการรับ-ส่งข้อมูล ค่า Throughput ค่า RSRP และ ค่า SINR

## 2. ทฤษฎีที่เกี่ยวข้อง

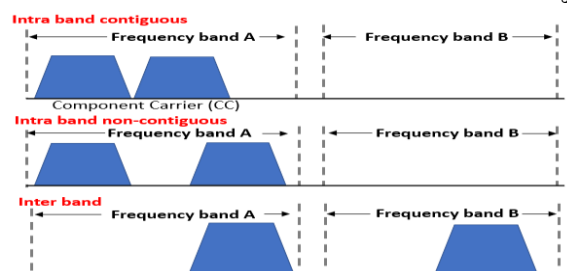
### 2.1 เทคโนโลยีการรวมความถี่ (Carrier aggregation)

เทคโนโลยีการรวมคลื่นความถี่เพื่อให้ช่องสัญญาณกว้างขึ้น [7] สามารถใช้ได้ทั้งการรับส่งสัญญาณประเภทการรับส่งข้อมูลด้วยการแยกความถี่หรือเรียกว่า FDD และการรับส่งข้อมูลด้วยการแบ่งช่วงเวลาเรียกว่า TDD โดยการรวมในระบบ FDD หรือระบบ TDD สามารถมีแบนด์วิธได้ตั้งแต่ 1.4 3 5 10 15 และ 20MHz ต่อหนึ่งความถี่ซึ่งได้สูงสุดถึง 5 ความถี่ แสดงดังรูป 1



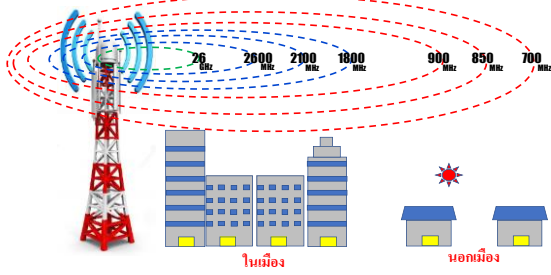
รูปที่ 1: การรวมความถี่ (carrier aggregation)

มีรูปแบบของการรวมความถี่มี 3 ประเภทได้แก่ 1. Intra-band contiguous เป็นการรวมจำนวนช่องสัญญาณของคลื่นความถี่ในย่านเดียวกันจากการขยายความถี่ที่ต่อเนื่องกัน 2. Intra-band non-contiguous เป็นการรวมช่องสัญญาณของความถี่ในย่านเดียวกันแต่เป็นช่วงความถี่ที่ไม่ติดต่อกันต้องใช้อุปกรณ์รับส่งสัญญาณสองตัวในการรับ-ส่ง ทำให้สูญเสียพลังงานของอุปกรณ์รับสัญญาณมากขึ้น 3. Inter-band non-contiguous เป็นการรวมจำนวนช่องสัญญาณของคลื่นความถี่ระหว่างย่านความถี่โดยมีตัวรับส่งหลายตัวทำให้มีผลกระทบต่อการใช้พลังงานความถี่เหมาะสำหรับผู้ใช้บริการที่มีความถี่มากกว่า 1 ย่านความถี่ เช่นระหว่างคลื่น 1800 และ 2100MHz แสดงดังรูป 2 และผู้ใช้บริการถือครองความถี่ที่จำกัดการจัดสรรและนำไปใช้ให้เหมาะสมจึงสำคัญ



รูปที่ 2: รูปแบบของการรวมความถี่ 3 ประเภท

## 2.2 การนำความถี่ไปใช้งาน (Frequency Deployment)



รูปที่ 3: การนำความถี่ไปใช้งาน

เนื่องจากผู้ให้บริการมีการถือครองความถี่แต่ละความถี่ที่หลากหลาย การจัดสรรและการนำไปใช้ให้เหมาะสมในแต่ละพื้นที่จึงสำคัญ รูปที่ 3 เป็นการแบ่งช่วงย่านความถี่เพื่อให้เหมาะสมกับการนำไปใช้งานได้ 3 ย่านความถี่ ได้แก่ 1. Low Frequency เป็นย่านความถี่ต่ำช่วง 700 ถึง 850MHz. ส่งข้อมูลได้ไกลและมีความทะลุทะลวงที่สูง ซึ่งเหมาะกับการใช้งานในชุมชนที่ห่างไกล หรือภายในอาคาร แต่ช่องสัญญาณจะไม่กว้าง จึงไม่เหมาะกับพื้นที่หนาแน่น 2. High Frequency เป็นย่านความถี่สูงช่วง 1800 ถึง 2600Mhz เป็นคลื่นความถี่ที่เหมาะสมกับใช้ในสถานที่ชุมชนหนาแน่น รองรับการส่งข้อมูลความเร็วสูง แต่ส่งข้อมูลได้ไม่ไกลเท่าความถี่ต่ำ 3. Millimeter wave ย่านความถี่สูงระหว่าง 24 ถึง 40GHz ใช้ในโรงงานอุตสาหกรรม และรับส่งข้อมูลด้วยความเร็วสูง เดินทางได้ระยะทางสั้นมาก แต่มีปัญหาในการทะลุทะลวงสิ่งกีดขวาง

ในงานวิจัยนี้ภายในอาคารที่ทดสอบนั้นมีสถานี่ฐานความถี่ 900MHz สัญญาณนั้นครอบคลุมภายในอาคาร แต่เนื่องจากมีความกว้างสัญญาณไม่กว้างทำให้ไม่สามารถรองรับการใช้งานที่เพิ่มขึ้นได้เพียงพอ จึงได้รวมความถี่เพื่อเพิ่มขนาดช่องสัญญาณรองรับการใช้งานที่เพิ่มมากขึ้น

### 2.3 พารามิเตอร์ตัวชี้วัดการวัดผลในงานวิจัย [8,9]

2.4.1 Speed Test คือการวัดค่าความเร็วในการรับ-ส่งข้อมูล โดยจะวัดออกมาเป็นค่าความเร็วดาวน์โหลดความเร็วอัปโหลดและค่า ping ที่จะใช้หลักการของความเร็วในการรับ - ส่ง ไฟล์เป็นหลัก โดยที่ดาวน์โหลดคือ ค่าความเร็วดาวน์โหลดข้อมูลจากเซิร์ฟเวอร์มาสู่ตัวเครื่อง และ อัปโหลด คือ ความเร็วการส่งข้อมูลจากเครื่องเข้าสู่เซิร์ฟ

เวอร์ และ ping คือความเร็วในการตอบสนองระหว่างเครื่องเรากับเซิร์ฟเวอร์

2.4.2 ทฤษฎี (Throughput) คือความสำเร็จในการนำส่งข้อมูลจากจุดหนึ่งไปยังอีกจุดหนึ่งในช่วงเวลาที่กำหนด มีหน่วยเป็นเมกะบิตต่อวินาที (Mbps) ซึ่งในงานวิจัยนี้จะวัดผลส่งค่านโวลด์สามารถคำนวณได้ตามสมการ (1)

$$Throughput_{DL} = \frac{L.Thrp.bits.DL}{L.Thrp.Time.Cell} \times 1000 \quad (1)$$

โดยที่

$L.Thrp.bits.DL$  คือ ปริมาณข้อมูลของดาวน์โหลด

$L.Thrp.Time.Cell$  คือ เวลาในการส่งข้อมูลของเซลล์ในความละเอียด 1 มิลลิวินาที

2.4.3 RSRP (Reference Signal Received Power) ค่าเฉลี่ยเชิงเส้นของสัญญาณอ้างอิงความถี่ผ่านช่องแบนด์วิธสามารถบอกความแรงของสัญญาณ มีหน่วยเป็น dBm เป็นการลดทอนของสัญญาณยิ่งลดทอนน้อยยิ่งดี สามารถคำนวณได้ตามสมการ (2)

$$RSRP = \frac{P_0}{N} \quad (2)$$

โดยที่

$P_0$  คือ กำลังสัญญาณที่ได้รับ

$N$  คือ จำนวน Resource Block ตามช่องสัญญาณ

2.4.4 SINR (Signal to interference plus noise ratio) อัตราส่วนของกำลังของสัญญาณที่เราต้องการสื่อสารเทียบกับกำลังของสัญญาณแทรกสอดและสัญญาณรบกวนหรือก็คือสามารถบอกคุณภาพของสัญญาณได้ สามารถคำนวณได้ตาม (3)

$$SINR = \frac{P_0}{\sum_1^n P_I + N_b} \quad (3)$$

โดยที่

$P_0$  คือ กำลังสัญญาณที่ได้รับ

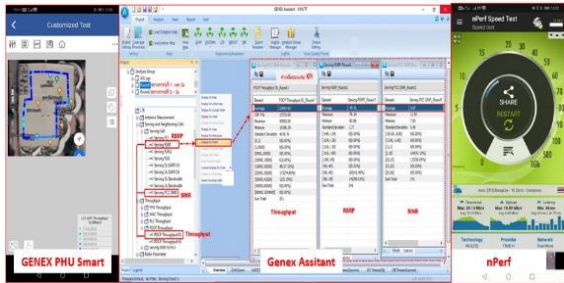
$P_I$  คือ กำลังรบกวนเฉลี่ยของเซลล์ข้างเคียง

$N_b$  คือ สัญญาณรบกวนพื้นหลัง

### 2.4 โปรแกรมที่เกี่ยวข้องในการทดลอง

GENEX PHU Smart ใช้ในการเก็บผลพารามิเตอร์ ค่าทฤษฎี ค่า RSRP และค่า SINR เป็นโปรแกรมใช้ร่วมกับมือถือสมาร์ทโฟนสามารถแสดงผลแบบเรียลไทม์ และ Genex Assistant เป็นโปรแกรมวิเคราะห์ผลหลังการเก็บ

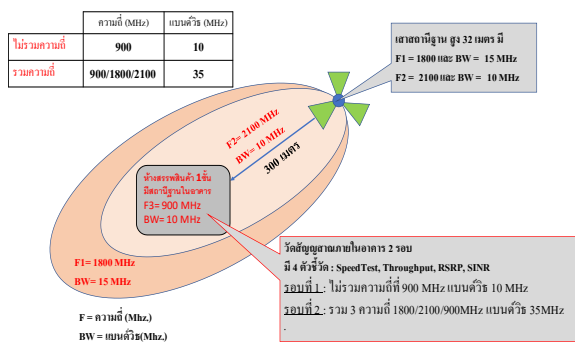
ผลโดยสามารถแสดงผลเป็นค่าเฉลี่ยของแต่ละพารามิเตอร์ได้ ซึ่งทั้งสองโปรแกรมนี้ถูกคิดค้นพัฒนาโดย หัวเว่ยสุดท้ายแอปพลิเคชัน nPerf ใช้ทดสอบความเร็วในการรับ-ส่งข้อมูลมีค่า ความเร็วดาวน์โหลด ความเร็วอัปโหลด และ ping แสดงดังรูปที่ 4



รูปที่ 4: โปรแกรมในการทดลอง

### 3. การดำเนินงานวิจัย

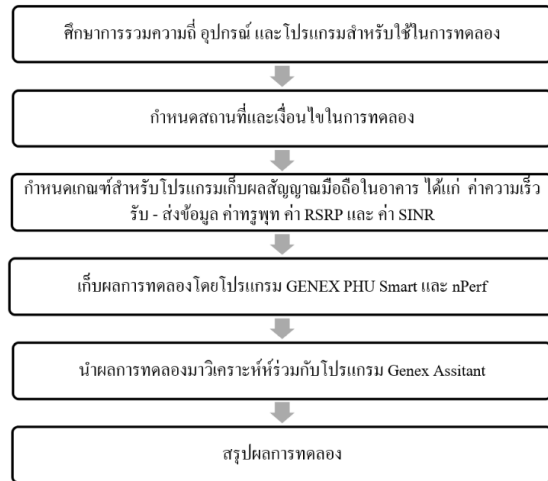
#### 3.1 แผนการดำเนินงาน



รูปที่ 5: รายละเอียดสถานที่ในการวิจัย

จากรูปที่ 5 แสดงถึงรายละเอียดสถานที่ที่วิจัยภายในอาคารห้างสรรพสินค้าแห่งหนึ่งมี 1 ชั้น มีสถานีฐานภายในอาคารกระจายความถี่ 900MHz แบนด์วิธ 10MHz และด้านนอกอาคารห่างออกไปประมาณ 300 เมตรมีเสาสถานีฐานสูง 32 เมตร มีความถี่ 1800 และ 2100MHz มีแบนด์วิธ 15 และ 10MHz ตามลำดับ โดยรวม 3 ความถี่ทำให้ขนาดแบนด์วิธเพิ่มขึ้นเป็น 35MHz

การวิจัยนี้ได้มีวัตถุประสงค์เพื่อวัดผลประสิทธิภาพสัญญาณมือถือในอาคารที่เพิ่มขึ้นจากการทำรวมความถี่ภายใต้เงื่อนไขต่างความถี่ (Inter-Band 3 ความถี่) และต่างสถานี โดยศึกษาการวัดผลสัญญาณ 4 ตัวชี้วัดได้แก่ ค่าความเร็วในการรับ - ส่งข้อมูล ค่าทราฟฟิค ค่า RSRP และค่า SINR ใช้อุปกรณ์และโปรแกรมที่สถานที่จริงซึ่งได้มีแผนการทดลองดังรูปที่ 6 ดังนี้



รูปที่ 6: แผนดำเนินงานวิจัย

#### 3.2 วิธีการเก็บผล

ในงานวิจัยนี้ทำการเดินเก็บผลในอาคาร 2 รอบ ได้แก่ รอบที่ 1 ไม่รวมความถี่ที่ความถี่ 900MHz แบนด์วิธ 10 MHz และรอบที่สองรวมความถี่ 900 1800 และ 2100 MHz แบนด์วิธรวม 35MHz รูปที่ 7 แสดงวิธีการเดินวัดผลสัญญาณในอาคารด้วยโปรแกรม GENEX PHU Smart เริ่มต้นที่หน้าโปรแกรม (1) กดที่คำสั่ง FTPDL ให้โปรแกรมทำงานจากนั้น (2),(3) กดแผนที่ในโปรแกรมตามตำแหน่งที่ยืนอยู่คือจุดที่ 1 และเดิน ไปจุดที่ 2 เพื่อบันทึกค่าสัญญาณ เมื่อเดินแล้วหยุดที่จุดไหนแล้วต้องกดในแผนที่เพื่อบันทึกค่าสัญญาณของเส้นทางที่ผ่านมา (4) จะเดินเก็บผลภายในอาคาร 1 รอบ (5) กดหยุดการทำงานของโปรแกรมเพื่อบันทึกผลโดยการเดิน 1 รอบโปรแกรมจะบันทึกผลค่า ทราฟฟิค ค่า RSRP และค่า SINR โปรแกรม nPerf เก็บค่าความเร็วในการรับ - ส่งข้อมูล



รูปที่ 7: ภาพประกอบการเดินเก็บผลทดลอง



### 3.2 กำหนดเกณฑ์ที่ใช้สำหรับวัดผลประสิทธิภาพสัญญาณมือถือในอาคาร

การวัดค่าสัญญาณมือถือในอาคารนั้นมีการกำหนดเกณฑ์ที่ใช้ดูในอุปกรณ์ระหว่างการเดินทดสอบ ทำให้ผู้ทดสอบนั้นสามารถประเมินได้ว่าค่าสัญญาณจุด ๆ นั้นเป็นอย่างไร โดยมีค่า ทรุษุท ค่า RSRP และ SINR แสดงดังตารางที่ 1

ตารางที่ 1: เกณฑ์ค่า Throughput, RSRP และ SINR

กำหนดเกณฑ์พารามิเตอร์ในการทดสอบ				
Throughput (Mbps)	RSRP (dB)	SINR (dB)	ความหมาย	สี
$x >= 60$	N/A	N/A	N/A	
$42 <= x < 60$	$x >= -80$	$x >= 20$	ยอดเยี่ยม	
$30 <= x < 42$	$-90 <= x < -80$	N/A	ดีมาก	
$20 <= x < 30$	$-95 <= x < -90$	$15 <= x < 20$	ดี	
$15 <= x < 20$	$-100 <= x < -95$	$10 <= x < 15$	ปานกลาง	
$10 <= x < 15$	$-110 <= x < -100$	$1 <= x < 10$	อ่อน	
$50 <= x < 10$	$-116 <= x < -110$	$-1 <= x < 1$	ไม่ดี	
$0 < x < 5$	$-124 <= x < -116$	$-3 <= x < -1$	ไม่ดีมาก	
$x < 0$ kbps	$x < -124$	$x < -3$	ไม่มีสัญญาณ	

### 4. ผลการวิจัย

#### 4.1 ผล 2 สถานการณ์ ไม่รวมความถี่และรวมความถี่

4.1.1 สถานการณ์ ไม่รวมความถี่พบว่าค่าเฉลี่ยของความเร็วในการรับ – ส่งข้อมูลได้ ค่าความเร็วดาวน์โหลด 13.74Mbps ความเร็วอัปโหลด 6.40Mbps และ ping 62ms. ในส่วนของค่าทรุษุท RSRP และ SINR มีค่า 23.99Mbps -79.71dBm และ 9.87dB ตามลำดับ แสดงดังรูปที่ 8



รูปที่ 8: ผลพารามิเตอร์ ไม่รวมความถี่จากโปรแกรม nPerf GENEX PHU Smart และ Genex Assistant

4.1.2 สถานการณ์ รวมความถี่พบว่าค่าเฉลี่ยของความเร็วในการรับ – ส่งข้อมูลได้ ค่าความเร็วดาวน์โหลด 70.06Mbps ความเร็วอัปโหลด 25.46Mbps และ ping 34ms ในส่วนค่าทรุษุท RSRP และ SINR มีค่า 83.89Mbps -85.78dBm และ 8.42dB ตามลำดับ แสดงดังรูปที่ 9



รูปที่ 9: ผลพารามิเตอร์ รวมความถี่จากโปรแกรม nPerf GENEX PHU Smart และ Genex Assistant

จากการเปรียบเทียบจาก 2 สถานการณ์ระหว่าง ไม่รวมความถี่และรวมความถี่ เพื่อวิเคราะห์ประสิทธิภาพที่เพิ่มขึ้นของ 4 ตัวชี้วัดพบว่าความเร็วในการรับ – ส่งข้อมูลมีค่าความเร็วดาวน์โหลดเพิ่มขึ้น 409.90% ความเร็วอัปโหลดเพิ่มขึ้น 297.81% และค่า ping ดีขึ้น 45.16% ทรุษุทเพิ่มขึ้น 249.69% RSRP ลดลง 7.68% และ SINR ลดลง 14.69% แสดงดังตารางที่ 2

ตารางที่ 2: สรุปผลค่าสัญญาณเปรียบเทียบระหว่าง 2 สถานการณ์

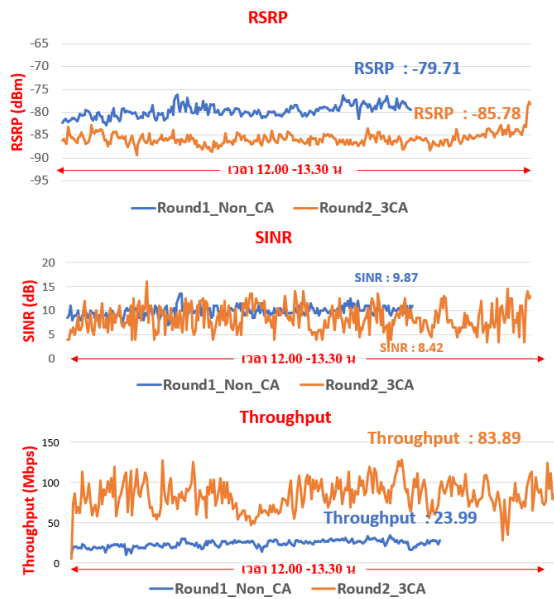
		รอบ 1	รอบ 2	ประสิทธิภาพที่เพิ่มขึ้น (%)		
รวมช่องสัญญาณ (CA)		ไม่มี	มี			
ความถี่ (Mhz)		900	900/1800 /2100			
Bandwidth (Mhz.)		10	35			
ตัวชี้วัด	ความเร็วในการรับ – ส่งข้อมูล	Download (Mbps)	13.74	70.06	409.90%	เพิ่มขึ้น
		Upload (Mbps)	6.4	25.46	297.81%	เพิ่มขึ้น
		Ping	62	34	45.16%	เพิ่มขึ้น
	ทรุษุท (Mbps)	23.99	83.89	249.69%	เพิ่มขึ้น	
	RSRP (dBm)	-79.71	-85.78	-7.68%	ลดลง	
SINR (dB)	9.87	8.42	-14.69%	ลดลง		

ผลเพิ่มเติมรอบนอกอาคารสามารถจับสัญญาณการรวม 2 ความถี่คือ 1800 และ 2100MHz ของเสาส่งด้านนอกเท่านั้นพบว่าค่าเฉลี่ยของความเร็วในการรับ – ส่งข้อมูลได้ ค่าความเร็วดาวน์โหลด 42.36Mbps ลดลง 39.54% ความเร็วอัปโหลด 20.81Mbps ลดลง 18.26% และ ping 46ms ลดลง 35.29% ในส่วนของค่าทรุษุทมีค่า 47.49Mbps ลดลง 43.39% ค่า RSRP -79.57dBm เพิ่มขึ้น 7.23% และค่า SINR 20.06dB เพิ่มขึ้น 138.24% เมื่อเทียบกับผลรวม 3 ความถี่ภายในอาคาร

#### 4.2 ผลกราฟเปรียบเทียบ 2 สถานการณ์

รูปที่ 10 แสดงความสัมพันธ์ระหว่างค่า ทรุษุท ค่า RSRP และค่า SINR มานำเสนอในรูปแบบกราฟเปรียบเทียบระหว่าง ไม่รวมความถี่ (เส้นสีฟ้า) และรวมความถี่ (เส้นสีส้ม) พบว่าค่าเฉลี่ย RSRP และ SINR จะลดลง 6.07dBm และ 1.45dB ตามลำดับ แต่ค่าทรุษุท

เพิ่มขึ้นเมื่อรวมความถี่ แม้ว่าในช่วงเวลาที่กราฟทรูพุตลดลง แต่ยังคงสูงกว่าการไม่รวมความถี่ตลอดผลการทดลอง



รูปที่ 10: กราฟเปรียบเทียบ RSRP ค่า SINR และทรูพุต

### 5. สรุปผลการทดลอง

ในงานวิจัยนี้ได้ใช้เทคโนโลยีการรวมความถี่ (Carrier aggregation) ที่ต่างความถี่กันระหว่างความถี่ 900 1800 และ 2100MHz และต่างสถานีกันระหว่างสถานีฐานภายในอาคารกับด้านนอกอาคารเพื่อเพิ่มประสิทธิภาพสัญญาณมือถือในอาคาร ซึ่งจากการรวม 3 ความถี่นั้นทำให้ขนาดช่องสัญญาณเพิ่มขึ้นเป็น 3.5 เท่า ส่งผลให้สามารถเพิ่มประสิทธิภาพสัญญาณมือถือภายในอาคารโดยค่าความเร็วในการรับ-ส่งข้อมูลมีค่าความเร็วดาวน์โหลดเพิ่มขึ้น 409.90% ความเร็วอัพโหลดเพิ่มขึ้น 297.81% ping ดีขึ้น 45.16% และในส่วนค่าทรูพุตเพิ่มขึ้น 249.69% มากกว่าการไม่ใช้เทคโนโลยีการรวมความถี่ จึงส่งผลให้ผู้ใช้งานสามารถรับส่งข้อมูลมือถือภายในอาคารประสิทธิภาพดีขึ้นด้วย แต่ที่ว่าค่า RSRP (ความแรงสัญญาณ) และ SINR (คุณภาพสัญญาณ) จะลดลง 7.68% และ 14.69% เนื่องจากจับได้สัญญาณความถี่หลักเป็น 1800MHz ของสถานีฐานนอกอาคาร ทั้งนี้หากมีการรวมจำนวนความถี่ที่เพิ่มขึ้น ทำให้ขนาดช่องสัญญาณเพิ่มขึ้นก็จะสามารถเพิ่มประสิทธิภาพสัญญาณในอาคารได้ดียิ่งขึ้น

### เอกสารอ้างอิง

- [1] A. Mihovska, and R. Prasad. "Overview of 5G New Radio and Carrier Aggregation: 5G and Beyond Networks." *2020 23rd International Symposium on Wireless Personal Multimedia Communications (WPMC)*. IEEE, 2020.
- [2] S. Lee, S. Hyeon, J. Kim, H. Roh, W. Lee. "The Useful Impact of Carrier Aggregation: A Measurement Study in South Korea for Commercial LTE-Advanced Networks. " *IEEE Vehicular Technology Magazine*, Vol.12, 2017.
- [3] A. Tanner. *LTE-A 3CC Carrier Aggregation*. Thesis Bachelor of Engineering Dept. Information Technology Helsinki Metropolia University of Applied Sciences, 2016.
- [4] L. Fadlan, T. Juhana. "Performance Analysis of Inter-band and Intra-band Carrier Aggregation on Planning and Dimensioning LTE-Advanced in Bandung City." *3rd International Conference on Wireless and Telematics (ICWT)*, 2017.
- [5] P. Subuh, D. Mustofa, A. Lia, E. Meiyanto. "Analysis and optimization of 4G long term evolution (LTE) network in urban area with carrier aggregation technique on 1800 MHz and 2100 MHz frequencies." *AIP Conference Proceedings*, Vol. 2217, 2020.
- [6] P. Pantham, S. Pattaramalai. "Simulation of LTE-Network for Throughput Improving Using FDD and TDD Carrier Aggregation." *In Proc. 8th International Electrical Engineering Congress (iEECON)*, 2020.
- [7] S. Ahmadi. *LTE-Advanced. A Practical Systems Approach to Understanding 3GPP LTE Releases 10 and 11 Radio Access Technologies*. 1, Academic Press, India, pp. 230-233, 2014.
- [8] A. Elnashar, and M. El-saidny. *Practical Guide to LTE-A, VoLTE and IoT*, John Wiley & Sons Ltd: UK, 2018.
- [9] A. L. Imoize, and O. D. Adegbite. "Measurements-based performance analysis of a 4G LTE network in and around shopping malls and campus environments in Lagos Nigeria." *Arid Zone Journal of Engineering, Technology and Environment*, Vol. 14, No. 2, pp. 208-225, June, 2018.

การประเมินการรับรู้ความปลอดภัยทางไซเบอร์ขององค์กร  
โดยการจำลองสถานการณ์ฟิชซิง  
The Assessment of an Organization's Cyber Security Awareness  
by Using a Phishing Simulation

ปรีชา หาญยืนยงสกุล (Preecha Harnyuenyongsakul)<sup>1</sup> และพงศ์ศรัณย์ บุญญูปกรณ์ (Pongsarun Boonyopakorn)<sup>2</sup>

ภาควิชาการบริหารเครือข่ายดิจิทัลและความมั่นคงปลอดภัยสารสนเทศ

คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

s6307031858051@kmutnb.ac.th, pongsarunb@kmutnb.ac.th

### บทคัดย่อ

บทความนี้มีวัตถุประสงค์เพื่อศึกษาการประเมินการรับรู้ความปลอดภัยทางไซเบอร์ขององค์กรโดยการจำลองสถานการณ์ฟิชซิง เพื่อทดสอบและศึกษาพฤติกรรมความตระหนักทางไซเบอร์ของบุคลากรในองค์กร ผู้วิจัยได้แบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่ม และมีการจำลองสถานการณ์ฟิชซิง 2 ครั้ง โดยมีเนื้อหาที่แตกต่างกัน และทั้ง 2 กลุ่มตัวอย่างมีรูปแบบการถ่ายทอดความรู้ที่แตกต่างกันอีกด้วย กลุ่มที่ 1 ฝายปฏิบัติการซึ่งเกี่ยวข้องกับทางด้านไอทีโดยตรง มีการถ่ายทอดความรู้ในรูปแบบ two-factor awareness และกลุ่มที่ 2 ฝายสนับสนุนซึ่งส่วนงานไม่ได้เกี่ยวข้องกับไอทีโดยตรง (Non-IT) มีการถ่ายทอดความรู้ในรูปแบบ single factor awareness พบว่าในกลุ่มที่ 1 มีจำนวนบุคลากรที่ไม่เปิดอีเมลมีจำนวนเพิ่มขึ้น 6.96% เปิดและอ่านอย่างเดียวก่อนเพิ่มขึ้น 9.57% เปิดอ่านและคลิกลิงก์ลดลง 16.52% คลิกและป้อนข้อมูลลดลง 20% นอกจากนี้ในกลุ่มตัวอย่างที่ 2 จาก ผลวิเคราะห์การทดสอบทั้ง 2 ครั้งพบว่าบุคลากรที่ไม่เปิดอีเมลมีจำนวนลดลง 22.22% เปิดและอ่านอย่างเดียวก่อนเพิ่มขึ้น 3.17% เปิดอ่านและคลิกลิงก์เพิ่มขึ้น 19.05% คลิกและป้อนข้อมูลเพิ่มขึ้น 19.05% นั้นแสดงว่าการถ่ายทอดความรู้ในรูปแบบ two-factor awareness สามารถให้ผลลัพธ์ที่ดีขึ้นจากเดิมในเรื่องของการคลิกและป้อนข้อมูล ซึ่งแตกต่างจากการถ่ายทอดความรู้ในรูปแบบ single factor awareness

**คำสำคัญ:** ฟิชซิง ความมั่นคงปลอดภัยทางไซเบอร์ ความตระหนัก

### Abstract

The objective of this paper was to evaluate an organization's cyber security awareness by using a phishing simulation. The participants were divided into two groups; operation, which deals with IT directly, and support, which deals with non-IT issues. The operating group was taught using a two-factor awareness technique. To educate the support group, single factor awareness was used. According to the findings, there was 6.96% increase in subjects who did not respond to email, 9.57% increase in subjects who opened email, 16.52% decrease in subjects who opened and clicked email, and 20% decrease in subjects who clicked and filled out the information on email in operation group. In support group, there were 22.22% decrease in subjects who did not respond to email, 3.17% increase in subjects who opened email, 19.05% increase in subjects who opened and clicked email, and 19.05% increase in subjects who clicked and filled out the information on email. From the result, we concluded

that two-factor awareness technique instruction was more effective than single-factor awareness instruction in raising cyber security awareness.

**Keywords:** phishing, cybersecurity, awareness.

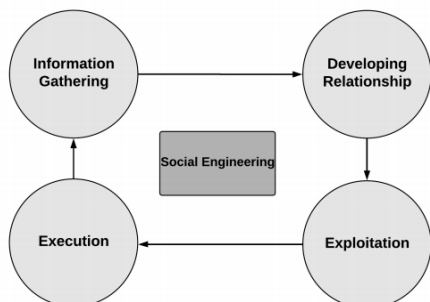
### 1. บทนำ

การโจมตีทางไซเบอร์ในปัจจุบัน ส่วนใหญ่มักเกิดจากตัวผู้ใช้งานในระบบสารสนเทศ การโจมตีแบบฟิชชิ่ง (Phishing) คือ การหลอกลวงทางอินเทอร์เน็ต เป็นการโจมตีด้วยวิธีทางวิศวกรรมสังคม (Social Engineering) นับเป็นการโจมตีที่ประสบความสำเร็จมาก และส่งผลกระทบต่อความมั่นคงปลอดภัยทางไซเบอร์มากเช่นกัน ซึ่งทางผู้วิจัยมีความสนใจที่จะนำการโจมตีแบบฟิชชิ่ง (Phishing) มาทดสอบในองค์กร เพื่อลดช่องโหว่ในการถูกโจมตี และสร้างความตระหนักให้แก่บุคลากรในองค์กรไม่ให้ตกเป็นเป้าหมายซ้ำได้อีก

### 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

#### 2.1 การโจมตีทางวิศวกรรมสังคม

การโจมตีทางวิศวกรรมสังคม คือ กลยุทธ์ทางจิตวิทยาที่ใช้ในกลอุบายหลอกล่อเป้าหมายให้หลงเชื่อและยอมเปิดเผยความลับ [1] มีการดำเนินการในรูปแบบต่าง ๆ จะเริ่มต้นจากการรวบรวมข้อมูลของเป้าหมายซึ่งเป็นหนึ่งในขั้นตอนที่สำคัญที่สุดในการโจมตี ในกรณีของสังคม การโจมตีทางวิศวกรรมสังคม ผู้โจมตีต้องมีความรู้เกี่ยวกับลักษณะพื้นฐานของเป้าหมาย หรือหลักจิตวิทยาที่สามารถใช้แสวงหาประโยชน์จากเป้าหมาย [2] ซึ่งวงจรชีวิตของการโจมตีทางวิศวกรรมสังคมแสดงในภาพที่ 1



ภาพที่ 1: องค์ประกอบวงจรชีวิตของการโจมตีทางวิศวกรรมสังคม

ขั้นตอนการรวบรวมข้อมูล (Information Gathering) [3,4]

ขั้นตอนแรกของการโจมตีทางวิศวกรรมสังคม คือ การรวบรวมข้อมูลของเป้าหมาย ในขั้นตอนนี้ผู้โจมตีจะรวบรวมข้อมูลเกี่ยวกับเป้าหมายเพื่อให้ได้มาซึ่งความน่าเชื่อถือ และเพื่อสร้างความสัมพันธ์ที่ไว้วางใจได้ ผู้โจมตีอาจได้รับข้อมูลเกี่ยวกับ ความเกี่ยวพันทางการเมือง ภูมิหลังทางการศึกษา ข้อมูลครอบครัว ข้อมูลทางการเงิน นอกจากนี้ยังรวมถึงข้อมูลเกี่ยวกับ โครงสร้างขององค์กร ลักษณะการทำงาน เพื่อนำมาวิเคราะห์กลยุทธ์ในการโจมตีเป้าหมายต่อไป

ขั้นตอนพัฒนาความสัมพันธ์ (Developing Relationships)

เมื่อมีการรวบรวมข้อมูลแล้ว ผู้โจมตีจะใช้ขั้นตอนพัฒนาความสัมพันธ์โดยใช้ข้อมูลที่เก็บรวบรวมมาดัดแปลงให้เสมือนจริง เพื่อเชื่อมโยงความสัมพันธ์ให้เป้าหมายหลงเชื่อ

ขั้นตอนแสวงหาประโยชน์ (Exploitation)

ผู้โจมตีจะใช้ทั้งข้อมูลและความสัมพันธ์เชื่อมโยงกัน เพื่อแทรกซึมเป้าหมาย หลอกล่อให้เป้าหมายเผยความลับด้วยวิธีต่าง ๆ

ขั้นตอนดำเนินการ (Execution)

ทำการโจมตีเป้าหมาย ให้ตรงกับวัตถุประสงค์ที่ผู้โจมตีต้องการ และอาจนำข้อมูลของเป้าหมายมาสร้างความเสียหายในด้านต่าง ๆ

#### 2.2 ฟิชชิ่ง (Phishing)

ฟิชชิ่ง (Phishing) หมายถึง การหลอกลวงทางอินเทอร์เน็ต โดยจะหลอกล่อเป้าหมายให้หลงเชื่อข้อความที่หลอกลวง เพื่อให้ได้มาซึ่งข้อมูลสำคัญของเป้าหมาย เช่น ข้อมูลส่วนบุคคล รหัสผ่าน บัตรเครดิต ฯลฯ โดยการปลอมแปลงให้เหมือนกับเว็บไซต์ผู้ให้บริการจริง [5] เพื่อเพิ่มโอกาสในการเข้าถึงข้อมูลส่วนตัว และเอกสารลับของเป้าหมาย หรือความลับทางการค้าของบริษัท จะมาในรูปแบบของอีเมลที่ไม่ได้ระบุเป้าหมายอย่างชัดเจน และจะถูกส่งอีเมลแบบกระจายโดยที่ไม่ได้หวังผลที่เป้าหมายเป็นพิเศษ [6]

#### 2.3 ฟิชชิ่งแบบเจาะจงเป้าหมาย (Spear-Phishing)

หมายถึง การโจมตีที่พุ่งเป้าไปยังเป้าหมายอย่างชัดเจน โดยผู้โจมตีจะค้นหาข้อมูลเบื้องต้นของพนักงาน

ในบริษัทที่เป็นเป้าหมายจากช่องทางต่าง ๆ เช่น จากนั้นผู้โจมตีจะสร้างอีเมล Phishing ที่ระบุเนื้อหาสอดคล้องกับเป้าหมายเพื่อให้เป้าหมายหลงเชื่อ กดลิงก์ที่แนบมากับอีเมล การโจมตีลักษณะนี้หากเป้าหมายเป็นบุคคลสำคัญในองค์กรจะเรียกว่า Whaling [7]

**2.4 ฟิชซิงด้วยเสียง (Vishing)**

หมายถึง เป็นการฟิชซิงด้วยเสียง ใช้การโน้มน้าวให้เหยื่อหลงเชื่อว่าเป็นสายที่โทรมาจากเจ้าหน้าที่หน่วยงานใดหน่วยงานหนึ่ง โดยเลือกที่จะขอข้อมูลบัญชี, รหัสผ่าน หรือข้อมูลธุรกรรมผ่านทางโทรศัพท์ เมื่อเป้าหมายหลงเชื่อให้ข้อมูลก็จะทำให้เกิดความสูญเสียเกิดขึ้น [8]

**2.5 Cyber Kill Chain**

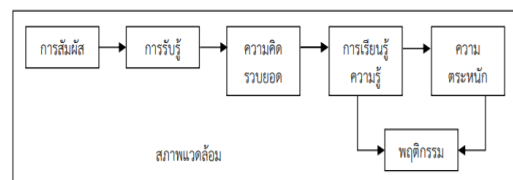
Cyber Kill Chain ถูกคิดค้นโดย Lockheed Martin ในปี 2011 [9] เพื่อใช้ในการอธิบายขั้นตอนในการเจาะระบบเพื่อโจมตีทางไซเบอร์ ซึ่งประกอบด้วย 7 ขั้นตอนต่อเนื่อง ได้แก่

1. Reconnaissance: อาชญากรทางไซเบอร์เริ่มเก็บข้อมูลของเป้าหมายก่อนเริ่มโจมตี ไม่ว่าจะเป็นทางอินเทอร์เน็ต หรือ ทางกายภาพ โดยหากสามารถเก็บข้อมูลของเป้าหมายได้มากก็จะส่งผลให้การโจมตีมีแนวโน้มที่จะประสบความสำเร็จมากขึ้น
2. Weaponization: เตรียมเครื่องมือที่เหมาะสมกับการโจมตีเป้าหมาย เช่น การเตรียมโปรแกรมสำหรับเจาะระบบที่เหมาะสมกับเป้าหมาย เป็นต้น
3. Delivery: ส่ง Payload ที่เตรียมไว้ไปยังเป้าหมายผ่านช่องทางต่างๆ เช่น ฟิชซิงอีเมล เว็บ หรือ USB เป็นต้น
4. Exploitation: อาชญากรทางไซเบอร์ทำการเจาะระบบของเป้าหมายผ่านช่องโหว่ต่างๆ ที่พบเจอจากการทำ Reconnaissance
5. Installation: ติดตั้ง Payload บนเครื่องของเป้าหมายเพื่อคอยรับคำสั่งจากอาชญากรทางไซเบอร์
6. Command & Control: หลังจากที่ได้ติดตั้ง Payload อาชญากรทางไซเบอร์สามารถทำการควบคุมหรือสั่งการเครื่องเป้าหมายได้
7. Action on Objectives: ดำเนินการตามเป้าหมายของตนเอง เช่น ขโมยข้อมูล หรือขัดขวางการทำงานทำให้ไม่

สามารถใช้งานได้ ส่วนใหญ่เมื่อโจมตีสำเร็จจะทำการลบร่องรอยเพื่อให้ไม่สามารถตรวจสอบย้อนกลับมาหาตนเองได้

**2.6 ทฤษฎีความตระหนัก**

ความตระหนัก (Awareness) เป็นแนวคิดเชิงจิตวิทยา (Psychological Approach) ผสมกับแนวคิดเชิงพฤติกรรมศาสตร์ (Behavior Science) เป็นผลมาจากกระบวนการทางปัญญา (Cognitive Process) กล่าวคือ เมื่อบุคคลได้รับการสัมผัสจากสิ่งเร้าในสภาพแวดล้อม จะทำให้เกิดกระบวนการการรับรู้ เมื่อบุคคลรับรู้จากสิ่งเร้าก็จะเกิดความเข้าใจในสิ่งเร้า นั้น ส่งผลให้เกิดความคิดรวบยอดนำไปสู่การเรียนรู้ และนำไปสู่การเกิดความตระหนักในที่สุด ซึ่งความรู้และความตระหนักต่างก็นำไปสู่การแสดงพฤติกรรมของบุคคลต่อสิ่งเร้า นั้นๆ [10] แสดงดังภาพที่ 2



ภาพที่ 2: ขั้นตอนและกระบวนการเกิดความตระหนัก

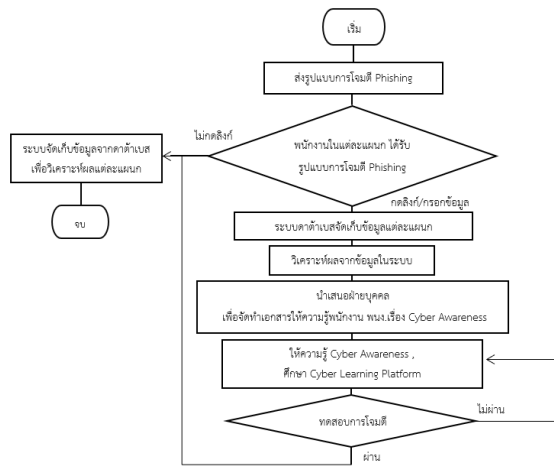
**3. วิธีการดำเนินงาน**

ในส่วนนี้จะกล่าวถึงขั้นตอนการดำเนินการวิจัยต่างๆ เพื่อให้การวิจัยดำเนินไปได้อย่างมีประสิทธิภาพและลุล่วง ผู้วิจัยจึงวางแผนขั้นตอนการวิจัยดังนี้

**3.1 ออกแบบขั้นตอนการวิจัย**

ผู้วิจัยเริ่มต้นจากการรวบรวมข้อมูลของเป้าหมายก่อนเริ่มทำการโจมตี มีการสร้างเนื้อหาอีเมล 2 ฉบับในการทดสอบแบ่งเป็น 2 รอบคือ ก่อนการโจมตีและหลังการโจมตี และได้มีการสร้างเว็บไซต์ปลอมเพื่อใช้สำหรับการจำลองสถานการณ์โจมตีในรูปแบบฟิชซิง เมื่อพนักงานในแต่ละแผนกได้รับอีเมล เปิดอ่าน กดลิงก์ กรอกข้อมูล ระบบจะจัดเก็บข้อมูลลงฐานข้อมูลเพื่อวิเคราะห์ผล และนำเสนอฝ่ายบุคคล เพื่อจัดทำโปสเตอร์ถ่ายทอดความรู้ทางด้าน Cyber Awareness หลังจากนั้นจะมีการส่งการจำลองสถานการณ์โจมตีในรูปแบบฟิชซิงทดสอบในรอบ

ที่ 2 หลังจากที่ได้ถ่ายทอดความรู้ในรูปแบบต่าง ๆ และนำข้อมูลจากฐานข้อมูลมาวิเคราะห์ผลเพื่อสรุปผล ดังภาพที่ 3



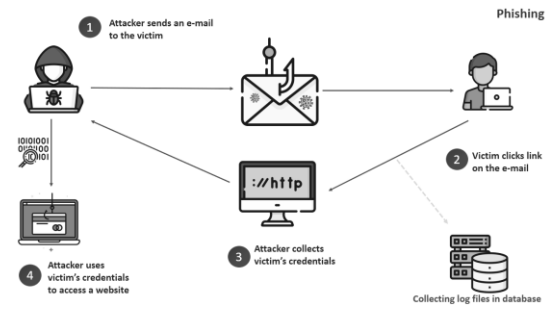
ภาพที่ 3: กระบวนการวิจัย

### 3.2 ดำเนินการวิจัย

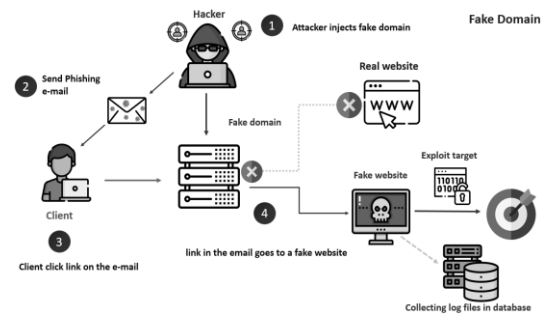
ผู้วิจัยได้มีการจำลองสถานการณ์การโจมตีในรูปแบบอีเมลฟิชชิ่งในองค์กรแห่งหนึ่ง เพื่อทดสอบและศึกษาพฤติกรรมความตระหนักรู้ทางไซเบอร์ของบุคลากรในองค์กร ในช่วงปลายไตรมาสที่ 4 ของปี พ.ศ. 2564 ถึงช่วง ต้นไตรมาสที่ 1 ของปี 2565 ซึ่งองค์กรแห่งนี้เกี่ยวข้องกับทางด้านเทคโนโลยี กลุ่มตัวอย่างที่ใช้ในการทดลองครั้งนี้มีจำนวนทั้งสิ้น 178 คน ผู้วิจัยได้แบ่งกลุ่มตัวอย่างเป็น 2 กลุ่ม และมีการจำลองสถานการณ์ฟิชชิ่ง 2 ครั้ง โดยมีเนื้อหาที่แตกต่างกัน และทั้ง 2 กลุ่มตัวอย่างมีรูปแบบการถ่ายทอดความรู้ที่แตกต่างกันอีกด้วย กลุ่มที่ 1 ฝ่ายปฏิบัติการซึ่งเกี่ยวข้องกับทางด้านไอทีโดยตรง มีการถ่ายทอดความรู้ในรูปแบบ two-factor awareness และกลุ่มที่ 2 ฝ่ายสนับสนุนซึ่งส่วนงานไม่ได้เกี่ยวข้องกับไอทีโดยตรง (Non-IT) มีการถ่ายทอดความรู้ในรูปแบบ single factor awareness โดยมีกระบวนการในการทดลองการโจมตี ดังนี้

ผู้วิจัยได้ศึกษาขั้นตอนการโจมตีแบบฟิชชิ่ง [11] เพื่อวางแผนการจำลองสถานการณ์ในรูปแบบอีเมลฟิชชิ่งแก่บุคลากรภายในองค์กร เพื่อต้องการที่จะการประเมินการรับรู้ความปลอดภัยทางไซเบอร์ขององค์กรได้ โดยมีขั้นตอน ดังภาพที่ 4 และ 5

1. ผู้โจมตีส่งอีเมลถึงเหยื่อ
2. ผู้ที่ตกเป็นเหยื่อคลิกลิงก์ฟิชชิ่งในอีเมล โดยอาจเป็นลิงก์ปลอมที่นำทางไปสู่เว็บไซต์ฟิชชิ่ง
3. ผู้โจมตีรวบรวมข้อมูลข้อมูลประจำตัวของเหยื่อ
4. ผู้โจมตีใช้ข้อมูลประจำตัวของเหยื่อเพื่อเข้าถึงเว็บไซต์



ภาพที่ 4: ขั้นตอนโจมตีแบบฟิชชิ่ง



ภาพที่ 5: ขั้นตอนโจมตีแบบฟิชชิ่ง (Fake Domain)

ผู้วิจัยได้ทำการออกแบบหน้าเว็บไซต์ โดยมีเนื้อหาเกี่ยวข้องกับการชี้ชวนให้บุคลากรในองค์กรลงทะเบียนเพื่อยืนยัน การเพิ่มพื้นที่อีเมลด้วยระบบอัตโนมัติ และได้ทำการออกแบบโปสเตอร์ infographic สำหรับเป็นสื่อในการถ่ายทอดความรู้ให้แก่บุคลากรแต่ละฝ่ายขององค์กร โดยเนื้อหาจะกล่าวถึงความหมายของการโจมตีในรูปแบบต่าง ๆ ของ ฟิชชิ่ง และ แรนซัมแวร์ รวมถึงวิธีการสังเกตอีเมลต้องสงสัย รวมถึงการป้องกันไม่ให้ตกเป็นเหยื่อ หลังจากที่ได้ผู้วิจัยได้ทดสอบสถานการณ์การโจมตีในรูปแบบฟิชชิ่งในครั้งที่ 1 และได้ถ่ายทอดความรู้ในรูปแบบ single factor awareness และ two-factor awareness โดยขั้นตอนต่อมาได้มีการออกแบบพัฒนา

เว็บไซต์สำหรับการจำลองสถานการณ์ในรูปแบบอีเมล ฟิชชิ่งในครั้งที่ 2 โดยอีเมลมีเนื้อหาแตกต่างจากครั้งแรก ซึ่งในครั้งนี้จะเกี่ยวข้องกับกิจกรรมการร่วมลงทะเบียน กลุ่มของขวัญ เพื่อลุ้นรับเงินรางวัลและรางวัลพิเศษ

ในระหว่างการทดสอบสถานการณ์การโจมตีในรูปแบบฟิชชิ่งในครั้งที่ 1 และ 2 ผู้วิจัยได้ใช้โปรแกรมที่ใช้ในการตรวจจับสถานะของอีเมลปลายทางไม่ว่าจะเป็นสถานะส่งสำเร็จ เปิดหรือไม่เปิดอีเมล หรือแม้แต่คลิกลิงก์ ตัวโปรแกรมจะระบบแจ้งเตือนผ่านอีเมลของผู้ดูแลระบบ หากบุคลากรในองค์กรคลิกลิงก์และทำการป้อนข้อมูลเข้ามาในระบบ ข้อมูลที่ป้อนเข้ามาในระบบนั้นจะถูกจัดเก็บไว้ในฐานข้อมูลของระบบ SQL Server เพื่อนำข้อมูลมาวิเคราะห์

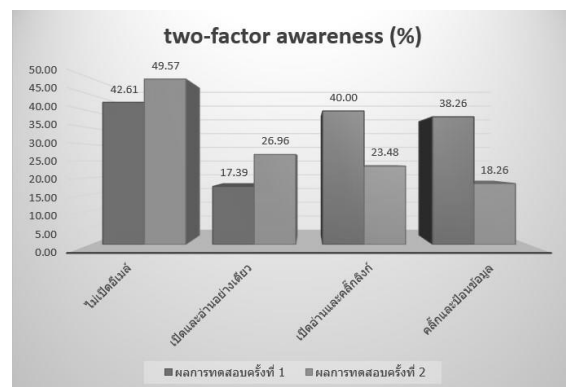
### 3.3 สรุปขั้นตอนการดำเนินการวิจัย

ผู้วิจัยออกแบบขั้นตอนการวิจัยได้ศึกษาขั้นตอนการโจมตีแบบฟิชชิ่ง เพื่อวางแผนในการจำลองสถานการณ์ในรูปแบบอีเมลฟิชชิ่งแก่บุคลากรภายในองค์กร เพื่อต้องการที่จะการประเมินการรับรู้ความปลอดภัยทางไซเบอร์ขององค์กร และกำหนดกระบวนการวิจัยให้เป็นไปอย่างมีประสิทธิภาพ โดยเริ่มรวบรวมข้อมูลของบุคลากรในองค์กร ออกแบบคอนเทนต์ตามสภาพแวดล้อมขององค์กร มีการมีการพัฒนาเว็บแอปพลิเคชันเพื่อใช้ในการจำลองสถานการณ์ในรูปแบบอีเมลฟิชชิ่ง รวมถึงการจัดทำสื่อ รวมถึงการประชาสัมพันธ์ถ่ายทอดความรู้ให้แก่บุคลากรแต่ละฝ่ายขององค์กร และมีการเก็บในส่วนของฐานข้อมูลในการจำลองสถานการณ์เพื่อนำไปวิเคราะห์ผลต่อไป

### 4. ผลการดำเนินงาน

จากการดำเนินงานการประเมินการรับรู้ความปลอดภัยทางไซเบอร์ขององค์กร โดยการจำลองสถานการณ์ในรูปแบบอีเมลฟิชชิ่ง โดยสามารถแบ่งผลการวิจัยแบ่งเป็น 2 กลุ่ม ได้แก่ ฝ่ายปฏิบัติการ มีการถ่ายทอดความรู้ในรูปแบบ two-factor awareness และฝ่ายสนับสนุน มีการถ่ายทอดความรู้ในรูปแบบ single factor awareness ในกลุ่มที่ 1 ฝ่ายปฏิบัติการ ผู้วิจัยมีการ

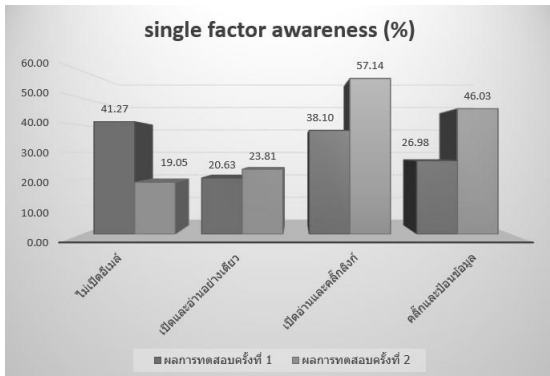
ถ่ายทอดความรู้ให้กับบุคลากร 2 ปัจจัย คือ ผู้วิจัยได้ทำบทความจำนวน 2 บทความ และเผยแพร่ให้แก่กลุ่มตัวอย่างผ่านอีเมล อีกทั้งมีการแลกเปลี่ยนความคิดเห็นเกี่ยวกับด้านความมั่นคงปลอดภัยทางไซเบอร์ในเรื่องต่างๆ ซึ่งจากกลุ่มตัวอย่าง 115 คนในกลุ่มที่ 1 มีสถานะไม่เปิดอีเมลในครั้งที่ 1 จำนวน 49 คน คิดเป็น 42.61% และในครั้งที่ 2 จำนวน 57 คน คิดเป็น 49.57% สถานะเปิดและอ่านอย่างเดียวในครั้งที่ 1 จำนวน 20 คน คิดเป็น 17.39% และในครั้งที่ 2 จำนวน 31 คน คิดเป็น 26.96% สถานะเปิดอ่านและคลิกลิงก์ในครั้งที่ 1 จำนวน 46 คน คิดเป็น 40% และในครั้งที่ 2 จำนวน 27 คน คิดเป็น 23.48% และสถานะคลิกและป้อนข้อมูลในครั้งที่ 1 จำนวน 44 คน คิดเป็น 38.26% และในครั้งที่ 2 จำนวน 21 คน คิดเป็น 18.26% แสดงดังภาพที่ 6



ภาพที่ 6: กราฟเปรียบเทียบ two-factor awareness

ในกลุ่มที่ 2 ฝ่ายสนับสนุน ผู้วิจัยมีการถ่ายทอดความรู้ให้กับบุคลากรเพียง 1 ปัจจัย คือ ผู้วิจัยได้ทำบทความจำนวน 2 บทความเผยแพร่ให้แก่กลุ่มตัวอย่างผ่านอีเมลเพียงอย่างเดียว ซึ่งจากกลุ่มตัวอย่าง 63 คนในกลุ่มที่ 2 มีสถานะไม่เปิดอีเมลในครั้งที่ 1 จำนวน 26 คน คิดเป็น 41.27% และในครั้งที่ 2 จำนวน 12 คน คิดเป็น 19.05% สถานะเปิดและอ่านอย่างเดียวในครั้งที่ 1 จำนวน 13 คน คิดเป็น 20.64% และในครั้งที่ 2 จำนวน 15 คน คิดเป็น 23.81% สถานะเปิดอ่านและคลิกลิงก์ในครั้งที่ 1 จำนวน 24 คน คิดเป็น 38.10% และในครั้งที่ 2 จำนวน 36 คน คิดเป็น 57.14% และสถานะคลิกและป้อนข้อมูลใน

ครั้งที่ 1 จำนวน 17 คน คิดเป็น 26.98% และในครั้งที่ 2 จำนวน 29 คน คิดเป็น 46.03% แสดงดังภาพที่ 7



ภาพที่ 7: กราฟเปรียบเทียบ single factor awareness

### 5. สรุป

ผลการศึกษการจำลองสถานการณ์การโจมตีในรูปแบบอีเมลฟิชชิ่งในองค์กรแห่งหนึ่งด้วยสถิติพื้นฐานจากผลวิเคราะห์การทดสอบทั้ง 2 ครั้งในกลุ่มฝ่ายปฏิบัติการพบว่าบุคลากรที่ไม่เปิดอีเมลมีจำนวนเพิ่มขึ้น 6.96% เปิดและอ่านอย่างเดียวเพิ่มขึ้น 9.57% เปิดอ่านและคลิกลิงก์ลดลง 16.52% คลิกและป้อนข้อมูลลดลง 20% ส่วนในกลุ่มฝ่ายสนับสนุนพบว่าบุคลากรที่ไม่เปิดอีเมลมีจำนวนลดลง 22.22% เปิดและอ่านอย่างเดียวเพิ่มขึ้น 3.17% เปิดอ่านและคลิกลิงก์เพิ่มขึ้น 19.05% คลิกและป้อนข้อมูลเพิ่มขึ้น 19.05% นั้นแสดงว่าการถ่ายทอดความรู้ในแบบ two-factor awareness สามารถให้ผลลัพธ์ที่ดีขึ้นจากเดิมในเรื่องของการคลิกและป้อนข้อมูล ซึ่งแตกต่างจากการถ่ายทอดความรู้ในแบบ single factor awareness

### เอกสารอ้างอิง

[1] Zwilling M, Klien G, Lesjak D, Wiechetek L, Cetin F, Basim HN. Cyber Security Awareness, Knowledge and Behavior: A Comparative Study. *Journal of Computer Information Systems*. 2020.

[2] Aloul FA. The Need for Effective Information Security Awareness. *JOURNAL OF ADVANCES IN INFORMATION TECHNOLOGY*, Vol. 3, No.3, 2012.

[3] Yeboah-Boateng EO, Amanor PM. Phishing, SMiShing & Vishing: an assessment of threats against mobile devices. *Journal of Emerging Trends in Computing and Information Sciences*, Vol. 5, No. 4, pp.297-307, 2014.

[4] Spinapolic M. *Mitigating the risk of social engineering attacks*. 2011.

[5] Jamil A, Asif K, Ghulam Z, Nazir MK, Alam SM, Ashraf R, editors. Mpmma: A mitigation and prevention model for social engineering based phishing attacks on facebook. *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, 2018.

[6] Mouton F, Malan MM, Leenen L, Venter HS, editors. *Social engineering attack framework*. 2014 *Information Security for South Africa*, IEEE, 2014.

[7] Zareapoor M, Seeja K. Feature extraction or feature selection for text classification: A case study on phishing email detection. *International Journal of Information Engineering and Electronic Business*, Vol. 7, No. 2, pp. 60, 2015.

[8] RSA, "Phishing, Vishing and Smishing: Old Threats Present New Risks," *RSA Monthly Online Fraud Report*, September, October & November 2009

[9] Applying Security Awareness to the Cyber Kill Chain, Available Online at <https://www.sans.org/blog/applying-security-awareness-to-the-cyber-kill-chain/>

[10] Good CV, Merkel WR. *Dictionary of education*, McGraw-Hill; 1973.

[11] Shabnam Sharma, Paper Study on Phishing Attacks, *Article in International Journal of Computer Applications*, December 2018.



# ผลกระทบของการรับรู้ความเสี่ยงด้านความมั่นคงปลอดภัยและความเป็นส่วนตัว การยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง

## The Impact of Perceived Security Risk and Privacy Risk on the Acceptance of IoT Technologies

มรุชัย วัฒนรังสรรค์ (Maruchai Wattanarungson)<sup>1</sup> และปราโมทย์ ลือนาม (Pramote Luenam)<sup>2</sup>

สาขาบริหารเทคโนโลยีสารสนเทศ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

<sup>1</sup>mcmaruchai@gmail.com, <sup>2</sup>pramote@nida.ac.th

### บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาอิทธิพลต่อความตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งของผู้บริโภคที่เคยใช้อุปกรณ์ IoT และมีถิ่นฐานในประเทศไทยจำนวน 396 คน โดยใช้แบบสอบถามออนไลน์เป็นเครื่องมือในการเก็บรวบรวมข้อมูล และทำการวิเคราะห์ข้อมูลที่ได้ในรูปแบบเชิงโครงสร้างด้วยโปรแกรม SPSS และ AMOS ในการทดสอบสมมติฐานเพื่อวิเคราะห์หาค่าอิทธิพลทั้งทางตรงและทางอ้อม โดยค่าความสอดคล้องของโมเดล (goodness-of-fit) ที่ได้เป็นดังนี้  $P\text{-value} = 0.406$ ,  $\chi^2/df = 1.035$ ,  $RMSEA = 0.009$ ,  $CFI = 0.999$  และ  $TLI = 0.998$  ผลการศึกษาพบว่าการรับรู้ถึงแนวทางการปฏิบัติด้านข้อมูลนั้นไม่มีผลกระทบต่อความตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง แต่การรับรู้ความเสี่ยงด้านความมั่นคงปลอดภัยของข้อมูลและการรับรู้ความเสี่ยงด้านความเป็นส่วนตัวของผู้บริโภคมีอิทธิพลเชิงลบทั้งทางตรงและทางอ้อมต่อความตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งผ่านการรับรู้ถึงแนวทางการปฏิบัติด้านข้อมูล

**คำสำคัญ:** อินเทอร์เน็ตทุกสรรพสิ่ง การรับรู้ถึงแนวทางการปฏิบัติด้านข้อมูล การรับรู้ถึงความเสี่ยงด้านความมั่นคงปลอดภัยของข้อมูล การรับรู้ถึงความเสี่ยงด้านความเป็นส่วนตัว

### Abstract

The purpose of this study was to examine the impact of perceived security risk and privacy risk on

the acceptance of IoT technologies. The study used online questionnaire to collect a sample of 396 people who have experience on IoT devices and live in Thailand. The data were analyzed through Structural Equation Model to test the hypotheses and derive model fit. The model showed a goodness-of-fit with  $P\text{-Value} = 0.406$ ,  $\chi^2/df = 1.035$ ,  $RMSEA = 0.009$ ,  $CFI = 0.999$  and  $TLI = 0.998$ . According to analysis results, the data practice perception has no significant effect on the intention to accept IoT technologies. The findings also show that the security risk perception and the privacy risk perception have both negative direct and indirect affect influence on the intention to accept IoT technologies through the data practice perception.

**Keywords:** Internet of things, Data Practice Perception, Security Risk Perception, Privacy Risk Perception.

### 1. บทนำ

ปัจจุบันเทคโนโลยีมีส่วนอย่างมากในการใช้ชีวิตประจำวัน ซึ่งในระหว่างวันนี้แทบจะทุกกิจกรรมเรามีการใช้อุปกรณ์อิเล็กทรอนิกส์เข้ามาช่วย ซึ่งอุปกรณ์หลายชนิดเป็นอุปกรณ์ที่มีการนำเทคโนโลยีเกี่ยวกับอินเทอร์เน็ตทุกสรรพสิ่งเข้ามาช่วย เพื่อให้อุปกรณ์สามารถตอบสนองหรือเพิ่มความสะดวกสบายในการใช้เทคโนโลยีมากขึ้น เช่น บ้านอัจฉริยะที่สามารถเปิดปิดไฟ

ได้อัตโนมัติ สามารถตรวจสอบตำแหน่งและทำนายเวลาที่คนเปิดหรือปิดแอร์ได้ ส่งผลให้เกิดความสะดวกสบายและลดการใช้พลังงานได้มาก [1] ซึ่งความสะดวกสบายและมีประโยชน์นี้นำไปสู่ความแพร่หลายในการใช้เทคโนโลยี โดยบางคนอาจจะไม่รู้ตัวด้วยซ้ำว่ากำลังใช้เทคโนโลยีอะไรอยู่บ้าง

ผู้ให้บริการอุปกรณ์อินเทอร์เน็ตทุกสรรพสิ่งรายใหญ่สามารถนำข้อมูลวิเคราะห์ เพื่อพัฒนาหรือออกแบบผลิตภัณฑ์ให้ตรงกับความต้องการของผู้บริโภคมากขึ้น ในราคาที่ถูกลง แต่ก็แลกมากับการนำข้อมูลส่วนบุคคลของผู้บริโภคมาวิเคราะห์ โดยทั่วไปหากนำข้อมูลมาวิเคราะห์โดยสุจริตและถูกกำกับด้วยกฎหมายและมาตรฐานในการปกป้องข้อมูลก็ยังไม่เป็นปัญหามากนัก แต่ในปัจจุบันมีข่าว เหตุการณ์ต่าง ๆ มากมายที่เกิดขึ้นโดยผู้ให้บริการนำข้อมูลส่วนบุคคลไปใช้ในทางที่ผิดหรือไม่มีการควบคุมที่ดี และจุดอ่อนของเทคโนโลยีนี้คือการเก็บข้อมูลที่ละเอียดอ่อนสามารถสื่อถึงตัวตนและพฤติกรรมของเจ้าของข้อมูลได้ เป็นจุดสำคัญที่องค์กร สามารถสร้างรายได้จากการรวบรวมและขายข้อมูลให้กับองค์กรอื่น ๆ ต่อไป ทำให้ข้อมูลเหล่านี้เป็นที่ต้องการของตลาด จนทำให้ข้อมูลส่วนบุคคลถูกกระจายออกไป [2] ซึ่งอาจถูกส่งต่อไปจนถึงมิถิลาชีพได้โดยวิธีใดวิธีหนึ่ง

การรับรู้และความมั่นคงปลอดภัยของข้อมูลส่วนบุคคลของผู้บริโภคในปัจจุบันมีผลอย่างมากในการดำรงชีวิตและอาจส่งผลกระทบต่อเหตุการณ์ใหญ่ ๆ ตัวอย่างเช่น เหตุการณ์ [3] ใช้ข้อมูลของสมาชิกผู้ใช้เฟซบุ๊กกว่า 50 ล้านคน อย่างไม่เหมาะสม ช่วยให้โดนัลด์ ทรัมป์ ชนะการเลือกตั้งประธานาธิบดีเมื่อปี ค.ศ. 2016

จะเห็นว่าเหตุการณ์ที่อ้างอิงนั้นเกิดขึ้นโดยที่ผู้ใช้งานขาดการรับรู้เกี่ยวกับความเสี่ยงด้านความมั่นคงปลอดภัย และข้อมูลส่วนบุคคล ดังนั้นปัญหาสำคัญที่งานวิจัยนี้จะมีส่วนร่วมในการช่วยลดและแก้ปัญหาที่อาจเกิดขึ้นในอนาคตก็คือ ปัญหาด้านการรับรู้และความมั่นคงปลอดภัยของข้อมูลส่วนบุคคลที่มากับเทคโนโลยีใหม่ ๆ ของผู้บริโภค

จึงเป็นที่มาของงานวิจัยนี้ที่ต้องการศึกษาระดับและปัจจัยด้านการรับรู้ถึงความมั่นคงปลอดภัยของข้อมูลและความเป็นส่วนตัวของผู้บริโภคที่ถูกนำมาวิเคราะห์นั้น มีผลอย่างไรต่อการยอมรับการใช้งานเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง

## 2. วรรณกรรมที่เกี่ยวข้อง

ในการศึกษาเรื่องการรับรู้ความเป็นส่วนตัวของผู้บริโภคและความมั่นคงปลอดภัยจากเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งมาใช้ในการวิเคราะห์ ผู้วิจัยได้ทำการทบทวนวรรณกรรมจากแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องพบว่า มีปัจจัยที่มีความเกี่ยวเนื่องและสำคัญ 3 สิ่ง คือ การรับรู้ถึงเทคโนโลยี ความกังวลเกี่ยวกับข้อมูลส่วนบุคคลและความมั่นคงปลอดภัยของข้อมูล [2]

ปัจจุบันเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งกลายเป็นเทคโนโลยีที่ถูกใช้กันอย่างแพร่หลายในทุกกลุ่มประชากร [4] เนื่องจากมีการนำเทคโนโลยีนี้มาประยุกต์เข้ากับอุปกรณ์สวมใส่ต่าง ๆ เครื่องจักรในอุตสาหกรรม รวมไปถึงระบบต่าง ๆ ภายในที่อยู่อาศัยโดยที่เทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งนี้มีประโยชน์มากเมื่อนำมาประยุกต์กับสิ่งอื่นใด ๆ ก็ตาม ก็จะทำให้เราสามารถเก็บข้อมูลมาเพื่อวิเคราะห์และช่วยให้การตัดสินใจที่ยากหรือซับซ้อนจนกลายเป็นแนวคิดที่ว่า การตัดสินใจของทุก ๆ คนควรขึ้นอยู่กับข้อมูลและการวิเคราะห์ [5] เพื่อให้สามารถตัดสินใจได้อย่างมีเหตุผลมากขึ้น [6]

จะเห็นได้ว่าเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งเริ่มแพร่หลายเข้ามาในชีวิตประจำวันซึ่งมีการทำวิจัยเกี่ยวกับการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งโดย [7] วิจัยจากตัวแปรต่าง ๆ เช่น การรับรู้ถึงประโยชน์และการใช้งาน ความง่ายในการใช้งาน ความน่าเชื่อถือ อิทธิพลทางสังคม ความเพลิดเพลินและการรับรู้การควบคุมพฤติกรรม ซึ่งผลการวิจัยได้สรุปออกมาว่ามีแนวโน้มก่อนข้างมากที่ผู้ถูกสำรวจจะยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง

เมื่อมองในด้านของการใช้งานและการตอบโต้ความต้องการของผู้ใช้ เทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง

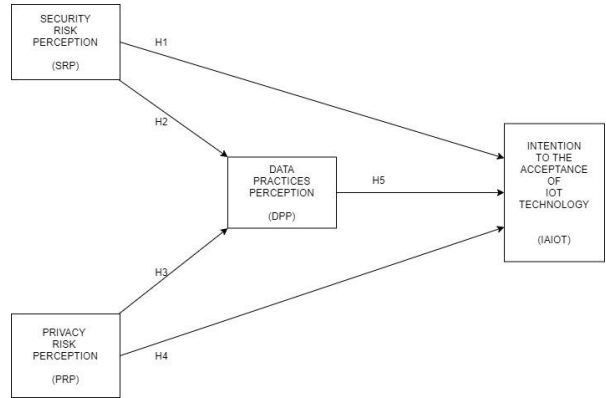
สามารถตอบสนองความคาดหวังได้เป็นอย่างดี แต่เมื่อพิจารณาในด้านความเข้าใจในเทคโนโลยี ความเสี่ยงด้านความมั่นคงปลอดภัยของข้อมูลผู้บริโภค [2], [8] และความเป็นส่วนตัว การวิจัยของ [9] ที่ทำการวิจัยโดยวัดผลกระทบของการรับรู้ความเสี่ยงของข้อมูล (security risk perception) และความเป็นส่วนตัว (privacy risk perception) ผ่านการเปรียบเทียบ เหตุการณ์จำลอง 2 เหตุการณ์และพบว่าได้ผลลัพธ์ที่แตกต่างกัน โดยที่บางเหตุการณ์ความเสี่ยงด้านความมั่นคงปลอดภัยมีผลต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งและบางเหตุการณ์ความเสี่ยงต่าง ๆ ไม่มีผลต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง ซึ่งผลสรุปของงานวิจัยกล่าวว่าผลกระทบของการรับรู้ความเสี่ยงของข้อมูลและความเป็นส่วนตัวนั้นอาจต้องมีตัวแปรอื่นเพื่อนำมาใช้ในการวิเคราะห์

จุดเด่นสำคัญของเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งคือความสะดวกสบาย แต่ความสะดวกสบายนั้นไม่ได้เกิดขึ้นเอง แต่เกิดขึ้นจากการเก็บข้อมูลการใช้งาน รูปแบบความต้องการรวมไปถึงอุปนิสัยส่วนบุคคลของผู้ใช้งานเพื่อนำมาวิเคราะห์และตอบสนองในสิ่งที่ควรจะเป็นความต้องการของผู้ใช้ โดยกระบวนการนี้เป็นส่วนหนึ่งในการศึกษาวิจัยความเข้าใจและการรับรู้ของสาธารณชนเกี่ยวกับการนำแนวทางปฏิบัติด้านข้อมูล (data practices) [15] ที่สื่อความหมายถึงการเก็บรวบรวม การวิเคราะห์และการแบ่งปันข้อมูลอย่างเป็นระบบ มาใช้ร่วมกับการวิเคราะห์ด้วยอัลกอริทึมต่าง ๆ เพื่อให้เข้าใจผู้คน ซึ่งผลลัพธ์ที่ได้ขึ้นอยู่กับอัลกอริทึมที่ใช้ [1] ดังนั้นแนวทางปฏิบัติด้านข้อมูลนั้นถือเป็นส่วนสำคัญของเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งซึ่งผู้วิจัยเล็งเห็นถึงประเด็นสำคัญจากประเด็นนี้ว่าผู้บริโภคได้รับรู้ถึงการนำแนวทางปฏิบัติด้านข้อมูล (data practices perception) มาใช้หรือไม่

ดังนั้นในงานวิจัยนี้จึงต้องการศึกษาเกี่ยวกับการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งโดยพิจารณาจากตัวแปรด้านการรับรู้ถึงความเสี่ยงของข้อมูล การรับรู้ความเสี่ยงของความเป็นส่วนตัว และการรับรู้ถึงแนวทาง

ปฏิบัติด้านข้อมูลว่ามีผลกระทบมากหรือน้อยแค่ไหนกับการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง

จากการที่ผู้วิจัยได้ศึกษาแนวคิด และทฤษฎีที่เกี่ยวข้องสามารถกำหนดกรอบแนวความคิดของการวิจัย ดังแสดงในภาพที่ 1



ภาพที่ 1: กรอบแนวคิดของงานวิจัย

สมมติฐานที่ 1 (H1) : การรับรู้ความเสี่ยงด้านความมั่นคงปลอดภัยส่งผลเชิงลบต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง

สมมติฐานที่ 2 (H2) : การรับรู้ความเสี่ยงด้านความมั่นคงปลอดภัยส่งผลเชิงลบต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งผ่านการรับรู้ถึงการนำแนวทางปฏิบัติด้านข้อมูล

สมมติฐานที่ 3 (H3) : การรับรู้ความเสี่ยงด้านข้อมูลส่วนบุคคลส่งผลเชิงลบต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งผ่านการรับรู้ถึงการนำแนวทางปฏิบัติด้านข้อมูล

สมมติฐานที่ 4 (H4) : การรับรู้ความเสี่ยงด้านข้อมูลส่วนบุคคลส่งผลเชิงลบต่อการตัดสินใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง

สมมติฐานที่ 5 (H5) : การรับรู้ถึงการนำแนวทางปฏิบัติด้านข้อมูลส่งผลโดยตรงต่อการตัดสินใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง

### 3. งานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้เป็น การวิจัยเชิงปริมาณ (Quantitative Research) โดยใช้แบบสอบถามเป็นเครื่องมือเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่าง ในช่วงระหว่างธันวาคม 2564 –

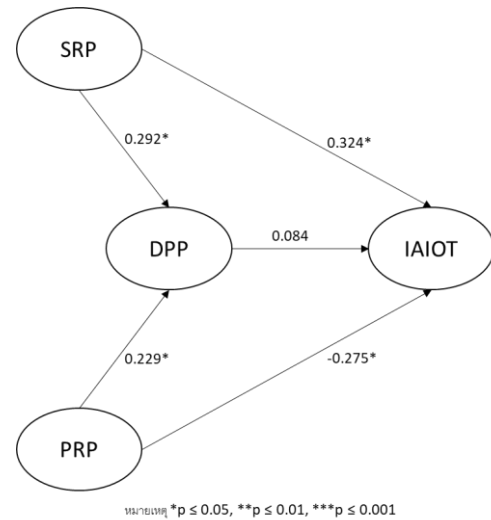
มกราคม 2565 ประชากรคือผู้อาศัยในประเทศไทยและเคยใช้อุปกรณ์ที่มีเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง ผู้วิจัยใช้โปรแกรม G\*Power ในการคำนวณขนาดของกลุ่มตัวอย่าง โดยกำหนดค่าเพาเวอร์ (1-β) เท่ากับ 0.95 ค่าอัลฟา (α) เท่ากับ 0.05 จำนวนตัวแปรทำนายเท่ากับ 4 ตัวแปร ขนาดของอิทธิพล (effect size) เท่ากับ 0.03 ผลที่ได้คือขนาดของกลุ่มตัวอย่างเท่ากับ 396 ตัวอย่าง

การวัดคุณภาพเครื่องมือแบบสอบถาม ทำโดยการวัดความเที่ยงตรง (validity) โดยนำไปให้ผู้ทรงคุณวุฒิ 3 ท่าน แล้วนำมาพิจารณาค่าความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์หรือเนื้อหาด้วยเทคนิค IOC (Index of Item Objective Congruent) พบว่าทุกข้อมีค่าผ่านเกณฑ์มาตรฐานที่ยอมรับได้ แล้วจึงนำมาวัดค่าความเที่ยง (stability) กับกลุ่มตัวอย่างที่มีลักษณะใกล้เคียงกับกลุ่มตัวอย่างที่ต้องการจะเก็บข้อมูลจำนวน 30 ชุด ได้ค่าสัมประสิทธิ์แอลฟา (Alpha Coefficient) ของครอนบาค (Cronbach) เท่ากับ 0.89 ถือได้ว่าแบบสอบถามที่ใช้มีความน่าเชื่อถือได้

จากการนำข้อมูลมาทำการวิเคราะห์ด้วยโมเดลสมการโครงสร้าง (Structural Equation Modeling: SEM) เพื่อทดสอบสมมติฐาน และวิเคราะห์อิทธิพลของตัวแปรอิสระ ตัวแปรตาม และตัวแปรคั่นกลาง (mediator) ในการทดสอบสมมติฐานจะทำการทดสอบที่ระดับนัยสำคัญทางสถิติ 0.05 การวิเคราะห์โมเดลสมการโครงสร้างอาศัยโปรแกรมสำเร็จรูป AMOS ซึ่งเป็นโปรแกรมที่ให้ความแม่นยำสูงในการพิสูจน์ข้อสมมติฐานของงานวิจัย และอธิบายระดับความสัมพันธ์ของตัวแปร

#### 4. ผลการดำเนินงานวิจัย

เมื่อนำข้อมูลมาวิเคราะห์ด้วยโมเดลสมการโครงสร้าง เพื่อทดสอบสมมติฐาน และหาความสัมพันธ์กับข้อมูลเชิงประจักษ์ โมเดลแสดงความเหมาะสมของด้วยค่า P-value = 0.252,  $\chi^2/df = 1.129$ , RMSEA = 0.018, CFI = 0.997 และ TLI = 0.994 ดังแสดงในภาพที่ 2 ส่วนผลการวิเคราะห์อิทธิพลเชิงสาเหตุภายใน โมเดลแสดงในตารางที่ 1



ภาพที่ 2: สัมประสิทธิ์เส้นทางมาตรฐานของโมเดล

ตารางที่ 1: ผลการวิเคราะห์อิทธิพลเชิงสาเหตุภายในโมเดล

ตัวแปรตาม	อิทธิพล	ตัวแปรต้น		
		SRP	PRP	DPP
DPP	ทางตรง	0.239*	0.281*	-
	ทางอ้อม	-	-	-
IAIOT	ทางตรง	0.352*	-0.278	0.033*
	ทางอ้อม	0.009*	0.008*	-

จากตารางที่ 1 สามารถตอบวัตถุประสงค์ของการวิจัย จากสมมติฐานของการวิจัยได้ดังแสดงในตารางที่ 2

ตารางที่ 2: สรุปผลการทดสอบสมมติฐานในการวิจัย

สมมติฐานงานวิจัย	ค่าอิทธิพล	ค่า P	ผลการทดสอบ
H1: การรับรู้ความเล็งด้านความมั่นคงปลอดภัย (SRP) ส่งผลเชิงลบต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง (IAIOT)	0.352	*	ยอมรับ
H2: การรับรู้ความเล็งด้านความมั่นคงปลอดภัย (SRP) ส่งผลเชิงลบต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง (IAIOT) ผ่านการรับรู้ถึงการนำแนวทางปฏิบัติด้านข้อมูล (DPP)	0.009	*	ยอมรับ
H3: การรับรู้ความเล็งด้านข้อมูลส่วนบุคคล (PRP) ส่งผลเชิงลบต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง (IAIOT) ผ่านการรับรู้ถึงการนำแนวทางปฏิบัติด้านข้อมูล (DPP)	0.008	>0.05	ไม่ยอมรับ
H4: การรับรู้ความเล็งด้านข้อมูลส่วนบุคคล (PRP) ส่งผลเชิงลบต่อการตัดสินใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง (IAIOT)	-0.278	*	ยอมรับ
H5: การรับรู้ถึงการนำแนวทางปฏิบัติด้านข้อมูล (DPP) ส่งผลโดยตรงต่อการตัดสินใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง (IAIOT)	0.033	*	ยอมรับ

หมายเหตุ \*p ≤ 0.05, \*\*p ≤ 0.01, \*\*\*p ≤ 0.001

## 5. สรุปผลงานวิจัยและอภิปราย

จากการศึกษาปัจจัยที่ส่งผลกระทบต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง ปัจจัยที่นำมาวิจัยจะเป็นปัจจัยด้านการรับรู้ทั้งหมด โดยที่จะนำระดับการรับรู้ความเล็งด้านความมั่นคงปลอดภัยของข้อมูลและความเป็นส่วนตัวของผู้บริโภคมาวิเคราะห์ร่วมกับการรับรู้ถึงการนำแนวทางการปฏิบัติข้อมูลของผู้ให้บริการอุปกรณ์ที่มีเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง

เมื่อวิเคราะห์ผลงานวิจัยครั้งนี้เปรียบเทียบกับงานวิจัยของ [7] ที่ทำการวิจัยโดยวัดผลกระทบของการรับรู้ความเล็งด้านความมั่นคงปลอดภัยของข้อมูลและความเป็นส่วนตัว ผ่านการเปรียบเทียบ เหตุการณ์จำลอง 2 เหตุการณ์ และพบว่าได้ผลลัพธ์ที่แตกต่างกัน โดยที่บางเหตุการณ์ความเล็งด้านความมั่นคงปลอดภัยของข้อมูลและความเป็นส่วนตัวส่งผลเชิงลบต่อการยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งซึ่งผลสรุปของงานวิจัยกล่าวว่าผลกระทบของการรับรู้ความเล็งของข้อมูลและความเป็นส่วนตัวนั้นอาจต้องมีตัวแปรอื่นเพื่อนำมาใช้ในการวิเคราะห์ ซึ่งสอดคล้องกับภาพรวมงานวิจัยฉบับนี้

โดยที่งานวิจัยฉบับนี้ทำการรับรู้ถึงแนวทางปฏิบัติด้านข้อมูลของผู้ให้บริการมาวิเคราะห์ร่วมกับการรับรู้ความเล็งด้านความมั่นคงปลอดภัยของข้อมูลและความเป็นส่วนตัวของผู้บริโภคนั้นส่งผลกระทบเชิงลบต่อการตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง โดยจะอภิปรายผลเป็นประเด็นต่าง ๆ ดังนี้

การรับรู้ความเล็งด้านความมั่นคงปลอดภัย (SRP) พบว่ามีอิทธิพลเชิงลบเล็กน้อยต่อความตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งทั้งในทางตรงและทางอ้อมผ่านทางรับรู้ถึงแนวทางปฏิบัติด้านข้อมูล (IAIOT)

การรับรู้ความเล็งด้านความเป็นส่วนตัว (PRP) พบว่ามีอิทธิพลเชิงลบเล็กน้อยต่อความตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งทั้งในทางตรงและทางอ้อมผ่านทางรับรู้ถึงแนวทางปฏิบัติด้านข้อมูล

การรับรู้ถึงแนวทางปฏิบัติด้านข้อมูล (DPP) พบว่าไม่มีอิทธิพลโดยตรงต่อความตั้งใจยอมรับเทคโนโลยี

โดยสรุปความตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง ผลการวิจัยพบว่า มีผู้บริโภคจำนวนมากที่มีความกังวลเกี่ยวกับความมั่นคงปลอดภัยของข้อมูลและความเป็นส่วนตัว แต่ผู้บริโภคก็ยังมีความตั้งใจที่จะยอมรับและใช้งานอุปกรณ์ที่มีเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งต่อไปโดยที่ผลกระทบเหล่านั้นสร้างอิทธิพลเชิงลบต่อการตัดสินใจเพียงเล็กน้อยทั้งทางตรงและทางอ้อม แต่สิ่งที่น่าสนใจคือเรื่องของการรับรู้ถึงแนวทางปฏิบัติด้านข้อมูลนั้น กลับไม่มีอิทธิพลต่อความตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง ซึ่งจากงานวิจัยของ [13] พบว่าคนทั่วไปจะเริ่มให้ความสำคัญกับสิ่งใดสิ่งหนึ่งมากขึ้นเมื่อเกิดผลกระทบต่อตนเองหรือเป็นข่าวที่ก่อให้เกิดความวิตกกังวลเป็นวงกว้างตัวอย่างเช่นวิกฤติการณ์โควิด-19 ทำให้ทุกคนเริ่มศึกษาและให้ความสำคัญต่อการใช้ชีวิตและปรับเปลี่ยนพฤติกรรม จึงสันนิษฐานว่าในประเทศไทยการรับรู้ถึงแนวทางปฏิบัติด้านข้อมูล ยังเป็นสิ่งที่ผู้บริโภคไม่ได้ให้ความสนใจเท่าที่ควร ซึ่งผลของงานวิจัยอาจมีความเปลี่ยนแปลงได้ หากว่าในอนาคตเกิดเหตุการณ์ [3], [10] และ [11] ซึ่งส่งผลกระทบอย่างรุนแรงกับสังคม อาจก่อให้เกิดการรับรู้และตื่นตัวที่มากขึ้นซึ่งในประเด็นนี้ยังรวมไปถึงความมั่นคงปลอดภัยของข้อมูลและความเป็นส่วนตัวด้วย เห็นได้จากผลของงานวิจัยที่เกิดผลกระทบเล็กน้อยต่อความตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่ง ซึ่งอาจเป็นผลมาจากข่าวที่เกี่ยวข้องกับการโจมตีทางอินเทอร์เน็ต ไม่ว่าจะเป็นเหตุการณ์โจมตีของแรนซัมแวร์ [12] หรือการออกพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 [14]

ข้อเสนอแนะสำหรับการเพิ่มความตั้งใจยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งนั้น เนื่องจากผู้บริโภคมีความกังวลในเรื่องความมั่นคงปลอดภัยของข้อมูลและความเป็นส่วนตัว ที่อาจส่งผลให้เกิดอันตรายหรือสิ่งที่ไม่พึงประสงค์ต่อตัวผู้บริโภคหรือคนรอบ ๆ ตัว ดังนั้นควรสร้างการรับรู้และความเชื่อมั่น ในด้านของมาตรฐานการเก็บข้อมูลและจริยธรรมด้านการนำข้อมูลไปใช้ว่ามีมีการกำกับและควบคุมอยู่ภายใต้พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล และเผยแพร่ให้สังคมรับรู้ผ่านทุกช่องทาง

ควบคู่กับการนำเสนอด้านความสามารถของผลิตภัณฑ์ เพื่อให้ผู้บริโภคลดความกังวลและเกิดความเชื่อมั่นยิ่งขึ้น

ข้อเสนอแนะในการวิจัยครั้งต่อไป ควรศึกษาในด้าน ระดับของความเข้าใจเทคโนโลยี ความมั่นคงปลอดภัยของ ข้อมูล ความเป็นส่วนตัวและแนวทางปฏิบัติด้านข้อมูลมา ศึกษา ว่าแท้จริงแล้วผู้บริโภคมีการรับรู้และเข้าใจถึงสิ่ง เหล่านี้มากน้อยแค่ไหน จากนั้นก็จะสามารถนำมาศึกษา ควบคู่กับงานวิจัยนี้เพื่อให้ได้ผลลัพธ์และแนวทางการเพิ่ม การยอมรับเทคโนโลยีอินเทอร์เน็ตทุกสรรพสิ่งให้มากขึ้น และยังเป็นการศึกษาเพื่อให้ผู้บริโภคได้สนใจและมีความรู้ ที่สำคัญต่อการเลือกใช้เทคโนโลยีในอนาคตได้มากขึ้น

### เอกสารอ้างอิง

- [1] Z. A. Almusaylim, & N. Zaman. "A review on smart home present state and challenges: linked to context-awareness internet of things (IoT)." *Wireless networks*, Vol. 25, No. 6, pp. 3193-3204, 2019.
- [2] M. Park, H. Oh, & K. Lee. "Security risk measurement for information leakage in IoT-based smart homes from a situational awareness perspective." *Sensors*, Vol. 19, No. 9, pp. 2148, 2019.
- [3] J. Isaak, & M. J. Hanna. "User data privacy: Facebook, Cambridge Analytica, and privacy protection." *Computer*, Vol. 51, No. 8, pp. 56-59, 2018.
- [4] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiq, & I. Yaqoob. "Big IoT data analytics: architecture, opportunities, and open research challenges." *IEEE Access*, Vol. 5, pp. 5247-5261, 2017.
- [5] J. Blythe, & S. Johnson. "The Consumer Security Index for IoT: A protocol for developing an index to improve consumer decision making and to incentivize greater security provision in IoT devices." *Living in the Internet of Things: Cybersecurity of the IoT-2018*, 2018.
- [6] Z. Ma, Y. Ren, X. Xiang, & Z. Turk. "Data-driven decision-making for equipment maintenance." *Automation in Construction*, Vol. 112, 103103, 2020.
- [7] L. Gao, & X. Ba. "A unified perspective on the factors influencing consumer acceptance of internet of things technology." *Asia Pacific Journal of Marketing and Logistics*, 2014.
- [8] D. Kaczorowska-Spychalska. "Digital technologies in the process of virtualization of consumer behaviour—awareness of new technologies." *Management*, Vol. 22, No. 2, 2018.
- [9] R. Ando, S. Shima, & T. Takemura. "Analysis of privacy and security affecting the intention of use in personal data collection in an IoT environment." *IEICE TRANSACTIONS on Information and Systems*, Vol. 99, No. 8, pp. 1974-1981, 2016.
- [10] C. D. McDermott, J. P. Isaacs, & A. V. Petrovski. "Evaluating awareness and perception of botnet activity within consumer internet-of-things (IoT) networks." *Informatics*, 2019.
- [11] วิชกรกุล พัฒนาประทีป. "ไปชม Shopee ดึงรูปผู้ใช้ไม่แจ้ง มีดีและได้พบแอปรุ่นใหม่อาจเก็บรูปในพื้นที่ของแอปแทน." [สืบค้นเมื่อ วันที่ 9 สิงหาคม 2563]. จาก <https://www.beartai.com/news/it-thai-news/473673>.
- [12] Q. Chen, & R. A. Bridges. "Automated behavioral analysis of malware: A case study of wannacry ransomware." *In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 454-460, December, 2017.
- [13] S. Rousseau, & N. Deschacht. "Public awareness of nature and the environment during the COVID-19 crisis." *Environmental and Resource Economics*, Vol. 76, No. 4, pp. 1149-1159, 2020.
- [14] T. Bumpenboon. "Thailand's Personal Data Protection Act: An Understanding from the Perspectives of the European Privacy Law." *Thammasat Review of Economic and Social Policy*, Vol. 6, No. 1, pp. 50-82, 2020.
- [15] H. Kennedy, S. Oman, M. Taylor, J. Bates, & R. Steedman. Public understanding and perceptions of data practices: a review of existing research. Sheffield: The University of Sheffield, 2020.

# การจัดทำชุดข้อมูลตรวจจับการบุกรุกของกองทัพอากาศ เพื่อเปรียบเทียบระหว่างชุดข้อมูลที่เป็นมาตรฐาน Generating an Air Force Intrusion Detection Dataset to Compare between Standardized Datasets

สัพพัญญู ชูแก้ว (Sappanyou Chukaew)<sup>1</sup> ประสงค์ ปราณีตพลกรัง (Prasong Praneetpolgrang)<sup>2</sup>

และ พายัพ ศิรินาม (Payap Sirinam)<sup>3</sup>

<sup>1,2,3</sup>หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีป้องกันประเทศ

สำนักบัณฑิตศึกษา โรงเรียนนายเรืออากาศนวมินทกษัตริยาธิราช

<sup>1</sup>sappanyou@gmail.com, <sup>2</sup>prasongspu@gmail.com, <sup>3</sup>p.sirinam@gmail.com

## บทคัดย่อ

ปัจจุบัน ภัยคุกคามทางไซเบอร์ ได้ส่งผลกระทบต่อหน่วยงานด้านความมั่นคง จึงมีความจำเป็นอย่างยิ่งที่จะต้องมีการตรวจจับการบุกรุก หนึ่งในปัจจัยที่ส่งผลต่อประสิทธิภาพของการตรวจจับการบุกรุก คือ การมีชุดข้อมูลที่มีมาตรฐานเพื่อใช้ในการฝึกฝน และสร้างโมเดล ชุดข้อมูลตรวจจับการบุกรุกมาตรฐาน โดยทั่วไปอาจจะไม่มีความเหมาะสม ล้าสมัย หรือไม่ครอบคลุมการโจมตีที่เกิดขึ้นต่อหน่วยงานความมั่นคง ดังนั้น หน่วยงานความมั่นคงจึงจำเป็นต้องมีกระบวนการจัดทำชุดข้อมูลตรวจจับการบุกรุกนำมาเปรียบเทียบกับชุดข้อมูลตรวจจับการบุกรุกมาตรฐาน เพื่อหาข้อแตกต่างในการโจมตีของแต่ละชุดข้อมูลนำมาสู่การพัฒนารูปแบบการป้องกันภัยคุกคามต่อไปในอนาคต

**คำสำคัญ:** ภัยคุกคามทางไซเบอร์ ระบบตรวจจับการบุกรุกชุดข้อมูล

## Abstract

Cyber threat has affected broad to the security agencies. Therefore, it is absolutely necessary to have intrusion detection system. One of the factors affecting the effectiveness of intrusion detection is to have a standardized dataset for training and modeling. Nowadays, dataset for intrusion detection system could be inappropriate, outdated, or not cover

an attack by security forces. Consequently, security agencies need to prepare for intrusion detection data sets to compare with standard intrusion detection data sets to detect the differences in the attack of each data set which lead to the development of future threat protection models.

**Keywords:** Cyber Threat, Intrusion Detection System (IDS), Dataset.

## 1. บทนำ

ความก้าวหน้าทางอินเทอร์เน็ตทำให้มิติทางไซเบอร์ได้กลายเป็นส่วนหนึ่งของชีวิตมนุษย์ และส่งผลกระทบต่อการใช้งานในชีวิตประจำวัน ปัจจุบันการใช้งานอินเทอร์เน็ตมีจำนวนเพิ่มขึ้นอย่างทวีคูณ ทั้งในด้านการใช้งานส่วนตัว ด้านเศรษฐกิจและสังคม ด้านความมั่นคงและการป้องกันประเทศ ด้วยเหตุนี้ทุกประเทศทั่วโลกจึงมีความเกี่ยวข้องกับมิติไซเบอร์อย่างหลีกเลี่ยงไม่ได้ จากผลสำรวจของ สำนักงานคณะกรรมการดิจิทัลเพื่อเศรษฐกิจและสังคมแห่งชาติ (สดช) ในปี 2564 พบว่า จำนวนประชากรประเทศไทยทั้งหมด 69.88 ล้านคน 85.1% มีการใช้งานอินเทอร์เน็ตระยะเวลาเฉลี่ยอยู่ที่ 6 -10 ชั่วโมงต่อวัน จากการใช้งานอินเทอร์เน็ตที่เพิ่มมากขึ้นทำให้เกิดการไหลผ่านของในมิติไซเบอร์ปริมาณมหาศาล ทำให้เกิดความเสียหายต่อการตกเป็นเหยื่อและถูกภัยคุกคามในมิติไซ

เบอร์จากผู้ไม่ประสงค์ดี นอกจากนั้นภัยที่จะเกิดต่อระบบควบคุมดูแลการใช้งานอินเทอร์เน็ตและเครื่องคอมพิวเตอร์ที่เกี่ยวข้อง และ ส่งผลกระทบต่อความมั่นคงของประเทศในภาพรวมได้เป็นอย่างมาก ตัวอย่างของการโจมตีที่มีผลกระทบต่อความมั่นคงของประเทศได้แก่ ในปี พ.ศ.2559 เว็บไซต์ของหน่วยงานภาครัฐในประเทศไทยถูกโจมตีด้วย Distributed Denial-of-Service (DDoS) จากกลุ่มพลเมืองต่อต้าน Single Gateway ซึ่งทำการรณรงค์ให้สมาชิกช่วยกันโจมตีเว็บไซต์ของรัฐบาล นอกจากนั้นยังตรวจพบการเจาะระบบเข้ามายังฐานข้อมูลเพื่อโจรกรรมข้อมูลมาเผยแพร่รวมถึงการใช้ปฏิบัติการข่าวสารในการลดความน่าเชื่อถือของรัฐบาล และ ในปี พ.ศ.2560 WannaCry มัลแวร์ที่ส่งผลกระทบต่อความมั่นคงในระดับประเทศ เข้าโจมตีหน่วยงานสาธารณสุขของอังกฤษ ผู้ป่วยเกือบ 7,000 รายไม่สามารถเข้ารับบริการได้ WannaCry ได้แพร่กระจายไปยังคอมพิวเตอร์ต่าง ๆ ทั่วโลกมากกว่า 150 ประเทศ รวมไปถึงหน่วยงานความมั่นคงในประเทศไทย การรั่วไหลของข้อมูลโดยกลุ่มผู้ไม่ประสงค์ดี ก่อให้เกิดผลกระทบต่อเสถียรภาพ ความน่าเชื่อถือของประเทศ และหน่วยงานความมั่นคงได้เช่นกัน

จากที่กล่าวมาข้างต้นทำให้เห็นถึงความรุนแรงและผลกระทบในมิติไซเบอร์ต่อความมั่นคงของประเทศและกองทัพอากาศเป็นอย่างมาก การใช้งานชุดข้อมูลการบุกรุกในมิติไซเบอร์ที่เป็นมาตรฐาน อาจจะไม่เพียงพอต่อการวิเคราะห์ ลักษณะ รูปแบบ วิธีการโจมตีในมิติไซเบอร์สำหรับกองทัพอากาศ ผู้วิจัยจึงได้เก็บรวบรวมชุดข้อมูลการบุกรุกในมิติไซเบอร์ของกองทัพอากาศ เพื่อเปรียบเทียบความแตกต่างของลักษณะ รูปแบบ วิธีการโจมตี ในมิติไซเบอร์ระหว่างชุดข้อมูลมาตรฐานและชุดข้อมูลของกองทัพอากาศ

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ระบบตรวจจับผู้บุกรุก (Intrusion Detection System)

ระบบตรวจหาการบุกรุกใช้ตรวจสอบการเข้าถึงการใช้งานในระบบเครือข่ายคอมพิวเตอร์ โดยระบบสามารถ

ทำงานทั้งในรูปแบบของฮาร์ดแวร์หรือซอฟต์แวร์ ซึ่งมีหน้าที่ในการติดตามตรวจสอบรวมถึงการเฝ้าระวัง ข้อมูลการใช้งานภายในระบบเครือข่ายที่มีความผิดปกติ ผ่าฝืนต่อความมั่นคงปลอดภัยภายในองค์กร นำมาวิเคราะห์รูปแบบพฤติกรรมในแพ็กเก็ตข้อมูล (Packet Data) เพื่อค้นหาสิ่งผิดปกติ (Anomaly) แล้วนำเข้าสู่ กระบวนการทำนาย (Prediction) เพื่อตัดสินว่าเป็นเหตุการณ์ บุกรุกจริงแล้วแจ้งเตือน (Alert) ให้ผู้รับผิดชอบระบบทราบเพื่อดำเนินการป้องกัน และแก้ไขต่อไป [1], [8]

โดยสถาปัตยกรรมระบบตรวจจับการบุกรุกโดยทั่วไปมี 2 ระบบ

ระบบที่ 1 ระบบตรวจจับผู้บุกรุกบนเครือข่าย (Network-based Intrusion Detection System : NIDS) [3], [4] เป็นระบบตรวจจับการบุกรุกและวิเคราะห์ข้อมูลที่รับส่งกันในเครือข่าย หากความผิดปกติหรือว่ามีผู้บุกรุกหรือไม่ โดยดักจับข้อมูลบน Network Segment และนำมาวิเคราะห์กับรูปแบบการบุกรุกที่กำหนดในฐานข้อมูลการบุกรุก

ระบบที่ 2 ระบบตรวจจับผู้บุกรุกภายในเครื่อง (Host-based Intrusion Detection System : HIDS) [3], [4] เป็นระบบตรวจจับการบุกรุกที่รวบรวมข้อมูลจากแต่ละเครื่องคอมพิวเตอร์ เพื่อตรวจสอบหาความผิดปกติหรือว่ามีผู้บุกรุกหรือไม่ เช่น มีการลบไฟล์ เปลี่ยนสิทธิ์ ในช่วงเวลาที่ผิดปกติ ระบบ HIDS

จะถูกติดตั้งในเครื่องคอมพิวเตอร์ เพื่อตรวจสอบความผิดปกติภายในเครื่องคอมพิวเตอร์ที่ถูกใช้งาน

### 2.2 ชุดข้อมูลทั่วไป

ในส่วนนี้เป็นชุดข้อมูลของระบบตรวจจับการบุกรุก โดยแต่ละชุดข้อมูลจะมีความแตกต่างกัน แสดงให้เห็นถึงความไม่ครอบคลุมของข้อมูลภายในชุดข้อมูลนั้น [2], [4]

KDD'99 (University of California, Irvine 1998-99): เป็นชุดข้อมูลที่บันทึกข้อมูลในระบบเครือข่ายที่พัฒนามาจากชุดข้อมูล Defense Advanced Research Projects Agency (DARPA) 98 บันทึกรูปแบบภัยคุกคาม 31 รายการ [5], [7]

Kyoto (Kyoto University 2009): เป็นชุดข้อมูลที่สร้างขึ้นผ่านการโจมตี Honeypot เป็นการรับส่งข้อมูล



ภายในระบบเครือข่ายบันทึกรูปแบบภัยคุกคาม 14 รายการแบบเดียวกับชุดข้อมูล KDD'99 บวกกับภัยคุกคามใหม่ 10 รายการที่เกี่ยวข้องกับแฮคเกอร์ของระบบตรวจจับการบุกรุกที่หลากหลาย

UNSW-NB15 (University of New South Wales, Canberra, Australia, 2017): เป็นชุดข้อมูลที่สร้างจากการขาดชุดข้อมูลที่ทันสมัยของภัยคุกคามทางไซเบอร์ การสร้างชุดข้อมูลใช้เครื่องสังเคราะห์รูปแบบการโจมตี โดยมีการโจมตีแตกต่างกัน 9 ชนิด ซึ่งชุดข้อมูลมีในรูปแบบไฟล์ PCAP รวมการบันทึกการโจมตีเครือข่ายในรูปแบบเดียวกับ ชุดข้อมูล KDD'99 [5]

### 2.3 นโยบายความมั่นคงแห่งชาติ พ.ศ. 2558 - 2564

นโยบายความมั่นคงแห่งชาติ พ.ศ. 2558 - 2564 ได้กล่าวถึงบริบทการเปลี่ยนแปลง ที่นำไปสู่ภัยคุกคามรูปแบบใหม่ไว้ว่า ประเด็นที่มีความเสี่ยง จากการถูกโจมตี และการจารกรรมทางไซเบอร์ เนื่องจากการกำหนดมาตรการป้องกันทำได้ยากและไม่ทันต่อความก้าวหน้าทางเทคโนโลยี ดังนั้น นโยบายความมั่นคงแห่งชาติฉบับนี้ จึงได้กำหนดนโยบายความมั่นคงไว้รองรับ ในส่วนที่ 2 นโยบายความมั่นคงแห่งชาติทั่วไป นโยบายที่ 10 เสริมสร้างความมั่นคงทางเทคโนโลยีสารสนเทศและไซเบอร์ ดังนี้

ข้อ 10.1 ปกป้อง ป้องกัน ภัยคุกคามด้านไซเบอร์ สงครามไซเบอร์ และเสริมสร้าง ความปลอดภัยระบบเทคโนโลยีสารสนเทศ โดยการบูรณาการจัดการความมั่นคงทางไซเบอร์ระหว่างหน่วยงานภาครัฐ การประสานความร่วมมือและเสริมสร้างเครือข่ายกับภาคเอกชน ภาควิชาการ บุคลากร องค์กรและผู้เชี่ยวชาญทางด้านการรักษาความมั่นคงทางไซเบอร์

ข้อ 10.2 พัฒนาการบังคับใช้กฎหมาย โดยการพัฒนาระเบียบและกฎหมายเพื่อความมั่นคงปลอดภัยไซเบอร์ และการพัฒนาเทคโนโลยีสำหรับงานสืบสวนและป้องกันอาชญากรรมไซเบอร์ให้สามารถลดภัยคุกคามหรืออันตรายที่ส่งผลกระทบต่อบุคคล ข้อมูลและระบบเทคโนโลยีสารสนเทศ

ข้อ 10.3 พัฒนาศักยภาพทางด้านเทคโนโลยีสารสนเทศ โดยส่งเสริมการวิจัย พัฒนา และ จดสิทธิบัตร

เทคโนโลยีสารสนเทศที่ผลิตโดยคนไทย การวิจัยและพัฒนาเพื่อความมั่นคงปลอดภัยไซเบอร์ การบูรณาการเชื่อมโยงระบบฐานข้อมูลภาครัฐ

### 2.4 ยุทธศาสตร์ป้องกันประเทศ กระทรวงกลาโหม พ.ศ. 2560 - 2579

ยุทธศาสตร์ป้องกันประเทศ กระทรวงกลาโหม พ.ศ.2560 – 2579 บทที่ 4 แนวความคิด และประเด็น ยุทธศาสตร์ ได้กำหนดมาตรการและขีดความสามารถที่ต้องการด้านการปฏิบัติการด้านไซเบอร์ ในประเด็น ยุทธศาสตร์ที่ 6 การปฏิบัติการทางทหารเพื่อรักษาอธิปไตย และผลประโยชน์แห่งชาติ ดังนี้

ด้านการปฏิบัติการด้านไซเบอร์ ระยะที่ 1 (พ.ศ.2560 - 2564) พัฒนากำลังพล โครงสร้างพื้นฐาน และเทคโนโลยี ให้มีความพร้อมในการปฏิบัติในมิติไซเบอร์ สร้างความตระหนักรู้ทางไซเบอร์ให้กับทุกภาคส่วนและ สร้างความร่วมมือทางไซเบอร์ทั้งในและต่างประเทศ ระยะที่ 2 (พ.ศ. 2565 - 2569) และ ระยะที่ 3 (พ.ศ.2570 - 2574) พัฒนาศักยภาพและ ขีดความสามารถในการปฏิบัติในมิติไซเบอร์ เพื่อให้มีพลังอำนาจทางไซเบอร์ที่มีประสิทธิภาพอย่างต่อเนื่อง ระยะที่ 4 (พ.ศ.2575 - 2579) เพิ่มศักยภาพและ ขีดความสามารถในการปฏิบัติในมิติไซเบอร์ให้อยู่ในแนวหน้าและเป็นที่ยอมรับในระดับภูมิภาคเอเชียตะวันออกเฉียงใต้

### 2.5 งานวิจัยที่เกี่ยวข้อง

ธริช พงษ์รชนี [7] นำเสนอสารนิพนธ์ฉบับนี้เป็นการออกแบบและจัดการศูนย์ปฏิบัติการความมั่นคงปลอดภัยระบบสารสนเทศ (Security Operation Center Design and Management) ขนาดเล็ก (Mini SOC : Web Server) โดยเก็บเฉพาะ Log ของ Web Server เท่านั้น เพื่อใช้สำหรับการวิเคราะห์และเฝ้าระวัง (Monitor) ภัยคุกคามทางไซเบอร์ที่อาจจะเกิดขึ้นกับระบบ และหาแนวทางรับมือ ยับยั้ง ป้องกันภัย คุกคาม ที่อาจจะเกิดขึ้นกับระบบ ก่อให้เกิดความเสียหายกับองค์กรได้ โดยการส่ง Log จาก Web Server มายังระบบวิเคราะห์ Log (SIEM) และใช้ AlienVault SIEM ในการวิเคราะห์ Log ซึ่ง AlienVault SIEM เป็นซอฟต์แวร์โอเพนซอร์ส (Open Source Software – OSS) ที่สามารถนำมาใช้ งานได้โดยไม่มี

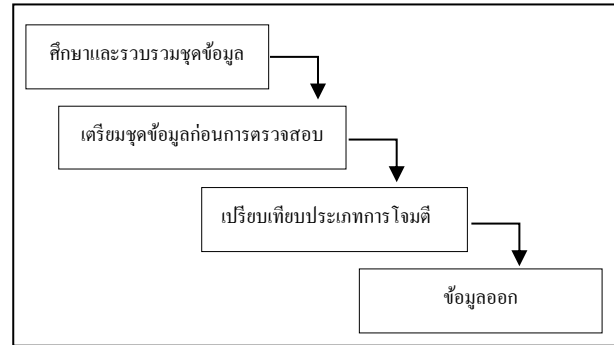
ค่าใช้จ่ายอีกทั้งช่วยลดต้นทุนในองค์กร และไม่เสี่ยงต่อการละเมิดลิขสิทธิ์

ธนกร มีหิнокง [8] นำเสนอแบบสถาปัตยกรรมความรู้ด้านความมั่นคงปลอดภัยไซเบอร์เพื่อสนับสนุนระบบตรวจหาการบุกรุกแบบปรับตัวใหม่ โดยใช้ตัวแบบประเมินความเสี่ยงด้านความมั่นคงปลอดภัยไซเบอร์ของสถาบันการพลศึกษาตามมาตรฐาน ISO/IEC 27005 มาตรฐาน ISO/DIS 31000 และมาตรฐาน OCTAVE เพื่อตรวจสอบวัดผลปัจจัยความเสี่ยงด้านต่าง ๆ ควบคู่ไปกับการตรวจสอบจากระบบตรวจหาการบุกรุกแบบปรับตัวด้วยเทคนิคเหมืองข้อมูลจากการวิเคราะห์ด้วยกฎความสัมพันธ์จากโครงข่ายประสาทเทียม จากผลการทดลองพบว่าระบบตรวจหาการบุกรุกที่ได้พัฒนาขึ้นนี้สามารถรายงานผลได้อย่างรวดเร็ว โดยมีค่าความเที่ยงที่ 97.4% และค่าเรียกคืนที่ 92.0% จึงสามารถนำไปใช้วิเคราะห์ และทำนายผลการรักษาความมั่นคงปลอดภัยไซเบอร์ได้ต่อไป

### 3. วิธีการดำเนินงาน

งานวิจัยนี้มีวัตถุประสงค์เพื่อการจัดทำชุดข้อมูลตรวจจับการบุกรุกของกองทัพอากาศเพื่อเปรียบเทียบระหว่างชุดข้อมูลที่เป็นมาตรฐานในบทนี้จะกล่าวถึงขั้นตอนการวิจัยและการใช้เครื่องมือที่ใช้ในการวิจัย โดยมีรายละเอียดตามภาพที่ 1 ดังนี้

- 1) ศึกษาและรวบรวมชุดข้อมูล
- 2) เตรียมชุดข้อมูลก่อนการตรวจสอบ เพื่อจัดลำดับในการประมวลผล
- 3) เปรียบเทียบประเภทการโจมตี
- 4) ข้อมูลออก คือ แสดงให้เห็นถึงความจำเป็นในการสร้างชุดข้อมูลภัยคุกคามของกองทัพอากาศว่ามีภัยคุกคามที่ต่างออกไปจากชุดข้อมูลทั่วไปจะได้นำมาเป็นแนวทางในการพัฒนาแบบจำลองภัยคุกคามทางไซเบอร์เพื่อส่งเสริมการพึ่งพาตนเองและรักษาความมั่นคงทางไซเบอร์ของกองทัพอากาศ ต่อไป



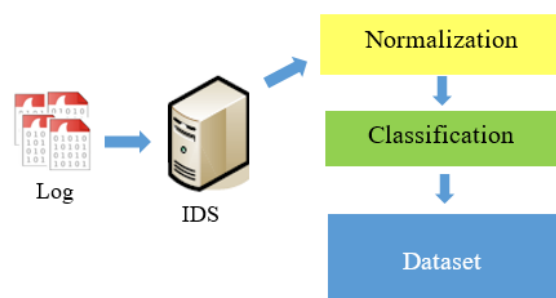
ภาพที่ 1: แสดงภาพรวมการทำงาน

### 3.1 วิธีการจัดเตรียมชุดข้อมูล

ซึ่งในงานวิจัยนี้ได้จัดเตรียมชุดข้อมูลโดยมีขั้นตอนการรวบรวมดังนี้

- 1) ดักจับข้อมูลการจราจรทางคอมพิวเตอร์ (Log) จากอุปกรณ์ภายในกองทัพอากาศ
- 2) นำข้อมูลจราจรทางคอมพิวเตอร์ (Log) เข้าวิเคราะห์ระบบตรวจจับผู้บุกรุก เพื่อเป็นข้อมูลตั้งต้นในการเตรียมชุดข้อมูล
- 3) ทำการจัดทำข้อมูลให้เป็นมาตรฐาน (Normalization)
- 4) จำแนกข้อมูลภัยคุกคาม (Classification)

จากขั้นตอนที่กล่าวมาข้างต้นจึงได้ชุดข้อมูลเพื่อนำมาทำการเปรียบเทียบประเภทการโจมตี ดังภาพที่ 2



ภาพที่ 2 : แสดงภาพรวมจัดเตรียมชุดข้อมูล

### 4. ผลการดำเนินงาน

การวิจัยครั้งนี้ ผู้วิจัยได้ใช้ชุดข้อมูล UNSW-NB15 ซึ่งเป็นชุดข้อมูลซึ่งเป็นมาตรฐาน นำมาเปรียบเทียบระหว่างชุดข้อมูลของกองทัพอากาศ โดยชุดข้อมูลทั้งสองชุดข้อมูล

นี้จะผ่านการเตรียมข้อมูลก่อนการตรวจสอบเพื่อปรับค่าให้มีความเหมาะสม แล้วจำแนกประเภทการโจมตี ดังตารางที่ 1 และตารางที่ 2 จะเห็นได้ว่าข้อมูลทั้งสองตารางมีความแตกต่างของประเภทการโจมตีอย่างมีนัยสำคัญ

ตารางที่ 1: ประเภทการโจมตีภายในชุดข้อมูล UNSW-NB15

ประเภทการโจมตี	จำนวน	อัตราส่วน
Normal	56,000	31.94%
Fuzzers	18,184	10.37%
Analysis	2,000	1.14%
Backdoors	1,746	1.00%
DoS	12,264	6.99%
Exploits	33,393	19.04%
Generic	40,000	22.81%
Reconnaissance	10,491	5.98%
Shellcode	1,133	0.65%
Worms	130	0.07%
ALL	175,341	100.00%

ตารางที่ 1 ชุดข้อมูล UNSW-NB15 มีการบันทึกประเภทการโจมตี ทั้งหมด 175,341 รายการ โดยจำแนกประเภทการโจมตี ได้แก่ Normal, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode และ Worms แบ่งส่วนการโจมตีเป็นเปอร์เซ็นต์ตามตารางที่ 1 โดยในชุดข้อมูลนี้มีข้อมูลที่เป็นปกติสูงและตามด้วยการโจมตีที่เป็นการ Generic หรือการโจมตีที่สร้างขึ้นมาจากเครื่องมือผู้จัดทำชุดข้อมูลนี้ [5]

ตารางที่ 2: ประเภทการโจมตีภายในชุดข้อมูลของกองทัพอากาศ

ประเภทการโจมตี	จำนวน	อัตราส่วน
Normal	1,647	0.47%
Script Control Command	308,627	87.17%
Script Control - Path	9,687	2.74%
Backdoors	2,576	0.73%
Dropper	4,797	1.35%
PUA	9,565	2.70%
Ransomware	3,209	0.91%
Spyware	4,187	1.18%
Virus	8,545	2.41%
Worm	1,226	0.35%
ALL	354,066	100.00%

ตารางที่ 2 ชุดข้อมูลของกองทัพอากาศ มีการบันทึกประเภทการโจมตี ทั้งหมด 354,066 รายการโดยจำแนกประเภทการโจมตี ได้แก่ Normal, Script Control - Command, Script Control - Path, Backdoors, Dropper, Potentially Unwanted Application (PUA), Ransomware, Spyware, Virus และ Worm แบ่งส่วนการโจมตีเป็นเปอร์เซ็นต์ตามตารางที่ 2 โดยในชุดข้อมูลนี้มีการโจมตีในรูปแบบ Script Control - Command สูงที่สุดมากถึง 87.17%

จากผลลัพธ์ของการวัดประสิทธิภาพชุดข้อมูลตรวจจับการบุกรุกของทั้ง 2 ชุดข้อมูลแล้ว เมื่อนำมาเปรียบเทียบกันจะเห็นได้ว่าชุดข้อมูลของกองทัพอากาศมีการโจมตีที่เป็นลักษณะเฉพาะ ทำให้เมื่อมีการวิเคราะห์หรือพัฒนารูปแบบการป้องกันภัยคุกคามที่ใช้ในหน่วยงานจึงจำเป็นต้องจัดทำชุดข้อมูลเป็นของตัวเอง

จากการวิเคราะห์เพิ่มเติม Script Control Command เป็นชุดคำสั่งในการประมวลผลซึ่งมีทั้งในรูปแบบการโจมตีที่เกิดขึ้นจริง และชุดคำสั่งปกติที่มีความเสี่ยงที่ถูกจับได้โดยระบบตรวจภัยคุกคามที่มีประสิทธิภาพของกองทัพอากาศ จึงเป็นแนวทางในการพัฒนารูปแบบเพื่อหาคำความแม่นยำในการจำแนกภัยคุกคามทางไซเบอร์ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

## 5. สรุป

งานวิจัยนี้ผู้วิจัยได้ศึกษาข้อมูลระบบตรวจจับผู้บุกรุกในมิติไซเบอร์ เพื่อเก็บรวบรวมข้อมูลการโจมตีภายในกองทัพอากาศ และ นำมาวิเคราะห์จำแนกประเภทการโจมตีชุดข้อมูลของกองทัพอากาศ หากการโจมตีที่ส่งผลกระทบต่อการทำงานในมิติไซเบอร์ของกองทัพอากาศแล้วนำผลที่ได้จากการจำแนกประเภทการโจมตีมาเปรียบเทียบกับชุดข้อมูลมาตรฐาน หากความเหมาะสมในการใช้ชุดข้อมูลหรือประสานชุดข้อมูล เพื่อเพิ่มประสิทธิภาพในการพัฒนาระบบป้องกันภัยคุกคามในมิติไซเบอร์ของกองทัพอากาศในอนาคต

จากผลการวิจัยทั้ง 2 ชุดข้อมูลพบว่า ชุดข้อมูล UNSW-NB15 มีรูปแบบการโจมตีที่แตกต่างจากชุดข้อมูลของกองทัพอากาศ ทางกองทัพอากาศจึงจำเป็นต้องจัดทำชุดข้อมูลของกองทัพอากาศ เพื่อให้เหมาะสมในการวิเคราะห์

รูปแบบการโจมตีที่ส่งผลเสียในมิติไซเบอร์ของ กองทัพอากาศ ตอบสนองนโยบายความมั่นคงแห่งชาติ พ.ศ.2558 - 2564 ปกป้อง ป้องกัน ภัยคุกคามด้านไซเบอร์ สงครามไซเบอร์ และเสริมสร้าง ความปลอดภัยระบบ เทคโนโลยีสารสนเทศ และ ยุทธศาสตร์ป้องกันประเทศ กระทรวงกลาโหม พ.ศ.2560 – 2579 จัดความสามารถ ด้านการปฏิบัติการด้านไซเบอร์

นอกจากนี้ ผู้วิจัยพบแนวทางในการพัฒนาเพิ่ม ประสิทธิภาพความแม่นยำการจำแนกการโจมตีใน มิติไซเบอร์ของชุดข้อมูลการบุกรุกกองทัพอากาศ โดยจะ ใช้ระบบปัญญาประดิษฐ์มาพัฒนารูปแบบการจำแนกการ โจมตีให้ได้ค่าความแม่นยำมากยิ่งขึ้นต่อไปในอนาคต

### เอกสารอ้างอิง

- [1] Muataz Salam Al-Daweri, Khairul Akram Zainol Arin, and Mohamad Firham Efendy Md. Senan, "An Analysis of the KDD99 and UNSW-NB15 Datasets for the Intrusion Detection System," *IEEE Trans. on Instrumentation and Measurement Centre for Artificial Intelligence Technology, Universiti Kebangsaan Malaysia*, Oct 13, 2020.
- [2] Carlos Garcia Cordero, Emmanouil Vasilomanolakis, Simin Nadjm-Tehrani, Aidmar Wainakh and Max Mühlhäuser, "On Generating Network Traffic Datasets with Synthetic Attacks for Intrusion Detection," *ACM Transactions on Privacy and Security*, Vol. 24, No. 2, Article 8, December, 2020.
- [3] Somayye Hajiheidari, Karzan Wakil, Maryam Badri, Nima Jafari Navimipour, "Intrusion detection systems in the Internet of things: A comprehensive investigation" *Computer Networks*, Vol. 160, Sep 4, pp 165-191, 2019.
- [4] Sharafaldin, I., Lashkari A., and Ghorbani, A., "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization" *IEEE Trans. on Industrial Electronics International Conference on Information Systems Security and Privacy (ICISSP 2018)*, pp. 108-116, 2018.
- [5] Atilla Ozg'ur and Hamit Erdem, "A review of KDD99 dataset usage in intrusion detection and machine learning between 2010 and 2015," *Baskent University, Turkey*, Apr 14, 2016.
- [6] V. Jaiganesh, S. Mangayarkarasi, and Dr. P. Sumathi, P. Meesad, "Intrusion Detection Systems: A Survey and Analysis of Classification Techniques," *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 2, Issue 4, April, 2013.
- [7] ชรัช พงษ์รัตน์, "ระบบเฝ้าระวังและตรวจจับภัยคุกคามเว็บเซิร์ฟเวอร์ Mini SOC : Web Server" วิทยาสตรมหาบัณฑิต สาขาวิศวกรรมเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ แขนงความมั่นคงปลอดภัยไซเบอร์ คณะวิทยาการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีมหานคร, 2652.
- [8] ธนกร มีหินกอง, "สถาปัตยกรรมความรู้ด้านความมั่นคงปลอดภัยไซเบอร์เพื่อสนับสนุนระบบตรวจหาการบุกรุกแบบปรับตัวด้วยเทคนิคกฎความสัมพันธ์" *วารสารวิชาการพระจอมเกล้าพระนครเหนือ* Vol. 25 No. 2 พฤษภาคม – สิงหาคม, pp. 277-288, 2558.
- [9] จุติชัย ทองกระจาย, "ระบบตรวจจับการบุกรุกเครือข่ายกรณีศึกษา บริษัท เวนต้า ซอฟต์แวร์ ดีเวลอปเมนท์ จำกัด", วิทยาสตรมหาบัณฑิต สาขาวิชาความมั่นคงทางระบบสารสนเทศ คณะวิทยาการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีมหานคร, 2557.

# กระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ

## Continuous Integration and Continuous Deployment on Cloud Computing using Gitlab

ธัญธิชา นิ่มนวน (Thanthicha Nimnuan)<sup>1</sup> และผุสดี บุญรอด (Pudsadee Boonrawd)<sup>2</sup>

<sup>1,2</sup>ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

s6307011858110@email.kmutnb.ac.th, pudsadee.b@itd.kmutnb.ac.th

### บทคัดย่อ

การพัฒนาเว็บแอปพลิเคชันรูปแบบเดิมใช้กระบวนการรวบรวมและนำส่งซอฟต์แวร์ที่มีหลายขั้นตอน มีความซ้ำซ้อน ทำให้อาจต้องใช้ระยะเวลาในการพัฒนา และเกิดข้อผิดพลาดได้ในแต่ละกระบวนการ วัตถุประสงค์ของงานวิจัยนี้จึงได้ทำการสร้างกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ ซึ่งมีกระบวนการหลัก 3 ขั้นตอน ได้แก่ การรวบรวมซอร์สโค้ดมาสร้างให้อยู่ในรูปแบบคอนเทนเนอร์โดยใช้ Docker การทดสอบซอฟต์แวร์อัตโนมัติ และการนำส่งซอฟต์แวร์แบบอัตโนมัติ ซึ่งกระบวนการเหล่านี้จะทำงานบนกิตแล็บ จากผลการทดสอบกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ พบว่า สามารถช่วยลดระยะเวลาและขั้นตอนการทำงาน ทำให้เกิดข้อผิดพลาดในการทำงานน้อยลง ทำให้ผู้ให้บริการสามารถแก้ไขปรับปรุงแอปพลิเคชันได้อย่างสะดวกรวดเร็ว และตรงตามความต้องการของผู้ใช้งาน

**คำสำคัญ:** คลาวด์คอมพิวติ้ง กิตแล็บ การรวบรวมซอฟต์แวร์แบบอัตโนมัติ การนำส่งซอฟต์แวร์แบบอัตโนมัติ

### Abstract

Traditional web application development uses Continuous Integration and Continuous Deployment, which are multi-step, redundant, time-consuming, and potentially faulty in each process. The objective of this research was to develop the process of Continuous Integration and Continuous Deployment

via Cloud Computing using Gitlab. The main process consists of three steps: Continuous Integration as a docker-based container, Automated Software Testing, and Continuous Deployment. The testing results of Continuous Integration and Continuous Deployment showed that it can help reduce the time and workflow, reduce the error of the operation, improve the application quickly and easily, and meet the needs of users.

**Keywords:** Cloud Computing, Gitlab, Continuous Integration, Continuous Deployment.

### 1. บทนำ

ปัจจุบันเทคโนโลยีเว็บแอปพลิเคชันมีการพัฒนาเพื่อให้ผู้ให้บริการและผู้ใช้บริการสามารถเข้าถึงได้อย่างรวดเร็ว โดยผู้ให้บริการสามารถแก้ไขและติดตั้งเว็บแอปพลิเคชันบนคลาวด์คอมพิวติ้ง โดยไม่ต้องคำนึงถึงระบบปฏิบัติการของผู้ให้บริการ จากการศึกษางานวิจัยพบว่า เทคโนโลยีเว็บแอปพลิเคชันได้ถูกพัฒนาขึ้น เพื่อให้เทคโนโลยีเว็บแอปพลิเคชันรูปแบบเดิมสามารถรองรับการเขียนโปรแกรมหรือการทำงานร่วมกันกับฐานข้อมูลให้มีประสิทธิภาพมากยิ่งขึ้น และด้วยปัจจัยด้านเทคโนโลยีที่ถูกพัฒนาไปอย่างมาก ไม่ว่าจะเป็นความเร็วของอินเทอร์เน็ต หรือประสิทธิภาพของเครื่องคอมพิวเตอร์แม่ข่าย ซึ่งเทคโนโลยีเว็บแอปพลิเคชันนั้นสามารถรองรับการทำงานกับบุคคลจำนวนมากได้ในเวลาเดียวกัน [1]

ดังนั้นคณะผู้พัฒนาจึงสังเกตเห็นปัญหาจากกระบวนการรวบรวมและนำส่งซอฟต์แวร์รูปแบบเดิม ที่มีขั้นตอนในแต่ละ

กระบวนการที่ช้าซ้อน ทำให้ไม่สามารถเข้าถึงผู้ใช้งานได้อย่างรวดเร็ว จึงทำให้งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ ซึ่งประกอบด้วย 3 ขั้นตอนได้แก่ การรวบรวมซอร์สโค้ดให้อยู่ในรูปแบบคอนเทนเนอร์โดยใช้ดีออกเกอร์ การทดสอบซอฟต์แวร์แบบอัตโนมัติ และการนำส่งซอฟต์แวร์แบบอัตโนมัติ ซึ่งในบทความวิจัยนี้ได้แบ่งเนื้อหาออกได้เป็น 5 ส่วน ได้แก่ ส่วนที่ 1 กล่าวถึงบทนำ ส่วนที่ 2 นำเสนอทฤษฎีและงานวิจัยที่เกี่ยวข้อง ส่วนที่ 3 วิธีดำเนินการวิจัย ส่วนที่ 4 ผลการดำเนินงานวิจัย และส่วนที่ 5 สรุปผลการวิจัย

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 กระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ

กระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ (Continuous Integration และ Continuous Deployment) เป็นกระบวนการรวบรวมซอฟต์แวร์ที่มีการพัฒนาซอร์สโค้ดแบบแยกส่วนจากกัน และนำมารวบรวมเป็นซอฟต์แวร์เดียว โดยมีการทดสอบซอฟต์แวร์แบบอัตโนมัติเพื่อหาข้อบกพร่องหรือข้อผิดพลาดของซอร์สโค้ด และเมื่อทดสอบผ่านจะนำส่งซอฟต์แวร์ทั้งหมดขึ้นไปยังเครื่องแม่ข่ายแบบอัตโนมัติโดยกิตแล็บ (Gitlab) ทำหน้าที่หลักในการจัดการเก็บซอร์สโค้ดของแต่ละระบบงาน และจัดการกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ ซึ่งแบ่งออกได้เป็น 2 ส่วนดังนี้

#### 2.1.1 การรวบรวมซอฟต์แวร์ (Continuous Integration)

กระบวนการรวบรวมซอร์สโค้ดของนักพัฒนาระบบเพื่อตรวจสอบข้อผิดพลาดของซอร์สโค้ดโดยทำงานแบบอัตโนมัติตั้งแต่การนำซอร์สโค้ดที่ได้จากหนึ่ง หรือหลายนักพัฒนามารวมในอยู่ในรูปแบบคอนเทนเนอร์ และทำการทดสอบซอร์สโค้ดเพื่อหาข้อผิดพลาดแบบอัตโนมัติ

#### 2.1.2 การนำส่งซอฟต์แวร์แบบอัตโนมัติ (Continuous Delivery หรือ Continuous Deployment)

การนำซอร์สโค้ดที่ได้จากกระบวนการรวบรวมซอฟต์แวร์แบบอัตโนมัติที่ผ่านการทดสอบผ่านเรียบร้อยแล้วนำส่งขึ้นไปยังเครื่องแม่ข่ายแบบอัตโนมัติ (Continuous Deployment)

จากการศึกษาเกี่ยวกับกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ (Continuous Integration และ Continuous Deployment) สามารถสรุปได้ว่ากิตแล็บ (Gitlab) มีหน้าที่หลักได้แก่ จัดเก็บซอร์สโค้ดของแต่ละระบบงาน และจัดการกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ ซึ่งมีเครื่องมือในการจัดการที่มีความสัมพันธ์กัน และทำงานเป็นขั้นตอนตามลำดับแบบอัตโนมัติ โดยการรวบรวมซอฟต์แวร์เป็นกระบวนการนำซอร์สโค้ดที่อยู่บนกิตแล็บมาสร้างเป็นคอนเทนเนอร์ จากนั้นนำไปทดสอบเมื่อผลการทดสอบผ่านจะเข้าสู่กระบวนการนำส่งซอฟต์แวร์ไปยังเครื่องแม่ข่าย

### 2.2 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้องเกี่ยวกับกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ ซึ่งมีแนวคิดและวิธีการนำมาใช้ได้แก่ การพัฒนาซอฟต์แวร์แบบอัตโนมัติและการบูรณาการโดยใช้ DevOps [2] การทดสอบกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติโดยใช้ Jenkins Ansible [3] กระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติสำหรับการจัดการโครงการซอฟต์แวร์แบบ Agile [4] กระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติโดยใช้ Asylo ซึ่งเป็นการจำลองสภาพแวดล้อมแบบเฟรมเวิร์กที่มีความน่าเชื่อถือ [5] โครงสร้างพื้นฐานแบบอัตโนมัติบนคลาวด์ โดยการใช้การรวบรวมและนำส่งแบบอัตโนมัติโดยใช้ดีออกเกอร์พร้อมการรักษาความปลอดภัยคอนเทนเนอร์ [6] การเปรียบเทียบเครื่องมือการรวบรวมและนำส่งอัตโนมัติ แบบบูรณาการด้วยแพลตฟอร์มคลาวด์ [7] การระบบรักษาความปลอดภัยของดีออกเกอร์ คอนเทนเนอร์ในการประมวลผลแบบคลาวด์ [8] และ กระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติในรูปแบบไปป์ไลน์ [9]

### 3. วิธีการดำเนินงาน

การพัฒนาเว็บแอปพลิเคชันรูปแบบเดิมใช้กระบวนการรวบรวมและนำส่งซอฟต์แวร์ที่มีความซับซ้อน หลายขั้นตอน ซึ่งอาจทำให้ใช้ระยะเวลา นาน และเกิดข้อผิดพลาดในแต่ละกระบวนการ งานวิจัยนี้จึงได้ทำการสร้างกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บที่มีกระบวนการทั้งหมด 3 ขั้นตอนและมีการดำเนินงานดังนี้

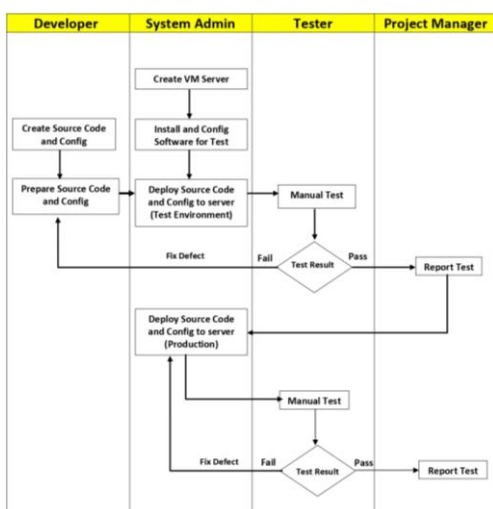
#### 3.1 การศึกษาและรวบรวมข้อมูล

จากการศึกษาและรวบรวมข้อมูลเกี่ยวกับกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ พบว่า กระบวนการรวบรวมและนำส่งซอฟต์แวร์รูปแบบเดิมมีขั้นตอนในแต่ละกระบวนการที่ซ้ำซ้อน ทำให้ไม่สามารถเข้าถึงผู้ใช้งานได้อย่างรวดเร็ว จึงได้ทดสอบกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ

#### 3.2 การออกแบบกรอบแนวคิดการวิจัย

##### 3.2.1 การวิเคราะห์กระบวนการรวบรวมและนำส่งซอฟต์แวร์รูปแบบเดิม

จากการวิเคราะห์กระบวนการรวบรวมและนำส่งซอฟต์แวร์รูปแบบเดิมโดยมีขั้นตอนทั้งหมด 5 ขั้นตอน และมีผู้จัดการแต่ละกระบวนการที่แตกต่างกัน ซึ่งอาจทำให้ใช้ระยะเวลาในแต่ละกระบวนการที่นาน และเกิดข้อผิดพลาดในการทำงาน โดยกระบวนการทำงานรูปแบบเดิมมีขั้นตอนดังภาพที่ 1



ภาพที่ 1: การวิเคราะห์กระบวนการรวบรวมและติดตั้งรูปแบบเดิม

จากภาพที่ 1 สามารถอธิบายการกระบวนการรวบรวมและติดตั้งรูปแบบเดิมโดยมีขั้นตอนกระบวนการดังนี้

1) นักพัฒนา (Developer) จะทำการจัดเตรียมซอร์สโค้ดและกำหนดคุณสมบัติ (Config) ของโปรแกรม เพื่อนำส่งให้ผู้ดูแลระบบ (System Administrator) ทำการติดตั้งซอฟต์แวร์ดังกล่าวบนเครื่องที่ใช้ในการทดสอบ

2) ผู้ดูแลระบบ (System Administrator) ทำการสร้างเครื่องแม่ข่ายที่ใช้ในการทดสอบ โดยติดตั้งและกำหนดคุณสมบัติของซอฟต์แวร์ที่จำเป็นต่อการติดตั้งเว็บแอปพลิเคชัน จากนั้นนำซอร์สโค้ดที่ได้จากนักพัฒนามาติดตั้งที่เครื่องแม่ข่ายสำหรับทำทดสอบในขั้นตอนต่อไป

3) นักทดสอบซอฟต์แวร์ (Tester) ทำการทดสอบซอฟต์แวร์ด้วยตนเอง

กรณีที่พบข้อบกพร่อง นักทดสอบซอฟต์แวร์จะรายงานข้อบกพร่องไปยังนักพัฒนา เพื่อทำการแก้ไขข้อบกพร่องดังกล่าว เมื่อนักพัฒนาแก้ไขเสร็จเรียบร้อยแล้ว จะทำการจัดเตรียมซอร์สโค้ด และกำหนดคุณสมบัติอีกครั้ง เพื่อนำส่งต่อให้ผู้ดูแลระบบทำการติดตั้งซอฟต์แวร์ หลังจากนั้นนักทดสอบซอฟต์แวร์จะทำการทดสอบซอฟต์แวร์อีกครั้ง

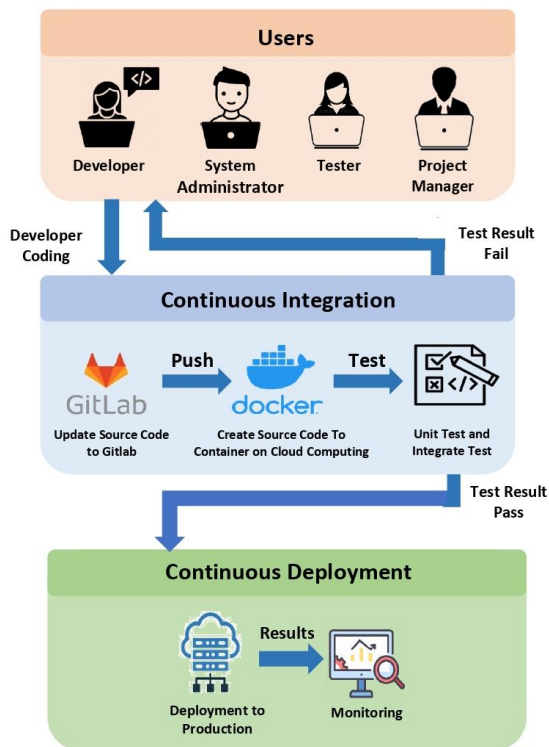
กรณีผลการทดสอบผ่าน นักทดสอบซอฟต์แวร์จะส่งเอกสารผลการทดสอบไปยังผู้จัดการโครงการเพื่อให้ทราบถึงผลการทดสอบ

4) ผู้จัดการโครงการ (Project Manager) เมื่อผู้จัดการโครงการได้รับผลการทดสอบจากนักทดสอบซอฟต์แวร์ จะพิจารณาผลการทดสอบและทำการวางแผนเพื่อติดตั้งซอฟต์แวร์บนเครื่องแม่ข่ายที่ใช้งานจริง

5) ผู้ดูแลต้องทำการสร้างและกำหนดคุณสมบัติบนเครื่องแม่ข่ายที่ใช้งานจริง เมื่อผู้ดูแลระบบติดตั้งซอฟต์แวร์เสร็จเรียบร้อยแล้ว นักทดสอบซอฟต์แวร์จะทำการทดสอบอีกครั้ง

##### 3.2.2 การออกแบบกรอบแนวคิดการวิจัย

จากการวิเคราะห์การทำงานรูปแบบเดิมจึงทำการออกแบบกรอบแนวคิดการวิจัยของกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ เพื่อปรับปรุงกระบวนการให้มีประสิทธิภาพ ซึ่งแสดงได้ดังภาพที่ 2



**ภาพที่ 2:** การออกแบบกรอบแนวคิดกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ

จากภาพที่ 2 เป็นอธิบายการออกแบบกรอบแนวคิดกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ ซึ่งมีรายละเอียดดังนี้

- 1) นักพัฒนา (Developer) จะทำการจัดเตรียมซอร์สโค้ดเพื่อนำขึ้นไปไว้ยังกิตแล็บเซิร์ฟเวอร์
  - 2) หลังจากนำซอร์สโค้ดขึ้นไปไว้ยังกิตแล็บเรียบร้อยแล้ว จะเริ่มทำงานตามกระบวนการที่จัดเตรียมไว้ โดยการนำซอร์สโค้ดมาทำให้อยู่ในรูปแบบคอนเทนเนอร์ และทำการทดสอบแบบอัตโนมัติ
  - 3) เมื่อผ่านขั้นตอนของการทดสอบเรียบร้อยแล้ว ขั้นตอนต่อไปเป็นการนำส่งคอนเทนเนอร์ที่ได้จากขั้นตอนการสร้างและนำไปติดตั้งแบบอัตโนมัติบนเครื่องแม่ข่าย
- กรณีที่ผู้พัฒนามีการเปลี่ยนแปลงแก้ไขซอร์สโค้ดไปยังกิตแล็บเซิร์ฟเวอร์ กระบวนการทดสอบและการติดตั้งแบบอัตโนมัติจะดำเนินตามกระบวนการเดิมอีกครั้ง

### 3.3 การออกแบบและพัฒนากระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ

การออกแบบและพัฒนากระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ สามารถแบ่งการออกแบบและพัฒนาได้เป็น 2 ขั้นตอนดังนี้

3.3.1 การเตรียมการติดตั้งและบริหารจัดการซอฟต์แวร์  
 ขั้นตอนนี้จะทำการเตรียมเครื่องคลาวด์คอมพิวติ้งสำหรับการติดตั้งและบริหารจัดการซอฟต์แวร์ โดยจะติดตั้งและคอนฟิกซอฟต์แวร์ดีออกเกอร์สำหรับการสร้างคอนเทนเนอร์ หลังจากนั้นทำการคอนฟิกกิตแล็บ

3.3.2 การสร้างกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ

การสร้างกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ ซึ่งมีรายละเอียดดังนี้

- 1) ขั้นตอนการสร้างซอฟต์แวร์ให้อยู่ในรูปแบบของคอนเทนเนอร์ โดยการนำซอร์สโค้ดที่ได้จากหนึ่งหรือหลายนักพัฒนาขึ้นไปวางไว้บนกิตแล็บ จากนั้นจะเริ่มกระบวนการรวบรวมซอร์สโค้ดที่อยู่บนกิตแล็บมาให้อยู่ในเป็นคอนเทนเนอร์
- 2) ขั้นตอนการทดสอบซอฟต์แวร์ โดยทำการทดสอบซอฟต์แวร์แบบทีละส่วนของโปรแกรม ในการค้นหาข้อผิดพลาด ซึ่งอาจเป็นรูปแบบหรือไวยากรณ์ภาษาที่มีผิดพลาด
- 3) ขั้นตอนการนำส่งซอฟต์แวร์แบบอัตโนมัติ โดยจะกระทำหลังจากทดสอบกระบวนการรวบรวมแบบอัตโนมัติผ่านเรียบร้อยแล้ว จะทำการนำส่งซอฟต์แวร์ที่ทดสอบนำขึ้นไปยังเครื่องแม่ข่ายโดยอัตโนมัติ

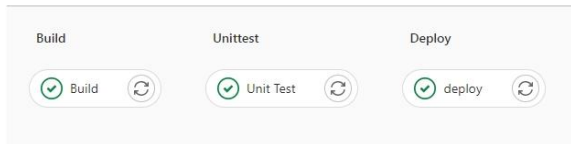
### 3.4 การประเมินประสิทธิภาพกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ

หลังจากทำการทดสอบกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บเรียบร้อยแล้ว จะทำการประเมินประสิทธิภาพกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ โดยการเก็บระยะเวลาการดำเนินงาน และผลการทดสอบของแต่ละขั้นตอน



#### 4. ผลการดำเนินงานวิจัย

ผลการดำเนินงานวิจัย ซึ่งจะทำการทดสอบกระบวนการรวบรวม และนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ จะแสดงผลของกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ โดยสามารถเข้าไปตรวจสอบหรือดูสถานะของแต่ละกระบวนการได้ ทุกขั้นตอนบนกิตแล็บได้ดังภาพที่ 3



ภาพที่ 3: หน้าจอแสดงกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ

จากภาพที่ 3 เป็นการแสดงกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ ซึ่งมีรายละเอียดดังนี้

##### 4.1 การสร้างคอนเทนเนอร์ (Build)

ขั้นตอนการสร้างคอนเทนเนอร์ โดยการรวบรวมซอร์สโค้ดทั้งหมด ซึ่งนำมาสร้างให้อยู่ในรูปแบบของคอนเทนเนอร์ก่อนนำไปทดสอบ โดยสามารถดูผลระยะเวลาการสร้างคอนเทนเนอร์ได้ดังภาพที่ 4

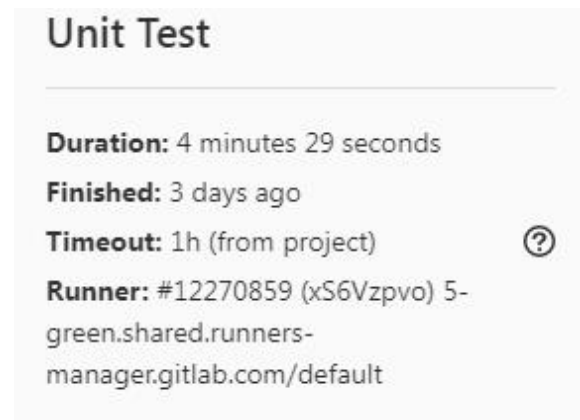


ภาพที่ 4: รายละเอียดของการสร้างคอนเทนเนอร์

##### 4.2 การทดสอบซอร์สโค้ด (Test)

ขั้นตอนการทดสอบซอร์สโค้ดเป็นการเขียนไปป์ไลน์ในการทดสอบให้ทำการทดสอบแต่ละโมดูลของโปรแกรมโดยมีเงื่อนไข หากพบข้อผิดพลาดจะไม่สามารถดำเนินการ

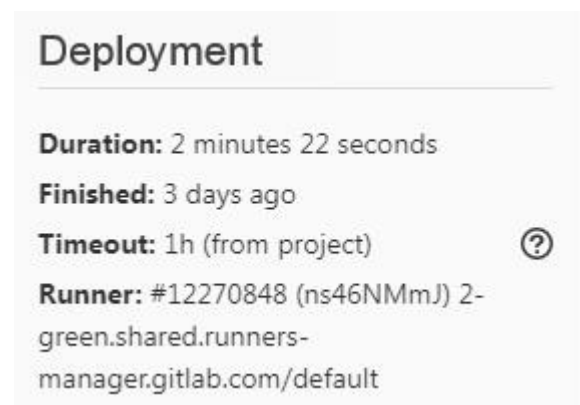
นำส่งซอฟต์แวร์แบบอัตโนมัติได้จะต้องทำการแก้ไขซอร์สโค้ดให้เรียบร้อย และนำเข้ากระบวนการใหม่ ซึ่งสามารถดูผลระยะเวลาการทดสอบได้ดังภาพที่ 5



ภาพที่ 5: รายละเอียดของการทดสอบซอร์สโค้ด

##### 4.3 การนำส่งซอฟต์แวร์แบบอัตโนมัติ (Deployment)

ขั้นตอนการนำส่งซอฟต์แวร์แบบอัตโนมัติจะเกิดขึ้นหลังจากดำเนินการสร้างและทดสอบแล้วไม่มีข้อผิดพลาดเท่านั้น จึงจะทำการนำส่งซอฟต์แวร์แบบอัตโนมัติเครื่องที่ใช้งานจริง ซึ่งสามารถดูผลเวลานำส่งซอฟต์แวร์แบบอัตโนมัติได้ดังภาพที่ 6



ภาพที่ 6: รายละเอียดของการนำส่งซอฟต์แวร์แบบอัตโนมัติ

เมื่อทดสอบกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บเรียบร้อยแล้ว พบว่า กระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติช่วยลดขั้นตอนในการรวบรวมและนำส่งซอฟต์แวร์จากเดิมที่มีหลายขั้นตอน และระยะเวลาในแต่ละกระบวนการ ซึ่งสามารถแสดงได้ดังภาพที่ 7

Status	Job	Stage	Name	Duration
	#2199520662 ▼ main → 4c2776f8	deploy	deploy	⌚ 00:02:27 📅 2 weeks ago
	#2199520661 ▼ main → 4c2776f8	unittest	Unit Test	⌚ 00:04:29 📅 2 weeks ago
	#2199520658 ▼ main → 4c2776f8	build	Build	⌚ 00:08:18 📅 2 weeks ago

ภาพที่ 7: ผลการทดสอบกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ

## 5. สรุปผลการวิจัย

งานวิจัยนี้นำเสนอการสร้างกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ ซึ่งประกอบด้วย 3 ขั้นตอนหลัก ได้แก่ การรวบรวมซอร์สโค้ดมาสร้างให้อยู่ในรูปแบบคอนเทนเนอร์โดยใช้ด็อกเกอร์ การทดสอบซอฟต์แวร์อัตโนมัติ และการนำส่งซอฟต์แวร์แบบอัตโนมัติ จากการทดสอบกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บ พบว่า กระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ มีขั้นตอนการทำงานเพียง 3 ขั้นตอน เมื่อเทียบกับกระบวนการรวบรวมและนำส่งซอฟต์แวร์รูปแบบเดิมที่มีขั้นตอนการทำงาน 5 ขั้นตอน และจากผลการทดสอบกระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติบนคลาวด์คอมพิวติ้งโดยใช้กิตแล็บสรุปได้ว่า กระบวนการรวบรวมและนำส่งซอฟต์แวร์แบบอัตโนมัติ สามารถลดขั้นตอนการทำงานระยะเวลาการทำงานแต่ละขั้นตอน ช่วยลดการเกิดข้อผิดพลาดในการทำงานให้น้อยลง และส่งผลให้อัปพลิเคชันมีประสิทธิภาพมากขึ้น

## เอกสารอ้างอิง

- [1] วรุตม์ พลอยสวยงาม. “เทคโนโลยีเว็บแอปพลิเคชัน (Web Application Technology)” Available Online at <http://comed.bsru.ac.th/>, สืบค้นเมื่อ 16 มีนาคม 2565.
- [2] Agarwal, A., Gupta, S, and Choudhury, T. “Continuous and Integrated Software Development using DevOps,” *In Proceeding of 2008 International Conference on Advances in Computing and Communication Engineering*. Paris, pp. 290-293, 2008.
- [3] Mysari, S, and Bejgam, V. (2020). “Continuous Integration and Continuous Deployment Pipeline Automation Using Jenkins Ansible” *In Proceeding of 2020 International Conference on Emerging Trends in Information Technology and Engineering*, pp. 1-4, 2020.
- [4] Perera, I., “Continuous Integration and Continuous Delivery Pipeline Automation for Agile Software Project Management” *In Proceeding of 2018 Moratuwa Engineering Research Conference*. pp. 156-161, 2018.
- [5] Mahboob, J, and Coffman J., “A Kubernetes CI/CD Pipeline with Asylo as a Trusted Execution Environment Abstraction Framework” *In Proceeding of 2021 Computing and Communication Workshop and Conference (CCWC)*, pp. 0529-0535, 2021.
- [6] Garg, S, and Garg, S., “Automated Cloud Infrastructure, Continuous Integration and Continuous Delivery Using Docker with Robust Container Security” *In Proceeding of 2019 Multimedia Information Processing and Retrieval (MIPR)*. pp. 467-470, 2019.
- [7] Singh, C, Gaba, N, and Kaur, M. , “Comparison of Different CI/CD Tools Integrated with Cloud Platform” *In Proceeding of 2019 9<sup>th</sup> International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. pp. 7-12, 2019.
- [8] Yasar, H., “Experiment : Sizing Exposed Credentials in GitHub Public Repositories for CI/CD” *In Proceeding of 2018 Development Conference*. pp. 143, 2018.
- [9] Cepuc, A, Botez, R, and Craciun, O., “Implementation of a Continuous Integration and Deployment Pipeline for Containerized Applications in Amazon Web Services Using Jenkins, Ansible and Kubernetes” *In Proceeding of 2020 19<sup>th</sup> RoEduNet Conference:Networking*, 2020.

# ระบบวินิจฉัยโรคกล้วยไม้ด้วยการเรียนรู้เชิงลึกผ่านแอปพลิเคชันบนโทรศัพท์มือถือ

## Diagnosis System of Orchid Disease using Deep Learning on Mobile Application

ณัฐวดี หงษ์บุญมี (Nattavadee Hongboonmee)<sup>1</sup> และกวินทิพย์ ทุมจรรย์ (Kawinthip Thumchon)<sup>2</sup>  
 ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนครสวรรค์  
<sup>1</sup>nattavadeeho@nu.ac.th, <sup>2</sup>kawinthipt60@nu.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ คือ (1) เพื่อศึกษาการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึกกับการวินิจฉัยโรคกล้วยไม้จากภาพใบและ (2) เพื่อพัฒนาระบบวินิจฉัยโรคกล้วยไม้จากภาพใบ โดยใช้งานผ่านแอปพลิเคชันบนโทรศัพท์มือถือระบบปฏิบัติการแอนดรอยด์ ในส่วนการวินิจฉัยโรคใช้เทคนิคการเรียนรู้เชิงลึกด้วยโครงข่ายประสาทเทียมคอนโวลูชันในการสร้างแบบจำลองจำแนกโรค ได้แก่ โรคใบจุด โรคใบปื้นเหลือง และใบปกติ จากนั้นนำแบบจำลองมาพัฒนาระบบ โดยใช้เครื่องมือ ได้แก่ โปรแกรมแอนดรอยด์สตูดิโอ ขั้นตอนสุดท้ายคือการตรวจสอบความแม่นยำของแอปพลิเคชันใช้วิธีการตรวจสอบโดยทดสอบทีละตัว (Leave-One-Out Cross Validation) ผลการทดลองพบว่า (1) แบบจำลองจำแนกภาพจากเทคนิคการเรียนรู้เชิงลึก มีค่าความถูกต้องสูงเท่ากับ 98.10% (2) ผลการวิเคราะห์ความแม่นยำของแอปพลิเคชัน มีความแม่นยำเท่ากับ 81.11% การทดลองทั้งหมดแสดงให้เห็นว่าระบบที่พัฒนาขึ้นนี้มีประสิทธิภาพและง่ายต่อการใช้งาน สามารถช่วยเกษตรกรวินิจฉัยโรคกล้วยไม้ได้อย่างเหมาะสม

**คำสำคัญ:** โรคกล้วยไม้ การเรียนรู้เชิงลึก แอปพลิเคชันบนโทรศัพท์มือถือ

### Abstract

The objectives of this research are (1) to study the deep learning technique for orchid disease diagnosis from leaf and (2) to develop an orchid disease diagnosis system from leaf on mobile application android. The diseases diagnosis process classified

diseases that have spot leaf, yellow spot leaf and healthy leaf by deep learning technique, by using convolutional neural network to make a classification model. Then the model was developed its system using a tool, i.e. Android Studio. In the final stage, to verify the accuracy of the application, by using Leave-One-Out Cross Validation. The study results found that (1) the model from deep learning technique with the accuracy of 98.10%, (2) the accuracy of the application was 81.11%. From all experimental results show that the proposed system is able to provide accurate diagnosis results efficiently and easy to diagnosis the orchid disease. The proposed system is very useful in assisting agriculturists to accomplish the task of diagnosing orchid disease efficiently.

**Keywords:** Orchid Disease, Deep Learning, Mobile Application.

### 1. บทนำ

กล้วยไม้มีชื่อวิทยาศาสตร์และชื่อสามัญว่า Orchid [1] ปัจจุบันกล้วยไม้เป็นพืชเศรษฐกิจที่มีความสำคัญทางด้านการส่งออกของประเทศไทย ข้อมูลจากสำนักส่งเสริมการค้าสินค้าเกษตรและอุตสาหกรรม กรมส่งเสริมการค้าระหว่างประเทศ ประจำปี พ.ศ. 2562 [2] ระบุว่าประเทศไทยเป็นผู้ส่งออกดอกกล้วยไม้อันดับหนึ่งของโลก มีมูลค่ารวม 213 ล้านบาท เนื่องจากประเทศไทยตั้งอยู่ในแหล่งภูมิศาสตร์ที่มีสภาพเหมาะสมในการผลิตกล้วยไม้เขตร้อน มีการเพาะเลี้ยงอย่างครบวงจร ตั้งแต่การผสมเกสรเพาะเลี้ยงเนื้อเยื่อ เลี้ยงต้นกล้วยไม้จนกระทั่งให้ดอก ตัดดอกบรรจุหีบห่อและส่งออก

กล้วยไม้เป็นพืชที่ต้องการการดูแลเอาใจใส่อย่างใกล้ชิดอย่างสม่ำเสมอ เพราะถ้าไม่ดูแลจะทำให้มีเชื้อราและเชื้อแบคทีเรียที่เป็นสาเหตุของการเกิดโรครอบคอบในกล้วยไม้ ซึ่งส่วนใหญ่จะแสดงอาการให้เห็นอย่างชัดเจนทางใบ [1] เช่น โรคใบปื้นเหลือง โรคใบจุด โรคยอดเน่า เป็นต้น โรคเหล่านี้จะส่งผลกระทบต่อเจริญเติบโตของกล้วยไม้อย่างมากทำให้ปริมาณผลผลิตลดลงและได้ผลผลิตคุณภาพต่ำทำให้เกษตรกรมีรายได้น้อยลงและยังส่งผลกระทบต่อภาพรวมการส่งออกสินค้าเกษตรของประเทศอีกด้วย ถึงแม้จะมีหน่วยงานที่เกี่ยวข้องคอยช่วยเหลือโดยการส่งผู้ที่มีความเชี่ยวชาญทางด้านกล้วยไม้ไปยังพื้นที่ในท้องถิ่นต่าง ๆ เพื่อให้คำแนะนำแก่ชาวสวนกล้วยไม้ แต่ก็ยังมีข้อจำกัดเนื่องจากจำนวนผู้เชี่ยวชาญยังมีน้อยและอาการของโรคนั้นก็มีหลายลักษณะ นอกจากนี้ลักษณะของโรคกล้วยไม้ยังมีความใกล้เคียงกันมากทำให้การให้คำแนะนำแก่ชาวสวนกล้วยไม้อาจจะยังไม่ทั่วถึงและไม่มีประสิทธิภาพเพียงพอต่อการแก้ไขปัญหาต่าง ๆ จากโรคกล้วยไม้ได้ทันทั่วทั้ง

จากปัญหาข้างต้น คณะผู้วิจัยจึงมีแนวคิดที่จะนำเทคโนโลยีคอมพิวเตอร์เข้ามาช่วยในการวินิจฉัยโรคกล้วยไม้ที่มีประสิทธิภาพ โดยนำเทคโนโลยีการเรียนรู้ของเครื่องซึ่งถูกพัฒนาไปจนทำให้คอมพิวเตอร์สามารถจำแนกภาพถ่ายได้ด้วย การเรียนรู้จากภาพเรียกว่าการเรียนรู้เชิงลึก (Deep Learning) ซึ่งเป็นวิธีการที่ได้รับการยอมรับด้านความแม่นยำและเป็นวิธีการจำแนกภาพที่ทันสมัยที่สุดแบบหนึ่งมาประยุกต์ใช้ในการสร้างแบบจำลองวินิจฉัยโรคกล้วยไม้และนำแบบจำลองจำแนกภาพที่ได้มาพัฒนาเป็นแอปพลิเคชันบนโทรศัพท์มือถือระบบปฏิบัติการแอนดรอยด์เพื่อให้ง่ายและสะดวกต่อการใช้งาน แอปพลิเคชันที่พัฒนาขึ้นนี้จะมีประโยชน์กับเกษตรกรชาวสวนกล้วยไม้และผู้ที่สนใจต้องการศึกษาค้นคว้าเกี่ยวกับโรคกล้วยไม้ สามารถนำแอปพลิเคชันนี้ไปใช้เป็นเครื่องมือช่วยวินิจฉัยโรคกล้วยไม้ได้ ทำให้เกษตรกรป้องกันและดูแลกล้วยไม้ได้อย่างมีประสิทธิภาพมากขึ้น

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 โรคกล้วยไม้

โรคกล้วยไม้ [1] หมายถึง การเปลี่ยนแปลงของกล้วยไม้ที่ผิดปกติไปจากปกติ โดยมีเชื้อโรคหรือปัจจัยอื่นเป็น

สาเหตุมีผลทำให้การเจริญเติบโตลดลงหรือเกิดการแพร่ระบาดของโรคทำให้กล้วยไม้ตายภายในระยะเวลารวดเร็ว ซึ่งอาการของโรคส่วนใหญ่จะเด่นชัด โดยดูได้จากลักษณะของใบกล้วยไม้ ดังนั้นการศึกษานี้จึงสนใจศึกษาโรคที่เกิดกับใบกล้วยไม้ ซึ่งพบได้บ่อยจำนวน 2 โรค ได้แก่ (1) โรคใบจุด (Spot Leaf Disease) ลักษณะอาการ คือ ใบมีจุดกลมสีน้ำตาลเข้มหรือดำ และ (2) โรคใบปื้นเหลือง (Yellow Spot Leaf Disease) ลักษณะอาการ คือ มีจุดกลมสีเหลืองที่ใบหรือปื้นสีเหลืองตามแนวยาวของใบ

### 2.2 การเรียนรู้เชิงลึก

การเรียนรู้เชิงลึก (Deep Learning) [3] เป็นเทคนิคในกลุ่มโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) ซึ่งอยู่บนพื้นฐานของระเบียบวิธีที่จะสร้างรูปแบบแนวคิดระดับสูงให้กับข้อมูลโดยใช้การประมวลผลบนโครงสร้างซับซ้อน ปัจจุบันมีสถาปัตยกรรม (Architecture) ที่สามารถนำมาประยุกต์ใช้งานได้หลากหลาย เช่น โครงข่ายประสาทเทียมคอนโวลูชันที่สามารถทำได้ทั้งสกัดคุณลักษณะ (Feature Extraction) และจำแนกประเภท (Classification) ซึ่งถือได้ว่าเป็นจุดเด่นของโครงข่ายนี้

### 2.3 โครงข่ายประสาทเทียมคอนโวลูชัน

โครงข่ายประสาทเทียมคอนโวลูชัน (Convolution Neural Network: CNN) [4] มีกระบวนการสกัดคุณลักษณะของภาพโดยใช้ Convolutional Layer ผ่านทางฟิลเตอร์ในแต่ละ Convolutional Layer ผู้ใช้ไม่จำเป็นต้องกำหนดรูปแบบในการสอนให้กับแบบจำลอง เพียงแค่ทำการเตรียมข้อมูลภาพตัวอย่างที่ต้องการใช้งานเอาไว้แล้วนำภาพเหล่านั้นป้อนเข้ากระบวนการ CNN จะทำเรียนรู้โดยอัตโนมัติและหากผู้ใช้ต้องการทำนายรูปภาพก็เพียงแค่ป้อนรูปภาพที่ต้องการทำนายเข้าไป CNN จะทำการเรียนรู้ เพื่อเปรียบเทียบรูปภาพที่ต้องการทำนายกับข้อมูลรูปภาพที่มีอยู่เพื่อแสดงผลลัพธ์ออกมา

### 2.4 งานวิจัยที่เกี่ยวข้อง

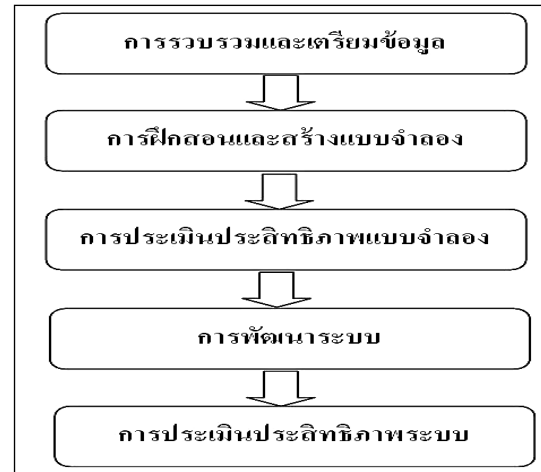
มีหลายงานวิจัยที่นำเทคนิคการเรียนรู้เชิงลึกมาประยุกต์ใช้เพื่อการวินิจฉัยโรคพืช เช่น งานวิจัย [5] ศึกษาการประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมคอนโวลูชัน

ชั้นเพื่อตรวจและวินิจฉัยโรคพืช ผลการทดลองพบว่าแบบจำลองได้ค่าประสิทธิภาพที่สูงถึง 99.53% งานวิจัย [6] ศึกษาการใช้ข้อมูลใบกล้วยที่มีลักษณะเป็นโรคและใบปกติ โดยใช้โครงข่ายประสาทเทียมคอนโวลูชัน ผลการทดลองได้ค่าความถูกต้องของการจำแนกภาพเท่ากับ 99.72% งานวิจัย [7] ศึกษาการใช้เทคนิคโครงข่ายประสาทเทียมเชิงลึกคอนโวลูชัน (D-CNN) สำหรับการจำแนกโรคข้าว 10 ประเภท พบว่าแบบจำลอง D-CNN มีค่าความแม่นยำ 95.48% งานวิจัย [8] นำเสนอการตรวจจับโรคจากใบแอปเปิ้ลจำนวน 4 ประเภท โดยใช้โครงข่ายประสาทเทียมคอนโวลูชัน ได้ค่าความถูกต้อง 97.62% งานวิจัย [9] ศึกษาการตรวจจับโรคของผลฝรั่งโดยใช้ข้อมูลจากลักษณะใบ ประยุกต์ใช้โครงข่ายประสาทเทียมคอนโวลูชัน ผลการทดลองได้ค่าความแม่นยำเท่ากับ 98.74%

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่าการเรียนรู้เชิงลึกด้วยโครงข่ายประสาทเทียมคอนโวลูชัน (CNN) เป็นวิธีการที่ได้รับความนิยมมากในการนำมาใช้ในการจำแนกรูปร่างภาพ เป็นแบบจำลองที่ได้รับการยอมรับด้านความแม่นยำที่สูง และเป็นวิธีการจำแนกภาพที่ทันสมัยที่สุดแบบหนึ่ง แต่ยังพบข้อจำกัดจากงานวิจัยก่อนหน้านี้ คือแบบจำลองวินิจฉัยโรคที่พัฒนาขึ้น ไม่สะดวกต่อการใช้งานของเกษตรกร งานวิจัยนี้จึงนำแบบจำลองมาประยุกต์ใช้ร่วมกับแอปพลิเคชันบนโทรศัพท์มือถือระบบปฏิบัติการแอนดรอยด์ เพื่อให้เกษตรกรหรือนักคิดทั่วไปที่ไม่มีความรู้เรื่องโรคกล้วยไม้สามารถใช้ได้จริง

### 3. วิธีการดำเนินการวิจัย

การพัฒนากระบวนการวินิจฉัยโรคกล้วยไม้เป็นการประยุกต์การเรียนรู้เชิงลึกมาใช้ในการจำแนกภาพใบกล้วยไม้ โดยใช้โครงข่ายประสาทเทียมคอนโวลูชันสำหรับวิเคราะห์ภาพ เพื่อให้สามารถจำแนกภาพได้อย่างมีประสิทธิภาพ และสามารถนำแบบจำลองไปใช้งานบนโทรศัพท์มือถือระบบปฏิบัติการแอนดรอยด์ได้ โดยขั้นตอนการดำเนินการวิจัย แสดงดังภาพที่ 1

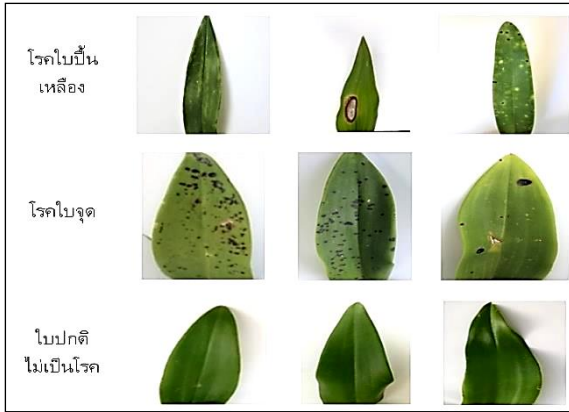


ภาพที่ 1: ขั้นตอนการดำเนินการวิจัย

#### 3.1 การรวบรวมและเตรียมข้อมูล

การรวบรวมข้อมูลรูปภาพกลุ่มตัวอย่างได้มาจากสวนกล้วยไม้ในจังหวัดพิษณุโลก การเก็บข้อมูลใช้การเด็ดใบที่เป็นโรคจากต้นกล้วยไม้นำมาถ่ายภาพด้วยกล้องถ่ายรูปบนโทรศัพท์มือถือ โดยห่างจากกล้องประมาณ 20-30 เซนติเมตร ใช้พื้นหลังเป็นกระดาษขาว รูปภาพที่ถ่ายเป็นภาพด้านหน้าใบกล้วยไม้ ขนาดของรูปภาพปรับขนาดให้เท่ากันทั้งหมด คือ 224x224 พิกเซล

ข้อมูลรูปภาพโรคกล้วยไม้ทั้ง 3 ประเภท ถูกจัดเก็บไว้ในแต่ละโฟลเดอร์ตามชื่อประเภทโรคกล้วยไม้ ได้แก่ (1) Spot Leaf (โรคใบจุด) (2) Yellow Spot Leaf (โรคใบปื้นเหลือง) และ (3) Healthy Leaf (ใบปกติไม่เป็นโรค) จากนั้นได้นำภาพกลุ่มตัวอย่างไปให้ผู้เชี่ยวชาญตรวจสอบอีกครั้ง เพื่อตรวจสอบความถูกต้องของภาพให้ตรงกับประเภทโรคกล้วยไม้ จำนวนรูปภาพที่ผ่านการตรวจสอบมีจำนวน 690 ภาพ ดังภาพที่ 2 จากนั้นนำมาแบ่งด้วยวิธีการสุ่มเพื่อใช้เป็นชุดข้อมูลฝึกสอน (Train Data Set) ชุดข้อมูลทดสอบ (Test Data Set) และชุดข้อมูลทดสอบแอปพลิเคชัน (Validate Data Set) จำนวน 510, 90 และ 90 ภาพ ตามลำดับ ดังตารางที่ 1



ภาพที่ 2: ตัวอย่างชุดข้อมูลรูปภาพที่นำมาใช้ในการทดลอง

ตารางที่ 1: จำนวนชุดข้อมูลรูปภาพโรคกล้วยไม้ในแต่ละประเภท

ประเภทของโรคกล้วยไม้	จำนวนรูปภาพ		
	Train	Test	Validate
โรคใบจุด	170	30	30
โรคใบเป็นเหลือง	170	30	30
ใบปกติ	170	30	30
รวม	510	90	90

### 3.2 การฝึกสอนและสร้างแบบจำลอง

งานวิจัยนี้สร้างแบบจำลองด้วยโครงข่ายประสาทเทียมคอนโวลูชัน (CNN) ฝึกสอนและทดสอบชุดข้อมูลด้วย TensorFlow ซึ่งเป็นไลบรารีสำหรับการเรียนรู้เชิงลึกผ่านภาษา Python กำหนดพารามิเตอร์ที่ใช้ในการทดลองคือขนาดรูปภาพนำเข้าเท่ากับ 224x224 พิกเซล อัตราการเรียนรู้ (Learning Rate) 0.005

```

Administrator Command Prompt
INFO:tensorflow:2020-10-12 16:12:19.571165: Step 920: Validation accuracy = 98.0% (N=100)
INFO:tensorflow:2020-10-12 16:12:20.495155: Step 930: Train accuracy = 100.0%
INFO:tensorflow:2020-10-12 16:12:20.495155: Step 930: Cross entropy = 0.000456
INFO:tensorflow:2020-10-12 16:12:20.592160: Step 930: Validation accuracy = 98.0% (N=100)
INFO:tensorflow:2020-10-12 16:12:21.588161: Step 940: Train accuracy = 100.0%
INFO:tensorflow:2020-10-12 16:12:21.588161: Step 940: Cross entropy = 0.000845
INFO:tensorflow:2020-10-12 16:12:21.612162: Step 940: Validation accuracy = 97.0% (N=100)
INFO:tensorflow:2020-10-12 16:12:22.447156: Step 950: Train accuracy = 100.0%
INFO:tensorflow:2020-10-12 16:12:22.447156: Step 950: Cross entropy = 0.001047
INFO:tensorflow:2020-10-12 16:12:22.530163: Step 950: Validation accuracy = 100.0% (N=100)
INFO:tensorflow:2020-10-12 16:12:23.364176: Step 960: Train accuracy = 100.0%
INFO:tensorflow:2020-10-12 16:12:23.364176: Step 960: Cross entropy = 0.000923
INFO:tensorflow:2020-10-12 16:12:23.492157: Step 960: Validation accuracy = 97.0% (N=100)
INFO:tensorflow:2020-10-12 16:12:24.312164: Step 970: Train accuracy = 100.0%
INFO:tensorflow:2020-10-12 16:12:24.312164: Step 970: Cross entropy = 0.003974
INFO:tensorflow:2020-10-12 16:12:24.422163: Step 970: Validation accuracy = 98.0% (N=100)
INFO:tensorflow:2020-10-12 16:12:25.086158: Step 980: Train accuracy = 100.0%
INFO:tensorflow:2020-10-12 16:12:25.086158: Step 980: Cross entropy = 0.001109
INFO:tensorflow:2020-10-12 16:12:25.198162: Step 980: Validation accuracy = 100.0% (N=100)
INFO:tensorflow:2020-10-12 16:12:26.070171: Step 990: Train accuracy = 100.0%
INFO:tensorflow:2020-10-12 16:12:26.070171: Step 990: Cross entropy = 0.000910
INFO:tensorflow:2020-10-12 16:12:26.152164: Step 990: Validation accuracy = 100.0% (N=100)
INFO:tensorflow:2020-10-12 16:12:26.948155: Step 999: Train accuracy = 100.0%
INFO:tensorflow:2020-10-12 16:12:26.948155: Step 999: Cross entropy = 0.001056
INFO:tensorflow:2020-10-12 16:12:27.020165: Step 999: Validation accuracy = 99.0% (N=100)
INFO:tensorflow:Final test accuracy = 98.1% (N=108)
INFO:tensorflow:Froze 2 variables.
Converted 2 variables to const ops.
    
```

ภาพที่ 3: การฝึกสอนชุดข้อมูลกับเทนเซอร์โฟลผ่านภาษาไพทอน

รอบการฝึกสอน (Epoch) จำนวน 500 รอบ เพื่อความแม่นยำในการทดสอบประสิทธิภาพ (ภาพที่ 3) จะสังเกตได้ว่ายังมีการฝึกสอนมากความแม่นยำก็จะสูงตามไปด้วย

### 3.3 การประเมินประสิทธิภาพแบบจำลอง

การประเมินประสิทธิภาพของแบบจำลองจำแนกโรคกล้วยไม้ด้วยค่าความถูกต้อง (Accuracy) [8] เป็นค่าที่ได้จากวิธีการทดสอบเพื่อหาค่าพยากรณ์ความถูกต้องของข้อมูล โดยคิดเป็นค่าร้อยละ (%) ใช้สูตรการคำนวณ ดังสมการที่ 1

$$Accuracy = \left( \frac{TP+TN}{TP+TN+FP+FN} \right) \times 100 \quad (1)$$

โดยที่ True Positive (TP) หมายถึง ค่าที่พยากรณ์ถูกต้องเชิงบวก True Negative (TN) หมายถึง ค่าที่พยากรณ์ถูกต้องเชิงลบ False Positive (FP) หมายถึง ค่าที่พยากรณ์ผิดพลาดเชิงบวก False Negative (FN) หมายถึง ค่าที่พยากรณ์ผิดพลาดเชิงลบ

### 3.4 การพัฒนาระบบ

เมื่อได้แบบจำลองที่เหมาะสมแล้ว จึงได้นำแบบจำลองไปพัฒนาระบบในรูปแบบแอปพลิเคชันบนโทรศัพท์มือถือหรือระบบปฏิบัติการแอนดรอยด์เพื่อให้ง่ายต่อการใช้งาน เครื่องมือในการพัฒนาใช้โปรแกรมแอนดรอยด์สตูดิโอและภาษาจาวา

### 3.5 การประเมินประสิทธิภาพระบบ

การประเมินประสิทธิภาพระบบ ได้มีการนำระบบติดตั้งในรูปแบบแอปพลิเคชันบนโทรศัพท์มือถือใช้วิธีการทดสอบความแม่นยำของแอปพลิเคชันในการจำแนกภาพโรคกล้วยไม้ ทำการสรุปผลการประเมินประสิทธิภาพของระบบด้วยการวัดค่าร้อยละความถูกต้อง

## 4. ผลการดำเนินงาน

### 4.1 ผลการสร้างแบบจำลอง

การสร้างแบบจำลองวินิจฉัยโรคเป็นการนำภาพตัวอย่าง 600 ภาพ มาทำการเรียนรู้กับโครงข่ายประสาทเทียมคอนโวลูชัน โดยทำการแบ่งข้อมูลเป็นข้อมูลฝึกสอนจำนวน 510 ภาพ และข้อมูลทดสอบจำนวน 90 ภาพ ซึ่งสามารถสรุปผลได้ ดังตารางที่ 2

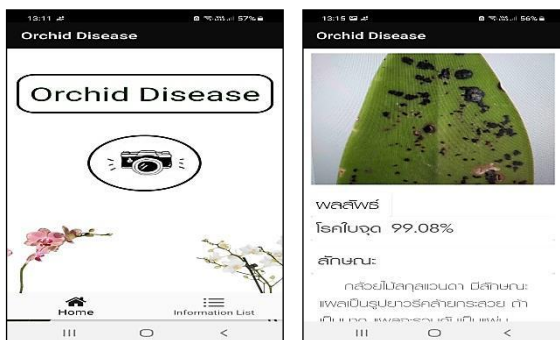
ตารางที่ 2: ผลการประเมินประสิทธิภาพของแบบจำลอง

ประเภทของโรคกล้วยไม้	Accuracy (%)
โรคใบจุด	99.45
โรคใบปื้นเหลือง	98.33
ใบปกติ	96.51
<b>Average</b>	<b>98.10</b>

จากตารางที่ 2 ผลประเมินความถูกต้องของแบบจำลองบนชุดข้อมูลฝึกสอน พบว่าแบบจำลองจำแนกภาพใบจุดได้ค่าความถูกต้องมากที่สุด คือ 99.45% รองลงมา คือ ภาพใบปื้นเหลืองและภาพใบปกติ ซึ่งได้ค่าความถูกต้องเท่ากับ 98.33% และ 96.51% ตามลำดับ ค่าความถูกต้องเฉลี่ยรวมเท่ากับ 98.10% แสดงให้เห็นว่าแบบจำลองจากเทคนิคการเรียนรู้เชิงลึกเป็นแบบจำลองที่มีประสิทธิภาพในการจำแนกภาพได้ค่าความถูกต้องที่สูง รวมทั้งมีค่าความผิดพลาดต่ำ ประมวลผลได้รวดเร็ว มีความเหมาะสมสำหรับนำมาใช้พัฒนาแอปพลิเคชันช่วยวินิจฉัยโรคกล้วยไม้แบบอัตโนมัติต่อไป

4.2 ผลการพัฒนาแอปพลิเคชัน

จากการพัฒนาแอปพลิเคชัน ได้แอปพลิเคชันวินิจฉัยโรคกล้วยไม้จากภาพใบบนโทรศัพท์มือถือระบบปฏิบัติการแอนดรอยด์ เมื่อเข้าสู่แอปพลิเคชันจะพบกับหน้าจอหลัก (ภาพที่ 4) เลือกที่กล้องถ่ายภาพ ผู้ใช้ทำการถ่ายภาพใบกล้วยไม้ จากนั้นแอปพลิเคชันจะใช้แบบจำลองเพื่อวินิจฉัยโรคกล้วยไม้และแสดงผลลัพธ์การวิเคราะห์ (ภาพที่ 4-5)



ภาพที่ 4: หน้าจอหลักแอปพลิเคชันและผลการวิเคราะห์โรคใบจุด



ภาพที่ 5: หน้าจอการวิเคราะห์โรคใบปื้นเหลืองและใบปกติ

จากผลการพัฒนาแอปพลิเคชัน พบว่าแอปพลิเคชันสามารถจำแนกและระบุภาพโรคกล้วยไม้จากใบแต่ละภาพได้อย่างอัตโนมัติ ซึ่งจะช่วยให้เพิ่มความสะดวกรวดเร็วในการวิเคราะห์โรคกล้วยไม้มากขึ้น

4.3 ผลการทดสอบประสิทธิภาพแอปพลิเคชัน

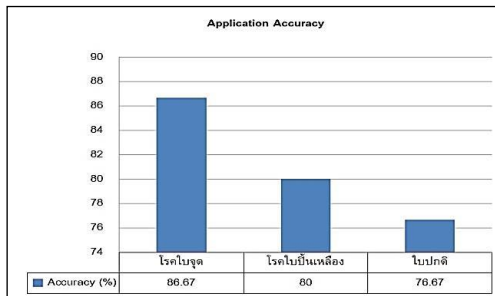
การทดสอบประสิทธิภาพการวินิจฉัยโรคกล้วยไม้ของแอปพลิเคชันเป็นการทดสอบด้วยภาพใบกล้วยไม้ผ่านกล้องโทรศัพท์มือถือ โดยดำเนินการการติดตั้งแอปพลิเคชันลงโทรศัพท์มือถือ ใช้ชุดข้อมูลสำหรับทดสอบ (Validate Data Set) จำนวน 90 ภาพ แบ่งเป็นภาพโรคใบจุดจำนวน 30 ภาพ โรคใบปื้นเหลืองจำนวน 30 ภาพ และใบปกติจำนวน 30 ภาพ โดยในงานวิจัยได้ใช้วิธีการทดสอบวัดประสิทธิภาพแอปพลิเคชันด้วยวิธี Leave One Out Cross Validation (LOOCV) [10] โดยข้อมูลจะถูกทดสอบทีละตัว

ตารางที่ 3: ผลการทดสอบประสิทธิภาพของแอปพลิเคชัน

ประเภทของโรคกล้วยไม้	จำนวนทั้งหมด	จำแนกถูกต้อง	จำแนกผิดพลาด	Accuracy (%)
โรคใบจุด	30	26	4	86.67
โรคใบปื้นเหลือง	30	24	6	80.00
ใบปกติ	30	23	7	76.67
<b>Average</b>				<b>81.11</b>

ตารางที่ 3 แสดงผลการทดสอบประสิทธิภาพการทำงานของแอปพลิเคชัน โดยสรุปค่าเฉลี่ยความถูกต้องในการวิเคราะห์ภาพใบกล้วยไม้ทั้ง 3 ประเภทเท่ากับ 81.11% ซึ่งมีประสิทธิภาพความแม่นยำอยู่ในระดับดี โรคกล้วยไม้ที่แอป

พลิกชันสามารถวิเคราะห์ได้ถูกต้องมากที่สุด ได้แก่ โรคใบจุด ได้ค่าความถูกต้อง 86.67% รองลงมาได้แก่ โรคใบปื้นเหลือง และใบปกติ ค่าความถูกต้อง 80.00% และ 76.67% ตามลำดับ



ภาพที่ 6: ผลการทดสอบข้อมูลแบ่งตามโรคของแอปพลิเคชัน

จากผลการทดสอบแอปพลิเคชันในภาพที่ 6 พบข้อสังเกต คือ ผลวิเคราะห์บางส่วนมีความคลาดเคลื่อน โดยเฉพาะภาพใบปกติที่ไม่เป็นโรค หากภาพจากกล้องสมาร์ทโฟนมีแสงมาก ผลวิเคราะห์จะมีความคลาดเคลื่อนสูง ส่วนโรคใบปื้นเหลือง พบว่าผลวิเคราะห์ของโรคใบปื้นเหลืองมีความคลาดเคลื่อนสูงกว่าโรคใบจุด เนื่องจากขนาดของโรคใบปื้นเหลืองมีขนาดเล็กมาก แอปพลิเคชันจึงวิเคราะห์ออกมาเป็นโรคใบจุด

## 5. สรุปผลการวิจัย

งานวิจัยนี้นำเสนอการพัฒนากระบวนการวินิจฉัยโรคกล้วยไม้จากรูปภาพในกล้องถ่ายรูปบนโทรศัพท์มือถือโดยใช้เทคนิคการเรียนรู้เชิงลึก มีวัตถุประสงค์เพื่อสร้างแบบจำลองวินิจฉัยโรคที่มีประสิทธิภาพและความสะดวกรวดเร็วในการวินิจฉัยโรคกล้วยไม้ โดยพัฒนาแบบจำลองจากโครงข่ายประสาทเทียมคอนโวลูชัน ซึ่งแบบจำลองที่พัฒนาขึ้นสามารถวิเคราะห์โรคกล้วยไม้ได้ 3 ประเภท คือ โรคใบจุด โรคใบปื้นเหลืองและใบปกติ ผ่านแอปพลิเคชันที่ถูกพัฒนาบนระบบปฏิบัติการแอนดรอยด์ ผลการทดลองพบว่าแบบจำลองการเรียนรู้เชิงลึกมีประสิทธิภาพในการจำแนกที่สูง โดยได้ค่าความถูกต้องเท่ากับ 98.10%

การตรวจสอบประสิทธิภาพการทำงานของแอปพลิเคชัน พบว่าแอปพลิเคชันสามารถวิเคราะห์ภาพและระบุโรคกล้วยไม้ มีค่าความถูกต้องเฉลี่ย 81.11% แสดงให้เห็นว่าแอปพลิเคชันที่นำเสนอสามารถวินิจฉัยโรคกล้วยไม้ได้อย่างมีประสิทธิภาพ สามารถนำไปประยุกต์ใช้งานได้จริง

ตามวัตถุประสงค์ที่ตั้งไว้ ช่วยให้เกษตรกรและผู้ที่สนใจสามารถวิเคราะห์โรคในกล้วยไม้ด้วยตนเองเบื้องต้นได้ทันทีก่อนที่จะโรคนกล้วยไม้จะรุนแรงและส่งผลกระทบต่อปริมาณผลผลิตที่ลดลง

## เอกสารอ้างอิง

- [1] สลิล สิทธิสังขธรรม. *คู่มือกล้วยไม้*. สำนักพิมพ์สารคดี: กรุงเทพฯ, 2558.
- [2] กรมส่งเสริมการค้าระหว่างประเทศ. *สินค้ากล้วยไม้*. สืบค้นเมื่อ 12 ธันวาคม 2020. Available Online at [http://www.ditp.go.th/contents\\_attach/539560/539560.pdf](http://www.ditp.go.th/contents_attach/539560/539560.pdf).
- [3] Y. LeCun, Y. Bengio, and G. Hinton. "Deep Learning." *Nature International Journal of Science*, Vol. 521, No. 7553, pp. 436-444, 2015.
- [4] T. Guo, J. Dong, H. Li, and Y. Gao. "Simple Convolutional Neural Network on Image Classification." *Proceedings of the IEEE International Conference on Big Data Analysis (ICBDA2017)*, Beijing, China, March 10-March 12, pp. 721-724, 2017.
- [5] K. Ferentinos. "Deep Learning Model for Plant Disease Detection and Diagnosis." *Computer and Electronics in Agriculture*, Vol. 145, No. 1, pp. 311-318, 2018.
- [6] J. Amara, B. Bouaziz, and A. Algergawy. "A Deep Learning Based Approach for Banana Leaf Diseases Classification." *Lecture Notes in Informatics*, Vol. 2017, No. 1, pp. 79-88, 2017.
- [7] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang. "Identification of Rice Diseases using Convolutional Neural Networks." *Neurocomputing*, Vol. 2017, No. 267, pp. 378-384, 2017.
- [8] B. Liu, Y. Zhang, D. He and Y. Li. "Identification of Apple Leaf Diseases based on Deep Convolutional Neural Networks." *Symmetry*, Vol. 10, No. 1, pp. 1-16, 2018.
- [9] M. Howlader, U. Habiba, R. Faisal and M. Rahman. "Automatic Recognition of Guava Leaf Diseases using Deep Convolutional Neural Network." *Proceedings of the International Conference on Electrical Computer and Communication Engineering (ECCE2019)*, Cox's Bazar, Bangladesh, February 7- February 9, pp. 1-5, 2019.
- [10] R. Kohavi. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Canada, August 20-August 25, pp. 1137-1143, 1995.



**ระบบสารสนเทศเพื่อการบริหารโครงการตามข้อกำหนดการจ้างงาน  
โดยใช้ทฤษฎีการจัดเรียงลำดับความสำคัญของงาน  
สำหรับหน่วยงานผู้ให้บริการที่ปรึกษาด้านสารสนเทศ  
Project Management Information System According to  
the Term of Reference by Using the Theory of Task Priorities for  
Information Consulting Service**

เดชทัต ศิริมัย (Taechathat Sirimai)<sup>1</sup> และณัฐวี อดุตฤกษ์ (Nattavee Utakrit)<sup>2</sup>  
ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล  
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ  
<sup>1</sup>s6207021810013@email.kmutnb.ac.th, <sup>2</sup>nattavee.u@itd.kmutnb.ac.th

### บทคัดย่อ

ปัจจุบันการดำเนินการของโครงการต่าง ๆ ของหน่วยงานรัฐบาลและรัฐวิสาหกิจในประเทศไทยมีแนวโน้มในการขยายตัวสูงกว่าปีก่อนหน้า จึงจำเป็นต้องมีการว่าจ้างที่ปรึกษาผู้เชี่ยวชาญเพื่อมาควบคุมดูแลคุณภาพของโครงการ โดยมีการนำระบบสารสนเทศเพื่อบริหารจัดการโครงการเข้ามาช่วยเพิ่มประสิทธิภาพในการทำงานและเป็นตัวควบคุมการประสิทธิภาพในการบริหารโครงการ ผู้วิจัยได้นำเสนอระบบบริหารโครงการจากข้อกำหนดการว่าจ้าง โดยประยุกต์ใช้ร่วมกับทฤษฎีการจัดเรียงลำดับความสำคัญของงานเพื่อควบคุมกำหนดการส่งงานในแต่ละงวดงานได้ด้วยระบบการจัดเรียงลำดับความสำคัญของงานแบบอัตโนมัติ โดยใช้วิธีการประมวลผลจากการขอบเขตการทำงานร่วมกับระยะเวลาการส่งงานของข้อกำหนดการส่งงานแต่ละโครงการที่ได้รับมา เพื่อเพิ่มประสิทธิภาพในการบริหารงานทำให้มองเห็นภาพรวมการดำเนินงานในปัจจุบันอย่างชัดเจนจากผลการประเมินประสิทธิภาพพบว่าระบบที่พัฒนาขึ้นสามารถจัดเรียงลำดับความสำคัญของงานที่ถูกสร้างขึ้นในแต่ละโครงการ จัดเรียงและแบ่งหมวดหมู่ของงานแต่ละประเภทได้อย่างชัดเจน รวมทั้งมีความถูกต้องในการ

บันทึกและแสดงผลข้อมูลอยู่ที่ 100% นอกจากนี้ผลการประเมินความพึงพอใจจากผู้ใช้งานอยู่ในระดับที่ดีอีกด้วย  
**คำสำคัญ:** ทฤษฎีการจัดลำดับความสำคัญของงาน ระบบบริหารโครงการ ข้อกำหนดการว่าจ้างโครงการ ที่ปรึกษา

### Abstract

At present, the implementation of various projects by government agencies and state enterprises in Thailand tends to expand higher than the previous year. Therefore, it is necessary to hire expert consultants to supervise project quality by bringing in project management information systems to enhance work efficiency and control efficiency in project management. The researcher presented the project management system based on the Term of Reference (TOR). Applying with the theory of prioritization of tasks to control the delivery schedule in each work period with an automated system of prioritization. Using the processing method from the scope of work together with the submission period of the submission requirements for each project received in order to increase the efficiency of management, the overview of current operations is

clearly visible. From the performance evaluation results, it was found that the developed system was able to prioritize the tasks created in each project. Clearly organized and categorized each type of work. Including the accuracy of recording and displaying data at 100%. In addition, the satisfaction assessment results from users are at a good level.

**Keywords:** Theory of prioritization of tasks, Project Management Information Systems, Term of Reference (TOR), Consultants.

## 1. บทนำ

ปัจจุบันการดำเนินการของโครงการต่าง ๆ ของหน่วยงานรัฐบาลและรัฐวิสาหกิจในประเทศไทยมีแนวโน้มในการขยายตัวสูงกว่าปีก่อนหน้า ปริยาเทศ นอก [1] ได้ให้ความเห็นไว้ว่าจากแรงหนุนของ Mega Projects ทั้งกลุ่มที่อยู่ระหว่างการดำเนินการก่อสร้าง รวมถึงโครงการที่ได้รับ การอนุมัติและมีความชัดเจนว่าจะได้รับการอนุมัติเพิ่มเติม และคาดว่าจะเริ่มก่อสร้างในไม่กี่ปีข้างหน้า จึงได้มีการนำระบบสารสนเทศเพื่อบริหารจัดการโครงการเข้ามาช่วยเพิ่มประสิทธิภาพในการทำงานและควบคุมการประสิทธิภาพในการบริหารโครงการเพื่อให้เกิดความถูกต้องตามกระบวนการต่าง ๆ และจากผลการวิจัยจากการใช้งานระบบ PM Software จำนวนกว่า 400 โครงการในงานวิจัยของ Eileen O'loughlin [2] ในปี 2562 พบว่า 70% ของผู้ใช้งานระบบ PM Software ให้ความเห็นว่าสามารถควบคุมประสิทธิภาพการบริหารโครงการด้านต่าง ๆ ได้ดีขึ้นและตรงตามเป้าหมายที่วางไว้ จากเครื่องมือบริหารจัดการที่มีให้ ผลที่ตามมาคือการดำเนินงานภายในโครงการมีประสิทธิภาพสูงขึ้น เทคโนโลยีสารสนเทศทำให้การจัดการกับกิจกรรมจำนวนมากและปริมาณข้อมูลมหาศาล แต่ในทางกลับกันผลสำรวจความพึงพอใจในการใช้งาน โดยรวมของระบบ PM Software พบว่ากว่า 40% จากผู้ใช้งานให้ความเห็นว่าใช้งานระบบได้ยาก และ 35% ใช้งานได้ไม่ตรงตามจุดประสงค์ของโครงการ

ปัจจุบันบริษัทที่ผู้วิจัยเลือกเป็นกรณีศึกษาได้ให้บริการที่ปรึกษาโครงการจำนวนหลายโครงการทั่วประเทศ และ

ได้มีการนำระบบ PM software เข้ามาใช้งาน แต่เมื่อทดลองใช้งานแล้วกลับพบว่าระบบมีความยุ่งยากซับซ้อน จึงจำเป็นต้องเพิ่มข้อมูลในการทำงานลงในระบบอยู่ตลอดเวลา รวมไปถึงการเรียงลำดับความสำคัญของแต่ละโครงการ บริษัทกรณีศึกษาให้ความเห็นว่า การนำระบบ PM Software เข้ามาช่วยบริหารจัดการนั้นยังไม่เหมาะสมกับการทำงานขององค์กรในปัจจุบันเท่าที่ควร

จากปัญหาที่ได้กล่าวมานั้นจึงทำให้บริษัทที่เลือกเป็นกรณีศึกษานี้จำเป็นต้องหาระบบเทคโนโลยีสารสนเทศและเทคนิคการจัดลำดับความสำคัญของงานเข้ามาใช้ในการบริหารโครงการ ผู้วิจัยจึงได้เลือกนำเสนอทฤษฎีที่จะนำมาใช้ร่วมกับการบริหารงานคือ วิธีการจัดเรียงความสำคัญของงานและบริหารเวลาด้วยการใช้เทคนิค Eisenhower Matrix ]3[ โดยนำมาจัดเรียงลำดับความสำคัญของงานโดยแบ่งออกเป็น 4 แนวทางปฏิบัติ ได้แก่ 1) งานเร่งด่วนและสำคัญ 2) สำคัญแต่ไม่เร่งด่วน 3) เร่งด่วนแต่ไม่สำคัญ 4) ไม่เร่งด่วนและไม่สำคัญ นำมาประยุกต์ใช้เข้ากับข้อกำหนดการว่าจ้างเพื่อควบคุมคุณภาพงานและระยะเวลาการส่งงาน โดยเปรียบเทียบและจัดลำดับความสำคัญของงานที่รับผิดชอบด้วยระบบอัตโนมัติ เพื่อเพิ่มประสิทธิภาพในการบริหารงานให้ได้มากยิ่งขึ้น

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

จากงานวิจัยที่ผ่านมาพบว่า การกำกับและติดตามโครงการ [4] คือการกำกับติดตามความก้าวหน้าโครงการตามระยะเวลาที่เหมาะสมในการติดตามที่ได้กำหนดไว้เพื่อทราบสภาพหรือ สถานการณ์ของโครงการโดยเปรียบเทียบกับสิ่งที่กำหนดไว้ นอกจากนี้ [5] ความสำคัญในการการควบคุมคุณภาพกระบวนการเพื่อให้เกิดความมั่นใจในการดำเนินงานทุกวันกระบวนการจะไม่ย้อนกลับไปเป็นแบบเดิม [5] หลักการบริหารเวลาที่ดียังจะต้องมีการวางแผนการใช้เวลาการตั้งเป้าหมายของภารกิจที่ต้องปฏิบัติตามการจัดลำดับความสำคัญของงานการกำหนดระยะเวลาในการทำกิจกรรม

ผู้วิจัยจึงได้ค้นคว้าข้อมูลเกี่ยวกับ การเรียงลำดับ ความสำคัญในการทำงานของ Dwight D. Eisenhower ที่มี ชื่อเรียกว่า Eisenhower Matrix ]6[ มีเกณฑ์ในการแบ่ง ประเภทของงาน ว่าสิ่งใดควรทำก่อน และอะไรควร ทำ ที่หลังเรียงลำดับตามความสำคัญและความเร่งด่วนของงาน เพื่อจัดเรียงกระบวนการทำงานให้ดียิ่งขึ้น โดยแบ่งประเภท ของการจัดลำดับงานหลัก ๆ ไว้ 2 แบบ คือ งานที่มีความ เร่งด่วน และ งานที่มีความสำคัญจากนั้นนำมาจับคู่เพื่อ ประเมินลำดับงานที่ต้องทำ โดยแบ่งจับคู่ตามความสำคัญ และความเร่งด่วน โดยความเร่งด่วนของงาน [7] มีเวลาเป็น ทรัพยากรที่มีความจำกัดที่สุด เพราะทรัพยากรอื่น ๆ สามารถหามาทดแทนหรือเพิ่มเติมได้แต่สำหรับเวลาแล้วมี เวลาอยู่ในจำนวนที่เท่ากัน ความสำคัญของงานในโครงการ ของบริษัทกรณีศึกษา ผู้เชี่ยวชาญจากบริษัท ๆ ได้ให้ ความเห็นเกี่ยวกับการจัดลำดับความสำคัญของงานได้แก่ ผู้ สั่งงาน มูลค่าและขนาดโครงการ และในส่วนขั้นตอนการ พิจารณาการจัดเรียงลำดับความสำคัญของงานนั้นผู้วิจัยได้ ทำการศึกษา [8] ระบบสนับสนุนการตัดสินใจ ซึ่งเป็น ระบบย่อยระบบหนึ่งในระบบสารสนเทศเพื่อการจัดการ โดยระบบสนับสนุนการตัดสินใจอาจจะใช้กับบุคคลเดียว หรือช่วยสนับสนุนการตัดสินใจเป็นกลุ่ม นอกจากนี้ยังเป็น ระบบที่เชื่อมโยงกันระหว่างทรัพยากรสมองของมนุษย์ ให้ทำงานร่วมกับความสามารถของคอมพิวเตอร์เพื่อ ต้องการการปรับปรุงคุณภาพของการตัดสินใจให้ได้ดีที่สุด ซึ่งผู้วิจัยได้นำเครื่องมือ [9] Decision Table เข้ามาใช้เป็น เกณฑ์ในการพิจารณาการจัดลำดับความสำคัญของงาน พร้อมกับพัฒนาระบบด้วย [10] วงจรการพัฒนา ระบบ ที่ ประกอบไปด้วยกิจกรรมต่าง ๆ สามารถนำมาประยุกต์ใช้ในการ พัฒนาเว็บแอปพลิเคชันได้ นอกจากนี้ยังได้ศึกษา งานวิจัยที่เกี่ยวข้องกับการประยุกต์ใช้ Eisenhower Matrix ]11[ ในหัวข้อ Eisenhower matrix \* Saaty AHP = การ จัดลำดับความสำคัญของการกระทำที่มีอิทธิพล โดยใช้ ทฤษฎีวรรณกรรมและบทเรียนจากหลักฐานเชิงประจักษ์ ผลการวิจัยพบว่า กลุ่มตัวอย่างสามารถใช้งานตารางที่ถูก ประยุกต์ขึ้นใหม่ได้จริง อีกทั้งยังประหยัดเวลาและเงินที่ใช้ ลงทุนได้อีกด้วย โดยเมทริกซ์ที่สร้างขึ้นใหม่มีชื่อเรียกว่า

Actions of Governance Matrix (AGM) นอกจากนั้น ผู้วิจัยยังศึกษางานวิจัยเกี่ยวกับระบบสนับสนุนการตัดสินใจ โดยใช้ตารางการตัดสินใจในหัวข้อ [12] ระบบวินิจฉัยโรค ชาติสซีเมียโดยใช้ตารางการตัดสินใจผลการทดลองพบว่า ระบบสามารถวินิจฉัยโรคชาติสซีเมียได้ใกล้เคียงกับผล วินิจฉัยของแพทย์ โดยให้ความแม่นยำสูงถึงร้อยละ 96.2

### 3. วิธีดำเนินการวิจัย

#### 3.1 ขั้นตอนการออกแบบโครงสร้างข้อมูล

กระบวนการทำงาน ของบริษัทกรณีศึกษา ปัจจุบันมีโครงการที่ดำเนินการอยู่เป็นจำนวนมากทำให้มีความจำเป็นอย่างยิ่งในการจัดสรรลำดับความสำคัญของ แต่ละโครงการเนื่องจากมีข้อกำหนดการว่าจ้างเข้ามาเป็น ตัวกำหนดการส่งงานของแต่ละโครงการประกอบกับ กระบวนการติดตามสถานะ โครงการที่ค่อนข้างไม่เป็น ระเบียบ การทำงานนั้นจะขึ้นอยู่กับแต่ละบุคคลว่าจะจัดการ งานที่ตนเองมี ส่งผลให้การทำงานขาดความแม่นยำและ ความคล่องตัวในกระบวนการทำงาน การตรวจสอบ สถานะข้อมูลที่สำคัญเพื่อให้ควบคุมคุณภาพงานนั้นถ้าเข้า จึงส่งผลให้เกิดความผิดพลาดตามไปด้วย ผู้วิจัยจึงได้ เริ่มต้นการออกแบบ โครงสร้างข้อมูล โดยดำเนินการ สำรวจและวิเคราะห์การทำงานและประสิทธิภาพของ ระบบบริหาร โครงการอื่น ๆ ที่มีการใช้งานกันอย่าง แพร่หลายและเปรียบเทียบประสิทธิภาพของการทำงาน เพื่อหาข้อดีและข้อเสียของแต่ละระบบ ผู้วิจัยจึงสรุปผล การวิเคราะห์ได้ว่าระบบการบริหารจัดการข้อมูลและ ผู้ร่วมโครงการนั้นมีหลากหลายประเภทงานแล้วแต่ โครงการที่ได้รับมอบหมาย การนำระบบเข้ามาเป็น ส่วนกลางการ รับ – ส่งข้อมูลภายในโครงการนั้นทำให้เกิด ความคล่องตัวในการ ได้มาซึ่งข้อมูลสูง จากนั้นผู้วิจัยได้ทำ การสำรวจความคิดเห็นจากผู้เชี่ยวชาญด้านการบริหาร โครงการจากบริษัทกรณีศึกษา ๆ โดยใช้ในหัวข้อ แบบสอบถามเกณฑ์การให้ความสำคัญของการบริหาร โครงการ หลังจากนั้นผู้วิจัยนำข้อมูลที่ได้จากผู้เชี่ยวชาญ มาทำการออกแบบหลักเกณฑ์การพิจารณาความสำคัญเพื่อ นำเข้าข้อมูลที่จำเป็นสู่ระบบสารสนเทศที่พัฒนาขึ้น เพื่อให้สอดคล้องกับการใช้เครื่องมือ ตารางการตัดสินใจ

)Decision Table) โดยสามารถแบ่งออกเป็น 3 หัวข้อ ดังนี้  
 1) ระยะเวลาการดำเนินงานโครงการ 2) ระดับของผู้ปฏิบัติงาน  
 3) มูลค่า / ขนาดโครงการ

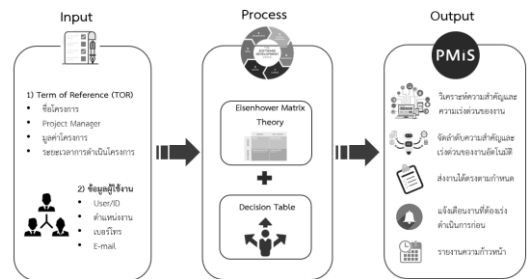
**3.2 ขั้นตอนการทำงานของระบบการสนับสนุนการตัดสินใจโดยใช้ทฤษฎีการจัดลำดับความสำคัญของงาน**

ผู้วิจัยได้ทำการออกแบบและจัดทำโครงสร้างของระบบการปรับปรุงประสิทธิภาพการบริหารโครงการตามข้อกำหนดการว่าจ้าง โดยนำระบบสนับสนุนการตัดสินใจเข้ามาประยุกต์เพื่อใช้ในการจัดลำดับความสำคัญตามทฤษฎีการจัดลำดับความสำคัญของงานแบบอัตโนมัติ โดยเลือกใช้เครื่องมือ Decision table โดยทำการจำแนกหมวดหมู่ของทฤษฎีการจัดลำดับความสำคัญของงานเพื่อสร้างเงื่อนไขสามารถแบ่งออกได้เป็น 4 หัวข้อ ได้แก่ แรงจูงใจและความสำคัญ, แรงจูงใจแต่ไม่สำคัญ, ไม่แรงจูงใจแต่สำคัญ, ไม่สำคัญและไม่แรงจูงใจ จากนั้นนำทฤษฎีการจัดเรียงลำดับความสำคัญของงานที่จำแนกหมวดหมู่ไว้มากำหนดในส่วนของ “เงื่อนไข” และ “การกระทำ” เพื่อใช้สำหรับการสร้างตารางการตัดสินใจโดยอ้างอิงแนวทางการสร้างเงื่อนไขการดำเนินงานโครงการจากผู้เชี่ยวชาญ หลังจากนั้นพิจารณาเงื่อนไขทั้งหมดและนำมากำหนดจำนวนแถวในส่วนของ “เงื่อนไข” ลงในตารางการตัดสินใจ และพิจารณาจำนวนทางเลือกที่เป็นไปได้ของแต่ละเงื่อนไข ซึ่งโดยทั่วไปจะมีเพียง 2 ทางคือ ใช่ หรือ ไม่ใช่ ซึ่งผู้วิจัยได้กำหนดเงื่อนไขการพิจารณาการตัดสินใจออกเป็น 14 เงื่อนไข เงื่อนไขละ 2 ทางเลือก และตรวจสอบความเป็นไปได้ของสถานการณ์ต่าง ๆ ในตารางที่มีความขัดแย้งและความซ้ำซ้อนของการกระทำต่าง ๆ และตัดทางเลือกที่เป็นไปไม่ได้ออกไป จึงสามารถสรุปและแปลผลของการกระทำ ตามกฎของการตัดสินใจ ในแต่ละเงื่อนไข

เงื่อนไข	237	238	239	240	241	242	243	244	245
เวลา < 50 %	Y	Y	Y	Y	Y	N	N	N	N
เวลา 50 - 70 %	N	N	N	N	N	Y	Y	Y	Y
เวลา > 70 %	N	N	N	N	N	Y	Y	Y	Y
ผู้บริหารสูงสุด	Y	N	N	N	N	Y	N	N	N
ผู้บริหารระดับกลาง	N	Y	Y	N	N	N	Y	Y	N
พนักงาน	N	N	N	Y	Y	N	N	N	Y
โครงการใหญ่	Y	Y	N	Y	N	Y	Y	N	Y
โครงการกลาง	Y	Y	N	N	Y	Y	Y	N	Y
โครงการเล็ก	Y	N	Y	N	Y	Y	N	Y	Y
การขอเงิน	X	X		X					
ไม่ส่งมอบผลิตภัณฑ์						X	X		
ส่งมอบผลิตภัณฑ์			X		X				
ไม่ส่งมอบโครงการ								X	X

ภาพที่ 1: ตารางการตัดสินใจ

**3.3 ขั้นตอนการวิเคราะห์และออกแบบระบบ (SDLC)**  
 เริ่มจากวิเคราะห์ระบบการทำงานเดิมในปัจจุบันขององค์กร ซึ่งการบริหารโครงการขององค์กรในปัจจุบันมีขั้นตอนในการรับ – ส่งงานที่ซ้ำซ้อน การบริหารจัดการงานในทีมยังไม่สามารถติดตามทั้งปริมาณงานและระยะเวลาการดำเนินงานโครงการที่แต่ละบุคคลภายในทีมเหลืออยู่แต่ละโครงการ ผู้วิจัยได้นำเสนอระบบบริหารจัดการโครงการด้วยระบบสารสนเทศโดยใช้ทฤษฎีการจัดเรียงลำดับความสำคัญของงานเข้ามาประยุกต์ใช้ เพื่อแก้ปัญหาการจัดลำดับงาน ซึ่งการแก้ปัญหาประกอบไปด้วย ระบบแจ้งเตือนและติดตามการดำเนินงานโครงการตามข้อกำหนดการว่าจ้าง และระบบการจัดลำดับความสำคัญของงาน ผู้วิจัยทำการวิเคราะห์ระบบงานใหม่ และได้นำระบบ PM software เข้ามาบริหารจัดการโครงการเข้ามาใช้ซึ่งจะทำให้การบริหารโครงการมีประสิทธิภาพมากยิ่งขึ้น โดยเน้นไปที่การบริหารจัดการเรื่องลำดับความสำคัญของงาน โดยนำเข้าข้อมูลก่อนเข้าระบบก่อนเริ่มบันทึกโครงการเพื่อให้สามารถประมวลผลได้อัตโนมัติ โดยหลักการทำงานของระบบจะจัดเรียงงานของผู้ใช้งานตามเงื่อนไขของระบบที่สร้างขึ้นอัตโนมัติเพื่อให้ผู้ใช้งานสามารถทราบถึงความสำคัญของงานว่างานไหนควรดำเนินการก่อน หรือ หลัง และระบบติดตามความคืบหน้าการดำเนินงานโครงการเข้ามาเป็นส่วนหลักของระบบ เพื่อเน้นการทำงานควบคู่กับระบบการจัดเรียงลำดับความสำคัญของงาน



ภาพที่ 2: กรอบแนวความคิดการออกแบบระบบ

**3.4 การพัฒนาระบบ**

การปรับปรุงประสิทธิภาพการบริหารโครงการตามข้อกำหนดการว่าจ้างด้วยระบบสารสนเทศ โดยใช้

ทฤษฎี การจัดเรียงลำดับความสำคัญของงาน สำหรับหน่วยงานผู้ให้บริการที่ปรึกษาด้านสารสนเทศ ผู้วิจัยได้สร้างฐานข้อมูลตามที่ได้ออกแบบไว้ โดยใช้ SQL Server ในการจัดเก็บโครงสร้างและข้อมูล และทำการพัฒนา User Interface (UI) จากการออกแบบส่วนต่อประสานงานผู้ใช้ โดยใช้โปรแกรม Visual Studio Code และใช้ Framework Vue JS ในการพัฒนา จากนั้น ผู้วิจัย ได้ทำการพัฒนา Service สำหรับเชื่อมต่อระหว่าง UI และ Database โดยใช้โปรแกรม Microsoft Visual Studio และ #C ในการพัฒนาระบบ



ภาพที่ 3: หน้าจอของระบบจริง

### 3.5 การประเมินผลระบบ

การประเมินผลระบบแบ่งออกเป็นการประเมินประสิทธิภาพโดยผู้ผู้เชี่ยวชาญในการพัฒนาโครงการจำนวน 3 คน โดยใช้แบบประเมิน UAT ซึ่งเป็นวิธีประเมินแบบ Black-Box Testing และ การประเมินผลความพึงพอใจของระบบโดยผู้ประเมินเป็นกลุ่มผู้ใช้งานระบบจำนวน 30 คน วิเคราะห์โดยใช้สถิติพื้นฐาน และการทดสอบสมมุติฐานด้วยวิธี One-Sample Test (t-test)

## 4. ผลการดำเนินงาน

### 4.1 ผลการทดสอบประสิทธิภาพของระบบ

ตารางที่ 1 ผลการทดสอบประสิทธิภาพของระบบ

ระบบ	ประสิทธิภาพ
1. ระบบจัดการผู้ใช้งานและข้อมูลพื้นฐานของผู้ใช้งาน	100%
2. ระบบการจัดการโครงการ	100%
3. ระบบบริหารโครงการโดยใช้ทฤษฎีการจัดเรียงลำดับความสำคัญของงาน	100%
4. ระบบติดตามความก้าวหน้าการดำเนินงาน	100%

จากตารางที่ 1 พบว่าผลการวัดประสิทธิภาพการทำงานทั้ง 4 ด้านมีผลการประเมินที่ผ่านการทดสอบที่ 100% โดย

การทดสอบประสิทธิภาพฟังก์ชันหลักระบบบริหารโครงการ โดยใช้ทฤษฎีการจัดเรียงลำดับความสำคัญของงาน สามารถจัดเรียงลำดับความสำคัญของงานที่ถูกสร้างขึ้นในแต่ละโครงการ ทำให้จัดเรียงและแบ่งหมวดหมู่ของงานแต่ละประเภทได้อย่างถูกต้อง โดยใช้เครื่องมือแบบทดสอบประสิทธิภาพจากผู้เชี่ยวชาญดังนี้

รหัสการทดสอบ	UAT_MP_002	กระบวนการที่ทดสอบ	ระบบบริหารโครงการโดยใช้ทฤษฎีการจัดเรียงลำดับความสำคัญของงาน
เมนูระบบ	ระบบการจัดลำดับความสำคัญและเร่งด่วนของงาน		
รายละเอียดการทดสอบระบบ	ทดสอบความถูกต้องในการแสดงผลและการจัดการข้อมูล ในหมวด "เร่งด่วนและสำคัญ"		
ผลที่คาดหวัง	ระบบสามารถจัดลำดับความสำคัญและเร่งด่วนของงานให้อยู่ในหมวด "เร่งด่วนและสำคัญ" ได้อย่างถูกต้อง		
ผู้ทดสอบคนที่	1	2	3
ผลลัพธ์	<input checked="" type="checkbox"/> ผ่าน <input type="checkbox"/> ไม่ผ่าน	<input checked="" type="checkbox"/> ผ่าน <input type="checkbox"/> ไม่ผ่าน	<input checked="" type="checkbox"/> ผ่าน <input type="checkbox"/> ไม่ผ่าน

ภาพที่ 4: ตัวอย่างแบบทดสอบประสิทธิภาพ

## 4.2 ผลการประเมินความพึงพอใจของผู้ใช้งาน

ตารางที่ 2 ผลการประเมินความพึงพอใจของผู้ใช้ระบบ

รายการประเมิน	$\bar{x}$	S.D.
1. ผลการประเมินด้านประสิทธิภาพของระบบสารสนเทศ	3.95	0.72
2. ผลการประเมินด้านการออกแบบของระบบสารสนเทศ	4.00	0.75
3. ผลการประเมินด้านความปลอดภัยของระบบสารสนเทศ	4.01	0.86
สรุปผลการประเมินความพึงพอใจที่มีต่อระบบสารสนเทศ	3.99	0.78

จากตารางที่ 2 พบว่าผลการประเมินความพึงพอใจของกลุ่มตัวอย่างผู้ใช้งานที่ทำการทดลองใช้ระบบจำนวน 30 คน มีผลการทดสอบอยู่ในระดับความพึงพอใจมาก โดยมีค่าเฉลี่ยเท่ากับ 3.99 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.78

## 5. สรุป อภิปรายผล

การศึกษาวิจัยนี้ได้นำเสนอการบริหารจัดการโครงการโดยใช้เทคโนโลยีสารสนเทศเข้ามาประยุกต์ใช้ร่วมกันกับทฤษฎีการจัดเรียงลำดับความสำคัญของงาน เพื่อเพิ่มความสะดวกในการประมวลผลความเร่งด่วนและความสำคัญของงานที่ได้รับมอบหมายในแต่ละโครงการและงวดงาน โดยมีระบบการจัดลำดับความเร่งด่วนและความสำคัญของงานที่จัดลำดับความสำคัญให้แก่ผู้ใช้งานได้แบบอัตโนมัติ ผู้ใช้สามารถกรอกข้อมูลที่จำเป็นเข้าสู่ระบบและสามารถ

พิจารณาตามเงื่อนไขของทฤษฎีการจัดเรียงลำดับความสำคัญได้ทันที ทำให้สามารถประเมินการทำงานของตนเองได้ว่าสิ่งใดควรเริ่มดำเนินการทันที นอกจากนั้นยังมีระบบติดตามความก้าวหน้าของโครงการที่คอยกำกับผู้ใช้งานให้ได้เห็นถึงความคืบหน้าของแต่ละโครงการที่กำลังดำเนินการอยู่อีกทั้งยังแจ้งเตือนถึงจำนวนงานที่ผู้ใช้งานกำลังดำเนินการซึ่งสอดคล้องกับความต้องการจัดลำดับความเร่งด่วนและความสำคัญที่ผู้ใช้มีทั้งหมดอีกด้วย โดยผลการประเมินที่ได้ผ่านการทดสอบโดยผู้เชี่ยวชาญอยู่ที่ 100% ทำให้ผลลัพธ์ที่ทำการจัดลำดับความเร่งด่วนและความสำคัญมีความถูกต้องและผลการประเมินความพึงพอใจของผู้ใช้งานอยู่ในระดับมีความพึงพอใจมาก

## 6. ข้อเสนอแนะในการวิจัย

ในการพัฒนาต่อไปเพื่อความสมบูรณ์ยิ่งขึ้นของระบบ ควรเพิ่มฟังก์ชันการจัดทำแผนการดำเนินงานบนระบบได้อย่างละเอียดและสามารถเรียกรายงานแผนการดำเนินงานออกมาได้ในหลากหลายรูปแบบ เพื่อให้สอดคล้องกับกิจกรรมที่กำลังดำเนินอยู่ในโครงการ ณ ปัจจุบันทำให้เห็นถึงภาพรวมการดำเนินงานทุกกิจกรรมทั้งหมดในโครงการได้อย่างชัดเจน รวมถึงสามารถนำเสนอรายงานแผนการดำเนินงานให้แก่ผู้ว่าจ้างหรือผู้บังคับบัญชาได้อย่างง่าย

ในอนาคตผลงานวิจัยนี้สามารถนำไปต่อยอดการศึกษาที่เกี่ยวข้องได้เช่น ระบบบริหารองค์กร ระบบติดตามความก้าวหน้าการดำเนินงานในสายงานอื่น หรือแม้กระทั่งระบบบริหารโครงการที่สามารถนำระบบจัดลำดับความสำคัญของงานเข้าไปประยุกต์ใช้เพื่อเพิ่มประสิทธิภาพในการทำงานให้ดียิ่งขึ้น

## เอกสารอ้างอิง

- [1] ปรีชา เทศนอก. “แรงหนุนของ Mega Projects” สืบค้นเมื่อ 20 สิงหาคม 2563. จากเว็บไซต์ Available Online at <https://www.prop2morrow.com>, 2560.
- [2] Eileen O’ loughlin. “Project Management User Research Report” สืบค้นเมื่อ 20 สิงหาคม 2563. Available Online at <https://www.capterra.com/project-management-software/user-research/>, 2562.
- [3] Eisenhower Matrix คืออะไร. สืบค้นเมื่อ 1 ตุลาคม 2563. จากเว็บไซต์ Available Online at [https://www.dropbox.com/th\\_TH/business/resources/eisenhower-matrix](https://www.dropbox.com/th_TH/business/resources/eisenhower-matrix), 2550.

- [4] ทวีป ศิริรัศมี. การวางแผนและการประเมินโครงการ. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร : สำนักงานกองทุนสนับสนุนงานวิจัย (สกว.). 2545.
- [5] สกาวรัตน์ จงพัฒนากร. การวิเคราะห์และออกแบบระบบสารสนเทศ. กรุงเทพมหานคร : สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์. 2550.
- [6] Campbell R. F. Corbally. J. E. and Nystrand. *Introduction to educational administration*. R.O. 6th ed. Boston : Allyn and Bacon. 2526.
- [7] กัทรพร นีรเนตร. ปัจจัยความเสี่ยงของการพัฒนาอสังหาริมทรัพย์ ประเภทอาคารชุดในกรุงเทพมหานคร. วิทยาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมการพัฒนาอสังหาริมทรัพย์ คณะสถาปัตยกรรมศาสตร์และการผังเมือง มหาวิทยาลัยธรรมศาสตร์. 2557.
- [8] ชีววรรณ เจริญสุข. (2557). บทที่ 12 ระบบสนับสนุนการตัดสินใจและการดำเนินการ (DECISION SUPPORT AND IMPLEMENTS). (บทเรียนหลักสูตรวิชาการบริหารธุรกิจ) [สืบค้น วันที่ 20 กรกฎาคม 2564]. จากเว็บไซต์ [HTTPS://MAYMAYNY.WORDPRESS.COM/2014/12/05/บทที่-12-ระบบสนับสนุนการตัดสินใจ/#\\_ENREF\\_2](https://maymayny.wordpress.com/2014/12/05/บทที่-12-ระบบสนับสนุนการตัดสินใจ/#_ENREF_2)
- [9] เอกสารประกอบ ส่วนการจัดการแบบจำลอง (Model Management). (2550). [ออนไลน์] สืบค้น วันที่ 20 กรกฎาคม 2564]. Available Online at <https://staff.informatics.buu.ac.th>
- [10] เทพฤทธิ์ สินธรรังค์. ระบบสนับสนุนการตัดสินใจสำหรับประเมินศักยภาพ ของการจัดการโลจิสติกส์และห่วงโซ่อุปทาน. วิทยาสตรมหาบัณฑิตวิศวกรรมอุตสาหการ มหาวิทยาลัยเชียงใหม่. 2553.
- [11] สุวัฒน์ ชาติวธ. ระบบวินิลัยโรคราษสีขีเมียโดยใช้ตารางการตัดสินใจ. วิทยาสตรมหาบัณฑิต สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะ วิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. 2547.
- [12] Alfred Homère NGANDAM MFONDOUM. (2562). “Eisenhower matrix \* Saaty AHP = Strong actions prioritization? Theoretical literature and lessons drawn from empirical evidences” สืบค้นเมื่อ 3 มีนาคม 2565. Available Online at [https://www.researchgate.net/publication/333516142\\_Eisenhower\\_matrix\\_Saaty\\_AHP\\_Strong\\_actions\\_prioritization\\_Theoretical\\_literature\\_and\\_lessons\\_drawn\\_from\\_empirical\\_evidences](https://www.researchgate.net/publication/333516142_Eisenhower_matrix_Saaty_AHP_Strong_actions_prioritization_Theoretical_literature_and_lessons_drawn_from_empirical_evidences)

**การจัดการระบบเครือข่ายไร้สายเพื่อการใช้งานโครงการมหาวิทยาลัยสีเขียว  
กรณีศึกษา ระบบควบคุมเครื่องปรับอากาศและระบบถังขยะกึ่งอัจฉริยะ  
มหาวิทยาลัยวลัยลักษณ์**

**Wireless Network Management for Green University Case Study Air  
Control System and Semi-Smart Bin System of Walailak University**

สาธิต ขวัญชุม (Satit Khwanchum)

แผนกสารสนเทศ โรงพยาบาลศูนย์การแพทย์มหาวิทยาลัยวลัยลักษณ์ มหาวิทยาลัยวลัยลักษณ์  
ksatit@gmail.com

### บทคัดย่อ

งานวิจัยนี้เป็นการศึกษาและพัฒนารูปแบบของระบบเครือข่ายไร้สายเพื่อใช้ในโครงการมหาวิทยาลัยสีเขียวเกี่ยวกับระบบควบคุมเครื่องปรับอากาศและระบบถังขยะกึ่งอัจฉริยะ ผ่านการควบคุมโดยระบบเครือข่ายไร้สาย การดำเนินการได้พิจารณาถึงความครอบคลุมของสัญญาณความปลอดภัยในการใช้งาน การตรวจสอบสถานะ และอัตราการใช้งานข้อมูล ผลการดำเนินการของระบบควบคุมเครื่องปรับอากาศพบว่ามีระดับความแรงของสัญญาณอยู่ในช่วง -35 dBm ถึง -47 dBm มีการเข้ารหัสข้อมูลแบบ Open การยืนยันตัวตนแบบ Mac Authentication อัตราการใช้งานข้อมูลมีค่าเฉลี่ยไม่เกิน 400 KB ส่วนระบบถังขยะกึ่งอัจฉริยะมีระดับความแรงของสัญญาณอยู่ในช่วง -49 dBm ถึง -54 dBm มีการเข้ารหัสข้อมูลแบบ WPA2 มีการยืนยันตัวตนแบบ Open มีอัตราการใช้งานข้อมูลมีค่าเฉลี่ยไม่เกิน 10 KB โดยจากการทดสอบและใช้งานพบว่าการทำงานของทั้งสองระบบสามารถทำงานได้อย่างถูกต้อง และมีประสิทธิภาพสอดคล้องกับความต้องการของโครงการ

**คำสำคัญ:** ระบบเครือข่ายไร้สาย มหาวิทยาลัยสีเขียว การยืนยันตัวตน การเข้ารหัสข้อมูล

### Abstract

*This research aims to study and develop the services of wireless network systems for supporting the green university project about the air control*

*system (ACS) and the semi-smart bin system (SBS). They are both controlled by wireless networks. The researcher considerations include; the received signal strength indicator (RSSI), wireless security, status monitoring, and data traffic usages. The results showed the ACS had RSSI in the range of -35 dBm to -47 dBm, encryption type was open, authentication type was Mac, and the average data traffic usages was less than 400 KB. The SBS had RSSI in the range of -49 dBm to -54 dBm, encryption type was wpa2, authentication type was open, and the average data traffic usages was less than 10 KB. Both systems can be monitored through the Wireless Controller, and Network Monitoring System. Test results and applications show that both systems can work properly and effectively in line with project requirements.*

**Keywords:** Wireless Network System, Encryption, Green University, Authentication.

### 1. บทนำ

มหาวิทยาลัยสีเขียว (Green University) เป็นการจัดลำดับมหาวิทยาลัยที่เป็นมิตรต่อสิ่งแวดล้อม เพื่อส่งเสริมให้ประยุกต์การอนุรักษ์พลังงานและสิ่งแวดล้อม เข้ากับการเรียนการสอน การวิจัย การทำงาน และการใช้ชีวิตประจำวันของนักศึกษา และบุคลากร เพื่อให้เกิดความ

ปลอดภัย ความสุขในที่ทำงาน เป็นมิตรกับสิ่งแวดล้อม ประหยัดพลังงาน และการพัฒนาที่ยั่งยืน [1], [2]

การจัดลำดับมหาวิทยาลัยสีเขียว โดยมหาวิทยาลัย อินโดนีเซียเริ่มในปี พ.ศ. 2553 ภายใต้อำนาจ การจัดลำดับ มหาวิทยาลัยสีเขียว โลก ยูไอ กรีน เมตริก (UI GreenMetric World University Ranking) มี เกณฑ์ ประกอบด้วย การจัดการ โครงสร้างพื้นฐาน 15% การจัดการ พลังงานและการเปลี่ยนแปลงสภาพภูมิอากาศ 21% การจัดการของเสีย 18% การจัดการน้ำ 10% การจัดการ ระบบขนส่ง 18% และ การจัดการศึกษาและวิจัยด้าน สิ่งแวดล้อม 18% [1], [3], [6]

มหาวิทยาลัยวลัยลักษณ์ได้กำหนดนโยบายการเป็น มหาวิทยาลัยสีเขียวไว้ในแผนยุทธศาสตร์ 20 ปี (พ.ศ. 2561-2580) ในประเด็นยุทธศาสตร์ที่ 5 การเสริมสร้าง ภาพลักษณ์เป็นเมืองมหาวิทยาลัยสีเขียวแห่งความสุข และ จัดตั้งคณะกรรมการขับเคลื่อนเพื่อการสร้างกระบวนการ เรียนรู้ การทำความเข้าใจและส่งเสริมความร่วมมือในการ ดำเนินการเพื่อเข้าร่วมโครงการมหาวิทยาลัยสีเขียว ซึ่งผล การจัดลำดับในปี 2563 อยู่อันดับที่ 16 ของประเทศ อันดับที่ 219 ของโลก และปี 2564 อยู่อันดับที่ 8 ของ ประเทศ อันดับที่ 116 ของโลก [4], [5]

ภายใต้ นโยบายการเป็น มหาวิทยาลัยสีเขียว คณะกรรมการได้จัดทำโครงการย่อยเกี่ยวกับพลังงาน โดย เป็นระบบควบคุมการทำงานของเครื่องปรับอากาศ (ACS) แบบอัตโนมัติ และ โครงการย่อยทางด้านการจัดการของ เสีย ซึ่งเป็นระบบแจ้งเตือนปริมาณถังขยะเต็มไปยัง ผู้จัดการขยะผ่านทางไลน์ เรียกว่า ระบบถังขยะกิ่งอัจฉริยะ (SBS) โดยระบบทั้งสองจะอาศัยระบบเครือข่ายไร้สายใน การสั่งการทำงาน ผู้วิจัยจึงได้ทำการศึกษาและทดสอบ รูปแบบการติดตั้ง การยืนยันตัวตนและการกำหนดค่า ระบบเครือข่ายไร้สายที่เหมาะสม เพื่อให้รองรับการใช้งาน ของระบบ ACS และระบบ SBS ในมหาวิทยาลัยให้ สามารถทำงานได้อย่างถูกต้องและมีประสิทธิภาพ

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การรักษาความปลอดภัยของเครือข่ายไร้สาย [7]

การรักษาความปลอดภัยเป็นการตรวจสอบและ ป้องกันระบบเครือข่ายไร้สายไม่ให้ถูกใช้งานหรือเข้าถึง โดยผู้ที่ไม่ได้รับสิทธิ์ ซึ่งจะประกอบด้วย 2 ส่วน คือการ เข้ารหัสข้อมูล (Encryption) และการตรวจสอบผู้ใช้งาน ระบบ (Authentication)

### 2.2 การเข้ารหัสข้อมูล [8], [9]

การเข้ารหัสข้อมูลมีหลายรูปแบบ โดยรูปแบบที่มีการ ใช้งานหลัก ๆ มีดังนี้ 1) แบบ Open 2) แบบ WEB 3) แบบ WPA และ 4) แบบ WPA2

### 2.3 การยืนยันตัวตนของระบบเครือข่าย [10]

การยืนยันตัวตน เป็นการตรวจสอบสิทธิ์เพื่อแสดง ตัวตนของอุปกรณ์หรือบุคคลก่อนที่จะเข้าใช้งานระบบ เครือข่าย ซึ่งมีดังนี้ 1) Open Authentication 2) Mac Authentication 3) Web Captive Portal และ 4) 802.1x

### 2.4 NodeMCU (ESP8266) [11], [12]

NodeMCU (ESP8266) เป็น Platform ที่ ใช้ใน โครงการ Internet of Things (IoT) ประกอบด้วยบอร์ด วงจร และ Software ที่ติดตั้งบนบอร์ด NodeMCU จะ นิยมติดตั้งโมดูล WiFi (ESP8266) เพื่อใช้สำหรับการ เชื่อมต่อกับเครือข่ายไร้สาย ปัจจุบันที่นิยมใช้รุ่น ESP8266 NodeMCU V2 โดยจะทำงานที่ ไฟฟ้า กระแสตรง แรงดันไฟฟ้า 3.0-3.6 V กระแสไฟฟ้าเฉลี่ย 80 mA

### 2.5 Sonoff [13], [14], [15]

Sonoff เป็น อุปกรณ์ IoT ทำหน้าที่เป็น Switch ควบคุมแบบไร้สายสำหรับการสั่งเปิด-ปิดอุปกรณ์ไฟฟ้า โดยใช้ ESP8266 ในการเชื่อมต่อ ใช้งานกับ ไฟฟ้า กระแสสลับที่ช่วงแรงดัน 90-250 V รองรับกระแสได้ สูงสุด 10 Am โดยอุปกรณ์รองรับการเปิด-ปิดแบบการตั้ง เวลา และสามารถเขียน โปรแกรมลงบนบอร์ดเพื่อสั่งการ ได้ นิยมใช้เป็น Relay ควบคุมจากระยะไกล

### 2.6 งานวิจัยที่เกี่ยวข้อง

ข้อมูลการศึกษาแนวทางในการจัดการระบบเครือข่าย และการประเมินประสิทธิภาพของระบบ ผู้ใช้บริการจะมีความคาดหวังเกี่ยวกับความเร็วในการใช้งานระบบที่ เหมาะสม ความครอบคลุมของสัญญาณในพื้นที่ใช้งาน และความต่อเนื่องในการให้บริการ [16], [17] โดยปัจจัย ต่อการใช้งานระบบและการให้บริการที่มีคุณภาพจะ

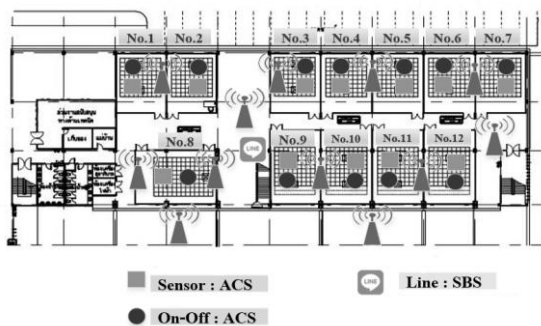


เน้นหนักไปทางด้านความครอบคลุมของสัญญาณ และการใช้งานระบบที่ต่อเนื่อง ซึ่งเงื่อนไขที่สำคัญประการหนึ่งของการให้บริการ คือระดับความแรงของสัญญาณไม่ควรจะต่ำกว่า -75 dBm นอกจากนี้ยังให้สำคัญเกี่ยวกับความปลอดภัยในการใช้งาน ขั้นตอนการใช้งานที่เป็นมิตรกับผู้ใช้ หากเป็นการเชื่อมต่อและตรวจสอบข้อมูลแบบอัตโนมัติหรือน้อยครั้งที่สุดท้ายจะมีผลดีต่อการใช้งานระบบ [18-21]

### 3. การดำเนินการ

#### 3.1 รูปแบบและพื้นที่การติดตั้งระบบ

ระบบควบคุมเครื่องปรับอากาศ (ACS) และระบบถึงขยะกิ่งอัจฉริยะ (SBS) มีอุปกรณ์ IoT ทำหน้าที่ควบคุมการทำงานผ่านเครือข่ายไร้สาย ทั้งนี้ระบบ ACS จะใช้ Sonoff ส่วนระบบ SBS จะใช้ NodeMCU ระบบ ACS ได้ติดตั้งไว้ในห้องเรียนขนาด 35 ที่นั่ง (8x10 ตร.ม.) จำนวน 40 ห้อง ส่วนระบบ SBS ติดตั้งไว้ที่โถงกลางของอาคาร จำนวน 6 จุด โดยระบบทั้งสองจะติดตั้งไว้ใกล้กับระบบเครือข่ายไร้สายที่มีให้บริการอยู่ในพื้นที่นั้น ๆ ดังภาพที่ 1 การติดตั้งระบบ ACS ในห้องเรียน และระบบ SBS ที่โถงกลางของอาคารสถาปัตยกรรม Zone A



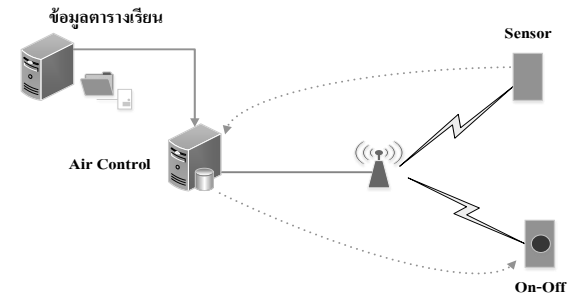
ภาพที่ 1: การติดตั้งระบบ ACS และ SBS ในอาคาร Zone A

#### 3.2 ระบบควบคุมเครื่องปรับอากาศ (ระบบ ACS)

ระบบ ACS เป็นระบบเปิด-ปิดการทำงานของเครื่องปรับอากาศ สั่งงานผ่าน Sonoff และเครือข่ายไร้สาย ซึ่งจะทำงานเมื่อ Server ตรวจพบว่ามีตารางเรียน (เงื่อนไขหลัก) และมีการเคลื่อนไหวภายในห้องเรียน (เงื่อนไขร่วม) ดังแสดงในตารางที่ 1 และภาพที่ 2

ตารางที่ 1: เงื่อนไขการทำงานของระบบ ACS

ตารางเรียน	การเคลื่อนไหวของ User	Sonoff
มี	มี	On
มี	ไม่มี	Off
ไม่มี	มี	Off
ไม่มี	ไม่มี	Off



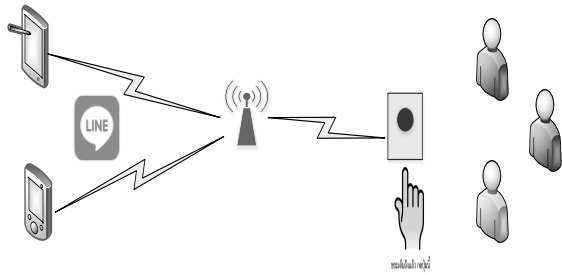
ภาพที่ 2: ภาพรวมของระบบ ACS

การทำงานของอุปกรณ์ตรวจจับความเคลื่อนไหว จะใช้คลื่นไมโครเวฟที่มีรัศมีการทำงานไม่เกิน 10 เมตร โดยจะตรวจสอบและส่งต่อข้อมูลอย่างต่อเนื่องไปยัง Server เพื่อให้เครื่อง Server ทำการประมวลผลและสั่งการผ่านเครือข่ายไร้สายมายังอุปกรณ์ Sonoff เพื่อเปิด-ปิดระบบ

#### 3.3 ระบบถึงขยะกิ่งอัจฉริยะ (ระบบ SBS)

ระบบ SBS เป็นระบบแจ้งเตือนผ่านไลน์ไปยังผู้จัดการขยะเมื่อถึงขยะเต็มด้วยการกดปุ่มแจ้งเตือน โดยอาศัย NodeMCU และเครือข่ายไร้สาย ข้อมูลการแจ้งเตือนจะประกอบด้วยสถานะของถังขยะ และจุดวางถังขยะส่งไปทางไลน์เพื่อแจ้งให้ผู้ดูแลได้รับทราบและดำเนินการ ซึ่งผู้ดูแลจะเปลี่ยนถุงใส่ขยะใหม่และเก็บขยะเดิมออกไปเพื่อให้สามารถนำขยะมาทิ้งในถังได้ต่อ ดังภาพที่ 3 สำหรับขั้นตอนการทำงานมีดังนี้

- 1) นักศึกษา/บุคลากร มองเห็นว่าถังขยะเต็ม
- 2) นักศึกษา/บุคลากร กดปุ่มแจ้ง “ถังขยะเต็ม”
- 3) ระบบรับทราบการแจ้งเตือนจากการกดปุ่ม และส่งข้อมูลการแจ้ง “ถังขยะเต็ม” ผ่านไลน์ไปยังกลุ่มผู้ดูแลและจัดการขยะประจำอาคาร โดยอาศัยระบบเครือข่ายไร้สาย
- 5) กลุ่มผู้ดูแลนำถุงใส่ขยะใบใหม่มาเปลี่ยน แล้วส่งข้อมูลแจ้งกลับไปในกลุ่มหลังเปลี่ยนถุงใส่ขยะแล้วเสร็จ
- 6) ระบบ Standby รอรับการกดปุ่มแจ้ง “ถังขยะเต็ม”



ภาพที่ 3: ภาพรวมการทำงานของระบบ SBS

**3.4 รูปแบบการยืนยันตัวตนและการเข้ารหัสข้อมูล**

รูปแบบในการยืนยันตัวตนของผู้ใช้งานระบบเครือข่าย มีอยู่ 4 แบบ เพื่อให้สอดคล้องกับเงื่อนไขในการใช้งานระบบเครือข่ายขององค์กร และความสามารถที่จะทำให้ระบบ ACS และระบบ SBS สามารถใช้งานได้ จึงได้มีการตรวจสอบรูปแบบการยืนยันตัวตนและการเข้ารหัสข้อมูลทั้งหมดแล้วพิจารณาเลือกรูปแบบที่เหมาะสม

**3.5 การตรวจสอบสถานะการทำงานของระบบ**

การทำงานของระบบ ACS และระบบ SBS สามารถตรวจสอบได้จากสถานะการเชื่อมต่อหรือการ Online ของอุปกรณ์ และปริมาณการใช้งาน Traffic ที่เกิดขึ้นบนระบบเครือข่ายไร้สาย โดยจะตรวจสอบผ่านอุปกรณ์บริหารจัดการกลาง (Wireless Controller) และระบบ Network Monitoring System หากระบบสามารถแสดงผลการเชื่อมต่อและการใช้งานเครือข่ายไร้สายได้ ก็จะสามารถตรวจสอบการทำงานของระบบ ACS และระบบ SBS ได้

**4. ผลการดำเนินการและการวิเคราะห์ผล**

**4.1 ผลการติดตั้งระบบ**

อุปกรณ์ Sonoff ที่ใช้สำหรับการเปิด-ปิดระบบ ACS จะติดตั้งภายในห้องเรียน ส่วนอุปกรณ์ NodeMCU ในการควบคุมการส่งไลน์ของระบบ SBS จะติดตั้งบริเวณโถงกลางอาคาร ซึ่งมีความเหมาะสม เนื่องจากอยู่ใกล้อุปกรณ์กระจายสัญญาณแบบไร้สาย โดยระบบ ACS มีความครอบคลุมของสัญญาณเครือข่ายระหว่าง -35 dBm ถึง -47 dBm และระบบ SBS มีความครอบคลุมของสัญญาณเครือข่ายระหว่าง -49 dBm ถึง -54 dBm ดังแสดงในภาพที่ 4 และภาพที่ 5

IP Address	RSSI	Packets to	Packets from
172.16.205.87 / ::	-35 dBm	20.3K	2.4K
172.16.205.89 / ::	-38 dBm	46.3K	6.3K
172.16.205.80 / ::	-38 dBm	28.8K	3.7K
172.16.205.40 / ::	-39 dBm	2.7K	154
172.16.205.84 / ::	-40 dBm	31.3K	4.0K
172.16.205.62 / ::	-43 dBm	2.4K	119
172.16.205.57 / ::	-44 dBm	692.8K	38.0K

ภาพที่ 4: ค่า RSSI ของระบบ ACS

IP Address	RSSI	Packets to	Packets from
10.12.0.12 / ::	-54 dBm	1.6K	249
10.12.0.11 / ::	-49 dBm	6.2K	4.0K
10.12.0.10 / ::	-52 dBm	29	5

ภาพที่ 5: ค่า RSSI ของระบบ SBS

ระดับความแรงของสัญญาณเครือข่ายไร้สายที่ทั้งสองระบบติดตั้งอยู่ มีความเหมาะสมและครอบคลุม โดยระดับความแรงที่เหมาะสมมีค่าได้ไม่เกิน -75 dBm [20], [21] ซึ่งเป็นไปตามเงื่อนไข ระบบ ACS จะมีความครอบคลุมมากกว่าเนื่องจากติดตั้ง Sonoff ไว้ในห้องเรียนที่มีอุปกรณ์กระจายสัญญาณแบบไร้สายไว้สำหรับบริการแก่นักศึกษา

**4.2 การรักษาความปลอดภัยเครือข่าย**

ผลการรักษาความปลอดภัยเครือข่าย ของระบบ ACS และระบบ SBS สามารถสรุปได้ดังแสดงในตารางที่ 2

ตารางที่ 2: การรักษาความปลอดภัยเครือข่ายของระบบ

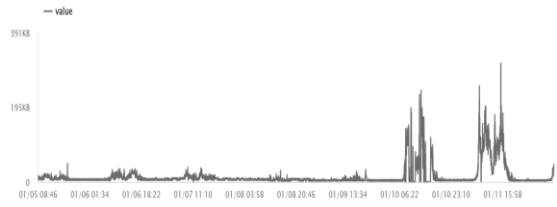
No.	System	Authentication	Encryption
1	ACS	Mac Authentication	Open
2	SBS	Open	WPA2

ระบบ ACS จะยืนยันตัวตนแบบ Mac Authentication มีการเข้ารหัสข้อมูลแบบ Open ส่วนระบบ SBS จะยืนยันตัวตนแบบ Open และมีการเข้ารหัสแบบ WPA2 ซึ่งระบบทั้ง 2 มีเงื่อนไขในการทำงานต่างกันเนื่องมาจากการเขียนโปรแกรมของระบบ แต่การยืนยันตัวตนที่นิยมใช้ และมีความปลอดภัยของเครือข่ายมากกว่า คือ การยืนยันตัวตนแบบ Mac Authentication โดยระบบ SBS ยังเป็นแบบ Open ถือว่าเป็นจุดอ่อนทางด้านนี้ ซึ่งเจ้าของโครงการจะดำเนินการแก้ไขเพิ่มเติมในอนาคต

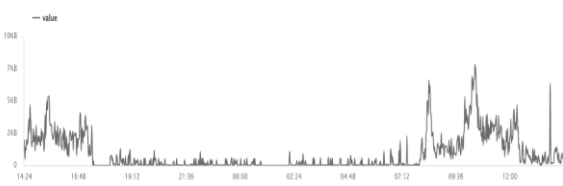
**4.3 การตรวจสอบสถานะการทำงานของระบบ**

การทำงานของระบบ ACS อาศัยอุปกรณ์ Sonoff ส่วนระบบ SBS อาศัยอุปกรณ์ NodeMCU ซึ่งอุปกรณ์ทั้ง

สองจะมีการเชื่อมต่อกับระบบเครือข่ายไร้สาย ทำให้ผู้ดูแลระบบเครือข่ายสามารถตรวจสอบการทำงานของระบบทั้งสองได้ โดยการตรวจสอบสถานะของอุปกรณ์จากระบบบริหารจัดการกลาง (Wireless Controller) และระบบ Network Monitoring System ซึ่งจะเป็นเครื่องมือในการตรวจสอบและดูแลระบบ ACS และระบบ SBS ได้ ทั้งนี้ อัตราการใช้งานข้อมูลของทั้งสองระบบควรอยู่ในระดับต่ำ เพื่อความครบถ้วนและรวดเร็วในการจัดส่งข้อมูล และไม่รบกวนระบบเครือข่ายในภาพรวม โดยอัตราการใช้งานข้อมูลเฉลี่ยของระบบ ACS ไม่เกิน 400 KB และระบบ SBS เฉลี่ยไม่เกิน 10 KB ดังตัวอย่างในภาพที่ 6 ถึงภาพที่ 7



ภาพที่ 6: ปริมาณ Traffic ของระบบ ACS รายวัน



ภาพที่ 7: ปริมาณ Traffic ของระบบ SBS รายชั่วโมง

#### 4.4 การทำงานของระบบหลังการติดตั้งและใช้งาน

ผลการใช้งานระบบ ACS จำนวน 40 จุด พบว่าระบบสามารถทำงานได้อย่างถูกต้อง แต่ยังมีอุปกรณ์ตรวจจับความเคลื่อนไหวที่มีความผิดพลาดบ้าง ผู้วิจัยได้แก้ไขโดยการกำหนดเวลาหน่วงในการปิดระบบ ACS ซึ่งมีรูปแบบการทำงาน คือ เครื่อง Server ตรวจพบตารางเรียน และมีการเคลื่อนไหว ระบบจะเปิด แต่หากไม่มีการเคลื่อนไหวภายในระยะเวลาหน่วง (5-10 นาที) ระบบจะปิด ทั้งนี้ อุปกรณ์ตรวจจับความเคลื่อนไหวจะทำงานตลอดเวลา หากภายในระยะเวลาหน่วงแล้วตรวจเจอการเคลื่อนไหว ระบบก็จะยกเลิกเวลาหน่วง ซึ่งการกำหนดค่าในรูปแบบนี้สามารถทำให้ระบบทำงานได้ดียิ่งขึ้น ไม่ประสบปัญหาเรื่องการเปิด-ปิดระบบบ่อยครั้ง ตัวอย่างรูปแบบการทำงาน of ระบบดังแสดงในภาพที่ 8

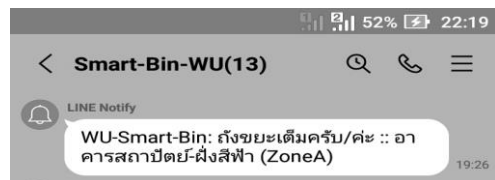
Room	Activities	Occupancy	Permission	A/C state			
5202	POS64-103	Occupied	Approved	On	Off	Off	Off
5204	1	Free	Vacant	Denied	Off	Off	
5206	1	Free	Vacant	Denied	Off	Off	
5207	2	Free	Vacant	Denied	Off	Off	Off
5211		Free	Vacant	Denied	Off	Off	Off
5212		Free	Vacant	Denied	Off	Off	
5214	CVE62-419	Vacant	Approved	Off	Off		
5216		Free	Vacant	Denied	Off	Off	Off

ภาพที่ 8: การทำงานของระบบควบคุมเครื่องปรับอากาศ

ส่วนผลการติดตั้งและใช้งานระบบ SBS ทั้ง 6 จุด พบว่าระบบทำงานได้ถูกต้องและแม่นยำ สามารถสนองตอบต่อความต้องการใช้งานได้ ทั้งนี้ตัวอย่างการทำงานของระบบดังแสดงในภาพที่ 9 ถึงภาพที่ 10 ส่วนประเด็นที่เป็นข้อควรพิจารณา คือ การแจ้งเตือนถึงขยะเต็ม นอกเวลาทำการ ซึ่งต้องรอการแจ้งเตือนในวันถัดไป



ภาพที่ 9: การจัดวางอุปกรณ์ของระบบถังขยะกิ่งอัจฉริยะ



ภาพที่ 10: ผลการทำงานของระบบถังขยะกิ่งอัจฉริยะ

### 5. สรุปผลการดำเนินการและข้อเสนอแนะ

การจัดการระบบเครือข่ายไร้สายที่เหมาะสมสำหรับระบบควบคุมเครื่องปรับอากาศและระบบถังขยะกิ่งอัจฉริยะ ประกอบด้วยการกำหนดจุดติดตั้งอุปกรณ์กระจายสัญญาณแบบไร้สายให้มีสัญญาณที่ครอบคลุมที่เหมาะสม ซึ่งระบบควบคุมเครื่องปรับอากาศมีระดับความแรงของสัญญาณอยู่ในช่วง -35 dBm ถึง -47 dBm ส่วนระบบถังขยะกิ่งอัจฉริยะมีระดับความแรงของสัญญาณอยู่ในช่วง -49 dBm ถึง -54 dBm ซึ่งมีความเหมาะสมมาก (ไม่เกิน -75 dBm) รูปแบบการรักษาความปลอดภัยของ

ระบบเครือข่าย ระบบควบคุมเครื่องปรับอากาศจะใช้การยืนยันตัวตนของระบบเครือข่ายเป็นแบบ Mac Authentication และการเข้ารหัสข้อมูลเป็นแบบ Open ส่วนระบบถังขยะกึ่งอัจฉริยะจะใช้การยืนยันตัวตนของระบบเครือข่ายเป็นแบบ Open และการเข้ารหัสแบบ WPA2 ซึ่งระบบเครือข่ายไร้สายสามารถให้บริการได้ทั้งหมด สำหรับการตรวจสอบสถานะการทำงานของระบบ สามารถตรวจสอบได้จากอุปกรณ์บริหารจัดการเครือข่ายกลาง (Wireless Controller) และระบบ Network Monitoring System โดยการทำงานของระบบจะมีอัตราการใช้งานข้อมูลเฉลี่ยของระบบควบคุมเครื่องปรับอากาศมีค่าไม่เกิน 400 KB และระบบถังขยะกึ่งอัจฉริยะมีค่าไม่เกิน 10 KB ซึ่งไม่มีผลกระทบต่อการใช้งานเครือข่ายไร้สายของมหาวิทยาลัยในภาพรวม

การยืนยันตัวตนในระบบเครือข่ายไร้สายของถังขยะกึ่งอัจฉริยะสามารถกำหนดเพิ่มเติมโดยการ ใช้ Mac Authentication ซึ่งจะทำการใช้งานระบบเครือข่ายไร้สายมีความปลอดภัยยิ่งขึ้น และควรปรับปรุงเป็นระบบถังขยะอัจฉริยะโดยสามารถแยกขยะแต่ละประเภท และแจ้งเตือนอัตโนมัติเมื่อถังขยะเต็ม ส่วนระบบ ACS ควรพิจารณาถึงอุปกรณ์ตรวจจับการเคลื่อนไหวที่ละเอียดขึ้นเพื่อช่วยให้เวลาหน่วยของการปิดระบบลดลง

### เอกสารอ้างอิง

- [1] สุขอนันต์ ชูกระชั้น. “แนวทางการพัฒนามหาวิทยาลัยราชภัฏภูเก็ตเป็นมหาวิทยาลัยสีเขียว.” มหาวิทยาลัยราชภัฏภูเก็ต, 2561.
- [2] อิศรี รอดทัศนาศ. “มหาวิทยาลัยสีเขียว.” *วารสาร มจร. วิชาการ*, ปีที่ 18, ฉบับที่ 36, หน้า 171-188, ม.ค.-มิ.ย., 2558.
- [3] UI GreenMetric. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <https://building.rmutr.ac.th/ui-green-metric/>
- [4] Green WU. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <https://green.wu.ac.th/>
- [5] UI GreenMetric. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <https://greenmetric.ui.ac.id>
- [6] จิตติกรณ์ ยาวินไชย จารึกศิลป์. “กลยุทธ์การบริหารสถาบันอุดมศึกษาตามแนวคิดการเป็นมหาวิทยาลัยสีเขียว.” *[วิทยานิพนธ์]*. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย, 2559.
- [7] การรักษาความปลอดภัยของเครือข่ายไร้สาย. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <https://sites.google.com/site/nutthawoot051b/kar-seth-security/kar-raksa-khwam-plxdphay>
- [8] การเข้ารหัสข้อมูล. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <https://ez-genius.com/เลือกการเข้ารหัส-wi-fi/>
- [9] มหาวิทยาลัยสุรนารี. ศูนย์เทคโนโลยีดิจิทัล. “การปรับปรุงระบบเครือข่ายไร้สายมหาวิทยาลัยให้มีความปลอดภัยตามมาตรฐาน IEEE802.1x.” [คู่มือการจัดการองค์ความรู้ ประจำปี 2560]. นครราชสีมา: มหาวิทยาลัยสุรนารี, 2560.
- [10] การยืนยันตัวตน. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <https://www.techtalkthai.com/4-basic-network-authentication-technologies/>
- [11] ESP8266. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <https://www.allnewstep.com/article/30/nodemcu-esp8266-esp8285-arduino-1-esp8266-คือ>
- [12] ชนาธิป อาจสันเทียะ และจิระศักดิ์ นำประดิษฐ์. “ระบบแจ้งเตือนความปลอดภัยภายในบ้านผ่านแอปพลิเคชันไลน์ โดยใช้แนวคิดอินเทอร์เน็ตออฟติงส์.” *การประชุม NCCIT 2018 ครั้งที่ 15*, ประเทศไทย, หน้า 226-231
- [13] Sonoff. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <http://www.raspberrypi.com/product/2>
- [14] Sonoff. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <https://www.unzeen.com/article/3407/>
- [15] Sonoff. [สืบค้นเมื่อวันที่ 1 มกราคม 2565]. จาก <https://www.blogsdit.com/2019/06/sonoff.html>
- [16] พัศกรพิไล หน่อคำ. “ประสิทธิผลของระบบเครือข่ายอินเทอร์เน็ตแบบไร้สายของเทศบาลตำบลหนองหอย อำเภอเมืองเชียงใหม่ จังหวัดเชียงใหม่” *[การค้นคว้าอิสระ]*. เชียงใหม่: มหาวิทยาลัยเชียงใหม่, 2561.
- [17] ศดานนท์ วัฒนธรรม. กรรณิกา ศรีวรรณไพศาล. ปิยกุล บุญญาศรีรัตน์. “แนวทางการจัดบริการระบบเครือข่ายไร้สายสาธารณะของเทศบาลนครสงขลา.” *[รายงานผลการวิจัย]*. สงขลา: มหาวิทยาลัยราชภัฏสงขลา, 2560.
- [18] กรณิการ์ จันทร์แย้ม. “การประเมินคุณภาพเครือข่ายไร้สายกรณีศึกษา มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.” *[สารนิพนธ์]*. กรุงเทพมหานคร: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2559.
- [19] พลธนธรณ์ กรยุทธพิพัฒน์. “ปัจจัยที่มีผลต่อการใช้ระบบเครือข่ายไร้สายของนักศึกษามหาวิทยาลัยรามคำแหงในกรุงเทพมหานคร.” *[วิทยานิพนธ์]*. กรุงเทพมหานคร: มหาวิทยาลัยรามคำแหง, 2558.
- [20] สาธิต ขวัญชุม. สุริยะ เมืองสุวรรณ. “การศึกษารูปแบบที่เหมาะสมสำหรับติดตั้งอุปกรณ์กระจายสัญญาณแบบไร้สายชนิดภายนอกอาคาร กรณีศึกษา กลุ่มอาคารบ้านพักบุคลากร

วลัยนิवास 7-10 มหาวิทยาลัยวลัยลักษณ์.” การประชุม  
วิชาการระดับชาติวลัยลักษณ์วิจัย ครั้งที่ 9, นครศรีธรรมราช,  
ประเทศไทย, วันที่ 31 มีนาคม 2560.

- [21] สาริต ขวัญชุม. และ สุริยะ เมืองสุวรรณ. “การเพิ่ม  
ประสิทธิภาพการให้บริการระบบเครือข่ายไร้สาย กรณีศึกษา  
กลุ่มอาคารหอพักนักศึกษา 16-18 มหาวิทยาลัยวลัยลักษณ์.”  
การประชุมวิชาการระดับชาติและนานาชาติ ราชภัฏสุราษฎร์  
ธานีวิจัย, ครั้งที่ 12, สุราษฎร์ธานี, หน้า 27 - 28, ตุลาคม,  
2559.

# บริการซอฟต์แวร์คลาวด์การรู้จำหลายใบหน้าพร้อมกันสำหรับการตรวจสอบการเข้าชั้นเรียน

## Simultaneous Multi-facial Recognition Software-as-a-Service for Class Attendance Checking

อริษา ไวยเจียรนัย (Arisa Vaichiaranai)<sup>1</sup> และ เทพฤทธิ์ บัณฑิตวัฒนาวงศ์ (Thepparit Banditwattanawong)<sup>2</sup>

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

<sup>1</sup>arisa.v@ku.th, <sup>2</sup>fscitrb@ku.ac.th

### บทคัดย่อ

การตรวจสอบรายชื่อและจำนวนผู้เข้าชั้นเรียนทำให้เวลาเรียนส่วนหนึ่งสูญเสียไปโดยเฉพาะชั้นเรียนที่มีผู้เรียนจำนวนมาก บทความนี้จึงเสนอการตรวจสอบรายชื่อผู้เรียนในรูปแบบบริการซอฟต์แวร์คลาวด์ด้วยการถ่ายรูปหมู่ของผู้เรียนแทนการถ่ายภาพทีละใบหน้าโดยทำงานผ่านโปรแกรมประยุกต์เว็บที่มีการจัดเก็บและประมวลผลข้อมูลบนคลาวด์ จากการทดลองใช้งานโดยผู้ใช้งาน 30 คน ได้แก่ผู้สอนจำนวน 2 คนและผู้เรียนจำนวน 28 คนพบว่า ระบบมีค่าความเที่ยงเท่ากับร้อยละ 73.33 ค่ารีคอลล์เท่ากับร้อยละ 97.77 ค่าเอฟเมซเซอร์เท่ากับร้อยละ 83.80 และมีค่าเฉลี่ยความพึงพอใจต่อภาพรวมระบบเท่ากับ 4.58

**คำสำคัญ:** การตรวจจับใบหน้า การรู้จำใบหน้า การคำนวณแบบคลาวด์

### Abstract

Checking the list and the number of attendees can waste some time, especially in a class with a large number of students. This paper proposes a cloud software service for checking student lists by taking a group photo of the students instead of one face at a time through a web application that stores and processes data on cloud. Evaluation results by 30 users, including 2 teachers and 28 students showed that precision, recall, and f-measure scores were 73.33%, 97.77%, and 83.80%, respectively, while the overall satisfaction score was 4.58.

**Keywords:** Face Detection, Face Recognition, Cloud Computing.

### 1. บทนำ

#### 1.1 ที่มาและปัญหา

ปัจจุบันเป็นยุคที่เร่งด่วนในการใช้ชีวิตประจำวันสำหรับประชากรทั่วโลก จึงทำให้มีข้อจำกัดในการทำกิจกรรมประจำวันในหลายๆ ด้าน รวมถึงด้านการศึกษา ในประเทศไทย การศึกษาในแต่ละวันมีระยะเวลาค่อนข้างจำกัด อุปสรรคอย่างหนึ่งให้เห็นผลชัดเจนที่ส่งผลกระทบต่อการเรียนการสอนในเวลานี้ คือระยะเวลาในการเรียนการสอนไม่เพียงพอต่อเนื้อหาและบทเรียนตามหลักสูตร เนื่องจากจำนวนผู้เรียนต่อหนึ่งห้องเรียนมีปริมาณค่อนข้างมากทำให้อาจารย์ผู้สอนใช้เวลาในการตรวจสอบรายชื่อและจำนวนผู้เข้าเรียนค่อนข้างนานจึงทำให้สูญเสียเวลาและมีเวลาในการเรียนการสอนไม่เต็มที่

ในยุคที่เทคโนโลยีเข้ามามีบทบาทสำคัญในการช่วยการจัดการศึกษาให้บรรลุอุดมการณ์ทางการศึกษาตามนโยบายการจัดการศึกษาของรัฐ เครื่องมือที่มีเทคโนโลยีสมัยใหม่ต่างๆ ได้เข้ามามีบทบาทและถูกนำมาประยุกต์ใช้ในการอำนวยความสะดวกในหลายๆ ด้าน จากการทบทวนวรรณกรรมที่ผ่านมาพบว่าเทคโนโลยีและโปรแกรมประยุกต์ที่สามารถช่วยในการตรวจสอบรายชื่อและจำนวนผู้เข้าเรียนผ่านการถ่ายรูปหน้าของผู้เรียนเพื่อลงชื่อคนเข้าเรียน โดยผู้เรียนเป็นผู้ดำเนินการถ่ายรูปหน้าของตนเองเข้าไปในระบบ หรืออีกหนึ่งเทคโนโลยีที่เป็นที่แพร่หลายในปัจจุบันคือการตั้งกล้องในจุดที่ผู้เรียนเดินผ่านก่อนเข้าสู่การเรียนการสอน แต่พบว่าเทคโนโลยีที่กล่าวมานั้นมีจุดอ่อน เช่น ต้องใช้พื้นที่ในการตั้งกล้องและต้นทุนที่สูง อีกทั้งยังพบว่าโปรแกรมประยุกต์เหล่านั้นในบางครั้งยังขาดประสิทธิภาพในการรู้จำใบหน้าได้อย่าง

ถูกต้องและเป็นเหตุให้ผู้เรียนขาดเรียน โดยไม่รู้ตัว เนื่องจากให้ผู้เรียนไม่สามารถตรวจสอบหรือแจ้งแก้ไขข้อมูลได้เมื่อเกิดข้อผิดพลาด

## 1.2 วัตถุประสงค์

เพื่อพัฒนาโปรแกรมประยุกต์การรู้จำหลายใบหน้า พร้อมกันสำหรับการตรวจสอบการเข้าชั้นเรียนในรูปแบบบริการซอฟต์แวร์คลาวด์

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การคำนวณแบบคลาวด์ (Cloud Computing) [1] : บริการซอฟต์แวร์คลาวด์ (Software as a Service : SaaS) คือบริการ โปรแกรมประยุกต์ของผู้ให้บริการที่ดำเนินงานอยู่บน โครงสร้างพื้นฐานคลาวด์ (cloud infrastructure) และสามารถเข้าถึงได้จากอุปกรณ์ ผู้รับบริการที่หลากหลายผ่านส่วนต่อประสานเครื่องรับบริการแบบบาง (thin client interface) เช่น เว็บเบราว์เซอร์ หรือส่วนต่อประสานโปรแกรม (program interface) ผู้ใช้งานไม่ต้องจัดการหรือควบคุมโครงสร้างพื้นฐานคลาวด์เบื้องหลังซึ่งรวมถึงเครือข่าย เครื่องบริการ ระบบปฏิบัติการและหน่วยเก็บข้อมูล ผู้ใช้บริการไม่จำเป็นต้องลงทุนเพิ่มเติมในส่วนอุปกรณ์ รวมไปถึงค่าใช้จ่ายในการบำรุงรักษาระบบ เพราะผู้ให้บริการจะเป็นผู้ดูแลระบบทั้งหมดทำให้สามารถควบคุมค่าใช้จ่ายได้ดี ทั้งนี้ผู้ใช้งานสามารถใช้งานได้ทุกที่ ทุกเวลาเพียงแค่เชื่อมต่ออินเทอร์เน็ต

สาเหตุที่เลือกใช้บริการรูปแบบนี้เนื่องจากบริการซอฟต์แวร์คลาวด์ใช้ต้นทุนต่ำและสามารถควบคุมค่าใช้จ่ายได้ดี ไม่ต้องเสียค่าใช้จ่ายไปกับการติดตั้ง บำรุงรักษาอุปกรณ์ และสามารถใช้งานได้ทุกที่ ทุกเวลาเพียงแค่เชื่อมต่ออินเทอร์เน็ต

2.1.2 การตรวจจับใบหน้า (Face Detection) [2] คือการค้นหาใบหน้าของบุคคลจากภาพหรือวิดีโอเพื่อทำการประมวลผลภาพใบหน้าเพื่อวัตถุประสงค์ต่างๆ เช่น เพื่อการจำแนก

2.1.3 การรู้จำใบหน้า (Face Recognition) [2] คือการนำภาพใบหน้าที่ตรวจจับได้มาเปรียบเทียบกับใบหน้าในฐานข้อมูลเพื่อระบุว่าใบหน้าที่ตรวจจับได้ตรงกับบุคคลใดแล้วจึงนำผลลัพธ์ที่ได้ส่งไปประมวลผลอื่นต่อไป หลักการทำงานที่สำคัญของเทคโนโลยีการรู้จำใบหน้าที่คือการสร้างตัวแบบการอ้างอิงที่เรียกว่า เฟซพริ้น (Face print) ขึ้นมาจากลักษณะเฉพาะต่างๆ บนใบหน้า เช่น โครงหน้า ความกว้างของจมูก ระยะห่างระหว่างตาทั้งสองข้าง ขนาดของ โหนกแก้ม ความลึกของเบ้าตา รวมถึงพื้นผิวบนใบหน้า (Facial texture) จากนั้นทำการสร้างจุดเชื่อมโยงบนใบหน้า (Nodal points) เพื่อเปรียบเทียบกับรูปภาพที่ถูกเก็บไว้ในฐานข้อมูล (Database) ทั้งในลักษณะภาพนิ่ง และภาพเคลื่อนไหว

2.1.4 เทคโนโลยีตรวจจับใบหน้า Faceapi.js [3] คือ ส่วนต่อประสาน โปรแกรมประยุกต์ (application program interface) ของจาวาสคริปต์สำหรับงานตรวจจับใบหน้าคนในรูปภาพใครเป็นใคร ทำหน้าที่ช่วยในเรื่อง การแสดงผลบนเว็บไซต์โดยทำงานภายใต้เบราว์เซอร์ และ สำหรับการตรวจจับใบหน้าในการศึกษานี้เลือกใช้วิธีการประมวลผลแบบโมบายเน็ตรุ่น 1 (SSD Mobilenet V1)

2.1.5 การประมวลผลแบบโมบายเน็ต (Mobilenet) [4] ตัวแบบขนาดเล็ก ที่ทำงานได้เร็ว ใช้พลังงานในการประมวลผลไม่มาก ถูกออกแบบมาสำหรับงานที่มีทรัพยากรจำกัด สามารถใช้งานได้ทั้ง การจดหมวดหมู่ การตรวจจับ การฝัง และการแบ่งกลุ่ม

### 2.2 งานวิจัยที่เกี่ยวข้อง

ระบบสำหรับบันทึกข้อมูลการเข้างานของพนักงานด้วยเทคโนโลยีรู้จำใบหน้า [5] ทำการระบุตัวบุคคลได้แบบทันทีทันใดระหว่างที่พนักงานเดินผ่านกล้องวงจรปิดและบันทึกการเข้างานของบุคคลนั้น การทำงานเริ่มต้นเมื่อกระทำการตรวจจับใบหน้า จากกล้องวงจรปิด และนำรูปภาพใบหน้าไปผ่านกระบวนการฮิสโตแกรมอีควอลไลซ์ (Histogram equalization) เพื่อใช้ โอ เก น เฟ ซ (Eigenface recognition) สำหรับการระบุตัวตน สำหรับการแจ้งเตือน ระบบมีบริการส่งข้อความสั้นๆ ไปยัง

โทรศัพท์เคลื่อนที่ โดยได้ผลการวิจัยมีความถูกต้องในการระบุตัวตน 68%

ระบบตรวจสอบนักศึกษาเข้าเรียนด้วยเทคนิคการรู้จำใบหน้า [6] สามารถตรวจสอบนักศึกษาระหว่างที่กำลังเรียนในห้องเรียน โดยติดตั้งกล้อง 2 ตัว ตัวที่ 1 ติดไว้ที่เพดานและตัวที่ 2 ติดไว้บริเวณหน้าห้องเรียน ภาพที่ได้มาจากกล้องตัวที่ 1 ที่ได้ทำการติดตั้งบนเพดาน ระบบจะนำมาใช้ในการระบุตำแหน่งที่นักศึกษานั่งอยู่ในห้องเพื่อมาวิเคราะห์เพื่อปรับตำแหน่งและมุมกล้องของกล้องตัวที่ 2 ไปหาผู้เรียน และเมื่อผู้เรียนหันมองมาที่หน้าห้อง กล้องจะทำการบันทึกภาพนั้นมาใช้ในการระบุตัวตนต่อไป แต่จากผลการวิจัยเกิดความคลาดเคลื่อนในการตรวจสอบหาความใกล้เคียงตำแหน่งที่นั่ง จึงทำให้เกิดเป็นผลกระทบต่อกระบวนการบันทึกภาพมาใช้ในการระบุตัวตน

ระบบสำหรับตรวจเช็คผู้เรียนที่เข้าเรียน [7] ใช้เทคนิครู้จำใบหน้าในการยืนยันการระบุตัวตนผู้เรียน โดยมีการเลือกใช้ขั้นตอนวิธีวิโอลาโจนส์ (Viola-Jones) จากนั้นทำการแปลงภาพให้เป็นระดับสีเทาและทำการฮิสโตแกรมอีควอไลซ์ ซึ่งเทคนิคนี้จะช่วยปรับปรุงความคมชัดของภาพเมื่อถูกยัดออก โดยการทำให้ช่วงของความเข้มในภาพมากขึ้น และจากนั้นทำการปรับขนาดภาพให้มีขนาด 100 x100 พิกเซล จากนั้นนำภาพดังกล่าวไปใช้ในกระบวนการจัดกลุ่มเพื่อใช้ในการระบุตัวตน โดยมีเทคนิค 3 เทคนิคให้เลือกใช้โปรแกรม เมื่อเสร็จสิ้นคาบเรียนจะมีการส่งสรุปผลการเข้าเรียนไปยังอีเมลของผู้สอนประจำวิชา งานวิจัยนี้มีปัญหาเรื่อง ความถูกต้องเมื่อนำไปใช้ในห้องที่มีปริมาณแสงไม่เท่ากันกับในภาพสำหรับการรู้จำและยังไม่มียุทธศาสตร์สำหรับการแก้ไขผลการระบุตัวตน เนื่องจากมีการส่งผลสรุปไปทางอีเมลเมื่อจบคาบเรียนเท่านั้น

### 3. วิธีการดำเนินการวิจัย

#### 3.1 เครื่องมือที่ใช้ในการพัฒนาและจัดเก็บข้อมูล

เครื่องมือที่ใช้ในการศึกษาครั้งนี้ประกอบด้วยเครื่องมือที่ใช้ในการพัฒนา และเครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลดังรายละเอียดดังต่อไปนี้

##### 3.1.1 ส่วนอุปกรณ์ (Hardware)

- คอมพิวเตอร์พกพา สำหรับพัฒนาระบบ
- โทรศัพท์เคลื่อนที่ สำหรับใช้ในการถ่ายภาพและทดสอบระบบ

##### 3.1.2 ส่วนชุดคำสั่ง (Software)

- โปรแกรมระบบจัดการฐานข้อมูลมองโก (MongoDB Atlas Database : MongoDB) เป็นฐานข้อมูลในการพัฒนาโปรแกรมประยุกต์
- โหนดเจเอส (Node.js) ในการเขียนโปรแกรมเพื่อการพัฒนาโปรแกรมประยุกต์
- เว็บเบราว์เซอร์ ได้แก่ โครม (Chrome), ไฟร์ฟ็อกซ์ (Firefox) และซาฟารี (Safari) ในการศึกษาได้ใช้ในการเปิดหน้าจอผลตัวระบบ
- เทคโนโลยีตรวจจับใบหน้า Faceapi. ในตรวจจับใบหน้าคนในรูปภาพ และสำหรับการตรวจจับใบหน้าในการศึกษานี้เลือกใช้วิธีการประมวลผลแบบโมบายเน็ตรุ่น 1 (SSD Mobilenet V1)

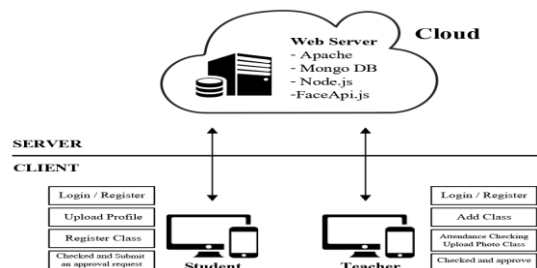
##### 3.1.3 บริการคลาวด์

- แรม 2 กิกะไบต์
- โซลิตัสเตดไครฟ์ 40 กิกะไบต์
- แบนด์วิดท์ 2 เทราไบต์

3.1.4 กูเกิ้ลฟอร์ม (Google form) เพื่อจัดทำผลสำรวจออนไลน์เพื่อเพื่อวัดประสิทธิภาพ ประสิทธิผลของการทำงานของระบบ และสอบถามเพื่อประเมินความพึงพอใจจากผู้ทดสอบใช้งานระบบ

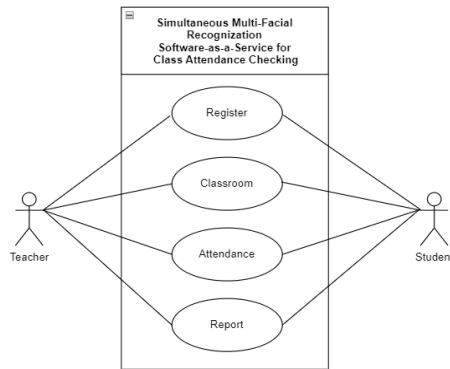
### 3.2 การวิเคราะห์ออกแบระบบ

โปรแกรมประยุกต์การศึกษานี้ มีการให้บริการในรูปแบบบริการซอฟต์แวร์คลาวด์ ดังภาพสถาปัตยกรรม ดังภาพที่ 1 และมีการแบ่งประเภทผู้ใช้งานออกเป็น 2 ส่วนหลัก ได้แก่ ผู้เรียน และผู้สอน โดยแต่ละส่วนมีการใช้งานดังภาพที่ 2



ภาพที่ 1: สถาปัตยกรรมระบบ





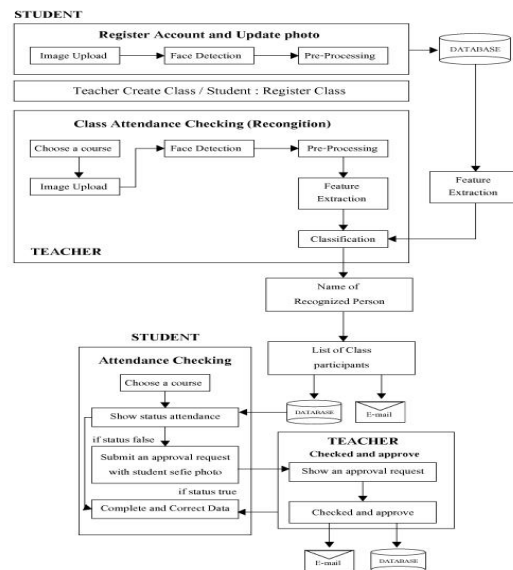
ภาพที่ 2: แผนภาพยูสเคส (Use Case Diagram)

ขั้นตอนการทำงานของระบบเริ่มต้นจากผู้ใช้งานทำการลงทะเบียนและอนุญาตให้ความยินยอมในการเข้าถึงข้อมูลตาม พ.ร.บ.คุ้มครองข้อมูลส่วนบุคคลและเข้าสู่ระบบด้วยบัญชีผู้ใช้งานที่ทำการสมัครไว้ จากนั้นผู้เรียนทำการเพิ่มรูปภาพของตนเองและลงทะเบียนในรายวิชาที่ผู้สอนทำการสร้างไว้

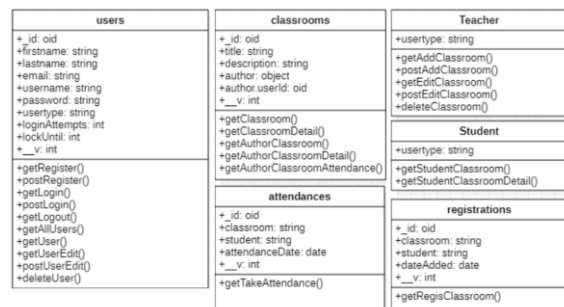
กระบวนการตรวจสอบเช็คชื่อการเข้าเรียนทำโดยผู้สอน ผู้สอนจะทำการถ่ายภาพและอัปโหลดภาพลงในรายวิชาที่ตนสอน การอัปโหลดภาพสามารถทำได้ไม่จำกัดจำนวนครั้งเพื่อให้รองรับในกรณีที่ห้องเรียนมีขนาดใหญ่ ผู้สอนต้องถ่ายภาพที่คมชัดและเห็นใบหน้าชัดจนเพื่อให้ได้ภาพที่คมชัดและเหมาะสมต่อการทำงานของระบบและเกิดข้อผิดพลาดน้อยที่สุด ระบบจะทำการตรวจสอบข้อมูลใบหน้าที่ถูกลงทะเบียนในฐานะข้อมูล เพื่อมาเปรียบเทียบกับใบหน้าของผู้สอนทำการอัปโหลดและทำการระบุตัวตนของผู้เรียนและบันทึกฐานข้อมูล ซึ่งในแต่ละครั้งระบบจะมีอีเมลแจ้งผลการเข้าเรียนไปยังอีเมลของผู้เรียนทุกครั้ง ทำให้ผู้เรียนทราบผลและหากเกิดข้อผิดพลาด ผู้เรียนสามารถแก้ไขปัญหาได้ทันที

ในกรณีที่ผู้เรียนพบว่าตนเองขาดเรียน เนื่องจากระบบเกิดข้อผิดพลาด ผู้เรียนสามารถทำการอัปโหลดภาพของตนเองเข้าสู่ระบบเพื่อส่งให้ผู้สอนอนุมัติ หรือทำการตรวจสอบใบหน้าของตนเองผ่านภาพถ่ายของผู้สอนและทำการระบุยืนยันใบหน้าตนเองลงไป จากนั้นส่งคำขออนุมัติไปยังผู้สอนเพื่อให้ผู้สอนอนุมัติ เมื่อได้รับผลอนุมัติระบบจะมีการข้อความอัตโนมัติแจ้งเตือนไปยังอีเมล ซึ่งขั้นตอนการทำงานของระบบมีภาพแนวคิดการวิจัยดังภาพที่ 3 มีภาพแผนภาพคลาซของระบบดังภาพที่ 4 และ

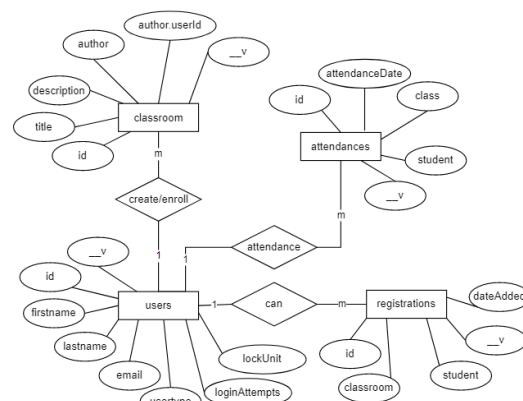
แผนภาพอีอาร์ของระบบดังภาพที่ 5 หมายเหตุ \_\_v คือ จำนวนครั้งของการจัดการระเบียบในตารางดังกล่าว



ภาพที่ 3: กรอบแนวคิดการวิจัย

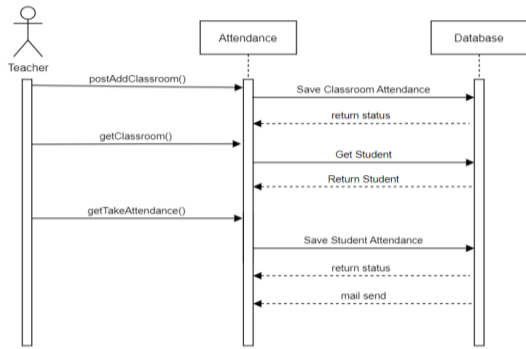


ภาพที่ 4: แผนภาพคลาซของระบบ

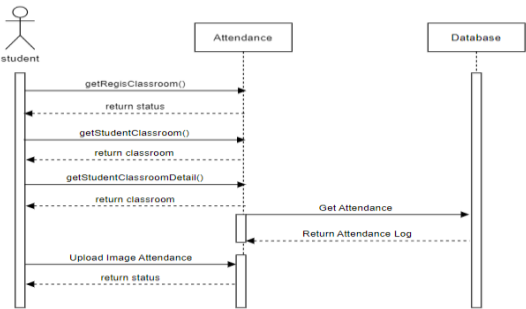


ภาพที่ 5: แผนภาพอีอาร์ของระบบ

ระบบมีแผนภาพลำดับการทำงานของผู้เรียนดังภาพที่ 6 และแผนภาพลำดับการทำงานของผู้สอนดังภาพที่ 7



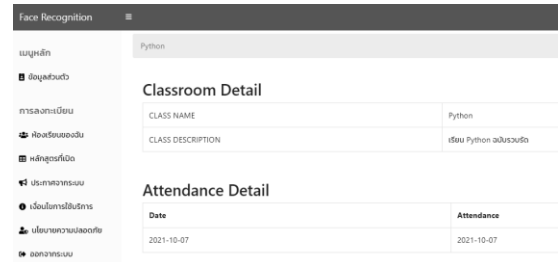
ภาพที่ 6: แผนภาพลำดับการสร้างรายวิชา การตรวจสอบรายชื่อผู้ลงทะเบียน และการตรวจสอบรายชื่อการเข้าชั้นเรียน



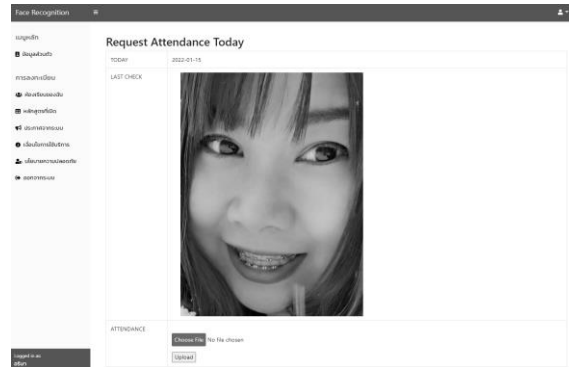
ภาพที่ 7: แผนภาพลำดับการลงทะเบียนรายวิชา การตรวจสอบผลการลงทะเบียนเข้าชั้นเรียนและการอัปโหลดภาพของตนเองเข้าสู่ระบบเพื่อส่งให้ผู้สอนอนุมัติ

### 3.3 การสร้างโปรแกรมประยุกต์

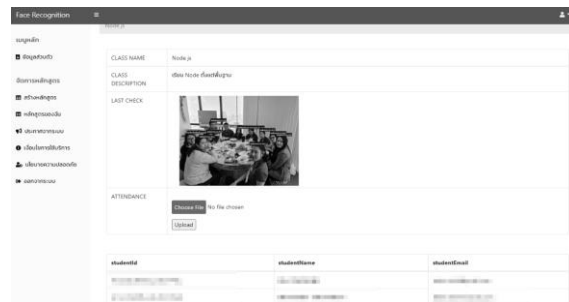
ในการสร้างโปรแกรมประยุกต์ได้เลือกใช้โปรแกรมระบบจัดการฐานข้อมูลมอองโก (MongoDB Atlas Database : MongoDB) เป็นฐานข้อมูล ใช้โหนดเจเอส (Node.js) ในการเขียนโปรแกรมเพื่อการพัฒนาโปรแกรมประยุกต์ ใช้เฟรมเวิร์กทูลสแตปในการพัฒนาระบบในส่วนการแสดงผลให้กับผู้ใช้ ให้ระบบสามารถแสดงผลและรองรับได้บนทุกอุปกรณ์ และใช้เทคโนโลยีตรวจจับใบหน้า Faceapi.js ของจาวาสคริปต์สำหรับงานตรวจจับใบหน้าคน ในรูปภาพและทำหน้าที่ช่วยในเรื่องการแสดงผลบนเว็บไซต์โดยทำงานภายใต้เบราว์เซอร์ ในส่วนสำหรับการตรวจจับใบหน้าในการศึกษานี้เลือกใช้วิธีการประมวลผลแบบโมบายเน็ตรุ่น 1 และทำการเปิดหน้าจอแสดงผลตัวระบบโดยเว็บเบราว์เซอร์ ได้แก่ โครม, ไฟร์ฟ็อกซ์ และซาฟารี ตัวอย่างหน้าจอระบบดังภาพที่ 8,9,10 และ 11



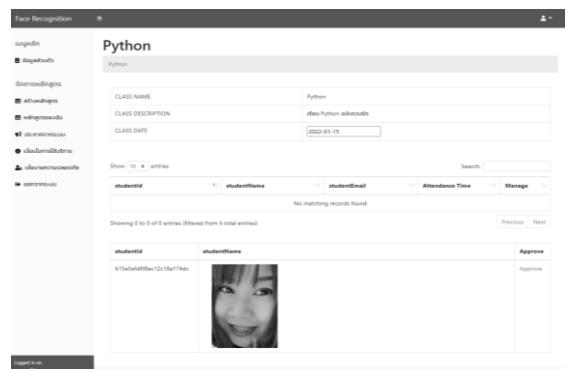
ภาพที่ 8: หน้าจอตรวจสอบประวัติการเข้าเรียนของผู้เรียน



ภาพที่ 9: หน้าจอการส่งคำขออนุมัติเข้าเรียนของผู้เรียน



ภาพที่ 10: หน้าจอการเช็คชื่อผู้เรียนสำหรับผู้สอน



ภาพที่ 11: หน้าจออนุมัติผลการเข้าเรียนสำหรับผู้สอน

ภาพที่สามารถตรวจจับใบหน้าและมีใบหน้าตรงกับภาพใบหน้าที่อยู่ในฐานข้อมูล จะปรากฏสี่เหลี่ยมบริเวณใบหน้า ดังภาพที่ 12 (ซ้าย) และ ภาพที่ 12 (ขวา) คือ ตัวอย่างใบหน้าที่อยู่ในฐานข้อมูล



ภาพที่ 12: ตัวอย่างภาพหน้าจอที่ตรงกับภาพหน้าจอที่อยู่ในฐานข้อมูล

3.4 การทดลองและการทดสอบประสิทธิภาพ

3.4.1 ขั้นตอนการทดลองและทดสอบประสิทธิภาพ

การทดสอบการใช้งานระบบ ใช้ผู้ทดสอบระบบจำนวน 30 คนซึ่งแบ่งผู้ทดสอบออกเป็น 2 ประเภทได้แก่ ผู้เรียนจำนวน 28 คน และผู้สอนจำนวน 2 คน โดยให้ผู้ทดสอบทุกคนเข้ามาทำการลงทะเบียน หลังจากการลงทะเบียนเสร็จสิ้น ผู้สอนคนที่ 1 ต้องทำการสร้างรายวิชา ส่วนผู้เรียนจำนวน 28 คน ต้องทำการอัปโหลดภาพถ่ายของตนจำนวน 1-5 ภาพและลงทะเบียนในรายวิชาที่กำหนด ส่วนผู้สอนคนที่ 2 ทำการทดสอบการใช้งานในส่วนของผู้สอนว่าเห็นข้อมูลวิชาเรียนและรายชื่อผู้ลงทะเบียนในวิชาของผู้สอนอีกคนหรือไม่

ถัดมาผู้สอนจะทำการทดสอบการเช็คชื่อจำนวน 3 ครั้ง โดยในแต่ละครั้งของการทดสอบ กำหนดให้มีภาพที่อัปโหลดต้องมีใบหน้าของผู้ทดสอบจริงจำนวนคน 20 คน

เมื่อการทดสอบเช็คชื่อเสร็จสิ้น ให้ผู้เรียนทั้ง 28 คน ตรวจสอบสถานะการเข้าร่วมของตนเองผ่านอีเมลหรือหน้าตรวจสอบรายงานของระบบ หากพบว่าผลในการเช็คชื่อของตนผิดพลาด ให้เข้าสู่ระบบเพื่อทดสอบการใช้เมนูส่งคำขออนุมัติโดยผู้สอนต่อไป

3.4.2 ขั้นตอนการหาความพึงพอใจ

ผู้วิจัยทำการส่งแบบฟอร์มสำรวจออนไลน์ ที่สร้างโดยกูเกิ้ลฟอร์ม (Google form) เพื่อสอบถามและประเมินผลความพึงพอใจ ข้อเสนอแนะ พร้อมระบุปัญหาการใช้งาน และประเภทการทดสอบของตนเอง จากกลุ่มผู้ทดสอบกลุ่มข้างต้น

4. ผลการดำเนินการวิจัย

ในการทดสอบประสิทธิภาพการใช้งานระบบ ผู้วิจัยใช้กลุ่มตัวอย่างจำนวน 30 คน ได้แก่ผู้สอนจำนวน 2 คนและผู้เรียนจำนวน 28 คนและจำนวนการทดสอบ 3 รอบ

จากการทดสอบแต่ละครั้งกำหนดให้มีภาพของใบหน้าของผู้ทดสอบจริงจำนวนคน 20 คน กำหนดแบ่งกลุ่มสถานะการทดสอบออกเป็น 4 ประเภทดังนี้

1. TP ผู้ที่เข้าเรียน ระบบแสดงสถานะเข้าเรียน
2. TN ผู้ที่ไม่เข้าเรียน ระบบแสดงสถานะไม่เข้าเรียน
3. FP ผู้ที่เข้าเรียน ระบบแสดงสถานะไม่เข้าเรียน
4. FN ผู้ที่ไม่เข้าเรียน ระบบแสดงสถานะเข้าเรียน

จากการทดสอบสามารถทำการสรุปผลการทดลองทั้งหมดที่ได้ดังตารางที่ 1, 2, 3 และ 4

ตารางที่ 1: เมทริกซ์ความสับสน (Confusion Matrix)

		ค่าผลลัพธ์จริง					
		การทดสอบรอบที่ 1		การทดสอบรอบที่ 2		การทดสอบรอบที่ 3	
ผลการทดสอบ		Positive	Negative	Positive	Negative	Positive	Negative
	Positive	15	8	14	7	15	8
Negative	0	5	1	6	0	5	

ตารางที่ 2: ผลการวัดความเที่ยงของข้อมูล (Precision)

	การทดสอบรอบที่ 1	การทดสอบครั้งที่ 2	การทดสอบครั้งที่ 3
ค่าความแม่นยำ (%)	0.75	0.7	0.75

ตารางที่ 3: ผลการวัดค่าความแม่นยำ (Accuracy)

	การทดสอบรอบที่ 1	การทดสอบครั้งที่ 2	การทดสอบครั้งที่ 3
ค่าความถูกต้อง (%)	0.82	0.75	0.82

ตารางที่ 4: ผลเอฟเมซเซอร์ (F-measure)

	การทดสอบรอบที่ 1	การทดสอบครั้งที่ 2	การทดสอบครั้งที่ 3
เอฟเมซเซอร์ (%)	85.71	80	85.71

ผลค่าเฉลี่ยผลการวัดค่าเฉลี่ยความพึงพอใจของผู้ใช้จากผู้ทดสอบใช้งานระบบจำนวน 30 คน โดยมีคะแนนเต็มสำหรับเกณฑ์แต่ละข้อเท่ากับ 5 โดยมีรายละเอียดของผลการประเมินดังตารางที่ 5

ตารางที่ 5: ผลการวัดค่าเฉลี่ยความพึงพอใจของผู้ใช้

เกณฑ์การประเมิน	ค่าเฉลี่ย ( $\bar{x}$ )	ค่าเฉลี่ยของส่วนเบี่ยงเบนมาตรฐาน (S.D.)
1. ความถูกต้องและน่าเชื่อถือของระบบ	4.67	0.55
2. ความง่ายในการใช้ระบบ		
2.1 ด้านการออกแบบ	4.77	0.57
2.2 ด้านการใช้งาน	4.47	0.57
3. ความรวดเร็วในการทำงานของระบบ	4.23	0.50
4. สิ่งที่คาดหวังและประโยชน์ที่ได้รับ	4.87	0.35
5. ความพึงพอใจต่อภาพรวมต่อระบบ	4.50	0.51
ค่าเฉลี่ย	4.58	0.51

## 5. สรุปผลการศึกษาและข้อเสนอแนะ

ผลประเมินโปรแกรมประยุกต์ในรูปแบบของบริการซอฟต์แวร์คลาวด์โดยผู้ใช้งานจริงพบว่าโปรแกรมประยุกต์จากการศึกษานี้สามารถใช้งานได้จริง และช่วยอำนวยความสะดวกได้อย่างมีประสิทธิภาพ มีความรวดเร็วในการทำงานของระบบในเกณฑ์ที่เหมาะสม และสามารถนำมาประยุกต์ต่อการใช้งานหลายในหลากหลายรูปแบบไม่เพียงแต่การใช้สำหรับการตรวจสอบรายชื่อผู้เข้าเรียนเท่านั้น ยังสามารถใช้ประยุกต์เข้ากับการตรวจสอบรายชื่อผู้เข้าร่วมประชุมหรือรวมถึงการประชุมออนไลน์ผ่านทางช่องทางออนไลน์ต่างๆ ได้

เนื่องจากปัจจุบันเป็นยุคที่มีสถานการณ์แพร่ระบาดของโรคติดเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ 2019 ทำให้การใช้ชีวิตของคนในปัจจุบันต้องสวมใส่หน้ากากอนามัยตลอดเวลา และผลกระทบการทำงานของตัวระบบจากการศึกษานี้จำเป็นต้องใช้รูปภาพที่มีส่วนประกอบของใบหน้าครบถ้วน ได้แก่ คิ้ว ตา จมูก และปาก จึงเป็นเหตุให้ไม่สามารถตรวจจับใบหน้าและยืนยันตัวตนได้เมื่อสวมใส่หน้ากากอนามัย

จากปัญหาและอุปสรรคที่พบในการศึกษานี้พบว่าผู้วิจัยควรต้องพัฒนาระบบให้สามารถตรวจจับและยืนยันตัวตนได้ถูกต้องแม่นยำมากยิ่งขึ้น โดยการเพิ่มจำนวนภาพต้นฉบับ หรือปรับเปลี่ยนขนาดภาพถ่ายและความละเอียดภาพที่เหมาะสม ไม่ให้สูงหรือต่ำไปเพื่อประสิทธิภาพและความรวดเร็วในการทำงานของระบบ และคำนึงถึงวิถีชีวิตของคนไทยโดยเฉพาะ ควรพัฒนาให้สอดคล้องกับและเหมาะสมกับการใช้งานในสถานการณ์ปัจจุบัน รวมถึงสถานการณ์โรคระบาด โดยปรับเปลี่ยนหรือหาวิธีการที่ทำให้สามารถทำให้ตรวจจับและยืนยันตัวตนได้เมื่อสวมใส่หน้ากากอนามัย

### เอกสารอ้างอิง

- [1] เทพฤทธิ์ บัณฑิตวัฒนาวงศ์. “การคำนวณแบบคลาวด์ (Cloud Computing)” *พื้นฐานการคำนวณแบบคลาวด์ (Cloud Computing Fundamental)*, โรงพิมพ์ Protex บริษัท แคนเน็กซ์ อินเทอร์เน็ตเซอร์เปอเรชั่น จำกัด, กรุงเทพมหานคร, 2562.

- [2] A.K. Datta, M. Datta, and P.K.Banerjee. “Face Detection and Recognition: Theory and Practice.” *CRC Press Taylor & Francis Group*, Florida, 2016.
- [3] C.G.E. Sandoval, “What is face-api.js ?.” *Multiple Face Detection and Recognition System Design Applying Deep Learning in Web Browsers using JavaScript*, Department of Computer Science and Computer Engineering College of Engineering University of Arkansas Fayetteville, 2019.
- [4] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam. “Mobilenets: Efficient convolutional neural networks for mobile vision applications.” *arXiv:1704.04861*, Apr, 2017.
- [5] J.G. RoshanTharanga, S. M. S.C.Samarakoon, T. A. P. Karunarathne, K. L. P. M. Liyanage. M.P.A.W. Gamage and D.Perera., “Smart Attendance using Real Time Face Recognition (SMART-FR).” 2013.
- [6] Y. Kawaguchi, T. Shoji, W. Lin, K. kakusho and M. Minoh. “Face Recognition-based Lecture Attendance System.” Oct, 2014.
- [7] S. Chintalapati, and M.V. Raghunadh. “Automated Attendance Management System Based on Face Recognition Algorithms.” *IEEE International Conference on Computational Intelligence and Computing Research*, 2013.

# การวิเคราะห์รูปแบบสื่อการสอนประเภทเกมวิชาการเขียนโปรแกรมภาษาจาวา เพื่อเพิ่มผลสัมฤทธิ์ทางการเรียนโดยใช้หลักการออกแบบประสบการณ์ผู้ใช้ An Analysis of Game-Based Learning Materials in Java Programming Language to Enhance Learning Achievement by Using User Experience

ภัทรธร เอกธรรมบดี (Phattaradon Ekthammabordee)<sup>1</sup> และอภิษฐา นิมคุ้มภัย (Apichaya Nimkoompai)<sup>2</sup>  
วิทยาศาสตร์มหาบัณฑิต (สาขาเทคโนโลยีสารสนเทศ) คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยี ไทย – ญี่ปุ่น  
<sup>1</sup>ek.phattaradon\_st@tni.ac.th, <sup>2</sup>apichaya@tni.ac.th

## บทคัดย่อ

ในปัจจุบันภาคธุรกิจส่วนใหญ่ได้ให้ความสำคัญกับงานด้านไอทีมากขึ้นส่งผลกระทบต่อความต้องการของตลาดแรงงานไอทีเพิ่มสูงขึ้น แต่จำนวนบุคลากรไอทีในตลาดกลับขาดแคลน เนื่องจากวิชาการเขียนโปรแกรมมีความซับซ้อน และเข้าใจได้ยาก การวิเคราะห์หารูปแบบสื่อการสอนที่น่าสนใจจะช่วยแก้ปัญหาดังกล่าวได้ ผู้วิจัยจึงนำเอาการเรียนรู้ผ่านการใช้สื่อการสอนประเภทเกมมือถือ (Mobile Game-Based Learning) มาวิเคราะห์ร่วมกับรูปแบบประสบการณ์ผู้ใช้ (User Experience : UX) เพื่อประเมินหา รูปแบบเกมของสื่อการสอนประเภทเกมที่เหมาะสมต่อผู้เรียนมากที่สุด และสามารถเพิ่มผลสัมฤทธิ์ทางการเรียนได้อย่างมีประสิทธิภาพ

**คำสำคัญ:** การสอนประเภทเกมมือถือ ประสบการณ์ผู้ใช้ ภาษาจาวา

## Abstract

Nowadays, most businesses are focusing more on IT jobs, affecting the demand of the IT labor market to increase. But the number of IT personnel in the market is scarce. Because programming is complicated and difficult to understand Therefore, from the aforementioned problems, the researcher has brought learning through the use of teaching media in the type of mobile games (Mobile Game-Based Learning) together with the user experience (UX) to analyze and evaluate. Best game style of teaching

materials and can increase my academic achievement effectively

**Keywords:** Mobile Game-Based Learning, User Experience, Java Language.

## 1. บทนำ

ในปัจจุบันความต้องการของตลาดแรงงานในปี 2021 มีความต้องการในสายอาชีพ IT เพิ่มขึ้น จากสถิติผลสำรวจความต้องการของตลาดแรงงานจาก LinkedIn โดยเก็บข้อมูลจากการเปิดรับสมัครตำแหน่งงานกว่า 15,000 ประเภท เผยว่าสายอาชีพซอฟต์แวร์และเทคโนโลยี คิด 1 ใน 15 อันดับตำแหน่งงานที่ตลาดต้องการมากที่สุด โดยเฉพาะในประเทศไทย มาเลเซีย และสิงคโปร์ [1] อีกทั้งในยุคที่โลกธุรกิจในปัจจุบันให้ความสำคัญกับการทรานส์ฟอร์มเข้าสู่ดิจิทัลมากขึ้นเนื่องจากสถานการณ์ COVID-19 ทำให้สายอาชีพนี้เป็นที่ต้องการสูงขึ้นไปอีก และยังมีแนวโน้มการขยายตัวอย่างต่อเนื่องเพิ่มขึ้นทุกปี ซึ่งสวนทางกับจำนวนบุคลากรไอทีในตลาดแรงงานที่ไม่เพียงพอและอยู่ในภาวะขาดแคลน

ด้วยความต้องการของตลาดแรงงานในประเทศไทยในปัจจุบันชี้ให้เห็นว่า ทักษะการเขียนโปรแกรม เป็นอีกทักษะหนึ่งที่มีความสำคัญมากในปัจจุบัน ส่งผลให้การศึกษาในระดับปริญญาตรี มีการนำวิชาการเขียนโปรแกรมเข้ามาอยู่ในหลักสูตรมากขึ้น แต่เนื่องจากเนื้อหาวิชานี้ค่อนข้างซับซ้อน และเข้าใจได้ยาก ทำให้การเรียนรู้ลักษณะการเขียนโปรแกรม (Syntax) เพียงอย่าง

เดียวไม่เพียงพอ ต้องอาศัยการทดลองและการฝึกฝนเพื่อทำให้เกิดเข้าใจมากยิ่งขึ้น

จากความเป็นมาและปัญหาที่กล่าวมานี้ จึงมีแนวคิดที่จะพัฒนาสื่อการสอนประเภทเกมมือถือ (Mobile Game-Based) เนื่องจากเป็นสื่อประเภทที่ทำให้ผู้เรียนสามารถทดลองฝึกฝน อีกทั้งสื่อการสอนบนอุปกรณ์มือถือจะทำให้ผู้เรียนสามารถเรียนได้ทุกที่ทุกเวลา อีกทั้งการพัฒนาสื่อการสอนในรูปแบบเกม จะทำให้ผู้เรียนได้รับประสบการณ์ใหม่ในการเรียน และเป็นการกระตุ้นผู้เรียนให้เรียนรู้ได้ดียิ่งขึ้น ซึ่งจะส่งผลให้ผู้เรียนสามารถเข้าถึงบทเรียนได้ง่ายมากยิ่งขึ้น

อย่างไรก็ตามสื่อการเรียนการสอนประเภทเกม (Mobile Game-Based Learning) ทุกรูปแบบไม่สามารถใช้ได้กับการเรียนวิชาการเขียนโปรแกรมภาษาจาวา (Java) และผู้เรียนทุกคนอาจจะไม่สามารถเรียนรู้จากสื่อการสอนในรูปแบบเดียวกันได้ จึงนำเอาหลักการออกแบบประสบการณ์ผู้ใช้ (UX) มาศึกษาและวิเคราะห์ร่วมด้วยเพื่อหารูปแบบประสบการณ์ผู้ใช้ที่เหมาะสมกับเนื้อหาของวิชาการเขียนโปรแกรมภาษาจาวา ในรูปแบบเกมมือถือ (Mobile Game-Based Learning) เพื่อหารูปแบบที่เหมาะสมกับผู้เรียน เพื่อช่วยเพิ่มผลสัมฤทธิ์ทางการเรียน

## 2. วรรณกรรมที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึง แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับเนื้อหาของงานวิจัย และนำเสนอตามหัวข้อดังต่อไปนี้

### 2.1 Java Programming Language

ภาษาจาวา (Java) เป็นภาษาโปรแกรมเชิงวัตถุ (Object Oriented Programming หรือ OOP) พัฒนาโดย เจมส์ กอสลิง และวิศวกรคนอื่น ๆ ที่ ซัน ไมโครซิสเต็มส์ ภาษาจาวาถูกพัฒนาขึ้นในปี พ.ศ. 1991 และสำเร็จออกสู่สาธารณะในปี ค.ศ. 1995 ในปัจจุบันภาษาจาวา (Java) ได้รับความนิยมอย่างมาก โดยติดเป็นอันดับ 3 จากสถิติ TIOBE (สถิติการค้นหาค้นหาผ่าน Search Engine) [2] นอกจากนี้บริษัท Coding Dojo ได้รวบรวมข้อมูล 10 อันดับภาษาโปรแกรมมิ่งที่นายจ้างกำลังมองหา โดยวิเคราะห์มาจากประกาศรับสมัครงานล่าสุดจาก

Indeed.com ซึ่งเป็นเว็บไซต์หางานอันดับต้น ๆ ของโลก ได้ระบุว่าภาษาจาวาติดเป็นอันดับที่ 3 จาก 10 อันดับภาษาโปรแกรมมิ่งที่นายจ้างกำลังมองหา [3]

**ตารางที่ 1:** แสดงสถิติ TIOBE ประจำเดือนตุลาคม 2021 และ ตุลาคม 2020

Oct 2021	Oct 2020	Programming Language
1	3	Python
2	1	Java
3	2	C
4	4	C++

จากตารางที่ 1 แสดงให้เห็นลำดับความนิยมของภาษาในการเขียนโปรแกรมปี 2021 และ 2020 จะเห็นได้ว่าภาษาจาวายังได้รับความนิยมอย่างต่อเนื่องซึ่งเป็นสาเหตุในการนำมาวิเคราะห์หารูปแบบสื่อการสอนในงานวิจัยนี้

### 2.2 ประสบการณ์ผู้ใช้ (User Experience : UX)

ประสบการณ์ของผู้ใช้งานในด้านความรู้สึกที่ตอบสนองต่อการใช้งานผลิตภัณฑ์, สื่อ หรือระบบต่าง ๆ เช่น ความสะดวกสบาย ใช้งานง่าย ความสนุกสนาน จนเกิดเป็นความพึงพอใจหรือเกิดประสบการณ์ที่ดีของผู้ใช้งานนั่นเอง [4] หรือในอีกแง่หนึ่งก็คือ ความยากลำบาก ใช้งานได้ยาก และเกิดเป็นความไม่พอใจหรือเกิดประสบการณ์ที่แย่ต่อผู้ใช้ ดังนั้นประสบการณ์ผู้ใช้ (UX) คือ การพยายามศึกษาและทำความเข้าใจว่าผู้ใช้งานต้องการอะไร แบบไหน กลุ่มเป้าหมายมีใครบ้าง มีอะไรน่าสนใจบ้าง อย่างละเอียด เพื่อให้ตอบโต้กับผู้ใช้งานมากที่สุด ฉะนั้นอาจเปรียบได้ว่าประสบการณ์ผู้ใช้ (UX) คือ “ศาสตร์แห่งความพยายามเข้าใจผู้อื่นเพื่อประโยชน์อันสูงสุด” [4]

ประสบการณ์ผู้ใช้เป็นอีกหนึ่งปัจจัยสำคัญที่ส่งผลถึงชื่อเสียงของแบรนด์ ความน่าดึงดูดของผลิตภัณฑ์ที่มีต่อลูกค้า และยอดขายสินค้า จากผลการวิจัยของ Alexandre [5] พบว่าปัจจัยที่ส่งผลต่อความประทับใจแรกของผู้ใช้ที่มีต่อเว็บไซต์ ประกอบไปด้วย 2 ปัจจัยด้วยกันคือ ความซับซ้อนของเว็บไซต์น้อย และการเข้าถึงเว็บไซต์ได้ง่าย โดยทั้ง 2 ปัจจัยนี้ผู้ใช้ใช้เวลาในการตัดสินใจเพียง 50 milliseconds (ms) [6] เท่านั้น นอกจากนี้จากงานวิจัยของ

Silence [7] กล่าวว่าความประทับใจครั้งแรกที่ผู้ใช้มีต่อการใช้งานของเว็บไซต์ที่ออกแบบแย่งส่งผลกระทบต่อความน่าเชื่อถือของเว็บไซต์ต่อผู้ใช้นั้นมากกว่าเนื้อหาหลักที่อยู่ในเว็บไซต์ถึง 94% ซึ่งจากผลดังกล่าวชี้ให้เห็นว่า หากผู้ใช้รู้สึกไม่ประทับใจต่อเว็บไซต์แล้ว จะก่อให้เกิดความรู้สึกในเชิงลบและส่งผลกระทบต่อความพึงพอใจของผู้ใช้ต่อแบรนด์อีกด้วย [8]

จากผลการวิจัยที่กล่าวมาข้างต้น แสดงให้เห็นว่าประสบการณ์ผู้ใช้ที่มีอิทธิพลต่อความรู้สึกของผู้ใช้อย่างมาก หากผู้ใช้ได้รับประสบการณ์ที่แย่งจากการใช้งานเว็บไซต์หรือผลิตภัณฑ์ใดๆ ก็ตาม อาจส่งผลทำให้เกิดความรู้สึกในเชิงลบต่อผลิตภัณฑ์ได้ ดังนั้นประสบการณ์ผู้ใช้จึงเป็นอีกหนึ่งปัจจัยสำคัญในการนำมาวิเคราะห์และประเมินความพึงพอใจของผู้ใช้งานที่ส่งผลกระทบต่อผลิตภัณฑ์ [9]

### 2.3 การเรียนรู้ผ่านการใช้สื่อการสอนประเภทเกมมือถือ (Mobile Game-Based Learning)

การเรียนรู้ผ่านการใช้สื่อการสอนประเภทเกมมือถือ (Mobile Game-Based Learning หรือ mGBL) เป็นรูปแบบการเรียนรู้ที่เพิ่มความสนุกสนานในรูปแบบของเกมมือถือเข้าไปเพื่อช่วยดึงดูดความสนใจของผู้เรียน อีกทั้งยังช่วยให้ผู้เรียนสามารถเรียนที่ไหนก็ได้ นอกเหนือไปจากที่โรงเรียน โดยแนวคิดหลักของการเรียนรู้ผ่านการใช้สื่อการสอนประเภทเกมมือถือ คือการปรับปรุงการเรียนการสอนใหม่ โดยนำเอาความบันเทิงเข้ามาเป็นส่วนหนึ่งในการออกแบบการเรียนรู้เพื่อให้มีความสอดคล้องกับการเรียนรู้ของผู้เรียนมากขึ้น จนเป็นที่มาของคำว่า Edutainment ซึ่งหมายถึงการศึกษาที่มาพร้อมกับความบันเทิง กับคำว่า Plearn ที่หมายถึง เล่นและเรียน โดยมาจากคำว่า Play รวมกับ Learn นั่นเอง

มีงานวิจัยเกี่ยวกับการเรียนรู้ผ่านการใช้สื่อการสอนประเภทเกมมือถือโดย Jordine [10], Nunohiro [11], Lin [12] และ Yadav [13] กล่าวตรงกันว่าสื่อการเรียนการสอนประเภทเกมมือถือมีผลต่อการเพิ่มแรงจูงใจในการเรียนรู้ต่อผู้เรียนอย่างมาก อีกทั้งการนำเสนอสื่อที่อยู่บนมือถือยังมีส่วนช่วยในการเพิ่มโอกาสการเรียนรู้ให้กับผู้เรียนได้อย่างดี ไม่ว่าจะผู้เรียนจะอยู่ที่ไหนก็สามารถเรียนได้ อีกทั้งใน

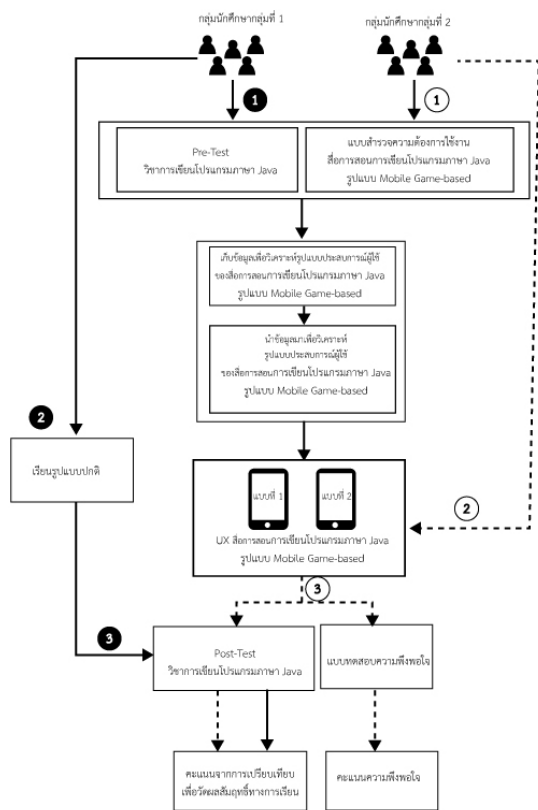
สถานการณ์ที่ COVID-19 แพร่ระบาดส่งผลกระทบต่อการศึกษาในประเทศไทย ณ ปัจจุบัน ทำให้การเรียนผ่านทางสื่อออนไลน์ต่างๆ มีบทบาทมากยิ่งขึ้น และส่งผลทำให้การเรียนรู้อ่านการใช้สื่อการสอนประเภทเกมมือถือมีความน่าสนใจมากขึ้น จากผลการวิจัยของ Krouska [14] พบว่า การเรียนรู้ผ่านการใช้สื่อการสอนประเภทเกมมือถือในช่วง COVID-19 ช่วยเพิ่มความสนใจให้กับผู้เรียนได้อย่างมีประสิทธิภาพ

สิ่งหนึ่งที่ต้องคำนึงถึงในการพัฒนาสื่อการเรียนการสอนประเภทเกมมือถือ นั่นคือการออกแบบส่วนต่อประสานผู้ใช้ (User Interface : UI) เพื่อให้ผู้ใช้ได้รับประสบการณ์ผู้ใช้ที่ดี จากผลการวิจัยของ Bartel [15] พบว่าหลังจากกลุ่มสำรวจทดลองใช้สื่อการเรียนการสอนประเภทเกมมือถือแล้ว แสดงความคิดเห็นว่าสื่อควรปรับปรุงในเรื่องของขนาดตัวอักษร และตำแหน่งของปุ่มในแอปพลิเคชัน

จากผลการวิจัยทั้งหมดข้างต้นสรุปได้ว่า สื่อการสอนประเภทเกมมือถือ มีศักยภาพในการทำให้ผู้เรียนมีแรงจูงใจต่อบทเรียนมากขึ้น อีกทั้งยังมีข้อดีคือเพิ่มโอกาสในการเรียนรู้บทเรียนเพิ่มเติม ไม่ว่าจะอยู่สถานที่ไหนเมื่อไหร่ และยังเป็นสื่อการเรียนการสอนที่เหมาะสมต่อยุคสมัยในปัจจุบันที่เกิดเหตุการณ์การแพร่ระบาดของ COVID-19 ทั้งนี้การพัฒนาสื่อการเรียนการสอนประเภทเกมมือถือยังมีปัจจัยที่ต้องคำนึงถึงคือ เรื่องของการออกแบบสื่อให้ง่ายต่อการใช้งาน

### 3. วิธีการดำเนินการวิจัย

สำหรับขั้นตอนการทำงานจะมีทั้งหมด 7 ขั้นตอน ดังภาพที่ 1 โดยในงานวิจัยนี้จะแบ่งกลุ่มสำรวจออกเป็น 2 กลุ่ม กลุ่มที่ 1 เป็นนักศึกษาที่เรียนวิชา Java Programming Language ผ่านการเรียนการสอนในห้องเรียน และกลุ่มที่ 2 กลุ่มนักศึกษาศึกษาที่เรียน Java Programming Language ผ่านสื่อการสอนประเภทเกมมือถือ โดยมีรายละเอียดขั้นตอนการทำงาน ดังนี้



ภาพที่ 1: ขั้นตอนการทำงาน

จากภาพที่ 1 แสดงให้เห็นถึงขั้นตอนการดำเนินการวิจัยดังต่อไปนี้

**ขั้นตอนที่ 1 เก็บข้อมูลเบื้องต้นและทำ Pre-Test**

ขั้นตอนที่ 1 การเก็บข้อมูลเบื้องต้นของนักศึกษาระดับปริญญาตรี คณะเทคโนโลยีสารสนเทศ (ทั้ง 2 กลุ่ม) เกี่ยวกับปัจจัยต่างๆของผู้เรียนที่อาจส่งผลกระทบต่อประสิทธิภาพในการเรียนรู้วิชา Java Programming Language และยังเป็นข้อมูลเพื่อกำหนดแนวทางการเลือกหัวข้อเนื้อหาในวิชา และกำหนดประเภทของเกมมือถือมาพัฒนา Prototype รวมไปถึงให้กลุ่มสำรวจทั้ง 2 กลุ่มทำข้อสอบ Pre-Test เพื่อนำไปใช้วิเคราะห์ข้อมูลในขั้นตอนต่อไป

เมื่อเก็บข้อมูลจากกลุ่มทดลองแล้ว นำผลจากแบบสอบถามที่ได้รับมาวิเคราะห์เพื่อนำไปเป็นแนวทางในการออกแบบและพัฒนาต้นแบบจำลอง สำหรับสื่อการเรียนการสอนประเภทเกม วิชา Java Programming Language ทั้งหมด 2 รูปแบบเกม เพื่อให้สื่อการเรียนการสอนนี้สามารถมอบความพึงพอใจและประสบการณ์ผู้ใช้ (UX) ที่ดีให้แก่ผู้เรียนได้อย่างมีประสิทธิภาพ

**ขั้นตอนที่ 2 ทดลองใช้สื่อการสอน**

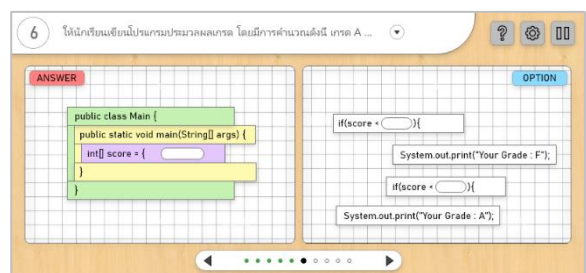
ในขั้นตอนนี้จะแบ่งออกเป็น 2 แบบโดยอิงตามกลุ่มสำรวจทั้ง 2 กลุ่ม โดยที่กลุ่มที่ 1 จะไปเรียนผ่านการสอนในห้องเรียน และกลุ่มที่ 2 จะเรียนผ่านการใช้สื่อการสอนประเภทเกมมือถือ

**ขั้นตอนที่ 3 ทำ Post-Test และแบบประเมินความพึงพอใจ**

ให้กลุ่มสำรวจทั้ง 2 กลุ่ม ทำข้อสอบ Post-Test และให้กลุ่มสำรวจกลุ่มที่ 2 แบบประเมินความพึงพอใจหลังการใช้งานสื่อการสอนประเภทเกมมือถือ เพื่อนำข้อมูลไปใช้ในการประเมินและวิเคราะห์ผลสัมฤทธิ์ทางการเรียน และประเมินความพึงพอใจของผู้เรียน

**4. ผลการดำเนินงาน**

จากการเก็บข้อมูลปัจจัยต่างๆ ของผู้เรียนที่อาจส่งผลกระทบต่อประสิทธิภาพในการเรียนรู้วิชา Java Programming Language ในขั้นตอนที่ 1 ทำให้สามารถกำหนดแนวทางการพัฒนาต้นแบบจำลอง โดยจะนำเนื้อหาในหัวข้อ If-Else มาพัฒนาเป็นสื่อการสอนเนื่องจากเป็นหัวข้อที่นักศึกษามีความเข้าใจค่อนข้างน้อย ทั้งนี้ผู้วิจัยจึงได้พัฒนาต้นแบบจำลองสื่อการสอนประเภทเกมบนมือถือ 2 รูปแบบ โดยแบบที่ 1 มีรูปแบบของเกมเป็นแบบ Drag & Drop ดังรูปภาพที่ 2 และแบบที่ 2 จะเป็นเกมแบบดำเนินเรื่องด้วยภาพเพื่อสร้างความเข้าใจตรรกะการเขียนโปรแกรม



ภาพที่ 2: แสดงตัวอย่างส่วนต่อประสานกับผู้ใช้งาน (UI) ของสื่อการสอนประเภทเกมมือถือแบบ Drag & Drop

จากภาพที่ 2 แสดงให้เห็นถึงรูปแบบสื่อการสอนประเภทเกมมือถือแบบ Drag & Drop ซึ่งเป็นรูปแบบที่มีความเหมาะสมกับลักษณะของเนื้อหาในบทเรียนและสามารถทำความเข้าใจได้ง่าย





ภาพที่ 3: แสดงตัวอย่างส่วนต่อประสานกับผู้ใช้งาน (UI) ของสื่อการสอนประเภทเกมมือถือแบบดำเนินเรื่องด้วยภาพ

จากภาพที่ 3 แสดงให้เห็นถึงสื่อการสอนประเภทเกมมือถือแบบดำเนินเรื่องด้วยภาพเพื่อสร้างความเข้าใจตรรกะการเขียนโปรแกรม ซึ่งเป็นรูปแบบที่สามารถเข้าใจได้จากการจำลองสถานการณ์ต่างๆ

จากการดำเนินการวิจัยพบว่า กลุ่มสำรวจที่เรียนรู้วิชา Java Programming Language ผ่านสื่อการสอนประเภทเกมมือถือทั้งสองรูปแบบมีผลการสอบ Post-Test ที่ดีมากกว่ากลุ่มสำรวจที่เรียนรู้ผ่านการสอนในห้องเรียนปกติ ดังที่แสดงในตารางที่ 2

ตารางที่ 2: แสดงข้อมูลคะแนนเฉลี่ยการสอบ Pre-Test และ Post-Test ของกลุ่มสำรวจทั้ง 2 กลุ่ม

	คะแนนสอบเฉลี่ย (คะแนนเต็ม 100)	
	กลุ่มสำรวจที่ 1	กลุ่มสำรวจที่ 2
Pre-Test	18	10
Post-Test	72	79
ผลต่างระหว่าง Pre-Test / Post-Test	54%	69%

จากตารางที่ 2 แสดงให้เห็นอัตราการเพิ่มขึ้นของคะแนนจะสังเกตได้ว่ากลุ่มที่ 2 ที่เรียนผ่านสื่อการสอนประเภทเกมมือถืออัตราการเพิ่มขึ้นของคะแนนมากกว่ากลุ่มที่ 1 ที่เรียนแบบปกติ

ตารางที่ 3: แสดงข้อมูลความพึงพอใจ

หัวข้อการประเมิน	คะแนนความพึงพอใจเฉลี่ย (เต็ม 5)	
	แบบที่ 1	แบบที่ 2
สร้างความน่าสนใจในการเรียน	4.15	4.23
สร้างความสนุกสนานในการเรียน	3.95	4.5
ความสะดวกในการใช้งาน	4.13	4.15
ช่วยให้เข้าใจบทเรียนมากขึ้น	4.72	3.98
การติดตั้งแอปพลิเคชัน	4.8	4.78
คะแนนโดยรวม	4.35	4.33

จากตารางที่ 3 แสดงผลการประเมินความพึงพอใจจากการใช้งานสื่อการสอนประเภทเกมมือถือ พบว่ากลุ่มสำรวจพึงพอใจในการใช้งานสื่อการสอนประเภทเกมมือถือรูปแบบที่ 1 คือรูปแบบ Drag & Drop มากกว่ารูปแบบที่ 2 รูปแบบดำเนินเรื่องด้วยภาพเล็กน้อย เนื่องจากมีความเหมาะสมกับรูปแบบของเนื้อหา สามารถเข้าใจถึงวิธีการใช้งานได้ง่าย ซึ่งจะทำให้เข้าใจบทเรียนได้มากขึ้น อีกทั้งยังสามารถนำเสนอบทเรียนในรูปแบบที่ให้ความเพลิดเพลินกับผู้เรียนได้อีกด้วย

## 5. สรุป

เนื่องจากปัญหาการเรียนรู้อวิชญา Java Programming Language ที่มีความซับซ้อน เข้าใจได้ยาก ซึ่งเป็นอุปสรรคต่อผู้เรียน จากปัญหาดังกล่าว จึงได้นำเอาสื่อการสอนประเภทเกมมือถือ (Mobile Game-Based Learning) มาวิเคราะห์ร่วมกับประสบการณ์ผู้ใช้ (UX) เพื่อเพิ่มผลสัมฤทธิ์ทางการเรียน และทำให้ผู้เรียนสามารถเข้าถึงบทเรียนได้ง่าย ลดความซับซ้อนของบทเรียน มีความสะดวกสบายในการใช้งานสื่อที่ไหนก็ได้ การนำเสนอสื่อการสอนประเภทเกมมือถือยังส่งเสริมให้ผู้เรียนมีแรงจูงใจในการเรียนรู้มากยิ่งขึ้น จากผลการวิจัยโดยการประเมินจากคะแนน Pre-Test และ Post-Test ของกลุ่มสำรวจทั้ง 2 กลุ่ม แสดงให้เห็นว่าการใช้สื่อการสอนแบบเกมมือถือ (69%) ให้ผลสัมฤทธิ์ทางการเรียนสูงกว่าการเรียนรู้ผ่านห้องเรียน (54%) มากกว่า 15% อีกทั้งยังสรุปได้ว่าผู้เรียนมีความพึงพอใจต่อสื่อการสอนประเภทเกมมือถือใน

รูปแบบของเกม Drag & Drop มากกว่า เนื่องจากมีความเหมาะสมกับรูปแบบเนื้อหามากกว่า

ทั้งนี้ในอนาคตผู้วิจัยจะนำเอาเนื้อหาอื่นๆ ในวิชา Java Programming Language มาทำเป็นเป็นสื่อการสอนประเภทเกมเช่นเดียวกันนอกเหนือจากเรื่อง If-Else และพัฒนาให้สื่อการสอนประเภทมื่อถื่อนี้สามารถนำไปใช้งานร่วมกับการสอนของอาจารย์ในห้องเรียนได้จริง

### เอกสารอ้างอิง

- [1] A. BrandInside. 2021. *LinkedIn เปิด 10 สายงานมาแรง ในอาเซียน คอนเทนต์-ไอที-อีคอมเมิร์ซต้องการสูง*. [Online] Brand Inside. Available at: <<https://brandinside.asia/fastest-growing-jobs-sea/>> [Accessed 23 October 2021].
- [2] Tiobe.com. 2000. *TIOBE Index for October 2021*. [online] Available at: <<https://www.tiobe.com/tiobe-index/>> [Accessed 8 October 2021].
- [3] thip. 2020. *Business Insider เผย 10 อันดับ ภาษาโปรแกรมมิ่ง ที่นายจ้างกำลังมองหา*. [Online] Medium. Available at: <<https://medium.com/thipwriteblog/business-insider-business-insider-เผย-10-อันดับ-ภาษาโปรแกรมมิ่ง-ที่นายจ้างกำลังมองหา-แถมลิงก์เรียนฟรีให้ทุกคน-56eca1302776>> [Accessed 7 November 2021].
- [4] DEMETER ICT. 2015. “UX / UI คืออะไร?”. [online] Available at: <<https://www.dmit.co.th/ข่าวสาร/ux-vs-ui/>> [Accessed 8 October 2021].
- [5] A. Tuch, E. Presslauer, M. Stöcklin, K. Opwis, and J. Bargas-Avila. “The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments.” *International Journal of Human-Computer Studies*, Vol. 70, No. 11, pp.794-811, 2012.
- [6] G. Lindgaard, G. Fernandes, C. Dudek, and J. Brown. “Attention web designers: You have 50 milliseconds to make a good first impression!” *Behaviour & Information Technology*, Vol. 25, No. 2, pp.115-126, 2006.
- [7] E. Sillence, P. Briggs, L. Fishwick, and P. Harris. “Trust and mistrust of online health sites.” *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, 2004.
- [8] E. Raita, and A. Oulasvirta. “Too good to be bad: Favorable product expectations boost subjective usability ratings.” *Interacting with Computers*, Vol. 23, No. 4, pp. 363-371, 2011.
- [9] Y. Wang, and X. Che. “How to Keep People Playing Mobile Games: An Experience Requirements Testing Approach.” *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCOM/IoP/SmartWorld)*, 2016.
- [10] T. Jordine, Y. Liang, and E. Ihler. “A Mobile Device Based Serious Gaming Approach for Teaching and Learning Java Programming.” *International Journal of Interactive Mobile Technologies (IJIM)*, Vol. 9, No. 1, p. 53, 2015.
- [11] E. Nunohiro, K. Matsushita, K. Mackin, and M. Ohshiro. “Development of game-based learning features in programming learning support system.” *Artificial Life and Robotics*, Vol. 17, No. 3-4, pp. 373-377, 2012.
- [12] W. Lin, J. Ho, C. Lai, and B. Jong. “Mobile game-based learning to inspire students learning motivation.” *2014 International Conference on Information Science, Electronics and Electrical Engineering*, 2014.
- [13] A. Yadav, and S. Oyelere. “Contextualized mobile game-based learning application for computing education.” *Education and Information Technologies*, Vol. 26, No. 3, pp. 2539-2562, 2020.
- [14] A. Krouska, C. Troussas, and C. Sgouropoulou. “Mobile game-based learning as a solution in COVID-19 era: Modeling the pedagogical affordance and student interactions.” *Education and Information Technologies*, Vol. 27, No. 1, pp. 229-241, 2021.
- [15] A. Bartel, and G. Hagel. “Engaging students with a mobile game-based learning system in university education.” *2014 IEEE Global Engineering Education Conference (EDUCON)*, 2014.

# การปรับแต่งประสิทธิภาพความพร้อมใช้งานแอปพลิเคชันงานทะเบียนและสถิติ

นักศึกษาด้วยมองโกดีบีเรพลิเคชัน

## An Efficiency Tuning for Registration and Educational Statistic Application with MongoDB Replication

วิโรจน์ เต็มบุญเลิศชัย (Wiroj Tamboonlertchai)<sup>1</sup> ณัฐสุรางค์ คำละกอ (Nutsurang Dumlakor)<sup>2</sup>

และชุมพล โมฆรัตน์ (Chumpol Mokarat)<sup>3</sup>

สาขาวิชาเทคโนโลยีสารสนเทศ คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก

<sup>1</sup>wiroj.tam@rmutto.ac.th, <sup>2</sup>nutsurang.dum@rmutto.ac.th, <sup>3</sup>chumpol\_mo@rmutto.ac.th

### บทคัดย่อ

งานทะเบียนและสถิตินักศึกษาเป็นหน่วยงานที่มีการดำเนินงานกับผู้เกี่ยวข้องหลายฝ่าย มีการจัดการและการประมวลผลข้อมูลที่หลากหลาย และผู้ใช้งานพร้อมกันจำนวนมากทำให้มีโอกาสสูงที่เครื่องแม่ข่ายเกิดข้อผิดพลาดอาจส่งผลให้ระบบมีความพร้อมใช้งานและความเชื่อถือในระดับค่อนข้างต่ำ ผู้วิจัยจึงนำฐานข้อมูลโนเอสคิวแอลมาใช้งาน ออกแบบและพัฒนาเป็นระบบฐานข้อมูลแบบกระจายด้วยการจัดเก็บข้อมูลบนเครื่องแม่ข่ายหลายเครื่อง ด้วยกระบวนการทำงานของเรพลิเคชัน ที่ซึ่งมีการดำเนินการทำสำเนาข้อมูลจากโหนดหลักไปยังโหนดรองอื่น หากโหนดหลักไม่สามารถทำงานได้ มีวัตถุประสงค์เพื่อเพิ่มประสิทธิภาพด้านความพร้อมใช้งานและความน่าเชื่อถือระบบงานทะเบียนและสถิตินักศึกษา ด้วยการประยุกต์ใช้งานมองโกดีบีเรพลิเคชัน พัฒนาแอปพลิเคชันด้วยดีจัง โทเฟรมเวิร์คกับฐานข้อมูลโนเอสคิวแอล ด้วยมองโกดีบี ประเมินผลการทำงานค่าเฉลี่ยของเวลาซ่อมแซม จากเวลาการซ่อมแซมระบบจำนวน 12 ครั้ง พบว่า ระบบมีเวลาเฉลี่ยในการซ่อมแซมประมาณ 4 วินาที ใช้เวลาค่อนข้างน้อยและส่งผลต่อการดำเนินงานของระบบให้มีความต่อเนื่อง

**คำสำคัญ:** งานทะเบียนและสถิตินักศึกษา ฐานข้อมูลโนเอสคิวแอล มองโกดีบี เรพลิเคชัน

### Abstract

The Registration and Educational Statistics Office is a comprehensive department with many divisions that operate cooperatively. There are several concurrent users on the server, that include various types of data management and processing, consequently there's a high likelihood it is always crashed. And efficiently in a poor performance to the system's availability and reliability. We're applying NoSQL database application, design, and development to collect data in distributed databases through replication techniques. When the primary node is cut off, data is efficiently duplicated from the primary node to secondary nodes and others. The goal of this study is to use the MongoDB replication approach to improve the application's availability and reliability based on the Django framework and the MongoDB database. When a system crashes, the task is evaluated in Mean Time to Repair, which is the average system repair time divided by the number of times the system must be fixed. The average system repair time was around 4 seconds after we turned it all 12 times. It takes less time and has a significant effect on the system's uptime.

**Keywords:** Registration and Educational Statistic Office, NoSQL Database, MongoDB, Replication.

## 1. บทนำ (Introduction)

ระบบงานทะเบียนและสถิติดักศึกษาเป็นหน่วยงานที่มีการดำเนินงานเกี่ยวกับผู้เกี่ยวข้องหลายฝ่าย ได้แก่ อาจารย์ เจ้าหน้าที่ นักศึกษาและผู้ที่มีส่วนเกี่ยวข้องอื่น ๆ ซึ่งการดำเนินการประกอบด้วยการจัดการตารางเรียน ตารางสอบ การลงทะเบียนรายวิชา การประมวลผลคะแนน และอื่น ๆ ประกอบกับมีผู้ใช้งานพร้อมกันเป็นจำนวนมาก โดยพบมากในช่วงเวลาที่มีการลงทะเบียนเรียนของนักศึกษาและการประมวลผลการศึกษาทั้งการส่งผลการศึกษาของอาจารย์ผู้รับชอบรายวิชาและการตรวจสอบผลการศึกษาของนักศึกษาในแต่ละปีการศึกษาเกิดปัญหาขึ้นในระหว่างการใช้งานระบบ ซึ่งปัญหาดังกล่าวทำให้มีโอกาสสูงที่เครื่องแม่ข่าย (Server) เกิดข้อผิดพลาดหรือหยุดทำงานได้ โดยมีสาเหตุมาจากการประมวลผลข้อมูลปริมาณมาก และถ้าหากเดิมที่ระบบงานทะเบียนใช้ระบบฐานข้อมูลแบบเดี่ยว (Centralized Database Systems) จะมีข้อจำกัดด้านการเพิ่มประสิทธิภาพการจัดเก็บและประมวลผลข้อมูลเนื่องจากประสิทธิภาพของซอฟต์แวร์และฮาร์ดแวร์ที่จำกัดหรือด้วยข้อจำกัดอื่น ๆ ทำให้ระบบอาจมีประสิทธิภาพไม่เพียงพอต่อความต้องการในการจัดเก็บและประมวลผลข้อมูลที่มากขึ้นในอนาคต จากปัญหาดังกล่าว อาจส่งผลให้ระบบมีความพร้อมใช้งาน (Availability) และความเชื่อถือ (Reliability) ในระดับค่อนข้างต่ำ ผู้วิจัยจึงนำฐานข้อมูลโนเอสคิวแอล (NoSQL) มาประยุกต์ใช้งาน โดยออกแบบและพัฒนาเป็นระบบฐานข้อมูลแบบกระจาย (Distributed Database Systems) [1] ด้วยการจัดเก็บข้อมูลไว้กับเครื่องแม่ข่ายหลายเครื่องแทนที่จะเก็บข้อมูลไว้ยังเครื่องเดียว โดยระบบฐานข้อมูลแบบกระจายจะเน้นเทคนิคการขยายแบบแนวนอน (Horizontal Scaling) โดยเป็นการเพิ่มทรัพยากรจำนวนโหนด (Node) แทนที่จะขยายแบบแนวตั้ง (Vertical Scaling) ที่เป็นการเพิ่มประสิทธิภาพของโหนด เนื่องจากข้อจำกัดในประสิทธิภาพของโหนดเดียว นอกจากนี้ยังนำกระบวนการทำงานของเรพลิเคชัน (Replication) [2] หรือร่วมและแลกเปลี่ยนข้อมูลข้ามเครื่องแม่ข่าย ที่ซึ่งเป็นหนึ่งในคุณสมบัติที่สำคัญของโนเอสคิวแอล ที่ซึ่งมีการ

ดำเนินการทำสำเนาข้อมูลจากโหนดหลักไปยังโหนดรองอื่น ๆ หากเกิดกรณีที่โหนดหลักไม่สามารถทำงานได้ หนึ่งในโหนดรองที่ถูกกำหนดไว้จะขึ้นมาทำหน้าที่เป็นโหนดหลักแทน ส่งผลให้ผู้ใช้งานสามารถใช้งานระบบต่อไปได้

ดังนั้น ผู้วิจัยจึงนำเสนอแนวคิดการดำเนินงานของเรพลิเคชัน เพื่อเพิ่มประสิทธิภาพด้านความพร้อมใช้งานและความน่าเชื่อถือของระบบงานทะเบียนและสถิติดักศึกษาด้วยการประยุกต์ใช้งานมอดูลโคดีบีด้วยเทคนิคเรพลิเคชัน ซึ่งออกแบบและพัฒนาแอปพลิเคชันด้วยดีเจ็งโกเฟรมเวิร์ค (Django Framework) กับฐานข้อมูลโนเอสคิวแอลด้วยมอดูลโคดีบี เพื่อนำเสนอการดำเนินการของเทคนิคดังกล่าว

## 2. วรณกรรมที่เกี่ยวข้อง (Theory and Related Works)

### 2.1 ปัญหาของระบบงานทะเบียนและสถิติดักศึกษา

เนื่องจากจำนวนของผู้ใช้บริการงานทะเบียนและสถิติดักศึกษาของมหาวิทยาลัยมีจำนวนหลายกลุ่ม ได้แก่ อาจารย์ เจ้าหน้าที่ นักศึกษาและผู้ใช้งานทั่วไป ทำให้การใช้งานระบบดังกล่าวสำหรับการประมวลผลของงานทะเบียนและสถิติดักศึกษามีข้อผิดพลาดเกิดขึ้นค่อนข้างบ่อย โดยพบมากในช่วงเวลาที่มีการลงทะเบียนเรียนของนักศึกษาและการประมวลผลการศึกษาทั้งการส่งผลการศึกษาของอาจารย์และตรวจสอบผลการศึกษาของนักศึกษาเกิดปัญหาในระหว่างการใช้งานทำให้เกิดข้อผิดพลาดและความล่าช้าในการดำเนินการ ซึ่งปัญหาดังกล่าวจำเป็นต้องดำเนินการปรับปรุงระบบเพื่อให้มีความพร้อมใช้งานตลอดเวลาและรองรับปริมาณผู้ใช้งานที่เพิ่มขึ้น ดังนั้นทางผู้วิจัยจึงเล็งเห็นถึงการนำระบบฐานข้อมูลมอดูลโคดีบีมาประยุกต์ใช้งานร่วมกับการพัฒนาระบบดังกล่าว

### 2.2 ระบบฐานข้อมูลโนเอสคิวแอล (NoSQL Database System)

โนเอสคิวแอล [3] (NoSQL) เป็นระบบจัดการฐานข้อมูลที่ไม่ใช่เชิงสัมพันธ์ เป็นฐานข้อมูลที่มีความสามารถประมวลผลได้อย่างรวดเร็ว มักจะใช้ภาษาที่ไม่มีโครงสร้างในการจัดเก็บข้อมูล และระบบฐานข้อมูล

โนเอสคิวแอลมีโครงสร้างที่ถูกต้องแบบมาเพื่อรองรับข้อมูลที่มีปริมาณจำนวนมาก ที่ซึ่งแบ่งเป็น 4 ประเภท ดังนี้ Key-Value Store, Document Database, Wide Column Store และ Graph Database ดังตัวอย่างการจัดเก็บข้อมูลแบบ Document Database ดังภาพที่ 1

```

1 {
2   "1001": {
3     "Name": "Chumpol",
4     "Surname": "Mokarat",
5     "Zipcode": "12120"
6   },
7   "1002": {
8     "Name": "Wiroj",
9     "Surname": "Tamboonlertchai",
10    "Zipcode": "10400"
11  }
12 }
    
```

ภาพที่ 1: การจัดเก็บข้อมูลแบบ Document Database

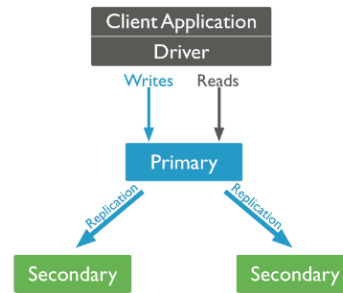
**2.3 ฐานข้อมูลมองโกดีบี (MongoDB Database)**

ฐานข้อมูลมองโกดีบี [4] เป็นฐานข้อมูลโอเพ่นซอร์สที่จัดเก็บข้อมูลแบบ Document Database โดยเป็นฐานข้อมูลแบบ โนเอสคิวแอล ไม่เน้นในการสร้างความสัมพันธ์ของข้อมูล แต่เน้นรูปแบบโครงสร้างที่เจ้าของโนเอสคิวแอลสร้างขึ้นมาและจัดเก็บข้อมูลแบบ JSON (JavaScript Object Notation) ซึ่งเก็บค่าข้อมูลในรูปแบบ Key-Value โดยจุดเด่นอยู่ที่ความเร็วในการทำงานเป็นหลักสามารถ Query ข้อมูลได้เร็ว และการทำงานในส่วนของฐานข้อมูลนั้นจะทำงานลดลง แต่จะไปเน้นการทำงานในส่วนของโปรแกรมที่พัฒนาขึ้นมาแทน ซึ่งเหมาะกับข้อมูลขนาดใหญ่ ที่ไม่ซับซ้อน การทำงานที่ไม่หนักมาก สามารถทำงานกับระบบที่เป็นการทำงานแบบเรียลไทม์ได้โดยฐานข้อมูลมองโกดีบีมีรูปแบบการจัดเก็บ 2 วิธี ดังนี้ Collections และ Schemeless

**2.4 เรพลีเคชัน (Replication)**

ชุดของเรพลีเก้ (Replica Set) [2] ในมองโกดีบีเป็นกลุ่มของกระบวนการ mongod ที่ซึ่ง processes ซึ่งจัดเก็บและบำรุงรักษาข้อมูลชุดเดียวกัน โดยชุดของเรพลีเก้มีการจัดเตรียมข้อมูลสำเนาที่ซ้ำซ้อน (Redundancy) และความน่าเชื่อถือได้ที่สูง (High availability) ถือเป็นพื้นฐานสำคัญสำหรับการกำหนดค่าและติดตั้งผลิตภัณฑ์เพื่อให้ระบบมีความน่าเชื่อถือและคงไว้ซึ่งข้อมูลที่ซ้ำซ้อนกันใน

หลายสำเนา โดยมีองค์ประกอบ ดังนี้ โหนดหลัก (Primary Node) โหนดรอง (Secondary Node) และอาร์บิเตอร์ โหนด (Arbiter) สามารถจำลองการออกแบบดังกล่าวด้วยมาสเตอร์สเลฟเรพลีเคชัน (Master-slave Replication) [5] ซึ่งช่วยให้สามารถจำลองข้อมูลจากเซิร์ฟเวอร์ฐานข้อมูลใด ๆ ไปยังเซิร์ฟเวอร์ฐานข้อมูลอื่นได้ แสดงสถาปัตยกรรมการออกแบบดังกล่าว ดังภาพที่ 2



ภาพที่ 2: สถาปัตยกรรมการออกแบบมองโกดีบีเรพลีเคชัน [2]

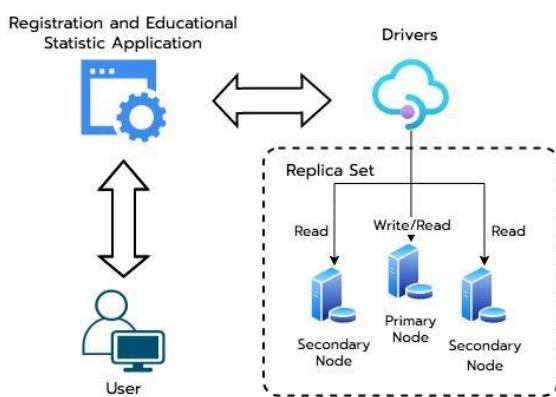
**2.5 งานวิจัยที่เกี่ยวข้อง (Related Works)**

ในปี 2557 นิรุทธ์ รวยรื่น และ เกรียงไกร ปอแก้ว [6] นำเสนอ การใช้แมพรีดิวซ์เชื่อมคอลเลกชันของฐานข้อมูลโนเอสคิวแอลบนมองโกดีบี โดยนำเทคโนโลยีฐานข้อมูลโนเอสคิวแอลเทคนิคใช้ในการจัดเก็บและจัดการข้อมูลที่มีปริมาณมาก เน้นหลักการเชื่อมความสัมพันธ์ของข้อมูลด้วยเทคนิคแมพรีดิวซ์และวัดประสิทธิภาพในการจัดเก็บข้อมูลดังกล่าว ต่อมาในปี พ.ศ. 2558 ศุภดี บุญรอด นำเสนอ [7] การค้นคืนข้อมูลขนาดใหญ่โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้างร่วมกับเทคโนโลยี เว็บเชิงความหมาย โดยดำเนินการวิจัยการค้นคืนข้อมูลให้ตรงตามความต้องการของผู้ใช้งาน การนำเสนอนี้ได้ทำการนำฐานข้อมูลไม่สัมพันธ์มาประยุกต์ใช้ โดยทำการออกแบบโครงสร้างออนโทโลยีและจัดเก็บลงใน MongoDB ซึ่งเป็นฐานข้อมูลไม่สัมพันธ์ ผลสรุปว่าการใช้ฐานข้อมูลไม่สัมพันธ์นั้นใช้เวลาในการประมวลผลได้เร็วขึ้นภายในเวลา 36.42 วินาที ซึ่งได้ประสิทธิภาพดีกว่าระบบฐานข้อมูลสัมพันธ์ ในปี พ.ศ. 2559 พบว่า โชติชัย พลอยอรุณรุ่ง และชุมพล โมฆรัตน์ [8] ได้นำเสนอการปรับแต่งประสิทธิภาพระบบงานทะเบียนกับฐานข้อมูลโนเอสคิว

แอลด้วยการจัดเก็บและประมวลผล ผู้วิจัยได้นำเสนอ แนวคิดการปรับแต่งประสิทธิภาพข้อมูลระบบงาน ทะเบียนกับฐานข้อมูลโนเอสคิวแอล เพื่อการแลกเปลี่ยน ข้อมูลข้ามเครื่องแม่ข่าย ที่มุ่งเน้นการจัดเก็บและ ประมวลผลข้อมูลให้มีประสิทธิภาพเพิ่มมากขึ้น และในปี พ.ศ. 2562 สมเกียรติ คอนทองแดง [9] นำเสนอการโจมตี ด้วย DDoS สามารถสร้างความเสียหายให้กับระบบ เครือข่ายอย่างมหาศาลเนื่องจากระบบเครือข่ายที่มีขนาด ใหญ่ ซึ่งใช้ Hadoop Platform ในการจัดการกับ Big Data จัดเก็บด้วย HDFS และคัดกรองข้อมูลด้วย MapReduce การทดลองจะเกิดขึ้นด้วยการนำ NetFlow log file ในระบบเครือข่าย UniNet ที่มีการโจมตีด้วย DDoS ประเมินผ่านอัลกอริทึม Vishal Masheshwari โดยจากงานวิจัยที่เกี่ยวข้องดังกล่าว ผู้วิจัยจึงได้สังเกตเห็น แนวทางในการปรับแต่งประสิทธิภาพความพร้อมใช้งาน แอปพลิเคชันงานทะเบียนและสถิตินักศึกษาด้วยมองโกดี บีเรพลิเคชัน

### 3. วิธีการดำเนินงานวิจัย (Research Methods)

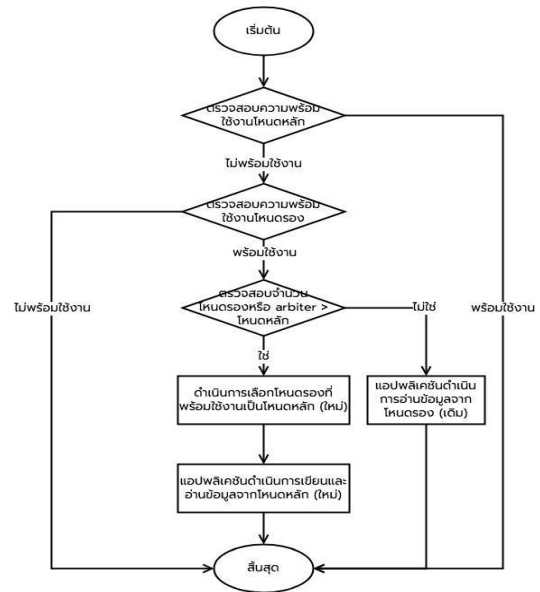
สำหรับวิธีการดำเนินการวิจัยผู้วิจัยได้นำเสนอภาพรวม การดำเนินงานและแสดงกระบวนการแก้ปัญหาจากการ ดำเนินการวิจัย ดังภาพที่ 3 และ 4 ตามลำดับ



ภาพที่ 3: ภาพรวมการดำเนินงานวิจัย

จากภาพที่ 3 แสดงภาพรวมการดำเนินงานวิจัย ซึ่งใน การดำเนินงานวิจัยการเพื่อเพิ่มความพร้อมใช้งาน

ระบบงานทะเบียนและสถิตินักศึกษาด้วยเทคนิคเรพลิเคชัน

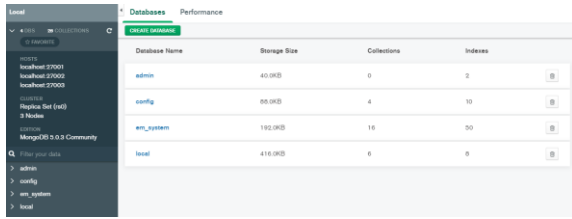


ภาพที่ 4: กระบวนการแก้ปัญหาจากการดำเนินการวิจัย

หรือกระบวนการทำงานร่วมและการเปลี่ยนข้อมูลข้าม เครื่องแม่ข่ายในฐานข้อมูลโนเอสคิวแอลนั้น ประกอบด้วย ขั้นตอน 3 ขั้นตอน ดังนี้ 1. การติดตั้ง Replica Set 2. การ ตั้งค่าการเชื่อมต่อเว็บแอปพลิเคชันกับฐานข้อมูล 3. ขั้นตอนการนำเสนอผลลัพธ์ ตามลำดับ และภาพที่ 4 กระบวนการแก้ปัญหาจากการดำเนินการวิจัย ดัง รายละเอียด

#### 3.1 การติดตั้ง Replica Set

ในขั้นตอนนี้จะเป็นการติดตั้งจัดวาง Replica Set สำหรับจับกลุ่มโหนดต่างๆ ของโครงสร้างการพัฒนาแอปพลิเคชันเพื่อให้สามารถทำงานร่วมกันและแลกเปลี่ยน ข้อมูลระหว่างโหนดได้ ซึ่งเริ่มจากการกำหนดชื่อของ Replica Set แล้วดำเนินการการเริ่มต้น Instance ที่เป็น โหนดหลัก จากนั้นเริ่มต้นการทำงานของ Replica Set จาก Instance ที่ทำหน้าที่เป็นโหนดหลัก และดำเนินการ เพิ่ม Instance ที่เป็นโหนดรองเข้าไปยัง Replica Set โดย เมื่อดำเนินการตามกระบวนการข้างต้นเสร็จสิ้น สามารถ ผลลัพธ์ดังภาพที่ 5



ภาพที่ 5: ผลลัพธ์จากการดำเนินการติดตั้งและจัดวาง Replica Set

### 3.2 การตั้งค่าการเชื่อมต่อเว็บแอปพลิเคชันกับฐานข้อมูล

ขั้นตอนนี้เป็นขั้นตอนการตั้งค่าการเชื่อมต่อของเว็บแอปพลิเคชันกับฐานข้อมูล โนเอสคิวแอลผ่านทาง Drivers โดยการกำหนดเส้นทางในการเชื่อมต่อไปยังโหนดต่าง ๆ เพื่อนำทางให้เว็บแอปพลิเคชันให้สามารถเชื่อมต่อกับโหนดต่าง ๆ ได้ โดยจะกำหนดให้เว็บแอปพลิเคชันสามารถเขียนบันทึกข้อมูลได้เพียงแค่โหนดหลักเท่านั้น และตั้งค่าให้เว็บแอปพลิเคชันติดต่อกับโหนดรองได้ ในกรณีที่โหนดหลักไม่สามารถใช้งานได้ ซึ่งพัฒนาโดยภาษา Python ด้วย Django Framework ดังภาพที่ 6



ภาพที่ 6: การตั้งค่าการเชื่อมต่อเว็บแอปพลิเคชันกับฐานข้อมูล

### 3.3 ขั้นตอนการนำเสนอผลลัพธ์

ในขั้นตอนการนำเสนอผลลัพธ์ของข้อมูลนั้น ทีมผู้วิจัยดำเนินการแสดงข้อมูลที่เป็นผลลัพธ์จากขั้นตอนการติดตั้งและจัดวาง Replica Set และขั้นตอนการตั้งค่าการเชื่อมต่อเว็บแอปพลิเคชันกับฐานข้อมูล โดยเน้นการนำเสนอผ่านทางเว็บแอปพลิเคชันในส่วนของฟังก์ชันการลงทะเบียนรายวิชาบนแอปพลิเคชันงานทะเบียนและสถิติ

นักศึกษา และแสดงผลข้อมูลที่ถูกจัดเก็บในฐานข้อมูล ดังภาพที่ 7

ภาพการศึกษาที่ 1/2564

รายวิชา	เรียนวิชา	หน่วยกิต	กลุ่ม	แบบ	เวลา
0033001-1[นศ]	Chinese for Communication	3	2	GD	WE 13:00-16:00 324
0406204-2[นศ]	Web Technology	3	1	GD	TH 13:00-16:00 866
0406205-2[นศ]	Database System	3	1	GD	TU 09:00-12:00 852
0406306-2[นศ]	Software Engineering	3	1	GD	WE 09:00-12:00 853
0406209-1[นศ]	Computer Network 1	3	1	GD	TU 13:00-16:00 842

จำนวนหน่วยกิตรวม 15

ภาพที่ 7: หน้าจอการลงทะเบียนของนักศึกษาในแอปพลิเคชัน

### 3.4 การประเมินความน่าเชื่อถือ

การประเมินความน่าเชื่อถือของของระบบนั้น ผู้วิจัยได้ประเมินค่าเฉลี่ยของเวลาซ่อมแซม (Mean Time to Repair: MTTR) ที่ซึ่งหาค่าเฉลี่ยในการซ่อมแซมระบบจากระยะเวลาการซ่อมแซมระบบต่อจำนวนครั้งที่ซ่อมแซมระบบในกรณีที่ระบบหยุดทำงานหรือเกิดปัญหาขึ้น จากภาพที่ 9 มีการซ่อมแซมระบบ 12 ครั้ง เนื่องจากระบบเกิดข้อผิดพลาด ดังสมการที่ 3-1

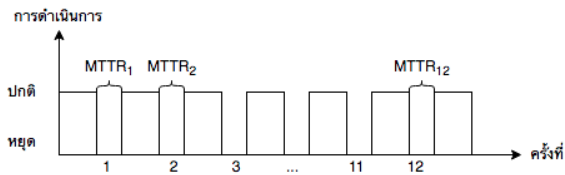
$$MTTR = T / N \tag{3-1}$$

โดย MTTR หมายถึง ค่าเฉลี่ยของเวลาการซ่อมแซมระบบ Total Time to Repair (T) หมายถึง ระยะเวลาการซ่อมแซมระบบทั้งหมด และ Number of Repairs (N) หมายถึง จำนวนการซ่อมแซมระบบที่เกิดขึ้น ตามลำดับ

## 4. ผลการดำเนินงานวิจัย (Research Results)

จากการดำเนินงานวิจัย ผู้วิจัยได้ทดสอบใช้งานระบบงานทะเบียนผ่านเว็บแอปพลิเคชันระบบงานทะเบียนและสถิตินักศึกษาที่เชื่อมต่อ Replica Set ด้วยการดำเนินการลงทะเบียนรายวิชา ภายใต้สถานการณ์เครื่องแม่ข่ายใด ๆ ไม่สามารถทำงานได้ พบว่าระบบยังคงสามารถดำเนินการลงทะเบียนได้ตามปกติและข้อมูลที่ถูกจัดเก็บภายในระบบฐานข้อมูลมีความถูกต้อง และสอดคล้องกัน จากนั้นผู้วิจัยได้ทดสอบหยุดการทำงาน Instance ที่เป็นโหนดหลัก และดำเนินการสังเกตโหนดรอง พิจารณาวามีโหนดใดขึ้นมาทำหน้าที่เป็นโหนดหลักแล้วดำเนินการเริ่มการทำงาน Instance เดิมที่ได้มีการหยุดการทำงานไปแล้วขึ้นมาใหม่อีกครั้ง จากนั้นจึงดำเนินการ

หยุดการทำงาน Instance โหนดหลัก พร้อมดำเนินการบันทึกเวลาการดำเนินการในรูปแบบวนซ้ำจำนวน 12 ครั้ง โดยระบบมีระยะเวลาเฉลี่ยในการซ่อมแซมระบบประมาณ 4 วินาที ซึ่งพบว่าใช้เวลาในการซ่อมแซมก่อนข้างต้นและระบบสามารถดำเนินการได้อย่างต่อเนื่องจากการทดสอบดังกล่าวได้ผลลัพธ์การทดสอบ ดังภาพที่ 8



ภาพที่ 8: ภาพเวลาดำเนินการเริ่ม/หยุดทำงาน Instance โหนดหลัก

### 5. สรุปผลการดำเนินงานวิจัย (Research Conclusions)

จากผลการดำเนินการวิจัยสรุป พบว่าการนำเทคนิคเรพลิเคชันและกระบวนการทำงานร่วมกันและแลกเปลี่ยนข้อมูลระหว่างเครื่องแม่ข่าย โดยใช้งานแอปพลิเคชันงานทะเบียนและสถิตินักศึกษาผ่านเว็บแอปพลิเคชันได้อย่างต่อเนื่อง แม้ว่ามีการใช้งานภายใต้สถานการณ์ที่โหนดหลักอย่างน้อยหนึ่งโหนดไม่สามารถทำงานได้ โดยประเมินค่าเฉลี่ยของเวลาซ่อมแซม ซึ่งพบว่าใช้เวลาในการซ่อมแซมก่อนข้างต้นและระบบสามารถดำเนินการได้อย่างต่อเนื่อง

### 6. ข้อเสนอแนะการดำเนินงานวิจัย (Research Suggestions)

จากบทความที่นำเสนอ ทีมผู้วิจัยได้นำเสนอเฉพาะส่วนของการดำเนินงานกับเครื่องคอมพิวเตอร์ใด ๆ ที่ตั้งค่าให้ใช้งานและทดสอบกระบวนการทำงานร่วมและแลกเปลี่ยนข้อมูลข้ามเครื่องแม่ข่าย ในอนาคตต้องดำเนินการทดสอบกับเครื่องแม่ข่ายจริงของระบบงานทะเบียนและสถิตินักศึกษา ตลอดจนการดำเนินการที่มุ่งเน้นการนำเสนอให้แอปพลิเคชันและทดสอบกับการดำเนินการกับฟังก์ชันอื่นของแอปพลิเคชัน มีความน่าเชื่อถือและสมบูรณ์ยิ่งขึ้น และครอบคลุมต่อการใช้งาน

### เอกสารอ้างอิง

- [1] L. Moore. “Distributed Database.” Available Online at <https://searchoracle.techtarget.com/definition/distributed-database>, accessed on November 2021.
- [2] MongoDB, Inc. “Replication – MongoDB Manual.” Available Online at <https://docs.mongodb.com/manual/replication/>, accessed on 17 January 2022.
- [3] A. Meysman. “NoSQL Database Types.” Available Online at <https://dzone.com/articles/nosql-database-types-1/>, accessed on November 2021.
- [4] PSU IT Blog. “What is MongoDB?.” Available Online at <https://sysadmin.psu.ac.th/2017/01/11/what-is-mongodb/>, accessed on November 2021.
- [5] MariaDB. “Database Master-Slave Replication in the Cloud.” Available Online at <https://mariadb.com/resources/blog/database-master-slave-replication-in-the-cloud/>, accessed on November 2021.
- [6] นิรุทธ์ รวยริน. และเกรียงไกร ปอแก้ว. “การใช้เมพริคิวซ์เชื่อมคอลเลคชันของฐานข้อมูลโนเอสคิวแอลบนมองโกดีบี.” *การประชุมวิชาการเสนอผลงานวิจัยระดับบัณฑิตศึกษา ครั้งที่ 15 (The 15th Graduate Research Conferences: GRC2014)*, เมื่อวันที่ 28 มีนาคม 2557, หน้า 342-350, มหาวิทยาลัยขอนแก่น, 2557.
- [7] ศุภดี บุญรอด. และประกายมาศ ศรีสุขทักษิณ. “การค้นคืนข้อมูลขนาดใหญ่โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้างร่วมกับเทคโนโลยีเว็บเชิงความหมาย.” *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 25, ฉบับที่ 2, หน้า 255 - 264, พ.ศ. - ส.ศ., 2558.
- [8] โชติชัย พลอยอรุณรุ่ง. และชุมพล โมฆรัตน์. “การปรับแต่งประสิทธิภาพระบบงานทะเบียนกับฐานข้อมูลโนเอสคิวแอลด้วยการจัดเก็บและประมวลผล.” *การประชุมวิชาการระดับปริญญาตรีด้านคอมพิวเตอร์ภูมิภาคอาเซียน ครั้งที่ 4 (AUCC 2016)*, วันที่ 27-29 เมษายน 2559, หน้า 997-1003, สระแก้ว: มหาวิทยาลัยบูรพา, 2559.
- [9] สมเกียรติ ดอนทองแดง. “การพัฒนาต้นแบบระบบวิเคราะห์เครือข่ายที่มีความซับซ้อนสูงด้วยข้อมูลขนาดใหญ่.” *วิทยานิพนธ์ปรัชญาดุษฎีบัณฑิต, นครปฐม: มหาวิทยาลัยศิลปากร, 2562.*



## การพัฒนาระบบสารสนเทศการให้บริการทางด้านเทคโนโลยี A Developing of IT Service Management System

วุฒินันท์ ไชยพรหมนิมิต (Wuttinun Chaipromnimit)<sup>1</sup> และณัฐวี อดตฤกษ์ (Nattavee Utakrit)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

s6307021858261@email.kmutnb.ac.th<sup>1</sup>, nattavee.u@itd.kmutnb.ac.th<sup>2</sup>

### บทคัดย่อ

ปัจจุบันองค์กรหลายแห่งต่างได้ให้ความสำคัญกับมาตรฐานการบริหารจัดการการให้บริการเทคโนโลยีสารสนเทศของฝ่าย IT Helpdesk และฝ่าย IT service ของตนเองโดยการปฏิบัติตามมาตรฐาน ITIL 8 กระบวนการงานวิจัยนี้ได้นำเสนอระบบสารสนเทศเพื่อสนับสนุนการให้บริการทางด้านเทคโนโลยีสำหรับองค์กรตามกระบวนการของ ITIL ในรูปแบบเว็บแอปพลิเคชัน เพื่อให้การดำเนินงานด้านการบริการดังกล่าวเป็นไปตามมาตรฐานอย่างมีประสิทธิภาพ เพื่อให้ผู้ใช้งานมีความสะดวกในการใช้งาน และมีความพึงพอใจ ซึ่งผลการประเมินประสิทธิภาพจากการทดสอบแบบ Black-Box Testing ด้วยวิธี UAT โดยผู้เชี่ยวชาญจำนวน 3 คน สรุปได้ว่าระบบสารสนเทศในการให้บริการทางด้านเทคโนโลยีมีประสิทธิภาพในระดับ 100% ในส่วนผลการประเมินความพึงพอใจ โดยผู้ใช้งานในภาพรวมอยู่ในระดับดีมากที่สุด

**คำสำคัญ:** การให้บริการทางด้านเทคโนโลยี กรอบงาน ITIL เว็บแอปพลิเคชัน

### Abstract

At the present, many organizations give priority to the standard of IT service management System of IT Helpdesk and IT Service by applying as the standard of ITIL of 8 processes. This research presented a web application to support IT service management for organizations based on ITIL in order to operate IT service effectively and to provide users with convenience and satisfaction. The results tested by the UAT method from 3 experts can be

concluded that this system is effective at 100% and the users' satisfactions were very high level.

**Keywords:** IT Service Management , ITIL Web Application.

### 1. บทนำ

การบริการลูกค้า (Customer Service) เป็นเรื่องสำคัญสำหรับทุกธุรกิจเพราะเป็นเครื่องมือหนึ่งในการสร้างความแตกต่างและความสามารถในการแข่งขันการบริการลูกค้า ที่มีคุณภาพยังช่วยผูกใจลูกค้าไว้ได้ในระยะยาว และเพิ่มคุณค่าให้กับกิจการ การบริการลูกค้าอย่างไรทิศทางจากการขาดการวางแผนที่ดี ทำให้กิจการถดถอยลงและยากที่จะรักษาลูกค้าเอาไว้ได้ในยุคที่ลูกค้ามีทางเลือกมากมายแทบจะไม่จำกัด เมื่อทุกระบบยังตอบสนองต่อความต้องการได้ตลอดเวลาที่ต้องการและการดำเนินการต่าง ๆ ในทางธุรกิจยังคงราบรื่น เทคโนโลยีสารสนเทศมักถูกมองข้ามความสำคัญเสมอ แต่เมื่อใดก็ตามที่เทคโนโลยีสารสนเทศเกิดปัญหาทำให้กิจกรรมต่าง ๆ ชงักหรือขั้นตอนต่าง ๆ ช้าลง เกิดผลกระทบต่อธุรกิจโดยตรงและทางอ้อม ถ้ามีระบบการบริการจัดการเทคโนโลยีสารสนเทศที่ดีปัญหาเหล่านั้นก็สามารถแก้ไขได้อย่างรวดเร็ว แต่หากไม่มีการกำหนดข้อตกลงต่าง ๆ กันมาก่อนปัญหาเล็กน้อยจะถูกสะสมกลายเป็นปัญหาใหญ่ [1]

ในองค์กรที่มีระบบสารสนเทศที่หลากหลายจำเป็นต้องมีการบริหารจัดการ การให้บริการเทคโนโลยีสารสนเทศที่ดีเพื่อให้ทุกระบบสามารถทำงานได้อย่างต่อเนื่องและมีประสิทธิภาพ ซึ่งเมื่อระบบเทคโนโลยีสารสนเทศทำงานได้ดีเป็นปกติ การดำเนินธุรกิจในส่วนที่พึ่งพาระบบสารสนเทศย่อมราบรื่น แน่ใจว่าการใช้งาน

ระบบสารสนเทศทุกระบบไม่สามารถใช้งานได้อย่างต่อเนื่องตลอดเวลา 24 ชั่วโมงไม่มีวันหยุด บางครั้งอาจมีปัญหาติดขัดเกิดขึ้นได้การบริหารจัดการ การให้บริการ เทคโนโลยีสารสนเทศในส่วนการแก้ปัญหา การตรวจสอบการเกิดปัญหาซ้ำ การดำเนินการแก้ไขหรือการปรับปรุงระบบต่างๆ การทำข้อตกลงการให้บริการ การกำหนดระดับของการให้บริการ อีกทั้งระบบที่มีอยู่ในปัจจุบันมีราคาค่อนข้างสูง เช่น ระบบ Service now เป็นต้น ระบบส่วนใหญ่เป็นของต่างประเทศและเป็นภาษาอังกฤษ และฟังก์ชันไม่ยืดหยุ่นจึงพัฒนาระบบมาให้เหมาะสมกับการใช้งานในองค์กรได้อย่างมีประสิทธิภาพ [2]

องค์กรในปัจจุบันนิยมนำเทคโนโลยีสารสนเทศมาใช้ในการให้บริการลูกค้าหรือผู้ใช้งานมากขึ้นเพื่อให้เกิดความพึงพอใจ แต่เนื่องจากเทคโนโลยีในปัจจุบันมีการเปลี่ยนแปลงอยู่ตลอดเวลาทำให้เทคโนโลยีสารสนเทศต้องมีการพัฒนาและปรับปรุงการให้บริการต่างที่กัน เพื่อตอบสนองความที่ต้องการและทำให้เกิดความพึงพอใจของผู้ใช้งานที่แตกต่างกัน โดยเน้นไปที่คุณภาพในการให้บริการและประสิทธิภาพในการปฏิบัติงาน จึงได้มีการนำเอากระบวนการทำงานทางเทคโนโลยีสารสนเทศมาใช้เพื่อปรับปรุงและพัฒนาในการปฏิบัติงานให้มีคุณภาพและประสิทธิภาพมากยิ่งขึ้นเพื่อเป็นการยกระดับการให้บริการด้านเทคโนโลยีสารสนเทศตามหลักการของ IT Service Management หรือ ITSM ซึ่งมุ่งเน้นการพัฒนากระบวนการบริหารจัดการทางเทคโนโลยีสารสนเทศให้มีประสิทธิภาพมากขึ้นมีหลากหลาย กรอบงาน (Framework) ที่องค์กรสามารถนำมาประยุกต์ใช้ได้แก่ ITIL Framework ของ OGC, CobiT และ Val IT Framework ของ ISACA, ITUP ของ IBM, และ MOF ของ Microsoft เป็นต้น [3]

เว็บแอปพลิเคชัน หรือ โปรแกรมปรับใช้บนเว็บไซต์ถูกพัฒนาขึ้นมาเพื่อใช้งานผ่านเบราว์เซอร์ การใช้งานในองค์กรทำได้ง่าย ข้อมูลจัดเก็บที่เดียว ง่ายต่อการจัดการ และไม่เกิดความซ้ำซ้อน สามารถทำงานได้ทุกที่ทุกเวลา เชื่อมต่อกับบริการออนไลน์อื่นๆ ได้ง่าย ซึ่งองค์กรขนาดเล็กสามารถใช้งานได้เพราะมีค่าใช้จ่ายต่ำ [4] โดยในการวิจัยนี้ผู้วิจัยได้นำเสนอรูปแบบวิธีการดำเนินการตามกระบวนการทำงานระบบสามารถทำงานได้ตรงตามมาตรฐาน ITIL สำหรับหน่วยงานในฝ่าย IT Helpdesk

และ IT service ขององค์กรขนาดเล็ก โดยมุ่งเน้นการแก้ไขกระบวนการ ตามมาตรฐานเพื่อแก้ไขข้อจำกัดในการให้บริการ และเป็นแนวทางในการพัฒนาระบบให้บริการทางด้านเทคโนโลยีสารสนเทศให้สามารถเข้าถึงบริการได้สะดวกรวดเร็ว สามารถแก้ไขปัญหาได้ตรงประเด็นตรงจุด มีความยืดหยุ่นในการใช้งาน ระบบสามารถปรับเปลี่ยนได้ตามความเปลี่ยนแปลงขององค์กรตรงตามความต้องการขององค์กรลดค่าใช้จ่ายรองรับภาษาไทย มีความปลอดภัยในการใช้งาน ข้อมูลสามารถตรวจสอบและติดตามงานได้ ลดภาระการทำงานของ IT Helpdesk และ IT Service ได้ โดยได้พัฒนาในรูปแบบเว็บแอปพลิเคชัน มุ่งเน้นการแก้ไขกระบวนการตามมาตรฐานเพื่อแก้ไขข้อจำกัดในการให้บริการ และเป็นแนวทางในการพัฒนาระบบให้บริการทางด้านเทคโนโลยีสารสนเทศให้สามารถเข้าถึงบริการได้สะดวก รวดเร็วสามารถแก้ไขปัญหาได้ตรงประเด็นตรงจุดมีความยืดหยุ่นในการใช้งาน ระบบสามารถปรับเปลี่ยนได้ตามความเปลี่ยนแปลงขององค์กรตรงตามความต้องการขององค์กรลดค่าใช้จ่าย รองรับภาษาไทย มีความปลอดภัยในการใช้งาน ข้อมูลสามารถตรวจสอบและติดตามงานได้ลดภาระการทำงานของ IT Helpdesk และ IT Service ได้รวมถึงกระบวนการทำงานระบบสามารถทำงานได้ตรงตามมาตรฐาน ITIL 8 กระบวนการ ประกอบด้วย Change Management, Release and Deployment Management, Access Management, Problem Management, Incident Management, Service Reporting, Service Improvement and Service Request

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

จากงานวิจัยที่ผ่านมาพบว่า ปัจจุบันเทคโนโลยีและกระบวนการจัดการมีการพัฒนาขึ้นเป็นจำนวน ดังนั้นผู้ประกอบการที่มีความต้องการเพิ่มศักยภาพในการพัฒนาคุณภาพการบริการ คุณภาพผลิตภัณฑ์ เพื่อสร้างความได้เปรียบในการแข่งขันต้องใช้เทคโนโลยีสารสนเทศและหลักการบริหารที่เน้นความพึงพอใจของผู้บริโภค [5]

ระบบสารสนเทศมีความสำคัญในกิจกรรมการจัดการทำงาน ทั้งการวางแผนการทำงาน การตั้งการควบคุมการทำงาน โดยระบบสารสนเทศนั้นสามารถจัดเก็บข้อมูล บันทึกข้อมูล วิเคราะห์ผล แปลผลผลลัพธ์ และ

ส่งกลับข้อมูลสารสนเทศได้อย่างสะดวกรวดเร็ว ระบบสารสนเทศนั้นเปรียบเสมือนเครื่องมือในการดำเนินการเชื่อมช่องทางข้อมูลของเครือข่ายสังคมออนไลน์ ให้ผู้ใช้ได้เข้าถึงข้อมูลทางสังคมร่วมกัน อีกทั้งเครือข่ายสื่อสารออนไลน์ช่วยส่งเสริมให้ผู้ใช้มีความต้องการแลกเปลี่ยนเรียนรู้กันมากขึ้น เทคโนโลยีสารสนเทศมีบทบาทอย่างยิ่งเนื่องจากสามารถทำให้องค์กรดำเนินงานได้อย่างมีประสิทธิภาพด้วยความรวดเร็ว [6] ระบบเทคโนโลยีถือว่ามีความสำคัญช่วยให้องค์กรทราบถึงสภาพและการดำเนินงานขององค์กร การพัฒนาและออกแบบระบบเพื่อให้เกิดการเปลี่ยนแปลงโครงสร้างขององค์กร [7]

### 3. วิธีดำเนินการวิจัย

#### 3.1. การศึกษาความเป็นไปได้

ผู้วิจัยได้ทบทวนความรู้เกี่ยวกับมาตรฐาน ITIL โดยนำมาประเมินโอกาสความเป็นไปได้ ในการนำมาตรฐาน ITIL มาใช้ในการพัฒนาระบบ ทบทวนความรู้เกี่ยวกับกระบวนการมาตรฐาน ITIL ทั้งงานวิจัยไทย และต่างประเทศ เพื่อนำมาประเมินโอกาสความเป็นไปได้มาประยุกต์ใช้ในองค์กรปัจจุบัน ศึกษาเครื่องมือที่นำมาใช้ในการพัฒนาระยะยาว

#### 3.2 การวิเคราะห์และออกแบบระบบ

ศึกษาปัญหา วิธีการ และความต้องการของผู้ใช้งาน ในขั้นตอนนี้ผู้วิจัยได้ทำการ ศึกษาปัญหาที่เกิดขึ้นจากแหล่งข้อมูล 2 แหล่ง

ศึกษาจากแหล่งข้อมูลทุติยภูมิ (Secondary Source) เป็นการศึกษาจากบทความหนังสือ และงานวิจัยที่เกี่ยวข้องเพื่อศึกษาปัญหาที่เกิดขึ้นกับระบบดังกล่าวในปัจจุบัน และแนวโน้มของเทคโนโลยีในการพัฒนาระบบสารสนเทศด้านให้บริการทางด้านเทคโนโลยีตามมาตรฐาน ITIL รวมถึงการเปลี่ยนแปลงที่อาจเกิดขึ้นในอนาคต

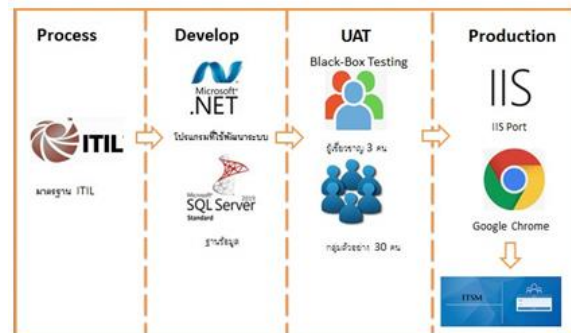
ศึกษาจากแหล่งข้อมูลปฐมภูมิ (Primary Source) ซึ่งเป็นแหล่งข้อมูลที่ได้โดยการออกแบบสอบถามไปยังพนักงาน และบุคลากรที่เป็นกลุ่มตัวอย่าง

วิเคราะห์ความต้องการของผู้ใช้งาน ในขั้นตอนนี้ผู้วิจัย ได้ทำการการออกแบบ สอบถามเกี่ยวกับปัญหาและความต้องการระบบไปยัง พนักงาน และบุคลากรที่เป็นกลุ่มตัวอย่าง

ขั้นต่อมา นำข้อมูลที่ได้รับมาวิเคราะห์โดยจัดลำดับตามความต้องการของผู้ใช้งานมาออกแบบ และออกแบบภาพรวมของระบบสารสนเทศดังภาพที่ 1.2 ก่อนที่จะทำการพัฒนาระบบและสร้างให้อยู่ในรูปแบบเว็บแอปพลิเคชัน

หัวข้อ	ITIL	ISO
ลดต้นทุนค่าใช้จ่ายด้านเทคโนโลยี	☑	☒
ลดความคิดพลาด	☑	☑
เป็นที่ยอมรับของลูกค้าและผู้ให้บริการ	☑	☒
ระบบมีการปรับปรุงพัฒนาอย่างต่อเนื่อง	☑	☑
ลดความซ้ำซ้อนในการทำงาน	☑	☑
ด้านความปลอดภัย	☑	☑
ปรับปรุงกระบวนการให้บริการด้านเทคโนโลยีสารสนเทศ	☑	☑
การใช้ทรัพยากรที่มีอยู่อย่างคุ้มค่า	☑	☑
มาตรฐานสากล	☑	☑
ช่วยให้องค์กร มีความคล่องตัวและประสิทธิภาพมากขึ้น	☑	☒

ภาพที่ 1.1: ภาพรวมเปรียบเทียบ ITIL และ ISO

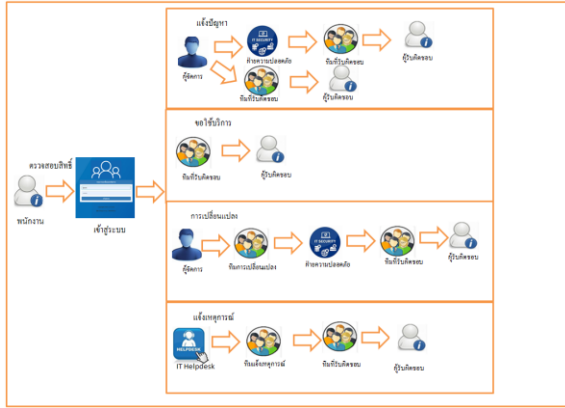


ภาพที่ 1.2: ภาพรวมกระบวนการพัฒนาระบบ



ภาพที่ 1.3: ภาพรวมกระบวนการพัฒนาระบบผ่าน ITIL

กระบวนการ ITIL ทั้ง 8 กระบวนการที่เลือกนำมาใช้ จะสามารถอธิบายความสัมพันธ์ของกระบวนการ ITIL และการทำงานของระบบ เช่น Assess Management กับ การตรวจสอบการเข้าใช้งานของระบบ Problem Management ในส่วนของเมนูการแจ้งปัญหาเป็นต้น



ภาพที่ 1.4: กระบวนการทำงานของระบบในส่วนต่าง ๆ

กระบวนการทำงานของระบบในการขอใช้บริการ มี 4 ส่วน 1) แจ้งปัญหา 2)ขอใช้บริการ 3) ขอการเปลี่ยนแปลง 4) แจ้งเหตุการณ์ ซึ่งการทำงานในแต่ละส่วน จะมีกระบวนการขั้นตอนที่มีความแตกต่างกัน เช่น การแจ้งเหตุการณ์ ( Incident) ก็จะมีลำดับขั้นตอนดังนี้ พนักงานแจ้งปัญหาเข้ามาในระบบ IT Helpdesk รับปัญหา ตรวจสอบประสานงานแจ้งทีม Incident ทางทีม Incident รับทราบปัญหาและส่งงานต่อให้ทีมที่รับผิดชอบส่วนงานที่รับผิดชอบ หลังจากดำเนินการแก้ไขเรียบร้อยแล้วก็จะสามารถปิดงานผ่านระบบ เป็นต้น

### 3.3 การประเมินผลระบบ

3.3.1 ผู้วิจัยทำการตรวจสอบการใช้งานของโปรแกรมเพื่อหาข้อผิดพลาดของโปรแกรม จากนั้นทำการตรวจสอบโปรแกรมที่พัฒนาเสร็จแล้วว่าสามารถดำเนินการได้ตาม Process และหลักการของ ITIL

#### 3.3.2 การประเมินผล

ประเมินผลระบบโดยผู้เชี่ยวชาญ นำผลที่ได้จากการทดสอบประสิทธิภาพกระบวนการเพิ่มประสิทธิภาพของระบบให้บริการทางด้านเทคโนโลยีสารสนเทศตามมาตรฐาน ITIL ในงานวิจัยนี้ใช้ User Acceptance Test (UAT)

3.3.3 ประเมินความพึงพอใจของระบบ โดยผู้ใช้งานทั่วไป

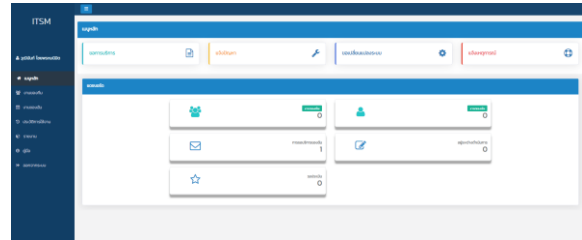
## 4. ผลการดำเนินงาน

### 4.1 ผลการพัฒนาาระบบสารสนเทศการให้บริการทางด้านเทคโนโลยี

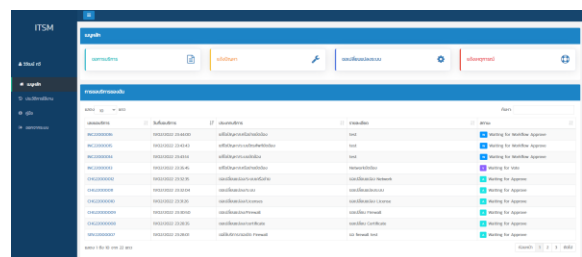
ระบบสารสนเทศที่ได้พัฒนาขึ้นประกอบด้วยระบบการทำงาน 6 หน้าจอเมนูหลักได้แก่ ขอใช้บริการ แจ้งปัญหา ขอเปลี่ยนแปลงระบบ แจ้งเหตุการณ์ ผู้ดูแลระบบ ซอฟต์แวร์ที่มีระยะเวลาในการดำเนินงาน 6 เดือน ดังแสดงในภาพที่ 1.5 ,ภาพที่ 1.6 ,ภาพที่ 1.7 และ ภาพที่ 1.8



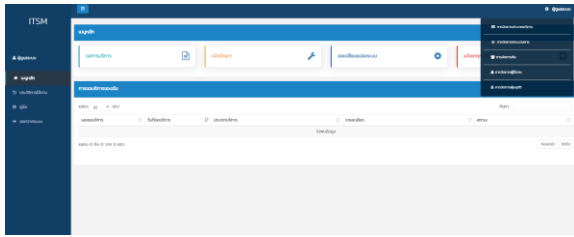
ภาพที่ 1.5: ภาพเข้าสู่ระบบ



ภาพที่ 1.6: ภาพแสดงเมนูระบบ



ภาพที่ 1.7: ภาพแสดงการขอใช้บริการระบบ



ภาพที่ 1.8: ภาพแสดงผู้ดูแลระบบ

#### 4.2 ผลการระบบสารสนเทศการให้บริการทางด้านเทคโนโลยี

4.2.1 ผลการประเมินประสิทธิภาพของระบบ โดยผู้เชี่ยวชาญ มีวัตถุประสงค์เพื่อทดสอบความถูกต้องของระบบระหว่างข้อมูลนำเข้า และผลลัพธ์ที่ได้จากระบบ ด้วยวิธี Black-Box-Testing แบบ UAT ซึ่งทำการประเมินโดยผู้เชี่ยวชาญ จำนวน 3 คน ผลที่ได้สามารถแสดงได้ดังนี้

ตารางที่ 1: ระบบสารสนเทศการให้บริการทางด้านเทคโนโลยี

ส่วนที่ 1 ระบบจัดการข้อมูลพื้นฐาน	คะแนน	ส่วนที่ 2 ระบบจัดการของบริหาร	คะแนน
การจัดการประเภทบริการ	100%	ขอการบริการ	100%
การจัดการกระบวนการ	100%	แจ้งปัญหา	100%
การจัดการทีม	100%	ขอเปลี่ยนแปลงระบบ	100%
การจัดการผู้ใช้งาน	100%	แจ้งเหตุการณ์	100%
การจัดการผู้อนุมัติ	100%	งานของทีม	100%
		การขอบริการของตนเอง	100%
		รอประเมิน	100%
		งานของตนเอง	100%
		การขอบริการของตนเอง	100%

4.2.2 ผลการประเมินความพึงพอใจโดยผู้ใช้งานระบบ 30 คน ทำการแบ่งการประเมินออกเป็น 3 ส่วน ได้แก่ 1 ความพึงพอใจในประสิทธิภาพของระบบสารสนเทศการให้บริการทางด้านเทคโนโลยี 2 การออกแบบของระบบ และส่วนที่ 3 ความปลอดภัยของระบบ

ตารางที่ 2: ผลการประเมินด้านความพึงพอใจของระบบสารสนเทศ

ที่	รายการประเมินความพึงพอใจ	ผลการวัด		ความหมาย
		$\bar{X}$	S.D.	
	ด้านประสิทธิภาพ	4.59	0.26	มากที่สุด
	ด้านการออกแบบ	50.4	14.0	มากที่สุด
	ด้านความปลอดภัย	53.4	24.0	มากที่สุด

#### 4.3 ข้อคิดเห็นจากการประเมินผลระบบระบบ

##### สารสนเทศการให้บริการทางด้านเทคโนโลยี

ผู้เชี่ยวชาญในการประเมินระบบสารสนเทศการให้บริการทางด้านเทคโนโลยี ได้ให้คำแนะนำว่า ระบบสารสนเทศที่ผู้วิจัยพัฒนาขึ้นควรนำไปต่อยอดในส่วนของการบริหารงานด้านต่าง ๆ เช่น การวางแผนงบประมาณ การบริหารงานบุคคล การบริหารงานด้านการขนส่ง เป็นต้น และระบบสารสนเทศที่พัฒนาขึ้นยังมีการใช้งานที่ซับซ้อนอาจจะส่งผลกระทบต่อความเข้าใจในการใช้งาน จึงควรมีการอบรมเกี่ยวกับการใช้งานสำหรับพนักงานใหม่ และควรมีการแจ้งเตือนเพิ่มเติมเมื่อมีกิจกรรมสำคัญ ที่ได้มีการบันทึก เช่น การบันทึกแผนโครงการ การส่งแผนดำเนินงาน เป็นต้น

### 5. สรุป อภิปรายผล

งานวิจัยนี้ ได้การพัฒนา ระบบสารสนเทศการให้บริการทาง ด้านเทคโนโลยีตามกรอบ ITIL ในรูปแบบเว็บแอปพลิเคชัน เพื่อช่วยให้การทำงานให้มีประสิทธิภาพ การพัฒนาระบบ โดยใช้ภาษา Microsoft dot net เชื่อมต่อจากระบบจัดการฐานข้อมูล SQL Server และทดสอบการขอใช้บริการผ่านเมนูต่าง ๆ ของระบบ ให้บริการทางด้านเทคโนโลยี ซึ่งสอดคล้องกับ นงลักษณ์ ซึ่งพบว่า มีการเปลี่ยนแปลงไปในทางที่ดีขึ้นเช่นกัน ตั้งแต่การจัดเก็บข้อมูลที่เป็นระบบมากขึ้น ง่ายต่อการสืบค้น และการช่วยในการตรวจสอบสถานะการดำเนินงานของบุคลากรภายในหน่วยงาน ซึ่งช่วยให้มีการส่งต่องานกันอย่างเป็นระบบ มีการตรวจสอบงานที่เข้ามาใหม่เสมอ เพื่อทำการมอบหมายงานให้กับผู้ดำเนินงานต่อไป ซึ่งส่วนใหญ่เป็นไปตามเงื่อนไขข้อตกลง (Service Level Agreement) การให้บริการที่ได้กำหนดไว้ และสามารถวัดผลได้ตาม KPI ที่ได้จัดเก็บไว้ใน Service Catalogue [8] ยังสอดคล้องกับ

Rudy Yandri and Suharjo พบว่า การปรับปรุงระดับวุฒิภาวะที่เพิ่มขึ้นใน SLM ด้วย ITIL เข้าใกล้สามารถจัดการได้อย่างมีประสิทธิภาพ [9] เช่นเดียวกับผลการทดลองของพิมพ์ลดา ที่กล่าวว่างานบริหารด้านสารสนเทศมีการเปลี่ยนแปลงไปในทางที่ดีขึ้น ตั้งแต่การจัดเก็บข้อมูลที่เป็นระบบมากขึ้น ง่ายต่อการสืบค้นและการช่วยในการตรวจสอบสถานะการดำเนินงาน ซึ่งช่วยให้มีการส่งต่องานกันอย่างเป็นระบบมีการตรวจสอบงานที่เข้ามาใหม่เสมอ เพื่อทำการมอบหมายงานให้กับผู้ดำเนินงานต่อไป ซึ่งส่วนใหญ่เป็นไปตามเงื่อนไขข้อตกลง (Service Level Agreement) การให้บริการที่ได้กำหนดไว้และสามารถวัดผลได้ตาม KPI ที่ได้จัดเก็บไว้ใน Service Catalogue [10]

## 6. ข้อเสนอแนะในการวิจัย

ในการพัฒนาระบบให้บริการทางด้านเทคโนโลยี คือ ได้ระบบสารสนเทศการให้บริการทางด้านเทคโนโลยีตาม ITIL ในรูปแบบเว็บแอปพลิเคชัน โดยในอนาคตควรมีการขยายกระบวนการอื่นเพิ่มขึ้น เป็นการพัฒนากระบวนการทรัพย์สินในองค์กร ระบบบริหารงานบุคลากร เพื่อเพิ่มประสิทธิภาพในการให้บริการของการพัฒนางานต่อไป

### เอกสารอ้างอิง

- [1] ชนาภา มีศรี .การนำกรอบ ITIL มาประยุกต์ใช้ในการบริหารจัดการระบบสารสนเทศ กรณีศึกษา บริษัท ทริสคอร์ปอเรชั่น จำกัด ,วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต . สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย. กรุงเทพฯ ; มหาวิทยาลัยเทคโนโลยีมหานคร. 2558.
- [2] ชัชวาลย์ ตูฬิจิตร. การปรับปรุงกระบวนการทำงานด้านไอทีด้วย ไอทีล กรณีศึกษา บริษัท คาด้า โปรดักส์ 3 ทอ ป ปี น ฟอรั่ม จำกัด ,วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต . สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย. กรุงเทพฯ ; มหาวิทยาลัยศรีปทุม. 2560.
- [3] วัฒนัยุทธิ์ ขวัญจันทร์. การบริหารจัดการการให้บริการเทคโนโลยีสารสนเทศโดยการประยุกต์ใช้ ITIL. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สาขาวิชาเทคโนโลยีสารสนเทศ, บัณฑิตวิทยาลัย. กรุงเทพฯ : มหาวิทยาลัยเทคโนโลยีมหานคร. 2555.
- [4] ณัฐพันธ์ เขจรนันท์ และไพบุณย์ เกียรติโกมล. ระบบสารสนเทศเพื่อการจัดการ. กรุงเทพฯ : ซีเอ็ดยูเคชั่น. 2552.
- [5] ปิยนันท์ เสนะโท. การพัฒนาระบบสารสนเทศเพื่อการบริหารของศูนย์เทคโนโลยีดิจิทัล. เพชรบุรี : มหาวิทยาลัยราชภัฏเพชรบุรี. 2563.
- [6] ศรีนัย นาคณอม. รูปแบบการบริหารบริการเทคโนโลยีสารสนเทศในสถาบันอุดมศึกษา. ปรัชญาดุษฎีบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสารเพื่อการศึกษา กรุงเทพฯ : มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. 2557.
- [7] ประสาน แก้วก้อน. การประยุกต์กระบวนการ ITIL กับกระบวนการให้บริการรับแจ้งเหตุแก้ไขปัญหา. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สาขาเทคโนโลยีสารสนเทศ, บัณฑิตวิทยาลัย. กรุงเทพฯ : มหาวิทยาลัยเทคโนโลยีมหานคร. 2559.
- [8] นงลักษณ์ ทรพน์นนท์. การบริหารจัดการด้านไอทีโดยการนำไอทีล 3 มาประยุกต์ใช้ในการสนับสนุนระบบการรับแจ้งปัญหา กรณีศึกษา สำนักงานเขต กรุงเทพมหานคร. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สาขาวิชาเทคโนโลยีสารสนเทศ, บัณฑิตวิทยาลัย. กรุงเทพฯ : มหาวิทยาลัยศรีปทุม. 2560.
- [9] Rudy Yandri and Suharjo Evaluation Model for the Implementation of Information Technology Service Management using Fuzzy ITIL, 4<sup>th</sup>. International Conference on Computer Science and Computational Intelligence (ICCSICI). 12–13 September 2019, Indonesia, 2019.
- [10] พิมพ์ลดา สร้อยสนธิ .การบริหารงานด้านการบริการเทคโนโลยีสารสนเทศด้วยกระบวนการ ITIL กรณีศึกษา : ศูนย์เทคโนโลยีสารสนเทศ มหาวิทยาลัยแม่โจ้ววิทยานิพนธ์ . วิทยาศาสตรมหาบัณฑิต, สาขาวิชาเทคโนโลยีสารสนเทศ, บัณฑิตวิทยาลัย. กรุงเทพฯ : มหาวิทยาลัยเทคโนโลยีมหานคร. 2557.

# การพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

## The Development of Asset Trading Platform using Non-Fungible Token

จิตภา คำรงค์สมบัติ (Jidapa Dumrongsombut)<sup>1</sup> และพุดสดี บุญรอด (Pudsadee Boonrawd)<sup>2</sup>

<sup>1,2</sup>ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

s6307011858152@email.kmutnb.ac.th, pudsadee.b@itd.kmutnb.ac.th

### บทคัดย่อ

การซื้อขายสินทรัพย์ดิจิทัลเริ่มเป็นที่นิยมมากขึ้นในปัจจุบัน แต่แพลตฟอร์มส่วนใหญ่ที่ให้บริกรายังไม่มีกระบวนการตรวจสอบความเป็นต้นฉบับของสินทรัพย์ดิจิทัล ส่งผลให้สินทรัพย์ดิจิทัลไม่มีมูลค่า และไม่มีการใช้ประโยชน์จากการเป็นเจ้าของ งานวิจัยนี้จึงได้พัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะบนเครือข่ายอีเธอเรียมที่พัฒนาขึ้นเพื่อรองรับการสร้าง การแสดงรายการขาย การซื้อ และการโอนกรรมสิทธิ์ความเป็นเจ้าของโดยไม่ผ่านตัวกลาง รวมทั้งการได้รับค่าสิทธิความเป็นเจ้าของ และผลตอบแทนจากการซื้อขายสินทรัพย์ดิจิทัล จากผลการวิจัยพบว่า แพลตฟอร์มที่พัฒนาขึ้นสามารถซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะได้ตรงตามความต้องการของผู้ใช้งาน อีกทั้งเป็นแนวทางในการประยุกต์ใช้กับงานวิจัยอื่น ๆ ที่เกี่ยวข้อง และสามารถนำไปต่อยอดในการพัฒนาแอปพลิเคชันแบบกระจายบนเครือข่ายอีเธอเรียมได้ในอนาคต

**คำสำคัญ:** แพลตฟอร์มซื้อขายสินทรัพย์ โทเคนดิจิทัล ลักษณะเฉพาะ อีเธอเรียม

### Abstract

Nowadays, Digital Asset Exchange has gained popularity for investors. Most Digital Asset Exchange Service platforms don't have an authenticity verification process which caused digital assets to have no value and no leverage on digital assets ownership. This research has developed an asset trading platform using Non-Fungible Token on the

Ethereum blockchain to support mint Non-Fungible Token, listing Non-Fungible Token, buying Non-Fungible Token, transferring ownership of assets without an intermediary, obtaining ownership rights, and earning benefits from trading digital assets. The research results found that the developed platform can exchange digital assets using a Non-Fungible Token which serves the needs of users, as well as a guideline for applying to other related research and can be used to develop distributed applications on the Ethereum Blockchain in the future.

**Keywords:** Asset Trading Platform, Non-Fungible Token, Ethereum.

### 1. บทนำ

การใช้เงินตราและสกุลเงินดิจิทัลเพื่อซื้อขายและแลกเปลี่ยนสินทรัพย์เป็นโทเคนดิจิทัลไม่มีลักษณะเฉพาะ (Fungible Token : FT) หรือโทเคนดิจิทัลทดแทนกันได้ ตัวอย่างเช่น บิทคอยน์ (Bitcoin) อีเธอเรียม (Ethereum) และสกุลเงินต่าง ๆ ที่สามารถแลกเปลี่ยนด้วยจำนวนเดียวกัน เนื่องจากมีมูลค่าทดแทนกันได้ ด้วยคุณสมบัติดังกล่าวจึงทำให้สินทรัพย์ดิจิทัลไม่มีมูลค่าและไม่มีความสามารถในการซื้อขาย [1]

ในการซื้อขายและแลกเปลี่ยนสินทรัพย์บนแพลตฟอร์มโดยใช้โทเคนดิจิทัลลักษณะเฉพาะ (Non-Fungible Token : NFT) เพื่อแสดงกรรมสิทธิ์ความเป็นเจ้าของสินทรัพย์ซึ่งผู้ใช้งานมีสิทธิ์ถือครอง และจัดการเคลื่อนย้ายสินทรัพย์ดิจิทัลได้ตามต้องการ รวมถึงการเพิ่มคุณสมบัติให้สินทรัพย์ดิจิทัลมีความเป็นเอกลักษณ์และแตกต่าง

ทำให้ต้องซื้อขายแบบเต็มจำนวนของโทเคนดิจิทัลเท่านั้น โดยพัฒนาแพลตฟอร์มในรูปแบบแอปพลิเคชันแบบกระจาย (Decentralized Applications : DApps) เพื่อจัดเก็บและตรวจสอบข้อมูลการทำธุรกรรมบนบล็อกเชนแบบสาธารณะ (Public Blockchain) ที่ใช้หลักการของสัญญาอัจฉริยะ (Smart Contract) บนเครือข่ายอีเธอร์เรียม (Ethereum Blockchain) [2] ทำให้การซื้อขายดังกล่าวมีความโปร่งใส ปลอดภัย และไร้ตัวกลาง ตัวอย่างผลงานซึ่งเป็นที่นิยมเป็นของสะสม และมีคุณค่าทางจิตใจ เช่น รูปภาพ คลิปวิดีโอ และไอเทมเกม เป็นต้น [3] โดยแต่ละแพลตฟอร์มจะมีค่าธรรมเนียมในการซื้อขาย (Gas Fee) แตกต่างกันไป ซึ่งใช้สกุลเงินอีเทอร์ (Ether : ETH) เป็นสกุลเงินดิจิทัลในการซื้อขายและการชำระค่าธรรมเนียม

จากปัญหาที่กล่าวมาข้างต้น งานวิจัยนี้จึงได้ทำการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัล ลักษณะเฉพาะบนเครือข่ายอีเธอร์เรียมในรูปแบบแอปพลิเคชันแบบกระจายที่ประยุกต์ใช้กระเป๋าเงินดิจิทัลของ Metamask และสกุลเงินอีเทอร์ในกระเป๋าเงินดิจิทัลสำหรับซื้อขายและชำระค่าธรรมเนียม โดยในบทความวิจัยแบ่งเนื้อหาออกได้เป็น 5 ส่วน ได้แก่ ส่วนที่ 1 บทนำ ส่วนที่ 2 ทฤษฎี และงานวิจัยที่เกี่ยวข้อง ส่วนที่ 3 วิธีการดำเนินการวิจัย ส่วนที่ 4 ผลการดำเนินงานวิจัย และส่วนที่ 5 สรุปผลการวิจัย

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 โทเคนดิจิทัลลักษณะเฉพาะ (Non-Fungible Token)

โทเคนดิจิทัลลักษณะเฉพาะแสดงถึงการระบุกรรมสิทธิ์ของสินทรัพย์ ซึ่งมีคุณสมบัติความเป็นเอกลักษณ์ มีคุณค่าทางจิตใจ และเหมาะแก่การเก็บสะสม ตัวอย่างเช่น คลิปวิดีโอ นักกีฬาชนะการแข่งขัน มีม (Meme) รูปภาพ งานศิลปะ และไอเทมในเกม เป็นต้น ซึ่งอยู่ในรูปแบบของสินทรัพย์ดิจิทัล โดยโทเคนดิจิทัลเปรียบเสมือนใบรับรองดิจิทัลที่มีความคล้ายคลึงกับ โฉนดที่ดินที่ได้รับบูรณายถะเอียดความเป็นเจ้าของ และถูกบันทึกบนบล็อกเชน จึงทำให้ธุรกรรมที่เกิดขึ้นมีความโปร่งใส ปลอดภัย และไร้ตัวกลาง

การซื้อขายแลกเปลี่ยนกรรมสิทธิ์ของสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ ผู้ซื้อได้รับกรรมสิทธิ์ความเป็น

เจ้าของสินทรัพย์ จึงเป็นการเปิดโอกาสสร้างรายได้ให้กับผู้สร้างและผู้ขาย ตัวอย่างเช่น การประมูลขายทวิตแรกของ Jack Dorsey โดยผู้ซื้อโทเคนดิจิทัลลักษณะเฉพาะของทวิตดังกล่าวจะได้รับกรรมสิทธิ์ความเป็นเจ้าของเท่านั้น แต่ผู้สร้างทวิตแรกของโลกยังคงเป็น Jack Dorsey รวมทั้งผู้อื่นยังสามารถเข้ารับชมทวิตดังกล่าวได้เช่นเดิม [1]

### 2.2 อีเธอร์เรียมบล็อกเชน (Ethereum Blockchain)

อีเธอร์เรียมบล็อกเชนเป็นแพลตฟอร์มโอเพนซอร์สที่สามารถปรับแก้ไขได้บนบล็อกเชน มีการพัฒนาขึ้นเป็นแอปพลิเคชันแบบกระจาย และดำเนินการธุรกรรมด้วยสกุลเงินอีเทอร์ในการทำธุรกรรม ซึ่งต้องมีการชำระค่าธรรมเนียมให้กับนักพัฒนา และนักขุดที่แก่สมการจนได้รับรางวัล [4]

#### 2.2.1 รูปแบบการใช้งานเครือข่ายอีเธอร์เรียม

รูปแบบการใช้งานเครือข่ายอีเธอร์เรียมสามารถแบ่งออกเป็น 3 รูปแบบ ได้แก่ รูปแบบที่ 1 ระบบการเงินแบบกระจาย (Decentralized Finance : Defi) เป็นรูปแบบการเงินไม่มีตัวกลาง ซึ่งใช้สัญญาอัจฉริยะ และทำงานบนบล็อกเชน ตัวอย่างแพลตฟอร์มของระบบการเงินแบบกระจาย ได้แก่ Uniswap Compound และ MakerDao เป็นต้น รูปแบบที่ 2 โทเคนดิจิทัลลักษณะเฉพาะ ซึ่งได้รับความนิยมและเป็นเรื่องใหม่ในช่วงของการพัฒนา ส่วนใหญ่ประยุกต์ใช้กับงานศิลปะหรือเกี่ยวข้องกับความคิดสร้างสรรค์ มีความเป็นเอกลักษณ์และแตกต่าง ตัวอย่างแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ ตัวอย่างเช่น Opensea Rarible และ SuperRare เป็นต้น [2] และรูปแบบที่ 3 การบริหารองค์กรแบบกระจาย เป็นแนวทางการบริหารองค์กร โดยสมาชิกในเครือข่ายสามารถตรวจสอบการทำงานของการบริหารองค์กร และมีส่วนร่วมในการบริหารด้วยการถือครองโทเคนดิจิทัล

### 2.3 งานวิจัยที่เกี่ยวข้อง

จากการศึกษาเกี่ยวกับโทเคนดิจิทัลลักษณะเฉพาะ รวมทั้งการหาความสัมพันธ์ระหว่างตลาดสกุลเงินดิจิทัลและตลาดโทเคนดิจิทัลลักษณะเฉพาะ [1] โดยในการสร้างโทเคนดิจิทัลลักษณะเฉพาะใช้มาตรฐาน ERC721



ซึ่งมีข้อจำกัดของมาตรฐานในการรองรับหลายโทเคนดิจิทัลบนสัญญาอัจฉริยะและค่าธรรมเนียมในการทำธุรกรรม [5] และมีการประยุกต์ใช้กับการจัดเก็บข้อมูลด้วย Interplanetary File System (IPFS) เพื่อซื้อขายสินทรัพย์ดิจิทัลบนแพลตฟอร์มตัวอย่างเช่น เกมสะสมแมว CryptoKitties เป็นต้น [3]

จากการศึกษาการประยุกต์ใช้งานบนเครือข่ายอีเธอเรียมในด้านต่าง ๆ ได้แก่ ด้านซื้อขายสินค้า ด้านสุขภาพ ด้านการจัดการลิขสิทธิ์ ด้านประกันภัย และด้านการเงิน [4] ตัวอย่างงานวิจัยได้มีการพัฒนาระบบจัดเก็บข้อมูลของกรรมที่ดิน ซึ่งใช้หลักการควบคุมการเข้าถึงข้อมูลที่มีความเชื่อถือ เพื่อความโปร่งใส และความน่าเชื่อถือ [6] ซึ่งจากงานวิจัยพบว่าการพัฒนาแอปพลิเคชันแบบกระจายมีความปลอดภัยมากกว่าแอปพลิเคชันแบบรวมศูนย์ เนื่องจากประยุกต์ใช้สัญญาอัจฉริยะในการประมวลผลและบันทึกข้อมูลบนบล็อกเชน [7]

### 3. วิธีการดำเนินการวิจัย

วิธีการดำเนินการวิจัยในการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ แบ่งออกเป็น 4 ขั้นตอนหลักดังนี้

#### 3.1 การศึกษาปัญหาและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาปัญหาที่เกี่ยวข้องกับแพลตฟอร์มซื้อขายสินทรัพย์ดิจิทัลในปัจจุบัน พบว่า สินทรัพย์ดิจิทัลทั่วไปไม่สามารถตรวจสอบความเป็นของแท้ได้ จึงทำให้สินทรัพย์ไม่มีมูลค่า และไม่มีการใช้ประโยชน์จากสิทธิการเป็นเจ้าของสินทรัพย์ดิจิทัล ดังนั้นวัตถุประสงค์ของงานวิจัยนี้จึงได้พัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

#### 3.2 การออกแบบกรอบแนวคิดการวิจัย

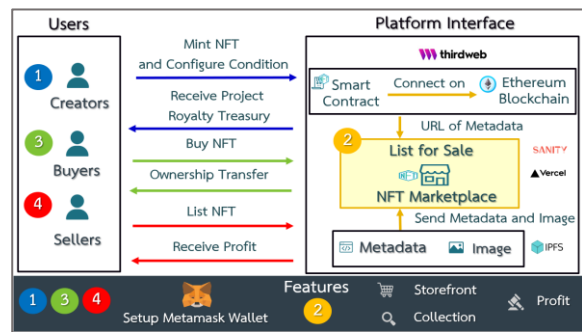
หลังจากการศึกษาปัญหาและงานวิจัยที่เกี่ยวข้อง คณะผู้วิจัยจึงได้ทำการออกแบบกรอบแนวคิดการวิจัยของการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะสามารถแสดงได้ดังภาพที่ 1 แบ่งออกเป็น 2 ส่วนหลัก ได้แก่ ส่วนของผู้ใช้งาน (Users) และส่วนของแพลตฟอร์ม (Platform Interface) มีรายละเอียดดังนี้

#### 1) ส่วนของผู้ใช้งาน

ในส่วนผู้ใช้งานสามารถแบ่งออกได้เป็น 3 ส่วน ได้แก่ หมายเลข 1 ผู้สร้าง (Creators) หมายเลข 3 ผู้ซื้อ (Buyers) และหมายเลข 4 ผู้ขาย (Sellers) โดยผู้ใช้งานจะต้องเชื่อมต่อกระเป๋าเงินดิจิทัลก่อนการใช้งานแพลตฟอร์ม ซึ่งในงานวิจัยนี้ใช้กระเป๋าเงินดิจิทัลของ Metamask และใช้สกุลเงินอีเทอร์ในกระเป๋าเงินดิจิทัลสำหรับซื้อขาย และการชำระค่าธรรมเนียม

#### 2) ส่วนของแพลตฟอร์ม

ในส่วนของแพลตฟอร์มแบ่งออกเป็น 2 ส่วนหลัก ได้แก่ ส่วนติดต่อผู้ใช้งาน (Frontend) และส่วนการทำงานเบื้องหลัง (Backend) โดยมีฟังก์ชันการทำงานของแพลตฟอร์ม ได้แก่ การสร้าง การแสดงรายการขาย การซื้อ และการโอนกรรมสิทธิ์ความเป็นเจ้าของระหว่างผู้ใช้งาน รวมทั้งการได้รับค่าลิขสิทธิ์ความเป็นเจ้าของ และผลตอบแทนจากการซื้อขายสินทรัพย์ดิจิทัล



ภาพที่ 1: กรอบแนวคิดการวิจัยการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

#### 3.3 การออกแบบและพัฒนาแพลตฟอร์ม

ในการออกแบบและพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์ดิจิทัลพัฒนาขึ้นในรูปแบบแอปพลิเคชันแบบกระจาย (Decentralized Applications : DApps) ซึ่งในงานวิจัยประยุกต์ใช้เครื่องมือ Thirdweb ในส่วนการทำงาน และตรวจสอบการทำธุรกรรมกับสินทรัพย์ดิจิทัลบนแพลตฟอร์มด้วยการสร้างสัญญาอัจฉริยะ (Smart Contract) ที่พัฒนาด้วยภาษาซอลิดิตี (Solidity) บนเครือข่ายอีเธอเรียม อีกทั้งทำการจัดเก็บรายละเอียดของสินทรัพย์และแพลตฟอร์มโดยใช้เครื่องมือ Sanity

ในส่วนการจัดเก็บและรักษาข้อมูลการทำธุรกรรมของผู้ใช้งานมีการเชื่อมต่อกับบล็อกเชนสาธารณะ (Public

Blockchain) ซึ่งสามารถตรวจสอบ ยืนยันความถูกต้อง เพื่อมีความโปร่งใสของข้อมูล โดยทำการจัดเก็บคำอธิบาย ข้อมูล และสินทรัพย์ดิจิทัลด้วย Interplanetary File System (IPFS) ในส่วนการพัฒนาเพื่อใช้งานจริงประยุกต์ใช้ เครื่องมือ Vercel โดยแบ่งการทำงานของแพลตฟอร์ม ออกเป็น 2 ส่วน ได้แก่ ส่วนของผู้ใช้งาน ได้แก่ ผู้สร้าง ผู้ซื้อ และผู้ขาย ในส่วนแพลตฟอร์มที่เชื่อมต่อกับเครือข่าย อีเธอเรียม ประกอบด้วยส่วนติดต่อกับผู้ใช้งาน และส่วน การทำงานเบื้องหลัง

จากการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้ โทเคนดิจิทัลลักษณะเฉพาะ ประกอบด้วย 6 ฟังก์ชัน การทำงานหลัก ได้แก่ การสร้าง (Mint Non-Fungible Token) การแสดงรายการขาย (List Non-Fungible Token) การซื้อ (Buy Non-Fungible Token) การโอนกรรมสิทธิ์ความเป็น เจ้าของ (Transfer Owner) การได้รับค่าสิทธิความเป็นเจ้าของ (Receive Project Royalty Treasury) และการได้รับผลตอบแทน (Receive Profit) ซึ่งตัวอย่างฟังก์ชันการตรวจสอบ และแสดง รายการขายสินทรัพย์ดิจิทัลสามารถแสดงได้ดังภาพที่ 2

```

List Non-Fungible Token Function
const NFTCard = ({nftItem, title, listings}) => {
  const [isListed, setIsListed] = useState(false)
  const [price, setPrice] = useState(0)
  const [listingId, setListingId] = useState(0)
  useEffect(() => {
    const listing = listings.find((listing) => listing.asset.name === nftItem.name)
    if (Boolean(listing)) {
      setIsListed(true)
      setPrice(listing.buyoutCurrencyValuePerToken.displayValue)
      setListingId(listing.id)
    }, [listings, nftItem])
  }
}
    
```

**ภาพที่ 2:** ฟังก์ชันการตรวจสอบและแสดงรายการขายสินทรัพย์ดิจิทัล

ผู้ซื้อสามารถดูรายละเอียดข้อมูลรายการขาย และดำเนินการซื้อสินทรัพย์ดิจิทัล (Buy Non-Fungible Token Function) ซึ่งสามารถแสดงฟังก์ชันการทำงานได้ดังภาพที่ 3

```

Buy Non-Fungible Token Function
const confirmPurchase = (toastHandler = toast) => {
  toastHandler.success('Purchase successful!', {
    style: {background: '#04111d',color: '#fff', }, })
  const buyItem = async (
    listingId = selectedMarketNft.id,
    quantityDesired = 1,
    module = marketPlaceModule) => {
    console.log(listingId)
    await module
      .buyoutDirectListing({
        listingId: listingId,
        quantityDesired: quantityDesired,})
    confirmPurchase()
  }
}
    
```

**ภาพที่ 3:** ฟังก์ชันการซื้อสินทรัพย์ดิจิทัล

### 3.4 การทดสอบการใช้งานของแพลตฟอร์ม

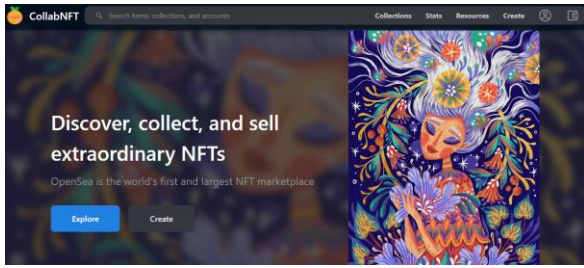
ในการวิจัยจะทำการทดสอบการใช้งานแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ บนเครือข่ายอีเธอเรียมที่พัฒนาขึ้นจากผู้ใช้งานที่เกี่ยวข้อง 3 กลุ่ม ซึ่งมีรายละเอียดดังนี้

- 1) ผู้สร้าง (Creators) สามารถสร้างและกำหนดเงื่อนไขการซื้อสินทรัพย์ดิจิทัล รวมถึงแสดงรายการขาย และได้รับค่าสิทธิความเป็นเจ้าของ
- 2) ผู้ขาย (Sellers) สามารถกำหนดสินทรัพย์ดิจิทัลเพื่อแสดงรายการขาย และได้รับผลตอบแทนจากการซื้อขายสินทรัพย์
- 3) ผู้ซื้อ (Buyers) สามารถรับชมรายการขายและซื้อ รวมทั้งได้รับการโอนกรรมสิทธิ์ความเป็นเจ้าของสินทรัพย์ดิจิทัลระหว่างผู้ใช้งานผ่านแพลตฟอร์มได้

## 4. ผลการดำเนินงานวิจัย

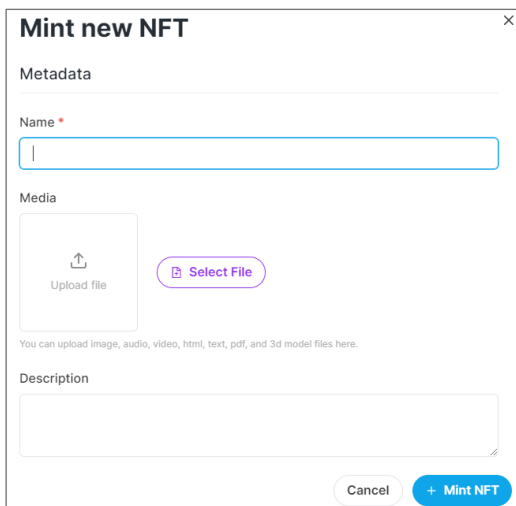
### 4.1 ผลการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

ผลการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ สามารถแสดงตัวอย่างหน้าแรกของแพลตฟอร์มได้ดังภาพที่ 4



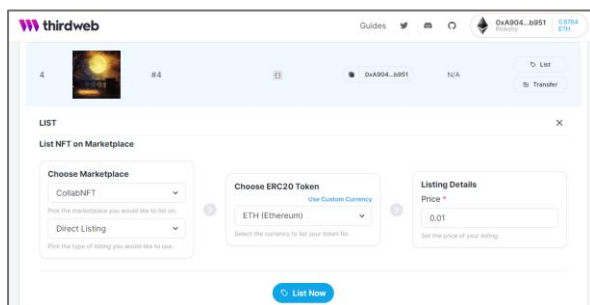
ภาพที่ 4: หน้าแรกของแพลตฟอร์มซื้อขายสินทรัพย์  
โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

ตัวอย่างหน้าจอในส่วนของการสร้างโดยใช้โทเคนดิจิทัลลักษณะเฉพาะ ซึ่งผู้สร้างสามารถกำหนดรายละเอียดได้แก่ ชื่อ ไฟล์สินทรัพย์ดิจิทัล และคำอธิบายรายละเอียด ซึ่งสามารถแสดงได้ดังภาพที่ 5



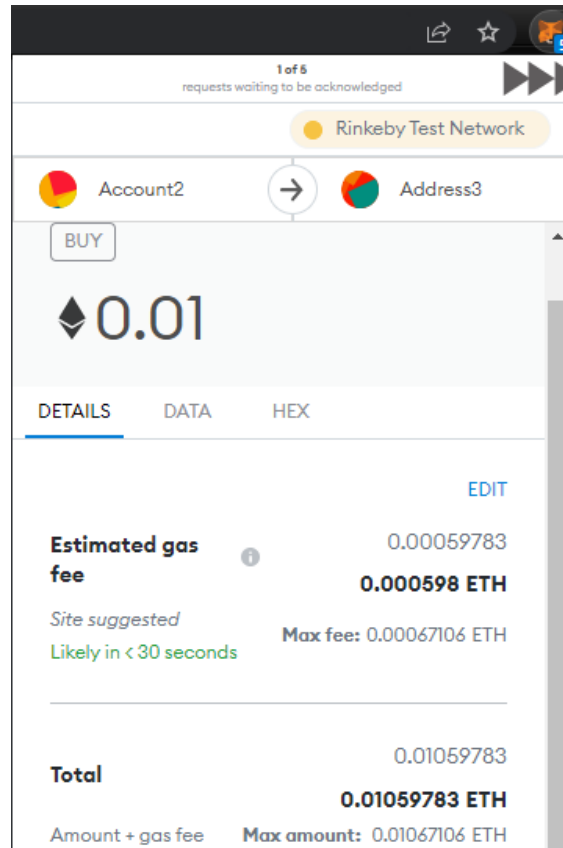
ภาพที่ 5: หน้าจอการสร้างโดยใช้โทเคนดิจิทัล  
ลักษณะเฉพาะ

ตัวอย่างหน้าจอในส่วนของผู้สร้างหรือผู้ขายที่ทำการกำหนดเงื่อนไขซื้อขายสินทรัพย์เพื่อแสดงรายการขาย ซึ่งสามารถแสดงได้ดังภาพที่ 6



ภาพที่ 6: หน้าจอการกำหนดเงื่อนไขของการซื้อขาย  
สินทรัพย์ดิจิทัล

ตัวอย่างหน้าจอในส่วนของการชำระค่าธรรมเนียมด้วยกระเป๋าเงินดิจิทัล Metamask ซึ่งสามารถแสดงได้ดังภาพที่ 7



ภาพที่ 7: หน้าจอสำหรับการชำระค่าธรรมเนียมด้วย Metamask

#### 4.2 ผลการทดสอบการใช้งานของแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

การทดสอบการใช้งานของแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะที่พัฒนาขึ้น สามารถแสดงได้ผลการทดสอบการใช้งานได้ดังตารางที่ 1 ซึ่งทำการทดสอบจากทุกบทบาท (Roles) ที่เกี่ยวข้องกับแพลตฟอร์ม ได้แก่ ผู้สร้าง ผู้ขาย และผู้ซื้อ โดยทำการทดสอบฟังก์ชันการทำงานของแพลตฟอร์ม 6 ฟังก์ชันหลัก ได้แก่ การสร้าง (Mint NFT) การแสดงรายการขาย (List NFT) การซื้อ (Buy NFT) การโอนกรรมสิทธิ์ความเป็นเจ้าของ (Transfer Owner) การได้รับค่าสิทธิความเป็นเจ้าของ (Receive Project Royalty Treasury) และการได้รับผลตอบแทน (Receive Profit) จากผลการทดสอบพบว่าทุกฟังก์ชันการทำงานสามารถซื้อขายแลกเปลี่ยนสินทรัพย์ แสดงผลรายการขายและการได้รับผลประโยชน์ที่ควรจะได้รับได้อย่างถูกต้อง

**ตารางที่ 1:** ผลการทดสอบการใช้งานของแพลตฟอร์ม  
ซื้อขายสินทรัพย์ดิจิทัล

ลำดับ	ฟังก์ชันการทำงาน (Features)	บทบาท (Role)	ผลการทดสอบ			
			ผลลัพธ์ที่คาดหวัง (Excepted Result)	ผลลัพธ์จากการทดสอบ (Actual Result)	ผ่าน (Pass)	ไม่ผ่าน (Fail)
1	การสร้าง	ผู้สร้าง	สร้างสินทรัพย์ด้วย NFT	สร้างสินทรัพย์ด้วย NFT	✓	
2	การแสดงผล	ผู้สร้างและผู้ขาย	กำหนดสินทรัพย์เพื่อแสดงรายการขาย	กำหนดสินทรัพย์เพื่อแสดงรายการขาย	✓	
3	การซื้อ	ผู้ซื้อ	รับชมรายการขายและซื้อสินทรัพย์ดิจิทัล	รับชมรายการขายและซื้อสินทรัพย์ดิจิทัล	✓	
4	การโอนกรรมสิทธิ์ความเป็นเจ้าของ	ผู้ซื้อ	ได้รับการโอนกรรมสิทธิ์ความเป็นเจ้าของสินทรัพย์ดิจิทัล	ได้รับการโอนกรรมสิทธิ์ความเป็นเจ้าของสินทรัพย์ดิจิทัล	✓	
5	การได้รับค่าสิทธิความเป็นเจ้าของ	ผู้สร้าง	ได้รับค่าสิทธิความเป็นเจ้าของจากการสร้างสรรค์สินทรัพย์ดิจิทัล	ได้รับค่าสิทธิความเป็นเจ้าของจากการสร้างสรรค์สินทรัพย์ดิจิทัล	✓	
6	การได้รับผลตอบแทน	ผู้ขาย	ได้รับผลตอบแทนจากการขายสินทรัพย์ดิจิทัล	ได้รับผลตอบแทนจากการขายสินทรัพย์ดิจิทัล	✓	

## 5. สรุปผลการวิจัย

จากการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะบนเครือข่ายอีเธอเรียมในรูปแบบแอปพลิเคชันแบบกระจาย ซึ่งในงานวิจัยใช้กระเป๋าเงินดิจิทัลของ Metamask และใช้สกุลเงินอีเทอร์ในกระเป๋าเงินดิจิทัลสำหรับซื้อขาย และการชำระค่าธรรมเนียม โดยแพลตฟอร์มที่พัฒนาขึ้นสามารถรองรับฟังก์ชันการทำงานต่าง ๆ ได้แก่ การสร้าง การแสดงผล รายการขาย การซื้อ การโอนกรรมสิทธิ์ความเป็นเจ้าของ การได้รับค่าสิทธิความเป็นเจ้าของและการได้รับผลตอบแทน จากผลการทดสอบการใช้งานแพลตฟอร์มพบว่าทุกฟังก์ชันการทำงานสามารถนำไปใช้งานได้จริง ช่วยเพิ่มช่องทางให้สินทรัพย์ดิจิทัลที่มีมูลค่าสามารถซื้อขาย

กันได้ง่ายมากขึ้น อีกทั้งเป็นแนวทางในการประยุกต์ใช้กับการลงทุนของสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะหรืองานวิจัยอื่น ๆ ที่เกี่ยวข้อง และสามารถนำไปต่อยอดพัฒนาแอปพลิเคชันแบบกระจายบนเครือข่ายอีเธอเรียมที่เน้นความมั่นคงปลอดภัยมากขึ้นในการใช้งาน

## เอกสารอ้างอิง

- [1] Ante, L. "The Non-Fungible Token (NFT) Market and Its relationship with Bitcoin and Ethereum," *SSRN Electronic Journal*, Vol. 20, 2021.
- [2] Thilagaraj, A., and Davis, J. "Non-Fungible Token (NFT) – The Game Changer in the Digital Art World," *In Proceeding of 2021 Ciencia Y Sociedad*, Vol. 51, pp. 190-194, 2021..
- [3] Karapapas, C., Pittaras, I. and Polyzos, G.C. "Fully Decentralized Trading Games with Evolvable Characters using NFTs and IPFS," *In Proceeding of 2021 IFIP Networking Conference (IFIP Networking)*, pp. 1-2, 2021.
- [4] Mohanta, B.K., Panda, S.S. and Jena, D. "An Overview of Smart Contract and Use Cases in Blockchain Technology," *In Proceeding of 2018 9th international conference on computing, communication and networking technologies (ICCCNT)*, pp. 1-4, 2018.
- [5] Chirtoaca, D., Ellul, J. and Azzopardi, G. "A Framework for Creating Deployable Smart Contracts for Non-Fungible Tokens on the Ethereum Blockchain," *In Proceeding of 2020 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS)*, pp. 100-105, 2020.
- [6] Nandi, M., Bhattacharjee, R.K., Jha, A. and Barbhuiya, F.A. "A Secured Land Registration Framework on Blockchain," *In Proceeding of 2020 Third ISEA Conference on Security and Privacy (ISEA-ISAP)*, pp. 130-138, 2020.
- [7] Kina-Kina, K.M., Cutipa-Arias, H.E. and Shiguihara-Juárez, P. "A Comparison of Performance between Fully and Partially Decentralized Applications," *In Proceeding of 2019 IEEE XXVI International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pp. 1-4, 2019.

# ระบบระบุตำแหน่งภายในอาคารโดยการวิเคราะห์ความแรงของ เครือข่ายไร้สายโดยใช้เทคนิคการเรียนรู้ของเครื่อง Indoor Positioning System by Analyzing Wireless of Network Using Machine Learning

ณัฐพล ไชยวงศ์ (Nattaphon Chaiwong)<sup>1</sup> อธิคม ศิริ (Atikhom Siri)<sup>2</sup>  
สำนักวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏเชียงราย  
<sup>1</sup>asd123\_789@hotmail.com, <sup>2</sup>atikhom.s@gmail.com

## บทคัดย่อ

ระบบระบุตำแหน่งภายในอาคารโดยการวิเคราะห์ความแรงของเครือข่ายไร้สาย โดยใช้เทคนิคการเรียนรู้ของเครื่องพัฒนาขึ้นเพื่อใช้งานในสภาพแวดล้อมภายในอาคารที่เป็นข้อจำกัดของระบบระบุตำแหน่งโดยใช้ระบบจีพีเอสที่ใช้สัญญาณจากดาวเทียมที่ไม่สามารถใช้งานได้ภายในอาคาร ปัจจุบันมีบริการตำแหน่งที่มีการระบุตำแหน่งโดยไวไฟเข้ามาช่วย แต่ก็ยังมีความคลาดเคลื่อนอยู่ โดยระบบต้องมีการจัดเก็บข้อมูลตำแหน่งและความแรงของสัญญาณเป็นชุดข้อมูลเพื่อสร้างวิธีการ ซึ่งระบบประกอบด้วยอุปกรณ์ที่ทำหน้าที่ค้นหาอุปกรณ์และค่าความแรงของสัญญาณเพื่อส่งค่าข้อมูลไปยังอุปกรณ์ควบคุมนำค่าความแรงของสัญญาณไปวิเคราะห์ทำนายตำแหน่งของอุปกรณ์โดยจัดการข้อมูลและแสดงผลตำแหน่งผ่านเว็บแอปพลิเคชัน สามารถนำมาประยุกต์ในเรื่องของความปลอดภัยเนื่องจากไม่จำเป็นต้องเชื่อมต่อกับไวไฟแค่เปิดไวไฟก็สามารถระบุตำแหน่งได้ หรือในห้างสรรพสินค้าในการให้ข้อมูล โชนผลิตภัณฑ์ตามตำแหน่งที่ถูกค้าอยู่ ซึ่งการทดลองผู้วิจัยได้ใช้ขนาดพื้นที่ 24x24 เมตร โดยแบ่งพื้นที่ย่อยขนาด 3x3, 5x5 และ 10x10 เมตร ใช้อุปกรณ์ค้นหาค่าความแรงสัญญาณ 3 และ 4 อุปกรณ์ โดยใช้ชุดข้อมูล 397 ข้อมูลเพื่อสร้างโมเดล

ผลการทดลอง อัลกอริทึมที่มีความเหมาะสมที่สุดในการนำมาทำนายคือ Decision Tree ความแม่นยำเฉลี่ยที่ 86.79 เปอร์เซ็นต์ หากเพิ่มอุปกรณ์ค้นหาค่าความแรงสัญญาณจาก 3 เป็น 4 จะทำให้เทคนิคการเรียนรู้ของเครื่องวิธีทุกวิธีมีความแม่นยำเพิ่มขึ้นเฉลี่ย 7.71 เปอร์เซ็นต์

คำสำคัญ: ระบบระบุตำแหน่ง การเรียนรู้ของเครื่อง เว็บแอปพลิเคชัน

## Abstract

Indoor positioning system by analyzing the strength of the wireless network. It uses machine learning techniques developed for use in indoor environments where GPS positioning systems are limited by satellite signals that cannot be used indoors. Wi-Fi-based location services are now available to help. There is still a discrepancy. System must store location and signal strength data as a dataset to create a method. The system consists of a device that looks for devices and the signal strength to send data to the control device, bringing the signal strength to analyze and predict the location of the device. by managing data and displaying location via web application It can be applied in terms of safety as it does not need to be connected to a Wi-Fi, just turn on the Wi-Fi can determine the location. or in a department store to provide product zone information according to the customer's location In the experiment, the researcher used the area size 24x24 meters, divided into 3x3, 5x5 and 10x10 meters sub-area, using 3 and 4 signal strength search devices, using 397 data sets to create a model.

Results The most suitable algorithm for making predictions was the Decision Tree, with an average accuracy of 86.79 percent. If the signal strength finder was increased from 3 to 4, all machine learning

*techniques would be more efficient. Accuracy increased by an average of 7.71 percent.*

**Keywords:** Location system, Web application, Machine learning.

## 1. บทนำ

บทปัจจุบันการติดต่อสื่อสารและเทคโนโลยีต่าง ๆ ได้เข้ามามีบทบาทต่อการดำเนินชีวิตของมนุษย์เป็นอย่างมาก และเป็นปัจจัยที่สำคัญต่อการทำงานขององค์กร หน่วยงานต่างๆ ให้มีการติดต่อสื่อสารแลกเปลี่ยนข้อมูลที่รวดเร็วและประสิทธิภาพในการทำงานมากยิ่งขึ้น เทคโนโลยีที่ได้รับความนิยมมากที่สุดในปัจจุบันคือเทคโนโลยีไวไฟ หรือ เครือข่ายไร้สาย (Wireless LAN Technology) เป็นเทคโนโลยีที่ได้รับความนิยมที่ช่วยให้อุปกรณ์อิเล็กทรอนิกส์สามารถแลกเปลี่ยนข้อมูลโดยการเชื่อมต่ออินเทอร์เน็ตแบบไร้สายโดยใช้คลื่นวิทยุ

ในยุคที่เทคโนโลยีสารสนเทศ มีการพัฒนาอย่างรวดเร็วและการแข่งขันที่สูง จึงทำให้องค์กรและหน่วยงานจำเป็นต้องมีการปรับตัวในหลายๆ ด้านและพัฒนาให้ทันต่อการเปลี่ยนแปลงในโลกของการสื่อสาร จึงได้มีการใช้เทคนิคกลยุทธ์เข้ามาช่วยในการทำให้องค์กรและหน่วยงานเติบโตมากยิ่งขึ้น เช่น โรงงานอุตสาหกรรมได้มีการตรวจสอบพนักงาน โดยการสังเกตพฤติกรรมของพนักงาน หรือ ภายในห้างสรรพสินค้ามีการเก็บข้อมูลของลูกค้า ไม่ว่าจะเป็นจำนวนหรือการเดินหิบบจับของภายในห้างสรรพสินค้า เพื่อนำไปใช้ในการโฆษณาสินค้าที่ลูกค้ามีความสนใจมากที่สุด โดยทั้งหมดที่ยกมาอย่างมานี้ทำโดยการสังเกตโดยให้พนักงานภายในเป็นคนเก็บข้อมูลทั้งหมด ซึ่งเป็นเรื่องที่ต้องใช้พนักงานในการเก็บข้อมูลที่เยอะเกินความจำเป็นและทำให้เกิดความผิดพลาดของข้อมูลค่อนข้างสูง

จากปัญหาที่กล่าวมานี้ทางผู้วิจัยจึงได้หาวิธีการประยุกต์ใช้เครือข่ายไร้สายเพื่อช่วยในเรื่องของทรัพยากรบุคคลที่มีการใช้เกินความจำเป็น ความรวดเร็วในด้านการเก็บข้อมูล โดยการนำความแรงของสัญญาณอุปกรณ์อิเล็กทรอนิกส์ที่สามารถรับสัญญาณเครือข่ายไร้สายได้ เช่น สมาร์ทโฟน โน้ตบุ๊ก จากอุปกรณ์กระจายสัญญาณ

(Access Point) ภายในบริเวณทั้งหมดมาทำการทำนายโดยใช้เทคนิคการเรียนรู้ของเครื่องเข้ามาช่วยในการระบุตำแหน่งและแสดงผลผ่านเว็บแอปพลิเคชัน เพื่อนำตำแหน่งของอุปกรณ์มาต่อยอด ไม่ว่าจะเป็นในเรื่องของการสังเกตพฤติกรรมของพนักงาน การเดินหิบบจับของภายในห้างสรรพสินค้าของลูกค้าว่าลูกค้าเดินไปบริเวณไหนมากที่สุด จนไปถึงในเรื่องของความปลอดภัย หรือของหาย ซึ่งระบบสามารถที่จะระบุตำแหน่งของอุปกรณ์อิเล็กทรอนิกส์ที่เข้ามาภายในบริเวณได้โดยไม่ต้องเชื่อมต่อกับอุปกรณ์กระจายสัญญาณ หากเกิดอุบัติเหตุสามารถที่จะนำตำแหน่งของผู้เกิดอุบัติเหตุที่อยู่ภายในบริเวณ ระบุต่อเจ้าหน้าที่ เพื่อเข้าช่วยเหลือได้อย่างทันท่วงที

ดังนั้นงานวิจัยนี้จึงได้ออกแบบขั้นตอนวิธีและพัฒนาระบบระบุตำแหน่งภายในอาคาร โดยการวิเคราะห์ความแรงของเครือข่ายไร้สายโดยใช้เทคนิคการเรียนรู้ของเครื่องเพื่อแก้ปัญหาที่เกิดขึ้นระบบระบุตำแหน่งภายในอาคาร โดยการวิเคราะห์ความแรงของเครือข่ายไร้สายโดยใช้เทคนิคการเรียนรู้ของเครื่องจึงช่วยแก้ไขในข้อบกพร่องได้อย่างถูกต้องแม่นยำมากขึ้น

## 2. ทฤษฎีที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบระบุตำแหน่งภายในอาคาร โดยการวิเคราะห์ความแรงของสัญญาณเครือข่ายไร้สายโดยใช้เทคนิคการเรียนรู้ของเครื่อง โดยมีรายละเอียดดังต่อไปนี้ 1) โครงสร้างพื้นฐานของระบบระบุตำแหน่งภายในอาคาร 2) เอ็มคิวทีที 3) การเรียนรู้ของเครื่อง 4) อุปกรณ์กระจายสัญญาณ 5) ดัชนีที่ใช้วัดการได้อินคลื่นสัญญาณ

**2.1 โครงสร้างพื้นฐานของระบบระบุตำแหน่งภายในอาคาร Location-Based Service (LBS)** เป็นบริการที่ใช้ประโยชน์จากเทคโนโลยีเครือข่ายไร้สาย ที่ทำให้คนสามารถที่ระบุตำแหน่งที่อยู่ผู้ใช้ได้ ผ่านทางโทรศัพท์มือถือ ซึ่งจะให้บริการออกมาในรูปแบบ Software Application โดยที่เทคโนโลยี LBS จะเป็นการ

นำตำแหน่งของคนที่อยู่ภายในอาคารหรือคนที่อยู่ภายในบริเวณมาต่อยอดให้เกิดประโยชน์ [5] [6]

**2.2 การเรียนรู้ของเครื่อง (Machine Learning)** การเรียนรู้ของเครื่อง คือระบบที่สามารถเรียนรู้ได้จากตัวอย่างด้วยตนเองโดยปราศจากการป้อนคำสั่งของโปรแกรมเมอร์ ความก้าวหน้าในครั้งนี้นำมาพร้อมกับความคิดที่ว่าเครื่องคอมพิวเตอร์สามารถเรียนรู้เพียงแค่ว่าจากข้อมูลอย่างเฉียวเพื่อที่จะผลิตผลลัพธ์ที่แม่นยำออกมาได้ ซึ่งใช้อัลกอริทึมทั้งหมด 5 อัลกอริทึม

**K-Nearest Neighbors** [1] เป็นวิธีการแบ่งคลาสสำหรับใช้จัดหมวดหมู่ข้อมูล (Classification) ใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงมากน้อยเพียงใด หากข้อมูลที่กำลังสนใจนั้นอยู่ใกล้ข้อมูลใดมากที่สุด ระบบจะให้คำตอบเป็นเหมือนคำตอบของข้อมูลที่อยู่ใกล้ที่สุดนั้น โดยใช้สูตรที่ใช้ในการคำนวณค่าความคล้ายคลึง

$$d_1(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2},$$

**Nearest Centroid** คือการจำแนกข้อมูลที่เป็นกลุ่มเพื่อค่าจุดที่ใกล้ที่สุด (NC) เป็นหนึ่งในการจำแนกที่ง่ายที่สุดแต่ทรงพลังที่สุดหลักการของ Nearest Centroid จะทำการสรุปชุดข้อมูลด้วยค่าเฉลี่ยแล้วคำนวณหาจุดกึ่งกลางของชุดข้อมูลที่แตกต่างกัน เมื่อได้จุดกึ่งกลางจากการหาค่าเฉลี่ยแล้ว จะทำการทำนายผลจาก จุดที่ใกล้ที่สุด โดยใช้สูตรที่ใช้ในการคำนวณหาจุดกึ่งกลาง ดังต่อไปนี้

$$\bar{x} = \frac{\sum x}{n_i}$$

**Decision Tree** [2] ต้นไม้การตัดสินใจจะทำการจัดกลุ่ม (classify) ชุดข้อมูลนำเข้าในแต่ละกรณี (Instance) แต่ละบัพ (node) ของต้นไม้การตัดสินใจคือตัวแปร (attribute) ต่างๆ ของชุดข้อมูล เช่นหากต้องการตัดสินใจว่าจะไปเล่นกีฬาหรือไม่ก็จะมีตัวแปรต้นที่จะต้องพิจารณา คือ ทัศนียภาพ ลม ความชื้น อุณหภูมิ เป็นต้น และมีตัวแปรตามซึ่งเป็นผลลัพธ์จากต้นไม้คือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่า ซึ่งแต่ละตัวแปรนั้นก็จะมีค่าของตัวเอง (value) เกิดเป็นชุดของตัวแปร-ค่าของตัวแปร (attribute-value pair) เช่น ทัศนียภาพเป็นตัวแปร ก็อาจมีค่าได้เป็น ฝนตก

แดดออก หรือการตัดสินใจว่าจะไปเล่นกีฬารึเปล่านั้นก็อาจมีค่าได้เป็นใช่ กับ ไม่ใช่ เป็นต้น

คำนวณหาค่า Entropy ของแต่ละกิ่งเป้าหมายโดยใช้สูตร

$$EN(b_i) = \sum (-P(b_i) * \log_2 P(b_i))$$

คำนวณหาค่า Entropy ของโหนดเป้าหมาย โดยใช้สูตร

$$EN(node) = \sum P(node) * EN(b_i)$$

คำนวณหาค่า Entropy ของโหนดเป้าหมาย โดยใช้สูตร

$$IG(node) = EN(PrevNode) - EN(node)$$

**Support Vector Machine** [3] หลักการทำงานโดยทั่วไปของ SVM คือ จะกำหนดเส้นแบ่งกลุ่มข้อมูลตามผลลัพธ์ (Class หรือ Label) ด้วยวิธีใดวิธีหนึ่ง หรือเรียกว่า เคอร์เนล (Kernel) เช่น Linear (เส้นตรง) Radial (ตามแนวรัศมี) หรือ Polynomial (ตามแนวโค้ง) เป็นต้น ทั้งนี้ขึ้นอยู่กับลักษณะการกระจายของข้อมูล เช่น ภาพต่างๆ ดังต่อไปนี้ โดย หากค่าที่เราทำนายผลอยู่ฝั่งใดของแนวเส้นแบ่ง ก็จะถือว่าเป็นคลาสเดียวกับข้อมูลในฝั่งนั้น การแบ่งเส้นตรงแบบ hyperplane สามารถใช้สูตรในสมการเส้นตรงที่แสดงความสัมพันธ์ระหว่าง  $x_1$  และ  $x_2$  เป็น

$$b + w_1x_1 + w_2x_2 = 0$$

เส้นตรงที่ใช้แทน Boundary ซึ่งอยู่เหนือ Haperplane สามารถเขียนในแบบสูตรเส้นตรงเป็น

$$b + w_1x_1 + w_2x_2 = 1$$

เส้นตรงที่ใช้แทน Boundary ซึ่งอยู่เหนือ Haperplane สามารถเขียนในแบบสูตรเส้นตรงเป็น

$$b + w_1x_1 + w_2x_2 = -1$$

หากแบ่งคลาส ลักษณะการทำนายผล คือ

$b + w_1x_1 + w_2x_2 \geq 1$  แสดงว่าข้อมูลตำแหน่ง จัดอยู่ในกลุ่มคลาส B

$b + w_1x_1 + w_2x_2 \leq -1$  แสดงว่าข้อมูลตำแหน่ง จัดอยู่ในกลุ่มคลาส A

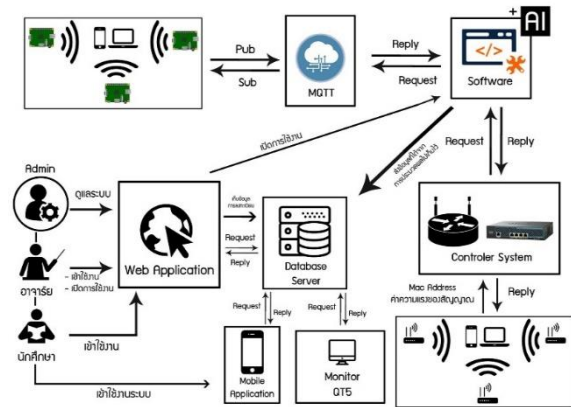
**Naive Bayes** [4] โดย Naive Bayes หลักการทำงานคำนวณหาความน่าจะเป็น โดยใช้ใช้อัลกอริทึม โดยอัลกอริทึมจะเปรียบเทียบข้อมูลที่ต้องการทำนาย โดยการนำข้อมูลที่ต้องการทำนายมาเปรียบเทียบในชุดข้อมูลในแอตทริบิวต์นั้นๆ และได้จะได้ค่ามาเปรียบเทียบในแต่ละผลลัพธ์เพื่อได้ผลการทำนาย โดยใช้สูตรของ Bayes

$$P(A | B) = \frac{P(B|A) * P(A)}{P(B)}$$

### 3. การดำเนินการวิจัย

#### 3.1 กรอบแนวคิดระบบ

กรอบแนวความคิดระบบระบุตำแหน่งภายในอาคาร โดยการวิเคราะห์ความแรงของสัญญาณเครือข่ายไร้สาย โดยใช้เทคนิคการเรียนรู้ของเครื่องจากการที่ผู้วิจัยได้วางแผนการดำเนินงานผู้วิจัยได้ออกแบบภาพรวมของระบบระบุตำแหน่ง



ภาพที่ 1: แผนผังแสดงถึงภาพรวมการออกแบบ

ระบบระบุตำแหน่งภายในอาคาร โดยการวิเคราะห์ความแรงของสัญญาณเครือข่ายไร้สายโดยใช้เทคนิคการเรียนรู้ของเครื่อง ในส่วนของสถาปัตยกรรมของระบบตรวจสอบ มีองค์ประกอบที่สำคัญ 8 ส่วนคือ 1) อุปกรณ์รับสัญญาณไวไฟของผู้ใช้งานซึ่งติดตั้งอุปกรณ์ไวไฟมาพร้อมกับตัวเครื่อง ทำหน้าที่เป็นอุปกรณ์รับสัญญาณเพื่อติดต่อกับระบบค้นหาสัญญาณของระบบ 2) เครื่องคอมพิวเตอร์ที่ติดตั้ง อุปกรณ์ไวไฟเพื่อเข้าใช้งานเว็บแอปพลิเคชัน (Web Application) เพื่อดูข้อมูลตำแหน่งภายในอาคารของผู้ใช้งานโดยรอบที่เปิดใช้งานไวไฟอยู่ 3) อุปกรณ์กระจายสัญญาณไร้สาย (Access Point) ซึ่งทำหน้าที่ในการกระจายสัญญาณไวไฟไปยังอุปกรณ์รับสัญญาณไวไฟของผู้ใช้งานโดยรอบ 4) อุปกรณ์ควบคุม (Controller) ในส่วนนี้ อุปกรณ์ควบคุมจะทำการควบคุมอุปกรณ์กระจายสัญญาณไร้สาย เพื่อเก็บรวบรวมข้อมูลจากอุปกรณ์กระจายสัญญาณไร้สาย และส่งข้อมูลไปยังซอฟต์แวร์ (Software) เรื่อยๆ หากซอฟต์แวร์ เปิดการใช้

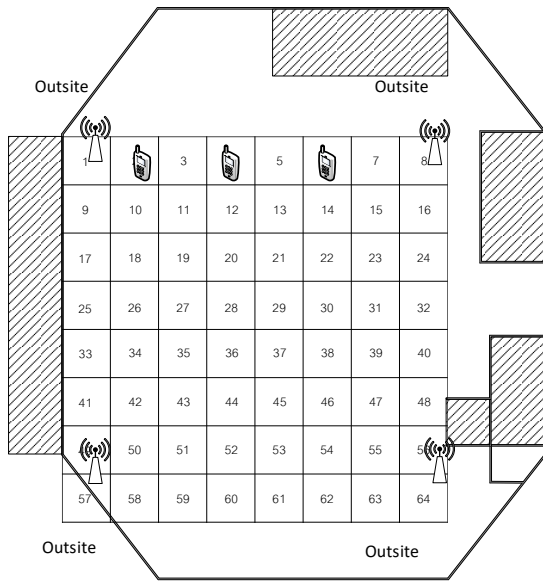
งานการรับข้อมูลซอฟต์แวร์ ก็จะได้รับข้อมูล 5) ซอฟต์แวร์ ในส่วนของซอฟต์แวร์จะทำการรองรับข้อมูลจากอุปกรณ์ควบคุม Mac Address และ ค่าความแรงของสัญญาณ มาทำการประมวลผล และส่งข้อมูล Output ไปเก็บไว้ในระบบฐานข้อมูล (Database Server) 6) ระบบฐานข้อมูล ทำหน้าที่จัดเก็บข้อมูลต่างๆ และเก็บข้อมูลตำแหน่งที่ผ่านการประมวลผลที่ถูกส่งมาจากซอฟต์แวร์ มีการร้องขอ 7) เว็บแอปพลิเคชัน (Web Application) ทำหน้าที่ในการลงทะเบียนข้อมูลของผู้ใช้งานผู้ดูแล ลงทะเบียนข้อมูลของตำแหน่งต่างๆ แสดงข้อมูลตำแหน่งของผู้ใช้งานเวลาเข้าออกโดยเทคนิคการเรียนรู้ของเครื่อง และเปิดใช้งานระบบ 8) Model ในการ ทำนาย (Predict) ซึ่งในส่วนนี้จะใช้อัลกอริทึมต่างๆ เข้ามาช่วยในการสร้างโมเดล

#### 3.2 ขั้นตอนการดำเนินการทดลอง

ผู้วิจัยได้ทำการออกแบบและพัฒนาระบบเรียบร้อยแล้ว ในส่วนของทดลองผู้วิจัยได้ ทำการทดลองเพื่อหาว่าการทำนาย โดยใช้เทคนิคการเรียนรู้ของเครื่องจะเปรียบเทียบกับอัลกอริทึม (Algorithm) ใดมีความแม่นยำมากกว่ากัน โดยแบ่งออกเป็น 2 ส่วนคือ การใช้อุปกรณ์กระจายสัญญาณ (Access Point) 3 อุปกรณ์และ 4 อุปกรณ์ในพื้นที่เปิดกว้างไม่มีสิ่งกีดขวาง ในการทดลองผู้วิจัยได้ทำการเก็บชุดข้อมูล (Dataset) เริ่มจากการสร้างข้อมูลตำแหน่งต่างๆ บนเว็บแอปพลิเคชันและนำโทรศัพท์มือถือที่เปิดใช้งานไวไฟไปอยู่ในตำแหน่งที่ต้องการเก็บชุดข้อมูล จากนั้นเปิดใช้งาน ซอฟต์แวร์บนเว็บแอปพลิเคชัน ซอฟต์แวร์จะทำการรับข้อมูลความแรงของสัญญาณไวไฟบนมือถือจากราสเบอร์รี่พายหรืออุปกรณ์ควบคุม (Controller) และนำข้อมูลมาแยกความแรงของสัญญาณ โดยเลือกข้อมูลจากแมคแอดเดรสที่ต้องการเก็บเท่านั้นไปเก็บไว้ในฐานข้อมูล (Database) เพื่อเตรียมที่จะใช้ในการทำนายลำดับต่อไป ลำดับต่อไปนำสมาร์ทโฟนไปอยู่ในตำแหน่ง ที่ต้องการหาความถูกต้อง จากนั้นให้โปรแกรมทำการทำนายผล 10 ผลการทำนายต่อ 1 ตำแหน่งโดยใช้อัลกอริทึมที่แตกต่างกันในการทำนายในแต่ละครั้งและหาว่าโปรแกรมได้ทำนายตำแหน่งได้ถูกต้องหรือไม่ และ



เก็บข้อมูลการทำนายผลและสรุปผลการทดลอง โดยการ ออกแบบพื้นที่ในการทดลองดังภาพที่ 2

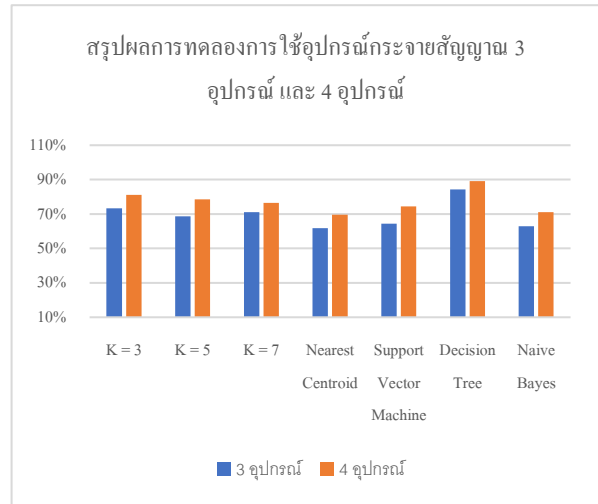


ภาพที่ 2: การออกแบบการทดลองความละเอียดที่ 3 เมตรคูณ 3 เมตร โดยใช้ 4 อุปกรณ์กระจายสัญญาณ

### 3.3 สรุปผลการทดลอง

จากผลการทดลองโปรแกรมระบบระบุตำแหน่งภายในอาคาร โดยการวิเคราะห์ความแรงของสัญญาณเครือข่ายไร้สาย โดยใช้เทคนิคการเรียนรู้ของเครื่อง แบ่งการทดลองออกเป็น 2 ส่วนคือส่วนที่ 1 การทดลองหาความแม่นยำของการทำนาย โดยใช้เทคนิคการเรียนรู้ของเครื่องเปรียบเทียบอัลกอริทึม Algorithm ใดมีความแม่นยำมากที่สุด โดยใช้อุปกรณ์กระจายสัญญาณ 3 อุปกรณ์ ส่วนที่ 2 การทดลองหาความแม่นยำของการทำนาย โดยใช้เทคนิคการเรียนรู้ของเครื่องเปรียบเทียบอัลกอริทึม Algorithm ใดมีความแม่นยำมากที่สุด โดยใช้อุปกรณ์กระจายสัญญาณ 4 อุปกรณ์ โดยแบ่งการทดลองในแต่ละส่วนออกเป็น 3 ส่วนย่อย การทดลองทำนายความแม่นยำในเขตพื้นที่ขนาดกว้าง 24 เมตร และยาว 24 เมตร โดยกำหนดความละเอียดที่ 3 เมตรคูณ 3 เมตร การทดลองทำนายความแม่นยำในเขตพื้นที่ขนาดกว้าง 24 เมตร และยาว 24 เมตร โดยกำหนดความละเอียดที่ 5 เมตรคูณ 5 เมตร การทดลองทำนายความแม่นยำในเขตพื้นที่ขนาดกว้าง 24 เมตร และยาว 24 เมตร โดยกำหนดความละเอียด

ที่ 10 เมตรคูณ 10 เมตร ซึ่งจะแบ่งการสรุปออกเป็น 2 ส่วนดังนี้



ภาพที่ 3: สรุปผลการทดลองการทำนายโดยใช้อุปกรณ์กระจายสัญญาณ 3 อุปกรณ์ และ 4 อุปกรณ์

จากภาพที่ 3 ผลการทดลองทำนายความแม่นยำในเขตพื้นที่ขนาดกว้าง 24 เมตร และยาว 24 เมตร โดยใช้อุปกรณ์กระจายสัญญาณภายในพื้นที่จำนวน 3 อุปกรณ์ และ 4 อุปกรณ์ ซึ่งใช้ Algorithm เช่น K-Nearest Neighbors, Decision Tree, Support Vector Machine, Nearest Centroid, Naive Bayes เป็นต้น โดยแบ่งการเปรียบเทียบออกเป็น 5 ส่วนได้แก่ ส่วนที่ 1 อัลกอริทึม K-Nearest Neighbors เมื่อ k = 3 อุปกรณ์กระจายสัญญาณภายในพื้นที่จำนวน 4 อุปกรณ์ มีค่าความแม่นยำมากกว่า 3 อุปกรณ์ 7.90 เปอร์เซ็นต์ เมื่อ k = 5 อุปกรณ์กระจายสัญญาณภายในพื้นที่จำนวน 4 อุปกรณ์ มีค่าความแม่นยำมากกว่า 3 อุปกรณ์ 9.84 เปอร์เซ็นต์ เมื่อ k = 7 อุปกรณ์กระจายสัญญาณภายในพื้นที่จำนวน 4 อุปกรณ์ มีค่าความแม่นยำมากกว่า 3 อุปกรณ์ 5.33 เปอร์เซ็นต์ ส่วนที่ 2 อัลกอริทึม Nearest Centroid อุปกรณ์กระจายสัญญาณภายในพื้นที่จำนวน 4 อุปกรณ์ มีค่าความแม่นยำมากกว่า 3 อุปกรณ์ 7.78 เปอร์เซ็นต์ ในส่วนที่ 3 อัลกอริทึม Support Vector Machine อุปกรณ์กระจายสัญญาณภายในพื้นที่จำนวน 4 อุปกรณ์ มีค่าความแม่นยำมากกว่า 3 อุปกรณ์ 10.22 เปอร์เซ็นต์ ส่วนที่ 4 อัลกอริทึม Decision Tree อุปกรณ์กระจายสัญญาณภายในพื้นที่จำนวน 4 อุปกรณ์ มีค่าความแม่นยำมากกว่า 3 อุปกรณ์ 4.80 เปอร์เซ็นต์ และ ส่วนที่ 5 อัลกอริทึม Naive Bayes อุปกรณ์กระจาย

สัญญาณภายในพื้นที่จำนวน 4 อุปกรณ์ มีค่าความแม่นยำมากกว่า 3 อุปกรณ์ 8.12 เปอร์เซ็นต์

#### 4. สรุปผลการวิจัย

จากการทดลองระบบระบุตำแหน่งภายในอาคารโดยการวิเคราะห์ความแรงของสัญญาณเครือข่ายไร้สาย ทำให้ได้ผลการทดลองครั้งนี้ อัลกอริทึมที่มีความแม่นยำมากที่สุดคือ Decision Tree ความแม่นยำเฉลี่ยที่ 86.79 เปอร์เซ็นต์ หากใช้อุปกรณ์กระจายสัญญาณ 3 อุปกรณ์มีค่าความแม่นยำเฉลี่ย 69.53 เปอร์เซ็นต์ จากอัลกอริทึมที่ใช้ในการทำนายทั้งหมด ในการทดลองโดยใช้อุปกรณ์กระจายสัญญาณ 4 อุปกรณ์มีค่าความแม่นยำเฉลี่ย 77.24 เปอร์เซ็นต์ จากอัลกอริทึมที่ใช้ในการทำนายทั้งหมด ซึ่งอุปกรณ์กระจายสัญญาณ 4 อุปกรณ์มีค่าความแม่นยำมากกว่า 3 อุปกรณ์ ถึง 7.71 เปอร์เซ็นต์

##### 4.1 ข้อเสนอแนะ

จากการทดลองโปรแกรมระบุตำแหน่งภายในอาคารโดยการวิเคราะห์ความแรงของสัญญาณเครือข่ายไร้สาย โดยใช้เทคนิคการเรียนรู้ของเครื่องสามารถชี้แนะทางในการพัฒนาให้สามารถนำไปใช้ประโยชน์ได้อย่างมีประสิทธิภาพมากขึ้น โดยมีข้อเสนอแนะในการทำวิจัยครั้งต่อไปดังนี้

1. พัฒนาเว็บแอปพลิเคชันให้ผู้ใช้สามารถดูรายละเอียดของตำแหน่งภายในอาคารได้ง่ายมากยิ่งขึ้น
2. พัฒนา ปรับปรุง เทคนิค วิธีการและแนวคิดใหม่สำหรับเพิ่มประสิทธิภาพ
3. ปรับปรุงความเสถียรของสัญญาณมากยิ่งขึ้น

#### เอกสารอ้างอิง

- [1] D. Ouyang, D. Li, Q. Li. "Cross-validation and non-parametric k nearestneighbour estimation." *Econom. J.* Vol. 9, No. 3, pp. 448–471, 2006.
- [2] B. Ding, Y. Zheng, S. Zang. "A New Decision Tree Algorithm Based on Rough Set Theory." *Asia-Pacific Conference on Information Processing*, IEEE, pp. 326-329, 2009.

- [3] I. Guyon, J. Weston Barnhill, S. Barnhill, & V. Vapnik. "Gene selection for cancer classification using support vector machines." *Machine learning*, Vol. 46, No. 1, pp. 389–422, 2002.
- [4] L. Jiang, D. Wang, Z. Cai, and X. Yan. "Survey of improving naive Bayes for classification." *Advanced Data Mining and Applications*, Vol. 4632, pp. 134–145, 2007.
- [5] S. Steiniger, M. Neun, A. Edwardes. "Foundations of Location Based Services Lesson 1 CartouChe 1-Lecture Notes on LBS, V. 1.0." 2011.
- [6] H. Lee, I. Park, K. Hong. "Design and implementation of a mobile devices-based real-time location tracking." *The Second International on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2008.

# การวิเคราะห์ข้อมูลภาพเพื่อติดตามปริมาณน้ำฝนและคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย Analysis of Image-Based Data to Track Rainfall and Forecast Flood Risk Areas

กัลยรัตน์ เจริญปุระ (Kanyarat Charoenpru)<sup>1</sup> และปรารณา ดีประเสริฐกุล (Prattana Deeprasertkul)<sup>2</sup>

สถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน)

<sup>1</sup>kanyarat@hii.or.th, <sup>2</sup>prattana@hii.or.th

## บทคัดย่อ

สถานการณ์อุทกภัยสามารถก่อให้เกิดความเสียหายกับผู้อยู่อาศัยในพื้นที่ที่ประสบภัยในบริเวณนั้น ดังนั้นการคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย (Forecasting flood risk areas) เพื่อที่จะสามารถเตรียมความพร้อมสำหรับการเกิดอุทกภัยเป็นสิ่งที่ช่วยลดโอกาสเกิดความเสียหายที่อาจเกิดจากอุทกภัยได้ บทความนี้กล่าวถึงระบบการคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย (Forecasting flood risk areas) โดยนำข้อมูลปริมาณน้ำฝนที่ได้จากการวิเคราะห์ข้อมูลภาพเรดาร์ตรวจอากาศ (Weather radar image) ซึ่งเป็นผลลัพธ์จากระบบการตรวจจับปริมาณน้ำฝนจากภาพเรดาร์เป็นข้อมูลนำเข้าของระบบ จากนั้นนำข้อมูลที่ได้ออกมาคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย โดยระบบที่พัฒนาขึ้นนี้ใช้ข้อมูลบันทึกเหตุการณ์อุทกภัยในอดีตเป็นชุดข้อมูลสำหรับระบบใช้เรียนรู้ และนำหลักการของเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ซึ่งเป็นหนึ่งในเทคนิคการเรียนรู้ของเครื่อง (Machine learning) มาใช้ในการออกแบบและพัฒนาระบบการคาดการณ์ให้มีประสิทธิภาพ โดยระบบนี้จะแสดงผลพื้นที่เสี่ยงภัยน้ำท่วมจากการตรวจจับค่าฝนของภาพเรดาร์และจะถูกพัฒนาต่อยอดให้สามารถแสดงผลรายชั่วโมงตามข้อมูลภาพเรดาร์ตรวจอากาศ

**คำสำคัญ:** การคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย ภาพเรดาร์ตรวจอากาศ เทคนิคต้นไม้ตัดสินใจ เทคนิคการเรียนรู้ของเครื่อง

## Abstract

The flood situation can cause damage to the residents living in the affected area. Therefore,

forecasting flood risk areas, in order to prepare for flooding, is a way to mitigate the damage that may be caused by floods. In this paper, we present forecasting flood risk areas system using rainfall from the analysis of weather radar image which is the result of rainfall detection system from radar images as input data of this system. Then use the data to predict flood risk areas. The developed system uses historical flood event as a data set for the system to learn and apply the principles of Decision Tree Technique, which is one of Machine learning techniques, to design and develop an efficient forecasting system. The system will display the results of flood risk areas from rainfall radar image detection and will be further developed to be able to display hourly results based on weather radar image data.

**Keywords:** Forecasting Flood Risk Areas, Weather Radar Image, Decision Tree, Machine Learning

## 1. บทนำ

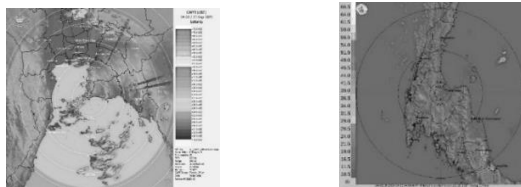
อุทกภัย [1] หรือน้ำท่วมเป็นภัยธรรมชาติที่เกิดจากน้ำเข้าท่วมพื้นที่ที่ซึ่งปกติไม่ได้อยู่ใต้อุณหภูมิระดับน้ำ หรือเกิดจากการสะสมน้ำบนพื้นที่ที่ระบายออกไม่ทันทำให้พื้นที่นั้นปกคลุมไปด้วยน้ำ สาเหตุหลักของการเกิดอุทกภัยนี้เกิดจากการที่ฝนตกหนักต่อเนื่องทำให้มีปริมาณฝนสะสมบนพื้นที่มาก ซึ่งเหตุการณ์อุทกภัยนี้สามารถก่อให้เกิดความเสียหายทั้งต่อชีวิต ทรัพย์สินและสิ่งแวดล้อมตามธรรมชาติ รวมไปถึงความเสียหายทางด้านเศรษฐกิจของพื้นที่ที่ประสบภัยได้อีกด้วย

เพื่อให้การติดตามสถานการณ์น้ำท่วมมีประสิทธิภาพ จึงได้พัฒนาระบบการวิเคราะห์ข้อมูลภาพเพื่อติดตามปริมาณน้ำฝนและคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย ซึ่งระบบนี้เป็นระบบคาดการณ์พื้นที่เสี่ยงภัยน้ำท่วมจากการตรวจจับปริมาณฝนของภาพเรดาร์ตรวจอากาศ [2] โดยใช้ข้อมูลบันทึกเหตุการณ์อุทกภัยในอดีตเป็นชุดข้อมูลสำหรับระบบได้เรียนรู้ และนำหลักการของเทคนิคต้นไม้ตัดสินใจซึ่งเป็นหนึ่งในเทคนิคการเรียนรู้ของเครื่องมาใช้ในการพัฒนาระบบนี้

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ภาพเรดาร์ตรวจอากาศ

ภาพเรดาร์ตรวจอากาศ [3] เป็นผลผลิตของเรดาร์ตรวจอากาศ โดยเรดาร์ตรวจอากาศนั้นคือเครื่องมือทางไฟฟ้าอิเล็กทรอนิกส์ที่ใช้สำหรับการตรวจวัดสภาพอากาศในระยะไกล ข้อดีของภาพเรดาร์ตรวจอากาศคือสามารถทราบถึงบริเวณที่มีฝนในพื้นที่ระยะรัศมีเรดาร์ทำการซึ่งเป็นระยะทางหลายร้อยกิโลเมตรรวมถึงตรวจหาความเข้มของฝนที่กำลังเกิดขึ้นจริงในพื้นที่ได้อีกด้วย



(ก) ภาพเรดาร์สดดิบ (ข) ภาพเรดาร์สุราษฎร์ธานี

ภาพที่ 1: ตัวอย่างข้อมูลภาพเรดาร์ตรวจอากาศ

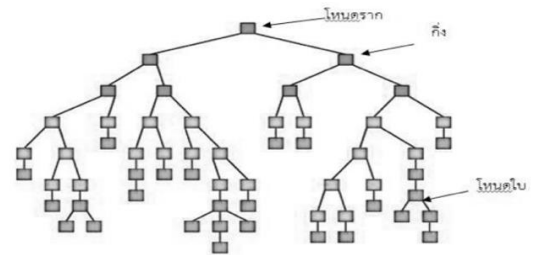
### 2.2 ระบบวิเคราะห์การตรวจจับปริมาณน้ำฝนจากภาพเรดาร์

ระบบวิเคราะห์การตรวจจับปริมาณน้ำฝนจากภาพเรดาร์ [2] เป็นระบบวิเคราะห์หาค่าปริมาณน้ำฝนในพื้นที่ที่สนใจจากภาพเรดาร์ให้มีประสิทธิภาพและมีความแม่นยำ โดยใช้เทคนิคการประมวลผลภาพในการหาค่าสัญญาณสะท้อนกลับของคลื่นเรดาร์ ใช้เทคนิคด้านภูมิสารสนเทศในการหาขอบเขตพื้นที่ที่สนใจบนภาพเรดาร์ และใช้สมการความสัมพันธ์ระหว่างค่าสัญญาณสะท้อนกลับของคลื่นเรดาร์กับค่าความเข้มของฝนในการหาค่า

ความเข้มของฝน ซึ่งผลลัพธ์ที่ได้ของระบบคือ ค่าปริมาณน้ำฝนของพื้นที่ที่สนใจบนภาพเรดาร์

### 2.3 เทคนิคต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ (Decision Tree) [4] คือเทคนิคการสร้างแบบจำลองโดยนำข้อมูลที่มีมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ เทคนิคนี้เป็นหนึ่งในเทคนิคของการเรียนรู้ของเครื่อง (Machine learning) ประเภทการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised learning) ซึ่งหลักการของต้นไม้ตัดสินใจคือจำแนกประเภทข้อมูลที่มีแบ่งออกเป็นกลุ่ม (class) โดยใช้คุณลักษณะข้อมูล (attribute) ในการจำแนกประเภท เทคนิคนี้ประกอบไปด้วย 3 ส่วนหลัก [5] ได้แก่ โหนดภายใน (internal node) คือคุณลักษณะของข้อมูล (attribute) โดยโหนดภายในที่เป็นจุดเริ่มต้นของต้นไม้เรียกว่า โหนดราก กิ่ง (branch) คือ ค่าของคุณลักษณะในโหนดภายใน และ โหนดใบ (leaf node) คือกลุ่ม (class) ซึ่งเป็นผลลัพธ์ในการจำแนกประเภทข้อมูล

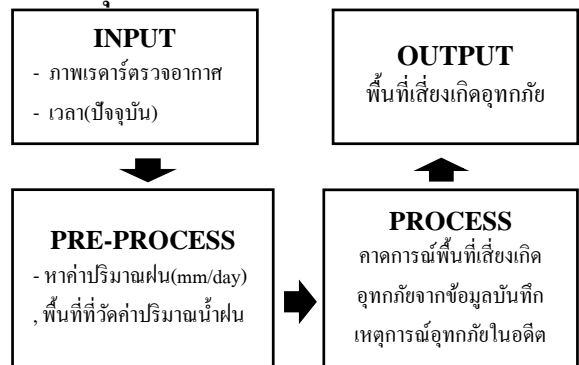


ภาพที่ 2: ส่วนประกอบของต้นไม้ตัดสินใจ

## 3. วิธีการดำเนินการวิจัย

### 3.1 กระบวนการทำงานระบบการวิเคราะห์

ข้อมูลภาพเพื่อติดตามปริมาณน้ำฝนและคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย



ภาพที่ 3: กระบวนการทำงานการวิเคราะห์ข้อมูลภาพ เพื่อติดตามปริมาณน้ำฝนและคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย

จากภาพที่ 3 แสดงถึงกระบวนการวิเคราะห์ ข้อมูลภาพเพื่อติดตามปริมาณน้ำฝนและคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย โดยเริ่มจากข้อมูลนำเข้าของระบบคือ ภาพเรดาร์และเวลาซึ่งมีรายละเอียดดังนี้

### 3.2 ข้อมูลนำเข้า (Input)

#### 3.2.1 ภาพเรดาร์ตรวจอากาศ (ปัจจุบัน)

ภาพเรดาร์ตรวจอากาศที่นำมาเป็นข้อมูลนำเข้าใน ระบบที่พัฒนาขึ้นนี้ เป็นภาพจากสถานีเรดาร์ ณ ตอนที่ เรดาร์ทำการตรวจจับปริมาณน้ำฝน (ปัจจุบัน)ในพื้นที่ 20 จังหวัด ได้แก่ เชียงราย ลำพูน ขอนแก่น ชัยนาท สกลนคร สมุทรสงคราม สุราษฎร์ธานี กรุงเทพมหานคร พิษณุโลก อุบลราชธานี ชุมพร ภูเก็ต เชียงใหม่ น่าน นราธิวาส นครสวรรค์ สมุทรปราการ นครราชสีมา ประจวบคีรีขันธ์ สงขลา

#### 3.2.2 เวลา (ปัจจุบัน)

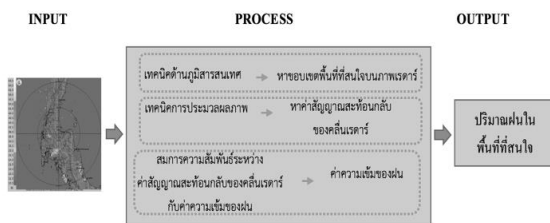
เวลา (ปัจจุบัน) คือเวลา ณ ตอนที่เรดาร์ทำการ ตรวจจับปริมาณน้ำฝน

จากนั้นทำการหาค่าปริมาณฝน (ปัจจุบัน) และพื้นที่ที่ วัดค่าปริมาณน้ำฝน (ปัจจุบัน) ซึ่งเป็นขั้นตอนการเตรียม ข้อมูลก่อนที่จะเข้าสู่กระบวนการหลัก (pre-process) โดยในการหาค่านั้นหาได้จากระบบวิเคราะห์การ ตรวจจับปริมาณน้ำฝนจากภาพเรดาร์ซึ่งมีกระบวนการ ดังนี้

### 3.3 การเตรียมข้อมูลกระบวนการหลัก (pre-process)

#### 3.3.1 ระบบวิเคราะห์การตรวจจับปริมาณน้ำฝนจาก

ภาพเรดาร์



ภาพที่ 4: ระบบวิเคราะห์การตรวจจับปริมาณน้ำฝนจากภาพเรดาร์

จากภาพที่ 4 แสดงระบบวิเคราะห์การตรวจจับ ปริมาณน้ำฝนจากภาพเรดาร์เริ่มจากการรับข้อมูล (Input) เป็นภาพเรดาร์ฝน จากนั้นหาพื้นที่ที่สนใจโดยใช้ เทคนิคด้านภูมิสารสนเทศซึ่งในการศึกษาครั้งนี้คือพื้นที่ จังหวัดที่เป็นที่ตั้งของสถานีเรดาร์ เมื่อได้พื้นที่จังหวัด

บนภาพที่ต้องการแล้วจึงหาค่าปริมาณฝนโดยใช้เทคนิค การประมวลผลภาพและสมการ ฉะนั้นผลลัพธ์ที่ได้จาก ระบบนี้คือ ปริมาณฝนในพื้นที่ที่สนใจ โดยพื้นที่ที่สนใจ ที่ได้ คือ พื้นที่ที่วัดค่าปริมาณน้ำฝน (ปัจจุบัน) และ ปริมาณฝนคือค่า ปริมาณน้ำฝน (ปัจจุบัน) หน่วย มิลลิเมตร / ชั่วโมง แต่เนื่องจากข้อมูลบันทึกเหตุการณ์ อุทกภัยในอดีตซึ่งเป็นชุดข้อมูลสำหรับกระบวนการหลัก ได้บันทึกข้อมูลปริมาณฝนในหน่วยมิลลิเมตร / วัน จึงได้ แปลงจากปริมาณฝน(ปัจจุบัน) หน่วยมิลลิเมตร / ชั่วโมง เป็น หน่วยมิลลิเมตร / วัน

### 3.4 กระบวนการหลัก (Process)

กระบวนการหลักของระบบที่พัฒนาขึ้นนี้คือการ คาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย โดยใช้ข้อมูลบันทึก เหตุการณ์อุทกภัยในอดีตเป็นชุดข้อมูลสำหรับระบบได้ เรียนรู้ และนำหลักการของเทคนิคต้นไม้ตัดสินใจใช้ ออกแบบและพัฒนาระบบนี้ ซึ่งมีรายละเอียดดังนี้

#### 3.4.1 ข้อมูลบันทึกเหตุการณ์อุทกภัยในอดีต

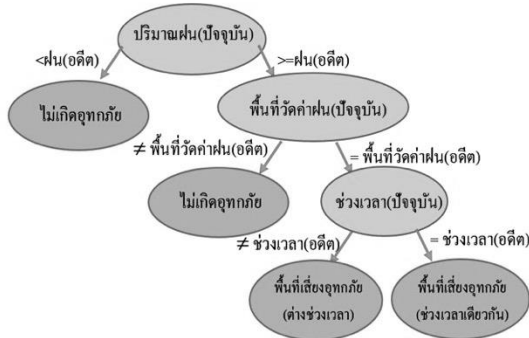
ข้อมูลได้บันทึกเหตุการณ์อุทกภัยตั้งแต่ปี 2545 – 2564 จากเดิมข้อมูลนี้ไม่มีรูปแบบของข้อมูลที่ชัดเจน (unstructured) จึงได้ทำการจัดและเรียบเรียงข้อมูลเพื่อ สามารถนำมาใช้ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น โดย ได้จัดทำขึ้นใหม่เป็น ไฟล์ excel ดังตัวอย่างต่อไปนี้

no.	name	year	month	mea_area	rainfall	flood_area
1	เหตุการณ์น้ำท่วมบริเวณภาคใต้ตอนล่าง	2554	1	นราธิวาส/ปัตตานี/สงขลา	>90	(นราธิวาส(ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์),ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์)
2	เหตุการณ์น้ำท่วมบริเวณจังหวัดปัตตานีและนราธิวาส	2553	12	ปัตตานี/นราธิวาส	>90	(ปัตตานี(ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์),ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์), (นราธิวาส(ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์),ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์,ปีเกิดเหตุการณ์)

ภาพที่ 5: ตัวอย่างข้อมูลบันทึกเหตุการณ์อุทกภัยในอดีต

จากภาพที่ 5 ตัวอย่างข้อมูลบันทึกเหตุการณ์อุทกภัย ในอดีตแสดงถึงการจัดและเรียบเรียงการบันทึกข้อมูล บนไฟล์ excel โดยแบ่งข้อมูลเป็น 6 ประเภท ได้แก่ ชื่อ เหตุการณ์อุทกภัยในอดีต (name) ปีที่เกิดเหตุการณ์ อุทกภัยในอดีต (year) เดือนที่เกิดเหตุการณ์อุทกภัยใน อดีต (month) พื้นที่ที่วัดค่าปริมาณน้ำฝนของเหตุการณ์ อุทกภัยในอดีต (mea\_area) ปริมาณน้ำฝนของเหตุการณ์ อุทกภัยในอดีต (rainfall) หน่วย มิลลิเมตร/วันและพื้นที่ เกิดอุทกภัยของเหตุการณ์อุทกภัยในอดีต (flood\_area) ซึ่งเป็นข้อมูลพื้นที่ในระดับจังหวัดและระดับอำเภอ

หลังจากทำการจัดและเรียบเรียงข้อมูลเสร็จสิ้นได้เป็นข้อมูลบันทึกเหตุการณ์อุทกภัยในอดีตตามรูปแบบดังภาพที่ 5 จากนั้นนำข้อมูลที่ได้ไปใช้คาดการณ์การหาพื้นที่เสี่ยงอุทกภัยซึ่งเป็นกระบวนการหลักของระบบที่พัฒนาขึ้น โดยกระบวนการทำงานดังนี้



ภาพที่ 6: กระบวนการหลักระบบการวิเคราะห์ข้อมูลภาพเพื่อติดตามปริมาณน้ำฝนและคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย

### 3.5 ผลลัพธ์

ผลลัพธ์ที่ได้จากระบบการวิเคราะห์ข้อมูลภาพเพื่อติดตามปริมาณน้ำฝนและคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัยมี 3 ประเภท คือ ไม่เกิดอุทกภัย พื้นที่เสี่ยงอุทกภัย (ช่วงเวลาเดียวกัน) และพื้นที่เสี่ยงอุทกภัย (ต่างช่วงเวลา) ตัวอย่างการแสดงผลของระบบมีดังนี้

#### ตัวอย่างที่ 1

- ข้อมูลนำเข้า : ปริมาณฝน (ปัจจุบัน) = 10 มิลลิเมตร / วัน พื้นที่ที่วัดปริมาณน้ำฝน (ปัจจุบัน) = จังหวัดนราธิวาส เวลา(ปัจจุบัน) = เดือนพฤศจิกายน
- ผลลัพธ์ : ไม่เกิดอุทกภัย

#### ตัวอย่างที่ 2

- ข้อมูลนำเข้า : ปริมาณฝน (ปัจจุบัน) = 90 มิลลิเมตร / วัน พื้นที่ที่วัดปริมาณน้ำฝน (ปัจจุบัน) = จังหวัดสงขลา เวลา (ปัจจุบัน) = เดือนธันวาคม
- ผลลัพธ์ :
  - พื้นที่เสี่ยงเกิดอุทกภัย(ช่วงเวลาเดียวกัน) ได้แก่ อ.ระโนด อ.กระแสดินธุ์ อ.สทิงพระ อ.สิงหนคร อ.ควนเนียง อ.บางกล่ำ อ.หาดใหญ่
  - พื้นที่เสี่ยงเกิดอุทกภัย(ต่างช่วงเวลา) ได้แก่ อ.หาดใหญ่ อ.ระโนด อ.ควนเนียง อ.สิงหนคร อ.รัตภูมิ อ.กระแสดินธุ์ อ.บางกล่ำ อ.สทิงพระ

- อ.คลองหอยโข่ง อ.สะเดา อ.เมือง อ.สะบ้าย้อย
- อ.นาทวี อ.จะนะ

### 4. ผลการทดสอบ

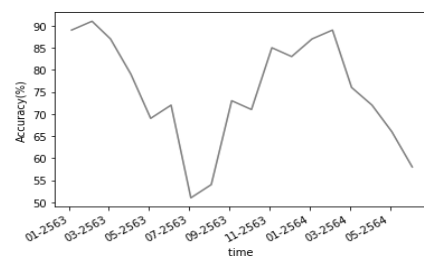
การทดสอบนี้เพื่อให้ทราบถึงความแม่นยำของระบบที่พัฒนาขึ้น โดยเริ่มจากแบ่งชุดข้อมูลบันทึกเหตุการณ์อุทกภัยในอดีตเป็น 2 ส่วน ส่วนที่ 1 บันทึกเหตุการณ์อุทกภัยในอดีต(พ.ศ. 2545 – พ.ศ. 2562) โดยส่วนนี้ใช้สำหรับระบบและส่วนที่ 2 บันทึกเหตุการณ์อุทกภัยในอดีต (พ.ศ. 2563 – พ.ศ. 2564) ส่วนนี้ใช้ในการทดสอบระบบ จากนั้นนำภาพเรดาร์ฝนจาก 20 จังหวัด ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึง 30 มิถุนายน พ.ศ. 2564 มาใช้เป็นข้อมูลนำเข้าของโปรแกรมและนำผลลัพธ์ที่ได้จากโปรแกรมมาทดสอบโดยการนำมาเทียบกับเหตุการณ์ที่เกิดขึ้นจริงจากบันทึกเหตุการณ์อุทกภัยในอดีต (ข้อมูลส่วนที่ 2) ณ วันที่เดียวกัน โดยเงื่อนไขมีดังนี้

- ถ้าผลลัพธ์ที่ได้จากโปรแกรม เท่ากับ เหตุการณ์ที่เกิดขึ้นจริงจากบันทึกเหตุการณ์อุทกภัยในอดีต
  - ให้ x เท่ากับ 1
- ถ้าผลลัพธ์ที่ได้จากโปรแกรม ไม่เท่ากับ เหตุการณ์ที่เกิดขึ้นจริงจากบันทึกเหตุการณ์อุทกภัยในอดีต
  - ให้ x เท่ากับ 0

จากนั้นนำค่าที่ได้มาหาค่าเปอร์เซ็นต์ความแม่นยำของโปรแกรมจากสมการดังต่อไปนี้

$$\%Accuracy = \frac{\sum x}{n} \times 100 \quad (4-1)$$

โดยที่ %Accuracy คือ ค่าเปอร์เซ็นต์ความแม่นยำข้อมูล  $\sum x$  คือ ผลรวมของผลการเทียบผลลัพธ์ที่ได้จากระบบกับบันทึกเหตุการณ์อุทกภัยในอดีต n คือ จำนวนครั้งการเทียบผลลัพธ์ที่ได้จากระบบกับบันทึกเหตุการณ์อุทกภัยในอดีตทั้งหมด โดยผลการทดสอบที่ได้มีดังนี้



ภาพที่ 7: ผลการทดสอบระบบการวิเคราะห์ข้อมูลภาพเพื่อติดตามปริมาณน้ำฝนและคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัย

จากภาพที่ 7 แสดงผลการทดสอบระบบการวิเคราะห์ข้อมูลภาพเพื่อติดตามปริมาณน้ำฝนและคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัยโดยนำภาพเรดาร์ฝนจาก 20 จังหวัด ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึง 30 มิถุนายน พ.ศ.2564 เป็นข้อมูลนำเข้าสำหรับการทดสอบนี้ ซึ่งค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของข้อมูลทั้งหมดคือ 73.42 เปอร์เซ็นต์ โดยเดือนกุมภาพันธ์ พ.ศ. 2563 มีค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของข้อมูลมากที่สุดคือ 91.2 เปอร์เซ็นต์และเดือนกรกฎาคม พ.ศ. 2563 มีค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของข้อมูลน้อยที่สุดคือ 52.17 เปอร์เซ็นต์ ซึ่งจากการวิเคราะห์ผลการทดสอบพบว่าวันที่เดือนกรกฎาคม พ.ศ. 2563 มีค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของข้อมูลน้อยที่สุด เนื่องจากเป็นเดือนที่มีฝนตกถี่มากที่สุด ฉะนั้นเมื่อทดสอบจึงเกิดโอกาสสะสมค่าผิดพลาดของข้อมูลมากกว่าเดือนอื่นและส่งผลให้ค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำน้อยที่สุดด้วย

## 5. สรุปผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบการคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัยโดยนำข้อมูลปริมาณน้ำฝนที่ได้จากการวิเคราะห์ข้อมูลภาพเรดาร์ตรวจอากาศ ซึ่งเป็นผลลัพธ์จากระบบการตรวจจับปริมาณน้ำฝนจากภาพเรดาร์เป็นข้อมูลนำเข้าของระบบ จากนั้นนำข้อมูลที่ได้ออกมาคาดการณ์พื้นที่เสี่ยงเกิดอุทกภัยโดยระบบที่พัฒนาขึ้นนี้ใช้ข้อมูลบันทึกเหตุการณ์อุทกภัยในอดีตเป็นชุดข้อมูลสำหรับระบบใช้เรียนรู้ และนำหลักการของเทคนิคต้นไม้ตัดสินใจมาใช้ในการออกแบบและพัฒนา ระบบ โดยผลลัพธ์ที่ได้จากระบบที่พัฒนาขึ้นนี้แสดงพื้นที่เสี่ยงภัยน้ำท่วมจากการตรวจจับค่าน้ำฝนของภาพเรดาร์ จากผลการทดสอบระบบสรุปได้ว่าค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของข้อมูลทั้งหมดคือ 73.42 เปอร์เซ็นต์ โดยในอนาคตจะพัฒนาต่อยอดให้ผลลัพธ์ที่

ได้จากระบบมีความแม่นยำเพิ่มขึ้นและให้ระบบสามารถแสดงผลพื้นที่รายชั่วโมงตามข้อมูลภาพเรดาร์ตรวจอากาศ

## เอกสารอ้างอิง

- [1] ศูนย์เตือนภัยพิบัติแห่งชาติ. “รู้จักภัยจาก อุทกภัย หรือ น้ำท่วม.” [สืบค้นเมื่อวันที่ 5 กันยายน 2564]. จาก [http://ndwc.disaster.go.th/cmsdetail.ndwc-9.283/cms/menu\\_7525/4214.3/รู้จักภัยจาก+อุทกภัย+หรือ+น้ำท่วม, 2561](http://ndwc.disaster.go.th/cmsdetail.ndwc-9.283/cms/menu_7525/4214.3/รู้จักภัยจาก+อุทกภัย+หรือ+น้ำท่วม, 2561).
- [2] กัลยรัตน์ เจริญปฐุ. และปรารณา ดิประเสริฐกุล. “ระบบวิเคราะห์การตรวจจับปริมาณน้ำฝนจากภาพเรดาร์.” *งานประชุมวิชาการด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 17*, 2564.
- [3] ส่วนติดตามสภาวะอากาศศูนย์อุตุนิยมวิทยาภาคตะวันออกเฉียงเหนือตอนล่าง. “องค์ความรู้เรื่องเรดาร์ตรวจอากาศเพื่อประชาชน.” [สืบค้นเมื่อวันที่ 11 กันยายน 2564]. จาก <http://www.ubonmet.tmd.go.th/index.php/km-base/71-km-base-2560, 2560>.
- [4] T. Mitchell. *Machine Learning*. McGraw Hill, 1997.
- [5] B. Kijisirikul. “Data Mining Algorithms.” *the final report on the Joint Government and Private Sectors*, Chulalongkorn University Press, 2004.

# การแก้ปัญหาความไม่สมดุลของข้อมูลสำหรับการพยากรณ์การเสียชีวิตจากอุบัติเหตุ

## Imbalanced Data Problem Solving in Predicting Accidental Deaths

ธนิษฐ์ วาสุพฐิติ (Thaninthorn Whasphutthisit)<sup>1</sup> และวัชรวิกรม จิตต์สกุล (Watchareewan Jitsakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858058@email.kmutnb.ac.th, <sup>2</sup>watchareewan.j@itd.kmutnb.ac.th

### บทคัดย่อ

บทความวิจัยฉบับนี้เป็นการนำเสนอการพยากรณ์การเสียชีวิตจากอุบัติเหตุ โดยทำการเปรียบเทียบประสิทธิภาพจำนวน 3 อัลกอริทึม ได้แก่ อัลกอริทึมป่าสุ่ม อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนและอัลกอริทึมแบร์นูลลีนาอิวเบย์ร่วมกับแก้ปัญหาความไม่สมดุลของข้อมูลด้วยการสังเคราะห์ข้อมูลเพิ่ม โดยข้อมูลที่ใช้เป็นข้อมูลอุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคม เดือนมกราคม – เมษายน พ.ศ. 2564 เป็นชุดข้อมูลเปิดภาครัฐ เผยแพร่ผ่านระบบบัญชีข้อมูลด้านคมนาคม เว็บไซต์ <https://datagov.mot.go.th> จัดทำโดยกระทรวงคมนาคม จำนวน 8,560 ระเบียบ จำนวน 46 คุณลักษณะ เพื่อศึกษาคุณลักษณะของตัวแปรที่มีผลต่อการพยากรณ์การเสียชีวิตจากอุบัติเหตุ ประเมินประสิทธิภาพของอัลกอริทึม โดยพิจารณาจากความถูกต้อง ทั้งนี้ผลการพยากรณ์พบว่า อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ร่วมกับการแก้ปัญหาความไม่สมดุลของข้อมูล ให้ประสิทธิภาพการพยากรณ์การเสียชีวิตจากอุบัติเหตุสูงที่สุด ค่าความถูกต้องร้อยละ 94.03 สามารถเพิ่มประสิทธิภาพในการพยากรณ์ร้อยละ 5.05

**คำสำคัญ:** ความไม่สมดุลของข้อมูล อัลกอริทึมป่าสุ่ม  
อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน  
อัลกอริทึมแบร์นูลลีนาอิวเบย์

### Abstract

*This research presents to compare prediction models for road deaths on road network by data mining techniques. In this work, the classifier is selected from four prediction algorithms: Random*

*Forest (RF), Support Vector Machine (SVM), and Bernoulli Naïve Bayes (BNB) with solving data imbalances using the Synthetic Minority Oversampling Technique (SMOTE). The road deaths on road network data in road accident data set of the Ministry of Transport, Thailand from January to April 2021. It has up to 8,560 records 46 attributes. This research has measured performance models with accuracy. The comparative results showed that the accuracy of SVM is the most appropriate for predicting road deaths on road network with accuracy 94.03% after solving data imbalance can improve forecasting efficiency by 5.05%.*

**Keywords:** Imbalanced Data, Random Forest, Support Vector Machine, Bernoulli Naïve Bayes.

### 1. บทนำ

ในปัจจุบันสถิติการเสียชีวิตจากอุบัติเหตุทางถนนเพิ่มสูงขึ้นอย่างต่อเนื่อง โดยแต่ละปีมีผู้เสียชีวิตจากอุบัติเหตุทางถนนทั่วโลกประมาณ 1.3 ล้านคน ซึ่งองค์การอนามัยโลก (World Health Organization) ได้มีการเผยแพร่สถิติการเกิดอุบัติเหตุทางถนนของประเทศต่าง ๆ โดยปี พ.ศ. 2562 ประเทศไทยมีสถิติการเสียชีวิตจากอุบัติเหตุทางถนนเป็นอันดับที่ 9 ของโลก มีผู้เสียชีวิตเฉลี่ยปีละ 22,491 ราย คิดเป็น 32.7 คน ต่อประชากร 1 แสนคน นับเป็นอันดับหนึ่งในเอเชีย และในภูมิภาคอาเซียน และสูงกว่าค่าเฉลี่ยของโลก [1] สำหรับถนนในประเทศไทยส่วนใหญ่ร้อยละ 86 อยู่ในความดูแลขององค์กรปกครองส่วนท้องถิ่น กระทรวงมหาดไทย และถนนประมาณร้อยละ 14 อยู่ใน



ความรับผิดชอบของกระทรวงคมนาคมประกอบด้วย

1) ทางหลวงแผ่นดิน คูแผลโดยกรมทางหลวง ระยะทาง 51,936.84 กิโลเมตร 2) ทางหลวงชนบท คูแผลโดยกรมทางหลวงชนบท ระยะทาง 48,597.75 กิโลเมตร และ 3) ทางพิเศษ คูแผลโดยการทางพิเศษแห่งประเทศไทย ระยะทาง 224.60 กิโลเมตร ในปี พ.ศ. 2562 พบว่า มีจำนวนการเกิดอุบัติเหตุบนถนนในความรับผิดชอบของกระทรวงคมนาคม จำนวน 20,007 ครั้ง มีผู้เสียชีวิต 3,277 ราย และมีผู้บาดเจ็บ 18,217 คน ซึ่งสูงกว่าปี พ.ศ. 2561 และมีแนวโน้มเพิ่มขึ้นในช่วง 5 ปีที่ผ่านมา [2] โดยการเกิดอุบัติเหตุทางถนนในแต่ละครั้งส่งผลให้มีผู้บาดเจ็บ ผู้เสียชีวิต ทรัพย์สินเสียหาย และเกิดการสูญเสียทางเศรษฐกิจจำนวนมาก

จากข้อมูลและสถิติอุบัติเหตุทางถนนของหน่วยงานต่าง ๆ ได้มีผู้วิจัยทำการสร้างแบบจำลองพยากรณ์ที่เกี่ยวข้องกับอุบัติเหตุ ด้วยอัลกอริทึมมัลติเลเยอร์เพอร์เซ็ปตรอน และอัลกอริทึม Meta bagging สำหรับการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศกาลปีใหม่ [3] หรืออัลกอริทึมโครงข่ายประสาทเทียมสำหรับการศึกษาปัจจัยที่ทำให้เกิดอุบัติเหตุรุนแรง [4] การพยากรณ์ประเภทพาหนะที่เกิดอุบัติเหตุชนแล้วหนี ด้วยอัลกอริทึมการวิเคราะห์ถดถอยโลจิสติก อัลกอริทึมการวิเคราะห์การจำแนกประเภทเชิงเส้น อัลกอริทึมต้นไม้ตัดสินใจแบบจำแนก และถดถอย อัลกอริทึมเพื่อนบ้านใกล้ที่สุด และอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน [5] การพยากรณ์การเสียชีวิตจากอุบัติเหตุบนโครงข่ายถนนด้วยอัลกอริทึมป่าสุ่ม อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน อัลกอริทึมเพื่อนบ้านใกล้ที่สุด และอัลกอริทึม โครงข่ายประสาทเทียม [6] รวมถึงการแก้ปัญหาความไม่สมดุลของข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งหูในชั้นกลาง หรือหูน้ำหนวกด้วยอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ชุดข้อมูลยอคคองเกลือในบัตรเครดิตของลูกค้าโดยใช้ อัลกอริทึมเพื่อนบ้านใกล้ที่สุดและชุดข้อมูลคุณภาพไวน์แดง โดยใช้อัลกอริทึม โครงข่ายประสาทเทียมโดยทั้ง 3 ชุดข้อมูลทำงานร่วมกับเทคนิค SMOTE ส่งผลให้ประสิทธิภาพของแบบจำลองมีประสิทธิภาพดีที่สุด [7]

จากปัญหา และงานวิจัยดังกล่าว ผู้วิจัยมีแนวคิดในการสร้างแบบจำลองพยากรณ์ร่วมกับการสังเคราะห์ข้อมูลเพิ่ม เพื่อแก้ปัญหาความไม่สมดุลของข้อมูล เพื่อช่วยในการพยากรณ์การเสียชีวิตจากอุบัติเหตุ

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้รวบรวมข้อมูลทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยแบ่งเป็น 2 หัวข้อหลัก ดังนี้

### 2.1 อัลกอริทึมการพยากรณ์

**2.1.1 อัลกอริทึมป่าสุ่ม** เป็น อัลกอริทึม ที่มีการพัฒนาจากอัลกอริทึมต้นไม้ตัดสินใจ ทำการเพิ่มจำนวนต้นไม้หลาย ๆ ต้น เพื่อเพิ่มประสิทธิภาพการพยากรณ์ โดยแต่ละต้นจะได้รับข้อมูล และคุณลักษณะที่ไม่เหมือนกัน เริ่มจากการสุ่มเลือกคุณลักษณะจากชุดข้อมูล ทำการสร้างต้นไม้ตัดสินใจ และหาค่าพยากรณ์ของต้นไม้แต่ละต้น เลือกจำนวนต้นไม้ตัดสินใจที่ต้องการหาเป็นการสร้างแบบจำลองการจำแนกใช้วิธีผลโหวตมากที่สุด [8]

**2.1.2 อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน** เป็นอัลกอริทึมที่มีประสิทธิภาพการจำแนกข้อมูลที่มีมิติจำนวนมาก รวมถึงข้อมูลที่มีความคลุมเครือได้ โดยใช้ฟังก์ชัน Kernel เพื่อนำค่าของข้อมูลไปวางไว้ Feature Space และคำนวณหาเส้นสำหรับแบ่งข้อมูลทั้งสองออกจากกัน โดยสร้างเส้นแบ่งเส้นตรง และคว่ำเส้นตรงใดเป็นเส้นตรงดีที่สุด โดยหาเส้นตรงที่มีการแบ่งกลุ่มที่กว้างที่สุดสามารถแบ่งข้อมูลเป็น 2 คลาส [9]

**2.1.3 อัลกอริทึม นาอ็ฟเบย์** เป็น อัลกอริทึม การจำแนกประเภท โดยใช้ความน่าจะเป็น และวิธีการทางสถิติตามทฤษฎีเบย์ เพื่อหาความน่าจะเป็นที่น่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้าร่วมกับข้อมูล เช่นความน่าจะเป็นที่ได้สำหรับสมมติฐานหนึ่ง ๆ เพื่อหาสมมติฐานที่ดีที่สุด [10]

**2.1.4 เทคนิคการแก้ปัญหาความไม่สมดุลของข้อมูล** การจัดการความไม่สมดุลของข้อมูล (Imbalanced Data) หมายถึง การแก้ปัญหาข้อมูลไม่สมดุลของคลาสที่มีจำนวนข้อมูลที่แตกต่างกัน ซึ่งอาจส่งผลให้การพยากรณ์

มีความคลาดเคลื่อนและโน้มเอียงไปทางข้อมูลกลุ่มมาก วิธีการแก้ปัญหาคือการไม่สมดุลของข้อมูลมีหลายวิธี เช่น การสังเคราะห์ข้อมูลด้วยวิธีการสุ่มเพิ่ม เพื่อให้ข้อมูลกลุ่มน้อยมีปริมาณใกล้เคียงกับข้อมูลกลุ่มมาก หรือการสังเคราะห์ข้อมูลด้วยวิธีการสุ่มลด เพื่อให้ข้อมูลกลุ่มมากมีปริมาณใกล้เคียงกับข้อมูลกลุ่มน้อย เป็นต้น [11]

**2.2 งานวิจัยที่เกี่ยวข้อง**

Al-Zadid and et. al. [12] ได้ทำการศึกษาการสร้างแบบจำลองพยากรณ์โรคหัวใจ และหลอดเลือดหัวใจ ทำการเปรียบเทียบประสิทธิภาพการจำแนกด้วยอัลกอริทึม LogReg อัลกอริทึม RF อัลกอริทึม Multilayer Perceptron อัลกอริทึม Gaussian Naïve Bayes และเทคนิค Ensemble Hard Voting ใช้ชุดข้อมูลผู้ป่วยโรคหัวใจและหลอดเลือดหัวใจ จำนวน 3,751 ระเบียบงาน แก้ปัญหาความไม่สมดุลข้อมูลด้วยวิธีการสุ่มลด ผลการวิจัยพบว่าเทคนิค Ensemble Hard Voting ให้ประสิทธิภาพสูงที่สุด ค่าความแม่นยำร้อยละ 100 ค่าระลอก 0.043 ค่าการถ่วงดุล 0.082 และค่า AUC 0.73 ตามลำดับ

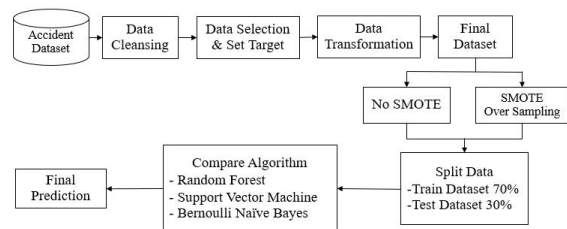
Abdullah and Mohammad [13] ได้ทำการศึกษาการพยากรณ์ข้อบกพร่องของซอฟต์แวร์ โดยทำการเปรียบเทียบประสิทธิภาพการจำแนกด้วยอัลกอริทึม Bagging อัลกอริทึม SVM อัลกอริทึม DT และอัลกอริทึม RF ข้อมูลที่ใช้เป็นชุดข้อมูลจาก NASA จำนวน 10 ชุดข้อมูล แก้ปัญหาความไม่สมดุลข้อมูลด้วยวิธีสังเคราะห์ข้อมูลเพิ่ม ผลการวิจัย พบว่า อัลกอริทึม RF ให้ประสิทธิภาพสูงที่สุดจากชุดข้อมูล 6 ใน 10 ชุดข้อมูล ทั้งหมด ได้แก่ ชุดข้อมูล PC1 ค่าความถูกต้องร้อยละ 91 ชุดข้อมูล PC3 ค่าความถูกต้องร้อยละ 84 PC4 ค่าความถูกต้องร้อยละ 90 ชุดข้อมูล KC2 ค่าความถูกต้องร้อยละ 82 ชุดข้อมูล MC1 ค่าความถูกต้องร้อยละ 97 ชุดข้อมูล CM1 ค่าความถูกต้องร้อยละ 83 ตามลำดับ

พัชรียา และคณะ [14] ได้ทำการศึกษาการทำนายผลการปรับความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล ทำการเปรียบเทียบประสิทธิภาพการจำแนกด้วย อัลกอริทึม KNN อัลกอริทึม DT อัลกอริทึม NN และอัลกอริทึม SVM ร่วมกับเทคนิค

การแก้ปัญหาคือการไม่สมดุลของข้อมูลจำนวน 4 วิธี ได้แก่ วิธีการสุ่มเกิน วิธีการสังเคราะห์ข้อมูลเพิ่ม วิธีการสุ่มลด และวิธีการผสมผสาน โดยใช้ชุดข้อมูลจำนวน 3 ชุด คือชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูชั้นนอก ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า และชุดข้อมูลคุณภาพไวน์แดง ผลการวิจัยพบว่าชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูชั้นนอก อัลกอริทึม SVM ร่วมกับการแก้ปัญหาคือการไม่สมดุลของข้อมูลด้วยวิธีการสังเคราะห์ข้อมูลเพิ่มให้ประสิทธิภาพสูงที่สุด ค่าความถูกต้อง 0.8509 ค่าความไว 0.8210 ค่าความจำเพาะ 0.8838 และค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย 0.1490 ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า อัลกอริทึม KNN ให้ประสิทธิภาพสูงที่สุด ร่วมกับการแก้ปัญหาคือการไม่สมดุลของข้อมูลด้วยวิธีการ วิธีการสังเคราะห์ข้อมูลเพิ่ม ค่าความถูกต้อง 0.7778 ค่าความไว 0.7834 ค่าความจำเพาะ 0.7680 และค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย 0.2157 และชุดข้อมูลคุณภาพไวน์แดง อัลกอริทึม NN ร่วมกับการแก้ปัญหาคือการไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน ค่าความถูกต้อง 0.5888 ค่าความไว 0.6703 ค่าความจำเพาะ 0.2861 และค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย 0.6772 ตามลำดับ

**3. วิธีการดำเนินการวิจัย**

การพยากรณ์การเสียชีวิตจากอุบัติเหตุ ผู้วิจัยมีกรอบแนวคิดแสดงดังภาพที่ 1



**ภาพที่ 1:** กรอบแนวคิดของงานวิจัย

จากภาพที่ 1 กรอบแนวคิดการวิจัย ประกอบด้วยขั้นตอนการดำเนินงานวิจัยจำนวน 4 ขั้นตอน ได้แก่ การรวบรวมข้อมูล (Data Gathering) การเตรียมข้อมูล (Data Preprocessing) การสร้างแบบจำลอง (Modeling)

โดยใช้อัลกอริทึม RF อัลกอริทึม SVM และ BNB และการประเมินผล (Evaluation) ดังนี้

### 3.1 การรวบรวมข้อมูล (Data Gathering)

การรวบรวมชุดข้อมูลสำหรับการพยากรณ์ เป็นข้อมูลอุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคม ตั้งแต่เดือนมกราคม – เมษายน พ.ศ. 2564 เป็นชุดข้อมูลเปิดภาครัฐ เผยแพร่ผ่านระบบบัญชีข้อมูลด้านคมนาคม เว็บไซต์ <https://datagov.mot.go.th> โดยกระทรวงคมนาคม จำนวน 8,560 ระเบียบ จำนวน 46 คุณลักษณะ แสดงดังตารางที่ 1

ตารางที่ 1: ตัวอย่างคุณลักษณะข้อมูล

คุณลักษณะ	ตัวอย่างข้อมูล
วันที่เกิดเหตุ	1/1/2021 14/2/2021
รถคันที่ 1	รถจักรยานยนต์ รถยนต์นั่งส่วนบุคคล/รถยนต์นั่งสาธารณะ
บริเวณที่เกิดเหตุ/ลักษณะทาง	ทางตรง+ไม่มีความลาดชัน ทางสามแยก (Y) ทางสามแยก (T) ทางสี่แยก
มูลเหตุสันนิษฐาน	เมาสุรา หลับใน คน/รถ/สัตว์ตัดหน้า กระชั้นชิด ขับเร็วเกินอัตรากำหนด
ลักษณะการเกิดอุบัติเหตุ	พลิกคว่ำ/ตกลงบนในทางตรง พลิกคว่ำ/ตกลงบนในทางโค้ง ชนท้าย
สภาพอากาศ	แจ่มใส มีหมอก/ควัน/ฝุ่น ฝนตก
DRIVER_SAFETY	ไม่สวมหมวกนิรภัย คาดเข็มขัดนิรภัย
เพศ	ชาย หญิง ไม่ระบุ

### 3.2 การเตรียมข้อมูล (Data Preprocessing)

กระบวนการเตรียมข้อมูลเพื่อนำเข้าสู่แบบจำลองประกอบด้วย 3 ขั้นตอน

**3.2.1 การทำความสะอาดข้อมูล** ได้แก่ การจัดการกับข้อมูลสูญหายโดยการแทนค่าข้อมูลจังหวัดที่เกิดอุบัติเหตุ การลบค่าว่างของข้อมูลรถคันที่ 1 ออกไป และการจัดการข้อมูลซ้ำซ้อน โดยทำการรวมแอททริบิวต์ข้อมูลบริเวณที่เกิดเหตุ/ลักษณะทาง มีความซ้ำซ้อนกับแอททริบิวต์บริเวณที่เกิดเหตุอื่น

**3.2.2 การคัดเลือกแอททริบิวต์** ทำการคัดเลือกโดยอ้างอิงจากแอททริบิวต์ของงานวิจัยที่เกี่ยวข้องกับอุบัติเหตุ สามารถคัดเลือกแอททริบิวต์ได้จำนวน 11 แอททริบิวต์ จำนวน 2,495 ระเบียบ ได้แก่ วันที่ เวลา จังหวัดที่เกิดเหตุ รถคันแรกที่เกิดอุบัติเหตุ บริเวณที่เกิดเหตุ ลักษณะการเกิด

อุบัติเหตุ สภาพอากาศ มาตรการความปลอดภัย เพศ จำนวนรถที่เกิดเหตุ และกำหนดจำนวนผู้เสียชีวิตเป็นค่าเป้าหมาย และลบแอททริบิวต์ที่ไม่ใช้งานออกไป

**3.2.3 การแปลงข้อมูล** โดยกำหนดค่าเป้าหมายเป็น 0 คือ เสียชีวิต และค่า 1 คือ ไม่เสียชีวิต และการจัดกลุ่มข้อมูล กำหนดข้อมูลเวลาในการเกิดอุบัติเหตุ เป็นช่วงข้อมูล เช่น ช่วงเวลา 05.00 น. – 10.00 น. กำหนดเป็นค่า 1 ช่วงเวลา 10.01 น. – 15.00 น. กำหนดเป็นค่า 2 ช่วงเวลา 15.01 น. – 19.00 น. กำหนดเป็นค่า 3 ช่วงเวลา 19.01 น. – 23.00 น. กำหนดเป็นค่า 4 ช่วงเวลา 23.01 น. เป็นต้นไป กำหนดเป็นค่า 5 เป็นต้น

### 3.3 การสร้างแบบจำลอง (Modeling)

อัลกอริทึมที่ใช้ในการสร้างแบบจำลองพยากรณ์ ประกอบด้วย 3 อัลกอริทึม ได้แก่ RF, SVM และ BNB แบ่งการดำเนินงานเป็น 2 ส่วน ดังนี้

**3.3.1 การกำหนดพารามิเตอร์ที่ใช้สำหรับการทดสอบอัลกอริทึมทั้ง 3 อัลกอริทึม** และทำการปรับค่าพารามิเตอร์ตามที่กำหนดไว้ในแต่ละอัลกอริทึม และเปรียบเทียบผลลัพธ์ แสดงดังตารางที่ 2

ตารางที่ 2: พารามิเตอร์สำหรับการทดสอบอัลกอริทึม

อัลกอริทึม	พารามิเตอร์
RF	n_estimators, criterion, random_state
SVM	c, kernel, gamma, random_state
BNB	alpha

**3.3.2 การทดสอบประสิทธิภาพอัลกอริทึม** ทำการแบ่งข้อมูลออกเป็น Training Dataset ร้อยละ 70 จำนวน 1,746 ระเบียบ และ Test Dataset ร้อยละ 30 จำนวน 749 ระเบียบ และทำซ้ำแต่ละรูปแบบจำนวน 5 ครั้ง โดยพิจารณาความถูกต้อง

### 3.4 การประเมินผล (Evaluation)

ทำการวัดประสิทธิภาพแบบจำลองการพยากรณ์ด้วยค่าความถูกต้อง

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

## 4. ผลการดำเนินงาน

ผลการดำเนินงานแสดงดังตารางที่ 2 ถึง 6 โดยมีรายละเอียดดังนี้

ตารางที่ 3: ตารางเปรียบเทียบจำนวนคลาส

คลาส "จำนวนผู้เสียชีวิต"	ก่อน SMOTE	หลัง SMOTE
0	2,178	2,178
1	317	2,178

จากตารางที่ 3 จำนวนคลาสนก่อนทำการสังเคราะห์ข้อมูลเพิ่มคลาส 0 มีจำนวน 2,178 ระเบียบ คลาส 1 จำนวน 317 ระเบียบ และหลังจากทำการสังเคราะห์ข้อมูลเพิ่มด้วยวิธี SMOTE คลาส 0 มีจำนวน 2,178 ระเบียบ คลาส 1 จำนวน 2,178 ระเบียบ

ตารางที่ 4: ผลลัพธ์การวัดค่าความถูกต้องของอัลกอริทึม RF

พารามิเตอร์			ก่อน SMOTE	หลัง SMOTE	เปลี่ยนแปลง
criterion	n_estimators	random state			
gini	10	4	89.45	88.59	-0.86
gini	20	5	89.05	88.59	-0.46
gini	30	5	89.18	88.67	-0.51
gini	40	5	89.05	89.90	0.85
gini	50	5	89.05	89.82	0.77
gini	60	5	89.05	89.90	0.85
gini	70	5	89.05	89.51	0.46
gini	80	5	89.05	89.74	0.69
gini	90	6	89.05	88.82	-0.23
gini	100	5	88.91	89.97	1.06
entropy	10	5	89.18	88.82	-0.36
entropy	20	3	89.05	89.67	0.62
entropy	30	6	89.18	88.82	-0.36
entropy	40	3	88.78	89.51	0.73
entropy	50	6	89.05	89.13	0.08
entropy	60	6	89.18	89.44	0.26
entropy	70	6	89.18	89.44	0.26
<b>entropy</b>	<b>80</b>	<b>5</b>	<b>89.05</b>	<b>90.28</b>	<b>1.23</b>
entropy	80	6	89.18	89.05	-0.13

จากตารางที่ 4 ค่าพารามิเตอร์ criterion = entropy n\_estimators = 80 และ randomstate = 5 ร่วมกับ SMOTE ให้ประสิทธิภาพดีที่สุด ค่าความถูกต้องร้อยละ 90.28 และการพยากรณ์ดีขึ้นร้อยละ 1.23

ตารางที่ 5: ผลลัพธ์การวัดค่าความถูกต้องของอัลกอริทึม SVM

พารามิเตอร์				ก่อน SMOTE	หลัง SMOTE	เปลี่ยนแปลง
gamma	kernel	c	random state			
scale	rbf	1	1	88.38	94.03	5.65
scale	poly	1	1	88.78	91.35	2.57
auto	rbf	100	1	88.65	93.57	4.92
<b>0.1</b>	<b>rbf</b>	<b>1</b>	<b>1</b>	88.38	<b>94.03</b>	<b>5.65</b>
0.2	rbf	1	1	88.78	93.80	5.02

พารามิเตอร์				ก่อน SMOTE	หลัง SMOTE	เปลี่ยนแปลง
gamma	kernel	c	random state			
scale	rbf	1	1	88.38	94.03	5.65
scale	poly	1	1	88.78	91.35	2.57
auto	rbf	100	1	88.65	93.57	4.92
<b>0.1</b>	<b>rbf</b>	<b>1</b>	<b>1</b>	88.38	<b>94.03</b>	<b>5.65</b>
<b>0.3</b>	<b>rbf</b>	<b>1</b>	<b>1</b>	88.78	<b>94.03</b>	<b>5.25</b>
0.4	rbf	1	1	88.65	93.80	5.15
0.5	rbf	10	1	88.65	93.95	5.3
0.6	rbf	10	1	88.65	93.72	5.07
0.7	rbf	1	1	88.38	93.72	5.34
0.8	rbf	10	1	88.38	93.72	5.34
0.9	rbf	1	1	88.38	93.64	5.26

จากตารางที่ 5 ค่าพารามิเตอร์ gamma = scale หรือ 0.1 หรือ 0.3 kernel = rbf c = 1 และ randomstate = 1 ร่วมกับ SMOTE ให้ประสิทธิภาพดีที่สุด ค่าความถูกต้องร้อยละ 94.03 และการพยากรณ์ดีขึ้นร้อยละ 5.25 - 5.65

ตารางที่ 6: ผลลัพธ์การวัดค่าความถูกต้องของอัลกอริทึม BNB

พารามิเตอร์	ก่อน SMOTE	หลัง SMOTE	เปลี่ยนแปลง
alpha			
0.01	87.18	88.59	1.41
<b>0.1</b>	<b>86.78</b>	<b>88.82</b>	<b>2.04</b>
1	85.58	87.68	2.1

จากตารางที่ 6 ค่าพารามิเตอร์ alpha = 0.1 ร่วมกับ SMOTE ให้ประสิทธิภาพดีที่สุด ค่าความถูกต้องร้อยละ 88.82 และการพยากรณ์ดีขึ้นร้อยละ 2.04

ผลการทดสอบประสิทธิภาพแบบจำลอง 3 แบบจำลอง แสดงดังตารางที่ 7

ตารางที่ 7: ตารางเปรียบเทียบผลการพยากรณ์ทั้ง 3 แบบจำลอง

อัลกอริทึม	ก่อน SMOTE	หลัง SMOTE	เปลี่ยนแปลง
RF	89.45	90.28	0.83
<b>SVM</b>	88.78	<b>94.03</b>	<b>5.25</b>
BNB	86.78	88.82	2.04

จากตารางที่ 7 พบว่าการพยากรณ์การเสียชีวิตจากอุบัติเหตุด้วยอัลกอริทึม RF ก่อนทำการทำการแก้ปัญหาความไม่สมดุลของข้อมูล ค่าความถูกต้องร้อยละ 89.45 และหลังแก้ปัญหาความไม่สมดุลของข้อมูล ค่าความถูกต้องร้อยละ 90.28 ส่งผลให้ประสิทธิภาพดีขึ้นร้อยละ 0.83 อัลกอริทึม SVM ก่อนทำการทำการแก้ปัญหาความไม่สมดุลของข้อมูล ค่าความถูกต้องร้อยละ 88.78 และหลังแก้ปัญหาความไม่สมดุลของข้อมูล ค่าความถูกต้องร้อยละ 94.03 ส่งผลให้ประสิทธิภาพดีขึ้นร้อยละ 5.25 และอัลกอริทึม BNB ก่อนทำการแก้ปัญหา

ความไม่สมดุลของข้อมูล ค่าความถูกต้องร้อยละ 86.78 และ  
หลังแก้ปัญหาความไม่สมดุลของข้อมูล ค่าความถูกต้อง  
ร้อยละ 88.82 ส่งผลให้ประสิทธิภาพดีขึ้นร้อยละ 2.04

## 5. สรุป

งานวิจัยฉบับนี้ผู้วิจัยได้นำเสนอการเปรียบเทียบ  
ประสิทธิภาพแบบจำลองการพยากรณ์การเสียชีวิตจาก  
อุบัติเหตุจากชุดข้อมูลอุบัติเหตุบนโครงข่ายถนนของ  
กระทรวงคมนาคม จำนวน 11 แอททริบิวต์ 2,495 ระเบียบ  
โดยเลือกอัลกอริทึมที่ได้รับความนิยม จำนวน 3 อัลกอริทึม  
ได้แก่ อัลกอริทึม RF, SVM และ BNB ผลการทดสอบ  
ประสิทธิภาพของแบบจำลองด้วยอัลกอริทึม SVM ร่วมกับ  
การแก้ปัญหาความไม่สมดุลของข้อมูล มีประสิทธิภาพ  
การพยากรณ์การเสียชีวิตจากอุบัติเหตุสูงที่สุดค่าความถูกต้อง  
ร้อยละ 94.03 เห็นได้ว่า การแก้ปัญหาความไม่สมดุล  
ของข้อมูลด้วยวิธีการสังเคราะห์ข้อมูลเพิ่มสามารถเพิ่ม  
ประสิทธิภาพการพยากรณ์ได้

## เอกสารอ้างอิง

- [1] World Health Organization. (2021). [online]. Road Traffic Injuries. [cited 13 June. 2021]. Available from : URL : <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
- [2] ฉบับล่าสุด! รายงานอุบัติเหตุถนนไทย คค. ชนทางตรงมากที่สุด 1.3 หมื่นครั้ง – ชายตายมากกว่าหญิง. (2563). [ออนไลน์]. [สืบค้นวันที่ 13 กรกฎาคม 2563]. จาก <https://www.isranews.org/article/isranews/93883-investigative742.html>
- [3] กัทธิดา สุวรรณโค และคณะ, “แบบจำลองการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศกาลปีใหม่ด้วยการทำเหมืองข้อมูล” *วารสารวิทยาการและเทคโนโลยีสารสนเทศ* ปีที่ 7 ฉบับที่ 2 พ.ศ. 2560 หน้า 10-19.
- [4] S. Ramya and et al, “Accident Severity Prediction Using Data Mining Methods” *International Journal of Scientific Research in Computer Science. on Engineering and Information Technology*. vol. 5, no. 2, pp. 528–536, 2019.
- [5] A. Nikhil Jha and et al, “A Performance Analysis of Prediction Techniques for Impacting Vehicles in Hit-and-Run Road Accidents” *Accident Analysis and Prevention*, vol.157, pp. 106-164, 2021.
- [6] T. Whasphutthisit and W. Jitsakul, “Comparison of Prediction Models for Road Deaths On Road Network” *In 14th International Conference on Knowledge and Smart Technology (KST)*, Thailand, pp. 137-142, 2022.
- [7] พัชรียา ทองพูล และคณะ, “การเปรียบเทียบประสิทธิภาพในการทำนายผลการปรับความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล”, *วารสารวิทยาศาสตร์และเทคโนโลยี* ปีที่ 8 ฉบับที่ 6 พ.ศ. 2562 หน้า 565-584.
- [8] The Art of Data Science for the Beginners using python and Azure ML. (2021). [online]. บทที่ 5 การจำแนก (Classification). [cited 22 December 2021]. Available from : URL : <https://csit.nu.ac.th/kraisak/ds/ds/chapter05/Chapter05.pdf>
- [9] SNRU Knowledge. (2018). [online]. ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM). [cited 25 December 2021]. Available from : URL : <https://knowledge.snru.ac.th/ซัพพอร์ตเวกเตอร์แมชชีน>
- [10] F. Nuraeni and et al, “Student Study Timeline Prediction Model Using Naïve Bayes Based Forward Selection Feature” *In 2021 International Conference on ICT for Smart Society (ICISS)*, Indonesia, 2021.
- [11] Automation Software (AutoSoft). (2021). [online]. เหมืองข้อมูล (data mining). [cited 19 December 2021]. Available from : URL : <https://www.autosoft.in.th/data-platform/เหมืองข้อมูล-data-mining>
- [12] A. -Z. S. Bin Habib and T. Tasnim, “An Ensemble Hard Voting Model for Cardiovascular Disease Prediction” *In 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, 2020.
- [13] A. Alsaeedi and M. Zubair Khan, “Software Defect Prediction Using Supervised Machine Learning and Ensemble Techniques: A Comparative Study” *Journal of Software Engineering and Applications*. vol.12, pp. 85-100, 2019.
- [14] พัชรียา ทองพูล และคณะ, “การเปรียบเทียบประสิทธิภาพในการทำนายผลการปรับความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล” *วารสารวิทยาศาสตร์และเทคโนโลยี*. ปีที่ 8 ฉบับที่ 6 หน้า 565-584.

# การพยากรณ์ความเสี่ยงการทุจริตของรายการบัตรเครดิตอิเล็กทรอนิกส์ Forecasting the Fraud Risk of Electronic Card Transactions

ศิริพงษ์ ประดิษฐ์ (Sirepong padit)<sup>1</sup> และวัชรวิวัฒน์ จิตต์สกุล(Watchareewan Jitsakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021856047@email.kmutnb.ac.th, <sup>2</sup>watchareewan.j@itd.kmutnb.ac.th

## บทคัดย่อ

งานวิจัยนี้ทำการศึกษาปัญหาเกี่ยวกับการทุจริตที่เกิดขึ้นกับบัตรเครดิตอิเล็กทรอนิกส์ โดยจะทำการศึกษาพฤติกรรมจากรายการใช้บัตรเครดิตอิเล็กทรอนิกส์ในอดีตที่ส่งผลต่อการทุจริต และหาแนวโน้มความเสี่ยงที่จะเกิดการทุจริตที่อาจเกิดขึ้นในอนาคต โดยเปรียบเทียบอัลกอริทึมจำนวน 3 อัลกอริทึม ได้แก่ อัลกอริทึมป่าสุ่ม อัลกอริทึมเพื่อนบ้านใกล้ที่สุด และอัลกอริทึมกาเดียนบูท ข้อมูลที่ใช้คือ รายการใช้บัตรเครดิตอิเล็กทรอนิกส์มีจำนวนทั้งสิ้น 110,233 รายการ ทำการคัดเลือกข้อมูลให้เหลือเพียง 11,000 รายการ ทำการแบ่งเป็น 2 ส่วน คือ ข้อมูลสำหรับสร้างแบบจำลองจำนวน 10,000 รายการ และข้อมูลสำหรับทดสอบแบบจำลองจำนวน 1,000 รายการ เมื่อได้แบบจำลองที่ดีที่สุดแล้วได้ทำการทดสอบแบบจำลองเพื่อเปรียบเทียบประสิทธิภาพสำหรับการพยากรณ์ความเสี่ยงที่จะเกิดการทุจริต ผลการวิจัยพบว่า อัลกอริทึมกาเดียนบูทให้ประสิทธิภาพสูงสุด โดยมีค่าความถูกต้องร้อยละ 80.83 ค่าความแม่นยำเท่ากับ 0.81 ค่าระลอกเท่ากับ 0.81 รองลงมาคือ อัลกอริทึมป่าสุ่ม โดยมีค่าความถูกต้องเท่ากับร้อยละ 74.49 ค่าความแม่นยำเท่ากับ 0.76 ค่าระลอกเท่ากับ 0.74 และอัลกอริทึมเพื่อนบ้านใกล้ที่สุด โดยมีค่าความถูกต้องเท่ากับร้อยละ 54.88 ค่าความแม่นยำเท่ากับ 0.59 และค่าระลอกเท่ากับ 0.55 ตามลำดับ

**คำสำคัญ:** รายการทุจริต อัลกอริทึมป่าสุ่ม

อัลกอริทึมเพื่อนบ้านใกล้ที่สุด

อัลกอริทึมกาเดียนบูท

## Abstract

This research investigates the problem of fraud occurring with electronic cards. It will study the behavior of electronic card transactions in the past that resulted in fraud and find out the potential risk of fraud that may occur in the future. by using three algorithms, which are Random Forest, K-Nearest Neighbor and Gradient Boosted Trees, make a model of 110,233 data, only 11,000 entries were selected, then the data that has been divided into 2 parts is 10,000 modeling data and data to test 1000 models, Once the best model was obtained the model was tested to compare performance for forecasting fraud risk. The findings reveal that the Gradient Boosted Trees algorithm is more effective. It has 80.83%, of accuracy precision is 0.81 and recall is 0.81 Next is the Random Forest algorithm. with an accuracy of 74.49%, an accuracy of 0.76 and a recall value of 0.74 K-Nearest Neighbor algorithms. The accuracy was 54.88%, the accuracy was 0.59 and the recall was 0.55, respectively.

**Keywords:** Transaction Fraud, Random Forest, K-Nearest Neighbor, Gradient Boosted Trees.

## 1. บทนำ

ปัจจุบันมีการใช้บัตรเครดิตอิเล็กทรอนิกส์เพื่อการชำระสินค้าหรือการบริการของคนไทยในแต่ละปีเพิ่มมากขึ้น เนื่องจากการชำระเงินด้วยบัตรเครดิตอิเล็กทรอนิกส์มีความสะดวกสบายมากมาย เช่น ช่วยให้ได้รับสินค้าก่อนแล้วสามารถผ่อนจ่ายในภายหลังตามอัตราดอกเบี้ยที่ได้ตกลง

ไว้ เป็นต้น นอกจากนี้ยังได้รับส่วนลดหรือสิทธิพิเศษในการซื้อสินค้า หรือบริการต่าง ๆ อีกด้วย อีกทั้งธนาคารต่าง ๆ มักหา กฤตยุทธ์ทางการตลาดมาใช้เพื่อดึงดูดให้ลูกค้าสนใจและใช้บัตรเครดิตอิเล็กทรอนิกส์ จึงทำให้บัตรเครดิตอิเล็กทรอนิกส์เป็นที่นิยมและมีผู้สนใจเพิ่มมากขึ้น การที่ธนาคาร (Bank) หรือผู้ประกอบการต่าง ๆ (Non-bank) ให้ความสนใจเข้ามาลงทุนกับธุรกิจบัตรเครดิตอิเล็กทรอนิกส์มากขึ้น เพราะเนื่องด้วยผลกำไรและการขยายตลาดในด้านอื่น ๆ ตามมานั้น ทำให้รายการการใช้บัตรเครดิตอิเล็กทรอนิกส์สูงขึ้นมาก แต่อย่างไรก็ตามก็ยังมีผู้ไม่หวังดีที่หาช่องทางทุจริตจากการใช้บัตรเครดิต ทำให้ธนาคาร หรือผู้ประกอบการต่าง ๆ หลายแห่งประสบปัญหาในการวิเคราะห์และติดตามการทุจริตหรือหาความผิดปกติของรายการใช้บัตรเครดิตที่มีความเสี่ยงเกิดการทุจริตจะเป็นเรื่องของการพยากรณ์ ซึ่งการพยากรณ์มีผู้วิจัยนำเสนออัลกอริทึมต่าง ๆ มาช่วยในการวิเคราะห์ข้อมูลไม่ว่าจะเป็นเรื่องของการเสี่ยงการเกิดทุจริต การเกิดอุบัติเหตุ การเจ็บป่วย หรือในเรื่องอื่น ๆ เช่น อัลกอริทึม Random Forest [1] อัลกอริทึม Artificial Neural Network อัลกอริทึม Random Forest และ อัลกอริทึม Support Vector Machine [2] อัลกอริทึม K-Nearest Neighbor [3] อัลกอริทึม Gradient Boosted [4] อัลกอริทึม Generalized Linear Model อัลกอริทึม Naive Bayes อัลกอริทึม Logistic Regression อัลกอริทึม Decision Tree อัลกอริทึม Random Forest และอัลกอริทึม Gradient Boosted Trees[5] เป็นต้น

จากปัญหา และงานวิจัยดังกล่าว ผู้วิจัยจึงมีแนวคิดในการนำอัลกอริทึมสำหรับการพยากรณ์มาทำการวิเคราะห์แนวโน้มความเสี่ยงที่จะเกิดการทุจริตบัตรเครดิตอิเล็กทรอนิกส์ และจัดทำรายงานเพื่อแสดงผลต่อไป

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้รวบรวมข้อมูลทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยแบ่งเป็น 2 หัวข้อหลัก ดังนี้

### 2.1 อัลกอริทึมการพยากรณ์ (Prediction Algorithm)

2.1.1 อัลกอริทึมป่าสุ่ม (Random Forest) เกิดจากการรวมกลุ่มกันของโครงสร้างต้นไม้โดยในการทำงานของ

อัลกอริทึม จะทำการจำแนกต้นไม้หลายต้น ซึ่งในต้นไม้แต่ละต้นจะมีการแบ่งเป็นคลาส โดยที่ต้นไม้แต่ละต้นจะถูกสร้างกลุ่มตัวอย่างที่แตกต่างกันจากกระบวนการของต้นไม้การตัดสินใจ และถูกสร้างขึ้นจนกลายเป็นป่า ซึ่งสามารถสรุปได้ว่าอัลกอริทึมป่าสุ่มเป็นอัลกอริทึมประเภทหนึ่งของอัลกอริทึมต้นไม้การตัดสินใจ [6]

2.1.2 อัลกอริทึมเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) เป็นวิธีการที่ใช้ในการจำแนกประเภทข้อมูลที่มีระยะห่าง 1 กลุ่ม (ใกล้ที่สุด) จะเรียกว่า “1NN (One Nearest Neighbor) ดังนั้น “k-NN” ค่า k จึงเป็นจำนวนของกลุ่มที่ต้องการเลือกเป็นกลุ่มเพื่อนบ้าน โดยควรกำหนดเป็นเลขคี่ [7]

2.1.3 อัลกอริทึมกาเดียนบูท (Gradient Boosted Trees) เป็นวิธีการพื้นฐานมาจาก Decision tree ซึ่งเป็นการปรับปรุงประสิทธิภาพของแบบจำลองให้มีค่าสูงขึ้น โดย,การสุ่มสร้าง Decision Tree หลายร้อยแบบจำลองและประเมินผลแต่ละแบบจำลองจนกว่าจะได้ Decision Tree ที่สมบูรณ์ และใช้วิธีการ Optimize Parameter เพื่อหาค่าพารามิเตอร์ที่ดีที่สุดของแต่ละแบบจำลองด้วยวิธีการ Evolutionary Optimization โดยมีพื้นฐานมาจาก Genetic Algorithms เป็นวิธีการหาค่าตอบจากการเลียนแบบพันธุกรรมและกระบวนการวิวัฒนาการของสิ่งมีชีวิต โดยมีหลักการในการหาปัจจัยที่ทำให้คำตอบดีขึ้นและทำการไขว้คำตอบเพื่อนำไปสู่ชุดปัจจัยที่ทำให้คำตอบดีขึ้นกว่าเดิม [8]

### 2.2 งานวิจัยที่เกี่ยวข้อง

Monika and et al. [6] การจำแนกประเภทข้อมูลการทุจริตบัตรเครดิตด้วยอัลกอริทึม Random Forest ข้อมูลที่ใช้เป็นเป็นชุดข้อมูลการใช้บัตรเครดิต โดยการแยกคุณลักษณะเป็นข้อมูลเพื่อศึกษา และข้อมูลเพื่อทดสอบต่อมาใช้ชุดเครื่องมือบนภาษา Python เพื่อทำการแยกประเภท สุดท้ายแสดงผลที่ Tableau ผลการวิจัยพบว่าความแม่นยำในการจำแนกเท่ากับ ร้อยละ 99.9

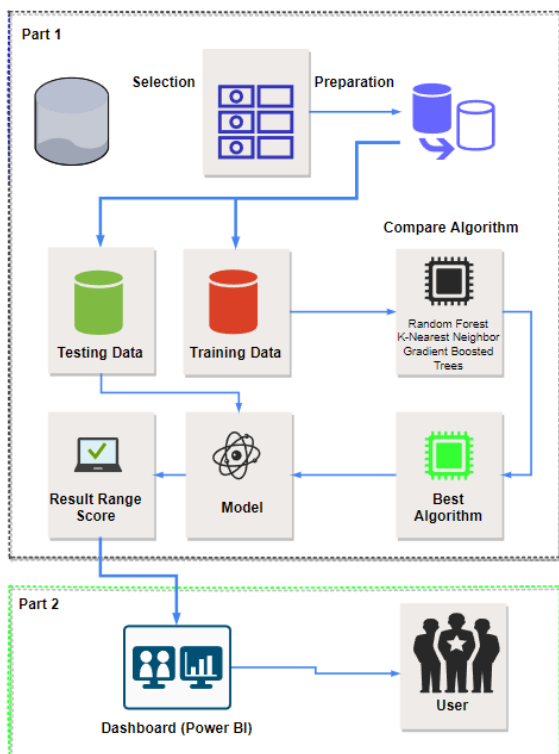
Devi and et al. [7] ทำการจำแนกการป้องกันการทุจริตเครดิตเพื่อพิสูจน์ว่าผู้ใช้บัตรเครดิตเป็นผู้ถือครองบัตรจริง ๆ โดยใช้อัลกอริทึม Random Forest และการใช้

โปรแกรม Anaconda โดยแบ่งข้อมูล 2 ส่วน คือ ข้อมูลสำหรับฝึกสอน และข้อมูลสำหรับทดสอบโดยทำ Cross Validation แบบ Hold-Out พบว่า ประสิทธิภาพของอัลกอริทึม Random Forest นั้นสามารถเพิ่มความแม่นยำในการป้องกันการทำทุจริตเครดิต ผลลัพธ์ที่ได้จากการประมวลผลชุดข้อมูลมีความแม่นยำประมาณร้อยละ 90 นำค่าที่ได้มาพัฒนาการแสดงผลในรูปแบบแดชบอร์ด

บริษัท จระณะสมบูรณ์ [8] ทำการทำนายการซื้อซ้ำของผู้ซื้อโดยใช้เทคนิคการเรียนรู้ของเครื่องจักร เพื่อจำแนกลูกค้าที่จะกลับมาซื้อซ้ำในอนาคต โดยใช้เทคนิคการจำแนกข้อมูล และการวิเคราะห์การถดถอย ได้แก่ Random Forest Classifier, Random Forest Regressor, XGBoost และ Gradient Boost ร่วมกับเทคนิค Leave One Out ผลการวิจัยพบว่า Random Forest Regressor ใช้ร่วมกับ Leave One Offer Out ให้ผลลัพธ์สูงที่สุดเมื่อเทียบกับอัลกอริทึมอื่นๆ

### 3. วิธีการดำเนินการวิจัย

การพยากรณ์ความเสี่ยงการทุจริตของรายการบัตรอิเล็กทรอนิกส์ ผู้วิจัยมีกรอบแนวคิดแสดงดังภาพที่ 1



ภาพที่ 1: กรอบแนวคิด

จากภาพที่ 1 กรอบแนวความคิด สามารถแบ่งขั้นตอนการดำเนินการวิจัยเป็น 2 ส่วน ดังนี้

#### 3.1 ขั้นตอนการคัดเลือกและทดสอบประสิทธิภาพอัลกอริทึม

นำข้อมูลที่ได้จากการเตรียมมาทำการคัดเลือกอัลกอริทึม โดยทำการเปรียบเทียบอัลกอริทึม 3 อัลกอริทึม ได้แก่ Random Forest K-Nearest Neighbor และ Gradient Boosted Trees โดยพิจารณาค่าความถูกต้อง (Accuracy: ACC) ค่าความแม่นยำ (Precision: P) และค่าระลึก (Recall: R) เพื่อคัดเลือกหาอัลกอริทึมที่ดีที่สุดเพื่อไปสร้างเป็นแบบจำลองในการจำแนกรายการเสี่ยงเกิดการทุจริต สำหรับการประเมินจะใช้วิธี Supplied Test Set คือ การแบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลสำหรับการเรียนรู้ (Training Set) 70% และข้อมูลสำหรับทดสอบ (Testing Set) 30% ซึ่ง Training Set จะกำหนดคะแนนความเสี่ยง ดังนี้ Class A: คะแนน 0-30 Class B: คะแนน 31-50 Class C: คะแนน 51-70 Class D: คะแนน 71-100

สำหรับข้อมูลที่นำมาใช้ในการวิจัย เป็นข้อมูลรายการใช้บัตรอิเล็กทรอนิกส์ ประกอบด้วยคุณลักษณะทั้งสิ้น 35 คุณลักษณะ จำนวน 110,233 รายการ รายละเอียดคุณลักษณะ แสดงดังตารางที่ 1

ตารางที่ 1: ตารางแสดงคุณลักษณะของข้อมูล

ลำดับ	คุณลักษณะ
1	Account Number
2	Transaction Amount (U.S. \$)
3	Acquirer Transaction Amount
4	Acquirer Currency Code
5	Authorization Timestamp (Asia/Bangkok)
6	Advanced Authorization Risk Score
7	Issuer BIN
8	Acquirer BIN
9	Merchant Category Code (MCC)
10	Merchant Location
11	Merchant Name
12	Authorization Response Code
13	STIP Code
14	POS Entry Mode
15	Compromised Account Risk Condition Code
16	Compromised Event Reference ID
17	Status
18	Status Timestamp (Asia/Bangkok)



ตารางที่ 1: (ต่อ)

ลำดับ	คุณลักษณะ
19	MOTO/ECI/Recurring
20	Visa Transaction ID
21	Terminal Capability
22	Issuer Currency Code
23	Issuer Amount
24	Merchant ID
25	Terminal ID
26	Terminal Type
27	Authorization Code
28	Stated By User ID
29	Stated By Last Name
30	Card Acceptor Country Code
31	Acquirer Country Code
32	CC Rule Name
33	RTD Rule Name
34	Alert Timestamp (Asia/Bangkok)
35	STIP Reason Code

จากนั้นทำการคัดเลือกคุณลักษณะโดยอ้างอิงจากของงานวิจัยที่เกี่ยวข้องกับการทุจริต สามารถคัดเลือกได้ 6 คุณลักษณะ และทำการคัดเลือกข้อมูลเพื่อใช้ในงานวิจัยจำนวน 11,000 รายการ แบ่งเป็น 2 ส่วน คือ ข้อมูลสำหรับสร้างแบบจำลองจำนวน 10,000 รายการ และข้อมูลสำหรับทดสอบแบบจำลองจำนวน 1,000 รายการ รายละเอียดคุณลักษณะที่ผ่านการคัดเลือก แสดงดังตารางที่ 2

ตารางที่ 2: ตารางแสดงคุณลักษณะของข้อมูลที่ผ่านการคัดเลือก

ลำดับ	คุณลักษณะ
1	Transaction Amount
2	Authorization Risk Score
3	Merchant Category Code (MCC)
4	Merchant Name
5	Card Acceptor Country Code
6	Acquirer Country Code

### 3.2 การออกแบบรายงาน

เมื่อสร้างแบบจำลองในการจำแนกรายการเสี่ยงเกิดการทุจริตเรียบร้อยแล้ว ในส่วนนี้จะเป็นการนำข้อมูลการใช้งานบัตรอิเล็กทรอนิกส์ มาวิเคราะห์ผ่านแบบจำลองเพื่อได้ผลลัพธ์ของการจำแนกออกมา นำผลลัพธ์ที่ได้มาแสดงผลผ่าน Power BI โดยแสดงผลข้อมูลในรูปแบบ Dashboard แบบภาพรวม

## 4. ผลการดำเนินงาน

ผลการดำเนินงานสามารถแบ่งเป็นหัวข้อได้ ดังนี้

### 4.1 ผลการเปรียบเทียบอัลกอริทึม

การสร้างแบบจำลองการพยากรณ์ สามารถสรุปผลการเปรียบเทียบอัลกอริทึมแสดงดังตารางที่ 3 ถึง 5

ตารางที่ 3: ผลการทดสอบอัลกอริทึม Random Forest

No.	N_Estimators	Criterion	Voting	Acc	P	R
1	80	Information_gain	majority vote	73.67	0.75	0.74
2	70	Information_gain	majority vote	73.57	0.75	0.74
3	60	Information_gain	majority vote	73.61	0.75	0.74
<b>4</b>	<b>100</b>	<b>Information_gain</b>	<b>confidence vote</b>	<b>74.49</b>	<b>0.76</b>	<b>0.74</b>
5	90	Information_gain	confidence vote	74.47	0.76	0.74
6	80	Information_gain	confidence vote	74.26	0.76	0.74
7	70	Information_gain	confidence vote	74.26	0.76	0.74
8	60	Information_gain	confidence vote	74.40	0.76	0.74

จากตารางที่ 3 ค่าพารามิเตอร์ N\_Estimators เท่ากับ 100 Criterion เท่ากับ Information\_gain และ Voting เท่ากับ confidence vote ให้ผลลัพธ์ที่ดีที่สุด ค่าความถูกต้องเท่ากับ ร้อยละ 74.49 ค่าความแม่นยำเท่ากับ 0.76 และค่าระลอกเท่ากับ 0.74 ตามลำดับ

ตารางที่ 4: ผลการทดสอบอัลกอริทึม K-Nearest Neighbor

No.	N_Neighbors	Measure Types	Mixed Measure	Acc	P	R
1	1	mixed measure	Mixed Euclidean Distance	52.83	0.57	0.53
2	2	mixed measure	Mixed Euclidean Distance	52.83	0.57	0.53
3	3	mixed measure	Mixed Euclidean Distance	54.11	0.59	0.54
4	4	mixed measure	Mixed Euclidean Distance	53.78	0.58	0.54
<b>5</b>	<b>5</b>	<b>mixed measure</b>	<b>Mixed Euclidean Distance</b>	<b>54.88</b>	<b>0.59</b>	<b>0.55</b>
6	6	mixed measure	Mixed Euclidean Distance	54.64	0.58	0.55

จากตารางที่ 4 ค่าพารามิเตอร์ N\_Neighbors เท่ากับ 5 Measure Types เท่ากับ mixed measure และ Mixed Measure เท่ากับ MixedEuclideanDistance ให้ผลลัพธ์ที่

ดีที่สุด ค่าความถูกต้องเท่ากับ ร้อยละ 54.88 ค่าความแม่นยำเท่ากับ 0.59 และค่าระลอกเท่ากับ 0.55 ตามลำดับ

ตารางที่ 5: ผลการทดสอบอัลกอริทึม Gradient Boosted Trees

No.	Number of Trees	Maximal Depth	Acc	P	R
1	50	10	79.76	0.80	0.80
2	60	10	79.78	0.80	0.80
3	70	10	79.90	0.80	0.80
4	80	10	79.97	0.81	0.80
5	90	10	80.08	0.81	0.80
6	100	10	80.12	0.81	0.80
7	50	15	80.43	0.81	0.80
8	60	15	80.40	0.81	0.80
9	70	15	80.60	0.81	0.81
10	80	15	80.60	0.81	0.81
11	90	15	80.76	0.81	0.81
<b>12</b>	<b>100</b>	<b>15</b>	<b>80.83</b>	<b>0.81</b>	<b>0.81</b>

จากตารางที่ 5 ค่าพารามิเตอร์ Number of Trees เท่ากับ 100 Maximal Depth และเท่ากับ 15 ให้ผลลัพธ์ที่ดีที่สุด ค่าความถูกต้องเท่ากับ ร้อยละ 80.83 ค่าความแม่นยำเท่ากับ 0.81 และค่าระลอกเท่ากับ 0.81 ตามลำดับ

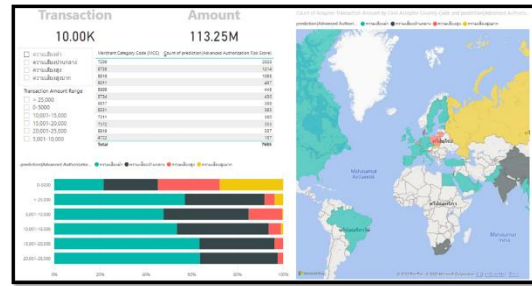
ตารางที่ 6: แสดงการเปรียบเทียบอัลกอริทึม

No.	Algorithm	Performance		
		Acc	P	R
1	Random Forest	74.49	0.76	0.74
2	K-Nearest Neighbor	54.88	0.59	0.55
3	<b>Gradient Boosted Trees</b>	<b>80.83</b>	<b>0.81</b>	<b>0.81</b>

จากตารางที่ 6 สรุปได้ว่า อัลกอริทึม Gradient Boosted Trees มีประสิทธิภาพการพยากรณ์สูงที่สุด โดยมีค่าความถูกต้องร้อยละ 80.83 ค่าความแม่นยำเท่ากับ 0.81 และค่าระลอกเท่ากับ 0.81

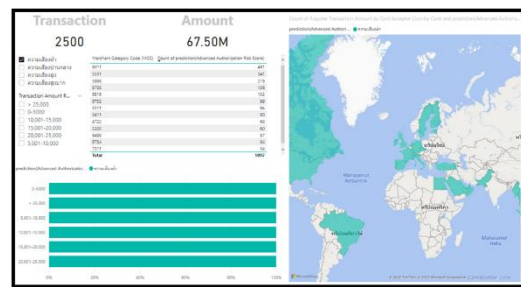
4.2 การออกแบบรายงาน Power BI

ภาพแสดงผลรวมการพยากรณ์ความเสี่ยงการทุจริตของรายการบัตรเครดิตทรอนิกส์จำนวน 10,000 รายการ เพื่อนำไปใช้ในการ Training และทำการกำหนดคะแนนความเสี่ยง ดังนี้ Class A: ความเสี่ยงต่ำ Class B: ความเสี่ยงปานกลาง Class C: ความเสี่ยงสูง Class D: ความเสี่ยงสูงมาก แสดงดังภาพที่ 2

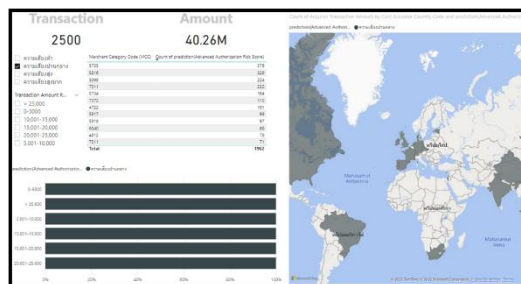


ภาพที่ 2: ภาพรวม Training Data 10,000 Transaction

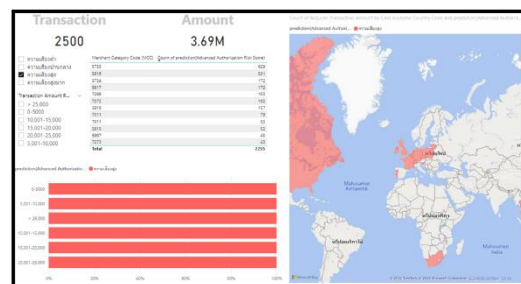
จาก Training Data 10,000 Transaction ผู้วิจัยได้ทำการแบ่งแต่ละ Class ให้เท่ากันจำนวน Class ละ 2,500 รายการ เพื่อแสดงให้เห็นว่าในแต่ละ Class รายละเอียดความเสี่ยงอยู่ในโซนประเทศใด มูลค่าเท่าไร และเป็นสินค้าประเภทอะไร แสดงดังภาพที่ 3 ถึง 6



ภาพที่ 3: Class A ความเสี่ยงต่ำ



ภาพที่ 4: Class B ความเสี่ยงปานกลาง

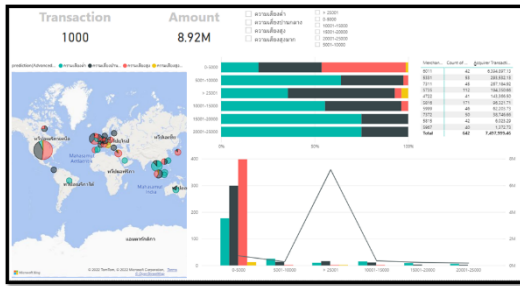


ภาพที่ 5: Class C ความเสี่ยงสูง



ภาพที่ 6: Class D ความเสี่ยงสูงมาก

จากนั้นผู้วิจัยได้นำ Testing Data 1,000 Transaction ทำการ Predict เพื่อแสดงให้เห็นว่า Transaction ดังกล่าว จะมีความเสี่ยงการทุจริตของรายการบัตรเครดิตอิเล็กทรอนิกส์อยู่ใน Class ความเสี่ยงระดับใด แสดงดังภาพที่ 7



ภาพที่ 7: Predict Testing Data 1,000 Transaction

#### 4. สรุป

จากการเปรียบเทียบอัลกอริทึมสำหรับการพยากรณ์ จำนวน 3 อัลกอริทึม คือ 1) อัลกอริทึม Random Forest 2) อัลกอริทึม K-Nearest Neighbor และ 3) อัลกอริทึม Gradient Boosted Trees โดยการสร้างแบบจำลองจาก ข้อมูลการใช้บัตรเครดิตอิเล็กทรอนิกส์ที่เป็นรายการใช้แบบปกติ และแบบที่เกิดการทุจริต ตั้งแต่ 1 พฤษภาคม 2563 – 31 ธันวาคม 2563 ซึ่งเป็นข้อมูลที่ผู้วิจัยได้ทำความสะอาด ข้อมูล (Data Cleansing) เพื่อจะได้อัลกอริทึมที่เหมาะสม สำหรับพยากรณ์

จากการทดสอบ และเปรียบเทียบประสิทธิภาพ อัลกอริทึม สรุปผลได้ว่า อัลกอริทึม Gradient Boosted Trees ให้ประสิทธิภาพการพยากรณ์ดีที่สุดโดยมีค่าความ ถูกต้องร้อยละ 80.83 ค่าความแม่นยำเท่ากับ 0.81 และ ค่าระลอกเท่ากับ 0.81 ตามลำดับ จากนั้นนำข้อมูลการ

พยากรณ์แสดงผลผ่าน Power BI เพื่อนำไปใช้ประโยชน์ ในการช่วยตรวจจบบัตรการทุจริตบัตรเครดิตอิเล็กทรอนิกส์ได้ อย่างมีประสิทธิภาพยิ่งขึ้นต่อไป

#### เอกสารอ้างอิง

- [1] V. Jonnalagadda, P. Gupta, and E. Sen, “Credit Card Fraud Detection using Random Forest Algorithm,” *International Journal of Advance Research, Ideas and Innovations in Technology*, Vol. 5, Issue. 2, pp. 1797-1801, 2019.
- [2] ศตวรรษ อ่อนมิ่ง และ วัชรวิวรรณ จิตต์สกุล, “การเปรียบเทียบ อัลกอริทึมเหมืองข้อมูลสำหรับพยากรณ์การทำทุจริต” การ ประชุมวิชาการระดับชาติด้านวิทยาศาสตร์และเทคโนโลยี ครั้งที่ 4, 2563.
- [3] พัชรียา ทองพูล และคณะ, “การเปรียบเทียบประสิทธิภาพใน การทำนายผลการปรับความไม่สมดุลของข้อมูลในการ จำแนกด้วยอัลกอริทึมการทำเหมืองข้อมูล” *วารสาร วิทยาศาสตร์และเทคโนโลยี ปีที่ 8 ฉบับที่ 6 พฤศจิกายน – ธันวาคม 2562* หน้า 565-584.
- [4] สกฤตน์ ขุนสูงเนิน, “การใช้การวิเคราะห์เชิงทำนายสำหรับการ การระบุสถานะการจำหน่ายและการรอดชีวิตในผู้ป่วยภาวะ ดิabetic และผู้ป่วยภาวะช็อกจากเหตุพิษติดเชื้อบนพื้นฐานของ ปัจจัยการผ่าตัดและการทำหัตถการทางการแพทย์” *วารสาร วิจัยมหาวิทยาลัยขอนแก่น ปีที่ 19 ฉบับที่ 2 เม.ย.-มิ.ย. 2019* หน้า 117-130.
- [5] วิไลลักษณ์ ศรีพีช, “การวิเคราะห์ข้อมูลผู้บาดเจ็บและเสียชีวิต จากอุบัติเหตุในช่วงเทศกาลสงกรานต์โดยใช้เทคนิคการทำ เหมืองข้อมูล” *วารสารวิจัยและพัฒนา สาขาวิชาวิทยาศาสตร์ และเทคโนโลยี วิทยาลัยสงฆ์ ในพระบรมราชูปถัมภ์ ปีที่ 14 ฉบับที่ 1 ม.ค.-เม.ย. 2562* หน้า 11-20.
- [6] S. Monika, K. Venkataramanamma, P. Pritto, and M. Usha, “Credit Card Fraud Detection using Random Forest Algorithm,” *International Journal of Research in Engineering, Science and Management*, Vol. 2, Issue 3, pp. 131-133, 2003.
- [7] D. Meenakshi, et al, “Credit Card Fraud Detection using Random Forest.” *International Research Journal of Engineering and Technology (IRJET)*, Vol. 6, Issue 3, pp. 6662-6666, 2019.
- [8] ธนัท จระสมบูรณ์, “การทำนายการซื้อซ้ำของผู้ซื้อโดยใช้ เทคนิคการเรียนรู้ของเครื่องจักร” *ปริญญาณิพนธ์วิทยาศาสตร มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะ วิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ*, 2561.

# การพัฒนาระบบธุรกิจอัจฉริยะ เพื่อการวิเคราะห์ข้อมูลจากระบบการรับแจ้ง เหตุขัดข้องโดยใช้เทคนิคเหมืองข้อมูล

## The Development of Business Intelligence for Analyzing Data from the Fault Reporting System using Data Mining Techniques

จิรรัฐ ศรีโชค (Jirarat Srichok)<sup>1</sup> และมณฑิธร รัตนศิริวงษ์วุฒิ (Montean Rattanasirivongwut)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

S6307021856063@email.kmutnb.ac.th<sup>1</sup>, montean.r@itd.kmutnb.ac.th<sup>2</sup>

### บทคัดย่อ

งานวิจัยนี้มีแนวคิดในการนำข้อมูลจากระบบรับแจ้งเหตุขัดข้องเพื่อนำมาสร้างแบบจำลองในการวิเคราะห์ประสิทธิภาพของการทำงานที่เป็นไปตามข้อตกลงระดับในการให้บริการระหว่างผู้ให้บริการ และลูกค้า หรือ SLA (Service Level Agreement) โดยเปรียบเทียบระหว่างอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน อัลกอริทึมเพื่อนบ้านใกล้ที่สุด อัลกอริทึมนาอิวเบย์ และเอนเซมเบิลเมธอด ข้อมูลที่ใช้คือ รายการแจ้งเหตุขัดข้องจำนวนทั้งสิ้น 306,075 รายการ โดยทำการข้อมูลให้เหลือ 11,000 รายการ โดยแบ่งเป็นข้อมูลสำหรับสร้างแบบจำลอง 10,000 รายการ และข้อมูลสำหรับทดสอบแบบจำลอง 1,000 รายการ เมื่อทดลองจนพบแบบจำลองที่ดีที่สุดแล้ว จึงทำการทดสอบแบบจำลองเพื่อเทียบประสิทธิภาพในการวิเคราะห์ข้อมูล การทดสอบประสิทธิภาพครั้งนี้พบว่าอัลกอริทึมที่ให้ประสิทธิภาพสูงที่สุดคือ อัลกอริทึมเพื่อนบ้านใกล้ที่สุดโดยมีค่าความถูกต้อง ร้อยละ 99.28 ค่าความแม่นยำเท่ากับ 0.99 ค่าความระลึกเท่ากับ 0.99 ตามลำดับ จากนั้นจึงนำข้อมูลที่ได้จากการวิเคราะห์มาพัฒนาระบบรายงานโดยใช้ Power BI ในการพัฒนาให้ได้รายงานที่สามารถเข้าใจได้ง่าย และสามารถนำมาช่วยในการตัดสินใจของผู้บริหารต่อไปได้ในอนาคต

**คำสำคัญ:** คุณภาพการให้บริการ อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน อัลกอริทึมเพื่อนบ้านใกล้ที่สุด อัลกอริทึมนาอิวเบย์ เอนเซมเบิลเมธอด

### Abstract

*In this research, the researcher has the idea to use the data information from the Fault Report system to create a model to analyze performance in accordance with the service level agreement between the service provider and the customer or SLA (Service Level Agreement). from various factors of information received from the Fault Report System It compares The Support Vector Machine algorithm, The K-Nearest Neighbor algorithm, The Naïve Bays algorithm, Ensemble Method. The information used is The total number of fault reports was 306,075. Only 11,000 of the data were selected, divided into two parts 10,000 modeling data and 1,000 model testing data. Therefore, the model was tested to compare the efficiency of the data analysis. This performance test revealed that the most efficient algorithm was Nearest neighbor algorithm with an accuracy of 99.28%, precision is 0.99, and recall is 0.99, respectively. Then use the data from the analysis to develop a reporting system using Power BI that is easy to understand. and can be used to assist in the decision of the management in the future*

**Keywords:** Service quality, Support Vector Machine, K-Nearest Neighbor, Naïve Bays, Ensemble Method.

## 1. บทนำ

ในปัจจุบันแทบจะไม่สามารถปฏิเสธได้ว่าการใช้อินเทอร์เน็ตเป็นส่วนหนึ่งในชีวิตของคนไทยและในแต่ละปีจะมียอดการใช้งานเพิ่มขึ้นมากเป็นเท่าตัว การให้บริการด้านอินเทอร์เน็ตความเร็วสูงผ่าน เทคโนโลยี FTTx ให้แก่ประชาชนที่มีความต้องการด้านการสื่อสาร และอินเทอร์เน็ตที่มีความเร็วสูงขึ้น เพื่อใช้กับ ธุรกิจ ความบันเทิง การศึกษา การแพทย์ และอื่น ๆ [1] ข้อมูลจาก We Are Social และ Hootsuit ได้เปิดเผยข้อมูล ที่มีความสำคัญสำหรับต่อตลาดธุรกิจการให้บริการออนไลน์ ซึ่งชี้ให้เห็นถึงพฤติกรรมการใช้งานอินเทอร์เน็ตของคนไทยที่ติดอันดับโลก โดยผลของการจัดอันดับการใช้งานอินเทอร์เน็ตของโลกปี พ.ศ. 2565 ประชากรไทยสามารถเข้าถึงอินเทอร์เน็ตคิดเป็น 77% ของประชากร การเข้าถึงอินเทอร์เน็ตของคนไทยคิดค่าเฉลี่ยทั่วโลกที่ 62.5% ขณะที่มียี่สิบหลายประเทศสามารถผลักดันให้การเข้าถึงอินเทอร์เน็ตทั่วประเทศอยู่ในระดับที่สูงกว่า 90% จึงแสดงให้เห็นถึงโอกาสในการพัฒนาโครงข่ายเพื่อให้คนไทยสามารถเข้าถึงอินเทอร์เน็ตได้ในระดับมากกว่า 90% จากรายงาน Data Report Digital Stat 2022 ซึ่งชี้ให้เห็นถึงการใช้งานอินเทอร์เน็ตของคนไทยที่มีระยะเวลาการใช้งานอินเทอร์เน็ตสูงติดอันดับ 7 ของโลก โดยมีการใช้งานอินเทอร์เน็ตในแต่ละวันสูงถึง 9.06 ชั่วโมง แต่พบว่าระยะเวลาใช้งานอินเทอร์เน็ตเฉลี่ยทั่วโลกอยู่ที่ 6.58 ชั่วโมง โดยในปี 2022 ความเร็วอินเทอร์เน็ตบ้านของประเทศไทยติดอันดับ 2 ของโลก โดยมีความเร็วเฉลี่ยอยู่ที่ 171.37 MBPS ขณะที่ผู้ให้บริการบางรายในประเทศไทยสามารถระบุความเร็วอินเทอร์เน็ตบ้าน สูงสุดที่ 1000 MBPS นอกจากนี้ความเร็วอินเทอร์เน็ตบ้านของไทยยังพัฒนาขึ้นจากปีที่ผ่านมาถึง 46% กลายเป็นเป็นอันดับ 3 ของโลกที่มีการพัฒนาความเร็วอินเทอร์เน็ตบ้าน [2] ดังนั้น

จากข้อมูลต่าง ๆ ความเติบโตที่เพิ่มขึ้น และความเสรีในตลาด ผู้ให้บริการอินเทอร์เน็ตแต่ละค่ายต่างจึงออกโปรโมชั่นแข่งขันกัน ทั้งด้านความเร็ว และอุปกรณ์ต่าง ๆ ที่ดึงดูดใจผู้ให้บริการให้หันมาใช้บริการของตน แต่ปัญหาหลักที่พบมักอยู่ที่การบริการและคุณภาพบริการ ความ

รวดเร็วในการแก้ไขปัญหาเมื่อมีเหตุขัดข้อง ในบางครั้งอาจใช้เวลานาน อันเกิดจากสาเหตุต่าง ๆ ซึ่งทำให้ลูกค้าเกิดความไม่สะดวกในการใช้บริการ เมื่อเกิดเหตุขัดข้องบ่อยขึ้น และใช้เวลาในการดำเนินการแก้ไขที่นาน จะนำมาซึ่งผลที่ทำให้ลูกค้ามีความต้องการเปลี่ยนไปใช้งานกับผู้ให้บริการเจ้าอื่นที่สามารถให้บริการได้ดีกว่า ข้อตกลงระดับการให้บริการ (Service Level Agreement) จึงเป็นข้อตกลงเพื่อรับประกันคุณภาพการให้บริการ ระหว่างผู้ให้บริการกับผู้รับบริการ เพื่อเพิ่มความมั่นใจแก่ผู้รับบริการตามระดับที่ตกลงกันได้ โดยธุรกิจการให้บริการโทรคมนาคมมีการรับประกันมาตรฐาน Service Level Agreement ในระดับที่แตกต่างกันตามความสามารถขององค์กรนั้น ๆ (เฉลี่ยแล้วสำหรับลูกค้าปกติ คือ ต้องแก้ไขคืนดีภายใน 24 ชม.) โดยกำหนดให้มีบริการทางโทรศัพท์ (On Call Service) ตลอด 24 ชม. 7 วัน ให้บริการแจ้งเหตุ และสอบถามแก้ไขปัญหาต่าง ๆ ผ่านโทรศัพท์ หากไม่สามารถแก้ไขได้จะต้องส่งจากช่างที่เชี่ยวชาญไปยังจุดที่เกิดปัญหาภายในเวลาที่กำหนด เพื่อแก้ไขให้กลับมาใช้งานได้อย่างเป็นปกติโดยเร็วที่สุดนำมาซึ่งการรักษาลูกค้าเดิมให้ใช้บริการต่อไปและเพิ่มฐานลูกค้าใหม่ให้มากยิ่งขึ้น[3]

จากปัญหาดังกล่าว ผู้วิจัยจึงมีแนวคิดในการนำอัลกอริทึมมาทำการวิเคราะห์ระดับของการให้บริการ (SLA) และจัดทำรายงานเพื่อแสดงผลต่อไป

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล เป็นการวิเคราะห์ข้อมูลจากข้อมูลจำนวนมาก เพื่อหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ โดยทำการจำแนกประเภทรูปแบบเชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน และหาความน่าจะเป็นที่จะเกิดขึ้น เพื่อให้ได้องค์ความรู้ใหม่[4] ซึ่งสามารถนำไปใช้ประโยชน์ในการประกอบการตัดสินใจในต่าง ๆ ได้

### 2.2 การจัดเตรียมข้อมูล (Data Preparation)

การจัดเตรียมข้อมูล (Data Preparation) หมายถึง กระบวนการใด ๆ ที่มีความจำเป็นต้องทำกับข้อมูลดิบ

(Raw Data) ที่ได้รับมา เพื่อปรับเปลี่ยนข้อมูลให้อยู่ในรูปแบบที่เหมาะสมที่จะไปโหลดเข้าฐานข้อมูลหรือนำไปวิเคราะห์ต่อไป ความหมายที่เจาะจงของการทำ Data Preparation อาจจะแตกต่างกันไปสำหรับแต่ละระบบผลลัพธ์จากการประมวลผลของระบบหนึ่งอาจกลายเป็นข้อมูลดิบของระบบต่อไป ดังนั้น จึงต้องมีการจัดเตรียมข้อมูล เพื่อให้ได้ข้อมูลที่สมบูรณ์และสามารถนำไปใช้งานต่อไป

### 2.3 เพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor)

เป็นการเรียนรู้โดยการเปรียบเทียบกันระหว่างเรคคอร์ดของข้อมูลที่ต้องการจำแนกหรือทำนายหมวดหมู่กับเรคคอร์ดทั้งหมดในชุดข้อมูลสอน(TrainingData) โดยข้อมูลเรคคอร์ดหนึ่งจะสามารถถูกมองว่าเป็นจุดหนึ่งในระนาบ  $n$  มิติ นำข้อมูลทุก ๆ เรคคอร์ดในชุดข้อมูลสอนมาวางในระนาบ  $n$  มิติ จากนั้นนำข้อมูลเรคคอร์ดที่ต้องการจำแนกหรือทำนายหมวดหมู่มาวางในระนาบเช่นกัน แล้วพิจารณาหาว่ามีข้อมูลใดบ้างที่มีคุณลักษณะใกล้เคียงกับเรคคอร์ดที่ต้องการจำแนกมากที่สุดเป็นจำนวน  $k$  เรคคอร์ดในการที่จะหาความเหมือนกัน ต่างกัน หรือค่าความใกล้เคียงกันระหว่างเรคคอร์ดใด ๆ [5]

### 2.4 เทคนิคซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine)

เป็นการจัดกลุ่มข้อมูล Classification โดยการแบ่ง Class ของข้อมูลออกจากกัน อัลกอริทึมซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine) เป็นอัลกอริทึมที่จัดให้อยู่ในกลุ่มของฐานการเรียนรู้ มีการสร้างเส้นแบ่งข้อมูลออกเป็นส่วน โดยที่เส้นแบ่งข้อมูลจะมีระยะของความห่างที่มากที่สุดจากขอบเขตของข้อมูลที่ใกล้ที่สุดจากทั้งสองส่วน [6]

### 2.5 นาอิวเบย์ (Naïve Bayes)

ทฤษฎีของเบย์ (Bayes Theorem) มีอัลกอริทึมที่ไม่ซับซ้อน เป็นขั้นตอนวิธีในการจำแนกข้อมูลโดยการเรียนรู้ปัญหาที่เกิดขึ้น เพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ ใช้การคำนวณหาความน่าจะเป็นในการทำนายผลเป็นเทคนิคในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ จะทำการวิเคราะห์ความสัมพันธ์

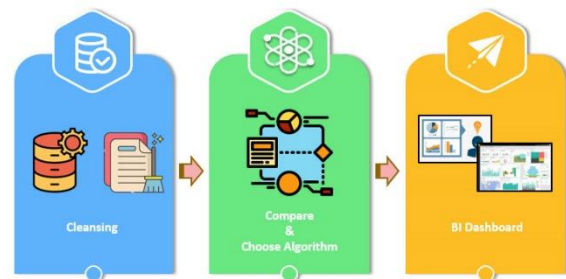
ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน [7]

### 2.6 เอนเซมเบิล (Ensemble Method)

Ensemble Method คือหนึ่งในเทคนิคของ Machine learning ที่นำ Model หลาย ๆ Model มารวมกันเพื่อที่จะทำการวิเคราะห์ให้ผลลัพธ์ออกมาดีที่สุด ซึ่ง Model ที่นำมาใช้สามารถใช้ Algorithm ของ Machine learning ได้หลากหลายแบบอย่าง หลักการสร้างโมเดล Ensemble คือโมเดลที่สร้างควรจะมีความหลากหลายเพื่อให้ทำนายข้อมูลแบบต่าง ๆ กัน [8]

## 3. วิธีดำเนินการวิจัย

กรอบแนวความคิดของการวิจัยงานวิจัยครั้งนี้เป็นการวิเคราะห์ เพื่อหาอัลกอริทึมที่ให้ประสิทธิภาพที่ดีและแม่นยำที่สุดในการนำมาใช้วิเคราะห์ข้อมูลระดับการให้บริการที่ถูกต้อง แสดงดัง ภาพที่ 1



ภาพที่ 1 : กรอบแนวคิด

โดยผู้ทำวิจัยได้นำข้อมูลการรับแจ้งเหตุจากระบบรับแจ้งเหตุขัดข้องของบริษัทโทรคมนาคมแห่งชาติจำนวน 306,075 รายการ จากนั้นนำเข้าสู่กระบวนการการเตรียมข้อมูล (Data Preparation) ให้อยู่ในรูปแบบที่เหมาะสมที่จะสามารถนำไปโหลดเข้าฐานข้อมูลหรือนำไปวิเคราะห์ต่อไป จากนั้นคัดเลือกข้อมูลที่จัดเตรียมเรียบร้อยแล้วมาทดสอบกับอัลกอริทึมที่สนใจที่จะนำมาใช้ในการวิเคราะห์ข้อมูลเพื่อเปรียบเทียบหาอัลกอริทึมที่มีประสิทธิภาพที่ดีที่สุด จากนั้นจึงนำข้อมูลที่ผ่านการวิเคราะห์ด้วยอัลกอริทึมที่ดีที่สุดนั้นมาจัดแสดงผลภาพรวมของข้อมูลโดยใช้โปรแกรม Power Bi มาพัฒนารายงาน เพื่อให้สามารถ

เข้าใจได้ง่าย โดยแบ่งขั้นตอนการดำเนินงานออกเป็น 4 ขั้นตอน ดังนี้

**3.1 การศึกษาและเปรียบเทียบเทคนิคของอัลกอริทึม**

ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องและศึกษาอัลกอริทึม 4 อัลกอริทึม ได้แก่ Support Vector Machine K-Nearest Neighbor Naïve Bays และ Ensemble Method มาใช้ในการเปรียบเทียบเพื่อหาอัลกอริทึมที่ดีที่สุด ในการเรียนรู้และวิเคราะห์ผลระดับการผ่านคุณภาพของการให้บริการ หรือ SLA (Service Level Agreement)

**3.2 การเตรียมข้อมูลและคัดเลือกข้อมูล**

ข้อมูลที่นำมาใช้ในงานวิจัยครั้งนี้ เป็นข้อมูลที่รวบรวมจากระบบรับแจ้งเหตุขัดข้องช่วงวันที่ 1 มกราคม พ.ศ.2564 จนถึง 30 มิถุนายน พ.ศ. 2564 โดยเป็นข้อมูลเกี่ยวกับการรับแจ้งเหตุขัดข้องในการใช้บริการ โดยคัดเลือกข้อมูลจำนวน 11,000 เพื่อใช้ในการวิเคราะห์ห้ออกแบบเรียนรู้และทดสอบอัลกอริทึมและสร้างโมเดลการวิเคราะห์ข้อมูลแสดงดัง ตารางที่ 1

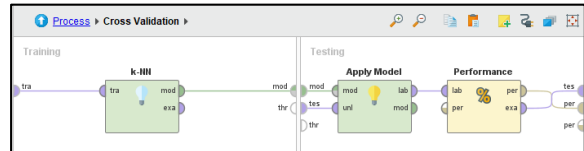
ตารางที่ 1 : ตัวอย่างข้อมูลที่นำมาใช้ในการวิเคราะห์

ลำดับ	ชื่อข้อมูล	คำอธิบาย
1	Region	พื้นที่
2	Department	สังกัด
3	Service_type	ประเภทบริการ
4	Fault_type	ประเภทการขัดข้อง
5	Day	วันที่เกิดเหตุขัดข้อง
6	Month	เดือนที่เกิดเหตุขัดข้อง
7	Repair_Time	ระยะเวลาในการแก้ไข
8	SLA	ผลคุณภาพการให้บริการ
9	SLA Rank	ระดับคุณภาพการให้บริการ

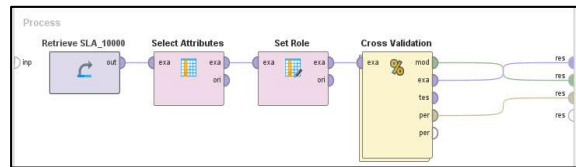
**3.3 ขั้นตอนการวิเคราะห์ข้อมูลโดยใช้เทคนิคเหมืองข้อมูล เปรียบเทียบ และคัดเลือกเทคนิคที่มีประสิทธิภาพที่สุดไปใช้งาน**

ผู้วิจัยได้ทำการวิเคราะห์ผลระดับการผ่านคุณภาพของการให้บริการ หรือ SLA (Service Level Agreement) โดยใช้ อัล ก อ ริ ทึม Support Vector Machine K-Nearest Neighbor Naïve Bays และ Ensemble Method มาใช้ในการเปรียบเทียบเพื่อหาอัลกอริทึมที่ดีที่สุดด้วยโปรแกรม

RapidMiner Studio โดยใช้ข้อมูลฝึกสอน (Training Data) จำนวน 10,000 รายการ เพื่อนำมาใช้ในการฝึกสอน และสร้างตัวแบบจำลองของแต่ละอัลกอริทึม เพื่อให้ได้ผลลัพธ์ออกมา เมื่อทำการวิเคราะห์แต่ละเทคนิคเรียบร้อยแล้ว จึงทำการสรุปผลและเลือกเทคนิคที่เหมาะสมและมีความแม่นยำมากที่สุด เพื่อนำมาวิเคราะห์กับข้อมูลทดสอบจำนวน 1,000 รายการ (Test Data) ต่อไป



ภาพที่ 2 : แสดงรูปขั้นตอนการหาประสิทธิภาพของอัลกอริทึม(1)



ภาพที่ 3 : แสดงรูปขั้นตอนการหาประสิทธิภาพของอัลกอริทึม(2)

**3.4 ขั้นตอนการแสดงผลการวิเคราะห์ข้อมูลจากแบบจำลอง และภาพรวมของข้อมูล**

ขั้นตอนนี้เป็นการนำผลที่ได้จากการวิเคราะห์ข้อมูลมาแสดงผลระดับการผ่านคุณภาพการให้บริการ หรือ SLA (Service Level Agreement) โดยใช้โปรแกรม Power Bi ในการพัฒนารายงาน เพื่อนำไปใช้วางแผนการดำเนินงาน และช่วยเสริมการตัดสินใจอย่างถูกต้องให้แก่ผู้บริหาร

**4. ผลการดำเนินงาน**

ผลการพัฒนาระบบธุรกิจอัจฉริยะ เพื่อการวิเคราะห์ข้อมูลจากระบบการรับแจ้งเหตุขัดข้องโดยใช้เทคนิคเหมืองข้อมูล สามารถแบ่งหัวข้อ ดังนี้

**4.1 ผลการเปรียบเทียบอัลกอริทึม**

ทำการเปรียบเทียบอัลกอริทึม โดยนำชุดข้อมูลตัวอย่างมาเพื่อเรียนรู้อัลกอริทึมที่คัดเลือกมาใช้ในการวิเคราะห์ จากนั้นนำข้อมูลมาวิเคราะห์ปัจจัยต่าง ๆ จากข้อมูลที่พบในการรับแจ้งเหตุขัดข้องที่ส่งผลกระทบต่อระดับคุณภาพการให้บริการ โดยทำการเปรียบเทียบด้วย อัลกอริทึม Support

Vector Machine K-Nearest Neighbor Naïve Bays และ Ensemble Method โดยผลลัพธ์การเปรียบเทียบแสดงดังตารางที่ 2

ตารางที่ 2 : ผลการเปรียบเทียบความแม่นยำ

ลำดับ	อัลกอริทึม	Performance		
		Accuracy	Precision	Recall
1	K-Nearest Neighbor	99.28%	0.99	0.99
2	Support Vector Machine	66.28%	0.66	0.66
3	Naïve Bays	98.81%	0.98	0.98
4	Ensemble Method	90.26%	0.88	0.90

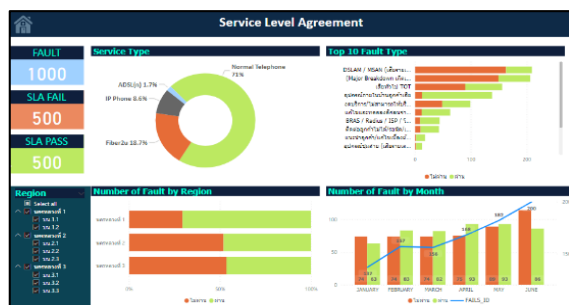
จากการทดสอบทั้ง 4 อัลกอริทึม พบว่า K-Nearest Neighbor ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่าความถูกต้องอยู่ที่ร้อยละ 99.28 ค่าความแม่นยำเท่ากับ 0.99 และค่าระลึกเท่ากับ 0.99

4.2 ผลการพัฒนากระบวนการรายงานเพื่อแสดงผล

ภาพแสดงหน้าจอเริ่มต้นของระบบรายงานการวิเคราะห์ข้อมูลจากระบบรับแจ้งเหตุขัดข้อง โดยสามารถเลือกเข้าใช้งานรายงานที่ต้องการได้ แสดงดังภาพที่ 5



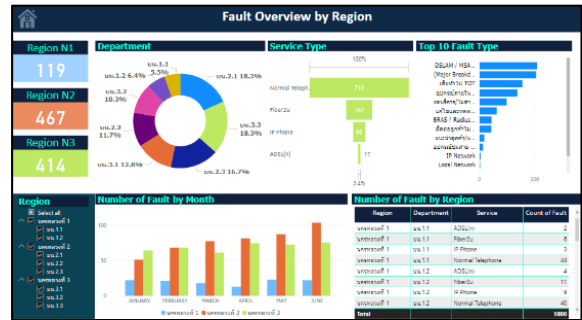
ภาพที่ 5 : หน้าแรกของระบบรายงาน



ภาพที่ 6 : ผลจากการวิเคราะห์

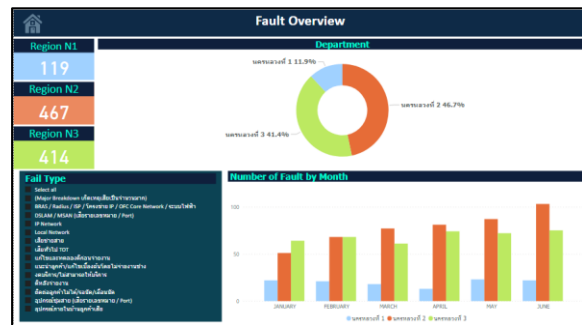
จากภาพที่ 6 แสดงผลการวิเคราะห์ข้อมูลระดับการให้บริการ จำนวน 1000 รายการที่ได้จากการทดสอบ สามารถกรองข้อมูลเป็นรายพื้นที่ และสังกัด จะแสดงข้อมูลจำนวนของเหตุขัดข้องที่เกิดขึ้นแยกตามบริการ

พื้นที่ ประเภทเหตุขัดข้อง ข้อมูลและผลการผ่านคุณภาพของการให้บริการ หรือ SLA (Service Level Agreement)



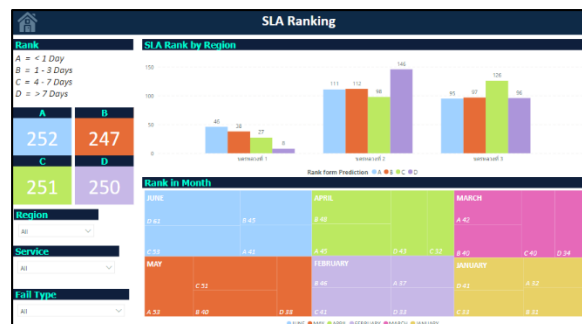
ภาพที่ 6 : ข้อมูลเหตุขัดข้องตามสังกัด

จากภาพที่ 6 แสดงข้อมูลเหตุขัดข้องแยกตามสังกัด และพื้นที่การให้บริการอย่างลงรายละเอียดมากขึ้น โดยสามารถกรองข้อมูล ตามพื้นที่ และสังกัดได้



ภาพที่ 7 : ข้อมูลเหตุการณ์เหตุขัดข้อง

จากภาพที่ 7 สามารถเลือกแสดงรายการเหตุขัดข้องที่พบ โดยจะสรุปจำนวนตามพื้นที่ และจำนวนการพบเหตุขัดข้องในแต่ละเดือนได้

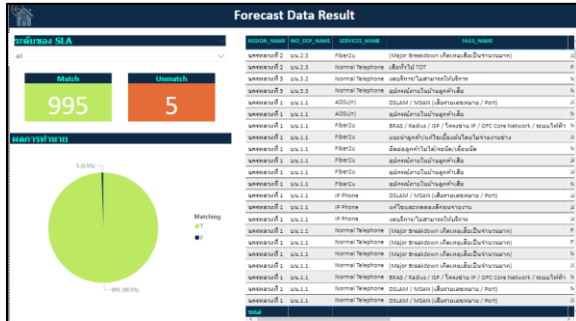


ภาพที่ 8 : ผลการวิเคราะห์ระดับการผ่านคุณภาพการให้บริการ

จากภาพที่ 8 แสดงระดับการผ่านคุณภาพการให้บริการ โดยจะแบ่งเป็นระดับที่ได้รับจากการวิเคราะห์โดยสามารถกรองข้อมูลด้วยสังกัด ประเภทบริการ ประเภท



เหตุขัดข้อง โดยแสดงระดับ และจำนวนเหตุขัดข้อง ตามพื้นที่บริหาร และแสดงผลรวมในแต่ละเดือนตามระดับที่ได้รับ



ภาพที่ 9 : ผลการทดสอบด้วยอัลกอริทึมที่มีประสิทธิภาพดีที่สุด จากภาพที่ 9 แสดงผลที่ได้จากการทดสอบข้อมูลตัวอย่างด้วยอัลกอริทึมที่ดีที่สุด



ภาพที่ 10 : ข้อมูลที่ใช้ในการสร้างแบบจำลอง จากภาพที่ 10 แสดงรายละเอียดข้อมูลที่นำเข้ามาเพื่อสร้างแบบจำลองด้วยอัลกอริทึมที่มีประสิทธิภาพดีที่สุด

5. สรุป อภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ระดับคุณภาพของการให้บริการ หรือ SLA (Service Level Agreement) โดยแสดงผลของระดับการแก้ไขเหตุขัดข้องที่ผ่านคุณภาพการให้บริการ และไม่ผ่านคุณภาพการให้บริการ แบ่งตามระดับคุณภาพบริการที่ได้ โดยเปรียบเทียบและหาอัลกอริทึมที่มีประสิทธิภาพที่ดีที่สุด เพื่อนำมาใช้ในการวิเคราะห์ระดับของคุณภาพบริการ

จากการทดสอบและเปรียบเทียบประสิทธิภาพอัลกอริทึม พบว่าอัลกอริทึม K-Nearest Neighbor มีผลลัพธ์ที่ดีที่สุด โดยมีค่าความถูกต้องอยู่ที่ร้อยละ 99.28 ค่าความ

แม่นยำเท่ากับ 0.99 และค่าระยะเท่ากับ 0.99 จากนั้นจึงนำข้อมูลที่นำมาพัฒนารายงาน โดยใช้โปรแกรม Power Bi เพื่อแสดงผลในรูปแบบรายงานต่าง ๆ ให้สามารถเข้าใจได้ง่าย และสามารถนำมาช่วยในการตัดสินใจของผู้บริหารต่อไปได้ในอนาคต

เอกสารอ้างอิง

- [1] บมจ.โทรคมนาคมแห่งชาติ Nation Telecom Public Company Limited.2565[อินเทอร์เน็ตสืบค้นเมื่อ] [9 ธันวาคม 2564].แหล่งข้อมูล : <https://www.ntplc.co.th/About/about>
- [2] WEARESOCIAL.COM.2565[อินเทอร์เน็ตสืบค้นเมื่อ] [11 มกราคม 2565].แหล่งข้อมูล:<https://wearesocial.com/us/blog/2022/01/digital-2022-another-year-of-bumper-growth-2>
- [3] บมจ.โทรคมนาคมแห่งชาติ Nation Telecom Public Company Limited.2565[อินเทอร์เน็ตสืบค้นเมื่อ] [11 มกราคม 2565].แหล่งข้อมูล : [https://www.ntplc.co.th/pdf\\_doc/BroadbandQuality](https://www.ntplc.co.th/pdf_doc/BroadbandQuality)
- [4] กมลวรรณ จันทรมิณี และพวง มีสีจ, “การพัฒนาระบบแนะนำด้วยเทคนิคกฎความสัมพันธ์และการจัดกลุ่มข้อมูลด้วย Cosine Similarity” บทความวิจัการประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 15, 2562.
- [5] ศศิธร โหราชิต และมณฑิยา รัตนศิริวงษ์วุฒิ, “ระบบธุรกิจอัจฉริยะเพื่อการพยากรณ์แนวโน้มความต้องการแรงงานของตลาดแรงงานในประเทศไทย โดยเทคนิคเหมืองข้อมูล”
- [6] วงศกร คำวงศ์และวัชรวิพรรณ จิตต์สกุล, “ระบบคัดกรองความเหมาะสมของเนื้อหาก่อนเผยแพร่สู่สาธารณะ” บทความวิจัการประชุมวิชาการระดับชาติด้านวิทยาศาสตร์ และเทคโนโลยีครั้งที่ 4, 2563.
- [7] อนันต์ชัย ชูดีภาสเจริญ และ ดร.จรัญ แสนราช "การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อพยากรณ์โอกาสความสำเร็จในการ โอนเงินข้ามประเทศของบุคคลทั่วไป" KRU RESEARCH JOURNAL OF HUMANITIES AND SOCIAL SCIENCES (GRADUATE STUDIES)Vol. 6, 2561.
- [8] การสร้างโมเดล Ensemble [อินเทอร์เน็ต] [สืบค้นเมื่อ 20 มกราคม 2565].แหล่งข้อมูล : <https://medium.com/cw-quantlab/ensemble-learning-method-98359636adf9>

# ระบบวิเคราะห์ระดับของภาษาด้วยเหมืองข้อความ

## Thai Language Level Analysis System using Text Mining Technique

สาริน สงครินทร์ (Sarin Songkarin)<sup>1</sup>, อภิสิตี ไกรยะโส (Apisit Kraiyaso)<sup>2</sup>

และวรัญญา วรณศรี (Warunya Wunnasri)<sup>3</sup>

สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

<sup>1</sup>rin\_sarin@kkumail.com, <sup>2</sup>apisit\_kra@kkumail.com, <sup>3</sup>waruwu@kku.ac.th

### บทคัดย่อ

เนื่องจากการเขียนหรือเรียบเรียงข้อมูลในรายงานจะต้องทำอย่างเป็นระบบและต้องมีการใช้ระดับภาษาเข้ามาช่วยในการเขียนรายงาน ซึ่งนักเรียนและนักศึกษา ยังไม่เข้าใจเรื่องระดับภาษามากนักจึงทำให้กระทบต่อการเขียนรายงาน จากปัญหานี้ผู้จัดทำโครงการได้พัฒนา ระบบวิเคราะห์ระดับของภาษาด้วยเหมืองข้อความ เพื่ออำนวยความสะดวกในการเขียนรายงาน โดยใช้เทคนิคการเรียนรู้ด้วยเครื่อง โดยเปรียบเทียบแบบจำลองการจำแนกข้อมูล 3 วิธี ประกอบด้วยการค้นหาเพื่อนบ้านใกล้ที่สุด ต้นไม้ตัดสินใจและนาอิวเบย์ พบว่านาอิวเบย์ให้ค่าความถูกต้องที่ดีที่สุดที่ร้อยละ 75.63 จึงได้เลือกใช้แบบจำลองนาอิวเบย์ มาพัฒนาเป็นระบบวิเคราะห์ระดับภาษาด้วยเหมืองข้อความ

**คำสำคัญ:** การจัดแบ่งประเภทของเอกสาร การประมวลผลภาษาธรรมชาติ ระดับภาษาไทย

### Abstract

Nowadays, writing reports and essays should be organized systematically. Especially, writing an academic report using the Thai language requires knowledge about language level. Therefore, the writers should have the skill to use language level appropriately. Unfortunately, there are many students and non-professional writers who lack this skill. The paper proposes the Thai language level analysis system using the text mining technique.

Three techniques: Naive Bays, K-Nearest Neighbor, and Decision Tree are compared to classify the sentences between formal and informal sentences. The result shows that the Naive Bays claim the highest accuracy with 75.63 percent. Hence, it is applied to develop the Thai language level analysis system, a web application to support the writers improve their language level skills.

**Keywords:** Classification of Documents, Natural Language Processing, Thai Language Level

### 1. บทนำ

การศึกษาปัจจุบันส่งเสริมให้ผู้เรียนศึกษาค้นคว้าด้วยตนเองมากขึ้น การเขียนรายงานจึงมีบทบาทและสำคัญยิ่งในการเรียนการสอน ซึ่งส่งเสริมให้ผู้เรียนทำงานให้เป็นระเบียบเรียบร้อย มีเหตุผล มีทักษะในการใช้ภาษา เพราะการเขียนหรือเรียบเรียงข้อมูลในรายงานจะต้องทำอย่างเป็นระบบและต้องมีการใช้ระดับภาษาเข้ามาช่วยในการเขียนรายงาน ภาษาที่ใช้ในการเขียนนั้นต้องเป็นภาษาระดับทางการ ซึ่งการใช้คำในภาษาไทยใช้คำในภาษาราชการ ไม่ใช่ภาษาพูด และให้เป็นคำภาษาไทยมากที่สุด ใช้คำที่สามารถเข้าใจง่าย ไม่ใช่คำภาษาถิ่นหรือพวน คำย่อจำเป็นต้องเขียนเป็นคำเต็ม เมื่อต้องใช้ศัพท์ทางวิชาการ ควรเลือกใช้ศัพท์บัญญัติของราชบัณฑิตยสถาน

จากปัญหาผู้จัดทำโครงการได้พัฒนาระบบวิเคราะห์ระดับของภาษาด้วยเหมืองข้อความ เพื่ออำนวยความสะดวกสำหรับผู้เรียนในการเขียนรายงาน โดยผู้จัดทำ

รวบรวมข้อมูลประโยคจากรายงานทั้งหมด 2,000 ประโยค ประกอบไปด้วยประโยคภาษาระดับทางการ 1,000 ประโยค และประโยคภาษาระดับไม่เป็นทางการ 1,000 ประโยค ซึ่งต้องเตรียมข้อมูลด้วยการตัดคำ นำมาสร้างคลังคำศัพท์ และนับความถี่ของคำศัพท์ว่าแต่ละคำนั้นปรากฏบนเอกสารมากน้อยเพียงใด โดยใช้เทคนิคการเรียนรู้ด้วยเครื่อง นำมาวิเคราะห์และเปรียบเทียบด้วยแบบจำลองการจำแนกข้อมูล 3 วิธี ได้แก่ เทคนิคต้นไม้ตัดสินใจ(Decision Tree) การหาเพื่อนบ้านใกล้ที่สุด(K-Nearest Neighbor) และ นาอิว เบย์(Naive Bayes) เพื่อหาวิธีที่ดีที่สุดมาพัฒนาเป็นระบบวิเคราะห์ของระดับภาษา

## 2. งานวิจัยและทฤษฎีที่เกี่ยวข้อง

จากงานวิจัยระดับภาษาและการใช้ภาษาที่ถูกต้อง [1] ได้ศึกษาการพิจารณาระดับภาษาและหลักการใช้ภาษาที่ถูกต้อง ตามกาลเทศะหรือ โอกาสในการใช้ภาษา เพื่อให้ผู้ใช้ภาษาสามารถเลือกใช้ภาษาในสถานการณ์ต่างๆ โดยมีหลักการดังนี้

- ภาษาใช้ในการสื่อสารที่เป็นทางการควรใช้ภาษาเขียนในระดับทางการ โดยต้องหลีกเลี่ยงการใช้ภาษาพูด และต้องคำนึงถึงความถูกต้องเรื่องความหมาย และแบบแผนการใช้ภาษาทั้งในด้านการใช้คำและประโยค เน้นความถูกต้อง กระชับ ชัดเจน ตรงประเด็น

- การใช้คำ การใช้คำต้องพิจารณาให้ถูกต้องตามหลักไวยากรณ์ ความหมาย และระดับของคำนั้น ๆ

- การใช้ประโยค การใช้ประโยคต้องพิจารณาให้ประโยคที่ใช้ในนั้นถูกต้อง กะทัดรัด ชัดเจน สละสลวย และลำดับคำในประโยคถูกต้อง

- การเปรียบเทียบคำระหว่างภาษาระดับทางการและภาษาระดับไม่เป็นทางการนั้น สามารถเปรียบเทียบได้ ดังนี้ ภาษาเป็นทางการมีลักษณะกระชับ ภาษาไม่เป็นทางการมีลักษณะไม่กระชับค่อนข้างเป็นภาษาพูด ดังตัวอย่างในตารางที่ 1 คำว่า “ไม่ระมัดระวัง” ในภาษาทางการจะกระชับและสุภาพกว่า “ไม่ดูตาม้าตาเรือ” ในภาษาไม่เป็นทางการ ภาษาเป็นทางการส่วนมากใช้คำบาลีสันสกฤต ตัวอย่าง ประโยคภาษาระดับไม่เป็น

ทางการ เช่น ใครไม่ทำตามกฎระเบียบ สามารถเขียนเป็นประโยคภาษาระดับทางการได้ คือ ผู้ใดไม่ปฏิบัติตามกฎระเบียบ

**ตารางที่ 1:** ตัวอย่างภาษาระดับทางการมีลักษณะกระชับ ภาษาระดับไม่เป็นทางการมีลักษณะไม่กระชับ

ภาษาระดับทางการ	ภาษาระดับไม่เป็นทางการ
ไม่ระมัดระวัง	ไม่ดูตาม้าตาเรือ
ไม่ทราบข้อเท็จจริง	ไม่รู้เรื่องรูราว
ซ้ำซาก	ซ้ำๆ ซากๆ
มีบุตร	มีลูกมีเต้า
น้อยใจ	น้อยอกน้อยใจ
คิดใจ	คิดอกคิดใจ

จากงานวิจัยเรื่องการจำแนกกลุ่มคำถามอัตโนมัติบนกระดานสนทนาโดยใช้เทคนิคเหมืองข้อความ[2] ได้ศึกษาการประมวลผลข้อความและการจำแนกข้อความด้วยการเรียนรู้ด้วยเครื่องทั้ง 3 เทคนิค ประกอบด้วย การหาเพื่อนบ้านที่ใกล้ที่สุด ต้นไม้ตัดสินใจ และ นาอิว เบย์ ซึ่งงานวิจัยนี้มีปัญหาด้านการตอบคำให้ชัดเจนและตรงประเด็น ทำให้ผู้พัฒนาได้นำเอา 3 เทคนิค มาใช้ในการพัฒนาระบบวิเคราะห์ระดับภาษา โดยมีหลักการดังนี้

### การประมวลผลข้อความ (Text processing)

การประมวลผลข้อความ เป็นกระบวนการอัตโนมัติในการวิเคราะห์และจัดเรียงข้อมูลข้อความที่ไม่มีโครงสร้างเพื่อให้ได้ข้อมูลเชิงลึก [2] ซึ่งมีขั้นตอนดังนี้

### การตัดคำ

การตัดคำภาษาไทยเป็นงานพื้นฐานสำหรับการประมวลผลภาษาไทย เนื่องจากลักษณะของข้อความจะติดกันและไม่มีการใช้ตัวอักษรหรือสัญลักษณ์นำมามีคั่นระหว่างคำเหมือนภาษาอังกฤษ ดังนั้นการตัดคำจึงเป็นการแบ่งคำแต่ละคำในประโยคออกจากกันเพื่อให้คอมพิวเตอร์สามารถคำนวณหรือค้นหาคำได้อย่างถูกต้องและแม่นยำ

การสร้างคลังคำศัพท์

เมื่อตัดคำแล้วจะนำคำที่ตัดมารวบรวมเป็นคลังคำศัพท์และแต่ละคำนั้น จะมีหมายเลขกำกับเพื่อระบุว่าจะแต่ละคำนั้นอยู่ลำดับที่เท่าไร

สร้างคุณลักษณะจากการนับคำ

การนับจำนวนครั้งที่แต่ละคำ ในคลังคำศัพท์ทั้งหมดที่ปรากฏในประโยค ซึ่งยิ่งคำในคลังคำศัพท์มีจำนวนมาก จำนวนคุณลักษณะของข้อมูล ก็จะมีจำนวนเพิ่มขึ้นและแปลงคำศัพท์ให้เป็นเวกเตอร์ เพื่อจะได้นำไปใช้ในขั้นตอนการเรียนรู้ของเครื่องต่อไป

เทคนิคการเรียนรู้ด้วยเครื่อง

- นาอ็ฟ เบย์

เป็นการจัดหมวดหมู่โดยใช้หลักความน่าจะเป็นเข้ามาช่วยคำนวณ ดังสมการที่ 1

$$P(c|X) = \frac{P(x|c)P(c)}{P(x)} \tag{1}$$

ตัวแปรในสมการมีดังนี้

X คือ ประโยคที่ต้องการวิเคราะห์

x คือ คำในประโยค

c คือ คลาสของข้อมูลในที่นี้ประกอบด้วย 2 คลาส คือ ภาษาระดับทางการและภาษาระดับไม่เป็นทางการ

$P(c|X)$  คือ ความน่าจะเป็นที่ประโยค X จะเป็นคลาสนั้นๆ

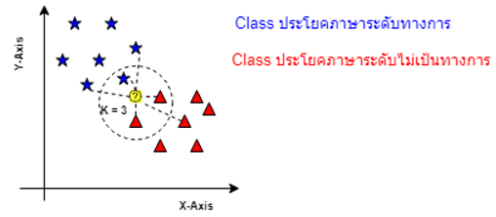
$P(x|c)$  คือ ความน่าจะเป็นที่คำ x ปรากฏอยู่ในคลาสนั้นๆ

$P(c)$  คือ ความน่าจะเป็นของการเกิดของคลาสนั้นๆ ในคลังข้อมูล

$P(x)$  คือ ความน่าจะเป็นของคำ x

- การหาเพื่อนบ้านที่ใกล้ที่สุด

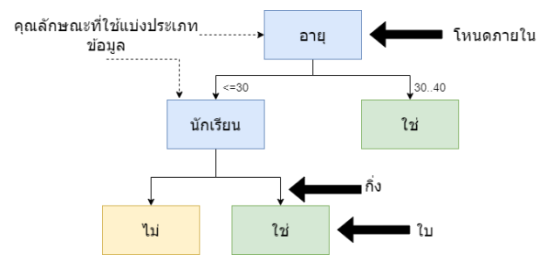
เป็นแบบจำลองในการจำแนกประเภทของข้อมูลจากการเรียนรู้ข้อมูลฝึกฝนที่ทราบคำตอบ วิธีการจำแนกของการหาเพื่อนบ้านที่ใกล้ที่สุด จะใช้วิธีการวิเคราะห์จากข้อมูลที่ใกล้เคียงที่สุดจำนวน k ตัว กับข้อมูลที่ต้องการจำแนกประเภทของข้อมูลหรือต้องการ ทำนายคลาสของข้อมูลใหม่โดยจะทำนายตามคลาสส่วนใหญ่ของข้อมูลฝึก k ตัว ดังภาพที่ 1



ภาพที่ 1: ข้อมูลที่ใกล้เคียงและข้อมูลที่ต้องการจำแนกประเภท [3]

- ต้นไม้ตัดสินใจ

การจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ จะสร้างต้นไม้ขึ้นเพื่อใช้ในการตัดสินใจจากข้อมูลที่มีหมวดหมู่ข้อมูลแนบอยู่ด้วย โดยโครงสร้างของต้นไม้ตัดสินใจจะประกอบด้วย ใบ (Leaf) เป็นส่วนของผลลัพธ์ที่ได้ถึง (Branch) เป็นตัวเชื่อมระหว่างโหนดราก (Root) เป็นจุดเริ่มต้นของเหตุการณ์ โหนดภายใน (Internal node) คือ คุณลักษณะต่างๆ ของข้อมูล ซึ่งเมื่อข้อมูลมาถึงโหนดจะใช้คุณลักษณะของข้อมูลเพื่อตัดสินใจว่าข้อมูลจะไปทิศทางใดต่อไป โดยโหนดภายในจะมีจุดเริ่มต้นคือโหนดราก ดังภาพที่ 2



ภาพที่ 2: โครงสร้างต้นไม้ตัดสินใจ [4]

จากงานวิจัย Automatic Classification of Documents by Formality [5] ได้ศึกษาความแตกต่างระหว่างประโยคภาษาระดับทางการและประโยคภาษาระดับไม่เป็นทางการของภาษาอังกฤษ โดยใช้การเรียนรู้ด้วยเครื่อง 3 เทคนิค ในการจำแนกประเภทประโยค ประกอบด้วย ต้นไม้ตัดสินใจ นาอ็ฟ เบย์ และ Support Vector Machine โดยใช้ชุดข้อมูล 1,000 รายการ คือ ประโยคภาษาระดับทางการ 500 ประโยค และ ประโยคภาษาระดับไม่เป็นทางการ 500 ประโยค ซึ่งงานวิจัยนี้ระบุว่า ชุดข้อมูลที่นำมาทดสอบยังมีจำนวนที่ไม่มากพอ

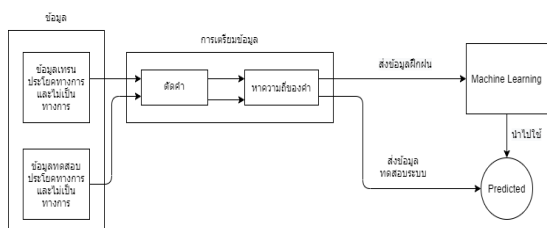
จากงานวิจัยการแบ่งข้อมูลและการจำแนกประเภทข้อมูล [6] งานวิจัยนี้ทำการศึกษา การเรียนรู้ด้วยเครื่อง

ด้วยเทคนิคต้นไม้ตัดสินใจ เป็นการเรียนรู้โดยการจำแนกประเภทข้อมูลออกเป็นกลุ่มโดยใช้คุณลักษณะ ในการจำแนกประเภท จากงานวิจัยนี้ผู้พัฒนาได้นำเอาเทคนิคต้นไม้ตัดสินใจ มาใช้ในการพัฒนาระบบเพื่อเปรียบเทียบแบบจำลองที่ดีที่สุด

จากบทความระดับภาษาและคำราชาศัพท์ [7] ผู้พัฒนาได้ทำการศึกษารูปแบบการใช้ภาษาและการสื่อสารที่มีลักษณะเป็นภาษาระดับทางการ บริบทของผู้สื่อสารเพื่อนำมาใช้ในการเตรียมชุดข้อมูล

### 3. วิธีการดำเนินงาน

ระบบวิเคราะห์ระดับของภาษาด้วยเหมืองข้อความ จะรับข้อมูลจากผู้ใช้งานและตรวจสอบว่าข้อมูลจากผู้ใช้มีคำอยู่ในระดับใดบ้างและเป็นประโยคภาษาระดับทางการหรือประโยคภาษาระดับไม่เป็นทางการแล้วแจ้งให้กับผู้ใช้



ภาพที่ 3: โครงสร้างในส่วนการฝึกฝนแบบจำลอง

ซึ่งในขั้นตอนการฝึกฝนแบบจำลองสามารถแสดงได้ดังภาพที่ 3 มีรายละเอียดการทำงานแต่ละขั้นตอนดังนี้

1. ข้อมูล ประกอบไปด้วยข้อมูลการฝึกฝนและข้อมูลทดสอบทั้งหมด 2,000 ประโยค
  - ข้อมูลการฝึกฝน 2,000 ประโยค ประกอบไปด้วย ประโยคภาษาระดับทางการ 1,000 ประโยคและประโยคภาษาระดับไม่เป็นทางการ 1,000 ประโยค
2. การเตรียมข้อมูล
  - การตัดคำ โดยการ ใช้ฟังก์ชัน word tokenize ของไลบรารี Pythainlp ตัดคำตามพจนานุกรม Pythainlp
  - Countvectorizer คือ การนำกลุ่มของประโยคมาสร้างเป็นเวกเตอร์ โดยใช้กลุ่มของคำที่ปรากฏในเอกสารเป็นตัวอ้างอิงคำที่มีในประโยคจะถูกตั้งค่าเป็น 1 คำที่ไม่

ปรากฏในเอกสารจะเป็น 0 และถ้ามีมากกว่า 1 คำ จะตั้งค่าเป็น 2

#### 3. การฝึกฝนแบบจำลอง

- แบบจำลองทำการจำแนกข้อมูลของประโยคภาษาระดับทางการและประโยคภาษาระดับไม่เป็นทางการ โดยแบบจำลองที่นำมาเปรียบเทียบประสิทธิภาพประกอบด้วย

#### นาอิว เบย์

ใช้ไลบรารี scikit – learn และเลือกใช้เป็น Multinomial Naïve Bayes โดยนำประโยคที่ทำการแปลงเป็น เวกเตอร์ แล้ว มาหาค่าความน่าจะเป็นเพื่อใช้พิจารณาค่าในประโยคนั้นว่าเป็นประโยคภาษาระดับทางการหรือประโยคภาษาระดับไม่เป็นทางการ

#### การหาเพื่อนบ้านที่ใกล้ที่สุด

ใช้ไลบรารี scikit – learn ในการแยกประเภทประโยค โดยใช้ข้อมูลบางชุดมาตรวจสอบ จากนั้นหาระยะห่างระหว่างข้อมูลที่ทำการรับเข้ามาและตัวข้อมูลที่จะใช้ในการตรวจสอบ เมื่อได้ค่าระยะห่างมาแล้วจะใช้เพื่อพิจารณาว่าประโยคที่รับเข้ามาเป็นประโยคภาษาระดับทางการหรือประโยคภาษาระดับไม่เป็นทางการ โดยได้กำหนดค่า k เท่ากับ 3 5 และ 7 ตามลำดับ

#### ต้นไม้ตัดสินใจ

ใช้ไลบรารี scikit – learn ทำการเลือก คุณลักษณะของข้อมูล ที่เหมาะสมจะนำมาใช้ในการแบ่ง โดยใช้ค่า Gini Index ในการเปรียบเทียบประโยคว่าเป็นประโยคภาษาระดับทางการหรือประโยคภาษาระดับไม่เป็นทางการ

ในส่วนการออกแบบและพัฒนาระบบในรูปแบบเว็บแอปพลิเคชัน ได้มีการออกแบบฟังก์ชันการทำงานดังนี้

- ระบบสามารถตรวจสอบประโยคที่ให้ผลลัพธ์เป็นภาษาระดับทางการหรือภาษาระดับไม่เป็นทางการโดยใช้แบบจำลองนาอิว เบย์ที่ได้ฝึกฝนจากชุดข้อมูลที่เตรียมไว้
- ระบบสามารถแนะนำคำที่เป็นภาษาระดับไม่เป็นทางการให้เป็นภาษาระดับทางการจากพจนานุกรมที่ผู้พัฒนาสร้างขึ้น

- ระบบสามารถรับข้อเสนอแนะจากผู้ใช้หากเว็บแอปพลิเคชันตรวจสอบประโยคผิดพลาด หรือผู้ใช้สามารถแนะนำคำที่เป็นภาษาระดับไม่เป็นทางการให้ผู้พัฒนาได้ทราบ

ซึ่งหน้าแรกของเว็บแอปพลิเคชัน ดังภาพที่ 4



ภาพที่ 4: หน้าแรกของเว็บแอปพลิเคชัน

เมื่อผู้ใช้ทำการกรอกประโยคและทำการกดปุ่มตรวจสอบ เว็บแอปพลิเคชันจะทำการวิเคราะห์ประโยคและแสดงผลลัพธ์ ดังภาพที่ 5



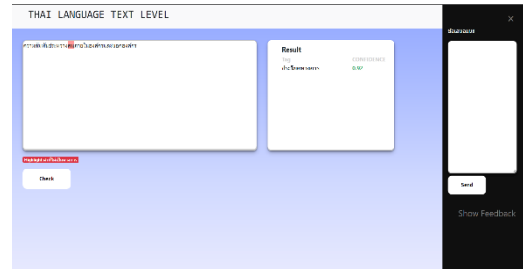
ภาพที่ 5: หน้าเว็บแอปพลิเคชันแสดงผลวิเคราะห์ประโยค

เมื่อผู้ใช้ทำการเลือกคำที่เป็นภาษาระดับไม่เป็นทางการจะแสดงคำแนะนำคำที่เป็นภาษาระดับทางการ ดังภาพที่ 6



ภาพที่ 6: การแนะนำประโยคภาษาระดับทางการ

หากผู้ใช้ทำการกดปุ่ม GIVE FEEDBACK ที่มุมบนขวาจะแสดงส่วนที่ให้กรอกข้อความเพื่อส่งข้อเสนอแนะให้ผู้พัฒนา ดังภาพที่ 7



ภาพที่ 7: ส่งข้อความเสนอแนะให้ผู้พัฒนา

#### 4. ผลการดำเนินงาน

จากการทดสอบแบบจำลองทั้ง 3 แบบจำลองโดยใช้ชุดข้อมูลในการฝึกฝนทั้งหมด 2,000 ประโยค เป็นประโยคภาษาระดับทางการ 1,000 ประโยค ประโยคภาษาระดับไม่เป็นทางการ 1,000 ประโยค

ตารางที่ 2: ผลการทดสอบ 10-fold Cross Validation ของ นาอ็ฟ เบย์ ดันไม้ตัดลิ้นใจ และการหาเพื่อนบ้านที่ใกล้ที่สุด

แบบจำลอง	Precision	Accuracy	Recall	F-Measure
นาอ็ฟ เบย์	75.63%	73.57%	0.713	0.724
ดันไม้ตัดลิ้นใจ	63.37%	68.33%	0.683	0.683
การหาเพื่อนบ้านที่ใกล้ที่สุด (k = 5)	58.37%	60.60%	0.665	0.634

จากผลการทดสอบ 10-fold Cross Validation ของแบบจำลอง ผลการจำแนกข้อมูลด้วย นาอ็ฟ เบย์ มีความถูกต้องอยู่ที่ร้อยละ 75.63 ซึ่งดีกว่าผลการจำแนกข้อมูลดันไม้ตัดลิ้นใจ มีความถูกต้องอยู่ที่ร้อยละ 63.37 เนื่องจาก นาอ็ฟ เบย์ วิเคราะห์จากความน่าจะเป็นของคำที่พบบ่อยในแต่ละคลาส แต่ดันไม้ตัดลิ้นใจวิเคราะห์จากคุณลักษณะยิ่งคุณลักษณะยิ่งเยอะ การตัดลิ้นใจของดันไม้ก็ยิ่งซับซ้อน ทำให้ความถูกต้องในการตัดลิ้นใจมีประสิทธิภาพที่น้อยลง และการหาเพื่อนบ้านที่ใกล้ที่สุดมีความถูกต้องอยู่ที่ร้อยละ 58.37 ซึ่งแบบจำลองการหาเพื่อนบ้านที่ใกล้ที่สุดมีความถูกต้องต่ำเพราะประโยคภาษาระดับทางการ และประโยคภาษาระดับไม่เป็นทางการที่ผ่านการเตรียมข้อมูลมีความแตกต่างกันไม่มาก

## 5. สรุปและข้อเสนอแนะ

จากการดำเนินโครงการ ทางผู้จัดทำได้เก็บรวบรวมชุดข้อมูลทดสอบแบบจำลองครบตามจำนวนที่ต้องการ และทำการฝึกแบบจำลองด้วยการเรียนรู้ของเครื่องโดยใช้เทคนิคทั้ง 3 โมเดลและทำการทดสอบด้วย 10-fold Cross Validation แล้วจึงนำมาเปรียบเทียบ โดยแบบจำลองที่ให้ค่าความถูกต้องที่ดีที่สุดคือแบบจำลองนาอิว เบย์ ให้ค่าความถูกต้องอยู่ที่ร้อยละ 75.63 จากการวิเคราะห์ผลเนื่องจากการหาค่าความน่าจะเป็นของแต่ละคำที่จะปรากฏในประโยค ซึ่งสามารถพิจารณาเพิ่มเติมได้ เช่นแนวทางการพัฒนาแบบจำลองโดยการเพิ่มประโยคที่จะใช้ในการฝึกแบบจำลองเพื่อให้มีความแม่นยำเพิ่มขึ้น

### เอกสารอ้างอิง

- [1] รัชย์พงษ์ ธรรมสุสนา. (2554). ระดับภาษาและการใช้ภาษาที่ถูกต้อง. สืบค้นจาก <http://gened.siam.edu/wp-content/uploads/2018/07/thaic-handout-03.pdf>
- [2] ราชวิทย์ทิพย์เสนา, ฉัตรเกล้า เจริญผล, แกมกาญจน์สมประเสริฐศรี. (2556). การจำแนกกลุ่มคำถามอัตโนมัติบนกระดานสนทนาโดยใช้เทคนิคเหมืองข้อความ. มหาวิทยาลัยมหาสารคาม. คณะวิทยาการสารสนเทศ
- [3] Avinash Navlani. (2018). KNN Classification usinScikit – learn. Retrieved from <https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>
- [4] เอกรัฐ บุญเชียง. data clustering and Data classification การแบ่งกลุ่มข้อมูลและการจำแนกประเภทข้อมูล [pdf file].
- [5] Fadi ABU SHEIKHA, Diana INKPEN. (2010). Automatic Classification of Documents by Formality. University of Ottawa, SITE University of Ottawa, SITE 800 King Edward, Ottawa, ON, Canada
- [6] เอกรัฐ บุญเชียง. (2561). data clustering and Data classification การแบ่งกลุ่มข้อมูลและการจำแนกประเภทข้อมูล มหาวิทยาลัยเชียงใหม่. คณะวิทยาศาสตร์. ภาควิชาวิทยาการคอมพิวเตอร์
- [7] อุมารินทร์ นอกตะแบก. (2555). ระดับภาษาและคำราชาศัพท์. สืบค้นจาก <https://sites.google.com/site/018schoolnet/bth-thi1-radab-phasa-laea-kha-rachasaphth/1>

# การเปรียบเทียบเทคนิคการวิเคราะห์อนุกรมเวลา สำหรับการพยากรณ์ค่าโฆษณาเฟซบุ๊ก Comparison of Time Series Analysis Techniques for Forecasting Facebook Ad Values

สุจรรยา ทับทอง(Sujanya Tubtong)<sup>1</sup> และวัชรวิวัฒน์ จิตต์สกุล(Watchareewan Jitsakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858139@emai.kmutnb.ac.th, <sup>2</sup>watchareewan.j@itd.kmutnb.ac.th

## บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อทำการศึกษาคัดเลือกแบบจำลองที่เหมาะสมในการวิเคราะห์การซื้อโฆษณาผ่านแพลตฟอร์มเฟซบุ๊กด้วยเทคนิคการวิเคราะห์อนุกรมเวลาจำนวน 3 วิธี ได้แก่ วิธีหาค่าเฉลี่ยเคลื่อนที่ วิธีเอ็กซ์โปเนนเชียล และวิธีเอ็กซ์โปเนนเชียลกำลังสอง ด้วยโปรแกรม Minitab ข้อมูลที่ใช้เป็นข้อมูลการซื้อโฆษณาจากแพลตฟอร์มเฟซบุ๊ก ในช่วงปี พ.ศ. 2560 จนถึงปี พ.ศ. 2564 รวมจำนวน 60 เดือน ทำการแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลสำหรับสร้างแบบจำลองพยากรณ์ จำนวน 48 เดือนและชุดข้อมูลสำหรับการทดสอบความถูกต้องของแบบจำลองพยากรณ์ จำนวน 12 เดือน ทำการเปรียบเทียบประสิทธิภาพด้วยค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ ค่าเบี่ยงเบนสัมบูรณ์เฉลี่ยที่ต่ำที่สุด

ผลการวิจัยพบว่า แบบจำลองเอ็กซ์โปเนนเชียลกำลังสอง ให้ประสิทธิภาพการพยากรณ์ที่ดีที่สุด เนื่องจากมีค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์(MAPE) ต่ำที่สุดเท่ากับร้อยละ 1.00 และค่าเบี่ยงเบนสัมบูรณ์เฉลี่ย(MAD) เท่ากับร้อยละ 3.26 ผลการทดสอบความถูกต้องของแบบจำลอง พบว่า แบบจำลองเอ็กซ์โปเนนเชียลกำลังสอง มีความคลาดเคลื่อนต่ำที่สุด โดยมีค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เท่ากับร้อยละ 1.00622

**คำสำคัญ:** เทคนิคการวิเคราะห์อนุกรมเวลา แบบจำลอง

ค่าเฉลี่ยเคลื่อนที่ แบบจำลองเอ็กซ์โปเนนเชียล

แบบจำลองเอ็กซ์โปเนนเชียลกำลังสอง

## Abstract

The purpose of this research is to select a suitable model for analyzing advertising purchases through the Facebook platform with three time series analysis. that are used by moving average, single exponential smoothing and double exponential smoothing by Minitab. The research uses the history of advertising purchase information from Facebook between 2017 and 2021 total of 60 months, spilt 2 datasets as 48 months are trainset and 12 month are validating set then compare performance for forecasting by lowest Mean absolute percentage error (MAPE) and Mean absolute deviation (MAD)

From the research results, double exponential smoothing gives the best forecasting efficiency. it was lowest Mean absolute percentage error is 1.00 percent and the Mean absolute deviation is 3.26 percent, The results of validating double exponential smoothing give lowest error Mean absolute percentage error is 1.00622 percent

**Keywords:** Time Series Analysis, Moving Average, Single Exponential Smoothing, Double Exponential Smoothing.



## 1. บทนำ

ปัจจุบันการซื้อโฆษณาผ่านช่องทางออนไลน์ได้รับความนิยมเป็นอย่างมากโดยเฉพาะแพลตฟอร์มของเฟซบุ๊ก ซึ่งสามารถตอบสนองความต้องการของผู้บริโภคได้ดี โดยเฟซบุ๊กมีฟังก์ชันในการซื้อโฆษณาที่สามารถตอบสนองความต้องการของผู้บริโภคได้อย่างง่ายดาย ผู้บริโภคสามารถซื้อโฆษณาได้อย่างรวดเร็วทันใจ เนื่องจากเครื่องมือการซื้อโฆษณาของเฟซบุ๊กเข้ามาช่วยวิเคราะห์ในเรื่องของกลุ่มเป้าหมาย เพื่อตอบสนองความต้องการและความสนใจในโฆษณาของผู้บริโภค

เนื่องจากซื้อโฆษณาผ่านเฟซบุ๊กเป็นที่นิยมอย่างมากในประเทศไทย จึงมีแนวคิดที่จะศึกษาทฤษฎีการให้บริการทางด้าน การซื้อโฆษณาของเฟซบุ๊ก ด้วยเทคนิค Time Series Analysis [1] จำนวน 3 วิธี ได้แก่ วิธี Moving Average [2] วิธี Single Exponential Smoothing และวิธี Double Exponential Smoothing [3] จากนั้นนำวิธีพยากรณ์ที่เหมาะสมกับข้อมูลมาเปรียบเทียบเพื่อหาวิธีการพยากรณ์ที่มีประสิทธิภาพสูงสุดกับข้อมูลการให้บริการทางด้าน การซื้อโฆษณาของเฟซบุ๊ก [4] เพื่อนำผลพยากรณ์มาประยุกต์ในการวางแผนบริหารจัดการระบบประมาณจากการซื้อโฆษณาผ่านเฟซบุ๊กได้อย่างมีประสิทธิภาพ ทั้งนี้เพื่อช่วยในเรื่องของการหาประสิทธิภาพในการซื้อโฆษณาผ่านเฟซบุ๊กเพื่อวิเคราะห์ต้นทุนและความพึงพอใจ [5] ถึงจุดคุ้มทุนที่ผู้ซื้อโฆษณาผ่านเฟซบุ๊กควรมีประสิทธิภาพมากน้อยเพียงใด

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้รวบรวมทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยแบ่งเป็น 2 หัวข้อหลัก ดังนี้

### 2.1 การพยากรณ์แบบอนุกรมเวลา

การพยากรณ์แบบอนุกรมเวลา เป็นเทคนิคการพยากรณ์ข้อมูลในอดีตเป็นหลัก ซึ่งเป็นเทคนิคที่ยอมรับโดยทั่วไป สามารถนำข้อมูลที่มีค่ามากมาคำนวณหาค่าพยากรณ์ได้ ลักษณะของอนุกรมเวลาจะต้องใช้ข้อมูลความต้องการหรือข้อมูลสถิติในอดีตที่ทำการรวบรวมในระยะเวลาหนึ่ง [6]

### 2.1.1. วิธีค่าเฉลี่ยเคลื่อนที่ (Moving Average)

วิธีค่าเฉลี่ยเคลื่อนที่ที่เหมาะสมสำหรับการพยากรณ์ในช่วงเวลาระยะสั้นๆ และข้อมูลที่มีลักษณะค่อนข้างแน่นอน เป็นเส้นตรง คงที่ และไม่เหมาะสมกับข้อมูลที่มีลักษณะเป็นแนวโน้ม ถูกลม หรือข้อมูลที่มีลักษณะเป็นการเปลี่ยนแปลงเป็นขั้นบันได มีหลักการในการหาค่าเฉลี่ย คือ ใช้ค่าเฉลี่ยจากการสังเกตข้อมูลในอดีตมาคำนวณหาค่าเฉลี่ยแล้วใช้ค่าเฉลี่ยที่ได้เป็นค่าพยากรณ์สำหรับช่วงเวลาถัดไป โดยมีสมการดังนี้ [7]

$$F_{t+1} = \frac{(Y_t + Y_{t-1} + Y_{t-2} + \dots + Y_{t-N+1})}{N} \quad (1)$$

โดย  $F_t$  = ค่าพยากรณ์ในช่วงเวลา  $t$   
 $t$  = เวลาที่ทำการพยากรณ์  
 $A_t$  = ค่าความต้องการในงวดที่  $t$   
 $N$  = จำนวนข้อมูลที่นำมาเฉลี่ย

### 2.1.2. วิธีปรับเรียบเอ็กซ์โปเนนเชียล (Single Exponential Smoothing)

เป็นวิธีที่เหมาะสมกับข้อมูลที่เคลื่อนไหวอยู่ในระดับคงที่ ไม่มีการเปลี่ยนแปลง หรือเป็นข้อมูลที่ไม่มีการประกอบของแนวโน้ม และไม่มีความผันแปรตามฤดูกาล มีเฉพาะความผันแปรเนื่องจากเหตุการณ์ที่ผิดปกติเพียงอย่างเดียว เหมาะกับการพยากรณ์ระยะสั้นสำหรับข้อมูลที่เหมาะสมที่จะใช้วิธีนี้ควรมีข้อมูลอย่างน้อย 5 ถึง 10 ข้อมูล ซึ่งวิธีการปรับเรียบแบบเอ็กซ์โปเนนเชียลให้น้ำหนักความสำคัญของข้อมูลในอดีต และข้อมูลที่ทำการพยากรณ์ซึ่งน้ำหนักที่ถ่วงให้กับค่าสังเกตแต่ละค่าจะมีค่าคงที่ของการปรับเรียบเรียกว่า  $\alpha$  โดยที่ค่าของ  $\alpha$  จะอยู่ในช่วงระหว่าง 0 ถึง 1)  $0 < \alpha < 1$ ) โดยมีสมการดังนี้ [8]

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t \quad (2)$$

โดย  $F_t$  = ค่าพยากรณ์ในช่วงเวลา  $t$   
 $F_{t-1}$  = ค่าพยากรณ์ในช่วงเวลา  $t-1$   
 $Y_{t-1}$  = ค่าจริงที่เกิดขึ้น ณ เวลา  $t$   
 $\alpha$  = มีค่าเท่ากับ  $0 < \alpha < 1$

### 2.1.3 วิธีปรับเรียบแบบเอ็กซ์โปเนนเชียลกำลังสอง (Double Exponential Smoothing)

เป็นเทคนิคที่เหมาะสมกับข้อมูลที่มีอิทธิพลของแนวโน้มเส้นตรงเป็นส่วนประกอบ โดยนำค่าของการพยากรณ์มาปรับซ้ำอีกครั้ง เพื่อพยายามลดปัจจัยอันเกิดจากการเปลี่ยนแปลงที่อธิบายไม่ได้ โดยมีสมการดังนี้ [9]

$$F_t = \alpha Y_t + (1 - \alpha)F_{t-1} \quad (3)$$

$$F'_t = \alpha F_t + (1 - \alpha)F'_{t-1} \quad (4)$$

โดย  $F_t$  = ค่าพยากรณ์ความต้องการจากวิธี SES

$Y_t$  = ค่าจริงที่เกิดขึ้น ณ เวลา t

$F'_t$  = ค่าพยากรณ์จากวิธีการ DES สำหรับเวลา t

## 2.2 งานวิจัยที่เกี่ยวข้อง

เมธินี ประสิทธิ์ และบุญอ้อม [10] [ทำการศึกษการพยากรณ์สำหรับดัชนีราคาหุ้นกลุ่มอสังหาริมทรัพย์ และก่อสร้างในประเทศไทยโดยการ Time Series Analysis ด้วยวิธี Single Exponential ผลการวิจัยพบว่า การพยากรณ์ล่วงหน้า 3 เดือน มีประสิทธิภาพสูงสุด ค่า MAPE ร้อยละ 1.9273 รองลงมา คือ การพยากรณ์ล่วงหน้า 2 เดือน ค่า MAPE ร้อยละ 2.2648 และการพยากรณ์ล่วงหน้า 6 เดือน ค่า MAPE ร้อยละ 3.4621 ตามลำดับ

อนุสรณ์ [11] [ทำการศึกษการพยากรณ์ความต้องการแวนตารันรักแวน โดยทำการเปรียบเทียบประสิทธิภาพด้วยเทคนิค Time Series Analysis จำนวน 4 วิธี ได้แก่ วิธี Moving Average วิธี Single Exponential วิธี Weighted และวิธี Simple Regression ผลการวิจัยพบว่า วิธี Simple Regression มีประสิทธิภาพสูงสุด ค่า MAPE เท่ากับ ร้อยละ 58 รองลงมาคือ วิธี Simple Exponential ค่า MAPE ร้อยละ 64.23 วิธี Moving Average ค่า MAPE ร้อยละ 68.02 และวิธี Weighted ค่า MAPE ร้อยละ 69.83 ตามลำดับ

สมฤดี กัญญา และอรวรรณ [12] ได้ทำการศึกษการพยากรณ์มูลค่าการส่งออกอัญมณีและเครื่องประดับของประเทศไทย ด้วยการวิเคราะห์อนุกรมเวลา 3 วิธี ได้แก่ วิธี Decomposition วิธี Exponential Smoothing และ

วิธีบอซเจนกินส์ ข้อมูลที่ใช้มาจากข้อมูลการส่งออกอัญมณี และเครื่องประดับของประเทศไทยรายเดือน ตั้งแต่เดือนมกราคม พ.ศ. 2556 ถึงเดือนธันวาคม พ.ศ. 2562 จำนวน 84 วัดประสิทธิภาพการพยากรณ์ด้วยค่า MAPE ผลการวิจัยพบว่า วิธี Decomposition ให้ประสิทธิภาพสูงที่สุด ค่า MAPE เท่ากับ 0.24 รองลงมา คือ Exponential Smoothing ค่า MAPE เท่ากับ 0.26 MAPE เท่ากับ 0.47 ตามลำดับ

พรฤดี และณัฐวิ [13] ได้ทำการศึกษการตัวแบบที่เหมาะสมสำหรับการพยากรณ์ยอดขายเสื้อผ้าด้วยการวิเคราะห์อนุกรมเวลา 2 วิธี ได้แก่ วิธี Decomposition และวิธี Linear Regression ข้อมูลที่ใช้มาจากยอดขายรายวัน ตั้งแต่วันที่ 1 มกราคม 2015 ถึงวันที่ 31 ธันวาคม 2017 จำนวน 1,096 ค่า วัดประสิทธิภาพการพยากรณ์ด้วยค่า Basic Error ค่า MAD ค่า MSE ค่า MAPE และค่า Theil's U ผลการวิจัยพบว่า วิธี Linear Regression ให้ประสิทธิภาพสูงสุด ค่า Basic Error เท่ากับ 24.441 ค่า MAD เท่ากับ 0.380 ค่า MSE เท่ากับ 0.266 ค่า MAPE เท่ากับ 7.850 และค่า Theil's U เท่ากับ 0.5555

## 3. วิธีการดำเนินการวิจัย

งานวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อศึกษาแบบจำลองการพยากรณ์ที่เหมาะสมสำหรับการพยากรณ์การซื้อโฆษณาผ่านทางเฟซบุ๊ก โดยใช้เทคนิคการวิเคราะห์อนุกรมเวลา โดยมีรายละเอียดและขั้นตอนการดำเนินงานดังนี้

### 3.1 การรวบรวมข้อมูลและการจัดการข้อมูล

งานวิจัยนี้มีการเก็บรวบรวมข้อมูลการซื้อโฆษณาจากแพลตฟอร์มเฟซบุ๊กในช่วงปี พ.ศ. 2560 ถึงปี พ.ศ.2564 จำนวน 60 เดือน โดยนำข้อมูลการซื้อโฆษณาจากเว็บไซต์ [www.facebook.com/adsmanager](http://www.facebook.com/adsmanager) ทำการจัดข้อมูลให้อยู่ในรูปแบบรายเดือน ด้วยโปรแกรม Microsoft Excel แบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ชุดข้อมูล Train Set สำหรับการสร้างแบบจำลองพยากรณ์ ใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2560 ถึงเดือนธันวาคม พ.ศ. 2564 จำนวน 48 เดือน และชุดข้อมูล Test Set ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2564 ถึงเดือนธันวาคม พ.ศ. 2564 จำนวน 12 เดือน

ใช้สำหรับการทวนสอบความถูกต้องของการพยากรณ์

**3.2 การสร้างแบบจำลองการพยากรณ์**

ผู้วิจัยได้ทำการเลือกใช้เทคนิค Time Series Analysis ในการสร้างแบบจำลองการพยากรณ์ จำนวน 3 วิธี ได้แก่ วิธี Moving Average โดยการกำหนดค่า MA Length 3,6,9 และ 12 วิธี Single Exponential Smoothing กำหนดค่าพารามิเตอร์  $\alpha$  (Alpha) เท่ากับ 0.2, 0.3, 0.6 และ 0.9 และวิธี Double Exponential Smoothing โดยการกำหนดค่าพารามิเตอร์  $\alpha$  เท่ากับ 0.1 ถึง 0.9 และค่าพารามิเตอร์  $\gamma$  (Gamma) เท่ากับ 0.1 ถึง 0.9

**3.3 การเปรียบเทียบประสิทธิภาพและคัดเลือกแบบจำลองการพยากรณ์**

ผู้วิจัยพิจารณาประสิทธิภาพของแบบจำลองการพยากรณ์ทั้ง 3 วิธีด้วยค่า ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ ค่าเบี่ยงเบนสัมบูรณ์เฉลี่ยที่ต่ำที่สุด

**3.4 การทวนสอบความถูกต้องของแบบจำลองพยากรณ์**

ในส่วนนี้เป็นการทวนสอบความถูกต้องโดยประเมินความถูกต้องของการพยากรณ์ของแบบจำลองด้วยข้อมูลการซื้อโฆษณาจากแพลตฟอร์มเฟสบุ๊กจำนวน 12 เดือน และประเมินค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

**4. ผลการดำเนินงาน**

ผลการดำเนินงานแบ่งออกเป็น 3 หัวข้อ ได้แก่ ผลการเปรียบเทียบประสิทธิภาพแบบจำลอง ผลการทวนสอบความถูกต้องของแบบจำลอง และการนำเสนอข้อมูล แสดงรายละเอียดได้ ดังนี้

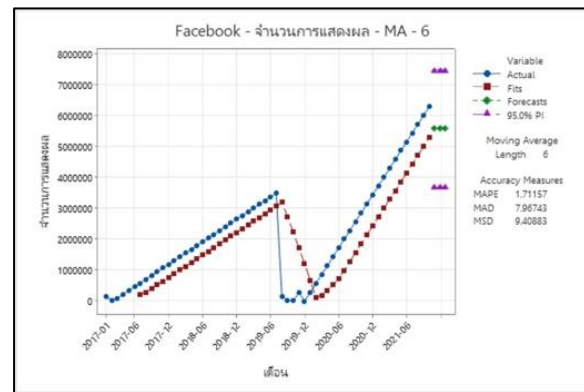
**4.1 ผลการเปรียบเทียบประสิทธิภาพแบบจำลอง**

ผลการเปรียบเทียบประสิทธิภาพแบบจำลองเทคนิค Time Series Analysis จำนวน 3 วิธี ได้แก่ วิธี Moving Average วิธี Single Exponential Smoothing และวิธี Double Exponential Smoothing ด้วยโปรแกรม Minitab แสดงดังตารางที่ 1 ถึงตารางที่ 3

**ตารางที่ 1:** การวิเคราะห์ด้วยวิธี Moving Average

วิธีพยากรณ์	ค่าพารามิเตอร์	MAPE	MAD
Moving Average	MA Length 3	6.43	4.68
	MA Length 6	1.71	7.96
	MA Length 9	2.30	1.07
	MA Length 12	2.61	1.29

จากตารางที่ 1 เมื่อพิจารณาค่าความคลาดเคลื่อนพบว่า MA Length 6 ให้ประสิทธิภาพที่ดีที่สุด โดยมีค่า MAPE เท่ากับ 1.71 และค่า MAD เท่ากับ 7.96 ซึ่งมีกราฟการพยากรณ์ แสดงดังภาพที่ 1

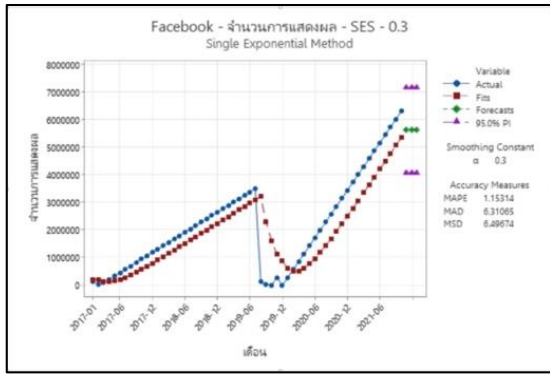


**ภาพที่ 1:** กราฟการพยากรณ์แบบจำลอง Moving Average

**ตารางที่ 2:** การวิเคราะห์ Single Exponential Smoothing

วิธีพยากรณ์	ค่าพารามิเตอร์	MAPE	MAD
Single Exponential Smoothing	$\alpha$ (Alpha) = 0.3	1.15	6.31
	$\alpha$ (Alpha) = 0.6	4.73	3.68
	$\alpha$ (Alpha) = 0.9	2.54	2.54
	$\alpha$ (Alpha) = 0.12	1.70	1.03

จากตารางที่ 2 เมื่อพิจารณาค่าความคลาดเคลื่อนพบว่า ค่า  $\alpha$  เท่ากับ 0.3 โดยมีค่า MAPE เท่ากับ 1.15 และค่า MAD เท่ากับ 6.31 ซึ่งมีกราฟการพยากรณ์ แสดงดังภาพที่ 2

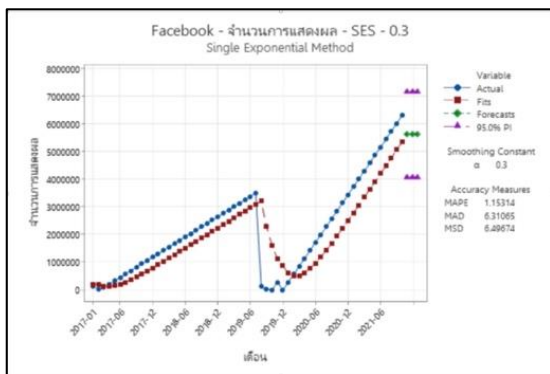


ภาพที่ 2: กราฟการพยากรณ์แบบจำลอง Single Exponential Smoothing

ตารางที่ 3: การวิเคราะห์ Double Exponential Smoothing

วิธีพยากรณ์	ค่าพารามิเตอร์	MAPE	MAD
Double Exponential Smoothing	$\alpha = 0.1, \gamma = 0.9$	1.21	7.58
	$\alpha = 0.2, \gamma = 0.1$	1.44	5.41
	$\alpha = 0.3, \gamma = 0.6$	1.00	3.26
	$\alpha = 0.4, \gamma = 0.9$	1.15	2.23

จากตารางที่ 3 เมื่อพิจารณาค่าความคลาดเคลื่อนพบว่าค่า  $\alpha$  เท่ากับ 0.3 และค่า  $\gamma$  เท่ากับ 0.6 โดยมีค่า MAPE เท่ากับ 1.00 และค่า MAD เท่ากับ 3.26 ซึ่งมีกราฟการพยากรณ์ แสดงดังภาพที่ 3



ภาพที่ 3: กราฟการพยากรณ์แบบจำลอง Double Exponential Smoothing

4.2 ผลการทวนสอบความถูกต้องของแบบจำลอง

ผลการทวนสอบความถูกต้องของแบบจำลอง เมื่อพิจารณาค่าความคลาดเคลื่อน วิธี Double Exponential Smoothing ให้ประสิทธิภาพดีที่สุด รองลงมาคือ วิธี Single Exponential Smoothing และวิธี Moving Average

ตามลำดับ นำแบบจำลองที่ได้มาทำการทวนสอบความถูกต้องกับข้อมูลการซื้อโฆษณาจากแพลตฟอร์มเฟซบุ๊กจำนวน 12 เดือน พบว่าวิธี Double Exponential Smoothing มีความคลาดเคลื่อนต่ำที่สุดเท่ากับ ร้อยละ 1.00 แสดงดังตารางที่ 4

ตารางที่ 4: ผลการทวนสอบความถูกต้องของแบบจำลอง

เดือน	จำนวนการแสดงผล	Double Exponential Smoothing		
		Test Set	Smooth	Predict
2021-01	3148265	3148265	3148265	8.72804
2021-02	3434759	3434759	3434759	3.40124
2021-03	3721253	3721253	3721253	3.98769
2021-04	4007747	4007747	4007747	4.28089
2021-05	4294241	4294241	4294241	3.71557
2021-06	4580735	4580735	4580735	2.65104
2021-07	4867229	4867229	4867229	1.42869
2021-08	5153723	5153723	5153723	2.98125
2021-09	5440217	5440217	5440217	2.19874
2021-10	5726711	5726711	5726711	2.669347
2021-11	6013205	6013205	6013205	2.91442
2021-12	6299699	6299699	6299699	2.79494
ร้อยละ				1.00622

4.3 การนำเสนอข้อมูล

ทางผู้วิจัยได้นำผลการพยากรณ์โดยใช้วิธี Double Exponential Smoothing เป็นตัวแบบในการพยากรณ์ค่าโฆษณาจากแพลตฟอร์มเฟซบุ๊ก นำเสนอข้อมูลด้วยรูปแบบของเว็บแอปพลิเคชัน แสดงดังภาพที่ 4



ภาพที่ 4: ภาพการแสดงผลการพยากรณ์ค่าโฆษณา

## 5. สรุป

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อทำการศึกษาและคัดเลือกแบบจำลองการพยากรณ์ที่เหมาะสมกับข้อมูลการซื้อโฆษณาจากแพลตฟอร์มเฟสบุ๊ก จำนวน 60 เดือน โดยใช้เทคนิค Time Series Analysis จำนวน 3 วิธี ได้แก่ วิธี Moving Average วิธี Single Exponential Smoothing และวิธี Double Exponential Smoothing ทำการทดสอบประสิทธิภาพของแบบจำลองการพยากรณ์ด้วยค่า MAPE และค่า MAD ผลการวิจัยพบว่า วิธี Double Exponential Smoothing มีความเหมาะสมสำหรับการพยากรณ์ที่สุด โดยมีค่า MAPE เท่ากับร้อยละ 1.00 และค่า MAD เท่ากับ 3.26 หลังจากนั้นนำข้อมูลมาทวนสอบกับข้อมูลจริงจำนวน 12 เดือน สรุปได้ว่า การพยากรณ์ค่าโฆษณาจากแพลตฟอร์มเฟสบุ๊ก ด้วยแบบจำลองการพยากรณ์วิธี Double Exponential Smoothing มีความแม่นยำ โดยมีค่าความคลาดเคลื่อนในการพยากรณ์ เท่ากับร้อยละ 1.00622

### เอกสารอ้างอิง

- [1] เมธินี กวินภาส, ประสิทธิ์ พัทฒพงษ์ และบุญอ้อม โจมที. “การเปรียบเทียบการพยากรณ์ระหว่างวิธีเครือข่ายเบย์เซียน และวิธีการปรับให้เรียบแบบเอ็กซ์โพเนนเชียล สำหรับดัชนีราคาหุ้นกลุ่มอสังหาริมทรัพย์และก่อสร้างในประเทศไทย.” *วารสารวิทยาศาสตร์และเทคโนโลยี*. ปีที่ 23 ฉบับที่ 2 พ.ศ. 2559 หน้า 203-211.
- [2] อภิชัย พรหมอ่อน, การศึกษาการพยากรณ์แบบอนุกรมเวลา (Time Series) เพื่อการวางแผนการสั่งซื้อวัตถุดิบ กรณีศึกษา บริษัทผลิตชิ้นส่วนต่อวงจรรถยนต์. สารนิพนธ์ปริญญาบริหารธุรกิจมหาบัณฑิต บัณฑิตวิทยาลัย สาขาวิชาการจัดการอุตสาหกรรม สถาบันเทคโนโลยีไทย-ญี่ปุ่น.
- [3] ประภัสสร คำสวัสดิ์. การพยากรณ์ปริมาณรถยนต์จดทะเบียนใหม่ในจังหวัดชลบุรี. สารนิพนธ์บัณฑิตวิทยาลัย คณะบริหารธุรกิจ มหาวิทยาลัยศรีปทุม วิทยาเขตชลบุรี
- [4] กนกกาญจน์ มูลผา และ เรืองศักดิ์ แก้วธรรมชัย. “ศึกษาเทคนิคการพยากรณ์ยอดขายสินค้าอุปโภคบริโภคที่เหมาะสมของบริษัทเอกชนแห่งหนึ่ง.” *วารสารวิชาการบริหารธุรกิจ สมาคมสถาบันอุดมศึกษาเอกชนแห่งประเทศไทย*. ปีที่ 3 ฉบับที่ 1 พ.ศ. 2560 หน้า 12-21.
- [5] วรางคณา เรียงสุทธิ์, “ตัวแบบพยากรณ์ราคามังคุดกละ” *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี* ปีที่ 19 ฉบับที่ 2 พฤษภาคม-สิงหาคม พ.ศ. 2560 หน้า 31-42.
- [6] จินตพร หนี่วอินปิ่น, บุญอ้อม โจมที และประสิทธิ์ พัทฒพงษ์. (2560).การเปรียบเทียบวิธีการพยากรณ์ 4 วิธี สำหรับความต้องการพลังงานไฟฟ้าสูงสุดในภาคกลางของประเทศไทย. *ปริญญาญานิพนธ์อุตสาหกรรมศาสตร์บัณฑิต สาขาวิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์*.
- [7] นิตินัย รุ่งจินดารัตน์ และ ศรัณย์ ทัศนาวุฒิน, “การพยากรณ์อนุกรมเวลาด้วยวิธีแบบคลาสสิก การส่งออกข้าวหอมมะลิของไทย” *วารสารวิทยาลัยดุสิตธานี* ปีที่ 13 ฉบับที่ 2 พฤษภาคม-สิงหาคม พ.ศ. 2562 หน้า 283-293.
- [8] อภิชัย พรหมอ่อน, การศึกษาการพยากรณ์แบบอนุกรมเวลา (Time Series) เพื่อการวางแผนการสั่งซื้อวัตถุดิบ กรณีศึกษา บริษัทผลิตชิ้นส่วนต่อวงจรรถยนต์. สารนิพนธ์ปริญญาบริหารธุรกิจมหาบัณฑิต บัณฑิตวิทยาลัย สาขาวิชาการจัดการอุตสาหกรรม สถาบันเทคโนโลยีไทย-ญี่ปุ่น.
- [9] สุคนันทา ร่วมประเสริฐ, “การพยากรณ์แนวโน้มมูลค่าการส่งออกน้ำสับประคของไทย” *วารสารวิชาการบริหารธุรกิจ* ปีที่ 2 ฉบับที่ 1 มกราคม – มิถุนายน 2556 หน้า 16-24.
- [10] เมธินี กวินภาส, ประสิทธิ์ พัทฒพงษ์ และบุญอ้อม โจมที. “การเปรียบเทียบการพยากรณ์ระหว่างวิธีเครือข่ายเบย์เซียน และวิธีการปรับให้เรียบแบบเอ็กซ์โพเนนเชียล สำหรับดัชนีราคาหุ้นกลุ่มอสังหาริมทรัพย์และก่อสร้างในประเทศไทย.” *วารสารวิทยาศาสตร์และเทคโนโลยี*. ปีที่ 23 ฉบับที่ 2 พ.ศ. 2559 หน้า 203-211.
- [11] อนุสรณ์ บุญสง่า, การพยากรณ์ความต้องการแวนดา กรณีศึกษาร้านรักแวน. การค้นคว้าอิสระวิทยาศาสตร์มหาบัณฑิต สาขาการจัดการโลจิสติกส์และโซ่อุปทาน วิทยาลัยโลจิสติกส์และซัพพลายเชน มหาวิทยาลัยศรีปทุม.
- [12] สมฤดี พงษ์เสนา กัญญา บวรโชคชัย และ อรวรรณ รุ่งทอง “ตัวแบบการพยากรณ์มูลค่าการส่งออกอัญมณีและเครื่องประดับของประเทศไทย” *Journal of Business Management, Burapha University*, Vol.9 No. 2 (July – December 2021) หน้า 66-85.
- [13] พรฤดี เนติโสภาคกุล และ ณัฐวิช สุภายา, “การวิเคราะห์ตัวแบบที่เหมาะสมสำหรับการพยากรณ์ยอดขายเสื้อผ้า” *วารสารบริหารธุรกิจเทคโนโลยีมหานคร* ปีที่ 16 ฉบับที่ 2 (กรกฎาคม-ธันวาคม) พ.ศ. 2562 หน้า 77-96.

# การวิเคราะห์ข้อมูลการขายผลิตภัณฑ์ให้แสงสว่างโดยใช้เทคนิคการจัดกลุ่ม

## Sales Data Analysis of Lighting Products using Clustering Technique

สุวพัชร จันทร์ไมตรี (Suwapat Janmaitree)<sup>1</sup> และศักดิ์ชาย ตั้งวรรณวิทย์ (Sakchai Tangwannawit)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>6307021858171@email.kmutnb.ac.th, <sup>2</sup>sakchai.t@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลการขายผลิตภัณฑ์ให้แสงสว่าง ด้วยวิธีการจัดกลุ่ม (K-means Clustering) ซึ่งข้อมูลที่น่ามาวิเคราะห์ได้แก่ รหัสสาขา ข้อมูลการขาย ราคาขายสินค้า และรหัสสินค้า โดยใช้ข้อมูลการขายสินค้าจำนวน 110,258 รายการ ตั้งแต่วันที่ 1 มกราคม 2562 – 31 ตุลาคม 2564 โดยวิธีวิจัยดำเนินการตามรูปแบบ CRISP-DM และนำไปวิเคราะห์ด้วยโปรแกรม RapidMiner ผลการวิจัย พบว่าสามารถแบ่งกลุ่มสินค้าได้ จำนวน 5 กลุ่ม โดยแต่ละกลุ่มมีส่วนของการขายตามรหัสสินค้า ที่แตกต่างกันอย่างชัดเจน ลำดับที่ 1 มีรหัสสินค้าที่เข้าเงื่อนไขสูงถึง 79 % นอกจากนี้จากผลการศึกษายังพบประเด็นที่น่าสนใจคือสามารถเพิ่มประสิทธิภาพในการบริหารจัดการสินค้า ซึ่งเห็นได้ว่าในแต่ละกลุ่มไม่จำเป็นต้องมีสินค้าขายเหมือนกันทุกสาขา ตัวอย่างเช่นสินค้าในกลุ่มสุดท้ายมีการขายเพียง 6 สาขาจากทั้งหมด 15 สาขา หลังจากสรุปข้อมูลได้และนำไปเทียบกับปริมาณการสั่งซื้อจากบริษัทผู้ค้าพบว่า ในกลุ่มนี้มีสัดส่วนการขายเพียง 24 % จากการสั่งซื้อทั้งหมด หากบริษัทนำข้อมูลการจัดกลุ่มสินค้ามาบริหารจัดการทำให้ลดปริมาณการสั่งซื้อได้สูงถึง 27 ล้านบาท

**คำสำคัญ:** เทคนิคการจัดกลุ่ม เคมีน ผลิตภัณฑ์ให้แสงสว่าง

### Abstract

This research aims to the sales data analysis of lighting products by K-means Clustering technique. The data consist of branches code, sales transactions, sales price, and product code. The

gathering of sales data at 110,258 items from January 1, 2019, to October 31, 2021. The research methodology according CRISP-DM model and to analyze with the RapidMiner program. The result of this research found that the sales data was to optimized cluster into 5 groups. Each group is different in sales according to the product code. The first cluster has product codes that meet the conditions as high as 79%. In addition, the results of this research study also found interesting issues that can increase efficiency in product management. It can be found that each group does not need to have the same products for sale in every branch. For example, the last clustering group of the same products found six of fifteen branches. The comparison of summary of sale volume from vender found that the last clustering group of sales is only 24% of all orders. If the company uses the product grouping information to manage, it can reduce the purchase order up to 27 million baht.

**Keywords:** Clustering technique, K-Means, Lighting product.

### 1. บทนำ

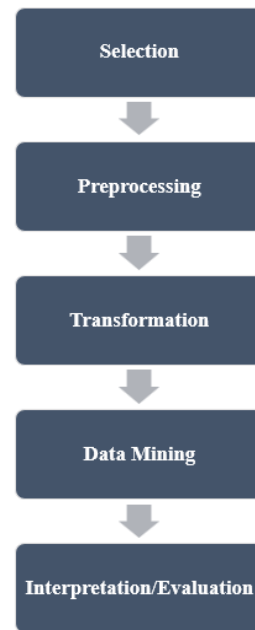
การบริหารจัดการของผู้บริหารในสถานการณ์เศรษฐกิจระหว่างปี พ.ศ. 2563 – 2564 ปัญหาหลักสำคัญที่ทุกบริษัทต้องเผชิญคือ สถานการณ์โควิด การปรับตัวให้เข้ากับสิ่งที่เกิดขึ้น ถือเป็นบทเรียนครั้งใหม่ที่ทุกบริษัทต้องปรับตัวหรือที่เรียกกันว่า ความปกติใหม่ (New Normal) การวิเคราะห์ข้อมูลของบริษัทที่เป็นข้อมูลในอดีตสามารถ

ช่วยให้บริษัท วางแผนกลยุทธ์ กำหนดนโยบาย ไว้รองรับ ปัญหาที่เข้ามาได้ทั้งระยะสั้น ระยะกลาง และระยะยาว

บริษัทที่ผู้วิจัยศึกษาอยู่พบปัญหาเรื่องการคุมค่าใช้จ่าย โดยที่ผ่านมา บริษัทไม่ได้สังเกตเห็นถึงความสำคัญในส่วนนี้ เพราะด้วยขาดทรัพยากรบุคคลที่มีความรู้ ความสามารถ และการเข้าถึงการวิเคราะห์ข้อมูลขั้นสูงยังไม่เป็นที่แพร่หลายในบริษัท ผู้วิจัยจึงร่วมศึกษาข้อมูลของบริษัท และพบว่าปัญหาหนึ่งที่บริษัทมองข้ามคือการเบิกสินค้าไป ยังแต่ละสาขาโดยไม่มีการควบคุมที่รัดกุมพอ เพราะในแต่ละสาขามีการจัดวางสินค้าบนชั้นวางสินค้าที่ขนาดเท่ากัน และไม่มีการควบคุมการจัดวางสินค้า จึงทำให้ปัญหาด้านทุนของสินค้าที่มีการเบิกไปในแต่ละสาขาเป็นต้นทุนที่บริษัทต้องรับรู้เกินความจำเป็น และไม่ได้คำนึงปริมาณการขายของแต่ละสาขา จึงเป็นที่มาของการนำข้อมูลการขายสินค้าย้อนหลังมาวิเคราะห์ด้วยการจัดกลุ่มประเภทสินค้าที่เหมาะสมและสามารถจัดวางสินค้าในแต่ละสาขาได้ และอย่างเหมาะสม ช่วยลดค่าใช้จ่าย และเพิ่มประสิทธิภาพในการบริหารเงินสดให้กับบริษัทได้

## 2. วรรณกรรมที่เกี่ยวข้อง

ทฤษฎีเกี่ยวกับ Data Mining [1] เป็นเทคนิคหนึ่ง สำหรับการวิเคราะห์ข้อมูล มาจากคำว่าเหมืองข้อมูล โดยเป็นการค้นคว้าหาข้อมูลจากฐานข้อมูลขนาดใหญ่ ซึ่งส่วนใหญ่เป็นกระบวนการสกัดข้อมูลจากในอดีต เพื่อนำผลลัพธ์ที่ได้มาช่วยในการวิเคราะห์และประกอบการตัดสินใจ จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายได้มีการปรับรูปแบบการจัดเก็บข้อมูลสารสนเทศ [2] เพื่อให้สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล โดยมีขั้นตอนการทำเหมืองข้อมูล [3] [6] ได้แก่ การคัดเลือกข้อมูล (Selection) การเตรียมข้อมูล (Preprocessing) การแปลงข้อมูล (Transformation) การวิเคราะห์และค้นรูปแบบข้อมูล (Data Mining) และการแปล/ประเมินผลการวิเคราะห์ข้อมูล (Interpretation/Evaluation)



ภาพที่ 1: ขั้นตอนการทำเหมืองข้อมูล

การทำเหมืองข้อมูลได้รับความนิยมและนำไปใช้งานในหลายด้าน [4] การทำเหมืองข้อมูลมีเทคนิคการค้นคว้าความรู้ได้หลายรูปแบบ โดยสามารถเลือกใช้ได้ตามลักษณะของปัญหาที่ต้องการนำการเหมืองข้อมูลไปใช้ในการแก้ไข [5] , [7] , [8] , [10] เช่นการจำแนกข้อมูลของข้อมูล (Classification) , การค้นหาคำความสัมพันธ์ของข้อมูล (Association rule) หรือ การแบ่งข้อมูล (Clustering) ประกอบกับการนำเทคนิคมาใช้ร่วมกันดังนี้

1. การวิเคราะห์ความสัมพันธ์ (Association Analysis) การนำกฎความสัมพันธ์ ซึ่งแสดงความสัมพันธ์ที่มีเงื่อนไขตรงกับข้อกำหนด และลักษณะของข้อมูลที่มีการเรียนรู้ของตระกร้าจ่ายตลาด (Market Basket) มาวิเคราะห์ความสัมพันธ์กฎ

2. เทคนิคเหมืองข้อมูลเพื่อค้นหากฎความสัมพันธ์ เป็นการค้นหากฎข้อมูลที่ปรากฏข้อมูลร่วมกันบ่อย

3. เทคนิคการวิเคราะห์เพื่อจัดกลุ่ม (Clustering Analysis) การวิเคราะห์เพื่อจัดกลุ่มจะแตกต่างกับการแบ่งประเภทและการทำนายซึ่งวิเคราะห์ข้อมูลที่มีความคล้ายกันมาก

ประเภทการพยากรณ์มี 2 ประเภทคือ

1.พยากรณ์เชิงคุณภาพ [12] คือพยากรณ์ที่ไม่อาศัยข้อมูลในอดีตแต่อาศัยประสบการณ์ของผู้บริหารหรือผู้มีหน้าที่เกี่ยวข้องเพื่อเปลี่ยนรูปแบบในขั้นพื้นฐาน

2.พยากรณ์เชิงปริมาณ [13] คือการพยากรณ์ที่ใช้หลักสถิติจากข้อมูลในอดีตเพื่อพิจารณารูปแบบและวางแผนกลยุทธ์ในอนาคต

### 3. วิธีการดำเนินการวิจัย

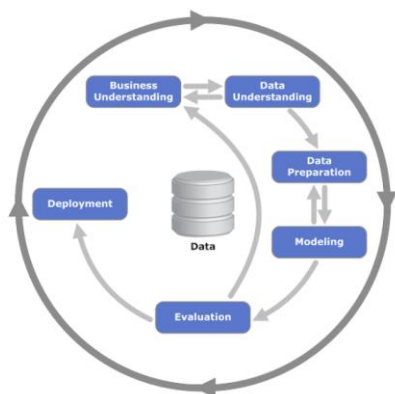
#### 3.1 เครื่องมือการวิจัย

3.1.1 โปรแกรมวิเคราะห์ข้อมูล RapidMiner

3.1.2 โปรแกรม Power BI Desktop

#### 3.2 ขั้นตอนการดำเนินการวิจัย

ผู้วิจัยได้เลือกวิธีการทำวิจัยโดยนำทฤษฎีเดิมมาประยุกต์ใช้เพื่อให้บริษัทสามารถนำวิธีการทำมาให้ได้ หลังจากการวิจัยสิ้นสุดลง เพื่อก่อให้เกิดประโยชน์สูงสุด และนำไปใช้งานได้จริง ผู้วิจัยจึงได้นำกระบวนการสำหรับการทำเหมืองข้อมูลที่มีขั้นตอนที่ไม่ซับซ้อนโดยประกอบไปด้วยกันทั้งสิ้น 6 ขั้นตอนดังนี้



ภาพที่ 2: Cross-industry standard process for data mining (CRISP-DM)

3.2.1 การทำความเข้าใจธุรกิจ (Business Understanding) เพื่อให้บรรลุวัตถุประสงค์ของการวิจัย ผู้วิจัยได้ศึกษาและทำความเข้าใจเกี่ยวกับบริษัท จากพนักงานของบริษัท โดยตรวจสอบกระบวนการทำงาน นโยบาย ต่างๆ เพื่อนำมาวิเคราะห์หาที่มาของปัญหาและนำไปสู่การวางแผนการดำเนินงานในเบื้องต้น

3.2.2 การทำความเข้าใจข้อมูล (Data Understanding) ผู้วิจัยได้ศึกษาข้อมูลในส่วนของคุณค่าของยอดขายบริษัท และต้นทุนค่าใช้จ่ายต่างๆ ของบริษัท ซึ่งผู้วิจัยได้สังเกตเห็นว่าที่มาของปัญหาคือกระบวนการในการเบิกของที่ไม่มีความชัดเจน ขาดการวางแผนการดำเนินงาน โดยไม่มีหลักการที่สอดคล้องกับการขายสินค้าจึงทำให้บริษัทรับภาระค่าใช้จ่ายเกินความจำเป็น

3.2.3 การเตรียมข้อมูล เริ่มจากการเก็บข้อมูลการขายสินค้าย้อนหลังทุกสาขา โดยในขั้นตอนแรก ผู้วิจัยได้นำข้อมูลสินค้าตั้งแต่ปี พ.ศ. 2562-2564 รวมทั้งสิ้น 110,258 รายการมาวิเคราะห์ และจัดลำดับยอดขายที่สูงสุด 3 ลำดับแรก และเลือกลำดับที่ 1 มาเพื่อใช้สำหรับการวิเคราะห์ในโปรแกรม RapidMiner โดยข้อมูลในลำดับที่ 1 แบ่งเป็นสินค้าประเภทย่อย จำนวน 1,431 รายการ ข้อมูลประกอบด้วย แอตทริบิวต์ ดังตารางที่ 1

ตารางที่ 1: ตารางอธิบายแอตทริบิวต์ทั้งหมด

ชื่อแอตทริบิวต์	คำอธิบาย
Date	วันที่ขายสินค้า
Month	เดือนที่ขายสินค้า
Year	ปีที่ขายสินค้า
Article Code	รหัสสินค้า
Article Name	คำอธิบายรหัสสินค้า
Watt	ข้อมูลแสดงความจุของสินค้า
Brand	ข้อมูลแสดงยี่ห้อสินค้า
Type	ข้อมูลแสดงประเภท
Sub Type	ข้อมูลแสดงประเภทย่อย
Sales Office	สาขาที่ขาย
Location	รหัสสาขาที่ขาย
Sales Qty	จำนวนสินค้าที่ขาย
Net Sales Inc Tax	มูลค่าสินค้าที่ขาย



จากข้อมูลตามตารางผู้วิจัยได้ตรวจสอบข้อมูลพบว่ามีลักษณะของข้อมูลที่ตรวจสอบเป็นรายการขายสินค้าตามใบรายการขาย และได้เตรียมข้อมูลโดยการนำ Pivot บนระบบ MS Excel เพื่อจัดข้อมูลให้ข้อมูลเป็นจำนวนการขายตาม รหัสสินค้าก่อนนำทุกรายการเข้าโปรแกรม RapidMiner Studio

3.2.4 การสร้างโมเดล (Modeling) ผู้วิจัยได้ประยุกต์ใช้ให้เข้ากับลักษณะของข้อมูลโดยได้เลือก การแบ่งกลุ่มข้อมูลแบบเคมีน(K-means Clustering) [9], [11] ซึ่งการใช้โมเดลนี้จะเป็นลักษณะของการจัดกลุ่มที่เป็นการกำหนดจำนวนกลุ่มก่อนนำไปวิเคราะห์แบบไม่เป็นขั้นตอน เมื่อนำข้อมูลไปสร้างโมเดลผ่าน ระบบ RapidMiner ตามภาพที่ 2 และภาพที่ 3 แสดงให้เห็นว่าก่อนการวิเคราะห์หาค่า K ผู้วิจัยได้จัดการข้อมูลตามลำดับดังนี้

3.2.4.1. การจัดการ Missing Data ให้มีค่าเป็น ศูนย์ เนื่องจากข้อมูลที่ต้องใช้วิเคราะห์เป็นการวิเคราะห์ข้อมูลด้วยลักษณะของตัวเลข

3.2.4.2. เลือกแอตทริบิวต์ จากตารางที่ 1 โดยมีแอตทริบิวต์ที่ใช้ดังตารางที่ 2

ตารางที่ 2: ตารางอธิบายแอตทริบิวต์ที่นำมาวิเคราะห์

ชื่อแอตทริบิวต์	คำอธิบาย
Article Code	รหัสสินค้า
Location	รหัสสาขาที่ขาย
Sales Qty	จำนวนสินค้าที่ขาย
Net Sales Inc Tax	มูลค่าสินค้าที่ขาย

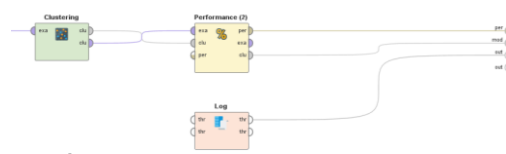
3.2.4.3. กำหนดค่า แอตทริบิวต์ชื่อ “Article Code” ให้มีค่าเป็น ID

3.2.4.4. กำหนดค่าพารามิเตอร์ในการทำคลัสเตอร์ริงให้ดีที่สุด (Customize) โดยกำหนดค่า K = 10, Max runs = 1,000, Measure types = Numerical Measures, Max Optimization steps = 1,000

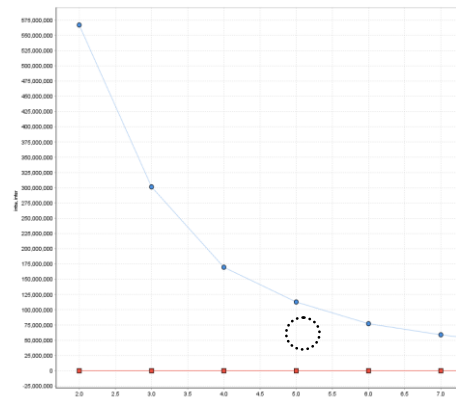
3.2.4.5. ระบุค่า Main Criterion เท่ากับ Avg. within centroid distance ดังภาพที่ 2 และภาพที่ 3



ภาพที่ 2: แสดงวิธีประมวลผลบนระบบ Rapid Miner



ภาพที่ 3: แสดงวิธีประมวลผลบนระบบ Rapid Miner



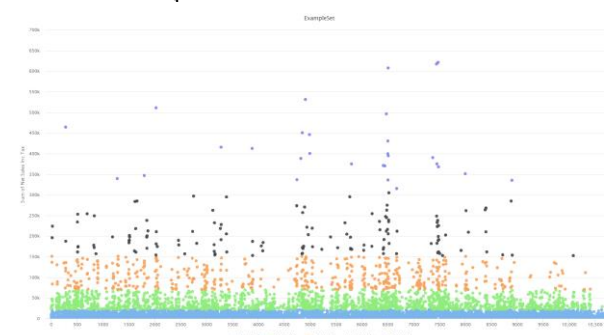
ภาพที่ 4: กราฟแสดงผลเพื่อระบุค่า K

จากข้อมูลตามภาพที่ 4 แสดงลักษณะของกราฟหลังจากประมวลผลบนระบบ Rapid Miner เพื่อหาจำนวนคลัสเตอร์ที่เหมาะสมที่สุด โดยเลือกค่า K ที่จุดก่อนที่ความเฉื่อย จะเริ่มลดลงเป็นเส้นตรง และจากกราฟสามารถระบุค่า K = 5 และนำค่า K ที่ได้จากการประมวลผลมาวิเคราะห์ ต่อเพื่อจัดกลุ่มตามภาพที่ 5



ภาพที่ 5: แสดงวิธีการประมวลผลหลังระบุค่า K

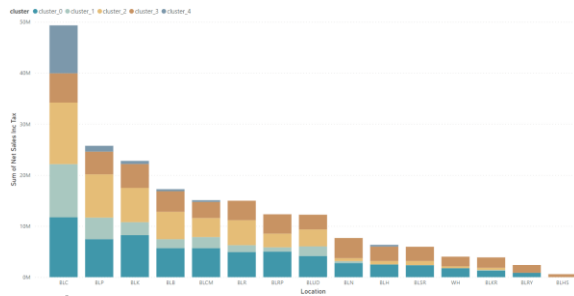
หลังจากระบบได้ประมวลผลเรียบร้อยแล้ว พบการจัดกลุ่มของข้อมูลที่มีลักษณะแยกกลุ่มได้อย่างชัดเจน โดยเห็นได้ชัดว่าในแต่ละกลุ่มมีลักษณะที่แตกต่างกันตามภาพที่ 6



ภาพที่ 6: แสดงผลการจัดกลุ่ม

### 3.2.5 การวัดผลประสิทธิภาพของโมเดล (Evaluation)

ในขั้นตอนที่ 4 ของการสร้างโมเดล ผู้วิจัยได้นำผลของการทำโมเดลวิเคราะห์ปริมาณสินค้าที่เพียงพอต่อการขายของแต่ละสินค้า ซึ่งเมื่อนำข้อมูลที่ผ่านมาการจัดกลุ่มมาทำเข้า โปรแกรม Power BI เพื่อแสดงให้เห็นสินค้าประเภทนี้ว่าในแต่ละสาขามีคลัสเตอร์ใดบ้างที่ขาย ดังภาพที่ 7



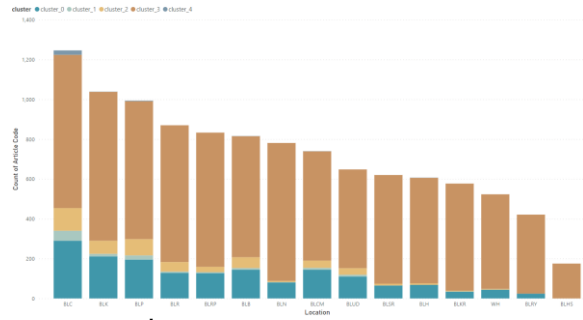
ภาพที่ 7: แสดงประเภทสินค้าตามคลัสเตอร์ในแต่ละสาขา

Location	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	Total
BLB	144	9	54	608	1	816
BLC	290	50	114	770	22	1246
BLCM	144	10	36	549	1	740
BLH	68		8	530	1	607
BLHS				175		175
BLK	211	13	66	748	1	1039
BLKR	34	1	4	538		577
BLN	80	2	6	693		781
BLP	195	21	82	693	3	994
BLR	128	7	48	687		870
BLRP	127	5	27	674		833
BLRY	24			397		421
BLSR	65		9	546		620
BLUD	110	9	33	496		648
WH	44		4	475		523
<b>Total</b>	<b>1664</b>	<b>127</b>	<b>491</b>	<b>8579</b>	<b>29</b>	<b>10890</b>

ภาพที่ 8: รายละเอียดประเภทสินค้าตามคลัสเตอร์ในแต่ละสาขา

## 4. ผลการดำเนินงาน

ผลการวิจัยได้นำข้อมูลการแบ่งแยกตามปริมาณการขายในแต่ละสาขาที่จัดกลุ่มได้ตามภาพที่ 8 โดยแต่ละคลัสเตอร์สามารถเป็นข้อมูลให้ผู้บริหารเพื่อประกอบการพิจารณา ซึ่งจากข้อมูลแต่ละคลัสเตอร์ สรุปผลการดำเนินงานได้ดังนี้



ภาพที่ 9: ปริมาณการขายสินค้าในแต่ละสาขา

จากภาพที่ 9 แสดงให้เห็นข้อมูลในแต่ละคลัสเตอร์ โดยพบว่าลำดับที่ 1 คลัสเตอร์ 3 มีกลุ่มรหัสสินค้าที่ขายได้สูงสุด ลำดับที่ 2 คือคลัสเตอร์ 0 และลำดับที่ 3 คือคลัสเตอร์ 2 และคลัสเตอร์ 1 และ 4 ตามลำดับ ผู้วิจัยได้ศึกษาพบว่าในแต่ละสาขาไม่มีความต่างกันในการเบิกสินค้าจึงทำให้เกิดการเบิกสินค้าที่เกินปริมาณการขาย ดังจะเห็นจากภาพที่ 9 ที่เห็นได้ว่ามีเพียงกลุ่มคลัสเตอร์ 3 ที่มีปริมาณการขายมาก ซึ่งหากดูจากการจัดกลุ่มแล้ว พบบางสาขาไม่มีความจำเป็นที่ต้องขายสินค้าในบางคลัสเตอร์ โดยสรุปข้อมูลได้ดังนี้

Location	cluster_4	cluster_3	cluster_2	cluster_1	cluster_0
BLC	✓	✓	✓	✓	✓
BLP	✓	✓	✓	✓	✓
BLK	✓	✓	✓	✓	✓
BLB	✓	✓	✓	✓	✓
BLCM	✓	✓	✓	✓	✓
BLH	✓	✓	✓		✓
BLR		✓	✓	✓	✓
BLUD		✓	✓	✓	✓
BLRP		✓	✓	✓	✓
BLN		✓	✓	✓	✓
BLKR		✓	✓	✓	✓
BLSR		✓	✓		✓
WH		✓	✓		✓
BLHS		✓			
BLRY		✓			✓

ภาพที่ 10: แสดงสาขาที่จำหน่ายสินค้าตามคลัสเตอร์

## 5. สรุป

จากผลการวิเคราะห์ เกี่ยวกับข้อมูลการขายสินค้าพบว่า มีสาขาที่ไม่มียอดขายสินค้า และขายสินค้าได้น้อยจึงเป็น

กลุ่มที่มีข้อมูลที่น่าสนใจเพราะเป็นกลุ่มที่ส่งผลต่อค่าใช้จ่ายของบริษัท เช่นค่าใช้จ่ายในการขนส่งสินค้าไปยังแต่ละสาขา ค่าเช่าพื้นที่ในการจัดวางสินค้า ค่าเสียโอกาสในการวางสินค้าบนชั้นวางสินค้า เป็นต้น โดยจากการประยุกต์ใช้วิธีการจัดกลุ่มด้วย K-means Clustering ถือว่าเป็นวิธีที่ง่ายที่บริษัทสามารถนำไปใช้งานได้หลังจากการวิจัยเสร็จสิ้น เพื่อนำข้อมูลมาวิเคราะห์และประยุกต์ใช้ให้เกินประสิทธิภาพสูงสุดกับบริษัท ดังเห็นได้จากข้อมูลที่สรุปได้เบื้องต้นว่าสินค้าที่ขายดีเป็นลำดับที่ 1 ของบริษัทคือคลัสเตอร์ 3 แต่ในส่วนของคลัสเตอร์ 4 ควรเสนอขายเพียง 6 สาขาจากทั้งสิ้น 15 สาขา และรองลงมาคือคลัสเตอร์ 1 ควรเสนอขายเพียง 10 สาขา จากทั้งหมด โดยเมื่อนำรหัสสินค้าไปวิเคราะห์เพิ่มพบว่าจากข้อมูลรหัสสินค้าในกลุ่มคลัสเตอร์ 4 รวมยอดซื้อสินค้าสูงถึง 51 ล้านบาท แต่สัดส่วนในการขายเพียงแค่ 24 % ของยอดซื้อ ซึ่งเมื่อบริษัทได้นำการจัดกลุ่มมาใช้สามารถบริหารเงินสดของบริษัทที่นำมาลงทุนเพื่อซื้อสินค้าได้ถึง 27 ล้านบาท

### เอกสารอ้างอิง

- [1] สุทธิพงษ์ ผ่องแก้ว, “การศึกษากระบวนการแบ่งกลุ่มเมลล์คัพด้วยข้อมูลโครงสร้างของเมลล์คัพ”, สารนิพนธ์ วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาการจัดการงานวิศวกรรม มหาวิทยาลัยศรีนครินทรวิโรฒ, 2561.
- [2] ชนกชนม์ ใจมั่นคง, “การวิเคราะห์ข้อมูลการเกิดอุบัติเหตุบนถนนในช่วงเทศกาลปีใหม่โดยเทคนิคเหมืองข้อมูล”, สารนิพนธ์ วิทยาการคอมพิวเตอร์ วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏกาญจนบุรี, 2562.
- [3] สัญญา พันธุ์แพง, “การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลสำหรับพยากรณ์การศึกษาต่อนักศึกษาใหม่ระดับปริญญาตรี ในมหาวิทยาลัยราชภัฏเชียงใหม่”, สารนิพนธ์ เทคโนโลยีสารสนเทศและการสื่อสาร วิทยาลัยแม่ฮ่องสอน มหาวิทยาลัยราชภัฏเชียงใหม่, 2562.
- [4] นันทภัก สุทธิเลิศ, “การวิเคราะห์ความรู้สึกผู้โดยสารที่ใช้บริการสายการบินของบริษัทในประเทศสหรัฐอเมริกา”, สารนิพนธ์ เทคโนโลยีสารสนเทศ วิทยาการสารสนเทศ มหาวิทยาลัยบูรพา, 2560.
- [5] พรนภา ชุมเชื้อ, “การใช้เทคนิคการจัดกลุ่มร่วมกับเทคนิคการจำแนกประเภทสำหรับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้จากบริบทพฤติกรรม”, สารนิพนธ์ เทคโนโลยีสารสนเทศ มหาวิทยาลัยทักษิณ, 2562.
- [6] วณิษา แผลงรักขม และ นิเวศ จิระวิจิตรชัย, “การแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน สำหรับการบริหารลูกค้าสัมพันธ์” วารสารวิชาการชาชนันท์ มจร.ภูเก็ต ปีที่ 3 ฉบับที่ 2, 2562.
- [7] ศุภามณ จันทร์สกุล, “เทคนิคเหมืองข้อมูลในการวิเคราะห์ข้อมูลทางการพยาบาล” *EAU HERITAGE JOURNAL Science and Technology* ปีที่ 12 ฉบับที่ 2, พ.ศ.-ศ.ศ. 2561 หน้า 83-96, 2561.
- [8] นิสานันท์ พลอาสา, “การสร้างแบบจำลองการขายผลิตภัณฑ์และพยากรณ์ยอดขายประกันชีวิตโดยเทคนิคเหมืองข้อมูลกรณีศึกษา บริษัทประกันชีวิตแห่งหนึ่ง” วิทยานิพนธ์ วิทยาลัยนวัตกรรมการศึกษา มหาวิทยาลัยธรรมศาสตร์, 2558.
- [9] สุธีรา วงศ์อนันท์พรีย์ ต้องใจเข้มผกา และ อรรรรณ มุสิกะ, “การใช้เทคนิคเหมืองข้อมูลในการประเมินความรู้ และหาความถนัดเพื่อพัฒนาสัถยภาพของนักศึกษา” *วารสารสังคมศาสตร์* ปีที่ 5 ฉบับที่ 1 ม.ศ. - มิ.ย. 2559 หน้า 12 - 19, 2559.
- [10] วัจนาวา ชาวฟ้า และ อรุณี ศิลาลัย, “การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการตัดสินใจ”, สารนิพนธ์ วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต, 2558.
- [11] รองศาสตราจารย์ ดร.กิตติศักดิ์ เกิดประสพ, “การพัฒนาการทำเหมืองข้อมูลแบบจัดกลุ่ม”, วิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีสุรนารี, 2552.
- [12] กิตติพงษ์ อินทร์ทอง, “การกำหนดเป้าหมายยอดขายที่เหมาะสม กรณีศึกษาสินค้าไฟเบอร์ซีเมนต์”, วิศวกรรมศาสตรมหาบัณฑิต วิทยาลัยธุรกิจบัณฑิต, 2556.
- [13] ชุมชัย บุญศรี, “การพยากรณ์ยอดขายสินค้าประเภทสายไฟฟ้าของลูกค้านักค้าประเภทตัวแทนจำหน่าย (Distributor) : กรณีศึกษาบริษัทสายไฟฟ้าแห่งหนึ่ง”, วิทยานิพนธ์ วิทยาศาสตร์ มหาวิทยาลัยบูรพา, 2559.

# ระบบการวิเคราะห์อุบัติเหตุทางถนนโดยใช้ธุรกิจอัจฉริยะ

## Road Accidents Analytics System using Business Intelligence

ธนินา สังข์แสง (Thanisa Sungsaeng)<sup>1</sup> และผุสดี บุญรอด (Pudsadee Boonrawd)<sup>2</sup>

<sup>1</sup>ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

s6307011858101@email.kmutnb.ac.th, pudsadee.b@itd.kmutnb.ac.th

### บทคัดย่อ

ปัจจุบันปัญหาอุบัติเหตุทางถนนของประเทศไทยเป็นปัญหาใหญ่ที่ยังคงต้องเผชิญ ซึ่งส่งผลกระทบต่อ การบาดเจ็บและเสียชีวิต รวมทั้งความสูญเสียทางเศรษฐกิจ สังคมของผู้ประสบอุบัติเหตุ ครอบครัว ชุมชน และ ประเทศชาติ งานวิจัยนี้จึงได้ทำการพัฒนาระบบการวิเคราะห์ อุบัติเหตุทางถนนที่ใช้กระบวนการวิทยาศาสตร์ข้อมูล (Data Science) ซึ่งแบ่งออกได้เป็น 5 ขั้นตอน ได้แก่ การทำความเข้าใจกำหนดปัญหา (Business Understanding) การเก็บรวบรวมข้อมูล (Data Acquisition) การเตรียมข้อมูล (Data Preparation) การวิเคราะห์ข้อมูล (Data Analysis) และการนำไปใช้งานจริง (Deployment) โดยงานวิจัย ประยุกต์ใช้ธุรกิจอัจฉริยะ (Business Intelligence) สำหรับ ออกรายงานในรูปแบบหลายมิติ และนำเสนอผลการวิเคราะห์ ในรูปแบบแดชบอร์ด (Dashboard) เพื่อรองรับการใช้งาน และสนับสนุนการตัดสินใจที่เกี่ยวข้องกับความปลอดภัย ทางถนนได้อย่างรวดเร็ว และหลากหลายมิติมากขึ้น

**คำสำคัญ:** ระบบการวิเคราะห์ อุบัติเหตุทางถนน วิทยาศาสตร์ ข้อมูล ธุรกิจอัจฉริยะ

### Abstract

At present, the problem of road accidents is still a major problem in Thailand which affects injuries and deaths including the socio-economic loss of the victims, their families, communities, and the nation. This research aims to develop a system for analyzing road accidents using data science which can be divided into 5 steps: Business Understanding, Data Acquisition, Data Preparation, Data Analysis, and Deployment.

This research also uses Business Intelligence to generate reports in multidimensional formats and present the results of the analysis in a dashboard to support the use and making of decisions related to road safety more quickly and multi-dimension.

**Keywords:** Analytic System, Road Accident, Data Science, Business Intelligence.

### 1. บทนำ

ปัจจุบันอุบัติเหตุทางถนนเป็นปัญหาที่มีความสำคัญ ซึ่งมีผู้บาดเจ็บและเสียชีวิตเป็นจำนวนมาก จากรายงาน สถานการณ์ความปลอดภัยทางถนนโลกปีพ.ศ. 2561 ของ องค์การอนามัยโลก (World Health Organization) ระบุ จำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนทั่วโลกสูงถึง 1.35 ล้านคนต่อปีทั่วโลก โดยส่วนใหญ่เป็นประชากรที่มีอายุ 5-29 ปี [1] รวมทั้งได้ระบุถึงประเทศไทยว่ามีอัตราการเสียชีวิต สูงสุดในเอเชียตะวันออกเฉียงใต้ ประมาณการมีจำนวน 22,491 ราย คิดเป็น 32.7 คน ต่อประชากรหนึ่งแสนคน ทั้งนี้จากสถานการณ์อุบัติเหตุทางถนนของประเทศไทย ในรอบ 10 ปีที่ผ่านมา ตั้งแต่ปีพ.ศ. 2554-2563 จำนวน ผู้เสียชีวิตสะสมถึง 206,588 ราย เฉลี่ยเสียชีวิต 20,659 ราย ต่อปี หรือ 58 คน/วัน

ดังนั้นคณะผู้วิจัยจึงเล็งเห็นว่าปัญหาอุบัติเหตุการจราจร ทางถนนถือเป็นปัญหาที่มีความสำคัญ และเป็นสาเหตุ ของการสูญเสียชีวิตและทรัพย์สิน รวมทั้งงบประมาณต่าง ๆ ดังนั้นวัตถุประสงค์ของงานวิจัยนี้จึงได้พัฒนาระบบการ วิเคราะห์อุบัติเหตุทางถนน โดยประยุกต์ใช้ธุรกิจอัจฉริยะ (Business Intelligence) บนพื้นฐานของกระบวนการ วิทยาศาสตร์ข้อมูล (Data Science) ซึ่งในบทความวิจัยแบ่ง

เนื้อหาออกได้เป็น 5 ส่วน ได้แก่ ส่วนที่ 1 บทนำ ส่วนที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง ส่วนที่ 3 วิธีดำเนินการวิจัย ส่วนที่ 4 ผลการดำเนินงานวิจัยและส่วนที่ 5 สรุปผลการวิจัย

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ความปลอดภัยทางถนน

ความปลอดภัยทางถนนเป็นสิ่งสำคัญที่ทุกหน่วยงานหรือทุกกลุ่มคนควรให้ความตระหนัก ปฏิบัติตามกฎหมายระเบียบข้อบังคับ หรือแนวทางปฏิบัติต่าง ๆ ที่เกี่ยวข้องกับการใช้รถใช้ถนน ซึ่งจะช่วยป้องกันและแก้ไขปัญหาอุบัติเหตุทางถนน ลดความเสี่ยงของการเกิดอุบัติเหตุ ลดจำนวนผู้บาดเจ็บหรือพิการ ตลอดจนลดอัตราการสูญเสียชีวิตและทรัพย์สินของผู้ใช้ถนนทุกกลุ่มหรือผู้ขับขี่ยานพาหนะทุกประเภท เพื่อให้มีความปลอดภัยในการเดินทาง ซึ่งความปลอดภัยทางถนน อาจกำหนดเป็นมาตรการ กฎหมาย ระเบียบข้อบังคับ แนวทาง วิธีปฏิบัติ หรือแผนการดำเนินงานในด้านต่าง ๆ สำหรับใช้เป็นกรอบการดำเนินงาน บริหารจัดการ หรือควบคุมผู้ใช้รถใช้ถนน ได้อย่างมีประสิทธิภาพและเกิดประสิทธิผล [2]

#### 2.1.1 ประเภทของอุบัติเหตุ

ประเภทของอุบัติเหตุสามารถจำแนกตามโครงข่ายคมนาคม ประกอบด้วย อุบัติเหตุทางถนน (ทางหลวง ทางหลวงชนบท ทางพิเศษ) อุบัติเหตุทางราง (ทางรถไฟ ไฟฟ้ามหานคร รถไฟฟ้าแอร์พอร์ต เรล ลิงก์) อุบัติเหตุทางน้ำ และอุบัติเหตุทางอากาศ

#### 2.1.2 สาเหตุที่ทำให้เกิดอุบัติเหตุทางถนน

สถิติอุบัติเหตุการจราจรทางถนน สามารถแบ่งสาเหตุการเกิดอุบัติเหตุได้เป็น 3 สาเหตุหลัก [3] คือ

- สาเหตุการเกิดอุบัติเหตุจากบุคคล เช่น เมาสุรา ไม่คาดเข็มขัดนิรภัย ใช้โทรศัพท์เคลื่อนที่ เป็นต้น
- สาเหตุจากสิ่งแวดล้อม เช่น ถนนลื่น ถนนชำรุด ถนนมืด คนตัดหน้ารถ หมอก น้ำท่วม เป็นต้น
- สาเหตุจากอุปกรณ์ที่ใช้ขับขี่ เช่น ระบบไฟฟ้าขัดข้อง ระบบเบรก ยางเสื่อมสภาพ เป็นต้น

### 2.2 ธุรกิจอัจฉริยะ (Business Intelligence : BI)

ธุรกิจอัจฉริยะเป็นเทคโนโลยีที่ช่วยให้ธุรกิจสามารถจัดระเบียบ วิเคราะห์ และประมวลผลข้อมูลทั้งหมดขององค์กร หรือกล่าวได้ว่าเป็นเครื่องมือที่ช่วยในการเก็บข้อมูลจากแหล่งต่าง ๆ แล้วเปลี่ยนเป็นสารสนเทศที่ช่วยสนับสนุนการตัดสินใจ เพื่อใช้ในการดำเนินธุรกิจ การบริหารจัดการ ตลอดจนการวางแผนงานต่าง ๆ

#### 2.2.1 ส่วนประกอบหลักของธุรกิจอัจฉริยะ

คลังข้อมูล (Data Warehouse) เป็นฐานข้อมูลสำหรับจัดเก็บข้อมูลจากแหล่งต่าง ๆ โดยอยู่ที่ส่วนกลาง และเข้าถึงได้ โดยเครื่องมือวิเคราะห์ธุรกิจ (Business Analytics) เป็นเครื่องมือสำหรับช่วยวิเคราะห์ข้อมูลในคลังข้อมูล เพื่อช่วยในการจัดการ ประสิทธิภาพธุรกิจ (Business Performance Management: BPM) สำหรับการตรวจสอบการวิเคราะห์เพื่อ ไปสู่เป้าหมายทางธุรกิจ และยังเป็นส่วนที่ผู้ใช้เชื่อมต่อกับผู้ใช้งาน (User Interface) สำหรับรายงานการแสดงผลข้อมูล (Dashboard) เพื่อช่วยให้ผู้ใช้เข้าถึงข้อมูลได้อย่างรวดเร็ว [4]

#### 2.2.2 รูปแบบของการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูล (Data Analytics) เป็นเครื่องมือสำหรับธุรกิจอัจฉริยะ (Business Intelligence) ซึ่งเป็นศาสตร์ของการใช้ข้อมูลต่าง ๆ จำนวนมากสำหรับการวิเคราะห์ข้อมูลที่มีอยู่ในปัจจุบัน อดีต และทำนายอนาคต โดยเริ่มต้นจากการนำข้อมูลมาทำให้อยู่ในรูปแบบที่พร้อมจะประมวลผลด้วยเทคโนโลยี ชุดคำสั่ง หรือแบบจำลองที่สร้างขึ้น เพื่อนำข้อมูลต่าง ๆ ที่ได้ผ่านการวิเคราะห์มาใช้เป็นรูปแบบของการวิเคราะห์ข้อมูล ซึ่งสามารถแบ่งได้ดังนี้

- 1) การวิเคราะห์แบบพรรณนา (Descriptive Analytics) เป็นการวิเคราะห์พื้นฐาน เพื่อแสดงผลของรายการ เหตุการณ์ หรือกิจกรรมต่าง ๆ ที่ได้เกิดขึ้น หรืออาจกำลังเกิดขึ้น ในลักษณะที่ง่ายต่อการเข้าใจหรือการตัดสินใจ
- 2) การวิเคราะห์แบบพยากรณ์ (Predictive Analytics) เป็นการวิเคราะห์เพื่อพยากรณ์หรือทำนายสิ่งที่กำลังจะเกิดขึ้น หรือน่าจะเกิดขึ้น โดยใช้ข้อมูลที่ได้เกิดขึ้นแล้วร่วมกับแบบจำลองทางสถิติ หรือเทคโนโลยีปัญญาประดิษฐ์ต่าง ๆ (Artificial Intelligence) นอกจากนี้ยังสามารถวิเคราะห์

หาโอกาส และความเสี่ยงต่าง ๆ ที่จะเกิดขึ้นในอนาคตได้ด้วย เช่น การพยากรณ์ยอดขาย และการพยากรณ์ผลประจําชาติ เป็นต้น

3) การวิเคราะห์แบบให้คำแนะนำ (Prescriptive Analytics) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนและยากที่สุด เพราะเป็นทั้งการพยากรณ์ หรือทำนายสิ่งต่าง ๆ ที่จะเกิดขึ้น รวมทั้งให้คำแนะนำในทางเลือกต่าง ๆ ข้อดี ข้อเสียของผลแต่ละทางเลือก และระยะเวลาของสิ่งที่จะเกิดขึ้น โดยแบบจำลองของการวิเคราะห์นี้ในลักษณะนี้สามารถปรับเปลี่ยนได้ตามข้อมูลที่เพิ่มเติมเข้ามา

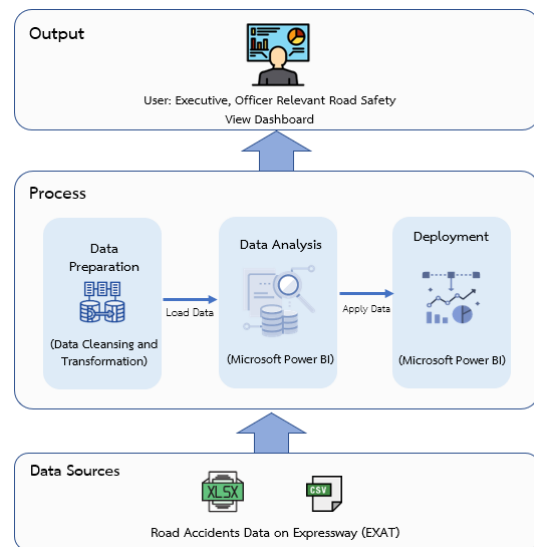
### 2.3 งานวิจัยที่เกี่ยวข้อง

จากการทบทวนงานวิจัยที่เกี่ยวข้องพบว่า การวิเคราะห์ความสัมพันธ์ของปัจจัยที่เกี่ยวข้องกับอุบัติเหตุโดยใช้เทคนิคกฎความสัมพันธ์ ซึ่งปัจจัยเสี่ยงสำคัญที่ก่อให้เกิดอุบัติเหตุจราจร ได้แก่ คน รถ ถนน และสภาพดินฟ้าอากาศ [5] และการใช้เครื่องมือการเรียนรู้ของเครื่อง (Machine learning) โดยใช้ต้นไม้ตัดสินใจ (Decision Tree) สำหรับวิเคราะห์ข้อมูลพยากรณ์การเสียชีวิต และมีการจัดกลุ่มประชากรที่มีโอกาสเสียชีวิตในช่วงเทศกาลปีใหม่แบ่งเป็น 2 กลุ่มประกอบด้วย เพศชายและเพศหญิง ซึ่งลักษณะข้อมูลที่มีความแตกต่างกัน ได้แก่ ช่วงอายุ การดื่มแอลกอฮอล์ และรถจักรยานยนต์ [6] การวิเคราะห์สาเหตุหลักของการเกิดอุบัติเหตุจราจร ซึ่งปัจจัยสาเหตุ ได้แก่ ความเร็ว พฤติกรรมที่ไม่มีความประมาท ข้อบกพร่องบนท้องถนน ยานพาหนะขัดข้อง อากาศมีหมอก และสภาพอากาศ [7] งานวิจัยที่ทำการวิเคราะห์อุบัติเหตุโดยใช้ธุรกิจอัจฉริยะ (Business Intelligence) นำเสนอในรูปแบบข้อมูลจินตภาพ (Data Visualization) สามารถแสดงภาพอุบัติเหตุในมิติเชิงพื้นที่และเวลา เพื่อสร้างความตระหนักการเกิดอุบัติเหตุ [8] และงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบรายงานรูปแบบหลายมิติและแดชบอร์ดในมุมมองต่าง ๆ เพื่อสนับสนุนการตัดสินใจ สามารถทำการวิเคราะห์ และจัดทำรายงานที่เกี่ยวข้องได้อย่างรวดเร็ว [9]

### 3. วิธีการดำเนินการวิจัย

คณะผู้วิจัยทำการพัฒนาระบบการวิเคราะห์อุบัติเหตุทางถนนบนพื้นฐานกระบวนการวิทยาศาสตร์ข้อมูล (Data

Science) แบ่งออกได้เป็น 5 ขั้นตอน ได้แก่ การทำความเข้าใจกำหนดปัญหา (Business Understanding) การเก็บรวบรวมข้อมูล (Data Acquisition) การเตรียมข้อมูล (Data Preparation) การวิเคราะห์ข้อมูล (Data Analysis) และการนำไปใช้งานจริง (Deployment) โดยกรอบแนวคิดงานวิจัยในการวิเคราะห์อุบัติเหตุทางถนนในประเทศไทยโดยใช้ธุรกิจอัจฉริยะ สามารถแสดงได้ดังภาพที่ 1



ภาพที่ 1: กรอบแนวคิดงานวิจัยของระบบการวิเคราะห์อุบัติเหตุทางถนนโดยใช้ธุรกิจอัจฉริยะ

#### 3.1 การทำความเข้าใจกำหนดปัญหา

จากการศึกษาปัญหาอุบัติเหตุทางถนนในประเทศไทยพบว่ามีความโน้มเกิดการเกิดอุบัติเหตุทางถนน ซึ่งมีจำนวนผู้บาดเจ็บและผู้เสียชีวิตอยู่ในเกณฑ์ที่สูง รวมทั้งยังส่งผลกระทบต่อความสูญเสียทรัพย์สิน ตลอดจนเศรษฐกิจสังคมของประเทศ ดังนั้นงานวิจัยนี้จึงได้พัฒนาระบบการวิเคราะห์อุบัติเหตุทางถนนโดยใช้ธุรกิจอัจฉริยะมาช่วยในการวิเคราะห์ และตัดสินใจกรณีศึกษาดังกล่าว

#### 3.2 การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในการวิจัยทำการเก็บรวบรวมข้อมูลอุบัติเหตุทางถนนบนทางพิเศษจากฐานข้อมูลของหน่วยงานการทางพิเศษแห่งประเทศไทย (Expressway Authority of Thailand : EXAT) ซึ่งอยู่ในรูปแบบไฟล์ Excel จำนวน 7 ปี ตั้งแต่ 1 มกราคม 2558 จนถึง 31 ธันวาคม 2564 ประกอบด้วยข้อมูลอุบัติเหตุต่าง ๆ เช่น ลักษณะอุบัติเหตุ บริเวณที่เกิดเหตุ สภาพอากาศ ประเภทรถที่เกิดอุบัติเหตุ เป็นต้น สามารถแสดงตัวอย่างข้อมูลได้ดังภาพที่ 2

สายทาง	รายละเอียดทาง	ลักษณะอุบัติเหตุ	สาเหตุ	ลักษณะช่องทาง	สภาพอากาศ
พญาไท-ศรีนครินทร์	อโศก 4	ชนกับคนเดิน	ขับรถเร็วเกินไป	ทางตรง	ปกติ
พญาไท-บางโคล่	00+600B	ชนขอบทาง	เปลี่ยนช่องทางการทับนิน	ทางตรง	ปกติ
พญาไท-บางโคล่	06+100B	ชนขอบทาง	ขับรถเร็วเกินไป	ทางตรง	ปกติ
พญาไท-บางโคล่	04+000B	อื่นๆ	อื่นๆ	ทางตรง	ปกติ
พญาไท-ศรีนครินทร์	อโศก3-1	ชนกับคนเดิน	ฝ่าฝืนสัญญาณป้ายจราจรทางเข้าด้านเกินเงิน	ปกติ	
บางนา-ฉลองศรี	19+200A	ชนขอบทาง	ขับรถเร็วเกินไป	ทางตรง	ฝนตก
บางนา-แจ้งวัฒนะ	OTH	อื่นๆ	เปลี่ยนช่องทางการทับนิน	ทางโค้ง	ฝนตก
พญาไท-ศรีนครินทร์	12+800A	ชนท้าย	ขับรถเร็วเกินไป	ทางตรง	ฝนตก
รามอินทรา-อารจรณ์	ปราชญ์ศักดิ์	ชนกับรถเก็บค่าผ่านทาง	บีบเบร่า	ทางเข้าด้านเกินเงิน	ปกติ
บางนา-ฉลองศรี	บางลำไย	ชนกับคนเดิน	บีบเบร่า	ทางเข้าด้านเกินเงิน	ปกติ
สุขสวัสดิ์-บางขุนเทียน	06+300B	ชนท้าย	ขับรถเร็วเกินไป	ทางตรง	ปกติ
แจ้งวัฒนะ-พญาไท	ปราชญ์ชัย(เข้า)	ชนกับคนเดิน	ฝ่าฝืนสัญญาณป้ายจราจรทางเข้าด้านเกินเงิน	ปกติ	

ภาพที่ 2: ตัวอย่างข้อมูลอุบัติเหตุที่ใช้ในการวิจัย

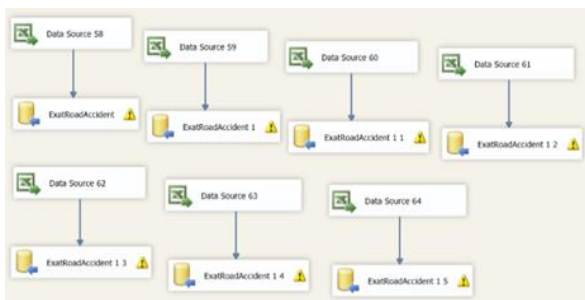
### 3.3 การเตรียมข้อมูล

ในงานวิจัยจะต้องเตรียมและทำความสะอาดข้อมูลก่อนนำไปใช้ในการวิเคราะห์ ซึ่งจะต้องทำการแปลงข้อมูล ตรวจสอบความครบถ้วนสมบูรณ์ ปรับรูปแบบข้อมูล (Data Transformation) ให้อยู่ในรูปแบบมาตรฐานเดียวกัน โดยนำข้อมูลเข้าสู่ฐานข้อมูลที่ใช้โปรแกรมจัดการฐานข้อมูล Microsoft SQL Server โดย Control Flow ของกระบวนการ ETL สามารถแสดงได้ดังภาพที่ 3



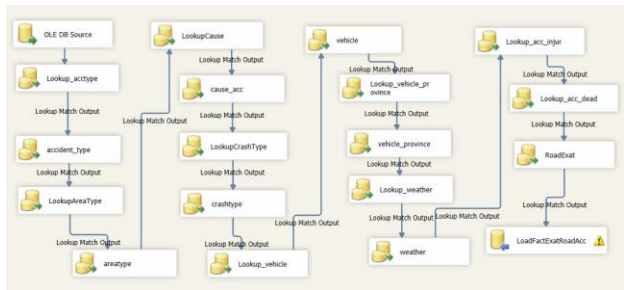
ภาพที่ 3: Control Flow ของกระบวนการ ETL

จากภาพที่ 3 Control Flow หมายเลข 1 ที่ใช้เครื่องมือ Execute SQL Task เพื่อสร้างการทำงาน โดยให้มีการลบข้อมูลออกจากฐานข้อมูลของตาราง ExatRoadAccident ก่อนที่จะทำการ Load ข้อมูลทุกครั้ง Control Flow หมายเลข 2 ใช้เครื่องมือ Data Flow Task เพื่อการควบคุมการไหลของข้อมูลใน Data Flow ซึ่งเป็นการดึงข้อมูล (Extract) จาก Data Sources โดยในงานวิจัยนี้ Data Sources อยู่ในรูปแบบไฟล์ Excel ซึ่งในส่วนการควบคุมการไหลของข้อมูลใน Data Flow สามารถแสดงได้ดังภาพที่ 4



ภาพที่ 4: Data Flow ภายใน Control Flow หมายเลข 2

จากภาพที่ 3 แสดง Control Flow หมายเลข 3 ที่ใช้เครื่องมือ Data Flow Task สำหรับแปลงข้อมูลให้ถูกต้องและปรับรูปแบบข้อมูล ซึ่งได้ปรับปรุงข้อมูลต่าง ๆ ได้แก่ วันที่เกิดอุบัติเหตุ ประเภทอุบัติเหตุ บริเวณที่เกิดเหตุ สาเหตุการเกิดอุบัติเหตุ ลักษณะอุบัติเหตุ ประเภทยานพาหนะที่ใช้ในการขับขี่ และสภาพอากาศ โดยทำการปรับรูปแบบข้อมูล ปรับปรุงข้อมูลให้ถูกต้องโดยใช้เครื่องมือ Lookup ตรวจสอบข้อมูล และทำการแทนที่ข้อมูลตามที่กำหนด จากนั้นนำข้อมูล (Load) เข้าสู่ตาราง เพื่อเตรียมข้อมูลสำหรับนำไปประมวลผลวิเคราะห์ข้อมูล ในการควบคุมการไหลของข้อมูลใน Data Flow สามารถแสดงได้ดังภาพที่ 5



ภาพที่ 5: Data Flow ภายใน Control Flow หมายเลข 3

### 3.4 การวิเคราะห์ข้อมูล

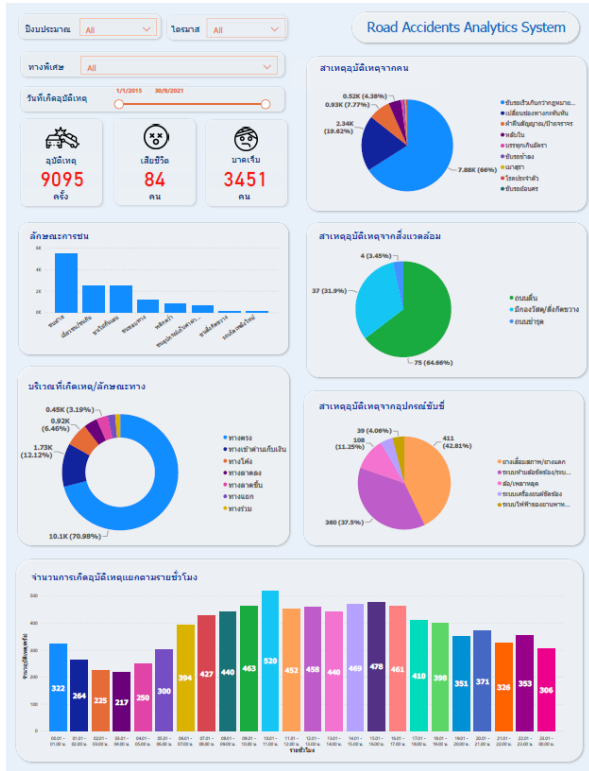
กระบวนการนี้เป็นการนำข้อมูลไปประมวลผลเพื่อวิเคราะห์ข้อมูลที่ใช้โปรแกรม Microsoft Power BI สำหรับวิเคราะห์เพื่อหาความสัมพันธ์ และมุมมองของข้อมูลต่าง ๆ ได้แก่ สาเหตุของการเกิดอุบัติเหตุ ความเสียหายต่อชีวิตและการบาดเจ็บ ช่วงเวลาการเกิดอุบัติเหตุ ลักษณะการเกิดอุบัติเหตุ และบริเวณที่เกิดเหตุ

### 3.5 การนำไปใช้งานจริง

ในการพัฒนาระบบการวิเคราะห์อุบัติเหตุทางถนน โดยใช้ธุรกิจอัจฉริยะ ผลลัพธ์ที่ได้ (Output) จะแสดงผลมุมมองของการวิเคราะห์อุบัติเหตุทางถนน ในรูปแบบข้อมูลจินตภาพ (Data Visualization) และนำเสนอในรูปแบบแดชบอร์ด (Dashboard) ที่สามารถสื่อสารให้ผู้ใช้งานเข้าใจได้ง่าย เพื่อช่วยให้ผู้บริหารและผู้มีส่วนเกี่ยวข้องในด้านความปลอดภัยทางถนนสามารถนำข้อมูลไปประกอบการตัดสินใจ และแก้ไขปัญหาอุบัติเหตุทางถนน

#### 4. ผลการดำเนินงานวิจัย

ผลการดำเนินงานวิจัยสามารถแสดงตัวอย่างผลการพัฒนาระบบการวิเคราะห์อุบัติเหตุทางถนนโดยใช้ธุรกิจอัจฉริยะในรูปแบบแดชบอร์ด (Dashboard) แสดงได้ดังภาพที่ 6-11



ภาพที่ 6: ตัวอย่างหน้าจอแสดงภาพรวมของรายงานการเกิดอุบัติเหตุ

ตัวอย่างหน้าจอการแสดงผลภาพรวมของรายงานการเกิดอุบัติเหตุ สามารถแสดงได้ดังภาพที่ 6

จากผลการพัฒนาระบบได้นำข้อมูลอุบัติเหตุทางถนนบนทางพิเศษมาทำการวิเคราะห์ และนำเสนอข้อมูลใน 5 รูปแบบดังนี้

##### 1) ข้อมูลสรุปจำนวนการเกิดอุบัติเหตุ

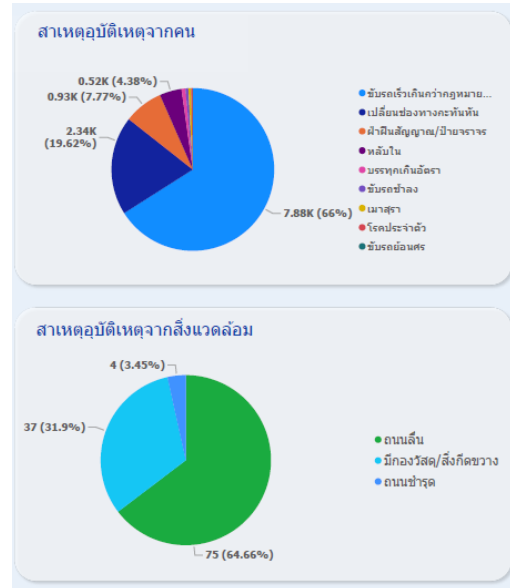
ข้อมูลสรุปจำนวนการเกิดอุบัติเหตุ จะแสดงจำนวนอุบัติเหตุ จำนวนผู้เสียชีวิตและจำนวนผู้บาดเจ็บ แสดงตัวอย่างรายงานได้ดังภาพที่ 7



ภาพที่ 7: ข้อมูลสรุปจำนวนการเกิดอุบัติเหตุ

##### 2) ข้อมูลสาเหตุการเกิดอุบัติเหตุ

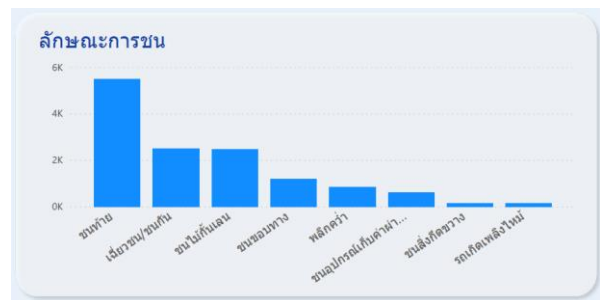
ข้อมูลสาเหตุการเกิดอุบัติเหตุบนทางพิเศษ แบ่งออกเป็น 3 สาเหตุ ประกอบด้วย สาเหตุอุบัติเหตุจากคน สาเหตุอุบัติเหตุจากสิ่งแวดล้อม และสาเหตุอุบัติเหตุจากอุปกรณ์ ขับขี่ แสดงตัวอย่างรายงานได้ดังภาพที่ 8



ภาพที่ 8: รายงานแสดงข้อมูลสาเหตุการเกิดอุบัติเหตุทางถนน

##### 3) ข้อมูลลักษณะการชน

ข้อมูลลักษณะการชนที่เกิดขึ้นบ่อยบนทางพิเศษ ตัวอย่างเช่น การชนท้าย เจี้ยวชน/ชนกัน ชนไม้กั้นเลน ชนขอบทาง พลิกคว่ำ ชนอุปกรณ์เก็บค่าผ่านทาง ชนสิ่งกีดขวาง และรถเกิดเพลิงไหม้ แสดงตัวอย่างรายงานได้ดังภาพที่ 9

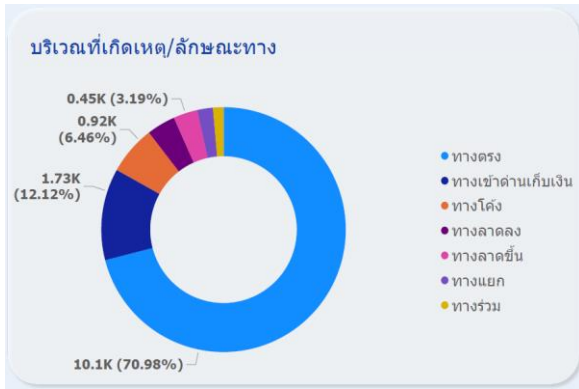


ภาพที่ 9: ข้อมูลลักษณะการชน

##### 4) ข้อมูลบริเวณที่เกิดเหตุ

ข้อมูลบริเวณที่เกิดเหตุ/ลักษณะทางที่เกิดอุบัติเหตุขึ้นบ่อยบนทางพิเศษ ได้แก่ บริเวณทางตรง ทางเข้าด่านเก็บเงิน และทางโค้ง เป็นต้น แสดงตัวอย่างรายงานได้ดังภาพที่ 10





ภาพที่ 10: ข้อมูลบริเวณที่เกิดเหตุ

5) ข้อมูลจำนวนการเกิดอุบัติเหตุแยกตามรายชั่วโมง ข้อมูลจำนวนการเกิดอุบัติเหตุแยกตามรายชั่วโมง พบว่า ช่วงเวลาที่เกิดอุบัติเหตุมากที่สุด คือ 10.01-11.00 น. รองลงมา คือ 15.01-16.00 น. และ 14.01-15.00 น. ตามลำดับ แสดง ตัวอย่างรายงานได้ดังภาพที่ 11



ภาพที่ 11: ข้อมูลจำนวนการเกิดอุบัติเหตุแยกตามรายชั่วโมง

### 5. สรุปผลการวิจัย

จากการพัฒนาระบบการวิเคราะห์อุบัติเหตุทางถนน ที่อิงตามกระบวนการวิทยาศาสตร์ข้อมูล (Data Science) แบ่งออกได้เป็น 5 ขั้นตอน ได้แก่ การทำความเข้าใจ กำหนดปัญหา (Business Understanding) การเก็บรวบรวมข้อมูล (Data Acquisition) การเตรียมข้อมูล (Data Preparation) การวิเคราะห์ข้อมูล (Data Analysis) และการนำไปใช้งานจริง (Deployment) โดยใช้ข้อมูลอุบัติเหตุทางถนนบนทางพิเศษของหน่วยงานทางพิเศษแห่งประเทศไทย ประยุกต์ใช้ธุรกิจอัจฉริยะ ซึ่งใช้โปรแกรม Power BI สำหรับออกรายงาน และนำเสนอผลการวิเคราะห์ข้อมูลอุบัติเหตุทางถนนในรูปแบบแดชบอร์ด (Dashboard) งานวิจัยในอนาคตจะนำเครื่องจักรเรียนรู้ (Machine Learning) มาประยุกต์ใช้ในการวิเคราะห์และสร้างแบบจำลอง เพื่อพยากรณ์หรือทำนายแนวโน้มของอุบัติเหตุทางถนน รวมทั้งหากมีการเพิ่มข้อมูลอื่น ๆ มาใช้ในการพัฒนาระบบ เช่น เรื่องความเร็ว

สภาพจิตใจของผู้ขับขี่ เป็นต้น จะทำให้สามารถออกรายงานได้อย่างครอบคลุม และหลากหลายมากขึ้น

### เอกสารอ้างอิง

- [1] World Health Organization. “Global Status Report on Road Safety 2018” [สืบค้นวันที่ 10 สิงหาคม 2564]. จาก <https://www.who.int/publications/i/item/9789241565684>, 2561.
- [2] ศูนย์อำนวยการความปลอดภัยทางถนน. “แผนแม่บทความปลอดภัยทางถนน” [สืบค้นวันที่ 7 สิงหาคม 2564]. จาก [http://roadsafety.disaster.go.th/upload/minisite/file\\_attach/196/5e8f159b2f84c.pdf](http://roadsafety.disaster.go.th/upload/minisite/file_attach/196/5e8f159b2f84c.pdf), 2562.
- [3] สำนักงานสถิติแห่งชาติ. “อุบัติเหตุการจราจรทางบก” [สืบค้นวันที่ 7 สิงหาคม 2564]. จาก <http://service.nso.go.th/nso/web/statseries/statseries21.html>
- [4] Oracle Netsuite. “23 Case Studies and Real-World Examples of How Business Intelligence Keeps Top Companies Competitive” [สืบค้นวันที่ 4 ตุลาคม 2564]. จาก <https://www.netsuite.com/portal/resource/articles/business-strategy/business-intelligence-examples.shtml>, 2564
- [5] ดวงใจ รุ่งพัฒนกิจชัย. (2561). ความสัมพันธ์ของปัจจัยที่เกี่ยวข้องกับอุบัติเหตุรถจักรยานยนต์เพื่อการสืบสวนคดีอุบัติเหตุจราจรเชิงลึกโดยใช้เทคนิคกฎความสัมพันธ์. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชานิติวิทยาศาสตร์ บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร.
- [6] พิเศษ ฤปดินทร. (2561). การพยากรณ์การเสียชีวิตจากอุบัติเหตุทางถนนในช่วงเทศกาลปีใหม่. การค้นคว้าอิสระวิทยาศาสตรมหาบัณฑิต สาขาวิชานโยบายและการบริหารจัดการ วิทยาลัยนวัตกรรม มหาวิทยาลัยธรรมศาสตร์.
- [7] Narasimhan, G., and Cheriyan, S., (2017). “Predictive Analytics of Road Accidents in Oman using Machine Learning Approach”. In Proceedings of 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). Kerala, India, 6-7 July. pp. 1058-1065.
- [8] Sunkpho, J., and Wipulanusat, W., (2020). “The Role of Data Visualization and Analytics of Highway Accidents,” Walailak Journal of Science and Technology.17(2), pp. 1379-1389.
- [9] รัตนา สุวรรณวิชณี. (2560). “การพัฒนาระบบรายงานรูปแบบหลายมิติเพื่อสนับสนุนการตัดสินใจเชิงนโยบายของสำนักงานคณะกรรมการวิจัยแห่งชาติ”. วารสารศรีปทุมปริทัศน์. ปีที่ 9 ฉบับที่ 1 มกราคม-ธันวาคม. หน้า 66-75.

# การเพิ่มประสิทธิภาพการวิเคราะห์ข้อมูลเพื่อการวินิจฉัยทางการแพทย์

## Enhancement Data Analytics for Medical Diagnosis

สุทธาทิพย์ ปรีดตะพงสาชัย (Suthatip Parittapongsachai)<sup>1</sup> และพยุ่ง มีดีจ (Phayung Meesad)<sup>2</sup>

<sup>1,2</sup>ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858155@email.kmutnb.ac.th, <sup>2</sup>phayung.m@itd.kmutnb.ac.th

### บทคัดย่อ

การเตรียมข้อมูลล่วงหน้าสำหรับการวิเคราะห์ข้อมูลด้วยเทคนิคเดียวอาจส่งผลให้การเรียนรู้ของเครื่องมีประสิทธิภาพไม่เพียงพอ งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อ 1) นำเสนอวิธีการเตรียมข้อมูลเพื่อเพิ่มประสิทธิภาพในการวิเคราะห์ และ 2) เปรียบเทียบประสิทธิภาพของวิธีการที่นำเสนอโดยวิธีการเตรียมข้อมูลที่นำเสนอเป็นการผสมผสานการปรับมาตรฐานขนาดข้อมูล การคัดเลือกคุณลักษณะ และการจัดการข้อมูลที่ไม่สมดุลกัน ในการวิจัยครั้งนี้ใช้ข้อมูลจาก UCI Machine Learning Repository เป็นข้อมูลการวินิจฉัยโรคทางการแพทย์ จำนวน 5 ชุด ประกอบด้วยข้อมูลแบบตัวเลข จำนวน 3 ชุด และข้อมูลตัวเลขร่วมกับข้อมูลที่ไม่ใช่ตัวเลข จำนวน 2 ชุด ผู้วิจัยใช้วิธีการเตรียมข้อมูลที่นำเสนอเพื่อป้อนเข้าสู่การเรียนรู้ของเครื่อง ได้แก่ 1) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน 2) อัลกอริทึมป่าสุ่ม 3) อัลกอริทึมเพื่อนบ้านใกล้เคียง 4) อัลกอริทึมเกาส์เซียนนาอิวเบย์ 5) อัลกอริทึมเพอร์เซ็ปตรอนหลายชั้น 6) อัลกอริทึมเพอร์เซ็ปตรอนหลายชั้นเชิงลึก ผลการวิจัยพบว่า การเตรียมข้อมูลด้วยวิธีผสมผสาน โดยส่วนใหญ่เข้ามาช่วยในการเพิ่มประสิทธิภาพสูงถึงร้อยละ 83-100 ด้วยรูปแบบที่แตกต่างกันขึ้นอยู่กับอัลกอริทึมและลักษณะของข้อมูล

**คำสำคัญ:** การเรียนรู้ของเครื่อง การจำแนกข้อมูล การเตรียมข้อมูล การวิเคราะห์ข้อมูล การวินิจฉัยทางการแพทย์

### Abstract

Data preprocessing for data analysis with a single technique may result in insufficient machine

learning efficiency. The objectives of this research were 1) to present data preparation methods to increase the analytical efficiency and 2) to compare the efficiency of the proposed methods. We propose data preprocessing based on the combination of feature scaling, feature selection, and imbalanced data management. To test the performance of the proposed methods, we used five datasets from the UCI Machine Learning Repository for medical diagnosis. The datasets were three numerical data and two other datasets containing mixed numerical and non-numeric data. The proposed inputs to machine learning are 1) Support Vector Machine, 2) Random Forest, 3) K Nearest Neighbor, and 4) Gaussian Naive Bayesian 5) Multilayer Perceptron 6) Deep Multilayer Perceptron. The results showed that predominantly integrated data preparation contributed to an 83-100% with different formats depending on the algorithm and nature of the data.

**Keywords:** Machine learning, Data classification, Data Preparation, Data Analytics, Medical Diagnosis.

### 1. บทนำ

ปัจจุบันการแพทย์มีการจัดเก็บข้อมูลที่ใช้ในการวินิจฉัยโรคต่างๆ อาทิเช่น โรคไฮเปอร์ไทรอยด์ [1] โรคเบาหวาน [10] โรคไตเรื้อรัง [9] เป็นต้น และยังมีจำนวนผู้ป่วยแต่ละโรคเพิ่มสูงขึ้นอย่างรวดเร็วทำให้ข้อมูลในฐานข้อมูลเพิ่มขึ้นเป็นจำนวนมากและหากมีการนำข้อมูลมาทำการวิเคราะห์จะต้องใช้เวลา [1] ซึ่งมี

การนำเทคโนโลยีต่าง ๆ มาช่วยในการวิเคราะห์ข้อมูลมากมาย อาทิเช่น การวิเคราะห์ข้อมูลด้วยการเรียนรู้ของเครื่องเพื่อให้สามารถวิเคราะห์และทำนายผลลัพธ์ได้แม่นยำมากขึ้นสำหรับการเรียนรู้ของเครื่องนั้นมีข้อจำกัดในเรื่องของการประมวลผลข้อมูลเนื่องจากไม่สามารถนำข้อมูลที่ถูกรวบรวมไว้มาวิเคราะห์ได้โดยตรงหากนำข้อมูลป้อนเข้าสู่การเรียนรู้ของเครื่องโดยตรงโดยไม่มีการทำให้ข้อมูลสมบูรณ์เสียก่อน จะทำให้การประมวลผลนั้นคลาดเคลื่อนไป

การเรียนรู้ของเครื่องมีประสิทธิภาพสูงก็ต่อเมื่อมีการเตรียมข้อมูลด้วยวิธีการที่เหมาะสม ผู้วิจัยได้ทำการศึกษาเทคนิคและวิธีการปรับปรุงประสิทธิภาพของการเรียนรู้ของเครื่องจากงานวิจัยที่เกี่ยวข้อง ได้แก่ การปรับปรุงประสิทธิภาพของการวิเคราะห์ข้อมูลการผ่าตัดหลังคลอดด้วยเทคนิคการคัดเลือกคุณลักษณะข้อมูล [8] การเตรียมข้อมูลด้วยการจัดการข้อมูลสูญหายและการเปลี่ยนแปลงรูปแบบข้อมูล [1], [5] การปรับขนาดของข้อมูลสำหรับการวิเคราะห์ทางชีวแพทย์ [4] การจำแนกโรคมะเร็งเต้านมด้วยการสุ่มลดตัวอย่างใหม่เพื่อแก้ปัญหาความไม่สมดุลของข้อมูล [7] ร่วมกับอัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอนเพื่อช่วยในการจำแนกโรค แต่ยังไม่มีการนำเทคนิคหลายๆ เทคนิคมาผสมผสานกันเพื่อให้ได้วิธีการเตรียมข้อมูลที่เหมาะสมของแต่ละชุดข้อมูล

ผู้วิจัยจึงได้นำเทคนิคการเตรียมข้อมูลมากกว่าหนึ่งเทคนิคมาประยุกต์ใช้ร่วมกับอัลกอริทึมการเรียนรู้ของเครื่อง อาทิเช่น การคัดเลือกคุณลักษณะ [5] การปรับมาตรฐานของข้อมูล [6] การสุ่มตัวอย่างข้อมูลใหม่ [3] และการปรับค่าพารามิเตอร์ของอัลกอริทึมเพื่อเพิ่มประสิทธิภาพในการวินิจฉัยโรคให้มีความแม่นยำมากยิ่งขึ้น

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การเรียนรู้ของเครื่องแบบมีผู้สอน

เป็นการสร้างแบบจำลองเพื่อจำแนกข้อมูลเป็นการจำแนกข้อมูลใหม่ซึ่งแบบจำลองในการจำแนกข้อมูลเป็น

การหาความสัมพันธ์ระหว่างตัวแปรนำเข้าเพื่อทำนายผลลัพธ์ที่เรียกว่าคลาส (Class) [2] จะประกอบด้วย 2 ส่วนคือ ส่วนที่เป็นข้อมูลชุดการเรียนรู้ และส่วนของข้อมูลชุดทดสอบสำหรับงานวิจัยนี้ใช้อัลกอริทึมการเรียนรู้ของเครื่องด้วยวิธีการจำแนกข้อมูล 6 อัลกอริทึม ได้แก่ 1) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector machine: SVM) [9],[10] 2) อัลกอริทึมป่าสุ่ม (Random Forest: RF) [9] 3) อัลกอริทึมเพื่อนบ้านใกล้เคียง (K Nearest Neighbor: KNN) [9],[10] 4) อัลกอริทึมเกาส์เซียนนาอิวเบย์ (Gaussian Naïve Bays: GNB) [8] 5) อัลกอริทึมเพอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron: MLP) [9],[1] 6) อัลกอริทึมเพอร์เซ็ปตรอนหลายชั้นเชิงลึก (Deep Multi-Layer Perceptron: DMLP) [9]

### 2.2 เทคนิคการสุ่มตัวอย่างใหม่

เป็นเทคนิคที่ใช้ในการแก้ปัญหาของการจำแนกข้อมูลที่ไม่สมดุลกัน [3] จำนวนข้อมูลที่มีความแตกต่างกันของคลาสมากจะส่งผลให้การเรียนรู้ของเครื่องจำแนกข้อมูลเอนเอียงไปทางข้อมูลกลุ่มของคลาสที่มากโดยเทคนิคการสุ่มตัวอย่างใหม่มี 2 วิธี ได้แก่ 1) Random Under-Sampling (RUS) [7] เป็นเทคนิคการสุ่มลดกลุ่มตัวอย่างใหม่ของข้อมูลที่มากเกินไปลดลงเท่ากับหรือใกล้เคียงกับกลุ่มข้อมูลตัวอย่างที่น้อยกว่าและ 2) Random Over-Sampling (ROS) [3] เป็นเทคนิคการสุ่มข้อมูลตัวอย่างใหม่ของข้อมูลที่มีจำนวนกลุ่มตัวอย่างน้อยให้เพิ่มขึ้นตามจำนวนใกล้เคียงหรือเท่ากับจำนวนกลุ่มข้อมูลของคลาสที่มาก

### 2.3 การปรับมาตรฐานข้อมูล

การปรับมาตรฐานข้อมูล (Feature Scaling: FSC) [4] เป็นการปรับขนาดของข้อมูลแต่ละแอททริบิวต์ให้อยู่ในช่วงเดียวกันซึ่งหากข้อมูลในแอททริบิวต์นั้นๆ แตกต่างกันมากจะทำให้แอททริบิวต์นั้นเกิดค่าผิดปกติหรือที่เรียกว่า Outliers จึงต้องมีการนำเทคนิคการลดขนาดของข้อมูลเข้ามาช่วยในการเตรียมข้อมูล โดยใช้เทคนิค ดังนี้

2.3.1 Normalization เป็นเทคนิคการปรับขนาดของข้อมูลโดยการคำนวณจากผลต่างของแอททริบิวต์แต่ละ

แอททริบิวต์และค่าต่ำสุดของแอททริบิวต์นั้นๆหารด้วยผลต่างของค่าของแอททริบิวต์ที่สูงสุดและต่ำสุดเพื่อลดขนาดของข้อมูลให้อยู่ในช่วง 0 ถึง 1 ดังสมการ 2-1

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2-1)$$

โดยที่

$x_{norm}$  คือ ค่าที่ทำการปรับช่วง Normalization

$x$  คือ ค่าก่อนการปรับช่วงของแอททริบิวต์

$x_{min}$  คือ ค่าที่ต่ำที่สุดของแต่ละแอททริบิวต์

$x_{max}$  คือ ค่าที่สูงที่สุดของแต่ละแอททริบิวต์

2.3.2 Standardization เป็นเทคนิคการปรับขนาดของข้อมูล โดยการนำค่าผลต่างระหว่างข้อมูลแต่ละแถวของแอททริบิวต์กับค่าเฉลี่ยของแอททริบิวต์ที่พิจารณาหารด้วยค่าความแปรปรวนของแอททริบิวต์ที่พิจารณา โดยมีค่าเฉลี่ยเท่ากับศูนย์และค่าความแปรปรวนเท่ากับหนึ่ง ดังสมการ 2-2

$$x_{std} = \frac{x - \mu}{\sigma} \quad (2-2)$$

โดยที่

$x_{std}$  คือ ค่าที่ทำการปรับช่วง Standardization

$x$  คือ ค่าก่อนการปรับช่วงของแอททริบิวต์

$\mu$  คือ ค่าเฉลี่ยของแต่ละแอททริบิวต์

$\sigma$  คือ ค่าความแปรปรวนของแต่ละแอททริบิวต์

2.3.3 Robust Scale เป็นเทคนิคการลดขนาดด้วยการคำนวณผลต่างระหว่างค่าของข้อมูลแต่ละแถวในแอททริบิวต์กับค่ามัธยฐานของแอททริบิวต์ที่พิจารณาหารด้วยค่าพิสัยระหว่างควอไทล์ [7] ดังสมการ 2-3

$$x_{rbs} = \frac{x - x_{median}}{IQR} \quad (2-3)$$

โดยที่

$x_{rbs}$  คือ ค่าที่ทำการปรับช่วง Robust Scale

$x$  คือ ค่าก่อนการปรับช่วงของแอททริบิวต์

$x_{median}$  คือ ค่ามัธยฐานของช่วงควอไทล์

$IQR$  คือ ความแตกต่างระหว่างควอไทล์ที่ 25

และ 75

## 2.4 การคัดเลือกคุณลักษณะข้อมูล

เป็นเทคนิคการคัดเลือกข้อมูลของแต่ละแอททริบิวต์ที่มีความซ้ำซ้อนกันหรือมีความสำคัญในการจำแนกข้อมูลน้อยออกจากการพิจารณาเพื่อลดระยะเวลาในการ

ประมวลผลการเรียนรู้ของเครื่อง โดยการคัดเลือกคุณลักษณะของข้อมูลนั้นมีด้วยกัน 3 วิธี [5] ได้แก่ การคัดเลือกคุณลักษณะแบบกรอง เป็นการคัดเลือกแอททริบิวต์ทั้งหมด โดยการใช้ค่าสถิติเพื่อจัดลำดับหาความสัมพันธ์เพื่อคัดเลือกแอททริบิวต์ที่ดีที่สุดซึ่งจะใช้เวลาในการประมวลผลเร็วและเกิดโอเวอร์ฟิตดิ่งน้อยกว่า (ที่สุด 2) การคัดเลือกคุณลักษณะแบบควมรวม เป็นการคัดเลือกแอททริบิวต์แบบเซตย่อยจากแอททริบิวต์ทั้งหมดโดยเพิ่มจำนวนแอททริบิวต์ทีละตัวหรือลดแอททริบิวต์ทีละตัวเพื่อหาเซตแอททริบิวต์ที่ดีที่สุดแต่จะใช้เวลาในการประมวลผลนานและเกิดโอเวอร์ฟิตดิ่งได้ง่าย 3) การคัดเลือกคุณลักษณะแบบฟิง เป็นการคัดเลือกแอททริบิวต์โดยการเรียนรู้ของอัลกอริทึมซึ่งเป็นวิธีที่ไม่ค่อยมีความยืดหยุ่นสำหรับบางอัลกอริทึมในการจำแนกข้อมูล ในงานวิจัยนี้จะเลือกใช้การคัดเลือกคุณลักษณะแบบกรองโดยใช้ค่าสถิติในการคัดเลือกแอททริบิวต์ด้วย SelectKbest เป็นการคัดเลือกค่า K ที่มีคะแนนสูงสุดซึ่งพิจารณาจาก ANOVA [11] เป็นการวิเคราะห์ความแปรปรวน 2 กลุ่มเพื่อจัดลำดับในการเลือกหากค่า F-Value มีค่ามากก็จะส่งผลต่อการคัดเลือกแอททริบิวต์ได้ดี ดังสมการ 2-4

$$F = \left( \frac{SSB}{df_b} \right) / \left( \frac{SSW}{df_w} \right) \quad (2-4)$$

โดยที่

$SSB$  คือ ผลรวมยกกำลังสองระหว่างกลุ่ม

$SSW$  คือ ผลรวมยกกำลังสองภายในกลุ่ม

$df_b$  คือ องศาอิสระระหว่างกลุ่ม

$df_w$  คือ องศาอิสระภายในกลุ่ม

## 2.5 การตรวจสอบแบบไขว้

เป็นวิธีการทดสอบโดยการแบ่งกลุ่มข้อมูลออกเป็น K กลุ่ม [1] แต่ละกลุ่มจะมีจำนวนเท่าๆ กันและข้อมูลแต่ละกลุ่มวนสับเปลี่ยนกันสำหรับการเรียนรู้และการทดสอบตามจำนวน K กลุ่ม ตัวอย่างเช่น ทำการแบ่งข้อมูลออกเป็น 5 ชุด โดยทำการคำนวณ 5 รอบ ซึ่งในแต่ละรอบจะทำการเลือกข้อมูล 4 ชุดสำหรับการเรียนรู้และ 1 ชุดสำหรับการทดสอบแสดงตัวอย่าง ดังภาพที่ 1

Iteration 1: train on	2	3	4	5	test on	1
Iteration 2: train on	1	3	4	5	test on	2
Iteration 3: train on	1	2	4	5	test on	3
Iteration 4: train on	1	2	3	5	test on	4
Iteration 5: train on	1	2	3	4	test on	5

ภาพที่ 1: การแบ่งชุดข้อมูลแบบ k-fold โดยที่ k=5 [1]

### 2.6 การวัดประสิทธิภาพ

การประเมินประสิทธิภาพการนำเสนอวิธีเตรียมข้อมูลของอัลกอริทึมการจำแนกข้อมูลด้วยค่าความถูกต้อง (Accuracy) ดังสมการ 2-5

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2-5)$$

โดยที่

TP คือ ข้อมูลผู้ป่วยที่เสี่ยงและทำนายว่าเสี่ยง

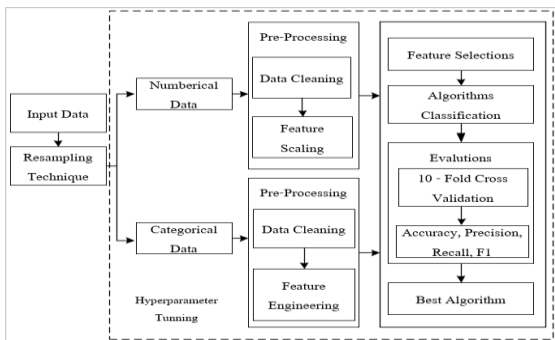
TN คือ ข้อมูลผู้ป่วยที่ไม่เสี่ยงและทำนายว่าไม่เสี่ยง

FP คือ ข้อมูลผู้ป่วยที่เสี่ยงและทำนายว่าไม่เสี่ยง

FN คือ ข้อมูลผู้ป่วยที่ไม่เสี่ยงและทำนายว่าเสี่ยง

### 3. วิธีการดำเนินการวิจัย

การดำเนินงานวิจัยในครั้งนี้เป็นการนำข้อมูลเกี่ยวกับทางด้านการแพทย์มาประยุกต์ใช้กับเทคนิคการเตรียมข้อมูลเพื่อเพิ่มประสิทธิภาพในการจำแนกข้อมูล โดยมีรายละเอียด ดังภาพที่ 2



ภาพที่ 2: กรอบแนวคิดการดำเนินงานวิจัย

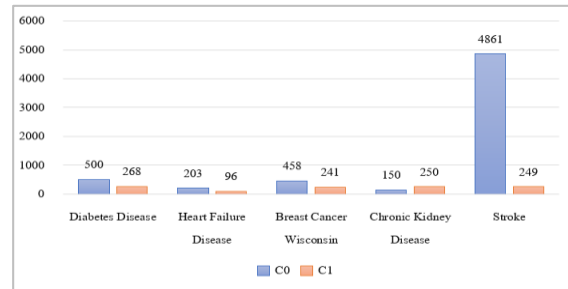
#### 3.1 การรวบรวมข้อมูล

ผู้วิจัยได้นำข้อมูลจากฐานข้อมูล UCI Machine Learning Repository จำนวน 5 ชุดเป็นข้อมูลแบบมีโครงสร้าง ประกอบด้วยข้อมูลแบบตัวเลข จำนวน 3 ชุด ได้แก่ 1) โรคเบาหวาน (Diabetes Disease) จำนวน 768 ระเบียบ 8 แอททริบิวต์ 2) โรคหัวใจล้มเหลว (Heart Failure Disease) จำนวน 299 ระเบียบ 12 แอททริบิวต์

และ 3) โรคมะเร็งเต้านม (Breast Cancer Wisconsin) จำนวน 699 ระเบียบ จำนวน 10 แอททริบิวต์ และข้อมูลตัวเลขร่วมกับข้อมูลที่ไม่ใช่ตัวเลข จำนวน 2 ชุด ได้แก่ 1) โรคไตเรื้อรัง (Chronic Kidney Disease) จำนวน 400 ระเบียบ 25 แอททริบิวต์ และ 2) โรคหลอดเลือดสมอง (Stroke Disease) จำนวน 5,110 ระเบียบ 11 แอททริบิวต์

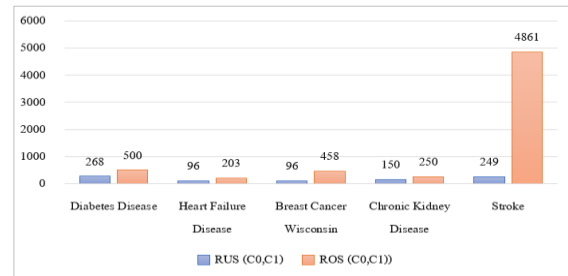
#### 3.2 การสุ่มกลุ่มตัวอย่างใหม่

ข้อมูลที่นำมาวิเคราะห์นั้นสามารถแสดงความแตกต่างของคลาสได้ ดังภาพที่ 3



ภาพที่ 3: ความแตกต่างกันของคลาสในแต่ละชุดข้อมูล

จากภาพที่ 3 ข้อมูลบางชุดข้อมูลมีความแตกต่างกันของคลาสค่อนข้างมากจึงทำการสุ่มเพิ่มกลุ่มตัวอย่างตามจำนวนคลาสมากและสุ่มลดกลุ่มตัวอย่างตามจำนวนคลาสน้อยในแต่ละชุดข้อมูล ดังภาพที่ 4



ภาพที่ 4: จำนวนข้อมูลแต่ละคลาสหลังสุ่มกลุ่มตัวอย่าง

#### 3.3 การเตรียมข้อมูล

งานวิจัยนี้แนะนำเสนอวิธีการเตรียมข้อมูลแบบผสมผสานก่อนป้อนข้อมูลเข้าสู่การเรียนรู้ของเครื่องสำหรับการจำแนกข้อมูลโดยแบ่งเป็น 2 ส่วน ดังนี้

##### 3.3.1 การเตรียมข้อมูลแบบไม่ใช่ตัวเลข

การจัดการข้อมูลสูญหายใช้วิธีการแทนค่าข้อมูลสูญหายด้วยค่าความถี่สูงสุดของข้อมูลแต่ละแอททริบิวต์ในชุดข้อมูลโรคไตเรื้อรังซึ่งมีจำนวนข้อมูลจัดหมวดหมู่สูญ

หาย 11 แอททริบิวต์ และชุดข้อมูลโรคหลอดเลือดสมอง ไม่มีจำนวนข้อมูลสูญหายสำหรับข้อมูลที่ไม่ใช่ตัวเลข

การแปลงข้อมูลแบบจัดหมวดหมู่ให้เป็นตัวเลขด้วยการเข้ารหัสแบบ One-Hot Encoding ในชุดข้อมูลโรคไตเรื้อรังมีแอททริบิวต์ที่ไม่ใช่ตัวเลขจำนวน 10 แอททริบิวต์และเมื่อผ่านการแปลงข้อมูลจะเพิ่มเป็น 20 แอททริบิวต์ สำหรับโรคหลอดเลือดสมองมีแอททริบิวต์ที่ไม่ใช่ตัวเลขจำนวน 5 แอททริบิวต์ และเมื่อแปลงข้อมูลเป็นตัวเลขแล้วจะเพิ่มมาเป็น 16 แอททริบิวต์ในรูปแบบของเลขฐานสองที่มีค่า 0 หรือ 1 เท่านั้น

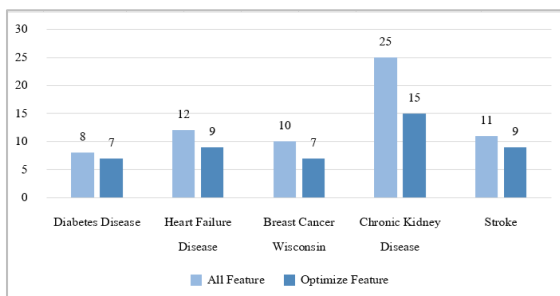
### 3.3.2 การเตรียมข้อมูลแบบตัวเลข (Numerical Data)

การจัดการข้อมูลสูญหายข้อมูลที่เป็ตัวเลขจะทำการจัดการข้อมูลโดยใช้วิธีการแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยในชุดข้อมูลโรคมะเร็งเต้านมจำนวน 1 แอททริบิวต์แทนค่าเฉลี่ยเท่ากับ 3.544656 จำนวน 16 ระเบียบ

การปรับมาตรฐานของข้อมูล (Feature Scaling: FSC) จะใช้ 3 เทคนิค ได้แก่ 1) MinMax Scale 2) เทคนิค Standard Scale และ 3) เทคนิค Robust Scale โดยทำการ Hyperparameter Tunning เพื่อค้นหาเทคนิคการปรับขนาดมาตรฐานที่ดีที่สุดสำหรับแต่ละชุดข้อมูล ซึ่งเป็นการปรับช่วงขนาดของแอททริบิวต์ให้มีช่วงใกล้เคียงกัน

### 3.4 การคัดเลือกคุณลักษณะข้อมูล

เมื่อดำเนินการเตรียมข้อมูลแบบไม่ใช่ตัวเลขและแบบตัวเลขเรียบร้อยแล้วนำแอททริบิวต์ทั้งหมดมา รวมกันและมาคัดเลือกคุณลักษณะของข้อมูลเพื่อหาจำนวนแอททริบิวต์ที่เหมาะสมของแต่ละชุดข้อมูลด้วยการทำ Hyperparameter Tunning ของแต่ละชุดข้อมูล ดังภาพที่ 5



ภาพที่ 5: จำนวนแอททริบิวต์ที่คัดเลือก

### 3.5 การสร้างแบบจำลอง

การสร้างแบบจำลองจะแบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลชุดการเรียนรู้ และข้อมูลชุดการทดสอบ การสร้างแบบจำลองจะแบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลชุดการเรียนรู้ และข้อมูลชุดการทดสอบโดยมีอัตราส่วนเป็น 70:30 และทำการเปรียบเทียบประสิทธิภาพวิธีนำเสนอการเตรียมข้อมูลซึ่งมีการปรับค่าพารามิเตอร์ (HPT) ของอัลกอริทึมโดยมีรายละเอียด ดังตารางที่ 1

ตารางที่ 1: กำหนดค่าพารามิเตอร์ของอัลกอริทึม

Algorithms	HPT	FSC
SVM	C : np.arange(0.1,1) gamma : ['scale','auto'] kernel : ['linear', 'poly', 'rbf', 'sigmoid']	Y
KNN	weights : ['uniform','distance'] n_neighbors : 1,3,5,7,9	Y
GNB	var_smoothing : np.logspace(0,-9, num=100)	N
RF	criterion : ['gini','entropy'] max_features : ['auto','sqrt','log2']	N
MLP	hidden_layer_sizes : [(10),(20),(30),(50),(100,)] activation : ['identity', 'logistic', 'tanh', 'relu'] solver : ['lbfgs','sgd', 'adam']	Y
DMLP	hidden_layer_sizes : [(10,20,10),(20,30,20), (50,50,50),(100,100,100)] activation : ['identity', 'logistic', 'tanh', 'relu'] solver : ['lbfgs','sgd', 'adam'] learning_rate : ['constant','adaptive','invscaling']	Y

### 3.6 การวัดประสิทธิภาพ

วัดประสิทธิภาพโดยการทดสอบด้วย 10-fold cross validations และวัดประสิทธิภาพด้วยค่าความถูกต้อง โดยเปรียบเทียบประสิทธิภาพของการเตรียมข้อมูลแบบดั้งเดิมและประสิทธิภาพของข้อมูลที่ผ่านการเตรียมข้อมูลแบบผสมผสานเทคนิคต่างๆ

## 4. ผลการดำเนินงาน

ผลการดำเนินงานแบ่งออกเป็น 4 รูปแบบของแต่ละชุดข้อมูล ดังนี้ 1) ผลการเปรียบเทียบของการเตรียมข้อมูลแบบดั้งเดิม 2) ผลการเปรียบเทียบด้วยการปรับมาตรฐานของข้อมูลร่วมกับการคัดเลือกคุณลักษณะ 3) ผลการเปรียบเทียบด้วยการปรับมาตรฐานของข้อมูลร่วมกับการคัดเลือกคุณลักษณะและสุ่มลดกลุ่มตัวอย่างใหม่ และ 4) ผลการเปรียบเทียบด้วยการปรับมาตรฐานของข้อมูลร่วมกับการคัดเลือกคุณลักษณะและสุ่มเพิ่มกลุ่มตัวอย่างใหม่โดยแสดงค่าความถูกต้องทั้ง 4 รูปแบบ ดังตารางที่ 2

**ตารางที่ 2:** ผลการเปรียบเทียบค่าความถูกต้องระหว่างการเตรียมข้อมูลแบบดั้งเดิมและวิธีการเตรียมข้อมูลแบบผสมผสาน

Data Set	Algorithms	Original	FSC+FST+HPT	RUS+FSC+FST+HPT	ROS+FSC+FST+HPT
Diabetes Disease	KNN	73.59	75.32	77.02	82.33
	RF	76.62	78.35	78.88	<b>83.33</b>
	SVM	77.92	77.06	73.29	71.67
	GNB	78.35	74.89	74.53	70.67
	MLP	74.03	79.22	77.02	79.33
	DMLP	64.50	78.35	79.50	78.33
Heart Failure Disease	KNN	65.56	76.67	81.03	84.43
	RF	85.56	<b>90.00</b>	84.48	84.43
	SVM	68.89	84.44	74.14	81.97
	GNB	81.11	85.56	79.31	77.05
	MLP	68.89	84.44	74.14	86.07
	DMLP	68.89	85.56	77.59	85.25
Breast Cancer Wisconsin	KNN	96.67	97.14	<b>98.62</b>	97.45
	RF	96.19	96.67	<b>100.00</b>	98.18
	SVM	97.14	96.67	<b>100.00</b>	97.45
	GNB	96.19	96.19	95.86	96.73
	MLP	96.19	97.14	98.62	97.82
	DMLP	96.67	98.10	99.31	97.45
Chronic Kidney Disease	KNN	74.17	94.17	<b>100.00</b>	98.67
	RF	99.17	96.67	98.89	98.00
	SVM	66.67	96.67	<b>100.00</b>	98.67
	GNB	99.17	94.17	95.56	96.00
	MLP	88.33	99.17	<b>100.00</b>	99.33
	DMLP	67.50	<b>100.00</b>	<b>100.00</b>	99.33
Stroke	KNN	94.26	94.26	79.33	96.81
	RF	94.32	94.32	73.33	<b>98.08</b>
	SVM	94.32	94.32	78.00	77.89
	GNB	91.19	94.32	78.00	76.45
	MLP	94.32	94.32	76.67	87.66
	DMLP	94.32	94.32	75.33	93.79

จากตารางที่ 2 สามารถอธิบายผลการทดลองเทคนิคการผสมผสานแบบที่หนึ่งการปรับมาตรฐานของข้อมูล ร่วมกับการคัดเลือกคุณลักษณะพบว่าโรคไตเรื้อรัง อัลกอริทึมเพอร์เซ็ปตรอนหลายชั้นมีค่าความถูกต้องสูงสุดร้อยละ 100 และโรคหัวใจล้มเหลวอัลกอริทึมป่าสุ่มมีค่าความถูกต้องสูงสุดร้อยละ 90 เทคนิคการผสมผสานแบบที่สองการปรับมาตรฐานของข้อมูล ร่วมกับการคัดเลือกคุณลักษณะและสุ่มลดกลุ่มตัวอย่างใหม่พบว่าโรคไตเรื้อรังอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน อัลกอริทึมเพื่อนบ้านใกล้เคียง อัลกอริทึมเพอร์เซ็ปตรอนหลายชั้น และอัลกอริทึมเพอร์เซ็ปตรอนหลายชั้นเชิงลึกมีค่าความถูกต้องสูงสุดร้อยละ 100 และโรคมะเร็งเต้านม อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน และอัลกอริทึมป่าสุ่มมีค่าความถูกต้องสูงสุดร้อยละ 100 และเทคนิคการผสมผสานแบบที่สามการปรับมาตรฐานของข้อมูลร่วมกับการคัดเลือกคุณลักษณะและสุ่มเพิ่มกลุ่มตัวอย่างใหม่พบว่าโรคเบาหวานและโรคหลอดเลือดสมองอัลกอริทึมป่าสุ่มมีค่าความถูกต้องสูงสุดร้อยละ 83.33 และ 98.08 ตามลำดับ

**5. สรุปผล อภิปรายผล และข้อเสนอแนะ**

จากผลการนำเสนอวิธีการเตรียมข้อมูลแบบผสมผสานหลายเทคนิคพบว่าวิธีการเตรียมข้อมูลด้วยการปรับ

มาตรฐานขนาดของข้อมูล การคัดเลือกคุณลักษณะ ร่วมกับการจัดการข้อมูลที่ไม่สมดุลกันด้วยการสุ่มตัวอย่างใหม่และการปรับค่าพารามิเตอร์ที่เหมาะสมนั้น เข้ามาช่วยในการเพิ่มประสิทธิภาพในการจำแนกข้อมูลทางการแพทย์ยกเว้นบางกรณี ได้แก่ 1) ข้อมูลโรคหลอดเลือดสมองมีความแตกต่างของจำนวนคลาสมากทำให้ประสิทธิภาพการเรียนรู้ของเครื่องมีค่าสูงเกินความเป็นจริงการเตรียมข้อมูลด้วยการสุ่มตัวอย่างด้วยการเพิ่มหรือลดกลุ่มตัวอย่างให้จำนวนคลาสมีค่าใกล้เคียงกันทำให้ประสิทธิภาพลดลง 2) ข้อมูลโรคมะเร็งเต้านมเป็นข้อมูลแบบตัวเลขที่มีค่าสูงสุดและค่าต่ำสุดของแต่ละแอททริบิวต์ใกล้เคียงกันจะไม่ส่งผลต่อการปรับช่วงขนาดข้อมูล 3) อัลกอริทึมเคาเซียนนาอ็อฟเซย์ไม่เหมาะสมกับการเตรียมข้อมูลแบบผสมผสานเนื่องจากหากข้อมูลมีความซับซ้อนมากจะไม่สามารถทำนายผลได้ดีเหมือนกับอัลกอริทึมอื่นๆ

**เอกสารอ้างอิง**

- [1] ณัฐวดี หงส์บุญมี และ ประภาศิริ ตรีพานิชกุล, “การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์โทรอยด์ด้วยเทคนิคเหมืองข้อมูล,” *Journal of Information Science and Technology*, vol. 9, no. 1, pp. 41–51, Jun. 2562.
- [2] รัชพล กลัดชื่น และ จริญญา แสนราช, “การเปรียบเทียบประสิทธิภาพอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาในระดับอาชีวศึกษา,” vol. 17, no. 1, pp. 1–10, 2561.
- [3] นิเวศ จิระวิจิตรชัย, “แบบจำลองการตรวจสอบการทุจริตสำหรับข้อมูลที่ไม่สมดุลโดยใช้เทคนิคการลดมิติข้อมูลร่วมกับการเรียนรู้ของเครื่อง,” *Science and Technology RMUTT Journal*, vol. 10, no. 1, pp. 215–225, 2564.
- [4] X. H. Cao, I. Stojkovic, and Z. Obradovic, “A robust data scaling algorithm to improve classification accuracies in biomedical data,” *BMC Bioinformatics*, pp. 1–10, 2016, doi: 10.1186/s12859-016-1236-x.
- [5] N. Sharma, H. V. Bhandari, N. S. Yadav, and H. V. Shroff, “Optimization of IDS using Filter-Based Feature Selection and Machine Learning Algorithms,” *International Journal of Innovative*

- Technology and Exploring Engineering (IJITEE)*, vol. 10, no. 2, pp. 96–102, Dec. 2020.
- [6] A. Md Manjurul, M. A. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, “Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance,” *MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.*, vol. 9, no. 3, pp. 1–17, Jul.2021.
- [7] A.-Z. S. B. Habib, K. T. Islam, M. M. H. Pranto, and M. Nooruddin, “Breast Cancer Classification Using Ensemble Hard Voting with Random Under-Sampling,” *IEEE*, Apr. 2021, doi: 10.1109/ICECE51571.2020.9393080.
- [8] D. T, S. A, and Y. Y, “Feature Selection Based on Naive Bayes for Caesarean Section Prediction,” *IOP Conference Series: Materials Science and Engineering*, vol. 879, no. 1, pp. 1–8, Jul. 2020, doi: 10.1088/1757-899X/879/1/012091.
- [9] ประยุทธ์ศิลป์ ชัยนาม, “การสร้างแบบจำลองจำแนกกลุ่มผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูลและวิชาพลเซชัน,” *การประชุมวิชาการ มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ ครั้งที่ 4*, pp. 1–8, 28 2562.
- [10] รุ่งโรจน์ บุญมา และ นิเวศ จิระวิจิตรชัย, “การจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล,” *วารสารวิชาการสายเคเบิล มจร.ภูเก็ต*, vol. 3, no. 2, Art. no. 2, 2562.
- [11] I. Kim and H. M. Deberneh, “Prediction of Type 2 Diabetes Based on Machine Learning Algorithm,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, pp. 2–14, Mar. 2021, doi: <https://doi.org/10.3390/ijerph18063317>.



# การพยากรณ์ราคาหุ้นระยะสั้นและระยะยาวด้วยวิธีการเรียนรู้เชิงลึก

## Short and Long-term Stock Price Forecasting with Deep Learning

สุดสรณ์ อยู่มาก (Sutasorn Yumak)<sup>1</sup> และ พยุง มีสัง (Phayung Meesad)<sup>2</sup>

<sup>1</sup>ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858147@email.kmutnb.ac.th, <sup>2</sup>phayung.m@itd.kmutnb.ac.th

### บทคัดย่อ

ราคาหุ้นมีความผันผวนมากทำให้ยากต่อการพยากรณ์ จำเป็นที่ต้องใช้วิธีการการเรียนรู้ของเครื่องที่มีความสามารถในการรองรับความผันผวน งานวิจัยนี้นำเสนอวิธีการสร้างโมเดลพยากรณ์ราคาหุ้นระยะสั้นและระยะยาวด้วยเทคนิคการเรียนรู้เชิงลึกเทคนิคที่เลือกใช้ ได้แก่ การเรียนรู้เชิงลึกชนิดความจำระยะสั้นระยะยาว ความจำระยะสั้นระยะยาวแบบสองทาง หน่วยเวียนกลับแบบมีประตู และหน่วยเวียนกลับแบบมีประตูแบบสองทาง โดยใช้ชุดข้อมูลอนุกรมเวลาของหุ้น 3 ตัว ได้แก่ BBL, KBANK และ SCB ตั้งแต่วันที่ 1 มกราคม 2561 ถึง 29 พฤศจิกายน 2564 ทำการแบ่งข้อมูลเป็น 3 ชุด ได้แก่ ชุดข้อมูลฝึกฝน ชุดข้อมูลตรวจสอบ และชุดข้อมูลทดสอบ ใช้เทคนิคหน้าต่างเคลื่อนที่เพื่อแบ่งชุดข้อมูลสำหรับพยากรณ์ราคาปีครระยะสั้น (พยากรณ์ 1 วันข้างหน้า) กับระยะยาว (พยากรณ์ 22 วันข้างหน้า) ในการสร้างโมเดลการพยากรณ์ราคาหุ้นงานวิจัยนี้จัดเตรียมข้อมูลนำเข้า 3 รูปแบบ ได้แก่ 1. ข้อมูลราคาพื้นฐาน 2. ข้อมูลราคาหุ้นร่วมกับข้อมูลปัจจัยทางเทคนิคและ 3. ข้อมูลราคาหุ้นร่วมกับ ปัจจัยทางเทคนิคและสัญญาณการซื้อขาย เมื่อได้โมเดลแล้วก็ทำการทดสอบโมเดล เพื่อเปรียบเทียบประสิทธิภาพในการพยากรณ์ราคาปีคของหุ้น ผลการวิจัยพบว่าการพยากรณ์ระยะสั้นประสิทธิภาพไม่มีความแตกต่างอย่างมีนัยสำคัญ ในขณะที่ระยะยาวโมเดลที่ใช้ข้อมูลนำเข้ารูปแบบที่ 3 มีประสิทธิภาพสูงที่สุด

**คำสำคัญ :** อนุกรมเวลา หน้าต่างเคลื่อนที่ หน่วยความจำระยะสั้นระยะยาว หน่วยความจำระยะสั้นระยะยาว

แบบสองทาง หน่วยเวียนกลับแบบมีประตู หน่วยเวียนกลับแบบมีประตูแบบสองทาง

### Abstract

Stock prices have been very volatile, making them difficult to forecast. Using machine learning models with the capability of capturing dynamic behaviors of stock prices could be a solution. This research presents a method for creating short-term and long-term stock price forecasting models using selected deep learning techniques, which are Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), Gated recurrent unit (GRU), and Bidirectional Gated recurrent unit (Bi-GRU). In this work, Time series datasets stock of 3 stock BBL, KBANK, and SCB from 1 January 2018 to 29 November 2021. Then normalize data, split data training datasets, validation datasets and testing datasets, and sliding window methodology for predict close price short-term (1 day in advance) and long-term (22 days in advance). This research uses three input data to build prediction models. The input data include stock price data, stock price data combined with technical data, and stock price data combined with technical data and trading signals. We split the data into training, validation, and test sets. The trained forecasting models were evaluated and compared based on the prediction of the stock closing price. The results showed that short-term forecasting efficacy was not significantly different, while a long-term model third was the most efficient.

**Keywords:** Time series, Sliding Window, Long Short-Term Memory, Bidirectional Long Short-Term Memory, Gated recurrent unit, Bidirectional Gated recurrent unit.

## 1. บทนำ

การพยากรณ์ที่มีประสิทธิภาพจะส่งผลให้การวางแผนและการตัดสินใจในการดำเนินการลงทุนได้อย่างถูกต้อง

ปัจจุบันงานวิจัยด้านการพยากรณ์ราคาหุ้นนั้นมีการนำเสนอในหลาย ๆ เทคนิควิธีการหนึ่งคือการใช้การเรียนรู้ของเครื่อง มาพยากรณ์ เช่น Ojo et al. [1] ได้ให้ข้อเสนอแนะว่าความผันผวนของตลาดหุ้นไม่สามารถวิเคราะห์ได้โดยใช้ข้อมูลราคาหุ้นในอดีตเพียงอย่างเดียว แต่ต้องวิเคราะห์รวมถึงข่าวปัจจุบันของโลก ข่าวการเมือง และเศรษฐกิจที่อาจจะส่งผลต่อพฤติกรรมของนักลงทุน และราคาของหุ้นได้ นอกจากนี้ Shahi et al. [2] ได้ทดลองพยากรณ์ราคาหุ้นโดยใช้ข้อมูลราคาหุ้นในอดีตร่วมกับข้อมูลข่าวสารด้านการเงินสร้างโมเดลโดยใช้เทคนิค LSTM พบว่าประสิทธิภาพของโมเดลที่ใช้ข้อมูลราคาหุ้นในอดีตร่วมกับข้อมูลข่าวสารด้านการเงินมีประสิทธิภาพมากกว่าการสร้างโมเดลที่ใช้ข้อมูลราคาหุ้นในอดีต และ Du et al. [3] ได้ทดลองเปรียบเทียบการพยากรณ์ราคาหุ้นโดยสร้างโมเดลด้วยเทคนิค LSTM โดยแบ่ง 2 รูปแบบ ได้แก่ รูปแบบแรกใช้ข้อมูลราคาปิดเพียงอย่างเดียวในการฝึกฝนโมเดล และรูปแบบที่สองใช้ข้อมูล ราคาเปิดตลาด ราคาสูงสุด ราคาต่ำสุด ราคาปิดตลาด และปริมาณการซื้อขายในการฝึกฝน โมเดลพบว่าในการป้อนข้อมูลคุณสมบัติหลายตัวแปรทำให้โมเดลมีประสิทธิภาพมากขึ้น

จากงานวิจัยที่ได้กล่าวมาขั้นต้นการพยากรณ์ราคาของหุ้นจะมีประสิทธิภาพมากขึ้นเมื่อนำข้อมูลที่เกี่ยวข้องกับตลาดร่วมกับราคาหุ้นในอดีตมาฝึกฝนโมเดล ดังนั้นผู้วิจัยจึงมีแนวคิดในการนำข้อมูลของราคาหุ้น และข้อมูลการวิเคราะห์ปัจจัยทางเทคนิค มาใช้ในการสร้างโมเดลการพยากรณ์ราคาของหุ้นโดยใช้เทคนิคอย่างน้อย 4 เทคนิคมาเปรียบเทียบประสิทธิภาพของโมเดลที่ได้พัฒนาขึ้น

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การวิเคราะห์หุ้นปัจจัยทางเทคนิค (Technical Analysis)

#### 2.1.1 Relative Strength Index (RSI)

Anbalagan et al. [6] RSI เป็นเครื่องมือวัดความเร็วในการเคลื่อนไหวและทิศทางของราคาหุ้นการหาค่า RSI จำเป็นต้องมีค่า RS ก่อนโดยดัง (1)

$$RS = \frac{\text{Average gain}}{\text{Average loss}} \quad (1)$$

เมื่อ Average gain คือ ราคาปิดบวกเฉลี่ย 14 วัน และ

Average loss คือ ราคาปิดลบเฉลี่ย 14 วัน

ค่า RSI สามารถหาได้ดัง (2)

$$RSI = 100 - \frac{100}{(1+RS)} \quad (2)$$

โดย RSI สามารถบอกสัญญาณการซื้อขายได้ดังนี้

กรณี RSI > 70 ขาย (Sell) กรณี 30 <= RSI <= 70 ถือ (Hold) กรณี RSI < 30 เข้าซื้อ (Buy)

#### 2.1.2 Exponential Moving Average (EMA)

Anbalagan et al. [6] EMA เป็นเครื่องมือทำให้เห็นถึงการเปลี่ยนแปลงของราคาโดยสูตร EMA ดัง (3)

$$EMA = EMA_{t-1} + SF(P_t - EMA_{t-1}) \quad (3)$$

เมื่อ EMA คือ ค่าของ Exponential Moving Average ณ

เวลาปัจจุบัน,  $EMA_{t-1}$  คือ ค่าของ Exponential Moving

Average ณ เวลาก่อนหน้า, SF คือ ค่าของตัวถ่วงน้ำหนัก

เท่ากับ  $2/(n+1)$ ,  $P_t$  คือ ราคาปัจจุบัน, n คือ จำนวนวัน

#### 2.1.3 Moving Average Convergence / Divergence (MACD)

Anbalagan et al. [6] เป็นเครื่องมือวัดหาสัญญาณการเข้าซื้อขายคำนวณหาค่า MACD ดัง (4) และ (5)

$$MACD = EMA(12) - EMA(26) \quad (4)$$

$$\text{Signal Line} = EMA(9) \quad (5)$$

โดย MACD สามารถบอกสัญญาณการซื้อขายได้ดังนี้

กรณีเส้น MACD ตัดขึ้นเส้น Signal line คือจุดซื้อ (Buy)

กรณีเส้น MACD ตัดลงเส้น Signal line คือจุดขาย (Sell)

### 2.2 การปรับช่วงขอบเขตของข้อมูล (Normalization)

วิธี Min-Max Scaler มาใช้ในการปรับข้อมูลให้อยู่ในช่วง 0 ถึง 1 ดังสมการที่ 6

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (6)$$

เมื่อ  $x_{scaled}$  คือ ค่าที่ปรับช่วงขอบเขตข้อมูลของตัวแปร  $x$ ,  $\min(x)$  คือ ค่าต่ำสุดของตัวแปร  $x$ ,  $\max(x)$  คือ ค่าสูงสุดของตัวแปร  $x$ ,  $x$  คือ ข้อมูลที่เราสนใจ

### 2.3 หลักการหน้าต่างเคลื่อนที่ (Sliding Window)

ปัทิตญา [7] Sliding Window เป็นหลักการแบ่งข้อมูลในอดีตเพื่อทำนายค่าในอนาคตโดยการกำหนดความกว้างของหน้าต่างและภายในความกว้างของหน้าต่างจะถูกแบ่งเป็น 2 ส่วนได้แก่ชุดข้อมูลการเรียนรู้ และชุดข้อมูลทดสอบ ทำให้สามารถกำหนดช่วงของวันที่ต้องการพยากรณ์ได้เช่นหากจำนวนข้อมูลมีทั้งหมด 8 วันสมมุติถ้าต้องการพยากรณ์ข้อมูล 2 วันข้างหน้า โดยกำหนดความกว้างของหน้าต่างเท่ากับ 5 ในการเลื่อนหน้าต่างรอบที่ 1 ภายในความกว้างของหน้าต่างแบ่งข้อมูล 3 วันแรกเป็นชุดข้อมูลการเรียนรู้และใช้ 2 วันถัดไปเป็นชุดข้อมูลทดสอบ ต่อจากนั้นจะเลื่อนไปจนถึงชุดข้อมูลที่ 8 ซึ่งจะใช้ทั้งหมดจำนวน 4 รอบ

### 2.4 Long Short-Term Memory (LSTM)

LSTM จะใช้ข้อมูล Cell state และ Hidden state ในการเก็บข้อมูลต่าง ๆ ในการคำนวณว่าควรจะเก็บรักษาข้อมูลภายใน Cell state และ Hidden state มากน้อยเพียงใด โดยประกอบด้วย Input gate, Output gate และ Forget gate แต่ละ Gate จะมีหน้าที่ในการตัดสินใจว่าจะให้ข้อมูลผ่านไปหรือไม่ ตามค่าความสำคัญของข้อมูลหากมีค่าน้อยก็จะไม่สามารถผ่าน Gate ไปได้ [8]

### 2.5 Bidirectional Long Short-Term Memory (BiLSTM)

BiLSTM เป็นการนำ LSTM สองเครือข่ายเชื่อมต่อกันเป็น เลเยอร์เดียวกัน โดย LSTM แรกจะทำหน้าที่คำนวณไปข้างหน้า (Forward LSTM) LSTM สองจะทำหน้าที่คำนวณไปข้างหลัง (Backward LSTM) สำหรับการฝึกแต่ละลำดับครั้งทำให้การฝึกแต่ละครั้งข้อมูลจะครบถ้วนมากขึ้น [9]

### 2.6 Gated Recurrent Unit (GRU)

GRU โดยพัฒนาต่อออกจาก LSTM ให้มีโครงสร้างที่ไม่ซับซ้อน โดย การปรับ Gate ใน LSTM เป็น Reset Gate และ Update Gate ซึ่ง Update Gate ทำหน้าที่พิจารณาว่าควรจะเก็บ ข้อมูล State ก่อนหน้าไว้มากน้อยเพียงใด และ Reset Gate ทำหน้าที่คำนวณว่าจะนำข้อมูลจากจาก State ก่อนหน้ามาพิจารณาพร้อมกับข้อมูลนำเข้าปัจจุบันมากน้อยเพียงใด [8]

### 2.7 Bidirectional Gated Recurrent Unit (BiGRU)

BiGRU มีความคำนวณที่เหมือนกับ GRU มีการคำนวณไปข้างหน้า (Forward GRU) และ คำนวณไปข้างหลัง (Backward GRU) แล้วนำผลที่ได้จากการคำนวณไปข้างหน้ากับข้างหลังมารวมกันเป็นผลลัพธ์สุดท้าย ซึ่งจะสามารถใช้ข้อมูลในอดีตและอนาคตได้พร้อมกัน [10]

### 2.8 ฟังก์ชันกระตุ้น (Active Function)

Linear Function มีจะให้ผลลัพธ์ตามค่าที่ส่งมาดัง (7)

$$f(x) = x \quad (7)$$

เมื่อ  $f(x)$  คือ ฟังก์ชันเส้นตรง,  $x$  คือ ข้อมูลนำเข้า

ReLU Function คือค่าที่เป็น ลบ จะทำให้เป็นศูนย์ ค่าเป็นบวก จะให้ผลลัพธ์ตามค่าที่ส่งเข้ามาดัง (8)

$$relu(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (8)$$

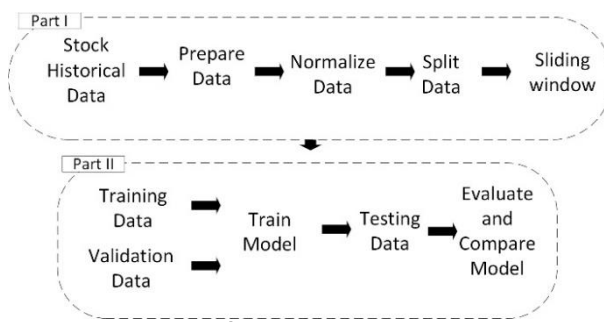
เมื่อ  $relu(x)$  คือ ReLU ฟังก์ชัน,  $x$  คือ ข้อมูลนำเข้า

#### งานวิจัยที่เกี่ยวข้อง

Gao et al. [4] ได้ทดลองวิจัยพยากรณ์ราคาปิดของหุ้น S&P500 โดยใช้เทคนิคหน้าต่างบานเลื่อนโดยกำหนดความกว้างเท่ากับ 20 วันพยากรณ์ราคาปิดของวันถัดไปงานวิจัยนี้ทดลอง 3 เทคนิคได้แก่ LSTM Moving average (MA), Exponential moving average (EMA) และ Support vector machine (SVM) พบว่าเทคนิค LSTM มีประสิทธิภาพสูงที่สุดที่ทดลองลงมา SVM EMA และ MA ตามลำดับฟังก์ชันกระตุ้นมีผลต่อประสิทธิภาพในการพยากรณ์ราคาหุ้น Rana et al. [5] ได้ทดลองพยากรณ์ราคาหุ้นโดยใช้เทคนิค LSTM โดยงานวิจัยนี้สนใจทดลองฟังก์ชันกระตุ้นกับตัวปรับพารามิเตอร์โมเดลพบว่าโมเดล LSTM ที่ใช้ฟังก์ชันกระตุ้นเป็น Linear function กับใช้ตัวปรับพารามิเตอร์เป็น Adamax optimizer และ ฟังก์ชันกระตุ้นเป็น tanh function กับใช้ตัวปรับพารามิเตอร์เป็น Adam optimizer ทั้งสอง โมเดลนี้ให้ความแม่นยำถึงร้อยละ 98.49

ผู้วิจัยจะนำแนวคิดกับทฤษฎีที่กล่าวมาข้างต้นเพื่อมาประยุกต์ใช้ในขั้นตอนการดำเนินงานวิจัยดังต่อไปนี้

### 3. วิธีดำเนินการวิจัย



ภาพที่ 1: วิธีดำเนินการวิจัย

จากภาพที่ 1 แสดงวิธีดำเนินการวิจัยโดยจะประกอบด้วย 2 ขั้นตอนหลักได้แก่ การเตรียมข้อมูลอธิบายไว้ในข้อ 3.1 ถึง 3.5 กับการทำโมเดลอธิบายไว้ในข้อ 3.6 ถึง 3.7 โดยมีรายละเอียดดังนี้

#### 3.1 ข้อมูลที่นำมาใช้

งานวิจัยนี้ ได้ใช้ข้อมูลราคาของหุ้นบนเว็บไซต์ Yahoo Finance ซึ่งจะเป็นข้อมูลในรูปแบบอนุกรมเวลา (Time Series) โดยใช้ข้อมูลหุ้น 3 ตัวได้แก่ BBL, KBANK และ SCB ตั้งแต่วันที่ 1 มกราคม 2561 ถึง วันที่ 29 พฤศจิกายน 2564 ข้อมูลจะประกอบไปด้วย 6 คุณลักษณะ ได้แก่ ราคาเปิด ราคาปิด ราคาสูงสุด ราคาต่ำสุด ปริมาณการซื้อขาย และเวลาดำเนินการซื้อขาย ทั้งหมด 951 แถว

#### 3.2 การเตรียมข้อมูล

##### 3.2.1 การทำดัชนี (Index) ข้อมูล

ข้อมูลจากข้อ 3.1 ทำการแทนค่าข้อมูลวันที่ด้วยดัชนี โดยดัชนีจะเรียงลำดับตามวันที่จากก่อนไปถึงหลัง

##### 3.2.2 การวิเคราะห์หุ้นปัจจัยทางเทคนิค (Technical Analysis)

จากข้อมูลในข้อ 3.2.1 ผู้วิจัยคำนวณหาค่า RSI กับค่า MACD และวิเคราะห์ข้อมูลสัญญาณการซื้อขายจาก 2 เทคนิคที่กล่าวมา ข้อมูลของเราจะประกอบไปด้วยคุณสมบัตินี้ ได้แก่ ราคาเปิด, ราคาปิด, ราคาสูงสุด, ราคาต่ำสุด, ปริมาณการซื้อขาย, RSI, สัญญาณการเข้าซื้อขายของ RSI, EMA 9, EMA 12, EMA 26 และสัญญาณการเข้าซื้อขายของ MACD

#### 3.3 การปรับช่วงขอบเขตของข้อมูล (Normalization)

ผู้วิจัยได้ใช้วิธีการ Min-Max Scaler ในการปรับข้อมูล

#### 3.4 การแบ่งข้อมูล

ข้อมูลจากข้อ 3.3 ผู้วิจัยจะแบ่งข้อมูลเป็น 3 ชุดได้แก่ชุดข้อมูลฝึกฝนร้อยละ 80 ชุดข้อมูลตรวจสอบร้อยละ 10 และชุดข้อมูลทดสอบร้อยละ 10

#### 3.5 การคำนวณแบบหน้าต่างเคลื่อนที่ (Sliding Window)

งานวิจัยนี้ผู้ได้กำหนดความกว้างของข้อมูลฝึกฝนเท่ากับ 30 วันและพยากรณ์ข้อมูล 1 กับ 22 วันข้างหน้า โดยข้อมูลที่แบ่งจากข้อ 3.4 จะถูกเปลี่ยนแปลงโดยใช้เทคนิคหน้าต่างเคลื่อนที่ทั้งหมด

#### 3.6 การสร้างโมเดล

##### 3.6.1 รูปแบบของโมเดล

งานวิจัยนี้แบ่งรูปแบบของโมเดลตามคุณสมบัติของข้อมูลนำเข้าไว้ 3 รูปแบบได้แก่ 1. ใช้ข้อมูลคุณสมบัติพื้นฐานของหุ้นได้แก่ ราคาเปิด ราคาปิด ราคาสูงสุด และราคาต่ำสุด มาใช้เป็นข้อมูลในการฝึกฝนโมเดล 2. ใช้รูปแบบที่ 1 เพิ่มคุณสมบัติ 5 คุณสมบัติได้แก่ ปริมาณการซื้อขาย, RSI, EMA 9, EMA 12 และ EMA 26 มาใช้เป็นข้อมูลในการฝึกฝนโมเดล และ 3. แบ่งโมเดลเป็น 2 ชนิดคือชนิดแรกใช้ข้อมูลของโมเดลรูปแบบที่ 1 ชนิดที่สองใช้ข้อมูลโมเดลรูปแบบที่ 2 เพิ่มคุณสมบัติของข้อมูล 2 ตัวได้แก่สัญญาณการซื้อขายของ MACD กับสัญญาณการซื้อขายของ RSI มาใช้เป็นข้อมูลในการฝึกฝนโมเดล

##### 3.6.2 โครงสร้างของโมเดล

จากข้อ 3.6.1 งานวิจัยนี้กำหนดชั้นเลเยอร์แรกของโมเดลแต่ละชนิดด้วยเทคนิค 4 เทคนิคได้แก่ LSTM, BiLSTM, GRU และ BiGRU โดยโครงสร้างรูปแบบโมเดลที่ 1 กับรูปแบบโมเดลที่ 2 กำหนดชั้น เลเยอร์แรกใช้เทคนิคที่กล่าวมาข้างต้นและ 2 ชั้นเลเยอร์ต่อมาใช้หน่วยการเรียนรู้เชิงลึกโดยกำหนดค่าพารามิเตอร์ตารางที่ 1

ตารางที่ 1: ข้อมูลพารามิเตอร์ของโมเดลรูปแบบที่ 1 กับ 2

Layer	Hyperparameter
1	500 Neurons, activation function: ReLu
2	100 Neurons, activation function: ReLu
3	Period predict Neurons, activation function: ReLu

โครงสร้างโมเดลรูปแบบที่ 3 จะประกอบไปด้วย 2 โมเดลคือโมเดลแรกกำหนดชั้นเลเยอร์กำหนดหน่วยการเรียนรู้เชิงลึกทั้ง 3 ชั้นเลเยอร์กำหนดค่าพารามิเตอร์ดังตาราง 2

**ตารางที่ 2:** พารามิเตอร์ โมเดลรูปแบบที่ 3 ใช้ข้อมูลเชิงเทคนิคเป็นข้อมูลนำเข้า

Layer	Hyperparameter
1	500 Neurons, activation function: ReLu, Dropout Layer (0.1 ratio)
2	250 Neurons, activation function: ReLu, Dropout Layer (0.1 ratio)
3	Period predict Neurons, activation function: Linear

โมเดลสองกำหนดชั้นเลเยอร์แรกใช้เทคนิคที่กล่าวมาข้างต้นและ 2 ชั้นเลเยอร์ต่อมาจะใช้หน่วยการเรียนรู้เชิงลึก โดยกำหนดค่าพารามิเตอร์ดังตาราง 3

**ตารางที่ 3:** พารามิเตอร์ โมเดลรูปแบบที่ 3 ใช้ข้อมูลหุ้นเป็นข้อมูลนำเข้า

Layer	Hyperparameter
1	500 Neurons, activation function: ReLu
2	250 Neurons, activation function: ReLu, Dropout Layer (0.1 ratio)
3	Period predict Neurons, activation function: Linear

จากนั้นนำโมเดลที่ 1 กับ 2 มาต่อกันก่อนที่ได้ผลลัพธ์กำหนด ชั้นเลเยอร์กำหนดหน่วยการเรียนรู้เชิงลึก 3 ชั้นเลเยอร์ กำหนดค่าพารามิเตอร์ดังตาราง 4

**ตารางที่ 4:** พารามิเตอร์ โมเดลรูปแบบที่ 3 หลังจากการต่อกัน

Layer	Hyperparameter
1	500 Neurons, activation function: ReLu, Dropout Layer (0.1 ratio)
2	250 Neurons, activation function: ReLu, Dropout Layer (0.1 ratio)
3	Period predict Neurons, activation function: Linear

3.6.3 เทคนิคที่ใช้ในการสอนโมเดล

จากข้อ 3.6.2 จะได้โมเดลพร้อมที่จะไปทำการสอน โดยงานวิจัยนี้กำหนดตัวปรับพารามิเตอร์แบบจำลอง (Optimizer) ด้วยวิธี Adam Optimizer ด้วย Learning Rate เท่ากับ 0.0008 และ กำหนดหยุดการสอนโมเดล (Early stopping) โดยวัดจากค่า Loss Function จากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE)

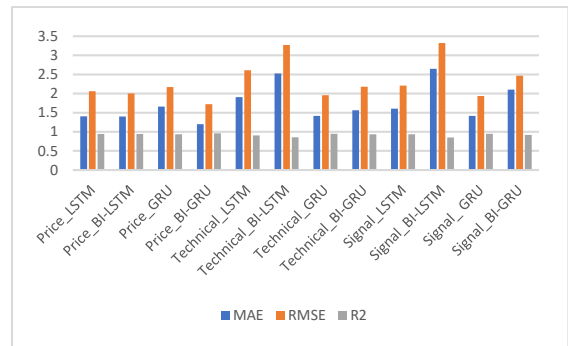
3.7 การวัดประสิทธิภาพของโมเดล

งานวิจัยนี้ทำการเปรียบเทียบรูปแบบของโมเดลทั้ง 3 รูปแบบโดยใช้ชุดข้อมูลทดสอบที่ได้แบ่งไว้ในข้อ 3.4 มาทำการทดสอบและวัดค่า 3 ค่าได้แก่ ค่าราคาที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute

Error: MAE) และค่าสัมประสิทธิ์การกำหนด (The Coefficient of determination) หรือ R<sup>2</sup>

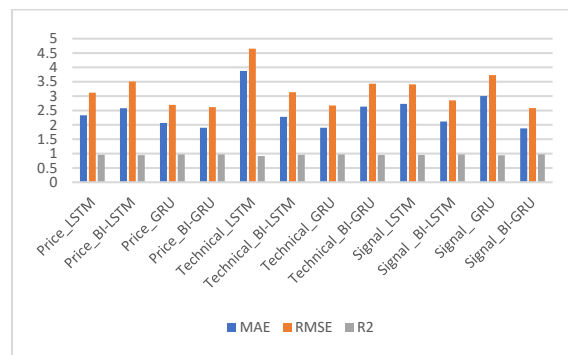
4. ผลการดำเนินการวิจัย

จากการวัดผลประสิทธิภาพของโมเดลทั้ง 3 รูปแบบโดยกำหนดความกว้างของหน้าต่างเท่ากับ 30 วันและพยากรณ์ราคาปิด 1 กับ 22 วันข้างหน้า โดยกำหนดชื่อของโมเดลที่แสดงในกราฟภาพที่ 2- 7 ดังนี้ โมเดลรูปแบบที่ 1 ชื่อขึ้นต้นด้วย Price\_ ลงท้ายด้วยชื่อเทคนิค โมเดลรูปแบบที่ 2 ชื่อขึ้นต้นด้วย Technical\_ ลงท้ายด้วยชื่อเทคนิค และ โมเดลรูปแบบที่ 3 ชื่อขึ้นต้นด้วย Signal\_ ลงท้ายด้วยชื่อเทคนิค



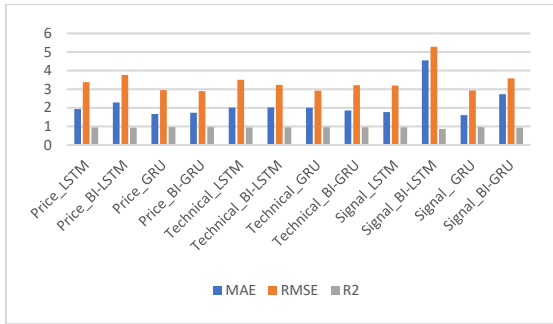
**ภาพที่ 2:** ประสิทธิภาพของ โมเดลใช้ข้อมูลหุ้น BBL พยากรณ์ จำนวน วันข้างหน้า 1

จากภาพที่ 2 โมเดลรูปแบบที่ 1 ใช้อัลกอริทึม BiGRU มีประสิทธิภาพสูงที่สุด โดยค่า MAE เท่ากับ 1.1995 RMSE เท่ากับ 1.7239 และค่า R<sup>2</sup> เท่ากับ 0.9597



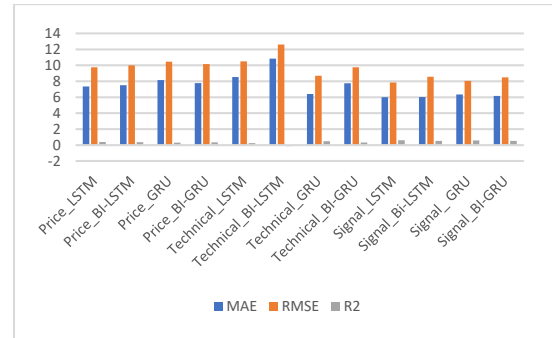
**ภาพที่ 3:** ผลการประเมินประสิทธิภาพของ โมเดลใช้ข้อมูลหุ้น KBANK พยากรณ์จำนวน วันข้างหน้า 1

จากภาพที่ 3 โมเดลรูปแบบที่ 3 ใช้อัลกอริทึม BiGRU มีประสิทธิภาพสูงที่สุด โดยค่า MAE เท่ากับ 1.8754 RMSE เท่ากับ 2.5864 และค่า R<sup>2</sup> เท่ากับ 0.9724



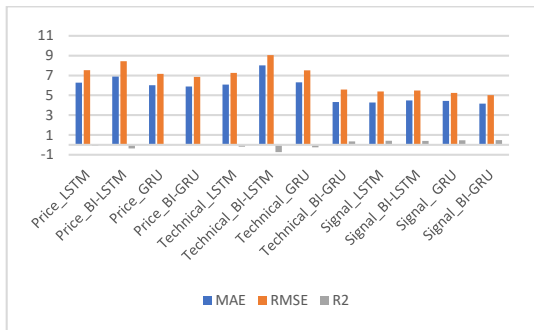
ภาพที่ 4: ผลการประเมินประสิทธิภาพของโมเดลใช้ข้อมูลหุ้น SCB พยากรณ์จำนวน วันข้างหน้า 1

จากภาพที่ 4 โมเดลรูปแบบที่ 1 ใช้อัลกอริทึม BiGRU มีประสิทธิภาพสูงที่สุด โดยค่า MAE เท่ากับ 1.7329 RMSE เท่ากับ 2.8971 และค่า R<sup>2</sup> เท่ากับ 0.9614



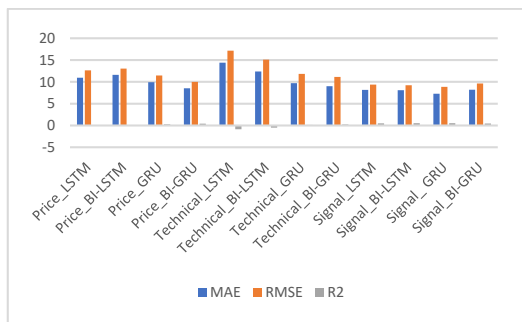
ภาพที่ 7: ผลการประเมินประสิทธิภาพของโมเดลใช้ข้อมูลหุ้น SCB พยากรณ์จำนวน 22 วันข้างหน้า

จากภาพที่ 7 โมเดลรูปแบบที่ 3 ใช้อัลกอริทึม LSTM มีประสิทธิภาพสูงที่สุด โดยค่า MAE เท่ากับ 5.9968 RMSE เท่ากับ 7.8501 และค่า R<sup>2</sup> เท่ากับ 0.6045



ภาพที่ 5: ผลการประเมินประสิทธิภาพของโมเดลใช้ข้อมูลหุ้น BBL พยากรณ์จำนวน 22 วันข้างหน้า

จากภาพที่ 5 โมเดลรูปแบบที่ 3 ใช้อัลกอริทึม BiGRU มีประสิทธิภาพสูงที่สุด โดยค่า MAE เท่ากับ 4.162 RMSE เท่ากับ 5.0191 และค่า R<sup>2</sup> เท่ากับ 0.4779



ภาพที่ 6: ผลการประเมินประสิทธิภาพของโมเดลใช้ข้อมูลหุ้น KBANK พยากรณ์จำนวน 22 วันข้างหน้า

จากตารางที่ 6 โมเดลรูปแบบที่ 3 ใช้อัลกอริทึม GRU มีประสิทธิภาพสูงที่สุด โดยค่า MAE เท่ากับ 7.2871 RMSE เท่ากับ 8.8589 และค่า R<sup>2</sup> เท่ากับ 0.5414

### 5. สรุป

งานวิจัยนี้แนะนำเสนอเกี่ยวกับการเปรียบเทียบประสิทธิภาพการพยากรณ์ราคาปิดหุ้นด้วยเทคนิคการเรียนรู้เชิงลึกโดยเลือกเทคนิคที่ต่างกัน 4 เทคนิค และรูปแบบของโมเดลที่ต่างกัน 3 รูปแบบ เพื่อพยากรณ์ราคาหุ้นระยะสั้น และระยะยาว ตัวบ่งชี้ MAE, RMSE และ R<sup>2</sup> ถูกนำไปใช้ในการประเมินประสิทธิภาพของการพยากรณ์ โดยค่า MAE กับ RMSE มีค่าน้อยแสดงให้เห็นว่าโมเดลที่แนะนำมีค่าความคลาดเคลื่อนต่ำ ในขณะที่ R<sup>2</sup> มีค่าเข้าใกล้ 1 แสดงถึงค่าความคลาดเคลื่อนต่ำ จากการทดลองพยากรณ์ระยะสั้นประสิทธิภาพของการพยากรณ์ไม่แตกต่างกันอย่างมีนัยสำคัญเนื่องจากค่า MAE, RMSE และ R<sup>2</sup> ไม่แตกต่างกันมาก เปรียบเทียบการพยากรณ์ระยะสั้นจะมีประสิทธิภาพสูงกว่าระยะยาวค่อนข้างมาก เมื่อพิจารณาการพยากรณ์ระยะยาวค่าความคลาดเคลื่อนของโมเดลที่ใช้ข้อมูลสัญญาณและมีโครงสร้างของการรวมโมเดล มีค่าความคลาดเคลื่อนต่ำกว่าโมเดลที่ใช้ราคาพื้นฐานหุ้นและโมเดลที่ใช้ราคาพื้นฐานหุ้นร่วมกับปัจจัยทางเทคนิคที่มีโครงสร้างชั้นเลเยอร์ 3 ชั้นอย่างเห็นได้ชัดเจน สามารถสรุปได้ว่าการนำข้อมูลเชิงเทคนิคมาเป็นข้อมูลนำเข้า และโครงสร้างของการรวมโมเดลสามารถเพิ่มประสิทธิภาพในการพยากรณ์หุ้นระยะยาวได้

## 6. กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณ รศ.ดร. พยุง มีสัง ที่ให้คำแนะนำในการดำเนินการวิจัย และคณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ที่ได้อำนวยความสะดวกสำหรับสถานที่ในการทำวิจัย

### เอกสารอ้างอิง

- [1] S. O. Ojo, P. A. Owolawi, M. Mphahlele, and J. A. Adisa, "Stock Market Behaviour Prediction using Stacked LSTM Networks\*," in *2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, Nov. 2019, pp. 1–5.
- [2] T. B. Shahi, A. Shrestha, A. Neupane, and W. Guo, "Stock Price Forecasting with Deep Learning: A Comparative Study," *Mathematics*, vol. 8, no. 9, Art. no. 9, Sep. 2020.
- [3] J. Du, Q. Liu, K. Chen, and J. Wang, "Forecasting stock prices in two ways based on LSTM neural network," in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Mar. 2019, pp. 1083–1086.
- [4] T. Gao, Y. Chai, and Y. Liu, "Applying long short term memory neural networks for predicting stock closing price," in *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2017, pp. 575–578.
- [5] M. Rana, Md. M. Uddin, and Md. M. Hoque, "Effects of Activation Functions and Optimizers on Stock Price Prediction using LSTM Recurrent Networks," in *Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence*, New York, NY, USA, 2019, pp. 354–358.
- [6] T. Anbalagan and S. U. Maheswari, "Classification and Prediction of Stock Market Index Based on Fuzzy Metagraph," *Procedia Computer Science*, vol. 47, pp. 214–221, 2015.
- [7] ปัทมฉญา บุญรักษา และ จารีย์ ทองคำ "Predicting the Occurrence of Accidents on the Road using Time Series Techniques," Thesis, Mahasarakham University, 2018. Accessed: Feb. 22, 2022.
- [8] T. Phaladisailoed and T. Numnonda, "การเปรียบเทียบโมเดลการเรียนรู้ของเครื่องแบบต่าง ๆ สำหรับการทำนายราคาบิทคอยน์," *วารสารเทคโนโลยีสารสนเทศลาดกระบัง*, vol. 6, no. 1, Art. no. 1, Nov. 2017, Accessed: Aug. 07, 2021.
- [9] Md. A. Istiaque Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, 2020, pp. 87–92.
- [10] Q. Zhou, C. Zhou, and X. Wang, "Stock prediction based on bidirectional gated recurrent unit with convolutional neural network and feature selection," *PLOS ONE*, vol. 17, no. 2, p. e0262501, 2022.

# การเปรียบเทียบโมเดลการเรียนรู้ของเครื่องสำหรับการคัดแยกผู้ป่วยโรคโควิด-19

## A Comparison of Machine Learning Models for COVID-19 Patients Screening

ศิวกร บรรลือทรัพย์ (Siwakorn Banluesapy)<sup>1</sup> และวราพร จิระพันธุ์ทอง (Waraporn Jirapanthong)<sup>2</sup>  
 สาขาวิชาวิศวกรรมเว็บและการพัฒนาแอปพลิเคชันบนอุปกรณ์พกพา วิทยาลัยครีเอทีฟดีไซน์แอนด์เอ็นเตอร์เทนเมนต์  
 เทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิตย์  
<sup>1</sup>siwakorn167@gmail.com, <sup>2</sup>waraporn.jir@dpu.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการทำนายการคัดแยกผู้ป่วยโควิด-19 โดยทำการศึกษาข้อมูลผู้ป่วย จำนวน 1,608,923 ราย จากกรมควบคุมโรค ทำการรวบรวมข้อมูล 1 มกราคม 2563 ถึง วันที่ 1 ตุลาคม 2564 โดยใช้เทคนิคการจำแนกข้อมูลทั้งหมด 3 เทคนิค ได้แก่ Random Forest, Neural network และ Naive Bayes ซึ่งทำการเปรียบเทียบประสิทธิภาพรูปแบบเทคนิคการทำนายระหว่างคุณลักษณะ การทดสอบประสิทธิภาพตัวแบบทำนายด้วยวิธีการ Split Test คือ การแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วนคือ Training Data (ร้อยละ 80) และ Test Data (ร้อยละ 20) โดยใช้โปรแกรม Orange Canvas จากนั้นทำการทดลองเพื่อหาผลการทดสอบประสิทธิภาพที่มีค่าความถูกต้องที่สูงที่สุด

ผลการวิจัยพบว่า Random Forest มีค่าความถูกต้องเท่ากับ 93.51% Neural network มีค่าความถูกต้องเท่ากับ 93.02% Naive Bayes มีค่าความถูกต้องเท่ากับ 27.54% จากผลการเปรียบเทียบในครั้งนี้สามารถนำ Random Forest ที่มีค่าความถูกต้องมากที่สุดไปใช้ในการพยากรณ์การคัดแยกผู้ป่วยโควิด-19 ได้ต่อไป

**คำสำคัญ:** การเรียนรู้ของเครื่อง ป่าสุ่ม โครงข่ายประสาทเทียม นาอิมเบย์ โควิด-19

### Abstract

The purpose of this research is to compare the effectiveness of algorithms in screening of COVID-19 patients. According to patients' data from the

department of disease control (1,608,923 patients), the data has been collected from January, 1st 2020 to October, 1st 2021. All three classification techniques were used: Random Forest, Neural network and Naive Bayes. We compare the performance of predictive techniques between features. We apply split test method to evaluate the performance of predictive techniques. The collection of data is randomly divided into two parts: Training Data (80%) and Test Data (20%). Additionally, we apply Orange Canvas program to evaluate the most accurate results.

The results show that Random Forest has an accuracy of 93.51%. Neural network has an accuracy of 93.02%. Naive Bayes has an accuracy of 27.54%. This presents the Random Forest with the highest accuracy figure for screening of COVID-19 patients.

**Keywords:** Machine Learning, Random Forest, Neural Network, Naive Bayes, Covid-19.

### 1. บทนำ

องค์การอนามัยโลกได้รับแจ้งเมื่อเดือนธันวาคม พ.ศ. 2562 ว่าพบกลุ่มผู้ป่วยมีอาการปอดบวมโดยไม่ทราบสาเหตุในเมืองอู่ฮั่น มณฑลหูเป่ย์ สาธารณรัฐประชาชนจีน ต่อมา มีรายงานว่าพบเชื้อไวรัสโคโรนาสายพันธุ์ใหม่หรือ Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-2) เป็นสาเหตุของโรคและได้แพร่ระบาดไปยังประเทศอื่น ๆ โดยโรคโควิด-19 เป็นโรคติดต่อสายพันธุ์ใหม่ และได้มีการคาดการณ์ว่าเป็นการติดเชื้อจาก



สัตว์สู่คน โดยมีค้างคาวเป็นพาหะเป็นโรคได้ทั้งคนและสัตว์ [1], [2] เชื้อสามารถติดต่อจากคนสู่คนผ่านทางเดินหายใจจากละอองของเสมหะจากการ ไอ จาม น้ำมูก น้ำลายในระยะประมาณ 1-2 เมตร [3]

สำหรับการระบาดของเชื้อไวรัสโควิด-19 รอบแรกในประเทศไทย พบผู้ป่วยต้องสงสัยรายแรกเมื่อวันที่ 21 มกราคม พ.ศ. 2563 เป็นนักท่องเที่ยวหญิงชาวจีนอายุ 74 ปี ซึ่งเดินทางมาถึงกรุงเทพมหานคร โดยเที่ยวบินจากนครอู่ฮั่น เมืองหลวงของมณฑลหูเป่ย์ สาธารณรัฐประชาชนจีน และ ณ วันที่ 8 เมษายน พ.ศ. 2563 พบอัตราการติดเชื้อในประเทศไทย จำนวน 2,369 คน [4] หลังจากนั้นพบการระบาดระลอกใหม่ช่วงปลายปี 2563 เริ่มตั้งแต่วันที่ 17 ธันวาคม 2563 พบหญิงไทย อายุ 67 ปี อาชีพค้าขายที่ตลาดกลางกุ้ง ในตำบลมหาชัยอำเภอมหาชัย จังหวัดสมุทรสาคร [5] จากรายงานสำรวจของผู้ติดเชื้อตั้งแต่เกิดการระบาดครั้งแรกจนถึงเดือน ตุลาคม 2564 มียอดติดเชื้อรวมที่ยืนยันแล้ว 1,730,364 ราย ยอดเสียชีวิต 17,835 ราย [6]

กระบวนการตรวจ วินิจฉัยโรคโควิด 19 เป็นการตรวจทางห้องปฏิบัติการ ต้องเสียเวลาในการเตรียมเชื้อและกระบวนการตรวจ เพราะต้องตรวจกับตัวเชื้อจริง ในช่วงแรกของการระบาด ต้องใช้เวลาไม่น้อยกว่า 2 วัน และทำได้เฉพาะในโรงพยาบาลใหญ่หรือโรงพยาบาลศูนย์ ค่าตรวจมีราคาแพง เพราะผลที่ได้ต้องได้ผลชัดเจน และทางกระทรวงสาธารณสุขมีข้อจำกัดในเรื่องจำนวนการตรวจ จึงต้องคัดกรองผู้ต้องสงสัยก่อนเก็บเชื้อส่งตรวจ และใช้วิธีนี้ตั้งแต่เริ่มการระบาด คือ ต้องเก็บเชื้อจากสารคัดหลั่งที่โพรงจมูก ดังนั้นในกระบวนการควบคุมโรค การค้นหาการตรวจหาผู้ป่วย จำเป็นต้องมีกระบวนการคัดกรองเพื่อคัดกรองเป็นลำดับในการหาผู้สงสัยก่อน [7] ดังนั้นผู้วิจัยจึงทำการศึกษาและวิเคราะห์ข้อมูลผู้ติดเชื้อ โควิด-19 จากกรมควบคุมโรคตั้งแต่ระลอกที่ 1-3 ซึ่งเป็นข้อมูลแบบเปิดโดยทำการรวบรวมข้อมูลตั้งแต่วันที่ 1 มกราคม 2563 ถึงวันที่ 1 ตุลาคม 2564 เป็นจำนวน 1,608,923 ราย (กรมควบคุมโรค, 2021) ซึ่งเป็นข้อมูลในลักษณะไฟล์คอมพิวเตอร์ (.csv) และ รูปแบบ json data api เพื่อช่วยคัดแยกประเภทผู้ป่วยได้เร็วขึ้น

การเรียนรู้ของเครื่อง (Machine Learning) [8] คือ การที่สามารถทำให้คอมพิวเตอร์สามารถเรียนรู้ได้ด้วยตนเอง โดยการเขียนโปรแกรมหรือการสร้างอัลกอริทึมที่สามารถเรียนรู้และพยากรณ์ข้อมูลได้ ทั้งนี้อัลกอริทึมจะต้องอาศัยโมเดลที่ได้จากการสร้างชุดของข้อมูลเพื่อพยากรณ์หรือตัดสินใจ

งานวิจัยนี้จึงสนใจวิเคราะห์ข้อมูลโควิด-19 โดยการคัดแยกประเภทผู้ป่วยโดยใช้การเรียนรู้ของเครื่อง เพื่อคัดกรองผู้ต้องสงสัยได้เร็วขึ้น และได้เตรียมการรักษาได้ทันเวลา

### วัตถุประสงค์การวิจัย

1. เพื่อศึกษาและสร้างโมเดลการเรียนรู้ของเครื่อง (Machine Learning Model) ที่มีประสิทธิภาพในการคัดแยกประเภทผู้ป่วยโดยพิจารณาจากค่าความแม่นยำที่สุดที่ได้จากโมเดล ระหว่าง 3 เทคนิค คือ Random Forest, Neural network และ Naive Bayes

2. เพื่อวิเคราะห์ ออกแบบและพัฒนาระบบตัดสินใจในการคัดแยกประเภทผู้ป่วยได้อย่างมีประสิทธิภาพ

### สมมติฐานการวิจัย

แบบจำลองเรียนรู้จากเครื่องสำหรับการคัดกรองข้อมูลผู้ป่วย โควิด แบบ เทคนิค Random Forest, Neural network และ Naive Bayes ให้ผลความถูกต้องที่แตกต่างกันอย่างมีนัยสำคัญ

### ขอบเขตการวิจัย

ศึกษาและทำการวิเคราะห์ข้อมูลเพื่อคัดแยกประเภทผู้ป่วย โควิด-19 จาก 1 มกราคม 2563 ถึง วันที่ 1 ตุลาคม 2564 เป็นจำนวน 1,608,923 ราย จากกรมควบคุมโรค เว็บไซต์ <https://covid19.ddc.moph.go.th/> [9]

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การประยุกต์ใช้การเรียนรู้ของเครื่องสำหรับโรคโควิด-19

ปัจจุบันมีการพัฒนาอย่างต่อเนื่องในเทคโนโลยีการเรียนรู้ของเครื่องและปัญญาประดิษฐ์ในงานการแพร่ระบาดของโรคโควิด-19 เพื่อวิเคราะห์การใช้จ่าย การคัดกรอง การคาดการณ์ การติดตามผู้สัมผัส ตลอดจนกระบวนการพัฒนาวัคซีนของโรคโควิด-19 อย่างมี

นัยสำคัญ เพื่อลดการแทรกแซงของมนุษย์ในการปฏิบัติการทางการแพทย์ แต่อย่างไรก็ตามโมเดลส่วนใหญ่ไม่ได้ถูกนำไปใช้งานมากพอที่จะสามารถแสดงการปฏิบัติงานในโลกแห่งความเป็นจริง แต่ก็ยังสามารถใช้รับมือการแพร่ระบาดของโรคโควิด-19 ได้ [10]

**2.2 สถานการณ์ผู้ติดเชื้อ COVID-19 อพเททรายวันรูปแบบ API (Json/CSV Data Format)**

เป็นข้อมูลแบบเปิดที่รวบรวมโดยกรมควบคุมโรคจากกระทรวงสาธารณสุข [9] ซึ่งรายละเอียดจะปรากฏในรูปแบบพจนานุกรมข้อมูล (Data dictionary) ดังตารางที่ 1

ตารางที่ 1: พจนานุกรมข้อมูลข้อมูลผู้ป่วยโควิด-19

Colum Name	Data Type	Comment	Status
txn_date	date	วันแถลง	remove
gender	varchar	เพศ	category
age_number	varchar	อายุ	remove
age_range	varchar	ช่วงอายุ	category
nationality	varchar	สัญชาติ	category
job	varchar	อาชีพ	remove
risk	varchar	ปัจจัยเสี่ยง	category
patient_type	text	ประเภทผู้ป่วย	target
province	varchar	จังหวัด	category
update_date	datetime	วันที่ปรับปรุงข้อมูลล่าสุด	remove

ตารางที่ 2: Class ในการวิจัย

ประเภทของผู้ป่วย	
Class	Description
1	การเฝ้าระวังในกลุ่มผู้ป่วยหรือผู้มีอาการเข้าได้กับนิยามผู้สงสัยติดเชื้อไวรัสโคโรนา 2019 ที่เข้าเกณฑ์สอบสวนโรค (patient under investigation : PUI)
2	การตรวจคัดกรองในประชากรเสี่ยงตามจุดคัดกรองและด่านเข้าออกระหว่างประเทศ (screening)
3	การเฝ้าระวังในกลุ่มเป้าหมายเฉพาะหรือพื้นที่เฉพาะ (sentinel surveillance)
4	การเฝ้าระวังเหตุการณ์ในสถานที่เสี่ยง เก็บตัวอย่างส่งตรวจเมื่อเข้านิยาม PUI หรือเป็นกลุ่มก้อนของผู้ป่วยทางเดินหายใจ

**2.3 การเรียนรู้ของเครื่อง (Machine Learning)**

การเรียนรู้ของเครื่อง (Machine Learning) เป็นหนึ่งในประเภทของปัญญาประดิษฐ์ ที่จะช่วยให้แอปพลิเคชันหรือซอฟต์แวร์เกิดการเรียนรู้จากชุดข้อมูลที่จะสามารถทำนายผลได้อย่างแม่นยำ โดยหลักการคือจะมีโมเดลทางคณิตศาสตร์และมีอัลกอริทึมในหลากหลายรูปแบบมาเรียนรู้จากชุดข้อมูล (Data set) จนทำให้คอมพิวเตอร์เกิดการเรียนรู้และฉลาดในการตัดสินใจบนข้อมูลที่ไม่เคยได้เห็นมาก่อนได้สำเร็จ [11]–[13] ประเภทการเรียนรู้ของเครื่องประกอบด้วย เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) เทคนิคในประเภทนี้จะเน้นการเรียนรู้จากข้อมูลที่มีอยู่ในอดีตเพื่อนำมาสร้างโมเดลสำหรับทำนายหรือคาดการณ์สิ่งที่จะเกิดขึ้นในอนาคต [14], [15] และเทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เทคนิคนี้จะเน้นการพิจารณาข้อมูลเป็นหลัก เช่น พิจารณาความสัมพันธ์ของข้อมูลกันในลักษณะแบบใดบ้าง โดยเทคนิคประเภทนี้จะแบ่งย่อยได้อีกคือ เทคนิคการค้นหากฎความสัมพันธ์ [15]

**2.4 Random Forest**

หลักการของ Random Forest คือการสร้าง model จาก Decision Tree หลายๆ model ย่อยๆ (ตั้งแต่ 10 model ถึง มากกว่า 1000 model) โดยแต่ละ model จะได้รับ data set ไม่เหมือนกัน ซึ่งเป็น subset ของ data set ทั้งหมด ตอนทำ prediction จะให้ Decision Tree ทำการ prediction ในแต่ละโหนด และคำนวณผล prediction ด้วยการ vote output ที่ ถูกเลือกโดย Decision Tree มากที่สุด [16]

**2.5 Naive Bayes**

ใช้การคำนวณความน่าจะเป็นซึ่งถูกใช้ในการทำนายผลจัดเป็นเทคนิคในการแก้ปัญหาแบบ Classification ที่สามารถคาดการณ์ผลลัพธ์ได้และสามารถอธิบายได้ โดยจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร เพื่อใช้ในการสร้างเงื่อนไข ความน่าจะเป็น สำหรับแต่ละความสัมพันธ์ [14]

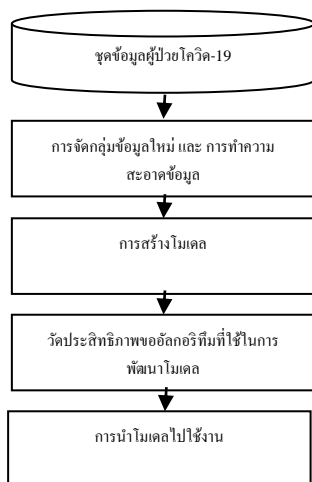
**2.6 Neural network**

ลักษณะทั่วไปของโครงข่ายประสาทเทียม คือ โหนดจำลองมาจากรอยประสานประสาท (Synapse) ระหว่างใยประสาทนำเข้า (Dendrite) ของเซลล์ประสาทตัวหนึ่งและแกนประสาทนำออก (Axon) ของเซลล์ประสาทอีกตัว

หนึ่ง โดยมีฟังก์ชันการแปลงเป็นตัวกำหนดสัญญาณส่งออก องค์ประกอบในการทำงานของระบบประสาทในสมองมนุษย์กับโครงข่ายประสาทเทียมมีความคล้ายคลึงกัน ในระบบประสาทของมนุษย์การรับกระแสประสาทจะอาศัยเดนไดรต์ (Dendrites) เป็นตัวนำพา เปรียบเทียบได้กับหน่วยข้อมูลป้อนเข้าในระบบโครงข่ายประสาทเทียม ส่วนการนำกระแสประสาทไปยังเซลล์ประสาทอื่นผ่านทางแอกซอน ที่เป็นตัวส่งออกกระแสประสาทเปรียบเทียบกับหน่วยผลลัพธ์ในโครงข่ายประสาทเทียม [17]

### 3. วิธีการดำเนินการวิจัย

ผู้วิจัยได้ทำการแบ่งการวิจัยออกเป็น 5 ระยะเวลา ประกอบด้วย Dataset, Data Preparation, Modeling, Evaluation และ Deployment มีรายละเอียดดังนี้



ภาพที่ 1: กรอบแนวคิดงานวิจัย

#### 3.1 Dataset (วิธีการรวบรวมข้อมูล)

ทำการรวบรวมข้อมูลจากศูนย์กลางข้อมูลภาครัฐ ซึ่งเป็นข้อมูลแบบเปิดจากกรมควบคุมโรค กระทรวงสาธารณสุข โดยจะเป็นข้อมูลผู้ป่วยโควิด-19 ระลอกที่ 1 ถึงระลอกที่ 3 [9] ตั้งแต่ 1 มกราคม 2563 ถึง วันที่ 1 ตุลาคม 2564 เป็นจำนวน 1,608,923 ราย

#### 3.2 Data Preparation (การเตรียมข้อมูล)

##### 3.2.1 Formatting (การจัดรูปแบบข้อมูล)

ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นตอนถัดไปได้ ในการแปลงข้อมูลนี้อาจจะ

ต้องมีการทำข้อมูลให้ถูกต้อง เช่นการแปลงข้อมูลให้อยู่ในช่วงเดียวกันโดยจะทำการแปลงข้อมูลจากตัวอักษรเป็นตัวเลข

##### 3.2.2 Data Cleaning (การทำความสะอาดข้อมูล)

เนื่องจากคะแนนหรือข้อมูลบางส่วนไม่สามารถใช้ในการคำนวณได้และข้อมูลที่ใส่จะต้องเป็นข้อมูลที่ไม่มีการสูญหายหรือ Missing Value

#### 3.3 Modeling (การสร้างโมเดล)

ทำการสร้างโมเดลด้วยวิธี Naive Bayes, Random Forest และ Neural network โดยใช้วิธีแบ่งข้อมูลแบบ Split Test จำนวน 80% เป็นข้อมูลสำหรับสร้างโมเดล และข้อมูลจำนวน 20% ใช้เป็นข้อมูลสำหรับทดสอบ ในงานวิจัยนี้ได้สร้างโมเดลแยกตามประเภทผู้ป่วยโดย Class ของข้อมูลจะแบ่งเป็น 4 Class ตามตารางที่ 2

#### 3.4 Evaluation (การวัดประสิทธิภาพโมเดล)

หลังจากที่แบ่งข้อมูลออกเป็น 2 ส่วนและใช้โมเดลที่สร้างได้จากข้อมูลจำนวน 80% มาทำการคัดแยกผู้ป่วยโควิด-19 กับข้อมูลผู้ป่วยที่อยู่ใน 20% ที่เหลือ และใช้วัดค่าความถูกต้องโดยใช้ Confusion matrix เพื่อดูว่าเทคนิควิธี Naive Bayes, Random Forest และ Neural network เทคนิคสร้างโมเดลไหนให้ค่าผลทำนายดีมากที่สุดจึงจะนำโมเดลนั้นไปใช้งานต่อไป

#### 3.5 Deployment (การนำโมเดลไปใช้งาน)

จากโมเดลที่สร้างโดย Naive Bayes, Random Forest และ Neural network ที่สร้างขึ้นมาได้นั้นสามารถนำไปทำการคัดแยกผู้ป่วยโควิด-19 ได้โดยเลือกเทคนิคที่มีค่าความถูกต้องมากที่สุดเพียง 1 เทคนิคเท่านั้น

### 4. ผลการดำเนินงาน

การวิจัยนี้ใช้ข้อมูลผู้ป่วยโควิด-19 ที่ได้ทำการรวบรวมไปทดสอบกับเทคนิคตัวทำนาย 3 แบบ คือ Naive Bayes, Random Forest และ Neural network โดยใช้วิธีแบ่งข้อมูลแบบ Split Test จำนวน 80% เป็นข้อมูลสำหรับสร้างโมเดลและข้อมูลจำนวน 20% ใช้เป็นข้อมูลสำหรับทดสอบ และผลลัพธ์ของการทำวิจัยดังตารางที่ 3 ถึงตารางที่ 6

**ตารางที่ 3:** Information Gain ของฟีเจอร์ในการวิจัย

Feature	Information Gain
gender	0.003
age rang	0.011
nationality	0.019
province	0.214
risk	0.817

จากตารางที่ 3 พบว่าการคัดเลือกฟีเจอร์ทั้งหมด โดยใช้เทคนิคคำนวณค่าน้ำหนักของฟีเจอร์ต่าง ๆ โดยฟีเจอร์ที่มีค่าน้ำหนักมากที่สุดคือ risk ค่าน้ำหนักที่ 0.817 และ ฟีเจอร์ที่มีค่าน้ำหนักน้อยที่สุดคือ gender ค่าน้ำหนักที่ 0.003

**ตารางที่ 4:** ตัววัดประสิทธิภาพของอัลกอริทึม Naive Bayes

Performance Classification (Naive Bayes)					
	true(1)	true(2)	true(3)	true(4)	Precision
pred.(1)	22211	0	113277	2243	16.13%
pred.(2)	29	2	31	204	0.75%
pred.(3)	8791	178	474	4002	3.53%
pred.(4)	638	10	3088	27667	88.09%
Recall	70.13%	1.05%	0.41%	81.09%	

จากตารางที่ 4 พบว่า Class 1 True Positive เท่ากับ 22,211 ค่า False Positive เท่ากับ 115,520 ค่า คิดเป็น Precision เท่ากับ 16.13% และ Recall เท่ากับ 70.13% Class 2 True Positive เท่ากับ 2 ค่า False Positive เท่ากับ 264 ค่า คิดเป็น Precision เท่ากับ 0.75% และ Recall เท่ากับ 1.05% Class 3 True Positive เท่ากับ 474 ค่า False Positive เท่ากับ 12,971 ค่า คิดเป็น Precision เท่ากับ 3.53% และ Recall เท่ากับ 0.41% Class 4 True Positive เท่ากับ 27,667 ค่า False Positive เท่ากับ 3,739 ค่า คิดเป็น Precision เท่ากับ 88.10 % และ Recall เท่ากับ 88.10% สรุปเป็นค่า Accuracy ได้ 27.54%

**ตารางที่ 5:** ตัววัดประสิทธิภาพของอัลกอริทึม Random Forest

Performance Classification (Random Forest)					
	true(1)	true(2)	true(3)	true(4)	Precision
pred.(1)	135543	0	999	1189	98.41%
pred.(2)	2	246	0	18	92.48%
pred.(3)	415	1	9614	3415	71.51%
pred.(4)	4601	0	1221	25581	81.46%
Recall	96.43%	99.60%	81.24%	84.70%	

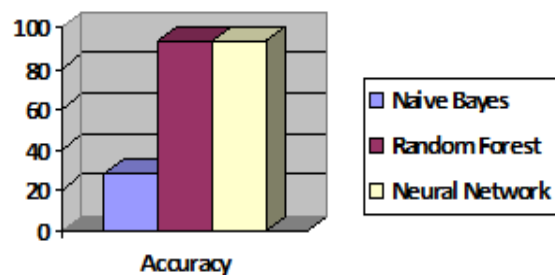
จากตารางที่ 5 พบว่า Class 1 True Positive เท่ากับ 135,543 ค่า False Positive เท่ากับ 2,188 ค่า คิดเป็น

Precision เท่ากับ 98.41% และ Recall เท่ากับ 96.43% Class 2 True Positive เท่ากับ 246 ค่า False Positive เท่ากับ 20 ค่า คิดเป็น Precision เท่ากับ 92.48% และ Recall เท่ากับ 99.60% Class 3 True Positive เท่ากับ 9614 ค่า False Positive เท่ากับ 3,831 ค่า คิดเป็น Precision เท่ากับ 71.51% และ Recall เท่ากับ 81.24% Class 4 True Positive เท่ากับ 25,581 ค่า False Positive เท่ากับ 30,203 ค่า คิดเป็น Precision เท่ากับ 81.46% และ Recall เท่ากับ 84.71% สรุปเป็นค่า Accuracy ได้ 93.51%

**ตารางที่ 6:** ตัววัดประสิทธิภาพของอัลกอริทึม Neural Network

Performance Classification (Neural Network)					
	true(1)	true(2)	true(3)	true(4)	Precision
pred.(1)	135546	0	1040	1145	98.41%
pred.(2)	1	247	0	18	92.48%
pred.(3)	375	0	9659	3411	71.51%
pred.(4)	4078	0	2698	24627	81.46%
Recall	96.43%	99.60%	81.24%	84.70%	

จากตารางที่ 6 พบว่า Class 1 True Positive เท่ากับ 135,546 ค่า False Positive เท่ากับ 2,185 ค่า คิดเป็น Precision เท่ากับ 98.41% และ Recall เท่ากับ 96.82% Class 2 True Positive เท่ากับ 247 ค่า False Positive เท่ากับ 19 ค่า คิดเป็น Precision เท่ากับ 92.86% และ Recall เท่ากับ 100% Class 3 True Positive เท่ากับ 9,659 ค่า False Positive เท่ากับ 13,397 ค่า คิดเป็น Precision เท่ากับ 71.84% และ Recall เท่ากับ 72.10% Class 4 True Positive เท่ากับ 24,627 ค่า False Positive เท่ากับ 29,201 ค่า คิดเป็น Precision เท่ากับ 78.42% และ Recall เท่ากับ 84.34% สรุปเป็นค่า Accuracy ได้ 93.02%



**ภาพที่ 2:** Performance Accuracy of Models

## 5. สรุปผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อทดลองสร้างโมเดลการทำนาย โดยใช้ข้อมูลผู้ป่วยโควิด-19 จากกรมควบคุมโรค ร่วมกับเทคนิคการเรียนรู้ของเครื่องได้แก่ Random Forest, Neural network และ Naive Bayes เพื่อใช้ในการคัดแยกผู้ป่วยโควิด-19 ในการเปรียบเทียบประสิทธิภาพความแม่นยำในการทำนาย โดยการทดลองมีเพียง 2 โมเดลที่มีค่าความถูกต้องในการทำนายผลมากกว่าร้อยละ 80 คือ Random Forest, Neural network โดยพิจารณาตามตัวชี้วัดคุณภาพ Accuracy, Precision และ Recall ซึ่งจากผลการทดลองผู้วิจัยจึงขอเสนอให้ใช้โมเดลจากเทคนิค Random Forest ในการประยุกต์ใช้เพื่อพัฒนาระบบคัดแยกผู้ป่วยโควิด-19 ต่อไป

### เอกสารอ้างอิง

- [1] สุรียา หนานมานะ, โสภณ เอี่ยมศิริถาวร, สุนนมาลย์ อุทยมกุล. “โรคติดเชื้อไวรัสโคโรนา 2019 (โควิด-19).” *วารสารสถาบันบำราศนราดูร*, ปีที่ 14, ฉบับที่ 2, หน้า 124–33, มีนาคม, 2563.
- [2] เสาวภา ทองงาม, สุพิศตรา ภูมุล, รณิษฐา รัตนะรัต. “การพยาบาลผู้ป่วยโรคโควิด-19 ในหอผู้ป่วยแยกโรค โรงพยาบาลศิริราช.” จาก <https://dx.doi.org/10.33192/Simedbull.2020.29>.
- [3] รัชดา โคนันต์. “โควิด-19 โดโรนาไวรัส สงครามที่มองไม่เห็น.” *วารสารอาหารและยา THAI FOOD AND DRUG JOURNAL*, ปีที่ 27, ฉบับที่ 2, พฤษภาคม, 2563.
- [4] ญัฐวรรณ คำแสน. “ความรู้ที่สั่นคลอนและพฤติกรรมในการป้องกันตนเองจากการติดเชื้อไวรัสโควิด-19 ของประชาชนในเขตอำเภออุ้มถ้อง จังหวัดสุพรรณบุรี.” *วารสารวิทยาลัยพยาบาลพระจอมเกล้า จังหวัดเพชรบุรี*, ปีที่ 4, เมษายน, 2564.
- [5] สุรัชย์ โชคกรรชิตไชย. “โควิด-19 : การระบาดระลอกใหม่ในประเทศไทยปลายปี 2563.” *วารสารสมาคมเวชศาสตร์ป้องกันแห่งประเทศไทย*, ปีที่ 10, ฉบับที่ 3, กันยายน-ธันวาคม, 2564.
- [6] “โรคติดเชื้อไวรัสโคโรนา 2019 (โควิด-19).” *World Health Organization Thailand*. Available Online at [https://cdn.who.int/media/docs/default-source/searo/thailand/2021\\_10\\_14\\_tha-sitrep-205-covid-19-final.pdf?sfvrsn=b19859cf\\_3](https://cdn.who.int/media/docs/default-source/searo/thailand/2021_10_14_tha-sitrep-205-covid-19-final.pdf?sfvrsn=b19859cf_3).
- [7] มหาวิทยาลัยเกษตรศาสตร์. “โควิด-19 และระบาดวิทยา.” จาก <https://learningcovid.ku.ac.th/course/?c=3&l=1> กรกฎาคม 2020.
- [8] A. Puitrakul. “ว่าด้วยเรื่อง Machine Learning.” *ว่าด้วยเรื่อง Machine Learning*, จาก <https://armondora.in.th/what-is-machine-learning/>.
- [9] กรมควบคุมโรค. “DDC OPENDATA Covid19 Thailand.” *DDC OPENDATA Covid19 Thailand*, Available Online at <https://covid19.ddc.moph.go.th/> [accessed on 22 October 2021].
- [10] S. Lalmuanawma, J. Hussain, and L. Chhakhuak. “Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review.” *Chaos, Solitons & Fractals*, Vol. 139, pp. 110059, Oct, 2020.
- [11] E. Alpaydin. *Introduction to Machine Learning*, New edition. MIT Press Ltd, Available Online at <https://www.bookdepository.com/Introduction-to-Machine-Learning-Ethem-Alpaydin-OEzyegin-University/9780262043793>.
- [12] Ian H. Witten, E. Frank, and M. A. Hakk. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, 2011.
- [13] สติชัยโชค โพธิ์สอาด, ปิติภูมิ โปสาวัง. “การจำแนกพฤติกรรม การจับขังโรคโดยสาธารณสุขโดยใช้วิธีการสกัดข้อความและเทคนิคการเรียนรู้ของเครื่อง.” *วารสารเทคโนโลยีสารสนเทศ*, ปีที่ 15, ฉบับที่ 1, มกราคม, 2562.
- [14] เอกสิทธิ์ พัชรวงศ์ศักดิ์. *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไม่นิ่งเบื้องต้น*. ครั้งที่ 2. กรุงเทพฯ : เอเชีย, ดิจิตอลการพิมพ์ จำกัด, 2557.
- [15] มานวิกา กิตติพร. “ระบบแนะนำทางการศึกษาและเทคนิคการเรียนรู้ของเครื่องจักร.” *วารสารสารสนเทศศาสตร์*, ปีที่ 37, ฉบับที่ 2, ธันวาคม, 2562.
- [16] ชนะวิชญ์ พัชรเจริญวงศ์, กันต์กวี เหิรเมฆ. *วางแผนงาน กัมปนาท*. “แบบจำลองการทำนายระยะเวลาในการเข้าเทียบท่าของเรือโดยสารสาธารณะ.” *วารสารวิทยาศาสตร์ลาดกระบัง*, ปีที่ 29, ฉบับที่ 2, กรกฎาคม-ธันวาคม, 2563.
- [17] อกันย์ฐ์ ทองจิตร, พูลพงศ์ สุขสว่าง, จตุภัทร เมฆพ่ายพ. “การพัฒนาวิธีจำแนกประเภทข้อมูลโดยใช้โครงข่ายประสาทเทียมแบบปรับเหมาะผสมผสานการหาค่าที่เหมาะสมที่สุดแบบกลุ่มอนุภาค สำหรับการจำแนกประเภทกลุ่มเสี่ยง.” *วิทยาการวิจัยและวิทยาการปัญญา*, ปีที่ 17, ฉบับที่ 2, กรกฎาคม-ธันวาคม, 2562.

# การวิเคราะห์ความรู้สึกของข้อความ กรณีศึกษาชุดข้อมูลของทวิตเตอร์ เรื่อง การศึกษาไทย

## Thai Sentiment Analysis about Education in Thailand on Twitter

วิภาดา ศิลาราช (Wipada Silarach)<sup>1</sup> พิทยารัตน์ พ่อมชภู (Phithayarat Phochompu)<sup>2</sup>

และธนพล ตั้งชูพงศ์ (Thanaphon Tangchoopong)<sup>3</sup>

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

<sup>1</sup>wipadasi@kkumail.com, <sup>2</sup>phithayarat.mwph@kkumail.com, <sup>3</sup>thanaphon@kku.ac.th

### บทคัดย่อ

เหมือนความคิดเห็นบนโลกออนไลน์ มีความจำเป็นมากในปัจจุบัน สามารถนำไปพัฒนาธุรกิจ หรือพิจารณาอารมณ์ของข้อความได้ งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างโมเดลในจำแนกข้อความภาษาไทยที่แสดงความคิดเห็นในแง่บวกของการศึกษาไทยบนทวิตเตอร์ออกมาเป็น 3 ข้อความคิดเห็นคือ ทางบวก ทางลบ และกลาง โดยนำชุดข้อมูลที่ได้จากการเลเบลไปคัดเลือกคำที่เป็นพีเจอร์ และพบปัญหาในเรื่องข้อมูลที่ไม่สมดุล จึงนำเสนอการจัดการข้อมูลที่ไม่สมดุลด้วยการปรับข้อมูลที่ไม่สมดุล 3 วิธีคือ การสุ่มตัวอย่างลด การสุ่มตัวอย่างเพิ่ม และการสังเคราะห์ข้อมูลเพิ่ม (SMOTE) แล้วฝึกโมเดลสำหรับจำแนกอารมณ์ของข้อความด้วยโมเดล Logistic regression, XGBoost, Decision tree, SVM, Random forest, K-NN โดยเลือกไฮเปอร์พารามิเตอร์ (Hyperparameter) ที่ดีที่สุดของแต่ละโมเดลจากการทำการค้นหาแบบกริด (Grid search) และเปรียบเทียบประสิทธิภาพของโมเดลก่อนและหลังการปรับข้อมูลที่ไม่สมดุลด้วยเมทริกซ์ความสับสน (Confusion matrix) พบว่าก่อนปรับข้อมูลค่าที่ได้ส่วนใหญ่ต่ำเกินไปที่คลาสลบ และค่าความถูกต้องในแต่ละโมเดลมีค่าต่ำเฉลี่ยที่ 41% และการทำนายในคลาสบวกมีผลที่น้อยจากค่าเฉลี่ย F1 score 17% หลังจากปรับข้อมูลทำให้ค่าเฉลี่ยความถูกต้องเพิ่มขึ้นทั้ง 3 วิธี โดยวิธีการสุ่มตัวอย่างลดมากที่สุดเฉลี่ยที่ 51% ซึ่งมีโมเดลที่ดีที่สุดคือ Random forest ที่ 55% อีกทั้งค่าเฉลี่ย F1 score

ในคลาสบวกสูงสุดที่ 58% และจาก 3 วิธี โมเดลที่ดีที่สุดคือ Logistic regression มีค่าเฉลี่ยความถูกต้องที่ 53%

**คำสำคัญ:** การวิเคราะห์ความรู้สึก เหมือนข้อความ การประมวลผลภาษาธรรมชาติ

### Abstract

Nowadays, text mining from social networks has become very important in order to thrive in business and sentiment monitoring. Therefore, this research aims to create a model that classifies the sentiments of Thai text comments on Twitter into three categories: negative, positive, and neutral. The collected data is labeled and selected as a vector feature. The major results are analyzed and indicate the negativity of the data. Thus, the research presents imbalanced data management and adjustment using three methods: random under sampling, random oversampling, and SMOTE. Afterward, the transformed training data is used to train models for analyzing the sentiment of data with the classification algorithms: Logistic regression, XGBoost, Decision tree, SVM, Random Forest, K-NN. The hyperparameters of each algorithm are well selected by the grid search method and evaluate the model's performance by a confusion matrix of result pre and post imbalanced data adaptation. Before the adaptation, the data have a tendency to negative classification. The model analysis has low accuracy, 41 percent on average, and the prediction

outcome of positive classification has 17 percent from the calculation of the F1 score. After the adaptation, the average accuracy has been increased in all three methods. The most effective method is random under sampling which has 51 percent accuracy, with the Random Forest model at 55 percent accuracy. Moreover, the average F1 score in positive classification has been raised to 58 percent. From all three methods combined with model utilization, the Logistic Regression is the best model.

**Keywords:** Sentiment Analysis, NLP, Text Mining.

## 1. บทนำ

เครือข่ายสังคมออนไลน์ในปัจจุบัน เป็นพื้นที่ที่ใช้ในการแสดงออกต่าง ๆ ของผู้คน ทั้งโพสต์กิจกรรมในชีวิตประจำวัน ติดต่อสื่อสาร ซึ่งทั้งหมดนี้สามารถโพสต์เป็นข้อความ เสียง หรือวิดีโอ แต่ส่วนใหญ่จะเกิดเป็นข้อความมากกว่า ทั้งการแสดงความเห็นได้โพสต์ การรับ-ส่งข้อความ ซึ่งมีทั้งคำพูดในทางที่ดีและไม่ดีปะปนกันไป แต่ข้อความที่แสดงความคิดเห็นบนสังคมออนไลน์มีอยู่จำนวนมาก การที่จะให้ผู้คนแต่ละคนมาทำความเข้าใจในข้อความแต่ละอันว่ามีความรู้สึกอย่างไรจึงเป็นไปได้ยาก การวิเคราะห์ความคิดเห็นบนสังคมออนไลน์โดยใช้ความสามารถของคอมพิวเตอร์สามารถช่วยลดปัญหาตามที่กล่าวมาได้ แต่ความท้าทายหนึ่งคือข้อความแสดงความคิดเห็นส่วนใหญ่ที่อยู่บนอินเทอร์เน็ตนิยมใช้ภาษาที่มีโครงสร้างไม่แน่นอน (Unstructured data) หรือเป็นภาษาธรรมชาติ (Natural language) ไม่ถูกต้องตามหลักไวยากรณ์ทางภาษาทำให้ยากต่อการวิเคราะห์ ในงานวิจัยนี้จึงได้นำเทคนิคการวิเคราะห์เหมืองข้อความ (Text mining) การประมวลภาษาธรรมชาติ (Natural Language Processing: NLP) เทคนิคการจัดการกับชุดข้อมูลไม่สมดุล (Imbalance data) และเทคนิคอื่น ๆ มาประยุกต์ใช้เพื่อวิเคราะห์ความคิดเห็นที่เป็นภาษาไทยของผู้คนบนเครือข่ายสังคมออนไลน์ ซึ่งเรียกว่าการวิเคราะห์ความรู้สึก (Sentiment analysis) ซึ่งงานวิจัยนี้จะอธิบายแนวคิด ทฤษฎี เทคนิค

ต่าง ๆ ที่เกี่ยวข้อง รวมถึงกระบวนการในการวิเคราะห์ การสร้างแบบจำลอง และแสดงตัวอย่างจากงานศึกษาวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกของข้อความ โดยจะทำการศึกษาผ่านเว็บไซต์บนสื่อสังคมออนไลน์ ผ่านทางทวิตเตอร์เอพีไอ (Twitter API) โดยกรณีศึกษาที่ผู้วิจัยเลือกมาคือเรื่องการศึกษาไทย ซึ่งหัวข้อนี้เป็นที่ถกเถียงกันในปัจจุบัน

## 2. ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง

**2.1 การประมวลผลภาษาธรรมชาติ [1]** เป็นการแปลภาษาที่มนุษย์ใช้สื่อสารกัน หรือภาษาธรรมชาติ ให้คอมพิวเตอร์สามารถเข้าใจได้ หรือทำให้เป็นโครงสร้าง

**2.2 การตัดคำ (Word segmentation) [2]** คือกระบวนการนำข้อความที่เป็นประโยค หรือข้อความที่มีความยาวมาตัดเป็นคำ ๆ เพื่อนำไปวิเคราะห์

**2.3 การปรับข้อมูลไม่สมดุล (Imbalance data)** การสุ่มตัวอย่างลด (Random Under Sampling: RUS) เป็นการสุ่มลดจำนวนข้อมูลตัวอย่างจากกลุ่มข้อมูลที่มีในกลุ่มมาก ให้มีขนาดน้อยเท่ากับกลุ่มข้อมูลที่น้อย การสุ่มตัวอย่างเกิน (Random Over Sampling: ROS) เป็นการสุ่มเกินจำนวนข้อมูลตัวอย่างจากกลุ่มข้อมูลที่มีขนาดน้อย ให้มีขนาดมากเท่ากับกลุ่มข้อมูลที่มีมากที่สุด การสังเคราะห์ข้อมูลเพิ่ม (Synthetic Minority Oversampling Technique: SMOTE) [3-4] เป็นการสังเคราะห์สุ่มเกินข้อมูลขึ้นมาใหม่จากข้อมูลที่มีอยู่

**2.4 การประเมินหรือวัดประสิทธิภาพของแบบจำลอง**

**การวิเคราะห์ข้อมูล (Validation)** งานวิจัยนี้ [5] จะใช้ค่าความถูกต้อง (Accuracy), ค่าความแม่นยำ (Precision), ค่าความครบถ้วน (Recall), ค่าความถ่วงดุล (F1-Score) โดย ค่าความถูกต้อง คือการวัดค่าความถูกต้องโดยรวมระหว่างค่าจริงและค่าการทำนายของทุก ๆ คลาส

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

ค่าความแม่นยำคือ อัตราส่วนที่โมเดลทำนายถูกหารด้วยจำนวนการทำนายที่ทำนายว่าเป็นคลาสที่กำลังพิจารณา ทั้งถูกและผิด

$$\text{Precision} = \frac{TP}{TP+FP}$$

ค่าความครบถ้วนคือ อัตราส่วนการวัดค่าการทำนายที่ทำนายได้ถูกต้องตรงกับค่าจริงจากจำนวนของค่าจริงทั้งหมดของคลาสที่กำลังพิจารณา

$$Recall = \frac{TP}{TP+FN}$$

ค่าความถ่วงดุลคือ ค่าเฉลี่ยของผลบวกของส่วนกลับแต่ละค่าของข้อมูลสถิติระหว่างค่าความแม่นยำและค่าความครบถ้วน (Harmonic mean)

$$F1 = 2 * \left( \frac{precision*recall}{precision+recall} \right)$$

### 2.5 งานวิจัยที่เกี่ยวข้อง

กานดา แก้วต้นนากุล และปราโมทย์ ลีอนาม [6] ได้ทำการวิเคราะห์ความเห็นจากเครือข่ายสังคมออนไลน์ ซึ่งได้ใช้ข้อความที่เป็นภาษาไทย และได้ตัดประโยคก่อนนำไปวิเคราะห์ โดยใช้โมเดล ต้นไม้ตัดสินใจ นาอ็ฟเบย์ และซัพพอร์ตเวกเตอร์แมชชีน มาสร้างโมเดลในการจำแนก

Younis, E. M. [7] วิเคราะห์ความเห็นจากข้อความบนสื่อสังคมออนไลน์ เกี่ยวกับผลิตภัณฑ์และงานบริการ ใช้อัลกอริทึมในการจัดหมวดหมู่ การทำคลัสเตอร์ และอื่น ๆ ซึ่งงานนี้ได้ลบค่าแสดงที่มักใช้ในออนไลน์ มีการตัดค่าแบบ Term - document matrix

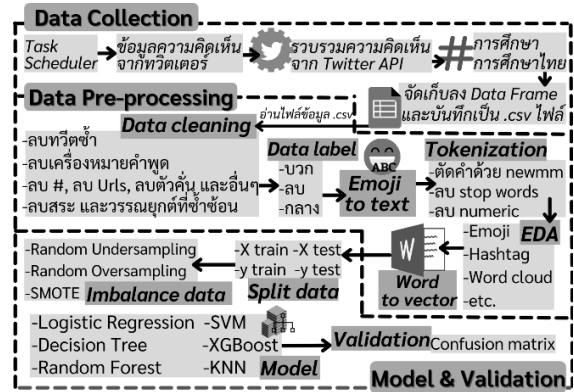
อนนต์ชัย ชูติภาสเจริญ [8] ได้เปรียบเทียบประสิทธิภาพของโมเดลที่ใช้ทำนายและคุณลักษณะ ที่มีต่อโอกาสความสำเร็จในการโอนเงินข้ามประเทศ โดยใช้ต้นไม้ตัดสินใจ นาอ็ฟเบย์ และเพื่อนบ้านใกล้สุด มาทำการเปรียบเทียบ และทดสอบประสิทธิภาพได้ค่าความถูกต้องเป็น 99.90% 96.71% และ 99.55% ตามลำดับ

Pong-Inwong, C. และ Songpan, W. [9] ได้รวบรวมความเห็นจากคำถามปลายเปิดจากนักศึกษาต่อการสอน เสนอวิธีการปรับปรุงด้วยการทำเหมืองข้อมูล การเชื่อมโยงเพื่อวิเคราะห์หัวข้อความรู้สึกร ซึ่งใช้เทคนิค SPPM ในการจำแนกทัศนคติการสอน มีความแม่นยำสูงสุด 87.94% เมื่อเทียบกับโมเดลลักษณะนามอื่น ๆ

พัชรียา ทองพูล และคณะ [10] ได้เปรียบเทียบวิธีการปรับข้อมูลไม่สมดุล 4 วิธีคือ การสุ่มเกิน SMOTE การสุ่มลด และการสุ่มผสมผสาน กับการจำแนก 4 วิธี เพื่อทดสอบว่าวิธีใดที่จะทำให้โมเดลจำแนกได้ดีที่สุด

### 3. วิธีการดำเนินงาน

ดังภาพที่ 1 งานวิจัยนี้ประกอบด้วย 3 ขั้นตอนใหญ่คือ การเก็บข้อมูล การเตรียมข้อมูล ท้ายสุดคือการสร้างโมเดล และการประเมินประสิทธิภาพของโมเดลรายละเอียด ดังนี้



ภาพที่ 1: ลำดับการทำงาน (Process)

**3.1 การเก็บข้อมูล (Data collection)** การเก็บข้อมูลแบบอัตโนมัติโดยใช้โปรแกรมตัวกำหนดตารางเวลางาน (Task Scheduler) โดยสร้างไฟล์ชุด (Batch Files) ด้านในมีไพธอนพาร์ทและไพธอนสคริปต์ไฟล์ จากนั้นตั้งค่าให้โปรแกรมรันสัปดาห์ละครั้ง ทุก ๆ วันจันทร์เวลา 01.00 น. ซึ่งภายในไฟล์มีการขอ Twitter API เพื่อเก็บข้อความบนทวิตเตอร์ โดยใช้แฮชแท็กในการค้นหาข้อมูล คือ #การศึกษาไทย และ#การศึกษา

**3.2 การเตรียมข้อมูล (Data pre-processing)** การทำความสะอาดข้อมูล (Data cleaning) เป็นการตรวจสอบ และแก้ไขข้อมูล เพื่อให้ข้อมูลอยู่ในรูปแบบที่ต้องการ การเลเบล (Label) ผู้วิจัยได้เลเบลข้อความด้วยตนเอง ซึ่งแบ่งออกเป็น 3 คลาสคือ คลาสบวก คลาสลบ และคลาสดกลาง โดยทำการเลเบลในระดับเอกสาร (Document level) การแปลงอิโมจิเป็นคำไทย เป็นการเปลี่ยนรูปแบบของอิโมจิให้เป็นคำภาษาไทย โดยใช้ไลบรารีของ Pythainlp ที่นำเข้า emoji\_to\_thai มาทำการแปลง เช่น 😊 เป็นคำว่า กลอกตา เป็นต้น การตัดคำเป็นการนำข้อความมาทำการตัดแบ่งเป็นคำ ๆ โดยใช้ newmm มาช่วยในการตัดคำ ซึ่งเป็นหนึ่งในไลบรารีของ Pythainlp การสำรวจข้อมูล (Exploratory Data



Analysis: EDA) เป็นการตรวจสอบข้อมูลที่ได้มาก่อน นำไปใช้สำรวจข้อมูลในมุมมองต่าง ๆ ในทุก ๆ ตัวแปร หรือ เปรียบเทียบกันระหว่างตัวแปร ดังตัวอย่างในภาพที่ 2 ที่แสดงให้เห็นถึงข้อความที่มักพบมากที่สุด ในคลาสบวก การแทนข้อความ คือก่อนการนำไปสร้างโมเดลต้องแปลงข้อมูลที่อยู่ในรูปแบบที่สามารถนำไปประมวลผลได้

**3.3 การแบ่งข้อมูล และการจัดการข้อมูลไม่สมดุล**

ผู้วิจัยได้แบ่งข้อมูลออกเป็นข้อมูลฝึกสอน และข้อมูลทดสอบจำนวน 4,096 และ 1,756 ข้อความ ตามลำดับ จาก 5,852 ข้อความ ซึ่งในแต่ละคลาสของข้อมูลฝึกสอน มีจำนวนไม่เท่ากัน จึงแก้ปัญหาของข้อมูลที่ไม่สมดุลนี้ จาก 3 วิธีคือ การสุ่มตัวอย่างลด การสุ่มตัวอย่างเกิน และ SMOTE

**3.4 การสร้างแบบจำลองข้อมูล เป็นขั้นตอนในการ**

สร้างโมเดลด้วยเทคนิคต่าง ๆ ได้แก่ Logistic Regression, XGBoost, Decision Tree, SVM, Random Forest, K-Nearest Neighbor ในการสร้างกฎ เพื่อทำนายแนวโน้มการเกิดขึ้นของข้อมูลที่ยังไม่มีเกิดขึ้น

**3.5 การประเมินประสิทธิภาพแบบจำลองข้อมูล**

ในงานวิจัยนี้จะใช้ Confusion matrix ในการวัดประสิทธิภาพของโมเดล ค่าที่นำมาวัดได้แก่ ค่าความถ่วงดุล, ค่าความครบถ้วน, ค่าความแม่นยำ และ ค่าความถูกต้อง



ภาพที่ 2: ความถี่ของคำที่ปรากฏในคลาสบวก

**Tweets :** บางทีครูก็สั่งงานเหมือนไม่อยากให้เด็กจบ 🙄🙄 #การศึกษาไทย #การบ้าน #เด็กมีคนเดียว #วิชาสามัญ #โรงเรียน

**Cleaning :** บางทีครูก็สั่งงานเหมือนไม่อยากให้เด็กจบ อดทน ไล่ออก

ภาพที่ 3: ตัวอย่างการทำความสะอาดข้อมูล

**4. ผลการศึกษា**

**4.1 การเก็บข้อมูล** เก็บข้อมูลแบบอัตโนมัติโดย

ค้นหาทวีตจาก #การศึกษาไทย และ #การศึกษา ทั้งหมด

5,852 ข้อความ มีคลาสลบ 3,346 ข้อความ คลาสบวก 257 ข้อความ และคลาสดกลาง 2,249 ข้อความ

**4.2 การเตรียมข้อมูล** ขั้นแรกทำความสะอาดข้อมูล

โดยลบข้อมูลที่ซ้ำกันออก ลบสัญลักษณ์ที่ไม่ได้ใช้ ลบ URLs ลบแฮชแท็ก ลบข้อความที่อยู่ใน <> ลบ @Users และอื่น ๆ หลังทำความสะอาดข้อมูลจึงจะทำการเลเวลข้อมูลในระดับประโยค แบ่งเป็น 3 คลาสคือ คลาสบวก คลาสลบ และคลาสดกลาง จากนั้นแปลงอิมojis ให้เป็นคำภาษาไทยโดยใช้ emoji\_to\_thai ตัวอย่างในภาพที่ 3 และตัดคำโดยใช้ newmm โดยลบตัวเลข คำฟุ่มเฟือย (Stop word) ทำการสำรวจข้อมูลเพื่อตรวจสอบข้อมูลก่อนนำไปใช้ และสุดท้ายเป็นการทำให้ข้อมูลอยู่ในรูปแบบที่คอมพิวเตอร์สามารถเข้าใจได้ โดยแปลงข้อความเป็นเวกเตอร์ ซึ่งใช้วิธีแบบ TF-IDF เพื่อให้โมเดลวิเคราะห์ได้

**4.3 การแบ่งข้อมูล การจัดการกับข้อมูลที่ไม่สมดุล**

การสร้างโมเดล และการวัดประสิทธิภาพ ทำการแบ่งชุดข้อมูลออกเป็นชุดฝึกสอน 4,096 ข้อความ และชุดทดสอบ 1,756 ข้อความ จากนั้นเปรียบเทียบค่าที่ได้ก่อน-หลังการปรับข้อมูลที่ไม่สมดุลของชุดข้อมูลฝึกสอน ด้วย 3 วิธีการคือ การสุ่มตัวอย่างลด การสุ่มตัวอย่างเพิ่ม และ SMOTE ซึ่งแบ่งข้อมูลได้ตามตารางที่ 1 จากตารางที่ 2-5 แสดงผลค่าความถ่วงดุล, ค่าความครบถ้วน, ค่าความแม่นยำ และ ค่าความถูกต้อง จากเมทริกซ์ความสับสน โดยกำหนดไฮเปอร์พารามิเตอร์ ให้แต่ละโมเดล ดังนี้ **Logistic Regression:** C=2, max\_iter=100, penalty='l2', solver='liblinear', **XGBoost:** max\_depth=5, gamma=2, subsample=0.6, colsample\_bytree=1.0, **Decision tree:** max\_depth=20, min\_samples\_leaf=20, criterion='entropy', **Random forest:** min\_samples\_leaf=3, min\_samples\_split=2, max\_depth=6, min\_sample\_leaf=3, max\_leaf\_node=5, n\_estimators=200, max\_samples=0.8, **KNN:** p=2, n\_jobs=-1, metric='minkowski', weights='distance', algorithm='auto', n\_neighbors=5, leaf\_size=30, **SVM:** C=100, gamma=0.001, kernel='rbf' ซึ่งกำหนดโดยใช้การค้นหาแบบกริด ในการค้นหาตัวไฮเปอร์พารามิเตอร์ ที่ดีที่สุดที่เหมาะสมกับโมเดลแต่ละตัว จากตาราง LGR คือ Logistic Regression, XGB คือ XGBoost, DT คือ Decision Tree, RFT คือ

Random Forest, KNN คือ K-Nearest Neighbor, SVM คือ Support Vector Machine และ Avg คือ ค่าเฉลี่ยของทุกโมเดล ซึ่ง (-1) แทนคลาสลบ (0) แทน คลาสกลาง และ (1) แทนคลาสบวก ซึ่งใช้ข้อมูลฝึกสอน ในการปรับข้อมูลที่ไม่สมดุล และใช้ข้อมูลทดสอบเป็น ชุดข้อมูลเดียวกันทั้งหมด จากตารางที่ 2 ค่าความถ่วงดุล ของคลาสบวกมีค่าที่ต่ำเฉลี่ย 17% เนื่องจากจำนวน ข้อมูลน้อยที่สุดจากทั้ง 3 คลาส ค่าความถูกต้องของแต่ละ โมเดลมีค่าที่ต่ำเช่นกันเฉลี่ยที่ 41% และค่าที่ได้ส่วน ใหญ่ทำนายได้ดีในคลาสลบ ทางผู้วิจัยจึงทำการปรับ ข้อมูลที่ไม่สมดุลของชุดข้อมูลฝึกสอน เพื่อเพิ่ม ประสิทธิภาพของค่าความถ่วงดุลในคลาสบวก

ตารางที่ 1: ข้อมูลในแต่ละคลาสก่อน-หลัง ข้อมูลที่ไม่สมดุล

	Pos	Neg	Neu
Train data (pre)	186	2,339	1,571
Test data	71	1007	678
Random under sampling	186	186	186
Random over sampling	2,339	2,339	2,339
SMOTE	2,339	2,339	2,339

ตารางที่ 2: แสดงค่าที่ได้ก่อนการปรับข้อมูลที่ไม่สมดุล

B4 Imbalance		LGR	XGB	DT	RFT	KNN	SVM	Avg
F1 Score	-1	0.59	0.56	0.48	0.62	0.29	0.53	0.51
	0	0.45	0.41	0.38	0.36	0.46	0.43	0.42
	1	0.05	0.14	0.19	0.1	0.33	0.23	0.17
Recall	-1	0.7	0.62	0.55	0.77	0.24	0.68	0.59
	0	0.58	0.54	0.48	0.44	0.63	0.48	0.53
	1	0.03	0.08	0.11	0.06	0.27	0.14	0.12
Precision	-1	0.51	0.52	0.43	0.51	0.38	0.43	0.46
	0	0.36	0.33	0.31	0.31	0.37	0.39	0.35
	1	1.00	0.5	0.57	0.57	0.42	0.67	0.62
Accuracy		0.44	0.41	0.38	0.42	0.38	0.43	0.41

ตารางที่ 3: การปรับข้อมูลที่ไม่สมดุลด้วยวิธี RUS

RUS		LGR	XGB	DT	RFT	KNN	SVM	Avg
F1 Score	-1	0.48	0.55	0.5	0.57	0.49	0.47	0.51
	0	0.42	0.44	0.42	0.43	0.41	0.39	0.42
	1	0.58	0.62	0.52	0.63	0.49	0.65	0.58
Recall	-1	0.46	0.54	0.55	0.56	0.49	0.41	0.5
	0	0.42	0.44	0.38	0.37	0.42	0.39	0.4
	1	0.59	0.63	0.52	0.72	0.46	0.73	0.61
Precision	-1	0.49	0.56	0.45	0.57	0.49	0.55	0.52
	0	0.42	0.44	0.47	0.51	0.39	0.39	0.44
	1	0.57	0.6	0.53	0.55	0.51	0.58	0.56
Accuracy		0.49	0.54	0.48	0.55	0.46	0.51	0.51

ตารางที่ 4: การปรับข้อมูลที่ไม่สมดุลด้วยวิธี ROS

ROS		LGR	XGB	DT	RFT	KNN	SVM	Avg
F1 Score	-1	0.63	0.58	0.44	0.62	0.51	0.56	0.56
	0	0.49	0.41	0.42	0.3	0.43	0.52	0.43
	1	0.53	0.35	0.56	0.61	0.27	0.54	0.48
Recall	-1	0.63	0.62	0.42	0.65	0.54	0.52	0.56
	0	0.59	0.52	0.41	0.24	0.56	0.68	0.5
	1	0.42	0.23	0.59	0.7	0.17	0.42	0.42
Precision	-1	0.62	0.54	0.46	0.59	0.49	0.62	0.55
	0	0.42	0.34	0.43	0.4	0.34	0.43	0.39
	1	0.71	0.76	0.53	0.54	0.63	0.73	0.65
Accuracy		0.55	0.46	0.47	0.53	0.42	0.54	0.5

จากตารางที่ 3-5 พบว่าโดยผลจากวิธี RUS วิธี ROS และ วิธี SMOTE ค่าความถ่วงดุลคลาสบวกโดยเฉลี่ยเพิ่มขึ้น ที่ 58% 48% และ 36% ตามลำดับ และค่าความถูกต้อง เพิ่มขึ้นเฉลี่ยในแต่ละวิธีที่ 51% 50% และ 47% ตามลำดับ เมื่อเปรียบเทียบการปรับข้อมูลที่ไม่สมดุล จากทั้ง 3 แบบพบว่าค่าความถูกต้องส่วนใหญ่มีค่า เพิ่มขึ้นหลังจากปรับแก้ข้อมูลแล้วโดยโมเดลที่ทำนาย แม่นยำน้อยส่วนใหญ่คือ K-NN เฉลี่ยจากทั้ง 3 วิธีที่ 43.33% ส่วนโมเดลที่ทำนายได้แม่นยำส่วนใหญ่จากทั้ง 3 วิธี เฉลี่ย 52.67% คือ Logistic Regression โดยวิธี RUS ให้ค่าความถูกต้องเฉลี่ยของการทำนายที่ดีที่สุด 51% โดยมีโมเดล Random forest ที่ให้ค่าความถูกต้อง ที่ 55%

ตารางที่ 5: การปรับข้อมูลที่ไม่สมดุลด้วยวิธี SMOTE

SMOTE		LGR	XGB	DT	RFT	KNN	SVM	Avg
F1 Score	-1	0.63	0.62	0.48	0.55	0.53	0.59	0.57
	0	0.48	0.46	0.43	0.46	0.42	0.44	0.45
	1	0.49	0.24	0.31	0.41	0.26	0.46	0.36
Recall	-1	0.65	0.7	0.49	0.56	0.58	0.61	0.6
	0	0.56	0.59	0.54	0.58	0.52	0.52	0.55
	1	0.39	0.14	0.23	0.3	0.17	0.37	0.27
Precision	-1	0.61	0.56	0.46	0.54	0.48	0.58	0.54
	0	0.43	0.38	0.36	0.38	0.35	0.38	0.38
	1	0.64	0.77	0.5	0.66	0.52	0.62	0.62
Accuracy		0.54	0.48	0.42	0.48	0.42	0.5	0.47

## 5. สรุป

ข้อมูลจากทวิตเตอร์ในหัวข้อการศึกษาไทย เป็น ข้อมูลที่รวบรวมมาทั้งสิ้น 5,852 ข้อความ แบ่งเป็นคลาส บวก คลาสลบ และคลาสกลาง เมื่อทดสอบและวัด ประสิทธิภาพแล้วพบว่าค่าความถูกต้องในแต่ละโมเดลมี ค่าโดยเฉลี่ยที่ 41% ผลการทำนายมีแนวโน้มไปทางคลา

สลับ ซึ่งคลาสบวกมีการทำนายที่แม่นยำได้น้อยที่สุดจากทั้ง 3 คลาส เนื่องจากข้อมูลมีเกิดอคติ (Bias) ดังนั้นผู้วิจัยจึงปรับข้อมูลให้สมดุลด้วยวิธี RUS วิธี ROS และ SMOTE เพื่อลดปัญหาดังกล่าว และการค้นหาแบบกริดเพื่อหาไฮเปอร์พารามิเตอร์ที่ดีที่สุดของแต่ละโมเดล เมื่อทดสอบและวัดประสิทธิภาพพบว่า สามารถเพิ่มประสิทธิภาพให้โมเดลได้จากทั้ง 3 วิธี ซึ่งสอดคล้องกับงานวิจัยของพัชรียา ทองพูล และคณะ [10] ส่งผลให้ค่าเฉลี่ยความถูกต้องเพิ่มขึ้น เมื่อพิจารณาที่ค่าเฉลี่ยวิธี RUS ให้ผลดีที่สุด ก่อนและหลังการปรับข้อมูลที่ไม่สมดุลโมเดล Logistic regression ให้ค่าความถูกต้องมากที่สุด เฉพาะวิธี RUS เท่านั้นที่โมเดลทำนายดีที่สุดคือ Random forest และยังคงผลให้ค่าความถ่วงดุลในคลาสบวกเพิ่มขึ้น พบว่าวิธี RUS ให้ประสิทธิภาพที่ดีที่สุดในการทำนายความถ่วงดุลในคลาสบวก

เมื่อดูจากความถี่คำที่ปรากฏในคลาสบวกพบคำที่ปรากฏถี่ เช่น “ชอบ”, “ดี” และ “น่ารัก” เป็นต้น ทั้งนี้ในชุดข้อมูลยังพบอุปสรรคในเรื่องของภาษา บนแพลตฟอร์มนี้ นิยมใช้ภาษาที่กระชับ พบคำย่อ คำที่เขียนผิดอาจส่งผลให้การทำให้เจอร์เวเตอร์ให้ค่าที่ต่างกัน ซึ่งทางผู้วิจัยจะนำประเด็นนี้ไปพัฒนาต่อไปเพื่อประสิทธิภาพที่ดีขึ้น

### เอกสารอ้างอิง

- [1] ยืน กุ้ววรรณ. *การประมวลผลภาษาธรรมชาติ*. กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี. 2535.
- [2] สุรศักดิ์ ตั้งสกุล. และธูปานิ เฮงสนันกุล. “การตัดคำภาษาไทยสำหรับข้อความในพีพธิภณฑป์ลาน้ำจืด.” *วารสารวิทยาศาสตร์บูรพา*, ปีที่ 16, ฉบับที่ 1, หน้า 84-93, 2554.
- [3] T. Fawcett. (2016). *Learning from Imbalanced Classes*. Available Online at [https://www.svds.com/learning-imbalanced-classes/?utm\\_source=kdnuggets&utm\\_medium=](https://www.svds.com/learning-imbalanced-classes/?utm_source=kdnuggets&utm_medium=), accessed on 1 November 2021.
- [4] Garbled. (2013). *Class Imbalance Problem*. Available Online at <http://www.chioka.in/class-imbalance-problem/>, accessed on 1 November 2021.
- [5] Chengz. *วัดประสิทธิภาพ Model จาก Confusion Matrix*. [สืบค้นเมื่อวันที่ 2 ธันวาคม 2563]. จาก <https://medium.com/@cheng-confusion-matrix>.
- [6] กานดา แผ้วพัฒนากุล, ปราโมทย์ ลีอนาม. “การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์.” *วารสารการจัดการสมัยใหม่*, ปีที่ 11, ฉบับที่ 2, หน้า 12-20, 2556.
- [7] E. Younis. “Sentiment analysis and text mining for social media microblogs using Opensource tools: an empirical study.” *International Journal of Computer Applications*, Vol. 112, No. 5, February, 2015.
- [8] A. Chutipascharoen, and C. Sanrach. *A Comparison of the Efficiency of Algorithms and Feature Selection Methods for Predicting the Success of Personal Overseas Money Transfer*. Available Online at <https://so04.tci-thaijo.org/index.php/gskkuhs/article/view/156370>, accessed on 1 August 2021.
- [9] C. Pong-Inwong, & W. Songpan. “Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (SPPM) based on association mining.” *International Journal of Machine Learning and Cybernetics*, Vol. 10, No. 8, pp. 2177-2186, March, 2019.
- [10] พชรียา ทองพูล, พิมพชนก จำเริญ, รรมย์นลิน บุญฤทธิ, และสายชลสิน สมบูรณ์ทอง. “การเปรียบเทียบประสิทธิภาพ ในการทำนายผลการปรับความไม่สมดุลของข้อมูล ในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล.” *Thai Journal of Science and Technology*, ฉบับที่ 8, ปีที่ 6, หน้า 565-584, 2019.

# การพยากรณ์รีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออกโดยใช้การเรียนรู้เครื่องจักร

## Forecasting Home Loan Refinance Out Using Machine Learning

พงศ์พันธุ์ แควดวง (Pongpan Kaeduang)<sup>1</sup> และ กาญจนา วิริยะพันธ์ (Kanchana Viriyapant)<sup>2</sup>

ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307011858225@email.kmutnb.ac.th, <sup>2</sup>kanchana.v@itd.kmutnb.ac.th

### บทคัดย่อ

ปัจจุบันการตรวจสอบความต้องการรีไฟแนนซ์ของลูกค้าสินเชื่อเพื่อที่อยู่อาศัยของลูกค้าเป็นเรื่องที่สถาบันทางการเงินมีความยากลำบาก ในการตรวจสอบเนื่องจากธนาคารต้นทางจะรับรู้ได้ก็ต่อเมื่อลูกค้าได้ติดต่อและทำสัญญากับธนาคารปลายทางเรียบร้อยแล้ว ทำให้เกิดความสูญเสียด้านการเงินและสูญเสียส่วนแบ่งการตลาดให้แก่คู่แข่งทางการตลาด

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาการพยากรณ์การรีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออกโดยสร้างแบบจำลอง ด้วยข้อมูลลูกค้าจากธนาคารแห่งหนึ่งในประเทศไทยย้อนหลัง ตั้งแต่ปี พ.ศ.2560 - 2564 จำนวน 70,954 รายการ นำมาใช้ในการจัดทำแบบจำลองโดยใช้ป่าสุ่ม (Random Forest) ในการสร้างแบบจำลอง พบว่าปัจจัย 13 ปัจจัย ได้แก่ เงินงวดที่เพิ่มจากงวดแรก อายุของผู้กู้ วัตถุประสงค์การกู้ เงินงวดปัจจุบัน วงเงินกู้ อาชีพของผู้กู้ อัตราดอกเบี้ย สวัสดิการของผู้กู้ จำนวนบุตร การเข้าร่วมมาตรการช่วยเหลือลูกค้าที่ได้รับผลกระทบด้านรายได้จากสถานการณ์การแพร่ระบาดของ (COVID-19) เพศของผู้กู้ ระยะเวลากู้ และสถานภาพสมรส ส่งผลต่อลูกค้าที่ทำการรีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออก และวัดประสิทธิภาพแบบจำลองโดยหาค่าความถูกต้องของข้อมูล การวิจัยสามารถพยากรณ์ลูกค้าที่มีแนวโน้มที่จะรีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออกเพื่อรักษาผลประโยชน์ด้านการเงินและรักษารฐานลูกค้าของธนาคารโดยมีค่าความถูกต้องของแบบจำลองอยู่ที่ 88.20 %

**คำสำคัญ:** พยากรณ์การรีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออก รีไฟแนนซ์ การเรียนรู้เครื่องจักร ป่าสุ่ม

### Abstract

Currently, it is difficult for financial institutions to verify the refinancing needs of their home loan customers. Because the origin bank will be recognized only after the customer has successfully contacted and entered into a contract with the receiving bank causing loss of market share to competitors.

This research aims to develop forecasting model for home loan refinance out by 70,954 records of bank case study in Thailand from 2017 to 2021 using random forest. The result found that 13 factors as increase of installment, age, loan purpose, current installment, loan amount, occupation, interest rate, welfare, number of children, financial assistance COVID-19 epidemic, gender, loan term, and marriage status Affects customers who refinance mortgage loans. The system performance testing from case study data show that the accuracy is 88.20 percent. It can be concluded that the system is effective in prediction in home loan refinance.

**Keywords:** Forecasting Home Loan Refinance Out, Refinance, Machine Learning, Random Forest

### 1. บทนำ

การรีไฟแนนซ์บ้าน คือ การขอขึ้นกู้สินเชื่อบ้านกับธนาคารแห่งใหม่ โดยการชำระเงินกู้ที่มีอยู่ในปัจจุบันทั้งหมดด้วยเงินกู้ใหม่ และใช้สินทรัพย์เดิมเป็นหลักประกัน ผู้ขอสินเชื่อสามารถเลือกธนาคารได้เองตามแต่ธนาคารใดจะให้ข้อเสนอที่น่าพอใจมากกว่ากัน

เพื่อลดภาระเงินกู้เก่าที่มีอยู่ ผู้ทำการรีไฟแนนซ์ มีจุดประสงค์เพื่อลดดอกเบี้ยบ้านให้ลดลง จากอัตราดอกเบี้ยปัจจุบัน และช่วยให้จำนวนเงินงวดการผ่อนชำระต่อเดือนลดน้อยลง

ปัจจุบันในประเทศไทยมียอดสินเชื่อที่อยู่อาศัยทั้งหมด 2,575,667 ล้านบาท แบ่งเป็นสินเชื่อที่อยู่อาศัยแนวราบ ประกอบด้วย บ้านเดี่ยว บ้านแฝด ทาวน์เฮาส์ และอาคารพาณิชย์ จำนวน 1,691,786 ล้านบาท สินเชื่อที่อยู่อาศัยประเภทอาคารชุด จำนวน 607,445 ล้านบาท สินเชื่อที่อยู่อาศัยประเภทอาคารพาณิชย์ จำนวน 185,365 ล้านบาท และสินเชื่อที่อยู่อาศัยประเภทอื่นๆ จำนวน 91,070 ล้านบาท และมีมูลค่าสินเชื่อปล่อยใหม่ในการรีไฟแนนซ์ จำนวน 16,931 ล้านบาท [1] โดยตลาดสินเชื่อที่อยู่อาศัย ในปัจจุบันมีส่วนแบ่งตลาดตามสถาบันการเงิน แบ่งเป็น ธนาคารอาคารสงเคราะห์ 31.65 % ธนาคารไทยพาณิชย์ 13.36 % ธนาคารกรุงไทย 10.32 % ธนาคารกสิกรไทย 9.25 % ธนาคารออมสิน 8.14% ธนาคารทีเอ็มบีธนชาต 7.10 % ธนาคารกรุงศรีอยุธยา 6.81 % ธนาคารกรุงเทพ 6.23 % ธนาคารอื่นๆ 4.41% ซึ่งการรีไฟแนนซ์ที่เป็นการยื่นกู้สินเชื่อบ้านกับธนาคารแห่งใหม่นั้น จะทำให้เกิดผลกระทบกับธนาคารที่เป็นเจ้าของสินเชื่อเดิม ทั้งการสูญเสียผลประโยชน์ด้านการเงิน และการสูญเสียส่วนแบ่งการตลาดให้แก่คู่แข่ง

จากสถานการณ์ในปัจจุบัน การแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) ทำให้เกิดผลกระทบต่อเศรษฐกิจของประเทศไทย ซึ่งส่งผลกระทบต่อหลายด้าน ไปจนถึงอัตราการจ้างงานภายในประเทศให้ลดน้อยลง ส่งผลต่อความเชื่อมั่นกับนักลงทุนในประเทศและต่างประเทศ เป็นผลให้ผู้กู้สินเชื่อสูญเสียรายได้ ส่งผลให้จำนวนคนตกงานที่เพิ่มสูงขึ้น ทำให้ผู้คนมีการวางแผนด้านการเงินมากขึ้น ประกอบกับกลยุทธ์ทางการตลาดของธนาคารต่างๆ ที่ออกมาเพื่อทำการตลาดแข่งขันกับคู่แข่ง ซึ่งปัจจัยเหล่านี้มีผลต่อความต้องการในการรีไฟแนนซ์สินเชื่อที่อยู่อาศัยเพิ่มมากขึ้น

การมีระบบการพยากรณ์รีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออกโดยใช้การเรียนรู้เครื่องจักร เป็นการช่วยในการ

ตัดสินใจ การทำการตลาดของธนาคารเพื่อจะได้วางแผนการดำเนินงานกับกลุ่มลูกค้า ที่มีความต้องการรีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยดังกล่าวให้ตรงตามความต้องการของลูกค้ามากขึ้น

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การเรียนรู้เครื่องจักร (Machine Learning)

การเรียนรู้เครื่องจักร คือ การเรียนรู้ด้วยตนเอง โดยการใช้อัลกอริทึม ประมวลผลข้อมูลของโปรแกรมเมอร์ หรือระบบที่สามารถเรียนรู้ได้จากตัวอย่างด้วยตนเอง ผู้เกี่ยวข้องไม่จำเป็นต้องป้อนคำสั่งสำหรับการทำงาน คอมพิวเตอร์สามารถเรียนรู้จากข้อมูล เพื่อที่จะแสดงผลลัพธ์ที่แม่นยำออกมาจากการฝึกสอน โดยสามารถแบ่งประเภทการเรียนรู้ออกเป็น 2 ประเภท คือ การเรียนรู้แบบมีผู้สอน (Supervised Learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) [2]

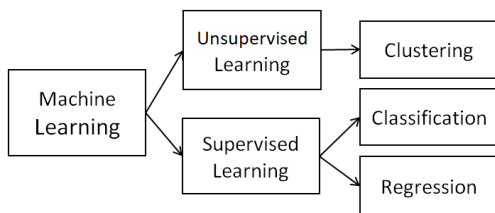
#### 2.1.1 การเรียนรู้แบบมีผู้สอน

การเรียนรู้แบบมีผู้สอนเป็นเทคนิคการเรียนรู้เครื่องจักร โดยการสร้างฟังก์ชันจากข้อมูลที่นำมาสอน ซึ่งข้อมูลที่นำมาทำการสอนจะต้องมีการนำเข้าข้อมูล และมีผลลัพธ์จากการทำนาย ที่สามารถแบ่งได้เป็น 2 ประเภท คือ การแบ่งแยกประเภทหรือการแยกหมวดหมู่ (Classification) และการวิเคราะห์การถดถอย (Regression) โดยการแบ่งแยกประเภทหรือการแยกหมวดหมู่ของข้อมูล เป็นการแบ่งแยกประเภทของข้อมูลให้อยู่ในรูปแบบของหมวดหมู่ที่แตกต่างกัน เป็นการจำแนกแบบ ไบนารี (Binary Classification) เพื่อใช้สำหรับแปลงผลลัพธ์ให้อยู่ในรูปแบบของหมวดหมู่แค่สองหมวดหมู่ ผลลัพธ์จะแสดง ใช่ หรือ ไม่ใช่ เท่านั้น โดยมีเทคนิคที่ใช้กับการจำแนกแบบ ไบนารี คือ ต้นไม้ตัดสินใจ (Decision Tree) [7] การค้นหาเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) [4] และการแบ่งกลุ่ม โดยใช้หลักการความน่าจะเป็น (Naive Bayes) [4] และการแบ่งประเภทอิกูรูปแบบ คือ การจำแนกประเภทหลายประเภท (Multi-Class Classification) แตกต่างจากการจำแนกแบบ ไบนารี การจำแนกประเภทหลายประเภทจะสามารถแบ่งประเภท โดยมีหมวดหมู่หรือ

ผลลัพธ์จากการแบ่งประเภทมากกว่า 2 หมวดหมู่ โดยมีเทคนิคที่ใช้กับการจำแนกประเภทหลายประเภท ตัวอย่างเช่น โครงข่ายประสาทเทียม (Neural Network) [6] ส่วนการวิเคราะห์การถดถอย เป็นวิธีการทางสถิติที่ใช้สำหรับศึกษาหาความสัมพันธ์ระหว่างตัวแปรตั้งแต่สองตัวแปรขึ้นไป ประกอบด้วย ตัวแปรต้น และตัวแปรที่ต้องการทราบค่า

### 2.1.2 การเรียนรู้แบบไม่มีผู้สอน

การเรียนรู้แบบไม่มีผู้สอนเป็นเทคนิคหนึ่งของการเรียนรู้เครื่องจักร โดยการสร้างแบบจำลองที่เหมาะสมกับข้อมูล การเรียนรู้แบบไม่มีผู้สอนจะไม่มีภาระผลที่ต้องการหรือประเภทไว้ก่อน การเรียนรู้แบบนี้จะพิจารณาวัตถุเป็นกลุ่มของตัวแปรสุ่ม แล้วจึงสร้างแบบจำลองเพื่อหาผลลัพธ์ที่ต้องการ จากชุดข้อมูล และมีผลลัพธ์จากการทำนาย เป็นการจัดกลุ่มโดยมีเทคนิคที่ใช้กับการจัดกลุ่ม เช่น การแบ่งข้อมูลแบบเคมีน (K-Means Clustering) และแบบจำลองการผสม (Gaussian Mixture Model) [2] เป็นต้น



ภาพที่ 1: เทคนิคการทำ Machine Learning [2]

### 2.2 ต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจเป็นแบบจำลองทางคณิตศาสตร์รูปแบบหนึ่ง ที่สร้างขึ้นเพื่อทำการค้นหาทางเลือกที่ดีที่สุด โดยสร้างแบบจำลองในรูปแบบ โครงสร้างต้นไม้ ใช้คุณลักษณะของข้อมูลในการจำแนกประเภทข้อมูล ทำให้ทราบถึงคุณลักษณะที่เป็นตัวกำหนด การจำแนกประเภทและความสำคัญที่แตกต่างกันของแต่ละคุณลักษณะ ต้นไม้ตัดสินใจประกอบด้วย โหนด (Node) ที่แสดงถึงคุณลักษณะต่างๆ ของข้อมูล และเป็นจุดที่ใช้กำหนดเงื่อนไข โดยมีโหนดที่เป็นจุดเริ่มต้นของต้นไม้ตัดสินใจเรียกว่า โหนดราก (Root Node) และมีโหนดลำดับสุดท้ายคือ โหนดใบ (Leaf Node) และกิ่ง (Branch) [7]

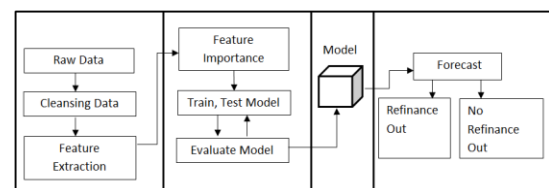
### 2.3 ป่าสุ่ม

ป่าสุ่มเป็นอัลกอริทึมที่ถูกพัฒนาต่อขุดมาจากต้นไม้ตัดสินใจ โดยการนำแบบจำลองจากต้นไม้ตัดสินใจหลายๆ แบบจำลอง มาเรียนรู้หลายครั้งบนข้อมูลชุดเดียวกัน โดยแต่ละครั้งของการเรียนรู้แบบจำลองจะเลือกส่วนของข้อมูลที่นำมาเรียนรู้ไม่เหมือนกัน และนำการตัดสินใจของแบบจำลองทั้งหมดที่ได้จากการเรียนรู้มาโหวตกัน เพื่อหาผลลัพธ์ที่ดีที่สุด โดยป่าสุ่มสามารถใช้กับปัญหาการแยกหมวดหมู่ และการวิเคราะห์การถดถอยด้วยข้อมูลที่มีโครงสร้างและข้อมูลที่ไม่มีโครงสร้าง [12]

## 3. วิธีดำเนินการวิจัย

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องพบว่าการเรียนรู้ และการหาคุณลักษณะสำคัญด้วยป่าสุ่มให้ผลลัพธ์ที่ถูกต้องและเหมาะสมกับข้อมูล ผู้วิจัยจึงได้ออกแบบและพัฒนาระบบการพยากรณ์รีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออกโดยใช้การเรียนรู้เครื่องจักร ดังนี้

### 3.1 กรอบแนวคิด



ภาพที่ 2: กรอบแนวคิดการวิจัย

### 3.2 การรวบรวมข้อมูลดิบ (Raw Data)

รวบรวมข้อมูลของลูกค้าลูกค้าที่ทำการรีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออกในอดีต และลูกค้าที่ยังคงมีสินเชื่อกับธนาคารปัจจุบันย้อนหลัง ตั้งแต่ปี พ.ศ. 2560 - 2564 จากธนาคารแห่งหนึ่งในประเทศไทย จำนวน 70,954 ราย

### 3.3 การทำความสะอาดข้อมูล (Cleansing Data)

ทำความสะอาดข้อมูลที่ไม่สามารถนำมาใช้งานได้ เนื่องจากข้อมูลอยู่ในรูปแบบของข้อมูลดิบ ที่ยังมีความผิดพลาดจากการนำเข้าข้อมูลโดยมนุษย์ เช่น อายุของผู้กู้ที่เป็นค่าว่าง ข้อมูลที่ไม่ระบุวัตถุประสงค์การกู้ และการจัดกลุ่มของอาชีพผู้กู้ จาก รัฐบาลการครู และรัฐบาลการตำรวจ เป็นรัฐบาลการ เป็นต้น

### 3.4 การสกัดคุณลักษณะ (Feature Extraction)

สกัดคุณลักษณะเด่นจากข้อมูลทำการจัดกลุ่มและแปลงข้อมูล ให้อยู่ในรูปแบบที่สามารถนำไปใช้งานกับการเรียนรู้เครื่องจักรได้ โดยการแปลงข้อมูลให้อยู่ในรูปของชุดตัวเลข ดังตารางต่อไปนี้

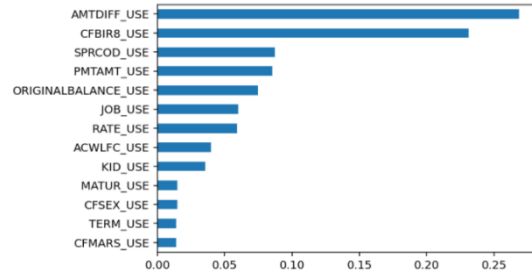
ตารางที่ 1: แสดงค่าคุณลักษณะของข้อมูล

เงินงวดที่เพิ่มขึ้นจากงวดแรก (พันบาท)	0 - 1	> 1 - 2	> 2 - 3	> 3 - 4	> 4 - 5	> 5
อายุของผู้กู้ (ปี)	18 - 30	31 - 40	41 - 50	51 - 60	> 60	
วัตถุประสงค์การกู้	ซื้อ	ปลูกสร้าง	ต่อเติม	ไถ่ถอน		
เงินงวดปัจจุบัน (หมื่นบาท)	0 - 1	> 1 - 2	> 2 - 3	> 3 - 4	> 4 - 5	> 5
วงเงินกู้ (ล้านบาท)	0 - 1	> 1 - 2	> 2 - 3	> 3 - 4	> 4 - 5	> 5
อาชีพของผู้กู้	เจ้าหน้าที่รัฐ	พนักงานบริษัทเอกชน	ธุรกิจส่วนตัว	อาชีพอื่นๆ		
อัตราดอกเบี้ย(%)	0 - 2	2 - 4	> 4			
สวัสดิการของผู้กู้	มี	ไม่มี				
จำนวนบุตร (คน)	0	1	2	> 2		
การเข้าร่วมมาตรการฯ	เข้าร่วม	ไม่เข้าร่วม				
ค่าคุณลักษณะ	0	1	2	3	4	5

### 3.5 การหาคุณลักษณะที่สำคัญของข้อมูล

นำข้อมูลที่ได้จากการสกัดคุณลักษณะ ด้านข้อมูลส่วนบุคคล ประกอบด้วย อาชีพของผู้กู้ อายุของผู้กู้ เพศ สถานภาพสมรส จำนวนบุตร และข้อมูลด้านการเงิน ประกอบด้วย วงเงินกู้ อัตราดอกเบี้ย ระยะเวลากู้ เงินงวดปัจจุบัน สวัสดิการของผู้กู้ การเข้าร่วมมาตรการช่วยเหลือลูกค้าที่ได้รับผลกระทบด้านรายได้จากสถานการณ์การแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) เงินงวดที่เพิ่มจากงวดแรก และวัตถุประสงค์การกู้ โดยมีตัวแปรตาม คือ ลูกค้าที่ทำการรีไฟแนนซ์สินเชื่อเพื่อที่อยู่

อาศัยออก และลูกค้าที่ไม่ทำการรีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออก นำไปทำการหาคุณลักษณะที่สำคัญของข้อมูลโดยใช้ป่าสุ่ม พบว่าปัจจัยที่ส่งผลต่อการตัดสินใจทำการรีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออก แสดงผลดังภาพที่ 3



ภาพที่ 3: คุณลักษณะที่สำคัญของข้อมูล

ตารางที่ 2: คุณลักษณะที่สำคัญของข้อมูล

ลำดับ	ปัจจัย	ค่าความสำคัญ
1	เงินงวดที่เพิ่มจากงวดแรก	0.2687
2	อายุของผู้กู้	0.2312
3	วัตถุประสงค์การกู้	0.0874
4	เงินงวดปัจจุบัน	0.0853
5	วงเงินกู้	0.0746
6	อาชีพของผู้กู้	0.0600
7	อัตราดอกเบี้ย	0.0591
8	สวัสดิการของผู้กู้	0.0396
9	จำนวนบุตร	0.0357
10	การเข้าร่วมมาตรการฯ	0.0149
11	เพศของผู้กู้	0.0148
12	ระยะเวลากู้	0.0142
13	สถานภาพสมรส	0.0139

### 3.6 การสร้างแบบจำลอง

นำข้อมูลที่ผ่านกระบวนการทำความสะอาดข้อมูล และผ่านการสกัดคุณลักษณะเด่นเรียบร้อยแล้ว นำมาสร้างแบบจำลอง โดยการแบ่งข้อมูลออกเป็น 2 ส่วน สำหรับฝึกสอนแบบจำลองและสำหรับทดสอบแบบจำลอง โดยมีรายละเอียดจากข้อมูลทั้งหมด 70,954 ราย แบ่งข้อมูลสำหรับฝึกสอนแบบจำลอง 80 % จำนวน 56,764 รายการ ข้อมูลสำหรับทดสอบแบบจำลอง 20 % จำนวน 14,190 รายการ นำมาสร้างแบบจำลองโดยใช้ต้นไม้ตัดสินใจ

และป่าสุ่ม ด้วยจำนวนต้นไม้ 1 ต้น 100 ต้น 200 ต้น และ 300 ต้น ตามลำดับ

4. ผลการทดสอบ

4.1 ผลการวัดประสิทธิภาพของแบบจำลอง

วัดประสิทธิภาพของแบบจำลอง โดยใช้เมทริกซ์ความสับสน (Confusion Matrix) โดยค่า 0 คือ ลูกค้าที่ทำ การรีไฟแนนซ์ที่อยู่อาศัยออก และค่า 1 คือ ลูกค้าที่ไม่ทำ การรีไฟแนนซ์ที่อยู่อาศัยออก ดังภาพที่ 5 และ 6

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

ภาพที่ 4: สูตรเมทริกซ์ความสับสน

	ต้นไม้ตัดสินใจ (จำนวน 1 ต้น)		ป่าสุ่ม (จำนวนต้นไม้ 100 ต้น)			
	Actual		Actual			
Prediction		0	1		0	1
	0	9364	635	0	9344	655
	1	1120	3071	1	1025	3166

ภาพที่ 5: เมทริกซ์ความสับสนของแบบจำลองที่ 1 และ 2

	ป่าสุ่ม (จำนวนต้นไม้ 200 ต้น)		ป่าสุ่ม (จำนวนต้นไม้ 300 ต้น)			
	Actual		Actual			
Prediction		0	1		0	1
	0	9354	645	0	9351	648
	1	1043	3148	1	1025	3166

ภาพที่ 6: เมทริกซ์ความสับสนของแบบจำลองที่ 3 และ 4

4.2 ผลวัดความถูกต้องของแบบจำลอง

วัดความถูกต้องของแบบจำลอง โดยการวัดค่าความแม่นยำ (Accuracy) ค่าความแม่นยำ (Recall) ค่าความแม่นยำ (Precision) และค่า F1-score

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

TP คือ ลูกค้าที่ทำ การรีไฟแนนซ์ที่อยู่อาศัยออก

และพยากรณ์ว่ามี การรีไฟแนนซ์ออก

FP คือ ลูกค้าที่ไม่ได้ทำ การรีไฟแนนซ์ที่อยู่อาศัยออก

แต่พยากรณ์ว่ามี การรีไฟแนนซ์ออก

FN คือ ลูกค้าที่ทำ การรีไฟแนนซ์ที่อยู่อาศัยออก

แต่พยากรณ์ว่าไม่มีการรีไฟแนนซ์ออก

TN คือ ลูกค้าที่ไม่ได้ทำ การรีไฟแนนซ์ที่อยู่อาศัยออก

และพยากรณ์ว่าไม่มีการรีไฟแนนซ์ออก

ตารางที่ 3: ตารางแสดงความถูกต้องของแบบจำลอง

	ต้นไม้ตัดสินใจ (จำนวน 1 ต้น)			ป่าสุ่ม (จำนวนต้นไม้ 100 ต้น)		
	0	1	Accuracy	0	1	Accuracy
precision	89.32%	82.87%	87.63%	90.11%	82.86%	88.16%
recall	93.65%	73.28%	87.63%	93.45%	75.54%	88.16%
f1-score	91.43%	77.78%	87.63%	91.75%	79.03%	88.16%
	ป่าสุ่ม (จำนวนต้นไม้ 200 ต้น)			ป่าสุ่ม (จำนวนต้นไม้ 300 ต้น)		
	0	1	Accuracy	0	1	Accuracy
precision	89.97%	83.00%	88.10%	90.12%	83.01%	88.20%
recall	93.55%	75.11%	88.10%	93.52%	75.54%	88.20%
f1-score	91.72%	78.86%	88.10%	91.79%	79.10%	88.20%

จากตารางที่ 3 พบว่าแบบจำลองที่ใช้อัลกอริทึมป่าสุ่ม โดยมีจำนวนต้นไม้ 300 ต้น มีค่าความถูกต้องอยู่ที่ 88.20 % ซึ่งมีความถูกต้องมากกว่าอัลกอริทึมอื่น ๆ

5. สรุปผล

งานวิจัยเพื่อพัฒนาการพยากรณ์การรีไฟแนนซ์สินเชื่อ เพื่อที่อยู่อาศัยออก โดยการสร้างแบบจำลองด้วยข้อมูล ลูกค้าจากธนาคารแห่งหนึ่งในประเทศไทยย้อนหลัง ตั้งแต่ปี พ.ศ. 2560-2564 จำนวน 70,954 รายการ พบว่า แบบจำลองที่มีค่าความถูกต้องที่สุด เป็นแบบจำลองที่ สร้างโดยใช้อัลกอริทึมป่าสุ่ม มีจำนวนต้นไม้ 300 ต้น มีค่า ความแม่นยำ (Precision) อยู่ที่ 90.12% ค่าความแม่นยำ (Recall) อยู่ที่ 93.52 % มีค่า F1- score อยู่ที่ 91.79 % และพบว่า แบบจำลองมีค่าความถูกต้องในการพยากรณ์อยู่ที่ 88.20 % โดยมีปัจจัยที่มีผลต่อการพยากรณ์ประกอบด้วย จำนวน เงินงวดที่เพิ่มจากงวดแรก อายุของผู้กู้ วัตถุประสงค์การ กู้เงินงวดปัจจุบัน วงเงินกู้ อาชีพของผู้กู้ อัตราดอกเบี้ย สวัสดิการของผู้กู้ จำนวนบุตร การเข้าร่วมมาตรการ ช่วยเหลือลูกค้าที่ได้รับผลกระทบด้านรายได้จาก สถานการณ์การแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) เพศ ระยะเวลากู้ และสถานภาพสมรส ส่งผล



ต่อการตัดสินใจรีไฟแนนซ์สินเชื่อเพื่อที่อยู่อาศัยออกตามลำดับ

ทั้งนี้สำหรับงานวิจัยในอนาคตสามารถนำผลการวิเคราะห์ และแนวทางที่ได้นี้ไปออกแบบแนวทางการพัฒนาการพยากรณ์ด้านสินเชื่อการเงิน เพื่อสร้างเครื่องมือที่ช่วยในการรักษาฐานลูกค้า และความได้เปรียบทางด้านธุรกิจการเงินต่อไปได้

### เอกสารอ้างอิง

- [1] ธนาคารแห่งประเทศไทย [สืบค้นวันที่ 20 กันยายน 2564] <https://www.bot.or.th/>
- [2] Kevin Bouchard, Lucas Gonzales, Julien Maitre, and Sébastien Gaboury, “Features Exploration for Grades Prediction using Machine Learning” *Association for Computing Machinery*. 2020.
- [3] F. Campomanes, A. V. Pada, and Sébastien Gaboury, “Mangrove Classification Using Support Vector Machines and Random Forest Algorithm: A Comparative Study” *GEOBIA*. 2016.
- [4] Yulia Ery Kurniawati, Adhistya Erna Permasari, and Silmi Fauziati, “Comparative Study on Data Mining Classification Methods for Cervical Cancer Prediction Using Pap Smear Results” *1st International Conference on Biomedical Engineering*. 2016.
- [5] วัชรวิวัฒน์ จิตตสกุล และสุนันทา สดสี, “Applying Random Forest Algorithm and Graph Theory for Text Analyzing”. *วารสารวิทยาศาสตร์และเทคโนโลยี*. 2019.
- [6] Meenu Sreedharan, Ahmed M. Khedr, and Magdi El Bannany, “A Comparative Analysis of Machine Learning Classifiers and Ensemble Techniques in Financial Distress Prediction” *17th International Multi-Conference on Systems, Signals & Devices*. 2020.
- [7] รุจิรา ธรรมสมบัติ, “Decision Support System for Selection the Mobile Internet Package Using Decision Tree” *ComBus*. 2011.
- [8] Yu Li, “Credit Risk Prediction Based on Machine Learning Methods” *The 14th International Conference on Computer Science & Education*. 2019.
- [9] Muhammad Imran Faisal, Saba Bashir, Zain Sikandar Khan, and Farhan Hassan Khan, “An Evaluation of Machine Learning Classifiers and Ensembles for Early Stage Prediction of Lung Cancer”. *3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST)*. 2018.
- [10] Sharmin Ara, Annesha Das, and Ashim Dey, “Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms”. *International Conference on Artificial Intelligence*. 2021.
- [11] Neli Kalcheva, Milena Karova, and Ivaylo Penev, “Comparison of the accuracy and the execution time of classification algorithms for Bulgarian literary works”. *International Conference Automatics and Informatics (ICAI)*. 2020.
- [12] Debadri Dutta, Kolkata, Debpriyo Paul, and Parthajeet Ghosh, “Analyzing Feature Importance’s for Diabetes Prediction using Machine Learning”. *IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. 2018.
- [13] Thee Juvitayapun, “Employee Turnover Prediction: The impact of employee event features on interpretable machine learning methods”. *13th International Conference on Knowledge and Smart Technology (KST)*. 2021.
- [14] Haru Kaneko, Tahera Hossain, and Sozo Inoue Ghosh, “Analysis of Feature Importances for Automatic Generation of Care Records”. *UbiComp-ISWC*. 2021.
- [15] Huanian Zhang, and Fanliang Bu, “Weighted KNN Algorithm Based on Random Forests”. *11th International Conference on Machine Learning and Computing*. 2019.

# การบูรณาการชุดข้อมูลทางไซเบอร์สำหรับตัวแบบการทำนายการบุกรุกทางไซเบอร์

## Integrating Cyber Dataset for Cyber Intrusion Prediction Model

สมบูรณ์ อุดนัน (Somboon Udnan)<sup>1</sup> ประสงค์ ปราณีตพลกรัง (Prasong Praneetaphonkrang)<sup>2</sup>

และพายัพ ศิรินาม (Payap Sirinam)<sup>3</sup>

<sup>1,2,3</sup>หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีป้องกันประเทศ

สำนักบัณฑิตศึกษา โรงเรียนนายเรืออากาศนวมินทกษัตริยาธิราช

<sup>1</sup>vit.peksayfa@gmail.com, <sup>2</sup>prasongspu@gmail.com, <sup>3</sup>p.sirinam@gmail.com

### บทคัดย่อ

การทำนายภัยคุกคามทางไซเบอร์ในปัจจุบันจะอาศัยข้อมูลผลการตรวจจับการบุกรุกทางไซเบอร์เพียงอย่างเดียว ซึ่งไม่ครอบคลุมมิติของภัยคุกคามทางไซเบอร์ที่เกิดขึ้นจริงกับองค์กร ถึงแม้ว่าข้อมูลเหล่านี้จะบอกถึงการโจมตีทางไซเบอร์ที่เกิดขึ้นในหน่วยงานได้ค่อนข้างดี แต่ยังคงไม่ครอบคลุมในบริบทความมั่นคงปลอดภัยทางไซเบอร์ขององค์กร แต่เนื่องจากข้อมูลดังกล่าว ยังขาดข้อมูลที่สำคัญที่เกี่ยวข้องกับสาเหตุที่ทำให้ระบบเกิดช่องโหว่จนถูกโจมตี รวมถึงระดับความเสี่ยงของการโจมตี

งานวิจัยนี้ เสนอวิธีการบูรณาการข้อมูลผลการตรวจจับการบุกรุกทางไซเบอร์ ข้อมูลการทดสอบเจาะระบบ และข้อมูลด้านข่าวกรองไซเบอร์ ให้เป็นชุดข้อมูล (Dataset) เพื่อใช้สำหรับการประมวลผล โมเดลพื้นฐานด้านการทำนายการบุกรุกทางไซเบอร์ด้วยปัญญาประดิษฐ์ โดยได้ผลลัพธ์เป็นชุดข้อมูลใหม่ (New Dataset) ที่มีข้อมูลครอบคลุมภัยคุกคามทางไซเบอร์มากยิ่งขึ้น พร้อมกับระบบแสดงข้อมูลผลการทำนายการบุกรุก ข้อมูลความเสี่ยงและความพร้อมทางไซเบอร์ขององค์กร อันจะเป็นต้นแบบของวิธีการใช้ชุดข้อมูลด้านการบุกรุกทางไซเบอร์ที่มีความครอบคลุมกับภัยคุกคามทางไซเบอร์ในปัจจุบัน

**คำสำคัญ:** ชุดข้อมูลทางไซเบอร์ ตัวแบบการทำนายการบุกรุกทางไซเบอร์

### Abstract

Currently, cyber threat predictions rely solely on cyber intrusion detection results. This does not cover the dimensions of actual cyber threats to organizations.

Although this information is a pretty good indication of a cyber attack taking place in an organization, But still not covered in the corporate cybersecurity context.

This research proposes a method for integrating cyber intrusion detection results data penetration testing data and cyber intelligence into datasets for processing AI based cyber intrusion prediction base models. Research results in a new dataset with comprehensive threat intelligence is used for the intrusion prediction, cyber risk and cyber readiness in organizations. Therefore, integrating cyber data would be a model of how to use a comprehensive cyber intrusion dataset with today's cyber threats.

**Keywords:** Cyber Dataset, Cyber Intrusion Prediction Model

### 1. บทนำ

ข้อมูลผลการตรวจจับการบุกรุกทางไซเบอร์ในปัจจุบันนั้น มักจะมาจากระบบตรวจจับการบุกรุก (Intrusion Detection System : IDS) และระบบป้องกันการบุกรุก (Intrusion Prevention System : IPS) หรือไฟร์วอลล์ (Firewall) เป็นหลัก โดยอาจเรียกว่าเป็นข้อมูลบันทึกเข้า-ออก (Log) ที่เป็นภัยคุกคามทางไซเบอร์ โดย IDS จะแจ้งเตือนเมื่อมีการบุกรุกเครือข่าย

อย่างไรก็ตาม IDS/IPS มีความสามารถเพียงแค่ตรวจจับและแจ้งเตือนการเข้าใช้งานเครือข่ายที่ผิดปกติ ดังนั้นการใช้ IDS/IPS ก็จะขึ้นอยู่กับว่าสภาพแวดล้อม

ในระบบที่ใช้งานอยู่นั้นมีอะไรที่จะแจ้งเตือนให้ทราบ ทั้งนี้ ข้อมูลที่ได้รับจากการตรวจจับภัยคุกคามทางไซเบอร์ของ IPS/IDS ยังไม่เพียงพอในการวิเคราะห์ข้อมูลภัยคุกคามทางไซเบอร์ให้ครอบคลุม

เราอาจทราบรูปแบบและปริมาณการโจมตีเข้ามาในระบบจาก Log แต่ไม่อาจทราบถึงสาเหตุการโจมตี หรือกรณีที่หน่วยงานอื่นถูกโจมตีทางไซเบอร์ ซึ่งการโจมตีดังกล่าวก็อาจจะเกิดขึ้นกับหน่วยงานเราได้ด้วยเช่นกัน แต่ก็ไม่อาจทราบรายละเอียดได้เพราะขาดการแลกเปลี่ยนข้อมูลภัยคุกคามระหว่างกัน ดังนั้นการบูรณาการชุดข้อมูลทางไซเบอร์จึงมีความจำเป็นอย่างยิ่งที่จะทำให้องค์กรมีความพร้อมสำหรับการรับมือกับภัยคุกคามทางไซเบอร์ได้อย่างทันท่วงที

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการทำวิจัยครั้งนี้ ผู้วิจัยได้ใช้ทฤษฎีที่เกี่ยวข้องดังต่อไปนี้

**2.1 ระบบตรวจจับการบุกรุกทางไซเบอร์ (IDS) เป็นระบบสำหรับแจ้งเตือนเมื่อพบการบุกรุกทางไซเบอร์ แต่ก็ไม่สามารถตรวจจับการบุกรุกกับสถานะแวดล้อมที่มี สวิตซ์ซิง (Switching) และไม่สามารถป้องกันการบุกรุกได้โดยทันทีหรือแบบเรียลไทม์ได้ ดังนั้นจึงได้มีระบบป้องกันการบุกรุก (IPS) ที่สามารถตรวจจับการบุกรุกและหยุดการบุกรุกได้อย่างทันที IPS รุ่นใหม่ ๆ จะมีการนำเทคโนโลยีในการวิเคราะห์ข้อมูลขั้นสูง เช่น Neural Networks และ Fuzzy Logic มาใช้ ซึ่งจะทำให้ปัญหาการแจ้งเตือนที่ผิดพลาดลดลงไปได้ [1, 2] หรือการใช้ Honeypot เพื่อดึงดูดและกำหนดลักษณะการรับส่งข้อมูลที่เป็นอันตราย [3]**

### 2.2 ระบบวิเคราะห์ภัยคุกคามทางไซเบอร์ด้วย Security Onion

Security Onion คือ SIEM ตัวหนึ่งที่มีความสามารถในการวิเคราะห์ข้อมูลภัยคุกคามทางไซเบอร์แบบ NIDS ได้ โดย Security Onion เริ่มต้นในปี 2008 โดยมีพื้นฐานมาจาก Ubuntu Linux มีการพัฒนาเป็นเวอร์ชันต่าง ๆ เรื่อยมา[4] กระบวนการวิเคราะห์ภัยคุกคามของ Security Onion2

จะเริ่มจากการนำข้อมูล Log ที่ได้รับเข้ามาผ่านกระบวนการตรวจสอบกับ Signature Base ที่มี หากตรงตามเงื่อนไขก็จะทำการแจ้งเตือนผ่านโปรแกรม Alerts โดยมีโปรแกรม Kibana เป็นตัวแสดงผลกราฟการโจมตีต่างๆ ในการคัดแยกรูปแบบภัยคุกคามทางไซเบอร์ที่ตรวจพบจะสามารถแบ่งออกได้ตามเอกสาร ET Category Descriptions ของบริษัท Proofpoint ซึ่งเป็นบริษัทรักษาความมั่นคงปลอดภัยระดับองค์กรของสหรัฐอเมริกา โดยได้ขยายความหมายของคำต่าง ๆ ที่ใช้ในการตั้งชื่อ Message Rule Alert ของโปรแกรม Security Onion เอาไว้ทั้งหมด 53 รูปแบบ [5] ดังตารางที่ 1

ตารางที่ 1: รายชื่อ Message Ruler ของโปรแกรม Security Onion

id	ET Category	Description
1	3COREsec	3COREsec team's IP block lists
2	ActiveX	Microsoft ActiveX controls and exploits
3	Adware-PUP	tracking or other types of spyware
4	Attack Response	compromise may have happened
5	Botcc	botnet and other Command and Control (C2)
6	Botcc Portgrouped	Botcc category but grouped by destination port
7	Chat	identify traffic related to numerous chat clients
8	CIArmy	IP rules for blocking
9	Coinmining	rules that detect malware which performs coin mining
10	Compromised	list of known compromised hosts
11	Current Events	Snort can experience performance
12	Deleted	deactivated state within their respective rule files
13	DNS	attacks and vulnerabilities regarding DNS
14	DO5	detect Denial of Service (DoS) attempts
15	Drop	block IP addresses on the Spamhaus DROP
16	Dshield	updated daily from the DShield top attackers list
17	Exploit	vulnerabilities such as against Microsoft Windows
18	Exploit-Kit	detect activity related to Exploit Kits
19	FTP	vulnerabilities regarding File Transfer Protocol (FTP)
20	Games	identify of gaming traffic
21	Hunting	threat hunting in an environment.
22	ICMP	signatures related to attacks ping
23	ICMP_info	ICMP protocol specific events
24	IMAP	vulnerabilities regarding Internet Message Access Protocol
25	Inappropriate	potentially activity related to sites that are pornographic
26	Info	downloading an Executable over HTTP by IP
27	JA3	parameters that are in the SSL handshake
28	Malware	detect Malware
29	Misc	category is for signatures not covered
30	Mobile Malware	detect Malware mobile
31	NETBIOS	includes rules that detect activity
32	P2P	attacks against it. Identified P2P traffic
33	Phishing	credentials into credential phishing sites
34	Policy	indicate violations to an organization's policy
35	POP3	vulnerabilities regarding mail
36	RPC	vulnerabilities regarding Remote Procedure Call (RPC)
37	SCADA	vulnerabilities regarding supervisory control
38	SCADA_special	Snort Digital Bond based SCADA
39	SCAN	detect Scanning
40	Shellcode	detect shellcode
41	SMTP	vulnerabilities regarding mail
42	SNMP	vulnerabilities regarding SNMP
43	SQL	vulnerabilities SQL
44	TELNET	vulnerabilities regarding TELNET
45	TFTP	vulnerabilities regarding Trivial File Transport Protocol
46	TOR	traffic to and from TOR exit nodes based on IP
47	Trojan	detect Trojan
48	User Agents	detect suspicious and anomalous user agents
49	VOIP	vulnerabilities regarding Voice over IP
50	Web Client	vulnerabilities regarding web browsers CURL, WGET
51	Web Server	detect attacks IIS, Tomcat, apache
52	Web Specific Apps	vulnerabilities in specific web applications
53	WORM	detect worm

### 2.3 งานวิจัยที่เกี่ยวข้อง

Robert C. Newman (2009) ได้นำเสนอเกี่ยวกับการจำแนกประเภทการตรวจจับการบุกรุกเครือข่าย (NIDS) และระบบตรวจจับการบุกรุกบนโฮสต์ (HIDS) โดยวิธีที่นิยมกันมากที่สุดคือการตรวจจับแบบ Signature Based เป็นการจดจำรูปแบบที่ไม่ดี และการตรวจจับตามความคิดปกติ เป็นการตรวจจับการเบี่ยงเบนจากการรับส่งข้อมูล ซึ่งมีกฏการเรียนรู้ของเครื่อง โดยผลิตภัณฑ์ IDS บางตัวสามารถตอบสนองต่อการบุกรุกที่ตรวจพบได้ ระบบที่มีความสามารถในการตอบสนองมักจะเรียกว่าระบบป้องกันการบุกรุก [6]

Proofpoint Company ได้นำเสนอเอกสารการแบ่ง Message Rule Alert ของโปรแกรม Security Onion เอาไว้ทั้งหมด 53 รูปแบบ ช่วยให้สามารถแยกรูปแบบภัยคุกคามทางไซเบอร์ได้สะดวกยิ่งขึ้น [5]

กอบเกียรติ สระอุบล (2020) ได้นำเสนอเกี่ยวกับกระบวนการสร้าง Dataset ที่มีทั้งหมด 6 ขั้นตอนคือ การกำหนดประเด็นหรือตั้งคำถาม (Business/Problem Definition) การเก็บรวบรวมข้อมูล (Data Acquisition) การเตรียมข้อมูล (Data Preparation) การสำรวจและสรุปหาสิ่งที่อยู่ในข้อมูล (Exploratory Data Analysis) การสร้างแบบจำลอง (Model Building) และการนำไปใช้งาน (Communication & Deployment) [7]

### 3. วิธีการดำเนินงาน

ระบบบูรณาการชุดข้อมูลทางไซเบอร์ ผู้วิจัยได้แบ่งขั้นตอนการดำเนินการออกเป็น 5 ขั้นตอน ดังต่อไปนี้

3.1 การรวบรวมข้อมูล แบ่งการทำงานออกเป็น 3 ขั้นตอนย่อย ๆ ดังนี้

3.1.1 ข้อมูลผลการตรวจจับภัยคุกคามทางไซเบอร์ หลังจากที่ทำการถอดรหัสด้วยอัลกอริทึม RSA แล้วก็จะนำข้อมูลมาทำการคำนวณจำนวน Even Count ของการโจมตี และให้ค่า Score ของการโจมตีนั้น ๆ ตามด้วยการเปลี่ยนคอลัมน์ Event\_Date เป็น Index ของ Data Table เพื่อให้สามารถประมวลผลแบบ Time Series ได้ และใช้เป็น Index ของตารางข้อมูลบูรณาการฯ โดยจะต้อง

แปลงผลการตรวจจับของโปรแกรม Security Onion ที่เดิมที่มี Rule ของการโจมตีทั้งหมด 53 รูปแบบ อ้างอิงตามมาตรฐานของ MITER ATT&CK [8] และ OWASP [9] ก็จะได้รูปแบบภัยคุกคามทั้งหมด 20 รูปแบบดังตารางที่ 2 ตารางที่ 2: ผลการเปรียบเทียบ Rule ของ Security Onion กับรูปแบบภัยคุกคามทางไซเบอร์ของ MITER ATT&CK และ OWASP

id	ET Category SO	Type Cyber (MITER & OWASP)	Description
1	3CORESec	warning	3CORESec team's IP block lists
2	ActiveX	Injection	Microsoft ActiveX controls and exploits
3	Adware-PUP	botnet	tracking or other types of spyware
4	Attack Response	compromise	compromise may have happened
5	Botcc	botnet	botnet and other Command and Control (C2)
6	Botcc Portgrouped	botnet	Botcc category but grouped by destination port
7	Chat	warning	identify traffic related to numerous chat clients
8	CIArmy	warning	IP rules for blocking
9	Coinmining	Bitcoin	rules that detect malware which performs coin mining
10	Compromised	compromise	list of known compromised hosts
11	Current Events	warning	Snort can experience performance
12	Deleted	warning	deactivated state within their respective rule files
13	DNS	Injection	attacks and vulnerabilities regarding DNS
14	DDoS	DDoS	detect Denial of Service (DoS) attempts
15	Drop	dropper	block IP addresses on the Spamhaus DROP
16	Dshield	Injection	updated daily from the DShield top attackers list
17	Exploit	compromise	vulnerabilities such as against Microsoft Windows
18	Exploit-Kit	compromise	detect activity related to Exploit Kits
19	FTP	data_leak	vulnerabilities regarding File Transfer Protocol (FTP)
20	Games	warning	identify of gaming traffic
21	Hunting	warning	threat hunting in an environment.
22	ICMP	scan	signatures related to attacks ping
23	ICMP_info	scan	ICMP protocol specific events
24	IMAP	phishing_web	vulnerabilities regarding Internet Message Access Protocol
25	Inappropriate	phishing_web	potentially activity related to sites that are pornographic
26	Info	malware	downloading an Executable over HTTP by IP
27	JAS	malware	parameters that are in the SSL handshake
28	Malware	malware	detect Malware
29	Misc	warning	category is for signatures not covered
30	Mobile Malware	malware	detect Malware mobile
31	NETBIOS	Injection	includes rules that detect activity
32	P2P	phishing_web	attacks against it. Identified P2P traffic
33	Phishing	phishing_web	credentials into credential phishing sites
34	Policy	warning	indicate violations to an organization's policy
35	POP3	phishing_email	vulnerabilities regarding mail
36	RPC	U2L	vulnerabilities regarding Remote Procedure Call (RPC)
37	SCADA	data_leak	vulnerabilities regarding supervisory control
38	SCADA_special	data_leak	Snort Digital Bond based SCADA
39	SCAN	scan	detect Scanning
40	Shellcode	trojan	detect shellcode
41	SMTP	phishing_email	vulnerabilities regarding mail
42	SNMP	warning	vulnerabilities regarding SNMP
43	SQL	Injection	vulnerabilities SQL
44	TELNET	U2L	vulnerabilities regarding TELNET
45	TFTP	data_leak	vulnerabilities regarding Trivial File Transport Protocol
46	TOR	warning	traffic to and from TOR exit nodes based on IP
47	Trojan	trojan	detect Trojan
48	User Agents	R2U	detect suspicious and anomalous user agents
49	VOIP	compromise	vulnerabilities regarding Voice over IP
50	Web Client	data_leak	vulnerabilities regarding web browsers CURL, WGET
51	Web Server	XSS	detect attacks IIS, Tomcat, apache
52	Web Specific Apps	XSS	vulnerabilities in specific web applications
53	WORM	worm	detect worm

จากนั้นจะนำข้อมูล Event\_Date กับ Event\_Count ไปบันทึกลงคอลัมน์ dateTime และ scoreType ของตารางชุดข้อมูลบูรณาการฯ และโมเดล LSTM ตามลำดับ

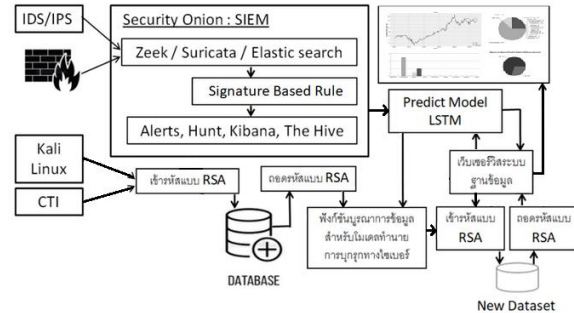
3.1.2 ข้อมูล Scan เป็นการหาข้อมูล IP เครื่องที่มีช่องโหว่ โดยค้นหาจากคอลัมน์ reak ซึ่งได้ข้อมูลมาจากผลการวิเคราะห์ช่องโหว่ที่ตรวจพบของเครื่องแต่ละ IP โดยมีเงื่อนไขว่า ถ้าหากตรวจพบว่ามีช่องโหว่เพียงหนึ่งใน

ช่องโหว่จาก ระบบปฏิบัติการ, Port, Service ที่เปิดอยู่, การแชร์ไฟล์ที่ไม่ปลอดภัยหรือการเปิด Wifi ที่ไม่ปลอดภัย และการไม่ติดตั้งโปรแกรม Antivirus จะกำหนดคะแนนในคอลัมน์ reack ของเครื่องนั้นเป็น 1 ซึ่งหมายความว่า เป็นเครื่องที่มีช่องโหว่ แต่ถ้าไม่พบเลยก็จะบันทึกเป็น 0 ซึ่งหมายความว่า เป็นเครื่องคอมพิวเตอร์ที่ปลอดภัย แล้วส่งไปสู่การเขียนรหัส สำหรับกระบวนการบันทึกผลลงตารางข้อมูลบูรณาการฯ จะเขียนเป็นรหัส โดยจะประกอบด้วย ลำดับระบบปฏิบัติการ, ลำดับหน่วยงาน และ หมายเลข Port เขียนเรียงติดกันก่อนที่จะบันทึกลงคอลัมน์ reack ของตาราง

**3.1.3 ข้อมูลข่าวกรองไซเบอร์** ข้อมูลจะถูกบันทึกโดยผู้ใช้งานผ่าน หน้าเว็บ โปรแกรมที่พัฒนาขึ้นประกอบด้วยข้อมูล วันเวลา, ชื่อเหตุการณ์, รายละเอียดข่าว, Type ของเหตุการณ์ซึ่งจะเป็นตัวเลือกอัตโนมัติที่อยู่ใน Type Cyber, ชื่อแหล่งที่มาของข่าว, Status สถานการณ์จะเป็นการบอกว่าเหตุการณ์นี้เคยเกิดขึ้นมาก่อนหรือไม่ ดำเนินต่อไปถึงเมื่อไหร่ หรือได้แก้ไขไปแล้ว ในการประมวลผลข้อมูลของฟังก์ชันบูรณาการข้อมูล ในส่วนนี้จะเป็นการนำข้อมูลของ Type ไปเขียนเป็นรหัสเพื่อเพิ่มคะแนนความเสี่ยงให้กับเหตุการณ์ภัยคุกคามทางไซเบอร์ ในคอลัมน์ reack ของตารางข้อมูลบูรณาการ ซึ่งจะหมายความว่า ถ้ามีการแจ้งข่าวที่มี Type ตรงกับเหตุการณ์โจมตี แสดงว่าเหตุการณ์ดังกล่าวได้ส่งผลกระทบต่อความเสี่ยงทางไซเบอร์ให้เพิ่มขึ้น แต่ต้องไม่เกิน 100 %

**3.2 ออกแบบฟังก์ชันการบูรณาการข้อมูลทางไซเบอร์** ข้อมูล Log ที่ Input เข้าสู่โปรแกรม Security Onion หลังจากได้รูปแบบภัยคุกคามทางไซเบอร์แล้วจะบันทึกลง Database ภายนอก เพราะ Security Onion จะเก็บข้อมูลไว้ไม่เกิน 7 วัน ตามด้วยการเข้ารหัสข้อมูลด้วยอัลกอริทึม RSA เช่นเดียวกับข้อมูลผลการทดสอบเจาะระบบและข้อมูลข่าวกรองทางไซเบอร์ โดยข้อมูลทั้งหมดจะแยกเก็บคนละตารางในฐานข้อมูล หลังจากนั้นข้อมูลจะถูกนำเข้าสู่ฟังก์ชันการถอดรหัสและฟังก์ชันบูรณาการข้อมูลฯ แล้วนำผลลัพธ์ที่ได้มาเก็บไว้ใน Database ตัวใหม่ ที่อยู่ในตารางเดียวกัน แล้วใช้โมเดล

ทำนายการบุกรุก LSTM ทำการทำนายข้อมูล ซึ่งจะทำงานควบคู่กับแสดงผลค่าความเสี่ยงและความพร้อมทางไซเบอร์ขององค์กร ดังภาพที่ 1



ภาพที่ 1: โครงสร้างฟังก์ชันบูรณาการข้อมูล

**3.3 การประมวลผลของโมเดล** เพื่อทดสอบว่า Dataset ที่ผ่านกระบวนการบูรณาการแล้ว สามารถทำงานกับ Model ทำนายการบุกรุกทางไซเบอร์ได้ เราจะนำข้อมูลจากคอลัมน์ dateTime และ scoreEven ซึ่งจะ เป็นข้อมูลวันที่และจำนวนรูปแบบการโจมตีที่เกิดขึ้นในแต่ละวัน ห้วงวันที่ 1 ม.ค. - 31 ธ.ค.2021 ทั้งหมด 337 แถว จากนั้นเปลี่ยนคอลัมน์ dateTime เป็น Index แล้วนำข้อมูลจากคอลัมน์ scoreEven พล็อตเป็นกราฟเพื่อดูโครงสร้างข้อมูลคร่าว ๆ แล้วทำ Data Scaling ให้ข้อมูลอยู่ในช่วง 0-1 เป็นการทำให้ข้อมูลไม่กระจายตัวมากเกินไปเวลาอยู่ในรูปของ Matrix แล้วทำ Data Transformation คือการเข้ารหัสข้อมูลหรือการแทนค่าข้อมูล เนื่องจาก ML ไม่สามารถประมวลผลข้อความแบบตรง ๆ ได้ เราต้องทำให้เป็นค่า Vector เสียก่อน โดยใช้อัลกอริทึม One-hot Encoding แล้วทำ Feature Selection หรือ Correlation Coefficients เพื่อคำนวณว่าข้อมูลที่เราที่มีอยู่มีฟีเจอร์ไหนที่ควรนำมาใช้ในการทดลองมากที่สุด โดยดูจากปริมาณข้อมูล ซึ่งมีค่าอยู่ระหว่าง -1 ถึง 1 ถ้าเข้าใกล้ -1 แสดงว่าข้อมูลเราน้อยเกินไป แต่ถ้าเข้าใกล้ 1 แสดงว่าข้อมูลเรามีมากพอที่จะใช้งาน หลังจากทำ Data Cleaning เสร็จแล้วก็จะทำการแบ่งข้อมูลเป็น Train, ปรับจูน และ Test เนื่องจากเรามีข้อมูลไม่มาก จึงใช้สัดส่วนข้อมูลเป็น 60% : 20% : 20% ตามลำดับ (200 : 68 : 69) โดย 20% เท่ากับเป็นการทำนายล่วงหน้าไปประมาณ 60 วัน (time step = 60) ซึ่งการเพิ่ม

หรือลดค่าดังกล่าวจะส่งผลต่อการ Overfitting หรือ Model มีความคลาดเคลื่อนสูง ก่อนจะนำเข้า model LSTM ต้องทำการเปลี่ยนจาก 2-dimensional ไปเป็น 3 - dimensional ก่อน โดยในกระบวนการ Train ข้อมูลจะใช้ optimizer = 'adam' และ Loss function = 'Mean Squared Error' เมื่อ Train เสร็จแล้วก็จะนำตัวโมเดล LSTM ไปทำการ Test ต่อ โดยใช้วิธีการเดียวกันกับ Train หลังจากการ Predict ต้องทำการ Inverse\_tranform กลับมาก่อน เนื่องจาก ขั้นตอนการ Train Data ได้ทำการ Scale Data ไปนั่นเอง

**3.4 การออกแบบส่วนแสดงผลการทำนาย (GUI)**  
แบ่งเป็น 3 ส่วนใหญ่ๆ คือ 1) กราฟเส้นแสดงผลการทำนาย ล่วงหน้า 5 วัน (Timestamp) 2) กราฟวงกลมแสดงค่าความเสี่ยงทางไซเบอร์ขององค์กร ณ เวลาปัจจุบันว่ามีความเสี่ยงเท่าใด โดยคำนวณจากปริมาณคอมพิวเตอร์ที่ถูกโจมตีหรือมีช่องโหว่ รวมทั้งถ้ามีการแจ้งเตือนด้านข่าวกรองฯ เข้ามาก็จะทำให้ความเสี่ยงเพิ่มขึ้น และ 3) กราฟวงกลมแสดงความพร้อมทางไซเบอร์ขององค์กร โดยคำนวณจากจำนวนจากปริมาณเครื่องคอมพิวเตอร์ทั้งหมดลบด้วยจำนวนเครื่องคอมพิวเตอร์ที่ถูกโจมตีหรือมีช่องโหว่

**3.5 การประเมินประสิทธิภาพการทำนาย LSTM**  
ใช้ฟังก์ชัน Root Mean Squared Error (RSME) เพื่อหาความแม่นยำ โดยได้ความแม่นยำอยู่ที่ 0.366 สำหรับ

โครงสร้างฟังก์ชันบูรณาการข้อมูลทางไซเบอร์และความสัมพันธ์ของแต่ละคอลัมน์ในตารางต่างๆ จะสามารถแสดงได้ดังภาพที่ 2

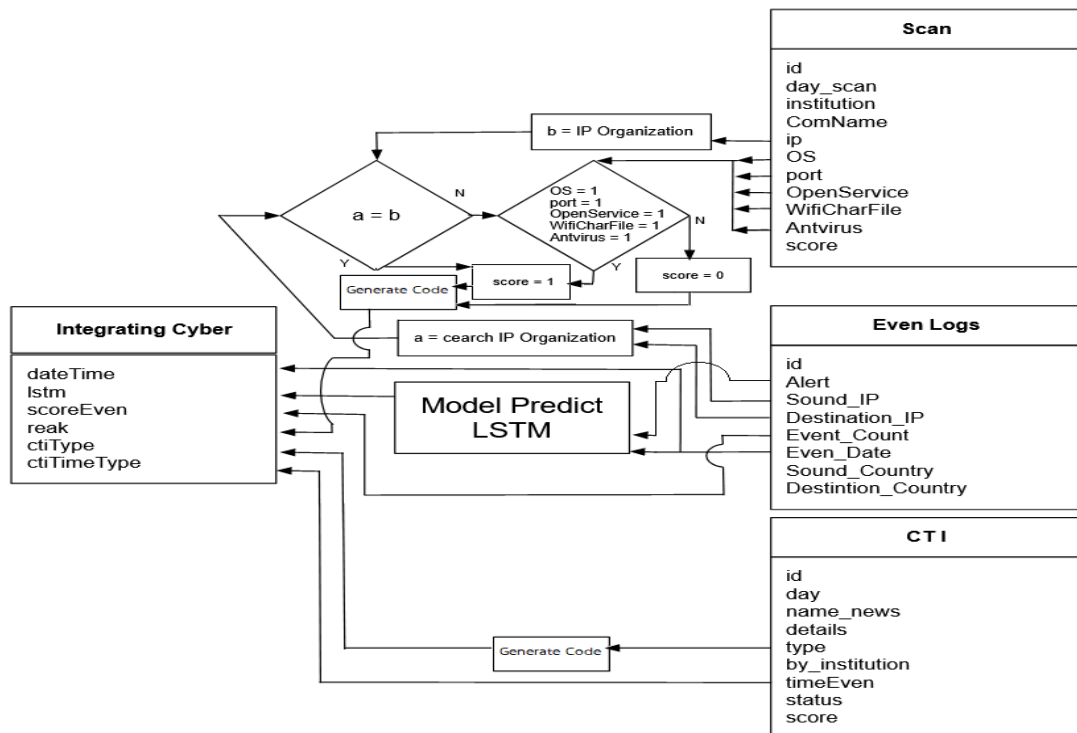
**4. ผลการดำเนินงาน**

**4.1 ผลการบูรณาการข้อมูลภัยคุกคามทางไซเบอร์** ซึ่งประกอบด้วยข้อมูลผลการทดสอบเจาะระบบ และข้อมูลด้านข่าวกรองไซเบอร์ จะได้ผลลัพธ์เป็น Dataset ใหม่สำหรับใช้ในการทดสอบโมเดลทำนายการบุกรุกทางไซเบอร์ ดังตารางที่ 3

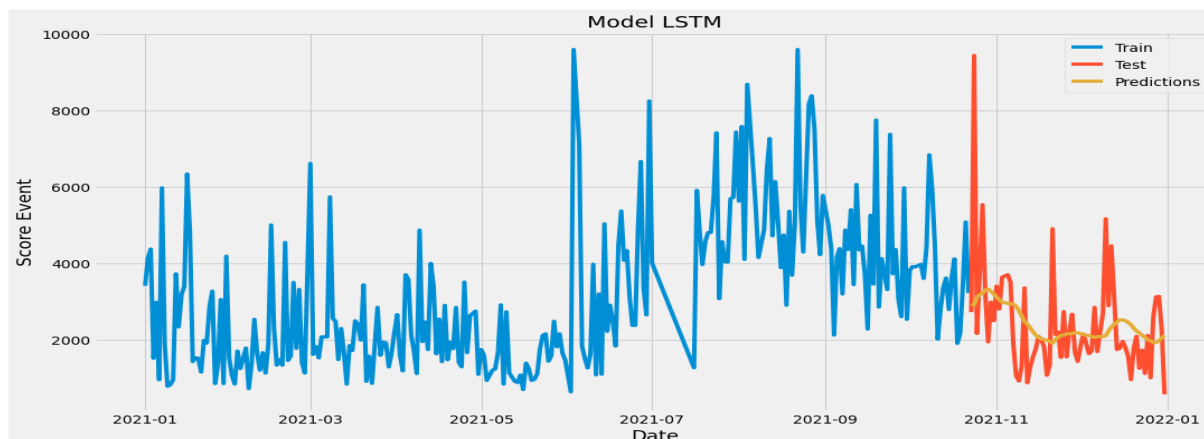
ตารางที่ 3: ข้อมูลบูรณาการทางไซเบอร์ บางส่วน

dateTime	lstm	scoreEven	ctiType
2021-01-01	5	3415	522071
2021-01-01	3	4138	510172
2021-01-01	14	1519	202161
.....	.....	.....	.....
2021-01-01	20	213445	202801
2021-01-01	12	1801234	202212

**4.2 ผลการทดลองกับโมเดลพื้นฐานสำหรับการทำนายการบุกรุกทางไซเบอร์ LSTM** โดยใช้ Dataset ที่ได้ผ่านการบูรณาการข้อมูลเรียบร้อยแล้ว จะแสดงผลออกมาพร้อมกับกราฟผลการทำนาย ดังภาพที่ 3



ภาพที่ 2: แผนผังฐานข้อมูลและฟังก์ชันบูรณาการข้อมูล



ภาพที่ 3: กราฟแสดงผลการทำนายด้วยโมเดลโครงข่ายประสาทเทียม LSTM

## 5. สรุป

งานวิจัยนี้ได้ค้นพบวิธีการบูรณาการชุดข้อมูลผลการตรวจจับการโจมตีทางไซเบอร์ ข้อมูลผลการทดสอบการเจาะระบบ และข้อมูลข่าวกรองไซเบอร์ เมื่อทำงานผ่านฟังก์ชันบูรณาการข้อมูลที่ได้พัฒนาขึ้นจะทำให้ได้ Dataset ตัวใหม่ ที่มีข้อมูลครอบคลุมทั้ง 3 ด้าน หากเป็น Dataset แบบเดิมจะสามารถบอกได้แค่ว่าภัยคุกคามคืออะไร ปริมาณเท่าใด และการโจมตีจากไหน หากแต่ Dataset ที่ผ่านกระบวนการบูรณาการข้อมูลฯ แล้ว จะสามารถบอกได้ถึงผลกระทบจากการโจมตีว่าจะสามารถกระจายไปได้มากน้อยเท่าใด และส่วนไหนขององค์กรจะมีความเสี่ยงทางไซเบอร์ที่เพิ่มขึ้นบ้าง จากการถอดรหัสคอลลัมน์ reat และ ctiType ซึ่งถูกสร้างความสัมพันธ์กับข้อมูลการโจมตี และข้อมูลข่าวกรองไซเบอร์แล้ว และเมื่อนำมาใช้กับโมเดลทำนายการบุกรุกโครงข่ายประสาทเทียม (LSTM) จะสามารถคาดการณ์เหตุการณ์ที่จะเกิดขึ้นในอนาคตได้ ทำให้องค์กรสามารถคาดการณ์ผลกระทบที่จะเกิดกับหน่วยงานและมีเวลาที่จะเตรียมแผนรับมือไว้ก่อนได้ผ่านฐานชุดข้อมูล (Dataset) เพียงตัวเดียว

## เอกสารอ้างอิง

- [1] Ahmed Elsharif, "Automatic Intrusion Detection System Using Deep Recurrent Neural Network Paradig," Naif Arab University for Security Sciences, Riyadh, Saudi Arabia, May 30 2018.
- [2] ธนกร มีหินกอง, "สถาปัตยกรรมความรู้ด้านความมั่นคงปลอดภัยไซเบอร์เพื่อสนับสนุนระบบตรวจหาการบุกรุกแบบปรับตัวด้วยเทคนิคกฎความสัมพันธ์" วารสารวิชาการพระจอมเกล้าพระนครเหนือ ปีที่ 25 ฉบับที่ 2 พ.ศ.-ศ.ศ. 2558 หน้า 277-288.
- [3] M.R.Amal, "Review of Cyber Attack Detection: Honeypot System," Noorul Islam Centre for Higher Education, India, Jan 1, 2022.
- [4] Pekka Lupari, "Detecting Anomalies in TLS Traffic Using Encrypted Traffic Analysis," JAMK University of Applied Sciences, May 2021, pp 49-51.
- [5] Proofpoint, "ET Category Descriptions." 2021, [Online]. Available: <https://tools.emergingthreats.net/docs/ETPro%20Rule%20Categories.pdf>.
- [6] Robert C. Newman, "Computer Security: Protecting Digital Resources", Jones and bartlett publishers, sudbury, massachusetts, usa, Jun 23, 2009, pp 268 - 413.
- [7] กอบเกียรติ สระอุบล, "เรียนรู้ Data Science และ AI: Machine Learning ด้วย Python," หสม มีเดีย เนทเวิร์ค, กรุงเทพฯ, 1 มิ.ย. 2563, หน้า 20-24.
- [8] Yonghyun Jo, "Cyberattack Models for Ship Equipment Based on the MITRE ATT&CK Framework," DSLAB Company Ltd, Seoul, Korea, Feb 26, 2022.
- [9] Kuncoro, Aditya Wibisono, S. T. Fayruz Rahma, and M. ENG, "Analisis Metode Open Web Application Security Project(OWASP) pada Pengujian Keamanan Website: Literature Review," Jurusan Informatika Fakultas Teknologi Industri, Universitas Islam Indonesia, Mar 1, 2022.

# ระบบการตรวจจับใบหน้าเพื่อการรู้จำใบหน้าที่มีประสิทธิภาพด้วยเทคโนโลยีปัญญาประดิษฐ์โดยใช้กล้องโทรทัศน์วงจรปิด

## Face Detection Systems for Effective Facial Recognition with Artificial Intelligence Using Closed Circuit Television Camera

อภิสิทธิ์ ซีลสัตย์ (Apisit Seelsat)<sup>1</sup> ประสงค์ ปราณีตพลกรัง (Prasong Praneetpolgrang)<sup>2</sup>

และพาทรรณ สงวนโภคัย (Pahron Sanguanbhokai)<sup>3</sup>

หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีป้องกันประเทศ

สำนักบัณฑิตศึกษา โรงเรียนนายเรืออากาศนวมินทกษัตริยาธิราช

<sup>1</sup>apisit.digital@gmail.com, <sup>2</sup>prasongspu@gmail.com, <sup>3</sup>pahron@gmail.com

### บทคัดย่อ

งานวิจัยนี้จึงมีจุดประสงค์เพื่อนำระบบรู้จำใบหน้า ที่พัฒนาด้วยเทคโนโลยีปัญญาประดิษฐ์มาประยุกต์ใช้กับ กล้องโทรทัศน์วงจรปิดที่คุณสมบัติพื้นฐานตรงตาม ประกาศของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม มาทดสอบหาประสิทธิภาพ ความแม่นยำในการตรวจจับ ใบหน้า และในขณะสวมใส่หน้ากากอนามัยหรือผ้าปิดปาก เทียบกับความสามารถในการติดตั้งกับระยะห่างของ ใบหน้าและความสูงของบุคคลว่าจะมีความแม่นยำในการ ระบุตัวตนบุคคลได้มากน้อยเพียงใด โดยผลการทดสอบ พบว่ากล้องโทรทัศน์วงจรปิดความละเอียด 2 ล้านพิกเซลมี ความแม่นยำในการตรวจจับใบหน้าตรงได้ ที่ระยะ 3 เมตร ได้ 100% สามารถรู้จำใบหน้ามนุษย์ที่ไม่ใส่หน้ากาก ได้สูงถึง 95.30% ในกรณีที่ใส่หน้ากากสามารถรู้จำได้ 90.70% โดยทำงานได้ดีในพื้นที่ทางเดินที่ถูกบังคับให้เดิน เป็นเส้นตรงผ่านไปยังกล้องเท่านั้นเพื่อให้กล้องสามารถ จับใบหน้าตรงได้ชัดเจน

**คำสำคัญ:** กล้องวงจรปิดตรวจจับใบหน้า การรู้จำใบหน้า การจดจำใบหน้าด้วยกล้องวงจรปิด

### Abstract

*This research had the objective to apply facial recognition system by artificial intelligence with CCTV having basic specifications consistent with the announcement of the Ministry of Digital Economy*

*and Society to test the efficiency of accuracy in facial detection and while wearing a mask with the height in the installation and distance and eight of people in order to see how accurate it can identify people. The test result found that CCTV with 2 Megapixels had accuracy in direct facial detection at 3 meters for 100%. It could recognize human faces with no mask up to 95.30%. In the case of wearing masks, it could recognize faces up to 90.70%. It could operate really well in the area forcing to walk in a direct line to the camera only so that the camera could detect the straight faces clearly.*

**Keywords:** Face Detection Camera, Face Recognition Camera, Facial Recognition, Face Capture.

### 1. บทนำ

ปัจจุบันกล้องโทรทัศน์วงจรปิดส่วนใหญ่ที่ใช้งานในประเทศไทย จะใช้สำหรับการบันทึกวิดีโอมากกว่าคอย แจ้งเตือนเหตุการณ์ที่ผิดปกติ จากปัญหาวิกฤติโรคติดต่อจากเชื้อไวรัสโคโรนา 2019 ผู้คนจำนวนมากต้องสวมใส่ หน้ากากอนามัยหรือผ้าปิดปาก เพื่อป้องกันการติดเชื้อ ส่งผลให้การจดจำใบหน้าเพื่อระบุตัวตนโดยใช้ กล้องโทรทัศน์วงจรปิดที่ไม่ได้ออกแบบจุดติดตั้งที่ เหมาะสมกับชนิดของกล้อง และขนาดของเลนส์นั้นไม่ สามารถทำงานได้จริง



ดังนั้น ผู้วิจัยจึงต้องการนำโมเดลในการตรวจจับใบหน้า และโมเดลในการรู้จำใบหน้า ที่พัฒนาจากเทคโนโลยีปัญญาประดิษฐ์มาทำงานร่วมกับกล้องโทรทัศน์วงจรปิดที่มีคุณลักษณะพื้นฐานตรงตามประกาศของกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมฉบับเดือนมิถุนายน 2564 [1] เพื่อทดสอบหาประสิทธิภาพ ความแม่นยำในการตรวจจับใบหน้า และในขณะสวมใส่หน้ากากอนามัย หรือผ้าปิดปาก เทียบกับความเสี่ยงในการติดตั้ง และระยะห่างของบุคคล ว่าจะมีความแม่นยำในการระบุตัวตนบุคคลได้มากน้อยเพียงใด เพื่อเป็นประโยชน์ในการออกแบบและติดตั้งระบบกล้องโทรทัศน์วงจรปิดที่ต้องการที่จะเก็บภาพใบหน้าบุคคล หรือใช้ระบุตัวอาชญากรบุคคลแปลกหน้าที่เข้ามาในพื้นที่ได้

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ระบบการตรวจจับใบหน้า (Face Detection)

การตรวจจับใบหน้า หมายถึง กระบวนการค้นหาใบหน้าของบุคคลจากภาพหรือวิดีโอ ซึ่งปัจจุบันก็มีอัลกอริทึมในการตรวจจับใบหน้าอยู่หลายวิธี เช่น เทคโนโลยี Deep Neural Network (DNN) [2] คือการจำลองรูปแบบการประมวลผลเลียนแบบสมองมนุษย์ โดยนำระบบโครงข่ายประสาทมาประมวลผลซ้อนกันหลาย ๆ ชั้น เพื่อหาจุดเด่นและจุดแตกต่างของข้อมูลในเชิงลึก ทำการเรียนรู้ข้อมูลตัวอย่างแล้วสรุปผลข้อมูลออกมา ซึ่งการตรวจจับใบหน้าจากรูปภาพ หรือวิดีโอ นั้น ตัวโมเดลที่ผ่านการเรียนรู้ใบหน้ามนุษย์แล้ว จะตรวจหาตาทั้งสองข้าง ปลายจมูก และริมฝีปาก โดยโมเดลนั้นจะกำหนดจุด Land Mark จำนวน 6 จุด ลงไปที่ภาพ ดังนี้ จุดที่ 1 ตาข้างซ้าย จุดที่ 2 ตาข้างขวา จุดที่ 3 ปลายจมูก จุดที่ 4 ริมฝีปาก จุดที่ 5 มุมปากซ้าย และจุดที่ 6 มุมปากขวา เมื่อกำหนดจุด Land Mark แล้วก็จะทำการตัดกรอบใบหน้านั้นออกมา เพื่อนำไปบันทึกเป็นรูปภาพใบหน้าที่ตรวจจับได้

### 2.2 ระบบการรู้จำใบหน้า (Face Recognition)

การรู้จำใบหน้า หมายถึง กระบวนการที่ได้นำภาพใบหน้าที่ตรวจจับได้และประมวลผลแล้วจากขั้นตอนการตรวจจับใบหน้ามาเปรียบเทียบกับฐานข้อมูลของใบหน้าเพื่อระบุว่าใบหน้าที่ตรวจจับได้ตรงกับบุคคลใด โดย

กระบวนการรู้จำใบหน้า นั้นจะต้องนำรูปภาพใบหน้ามา มาร์คจุดบนใบหน้า บริเวณรอบโครงสร้างของใบหน้า และหาระยะทางในแต่ละจุด เช่น ตำแหน่งจุดรอบ ๆ คาง ตำแหน่งจุดตามความยาวของคิ้วซ้าย และ คิ้วขวา ตำแหน่งจุดรอบ ๆ ดวงตาซ้าย และ ดวงตาขวา ตำแหน่งจุดของสันจมูกถึงปลายจมูก ตำแหน่งจุดรอบ ๆ ริมฝีปากบนและล่าง จากนั้นก็แปลงเป็นภาพ 3 มิติ เพื่อจัดองค์ประกอบของใบหน้าให้เป็นใบหน้าตรง และเข้าสู่กระบวนการ Deep Learning เพื่อหาเอกลักษณ์ของใบหน้า จากนั้นจึงนำมาเปรียบเทียบหาเอกลักษณ์ของใบหน้าที่ระหว่างภาพใบหน้าที่ตรวจจับได้ กับ ภาพใบหน้าในฐานข้อมูล เพื่อหาค่าที่ใกล้เคียงกันมากที่สุด หากค่าดังกล่าวมีค่ามากกว่า 85 % โมเดลจะทำนายภาพใบหน้านั้นว่าเป็นภาพใบหน้าบุคคลเดียวกัน

### 2.3 งานวิจัยที่เกี่ยวข้อง

Florian Schroff et al. [3] ได้นำเสนอผลงานวิจัยเรื่อง Facenet: A Unified Embedding for Face Recognition and Clustering โดยใช้โครงข่ายประสาทแบบคอนโวลูชัน (DCNN) ในการระบุรูปแบบการหาเอกลักษณ์ของวัตถุในรูปภาพและวิดีโอ โดยจะจำลองการมองเห็นของมนุษย์ที่มองเห็นที่เป็นที่ย่อย ๆ เพื่อดูว่าสิ่งที่เห็นอยู่เป็นอะไรกันแน่ เช่น พื้นที่ของภาพดวงตา, พื้นที่ของภาพจมูก, พื้นที่ของภาพปาก เป็นต้น และนำกลุ่มของพื้นที่ย่อย ๆ มาผสมกันเพื่อเข้าสู่กระบวนการ DCNN และ Triplet Loss Model ซึ่งเป็นการนำค่า Weights ผลลัพธ์ที่ต่ำสุดของ Distant ของภาพหลัก (Anchor) กับภาพของบุคคลเดียวกัน และค่า Weights ผลลัพธ์ที่สูงสุดของ Distant ของภาพหลัก (Anchor) กับภาพบุคคลอื่น มาเข้าสู่ตรรกะคำนวณของ Triplet Loss เพื่อให้ Model ได้เรียนรู้เปรียบเทียบภาพหลัก Anchor ว่าค่า Distant ที่ได้จากภาพ ที่เป็นบุคคลคนเดียวกัน กับภาพบุคคลอื่น มีความแตกต่างกันอย่างไร ผลการวิจัยพบว่าเมื่อทำการทดสอบหาประสิทธิภาพ LFW มีความแม่นยำอยู่ที่ 98.87% และเมื่อใช้วิธีการครอบตัดใบหน้าและจัดแต่งองค์ประกอบของใบหน้าทำให้สามารถทำลายสถิติของ LFW ที่ 99.63% โดยใช้เพียง 128 ไบต์ต่อหน้า

Jiankang Deng et al. [4] ได้นำเสนอผลงานวิจัยเรื่อง ArcFace: Additive Angular Margin Loss for Deep Face

Recognition โดยใช้หลักการของ Deep Convolutional Neural Network (DCNN) ของ FaceNet มาออกแบบฟังก์ชันลดการสูญเสีย (Softmax) ทำให้ระยะทางระหว่าง Deep Features และ Corresponding Class รวมกลุ่มกันมากยิ่งขึ้น มีขอบเขต (Boundary) ระหว่าง Class ชัดเจน ทำให้โมเดลสามารถจำแนกประเภทของ Class ได้ดีขึ้น ดังนั้นขั้นตอนวิธีแบบ Angular Distance จึงทำให้การจำแนกประเภททำได้ดียิ่งขึ้น และมีประสิทธิภาพเหนือกว่า FaceNet โดยมีความแม่นยำสูงสุดอยู่ที่ 99.83% (LFW)

### 3. วิธีการดำเนินการวิจัย

#### 3.1 วิเคราะห์และออกแบบการทำงานของระบบ

ระบบรู้จำใบหน้าในการวิจัยนี้จะประกอบไปด้วย 3 ส่วนหลัก ๆ ดังนี้

1) Input จะเป็นส่วนที่นำเหตุการณ์ต่าง ๆ ที่เกิดขึ้นบันทึกเป็นวิดีโอ และส่งต่อข้อมูลภาพและวิดีโอไปยังส่วนถัดไปผ่านระบบ Network โดยในสิ่งนี้จะประกอบไปด้วย

1.1) กล้องโทรทัศน์วงจรปิดชนิดเครือข่าย (IP Camera) ความละเอียดขนาด 1980\*1080 Pixel (2 MP)

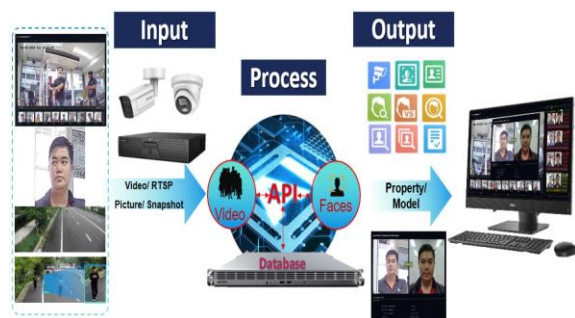
1.2) อุปกรณ์บันทึกภาพผ่านเครือข่าย

2) Process จะเป็นส่วนที่วิเคราะห์ข้อมูลใบหน้าเพื่อทำ Face Recognition โดยโมเดลการเรียนรู้เชิงลึก (Deep Learning) ที่ผ่านการ Training แล้ว จะถูกเขียนโปรแกรมไว้ที่ฝั่ง Process นี้ด้วยภาษา Python โดยใช้เครื่องมือ Anaconda เป็นตัวช่วยในการพัฒนาระบบ โดยเครื่องมือดังกล่าวสามารถติดตั้ง Tensor Flow ซึ่งเป็น Deep Learning Library จาก Google ที่สามารถใช้พัฒนาโปรแกรมที่มี Model Face Recognition ฝังลงใน Hardware ที่ทำงานภายใต้ระบบปฏิบัติการ Linux ได้ โดยข้อมูลจากฝั่ง Input จะถูกส่งมาประมวลผลที่ฝั่ง Process นี้ในรูปแบบ Video Streaming โดยใช้ RTSP Protocol พร้อมกับบันทึกข้อมูลภาพใบหน้าที่ตรวจจับได้เก็บไว้ใน Storage ของเครื่อง และมีระบบฐานข้อมูลเพื่อใช้เก็บข้อมูลใบหน้าที่ใช้สำหรับเปรียบเทียบ พร้อมทั้งบันทึกช่วงเวลาในระบบตรวจจับใบหน้าได้เพื่อใช้ในการเปิดดู (Playback) เหตุการณ์ย้อนหลังในรูปแบบไฟล์วิดีโอที่ถูกบันทึกไว้

ถ้าระบบตรวจจับใบหน้า ตรวจพบใบหน้ามนุษย์จากกล้องโทรทัศน์วงจรปิด ตัวโมเดลจะทำการประมวลเทียบกับภาพใบหน้าที่เก็บไว้ในฐานข้อมูลและทำนายผลออกมาเป็นค่าเปอร์เซ็นต์ความเหมือน ถ้ามีค่าเปอร์เซ็นต์มากกว่าค่าที่ตั้งไว้ในระบบ ระบบจะทำนายใบหน้าที่ตรวจจับได้เทียบกับภาพใบหน้าในฐานข้อมูลนั้นด้วยค่าเปอร์เซ็นต์ความเหมือนที่สูงที่สุด

3) Output จะเป็นส่วนของโปรแกรมที่เชื่อมต่อกับผู้ใช้งาน โดยจะถูกออกแบบ User Interface ให้มีหน้าในการแสดงผลวิดีโอที่ถูกส่งมาจากกล้องโทรทัศน์วงจรปิด พร้อมทั้งโชว์ข้อมูลภาพใบหน้าที่ระบบตรวจจับได้ จะนำภาพใบหน้ามาแสดงเพื่อเปรียบเทียบใบหน้าที่ตรวจจับได้ และ ภาพ ใบหน้าที่มีเปอร์เซ็นต์ความเหมือนสูงที่สุด พร้อมแสดงค่าและข้อมูลรายละเอียดของบุคคลในภาพที่เก็บไว้ในฐานข้อมูลออกมาแสดงผลทางฝั่งขวามือ

โปรแกรมที่ออกแบบมานี้จะสามารถกำหนด Library เพื่อจัดกลุ่มข้อมูลภาพที่ใช้ในการเปรียบเทียบ และสามารถที่จะกำหนดกลุ่มของภาพใน Library ให้แยกประเภทได้ เช่น กลุ่ม VIP, Black List, Normal เป็นต้น ถ้าภาพใบหน้าที่ตรวจจับได้ไม่มีความคล้ายคลึงกับรูปภาพใน Library ที่กำหนดไว้ สามารถจัดประเภทให้เป็นกลุ่มของบุคคลแปลกหน้าได้ หากรูปภาพที่ตรวจจับได้เมื่อเปรียบเทียบแล้วพบว่าอยู่ในกลุ่ม Black List สามารถกำหนดให้ระบบทำการแจ้งเตือนในรูปแบบ Pop-Up ได้ ทั้งนี้ฝั่ง Output นอกจากโปรแกรมที่จะใช้เป็นหน้า Monitoring แล้ว ยังมีหน้า History Playback, Log Search ที่สามารถค้นหาข้อมูลใบหน้าที่ตรวจจับได้ และเชื่อมโยงเพื่อดูวิดีโอเหตุการณ์ที่ตรวจจับใบหน้าย้อนหลังได้ดังภาพที่ 1



ภาพที่ 1: แสดง Diagram การทำงานของระบบ

### 3.2 การคัดเลือกโมเดลรู้จำใบหน้า

ปัจจุบันมีโมเดล Open Source ให้เลือกหลากหลายโมเดลที่สามารถตรวจจับใบหน้าและการรู้จำใบหน้าด้วยเทคโนโลยีปัญญาประดิษฐ์ มีนักพัฒนาสามารถเข้าถึงและนำไปพัฒนาปรับปรุงเอาไปใช้งานได้ ซึ่งผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องในการตรวจจับใบหน้าและการรู้จำใบหน้าจึงได้คัดเลือกโมเดล ArcFace ที่มีประสิทธิภาพสูงสุดในปัจจุบันมาพัฒนา และใช้ Pretrained Model ที่ได้ผ่านการ Training ด้วย Dataset รูปภาพใบหน้ามากกว่า 5,800,000 รูป มาพัฒนาต่อด้วย Dataset รูปภาพใบหน้าที่ได้เตรียมไว้

### 3.3 การเตรียมสภาพแวดล้อมของระบบในการพัฒนา

ในขั้นตอนนี้จะต้องมีการเตรียมระบบที่จะใช้พัฒนาโดยใช้เครื่องมือ Anaconda ที่ติดตั้ง Python 3.6 และ ติดตั้งเฟรมเวิร์ค Library เช่น Tensor Flow 1.7.0 และ Mxnet 1.3.1 จากนั้นนำโมเดล MSIM ArcFace [5] ที่ผ่านการ Training Dataset ใบหน้าไว้เรียบร้อยแล้ว ซึ่งเรียกว่า Pretrained Model โดยโมเดลดังกล่าวจะอยู่ในรูปแบบไฟล์ Mxrec และทำการแปลงให้เป็นรูปแบบ Tfrecored ได้ โดยมีรูปภาพใบหน้าทั้งหมด 5.8 ล้านรูป โดยมีรูปที่ระบุตัวตนได้จำนวน 85,000 รูป ซึ่ง Model ดังกล่าวมีความแม่นยำในการทำนายอยู่ที่ 99.53% (LFW) ในตารางที่ 1 จะเป็นตัวอย่าง Dataset ที่นำมาใช้เป็น Pretrained Model

ตารางที่ 1: ตัวอย่าง Pretrained Model

Face Recognition Datasets	Identities	Image
IMDB-Face [6]	59K	1.7M
Glnt360K [7]	360K	17M
MS1M-IBUG [8]	85K	3.8M
MegaFace(train) [9]	672K	4.7M

### 3.4 ขั้นตอนการเก็บข้อมูล Dataset และ Training Model

การรวบรวมข้อมูล Dataset จะใช้กล้อง IP Camera ยี่ห้อ Hikvision ที่มีความละเอียด 4 MP มาเก็บภาพใบหน้ามนุษย์ โดยนำกล้องไปติดตั้งบริเวณทางเดินที่ผู้คนจะต้องเดินตรงผ่านไปยังกล้องเพื่อให้ได้รูปภาพใบหน้าตรง

เครื่องมือที่ใช้ในการรวบรวมข้อมูล Dataset จะเลือกใช้งานกล้องวงจรปิดที่มีเลนส์ขนาด 2.8 mm โดยระยะห่าง

ของกล้องกับใบหน้าต้องไม่เกิน 4 เมตร ที่ความละเอียด 4MP โดยระยะตำแหน่งจุดยืนที่กล้องตรวจจับได้ไปยังจุดที่ติดตั้งกล้อง Distance (D) สามารถคำนวณได้จากสูตรทฤษฎีบทพีทาโกรัส ตามสมการ (1) ดังนี้

$$D^2 = I^2 - L^2 \tag{1}$$

โดยที่

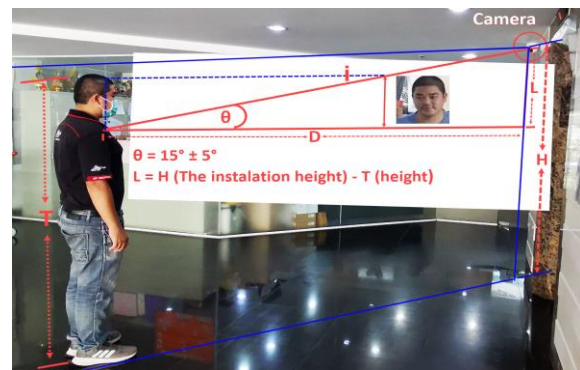
H คือ ความสูงของของกล้องที่ติดตั้ง

T คือ ความสูงจากหัวจรดพื้น

L คือ H - T

I คือ Identification Distance

D คือ Distance ระยะตำแหน่งจุดยืน



ภาพที่ 2: การหาระยะการตรวจจับใบหน้าที่ชัดเจนจากกล้อง

ในขั้นตอนการ Training Model จะดำเนินการ Training Dataset ชุดใหม่ และ Fine Tuning ด้วยการ ใช้รูปภาพ Dataset ใบหน้าที่มีขนาดของภาพที่ 500 Pixel โดยมีการกำหนดสัดส่วนของใบหน้าที่ความกว้าง 2 ส่วน ความสูงส่วนหัว 1.5 ส่วน และส่วนใต้คาง (Body) 0.5 ส่วน ซึ่งภาพ Dataset นั้นเป็นภาพใบหน้าของพนักงานในบริษัท ดิจิตอล โฟกัส จำกัด และรูปภาพบุคคลที่ปกอาศัยอยู่ในบางกอก-ฮอไรซอนคอนโดมิเนียม รวมค่าแห่ง จำนวน 10,000 รูป เมื่อเตรียม Dataset เรียบร้อยแล้ว นำไปเทรนนิ่งโมเดล ที่ใช้สถาปัตยกรรม ResNet-100 ร่วมกับฟังก์ชันลดการสูญเสียของ ArcFace เพื่อสร้าง Model Face Recognition ขึ้นมาใหม่ ซึ่งการ Training ในแต่ละครั้งกับชุด Dataset จะได้อ่านน้ำหนัก (Weight) เพื่อใช้สำหรับการทำ Fine Tuning เพิ่มเติมในครั้งถัดไป



ภาพที่ 3: ตัวอย่างภาพ Dataset ใบหน้าบุคคล

### 3.5 อัลกอริทึมในการตรวจจับใบหน้าและการรู้จำใบหน้า

ผู้วิจัยได้นำ DCNN มาใช้ในการตรวจจับใบหน้า ซึ่งจะตรวจหาดวงตา จมูก และ ปาก และทำการตัดกรอบพื้นที่ที่ตรวจเจอ เช่น พื้นที่ของภาพดวงตา, พื้นที่ของภาพจมูก และนำกลุ่มของพื้นที่ย่อย ๆ มาผสานกันเป็นภาพใบหน้า วิธีนี้จะช่วยให้การทำงานได้รวดเร็วขึ้น ในการจำแนกใบหน้าเพื่อหาเอกลักษณ์บนใบหน้า จะทำการกำหนดจุดบนใบหน้า Face Alignment [10] จากนั้นเข้าสู่การครอบตัดเฉพาะใบหน้าออกมาและแปลงภาพใบหน้า 2 มิติ ให้กลายเป็น 3 มิติ ในขั้นตอนนี้จะสามารถนำภาพใบหน้าที่มีการหันซ้าย-ขวา หรือ ก้ม-เงย จากการหามุมองศาของใบหน้าที่ได้จากโมเดล 3 มิติ เพื่อให้ภาพใบหน้าที่ตรวจจับได้นั้นมีเอกลักษณ์มากพอกับภาพใบหน้าที่ตรงที่จะนำมาเปรียบเทียบ จากนั้นใช้โมเดล ArcFace เข้าสู่กระบวนการ DCNN เพื่อหาระยะทางจากจุดต่าง ๆ ที่เป็นเอกลักษณ์บนใบหน้า ซึ่งทำให้การจำแนกใบหน้าได้แม่นยำมาก

### 3.6 การเชื่อมต่อการทำงานของระบบ Face Recognition กับกล้องโทรทัศน์วงจรปิด

ในขั้นตอนนี้ จะนำกล้องโทรทัศน์วงจรปิดชนิดเครือข่าย ความละเอียด 2MP รุ่น DS-2CD1327G0-L ยี่ห้อ Hikvision มาเชื่อมต่อกับเครื่องบันทึกแบบ NVR ผ่านระบบเครือข่าย LAN

## 4. ผลการดำเนินงาน

### 4.1 ผลการทดสอบประสิทธิภาพโมเดลด้วย Test Dataset

Test Dataset ที่นำมาใช้ทดสอบประสิทธิภาพของ Model ได้เลือกใช้ Dataset จาก LFW ซึ่งมีรูปภาพที่ระบุตัวตนได้ จำนวน 5,749 รูป จากจำนวน Dataset ทั้งหมด 13,233 รูป โดยกำหนดค่าความมั่นใจให้กับโมเดลที่ 75% ขึ้นไป ซึ่งผลลัพธ์การทดสอบ ตัวโมเดลมีความแม่นยำในการทำนายรูปภาพที่ระบุตัวตนได้ จากจำนวน 5,749 รูป

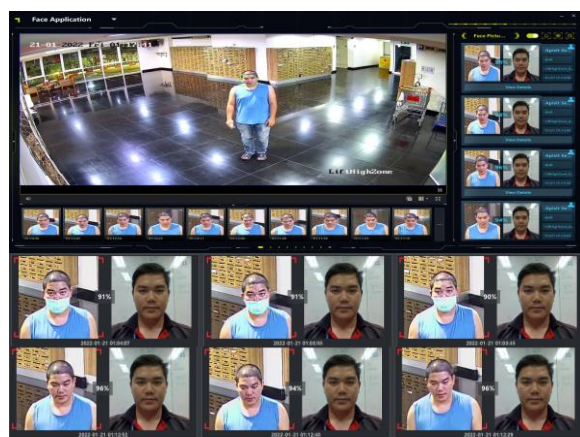
สามารถทำนายได้ถูกต้องจำนวน 5,722 รูป คิดเป็น 99.53% ซึ่งเท่ากับค่าความแม่นยำของ Pretrained Model ในส่วนที่โมเดลทำนายผิด เนื่องจากเอกลักษณ์บนใบหน้าของ Test Dataset มีความคล้ายกับ Verify Dataset มากกว่า 85%



ภาพที่ 4: ตัวอย่างภาพแสดงผลลัพธ์ที่โมเดลได้ทำนายผลถูกต้อง

### 4.2 ผลการทดสอบประสิทธิภาพของระบบโดยการใช้งานร่วมกับกล้องโทรทัศน์วงจรปิด

การทดสอบประสิทธิภาพจะใช้กล้องโทรทัศน์วงจรปิดชนิดเครือข่าย ที่มีความละเอียดขนาด 2 MP และกล้องความละเอียด 1 MP โดยทำการทดสอบตรวจจับใบหน้าที่ใส่หน้ากากเปรียบเทียบกับใบหน้าที่ไม่ได้ใส่หน้ากากที่ระยะห่างจากจุดที่กล้องโทรทัศน์วงจรปิดได้ถูกติดตั้งไว้ที่เพดาน ที่ความสูง 2.5 เมตร จนถึงจุดที่ผู้วิจัยได้ยื่นทดสอบที่ระยะ 1.5 เมตร, 3 เมตร, และ 6 เมตร โดยได้ยื่นหน้าเข้าไปในทิศทางที่กล้องโทรทัศน์วงจรปิดได้ถูกติดตั้งไว้และมองหน้าตรงผ่านกล้องไปยังด้านหน้า ดังภาพที่ 5



ภาพที่ 5: แสดงภาพใบหน้าที่ตรวจจับได้อย่างถูกต้อง ทั้งแบบใส่หน้ากากและไม่ใส่หน้ากาก

ในการตรวจจับใบหน้าที่ตรงที่ไม่ใส่หน้ากาก ที่ระยะห่างจากกล้องโทรทัศน์วงจรปิดที่ติดตั้งอยู่บนเพดานที่ความสูง 2.5 เมตร มายังตำแหน่งที่ผู้วิจัยได้ยื่นทำการทดสอบห่างจาก

กล้องโทรทัศน์วงจรปิดในระยะ 3 เมตรนั้น จะ ได้ความแม่นยำที่สูงที่สุด โดยมีความแม่นยำในการรู้จำใบหน้า 95.30% เมื่อเปรียบเทียบกับการใช้หน้ากากพบว่าความแม่นยำในการรู้จำใบหน้า 92.60% ในส่วนระยะห่าง 1.5 เมตร ที่ใกล้ที่สุดแต่ค่าความแม่นยำได้น้อยกว่า ซึ่งเป็นผลมาจากกล้องถูกติดตั้งสูงกว่าใบหน้าจึงทำให้กล้องจับใบหน้าได้มุมมองก้มเป็นไปตามตารางที่ 2

ตารางที่ 2: ค่าความแม่นยำในการตรวจจับใบหน้าและรู้จำใบหน้า

ความละเอียดของกล้อง	ระยะห่าง	ตรวจจับการใส่หน้ากาก	ความแม่นยำในการตรวจจับใบหน้า	ความแม่นยำในการรู้จำใบหน้า		
				ต่ำสุด	สูงสุด	ค่าเฉลี่ย
1080P (2MP)	3	ไม่ใส่หน้ากาก	100.00%	94.00%	96.00%	95.30%
720P (1MP)	3	ไม่ใส่หน้ากาก	100.00%	93.00%	96.00%	94.30%
1080P (2MP)	1.5	ไม่ใส่หน้ากาก	100.00%	91.00%	95.00%	92.60%
720P (1MP)	1.5	ไม่ใส่หน้ากาก	100.00%	91.00%	94.00%	92.10%
1080P (2MP)	6	ไม่ใส่หน้ากาก	100.00%	83.00%	94.00%	92.00%
1080P (2MP)	3	ใส่หน้ากาก	100.00%	90.00%	92.00%	90.70%
1080P (2MP)	1.5	ใส่หน้ากาก	83.33%	74.00%	91.00%	82.60%
720P (1MP)	1.5	ใส่หน้ากาก	58.82%	71.00%	90.00%	79.20%
720P (1MP)	3	ใส่หน้ากาก	83.33%	70.00%	84.00%	77.00%
1080P (2MP)	6	ใส่หน้ากาก	90.90%	70.00%	85.00%	71.18%
720P (1MP)	6	ไม่ใส่หน้ากาก	0.00%	0.00%	0.00%	0.00%

### 4.3 ผลการทดสอบประสิทธิภาพของระบบโดยการใช้งานร่วมกับกล้องโทรทัศน์วงจรปิดรุ่นอื่นที่แตกต่างกัน

การทดสอบนี้ได้นำกล้องกล้องโทรทัศน์วงจรปิดที่มีความละเอียด 2 MP ที่สามารถรองรับ Protocol RTSP หรือ Onvif มาทำการทดสอบ ผลการทดสอบ พบว่ากล้องโทรทัศน์วงจรปิดที่แตกต่างกันนั้นถึงแม้จะมีความละเอียดที่เท่ากัน แต่จะให้คุณภาพของภาพที่แตกต่างกัน ซึ่งส่งผลทำให้ประสิทธิภาพในการตรวจจับไม่เท่ากัน ดังนั้นกล้องโทรทัศน์วงจรปิดที่ไม่สามารถกำหนดค่า Bitrate, Frame Rate, Shutter Speed รวมถึงมีค่าของรูรับแสงที่แตกต่างกันจะมีผลต่อประสิทธิภาพในการรู้จำใบหน้า

ตารางที่ 3: ประสิทธิภาพในการรู้จำใบหน้าของระบบ

ยี่ห้อ	รุ่น	ความแม่นยำในการตรวจจับ	ความแม่นยำในการรู้จำใบหน้า		
			ต่ำสุด	สูงสุด	ค่าเฉลี่ย
Hikvision	DS-2CD1327G0-L	100.00%	94.00%	96.00%	95.30%
Hilook	THC-B129-M	100.00%	89.00%	95.00%	94.50%
Dahua	IPC-HDW2230T-AS-S2	100.00%	86.00%	94.00%	91.80%
Hiview	HP-97D20STPE	100.00%	82.00%	93.00%	90.50%

## 5. สรุป

การนำระบบรู้จำใบหน้ามาประยุกต์ใช้กับกล้องโทรทัศน์วงจรปิดที่มีคุณลักษณะพื้นฐานตรงตามประกาศของ

กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมฉบับ พ.ศ. 2564 พบว่า ความแม่นยำในการตรวจจับใบหน้าและรู้จำใบหน้ามนุษย์ที่ใส่หน้ากากนั้นสามารถทำได้ในพื้นที่ทางเดินที่ถูกบังคับให้เดินเป็นเส้นตรงผ่านไปยังกล้อง สำหรับภาพใบหน้าที่ไม่ใส่หน้ากากนั้น ระบบยังคงรู้จำใบหน้าได้ โดยที่ใบหน้านั้นจะต้องทำมุมเฉียงซ้ายขวาไม่เกิน 60° และก้มเงยไม่เกิน 45° ทั้งนี้ จุดติดตั้งกล้อง ขนาดของเลนส์ที่เลือกใช้ และมุมมองของกล้องที่ติดตั้งจะมีผลต่อความแม่นยำในการรู้จำใบหน้าด้วยเช่นกัน

## เอกสารอ้างอิง

- [1] กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม. เกณฑ์ราคากลางและคุณลักษณะพื้นฐานของระบบกล้องโทรทัศน์วงจรปิดฉบับเดือนมิถุนายน 2564, สืบค้นเมื่อ 29 พฤศจิกายน 2564. จาก. <https://www.mdes.go.th/service/87>
- [2] N. Prasad and B. Rajpal, "Frontal and non-frontal face detection using deep neural networks (DNN)," *International Journal of Research in Industrial Engineering*, Vol. 10, No. 1, pp. 9-21, 2021.
- [3] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *CVPR*, pp. 815–823, 2015.
- [4] J. Deng, J. Guo, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," *arXiv preprint arXiv:1801.07698*, 2018.
- [5] Y. Guo, L. Zhang, Y. Hu, X. He and J. Gao, "Mscceleb-1m: A dataset and benchmark for large-scale face recognition," *ECCV*, 2016.
- [6] W. Fei, C. Liren, L. Cheng, H. Shiyao, C. Yanjie, Q. Chen, Loy and C. Change, "The devil of face recognition is in the noise," *ECCV*, 2018.
- [7] X. An, X. Zhu, Y. Xiao, L. Wu, M. Zhang, Y. Gao, B. Qin, D. Zhang and Y. Fu, "Partial FC: Training 10 million identities on a single machine," *arxiv:2010.05222*, 2020. s
- [8] J. Deng, Y. Zhou, S. Zafeiriou, "Marginal loss for deep face recognition," *CVPRW*, 2017.
- [9] N. Aaron, K. Ira, "Level playing field for million scale face recognition," *CVPR*, 2017.
- [10] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, pp. 1499-1503, 2016.

# การเปรียบเทียบแบบจำลองการเรียนรู้เชิงลึกเพื่อการระบุภาพทางสัณฐานวิทยา ของเห็ดป่า

## Comparison of Deep Learning Models for Morphological Image Identification of Wild Mushrooms

วชรพล เกตวงษา (Wacharaphol Ketwongsa)<sup>1</sup> และอุรฉัตร โคแก้ว (Urachart Kokaew)<sup>2</sup>

<sup>1</sup>วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

<sup>1</sup>k.wacharaphol@kkumail.com, <sup>2</sup>urachart@kku.ac.th

### บทคัดย่อ

งานวิจัยฉบับนี้เพื่อเปรียบเทียบความแม่นยำในการจำแนกประเภทของเห็ดมีพิษและไม่มีพิษโดยใช้ CNN และความแม่นยำในการระบุตำแหน่งและจำแนกประเภทของเห็ดมีพิษและไม่มีพิษโดยใช้ R-CNN เพื่อที่จะลดอัตราความเสี่ยงที่อาจทำให้เกิดการเสียชีวิตจากการบริโภคเห็ดที่มีพิษ โดยใช้สถาปัตยกรรมทั้งหมด 3 แบบ ได้แก่ AlexNet, ResNet-50 และ GoogLeNet ใช้เห็ดทั้งหมด 5 ชนิดแบ่งออกเป็นเห็ดมีพิษ 2 ชนิดและเห็ดไม่มีพิษ 3 ชนิด ผลการทดลองในการจำแนกประเภทของเห็ดมีพิษและไม่มีพิษโดยใช้ CNN สถาปัตยกรรม ResNet-50 และ GoogLeNet มีอัตราความถูกต้องมากที่สุด 99.50% ResNet-50 มีค่า Precision 99.33%, Recall 100%, F1 score 99.66% GoogLeNet มีค่า Precision 99.35%, Recall 100%, F1 score 99.67% และอัตราความถูกต้องในการระบุตำแหน่งและจำแนกประเภทของเห็ดมีพิษและไม่มีพิษโดยใช้ R-CNN สถาปัตยกรรม ResNet-50 มีอัตราความถูกต้องมากที่สุดที่ 96.50% มีค่า Precision 100%, Recall 95.24%, F1 score 97.56%

คำสำคัญ: AlexNet, GoogLeNet, เห็ด, ResNet-50

### Abstract

This study was to compare the classification accuracy of poisonous and edible mushrooms using convolutional neural networks (CNN) and region convolutional neural networks (R-CNN). In order to reduce the risk of death from eating poisonous mushrooms. Compare the testing time and accuracy

of three pretrained models: AlexNet, ResNet-50, and GoogLeNet. Mushroom datasets are divided into 2 poisonous and 3 eating. From the experiments in the mushroom classification experiments using CNN ResNet-50 and GoogLeNet have the highest accuracy at 99.50%, ResNet-50 has a precision 99.33%, recall 100%, F1 score 99.66%, GoogLeNet has a precision 99.35%, recall 100%, F1 score 99.67%. And using R-CNN ResNet-50 has the highest accuracy at 95.50%. precision 100%, recall 95.24%, F1 score 97.56%.

**Keywords:** AlexNet, GoogLeNet, Mushroom, ResNet-50.

### 1. บทนำ

ประเทศไทยเป็นประเทศที่มีความหลากหลายทางชีวภาพซึ่งมีสิ่งมีชีวิตและพันธุ์พืชหลากหลายรูปแบบ หนึ่งในนั้นคือเห็ด ซึ่งเป็นพืชผักที่มีคุณค่าทางโภชนาการสูงและเป็นอาหารของคนไทยมาช้านาน คนไทยนิยมนำเห็ดมาบริโภคเพราะว่ามีโปรตีนสูง แคลอรีต่ำ มีไขมันที่มีประโยชน์ต่อร่างกาย มีวิตามินหลายชนิดและมีธาตุเหล็ก และยังเป็นพืชเศรษฐกิจสร้างรายได้ให้กับเกษตรกร เห็ดจะพบได้ช่วงหน้าฝนเดือนพฤษภาคมถึงกันยายนหลังฝนตก 2-3 วัน และจะพบได้มากในป่าเต็งรัง ป่าเบญจพรรณ และป่าดิบแล้งทางภาคเหนือและภาคตะวันออกเฉียงเหนือของประเทศไทย เห็ดทั่วโลกมีประมาณล้านถึงสองล้านชนิด [1-3] สามารถแบ่งออกเป็น 2 ประเภทคือ เห็ดมีพิษและเห็ดไม่มีพิษ ซึ่งเห็ดมีพิษบางชนิดจะมัลักษณะใกล้เคียง

กับเห็ดไม่มีพิษ การจำแนกเห็ดพิษตามภูมิปัญญาชาวบ้านคือการนำเห็ดมาต้มกับข้าวหากข้าวเปลี่ยนสีแสดงว่าเป็นเห็ดมีพิษ หรือการต้มเห็ดแล้วนำช้อนเงินมาคนถ้าช้อนเป็นสีดำแสดงว่าเป็นเห็ดพิษ ซึ่งวิธีที่กล่าวมาไม่สามารถแยกเห็ดมีพิษได้ทุกชนิดเนื่องจากเห็ดมีพิษบางชนิดไม่มีปฏิกิริยากับการทดลองดังกล่าว ลักษณะทางสัณฐานวิทยาพื้นฐานของเห็ดมีพิษจะมีสีสันสดใส มีเกล็ดบนหมวกเห็ด และมีวงแหวนอยู่ใต้หมวกเป็นต้น ซึ่งหากคนเก็บไม่มีความชำนาญอาจก่อให้เกิดอันตรายถึงชีวิตหากนำไปบริโภค เห็ดมีพิษจะส่งผลกระทบต่อประสาท และทำให้เสียชีวิตได้หากบริโภคในปริมาณมากหรือคนที่มียาแพ้ และคนไทยเสียชีวิตจากการบริโภคเห็ดมีพิษต่อเนื่องทุกปี

การเรียนรู้เชิงลึก (Deep learning) เป็นส่วนหนึ่งของการเรียนรู้ด้วยเครื่อง (Machine learning) ซึ่งทำการเรียนรู้โดยอัตโนมัติโดยเลียนแบบการทำงานของสมองของมนุษย์ โดยเป็นการสร้างโครงข่ายประสาทเทียมหลายๆชั้นเพื่อที่จะให้สามารถประมวลผลที่มีความซับซ้อนมากได้และให้ความแม่นยำสูง ซึ่งความแม่นยำจะขึ้นอยู่กับข้อมูลที่ใช้ในการฝึกสอน ยิ่งชุดข้อมูลที่ใช้ฝึกสอนมีขนาดใหญ่ยิ่งความแม่นยำมาก ปัจจุบันการเรียนรู้เชิงลึกถูกนำมาใช้สำหรับการจำแนกวัตถุเช่น การจดจำใบหน้า [4,5] การตรวจหาโรคพืช [6,7] ยานยนต์อัตโนมัติ [8,9] เป็นต้น การเรียนรู้เชิงลึกประกอบไปด้วย 3 ส่วน ได้แก่ input layer, hidden layer และ output layer

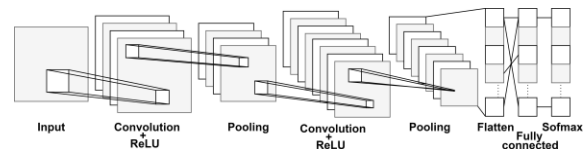
การศึกษาการเรียนรู้เชิงลึกนี้เพื่อเปรียบเทียบความแม่นยำและเวลาในการฝึกสอนของ 3 สถาปัตยกรรมได้แก่ AlexNet, ResNet และ GoogLeNet โดยใช้ชุดข้อมูลเห็ดทั้งหมด 5 ชนิดแบ่งออกเป็นเห็ดไม่มีพิษ 3 ชนิดได้แก่ เห็ดระโงก เห็ดตับเต่า และเห็ดตะไค เห็ดมีพิษ 2 ชนิดได้แก่ เห็ดระโงกขาว และเห็ดหมวกเงิน

## 2. งานที่เกี่ยวข้อง

### 2.1 Convolutional Neural Network (CNN)

เป็นอัลกอริทึมการเรียนรู้เชิงลึกที่นิยมนำมาใช้ในการวิเคราะห์ภาพและการจดจำวัตถุ โดยนำภาพมาแยกคุณลักษณะเพื่อที่จะดูว่าของสิ่งนั้นคืออะไร สถาปัตยกรรม

ของ CNN ประกอบไปด้วย Input, Convolution Rectified Linear Unit (ReLU), Pooling, Flattening, Fully Connected และ Softmax [10,11] เลขอร์ที่พบได้บ่อยคือ Convolution, activation หรือ ReLU และ pooling ในหนึ่งโครงสร้างอาจจะมีชั้นพวกนี้ได้มากกว่า 1 ชั้นเพื่อเพิ่มความซับซ้อนในการคำนวณ



ภาพที่ 1: สถาปัตยกรรมการเรียนรู้เชิงลึกแบบ CNN

Input layer คือชั้นนำเข้าข้อมูล ทำหน้าที่กำหนดขนาดและมีมิติของภาพ เช่น ภาพขนาด  $227 \times 227 \times 3$  pixels

Convolution layer จะใช้ตัวกรองตรวจ (filter) จับคุณลักษณะต่างๆ ตัวกรองจะเคลื่อนผ่านแต่ละส่วนในภาพเพื่อหาลักษณะเด่น [12] เช่น ขอบ สี รูปทรง เป็นต้น ขนาดของตัวกรองที่นิยมใช้คือ  $3 \times 3$  และ  $5 \times 5$

Rectified Linear Unit (ReLU) layer เป็นฟังก์ชันการเปิดใช้งานแบบไม่เชิงเส้น เป็นชั้นที่ทำให้การเรียนรู้ข้อมูลมีความรวดเร็ว และได้ประสิทธิภาพมากขึ้น ชั้นนี้จะคืนค่า 0 หากอินพุตเป็นลบ และจะคืนค่าเท่ากับอินพุตที่เข้ามามากกว่า 0 เมื่ออินพุตเป็นบวก

Pooling layer คือการลดความซับซ้อนของเอาต์พุตเพื่อลดเวลาในการฝึก โดยทำได้ 2 วิธีคือ Max Pooling และ Average Pooling โดย Max Pooling จะเอาค่าที่มากที่สุด Average Pooling จะหาเอาค่าเฉลี่ยทั้งหมด

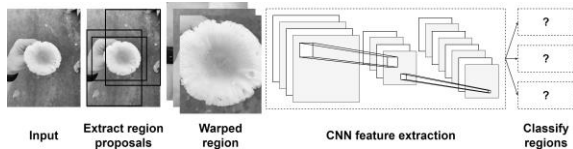
Flattening layer เป็นการทำให้ Pooling Feature Map ที่ได้ทำเป็นคอลัมน์เดียวกัน เพื่อความสะดวกในการวิเคราะห์ข้อมูล

Fully Connected layer เป็นส่วนที่นำเอาลักษณะสำคัญที่เราได้มาจากการหาลักษณะเด่นของภาพในชั้นก่อนหน้านี้มาทำการสร้างเป็น Neural Network สำหรับการเรียนรู้และทำนายประเภทของรูปภาพ

Softmax layer เป็นฟังก์ชันกำหนดความน่าจะเป็นให้การทำนายสำหรับแต่ละคลาส เช่น นำภาพสุนัขเข้าไปถามว่าคืออะไร Softmax จะตอบมาว่าโอกาสที่เป็น สุนัขกี่เปอร์เซ็นต์ แมวกี่เปอร์เซ็นต์ เป็นต้น

### 2.2 Region Convolutional Neural Network (RCNN)

CNN ถูกออกแบบมาเพื่อการจำแนกวัตถุ แต่ไม่สามารถแก้ปัญหาการตรวจจับวัตถุและวาดกรอบล้อมรอบวัตถุได้ [13] จึงมีการเสนอวิธี R-CNN เพื่อใช้ในการตรวจจับวัตถุต่างๆ หลังจากมี R-CNN ก็มีอีกหลายโมเดล เช่น Fast-RCNN, Faster-RCNN เป็นต้น R-CNN ประกอบด้วย 4 อย่าง อย่างแรก การเสนอพื้นที่ในภาพที่อาจจะมีวัตถุที่สนใจด้วยวิธี Selective Search จำนวน 2,000 region proposals แต่ละ region proposals จะทำการปรับขนาดให้มีความเท่ากันและส่งไปยัง CNN อย่างที่สอง เมื่อส่งไป CNN จะทำการสกัดคุณลักษณะเด่นของแต่ละ region proposals [14] อย่างที่สาม เมื่อได้คุณลักษณะมาแล้วจะนำไปจำแนกโดยใช้ SVM อย่างสุดท้ายคือการสร้างกรอบของวัตถุ



ภาพที่ 2: สถาปัตยกรรมการเรียนรู้เชิงลึกแบบ R-CNN

### 3. วิธีการดำเนินการวิจัย

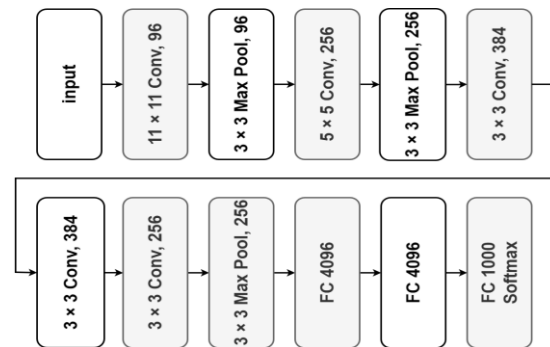
งานที่เสนอดำเนินการโดยใช้ MATLAB R2021b ระบบปฏิบัติการ Windows 10 64 bit, CPU Intel core i5-12600, GPU NVIDIA GeForce RTX 3060 Ram 12 GB.

#### 3.1 สถาปัตยกรรม AlexNet

AlexNet เป็นโครงข่ายประสาทเทียมที่มีความลึก 8 ชั้น เป็นโครงข่ายประสาทเทียมที่ได้รับการฝึกฝนมาแล้วจากฐานข้อมูล ImageNet ที่มีรูปภาพมากกว่า 14 ล้านภาพ จำนวนมากกว่า 1000 ประเภท AlexNet ประกอบด้วย 5 ชั้น Convolutional และ 3 ชั้น Fully Connected [15,16] Input มีขนาด  $227 \times 227$  Pixels Convolutional ชั้นแรกประกอบไปด้วย filter ขนาด  $11 \times 11$  จำนวน 96 filter ชั้นที่สอง filter ขนาด  $5 \times 5$  จำนวน 256 filter ชั้นที่สาม filter ขนาด  $3 \times 3$  จำนวน 384 filter ชั้นที่สี่ filter ขนาด  $3 \times 3$  จำนวน 384 filter ชั้นที่ห้า filter ขนาด  $3 \times 3$  จำนวน 256 filter หลังจากการทำ Convolutional แต่ละครั้งจะทำการ Normalized และ Pooling (Max Pooling) PoolSize  $3 \times 3$  Stride 2,2 โดย Convolutional ทุกชั้นใช้ฟังก์ชัน Rectified Linear Units

(ReLU) เป็น ฟังก์ชัน กระตุ้น (Activation Function) สถาปัตยกรรม AlexNet แสดงดังภาพที่ 3

ในงานวิจัยของ Y. Beeharry และ V. Bassoo ได้เสนอเรื่อง ประสิทธิภาพของ ANN และ AlexNet สำหรับการตรวจจับพืชโดยใช้ภาพแบบ UAV ใช้ข้อมูลในการทดลองทั้งหมด 15,336 ภาพ แบ่งเป็น ดิน 3,249 ภาพ ถั่วเหลือง 7,376 ภาพ หญ้า 1,191 ภาพและวัชพืช 3,520 ภาพ ใช้ฝึกสอนและทดสอบเป็น 70:30 ผลการทดลองพบว่า ANN และ AlexNet มีความถูกต้องที่ 48.09% และ 99.80% ตามลำดับ [16]



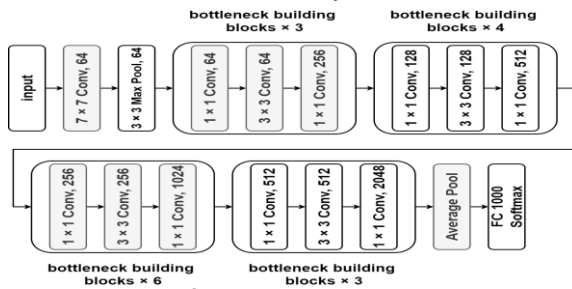
ภาพที่ 3: สถาปัตยกรรม AlexNet

#### 3.2 สถาปัตยกรรม ResNet-50

ResNet-50 เป็นโครงข่ายประสาทเทียมที่มีความลึก 50 ชั้นประกอบไปด้วย ชั้นอินพุตขนาด  $224 \times 224 \times 3$  pixels และมี bottleneck building blocks ทั้งหมด 16 ชั้น convolution ทั้งหมด 49 ชั้นและชั้น fully connected layer [17,18] โดยมีการสร้าง bottleneck building blocks ที่แตกต่างกัน bottleneck building blocks หมายเลข 1 ถึง 3 ประกอบไปด้วยชั้น convolution 3 ชั้น ชั้นที่หนึ่งและชั้นที่สองประกอบไปด้วย filter 64 อัน filter ขนาด  $1 \times 1$  และ  $3 \times 3$  ชั้นที่สามประกอบไปด้วย filter 256 อัน filter ขนาด  $1 \times 1$  หมายเลข 4 ถึง 7 ประกอบไปด้วยชั้น convolution 3 ชั้น ชั้นที่หนึ่งและชั้นที่สองประกอบไปด้วย filter 128 อัน filter ขนาด  $1 \times 1$  และ  $3 \times 3$  ชั้นที่สามประกอบไปด้วย filter 512 อัน filter ขนาด  $1 \times 1$  หมายเลข 8 ถึง 13 ประกอบไปด้วยชั้น convolution 3 ชั้น ชั้นที่หนึ่งและชั้นที่สองประกอบไปด้วย filter 256 อัน filter ขนาด  $1 \times 1$  และ  $3 \times 3$  ชั้นที่สามประกอบไปด้วย filter 1024 อัน filter ขนาด  $1 \times 1$  หมายเลข 14 ถึง 16 ประกอบไปด้วยชั้น convolution 3



ชั้น ชั้นที่หนึ่งและชั้นที่สองประกอบไปด้วย filter 512 อัน filter ขนาด 1×1 และ 3×3 ชั้นที่สามประกอบไปด้วย filter จำนวน 2048 อัน filter ขนาด 1×1 แสดงดังภาพที่ 4 ในงานวิจัยของ I. Z. Mukti and D. Biswas ได้เสนอเรื่อง การถ่ายทอดการเรียนรู้ ResNet-50 เพื่อตรวจหาโรคพืช ใช้ข้อมูลในการฝึกสอนทั้งหมด 87,867 ภาพ ทั้งหมด 38 คลาส ใช้ฝึกสอน 70,295 ภาพ และทดสอบ 17,572 ภาพ ผลการทดลองพบว่ามีความถูกต้องที่ 98.80% [18]

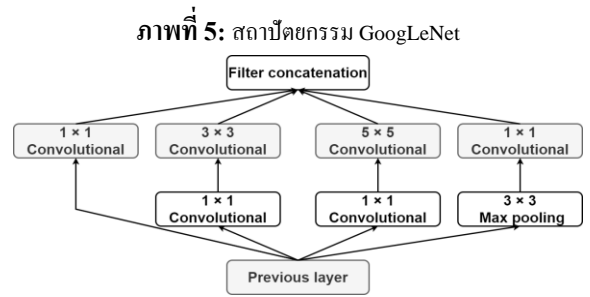
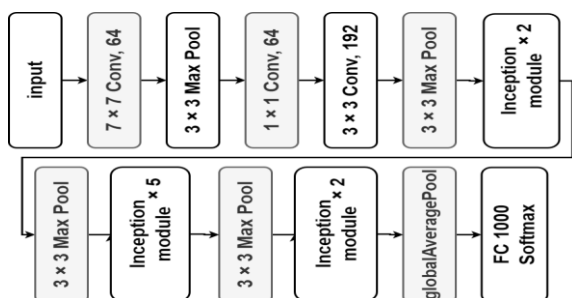


ภาพที่ 4: สถาปัตยกรรม ResNet-50

3.3 สถาปัตยกรรม GoogLeNet

GoogLeNet เป็น โครงสร้างที่มีการทำงานแบบ Inception Module [19] โดยในแต่ละโมดูลจะประกอบไปด้วย Convolutions 1×1, 3×3, 5×5 และ Max Pooling 3×3 [20,21] แสดงดังภาพที่ 6 ซึ่งมีการทำงานแบบขนาด GoogLeNet ออกแบบมาให้มี Inception Module จำนวน 9 ชั้น จากทั้งหมด 22 ชั้น เมื่อข้อมูลถูกส่งเข้า Module ข้อมูลจะถูกแบ่งออกเป็น 4 ชุดและจะรวมข้อมูลเป็นชุดเดียวเมื่อออกจาก Module ทำให้การฝึกสอนใช้เวลาที่สั้นลงแต่ใช้ทรัพยากรในการคำนวณมากขึ้น สถาปัตยกรรม GoogLeNet แสดงดังภาพที่ 5

ในงานวิจัยของ P. Jasitha, et al. ได้เสนอเรื่อง การจำแนกใบพืชตามลายเส้นโดยใช้ GoogLeNet และ VGG ใช้ชุดข้อมูล Dleaf, Flavia และ Leaf1 ผลการทดลองโมเดล GoogLeNet ที่ปรับปรุงแล้วมีความถูกต้องที่ 99.20% [20]



ภาพที่ 6: Inception Module

3.4 ชุดข้อมูลเห็ด

ชุดข้อมูลที่ใช้ในงานวิจัยทั้งหมด 5 ชนิดแบ่งออกเป็น 2 ประเภทคือเห็ดมีพิษและไม่มีพิษจำนวนทั้งหมด 623 ภาพ ขนาด 227×227×3 pixels เห็ดไม่มีพิษ 3 ชนิดได้แก่ เห็ดระโงกขาว (A) เห็ดตะโล (B) และ เห็ดดับเตา (C) เห็ดมีพิษ 2 ชนิดได้แก่ เห็ดหมวกจีน (D) และ เห็ดระโงกหิน (E)

ตารางที่ 1: ชุดข้อมูลเห็ด

เห็ดไม่มีพิษ		เห็ดมีพิษ		
A	B	C	D	E
248	88	155	76	56

การเรียนรู้เชิงลึกนั้นมีความแม่นยำสูงเพราะใช้ชุดข้อมูลการฝึกอบรมจำนวนมาก อย่างไรก็ตามหากชุดข้อมูลที่ใช้ฝึกอบรมมีขนาดเล็ก ก็จะทำให้ความถูกต้องลดน้อยลง ดังนั้นจึงใช้การทำ Data Augmentation เพื่อเพิ่มข้อมูลในการฝึกอบรมและทดสอบ

Data Augmentation คือการเพิ่มปริมาณข้อมูลในการฝึกสอน โดยการนำภาพมา ย่อ/ขยาย หมุนซ้าย/ขวา/บน/ล่าง พลิกภาพในมุมต่างๆ ปรับสีเข้ม/อ่อน เป็นต้น ทำให้ได้ภาพที่มีความแตกต่างกัน

จากชุดข้อมูลเห็ดจำนวน 623 ภาพ เราสร้างชุดข้อมูลที่มี 2,000 ภาพและแบ่งออกเป็น 10 กลุ่มเท่าๆ กัน สิ่งเหล่านี้ถูกใช้ในการฝึกอบรมและการทดสอบตามวิธี K-Folds Cross-Validation

3.5 Confusion Matrix

เป็นเครื่องมือการวัดประสิทธิภาพ Model ว่ามีความแม่นยำมากเพียงใด

$$F1 \text{ Score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3-1}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3-2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3-3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3-4)$$

#### 4. ผลการดำเนินงาน

ชุดข้อมูลเห็นทั้งหมด 2,000 ภาพแบ่งออกเป็น 2 ประเภทคือเห็นมีพิษ 527 ภาพ และเห็นไม่มีพิษ 1,473 ภาพ ในการฝึกอบรมและทดสอบนั้นจะใช้ 10-fold cross validation

##### 4.1 Mushroom Classification Using CNN

พารามิเตอร์ที่ใช้ในการทดลอง CNN ใช้ Learning rate ที่ 0.001 MaxEpochs ที่ 10 รอบ MiniBatchSize ที่ 40 ภาพและใช้การประมวลผลแบบ single GPU

ตารางที่ 2: ผลลัพธ์การทดลอง CNN

Architecture	Accuracy	Time
ResNet-50	99.50%	5.50 นาที
GoogLeNet	99.50%	2.20 นาที
AlexNet	99.00%	1.42 นาที

จากการทดลองสถาปัตยกรรม ResNet-50 มีความแม่นยำมากที่สุด Accuracy 99.50%, Precision 99.33%, Recall 100%, F1 score 99.66% และใช้เวลาในการฝึกสอน 5 นาที 50 วินาที ตามมาด้วยสถาปัตยกรรม GoogLeNet ที่ Accuracy 99.50%, Precision 99.35%, Recall 100%, F1 score 99.67% และใช้เวลาในการฝึกสอน 2 นาที 20 วินาทีและสถาปัตยกรรม AlexNet ที่ Accuracy 99.00%, Precision 98.61%, Recall 100%, F1 score 99.30% และใช้เวลาในการฝึกสอน 1 นาที 42 วินาที

##### 4.2 Mushroom Classification Using R-CNN

พารามิเตอร์ที่ใช้ในการทดลอง CNN ใช้ Learning rate ที่ 0.001 MaxEpochs ที่ 1 รอบ MiniBatchSize ที่ 20 ภาพและใช้การประมวลผลแบบ single GPU

ตารางที่ 3: ผลลัพธ์การทดลอง R-CNN

Architecture	Accuracy	Time
ResNet-50	96.50%	13.44 นาที
GoogLeNet	96.00%	7.28 นาที
AlexNet	95.00%	5.07 นาที

จากการทดลองสถาปัตยกรรม ResNet-50 มีความแม่นยำมากที่สุด Accuracy 96.50%, Precision 100%,

Recall 95.24%, F1 score 97.56% และใช้เวลาในการฝึกสอน 13 นาที 44 วินาที ตามมาด้วยสถาปัตยกรรม GoogLeNet ที่ Accuracy 96.00%, Precision 99.30%, Recall 95.27%, F1 score 97.24% และใช้เวลาในการฝึกสอน 7 นาที 28 วินาทีและสถาปัตยกรรม AlexNet ที่ Accuracy 95.00%, Precision 97.24%, Recall 95.92%, F1 score 96.58% และใช้เวลาในการฝึกสอน 5 นาที 7 วินาที

#### 5. สรุป

งานวิจัยฉบับนี้ผู้วิจัยได้หาค่าความแม่นยำในการจำแนกประเภทเห็นมีพิษและไม่มีพิษโดยใช้วิธี CNN และ R-CNN ของสถาปัตยกรรม 3 แบบ ได้แก่ AlexNet, ResNet-50 และ GoogLeNet

การจำแนกประเภทเห็นโดยใช้ CNN สถาปัตยกรรม ResNet-50 และ GoogLeNet มีค่าความแม่นยำมากที่สุดที่ 99.50% ใช้เวลาในการฝึกสอนที่ 5 นาที 50 วินาที และ 2 นาที 20 วินาทีตามลำดับ

การจำแนกประเภทเห็นโดยใช้ R-CNN สถาปัตยกรรม ResNet-50 ยังคงมีค่าความแม่นยำมากที่สุดที่ 96.50% ใช้เวลาในการฝึกสอนนานที่สุดที่ 13 นาที 44 วินาที สถาปัตยกรรม AlexNet ค่าความแม่นยำที่ 95.00% ใช้เวลาในการฝึกสอนที่ 5 นาที 7 วินาทีเป็นสถาปัตยกรรมที่ใช้เวลาในการฝึกอบรมน้อยที่สุด

#### เอกสารอ้างอิง

- [1] A. Wibowo, et al. "Classification algorithm for edible mushroom identification." *International Conference on Information and Communications Technology*, Yogyakarta, pp. 250–253, 2018.
- [2] A. Mešić et al. "Integrated morphological with molecular identification and bioactive compounds of 23 Croatian wild mushrooms samples." *Food Biosci*, p. 100720, 2020.
- [3] N. Chitayae. and A. Sunyoto. "Performance Comparison of Mushroom Types Classification Using K-Nearest Neighbor Method and Decision Tree Method." *International Conference on*

- Information and Communications Technology*, Yogyakarta, Indonesia, pp. 308–313, 2020.
- [4] S. Khan, et al. “Transfer Learning of a Neural Network Using Deep Learning to Perform Face Recognition.” *International Conference on Electrical, Communication and Computer Engineering*, Swat, Pakistan, pp. 1–5, 2019.
- [5] M. Lin, et al. “A Small Sample Face Recognition Method Based on Deep Learning.” *IEEE 20th International Conference on Communication Technology*, Nanning, China, pp. 1394–1398, 2020.
- [6] Md. A. Rahman, et al. “Improved Segmentation Approach for Plant Disease Detection.” *International Conference on Advances in Science, Engineering and Robotics Technology*, Dhaka, Bangladesh, pp. 1–5, 2019.
- [7] S. V. Militante, et al. “Plant Leaf Detection and Disease Recognition using Deep Learning.” *IEEE Eurasia Conference on IOT, Communication and Engineering*, Yunlin, Taiwan, pp. 579–582, 2019.
- [8] S. M. Alhabshee. and A. U. bin Shamsudin. “Deep Learning Traffic Sign Recognition in Autonomous Vehicle.” *IEEE Student Conference on Research and Development*, Batu Pahat, Malaysia, pp. 438–442, 2020.
- [9] I. A. Tarmizi. and A. A. Aziz. “Vehicle Detection Using Convolutional Neural Network for Autonomous Vehicles.” *International Conference on Intelligent and Advanced System*, Kuala Lumpur, pp. 1–5, 2018.
- [10] C. Lee, S et al. “Performance analysis of local exit for distributed deep neural networks over cloud and edge computing.” *ETRI J.*, no. 5, pp. 658–668, 2020.
- [11] T. D. Sajanraj. and M. Beena. “Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network.” *International Conference on Inventive Communication and Computational Technologies*, Coimbatore, pp. 636–640, 2018.
- [12] J. Dong. and L. Zheng. “Quality Classification of Enoki Mushroom Caps Based on CNN.” *IEEE 4th International Conference on Image. Vision and Computing*, Xiamen, China, pp. 450–454, 2019.
- [13] S. Kido, et al. “Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN).” *International Workshop on Advanced Image Technology*, Chiang Mai, pp. 1–4, 2018.
- [14] H. Yanagisawa, et al. “A study on object detection method from manga images using CNN.” *International Workshop on Advanced Image Technology*, Chiang Mai, pp. 1–4, 2018.
- [15] S. Sun, et al. “Fault Diagnosis of Conventional Circuit Breaker Contact System Based on Time–Frequency Analysis and Improved AlexNet.” *IEEE Trans. Instrum. Meas.*, pp. 1–12, 2021.
- [16] Y. Beeharry. and V. Bassoo. “Performance of ANN and AlexNet for weed detection using UAV-based images.” *International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering*, Balaclava, Mauritius, pp. 163–167, 2020.
- [17] A. P. Rahmathunneesa. and K. V. Ahammed Muneer. “Performance Analysis of Pre-trained Deep Learning Networks for Brain Tumor Categorization.” *International Conference on Advances in Computing and Communication*, Kochi, India, pp. 253–257, 2019.
- [18] I. Z. Mukti. and D. Biswas. “Transfer Learning Based Plant Diseases Detection Using ResNet50.” *International Conference on Electrical Information and Communication Technology*, Khulna, Bangladesh, pp. 1–6, 2019.
- [19] L. Balagourouchetty, et al. “GoogLeNet-Based Ensemble FCNet Classifier for Focal Liver Lesion Diagnosis.” *IEEE J. Biomed. Health Inform.*, no. 6, pp. 1686–1694, 2020.
- [20] P. Jasitha, et al. “Venation Based Plant Leaves Classification Using GoogLeNet and VGG.” *International Conference on Recent Trends on Electronics, Information, Communication & Technology*, Bangalore, India, pp. 715–719, 2019.
- [21] D. Haritha, et al. “Prediction of COVID-19 Cases Using CNN with X-rays.” *International Conference on Computing, Communication and Security*, Patna, India, pp. 1–6, 2020.

# วิเคราะห์ความรู้สึกจากความคิดเห็นของโรคโควิด 19 บนทวิตเตอร์โดยใช้เทคนิคเหมืองข้อความ

## Sentiment Analysis of Twitter Comments About COVID-19 Using Text Mining Techniques

ณัฐรดา คีอามรัมย์ (Natarda deearom)<sup>1</sup> และธนพล เจนสุทธิเวชกุล (Tanapon Jensuttiwetchakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

s6307021858341@email.kmutnb.ac.th, tanapon.j@itd.kmutnb.ac.th

### บทคัดย่อ

ความคิดเห็นของผู้ใช้งานบนทวิตเตอร์ถือว่ามี ความสำคัญ เนื่องจากทำให้ผู้ประกอบการและนักธุรกิจ สามารถนำไปใช้ในการจัดโปรโมชั่น หรือออกแคมเปญ ให้ตรงกับช่วงอารมณ์ของผู้บริโภคได้ งานวิจัยนี้มี วัตถุประสงค์ในการหาประสิทธิภาพของอัลกอริทึมในการ ใช้จำแนกอารมณ์ของผู้ที่แสดงความคิดเห็นของโรคโควิด 19 ด้วยเทคนิคการตัดคำ และการวิเคราะห์ข้อความเพื่อ ค้นหาคำภายในข้อความแสดงความคิดเห็น โดยค้นหาจาก กลุ่มคำที่จะใช้ในการวิเคราะห์ห่างของคำที่อยู่ใกล้เคียงกับ อารมณ์มากที่สุด โดยมีการจำแนกอารมณ์ที่จะใช้ในการ วิเคราะห์ข้อมูลทั้งหมด 7 อารมณ์ ได้แก่ ความทุกข์ (Bad) ความกลัว (Fearful) โกรธ (Angry) รังเกียจ (Disgusted) เสียใจ (Sad) ความสุข (Happiness) และ ประหลาดใจ (surprised)

ผลการทดลองในการจำแนกข้อความ โดยใช้การ ตัดคำและการนับคำคุณลักษณะที่พบอยู่ในประโยคของ ผู้ที่แสดงความคิดเห็นบนทวิตเตอร์ของโรคโควิด 19 นั้น สามารถสรุปได้ว่า อัลกอริทึมมีประสิทธิภาพในการ จำแนก โดยมีค่าความถูกต้องเท่ากับ 97.24% ความแม่นยำ เท่ากับ 96.41% และ ค่าความลึก เท่ากับ 97.24%

คำสำคัญ: โรคโควิด 19 การตัดคำ อารมณ์

### Abstract

The objective of this research is to develop an algorithm for using to classify the emotions of those

who express their opinions about the disease Covid-19 with word cutting technique and text analysis. In the algorithm, searching for sub-sentences and measuring the distance of within the comment text were used in data analytic process. The classification of emotions was used to analyze the data for 7 emotions, namely Suffering, Fearful, Angry, Disgusted, Sad, Happiness and Surprised.

The results indicated that the algorithm is efficient in classifying emotion about COVID-19 with the accuracy at 97.24% , precision at 96.41% , and recall at 97.24%, respectively.

**Keyword:** Covid-19, word cutting, emotions.

### 1. บทนำ

ในปัจจุบันมีการแพร่ระบาดของเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ (Covid-19) เกิดขึ้นเป็นวงกว้างทั่วโลก รวมถึงประเทศไทยด้วย ซึ่งสถานการณ์การแพร่ระบาดนับเป็นวิกฤตที่เกิดขึ้นอย่างกะทันหันไม่สามารถคาดการณ์ล่วงหน้าและเตรียมรับมือกับสถานการณ์เช่นนี้ได้ ทำให้หลายประเทศต้องปิดประเทศลง เพื่อป้องกันการแพร่ระบาดของเชื้อไวรัส เนื่องจากเชื้อกระจายเข้าสู่คนต่อคนได้ง่าย ในส่วนประเทศไทยได้มีการประกาศมาตรการล็อกดาวน์ (พ.ร.ก. ฉุกเฉิน) โดยงดการทำกิจกรรมบางประเภทที่มีความเสี่ยงต่อการแพร่กระจายของเชื้อไวรัส พร้อมทั้ง

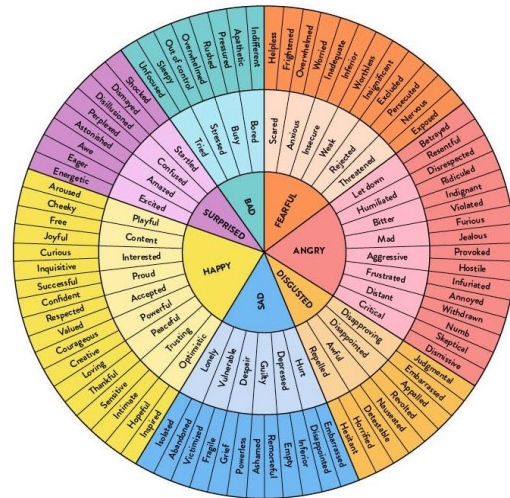
ประกาศห้ามออกนอกเคหสถาน ห้ามเดินทางข้ามจังหวัด ห้ามไม่ให้ชาวต่างชาติและคนไทยที่อยู่ต่างประเทศเดินทางเข้าประเทศไทย

จากสถานการณ์การแพร่ระบาดที่เกิดขึ้นนั้นกระทบกับการดำรงชีวิตต่อประชากรเป็นอย่างมากทั้งต่อเศรษฐกิจ สังคมและสิ่งแวดล้อม ประชากรต้องปรับเปลี่ยนพฤติกรรมในการใช้ชีวิตในหลาย ๆ ด้าน เรียนรู้ในการดูแลตนเอง การเว้นระยะห่างทางสังคม (Social distancing) จึงทำให้ผู้คนเริ่มแสดงความคิดเห็นบนสื่อสังคมออนไลน์ เนื่องจากเป็นช่องทางที่สำคัญในปัจจุบันที่ไว้ใช้สำหรับการติดต่อสื่อสาร และเป็นช่องทางที่สามารถแชร์ประสบการณ์การนึกตา การรักษาตัว การตรวจเชื้อของโรคโควิด 19 หรือวิจารณ์เรื่องราวต่างๆของผู้ใช้งานผ่านโลกอินเทอร์เน็ต โดยผ่านแอปพลิเคชันเช่น twitter Facebook และ Instagram ซึ่งเป็นแอปพลิเคชันที่สามารถเข้าถึงผู้ใช้งานได้ง่ายและรวดเร็ว โดยสามารถนำมาใช้ประโยชน์โดยการรวบรวมข้อมูล สืบค้นข้อมูล นำเสนอความรู้ เพื่อให้ผู้ใช้ที่มีความสนใจในเรื่องเดียวกันหรือคล้ายกัน มาแชร์เรื่องราวในสื่อสังคมออนไลน์ ในปัจจุบันคนนิยมใช้สื่อออนไลน์เป็นจำนวนมาก เนื่องจากมีข้อมูลที่หลากหลาย มีความน่าเชื่อถือ และสามารถแสดงความคิดเห็น หรือวิจารณ์เรื่องต่างๆ ในสังคมไทยที่เกิดขึ้นได้อย่างอิสระ

ทวิตเตอร์ (Twitter) เป็นการบริการเครือข่ายสังคมออนไลน์จำพวกไมโครบล็อก โดยผู้ใช้สามารถส่งข้อความยาวเพื่อบอกเล่าเรื่องราวในชีวิตประจำวันหรือแสดงความคิดเห็น โดยจุดเด่นของทวิตเตอร์คือ ข้อความยาวไม่เกิน 280 ตัวอักษรจึงทำให้ผู้อ่านสามารถเข้าใจความหมายได้ง่ายโดยเรียกการส่งข้อความนี้ว่า ทวิต (Tweet) [1] นับเป็นสังคมออนไลน์ประเภทหนึ่งที่ได้รับคามนิยมเป็นอย่างมาก โดยเฉพาะกลุ่มวัยรุ่นจนถึงวัยทำงาน โดยที่ข้อความเหล่านี้มักจะถ่ายทอดถึงอารมณ์และความรู้สึกของผู้ใช้งานทวิตเตอร์

จากเหตุการณ์ที่กล่าวข้างต้น ผู้วิจัยจึงมีแนวคิดที่จะทำการศึกษาวิจัยในการทำเหมืองข้อความจากข้อความทวิตเพื่อวิเคราะห์ความรู้สึกของผู้ใช้งานที่มีการแสดง

ความคิดเห็นของโรคโควิด 19 โดยนำเทคนิคการทำเหมืองข้อความและวิธีการตัดคำมาใช้ในการจำแนกข้อความ เพื่อวิเคราะห์ความรู้สึกจากข้อความทวิตที่มีการแสดงความคิดเห็นนั้นอยู่ในช่วงอารมณ์ใด ได้แก่ ความทุกข์ ความกลัว โกรธ รังเกียจ เสียใจ ความสุข และประหลาดใจ เพื่อให้ทราบถึงการแสดงอารมณ์ของผู้แสดงความคิดเห็นบนทวิตเตอร์ ณ ช่วงเวลานั้น เพื่อลดความเข้าใจผิดระหว่างผู้อ่านกับผู้แสดงความคิดเห็น และธุรกิจต่างๆสามารถนำไปดู ช่วงอารมณ์ของลูกค้าเพื่อพิจารณาการจัดโปรโมชั่น



ภาพที่ 1: Feeling Circle[8]

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ทวิตเตอร์ (Twitter)

สื่อสังคมออนไลน์จำพวกไมโครบล็อก (Micro Blogging) โดยผู้ใช้สามารถส่งข้อความยาวไม่เกิน 280 ตัวอักษร ว่าตัวเองกำลังทำอะไรอยู่หรือ ธิทวิตข่าวสารที่น่าสนใจ และข้อความที่ส่งถึงกันมีเรียกว่า ทวิต (tweet) เปรียบเหมือนเสียงนกหรือ ข้อความที่ส่งจะต้องเป็น Plain Text เท่านั้น จะไม่สามารถแทรกคำสั่งอะไรได้ ยกเว้นแต่เป็นไฮเปอร์ลิงค์ [1]

### 2.2 Twitter Application programming Interface

การบริการที่ช่วยให้ผู้ใช้งานและองค์กรต่างๆ สามารถเรียกดู Feedback ยอดการเข้าถึงทวิตต่างๆ และสามารถวิเคราะห์ข้อมูลเบื้องต้นของ Follower เช่น เพศ ช่วงอายุ และเวลาในการเข้าถึงของ Twitter เป็นต้น เพื่อ

นำไปใช้ในการปรับปรุงลักษณะของการทวิตเพื่อโปรโมตสินค้า กิจกรรมของศิลปินหรือบริการต่างๆ ขององค์กรได้สะดวกขึ้น โดยปัจจุบัน Twitter API มี Version 2 ที่มีการเพิ่มเครื่องมือต่างๆ เข้ามาเสริมเพื่อช่วยให้ใช้งานได้สะดวกมากขึ้น และเพิ่มสิทธิ์การเข้าถึงข้อมูลให้ครอบคลุม [4]

**2.3 อารมณ์และการจำแนกอารมณ์ (classification of emotions)**

อารมณ์ คือการแสดงความรู้สึกออกมาเมื่อร่างกายถูกกระตุ้นทั้งภายในและภายนอก และอารมณ์มีลักษณะทั้งที่เป็นด้านบวกและด้านลบ ซึ่งอาจจะช่วยส่งเสริมให้ชีวิตมีความสุขและอาจเป็นสิ่งที่บั่นทอนทำลายชีวิตได้เช่นกัน โดยอารมณ์นั้นสามารถแบ่งออกเป็นหลายชนิด แต่ส่วนใหญ่จัดการแบ่งอารมณ์ที่เป็นสากล ซึ่งได้จากการแสดงสีหน้าออกเป็น 6 แบบ ได้แก่ ความสุข ประหลาดใจ ความกลัว ความเศร้า ความโกรธ และน่ารังเกียจ เป็นต้น ยังสามารถแยกย่อยได้อีก รวมทั้งยังมีอารมณ์ที่ไม่สามารถสังเกตเห็นสีหน้าได้อีก [9]

**2.4 การตัดคำ (Word Segmentation)**

ประมวลผลการจำแนกหมวดหมู่ข้อความเพื่อให้มีประสิทธิภาพในการจำแนกมากขึ้น สำหรับการตัดคำในภาษาไทยยังพบปัญหาในการตัดคำเนื่องจากภาษาไทยมีลักษณะการเขียนติดต่อกันแบบไม่มีเว้นวรรค หรือไม่มีช่องว่างคั่นระหว่างคำ ที่แสดงให้เห็นถึงขอบเขตของคำ โดยสามารถแบ่งตามลักษณะฐานข้อมูลที่น่ามาใช้ในการตัดคำมีดังนี้ หลักการตัดคำโดยใช้พจนานุกรม (Dictionary-Based Approach) หลักการตัดคำโดยใช้คลังข้อมูล (Corpus-Based Approach) เป็นต้น [7] ในงานวิจัยนี้จะใช้อัลกอริทึม Dictionary-Based longest matching ซึ่งเป็นการตัดคำโดยการเปรียบเทียบกับคำยาวที่สุดที่มีอยู่ในพจนานุกรม

**2.5 การกำจัดคำหยุด (Stop word removal)**

การนำคำที่ไม่มีนัยสำคัญออกจากเอกสาร โดยไม่ทำให้ความหมายของข้อความเปลี่ยนแปลง ตัวอย่างเช่น คำบุพบทเป็นคำที่ใช้เชื่อมคำหรือกลุ่มคำให้สัมพันธ์กัน คำสันธานเป็นคำที่ทำหน้าเชื่อมต่อกับคำ คำสรรพนามเป็นคำที่ใช้

แทนคำนามที่กล่าวถึงมาแล้วในประโยค เป็นต้น คำที่ใช้กันทั่วไปไม่มีความหมายสำคัญต่อข้อความในเอกสาร เมื่อตัดออกจากเอกสารแล้วไม่ทำให้ใจความของเอกสารเปลี่ยนแปลง ประโยชน์ในการกำจัดคำหยุดจะช่วยให้ขนาดของเอกสารนั้นลดลง ยังลดขนาดพื้นที่ที่อีกด้วยและระยะเวลาในการประมวลผล ซึ่งคำที่เป็นคำหยุดในภาษาไทย [3]

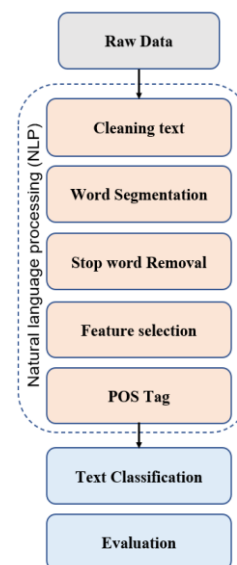
ตารางที่ 1: ตัวอย่างการกำจัดคำหยุด

คำบุพบท	ตั้งแต่ ถึง ที่ ใน เพื่อ เมื่อ
คำสันธาน	ด้วย ได้แก่ แต่ แต่ว่า ถ้า ถึง
คำสรรพนาม	นั่น โน่น นี่ ใคร อะไร ที่ไหน
คำวิเศษณ์	ทั้งนี้ ทั้งนี้ ทำไม ที่ ซึ่ง อัน
คำอุทาน	ชะงัน มิน่า ้วย ว้าว หือ อ้อ

**2.6 การสกัดคุณลักษณะ (Feature Extraction)**

การสกัดคุณลักษณะ คือการดึงคุณลักษณะของเอกสาร กับการลดขนาดของเอกสารลง ซึ่งต้องกำหนดว่าใช้คำใดแทนคุณลักษณะ จากการสำรวจบทความที่ผ่านมา จะใช้คำเดี่ยว แล้วยังสามารถใช้คำวลี หรือกลุ่มของคำ ประโยค แทนคุณลักษณะของเอกสารได้เช่นกัน ตัวแทนคุณลักษณะของเอกสารนิยมใช้ในการจัดหมวดหมู่ประเภทข้อความ คือ ถุงคำ (bag of word) ซึ่งจะเป็นการเก็บอยู่ในรูปแบบของเวกเตอร์ โดยองค์ประกอบของเวกเตอร์อาจจะแทน ด้วยคุณลักษณะของค่าความจริง (Boolean) [4]

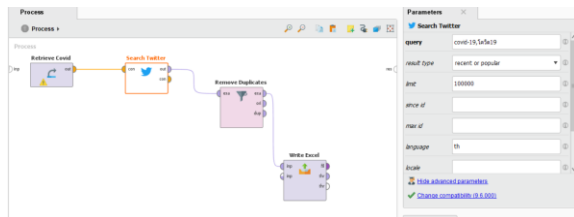
**3. วิธีการดำเนินงาน**



ภาพที่ 2: ขั้นตอนการดำเนินการ

### 3.1 การรวบรวมข้อมูล

ผู้วิจัยได้ทำการรวบรวมข้อมูลการแสดงความคิดเห็นที่มาจากทวิตเตอร์ ผ่าน API โดยใช้โปรแกรม RapidMiner ในการดึงข้อมูลจากทวิตเตอร์ที่มีคำเกี่ยวข้องกับโรคโควิด 19 ในประเทศไทย เช่นคำว่า Covid-19 โควิด 19, โควิด19วันนี้ และ โควิด โดยเรียกใช้ operation search twitter และ Remove Duplicate ข้อความ เพื่อรวบรวมข้อมูลการแสดงความคิดเห็นในรูปแบบไฟล์ Excel โดยรวบรวมข้อมูลตั้งแต่ มกราคม 2563 - มีนาคม 2564 ดังภาพที่ 3



ภาพที่ 3: รวบรวมชุดข้อมูล

### 3.2 การเตรียมข้อมูล (Data Processing)

หลังจากรวบรวมชุดข้อมูลเรียบร้อยแล้ว เนื่องจากชุดข้อมูลที่แสดงความคิดเห็นของโรคโควิด 19 นั้นยังอยู่ในรูปแบบภาษาธรรมชาติ จากนั้นจึงทำการจำแนกอารมณ์ต่างๆ ที่กล่าวถึงในแต่ละข้อความที่แสดงความคิดเห็น โดยแบ่งออกเป็น 7 อารมณ์ ได้แก่ ความทุกข์ ความกลัว โกรธ รังเกียจ เสียใจ ความสุข และประหลาดใจ ซึ่งในหนึ่งข้อความสามารถแสดงอารมณ์ได้มากกว่าหนึ่งอารมณ์ ยกตัวอย่างเช่น Pfizer ชีวิตไม่เคยกลัวและกังวลกับอะไรเท่าการฉีดไฟเซอร์ครั้งนี้เลย กลัวฉีดแล้วตายอ่า นั่งจิตตกมาหลายวันแล้ว ใจอออ จากนั้นก็จะทำความสะอาดข้อความเป็นวิธีการกำจัด Mention Retweet hyperlink และลบสัญลักษณ์ต่างๆที่ไม่มีผลต่อการประมวลผล

### 3.3 การประมวลผลภาษาธรรมชาติ (Text process)

การตัดคำ (Word segmentation) เป็นการแบ่งข้อความออกมาเป็นคำ โดยการใช้เทคนิคการเปรียบเทียบคำยาวที่สุด จากนั้นจะเลือกคำที่ยาวที่สุดในพจนานุกรมมาใช้และทำการเพิ่มคำที่มีผลกับการประมวลผลเข้าในพจนานุกรม ยกตัวอย่างจากข้อความที่แสดงความคิดเห็นของอารมณ์ คือ Pfizer ชีวิตไม่เคยกลัวและกังวลกับอะไรเท่าการฉีดไฟเซอร์ครั้งนี้เลย กลัวฉีดแล้วตายอ่า นั่งจิตตก

มาหลายวันแล้ว ใจอออ จะสามารถตัดคำออกมาเป็น “Pfizer|ชีวิต|ไม่|เคย|กลัว|และ|กังวล|กับ|อะไร|เท่า|การ|ฉีด|ไฟ|เซอร์|ครั้ง|นี้|เลย|กลัว|ฉีด|แล้ว|ตาย|อ่า|นั่ง|จิต|ตก|มา|หลาย|วัน|แล้ว|ใจ|อออ”

จากนั้นจึงทำการกำจัดคำหยุด (Stop word Removal) เป็นการนำคำที่ไม่มีความหมายออกหรือคำที่พบเจอได้บ่อย โดยที่ไม่ทำให้ความหมายเปลี่ยนแปลงและไม่มีผลต่อการนำมาประมวลผล

การสกัดคุณลักษณะ (Feature Extraction) ในงานวิจัยนี้จะเป็นการสกัดคำคุณลักษณะเอาคำคุณลักษณะของอารมณ์ที่มีผลต่อการแสดงความคิดเห็นของโรคโควิด 19 ด้วยกันทั้งหมด 7 คุณลักษณะ และจะทำการสกัดคำที่เป็นคำที่ใช้ในการบ่งบอกคุณลักษณะของอารมณ์ออกมา ดังตารางที่ 2

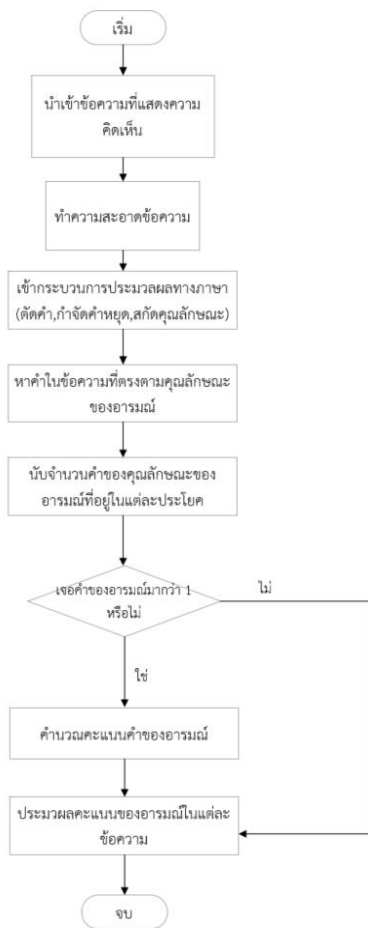
ตารางที่ 2: การสกัดคำคุณลักษณะ

คุณลักษณะ	คำคุณลักษณะ
โกรธ	โมโห โกรธ ใจร้อน หงุดหงิด หัวเสีย ข้องใจ บ้าตาย เจ็บใจ
รังเกียจ	เกลียด ขี้บไล้ หมั่นไส้ ทุเรศ ขยะเขยง ออกติ ชิงชัง ดูถูก
ประหลาดใจ	ตกใจ ตระหนก แปลกใจ โคตรเจ๋ง ตื่น เต็น ตกตะลึง อึ้งจรรย กระตือรือร้น
ทุกข์	เบื่อ เกรียด ยุง ห่วย แยก ล้มเหลว ห่วยแตก คิดมาก
กลัว	กังวล ลังเล วิดก น่ากลัว ใจแป้ว ปอด แหก หัวันใจ กลัว
เสียใจ	เศร้า ร่ำลา อ่อนแอ เศร้าใจ สงสาร เสียใจ น้อยใจ ปวดใจ
ความสุข	ดีใจ โชคดี สบายใจ ชื่นชม ร่าเริง ดี เยี่ยม ชื่นใจ โลงใจ

การกำกับคำ (Pos tagging) เป็นการระบุค่าของแต่ละคำ โดยจะแยกตามองค์ประกอบของประโยค (part of speech tagging) ในงานวิจัยนี้ผู้วิจัยสนใจเลือกใช้เป็นคำนาม คำกริยา คำวิเศษณ์ เป็นหลัก

### 3.1 การจัดกลุ่มเพื่อจำแนกคำในข้อความ

เป็นการจำแนกข้อความ โดยการวิเคราะห์คำที่อยู่ในประโยค ซึ่งในที่นี้เป็นการหาคำในประโยคที่ตรงกับหมวดหมู่ของอารมณ์ โดยจะเลือกคำที่มีคุณลักษณะเดียวกันและอยู่ใกล้กับคำคุณลักษณะที่กำลังพิจารณาอยู่มาใช้ในการระบุคำนั้น อยู่ในช่วงอารมณ์ใด ยกตัวอย่าง จากข้อความที่แสดงความคิดเห็นของอารมณ์ คือ “Pfizer | ชีวิต | ไม่เคยกลัว | และ | กังวล | กับ | อะไร | เท่า | การ |ฉีด |ไฟ |เซอร์ | ครั้ง |นี้ | เลย | กลัว | ฉีด |แล้ว |ตาย |อ่า | นิ่ง |จิตตก |มา |หลาย |วัน |แล้ว |จื่อ |ออ” จะพบว่าประโยคนี้ มีคำคุณลักษณะของอารมณ์ อยู่ 2 คำ คือ ‘กลัว’ และ ‘กังวล’ ดังนั้นจึงสามารถจำแนกได้ว่าความคิดเห็นนี้มีอารมณ์ของกลัว



ภาพที่ 4: ขั้นตอนการดำเนินการ

### 3.4 การประเมินผล

ค่าความถูกต้อง (Accuracy) เป็นการวัดความถูกต้องของแบบจำลองในการจำแนกอารมณ์ โดยพิจารณาจากทุกข้อความ ดังสมการที่ 1

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

ค่าความแม่นยำ (Precision) ความแม่นยำ จะสามารถบอกได้ว่าการจำแนก ดังสมการที่ 2

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

ค่าการค้นคืน (Recall) เป็นการตรวจสอบความถูกต้องของการจำแนกอารมณ์ โดยมีสูตรในการคำนวณดังสมการที่ 3

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

โดย TP คือ จำนวนข้อมูลที่ถูกนำมาใช้อย่างถูกต้อง  
 TN คือ จำนวนข้อมูลที่ผิดและถูกนำมาใช้  
 FP คือ จำนวนข้อมูลที่ถูกต้องแต่ไม่นำมาใช้  
 FN คือ จำนวนข้อมูลที่ผิดแต่ไม่นำมาใช้

### 4. ผลการดำเนินการ

จากการจำแนกอารมณ์ต่างๆ แบ่งออกเป็น 7 อารมณ์ ได้แก่ ความทุกข์ ความกลัว โกรธ รังเกียจ เสียใจ ความสุข และประหลาดใจ โดยใช้ข้อความแสดงความคิดเห็นจากทวีตเตอร์เป็นจำนวน 110 ข้อความ ผ่านกระบวนการวิเคราะห์ข้อความ จากนั้นจะสามารถจำแนกคุณลักษณะของอารมณ์ ในข้อความออกมาถูกต้องจำนวน 107 ข้อความและพบว่า วิเคราะห์ข้อความผิดพลาดจำนวน 3 ข้อความ โดยตัวอย่าง ผลลัพธ์ที่ได้จากการจำแนกข้อความนั้นเป็นไปตาม ดังภาพที่ 5 โดยกำหนดค่า ‘0’ แทนการจำแนกอารมณ์ที่ไม่พบคำที่ แสดงอารมณ์ของคุณลักษณะ และค่าตัวเลข 1 แทนการ แสดงอารมณ์ต่อคุณลักษณะนั้นๆ ดังภาพที่ 5

Review	result						
	โกรธ	กลัว	ทุกข์	รังเกียจ	เสียใจ	ความสุข	ประหลาดใจ
เพิ่งรู้ว่าน้องสาวได้ pfizer เข็มแรก ดีใจกับน้องสาวครับ	0	0	0	0	1	0	0
Pfizer ชีวิตไม่เคยกลัวและกังวลกับอะไรเท่าการฉีดไฟเซอร์ครั้งนี้เลย กลัวฉีดแล้วตายอ่า นิ่ง	0	0	0	1	0	0	0
Pfizer คือเราต้องกรอกจริงๆนะ จะได้ฉีดPfizerจริงๆใช่ไหม กลัวนะ	0	0	1	0	0	0	0
จ๊ว pfizer ได้ตามกุนนะละ แบบดีมากเอพีเห็นไม่แะะจริงๆ ที่มีอาการเวียนหัวเพราะว่า	0	0	0	0	0	1	0
นอนดีเพราะต้องเตรียมตัวสองห้าก่อนลาคะฮือออ	0	0	0	0	0	0	0
ขอไป4ชม.เอง แต่ไม่มีอาการข้างเคียงจริงๆกร๊กล้มได้เลยละ สำหรับเราคือดีมากสามารถ	0	0	0	0	0	0	0
มาทำงานหนักต่อไปได้สบายๆเลยแต่ตั้งที่ฉีดเอง	0	0	1	0	0	0	0

ภาพที่ 5: ตารางแสดงผลการจำแนก

ดังนั้นจึงสามารถคำนวณหาประสิทธิภาพในการจำแนกได้ตามตารางที่ 3



### ตารางที่ 3: ตารางแสดงประสิทธิภาพในการจำแนก

Performance		
Accuracy	Precision	Recall
97.24%	96.41%	97.24%

จากตารางที่ 3 จะเห็นได้ว่างานวิจัยนี้มีประสิทธิภาพในการจำแนกข้อความในแต่ละคุณลักษณะของอารมณ์ได้ ดังนี้ Accuracy มีค่าอยู่ที่ 97.24 % Precision มีค่าอยู่ที่ 96.41% และ Recall อยู่ที่ 97.24%

## 5. สรุป

งานวิจัยนี้ผู้วิจัยได้นำข้อความแสดงความคิดเห็นบนทวีตเตอร์ของผู้แสดงความคิดเห็นของโรคโควิด 19 บนทวีตเตอร์ มาทำการจำแนกอารมณ์ของผู้แสดงความคิดเห็น ที่แสดงอารมณ์ในช่วงนั้นๆ โดยทดลองใช้เทคนิคการตัดคำ และการนับคำในประโยค มาช่วยในการจำแนกข้อความ จาการทดลองพบว่าสามารถจำแนกข้อความ โดยให้ค่าความถูกต้อง (Accuracy) ที่ 97.24% ดังนั้นงานวิจัยนี้สามารถนำมาวิเคราะห์ความคิดเห็นของผู้ใช้งานบนทวีตเตอร์ และนำไปประยุกต์ใช้เป็นเครื่องมือสำหรับการตัดสินใจในการจัดโปรแกรมชั้นหรือเกมเป็ญสินค้าได้ตรงกับความต้องการมากขึ้น

## เอกสารอ้างอิง

- [1] มะลิวัลย์ นิตย์สุวรรณ และวัชรวิพรรณ จิตต์สกุล. (2562). [ออนไลน์]. ระบบเว็บแอปพลิเคชันเหมืองข้อความจากทวีตเตอร์ในการจำแนกความคิดเห็นต่อตราสินค้ากาแฟ. [สืบค้นวันที่ 25 กรกฎาคม 2564]. จาก [https://tdc.thailis.or.th/tdc/dccheck.php?Int\\_code=52&RecId=37932&obj\\_id=206911](https://tdc.thailis.or.th/tdc/dccheck.php?Int_code=52&RecId=37932&obj_id=206911)
- [2] นันทก สุธธิเลิศ. (2560). [ออนไลน์]. การวิเคราะห์ความรู้สึกผู้โดยสารที่ใช้บริการสายการบินของบริษัทในประเทศไทยสหรัฐอเมริกา. [สืบค้นวันที่ 10 สิงหาคม 2564]. จาก [https://tdc.thailis.or.th/tdc/dccheck.php?Int\\_code=95&RecId=28240&obj\\_id=184216&showmenu=no&userid=0](https://tdc.thailis.or.th/tdc/dccheck.php?Int_code=95&RecId=28240&obj_id=184216&showmenu=no&userid=0)
- [3] Nivet Chirawichitcha. (2557). [ออนไลน์]. Emotion classification of Thai text based using term weighting and machine learning techniques. [สืบค้นวันที่ 4 สิงหาคม 2564]. จาก <https://ieeexplore.ieee.org/document/6841848>
- [4] นภัตสร สัจจวิตร. (2562). [ออนไลน์]. การประยุกต์ใช้เทคนิคอัลกอริทึมในการทำเหมืองข้อความเพื่อวิเคราะห์ทัศนคติของผู้ใช้ที่มีต่อการเลือกซื้อโคมไฟสำนักงาน ทวิตเตอร์. [สืบค้นวันที่ 30 กรกฎาคม 2564]. จาก [https://tdc.thailis.or.th/tdc/dccheck.php?Int\\_code=52&RecId=37659&obj\\_id=206636&showmenu=no](https://tdc.thailis.or.th/tdc/dccheck.php?Int_code=52&RecId=37659&obj_id=206636&showmenu=no)
- [5] Siti Sendari and Ilham Ari Elbaith Zaeni (2563). [ออนไลน์]. Opinion Analysis for Emotional Classification on Emoji Tweets using the Naïve Bayes Algorithm. [สืบค้นวันที่ 3 สิงหาคม 2564]. จาก <http://journal2.um.ac.id/index.php/keds/article/view/15449>
- [6] ชวิต ภูมิพนาและณัฏฐ์ คิยเจริญ. (2563). [ออนไลน์]. การวิเคราะห์ความรู้สึกจากข้อความภาษาไทยของข้อคิดเห็นในเฟสบุ๊คด้วยนาอ็อบเบย์. [สืบค้นวันที่ 26 กรกฎาคม 2564]. จาก [https://tdc.thailis.or.th/tdc/dccheck.php?Int\\_code=52&RecId=37664&obj\\_id=206641&showmenu=no&userid=0](https://tdc.thailis.or.th/tdc/dccheck.php?Int_code=52&RecId=37664&obj_id=206641&showmenu=no&userid=0)
- [7] ชนิดา ลิสิริกุล. (2560). [ออนไลน์]. ระบบวิเคราะห์ความคิดเห็นต่อธุรกิจจากการทำเหมืองข้อความบนทวีตเตอร์. คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. [สืบค้นวันที่ 30 กรกฎาคม 2564]. จาก [https://tdc.thailis.or.th/tdc/dccheck.php?Int\\_code=52&RecId=32326&obj\\_id=200777&showmenu=no&userid=0](https://tdc.thailis.or.th/tdc/dccheck.php?Int_code=52&RecId=32326&obj_id=200777&showmenu=no&userid=0)
- [8] Gloria Willcox. (2560) [ออนไลน์]. The Feelings Wheel". [สืบค้นวันที่ 26 สิงหาคม 2564]. จาก <https://blog.calm.com/blog/the-feelings-wheel>
- [9] นพดล สิทธิเลิศ (2556). [ออนไลน์]. การพยากรณ์อารมณ์ความรู้สึกจากการโพสต์ข้อความบนเฟสบุ๊ค. [สืบค้นวันที่ 20 สิงหาคม 2564]. จาก [https://tdc.thailis.or.th/tdc/browse.php?option=show&institute\\_code=52&bib=26963](https://tdc.thailis.or.th/tdc/browse.php?option=show&institute_code=52&bib=26963)

## การจำแนกประเภทปัญหาด้านสารสนเทศจากข้อความ Classification of information problems from text

ธนกร ลีวเวหา (Thanakorn Liwveha)<sup>1</sup> และณัฐวี ฤกษ์ฤกษ์ (Nattavee Utakrit)<sup>2</sup>  
ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล  
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ  
s6307021856071@email.kmutnb.ac.th, nattavee.u@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพอัลกอริทึมในการจำแนกข้อมูลปัญหาด้านสารสนเทศด้วยเทคนิคเหมืองข้อความ ผู้วิจัยแบ่งงานวิจัยออกเป็น 2 ส่วน 1) ส่วนของการเปรียบเทียบประสิทธิภาพความถูกต้องในการจำแนก ของแต่่อัลกอริทึม ได้แก่ Decision Tree, Support Vector Machine และ Naive Bayes โดยข้อมูลที่ได้นำมาใช้งานในวิจัยอยู่รูปแบบของข้อความภาษาไทยเป็นจำนวนประมาณ 17,000 ตัวอย่าง จากนั้นนำข้อมูลตัวอย่างมาผ่านกระบวนการ Text Processing ต่างๆ จนไปถึงขั้นตอนการทดสอบแบบจำลองเพื่อเปรียบเทียบประสิทธิภาพความถูกต้องในการจำแนกข้อมูล พบว่า อัลกอริทึม Support Vector Machine มีประสิทธิภาพความถูกต้องในการจำแนกสูงที่สุด โดยมีค่าเฉลี่ยความถูกต้องอยู่ที่ร้อยละ 89 2) ส่วนของการพัฒนาระบบ โดยนำแบบจำลองที่ได้จากอัลกอริทึม Support Vector Machine ไปประยุกต์ใช้งานในรูปแบบเว็บเพจ เพื่อทดสอบการจำแนกประเภทปัญหาด้านสารสนเทศ และประเมินผลการทำงานจากผู้เชี่ยวชาญด้านสารสนเทศพบว่าประสิทธิภาพความถูกต้องเฉลี่ยอยู่ที่ร้อยละ 81.6

**คำสำคัญ:** การจำแนก ข้อความ เหมืองข้อมูล เหมืองข้อความ

### Abstract

The objective of this research was to an effective comparison of the algorithm for classifying information problems with text mining techniques. The researcher divided the research into 2 parts. 1) The part of the classification accuracy efficiency comparison of each algorithm, including Decision Tree, Support Vector Machine, and Naive Bayes, the

data used in the research are in the form of Thai text for approximately 17,000 samples, then take the sample data through various text processing processes. Until the process of testing the model to compare the performance of classification accuracy, it was found that the Support Vector Machine algorithm had the highest efficiency of classification accuracy. with an average accuracy of 89 percent. 2) System development part The model derived from the Support Vector Machine algorithm is applied in the form of a web page. to test the classification of information problems and evaluating the results of work from information experts found that the average accuracy efficiency was 81.6%.

**Keywords:** Classification, Text, Data Mining, Text Mining

### 1. บทนำ

ในปัจจุบัน ฝ่ายสารสนเทศขององค์กรต่างๆ จะได้รับปัญหาเข้ามาตลอดเวลา โดยองค์กรของผู้วิจัยได้มีคณะทำงานจัดสร้างหมวดหมู่ของปัญหาด้านสารสนเทศเป็นจำนวน 33 ประเภทปัญหาที่เผชิญอยู่คือการคัดแยกหมวดหมู่ของปัญหาที่ยังไม่ตรงกันเหตุมาจากใช้ทรัพยากรมนุษย์ในการจำแนก อาจทำให้เกิดข้อผิดพลาดต่างๆ ซึ่งมีผลต่อการทำงานในลำดับถัดไปของหน่วยงานที่เกี่ยวข้องซึ่งปัญหาในจุดนี้หากผู้จำแนกไม่ใช่มนุษย์แต่เป็นระบบอัตโนมัติก็จะสามารถเพิ่มความถูกต้องในการจำแนก ทั้งนี้การเก็บข้อมูลปัญหาต่างๆขององค์กร จะถูกเก็บอยู่ในรูปแบบของข้อความ ฉะนั้นการนำเทคนิคเหมืองข้อความมาประยุกต์ใช้งาน โดยสร้างแบบจำลองการจำแนกที่มีประสิทธิภาพความถูกต้องสูง จึงเป็นการดำเนินที่สามารถ

แก้ไขปัญหาค่าที่เผชิญอยู่ อีกทั้งยังมีทรัพยากรข้อมูลเพียงพอต่อความต้องการในการดำเนินการ ปัจจุบันมีผู้วิจัยได้นำอัลกอริทึม Decision Tree Support Vector Machine และ Naive Bayes [1-6] ถูกนำมาใช้งานในสร้างแบบจำลองเพื่อจำแนกข้อมูลต่างๆ ที่มีบริบทใกล้เคียงกับประเภทข้อมูลที่ผู้วิจัยได้ทำการศึกษาในครั้งนี้ ดังนั้นผู้วิจัยได้เลือกทำการเปรียบเทียบประสิทธิภาพความถูกต้องอัลกอริทึมทั้ง 3 และเลือกอัลกอริทึมที่ดีที่สุดมาประยุกต์ใช้ในลักษณะการงานในจริงในรูปแบบเว็บเพจ และทดสอบการจำแนกประเภทปัญหาด้านสารสนเทศต่อไป

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 เหมือนข้อความ (Text mining)

การทำเหมืองข้อความ คือการค้นหาคำหรือความรู้รูปแบบ แนวทางหรือความสัมพันธ์ในชุดข้อความนั้น ซึ่งข้อความนั้นไม่มีโครงสร้างของข้อมูล (Unstructured Data) การทำเหมืองข้อความถูกนำไปใช้ในการแก้ปัญหาที่หลากหลาย เช่น การจำแนกเอกสาร (Document Classification) การจัดกลุ่มและจัดระบบเอกสาร (Clustering and Organizing Documents) เป็นต้น เนื่องจากข้อมูลไม่มีโครงสร้าง ดังนั้นในการทำเหมืองข้อความจึงต้องมีการจัดการข้อความให้มีโครงสร้างที่สะดวกต่อการนำไปประมวลผลแบบเหมืองข้อความได้ก่อน [7]

### 2.2 กระบวนการตัดคำ

เป็นกระบวนการตัดแบ่งข้อความ (String) เพื่อให้ได้หน่วยคำที่ถูกตัด (Morpheme) การเขียนภาษาไทยโดยปกติแล้วจะเขียนข้อความติดกัน โดยไม่มีการเว้นวรรคเพื่อแบ่งคำ แต่จะเว้นวรรคเพื่อช่องว่างสำหรับการอ่านและการตีความซึ่งต่างกับภาษาอังกฤษที่จะมีการเว้นวรรคระหว่างคำ ซึ่งทำให้ง่ายต่อการนำภาษาไทยมาประมวลผล [8]

### 2.3 อัลกอริทึม ต้นไม้ตัดสินใจ (Decision Tree)

เป็นอัลกอริทึมการจำแนกประเภท (Classification) ข้อมูลที่มีอยู่ออกเป็นประเภทต่างที่มีการกำหนดไว้ โดยใช้คุณลักษณะ (Attribute) เป็นเงื่อนไขสำหรับการจำแนกประเภท ต้นแบบของอัลกอริทึม Decision Tree ที่ได้จากข้อมูลสำหรับสอน (Training Data) จะทำให้ทราบได้ว่า ใช้

คุณลักษณะใดเป็นเงื่อนไขการจำแนก และความสำคัญของแต่เงื่อนไขมีความสำคัญมากน้อยแตกต่างกันอย่างไร [9]

### 2.4 อัลกอริทึม ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

เป็นอัลกอริทึมสำหรับการใช้จำแนกประเภทหรือการจัดหมวดหมู่ ในลักษณะการเรียนรู้แบบมีผู้สอน (Supervised Learning) การทำงานของอัลกอริทึมนี้ก็คือหาค่าสัมประสิทธิ์ของสมการ เพื่อหาเส้นแบ่งกลุ่มชุดข้อมูลที่ได้นำเข้ามาให้อัลกอริทึมได้เรียนรู้ เพื่อเป็นการหาเส้นแบ่งระหว่างกลุ่มข้อมูลที่เหมาะสมที่สุด อัลกอริทึมจะประมวลผลหาระยะห่างที่มีความกว้างมากที่สุดระหว่างชุดข้อมูล จุดที่ใกล้เส้นแบ่งมากที่สุดของแต่ละชุดข้อมูลจะถูกกำหนดเป็นระยะขอบ ข้อมูลใหม่ที่รับเข้ามาหลังจากที่ผ่านตัวอัลกอริทึมแล้วจะถูกพล็อตและวางลงในกราฟ [8]

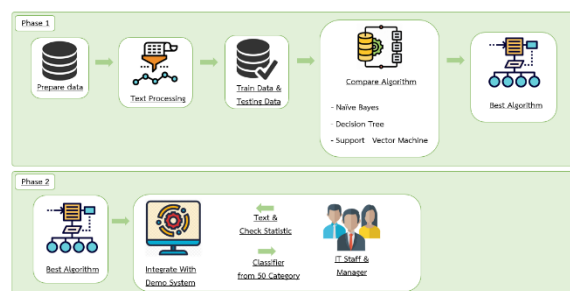
### 2.5 อัลกอริทึม นาอิวเบย์ (Naive Bayes)

มีความคิดพื้นฐานจากทฤษฎีของเบย์ (Bay theorem) เป็นการสร้างตัวจำแนกข้อมูล ใช้สำหรับวิเคราะห์หาความน่าจะเป็นการเกิดของเหตุการณ์ในอนาคตที่ไม่เคยเกิดขึ้นมาก่อน โดยใช้เหตุการณ์ที่เคยเกิดขึ้นมาแล้วในอดีตเป็นฐานในการคาดเดาเหตุการณ์ในอนาคต ตัวแปรแต่ละตัวที่เกี่ยวข้องนั้นมีความเป็นอิสระต่อกัน [9]

## 3. วิธีดำเนินการวิจัย

### 3.1 การออกแบบกรอบแนวคิดและการศึกษาข้อมูลของอัลกอริทึม

จากการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับการจำแนกประเภทของข้อความ ผู้วิจัยได้จัดทำกรอบแนวคิดของงานวิจัยไว้ดังภาพ



ภาพที่ 1: ภาพแสดงกรอบแนวคิดการดำเนินงาน

กรอบแนวคิดในการดำเนินงานวิจัยการจำแนกประเภทปัญหาสารสนเทศจากข้อความ มีการดำเนินงาน 2 ส่วนคือ 1) การเปรียบเทียบและคัดเลือกอัลกอริทึมที่มีประสิทธิภาพความถูกต้องสูงสุด 2) การพัฒนาระบบเพื่อทดสอบอัลกอริทึม และการสรุปผลงานวิจัย

ในส่วนที่ 1 ขั้นตอน Prepare Data เป็นขั้นตอนการส่งออกข้อมูล (Export) ข้อมูลออกจากระบบสารสนเทศ ข้อมูลที่ได้รับมาอยู่ในลักษณะข้อความ ขั้นตอน Text Processing เป็นขั้นตอนนำข้อมูลไปผ่านขั้นตอนการประมวลผลต่างๆ เพื่อลดมิติของข้อมูล เช่น การลบข้อมูลไม่สมบูรณ์ การตัดคำ การหยุดคำ และการลบคำศัพท์เฉพาะที่ไม่มีผลต่อการประมวลผล ขั้นตอน Training Data & Testing Data เป็นขั้นตอนนำข้อมูลที่ผ่านการประมวลผลมาทำแบ่งออกเป็น 2 ชุดคือ Training Data & Testing Data ในอัตราส่วนร้อยละ 75 และ ร้อย 25 โดยขั้นตอนข้างต้นจะดำเนินการผ่านชุดคำสั่งในภาษา Python ทั้งหมด

### 3.2 การเปรียบเทียบและคัดเลือกอัลกอริทึมที่มีประสิทธิภาพความถูกต้องที่สุด

การเปรียบเทียบประสิทธิภาพความถูกต้องของอัลกอริทึม Decision Tree , Support Vector Machine และ Naïve Bayes ทำโดยการกำหนดค่าพารามิเตอร์ที่ต่างออกไปในแต่ละแบบจำลองแล้วนำค่าที่ได้มาเปรียบเทียบกัน การกำหนดค่าพารามิเตอร์แต่ละแบบจำลอง มีดังนี้ อัลกอริทึม Decision Tree กำหนดค่าพารามิเตอร์ Criterion คือ Gini และ entropy อีกหนึ่งพารามิเตอร์คือ Max depth โดยกำหนดในช่วง 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 และ None อัลกอริทึม Support Vector Machine กำหนดค่าพารามิเตอร์ Kernel มี 4 ประเภทคือ Sigmoid, Poly, Rbf, linear และกำหนดค่าพารามิเตอร์ C ในช่วงต่างๆ คือ 0.1 1.0 10.0 100.0 และอัลกอริทึม Naïve Bayes ประเภทที่ใช้คือ Multinomial Naïve Bayes กำหนดพารามิเตอร์คือ Alpha ตั้งแต่ 0.001 ถึง 1 โดยปรับพารามิเตอร์ในช่วงต่างๆ คือ 0.001 0.01 1 10 100

### 3.3 ขั้นตอนการออกแบบระบบ

ในการออกแบบระบบ ผู้วิจัยจะนำอัลกอริทึมที่มีประสิทธิภาพความถูกต้องสูงที่สุด มาประยุกต์ใช้งานร่วมกับเว็บเพจดังภาพ



ภาพที่ 2 : ภาพการดำเนินงานใน Phase ที่ 2

จากภาพที่ 2 การพัฒนาเว็บเพจทดสอบจำแนกข้อมูลสำหรับการบริหารงานหน่วยงานสารสนเทศ สามารถแบ่งการดำเนินการเป็น 2 ส่วน ดังนี้ ส่วนที่ 1 : การออกแบบ User Interface เป็นส่วนที่ใช้ติดต่อกับผู้ใช้งานซึ่งเป็นเจ้าหน้าที่หน่วยงานสารสนเทศและผู้บริหาร ส่วนที่ 2 : การรับส่งข้อมูลระหว่างเว็บเพจและ server จำลอง ในส่วนนี้ผู้วิจัยได้ใช้ Web Framework : Flask เข้ามาตัวเชื่อมการทำงาน

### 3.4 ขั้นตอนการพัฒนาระบบ

การพัฒนาเว็บเพจสำหรับทดสอบใช้งานอัลกอริทึมจำแนกข้อมูล แบ่งการดำเนินการเป็น 2 ส่วนดังนี้ ส่วนที่ 1 : การสร้าง User Interface สำหรับติดต่อกับผู้ใช้งาน เป็นการพัฒนาโดยใช้ภาษา HTML และ CSS ส่วนที่ 2 : การประมวลผลและการติดต่อกับ Server จำลอง ในส่วนนี้เป็นการพัฒนาชุดคำสั่งต่างๆ สำหรับรับข้อมูลจากส่วนผู้ใช้งานส่งไปยัง Server จำลอง โดยนำแบบจำลองที่มีประสิทธิภาพความถูกต้อง (Accuracy) สูงที่สุดในขั้นตอนที่ 3.2 มาพัฒนาร่วมกับภาษา Python, JavaScript โดยโปรแกรมสำหรับการพัฒนาชุดคำสั่ง (Editor) คือ Anaconda และ Visual Studio

### 3.5 การประเมินผลระบบ

ในส่วนนี้ เป็นการทดสอบการใช้งานจริงของ เว็บเพจทดสอบจำแนกข้อมูล โดยการทดสอบการจำแนกประเภทปัญหาสารสนเทศประเภทต่างๆ จากผู้เชี่ยวชาญจำนวน 3 ท่าน โดยผู้เชี่ยวชาญที่ได้เชิญมาทดสอบนั้นมีเกณฑ์การคัดเลือกคือ 1) มีประสบการณ์ทำงานสูง 2) มีผลประเมินสมรรถนะการทำงานอยู่เกณฑ์ดีมาก 3) เป็นส่วนหนึ่งของคณะทำงานด้านสารสนเทศในโครงการต่าง ๆ

#### 4. ผลการดำเนินงาน

##### 4.1 ผลการเปรียบเทียบประสิทธิภาพด้านต่างๆของอัลกอริทึม

จากการทดสอบอัลกอริทึม Decision Tree มีการปรับค่าพารามิเตอร์เพื่อหาค่าความถูกต้องที่สูงที่สุด โดยการปรับพารามิเตอร์ 2 พารามิเตอร์คือ Criterion และ Max\_depth มีจำนวนครั้งในการทดสอบเพื่อให้ครอบคลุมทุกการปรับค่าพารามิเตอร์ทั้งสิ้นจำนวน 16 ครั้ง โดยที่ค่า Criterion เท่ากับ Gini และ Max\_depth เท่ากับ 100 จะให้ประสิทธิภาพความถูกต้องสูงที่สุดตามตารางที่ 1

ตารางที่ 1: ผลดำเนินการทดสอบ Decision Tree

Parameter		Performance		
Criterion	Max_Dept	Accuracy	Preci	ROC
	h		sion	
gini	80	0.85	0.86	0.86
gini	100	<u>0.86</u>	0.86	0.86
entropy	40	0.82	0.83	0.82
entropy	60	0.83	0.83	0.83

จากการทดสอบอัลกอริทึม Support Vector Machine มีการปรับค่าพารามิเตอร์เพื่อหาค่าความถูกต้องที่สูงที่สุด โดยการปรับพารามิเตอร์ 2 พารามิเตอร์คือ Kernel และ C มีจำนวนครั้งในการทดสอบเพื่อให้ครอบคลุมทุกการปรับค่าพารามิเตอร์ทั้งสิ้นจำนวน 16 ครั้ง โดยที่ค่า Kernel เท่ากับ linear และ C เท่ากับ 1.0 ให้ประสิทธิภาพความถูกต้องสูงที่สุดตามตารางที่ 2

ตารางที่ 2: ผลดำเนินการทดสอบ Support Vector Machine

Parameter		Performance		
Kernel	C	Accuracy	Precision	ROC
sigmoid	1.0	0.88	0.88	0.88
sigmoid	10.0	0.85	0.86	0.85
rbf	1.0	0.75	0.81	0.72
rbf	10.0	0.88	0.88	0.88
polynor	1.0	0.81	0.85	0.81
mial				
polynor	10.0	0.82	0.84	0.80
mial				

linear	1.0	<u>0.89</u>	0.89	0.89
linear	10.0	0.87	0.88	0.88

จากการทดสอบอัลกอริทึม Naïve Bayes มีการปรับค่าพารามิเตอร์เพื่อหาค่าความถูกต้องที่สูงที่สุด โดยการปรับพารามิเตอร์ Alpha มีจำนวนครั้งในการทดสอบเพื่อให้ครอบคลุมทุกการปรับค่าพารามิเตอร์ทั้งสิ้นจำนวน 20 ครั้ง โดยที่ค่า Alpha เท่ากับ 0.03 ให้ประสิทธิภาพความถูกต้องสูงที่สุดตามตารางที่ตามตารางที่ 3

ตารางที่ 3: ผลดำเนินการทดสอบ Naïve Bayes

Parameter	Performance		
Alpha	Accuracy	Precision	ROC
0.01	0.84	0.85	0.84
0.02	0.84	0.85	0.84
0.03	<u>0.85</u>	0.85	0.84
0.04	0.84	0.85	0.83
0.05	0.84	0.85	0.83
0.06	0.84	0.85	0.83

ตารางที่ 4: ผลการทดสอบประสิทธิภาพความถูกต้อง (Accuracy) ของแต่ละอัลกอริทึม

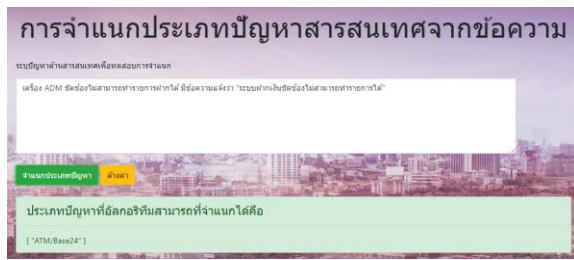
การประเมิน	Accuracy	Precision	ROC
1. Decision Tree	0.86	0.86	0.86
2. Support Vector Machine	<u>0.89</u>	0.89	0.89
3. Naïve Bayes	0.85	0.85	0.85

จากตารางที่ 4 พบว่าผลการวัดประสิทธิภาพความถูกต้อง (Accuracy) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ความค่าความถูกต้อง (Accuracy) สูงที่สุด ผู้วิจัยจึงเลือกนำไปประยุกต์ใช้กับเว็บเพจทดสอบจำแนกประเภทปัญหา ด้านสารสนเทศ

##### 4.2 ตัวอย่างเว็บเพจทดสอบการจำแนกข้อมูล



ภาพที่ 3: ภาพตัวอย่างของเว็บเพจ



ภาพที่ 4 : ทดสอบการจำแนกข้อมูล

ประเภทปัญหา	จำนวนครั้ง
ATM/Bank24	3
ATM/Bank24	2
ATM/Bank24	2
ATM/Bank24	1
ATM/Bank24	1
ATM/Bank24	1
ATM/Bank24	1
ATM/Bank24	1
ATM/Bank24	1
ATM/Bank24	1

ภาพที่ 5 : แสดงสถิติประเภทปัญหาต่างๆย้อนหลัง

#### 4.2 ผลการประเมินประสิทธิภาพของผู้ใช้งาน

ตารางที่ 5: ผลการประเมินประสิทธิภาพความถูกต้องจากผู้เชี่ยวชาญ

ผู้เชี่ยวชาญ	ผลที่ถูกต้อง	จำนวนทั้งหมด	ร้อยละ
ผู้เชี่ยวชาญคนที่ 1	16	20	80
ผู้เชี่ยวชาญคนที่ 2	16	20	80
ผู้เชี่ยวชาญคนที่ 3	17	20	85
ผลลัพธ์ที่จำแนกถูกต้องคิดเป็นค่าเฉลี่ยร้อยละ			81.6

จากตารางที่ 5 พบว่าค่าเฉลี่ยความถูกต้อง ของการจำแนกปัญหาจากอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนจากผู้เชี่ยวชาญมีค่าเฉลี่ยอยู่ที่ร้อยละ 81.6

โดยการทดสอบมีวิธีการคือ ให้ผู้เชี่ยวชาญได้ทดสอบระบบปัญหาต่างๆ ที่พบเจอในองค์กรในลักษณะประโยคลงในเว็บเพจทดสอบการจำแนก ซึ่งทางผู้เชี่ยวชาญจะเป็นผู้เตรียมข้อมูลทดสอบมาด้วยตนเอง และตรวจสอบผลลัพธ์ที่อัลกอริทึมสามารถจำแนกออกมาได้ มีความถูกต้องจากความเป็นจริงหรือไม่ โดยผู้เชี่ยวชาญแต่ละท่านทดสอบเป็นจำนวน 20 ครั้งจากนั้นนำมาคิดความถูกต้องเฉลี่ย

### 5. สรุป อภิปรายผล

จากผลสรุปการวิจัยสามารถอภิปรายผล ได้ดังนี้ อัลกอริทึมที่เหมาะสมกับข้อมูลชุดนี้ซึ่งเป็นลักษณะของข้อความคือ อัลกอริทึม Support Vector Machine ที่กำหนด

ค่าพารามิเตอร์ Kernel เท่ากับ linear และ C เท่ากับ 1.0 จะให้ประสิทธิภาพความถูกต้องสูงสุดคือร้อยละ 89 จากนั้นนำมาสร้างต้นแบบ การจำแนกประเภทปัญหาด้านสารสนเทศจากข้อความ เพื่อช่วยเพิ่มประสิทธิภาพการทำงานของหน่วยงานสารสนเทศ พบว่าสามารถช่วยในการจำแนกประเภทปัญหาด้านสารสนเทศในเบื้องต้นได้ดี และสามารถเรียกดูสถิติย้อนหลังได้อย่างถูกต้อง ซึ่งสอดคล้องกับงานวิจัยต่างๆ ที่ได้มีการนำอัลกอริทึม Support Vector Machine ซึ่งมีประสิทธิภาพความถูกต้องสูงสุด มาสร้างแบบจำลอง และนำไปจำแนกประเภทปัญหาจากข้อความในลักษณะต่างๆ เช่น การจำแนกประเภทการให้ความช่วยเหลือเด็กที่ถูกทารุณ โดยเทคนิคเหมืองข้อความ [10] ระบบวิเคราะห์ความคิดเห็นต่อธุรกิจด้วยการทำเหมืองข้อความบนทวิตเตอร์ [11] และการสกัดปัจจัยที่ส่งผลต่อการผิดพลาดที่ก่อให้เกิดอุบัติเหตุในระบบรถไฟด้วยเหมืองข้อความ [12] ทั้งนี้การจำแนกประเภทปัญหาด้านสารสนเทศได้ถูกต้อง ส่งผลการประสานงานกับหน่วยงานต่างๆ มีความรวดเร็วยิ่งขึ้น และทำให้ประสิทธิภาพการทำงานของหน่วยงานสารสนเทศมีมากขึ้นตามไปด้วย

### 6. ข้อเสนอแนะในการวิจัย

การดำเนินการวิจัยสร้างแบบจำลอง การจำแนกประเภทสารสนเทศจากข้อความ ที่ผ่านมานั้น พบว่า ในด้านปริมาณที่ได้นำมาใช้เป็นข้อมูลในการสอนอัลกอริทึมแบบภาพรวม มีจำนวนที่มากพอและไม่เป็นปัญหา แต่ในส่วนที่พบว่ายังเป็นปัญหาคือ จำนวนตัวอย่างที่แตกต่างกันอย่างมาก เนื่องด้วยปัญหาด้านสารสนเทศที่เกิดในการดำเนินการประจำวัน จะมีบางประเภทที่จะเกิดขึ้นค่อนข้างบ่อยจึงทำให้ มีตัวอย่างให้อัลกอริทึมในปริมาณที่มาก แต่ในบางประเภทจะเกิดขึ้นค่อนข้างน้อย จึงมีตัวอย่างให้อัลกอริทึมน้อยตามลงไปเช่นเดียว เนื่องจากผู้วิจัยต้องการให้เป็นไปตามสถานการณ์จริงขององค์กร จึงไม่สามารถสร้างข้อมูลขึ้นมาเองได้ และมีประเภทปัญหาบางอย่าง มีการกำหนดที่ค่อนข้างกว้างจึงเกิดการทับซ้อนของปัญหาได้ ดังนั้นหากต้องการให้ประสิทธิภาพความถูกต้องของต้นแบบสูงที่สุด ต้องแบ่งประเภทปัญหาให้ชัดเจน ไม่ทับ

ซ้อน รวมไปถึงข้อมูลตัวอย่างที่มีอยู่ในระดับใกล้เคียงกัน จะช่วยให้การสร้างแบบจำลองมีประสิทธิภาพความถูกต้องสูงสุดได้

นอกจากนี้ในอนาคตสามารถนำข้อสรุปต่างๆ ไปต่อยอดในลักษณะของการจำแนกความเร่งด่วนของปัญหาที่เกิดขึ้นได้เช่นกัน เพราะปัญหาที่เกิดขึ้นมีใบงานในลักษณะของข้อความเช่นเดียวกัน แต่กรณีที่เกิดขึ้นมีความเร่งด่วนในการแก้ไขที่ไม่เท่ากันบางระบบต้องได้รับการแก้ไขทันที บางระบบมีห้วงเวลาที่สามารถรอคอยได้

### เอกสารอ้างอิง

- [1] พรรษี จตุรชาติพิชัย. (2558). แบบจำลองการจำแนกไอซีดี-เทน ทีเอ็ม ซ้ำภาษาโดยใช้เหมืองข้อความ. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- [2] ณัฐกิจ เจนการ และมหศักดิ์ เกตุกล้า. (2562). การพัฒนาแบบจำลองในการตรวจจับข้อความภาษาไทยที่เป็นการกลั่นแกล้งทางไซเบอร์โดยใช้วิธีซัพพอร์ตเวกเตอร์แมชชีน. The 15th National Conference and 2019-1 International Conference on Applied Computer Technology and Information Systems and 2019-1 National Conference on Business Administration. 144-149.
- [3] อุดินดี วัฒนบูรานนท์. (2560). การจำแนกคุณภาพการใช้งานและปัญหาของเกมแอ็กชัน จากบทวิจารณ์ของผู้ใช้ด้วยการทำเหมืองข้อความ. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- [4] ปราจีน จิตรเย็น. (2559). การจำแนกหัวข้อการอบรมด้วยเทคนิคเนอโฟเบย์. สารนิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- [5] ศศิมา มณฑาสวรรณ. (2557). การพัฒนาระบบค้นหารหัส ICD-10 สำหรับงานเวชระเบียน. วิทยานิพนธ์ ศิลปศาสตรมหาบัณฑิต สาขาวิชาสารสนเทศเพื่อการศึกษา มหาวิทยาลัยศิลปากร.
- [6] ชัญญานุช แก้วทิพย์. (2560). การเปรียบเทียบประสิทธิภาพอัลกอริทึมสำหรับการจำแนกข้อผิดพลาดของกระบวนการทดสอบซอฟต์แวร์ด้วยเทคนิคการทำเหมืองข้อมูล. ปัญหาพิเศษ วิทยาศาสตร์มหาบัณฑิต ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- [7] อุดินดี วัฒนบูรานนท์. (2560). การจำแนกคุณภาพการใช้งานและปัญหาของเกมแอ็กชัน จากบทวิจารณ์ของผู้ใช้ด้วยการทำเหมืองข้อความ. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิชา วิศวกรรมซอฟต์แวร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- [8] จิรายุ ชัยมีบุญ. (2563). การประยุกต์ใช้ปัญญาประดิษฐ์กับหุ่นยนต์สนทนาเพื่อช่วยในการบริหารงาน ฝ่ายเทคโนโลยีสารสนเทศ ยุค 4.0. การค้นคว้าอิสระ วิทยาศาสตร์มหาบัณฑิต. สาขาวิชาการบริหารสารสนเทศเพื่อการจัดการ. ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- [9] ประเดิม วงศ์กระโซ่. (2560). ระบบจำแนกหมวดหมู่การแจ้งซ่อมบ้านออนไลน์โดยใช้เหมืองข้อความ. สารนิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ ภาควิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- [10] ปราณีปราณี ล้าเลิศพระคุณ. (2561). การจำแนกประเภทการให้ความช่วยเหลือเด็กที่ถูกทารุณโดยเทคนิคเหมืองข้อความ. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมเว็บวิทยาลัยครีเอทีฟดีไซน์ แอนด์ เอ็นเตอร์เทนเมนต์เทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิต
- [11] ชนิตา ลิสิริกุล. (2560). ระบบวิเคราะห์ความคิดเห็นต่อธุรกิจด้วยการทำเหมืองข้อความบนทวิตเตอร์. ปัญหาพิเศษ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาการบริหารสารสนเทศเพื่อการจัดการ ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
- [12] ชันยพร ทรานสุข. (2560). การสกัดปัจจัยที่ส่งผลต่อการผิดพลาดที่ก่อให้เกิดอุบัติเหตุในระบบรถไฟด้วยเหมืองข้อความ. ปัญหาพิเศษ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาการบริหารสารสนเทศเพื่อการจัดการ ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

# การวิเคราะห์ปัจจัยความคิดเห็นและแสดงผลการค้นหาด้วยกราฟความรู้สำหรับ

## ธุรกิจโรงแรมในประเทศไทย

### Information Extraction with Knowledge Graph for Hotel Business in Thailand

ต้นติกร ปุญญกริยากร (Tantikorn Poonyakariyakorn)<sup>1</sup> และธนชาติย์ ฤทธิบำรุง (Thanachart Ritbumroong)<sup>2</sup>

ภาควิชาวิทยาการข้อมูล คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

<sup>1</sup>tantikorn.poo@stu.nida.ac.th, <sup>2</sup>thanachart.rit@nida.ac.th

#### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและวิเคราะห์ปัจจัยความคิดเห็นจากนักท่องเที่ยวและทำการค้นหาและแสดงผลด้วยกราฟความรู้ ซึ่งมีจำนวนข้อความจากความคิดเห็นทั้งสิ้น 50,666 ข้อความจาก 15 โรงแรม แบบจำลอง Latent Dirichlet Allocation (LDA) ถูกนำมาใช้เพื่อการจัดกลุ่มของความคิดเห็น พบว่า แรงจูงใจ ความเป็นมืออาชีพของพนักงาน และมื้ออาหาร เป็นหัวข้อที่ถูกพูดถึงมากที่สุด และได้มีการนำเทคนิค Name Entity Recognition (NER) เพื่อระบุสิ่งที่น่าสนใจในข้อความ พบว่านักท่องเที่ยวมีการแสดงความคิดเห็นเกี่ยวกับสิ่งอำนวยความสะดวก องค์กรประกอบของโรงแรม และมื้ออาหาร มากที่สุดตามลำดับ กราฟความรู้ จะประกอบไปด้วย 58,677 โหนดจากข้อมูลของนักท่องเที่ยวและข้อมูลของแบบจำลอง Latent Dirichlet Allocation (LDA) และ Name Entity Recognition (NER) ทำให้มีความสัมพันธ์ระหว่างโหนดรวมทั้งสิ้น 877,976 ความสัมพันธ์ ผลลัพธ์ของการศึกษานี้คาดว่าจะประโยชน์สำหรับผู้ประกอบการโรงแรมที่ได้รับข้อมูลเพิ่มเติมสามารถสร้างกราฟความรู้เพื่อที่จะเข้าใจนักท่องเที่ยวได้มากยิ่งขึ้น

**คำสำคัญ:** การจัดสรรของดีริเคลแ่ง การสกัดนิพจน์เฉพาะ กราฟความรู้ ค้นหาด้วยกราฟ การฟังเสียงของลูกค้า การสกัดข้อมูล

#### Abstract

This research aims to study and analyze the factors of tourists' opinions and display the results with knowledge graphs. A total of 50,666 comments from 15 hotels were

obtained from TripAdvisor. A Latent Dirichlet Allocation (LDA) model was used to group comments. The results reveal three interesting topics including travel motivation, staff professionalism, and meals. The Name Entity Recognition (NER) technique was also employed to extract entities of interest. It was found that travelers expressed their opinions about hotel facilities, hotel elements, and meals respectively. The results were then represented in the forms of Knowledge Graph consisting of 58,677 nodes from the traveler data and the results of the Latent Dirichlet Allocation (LDA) models and Name Entity Recognition (NER) models, resulting in a total of 877,976 correlations between the nodes. This can be seen as helpful for management to gain insightful information in better understanding travelers.

**Keywords:** Latent Dirichlet Allocation, Name Entity Recognition, Knowledge Graph, Graph Query, Voice Customer, Data Extraction.

#### 1. บทนำ

ปัจจุบันอุตสาหกรรมการท่องเที่ยวได้กลายเป็นปัจจัยสำคัญประการหนึ่งต่อเศรษฐกิจของประเทศไทย และมีส่วนสำคัญในการขยายตัวของผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) ในช่วงการระบาดของ COVID-19 อุตสาหกรรมบริการท่องเที่ยวได้รับผลกระทบอย่างมาก เนื่องจากการปิดประเทศจากมาตรการภาครัฐ ทำให้ธุรกิจโรงแรมที่พึ่งพิงรายได้จากนักท่องเที่ยวได้รับผลกระทบทั้งทางตรงและทางอ้อม นักท่องเที่ยวนับว่าเป็นส่วนสำคัญในการสร้างความเติบโตให้กับอุตสาหกรรม



ท่องเที่ยว การท่องเที่ยวในสถานที่ข่มส่นผลกระทบ โดยตรงต่อธุรกิจหลากหลายประเภทเช่น สายการบิน ร้านอาหาร ร้านขายของชำ และโดยเฉพาะอย่างยิ่ง ธุรกิจที่พักอาศัยอย่างโรงแรม ปัจจัยที่มีอิทธิพลต่อการตัดสินใจเลือกสถานที่พักอาศัยของนักท่องเที่ยวนั้นมีด้วยกันหลายปัจจัยด้วยกันตั้งแต่ปัจจัยที่จับต้องได้และปัจจัยที่จับต้องไม่ได้ เช่น ความต้องการ แรงจูงใจ ทักษะคติ ประสบการณ์ ความพอใจส่วนบุคคล ที่ทำให้นักท่องเที่ยวเดินทางไปยังสถานที่ท่องเที่ยว ส่งผลต่อการตัดสินใจเลือกโรงแรมของนักท่องเที่ยวทั้งสิ้น แต่ทั้งนี้ นักท่องเที่ยวแต่ละคนมีความสนใจหรือมีความต้องการที่แตกต่างกัน

ในปัจจุบันนี้การการตัดสินใจเข้าพักโรงแรมได้รับอิทธิพลจากความคิดเห็นของนักท่องเที่ยวคนอื่นมากขึ้น จากการอ่านบทวิจารณ์ และความคิดเห็น ของนักท่องเที่ยวคนก่อนหน้าที่เคยเข้าพักโดยตรง ทำให้บทวิจารณ์และความคิดเห็นเป็นตัวบอกลถึงคุณภาพ และคุณลักษณะต่างๆของที่พักได้เป็นอย่างดี อย่างความคิดเห็นของนักท่องเที่ยวที่โพสต์ไปยังแพลตฟอร์ม Tripadvisor.com ที่มีข้อมูลการความคิดเห็นเป็นจำนวนมาก ก็สามารถนำมาวิเคราะห์และหาปัจจัยที่นักท่องเที่ยวได้วิจารณ์ และแสดงความคิดเห็นไว้ได้

งานวิจัยนี้มุ่งเน้นการศึกษาเพื่อค้นหาปัจจัยที่ส่งผลกระทบต่อความรู้สึกและประสบการณ์ของนักท่องเที่ยว เพื่อจัดเก็บและแสดงผลในรูปแบบ knowledge graph โดยทางผู้วิจัยมีความต้องการที่จะทำการสรุปประเด็น (Topic Modelling) จาก ข้อมูลรีวิวทั้งหมดที่เป็นภาษาอังกฤษ ร่วมกับการประยุกต์ใช้การสกัดนิพจน์เฉพาะ (Named Entity Recognition) เพื่อค้นหาปัจจัยจากความ คิดเห็นของนักท่องเที่ยวต่อสถานที่พัก โดยผลลัพธ์ที่คาดว่าจะ เป็นประโยชน์สำหรับผู้ประกอบการ โรงแรมที่ ต้องการสร้างความแตกต่างจากความต้องการของ นักท่องเที่ยว และ พัฒนาบริการ แผนการปรับปรุง หรือ กลยุทธ์ทางการตลาดให้สอดคล้องกับกลุ่มนักท่องเที่ยว

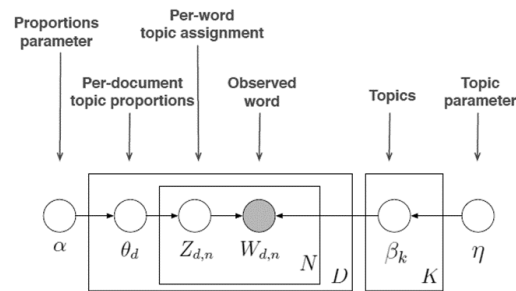
## 2. เทคโนโลยีและงานวิจัยที่เกี่ยวข้อง

แนวคิดและทฤษฎีที่เกี่ยวข้องงานวิจัยนี้เป็นการนำแนวคิดการใช้แบบจำลอง Latent Dirichlet Allocation

ในการวิเคราะห์หัวข้อแฝง Name Entity Recognition ในการสกัดนิพจน์เฉพาะ และ Knowledge graph ในการแสดงผลและจัดเก็บข้อมูล มาผสมผสานกันในการค้นหาปัจจัยจากพฤติกรรมของนักท่องเที่ยว เพื่อก่อให้เกิดความเข้าใจในประสบการณ์ของนักท่องเที่ยวให้มากยิ่งขึ้น

### 2.1 การจัดหมวดหมู่ของข้อความด้วยเทคนิค Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation เป็นเทคนิควิเคราะห์และสร้างแบบแผนจากข้อมูลที่ได้รับ ในการระบุหัวข้อที่แฝงอยู่จากเอกสาร โดยที่เอกสารคือเวกเตอร์ที่มีการเปลี่ยนรูปของประโยค ให้อยู่ในรูปแบบของการนับจำนวนคำ โดยที่ไม่สนใจลำดับของคำที่เกิดขึ้นในรูปประโยค ในระบบการเรียนรู้ของ LDA ใช้รูปแบบคำศัพท์ที่มีการเกิดร่วมกันจากคลังข้อมูล (Corpus) และเอกสารเทอมเมทริกซ์ (Document-term-matrix) เพื่อค้นหาหัวข้อแฝง ทำให้มีความเป็นไปได้ที่จะเกิดคำสองคำขึ้นในเอกสารและมีแนวโน้มว่าคำศัพท์สองคำนั้นจะอยู่ในหัวข้อเดียวกันไปวิเคราะห์ผ่านการแจกแจงความน่าจะเป็นของแต่ละหัวข้อ และความน่าจะเป็นของคำหลายๆคำแล้วแบ่งออกเป็นประเด็นสำคัญตามจำนวนกลุ่มที่กำหนดไว้ [1] โดยกระบวนการสร้างสามารถอธิบายตามภาพที่ 1



ภาพที่ 1: กราฟพิกโมเดลของ Latent Dirichlet Allocation

มีงานวิจัยมากมายที่ศึกษาการประยุกต์ใช้แบบจำลอง Latent Dirichlet Allocation เพื่อค้นหาปัจจัยแฝงที่เกิดขึ้นในข้อความหรือเอกสารไม่ว่าจะเป็นการค้นหาทางด้านชีวสารสนเทศจากการสะสมชุดข้อมูลทางชีววิทยา [2] การประยุกต์ใช้ในประวัติศาสตร์ [3] หรือการจัดประเภทภาพยนตร์อัตโนมัติ [4] นอกจากนี้ Latent Dirichlet Allocation ยังสามารถใช้ค้นหาปัจจัยแฝงที่เกิดขึ้นและทำการจัดประเภทความรู้สึกเพื่อให้เข้าใจความต้องการของ

ผู้ให้บริการมากยิ่งขึ้นเช่น การค้นหาปัจจัยแฝงที่เกิดขึ้นจาก ผู้ใช้บริการบน Airbnb [5] หรือการวัดคุณภาพของสายการบินจากผู้ใช้บริการ [6] ในอุตสาหกรรมโรงแรมอย่างประเทศจีน [7] สิงคโปร์ [8] และ อินโดนีเซีย [9] มีการประยุกต์ใช้ LDA เพื่อวิเคราะห์ปัจจัยแฝงในการวางแผนและมาตรการปรับปรุงจากความต้องการของนักท่องเที่ยวเหมือนกัน

## 2.2 Name Entity Recognition (NER)

Name Entity Recognition (NER) สร้างขึ้น โดยใช้เทคนิคที่เกี่ยวข้องกับไวยากรณ์ภาษาศาสตร์ตลอดจนประยุกต์ด้วยแบบจำลองทางสถิติ ระบบของ Name Entity Recognition เป็นกระบวนการค้นหาในการระบุเอนทิตีขึ้นอยู่กับคุณสมบัติที่หลากหลายของข้อกำหนดประโยคข้อความ และวากยสัมพันธ์ของประโยค [10] ทำให้การรับรู้ชื่อเอนทิตีมีบทบาทสำคัญในแอปพลิเคชันต่างๆ ไม่ว่าจะเป็น การสกัดข้อมูล การตอบคำถาม และการแปลคำผิด [11]

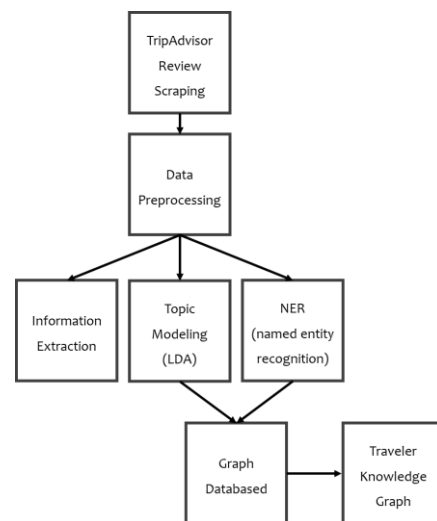
Name Entity Recognition มีทั้งระบบที่เน้นการแยกชื่อโดยใช้กฎเกณฑ์ที่มนุษย์สร้างขึ้นที่ประกอบด้วยชุดของรูปแบบจากไวยากรณ์และวากยสัมพันธ์ หรือระบบที่ใช้การเรียนรู้ของเครื่องจักร ที่มีวัตถุประสงค์เพื่อเปลี่ยนจากการระบุ Name Entity เป็นการจำแนก Name Entity โดยใช้แบบจำลองทางสถิติและอัลกอริทึม จากการมองหารูปแบบและความสัมพันธ์ของข้อความ แล้วทำการระบุและจำแนกคำเป็นคลาสเฉพาะเช่น บุคคล สถานที่ เวลา ฯลฯ โดยอ้างอิงจากข้อมูลที่ระบบได้เรียนรู้ [12]

## 2.3 Knowledge Graph

Knowledge Graph หรือ กราฟความรู้ เป็นกราฟที่แสดงถึงเครือข่ายของข้อมูลในโลกแห่งความเป็นจริง เช่น วัตถุ เหตุการณ์ สถานการณ์ หรือแนวคิด และแสดงความสัมพันธ์ระหว่างข้อมูลเหล่านั้น โดยทั่วไปจะถูกเก็บไว้ในฐานข้อมูลด้วยชุดข้อมูลจากแหล่งต่างๆ ที่มีโครงสร้างแบบสคีมา เพื่อจัดการ โครงสร้างข้อมูลที่หลากหลาย Knowledge Graph ประกอบไปด้วยโหนดพื้นฐานและความสัมพันธ์ ที่ถูกสร้างขึ้นด้วยชุดข้อมูลประจำตัวเพื่อจำแนกข้อมูลในโหนดให้มีความเหมาะสม และความสัมพันธ์จะกำหนดความรู้ที่เกิดขึ้นของโหนดแต่ละโหนด [13]

## 3. วิธีการดำเนินการวิจัย

ข้อมูลทางผู้วิจัยได้ใช้ทดลองในการศึกษาการค้นหาปัจจัยจากความคิดเห็นของนักท่องเที่ยวต่อสถานที่พัก โดยรวบรวมข้อมูลจากเว็บไซต์ TripAdvisor.com โดยเลือกโรงแรมที่อยู่ในอันดับ 30 โรงแรมที่คุ้มค่าที่สุดจากการจัดอันดับของ TripAdvisor มาคัดกรองโดยใช้ค่าเฉลี่ยทางสถิติของจำนวนความคิดเห็น พบว่าสามารถคัดกรองได้ 15 โรงแรมและมีข้อความความคิดเห็นทั้งหมด 50,666 ข้อความ และภาษา Python ได้ถูกนำมาใช้เป็นภาษาโปรแกรมหลักในงานวิจัยชิ้นนี้ โดยมีแนวทางการวิจัยดังภาพที่ 2



ภาพที่ 2: แผนผังแสดงกระบวนการทำวิจัย

### 3.1 การเตรียมข้อมูล

ขั้นตอนการเตรียมข้อมูลเป็นขั้นตอนที่สำคัญสำหรับการประมวลผลภาษาธรรมชาติ (National language Processing) เพื่อที่จะแปลงข้อความจากความคิดเห็นให้อยู่ในรูปแบบที่เข้าใจง่ายขึ้น ผู้วิจัยทำการเตรียมข้อมูลผ่านกระบวนการทั้งหมด 3 ขั้นตอนประกอบไปด้วย การเข้ารหัสและการถอดรหัส การจัดรูปประโยค การตัดคำ และแทนที่คำด้วยรากคำศัพท์

การเข้ารหัสและการถอดรหัสของประโยคเป็นการคัดกรองประโยคในเอกสาร เพื่อให้ได้ประโยคภาษาอังกฤษก่อนที่จะทำการแปลงอักษรเป็นตัวพิมพ์เล็กทั้งประโยคเพื่อลดอคติจากตัวอักษรพิมพ์ใหญ่หรือตัวอักษรพิมพ์เล็ก รวมถึงการลบเครื่องหมายวรรคตอน อักขระพิเศษ และ

ตัวเลขออก ซึ่งเป็นการลบตัวอักษรที่ไม่มีความหมายหรือเกี่ยวข้องกับบริบทของประโยค

ในการสรุปประเด็นสำคัญผู้วิจัยได้ดำเนินการลบคำที่ไม่เกี่ยวข้องกับบริบทของประโยคผ่านการตัดคำ (Tokenization) และการเปลี่ยนคำเป็นรูปคำศัพท์ปกติด้วย Natural Language Toolkit (NLTK) Library ซึ่งเป็นเครื่องมือสำหรับการประมวลผลทางภาษากับภาษาอังกฤษ และทำการระบุคุณลักษณะของกลุ่มคำที่อยู่ในประโยค (Part of Speech Tagging) ด้วยเครื่องมือ Spacy Library เพื่อทำการคัดกรองเฉพาะหมวดหมู่ที่อยู่ใน คำนาม กริยา คำคุณศัพท์ และคำวิเศษณ์ เนื่องจากเป็นกลุ่มคำที่สื่อความหมายจากรูปประโยคมากที่สุด ก่อนที่จะทำการค้นหากรณาคำศัพท์หรือรูปคำศัพท์ปกติที่เกิดขึ้นในประโยค

### 3.2 การค้นหาประเด็นสำคัญจากข้อความ

การค้นหาประเด็นสำคัญจากข้อความจากความคิดเห็นทางผู้วิจัยได้ใช้แบบจำลอง Bigram เพื่อหาคู่คำศัพท์ที่มีการเกิดร่วมกันด้วยมาตราวัดทั้งหมด 5 แบบ ประกอบด้วย ความถี่ที่อยู่ติดกัน (Frequencies of Adjacent Words) ข้อมูลร่วมกันแบบ pointwise (Pointwise Mutual Information) การทดสอบสถิติแบบ t (Student's t Test) การทดสอบไคสแควร์ (Chi-square) และภาวะน่าจะเป็น (Likelihood Ratio) ซึ่งแบบจำลองมีการประยุกต์ร่วมกับโครงสร้างพื้นฐานของคำนามในภาษาอังกฤษ (Noun-Type Structure) เพื่อทำการคัดกรองเฉพาะคำศัพท์ที่มีความหมายและมีกรณเกิดขึ้นในแบบจำลองมากที่สุด

### 3.3 การประมวลผลด้วยเทคนิค Latent Dirichlet Allocation (LDA)

ในการประมวลผลด้วยเทคนิค LDA นั้น ผู้วิจัยได้ทำการใช้ Gensim Library ในการสร้างแบบจำลองขึ้นมา โดยเริ่มจากการสร้างพจนานุกรม (Dictionary) และเอกสารตัวอย่างที่ถูกรวบรวม (Corpus) ที่เป็นข้อมูลสำคัญในการเข้ารหัสคำศัพท์จากเอกสารไปยังดัชนี รวมถึงการใช้เพื่อถอดรหัสดัชนีเป็นคำศัพท์ (id2word) ผ่านการตรวจสอบความสมเหตุสมผลของกฎเกณฑ์ทางภาษาศาสตร์ และการวิเคราะห์ทางสถิติ

ในการปรับปรุงไฮเปอร์พารามิเตอร์ของโมเดล LDA เพื่อให้ได้โมเดลที่มีประสิทธิภาพมากที่สุด ผู้วิจัยได้ทำการ

ปรับพารามิเตอร์ในส่วนจำนวนหัวข้อของแบบจำลอง ในช่วง 2 ถึง 25 หัวข้อโดยที่ไม่มีการเปลี่ยนแปลงไฮเปอร์พารามิเตอร์อื่น เพื่อหาค่าความเหมาะสมของจำนวนปัจจัยแฝงที่เหมาะสมที่สุด โดยใช้คะแนนความปนกันของข้อมูล (Perplexity Score) และคะแนนความสอดคล้องของข้อมูล (Coherence Score) เป็นเกณฑ์ในการประเมินคุณภาพของแบบจำลอง ซึ่งแบบจำลองที่มีคะแนนความสอดคล้องของข้อมูลที่น้อยลงจะเป็นแบบจำลองที่ดีกว่า โดยที่คะแนนความปนกันของข้อมูลที่มีค่าต่ำสุดแสดงถึงโมเดลที่ดีที่สุดจากชุดข้อมูล ในทางตรงกันข้ามคะแนนความสอดคล้องของข้อมูลที่มีค่าสูงสุดถือเป็นแสดงถึงโมเดลที่ดีที่สุดจากชุดข้อมูล

### 3.4 การสร้าง Name Entity Recognition

Name Entity Recognition (NER) ประกอบไปด้วย ไปป์ไลน์หรือรูปแบบที่ใช้ระบุเอนทิตี และ NER โมเดลที่ถูกริเริ่มขึ้น ในการระบุเอนทิตีมีการนำชุดความรู้ของ อาหาร และเครื่องดื่ม [14] ชุดความรู้ของสิ่งอำนวยความสะดวก [15] และชุดความรู้ของคำศัพท์พื้นฐานของนักท่องเที่ยว (Ontology of Vocabulary) [16] มาสร้างเป็นรูปแบบที่สามารถระบุเอนทิตีด้วยโมเดล NER ใน spacy library ดังภาพที่ 3

ภาพที่ 3: การระบุเอนทิตีด้วยโมเดล NER ใน spacy library ด้วยชุดความรู้ของคำศัพท์พื้นฐานของนักท่องเที่ยว

### 3.5 การสร้างกราฟความรู้

ในการแสดงผลลัพธ์ผู้วิจัยได้ทำการสร้างกราฟความรู้ด้วยโปรแกรม Neo4j จากชุดข้อมูลของ TripAdvisor ที่ถูกจัดเก็บตามตารางที่ 1 ชุดข้อมูลของโมเดล LDA ที่ถูกจัดเก็บ ตามตารางที่ 2 และ ชุดข้อมูลของโมเดล NER ที่ถูกจัดเก็บ ตามตารางที่ 3 ซึ่งข้อมูลทั้งหมดอยู่ในรูปแบบไฟล์ .csv

ตารางที่ 1: ตารางแสดงตัวอย่างชุดข้อมูลของ TripAdvisor ที่ถูกจัดเก็บในรูปแบบไฟล์ .csv

CUST_ID	COUNTRY	CITY	TRAVE LTYPE	PLACE
---------	---------	------	----------------	-------

CU00001	England	Oxford	Couple	Anantara Hua Hin Resort
CU00002	France	Paris	-	The Surin Phuket
CU00003	Germany	Munich	Solo	The Okura Prestige Bangkok

ตารางที่ 2: ตารางแสดงตัวอย่างชุดข้อมูลของ LDA ที่ถูกจัดเก็บในรูปแบบไฟล์ .csv

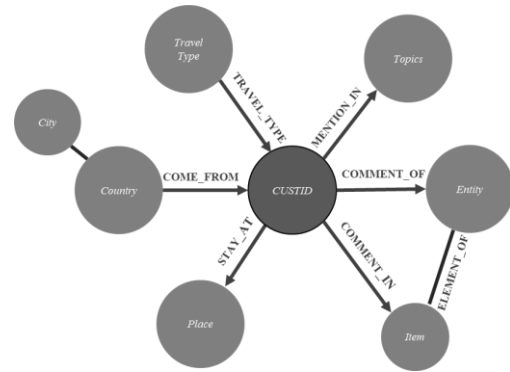
CUST_ID	REVIEW_TEXT	DM_TOPICS
CU00001	'swimming', 'pool', 'look', 'bit', 'small', 'however', 'many', 'people'	Facilities
CU00002	'take', 'hotel', 'shuttle', 'town', 'day', 'planning', 'get', 'taxi', 'take'	Mobility
CU00003	'breakfast', 'great', 'sticky', 'coconut', 'sauce'	Victuals

ตารางที่ 3: ตารางแสดงตัวอย่างชุดข้อมูลของ NER ที่ถูกจัดเก็บในรูปแบบไฟล์ .csv

CUST_ID	ENTITY	ITEM
CU00001	FAC	swimming pool
CU00002	ROAD_TRAVEL	taxi
CU00003	MEAL	breakfast

การสร้าง knowledge graph databased ผู้วิจัยทำการสร้าง databased ผ่าน neo4j ด้วยการระบุชื่อโหนดตามข้อมูลที่จัดเก็บและระบุชื่อความสัมพันธ์ระหว่างโหนดที่เกิดขึ้นได้โหนดทั้งสิ้น 7 ชื่อ ได้แก่ รหัสนักท่องเที่ยว (CUST\_ID) นักท่องเที่ยวจากประเทศ (COUNTRY) โรงแรมที่นักท่องเที่ยวพัก (PLACE) ประเภทของนักท่องเที่ยว (TRAVELTYPE) หัวข้อที่นักท่องเที่ยวพูดถึง จาก โมเดล LDA (DM\_TOPICS) เอนทิตีที่นักท่องเที่ยวพูดถึงจากโมเดล NER (ENTITY) และรายการที่เกิดขึ้นในแต่ละเอนทิตี (ITEM) ประกอบไปด้วยความสัมพันธ์ระหว่างโหนด 7 ความสัมพันธ์ได้แก่นักท่องเที่ยวมาจากประเทศ (COME\_FROM)

นักท่องเที่ยวพักที่โรงแรมไหน (STAY\_AT) นักท่องเที่ยวเป็นนักท่องเที่ยวประเภทไหน (TRAVEL\_TYPE) นักท่องเที่ยวพูดถึงเกี่ยวกับหัวข้อไหน (MENTION\_IN) นักท่องเที่ยวพูดถึงรายการอะไร (COMMENT\_IN) นักท่องเที่ยวพูดถึงเกี่ยวกับเอนทิตีอะไร (COMMENT\_OF) รายการที่เกิดขึ้นเกี่ยวข้องกับเอนทิตีไหน (ELEMENT\_OF) ดังที่เห็นในภาพที่ 4



ภาพที่ 4: ความสัมพันธ์ระหว่างโหนดที่เกิดขึ้น

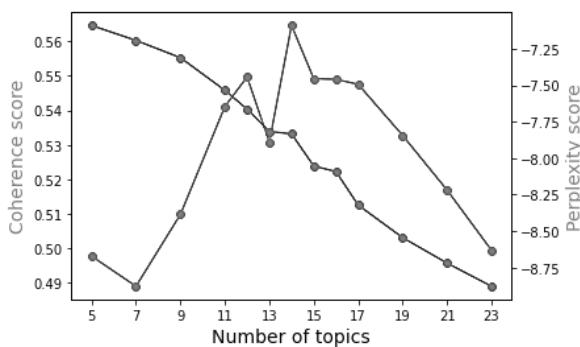
#### 4. ผลการดำเนินการวิจัย

จากการทำ Bigram เพื่อค้นหาประเด็นสำคัญจากข้อความทั้งหมด 5 มาตราวัด พบว่าการแสดงผล จากการคำนวณความถี่ที่อยู่ติดกัน (Frequencies of Adjacent Words) การทดสอบไคสแควร์ (Chi-square) และภาวะน่าจะเป็น (Likelihood Ratio) ให้ผลลัพธ์ Bigram ที่มีความใกล้เคียงกันโดยที่ผลลัพธ์ส่วนใหญ่จะเป็นคำนามทั่วไปเช่นคำว่า "staff, friendly" "swimming, pool" หรือ "sky, train" แต่หากวัดด้วยมาตราวัดข้อมูลร่วมกันแบบ pointwise (Pointwise Mutual Information) และการทดสอบไคสแควร์ (Chi-square) ผลลัพธ์ Bigram ส่วนใหญ่จะเป็นคำนามเฉพาะเช่นคำว่า "flip, flop" "butterfly, pea" หรือ "city, bistro"

ในการระบุจำนวนหัวข้อที่เหมาะสมของแบบจำลอง Latent Dirichlet Allocation (LDA) ผู้วิจัยได้ทำการปรับเปลี่ยนค่าตัวแปรในการสร้างแบบจำลอง โดยใช้วิธีการวนซ้ำ โดยปรับจำนวนหัวข้อตั้งแต่ 2 ถึง 25 หัวข้อ โดยเป็นการเพิ่มขึ้นทีละ 2 หน่วย เพื่อนำคะแนนความปนกันของข้อมูลและคะแนนความสอดคล้องของข้อมูลมา

เปรียบเทียบกันเพื่อหาจำนวนหัวข้อที่เหมาะสมของแบบจำลอง ดังผลลัพธ์ที่แสดงในภาพที่ 5

จากภาพที่ 5 ซึ่งแสดงค่าคะแนนความปนกันของข้อมูลและคะแนนความสอดคล้องของข้อมูลจะเห็นว่า คะแนนความปนกันของข้อมูลยังมีค่าน้อยเมื่อมีจำนวนหัวข้อมากในทางกลับกันคะแนนความสอดคล้องของข้อมูลให้จำนวนหัวข้อที่ชัดเจนกว่า จึงสามารถระบุจำนวนหัวข้อที่เหมาะสมที่ 14 หัวข้อ เนื่องจากเป็นจำนวนหัวข้อที่คะแนนความสอดคล้องของข้อมูลมากที่สุด เมื่อเปรียบเทียบคะแนนความสอดคล้องของข้อมูลระหว่างคู่หัวข้อทั้งหมด



ภาพที่ 5: ความสัมพันธ์ระหว่างจำนวนหัวข้อกับค่าคะแนนความสอดคล้องของข้อมูลและค่าความปนกันของข้อมูลที่พบโดยใช้แนวทางของ Latent Dirichlet Allocation

สามารถสรุปหัวข้อจากแบบจำลอง LDA ด้วยคีย์เวิร์ดที่เกิดขึ้นได้ตามตารางที่ 4 ประกอบไปด้วย ระดับการบริการ (Service Level) แผนกต้อนรับ (Reception) ความเป็นมืออาชีพของพนักงาน (Staff Professionalism) ความพึงพอใจ (Satisfactory) มื้ออาหาร (Meal) สิ่งอำนวยความสะดวก (Facilities) บรรยากาศ (Ambiance) เสบียงอาหาร (Victuals) คุณค่า (Value) การเข้าถึง (Accessibility) การประเมินค่า (Evaluation) องค์ประกอบห้องพัก (Room Condition) แรงจูงใจ (Motivation) ความสะดวกในการเดินทาง (Mobility)

ตารางที่ 4: ตารางแสดง Keywords ในการแยกหัวข้อจากแบบจำลอง Latent Dirichlet Allocation

Topic	Keywords	Related topic
1	well, hotel, high, star, service, keep, class, standard, expect,	Service Level
2	check, day, leave, room, get, arrive, ask, even, problem, book	Reception
3	staff, service, friendly, helpful, always, great, hotel, thank, help,	Staff Professionalism
4	view, amazing, fantastic, floor, great, pool, absolutely, villa,	Satisfactory
5	good, breakfast, food, restaurant, excellent, service, buffet, eat,	Meal
6	pool, area, spa, temper, massage, also, gym, kid, nice, lot	Facilities
7	beach, resort, place, beautiful, perfect, surin, private, set, relax,	Ambiance
8	great, bar, drink, free, price, worth, pay, evening, include,	Value
9	make, feel, special, fruit, guest, fresh, japanese, really, sure,	Victuals
10	location, hotel, walk, shopping, close, locate, station, right, bts,	Accessibility
11	go, back, time, come, stay, hotel, say, year, definitely, visit	Evaluation
12	room, clean, bed, comfortable, cottage, nice, spacious,	Room Condition
13	stay, hotel, night, recommend, experience, phuket, good, trip,	Motivation
14	take, get, hotel, way, boat, go, airport, mall, taxi, convenient	Mobility

ผลลัพธ์การสร้าง Name Entity Recognition ในการระบุเอนทิตีของคำพบว่าสิ่งอำนวยความสะดวก (FACILITY) องค์ประกอบของโรงแรม (HOTEL) มื้ออาหาร (MEAL) ความสนใจในธรรมชาติ (NATURAL ATTRACTION) และอาหารและเครื่องดื่ม (FOOD AND BEVERAGE) เป็น 5 อันดับแรกที่นักท่องเที่ยวมีการพูดถึงตามลำดับที่แสดงในตารางที่ 5

หากจำแนกเป็นรายการในสิ่งอำนวยความสะดวกพบว่านักท่องเที่ยวแสดงความคิดเห็นต่อเกี่ยวข้องกับสิ่งอำนวยความสะดวกพื้นฐานอย่าง ง่ายๆ น้ำคิดเป็น 21.0% ร้านอาหารคิดเป็น 8.5% บุฟเฟ่ต์คิดเป็น 5.8% บาร์คิดเป็น 5.4% และสปาคิดเป็น 5.0% จากการพูดถึงสิ่งอำนวยความสะดวก

**ตารางที่ 5:** ตารางสรุปการกล่าวถึงเอนทิตีจากแบบจำลอง Name Entity Recognition (NER) ใน 10 อันดับแรก

Name Entity Recognition (NER)	% Mentions
FACILITY	35.1%
HOTEL	21.5%
MEAL	8.8%
NATURAL ATTRACTION	6.9%
FOOD AND BEVERAGE	6.5%
ROOM FACILITIES	4.9%
MOTIVATION	2.3%
PERSON	2.3%
AIR TRAVEL	2.0%
WATER TRAVEL	2.0%

หากจำแนกเป็นรายการในองค์ประกอบของโรงแรม ออกมาเป็นปัจจัยจะสามารถเห็นได้ว่าความคิดเห็นของนักท่องเที่ยวที่มีการพูดถึงพนักงานมากที่สุดคิดเป็น 46.8% ตามด้วยห้องพัก 36.7% สิ่งอำนวยความสะดวก 6.3% แผนกต้อนรับ 5.3% และล็อบบี้มี 4.9%

หากจำแนกเป็นรายการในมื้ออาหารพบว่า นักท่องเที่ยวแสดงความเห็นต่อ อาหารเช้ามากที่สุดคิดเป็น 79.0% จากการพูดถึงมื้ออาหารทั้งหมด อาหารกลางวันคิดเป็น 14.3% และอาหารเย็นถูกพูดถึงน้อยที่สุดคิดเป็น 6.7%

หากจำแนกเป็นรายการในเอนทิตีพบว่า การแสดงความเห็นต่อเอนทิตีที่เกี่ยวข้องกับอาหารและเครื่องดื่ม จะแสดงผลลัพธ์ที่แตกต่างกันสำหรับนักท่องเที่ยวแต่ละประเทศดังที่แสดงในภาพที่

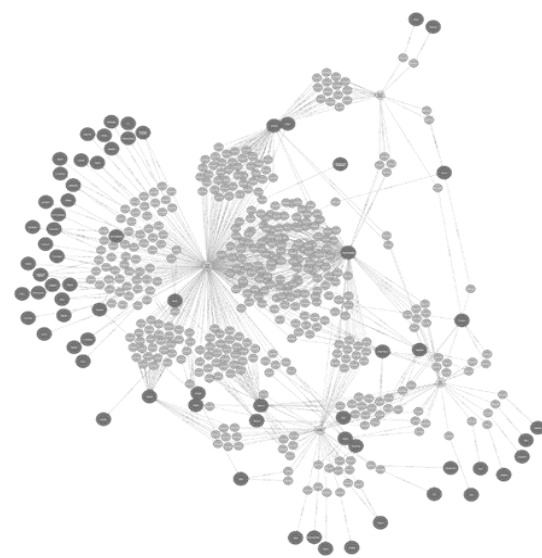
Food Type	United Kingdom	Australia	United States	Thailand	Singapore	Hong Kong	India	Canada	Germany	Switzerland
Beverage	1.0%	1.0%	1.0%	1.1%	1.0%	1.0%	1.0%	0.9%	0.9%	0.9%
Breakfast	4.2%	3.2%	2.0%	3.0%	1.0%	2.0%	1.0%	1.1%	3.2%	2.0%
Breakfast buffet	0.2%	0.2%	0.2%	1.1%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%
Chicken	1.1%	1.1%	1.1%	1.1%	1.0%	1.0%	1.0%	0.9%	0.9%	0.9%
Coffee	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%
Continental	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%
Italian	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Mediterranean	1.1%	1.1%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%
Polish	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Portuguese	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Russian	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Seafood	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Snack	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Spanish	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Swiss	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Thai	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Vegetarian	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Wine	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Other	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Unknown	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

**ภาพที่ 6:** ตารางการกล่าวถึงเอนทิตีอาหารและเครื่องดื่มแยกตามรายการสินค้าและประเทศนักท่องเที่ยว

ในการสร้างกราฟความรู้จากชุดข้อมูลของ TripAdvisor ชุดข้อมูลของ แบบจำลอง LDA และ ชุดข้อมูลของ NER มีจำนวนโหนดรวมทั้งสิ้น 58,677 โหนด และมีจำนวนความสัมพันธ์ระหว่างโหนดรวมทั้งสิ้น 877,976 ความสัมพันธ์

จากผลลัพธ์ในภาพที่ 6 ผู้วิจัยพบว่ารายการอาหารและเครื่องดื่มที่เกี่ยวข้งกับชาเป็นรายการที่นักท่องเที่ยวมีการกล่าวถึงมากที่สุด

ในการค้นหากราฟความรู้ที่เกี่ยวข้องกับชาตามที่แสดงในภาพที่ 7 พบว่าโรงแรม Centara Grand Beach Resort & Villas Hua Hin มีนักท่องเที่ยวพูดถึงชามากที่สุดเป็นอันดับหนึ่ง และ Mandarin Oriental, Bangkok มีนักท่องเที่ยวพูดถึงชามากที่สุดเป็นอันดับสอง โดย Chatrium Hotel Riverside Bangkok The Okura Prestige Bangkok และ Pathumwan Princess Hotel มีนักท่องเที่ยวกล่าวถึงชาเป็นลำดับที่สาม สี่ ห้าตามลำดับ และมีสามจากห้าโรงแรมที่ถูกพูดถึงและมีการเขียนรีวิวเกี่ยวกับ afternoon tea ตามเว็บไซต์ต่างๆ ไม่ว่าจะเป็นภาษาไทยหรือภาษาอังกฤษ โดยประกอบไปด้วยโรงแรมในกรุงเทพ สอง โรงแรมคือ Mandarin Oriental, Bangkok และ The Okura Prestige Bangkok และ ในหัวหินหนึ่ง โรงแรม คือ Centara Grand Beach Resort & Villas Hua Hin



**ภาพที่ 7:** กราฟความรู้ของนักท่องเที่ยวที่กล่าวถึง ชา ในเอนทิตีอาหารและเครื่องดื่ม

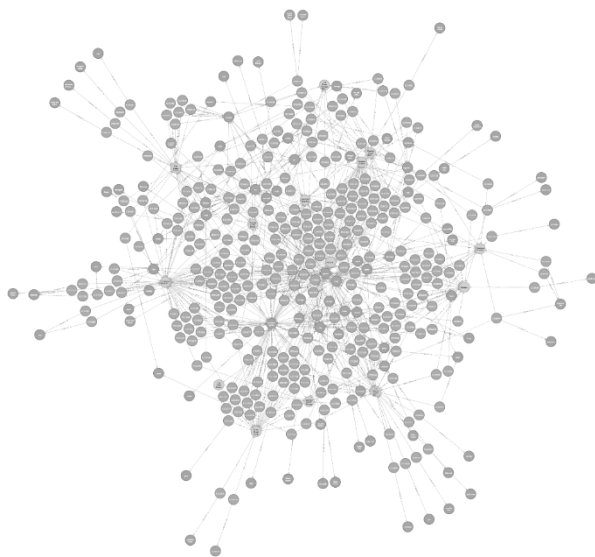
และกราฟความรู้ยังสามารถบอกได้ว่านักท่องเที่ยวที่มีการกล่าวถึงชาในหลายๆ โรงแรมคือ นักท่องเที่ยวผู้รักและนักท่องเที่ยวผู้รักเหล่านั้นยังมีการพูดถึงหัวข้อจากโมเดล LDA ที่เกี่ยวข้องกับ คุณค่า (Value) แรงจูงใจ (Motivation) ความเป็นมืออาชีพของพนักงาน (Staff

Professionalism) การประเมินค่า (Evaluation) และองค์ประกอบห้องพัก (Room Condition) มากที่สุด 5 อันดับแรกตามลำดับที่แสดงในตาราง 6

**ตารางที่ 6:** ตารางสรุปกราฟความรู้ของนักท่องเที่ยวที่รักที่พักในโรงแรมจำแนกตามหัวข้อจากแบบจำลอง Latent Dirichlet Allocation 5 อันดับแรกจาก Neo4j

TRAVELTYPE	TOPIC	TOTAL
COUPLE	Value	612
COUPLE	Motivation	526
COUPLE	Staff Professionalism	495
COUPLE	Evaluation	472
COUPLE	Room Condition	449

นอกจากรายการอาหารที่เกี่ยวข้องกับชาแล้ว ผู้วิจัยพบว่านักท่องเที่ยวอินเดียวมักแสดงความคิดเห็นต่อรายการอาหารสปรดเป็นส่วนใหญ่ในทุกโรงแรมดังที่แสดงในภาพที่ 8 ซึ่งมีความแตกต่างจากนักท่องเที่ยวประเทศอื่นๆ



**ภาพที่ 8:** กราฟความรู้ของนักท่องเที่ยวอินเดียที่กล่าวถึงเมนูอาหารและเครื่องดืมในโรงแรมต่างๆ

## 5. ผลการดำเนินการวิจัย

ความคิดเห็นของนักท่องเที่ยวคือข้อมูลที่เป็นประโยชน์และเป็นข้อมูลเชิงลึก ที่สามารถทำความเข้าใจและสกัดข้อมูลได้อย่างมีประสิทธิภาพ จากการประยุกต์แบบจำลอง Latent Dirichlet Allocation (LDA) และ Name Entity Recognition (NER) ทำให้การวิเคราะห์ข้อความง่ายขึ้นเป็นอย่างมาก และด้วยกราฟความรู้ทำให้

งานวิจัยนี้สามารถที่จะค้นหาปัจจัยและข้อมูลของนักท่องเที่ยวได้ละเอียดมากยิ่งขึ้นและสามารถแสดงผลการค้นหาโดยใช้คำสั่งที่เกี่ยวข้องกับกราฟความรู้แทนการใช้คำสั่งค้นหาแบบปกติ โดยปกติแล้วนั้น ข้อมูลเกี่ยวกับความชอบของนักท่องเที่ยวจะถูกจัดเก็บไว้ในรูปแบบตารางเพื่อให้ผู้ใช้งานไปวิเคราะห์ด้วยตนเอง

งานวิจัยนี้แสดงให้เห็นถึงกระบวนการในการสกัดองค์ความรู้จากข้อมูลโดยใช้ แบบจำลอง Latent Dirichlet Allocation (LDA) และ Name Entity Recognition (NER) ในการสกัดหัวข้อและประเด็นสำคัญจากข้อมูลและนำผลการวิเคราะห์ข้อมูลเบื้องต้น มาจัดเก็บไว้ในฐานข้อมูลกราฟเพื่อให้ผู้ใช้สามารถเรียกดูผลวิเคราะห์ได้อย่างรวดเร็ว ส่งผลให้เกิดความสะดวกและมีประสิทธิภาพในการวิเคราะห์ข้อมูล

ผลของงานวิจัยนี้แสดงให้เห็นแนวทางการนำผลจากการประมวลผลข้อมูล ไปใช้ในการนำเสนอปัจจัยที่ส่งผลกระทบต่อนักท่องเที่ยวเพื่อเพิ่มเอกลักษณ์เฉพาะสำหรับโรงแรมแต่ละแห่ง และความเข้าใจในนักท่องเที่ยวได้มากยิ่งขึ้นซึ่งเป็นประโยชน์ต่อผู้ให้บริการโรงแรมและอุตสาหกรรมท่องเที่ยวในการประยุกต์ใช้ข้อมูล

อย่างไรก็ตามการใช้ Name Entity Recognition (NER) ยังมีข้อจำกัดในส่วนของการทำงาน Rule-based ที่ต้องใช้ทรัพยากรอย่างมากการสร้างรูปแบบไปป์ไลน์ของแต่ละเอนทิตี รวมถึง LDA เทคนิคที่ใช้ยังเป็นแนวทางการค้นหาประเด็นในรูปแบบ unsupervised learning ซึ่งอาจจะไม่ครอบคลุมทุกประเด็นได้

งานวิจัยนี้ได้ระบุนหุมวดหมู่ของเอนทิตีผ่านการใช้การใช้คำศัพท์พื้นฐานของนักท่องเที่ยวเท่านั้น แต่สามารถเพิ่มแต่สามารถเพิ่มความหลากหลายให้มีหลายขั้นมากขึ้นได้เช่น การเพิ่มเอนทิตีของส่วนพักเป็น ห้องน้ำ ห้องนอน หรือระเบียบเป็นต้น หรือการจำแนกลักษณะอาหารและเครื่องดื่มเป็น ชนิดของอาหาร และชนิดของเครื่องดื่มเป็นต้น หรือนำไปประยุกต์ใช้ในการพัฒนาระบบแนะนำสินค้า (Recommendation System) ด้วยข้อมูลที่ที่เกิดขึ้นภายใต้ Name Entity Recognition

ในงานวิจัยนี้หวังผลเพื่อให้โรงแรมมีความเข้าใจถึงปัจจัยต่างๆ ที่นักท่องเที่ยวให้ความสนใจ เพื่อที่จะเพิ่มเอกลักษณ์และบริการของโรงแรมให้ดียิ่งขึ้นได้ ด้วยการใช้อรรถศาสตร์และความชื่นชอบในการท่องเที่ยวที่นักท่องเที่ยวแสดงความคิดเห็นไว้จากการค้นหาและแสดงผลด้วยกราฟความรู้ และยังรวมไปถึงผู้ดูแลอุตสาหกรรมการท่องเที่ยวที่จะสามารถนำกระบวนการดำเนินงานหรือผลลัพธ์ของงานวิจัยนี้ไปเป็นแนวทางเพื่อเริ่มต้นการพัฒนาแนวทางการส่งเสริมภาคการท่องเที่ยวให้มากยิ่งขึ้น

### เอกสารอ้างอิง

- [1] David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent Dirichlet Allocation", Journal of Machine Learning Research 3, 2003
- [2] Liu, L., Tang, L., Dong, W. et al. "An overview of topic modeling and its current applications in bioinformatics". SpringerPlus 5, 1608 (2016).
- [3] Yang, T.I., Torget, A.J. and Mihalcea, R, "Topic modeling on historical newspapers", Proceeding of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, pp.96–104, 2011
- [4] Chao, B. and Sirmorya, A., "Automated movie genre classification with LDA-based topic modeling", International Journal of Computer Applications, Vol. 145, No. 13, pp.1–5, 2016
- [5] Sutherland, Ian. and Kiatkawsin, K., "determinants of guest experiences in Airbnb: a topic modeling approach using LDA", Sustainability, Vol. 12, No. 8, pp.1–16, 2020
- [6] Korfiatis, N., Kourouthanassis, P.E., Stamolampros, P. and Sagiadinos, V., "Measuring service quality from unstructured data: a topic modelling application on airline passengers online reviews", Expert Systems with Applications, Forthcoming, 2018
- [7] Tian, X., He, W., Tao. R. and Akula, "Mining online hotel reviews: a case study from hotels in China", 22nd Americas Conference on Information Systems, pp.2851–2858, 2016
- [8] Hargreaves, C.A., "Analysis of hotel guest satisfaction ratings and reviews: an application in Singapore", American Journal of Marketing Research, Vol. 1, No. 4, pp.208–214, 2015
- [9] Irawan, H., Akmalia, G. and Masrury, R.A. (2019a) "Mining tourist's perception toward Indonesia tourism destination using sentiment analysis and topic modeling", Proceeding of the 2019 4th International Conference on Cloud Computing and Internet of Things, pp.7–12, 2019
- [10] I. Budi and S. Bressan, "Association rules mining for name entity recognition", Proceedings of the Fourth International Conference on Web Information Systems Engineering, 2003. pp. 325-328, 2003
- [11] Y.C. Wu, T.K. Fan, Y.S. Lee, S.J Yen, "Extracting Named Entities Using Support Vector Machines", Springer-Verlag, Berlin Heidelberg, 2006.
- [12] Alireza Mansouri, Lilly Suriani Affendey, Ali Mamat, "Named Entity Recognition Approaches", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.2, 2008
- [13] IBM, Learn about knowledge graphs, networks of semantic metadata which represent a collection of related entities [ออนไลน์]. สืบค้นจาก <https://www.ibm.com/cloud/learn/knowledge-graph> (วันที่สืบค้น 23 ธันวาคม 2564).
- [14] TMIC, FOODB [ออนไลน์]. สืบค้นจาก <https://foodb.ca/downloads> (วันที่สืบค้น 12 พฤศจิกายน 2564).
- [15] Chantana Chantrapornchai, Aphisit Tunsakul, "Information Extraction on Tourism Domain using SpaCy and BERT", ECTI Transactions on Computer and Information Technology, Vol.15, No.1, 2021.
- [16] Patrícia Tosqui-Lucks, Bento Carlos Dias Da Silva, "Structuring an Ontology of the Basic Vocabulary of Tourism", International Journal of Information and Education Technology, Vol. 2, No. 4, 2012.



# การจำแนกความคิดเห็นเกี่ยวกับรถยนต์ไฟฟ้าด้วยเทคนิคเหมืองข้อความ

## Sentiment Classification of Electric Vehicle using Text Mining

ดำรงพล สุวรรณชาติ (Damrongpol Suwanchatree)<sup>1</sup> และวัชรวิธรรม จิตต์สกุล (Watchareewan Jitsakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021856021@email.kmutnb.ac.th, <sup>2</sup>watchareewan.j@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อจำแนกความคิดเห็นเกี่ยวกับรถยนต์ไฟฟ้าบนสื่อสังคมออนไลน์ด้วยเทคนิคเหมืองข้อความ โดยเลือกจำนวน 4 อัลกอริทึม ได้แก่ อัลกอริทึมหน่วยความจำระยะสั้นแบบยาว อัลกอริทึมคอมพลิเมนต์แนฟเบย์ อัลกอริทึมเพื่อนบ้านใกล้ที่สุด และอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ข้อมูลที่ใช้ข้อความที่เป็นภาษาไทยและภาษาอังกฤษ จำนวน 1,000 ข้อความ นำข้อมูลมาเตรียมความพร้อม โดยผ่านกระบวนการทำความสะอาดข้อมูล กระบวนการตัดคำ การกำจัดคำหยุด และการหารากศัพท์ จากนั้นแปลงเป็นรูปแบบของเวกเตอร์ แบ่งข้อมูลเป็น 2 ชุด คือ ข้อมูลสำหรับสร้างและทดสอบแบบจำลอง ทำการเปรียบเทียบประสิทธิภาพด้วยค่าความถูกต้อง ค่าความแม่นยำ ค่าระลอก ค่าการถ่วงดุล และค่าการวิเคราะห์เส้นโค้งคุณลักษณะการทำงาน พบว่าอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพสูงที่สุด โดยมีค่าความถูกต้องเท่ากับร้อยละ 87.00 ค่าความแม่นยำเท่ากับ 0.89 ค่าระลอกเท่ากับ 0.84 ค่าการถ่วงดุลเท่ากับ 0.87 และค่าการวิเคราะห์เส้นโค้งคุณลักษณะการทำงานเท่ากับ 0.87

**คำสำคัญ:** อัลกอริทึมหน่วยความจำระยะสั้นแบบยาว อัลกอริทึมคอมพลิเมนต์แนฟเบย์ อัลกอริทึมเพื่อนบ้านใกล้ที่สุด อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน

### Abstract

This research aims to classify the opinions of electric cars on social media using text mining techniques. Four algorithms were selected: Long Short Term Memory, Complement Naive Bay, K-Nearest Neighbor, and Support Vector Machine. In this work, sentiment datasets in Thai and English language 1,000 samples. Then prepare data before modeling, cleansing data, tokenize, transform case, stop word, stemmer, transform to vector and split data to training and testing to measure the text classification efficiency. The results of the classification of electric vehicle comments on social media by concerning the accuracy, precision, recall, f-measure, and receiver operating characteristic (roc), Support Vector Machine presented the greatest performance accuracy was 87.50%, precision 0.90, recall 0.84, f-measure 0.87 and roc 0.87

**Keywords:** Long Short Term Memory, Complement Naive Bay, K-Nearest Neighbor, Support Vector Machine.

### 1. บทนำ

ปัจจุบันปัญหาภาวะโลกร้อนและปัญหามลพิษ เกิดจากการที่โลกมีปริมาณก๊าซคาร์บอนไดออกไซด์เพิ่มขึ้น ซึ่งส่วนหนึ่งเกิดจากความแออัดของการจราจรบนท้องถนน ส่งผลให้เกิดมลภาวะทางถนนขึ้น สาเหตุมาจากการที่ยานพาหนะต่าง ๆ ใช้พลังงานการเผาไหม้ของน้ำมันเพื่อการขับเคลื่อนบนท้องถนน สำหรับ

กรุงเทพมหานครซึ่งเป็นเมืองหลวงของประเทศไทย มีจำนวนยานพาหนะเพิ่มมากขึ้นทุกปี โดยมีสาเหตุอันเนื่องมาจากค่านิยมบางอย่าง รวมทั้งการคมนาคมซึ่งเป็นปัจจัยอันสำคัญของชีวิตมนุษย์กรุงเทพ ดังนั้นการเข้ามาแทนที่ของรถยนต์ไฟฟ้าจึงเป็นตัวแปรที่จะช่วยลดมลภาวะดังกล่าวได้ ดังนั้นการตัดสินใจมาเลือกซื้อรถยนต์ไฟฟ้าก็เป็นส่วนสำคัญสำหรับผู้บริโภคเช่นกัน

ในการตัดสินใจเลือกซื้อรถยนต์ไฟฟ้าผู้บริโภคจะทำการศึกษาข้อมูลผ่านความคิดเห็นบนสื่อสังคมออนไลน์ เช่น Facebook Twitter หรือ Web Board เป็นต้น เพื่อประกอบการตัดสินใจ ซึ่งความคิดเห็นเหล่านี้มีความสำคัญมากกับทั้งเจ้าของสินค้า และผู้บริโภค โดยความคิดเห็นนั้นมีทั้งความคิดเห็นเชิงบวก ความคิดเห็นเชิงลบ ในการดึงความคิดเห็น จำแนกความคิดเห็น หรือวิเคราะห์ความคิดเห็นเหล่านี้ทำให้เสียเวลา เนื่องจากข้อมูลมีการเพิ่มอยู่ตลอดเวลา และมีความซับซ้อน ได้มีผู้วิจัยนำเสนออัลกอริทึมเหมือนข้อความมาช่วยในการวิเคราะห์ข้อมูลความคิดเห็นในเรื่องต่าง ๆ เช่น อัลกอริทึม Decision Tree อัลกอริทึม Support Vector Machine อัลกอริทึม K-Nearest Neighbor ในการวิเคราะห์ความรู้สึกต่อโรงแรมในประเทศไทย [1] อัลกอริทึม Conjunctive Rule อัลกอริทึม Random Forest อัลกอริทึม Bayesian Logistic Regression และอัลกอริทึม Support Vector Machine ในจำแนกความคิดเห็นเกี่ยวกับภาพยนตร์ ร้านอาหาร และสินค้า [2] อัลกอริทึม Random Forest อัลกอริทึม Neural Network และอัลกอริทึม Naïve Bayes สำหรับการจำแนกความรู้สึกจากข้อความแสดงความคิดเห็นของผู้โดยสารที่ใช้บริการสายการบินของบริษัทในประเทศสหรัฐอเมริกา [3] อัลกอริทึม Naïve Bayes และอัลกอริทึม k-Nearest Neighbor สำหรับการวิเคราะห์ความคิดเห็นเกี่ยวกับภาพยนตร์และโรงแรม [4] หรืออัลกอริทึม Long Short-Term Memory: LSTM ในการเปรียบเทียบความคิดเห็นเกี่ยวกับภาพยนตร์อิหร่าน [5] เป็นต้น

จากปัญหา และงานวิจัยต่าง ๆ ข้างต้นผู้วิจัยจึงมีแนวคิดในการนำเหมืองข้อความมาช่วยในการประมวลผลข้อมูลเกี่ยวกับความคิดเห็นของรถยนต์ไฟฟ้าบนสื่อสังคมออนไลน์ เพื่อลดระยะเวลาในการวิเคราะห์ข้อมูลเกี่ยวกับรถยนต์ไฟฟ้าสำหรับผู้บริโภคใช้ในการประกอบการตัดสินใจซื้อรถยนต์ไฟฟ้า และเป็นข้อมูลสำหรับบริษัทผู้ผลิตรถยนต์ไฟฟ้าเพื่อใช้ในการปรับปรุงพัฒนารถยนต์ไฟฟ้าต่อไป

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้รวบรวมข้อมูลทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยแบ่งเป็น 2 หัวข้อดังนี้

### 2.1 อัลกอริทึมสำหรับบริการจำแนกข้อความ

#### 2.1.1 อัลกอริทึมหน่วยความจำระยะสั้นแบบยาว (Long Short Term Memory: LSTM)

เป็นอัลกอริทึมที่ถูกพัฒนาต่อออกมาจาก RNN โดยทำการแก้ปัญหาในส่วนของ Gradient Vanishing ด้วยการออกแบบการทำงานในส่วนของ Cell ใหม่ให้สามารถเก็บสถานะของการคำนวณได้ โดยใน Cell ของ LSTM นั้น มีหน่วยคำนวณย่อยเรียกว่า Gate ซึ่งประกอบด้วย Input Gate, Forget Gate, Memory Cell State Gate และ Output Gate [6]

#### 2.1.2 อัลกอริทึมคอมพลีเมนต์นาอิวเบย์ (Complement Naive Bay: CNB)

เป็นอัลกอริทึมใช้ทฤษฎีความน่าจะเป็นตามกฎของเบย์ เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่ง ๆ ร่วมกับข้อมูล เช่น ความน่าจะเป็นที่สังเกตได้สำหรับสมมติฐานหนึ่ง ๆ เพื่อหาสมมติฐานที่ดีที่สุด การเรียนรู้แบบเบย์อาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละสมมติฐาน [7]

#### 2.1.3 อัลกอริทึมเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor: KNN)

เป็นอัลกอริทึมที่มีการใช้วิธีการจัดแบ่งคลาส โดยอัลกอริทึมนี้จะตัดสินใจว่า คลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน (“K” ในขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุด) ของ

กรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไข หรือกรณีต่างๆ สำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด [8]

#### 2.1.4 อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน

##### (Support Vector Machine: SVM)

เป็นอัลกอริทึมที่สามารถนำมาช่วยแก้ปัญหาการจำแนกข้อมูล ใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูล โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกแยกกลุ่มข้อมูลได้ดีที่สุด [7]

#### 2.2 งานวิจัยที่เกี่ยวข้อง

Murthy and et al. [6] ทำการวิเคราะห์ความคิดเห็นด้วยอัลกอริทึม Long Short-Term Memory ข้อมูลที่ใช้เป็นข้อมูลแสดงความคิดเห็นเกี่ยวกับภาพยนตร์จากฐานข้อมูล IMDB จำนวน 50,000 ความคิดเห็น แบ่งเป็นความคิดเห็นเชิงบวก 25,000 ความคิดเห็น ความคิดเห็นเชิงลบ 25,000 ความคิดเห็น และสินค้าจากฐานข้อมูล Amazon จำนวน 50,000 ความคิดเห็น แบ่งเป็นความคิดเห็นเชิงบวก 25,000 ความคิดเห็น ความคิดเห็นเชิงลบ 25,000 ความคิดเห็นเช่นกัน ทำการวัดประสิทธิภาพด้วยค่าความถูกต้อง ผลการวิจัยพบว่า อัลกอริทึม Long Short Term Memory ให้ประสิทธิภาพดีที่สุดเมื่อมีค่า Dropout เท่ากับ 0.2 ค่า Batch Size เท่ากับ 500 โดยรันรอบที่ 1-10 มีค่าเฉลี่ยความถูกต้องเท่ากับร้อยละ 90.34

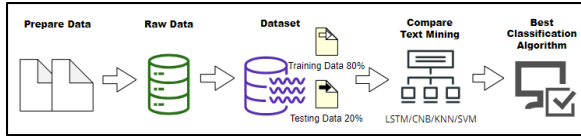
ดารงเดช ฉัตรเกล้า และจรรยา [7] ทำการเปรียบเทียบประสิทธิภาพโครงสร้างเหมืองข้อมูลเพื่อจำแนกโรคซึมเศร้าจากพฤติกรรมกรโพสต์ข้อความบนทวิตเตอร์ โดยทำการเปรียบเทียบประสิทธิภาพการจำแนก 2 ระดับ การจำแนกระดับหนึ่ง ด้วย 2 อัลกอริทึม คือ อัลกอริทึม Bayes อัลกอริทึม Support Vector Machine การจำแนกระดับสองด้วยอัลกอริทึม Support Vector Machine อัลกอริทึม Support Vector Machine

ร่วมกับอัลกอริทึม Bayes และอัลกอริทึม Support Vector Machine ร่วมกับอัลกอริทึม Random Forest ข้อมูลที่ใช้เป็นข้อความการ Retweet ที่เกี่ยวกับโรคซึมเศร้าบนทวิตเตอร์ที่บ่งบอกลักษณะอาการตามแบบสอบถาม DSM-5 9 รวบรวมข้อมูลจากผู้ใช้งานที่เป็นดาราต่างประเทศที่ป่วยเป็นโรคซึมเศร้า จำนวน 15 คน และผู้ใช้งานที่เป็นดาราต่างประเทศที่ไม่เป็นโรคซึมเศร้า จำนวน 15 คน โดยผู้ใช้งานแต่ละคนมีการโพสต์ข้อความมากกว่า 2 สัปดาห์ขึ้นไป ผลการวิจัยพบว่า ข้อมูล Training Set การจำแนกระดับหนึ่ง อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพดีที่สุด ค่าความถูกต้องร้อยละ 96.18 การจำแนกระดับสอง ผลการวิจัยพบว่า อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพดีที่สุด ค่าความถูกต้องร้อยละ 96.20 ข้อมูล Test Set การจำแนกระดับหนึ่ง อัลกอริทึม Bayes ให้ประสิทธิภาพดีที่สุด ค่าความถูกต้องร้อยละ 76.67 การจำแนกระดับสอง อัลกอริทึม Support Vector Machine ร่วมกับอัลกอริทึม Bayes ให้ประสิทธิภาพดีที่สุด ค่าความถูกต้องร้อยละ 73.33

สมศักดิ์ และสมัย [8] ทำการเปรียบเทียบการจำแนกบทวิจารณ์ออนไลน์ของนักท่องเที่ยว โดยทำการเปรียบเทียบด้วย 3 อัลกอริทึม คือ อัลกอริทึม Naïve Bayes อัลกอริทึม K-Nearest Neighbors และอัลกอริทึม Decision Tree ข้อมูลที่ใช้เป็นข้อมูลแสดงความคิดเห็นในกลุ่ม เฟสบุ๊กโรงแรมและที่พักจังหวัดลำปาง จำนวน 13,564 ความคิดเห็น นำมาจำแนกความคิดเห็นเชิงบวก และเชิงลบ ทำการวัดประสิทธิภาพด้วยค่าความถูกต้อง ผลการวิจัยพบว่า อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพดีที่สุด ค่าความถูกต้องร้อยละ 82.97

### 3. วิธีการดำเนินการวิจัย

การดำเนินงานวิจัยเพื่อจำแนกข้อความความคิดเห็นเกี่ยวกับรถยนต์ไฟฟ้าบนสื่อสังคมออนไลน์ ผู้วิจัยมีกรอบแนวคิดแสดงดังภาพที่ 1



ภาพที่ 1: กรอบแนวความคิด

จากภาพที่ 1 กรอบแนวความคิด สามารถแบ่งขั้นตอนการดำเนินการวิจัยเป็น 2 ส่วน ดังนี้

### 3.1 ขั้นตอนการรวบรวมและเตรียมข้อมูล

ทำการรวบรวมข้อมูลจากสื่อสังคมออนไลน์ คือ Twitter, Facebook และ Web Board โดยผ่านทาง <https://www.mediatoolkit.com> และ <https://pantip.com> ข้อมูลที่รวบรวมเป็นความคิดเห็นเห็นภาษาไทย และภาษาอังกฤษเกี่ยวกับรถยนต์ไฟฟ้าซึ่งเป็นทั้งความคิดเห็นเชิงบวก และเชิงลบ เมื่อได้ข้อมูลมาแล้วต้องนำมาคัดเลือก โดยใช้ python ในการจัดการตัดข้อมูลที่ไม่สามารถนำมาใช้ประกอบการวิเคราะห์ เช่น ข้อมูลที่เป็นข้อมูลรูปภาพ ข้อมูลที่เป็นข้อความของ Link URL หรือข้อมูลที่เป็นลักษณะเชิงโฆษณา เป็นต้น เมื่อได้ข้อมูลที่คัดเลือกแล้วนำข้อมูลมาทำการตัดคำภายในประโยค (Tokenize) เป็นการแบ่งข้อความออกมาเป็นคำ พิจารณาตัวอักษรพิเศษ และทำการปรับเป็นตัวพิมพ์เล็ก (Transform Case) ต่อจากนั้นมาทำการกำจัดคำหยุด (Stop Word) เมื่อตัดคำและกำจัดคำหยุดได้แล้วต่อมาทำการหารากศัพท์ (Stemmer) สำหรับความคิดเห็นภาษาไทยใช้ Pythainlp ในการประมวลผล ตัดคำด้วย attacut หลังจากนั้นจะนำคำที่ได้มาแปลงให้อยู่ในรูปแบบของเวกเตอร์ เตรียมนำเข้าสู่กระบวนการคัดเลือก และทดสอบประสิทธิภาพอัลกอริทึมเพื่อสร้างแบบจำลอง ตัวอย่างของข้อมูล แสดงดังตารางที่ 1

ตารางที่ 1: ตัวอย่างข้อมูล

ข้อความ	ความคิดเห็น
Ford has a really nice electric vehicle here. 🍌	Positive
Yeah, both brands have made huge leaps in quality and their warranties are best in class. I also like the Kia electric vehicle.	Positive
The electric vehicle charging problem.	Negative
I'm not anti-electric-vehicle in any way. But dismissing concerns over lithium battery fires	Negative

because ICE cars also catch fire is disingenuous and helps no one.	
ต้องขอบคุณ MG ที่ทำให้เมืองไทยได้แจ้งเกิดตลาดรถ EV แบบชาวบ้านเข้าถึงได้ อีกซัพพาก็จะโตขึ้นเรื่อยๆจนกลายเป็นประชากรส่วนใหญ่บนท้องถนน	Positive
อยากลองไฟฟ้าขึ้นมาเลยครับ	Positive
EV ยังอีกนานครับ ไม่ใช่เพราะเทคโนโลยีรถ แต่เพราะจุดขายยังไม่พร้อมมากกว่า  ถ้าใช้แค่ในเมือง กลับมาชาร์จบ้าน ก็ไม่มีปัญหา แต่ถ้าออกเดินทางท่องเที่ยว ก็เรื่องใหญ่แน่	Negative
เปลี่ยนแบตเตอรี่กะเป๋ากงแห้งกันเลยทีเดียว  วิถีแห่งการรักโลกมักจะไม่ค่อยรักเงินในกระเป๋า	Negative

### 3.2 ขั้นตอนการเปรียบเทียบอัลกอริทึม

นำข้อมูลที่ได้จากการเตรียมมาทำการคัดเลือกอัลกอริทึม โดยทำการเปรียบเทียบอัลกอริทึม 4 อัลกอริทึม ได้แก่ อัลกอริทึม LSTM อัลกอริทึม CNB อัลกอริทึม KNN และอัลกอริทึม SVM ทำการวัดประสิทธิภาพของอัลกอริทึม โดยพิจารณาค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึก ค่าความถ่วงดุล และค่าการวิเคราะห์เส้นโค้งคุณลักษณะสมบัติการทำงาน เพื่อคัดเลือกหาอัลกอริทึมที่ดีที่สุดเพื่อไปสร้างเป็นแบบจำลอง ซึ่งแต่ละอัลกอริทึมกำหนดเลือกพิจารณาพารามิเตอร์ที่สนใจ ทำการแบ่งข้อมูล ออกเป็น 2 ส่วน คือ ข้อมูลสำหรับการเรียนรู้ ร้อยละ 80 และข้อมูลสำหรับทดสอบ ร้อยละ 20 โดยค่าพารามิเตอร์ของแต่ละอัลกอริทึม แสดงดังตารางที่ 2

ตารางที่ 2: แสดงค่าพารามิเตอร์ของแต่ละอัลกอริทึม

LSTM	dense = 64, 128, 256, 512 activation function = sigmoid, relu
CNB	alpha = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0
KNN	n_neighbors = 1,2,3,4,5,6,7,8,9,10, weights= uniform, distance algorithm = auto, ball_tree, kd_tree, brute
SVM	kernel = linear, rbf, poly, sigmoid C = 0.1, 1, 10, 100 Gamma =0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0

## 4. ผลการดำเนินงาน

ผลการทดสอบประสิทธิภาพอัลกอริทึม แสดงดังตารางที่ 3 ถึง 6 โดยมีรายละเอียดดังนี้

ตารางที่ 3: ผลการทดสอบอัลกอริทึม LSTM

Parameter			Result				
Layer	Dense	Activation	Acc	P	R	F	ROC
64	64	sig	68.18	0.67	0.91	0.77	0.82

64	64	relu	79.87	0.84	0.80	0.82	0.89
64	128	sig	78.57	0.75	0.95	0.84	0.89
64	128	relu	75.32	0.78	0.74	0.76	0.82
64	256	sig	79.22	0.80	0.82	0.81	0.86
64	256	relu	75.97	0.84	0.79	0.82	0.85
<b>64</b>	<b>512</b>	<b>sig</b>	<b>83.12</b>	<b>0.82</b>	<b>0.91</b>	<b>0.86</b>	<b>0.93</b>
64	512	relu	79.22	0.77	0.83	0.80	0.82
128	64	sig	79.22	0.81	0.88	0.84	0.87

จากตารางที่ 3 ค่าพารามิเตอร์ LSTM ที่ให้ประสิทธิภาพดีที่สุดคือ ค่า Layer เท่ากับ 64 ค่า Dense เท่ากับ 512 และค่า Activation เท่ากับ sig โดยให้ค่าความถูกต้องเท่ากับร้อยละ 83.12 ค่าความแม่นยำเท่ากับ 0.82 ค่าระลอกเท่ากับ 0.91 ค่าการถ่วงดุลเท่ากับ 0.86 และค่าการวิเคราะห์เส้นโค้งคุณลักษณะการทำงานเท่ากับ 0.93 ตามลำดับ

ตารางที่ 4: ผลการทดสอบอัลกอริทึม CNB

Parameter	Result				
Alpha	Acc	P	R	F	ROC
0.1	82.50	0.80	0.87	0.83	0.82
0.2	82.50	0.80	0.87	0.83	0.82
0.3	82.50	0.80	0.87	0.83	0.82
0.4	83.00	0.80	0.88	0.83	0.82
<b>0.5</b>	<b>84.00</b>	<b>0.80</b>	<b>0.90</b>	<b>0.85</b>	<b>0.83</b>
0.6	83.00	0.80	0.88	0.83	0.82
0.7	83.00	0.80	0.88	0.83	0.82
0.8	83.00	0.80	0.88	0.83	0.82
0.9	83.00	0.80	0.88	0.83	0.82
1	83.00	0.80	0.88	0.83	0.82

จากตารางที่ 4 ค่าพารามิเตอร์ CNB ที่ให้ประสิทธิภาพดีที่สุดคือ ค่า Alpha เท่ากับ 0.5 โดยให้ค่าความถูกต้องเท่ากับร้อยละ 84.00 ค่าความแม่นยำเท่ากับ 0.80 ค่าระลอกเท่ากับ 0.90 ค่าการถ่วงดุลเท่ากับ 0.85 และค่าการวิเคราะห์เส้นโค้งคุณลักษณะการทำงานเท่ากับ 0.83 ตามลำดับ

ตารางที่ 5: ผลการทดสอบอัลกอริทึม KNN

Parameter			Result				
Neighbors	Weight	Algo	Acc	P	R	F	ROC
1	dis	auto	84.50	0.82	0.88	0.85	0.84

1	dis	ball	84.50	0.82	0.88	0.85	0.84
1	dis	kd	84.50	0.82	0.88	0.85	0.84
1	dis	brute	84.50	0.82	0.88	0.85	0.84
2	uni	auto	77.50	0.83	0.70	0.76	0.78
2	uni	ball	77.50	0.83	0.70	0.76	0.78
2	uni	kd	78.00	0.83	0.71	0.77	0.78
2	uni	brute	77.50	0.83	0.70	0.76	0.78
<b>2</b>	<b>uni</b>	<b>auto</b>	<b>85.00</b>	<b>0.83</b>	<b>0.88</b>	<b>0.86</b>	<b>0.85</b>

จากตารางที่ 5 ค่าพารามิเตอร์ KNN ที่ให้ประสิทธิภาพดีที่สุดคือ ค่า Neighbor เท่ากับ 2 ค่า Weight เท่ากับ distance และค่า Algorithm เท่ากับ auto โดยให้ค่าความถูกต้องเท่ากับร้อยละ 85.00 ค่าความแม่นยำเท่ากับ 0.83 ค่าระลอกเท่ากับ 0.88 ค่าการถ่วงดุลเท่ากับ 0.86 และค่าการวิเคราะห์เส้นโค้งคุณลักษณะการทำงานเท่ากับ 0.85 ตามลำดับ

ตารางที่ 6: ผลการทดสอบอัลกอริทึม SVM

Parameter			Result				
kerne l	Gamm a	C	Acc	P	R	F	RO C
rbf	0.6	1	49.50	0.00	0.00	0.00	0.50
rbf	0.6	1	84.50	0.88	0.80	0.84	0.85
rbf	0.6	10	87.00	0.89	0.84	0.87	0.85
rbf	0.7	1	49.50	0.00	0.00	0.00	0.50
rbf	0.7	1	85.00	0.89	0.80	0.84	0.85
<b>rbf</b>	<b>0.7</b>	<b>10</b>	<b>87.00</b>	<b>0.89</b>	<b>0.84</b>	<b>0.87</b>	<b>0.87</b>
rbf	0.8	1	49.50	0.00	0.00	0.00	0.50
rbf	0.8	1	85.00	0.89	0.80	0.84	0.85
rbf	0.8	10	86.50	0.89	0.83	0.86	0.87

จากตารางที่ 6 ค่าพารามิเตอร์ SVM ที่ให้ประสิทธิภาพดีที่สุดคือ ค่า Kernel เท่ากับ rbf ค่า Gamma เท่ากับ 0.7 และค่า C เท่ากับ 10 โดยให้ค่าความถูกต้องเท่ากับร้อยละ 87.00 ค่าความแม่นยำเท่ากับ 0.89 ค่าระลอกเท่ากับ 0.84 ค่าการถ่วงดุลเท่ากับ 0.867 และค่าการวิเคราะห์เส้นโค้งคุณลักษณะการทำงานเท่ากับ 0.87 ตามลำดับ

จากผลการทดสอบทั้ง 4 อัลกอริทึมสามารถสรุปผลประสิทธิภาพที่ดีที่สุดของแต่ละอัลกอริทึมได้ดังตารางที่ 7

ตารางที่ 7: ผลการทดสอบอัลกอริทึม SVM

Algorithm	Result				
	Acc	P	R	F	ROC
LSTM	83.12	0.82	0.91	0.86	0.93
CNB	84.00	0.80	0.90	0.85	0.83
KNN	85.00	0.83	0.88	0.86	0.85
SVM	<b>87.00</b>	<b>0.89</b>	<b>0.84</b>	<b>0.87</b>	<b>0.87</b>

จากตารางที่ 7 พบว่า อัลกอริทึม Support Vector Machine ให้ประสิทธิภาพในการวิเคราะห์ความคิดเห็นที่ดีที่สุด

## 5. สรุป

งานวิจัยฉบับนี้ผู้วิจัยได้นำเสนอการจำแนกข้อความความคิดเห็นเกี่ยวกับรถยนต์ไฟฟ้าบนสื่อสังคมออนไลน์จำนวน 1,000 ข้อความ แบ่งเป็นการแสดงความคิดเห็นเชิงบวก 500 ข้อความ และความเห็นเชิงลบ 500 ข้อความ นำข้อมูลไปทำความสะอาด ผ่านกระบวนการตัดคำ ต่อจากนั้นมาทำการกำจัดคำหยุด การหารากศัพท์ หลังจากนั้นจะนำคำที่ได้มาทำให้อยู่ในรูปแบบของเวกเตอร์เพื่อสร้างโมเดล ทำการแบ่งชุดข้อมูลเป็น 2 ชุด คือ ข้อมูลสำหรับการเรียนรู้ ร้อยละ 80 และข้อมูลสำหรับทดสอบ ร้อยละ 20 ทำการสร้างโมเดล และทดสอบประสิทธิภาพ 4 อัลกอริทึม คือ อัลกอริทึม Long Short Term Memory อัลกอริทึม Complement Naïve Bayes อัลกอริทึม K-Nearest Neighbor และอัลกอริทึม Support Vector Machine ผลการทดสอบประสิทธิภาพ พบว่า อัลกอริทึม Support Vector Machine มีประสิทธิภาพการจำแนกข้อความความคิดเห็นบนสื่อสังคมออนไลน์ที่ดีที่สุด โดยมีค่าความถูกต้องร้อยละ 87.00 ค่าความแม่นยำเท่ากับ 0.89 ค่าระลึกเท่ากับ 0.84 ค่าการถ่วงดุลเท่ากับ 0.867 และค่าการวิเคราะห์เส้นโค้งคุณลักษณะการทำงานเท่ากับ 0.87 ตามลำดับ

## เอกสารอ้างอิง

- [1] วาทีศย์ คำพรหมมา จักรชัย ไสอินทร์ และเพชร อิ่มทองคำ, “แบบจำลองการวิเคราะห์ความรู้สึกแบบผสมสำหรับความคิดเห็นต่อโรงแรมในประเทศไทยโดยใช้ K-Means และ

K-NN” ใน การประชุมวิชาการระดับชาติ สารสนเทศศาสตร์ วิชาการ 2019. (ม.ป.ท.), (1-11).

- [2] วัชรวิวัฒน์ จิตต์สกุล และสุนันทา สดสี, “การวิเคราะห์การจำแนกข้อความด้วยการเปรียบเทียบความเสถียรของอัลกอริทึม” *วารสารฉบับวิทยาศาสตร์และเทคโนโลยี ปีที่ 9* มกราคม – ธันวาคม 2560 หน้า 19-31.
- [3] นันทกัล สุทธิเลิศ, “การวิเคราะห์ความรู้สึกผู้โดยสารที่ใช้บริการสายการบินของบริษัทในประเทศสหรัฐอเมริกา” วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยบูรพา, 2560.
- [4] Lopamudra Dey, et al., “Sentiment Analysis of Review Datasets using Naïve Bayes and K-NN Classifier.” *International Journal of Information Engineering and Electronic Business*. Vol.8 No.4, pp. 54-62, 2016.
- [5] Dashtipour, K., et al., “Sentiment Analysis of Persian Movie Reviews Using Deep Learning.” *Entropy*. Vol.23 No.5, pp. 1-16, 2021.
- [6] Murthy, G. S. N., et al. (2020). “Text based Sentiment Analysis using LSTM.” *International Journal of Engineering Research & Technology (IJERT)*. Vol.9 Issue 5, pp. 299-303, 2020.
- [7] ดำรงเดช เคนริบรัมย์ ฉัตรเกล้า เจริญผล และจริยา จิรานุกูล. “การเปรียบเทียบประสิทธิภาพโครงสร้างเหมือนข้อมูลเพื่อจำแนกโรคซึมเศร้าจากพฤติกรรมกรมการโพสต์ข้อความบนทวิตเตอร์.” *วารสารวิทยาศาสตร์และเทคโนโลยี ปีที่ 39* ฉบับที่ 3 พฤษภาคม – มิถุนายน 2563 หน้า 1-13.
- [8] สมศักดิ์ ศรีสวกรย์ และ สมัย ศรีสว, “การวิเคราะห์เหมือนความคิดเห็นโดยใช้เทคนิคการสกัดคำ.” *วารสารวิชาการการประยุกต์ใช้เทคโนโลยีสารสนเทศ ปีที่ 6* ฉบับที่ 2 กรกฎาคม – ธันวาคม 2563 หน้า 96-104.

# ปัจจัยที่มีอิทธิพลต่อความผูกพันของผู้บริโภคในการใช้บริการแอปพลิเคชันแชทบอท ในเฟซบุ๊กเมสเซนเจอร์

## Factors Influencing Customer Engagement in Using Facebook Messenger ChatBot Services

ศุภวรรณ พงศ์ศุภวรรณ (Supawan Pongsupawan)<sup>1</sup> และปราโมทย์ ลือนาม (Pramote Luenam)<sup>2</sup>

สาขาบริหารเทคโนโลยีสารสนเทศ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

<sup>1</sup>supawan.pong@stu.nida.ac.th, <sup>2</sup>pramote@nida.ac.th

### บทคัดย่อ

ในปัจจุบันนี้องค์กรต่าง ๆ ได้นำเทคโนโลยีมาส่งเสริมการตลาดและรวมถึงการเข้าถึงลูกค้าของบริษัทตนเอง ซึ่งความเร็วในการตอบสนองนับเป็นปัจจัยหนึ่งในการสร้างความพึงพอใจให้แก่ลูกค้า ดังนั้นหลายบริษัทจึงได้นำแพลตฟอร์มแชทบอท เข้ามาใช้ในบริษัทตนเองเนื่องจากแชทบอท สามารถทำงานได้ 24 ชม ต่างจากมนุษย์ที่จะมีความเหนื่อยล้าเกิดขึ้นและรวมถึงการลดต้นทุนของบริษัทในด้านการจ้างพนักงาน นอกจากนี้ยังสร้างความพึงพอใจให้แก่ลูกค้า และการมีส่วนร่วมของลูกค้า งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษา และวิเคราะห์ปัจจัยที่มีอิทธิพลต่อความผูกพันของผู้บริโภคในการใช้บริการแอปพลิเคชันแชทบอทในเฟซบุ๊กเมสเซนเจอร์ โดยใช้แบบสอบถามออนไลน์เป็นเครื่องมือในการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างที่เคยใช้บริการแอปพลิเคชันแชทบอทในเฟซบุ๊กเมสเซนเจอร์จำนวน 348 คน เมื่อนำข้อมูลมาวิเคราะห์ด้วยโมเดลสมการโครงสร้าง เพื่อทดสอบสมมติฐาน และหาความสอดคล้องกับข้อมูลเชิงประจักษ์ โมเดลแสดงความเหมาะสมของด้วยค่า  $P\text{-value}=0.966$ ,  $\chi^2/df=0.070$ ,  $RMSEA=0.000$ ,  $GFI=1.000$ ,  $CFI=1.00$ ,  $NFI=1.00$  และ  $TLI=1.017$  จากผลการวิจัยพบว่า ปัจจัยด้านองค์ประกอบคุณภาพของแชทบอทอันได้แก่ความมีประสิทธิภาพ ความปลอดภัย และการตอบสนองล้วนมีอิทธิพลต่อทัศนคติของลูกค้าซึ่งก่อให้เกิดความผูกพันอันดีของลูกค้าต่อตัวสินค้า รวมถึงเพศสภาพของแชทบอทมีผลต่อความผูกพันของลูกค้าเช่นกัน ส่วนความสัมพันธ์ระหว่างความคาดหวังของลูกค้าต่อเพศสภาพของแชทบอทกับความผูกพันของลูกค้าพบว่าไม่แปรผันตามประเภทสินค้า

บอทกับความผูกพันของลูกค้าพบว่าไม่แปรผันตามประเภทสินค้า

**คำสำคัญ:** แชทบอทในเฟซบุ๊กเมสเซนเจอร์ ความคาดหวังของลูกค้าต่อเพศสภาพของแชทบอท องค์ประกอบด้านคุณภาพของแชทบอท

### Abstract

At present, various organizations incorporate new technologies to promote marketing activities and serves as a communication channel for businesses. The speed of response is one of the factors that contribute to customer satisfaction. Therefore, many companies have adopted the chatbot technology. A chatbot can work 24/7 without breaks and it would never ask for a leave or does not get tired like human. Chatbots lead to a reduction in operating cost and an increase in customer satisfaction and customer engagement. This research aims to study and analyze the factors that influence consumer engagement in using Facebook Messenger chatbot services. The study used online questionnaire survey to collect a sample of 348 of people who have used Facebook Messenger chatbot services. The data were analyzed through Structural Equation Model to test the hypotheses and derive model fit. The model showed a goodness-of-fit with  $P\text{-Value}=0.966$ ,  $\chi^2/df = 0.070$ ,  $RMSEA=0.00$ ,  $GFI=1.000$ ,  $CFI=1.00$ ,  $NFI=1.00$  and  $TLI=1.017$ . The findings show that the quality components of the chatbot (i.e., efficiency

and security) have a positive effect on customers' attitude which leads to an increase in customer's engagement. The chatbot's gender also has a significant effect on customer engagement. However, the relationship between the expectation of customer for chatbot perceived gender and customer's engagement does not vary by the product type.

**Keywords:** Facebook messenger chatbot, Customer expectations of the gender of chatbot, Quality component of the chatbot.

## 1. บทนำ

ปัจจุบันบทบาทความสำคัญของเทคโนโลยีได้เข้ามามีส่วนสำคัญต่อธุรกิจ เนื่องจากความมีประสิทธิภาพ รวมไปถึงถึงความรวดเร็ว นับว่าเป็นปัจจัยที่ทำให้เกิดความสัมพันธ์อันดีต่อลูกค้า

แชทบอท คือ โปรแกรมคอมพิวเตอร์ที่ประมวลผลการนำเข้าภาษาธรรมชาติจากผู้ใช้งานและมีความสัมพันธ์ในการตอบสนองในการตอบกลับอย่างชาญฉลาด ขับเคลื่อนโดยปัญญาประดิษฐ์ ที่ได้ตอบกับผู้ใช้ผ่านอินเทอร์เฟซแบบข้อความเป็นหลัก และได้รับความนิยมจากหลายๆ บริษัทเนื่องจาก แชทบอท สามารถทำงานได้ 24 ชม ต่างจากมนุษย์ที่จะมีความเหนื่อยล้า รวมถึงการลดต้นทุนของบริษัทในด้านการจ้างพนักงาน สร้างความพึงพอใจให้แก่ลูกค้า (customer satisfaction) และการมีส่วนร่วม ของลูกค้า (customer engagement) เหตุผลหลักที่ทำให้มีการนำแชทบอทมาใช้งานในธุรกิจ เพราะแชทบอทมีความรวดเร็วในการตอบคำถาม ได้ทันที ร้อยละ 83 ของผู้ซื้อสินค้าบนโลกออนไลน์ต้องการคำปรึกษาระหว่างการตัดสินใจซื้อ เพื่อค้นหาสินค้าที่ตนต้องการ การนำแชทบอทมาประยุกต์กับธุรกิจจะทำให้ลูกค้าได้รับประสบการณ์ที่ดีและตัดสินใจซื้อสินค้าหรือรับบริการได้ง่ายขึ้น และพบว่าร้อยละ 67 ของผู้ใช้อินเทอร์เน็ตซื้อผลิตภัณฑ์หรือบริการจากแบรนด์ที่ใช้แชทบอท เพราะแชทบอทสามารถจัดเก็บข้อมูลของลูกค้าได้โดยอัตโนมัติ ซึ่งเป็นประโยชน์ต่อธุรกิจเพื่อที่จะเข้าใจพฤติกรรมของลูกค้าในแต่ละราย เพื่อส่งเสริมให้เกิดการขายรวมถึงโปรโมชั่น

ส่งเสริมการขาย พบว่าร้อยละ 77 ของแชทบอททำให้ประสิทธิภาพการขายหรือการบริการหลังการขายเพิ่มขึ้น อีกทั้งยังมีศักยภาพในการจัดการความต้องการของลูกค้าได้มีประสิทธิภาพมากกว่ามนุษย์ รวมถึงการลดต้นทุนการทำ Customer Service ลดลงร้อยละ 30 อันเนื่องมาจากการจ้างพนักงานทำให้ธุรกิจสามารถลดจำนวนพนักงานสำหรับการตอบคำถามได้ [15] ดังนั้นผู้วิจัยจึงสนใจในการศึกษาปัจจัยที่มีผลต่อการมีส่วนร่วมของลูกค้าในการใช้บริการแชทบอท ทั้งนี้งานวิจัยนี้จะศึกษาเฉพาะเฟซบุ๊ก เมสเซนเจอร์ แชทบอท เนื่องจากปัจจุบัน (อ้างอิงข้อมูลเดือน มกราคม 2021) ประเทศไทยมีผู้ใช้งานเฟซบุ๊กจำนวน 51 ล้านบัญชีมีการใช้ เฟซบุ๊กเมสเซนเจอร์ จำนวนถึง 37 ล้านบัญชี [16] งานวิจัยนี้จัดทำขึ้นเพื่อศึกษาและวิเคราะห์ปัจจัยต่าง ๆ ที่ส่งผลต่อการมีส่วนร่วมของลูกค้า ความพึงพอใจของลูกค้าซึ่งมีผลต่อการตัดสินใจ กลับมาใช้บริการซ้ำซึ่งแสดงถึงความภักดีของลูกค้าที่มีต่อแบรนด์ เพื่อเป็นแนวทางในการปรับปรุงการให้บริการแชทบอทให้เหมาะสมกับกลุ่มเป้าหมาย

## 2. งานวิจัยและทฤษฎีที่เกี่ยวข้อง

ในการศึกษาครั้งนี้ ผู้วิจัยทำการศึกษาค้นคว้า รวบรวมแนวคิดและทฤษฎีต่าง ๆ ที่เกี่ยวข้อง ดังนี้

แนวคิดเกี่ยวกับความผูกพันของลูกค้า (Customer engagement) คือ การทำให้ลูกค้าเกิดความผูกพันระยะยาวกับองค์กรหรือแบรนด์ [1] ซึ่งความผูกพันของลูกค้าเป็นกระบวนการทางจิตวิทยาที่ลูกค้าแสดงออกมาทำให้เกิดการภักดีและการซื้อซ้ำ [2] ซึ่ง [3] ได้มีการระบุว่าความผูกพันของลูกค้ามีองค์ประกอบคือ ส่วนของความคิด ได้แก่ ความกระตือรือร้น ส่วนของอารมณ์ ได้แก่ การตระหนักในการมีส่วนร่วม และส่วนของพฤติกรรม ได้แก่ ปฏิสัมพันธ์ทางสังคมการที่จะประสบความสำเร็จในการมีปฏิสัมพันธ์กับลูกค้า รวมถึงการตอบสนองต่อความต้องการ ควรมีกระบวนการที่ทำให้ลูกค้ารับรู้ถึงเป็นส่วนตัวซึ่งกระบวนการนี้ควรมีทั้งความเร็วและความฉลาด รวมถึงสามารถส่งมอบความต้องการที่เฉพาะของลูกค้ารายใดรายหนึ่งได้ตลอดเวลาเพื่อให้บรรลุถึงความสำเร็จของ



เป้าหมาย โดยผลลัพธ์นี้เน้นลูกค้าเป็นศูนย์กลางที่มีประสิทธิภาพ ดังนั้นเพื่อตอบสนองถึงความเร็วและมีการบริการที่เหมาะสมในแต่ละบุคคลจึงเป็นการดีที่นำแชทบอทเข้ามามีส่วนร่วมแก่ลูกค้าซึ่งสามารถช่วยให้เข้าถึงกลุ่มลูกค้าขนาดใหญ่และโอกาสทางธุรกิจที่มีมากขึ้น รวมถึงปฏิสัมพันธ์ที่แน่นแฟ้นของทั้งลูกค้าและบริษัท ซึ่งการมีส่วนร่วมของลูกค้า ได้แก่ การเชื่อมต่อ การมีปฏิสัมพันธ์ ความพึงพอใจ การเก็บรักษา ความมุ่งมั่น และการสนับสนุน เป็นส่วนสำคัญและเป็นแรงผลักดันให้ลูกค้าเกิดความรู้สึกพึงพอใจจนเกิดความจงรักภักดี [4]

ทัศนคติ คือ ความชอบ ความพึงพอใจ หรือไม่ชอบ ที่บุคคลแสดงออกมาต่อสิ่งต่าง ๆ [5] ซึ่งการวัดความพึงพอใจของลูกค้าจะนำแนวคิดและทฤษฎีที่เกี่ยวกับคุณภาพการบริการเข้ามาเนื่องจากหากลูกค้าได้รับการบริการที่ต้องการหรือเกินกว่าความคาดหวังจะทำให้เกิดความรู้สึกอันดี ซึ่งงานวิจัยได้ประยุกต์ในการนำโมเดล SERVQUAL ซึ่งเป็น โมเดลที่ศึกษาคุณภาพการให้บริการที่จะส่งผลต่อความคาดหวัง ความพึงพอใจ และการรับรู้ของผู้บริโภค โดย [6] นำหลักของการประเมินคุณภาพของการบริการมาประยุกต์ซึ่งปัจจัยนี้จะส่งผลต่อความพึงพอใจในการได้รับบริการ โดยผู้วิจัยได้นำด้านประสิทธิภาพของการใช้งาน (Efficiency) มาประยุกต์ใช้กับการประเมินคุณภาพของแชทบอทมา 3 ด้านอันจะก่อให้เกิดความพึงพอใจได้แก่ (1) ประสิทธิภาพของการใช้งาน (Efficiency: EF) ที่นำมาใช้กับแชทบอท คือ ความมีประสิทธิภาพต่อการใช้งานของแชทบอท สามารถใช้งานได้ง่ายและมีความพร้อมต่อการใช้งาน (2) ความปลอดภัย (Security: SE) ทำให้ผู้ใช้บริการรู้สึกปลอดภัยในการใช้บริการแชทบอท (3) การตอบสนองลูกค้า (Responsiveness: RE) ซึ่งแชทบอทควรมีความพร้อมและความเร็วต่อการตอบสนองแก่ผู้รับบริการรวมถึงคำตอบที่น่าเชื่อถือซึ่งจะทำให้ผู้รับบริการเกิดความมั่นใจ ซึ่งสำหรับการตอบสนองนับเป็นหนึ่งในคุณลักษณะด้านคุณภาพของแชทบอทที่สำคัญและสามารถใช้ในการรับรองคุณภาพของแชทบอท [7] มาประยุกต์ใช้ในการวัด

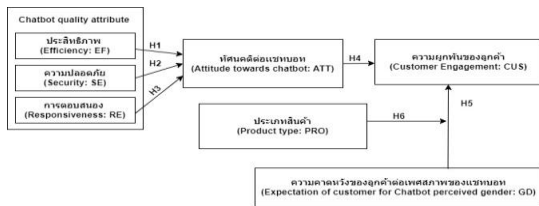
มิติคุณภาพด้านความพึงพอใจของผู้รับบริการที่มีต่อแชทบอท เพื่อศึกษาว่าปัจจัยใดส่งผลต่อความพึงพอใจในการใช้บริการแชทบอทเพื่อจะได้นำไปพัฒนาประสิทธิภาพให้ตรงต่อความต้องการของลูกค้าซึ่งทำให้เกิดความพึงพอใจในด้านการใช้บริการมากยิ่งขึ้น

นอกจากนี้งานวิจัยได้ทำการศึกษาด้านความคาดหวังของลูกค้าต่อเพศสภาพของแชทบอท (Expectation of customer for Chabot gender: GD) เพื่อศึกษาว่ามีผลต่อความผูกพันของลูกค้ามากน้อยเพียงใด พบว่าแบบแผนทางเพศของแชทบอทอิงจากรูปลักษณะของแชทบอท [8] และเสียง [10] เพื่อบ่งบอกถึงเพศสภาพของแชทบอท แต่เนื่องจากงานวิจัยนี้เป็นการบริการของแชทบอทในเฟซบุ๊กเมสเซนเจอร์ ดังนั้นจะไม่มีรูปลักษณะในการให้บริการและกลุ่มประชากรที่ศึกษาวิจัยอาศัยอยู่ที่ประเทศไทย จึงมีคำลงในภาษาไทยเช่น “ค่ะ” แทนเพศสภาพผู้หญิง “ครับ” แทนเพศสภาพผู้ชาย หรือคำขึ้นต้นนาม เช่น “กระผม” แทนเพศสภาพผู้ชาย “ดิฉัน” แทนเพศสภาพผู้หญิง เป็นต้น [10] ในส่วนของความคาดหวังต่อเพศที่มีต่อแชทบอท ผู้วิจัยได้นำแนวคิดเรื่องแบบแผนทางเพศซึ่งเป็นความเชื่อต่อบทบาททางเพศ กล่าวคือ เป็นความเชื่อที่ผู้หญิงและผู้ชายมีลักษณะแตกต่างกันและสามารถนำมาประยุกต์ใช้ได้ในส่วนประกอบประเภทของสินค้าที่แตกต่างกัน มีผลต่อความคาดหวังเพศสภาพของแชทบอทมากน้อยเพียงใด พบว่าคอมพิวเตอร์และหุ่นยนต์จากการวิจัยด้านปฏิสัมพันธ์ระหว่างมนุษย์กับหุ่นยนต์ โดยหุ่นยนต์เพศชายเหมาะกับงานลักษณะที่เป็นตัวแทนและรวมถึงงานที่เป็นส่วนของผู้ชาย อาทิ งานทางด้านเทคนิค ส่วนหุ่นยนต์ผู้หญิงจะเป็นที่ยอมรับในด้านของงานลักษณะ เช่น การดูแลผู้สูงอายุ เด็ก เป็นต้น [11] และความเหมาะสมของเพศสภาพแชทบอทต่อประเภทสินค้า เกี่ยวข้องกับความแตกต่างของแบบแผนทางเพศซึ่งสามารถพิจารณาได้ทางซอฟต์แวร์คอมพิวเตอร์หรือรวมถึงแชทบอท เช่น ด้านการบริการหรือสอบถามด้านเทคนิค ผู้รับบริการจะมีความมั่นใจหากเพศสภาพของแชทบอทเป็นเพศชายมากกว่าเพศหญิง ตรงข้ามกับงานส่วนบริการลูกค้าหรือสายด่วน การท่องเที่ยว แฟชั่น ความงาม จะเกิดปฏิสัมพันธ์หากแช

ทบอทนั้นเป็นเพศหญิง [14] ซึ่งนำมาประยุกต์ใช้กับธุรกิจ โดยอิงจากประเภทสินค้าโดยใช้เพศสภาพที่เหมาะสมของ แชนบอทเพื่อให้ผู้รับบริการเกิดความมั่นใจในตัวสินค้า เช่นธุรกิจด้านความงาม เพศสภาพของแชนบอทควรเป็น เพศหญิง แต่หากธุรกิจด้านอุปกรณ์ตกแต่งรถยนต์เพศสภาพ ของแชนบอทควรเป็นเพศชาย ด้านของสมมติฐานด้าน ความคาดหวังเพศสภาพของแชนบอท ได้นำทฤษฎีที่ใช้ สำหรับการอธิบายผลของความสอดคล้องทางเพศในการ มีปฏิสัมพันธ์กัน ซึ่งกล่าวว่าผู้คน มักดึงดูดคนที่ มี บุคลิกลักษณะคล้ายคลึงกับตนเอง [12] และรวมถึง สมมติฐานการมีปฏิสัมพันธ์ระหว่างเพศเดียวกันจะส่งผล ให้เกิดบริบททางการขายหรือบริการเกิดขึ้น [13]

### 3. กรอบแนวคิดของการวิจัย

จากการศึกษาแนวคิด และทฤษฎีที่เกี่ยวข้องสามารถ กำหนดกรอบแนวคิดงานวิจัยที่จะศึกษาได้ ดังแสดงใน ภาพที่ 1



ภาพที่ 1: กรอบแนวคิดงานวิจัย

**สมมติฐานที่ 1 (H1) :** ประสิทธิภาพของการใช้งาน แชนบอท (EF) มีอิทธิพลต่อความผูกพันของลูกค้า (CUS) ผ่านทัศนคติต่อแชนบอท (ATT)

**สมมติฐานที่ 2 (H2) :** ความปลอดภัย (SE) มีอิทธิพลต่อความผูกพันของลูกค้า (CUS) ผ่านทัศนคติต่อแชนบอท (ATT)

**สมมติฐานที่ 3 (H3) :** การตอบสนอง (RE) มีอิทธิพลต่อความผูกพันของลูกค้า (CUS) ผ่านทัศนคติต่อแชนบอท (ATT)

**สมมติฐานที่ 4 (H4) :** ทัศนคติต่อแชนบอท (ATT) มีผลต่อความผูกพันของลูกค้า (CUS)

**สมมติฐานที่ 5 (H5) :** ความคาดหวังของลูกค้าต่อเพศสภาพของแชนบอท (GD) มีผลต่อความผูกพันของลูกค้า (CUS)

**สมมติฐานที่ 6 (H6) :** ความคาดหวังของลูกค้าต่อเพศสภาพของแชนบอท (GD) มีผลต่อความผูกพันของลูกค้า (CUS) แปรผันตามประเภทสินค้า (PRO)

### 4. วิธีการดำเนินการวิจัย

ในการศึกษางานวิจัยนี้เป็นการวิจัยเชิงปริมาณ (quantitative research) ใช้แบบสอบถามออนไลน์ในการรวบรวมข้อมูลจากกลุ่มตัวอย่าง ประชากรของงานศึกษานี้ เป็นบุคคลทั่วไปที่อาศัยที่ประเทศไทยและเป็นผู้ใช้บริการแอปพลิเคชันแชนบอทในเฟซบุ๊ก เมสเซนเจอร์ โดยผู้วิจัยใช้โปรแกรม G\*Power ในการคำนวณหาขนาดตัวอย่างประชากร โดยกำหนดค่าเพาเวอร์ (1-β) เท่ากับ 0.95 ค่าอัลฟา (α) เท่ากับ 0.05 ขนาดของอิทธิพล (effect size) เท่ากับ 0.3 จำนวนตัวแปรทำนาย เท่ากับ 7 ตัวแปร ผลที่ได้คือ ขนาดของกลุ่มตัวอย่างเท่ากับ 134 ตัวอย่างซึ่งผู้วิจัยได้ปรับจำนวนขนาดตัวอย่างเพิ่มเป็น 318 ตัวอย่างและเพิ่มการเก็บข้อมูลที่ใช้สำหรับการศึกษานำร่อง (pilot study) เพื่อหาความเชื่อมั่นอีก 30 ตัวอย่าง รวมเป็น 348 ตัวอย่าง ดังนั้นขนาดของกลุ่มตัวอย่างที่คำนวณได้จากสูตรคือ 348 คน โดยใช้วิธีการเลือกกลุ่มตัวอย่างแบบสะดวก (Convenience Sampling)

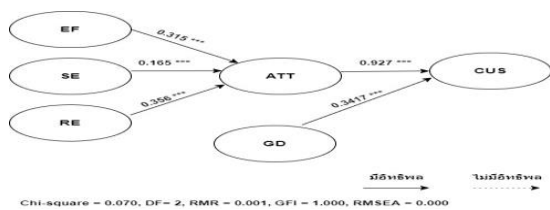
ผู้วิจัยได้ทำการทดสอบสมมติฐานของงานวิจัยโดยใช้โปรแกรม SPSS และ AMOS ในการวิเคราะห์รูปแบบเชิงโครงสร้าง (Structural Equation Modelling: SEM) เพื่อวิเคราะห์หาค่าอิทธิพลทั้งทางตรงและทางอ้อม และได้มีการปรับรูปแบบของโมเดล เพื่อสอดคล้องกับข้อมูลเชิงประจักษ์ (Model Fit)

### 5. ผลการดำเนินงานวิจัย

ในการศึกษาวิจัยครั้งนี้มีผู้ตอบแบบสอบถามทั้งหมด 348 คน โดยมีผู้ใช้บริการแอปพลิเคชันแชนบอทในเฟซบุ๊กเมสเซนเจอร์ที่สามารถนำมาวิเคราะห์ข้อมูลได้เป็นจำนวน 348 คน แบ่งเป็นเพศชาย 106 คน เพศหญิง 242 คน หากแบ่งตามช่วงอายุสามารถแบ่งเป็น ช่วงอายุต่ำกว่า 25 ปี จำนวน 104 คน ช่วงอายุ 25-30 ปี จำนวน 107 คน ช่วงอายุ 31-40 ปี จำนวน 97 คน ช่วงอายุ 41-50 ปี จำนวน 27 คน และมากกว่า 51 ปี จำนวน 13 คน

การวิจัยใช้การวิเคราะห์รูปแบบเชิงโครงสร้าง โดยเป็นการวิเคราะห์เส้นทาง (path analysis) เพื่อหาค่าอิทธิพลทั้งทางตรงและทางอ้อมของตัวแปรอิสระว่ามีผลต่อตัวแปรตามมากน้อยเพียงใด โดยค่าความสอดคล้องของโมเดล (goodness-of-fit) ที่ได้เป็นดังนี้  $\chi^2/df = 0.035$ , RMSEA = 0.000, GFI = 1.000, CFI = 1.00, NFI = 1.000 และ TLI = 1.017 ดังแสดงในภาพที่ 2

รวมถึงการทดสอบอิทธิพลกำกับ (moderating effect) ด้วยเทคนิคการวิเคราะห์กลุ่มพหุ (Multiple-Group SEM) โดยดำเนินการวิเคราะห์อิทธิพลของตัวแปรชนิดสินค้า ดังแสดงในตารางที่ 2



ภาพที่ 2: โมเดลจากการวิจัย

ตารางที่ 1: ผลการวิเคราะห์เชิงสาเหตุของโมเดล

ตัวแปรผล	อิทธิพล	ตัวแปรต้น					ตัวแปรกำกับ
		EF	SE	RE	GD	ATT	
ATT	ทางตรง	0.315***	0.165***	0.356***	-	-	-
	ทางอ้อม	-	-	-	-	-	-
CUS	ทางตรง	-	-	-	0.3417***	0.927***	0.001
	ทางอ้อม	0.292*	0.153*	0.330*	-	-	-

หมายเหตุ \* $p \leq 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

ผลที่ได้จากการวิเคราะห์อิทธิพลเชิงสาเหตุภายในโมเดล สามารถนำมาสรุปผลการทดสอบของสมมติฐานต่าง ๆ ที่ตั้งไว้ มีรายละเอียดที่แสดงในตารางที่ 3

ตารางที่ 2: สรุปผลการทดสอบสมมติฐาน

สมมติฐาน	ค่าอิทธิพล	ค่า P	ผลการทดสอบ
H1: ประสิทธิภาพของการใช้งานแชทบอท (EF) มีอิทธิพลต่อความผูกพันของลูกค้า (CUS) ผ่านทัศนคติต่อแชทบอท (ATT)	0.315	$\leq 0.001$	ยอมรับ
H2: ความปลอดภัย (SE) มีอิทธิพลต่อความผูกพันของลูกค้า (CUS) ผ่านทัศนคติต่อแชทบอท (ATT)	0.165	$\leq 0.001$	ยอมรับ
H3: การตอบสนอง (RE) มีอิทธิพลต่อความผูกพันของลูกค้า (CUS) ผ่านทัศนคติต่อแชทบอท (ATT)	0.356	$\leq 0.001$	ยอมรับ

สมมติฐาน	ค่าอิทธิพล	ค่า P	ผลการทดสอบ
H4: ทัศนคติต่อแชทบอท (ATT) มีผลต่อความผูกพันของลูกค้า (CUS)	0.927	$\leq 0.001$	ยอมรับ
H5: ความคาดหวังของลูกค้าต่อประสิทธิภาพของแชทบอท (GD) มีผลต่อความผูกพันของลูกค้า (CUS)	0.106	$\leq 0.001$	ยอมรับ
H6: ความสัมพันธ์ระหว่างความคาดหวังของลูกค้าต่อประสิทธิภาพของแชทบอท (GD) กับความผูกพันของลูกค้า (CUS) แปรผันตามประเภทสินค้า (PRO)	0.001	0.991	ปฏิเสธ

## 6. สรุป

ผลการศึกษาปัจจัยที่มีอิทธิพลต่อความผูกพันของผู้บริโภคในการใช้บริการแอปพลิเคชันแชทบอทในเฟซบุ๊กเมสเซนเจอร์ พบว่า ประสิทธิภาพของการใช้งานแชทบอทมีอิทธิพลต่อความผูกพันของลูกค้าผ่านทัศนคติต่อแชทบอทแสดงเป็นการอนุมานให้เห็นว่าการที่ผู้รับบริการได้รับการบริการจากแชทบอทที่มีความถูกต้องแม่นยำ รวมถึงมีประสิทธิภาพและความน่าเชื่อถือ จะส่งผลต่อทัศนคติด้านความพึงพอใจในการใช้บริการแชทบอท

ความปลอดภัย มีอิทธิพลต่อความผูกพันของลูกค้าผ่านทัศนคติต่อแชทบอทแสดงให้เห็นว่า ความปลอดภัยในการใช้บริการ ส่งผลให้เกิดทัศนคติด้านความพึงพอใจในการใช้บริการแชทบอท ซึ่งสอดคล้องกับแนวคิดคำตอบที่น่าเชื่อถือจะทำให้ผู้รับบริการเกิดความมั่นใจและพอใจ

การตอบสนองมีอิทธิพลต่อความผูกพันของลูกค้า (ผ่านทัศนคติต่อแชทบอท แสดงให้เห็นว่าปัจจัยด้านการตอบสนองของแชทบอทซึ่งเน้นความเร็วถูกต้องซึ่งเป็นคุณสมบัติเด่นของแชทบอทส่งผลทำให้เกิดความพึงพอใจซึ่งสอดคล้องกับสมมติฐานความเร็วต่อการตอบสนองแก่ผู้รับบริการ รวมถึงคำตอบที่น่าเชื่อถือซึ่งจะทำให้ผู้รับบริการเกิดความมั่นใจและพอใจ

ซึ่งทั้ง 3 ปัจจัยข้างต้นมีอิทธิพลต่อความผูกพันของลูกค้าผ่านทัศนคติต่อแชทบอทซึ่งสอดคล้องกับแนวคิดการวัดความพึงพอใจของลูกค้าซึ่ง 3 ปัจจัยข้างต้นเป็น 1 ในองค์ประกอบของโมเดล SERQUAL เพื่อวัดคุณภาพการให้บริการที่จะส่งผลต่อความคาดหวัง ความพึงพอใจ และการรับรู้ของลูกค้า ซึ่งจากคุณลักษณะด้านคุณภาพของ

แทบบอทข้างต้น ได้มีอิทธิพลส่งผลให้เกิดความผูกพันของลูกค้าในการใช้ บริการแทบบอทตามมา [6]

ทัศนคติต่อแทบบอท มีผลต่อความผูกพันของลูกค้า แสดงให้เห็นว่า ทัศนคติของลูกค้าส่งผลต่อความผูกพันในการใช้บริการ ซึ่งสอดคล้องกับงานวิจัยด้านวงจรการมีส่วนร่วมของลูกค้า ซึ่งความพึงพอใจ เป็นแรงผลักดันให้ลูกค้าเกิดความรู้สึกพึงพอใจจนเกิดความจงรักภักดี [4] และเกิดความผูกพัน

ในส่วนของความคาดหวังของลูกค้าต่อเพศสภาพของแทบบอทมีผลต่อความผูกพันของลูกค้า แสดงให้เห็นว่าลูกค้ามีความคาดหวังต่อเพศสภาพของแทบบอทในการให้บริการซึ่งสอดคล้องกับทฤษฎีที่ใช้สำหรับการอธิบายผลของความสอดคล้องทางเพศในการมีปฏิสัมพันธ์กันซึ่งกล่าวว่าผู้คนมักดึงดูดคนที่มีความคล้ายคลึงกับตนเอง [12] และรวมถึงสมมุติฐานการมีปฏิสัมพันธ์ระหว่างเพศเดียวกันจะส่งผลให้เกิดบริบททางการขายหรือบริการเกิดขึ้น [13]

ส่วนความสัมพันธ์ระหว่างความคาดหวังของลูกค้าต่อเพศสภาพของแทบบอทกับความผูกพันของลูกค้าไม่ได้แปรผันตามประเภทสินค้า ซึ่งไม่ได้สอดคล้องกับแนวคิดลักษณะความเหมาะสมของงานที่แตกต่างกันของหุ่นยนต์หญิงและชาย [11] รวมถึงความเหมาะสมของเพศสภาพของแทบบอทต่อประเภทสินค้า[14] ซึ่งอาจเป็นไปได้ว่างานวิจัยชิ้นนี้ได้ศึกษาเฉพาะแทบบอทในเฟซบุ๊กเมสเซนเจอร์ ซึ่งจะไม่มิลักษณะทางกายภาพที่บ่งบอกของเพศสภาพแทบบอทชัดเจนจึงทำให้ประเภทสินค้าไม่ได้มีอิทธิพลกำกับต่อความคาดหวังของเพศสภาพแทบบอทต่อความผูกพันของลูกค้า

**ข้อเสนอแนะ**ในการวิจัยครั้งต่อไป ควรนำการศึกษาของแทบบอทที่มีลักษณะทางกายภาพที่แสดงออกถึงเพศสภาพที่ชัดเจน อาทิ มีหน้าตา เสียง ซึ่งบ่งบอกถึงเพศสภาพได้มากกว่าแทบบอทในเฟซบุ๊กเมสเซนเจอร์ซึ่งอาศัยข้อความและคำลงท้ายเพื่อบ่งบอกเพศสภาพว่าเป็นเพศหญิงหรือชาย และควรมีการเจาะจงลักษณะประเภทของชนิดสินค้าเพื่อความแม่นยำในการแปลงผลของข้อมูลที่ได้ รวมถึงกำหนดช่วงกลุ่มวัยที่แตกต่างกันเพื่อศึกษาว่าปัจจัยใด

ส่งผลกับคนกลุ่มใดมากกว่ากัน และนำปัจจัยที่ได้มาวิเคราะห์เพื่อนำไปพัฒนาระบบให้ดียิ่งขึ้น

### เอกสารอ้างอิง

- [1] J.W. Kincaid. "Customer relationship management: getting it right!." *Prentice Hall Professional*, 2003.
- [2] J.L.H. Bowden. "The process of customer engagement: A conceptual framework." *Journal of marketing theory and practice*, Vol. 17, No. 1, pp. 63-74, 2009.
- [3] P. Patterson, and T.,Yu, and K. De Ruyter. "Understanding customer engagement in services." *Advancing Theory, Maintaining*, 2006.
- [4] C.M. Sashi. "Customer engagement, buyer-seller relationships, and social media." *Management decision*, 2012.
- [5] P.Leeson, and P.C.Heaven. "Social attitudes and personality." *Australian Journal of Psychology*, Vol. 51, No. 1, pp. 19-24, 1999.
- [6] A. Parasuraman, and V.A Zeithaml,A. Malhotra. "ES-QUAL: A multiple-item scale for assessing electronic service quality." *Journal of service research*, Vol. 7, No. 3, pp. 213-233, 2005.
- [7] H. Meerschman, and J. Verkeyn. "Towards a better understanding of service quality attributes of a chatbot." *Doctoral dissertation, Master's dissertation. Ghent University*, 2019.
- [8] S. Brahmam, and A. De Angeli. "Gender affordances of conversational agents." *Interacting with Computers*, Vol. 24, No. 3, pp. 139-153, 2012.
- [9] C. Nass, and Y.Moon. "Machines and mindlessness: Social responses to computers." *Journal of social issues*, Vol. 56, No. 1, pp. 81-103, 2000.
- [10] S. Iwasaki, and P. Ingkaphirom, and I. P. Horie. "A reference grammar of Thai." *Cambridge University Press*, 2005.
- [11] W.Wood, and A.H. Eagly. "Biosocial construction of sex differences and similarities in behavior. In Advances in experimental social psychology." *In Advances in experimental social psychology*, Vol. 46, pp. 55-123, 2012.
- [12] R.M. Montoya, and R.S. Horton. "A meta-analytic investigation of the processes underlying the

- similarity-attraction effect.” *Journal of Social and Personal Relationships*, Vol.30, Vo.1, pp.64-94, 2013.
- [13] L.L. Bove, and D.A Smith. “Relationship strength between a customer and service worker: does gender dyad matter?.” *Services Marketing Quarterly*, vol.27, no.3, pp.17-34, 2006.
- [14] B. Reeves, and C. Nass. “The media equation: How people treat computers, television, and new media like real people.” *Cambridge, UK*, 1996.
- [15] Customer Relationship Management. สืบค้น 10 มีนาคม 2565 เข้าถึงได้จาก <https://nipa.co.th/th/solution/crm>.
- [16] โลกโซเชียลมีเดียของคนไทย ในปี 2564 ที่น่าสนใจสำหรับนักการตลาด. สืบค้น 18 ธันวาคม 2565, เข้าถึงได้จาก <https://marketeeronline.co/archives/209273>.

# ปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยีหุ่นยนต์อัตโนมัติ RPA

## Factors Influencing the Intention to Use Robotic Process Automation Technology

จิราพร เทพทอง (Chiraporn Thepthong)<sup>1</sup> และ ธนพล เจนสุทธิเวชกุล (Tanapon Jensuttiwelchakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858325@email.kmutnb.ac.th, <sup>2</sup>tanapon.j@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยีหุ่นยนต์อัตโนมัติ RPA โดยใช้แบบสอบถามเป็นเครื่องมือในการเก็บรวบรวมข้อมูล ทำการเก็บข้อมูลจากกลุ่มตัวอย่างจำนวน 143 ราย โดยใช้สถิติในการวิเคราะห์ข้อมูล วิเคราะห์องค์ประกอบเชิงสำรวจ และวิเคราะห์ถดถอยพหุคูณแบบขั้นตอน

ผลการศึกษาพบว่าปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยีหุ่นยนต์อัตโนมัติ RPA ของผู้ใช้งาน คือ ปัจจัยด้านการรับรู้ถึงประโยชน์ ปัจจัยด้านความเหมาะสมระหว่างงานและเทคโนโลยี และปัจจัยด้านความน่าเชื่อถือ ซึ่งงานวิจัยนี้มีค่าที่ยอมรับได้ 30 % ( $R - Squared = 0.30$ )

**คำสำคัญ:** เทคโนโลยีหุ่นยนต์อัตโนมัติ RPA

### Abstract

The objective of this research is to explore factors influencing the intention to use Robotic Process Automation Technology (RPA). Data were collected by using survey from 143 respondents. The statistics used in this research were Exploratory Factor Analysis, and multiple regression analysis.

The result indicated that Factors Influencing the Intention to Use Robotic Process Automation Technology are Perceived of Use, Task Technology fit and Reliability these factors can explain the dependent variable about 30% ( $R - Squared = 0.30$ )

**Keywords:** Robotic Process Automation Technology.

### 1. บทนำ

เนื่องจากเพื่อให้ธุรกิจสามารถดำเนินการแข่งขันกับคู่แข่งได้ การปรับตัวและนำเทคโนโลยีใหม่ๆ มาใช้เพื่อให้การทำงานมีความคล่องตัว ลดข้อผิดพลาด เพิ่มประสิทธิภาพ ลดต้นทุน สนับสนุนการทำงานของมนุษย์ จึงมีการนำแรงงานดิจิทัล Robotic Process Automation (RPA) หรือการทำกระบวนการอัตโนมัติด้วยหุ่นยนต์ซอฟต์แวร์เข้ามาทำงานแทนมนุษย์มากขึ้น โดย RPA สามารถช่วยให้การทำงานมีประสิทธิภาพ และได้ปริมาณงานที่เพิ่มมากขึ้นในเวลาที่สั้นลง ช่วยจัดการงานที่ซ้ำซากในหลายอุตสาหกรรม ทั้งยังสามารถทำงานได้เร็วกว่าแรงงานมนุษย์ถึง 40-60% มีค่าความผิดพลาดเกือบเป็นศูนย์ และด้วยการทำงานที่ถูกต้องแม่นยำตามที่กำหนดได้ 100% ไม่ใช้อารมณ์ในการทำงาน ไม่มีการขาด ลา มาสาย ที่สำคัญคือช่วยลดต้นทุนการดำเนินงาน และคุ้มค่าต่อการลงทุนทำให้ผู้ประกอบการหันมาใช้งานเทคโนโลยี RPA กันมากขึ้น

ด้วยสภาพเศรษฐกิจในปัจจุบันที่ผู้ประกอบการต้องการลดต้นทุน เพิ่มศักยภาพ สร้างคุณภาพที่ดี ให้กับองค์กร เพื่อให้เกิดความได้เปรียบทางการแข่งขัน ผู้วิจัยจึงสนใจที่จะศึกษาปัจจัยที่มีอิทธิพลต่อการยอมรับเทคโนโลยีหุ่นยนต์กระบวนการอัตโนมัติ RPA เพื่อเป็นประโยชน์กับผู้ประกอบการที่ให้บริการด้านการติดตั้ง และพัฒนาระบบ RPA ในการวางแผนบริหารจัดการการ Implement ให้สำเร็จลุล่วงตามวัตถุประสงค์อย่างราบรื่น เนื่องจากการ Implement ระบบ RPA จะสำเร็จได้ด้วยดีจะต้องอาศัยความร่วมมือของพนักงานในการอธิบายกระบวนการ

ทำงานปัจจุบันพร้อมให้เงื่อนไขการตัดสินใจโดยละเอียด และการเตรียมข้อมูลตามเทมเพลต เพื่อให้ Robot สามารถทำงานแทนมนุษย์ในส่วนนั้นได้ โดยความเต็มใจที่จะร่วมมือ และความตั้งใจใช้งานเทคโนโลยี RPA นั้น จะเป็นปัจจัยสำคัญที่ทำให้การดำเนินงานสำเร็จ หากทีมผู้พัฒนาและผู้ประกอบการมีแผนการรับมือที่ดี ก็จะช่วยให้พนักงานเกิดความตั้งใจใช้เทคโนโลยีได้อย่างมีประสิทธิภาพ และเกิดประโยชน์สูงสุด

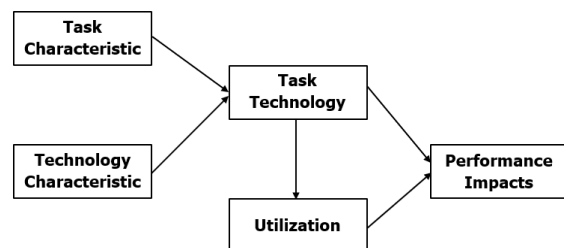
## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 Robotic Process Automation (RPA)

Robotic Process Automation (RPA) หรือเทคโนโลยีหุ่นยนต์กระบวนการอัตโนมัติ คือ ซอฟต์แวร์โรบอทที่เลียนแบบพฤติกรรมของมนุษย์ โดยที่มนุษย์เป็นผู้ออกแบบกระบวนการ (Process) และขั้นตอน (Workflow) การตัดสินใจต่าง ๆ เพื่อให้ซอฟต์แวร์โรบอทสามารถทำงานซ้ำ ๆ แทนมนุษย์ได้อย่างมีประสิทธิภาพ ลดข้อผิดพลาดในการทำงานที่เกิดจาก Human error และยังสามารถทำงานได้ตลอด 24 ชั่วโมง หากงานที่เป็นลักษณะงานที่ต้องทำซ้ำ ๆ เป็นประจำ มีปริมาณมาก และต้องใช้เวลามากในการทำให้เสร็จ ทำให้ต้องเสียเวลากับงานเหล่านั้น การนำเทคโนโลยี RPA มาเป็นเครื่องมือช่วยทำกระบวนการซ้ำ ๆ โดยอัตโนมัติจะสามารถลดเวลาการทำงาน โดยได้ผลของงานที่มากขึ้น ด้วยข้อผิดพลาดที่น้อยลงจึงมีส่วนช่วยในการเพิ่มประสิทธิภาพการทำงาน ลดเวลา และต้นทุนในการดำเนินงานได้ งานที่เหมาะสมจะนำระบบ RPA มาใช้ เช่น งานแยกไฟล์แนบจาก Email เพื่อจัดเก็บไว้ตามโฟลเดอร์, งานที่ต้องทำซ้ำ ๆ ใน Excel เช่น การแยกไฟล์ลูกค้าใหม่ ออกเป็นรายจังหวัด เพื่อส่งให้กับ Sales ในพื้นที่นั้น ๆ, การนำไฟล์ที่อยู่ในโฟลเดอร์ส่ง Email อัตโนมัติ ให้กับคนที่เกี่ยวข้อง, งาน Copy และ Paste ไม่ว่าจะ เป็น ใน Application หรือ ระหว่าง Application, เก็บข้อมูลสถิติของ Social Media, การเก็บข้อมูลจากเว็บไซต์ (Web Scraping) เป็นต้น

### 2.2 ทฤษฎีความเหมาะสมของงานและลักษณะของเทคโนโลยี

ทฤษฎีความเหมาะสมของงานและลักษณะของเทคโนโลยี Task Technology Fit หรือ TTF นำเสนอโดย Goodhue and Thompson เป็นแนวคิดที่ใช้สำหรับประเมินประสิทธิภาพ ของเทคโนโลยีที่ส่งผลกระทบต่อการใช้งาน และอธิบายความเหมาะสมระหว่างลักษณะของงาน และลักษณะของเทคโนโลยี โดยการที่ผู้ใช้งานจะยอมรับ เทคโนโลยีนั้น ๆ ได้ ก็ต่อเมื่อ เทคโนโลยีนั้นมีฟังก์ชันที่เหมาะสม และสนับสนุนผู้ใช้งานสามารถทำงานได้อย่างมีประสิทธิภาพ และเกิดประโยชน์สูงสุด เทคโนโลยีที่ไม่มีประโยชน์เพียงพอจะไม่ถูกนำมาใช้งาน โดย TTF ประกอบด้วยปัจจัยด้าน 2 ปัจจัย คือคุณลักษณะของงาน (Task Characteristics) และคุณลักษณะของเทคโนโลยี (Technology Characteristics) และความเหมาะสมของงาน และเทคโนโลยี (Task Technology Fit) จะส่งต่อพฤติกรรมกระตุ้นใจ ใช้เทคโนโลยีให้เป็นประโยชน์ (Utilization) และส่งผลกระทบต่อสัภาพของงาน (Performance Impacts)

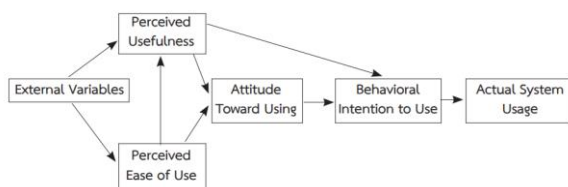


ภาพที่ 1: โมเดล Task – Technology Fit (TTF) [3]

### 2.3 ทฤษฎีการยอมรับเทคโนโลยี

ทฤษฎีการยอมรับเทคโนโลยี Technology Acceptance Model หรือ TAM เป็นทฤษฎีที่มีการยอมรับและมีชื่อเสียงในการเป็นตัวชี้วัดความสำเร็จของการใช้เทคโนโลยี นำเสนอโดย Davis, Bagozzi & Warshaw ซึ่งเป็นการพัฒนาเพิ่มเติมมาจากทฤษฎีการกระทำตามหลักเหตุและผล (The Theory of Reasoned Action) หรือ TRA ที่ใช้ในการศึกษาและอธิบายพฤติกรรมว่าบุคคลหนึ่ง จะมีการยอมรับเทคโนโลยีได้เกิดจากปัจจัยหลัก 2 ประการได้แก่ 1) การรับรู้ถึงประโยชน์ (Perceived Usefulness) และ 2) การรับรู้ความง่ายในการใช้งาน (Perceived Ease of Use) เนื่องจากการรับรู้ประโยชน์ และการรับรู้ความง่าย ในการใช้งานมี

ความสัมพันธ์กับพฤติกรรม การยอมรับเทคโนโลยี และมีความเชื่อมโยงกับทัศนคติต่อการใช้เทคโนโลยี และทัศนคติที่มีต่อการใช้งานเทคโนโลยีของบุคคลใด บุคคลหนึ่ง มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยีนั้น ซึ่งความตั้งใจแสดงพฤติกรรม ของบุคคลหนึ่งเป็นการแสดงออกตามทัศนคติของบุคคลนั้น ที่ส่งผลต่อการยอมรับเทคโนโลยี และใช้ศึกษาในบริบทการยอมรับการใช้ระบบสารสนเทศ



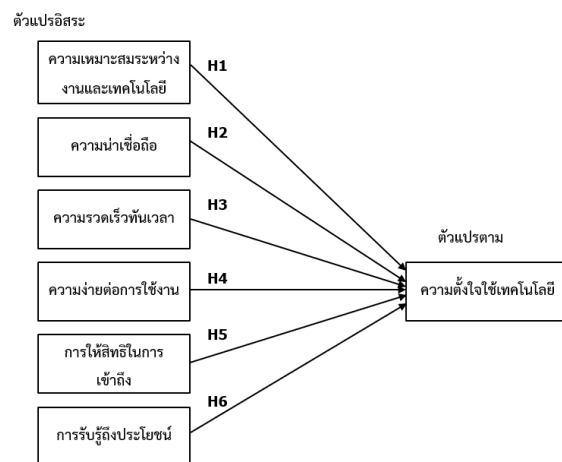
ภาพที่ 2: แบบจำลองการยอมรับเทคโนโลยี TAM [7]

จากภาพที่ 2 ตัวแปรภายนอก เช่น ข้อมูลประชากรศาสตร์ (Demographic) ประสบการณ์ เป็นต้น มีอิทธิพลต่อการรับรู้ถึงประโยชน์ที่จะได้รับจากเทคโนโลยีสารสนเทศ และการรับรู้ว่าเป็นระบบที่ง่ายต่อการใช้งานการรับรู้ถึงประโยชน์ที่ได้รับจากเทคโนโลยีสารสนเทศคือปัจจัยที่กำหนดการรับรู้ในแต่ละบุคคลว่าเทคโนโลยีสารสนเทศมีส่วนช่วยพัฒนาประสิทธิภาพการปฏิบัติงานได้อย่างไร และเป็นปัจจัยที่ส่งผล โดยตรงต่อความตั้งใจแสดงพฤติกรรม การใช้ด้วยการรับรู้ว่าเป็นระบบที่ง่ายต่อการใช้งานคือปัจจัยที่กำหนดในแง่ปริมาณหรือความสำเร็จที่ได้รับว่าตรงกับความต้องการหรือที่คาดหวังไว้หรือไม่ ซึ่งเป็นปัจจัยที่ส่งผลต่อการรับรู้ถึงประโยชน์ที่ได้รับจากเทคโนโลยีสารสนเทศด้วย

**2.4 การกำหนดกรอบแนวคิดการวิจัย**

จากศึกษาค้นคว้าทฤษฎีและงานวิจัยเกี่ยวกับปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยี ทฤษฎีการยอมรับเทคโนโลยี Technology Acceptance Model (TAM) และทฤษฎีความเหมาะสมของงานและลักษณะของเทคโนโลยี Task Technology Fit (TTF) ซึ่งทฤษฎีการยอมรับเทคโนโลยี ทฤษฎี (TAM) มีจุดประสงค์คือการอธิบาย และคาดการณ์การยอมรับระบบสารสนเทศ ของผู้ใช้งาน โดยมีสมมติฐานว่าการยอมรับเทคโนโลยีของแต่ละบุคคลนั้นมาจากปัจจัยหลัก 2 ประการคือ 1. การรับรู้ถึง

ประโยชน์ที่ได้รับ (Perceived usefulness) เนื่องจากการรับรู้ประโยชน์ และการรับรู้ความง่ายในการใช้งานมีความสัมพันธ์กับพฤติกรรมการยอมรับเทคโนโลยี และมีความเชื่อมโยงกับทัศนคติต่อการใช้เทคโนโลยี และทัศนคติที่มีต่อการใช้งาน มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยีนั้น และได้ศึกษาทฤษฎีความเหมาะสมของงานและลักษณะของเทคโนโลยี (TTF) มีจุดประสงค์คือการอธิบายว่าการที่ผู้ใช้งานจะยอมรับเทคโนโลยีก็ต่อเมื่อเทคโนโลยีนั้น มีฟังก์ชันที่เหมาะสม และสนับสนุนผู้ใช้ให้สามารถทำงาน ได้อย่างมีประสิทธิภาพ และเกิดประโยชน์สูงสุด ประกอบด้วยปัจจัยด้าน 2 ปัจจัย คือ คุณลักษณะของงาน (Task Characteristics) และคุณลักษณะของเทคโนโลยี (Technology Characteristics) และความเหมาะสมของงาน และเทคโนโลยียังส่งผลต่อพฤติกรรมการใช้เทคโนโลยี ให้เป็นประโยชน์ (Utilization) และส่งผลกระทบต่อศักยภาพของงาน (Performance Impacts) และจากการศึกษางานวิจัยต่าง ๆ ทำให้ทราบว่าปัจจัยภายนอกอื่น ๆ ที่เกี่ยวข้อง คือ ความน่าเชื่อถือของเทคโนโลยี ความรวดเร็วทันเวลา และการให้สิทธิ์การเข้าถึงเทคโนโลยี ทำให้ได้กรอบแนวคิด ดังนี้



ภาพที่ 3: การกำหนดกรอบแนวคิด

**3. วิธีดำเนินงานวิจัย**

ขั้นตอนการดำเนินงานของการวิจัย มีวัตถุประสงค์เพื่อวิเคราะห์หาปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยีหุ่นยนต์อัตโนมัติ RPA ซึ่งผู้วิจัยได้ศึกษา แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องต่างๆ เน้นการศึกษาคำอธิบายเชิง

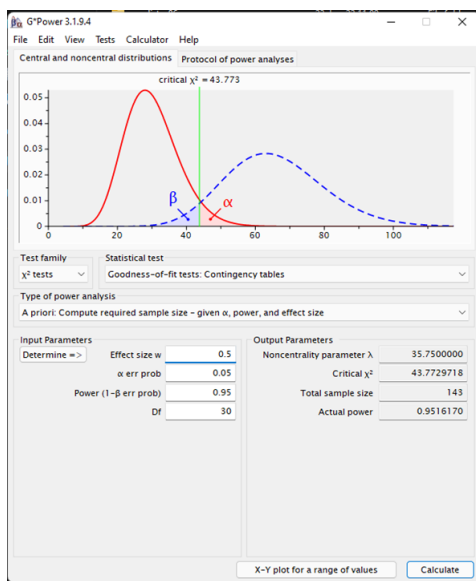


ปริมาณ (Quantitative-Research) เก็บรวบรวมความคิดเห็น ด้วยเครื่องมือแบบสอบถามและนำมาวิเคราะห์ด้วยวิธีการทางสถิติ โดยผู้วิจัยได้แบ่งขั้นตอนการดำเนินงานดังนี้

**3.1 ประชากรและกลุ่มตัวอย่าง**

3.1.1 ประชากรที่ใช้ในการวิจัย ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็น ประชากรผู้ใช้งานระบบ RPA ในชีวิตประจำวัน

3.1.2 กลุ่มตัวอย่าง ที่ใช้ในการวิจัยครั้งนี้ คือ กลุ่มตัวอย่างผู้ใช้งานระบบ RPA ที่ได้จากการสุ่ม โดยวิธีการสุ่มอย่างง่าย การคำนวณขนาดตัวอย่างใช้โปรแกรมสำเร็จรูป G\*Power สำหรับการทดสอบ Chi-square การทดสอบข้อมูลตารางการแจกแจง (contingency table) ด้วย Chi-square test ประกอบ ด้วย Goodness of fit test, Independence test (association), Chi-square Degree of Freedom (df) และ Conventional effect size = 0.50 จะได้ขนาดตัวอย่าง 143 ตัวอย่าง



ภาพที่ 4: ผลคำนวณขนาดตัวอย่างด้วย G\*Power

**3.2 เครื่องมือที่ใช้ในการวิจัย**

เครื่องมือที่ใช้ในการศึกษาการวิจัยคือแบบสอบถามในการเก็บรวบรวมข้อมูล แบบสอบถามแบ่งออกเป็น 3 ส่วน ดังนี้ ส่วนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม ส่วนที่ 2 ปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยี RPA และ ส่วนที่ 3 ความตั้งใจใช้งานเทคโนโลยี RPA โดยใช้การวัดระดับข้อมูลประเภทอันตรภาคชั้น (Interval Scale) วัด

ระดับความคิดเห็นเป็นมาตรฐานประมาณค่า (Rating Scales) เป็นการวัดแบบ Likert Scale มี 5 ระดับ

**3.3 การดำเนินและเก็บรวบรวมข้อมูล**

3.3.1 สร้างเครื่องมือในการเก็บรวบรวมข้อมูล ผู้วิจัยใช้ Google Form ในการสร้างแบบสอบถามออนไลน์ ซึ่งเป็นเครื่องมือในการเก็บรวบรวมข้อมูลที่มีความนิยม และมีความสะดวกในการส่งให้ผู้ใช้งานทำแบบประเมิน เพื่อสำรวจความเห็นปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยี RPA

3.3.2 เก็บรวบรวมข้อมูล ผู้วิจัยได้ทำการกระจายแบบสอบถาม ปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยี RPA ผ่านทาง Email, Facebook และ LINE Application ในการส่งแบบสอบถามออนไลน์ให้กับผู้ใช้งานเพื่อทำการเก็บรวบรวมข้อมูล

**4. ผลการดำเนินงานวิจัย**

การวิจัยครั้งนี้เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยี RPA ผู้วิจัยได้จัดทำแบบสอบถามเพื่อเป็นเครื่องมือในการรวบรวมข้อมูลจากกลุ่มตัวอย่างจำนวน 143 คน และนำมาวิเคราะห์ด้วยโปรแกรม SPSS โดยผู้วิจัยแบ่งการวิจัยได้เป็นหัวข้อดังต่อไปนี้

**4.1 ผลจากการวิเคราะห์ข้อมูลทั่วไป**

การวิเคราะห์ข้อมูลทั่วไปของกลุ่มตัวอย่างจำนวน 143 คน โดยใช้สถิติเชิงพรรณนา ซึ่งได้แก่ วิเคราะห์ข้อมูลโดยใช้การแจกแจงความถี่ และนำเสนอเป็นคำร้อยละ ผลการวิเคราะห์ข้อมูลดังนี้

ตารางที่ 4-1 ผลการวิเคราะห์ข้อมูลทั่วไป

ข้อความ	ข้อมูลทั่วไป	จำนวน (คน)	ร้อยละ (%)
อายุ	• 21 – 25 ปี	24	16.78
	• 26 – 30 ปี	59	41.26
	• 31 – 40 ปี	46	32.17
	• 41 – 45 ปี	11	7.69
	• มากกว่า 46 ปี	3	2.10
<b>รวม</b>		<b>143</b>	<b>100.00</b>
ประสบการณ์ทำงาน	• น้อยกว่า 1 ปี	9	6.92
	• 1 - 5 ปี	37	25.87

	• 6 - 10 ปี	59	41.26
	• มากกว่า 10 ปี	38	26.57
	<b>รวม</b>	<b>143</b>	<b>100.00</b>
ประเภทธุรกิจ	• ธุรกิจโลจิสติกส์	17	11.89
	• ธุรกิจบริการ	24	16.78
	• ธุรกิจอาหาร	19	13.29
	• ธุรกิจการผลิต	23	16.08
	• ธุรกิจน้ำมัน	16	11.19
	• ธุรกิจซื้อขายไป	40	27.97
	• อื่นๆ (ระบุ).....	4	2.80
	<b>รวม</b>	<b>143</b>	<b>100.00</b>

**4.2 ผลการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis)**

**KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.879
Bartlett's Test of Sphericity	Approx. Chi-Square 1494.150
df	253
Sig.	<.001

ภาพที่ 5: KMO และ Bartlett's Test

ตารางที่ 4-2 ผลการวิเคราะห์ Factor Matrix ครั้งที่ 2

รหัส	Factor					
	1	2	3	4	5	6
TTF1	0.651					
TTF2	0.672					
TTF3	0.768					
TTF4	0.609					
CER1		0.656				
CER2		0.603				
CER3		0.532				
CER4		0.636				
EOU1			0.676			
EOU2			0.620			
EOU3			0.629			
ACR1					0.724	
ACR2					0.824	
POU2				0.626		
POU3				0.663		
POU5				0.577		
IOU1						0.785
IOU2						0.591

จากตารางที่ 4-2 การวิเคราะห์ Factor Matrix ครั้งที่ 2 แสดงค่าสัมประสิทธิ์ ซึ่งเป็นค่าที่แสดงความสัมพันธ์ของตัวแปรกับองค์ประกอบ โดยการหมุนแกนปัจจัยด้วยเทคนิค Principal axis factoring Analysis หลังจากตัดตัวแปร INT1, INT2, POU1, POU4 และ IOU3 แล้วได้ค่า Factor 6 ตัวแปร ได้แก่ 1. TTF หมายถึง ตัวแปรปัจจัยด้านความเหมาะสมระหว่างงานและเทคโนโลยีของระบบ RPA 2. CER หมายถึง ตัวแปรปัจจัยด้านความน่าเชื่อถือของระบบ RPA 3. POU หมายถึง ตัวแปรปัจจัยด้านการ

รับรู้ถึงประโยชน์ของระบบ RPA 4. IOU หมายถึงตัวแปรความตั้งใจใช้งานระบบ RPA 5. EOU ตัวแปรปัจจัยด้านความง่ายต่อการใช้งานระบบ RPA และ 6. ACR หมายถึงตัวแปรปัจจัยด้านการให้สิทธิในการเข้าถึง

**4.3 การวิเคราะห์ถดถอยพหุคูณแบบขั้นตอน (Stepwise multiple regression analysis)**

เป็นการวิเคราะห์เพื่อทดสอบว่าตัวแปรอิสระทุกตัว จะไม่มีความสัมพันธ์กันเองสูง เพื่อป้องกันปัญหาความสัมพันธ์ภายในของตัวแปรพยากรณ์

Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
		B	Std. Error			
1	(Constant)	1.984	.374		5.251	<.001
	ปัจจัยด้านกำลังคนที่มีอยู่ในระบบ	.558	.090	.461	6.181	<.001
	(Constant)	1.337	.400		3.341	.001
2	ปัจจัยด้านกำลังคนที่มีอยู่ในระบบ	.418	.095	.345	4.376	<.001
	ปัจจัยด้านความสัมพันธ์ระหว่างงานและเทคโนโลยี	.297	.084	.281	3.560	<.001
	(Constant)	.975	.430		2.287	.025
3	ปัจจัยด้านกำลังคนที่มีอยู่ในระบบ	.349	.099	.290	3.527	<.001
	ปัจจัยด้านความสัมพันธ์ระหว่างงานและเทคโนโลยี	.226	.089	.213	2.532	.012
	ปัจจัยด้านความน่าเชื่อถือ	.228	.107	.183	2.129	.035

a. Dependent Variable: ความตั้งใจใช้งานระบบ RPA

ภาพที่ 6: Dependent Variable ความตั้งใจใช้งานระบบ RPA

จากรูป พบว่าการวิเคราะห์ถดถอยพหุคูณแบบขั้นตอนมีลำดับขั้นที่ทำการเลือกตัวแปรเข้าสู่สมการอยู่ 3 ขั้น โดยลำดับแรกคือ ปัจจัยด้านการรับรู้ถึงประโยชน์ ปัจจัยด้านความเหมาะสมระหว่างงานและเทคโนโลยี และปัจจัยด้านความน่าเชื่อถือ เป็นตัวแปรที่ส่งผลต่อความตั้งใจใช้ระบบ RPA โดยมีค่า Sig น้อยกว่า 0.05 มีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ เท่ากับ 0.548 และค่าสัมประสิทธิ์สหสัมพันธ์ความตั้งใจใช้งานระบบ RPA เท่ากับ 0.300 หรือร้อยละ 30

**5. สรุปผล อภิปรายผล และข้อเสนอแนะ**

**5.1 สรุปผลการดำเนินงานวิจัย**

จากการศึกษา รวบรวม และวิเคราะห์ข้อมูลโดยสำรวจจากแบบสอบถามจำนวน 143 ชุด โดยมีจุดประสงค์ในการทำวิจัยเพื่อศึกษาปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยีหุ่นยนต์กระบวนการอัตโนมัติ RPA ได้ทำการตรวจสอบความถูกต้องจากปัจจัยที่เป็นตัวแปรอิสระทั้งหมด 6 ปัจจัย มีเพียง 3 ปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยีหุ่นยนต์กระบวนการอัตโนมัติ RPA โดยเรียงจากองค์ประกอบที่มีความสำคัญมากที่สุดดังนี้ 1. ปัจจัยด้านการรับรู้ถึงประโยชน์ของเทคโนโลยี RPA 2. ปัจจัย

ด้านความเหมาะสมระหว่างงานและเทคโนโลยี RPA และ

### 3. ปัจจัยด้านความน่าเชื่อถือ

#### 5.2 อภิปรายผลการวิจัย

จากผลการวิเคราะห์ถดถอยพหุคูณแบบขั้นตอน และสมมติฐานที่ผู้วิจัยได้ตั้งสมมติฐานไว้ในเบื้องต้นผลปรากฏว่าปัจจัยที่มีอิทธิพลต่อความตั้งใจใช้เทคโนโลยี RPA มากที่สุดได้แก่ ปัจจัยด้านการรับรู้ถึงประโยชน์ ซึ่งสอดคล้องกับงานวิจัยของศศิจันทร์ ปัญจทวี (2560) ซึ่งได้ผลการวิจัยตรงกันว่า การรับรู้ถึงประโยชน์มีอิทธิพลต่อการยอมรับเทคโนโลยีของบุคคล สอดคล้องกับทฤษฎีการยอมรับเทคโนโลยี TAM ลำดับต่อมา คือ ปัจจัยด้านความเหมาะสมระหว่างงานและเทคโนโลยี ซึ่งสอดคล้องกับงานวิจัยของปรัชญา ชันธิยม (2557) ซึ่งได้ผลการวิจัยตรงกันว่าความเหมาะสมระหว่างลักษณะของงานและลักษณะเทคโนโลยี (TTF) มีอิทธิพลต่อการความตั้งใจใช้งานเทคโนโลยี และลำดับที่ 3 คือปัจจัยด้านความน่าเชื่อถือ ซึ่งตรงกับงานวิจัยของสุวิทย์ สมสุภาพรุ่งยศ, แสงทอง บุญยิ่ง และ อเนก พุทธิเดช (2563) พบว่า ความเหมาะสมของระบบสารสนเทศที่ใช้ปฏิบัติงานมีปัจจัยที่ส่งผลต่อความตั้งใจใช้เทคโนโลยี ได้แก่ ความน่าเชื่อถือของระบบ ทั้ง 3 ปัจจัยเป็นตัวแปรต้นที่ส่งผลต่อความตั้งใจใช้เทคโนโลยี RPA อย่างมีนัยสำคัญที่ระดับ 0.05

#### เอกสารอ้างอิง

- [1] Vante Zira Vanduhe, Muesser Nat and Hasan Fahmi Hasan. (2020). [ออนไลน์]. Continuance Intentions to Use Gamification for Training in Higher Education: Integrating the Technology Acceptance Model (TAM), Social Motivation, and Task Technology Fit (TTF). [สืบค้นวันที่ 6 สิงหาคม 2564]. จาก <https://ieeexplore.ieee.org/document/8957143>
- [2] Ibrahim Youssef Alyoussef. (2021). [ออนไลน์]. E-Learning Acceptance: The Role of Task-Technology Fit as Sustainability in Higher Education. [สืบค้นวันที่ 30 กรกฎาคม 2564]. <https://www.mdpi.com/2071-1050/13/11/6450>
- [3] Ibrahim Youssef Alyoussef. (2021). [ออนไลน์]. Massive Open Online Course (MOOCs) Acceptance: The Role of Task-Technology Fit (TTF) for Higher Education Sustainability. [สืบค้นวันที่ 13 สิงหาคม 2564]. <https://www.mdpi.com/2071-1050/13/13/7374>
- [4] Sayyed Khawar Abbas, Hafiz Ali Hassan, Jawad Asif, Bilal Ahmed, Fahad Hassan, Syed Salman Haider. (2018). [ออนไลน์]. Integration of TTF, UTAUT, and ITM for mobile Banking Adoption. [สืบค้นวันที่ 30 กรกฎาคม 2564]. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3509727](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3509727).
- [5] ชัชฎาภรณ์ เดชาเสถียร. (2557). [ออนไลน์] ปัจจัยที่ส่งผลต่อความพึงพอใจของผู้ใช้บริการ KTB Netbank ในเขตกรุงเทพมหานคร. [สืบค้นวันที่ 29 กรกฎาคม 2564]. จาก [http://dspace.bu.ac.th/bitstream/123456789/1296/1/c\\_hatchadaporn\\_daec.pdf](http://dspace.bu.ac.th/bitstream/123456789/1296/1/c_hatchadaporn_daec.pdf)
- [6] ณัฐธรา ธิรโสภี และชัชวรัตน์ อุดตมากร. (2562). [ออนไลน์]. ปัจจัยที่มีอิทธิพลต่อความตั้งใจเชิงพฤติกรรมในการใช้เทคโนโลยีเว็บพอร์ทัลของการเรียนการสอนในระบบเปิด Thai MOOC. [สืบค้นวันที่ 29 กรกฎาคม 2564]. จาก <https://so05.tci-thaijo.org/index.php/sujthai/article/view/181174>.
- [7] เมทินี จันทร์กระแจะ. (2558). [ออนไลน์]. คุณภาพของระบบสารสนเทศและการยอมรับเทคโนโลยีที่ส่งผลต่อการตัดสินใจใช้บริการธนาคารผ่านโทรศัพท์เคลื่อนที่ของประชาชน ในกรุงเทพมหานคร. [สืบค้นวันที่ 29 กรกฎาคม 2564]. จาก <http://dspace.bu.ac.th/handle/123456789/2464165778/>
- [8] ปรัชญา ชันธิยม. (2557). [ออนไลน์] ชื่อนำและผลของการใช้เทคโนโลยีการทำงานร่วมกันผ่านสื่ออิเล็กทรอนิกส์ในสถาบันการศึกษา. [สืบค้นวันที่ 29 กรกฎาคม 2564]. จาก <https://he02.tci-thaijo.org/index.php/RHPC9Journal/article/download/242473/165778/>
- [9] ศศิจันทร์ ปัญจทวี. (2560). [ออนไลน์]. ปัจจัยที่ส่งผลต่อการยอมรับการใช้ระบบสารสนเทศ กรณีศึกษา สถาบันการพลศึกษา วิทยาเขตเชียงใหม่. [สืบค้นวันที่ 9 สิงหาคม 2564]. จาก [http://www.graduate.cmru.ac.th/core/km\\_file/365.pdf](http://www.graduate.cmru.ac.th/core/km_file/365.pdf)
- [10] สุวิทย์ สมสุภาพรุ่งยศ, แสงทอง บุญยิ่ง และ อเนก พุทธิเดช. (2563). [ออนไลน์]. ปัจจัยที่เหมาะสมของระบบสารสนเทศที่ใช้ในการปฏิบัติงานของหน่วยงาน. [สืบค้นวันที่ 28 กรกฎาคม 2564]. จาก <https://ph02.tci-thaijo.Org/index.php/itm-journal/article/download/122028/129993/>

# ปัจจัยที่มีอิทธิพลต่อการยอมรับเทคโนโลยี แอปพลิเคชันติดตามผู้ป่วย

## Factors Influencing Patient Tracking Application Acceptance

นวกาญจน์ นิลประวิทย์ (Nawakarn Ninprawit)<sup>1</sup> และ ธนพล เจนสุทธิเวชกุล (Tanapon Jensuttiwetachakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858350@email.kmutnb.ac.th, <sup>2</sup>tanapon.j@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อการยอมรับเทคโนโลยีแอปพลิเคชันติดตามผู้ป่วยบนระบบปฏิบัติการแอนดรอยด์ เพื่อส่งเสริมการใช้งานเทคโนโลยีเพื่อสุขภาพโดยประยุกต์ใช้แบบจำลองการยอมรับเทคโนโลยี (Technology Acceptance Model หรือ TAM) โดยใช้แบบสอบถามเก็บข้อมูลกับกลุ่มประชากรตัวอย่างที่มีอายุ 20 – 70 ปี จำนวน 350 ตัวอย่าง

ผลการวิจัยพบว่าปัจจัยที่มีอิทธิพลต่อการยอมรับเทคโนโลยีแอปพลิเคชันติดตามผู้ป่วยประกอบด้วย โมเดลที่ 1 : การรับรู้ว่าการใช้เทคโนโลยีนั้นใช้งานง่าย (Perceived Ease of Use (PE)) กับ การรับรู้ถึงประโยชน์จากการใช้งาน (Perceived Usefulness (PU)) ส่งผลต่อปัจจัยทัศนคติที่มีต่อการใช้งานระบบ (Attitude Toward Using (A)) โดยอธิบายได้ร้อยละ 65 ( $R^2 = .651$ ) และ โมเดลที่ 2: การรับรู้ว่าการใช้เทคโนโลยีนั้นใช้งานง่าย (Perceived Ease of Use (PE)), การรับรู้ถึงประโยชน์จากการใช้งาน (Perceived Usefulness (PU)) และ ทัศนคติที่มีต่อการใช้งานระบบ (Attitude Toward Using (A)) ส่งผลต่อปัจจัยความตั้งใจและเจตนาที่มีต่อการใช้งานระบบ (Behavioral Intention to Use (BI)) โดยอธิบายได้ร้อยละ 57 ( $R^2 = .575$ )

**คำสำคัญ:** การยอมรับเทคโนโลยี ปัจจัยการรับรู้ว่าการใช้เทคโนโลยีนั้นใช้งานง่าย ปัจจัยการรับรู้ถึงประโยชน์จากการใช้งาน ทัศนคติที่มีต่อการใช้งานระบบ

### Abstract

The purpose of this research was to examine the factors influencing the adoption of patient tracking

application technology on Android operating system. To promote the use of technology for health by applying the Technology Acceptance Model (TAM) by using a questionnaire to collect data with a sample population aged 20 - 70 years, 350 samples.

The findings revealed that the influencing factors influencing the adoption of patient tracking application technology were Model 1: Perceived Ease of Use (PE) versus perceived benefit. Perceived Usefulness (PU) affects Attitude Toward Using (A) by 65 percent ( $R^2 = .651$ ) and Model 2: Perceived. that the technology is Perceived Ease of Use (PE), Perceived Usefulness (PU) and Attitude Toward Using (A). The effect on the intention and intent factors on the use of the system (Behavioral Intention to Use (BI)) was described by 57 percent ( $R^2 = .575$ ).

**Keywords:** Technology Acceptance Model, Perceived Ease of Use, Perceived Usefulness, Attitude Toward Using.

### 1. บทนำ

เนื่องจากในปัจจุบันประเทศไทยมีผู้ป่วยที่ป่วยเป็นโรคนิ่วในระบบทางเดินปัสสาวะ (Urolithiasis) เป็นจำนวนมากจากข้อมูลสถิติของกระทรวงสาธารณสุขพบว่าอัตราการเกิดโรคนิ่วปี พ.ศ. 2551 เท่ากับ 108.41 ต่อประชากร 100,000 คน ในปี พ.ศ. 2552 เท่ากับ 107.75 ในปี พ.ศ. 2553 เท่ากับ 118.97 ต่อประชากร 100,000 คน และในปี พ.ศ. 2553 เท่ากับ 122.46 ต่อประชากร 100,000 คน และมีแนวโน้มจะเพิ่มมากขึ้น ดังนั้นการให้บริการกับผู้ป่วยโรงพยาบาลเลือกใช้วิธีการนัดแบบบัตรนัดกระดาษซึ่ง

เจ้าหน้าที่และทีมแพทย์จะต้องการสร้างบัตรนัดและโทรตามเมื่อใกล้ถึงวันนัด ปัญหาที่พบในปัจจุบันกับกลุ่มผู้ป่วยที่ต้องใช้วิธีการรักษาแบบใส่สายสวนในท่อไต (Cystoscope c double j stent insertion) คือ บางครั้งผู้ป่วยทำบัตรนัดหาย หรือ ลืมวันไม่มาตามนัดเพื่อเอาสายสวนในท่อไตออก ซึ่งพบมากในกลุ่มผู้สูงอายุตลอดไปจนถึงกลุ่มผู้ป่วยวัยทำงาน การไม่มาตามนัดและไม่ปฏิบัติตามคำแนะนำของแพทย์ ส่งผลให้เกิดอันตรายกับผู้ป่วยและอาจเกิดภาวะแทรกซ้อนขึ้นภายหลังตลอดขั้นตอนการรักษา

ปัจจุบันความก้าวหน้าทางเทคโนโลยีค่อนข้างพัฒนาไปได้ไกลและมีเทคโนโลยีที่หลากหลายให้นักพัฒนาซอฟต์แวร์ได้เลือกใช้ เพื่อเพิ่มประสิทธิภาพในการพัฒนาซอฟต์แวร์ด้านสุขภาพ (eHealth) เครื่องมือและบริการทุกชนิดที่ใช้เทคโนโลยีสารสนเทศและการสื่อสาร (ICT) ในการดูแลสุขภาพที่เชื่อมโยงระหว่างผู้ป่วยและผู้ให้บริการด้านสุขภาพกับโรงพยาบาล ตามแผนพัฒนายุทธศาสตร์เทคโนโลยีสารสนเทศสุขภาพ (eHealth strategy) ของกระทรวงสาธารณสุข

ดังนั้นผู้วิจัยจึงมีความสนใจที่จะนำเอาแบบจำลองการยอมรับเทคโนโลยี (Technology Acceptance Model หรือ TAM) มาช่วยในการออกแบบและศึกษาปัจจัยที่ส่งผลต่อการยอมรับเทคโนโลยีแอปพลิเคชันติดตามผู้ป่วยบนระบบปฏิบัติการแอนดรอยด์แทนวิธีการใช้บัตรนัดแบบกระดาษ

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

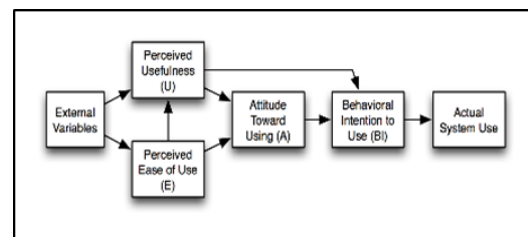
### 2.1 Mobile Health หรือ mHealth [5]

Mobile Health หรือในปัจจุบันเรียกว่า mHealth เป็นคำที่ใช้สำหรับการปฏิบัติทางการแพทย์และสาธารณสุขโดยการใช้เทคโนโลยีระบบโทรศัพท์มือถือเป็นเครื่องมือในการสนับสนุนซึ่งอุปกรณ์ที่ใช้เช่น โทรศัพท์มือถือ และคอมพิวเตอร์แท็บเล็ต ในการเก็บรวบรวมข้อมูลทางคลินิกสุขภาพการจัดส่งข้อมูลการดูแลสุขภาพของนักปฏิบัติงาน นักวิจัย เพื่อตรวจสอบ

สถานการณ์ปัจจุบันผู้ป่วย เช่น สัญญาณชีพของผู้ป่วย และการให้การดูแลโดยตรง

### 2.2 แบบจำลองการยอมรับเทคโนโลยี (Technology Acceptance Model หรือ TAM) [5]

แบบจำลองการยอมรับเทคโนโลยีได้รับความนิยมอย่างแพร่หลายในการนำมาใช้อธิบายพฤติกรรมของผู้ใช้เทคโนโลยีสารสนเทศ วัตถุประสงค์ของการใช้แบบจำลองคือเพื่ออธิบายถึงปัจจัยที่มีผลต่อการยอมรับในเทคโนโลยี และใช้สำหรับพยากรณ์การใช้งานของระบบสารสนเทศ ผู้วิจัยได้นำแบบจำลองนี้มาช่วยในการออกแบบแอปพลิเคชันติดตามผู้ป่วยให้เหมาะสมกับกลุ่มผู้ใช้งาน โดยใช้หลักการ “Attitude = Perceived Usefulness + Perceived Ease of Use” ซึ่งเกิดจากการที่ผู้ใช้รู้สึกและรับรู้ได้ว่าระบบนั้นมีประโยชน์ อธิบายได้ดังรูปที่ 1



รูปที่ 1: แบบจำลองการยอมรับเทคโนโลยี (TAM) [5]

### 2.3 เทคนิคการวิเคราะห์องค์ประกอบหรือการวิเคราะห์ตัวประกอบเชิงสำรวจ (Exploratory Factor Analysis (EFA)) [1]

การวิเคราะห์ปัจจัยเชิงสำรวจ (Exploratory Factor Analysis: EFA) เป็นการทดสอบความสัมพันธ์ว่าตัวแปรที่สังเกตได้นั้น สามารถแบ่งออกได้ เป็นกี่กลุ่มปัจจัย และการแสดงถึงปัจจัยแฝงทั้งนี้ เกณฑ์ที่ใช้ในการพิจารณาได้แก่ (1) KMO ใช้อธิบายความเหมาะสมของข้อมูล (2) Total Variance Explained โดย ใช้ เพื่ อ ยี น ชั น องค์ประกอบว่าสามารถอธิบายข้อมูลได้มากเพียงใด โดยต้องมีค่า มากกว่า 65% ขึ้น ไป (3) Rotated Component Matrix ใช้เพื่ออธิบายถึงตัวแปรต่าง ๆ ว่าควรอยู่ องค์ประกอบใด ค่าที่ได้ต้องมากกว่าไม่น้อยกว่า 0.5 (4) Communality โดยค่าที่ได้ต้องไม่น้อยกว่า 0.5 เพื่อวัดความแปรปรวน

## 2.4 เทคนิคการหาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ (Multiple Regression) [1]

การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) กับตัวแปรตาม (Dependent Variable) การวิเคราะห์ถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Analysis) ถ้าตัวแปรอิสระมีมากกว่าหนึ่งตัวกับตัวแปรตามหนึ่งตัว เรียกว่า การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) การวิเคราะห์การถดถอยเชิงพหุคูณเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตาม 1 ตัว เพื่อศึกษาว่ามีตัวแปรอิสระตัวใดบ้างที่ร่วมกันทำนายหรือพยากรณ์ หรืออธิบายการผันแปรของตัวแปรตามได้โดยเขียนความสัมพันธ์ในรูปแบบของสมการได้ดังนี้

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \dots + \beta_n \times X_n$$

สัญลักษณ์ที่ใช้สามารถอธิบายได้ดังนี้

$X_i$  คือ ค่าของตัวแปรอิสระแต่ละตัว (จะใช้สัญลักษณ์  $x_i$  สำหรับค่าที่ได้จากตัวอย่างและสำหรับค่าประมาณหรือตัวทำนาย)

$Y$  คือ ค่าของตัวแปรตาม (จะใช้สัญลักษณ์  $y$  สำหรับค่าที่ได้จากตัวอย่าง)

$n$  คือ จำนวนตัวแปรอิสระในสมการถดถอย

$\beta_0$  คือ ค่าคงที่ของสมการถดถอย

$\beta_n$  คือ ค่าสัมประสิทธิ์การของตัวแปรอิสระ

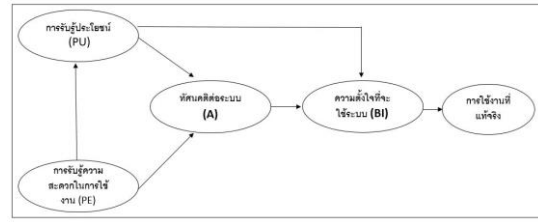
$\epsilon$  คือ ค่าความคลาดเคลื่อนระหว่างค่า  $Y$

## 3. วิธีการดำเนินงาน

หลังจากที่ผู้วิจัยได้ศึกษาทฤษฎี และงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้กำหนดกลุ่มประชากรตัวอย่างจำนวน 350 คน และได้กำหนดวิธีการดำเนินงานดังนี้

### 3.1 กำหนดกรอบแนวคิดของงานวิจัย

ผู้วิจัยได้วางแผนกรอบแนวคิดในการนำทฤษฎีแบบจำลองการยอมรับเทคโนโลยี (Technology Acceptance Model) มาใช้ในงานวิจัยดังรูปที่ 3



รูปที่ 3 : กรอบแนวคิดของงานวิจัย

ผู้วิจัยได้กำหนดตัวแปรดังนี้

1. Perceived Ease of Use (PE) คือ การรับรู้ว่าการใช้เทคโนโลยีนั้นใช้งานง่ายสะดวก ประกอบด้วย (1) ความง่ายของระบบปฏิบัติการ (2) ความชัดเจนของแอปพลิเคชัน (3) ความสะดวกสบายในการค้นหา
2. Perceived Usefulness (PU) คือ การรับรู้ถึงประโยชน์จากการใช้งานแอปพลิเคชันติดตามผู้ป่วย ประกอบด้วย (1) คุณภาพข้อมูลช่องทางติดต่อเจ้าหน้าที่ (2) การรับรู้ประโยชน์ของระบบ (3) ความรวดเร็วในการเข้าถึงข้อมูลทางการแพทย์
3. Attitude Toward Using (A) คือ ทัศนคติที่มีต่อการใช้งานระบบ ได้แก่ คุณภาพการบริการสุขภาพ (Health Service Quality) , ความเชื่อมั่นในระบบ , ความปลอดภัยของระบบปฏิบัติการ
4. Behavioral Intention to Use (BI) คือ ความตั้งใจ/เจตนาที่มีต่อการใช้ระบบ

### 3.2 พัฒนาข้อคำถามแบบเครื่องมือสำหรับงานวิจัย

ผู้วิจัยได้ใช้แบบสอบถามออนไลน์ (Online Questionnaire) ในการเก็บข้อมูลและเป็นแบบสอบถามปลายปิดโดยผู้วิจัยได้ตั้งคำถามในแบบสอบถามดังนี้

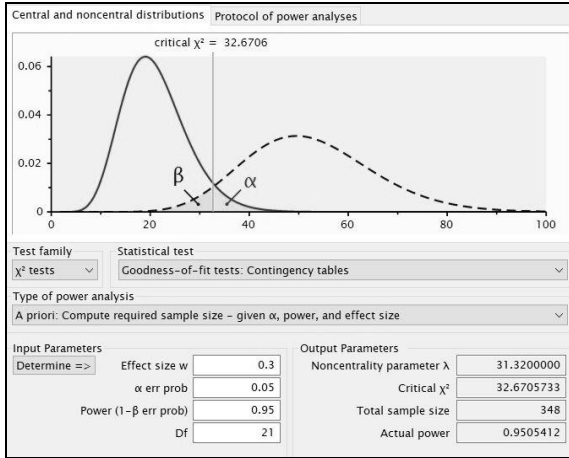
ส่วนที่ 1 : ข้อคำถามข้อมูลทั่วไป ได้แก่ เพศ อายุ ระดับการศึกษา และอาชีพ

ส่วนที่ 2 : ข้อคำถามที่เกี่ยวกับปัจจัยที่ส่งผลต่อการยอมรับเทคโนโลยี

### 3.3 การเก็บรวบรวมข้อมูลกับกลุ่มตัวอย่างที่เกี่ยวข้อง

การกำหนดกลุ่มประชากรตัวอย่างที่ใช้ในงานวิจัยครั้งนี้ ผู้วิจัยได้ขอคำแนะนำจากผู้เชี่ยวชาญ และใช้โปรแกรม g\*power ช่วยในการคำนวณขนาดตัวอย่าง พบว่าควรมี

กลุ่มประชากรอย่างน้อย 348 ตัวอย่างโดยเป็นกลุ่มอาสาสมัครในโรงพยาบาลกรณีศึกษา



รูปที่ 4 : ผลการคำนวณกลุ่มตัวอย่าง

**3.4 รวบรวมข้อมูลวิเคราะห์ผล**

เมื่อได้ค่าข้อมูลจากกลุ่มตัวอย่างของงานวิจัย ผู้วิจัยจะนำข้อมูลมาตรวจสอบความครบถ้วน และนำข้อมูลไปประมวลผลทางสถิติโดยแบ่งวิธีการวิเคราะห์ข้อมูลดังนี้

**3.4.1 การวิเคราะห์องค์ประกอบเชิงสำรวจ ((Exploratory Factor Analysis (EFA))** โดยนำข้อมูลจากกลุ่มตัวอย่างที่ตอบแบบสอบถามและนำเสนอข้อมูลในรูปแบบดังนี้ ค่าเฉลี่ยขององค์ประกอบ (Mean), ส่วนเบี่ยงเบนมาตรฐาน (Std. Deviation), ค่าวิเคราะห์องค์ประกอบ (Analysis N) เพื่ออธิบายลักษณะขององค์ประกอบ

**3.4.2 การวิเคราะห์น้ำหนักขององค์ประกอบ (Weight Score)** โดยการคัดองค์ประกอบเพื่อศึกษาถึงองค์ประกอบร่วมที่แสดงถึงความสัมพันธ์ระหว่างตัวแปรต่างๆ หากตัวแปรใดไม่เข้ากลุ่มจะถูกตัดออกไปจากงานวิจัย

**3.4.3 การวิเคราะห์สถิติเชิงปริมาณ** โดยใช้เทคนิคการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression)

**3.5 สรุปผลและข้อเสนอแนะ**

เพื่อให้ได้ข้อมูลที่น่าเชื่อถือและมีความถูกต้องผู้วิจัยได้นำผลการสังเคราะห์จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง รวมถึงผลที่ได้จากการตอบแบบสอบถามจากอาสาสมัครมาสอบถามกับผู้เชี่ยวชาญ

**4. ผลการดำเนินงาน**

**4.1 ผลจากการวิเคราะห์ข้อมูลประชากรทั่วไป**

การวิเคราะห์ข้อมูลทั่วไปกับกลุ่มประชากรตัวอย่างจำนวน 350 คน โดยเป็นกลุ่มอาสาสมัครในโรงพยาบาลกรณีศึกษาได้ผลการวิเคราะห์ข้อมูลดังตารางที่ 1

ตารางที่ 1 : ตารางผลการวิเคราะห์ข้อมูลทั่วไป

ข้อคำถาม	ข้อมูลทั่วไป	จำนวนคน	ร้อยละ
เพศ	● ชาย	141	40.3
	● หญิง	209	59.7
รวม		350	100
อายุ	● น้อยกว่า 20 ปี	8	2.3
	● 31 - 40 ปี	100	28.6
	● 41 - 50 ปี	131	37.4
	● 51 - 60 ปี	92	26.3
	● 61 - 70 ปี	19	5.4
รวม		350	100
ระดับการศึกษา	● มัธยมศึกษา	2	0.6
	● ประกาศนียบัตรวิชาชีพ	3	0.9
	● ประกาศนียบัตรวิชาชีพชั้นสูง	17	4.9
	● ปริญญาตรี	115	32.9
	● อื่นๆ	213	60.9
รวม		350	100
อาชีพ	● รับจ้างอิสระ	66	18.9
	● รับราชการ	76	21.7
	● ธุรกิจส่วนตัว	73	20.9
	● อื่นๆ	135	38.6
รวม		350	100

**4.2 การวิเคราะห์องค์ประกอบเชิงสำรวจ ((Exploratory Factor Analysis (EFA))**

การวิเคราะห์องค์ประกอบเชิงสำรวจ ((Exploratory Factor Analysis (EFA)) ผู้วิจัยใช้โปรแกรม IBM SPSS และใช้เกณฑ์การพิจารณาดังนี้

1. ค่า KMO (Kaiser-Meyer-Olkin) ใช้อธิบายความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์มีค่ามากกว่า 0.05 ซึ่งเป็นค่ายอมรับได้

2. ค่า Rotated Component Matrix ใช้อธิบายการจัดกลุ่มองค์ประกอบ โดยผู้วิจัยตัดองค์ประกอบที่ไม่เป็นองค์ประกอบร่วมและจัดเข้ากลุ่มใหม่ได้ดังตารางที่ 2

ตารางที่ 2 : ตารางค่าน้ำหนักองค์ประกอบหลังหมุนแกน

	Component			
	1	2	3	4
PU1	.706			
PU2	.652			
PU3	.603			
PE4		.736		
PE5		.681		
PE6		.606		
A7			.739	
A8			.566	
A9			.563	
BI10				.721
BI11				.533

Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization. <sup>a</sup>  
a. Rotation converged in 8 iterations.

ผลจากตารางสามารถอธิบายได้ว่าองค์ประกอบที่ 1 ได้แก่ การรับรู้ถึงประโยชน์จากการใช้งานระบบ องค์ประกอบที่ 2 การรับรู้ว่าคุณภาพเทคโนโลยีนั้นใช้งานง่ายและสะดวกสบาย องค์ประกอบที่ 3 ทักษะการที่มีต่อการใช้งานระบบ และองค์ประกอบที่ 4 ความตั้งใจ/เจตนาที่มีต่อการใช้งานระบบ

4.3 การวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression)

ผลที่ได้จากการวิเคราะห์ สามารถอธิบายได้ดังนี้ การรับรู้ว่าคุณภาพเทคโนโลยีนั้นใช้งานง่ายสะดวก (Perceived Ease of Use (PE)) กับ การรับรู้ถึงประโยชน์จากการใช้งาน (Perceived Usefulness (PU)) มีอิทธิพลกับทัศนคติที่มีต่อการใช้งานระบบ (Attitude Toward Using (A)) โดยมีค่า Sig น้อยกว่า 0.05 ซึ่งที่เป็นค่าที่ยอมรับได้ดังตารางที่ 3

ตารางที่ 3 : ตารางการวิเคราะห์อิทธิพลที่มีต่อทัศนคติ

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.897	.244		3.675	<.001
	PU	.446	.053	.423	8.409	<.001
	PE	.388	.061	.322	6.406	<.001

a. Dependent Variable: A

และการรับรู้ว่าคุณภาพเทคโนโลยีนั้นใช้งานง่ายสะดวก (Perceived Ease of Use (PE)) , การรับรู้ถึงประโยชน์จากการใช้งาน (Perceived Usefulness (PU)) และทัศนคติที่มีต่อการใช้งานระบบ (Attitude Toward Using (A)) มีอิทธิพลต่อ ความตั้งใจหรือเจตนาที่มีต่อการใช้งานระบบ (Behavioral Intention to Use (BI)) โดยมีค่า Sig น้อยกว่า 0.05 ซึ่งที่เป็นค่าที่ยอมรับได้ดังตารางที่ 4

ตารางที่ 4 : ตารางอิทธิพลที่มีต่อความตั้งใจหรือเจตนาที่มีต่อการใช้งานระบบ

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.074	.273		3.931	<.001
	PU	.338	.065	.309	5.230	<.001
	PE	.277	.071	.222	3.920	<.001
	A	.188	.064	.181	2.956	.003

a. Dependent Variable: BI

5. สรุปผล อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการดำเนินงานวิจัย

ตารางที่ 5 : ตารางสรุปผลการวิเคราะห์

Step		Model 1 (R <sup>2</sup> = .651)				Model 2 (R <sup>2</sup> = .575)			
		B	Coefficients Std. Error	Standardized Coefficients Beta	Sig.	B	Coefficients Std. Error	Standardized Coefficients Beta	Sig.
1	(Constant)	.897	.244		<.001				
	PU	.446	.053	.423	<.001				
	PE	.388	.061	.322	<.001				
2	(Constant)					1.074	.273		<.001
	PU					.338	.065	.309	<.001
	PE					.277	.071	.222	<.001
	A					.188	.064	.181	.003

โมเดลที่ 1 : ปัจจัยการรับรู้ว่าคุณภาพเทคโนโลยีนั้นใช้งานง่าย (Perceived Ease of Use (PE)) ซึ่งประกอบด้วย (1) ความง่ายของระบบปฏิบัติการ (2) ความชัดเจนของแอปพลิเคชัน (3) ความสะดวกสบายในการค้นหา กับปัจจัยการรับรู้ถึงประโยชน์จากการใช้งาน (Perceived Usefulness (PU)) ซึ่งประกอบด้วย (1) คุณภาพข้อมูลช่องทางติดต่อเจ้าหน้าที่ (2) การรับรู้ประโยชน์ของระบบ (3) ความรวดเร็วในการเข้าถึงข้อมูลทางการแพทย์ ส่งผลต่อทัศนคติที่มีต่อการใช้งานระบบ (Attitude Toward Using (A)) โดยอธิบายได้ร้อยละ 65 (R<sup>2</sup> = .651) ดังตารางที่ 5

โมเดลที่ 2 : ปัจจัยการรับรู้ว่าคุณภาพเทคโนโลยีนั้นใช้งานง่าย (Perceived Ease of Use (PE)) ซึ่งประกอบด้วย (1) ความง่ายของระบบปฏิบัติการ (2) ความชัดเจนของแอปพลิเคชัน (3) ความสะดวกสบายในการค้นหา , การรับรู้ถึงประโยชน์จากการใช้งาน (Perceived Usefulness (PU)) ซึ่งประกอบด้วย (1) คุณภาพข้อมูลช่องทางติดต่อเจ้าหน้าที่ (2) การรับรู้ประโยชน์ของระบบ (3) ความรวดเร็วในการเข้าถึงข้อมูลทางการแพทย์ และทัศนคติที่มีต่อการใช้งานระบบ (Attitude Toward Using (A)) ซึ่งประกอบไปด้วย (1) ทักษะการด้านความถูกต้องของข้อมูล



ในระบบ (2) ทักษะด้านความปลอดภัย (3) ทักษะติดต่อระบบปฏิบัติการส่งผลต่อความตั้งใจและเจตนาที่มีต่อการใช้ระบบ (Behavioural Intention to Use (BI)) โดยอธิบายได้ร้อยละ 57 ( $R^2 = .575$ ) ดังตารางที่ 5

## 5.2 อภิปรายผลการวิจัย

จากผลการศึกษานี้ผู้วิจัยยังพบว่า ปัจจัยที่ส่งผลกับการรับรู้เทคโนโลยีนั้นใช้งานง่าย ได้แก่ ความง่ายของระบบปฏิบัติการ ความชัดเจนของแอปพลิเคชัน และ ปัจจัยการรับรู้ถึงประโยชน์จากการใช้งาน ได้แก่ คุณภาพข้อมูลช่องทางติดต่อเจ้าหน้าที่ การรับรู้ประโยชน์ของระบบ และ ความรวดเร็วในการเข้าถึงข้อมูลทางการแพทย์ ซึ่งสอดคล้องกับงานวิจัยของพัชรนันท์มารีรักษ์ สังกะกุลลาบ (2561) สอดคล้องกับทฤษฎีการยอมรับเทคโนโลยี TAM ซึ่งทั้ง 2 ปัจจัยเป็นตัวแปรที่ส่งผลต่อการยอมรับเทคโนโลยีแอปพลิเคชันติดตามผู้ป่วยบนระบบปฏิบัติการแอนดรอยด์ อย่างมีนัยสำคัญที่ระดับ 0.05

ผลการวิจัยครั้งนี้ชี้ให้เห็นถึงความสำคัญของปัจจัยที่มีอิทธิพลต่อการยอมรับเทคโนโลยีแอปพลิเคชันติดตามผู้ป่วย ซึ่งเป็นพื้นฐานในการพัฒนาผลิตภัณฑ์ของธุรกิจบริการด้านสุขภาพในอนาคตต่อไปได้

## เอกสารอ้างอิง

- [1] Phatthanan Mariahlert. *Technology and science learning equipment*, Thammasat University, 2018.
- [2] Ahmed Alsswey , Hosam Al-Samarraie. Elderly users' acceptance of mHealth user interface (UI) design-based culture: the moderator role of age. *Journal on Multimodal User Interfaces*, 2019
- [3] Hanna O. Woldeyohannes and Ojelanki K. Ngwenyama. *Factors Influencing Acceptance and Continued Use of mHealth Apps Institute for Innovation and Technology Management* Ryerson University Toronto Canada, 2017
- [4] Muhammad Zubair Elahi. Fear of covid-19 and intentions towards adopting e-health services: exploring the technology acceptance model in the scenario of pandemic *International Journal of Business, Economics and Management University of Science and Technology of China*, 2021.
- [5] [https://www.researchgate.net/figure/Illustration-of-the-Technology-Acceptance-Model-TAM\\_fig1\\_228631114](https://www.researchgate.net/figure/Illustration-of-the-Technology-Acceptance-Model-TAM_fig1_228631114)
- [6] [https://www.researchgate.net/figure/UTAUT-model-Venkatesh-et-al-2003-In-the-UTAUT-model-performance-expectance-is-driven\\_fig1\\_281321623](https://www.researchgate.net/figure/UTAUT-model-Venkatesh-et-al-2003-In-the-UTAUT-model-performance-expectance-is-driven_fig1_281321623)
- [7] Rungrawee Sinla. *Self-care behavior of urinary tract stones disease patients at Samut Sakhon Hospital*. Professional Registered Nurse, Surgical Patient Examination Department Samut Sakhon Hospital, 2020.
- [8] Wiriyaoporn Techakritteerapong. *Factors for accepting the use of collaborative technology among personnel in the organization* Master of Science (Management Information Systems), Major Field: Management Information Systems Faculty of Commerce and Accountancy, Thammasat University, 2015.
- [9] Wiley, W. Souba, Chris, A. Haluck, & Melvyn, A. J. (2020). Marketing strategy: An essential component of business development for academic health centers. *The American Journal of Surgery*, 181, 105-114.
- [10] Wang, Y.H., Timmy H.T, Wang, W.T., Shih, Y.W., Chan, P.Y., et al. (2018). Developing and validating a mobile catering app success model. *Elsevier*, 1-12.

# ปัจจัยแบบกึ่งมีส่วนร่วมทางสังคมที่มีผลต่อความตั้งใจติดตาม บนแอปพลิเคชันติ๊กต็อก

## Parasocial Factors and Their Impact to the Intention to Follow Influencers on TikTok

แจ่มนภายอดหอ (Jamnapa Yodho)<sup>1</sup> และปราโมทย์ ลื่อนาม (Pramote Luenam)<sup>2</sup>  
สาขาบริหารเทคโนโลยีสารสนเทศ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์  
<sup>1</sup>jamnapa.yod@stu.nida.ac.th, <sup>2</sup>pramote@nida.ac.th

### บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยแบบกึ่งมีส่วนร่วมทางสังคมที่มีผลต่อความตั้งใจติดตามบนแอปพลิเคชันติ๊กต็อก เครื่องมือที่ใช้คือ แบบสอบถามออนไลน์ โดยศึกษาจากกลุ่มประชากรที่มีพฤติกรรมการใช้แอปพลิเคชัน TikTok และมีพฤติกรรมติดตามบัญชีผู้ใช้งาน TikTok ที่เนื้อหาวิดีโอแสดงด้วยบุคคลจำนวน 418 คน เมื่อนำข้อมูลมาวิเคราะห์ด้วยโมเดลสมการโครงสร้าง เพื่อทดสอบสมมติฐาน และหาความสัมพันธ์กับข้อมูลเชิงประจักษ์ โมเดลแสดงความเหมาะสมของด้วยค่า  $P\text{-value} = 0.105$ ,  $\chi^2/df = 196.616$ ,  $RMSEA = 0.018$ ,  $GFI = 0.960$ ,  $CFI = 0.995$ ,  $NFI = 0.963$  และ  $TLI = 0.993$  จากผลการศึกษาพบว่า ปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคมมีอิทธิพลทางตรงเชิงบวกต่อการความตั้งใจติดตามบนแอปพลิเคชัน TikTok ส่วนเนื้อหาของบัญชี TikTok มีอิทธิพลเชิงบวกทางอ้อมต่อความตั้งใจติดตามบนแอปพลิเคชัน TikTok ซึ่งเนื้อหาที่ให้ความบันเทิงมีอิทธิพลเชิงบวกทางอ้อมมากกว่าเนื้อหาที่มีประโยชน์ ส่วนความน่าเชื่อถือของแหล่งที่มาพบว่ามีผลต่อความตั้งใจติดตามบนแอปพลิเคชัน TikTok สำหรับผู้แสดงเนื้อหาวิดีโอที่มีความเชี่ยวชาญและมีความน่าเชื่อถือไม่มีอิทธิพลต่อความตั้งใจติดตามบนแอปพลิเคชัน TikTok

**คำสำคัญ:** TikTok ปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม มูลค่าของความบันเทิง มูลค่าของข้อมูล ความน่าเชื่อถือของแหล่งที่มา

### Abstract

The objective of this research is to investigate the parasocial factors and their impact to the intention to follow influencers on TikTok. The study used online questionnaire survey to collect a sample of 418 of people who have used TikTok and have followed influencers on TikTok. The data were analyzed through Structural Equation Model to test the hypotheses and derive model fit. The model showed a goodness-of-fit with  $P\text{-Value} = 0.105$ ,  $\chi^2/df = 196.616$ ,  $RMSEA = 0.018$ ,  $GFI = 0.960$ ,  $CFI = 0.995$ ,  $NFI = 0.963$ , and  $TLI = 0.993$ . The result of this research indicated that the parasocial relationship has a significant influence on the followers' intention to follow influencers on TikTok. The contents of a TikTok account have a positive indirect effect on the followers' intention to follow influencers on TikTok. When compared to the informative value, the entertainment value has a greater positive influence on the parasocial relationship. The credibility of the source has a positive indirect effect on the followers' intention to follow influencers on TikTok. The findings also show that the content contributors, expertise, and trustworthiness have no significant effect on the followers' intention to follow influencers on TikTok.

**Keywords:** TikTok, parasocial relationship, entertainment value, informative value, credibility.

## 1. บทนำ

เนื่องจาก TikTok เป็นแอปพลิเคชันที่ได้รับความนิยมมียอดดาวน์โหลดสูงที่สุดในโลกภายในระยะเวลาไม่กี่ปี หลังจากเปิดตัวแอปพลิเคชัน หลายบุคคลใช้แอปพลิเคชัน TikTok สร้างแบรนด์ส่วนบุคคล นำเสนอเนื้อหาวิดีโอผสมผสานทักษะประสบการณ์ บุคลิกภาพที่น่าดึงดูดใจ และความน่าเชื่อถือของตัวบุคคลส่งผลให้ได้รับความสนใจสร้างเครือข่ายมีผู้ติดตามจำนวนมากกลายเป็นอินฟลูเอนเซอร์ ซึ่งจากข้อมูลสถิติของ [1] พบว่าอินฟลูเอนเซอร์ของแอปพลิเคชัน TikTok ส่วนใหญ่เป็นบุคคลทั่วไป ไม่ได้เป็นผู้ที่มีชื่อเสียงโด่งดังมาก่อน แตกต่างจากโซเชียลมีเดียอื่นที่เป็นการยกย่องสำหรับผู้ไม่ได้มีชื่อเสียงมาก่อนจะได้รับความนิยมและมีผู้ติดตามเป็นจำนวนมาก จากการศึกษาพบว่าบางบัญชี TikTok มีจำนวนวิดีโอไม่ถึง 1,000 วิดีโอแต่มีจำนวนผู้ติดตามมากกว่าล้านผู้ติดตามแต่บางบัญชี TikTok มีจำนวนผู้ติดตามไม่ถึง 100 คน หลายบัญชี TikTok เป็นผู้คนทั่วไปที่มีจำนวนผู้ติดตามมากกว่าผู้ที่มีชื่อเสียงโด่งดังมาก่อน ทำให้เป็นที่น่าสนใจว่าผู้ผลิตสื่อสร้างเนื้อหาวิดีโอในการสื่อสารกับผู้ชมวิดีโออย่างไรที่ทำให้เกิดความรู้สึกพึงพอใจที่อยากติดตาม รับชมผลงานผ่านสื่อกลางแอปพลิเคชัน TikTok

งานวิจัยนี้จึงต้องการศึกษาว่าศึกษาปัจจัยที่มีอิทธิพลต่อความตั้งใจกดติดตามบนแอปพลิเคชัน TikTok เพื่อใช้เป็นแนวทางสร้างกลยุทธ์สำหรับผู้ที่ต้องการเป็นอินฟลูเอนเซอร์ ปัจจัยที่ผู้วิจัยนำมาใช้ในการศึกษาได้แก่ เนื้อหา (Content) ความน่าเชื่อถือของแหล่งที่มา (Credibility) ปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (Parasocial Relationship) และความตั้งใจกดติดตามแอปพลิเคชัน TikTok

## 2. งานวิจัยและทฤษฎีที่เกี่ยวข้อง

ความตั้งใจกดติดตามบนแอปพลิเคชัน TikTok (Follower: FL) ได้แนวคิดมาจาก [2] การเป็นอินฟลูเอนเซอร์คือบุคคลธรรมดาที่ใช้โซเชียลมีเดียในการแสดงความคิดเห็นเกิดเป็นชุมชนออนไลน์ ซึ่งอินฟลูเอนเซอร์มีความเป็นผู้นำต่อผู้ติดตามของเขา เนื่องจากผู้ติดตามที่ชื่นชอบเขาจะมีความไว้วางใจ เพื่อรักษาจำนวนผู้ติดตาม

เนื้อหาจึงต้องได้รับการเผยแพร่อย่างสม่ำเสมอและการนำเสนอเนื้อหาที่มีคุณภาพ ซึ่งผู้ติดตามเกิดความตั้งใจกดติดตามจากทฤษฎีการกระทำตามหลักเหตุผล [3] กล่าวว่าการกำหนดความตั้งใจเชิงพฤติกรรมเกิดจากพฤติกรรมทางด้านทัศนคติของบุคคลและบรรทัดฐานหรือความเชื่อของบุคคล โดยแตกต่างออกไปตามพฤติกรรม สถานการณ์ และความแตกต่างของแต่ละบุคคล ในงานวิจัยนี้วัดความตั้งใจกดติดตามบนแอปพลิเคชัน TikTok เพื่อนำไปสู่การเป็นอินฟลูเอนเซอร์ จากความน่าจะเป็นในการแสดงออกถึงพฤติกรรมหรือแนวโน้มของการกดติดตามแอปพลิเคชัน TikTok ที่บ่งบอกถึงความต้องการหรือความสนใจกดติดตาม การคาดคะเนกดติดตามในอนาคต และการแนะนำเพื่อนเพื่อกดติดตาม โดยนำปัจจัยที่มีอิทธิพลต่อความตั้งใจกดติดตามบนแอปพลิเคชัน TikTok ดังนี้

เนื้อหา (Content) ของบัญชี TikTok กล่าวถึงลักษณะเนื้อหาวิดีโอของบัญชี TikTok ที่โพสต์ โดยนำแนวคิดเนื้อหาของอินฟลูเอนเซอร์ (Content influence) ของ [7] มาปรับใช้ ซึ่งประกอบด้วย มูลค่าของข้อมูล (Informative value: IV) ความหมายคือทัศนคติของผู้รับชมเนื้อหาวิดีโอของบัญชี TikTok ที่ได้รับข้อมูลเป็นประโยชน์ ตรงประเด็น เป็นแหล่งข้อมูลที่ดีในการอัพเดทข้อมูลข่าวสาร และมูลค่าของความบันเทิง (Entertainment value: EV) คือทัศนคติของผู้รับชมเนื้อหาวิดีโอของบัญชี TikTok ที่ทำให้เกิดความรื่นรมย์ ความน่าตื่นเต้นและความเพลิดเพลิน เนื่องจากลักษณะเนื้อหาของวิดีโอของบัญชี TikTok เป็นวิดีโอสั้นๆช่วยให้รับรู้ถึงประโยชน์ที่กระชับและตรงประเด็นนอกจากนั้นผู้ชมวิดีโอยังคาดหวังการชมวิดีโอเพื่อความบันเทิง [8] หากสร้างเนื้อหาวิดีโอใหม่ๆที่สอดคล้องกับผู้ชมวิดีโอจะทำให้ขยายจำนวนผู้ติดตามได้ [9] จากข้อมูลดังกล่าว ผู้วิจัยจึงคิดว่าเนื้อหาของบัญชี TikTok มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามบนแอปพลิเคชัน TikTok ผ่านการมีปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม และตั้งเป็นสมมติฐานดังนี้

**สมมติฐานที่ 1 (H1):** มูลค่าของข้อมูล (IV) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR)

**สมมติฐานที่ 2 (H2):** มูลค่าของความบันเทิง (EV) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม(PR)

ความน่าเชื่อถือของแหล่งที่มา (Credibility) กล่าวถึงความรู้สึกของผู้ชมเนื้อหาวิดีโอที่มีต่อผู้แสดงเนื้อหาในบัญชี TikTok โดยใช้มาตราส่วนในการวัดความน่าเชื่อถือของแหล่งที่มาจากงานวิจัย [10] ที่สร้างมาตราส่วนเพื่อวัดความน่าเชื่อถือของแหล่งที่มาจากการวัดความเชี่ยวชาญ ความน่าเชื่อถือและความน่าดึงดูดใจของผู้ที่มีชื่อเสียง ซึ่งเป็นงานวิจัยที่ได้รับการยอมรับและถูกนำไปใช้อย่างแพร่หลาย โดยกำหนดมาตราส่วนวัดความน่าเชื่อถือของแหล่งที่มาไว้ ดังนี้ (1) มาตราส่วนวัดความน่าดึงดูดใจ (Attractiveness: AN) หล่อ/สวย-น่าเกลียด มีระดับ-ไม่มีระดับเรียบหรู-ธรรมดา เช็คซ์-ไม่เช็คซ์ (2) มาตราส่วนวัดความเชี่ยวชาญ (Expertise: ET) คือ เป็นผู้เชี่ยวชาญ-ไม่ใช่ผู้เชี่ยวชาญมีความรู้-ไม่มีความรู้ เก่ง-ไร้ฝีมือ (3) มาตราส่วนวัดความน่าเชื่อถือ (Trustworthiness: TN) ฟังพาได้-ฟังพาไม่ได้ ซื่อสัตย์-ไม่ซื่อสัตย์ เชื่อถือได้-ไม่น่าเชื่อถือ ซึ่งแนวคิดนี้ผู้วิจัยได้ดัดแปลงจากการวัดผลความตั้งใจซื้อและการวัดผลโฆษณาผ่านความน่าเชื่อถือของอินฟลูเอนเซอร์ที่รับรองแบรนด์ผลิตภัณฑ์ เนื่องจากการเป็นอินฟลูเอนเซอร์คือการสร้างแบรนด์ส่วนบุคคล การวัดผลความตั้งใจซื้อสินค้าหรือโฆษณาของแบรนด์สินค้าดัดแปลงนำมาวัดผลความตั้งใจกดติดตามของแบรนด์ส่วนบุคคล ซึ่งในบริบทนี้หมายถึงบัญชี TikTok ซึ่งจากงานวิจัย [11] พบว่าการสร้างแบรนด์ส่วนบุคคลคือการอธิบายตัวตนดั้งเดิมของใครบางคนและเพิ่มศักยภาพขึ้นมาด้วยการนำเสนอเนื้อหาแบรนด์ส่วนบุคคลอย่างสม่ำเสมอซึ่ง TikTok เป็นแพลตฟอร์มโซเชียลมีเดียที่ง่ายในการสร้างความรับรู้ สื่อสารและพัฒนาแบรนด์ส่วนบุคคลสอดคล้องกับงานวิจัย [12] ที่พบว่าผู้ติดตามจำนวนมากชื่นชมบุคลิกภาพของอินฟลูเอนเซอร์และทำให้ชื่นชอบในผลิตภัณฑ์ที่อินฟลูเอนเซอร์รับรอง สอดคล้องกับ [13] ที่พบว่าองค์ประกอบความน่าเชื่อถือของอินฟลูเอนเซอร์ ซึ่งประกอบด้วยความเชี่ยวชาญ ความน่าเชื่อถือ ความดึงดูด

ใจและความคล้ายคลึงกันของผู้ตามอินฟลูเอนเซอร์ส่งผลดีต่อความไว้วางใจของเนื้อหาแบรนด์สินค้าที่อินฟลูเอนเซอร์โฆษณา ซึ่งส่งผลต่อการโน้มน้าวใจและความตั้งใจซื้อของผู้ติดตามอินฟลูเอนเซอร์ ดังนั้นผู้วิจัยจึงคิดว่าความน่าเชื่อถือของแหล่งที่มาที่มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามบนแอปพลิเคชัน TikTok ผ่านการมีปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม และตั้งเป็นสมมติฐานดังนี้

**สมมติฐานที่ 3 (H3):** ความน่าดึงดูดใจ (AN) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR)

**สมมติฐานที่ 4 (H4):** ความเชี่ยวชาญ (ET) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR)

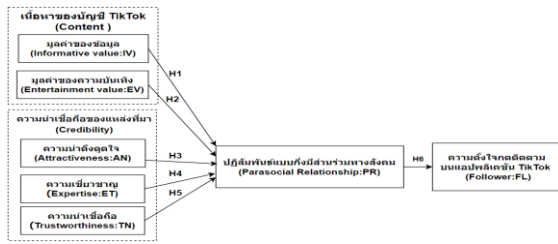
**สมมติฐานที่ 5 (H5):** ความน่าเชื่อถือ (TN) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR)

ปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (Parasocial Relationship: PR) กล่าวถึงความรู้สึกของผู้รับชมเนื้อหาวิดีโอที่มีปฏิสัมพันธ์กับผู้แสดงเนื้อหาวิดีโอของบัญชี TikTok เป็นการสื่อสารของผู้แสดงเนื้อหาวิดีโอที่กำลังสื่อสารกับผู้ชมวิดีโอ ซึ่งความสำเร็จของบุคคลสาธารณะในโซเชียลมีเดียขึ้นอยู่กับความสามารถในการสร้างความสัมพันธ์กับผู้ชม [4] ก่อให้เกิดการมีปฏิสัมพันธ์แบบฝ่ายเดียวที่ผู้รับชมวิดีโอรู้สึกมีปฏิสัมพันธ์กับผู้แสดงเนื้อหาวิดีโออยากติดตามข่าวสารและผลงานของผู้ที่ชื่นชอบ [5] มีความรู้สึกห่วงใยและรู้สึกว่าผู้แสดงเนื้อหาวิดีโอเข้าอกเข้าใจในสิ่งที่อยากรู้ โดยใช้แอปพลิเคชัน TikTok เป็นสื่อกลางในการสื่อสารทางโซเชียลมีเดีย เช่นการมีปฏิสัมพันธ์แบบแฟนคลับดารา นักร้องนักแสดง เป็นต้น [6] จากข้อมูลดังกล่าว ผู้วิจัยจึงคิดว่าการมีปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคมมีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามบนแอปพลิเคชัน TikTok และตั้งเป็นสมมติฐานดังนี้

**สมมติฐานที่ 6 (H6):** ปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR) อิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL)

### 3. กรอบแนวคิดของการวิจัย

จากการศึกษาแนวคิด และทฤษฎีที่เกี่ยวข้องสามารถกำหนดกรอบแนวคิดงานวิจัย ดังแสดงในภาพที่ 1



ภาพที่ 1: กรอบแนวคิดงานวิจัย

### 4. วิธีการดำเนินการวิจัย

ในการศึกษางานวิจัยนี้เป็นการวิจัยเชิงปริมาณ (Quantitative Research) ใช้แบบสอบถามออนไลน์ ในการรวบรวมข้อมูลจากกลุ่มตัวอย่างประชากรที่เป็นบุคคลทั่วไปที่มีพฤติกรรมการใช้ และพฤติกรรมการติดตามบัญชีผู้ใช้งาน TikTok ที่เนื้อหาวิดีโอแสดงด้วยบุคคลอยู่แล้ว ผู้วิจัยใช้โปรแกรม G\*Power ในการคำนวณหาขนาดตัวอย่างประชากร โดยกำหนดค่าเพาเวอร์ (1-β) เท่ากับ 0.95 ค่าอัลฟา (α) เท่ากับ 0.05 ขนาดของอิทธิพล (Effect size) เท่ากับ 0.03 จำนวนตัวแปรทำนายเท่ากับ 7 ตัวแปร ผลที่ได้คือ ขนาดของกลุ่มตัวอย่างเท่ากับ 363 ตัวอย่าง ซึ่งผู้วิจัยได้ปรับจำนวนขนาดตัวอย่างจากแบบสอบถามเพิ่มเติมเป็น 388 และเพิ่มการเก็บข้อมูลที่ใช้สำหรับการศึกษานำร่อง (Pilot study) เพื่อหาความเชื่อมั่นอีก 30 ตัวอย่าง รวมเป็น 418 ตัวอย่าง เพื่อให้การวิเคราะห์หมีองค์ประกอบที่น่าเชื่อถือมากขึ้น โดยใช้การสุ่มตัวอย่างแบบสะดวก (Convenient Sampling) ซึ่ง แบ่งแบบสอบถามออกเป็น 7 ส่วน ประกอบด้วย ส่วนที่ 1 แบบสอบถามคัดกรองผู้ใช้งานแอปพลิเคชัน TikTok ส่วนที่ 2 แบบสอบถามข้อมูลทั่วไป ส่วนที่ 3 แบบสอบถามเกี่ยวกับพฤติกรรมการใช้แอปพลิเคชัน TikTok ส่วนที่ 4 แบบสอบถามเกี่ยวกับเนื้อหาของบัญชี TikTok (Content) ส่วนที่ 5 แบบสอบถามเกี่ยวกับความน่าเชื่อถือของแหล่งที่มา (Credibility) ส่วนที่ 6 แบบสอบถามเกี่ยวกับปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (Parasocial Relationship) และ ส่วนที่ 7 แบบสอบถามเกี่ยวกับความตั้งใจในการกดติดตามบนแอป

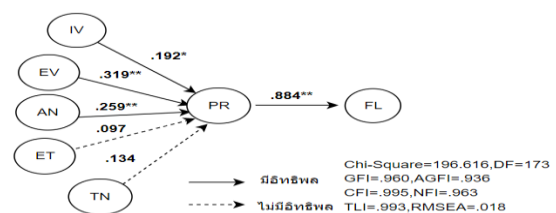
พลิเคชัน TikTok (Follower) ซึ่งรูปแบบคำถามประกอบด้วยคำถามแบบตรวจรายการ (Check List) ลักษณะคำถามปลายปิด (Close-Ended Questions) คำถามหลายตัวเลือก (Multiple Choices Question) และรูปแบบลักษณะคำถามแบบมาตราวัด (Interval Scale) โดยกำหนดค่าคะแนนช่วงน้ำหนัก 5 ระดับ

ผู้วิจัยได้ทำการทดสอบสมมติฐานของงานวิจัยโดยใช้โปรแกรม SPSS และ AMOS ในการวิเคราะห์รูปแบบเชิงโครงสร้าง (Structural Equation Modeling: SEM) เพื่อวิเคราะห์หาค่าอิทธิพลทั้งทางตรงและทางอ้อม และได้มีการปรับรูปแบบของโมเดล เพื่อทดสอบความสอดคล้องของโมเดลกับกับข้อมูลเชิงประจักษ์ (Model Fit)

### 5. ผลการดำเนินงานวิจัย

ในการศึกษาวิจัยครั้งนี้มีผู้ตอบแบบสอบถามทั้งหมด 490 คน เป็นผู้บุคคลทั่วไปที่มีพฤติกรรมการใช้แอปพลิเคชัน TikTok และมีพฤติกรรมการติดตามบัญชีผู้ใช้งาน TikTok ที่เนื้อหาวิดีโอแสดงด้วยบุคคล ข้อมูลที่สามารถนำมาวิเคราะห์ข้อมูลได้เป็นจำนวน 418 คน แบ่งเป็นเพศชาย 147 คน หญิง 264 คน และไม่ระบุเพศ 7 คน หากแบ่งตามช่วงอายุเป็น 15-30 ปี จำนวน 175 คน อายุ 31-45 ปี จำนวน 230 คน และอายุ 45-60 ปี จำนวน 13 คน

การวิจัยใช้การวิเคราะห์รูปแบบเชิงโครงสร้าง โดยเป็นการวิเคราะห์เส้นทาง (Path Analysis) เพื่อหาค่าอิทธิพลทั้งทางตรงและทางอ้อมของตัวแปรอิสระว่ามีผลต่อตัวแปรตามมากน้อยเพียงใด โดยค่าความสอดคล้องของโมเดล (Goodness-of-Fit) ที่ได้มีค่าดังนี้ P-value=0.105,  $\chi^2/df=196.616$ , RMSEA=0.18, GFI=0.960, CFI=0.995, NFI=0.963 และ TLI=0.993 ตามผลสถิติสามารถแสดงผลลัพธ์ที่ได้ในโมเดลการวิจัยในภาพที่ 2 และแสดงผลการวิเคราะห์เชิงสาเหตุของโมเดลในตารางที่ 1



ภาพที่ 2: โมเดลจากการวิจัย

ผลที่ได้จากการวิเคราะห์อิทธิพลเชิงสาเหตุภายในโมเดล สามารถนำมาสรุปผลการทดสอบของสมมติฐานต่าง ๆ ที่ตั้งไว้มีรายละเอียดที่แสดงในตารางที่ 2

ตารางที่ 1: ผลการวิเคราะห์เชิงสาเหตุของโมเดล

ตัวแปร ผล	อิทธิพล	ตัวแปรต้น					
		IV	EV	AN	ET	TN	PR
PR	ทางตรง	0.192 *	0.319 ***	0.259***	0.097	0.134	-
	ทางอ้อม	-	-	-	-	-	-
FL	ทางตรง	-	-	-	-	-	0.884***
	ทางอ้อม	0.170*	0.282**	0.229**	0.086	0.118	-

หมายเหตุ \*p<0.05, \*\*p<0.01, \*\*\*p<0.001

ตารางที่ 2: สรุปผลการทดสอบสมมติฐาน

สมมติฐาน	ค่าอิทธิพล	ค่า P	ผลการทดสอบ
H1: มูลค่าของข้อมูล (IV) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR)	0.170	≤0.05	ยอมรับ
H2: มูลค่าของความบันเทิง (EV) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR)	0.282	<0.001	ยอมรับ
H3: ความน่าดึงดูดใจ (AN) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR)	0.229	<0.001	ยอมรับ
H4: ความเชี่ยวชาญ (ET) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR)	0.086	0.394	ปฏิเสธ
H5: ความน่าเชื่อถือ (TN) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL) ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR)	0.118	0.120	ปฏิเสธ
H6: ปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม (PR) มีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok (FL)	0.884	<0.001	ยอมรับ

## 6. สรุป

ผลการศึกษาปัจจัยที่มีอิทธิพลต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok ผู้วิจัยได้ดำเนินการวิเคราะห์และอภิปรายผลการทดสอบเป็นประเด็นต่าง ๆ ได้ดังนี้

เนื้อหาของบัญชี TikTok พบว่า เนื้อหาของบัญชี TikTok ทั้งสองด้านคือด้านมูลค่าของข้อมูลและมูลค่าความบันเทิงมีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามบัญชี TikTok ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทาง

สังคม ซึ่งสอดคล้องกับงานวิจัยของ Virkkunen และ Norhio [2] ที่พบว่าการนำเสนอเนื้อหาที่มีคุณภาพมีผลต่อความตั้งใจในการกดติดตามบัญชี TikTok โดยจากงานวิจัยแสดงให้เห็นถึงมูลค่าความบันเทิงมีอิทธิพลเชิงบวกมากกว่ามูลค่าของข้อมูลสอดคล้องกับงานวิจัยของ Xiao, Wang, และ Wang [8] ที่พบว่าวิดีโอสั้นๆช่วยให้ถูกคำรับรู้อิงประโยชน์ที่กระชับและตรงประเด็น นอกจากนี้ถูกคำมักคาดหวังการชมวิดีโอเพื่อความบันเทิง ซึ่งจากงานวิจัยแสดงให้เห็นพฤติกรรมกดติดตามแอปพลิเคชัน TikTok ผู้ชมวิดีโอมีความตั้งใจกดติดตามเนื้อหาที่ให้ความบันเทิงมากที่สุด การกดติดตามเนื้อหาที่มีประโยชน์เพื่อเป็นแหล่งอัปเดตข้อมูลข่าวสารมีอิทธิพลเพียงเล็กน้อย ดังนั้น ผู้ที่ต้องการเป็นอินฟลูเอนเซอร์บนแอปพลิเคชัน TikTok ควรเน้นสร้างเนื้อหาวิดีโอที่ให้ความบันเทิงแก่ผู้รับชมวิดีโอ ให้ผู้รับชมวิดีโอรู้สึกเพลิดเพลิน รื่นรมย์ หรือตื่นเต้นกับเนื้อหาวิดีโอ โดยลักษณะเนื้อหาวิดีโอควรสอดแทรกให้ผู้รับชมวิดีโอรู้สึกถึงประโยชน์ของเนื้อหา สั้นๆตรงประเด็นให้เป็นแหล่งอัปเดตข้อมูลข่าวสารที่ดีแก่ผู้รับชมเนื้อหาวิดีโอ เพื่อเพิ่มโอกาสให้เกิดความตั้งใจกดติดตามบัญชี TikTok

ความน่าเชื่อถือของแหล่งที่มา พบว่า ความน่าดึงดูดใจของผู้แสดงเนื้อหาวิดีโอมีอิทธิพลเชิงบวกต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม สอดคล้องกับ Janssen, Schouten, และ Croes [12] ที่พบว่าผู้ติดตามจำนวนมากชื่นชมบุคลิกภาพของอินฟลูเอนเซอร์ ซึ่งจากงานวิจัยบ่งบอกผู้แสดงเนื้อหาที่มีความเชี่ยวชาญและความน่าเชื่อถือไม่มีอิทธิพลต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok ผ่านปฏิสัมพันธ์แบบกึ่งมีส่วนร่วมทางสังคม แสดงให้เห็นว่า ผู้ที่ต้องการเป็นอินฟลูเอนเซอร์บนแอปพลิเคชัน TikTok ควรให้ความสำคัญเรื่องบุคลิกภาพที่มีความน่าดึงดูดใจ เนื่องจากผลงานวิจัยพบว่าความตั้งใจในการกดติดตามนั้น ความน่าดึงดูดใจเป็นปัจจัยสำคัญและมีอิทธิพลเชิงบวกรองจากเนื้อหาวิดีโอที่มีความบันเทิงที่ส่งผลต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok ดังนั้น บุคลิกภาพที่ทำให้ผู้ชมวิดีโอรู้สึกกับผู้แสดงเนื้อหาวิดีโอ

เป็นผู้ที่มีบุคลิกภาพสวย/หล่อหรือบุคลิกภาพที่ดูดีมีระดับ หรือบุคลิกภาพที่ธรรมดาที่เข้าถึงง่ายหรือบุคลิกภาพที่มีความเชื่อกันประทับใจต่อผู้ชมวิดีโอสร้างความดึงดูดใจผู้ชมวิดีโอ และทำให้เกิดความตั้งใจในการกดติดตามบัญชี TikTok

โดยสรุปแล้วปัจจัยที่มีอิทธิพลต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok ผลงานวิจัยพบว่าเนื้อหาที่มีความบันเทิงมีอิทธิพลสูงสุดต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok รองลงมาคือความน่าดึงดูดใจของผู้แสดงเนื้อหาวิดีโอที่มีบุคลิกภาพสวย/หล่อ ดูดีมีระดับ ธรรมดา เข้าถึงง่าย เชื่อกันประทับใจ และสุดท้ายคือเนื้อหาที่มีประโยชน์มีอิทธิพลเชิงบวกเพียงเล็กน้อยต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok สำหรับผู้แสดงเนื้อหาที่มีความเชี่ยวชาญและมีความน่าเชื่อถือไม่มีอิทธิพลต่อความตั้งใจกดติดตามแอปพลิเคชัน TikTok

ข้อเสนอแนะสำหรับผู้ต้องการเป็นอินฟลูเอนเซอร์บนแอปพลิเคชัน TikTok ควรเน้นสร้างเนื้อหาให้มีความบันเทิง นำเสนอผ่านผู้แสดงเนื้อหาที่มีความดึงดูดใจต่อผู้รับชม เพิ่มเติมเนื้อหาที่มีประโยชน์ สั้น ๆ ตรงประเด็น เพื่อเพิ่มโอกาสในการติดตามแอปพลิเคชัน TikTok

ข้อเสนอแนะสำหรับงานวิจัยในครั้งถัดไป การศึกษาคควรนำเรื่องของการกดติดตามสัตว์น่ารักต่าง ๆ บนแอปพลิเคชัน TikTok มาเป็นปัจจัยเพิ่มเติมว่า มีอิทธิพลต่อความตั้งใจกดติดตามบนแอปพลิเคชัน TikTok อย่างไร

### เอกสารอ้างอิง

- [1] tikrank. (2021). "Top100 TikTok Influencer in Thailand." สืบค้นเมื่อวันที่ 9 กรกฎาคม 2564 จาก <https://tikrank.com/tiktok-influencer-rank/top-100-influencer-in-tiktok-Thailand>
- [2] P. Virkkunen. and E. Norhio. "Becoming a Social Media Influencer: Describing the journey of becoming a successful social media influencer." p.11, 2019.
- [3] I. Ajzen. "The Theory of Planned Behavior. *Organizational Behavior & Human Decision Processes.*" Vol.50, No. 2, p.179-212, 1991.
- [4] A. M. Ledbetter, and C. Meisner. "Extending the personal branding affordances typology to parasocial interaction with public figures on social media: Social presence and media multiplexity as mediators." *Computers in Human Behavior*, Vol. 115, p. 106610, 2021.
- [5] S. F. Rosaen, and J. L. Dibble. "Clarifying the role of attachment and social compensation on parasocial relationships with television characters." *Communication Studies*, Vol. 67 (2), p. 147-162, 2016.
- [6] รุฬพัศวัตต์ ครอบภูมินทร์. "สถานภาพ องค์ความรู้ ของ การ สื่อสาร ปฏิสัมพันธ์ แบบ กึ่ง มี ส่วน ร่วม ทาง สังคม." *วารสาร สังคมศาสตร์ และ มนุษยศาสตร์*, ปีที่ 47, ฉบับที่ 2, 2021.
- [7] C. Lou, and H. K. Kim. "Fancying the new rich and famous? Explicating the roles of influencer content, credibility, and parental mediation in adolescents' parasocial relationship, materialism, and purchase intentions." *Frontiers in Psychology*, Vol. 10, p. 2567, 2019.
- [8] Y. Xiao, L. Wang, and P. Wang. "Research on the influence of content features of short video marketing on consumer purchase intentions." in 4th International Conference on Modern Management, Education Technology and Social Science (MMETSS 2019), *Atlantis Press*, pp. 415-422, 2019.
- [9] Y. Y. U. Ishihara, and R. Oktavianti. "Personal Branding Influencer di Media Sosial TikTok." *Koneksi*, Vol. 5 (1), pp. 76-82, 2020.
- [10] R. Ohanian. "Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness." *Journal of advertising*, Vol. 19 (3), pp. 39-52, 1990.
- [11] E. D. S. Watie, and F. A. Eryad. "Personal Branding Mahasiswa Ilmu Komunikasi Usm Geri Da Pinto Pada Akun Creator Tiktok@ Omagaaakkk." *Journal Dinamika Sosial Budaya*, Vol. 23, No. 1, pp. 10-21, 2021.
- [12] L. Janssen, A. P. Schouten, and E. A. Croes. "Influencer advertising on Instagram: product-influencer fit and number of followers affect advertising outcomes and influencer evaluations via credibility and identification." *International Journal of Advertising*, pp. 1-27, 2021.
- [13] C. Lou, and S. Yuan. "Influencer marketing: how message value and credibility affect consumer trust of branded content on social media." *Journal of Interactive Advertising*, Vol. 19 (1), pp. 58-73, 2019.

# การวิเคราะห์องค์ประกอบที่มีอิทธิพลต่อการลงทุนสกุลเงินดิจิทัล

## An Analysis of Factors Influencing Cryptocurrency Investments

อนงวรรณิญา สุวยรูป (Anongwanneeya Suayroop)<sup>1</sup> และมมเทียร รัตน์ศิริวงศ์วุฒิ (Montean Rattanasiriwongwut)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021856110@email.kmutnb.ac.th, <sup>2</sup>montean.r@it.kmutnb.ac.th

### บทคัดย่อ

ในปัจจุบันเทคโนโลยีเริ่มเข้ามามีบทบาทในชีวิตประจำวันของมนุษย์รวมถึงสกุลเงินดิจิทัลที่ได้รับความนิยมอย่างมากเนื่องจากความผันผวนของราคาสูงแต่ถึงจะมีความเสี่ยงแต่ก็สามารถทำมูลค่าได้สูงในระยะเวลาอันสั้นทำให้นักลงทุนหันมาสนใจสกุลเงินดิจิทัลอย่างมากโดยเฉพาะในอนาคต การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์องค์ประกอบที่มีอิทธิพลต่อการเลือกการลงทุนสกุลเงินดิจิทัลและหาความสัมพันธ์ขององค์ประกอบที่มีอิทธิพลต่อการเลือกการลงทุนสกุลเงินดิจิทัลกลุ่มตัวอย่างในงานวิจัยครั้งนี้คือ กลุ่มคนที่เคยลงทุนสกุลเงินดิจิทัล จำนวน 360 คน โดยใช้เครื่องมือเป็นแบบสอบถาม ใช้สถิติในการวิเคราะห์ข้อมูล ได้แก่ ค่าความถี่ ร้อยละ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน การวิเคราะห์องค์ประกอบเชิงสำรวจและการวิเคราะห์องค์ประกอบเชิงยืนยัน

ผลการวิจัยพบว่า ในการวิเคราะห์องค์ประกอบเชิงสำรวจ ได้องค์ประกอบที่มีผลต่อการตัดสินใจลงทุนสกุลเงินดิจิทัล จำนวน 5 องค์ประกอบ ได้แก่ องค์ประกอบด้านสังคม องค์ประกอบด้านผลตอบแทน องค์ประกอบด้านบริการ องค์ประกอบด้านสกุลเงินดิจิทัล และองค์ประกอบด้านตัวบุคคล การวิเคราะห์องค์ประกอบเชิงยืนยันพบว่ามีค่า  $CMIN/DF$  (1.120),  $GFI$  (0.927),  $AGFI$  (0.913),  $CFI$  (0.990),  $RMSEA$  (0.018),  $SRMR$  (0.0380) แสดงว่าการทดสอบตามทฤษฎีมีความสอดคล้องกับข้อมูลเชิงประจักษ์

**คำสำคัญ:** การวิเคราะห์องค์ประกอบเชิงสำรวจ  
การวิเคราะห์องค์ประกอบเชิงยืนยัน  
สกุลเงินดิจิทัล

### Abstract

Nowadays, technology has started to play a role in human daily life. Including cryptocurrencies that are very popular due to their high price volatility. But even with the risk, it can make a high value. In a relatively short period of time, investors are increasingly interested in cryptocurrencies, especially in the future. The purpose of this research was to analyze the components influencing cryptocurrency investment choices and determine the coherence of the components influencing cryptocurrency investment choices. The sample group in this research is people who have previously invested. 360 digital currency users using the tool as a questionnaire. The statistics were used to analyze the data such as frequency, percentage, mean, standard deviation. Exploratory component analysis and corroborative component analysis

The results showed that in the exploratory component analysis There are 5 elements that affect the decision to invest in cryptocurrencies: social component, return component. Service component Cryptocurrency Components and individual components Confirmative component analysis revealed  $CMIN/DF$  (1.120),  $GFI$  (0.927),  $AGFI$  (0.913),  $CFI$  (0.990),  $RMSEA$  (0.018),  $SRMR$  (0.0380) Show that the theoretical test is consistent with the empirical data.

**Keywords:** Exploratory Factor Analysis,  
Confirmatory Factor Analysis,  
Cryptocurrency



## 1. บทนำ

ในอดีตมนุษย์ได้ใช้สื่อกลางในการแลกเปลี่ยนมาหลายรูปแบบจนมาถึงปัจจุบันที่ใช้เป็นเหรียญกษาปณ์และธนบัตรเป็นสื่อกลางแลกเปลี่ยนที่ใช้กันมาอย่างยาวนาน แต่เนื่องจากเทคโนโลยีได้เปลี่ยนแปลงอย่างรวดเร็ว รวมถึงวงการทางการเงินทำให้เกิดสิ่งที่เรียกว่า เงินดิจิทัล ถูกพัฒนาขึ้นมาเพื่อลดข้อจำกัดต่างๆ ของรูปแบบที่ใช้อยู่ในปัจจุบัน รูปแบบของดิจิทัลแทน การใช้เงินดิจิทัลดังกล่าวมีการซื้อขายในชีวิตประจำวันของมนุษย์ทุกคน เช่น บัตรเครดิต บัตรเติมเงิน นอกเหนือจากเงินดิจิทัลที่เป็นที่รู้จักโดยทั่วไปแล้วยังได้มีสื่อกลางการแลกเปลี่ยนทางดิจิทัลอีกรูปแบบหนึ่งที่ได้รับคามนิยมนิยมอยู่ตอนนี้ นั่นคือ สกุลเงินดิจิทัลหรือคริปโทเคอร์เรนซี [1]

สกุลเงินดิจิทัลหรือคริปโทเคอร์เรนซี เป็นสินทรัพย์ดิจิทัลประเภทหนึ่ง หลายคนอาจมองว่าเป็นสกุลเงินประเภทหนึ่งในอนาคตอาจเข้ามามีบทบาทในการซื้อขายสินค้าหรือแลกเปลี่ยนบริการระหว่างสินทรัพย์ด้วยกัน หลักการทำงานของสกุลเงินดิจิทัลจะถูกบันทึกในระบบที่เรียกว่า บล็อกเชน ช่วยบันทึกข้อมูลว่าใครคือเจ้าของที่แท้จริงของสกุลเงิน รวมไปถึงการทำธุรกรรมระหว่างบุคคลให้มีความน่าเชื่อถือและรับรองความถูกต้องได้ไม่จำเป็นต้องผ่านคนกลางใดๆ [2]

สำหรับในปีที่ผ่านมา 2021 ในวงการสกุลเงินดิจิทัลหรือคริปโทเคอร์เรนซีถือว่ามีเปลี่ยนแปลงไปอย่างมาก ทั้งการพัฒนาเทคโนโลยีและมีผลิตภัณฑ์ใหม่ๆ เกิดขึ้นหลากหลายและยังเป็นปีที่คนสนใจและยอมรับแพร่หลายมากยิ่งขึ้น เนื่องจากคนเริ่มหันมาสนใจทำให้หน่วยงานรัฐเข้ามาดูแลและควบคุมกำหนดกฎเกณฑ์ต่างๆ เพื่อเพิ่มความน่าเชื่อถือและยังทำให้ได้รับความไว้วางใจจากนักลงทุนทั่วทั้งประเทศ จึงทำให้มีนักลงทุนเข้ามาลงทุนในสกุลเงินดิจิทัลมากขึ้นและจะยิ่งมากขึ้นในอนาคต [3]

ในแง่ของนักวิชาการศึกษาเรื่องสกุลเงินดิจิทัลจึงเป็นเรื่องที่น่าสนใจ โดยเริ่มจากที่ศึกษาองค์ประกอบในด้านต่างๆ ว่ามีอะไรบ้างที่มีอิทธิพลให้คนหันมาลงทุนสกุลเงินดิจิทัลและองค์ประกอบเหล่านี้มีความสอดคล้องกับองค์ประกอบเชิงประจักษ์อย่างไร

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 แนวคิดเกี่ยวกับสกุลเงินดิจิทัล

สกุลเงินดิจิทัลมีลักษณะคล้ายสกุลเงินทั่วไป แตกต่างกันตรงที่ไม่มีรูปแบบทางกายภาพชัดเจน ไม่มีการพิมพ์ธนบัตรหรือผลิตเหรียญกษาปณ์ มีการจัดเก็บหรือแลกเปลี่ยนบนระบบคอมพิวเตอร์แบบดิจิทัลโดยใช้เทคโนโลยี บล็อกเชน ตัวอย่างของสกุลเงินดิจิทัล ได้แก่ คริปโทเคอร์เรนซี (cryptocurrency) [6]

### 2.2 การวิเคราะห์องค์ประกอบ (Factor Analysis)

เป็นเทคนิคที่ใช้ในการลดจำนวนตัวแปรเทคนิคหนึ่ง โดยการวิเคราะห์องค์ประกอบมี 2 ประเภท คือ [7]

1. การวิเคราะห์องค์ประกอบเชิงสำรวจ (EFA) การวิเคราะห์องค์ประกอบเชิงสำรวจจะใช้ในกรณีที่ไม่มีความรู้ หรือมีความรู้ น้อยมากเกี่ยวกับ โครงสร้างความสัมพันธ์ของตัวแปรเพื่อศึกษาโครงสร้างของตัวแปรและลดจำนวนตัวแปรที่มีอยู่เดิมให้มีการรวมกันได้

เพื่อศึกษาว่าองค์ประกอบรวมที่จะสามารถอธิบายความสัมพันธ์ร่วมกันระหว่างตัวแปรต่างๆ โดยที่จำนวนองค์ประกอบรวมที่หาได้จะมีจำนวนน้อยกว่าจำนวนตัวแปรนั้น

2. การวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) จะใช้กรณีที่ผู้ศึกษาทราบ โครงสร้างความสัมพันธ์ของตัวแปรหรือคาดว่าตัวแปรใดบ้างที่มีความสัมพันธ์กันมากและควรอยู่ในองค์ประกอบเดียวกัน หรือคาดว่าตัวแปรใดที่ไม่มี ความสัมพันธ์กัน ควรจะอยู่ต่างองค์ประกอบกัน หรือกล่าวได้ว่า ผู้ศึกษาทราบ โครงสร้างความสัมพันธ์ของตัวแปร หรือคาดว่าโครงสร้างความสัมพันธ์ของตัวแปรเป็นอย่างไรและจะใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยันมาตรวจสอบหรือยืนยันความสัมพันธ์ว่าเป็นอย่างไรหรือไม่ โดยการวิเคราะห์หาความตรงเชิงโครงสร้างนั่นเอง

เพื่อต้องการทดสอบสมมุติฐานเกี่ยวกับ โครงสร้างขององค์ประกอบว่า องค์ประกอบแต่ละองค์ประกอบด้วยตัวแปรอะไรบ้าง และตัวแปรแต่ละตัวควรมีน้ำหนักหรืออัตราความสัมพันธ์กับองค์ประกอบมากน้อยเพียงใด [8]

### 3. วิธีการดำเนินงาน

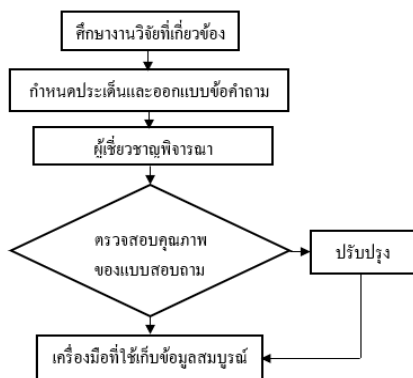
#### 3.1 ประชากรและกลุ่มตัวอย่าง

ประชากร คือ บุคคลที่เลขลงทุนในสกุลเงินดิจิทัล และกลุ่มตัวอย่าง คือ บุคคลที่เลขลงทุนในสกุลเงินดิจิทัลจำนวน 360 คน ซึ่งเป็นขนาดกลุ่มตัวอย่างที่เหมาะสมในการวิเคราะห์องค์ประกอบที่กำหนดตามอัตราส่วน 10-20 ต่อ 1 ตัวแปร [4] เพื่อให้ได้ผลการวิเคราะห์ที่สามารถประมาณค่าได้อย่างถูกต้อง จึงกำหนดกลุ่มตัวอย่างจำนวน 360 กลุ่มตัวอย่าง

#### 3.2 เครื่องมือในการวิจัย

1) เครื่องมือ คือ แบบสอบถาม มีทั้งหมด 2 ตอน ตอนที่ 1 เป็นแบบสอบถามทั่วไป จำนวน 10 ข้อ ตอนที่ 2 เป็นแบบสอบถามที่เกี่ยวกับองค์ประกอบที่มีอิทธิพลต่อการลงทุนสกุลเงินดิจิทัล เป็นคำถามแบบมาตราส่วน ชนิดการประมาณค่า 5 ระดับ จำนวน 30 ข้อ

2) การหาคุณภาพเครื่องมือ โดยผู้เชี่ยวชาญ จำนวน 5 ท่าน ได้ค่าความเที่ยงตรงทุกข้อมีค่ามากกว่า 0.5 และค่าความเชื่อมั่นอยู่ระหว่าง 0.844 ถึง 0.907 สรุปได้ว่าข้อคำถามตรงตามวัตถุประสงค์และมีค่าความเชื่อมั่นอยู่เกณฑ์ที่สามารถนำแบบสอบถามไปใช้เก็บข้อมูลได้

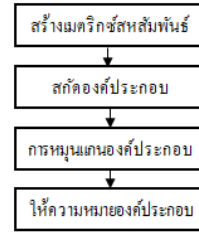


ภาพที่ 1 : ขั้นตอนการตรวจสอบคุณภาพแบบสอบถาม

#### 3.3 การวิเคราะห์องค์ประกอบเชิงสำรวจ

นำข้อมูลที่ได้เก็บจากกลุ่มตัวอย่างนำมาวิเคราะห์องค์ประกอบเชิงสำรวจโดยใช้โปรแกรม SPSS เพื่อสำรวจและระบุองค์ประกอบร่วมที่สามารถอธิบายความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้ โดยสร้างตัวแปรใหม่ในรูปแบบขององค์ประกอบร่วม ใช้ในกรณีที่ยังไม่มีทฤษฎีไหนสนับสนุนชัดเจนการวิเคราะห์องค์ประกอบเชิงสำรวจ

มีทั้งหมด 4 ขั้นตอนดังนี้ การเตรียมเมทริกซ์สหสัมพันธ์, การสกัดองค์ประกอบ, ประมาณค่าพารามิเตอร์และตรวจหาความสอดคล้องของโมเดล และการหมุนแกนองค์ประกอบ

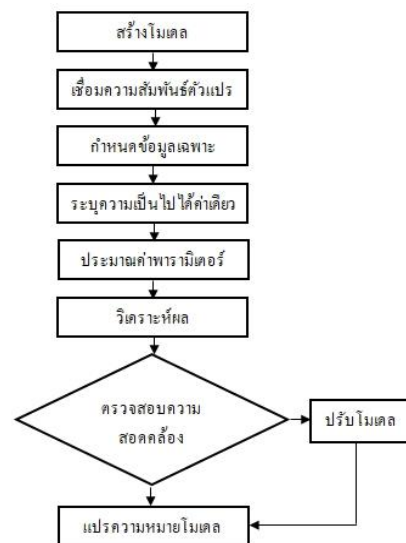


ภาพที่ 2 : ขั้นตอนการวิเคราะห์องค์ประกอบเชิงสำรวจ

#### 3.4 การวิเคราะห์องค์ประกอบเชิงยืนยัน

เป็นขั้นตอนการยืนยันความสอดคล้องของผลการวิเคราะห์องค์ประกอบเชิงสำรวจตามทฤษฎีกับข้อมูลเชิงประจักษ์ โดยใช้โปรแกรมสำเร็จรูป IBM SPSS AMOS 26 สามารถทดสอบความสอดคล้องตามเกณฑ์การพิจารณา ดังนี้ CMIN/DF ไคสแควร์สัมพัทธ์ (<2), GFI ดัชนีความกลมกลืน (>0.90), AGFI ดัชนีความกลมกลืนที่แก้ไขแล้ว (>0.90), CFI ดัชนีความสอดคล้องเปรียบเทียบ (>0.90), RMSEA ดัชนีรากที่สองความคลาดเคลื่อน (<0.08), SRMR ดัชนีรากที่สองของกำลังสองของเศษ (<0.08), Chi square ความน่าจะเป็น (>0.05) [5]

การวิเคราะห์องค์ประกอบเชิงยืนยันมีทั้งหมด 5 ขั้นตอนดังนี้ สร้างโมเดลตามสมมุติฐาน, ประมาณค่าขนาดตัวอย่าง, ประมาณค่าพารามิเตอร์, ตรวจหาความสอดคล้องของโมเดล และปรับปรุงโมเดล



ภาพที่ 3 : ขั้นตอนการวิเคราะห์องค์ประกอบเชิงยืนยัน

4. ผลการดำเนินงาน

4.1 การวิเคราะห์ข้อมูลทั่วไป

รวบรวมข้อมูลที่ได้เก็บแบบสอบถามผ่านช่องทางออนไลน์ ใช้เวลาเก็บข้อมูล 1 เดือน (ธันวาคม 2564) ได้รับแบบสอบถามที่สมบูรณ์ทั้งหมด 360 ชุด ได้ผลสรุป

ข้อมูลทั่วไป		จำนวน	ร้อยละ
เพศ	หญิง	207	57.50%
	ชาย	153	42.50%
รวม		360	100.00%
อายุ	น้อยกว่า 20 ปี	50	13.90%
	21-40 ปี	273	75.80%
	41-60 ปี	36	10%
	61 ปีขึ้นไป	1	0.30%
รวม		360	100.00%
การศึกษา	ต่ำกว่าปริญญาตรี	88	24.40%
	ปริญญาตรีหรือเทียบเท่า	237	65.80%
	สูงกว่าปริญญาตรี	35	9.70%
รวม		360	100.00%
อาชีพ	พนักงานบริษัทเอกชน	151	41.90%
	ข้าราชการ/พนักงานรัฐวิสาหกิจ	100	27.80%
	นักเรียน/นักศึกษา	38	10.60%
	ธุรกิจส่วนตัว	48	13.30%
	รับจ้าง	23	6.40%
รวม		360	100.00%
รายได้	น้อยกว่า 10000 บาท	67	18.60%
	10001-30000 บาท	203	56.40%
	30001-50000 บาท	63	17.50%
	50001 บาทขึ้นไป	27	7.50%
รวม		360	100.00%
ช่องทาง	เพื่อนหรือคนรู้จัก	179	49.70%
	สื่อออนไลน์	131	36.40%
	ค้นพบด้วยตนเอง	50	13.90%
รวม		360	100.00%
ที่มาของเงิน	เงินเดือน	135	37.50%
	เงินออม	144	40.00%
	เงินจากการกู้ยืม	17	4.70%
	เงินจากรายได้เสริมอื่นๆ	64	17.80%
รวม		360	100.00%
จำนวนเงิน	น้อยกว่า 10000 บาท	118	32.80%
	10000-50000 บาท	153	42.50%
	50001-100000 บาท	49	13.60%
	100001 ขึ้นไป	40	11.10%
	รวม		360
ระยะเวลา	ไม่เกิน 6 เดือน	97	26.90%
	6 เดือน - 1 ปี	124	34.40%
	1-3 ปี	70	19.40%
	3-5 ปี	42	12%
	5-10 ปี	18	5.00%
	มากกว่า 10 ปี ขึ้นไป	9	2.50%
รวม		360	100.00%
วัตถุประสงค์	เพื่อการลงทุนระยะยาว	158	43.90%
	เพื่อการลงทุนระยะสั้น	66	18.30%
	เพื่อเป็นรายได้เสริม	102	28.30%
	เพื่อนกระจายความเสี่ยง	34	9.40%
รวม		360	100.00%

ตารางที่ 3 : ตารางการคำนวณค่าเฉลี่ยข้อมูลทั่วไป

4.2 การวิเคราะห์องค์ประกอบเชิงสำรวจ

1) การวิเคราะห์ตัวแปรทั้งหมด 30 ตัวแปร พบได้ว่ามีค่าที่ใช้วัดความเหมาะสมของข้อมูลตัวอย่าง (KMO) เท่ากับ 0.865 มีค่ามากกว่า 0.5 และมีความเข้าใกล้ 1 แสดงถึงเหมาะสมของข้อมูลอยู่ในระดับที่ดีและค่าสถิติที่ใช้ทดสอบสมมติฐาน (Bartlett's Test Of Sphericity) มีค่า

นัยสำคัญที่ 0.05 สรุปได้ว่าข้อมูลที่ได้อาจสามารถนำไปวิเคราะห์องค์ประกอบเชิงสำรวจต่อไป

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling		0.864
Bartlett's Test of	Approx. Chi-Square	5020.288
	df	435
Sphericity	Sig.	0

ตารางที่ 4 : ตาราง KMO and Bartlett's Test Of Sphericity

2) เมื่อหมุนแกนองค์ประกอบพบว่าค่าน้ำหนักตัวแปรทั้งหมด 30 ตัวแปร ที่มีค่ามากกว่า 0.5 แสดงความสัมพันธ์ได้ทั้งหมด 5 องค์ประกอบ แต่ละองค์ประกอบตั้งชื่อตามความเหมาะสมได้ดังนี้ องค์ประกอบด้านสังคม องค์ประกอบด้านผลตอบแทน องค์ประกอบด้านบริการ องค์ประกอบด้านสกุลเงินดิจิทัล และองค์ประกอบด้านตัวบุคคล พบว่าค่าแสดงความสัมพันธ์ที่มีค่ามากที่สุดคือ องค์ประกอบด้านสกุลเงินดิจิทัล มีค่าเท่ากับ 0.883

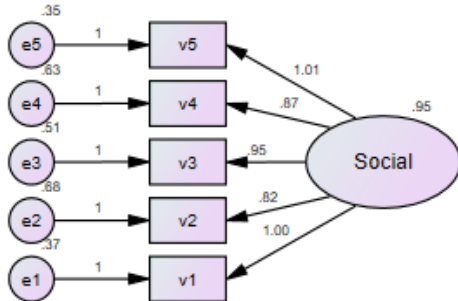
	ตารางหมุนองค์ประกอบเมทริกซ์				
	องค์ประกอบ				
	1	2	3	4	5
V1	0.867				
V2	0.778				
V3	0.839				
V4	0.8				
V5	0.872				
V6		0.825			
V7		0.72			
V8		0.753			
V9		0.791			
V10		0.829			
V11			0.770		
V12			0.586		
V13			0.684		
V14			0.69		
V15			0.712		
V16				0.86	
V17				0.817	
V18				0.849	
V19				0.806	
V20				0.883	
V21					0.732
V22					0.739
V23					0.796
V24					0.707
V25					0.64
V26					0.712
V27					0.599
V28					0.827
V29					0.767
V30					0.836

ตารางที่ 5 : ตาราง Rotated Component Matrix

4.3 การวิเคราะห์องค์ประกอบเชิงยืนยัน

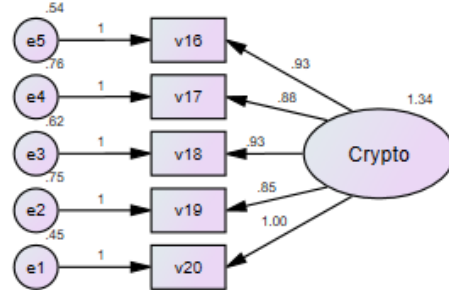
ผลการดำเนินการวิจัยในส่วนของการวิเคราะห์องค์ประกอบเชิงยืนยันในแต่ละด้าน จากการสร้างโมเดลการวัดและทดสอบความสอดคล้องของโมเดล ในโปรแกรม IBM SPSS AMOS 26 ผลสรุปได้ดังนี้

1) โมเดลองค์ประกอบด้านสังคม ดูได้จากค่าความสอดคล้องของโมเดล คือค่า CMIN/DF (1.246), GFI (0.993), AGFI (0.980), CFI (0.999), RMSEA (0.026), SRMR (0.0136), และค่า Chi square (0.284) พบว่าโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์อยู่ในเกณฑ์ดี



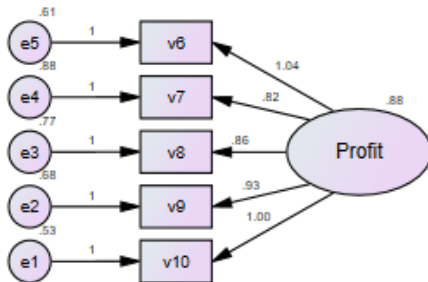
ภาพที่ 4 : โมเดลองค์ประกอบด้านสังคม

4) โมเดลองค์ประกอบด้านสกุลเงินดิจิทัล ดูได้จากค่าความสอดคล้องของโมเดล คือค่า CMIN/DF (1.054), GFI (0.994), AGFI (0.982), CFI (1.000), RMSEA (0.069), SRMR (0.0112), และค่า Chi square (0.384) พบว่าโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์อยู่ในเกณฑ์ดี



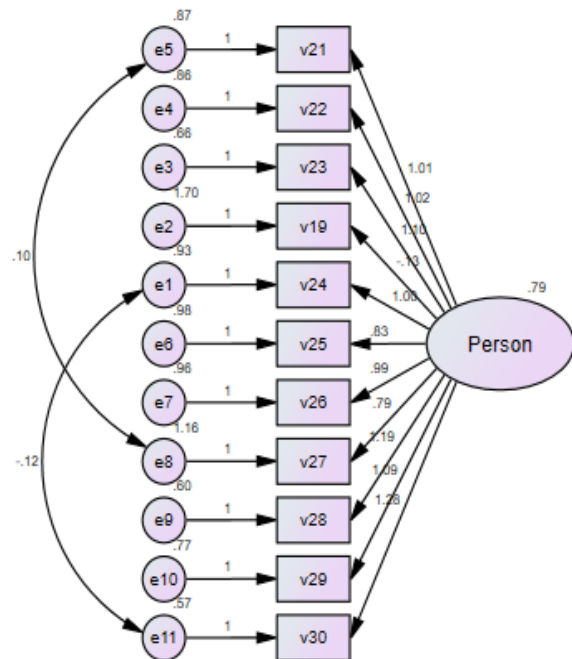
ภาพที่ 7 : โมเดลองค์ประกอบด้านสกุลเงินดิจิทัล

2) โมเดลองค์ประกอบด้านผลตอบแทน ดูได้จากค่าความสอดคล้องของโมเดล คือค่า CMIN/DF (2.689), GFI (0.984), AGFI (0.953), CFI (0.987), RMSEA (0.069), SRMR (0.0295), และค่า Chi square (0.20) พบว่าโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์อยู่ในเกณฑ์ดี



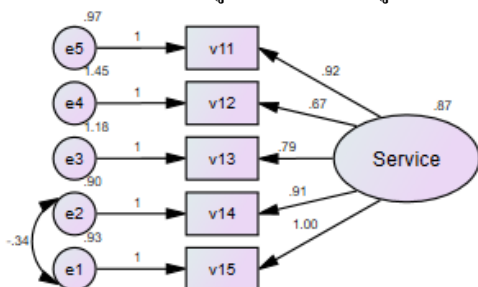
ภาพที่ 5 : โมเดลองค์ประกอบด้านผลตอบแทน

5) โมเดลองค์ประกอบด้านบุคคล ดูได้จากค่าความสอดคล้องของโมเดล คือค่า CMIN/DF (1.351), GFI (0.972), AGFI (0.957), CFI (0.991), RMSEA (0.031), SRMR (0.0258), และค่า Chi square (0.64) พบว่าโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์อยู่ในเกณฑ์ดี



ภาพที่ 8 : โมเดลองค์ประกอบด้านบุคคล

3) โมเดลองค์ประกอบด้านบริการ ดูได้จากค่าความสอดคล้องของโมเดล คือค่า CMIN/DF (1.351), GFI (0.994), AGFI (0.977), CFI (0.996), RMSEA (0.031), SRMR (0.0223), และค่า Chi square (0.248) พบว่าโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์อยู่ในเกณฑ์ดี



ภาพที่ 6 : โมเดลองค์ประกอบด้านบริการ

## 5. สรุปอภิปรายผล

จากการวิเคราะห์ห้องค์ประกอบเชิงสำรวจ พบว่า องค์ประกอบที่มีอิทธิพลต่อการลงทุนสกุลเงินดิจิทัล มีทั้งหมด 5 องค์ประกอบ ได้แก่ องค์ประกอบด้านสังคม, องค์ประกอบด้านผลตอบแทน, องค์ประกอบด้านบริการ, องค์ประกอบด้านสกุลเงินดิจิทัล และองค์ประกอบด้านตัวบุคคล และวิเคราะห์ห้องค์ประกอบเชิงยืนยันสำหรับทดสอบความสอดคล้องของโมเดลตามทฤษฎีกับข้อมูลเชิงประจักษ์ของแต่ละด้านมีดังนี้

องค์ประกอบด้านสังคม สอดคล้องกับกฤษฎี สังกี เพชร (2021) “การตัดสินใจลงทุนในสกุลเงินดิจิทัลฯ” นักลงทุนสนใจโอกาสที่จะได้ผลตอบแทนในระดับสูง แต่ก็มีความเสี่ยงที่สูงเช่นกัน โดยนักลงทุนเองจะต้องพิจารณาถึงความเสี่ยงอื่นๆให้รอบด้าน เป็นพื้นฐานจากการลงทุน [9]

องค์ประกอบด้านผลตอบแทน สอดคล้องกับกฤษฎี สังกี เพชร (2021) “การตัดสินใจลงทุนในสกุลเงินดิจิทัลฯ” ในหลายองค์กรสนใจลงทุนและเก็งกำไรในสินทรัพย์ดิจิทัลกันมากขึ้น โดยการเป็นที่ยอมรับของสังคมรวมไปถึงการเปิดให้ซื้อเงินดิจิทัลบนแพลตฟอร์มระดับโลก [9]

องค์ประกอบด้านบริการ สอดคล้องกับสมใจ ฟองธวัช “นวัตกรรมและเทคโนโลยีทางการเงินในสกุลเงินดิจิทัลฯ” (2019) นักลงทุนตัดสินใจที่จะใช้เงินสกุลดิจิทัล โดยให้ความสำคัญกับความสะดวกรวดเร็วในการใช้งานมากที่สุด ช่วยส่งเสริมการตัดสินใจใช้เงินสกุลดิจิทัล [10]

องค์ประกอบด้านสกุลเงินดิจิทัล สอดคล้องกับสมใจ ฟองธวัช “นวัตกรรมและเทคโนโลยีทางการเงินในสกุลเงินดิจิทัลฯ” (2019) การส่งต่อข้อมูลบันทึกทางธุรกรรมต่างๆที่ถูกเข้ารหัสแล้ว ไม่มีการเปิดเผยตัวตนของบัญชีผู้ใช้งาน เพื่อเป็นการรักษาความมั่นคง ปลอดภัยให้แก่ผู้ใช้งาน [10]

องค์ประกอบด้านตัวบุคคล สอดคล้องกับกฤษฎี สังกี เพชร (2021) “การตัดสินใจลงทุนในสกุลเงินดิจิทัลฯ” นักลงทุนมีการวางแผนก่อนตัดสินใจ โดยให้ความสำคัญกับการศึกษาหาความรู้พื้นฐานของเหรียญสกุลเงินดิจิทัลที่สนใจจะลงทุน [9]

การทดสอบ โมเดล โดยรวมมีค่า CMIN/DF (1.120), GFI (0.927), AGFI (0.913), CFI (0.990), RMSEA (0.018), SRMR (0.0380) ซึ่งค่าที่ได้อยู่ในเกณฑ์ดี มีค่า P-Value (0.51) มีค่ามากกว่าระดับนัยสำคัญที่ .05 แสดงว่าเป็นไปตามสมมุติฐานที่ตั้งไว้ หมายความว่าทฤษฎีมีความสอดคล้องกับข้อมูลเชิงประจักษ์

## เอกสารอ้างอิง

- [1] จิรายุทธ์ ธารารุ่งเรือง และ พัฒน์ พัฒนรังสรรค์, “ปัจจัยที่ส่งผลต่อความสนใจใช้งานสกุลเงินดิจิทัลลิบรา” *วารสารธุรกิจเศรษฐศาสตร์ และการสื่อสาร* ปีที่ 10 (พฤษภาคม - สิงหาคม 2564) หน้า 1-2.
- [2] ไทยรัฐ, (2564) “Cryptocurrency สกุลเงินดิจิทัลที่คนรุ่นใหม่เลือกที่จะลงทุน” <https://www.thairath.co.th/lifestyle/money/> (วันที่สืบค้น 8 กุมภาพันธ์ 2565) หน้า 1.
- [3] ชนาวิ บงวัฒนกุล, (2564) “สรุปเหตุการณ์คริปโตในตลอดช่วงปี 2021” <https://www.tnmtailand.com/news/columnist/> (วันที่สืบค้น 8 กุมภาพันธ์ 2565) หน้า 1.
- [4] Schumackr & Lomax. Hair, “A Beginner's Guide to Structural Equation Modeling” 1996 London: Routledge.
- [5] ชญานี นุธรรม และ สุมาลย์ปาน กา. Hair, “โมเดลความสัมพันธ์เชิงสาเหตุพฤติกรรมการตั้งใจซื้ออาหารฯ” *การประชุมนำเสนอผลงานวิจัยระดับบัณฑิตศึกษา* ปีที่ 18 (สิงหาคม 2563) หน้า 17-18.
- [6] ณทชัย สุขเสนา, “มาตรการกำกับดูแลเงินสกุลดิจิทัลและการปรับใช้กฎหมายไทยกับเงินสกุลดิจิทัล : ปัททอนน์” *วารสาร กสทช.* ปีที่ 1 (ธันวาคม 2560) หน้า 5-6.
- [7] ศศิวิมล ว่องวิไล และ ธนินทร์ รัตน์ โอปาร, “การวิเคราะห์องค์ประกอบของกระบวนการจัดการความรู้ของมหาวิทยาลัยในกำกับของรัฐ” *วารสารวิชาการ Veridian E-Journal มนุษยศาสตร์* ปีที่ 8 (กันยายน - ธันวาคม 2558) หน้า 4-5.
- [8] ปิยะพร อิศรารักษ์ และ ฉลองชัย ชิวสุทรสกุล, “การวิเคราะห์องค์ประกอบเชิงสำรวจที่ส่งผลต่อการเรียนฯ” *การประชุมวิชาการระดับชาติมหาวิทยาลัยวลัยลักษณ์* ปีที่ 9 (15-16 สิงหาคม 2562) หน้า 1-2.
- [9] กฤษฎี สังกี เพชร, (2564) “การตัดสินใจลงทุนในสกุลเงินดิจิทัลฯ” หน้า 12-13.
- [10] สมใจ ฟองธวัช, “นวัตกรรมและเทคโนโลยีทางการเงินในสกุลเงินดิจิทัลฯ” *วารสารเศรษฐศาสตร์และกลยุทธ์การจัดการ* ปีที่ 6 (กรกฎาคม - ธันวาคม 2562) หน้า 55-71.

# จำแนกปัจจัยที่มีผลต่อการตอบสนองของลูกค้าต่อคุณภาพการบริการของผู้ใช้บริการ โรงแรมในช่วงการแพร่ระบาดของเชื้อไวรัสโคโรนา 2019

## Classification Factors Affecting Customer's Responsiveness Toward Service Quality Behavior During Covid-19 Pandemic

ธนพร เทพทิพา (Thanaporn Tepkila)<sup>1</sup> และ ธนพล เจนสุทธิเวชกุล (Tanapon Jentsuttiwetchakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858040@email.kmutnb.ac.th, <sup>2</sup>tanapon.j@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาประสิทธิภาพของอัลกอริทึมที่ใช้ในการจำแนกปัจจัยที่มีผลต่อการตอบสนองของผู้ใช้บริการโรงแรมในช่วงการแพร่ระบาดของเชื้อไวรัสโคโรนา 2019 จากการแสดงความคิดเห็นบนเว็บไซต์ออนไลน์ ด้วยการวิเคราะห์ข้อความและการประมวลผลภาษาธรรมชาติ ซึ่งจะใช้เทคนิคโครงสร้างประโยคในภาษาไทย เพื่อการค้นหาประโยคและกลุ่มคำภายในข้อความแสดงความคิดเห็น โดยการค้นหากลุ่มคำนั้นจะใช้การวัดระยะห่างของส่วนขยายที่อยู่ใกล้กับคุณลักษณะในปัจจัยเดียวกันมากที่สุด โดยมีปัจจัยที่ใช้ในการวิเคราะห์ข้อมูลทั้งหมด 5 ปัจจัย ได้แก่ ทำเลสถานที่ตั้ง ความสะอาดและห้องพัก ความพร้อมของสิ่งอำนวยความสะดวก การบริการของพนักงาน และความคุ้มค่าของราคา

ผลจากการจำแนกข้อความแบบหลายป้ายกำกับโดยใช้การวัดระยะห่างระหว่างคำคุณลักษณะและส่วนขยายของคำนั้นสามารถสรุปได้ว่าอัลกอริทึมมีประสิทธิภาพในการจำแนกโดยมีค่าความถูกต้องของปัจจัยด้านทำเลสถานที่ตั้งอยู่ที่ 95.50% ด้านความสะอาดและห้องพักอยู่ที่ 89.50% ด้านความพร้อมของสิ่งอำนวยความสะดวกอยู่ที่ 92.50% ด้านการบริการของพนักงานอยู่ที่ 95.00% และด้านความคุ้มค่าของราคาอยู่ที่ 95.00% ตามลำดับ

**คำสำคัญ:** เชื้อไวรัสโคโรนา 2019, การประมวลผลภาษาธรรมชาติ

### Abstract

This research aims to develop an algorithm to classify user's opinions using responses of hotel guests during the corona virus disease 2019 epidemic. Text Analytics and Natural Language Processing, which uses sentence structure techniques in Thai to search for sub-sentences and word groups within the comment text were used in data analytic process. Moreover, the closet distance was used to identify user's opinions about related factor. The factors used in analytics consist of location, cleanliness and room, availability of facilities, service, and price.

The results indicated that the algorithm is efficient in classifying with the accuracy of location factor is 95.50%, the cleanliness and room factor is 89.50%, the availability of facilities factor is 92.50%, the service factor is 95.00%, and the price factor is 95.00%, respectively.

**Keyword:** COVID-19, Text Analytics, Natural Language Processing

### 1. บทนำ

อุตสาหกรรมการท่องเที่ยวของประเทศไทยมีบทบาทสำคัญไม่น้อยในระบบเศรษฐกิจ ซึ่งธุรกิจที่เกี่ยวข้องกับอุตสาหกรรมท่องเที่ยวขึ้นเป็นธุรกิจที่สามารถทำรายได้ให้กับประเทศได้จำนวนมากมหาศาล แต่ด้วยปัญหา

การแพร่ระบาดของเชื้อไวรัสโคโรนา 2019 ที่เกิดขึ้นในหลายปีที่ผ่านมา ทำให้หลายธุรกิจในระบบอุตสาหกรรมการท่องเที่ยวได้รับผลกระทบอย่างรุนแรงจากการที่ไม่สามารถเปิดรับนักท่องเที่ยวชาวต่างชาติเข้ามาภายในประเทศได้ ส่งผลให้ผู้ประกอบการจำเป็นต้องเร่งปรับตัวเพื่อให้สามารถดำเนินธุรกิจต่อไปได้อย่างยั่งยืน อาทิ การนำเทคโนโลยีเข้ามาช่วยในการวิเคราะห์ปัจจัยที่เกี่ยวข้องในการเลือกซื้อบริการห้องพักโรงแรมของนักท่องเที่ยวที่อาจจะเปลี่ยนแปลงไปจากเดิม เพื่อนำมาปรับปรุงและนำเสนอการบริการที่ตอบสนองต่อความต้องการผู้ใช้บริการได้มากยิ่งขึ้น[1]

ซึ่งในปัจจุบันการจำแนกข้อมูลหลายหมวดหมู่ในหนึ่งข้อความนั้นยังมีความคลาดเคลื่อนอยู่มาก โดยวิธีการใช้เทคนิคการวิเคราะห์ข้อความด้วยการประมวลผลภาษาธรรมชาตินั้นเป็นวิธีหนึ่งที่ได้รับการนิยมนิยม แต่เนื่องจากภาษาไทยที่ไม่ใช่ภาษาทางกรนั้นไม่มีรูปแบบโครงสร้างประโยคที่แน่นอนหรือไม่เป็นไปตามหลักของภาษาที่ถูกต้อง ทำให้ในบางครั้งการประมวลผลข้อความจึงเกิดความคลาดเคลื่อน จึงจำเป็นต้องศึกษาหากระบวนการที่สามารถนำมาใช้ในการจำแนกข้อความที่มีประสิทธิภาพมากยิ่งขึ้น

จากปัญหาดังกล่าวข้างต้น ผู้วิจัยจึงได้ทำการวิเคราะห์ข้อความแสดงความคิดเห็นบนเว็บไซต์ออนไลน์ของผู้เข้าใช้บริการโรงแรมในช่วงการแพร่ระบาดของเชื้อไวรัสโคโรนา 2019 โดยการใช้การประมวลผลภาษาธรรมชาติสำหรับจำแนกข้อความ ซึ่งจะใช้รูปแบบโครงสร้างประโยคที่เป็นหน่วยย่อยที่สุดในการจำแนกหมวดหมู่ ได้แก่ 1) ประโยคความเดียว(นาม+กริยา+ส่วนขยาย) ตัวอย่างเช่น พนักงาน+พูดจา+ไม่ดี และ 2) กลุ่มคำ(นาม+ส่วนขยาย) ตัวอย่างเช่น วิว+สวย เป็นต้น โดยจะใช้วิธีการวิเคราะห์ห่างระหว่างคำนามและส่วนขยายที่อยู่ในหมวดหมู่เดียวกันเพื่อทำการจับคู่คำและวิเคราะห์ความหมายเชิงบวกเชิงลบ

## 2. หลักการและทฤษฎีที่เกี่ยวข้อง

### 2.1 ปัจจัยที่มีผลต่อการเลือกใช้บริการที่พัก

ปัจจัยที่ส่งผลต่อการตัดสินใจในการเลือกใช้บริการที่พักนั้น ถูกนำมาใช้เพื่อเป็นประโยชน์สำหรับผู้ประกอบการธุรกิจที่พัก และผู้ที่สนใจลงทุน โดยนำมาประยุกต์ใช้สำหรับการเตรียมความพร้อม และวางแผนกลยุทธ์ให้สอดคล้องกับความต้องการของผู้บริโภคในปัจจุบัน โดยปัจจัยที่มีผลต่อการตัดสินใจเลือกใช้บริการนั้นสามารถเรียงจากมากไปน้อย ดังนี้ ปัจจัยด้านสิ่งอำนวยความสะดวกภายในห้องพัก ปัจจัยด้านบุคลากร ปัจจัยด้านราคา ปัจจัยด้านความสะดวกสบายรอบๆของห้องพัก และปัจจัยด้านสถานที่ [2]

### 2.2 การประมวลผลข้อความ (Text Processing)

การประมวลผลข้อความนั้นคือการช่วยทำให้อุปกรณ์คอมพิวเตอร์สามารถทำความเข้าใจและตีความภาษาของมนุษย์ได้ ไม่ว่าจะเป็นภาษาพูด ภาษาเขียน หรือแม้แต่คำสแลง เนื่องจากภาษาที่มนุษย์ใช้กันนั้น มีความซับซ้อนและหลากหลาย อีกทั้งยังมีความรู้สึกแฝงที่อยู่ภายในข้อความ ดังนั้นจึงจำเป็นต้องมีเทคนิคต่างๆ เพื่อช่วยในการวิเคราะห์ใจความสำคัญของข้อความออกมาได้อย่างถูกต้อง ยกตัวอย่างเช่น การใช้เทคนิคการประมวลผลภาษาธรรมชาติเข้ามาช่วยในการประมวลผลข้อความเพื่อลดความสับสนในการวิเคราะห์ภาษาลง[3] โดยการใช้การประมวลผลภาษาธรรมชาติ มีหลักการทำงานดังนี้ 1) การวิเคราะห์ทางองค์ประกอบ 2) การวิเคราะห์ทางไวยากรณ์ 3) การวิเคราะห์ทางความหมาย 4) นูรณ์การทางวจนิพนธ์ 5) การวิเคราะห์ทางปฏิบัติ

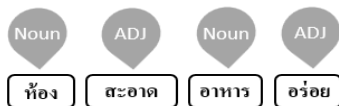
### 2.3 การประมวลผลจำแนกหมวดหมู่ข้อความ

การตัดคำ (Word Segmentation) เนื่องจากลักษณะการเขียนภาษาไทยมีการเขียนติดต่อกันโดยไม่มีเครื่องหมายวรรคตอนในการแสดงการแบ่งคำเหมือนกับภาษาอังกฤษ จึงเป็นอุปสรรคอย่างหนึ่งในการแบ่งคำ โดยปัจจุบันมีลักษณะการตัดคำแบ่งเป็น 3 กลุ่มหลัก คือ การตัดคำโดยใช้กฎ (Rule-Based Approach) การตัดคำโดยใช้พจนานุกรม (Dictionary-Based Approach) และการตัดคำโดยใช้คลังข้อมูล (Corpus-Based Approach) [4] ซึ่งในงานวิจัยนี้จะใช้อัลกอริทึม Dictionary-Based

Longest-Matching ซึ่งเป็นการตัดคำโดยการเปรียบเทียบกับคำที่ยาวที่สุดที่มีอยู่ในพจนานุกรม

การกำจัดคำหยุด (Stop-Word Removal) เป็นการนำคำที่ไม่มีความหมายสำคัญกับการวิเคราะห์ข้อความออก ซึ่งเมื่อนำออกแล้วต้องไม่ทำให้ความหมายเปลี่ยนไปจากเดิม โดยคำหยุดมักเป็นคำที่ปรากฏขึ้นบ่อยครั้ง จึงถือได้ว่าเป็นคำที่ไม่เกี่ยวข้องหรือไม่มีประโยชน์ในการจำแนกหมวดหมู่ [5] ยกตัวอย่างเช่น ดังเก่า, หากว่า, แห่งนี้, จ๊ะ, เสียจนกระทั่ง เป็นต้น ซึ่งจากตัวอย่างจะเห็นว่าการตัดคำว่า ‘เสียจนกระทั่ง’ ออกจากประโยคนั้นไม่ได้ทำให้ความหมายเปลี่ยนไปแต่อย่างใด

การกำกับหน้าที่ของคำ (POS Tagging) เป็นการแปะป้ายหน้าที่ของคำเพื่อใช้ในการบ่งบอกหน้าที่ของคำในประโยค ซึ่งหน้าที่ของคำสามารถแบ่งได้หลายประเภท ยกตัวอย่างเช่น คำนาม คำกริยา คำวิเศษณ์ เป็นต้น



ภาพที่ 1 : ตัวอย่างการแปะป้ายกำกับหน้าที่ของคำ

การสกัดคุณลักษณะ (Feature Extraction) เป็นการคัดเลือกเอาคำคุณลักษณะที่เหมาะสมสำหรับการพิจารณาจำแนกบทความ เพื่อนำมาจับใจความสำคัญของความหมายที่แท้จริงในบทความ อย่างเช่นการค้นหาความหมายที่แท้จริงของการแสดงความคิดเห็นของลูกค้าที่ซื้อผลิตภัณฑ์ออนไลน์ โดยการสกัดเอาคำคุณลักษณะของลูกค้าที่แสดงความคิดเห็นและความคิดเห็นนั้นเป็นบวกหรือลบเท่านั้น [6][7]

2.4 การวิเคราะห์ทัศนคติ (Sentiment Analysis)

คือการนำข้อความแสดงความคิดเห็น หรือบทวิจารณ์มาเข้าสู่กระบวนการประมวลผลเพื่อวิเคราะห์หาอารมณ์ความรู้สึกที่ผู้แสดงความคิดเห็นต้องการจะสื่อออกมาผ่านข้อความ โดยอาจจะวิเคราะห์ออกมาว่าความคิดเห็นนั้นเป็นความคิดเห็นเชิงบวกหรือความคิดเห็นเชิงลบ ซึ่งในปัจจุบันมีผู้ประกอบการหลายรายที่ใช้เทคโนโลยีในการวิเคราะห์ทัศนคติของผู้บริโภคเพื่อนำ

ข้อมูลที่ได้ไปทำการปรับปรุงแก้ไข และพัฒนาสินค้าและบริการของตนเองให้ตรงกับความต้องการของท้องตลาด

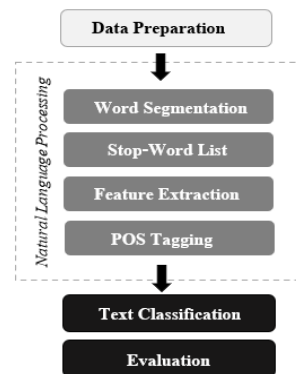
2.5 การจำแนกประเภทแบบหลายป้ายกำกับ (Multi-Label Classification)

เป็นการจำแนกประเภทของข้อมูลที่มีอาจจะมากกว่าหนึ่งคำตอบ ยกตัวอย่างเช่น การจำแนกว่ารูปภาพที่กำลังพิจารณาอยู่นั้นมีองค์ประกอบอะไรบ้าง ในรูป ซึ่งองค์ประกอบที่ใช้ในการพิจารณามีดังนี้ รูปสามเหลี่ยม รูปดอกไม้ และรูปหัวใจ จากภาพตัวอย่างจะเห็นได้ว่ารูปภาพหนึ่งภาพนั้นสามารถมีองค์ประกอบได้มากกว่า 1 องค์ประกอบในการจำแนก



ภาพที่ 2 : ตัวอย่างการจำแนกประเภทหลายป้ายกำกับ

3. การดำเนินการงานวิจัย



ภาพที่ 3 : ขั้นตอนการดำเนินงาน

3.1 การรวบรวมและการเตรียมข้อมูล

งานวิจัยขั้นนี้ได้ทำการรวบรวมข้อมูลการแสดงความเห็นจากผู้เข้าใช้บริการที่ผ่านการจองบนเว็บไซต์ ตั้งแต่ช่วงเดือนมกราคม ปีพุทธศักราช 2562 จนถึง ปีพุทธศักราช 2565 ซึ่งอยู่ในช่วงระยะเวลาที่มีการแพร่ระบาดของเชื้อ โควโรนา 2019 โดยใช้ซอฟต์แวร์ WebHarvy ในการรวบรวมข้อมูลจากเว็บไซต์มาเก็บไว้ในไฟล์รูปแบบตาราง



การเตรียมข้อมูลในที่นี้เป็นการนำข้อมูลที่นำมาทำ ความเข้าใจด้วยการประมวลผลโดยมนุษย์ จากนั้นจึงทำ การจำแนกปัจจัยต่างๆที่ถูกกล่าวถึงในแต่ละข้อความ แสดงความคิดเห็น ออกเป็น 5 ปัจจัยด้วยกัน ได้แก่ ทำเล สถานที่ตั้ง ความสะอาดและห้องพัก ความพร้อมของสิ่ง อำนวยความสะดวก การบริการของพนักงาน และความ คุ่มค่าของราคา ซึ่งในหนึ่งข้อความแสดงความคิดเห็น นั้นสามารถมีปัจจัยที่เกี่ยวข้องได้มากกว่าหนึ่งปัจจัย ยกตัวอย่างเช่น “ห้องเก่าไปหน่อย แต่คุ้มค่างับราคาที่ จ่าย” สามารถจำแนกได้ 2 ปัจจัยด้วยกัน คือ 1) ปัจจัยด้าน ความสะอาดและห้องพัก 2) ปัจจัยด้านความคุ้มค่าของ ราคา

### 3.2 การประมวลผลภาษาธรรมชาติ

เริ่มจากการตัดคำ (Word Segmentation) โดยการ แบ่งข้อความออกเป็นคำ ด้วยเทคนิคการเทียบคำที่ยาว ที่สุด ซึ่งจะเลือกคำที่ยาวที่สุดในพจนานุกรมมาใช้และ ทำการเพิ่มคำที่มีผลต่อการประมวลผลเข้าใน พจนานุกรม ยกตัวอย่างจากข้อความ “ห้องเก่าไปหน่อย แต่คุ้มค่างับเงินที่จ่าย เหมาะสำหรับแวะพักระหว่างทาง” จะสามารถตัดคำออกมาได้เป็น “ห้อง |เก่า| ไป | หน่อย | แต่ | คุ้มค่า | กับ | เงิน | ที่ | จ่าย | เหมาะสำหรับ | แวะ | พัก |ระหว่างทาง”

จากนั้นจึงทำการกำจัดคำหยุด (Stop-Word List Removal) ซึ่งเป็นคำที่พบเจอได้บ่อยในข้อความแต่ไม่มี ผลต่อการนำมาประมวลผล

การสกัดคำคุณลักษณะ (Feature Extraction) ใน งานวิจัยชิ้นนี้จะเป็นการสกัดเอาคำคุณลักษณะของปัจจัย ที่มีผลต่อการตอบสนองของผู้ใช้บริการด้วยกัน ทั้งหมด 5 คุณลักษณะ และได้ทำการสกัดคำที่ทำหน้าที่ เป็นส่วนขยายสำหรับใช้บ่งบอกความรู้สึกขั้วบวกและ ความรู้สึกขั้วลบในแต่ละคุณลักษณะออกมา โดยเทคนิค การสกัดคำที่ใช้เป็นลักษณะของ handcrafted features คือ ผู้วิจัยกำหนดคำคุณลักษณะที่ต้องการในแต่ละกลุ่มด้วย ตนเอง จากข้อความแสดงความคิดเห็นบนเว็บไซต์ จำนวน 1,000 ข้อความ โดยมีตัวอย่างคำคุณลักษณะที่ สกัดได้ ดังตารางที่ 1

ตารางที่ 1: แสดงตัวอย่างการสกัดคำคุณลักษณะ

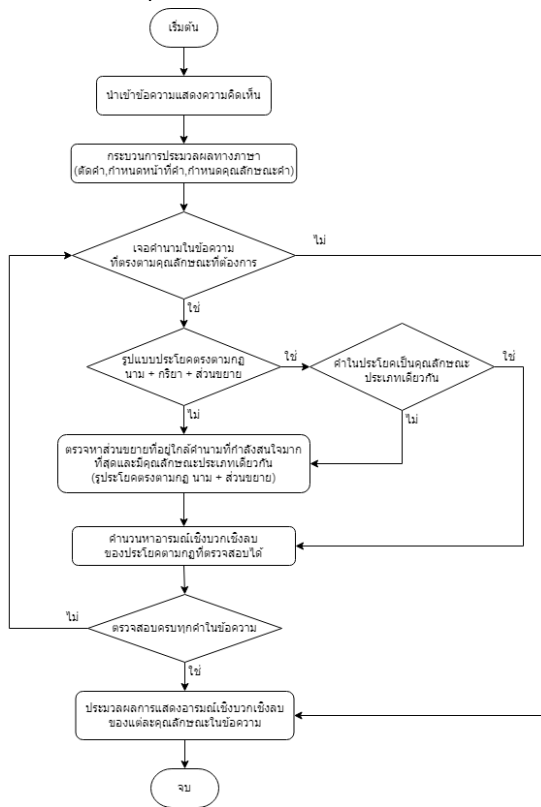
ทำเลสถานที่ตั้ง	
คำคุณลักษณะ	ท่าเล, การเดินทาง, ชายหาด, ห้อง
ส่วนขยาย	ใกล้(+), ไกล(-),สวย(+), สงบ(+), ชับซ้อน (-), รุนววย(-), สะดวก(+)
ความสะอาดและห้องพัก	
คำคุณลักษณะ	การตกแต่ง, ระเบียบ, ห้องพัก, ความ สะอาด, ห้องน้ำ, พื้น, เพดาน, ผนัง
ส่วนขยาย	ตรงปก(+), ทันสมัย(+), สะอาด(+), กว้างขวาง(+), สกปรก(-), โทรม(-)
ความพร้อมของสิ่งอำนวยความสะดวก	
คำคุณลักษณะ	ที่จอดรถ, สระว่ายน้ำ, ผ้าเช็ดตัว
ส่วนขยาย	สะดวก(+), อร่อย(+), เหม็นอับ(-), หลากหลาย(+), พัง(-), น้อย(-)
การบริการของพนักงาน	
คำคุณลักษณะ	การบริการ, พนักงาน, แม่บ้าน
ส่วนขยาย	ยิ้มแย้ม(+), เต็มใจบริการ(+), เสียใจ(-), ละเลย(-), เสียใจ(-)
ความคุ้มค่าของราคา	
คำคุณลักษณะ	ราคา, เงิน, ค่าใช้จ่าย, โปรโมชั่น, ส่วนลด
ส่วนขยาย	ถูก(+), แพง(-), ช่อมเยา(+), คุ้ม(+)

การกำกับคำ (POS Tagging) เป็นการระบุหน้าที่ของ คำแต่ละคำ ซึ่งในงานวิจัยชิ้นนี้จะสนใจคำที่มีหน้าที่ของ คำเป็นคำนาม คำกริยา คำขยาย และคำปฏิเสธ เป็นหลัก

### 3.3 การจัดกลุ่มคำเพื่อจำแนกข้อความ

เป็นการจำแนกข้อความโดยการวิเคราะห์ โครงสร้างของประโยค ซึ่งในที่นี้แบ่งเป็น 2 กรณี คือ โครงสร้างที่อยู่ในรูปของประโยค คือ คำนาม+คำกริยา+ ส่วนขยาย และโครงสร้างที่อยู่ในรูปของกลุ่มคำ คือ คำนาม+ส่วนขยาย โดยเลือกจากส่วนขยายที่มี คุณลักษณะเดียวกันและอยู่ใกล้กับคำคุณลักษณะที่ พิจารณาอยู่มากที่สุด มาระบุว่ากลุ่มคำนั้นมีความคิดเห็น เชิงบวกหรือลบ ตัวอย่างเช่น “ห้องเก่าไป แต่คุ้มค่างับ ราคาที่จ่าย” ซึ่งจะเห็นว่าคำคุณลักษณะ ‘ห้อง’ เป็น คำนามที่อยู่ใกล้กับส่วนขยายที่มีคุณลักษณะเดียวกันคือ ‘เก่า’ เป็นคำเชิงลบ และคำคุณลักษณะ ‘ราคา’ เป็น คำนามที่อยู่ใกล้กับส่วนขยายที่มีคุณลักษณะเดียวกันคือ ‘คุ้มค่า’ เป็นคำเชิงบวก ดังนั้นจึงสามารถจำแนกได้ว่า ความคิดเห็นนี้มีความรู้สึกเชิงลบกับปัจจัยทางด้านความ

สะอาดและห้องพัก แต่มีความรู้สึกเชิงบวกกับปัจจัยทางด้านความคุ้มค่าของราคา



ภาพที่ 4 : ฟังงานแสดงกระบวนการจำแนกข้อความ

### 3.4 การประเมินผล

ค่าความถูกต้อง (Accuracy) จะสามารถบอกได้ว่าสามารถคาดหวังผลลัพธ์ในการจำแนกได้กี่เปอร์เซ็นต์

ค่าความแม่นยำ (Precision) จะสามารถบอกได้ว่าจากการจำแนกที่คาดการณ์ว่าเป็นบวกมีกี่จุดที่เป็นบวกจริงๆ

ค่าความระลึก (Recall) เป็นการเช็ความถูกต้องของการจำแนก

ค่าประสิทธิภาพโดยรวม (F1-measure) เป็นการนำค่าความแม่นยำและค่าความระลึกมาเฉลี่ยหาประสิทธิภาพ

โดยมีสูตรในการคำนวณ ดังนี้

TP คือ ข้อมูลที่ทำนายแล้วถูกต้องเมื่อเทียบกับเฉลย

TN คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสไม่ได้สนใจ

FP คือ ข้อมูลที่อยู่ในเฉลยแต่ไม่มีในการทำนาย

FN คือ ข้อมูลที่ทำนายไม่ถูกต้องเมื่อเทียบกับเฉลย

$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

$$Precision = TP/(TP+FP)$$

$$Recall = TP/(TP+FN)$$

$$F1\text{-measure} = 2 * [(Precision * Recall) / (Precision + Recall)]$$

## 4. ผลการดำเนินงานวิจัย

จากการจำแนกอารมณ์เชิงบวกเชิงลบของคุณลักษณะต่างๆจำนวน 5 คุณลักษณะ โดยใช้ข้อความแสดงความคิดเห็นจากเว็บไซต์จำนวน 200 ข้อความ พบว่าเมื่อผ่านกระบวนการวิเคราะห์ข้อความแล้วมีจำนวนข้อความที่สามารถจำแนกอารมณ์เชิงบวกเชิงลบของคุณลักษณะในข้อความออกมาได้ถูกต้องดังต่อไปนี้ ปัจจัยด้านทำเลสถานที่ตั้งถูกต้อง 191 ข้อความ ปัจจัยด้านความสะอาดและห้องพักถูกต้อง 179 ข้อความ ปัจจัยด้านความพร้อมของสิ่งอำนวยความสะดวกถูกต้อง 185 ข้อความ ปัจจัยด้านการบริการของพนักงานถูกต้อง 190 ข้อความ และปัจจัยด้านความคุ้มค่าของราคาถูกต้อง 190 ข้อความ ซึ่งตัวอย่างผลลัพธ์ที่ได้จากการจำแนกนั้นเป็นไปตาม ภาพที่ 6 โดยกำหนดให้ค่า '0' แทนการไม่แสดงอารมณ์ต่อคุณลักษณะหรือแสดงอารมณ์เป็นกลาง ค่า '1' แทนการแสดงอารมณ์ต่อคุณลักษณะนั้นๆ ในเชิงบวกและค่า '-1' แทนการแสดงอารมณ์ต่อคุณลักษณะนั้นๆ ในเชิงลบ

review	Result				
	location	service	facility	worthiness	room
ห้องกว้างขวางดีมาก อีแคร์เน็คดี ห้องสะอาดพอใช้ อาหารเช้าอร่อย เกา และไกลกว่าที่คิด ห้องที่อยู่ ส่วนจะเสียบ กด ชักโครก จากใช้เวลานานจนกลายเป็นกดไม่ลง	0	0	1	0	1
พนักงานทุกแผนก บริการดีมากค่ะ ห้องพักสะอาด อาจมีบางอย่างบกพร่องแต่ได้รับการแก้ไขไปโดยดีด้วยดีค่ะ อาหารชาติตะ โดยเฉพาะพนักงานแผนกกาแฟน่ารักมาก สำหรับการบริการ	-1	0	0	0	-1
	0	1	1	0	1

ภาพที่ 5 : ผลลัพธ์ที่ได้จากการจำแนกข้อความ

ดังนั้นจึงสามารถคำนวณหาประสิทธิภาพในการจำแนกได้ตามตารางที่ 2

ตารางที่ 2: ตารางแสดงประสิทธิภาพในการจำแนก

ปัจจัยด้านทำเลสถานที่ตั้ง			
Accuracy	Precision	Recall	F1-measure
0.9550	0.9576	0.9550	0.9292
ปัจจัยด้านความสะอาดและห้องพัก			
Accuracy	Precision	Recall	F1-measure
0.8950	0.8961	0.8950	0.8947
ปัจจัยด้านความพร้อมของสิ่งอำนวยความสะดวก			
Accuracy	Precision	Recall	F1-measure
0.9250	0.9283	0.9250	0.9138

ปัจจัยด้านการบริการของพนักงาน			
Accuracy	Precision	Recall	F1-measure
0.9500	0.9588	0.9500	0.9097
ปัจจัยด้านความคุ้มค่าของราคา			
Accuracy	Precision	Recall	F1-measure
0.9500	0.9522	0.9500	0.8516

จากตารางที่ 2 จะเห็นได้ว่างานวิจัยชิ้นนี้มีประสิทธิภาพในการจำแนกอารมณ์เชิงบวกและอารมณ์เชิงลบของแต่ละคุณลักษณะดังนี้

ด้านทำเลสถานที่ตั้งมีค่าความแม่นยำในการจำแนกอยู่ที่ 95.50% ค่าความแม่นยำ 95.76% % ค่าความระลึก 95.50% และค่าประสิทธิภาพโดยรวม 92.92%

ด้านความสะอาดและห้องพักรมมีค่าความแม่นยำในการจำแนกอยู่ที่ 89.50% ค่าความแม่นยำ 89.61% ค่าความระลึก 89.50% และค่าประสิทธิภาพโดยรวม 89.47%

ด้านความพร้อมของสิ่งอำนวยความสะดวกมีค่าความแม่นยำในการจำแนกอยู่ที่ 92.50% ค่าความแม่นยำ 92.83% ค่าความระลึก 92.50% และค่าประสิทธิภาพโดยรวม เท่ากับ 91.38%

ด้านการบริการของพนักงานมีค่าความแม่นยำในการจำแนกอยู่ที่ 95.00% ค่าความแม่นยำ 95.88% ค่าความระลึก 95.00% และค่าประสิทธิภาพ โดยรวมเท่ากับ 90.97%

ด้านความคุ้มค่าของราคามีค่าความแม่นยำในการจำแนกอยู่ที่ 95.00% ค่าความแม่นยำ 95.22% ค่าความระลึก 95.00% และค่าประสิทธิภาพโดยรวม 85.16%

## 5. สรุป

งานวิจัยนี้ได้นำข้อความแสดงความคิดเห็นบนเว็บไซต์ออนไลน์ของผู้ใช้บริการโรงแรมในช่วงการแพร่ระบาดของเชื้อไวรัสโคโรนา 2019 มาทำการจำแนกอารมณ์เชิงบวกเชิงลบของแต่ละปัจจัยที่ส่งผลต่อการตอบสนองของผู้ใช้บริการต่อคุณภาพการบริการ โดยทดลองใช้เทคนิคการวิเคราะห์โครงสร้างของประโยคในภาษาไทยมาช่วยในการจำแนกข้อความ ซึ่งสามารถจำแนกข้อความเกี่ยวกับปัจจัยด้านทำเลสถานที่ตั้ง ด้าน

ความสะอาดและห้องพัก ด้านความพร้อมของสิ่งอำนวยความสะดวก ด้านการบริการของพนักงาน และด้านความคุ้มค่าของราคา โดยมีค่าความถูกต้องอยู่ที่ 95.50%, 89.50%, 92.50%, 95.00% และ 95.00% ตามลำดับ ดังนั้นงานวิจัยชิ้นนี้สามารถนำมาวิเคราะห์ความคิดเห็นส่วนใหญ่ที่ผู้เข้ารับบริการมีต่อโรงแรม และนำมาประยุกต์ใช้เป็นเครื่องมือสำหรับช่วยตัดสินใจในการบริหารจัดการคุณภาพของโรงแรมให้มีประสิทธิภาพได้มากยิ่งขึ้น

## เอกสารอ้างอิง

- [1] ชวฤทธิ์ โยศรีคุณ, “ปัจจัยที่มีผลต่อการเลือกพักโรงแรมในเขตพื้นที่ท่องเที่ยวทางธรรมชาติ ในจังหวัดนครราชสีมา,” *วารสารวิชาการมหาวิทยาลัยนอร์ทกรุงเทพ* ปีที่ 10 ฉบับที่ 1 มกราคม-มิถุนายน พ.ศ.2564
- [2] เตือนใจ ศรีชะฎา, “ปัจจัยที่ส่งผลต่อการตัดสินใจเลือกที่พักแรมของนักท่องเที่ยวชาวไทยในเขตพื้นที่เมืองพัทยา,” *วารสารวิจัยและพัฒนา วไลยอลงกรณ์ ในพระบรมราชูปถัมภ์* ปีที่ 12 ฉบับที่ 1 มกราคม-เมษายน พ.ศ.2560
- [3] นงเยาว์ สอนจะโปะ, “รูปแบบการจำแนกกลุ่มข้อความภาษาไทยแบบอัตโนมัติ โดยใช้การเรียนรู้ของเครื่องด้วยเทคนิค Unsupervised Learning ร่วมกับการประมวลผลภาษาธรรมชาติ,” *วารสารวิชาการศรีปทุม ชลบุรี* ปีที่ 14 ฉบับที่ 4 เมษายน-มิถุนายน พ.ศ.2561
- [4] กานดา แผ้วพัฒนากุล และ ปราโมทย์ ลีอนาม, "การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์," *วารสารการจัดการสมัยใหม่* ฉบับที่ 2 กรกฎาคม-ธันวาคม พ.ศ.2556
- [5] พยุง มีสัง นิเวศ จิระวิจิตรชัย และปริญญา สงวนศักดิ์, “การพัฒนาประสิทธิภาพการจัดหมวดหมู่เอกสารภาษาไทยแบบอัตโนมัติ,” *วารสารพัฒนบริหารศาสตร์ NIDA Development Journal* ปีที่ 51 ฉบับที่ 3 กรกฎาคม-กันยายน พ.ศ.2554
- [6] Hu, M. and Liu, B. “Mining and Summarizing Customer Reviews,” In Proceeding of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, USA, 2004.
- [7] สมศักดิ์ ศรีสวาทย์ และ สมัย ศรีสว, “การวิเคราะห์เหมืองความคิดเห็นโดยใช้เทคนิคการสกัดคำ,” *วารสารวิชาการ การประยุกต์ใช้เทคโนโลยีสารสนเทศ* ปีที่ 6 ฉบับที่ 2 กรกฎาคม-ธันวาคม พ.ศ.2563

# การศึกษาปัจจัยที่ส่งผลให้ผู้บริโภคซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ A study of Factors Affecting Consumers' Purchases on the E-Marketplace

ศศิมา ธนภัทร์พิบูล (Sasima Tanapatpiboon)<sup>1</sup> และธนพล เจนสุทธิเวชกุล (Tanapon Jentsuttiwetachakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ<sup>1,2</sup>

S6307021856039@email.kmutnb.ac.th, tanapon.j@itd.kmutnb.ac.th

## บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ศึกษาหาปัจจัยที่ส่งผลให้ผู้บริโภคซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ ได้แก่ ปัจจัยด้านความเชื่อมั่น ความง่ายต่อการใช้งาน ความมีส่วนร่วม ความตั้งใจซื้อ และพฤติกรรมผู้บริโภค การเก็บรวบรวมข้อมูลใช้แบบสอบถามออนไลน์จากผู้บริโภคที่ใช้บริการตลาดกลางพาณิชย์อิเล็กทรอนิกส์จำนวนตัวอย่าง 550 คน และมีการใช้สถิติในการวิเคราะห์ข้อมูลด้วย การวิเคราะห์หองค์ประกอบเชิงสำรวจ และการวิเคราะห์สมการถดถอยพหุคูณ

ผลการศึกษาพบว่าปัจจัยการรับรู้ถึงความง่ายในการใช้งาน และปัจจัยความมีส่วนร่วมมีผลบวกต่อการซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ โดยวิจัยนี้มีค่าความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามร้อยละ 10.9 และ 3 ( $R\text{-Squared}=0.109$  และ  $0.039$ )

**คำสำคัญ:** ตลาดกลางพาณิชย์อิเล็กทรอนิกส์ การซื้อสินค้าออนไลน์ ความตั้งใจซื้อ

## Abstract

The objective of this research was to analyze and study the factors affecting consumers' purchases on the e-commerce marketplace, such as trust, ease of use, engagement, purchase intention and consumer behavior. The data was collected using online questionnaires from 550 consumers who use e-commerce marketplace services, and statistics were calculated from Exploratory Factor Analysis and Multiple Regression.

The results showed that the perceived ease of use and the engagement factor has a positive effect on

purchasing products on the e-commerce marketplace. In this research, the R squared values were 10.9 percent and 3 percent ( $R\text{-Squared}=10.9$  and  $0.039$ ).

**Keywords:** E - marketplace, Online purchasing, Purchase Intention

## 1. บทนำ

การเลือกซื้อสินค้าผ่านช่องทางออนไลน์เริ่มแพร่หลายอย่างรวดเร็ว สืบเนื่องจากสถานการณ์การแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 ตั้งแต่ปี พ.ศ. 2562 มีมาตรการการป้องกันการแพร่ระบาดต่างๆ ที่ส่งผลให้ประชาชนต้องอยู่บ้านกักตัว นำไปสู่พฤติกรรมของผู้บริโภคที่เปลี่ยนแปลงไป ทั้งการทำงาน การซื้อสินค้า การบริโภคที่ต้องหลีกเลี่ยงการออกจากบ้าน ช่องทางออนไลน์จึงเป็นทางเลือกในการปรับพฤติกรรมการใช้ชีวิต และยังผลักดันให้ประชาชนเข้าสู่โลกออนไลน์มากขึ้นกว่าเดิมส่งผลให้ตัวเลขตลาดอีคอมเมิร์ซไทยมีอัตราการเติบโตจากปี พ.ศ. 2562 มูลค่า 163,300 ล้านบาทพุ่งสูงถึง 220,000 ล้านบาท อัตราการเติบโตสูงขึ้นร้อยละ 35 จากปีก่อนหน้า (priza.com, 2563) ทั้งนี้ โดยมีสัดส่วนยอดขายที่เว็บไซต์เจ้าของสินค้า หรือ brand.com ที่ ร้อยละ 15 การค้าขายสินค้าหรือบริการผ่านโซเชียลมีเดียโดยตรง หรือ Social Commerce ร้อยละ 38 ตลาดกลางพาณิชย์อิเล็กทรอนิกส์หรือ Marketplace ร้อยละ 47

จากสถิติแสดงให้เห็นว่าผู้คนให้ความสนใจที่จะซื้อสินค้าในเว็บไซต์เจ้าของสินค้าน้อยและให้ความสนใจที่จะซื้อผ่านโซเชียลมีเดียหรือตลาดกลางพาณิชย์

อิเล็กทรอนิกส์มากกว่า เมื่อตัวเลขสถิติพาณิชย์อิเล็กทรอนิกส์แสดงชัดเจนว่าการดำเนินชีวิตของมนุษย์ในปัจจุบันได้เปลี่ยนแปลงไปและสถิติการเข้าไปซื้อสินค้าในแต่ละช่องทางต่างกัน ภาคธุรกิจจึงจำเป็นต้องปรับเปลี่ยนช่องทางการขายสินค้า ช่องทางการนำเสนอสินค้า ตลอดจนช่องทางการทำการตลาดให้ลูกค้าสนใจ

อย่างไรก็ตามเมื่อการแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 พัฒนากลายเป็นโรคที่ผู้คนคุ้นเคยและสามารถรับมือได้เป็นอย่างดี สถิติแสดงให้เห็นชัดเจนแล้วว่าต่อให้โรคติดเชื้อไวรัสโคโรนา 2019 แพร่ระบาดน้อยลงหรือหมดไป ผู้บริโภคจะยังช้อปออนไลน์ต่อไป เนื่องจากผู้บริโภคได้ตระหนักแล้วว่าไ้ออกจากบ้าน ยังสามารถซื้อสินค้าผ่านช่องทางออนไลน์ได้

ดังนั้นเพื่อให้ภาคธุรกิจสามารถปรับตัวก้าวทันการเปลี่ยนแปลงของเทคโนโลยีและพฤติกรรมผู้บริโภคผู้วิจัยจึงตัดสินใจศึกษาปัจจัยที่มีอิทธิพลต่อการตัดสินใจซื้อสินค้าออนไลน์ผ่านตลาดกลางพาณิชย์อิเล็กทรอนิกส์ (Marketplace)

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ข้อมูลเกี่ยวกับตลาดกลางพาณิชย์อิเล็กทรอนิกส์ (e-Marketplace)

ตลาดกลางพาณิชย์อิเล็กทรอนิกส์ คือ ระบบเทคโนโลยีสารสนเทศ (Information Technology) ที่เปิดบริการในรูปแบบการทำธุรกรรมทางพาณิชย์อิเล็กทรอนิกส์ ซึ่งเป็นหนึ่งในประเภทธุรกิจของ e-Commerce และยังเป็นหนึ่งในช่องทางการเปิดตลาดของผู้ประกอบการที่มีต้นทุนถูก บทบาทในการให้บริการมี 3 รูปแบบคือ

1. เป็นตัวกลางให้ผู้ซื้อ และ ผู้ขาย ได้ทราบถึงโอกาสและทำธุรกิจร่วมกัน
2. เป็นตัวกลางในการแลกเปลี่ยนข้อมูลการบริการสินค้า และการชำระเงินผ่านระบบจากธุรกรรมที่เกิดขึ้น
3. เป็นผู้กำหนดกติกา กฏระเบียบ ในการเข้าใช้บริการของระบบ เช่นการแลกเปลี่ยนสินค้า บริการ และการชำระเงิน เพื่ออำนวยความสะดวกให้แก่ผู้เข้ามาใช้

บริการทั้งผู้ขายและผู้ซื้อให้มีความมั่นใจและความสะดวกในการใช้บริการมากขึ้น

ซึ่งตลาดกลางพาณิชย์อิเล็กทรอนิกส์ (e-Marketplace) เป็นการดำเนินงานที่เปลี่ยนแปลงรูปแบบของการทำธุรกรรมเชิงพาณิชย์การค้าขาย แลกเปลี่ยน และห่วงโซ่อุปทาน (Supply Chain) ของตลาดรูปแบบดั้งเดิม

### 2.2 ทฤษฎีพฤติกรรมผู้บริโภค

พฤติกรรมผู้บริโภค (Consumer behavior) หมายถึงลักษณะการซื้อของผู้บริโภค เช่น ซื้อที่ไหน ซื้อเมื่อใด ซื้อมากน้อยเพียงใด ใครเป็นผู้ซื้อ และใช้มาตรการอะไรในการตัดสินใจซื้อ และพฤติกรรมการใช้ (Using Behavior) หมายถึงลักษณะการบริโภคสินค้าของผู้บริโภค เช่น บริโภคที่ไหน บริโภคกับใครในอัตราส่วนน้อยเพียงใด พฤติกรรมของผู้บริโภค มีลักษณะเป็นกระบวนการประกอบด้วยกิจกรรมที่เป็นขั้นตอน ในการซื้อผลิตภัณฑ์ของผู้บริโภคก่อนตัดสินใจซื้อจะผ่านกระบวนการยืคาวหลายขั้นตอน เริ่มด้วยการแสวงหาข้อมูลเกี่ยวกับผลิตภัณฑ์ที่จะซื้อนั้นอย่างกว้างขวาง วางหลายๆ ตราสินค้า แล้วนำมาเปรียบเทียบและประเมินข้อดีข้อเสียแต่ละตราสินค้า และกิจกรรมอื่นๆ อีกมาก่อนตัดสินใจซื้อขั้นสุดท้าย แต่อย่างไรก็ตามในการซื้อผลิตภัณฑ์และบริการบางชนิด การตัดสินใจอาจกระทำได้อย่างรวดเร็ว เกือบจะทันทีภายหลังจากที่ได้เห็นผลิตภัณฑ์นั้น

โมเดลพฤติกรรมผู้บริโภค (Consumer Behavior Model) เป็นการศึกษาถึงเหตุจูงใจที่ทำให้เกิดการตัดสินใจซื้อผลิตภัณฑ์ โดยมีจุดเริ่มต้นจากการเกิดสิ่งกระตุ้น (Stimulus) ในสภาพแวดล้อมต่างๆ ทำให้เกิดความต้องการ เมื่อสิ่งกระตุ้นนั้นผ่านเข้ามาในความรู้สึกนึกคิดของผู้บริโภค (Buying's Black Box) ซึ่งเปรียบเสมือนกล่องดำที่ผู้ผลิตหรือผู้ขายไม่สามารถคาดคะเนได้ ซึ่งความรู้สึกนึกคิดต่างๆ จะได้รับอิทธิพลต่างๆ ภายในใจผู้บริโภค แล้วจะมีการตอบสนองของผู้บริโภค (Buyer's Response) หรือ การตัดสินใจของผู้บริโภค (Buyer's Purchase Decision) โดยการซื้อ หรือ ไม่ซื้อสรุปแล้วจุดเริ่มต้นของโมเดลนี้อยู่ที่การมีสิ่งกระตุ้นให้เกิดความ

ต้องการก่อน แล้วทำให้เกิดการตอบสนอง ดังนั้นเรียกว่า S - R Theory

**2.3 การรับรู้ประโยชน์ (Perceived Usefulness)**

การรับรู้ถึงความง่ายในการใช้งาน เป็นตัวแปรหลักที่สำคัญของทฤษฎีการยอมรับเทคโนโลยี (Technology Acceptance Model: TAM) เป็นทฤษฎีที่ได้พัฒนาต่อจากทฤษฎีการกระทำด้วยเหตุผล (Theory of Reasoned Action: TRA) ของ Fishbein and Ajzen (1975) โดย Davis (1993) ซึ่งมีจุดประสงค์เพื่อเป็นแบบแผนในการทำนายการยอมรับด้านเทคโนโลยีสารสนเทศ โดยตัวแปรหลักที่สำคัญของ TAM ซึ่งได้แก่ ทศนคติต่อการใช้ (Attitude Toward Using) พฤติกรรมความตั้งใจในการใช้ (Behavioural Intentions to Use) และการใช้จริง (Actual System Use) การรับรู้ว่ามีประโยชน์ (Perceived Usefulness) การรับรู้ความง่ายในการใช้งาน (Perceive Ease of Use) หมายถึงระดับความเชื่อคาดหวังของผู้ที่จะใช้ระบบสารสนเทศว่าระบบดังกล่าวเป็นระบบที่สามารถเรียนรู้ได้ง่ายไม่ต้องใช้ความพยายามอย่างมากในการเรียนรู้ที่จะใช้ระบบหรือในการเข้าใช้ระบบ โดย Davis (1989) ได้นิยามการรับรู้ความง่ายตามคำจำกัดความของคำว่า “ง่าย” และ “ปราศจากความยากหรือความพยายาม”

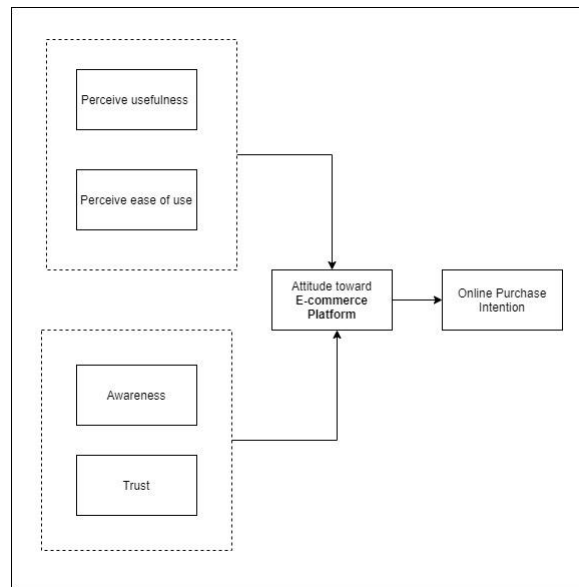
**3. วิธีการดำเนินงาน**

ผู้วิจัยได้ศึกษารวบรวมทฤษฎี งานวิจัยที่เกี่ยวข้อง และแนวทางในการศึกษาและวิเคราะห์ปัจจัยที่ส่งผลให้ผู้บริโภคซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ ผู้วิจัยได้ใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงสำรวจ (Exploratory Factor Analysis - EFA) เทคนิคการหาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ (Multiple Regression) และแบบสอบถามออนไลน์ (Questionnaire) เพื่อนำผลมาวิเคราะห์หาความสำคัญของตัวแปร โดยกำหนดวิธีการวิจัยดังนี้

**3.1 กำหนดกรอบแนวคิดของงานวิจัย**

งานวิจัยเรื่องการศึกษาหาปัจจัยที่ส่งผลให้ผู้บริโภคซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ โดยผู้วิจัยได้วางแผนกรอบแนวคิดในการนำทฤษฎีแบบจำลองการยอมรับเทคโนโลยี (Technology Acceptance Model) มาใช้

ในงานวิจัยดังรูปที่ 1



รูปที่ 1 แบบจำลองปัจจัยที่ส่งผลให้ซื้อสินค้า

**3.2 พัฒนาข้อคำถามออกแบบเครื่องมืองานวิจัย**

**3.2.1 พัฒนาและออกแบบข้อคำถามเพื่อใช้ในการเก็บข้อมูล**

ผู้วิจัยใช้แบบสอบถามออนไลน์เป็นเครื่องมือในการเก็บรวบรวมข้อมูลของกลุ่มเป้าหมาย แบบสอบถามเป็นแบบประเภทปลายปิด โดยการพัฒนาข้อคำถามนั้นผู้วิจัยได้นำมาจากการทบทวนวรรณกรรม งานวิจัย แนวคิดทฤษฎีที่เกี่ยวข้องและ การสัมภาษณ์ผู้เชี่ยวชาญ เพื่อให้คำถามของงานวิจัยนั้น ได้รับคำตอบที่ถูกต้องตรงกับวัตถุประสงค์ที่ตั้งไว้อย่างครบถ้วน ดังตัวอย่างตารางสรุปข้อคำถามแบบสอบถามแบ่งออกเป็น 3 ส่วนดังนี้

ส่วนที่ 1 : ข้อคำถามข้อมูลทั่วไป ได้แก่ เพศ อายุ อาชีพ

ส่วนที่ 2 : คำถามเกี่ยวกับผู้บริโภคที่ซื้อสินค้าออนไลน์บนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ (e-Marketplace)

ส่วนที่ 3 : คำถามเพื่อวิเคราะห์ปัจจัยที่มีผลต่อการตัดสินใจเลือกซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ (e-marketplace) โดยแบ่งออกเป็น 4 หัวข้อหลักดังนี้

1. Perceived Usefulness (PU) การรับรู้ถึงประโยชน์ของการใช้เทคโนโลยี เช่น การสั่งซื้อสินค้าผ่านระบบบนคอมพิวเตอร์หรือโทรศัพท์มือถือได้โดยปราศจากการเดินทางไปที่ร้านค้า
2. Perceived Ease of Use (PE) การรับรู้ความง่ายและสะดวกต่อการใช้งานเทคโนโลยีสารสนเทศ

เช่น การสั่งซื้อสินค้าโดยกดปุ่มสามครั้ง หรือรูปภาพสินค้าที่สามารถลากดูเพื่อความถนัดในการเลือกซื้อสินค้า

3. Engaging (E) การมีส่วนร่วมกับระบบเทคโนโลยีที่ส่งผลให้เกิดความถี่ในการซื้อสินค้า จากโฆษณาบนโซเชียลมีเดีย การลดราคา ความง่ายในการเก็บส่วนลดในระบบเพื่อใช้ในครั้งต่อไป
4. Attitude toward Ecommerce Platform (A) พฤติกรรมของผู้บริโภคที่ต้องซื้อสินค้าตามเทศกาลต่างๆ เปรียบเทียบความคุ้มค่าของราคาสินค้า ขึ้นชอบสินค้าที่จัดโปรโมชั่น
5. Trust (T) ความเชื่อมั่น ต่อการทำงาน ของเทคโนโลยีที่ใช้พัฒนาระบบ เช่นระบบชำระเงิน ระบบเก็บข้อมูล ความปลอดภัยในการเก็บข้อมูลของผู้ใช้บริการ
6. Purchase Intension (I) ความตั้งใจที่จะซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ (e-marketplace) ทั้งในปัจจุบัน อนาคต และการชักชวนผู้อื่นให้มาใช้งาน

### 3.2.2 ทดสอบเครื่องมือในการเก็บข้อมูล

พิจารณาความถูกต้องของข้อมูลโดยผู้เชี่ยวชาญ

### 3.3 วิธีการดำเนินการวิจัย

กำหนดกลุ่มประชากร หรือ กลุ่มเป้าหมายตัวอย่างที่ใช้ในงานวิจัยครั้งนี้ ผู้วิจัยได้ใช้โปรแกรม g\*power เพื่อช่วยในการคำนวณหาขนาดตัวอย่างโดยคำแนะนำจากผู้เชี่ยวชาญ ซึ่งจากผลการคำนวณพบว่าควรมี 371 คำตอบ โดยเป็นกลุ่มประชากรทั่วไปวัยทำงาน

### 3.4 ผลการดำเนินงาน

เมื่อได้คำตอบหรือข้อมูลจากกลุ่มตัวอย่างของงานวิจัยที่ส่งผลให้ผู้บริโภคซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ ผู้วิจัยได้นำข้อมูลดังกล่าวมาวิเคราะห์ด้วยขั้นตอนดังนี้

#### 3.4.1 การวิเคราะห์องค์ประกอบเชิงสำรวจ

(Exploratory Factor Analysis (EFA)) โดยนำข้อมูลที่ได้รับจากแบบสอบถามมาคำนวณตามสูตรในระบบและนำเสนอข้อมูลในรูปแบบดังนี้ ค่าเฉลี่ยขององค์ประกอบ

(Mean), ส่วนเบี่ยงเบนมาตรฐาน (Std. Deviation), ค่าวิเคราะห์องค์ประกอบ (Analysis N) เพื่ออธิบายลักษณะขององค์ประกอบ

#### 3.4.2 การวิเคราะห์หาค่าน้ำหนักขององค์ประกอบ

(Weight Score) โดยการตัดองค์ประกอบเพื่อศึกษาถึงองค์ประกอบร่วมที่แสดงถึงความสัมพันธ์ระหว่างตัวแปรต่างๆ

3.4.3 การวิเคราะห์สถิติเชิงปริมาณ โดยใช้เทคนิคการวิเคราะห์การถดถอยพหุคูณ (Multi Linear Regression)

### 3.5 สรุป

จากการดำเนินงานการศึกษาวิจัยเรื่อง “ปัจจัยที่ส่งผลให้ผู้บริโภคซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์” ผู้วิจัยได้นำผลคำนวณจากขั้นตอนต่างๆ และพบว่าคำถามบางส่วนนั้นกำกวม เนื่องจากผลลัพธ์ยังไม่มีความชัดและไม่มีความสัมพันธ์กับตัวแปรตาม

## 4. ผลการดำเนินงาน

### 4.1 การกำหนดตัวแปรที่ใช้ในการวิจัย (Research Model)

ผู้วิจัยได้กำหนดตัวแปรที่ใช้ในการวิจัยทั้งหมด 5 ตัวแปร ได้แก่ Perceived Ease of Use (PE), Trust, Engaging, Attitude toward E-commerce Platform and Purchase Intension มีค่าความเชื่อมั่น ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานดังนี้ ตารางที่ 1: ผลการวิเคราะห์ความเชื่อมั่น ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

ลำดับ	ปัจจัย	Mean	S.D.	Cronbach's Alpha
1	ปัจจัยความมีประโยชน์ของระบบ	26.49	4.635	0.709
2	ปัจจัยความง่ายในการใช้งานระบบ	26.58	4.696	0.764
3	ปัจจัยการมีส่วนร่วมกับระบบ	19.52	4.498	0.758
4	ปัจจัยพฤติกรรมของผู้บริโภค	21.35	18.642	0.741
5	ปัจจัยความเชื่อมั่น	18.92	4.628	0.805
6	ปัจจัยความตั้งใจซื้อ	16.85	5.985	0.874

4.2 ตัวอย่างแบบสอบถาม

ผู้วิจัยได้พัฒนาแบบสอบถามโดยแบ่งตามองค์ประกอบที่ใช้ในการวิจัย โดยมีรายละเอียดตามตารางที่ 1

ตารางที่ 1 : ผลการวิเคราะห์ข้อมูลทั่วไป

ข้อคำถาม	ข้อมูลทั่วไป	จำนวน (คน)	ร้อยละ (%)
เพศ	• ชาย	305	55.45
	• หญิง	245	44.54
รวม		550	100
อายุ	• น้อยกว่า 20 ปี	15	1.08
	• 21 – 30 ปี	220	40.00
	• 31 – 40 ปี	198	36.00
	• 41 – 50 ปี	89	16.18
	• 51 – 60 ปี	28	5.09
	• มากกว่า 60 ปี	0	0
รวม		550	100
ระดับการศึกษา	• ต่ำกว่าปริญญาตรี	48	8.72
	• ปริญญาตรี	315	57.27
	• ปริญญาโท	170	30.91
	• ปริญญาเอก	17	3.09
รวม		550	100
อาชีพปัจจุบัน	• ธุรกิจส่วนตัว	72	13.09
	• พนักงานบริษัท	265	48.18
	• นักเรียน/นักศึกษา	78	14.18
	• ข้าราชการ	75	13.63
	• อาจารย์	14	2.54
	• อื่นๆ	46	8.36
รวม		550	100

รวมข้อคำถามแบบสอบถามทั้งหมด 21 ข้อที่ใช้ในการเก็บข้อมูลกับกลุ่มประชากรตัวอย่าง 550 คน

4.3 การวิเคราะห์องค์ประกอบเชิงสำรวจ

(Exploratory Factor Analysis (EFA))

1. ค่า KMO (Kaiser-Meyer-Olkin) ใช้อธิบายความเหมาะสมของข้อมูลที่จะนำมาวิเคราะห์ต้องมีค่ามากกว่า 0.05 ได้ผลวิเคราะห์ดังรูปที่ 2

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.813
Bartlett's Test of Sphericity	Approx. Chi-Square	4.003E3
	df	325
	Sig.	.000

รูปที่ 2: KMO (Kaiser-Meyer-Olkin)

2. ค่า Rotated Component Matrix ใช้คำอธิบายการจัดกลุ่มองค์ประกอบโดยผู้วิจัยตัดองค์ประกอบที่เป็นองค์ประกอบร่วมและจัดเข้ากลุ่มได้ดังรูปที่ 3

	Rotated Component Matrix <sup>a</sup>					
	1	2	3	4	5	6
T1		.831				
T2		.783				
T3		.855				
T4		.738				
U1	.769					
U2	.715					
U3	.714					
U4	.689					
U5	.766					
A1			.825			
A2			.726			
A3			.754			
A4			.786			
E1				.684		
E2				.667		
E3				.649		
E4				.641		
E5				.623		
En1						.777
En2						.695
En3						.645
En4						.741
I1					.771	
I2					.654	
I3					.658	
I4					.819	

Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization.  
a. Rotation converged in 6 iterations.

รูปที่ 3: ค่า Rotated Component Matrix

ผลจากตาราง Rotated Component Matrix สามารถสรุปได้ว่า องค์ประกอบที่ 1 Trust (T) ความเชื่อมั่นต่อระบบ ประกอบด้วย 4 ตัวแปร ได้แก่ T1 T2 T3 และ T4 องค์ประกอบที่ 2 Perceive Usefulness (U) ความง่ายในการใช้งานระบบ ประกอบด้วย 4 ตัวแปร ได้แก่ U1 U2 U3 U4 และ U5 องค์ประกอบที่ 3 Attitude toward E-commerce Platform ประกอบด้วย 3 ตัวแปร ได้แก่ A1 A2 A3 และ A4 องค์ประกอบที่ 4 Perceive ease of use (E) ความง่ายในการใช้งานระบบ ประกอบด้วย 4 ตัวแปร ได้แก่ E1 E2 E3 E4 และ E5 องค์ประกอบที่ 5 Engaging (En) การมีส่วนร่วมกับระบบ ประกอบด้วย 3 ตัวแปร ได้แก่ En2 En3 และ En4 องค์ประกอบที่ 6 Purchase Intention (I) ความตั้งใจซื้อสินค้า ประกอบด้วย 4 ตัวแปร ได้แก่ I1 I2 I3 และ I4

4.4 การวิเคราะห์สถิติเชิงปริมาณโดยนำเทคนิคการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression)

ผลที่ได้จากการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) สามารถอธิบายได้ดังนี้



การรับรู้ความง่ายในการใช้งานเทคโนโลยีสารสนเทศ (E) และความมีประโยชน์ (U) ของระบบ ไม่มีผลต่อพฤติกรรมที่มีต่อตลาดกลางพาณิชย์อิเล็กทรอนิกส์ เนื่องจากมีค่า Sig มากกว่า 0.05 ความเชื่อมั่น (T) และความมีส่วนร่วม (En) ต่อตลาดกลางพาณิชย์อิเล็กทรอนิกส์ มีผลต่อพฤติกรรมที่มีต่อตลาดกลางพาณิชย์อิเล็กทรอนิกส์ (A) โดยที่ค่า Sig น้อยกว่า 0.05 ซึ่งเป็นค่าที่ยอมรับได้ในทางสถิติ ดังรูปที่ 4

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.223	.342		6.495	.000
	avg_e	.084	.047	.077	1.786	.075
	avg_t	.255	.040	.263	6.418	.000
	avg_en	.111	.043	.107	2.558	.011
	avg_u	.053	.050	.046	1.057	.291

a. Dependent Variable: avg\_a

รูปที่ 4: Dependent Variable: avg\_a

พฤติกรรมที่มีต่อตลาดกลางพาณิชย์อิเล็กทรอนิกส์ (A) มีผลต่อความตั้งใจในการซื้อสินค้า โดยที่ค่า Sig น้อยกว่า 0.05 ซึ่งเป็นค่าที่ยอมรับได้ในทางสถิติ ดังรูปที่ 5

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.750	.190		19.737	.000
	avg_a	.189	.040	.198	4.731	.000

a. Dependent Variable: avg\_i

รูปที่ 5: Dependent Variable: avg\_i

### 5. สรุป

จากกระบวนการพัฒนาและศึกษาทฤษฎีที่เกี่ยวข้องและคำปรึกษาจากผู้เชี่ยวชาญ กลุ่มเป้าหมายประชากรตัวอย่าง 550 คน ได้อंकประกอบที่คาดว่าเกี่ยวข้องกับปัจจัยที่ส่งผลให้ให้ผู้บริโภคซื้อสินค้าบนตลาดกลางพาณิชย์อิเล็กทรอนิกส์ดังรูปที่ 6

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1 R Square = 0.109	Constant	2.223	0.342		6.495	0.000
	avg_e	0.084	0.047	0.077	1.789	0.075
	avg_t	0.255	0.040	0.263	6.418	0.000
	avg_en	0.111	0.043	0.107	2.558	0.011
	avg_u	0.053	0.050	0.046	1.057	0.291
2 R Square = 0.039	Constant	3.750	0.190		19.737	0.000
	avg_a	0.189	0.040	0.198	4.731	0.000

รูปที่ 6 ตารางสรุปผล

อธิบายได้ว่าปัจจัยความเชื่อมั่น (Trust) ความมีส่วนร่วม (Engaging) และ ความง่ายในการใช้เทคโนโลยี

(Perceive ease of use) ต่อตลาดกลางพาณิชย์อิเล็กทรอนิกส์มีอิทธิพลกับทัศนคติหรือพฤติกรรมผู้บริโภคที่มีต่อการใช้งานตลาดกลางพาณิชย์อิเล็กทรอนิกส์ โดยมีค่า R square = 0.109 และ ทัศนคติหรือพฤติกรรมผู้บริโภคที่มีต่อการใช้งานตลาดกลางพาณิชย์อิเล็กทรอนิกส์มีอิทธิพลกับความตั้งใจซื้อสินค้า (Purchase Intention) โดยมีค่า R square = 0.039

### เอกสารอ้างอิง

[1] สมาคมการตลาดแห่งประเทศไทย. (2563). *ไพร์ซซ่า (Priceza) สรุปภาพรวมตลาดอีคอมเมิร์ซไทยปี 2019 พร้อมชี้ทิศทางปี 2020 ในช่วงสถานการณ์โควิด-19*. [สืบค้นวันที่ 12 สิงหาคม 2564]. <https://www.marketingthai.or.th/priceza-ecommerce-2020/>

[2] Ramzan Sama และ Jay Trivedi. (2562). *The Effect of Influencer Marketing on Consumers' Brand Admiration and Online Purchase Intentions: An Emerging Market Perspective*. [สืบค้นวันที่ 12 สิงหาคม 2564]. <https://www.tandfonline.com/action/showCitFormats?doi=10.1080%2F15332861.2019.1700741>

[3] Naseri และ Roszi Naszariah Nasni. (2564). *An Overview of Online Purchase Intention of Halal Cosmetic Product: A Perspective From Malaysia*. [สืบค้นวันที่ 12 สิงหาคม 2564]. <https://www.turcomat.org/index.php/turkbilmat/article/view/5679>

[4] Amir Reza Konjkav Monfared, Ghaffari Mohammad, Barootkoob Mohammadreza และ Malmiri Milad Mohebali. (2563). *The role of social commerce in online purchase intention: mediating role of social interactions, trust, and electronic word of mouth*. [สืบค้นวันที่ 12 สิงหาคม 2564]. <https://www.inderscienceonline.com/doi/abs/10.1504/JIBED.2021.112264>

# ปัจจัยด้านความเครียดที่มีอิทธิพลต่อประสิทธิภาพการสอนในระดับการศึกษา ขั้นพื้นฐาน ในช่วงสถานการณ์โควิด-19

## Stress Factors and Their Impact on Teaching Efficiency at Fundamental Education Level during the COVID-19 Pandemic

วิโรจน์ เชิดสวัสดิ์ (Wirote Cherdsawat)<sup>1</sup> และปราโมทย์ ลื่อนาม (Pramote Luenam)<sup>2</sup>

สาขาบริหารเทคโนโลยีสารสนเทศ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

<sup>1</sup>wirote.che@nida.ac.th, <sup>2</sup>pramote@nida.ac.th

### บทคัดย่อ

ปัจจุบันการดำเนินชีวิตของผู้คนมีการเปลี่ยนแปลงไปอย่างรวดเร็วและก้าวเข้าสู่ยุคสังคมดิจิทัล เทคโนโลยีที่มีความก้าวหน้าไม่หยุดยั้งเข้ามามีบทบาทในชีวิตประจำวันเป็นอย่างมาก โดยเฉพาะอย่างยิ่งในสถานการณ์การแพร่ระบาดของเชื้อไวรัสโควิด-19 ที่ส่งผลกระทบทำให้รูปแบบการเรียนการสอนของทั้งผู้สอนและผู้เรียนเปลี่ยนแปลงไปอย่างมาก งานวิจัยชิ้นนี้มีวัตถุประสงค์เพื่อตรวจสอบปัจจัยด้านความเครียดที่มีอิทธิพลต่อประสิทธิภาพการสอนในระดับการศึกษาขั้นพื้นฐานในช่วงสถานการณ์โควิด-19 ในการศึกษาครั้งนี้ได้รวบรวมแบบสอบถามจากกลุ่มตัวอย่างที่เป็นครูจำนวน 400 คนที่สอนอยู่ในช่วงระดับการศึกษาขั้นพื้นฐานและมีประสบการณ์สอนออนไลน์ในช่วงการแพร่ระบาด เมื่อนำข้อมูลมาวิเคราะห์ด้วยโมเดลสมการโครงสร้าง เพื่อทดสอบสมมติฐาน และหาความสัมพันธ์กับข้อมูลเชิงประจักษ์ โมเดลแสดงความเหมาะสมของด้วยค่า  $P\text{-Value} = 0.020$ ,  $\chi^2/df = 1.208$ ,  $RMSEA = 0.023$ ,  $CFI = 0.989$  และ  $TLI = 0.986$  ผลการศึกษาพบว่าความเครียดของบุคลากรครูมีผลต่อประสิทธิภาพการสอนและความพร้อมด้านเทคโนโลยีมีส่วนช่วยเพิ่มประสิทธิภาพการสอนด้วยการวิจัยนี้ค้นพบว่าความเครียดจากเทคโนโลยีทั้ง 3 ด้านมี 1 ด้านที่ส่งผลกระทบต่อความเครียดของครู คือ การถูกบังคับใช้เทคโนโลยี อีก 2 ด้านที่ไม่ส่งผลคือ ความซับซ้อนของเทคโนโลยีและความไม่แน่นอนของเทคโนโลยี

**คำสำคัญ:** การเรียนการสอนแบบออนไลน์ โควิด-19  
ความเครียด ครูระดับการศึกษาขั้นพื้นฐาน

ความเครียดทางเทคโนโลยี ความพร้อมด้านเทคโนโลยี

### Abstract

Nowadays, people's lifestyle is changing rapidly and is becoming more of a digital society. In such a society, the ongoing technologies play an important role in everyday life especially in the epidemic situation of the COVID-19 virus. The pandemic causes a big change the way in which teachers teach and students undertake study. The purpose of this study is to investigate the stress factors and their impact on teaching efficiency at fundamental education level during the COVID-19 pandemic. The study used online questionnaire survey to collect a sample of 400 teachers, who have experience in online teaching during the pandemic. The data were analyzed through Structural Equation Model to test the hypotheses and derive model fit. The model showed a goodness-of-fit with  $P\text{-Value}=0.020$ ,  $\chi^2/df = 1.208$ ,  $RMSEA = 0.023$ ,  $CFI = 0.989$  and  $TLI = 0.986$ . Results indicate that teachers' stress has a significant direct effect on teaching efficiency and technological readiness helps increasing the efficiency of teaching. Results also show that techno-overload has a positive direct effect on the teachers' stress. On the other hand, the techno-complexity and techno-insecurity have no effect on the teachers' stress.

**Keywords:** Online Teaching, COVID-19, Stress, Basic education teacher, Technostress, Technology Readiness.

## 1. บทนำ

ประเทศไทยเป็นหนึ่งในประเทศที่ได้รับผลกระทบจากการระบาดของเชื้อไวรัสโคโรนา-19 ในวันที่ 13 มกราคม 2563 พบผู้ป่วยคนแรกในประเทศไทยและนับจากนั้นเป็นต้นมาการระบาดในประเทศไทยก็ยิ่งรุนแรงขึ้น ส่งผลทำให้ สถานที่ท่องเที่ยว สวนสาธารณะ กิจกรรมต่างๆ ห้างสรรพสินค้า ต้องปิดตัวลง และสถานศึกษาต้องเลื่อนการเปิดภาคเรียนออกไป กระทรวงศึกษาธิการได้ประกาศเลื่อนการเปิดภาคเรียนของปีการศึกษา 2564 และกำหนดแนวปฏิบัติการเปิดภาคเรียนการสอนในรูปแบบการจัดการศึกษาทางไกล 4 รูปแบบ คือ (1) On-Air การเรียนผ่านมูลนิธิการศึกษาทางไกลผ่านดาวเทียม ในพระบรมราชูปถัมภ์ (2) Online การเรียนผ่านอินเทอร์เน็ต (3) On-Demand การเรียนผ่านแอปพลิเคชัน และ (4) On-hand การเรียนที่บ้านด้วยเอกสารผ่านการจัดส่งทางไปรษณีย์ [1]

ปัจจุบันแนวโน้มผู้ติดเชื้อไวรัสโคโรนา-19 เพิ่มสูงขึ้นทั่วโลก กระทรวงศึกษาธิการจึงไม่สามารถจัดการเรียนการสอนในรูปแบบปกติได้ ยังคงจัดการเรียนการสอนในรูปแบบการศึกษาทางไกลต่อไป ทำให้เทคโนโลยีเข้ามามีบทบาทในด้านการศึกษามากขึ้น ภาระหน้าที่ของครูผู้สอนก็มากขึ้นตามไปด้วย การเปลี่ยนแปลงรูปแบบการเรียนการสอนแบบกระทันหัน ทำให้การเตรียมการเรียนการสอนต้องเปลี่ยนแปลงใหม่ทั้งหมด ไม่ว่าจะเป็นเรื่องของสื่อการเรียนการสอน อุปกรณ์ การตั้ง-ส่งงาน การสอบ และแอปพลิเคชันที่นำมาใช้งาน ต้องมีการเรียนรู้ใหม่เพื่อนำมาปรับใช้ให้เข้ากับรูปแบบที่เปลี่ยนไป สร้างภาระและแรงกดดันกับครูผู้สอนเป็นอย่างมาก ยิ่งไปกว่านั้นยังต้องกังวลต่อการติดเชื้อไวรัส ทำให้เกิดปัญหาหลายด้านตามมา หนึ่งในนั้นคือความเครียด เมื่อครูผู้สอนเกิดความเครียดมากขึ้น จะส่งผลต่อประสิทธิภาพการสอน ทำให้เกิดผลกระทบเป็นลูกโซ่ต่อไป

งานวิจัยชิ้นนี้จัดทำขึ้นเพื่อศึกษาและวิเคราะห์ ปัจจัยด้านความเครียดที่มีอิทธิพลต่อประสิทธิภาพการสอนในระดับการศึกษาขั้นพื้นฐาน ในช่วงสถานการณ์โควิด-19

เพื่อเป็นประโยชน์ต่อแนวทางการวางแผนการเรียนการสอน และเป็นแนวทางลดความตึงเครียดของครูผู้สอน

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ความเครียด (Stress) จากการทำงานของครูสะท้อนถึงประสบการณ์อารมณ์อันไม่พึงประสงค์อันเป็นผลจากการสอน [2] ยิ่งในสถานการณ์ปัจจุบันที่มีข้อจำกัดในการเรียนการสอนรูปแบบออนไลน์ สิ่งนี้ไม่เพียงแต่มีความเกี่ยวข้องกับครูเท่านั้น ยังรวมไปถึงผู้บริหารโรงเรียนและผู้กำหนดนโยบายด้วย เนื่องจากอาชีพการสอนได้รับการระบุว่ามีความเครียดสูง โดยนักวิจัยหลายคน [3], [4], [5] ผลการศึกษาหลายชิ้นแสดงให้เห็นว่าหนึ่งในสามของประชากรครูมีความเครียดถึงเครียดมาก [6] สาเหตุของความเครียดของครู ได้แก่ (1) ความเครียดที่เกี่ยวข้องกับพฤติกรรมและวินัยของนักเรียน และ (2) ความเครียดที่เกี่ยวข้องกับภาระงาน [7] งานวิจัยแสดงให้เห็นว่าความเครียดทั้งสองประเภทนี้เกี่ยวข้องกับผลลัพธ์เชิงลบหลายประการสำหรับครู รวมถึงการหมดไฟในการสอน [8] และความมุ่งมั่นในการทำงาน

ประสิทธิภาพการสอน (Teaching Efficiency) ประกอบด้วย (1) ประสิทธิภาพในการมีส่วนร่วมของนักเรียน หมายถึง ความมั่นใจในความสามารถในการสร้างแรงจูงใจ ความเข้าใจ และคุณค่าของการเรียนรู้ ให้แก่นักเรียน (2) ประสิทธิภาพในการจัดการห้องเรียน หมายถึง ความมั่นใจในการควบคุมพฤติกรรมนักเรียนให้ปฏิบัติตามกฎของห้องเรียน (3) ประสิทธิภาพสำหรับกลยุทธ์การสอน หมายถึง ความมั่นใจในการใช้กลยุทธ์การสอนอย่างมีประสิทธิภาพ [9] นอกจากนี้ [7] พบว่าความเครียดของครูมีส่วนสำคัญต่อประสิทธิภาพการสอนอย่างมาก

ความเครียดทางเทคโนโลยี (Technostress) เป็นภาวะที่เกิดจากการไม่สามารถปรับตัวหรือรับมือกับเทคโนโลยีสมัยใหม่ในลักษณะที่เหมาะสมได้ [10] ได้อธิบายความหมายของความเครียดทางเทคโนโลยีว่า “เป็นความล้มเหลวของแต่ละบุคคลในการพยายามหาทางสู้และวิธีการจัดการกับเทคโนโลยีสารสนเทศที่มีการพัฒนาอย่างต่อเนื่อง ซึ่งจำเป็นที่จะต้องมีความรู้ความสามารถที่เพียงพอทันต่อเทคโนโลยีต่างๆ ตามการเปลี่ยนแปลงของ

บทบาทและความสำคัญของเทคโนโลยีที่มีต่อสังคม” โดยแบ่งความเครียดจากเทคโนโลยีออกเป็น 5 มิติ ในที่นี้ผู้วิจัยได้นำมาศึกษา 3 มิติที่มีอิทธิพลและเกี่ยวข้องกับกระบวนการสอนแบบออนไลน์ ได้แก่ (1) การถูกบังคับใช้เทคโนโลยี (techno-overload) หมายถึง สถานการณ์ที่เทคโนโลยีบังคับให้ผู้ใช้ทำงานได้เร็วขึ้นและต้องทำงานได้มากขึ้น ในช่วงการเรียนการสอนแบบออนไลน์ (2) ความซับซ้อนของเทคโนโลยี (techno-complexity) หมายถึง ความซับซ้อนของเทคโนโลยี ทำให้ผู้ใช้รู้สึกว่าทักษะที่มีไม่เพียงพอต่อการใช้งานและถูกบังคับให้ต้องเรียนรู้และความเข้าใจในเทคโนโลยีใหม่ๆ มากขึ้น ในช่วงการเรียนการสอนแบบออนไลน์ (3) ความไม่แน่นอนของเทคโนโลยี (techno-uncertainty) หมายถึง การเปลี่ยนแปลงอย่างต่อเนื่องและการพัฒนาของเทคโนโลยีสร้างความไม่แน่นอนสำหรับผู้ใช้งาน ทำให้ต้องศึกษาเรียนรู้เทคโนโลยีใหม่ๆ ในช่วงการเรียนการสอนแบบออนไลน์ [10] ได้ศึกษาวิจัยในเรื่องของความเครียดทางเทคโนโลยีจากความไม่สอดคล้องกับสภาพแวดล้อมของบุคคลในหลายมิติที่ส่งผลต่อเทคโนโลยีทางการศึกษาพบว่า ระหว่างบุคคลกับสิ่งแวดล้อม บุคคลกับเทคโนโลยีเพื่อการศึกษา และบุคคลกับบุคคล ส่งผลในเชิงบวกกับความเมือหน่ายในการเรียนรู้ของนักเรียน

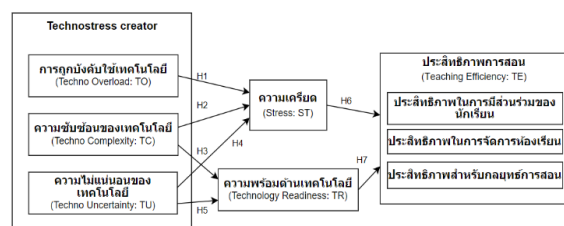
ความพร้อมด้านเทคโนโลยี (Technology Readiness) เป็นสภาพจิตใจโดยรวมของแต่ละบุคคลเกี่ยวกับเทคโนโลยี โดยทั่วไป [12] ให้คำจำกัดความของความพร้อมด้านเทคโนโลยีว่า “แนวโน้มของผู้คนในการยอมรับและใช้เทคโนโลยีใหม่ๆ เพื่อบรรลุเป้าหมายในชีวิตที่บ้านและที่ทำงาน” ความพร้อมด้านเทคโนโลยีประกอบด้วยสี่มิติ คือ (1) การมองโลกในแง่ดี หมายถึง มุมมองเชิงบวกของเทคโนโลยีและความเชื่อที่ว่าสิ่งนี้ทำให้ผู้คนมีการควบคุมความยืดหยุ่น และประสิทธิภาพในชีวิตเพิ่มขึ้น (2) นวัตกรรม หมายถึง แนวโน้มที่จะเป็นผู้บุกเบิกเทคโนโลยีและผู้นำทางความคิด (3) ความรู้สึกไม่สบาย หมายถึง การรับรู้ว่าขาดการควบคุมทางด้านเทคโนโลยีและความรู้สึกถูกครอบงำโดยเทคโนโลยี (4) ความไม่มั่นคง หมายถึง ความไม่ไว้วางใจในเทคโนโลยีและความกังวลเกี่ยวกับความสามารถ

ในการทำงานอย่างถูกต้อง มิติของการมองโลกในแง่ดีและนวัตกรรมถือเป็นตัวขับเคลื่อนของความพร้อมด้านเทคโนโลยี ในขณะที่ความรู้สึกไม่สบายและความไม่มั่นคงถือเป็นตัวยับยั้ง ในกลุ่มคนที่ทำงาน ผู้ใช้ที่แสดงทัศนคติที่มองโลกในแง่ดีมากขึ้นต่อความพร้อมด้านเทคโนโลยีก็มีความตั้งใจที่จะใช้งานไอทีมาก

การระบาดใหญ่ทั่วโลกส่งผลกระทบต่อเด็กและเยาวชน โรงเรียนและมหาวิทยาลัยปิดในหลายพื้นที่ของโลก ระบบการศึกษาได้หยุดการเรียนการสอนในห้องเรียนลงและเปลี่ยนรูปแบบเป็นการสอนออนไลน์แทน และการเรียนการสอนแบบออนไลน์ก่อให้เกิดปัญหาขึ้นมากมาย เช่น (1) ความพร้อมด้านเทคโนโลยี เนื่องจากไม่ใช่ นักเรียนและครูทุกคนที่มีอุปกรณ์ที่จำเป็นสำหรับการเรียนรู้ออนไลน์ ปัญหานี้นำไปสู่ความวิตกกังวล ความเครียด หงุดหงิด และโดดเดี่ยว (2) ความรู้ในด้านเทคโนโลยี ที่ครูต้องการการฝึกอบรมเพื่อทำความเข้าใจวิธีการใช้เทคโนโลยีเหล่านี้ (3) สิ่งอำนวยความสะดวกการเชื่อมต่ออินเทอร์เน็ต (4) สภาพแวดล้อมที่เหมาะสมสำหรับการเรียนรู้ออนไลน์ [13] ผลการวิจัยของ [14] กล่าวว่า “บุคคลบางคนแสดงความวิตกกังวลและกลัวเทคโนโลยีต่อเทคโนโลยีในอุตสาหกรรมของตนเอง พวกเขา รู้สึกอึดอัดและท้อแท้ในการใช้เทคโนโลยี โดยเฉพาะบุคคลที่ไม่มีคุณลักษณะ เช่นความพร้อมด้านเทคโนโลยี จะเผชิญกับความรู้สึกไม่สบายใจ ความวิตกกังวลและความหวาดกลัวเทคโนโลยีเมื่อต้องเผชิญกับเทคโนโลยีใหม่และมีความซับซ้อน”

### 3. กรอบแนวความคิดการวิจัย

จากทฤษฎีและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้กำหนดกรอบแนวคิดงานวิจัย ดังแสดงในภาพที่ 1



ภาพที่ 1: กรอบแนวคิดงานวิจัย

**สมมติฐานที่ 1 (H1):** การถูกบังคับใช้เทคโนโลยี (TO) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความเครียด (ST)

**สมมติฐานที่ 2 (H2):** ความซับซ้อนของเทคโนโลยี (TC) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความเครียด

**สมมติฐานที่ 3 (H3):** ความซับซ้อนของเทคโนโลยี (TC) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความพร้อมด้านเทคโนโลยี (TR)

**สมมติฐานที่ 4 (H4):** ความไม่แน่นอนของเทคโนโลยี (TU) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความเครียด

**สมมติฐานที่ 5 (H5):** ความไม่แน่นอนของเทคโนโลยี (TU) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความพร้อมด้านเทคโนโลยี (TR)

**สมมติฐานที่ 6 (H6):** ความเครียด (ST) มีอิทธิพลต่อประสิทธิภาพการสอน (TE)

**สมมติฐานที่ 7 (H7):** ความพร้อมด้านเทคโนโลยี (TR) มีอิทธิพลต่อประสิทธิภาพการสอน (TE)

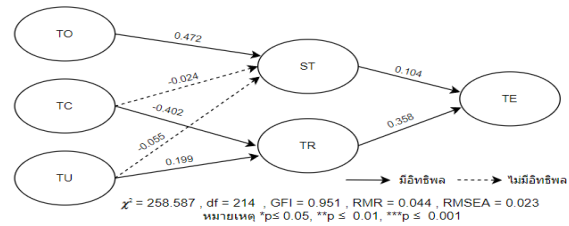
งานวิจัยนี้เป็นการวิจัยเชิงปริมาณ (Quantitative Research) โดยใช้แบบสอบถามเป็นเครื่องมือเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างในช่วงพฤศจิกายน-มกราคม 2564 ประชากรคือนุศลากรครูที่สอนอยู่ในช่วงระดับการศึกษาระดับชั้นพื้นฐานที่มีประสบการณ์สอนออนไลน์ในช่วงของการแพร่ระบาดของโควิด-19 ผู้วิจัยใช้โปรแกรม G\*Power ในการคำนวณขนาดของกลุ่มตัวอย่าง โดยกำหนดค่าเพาเวอร์ (1-β) เท่ากับ 0.95 ค่าอัลฟา (α) เท่ากับ 0.05 จำนวนตัวแปรทำนายเท่ากับ 8 ตัวแปร ขนาดของอิทธิพล (Effect Size) เท่ากับ 0.033 ผลที่ได้คือขนาดของกลุ่มตัวอย่างเท่ากับ 396 ตัวอย่าง ซึ่งผู้วิจัยได้ปรับจำนวนขนาดของกลุ่มตัวอย่างเพิ่มเติมเป็น 400 ตัวอย่าง

ผู้วิจัยได้ทำการทดสอบสมมติฐานของงานวิจัยโดยใช้โปรแกรม SPSS และ AMOS ในการวิเคราะห์รูปแบบเชิงโครงสร้าง (Structural Equation Modeling: SEM) เพื่อวิเคราะห์หาค่าอิทธิพลทั้งทางตรงและทางอ้อม และได้มีการปรับรูปแบบของโมเดล เพื่อสอดคล้องกับข้อมูลเชิงประจักษ์ (Model Fit)

**4. ผลการดำเนินการวิจัย**

ผู้ตอบแบบสอบถามในภาพรวมพบว่าเป็นผู้หญิงร้อยละ 78.5 และชาย ร้อยละ 21.5 ส่วนใหญ่ระดับ

การศึกษาอยู่ในระดับปริญญาตรี ร้อยละ 69 และมีช่วงอายุ 25 - 30 ปี ร้อยละ 37 เมื่อนำข้อมูลมาวิเคราะห์ด้วยโมเดลสมการโครงสร้างเพื่อทดสอบสมมติฐานและหาความสอดคล้องกับข้อมูลเชิงประจักษ์ โมเดลแสดงความเหมาะสมของด้วยค่า P-value = 0.020,  $\chi^2/df = 1.208$ , RMSEA = 0.023, CFI = 0.989 และ TLI = 0.986 ดังแสดงในภาพที่ 2



ภาพที่ 2: สัมประสิทธิ์เส้นทางมาตรฐานของโมเดล

ตารางที่ 1: ผลการวิเคราะห์อิทธิพลเชิงสาเหตุภายในโมเดล

ตัวแปรตาม	อิทธิพล	ตัวแปรต้น				
		TO	TC	TU	ST	TR
ST	ทางตรง	0.472***	-0.024	-0.055	-	-
	ทางอ้อม	-	-	-	-	-
TR	ทางตรง	-	-0.402***	0.199***	-	-
	ทางอ้อม	-	-	-	-	-
TE	ทางตรง	-	-	-	0.104*	0.358***
	ทางอ้อม	0.049**	-0.143**	0.071*	-	-

หมายเหตุ \*p≤0.05, \*\*p≤ 0.01, \*\*\*p≤ 0.001

จากผลการวิเคราะห์อิทธิพลเชิงสาเหตุภายในโมเดลในตารางที่ 1 สามารถนำมาหาผลสรุปตามข้อสมมติฐานในงานวิจัยได้ดังแสดงในตารางที่ 2

ตารางที่ 2: สรุปผลการทดสอบสมมติฐานในการวิจัย

สมมติฐานงานวิจัย	ค่าอิทธิพล	ค่า P	ผลการทดสอบ
H1: การถูกบังคับใช้เทคโนโลยี (TO) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความเครียด (ST)	0.049	≤ 0.01	ยอมรับ
H2: ความซับซ้อนของเทคโนโลยี (TC) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความเครียด (ST)	-0.002	0.620	ปฏิเสธ
H3: ความซับซ้อนของเทคโนโลยี (TC) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความพร้อมด้านเทคโนโลยี (TR)	-0.143	≤ 0.01	ยอมรับ

สมมติฐานงานวิจัย	ค่า อิทธิพล	ค่า P	ผลการ ทดสอบ
H4: ความไม่แน่นอนของเทคโนโลยี (TU) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความเครียด (ST)	-0.006	0.251	ปฏิเสธ
H5: ความไม่แน่นอนของเทคโนโลยี (TU) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความพร้อมด้านเทคโนโลยี (TR)	0.071	$\leq 0.05$	ยอมรับ
H6: ความเครียด (ST) มีอิทธิพลต่อประสิทธิภาพการสอน (TE)	0.104	$\leq 0.05$	ยอมรับ
H7: ความพร้อมด้านเทคโนโลยี (TR) มีอิทธิพลต่อประสิทธิภาพการสอน (TE)	0.358	$< 0.0001$	ยอมรับ

## 5. สรุปและการอภิปราย

จากการศึกษาวิจัยเรื่อง ปัจจัยด้านความเครียดที่มีอิทธิพลต่อประสิทธิภาพการสอน ในระดับการศึกษาขั้นพื้นฐานในช่วงสถานการณ์โควิด-19 ประกอบไปด้วยสมมติฐานการวิจัย 7 ข้อ ดังแสดงในตารางที่ 2

การถูกบังคับใช้เทคโนโลยี (TO) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความเครียด (ST) พบว่าข้อมูลเชิงประจักษ์สนับสนุนสมมติฐานการวิจัยอย่างมีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับงานวิจัยของ [10] ทั้งนี้เนื่องจากปัจจุบันครูในประเทศไทยมีภาระงานเอกสารมากกว่าภาระการสอนและเมื่อรูปแบบการเรียนการสอนเปลี่ยนเป็นแบบออนไลน์ จึงเป็นการเพิ่มภาระงานของครูเป็นอย่างมาก โดยปกติแล้วครูช่วงระดับการศึกษาขั้นพื้นฐานมีภาระการสอนมากกว่า 1 รายวิชาอยู่แล้ว บางโรงเรียนที่มีขนาดเล็กอาจมากถึง 8-9 รายวิชา สำหรับการสอนใน 1 ชั่วโมงนั้นครูต้องเตรียมการสอน 1-2 ชั่วโมง แต่ภาระการสอนก็ใช้ระยะเวลาพอสมควรแล้วในสถานการณ์โควิด-19 ครูต้องใช้เวลาในการศึกษาเทคโนโลยีในการสอนเพิ่มขึ้นทั้งในเรื่องของการสอน การส่งงาน สอบ ล้วนแล้วแต่เป็นเรื่องที่ต้องศึกษาเพิ่มเติม อีกทั้งยังต้องมีการประชุมกับทางผู้บริหาร ทำให้ภาระงานของครูเพิ่มขึ้น โดยข้อเสนอแนะงานวิจัยที่ผู้วิจัยได้รับมา กล่าวถึงเรื่องภาระงานที่สิ้นมือของครูเป็นส่วนใหญ่ ผู้วิจัยมีความเห็นว่า การลดภาระงานเอกสารจะช่วยลดความเครียดในการทำงานของครูลงด้วย และช่วยเพิ่มประสิทธิภาพการสอนให้สูงขึ้น

ความซับซ้อนของเทคโนโลยี (TC) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความเครียด (ST) พบว่าข้อมูลเชิงประจักษ์ไม่สนับสนุนสมมติฐานการวิจัย แสดงให้เห็นว่าครูมีทักษะด้านเทคโนโลยีพอสมควรจึงไม่ก่อให้เกิดความเครียดจากการใช้เทคโนโลยี ซึ่งขัดแย้งกับผลการวิจัยของ [10] ที่กล่าวว่า ความซับซ้อนของเทคโนโลยี ทำให้ผู้ใช้รู้สึกว่ามีไม่เพียงพอต่อการใช้งานและถูกบังคับให้ต้องเรียนรู้และเข้าใจเทคโนโลยีใหม่ๆ มากขึ้น

ความซับซ้อนของเทคโนโลยี (TC) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความพร้อมด้านเทคโนโลยี (TR) พบว่าข้อมูลเชิงประจักษ์สนับสนุนสมมติฐานการวิจัยอย่างมีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับงานวิจัยของ [14] ที่กล่าวว่า บุคคลบางคนแสดงความวิตกกังวลและกลัวเทคโนโลยีต่อเทคโนโลยีในอุตสาหกรรมของตนเอง พวกเขาเรียนรู้ที่อดทนและท้าทายในการใช้เทคโนโลยี โดยเฉพาะบุคคลที่ไม่มีคุณลักษณะ เช่นความพร้อมด้านเทคโนโลยีจะเผชิญกับความวิตกกังวลไม่สบายใจ ความวิตกกังวลและความหวาดกลัวเทคโนโลยีเมื่อต้องเผชิญกับเทคโนโลยีใหม่และมีความซับซ้อน ผู้วิจัยมีความเห็นว่า ควรมีการจัดอบรมและให้ความรู้เชิงปฏิบัติด้านการใช้งานเทคโนโลยีและการใช้งานสื่อการสอนแบบออนไลน์ เพื่อให้ครูสามารถนำไปปรับใช้และปฏิบัติได้จริง

ความไม่แน่นอนของเทคโนโลยี (TU) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความเครียด (ST) พบว่าข้อมูลเชิงประจักษ์ไม่สนับสนุนสมมติฐานการวิจัย แสดงให้เห็นว่าการเปลี่ยนแปลงอย่างรวดเร็วของเทคโนโลยีไม่ก่อให้เกิดความเครียดแก่ครู ซึ่งขัดแย้งกับผลการวิจัยของ [10] ที่กล่าวว่า การเปลี่ยนแปลงอย่างต่อเนื่องและการพัฒนาของเทคโนโลยี สร้างความไม่แน่นอนสำหรับผู้ใช้งาน ทำให้ต้องศึกษาเรียนรู้เทคโนโลยีใหม่ๆ

ความไม่แน่นอนของเทคโนโลยี (TU) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) ผ่านความพร้อมด้านเทคโนโลยี (TR) พบว่าข้อมูลเชิงประจักษ์สนับสนุนสมมติฐานการวิจัยอย่างมีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับงานวิจัยของ [13] กล่าวคือ การพัฒนาอย่างรวดเร็วของ

เทคโนโลยี ก่อให้เกิดปัญหาในเรื่องของสิ่งอำนวยความสะดวก การเชื่อมต่ออินเทอร์เน็ต และการสอนออนไลน์ ครูบางโรงเรียนไม่ได้รับการสนับสนุนในเรื่องของอุปกรณ์ในการสอนออนไลน์ ทำให้ต้องใช้ทรัพยากรส่วนตัวในจัดหา ทั้งในเรื่องของอินเทอร์เน็ตก็ต้องใช้ทรัพยากรส่วนตัวในการสอนออนไลน์ ผู้วิจัยมีความเห็นว่า คณะผู้บริหารโรงเรียนควรมองเห็นความสำคัญของเทคโนโลยีมากขึ้นและจัดสรรงบประมาณไว้รองรับการเปลี่ยนแปลงของเทคโนโลยี

ความเครียด (ST) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) พบว่าข้อมูลเชิงประจักษ์สนับสนุนสมมติฐานการวิจัยอย่างมีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับงานวิจัยของ [7] ที่กล่าวว่า ความเครียดของครูมีส่วนสำคัญต่อประสิทธิภาพการสอนของพวกเขาอย่างมาก ผลการวิจัยแสดงให้เห็นว่า เมื่อครูเกิดความเครียดจะส่งผลทำให้ประสิทธิภาพการสอนของพวกเขาตกลงไปด้วย

ความพร้อมด้านเทคโนโลยี (TR) มีอิทธิพลต่อประสิทธิภาพการสอน (TE) พบว่าข้อมูลเชิงประจักษ์สนับสนุนสมมติฐานการวิจัยอย่างมีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับงานวิจัยของ [12] ที่กล่าวว่า แนวโน้มของผู้คนในการยอมรับและใช้เทคโนโลยีใหม่ ๆ เพื่อบรรลุเป้าหมายในชีวิตที่บ้านและที่ทำงาน ผลการวิจัยแสดงให้เห็นว่าความพร้อมทางด้านเทคโนโลยีไม่ว่าจะเป็นการใช้งานเทคโนโลยีหรือความรู้ด้านเทคโนโลยีหากครูมีความพร้อมจะช่วยส่งผลทำให้ประสิทธิภาพการสอนรูปแบบออนไลน์ในช่วงสถานการณ์โควิด-19 ดียิ่งขึ้น

ข้อเสนอแนะในการวิจัยครั้งต่อไป กรณีสถานการณ์โควิด-19 ยังไม่คลี่คลายและการเรียนการสอนยังคงเป็นรูปแบบออนไลน์ ผู้วิจัยเสนอให้เพิ่มตัวแปรในเรื่องของกลุ่มสาระการเรียนรู้เข้ามาเกี่ยวข้องด้วย เนื่องจากแต่ละรายวิชาอาจมีความเครียดแตกต่างกันไป หรือเพิ่มตัวแปรในเรื่องปัจจัยภายนอกที่อาจส่งผลต่อความเครียดได้ เช่น สภาพแวดล้อมในการสอน นอกจากนี้การเพิ่มการวิเคราะห์ความสามารถในการใช้เทคโนโลยีของกลุ่มตัวอย่างก็เป็นอีกสิ่งที่น่าสนใจ เนื่องจากกลุ่มตัวอย่างอาจมีความสามารถในการใช้เทคโนโลยีที่แตกต่างกัน

## เอกสารอ้างอิง

- [1] ศษ.360 องค์ฯ (2564). ประกาศ ศษ. เรื่อง การเลื่อนเวลาเปิดภาคเรียนที่ 1 ประจำปีการศึกษา 2564 สืบค้นวันที่ 13 กรกฎาคม 2564 จาก <https://moe360.blog/2021/05/19/19-may-2564/>
- [2] C. Kyriacou. "Teacher stress: Directions for future research." *Educational review*, Vol. 53, No. 1, pp. 27-35, 2001.
- [3] C. Kyriacou. *Stress-busting for teachers*. Nelson Thornes, 2000.
- [4] D. Nobile, J. John, and J. McCormick. "Job satisfaction and occupational stress in Catholic primary schools: Implications for school leadership." *Leading and managing*, Vol. 13, No. 1, pp. 31-48, 2007.
- [5] A. Fudail, Mohammed, and H. Mellar. "Investigating teacher stress when using technology." *Computers & Education*, Vol. 51, No. 3, pp. 1103-1110, 2008.
- [6] Geving, M. Allison. "Identifying the types of student and teacher behaviours associated with teacher stress." *Teaching and Teacher Education*, Vol. 23, No.5, pp. 624-640, 2007.
- [7] Klassen, M. Robert, and M. M. Chiu. "Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress." *Journal of educational Psychology*, Vol. 102, No. 3, pp. 741, 2010.
- [8] McCarthy, C. J., Lambert, R. G., O'Donnell, M., & Melendres, L. T. "The relation of elementary teachers' experience, stress, and coping resources to burnout symptoms". *The Elementary School Journal*, Vol. 109, No. 3, pp. 282-300, 2009.
- [9] Tschannen-Moran, Megan, and Anita Woolfolk Hoy. "Teacher efficacy: Capturing an elusive construct." *Teaching and teacher education*, Vol. 17, No. 7, pp.783-805, 2001.
- [10] Tarafdar, M., Tu, Q., Ragu-Nathan, B. S., & Ragu-Nathan, T. S. "The impact of technostress on role stress and productivity". *Journal of management information systems*, Vol. 24, No. 1, pp. 301-328, 2007.
- [11] Wang, Xinghua, Seng Chee Tan, and Lu Li. "Technostress in university students' technology-enhanced learning: An investigation from multidimensional person-environment misfit." *Computers in Human Behavior*, Vol. 105, 106208, 2022.
- [12] Parasuraman, Ananthanarayanan. "Technology Readiness Index (TRI) a multiple-item scale to measure readiness to embrace new technologies." *Journal of service research* 2.4, 307-320, 2000.
- [13] Ahmadon, Fadzlin, Hajar Izzati Mohd Ghazalli, and Hazlifah Mohd Rusli. "Studying during Pandemic: A Review of Issues from Online Learning in the Middle of COVID-19." 2020 6th International Conference on Interactive Digital Media (ICIDM). IEEE, 2020.
- [14] Meuter, M. L., Ostrom, A. L., Bitner, M. J., & Roundtree, R. The influence of technology anxiety on consumer use and experiences with self-service technologies. *Journal of Business Research*, 56(11), 899-906, 2003

**การวิเคราะห์องค์ประกอบที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศ  
e-Government (G2E) ของข้าราชการ สังกัดกระทรวงยุติธรรม  
An Factor Analysis of components influencing the acceptance of  
e-Government (G2E) information technology among civil servants under  
the Ministry of Justice**

กฤตมุข เล็กเล็ง (Krittamuk Lekseng)<sup>1</sup> และมณเฑียร รัตนศิริวงศ์วุฒิ<sup>2</sup> (Montean Rattanasiriwongwut)

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ<sup>1,2</sup>

S6307021856128@email.kmutnb.ac.th<sup>1</sup>, montean.r@itd.kmutnb.ac.th<sup>2</sup>

### บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อวิเคราะห์องค์ประกอบที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศ e-Government (G2E) ของข้าราชการ สังกัดกระทรวงยุติธรรม 2) เพื่อหาความสอดคล้องขององค์ประกอบที่ส่งผลต่อการยอมรับเทคโนโลยีสารสนเทศ e-Government (G2E) ของข้าราชการ สังกัดกระทรวงยุติธรรม ประชากรที่ทำการวิจัยในครั้งนี้ คือ ข้าราชการ สังกัดกระทรวงยุติธรรม จำนวน 1,176 คน ใช้สูตร Taro Yamane ได้จำนวนตัวอย่างที่ต้องเก็บข้อมูล เท่ากับ 394.25 หรือ 395 ตัวอย่าง เพื่อความเที่ยงตรงในการวิเคราะห์ข้อมูล ผู้วิจัยจึงกำหนดกลุ่มตัวอย่าง เท่ากับ 440 ตัวอย่าง โดยใช้วิธีการสุ่มตัวอย่างอย่างง่าย แบบสอบถามเป็นเครื่องมือที่ใช้ในการเก็บข้อมูล โดยผ่านการประเมิน IOC จากผู้เชี่ยวชาญจำนวน 5 ท่าน และนำข้อมูลมาวิเคราะห์ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน การวิเคราะห์องค์ประกอบเชิงสำรวจ และการวิเคราะห์องค์ประกอบเชิงยืนยัน ผลการวิเคราะห์ข้อมูลพบว่า การวิเคราะห์องค์ประกอบเชิงสำรวจ ได้องค์ประกอบที่มีผลต่อการยอมรับเทคโนโลยีสารสนเทศ ได้แก่ การรับรู้ประโยชน์ทัศนคติ การรับรู้ความยากง่าย ความคาดหวัง การยอมรับ การวิเคราะห์องค์ประกอบเชิงยืนยันพบว่า ค่าที่ได้และโมเดลมีความสอดคล้องกันเชิงประจักษ์และอยู่ในเกณฑ์ดี

**คำสำคัญ:** การยอมรับเทคโนโลยี เทคโนโลยีสารสนเทศ

### Abstract

The purposes of this research are 1) to analyze the components influencing the acceptance of e-Government (G2E) information technology among civil servants under the Ministry of Justice. 2) to find the consistency of components affecting the acceptance of e-Government (G2E) information technology among civil servants under the Ministry of Justice. The population of this research was 1,176 civil servants under the Ministry of Justice. Using the Taro Yamane formula, the number of samples to be collected is 394.25 or 395 samples. For the accuracy of the data analysis, the researcher determined the sample group equal to 440 samples. using a simple random sampling method. A questionnaire is a tool used to collect data. through the IOC assessment from 5 experts and the data were analyzed for percentage, mean, standard deviation Exploratory factor analysis and Confirmatory Factor Analysis. The results of the data analysis revealed that exploratory factor analysis have elements that affect the acceptance of information technology including perceived benefits, attitudes perception of difficulty, expectation, acceptance. Confirmatory factor analysis revealed that the values obtained and the model were empirically consistent and favorable.

**Keywords:** Technology Acceptance, Information Technology



## 1. บทนำ

ปัจจุบันเทคโนโลยีสารสนเทศต่าง ๆ มีความสำคัญกับชีวิตประจำวันเป็นอย่างมากที่ต้องใช้ในการติดต่อสื่อสารกัน รวมทั้งการใช้อินเทอร์เน็ต ค้นหาข้อมูลต่าง ๆ ทุกวันนี้เราต้องพึ่งพาเทคโนโลยีสารสนเทศและการสื่อสารหรือไอซีที คอมพิวเตอร์ช่วยให้เราเพิ่มขีดความสามารถในการคำนวณและประมวลผลข้อมูลได้เร็วถูกต้อง แม่นยำ อีกทั้งยังเก็บข้อมูลได้จำนวนมากเครือข่ายสื่อสารทำให้การติดต่อถึงกันได้ยากก่อให้เกิดการเปลี่ยนแปลงทางสังคมอย่างรวดเร็ว สังคมใหม่จึงเป็นสังคมที่ต้องพึ่งพาไอซีที ส่งผลให้ผู้ประกอบการในยุคสังคมในปัจจุบันต่างต้องมองหากลยุทธ์ต่าง ๆ ที่จะสามารถช่วยให้อุรกิจนั้นประสบความสำเร็จได้ง่ายและเร็วขึ้น ทั้งนี้ยังรวมไปถึงเครื่องมือในการใช้งานในองค์กรให้มีความก้าวล้ำและทันสมัย เพื่อให้ต่อสู้กับคู่แข่งได้ ดังนั้นจึงหลีกเลี่ยงไม่ได้ที่เรื่องของเทคโนโลยีสารสนเทศที่จะเข้ามามีบทบาทหรือตัวช่วยสำคัญของการจัดการบริหารองค์กรให้มีประสิทธิภาพนั้น จึงเป็นเหตุผลสำคัญ ที่ผู้ประกอบการหรือผู้บริหารองค์กรต่างก็ต้องตระหนักถึงความจำเป็นของการนำเทคโนโลยีสารสนเทศเข้า มาช่วยในการปฏิบัติงาน

รัฐบาลอิเล็กทรอนิกส์มีที่มาจากมติคณะรัฐมนตรีเมื่อปี พ.ศ. 2539 เห็นชอบกรอบนโยบายเทคโนโลยีสารสนเทศปี ค.ศ. 2000 หรือที่เรียกว่า IT 2000 (พ.ศ. 2534-2543) จัดทำโดยคณะกรรมการเทคโนโลยีสารสนเทศแห่งชาติ(กทสช. หรือ NITC) ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ซึ่งกำหนดให้มีการพัฒนาด้านเทคโนโลยีสารสนเทศของประเทศในระยะเวลา 5 ปีแรกโดยมีเป้าหมายในการพัฒนาที่สำคัญ 3 ด้าน คือ 1) ด้านโครงสร้างพื้นฐานสารสนเทศแห่งชาติ (National Information Infrastructure) 2) ด้านการพัฒนาทรัพยากรมนุษย์ (Human Resource Development) 3) ด้านการปรับปรุงกระบวนการปฏิบัติงานและการให้บริการของรัฐ (IT for Good Governance)

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

เทคโนโลยีสารสนเทศ หมายถึง เทคโนโลยีที่เกี่ยวข้องกับการดำเนินการทั้งปวง เพื่อจัดทำสารสนเทศไว้ใช้งาน ซึ่ง

ประกอบไปด้วยเทคโนโลยีคอมพิวเตอร์ เทคโนโลยีคมนาคมเป็นหลักและรวมถึงเทคโนโลยีอื่น ๆ ที่เกี่ยวข้องกับการนำข้อมูลข่าวสารมาใช้ให้เป็นประโยชน์ โดยคอมพิวเตอร์เป็นเครื่องมือในการจัดการและจัดเก็บข้อมูล ส่วนการสื่อสารโทรคมนาคมใช้เป็นสื่อในการจัดส่งข้อมูลเผยแพร่และเสียงออกไปเพื่อสื่อสารกันตลอดจนถึงแนวความคิด ระบบ วิธี เครื่องมือ เครื่องใช้ทางการสื่อสารหรืออุปกรณ์ที่ใช้ ในการจัดเก็บ ประมวลผล ค้นคืน และเผยแพร่ข้อสนเทศ ไม่ว่าจะเป็นอุปกรณ์คอมพิวเตอร์ ระบบข้อมูลและโทรคมนาคม รวมทั้งการประยุกต์ใช้เครื่องมือหรืออุปกรณ์เหล่านั้นในงานสารสนเทศและงานบริการด้านอื่น ๆ (วิทวัส วรินทร์เวช 2541)

แบบจำลองการยอมรับเทคโนโลยี เป็นทฤษฎีที่กล่าวถึงการยอมรับเทคโนโลยีของผู้ใช้งาน โดยกล่าวถึงปัจจัยหลักได้แก่ ตัวแปรภายนอก(External Variables) ที่มีผลต่อการรับรู้ถึงความง่ายในการใช้งาน (Perceived Ease of Use) ซึ่งหมายถึงระดับความเชื่อว่าเทคโนโลยีนั้นไม่ต้องการความพยายามในการใช้งาน (Ease of Use) และการรับรู้ถึงประโยชน์ (Perceived Usefulness) ซึ่งหมายถึงระดับความเชื่อว่าจะสามารถสร้างประโยชน์ต่อผู้ใช้งานได้ โดยการรับรู้ 2 อย่างนี้จะส่งผลไปยัง ทศนคติในการใช้งานเทคโนโลยี (Attitude Toward Using) และส่งผลต่อเนื่องไปยังพฤติกรรมที่มีแนวโน้มจะใช้งาน (Behavioral Intention to use) จากนั้นจะนำไปสู่การใช้งานจริงในที่สุด (Actual System Use) (Technology Acceptance Model: TAM, Davis, (1989)

การรับรู้ถึงประโยชน์จากการใช้เทคโนโลยี (Perceived Usefulness) หมายถึง ทศนคติความเชื่อของบุคคลที่มีต่อการจะใช้เทคโนโลยีหรือระบบใดระบบหนึ่งเพื่อเพิ่มศักยภาพการทำงานของบุคคลนั้น (Davis 1989)

การรับรู้ความง่ายของการใช้เทคโนโลยี (Perceived Ease of Use) หมายถึงการรับรู้ว่าเทคโนโลยีนั้นใช้งานง่าย โดยการใช้นั้นไม่ต้องการความพยายามมาก ไม่มีความซับซ้อนในการใช้งาน จึงมีความเป็นไปได้ที่จะเกิดการยอมรับจากผู้ใช้งาน การรับรู้ความง่ายนั้นมีอิทธิพลทางตรงต่อพฤติกรรมการยอมรับ หรือความตั้งใจที่จะใช้

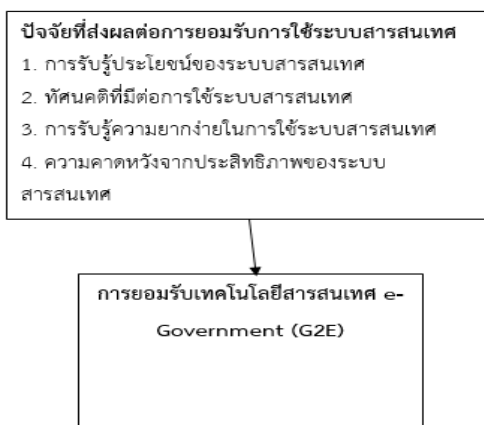
และมีอิทธิพลทางอ้อมต่อการใช้โดยส่งผ่านพฤติกรรมการยอมรับ (Agarwal et. al, 1999; Karrahanna, 1999; Toe et. al.,1999; Venkatesh,2000)

ปัจจัยที่ส่งผลต่อการยอมรับและเจตนาการนำไปใช้โปรแกรมสำเร็จรูป SCM (SUPPLY CHAIN MANAGMENT) ตามทฤษฎีของ TAM (TECHNOLOGY ACCEPTANCE MODEL) แบบจำลองการยอมรับเทคโนโลยี TAM โดยมีปัจจัยดังนี้ 1) ด้านการรับรู้ถึงประโยชน์ที่ได้รับ 2) ด้านการรับรู้ถึงความง่ายในการใช้งาน 3) ด้านทัศนคติที่มีผลต่อพฤติกรรมการใช้ 4) ด้านเจตนาที่จะใช้งาน และปัจจัยภายนอกที่ส่งผลต่อความสำเร็จของการนำระบบ SCM มาใช้ โดยมีปัจจัย ดังนี้ 1) ด้านการฝึกอบรมการใช้งาน 2) ด้านการสนับสนุนจากผู้บริหารระดับสูง 3) ด้านการประสานงาน 4) ด้านการสื่อสารข้อมูล (พรรณานรณ์ แสงดี 2554)

Roger, & Shoemaker (1978) กล่าวว่า การยอมรับของผู้บริโภคที่มีต่อนวัตกรรมและเทคโนโลยี (Adoption and Innovation Theory) เรียกว่า กระบวนการยอมรับ ซึ่งกล่าวถึงพฤติกรรมของบุคคลที่แสดงออกถึงการยอมรับและนำไปปฏิบัติ โดยแบ่งเป็น 5 ขั้นตอน

- ขั้นที่ 1 รับรู้ (Awareness Stage)
- ขั้นที่ 2 สนใจ (Interest Stage)
- ขั้นที่ 3 ประเมินค่า (Evaluation Stage)
- ขั้นที่ 4 ทดลอง (Trial Stage)
- ขั้นที่ 5 ยอมรับ (Adoption Stage)

**กรอบวิจัย**



ภาพที่ 1 กรอบแนวคิดการวิจัย [6]

**3.วิธีดำเนินการวิจัย**

การศึกษาวิจัยนี้เป็นการวิจัยเชิงปริมาณ (Quantitative Research) โดยใช้แบบสอบถามเป็นเครื่องมือในการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างเพื่อวิเคราะห์และแสดงผลของการวิจัยให้ออกมาเป็นจำนวนและร้อยละ โดยทำการวิจัยแยกตามวัตถุประสงค์เพื่อให้สอดคล้องกับวัตถุประสงค์ของการวิจัย

กลุ่มตัวอย่างที่ใช้ในการศึกษา คือ ข้าราชการ สังกัดกระทรวงยุติธรรม โดยกำหนดจำนวนกลุ่มตัวอย่างจากการใช้สูตร Taro Yamane (Taro Yamane 1970) ที่ระดับความเชื่อมั่นร้อยละ 95 ค่าความคลาดเคลื่อนหรือผิดพลาดได้ไม่เกินร้อยละ 5 หรือที่ระดับนัยสำคัญ 0.05 ประชากรที่ทำการศึกษาในครั้งนี้ คือ ข้าราชการ สังกัดกระทรวงยุติธรรม จำนวน 1,176 คน (กองบริหารทรัพยากรบุคคล สำนักงานปลัดกระทรวงยุติธรรม,2563) เมื่อคำนวณแล้วได้จำนวนตัวอย่างที่ต้องเก็บข้อมูล เท่ากับ 394.25 หรือ 395 ตัวอย่าง ทั้งนี้ เพื่อให้การวิเคราะห์ข้อมูลมีความเที่ยงตรง จึงกำหนดกลุ่มตัวอย่าง เท่ากับ 440 ตัวอย่าง โดยวิธีการเก็บข้อมูลแบบสะดวก

แบบสอบถามที่นำมาใช้เก็บข้อมูลผ่านการประเมิน IOC จากผู้เชี่ยวชาญจำนวน 5 ท่าน แบบสอบถามได้แบ่งออกเป็น 4 ตอน ได้แก่ ตอนที่ 1 : เป็นแบบสอบถามเกี่ยวกับปัจจัยด้านข้อมูลส่วนบุคคล ตอนที่ 2 : เป็นแบบสอบถามเกี่ยวกับพฤติกรรมการใช้เทคโนโลยีสารสนเทศ ตอนที่ 3 : เป็นแบบสอบถามเกี่ยวกับปัจจัยที่ส่งผลต่อการยอมรับเทคโนโลยีสารสนเทศ ตอนที่ 4 : เป็นแบบสอบถามเกี่ยวกับการยอมรับเทคโนโลยีสารสนเทศ โดยค่า Cronbach alpha เท่ากับ 0.87

การวิเคราะห์ข้อมูลมีขั้นตอนทางสถิติต่าง ๆ โดยใช้โปรแกรมสำเร็จรูปคอมพิวเตอร์ (SPSS) ในการคำนวณค่าสถิติ ดังนี้

**สถิติเชิงพรรณนา (Descriptive Statistics)**

1. แจกแจงความถี่ (Frequency) และค่าร้อยละ (Percentage) ใช้อธิบายถึงลักษณะข้อมูลส่วนบุคคลของผู้ตอบแบบสอบถาม ได้แก่ เพศ อายุ ระดับการศึกษา ระยะเวลาทำงาน รายได้ต่อเดือน

2. ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ใช้อธิบายปัจจัยที่ส่งผลกระทบต่อ การยอมรับเทคโนโลยีสารสนเทศ และการยอมรับเทคโนโลยีสารสนเทศ e-Government (G2E)

**สถิติเชิงอนุมาน (Inferential Statistics)**

1) การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) เป็นส่วนหนึ่งของเทคนิคการวิเคราะห์องค์ประกอบ (Factor Analysis) เพื่อทำการศึกษาค้นหาความสัมพันธ์ของโครงสร้างของตัวแปร ด้วยโปรแกรมสำเร็จรูปคอมพิวเตอร์ (SPSS)

2) การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) เพื่อตรวจสอบความเที่ยงตรงของโมเดลการวัดในแต่ละองค์ประกอบ ด้วยโปรแกรม IBM AMOS

3) วิเคราะห์โมเดลสมการโครงสร้าง (Structural Equation Model Analysis: SEM) ในการทดสอบแบบจำลอง และสมมติฐานการวิจัย เป็นการวิเคราะห์ความสัมพันธ์เชิงสาเหตุระหว่างตัวแปร ด้วยโปรแกรม IBM AMOS

ดัชนีตรวจสอบความสอดคล้อง	ค่าที่เป็นไปได้	เกณฑ์การพิจารณา
ค่าสถิติทดสอบไคสแควร์ (Chi-square)	0 (perfect fit) to positive value (poor fit)	ค่าไคสแควร์ที่ค่าจำนวนได้น้อยกว่าไคสแควร์ ตาราง หรือพิจารณา ค่า p-value ต้องมากกว่า 0.05
ค่าไคสแควร์สัมพันธ์ (CMIN /df model)	0 (perfect fit) to positive value (poor fit)	มีค่าน้อยกว่า 2.00
Goodness of Fit Index (GFI)	0 (no fit) to 1 (perfect fit)	มีค่ามากกว่า 0.95
Adjusted Goodness of Fit Index (AGFI)	0 (no fit) to 1 (perfect fit)	มีค่ามากกว่า 0.95
Comparative Fit Index (CFI)	0 (no fit) to 1 (perfect fit)	มีค่ามากกว่า 0.95
Tucker - Lewis Index (TLI)	0 (no fit) to 1 (perfect fit)	มีค่ามากกว่า 0.95
Norm Fit Index (NFI)	0 (no fit) to 1 (perfect fit)	มีค่ามากกว่า 0.95
Root mean square error of approximation (RMSEA)	0 (perfect fit) to 0.1 (no fit)	มีค่าน้อยกว่า 0.08
Adjusted goodness of fit index (AGFI)	0 (no fit) to 1 (perfect fit)	มีค่ามากกว่า 0.95
P of close fit (PCLOSE)	มีค่ามากกว่า 0.95	มีค่ามากกว่า 0.5

**ตารางที่ 1 กรอบแนวคิดการวิจัย**

จากตารางที่ 1 ผลจากการวิเคราะห์ข้อมูลค่าที่ได้ต้องอยู่ในเกณฑ์ที่กำหนดซึ่งปรากฏอยู่ในตาราง ดังนี้ (CHI-SQUARE) มากกว่า 0.05 (CMIN /DF MODEL) น้อยกว่า

2.00 (GFI) มากกว่า 0.95 (AGFI) มากกว่า 0.95 (CFI) มากกว่า 0.95 (TLI) มากกว่า 0.95 (NFI) มากกว่า 0.95 (RMSEA) น้อยกว่า 0.08 (AGFI) มากกว่า 0.95 (PCLOSE) มากกว่า 0.5

**4. ผลการดำเนินงาน**

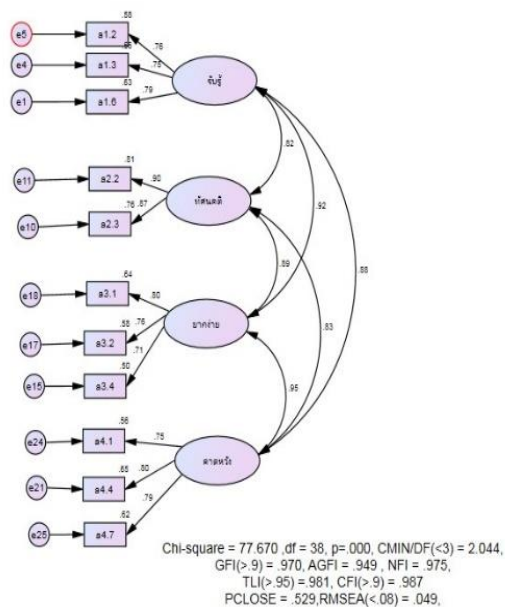
จากการสำรวจข้อมูลการวิเคราะห์องค์ประกอบที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศ e-Government (G2E) ของข้าราชการ สังกัดกระทรวงยุติธรรม สามารถเก็บข้อมูลตัวอย่างได้ทั้งหมดจำนวน 440 ตัวอย่าง แล้วนำมาวิเคราะห์ข้อมูลได้ผล ดังนี้

การวิเคราะห์ปัจจัยส่วนบุคคลกลุ่มตัวอย่าง กลุ่มตัวอย่างเพศชาย จำนวน 262 ตัวอย่าง คิดเป็นร้อยละ 59.5 และกลุ่มตัวอย่างที่เป็นเพศหญิง จำนวน 178 ตัวอย่าง คิดเป็นร้อยละ 40.5 กลุ่มตัวอย่างส่วนใหญ่มีช่วงอายุระหว่าง 21-30 ปี จำนวน 373 ตัวอย่าง คิดเป็นร้อยละ 84.8 กลุ่มตัวอย่างที่มีช่วงอายุระหว่าง 31-40 ปี จำนวน 51 ตัวอย่าง คิดเป็นร้อยละ 11.6 กลุ่มตัวอย่างที่มีช่วงอายุระหว่าง 41-50 ปี จำนวน 6 ตัวอย่าง คิดเป็นร้อยละ 1.4 กลุ่มตัวอย่างที่มีช่วงอายุระหว่าง 51-60 ปี จำนวน 10 ตัวอย่าง คิดเป็นร้อยละ 2.3 ระดับการศึกษาต่ำกว่าปริญญาตรี จำนวน 48 ตัวอย่าง คิดเป็นร้อยละ 10.9 กลุ่มตัวอย่างส่วนใหญ่ที่มีระดับการศึกษาปริญญาตรี จำนวน 377 ตัวอย่าง คิดเป็นร้อยละ 85.7 กลุ่มตัวอย่างที่มีระดับการศึกษาปริญญาโท จำนวน 13 ตัวอย่าง คิดเป็นร้อยละ 3 กลุ่มตัวอย่างที่มีระดับการศึกษาปริญญาเอก จำนวน 2 ตัวอย่าง คิดเป็นร้อยละ 0.5 ระยะเวลาในการปฏิบัติงาน 1-2 ปี จำนวน 67 ตัวอย่าง คิดเป็นร้อยละ 15.2 กลุ่มตัวอย่างส่วนใหญ่มีระยะเวลาในการปฏิบัติงาน 2-3 ปี จำนวน 236 ตัวอย่าง คิดเป็นร้อยละ 53.6 กลุ่มตัวอย่างมีระยะเวลาในการปฏิบัติงาน 3-4 ปี จำนวน 96 ตัวอย่าง คิดเป็นร้อยละ 21.8 กลุ่มตัวอย่างมีระยะเวลาในการปฏิบัติงานมากกว่า 4 ปีขึ้นไป จำนวน 41 ตัวอย่าง คิดเป็นร้อยละ 9.3 รายได้เฉลี่ยไม่เกิน 10,000 บาท จำนวน 17 ตัวอย่าง คิดเป็นร้อยละ 3.9 กลุ่มตัวอย่างส่วนใหญ่มีรายได้เฉลี่ย 10,001-20,000 บาท จำนวน 400 ตัวอย่าง คิดเป็นร้อยละ 90.9 กลุ่มตัวอย่างที่มีรายได้เฉลี่ย 20,001-

30,000 บาท จำนวน 3 ตัวอย่าง คิดเป็นร้อยละ 0.7 กลุ่มตัวอย่างที่มีรายได้เฉลี่ย 30,001-40,000 บาท จำนวน 8 ตัวอย่าง คิดเป็นร้อยละ 1.8 กลุ่มตัวอย่างที่มีรายได้เฉลี่ย 50,000 บาทขึ้นไป จำนวน 12 ตัวอย่าง คิดเป็นร้อยละ 2.7

การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory factor analysis) โดยการแยก 6 ตัวแปรย่อยหรือข้อคำถาม ออกมาเป็น 3 องค์ประกอบ โดยองค์ประกอบที่เหมาะสมกับการนำไปใช้ในการวิเคราะห์องค์ประกอบเชิงยืนยัน ประกอบไปด้วย ด้านการรับรู้ประโยชน์ a1.2 a1.3 a1.6 ด้านทัศนคติ a2.2 a2.3 ด้านการรับรู้ความยากง่าย a3.1 a3.2 a3.4 ด้านความคาดหวัง a4.1 a4.4 a4.7

การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis)



ภาพที่ 2 : ผลการวิเคราะห์ข้อมูลด้วยโปรแกรม AMOS จากภาพที่ 2 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยัน ประกอบด้วย 4 องค์ประกอบ ได้แก่ การรับรู้ประโยชน์ของระบบสารสนเทศ การรับรู้ความยากง่ายในการใช้ระบบสารสนเทศ ความคาดหวังจากประสิทธิภาพของระบบสารสนเทศ ทัศนคติที่มีอิทธิพลต่อการใช้ระบบสารสนเทศ ค่าที่ได้เท่ากับ Chi-square = 77.670 , df = 38 , p = 0.000 , CMIN/DF(<3) = 2.044 , GFI(>0.9) = 0.970 , AGFI = 0.949 , NFI = 0.975 , TLI(>0.95) = 0.981 , CFI(>0.9) = 0.987 , PCLOSE = 0.529 , RMSEA(<0.08) =

0.049 โมเดลมีความสอดคล้องกันเชิงประจักษ์และอยู่ในเกณฑ์ดี

### 5. สรุป อภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์องค์ประกอบที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศ e-Government (G2E) ของข้าราชการ สังกัดกระทรวงยุติธรรม และเพื่อหาความสอดคล้องขององค์ประกอบที่ส่งผลต่อการยอมรับเทคโนโลยีสารสนเทศ e-Government (G2E) ของข้าราชการ สังกัดกระทรวงยุติธรรม ซึ่งสามารถสรุปผล ดังนี้

จากกลุ่มตัวอย่างจำแนกเป็น เพศชาย 59.5% เพศหญิง 40.5% ส่วนใหญ่มีช่วงอายุระหว่าง 21-30 ปี ระดับการศึกษาส่วนใหญ่มีวุฒิปริญญาตรี ที่กล่าวมานี้มีความคิดเห็นที่สัมพันธ์ไปในทิศทางเดียวกัน คือ ปัจจัยที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศทั้ง 4 ด้านที่เป็นตัวแปรต้น มีผลต่อการยอมรับเทคโนโลยีสารสนเทศ

ด้านการรับรู้ประโยชน์ของระบบสารสนเทศ เป็นตัวแปรต้นที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศที่เกี่ยวกับการ ช่วยให้ผู้ใช้งานได้รับข้อมูลข่าวสารอย่างรวดเร็ว ช่วยให้ผู้ใช้งานได้รับข่าวสารที่ถูกต้องมีความน่าเชื่อถือ และระบบสารสนเทศสามารถช่วยลดขั้นตอนในการปฏิบัติงานได้ ซึ่งสอดคล้องกับ สุพจน์ อิงอาจ(2561) พบว่านักศึกษาที่มีการรับรู้ความง่าย และประโยชน์ของการใช้ไอซีทีอยู่ในระดับมาก แสดงให้เห็นว่านักศึกษาในระดับอุดมศึกษาในเขตภาคกลางส่วนใหญ่ให้การยอมรับต่อเทคโนโลยีสารสนเทศและการสื่อสาร หรือ ไอซีที

ด้านการรับรู้ความยากง่ายในการใช้ระบบสารสนเทศ เป็นตัวแปรต้นที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศในส่วนของ การใช้งานง่ายไม่ซับซ้อน สามารถช่วยลดขั้นตอนในการทำงานได้ และระบบสารสนเทศสามารถประยุกต์ใช้กับวิธีการทำงานแบบดั้งเดิมได้

ด้านความคาดหวังจากประสิทธิภาพของระบบสารสนเทศ เป็นตัวแปรต้นที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศในแง่ของ 1.ระบบสารสนเทศที่นำมาใช้ช่วยให้ปฏิบัติงานได้ง่ายขึ้น 2.ระบบสารสนเทศเข้ากับวิถี

การทำงานมากกว่าที่คาดไว้ 3.การใช้งานระบบสารสนเทศช่วยให้การบริการของทางภาครัฐสามารถเข้าถึงได้มากขึ้น

**ด้านทัศนคติที่มีอิทธิพลต่อการใช้ระบบสารสนเทศ** เป็นตัวแปรต้นที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศน้อยที่สุดจากทั้งหมด 4 ด้าน แต่ในด้านของทัศนคติจะมี 2 เหตุผลที่ส่งผลต่อการยอมรับเทคโนโลยีสารสนเทศในด้านทัศนคติ คือ 1.ผู้ใช้จะสามารถค้นคว้าหาข้อมูลในระบบสารสนเทศได้สะดวกและรวดเร็ว 2.การนำระบบสารสนเทศมาใช้งานจะช่วยเอื้อประโยชน์ต่อการปฏิบัติงานของผู้ใช้

โดยที่โมเดลมีความสอดคล้องเชิงประจักษ์และอยู่ในเกณฑ์ดี

## 6. ข้อเสนอแนะในการวิจัย

ในการศึกษาค้นคว้าครั้งนี้เป็นการหาค่าประกอบเชิงยืนยัน (CFA) และการวิเคราะห์องค์ประกอบเชิงของปัจจัยที่มีอิทธิพลต่อการยอมรับเทคโนโลยีสารสนเทศ e-Government (G2E) ของข้าราชการสังกัดกระทรวงยุติธรรม เพื่อให้ทราบถึงปัจจัยสำคัญที่สามารถส่งผลต่อการยอมรับเทคโนโลยีสารสนเทศ

ผลวิจัยนี้เป็นแนวทางเพื่อปรับเปลี่ยนหรือปรับปรุงระบบสารสนเทศให้สามารถรองรับการใช้งานได้สำหรับผู้ปฏิบัติงานทุกเพศ ทุกวัย ได้อย่างเหมาะสม และการบอกกล่าวถึงประโยชน์ของการนำระบบสารสนเทศมาใช้งาน ประโยชน์ที่ได้รับจากระบบสารสนเทศในส่วนของสถาบัน องค์กร พนักงาน ประสิทธิภาพในการปฏิบัติงาน และการลดระยะเวลาในการปฏิบัติงาน

## เอกสารอ้างอิง

- [1] Roger, & Shoemaker (1978) Communication of innovations: A cross-cultural approach. New York: Free Press.
- [2] เฉลิมศักดิ์ บุญประเสริฐ (2562) ได้ศึกษาเรื่อง ปัจจัยที่มีอิทธิพลต่อการยอมรับการใช้แอปพลิเคชันไลน์ของผู้สูงอายุ
- [3] ฉัญฉวนันท์ พิธีวัตโชติกุล (2562) ได้ศึกษาเรื่อง การยอมรับเทคโนโลยีโทรศัพท์มือถือ การตลาดผ่านสื่อสังคมออนไลน์ และพฤติกรรมผู้บริโภคออนไลน์ที่ส่งผลต่อความตั้งใจซื้อ

- ลินค้าออนไลน์ผ่านแอปพลิเคชันของผู้บริโภคในกรุงเทพมหานคร
- [4] อรุโณทัย พยัคฆพงษ์ (2561) ได้ศึกษาเรื่อง แบบจำลองการยอมรับเทคโนโลยีในการแข่งขันทางการตลาด
- [5] วัฒนันท์ วัฒนสังสุทธิ์ (2554) ได้ศึกษาปัจจัยที่ส่งผลต่อการยอมรับเทคโนโลยีนิเวศลิษฐ์ ด้วยการวิจัยเชิงสำรวจเพื่อประโยชน์ทางการเกษตรในประเทศไทย โดยอาศัยพื้นฐานของทฤษฎีความมีเหตุผล (TRA) กลุ่มตัวอย่างเป็นผู้บริโภคสินค้าเกษตรในเขตกรุงเทพมหานคร
- [6] ศศิจันทร์ ปัญจทวี (2560) ปัจจัยที่ส่งผลต่อการยอมรับการใช้ระบบสารสนเทศ กรณีศึกษา สถาบันการพลศึกษา วิทยาเขตเชียงใหม่
- [7] วิทวัส วรินทร์เวช (2545) ความพึงพอใจของหัวหน้างานต่อการรับบริการเทคโนโลยีสารสนเทศ
- [8] Fred D. Davis (1989) การรับรู้ถึงประโยชน์ การรับรู้ถึงความง่ายในการใช้งาน และการยอมรับของผู้ใช้เทคโนโลยีสารสนเทศ
- [9] Ritu Agarwal, Jayesh Prasad (2007) ความแตกต่างระหว่างบุคคลกับการยอมรับเทคโนโลยีสารสนเทศ
- [10] พรรณาภรณ์ แสงดี (2554) ศึกษาวิจัยเรื่อง ปัจจัยที่มีผลต่อการยอมรับเทคโนโลยีโปรแกรมสำเร็จรูป SCM
- [11] สุพจน์ อิงอาจ (2561) อิทธิพลของระบบการจัดการเรียนการสอนที่มีต่อการใช้เทคโนโลยีสารสนเทศและการสื่อสารเพื่อการเรียนรู้ของนักศึกษาระดับอุดมศึกษา

# การวิเคราะห์การออกแบบประสบการณ์ผู้ใช้ด้านมืดบน

## แอปพลิเคชันขายสินค้าออนไลน์ของไทย

### An Analysis of the Dark Pattern of User Experience Online Shopping Application in Thailand

ภูมรินทร์ เตียงพานิชย์ (Pumarin Tiangpanich)<sup>1</sup> และอภิษฐา นิมคุ้มภัย (Apichaya Nimkoompai)<sup>2</sup>

วิทยาศาสตร์มหาบัณฑิต (สาขาเทคโนโลยีสารสนเทศ) คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยี ไทย – ญี่ปุ่น

<sup>1</sup>ti.pumarin\_st@tmi.ac.th, <sup>2</sup>apichaya@tmi.ac.th

#### บทคัดย่อ

ข้อมูลเป็นสิ่งสำคัญสำหรับการตลาดในปัจจุบัน ก่อให้เกิดการต่าง ๆ ต้องการครอบครองข้อมูลโดยไม่คำนึงถึงความยินยอมของผู้ใช้งาน ผู้วิจัยจึงรวบรวมข้อมูลเพื่อสร้างองค์ความรู้ให้ผู้ใช้งานรู้เท่าทันและป้องกันความเข้าใจผิด พร้อมทั้งเข้าใจในผลกระทบที่อาจเกิดขึ้นจากการใช้งานงานออกแบบด้านมืด หรือ Dark Pattern of UX จากการทดสอบและประเมินกลุ่มเป้าหมายพบว่าร้อยละ 58.3 ไม่รู้จัก Dark Pattern of UX และมีความเสี่ยงต่อการเข้าใจผิดจนเกิดความเสียหายสูงถึงร้อยละ 31.3 หรือเป็นจำนวน 32 คนจาก 100 คน

**คำสำคัญ:** ประสบการณ์ผู้ใช้ ปฏิสัมพันธ์ระหว่างมนุษย์และคอมพิวเตอร์ อีคอมเมิร์ซ งานออกแบบด้านมืด โมบายแอปพลิเคชัน

#### Abstract

Nowadays information is principal for marketing. Lead to an organization trying to extract users' personal information without their consent. Thus, the researcher wants to create the knowledge to prevent any effect using the Dark pattern of UX. As testing and evaluating the target group found that 58.3% were unaware of the Dark Pattern of UX and had a 31.3% risk of misunderstanding, or 32 out of 100 people.

**Keywords:** Dark Pattern of UX, User Experience, Human Computer Interaction (HCI), Mobile Application, E-Commerce.

#### 1. บทนำ

จากการเติบโตของเทคโนโลยีและการตลาดรูปแบบใหม่ ส่งผลให้เกิดการเปลี่ยนแปลงทิศทางของธุรกิจต่าง ๆ เพื่อให้ทันต่อสถานการณ์ปัจจุบัน โดยร้อยละ 46 ของคนไทย หลีกเลี่ยงการออกไปซื้อสินค้าในรูปแบบออนไลน์ และผลสำรวจมุมมองผู้บริโภคทั่วโลกประจำปี 2564 ในเดือนมีนาคม 2564 ที่ผ่านมา PricewaterhouseCoopers Legal & Tax Consultants (PwC) กล่าวว่าร้อยละ 62 ของคนไทยเริ่มมีการซื้อสินค้าอุปโภคและบริโภคผ่านแอปพลิเคชันออนไลน์มากขึ้น

จากประเด็นข้างต้นทำให้ผู้วิจัยตระหนักถึงความสำคัญของการศึกษารูปแบบการซื้อสินค้าผ่านแอปพลิเคชันบนมือถือที่ได้มีการนำรูปแบบประสบการณ์ผู้ใช้งาน (User Experience : UX) มาร่วมพิจารณาในการออกแบบ เพื่อให้ได้ผลกำไรทางการค้ามากขึ้น จึงจำเป็นต้องมีการศึกษาถึงการเข้าถึงและกระบวนการการเก็บข้อมูลและประสบการณ์ของผู้ใช้งานจากกลุ่มเป้าหมาย แต่อย่างไรก็ตามในการใช้งานแอปพลิเคชันซื้อขายสินค้าในปัจจุบันยังมีรูปแบบที่ทำให้ผู้บริโภคเกิดความเข้าใจที่ผิดพลาด ซึ่งส่งผลเสียให้กับผู้ใช้งาน เรียกว่า งานออกแบบด้านมืด หรือ Dark Pattern of UX ที่มีขึ้นเพื่อแสดงถึงกลวิธีของการออกแบบส่วนต่อประสานกับผู้ใช้งาน (User Interface : UI) ที่นักออกแบบส่วนใหญ่ใช้เพื่อล่อลวงให้ผู้ใช้งานตัดสินใจผิดอันเกิดจากกระบวนการทางความคิดของผู้ใช้งาน จนทำให้ผู้ใช้งานทำอะไรบางอย่างเพื่อให้ได้สิ่งที่ต้องการ [1][2][3] “การที่ UI ได้รับการออกแบบมาอย่างพิถีพิถันเพื่อหลอกให้ผู้ใช้กระทำสิ่งใดสิ่งหนึ่ง

เพื่อให้บรรลุความต้องการนั้น... ไม่ใช่ความคิดพลาดแต่อย่างใด แต่มันถูกสร้างขึ้นอย่างรอบคอบด้วยความเข้าใจอย่างถ่องแท้เกี่ยวกับจิตวิทยาของมนุษย์โดยไม่คำนึงถึงความสนใจของผู้ใช้งาน” [3][6] จากข้อมูลดังกล่าวมา ทำให้ผู้วิจัยมีแนวคิดในการศึกษาเกี่ยวกับ Dark Pattern of UX เพื่อหารูปแบบที่มีการใช้งานบนแอปพลิเคชันอีคอมเมิร์ซในปัจจุบัน และจัดทาคำอธิบายให้กับกลุ่มนักออกแบบระดับมัธยมศึกษาในการออกแบบที่อาจทำให้ผู้ใช้งานเข้าใจผิดและใช้งานผิดพลาด ซึ่งอาจส่งผลกระทบต่อผู้ใช้เล็กใช้งานแอปพลิเคชันนั้นได้ และเพื่อเป็นแนวทางหรือองค์ความรู้ให้กับผู้ใช้ทั่วไปที่ไม่ได้มีความรู้ในด้านการใช้งานมากนักไม่ให้เกิดความเข้าใจผิดพลาดในการใช้งานจนทำให้เกิดความเสียหายอย่างใหญ่หลวงในอนาคต

## 2. วรรณกรรมที่เกี่ยวข้อง

การออกแบบใด ๆ ที่ทำขึ้นเพื่อให้ได้ข้อมูลส่วนตัวของผู้ใช้งานอันเป็นผลประโยชน์แก่บริษัทและองค์กรและเพื่อให้ผู้ใช้งานดำเนินการตามเป้าหมายที่องค์กรวางไว้ เช่น การซื้อสินค้า, การสมัครสมาชิกบริการ, เป็นต้น โดย Dark Pattern of UX มีผลเมื่อผู้ใช้งานไม่ได้ระมัดระวังในการอ่านข้อมูลให้ครบถ้วนหรือการกระทำกิจกรรมใด ๆ ที่อาจเกิดช่องโหว่อันส่งผลให้องค์กรต่าง ๆ สามารถถือเป็นผลประโยชน์ส่วนตัวได้ ทั้งนี้สามารถพบ Dark Pattern of UX ได้ผ่านธุรกิจ 3 ประเภท คือ (1) ธุรกิจขายสินค้าและบริการออนไลน์ (2) ความเป็นส่วนตัวและการยินยอม และ (3) ความสนใจและเกมส์ ทั้งนี้ผลลัพธ์ของประเภท Dark Pattern of UX [4] ประกอบกับการศึกษาเพิ่มเติม [5] ทำให้สามารถจำแนกออกได้ดังตารางที่ 1

ตารางที่ 1 : ประเภทของ Dark Pattern of User Experience ที่สามารถพบเจอได้ในปัจจุบัน

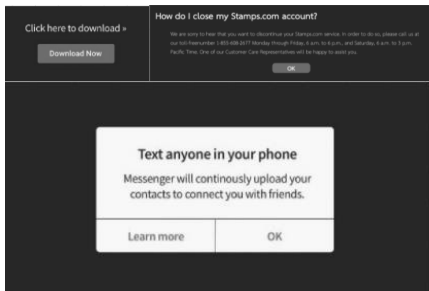
ประเภทของ Dark Patterns	คำอธิบาย
เหยื่อและสวิตช์ (Bait and Switch)	การตั้งใจทำสิ่งหนึ่ง แต่สิ่งอื่นไม่พึงปรารถนากลับเกิดขึ้นแทน

การยอมรับความอับอาย (Confirm shaming)	การกระทำที่ทำให้ผู้ใช้งานรู้สึกผิด มักใช้ข้อความที่ทำให้ผู้ใช้งานอับอาย
การปลอมโฆษณา (Disguised Ad)	โฆษณาที่ปลอมเนื้อหาหรือรูปแบบให้ผู้ใช้งานสนใจ
การบังคับให้ใช้งานต่อเนื่อง (Forced Continuity)	เมื่อช่วงทดลองสิ้นสุดลง ผู้ใช้งานถูกเรียกเก็บเงิน แบบไม่มีการแจ้งเตือนและยกเลิกการต่ออายุอัตโนมัติไม่ได้
สแปมเพื่อน (Friend Spam)	การขอสิทธิ์ข้อมูลส่วนตัว เพื่อสแปมรายชื่อผู้ติดต่อของผู้ใช้งาน
การซ่อนราคา (Hidden Costs)	การเรียกเก็บเงินเพิ่มเติมโดยไม่แจ้งให้ทราบล่วงหน้า ในขั้นตอนสุดท้ายของการทำธุรกรรม
การนำไปสู่ทิศทางที่ผิด (Misdirection)	การออกแบบเพื่อเบี่ยงเบนความสนใจของผู้ใช้งาน ไปในทางที่ไม่ถูกต้องจนก่อให้เกิดความเสียหายต่าง ๆ
การป้องกันการเปรียบเทียบราคา (Price Comparison prevention)	การซ่อนราคาสินค้าที่แท้จริง (Hidden Costs) ทำให้ผู้ใช้งานไม่สามารถเปรียบเทียบราคาได้
ความเป็นส่วนตัว - ซักเคอร์ริง (Privacy Zuckering)	การถูกหลอกให้เผยแพร่ข้อมูลส่วนตัวมากเกินไปจนเกิดความจำเป็น
โรช โมเต็ล (Roach Motel)	การนำผู้ใช้งานเข้าสู่ระบบได้ง่าย แต่ยากต่อการออกจากระบบเหล่านั้น ยกตัวอย่างเช่น การลบบัญชี เป็นต้น
การแอบเข้าไปในตะกร้า (Sneak into Basket)	การหลอกล่อผู้ใช้งานด้วยข้อมูลที่น่าสนใจต่าง ๆ เช่น โปรโมชั่น เพื่อเพิ่มจำนวนสินค้าเข้าสู่ตะกร้า
คำถามหลอกหลวง (Trick Questions)	การตั้งคำถามกำกวม อันส่งผลให้ผู้ใช้งานตอบคำถามโดยไม่ได้ตั้งใจ

จากตารางที่ 1 สามารถสรุปรูปแบบของ Dark Pattern of UX ที่นิยมใช้ในปัจจุบัน โดยการศึกษากระบวนการสร้าง UI ที่ได้รับการออกแบบมาอย่างพิถีพิถันเพื่อ

นำมาใช้ใน Dark Pattern of UX [6] สามารถจำแนกออกได้เป็น 3 ประเภทการใช้งานหลัก ๆ ดังนี้

1. General Dark patterns รูปแบบที่สามารถพบได้ทั้งในเว็บไซต์และแอปพลิเคชันทั่วไป
2. E-Commerce related Dark patterns รูปแบบที่สามารถพบได้เฉพาะในเว็บไซต์หรืออีคอมเมิร์ซแอปพลิเคชันเท่านั้น
3. Privacy-related Dark patterns รูปแบบที่สามารถพบได้เฉพาะในโซเชียลคอมมูนิตีเท่านั้น



ภาพที่ 1: ตัวอย่างของแต่ละประเภทการใช้งาน

จากภาพที่ 1 เป็นการแสดงถึงภาพตัวอย่างของ Dark Pattern of UX ที่จำแนกตามประเภทการใช้งาน โดยในตัวอย่างเป็น (1) การปลอมโฆษณา (2) โฆษ โมเติ้ล และ (3) ความเป็นส่วนตัว-ซั๊กเคอร์รี่ง

เมื่ออุปกรณ์เทคโนโลยีแบบพกพาอย่าง โทรศัพท์มือถือ กลายเป็นปัจจัยสำคัญในการใช้ชีวิต ทั้งเพื่อติดต่อสื่อสาร ความบันเทิง และอื่น ๆ จากการวิเคราะห์และค้นคว้าพบว่าการออกแบบ UI สามารถตอบสนองต่อความต้องการ ของผู้ใช้ สำหรับแอปพลิเคชันมือถือได้ [7][8][9] ก่อให้เกิดรูปแบบ Dark Pattern of UX [10][11][12] ที่มักนิยมใช้ในโทรศัพท์มือถือออกมาได้ดังตารางที่ 2

ตารางที่ 2 : รูปแบบที่นิยมใช้ในโทรศัพท์

รูปแบบนิยมใช้ในโทรศัพท์	รายละเอียด
การร้องขอข้อมูลสถานที่ที่มากเกินไปจนเกินความจำเป็น (Too Much Location Information)	การร้องขอข้อมูลที่อยู่และข้อมูลส่วนตัวโดยไม่จำเป็น โดยผู้ใช้งานไม่สามารถปฏิเสธได้
การแจ้งเตือนที่มากเกินไป	การแจ้งเตือนมากเกินไปจนเกินความจำเป็น

ความจำเป็น (Notification Fatigue)	เพื่อซ่อนหรือปิดข้อมูลสำคัญที่ผู้ใช้งานควรรู้
ข้อจำกัดในการใช้งานผ่านมือถือ (Limiting Mobile Functionality)	การทำให้ระบบที่เจอร์สามารถใช้งานได้ผ่านคอมพิวเตอร์เพียงอย่างเดียวหรือเปิดให้บริการฟรีเฉพาะบนคอมพิวเตอร์เท่านั้น
สิทธิ์ที่จะทำหรือไม่ทำ (Do-or-Die Permission)	เมื่อผู้ใช้งานไม่ยินยอมให้ข้อมูลระบบแอปพลิเคชันปิดตัวลงเพื่อให้เข้าสู่ระบบได้ ผู้ใช้งานจำเป็นต้องยินยอมให้ข้อมูล
ไม่ชัดเจนหรือยากที่จะสัมผัสปุ่ม (Unclear or Impossible Touch Interactions)	การทำให้ปุ่มปฏิเสธหรือยกเลิกการซื้อหรือการใช้งานเพื่อขัดขวางการทำกิจกรรมใด ๆ ของผู้ใช้งานให้ไม่สามารถดำเนินการได้สมบูรณ์
การเปลี่ยนการแสดงผลแบบโทรศัพท์มือถือแบบฉับพลัน (Suddenly not mobile)	การออกแบบให้แสดงผลในรูปแบบของเว็บไซต์สำหรับคอมพิวเตอร์แทนแอปพลิเคชันเพื่อปกปิดข้อมูลที่ผู้ใช้งานควรรู้
จรรยาบรรณและการออกแบบตามกฎหมาย (Ethical and Legal Design)	การตั้งค่าให้ระบบจดจำข้อมูลที่อาจผิดต่อจรรยาบรรณและกฎหมาย เพื่อหลอกให้ผู้ใช้ยืนยันข้อมูลจนเกิดความเสียหาย

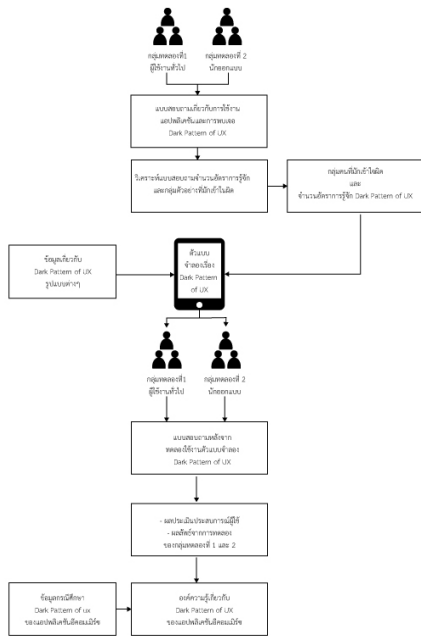
จากตารางที่ 2 สามารถสรุปรูปแบบที่ Dark Pattern of UX [10][11][12] ที่มักนิยมใช้ใน โทรศัพท์มือถือได้เป็น 8 ประเภท ขึ้นอยู่กับสถานการณ์ต่าง ๆ ที่ผู้ใช้งานพบเจอ เช่น โฆษณาที่ขอกดปิดบนแอปพลิเคชันที่ให้บริการฟรี เป็นต้น

### 3. วิธีการดำเนินการวิจัย

#### 3.1 รูปแบบวิธีการดำเนินงานวิจัย

ผู้วิจัยได้วางแผนวิธีการดำเนินการวิจัยเพื่อจัดทำองค์ความรู้เกี่ยวกับ Dark Pattern of UX ของกลุ่มตัวอย่าง โดยแบ่งออกเป็น 2 กลุ่ม ได้แก่ (1) กลุ่มบุคคลทั่วไป ที่ไม่มี ความรู้ ความเข้าใจในการออกแบบ โดยอิงประสบการณ์ ผู้ใช้งาน จำนวน 48 คน และ (2) กลุ่มนักออกแบบที่มีความสามารถในการออกแบบและเข้าใจถึงการออกแบบโดยอิงประสบการณ์ผู้ใช้งานเป็นหลัก จำนวน 48 คน ซึ่งสามารถนำเสนอวิธีการดำเนินการวิจัยได้ดังภาพที่ 2



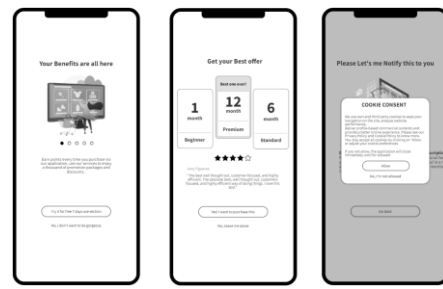


ภาพที่ 2: ภาพขั้นตอนการดำเนินงานวิจัย

จากภาพที่ 2 ผู้วิจัยได้จำแนกขั้นตอนเป็น (1) การรวบรวมข้อมูลเบื้องต้นเกี่ยวกับ Dark Pattern of UX จากกลุ่มตัวอย่างทั้งสองกลุ่ม (2) วิเคราะห์ข้อมูลเพื่อหาบุคคลที่มีแนวโน้มถูก Dark Pattern of UX ทำให้เข้าใจผิดได้ง่าย (3) ค้นคว้าและวางโครงสร้างของข้อมูลและศึกษาถึงประสบการณ์ของผู้ใช้งานในการใช้งานแอปพลิเคชันอเมริกา เพื่อจัดทำแอปพลิเคชันต้นแบบสำหรับทดสอบกลุ่มตัวอย่าง (4) ทดสอบกลุ่มตัวอย่างและประเมินถึงผลความเสี่ยงของการใช้งานแอปพลิเคชันและความเข้าใจของการออกแบบที่ก่อให้เกิดความเข้าใจผิดต่อผู้ใช้งาน โดยทดสอบในหน้าต่าง ๆ ของแอปพลิเคชันจำลอง และ (5) จัดทำองค์ความรู้และสร้างความเข้าใจให้แก่กลุ่มตัวอย่างทั้งสองกลุ่มถึง Dark Pattern of UX และผลกระทบของการใช้งานและการนำมาออกแบบภายในแอปพลิเคชันอเมริกา

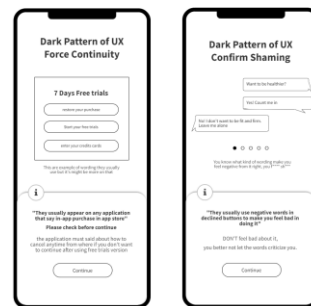
**3.2 การพัฒนาแอปพลิเคชันองค์ความรู้สำหรับกลุ่มตัวอย่างที่ 1 กลุ่มบุคคลทั่วไป**

ผู้วิจัยได้วางแผนการนำเสนอเพื่อจัดทำองค์ความรู้แก่กลุ่มตัวอย่างที่ 1 กลุ่มบุคคลทั่วไป ด้วยวิธีการให้ผู้ทดลองเข้าใช้งานแอปพลิเคชันและดำเนินกิจกรรมจนจบกระบวนการการใช้งาน



ภาพที่ 3: แอปพลิเคชันเพื่อใช้ทดลองกลุ่มตัวอย่างที่ 1

จากภาพที่ 3 ผู้วิจัยได้ยกตัวอย่างรูปแบบ Dark Pattern of UX ในอเมริกาแอปพลิเคชันที่ใช้ในการทดลอง โดยมีประเภทต่าง ๆ โดยเรียงจากซ้ายไปขวาดังนี้ (1) การยอมรับความอับอาย (2) การบังคับให้ใช้งานต่อเนื่อง (3) การป้องกันการเปรียบเทียบราคา (4) Suddenly not mobile (5) การร้องขอข้อมูลสถานที่ที่มากเกินไปจนจำใจเป็น และ (6) โรซ โมเต็ล

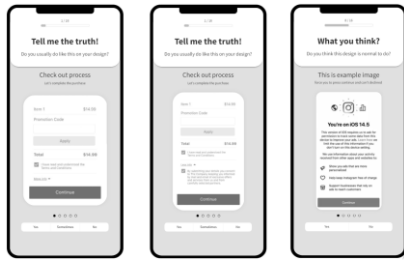


ภาพที่ 4: การนำเสนอองค์ความรู้ผ่านแอปพลิเคชันหลังการทดลองของกลุ่มตัวอย่างที่ 1

จากภาพที่ 4 ผู้วิจัยได้ยกตัวอย่างรูปแบบการนำเสนอองค์ความรู้เกี่ยวกับ Dark Pattern of UX ของรูปแบบการบังคับให้ใช้งานต่อเนื่องและการยอมรับความอับอาย โดยเรียงจากซ้ายไปขวา ด้วยการนำรูปภาพตัวอย่าง ข้อความหรือตัวอย่างการใช้งานจริงมาอธิบายเพิ่มเติม ถึงรูปแบบวิธีการสังเกตและวิธีการป้องกันเบื้องต้นจาก Dark Pattern of UX ของแต่ละประเภท โดยอิงผลลัพธ์จากการทดลองของผู้ใช้งานผ่านแอปพลิเคชันในขั้นตอนก่อนหน้า

**3.3 การพัฒนาแอปพลิเคชันองค์ความรู้สำหรับกลุ่มตัวอย่างที่ 2 กลุ่มนักออกแบบ**

ที่มีความรู้ความเข้าใจในการออกแบบเป็นอย่างดี ทางผู้วิจัย ได้มีวิธีการนำเสนอการทดลองดังภาพ

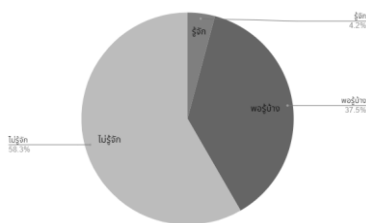


ภาพที่ 5: แอปพลิเคชันเพื่อใช้ทดสอบกลุ่มตัวอย่างที่ 2

จากภาพที่ 5 ผู้วิจัยได้ยกตัวอย่างรูปแบบแอปพลิเคชันที่ใช้ในการทดลองกลุ่มตัวอย่างที่ 2 โดยนำภาพที่มีการใช้ Dark Pattern of UX มาใช้ในการออกแบบมาสอบถามกลุ่มตัวอย่างเพื่อเก็บข้อมูลและประเมินผลลัพธ์ถึงความเข้าใจในการออกแบบของกลุ่มตัวอย่างที่ 2 อันส่งผลให้ผู้ใช้งานทั่วไปเกิดความเข้าใจผิดหรือเกิดประสบการณ์ที่ไม่ดีต่อการใช้งานและอาจส่งผลเสียต่าง ๆ ได้ในอนาคต

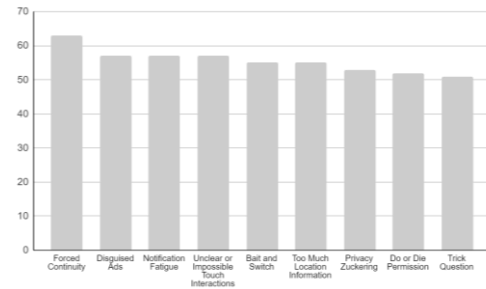
#### 4. ผลการดำเนินการ

จากการทดสอบและประเมินผลลัพธ์ของผู้เข้าร่วมการทดสอบจำนวน 96 คน โดยแบ่งกลุ่มตัวอย่างที่ 1 เป็นจำนวน 48 คนและกลุ่มตัวอย่างที่ 2 จำนวน 48 คน เพื่อให้สามารถสรุปผลการทดสอบของผู้ใช้งานผ่านการทดลองใช้แอปพลิเคชันและแบบสอบถาม โดยแบบสอบถามได้มีการใช้ผู้เชี่ยวชาญ 3 คนในการประเมิน โดยมีค่า IOC มากกว่าหรือเท่ากับ 0.05 ทั้งนี้สามารถแยกข้อมูลออกเป็นประเด็นต่าง ๆ ได้ดังนี้



ภาพที่ 6: ผลลัพธ์อัตราความรู้จัก Dark Pattern of UX ก่อนการเข้าทดสอบของกลุ่มตัวอย่าง

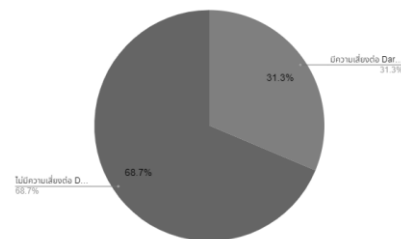
จากภาพที่ 6 สามารถสรุปได้ว่าร้อยละ 58.3 หรือกลุ่มตัวอย่างจำนวน 56 คน จากทั้งหมดจำนวน 96 คน ไม่รู้จัก Dark Pattern of UX มาก่อนและไม่รู้ถึงผลกระทบจากการเข้าใจผิดที่อาจเกิดขึ้นกับตัวเองได้ ทั้งนี้ผลลัพธ์การทำแบบทดสอบสามารถสรุปข้อมูลของกลุ่มตัวอย่างได้ดังภาพที่ 7



ภาพที่ 7: รูปแบบที่ผู้ใช้งานมักเข้าใจผิดและนักออกแบบมักใช้วิธีการนี้ออกแบบเป็นประจำ

จากภาพที่ 7 สามารถสรุป 3 อันดับรูปแบบที่ผู้ใช้งานมักเข้าใจผิดและนักออกแบบมักใช้วิธีการนี้ออกแบบเป็นประจำได้แก่ (1) การบังคับให้ใช้งานต่อเนื่องจำนวน 63 คน (2) การปลอมโฆษณาจำนวน 57 คน และ (3) การแจ้งเตือนที่มากเกินไปจนเกิดความจำเป็นจำนวน 57 คน และรูปแบบที่สามารถสร้างความเข้าใจผิดได้น้อยที่สุดต่อผู้ใช้งาน 3 อันดับ ได้แก่ การแอบเข้าไปในตะกร้าจำนวน 39 คน การป้องกันการเปรียบเทียบราคา จำนวน 40 คน และ โฆษ โหม เตล จำนวน 42 คน ตามลำดับ

ในการประเมินความเสี่ยงของกลุ่มตัวอย่างผ่านการใช้แบบทดสอบหลังการทดสอบ Dark Pattern of UX ของแต่ละประเภท ทำให้สรุปได้ดังภาพที่ 8



ภาพที่ 8: ความเสี่ยงจากการทดสอบ Dark Pattern of UX

จากภาพที่ 8 การประเมินการผลทดสอบเกี่ยวกับ Dark Pattern of UX ของกลุ่มตัวอย่าง พบว่ามีโอกาสเข้าใจผิดจนก่อให้เกิดความเสี่ยงในการสูญเสียข้อมูลทางทรัพย์สินมากถึงร้อยละ 31.3 หรือในจำนวน 100 คน มีโอกาสเสี่ยงต่อ Dark Pattern of UX จำนวน 32 คน ทั้งนี้หากเทียบกับจำนวนของประชากรในประเทศไทยปี 2565 มีจำนวน 67 ล้านคน ตามการประกาศของราชกิจจานุเบกษา เมื่อวันที่ 19 มกราคม 2565 โดยโอกาสที่คนไทยสามารถพบเจอและเกิดความเสี่ยงขึ้นมีเท่ากับ 1 ใน 3 ของประชากร หรือมีจำนวนเทียบเท่ากับ 21 กว่าล้านคน

## 5. สรุป

เพื่อป้องกันและเพื่อไม่ให้เกิดความเข้าใจผิดพลาดในการใช้งานจนทำให้เกิดความเสียหายในอนาคต โดยจากผลสำรวจเกี่ยวกับ Dark Pattern of UX ที่ผู้วิจัยนำไปสอบถามกลุ่มเป้าหมายพบว่าร้อยละ 58.3 ไม่รู้จัก Dark Pattern of UX มาก่อน และไม่รู้จักผลกระทบจากการเข้าใจผิดที่อาจเกิดขึ้นกับตัวเอง หากการเติบโตของเทคโนโลยียังพัฒนาต่อและไร้ซึ่งกฎหมาย ผู้ใช้งานอาจเกิดความเสี่ยงต่อการสูญเสียทรัพย์สินได้ถึงร้อยละ 31.3 ผู้วิจัยจึงสร้างองค์ความรู้ให้แก่กลุ่มเป้าหมายในรูปแบบของแอปพลิเคชัน ทั้งนี้ในอนาคตผู้วิจัยได้วางแผนจะนำเสนอในรูปแบบของเกมแอปพลิเคชันเพื่อสร้างองค์ความรู้ให้แก่เยาวชนที่อาจตกเป็นเหยื่อของ Dark Pattern of UX ได้ในอนาคตตามการเติบโตของยุคสมัยจักรวาลนฤมิต (Metaverse) ในปัจจุบัน

## เอกสารอ้างอิง

- [1] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, S. Wilson. "Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online." *ACM Computing Surveys*, vol. 50(3), no. 44, pp 1-41, 2018.
- [2] T. Mirsch, C. Lehrer, R. Jung. "Digital Nudging: Altering User Behavior in Digital Environments." *association for Information Systems. AIS Electronic Library (AISeL)*, pp 634-648, 2017.
- [3] M. Jesse, and D. Jannach. "Digital nudging with recommender systems: Survey and future directions." *Computers in Human Behavior Reports*, vol. 3, pp. 65-67, 2021.
- [4] C. M. Gray. "The Dark (Patterns) Side of UX Design." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, no. 534, pp. 1-14, 2018.
- [5] C. Bösch, B. Erb, F. Kargl, S. K. Pfattheicher. "Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns." *Proceedings on Privacy Enhancing Technologies*, pp. 237-254, 2016.
- [6] N. Emma. "Dark patterns and their use in E-Commerce." *Jyväskylä: University of Jyväskylä*, pp. 1-33, 2020.
- [7] Interaction Design Foundation. (2020 Sep) "Getting into the Mobile App User's Mind – User Research for Mobile Applications" Available Online at <https://www.interaction-design.org/literature/article/getting-into-the-mobile-app-user-s-mind-user-research-for-mobile-applications>, accessed on 6 Aug 2021.
- [8] M. A. Yazid, and A. H. Jantan. "User Experience Design (UXD) of Mobile Application: An Implementation of a Case Study." *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 9, no. 3, pp. 197-200, 2017.
- [9] Interaction Design Foundation. (2018 Jun). "Mobile User Experience (UX) Design." Available Online at <https://www.interaction-design.org/literature/topics/mobile-ux-design>, accessed on 5 Nov 2021.
- [10] S. Hooper. (2019 Nov). "Mobile Dark Patterns." Available Online at <https://www.uxmatters.com/mt/archives/2019/11/mobile-dark-patterns.php>, accessed on 28 Oct 2021.
- [11] L. Di Geronimo, L. Braz, E. Fregnan, F. Palomba, A. Bacchelli. "UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception." *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1-14, 2020.
- [12] S. Saetang. "The E-Commerce strategies responding to the UX design." *10th International Conference on Ubi-media Computing and Workshops (Ubi-Media)*, pp. 1-6, 2017.

# การพยากรณ์ปริมาณผลผลิตลำไยในภาคเหนือของประเทศไทย

## โดยใช้ข้อมูลแบบอนุกรมเวลา

### Predicting Longan Yield in Northern of Thailand using Time Series Data

ธัญชนก รังแก้ว (Thanchanok Rangkaew)<sup>1</sup> เอกสิทธิ์ พัทธวงษ์ศักดิ์ (Eakasit Pacharawongsakda)<sup>2</sup>

กันทรกร จิตต์หาญ (Kantarakorn Jitharn)<sup>3</sup> สกลภักดิ์ เจียรวัฒนสวัสดิ์ (Sakolpuck Jearawattanasawadi)<sup>4</sup>

และชุติมณฑน์ ภัคคีสิโรตม์ (Chutimon Pakdeesirote)<sup>5</sup>

<sup>1</sup>บริษัท คิวบี อนาคติกส์ คอนซัลติงจำกัด

<sup>2,3,4,5</sup>สาขาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม

61010912559@msu.ac.th, eakasit@datacubeth.ai, kantarakorn@datacubeth.ai,

sakolpuck@datacubeth.ai, chutimon@datacubeth.ai

#### บทคัดย่อ

งานวิจัยนี้เป็นการพยากรณ์ปริมาณผลผลิตลำไยในภาคเหนือของประเทศไทยที่เกี่ยวข้องกับประเภทดิน และสภาพอากาศที่เป็นอนุกรมเวลา (Time Series) โดยประยุกต์ใช้การเรียนรู้ของเทคนิคการวิเคราะห์การถดถอย (Regression) ด้วยวิธี  $k$ -Nearest Neighbors ( $k$ -NN), Gradient Boosted Trees (GBT) และ Support Vector Machine (SVM) ในการพยากรณ์ปริมาณผลผลิตลำไยรายเดือนของภาคเหนือ โดยสามารถใช้ข้อมูลดังกล่าวในการวางแผนการผลิตและจัดทำนโยบายส่งเสริมกลุ่มผู้ผลิตได้อย่างเหมาะสมมากและใกล้เคียงกับภาวะการณ์ที่เกิดขึ้นจริง

**คำสำคัญ:** ปริมาณผลผลิตลำไย อนุกรมเวลา การพยากรณ์

#### Abstract

This research is about a forecasting of longan yield in northern of Thailand by having soil type and time series weather as input data. Machine learning techniques have been used by applying the regression model of  $k$ -Nearest Neighbors ( $k$ -NN), Gradient Boosted Trees (GBT), and Support Vector Machine (SVM) for predicting monthly longan yield. The prediction result can be used to plan production and

formulate a policy for promoting more appropriately producer groups related to actual situation.

**Keywords:** longan yield, Time Series, Predicting.

#### 1. บทนำ

ลำไยเป็นพืชเศรษฐกิจที่สำคัญชนิดหนึ่งของประเทศไทย อีกทั้งเมื่อพิจารณาข้อมูลทางเศรษฐกิจการเกษตรของประเทศพบว่าประเทศไทยติดอันดับประเทศที่มียอดการส่งออกลำไยเป็นลำดับ 1 ของโลก และผลผลิตลำไยได้เป็นอันดับ 2 ของโลก [1] โดยแหล่งเพาะปลูกที่สำคัญอยู่ทางภาคเหนือ ซึ่งมีภูมิประเทศและภูมิอากาศเหมาะสำหรับการปลูกลำไยเป็นอย่างมาก การปลูกลำไยตามธรรมชาติมีปัญหาที่สำคัญ คือ การออกดอกและการติดผลไม่สม่ำเสมอสาเหตุอาจเนื่องมาจากอิทธิพลของสภาพภูมิอากาศในแต่ละเดือนไม่เหมาะสมหรือไม่เพียงพอที่จะทำให้ลำไยออกดอกและติดผลได้ [2] จากงานวิจัยในจังหวัดเชียงใหม่ พบว่าการเปลี่ยนแปลงของอุณหภูมิ ความชื้นสัมพัทธ์ รวมถึงปริมาณน้ำฝน ถือว่าเป็นปัจจัยหลักที่ส่งผลกระทบต่อผลผลิตลำไย [3] นอกจากนี้ยังมีสาเหตุอื่น ๆ อีก เช่น ปริมาณธาตุอาหารและความอุดมสมบูรณ์ของดิน [4]

เนื่องจากในช่วง 3 ปีที่ผ่านมา ทั่วโลกประสบกับการแพร่ระบาดของโรคโควิด-19 ซึ่งส่งผลกระทบต่อ

หลีกเลี่ยงไม่ได้กับทุกภาคส่วน แม้แต่กระบวนการการผลิตลำไย ปัญหาที่มีความสำคัญคือการคาดการณ์ปริมาณผลผลิตและปริมาณความต้องการของผู้บริโภค ให้มีความสมดุลกันทั้งด้านของราคาที่เหมาะสมสามารถยอมรับได้เมื่อเปรียบเทียบกับต้นทุนการผลิต และความเหมาะสมกับปัจจัยต่าง ๆ ในตลาดนั้นอาจจะทำได้ยาก การนำระบบสารสนเทศเข้ามาช่วยในการวิเคราะห์ข้อมูล โดยนำทฤษฎีการทำเหมืองข้อมูล (Data Mining) เทคนิคต่าง ๆ เช่น การประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียม (Neural Network) [5], แบบจำลองเกรเดียนบูททรี (Gradient Boosted Trees) [6] และแบบจำลองซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) [7] เพื่อคาดการณ์ปริมาณผลผลิตลำไย จึงถูกพัฒนาขึ้น เพื่อให้การวางแผนกระบวนการผลิตลำไยของภาคเหนือให้มีประสิทธิภาพมากยิ่งขึ้น

งานวิจัยนี้ได้ประยุกต์ใช้เทคนิคการวิเคราะห์การถดถอย (Regression) ด้วยวิธีเคเนียร์สเนเบอร์ (k-Nearest Neighbours), เกรเดียนบูททรี (Gradient Boosted Trees) และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) กับข้อมูลปริมาณผลผลิตลำไย ข้อมูลประเภดิน และข้อมูลสภาพอากาศที่เป็นอนุกรมเวลา (Time Series) เพื่อสร้างแบบจำลองในการทำนายปริมาณผลผลิตลำไยในภาคเหนือของประเทศไทย

## 2. แนวคิดและทฤษฎีที่เกี่ยวข้อง

**2.1 ข้อมูลอนุกรมเวลา (Time Series)** คือ ชุดของข้อมูลที่เก็บรวบรวมตามระยะเวลาเป็นช่วง ๆ อย่างต่อเนื่องกัน อาจอยู่ในลักษณะที่เป็นข้อมูลรายปี รายไตรมาส หรือรายเดือนก็ได้ ทั้งนี้ขึ้นอยู่กับความเหมาะสมในการนำไปใช้ประโยชน์

**2.2 การทำเหมืองข้อมูล (Data Mining)** เป็นการค้นหาความรู้ที่เป็นประโยชน์และน่าสนใจบนฐานข้อมูลขนาดใหญ่ (Knowledge discovery in databases - KDD) โดยนำข้อมูลที่มีอยู่มาวิเคราะห์แล้วดึงความรู้ในส่วนที่สำคัญออกมาประยุกต์ใช้งานหรือคาดการณ์สิ่งต่าง ๆ ที่อาจจะเกิดขึ้น เพื่อใช้ประกอบการตัดสินใจ [8]

### 2.3 การวิเคราะห์การถดถอย (Regression)

2.3.1 เคเนียร์สเนเบอร์ (k-Nearest Neighbors : k-NN) เป็นเทคนิคของการทำเหมืองข้อมูล ที่สามารถใช้ได้ทั้งการจำแนกประเภทข้อมูล (Classification) และการถดถอย (Regression) k-NN เป็นอัลกอริทึมที่จัดเก็บข้อมูลที่มีอยู่ และจำแนกข้อมูล ตามการวัดความคล้ายคลึงกัน โดยกระบวนการทำนายเป้าหมายที่เป็นตัวเลข จะทำการคำนวณค่าเฉลี่ยของเป้าหมายที่เป็นตัวเลข โดยใช้ฟังก์ชันระยะทาง เช่น ค่าเฉลี่ยอย่างง่าย และค่าเฉลี่ยถ่วงน้ำหนักระยะทาง ผกผัน [9]

2.3.2 เกรเดียนบูททรี (Gradient Boosted Trees : GBT) มีพื้นฐานมาจากเทคนิคคิซึซึทรี (Decision Tree : DT) โดยจะใช้ผลจากคิซึซึทรีมากกว่า 1 แบบจำลอง มาประมวลร่วมกัน ช่วยให้ผลลัพธ์คาดการณ์ผ่านการประมวลค่าที่ค่อย ๆ ดีขึ้น เทคนิคการถดถอยแบบเกรเดียนบูททรี เป็นกระบวนการถดถอยแบบไม่เป็นเชิงเส้น (Non-Linear Regression) โดยการปรับปรุงความแม่นยำของคิซึซึทรี ภายในแบบจำลอง และสร้างแผนผังการตัดสินใจของแบบจำลองที่ประมวลผลรวมจากคิซึซึทรีทุกต้นออกมา การเพิ่มคิซึซึทรี สามารถเพิ่มความแม่นยำ แต่ก็ลดความเร็วในการสร้างแบบจำลองลง [10]

2.3.3 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) เป็นเทคนิคที่ประยุกต์ใช้ไฮเปอร์เพลน (Hyperplane) หรือชุดของไฮเปอร์เพลนเพื่อสร้างขอบเขตในแบ่งแยก หรือ จัดกลุ่มข้อมูล ซึ่งสามารถนำมาใช้สำหรับการจำแนกประเภทข้อมูล (Classification) การถดถอย (Regression) หรืองานด้านอื่น ๆ [11]

## 3. วิธีการดำเนินการวิจัย

ผู้วิจัยได้ดำเนินการตามกระบวนการมาตรฐานของการทำเหมืองข้อมูล (Cross-Industry Standard Process for Data Mining : CRISP-DM) ประกอบด้วย 6 ขั้นตอน [12],[13] ดังนี้

### 3.1 การทำความเข้าใจเกี่ยวกับธุรกิจ (Business Understanding)

การปลูกลำไยตามธรรมชาติจะมีปัญหาที่สำคัญ คือ การออกดอกและการติดผลไม่สม่ำเสมอ สาเหตุอาจเนื่องมาจากอิทธิพลของสภาพภูมิอากาศในแต่ละเดือนไม่เหมาะสม

หรือไม่เพียงพอที่จะทำให้ลำไยออกดอกและติดผลได้นอกจากนี้ยังมีสาเหตุอื่น ๆ เช่น ปริมาณธาตุอาหารและความอุดมสมบูรณ์ของดิน เป็นต้น อีกทั้งปัจจุบันทั่วโลกประสบภาวะการแพร่ระบาดของโรคโควิด-19 ซึ่งส่งผลกระทบต่อกระบวนการผลิตลำไย โดยปัญหาที่สำคัญคือการคาดการณ์ปริมาณผลผลิตและปริมาณความต้องการของผู้บริโภคให้มีความสมดุลกันในด้านของราคาที่เหมาะสมสามารถรับยอมได้ซึ่งต้องครอบคลุมต้นทุนต่าง ๆ และมีความเหมาะสมกับปัจจัยอื่น ๆ ในท้องตลาด เพื่อแก้ปัญหาดังกล่าวผู้วิจัยจึงนำเอาระบบสารสนเทศเข้ามาช่วยในการวิเคราะห์ข้อมูล โดยนำทฤษฎีการทำเหมืองข้อมูลมาเพื่อใช้ในการวางแผนกระบวนการการผลิตลำไยให้มีประสิทธิภาพมากยิ่งขึ้น

### 3.2 การทำความเข้าใจเกี่ยวกับข้อมูล (Data Understanding)

จากการศึกษาพบว่าปริมาณผลผลิตลำไยของภาคเหนือจะขึ้นอยู่กับประเภทดิน และสภาพอากาศในช่วงเวลาก่อนหน้า และช่วงเวลาปัจจุบัน ข้อมูลที่นำมาใช้ในงานวิจัยประกอบด้วย 3 ส่วนดังนี้

#### 3.2.1 ข้อมูลปริมาณผลผลิตลำไย

ข้อมูลปริมาณผลผลิตลำไยที่ใช้ในงานวิจัย คือ ข้อมูลร้อยละและปริมาณผลผลิตสินค้าไม้ผล 6 ชนิด รายเดือน รายจังหวัด ระหว่างปี พ.ศ. 2551 - 2562 ประกอบด้วยข้อมูลที่มี 17 ตัวแปร และมีจำนวนข้อมูลทั้งหมด 4,693 แถว ซึ่งเป็นข้อมูลที่รวบรวมโดยศูนย์กลางข้อมูลเปิดภาครัฐประเทศไทย จากเจ้าของข้อมูลและนำมาเผยแพร่ [14] งานวิจัยนี้ทำการดึงเฉพาะข้อมูลปริมาณผลผลิตลำไยรายเดือนของจังหวัดในภาคเหนือ ได้แก่ น่าน, ลำปาง, อุตรดิตถ์, พะเยา, เชียงราย, แม่ฮ่องสอน, ลำพูน, กำแพงเพชร, พิชณุโลก, ตาก และเชียงใหม่

#### 3.2.2 ข้อมูลประเภทดิน

ข้อมูลประเภทดิน คือ ข้อมูลชุดดินของตำบล อำเภอ และจังหวัดนั้น ๆ ซึ่งเป็นแหล่งข้อมูลจากกรมป้องกันและบรรเทาสาธารณภัย (ปภ.) ประกอบด้วยข้อมูลที่มี 15 ตัวแปร และมีจำนวนข้อมูลทั้งหมด 282,873 แถว โดยชุดข้อมูลจะมีการปรับปรุงนาน ๆ ครั้ง หรือจัดทำขึ้นเพียงครั้ง

เดียว ข้อมูลส่วนที่นำมาใช้ในการสร้างแบบจำลอง ได้แก่ พื้นที่จังหวัด (P\_area) และข้อมูลประเภทดินที่ใช้ (SERISE\_USE)

#### 3.2.3 ข้อมูลสภาพอากาศ

ข้อมูลสภาพอากาศ คือ ข้อมูลอากาศในแต่ละช่วงเวลา โดยได้เลือกใช้รายงานสภาพอากาศในภาคเหนือของประเทศไทย ระหว่างปี พ.ศ. 2559-2564 ประกอบด้วยข้อมูลที่มี 21 ตัวแปร และมีจำนวนข้อมูลทั้งหมด 23,875,000 แถว ซึ่งเป็นข้อมูลจากกรมอุตุนิยมวิทยา ที่มีการรายงานค่าทุก ๆ 1 นาที ข้อมูลส่วนนี้มีการปรับปรุงข้อมูลอย่างสม่ำเสมอและมีช่วงเวลาที่แน่นอน โดยจะปรับเปลี่ยนรูปแบบข้อมูลให้เป็นรายเดือน และเป็นข้อมูลอนุกรมเวลาที่ย้อนหลังข้อมูลไป 6 เดือน เพื่อนำข้อมูลมาใช้สร้างแบบจำลอง ข้อมูลส่วนที่ถูกปรับเปลี่ยนเป็นข้อมูลอนุกรมเวลารายเดือน ได้แก่ ปริมาณน้ำฝนสะสม 1 นาที (Rainfall1Min - mm), ความชื้นสัมพัทธ์ (RelativeHumidity - %), อุณหภูมิอากาศ (AirTemperature - °C), ความเร็วลมกระโชก (GustWindSpeed - km/h) และความเร็วลม (WindSpeed - km/h)

#### 3.3 การเตรียมข้อมูล (Data Preparation)

จากรายละเอียดของข้อมูลและการทดลองปรับเปลี่ยนรูปแบบข้อมูลหลาย ๆ แบบทำให้ได้รูปแบบปัจจัยสำหรับสร้างแบบจำลองในการพยากรณ์ปริมาณผลผลิตลำไยของภาคเหนือในแต่ละเดือน โดยแบ่งเป็นปัจจัยนำเข้า (Input Features) และปัจจัยที่เป็นคำตอบ (Label)

**ปัจจัยข้อมูลนำเข้า (Input Features)** คือ ปัจจัยที่บอกให้แบบจำลองทราบถึงรูปแบบข้อมูล ณ เดือนปัจจุบัน ซึ่งประกอบด้วยตัวแปรต่าง ๆ ดังนี้

year – ค่าปีของเดือนที่ต้องการพยากรณ์

month – ค่าเดือนของวันเวลาที่ต้องการพยากรณ์

P\_area – ค่าพื้นที่จังหวัดที่ต้องการพยากรณ์

SeriseUse\_mode – ค่าฐานนิยมของค่าประเภทดินที่ใช้ของแต่ละจังหวัด

Rainfall\_max, Rainfall\_sum, Rainfall\_avg – ค่าปริมาณน้ำฝนสะสมรายเดือนที่มากที่สุด, ค่าผลรวมปริมาณน้ำฝนสะสมรายเดือน และค่าเฉลี่ยปริมาณน้ำฝนสะสมรายเดือน ในช่วงเวลา 6 เดือน ย้อนหลังนับจากเดือนที่ต้องการพยากรณ์

RelativeHumidity\_max, RelativeHumidity\_avg – ค่าความชื้นสัมพัทธ์รายเดือนที่มากที่สุด และค่าเฉลี่ยความชื้นสัมพัทธ์รายเดือน ในช่วงเวลา 6 เดือน ย้อนหลังนับจากเดือนที่ต้องการพยากรณ์

AirTemperature\_median – ค่ามัธยฐานของอุณหภูมิรายเดือนในช่วงเวลา 6 เดือน ย้อนหลังนับจากเดือนที่ต้องการพยากรณ์

GustWindSpeed\_max – ค่าความเร็วลมกระโชกรายเดือนที่มากที่สุดในช่วงเวลา 6 เดือน ย้อนหลังนับจากเดือนที่ต้องการพยากรณ์

WindSpeed\_max – ค่าความเร็วลมรายเดือนที่มากที่สุดในช่วงเวลา 6 เดือน ย้อนหลังนับจากเดือนที่ต้องการพยากรณ์

**ปัจจัยคำตอบ (Label)** คือ ค่าปริมาณผลผลิตลำไยรายเดือน ซึ่งขึ้นอยู่กับรูปแบบของปัจจัยนำเข้า (Input Features) โดยในงานวิจัยนี้ใช้ชื่อตัวแปรสำหรับปัจจัยคำตอบว่า prediction(yield)

### 3.4 การสร้างแบบจำลอง (Modeling)

จากการเชื่อมโยงข้อมูลปริมาณผลผลิตลำไย ข้อมูลประเภทดิน และข้อมูลสภาพอากาศ โดยใช้ เดือน ปี และชื่อจังหวัดจะได้ชุดข้อมูลที่พร้อมจะนำไปใช้สร้างแบบจำลองทั้งหมด 350 ตัวอย่าง และนำข้อมูลมาวิเคราะห์ตามเทคนิคการทำเหมืองข้อมูล โดยประยุกต์ใช้การเรียนรู้ของเทคนิคการจำแนกประเภทข้อมูลด้วยวิธีเคเนียร์สเนเบอร์ (k-NN), เกรเดียนบูททรี (GBT) และ ซัพพอร์ตเวกเตอร์แมชชีน (SVM) เริ่มต้นจากการแบ่งข้อมูลออกเป็นชุดเรียนรู้ (Training Dataset) และข้อมูลชุดทดสอบ (Testing Dataset) ด้วยวิธี Cross Validation เพื่อใช้ปรับพารามิเตอร์ของแบบจำลองระหว่างขั้นตอนการเรียนรู้ ซึ่งจะแบ่งข้อมูลออกเป็น k ชุดย่อยที่มีจำนวนข้อมูลเท่ากัน ข้อมูลย่อย 1 ชุด จะถูกเก็บไว้เป็นชุดข้อมูลทดสอบ (Testing Dataset) และข้อมูลย่อย ย่อย k - 1 ชุดที่เหลือใช้เป็นชุดข้อมูลชุดเรียนรู้ (Training Dataset) โดยในงานวิจัยนี้กำหนดค่า k = 10 (10-Fold Cross-Validation)

### 3.5 การประเมินผล (Evaluation)

การทดสอบประสิทธิภาพของแบบจำลอง คือ ขั้นตอนประเมินการทำงานของแบบจำลองที่ผ่านการเรียนรู้จัดจํารูปแบบที่เกิดขึ้นของข้อมูลในอดีต ว่าสามารถใช้ประเมินค่า

ข้อมูลที่ไม่เคยเรียนรู้ได้ถูกต้องเพียงใด ซึ่งค่าที่เลือกใช้เป็นดัชนีในการประเมินสำหรับงานวิจัยนี้ ได้แก่

- ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Squared Error - RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum (\text{prediction-actual})^2}$$

- ค่าความคลาดเคลื่อนสัมบูรณ์ (Absolute Error - AE)

$$AE = |\text{prediction-actual}|$$

- ค่ามาตรฐานของความคลาดเคลื่อนสัมบูรณ์

(Normalized Absolute Error - NAE)

$$NAE = \frac{\sum |\text{prediction-actual}|}{\sum \text{actual}}$$

เมื่อ n คือ จำนวนข้อมูลทั้งหมด

prediction คือ ค่าที่ได้จากการทำนาย

actual คือ คำตอบที่ถูกต้อง

### 3.6 การนำแบบจำลองไปใช้งาน (Deployment)

หลังจากที่ได้แบบจำลองที่มีคุณภาพและความถูกต้องแม่นยำสำหรับการพยากรณ์ประมาณผลผลิตลำไยในภาคเหนือของประเทศไทย การจะนำแบบจำลองไปใช้งานจริง อาจต้องมีการปรับแต่งแบบจำลองเพื่อความเหมาะสมในสถานะจริง อย่างไรก็ตามหลังจากติดตั้งแล้วควรมีการปรับปรุงแบบจำลองเป็นระยะ ๆ (Periodic Update) เพื่อให้แบบจำลองมีความถูกต้อง และทันกับสถานการณ์ที่เกิดขึ้นจริง

## 4. ผลการดำเนินงาน

ผลการประเมินประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ปริมาณผลผลิตลำไยในภาคเหนือของประเทศไทยทั้ง 3 วิธี โดยใช้พารามิเตอร์ที่เป็นค่าตั้งต้น (Default Parameters) สำหรับโปรแกรม Rapidminer Studio [10] พบว่า เทคนิคเคเนียร์สเนเบอร์ (k-NN) เป็นแบบจำลองที่ให้ผลการทดสอบประสิทธิภาพสูงสุด ดังตารางที่ 1 และเพื่อปรับปรุงแบบจำลองให้เกิดประสิทธิภาพในการประมวลผลมากขึ้น จึงได้ทดลองปรับเปลี่ยนพารามิเตอร์ของแบบจำลองเพิ่มเติม โดยผลที่ดีที่สุดสำหรับแต่ละทั้ง 3 เทคนิค มีพารามิเตอร์ดังนี้

- **แบบจำลองเคเนียร์เซนเนอร์ (k-NN)** มีพารามิเตอร์เป็น  $k = 5$ , Measure type = NumericalMeasures, numerical measure = ManhattanDistance
  - **แบบจำลองเกรเดียนบูทรี (GBT)** มีพารามิเตอร์เป็น number of trees = 200, maximal depth = 25, min rows = 2.0, number of bins = 50, learning rate = 0.04 และ sample rate = 1.0
  - **แบบจำลองซัพพอร์ตเวกเตอร์แมชชีน (SVM)** มีพารามิเตอร์เป็น kernel cache = 200,  $c = 0.0$ , convergence epsilon = 0.001 max iterations = 100,000
- หลังจากการปรับพารามิเตอร์ของแบบจำลอง พบว่า เทคนิคเกรเดียนบูทรี (GBT) มีประสิทธิภาพที่ดีที่สุดสำหรับแบบจำลองที่ได้นำมาศึกษา ดังตารางที่ 2

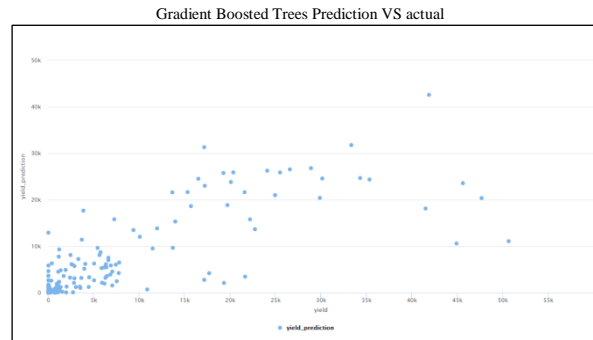
**ตารางที่ 1:** ผลการทดสอบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณผลผลิตลำไยในภาคเหนือของประเทศไทย

	ค่าทดสอบประสิทธิภาพ		
	root mean squared error (RMSE)	absolute error (AE)	normalized absolute error (nAE)
k-Nearest Neighbors	7767.676	3573.444	0.522
Gradient Boosted Trees	9388.638	5407.330	0.799
Support Vector Machine	12602.418	5057.226	0.692

**ตารางที่ 2:** ผลการพัฒนาแบบจำลองการพยากรณ์ปริมาณผลผลิตลำไยในภาคเหนือของประเทศไทย

	ค่าทดสอบประสิทธิภาพ		
	root mean squared error (RMSE)	absolute error (AE)	normalized absolute error (nAE)
k-Nearest Neighbors	6867.099	2871.707	0.409
Gradient Boosted Trees*	6925.252	2651.584	0.355
Support Vector Machine	12602.418	5057.226	0.692

\* คือวิธีที่มีความเหมาะสมสำหรับการพยากรณ์ปริมาณผลผลิตลำไยในภาคเหนือของประเทศไทย



**ภาพที่ 1:** ผลการประเมินประสิทธิภาพการพยากรณ์ค่าปริมาณผลผลิตลำไยของวิธี Gradient Boosted Trees (GBT)

### 5. อภิปรายและสรุปผล

จากการประเมินและพัฒนาประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ปริมาณผลผลิตลำไยในภาคเหนือของประเทศไทย โดยนำทฤษฎีการทำเหมืองข้อมูลมาประยุกต์ใช้การเรียนรู้ด้วยเทคนิคการถดถอย (Regression) 3 วิธี ได้แก่ เทคนิคเคเนียร์เซนเนอร์ (k-NN), เทคนิคเกรเดียนบูทรี (GBT) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) กับข้อมูลปริมาณผลผลิตลำไย ข้อมูลประเภทดิน และข้อมูลสภาพอากาศที่เป็นอนุกรมเวลา (Time Series) พบว่า แบบจำลองที่ดีที่สุดคือแบบจำลองเทคนิคเกรเดียนบูทรี (GBT) ซึ่งมีพารามิเตอร์ภายในแบบจำลองดังนี้ number of trees = 200, maximal depth = 25, min rows = 2.0, number of bins = 50, learning rate = 0.04 และ sample rate = 1.0 และมีประสิทธิภาพการทำงาน ของแบบจำลองที่ต่ำกว่าที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Squared Error - RMSE) เป็น 6925.252, ค่าความคลาดเคลื่อนสัมบูรณ์ (Absolute Error - AE) เป็น 2651.584, และ ค่ามาตรฐานของความคลาดเคลื่อนสัมบูรณ์ (Normalized Absolute Error - NAE) เป็น 0.355

แบบจำลองเกรเดียนบูทรี (GBT) คือวิธีที่มีความเหมาะสมสำหรับการพยากรณ์ปริมาณผลผลิตลำไยในภาคเหนือของประเทศไทย ตามลักษณะโครงสร้างของข้อมูลนำเข้า (Input Data) ที่นำมาใช้ในการวิจัยนี้เท่านั้น การนำแบบจำลองไปใช้งาน จึงจำเป็นต้องปรับเปลี่ยนรูปแบบข้อมูลนำเข้าให้มีลักษณะเดียวกันกับในงานวิจัย ผล



การคาดการณ์จึงจะมีความถูกต้อง เพื่อการวางแผนการผลิต และจัดทำนโยบายส่งเสริมกลุ่มผู้ผลิตที่เหมาะสม

## 6. ข้อเสนอแนะ

6.1 เนื่องจากจำนวนข้อมูลที่ใช้ในการวิจัย มีจำนวน ตัวอย่างค่อนข้างน้อย จึงส่งผลให้ประสิทธิภาพของ แบบจำลองมีค่าค่อนข้างต่ำ ดังนั้นหากมีการต่อยอดสำหรับ งานวิจัยควรมีการเพิ่มจำนวนชุดข้อมูลตัวอย่าง

6.2 แบบจำลองที่ใช้ในการพยากรณ์ปริมาณผลผลิต ลำไยควรมีการปรับปรุง โดยนำข้อมูลที่เกิดขึ้นใหม่มาเป็น ปัจจัยสำหรับให้แบบจำลองเรียนรู้อีกครั้ง เพื่อให้ทัน สถานการณ์ที่เปลี่ยนแปลงไป

### เอกสารอ้างอิง

- [1] BBC NEWS, “โควิด-19: ลำไยไทยในวิกฤตไวรัส เมื่อจีน ไม่ขยับ เกษตรกรไทยจะอยู่อย่างไร”, *สำนักวิจัยเศรษฐกิจ การเกษตร กระทรวงเกษตรและสหกรณ์ และส่วนเศรษฐกิจ ภาคเหนือ ธนาคารแห่งประเทศไทย*, 30 เม.ย. 2563.
- [2] กรมพัฒนาที่ดิน กระทรวงเกษตรและสหกรณ์, “ลำไย”, *แนวทางการส่งเสริมการเกษตรที่เหมาะสมตามฐานข้อมูลแผนที่ เกษตรเชิงรุก*, 2564, หน้า 22.
- [3] สถาบันถ่ายทอดเทคโนโลยีสู่ชุมชน มหาวิทยาลัยเทคโนโลยี ราชมงคลล้านนา, “การจัดการธาตุอาหารและให้น้ำลำไย”, *คู่มือการจัดการสวนลำไย การจัดการความรู้ และถ่ายทอด เทคโนโลยีการผลิตลำไยคุณภาพดีต้นทุนต่ำ*, โรงพิมพ์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา, 2564, หน้า 7.
- [4] บุญรักษ์ คู่ยศิริ และมงคล ทรัพย์ยิ่ง, “การศึกษาขั้นตอนอิทธิ ผลของอุณหภูมิความชื้นสัมพัทธ์ และปริมาณน้ำฝนต่อ ผลผลิตของลำไยในจังหวัดเชียงใหม่”, *วารสารวิชาการ กระทรวงเกษตรและสหกรณ์ (ประเทศไทย)*, ฉบับที่ 2, หน้า 106-113, พ.ศ.2527.
- [5] ชมพูนุท เกษมเศรษฐ์ และคณะ, “การประยุกต์ใช้ตัวแบบการ พยากรณ์โดยเทคนิคโครงข่ายประสาทเทียม: กรณีศึกษาการ พยากรณ์ผลผลิตลำไยนอกฤดู”, *วิศวกรรมสารฉบับวิจัยและ พัฒนา*, ปีที่ 23, ฉบับที่ 4, พ.ศ. 2555.
- [6] Waqas Ahmed Khan Muhammad and Fahad Irfan Siddiqui “Predicting Calorific Value of Thar Lignite Deposit: A Comparison between Back-propagation Neural Networks (BPNN), Gradient Boosting Trees (GBT), and Multiple Linear Regression (MLR)”, *Applied Artificial Intelligence An International Journal*, Vol.34, No.14, 2020.
- [7] ชรณินทร์ สัจจวิทย์ทรัพย์, “ตัวแบบซัพพอร์ตเวกเตอร์แมชชีน สำหรับการพยากรณ์ปริมาณการส่งออกยางคอมปาวด์ราย เดือนของประเทศไทย”, *วารสารวิทยาศาสตร์ มหาวิทยาลัย ศรีนครินทรวิโรฒ*, ปีที่ 33, ฉบับที่ 1, พ.ศ.2560.
- [8] สายชล สิ้นสมบูรณ์ทอง, *การทำเหมืองข้อมูลเล่ม 1 : การ ค้นหาความรู้จากข้อมูล*, พิมพ์ครั้งที่ 2, ปีที่พิมพ์ 2560.
- [9] Saed Saya, “K Nearest Neighbors - Regression”, *An Introduction to Data Science*, 2010-2022.
- [10] Rapidminer Studio, *Operator Reference Guide*, RapidMiner Documentation
- [11] Mariette Awad and Rahul, “Support Vector Regression”, *Efficient Learning Machines Theories, Concepts, and Applications for Engineers and System Designers*, pp. 67.
- [12] เอกสิทธิ์ พชรวงศ์ศักดิ์, “กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM), *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิ่ง เบื้องต้น*, บริษัท เอเชีย ดิจิตอลการพิมพ์ จำกัด, ปีที่พิมพ์ 2557, ฉบับที่ 2.
- [13] อนุพงศ์ สุขประเสริฐ, “กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM), *คู่มือการทำเหมืองข้อมูลด้วย โปรแกรม RapidMiner Studio*, สาขาคอมพิวเตอร์ธุรกิจ คณะการ บัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม, ปีที่พิมพ์ 2563, ฉบับที่ 3.
- [14] ศูนย์กลางข้อมูลเปิดภาครัฐ ประเทศไทย, *ข้อมูลร้อยละและ ปริมาณผลผลิตสินค้าไม้ผล 6 ชนิด รายเดือน รายจังหวัด*, 2551-2562, จาก <https://data.go.th/dataset/oa0015-2552>.

# ระบบธุรกิจอัจฉริยะสำหรับการให้ความยินยอมของลูกค้าตามวัตถุประสงค์ เพื่อเจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล

## Business Intelligence System for Customer Consent of Objectives for Data Protection Officer

ภัสสร สุขอมญ (Passorn Sukmon)<sup>1</sup> ศักดิ์ชัย ตั้งวรรณวิทย์ (Sakchai Tangwannawit)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>S6307021858376@email.kmutnb.ac.th, <sup>2</sup>sakchai.t@itd.kmutnb.ac.th

### บทคัดย่อ

ความสำคัญของพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคลธนาคารนับเป็นกลุ่มเป้าหมายที่จะเป็นตัวขับเคลื่อนธุรกิจและธุรกรรมทางการเงิน จึงต้องมีการเตรียมความพร้อมและฝึกทักษะในการปฏิบัติพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล ในเรื่องการขอความยินยอมจากเจ้าของข้อมูลส่วนบุคคล เพื่อให้การคุ้มครองข้อมูลส่วนบุคคลสอดคล้องกับกฎหมายและมีประสิทธิภาพให้สูงสุด ทั้งนี้ การขอความยินยอม การเก็บรวบรวมใช้เปิดเผยข้อมูลส่วนบุคคล ต้องได้รับการยินยอมจากเจ้าของข้อมูล โดยต้องมีการยอมรับเป็นลายลักษณ์อักษรและต้องไม่ใช่ข้อมูลนอกเหนือจากวัตถุประสงค์ที่แจ้งวัตถุประสงค์งานวิจัย

1) เพื่อวิเคราะห์การจัดกลุ่มลูกค้าที่ให้ความยินยอมตามวัตถุประสงค์ 2) เพื่อหาความสัมพันธ์ของแต่ละกลุ่มจากการให้ความยินยอมตามวัตถุประสงค์ของลูกค้า และ 3) เพื่อนำเสนอรายงานธุรกิจอัจฉริยะจากกฎความสัมพันธ์ทำการเก็บข้อมูลการให้ความยินยอมของลูกค้าธนาคารจากกลุ่มตัวอย่าง จำนวน 10,000 คน และวิเคราะห์ข้อมูลโดยใช้การหาความสัมพันธ์ (Association Rules Discovery) ด้วย โปรแกรม Rapid Miner Studio ผลการวิจัยพบว่า ได้จำนวน 46 กฎ

คำสำคัญ: การให้ความยินยอมของลูกค้าตามวัตถุประสงค์  
กฎความสัมพันธ์ ธนาคาร

### Abstract

The Personal Data Protection Act (PDPA) is important for the bank, to drives business and financial transactions. Therefore, it is necessary to prepare and practice skills in the implementation of the PDPA about obtaining the consent of the personal data to ensure the protection of personal data in accordance with the law and with maximum efficiency. Collection, use, and disclosure of personal information consent required from the owner of the data which must be accepted in writing and must not use the information other than the stated purpose.

The objectives of the research were 1) to analyze objective-consent of customer segmentation, 2) to determine the association rules for each group based on the objective consent of the customers, and 3) to present a business intelligence report based on the consent rules.

This research to collecting data on the consent of customers of the bank from a sample of 10,000 customers, the data was analyzed using association rules by rapid miner studio. The results of this research showed that 46 rules from the analysis of the association rule technique.

**Keywords:** Personal Data Protection Act, Association rule, Bank.

## 1. บทนำ

เนื่องจากในสถานการณ์ปัจจุบันมีการเปลี่ยนแปลงเข้าสู่ยุคดิจิทัลและความก้าวหน้าทางเทคโนโลยี ซึ่งส่งผลให้มีการล่วงละเมิดสิทธิความเป็นส่วนตัวและข้อมูลส่วนบุคคลเพิ่มมากขึ้น และสามารถทำได้โดยง่ายและรวดเร็ว จึงสร้างความเดือดร้อน ราคายู หรือความเสียหายแก่เจ้าของข้อมูลส่วนบุคคล [1] รวมถึงการนำข้อมูลส่วนบุคคลไปใช้โดยไม่ได้รับความยินยอมจากเจ้าของข้อมูลส่วนบุคคล การคุ้มครองความเป็นส่วนตัวและข้อมูลส่วนบุคคล ซึ่งจะต้องมีการเก็บรวบรวม ใช้ และเปิดเผย ข้อมูลส่วนบุคคล ให้เป็นไปตามวัตถุประสงค์ของเจ้าของข้อมูลส่วนบุคคล โดยขอบเขตการบังคับใช้ยังครอบคลุมถึงกรณีผู้ควบคุมข้อมูล และผู้ประมวลผลข้อมูลส่วนบุคคล อีกทั้งวัตถุประสงค์ในการใช้ข้อมูล และต้องไม่ใช่ข้อมูลนอกเหนือจากวัตถุประสงค์ที่แจ้ง ส่วนของบทลงโทษมีทั้งโทษทางแพ่ง ทางอาญา และทางปกครอง หากมีการฝ่าฝืนมีโทษจำคุกหรือปรับ หรือทั้งจำทั้งปรับ ฉะนั้น โดยเฉพาะอย่างยิ่งกับธนาคารและหน่วยงานของรัฐต้องพัฒนาระบบการจัดเก็บข้อมูลส่วนบุคคลให้มีประสิทธิภาพมากขึ้น [2] อีกทั้งธนาคารหลายแห่งพัฒนาและปรับปรุงกระบวนการแก้ไขปัญหาดังกล่าว เช่น ระบบส่งเสริมการขายจากลูกค้า ความสัมพันธ์ของสินค้า เช่นเดียวกับผู้วิจัยในอดีตที่มีการวิจัยด้านการหาความสัมพันธ์จากฐานข้อมูลการซื้อผลิตภัณฑ์ค่อนข้างมาก แต่ส่วนใหญ่การสร้างแบบจำลองความสัมพันธ์สำหรับฐานข้อมูลการสั่งซื้อสินค้า และมักให้ข้อเสนอแนะว่าสามารถนำไปพัฒนาเป็นในการออกรายงาน การจับคู่สินค้าที่มีความสัมพันธ์ที่มีค่าความเชื่อมั่นสูง ตลอดจนสามารถนำวิจัยนี้ไปประยุกต์ใช้ในระบบคลังข้อมูล และระบบสนับสนุนการตัดสินใจ รวมถึงการนำเสนอสินค้าได้ตามความต้องการของลูกค้าในแต่ละราย เป็นการสร้างความพึงพอใจให้กับลูกค้า ซึ่งเป็นกลยุทธ์หนึ่งของการวางแผนการตลาดได้อย่างเหมาะสม และมีประสิทธิภาพ เช่น การสร้างแบบจำลองความสัมพันธ์สำหรับฐานข้อมูล การสั่งซื้อสินค้าโดยใช้เทคนิค เอฟพีโกรธ [4] เป็นต้น

ผู้วิจัยจึงมีแนวคิดในการศึกษาทฤษฎีการสร้างแบบจำลองกฎความสัมพันธ์จากฐานข้อมูลลูกค้า เพื่อช่วยให้เพื่อเจ้าหน้าที่คุ้มครองข้อมูลส่วนบุคคล รวมถึงผู้บริหารและพนักงานสามารถนำไปใช้ในการสนับสนุนการตัดสินใจ การนำเสนอผลิตภัณฑ์ได้ตามความต้องการของลูกค้าและตามวัตถุประสงค์ของลูกค้าในแต่ละราย อีกทั้งยังเป็นการสร้างความพึงพอใจให้กับลูกค้า ซึ่งเป็นกลยุทธ์หนึ่งของการวางแผนการตลาดได้อย่างเหมาะสม และมีประสิทธิภาพ รวมถึงการนำข้อมูลการให้ความยินยอมไปใช้ตามวัตถุประสงค์

## 2. แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

### 2.1 พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562

ประกาศในราชกิจจานุเบกษา เมื่อวันที่ 27 พฤษภาคม พ.ศ. 2562 เป็นมาตรการในการกำกับดูแลการคุ้มครองข้อมูลส่วนบุคคล มีเครื่องมือในการกำกับการดำเนินงานด้านการควบคุมข้อมูลส่วนบุคคล รวมถึงส่งเสริมภาพลักษณ์ที่ดีของประเทศ สร้างสังคมที่เข้มแข็ง เนื่องจากสามารถตรวจสอบการดำเนินงานภาครัฐและภาคธุรกิจเกี่ยวกับการคุ้มครองข้อมูลส่วนบุคคลให้มีความถูกต้องเหมาะสม

### 2.2 แหล่งที่มาของข้อมูล (Source of Data)

กระบวนการข้อมูลสถิติอาจจำแนกตามแหล่งที่มาได้ 2 ทาง คือ 1) ข้อมูลปฐมภูมิ (Primary Data) เป็นข้อมูลที่ใช้หรือหน่วยงานที่ใช้เป็นผู้ทำการเก็บข้อมูลด้วยตนเอง 2) ข้อมูลทุติยภูมิ (Secondary Data) เป็นข้อมูลที่ใช้ไม่ได้เก็บรวบรวมเอง แต่มีผู้อื่น หรือ หน่วยงานอื่น ๆ ทำการเก็บรวบรวมไว้แล้ว เช่น จากระบบงาน ที่พิมพ์แล้ว

### 2.3 เอฟพี-โกรธ (FP-Growth Algorithms)

เป็นขั้นตอนและวิธีที่มีลักษณะการค้นหาหาข้อมูลที่น่าปรากฏบ่อยแบบการเติบโตอย่างเป็นรูปแบบ (Pattern Growth) โดยการทำงานของขั้นตอนวิธี FP-Growth สามารถอธิบายหลักการการทำงานได้ดังนี้

- อ่านข้อมูลจากฐานข้อมูลครั้งแรกเพื่อนับค่าความถี่ของแต่ละชั้น แล้วนำชั้นข้อมูลที่ไม่บ่อยกว่าค่าสนับสนุน

ขั้นต่ำ มาเรียงลำดับความถี่ของแต่ละชั้นข้อมูลจากมากไปหาน้อยแล้ว

นำมาสร้าง Header Table

- อ่านข้อมูลครั้งที่สองเพื่อสร้างต้นไม้ FP-Tree
- สร้าง Conditional Pattern Base และ สร้าง Conditional FP-Tree ของแต่ละชั้นข้อมูล เพื่อใช้ในขั้นตอนการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย โดยการพิจารณาจะเริ่มจากชั้นข้อมูลต่ำสุดจนถึงชั้นข้อมูลที่อยู่บนสุดในตาราง Header

- ค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยจากการสร้าง Conditional Pattern Base และ สร้าง Conditional FP-Tree ของแต่ละชั้นข้อมูล โดยใช้หลักการทำงานแบบแบ่งแยกแล้วเอาชนะ (Divide and Conquer)

## 2.4 การหาความสัมพันธ์ (Association Rules Discovery)

การค้นหากฎความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปหารูปแบบที่เกิดขึ้นบ่อย ๆ (Frequent Pattern) และใช้ในการวิเคราะห์ความสัมพันธ์หรือทำนายปรากฏการณ์ต่าง ๆ ฐานข้อมูลที่ใช้ในการทำเหมืองความสัมพันธ์ (Association Mining) ส่วนมากจะเป็นฐานข้อมูลประเภท Transaction Database ผลลัพธ์ที่ได้เป็นกฎความสัมพันธ์ (Association Rule) สามารถเขียนได้ในรูปเซตของรายการที่เป็นเหตุ ไปสู่เซตของรายการที่เป็นผล ซึ่งมีรากฐานมาจากการวิเคราะห์ตะกร้าตลาด (Market Basket Analysis) เช่น ลูกค้ายี่ห้อผ้าอ้อมส่วนใหญ่จะซื้อเบียร์ด้วย ข้อมูลที่นำมาใช้จะอยู่ในรูปแบบ Nominal หรือ Ordinal เท่านั้น

## 2.5 แผนภาพข้อมูล (Data Visualization)

การสรุปและแสดงข้อมูลออกมาให้อยู่ในรูปแบบของแผนภาพ กราฟ หรือวิดีโอที่อธิบายเกี่ยวกับข้อมูลที่ต้องการจะแสดง เนื่องจากการทำงานในปัจจุบัน ข้อมูลที่มีนั้นไม่ได้เป็นเพียงข้อความอย่างเดียว แต่มีข้อมูลที่หลากหลายมากขึ้น ดังนั้น จึงต้องมีการรองรับข้อมูลประมวลผล วิเคราะห์ข้อมูลออกมาเพื่อให้เห็นเป็นภาพที่ชัดเจนและเข้าใจง่ายกว่าเดิม การใช้ Data Visualization จะทำให้รู้ว่าข้อมูลในที่ที่น่าสนใจหรือทำให้เจอแนวทางใหม่ ๆ ในการทำการตลาดได้ด้วย

## 2.6 งานวิจัยที่เกี่ยวข้อง

เมธพร คงทอง (2556) ได้ทำการศึกษาวิจัยเรื่องการจัดกลุ่มลูกค้าค้างชำระของสินเชื่อบริษัทของธนาคารพาณิชย์แห่งหนึ่ง เพื่อจัดกลุ่มลูกค้าค้างชำระของสินเชื่อบริษัทและหาลักษณะปัจจัยสำคัญประจำกลุ่ม วิธีการศึกษาเป็นการวิจัยเชิงทดลอง โดยใช้ข้อมูลลูกค้าที่ค้างชำระจำนวน 744 ราย ทำการวิเคราะห์การจัดกลุ่ม (Cluster Analysis) แบบวิธี K-Means โดยใช้โปรแกรมสำเร็จรูปทางสถิติ ผลการศึกษาพบว่าสามารถจัดกลุ่มลูกค้าค้างชำระได้เป็น 3 กลุ่ม ปัจจัยสำคัญประจำกลุ่มสามารถนำมาใช้หลากหลาย เช่น เป็นข้อมูลสนับสนุนในการกำหนดแนวทางและนโยบายเพื่อการจัดกลุ่มลูกค้าค้างชำระของสินเชื่อบริษัทอนุมัติสินเชื่อบริษัท และลดการเกิดหนี้สูญ ซึ่งจากงานวิจัยนี้สามารถนำมาพัฒนาและสนับสนุนในการกำหนดแนวทางและนโยบายเพื่อการจัดกลุ่มลูกค้าและนำเสนอเงื่อนไข/ผลิตภัณฑ์ให้เหมาะสมกับลูกค้าแต่ละกลุ่ม [5]

กาญจนา หุทธิธรรมพงศ์, ปิยะมาศ จิตตระ 2 (2561). ได้ทำการศึกษาวิจัยเรื่อง การวิเคราะห์แบบ RFM ในการแบ่งกลุ่มผู้ใช้งานเครื่องพิมพ์และเครื่องถ่ายเอกสารในองค์กร กรณีศึกษาสำนักวิชาสารสนเทศศาสตร์ มหาวิทยาลัยวลัยลักษณ์ งานวิจัยนี้มีวัตถุประสงค์เพื่อทำการค้นหาคำแนะนำจากการวิเคราะห์ข้อมูลเพื่อจัดกลุ่มผู้ใช้งานเครื่องพิมพ์และเครื่องถ่ายเอกสารในองค์กร โดยการทำเหมืองข้อมูลด้วยวิธีการจัดกลุ่ม ข้อมูลที่ใช้ในการวิเคราะห์ประกอบด้วย 3 ส่วน คือ การพิมพ์งานด้วยเครื่องพิมพ์ การพิมพ์งานด้วยเครื่องถ่ายเอกสาร และการถ่ายเอกสาร งานวิจัยนี้นำเสนอวิธีการแบ่งกลุ่มผู้ใช้งานพฤติกรรมการใช้งานเครื่องพิมพ์และเครื่องถ่ายเอกสารด้วยแบบจำลอง Recency, Frequency and Monetary หรือ RFM ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลการใช้งานเครื่องพิมพ์และเครื่องถ่ายเอกสารของบุคลากรสำนักวิชาสารสนเทศศาสตร์ มหาวิทยาลัยวลัยลักษณ์ จำนวน 84,764 รายการ ตั้งแต่ปี 2554 - 2559 ผู้ใช้งานทั้งหมด 78 คน วิเคราะห์ด้วย RapidMiner Studio 7.4 จากผลการวิจัยทำให้ทราบรูปแบบการใช้งานและนำไป

ใช้ในการตรวจสอบพฤติกรรมการใช้งานเครื่องพิมพ์และเครื่องถ่ายเอกสารของบุคลากรว่ามีการใช้งานที่มีความสอดคล้องกับงานที่ทำอยู่หรือไม่เพื่อช่วยลดค่าใช้จ่าย เมื่อนำมาตรวจสอบไม่พบความคิดปกติในการใช้งานเนื่องจากกลุ่มผู้ใช้ในกลุ่มดังกล่าวเป็นเจ้าหน้าที่ที่รับผิดชอบทำงานด้านเอกสาร นอกจากนี้ยังพบว่าพฤติกรรมของบุคลากรในการพิมพ์เอกสารส่วนใหญ่นิยมพิมพ์ทางเครื่องพิมพ์มากกว่าการพิมพ์ทางเครื่องถ่ายเอกสาร ซึ่งงานวิจัยดังกล่าวสามารถนำมาพัฒนาและนำไปประยุกต์ใช้ประโยชน์กับการวิเคราะห์ข้อมูลองค์กรอื่นๆ ในลักษณะงานที่คล้ายคลึงกันได้ [10]

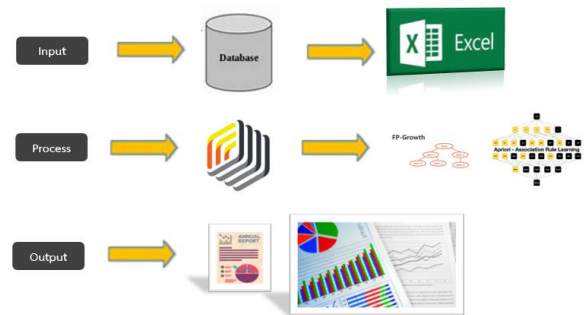
ณัฐญาพร ชื่นมัจฉา (2559) ได้ทำการศึกษาวิจัยเรื่อง การสร้างแบบจำลองกฎความสัมพันธ์สำหรับฐานข้อมูลการสั่งซื้อสินค้า โดยใช้เทคนิค เอฟพี-โกรธ เพื่อหาความสัมพันธ์การซื้อสินค้าของซูเปอร์มาเก็ตขนาดใหญ่ โดยการวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองความสัมพันธ์ของการซื้อสินค้าโดยใช้กฎความสัมพันธ์ (Association Rules) ด้วยเทคนิค เอฟพี โกรธ (FP-Growth) โดยวิเคราะห์จากการซื้อสินค้าของลูกค้าในแต่ละรายบุคคล ซึ่งผลลัพธ์ที่ได้จากงานวิจัยนี้ สามารถพยากรณ์ความต้องการของลูกค้าออกมาเป็นกฎความสัมพันธ์ได้ 4 รูปแบบ สามารถพยากรณ์ความต้องการของลูกค้าออกมาเป็นกฎความสัมพันธ์ซึ่งสามารถนำไปใช้ในการสนับสนุนการตัดสินใจการนำเสนอสินค้าได้ตามความต้องการของลูกค้าในแต่ละราย เป็นการสร้างความพึงพอใจให้กับลูกค้า ซึ่งเป็นกลยุทธ์หนึ่งของการวางแผนการตลาดได้อย่างเหมาะสม และมีประสิทธิภาพ ข้อเสนอแนะของงานวิจัยนี้คือการหาความสัมพันธ์ของการซื้อสินค้าของลูกค้าควรค้นหาความสัมพันธ์จากฐานข้อมูลขนาดใหญ่เพื่อที่จะได้กฎความสัมพันธ์ที่มีความน่าเชื่อถือที่สูง [4]

### 3. วิธีดำเนินการวิจัย

งานวิจัยฉบับนี้เป็นงานวิจัยเชิงปฏิบัติ โดยมีวัตถุประสงค์เพื่อ 1) วิเคราะห์การจัดกลุ่มลูกค้าที่ให้ความยินยอมตามวัตถุประสงค์ 2) หากฎความสัมพันธ์ของแต่ละ

กลุ่มจากการให้ความยินยอมตามวัตถุประสงค์ของลูกค้า และ 3) นำเสนอรายงานธุรกิจอัจฉริยะจากความสัมพันธ์

#### 3.1 กรอบแนวคิดงานวิจัย



ภาพที่ 1: กรอบแนวคิดงานวิจัย

จากภาพที่ 1 แสดงกรอบแนวคิดการสร้างตัวแบบจำลองการหาความสัมพันธ์ข้อมูลการให้ความยินยอมตามวัตถุประสงค์ของลูกค้า และการแสดงผลรายงานการหาความสัมพันธ์

#### 3.2 การรวบรวมและวิเคราะห์ข้อมูลเบื้องต้น

ในการวิจัยครั้งนี้ใช้ข้อมูลกลุ่มตัวอย่างในการวิจัยครั้งนี้เป็นข้อมูลกลุ่มตัวอย่างการให้ความยินยอมของลูกค้าธนาคารออมสิน โดยการรวบรวมข้อมูลจากแหล่งที่มาของข้อมูล (Source of Data) เป็นข้อมูลทุติยภูมิ (Secondary Data) จำนวน 10,000 รายการ

#### 3.3 การสร้างแบบจำลอง และการแสดงผล

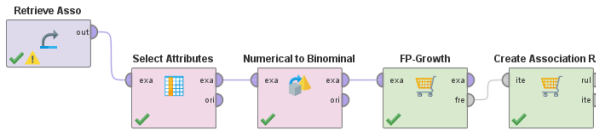
นำข้อมูลการให้ความยินยอมของลูกค้าธนาคารออมสินจำนวน 10,000 รายการ นำมาสร้างแบบจำลองเพื่อการหาความสัมพันธ์การให้ความยินยอมของลูกค้าธนาคารออมสิน โดยแบ่งเป็น 2 กิจกรรม ประกอบด้วย 1) การวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองความสัมพันธ์ของการให้ความยินยอมโดยใช้กฎความสัมพันธ์ (Association Rules) ด้วยเทคนิค เอฟพี โกรธ (FP-Growth) 2) จัดทำรายงานธุรกิจอัจฉริยะจากกฎความสัมพันธ์ ผู้จัดทำได้เลือกใช้เครื่องมือ Rapid Miner Studio ในการวิเคราะห์ และ Power BI ในการจัดทำรายงานตามรูปแบบ Data Visualization

#### 3.3 ระยะเวลาดำเนินการวิจัย

การดำเนินงานวิจัยในครั้งนี้ ใช้ระยะเวลา 6 เดือน

## 4. ผลการดำเนินงาน

### 4.1 ผลการวิเคราะห์ข้อมูล



ภาพที่ 2: ผลการวิเคราะห์ข้อมูลการให้ความยินยอมของลูกค้า

จากการวิเคราะห์ข้อมูล พบว่า สามารถพยากรณ์การให้ความยินยอมของลูกค้าตามวัตถุประสงค์ของลูกค้าเป็น กฎความสัมพันธ์ ได้ 4 กฎ ประกอบด้วย

1. กำหนดค่าพารามิเตอร์ Support ไว้ที่ 0.1 และ ค่าพารามิเตอร์ Confidence ไว้ที่ 0.8 ของรายการ การให้ความยินยอมลูกค้าทั้งหมด 11 กฎ
2. กำหนดค่าพารามิเตอร์ Support ไว้ที่ 0.1 และ ค่าพารามิเตอร์ Confidence ไว้ที่ 0.9 ของรายการ การให้ความยินยอมลูกค้าทั้งหมด 31 กฎ
3. กำหนดค่าพารามิเตอร์ Support ไว้ที่ 0.2 และ ค่าพารามิเตอร์ Confidence ไว้ที่ 0.8 ของรายการ การให้ความยินยอมลูกค้าทั้งหมด 9 กฎ
4. กำหนดค่าพารามิเตอร์ Support ไว้ที่ 0.2 และ ค่าพารามิเตอร์ Confidence ไว้ที่ 0.9 ของรายการ การให้ความยินยอมลูกค้าทั้งหมด 28 กฎ

ซึ่งสามารถนำไปใช้ในการสนับสนุนการตัดสินใจ การนำเสนอผลิตภัณฑ์ได้ตามความต้องการของลูกค้าและตามวัตถุประสงค์ของลูกค้าในแต่ละราย อีกทั้งยังเป็นการสร้างความพึงพอใจให้กับลูกค้า ซึ่งเป็นกลยุทธ์หนึ่งของการวางแผนการตลาดได้อย่างเหมาะสม และมีประสิทธิภาพ รวมถึงการนำข้อมูลการให้ความยินยอมไปใช้ตามวัตถุประสงค์ ดังแสดงในตารางที่ 1

ตารางที่ 1: สรุปผลการทดลองงานวิจัย

การทดลองที่	ค่าพารามิเตอร์ Min Support	ค่าพารามิเตอร์ Min Confidence	กฎความสัมพันธ์
1	10%	70%	-
2		80%	11
3		90%	31
4	20%	70%	-
5		80%	9
6		90%	28

จากตารางที่ 1 พบว่าการสร้างแบบจำลองเพื่อการหาความสัมพันธ์การให้ความยินยอมของลูกค้าของธนาคาร

ออมสิน โดยการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ของการให้ความยินยอมโดยใช้กฎความสัมพันธ์ด้วยเทคนิคเอพี-ไทร พบกฎความสัมพันธ์ เช่น

#### การทดลองที่ 1

แสดงกฎความสัมพันธ์จากค่าพารามิเตอร์ Min Support ไว้ที่ 0.1 หรือ 10% และ โอเปอเรเตอร์ Create Association Rules

เพื่อสร้างกฎความสัมพันธ์ทำการกำหนดค่าพารามิเตอร์ Min Confidence ไว้ที่ 0.8 หรือ 80% ของรายการ การให้ความยินยอมทั้งหมด 11 กฎ สามารถอธิบายความหมายได้ว่า Rule.1 การให้ความยินยอมข้อที่ 1 มีความสัมพันธ์กับสินค้าการให้ความยินยอมข้อที่ 3 และ 2 ด้วยความเชื่อมั่น 86% วัตถุประสงค์ของการให้ความยินยอม

ข้อที่ 1 เพื่อประโยชน์ในการพัฒนาและปรับปรุงการให้บริการของธนาคาร

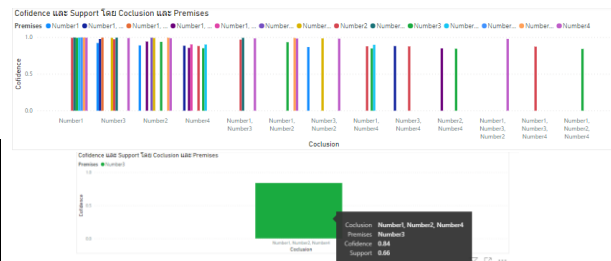
ข้อที่ 2 เพื่อวัตถุประสงค์ทางการตลาด การส่งเสริมการขาย การประชาสัมพันธ์ผลิตภัณฑ์

ข้อที่ 3 เพื่อการสถิติ ศึกษาวิจัย วิเคราะห์ และประเมินผลข้อมูล

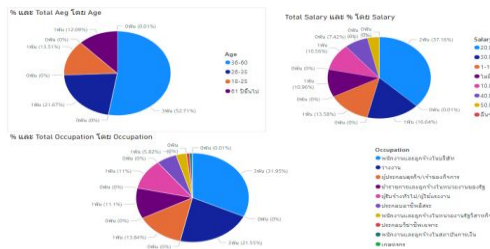
ข้อที่ 4 เพื่อการนำเสนอและขายผลิตภัณฑ์ประกันภัยของพันธมิตร

#### 4.2 การพัฒนาระบบธุรกิจอัจฉริยะเพื่อการแสดงผล

เมื่อ ได้กฎความสัมพันธ์เรียบร้อยแล้วผู้วิจัยทำนำผลวิเคราะห์ดังกล่าวมาจัดทำระบบธุรกิจอัจฉริยะ ด้วยโปรแกรม Power BI ในการจัดทำรายงานตามรูปแบบ Data Visualization โดยมีตัวอย่างระบบ ดังภาพที่ 3 และภาพที่ 4



ภาพที่ 3: รายงานจากการหาความสัมพันธ์



ภาพที่ 4: ความยินยอมข้อ 4 เปรียบเทียบกับช่วงอายุ รายได้ และอาชีพ

## 5. อภิปรายผลและข้อเสนอแนะ

### 5.1 อภิปรายผล

จากการดำเนินงานวิจัยโดยใช้การให้ความยินยอมของลูกค้านาคารออมสินจำนวน 10,000 รายการ สร้างตัวแบบจำลองเพื่อการหาความสัมพันธ์การให้ความยินยอมของลูกค้านาคารออมสินความสัมพันธ์การให้ความยินยอมของลูกค้านาคารออมสิน โดยใช้กฎความสัมพันธ์ (Association Rules) ด้วยเทคนิค เอฟพี กโรธ (FP-Growth 2) เพื่อสนับสนุนการตัดสินใจ การนำเสนอผลิตภัณฑ์ได้ตามความต้องการของลูกค้าและตามวัตถุประสงค์ของลูกค้าในแต่ละราย รวมถึงการนำข้อมูลการให้ความยินยอมไปใช้ตามวัตถุประสงค์ต่อไป

### 5.2 ข้อเสนอแนะ

งานวิจัยนี้ได้การให้ความยินยอมตามวัตถุประสงค์ที่ธนาคารกำหนด ซึ่งจากการวิเคราะห์ข้อมูล ลูกค้ายังขาดความเข้าใจในเรื่องดังกล่าว ดังนั้นงานวิจัยในอนาคตจึงควรศึกษาเพิ่มเติมให้ครอบคลุมถึงระบบงานเพื่อสนับสนุนการปฏิบัติงานของพนักงานให้ทราบถึงความตระหนักรู้ในการสื่อสารแก่ลูกค้าอย่างครบถ้วน

### เอกสารอ้างอิง

[1] ณัฐพร วิริยะดีพะทะ และธเนศ สุจริกุล, “ปัญหาทางกฎหมายเกี่ยวกับพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคลพ.ศ.2562 ศึกษากรณีหน้าที่ของผู้ควบคุมข้อมูลส่วนบุคคลตามมาตรา 39” *นิติศาสตร์มหาบัณฑิต มหาวิทยาลัยรังสิต.*

[2] ทัชชกร มหาแดง, “มาตรการในการคุ้มครองข้อมูลชีวมาตรตามพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562” *นิติศาสตร์มหาบัณฑิต มหาวิทยาลัยศรีปทุม.*

[3] คณาธิป ทองรวีวงศ์, “ผลกระทบทางลบอันเกิดจากกฎหมายคุ้มครองข้อมูลส่วนบุคคลของสหภาพยุโรปและพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562” *วารสารรัฐศาสตร์ภาคย์ ปีที่ 15 ฉบับที่ 38 : 2564. มหาวิทยาลัยเกษมบัณฑิต.*

[4] ฉัญญาพร ชื่นมัจฉา, “การสร้างแบบจำลองความสัมพันธ์สำหรับฐานข้อมูลการสั่งซื้อสินค้า โดยใช้เทคนิค เอฟพี – กโรธ” *วิทยาศาสตร์มหาบัณฑิต. มหาวิทยาลัยศรีปทุม 2559.*

[5] เมธิชา ยুবลชิต, “มาตรการคุ้มครองความเป็นส่วนตัวในข้อมูลไบโอเมตริกซ์” *นิติศาสตร์มหาบัณฑิต มหาวิทยาลัยศรีปทุม.*

[6] หทัยชนก หรัยวงศ์, “ปัญหาเกี่ยวกับการคุ้มครองข้อมูลส่วนบุคคลด้านสุขภาพ” *วารสารบัณฑิตศึกษาศาสตร์ นิติศาสตร์มหาบัณฑิต มหาวิทยาลัยธรรมศาสตร์.*

[7] กัณฑ์ร พองสินธุ์ สิทธิพันธ์ ทองศิริ และสุรพงษ์ ธรรมนันท์, “การพัฒนากระบวนการสนับสนุนการตัดสินใจสำหรับการเลือกวิธีการพยากรณ์ที่เหมาะสม”, *มหาวิทยาลัยราชภัฏอุดรดิตถ์ พ.ศ. 2562.*

[8] เมธพร คงทอง, “การจัดกลุ่มลูกค้าค้างชำระของสินเชื่อที่อยู่อาศัยของธนาคารพาณิชย์แห่งหนึ่ง”, *การค้นคว้าอิสระ. บริหารธุรกิจมหาบัณฑิต. มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี พ.ศ. 2556.*

[9] กิตติชัย ศรีธรรมพงษ์ มัลลิกา วัฒนะ วรรัตน์ สงฆ์เป็น ปวีณา วันชัย และอภิศักดิ์ พัฒนจักร, “ระบบช่วยการตัดสินใจสั่งซื้อโดยใช้เทคนิค ธุรกิจอัจฉริยะ กรณีศึกษา บริษัท มีสเตอร์ ซูชิ จำกัด”, *วารสารวิชาการและเทคโนโลยีสารสนเทศ (JIST) ปีที่ 8 ฉบับที่ 2 (2018) : [ก.ค. 2561 - ธ.ค. 2561]. มหาวิทยาลัยขอนแก่น.*

[10] กาญจนา หุทธิพงษ์, ปิยะมาศ จิตตระ, “การวิเคราะห์แบบ RFM ในการแบ่งกลุ่มผู้ใช้งานเครื่องพิมพ์และเครื่องถ่ายเอกสารในองค์กร”, *กรณีศึกษาสำนักวิชาสารสนเทศศาสตร์ มหาวิทยาลัยวลัยลักษณ์. มหาวิทยาลัยวลัยลักษณ์ ศูนย์เทคโนโลยีดิจิทัล. วารสารวิชาการการจัดการเทคโนโลยีสารสนเทศและนวัตกรรม ปีที่ 5 ฉบับที่ 1 [มกราคม – มิถุนายน 2561]. มหาวิทยาลัยราชภัฏมหาสารคาม.*

[11] Yun Chen, Guozheng Zhang, Dengfeng Hu, Shanshan Wang, “Customer Segmentation in Customer Relationship Management Based on Data Mining. School of Public Economy Administration” *Shanghai University, (2004).*

[12] Rudiger Wirth, Jochen Hipp, “CRISP-DM: Towards a Standard Process Model for Data Mining” *University of Tubingen.*

- [13] Kittkorn Dowpiset, Chompu Nuangjamnon, “An Investigation of Factors Affecting Intention to Comply Thailand PDPA with E-Services in Private” *University towards Social Media. International Journal of Economics and Business Administration Volume IX, Issue 2, 2021.*
- [14] Aurimas Sidlauskas, “The Role and Significance of the Data Protection Officer in the Organization” Faculty of Economics and Business, *PhD Student. Mykolo Romerio Universiteto, El. P, 2021.*
- [15] Mark Sheehan. (2011), “Can Broad Consent be Informed Consent?. Research Fellow, Oxford NIHR Biomedical Research Centre, the Ethox Centre and the Institute for Science and Ethics” *University of Oxford.*
- [16] Raghav Bhaskar, Srivatsan Laxman, Adam Smith, “Discovering Frequent Patterns in Sensitive Data. Microsoft Research Bangalore, India. Pennsylvania State” *University University Park, PA, USA.*
- [17] Pandaree Soonthonwarapas. (2021). “RFM Analysis to Segment Users of Circulation at Khunyung Long Athakravisunthorn Learning Resource Center, Prince of Songkla” *University. Journal of Library and Information Science Srinakharinwirot University Vol. 13 No. 1 January – June 2020.*
- [18] Mohammed J. Zaki, “Srinivasan Parthasarathy, Mitsunori Ogihara & Wei Li. (1997). Arallel Algorithms for Discovery of Association Rules. Department of Computer Science” *University of Rochester, Rochester, NY 14627.*
- [19] Christian Borgelt, “An Implementation of the FP-growth Algorithm. Department of Knowledge Processing and Language Engineering School of Computer Science, Otto-von-Guericke-University of Magdeburg” *Universitätsplatz 2, 39106 Magdeburg, Germany.*
- [20] David Neumark. (2018), “The effects of feedback on task group behavior: A review of the experimental” *research. Journal of Economic Literature 2018, 56(3), 799 – 866.*



# ระบบสนับสนุนการตัดสินใจเพื่อการเปิดหรือปิดสาขาโดยใช้การวิเคราะห์เครือข่าย

## Decision Support System for Opening or Closing Branches using Network Analysis

กรณทิพย์ หวานชะเอม (Kornthip Whanchaaim)<sup>1</sup> ธนัสณี เพียรตระกูล (Tanasanee Phienthrakul)<sup>2</sup>

และ คงฤทธิ หันจางสิทธิ์ (Konglit Hunchangsith)<sup>3</sup>

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล

<sup>1</sup>kornthip.wha@student.mahidol.edu, <sup>2</sup>tanasanee.ph@gmail.com, <sup>3</sup>konglit.hun@mahidol.ac.th

<sup>3</sup>konglit.hun@mahidol.ac.th

### บทคัดย่อ

การเปิดสาขาใหม่เพื่อขยายกิจการหรือเพิ่มการบริการแก่ลูกค้า มักต้องคำนึงถึงปัจจัยหลายด้าน เช่น ทำเลที่ตั้งของสาขาใหม่ ระยะห่างจากสาขาเดิม ต้นทุนการให้บริการ และ จำนวนลูกค้าที่คาดว่าจะมาใช้บริการ ดังนั้นในการเปิดสาขาใหม่แต่ละที่ จำเป็นศึกษาข้อมูลอย่างรอบคอบ การมีระบบสนับสนุนการตัดสินใจ เพื่อช่วยในการวางแผนเปิดหรือปิดสาขา จะช่วยทำให้การตัดสินใจเป็นไปได้อย่างสะดวกและแม่นยำมากขึ้น ในงานวิจัยนี้ได้ นำวิธีการวิเคราะห์เครือข่ายในรูปแบบกราฟ มาช่วยคาดการณ์จำนวนผู้ที่จะเข้าใช้บริการในสาขาเปิดใหม่ของการไฟฟ้าส่วนภูมิภาค โดยพิจารณาจากตำแหน่งที่ตั้งร่วมกับระยะห่างจากสาขาใกล้เคียง เพื่อคาดการณ์การเปลี่ยนแปลงของจำนวนลูกค้าที่จะเกิดขึ้นเมื่อต้องการเปิดหรือปิดสาขาที่ตำแหน่งนั้น จากผลการทดลองกรณีเปิดสาขาใหม่พบว่าจำนวนสาขาที่จะได้รับผลกระทบคือ 3 สาขาที่ใกล้เคียง และจำนวนลูกค้าของสาขาใหม่ที่คาดการณ์ได้มีความคลาดเคลื่อนเมื่อวัดด้วยค่าเฉลี่ยของร้อยละความผิดพลาดสัมบูรณ์ (MAPE) อยู่ที่ 15.09% สำหรับกรณีที่มีการปิดสาขา สาขาที่ได้รับผลกระทบจะอยู่ในระยะ 15 กิโลเมตรโดยรอบที่จะมีจำนวนลูกค้าเพิ่มมากขึ้น โดยค่าคาดการณ์ของจำนวนลูกค้าที่เพิ่มขึ้น ประเมินด้วยค่าเฉลี่ยของร้อยละความผิดพลาดสัมบูรณ์อยู่ที่ 3.28%

**คำสำคัญ:** ระบบสนับสนุนการตัดสินใจ การวิเคราะห์เครือข่าย การจัดการสาขา

### Abstract

Branch opening to extend business or increase customer services often have to consider many factors such as the location of the new branch, distance from the existing branches, cost of services, and the number of expected customers. Therefore, in each new branch opening, the information must be carefully studied. A decision support system for planning on branch opening or closing will help to make decision in more convenience and accurate way. In this research, graph-based network analysis method is applied to predict the number of customers in each new branch of the Provincial Electricity Authority (PEA). This prediction uses the location combined with the distance from nearby branches in order to consider the change in the number of customers when the branch is opened or closed at that location. From the branch opening experiments, 3 nearby branches are affected. The Mean Absolute Percentage Error (MAPE) of the expected number of customers in new branches is 15.09%. For branch closing, the affected branches will be located within 15 kilometers of surrounding areas where the number of customers will be increased. The MAPE of the predicted number of customers is 3.28%.

**Keywords:** Decision Support System, Network Analysis, Branch Management.

## 1. บทนำ

การเปิดสาขาใหม่ เป็นรูปแบบหนึ่งของการขยายกิจการของธุรกิจและบริการหลายประเภท เป็นการเพิ่มช่องทางให้ลูกค้าเข้าถึงสินค้าและบริการที่มากขึ้น ในบางธุรกิจ การเปิดสาขาใหม่ทำให้เป็นที่รู้จักมากขึ้น แต่ในกิจการบางประเภท การเปิดสาขาใหม่เป็นการเพิ่มความสะดวกให้กับลูกค้าในการรับบริการ ลดความแออัด และเพิ่มความสะดวกในการเดินทาง แม้การเปิดสาขาใหม่จะมีข้อดี แต่จำเป็นต้องมีงบประมาณในการลงทุน ดังนั้นก่อนการเปิดสาขาใหม่แต่ละครั้ง จำเป็นต้องคำนึงถึงต้นทุนและความคุ้มค่า

การไฟฟ้าส่วนภูมิภาค (กฟภ.) เป็นตัวอย่างของหน่วยงานที่มีการขยายสาขา เพื่ออำนวยความสะดวกแก่ลูกค้า เพื่อให้ลูกค้าสามารถใช้บริการชำระค่าไฟฟ้า หรือขอรับบริการอื่น ๆ โดยไม่ต้องเดินทางไปยังสำนักงาน ลดระยะเวลาารอคอย และลดความแออัดของการรับบริการ ดังนั้น การเลือกสถานที่ตั้งของแต่ละสาขา จึงมีความสำคัญเป็นอย่างยิ่ง สาขาของ กฟภ. มักเปิดให้บริการในศูนย์การค้า หรือ แหล่งชุมชน เพื่อให้ลูกค้าเข้าถึงได้โดยสะดวก แต่ก็มีความท้าทายในการเปิดสาขาและการดำเนินการสูง ความคุ้มค่าในการลงทุนจึงต้องได้รับการพิจารณาอย่างรอบคอบ

กฟภ. ได้มีเกณฑ์พื้นฐานสำหรับการเปิดสาขาอยู่ส่วนหนึ่ง อาทิเช่น จำนวนลูกค้าที่ใช้บริการที่สำนักงานมีมากกว่า 30,000 คน ติดต่อกัน 3 เดือน หรือ มีจำนวนการทำรายการมากกว่า 2,000 รายการต่อวัน หรือ ลูกค้าใช้ระยะเวลาารอคอยมากกว่า 15 นาที ในช่วงเวลาที่มีผู้ใช้บริการจำนวนมาก หากเกิดสถานการณ์เหล่านี้ขึ้น ถือเป็นเหตุให้ควรตัดสินใจเปิดสาขาเพิ่ม เพื่อให้บริการแก่ลูกค้าได้ดียิ่งขึ้น แต่หากต้นทุนเฉลี่ยในการบริการต่อจำนวนลูกค้าของสาขาใด มากกว่า 10 บาทต่อราย หรือ จำนวนลูกค้าที่มาใช้บริการไม่มากนัก ก็เป็นเหตุให้สมควรปิดสาขานั้นลง

ดังนั้นการพิจารณาจะเลือกตั้งสาขาที่ตำแหน่งใด จึงต้องอาศัยข้อมูลประกอบการตัดสินใจ ทั้งเรื่องของทำเลที่ตั้ง ตำแหน่งที่ตั้งของสาขาเดิม จำนวนผู้ใช้บริการในแต่ละ

สาขา และ ยังต้องอาศัยการคาดการณ์เกี่ยวกับปริมาณลูกค้าที่จะเข้าใช้บริการอีกด้วย ด้วยเหตุนี้การพัฒนากระบวนการสนับสนุนการตัดสินใจเพื่อการเปิดหรือปิดสาขา จะช่วยให้การวางแผนและการตัดสินใจเป็นไปได้โดยสะดวกมากยิ่งขึ้น นอกจากนี้จะสามารถรวบรวมข้อมูลที่เกี่ยวข้องแล้ว ยังสามารถที่จะคาดการณ์จำนวนลูกค้าที่จะเข้ามาใช้บริการในอนาคต โดยใช้วิธีการวิเคราะห์เครือข่ายในรูปแบบกราฟ เพื่อเพิ่มประสิทธิภาพในการพยากรณ์

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในหัวข้อนี้จะทบทวนทฤษฎีสำคัญ ที่ถูกนำมาใช้ในงานวิจัยนี้ ซึ่งได้แก่ ระบบสนับสนุนการตัดสินใจ และการวิเคราะห์เครือข่าย

### 2.1 ระบบสนับสนุนการตัดสินใจ

ระบบสนับสนุนการตัดสินใจ เป็นระบบที่ประมวลผลข้อมูลเพื่อให้ตัวเลือกในการตัดสินใจ โดยประกอบด้วย 3 ส่วนหลัก คือ ข้อมูล แบบจำลอง และส่วนต่อประสานกับผู้ใช้ [1] แบบจำลองเป็นส่วนประกอบที่สำคัญอันหนึ่ง ที่จะช่วยให้ได้ทางเลือกเหมาะสม หรือ สามารถประเมินโอกาสหรือความเสี่ยงของแต่ละทางเลือกที่เป็นไปได้

มีหลายงานวิจัยที่ได้นำระบบสนับสนุนการตัดสินใจไปใช้ เช่น การพัฒนาระบบสนับสนุนการตัดสินใจสำหรับการรักษาและการจัดการผู้ป่วยที่มีการใช้เครื่องช่วยหายใจ [2] ระบบให้ข้อมูลเกี่ยวกับการศึกษาต่อสำหรับนักเรียนชั้นมัธยมปลายและนักศึกษาใหม่ที่เพิ่งเข้าศึกษาในระบบมหาวิทยาลัย [3] นอกจากนี้ยังมีระบบสนับสนุนการตัดสินใจที่มีการใช้ข้อมูลเชิงพื้นที่เข้ามาช่วยประกอบการตัดสินใจ อาทิ ระบบสนับสนุนการตัดสินใจเพื่อการวางแผนงานก่อสร้างทางรถไฟ [4] การวินิจฉัยโควิด-19 ในอินโดนีเซีย [5] ที่มีการนำเสนอข้อมูลผ่านแอปพลิเคชันบนอุปกรณ์เคลื่อนที่

### 2.2 การวิเคราะห์เครือข่าย

การวิเคราะห์เครือข่าย (Network Analysis) [6] เป็นกระบวนการของการวิเคราะห์โครงสร้างเครือข่ายโดยใช้ทฤษฎีกราฟ ส่วนประกอบของเครือข่าย คือ โหนด ที่หมายถึงบุคคลหรือสิ่งของภายในเครือข่าย และ เส้น

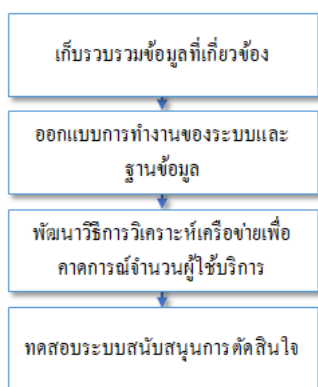
เชื่อมโยงที่แสดงความสัมพันธ์ระหว่าง โหนด การระบุ โหนดที่มีความสำคัญเป็นลักษณะพื้นฐานของการวิเคราะห์ [7]

ตัวอย่างงานวิจัยที่มีการใช้การวิเคราะห์เครือข่าย เช่น ในการศึกษาเกี่ยวกับการวิเคราะห์ความสัมพันธ์ของคำในหนังสือวรรณกรรม [8] และมีการวิเคราะห์เครือข่ายเพื่อของโครงสร้างระบบเศรษฐกิจแบบดิจิทัล [9] โดยใช้ข้อมูลจากบทความเพื่อพิจารณาว่าระบบเศรษฐกิจดิจิทัลเป็นแค่อุดมคติหรือมีความสำคัญ

ในงานวิจัยนี้จะใช้การวิเคราะห์เครือข่าย เพื่อช่วยในการพยากรณ์จำนวนลูกค้าของแต่ละสาขา เมื่อมีการเปิดหรือปิดสาขาใดไป โดยมีสมมุติฐานว่าสาขาที่อยู่ใกล้กับสาขาที่กำลังเปิดหรือปิดตัว จะได้รับผลกระทบที่มากกว่าสาขาที่อยู่ห่างไกลออกไป และจะพยากรณ์การเปลี่ยนแปลงที่เกิดขึ้นเหล่านั้น

### 3. ระบบสนับสนุนการตัดสินใจเพื่อการเปิดหรือปิดสาขา

การพัฒนา ระบบสนับสนุนการตัดสินใจเพื่อการเปิดหรือปิดสาขาได้แบ่งขั้นตอนการดำเนินงานออกเป็น 4 ขั้นตอน ดังแสดงในภาพที่ 1



ภาพที่ 1: ขั้นตอนในการพัฒนาระบบ

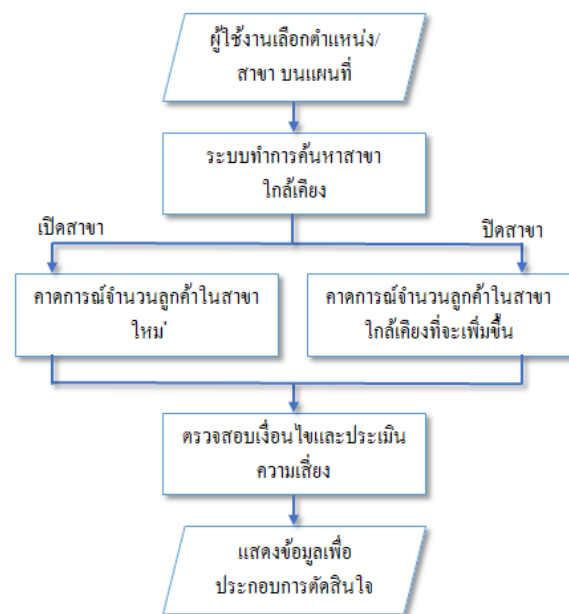
#### 3.1 เก็บรวบรวมข้อมูลที่เกี่ยวข้อง

ข้อมูลที่นำมาใช้ในงานวิจัยนี้ ได้รับความอนุเคราะห์จากไฟฟ้าส่วนภูมิภาค โดยได้ตัวอย่างข้อมูลจำนวนผู้ใช้บริการในแต่ละสาขา ในพื้นที่ 17 จังหวัดภาคกลาง ได้แก่ นครนายก ปทุมธานี ปราจีนบุรี พระนครศรีอยุธยา

สระแก้ว สระบุรี อ่างทอง ชลบุรี ฉะเชิงเทรา จันทบุรี ตราด ระยอง กาญจนบุรี นครปฐม ราชบุรี สมุทรสาคร และสุพรรณบุรี มีสาขาให้บริการทั้งสิ้น 2 แห่ง 138 แห่ง และสำนักงาน 8 แห่ง นอกจากนี้ยังมีหลักเกณฑ์หลักในการพิจารณาสถานที่ตั้งที่ใช้กันอยู่ในปัจจุบัน นอกจากนั้นยังได้ทำการรวบรวมข้อมูลอื่น ๆ ที่เกี่ยวข้อง เช่น ประมาณการค่าเช่าในแต่ละพื้นที่ ค่าสาธารณูปโภค ระยะห่างจากธนาคาร ระยะห่างจากสำนักงาน ระยะห่างจากสาขาที่ใกล้เคียง และจำนวนประชากรในพื้นที่

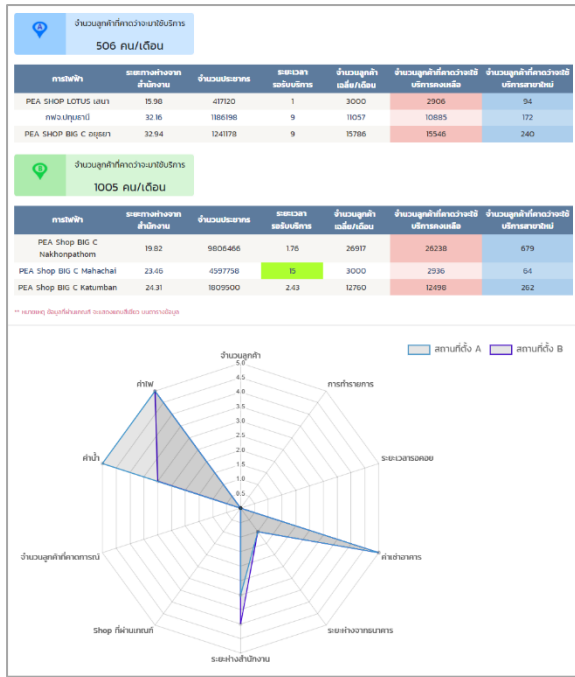
#### 3.2 ออกแบบการทำงานของระบบ

ระบบสนับสนุนการตัดสินใจที่พัฒนาขึ้นจะรองรับทั้งการเลือกที่ตั้งสาขาใหม่หรือการยกเลิกสาขาเดิม ในขั้นตอนการออกแบบ จะแบ่งออกเป็น ส่วนการทำงานของระบบ ฐานข้อมูล และการวิเคราะห์เครือข่ายเพื่อพิจารณาผลกระทบของสาขารอบ ๆ เมื่อมีการเปิดหรือปิดสาขา โดยระบบจะนำเสนอข้อมูลในรูปแบบแผนที่ และตารางภาพที่ 2 แสดงขั้นตอนการทำงานของระบบ



ภาพที่ 2: แผนภาพของระบบ

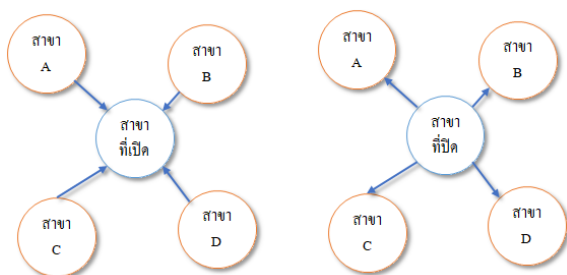
ผู้ใช้งานสามารถเลือกตำแหน่งที่ตั้ง จากนั้นระบบแสดงสาขาที่ใกล้ที่สุด คาดการณ์จำนวนลูกค้า ตรวจสอบเงื่อนไข และประเมินความเสี่ยงตามเงื่อนไขที่ตั้งไว้ จากนั้นจะแสดงผลการวิเคราะห์สาขาที่จะได้รับผลกระทบและแสดงผลในรูปแบบตารางและแผนภาพ ดังภาพที่ 3



ภาพที่ 3: ตัวอย่างผลการวิเคราะห์และการแสดงผลบนระบบ

### 3.3 พัฒนาระบบการวิเคราะห์เครือข่าย

ในระบบที่นำเสนอนี้ได้ใช้การวิเคราะห์เครือข่ายเพื่อวิเคราะห์ความสัมพันธ์ของจำนวนลูกค้าเมื่อมีสาขาใหม่เกิดขึ้นหรือสาขาที่มีอยู่ยกเลิก โดยโหนดในเครือข่ายถูกกำหนดให้เป็นสำนักงานและสาขา สำหรับเส้นเชื่อมจะเป็นการแสดงความสัมพันธ์ระหว่างสาขาที่อยู่ใกล้เคียงและตรงตามเงื่อนไข แนวคิดของจำนวนที่เปลี่ยนไปจะสัมพันธ์กับระยะทางและมีทิศทางดังแสดงในภาพที่ 4



ภาพที่ 4: ภาพแสดงทิศทางการไหลของลูกค้าเมื่อเปิดและปิดสาขา

จำนวนผู้ใช้บริการเป็นปัจจัยสำคัญในการเปิดสาขาใหม่และปิดสาขาคเดิมหากพบว่าจำนวนผู้ใช้บริการน้อยจะไม่คุ้มค่าต่อการลงทุน หรือหากปิดสาขาใดไปอาจทำให้สาขาใกล้เคียงมีผู้ใช้บริการมากขึ้นจนส่งผลกระทบต่อการใช้งานได้ และอาจส่งผลกระทบต่อสาขาอื่นเป็นทอดๆ

ไปด้วยเหตุนี้ผู้วิจัยจึงต้องการพยากรณ์จำนวนลูกค้าที่เปลี่ยนไปของสาขาที่เกี่ยวข้อง

เมื่อมีการเปิดสาขาใหม่ ลูกค้าจากสาขาที่อยู่ใกล้เคียงจะลดลง จำนวนที่ลดลงนั้นถูกนำไปคาดการณ์เป็นจำนวนลูกค้าของสาขาใหม่ซึ่งสามารถคำนวณได้จากสมการ

$$P_i = \frac{1}{d_i}, \tag{1}$$

$$W_i = \frac{P_i}{\sum_{i=1}^n P_i}, \tag{2}$$

$$C_{new} = \sum_{i=1}^n \left( \frac{C_i}{2} \times W_i \right) \tag{3}$$

เมื่อ  $d_i$  เป็นระยะห่างของสาขาใกล้เคียงลำดับที่  $i$   
 $P_i$  คือ ค่าความใกล้เคียงของสาขาใกล้เคียงลำดับที่  $i$   
 $W_i$  คือ ค่าน้ำหนักของสาขาที่ใกล้เคียงลำดับที่  $i$   
 $n$  เป็นจำนวนสาขาใกล้เคียงที่ใช้ในการทดสอบ  
 $C_i$  คือ จำนวนผู้ใช้บริการเฉลี่ยต่อเดือนของแต่ละสาขาใกล้เคียงลำดับที่  $i$  และ

$C_{new}$  คือ จำนวนผู้ใช้บริการที่ทำนายได้ของสาขาใหม่ จากสมการ [3] จะเห็นว่ามีการหาร  $C_i$  ด้วย 2 เนื่องจากลูกค้าแต่ละคนมีความน่าจะเป็น 0.5 ที่จะย้ายไปใช้บริการในสาขาใหม่ ดังนั้นจึงคำนวณค่าคาดหวัง (Expected Value) ของจำนวนลูกค้าที่จะย้ายไปใช้บริการในสาขาใหม่ได้เป็น  $C_i/2$  ก่อนนำไปคูณกับค่าถ่วงน้ำหนักที่คำนวณจากระยะทางอีกครั้ง

ในกรณีที่มีการยกเลิกบริการของสาขาใด จำนวนลูกค้าในสาขาอื่นๆ จะเพิ่มขึ้น โดยลูกค้าจากสาขาที่ปิดจะกระจายไปยังสาขาใกล้เคียงโดยใช้ค่าถ่วงน้ำหนักที่คิดจากระยะทางแบบเดียวกับตอนเปิดสาขาใหม่ จำนวนลูกค้าที่จะไปเพิ่มให้กับสาขาต่างๆ คำนวณได้จาก

$$C_i = W_i \times C_{close} \tag{4}$$

เมื่อ  $C_{close}$  เป็นจำนวนผู้ใช้บริการเฉลี่ยต่อเดือนของสาขาที่จะดำเนินการปิด

### 3.4 ทดสอบระบบสนับสนุนการตัดสินใจ

หลังจากทำนายจำนวนผู้ใช้บริการ จะทำการประเมินผลเพื่อดูความคลาดเคลื่อนของการทำนาย โดยใช้ค่าเฉลี่ยของร้อยละความผิดพลาดสัมบูรณ์ (MAPE) ตาม

สมการ [5] เพื่อเปรียบเทียบระหว่างค่าที่ทำนายได้กับข้อมูลที่เกิดขึ้นจริง โดยใช้สถานการณ์ที่เคยเกิดขึ้นในอดีต

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad [5]$$

เมื่อ  $n$  คือ จำนวนสาขาที่นำมาพิจารณา

$y_i$  คือ จำนวนลูกค้าที่มาใช้บริการสาขา  $i$  จริง

$\hat{y}_i$  คือ จำนวนลูกค้าที่มาใช้บริการสาขา  $i$  ที่ทำนายได้

ในการทดสอบจะทำนายจำนวนผู้ใช้บริการของเดือนที่เปิดสาขาจำนวน 28 แห่ง และทำนายจำนวนผู้ใช้บริการของเดือนหลังจากที่มีการยกเลิกสาขาจำนวน 4 แห่ง นำมาทดสอบโดยพิจารณาสถานที่ใกล้เคียงที่มีระยะทางแตกต่างกันไป หรือ จำนวนสาขาใกล้เคียงที่แตกต่างกันไป เพื่อเลือกระยะที่เหมาะสมสำหรับการแสดงผลบนระบบสนับสนุนการตัดสินใจ

#### 4. ผลการทดสอบ

เพื่อทำการทดสอบระยะใกล้เคียงที่เหมาะสมในการคาดการณ์จำนวนผู้ใช้บริการที่เปลี่ยนไปได้ทำการทดสอบโดยกำหนดระยะใกล้เคียงไว้ 6 แบบคือ จำนวนสาขาที่ใกล้เคียงกับสถานที่เลือกหรือสาขาที่ต้องการยกเลิกจำนวน 3 แห่ง 5 แห่ง 7 แห่ง และ ระยะห่างของสถานที่เลือกหรือสาขาที่ต้องการยกเลิกกับสาขาใกล้เคียง เป็นระยะทาง 5 กม. 10 กม. และ 15 กม.

สำหรับจำนวนผู้ใช้บริการเฉลี่ยต่อเดือนของแต่ละสาขา หรือ  $C_i$  มีการเปรียบเทียบเพื่อเลือกใช้งาน 3 รูปแบบได้แก่

- C1: จำนวนผู้ใช้บริการเฉลี่ย 1 ปี ย้อนหลัง นำมาเฉลี่ย
- C2: จำนวนผู้ใช้บริการเฉลี่ย ย้อนหลัง 1 เดือน และ
- C3: จำนวนผู้ใช้บริการเดือนเดียวกันของปีก่อนหน้า

ทำการแบ่งการทดสอบออกเป็น 2 ส่วน คือการเปิดสาขาใหม่ และ การปิดสาขา โดยทำการทดสอบที่มีการเลือกสถานที่ใกล้เคียงในลักษณะเดียวกัน และใช้ข้อมูลย้อนหลังเพื่อพยากรณ์ค่าในลักษณะเดียวกัน

##### 4.1 การเปิดสาขาใหม่

สำหรับการทดสอบการจำนวนลูกค้าที่จะมาใช้บริการในสาขาใหม่ จะใช้ข้อมูลในอดีตของสาขาที่อยู่ใกล้เคียงเพื่อทำนายว่าเมื่อเปิดสาขาใหม่จะมีผู้มาใช้บริการเท่าใด

เปรียบเทียบกับข้อมูลจริงเมื่อได้ทำการเปิดสาขาเดือนแรกทำการทดสอบทำนายจำนวนลูกค้าของสาขาใหม่จำนวน 28 สาขา และนำมาคำนวณค่า  $MAPE$  เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์โดยใช้เงื่อนไขที่ต่างกัน พบว่าเป็นไปตามตารางที่ 1

ตารางที่ 1: ค่า  $MAPE$  จากการทดสอบการเปิดสาขาใหม่

ระยะใกล้เคียง	C1	C2	C3
สถานที่ใกล้เคียง 3 แห่ง	<b>15.09</b>	24.59	19.84
สถานที่ใกล้เคียง 5 แห่ง	20.03	22.31	33.66
สถานที่ใกล้เคียง 7 แห่ง	28.68	26.14	44.97
ระยะห่างไม่เกิน 5 กิโลเมตร	19.63	32.39	25.16
ระยะห่างไม่เกิน 10 กิโลเมตร	16.57	28.48	21.23
ระยะห่างไม่เกิน 15 กิโลเมตร	20.70	25.13	32.16

จากผลการทดสอบในตารางที่ 1 เมื่อเลือกใช้สถานที่ใกล้เคียง 3 แห่ง และใช้จำนวนผู้ใช้บริการเฉลี่ย 1 ปี ย้อนหลัง นำมาเฉลี่ยกัน เพื่อทำนายจำนวนผู้ใช้บริการในสาขาใหม่ พบว่าให้ผลการทำนายที่มีค่าผิดพลาดต่ำสุด คือ 15.09% ในขณะที่การใช้จำนวนผู้ใช้บริการหนึ่งเดือน ย้อนหลัง หรือ เดือนเดียวกันของปีก่อนหน้ามีความคลาดเคลื่อนสูงกว่า นั่นอาจเป็นเพราะการใช้ค่าเฉลี่ยของทั้งปีเป็นตัวแทนจะช่วยลดความไม่แน่นอนที่ลูกค้าจะไปใช้บริการสาขาเยอะหรือน้อยเป็นพิเศษในบางเดือนได้

##### 4.2 การปิดสาขาเดิม

การพิจารณาปิดหรือยกเลิกสาขาเดิม จะเป็นการกระจายผู้ใช้บริการในสาขาเดิมออกไปยังสาขาที่อยู่ใกล้เคียง โดยสัดส่วนของลูกค้าที่จะกระจายไปยังแต่ละสาขาจะแปรผกผันกับระยะทาง ค่าที่ทำนายได้จะถูกนำไปเปรียบเทียบกับค่าจริงหลังมีการปิดสาขาแล้ว โดยที่มีสาขาที่เคยปิด และนำมาใช้ในการทดสอบนี้จำนวน 4 สาขา

สำหรับจำนวนลูกค้าในสาขาเดิมก่อนปิด ที่จะนำไปกระจายให้สาขาที่อยู่ใกล้เคียง จะมีการทดสอบโดยใช้ค่าเฉลี่ยย้อนหลัง 1 ปีก่อนปิด (C1) จำนวนลูกค้า 1 เดือนก่อนปิด (C2) และ จำนวนลูกค้าในเดือนที่ปิดของปีก่อนหน้า (C3) ผลการทดสอบวัดค่าเฉลี่ย  $MAPE$  ของสาขาใกล้เคียงหลังจากมีการปิดสาขา แสดงในตารางที่ 2

ตารางที่ 2: ผลการทดสอบแสดงค่า *MAPE* ของการปิดสาขา

ระยะใกล้เคียง	C1	C2	C3
สถานที่ใกล้เคียง 3 แห่ง	7.49	7.86	9.57
สถานที่ใกล้เคียง 5 แห่ง	80.14	64.50	65.07
สถานที่ใกล้เคียง 7 แห่ง	37.06	37.21	37.30
ระยะห่างไม่เกิน 5 กิโลเมตร	145.09	46.96	62.96
ระยะห่างไม่เกิน 10 กิโลเมตร	9.56	9.98	12.60
ระยะห่างไม่เกิน 15 กิโลเมตร	<b>3.28</b>	3.43	3.48

จากผลการทดสอบพบว่า สาขาใกล้เคียงที่เหมาะสมในการคาดการณ์จำนวนลูกค้าที่เพิ่มขึ้น คือสาขาที่มีระยะห่างไม่เกิน 15 กิโลเมตร โดยใช้ข้อมูลจำนวนผู้ใช้บริการเฉลี่ยย้อนหลังหนึ่งปีในการทำนายค่า ซึ่งให้ค่าความผิดพลาดต่ำที่สุดคือ *MAPE* ที่ 3.28% ดังแสดงในตาราง

## 5. สรุปผล

งานวิจัยนี้ได้นำเสนอระบบสนับสนุนการตัดสินใจสำหรับพิจารณาเปิดสาขาใหม่และยกเลิกสาขาเดิม ผู้ใช้งานสามารถเลือกสถานที่ตั้งสาขาใหม่ได้มากกว่า 1 เพื่อทำการเปรียบเทียบ หรือสามารถเลือกจำลองการยกเลิกสาขา ระบบจะทำการวิเคราะห์และนำเสนอข้อมูลเพื่อประกอบการตัดสินใจ ทั้งข้อมูลเชิงภูมิศาสตร์ ข้อมูลการใช้บริการในอดีต และ ข้อมูลที่ระบบคาดการณ์จำนวนผู้ใช้งานในอนาคต

ในขั้นตอนการวิเคราะห์ได้นำเอาวิธีการวิเคราะห์เครือข่ายในรูปแบบกราฟ มาช่วยวิเคราะห์ความคุ้มค่าของการเปิดสาขาใหม่ และผลกระทบของสาขาที่อยู่บริเวณใกล้เคียง จากการทดสอบพบว่าหาทำการเปิดสาขาใหม่สาขาที่จะได้รับผลกระทบจะเป็นสาขาใกล้เคียง 3 แห่ง แต่หากทำการปิดสาขา สาขาที่ได้รับผลกระทบจะเป็นสาขาใกล้เคียงในระยะ 15 กิโลเมตร โดยได้รับผลกระทบแตกต่างกันไปตามระยะห่าง

## กิตติกรรมประกาศ

ขอขอบคุณ การไฟฟ้าส่วนภูมิภาค ที่สนับสนุนข้อมูลหลักในงานวิจัยนี้ ซึ่งมีส่วนสำคัญที่ทำให้งานวิจัยสามารถดำเนินงานได้อย่างราบรื่น

## เอกสารอ้างอิง

- [1] N. Ghorbani Khodashahri, M. Mir Hassannia Sarabi. "Decision support system (DSS)." *Singaporean journal of business economics and management studies*, Vol. 1, No. 6, 2013.
- [2] T. Man, Z. Nataly, T. Georgi. "A Multilevel Intelligent Assistant for Multilevel Social Network Analysis." *IEEE 10th International Conference on Intelligent Systems*, Varna, Bulgaria, Aug 28-30, 2020.
- [3] K. Vassilis, P. Nikolaos, and V. Michael. "Analysis and design of a web-based decision support system for choosing higher education studies." *Yugoslav Journal of Operations Research*, Vol. 24, No. 3, pp. 399-414, January, 2014.
- [4] S. Qin, L. Zhou, Y. Yue. "A Web-based Distributed Group Decision Support System for Railway Construction Organization." *2nd International Asia Conference on Informatics in Control*, Wuhan, China, pp. 362-364, Mar 6-7, 2010.
- [5] Henderi, S. Didik, M. Miftah, and Q. Taufik. "Model Decision Support System for Diagnosis COVID-19 Using Forward Chaining A Case in Indonesia." *8th International Conference on Cyber and IT Service Management (CITSM)*, Pangkal, Indonesia, Oct 23-24, 2020.
- [6] T. Shazia, S. F. P. Fabiola, F. Sofia, and G. João. "Social Network Analysis an Overview." *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, Porto, Portugal, April, 2018.
- [7] Z. Yanping, B. Yuanyuan, Z. Shu, C. Jie, T. Jie, "Identifying Node Importance by Combining Betweenness Centrality and Katz Centrality." *International Conference on Cloud Computing and Big Data*, Shanghai, China, Nov 4-6, 2015.
- [8] D. Radha, and K. P. T. Nithia. "A Case Study on Social Network Analysis Thesaurus Book." *2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions*, Bengaluru, India, Dec 21- 22, 2017.
- [9] M. Oxana, G. Galina, K. Alexander. "Social Network Analysis of the Functional Meaning of the Term Digital Economy." *Eleventh International Conference "Management of large-scale system development" (MLSD)*, Moscow, Russia, Oct 1-3, 2018.

# การฝึกอบรมโครงข่ายประสาทเทียมเพื่อทำนายผลข้างเคียงหลังการฉีดวัคซีนโควิด-19 โดยใช้ขั้นตอนวิธีการหาค่าที่เหมาะสมซึ่งได้รับแรงบันดาลใจจากธรรมชาติ

## Neural Network Training for Prediction of Side Effects After COVID-19 Vaccination Using Optimization Algorithms Inspired by Nature

สุภาพร แคนกาไสย (Supaporn Deankasai)<sup>1</sup> หนึ่งฤทัย ดีมีวงษ์ (Neungrutai Deemeewong)<sup>2</sup>

ภูมินทร์ ดวนจันทร์ (Phumin Duankhan)<sup>3</sup> และคำรณ สุนิติ (Khamron Sunat)<sup>4</sup>

สาขาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

<sup>1</sup>desupaporn@kkumail.com, <sup>2</sup>neungrutai\_d@kkumail.com, <sup>3</sup>mini@kkumail.com, <sup>4</sup>skhamron@kku.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาขั้นตอนวิธีสอนตัวแบบโครงข่ายประสาทเทียมชนิดป้อนไปข้างหน้าสำหรับทำนายผลข้างเคียงหลังการฉีดวัคซีนโควิด-19 มีขั้นตอนวิธีที่นำมาทำการทดสอบครั้งนี้ ขั้นตอนวิธี Adaptive Multi-Population Optimization (AMPO) ขั้นตอนวิธี Grey wolf (GWO) ขั้นตอนวิธี An elite-guided hierarchical differential evolution algorithm (EHDE) และ White Shark Optimizer (WSO) ชุดข้อมูลคือข้อมูลการรับวัคซีนของผู้ป่วย 7 โรคเรื้อรัง จากผลการทดสอบพบว่าตัวแบบจำลองที่สอนด้วย GWO ให้ค่าความแม่นยำที่ดีที่สุดตามด้วยตัวแบบจำลองที่สอนด้วย AMPO EHDE และ WSO ตามลำดับ

**คำสำคัญ:** การวิเคราะห์ข้อมูล การจำแนกข้อมูลหลายกลุ่ม ประชากร วัคซีนโควิด-19 โครงข่ายประสาทเทียม

### Abstract

This research aims to determine an algorithm to teach a forward-feed neural network model for predicting side effects after COVID-19 vaccination. The following algorithms were tested: Adaptive Multi-Population Optimization (AMPO), Gray wolf (GWO) algorithm, An elite-guided hierarchical differential evolution algorithm (EHDE), and White Shark Optimizer (WSO). The dataset is immunization data for patients with 7 chronic diseases. As a result of testing, the GWO-trained model gave the best

accuracy. This was followed by models taught with AMPO EHDE and WSO, respectively.

**Keywords:** Data analyzes, multi-population, Nature-inspired Algorithm, Classification, Covid-19 vaccine, Neural network.

### 1. บทนำ

การใช้ขั้นตอนวิธีเมตาฮีริสติก (Metaheuristic) Grey wolf (GWO) [5] [7] สามารถช่วยหาคำตอบตัวแบบสำหรับแบ่งกลุ่มของข้อมูลที่มีประชากรหลายกลุ่มได้ก่อนข้างดี ปัจจุบันมีขั้นตอนวิธีอื่นที่ถูกนำเสนอใหม่หลายวิธี เช่น ขั้นตอนวิธี เช่น Adaptive Multi-Population Optimization Algorithm (AMPO) [4] ขั้นตอนวิธี An elite-guided hierarchical differential evolution algorithm (EHDE) [11] และขั้นตอนวิธี White Shark Optimizer (WSO) [1] บทความนี้ศึกษาเปรียบเทียบประสิทธิภาพการทำนายผลข้างเคียงหลังการฉีดวัคซีนเมื่อใช้ทั้ง 4 ขั้นตอนวิธีทำการฝึกสอนตัวแบบโครงข่ายประสาทเทียมแบบป้อนไปข้างหน้า

### 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

#### 2.1 ข้อมูลที่นำมาใช้ในการวิจัย

ข้อมูลคือชุดข้อมูลผู้เข้ารับการได้รับวัคซีนต้านโควิด 19 ของผู้ป่วย 7 โรคเรื้อรัง ที่ได้รับการฉีดวัคซีนจากสำนักงานป้องกันควบคุมโรคที่ 7 ขอนแก่น โดยตัวแปรต้น มี 3 ตัวแปร ได้แก่ ชนิดวัคซีนที่ฉีด โรคเรื้อรังของ

ผู้ป่วย อายุของผู้ป่วย ตัวแปรตาม 1 ตัวแปร คือ อาการ หลังจากการฉีดวัคซีนแล้ว จำนวนข้อมูลมีทั้งสิ้น 3969 รายการ

โดยผู้วิจัยจะแปลงข้อมูลจากตัวแปรต้นที่ 1 ซึ่งเป็น ข้อมูลชนิดของวัคซีน 1 คือ Sinovac Life Sciences, 2 คือ Sinopharm, 3 คือ AstraZeneca, 4 คือ Pfizer, BioNTech ตัวแปรต้นที่ 2 คือ กลุ่มโรคที่เสี่ยงโดยจะเลือกเอากลุ่มเสี่ยงจากโรคเรื้อรัง 7 ชนิด ประกอบไปด้วย 1 คือ โรคทางเดินหายใจเรื้อรัง 2 คือ โรคหัวใจและโรคหลอดเลือด 3 คือ โรคไตเรื้อรังที่อยู่ระยะที่ 5 ขึ้นไป 4 คือ โรคหลอดเลือดสมอง 5 คือ โรคมะเร็งทุกชนิด 6 คือ โรคเบาหวาน 7 คือ โรคอ้วน เนื่องจากข้อมูลในกลุ่มนี้เป็นข้อมูลที่สมบูรณ์กว่ากลุ่มอื่นทางผู้วิจัยคิดว่าสามารถใช้ในการทำนายได้

ตัวแปรต้นที่ 3 วันเกิด ผู้วิจัยจะแปลงเป็นอายุและแบ่งข้อมูลตามช่วงอายุดังนี้ 1 คือ อายุต่ำกว่า 18 ปี 2 คือ อายุ 18-35 18-35 ปี 3 คือ อายุ 36-39 ปี 4 คือ อายุ 60 ปีขึ้นไป

ตัวแปรตาม เป็นอาการข้างเคียงหลังจากการฉีดวัคซีน โดยผู้วิจัยจะแบ่งอาการข้างเคียง[10]เป็น 3 ระดับ คือ 1 คือ อาการไม่รุนแรง ประกอบไปด้วยอาการ ปวด บวม แดง บริเวณที่ฉีด ปวดศีรษะ ไข้ต่ำ น้อย อ่อนเพลีย ไม่มีแรง ปวดเมื่อยกล้ามเนื้อ 2. อาการรุนแรง ประกอบไปด้วย อาการ มีไข้สูง กล้ามเนื้ออ่อนแรง อาเจียน ท้องเสีย ปวดข้อ แขนขบชา ตาพร่ามัว 3. อาการแพ้วัคซีน ประกอบไปด้วยอาการ มีผื่น มีจุดจ้ำเลือด หายใจลำบาก หัวใจเต้นผิดปกติ หงุดหงิด ซักหมดสติ

## 2.2 Adaptive Multi-Population Optimization (AMPO) [4]

ขั้นตอนวิธี AMPO ประกอบไปด้วยประชากร ซึ่งถูกแบ่งเป็น 5 กลุ่มย่อย มีดังนี้ 1) กลุ่มการค้นหาแบบสุ่ม 2) กลุ่มการค้นหาแบบกว้าง 3) กลุ่มการค้นหาในท้องถิ่น 4) กลุ่มผู้นำ 5) กลุ่มการขยับขนาดของกลุ่มประชากรย่อย เหล่านี้เป็นแบบพลวัตระหว่างการค้นหา นอกจากนี้ยังมีการกำหนดกลยุทธ์การค้นหาที่แตกต่างกันให้แต่ละประชากรในแต่ละกลุ่มย่อยเพื่อความสะดวกกลุ่มการค้นหาแบบกว้าง กลุ่มการค้นหาท้องถิ่น และกลุ่มผู้นำจะเรียกว่ากลุ่มต้นทางเสมอ

(1) ประชากร: ใน AMPO ขนาดของประชากรจะแสดงด้วย  $N_{pop}$  แต่ละตัววิธีแก้ไขปัญหาคือเป็นไปได้อย่าง แสดงด้านล่าง

$$X_i^t = [x_{i,1}^t, x_{i,2}^t, \dots, x_{i,n}^t]^T \quad (1)$$

โดยที่  $X_i^t$  คือ วิธีการแก้ไขปัญหาคือเป็นไปได้อย่าง

$i \in \{1, 2, \dots, N_{pop}\}$  คือ ดัชนีของประชากร

$n$  คือ จำนวนรวมของตัวแปร

(2) รูปแบบการค้นหาแบบสุ่ม:

$$X_i^t = U(LB, UB) \quad (2)$$

โดยที่  $U$  คือ ตัวสร้างเลขสุ่มตามการแจกแจงแบบสม่ำเสมอ,  $LB$  คือ ขอบล่าง,  $UB$  คือ ขอบบน

ประชากรของกลุ่มการค้นหาแบบสุ่มสามารถเปลี่ยนเป็นกลุ่มการค้นหาแบบกว้างหรือกลุ่มการค้นหาแบบท้องถิ่นได้

(3) รูปแบบการค้นหาแบบกว้าง:

ประชากรของกลุ่มการค้นหาแบบกว้างที่มีความสัมพันธ์ขนาดขั้นตอนการค้นหาที่ใหญ่ขึ้นอย่างมาก ทำการค้นหาแบบกว้างมีการเคลื่อนที่ดังนี้

$$GS_i^{t+1} = w \times GS_i^t + r \times (gbest - x_i^t) \quad (3)$$

$$X_i^{t+1} = X_i^t + G_i^{t+1} \quad (4)$$

โดยที่  $GS_i^t$  คือ ขนาดขั้นตอนการค้นหาของกลุ่มการค้นหาแบบกว้างแต่ละกลุ่ม,  $w$  คือ ปัจจัยถ่วงน้ำหนักมีค่าเป็น 0 หรือ 1,  $gbest^t$  คือ ค่าที่เหมาะสมที่ประชากรจะได้รับจนถึง  $t^{th}$ ,  $r$  คือ ตัวเลขสุ่มที่เกิดจากการกระจายตัวแบบมาตรฐานของ  $U(0,1)$

(4) รูปแบบการค้นหาท้องถิ่น: มีการเคลื่อนที่เป็นดังนี้

$$\sigma_i^{t+1} = \gamma \times \sigma_i^t \quad (5)$$

$$LS_i^{t+1} = \text{Gaussian}(0, \sigma_i^{t+1}) \times X_i^t \quad (6)$$

$$X_{i+1} = X_i + LS_i^{t+1} \quad (7)$$

โดยที่  $LS_i^{t+1}$  คือ ขนาดขั้นตอนการค้นหาพื้นที่กลุ่มประชากรที่  $i$  ที่  $t^{th}$  วันซ้ำ,  $\gamma$  คือ อัตราการสลายตัวและคงที่มีค่าอยู่ระหว่าง 0 ถึง 1,  $\text{Gaussian}(0, \sigma_i^{t+1})$  คือ การสุ่มเลขตามการแจกแจงแบบ Gaussian ด้วยค่าเฉลี่ยของศูนย์และค่าเบี่ยงเบนมาตรฐานของ  $0, \sigma_i^{t+1}$



### 2.3 Grey Wolf Optimizer (GWO)

GWO เป็น เมตาฮิวริสติก แบบกลุ่มเลียนแบบลำดับชั้นทางสังคมและพฤติกรรมการล่าเหยื่อของหมาป่าสี่เท้า แบ่งประชากรเป็น 4 กลุ่ม คือ  $\alpha$   $\beta$   $\delta$  และ  $\omega$  โดยหมาป่าอันดับสูงที่สุดสามตัวแรก คือ  $\alpha$   $\beta$  และ  $\delta$  ตามลำดับ มีหน้าที่ชักนำและสั่งการหมาป่าตัวอื่น ๆ ซึ่งเรียกว่า  $\omega$  ไปพื้นที่ ที่มีแนวโน้มว่าจะเป็นตำแหน่งของเหยื่อฝูงหมาป่าสี่เท้าแต่ละตัวจะปรับตำแหน่งรอบ ๆ โดยการขึ้นนำของผู้นำทั้งสาม ดังนี้

$$D_z = |C \cdot X_p(t) - X(t)| \quad (8)$$

$$X(t+1) = X_p(t) - A \cdot D \quad (9)$$

$$X(t+1) = \frac{\sum_{i=1}^n X_i}{3} \quad (10)$$

โดยที่  $X_z$  คือ ตำแหน่งของ  $\alpha$   $\beta$  และ  $\delta$ ,

$A_i$  คือ เวกเตอร์สุ่ม,  $t$  คือ การวนซ้ำ

$A = 2a \times r1a$ ,  $C = 2 \cdot r2$ ,  $X_p$  คือ ตำแหน่งของเหยื่อ  $X$  คือ ตำแหน่งของหมาป่าสี่เท้า  $a$  คือ การลดเส้นจาก 2 เป็น 0,  $r1, r2$  คือ เวกเตอร์สุ่มใน [0,1]

พารามิเตอร์สุ่ม  $A$  และ  $C$  ช่วยให้ตัวหมาป่าย้ายไปตำแหน่งรอบ ๆ เหยื่อได้ ดังนั้นการปรับปรุงตำแหน่งตามสมการที่ 8 และสมการที่ 9 ทำให้หมาป่าทั้งฝูงสามารถล้อมเหยื่อและโจมตีเหยื่อได้

### 2.4 White Shark Optimizer (WSO) [1]

แนวคิดหลักของ WSO ได้รับแรงบันดาลใจจากพฤติกรรมของฉลามขาว รวมทั้งประสาทสัมผัสพิเศษในการไต่กลิ่นและการดมกลิ่นขณะนำทางและหาอาหาร ขั้นตอนย่อยมีดังนี้

(1) ติดตามเหยื่อ: ฉลามขาวท่องทะเลไปพร้อมกับการล่าเหยื่อมีประสาทสัมผัสที่ดีหลายอย่างไม่ว่าจะเป็น การมองเห็น การดมกลิ่น การไต่กลิ่น

(2) ค้นหาเหยื่อ (สำรวจ): ค้นหาเหยื่อโดยใช้ประสาทสัมผัสเสียงจากการเปลี่ยนแปลงของแรงดันน้ำเมื่อฉลามขาวอยู่ใกล้เป้าหมาย มันจะสามารถจับสนามแม่เหล็กไฟฟ้า และเมื่อมันพบเหยื่อ มันจะเคลื่อนที่หาเหยื่อด้วยการเคลื่อนไหวเป็นลูกคลื่น

$$v = xf \quad (11)$$

โดยที่  $v$  คือ ความเร็วของการเคลื่อนที่แบบคลื่น,  $x$  คือ ความยาวคลื่นที่กำหนดระยะเวลาทางที่ฉลามขาวเดินทางเป็นลูกคลื่น,  $f$  คือ ความถี่ของการเคลื่อนที่แบบคลื่นที่กำหนดโดยจำนวนรอบที่ฉลามขาวทำสำเร็จต่อวินาที (Hz)

(3) ค้นหาเหยื่อ (การเอารัดเอาเปรียบ): ใช้ประสิทธิภาพรับกลิ่นเฉพาะตอนเข้าใกล้เหยื่อ ขณะเคลื่อนตัวไปหาเหยื่อสามารถเคลื่อนที่ด้วยความเร่งคงตัวสามารถคิดเป็นสมการได้ดังนี้

$$x = x_i + v_i \cdot \Delta t + \frac{1}{2} a (\Delta t)^2 \quad (12)$$

โดยที่  $x$  คือ ตำแหน่งของฉลามขาว,  $x_i$  คือ ตำแหน่งดั้งเดิม,  $v_i$  คือ ความเร็วต้นของฉลามขาว,  $\Delta t$  คือ ระยะเวลาหาค่าเริ่มต้นและปัจจุบัน,  $a$  คือ ค่าคงที่

### 2.5 An elite-guided hierarchical differential evolution algorithm (EHDE) [11]

ขั้นตอนวิธี EHDE มีโครงสร้างประชากรแบบลำดับชั้นในแต่ละรุ่นประชากรจะแบ่งเป็น 3 กลุ่ม ตามค่า fitness โดยแต่ละกลุ่มมีบทบาทเฉพาะตัวตามลำดับชั้น โดยประชากรชั้นนำค้นหาโดยใช้การเคลื่อนที่แบบ Lévy flight เพื่อค้นหาเชิงแคบ ประชากรชั้นกลางมุ่งเน้นไปที่การค้นหาในท้องถิ่นรอบ ๆ ประชากรที่อยู่ชั้นล่างให้ความสำคัญกับการค้นหาเชิงกว้างมาก ตามคำแนะนำของประชากรชั้นที่กลาง กลยุทธ์การปรับพารามิเตอร์ควบคุมของ EHDE ได้เพิ่มส่วนประกอบแบบสุ่มและส่วนประกอบที่แปรผันตามเวลาด้วย เพื่อประโยชน์ของการประเมินประสิทธิภาพของ EHDE ประชากรชั้นนำเคลื่อนย้าย ดังนี้

$$V_{i,g+1} = \begin{cases} X_{i,g} + rand_{i,g} \cdot Levy(\lambda), & rand < 0.5 \\ L + rand_{i,g} \cdot (UB - LB), & otherwise \end{cases} \quad (13)$$

โดยที่  $X_{i,g}$  คือประชากรที่ดีที่สุด,

$rand_{i,g}, rand$  คือ การสุ่ม

ประชากรชั้นกลางเน้นการหาประโยชน์จากการที่ประชากรชั้นบนสุดเคลื่อนย้ายดังนี้:

$$V_{i,g} = X_{1,g} + F_{i,g} (X_{r2,g} - X_{r3,g}) \quad (14)$$

โดยที่  $X_{1,g}$  คือประชากรที่ดีที่สุด,  $F_{i,g}$  คือ ปรับขนาดปัจจัย,  $X_{r2,g}, X_{r3,g}$  คือ ได้จากการสุ่ม

ประชากรชั้นล่างเคลื่อนย้ายโดยยึดตามกลุ่มชั้นบนสุด  
 สุ่มประสิทธิภาพถ่วงน้ำหนักเพื่อรักษาความหลากหลาย  
 ของประชากร รักษาสมดุลระหว่างการสำรวจและ  
 เคลื่อนย้าย ดังนี้:

$$V_{i,g+1} = \alpha_{i,g} \cdot X_{r1,g}^{Ke} + F_{i,g} (X_{r2,g} - X_{r3,g}) \quad (15)$$

โดยที่  $K_e$  คือ เซตที่ดีที่สุด,  $X_{r1,g}^{Ke}$  คือ การถูกสุ่มจาก  
 $K_e$ ,  $\alpha_{i,g}$  คือ สัมประสิทธิ์ถ่วงน้ำหนัก

### 3. วิธีการดำเนินการวิจัย

#### 3.1 การแก้ไขปัญหาข้อมูลไม่สมดุล

เนื่องจากข้อมูลที่ได้อาจมีปัญหาข้อมูลไม่สมดุล  
 (Imbalanced data) เพราะ ผลข้างเคียงของการฉีดวัคซีน  
 อาการแพ้วัคซีนมีค่าต่างจากผลข้างเคียงที่ไม่รุนแรงดังนั้น  
 ผู้วิจัยจึงได้นำเทคนิคการแก้ไขข้อมูลที่ไม่สมดุล 2 วิธีการ  
 ดังนี้

(1) วิธีสุ่มเกิน (Oversampling) เป็นการเพิ่มจำนวน  
 ข้อมูลที่อยู่ในกลุ่มที่มีข้อมูลส่วนน้อยให้มีจำนวนใกล้เคียง  
 หรือเท่ากับจำนวนของข้อมูลที่อยู่ในกลุ่มส่วนมาก ในการ  
 วิจัยครั้งนี้จะใช้วิธีการสุ่มอย่างเป็นระบบ ซึ่งการเพิ่มข้อมูล  
 นั้นจะเพิ่มโดยการสุ่มเลือกจากข้อมูลเดิม (อาการความ  
 รุนแรงหลังจากฉีดวัคซีนระดับ 1-3 ตามลำดับ) ซึ่งผลการ  
 เพิ่มจำนวนข้อมูลที่อยู่ในกลุ่มที่มีข้อมูลส่วนน้อยให้มีข้อมูล  
 ที่ใกล้เคียงหรือเท่ากับจำนวนข้อมูลที่อยู่ในกลุ่มส่วนมาก

(2) วิธีสังเคราะห์ข้อมูลเพิ่ม (Synthetic Minority  
 Oversampling Technique: SMOTE) มาใช้ปรับสมดุล  
 ข้อมูลด้วยวิธีการที่นำมาใช้จะทำการเพิ่มจำนวนข้อมูล  
 กลุ่มน้อย ซึ่งก็คือจำนวนข้อมูลของ คลาสที่ 1-4 ก่อนจะ  
 เข้าสู่กระบวนการทำนายข้อมูลและนำข้อมูลที่แก้ปัญห  
 ความไม่สมดุลของข้อมูลแล้ว มาทำนาย เป็นวิธีการสุ่ม  
 ตัวอย่างเพิ่ม โดยทำการสังเคราะห์ข้อมูลขึ้นมาใหม่จาก  
 ข้อมูลเดิมที่มีอยู่ดังตารางที่ 2

ตารางที่ 1: ตารางแสดงจำนวนอาการข้างเคียงหลังฉีดวัคซีน

กลุ่มโรค	ความรุนแรงของอาการข้างเคียง		
	ไม่รุนแรง	รุนแรง	แพ้วัคซีน
โรคทางเดินหายใจ	647	68	37

โรคหัวใจและโรคหลอดเลือด	254	32	31
โรคไตเรื้อรังระยะที่รุนแรง	488	49	26
โรคหลอดเลือดในสมอง	177	15	12
โรคมะเร็งทุกชนิด	230	33	8
โรคเบาหวาน	589	330	44
โรคอ้วน	820	89	188
รวม	3165	616	188

ตารางที่ 2: เปรียบเทียบข้อมูลก่อนและหลังการทำ SMOTE

ชนิดข้อมูล	จำนวนของรายการข้อมูลข้างเคียงแยกตามอาการ		
	ไม่รุนแรง	รุนแรง	แพ้วัคซีน
ข้อมูลที่ไม่ทำ SMOTE	3165	616	187
ข้อมูลทำ SMOTE	3165	3080	2992

#### 3.2 การสร้างตัวแบบจำลองการพยากรณ์

การสร้างตัวแบบจำลองการพยากรณ์ได้ใช้งาน  
 โปรแกรม MATLAB โดยมีขั้นตอนการดำเนินการดังนี้  
 ใช้ตัวแบบ MLP ในการทำงาน โดยใช้ขั้นตอนวิธีทั้ง 4  
 ขั้นตอนวิธี ในกาค่าน้ำหนักของ MLP การเปรียบเทียบ  
 ขั้นตอนฝึกสอนทั้ง 4 วิธี กำหนดให้ Max Run = 30,  
 Search Agents No = 30, Max NFE = 3000, Lower  
 bounds ของค่าน้ำหนัก = -10, Upper bounds ของค่า  
 น้ำหนัก = 10 เท่ากันทั้ง 4 ขั้นตอนวิธี โดยจะนำค่าที่ได้ไป  
 ใช้ร่วมกับ MLP เพื่อหาค่าของความแม่นยำในการ  
 ทำนายข้อมูล

#### 3.3 การวัดประสิทธิภาพของตัวแบบการพยากรณ์

การทดสอบความแม่นยำของตัวแบบการพยากรณ์จะ  
 ทำการทดสอบความแม่นยำด้วยวิธีการสุ่มหาค่าที่ดีที่สุด  
 ของกลุ่มข้อมูลที่มีอยู่ ด้วยการวัดประสิทธิภาพด้วย  
 Accuracy [9] ด้วยสมการดังนี้:

$$Accuracy = \frac{Correct Prediction}{Total Prediction} \quad (16)$$

โดยที่ *Correct Prediction* คือ ค่าทำนายที่ถูกต้อง  
*Total Prediction* คือ ค่าทำนายทั้งหมด

และให้นำเอา Geometric Mean มาใช้ในส่วนของ *Correct Prediction* เพื่อทำนายค่าความแม่นยำได้ดีขึ้น และหาค่าความผิดพลาดด้วยสมการดังนี้:

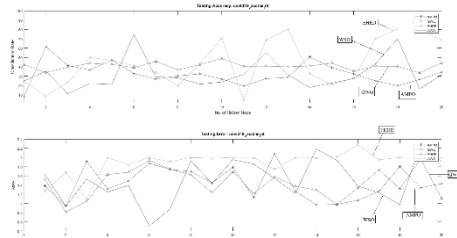
$$Error = \frac{Incorrect Prediction}{Total Prediction} \quad (17)$$

โดยที่ *Incorrect Prediction* คือค่าทำนายที่ไม่ถูกต้อง

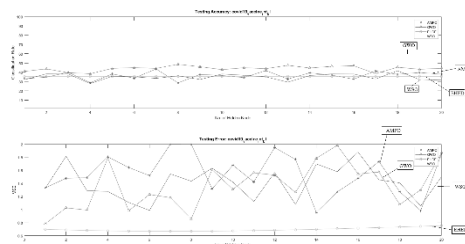
#### 4. ผลการดำเนินงาน

จากการดำเนินงานข้อมูลที่ถูกจัดให้นำมาใช้ทดสอบจะแบ่งออกเป็น 4 ประเภท คือข้อมูลนำเข้าที่เป็น Integer ข้อมูลนำเข้าที่เป็น Binary ข้อมูลที่เป็น Category ที่ทำการ SMOTE ข้อมูลแล้ว และข้อมูล Binary ที่ทำการ SMOTE ข้อมูลแล้ว จากภาพที่ 1 ภาพที่ 2 ภาพที่ 3 ภาพที่ 4 แนวแกน x คือ จำนวนโหนดในชั้นซ่อน (1-20) แนวแกน y คือ ค่า Accuracy (รูปบน) และ ค่า MSE (รูปล่าง) ข้อมูลที่เป็น Category เมื่อผ่านขั้นตอนวิธีทั้ง 4 ขั้นตอนวิธี จะเห็นได้ว่าตัวแบบจำลองที่มีค่าความแม่นยำที่ดีที่สุดคือตัวแบบจำลอง EHDE มีค่าความถูกต้อง 68.4719 เปอร์เซ็นต์ ตัวแบบจำลอง GWO มีค่าความถูกต้อง 44.2146 เปอร์เซ็นต์ ตัวแบบจำลอง AMPO มีค่าความถูกต้อง 34.2863 เปอร์เซ็นต์ ตัวแบบจำลอง WSO มีค่าความถูกต้อง 31.3266 เปอร์เซ็นต์ ตามลำดับ ดังรูปภาพที่ 1 ข้อมูล Category ที่ทำการ SMOTE ข้อมูลค่าความแม่นยำของตัวแบบจำลองจะมีค่าใกล้เคียงกันและตัวแบบจำลองที่มีค่าความแม่นยำที่สุดคือตัวแบบจำลอง GWO มีค่าความถูกต้อง 44.3223 ตัวแบบจำลอง AMPO มีค่าความถูกต้อง 37.9083 เปอร์เซ็นต์ ตัวแบบจำลอง WSO มีค่าความถูกต้อง 35.5712 เปอร์เซ็นต์ ตัวแบบจำลอง EHDE มีค่าความถูกต้อง 31.6513 เปอร์เซ็นต์ ตามลำดับ ดังภาพที่ 2 ข้อมูล Binary เมื่อผ่านขั้นตอนวิธีทั้ง 4 ขั้นตอนวิธีแล้ว จะเห็นได้ว่าแบบจำลองที่มีค่าแม่นยำที่ดีที่สุดคือตัวแบบจำลอง EHDE มีค่าความถูกต้อง 90.8228 เปอร์เซ็นต์ ตัวแบบจำลอง GWO เปอร์เซ็นต์ มีค่าความถูกต้อง 80.4786 เปอร์เซ็นต์ ตัวแบบจำลอง AMPO มีค่าความถูกต้อง 31.4247 เปอร์เซ็นต์ และตัวแบบจำลอง WSO มีค่าความถูกต้อง 22.0823

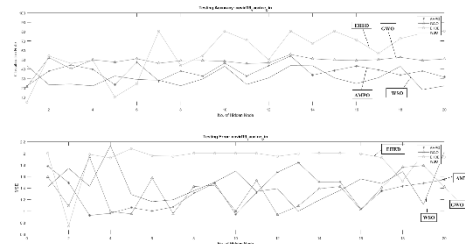
เปอร์เซ็นต์ เรียงตามลำดับ ดังรูปภาพที่ 3 ส่วนข้อมูล Binary ที่ทำการ SMOTE ข้อมูล จะเห็นได้ว่าตัวแบบจำลองมีความใกล้เคียงกัน ตัวแบบจำลองที่มีค่าความแม่นยำที่ดีที่สุดคือตัวแบบจำลอง GWO มีค่าความถูกต้อง 49.8087 เปอร์เซ็นต์ ตัวแบบจำลอง AMPO มีค่าความถูกต้อง 41.3914 เปอร์เซ็นต์ ตัวแบบจำลอง WSO มีค่าความถูกต้อง 39.5596 เปอร์เซ็นต์ ตัวแบบจำลอง EHDE มีค่าความถูกต้อง 35.5712 เปอร์เซ็นต์ ตามลำดับดังรูปภาพที่ 4 และค่า Error ของข้อมูลทั้ง 2 ประเภท เกิดขึ้นเนื่องจากข้อมูลแต่ละ Attribute แตกต่างกันไป ส่งผลให้ค่า Error ที่แสดงในกราฟเกิดการ Overfitting



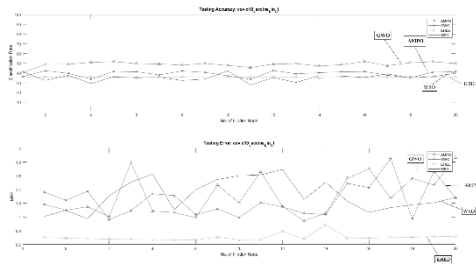
ภาพที่ 1: ข้อมูลที่แบ่งเป็น Category ที่ยังไม่ทำการ SMOTE ข้อมูล



ภาพที่ 2: ข้อมูลที่แบ่งเป็น Category ที่ทำการ SMOTE ข้อมูลแล้ว



ภาพที่ 3: ข้อมูลแบบ Binary ที่ยังไม่ทำการ SMOTE ข้อมูล



ภาพที่ 4: ข้อมูลแบบ Binary ที่ทำการ SMOTE ข้อมูล

## 5. สรุป

ในการทดลองครั้งนี้ผู้วิจัยได้ใช้ข้อมูลจริงมาทำการทดสอบ โดยธรรมชาติของข้อมูลจริงสิ่งที่ผิดปกติจะมีน้อยกว่าสิ่งที่ปกติคือ ความรุนแรงระดับที่ 3 การแพ้วัคซีนจะน้อยกว่าความรุนแรงระดับ 1 อาการปกติที่พบได้มากกว่าหลายเท่าทำให้ข้อมูลเกิดการไม่สมดุล (Imbalanced data) ทำให้เกิดผลกระทบกับตัวแบบจำลองที่ใช้ ผู้วิจัยได้ใช้วิธี SMOTE ข้อมูล เพื่อแก้ไขข้อมูลไม่สมดุล โดยทำการสังเคราะห์ข้อมูลกลุ่มน้อยให้มีข้อมูลที่ใกล้เคียงกันมากขึ้น จากการผ่านกระบวนการทั้งหมดจะแสดงได้ว่า ตัวแบบจำลอง GWO ที่ใช้ข้อมูล Binary มีความแม่นยำมากที่สุด

งานวิจัยนี้ตัวแบบจำลองที่ใช้สามารถนำไปพัฒนาตัวแบบจำลองให้มีการจำแนกข้อมูลที่ดีกว่านี้ ถึงแม้ว่าตัวแปรที่ใช้ในการพัฒนาจะน้อยมากก็ตามแต่ถ้าข้อมูลมีความสมดุลตัวแบบจำลองอาจจะมีประสิทธิภาพของความแม่นยำในการทำนายสูงขึ้น

## เอกสารอ้างอิง

- [1] M. Braik, A. Hammouri, J. Atwan, M. A. Al-Betar, & M. A. Awadallah. "White Shark Optimizer: A novel bio-inspired meta-heuristic algorithm for global optimization problems." *Knowledge-Based Systems*, 108457, 2022.
- [2] N. V. Chawla, K. W. Bowyer, L. O. Hall, & W. P. Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research*, Vol. 16, pp. 321-357, 2002.
- [3] A. F. Hernández, D. Calina, K. Poulas, A. O. Docea, & A. M. Tsatsakis. "Safety of COVID-19 vaccines administered in the EU: Should we be

concerned?." *Toxicology Reports*, Vol. 8, pp. 871-879, 2021.

- [4] Z. Li, V. Tam, & L. K. Yeung. "An Adaptive Multi-Population Optimization Algorithm for Global Continuous Optimization." *IEEE Access*, Vol. 9, pp. 19960-19989, 2021.
- [5] S. Mirjalili, S. M. Mirjalili, & A. Lewis. "Grey wolf optimizer." *Advances in engineering software*, Vol. 69, pp. 46-61, 2014.
- [6] P. K. Simpson, "Fuzzy min-max neural networks-part 1: classification," *IEEE Trans. Neural Networks*, vol. 3, no. 5, pp. 776-786, 1992.
- [7] S. Mirjalili. "How effective is the Grey Wolf optimizer in training multi-layer perceptrons." *Appl Intell*, Vol. 43, pp. 150-161, 2015.
- [8] S. Wu, and T. W. S. Chow, "Induction machine fault detection using SOM-based RBF neural networks." *IEEE Trans. on Industrial Electronics*, Vol. 51, No. 1, pp. 183-194, 2004.
- [9] J. Brownlee. Tour of Evaluation Metrics for Imbalanced Classification. Available Online at <https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification>.
- [10] World Health Organization. COVID-19 Vaccines Advice. Available Online at <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines/advice>.
- [11] X. Zhong, & P. Cheng. "An elite-guided hierarchical differential evolution algorithm." *Applied Intelligence*, Vol. 51, No. 7, pp. 4962-4983, 2021.

# การพัฒนาระบบธุรกิจอัจฉริยะ เพื่อพยากรณ์พฤติกรรมของประชากรวัยทำงาน ที่ส่งผลให้เกิดโรคไม่ติดต่อเรื้อรัง

## The Development of Business Intelligence to Predict the Behaviors of the Working-Age Population Resulting in Chronic Non-Communicable Diseases

สุภลักษณ์ เจริญธรรม (Suphaluck Charoenthum)<sup>1</sup> และมณฑิธร รัตนศิริวงศ์วุฒิ (Montean Rattanasiriwongwut)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>S6307021856055@email.kmutnb.ac.th, <sup>2</sup>montean.r@itd.kmutnb.ac.th

### บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมที่เหมาะสมในการวิเคราะห์พฤติกรรมที่มีแนวโน้มส่งผลให้เกิดโรคความดันโลหิต โรคไขมันในเส้นเลือด , โรคไต โรคเบาหวานและโรคมะเร็ง 2) เพื่อพัฒนาระบบธุรกิจอัจฉริยะเพื่อสนับสนุนการวิเคราะห์ความสัมพันธ์ของพฤติกรรมประชากรวัยทำงานและการช่วยตัดสินใจ งานวิจัยนี้ผู้วิจัยมีแนวคิดในการนำข้อมูลจากกรมอนามัย กระทรวงสาธารณสุข ที่มีการเก็บแบบสอบถามประเมินพฤติกรรมสุขภาพ ความรอบรู้ และปัจจัยแวดล้อมที่ส่งผลต่อสุขภาพของประชากรวัยทำงานจำนวน 2,520 รายการ เพื่อสร้างแบบจำลองในการวิเคราะห์พฤติกรรมความเสี่ยงของประชากรวัยทำงานที่มีแนวโน้มส่งผลให้เกิดโรคความดันโลหิตสูง โรคหัวใจและหลอดเลือด ไขมันในเลือดสูง และโรคเบาหวาน โดยเปรียบเทียบระหว่างอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine), อัลกอริทึมเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor), อัลกอริทึมโครงข่ายประสาทเทียม (Neural Network), อัลกอริทึมต้นไม้ป่าสุ่ม (Random Forest) และอัลกอริทึมนาอิวเบย์ (Naïve Bays) เพื่อพยากรณ์พฤติกรรมในชีวิตประจำวันที่มีแนวโน้มส่งผลให้เกิดโรคในกลุ่มโรคไม่ติดต่อเรื้อรัง(NCDs) จากนั้นนำข้อมูลที่มาแสดงผลรายงานในรูปแบบ Power BI Desktop โดยในการทดสอบประสิทธิภาพครั้งนี้พบว่าอัลกอริทึมที่ให้ประสิทธิภาพสูง

ที่สุดคือ อัลกอริทึมต้นไม้ป่าสุ่ม (Random Forest) โดยมีค่าความถูกต้อง (Accuracy) ร้อยละ 94.07

คำสำคัญ: พฤติกรรมสุขภาพ ประชากรวัยทำงาน โรคไม่ติดต่อเรื้อรัง

### Abstract

This research has subjects to 1) To Compare Efficiency of Algorithm which are suitable to analyze trend of behaviors to cause Hypertension, Hyperlipidemia, Nephropathy, Diabetes and Cancer. 2) To develop intelligence business system to support analysis of relevant between working age population's behaviors and decision making. This research has used database from department of Health, Ministry of Public Health which did the survey of working age population about Health Behaviors, knowledge, and environmental factors affected to working age population's health 2,520 record to make simulation model that can be analyze behavior risk incidence of Hypertension, Hyperlipidemia, Nephropathy, Diabetes and Cancer. The research have compared Algorithms such as Support Vector Machine, K-Nearest Neighbor, Neural Network, Random forest and Naïve Bays to forecast behaviors in everyday life which trend to be incident risk of non-communicable diseases (NCDs) and then use Power BI Desktop to make reports which found the most effective Algorithm is Random Forest (Accuracy Rate 94.07%).

**Keywords:** Healthy Behavior, Working Age Population, Chronic non-communicable disease

## 1. บทนำ

ความรู้ด้านสุขภาพ หรือ Health Literacy คือ ทักษะในการเข้าใจในเรื่องของสุขภาพทั้งในเรื่องของการดูแลสุขภาพ การปฏิบัติตนเมื่อเจ็บป่วย และยังรวมถึงการใช้ข้อมูลข่าวสารในด้านสุขภาพเพื่อส่งเสริมและคงไว้ซึ่งสุขภาพที่ดี เพื่อก่อให้เกิดแรงจูงใจในการดูแลสุขภาพสุขภาพของตนเอง ครอบครัวและบุคคลใกล้ชิด จากผลการศึกษาที่ผ่านมาพบว่าประชากรโดยส่วนใหญ่มีความรอบรู้ด้านสุขภาพต่ำ[1] เมื่อมีความรอบรู้ด้านสุขภาพต่ำทำให้ไม่รู้วิธีการรับมือเมื่อตนเองหรือบุคคลใกล้ชิดเจ็บป่วย และยังมีปัจจัยเสี่ยงต่อการมีสุขภาพที่ไม่ดีอันเป็นผลอันเนื่องมาจากการไม่มีความรู้เพียงพอที่จะสามารถดูแลตนเองเพื่อป้องกันพฤติกรรมเสี่ยงที่อาจก่อให้เกิดโรคได้อย่างเหมาะสม กล่าวคือความรู้ด้านสุขภาพเป็นปัจจัยสำคัญที่มีผลต่อมุมมองของการดูแลสุขภาพสุขภาพ

พฤติกรรมเสี่ยงของประชากรวัยทำงาน เป็นพฤติกรรมทางสุขภาพใด ๆ ของบุคคล ที่เมื่อปฏิบัติแล้วก่อให้เกิดผลเสียต่อสุขภาพ ซึ่งนำไปสู่การเจ็บป่วยต่อตนเอง ครอบครัว หรือบุคคลใกล้ชิด พฤติกรรมที่เคซินจะทำให้ติดจนเป็นนิสัยโดยไม่รู้ตัว จนก่อนให้เกิดโรคประจำตัวต่าง ๆ ได้ องค์การอนามัยโลก (World Health Organization : WHO) กล่าวว่า จากสถิติปัจจุบันสาเหตุการเสียชีวิตจากทั่วโลกร้อยละเกือบ 70 มีสาเหตุมาจากการเสียชีวิตจากโรคในกลุ่มโรคไม่ติดต่อเรื้อรัง(NCDs) และร้อยละ 82 ของ 16 ล้านคนที่เสียชีวิตก่อนวัยอันควรหรือก่อนอายุครบ 70 ปี[2] เห็นได้ว่าในปัจจุบันปัญหาด้านสุขภาพถือเป็นปัญหาอันดับหนึ่งของโลก โดยสาเหตุของการเพิ่มขึ้นของโรคไม่ติดต่อเรื้อรังเกิดจากปัจจัยต่าง ๆ จากพฤติกรรมเสี่ยงหลักสี่ประการ ได้แก่ การสูบบุหรี่ การไม่ออกกำลังกาย การดื่มเครื่องดื่มแอลกอฮอล์อย่างเป็นอันตราย และพฤติกรรมการรับประทานอาหารที่ไม่ดีต่อสุขภาพ[2]

จากข้อมูลข้างต้นผู้วิจัยมีความสนใจในการศึกษาและวิเคราะห์พฤติกรรมเสี่ยงของประชากรวัยทำงานที่มี

แนวโน้มส่งผลให้เกิดโรคในกลุ่มโรคไม่ติดต่อเรื้อรัง (NCDs) โดยเฉพาะ 3 กลุ่มโรค ต่อไปนี้ โรคความดันโลหิต โรคไขมันในเส้นเลือด , โรคไต โรคเบาหวานและโรคมะเร็ง เนื่องจากเป็นโรคหลักของสาเหตุการเสียชีวิตอันเกิดจากพฤติกรรมความเสี่ยงที่ไม่พึงประสงค์ และเป็นโรคที่มีแนวโน้มการเกินสูงขึ้นทุกปี ถึงแม้จากค่าสถิติการป่วยและการเสียชีวิตจากโรคในกลุ่มโรคไม่ติดต่อเรื้อรัง (NCDs) จะอยู่ในระดับที่สูงมาก แต่โรคในกลุ่มโรคนี้สามารถป้องกันได้ จากการปรับพฤติกรรมเสี่ยงต่าง ๆ เพราะสาเหตุหลักของการเกิดโรคในกลุ่มโรคไม่ติดต่อเรื้อรัง (NCDs) เกิดจากพฤติกรรมความเสี่ยงของเรา นั่นเอง อาทิ การรับประทานอาหารที่มีรสหวานมันเค็มจัด การขาดการออกกำลังกาย ความเครียดสะสม การดื่มเครื่องดื่มแอลกอฮอล์ สูบบุหรี่ พักผ่อนไม่เพียงพอ ซึ่งหากสามารถลดพฤติกรรมเสี่ยงเหล่านี้ได้ ก็สามารถลดโอกาสในการเป็นโรคในกลุ่มโรคไม่ติดต่อเรื้อรัง (NCDs) ได้มากถึงร้อยละ 80 และยังคงลดโอกาสในการเป็นโรคมะเร็งได้มากถึงร้อยละ 40 ลดโอกาสเป็นโรคหลอดเลือดและหัวใจ รวมถึงโรคเบาหวาน ประเภทที่ 2 ได้มากถึงร้อยละ 80[3] ดังนั้น เป็นสิ่งที่สำคัญมากที่ต้องให้ประชากรได้มีความรอบรู้ด้านสุขภาพและตระหนักถึงความสำคัญของการรับรู้เข้าใจ และรู้การปฏิบัติตนเพื่อหลีกเลี่ยงพฤติกรรมเสี่ยงที่จะส่งผลให้เกิดโรคเหล่านี้

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) คือ ขั้นตอนในการกรองข้อมูล เป็นการวิเคราะห์ข้อมูลจำนวนมาก เพื่อค้นหา รูปแบบและความสัมพันธ์ ที่ซ่อนอยู่ในชุดข้อมูล[4] เพื่อให้ข้อมูลกลายเป็นข้อมูลสารสนเทศที่สามารถนำไปใช้งานได้หรือเพื่อให้ได้องค์ความรู้ใหม่ ที่สามารถนำไปใช้ประกอบการตัดสินใจในด้านต่าง ๆ

### 2.2 การทำความสะอาดข้อมูล (Data Cleansing)

เป็นกระบวนการตรวจสอบและแก้ไข รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ที่มีความไม่สอดคล้องกันของข้อมูล อาจเกิดจากข้อผิดพลาดของการบันทึกข้อมูล

หรือการให้ความหมายของข้อมูลที่จัดเก็บแตกต่างกัน เช่น คีย์ข้อมูลเข้ามาผิด การเว้นว่างไม่กรอกข้อมูล หรือเป็นตัวเลขที่ไม่มีทางเป็นจริงได้ในความเป็นจริง ดังนั้น จึงต้องมีการทำความสะอาดข้อมูลเพื่อให้ได้ข้อมูลที่สมบูรณ์ขึ้น และสามารถนำไปใช้งานต่อไปได้

### 2.3 เทคนิคซ์พอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

เป็นอัลกอริทึม แบบ Supervised Learning ที่สามารถนำมาช่วยแก้ปัญหาในการจำแนกข้อมูลโดยมีการใช้อย่างกว้างขวาง ซึ่งใช้ในการวิเคราะห์และจำแนกข้อมูล โดยอาศัยหลักการทางคณิตศาสตร์ในการหาสัมประสิทธิ์ของสมการเพื่อมาหาเส้นแบ่งกลุ่มที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเลือกเส้นแบ่งที่แบ่งแยกกลุ่มข้อมูลได้ดีที่สุด โดยไม่จำเป็นต้องเป็นเส้นตรงเสมอไป

### 2.4 เพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor)

เป็นการแบ่งคลาสจากการหาระยะห่างระหว่างแต่ละตัวแปร ใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงกันมากน้อยเพียงใด โดยหาผลรวมของจำนวนเงื่อนไขหรือกรณีต่าง ๆ สำหรับแต่ละคลาสและกำหนดเงื่อนไขใหม่ ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด

### 2.5 อัลกอริทึมโครงข่ายประสาทเทียม (Neural Network)

โครงข่ายประสาทเทียม เป็นสาขาหนึ่งของปัญญาประดิษฐ์ Artificial Intelligence (AI) เป็นแนวคิดที่ออกแบบระบบโครงข่ายคอมพิวเตอร์ ให้เรียนแบบการทำงานของสมองมนุษย์ โครงข่ายประสาทเทียมสามารถเรียนรู้ที่จะทำงานที่มอบหมายได้ จากการเรียนรู้ผ่านตัวอย่าง โดยไม่ถูกโปรแกรมด้วยกฎเกณฑ์ตายตัวแบบระบบอัตโนมัติ

### 2.6 ต้นไม้ป่าสุ่ม (Random Forest)

ต้นไม้ป่าสุ่ม เป็นอัลกอริทึมการเรียนรู้ของเครื่องที่มีความยืดหยุ่นและใช้งานง่าย เป็นอีกหนึ่งอัลกอริทึมที่ใช้กันอย่างแพร่หลาย สามารถใช้สำหรับจำแนกประเภทและงานถดถอย โดยหลักการของต้นไม้ป่าสุ่ม คือ จะสร้าง

ต้นไม้ตัดสินใจหลาย ๆ แบบและรวมเข้าด้วยกันเพื่อให้ได้การคาดการณ์ที่แม่นยำและเสถียรยิ่งขึ้น

### 2.7 นาอิว เบย์ (Naïve Bayes)[5]

นาอิว เบย์ มีหลักการตามความน่าจะเป็นแบบมีเงื่อนไขมาพัฒนาตามทฤษฎีของเบย์ ซึ่งมีอัลกอริทึมที่ไม่ซับซ้อน เป็นขั้นตอนในการจำแนกข้อมูล โดยอาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละสมมติฐาน หรือก็คือคลาสเป้าหมายหรือผลลัพธ์การทำนาย โดยจะทำการการสร้างเงื่อนไขของความน่าจะเป็นของแต่ละความสัมพันธ์ จากการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร เหมาะกับข้อมูลที่มีจำนวนมากและมีคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน

### 2.8 Cross-validation Test

เป็นวิธีที่นิยมนำมาใช้ในการทดสอบประสิทธิภาพของอัลกอริทึม เนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพของ Cross-validation จะเป็นการแบ่งข้อมูลออกเป็นหลายส่วน (K) โดยแต่ละส่วนจะมีจำนวนข้อมูลเท่ากัน เช่น แบ่งข้อมูลออกเป็น 10-fold Cross-validation ตัว Cross-validation จะทำการแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กัน จากนั้นข้อมูลส่วนหนึ่งจะเป็นตัวทดสอบประสิทธิภาพของอัลกอริทึม และจะเป็นการวนการทำงานแบบนี้ไปเรื่อย ๆ จนครบ 10 ส่วนตามที่กำหนด

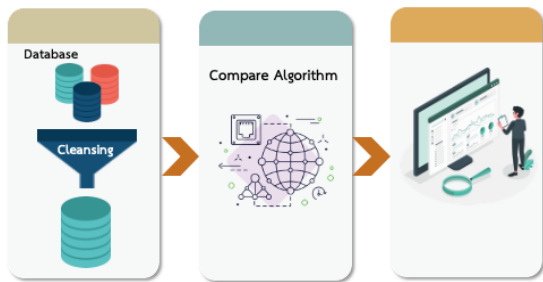
### 2.9 งานวิจัยที่เกี่ยวข้อง

ณัฐธิดา และวัชรวิวัฒน์ [4] นำเสนองานวิจัยเกี่ยวกับการวิเคราะห์ปัจจัยและพยากรณ์การโอนของข้าราชการเพื่อหาแนวโน้มการโอนของข้าราชการ โดยการพยากรณ์แนวโน้มด้วยการทำเหมืองข้อมูลโดยใช้เทคนิคการวิเคราะห์ 3 เทคนิค คือ ต้นไม้ตัดสินใจ เบย์เขียนแบบง่าย และเพื่อนบ้านใกล้ที่สุด ผลการวิจัยพบว่า ผลการทดสอบด้วยตัวแบบพยากรณ์ต้นไม้ตัดสินใจ ให้ค่าความถูกต้องในการพยากรณ์สูงที่สุดที่ร้อยละ 85.42

นภาพร [6] นำเสนองานวิจัยเกี่ยวกับการพัฒนาโมเดลสำหรับการวินิจฉัยการเกิดโรคไส้ติ่งอักเสบ และศึกษาปัจจัยที่ส่งผลต่อการทำนายการเกิดโรคไส้ติ่งอักเสบทั้งด้านประชากรศาสตร์ ด้านพฤติกรรม และด้านอาการ เพื่อนำไปใช้ในการช่วยการวินิจฉัยโรคที่รวดเร็วขึ้นและช่วย

ในการตัดสินใจ ผู้วิจัยใช้เทคนิคการเรียนรู้ของเครื่องในการวิเคราะห์ 6 เทคนิค คือ การทอดอยโลจิสติก, นาอ์ฟเบย์, ต้นไม้ตัดสินใจ, เพื่อนบ้านที่ใกล้ที่สุด, ซัพพอร์ตเวกเตอร์แมชชีน และมัลติเลเยอร์เพอร์เซ็ปตรอน ผลการวิจัยพบว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ประสิทธิภาพสูงสุด ซึ่งมีค่า Accuracy ร้อยละ 95, Precision ร้อยละ 94 , Recall ร้อยละ 94, และ F1-score ร้อยละ 94

### 3. วิธีดำเนินการวิจัย



ภาพที่ 1 : กรอบแนวคิดในการพยากรณ์พฤติกรรมของประชากรวัยทำงานที่ส่งผลให้เกิดโรคไม่ติดต่อเรื้อรัง

จากภาพที่ 1 แสดงกรอบแนวคิดของงานวิจัย งานวิจัยครั้งนี้เป็นงานวิจัยเชิงทดลอง เป็นการวิเคราะห์หาอัลกอริทึมที่ให้ประสิทธิภาพที่ดีที่สุดและแม่นยำที่สุดในการพยากรณ์ โดยผู้ทำวิจัยได้นำข้อมูลพฤติกรรมของประชากรวัยทำงาน จากกรมอนามัย กระทรวงสาธารณสุขจำนวน 2,520 รายการ โดยแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ข้อมูลกลุ่มศึกษาจำนวน 2,120 รายการ กลุ่มข้อมูลชุดทดสอบจำนวน 400 รายการ มาใช้ในการวิเคราะห์และพยากรณ์พฤติกรรมที่ส่งผลให้เกิดโรคไม่ติดต่อเรื้อรัง จากนั้นนำข้อมูลมาทำความสะอาด(Data Cleansing) เมื่อทำความสะอาดข้อมูลเรียบร้อยแล้วนำข้อมูลที่เข้ามาเข้าอัลกอริทึมเพื่อหาอัลกอริทึมที่ดีที่สุด จากนั้นนำมาแสดงผลภาพรวมของข้อมูลโดยใช้ Power BI Desktop ในการแสดงผล สามารถแบ่งขั้นตอนการดำเนินงานออกเป็น 4 ขั้นตอน ดังนี้

#### 3.1 การศึกษาและเปรียบเทียบเทคนิคของอัลกอริทึม

ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องและศึกษาอัลกอริทึม 5 อัลกอริทึม ได้แก่ อัลกอริทึมซัพพอร์ต

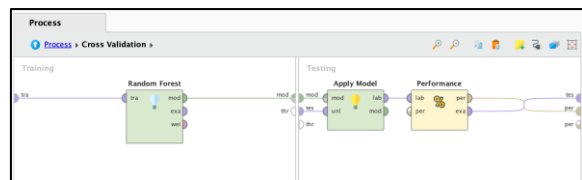
เวกเตอร์แมชชีน (Support Vector Machine), อัลกอริทึมเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor), อัลกอริทึมโครงข่ายประสาทเทียม (Neural Network), อัลกอริทึมต้นไม้ป่าสุ่ม (Random Forest) และอัลกอริทึมนาอ์ฟเบย์ (Naive Bays) มาใช้ในการเปรียบเทียบเพื่อหาอัลกอริทึมที่ดีที่สุด ในการเรียนรู้พฤติกรรมของประชากรวัยทำงาน

#### 3.2 การเตรียมข้อมูลและคัดเลือกข้อมูล

ข้อมูลที่นำมาใช้ในงานวิจัยครั้งนี้ เป็นข้อมูลที่รวบรวมจากแบบสอบถามพฤติกรรมอันพึงประสงค์ของประชากรวัยทำงาน สังกัดสำนักงานปลัดกระทรวงสาธารณสุข อยู่ในช่วงวันที่ 1 ตุลาคม พ.ศ.2563 จนถึงวันที่ 31 กันยายน พ.ศ.2564 ซึ่งเป็นข้อมูลเกี่ยวกับพฤติกรรมในชีวิตประจำวันของประชากรวัยทำงาน ได้แก่ จังหวัด อำเภอ อาชีพ เพศ อายุ ส่วนสูง น้ำหนัก เส้นรอบเอว โรคประจำตัว พฤติกรรมการทานอาหารและการปรุงอาหาร พฤติกรรมการออกกำลังกายและการพักผ่อน พฤติกรรมการสูบบุหรี่และการดื่มเครื่องดื่มแอลกอฮอล์

#### 3.3 ขั้นตอนการสร้างโมเดลพยากรณ์ เปรียบเทียบ และคัดเลือกเทคนิคการพยากรณ์ที่เหมาะสมที่สุดไปใช้งาน

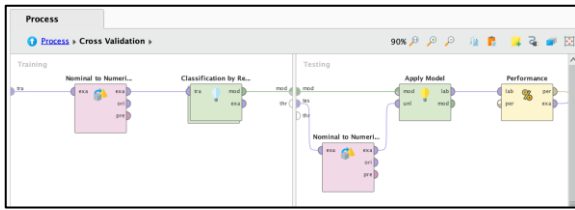
อัลกอริทึมต้นไม้ป่าสุ่ม (Random Forest) โดยใช้รูปแบบ Information gain ปรับค่าพารามิเตอร์ maximal depth = 11, confidence = 0.2, minimal gain = 0.03, minimal leaf size = 2, minimal size for split = 5 และ number of prepruning alternatives = 3 ได้ค่าความถูกต้อง (Accuracy) ร้อยละ 94.07 และมีค่าสัมประสิทธิ์ตัวชี้วัดทางสถิติ(kappa)ที่ 0.921



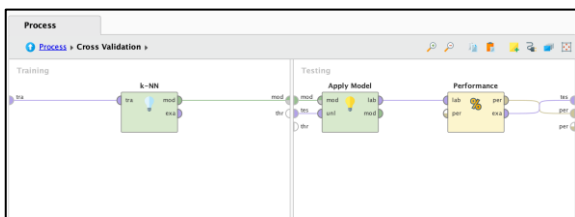
ภาพที่ 2 : แสดงรูปขั้นตอนการหาประสิทธิภาพของอัลกอริทึม(1)

อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) โดยเลือก Kernel type เป็น anova ปรับค่าพารามิเตอร์ Kernel Gramma = 1.0, Kernel Degree = 2.2 ได้ค่าความถูกต้อง(Accuracy) ร้อยละ 91.90 และมีค่าสัมประสิทธิ์ตัวชี้วัดทางสถิติ(kappa)ที่ 0.892

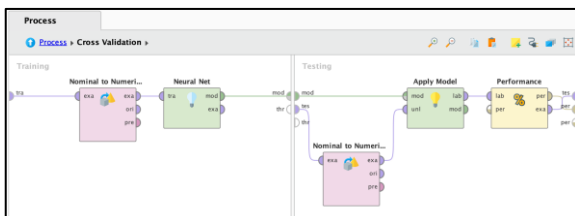




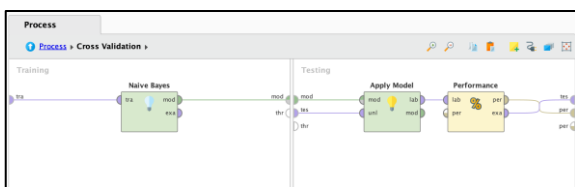
ภาพที่ 3 : แสดงรูปขั้นตอนการหาประสิทธิภาพของอัลกอริทึม(2)  
อัลกอริทึมเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) ปรับค่าพารามิเตอร์ K = 13 ได้ค่าความถูกต้อง (Accuracy) ร้อยละ 47.69 และมีค่าสัมประสิทธิ์ตัวชี้วัดทางสถิติ(kappa)ที่ 0.303



ภาพที่ 4 : แสดงรูปขั้นตอนการหาประสิทธิภาพของอัลกอริทึม(3)  
อัลกอริทึมโครงข่ายประสาทเทียม (Neural Network) ปรับค่าพารามิเตอร์ training cycles = 200, learning = 0.02, momentum = 0.8 ได้ค่าความถูกต้อง (Accuracy) ร้อยละ 92.56 และมีค่าสัมประสิทธิ์ตัวชี้วัดทางสถิติ(kappa)ที่ 0.901



ภาพที่ 5 : แสดงรูปขั้นตอนการหาประสิทธิภาพของอัลกอริทึม(4)  
อัลกอริทึมนาอิวเบย์ (Naïve Bays) ได้ค่าความถูกต้อง(Accuracy) ร้อยละ 89.17 และมีค่าสัมประสิทธิ์ตัวชี้วัดทางสถิติ(kappa)ที่ 0.856



ภาพที่ 6 : แสดงรูปขั้นตอนการหาประสิทธิภาพของอัลกอริทึม(5)  
งานวิจัยฉบับนี้ ผู้วิจัยได้ทำการพยากรณ์พฤติกรรมที่ส่งผลให้เกิดโรคไม่ติดต่อเรื้อรัง โดยใช้อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine), อัลกอริทึมเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor),

อัลกอริทึมโครงข่ายประสาทเทียม (Neural Network) , อัลกอริทึมต้นไม้ป่าสุ่ม (Random Forest) และอัลกอริทึมนาอิวเบย์ (Naïve Bays) มาใช้ในการเปรียบเทียบเพื่อหาอัลกอริทึมที่ดีที่สุดด้วยโปรแกรม RapidMiner Studio โดยเริ่มจากรวบรวมข้อมูล ทำความสะอาดข้อมูลที่ได้ และแบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลฝึกฝน (Training Data) ใช้ในการฝึกฝนตัวแบบพยากรณ์ และสร้างตัวแบบของแต่ละวิธีออกมา และข้อมูลทดสอบ (Test Data) หลังจากนั้นนำข้อมูลที่ได้มาใช้เป็นข้อมูลในการพยากรณ์ และสร้างแบบจำลองเมื่อทำการพยากรณ์แต่ละเทคนิคเรียบร้อยแล้ว ได้ทำการสรุปผลและเลือกเทคนิคที่เหมาะสมและแม่นยำที่สุด โดยวัดประสิทธิภาพจากความแม่นยำ(Accuracy) และค่าสัมประสิทธิ์ตัวชี้วัด(kappa) ตามสมการในที่แสดงในภาพที่ 7 และ ภาพที่ 8

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

ภาพที่ 7 : แสดงรูปสมการในการคำนวณหาค่า Accuracy

$$Cohen's\ kappa = \frac{N \sum_{i=1}^m CM_{ii} - \sum_{i=1}^m C_{i\text{corr}} C_{i\text{pred}}}{N^2 - \sum_{i=1}^m C_{i\text{corr}} C_{i\text{pred}}}$$

ภาพที่ 8 : แสดงรูปสมการในการคำนวณหาค่า kappa

3.4 ขั้นตอนการแสดงผลและแสดงภาพรวมของข้อมูล

ขั้นตอนนี้เป็นการนำข้อมูลที่ได้จากการพยากรณ์มาแสดงผลรายงานในรูปแบบต่าง ๆ เพื่อนำไปใช้งานแผนและกำหนดกลยุทธ์ ช่วยในการตัดสินใจอย่างถูกต้อง



ภาพที่ 9 : แสดงรูปการนำออกแบบหน้ารายงานแสดงผล

### 4. ผลการดำเนินงาน

ผลการพัฒนาระบบธุรกิจอัจฉริยะ เพื่อพยากรณ์พฤติกรรมของประชากรวัยทำงานที่ส่งผลให้เกิดโรคไม่ติดต่อเรื้อรัง มารดแบ่งหัวข้อได้ดังนี้

#### 4.1 ผลการเปรียบเทียบอัลกอริทึม

ทำการเปรียบเทียบอัลกอริทึม โดยนำชุดข้อมูลตัวอย่างมาเพื่อทดสอบประสิทธิภาพของอัลกอริทึม ตัวอย่างชุดข้อมูลแสดงดังตารางที่ 1

ตารางที่ 1 : ตัวอย่างข้อมูลที่นำมาใช้ในการวิเคราะห์

ที่	ข้อมูลที่นำมาใช้ในการวิเคราะห์	
1	province_key	จังหวัด
2	disease_key	โรคประจำตัว
3	height	ส่วนสูง
4	age_key	อายุ
5	sex_key	เพศ
6	waistline	เส้นรอบเอว
7	weight	น้ำหนัก
8	career_key	อาชีพ
9	algho_key	พฤติกรรมการดื่มเครื่องดื่มแอลกอฮอล์
10	sic_key	พฤติกรรมการสูบบุหรี่
11	exercise_key	พฤติกรรมการออกกำลังกาย
12	sleep_key	พฤติกรรมการนอน
13	loll_key	พฤติกรรมการนั่งหรือเอนกายติดต่อกันนาน 2 ชั่วโมงขึ้นไป โดยไม่รวมการนอนหลับ คิดเป็นกัวันใน 1 สัปดาห์
14	salty_key	พฤติกรรมการทานอาหารรสเค็มหรือการปรุงเค็ม
15	sweet_key	พฤติกรรมการดื่มเครื่องดื่มที่มีรสหวาน
16	vegetables_key	พฤติกรรมการทานผัก

จากนั้นนำข้อมูลวิเคราะห์พฤติกรรมที่ส่งผลให้เกิดโรคไม่ติดต่อเรื้อรัง โดยทำการเปรียบเทียบด้วย อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine), อัลกอริทึมเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor), อัลกอริทึมโครงข่ายประสาทเทียม (Neural Network), อัลกอริทึมต้นไม้ป่าสุ่ม (Random Forest) และอัลกอริทึมนาอิวเบย์ (Naïve Bays) โดยผลการเปรียบเทียบแสดงในตารางที่ 2

ตารางที่ 2 : ผลการเปรียบเทียบความแม่นยำ

ที่	อัลกอริทึม	Performance	
		Accuracy	Kappa
1	Random Forest	94.07	0.921
2	Neural Network	92.56	0.901
3	Support Vector Machine	91.90	0.892
4	K-Nearest Neighbor	47.69	0.303
5	Naïve Bays	89.17	0.856

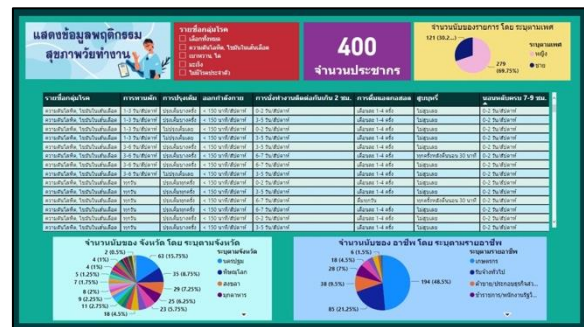
จากการทดสอบทั้ง 5 อัลกอริทึม พบว่า อัลกอริทึมต้นไม้ป่าสุ่ม ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่าความถูกต้อง (Accuracy) ร้อยละ 94.07 และมีค่าสัมประสิทธิ์ตัวชี้วัดทางสถิติ(kappa) ที่ 0.921

#### 4.2 ผลการพัฒนาหน้ารายงานแสดงผล



ภาพที่ 10 : หน้ารายงานแสดงผลภาพรวม

จากภาพที่ 10 แสดงตัวอย่างการแสดงผลภาพรวมการเกิดโรคไม่ติดต่อเรื้อรัง 3 กลุ่มโรค อันได้แก่ โรคความดันโลหิต โรคไขมันในเส้นเลือด, โรคเบาหวาน โรคไต, โรคกระเพาะ และไม่มีโรคประจำตัว เมื่อใช้ตัวกรองข้อมูลเป็นรายจังหวัด และอาชีพ โดยจะแสดง จำนวนคนทั้งหมด แสดงจำนวนผู้ที่มีโรคประจำตัวโดยแยกตามกลุ่มโรค และเพศ



ภาพที่ 11 : รายงานแสดงข้อมูลพฤติกรรมสุขภาพของวัยทำงาน

ภาพที่ 11 แสดงรายการพฤติกรรมโดยเลือกจากกลุ่มโรค ในตารางจะแสดงพฤติกรรมที่ส่งผลให้เกิดโรคตามที่เลือก โดยมีพฤติกรรม ดังนี้ การทานอาหารรสเค็มหรือการปรุงรสเค็ม การดื่มเครื่องดื่มที่มีรสหวานหรือน้ำอัดลม พฤติกรรมการทานผัก พฤติกรรมการออกกำลังกาย พฤติกรรมการ

เคลื่อนไหวและการนั่ง พฤติกรรมการนอน พฤติกรรมการดื่ม เครื่องดื่มแอลกอฮอล์ และพฤติกรรมการสูบบุหรี่



ภาพที่ 12 : หน้ารายงานแสดงผลการพยากรณ์พฤติกรรมที่ส่งผลให้เกิดโรคไม่ติดต่อเรื้อรัง

จากภาพที่ 12 แสดงผลการพยากรณ์ว่าตัวอัลกอริทึมพยากรณ์โรคโดยวิเคราะห์จากพฤติกรรมได้ถูกต้องหรือไม่ โดยสามารถเลือกดูได้ด้วยตัวเลือก อำเภอ จังหวัด และกลุ่มโรค โดยจะแสดงผลเป็นตาราง และแสดงจำนวนคนทั้งหมด ค่าเฉลี่ยอายุ และเพศ

5. สรุป อภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาพฤติกรรมที่ส่งผลให้เกิดโรคไม่ติดต่อเรื้อรัง โดยจะเน้นไปที่ 3 กลุ่มโรค อันได้แก่ 1)โรคความดันโลหิต โรคไขมันในเส้นเลือด 2)โรคไต โรคเบาหวาน 3)โรคมะเร็ง เพื่อเปรียบเทียบและหาอัลกอริทึมที่เหมาะสมที่สุดในการวิเคราะห์พฤติกรรมที่ส่งผลให้เกิดโรค ไม่ติดต่อเรื้อรัง

ในการสร้างตัวแบบพยากรณ์ได้นำเทคนิคการพยากรณ์ 5 เทคนิคมาวิเคราะห์ พบว่าตัวแบบพยากรณ์ต้นไม้ป่าสุ่มเป็นตัวแบบที่มีประสิทธิภาพสำหรับข้อมูลชุดนี้มากที่สุด โดยมีค่าความถูกต้องอยู่ที่ ร้อยละ 92.56 หลังจากสร้างตัวแบบพยากรณ์เสร็จสิ้นแล้ว ได้นำข้อมูลมาแสดงผลในรูปแบบต่างๆ ด้วยโปรแกรม Power BI Desktop

6. ข้อเสนอแนะในการวิจัย

6.1 ในการศึกษาครั้งนี้เป็นการศึกษาพฤติกรรมในชีวิตประจำวันของประชากรวัยทำงานจากการตอบแบบสอบถาม ที่ได้รวบรวมมาจากกรมอนามัย กระทรวงสาธารณสุข เพื่อวิเคราะห์พฤติกรรมที่ส่งผลให้เกิดโรคไม่ติดต่อเรื้อรังโดยวิเคราะห์จากโรคประจำตัวและพฤติกรรมของ

ผู้ตอบแบบสอบถามเพื่อให้เกิดความตระหนักรู้เบื้องต้นเท่านั้น แต่ปัจจัยที่ก่อให้เกิดโรคไม่ติดต่อเรื้อรัง มีมากกว่าพฤติกรรมที่ผู้วิจัยได้นำมาใช้วิเคราะห์ เช่น โรคที่สามารถส่งต่อผ่านทางพันธุกรรม ปัจจัยจากสภาพแวดล้อม หรือการมีความเครียดสะสม[7] ในการศึกษาครั้งต่อไปควรทำการศึกษาโดยนำปัจจัยต่างๆ มาใช้ในการวิเคราะห์ให้ครอบคลุมยิ่งขึ้น

6.2 ในการศึกษาครั้งนี้เป็นเพียงการวิจัยเฉพาะ 3 กลุ่มโรค ในโรคไม่ติดต่อเรื้อรัง อันได้แก่ 1)โรคความดันโลหิต โรคไขมันในเส้นเลือด 2)โรคไต โรคเบาหวาน 3)โรคมะเร็ง ซึ่งโรคในกลุ่มโรคไม่ติดต่อเรื้อรัง ยังมีอีกหลายโรคที่อยู่ในกลุ่มของโรคไม่ติดต่อเรื้อรัง เช่น โรคหลอดเลือดสมอง และหัวใจ โรคอ้วนลงพุง และอื่นๆ

เอกสารอ้างอิง

[1] Quinlan P, Price KO, Magid SK, Lyman S, Mandl LA, Stone PW. The relationship among health literacy, health knowledge, and adherence to treatment in patients with rheumatoid arthritis. HSS J 2013;9(1):42-9.

[2] องค์การอนามัยโลก World Health Organization, Noncommunicable diseases .2564 [อินเทอร์เน็ต] [สืบค้นเมื่อ 8 สิงหาคม 2564]. แหล่งข้อมูล: [https://www.who.int/health-topics/noncommunicable-diseases#tab=tab\\_1](https://www.who.int/health-topics/noncommunicable-diseases#tab=tab_1)

[3] สำนักงานกองทุนสนับสนุนการส่งเสริมสุขภาพ 2564 [อินเทอร์เน็ต] [สืบค้นเมื่อ 8 สิงหาคม 2564], <https://www.thaihealth.or.th/microsite/categories/5/ncds/2/173/176-กลุ่มโรค+NCDs.html>

[4] ณัฐธิดา ศรีเพชร และ วชิรวิระ จิตต์สกุล, “การวิเคราะห์ปัจจัยและพยากรณ์การไอของข้าราชการ” บทความวิชาการประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 17, 2564.

[5] ปัทมัทธนา หนูดี และ มณฑิลา รัตนศิริวงศ์, “การพยากรณ์ความสัมพันธ์ของปัจจัยข้อมูลผู้เรียนกับระดับผลการเรียนด้วยเทคนิคเหมืองข้อมูล” บทความวิชาการประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 17, 2564.

[6] นภาพร ดวงแก้ว, “การวินิจฉัยการเกิดโรคไตตั้งแต่อักเสบโดยใช้เทคนิคการเรียนรู้ของเครื่อง” บทความวิชาการประชุม

[7] กรมสุขภาพจิต กระทรวงสาธารณสุข.2564 [อินเทอร์เน็ต] [สืบค้นเมื่อ 8 สิงหาคม 2564]. แหล่งข้อมูล: <https://www.dmh.go.th/news-dmh/view.asp?id=30581>

# ระบบวิเคราะห์ภัยคุกคามบนเครือข่าย ด้วยวิธีโครงข่ายประสาทเทียม และ K-Means Network Intrusions Analysis with Neural Network and K-Means

ศักดิ์สิทธิ์ จันทรปลิน (Saksit Jantaraplin)<sup>1</sup> และพงศ์ศรัณย์ บุญโญปกรณ์ (Pongsarun Boonyopakorn)<sup>2</sup>

<sup>1</sup>ภาควิชาการบริหารเครือข่ายดิจิทัลและความมั่นคงปลอดภัยสารสนเทศ

คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307031858123@kmutnb.ac.th, <sup>2</sup>pongsarunb@kmutnb.ac.th

## บทคัดย่อ

บทความนี้มีวัตถุประสงค์เพื่อศึกษา ออกแบบ และเปรียบเทียบแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning Model) ต่อการวิเคราะห์ภัยคุกคามในระบบเครือข่าย ด้วยการหาความสัมพันธ์ของแพ็กเก็ตการสื่อสารข้อมูลบนเครือข่าย ที่มีรูปแบบใกล้เคียงกับลักษณะความผิดปกติหรือการโจมตีประเภทต่างๆ การวิจัยนี้ใช้ชุดข้อมูล KDDCUP99, NSL-KDD และ UNSW-NB15 ในการฝึกฝน และทดสอบแบบจำลอง เพื่อเปรียบเทียบผลลัพธ์ความเร็ว และความแม่นยำในการวิเคราะห์ชุดข้อมูล โดยออกแบบโครงสร้างแบบจำลอง 2 ลักษณะคือ ใช้อัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียว และใช้อัลกอริทึมโครงข่ายประสาทเทียมร่วมกับ K-Means

ผลการวิจัยในทุกชุดข้อมูลพบว่า เมื่อใช้อัลกอริทึมโครงข่ายประสาทเทียมร่วมกับ K-Means จะใช้เวลาในการประมวลผลน้อยกว่าการใช้อัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียวเฉลี่ย 1.43 นาที แต่อัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียวจะให้ความแม่นยำสูงกว่าการใช้อัลกอริทึมโครงข่ายประสาทเทียมร่วมกับ K-Means เฉลี่ย 0.22% ทั้งนี้เมื่อพิจารณาผลการวิจัย จากเวลาการประมวลผลซึ่งใช้น้อยลงเพียงเล็กน้อย ประกอบกับความแม่นยำที่ลดลงแปรผันตามกัน ผู้วิจัยจึงตัดสินใจเลือกใช้แบบจำลองอัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียวมาพัฒนาเว็บแอปพลิเคชันสำหรับวิเคราะห์ภัยคุกคามในระบบเครือข่าย

**คำสำคัญ:** ภัยคุกคามในระบบเครือข่าย การเรียนรู้ของเครื่อง  
โครงข่ายประสาทเทียม K-Means

## Abstract

The objective of this paper is to study, design and compare machine learning model for analyze intrusion in network system by finding correlation between network data communication packets, with a pattern that is close to nature of anomalies or attacks of various types. This research used the KDDCUP99, NSL-KDD and UNSW-NB15 datasets to train and test the model to compare the speed and accuracy results. The designed model structure is divided into 2 types, using only Neural Network algorithm and using a Neural Network algorithm along with K-Means.

The results of the research in all datasets found that when using a Neural Network algorithm along with K-Means, the processing time was an average of 1.43 minutes less than using only Neural Network algorithm, but only Neural Network algorithm provides 0.22 % higher accuracy than using a Neural Network algorithm along with K-Means. However, when considering the result, the processing time was slightly less and the accuracy is also reduced. The researcher decided to use only Neural Network algorithm model to develop a web application for analyzing network threats.

**Keywords:** Network intrusion, Machine Learning, Neural Network, K-Means.

## 1. บทนำ

ปัจจุบันการเข้าถึงเครือข่ายอินเทอร์เน็ตอาจเรียกได้ว่าเป็นปัจจัยขั้นพื้นฐานอย่างหนึ่งในการดำเนินชีวิตของประชากรส่วนใหญ่ เนื่องจากโอกาสในการเข้าถึงที่สะดวกมากขึ้น ตลอดจนการนำมาซึ่งแหล่งข้อมูลความรู้และความสามารถในการดำเนินธุรกิจ ส่งผลให้ปริมาณการเรียกใช้ข้อมูลบนเครือข่ายอินเทอร์เน็ตเพื่อเข้าถึงแพลตฟอร์มออนไลน์มีมากขึ้นทุกวัน ครอบคลุมประชากรทุกเพศ ทุกวัย ทั้งในประเทศไทยและนานาชาติ

กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมเผยแพร่ผลสำรวจพฤติกรรมผู้ใช้งานอินเทอร์เน็ตในปี 2563 [1] โดยระบุจุดประสงค์หนึ่งของการทำแบบสำรวจคือ “ดูแลคนไทยให้ใช้อินเทอร์เน็ตกันอย่างมั่นคงปลอดภัย และได้รับประโยชน์สูงสุด” ผลสำรวจระบุถึงจำนวนผู้ใช้งานอินเทอร์เน็ตที่เพิ่มมากขึ้นเฉลี่ย 11.5% ต่อปี ขณะที่ปี 2562 มีผู้ใช้งานอินเทอร์เน็ตคิดเป็น 75% ของประชากรในประเทศไทย และจากผลสำรวจระบุถึงร้อยละของผู้ตอบแบบสำรวจจำแนกตามปัญหาจากการใช้งานอินเทอร์เน็ตพบว่า 72.9% พบปัญหาความล่าช้าในการเชื่อมต่อ/ใช้งานอินเทอร์เน็ต, 48.9% ไม่มั่นใจว่าข้อมูลที่ปรากฏบนอินเทอร์เน็ตจะเชื่อถือได้, 34% ถูกรบกวนด้วยอีเมลขยะ, 23.3% การถูกละเมิดข้อมูลส่วนบุคคลหรือความเป็นส่วนตัว ฯลฯ

การเปลี่ยนแปลงที่รวดเร็วของความต้องการในการใช้งานเครือข่าย ส่งผลถึงปริมาณการจราจรของข้อมูลบนเครือข่ายที่เพิ่มมากขึ้น และนำมาซึ่งภัยคุกคามของผู้ไม่ประสงค์ดีในการแสวงหาประโยชน์จากการใช้งานเครือข่ายของผู้ใช้ โดยการประยุกต์ใช้หลากหลายเทคนิควิธีการโจมตีเครือข่าย เพื่อให้ได้มาซึ่งผลประโยชน์ อันเป็นผลเสียต่อผู้ใช้งาน และการให้บริการบนเครือข่าย

จากข้อมูลข้างต้น ผู้วิจัยจึงมีความสนใจจะใช้เทคโนโลยีการเรียนรู้ของเครื่อง ในการเรียนรู้คุณลักษณะการรับ-ส่งข้อมูลอันเป็นภัยคุกคามในระบบเครือข่าย และพัฒนาเป็นระบบวิเคราะห์เพื่อตรวจจับภัยคุกคามในระบบเครือข่าย เพื่อให้ระบบเครือข่ายและผู้ใช้งานปลอดภัยจากภัยคุกคามที่อาจเกิดขึ้น

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง (Machine Learning) [2] คือสาขาหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) และวิทยาการทางคอมพิวเตอร์ที่มุ่งเน้นการใช้ข้อมูลและอัลกอริทึมเพื่อเลียนแบบวิธีการเรียนรู้ของมนุษย์แบบค่อยๆ พัฒนาความแม่นยำ การเรียนรู้ของเครื่องแบ่งออกเป็น 3 รูปแบบหลักคือ

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นการเรียนรู้จากชุดข้อมูลนำเข้า (Input) และป้ายกำกับ (Label) ในการฝึกการเรียนรู้ของอัลกอริทึมเพื่อจัดประเภทข้อมูลหรือทำนายผลลัพธ์ โดยจะมีการปรับค่าน้ำหนักของแบบจำลองอย่างเหมาะสม
2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการเรียนรู้โดยไม่มีป้ายกำกับ เพื่อวิเคราะห์ความเหมือนและแตกต่างในชุดข้อมูล แล้วจึงจัดกลุ่มชุดข้อมูลในแต่ละรูปแบบ
3. การเรียนรู้แบบกึ่งมีผู้สอน (Semi-Supervised Learning) เป็นการเรียนรู้แบบใช้ป้ายกำกับส่วนหนึ่งเพื่อดึงเอาคุณลักษณะของข้อมูลตามป้ายกำกับนั้นๆ ออกมา เพื่อเป็นแนวทางในการจำแนกประเภทของข้อมูล

### 2.2 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (Neural Network) [3] เป็นอัลกอริทึมที่มีแนวคิดจำลองมาจากการทำงานของสมองมนุษย์ เช่น เมื่อมนุษย์รับภาพผ่านดวงตา สมองจะรับรู้และส่งต่อไปยังสมองส่วนต่างๆ เพื่อวิเคราะห์ออกมาเป็นสารให้มนุษย์รับรู้ โดยส่วนต่างๆ ของสมองที่วันนี้แบ่งเป็น 3 ส่วนหลัก ได้แก่

1. ชั้นข้อมูลเข้า (Input Layer) เป็นส่วนแรกที่ได้รับข้อมูลเข้าสู่โมเดล โดยจะมีจำนวนโหนดในชั้นข้อมูลเท่ากับจำนวนพีเจอร์ที่รับเข้ามา
2. ชั้นซ่อน (Hidden Layer) เป็นส่วนที่สำคัญในการวิเคราะห์โมเดล ภายในจะแบ่งออกเป็นกี่ชั้นการทำงานก็ได้ สามารถเพิ่มชั้นภายในโมเดลโดยการ

ปรับค่า Parameter ซึ่งจะส่งผลโดยตรงต่อประสิทธิภาพการทำงานของโมเดล

3. ชั้นข้อมูลออก (Output Layer) เป็นส่วนที่ส่งออกข้อมูลจากการวิเคราะห์ โดยจำนวนโหนดในชั้นนี้จะขึ้นอยู่กับรูปแบบและความเหมาะสมของข้อมูลที่จะส่งออก

### 2.3 K-Means

K-Means [4] เป็นอัลกอริทึมสำหรับการแบ่งกลุ่มชนิดหนึ่งที่มีกลไกเรียบง่ายไม่ซับซ้อน และสามารถใช้กับชุดข้อมูลที่มีฟีเจอร์เป็นตัวเลข (Numeric Feature) ได้เท่านี้ การทำงานของ K-Means เริ่มต้นจากผู้ใช้ต้องกำหนดค่า  $k$  ซึ่งในที่นี้หมายถึงจำนวนกลุ่มที่ต้องการออกมาเป็นผลลัพธ์ อัลกอริทึมจะนำค่า  $k$  ไปวัดระยะห่างของการกระจายตัวของกลุ่มข้อมูล แล้วจึงคำนวณหาค่าเฉลี่ยของกลุ่ม และเอาตัวเองไปอยู่ตรงกลางกลุ่ม ทำอย่างเดิมซ้ำจนกว่าจะได้ระยะห่างของกลุ่มที่สมบูรณ์หรือไม่สามารถคำนวณต่อได้

### 2.4 RapidMiner

RapidMiner [5] เป็นแพลตฟอร์มวิทยาศาสตร์ข้อมูล (Data Science Platform) สำหรับออกแบบการทำงานอย่างอัตโนมัติในงานเหมืองข้อมูล โดยไม่จำเป็นต้องเขียนโปรแกรม

RapidMiner [6] เป็นเครื่องมือที่ได้รับความนิยมในการนำไปใช้กับงาน Data Mining เนื่องจากมี GUI ที่ใช้งานง่าย และเป็น Open-Source ที่สามารถสร้าง Plugin หรือเลือกใช้ Plugin ที่มีอยู่ได้หลากหลาย ด้วยคุณสมบัติในการ Drag & Drop ทำให้สามารถสร้างแบบจำลองได้สะดวก โดยซ่อนการทำงานที่ซับซ้อนเอาไว้เบื้องหลัง

## 3. วิธีการดำเนินงาน

### 3.1 เตรียมเครื่องมือและปรับตั้งค่า

ในการวิจัยนี้ ผู้วิจัยใช้โปรแกรม RapidMiner version 9.10.001 บนเครื่องที่ใช้หน่วยประมวลผล Intel i7-7700HQ Ram 12 GB โดยอัลกอริทึมที่เลือกใช้จะปรับตั้งค่าพารามิเตอร์ตามค่าตั้งต้นของ โปรแกรม RapidMiner แต่

จะปรับค่า  $K$  ให้กับอัลกอริทึม K-Means ที่ 23 และ 10 ตามจำนวน Label ของชุดข้อมูล

### 3.2 รวบรวมชุดข้อมูล

การศึกษานี้ใช้ชุดข้อมูลมาตรฐานสำหรับการตรวจจับภัยคุกคาม (Benchmark Datasets) ทั้งสิ้น 3 ชุดในการฝึกฝนและทดสอบ ประกอบด้วย

- 1) KDDCUP99 [7] เป็นบันทึกการรับ-ส่งข้อมูลบนโปรโตคอล TCP ในระยะเวลา 7 สัปดาห์ รวมจำนวนบันทึกข้อมูลประมาณ 5 ล้านรายการ แบ่งการโจมตีออกเป็น 4 ประเภทหลัก คือ DOS (การปฏิเสธการให้บริการ), R2L (การเข้าถึงโดยไม่ได้รับอนุญาตจากเครื่องระยะไกล), U2R (การเข้าถึงสิทธิ์ Superuser โดยไม่ได้รับอนุญาต และ Probing (ตรวจสอบค้นหาข้อมูล เช่น การสแกนพอร์ต)
- 2) NSL-KDD [7] เป็นชุดข้อมูลที่ได้รับการพัฒนามาจาก KDDCUP99 โดยการลบรายการที่ซ้ำกันออก , ปรับจำนวนรายการในแต่ละรูปแบบการโจมตีให้มีความเท่ากัน เพื่อไม่ให้เกิดการเอนเอียงในการเรียนรู้การโจมตีรูปแบบใดรูปแบบหนึ่ง
- 3) UNSW-NB15 [8] เป็นชุดข้อมูลซึ่งสร้างขึ้นโดยเครื่องมือ IXIA PerfectStorm ชุดข้อมูลจะประกอบไปด้วยรายการปกติแบบสมัยใหม่ และพฤติกรรมโจมตีแบบใหม่ๆ 9 ประเภท ได้แก่ Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode และ Worms
- 4) ฟีเจอร์ทั้งหมดที่ใช้เป็นส่วนประกอบของการวิเคราะห์แบบจำลองในงานวิจัยนี้ ได้แก่ ข้อมูลดังตารางที่ 1

ตารางที่ 1: รายชื่อฟีเจอร์จากแต่ละชุดข้อมูลที่ใช้ในการทดสอบ

ชื่อฟีเจอร์ และคำอธิบายบางส่วน	ชุดข้อมูล		
	KDD CUP 99	NSL- KDD	UNSW -NB15
duration (ความยาวการเชื่อมต่อ)	✓	✓	
protocol_type (tcp, udp, etc.)	✓	✓	
service (http, telnet, etc.)	✓	✓	✓
flag (สถานะการเชื่อมต่อ)	✓	✓	
src_bytes (จำนวนไบต์จากต้นทาง)	✓	✓	
dst_bytes (จำนวนไบต์จากปลายทาง)	✓	✓	
land	✓	✓	
wrong_fragment	✓	✓	
urgent	✓	✓	
count	✓	✓	
srv_count	✓	✓	
error_rate	✓	✓	
srv_error_rate	✓	✓	
rerror_rate	✓	✓	
srv_rerror_ra	✓	✓	
same_srv_rate	✓	✓	
diff_srv_rate	✓	✓	
srv_diff_host_rate	✓	✓	
dst_host_count	✓	✓	
dst_host_srv_count	✓	✓	
dst_host_same_srv_rate	✓	✓	
dst_host_diff_srv_rate	✓	✓	
dst_host_same_src_port_rate	✓	✓	
dst_host_srv_diff_host_rate	✓	✓	
dst_host_serror_rate	✓	✓	
dst_host_srv_serror_rate	✓	✓	
dst_host_rerror_rate	✓	✓	
dst_host_srv_rerror_rate	✓	✓	
dur (ความยาวการเชื่อมต่อ)			✓
proto (tcp, udp, etc.)			✓
state			✓
spkts (จำนวนแพ็กเก็ตจากต้นทาง)			✓
dpkts (จำนวนแพ็กเก็ตจากปลายทาง)			✓
sbytes (จำนวนไบต์จากต้นทาง)			✓
dbytes (จำนวนไบต์จากปลายทาง)			✓

ชื่อฟีเจอร์ และคำอธิบายบางส่วน	ชุดข้อมูล		
	KDD CUP 99	NSL- KDD	UNSW -NB15
rate			✓
sttl (TTL จากต้นทางไปปลายทาง)			✓
dttl (TTL จากปลายทางไปต้นทาง)			✓
sload (จำนวนบิตต้นทางต่อวินาที)			✓
dload (จำนวนบิตปลายทางต่อวินาที)			✓
sloss (แพ็กเก็ตต้นทางลดลง)			✓
dloss (แพ็กเก็ตปลายทางลดลง)			✓
sinpkt (เวลามาถึงแพ็กเก็ตต้นทาง)			✓
dinpkt (เวลามาถึงแพ็กเก็ตปลายทาง)			✓
sjit			✓
djit			✓
swin			✓
stcpb			✓
dtcpb			✓
dwin			✓
tcprtt (เวลาไป-กลับการเชื่อมต่อ TCP)			✓
synack			✓
ackdat			✓
smean			✓
dmean			✓
trans_depth			✓
response_body_len			✓
ct_srv_src			✓
ct_state_ttl			✓
ct_dst_ltm			✓
ct_src_dport_ltm			✓
ct_dst_sport_ltm			✓
ct_dst_src_ltm			✓
is_fip_login			✓
ct_fip_cmd			✓
ct_flw_http_mthd			✓
ct_src_ltm			✓
ct_srv_dst			✓
is_sm_ips_ports			✓
attack_cat			✓

### 3.3 ศึกษาและทดลองโมเดล

ศึกษาและสร้างองค์ประกอบการทำงานของ โมเดล 2 รูปแบบ ได้แก่

3.2.1 แบบใช้อัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียว

3.2.2 แบบใช้อัลกอริทึมโครงข่ายประสาทเทียมร่วมกับ K-Means

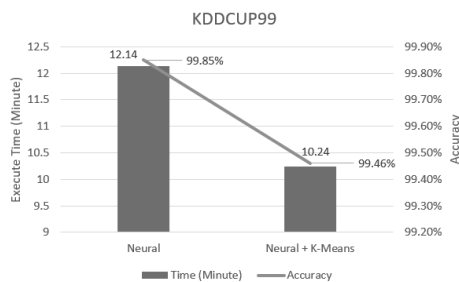
องค์ประกอบการทำงานของแต่ละโมเดลประกอบด้วย 4 ส่วนหลัก ได้แก่

- 1) Benchmark Datasets การเรียกใช้ชุดข้อมูลที่ได้ Import ไว้ก่อนหน้านี้
- 2) Cleansing Data การเตรียมข้อมูล ให้เหมาะสมกับการทำงานของอัลกอริทึม แบ่งเป็น 2 ขั้นตอน คือ
  - 2.1 แทนที่พีเจอร์ทที่หายไปในบางรายการของชุดข้อมูลโดยใช้ค่าเฉลี่ย
  - 2.2 แทนที่ข้อมูลประเภท Polynominal โดยการ Mapping กับค่าตัวเลขที่ไม่ซ้ำกัน
- 3) Train & Test Model ฝึกฝนและทดสอบโมเดล โดยแบ่งชุดข้อมูลเป็น 70% สำหรับฝึกฝน และ 30% สำหรับทดสอบ
- 4) Compare Result เปรียบเทียบความแม่นยำในการทำนายผล

## 4. ผลการดำเนินงาน

### 4.1 ผลการทดสอบด้วยชุดข้อมูล KDDCUP99

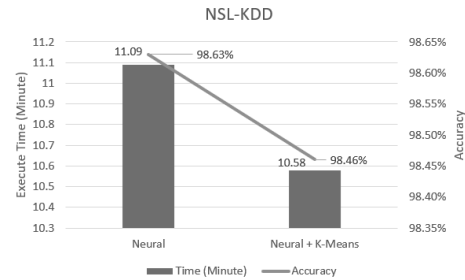
จากการทดสอบพบว่า อัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียวใช้เวลาในการประมวลผล 12.14 นาที มีความแม่นยำที่ 99.85% และอัลกอริทึมโครงข่ายประสาทเทียมร่วมกับ K-Means ใช้เวลาในการประมวลผล 10.24 นาที มีความแม่นยำที่ 99.46%



ภาพที่ 2: ผลการทดสอบโมเดลด้วยชุดข้อมูล KDDCUP99

### 4.2 ผลการทดสอบด้วยชุดข้อมูล NSL-KDD

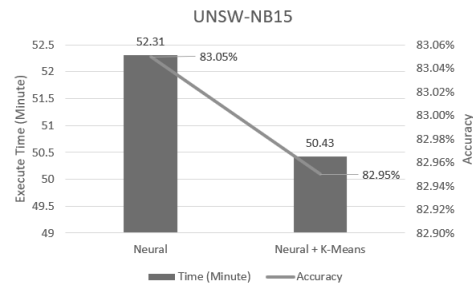
จากการทดสอบพบว่า อัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียวใช้เวลาในการประมวลผล 11.09 นาที มีความแม่นยำที่ 98.63% และอัลกอริทึมโครงข่ายประสาทเทียมร่วมกับ K-Means ใช้เวลาในการประมวลผล 10.58 นาที มีความแม่นยำที่ 98.46%



ภาพที่ 3: ผลการทดสอบโมเดลด้วยชุดข้อมูล NSL-KDD

### 4.3 ผลการทดสอบด้วยชุดข้อมูล UNSW-NB15

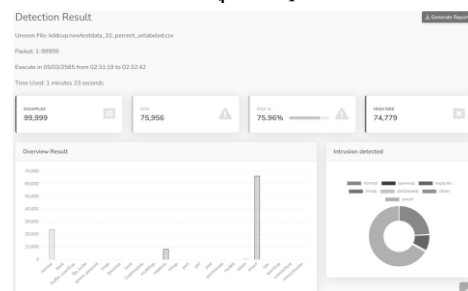
จากการทดสอบพบว่า อัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียวใช้เวลาในการประมวลผล 52.31 นาที มีความแม่นยำที่ 83.05% และอัลกอริทึมโครงข่ายประสาทเทียมร่วมกับ K-Means ใช้เวลาในการประมวลผล 50.43 นาที มีความแม่นยำที่ 82.95%



ภาพที่ 4: ผลการทดสอบโมเดลด้วยชุดข้อมูล UNSW-NB15

### 4.4 การพัฒนาเว็บแอปพลิเคชัน

ผู้วิจัยได้พัฒนาเว็บแอปพลิเคชันด้วยภาษา Python เพื่อสังการวิเคราะห์และแสดงผลการวิเคราะห์ภัยคุกคามผ่านการเรียกใช้โปรแกรมประยุกต์ RapidMiner



ภาพที่ 5: แสดงผลการวิเคราะห์ชุดข้อมูลบนเว็บแอปพลิเคชัน



## 5. สรุป

การวิจัยนี้ผู้วิจัยได้พัฒนาโมเดลการตรวจจับภัยคุกคามในระบบเครือข่าย จากข้อมูลการสื่อสารภายในเครือข่าย เพื่อวัดความเร็วในการประมวลผลและความแม่นยำในการวิเคราะห์ผลลัพธ์ โดยโมเดลที่ได้ทดสอบประกอบด้วย โมเดลแบบใช้อัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียว และโมเดลแบบใช้อัลกอริทึมโครงข่ายประสาทเทียมร่วมกับ K-Means โดยใช้ชุดข้อมูล KDDCUP99, NSL-KDD และ UNSW-NB15 สำหรับฝึกฝนและทดสอบ

ผลการวิจัยพบว่า โมเดลแบบใช้อัลกอริทึมโครงข่ายประสาทเทียมร่วมกับ K-Means ใช้เวลาในการประมวลน้อยกว่าโมเดลแบบใช้อัลกอริทึมโครงข่ายประสาทเทียมวิธีเดียว เฉลี่ยเพียง 1.43 นาที และวิธีการแบบใช้อัลกอริทึมร่วมกันนี้ยังทำให้สูญเสียความแม่นยำไปโดยเฉลี่ย 0.22% ทั้งนี้ผู้วิจัยจึงตัดสินใจเลือกใช้แบบจำลองอัลกอริทึมโครงข่ายประสาทเทียมเพียงวิธีเดียวมาพัฒนาเว็บแอปพลิเคชันสำหรับวิเคราะห์ภัยคุกคามในระบบเครือข่าย

## เอกสารอ้างอิง

- [1] “รายงานผลการสำรวจพฤติกรรมผู้ใช้อินเทอร์เน็ตในประเทศไทย ปี 2563”. <https://www.etcha.or.th/th/Useful-Resource/publications/Thailand-Internet-User-Behavior-2020.aspx> (สืบค้น 1 เมษายน 2022).
- [2] T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.
- [3] F. Günther and S. Fritsch, “neuralnet: Training of Neural Networks”, *R J.*, Year 2, No. 1, P. 30, 2010, doi: 10.32614/RJ-2010-006.
- [4] S. Na, L. Xumin, and G. Yong, “Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm”, in *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, Apr. 2010, P. 63–67. doi: 10.1109/IITSI.2010.74.
- [5] M. Hofmann and R. Klinkenberg, *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. CRC Press, 2016.
- [6] N. Kitcharoen, S. Kamolsantisuk, R. Angsomboon, and T. Achalakul, “RapidMiner framework for manufacturing data analysis on the cloud”, in *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, May 2013, P. 149–154. doi: 10.1109/JCSSE.2013.6567336.
- [7] A. Divekar, M. Parekh, V. Savla, R. Mishra, and M. Shirole, “Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives”, in *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*, Oct. 2018, P. 1–8. doi: 10.1109/ICCCS.2018.8586840.
- [8] N. Moustafa and J. Slay, “UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)”, in *2015 Military Communications and Information Systems Conference (MilCIS)*, May 2015, P. 1–6. doi: 10.1109/MilCIS.2015.7348942.
- [9] H. Xu, C. Fu, S. Yao, และ X. Zong, “An Improved K-means Algorithm Based on Intersection over Union for Network Security”, in *2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN)*, Jun. 2019, P. 514–517. doi: 10.1109/ICCSN.2019.8905279.
- [10] J. Woo, J.-Y. Song, and Y.-J. Choi, “Performance Enhancement of Deep Neural Network Using Feature Selection and Preprocessing for Intrusion Detection”, in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, Feb. 2019, P. 415–417. doi: 10.1109/ICAIIIC.2019.8668995.
- [11] D. Preethi and N. Khare, “Performance Evaluation of Shallow learning techniques and Deep Neural Network for Cyber Security”, in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, Feb. 2020, P. 1–5. doi: 10.1109/ic-ETITE47903.2020.128.
- [12] A. Pattawaro and C. Polprasert, “Anomaly-Based Network Intrusion Detection System through Feature Selection and Hybrid Machine Learning Technique”, in *2018 16th International Conference on ICT and Knowledge Engineering (ICT KE)*, May 2018, P. 1–6. doi: 10.1109/ICTKE.2018.8612331.

# การพัฒนาระบบสนับสนุนการตัดสินใจการพยากรณ์ความเสี่ยงในการเกิด โรคเบาหวานด้วยเทคนิคเหมืองข้อมูล

## The Development of Decision Support System for Diabetes Risk Prediction Using Data Mining Technology

ณัฐพงศ์ แผลหะดี (Nattapong Laehte) <sup>1</sup> และมณฑิธร รัตนศิริวงษ์วุฒิ (Montean Rattanasiriwongwut) <sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ<sup>1,2</sup>

s6307021858031@email.kmutnb.ac.th<sup>1</sup>, montean.r@itd.kmutnb.ac.th<sup>2</sup>

### บทคัดย่อ

การวิจัยครั้งนี้มีจุดประสงค์เพื่อทำการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวาน และทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึมเพื่อนำใช้ในการพัฒนาระบบสนับสนุนการตัดสินใจการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูล ซึ่งอัลกอริทึมที่ใช้ในการวิเคราะห์และพยากรณ์นั้นมีทั้งหมด 5 อัลกอริทึม ได้แก่ เทคนิคต้นไม้ป่าสุ่ม (Random Forest), เทคนิคต้นไม้ตัดสินใจ (Decision tree), เทคนิคโครงข่ายประสาทเทียม (Neural Network), เทคนิคโครอนาอีฟเบย์ (Naïve Bayes) และเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) ซึ่งข้อมูลที่นำมาใช้ในการทดลองครั้งนี้เป็นข้อมูลจากกรมอนามัย กระทรวงสาธารณสุข จำนวน 45462 รายการ เพื่อหาผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึม โดยพบว่าเทคนิคต้นไม้ป่าสุ่มให้ค่าประสิทธิภาพสูงสุด โดยให้ค่าความถูกต้องอยู่ที่ 99.68% ค่าความแม่นยำอยู่ที่ 99.85% ค่าความระลึกอยู่ที่ 99.11% และค่าประสิทธิภาพอยู่ที่ 99.48% จากนั้นนำข้อมูลและอัลกอริทึมที่ได้มาแสดงผลรายงานให้อยู่ในรูปแบบของระบบ

**คำสำคัญ:** การพยากรณ์ เบาหวาน เหมืองข้อมูล

### Abstract

The purpose of this research was to predict the risk of developing diabetes and to compare the performance of the algorithms for use in the

development of decision support system for diabetes risk prediction using data mining technology. There are 5 algorithms used for analysis and forecasting: Random Forest Technique (Random Forest), Decision tree Technique (Decision tree), Neural Network Technique (Neural Network), Naïve Bayes Technique (Naïve Bayes) and K-Nearest Neighbor Technique (K-Nearest Neighbor). The 45,462 dataset items from the Department of Health, Ministry of Public Health were used in this experiment in order to compare the algorithm efficiency, with a accuracy of 99.68%, a precision of 99.85%, a recall of 99.11% and an efficiency of 99.48%. The resulting report is displayed in the form of a system.

**Keywords:** Prediction, Diabetes, Data Mining

### 1. บทนำ

ในปัจจุบันเรามักพบเห็นได้ว่าการใช้ชีวิตของแต่ละบุคคลมีการดำเนินชีวิตที่แตกต่างกันออกไปจึงไม่สามารถทำการบ่งบอกได้ว่าการดำเนินชีวิตของแต่ละคนนั้นมีภาวะและมีความเสี่ยงในการเกิดโรคมามากน้อยแค่ไหน และไม่จำเป็นจะต้องเป็นโรคเบาหวานเสมอไป อีกทั้งโรคเบาหวานไม่จำเป็นจะต้องเกิดที่เฉพาะคนอ้วนหรือคนสูงวัยเท่านั้น แต่ยังสามารถทำการพบเห็นได้จากบุคคลที่มีร่างกายสมส่วน หรืออาจจะเกิดจากกรรมพันธุ์

ตั้งแต่กำเนิด และยังสามารถทำการสังเกตได้จากการรับประทานอาหารที่มีน้ำตาลสูงเพิ่มขึ้นในทุกๆวัน ซึ่งในปัจจุบันเราสามารถพบเห็นได้ว่าการเป็นโรคเบาหวานนั้นเกิดจากความผิดปกติของร่างกายที่มีการผลิตฮอร์โมนอินซูลินไม่เพียงพอ จึงทำให้เกิดโรคเบาหวานและอีกทั้งร่างกายไม่สามารถนำน้ำตาลในร่างกายไปใช้งานได้อย่างมีประสิทธิภาพจึงทำให้ระดับน้ำตาลในเลือดสูงขึ้น จึงทำให้เกิดโรคเบาหวานได้ในแต่ละบุคคล

จากการใช้ชีวิตในปัจจุบันของคนรุ่นใหม่มีการใช้ชีวิตที่เปลี่ยนแปลงไป ซึ่งมักเห็นได้ว่าการใช้ชีวิตของผู้คนมีความเสี่ยงที่อาจจะเกิดโรคเบาหวานเป็นอย่างมาก การเกิดโรคเบาหวานไม่จำเป็นจะต้องเป็นคนที่มีร่างกายอ้วน หรือเป็นผู้สูงอายุ แต่ในปัจจุบันทุกคนมีความเสี่ยงในการเป็นโรคเบาหวานเป็นอย่างมาก เนื่องจากการใช้ชีวิตในปัจจุบันค่อนข้างที่จะเปลี่ยนแปลง และผู้คนส่วนใหญ่ไม่ค่อยที่จะหันมาออกกำลังกายกัน ซึ่งปัญหาเหล่านี้มีความเสี่ยงอย่างยิ่งที่จะทำให้เกิดโรคเบาหวานได้ง่าย และสามารถเกิดขึ้นได้ง่ายกับหมู่คนวัยรุ่น วัยทำงาน แม้กระทั่งเด็กก็มีความเสี่ยงในการเกิดโรคเบาหวานได้ง่ายยิ่งขึ้นอีกด้วย

จากประเด็นและปัญหาดังกล่าวผู้วิจัยมีความประสงค์ที่จะนำแนวคิดและเทคนิคในการจัดทำเหมืองข้อมูล (Data Mining) มาประยุกต์ใช้ร่วมกับการพัฒนาระบบสนับสนุนการตัดสินใจความเสี่ยงในการเกิดโรคเบาหวาน เพื่อช่วยให้ผู้ที่เข้ามาใช้ระบบได้ทราบถึงความเสี่ยงในการเกิดโรคได้อย่างเบื้องต้น อีกทั้งสามารถใช้งานของระบบเป็นช่องทางในการทราบผลความเสี่ยงได้อย่างรวดเร็ว สะดวก และแม่นยำ และช่วยลดขั้นตอนในการบันทึกข้อมูลลงบนแบบฟอร์ม อีกทั้งยังช่วยให้เจ้าหน้าที่สามารถทำการตรวจสอบผลของผู้ที่ทำการบันทึกข้อมูลและทราบถึงผลลัพธ์รวมทั้งหมดเพื่อทำการตัดสินใจในการติดต่อผู้ที่บันทึกเพื่อเข้าทำการรักษาโรคเบาหวานเบื้องต้นได้อีกด้วย และยังช่วยให้เจ้าหน้าที่สามารถดูข้อมูลที่แบ่งออกตามแต่ละรายการได้

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) คือการค้นหาคำตอบในฐานข้อมูลเพื่อเป็นกระบวนการที่กระทำกับข้อมูลที่มีจำนวนมาก เพื่อค้นหารูปแบบแนวทางและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักทางสถิติ การเรียนรู้ของเครื่อง (Machine Learning) และการจดจำรูปแบบ (Pattern Recognition) [2]

### 2.2 เทคนิคต้นไม้ป่าสุ่ม (Random Forest)

เทคนิคต้นไม้ป่าสุ่มเป็นเทคนิคที่ใช้ตัวทำแบบ (Model) การจำแนกข้อมูลหลายๆตัวแบบมาช่วยในการค้นหาคำตอบซึ่งจะทำให้มีประสิทธิภาพมากกว่าการใช้ตัวแบบเดียวกัน ซึ่งเป็นการสุ่มตัวอย่างข้อมูลในชุดข้อมูล ออกเป็นหลายๆชุด โดยในแต่ละชุดมีการสุ่มแอตทริบิวต์มาใช้สร้างตัวแบบโดยเทคนิคที่ใช้ในการสร้างตัวแบบเป็น Decision Tree เดียวกันทั้งหมด ถึงแม้ว่าจะเป็นเทคนิคเดียวกันแต่ข้อมูลและแอตทริบิวต์ที่ใช้ในการสร้างต่างกัน ก็ทำให้ผลลัพธ์ที่สร้างมีลักษณะต่างกันออกไป [1]

### 2.3 เทคนิคต้นไม้ตัดสินใจ (Decision tree)

เทคนิคต้นไม้ตัดสินใจ เป็นเทคนิคที่ใช้ในงานด้านเหมืองข้อมูลประเภทการจำแนกข้อมูลที่มีจุดเด่นคือแสดงผลลัพธ์เป็นผังต้นไม้ ซึ่งเป็นวิธีที่ได้รับความนิยมเนื่องจากอัลกอริทึมไม่ซับซ้อน ติความและเข้าใจลักษณะของรูปแบบข้อมูลได้ง่าย [3]

### 2.4 เทคนิคโครงข่ายประสาทเทียม (Neural Network)

โครงข่ายประสาทเทียมเป็นโมเดลทางคณิตศาสตร์หรือโมเดลทางคอมพิวเตอร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนกชันนิสต์ (Connectionist) ที่มีการรวมกลุ่มแบบขนานของหน่วยประมวลผลย่อย โดยได้แนวคิดมาจากการจำลองการทำงานของเซลล์สมองมนุษย์ที่แต่ละเซลล์ประสาทจะประกอบไปด้วยเดนไดรต์ หรือปลายในรับกระแสประสาท [5]

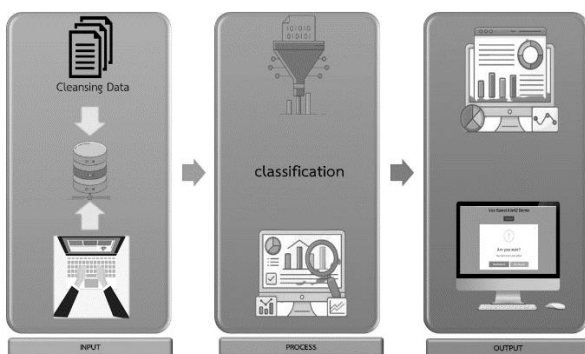
## 2.5 เทคนิคนาอิวเบย์ (Naïve Bayes)

นาอิวเบย์เป็นเครื่องมือที่อาศัยหลักการของความน่าจะเป็นตามทฤษฎีของเบย์ (Bayes theorem) เป็นอัลกอริทึมที่ไม่ซับซ้อน จำแนกข้อมูลโดยการเรียนรู้ปัญหาที่เกิดขึ้นเพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ หลักการของนาอิวเบย์จะทำการคำนวณหาความน่าจะเป็นซึ่งถูกใช้ในการแก้ปัญหาแบบการจำแนกที่สามารถคาดการณ์ผลลัพธ์ได้ และวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ [4]

## 2.6 เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor)

เทคนิคนี้จะตัดสินใจว่าคลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ โดยทำการตรวจสอบจำนวนบางจำนวนในขั้นตอนวิธีการหาเพื่อนบ้านใกล้ที่สุดของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุดโดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไขหรือกรณีต่างๆ สำหรับแต่ละคลาสและกำหนดเงื่อนไขใหม่ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด [6]

## 3. วิธีการดำเนินการวิจัย



ภาพที่ 1 : กรอบแนวคิดในการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวานด้วยเทคนิคเหมืองข้อมูล

จากภาพที่ 1 แสดงกรอบแนวคิดของงานวิจัย การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อเป็นการวิเคราะห์และทำการหาประสิทธิภาพของอัลกอริทึมที่ดีที่สุดและแม่นยำที่สุดในการพยากรณ์ โดยผู้วิจัยได้นำข้อมูลจากกรมอนามัย

กระทรวงสาธารณสุข มาใช้ในการวิเคราะห์และพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวาน โดยกรอบแนวคิดจะแบ่งออกด้วยกันเป็น 2 ส่วน คือ 1. นำข้อมูลที่ได้มาทำความสะอาด (Cleansing Data) เมื่อทำความสะอาดข้อมูลเรียบร้อยแล้วนำข้อมูลที่ได้มาหาอัลกอริทึมที่ดีที่สุด จากนั้นนำมาแสดงภาพรวมของข้อมูล ผ่านระบบที่ถูกพัฒนาขึ้นมา 2. ผู้ใช้งานระบบทำการกรอกบันทึกข้อมูลลงบนฟอร์มของเว็บไซต์ และเข้าสู่กระบวนการการวิเคราะห์โดยใช้อัลกอริทึมที่ดีที่สุดมาวิเคราะห์ข้อมูล จากนั้นระบบจะแสดงผลลัพธ์ของความเสี่ยงทันทีบนระบบหลังจากบันทึกข้อมูลซึ่งสามารถแบ่งขั้นตอนในการดำเนินงานออกเป็น 4 ขั้นตอน ดังนี้

### 3.1 การศึกษาและเปรียบเทียบเทคนิคของอัลกอริทึม

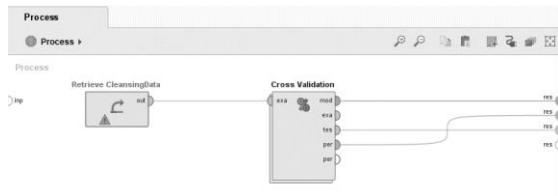
ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องและทำการศึกษาอัลกอริทึมที่จะใช้ในการวิเคราะห์และการพยากรณ์ข้อมูล โดยใช้อัลกอริทึม 5 อัลกอริทึม ได้แก่ เทคนิคต้นไม้ป่าสุ่ม, เทคนิคต้นไม้ตัดสินใจ, เทคนิคโครงข่ายประสาทเทียม, เทคนิคนาอิวเบย์ และเทคนิคเพื่อนบ้านใกล้ที่สุด มาใช้ในการเปรียบเทียบเพื่อหาอัลกอริทึมที่ดีที่สุดในการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวาน

### 3.2 การเตรียมข้อมูลและคัดเลือกข้อมูล

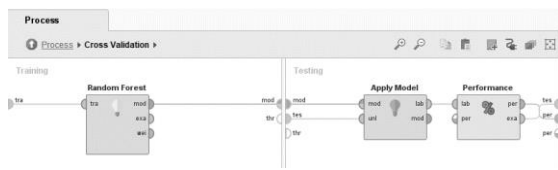
ข้อมูลที่นำมาใช้ในงานวิจัยครั้งนี้ เป็นข้อมูลที่รวบรวมจากแบบสอบถามพฤติกรรมอันพึงประสงค์ของประชากรวัยทำงาน สังกัดสำนักงานปลัดกระทรวงสาธารณสุข อยู่ในช่วงวันที่ 1 ตุลาคม พ.ศ.2563 จนถึงวันที่ 31 กันยายน พ.ศ.2564 ซึ่งเป็นข้อมูลเกี่ยวกับพฤติกรรมในชีวิตประจำวันของประชากรวัยทำงาน ได้แก่ เพศ อายุ ส่วนสูง น้ำหนัก เส้นรอบเอว พฤติกรรมการทานอาหารและการปรุงอาหาร และพฤติกรรมการออกกำลังกายและการพักผ่อน พฤติกรรมการสูบบุหรี่ และการดื่มเครื่องดื่มแอลกอฮอล์ จากนั้นนำข้อมูลที่

ได้มา มาทำการคัดสรรและทำการคัดเลือกข้อมูลที่จะในการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวาน

**3.3 ขั้นตอนการสร้างโมเดลพยากรณ์ เปรียบเทียบ และคัดเลือกเทคนิคการพยากรณ์ที่เหมาะสมที่สุดไปใช้งาน**



ภาพที่ 2 : แสดงรูปขั้นตอนการหาประสิทธิภาพของอัลกอริทึม(1)



ภาพที่ 3 : แสดงรูปขั้นตอนการหาประสิทธิภาพของอัลกอริทึม(2)

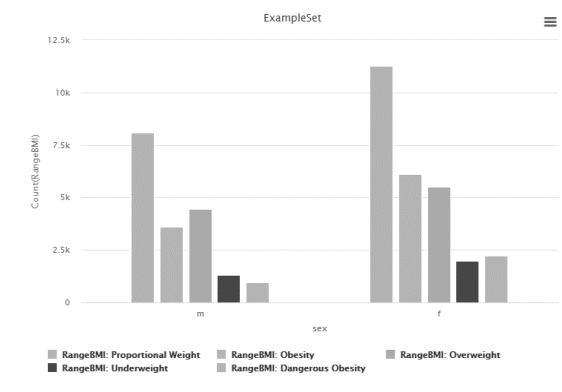
จากภาพที่ 2 และ 3 เป็นการแสดงขั้นตอนของการหาประสิทธิภาพของอัลกอริทึมที่ใช้ในงานวิจัยฉบับนี้ ซึ่งใช้ RapidMiner Studio เป็นเครื่องมือในการหาประสิทธิภาพของอัลกอริทึมทั้งหมดและนำผลลัพธ์ที่ได้ไปทำการเปรียบเทียบสรุปผลและคัดเลือกผลที่แม่นยำที่สุด

งานวิจัยฉบับนี้ ผู้วิจัยได้ทำการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูล 5 อัลกอริทึม ได้แก่ เทคนิคต้นไม้ป่าสุ่ม, เทคนิคต้นไม้ตัดสินใจ, เทคนิคโครงข่ายประสาทเทียม, เทคนิคนาอิวเบย์ และเทคนิคเพื่อนบ้านใกล้ มาใช้ในการเปรียบเทียบเพื่อหาอัลกอริทึมที่ดีที่สุด โดยใช้โปรแกรม RapidMiner Studio ในการวิเคราะห์ โดยเริ่มจากการเก็บและรวบรวมข้อมูล ทำความสะอาดข้อมูลที่ได้และทำการแบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลฝึกฝน (Training Data) ใช้ข้อมูล 70 เปอร์เซ็นต์จากข้อมูลทั้งหมดเพื่อใช้ในการฝึกฝนอัลกอริทึมทั้งหมดที่ใช้ในการพยากรณ์ และข้อมูลทดสอบ (Test Data) ซึ่งใช้

ข้อมูล 30 เปอร์เซ็นต์จากข้อมูลทั้งหมด เพื่อใช้ในการทดสอบอัลกอริทึม หลังจากนั้นนำข้อมูลที่ได้มาใช้เป็นข้อมูลในการพยากรณ์และสร้างแบบจำลอง เมื่อทำการพยากรณ์แต่ละเทคนิคเรียบร้อยแล้ว ทำการสรุปผลและเลือกเทคนิคที่เหมาะสมและแม่นยำที่สุด

**3.4 ขั้นตอนการแสดงผลและแสดงภาพรวมของข้อมูล**

ขั้นตอนนี้เป็นขั้นตอนที่ทำการแสดงผลการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวาน เพื่อนำไปใช้ในการวิเคราะห์และหาสาเหตุในการเกิดโรคเบาหวาน



ภาพที่ 4 : แสดงรูปการเปรียบเทียบความเสี่ยงในการเกิดโรคเบาหวานโดยแบ่งตามเพศ

จากภาพที่ 4 แสดงตัวอย่างการเปรียบเทียบความเสี่ยงในการเกิดโรคเบาหวาน โดยแบ่งตามช่วงของค่าดัชนีทั้ง 5 ระดับ โดยแบ่งตามเพศชายและเพศหญิง เพื่อทำการเปรียบเทียบและวิเคราะห์หาสาเหตุของความเสี่ยงในการเกิดโรคเบาหวานของเพศชายและเพศหญิง

**4. ผลการดำเนินงาน**

ผลการพัฒนาระบบสนับสนุนการตัดสินใจการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวานด้วยเทคนิคเหมืองข้อมูล สามารถแบ่งออกได้ดังนี้

**4.1 ผลการเปรียบเทียบอัลกอริทึม**

นำข้อมูลมาทำการวิเคราะห์ความเสี่ยงในการเกิดโรคเบาหวาน โดยทำการเปรียบเทียบด้วยอัลกอริทึม

ทั้งหมด 5 อัลกอริทึม ได้แก่ เทคนิคต้นไม้ป่าสุ่ม, เทคนิคต้นไม้ตัดสินใจ, เทคนิคโครงข่ายประสาทเทียม, เทคนิคนาอิวเบย์, และเทคนิคเพื่อนบ้านใกล้ที่สุด

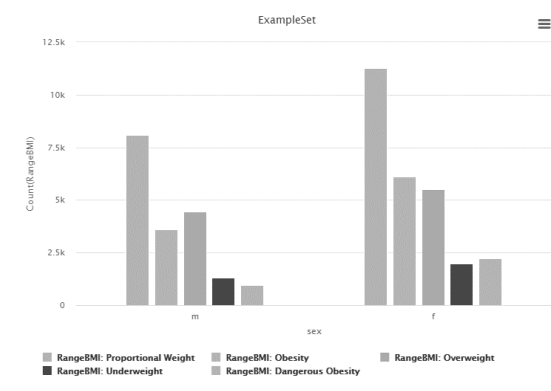
ตารางที่ 1 ผลการเปรียบเทียบอัลกอริทึม

เทคนิค	Performance			
	Accuracy	Precision	Recall	F-Measure
Random Forest	99.68	99.85	99.11	99.48
Decision Tree	94.70	97.84	85.90	91.48
KNN	94.16	94.68	92.53	93.59
Naïve Bayes	89.94	89.06	83.92	86.41
Neural Network	55.40	49.13	40.29	44.27

จากการทดสอบทั้ง 5 อัลกอริทึม พบว่าอัลกอริทึมต้นไม้ป่าสุ่มให้ผลลัพธ์ที่ดีที่สุด โดยให้ค่าความถูกต้องอยู่ที่ 99.68% ค่าความแม่นยำอยู่ที่ 99.85% ค่าความระลึกอยู่ที่ 99.11% และค่าประสิทธิภาพอยู่ที่ 99.48% รองลงมาอัลกอริทึมต้นไม้ตัดสินใจ โดยให้ค่าความถูกต้องอยู่ที่ 94.70% ค่าความแม่นยำอยู่ที่ 97.84% ค่าความระลึกอยู่ที่ 85.90% และค่าประสิทธิภาพอยู่ที่ 91.48%

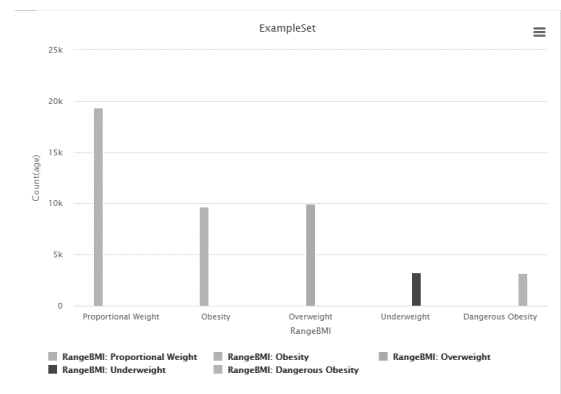
4.2 ผลการพัฒนาหน้ารายงานแสดงผล

การแสดงผลข้อมูลได้ทำการกำหนดค่าของช่วงดัชนี 5 ระดับ ได้แก่ น้ำหนักต่ำกว่าเกณฑ์, น้ำหนักตามสัดส่วน, น้ำหนักเกิน, โรคอ้วน และ โรคอ้วนอันตราย



ภาพที่ 5 : แสดงรูปการเปรียบเทียบความเสี่ยงในการเกิดโรคเบาหวานโดยแบ่งตามเพศ

จากภาพที่ 5 แสดงตัวอย่างการเปรียบเทียบความเสี่ยงในการเกิดโรคเบาหวาน โดยแบ่งตามช่วงของค่าดัชนีทั้ง 5 ระดับ โดยทำการแสดงจำนวนข้อมูลระหว่างเพศชายและเพศหญิงเพื่อทำการเปรียบเทียบและการวิเคราะห์สาเหตุในการเกิดโรคเบาหวาน พบว่าเพศหญิงมีจำนวนข้อมูลในการวิเคราะห์ และมีข้อมูลความเสี่ยงมากกว่าเพศชาย และมีความเสี่ยงในการเกิดโรคเบาหวานมากกว่าเพศชายอีกด้วย



ภาพที่ 6 : แสดงรูปการเปรียบเทียบความเสี่ยงในการเกิดโรคเบาหวานโดยแบ่งตามช่วงค่าดัชนี

จากภาพที่ 6 แสดงตัวอย่างการเปรียบเทียบความเสี่ยงในการเกิดโรคเบาหวาน โดยแบ่งตามช่วงของค่าดัชนีทั้ง 5 ระดับ โดยทำการแสดงและแบ่งจำนวนความเสี่ยงในการเกิดโรคเบาหวาน ซึ่งพบว่าข้อมูลที่ใช้ในการวิเคราะห์นั้นมีข้อมูล น้ำหนักต่ำกว่าเกณฑ์มากที่สุด รองลงมาน้ำหนักเกิน น้ำหนักตามสัดส่วน โรคอ้วน และโรคอ้วนอันตราย เรียงข้อมูลตามลำดับ

5. สรุปอภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อทำการศึกษาของผู้ที่มีความเสี่ยงในการเกิดโรคเบาหวาน โดยการพยากรณ์ความเสี่ยงซึ่งใช้เทคนิคเหมืองข้อมูลในการพยากรณ์และทำการศึกษาเพื่อทำการเปรียบเทียบอัลกอริทึมที่ดีที่สุดเพื่อนำอัลกอริทึมไปใช้ในการพัฒนาระบบสนับสนุนการตัดสินใจการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวานด้วยเทคนิคเหมืองข้อมูล

ในการสร้างแบบการพยากรณ์ข้อมูลโดยใช้เทคนิคเหมืองข้อมูลในการพยากรณ์ทั้ง 5 เทคนิคที่มาใช้

ในการวิเคราะห์ พบว่าสำหรับข้อมูลชุดนี้เหมาะสมกับเทคนิคต้นไม้ป่าสุ่มมากที่สุด โดยให้ค่าความถูกต้องอยู่ที่ 99.68% ค่าความแม่นยำอยู่ที่ 99.85% ค่าความระลึกละเอียดอยู่ที่ 99.11% และค่าประสิทธิภาพอยู่ที่ 99.48% หลังจากนั้นนำเทคนิคนี้ไปใช้ในการพัฒนาระบบและแสดงผลรูปแบบต่างๆ บนระบบที่ถูกพัฒนาขึ้นมา

## 6. ข้อเสนอแนะในการวิจัย

6.1 ในการวิจัยครั้งนี้เป็นการศึกษาความเสียหายในการเกิดโรคเบาหวาน โดยได้รับข้อมูลจากกรมอนามัย กระทรวงสาธารณสุข เพื่อนำมาวิเคราะห์และพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวาน เพื่อให้ทราบถึงความเสียหายเบื้องต้นของการเกิดโรคเบาหวาน

### เอกสารอ้างอิง

- [1] ณัฐพล แสนคำ, ทิพวัลย์ แสนคำ และชนากร ปุรารัมย์ “ระบบ สนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูล” วารสารวิชาการ โรงเรียนนายร้อยพระจุลจอมเกล้า ปีที่ 15, 2560, หน้า 161 – 170
- [2] ณัฐวดี หงษ์บุญมี และ ประภาสิริ ศรีพานิชกุล, “การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ด้วยเทคนิคเหมืองข้อมูล” บทความวิจัยวารสารวิทยาศาสตร์และเทคโนโลยี เล่มที่ 9 ครั้งที่ 1, 2562, หน้า 41 – 51
- [3] วนิตา พงษ์สงวน, ทิพย์ถินสูงเนิน และมาโนช ถินสูงเนิน “การพัฒนาแบบจำลองปัจจัยที่มีผลต่อการเป็นโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจ” วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม ปีที่ 1 ฉบับที่ 1, 2561
- [4] ศิรินทรา เสือพิทักษ์, บุญมี นิสสัยดี และ วิภาวรรณ บัวทอง “การเปรียบเทียบเทคนิคการจำแนกประเภทข้อมูลสำหรับทำนายแนวโน้มการสำเร็จการศึกษาของนักเรียน” วารสารวิชาการชายันท์ มจร. ภูเก็ต ปีที่ 5 ฉบับที่ 2, 2564, หน้า 42 – 50
- [5] อติสรณ์ กำลิ่งเพชร, พนิ หล่อวงศ์ตระกูล และ อริยะ นามวงศ์ “การประยุกต์ใช้โครงข่ายประสาทเทียมโดยใช้ตัวชี้วัดทางเทคนิคเพื่อการลงทุนในตลาด

หลักทรัพย์แห่งประเทศไทย” วารสารเทคโนโลยีสารสนเทศ ปีที่ 11 ฉบับที่ 1, 2558 หน้า 53 – 63

- [6] อนันต์ ชัยชูติภาสเจริญ และ ดร.จรัญ แสนราช “การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไป” วารสารวิจัยมหาวิทยาลัยขอนแก่น(ฉบับบัณฑิตศึกษา) สาขามนุษยศาสตร์และสังคมศาสตร์ ปีที่ 6 ฉบับที่ 3, 2561, หน้า 105 - 113

# การพัฒนาไลน์แชทบอทเพื่อให้ข้อมูลข่าวสารโรคติดต่อไวรัสโคโรนา

## The Development of ChatBot Line to Provide Information about Coronavirus

กัญจน์ เข็มมณ (Kan Khemnak)<sup>1</sup> และณัฐวี อดุกฤษฎ์ (Nattavee Utakrit)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>S6307021856012@email.kmutnb.ac.th, <sup>2</sup>nattavee.u@itd.kmutnb.ac.th

### บทคัดย่อ

ปัจจุบันการเผยแพร่ข้อมูลข่าวสารมีการเผยแพร่ผ่านช่องทางต่างๆมากมาย เทคโนโลยีทำให้ผู้คนสามารถรับรู้ข้อมูลข่าวสารได้รวดเร็ว และง่ายต่อการเข้าถึง โดยเฉพาะอย่างยิ่งในช่วงสถานการณ์การแพร่ระบาดของไวรัสโคโรนา (Covid-19) งานวิจัยนี้นำเสนอไลน์แชทบอทเพื่อให้ข้อมูลข่าวสารโรคติดต่อไวรัสโคโรนา โดยได้ทำการเปรียบเทียบวิธีการประมวลผลข้อความ 3 อัลกอริทึม คือ 1) ฟัซซี่วูซซี่ (Fuzzy Wuzzy) 2) แจ็กการ์ด (Jaccard) และ 3) การประมวลผลภาษาทางธรรมชาติ โมดูล ไพไทยเอ็นแอลพี (PythaiNLP) เพื่อนำมาพัฒนาการตอบคำถามผ่านไลน์แชทบอทให้มีความแม่นยำ ซึ่งพบว่า ฟัซซี่ วูซซี่ ให้ค่าความแม่นยำมากที่สุด จากนั้นจึงได้นำมาพัฒนาร่วมกับแอปพลิเคชันแชทบอทเพื่อตอบคำถามและเผยแพร่ข้อมูลข่าวสารที่มีความรวดเร็ว ถูกต้องและแม่นยำ ซึ่งผลการทดสอบประสิทธิภาพการทำงานของไลน์แชทบอทโดยผู้เชี่ยวชาญพบว่า มีประสิทธิภาพอยู่ในระดับดีมากที่ร้อยละ 82.75 และผู้ใช้งานงานทั่วไปมีความพึงพอใจอยู่ในระดับดีมาก

คำสำคัญ: ไลน์แชทบอท เว็บไซต์ ไวรัสโคโรนา ฟัซซี่วูซซี่

### Abstract

Currently, the dissemination of information is distributed through many channels. Technology has made people aware of information faster and easier to access. Especially during the coronavirus (Covid-19) epidemic, this research presents a ChatBot Line to provide information on coronavirus by comparing three algorithms for processing messages 1) Fuzzy

Wuzzy Algorithm, 2) Jaccard Algorithm, and 3) Natural Language Processing PythaiNLP to improve the accuracy of answering questions via Chatbot Line which found the Fuzzy Wuzzy Algorithm to be the most accurate. Then developed with a chatbot application to answer questions and disseminate information with speed, accuracy and precision. The results of testing the performance of the ChatBot Line by experts showed that the efficiency was very good at 82.75 percent and the users' satisfaction rating were very good.

**Keywords:** ChatBot Line, Web page, Coronavirus, Fuzzy Wuzzy Algorithm.

### 1. บทนำ

จากสถานการณ์การระบาดของโรคติดต่อไวรัสโคโรนา ได้ระบาดไปทั่วโลก รวมถึงในประเทศไทยทำให้ความรู้และข้อมูลที่ต้องการแม่นยำกับประชาชนเป็นเรื่องที่สำคัญในการที่จะให้ประชาชนเตรียมพร้อมเพื่อรับมือกับสถานการณ์การแพร่ระบาดได้อย่างเหมาะสม ปัจจุบันหลายหน่วยงานได้นำเสนอข้อมูลในรูปแบบต่างๆ ผ่านช่องทางที่หลากหลาย เช่น ศูนย์บริหารสถานการณ์แพร่ระบาดของโรคติดต่อไวรัสโคโรนา (ศบค.มท.) หรือผ่านทางเว็บไซต์ต่างๆ การสื่อสาร 2019 ข้อมูลผ่านไลน์ เป็นอีกช่องทางหนึ่ง queเข้าถึงผู้คนได้โดยง่าย เนื่องจากเป็นแอปพลิเคชันที่มีผู้นิยมใช้กันเป็นจำนวนมาก

ปัญญาประดิษฐ์ (Artificial Intelligence) เป็น วิทยาการด้านปัญญาที่จะมาช่วยให้นักวิจัยแก้ปัญหา วิเคราะห์ และให้ข้อมูลต่างๆ ที่แม่นยำ รวดเร็ว ในปัจจุบันมีการศึกษาและนำเทคนิคการประมวลผลตามหลักการของปัญญาประดิษฐ์มาใช้



หลายอัลกอริทึม แต่จากการศึกษาของผู้วิจัยพบว่า อัลกอริทึม ฟัซซี่วูซซี่ (Fuzzy Wuzzy), แจ็คการ์ด (Jaccard), การประมวลผลภาษาทางธรรมชาติโมดูลไพไทยเอ็นแอลพี (PyThaiNLP) มีการนำประยุกต์ใช้เพื่อวิเคราะห์และให้ข้อมูล โดยเฉพาะอย่างยิ่งการเปรียบเทียบความเหมือนทุกตัวอักษรของ ฟัซซี่วูซซี่ (Fuzzy Wuzzy) ให้ค่าเฉลี่ยอยู่ที่ร้อยละ 80.65 เมื่อเปรียบเทียบกับ การประมวลผลแบบภาษาธรรมชาติ (NLP) ให้ค่าเฉลี่ยอยู่ที่ร้อยละ 46.47 [1] แต่จากการศึกษาพบว่า การประมวลผลภาษาธรรมชาติโมดูล PyThaiNLP นั้นเป็นการประมวลผลแบบภาษาธรรมชาติที่ทำมาเพื่อใช้กับภาษาไทยโดยเฉพาะและเป็นอัลกอริทึมประเภท Library คล้ายกัน ในด้านการสืบค้นข้อมูล Jaccard ใช้ในการเปรียบเทียบความเหมือนของตัวอักษรที่ไม่ซับซ้อนมากสามารถทำได้รวดเร็วและแม่นยำ [2] ผู้วิจัยจึงได้นำ 3 อัลกอริทึมนี้มาศึกษาวิเคราะห์เปรียบเทียบเพื่อหาอัลกอริทึมที่เปรียบเทียบตัวอักษรได้แม่นยำที่สุดให้เหมาะสมกับการให้ข้อมูลข่าวสารจากนั้นนำไปประยุกต์ใช้ในการพัฒนาไลน์แชทบอท ซึ่งการให้ข้อมูลจะใช้ข้อมูลจากทุกโรงพยาบาลที่อยู่ในสังกัดสำนักงานปลัดกระทรวงสาธารณสุข

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

พงศกร อุทราจิตต์ [1] นำเสนอการพัฒนาหุ่นยนต์สนทนาสำหรับแนะนำคุณแม่่มือใหม่สำหรับให้คุณแม่่มือใหม่ได้มีผู้ช่วยแนะนำเกี่ยวกับการตั้งครรภ์และได้รับข้อมูลข่าวสารการประชาสัมพันธ์ต่างๆที่จำเป็นอีกทั้งยังสะดวกและรวดเร็วในการเข้าถึงได้ทุกที่ทุกเวลา ช่วยลดปัญหาในการขอคำแนะนำและเพื่ออำนวยความสะดวกในการให้ข้อมูลต่างๆ โดยใช้ฟัซซี่วูซซี่ (Fuzzy Wuzzy) ในการประมวลผลข้อความเพื่อให้หุ่นยนต์สนทนาตอบคำถามได้เหมาะสม

การประมวลผลภาษาธรรมชาติ (Natural language processing -NLP) เป็นการประมวลผลข้อความระหว่างภาษาคอมพิวเตอร์และภาษาธรรมชาติมนุษย์ ในปัจจุบันการประมวลผลภาษาธรรมชาติมีการนิยมใช้ขั้นตอนการเรียนรู้เชิงคุณลักษณะ [3] [4] ที่สามารถให้ผลลัพธ์ที่ออกมาได้ดี ตัวอย่างเช่นในแบบจำลองภาษา โดยได้มีการเปรียบเทียบกับวิธีการดั้งเดิมที่

ใช้การประมวลผลภาษาธรรมชาติในด้านสถิติ หรือการประมวลผลตามกฎที่วางไว้

User Interface Design หรือ Human-Computer Interaction คือ การออกแบบส่วนต่อประสานกับผู้ใช้ ระหว่างผู้ใช้กับคอมพิวเตอร์ เป็นหลักและวิธีการที่ทำให้มนุษย์และคอมพิวเตอร์สามารถสื่อสารกันได้ง่ายดายเสมือนสื่อสารกับมนุษย์ ผ่านการมองเห็น การได้ยิน การสัมผัส การพุดคุย หรือวิธีการใดๆ [5] ซึ่งผู้วิจัยได้นำหลักการนี้มาประยุกต์ในการออกแบบ User Interface ของ Chatbot Line ในงานวิจัยนี้

ฟัซซี่วูซซี่ (Fuzzy Wuzzy) เป็นไลบรารีของ Python ซึ่งใช้สำหรับเปรียบเทียบความเหมือนของตัวอักษรโดยการตรวจสอบตัวอักษรกับข้อความต้นทางเพื่อคำนวณความแตกต่างระหว่างลำดับของตัวอักษร โดยส่วนมากนิยมใช้กับการวิเคราะห์ข้อมูลขนาดใหญ่ หรือด้าน Data Science [6] ฟัซซี่วูซซี่ (Fuzzy Wuzzy) มีการจับคู่ประมวลผล 4 รูปแบบได้แก่ 1) Simple Ratio เปรียบเทียบทุกตัวอักษร 2) Partial Ratio เปรียบเทียบข้อความที่มีความยาวไม่เท่ากันได้ 3) Token Sort Ratio เปรียบเทียบข้อความที่สลับคำได้ 4) Token Set Ratio เปรียบเทียบข้อความที่มีคำซ้ำกันได้

แจ็คการ์ด (Jaccard) เป็นอัลกอริทึมเปรียบเทียบข้อความที่เป็นประเภทเดียวกับฟัซซี่วูซซี่ (Fuzzy Wuzzy) และเป็นไลบรารีหนึ่งของ Python ที่นิยมนำมาใช้ในการค้นคืนข้อมูลหรือทำ Search Engine เช่น Google, Yahoo, Bing เป็นต้น [2] แจ็คการ์ด (Jaccard) มีลักษณะการประมวลผลเปรียบเทียบข้อความในรูปแบบการแบ่งคำหาคำแทนที่เหมือนกันเพื่อนำมาคำนวณหาความเหมือนของข้อความที่เปรียบเทียบกับข้อความต้นทาง ค่าที่ได้จะมีตั้งแต่ 0-1 เรียงจากความเหมือนของข้อความจากน้อยไปมาก

## 3. วิธีดำเนินการวิจัย

### 3.1 การศึกษาและเปรียบเทียบเทคนิคของอัลกอริทึม

ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับขั้นตอนการประมวลผลของภาษาธรรมชาติ 3 อัลกอริทึม ได้แก่ 1) ฟัซซี่วูซซี่ (Fuzzy Wuzzy) 2) แจ็คการ์ด (Jaccard) และ 3) การประมวลผลภาษาทางธรรมชาติโมดูลไพไทยเอ็นแอลพี (PyThaiNLP) เปรียบเทียบหาความแม่นยำในการเปรียบเทียบ

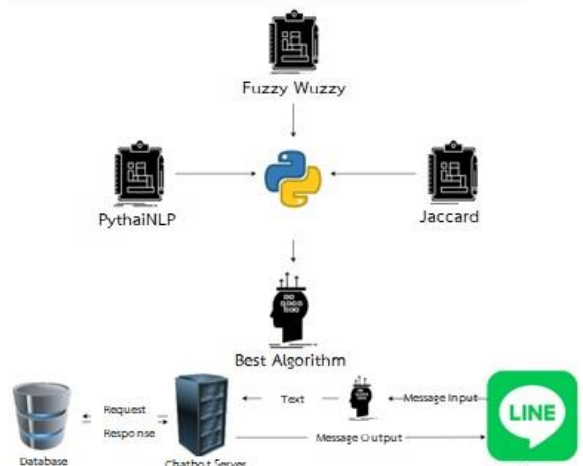
ความเหมือนของตัวอักษรเพื่อให้ไลน์แชทบอทสามารถโต้ตอบและให้ข้อมูลโรคติดเชื้อไวรัสโคโรนาได้อย่างถูกต้องและตรงตามที่ผู้ใช้ต้องการ โดยใช้การกำหนดรูปแบบการพิมพ์ที่มีความคิดพลาดจากการพิมพ์ตัวอักษรที่ใกล้เคียงบนเป็นพิมพ์โทรศัพท์มือถือให้ความหมายอยู่ในขอบเขตของเมนูบนไลน์แชทบอทผ่านการประมวลผลภาษา Python

**3.2 การเตรียมข้อมูลและคัดเลือกข้อมูล**

ข้อมูลที่นำมาใช้งานวิจัยครั้งนี้ เป็นการรวบรวมข้อมูลจากโรงพยาบาลสังกัดสำนักงานปลัดกระทรวงสาธารณสุขทั่วประเทศตั้งแต่เริ่มระบบ Co-ward อยู่ในช่วงเดือนมิถุนายน พ.ศ.2563 จนถึงปัจจุบัน ซึ่งเป็นข้อมูลการเข้ารับการรักษาของผู้ป่วยโรคติดเชื้อไวรัสโคโรนาได้แก่ข้อมูลอาการของผู้ป่วยในแต่ละวัน ข้อมูลการใช้เตียงของผู้ป่วยในแต่ละวัน ข้อมูลโรงพยาบาลในสังกัดสำนักงานปลัดกระทรวงสาธารณสุข ข้อมูลวัคซีนที่ได้รับ ข้อมูลยาที่ได้รับ ข้อมูลเครื่องช่วยหายใจของผู้ป่วยในแต่ละวัน จำนวน 1 ล้านรายการ

**3.3 กรอบแนวคิดการวิจัย**

ผู้วิจัยได้วางกรอบแนวคิดในการพัฒนาไลน์ แชทบอทซึ่งประกอบไปด้วย การเตรียมข้อมูลที่ใช้ในการตอบโต้กับผู้ใช้ งาน การเปรียบเทียบ ฟัซซี่วูซซี่ (Fuzzy Wuzzy), แจ็คการ์ด (Jaccard), การประมวลผลภาษาทางธรรมชาติ โมดูลไพไทยเอ็นแอลพี (PythaiNLP) เพื่อหาความเหมือนของตัวอักษรของผู้ใช้งานที่พิมพ์เข้ามา การนำอัลกอริทึมที่มีความแม่นยำมากที่สุดมาประยุกต์ใช้กับไลน์แชทบอท การแสดงผลข้อมูลผู้ติดเชื้อไวรัสโคโรนาผ่านทางไลน์แชทบอท 2019



ภาพที่ 1 : แสดงรูปกรอบแนวคิดการวิจัย

**3.4 การออกแบบและพัฒนาระบบ**

นำอัลกอริทึมที่มีความแม่นยำในการเปรียบเทียบความเหมือนของตัวอักษรมากที่สุดมาใช้ในการออกแบบและพัฒนาไลน์แชทบอท Chatbot Server จะถูกเชื่อมต่อกับฐานข้อมูล Replicate ของระบบ Co-ward ที่เก็บข้อมูลผู้ติดเชื้อไวรัสโคโรนา แบบ 2019 Real-Time รวมถึงมี Security Rules และระบบ Authentication ที่สามารถระบุตัวตนเพื่อเข้าถึงข้อมูลของระบบในแต่ละระดับเพื่อป้องกันไม่ให้ผู้ไม่ประสงค์ดีทำการโจรกรรมข้อมูลจากฐานข้อมูลกลางได้ ศึกษาการทำงานของ Line Messaging API เพื่อให้สามารถรับส่งข้อมูลระหว่างเซิร์ฟเวอร์ การส่งคำขอจะใช้ API Restful Service แบบ JSON โดย Messaging API ทำการเชื่อมต่อข้อมูลผู้ใช้งานผ่านทาง LINE Official Account ซึ่ง Messaging API จะสามารถยอมรับคำขอเป็นเพื่อน โดยผ่านหน้า LINE Manager การโต้ตอบการใช้งานกับผู้ใช้จะส่งข้อมูลผ่าน Chatbot Server ของเราไปยังผู้ใช้งาน LINE การออกแบบส่วนติดต่อกับผู้ใช้งาน GUI บน Line Application (Rich Menu) โดยใช้เครื่องมือ LINE Bot Designer ของโปรแกรม LINE สำหรับออกแบบหน้าเมนูไลน์แชทบอท สามารถออกแบบรูปแบบของเมนูได้ตามความต้องการ หลังจากนั้นให้เชื่อมต่อ API Restful Service ผ่านทาง Url Rich Menu ของ LINE ที่ทำการออกแบบไว้ในแต่ละปุ่มการทำงาน

**4. ผลการดำเนินงาน**

ผลการดำเนินงานไลน์แชทบอทเพื่อให้ข้อมูลโรคติดต่อไวรัสโคโรนาสามารถแบ่งหัวข้อได้ดังนี้

**4.1 ผลการเปรียบเทียบอัลกอริทึม**

ทำการเปรียบเทียบอัลกอริทึม โดยนำชุดข้อมูลตัวอย่างมาเพื่อทดสอบประสิทธิภาพ โดยการเปรียบเทียบข้อความของเมนูจำนวนผู้ติดเชื้อวันนี้ ตัวอย่างชุดข้อมูลแสดงดังตารางที่ 1

ตารางที่ 1 : ตัวอย่างข้อมูลที่นำมาทดสอบ

ที่	คำค้นหา
1	จำนวนผู้ติดเชื้อของวันนี้
2	จำนวนผู้ติดเชื้อวันนี้
3	จำนวนผู้ติดเชื้อของวันนี้
4	จำนวนผู้ติดเชื้อวันนี้

ที่	คำค้นหา	Fuzzy Wuzzy	Jaccard	PythaiNLP
1	จำนวนผู้ติดเชื้อของวันนี้	72	70	42.48
2	จำนวนผู้ติดเชื้อวันนี้	74	73.68	41.23
3	จำนวนผู้ติดเชื้อของวันนี้	72	78.95	37.35
4	ยำนวนผู้ติดเชื้อวันนี้	86	60	17.6
5	จำนวนผู้ติดเชื้อของวันนี้	72	77.78	62.84
6	คำนวนผู้ติดเชื้อวรัว	83	77.78	28.72
7	จำนวนผู้ติดเชื้อวันนี้	100	100	99.99
8	โรงพยาบาลที่ใกล้เคียง	25	10.34	20.35
9	โนงพยาบาลที่ใกล้เคียง	28	18.52	25.95
10	ผู้ติดเชื้อ	100	62.5	71.91
11	จำนวนผู้ติดเชื้อในวันนี้	89	88.89	84.38
12	จนวนผู้ติดเชื้อวันนี้	94	93.75	56.04
13	จำนวนคนติดเชื้อ	86	70.59	74.36
14	ผู้ติดเชื้อล่าสุด	85	70.59	55.6
15	จำนวนคนติดเชื้อล่าสุด	71	47.62	66.53
16	จำนวนคนติดเชื้อ Covid ล่าสุด	72	54.55	72.13
17	จำนวนผู้ติดเชื้อ Covid ล่าสุด	56	25	63.45
18	ผู้ติดเชื้อรายวัน	83	51.58	87.97
19	จำนวนผู้ติดเชื้อรายวัน	86	68.42	69.38
20	คนติดเชื้อวันนี้	83	78.95	90.16
5	จำนวนผู้ติดเชื้อของวันนี้			
6	คำนวนผู้ติดเชื้อวรัว			
7	จำนวนผู้ติดเชื้อวันนี้			
8	โรงพยาบาลที่ใกล้เคียง			
9	โนงพยาบาลที่ใกล้เคียง			
10	ผู้ติดเชื้อ			
11	จำนวนผู้ติดเชื้อในวันนี้			
12	จนวนผู้ติดเชื้อวันนี้			
13	จำนวนคนติดเชื้อ			
14	ผู้ติดเชื้อล่าสุด			
15	จำนวนคนติดเชื้อล่าสุด			
16	จำนวนคนติดเชื้อ Covid ล่าสุด			
17	จำนวนผู้ติดเชื้อ Covid ล่าสุด			
18	ผู้ติดเชื้อรายวัน			
19	จำนวนผู้ติดเชื้อรายวัน			
20	คนติดเชื้อวันนี้			

จากนั้นนำชุดข้อมูลตัวอย่างมาทำการเปรียบเทียบด้วย ฟัซซี่ วูซซี่ (Fuzzy Wuzzy), แจ็คการ์ด (Jaccard), การประมวลผลภาษาทางธรรมชาติ โมดูลไพไทยเอ็นแอลพี (PythaiNLP) โดยผลการเปรียบเทียบแสดงในตารางที่ 2

ตารางที่ 2 : ผลการเปรียบเทียบความแม่นยำจากการเปรียบเทียบข้อความกับข้อความค้นหา “จำนวนผู้ติดเชื้อวันนี้” โดยใช้การประมวลผลของอัลกอริทึม

จากตารางที่ 2 ฟัซซี่ วูซซี่ (Fuzzy Wuzzy) ประมวลผลความเหมือนของข้อความด้วยการเปรียบเทียบข้อความค้นหากับคำค้นหาได้ค่ามากที่สุดจากการเปรียบเทียบทีละตัวอักษรและคำนวณออกมาเป็นค่าความเหมือนที่มีค่า 0-100 ทำให้ไลน์แชทบอทรู้ว่าผู้ใช้งานต้องการดูข้อมูลในส่วนไหนของเมนู โดยกำหนดเมนูต่างๆ ไว้เป็นข้อความ เช่น ผู้ติดเชื้อล่าสุด ข้อมูลวัคซีนที่ได้รับ โรงพยาบาลที่ใกล้เคียง หากผู้ใช้งานพิมพ์ข้อความที่ใกล้เคียงกับเมนูที่กำหนดไว้โดยค่าความเหมือนมากกว่า 60 ไลน์แชทบอทจะตอบกลับไปยังผู้ใช้งาน โดยให้ข้อมูลตามความเหมาะสม

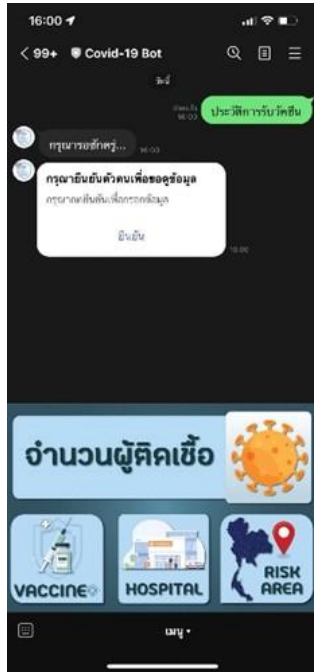
#### 4.2 ผลการพัฒนาระบบ

การพัฒนาไลน์แชทบอทแบ่งได้เป็น 2 ส่วน ได้แก่ หน้าเว็บไซต์กรอกข้อมูลเพื่อใช้ยืนยันตัวตน และส่วนของการประมวลผลข้อความโดยใช้อัลกอริทึมที่มีความแม่นยำมากที่สุด เมื่อผู้ใช้งานครั้งแรกเพิ่มเพื่อนไลน์แชทบอทจะไม่สามารถดูข้อมูลต่างๆ ได้ ผู้ใช้จะต้องทำการยืนยันตัวตนเพื่อพิสูจน์ว่าเป็นเจ้าของข้อมูลจริง



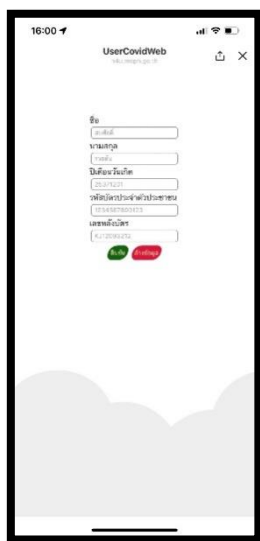
ภาพที่ 2 : ไลน์แชทบอทส่งข้อความต้อนรับให้กับผู้ใช้งาน

เมื่อผู้ใช้งานได้รับข้อความต้อนรับแล้ว ในขั้นตอนต่อไปจะต้องทำการยืนยันตัวตน โดยกดปุ่มยืนยันเพื่อเปิดเว็บไซต์สำหรับกรอกข้อมูลพิสูจน์ตัวตน ดังภาพที่ 3



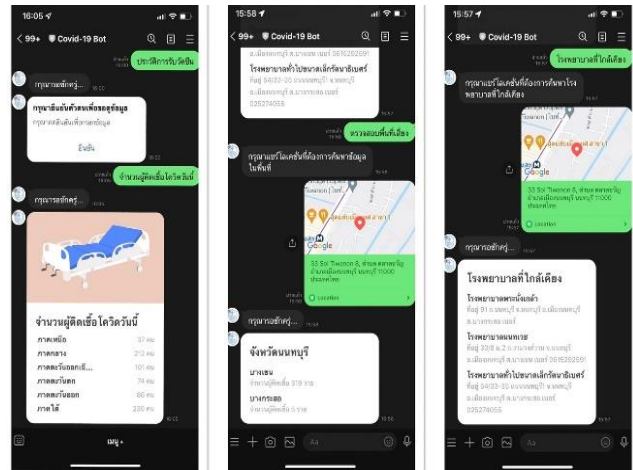
ภาพที่ 3 : โลกแชทบอทแสดงปุ่มยืนยันตัวตน

การยืนยันตัวตนจะประกอบไปด้วยเลขบัตรประจำตัวประชาชน ชื่อ นามสกุล เลขหลังบัตรประจำตัวประชาชน วันเดือนปีเกิด โดยนำข้อมูลดังกล่าวตรวจสอบกับฐานข้อมูลกรมการปกครองตรวจสอบความถูกต้องเพื่อพิสูจน์ตัวตนดังภาพที่ 4



ภาพที่ 4 : หน้ากรอกข้อมูลยืนยันตัวตน

เมื่อยืนยันตัวตนสำเร็จข้อมูลจะถูกจัดเก็บในฐานข้อมูลเพื่อบันทึกว่าผู้ใช้งานผ่านการยืนยันตัวตนเรียบร้อยแล้ว จากนั้นสามารถเลือกเมนูการแสดงผลต่างๆ ได้ตามภาพที่ 5



ภาพที่ 5 : โลกแชทบอทแสดงผลเมนูต่างๆ

โลกแชทบอทสามารถให้ข้อมูลโรคติดต่อไวรัสโคโรนาได้ 2 รูปแบบคือคลิกที่เมนูที่ต้องการหรือการพิมพ์ข้อความ การคลิกที่เมนูโลกแชทบอทเป็นการส่งข้อความที่ตั้งค่าไว้บน Lich Menu เพื่อให้โลกแชทบอทส่งข้อความกลับมาเป็นข้อมูลที่ต้องการ หรือการเลือกพิมพ์ข้อความเพื่อสอบถามข้อมูลโลกแชทบอทจะทำการเปรียบเทียบข้อความหาชื่อเมนูที่ตั้งค่าไว้ที่มีความเหมือนกับข้อความของผู้ใช้มากที่สุดและให้ข้อมูลกลับมาในรูปแบบข้อความที่ผู้ใช้ต้องการ เช่น ผู้ติดเชื้อล่าสุด ข้อมูลวัคซีนที่ได้รับ โรงพยาบาลที่ใกล้เคอ

### 5. สรุป อกิปรายผล

จากการพัฒนาโลกแชทบอทเพื่อให้ข้อมูลโรคติดต่อไวรัสโคโรนาแบ่งงานวิจัยออกเป็น 2 ส่วน ซึ่งจากการเปรียบเทียบและคัดเลือกอัลกอริทึมที่เหมาะสม 3 อัลกอริทึมได้แก่ ฟัซซี่ วูซซี่ ( Fuzzy Wuzzy), แจ็คการ์ด ( Jaccard), การประมวลผลภาษาทางธรรมชาติ โมดูลไพลีเอินแอลพี (PythaNLP) โดยการสร้างแบบจำลองทดสอบจากชุดข้อมูลตัวอย่างด้วยวิธีประมวลผลการเปรียบเทียบข้อความแล้วซึ่งผลการวิจัยพบว่าอัลกอริทึมฟัซซี่ วูซซี่ ( Fuzzy Wuzzy) ให้ค่าความเหมือนได้มากที่สุดโดยมีค่าเฉลี่ยความเหมือนของข้อความร้อยละ 75.85 ซึ่งผลที่ได้สอดคล้องกับงานวิจัยของพงศกร [1] ที่สรุปได้ว่าการเปรียบเทียบเทคนิคอัลกอริทึม

Fuzzy Wuzzy ให้คำร้อยละในการเปรียบเทียบระหว่างตัวแปรต้นกับตัวแปรตามได้เหมือนมากที่สุดเนื่องจากฟัซซี่วูซซี่ (Fuzzy Wuzzy) เป็นการเปรียบเทียบข้อความโดยแบ่งการเปรียบเทียบความเหมือนที่ละตัวอักษรเริ่มจากอักษรตัวแรกถึงอักษรตัวสุดท้ายและจาก อักษรตัวสุดท้ายกลับไปที่อักษรตัวแรกต่างจากแจ็การ์ดี (Jaccard) ที่เป็นการเปรียบเทียบโดยใช้ตัวอักษรที่มีร่วมกันระหว่างข้อความต้นทางกับข้อความค้นหา และการประมวลผลภาษาธรรมชาติ โมดูล PythaiNLP เป็นการแบ่งคำตามพจนานุกรมภาษาไทยเพื่อเปรียบเทียบข้อความต้นทางเมื่อนำวิธีฟัซซี่วูซซี่ (Fuzzy Wuzzy) ไปทำการพัฒนาในไลน์แชทบอทเพื่อให้ข้อมูลโรคติดต่อไวรัสโคโรนาจึงส่งผลให้ไลน์แชทบอทที่พัฒนาขึ้นมีประสิทธิภาพในให้ข้อมูลตอบกลับ ในรูปแบบข้อความที่ผู้ใช้ต้องการโดยผลทดสอบและการประเมินประสิทธิภาพการทำงานไลน์แชทบอท โดยผู้เชี่ยวชาญด้านฐานข้อมูลโรคติดต่อไวรัสโคโรนาจำนวน 3 ท่านประเมินจากความถูกต้องของการตอบคำถามของไลน์แชทบอทยืนยันว่ามีประสิทธิภาพอยู่ในระดับดีมากที่ร้อยละ 82.75 ในส่วนของการทดสอบโดยผู้ใช้งานทั่วไปพบว่าต่างพึงพอใจต่อระบบไลน์แชทบอทนี้ในระดับดีมาก โดยมีค่าเฉลี่ยความพึงพอใจในภาพรวมเท่ากับ 4.37 และ SD เท่ากับ 0.75

## 6. ข้อเสนอแนะในการวิจัย

การวิจัยนี้ได้แสดงให้เห็นถึงการทำงานของนแชทบอทร่วมกับอัลกอริทึมฟัซซี่วูซซี่ มีประสิทธิภาพการทำงานที่แม่นยำในการให้ข้อมูล แต่อย่างไรก็ตามการวิจัยครั้งนี้ยังพบปัญหาในส่วนของข้อความที่มีความยาวและมีความซับซ้อนของข้อความสูง เนื่องด้วยข้อความเป็นภาษาไทยมีความยากมากกว่าภาษาอังกฤษซึ่งในการประมวลผลเปรียบเทียบข้อความที่ใช้การประมวลผลเปรียบเทียบตัวอักษรอาจจะให้ค่าความแม่นยำที่น้อยสำหรับการเปรียบเทียบข้อความที่มีความซับซ้อนสูง ยังต้องศึกษาพัฒนาเพิ่มเติมในส่วนของการทำงานเปรียบเทียบคำภาษาไทยที่ถูกหลักตามพจนานุกรมเพื่อแก้ไขข้อความที่ผิดให้ถูกต้อง ก็จะทำให้การตอบคำถามไลน์แชทบอทมีประสิทธิภาพดียิ่งขึ้น

## เอกสารอ้างอิง

- [1] พงศกร อุทราจิตต์. การพัฒนาหุ่นยนต์ สนทนาสำหรับแนะนำคุณแม่่มือใหม่, 2562. ปัญหาพิเศษวิทยาศาสตรมหาบัณฑิต (ระบบสารสนเทศเพื่อการจัดการ) คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- [2] Suphakit Niwattanakul. Using of Jaccard Coefficient for Keywords Similarity, 2556. Proceedings of the International MultiConference of Engineers and Computer Scientists.
- [3] รัตนาวัลย์ ไม้สัก. แอปพลิเคชันแชทบอทเพื่อการส่งเสริมการท่องเที่ยวเชิงเกษตรและวัฒนธรรม, 2562. คณะบริหารธุรกิจ มหาวิทยาลัยรามคำแหงพระนคร.
- [4] สันศักดิ์ พรหมแทน. ระบบแนะนำอย่างเชี่ยวชาญสำหรับหญิงตั้งครรภ์โดยใช้กระบวนการประมวลผลภาษาธรรมชาติ, 2563. วิทยานิพนธ์ นวัตกรรมและการเปลี่ยนแปลงสู่ยุคดิจิทัล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์.
- [5] Shneiderman Ben, Plaisant Catherine. Design the user interface : Strategic for effective human- computer interaction. Pearson Education, 2005. University of Maryland.
- [6] Mohit Kumar (2563). What is a fuzzy-wuzzy algorithm?. Available on line at <https://www.quora.com/What-is-a-fuzzy-wuzzy-algorithm>

# การพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

## The Development of Asset Trading Platform using Non-Fungible Token

จิตภา คำรงค์สมบัติ (Jidapa Dumrongsombut)<sup>1</sup> และพุดสดี บุญรอด (Pudsadee Boonrawd)<sup>2</sup>

<sup>1,2</sup>ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

s6307011858152@email.kmutnb.ac.th, pudsadee.b@itd.kmutnb.ac.th

### บทคัดย่อ

การซื้อขายสินทรัพย์ดิจิทัลเริ่มเป็นที่นิยมมากขึ้นในปัจจุบัน แต่แพลตฟอร์มส่วนใหญ่ที่ให้บริกาวยังไม่มีกระบวนการตรวจสอบความเป็นต้นฉบับของสินทรัพย์ดิจิทัล ส่งผลให้สินทรัพย์ดิจิทัลไม่มีมูลค่าและไม่มีการใช้ประโยชน์จากการเป็นเจ้าของ งานวิจัยนี้จึงได้พัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะบนเครือข่ายอีเธอเรียมที่พัฒนาขึ้นเพื่อรองรับการสร้าง การแสดงรายการขาย การซื้อ และการโอนกรรมสิทธิ์ความเป็นเจ้าของโดยไม่ผ่านตัวกลาง รวมทั้งการได้รับค่าสิทธิความเป็นเจ้าของ และผลตอบแทนจากการซื้อขายสินทรัพย์ดิจิทัล จากผลการวิจัยพบว่า แพลตฟอร์มที่พัฒนาขึ้นสามารถซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะได้ตรงตามความต้องการของผู้ใช้งาน อีกทั้งเป็นแนวทางในการประยุกต์ใช้กับงานวิจัยอื่น ๆ ที่เกี่ยวข้อง และสามารถนำไปต่อยอดในการพัฒนาแอปพลิเคชันแบบกระจายบนเครือข่ายอีเธอเรียมได้ในอนาคต

**คำสำคัญ:** แพลตฟอร์มซื้อขายสินทรัพย์ โทเคนดิจิทัล ลักษณะเฉพาะ อีเธอเรียม

### Abstract

Nowadays, Digital Asset Exchange has gained popularity for investors. Most Digital Asset Exchange Service platforms don't have an authenticity verification process which caused digital assets to have no value and no leverage on digital assets ownership. This research has developed an asset trading platform using Non-Fungible Token on the

Ethereum blockchain to support mint Non-Fungible Token, listing Non-Fungible Token, buying Non-Fungible Token, transferring ownership of assets without an intermediary, obtaining ownership rights, and earning benefits from trading digital assets. The research results found that the developed platform can exchange digital assets using a Non-Fungible Token which serves the needs of users, as well as a guideline for applying to other related research and can be used to develop distributed applications on the Ethereum Blockchain in the future.

**Keywords:** Asset Trading Platform, Non-Fungible Token, Ethereum.

### 1. บทนำ

การใช้เงินตราและสกุลเงินดิจิทัลเพื่อซื้อขายและแลกเปลี่ยนสินทรัพย์เป็นโทเคนดิจิทัลไม่มีลักษณะเฉพาะ (Fungible Token : FT) หรือโทเคนดิจิทัลทดแทนกันได้ ตัวอย่างเช่น บิทคอยน์ (Bitcoin) อีเธอเรียม (Ethereum) และสกุลเงินต่าง ๆ ที่สามารถแลกเปลี่ยนด้วยจำนวนเดียวกัน เนื่องจากมีมูลค่าทดแทนกันได้ ด้วยคุณสมบัติดังกล่าวจึงทำให้สินทรัพย์ดิจิทัลไม่มีมูลค่าและไม่มีความสามารถในการซื้อขาย [1]

ในการซื้อขายและแลกเปลี่ยนสินทรัพย์บนแพลตฟอร์มโดยใช้โทเคนดิจิทัลลักษณะเฉพาะ (Non-Fungible Token : NFT) เพื่อแสดงกรรมสิทธิ์ความเป็นเจ้าของสินทรัพย์ซึ่งผู้ใช้งานมีสิทธิ์ถือครอง และจัดการเคลื่อนย้ายสินทรัพย์ดิจิทัลได้ตามต้องการ รวมถึงการเพิ่มคุณสมบัติให้สินทรัพย์ดิจิทัลมีความเป็นเอกลักษณ์และแตกต่าง

ทำให้ต้องซื้อขายแบบเต็มจำนวนของโทเคนดิจิทัลเท่านั้น โดยพัฒนาแพลตฟอร์มในรูปแบบแอปพลิเคชันแบบกระจาย (Decentralized Applications : DApps) เพื่อจัดเก็บและตรวจสอบข้อมูลการทำธุรกรรมบนบล็อกเชนแบบสาธารณะ (Public Blockchain) ที่ใช้หลักการของสัญญาอัจฉริยะ (Smart Contract) บนเครือข่ายอีเธอร์เรียม (Ethereum Blockchain) [2] ทำให้การซื้อขายดังกล่าวมีความโปร่งใส ปลอดภัย และไร้ตัวกลาง ตัวอย่างผลงานซึ่งเป็นที่นิยมเป็นของสะสม และมีคุณค่าทางจิตใจ เช่น รูปภาพ คลิปวิดีโอ และไอเทมเกม เป็นต้น [3] โดยแต่ละแพลตฟอร์มจะมีค่าธรรมเนียมในการซื้อขาย (Gas Fee) แตกต่างกันไป ซึ่งใช้สกุลเงินอีเทอร์ (Ether : ETH) เป็นสกุลเงินดิจิทัลในการซื้อขายและการชำระค่าธรรมเนียม

จากปัญหาที่กล่าวมาข้างต้น งานวิจัยนี้จึงได้ทำการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัล ลักษณะเฉพาะบนเครือข่ายอีเธอร์เรียมในรูปแบบแอปพลิเคชันแบบกระจายที่ประยุกต์ใช้กระเป๋าเงินดิจิทัลของ Metamask และสกุลเงินอีเทอร์ในกระเป๋าเงินดิจิทัลสำหรับซื้อขายและชำระค่าธรรมเนียม โดยในบทความวิจัยแบ่งเนื้อหาออกได้เป็น 5 ส่วน ได้แก่ ส่วนที่ 1 บทนำ ส่วนที่ 2 ทฤษฎี และงานวิจัยที่เกี่ยวข้อง ส่วนที่ 3 วิธีการดำเนินการวิจัย ส่วนที่ 4 ผลการดำเนินงานวิจัย และส่วนที่ 5 สรุปผลการวิจัย

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 โทเคนดิจิทัลลักษณะเฉพาะ (Non-Fungible Token)

โทเคนดิจิทัลลักษณะเฉพาะแสดงถึงการระบุกรรมสิทธิ์ของสินทรัพย์ ซึ่งมีคุณสมบัติความเป็นเอกลักษณ์ มีคุณค่าทางจิตใจ และเหมาะแก่การเก็บสะสม ตัวอย่างเช่น คลิปวิดีโอ นักกีฬาชนะการแข่งขัน มีม (Meme) รูปภาพ งานศิลปะ และไอเทมในเกม เป็นต้น ซึ่งอยู่ในรูปแบบของสินทรัพย์ดิจิทัล โดยโทเคนดิจิทัลเปรียบเสมือนใบรับรองดิจิทัลที่มีความคล้ายคลึงกับ โฉนดที่ดินที่ได้รับบูรณายถะเอียดความเป็นเจ้าของ และถูกบันทึกบนบล็อกเชน จึงทำให้ธุรกรรมที่เกิดขึ้นมีความโปร่งใส ปลอดภัย และไร้ตัวกลาง

การซื้อขายแลกเปลี่ยนกรรมสิทธิ์ของสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ ผู้ซื้อได้รับกรรมสิทธิ์ความเป็น

เจ้าของสินทรัพย์ จึงเป็นการเปิดโอกาสสร้างรายได้ให้กับผู้สร้างและผู้ขาย ตัวอย่างเช่น การประมูลขายทวิตแรกของ Jack Dorsey โดยผู้ซื้อโทเคนดิจิทัลลักษณะเฉพาะของทวิตดังกล่าวจะได้รับกรรมสิทธิ์ความเป็นเจ้าของเท่านั้น แต่ผู้สร้างทวิตแรกของโลกยังคงเป็น Jack Dorsey รวมทั้งผู้อื่นยังสามารถเข้ารับชมทวิตดังกล่าวได้เช่นเดิม [1]

### 2.2 อีเธอร์เรียมบล็อกเชน (Ethereum Blockchain)

อีเธอร์เรียมบล็อกเชนเป็นแพลตฟอร์มโอเพนซอร์สที่สามารถปรับแก้ไขได้บนบล็อกเชน มีการพัฒนาขึ้นเป็นแอปพลิเคชันแบบกระจาย และดำเนินการธุรกรรมด้วยสกุลเงินอีเทอร์ในการทำธุรกรรม ซึ่งต้องมีการชำระค่าธรรมเนียมให้กับนักพัฒนา และนักขุดที่แก่สมการจนได้รับรางวัล [4]

#### 2.2.1 รูปแบบการใช้งานเครือข่ายอีเธอร์เรียม

รูปแบบการใช้งานเครือข่ายอีเธอร์เรียมสามารถแบ่งออกเป็น 3 รูปแบบ ได้แก่ รูปแบบที่ 1 ระบบการเงินแบบกระจาย (Decentralized Finance : Defi) เป็นรูปแบบการเงินไม่มีตัวกลาง ซึ่งใช้สัญญาอัจฉริยะ และทำงานบนบล็อกเชน ตัวอย่างแพลตฟอร์มของระบบการเงินแบบกระจาย ได้แก่ Uniswap Compound และ MakerDao เป็นต้น รูปแบบที่ 2 โทเคนดิจิทัลลักษณะเฉพาะ ซึ่งได้รับความนิยมและเป็นเรื่องใหม่ในช่วงของการพัฒนา ส่วนใหญ่ประยุกต์ใช้กับงานศิลปะหรือเกี่ยวข้องกับความคิดสร้างสรรค์ มีความเป็นเอกลักษณ์และแตกต่าง ตัวอย่างแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ ตัวอย่างเช่น Opensea Rarible และ SuperRare เป็นต้น [2] และรูปแบบที่ 3 การบริหารองค์กรแบบกระจาย เป็นแนวทางการบริหารองค์กร โดยสมาชิกในเครือข่ายสามารถตรวจสอบการทำงานของการบริหารองค์กร และมีส่วนร่วมในการบริหารด้วยการถือครองโทเคนดิจิทัล

### 2.3 งานวิจัยที่เกี่ยวข้อง

จากการศึกษาเกี่ยวกับโทเคนดิจิทัลลักษณะเฉพาะ รวมทั้งการหาความสัมพันธ์ระหว่างตลาดสกุลเงินดิจิทัลและตลาดโทเคนดิจิทัลลักษณะเฉพาะ [1] โดยในการสร้างโทเคนดิจิทัลลักษณะเฉพาะใช้มาตรฐาน ERC721

ซึ่งมีข้อจำกัดของมาตรฐานในการรองรับหลายโทเคนดิจิทัลบนสัญญาอัจฉริยะและค่าธรรมเนียมในการทำธุรกรรม [5] และมีการประยุกต์ใช้กับการจัดเก็บข้อมูลด้วย Interplanetary File System (IPFS) เพื่อซื้อขายสินทรัพย์ดิจิทัลบนแพลตฟอร์มตัวอย่างเช่น เกมสะสมแมว CryptoKitties เป็นต้น [3]

จากการศึกษาการประยุกต์ใช้งานบนเครือข่ายอีเธอเรียมในด้านต่าง ๆ ได้แก่ ด้านซื้อขายสินค้า ด้านสุขภาพ ด้านการจัดการลิขสิทธิ์ ด้านประกันภัย และด้านการเงิน [4] ตัวอย่างงานวิจัยได้มีการพัฒนาระบบจัดเก็บข้อมูลของกรรมที่ดิน ซึ่งใช้หลักการควบคุมการเข้าถึงข้อมูลที่มีความเชื่อถือ เพื่อความโปร่งใส และความน่าเชื่อถือ [6] ซึ่งจากงานวิจัยพบว่าการพัฒนาแอปพลิเคชันแบบกระจายมีความปลอดภัยมากกว่าแอปพลิเคชันแบบรวมศูนย์ เนื่องจากประยุกต์ใช้สัญญาอัจฉริยะในการประมวลผลและบันทึกข้อมูลบนบล็อกเชน [7]

### 3. วิธีการดำเนินการวิจัย

วิธีการดำเนินการวิจัยในการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ แบ่งออกเป็น 4 ขั้นตอนหลักดังนี้

#### 3.1 การศึกษาปัญหาและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาปัญหาที่เกี่ยวข้องกับแพลตฟอร์มซื้อขายสินทรัพย์ดิจิทัลในปัจจุบัน พบว่า สินทรัพย์ดิจิทัลทั่วไปไม่สามารถตรวจสอบความเป็นของแท้ได้ จึงทำให้สินทรัพย์ไม่มีมูลค่า และไม่มีการใช้ประโยชน์จากสิทธิการเป็นเจ้าของสินทรัพย์ดิจิทัล ดังนั้นวัตถุประสงค์ของงานวิจัยนี้จึงได้พัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

#### 3.2 การออกแบบกรอบแนวคิดการวิจัย

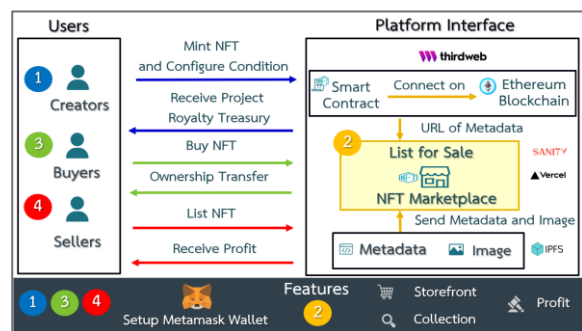
หลังจากการศึกษาปัญหาและงานวิจัยที่เกี่ยวข้อง คณะผู้วิจัยจึงได้ทำการออกแบบกรอบแนวคิดการวิจัยของการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะสามารถแสดงได้ดังภาพที่ 1 แบ่งออกเป็น 2 ส่วนหลัก ได้แก่ ส่วนของผู้ใช้งาน (Users) และส่วนของแพลตฟอร์ม (Platform Interface) มีรายละเอียดดังนี้

#### 1) ส่วนของผู้ใช้งาน

ในส่วนผู้ใช้งานสามารถแบ่งออกได้เป็น 3 ส่วน ได้แก่ หมายเลข 1 ผู้สร้าง (Creators) หมายเลข 3 ผู้ซื้อ (Buyers) และหมายเลข 4 ผู้ขาย (Sellers) โดยผู้ใช้งานจะต้องเชื่อมต่อกระเป๋าเงินดิจิทัลก่อนการใช้งานแพลตฟอร์ม ซึ่งในงานวิจัยนี้ใช้กระเป๋าเงินดิจิทัลของ Metamask และใช้สกุลเงินอีเทอร์ในกระเป๋าเงินดิจิทัลสำหรับซื้อขาย และการชำระค่าธรรมเนียม

#### 2) ส่วนของแพลตฟอร์ม

ในส่วนของแพลตฟอร์มแบ่งออกเป็น 2 ส่วนหลัก ได้แก่ ส่วนติดต่อผู้ใช้งาน (Frontend) และส่วนการทำงานเบื้องหลัง (Backend) โดยมีฟังก์ชันการทำงานของแพลตฟอร์ม ได้แก่ การสร้าง การแสดงรายการขาย การซื้อ และการโอนกรรมสิทธิ์ความเป็นเจ้าของระหว่างผู้ใช้งาน รวมทั้งการได้รับค่าลิขสิทธิ์ความเป็นเจ้าของ และผลตอบแทนจากการซื้อขายสินทรัพย์ดิจิทัล



ภาพที่ 1: กรอบแนวคิดการวิจัยการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

#### 3.3 การออกแบบและพัฒนาแพลตฟอร์ม

ในการออกแบบและพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์ดิจิทัลพัฒนาขึ้นในรูปแบบแอปพลิเคชันแบบกระจาย (Decentralized Applications : DApps) ซึ่งในงานวิจัยประยุกต์ใช้เครื่องมือ Thirdweb ในส่วนของการทำงาน และตรวจสอบการทำธุรกรรมกับสินทรัพย์ดิจิทัลบนแพลตฟอร์มด้วยการสร้างสัญญาอัจฉริยะ (Smart Contract) ที่พัฒนาด้วยภาษาซอลิดิตี (Solidity) บนเครือข่ายอีเธอเรียม อีกทั้งทำการจัดเก็บรายละเอียดของสินทรัพย์และแพลตฟอร์มโดยใช้เครื่องมือ Sanity

ในส่วนการจัดเก็บและรักษาข้อมูลการทำธุรกรรมของผู้ใช้งานมีการเชื่อมต่อกับบล็อกเชนสาธารณะ (Public



Blockchain) ซึ่งสามารถตรวจสอบ ยืนยันความถูกต้อง เพื่อมีความโปร่งใสของข้อมูล โดยทำการจัดเก็บคำอธิบาย ข้อมูล และสินทรัพย์ดิจิทัลด้วย Interplanetary File System (IPFS) ในส่วนการพัฒนาเพื่อใช้งานจริงประยุกต์ใช้ เครื่องมือ Vercel โดยแบ่งการทำงานของแพลตฟอร์ม ออกเป็น 2 ส่วน ได้แก่ ส่วนของผู้ใช้งาน ได้แก่ ผู้สร้าง ผู้ซื้อ และผู้ขาย ในส่วนแพลตฟอร์มที่เชื่อมต่อกับเครือข่าย อีเธอเรียม ประกอบด้วยส่วนติดต่อกับผู้ใช้งาน และส่วน การทำงานเบื้องหลัง

จากการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้ โทเคนดิจิทัลลักษณะเฉพาะ ประกอบด้วย 6 ฟังก์ชัน การทำงานหลัก ได้แก่ การสร้าง (Mint Non-Fungible Token) การแสดงรายการขาย (List Non-Fungible Token) การซื้อ (Buy Non-Fungible Token) การโอนกรรมสิทธิ์ความเป็น เจ้าของ (Transfer Owner) การได้รับค่าสิทธิความเป็นเจ้าของ (Receive Project Royalty Treasury) และการได้รับผลตอบแทน (Receive Profit) ซึ่งตัวอย่างฟังก์ชันการตรวจสอบ และแสดง รายการขายสินทรัพย์ดิจิทัลสามารถแสดงได้ดังภาพที่ 2

```

List Non-Fungible Token Function
const NFTCard = ({nftItem, title, listings}) => {
  const [isListed, setIsListed] = useState(false)
  const [price, setPrice] = useState(0)
  const [listingId, setListingId] = useState(0)
  useEffect(() => {
    const listing = listings.find((listing) => listing.asset.name === nftItem.name)
    if (Boolean(listing)) {
      setIsListed(true)
      setPrice(listing.buyoutCurrencyValuePerToken.displayValue)
      setListingId(listing.id)
    }, [listings, nftItem])
  }
}
    
```

ภาพที่ 2: ฟังก์ชันการตรวจสอบและแสดงรายการขายสินทรัพย์ดิจิทัล

ผู้ซื้อสามารถดูรายละเอียดข้อมูลรายการขาย และดำเนินการซื้อสินทรัพย์ดิจิทัล (Buy Non-Fungible Token Function) ซึ่งสามารถแสดงฟังก์ชันการทำงานได้ดังภาพที่ 3

```

Buy Non-Fungible Token Function
const confirmPurchase = (toastHandler = toast) => {
  toastHandler.success('Purchase successful!', {
    style: {background: '#04111d',color: '#fff', }, })
const buyItem = async (
  listingId = selectedMarketNft.id,
  quantityDesired = 1,
  module = marketPlaceModule) => {
  console.log(listingId)
  await module
    .buyoutDirectListing({
      listingId: listingId,
      quantityDesired: quantityDesired,})
  confirmPurchase()}
    
```

ภาพที่ 3: ฟังก์ชันการซื้อสินทรัพย์ดิจิทัล

### 3.4 การทดสอบการใช้งานของแพลตฟอร์ม

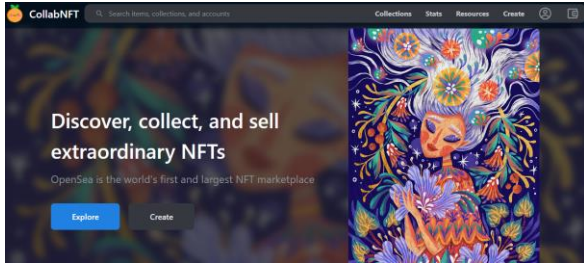
ในการวิจัยจะทำการทดสอบการใช้งานแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ บนเครือข่ายอีเธอเรียมที่พัฒนาขึ้นจากผู้ใช้งานที่เกี่ยวข้อง 3 กลุ่ม ซึ่งมีรายละเอียดดังนี้

- 1) ผู้สร้าง (Creators) สามารถสร้างและกำหนดเงื่อนไข การซื้อขายสินทรัพย์ดิจิทัล รวมถึงแสดงรายการขาย และได้รับค่าสิทธิความเป็นเจ้าของ
- 2) ผู้ขาย (Sellers) สามารถกำหนดสินทรัพย์ดิจิทัล เพื่อแสดงรายการขาย และได้รับผลตอบแทนจากการซื้อขายสินทรัพย์
- 3) ผู้ซื้อ (Buyers) สามารถรับชมรายการขายและซื้อ รวมทั้งได้รับการโอนกรรมสิทธิ์ความเป็นเจ้าของสินทรัพย์ดิจิทัลระหว่างผู้ใช้งานผ่านแพลตฟอร์มได้

## 4. ผลการดำเนินงานวิจัย

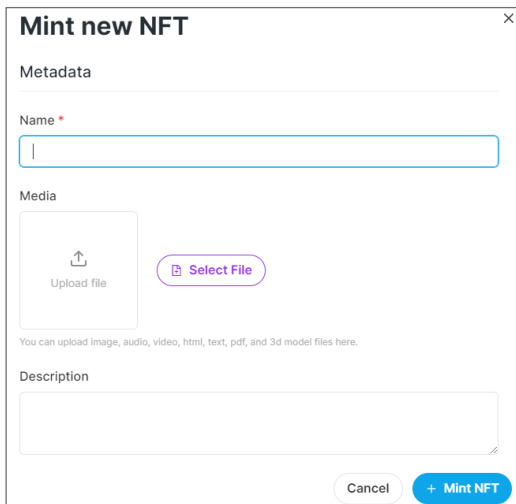
### 4.1 ผลการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

ผลการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ สามารถแสดงตัวอย่างหน้าแรกของแพลตฟอร์มได้ดังภาพที่ 4



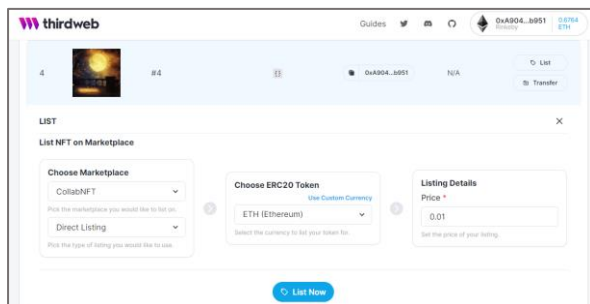
ภาพที่ 4: หน้าแรกของแพลตฟอร์มซื้อขายสินทรัพย์  
โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

ตัวอย่างหน้าจอในส่วนของการสร้างโดยใช้โทเคนดิจิทัลลักษณะเฉพาะ ซึ่งผู้สร้างสามารถกำหนดรายละเอียดได้แก่ ชื่อ ไฟล์สินทรัพย์ดิจิทัล และคำอธิบายรายละเอียด ซึ่งสามารถแสดงได้ดังภาพที่ 5



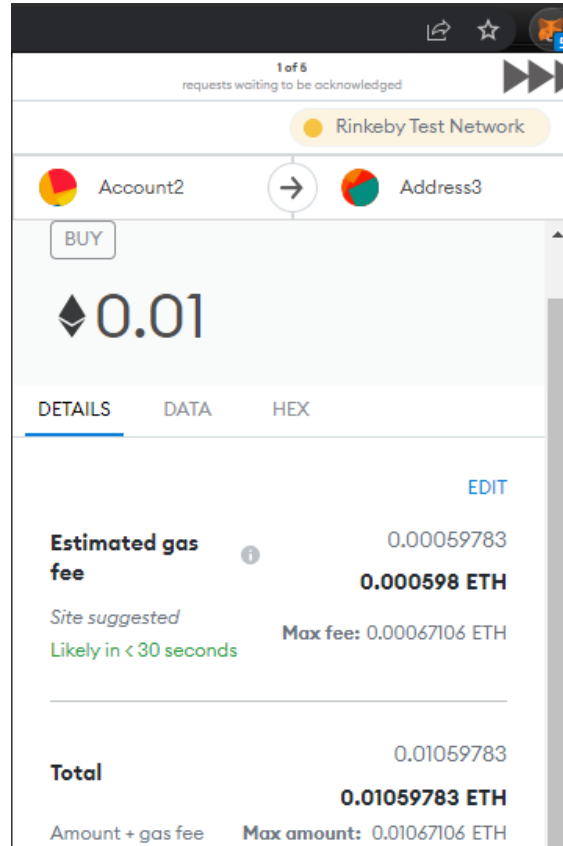
ภาพที่ 5: หน้าจอการสร้างโดยใช้โทเคนดิจิทัล  
ลักษณะเฉพาะ

ตัวอย่างหน้าจอในส่วนของผู้สร้างหรือผู้ขายที่ทำการกำหนดเงื่อนไขซื้อขายสินทรัพย์เพื่อแสดงรายการขาย ซึ่งสามารถแสดงได้ดังภาพที่ 6



ภาพที่ 6: หน้าจอการกำหนดเงื่อนไขของการซื้อขาย  
สินทรัพย์ดิจิทัล

ตัวอย่างหน้าจอในส่วนของการชำระค่าธรรมเนียมด้วยกระเป๋าเงินดิจิทัล Metamask ซึ่งสามารถแสดงได้ดังภาพที่ 7



ภาพที่ 7: หน้าจอสำหรับการชำระค่าธรรมเนียมด้วย Metamask

#### 4.2 ผลการทดสอบการใช้งานของแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะ

การทดสอบการใช้งานของแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะที่พัฒนาขึ้น สามารถแสดงได้ผลการทดสอบการใช้งานได้ดังตารางที่ 1 ซึ่งทำการทดสอบจากทุกบทบาท (Roles) ที่เกี่ยวข้องกับแพลตฟอร์ม ได้แก่ ผู้สร้าง ผู้ขาย และผู้ซื้อ โดยทำการทดสอบฟังก์ชันการทำงานของแพลตฟอร์ม 6 ฟังก์ชันหลัก ได้แก่ การสร้าง (Mint NFT) การแสดงรายการขาย (List NFT) การซื้อ (Buy NFT) การโอนกรรมสิทธิ์ความเป็นเจ้าของ (Transfer Owner) การได้รับค่าสิทธิความเป็นเจ้าของ (Receive Project Royalty Treasury) และการได้รับผลตอบแทน (Receive Profit) จากผลการทดสอบพบว่าทุกฟังก์ชันการทำงานสามารถซื้อขายแลกเปลี่ยนสินทรัพย์ แสดงผลรายการขายและการได้รับผลประโยชน์ที่ควรจะได้รับได้อย่างถูกต้อง

**ตารางที่ 1: ผลการทดสอบการใช้งานของแพลตฟอร์ม  
ซื้อขายสินทรัพย์ดิจิทัล**

ลำดับ	ฟังก์ชันการทำงาน (Features)	บทบาท (Role)	ผลการทดสอบ			
			ผลลัพธ์ที่คาดหวัง (Expected Result)	ผลลัพธ์จากการทดสอบ (Actual Result)	ผ่าน (Pass)	ไม่ผ่าน (Fail)
1	การสร้าง	ผู้สร้าง	สร้างสินทรัพย์ด้วย NFT	สร้างสินทรัพย์ด้วย NFT	✓	
2	การแสดงผล	ผู้สร้างและผู้ขาย	กำหนดสินทรัพย์เพื่อแสดงรายการขาย	กำหนดสินทรัพย์เพื่อแสดงรายการขาย	✓	
3	การซื้อ	ผู้ซื้อ	รับชมรายการขายและซื้อสินทรัพย์ดิจิทัล	รับชมรายการขายและซื้อสินทรัพย์ดิจิทัล	✓	
4	การโอนกรรมสิทธิ์ความเป็นเจ้าของ	ผู้ซื้อ	ได้รับการโอนกรรมสิทธิ์ความเป็นเจ้าของสินทรัพย์ดิจิทัล	ได้รับการโอนกรรมสิทธิ์ความเป็นเจ้าของสินทรัพย์ดิจิทัล	✓	
5	การได้รับค่าลิขสิทธิ์ความเป็นเจ้าของ	ผู้สร้าง	ได้รับค่าลิขสิทธิ์ความเป็นเจ้าของจากการสร้างสรรค์สินทรัพย์ดิจิทัล	ได้รับค่าลิขสิทธิ์ความเป็นเจ้าของจากการสร้างสรรค์สินทรัพย์ดิจิทัล	✓	
6	การได้รับผลตอบแทน	ผู้ขาย	ได้รับผลตอบแทนจากการขายสินทรัพย์ดิจิทัล	ได้รับผลตอบแทนจากการขายสินทรัพย์ดิจิทัล	✓	

## 5. สรุปผลการวิจัย

จากการพัฒนาแพลตฟอร์มซื้อขายสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะบนเครือข่ายอีเธอเรียมในรูปแบบแอปพลิเคชันแบบกระจาย ซึ่งในงานวิจัยใช้กระเป๋าเงินดิจิทัลของ Metamask และใช้สกุลเงินอีเทอร์ในกระเป๋าเงินดิจิทัลสำหรับซื้อขาย และการชำระค่าธรรมเนียม โดยแพลตฟอร์มที่พัฒนาขึ้นสามารถรองรับฟังก์ชันการทำงานต่าง ๆ ได้แก่ การสร้าง การแสดงผล รายการขาย การซื้อ การโอนกรรมสิทธิ์ความเป็นเจ้าของ การได้รับค่าลิขสิทธิ์ความเป็นเจ้าของและการได้รับผลตอบแทน จากผลการทดสอบการใช้งานแพลตฟอร์มพบว่าทุกฟังก์ชันการทำงานสามารถนำไปใช้งานได้จริง ช่วยเพิ่มช่องทางให้สินทรัพย์ดิจิทัลที่มีมูลค่าสามารถซื้อขาย

กันได้ง่ายมากขึ้น อีกทั้งเป็นแนวทางในการประยุกต์ใช้กับการลงทุนของสินทรัพย์โดยใช้โทเคนดิจิทัลลักษณะเฉพาะหรืองานวิจัยอื่น ๆ ที่เกี่ยวข้อง และสามารถนำไปต่อยอดพัฒนาแอปพลิเคชันแบบกระจายบนเครือข่ายอีเธอเรียมที่เน้นความมั่นคงปลอดภัยมากขึ้นในการใช้งาน

## เอกสารอ้างอิง

- [1] Ante, L. (2021). "The Non-Fungible Token (NFT) Market and Its relationship with Bitcoin and Ethereum," *SSRN Electronic Journal*, Vol. 20.
- [2] Thilagaraj, A., and Davis, J. (2021). "Non-Fungible Token (NFT) – The Game Changer in the Digital Art World," *In Proceeding of 2021 Ciencia Y Sociedad*, Vol. 51, pp. 190-194.
- [3] Karapapas, C., Pittaras, I. and Polyzos, G.C. (2021). "Fully Decentralized Trading Games with Evolvable Characters using NFTs and IPFS," *In Proceeding of 2021 IFIP Networking Conference (IFIP Networking)*, pp. 1-2.
- [4] Mohanta, B.K., Panda, S.S. and Jena, D. (2018). "An Overview of Smart Contract and Use Cases in Blockchain Technology," *In Proceeding of 2018 9th international conference on computing, communication and networking technologies (ICCCNT)*, pp. 1-4.
- [5] Chirtoaca, D., Ellul, J. and Azzopardi, G. (2020). "A Framework for Creating Deployable Smart Contracts for Non-Fungible Tokens on the Ethereum Blockchain," *In Proceeding of 2020 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS)*, pp. 100-105.
- [6] Nandi, M., Bhattacharjee, R.K., Jha, A. and Barbhuiya, F.A. (2020). "A Secured Land Registration Framework on Blockchain," *In Proceeding of 2020 Third ISEA Conference on Security and Privacy (ISEA-ISAP)*, pp. 130-138.
- [7] Kina-Kina, K.M., Cutipa-Arias, H.E. and Shiguihara-Juárez, P. (2019). "A Comparison of Performance between Fully and Partially Decentralized Applications," *In Proceeding of 2019 IEEE XXVI International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pp. 1-4.

# โมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์ด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ

## Predictive Models for Auto Loan Approval Using the Decision Tree Algorithm

ศุภชานันท์ วนภู (Supachanun Wanapu)<sup>1</sup> นริศ มิ่งโมรา (Naris Mingmora)<sup>2</sup> เจนจิรา อักษรพิมพ์ (Janejeera Aksornpim)<sup>3</sup>  
 เกศสุดา ชัยนอก (Ketsuda Chainok)<sup>4</sup> อวิษรา คอกอบล (Awissara Dokauboi)<sup>5</sup> และธิตี พิพัฒน์พฤกษา (Thati Phiphatprueksa)<sup>6</sup>

<sup>1,2,3,4,5,6</sup> หลักสูตรบริหารธุรกิจบัณฑิต สาขาวิชาคอมพิวเตอร์ธุรกิจ

คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏนครราชสีมา

<sup>1</sup>supachanun.w@nrru.ac.th, <sup>2</sup>naris.m@nrru.ac.th, <sup>3</sup>janejeera.a@nrru.ac.th,

<sup>4</sup>6040505302@nrru.ac.th, <sup>5</sup>6040505312@nrru.ac.th, <sup>6</sup>6040505319@nrru.ac.th

### บทคัดย่อ

ผลกระทบของหนี้ที่ไม่ก่อให้เกิดรายได้ (NPLs) อันเนื่องมาจากการอนุมัติสินเชื่อให้กับลูกค้าโดยขาดระบบสนับสนุนการตัดสินใจที่ดีพอ ก่อให้เกิดปัญหาหลายประการในการดำเนินธุรกิจซื้อขายรถยนต์มือสอง ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาและวิเคราะห์ข้อมูลเชิงลึกของลูกค้าที่ขอสินเชื่อรถยนต์ และออกแบบ พัฒนา โมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์ด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ โดยมี 4 ขั้นตอนคือ 1) การทำความเข้าใจกับปัญหาและข้อมูล 2) การจัดเตรียมข้อมูล 3) การสร้างและประเมินโมเดล และ 4) การนำโมเดลไปใช้งาน โดยใช้ข้อมูลจากผู้ประกอบการธุรกิจซื้อขายรถยนต์มือสอง บริษัท สรรธร 99 จำกัด ระหว่างปี พ.ศ. 2560–2562 จำนวน 11 แอททริบิวต์ มีข้อมูล 400 รายการ จากการศึกษาและวิเคราะห์ข้อมูลเชิงลึกของลูกค้าที่ขอสินเชื่อรถยนต์ พบว่า มีคุณลักษณะสำคัญ 6 แอททริบิวต์ที่ส่งผลต่อการอนุมัติให้สินเชื่อรถยนต์ หลังจากได้ข้อมูลที่ผ่านกระบวนการคัดเลือกข้อมูล ไปออกแบบและพัฒนาโมเดลเพื่อทำนายพบว่า โมเดลที่ได้มีความถูกต้อง ร้อยละ 79.54 ค่าความแม่นยำ ร้อยละ 92.15 ค่าความระลึก ร้อยละ 76.12 และค่าเอฟเมซเซอร์ ร้อยละ 83.11 จากนั้นนำโมเดลนี้ไปประยุกต์ใช้งานจริงบนโปรแกรม Microsoft Excel พบว่าผู้ประกอบการมีความพึงพอใจในระดับมากในด้านความง่ายต่อการใช้งาน

คำสำคัญ: เหมือนข้อมูล ต้นไม้ตัดสินใจ สินเชื่อรถยนต์

### Abstract

The impact of non-performing debts (NPLs) is due to credit approvals for customers without a good decision support system. It has caused many problems in the used car trading business. Therefore, the objectives of this research were to study and conduct business insight analysis about customers applying for car loans and to design and develop a model to predict auto loan approval patterns using a decision tree algorithm. The steps are 1) Business and Data Understanding, 2) Data Preparation, 3) Modeling and Evaluation, and 4) Model Implementation. Using data from used car dealers, Santhorn 99 Co., Ltd., during the years 2017–2019, amounting to 11 attributes and 400 records of customers who applied for the car loan, it was found that 6 key attributes affected the approval of the car loan. After obtaining the data that was selected to design and develop the predictive model, it was found that the efficient predictive model's accuracy was 79.54% and precision was 92.15%. The recall was 76.12% and the F-measure was 83.11%. Finally, the prediction model was applied to real applications using the Microsoft Excel program. The results showed that owner was satisfied with a high level of usability.

**Keywords:** Data Mining, Decision Tree, Car Loan.

## 1. บทนำ

จากสถานการณ์การแพร่ระบาดของโควิดทั่วโลก ส่งผลกระทบต่อยอดการผลิตและยอดขายรถยนต์ใหม่ แต่ในอีกมุมหนึ่งกลับพบว่า ความต้องการใช้รถยนต์ส่วนตัวเพิ่มขึ้น ในขณะที่ปริมาณการผลิตรถยนต์มีกำลังการผลิตที่ลดลง เนื่องจากปัญหาการขาดแคลนชิ้นส่วนสำคัญ ส่งผลให้ตลาดรถมือสองเติบโตขึ้น มีการคาดว่าราคารถยนต์นั่งส่วนบุคคลมือสองในประเทศไทยจะมีโอกาสขยายตัวร้อยละ 5.5- 7.5 [1] ธุรกิจซื้อขายรถยนต์มือสองจึงต้องเตรียมความพร้อมในการพิจารณาการให้สินเชื่อแก่ลูกค้า โดยคำนึงถึงทั้งความพึงพอใจในการให้บริการและคำนึงถึงความเสี่ยงที่จะเกิดขึ้นจากการให้สินเชื่อซึ่งเป็นสิ่งสำคัญที่จะกระทบต่อการดำเนินธุรกิจ

ปัจจัยความเสี่ยงที่อาจจะเกิดขึ้นและมีผลกระทบต่อ การดำเนินงานของธุรกิจซื้อขายรถยนต์มือสอง คือ ความเสี่ยงจากการปล่อยสินเชื่อรถยนต์มือสอง ความเสี่ยงจากการดำเนินงานอันเกิดจากสินเชื่อที่ไม่ก่อให้เกิดรายได้ และความเสี่ยงจากการขายรถที่ซื้อมา [2] ดังนั้น การวิเคราะห์ข้อมูลการให้สินเชื่อ จึงเป็นสิ่งสำคัญมากเพราะจะส่งผลต่อรายได้และสภาพคล่องของธุรกิจ พบว่าผู้ประกอบการธุรกิจหลายรายประสบปัญหาสูญเสียรายได้จากลูกค้าที่ไม่ดี ทำให้เกิดหนี้สูญ โดยวิธีการพิจารณาการให้สินเชื่อแก่ลูกค้าจะพิจารณาจากคุณสมบัติของลูกค้า เช่น เพศ อายุ อาชีพ อายุงาน รายได้ ค่าใช้จ่าย และรายได้สุทธิ เป็นต้น ซึ่งเป็นข้อมูลที่สามารถบ่งบอกความสามารถในการชำระหนี้ การวิเคราะห์ข้อมูลส่วนใหญ่จะใช้สูตรการคำนวณผ่านโปรแกรมการคำนวณอย่างง่ายด้วยโปรแกรมไมโครซอฟท์เอ็กเซล ซึ่งทำให้ไม่สามารถวิเคราะห์เชิงลึกในการพิจารณาให้สินเชื่อได้อย่างถูกต้องแม่นยำ ทำให้เกิดความคลาดเคลื่อนในการพิจารณาการให้สินเชื่อ และส่งผลต่อการเกิดหนี้สูญ รายได้และสภาพคล่องของธุรกิจ

การทำเหมืองข้อมูล (Data Mining) เป็นเทคนิคการวิเคราะห์ข้อมูลโดยอาศัยหลักการทางสถิติ และหลักการทางคณิตศาสตร์ เพื่อค้นหารูปแบบและความสัมพันธ์ของข้อมูลจากข้อมูลที่มีจำนวนมาก สามารถนำไปประยุกต์ใช้งานด้านต่างๆ ได้ โดยอาศัยการนำข้อมูลจากอดีตหรือข้อมูลที่มีอยู่แล้วนำมาสร้างตัวแบบพยากรณ์การทำนาย

หรือคาดการณ์สิ่งที่จะเกิดขึ้นในอนาคต อีกทั้งสามารถสรุปผลออกมาเป็นข้อมูลที่มีประโยชน์ [3] ทั้งนี้การใช้เหมืองข้อมูลในการวิเคราะห์ข้อมูลเพื่อการทำนายการอนุมัติสินเชื่อรถยนต์มือสองนี้ได้มีการนำวิธีการจำแนกประเภทข้อมูล (Data Classification) เป็นการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ ทำให้ได้กฎเพื่อช่วยในการตัดสินใจจากข้อมูลที่มีอยู่ และสามารถทำนายแนวโน้มการเกิดขึ้นของข้อมูลที่ยังไม่เกิดขึ้น พบว่าเทคนิคการจำแนกประเภทข้อมูล นิยมนำเสนอในรูปแบบของขั้นตอนวิธีต้นไม้ช่วยในการตัดสินใจ [4]

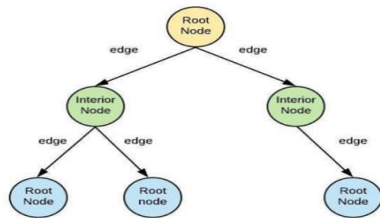
จากปัญหาผลกระทบของหนี้ที่ไม่ก่อให้เกิดรายได้ (NPLs) อันเนื่องมาจากการอนุมัติสินเชื่อให้กับลูกค้าโดยขาดระบบสนับสนุนการตัดสินใจที่ดี ผู้วิจัยจึงนำเทคนิคการทำเหมืองข้อมูล ด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ โดยมีวัตถุประสงค์ของการวิจัยเพื่อ 1) ศึกษาและวิเคราะห์ข้อมูลเชิงลึกของลูกค้าที่ขอสินเชื่อรถยนต์ 2) ออกแบบและพัฒนาโมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์ ด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ รวมทั้งนำโมเดลที่ได้ไปประยุกต์ใช้งานเพื่อสนับสนุนการตัดสินใจอนุมัติสินเชื่อซื้อขายรถยนต์มือสองในรูปแบบที่ง่ายและสะดวกต่อการใช้งาน ช่วยให้สามารถอนุมัติสินเชื่อได้อย่างรวดเร็วทันสมัย ลดความเสี่ยงด้านหนี้สูญ ทำให้เกณฑ์ในการคัดเลือกลูกค้าชั้นดีที่มีประสิทธิภาพให้กับธุรกิจมากยิ่งขึ้น

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคหนึ่งของการทำเหมืองข้อมูลที่ได้รับนิยมนำมาสร้างโมเดลเพื่อการทำนาย (Predictive Modeling) ซึ่งใช้หลักการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยเทคนิคต้นไม้ตัดสินใจใช้หลักการจำแนกประเภทข้อมูลออกเป็นกลุ่มของข้อมูล (Class) เพื่อทำนายผลลัพธ์ที่ได้ ที่ขึ้นอยู่กับคุณลักษณะ (Feature) หรือปัจจัยนำเข้าว่าควรใช้คุณลักษณะใดเป็นหลักในการแบ่งกลุ่มข้อมูล ลักษณะของโครงสร้างต้นไม้ที่ภายในต้นไม้จะประกอบไปด้วยโหนด (Node) โดยแต่ละโหนดจะมีคุณลักษณะเป็นตัวทดสอบ

กิ่งของต้นไม้ (Branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (Leaf) ซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจแสดงถึงกลุ่มของข้อมูล ก็คือผลลัพธ์ที่ได้จากการทำนาย โหนดที่อยู่บนสุดของต้นไม้เรียกว่าโหนดราก (Root Node) [5] ลักษณะ โครงสร้างของต้นไม้ตัดสินใจแสดง ดังภาพที่ 1



ภาพที่ 1: โครงสร้างต้นไม้ตัดสินใจ [6]

### 2.2 ขั้นตอนวิธีของต้นไม้ตัดสินใจ

ขั้นตอนวิธีหรืออัลกอริทึมที่นิยมนำมาใช้ในการสร้างต้นไม้ตัดสินใจ มีดังต่อไปนี้

#### 2.2.1 อัลกอริทึม ID3 เป็นการสร้างต้นไม้ตัดสินใจที่ใช้

หลักการทฤษฎีสารสนเทศ ค่าที่วัดได้จะนำมาใช้ตัดสินใจว่าจะใช้ตัวแปรใดในการแบ่งข้อมูลโดยวิธีกำหนด โครงสร้างต้นไม้ตัดสินใจจะเป็นการเลือกข้อมูลตามลำดับของตัววัดหรือค่าเกน (Gain) สูงที่สุดเป็นข้อมูลเริ่มต้นและข้อมูลถัดไปมีค่าลดหลั่นกันตามลำดับ [5]

#### 2.2.2 อัลกอริทึม C4.5 คล้ายกับอัลกอริทึม ID3 แต่มี

การพัฒนาเพิ่มเติม โดยจะใช้ค่ามาตรฐานเกนในการตัดสินใจเลือกคุณสมบัติที่ใช้เป็นรากหรือโหนดในต้นไม้ โดยการคำนวณค่าเกณฑ์จะเลือกคุณสมบัติที่มีค่าเกนสูงที่สุดมาเป็นรากหรือโหนด ค่าเกนนี้คำนวณโดยใช้ความรู้จากทฤษฎีสารสนเทศ คือค่าสารสนเทศของข้อมูลขึ้นอยู่กับความน่าจะเป็นของข้อมูล โดยค่าสารสนเทศที่น้อย หมายถึงข้อมูลชุดนั้นมีความแตกต่างกันน้อย แต่ถ้าค่าสารสนเทศมาก หมายถึง ข้อมูลชุดนั้นมีความแตกต่างกันมาก หรือประกอบด้วยตัวอย่างหลายพวกที่มีจำนวนใกล้เคียงกัน [7]

#### 2.2.3 อัลกอริทึม Classification and Regression

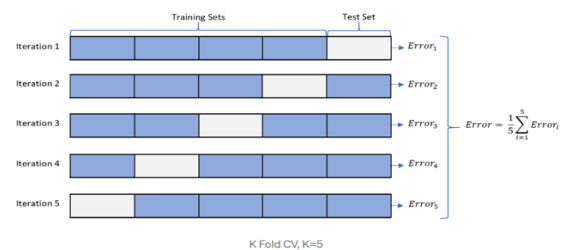
Trees หรือ CART เป็นการสร้างต้นไม้ตัดสินใจแบบ Binary ซึ่งประกอบด้วย กิ่งหรือแขนง 2 กิ่งสำหรับแต่ละโหนด เทคนิคนี้จะทำการแบ่งระเบียบในชุดข้อมูลฝึกสอนออกเป็น ระเบียบย่อยที่ให้ค่าเป้าหมายที่เหมือนกัน [5]

### 2.3 การประเมินประสิทธิภาพของต้นไม้ตัดสินใจ

การประเมินประสิทธิภาพของโมเดลเพื่อทำนายด้วยต้นไม้ตัดสินใจนั้น จะขึ้นอยู่กับวิธีการจัดการชุดข้อมูลตามหลักการเรียนรู้แบบมีผู้สอนและวิธีการวัดประสิทธิภาพของโมเดล ดังต่อไปนี้

#### 2.3.1 วิธีการจัดชุดข้อมูลแบบ K-fold Cross-Validation

เป็นการวิเคราะห์โดยจัดชุดข้อมูลเพื่อประเมินโมเดลในลักษณะการตรวจสอบไขว้กัน (Cross Validation) [6] โดยใช้หลักการสุ่มตัวอย่างเพื่อจัดชุดข้อมูลเป็น K ส่วนเท่า ๆ กัน ด้วยการนำบางส่วนจากชุดข้อมูลนั้นมาตรวจสอบผลลัพธ์จากการทำการตรวจสอบไขว้กัน เช่น กำหนดให้  $K = 5$  ก็จะได้ชุดข้อมูล 5 ชุดเพื่อคำนวณค่า 5 รอบ โดยแต่ละรอบการคำนวณ ข้อมูลชุดหนึ่งจากข้อมูล 5 ชุด จะถูกเลือกออกมาเพื่อใช้เป็นข้อมูลทดสอบ และข้อมูลที่เหลืออีก 4 ชุด จะถูกใช้เป็นข้อมูลสำหรับการเรียนรู้ ตัวอย่างดังภาพที่ 2



ภาพที่ 2: การวิเคราะห์ 5 - fold Cross Validation [8]

#### 2.3.2 วิธีการวัดประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูล

คำนวณจากค่าที่เกี่ยวข้อง 4 ค่า [3] ได้แก่

1. ค่าความแม่นยำ (Precision) เป็นการวัดความแม่นยำของโมเดล โดยพิจารณาแยกทีละคลาส ดังสมการที่ 2.1
2. ค่าความระลึก (Recall) เป็นการวัดความถูกต้องของโมเดล โดยพิจารณาแยกทีละคลาส ดังสมการที่ 2.2
3. ค่าความถูกต้อง (Accuracy) เป็นการวัดค่าความถูกต้องของโมเดล ซึ่งจะพิจารณารวมทุกคลาสดังสมการที่ 2.3
4. ค่าเอฟเมซเชอร์ (F-Measure) เป็นการวัดค่า Precision และ Recall พร้อมกันของโมเดล โดยพิจารณาแยกทีละคลาส ดังสมการที่ 2.4

$$Precision = \left( \frac{TP}{TP + FP} \right) \times 100\% \tag{2.1}$$

$$Recall = \left( \frac{TP}{TP + FN} \right) \times 100\% \tag{2.2}$$

$$Accuracy = \left( \frac{TP + TN}{TP + FP + FN + TN} \right) \times 100\% \tag{2.3}$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{2.4}$$

โดยที่ True Positive (TP) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งกำลังสนใจอยู่

True Negative (TN) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งไม่ได้สนใจอยู่

False Positive (FP) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งกำลังสนใจอยู่

False Negative (FN) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งไม่ได้สนใจอยู่

#### 2.4 งานวิจัยที่เกี่ยวข้อง

การทำเหมืองข้อมูลเป็นเทคนิคที่สำคัญในการวิเคราะห์ข้อมูลเพื่อสนับสนุนการตัดสินใจทางธุรกิจพบว่ามิจานวิจัย เช่น การศึกษาระบบสนับสนุนการตัดสินใจในการอนุมัติสินเชื่อรถจักรยานยนต์ โดยการใช้เทคนิคการทำเหมืองข้อมูล ด้วยโมเดลต้นไม้ตัดสินใจ ผลจากการนำข้อมูลมาทดสอบประสิทธิภาพของโมเดลพบว่า มีผลการทำนายข้อมูลถูกต้องร้อยละ 82.80 จึงนำมาพัฒนาเป็นเว็บแอปพลิเคชัน เพื่อให้ได้ระบบสนับสนุนการตัดสินใจที่มีประสิทธิภาพ [9] นอกจากนี้ยังมีการศึกษาระบบสนับสนุนการตัดสินใจในการช่วยเลือกสมัครบัตรเครดิต กรณีศึกษาร้านค้าแห่งหนึ่งในประเทศไทย เพื่อช่วยให้ลูกค้าสามารถเลือกสมัครบัตรเครดิตได้ตรงกับความต้องการ และมีความเหมาะสมกับวิถีการดำเนินชีวิตมากยิ่งขึ้น ผลการศึกษาพบว่า อัลกอริทึมการจำแนกต้นไม้ตัดสินใจมีความถูกต้องอยู่ที่ร้อยละ 86.47 ซึ่งเป็นอัลกอริทึมที่มีประสิทธิภาพสูงที่สุด [10]

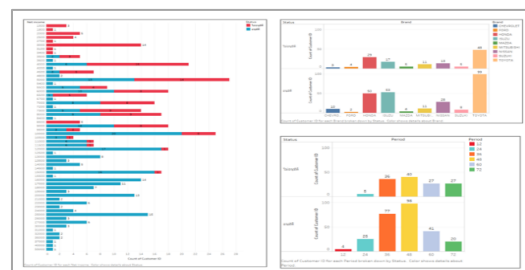
### 3. วิธีดำเนินการ

เพื่อให้บรรลุตามวัตถุประสงค์ของการวิจัยนี้ ผู้วิจัยแบ่งขั้นตอนการวิจัยออกเป็น 4 ขั้นตอน โดยประยุกต์มาจากกระบวนการทำเหมืองข้อมูลแบบ CRISP-DM Process ดังต่อไปนี้

3.1 การทำความเข้าใจกับปัญหาและข้อมูล (Business and Data Understanding) เป็นขั้นตอนที่เกี่ยวข้องกับการทำความเข้าใจกับปัญหาของธุรกิจซื้อขายรถยนต์มือสอง และทำความเข้าใจกับข้อมูลเพื่อตรวจสอบคุณสมบัติของข้อมูล ทั้งในด้านความถูกต้องและความสมบูรณ์ของข้อมูล

โดยใช้วิธีการสัมภาษณ์และประชุมระดมสมองร่วมกับผู้ที่เกี่ยวข้องของบริษัท สรรธร 99 จำกัด จากการทำความเข้าใจกับข้อมูล พบว่า ผู้ประกอบการได้มีการเก็บรวบรวมข้อมูลรายละเอียดของลูกค้าระหว่างปี พ.ศ. 2560 – 2562 จำนวนทั้งสิ้น 400 รายการ ซึ่งแต่ละรายการประกอบด้วยคุณลักษณะ จำนวน 11 แอททริบิวต์ ได้แก่ เพศของลูกค้า อายุของลูกค้า อาชีพของลูกค้า อายุงานของลูกค้า รายได้ของลูกค้า ค่าใช้จ่ายของลูกค้า รายได้สุทธิของลูกค้า ยี่ห้อของรถ ประเภทของรถ จำนวนวงของรถ และสถานะการอนุมัติสินเชื่อ โดยบริษัทมีรถที่นำมาขายสต็อกขายไว้ประมาณ 70-100 คัน ที่ผ่านมาจากบริษัทยังไม่มีระบบคาดการณ์ในการอนุมัติสินเชื่อ ดังนั้น จึงมีความเสี่ยงที่จะเกิดหนี้ที่ไม่ก่อให้เกิดรายได้ เนื่องจากมีการอนุมัติสินเชื่อให้แก่ลูกค้าที่ไม่เหมาะสม รวมถึง มีการอนุมัติขาดประสิทธิภาพส่งผลให้มีลูกค้าค้างชำระจำนวนมาก

3.2 การจัดเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนของการนำข้อมูลที่ได้นำมาทำความสะอาดข้อมูล (Data Cleaning) เพื่อศึกษาและวิเคราะห์ข้อมูลเชิงลึกการขอสินเชื่อรถยนต์ของลูกค้าด้วยโปรแกรม Tableau รวมทั้งดัดแปลงข้อมูล (Data Transformation) ให้อยู่ในรูปแบบกลุ่มของข้อมูล ให้สามารถสร้าง โมเดลด้วยเทคนิคต้นไม้ตัดสินใจได้ จากผลการศึกษาและวิเคราะห์ข้อมูลเชิงลึกของลูกค้าที่ขอสินเชื่อรถยนต์ พบว่า จากข้อมูลทั้งหมด 11 แอททริบิวต์ มีคุณลักษณะสำคัญที่เกี่ยวข้องกับการอนุมัติสินเชื่อ จำนวน 6 แอททริบิวต์ ได้แก่ 1) รายได้ของลูกค้า 2) ค่าใช้จ่ายของลูกค้า 3) อายุงานของลูกค้า 4) อายุของลูกค้า 5) ประเภทของรถ และ 6) ยี่ห้อของรถ ที่มีความสัมพันธ์กับสถานะการอนุมัติสินเชื่ออย่างเห็นได้ชัด ดังตัวอย่างผลการวิเคราะห์ข้อมูลเชิงลึกของข้อมูล ในภาพที่ 3



ภาพที่ 3: แดชบอร์ดผลการวิเคราะห์ข้อมูลเชิงลึกของข้อมูลที่เกี่ยวข้องกับสถานะการอนุมัติสินเชื่อ

สำหรับคุณลักษณะสำคัญที่เกี่ยวข้องกับการอนุมัติสินเชื่อ ที่มีลักษณะเป็นข้อมูลเชิงปริมาณ (Quantitative Data) และจำเป็นต้องตัดแปลงข้อมูลให้อยู่ในรูปแบบของกลุ่ม (Class) นั้น ผู้ประกอบการมีความต้องการให้จัดกลุ่มตามหลักเกณฑ์การจัดกลุ่มลูกค้าของบริษัท ทำให้ข้อมูลใหม่ ของทั้ง 4 แอททริบิวท์ ได้แก่ 1) ช่วงรายได้ 2) ช่วงค่าใช้จ่าย 3) ช่วงอายุงาน และ 4) ช่วงอายุ ได้ค่าของช่วงข้อมูลใหม่ที่ประกอบด้วย 5 Class สำหรับนำไปใช้สร้างโมเดลทำนายด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ

3.3 การสร้างและประเมิน โมเดล (Modeling and Evaluation) เป็นขั้นตอนการสร้างโมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์ด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ เนื่องจากเป็นขั้นตอนวิธีที่ให้ผลลัพธ์ในรูปแบบเงื่อนไขที่สามารถนำไปประยุกต์ใช้งานได้ง่าย โดยงานวิจัยนี้จะใช้โปรแกรม Rapid Miner เป็นเครื่องมือในการเรียนรู้แบบมีผู้สอนและวัดประสิทธิภาพของโมเดลด้วยการวิเคราะห์แบบ 10-fold Cross-Validation

3.4 การนำโมเดลไปใช้งาน (Model Implementation) เป็นขั้นตอนของการนำโมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์ด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ ไปประยุกต์ใช้งานจริงตามความต้องการของผู้ประกอบการ โดยจะใช้ฟังก์ชัน VLOOKUP ใน โปรแกรม Microsoft Excel เพื่อทำนายสถานะอนุมัติให้สินเชื่อรถยนต์ จากนั้นดำเนินการประเมินความพึงพอใจของผู้ใช้งานที่มีผลต่อการนำโมเดลไปประยุกต์ใช้งานทั้ง 5 ด้าน ได้แก่ 1) ด้านตรงตามความต้องการ (Function Requirement) 2) ด้านสามารถทำงานได้ตามหน้าที่ (Function) 3) ด้านความง่ายต่อการใช้งาน (Usability) 4) ด้านประสิทธิภาพ (Performance) และ 5) ด้านการรักษาความปลอดภัยของข้อมูล (Security) รวมทั้งสิ้น 25 ข้อคำถาม โดยกำหนดเกณฑ์การวัดระดับแต่ละข้อคำถาม 5 คือ พึงพอใจมากที่สุด และ 0 คือ ไม่พึงพอใจ ทำการเก็บรวบรวมข้อมูลจากผู้ที่เกี่ยวข้องในสถานประกอบการ และนำมาหาค่าเฉลี่ยรวมของระดับความพึงพอใจ

4. ผลการดำเนินการ

4.1 ผลการออกแบบและพัฒนาโมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์ด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ

ผลการศึกษาในขั้นตอนวิจัยที่ 3.3 เป็นผลที่ได้การออกแบบและพัฒนาโมเดลต้นไม้ตัดสินใจ แบบ 10-fold Cross-Validation ด้วยโปรแกรม Rapid Miner ดังภาพที่ 4



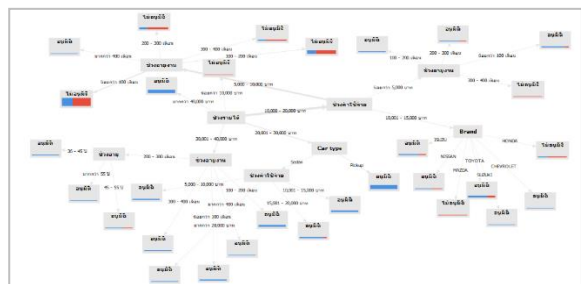
ภาพที่ 4: การสร้างโมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์ด้วย โปรแกรม Rapid Miner

ผลการออกแบบและพัฒนาโมเดลด้วยอัลกอริทึม C4.5 ได้กฎการจำแนก (Classification Rule) จำนวนทั้งสิ้น 30 กฎ (ภาพที่ 5) และแผนภูมิต้นไม้ตัดสินใจ ดังภาพที่ 6

```

Tree
|
| ช่วงรายได้ = 10,000 - 20,000 บาท
| | ช่วงค่าใช้จ่าย = 10,001 - 15,000 บาท
| | | Brand = CHEVROLET: อนุมัติ {อนุมัติ=1, ไม่อนุมัติ=0}
| | | | Brand = HONDA: ไม่อนุมัติ {อนุมัติ=2, ไม่อนุมัติ=4}
| | | | Brand = SUZUKI: อนุมัติ {อนุมัติ=0, ไม่อนุมัติ=2}
| | | | Brand = MAZDA: ไม่อนุมัติ {อนุมัติ=0, ไม่อนุมัติ=2}
| | | | Brand = NISSAN: อนุมัติ {อนุมัติ=3, ไม่อนุมัติ=2}
| | | | Brand = SUZUKI: อนุมัติ {อนุมัติ=1, ไม่อนุมัติ=0}
| | | | Brand = TOYOTA: อนุมัติ {อนุมัติ=14, ไม่อนุมัติ=6}
| | | ช่วงรายได้ = 5,000 บาท
| | | | ช่วงอายุงาน = 100 - 200 เดือน: ไม่อนุมัติ {อนุมัติ=12, ไม่อนุมัติ=28}
| | | | | ช่วงอายุงาน = 200 - 300 เดือน: ไม่อนุมัติ {อนุมัติ=2, ไม่อนุมัติ=13}
| | | | | ช่วงอายุงาน = 300 - 400 เดือน: ไม่อนุมัติ {อนุมัติ=2, ไม่อนุมัติ=5}
| | | | | ช่วงอายุงาน = มากกว่า 100 เดือน: ไม่อนุมัติ {อนุมัติ=34, ไม่อนุมัติ=56}
| | | | | ช่วงอายุงาน = มากกว่า 400 เดือน: อนุมัติ {อนุมัติ=3, ไม่อนุมัติ=0}
| | | | ช่วงอายุงาน = น้อยกว่า 5,000 บาท
| | | | | ช่วงอายุงาน = 200 - 300 เดือน: อนุมัติ {อนุมัติ=6, ไม่อนุมัติ=0}
| | | | | | ช่วงอายุงาน = 200 - 300 เดือน: อนุมัติ {อนุมัติ=3, ไม่อนุมัติ=1}
| | | | | | ช่วงอายุงาน = 300 - 400 เดือน: ไม่อนุมัติ {อนุมัติ=0, ไม่อนุมัติ=1}
| | | | | | ช่วงอายุงาน = น้อยกว่า 100 เดือน: อนุมัติ {อนุมัติ=0, ไม่อนุมัติ=1}
| | | | | | ช่วงอายุงาน = น้อยกว่า 400 เดือน: อนุมัติ {อนุมัติ=3, ไม่อนุมัติ=0}
| | | | | | ช่วงอายุงาน = 20,001 - 30,000 บาท
| | | | | | | Car type = Pickup: อนุมัติ {อนุมัติ=43, ไม่อนุมัติ=0}
| | | | | | | | Car type = Sedan
| | | | | | | | | ช่วงค่าใช้จ่าย = 10,001 - 15,000 บาท: อนุมัติ {อนุมัติ=14, ไม่อนุมัติ=0}
| | | | | | | | | | ช่วงค่าใช้จ่าย = 15,001 - 20,000 บาท: อนุมัติ {อนุมัติ=8, ไม่อนุมัติ=1}
| | | | | | | | | | | ช่วงค่าใช้จ่าย = 5,000 - 10,000 บาท: อนุมัติ {อนุมัติ=7, ไม่อนุมัติ=0}
| | | | | | | | | | | | ช่วงค่าใช้จ่าย = มากกว่า 20,000 บาท: อนุมัติ {อนุมัติ=3, ไม่อนุมัติ=0}
| | | | | | | | | | | | | Car type = Sedan
| | | | | | | | | | | | | | ช่วงอายุงาน = 100 - 200 เดือน: อนุมัติ {อนุมัติ=17, ไม่อนุมัติ=0}
| | | | | | | | | | | | | | | ช่วงอายุงาน = 200 - 300 เดือน
| | | | | | | | | | | | | | | | ช่วงอายุ = 35 - 45 ปี: อนุมัติ {อนุมัติ=7, ไม่อนุมัติ=0}
| | | | | | | | | | | | | | | | | ช่วงอายุ = 45 - 55 ปี: อนุมัติ {อนุมัติ=2, ไม่อนุมัติ=1}
| | | | | | | | | | | | | | | | | | ช่วงอายุ = น้อยกว่า 55 ปี: อนุมัติ {อนุมัติ=1, ไม่อนุมัติ=0}
| | | | | | | | | | | | | | | | | | | ช่วงอายุงาน = 300 - 400 เดือน: อนุมัติ {อนุมัติ=7, ไม่อนุมัติ=0}
| | | | | | | | | | | | | | | | | | | | ช่วงอายุงาน = น้อยกว่า 100 เดือน: อนุมัติ {อนุมัติ=10, ไม่อนุมัติ=0}
| | | | | | | | | | | | | | | | | | | | | ช่วงอายุงาน = มากกว่า 400 เดือน: อนุมัติ {อนุมัติ=4, ไม่อนุมัติ=0}
| | | | | | | | | | | | | | | | | | | | | | ช่วงรายได้ = น้อยกว่า 10,000 บาท: ไม่อนุมัติ {อนุมัติ=0, ไม่อนุมัติ=2}
| | | | | | | | | | | | | | | | | | | | | | | ช่วงรายได้ = มากกว่า 40,000 บาท: อนุมัติ {อนุมัติ=35, ไม่อนุมัติ=0}
    
```

ภาพที่ 5: กฎการจำแนกที่ได้ของโมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์



ภาพที่ 6: แผนภูมิต้นไม้ตัดสินใจของโมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์

ผลการวัดประสิทธิภาพของโมเดลทำนายรูปแบบการอนุมัติให้สินเชื่อรถยนต์ทั้ง 4 ค่า พบว่า โมเดลที่ได้มีค่าความถูกต้อง ร้อยละ 79.54 ค่าความแม่นยำ ร้อยละ 92.15 ค่าความระลึกร้อยละ 76.12 และค่าเอฟเมซเซอร์ ร้อยละ 83.11



#### 4.2 ผลการนำโมเดลไปใช้งาน

ผลการนำโมเดลไปใช้งาน เป็นการนำผลการจำแนกที่ได้จากโมเดลไปประยุกต์ใช้งานเพื่อทำนายสถานะอนุมัติให้สินเชื่อรถยนต์ด้วยโปรแกรม Microsoft Excel โดยใช้ฟังก์ชัน VLOOKUP ได้หน้าจอการใช้งาน ดังภาพที่ 7

ช่วงรายได้	ช่วงค่าไม่จ่าย	ช่วงอายุงาน	ช่วงอายุ	Car type	Brand	ผลลัพธ์
10,000 - 20,000 บาท	10,001 - 15,000 บาท	301 - 400 เดือน	36 - 45 ปี	Sedan	ISUZU	อนุมัติ
					CHEVROLET	
					HONDA	
					ISUZU	
					MAZDA	
					NISSAN	
					SUZUKI	
					TOYOTA	

ภาพที่ 7: หน้าจอการประยุกต์ใช้งานโปรแกรม Microsoft Excel เพื่อทำนายสถานะอนุมัติให้สินเชื่อรถยนต์

ผู้ประกอบการมีความพึงพอใจในภาพรวมเท่ากับ 3.68 เมื่อพิจารณาเป็นรายด้าน พบว่า ด้านความง่ายต่อการใช้งาน มีค่าเฉลี่ยความพึงพอใจในระดับมาก ( $\bar{x} = 4.8$ ) ส่วนด้านอื่นๆ มีค่าเฉลี่ยความพึงพอใจความพึงพอใจในระดับปานกลาง ( $3.2 < \bar{x} < 3.8$ )

#### 5. สรุป

งานวิจัยนี้มีวิธีดำเนินการ 4 ขั้นตอนคือ 1) การทำความเข้าใจกับปัญหาและข้อมูล 2) การจัดเตรียมข้อมูล 3) การสร้างและประเมินโมเดล และ 4) การนำโมเดลไปใช้งาน ผลการศึกษาพบว่า การวิเคราะห์ข้อมูลเชิงลึกของข้อมูลทางธุรกิจ และการคัดแปลงข้อมูลบนพื้นฐานความต้องการทางธุรกิจ ก่อนนำข้อมูลไปออกแบบและพัฒนาโมเดลทำนายรูปแบบการตัดสินใจด้วยขั้นตอนวิธีต้นไม้ เป็นอีกแนวทางหนึ่งในการคัดเลือกคุณลักษณะที่สำคัญของการพัฒนาโมเดลทำนาย ทำให้ได้ค่าประสิทธิภาพของโมเดลที่เหมาะสม อีกทั้งการนำโมเดลไปประยุกต์ใช้งานด้วยโปรแกรมที่ง่ายต่อการใช้งาน นอกจากจะสามารถรองรับข้อมูลที่มีการเปลี่ยนแปลงได้แล้ว ยังสามารถช่วยสนับสนุนการตัดสินใจทางธุรกิจได้ อย่างไรก็ตาม การพัฒนาโมเดลทำนายภายใต้เงื่อนไขและข้อจำกัดของแหล่งข้อมูลที่ได้รับมานั้น ย่อมส่งผลต่อประสิทธิภาพของโมเดล ดังนั้น ควรจะต้องแนะนำผู้ประกอบการในเรื่องการเก็บข้อมูลให้มีความหลากหลาย ครอบคลุมและเก็บข้อมูลอย่างต่อเนื่อง

#### เอกสารอ้างอิง

- [1] ศูนย์วิจัยกสิกรไทย ,ตลาดรถมือสอง โควิตราคาพุ่ง แพลตฟอร์มออนไลน์มาแรง, สืบค้นเมื่อ 5 กุมภาพันธ์ 2565, จาก <https://kasikomresearch.com/th/analysis/k-social-media/Pages/used-car-FB-20-10-21.aspx> ,2565.
- [2] (มหาชน)บริษัท ไมค์ ลิสซิ่ง จำกัด, บัญชีเสี่ยง, สืบค้นเมื่อ 5 กุมภาพันธ์ 2565, จาก [http://capital.sec.or.th/webapp/corp\\_fin/datafile/dsfi/0786130000-00-00TH204L01007.PDF](http://capital.sec.or.th/webapp/corp_fin/datafile/dsfi/0786130000-00-00TH204L01007.PDF)
- [3] เอกสิทธิ์ พืชรวงศ์ศักดิ์,การวิเคราะห์ข้อมูลด้วยเทคนิคค้ำค่าไมน์นิงเบื้องต้น, บริษัทเอเชีย ดิจิตอลกราฟิมพ์, พิมพ์ครั้งที่ 2, กรุงเทพฯ, 2557.
- [4] กฤษณะ ไวยมัย, การใช้เทคนิคค้ำค่าไมน์นิงเพื่อพัฒนาคุณภาพการศึกษาคณะวิศวกรรมศาสตร์, วิทยานิพนธ์ปริญญาโทมหาบัณฑิต มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพมหานคร, 2544.
- [5] Paranya Palwisut, "Improving Decision Tree Technique in Imbalanced Data Sets Using SMOTE for Internet Addiction Disorder Data," *Information Technology Journal*, Vol. 12, No. 1, pp. 54-63, 2016.
- [6] F.M. Javed Mehedi Shamrat, Rumesh Ranjan, Khan Md. Hasib, Amit Yadav, and Abdul Hasib Siddique, "Performance Evaluation among ID3, C4.5, and CART Decision Tree Algorithms," International Conference on Pervasive Computing and Social Networking [ICPCSN 2021] Salem, Tamil Nadu, India, pp. 127-142, March 19-20, 2021.
- [7] อโชนัย ศิลเทพาเวทย์, "แบบจำลองเพื่อพัฒนาคุณภาพของผลิตภัณฑ์เอชซีไอในโรงงานอุตสาหกรรมสารคดีสก็ด้วยเทคนิคต้นไม้ตัดสินใจ," วิทยานิพนธ์ ปริญญาโทมหาบัณฑิต มหาวิทยาลัยจุฬาลงกรณ์ กรุงเทพมหานคร, 2554.
- [8] Towards Data Science, "Cross-Validation: K Fold vs Monte Carlo," Jan, 2021. [Online]. Available: <https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b>. [Accessed Feb. 10,2022]
- [9] ชรินทร์พิทย์ วรรณกลาง, "ระบบสนับสนุนการตัดสินใจในการอนุมัติสินเชื่อรถยนต์จักรยานยนต์," วิทยานิพนธ์ปริญญาโทมหาบัณฑิต มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพมหานคร, 2560.
- [10] จันทราภรณ์ ชุนอักษร, "ระบบสนับสนุนการตัดสินใจในการช่วยเลือกสมัครบัตรเครดิต กรณีศึกษาธนาคารแห่งหนึ่งในประเทศไทย," วิทยานิพนธ์ปริญญาโทมหาบัณฑิต มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพมหานคร, 2560.

# แบบจำลองเพื่อการพยากรณ์การส่งออกสินค้าเกษตรโดยใช้เทคนิคเหมืองข้อมูล

## Forecasting Model for Agricultural Export using Data Mining Techniques

สุภาพร จำรูราย (Suphapon Jamrurai)<sup>1</sup> และศักดิ์ชาย ตั้งวรรณวิทย์ (Sakchai Tangwannawit)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858287@email.kmutnb.ac.th, <sup>2</sup>sakchai.t@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้ทำการพัฒนาแบบจำลองเพื่อการพยากรณ์การส่งออกสินค้าเกษตรโดยใช้เทคนิคเหมืองข้อมูลมีวัตถุประสงค์เพื่อ 1) เปรียบเทียบเทคนิคการพยากรณ์ที่เหมาะสม 2) พัฒนาแบบจำลองการพยากรณ์ 3) พัฒนาระบบธุรกิจอัจฉริยะเพื่อการพยากรณ์การส่งออกข้าว ยางพารา และมันสำปะหลัง โดยสร้างตัวแบบพยากรณ์ด้วยเทคนิคอนุกรมเวลา 3 วิธี ได้แก่ วิธีการปรับให้เรียบเอ็กซ์โปเนนเชียลแบบสองครั้ง วิธีการปรับให้เรียบเอ็กซ์โปเนนเชียลด้วยวิธีของวินเตอร์ และวิธีแยกองค์ประกอบ เปรียบเทียบแบบจำลองด้วยค่าร้อยละความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAPE) และค่าความผิดพลาดสัมบูรณ์เฉลี่ย (MAD) ที่ต่ำที่สุด ผลการวิจัยพบว่าการพยากรณ์ปริมาณการส่งออกข้าวและยางพารา วิธีปรับให้เรียบเอ็กซ์โปเนนเชียลด้วยวิธีของวินเตอร์แบบคูณให้ความแม่นยำมากที่สุด และการพยากรณ์ปริมาณการส่งออกมันสำปะหลัง วิธีปรับให้เรียบเอ็กซ์โปเนนเชียลด้วยวิธีของวินเตอร์แบบบวกมีความแม่นยำมากที่สุด เนื่องจากให้ค่าร้อยละความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAPE) และค่าความผิดพลาดสัมบูรณ์เฉลี่ย (MAD) ต่ำที่สุด

**คำสำคัญ:** เหมืองข้อมูล เทคนิคอนุกรมเวลา วิธีการปรับให้เรียบเอ็กซ์โปเนนเชียลแบบสองครั้ง วิธีการปรับให้เรียบเอ็กซ์โปเนนเชียลด้วยวิธีของวินเตอร์ วิธีแยกองค์ประกอบ

### Abstract

*This research was to develop a model for forecasting agricultural exports using data mining techniques. The step of the research was to consist of*

1) compare data mining models, 2) develop a forecasting model, and 3) develop a business intelligence system for forecasting the export volume of rice, rubber, and cassava. This research used three models 1) double exponential smoothing, 2) winter exponential smoothing, and 3) decomposition. The models were compared with the lowest mean absolute percentage error (MAPE) and mean absolute deviation (MAD). The results showed that the winter exponential smoothing method with multiplicative is the most accurate of rice and rubber exports forecast and the winter exponential smoothing method with additive is the most accurate of cassava export forecast. Because it gives the lowest the mean absolute percentage error (MAPE) and mean absolute deviation (MAD).

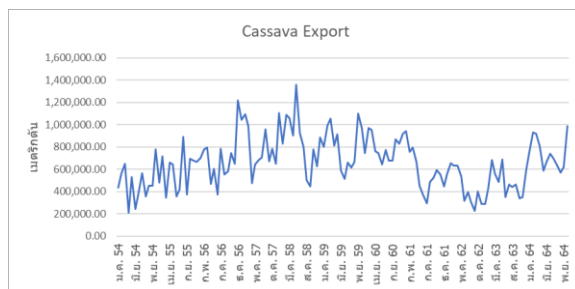
**Keywords:** Data Mining, Time Series Model, Double Exponential Smoothing, Winter Exponential Smoothing, Decomposition.

### 1. บทนำ

เกษตรกรรมเป็นภาคการผลิตที่สำคัญของไทยมาเป็นเวลานาน เนื่องจากมีพื้นที่ที่เอื้อต่อการทำการเกษตรส่งผลให้ในแต่ละภูมิภาคมีผลผลิตที่สำคัญได้แก่ ภาคตะวันออกเฉียงเหนือการผลิตสำคัญ ได้แก่ ข้าว มันสำปะหลังและยางพารา ภาคเหนือ การผลิตที่สำคัญ ได้แก่ ข้าว ลำไย และยางพารา ในขณะที่ภาคใต้มีการผลิตที่สำคัญ ได้แก่ ข้าว และยางพารา และภาคกลางมีการผลิตที่สำคัญ ได้แก่ ข้าว มันสำปะหลัง และยางพารา [1] ทำให้การส่งออกสินค้าเกษตรเป็นหนึ่งในรายได้สำคัญที่ส่งผล

ต่อการพัฒนาเศรษฐกิจของไทยเป็นอย่างมาก โดยผลผลิตที่ส่งออกและสร้างมูลค่าเพิ่มเป็นพืชเศรษฐกิจของไทย ได้แก่ ข้าว ยางพารา และมันสำปะหลัง [2]

แม้ว่าการส่งออกสินค้าเกษตรนั้นจะสร้างรายได้เป็นจำนวนมาก แต่เมื่อผู้วิจัยศึกษาข้อมูลในอดีตแล้วพบว่า ปริมาณการส่งออก มีความผันผวนค่อนข้างมาก ยกตัวอย่างเช่น ปริมาณการส่งออกมันสำปะหลังที่มีแนวโน้มเพิ่มขึ้นตั้งแต่ปีพ.ศ. 2554 ถึงพ.ศ. 2556 มีความผันผวนค่อนข้างคงที่ในปีพ.ศ. 2557 ถึงพ.ศ. 2558 จากนั้นในปีพ.ศ. 2559 ถึงพ.ศ. 2562 มีแนวโน้มลดลงอย่างเห็นได้ชัด และมีแนวโน้มเพิ่มขึ้นอีกครั้งในปีพ.ศ. 2563 จนถึงปัจจุบัน โดยความผันผวนที่กล่าวมานี้เมื่อเปรียบเทียบปริมาณการส่งออกในเดือนที่มากที่สุดและเดือนที่น้อยที่สุดแล้วพบที่มีความแตกต่างกันมากถึง 1,149,002.36 เมตริกตัน



ภาพที่ 1: ปริมาณการส่งออกมันสำปะหลัง

ด้วยปัจจุบันเทคโนโลยีเข้ามามีบทบาทที่ส่งผลกระทบต่อการแข่งขันและการดำเนินธุรกิจเป็นอย่างมาก เพื่อให้สามารถแข่งขันกับคู่แข่งในยุคปัจจุบันและวางแผนการดำเนินงานในอนาคตได้จึงต้องนำเทคโนโลยีมาประยุกต์ใช้ให้เหมาะสม จึงทำให้ผู้ประกอบการหลายรายพัฒนาและปรับปรุงกระบวนการเพื่อสร้างโอกาสทางธุรกิจโดยการคิดค้นเครื่องมือและกลยุทธ์ต่างๆ เช่นการพยากรณ์ราคาสินค้าและยอดขายสินค้า เช่นเดียวกับผู้วิจัยในอดีตที่มีการวิจัยในด้านการพยากรณ์การส่งออกสินค้าค่อนข้างมาก แต่ส่วนใหญ่จะทำการพยากรณ์มูลค่าการส่งออก และมักมีข้อเสนอแนะว่าเมื่อเวลาผ่านไปและปริมาณข้อมูลมีจำนวนมากขึ้นควรนำมาพัฒนาตัวแบบพยากรณ์ให้เหมาะสมกับปัจจุบันเพื่อเพิ่มประสิทธิภาพของตัวแบบพยากรณ์ เช่นตัวแบบพยากรณ์มูลค่าการส่งออกข้าวหอมมะลิ [3] การพยากรณ์มูลค่าการส่งออกน้ำยางชั้น [4] เป็นต้น

ผู้วิจัยจึงมีแนวคิดในการศึกษาทฤษฎีการพยากรณ์ การสร้างตัวแบบพยากรณ์และการคัดเลือกตัวแบบพยากรณ์ที่เหมาะสมที่สุดสำหรับชุดข้อมูลปริมาณการส่งออกข้าว ยางพารา และมันสำปะหลัง เพื่อช่วยให้เกษตรกรใช้เป็นแนวทางวางแผนการเพาะปลูกหรือการเก็บเกี่ยวผลผลิต รวมทั้งสามารถใช้เป็นข้อมูลประกอบการวางแผนการดำเนินงาน การผลิตและการส่งออกสำหรับผู้ประกอบการหรือหน่วยงานที่เกี่ยวข้องได้

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง เพื่อนำมาประยุกต์ใช้ โดยมีรายละเอียดดังต่อไปนี้

### 2.1 การทำเหมืองข้อมูล (Data mining)

กระบวนการในการค้นหารูปแบบหรือความสัมพันธ์ของข้อมูล ด้วยเทคนิคต่างๆ ทั้งทางสถิติ ทางคณิตศาสตร์ หรือเทคนิคทางวิทยาการคอมพิวเตอร์ เพื่อให้ได้สารสนเทศที่ไม่เคยรู้ล่วงหน้ามาก่อน เพื่อนำไปประยุกต์ใช้ในการใช้งาน เช่น ช่วยในการตัดสินใจวางแผนงาน เพื่อสร้างความได้เปรียบในการแข่งขัน โดยการทำเหมืองข้อมูลมีขั้นตอนดังนี้ 1) Data Cleaning 2) Data Integration 3) Data Selection 4) Data Transformation 5) Data Mining 6) Pattern Evaluation 7) Knowledge Representation

### 2.2 แบบจำลองอนุกรมเวลา (Time Series Models)

การใช้ข้อมูลในอดีตมาพยากรณ์ข้อมูลในอนาคต โดยมีเวลาเป็นตัวแปรอิสระซึ่งอาจกำหนดเป็นวัน เดือนปี หรืออื่นๆ โดยมีส่วนประกอบพื้นฐาน ได้แก่ แนวโน้ม ความผันแปรตามฤดูกาล การผันแปรตามวัฏจักร และการผันแปรที่ผิดปกติ [5]

2.2.1 วิธีปรับเรียบเอ็กซ์โปเนนเชียลแบบสองครั้ง (Double Exponential Smoothing) ให้ผลการพยากรณ์ได้ดีกับข้อมูลที่มีแนวโน้มและไม่มี ความแปรผันตามฤดูกาล ลักษณะของวิธีการนี้คือ ในครั้งแรกจะทำการปรับเรียบข้อมูลเดิม และครั้งที่สองทำการปรับเรียบข้อมูลที่ได้จากการปรับเรียบในครั้งแรก

2.2.2 วิธีปรับเรียบแบบเอ็กซ์โปเนนเชียลด้วยวิธีของวินเตอร์ (Winter's Method) มีวิธีการเดียวกันกับวิธีปรับเรียบเอ็กซ์โปเนนเชียลแบบสองครั้ง แต่จะทำการปรับ

เรียบเพิ่มอีก 1 ครั้ง โดยมี 2 รูปแบบ คือ รูปแบบบวก และ รูปแบบคูณ ซึ่งจะให้ผลการพยากรณ์ได้ดีกับข้อมูลที่มีแนวโน้มและมีความผันแปรตามฤดูกาล

2.2.3 วิธีแยกส่วนประกอบ (Decomposition Method) เป็นการแยกส่วนประกอบของข้อมูลเป็น 4 ส่วน คือ แนวโน้ม ฤดูกาล วัฏจักร และความคิดปกติ เพื่อพิจารณาว่าส่วนประกอบใดมีอิทธิพลต่อการเปลี่ยนแปลงของข้อมูลบ้าง แล้วนำค่าความผันแปรที่ได้นั้นไปพยากรณ์ข้อมูลในอนาคต โดยมี 2 รูปแบบ คือ รูปแบบบวก และ รูปแบบคูณ

### 2.3 การวัดความถูกต้องของการพยากรณ์

2.3.1 ค่าร้อยละความคลาดเคลื่อนสมบูรณ์เฉลี่ย Mean Absolute Percentage Error : MAPE) เป็นตัวชี้วัดความแม่นยำของการพยากรณ์ โดยค่าที่ได้จะเป็นค่าร้อยละ ซึ่งหากมีค่าต่ำแสดงว่าตัวแบบพยากรณ์มีความแม่นยำสูง

2.3.2 ค่าความผิดพลาดสัมบูรณ์เฉลี่ย (Mean Absolute Deviation: MAD) เป็นการพิจารณาความแตกต่างของค่าจริงกับค่าพยากรณ์ โดยไม่คำนึงถึงทิศทางของความคลาดเคลื่อน โดยค่าที่ได้จะมีหน่วยเดียวกับค่าที่นำมาพยากรณ์ ซึ่งหากมีค่าต่ำแสดงว่าตัวแบบพยากรณ์มีความแม่นยำสูง

### 2.4 งานวิจัยที่เกี่ยวข้อง

บุญฤทธิ์ [6] ทำการวิจัยตัวแบบพยากรณ์ปริมาณการส่งออกมะม่วงของไทย สร้างตัวแบบพยากรณ์ด้วยวิธีการทางสถิติ 3 วิธี ได้แก่ วิธีแยกองค์ประกอบแบบคูณ (Decomposition) วิธีโฮลต์ - วินเตอร์ (Holt - Winters) และวิธีบ็อกซ์ - เจนกินส์ (Box - Jenkins) เปรียบเทียบความถูกต้องด้วยค่าร้อยละความคลาดเคลื่อนสมบูรณ์เฉลี่ย Mean Absolute Percentage Error : MAPE) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ผลการวิจัยพบว่าวิธีบ็อกซ์-เจนกินส์ เป็นวิธีที่มีความแม่นยำมากที่สุด โดยมีตัวแบบเป็น ARIMA (0,1,2) (0,1,1) เนื่องจากลักษณะของข้อมูลมีแนวโน้มเพิ่มขึ้นและมีความผันแปรของฤดูกาล ทั้งนี้มีข้อเสนอแนะว่าในการวิจัยครั้งต่อไปให้พิจารณาข้อมูลที่เป็นปัจจุบันเพื่อนำมาปรับปรุงตัวแบบพยากรณ์ให้มีความเหมาะสม

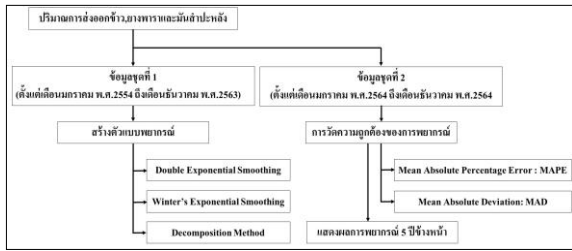
วรางคณา [7] ทำการวิจัยตัวแบบพยากรณ์มูลค่าการส่งออกไม้ยางพาราแปรรูปและเฟอร์นิเจอร์ของประเทศ ไทย สร้างตัวแบบพยากรณ์ด้วยวิธีบ็อกซ์-เจนกินส์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบคูณและวิธีการพยากรณ์รวม เปรียบเทียบความถูกต้องของค่าร้อยละความคลาดเคลื่อนสมบูรณ์เฉลี่ย Mean Absolute Percentage Error : MAPE) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ผลการศึกษาพบว่าภายใต้ค่า MAPE วิธีการพยากรณ์รวมมีความแม่นยำมากที่สุด ขณะที่ภายใต้ค่า RMSE วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบคูณมีความแม่นยำมากที่สุด เนื่องจากข้อมูลที่น่ามาใช้นั้นมีลักษณะที่มีแนวโน้มและมีความผันแปรตามฤดูกาล อย่างไรก็ตามค่าพยากรณ์ทั้ง 2 วิธี มีความน่าเชื่อถือ เนื่องจากไม่มีความแตกต่างกันอย่างมีนัยสำคัญ ทั้งนี้มีข้อเสนอแนะว่าเมื่อมูลค่าการส่งออกไม้ยางพาราแปรรูปและเฟอร์นิเจอร์เป็นปัจจุบันมากขึ้น ควรนำมาปรับปรุงตัวแบบเพื่อให้ได้ตัวแบบพยากรณ์ที่เหมาะสมในอนาคตต่อไป

วรางคณา [8] ทำการพยากรณ์ปริมาณการส่งออกกาแฟสำเร็จรูป สร้างตัวแบบพยากรณ์ด้วยวิธีการทางสถิติทั้งหมด 7 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของโฮลต์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของบราวน์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบเดิม วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังที่มีฤดูกาลอย่างง่าย วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบบวก และวิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบคูณ เปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ด้วยค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ผลการศึกษาพบว่าวิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบคูณมีความแม่นยำมากที่สุด

### 3. วิธีดำเนินการวิจัย

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้นำวิธีการทำเหมืองข้อมูลเพื่อการพยากรณ์ ด้วยเทคนิคการพยากรณ์อนุกรมเวลามาใช้ในการสร้างแบบจำลอง โดยมีรายละเอียดดังนี้

### 3.1 กรอบแนวคิดงานวิจัย



ภาพที่ 2: กรอบแนวคิดงานวิจัย

จากภาพที่ 2 แสดงกรอบแนวคิดการสร้างตัวแบบ การคัดเลือกตัวแบบ และการแสดงผลพยากรณ์ 5 ปีข้างหน้า ของปริมาณการส่งออกข้าว ยางพารา และมันสำปะหลัง

### 3.2 การรวบรวมและวิเคราะห์ข้อมูลเบื้องต้น

ในการวิจัยครั้งนี้ใช้ข้อมูลปริมาณการส่งออกข้าว ยางพารา และมันสำปะหลัง ซึ่งเป็นข้อมูลรายเดือนจากกรมศุลกากร ประมวลผลโดยธนาคารแห่งประเทศไทย ระหว่างปี พ.ศ. 2554 ถึง พ.ศ. 2564 โดยการรวบรวมและจัดเตรียมข้อมูลได้ดำเนินการตามขั้นตอนการทำเหมืองข้อมูลดังที่กล่าวไว้ข้างต้น ซึ่งผู้วิจัยได้นำข้อมูลทั้ง 3 ชุด มาทำการวิเคราะห์ข้อมูลเบื้องต้นด้วย Time Series Plot ร่วมกับกราฟด้วยโปรแกรม Microsoft Excel และหาค่าผิดปกติ (Outliers) ด้วยวิธี Boxplots เมื่อพบค่าผิดปกติจะแทนที่ด้วยค่าเฉลี่ยการส่งออกของเดือนและปีนั้นๆ แทน

### 3.3 การสร้าง การคัดเลือกแบบจำลอง และการแสดงผล

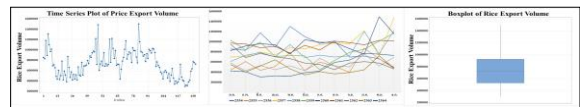
นำข้อมูลปริมาณการส่งออกทั้ง 3 ชนิด ตั้งแต่เดือน มกราคม พ.ศ. 2554 ถึงเดือนธันวาคม พ.ศ. 2563 มาสร้างแบบจำลองด้วยเทคนิคการวิเคราะห์อนุกรมเวลา 3 วิธี ประกอบด้วย 1) วิธีปรับเรียบเอ็กซ์โปเนนเชียลแบบสองครั้ง 2) วิธีปรับเรียบแบบเอ็กซ์โปเนนเชียลด้วยวิธีของวินเตอร์ และ 3) วิธีแยกส่วนประกอบ เปรียบเทียบความแม่นยำด้วยค่าร้อยละความคลาดเคลื่อนสมบูรณ์เฉลี่ย และค่าความผิดพลาดสัมบูรณ์เฉลี่ย หากแบบจำลองใดให้ค่าต่ำที่สุดให้แบบจำลองนั้นเป็นแบบจำลองที่เหมาะสมเพื่อนำไปทวนสอบความแม่นยำและทำการพยากรณ์ข้อมูล 5 ปีข้างหน้า โดยใช้โปรแกรม Minitab เป็นเครื่องมือในการวิเคราะห์

### 3.4 การทวนสอบความแม่นยำของแบบจำลอง

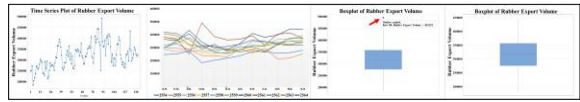
ประเมินความสามารถในการพยากรณ์โดยใช้ข้อมูลการส่งออกสินค้าทั้ง 3 ชนิดตั้งแต่เดือนมกราคม ถึงเดือน ธันวาคม พ.ศ. 2564 มาทวนสอบความแม่นยำและพิจารณาด้วยค่าร้อยละความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAPE)

## 4. ผลการดำเนินงาน

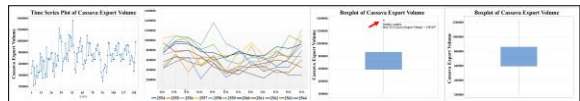
### 4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น



ภาพที่ 3: ผลการวิเคราะห์ข้อมูลเบื้องต้นของการส่งออกข้าว



ภาพที่ 4: ผลการวิเคราะห์ข้อมูลเบื้องต้นของการส่งออกยางพารา



ภาพที่ 5: ผลการวิเคราะห์ข้อมูลเบื้องต้นการส่งออกมันสำปะหลัง

จากผลการวิเคราะห์ข้อมูลเบื้องต้นดังภาพที่ 3, 4 และ 5 พบว่าข้อมูลทั้ง 3 ชุด มีความเป็นแนวโน้มและมีความเป็นฤดูกาลเข้ามาเกี่ยวข้อง ประกอบกับชุดข้อมูลการส่งออก ยางพาราและมันสำปะหลังพบค่าผิดปกติ ผู้วิจัยจึงแทนที่ค่าผิดปกติด้วยค่าเฉลี่ยการส่งออกของสินค้า ในเดือนและปีนั้นแทน และตรวจสอบอีกครั้งผลแสดงไม่พบค่าผิดปกติ จึงนำไปสร้างและเปรียบเทียบแบบจำลองการพยากรณ์

### 4.2 ผลการเปรียบเทียบแบบจำลอง

ในการพยากรณ์ด้วยวิธีปรับเรียบเอ็กซ์โปเนนเชียลแบบสองครั้ง และวิธีปรับเรียบแบบเอ็กซ์โปเนนเชียลด้วยวิธีของวินเตอร์ จำเป็นต้องกำหนดค่าพารามิเตอร์ผู้วิจัยจึงทดลองกำหนดค่าพารามิเตอร์เพื่อให้ได้ค่าที่เหมาะสมมากที่สุด ซึ่งได้ค่าพารามิเตอร์ดังตารางที่ 1 และแสดงผลการวิเคราะห์และการเปรียบเทียบแบบจำลองได้ดังตารางที่ 2

ตารางที่ 1: แสดงผลค่าพารามิเตอร์ที่กำหนด

สินค้า	DES		Winter		
	$\alpha$ (Level)	$\gamma$ (Trend)	$\alpha$ (Level)	$\gamma$ (Trend)	$\beta$ (Seasonal)
ข้าว	0.737	0.160	0.50	0.12	0.00
ยางพารา	1.065	0.023	0.50	0.00	0.00

มัน	0.889	0.022	0.22	0.10	0.05
ลำปะหลัง					

ยางพารา	9.48
มันสำปะหลัง	16.75

ตารางที่ 2: แสดงผลค่า MAPE และ MAD ของแบบจำลอง

แบบจำลอง	MAPE	MAD
<b>ข้าว</b>		
Double Exponential Smoothing	19.42	141642
Winter's Method รูปแบบคูณ	16.96	125588
Winter's Method รูปแบบบวก	17.14	126680
Decomposition รูปแบบคูณ	29.12	193197
Decomposition รูปแบบบวก	29.13	193130
<b>ยางพารา</b>		
Double Exponential Smoothing	10.39	31471
Winter's Method รูปแบบคูณ	7.77	23361
Winter's Method รูปแบบบวก	7.82	23524
Decomposition รูปแบบคูณ	9.72	29983
Decomposition รูปแบบบวก	9.77	30194
<b>มันสำปะหลัง</b>		
Double Exponential Smoothing	28.77	159067
Winter's Method รูปแบบคูณ	18.21	112638
Winter's Method รูปแบบบวก	17.43	107830
Decomposition รูปแบบคูณ	30.58	161816
Decomposition รูปแบบบวก	30.50	161713

จากตารางที่ 2 แสดงให้เห็นว่าการพยากรณ์ปริมาณการส่งออกข้าวและยางพาราด้วยวิธีปรับให้เรียบเอ็กซ์โปเนนเชียลของวินเตอร์ รูปแบบคูณ มีความแม่นยำมากที่สุด และการพยากรณ์ปริมาณการส่งออกมันสำปะหลังด้วยวิธีปรับให้เรียบเอ็กซ์โปเนนเชียลของวินเตอร์ รูปแบบบวก มีความแม่นยำมากที่สุด เนื่องจากให้ค่า MAPE และ MAD ต่ำที่สุด จึงนำแบบจำลองดังกล่าว ไปพยากรณ์ค่าในปี พ.ศ. 2564 เพื่อทดสอบค่าความแม่นยำของแบบจำลอง

4.3 ผลการทดสอบความแม่นยำของแบบจำลอง

จากการทดสอบความแม่นยำของแบบจำลองการพยากรณ์กร โดยใช้ข้อมูลปริมาณการส่งออกในปี พ.ศ. 2564 เป็นข้อมูลชุดทดสอบ เปรียบเทียบกับค่าพยากรณ์ ได้ผลการทดสอบ ดังนี้

ตารางที่ 3: แสดงผลการทดสอบค่าความแม่นยำของแบบจำลอง

สินค้าส่งออก	ค่า MAPE จากการทดสอบแบบจำลองการพยากรณ์
ข้าว	27.22

จากการทวนสอบค่าความแม่นยำของแบบจำลองการพยากรณ์ด้วยวิธีปรับให้เรียบเอ็กซ์โปเนนเชียลของวินเตอร์ (Winter's Method) รูปแบบคูณ พบว่า การพยากรณ์ปริมาณการส่งออกข้าวและยางพารา มีค่าร้อยละความคลาดเคลื่อนสมบูรณ์เฉลี่ย เท่ากับ 27.22 และ 9.48 ตามลำดับ และการพยากรณ์ปริมาณการส่งออกมันสำปะหลัง ด้วยวิธีปรับให้เรียบเอ็กซ์โปเนนเชียลของวินเตอร์ (Winter's Method) รูปแบบบวก มีค่าร้อยละความคลาดเคลื่อนสมบูรณ์เฉลี่ย เท่ากับ 16.75

4.4 การพัฒนาระบบธุรกิจอัจฉริยะเพื่อการแสดงผล

เมื่อได้ตัวแบบที่เหมาะสมเรียบร้อยแล้วผู้วิจัยจึงทำการพยากรณ์ปริมาณการส่งออกสินค้าทั้ง 3 ชนิด ในอีก 5 ปี ข้างหน้า และนำผลการพยากรณ์มาจัดทำระบบธุรกิจอัจฉริยะ ด้วยโปรแกรม Microsoft Power BI Desktop โดยมีตัวอย่างระบบ ดังภาพที่ 6 และภาพที่ 7



ภาพที่ 6: ตัวอย่างระบบธุรกิจอัจฉริยะ (หน้ารายงานการส่งออก)



ภาพที่ 7: ตัวอย่างระบบธุรกิจอัจฉริยะ (หน้ารายงานการพยากรณ์)

5. สรุปผลการวิเคราะห์และข้อเสนอแนะ

จากการดำเนินงานวิจัยโดยใช้ข้อมูลปริมาณการส่งออกสินค้าเกษตร 3 ชนิด ได้แก่ ข้าว ยางพารา และมันสำปะหลัง ตั้งแต่เดือนมกราคม พ.ศ. 2554 ถึงเดือนธันวาคม พ.ศ. 2564 จำนวน 132 เดือน สร้างตัวแบบพยากรณ์ด้วยเทคนิคการวิเคราะห์อนุกรมเวลา 3 วิธี ประกอบด้วย วิธีปรับเรียบ

เอ็กซ์โปเนนเชียลแบบสองครั้ง (Double Exponential Smoothing) วิธีปรับเรียบแบบเอ็กซ์โปเนนเชียลด้วยวิธีของวินเตอร์ (Winter's Method) และวิธีแยกส่วนประกอบ (Decomposition Method) เปรียบเทียบความแม่นยำของการพยากรณ์ด้วยค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) และค่าความผิดพลาดสัมบูรณ์เฉลี่ย (MAD) ที่ต่ำที่สุด ผลการทดสอบพบว่าการพยากรณ์ปริมาณการส่งออกข้าวและมันสำปะหลังด้วยวิธีปรับให้เรียบเอ็กซ์โปเนนเชียลของวินเตอร์ (Winter's Method) รูปแบบคูณ มีความแม่นยำมากที่สุด โดยมีค่า MAPE เท่ากับ 29.70 และ 9.14 ตามลำดับ ส่วนการพยากรณ์ปริมาณการส่งออกมันสำปะหลังวิธีปรับให้เรียบเอ็กซ์โปเนนเชียลของวินเตอร์ (Winter's Method) รูปแบบบวก ให้ความแม่นยำมากที่สุด โดยมีค่า MAPE เท่ากับ 21.26 สอดคล้องกับงานวิจัยของวารงคณา [7] วารงคณา [8] Boonlao [6] และปทุมชญา [5] ที่กล่าวไว้ว่าวิธีปรับให้เรียบเอ็กซ์โปเนนเชียลของวินเตอร์ (Winter's Method) เหมาะกับชุดข้อมูลที่มีการเคลื่อนไหว โดยมีแนวโน้มและมีความแปรผันตามฤดูกาล ซึ่งผลการวิจัยที่ได้ศึกษาในครั้งนี้สามารถนำไปใช้เป็นแนวทางวางแผนการเพาะปลูกหรือการเก็บเกี่ยวผลผลิต ใช้เป็นข้อมูลประกอบการวางแผนการผลิตและการส่งออกสำหรับผู้ประกอบการหรือหน่วยงานที่เกี่ยวข้องได้

อย่างไรก็ตามค่าความคลาดเคลื่อนที่ได้จากการวิเคราะห์ในการวิจัยครั้งนี้ ยังมีค่าสูงสืบเนื่องมาจากปริมาณการส่งออกสินค้าเกษตรของไทยในช่วงปีพ.ศ. 2562-2564 มีความเปลี่ยนแปลงจากในอดีตค่อนข้างมาก โดยมีสาเหตุจากการแพร่ระบาดของเชื้อไวรัส COVID-19 และนโยบายการบริหารประเทศเพื่อป้องกันการแพร่ระบาดของเชื้อไวรัส ที่มีการกำหนดมิให้มีการเข้า-ออกประเทศอยู่ช่วงหนึ่ง ทำให้ตลาดการส่งออกและนำเข้าได้รับผลกระทบและมีปริมาณลดลงด้วย ซึ่งส่งผลกระทบต่อพยากรณ์ด้วยเทคนิคอนุกรมเวลาเนื่องจากอาศัยเพียงค่าในอดีตมาพยากรณ์ค่าในอนาคตเพียงปัจจัยเดียว ในการวิจัยครั้งต่อไปจึงขอเสนอแนะให้ศึกษาปัจจัยอื่นๆ ร่วมด้วย เช่น ปัจจัยทางการผลิตสินค้าเกษตร ปัจจัยทางเศรษฐกิจ

และปัจจัยจากเหตุฉุกเฉินอื่นที่อาจเกิดขึ้น เช่นเดียวกับการแพร่ระบาดของโรคดังกล่าว เพื่อพัฒนาให้แบบจำลองการพยากรณ์มีความแม่นยำที่มากขึ้นต่อไป

### เอกสารอ้างอิง

- [1] สรุปผลสัมมนาการเกษตร พ.ศ. 2556. (2557). [ออนไลน์] จาก <https://www.ryt9.com/s/cabt/1820453>
- [2] พิเศษชูกิจ สินค้าสร้างรายได้ในครัวเรือนและประเทศ. (2564). [ออนไลน์]. จาก <https://www.arda.or.th/>
- [3] วารงคณา กิรติวิบูลย์. “ตัวแบบพยากรณ์มูลค่าการส่งออกข้าวหอมมะลิ” *วารสารวิทยาศาสตร์บูรพา*, ปีที่ 19, ฉบับที่ 1, หน้า 78-90.
- [4] สุรรัตน์ อนุเพ็ญ และคณะ. “การพยากรณ์มูลค่าการส่งออกน้ำยางพาราชั้น โดยเปรียบเทียบด้วยวิธีการทำให้เรียบเอ็กซ์โปเนนเชียลที่มีฤดูกาลอย่างง่าย และวิธีบอซ-เจนกินส์.” *การประชุมวิชาการระดับชาติราชภัฏหมู่บ้านจอมบึงวิจัย*, ครั้งที่ 9, หน้า 1211-1226.
- [5] ปทุมชญา กันธิยะ และชนพล เชนสิทธิเวชกุล. “แบบจำลองอนุกรมเวลาสำหรับการพยากรณ์จำนวนผู้ป่วยในสถานพยาบาล.” *การประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ*, ครั้งที่ 16, หน้า 431-436.
- [6] B. Chooptadit and S. Chaipitak. “Forecasting Model for Mango Export Volumes of Thailand.” *Naresuan University Journal: Science and Technology*, Vol. 26, No. 2, pp. 74-85, 2018.
- [7] วารงคณา เรียงสุทธิ. “ตัวแบบพยากรณ์มูลค่าการส่งออกไม้ยางพาราแปรรูปและเฟอร์นิเจอร์ของประเทศไทย.” *วารสารมหาวิทยาลัยนเรศวร วิทยาศาสตร์และเทคโนโลยี*, ปีที่ 24 ฉบับที่ 3, หน้า 108-122.
- [8] วารงคณา เรียงสุทธิ. “การพยากรณ์ปริมาณการส่งออกกาแฟสำเร็จรูป.” *วารสารหน่วยวิจัยวิทยาศาสตร์ เทคโนโลยี และสิ่งแวดล้อมเพื่อการเรียนรู้*, ปีที่ 11, ฉบับที่ 2, หน้า 238-251.
- [9] L. Ruekkasaem and M. Sasananan. “Forecasting Agricultural Products Prices using Time Series Methods for Crop Planning.” *International Journal of Mechanical Engineering and Technology (IJMET)*, July, 2018, Vol. 9, pp.957-971.
- [10] ศจี วานิช. “Data Mining.” (2558). [ออนไลน์]. จากเว็บไซต์ <http://sajeegm301.blogspot.com>

การพยากรณ์จำนวนงานจดทะเบียนของสำนักงานที่ดิน  
ด้วยเทคนิคการวิเคราะห์อนุกรมเวลา  
**Forecasting the Number Registration of the Land officer  
Using Time Series Techniques**

ชัชฎาวรรณ จันทร์ประเสริฐ (Chatchadawan Chanprasert)<sup>1</sup> และวัชรวิวัฒน์ จิตต์สกุล (Watchareewan Jitsakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858228@email.kmutnb.ac.th, <sup>2</sup>watchareewan.j@itd.kmutnb.ac.th

### บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อคัดเลือกแบบจำลองพยากรณ์ที่เหมาะสมที่สุดในการพยากรณ์จำนวนงานจดทะเบียนที่เกิดขึ้นในสำนักงาน โดยเปรียบเทียบแบบจำลองพยากรณ์จำนวน 3 วิธี ได้แก่ วิธีหาค่าเฉลี่ยเคลื่อนที่ วิธีการแยกส่วนประกอบ และวิธีการพยากรณ์แบบฤดูกาล ทำการสร้างแบบจำลองพยากรณ์ด้วยโปรแกรมวิเคราะห์ผลทางสถิติ Minitab วัดประสิทธิภาพแบบจำลองพยากรณ์ด้วยค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่าเบี่ยงเบนสัมบูรณ์เฉลี่ยที่ต่ำที่สุด ข้อมูลที่ใช้เป็นข้อมูลจำนวนงานจดทะเบียนในสำนักงานที่ดินระหว่างเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2563 รวม 72 เดือน และแบ่งข้อมูลเป็น 2 ชุด คือ ชุดข้อมูลฝึกฝน (Train Set) จำนวน 60 เดือน และชุดข้อมูลทดสอบ (Test Set) จำนวน 12 เดือน ผลการวิจัยพบว่า แบบจำลองการพยากรณ์ด้วยวิธีแยกส่วนประกอบให้ประสิทธิภาพการพยากรณ์สูงสุด โดยมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเท่ากับ ร้อยละ 8 และค่าเบี่ยงเบนสัมบูรณ์เฉลี่ยเท่ากับ 257 ซึ่งได้นำผลการพยากรณ์วิธีที่ดีที่สุดมานำเสนอบน Power BI สำหรับการวางแผนการให้บริการของสำนักงานที่ดินได้อย่างมีประสิทธิภาพ

**คำสำคัญ:** เทคนิคการวิเคราะห์อนุกรมเวลา วิธีหาค่าเฉลี่ยเคลื่อนที่ วิธีการพยากรณ์แบบฤดูกาล วิธีการแยกส่วนประกอบ

### Abstract

The objective of this research is to select the most suitable model for Forecasting the Number Registration of the Land officer using Time Series Techniques. Time series analysis had been applied to compare performance among 3 statistical models as following; 1) Moving Average Model 2) Winter's Method and 3) Decomposition Model. By using Minitab, Also perform model performance measurements in forecasting. The mean absolute percentage error (MAPE) and the mean absolute deviation (MAD) were the lowest. Dataset for creating forecasting Model is collected information of actual Registration of the Land officer since January 2015 to December 2020. The 72 months were used and divided into 2 sets. The first set had 60 months for identification models. The second set had 12 months for checking the accuracy of the forecasting model. The result showed that the best forecasting model was Decomposition Model due to its MAPE value of 8 and MAD value of 257. And It brings the best forecasting results to Power BI for effective land office service planning.

**Keywords:** Time Series Analysis, Moving Average, Winter's Method, Decomposition.



## 1. บทนำ

สำนักงานที่ดินมีหน้าที่ให้บริการในด้านการจดทะเบียนสิทธิและนิติกรรมเกี่ยวกับอสังหาริมทรัพย์ตามประมวลกฎหมายแพ่งและพาณิชย์ เช่น การจดทะเบียนขาย จำนอง โฉนด เป็นต้น การให้บริการจดทะเบียนของสำนักงานที่ดินมีจำนวนปริมาณงานที่ต้องจดทะเบียนเพิ่มขึ้นทุกปี ส่งผลให้การบริการของสำนักงานที่ดินมีความล่าช้า เนื่องจากปริมาณเจ้าหน้าที่ให้บริการมีไม่เพียงพอกับปริมาณงานในแต่ละวัน ที่ผ่านมการบริหารวางแผนจำนวนเจ้าหน้าที่ให้บริการในสำนักงานที่ดินนั้นเป็นการอาศัยจากประสบการณ์ในอดีตในการคาดเดาจำนวนงานจดทะเบียน ซึ่งเป็นการพยากรณ์แบบไม่มีรูปแบบทำให้มีความคลาดเคลื่อนของการพยากรณ์ได้สูง ดังนั้นเพื่อให้การบริการของสำนักงานที่ดินมีความรวดเร็ว มีประสิทธิภาพและสร้างความพึงพอใจกับผู้รับบริการ จึงจำเป็นต้องมีการพยากรณ์ปริมาณงานจดทะเบียนเพื่อบริหารจัดการจำนวนเจ้าหน้าที่ประจำสำนักงานให้เหมาะสมและเพียงพอต่อปริมาณงานจดทะเบียน

การพยากรณ์ด้วยเทคนิคอนุกรมเวลาเป็นการรวบรวมข้อมูลตามช่วงเวลาแบบสัปดาห์ เดือน หรือปี ที่ต่อเนื่องกันมาทำการวิเคราะห์แนวโน้มของปริมาณข้อมูลที่จะเกิดขึ้นในอนาคต โดยวิธีการพยากรณ์ของอนุกรมเวลามีหลายวิธี เช่น วิธีค่าเฉลี่ยเคลื่อนที่ วิธีปรับให้เรียบเอ็กซ์โปเนนเชียลแบบโฮลท์และวินเทอร์ และวิธี SARIMA ในการพยากรณ์มูลค่าการส่งออกรถยนต์ อุปกรณ์และส่วนประกอบของประเทศไทย [1] การพยากรณ์มูลค่าการส่งออกยางพาราของประเทศไทย ด้วยวิธีค่าเฉลี่ยเคลื่อนที่ วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลซ้ำสองครั้ง และวิธีบ็อกซ์-เจนกินส์ [2] การพยากรณ์ราคามังคุดละใช้วิธีพยากรณ์ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการแยกส่วนประกอบอนุกรมเวลา และวิธีปรับให้เรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบคูณ [3] การสร้างแบบจำลองการพยากรณ์ความต้องการแวนตา กรณีศึกษา ร้านรักแวนที่ใช้วิธีหาค่าเฉลี่ยเคลื่อนที่ วิธีปรับให้เรียบเอ็กซ์โปเนนเชียลแบบง่าย และวิธีแยกส่วนประกอบ [4] หรือวิธีบ็อกซ์-เจนกินส์ และวิธีการพยากรณ์อนุกรมเวลาโดยวิธีการของ

วินเทอร์ ที่ใช้ในการพยากรณ์ทำนายอุบัติเหตุการเกิดโรคไข้เลือดออกของประชาชน [5] เป็นต้น

จากปัญหา และงานวิจัยดังกล่าว ผู้วิจัยจึงมีแนวคิดที่จะศึกษาทฤษฎีการพยากรณ์ด้วยเทคนิคอนุกรมเวลา และคัดเลือกวิธีที่เหมาะสมที่สุด มาใช้ในการพยากรณ์จำนวนงานจดทะเบียนที่จะเกิดขึ้นในสำนักงานที่ดิน เพื่อนำผลพยากรณ์มาบริหารจัดการจำนวนเจ้าหน้าที่ประจำสำนักงานที่ดิน เพื่อให้การบริการจดทะเบียนของสำนักงานที่ดินนั้นมีความรวดเร็ว และสร้างความพึงพอใจให้แก่ผู้เข้ารับบริการ

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้รวบรวมข้อมูลทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยแบ่งเป็น 5 หัวข้อหลัก ดังนี้

### 2.1 การพยากรณ์แบบอนุกรมเวลา

การพยากรณ์แบบอนุกรมเวลา เป็นเทคนิคที่นำข้อมูลในอดีตที่เก็บรวบรวมเป็นรายวัน รายสัปดาห์ รายเดือน รายไตรมาส หรือรายปีอย่างต่อเนื่อง นำข้อมูลมาทำการวิเคราะห์เพื่อหารูปแบบการเปลี่ยนแปลงของข้อมูลเพื่อการพยากรณ์การเปลี่ยนแปลงข้อมูลนั้นในอนาคต การพยากรณ์แบบอนุกรมเวลานั้นสามารถทำได้หลายวิธี เช่น วิธีค่าเฉลี่ยเคลื่อนที่ วิธีพยากรณ์แบบฤดูกาล วิธีแยกส่วนประกอบ เป็นต้น [6]

### 2.2 วิธีค่าเฉลี่ยเคลื่อนที่ (Moving Average)

วิธีค่าเฉลี่ยเคลื่อนที่ที่เหมาะสมสำหรับการพยากรณ์ในช่วงระยะเวลาสั้น ๆ และข้อมูลที่มีลักษณะค่อนข้างแน่นอนและเป็นเส้นตรงและคงที่ตามแนวโน้ม ไม่เหมาะสมกับข้อมูลลักษณะเป็นแนวโน้ม ฤดูกาล หรือข้อมูลที่มีลักษณะเป็นการเปลี่ยนแปลงเป็นขึ้นบันได เทคนิคนี้ใช้หลักการในการหาค่าเฉลี่ย คือนำค่าจากการสังเกตหรือข้อมูลในอดีตมาคำนวณหาค่าเฉลี่ยที่ได้นี้เป็นค่าพยากรณ์สำหรับช่วงเวลาที่ถัดไป โดยมีสมการดังนี้ [7]

$$F_{t+1} = \frac{X_t + X_{t+1} + \dots + X_{t+n+1}}{n+1} \quad (1)$$

โดย  $F_t$  = ค่าพยากรณ์ในงวดที่  $t$

$X_t$  = ข้อมูลในงวดที่  $t$

$n$  = จำนวนงวดที่มาทำการหาค่าเฉลี่ย

**2.3 วิธีการพยากรณ์แบบฤดูกาล (Winter’s Method)**

เป็นวิธีการพยากรณ์ที่เหมาะสมกับข้อมูลที่มีลักษณะรูปแบบแนวโน้มอย่างมีทิศทาง และเป็นฤดูกาล ซึ่งการพยากรณ์วิธีวินเทอร์จะต้องใช้ข้อมูลอย่างน้อย 2 ฤดูกาล และมีรูปแบบที่ประกอบด้วย 3 ส่วน คือ ส่วนของการปรับเรียบ ส่วนของแนวโน้ม และส่วนของฤดูกาล โดยมีสมการดังนี้ [7]

สำหรับการปรับเรียบ

$$S_t = \alpha \frac{X_t}{I_{t-L}} (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (2)$$

สำหรับการทำให้เป็นแนวโน้ม

$$b_t = \gamma (S_t - S_{t-1}) + (1 - \gamma) b_{t-1} \quad (3)$$

สำหรับทำให้เป็นฤดูกาล

$$I_t = \beta \frac{X_t}{S_t} + (1 - \beta) I_{t-L} \quad (4)$$

สมการการพยากรณ์คือ

$$F_{t+m} = (S_t - b_t m) I_{t-L+m} \quad (5)$$

โดยที่ L = ช่วงของฤดูกาล

S<sub>t</sub> = ค่าการปรับเรียบ

b<sub>t</sub> = ค่าการประมาณแนวโน้ม

I<sub>t</sub> = ค่าการประมาณฤดูกาล

α = ค่าคงที่สำหรับการปรับเรียบ (0 < α < 1)

γ = ค่าคงที่สำหรับแนวโน้ม (0 < γ < 1)

β = ค่าคงที่สำหรับฤดูกาล (0 < β < 1)

X<sub>t</sub> = ค่าสังเกตที่เวลา t

m = ช่วงเวลาการพยากรณ์ในอนาคต

**2.4 วิธีแยกส่วนประกอบ (Decomposition)**

เป็นวิธีที่แยกส่วนประกอบต่าง ๆ ของอนุกรมเวลาออกจากกัน โดยที่จะทำให้พบการเคลื่อนไหวของอนุกรมเวลา และนำไปสร้างตัวแบบพยากรณ์ได้ ส่วนประกอบของอนุกรมเวลาได้แก่ แนวโน้ม ความผันแปรของฤดูกาล ความผันแปรตามวัฏจักร และส่วนประกอบที่ผิดปกติหรือความคลาดเคลื่อน ซึ่งวิธีการแยกส่วนประกอบสามารถแบ่งเป็น 2 แบบ ได้แก่ แบบบวก เหมาะกับลักษณะอนุกรมเวลาที่ความผันแปรของฤดูกาลคงที่ และแบบคูณ เหมาะกับลักษณะอนุกรมเวลาที่มีความผันแปรของฤดูกาลไม่คงที่ สำหรับวิธีนี้ได้ใช้วิธีแยกส่วนประกอบแบบคูณ โดยมีสมการดังนี้ [8]

$$F_t = T_t \times S_t \quad (6)$$

โดย F<sub>t</sub> = ค่าพยากรณ์ของช่วงเวลาปัจจุบัน

Y<sub>t</sub> = ค่าของข้อมูลจริงในช่วงเวลาปัจจุบัน

T<sub>t</sub> = ค่าประมาณของแนวโน้ม

S<sub>t</sub> = ค่าประมาณของการผันแปรตามฤดูกาล

C<sub>t</sub> = ค่าประมาณของการแปรผันตามวัฏจักร

I<sub>t</sub> = ค่าประมาณของความไม่แน่นอน

N = จำนวนข้อมูล

**2.5 งานวิจัยที่เกี่ยวข้อง**

สมฤดี กัญญา และอรุวรรณ [9] ได้ทำการศึกษการพยากรณ์มูลค่าการส่งออกอัญมณีและเครื่องประดับของประเทศไทย ด้วยการวิเคราะห์อนุกรมเวลา 3 วิธี ได้แก่ วิธี Decomposition วิธี Exponential Smoothing และวิธีบ็อกซ์ เจนกินส์ ข้อมูลที่ใช้มาจากข้อมูลการส่งออกอัญมณี และเครื่องประดับของประเทศไทยรายเดือน ตั้งแต่เดือนมกราคม พ.ศ. 2556 ถึงเดือนธันวาคม พ.ศ. 2562 จำนวน 84 วัดประสิทธิภาพการพยากรณ์ด้วยค่า MAPE ผลการวิจัยพบว่า วิธี Decomposition ให้ประสิทธิภาพสูงที่สุด ค่า MAPE เท่ากับ 0.24 รองลงมาคือ Exponential Smoothing ค่า MAPE เท่ากับ 0.26 MAPE เท่ากับ 0.47 ตามลำดับ

พรฤดี และณัฐทวี [10] ได้ทำการศึกษาตัวแบบที่เหมาะสมสำหรับการพยากรณ์ยอดขายเสื้อผ้าด้วยการวิเคราะห์อนุกรมเวลา 2 วิธี ได้แก่ วิธี Decomposition และวิธี Linear Regression ข้อมูลที่ใช้มาจากยอดขายรายวัน ตั้งแต่วันที่ 1 มกราคม 2015 ถึงวันที่ 31 ธันวาคม 2017 จำนวน 1,096 ค่า วัดประสิทธิภาพการพยากรณ์ด้วยค่า Basic Error ค่า MAD ค่า MSE ค่า MAPE และค่าTheil’s U ผลการวิจัยพบว่า วิธี Linear Regression ให้ประสิทธิภาพสูงสุด ค่า Basic Error เท่ากับ 24.441 ค่า MAD เท่ากับ 0.380 ค่า MSE เท่ากับ 0.266 ค่า MAPE เท่ากับ 7.850 และค่า Theil’s U เท่ากับ 0.5555

Lima and et al. [11] ได้ทำการศึกษการพยากรณ์ข้อมูลเศรษฐกิจทางด้านการขายปลีกทางอิเล็กทรอนิกส์ของโปรตุเกส ด้วยการวิเคราะห์อนุกรมเวลา 2 วิธี ได้แก่ วิธี Holt-Winters Models Additive และวิธี Holt-Winters Models Multiplicative ข้อมูลที่ใช้มาจากข้อมูลจำนวน

รายการขายสินค้าทางอีคอมเมิร์ซ จาก Eurostat Platform ตั้งแต่เดือนมกราคม 2000 ถึงเดือนกุมภาพันธ์ 2018 รวม 218 เดือน วัดประสิทธิภาพการพยากรณ์ด้วยค่า MSE ผลการวิจัยพบว่า วิธี Holt-Winters Models Multiplicative ให้ประสิทธิภาพสูงสุด ค่า MSE เท่ากับ 420.3238 รองลงมา คือ วิธี Holt-Winters Models Additive ค่า MSE เท่ากับ 452.3238

### 3. วิธีการดำเนินการวิจัย

งานวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อศึกษาแบบจำลองการพยากรณ์ที่เหมาะสมสำหรับการพยากรณ์จำนวนงานจดทะเบียนในสำนักงานที่ดิน โดยใช้เทคนิคการวิเคราะห์อนุกรมเวลา โดยมีรายละเอียดและขั้นตอนการดำเนินงานดังนี้

#### 3.1 การรวบรวมและการจัดการข้อมูล

เก็บรวบรวมข้อมูลจำนวนผู้เข้ารับบริการในสำนักงานที่ดินจังหวัดนนทบุรี สาขาปากเกร็ด ในช่วงตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึง เดือนธันวาคม พ.ศ. 2563 รวม 72 เดือน ทำการตรวจสอบข้อมูลและจัดการข้อมูลด้วยโปรแกรม Microsoft Excel แบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลฝึกฝน (Train Set) เป็นข้อมูลของจำนวนงานจดทะเบียนในสำนักงานที่ดินตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2562 รวม 60 เดือน เพื่อใช้ในการสร้างแบบจำลองและเปรียบเทียบประสิทธิภาพในการพยากรณ์ของแบบจำลอง และชุดข้อมูลทดสอบ (Test Set) เป็นข้อมูลจำนวนงานจดทะเบียนในสำนักงานที่ดินตั้งแต่เดือนมกราคม พ.ศ. 2563 ถึงเดือนธันวาคม พ.ศ. 2563 รวม 12 เดือน เพื่อใช้เป็นข้อมูลสำหรับทดสอบความถูกต้องของการพยากรณ์

#### 3.2 การสร้างแบบจำลองพยากรณ์

สร้างแบบจำลองการพยากรณ์ด้วยโปรแกรมวิเคราะห์ผลทางสถิติ Minitab โดยใช้เทคนิคการวิเคราะห์อนุกรมเวลา จำนวน 3 วิธี ได้แก่ วิธี Moving Average โดยทำการกำหนดค่า MA Length 3, 6, 9, 12 ตามลำดับ วิธี Winter's Method โดยทำการกำหนดค่า  $\alpha$  (Level) ค่า  $\gamma$  (Trend) และ ค่า  $\beta$  (Seasonal) เท่ากับ 0.1, 0.2, 0.3, 0.4, 0.5 และ

วิธี Decomposition โดยกำหนดค่า Seasonal Length 3, 6, 9, 12 ทำการเก็บข้อมูลและคัดเลือกแบบจำลองในการพยากรณ์ที่ดีที่สุด

#### 3.3 การคัดเลือกแบบจำลองในการพยากรณ์

พิจารณาประสิทธิภาพของแบบจำลองพยากรณ์ทั้ง 3 วิธี ด้วยค่าเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAPE) และค่าเบี่ยงเบนสมบูรณ์เฉลี่ย (MAD) ที่ต่ำที่สุด

#### 3.4 การทดสอบความถูกต้องของแบบจำลอง

ทดสอบความถูกต้องของแบบจำลองพยากรณ์นำข้อมูลการพยากรณ์ที่ได้จากแบบจำลองมาทดสอบความถูกต้องและแม่นยำกับข้อมูลชุดทดสอบ (Test Set) จำนวน 12 เดือน พิจารณาค่าความคลาดเคลื่อนในการพยากรณ์ ด้วยค่า MAPE

#### 3.5 นำเสนอข้อมูลการพยากรณ์

นำเสนอข้อมูลการพยากรณ์งานจดทะเบียนในสำนักงานที่ดินในรูปแบบ Dashboard ด้วยโปรแกรม Power BI

## 4. ผลการดำเนินงาน

ผลการดำเนินงานวิจัยแบ่งออกเป็น 4 หัวข้อ แสดงรายละเอียดได้ดังนี้

#### 4.1 ผลการดำเนินงานสร้างแบบจำลองพยากรณ์

ผลการสร้างแบบจำลองพยากรณ์งานจดทะเบียนโดยเทคนิคการวิเคราะห์แบบอนุกรมเวลา จำนวน 3 วิธี ได้แก่ วิธี Moving Average วิธี Winter's Method และวิธี Decomposition วิเคราะห์ค่าสถิติด้วยโปรแกรม Minitab แสดงรายละเอียดดังตารางที่ 1 ถึง 3

ตารางที่ 1: การวิเคราะห์อนุกรมเวลา วิธี Moving Average

วิธีพยากรณ์	พารามิเตอร์	MAPE	MAD
Moving Average	<b>MA Length 3</b>	<b>10</b>	<b>327</b>
	MA Length 6	11	336
	MA Length 9	11	354
	MA Length 12	12	374

จากตารางที่ 1 เมื่อพิจารณาค่าความคลาดเคลื่อน พบว่าพารามิเตอร์ MA Length 3 ให้ประสิทธิภาพดีที่สุด โดยค่า MAPE เท่ากับ 10 และค่า MAD เท่ากับ 327

ตารางที่ 2: การวิเคราะห์หอนุกรมเวลา วิธี Winter's Method

วิธีพยากรณ์	พารามิเตอร์	MAPE	MAD
Winter's Method	$\alpha = 0.5, \gamma = 0.1, \beta = 0.1$	<b>10</b>	<b>315</b>
	$\alpha = 0.5, \gamma = 0.2, \beta = 0.1$	10	326
	$\alpha = 0.5, \gamma = 0.3, \beta = 0.1$	11	333

จากตารางที่ 2 เมื่อพิจารณาค่าความคลาดเคลื่อนพบว่า พารามิเตอร์  $\alpha = 0.5, \gamma = 0.1, \beta = 0.1$  ให้ประสิทธิภาพที่ดีที่สุด โดยค่าค่า MAPE เท่ากับ 10 และค่า MAD เท่ากับ 315

ตารางที่ 3 การวิเคราะห์หอนุกรมเวลา วิธี Decomposition

วิธีพยากรณ์	พารามิเตอร์	MAPE	MAD
Decomposition	Seasonal Length 3	10	304
	Seasonal Length 6	9	302
	Seasonal Length 9	9	282
	<b>Seasonal Length 12</b>	<b>8</b>	<b>257</b>

จากตารางที่ 3 เมื่อพิจารณาค่าความคลาดเคลื่อนพบว่า พารามิเตอร์ Seasonal Length 12 ให้ประสิทธิภาพที่ดีที่สุด โดยมีค่า MAPE เท่ากับ 8 และค่า MAD เท่ากับ 257

#### 4.2 การเปรียบเทียบและคัดเลือกแบบจำลอง

การเปรียบเทียบและคัดเลือกแบบจำลอง นำค่าความคลาดเคลื่อนของตัวแบบการพยากรณ์ ได้แก่ ค่า MAPE และค่า MAD ของวิธี Moving Average วิธี Winter's Method และวิธี Decomposition มาเปรียบเทียบเพื่อหาแบบจำลองพยากรณ์ที่เหมาะสมกับข้อมูลจำนวนงานจดทะเบียนสำนักงานที่ดิน โดยพิจารณาจากวิธีพยากรณ์ที่ให้ค่าความคลาดเคลื่อนของแบบจำลองพยากรณ์ต่ำที่สุดพบว่า วิธี Decomposition เป็นวิธีที่ค่าความคลาดเคลื่อนต่ำที่สุด ซึ่งเหมาะสมที่จะใช้ในการพยากรณ์จำนวนงานจดทะเบียนในสำนักงานที่ดิน รายละเอียดแสดงดังตารางที่ 4

ตารางที่ 4: การเปรียบเทียบความแม่นยำของแบบจำลองพยากรณ์

วิธีพยากรณ์	พารามิเตอร์	MAPE	MAD
Moving Average	MA Length 3	10	327
Winter's Method	$\alpha = 0.5, \gamma = 0.1, \beta = 0.1$	10	315
<b>Decomposition</b>	<b>Seasonal Length 12</b>	<b>8</b>	<b>257</b>

จากตารางที่ 4 เมื่อพิจารณาค่าความคลาดเคลื่อนพบว่า วิธี Decomposition ให้ประสิทธิภาพที่ดีที่สุด โดยมีค่า MAPE เท่ากับ 8 และค่า MAD เท่ากับ 257

#### 4.3 การทดสอบความถูกต้องของแบบจำลอง

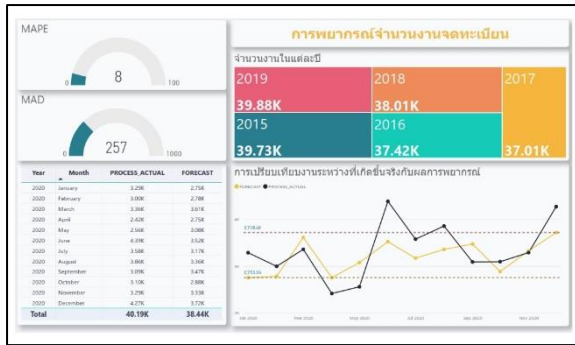
ผลการทดสอบความถูกต้องของแบบจำลอง เมื่อพิจารณาค่าความคลาดเคลื่อน วิธี Decomposition ให้ประสิทธิภาพที่ดีที่สุด รองลงมา วิธี Winter's Method และวิธี Moving Average ตามลำดับ นำแบบจำลองที่ได้มาทำการทดสอบความถูกต้องกับข้อมูลจำนวนงานจดทะเบียนในสำนักงานที่ดิน จำนวน 12 เดือน ซึ่งพบว่าแบบจำลองพยากรณ์วิธี Decomposition มีค่าความคลาดเคลื่อนต่ำที่สุดที่ร้อยละ 11.87 รายละเอียดแสดงดังตารางที่ 5

ตารางที่ 5: ผลการทดสอบความถูกต้องของแบบจำลอง

เดือน	จำนวนงาน	Decomposition		
		Test Set	Forecast	Error
JAN	3288	2753.06	534.94	16.27
FEB	2996	2783.91	212.09	7.1
MAR	3360	3613.01	-253.01	7.5
APR	2415	2753.92	-338.92	14.0
MAY	2559	3076.2	-517.2	20.2
JUN	4386	3523.64	862.36	19.7
JUL	3577	3171.71	405.29	11.3
AUG	3858	3358.15	499.85	13.0
SEP	3090	3474.7	-384.7	12.4
OCT	3095	2884.41	210.59	6.8
NOV	3290	3327.9	-37.9	1.2
DEC	4274	3718.4	555.6	13.0
ร้อยละ				<b>11.87</b>

#### 4.4 การนำเสนอข้อมูล

ผู้วิจัยนำเสนอผลการพยากรณ์ โดยใช้วิธี Decomposition เป็นตัวแบบในการพยากรณ์จำนวนงานจดทะเบียนในอนาคต โดยใช้โปรแกรม Power BI Desktop เป็นเครื่องมือช่วยในการแสดงผลพยากรณ์ในรูปแบบของรายงาน แสดงดังภาพที่ 1



ภาพที่ 1: แผนภูมิภาพแสดงจำนวนงานจดทะเบียน

จากภาพที่ 1 แสดงจำนวนงานจดทะเบียนในสำนักงานที่ดินระยะเวลา 5 ปี ตั้งแต่ พ.ศ. 2558 – พ.ศ. 2562 และการเปรียบเทียบข้อมูลจำนวนงานจดทะเบียนที่เกิดขึ้นจริงจำนวน 12 เดือน ตั้งแต่เดือนมกราคม ถึง เดือนธันวาคม พ.ศ. 2563 กับจำนวนงานจดทะเบียนที่มาจาก การสร้างแบบจำลองพยากรณ์ด้วยวิธี Decomposition

5. สรุป

งานวิจัยฉบับนี้ผู้วิจัยได้ศึกษาการพยากรณ์ด้วยเทคนิคอนุกรมเวลาเพื่อให้ได้วิธีการพยากรณ์ที่ดีและเหมาะสมที่สุดสำหรับการพยากรณ์จำนวนงานจดทะเบียนในสำนักงานที่ดิน ผู้วิจัยได้คัดเลือกวิธีการพยากรณ์ที่ได้รับความนิยม จำนวน 3 วิธี ได้แก่ วิธี Moving Average วิธี Winter’s Method และวิธี Decomposition ทดสอบประสิทธิภาพของแบบจำลองการพยากรณ์ด้วยค่า MAPE และค่า MAD ผลการวิจัยพบว่า วิธี Decomposition มีความเหมาะสมที่สุด ค่า MAPE เท่ากับ 8 ค่า MAD เท่ากับ 257 หลังจากนั้นนำข้อมูลมาทวนสอบกับข้อมูลจริงจำนวน 12 เดือน สรุปได้ว่า การพยากรณ์จำนวนงานจดทะเบียนในสำนักงานที่ดินโดยสร้างแบบจำลองการพยากรณ์ด้วยวิธี Decomposition มีความแม่นยำ โดยค่าความคลาดเคลื่อนในการพยากรณ์อยู่ที่ร้อยละ 11.87

เอกสารอ้างอิง

[1] อรุษา จันทรภา และ นันทชัย กานตานันท์, “การพยากรณ์มูลค่าการส่งออกรถยนต์ อุปกรณ์และส่วนประกอบของประเทศไทย ด้วยวิธีอนุกรมเวลา” *Thai Journal of*

*Operation Research: TJOR*, Vol. 15, No. 2 (July – December 2021) หน้า 12-20.

[2] จารุวรรณ สิงห์ม่วง และ ธิดาพร ศุภกากร, “ตัวแบบพยากรณ์มูลค่าการส่งออกยางพาราของประเทศไทย” *วารสารวิชาการวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครสวรรค์* ปีที่ 12 ฉบับที่ 15 มกราคม-มิถุนายน พ.ศ. 2563 หน้า 58-82.

[3] วรางคณา เรียนสุทธิ, “ตัวแบบพยากรณ์ราคามังคุดละ” *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี* ปีที่ 19 ฉบับที่ 2 พฤษภาคม-สิงหาคม พ.ศ. 2560 หน้า 31-42.

[4] อนุสรณ์ บุญสง่า, การพยากรณ์ความต้องการวันตากรณีศึกษา ร้านรักแฉ่น. การค้นคว้าอิสระวิทยาศาสตร์มหบัณฑิต สาขาการจัดการ โลจิสติกส์และโซ่อุปทาน วิทยาลัยโลจิสติกส์และซัพพลายเชน มหาวิทยาลัยศรีปทุม.

[5] วณิกเกียรติ อึ้งพันธ์ บัววรรณ ศรีชัยกุล จตุพร เหลืองอุบล และ พุทธิไกร ประมวล, “ตัวแบบทำนายอุบัติการณ์การเกิดโรคไข้เลือดออกของประชาชน จังหวัดอุบลราชธานี” *วารสารวิจัยสาธารณสุขศาสตร์ มหาวิทยาลัยราชภัฏอุบลราชธานี* ปีที่ 9 ฉบับที่ 2 กรกฎาคม-ธันวาคม พ.ศ. 2563 หน้า 174-184.

[6] นิติชัย รุ่งจินดารัตน์ และ ศรีณย์ ทัศนิกานเวสร์, “การพยากรณ์อนุกรมเวลาด้วยวิธีแบบคลาสสิก การส่งออกข้าวหอมมะลิของไทย” *วารสารวิทยาลัยดุสิตธานี* ปีที่ 13 ฉบับที่ 2 พฤษภาคม-สิงหาคม พ.ศ. 2562 หน้า 283-293.

[7] อภิชัย พรหมอ่อน, การศึกษาการพยากรณ์แบบอนุกรมเวลา (Time Series) เพื่อการวางแผนการสั่งซื้อวัตถุดิบ กรณีศึกษาบริษัทผลิตชิ้นส่วนท่อขยายรถยนต์. สารนิพนธ์ปริญญาบริหารธุรกิจมหาบัณฑิต บัณฑิตวิทยาลัย สาขาวิชาการจัดการอุตสาหกรรม สถาบันเทคโนโลยีไทย-ญี่ปุ่น.

[8] สุกัญญา ร่วมประเสริฐ, “การพยากรณ์แนวโน้มมูลค่าการส่งออกน้ำสับประรดของไทย” *วารสารวิชาการบริหารธุรกิจ* ปีที่ 2 ฉบับที่ 1 มกราคม – มิถุนายน 2556 หน้า 16-24.

[9] สมฤดี พงษ์เสนา กัญญา บวรโชคชัย และ อรรธรณ รุ่งทอง “ตัวแบบการพยากรณ์มูลค่าการส่งออกอัญมณีและเครื่องประดับของประเทศไทย” *Journal of Business Management, Burapha University*, Vol.9 No. 2 (July – December 2021) หน้า 66-85.

[10] พรฤดี เนติโสภาคกุล และ ณัฐวิษ สุภษา, “การวิเคราะห์ตัวแบบที่เหมาะสมสำหรับการพยากรณ์ยอดขายเสื้อผ้า” *วารสารบริหารธุรกิจเทคโนโลยีมหานคร* ปีที่ 16 ฉบับที่ 2 (กรกฎาคม-ธันวาคม) พ.ศ. 2562 หน้า 77-96.

[11] Lima, S., Goncalves, M. and Costa, M, “Time Series Forecasting Using Holt-Winter Exponential Smoothing : An Application to Economic Data,” *In AIP Conference Proceedings* 10 December 2019. (1-4).

# การพยากรณ์หุ้นเติบโตจากการจัดกลุ่มหลักทรัพย์เอ็มเอไอของตลาดหลักทรัพย์ แห่งประเทศไทย

## Prediction mai Growth stock in The Securities Exchange of Thailand

สุทธิดา ช่างปั้น (Sudthida Changpan)<sup>1</sup> และธนพล เจนสุทธิเวชกุล(Tanapon Jensuttiwetchakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>1s6307021856152@email.kmutnb.ac.th, <sup>2</sup>tanapon.j@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอการพัฒนาแบบจำลองการพยากรณ์หุ้นเติบโตจากการจัดกลุ่มหลักทรัพย์เอ็มเอไอ (mai) ของตลาดหลักทรัพย์แห่งประเทศไทยเพื่อช่วยในการประกอบการตัดสินใจในการลงทุนหลักทรัพย์จึงมีความสำคัญอย่างมากกับนักลงทุนในหุ้น โดยข้อมูลที่ใช้ในการจำแนกข้อมูลเป็นข้อมูลงบการเงิน และชื่อหลักทรัพย์ จำนวน 1,101 รายการ ระหว่างปี พ.ศ.2555 -2559 นำมาทำการวิเคราะห์ข้อมูลเพื่อทดลองแบบจำลองเมื่อผ่านขั้นตอนการเตรียมข้อมูลเหลือจำนวน 138 รายการ พบว่าข้อมูลในแต่ละคลาสมีจำนวนที่ไม่สมดุลกัน จึงใช้เทคนิค SMOTE Upsampling ซึ่งช่วยในการแก้ไข Unbalanced data ในการวิเคราะห์ข้อมูลและหาอัลกอริทึมผู้วิจัยใช้โปรแกรม Rapid Miner ในการวิเคราะห์และทดลองอัลกอริทึม โดยผู้วิจัยได้ใช้ทั้งหมด 4 อัลกอริทึม ได้แก่ Decision Tree, Random Forest, K-nearest neighbor และ Naïve Bayes เพื่อเปรียบเทียบหาความแม่นยำในการพยากรณ์การเติบโตของหลักทรัพย์เอ็มเอไอ โดยผู้วิจัยได้ใช้ Optimize Parameter (Grid) มาช่วยในการหาค่าพารามิเตอร์ที่เหมาะสมที่สุดทั้ง 4 อัลกอริทึม ซึ่งผลการทดลองพบว่าอัลกอริทึม Random Forest มีประสิทธิภาพและมีความถูกต้องแม่นยำมากที่สุดสำหรับชุดข้อมูลนี้ โดยมีค่าความถูกต้อง (Accuracy) เท่ากับ 90.42% ค่าความแม่นยำ (Precision) เท่ากับ 87.29% และค่าระลึก (Recall) เท่ากับ 96.33% จากผลการวิจัยสรุปได้ว่าแบบจำลองสำหรับการพยากรณ์หุ้นเติบโตจากการจัดกลุ่มหลักทรัพย์เอ็มเอไอ ของตลาดหลักทรัพย์แห่งประเทศไทย

ไทยที่ทดลองโดยใช้อัลกอริทึม Random Forest นั้นมีประสิทธิภาพและความแม่นยำอยู่ในเกณฑ์ดี

**คำสำคัญ:** หุ้นเติบโต, หลักทรัพย์, อัลกอริทึม Random Forest, ลดความไม่สมดุลของข้อมูล, การหาค่าพารามิเตอร์ที่เหมาะสม

### Abstract

The objective of this research is to develop model for predicting growth stocks based on their financial statements. Financial data of stocks in MAI between 2012 and 2016 that used as data of the model to clean data stage which is reduce and prepare raw data from 1,101 to 138 rows. From the data preparation found the amount of data in each class was not balanced. Therefore, the method of balancing the data that can solve the Unbalanced data is the SMOTE Upsampling technique. Rapid Miner software with four different algorithms - Decision Tree, Random Forest, K-nearest neighbors, and Naive Bayes - were used to develop comparative models. To compare the accuracy of forecasting the growth of MAI securities used Optimize Parameter (Grid). The results indicated that Random Forrest have the highest efficiency and accuracy. It has accuracy at 90.42%, Precision at 87.29% and Recall at 96.33%.

**Keyword:** Growth Stock, Securities, Random Forest Algorithm, SMOTE Upsampling, Optimize Parameters (Grid)

## 1. บทนำ

ปัจจุบันมีผู้คนเป็นจำนวนมากที่สนใจในเรื่องการลงทุนและการซื้อขายหลักทรัพย์เป็นหนึ่งในช่องทางการลงทุนต่อออกไปสู่ความมั่งคั่งและป้องกันความเสี่ยงในด้านการเงินที่ได้รับการยอมรับในวงกว้าง การลงทุนในหลักทรัพย์สามารถเพิ่มมูลค่าทางการเงินด้วยผลตอบแทนทั้งจากกำไรการขายและเงินปันผลจากจำนวนเงินที่ลงทุนในตลาดหลักทรัพย์ที่อยู่ภายใต้การกำกับของตลาดหลักทรัพย์แห่งประเทศไทย ซึ่งในตลาดหลักทรัพย์แห่งประเทศไทยได้รับรองบริษัทจดทะเบียนประกอบไปด้วยบริษัทที่จดทะเบียนในตลาดหลักทรัพย์ (The Securities Exchange of Thailand : SET) และตลาดเอ็มเอไอ (Market for Alternative Investment : mai)

ทุกวันนี้การคาดการณ์เพื่อหาแนวโน้มการเติบโตของกิจการในตลาดหลักทรัพย์เป็นประเด็นที่มีความน่าสนใจเนื่องจากได้ผลตอบแทนกำไรที่มีจำนวนมาก โดยการพยากรณ์เพื่อจัดหากลุ่มหลักทรัพย์ที่มีการเติบโตของกิจการได้อย่างแม่นยำและได้กำไรสิ่งนี้จึงเป็นสิ่งที่สามารถดึงดูดความสนใจของนักลงทุนหลายคน และเป็นวิธีที่รวดเร็วกว่าการวิเคราะห์จากงบการเงินที่ละตัว โดยใช้อัลกอริทึมในการสร้างแบบจำลองมาเป็นตัวช่วยในการพยากรณ์ ในบทความและงานวิจัยส่วนใหญ่จะให้ความสนใจไปที่ธุรกิจกลุ่ม SET ซึ่งเป็นกลุ่มบริษัทในตลาดหลักทรัพย์ที่ใหญ่ และมีการแบ่งกลุ่มธุรกิจที่ชัดเจน แต่ในทางกลับกันตลาดเอ็มเอไอ ยังไม่มีงานวิจัยหรือบทความที่น่าไปสู่การจัดกลุ่มหุ้นในลักษณะเช่นเดียวกับ SET มากนักซึ่งสาเหตุใหญ่มาจากความนิยมและปริมาณในการซื้อขายที่ไม่สูง และตลาดเอ็มเอไอส่วนมากเป็นธุรกิจ SMEs ซึ่งเป็นบริษัทที่มีโอกาสที่จะเติบโตสูง แต่ในทางกลับกันก็อาจมีบางบริษัทมีโอกาสล้มละลายได้เหมือนกัน แต่ถ้าหากมีอัลกอริทึมที่ช่วยพยากรณ์สนับสนุนการตัดสินใจก่อนในการลงทุนได้ เพื่อมองหาบริษัทที่มีโอกาสในการเติบโตของกิจการในอีก 3 ปีข้างหน้า อีกทั้งยังช่วยลดความเสี่ยงในการลงทุนที่จะเกิดขึ้น โดยผู้วิจัยมองจากงบทางการเงินของกิจการครอบคลุมตั้งแต่การเติบโตของสินทรัพย์ รายได้ กำไรและ

การเติบโตของกิจการ จะอยู่ในระดับที่สูงกว่าค่าเฉลี่ยของตลาด โดยมีการเติบโตของรายได้และกำไรเติบโตเร็วปีละ 15% ขึ้นไป ต่อเนื่องกันหลาย ๆ ปี ซึ่งสอดคล้องกับหลักเกณฑ์ที่นักลงทุนส่วนใหญ่เลือกซื้อหุ้นคือการมองหาหุ้นที่มีการเติบโตที่อิงจากกำไรของกิจการในระยะยาว ซึ่งเป็นจุดเด่นที่น่าสนใจของหุ้นขนาดเล็กอย่างเอ็มเอไอ

ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์เพื่อจัดกลุ่มหลักทรัพย์เอ็มเอไอ ในการพยากรณ์หุ้นตัวไหนจะเป็นหุ้นเติบโตในอีก 3 ปีข้างหน้าโดยอัลกอริทึมที่ใช้ทั้งหมด 4 อัลกอริทึมในการสร้างแบบจำลองเพื่อเปรียบเทียบแบบจำลองที่แม่นยำที่สุด ได้แก่ Decision Tree, Random Forest, K-nearest neighbor และ Naive Bayes มาเป็นตัวช่วยในการพยากรณ์ ซึ่งงานวิจัยนี้จะสามารถนำผลลัพธ์ที่ได้จากการพยากรณ์มาช่วยในการตัดสินใจให้กับนักลงทุนในการลงทุนหลักทรัพย์ในอนาคตได้

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การเติบโตของกิจการ

การเติบโตของกิจการมาจากยอดขายที่เพิ่มขึ้น ต้นทุนที่ลดลง ทำให้กำไรสุทธิมากขึ้น ซึ่งบริษัทจะต้องทำการลงทุนเพื่อการเติบโต ถ้าลงทุนแล้วให้ผลตอบแทนจากการลงทุนที่ดีก็จะทำให้ผลกำไรเติบโตมากขึ้น ดังนั้นปัจจัยหลักของการเติบโต คือ หากมีการลงทุนมาก และการลงทุนนั้นทำผลตอบแทนได้ดีก็จะทำให้กิจการนั้นมีการเติบโต [1]

### 2.2 อัตราส่วนทางการเงินที่สำคัญ

การนำตัวเลขที่อยู่ในงบการเงิน มาคำนวณเป็นอัตราส่วนเพื่อใช้ในการวิเคราะห์เปรียบเทียบผลการดำเนินงานของกิจการในอดีต ได้แก่

อัตรากำไรขั้นต้น (%) เป็น การ วัด อัตราส่วนเปรียบเทียบผลกำไรขั้นต้นกับยอดขาย สามารถประเมินประสิทธิภาพในการดำเนินงานของกิจการ เพื่อวัดความสามารถของบริษัทในการควบคุมต้นทุนสินค้า และความสามารถในการปรับราคาขายสินค้า [2]

อัตรากำไรจากการดำเนินงาน (%) เพื่อประเมินว่าบริษัทมีกำไรจากการดำเนินงานเป็นกี่เปอร์เซ็นต์ของรายได้ [3]

อัตรากำไรสุทธิ (%) อัตรารส่วนทางการเงินที่วัดความสามารถในการทำกำไรของบริษัท ว่าบริษัทมีกำไรสุทธิ คิดเป็นสัดส่วนเท่าไรจากรายรับทั้งหมดของบริษัท

อัตรารส่วนผลตอบแทนจากสินทรัพย์ (%) สัดส่วนระหว่างกำไรสุทธิ (Net Profit) และสินทรัพย์รวม (Asset) สามารถบอกได้ว่าบริษัททำกำไรได้มากน้อยเพียงใด เมื่อเทียบกับสินทรัพย์รวมของบริษัทนั้น

อัตรากำไรเติบโตของรายได้ (%) เป็นกระแสเงินสดที่กิจการได้รับจากการขายสินค้าหรือบริการ

อัตรากำไรเติบโตของกำไร (%) อัตรารส่วนการเติบโตของกำไรสุทธิเมื่อเทียบกับปีก่อนหน้าเพื่อแสดงให้เห็นว่าบริษัทมีความสามารถทำกำไรหรือขาดทุนเพิ่มขึ้นหรือลดลง [4]

อัตรานี้สินรวม ต่อ สินทรัพย์รวม (%) อัตรารส่วนทางการเงินที่วัดระหว่างหนี้สินรวมกับสินทรัพย์รวม ซึ่งเป็นอัตรารส่วนที่บ่งบอกว่ากิจการมีการกู้ยืมเงินเป็นส่วนส่วนเท่าไรต่อสินทรัพย์รวม

อัตรารส่วนเงินปันผลตอบแทน (%) อัตรารส่วนทางการเงินที่เป็นการเปรียบเทียบระหว่าง เงินปันผลต่อหุ้นเทียบกับ ราคา โดยพิจารณาบริษัทที่มี อัตรารเงินปันผลตอบแทนที่ค่อนข้างสูง

อัตรารส่วนผลตอบแทนจากผู้ถือหุ้น (%) อัตรารส่วนที่ช่วยวัดประสิทธิภาพของบริษัท ว่าสามารถสร้างผลตอบแทนจากส่วนของผู้ถือหุ้นหรือส่วนของผู้ถือหุ้นได้มากน้อยเพียงใด เพื่อวัดการเติบโตของกำไรได้ [5]

อัตรากำไรเติบโตของกำไรต่อหุ้น (%) คืออัตรากำไรเปลี่ยนแปลงกำไรต่อหุ้น (EPS) เมื่อวัดข้ามช่วงเวลา เพื่อเปรียบเทียบว่าในแต่ละช่วงเวลาบริษัทมีอัตรารผลตอบแทนให้แก่ผู้ถือหุ้น 1 หุ้น เพิ่มขึ้นหรือน้อยลง

### 2.3 เทคนิคการจำแนกข้อมูลด้วยวิธี Decision Tree

เป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล โดยต้นไม้ช่วยการตัดสินใจจะมีลักษณะคล้ายโครงสร้างต้นไม้ ที่แต่ละโหนดแสดงคุณลักษณะ (attribute) แต่ละกิ่งแสดงเงื่อนไขในการทดสอบ และโหนดปลาย (leaf node) แสดงกลุ่มที่กำหนดไว้ โดยการสร้างโมเดล (Model) จะทำโดยวิธีการขั้นตอนดังนี้ 1. หาคุณลักษณะ (Attribute) ที่สำคัญที่สุดมาแบ่ง

ข้อมูลซึ่งจะถูกนำมาสร้างเป็น Root Node โดยมี Target Attribute เป็นผลลัพธ์ซึ่งเป็น Leaf Node 2. นำค่าที่เป็นไปได้ใน Attribute ที่ถูกเลือกมาแตกออกเป็นกลุ่ม และทำการแบ่งข้อมูลทั้งหมดตามกลุ่มที่แตกออกจาก Root Node 3. ทำซ้ำที่ขั้นตอนแรก โดยหา Attribute ที่สำคัญที่สุดจากข้อมูลที่เข้ามาเพื่อหาตัวแบ่งต่อไป [6]

### 2.4 เทคนิคการจำแนกข้อมูลด้วยวิธี Random Forest

Random Forest เป็นหนึ่งในเทคนิคที่เรียกว่า Bagging หลักการจะคล้ายกับ Decision Tree แต่จะสุ่มเอาข้อมูล (Instance) ไปสร้างเป็นต้นไม้หลายๆต้น แต่ละต้นเรียกว่า Subset เหมือนว่ามีป่าที่มีต้นไม้จำนวนมาก ซึ่งแต่ละต้นจะมีรูปแบบสุ่มไม่เหมือนกัน โดยในการทำนายแต่ละต้นจะทำนายแล้วเลือกผลที่ได้จากการโหวตที่มากที่สุดมาเป็นผลลัพธ์ [7]

### 2.5 เทคนิคการจำแนกข้อมูลด้วยวิธี K-Nearest Neighbor (KNN)

เป็นเทคนิคหนึ่งที่ใช้ในการแก้ปัญหาการจำแนกข้อมูล (Classification) ใช้ในการจำแนกกลุ่มข้อมูลและเป็นวิธีการที่เหมาะสมกับข้อมูลที่กระจัดกระจาย โดยหลักการ คือ จำแนกกลุ่มข้อมูลด้วยระยะทางที่มีค่าใกล้เคียงที่สุดกับข้อมูลในแต่ละกลุ่มซึ่งกำหนดโดยค่าของข้อมูลในรอบการวนซ้ำก่อนหน้า เมื่อได้ระยะทางที่ถูกคำนวณแล้วจะถูกจัดเรียงลำดับและกำหนดข้อมูลเพื่อนบ้านตามระยะทางที่ใกล้ที่สุดจำนวน k ตามที่กำหนดไว้ การคำนวณระยะทางมีหลากหลายวิธีซึ่งการนำมาใช้งานขึ้นอยู่กับลักษณะของตัวแปร ได้แก่ การคำนวณระยะทาง Euclidean และการคำนวณระยะทาง Manhattan เหมาะสำหรับตัวแปรเป็นตัวเลข (continuous variable) ซึ่งเป็นวิธีที่นำมาใช้มากที่สุด การคำนวณระยะทาง Hamming เหมาะสำหรับตัวแปรที่ไม่ใช่ตัวเลข [8]

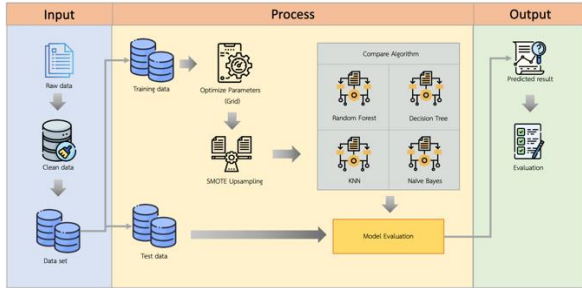
### 2.6 เทคนิคการจำแนกข้อมูลด้วยวิธี Naïve Bayes

เป็นเทคนิคที่อาศัยหลักการความน่าจะเป็นตามทฤษฎีของเบย์ (Bayes's theorem) ซึ่งเป็นอัลกอริทึมที่ไม่ซับซ้อน หลักการของ Naïve Bayes ใช้หลักการของความน่าจะเป็น โดยมีสมมติฐานว่าปริมาณของความสนใจขึ้นอยู่กับการกระจายความน่าจะเป็น อัลกอริทึมช่วยในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ [7]



### 3.วิธีการดำเนินการวิจัย

วิจัยนี้เป็นการพยากรณ์หุ้นเติบโตจากการจัดกลุ่มหลักทรัพย์เอ็มเอไอ โดยมีกรอบแนวความคิด และขอบเขตในการดำเนินวิจัย ดังภาพที่ 1



ภาพที่ 1: กรอบแนวความคิดในการดำเนินวิจัย

โดยผู้วิจัยได้ดำเนินการตามขั้นตอน และมีกระบวนการทำงานดังนี้

#### 3.1 ศึกษาเก็บรวบรวม และเตรียมข้อมูล

ผู้วิจัยได้ศึกษาและรวบรวมข้อมูลหลักทรัพย์เอ็มเอไอ จากตลาดหลักทรัพย์แห่งประเทศไทยย้อนหลัง 5 ปี ตั้งแต่ พ.ศ.2555 – 2559 จำนวน 1,101 รายการ ซึ่งเป็นข้อมูลงบการเงินของแต่ละบริษัทที่อยู่ในตลาดหลักทรัพย์เอ็มเอไอ ปัจจัยสำหรับการสร้างแบบจำลองดังตารางที่ 1

ตารางที่ 1: ปัจจัยสำหรับการสร้างแบบจำลอง

แอตทริบิวต์	ความหมาย
Securities (ID)	ชื่อหลักทรัพย์
GPM	อัตรากำไรขั้นต้น
EBIT Margin	อัตรากำไรจากการดำเนินงาน
NPM	อัตรากำไรสุทธิ
ROA	อัตราส่วนผลตอบแทนจากสินทรัพย์
Revenue Growth Rate	อัตราการเติบโตของรายได้
Net Profit Growth Rate	อัตราเติบโตของกำไร
Debt Ratio	อัตราหนี้สินรวมต่อสินทรัพย์รวม
Dividend Yield	อัตราส่วนเงินปันผลตอบแทน
ROE	อัตราส่วนผลตอบแทนจากผู้ถือหุ้น
EPS Growth Rate	อัตราการเติบโตของกำไรต่อหุ้น
Growth Result (Label)	ผลการเติบโตของกิจการ

โดยแต่ละแอตทริบิวต์จะใช้ของปี พ.ศ.2555 – 2559 จากนั้นนำข้อมูลเข้าสู่กระบวนการแปลงข้อมูลคือการทำให้ข้อมูลให้ถูกต้องโดยทำขั้นตอน Data Cleaning และนำข้อมูลมาทำการ Pivot เพื่อให้ได้ข้อมูลที่สมบูรณ์ และถูกต้องที่สุดในการนำไปพยากรณ์ ทำให้ตัวอินพุตมีทั้งหมด 50 ตัวจากการนำข้อมูลมาทำการ Pivot

#### 3.2 การลดความไม่สมดุลของข้อมูลด้วยวิธี SMOTE

##### Upsampling

จากการที่ข้อมูลได้ผ่านการเตรียมข้อมูล และทำการ pivot เรียบร้อยแล้วผู้วิจัยพบว่าข้อมูลในแต่ละคลาสมีจำนวนที่ไม่สมดุลกันหรือเรียกว่าเป็น Unbalanced data ซึ่งข้อมูล เมื่อผ่านการเตรียมข้อมูลเหลือทั้งหมด 138 รายการ โดยข้อมูลหลักทรัพย์ที่เป็นคลาส Growth จำนวน 23 รายการ และข้อมูลหลักทรัพย์ที่เป็นคลาส Nongrowth จำนวน 115 รายการสำหรับการแก้ปัญหา คือการทำ sampling ข้อมูล จากปัญหาดังกล่าวผู้วิจัยจึงทำการปรับความไม่สมดุลของข้อมูลโดยใช้แบบ Oversampling สร้างข้อมูลคลาสที่มีจำนวนน้อยกว่าให้มีปริมาณมากขึ้น เพราะการทำเช่นนี้ไม่ได้ทำให้ข้อมูลมีการเปลี่ยนแปลง ดังนั้นวิธีการที่ใช้คือการสร้างข้อมูลใหม่มาไว้ใกล้กับข้อมูลเดิมที่มีอยู่เทคนิคที่นิยมใช้ในกรณีนี้คือ SMOTE Upsampling เป็น Operators ใน Rapidminer ใช้ช่วยในการแก้ไข Unbalanced data ผู้วิจัยได้ทำการหาค่าพารามิเตอร์โดยใช้ Optimize Parameter (Grid) มาช่วยในการหาค่าพารามิเตอร์ที่เหมาะสมที่สุดโดยวัดจากค่าความแม่นยำ (Accuracy) ที่สูงที่สุด แล้วพบว่าค่า SMOTE - Number of neighbours ที่เหมาะสมที่สุดของแต่ละอัลกอริทึมดังตารางที่ 2 โดยจำนวนข้อมูลทั้งหมดที่ได้จากการทำ SMOTE Upsampling เพิ่มเป็นจำนวน 230 รายการ

ตารางที่ 2: ผลประเมินค่า Number of neighbours (k) ที่ดีที่สุดของแต่ละอัลกอริทึม

อัลกอริทึม	Number of neighbours
Decision Tree	5
Random Forest	20
K-nearest neighbor	20
NaiveBayes	5

### 3.3 การสร้างแบบจำลอง

ขั้นตอนการนำสู่การเรียนรู้เพื่อสร้างแบบอัลกอริทึม โดยใช้ 4 อัลกอริทึม ได้แก่ Decision Tree, Random Forest, K-nearest neighbor และ NaïveBayes เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์การเติบโตของหลักทรัพย์เอ็มเอไอ ได้ความแม่นยำมากที่สุด ผู้วิจัยได้ทดลองกำหนดค่าพารามิเตอร์ในการสร้างแบบอัลกอริทึมทั้ง 4 โดยการใช้ Optimize Parameter (Grid) ใน Rapidminer มาช่วยในการหาค่าพารามิเตอร์ที่เหมาะสมที่สุด โดยวัดจากค่าความแม่นยำ (Accuracy) ที่สูงที่สุดของทั้ง 4 อัลกอริทึม โดยแต่ละอัลกอริทึมมีค่าพารามิเตอร์ดังตารางที่ 3

ตารางที่ 3: ตารางแสดงค่าพารามิเตอร์ในการกำหนดอัลกอริทึม

อัลกอริทึม	พารามิเตอร์	ค่าที่ใช้
Decision Tree	Maximal depth	20
	Confidence	0.01
	Criterion	information_gain
Random Forest	Number of trees	30
	Maximal depth	60
	Confidence	0.01
	Criterion	gain_ratio
K-nearest neighbor	K	6
NaïveBayes	Laplace correction	Ture

### 3.4 การทดสอบ และวัดประสิทธิภาพของอัลกอริทึม

ในการทดสอบ และวัดประสิทธิภาพผู้วิจัยใช้วิธี Cross validation เป็นการจำแนกข้อมูลออกเป็นหลายส่วน (K) โดยแต่ละส่วนจะมีจำนวนข้อมูลเท่ากัน เมื่อแบ่งข้อมูลได้จะทำการฝึกสอนอัลกอริทึม โดยเหลือข้อมูลชุดหนึ่งเอาไว้สำหรับเป็นส่วนทดสอบ เมื่อทดสอบอัลกอริทึม รอบแรกเสร็จข้อมูลที่เหลือไว้จะนำกลับเข้าไปเป็นชุดข้อมูลฝึกสอนแล้วแบ่งอีกส่วนไว้สำหรับเป็นชุดข้อมูลทดสอบ โดยผู้วิจัยได้ทดลองหาค่าพารามิเตอร์ค่า k ทั้ง 4 อัลกอริทึม โดยการใช้ Optimize Parameter (Grid) ใน Rapidminer มาช่วยในการหาค่า k ที่เหมาะสมที่สุด โดยวัดจากค่าความแม่นยำ (Accuracy) ที่สูงที่สุดของทั้ง 4 อัลกอริทึม ได้ดังตารางที่ 4

ตารางที่ 4: ตารางแสดงค่า k-fold cross validation

อัลกอริทึม	k-fold cross validation
Decision Tree	40
Random Forest	20
K-nearest neighbor	40
NaïveBayes	10

วัดประสิทธิภาพแบบจำลองด้วยวิธี Confusion matrix ตัวชี้วัด ได้แก่ ค่าความถูกต้อง (Accuracy) ดังสมการที่ 2 ค่าความแม่นยำ (Precision) ดังสมการที่ 3 และค่าระลึก (Recall) ดังสมการที่ 4

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (4)$$

โดยที่ TP : ข้อมูลที่ทำนายตรงกับข้อมูลจริงในคลาสที่กำลังพิจารณา

TN: ข้อมูลที่ทำนายตรงกับข้อมูลจริงในคลาสที่ไม่ได้พิจารณา

FP: ข้อมูลที่ทำนายผิดเป็นคลาสที่กำลังพิจารณา

FN: ข้อมูลที่ทำนายผิดเป็นคลาสที่ไม่ได้พิจารณา

## 4.ผลการวิจัย

ผลการวิจัยการพยากรณ์หุ้นเติบโตจากการจัดกลุ่มหลักทรัพย์เอ็มเอไอของตลาดหลักทรัพย์แห่งประเทศไทย ได้ผลการเปรียบเทียบของทั้ง 4 อัลกอริทึมดังนี้

### 4.1 ผลทดสอบอัลกอริทึม Decision Tree

	true Non Growth	true Growth	class precision
pred. Non Growth	95	17	84.82%
pred. Growth	20	98	83.05%
class recall	82.61%	85.22%	

ภาพที่ 2: ผลทดสอบอัลกอริทึม Decision Tree

### 4.2 ผลทดสอบอัลกอริทึม Random Forest

	true Non Growth	true Growth	class precision
pred. Non Growth	97	4	96.04%
pred. Growth	18	111	86.05%
class recall	84.35%	96.52%	

ภาพที่ 3: ผลทดสอบอัลกอริทึม Random Forest

### 4.3 ผลทดสอบอัลกอริทึม K-nearest neighbor

	true Non Growth	true Growth	class precision
pred. Non Growth	85	11	88.54%
pred. Growth	30	104	77.61%
class recall	73.91%	90.43%	

ภาพที่ 4: ผลทดสอบอัลกอริทึม K-nearest neighbor

### 4.4 ผลทดสอบอัลกอริทึม NaïveBayes

	true Non Growth	true Growth	class precision
pred. Non Growth	53	3	94.64%
pred. Growth	62	112	64.37%
class recall	46.09%	97.39%	

ภาพที่ 5: ผลทดสอบอัลกอริทึม NaïveBayes

### 4.5 ผลการเปรียบเทียบอัลกอริทึม

ผลจากการนำชุดข้อมูลที่เตรียมไว้นามาทดสอบกับอัลกอริทึมที่ได้ทำการกำหนดค่าพารามิเตอร์ที่ดีที่สุดและมีความแม่นยำที่สุดของ 4 อัลกอริทึมเปรียบเทียบได้ดังตารางที่ 5

ตารางที่ 5: แสดงผลการทดสอบอัลกอริทึม

อัลกอริทึม	Accuracy	Precision	Recall
Decision Tree	84.33%	85.83%	85.62%
Random Forest	90.42%	87.29%	96.33%
K-nearest neighbor	82.33%	81.25%	90.83%
NaïveBayes	71.74%	65.08%	97.27%

จากตารางที่ 5 แสดงผลการประเมินการทดสอบอัลกอริทึมด้วยเทคนิค Cross validation ของทั้ง 4 อัลกอริทึม ผู้วิจัยพบว่า อัลกอริทึม Random Forest มีประสิทธิภาพมากที่สุด โดยมีค่าความถูกต้อง(Accuracy) เท่ากับ 90.42% ค่าความแม่นยำ (Precision) เท่ากับ 87.29% และค่าระลึก (Recall) เท่ากับ 96.33% ถัดมาคืออัลกอริทึม K-nearest neighbor, อัลกอริทึม Decision Tree และ อัลกอริทึม NaïveBayes ตามลำดับ

## 5.สรุปผลการวิจัย

จากการสร้างแบบจำลองการพยากรณ์หุ้นเติบโตจากการจัดกลุ่มหลักทรัพย์เอ็มเอไอของตลาดหลักทรัพย์แห่งประเทศไทยได้ทำการปรับการลดความไม่สมดุลของข้อมูลด้วยวิธี SMOTE Upsampling เพื่อแก้ปัญหาจำนวนข้อมูลในแต่ละคลาสมีจำนวนที่ไม่สมดุลกัน และทำการเปรียบเทียบอัลกอริทึมของทั้ง 4 อัลกอริทึม ได้แก่ Decision

Tree, Random Forest , K-nearest neighbor และ NaïveBayes เพื่อหาอัลกอริทึมที่มีประสิทธิภาพ และความแม่นยำที่สุด และทำการประเมินประสิทธิภาพอัลกอริทึมด้วยเทคนิค Cross validation จากการทดลองผลการวิจัยสรุปได้ว่าอัลกอริทึม Random Forest มีประสิทธิภาพมากที่สุด และเหมาะสมกับชุดข้อมูลที่สุด โดยมีค่า k-fold cross validation ที่ดีที่สุดเท่ากับ 50 การปรับการลดความไม่สมดุลของข้อมูลที่เหมาะสมที่สุด ค่า SMOTE - Number of neighbours เท่ากับ 20 โดยมีตัวแปรอินพุตที่ดีที่สุดจำนวน 50 ตัว และเมื่อนำชุดข้อมูลที่ได้มาทำการสร้างแบบจำลอง ผู้วิจัยพบว่าอัลกอริทึม Random Forest มีค่าความถูกต้อง(Accuracy) เท่ากับ 90.42% ค่าความแม่นยำ (Precision) เท่ากับ 87.29% และค่าระลึก (Recall) เท่ากับ 96.33% ซึ่งสามารถนำมาพัฒนาพยากรณ์การเติบโตของหุ้นเอ็มเอไอของตลาดหลักทรัพย์ได้

## เอกสารอ้างอิง

- [1] Investment Reader. (2561). เราจะประมาณการเติบโต (Growth) อย่างไร, สืบค้นเมื่อ 7 กุมภาพันธ์ 2565. จาก <https://www.finnomena.com/investment-reader/growth/>
- [2] Mr.LikeStock. (2561). อัตรากำไรขั้นต้น (Gross Profit Margin), สืบค้นเมื่อ 7 กุมภาพันธ์ 2565. จาก [www.mrlikestock.com](http://www.mrlikestock.com)
- [3] อธิป กิรติพัชญ์, ดิวหุ้นรวยด้วยวีไอ, พิมพ์ครั้งที่ 1 (สำนักพิมพ์ สตีลคทูนอร์โรว์ จำกัด,2556), 250 หน้า
- [4] Online Trading. (2564). รู้จัก ROE คืออะไร และวัดผลตอบแทนของผู้ถือหุ้นได้อย่างไร, สืบค้นเมื่อ 7 กุมภาพันธ์ 2565. จาก [www.knowledge.bualuang.co.th](http://www.knowledge.bualuang.co.th)
- [5] moneybuffalo. (2563). "Debt Ratio" คืออะไร ?, สืบค้นเมื่อ 7 กุมภาพันธ์ 2565. จาก [www.moneybuffalo.in.th](http://www.moneybuffalo.in.th)
- [6] ธนวัฒน์ ปัญญาวงศ์. (2555). การทำเหมืองข้อมูลด้วยเทคนิคการจำแนกข้อมูล (Classification).เทคนิค WEKA.
- [7] Han, J, Kamber Mi. and Pei J. (2 0 1 2 ). Data Mining Concepts and Techniques. 3rd ed. USA: Morgan Kaufmann.
- [8] Pimporn Pimpim. (2562). 0 4 . 1 Classification using K-Nearest Neighbor (K-NN). Medium.com.

# ระบบแนะนำการทำประกันภัยรถยนต์โดยใช้การสร้างกฎเฉพาะบุคคล ร่วมกับการแบ่งกลุ่มแบบเคมีน

## Car Insurance Recommendation System Using Rules-Based Personalization with K-Means Clustering

ปีใหม่ กีฬากิจไพบูลย์ (Peemai Keelakitpaibool)<sup>1</sup> และศุสดี บุญรอด (Pudsadee Boonrawd)<sup>2</sup>

ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307011858217@email.kmutnb.ac.th, <sup>2</sup>pudsadee.b@itd.kmutnb.ac.th

### บทคัดย่อ

ธุรกิจประกันภัยที่ให้บริการในปัจจุบันมีหลายประเภท ตัวอย่างเช่น สุขภาพ ยานพาหนะ ทรัพย์สิน ที่อยู่อาศัย ขนส่งสินค้า และอุบัติเหตุ เป็นต้น ธุรกิจประกันภัยรถยนต์เป็นหนึ่งในธุรกิจที่เติบโตอย่างรวดเร็วและมีการแข่งขันกันสูง จึงเป็นเรื่องง่ายที่ผู้ทำประกันภัยจะขอยกเลิกกรมธรรม์เพื่อทำประกันกับบริษัทอื่นที่ให้เงื่อนไขที่ดีกว่าหรือราคาถูกกว่า บางครั้งแผนความคุ้มครองที่เลือกอาจไม่เหมาะสมกับผู้ทำประกันภัยอีกด้วย งานวิจัยนี้จึงได้พัฒนาระบบแนะนำการทำประกันภัยรถยนต์โดยใช้การสร้างกฎเฉพาะบุคคล (Rules-Based Personalization) ที่พิจารณาจากข้อมูล และกฎต่าง ๆ ที่สร้างขึ้นเพื่อรองรับข้อมูลผู้ขับขี่ ข้อมูลรถยนต์ ทุนประกันภัย ประเภทการใช้งาน และแผนความคุ้มครองที่เกี่ยวข้อง ซึ่งในงานวิจัยยังประยุกต์ใช้การแบ่งกลุ่มแบบเคมีน (K-Means Clustering) เพื่อช่วยแนะนำแผนความคุ้มครองที่คาดว่าผู้ทำประกันภัยสนใจ จากการทดสอบประสิทธิภาพของระบบ พบว่า ระบบที่พัฒนาขึ้นสามารถให้คำแนะนำสำหรับการทำประกันภัยรถยนต์ตามความต้องการเฉพาะบุคคลได้อย่างมีประสิทธิภาพ ง่ายต่อการตัดสินใจเลือกแผนความคุ้มครอง อีกทั้งยังช่วยลดขั้นตอน และระยะเวลาในการทำประกันภัยรถยนต์ของลูกค้า

**คำสำคัญ:** ระบบแนะนำการทำประกันภัยรถยนต์ การสร้างกฎเฉพาะบุคคล การแบ่งกลุ่มแบบเคมีน

### Abstract

There are various types of insurance businesses that offer services today, for example, health insurance, car insurance, property insurance, householder insurance, and accident insurance. Car insurance is one of the fastest-growing businesses with a highly competitive and low average of brand royalty. Therefore, the customers can switch to the competitor product easily if the coverage plan or condition is more suitable and better. The objective of this research was to develop the Car Insurance Recommendation System Using Rules-Based Personalization that considered the data and rules created to analyze driver data, vehicle data, sum insured, usage type, and coverage plans. The research applied K-Means Clustering to recommend insurance plans that insurers are expected to be interested in. System performance measurements show that the developed system can provide effective recommendations for car insurance according to the needs of customers efficiently, making it easier to decide on a coverage plan, reducing the process and duration of car insurance for customers.

**Keyword:** Car Insurance Recommendation System, Rules-Based Personalization, K-Means Clustering

## 1. บทนำ

ในปัจจุบันบริษัทประกันภัยรถยนต์มีการแข่งขันกันสูง และมีหลากหลายช่องทางในการให้บริการ เช่น ผ่านทาง นายหน้าประกันภัย เทเลเชลล์ และเว็บไซต์ เป็นต้น ในการนำเสนอแผนความคุ้มครองให้กับลูกค้าจาก นายหน้า เพื่อให้ตัดสินใจในการทำประกันภัยรถยนต์ ลูกค้าบางคนอาจตัดสินใจเลือกซื้อแผนความคุ้มครองที่ไม่เหมาะสม หรือไม่ตรงตามความต้องการของลูกค้า ในบางกรณีอาจได้รับแผนความคุ้มครองที่มีค่าเบี้ยประกันภัย แพงเกินความจำเป็น หรือการแนะนำผ่านช่องทาง เทเลเชลล์ลูกค้าอาจฟังรายละเอียดของแผนความคุ้มครอง ไม่ครบถ้วน ทำให้อาจเสียผลประโยชน์ที่จะได้รับ

จากปัญหาที่กล่าวมาข้างต้น งานวิจัยนี้จึงได้พัฒนาระบบแนะนำการทำประกันภัยรถยนต์โดยใช้การสร้างกฎ เฉพาะบุคคลร่วมกับการแบ่งกลุ่มแบบเคมีน โดยระบบจะ ช่วยในการวิเคราะห์ และให้คำแนะนำแผนความคุ้มครอง ที่มีความเหมาะสมกับผู้ทำประกันรถยนต์แต่ละบุคคล ง่ายต่อการตัดสินใจเลือกแผนความคุ้มครอง ช่วยลดความ เสี่ยงที่ลูกค้าจะยกเลิกกรมธรรม์ อีกทั้งยังช่วยลดขั้นตอน และระยะเวลาในการทำประกันภัยรถยนต์ของลูกค้า โดย ในบทความวิจัยแบ่งเนื้อหาออกเป็น 5 ส่วน ได้แก่ ส่วนที่ 1 บทนำ ส่วนที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง ส่วนที่ 3 วิธีการดำเนินงานวิจัย ส่วนที่ 4 ผลการดำเนินงานวิจัย และ ส่วนที่ 5 สรุปผลการวิจัย

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การประกันภัยรถยนต์

ธุรกิจประกันภัยรถยนต์เป็นธุรกิจที่คุ้มครองผู้เอา ประกันภัยหรือผู้ที่ได้รับการคุ้มครองจากกรมธรรม์เมื่อเกิด ความสูญเสีย หรือเสียหายจากการใช้รถยนต์ โดยที่บริษัท ประกันภัยจะต้องบริหารความเสี่ยงที่อาจจะเกิดขึ้นที่ อยู่ภายใต้ข้อกำหนดของสำนักงานคณะกรรมการกำกับ และส่งเสริมการประกอบธุรกิจประกันภัย (คปภ.) ซึ่งการ ประกันภัยรถยนต์ประกอบด้วยประเภทการรับประกันภัย และความเสียหายที่อาจจะเกิดขึ้นดังนี้

#### 2.1.1 ประเภทประกันภัยรถยนต์

ประกันภัยรถยนต์ประเภทภาคบังคับ (Compulsory) หรือประกันภัย พ.ร.บ. ซึ่งรถยนต์ทุกชนิดต้องทำประกัน ประเภทนี้ เนื่องจากถูกบังคับโดยพระราชบัญญัติคุ้มครอง ผู้ประสบภัยจากรถ พ.ศ. 2535 คุ้มครองความสูญเสียชีวิต ร่างกาย หรืออนามัยของบุคคลประสบภัยจากรถยนต์

ประกันภัยรถยนต์ประเภทภาคสมัครใจ (Voluntary) เป็นประกันรถยนต์ที่ได้รับความยินยอมระหว่างผู้ทำ ประกันภัย และบริษัทประกันภัยในการเลือกซื้อแผนความ คุ้มครองตามความต้องการของผู้ทำประกันภัย

#### 2.1.2 ความเสี่ยงของประกันภัยรถยนต์

ในส่วนความเสี่ยงต่อการยกเลิกกรมธรรม์บริษัท ประกันภัยได้มีปรับเปลี่ยนแผนความคุ้มครองอยู่ ตลอดเวลา เพื่อตอบสนองความต้องการของผู้ทำ ประกันภัย เนื่องจากผู้ทำประกันสามารถเปลี่ยนแปลง ยกเลิกประกันได้ตามความต้องการ [1] เมื่อผู้ให้บริการอื่น ตอบสนองความต้องการได้ดีกว่าในด้านต่าง ๆ หรือความ เสี่ยงจากการถูกฉ้อโกง จากรายงานสำนักงานสอบสวน กลาง (Federal Bureau of Investigation: FBI) พบว่าการ ฉ้อโกงทางประกันภัยนำไปสู่การสูญเสียทางเศรษฐกิจ ประมาณ 80,000 ล้านดอลลาร์สหรัฐ [2]

จากความเสี่ยงที่อาจส่งผลกระทบต่อธุรกิจประกันภัย รถยนต์ได้มีงานวิจัยที่นำเทคโนโลยีมาประยุกต์ใช้เพื่อลด ความเสี่ยง เช่น การพยากรณ์ลูกค้าที่มีความเสี่ยงจะยกเลิก ประกันภัยรถยนต์โดยใช้หลักการเรียนรู้ของเครื่องด้วย วิธีการ Random Forest ในการพยากรณ์ [1] และประเมิน ระดับความเสียหายภาพสามมิติของรถยนต์ เพื่อป้องกันการ ฉ้อโกงด้วย Augmented Reality [3] เป็นต้น

#### 2.2 ระบบแนะนำ (Recommendation System)

ระบบแนะนำเป็นระบบที่ออกแบบมาเพื่อให้คำแนะนำ สิ่งใดสิ่งหนึ่งกับผู้ใช้ โดยพิจารณาจากปัจจัยต่าง ๆ เพื่อ คาดการณ์ถึงสิ่งที่ผู้ใช้สนใจ ซึ่งในปัจจุบันระบบแนะนำได้ นำไปใช้อย่างแพร่หลาย เช่น ด้านสื่อบันเทิง ซื้อขายสินค้า และหลักสูตรการเรียนรู้ [4] เพื่อช่วยให้ผู้ใช้งานง่ายต่อการตัดสินใจในการเลือกผลิตภัณฑ์หรือบริการ โดยมีวิธีการ ต่าง ๆ ที่นำมาประยุกต์กับระบบแนะนำดังนี้

2.2.1 วิธีการกรองแบบมีเนื้อหา (Content-Based Filtering)

วิธีการคัดกรองแบบมีเนื้อหาที่ประยุกต์ใช้ในระบบแนะนำเพื่อทำการวิเคราะห์พฤติกรรมของผู้ใช้งานในอดีต โดยทำการคำนวณเปรียบเทียบคุณลักษณะของผลิตภัณฑ์หรือบริการ เช่น ผู้ใช้งานเคยซื้อสินค้าออนไลน์มีการให้คะแนนสูง หรือชื่นชอบ ซึ่งระบบจะแนะนำผลิตภัณฑ์ที่เป็นประเภทเดียวกันให้ผู้ใช้โดยอ้างอิงจากประวัติของผู้ใช้เป็นหลัก [5]

2.2.2 วิธีการกรองแบบร่วมมือ (Collaborative Filtering)

วิธีการคัดกรองแบบร่วมมือเป็นวิธีการวิเคราะห์พฤติกรรมของผู้ใช้คนอื่นที่มีความคล้ายกับผู้ใช้ที่เป็นเป้าหมาย โดยดูจากการชื่นชอบผลิตภัณฑ์หรือบริการที่คล้ายกัน ซึ่งทำการคำนวณจากคุณลักษณะในการให้คะแนนของผลิตภัณฑ์หรือบริการ ดังนั้นระบบจะแนะนำผลิตภัณฑ์ที่คาดว่าผู้ใช้เป้าหมายจะสนใจ [5]

2.2.3 วิธีการแบบผสมผสานเทคโนโลยี

การนำวิธีการอื่น ๆ มาช่วยเพิ่มความสามารถในระบบแนะนำ เนื่องจากการแนะนำแบบพื้นฐานข้างต้นยังมีปัญหา และความถูกต้องที่ต้องปรับปรุง โดยมีวิธีการที่นำมาประยุกต์ใช้ เช่น โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN) เป็นแนวคิดการเลียนแบบสมองของมนุษย์ ซึ่งประกอบด้วย นิวรอน (Neurons) เป็นหน่วยประมวลผลพื้นฐานที่ทำหน้าที่ปรับเปลี่ยนสัญญาณข้อมูลระหว่างกัน [5] ซึ่งนำวิธีการนี้พัฒนาร่วมกับระบบแนะนำ เช่น การแนะนำสำหรับการออกกำลังกาย [6] และการแนะนำหลักสูตรการเรียนของนักเรียนมหาวิทยาลัยในประเทศไทย [7] วิธีการเชิงความหมาย (Semantic Technology) เป็นการพิจารณาจากเนื้อหาและความหมายของคำ โดยการพัฒนาแนะนำจะใช้ร่วมกับอนุกรมวิธาน (Taxonomy) และออนโทโลยี (Ontology) โดยทำการแบ่งกลุ่มสินค้า หรือประเภทสินค้าจากความหมาย [5] และวิธีการจัดกลุ่ม (Clustering) เป็นการพิจารณาจากข้อมูลของผู้ใช้งาน โดยมีการนำวิธีการจัดกลุ่มมาประยุกต์ใช้กับระบบแนะนำ เพื่อแนะนำสินค้าหรือบริการให้ตรงกับกลุ่มผู้ใช้งานที่มีความชอบคล้ายกัน

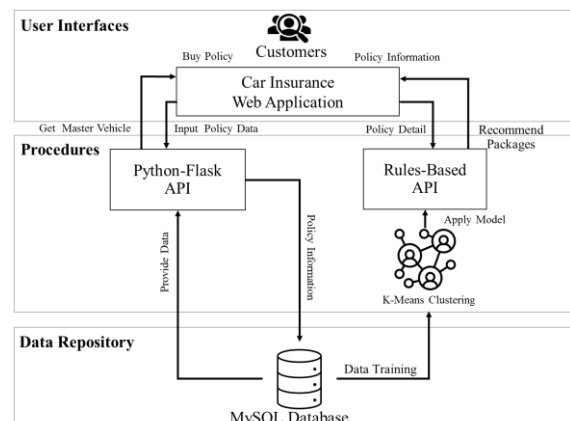
3. วิธีการดำเนินการวิจัย

3.1 การศึกษาปัญหาและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาปัญหาและงานวิจัยที่เกี่ยวข้องกับการทำประกันภัยรถยนต์ พบว่า ผู้ทำประกันอาจเลือกแผนความคุ้มครองที่ไม่เหมาะสม ทำให้อาจมีการยกเลิกกรมธรรม์ อีกทั้งในการทำประกันภัยมีหลายขั้นตอนทำให้ผู้ทำประกันได้รับกรมธรรม์ล่าช้า ใช้เวลานานกว่าจะทราบความคุ้มครองโดยละเอียด ซึ่งอาจจะทำให้เสียผลประโยชน์ในกรณีที่เกิดเหตุไม่คาดฝัน กรมธรรม์ไม่คุ้มครอง หรือคุ้มครองไม่เป็นไปตามที่ต้องการ งานวิจัยนี้จึงได้ทำการพัฒนาระบบแนะนำการทำประกันภัยรถยนต์ เพื่อแก้ปัญหาดังกล่าว

3.2 การวิเคราะห์และออกแบบกรอบแนวคิดการวิจัย

หลังจากศึกษาปัญหา และงานวิจัยที่เกี่ยวข้อง คณะผู้วิจัยได้ทำการวิเคราะห์และออกแบบกรอบแนวคิดการวิจัยของระบบแนะนำการทำประกันภัยรถยนต์ที่แบ่งการทำงานออกได้เป็น 3 ส่วน ได้แก่ ส่วนหน้าจของผู้ใช้งาน (User Interfaces) เพื่อระบุข้อมูลของลูกค้าที่จำเป็นต่อการทำประกันภัยรถยนต์ ส่วนกระบวนการ (Procedures) สำหรับตรวจสอบเงื่อนไขข้อมูลของผู้ทำประกันภัย เพื่อช่วยในการแนะนำแผนความคุ้มครอง และเชื่อมต่อการรับส่งข้อมูลในการทำกรมธรรม์กับฐานข้อมูล และส่วนเก็บข้อมูล (Data Repository) ที่รองรับการจัดเก็บข้อมูลต่าง ๆ ของระบบ ซึ่งสามารถแสดงได้ดังภาพที่ 1



ภาพที่ 1: กรอบแนวคิดการวิจัยของระบบแนะนำการทำประกันภัยรถยนต์

### 3.3 การเก็บข้อมูลและเตรียมข้อมูล

งานวิจัยใช้ข้อมูลจากสำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย (คปภ.) ปีพ.ศ. 2563 ที่นำข้อมูลเกี่ยวกับการทำประกันภัยรถยนต์ภาคสมัครใจทั้งรถเก๋ง และรถกระบะ จำนวน 10 ยี่ห้อรถยนต์ ยอดนิยมที่รวบรวมจาก Rabbit Finance ได้แก่ โตโยต้า ฮอนด้า มาสด้า นิสสัน อีซูซุ เซฟโรเลต มิตซูบิชิ ฟอर्ड ชูซูกิ และเอ็มจี ซึ่งข้อมูลก็นำมาใช้ในการพัฒนาระบบ ประกอบด้วย ข้อมูลกรมธรรม์ ข้อมูลรถยนต์ ข้อมูลแผนความคุ้มครอง ข้อมูลยี่ห้อ รุ่น รุ่นย่อย ข้อมูลเรียกร้อยสินไหมทดแทน ซึ่งสามารถแสดงสรุปจำนวนรายการของข้อมูลได้ดังตารางที่ 1

ตารางที่ 1 : สรุปจำนวนรายการของข้อมูลที่ใช้ในการวิจัย

ลำดับ	ข้อมูล	จำนวน (รายการ)
1	กรมธรรม์	218,542
2	รถยนต์	218,542
3	แผนความคุ้มครอง	468
4	ยี่ห้อ รุ่น รุ่นย่อย	5,766
5	เรียกร้อยสินไหมทดแทน	3,665

ข้อมูลการทำประกันภัยรถยนต์ภาคสมัครใจที่นำมาใช้ในการวิจัย มีจำนวนทั้งสิ้น 218,542 รายการ ซึ่งคณะผู้วิจัยจะต้องรวบรวมและจัดเตรียมข้อมูลโดยสร้างข้อมูลกรมธรรม์ และข้อมูลรถยนต์ต่าง ๆ ที่เกี่ยวข้อง จากนั้นทำความสะอาดข้อมูล เช่น ขจัดข้อมูลที่เป็นค่าว่าง และข้อมูลที่ไม่มีถูกต้อง เป็นต้น โดยแผนความคุ้มครองที่ใช้ในการวิจัยมีจำนวน 468 แผน ที่ครอบคลุมข้อมูลและประกันภัยทั้ง 4 ประเภท ได้แก่ ประเภทที่1 ประเภท 2+ ประเภท 3+ และประเภท 3 ซึ่งเป็นแผนคุ้มครองที่รองรับการใช้งานรถยนต์ส่วนบุคคล และใช้เพื่อการพาณิชย์ ข้อมูลยี่ห้อ รุ่น รุ่นย่อย คัดเลือกเฉพาะรุ่นที่อายุไม่เกิน 15 ปี โดยแต่ละรุ่นจะมีปีที่ผลิต ขนาดเครื่องยนต์ น้ำหนัก ที่นั่ง และราคารถยนต์ จำนวน 5,766 รายการ และข้อมูลรายการเรียกร้อยค่าสินไหมทดแทนจากกรมธรรม์ที่สร้างขึ้น จำนวน 3,665 รายการ จากนั้น ทำการตรวจสอบจำนวนเงินค่าสินไหม ซึ่งจะต้องไม่เกินทุนประกันภัยของกรมธรรม์

### 3.4 การออกแบบและพัฒนาระบบ

ในการออกแบบและพัฒนาระบบแนะนำประกันภัยรถยนต์ ซึ่งสามารถแบ่งออกได้เป็น 2 ส่วนหลัก ได้แก่ ส่วนที่ 1 การสร้างกฎเฉพาะบุคคล (Rules-Based Personalization) เป็นกฎที่ใช้ตรวจสอบแผนความคุ้มครองให้เป็นตามข้อกำหนดจากข้อมูลที่ได้รับของลูกค้า เพื่อนำไปใช้ในการให้คำแนะนำแผนความคุ้มครองที่เหมาะสม โดยสามารถแสดงตัวอย่างกฎที่ใช้กรองแผนความคุ้มครองเฉพาะบุคคลได้ดังตารางที่ 2

ตารางที่ 2: กฎที่ใช้กรองแผนความคุ้มครองเฉพาะบุคคล

Rules-Based Personalization
<pre> PackageList = [] for package in PackageData {     If (package.Suminsure &lt;= Car Price)     and (package.CarAgeMin &lt;= Car Age)     and (package.CarAgeMax &gt;= Car Age)     and (package.CarCode &lt;= Car Code)     {         PackageList.append(package)     } } return PackageList                     </pre>

จากตารางที่ 2 แสดงกฎที่ใช้กรองแผนความคุ้มครองจากข้อมูลเฉพาะบุคคล ได้แก่ ข้อมูลแผนความคุ้มครอง (Package Data) ราคารถยนต์ (Car Price) อายุรถยนต์ (Car Age) และลักษณะรถยนต์ (Car Code) โดยเงื่อนไขของกฎมีรายละเอียดดังนี้ คือ ทุนประกันภัยของแผนจะต้องไม่เกินราคารถยนต์ของลูกค้า อายุรถยนต์ของลูกค้าอยู่ระหว่างปีที่กำหนดของแต่ละแผน ซึ่งอายุรถยนต์คำนวณจากปีของวันที่คุ้มครองลบด้วยปีผลิตของรุ่นรถยนต์ และลักษณะการใช้งานรถยนต์เป็นรหัสเฉพาะที่ถูกกำหนดตามหลัก พ.ร.บ.ประกันวินาศภัย พ.ศ.2535 เช่น การใช้ส่วนบุคคล (110) การใช้เพื่อการพาณิชย์ (320) เป็นต้น จะต้องตรงกับแผนที่รับประกัน ซึ่งเงื่อนไขทุกข้อจะต้องเป็นจริง ซึ่งแผนความคุ้มครองที่กรองตามกฎเกณฑ์ (PackageList) จะนำไปใช้ในการแนะนำให้กับผู้ทำประกันภัย

ส่วนที่ 2 การแบ่งกลุ่มแบบเคมีน (K-Means Clustering) เพื่อช่วยในการแนะนำแผนความคุ้มครอง

สำหรับลูกค้า โดยใช้ข้อมูลกรรมธรรม์ และข้อมูลรถยนต์ที่รวบรวมในการทำวิจัย ได้แก่ ขนาดเครื่องยนต์ (Engine Size) จำนวนที่นั่ง (Seat) น้ำหนักรถยนต์ (Weight) อายุรถยนต์ (Car Age) ราคารถยนต์ (Car Price) ประเภทรถยนต์ (Car Type) อายุผู้ทำประกันภัย (Insured Age) เพศ (Gender) และลักษณะรถยนต์ (Car Code) ซึ่งในการสร้างแบบจำลองเคมีนประยุกต์ใช้ไลบรารี KMeans Cluster ของ SciKit Learn และทำการปรับแบบจำลองให้อยู่ในรูปแบบที่เหมาะสม ด้วย Elbow Method โดยหาผลรวมของระยะห่างระหว่างกลุ่มกับเซนทรอยด์ที่เรียกว่า Within Cluster Sum of Squares (WCSS) ในการหาจำนวนกลุ่ม (Cluster K) ซึ่งแสดงตัวอย่างได้ดังตารางที่ 3

ตารางที่ 3: การแบ่งกลุ่มข้อมูลโดยใช้วิธีการแบบเคมีน

K-Means Clustering	
<b>Feature:</b>	Engine Size, Vehicle Seat, Vehicle Weight, Car Age, Car Price, Car Type, Insured Age, Gender, Car Code
<b>Optimize</b>	Elbow method (return Cluster K)
<b>Output:</b>	Package Clustering Model (n_clusters = cluster K)

จากตารางที่ 3 จะได้แบบจำลองการแบ่งกลุ่มเพื่อนำไปใช้เพื่อแนะนำแผนความคุ้มครองให้ผู้ทำประกันภัย จากนั้นนำข้อมูลที่ได้จากการสร้างรายการกรรมธรรม์นำเข้ากฎเฉพาะบุคคล และปรับใช้ด้วยแบบจำลองของเคมีนที่สร้างไว้ก่อนหน้านี้เพื่อแนะนำแผนความคุ้มครองที่คิดว่าผู้ทำประกันสนใจ และพัฒนาเอพีไอ (Application Programming Interface: API) สำหรับรับส่งข้อมูลกรรมธรรม์ระหว่างเว็บแอปพลิเคชัน และฐานข้อมูลด้วยภาษาไพทอนในรูปแบบของ RESTful API

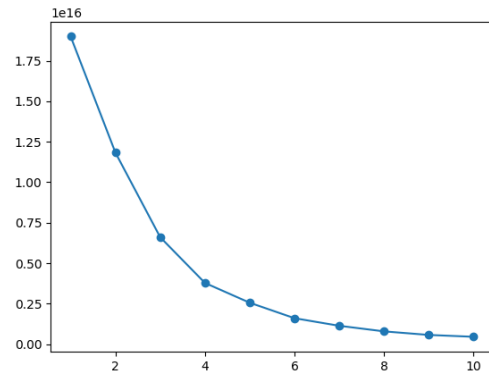
### 3.5 การทดสอบประสิทธิภาพของระบบ

สำหรับการพัฒนาระบบใน Phase แรก คณะผู้วิจัยได้ทำการทดสอบประสิทธิภาพระบบแนะนำแผนความคุ้มครองประกันภัยรถยนต์ภาคสมัครใจโดยวัดความแม่นยำในการแนะนำด้วยค่า Elbow Method

## 4. ผลการดำเนินงานวิจัย

### 4.1 ผลการทดสอบประสิทธิภาพระบบแนะนำการรับประกันภัยรถยนต์

จากการทดสอบประสิทธิภาพในการหาจำนวนกลุ่มที่เหมาะสมในการแบ่งกลุ่มเคมีนด้วย Elbow Method โดยทำการทดสอบการแบ่งกลุ่มเคมีนเริ่มจาก 1 และเพิ่มขึ้นตามลำดับ พบว่าผลลัพธ์ค่าระยะห่างระหว่างกลุ่มกับเซนทรอยด์ที่ลดลงต่อเนื่องจนถึง 10 กลุ่ม ซึ่งกลุ่มที่ 8 ถึง 10 มีค่าระยะห่างที่ไม่ส่งผลต่อการแบ่งกลุ่มแบบเคมีนและนำผลลัพธ์มาสร้างแผนภูมิ Elbow Method ซึ่งสามารถแสดงผลลัพธ์ได้ดังภาพที่ 2



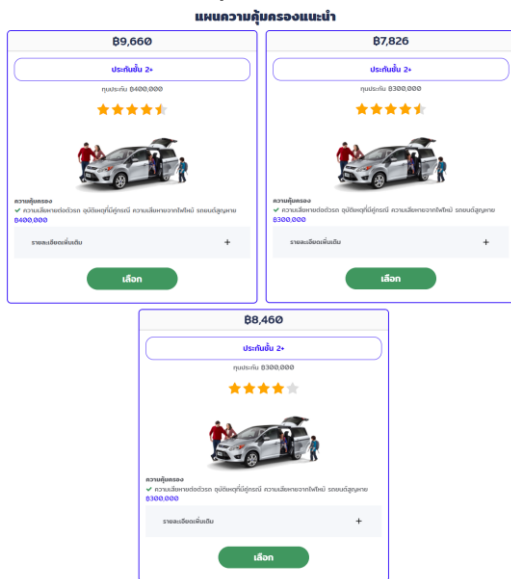
ภาพที่ 2: แผนภูมิการวัดประสิทธิภาพด้วย Elbow Method

จากการทดสอบหาจำนวนกลุ่มด้วย Elbow Method เพื่อพิจารณาความเหมาะสมของการแบ่งกลุ่ม โดยได้ทดสอบแบบจำลองที่สร้างขึ้นจากการแบ่งจำนวนกลุ่มตั้งแต่ 2 ถึง 5 กลุ่ม คณะผู้วิจัยได้เลือกการแบ่งกลุ่มที่เหมาะสมเป็น 3 กลุ่ม เนื่องจากมีความแม่นยำสูงสุด โดยแต่ละกลุ่มมีคุณลักษณะของข้อมูลที่แตกต่างกัน ได้แก่ กลุ่มที่ 1 ข้อมูลรถยนต์ที่มีอายุระหว่าง 3 – 7 ปี และราคาของรถยนต์ไม่เกิน 1,000,000 บาท กลุ่มที่ 2 ข้อมูลรถยนต์ที่มีอายุระหว่าง 1 – 5 ปี และราคาของรถยนต์ 1,000,000 บาทขึ้นไป และ กลุ่มที่ 3 ข้อมูลรถยนต์ที่มีอายุระหว่าง 7 ปีขึ้นไป และราคาของรถยนต์ไม่เกิน 500,000 บาท เป็นต้น หลังจากนั้นจะได้แบบจำลองแบ่งกลุ่มแบบเคมีนเพื่อใช้ในการแบ่งกลุ่มลูกค้าที่มีความคล้ายกัน ซึ่งใช้ในการแนะนำแผนความคุ้มครองร่วมกับกฎเฉพาะบุคคลที่สร้างไว้ก่อนหน้านี้



#### 4.2 ผลการพัฒนาระบบแนะนำการทำประกันภัยรถยนต์

จากการพัฒนาระบบแนะนำการทำประกันภัยรถยนต์ซึ่งผู้ทำประกันภัยจะต้องกรอกข้อมูลที่จำเป็นในการทำกรมธรรม์ต่าง ๆ ได้แก่ ยี่ห้อ รุ่น รุ่นย่อย ปีรุ่น เพศ สถานะวันเกิด จำนวนเคลมใน 12 เดือน จังหวัดจดทะเบียน วันที่เริ่มความคุ้มครองกรมธรรม์ ซึ่งระบบจะนำข้อมูลที่ได้มาประมวลผลผ่านกฎเฉพาะบุคคล ซึ่งจะได้แผนความคุ้มครองที่ถูกต้องในการนำเสนอลูกค้า และประยุกต์ใช้การแบ่งกลุ่มแบบเคมีน (K-Means Clustering) เพื่อช่วยในการแนะนำแผนความคุ้มครองที่คาดว่าผู้ทำประกันภัยสนใจ โดยสามารถแสดงตัวอย่างหน้าจอแผนความคุ้มครองที่แนะนำให้กับลูกค้า ได้ดังภาพที่ 3



ภาพที่ 3: ตัวอย่างหน้าจอแผนความคุ้มครองแนะนำ

#### 5. สรุป

จากการพัฒนาระบบแนะนำการทำประกันภัยรถยนต์โดยประยุกต์ใช้การสร้างกฎเฉพาะบุคคลร่วมกับการแบ่งกลุ่มแบบเคมีน ซึ่งรองรับรถกระบะ และรถเก๋งจำนวน 10 ยี่ห้อ จากแผนความคุ้มครองประกันภัยรถยนต์ 5 อันดับแรก และสามารถแนะนำประกันภัยรถยนต์ภาคสมัครใจ 4 ประเภท ได้แก่ ประกันประเภทที่ 1 ประกันประเภทที่ 2+ ประกันประเภทที่ 3+ และประกันประเภทที่ 3 จากผลการทดสอบประสิทธิภาพระบบสารสนเทศ

พบว่า สามารถให้คำแนะนำการทำประกันภัยรถยนต์เฉพาะบุคคลได้อย่างมีประสิทธิภาพ ลดขั้นตอนและระยะเวลาการทำประกันภัยรถยนต์ และง่ายต่อการตัดสินใจเลือกแผนความคุ้มครองของลูกค้าที่ต้องการทำประกันภัย ซึ่งงานวิจัยในอนาคตจะนำสัญญาอัจฉริยะ (Smart Contract) เพื่อช่วยในการตรวจสอบ อนุมัติการทำกรมธรรม์อัตโนมัติ และป้องกันการฉ้อโกงที่อาจเกิดขึ้น

#### เอกสารอ้างอิง

- [1] Spiteri, M. and Azzopardi, G. (2018). “Customer Churn Prediction for a Motor Insurance Company,” *Thirteenth International Conference on Digital Information Management (ICDIM 2018)*, pp. 173-178.
- [2] Dhieb, N., et al., (2020). “A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement,” *IEEE Access*, Vol.8, pp.58546-58558.
- [3] Jayanthi, N., et al., (2019). “ Vehicle Insurance Calculator Using Augmented Reality,” *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pp. 766-770.
- [4] C K, R., and Srikantaiah, K. C. (2021) “Similarity Based Collaborative Filtering Model for Movie Recommendation Systems,” *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1143-1147.
- [5] ธนพล ทุกสิ่ง สุชา สมานชาติ และสุนันชา สดสี. (2558). “ระบบผู้แนะนำโดยอาศัยความไว้วางใจร่วมกับผู้เชี่ยวชาญ” *วารสารวิทยาศาสตร์และเทคโนโลยี*. ปีที่ 23 ฉบับที่ 3 กรกฎาคม-กันยายน. หน้า 517-537.
- [6] Tran, T. T., et al., (2018). “Recommender System with Artificial Intelligence for Fitness Assistance System.” *15th International Conference on Ubiquitous Robots (UR)*, pp. 489-492.
- [7] Kongsakun, K., Kajornrit, J., and Fung C.C. (2012). “Neural Network Modeling for an Intelligent Recommendation System Supporting SRM for Universities in Thailand,” *Applied Science and Engineering Progress*, 5(3), pp. 67-75.

# การประเมินความเสี่ยงโรคภาวะถุงน้ำในรังไข่หลายใบ

## Assessing the Risk of Polycystic Ovary Syndrome

กมลฉัตร ดีห์ล่า (Kamonchat Deela)<sup>1</sup> ปวีตรา สาริวงษ์ (Pawitra Sariwong)<sup>2</sup>

และธนพล ตั้งชูพงศ์ (Thanaphon Tangchoopong)<sup>3</sup>

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

<sup>1</sup>kamonchat@kkumail.com, <sup>2</sup>pawitra.sa@kkumail.com, <sup>3</sup>thanaphon@kku.ac.th

### บทคัดย่อ

โรคภาวะถุงน้ำในรังไข่หลายใบ (PCOS) เป็นโรคที่สามารถเกิดขึ้นได้กับสตรีในวัยเจริญพันธุ์ ทางผู้ศึกษาจึงสนใจพัฒนาตัวแบบพยากรณ์ การเกิดโรคถุงน้ำในรังไข่หลายใบ สำหรับบุคคลทั่วไป ที่ไม่จำเป็นต้องมีผลตรวจจากห้องทดลอง โดยได้ศึกษาอาการและเลือกใช้อาการที่ทำให้มีโอกาสในการเกิดโรคนี้อันสูง จากข้อมูลโรค PCOS จากฐานข้อมูลเปิดเค็กเกิล [9] โดยมีทั้งข้อมูลทางกายภาพ และ ข้อมูลผลตรวจจากห้องทดลอง โดยอาศัยการคัดเลือกคุณลักษณะสำคัญที่สุด 11 ลำดับ (Feature selection) จากเกณฑ์การจำแนกของวิธีการป่าสุ่ม (Random forest) ที่สามารถสังเกตและ ตรวจสอบเองได้เบื้องต้นมาใช้ทดสอบ จากผลการดำเนินงานโมเดลที่เลือกใช้คุณลักษณะที่สามารถสังเกตอาการเองได้ มีประสิทธิภาพจาก พื้นที่ใต้เส้นโค้ง (AUC) ไม่แตกต่างจาก โมเดลที่ใช้ข้อมูลคุณลักษณะทั้งหมด มีค่าเท่ากับคือ 0.97 ซึ่งจากคุณลักษณะที่เลือกมา ได้นำมาฝึกสอนอีก 6 โมเดลได้แก่ วิธีต้นไม้แห่งการตัดสินใจ (Decision tree) วิธีเนออีฟเบย์ (Naive Bayes) วิธีการถดถอยโลจิสติก (Logistic Regression) วิธีรองรับเวกเตอร์ลักษณะนาม (Support Vector Classifiers : SVC) วิธีเพื่อนบ้านที่ใกล้ที่สุด (K Nearest Neighbor : KNN) และวิธีแคทบูส (CatBoost Classifier) เพื่อหาโมเดลที่มีประสิทธิภาพที่มีค่าคะแนนทดสอบ มากกว่าค่าคะแนนฝึกสอน สรุปได้ว่า มี 3 โมเดล คือ โมเดลป่าสุ่ม โมเดลการถดถอยโลจิสติก และ โมเดลรองรับเวกเตอร์ลักษณะนาม

คำสำคัญ: การประเมินความเสี่ยง โรคถุงน้ำในรังไข่

### Abstract

Polycystic ovary syndrome (PCOS) is a disease that can occur in women of childbearing age. As a result, the researcher is interested in developing a self-assessment predictive model for PCOS that does not require laboratory data by examining the disease's symptoms and selecting those that indicate a high risk of acquiring the disease. The Kaggle Database of PCOS [9] has both physical and laboratory data. By selecting the 11 most important self-observable features from the random forest method. It is found that the Random Forest model before and after selecting the most important self-observable from Area under the curve (AUC) is the same value of 0.97. The selected features were used to create six other models: Decision Tree, Nave Bayes, Logistic Regression, Support Vector Classifiers (SVC), K-Nearest Neighbor (KNN), and CatBoost Classifier to find efficient models with test scores higher than the training score, it was concluded that there were three models, the Random Forest, the Logistic Regression and Support Vector Classifiers (SVC).

**Keywords:** Risk assessment, PCOS, Polycystic ovary syndrome.

### 1. บทนำ

ในปัจจุบันอัตราผู้ป่วยในโรคภาวะถุงน้ำในรังไข่หลายใบ (Polycystic ovary syndrome : PCOS) ในสตรีมีเพิ่มมากขึ้น อีกทั้งยังไม่มีใครรู้จักโรคนี้นักมากนัก จึงยังไม่รู้

ว่าตนเองมีโอกาสเสี่ยงที่จะเป็นโรคหรือไม่ โดยสาเหตุหลักของการเกิดโรค PCOS นั้นคือ การมีฮอร์โมนเพศชายสูง ดังแสดงอาการภายนอกได้จากการมีขนขึ้นมากกว่าปกติตามจุดต่าง ๆ ของร่างกาย และการมีประจำเดือนมากกว่าปกติหรือน้อยกว่าปกติ เพื่อที่จะให้บุคคลธรรมดาสามารถตรวจสอบอาการและความเสี่ยงในการเกิดโรคของตนเองได้ ทางผู้ศึกษาจึงสังเกตเห็นว่าการที่จะศึกษาคุณลักษณะสำคัญในการเกิดโรค และจำแนกออกมาเปรียบเทียบระหว่างข้อมูลการเกิดโรคที่มีอยู่ว่าจะสามารถทำนายความเสี่ยงจากการเลือกคุณลักษณะสำคัญที่บุคคลธรรมดาสามารถตรวจสอบเองได้หรือไม่ แล้วทดสอบว่าโมเดลใดจะมีประสิทธิภาพในการทำนายมากที่สุด

## 2. งานวิจัยและทฤษฎีที่เกี่ยวข้อง

### 2.1 ความเสี่ยงของโรคถุงน้ำรังไข่

โรคภาวะถุงน้ำในรังไข่หลายใบ ที่มีอัตราผู้ป่วยเพิ่มขึ้นเรื่อย ๆ โดยบางคนจะไม่สามารถทราบเลยว่าตนเองเป็นโรคนี้ และหลายคนขาดความรู้เกี่ยวกับโรค PCOS ทำให้ไม่ทราบว่าตนเอง มีโอกาสเสี่ยงที่จะเกิดโรคหรือไม่ สาเหตุที่เกิดโรคจะมีอยู่หลายปัจจัย ซึ่งโรคนี้จะก่อให้เกิดโรคแทรกซ้อนตามมา ปัจจุบันได้มีวิจัยออกมามากมาย จึงสามารถนำมาเป็นขั้นตอนในการประเมินความเสี่ยง

#### 2.2.1 การประเมินความเสี่ยงโรคถุงน้ำรังไข่

การประเมินความเสี่ยงของโรคถุงน้ำรังไข่ โรคถุงน้ำรังไข่ โรคถุงน้ำรังไข่ เกิดจากความผิดปกติของต่อมไร้ท่อที่หยุดชะงัก และทำให้เอสโตรเจนที่มากเกินไป และทำให้ผู้หญิงส่วนใหญ่มีอินซูลิน ที่ผิดปกติ โรคถุงน้ำในรังไข่ ส่งผลให้เกิดภาวะแทรกซ้อนทางสุขภาพหลายอย่าง เช่น ความผิดปกติของประจำเดือน ภาวะมีบุตรยาก ขนดก สิวขึ้น โรคอ้วน และโรคเมตาบอลิก ผู้หญิงที่เป็นโรคนี้จะพบกับความเสี่ยงที่เพิ่มขึ้นในการเป็นโรคเบาหวาน โรคหัวใจ และหลอดเลือดความเสี่ยงการเผาผลาญที่ผิดปกติที่อาจจะเป็น Hyperandrogenism, ความผิดปกติของแอนโดรเจน, ความผิดปกติของการสร้างรูขุมขน, ความผิดปกติของ Gonadotropin, เมื่อทราบสาเหตุของโรคก็จะสามารถให้ผู้ป่วยได้ไปตรวจร่างกายเพื่อทำการรักษา แต่เมื่อผู้ป่วยเป็น

โรคนี้ในระยะยาวแล้วจะรักษาไม่หายขาด จะต้องปรับเปลี่ยนวิถีชีวิต จากเดิมให้ดีขึ้นเพื่อลดความเสี่ยงที่จะเกิดโรคแทรกซ้อนเช่น การดูแลอาหารการกิน ควบคุมน้ำหนักให้อยู่ในเกณฑ์มาตรฐาน ซึ่งโรคนี้อาจจะทำให้เกิดขนดก มีสิว จึงต้องใช้ยาคุมกำเนิดเพื่อรักษา ประจำเดือนมาไม่ปกติ การใช้ยาคุมเพื่อให้รอบประจำเดือนมาตรงมากขึ้น และปรึกษา หรือตรวจกับทางแพทย์ผู้เชี่ยวชาญ [1,2,3]

### 2.2 การเลือกคุณลักษณะ

การเลือกคุณลักษณะ จะอิงจากอาการและสาเหตุของโรค ซึ่งสาเหตุของโรคยังบอกไม่ได้แน่ชัดจึงต้องสังเกตอาการที่มี และอาการส่วนใหญ่เกิดขึ้นได้ในวัยเจริญพันธุ์

- รอบระดูผิดปกติ (Menstrual irregularity): คือมีการขาดระดู (secondary amenorrhea) หรือมีระดูนาน ๆ ครั้ง (oligomenorrhea) คือ มีรอบระดูห่างกันมากกว่า 35 วัน

- อาการแสดงของภาวะแอนโดรเจนเกิน (hyperandrogenism): ได้แก่ ภาวะขนดก (hirsutism) มีสิวจนขึ้น หน้ามัน และมีหัวล้านแบบเพศชาย (androgenic alopecia)

- ภาวะอ้วน (obesity): โดยจะมีลักษณะอ้วนแบบกลางลำตัว (central obesity) หรือเรียกอีกอย่างหนึ่งว่า android obesity

- ภาวะมีบุตรยาก (infertility): ซึ่งสาเหตุหลักเกิดจากการที่ไข่ตกลดลงหรือมีภาวะไข่ไม่ตกเรื้อรัง (Chronic anovulation)

- ภาวะดื้อต่ออินซูลินและมีอินซูลินในเลือดสูง (insulin resistant and hyperinsulinemia): มีผลทำให้ผิวหนังหนาตัว และผิวคล้ำดำ [4]

### 2.3 เหมืองข้อมูล (Datamining)

เทคนิคการจำแนกประเภทข้อมูล เป็นเทคนิคที่สำคัญในการสืบค้นความรู้ บนฐานข้อมูลขนาดใหญ่ ข้อมูลที่นำมาจำแนกอาจเป็นข้อมูล ที่มีโครงสร้างหรือไม่มีโครงสร้างก็ได้ การจำแนกข้อความเพื่อทำการค้นหา องค์ความรู้ ทำให้ได้ข้อมูลที่ง่ายต่อการสืบค้นต่อไป

- เทคนิคเนออีฟ เบย์ส์ (Naive Bayes) คือ โมเดลการคัดแยกประเภทข้อมูล ที่ใช้หลักการความน่าจะเป็นซึ่งอยู่บนพื้นฐานและสมมติฐาน ที่ให้การเกิดของเหตุการณ์ต่าง ๆ เป็นอิสระต่อกัน

- เทคนิควิธีการเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) เป็นอัลกอริทึมที่อยู่ในกลุ่มของข้อมูลที่มีฉลากหรือป้ายกำกับ (Label) คอยบอกว่าคือข้อมูลอะไร จึงเป็นการจัดกลุ่มข้อมูลที่อยู่ใกล้กันเข้าไว้ด้วยกัน โดยมีจำนวนข้อมูลที่จะนำมาพิจารณา

- เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เป็นโครงสร้างข้อมูลที่อยู่ในรูปแบบลำดับชั้น ใช้ช่วยในการตัดสินใจ โดยจะมีรูปแบบเหมือนกับต้นไม้ที่กลับหัวอยู่ จะมีคุณลักษณะ (Attribute) ซึ่งเป็นตัวทดสอบที่อยู่ภายในโหนด (Node) แต่ละตัว และมีกิ่งของต้นไม้ (Branch) ที่แสดงถึงค่าความเป็นไปได้ของคุณลักษณะที่ถูกเลือกใช้ในการทดสอบ และมีใบ (Leaf) ซึ่งเป็นสิ่งที่อยู่ในส่วนล่างสุดของต้นไม้ให้เห็นถึงกลุ่มของข้อมูล (Class) หรือก็คือผลลัพธ์ที่ได้จากการทำนาย

- เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine หรือ SVM) เป็นอัลกอริทึม ในรูปแบบการคัดแยกในการประมวลผลเป็นภาพดิจิทัล มีหลักการคือ การให้ข้อมูลนำเข้าใช้ฝึก เป็นเวกเตอร์ในพื้นที่ N มิติ เช่นถ้าในกรณีของ 2 มิติ และ 3 มิติ จะเป็นจุดที่อยู่ในระนาบ x, y และพื้นที่ x, y, z ตามลำดับ จากนั้นทำการสร้างไฮเปอร์เพลน (Hyperplane) ที่จะแยกกลุ่มของเวกเตอร์อื่นๆ ออกเป็นประเภทต่าง ๆ ในกรณีที่ เป็น 2 มิติ และ 3 มิติ ไฮเปอร์-เพลน คือเส้นตรงและระนาบตามลำดับ ข้อเด่นของ SVM จะทำการเก็บแผนที่ (Map) เวกเตอร์ในพื้นที่ให้นำเข้าให้ผู้ใช้ฟีเจอร์สเปซ (Feature Space) โดยใช้ฟังก์ชันหรือเรียกว่าเคอร์เนล (Kernel) ชนิดต่างๆ [5]

- เทคนิคการถดถอยโลจิสติกการวิเคราะห์การถดถอย- โลจิสติก (logistic regression analysis) เป็นเทคนิคการวิเคราะห์สถิติเชิงคุณภาพ (qualitative statistical techniques) ที่แตกต่างไปจากเทคนิคการวิเคราะห์เชิงปริมาณ (quantitative techniques) อย่างน้อยก็เรื่องของ ข้อมูลที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพ ซึ่งก็คือ เป็นตัวแปรเชิงกลุ่มนั่นเอง การวิเคราะห์การถดถอยโลจิสติกแบ่งเป็น 2 ประเภท คือ (1) การวิเคราะห์การถดถอยโลจิสติกทวิ (binary logistic regression analysis) และ (2) การวิเคราะห์การถดถอยโลจิสติกพหุกลุ่ม (multinomial lo-

gistic regression analysis) การวิเคราะห์การถดถอยโลจิสติกทั้ง 2 ประเภท แตกต่างกันในด้านตัวแปร ตามโดยที่การวิเคราะห์การถดถอยโลจิสติกทวิ ใช้กับตัวแปรตามที่แบ่งออกเป็น 2 กลุ่มย่อย (dichotomous variable) มี 2 ค่าคือมีค่าเป็น 0 กับ 1 เช่น กลุ่มที่มีเหตุการณ์กับกลุ่มที่ไม่มีเหตุการณ์ ส่วนการวิเคราะห์โลจิสติกแบบพหุกลุ่มใช้กับ ตัวแปรตามที่มีหลายค่ามากกว่า 2 กลุ่ม (polytomous variable) [6]

- เทคนิคป่าสุ่ม (Random Forest Model) ตัวแบบป่าสุ่ม พัฒนาจากวิธีการแบ็กกิ้ง (bagging) ซึ่งเป็นการสร้างต้นไม้ตัดสินใจหลาย ๆ ต้น โดยใช้ชุดตัวอย่างซ้ำด้วยหลักการบูตสแตรป (bootstrap) จากข้อมูลชุดฝึกฝนเพียงชุดเดียว โดยการสร้างต้นไม้ ใช้หลักการแบ่งต้นไม้ด้วยตัวอย่างสุ่มขนาด m จากตัวแปรอิสระทั้งหมด p ตัว (โดยทั่วไปจะให้ m เท่ากับรากที่สองของ p) เพื่อเป็นเกณฑ์ในการแบ่งต้นไม้ ซึ่งเมื่อนำต้นไม้ตัดสินใจมาทำนายร่วมกัน โดยใช้ค่าเฉลี่ยร่วมจากผลการทำนายจากต้นไม้แต่ละต้น ในกรณีที่ตัวแปรตอบสนองเป็นข้อมูลเชิงปริมาณ หรือใช้ฐานนิยมจากผลการทำนายจากต้นไม้แต่ละต้น ในกรณีที่ตัวแปรตอบสนองเป็นข้อมูลเชิงคุณภาพ ทำให้ได้ตัวแบบที่มีความแม่นยำมากกว่าการทำนายโดยใช้ต้นไม้ตัดสินใจเพียงต้นเดียว [7]

- เทคนิคทฤษฎีการเรียนรู้เชิงลึกเช่น Google's Tensor-Flow และ Apple's Core ML ได้ซึ่งอัลกอริทึม CatBoost ได้เพิ่มประสิทธิภาพการทำงานจาก GBM โดยสามารถจัดการกับตัวแปรโดยอัตโนมัติ และไม่ต้องมีการแปลงชุดข้อมูลให้เป็นรูปแบบเฉพาะนอกจากนี้ CatBoost ยังสามารถจัดการกับตัวแปร และค่าข้อมูลบางส่วนของที่หายไป (Missing Values) ได้อย่างมีประสิทธิภาพอีกด้วย [8]

### 3. วิธีการดำเนินงาน

การศึกษางานวิจัยโรคภาวะถุงน้ำในรังไข่หลายใบในสตรีผู้ศึกษาได้ทำการศึกษาค้นคว้า และรวบรวมข้อมูลผู้ที่เกี่ยวข้องไปตรวจสอบสุขภาพและ ศึกษาการวิเคราะห์ข้อมูลประเมินข้อมูล และทำนายอาการจากวิธีการต่าง ๆ ตลอด - จนวิธีการประยุกต์ใช้ในการดำเนินงานวิจัย และวิธีการดำเนินงาน โดยมีรายละเอียด ดังนี้

### 3.1 หาชุดข้อมูลของผู้ป่วยโรค PCOS

ซึ่งเป็นชุดข้อมูล (Dataset) ของสตรี ช่วงอายุ 20 - 50 ปี จำนวน 540 ราย ที่เข้ามาตรวจสุขภาพภายในโรงพยาบาล 10 แห่ง ในประเทศอินเดีย โดยที่ชุดข้อมูลอ้างอิงและคุณลักษณะทั้งหมด 32 คุณลักษณะ สามารถอ้างอิงได้จากเว็บไซต์ Kaggle [9]

### 3.2 ทำการจัดการชุดข้อมูลที่ได้มา

โดยใช้วิธีการเตรียมข้อมูล (Data Preparation) โดยทำการนำข้อมูลที่ซ้ำกันออก และข้อมูลที่เกี่ยวข้องกับการเกิดโรคมาใช้ และทำการสำรวจข้อมูลโดยวิธีการทางวิทยาการข้อมูล โดยใช้วิธีการคัดเลือกคุณลักษณะสำคัญ จากอัลกอริทึมป่าสุ่ม แล้วคัดเลือกคุณลักษณะที่สามารถสังเกตและตรวจสอบเองได้เบื้องต้นมาใช้ทดสอบ

### 3.3 พัฒนาโมเดล

ใช้วิธีการแบ่งข้อมูลออกเป็น 2 ส่วนคือ ส่วนของชุดข้อมูลฝึกสอน 70% และส่วนของข้อมูลทดสอบ 30% ทำการพัฒนาโมเดลป่าสุ่ม (Random forest), ต้นไม้แห่งการตัดสินใจ (Decision tree), เนออีฟ เบย์ (Naïve Bayes), การถดถอยโลจิสติก (Logistic Regression), รองรับเวกเตอร์ลักษณะนาม (Support Vector Classifiers : SVC), เพื่อนบ้านที่ใกล้ที่สุด (K Nearest Neighbor : KNN) และวิธีแคทบูส (CatBoost Classifier) จากคุณลักษณะที่ได้คัดเลือกไว้ เพื่อหาค่าคะแนนพื้นที่ใต้เส้นโค้งมาประเมินประสิทธิภาพความแม่นยำของแต่ละโมเดล

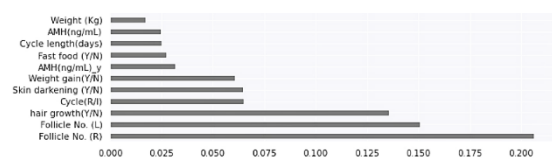
### 3.4 การประเมินประสิทธิภาพ

ทำการประเมินประสิทธิภาพ ของโมเดลว่ามีความแม่นยำและความน่าเชื่อถือได้หรือไม่ โดยวัดจากค่าคะแนนการฝึกสอน (Train) และคะแนนการทดสอบ (Test) ของคะแนนพื้นที่ใต้เส้นโค้ง (Area Under the Curve : AUC) ของกราฟลักษณะการทำงานของเครื่องรับ (Receiver Operating Characteristics : ROC Curve) นำคะแนนที่ได้ที่เปรียบเทียบกัน และเลือกโมเดลที่วัดประสิทธิภาพได้ดีที่สุดมาใช้งาน

## 4. ผลการดำเนินงาน

### 4.1 สำรวจคุณลักษณะที่สำคัญในการเกิดโรค

ทำการสำรวจคุณลักษณะ ที่สำคัญของสตรีที่เข้ามาตรวจสุขภาพแล้วเป็นโรค PCOS แล้วทำการประเมินความเสี่ยงโดยใช้วิธีการป่าสุ่ม (random forest) เพื่อทดสอบความแม่นยำ ในการทำนายความน่าเชื่อถือจากการสำรวจคุณลักษณะสำคัญ (feature importance) จากภาพที่ 1 จะเห็นได้ว่าอาการที่ผู้ป่วยและผู้ที่มีความเสี่ยงเป็นโรค PCOS ส่วนใหญ่มีอาการสำคัญ มีรูขุมขนด้านซ้ายและด้านขวากว้าง มีปริมาณการเกิดขนเพิ่มขึ้นมากกว่าปกติ มีสีผิวที่คล้ำขึ้น มีน้ำหนักเพิ่มขึ้น มีปริมาณฮอร์โมนเกินต่ำกว่าปกติ



ภาพที่ 1: แสดงถึงค่าความสำคัญของแต่ละคุณลักษณะ จากข้อมูลที่ได้รับ (information gain) ของอัลกอริทึมป่าสุ่ม

### 4.2 เลือกคุณลักษณะสำคัญ

คุณลักษณะสำคัญของสตรีที่เข้ามาตรวจสุขภาพนั้น มีทั้งส่วนที่ไม่เกี่ยวข้องกับการเกิดโรค และต้องไปทำการตรวจสอบอาการกับแพทย์ผู้เชี่ยวชาญที่โรงพยาบาลเท่านั้น ถึงจะทราบได้ว่าตนเองมีโอกาสป่วยเป็นโรค PCOS หรือไม่มีโอกาส ทางผู้ศึกษาจึงได้ทำการเลือกเฉพาะคุณลักษณะสำคัญที่ทำให้เกิดโรคสูง และบุคคลธรรมดาสามารถตรวจสอบสุขภาพของตนเอง ได้โดยไม่ต้องเข้าพบแพทย์ผู้เชี่ยวชาญ

4.2.1 ช่วงอายุของผู้ป่วย (Age): ข้อมูลป่วย จะมีอายุระหว่าง 20 – 48 ปี ซึ่งเป็นช่วงอายุที่จะมีโอกาสเสี่ยงในการเป็นโรค PCOS สูง ซึ่งก็คือในช่วงวัยเจริญพันธุ์ เกือบเป็นตัวเลข เช่น 55, 60

4.2.2, 3 การเพิ่มขึ้นของขน (Follicle No. (L, R)): การเกิดหรือการเติบโตของขน ในแต่ละคนจะแตกต่างกันออกไป ผู้ที่ทำการประเมินจะต้องสังเกตตนเอง ว่ามีขนเพิ่มขึ้นหรือไม่ และเก็บข้อมูลเป็นใช่ (1) หรือไม่ใช่ (0) หากสังเกตเห็นว่า ขนของตนเองขึ้นมากจนผิดปกติให้ตอบใช่

4.2.4 การเพิ่มขึ้นของสีผิว (Skin darkening): ความเข้มของผิวจะเกิดจากภาวะดื้ออินซูลิน จึงต้องสังเกต ผิวว่าสีผิวเข้มและมีการหนาตัวหรือไม่ โดยเก็บข้อมูลเป็นใช่ (1) หรือไม่ใช่ (0)

4.2.5 การเกิดสิว (Pimples): พื้นผิวบนใบหน้าของแต่ละคนมีลักษณะแตกต่างกัน ผู้ที่มาทำการประเมินจะต้องสังเกตตนเอง ว่าสิ่วที่เกิดขึ้นบนใบหน้ามีความผิดปกติหรือไม่ เป็นสิ่วอักเสบหรือผดผื่นทั่วไป เก็บข้อมูลเป็นใช่ (1) หรือไม่ใช่ (0)

4.2.6 การรับประทานอาหารที่มีไขมันสูง (Fast food): ในช่วงระยะเวลาที่มีอาการผิดปกติของร่างกาย หรือในช่วงระยะเวลาที่ทำการตรวจสอบสุขภาพ ผู้ทำการประเมินได้รับประทานอาหารที่มีไขมันสูงหรือไม่ โดยใช้การสังเกตตนเอง โดยเก็บข้อมูลเป็นใช่ (1) หรือไม่ใช่ (0)

4.2.7 ระยะเวลาของวันที่เกิดประจำเดือน (Cycle length): เก็บข้อมูลของวันที่มีรอบเดือนมา เก็บประจำเดือนมากี่วัน โดยให้นับการมาของประจำเดือนว่ามีระยะการมากี่วัน โดยข้อมูลส่วนใหญ่ จะอยู่ที่ 0 - 9 วัน และเก็บข้อมูลเป็นตัวเลข

4.2.8 น้ำหนัก (Weight): เก็บข้อมูล น้ำหนักของผู้ใช้งาน โดยเก็บข้อมูลเป็นตัวเลข มีหน่วยเป็นกิโลกรัม

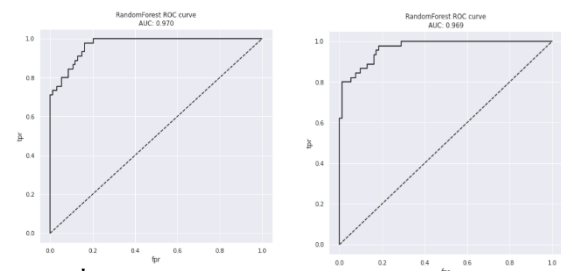
4.2.9 การเพิ่มขึ้นของน้ำหนัก (Weight gain): ให้สังเกตว่าในแต่ละเดือนมีน้ำหนักเพิ่มขึ้น หรือน้อยลงบ่อยจนผิดปกติหรือไม่ โดยเก็บข้อมูลเป็นใช่ (1) หรือไม่ใช่ (0)

4.2.10 การเจริญเติบโตของเส้นผม (hair growth): การเจริญเติบโตของเส้นผม ให้สังเกตว่าผมมีการยาวเร็วขึ้นมากกว่าปกติหรือไม่ โดยเก็บข้อมูลเป็นใช่ (1) หรือไม่ใช่ (0)

4.2.11 ระยะห่างของรอบเดือนในแต่ละเดือน (Cycle): ระยะห่างของการเกิดประจำเดือนในแต่ละครั้งห่างกันกี่วันถึงจะมีประจำเดือนในครั้งถัดไป เก็บข้อมูลเป็นตัวเลข

การเลือกคุณลักษณะ (Feature Selection) ทำการสำรวจคุณลักษณะ ที่สามารถทำให้เกิดโรคจากโมเดล Random forest และทำการหาค่าคะแนนพื้นที่ใต้เส้นโค้ง (AUC) จากโมเดลนี้ จะได้คะแนนในส่วนของคุณลักษณะ

ฝึกสอนเป็น 0.98 ชุดข้อมูลทดสอบเป็น 0.97 ซึ่งจากค่าที่ได้ชุดข้อมูลทดสอบมีค่าคะแนนต่ำกว่าชุดข้อมูลฝึกสอน ผู้ศึกษาจึงได้เลือกเอาเฉพาะข้อมูล ที่เป็นลักษณะอาการที่ทำให้เกิดโรค เลือกจากลักษณะที่สามารถตรวจสอบเองได้ด้วยตนเอง แล้วทำการหาค่าคะแนนพื้นที่ใต้เส้นโค้งจากโมเดล Random forest อีกครั้ง ผลปรากฏว่าค่าของคุณลักษณะฝึกสอนเป็น 0.96 ชุดข้อมูลทดสอบเป็น 0.97 ดังตารางที่ 1 ซึ่งเป็นตารางการเปรียบเทียบค่าของพื้นที่ใต้เส้นโค้งก่อนการเลือกคุณลักษณะ และค่าหลังจากเลือกคุณลักษณะแล้ว คุณภาพของค่าที่ได้หลังจากเลือกคุณลักษณะไม่ต่างจากค่าที่ยังไม่ได้เลือกคุณลักษณะมากนัก และยังเป็นคุณลักษณะที่สามารถตรวจสอบเองได้โดยไม่ต้องใช้ผลจากห้องทดลอง



ภาพที่ 2: กราฟพื้นที่ใต้เส้นโค้ง

โค้งของโมเดล Random forest ในช่วงก่อนเลือกคุณลักษณะ (ซ้าย) และเลือกคุณลักษณะแล้ว (ขวา)

ตารางที่ 1: ค่าคะแนนพื้นที่ใต้เส้นโค้งของโมเดล Random forest ในช่วงก่อนเลือกคุณลักษณะ และเลือกคุณลักษณะแล้ว

Model	ก่อนทำ		หลังทำ	
	Feature importance		Feature importance	
	Training score	Test score	Training score	Test score
Random Forest	0.98	0.97	0.96	0.97

### 4.3 ประเมินความเสี่ยงของโรค PCOS

4.3.1 เลือกโมเดลที่ใช้ในการประเมิน: ใช้โมเดลในการมาประเมินความเป็นไปได้ ว่ามีโอกาสในการเกิดโรคหรือไม่และตรวจสอบความแม่นยำของโมเดลว่ามีความน่าเชื่อถือได้หรือไม่

4.3.2 เปรียบเทียบประสิทธิภาพของโมเดลจากการ : คัดเลือกคุณลักษณะสำคัญแล้วนำค่าคะแนนพื้นที่ใต้เส้นโค้งของแต่ละโมเดลมาเปรียบเทียบกัน สรุปได้ว่าโมเดล

ป่าสุ่ม โมเดลการถดถอยโลจิสติก และ โมเดลรองรับ  
 เวกเตอร์ลักษณะนาม เป็น โมเดลที่มีค่าชุดข้อมูลทดสอบ  
 มากกว่าค่าชุดข้อมูลฝึกสอน

**ตารางที่ 2:** ค่าคะแนนพื้นที่ใต้เส้นโค้งของแต่ละโมเดล หลังจาก  
 คัดเลือกคุณลักษณะสำคัญแล้ว

Model	Training score	Test score
Random Forest	0.96	0.97
Decision Tree	1.00	0.78
Naïve bayes	0.95	0.95
Logistic Regression	0.95	0.97
Support Vector Classifiers	0.89	0.95
K Nearest Neighbor	0.94	0.87
CatBoost Classifier	1.00	0.96

### 5. สรุปผล

งานวิจัยนี้ผู้ศึกษา ได้ศึกษาและเลือกคุณลักษณะสำคัญ  
 ที่เป็นสาเหตุของโรคภาวะถุงน้ำในรังไข่หลายใบ 11  
 คุณลักษณะที่สามารถ ให้บุคคลธรรมดาตรวจสอบด้วย  
 ตนเองได้ โดยไม่จำเป็นต้องรอผลจากห้องทดลอง โดย  
 คุณลักษณะสำคัญที่ได้เลือกมาดังนี้ รุขุมขนฝั่งซ้าย รุขุมขน  
 ฝั่งขวา การเจริญเติบโตของเส้นผม ผิวคล้ำ น้ำหนักที่  
 เพิ่มขึ้น อาหารที่มีไขมันสูง ระยะห่างของรอบเดือนในแต่ละ  
 เดือน ระยะเวลาของรอบเดือน น้ำหนัก อายุ และ สิว ซึ่ง  
 โมเดลที่ได้ทำการฝึกสอนด้วยคุณลักษณะดังกล่าว  
 มีประสิทธิภาพทัดเทียม กับโมเดลที่ใช้ข้อมูลฝึกสอนที่มี  
 ผลตรวจจากห้องทดลอง ซึ่งโมเดลที่เหมาะสมนำไปใช้  
 งานคือ โมเดลป่าสุ่ม โมเดลการถดถอยโลจิสติก และโมเดล  
 รองรับเวกเตอร์ลักษณะนาม แนวทางในการดำเนินงาน  
 ขึ้นต่อไป ควรเก็บข้อมูลลักษณะอาการของผู้ป่วย PCOS  
 ภายในประเทศไทย เพื่อที่จะสามารถนำมาสร้างโมเดลที่  
 เหมาะสม สำหรับการประเมินโรคภาวะถุงน้ำในรังไข่  
 หลายใบของประชากรในประเทศไทย โดยใช้หลักการ  
 คัดเลือกคุณลักษณะตามรูปแบบที่นำเสนอมาข้างต้น

### เอกสารอ้างอิง

[1] R. J Norman, D. Dewailly, R. S Legro, T. E. Hickey.  
 "Polycystic ovary syndrome." *Lancet LondEngl*, Vol.  
 370, No. 9588, August, 2007.

[2] N. Afiqah-Aleng, S. Harun, M. R. A. A-Rahman, N.  
 A. N. Muhammad, Z. A. Mohamed-Hussein.  
 "PCOSBase: a manually curated database of  
 polycystic ovarian syndrome." Available Online at  
<https://doi.org/10.1093/database/bax098>, accessed on  
 January 2021.

[3] J. Avery, S. Ottey, R. Morman, M. Cree-Green,  
 M. Gibson-Helm. "Polycystic ovary syndrome  
 support groups and their role in awareness, advocacy  
 and peer support: A systematic search and narrative  
 review." *Current Opinion in Endocrine and Metabolic  
 Research*, Vol. 12, No. 7, pp. 98-104, June, 2020.

[4] อรสา เหมะจันทร์. สุเมธ พัฒนาสุทธินนท์. "ภาวะถุงน้ำ  
 จำนวนมากที่รังไข่." *วารสารการแพทย์และวิทยาศาสตร์  
 สุขภาพ*, ปีที่ 14, ฉบับที่ 1, 2550.

[5] ประพัฒน์ พรมน้ำอ่าง. วสุวรรณ์ พงศ์จักร. นิเวศ จิระวิจิ  
 ชัย. "Text Review using data mining Classification  
 Technique." *Information and Communications  
 Technology*, Vol. 6, No. 1, January-June, 2016.

[6] ยุทธ ไกยวรรณ. "หลักการและการใช้การวิเคราะห์การ  
 ถดถอยโลจิสติกสำหรับการวิจัย." *วารสารวิจัยมหาวิทยาลัย  
 เทคโนโลยีราชมงคลศรีวิชัย*, ฉบับที่ 4, ปีที่ 1, หน้า 1-2,  
 2555.

[7] กมล นุษยา. มทนาลัย สุดสายสาคร. ภัณจิรา เตชะพรสิน.  
 และสรวิพร โกมลทองทิพย์. "การเปรียบเทียบการทำนายการ  
 ต่ออายุกรรมธรรม์ประกันภัย รดยนต์ ประเภท 1 ระหว่างตัว  
 แบบป่าสุ่มและตัวแบบนิวส์ติงปรับได้." *Thai Journal of  
 Science and Technology*, Vol.10, No. 2, March-April,  
 2021.

[8] S. Ray. "CatBoost: A machine learning library to  
 handle categorical (CAT) data automatically." Available Online at [https://www.analyticsvidhya.com/  
 blog/2017/08/catboost-automated-categorical-data/](https://www.analyticsvidhya.com/blog/2017/08/catboost-automated-categorical-data/),  
 accessed on 19 March 2019.

[9] P. Kottarathil. "Polycystic ovary syndrome (PCOS)." Available Online at [https://www.kaggle.com/prasoon  
 kottarathil/polycystic-ovary-syndrome-pcos](https://www.kaggle.com/prasoonkottarathil/polycystic-ovary-syndrome-pcos), 2020.

## การพยากรณ์ยอดขายสินค้าด้วยเทคนิคการวิเคราะห์อนุกรมเวลา Sales Forecasting using Time Series Techniques

อดิเทพ ปานวณิชยกิจ (Adithep Panwanichayakij)<sup>1</sup> และ วัชรวิวัฒน์ จิตต์สกุล (Watchareewan Jitsakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021856101@email.kmutnb.ac.th, <sup>2</sup>watchareewan.j@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อทำการศึกษาคัดเลือกแบบจำลองที่เหมาะสมในการพยากรณ์ยอดขายที่ดีที่สุดขององค์กรคือ Tapioca Alpha Starch โดยใช้เทคนิคอนุกรมเวลา ประกอบด้วยวิธีทางสถิติ 3 วิธี ได้แก่ แบบจำลองค่าเฉลี่ยเคลื่อนที่ แบบจำลองเอ็กซ์โปเนนเชียลกำลังสอง และแบบจำลองการแยกส่วนประกอบในการพยากรณ์และเลือกเทคนิคที่เหมาะสมโดยใช้โปรแกรม Minitab เป็นเครื่องมือช่วยวิเคราะห์หาค่าทางสถิติทำการเปรียบเทียบประสิทธิภาพแบบจำลองในการพยากรณ์ โดยใช้เกณฑ์ในการคัดเลือกคือ ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ต่ำที่สุด ทำการเก็บรวบรวมข้อมูลยอดขายสินค้า Tapioca Alpha Starch จำนวน 60 เดือน ชุดแรกจำนวน 48 เดือน ตั้งแต่เดือนมกราคม พ.ศ. 2559 ถึงเดือนธันวาคม พ.ศ. 2562 เพื่อทำการทดสอบแบบจำลองการพยากรณ์ทั้ง 3 วิธี และใช้ข้อมูลชุดที่ 2 จำนวน 12 เดือน ตั้งแต่เดือนมกราคม ถึงเดือนธันวาคม พ.ศ. 2563 เพื่อทำการทวนสอบความถูกต้องของแบบจำลอง จากผลการวิจัยพบว่า ยอดขายสินค้าที่นำมาพยากรณ์ด้วยวิธีค่าเฉลี่ยเคลื่อนที่ให้ผลที่มีประสิทธิภาพสูงสุด โดยมีค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยร้อยละ 31.46

**คำสำคัญ:** เทคนิคการวิเคราะห์อนุกรมเวลา วิธีหาค่าเฉลี่ย

เคลื่อนที่ วิธีแบบจำลองเอ็กซ์โปเนนเชียล

กำลังสอง วิธีแบบจำลองการแยก

ส่วนประกอบ

### Abstract

The objective of this research is to select appropriate models, Tapioca Alpha Starch, for forecasting the best sales product by using Time Series Analysis. This analysis method is applied to compare performance among 3 statistical models as following; 1) Moving Average 2) Exponential Smoothing and 3) Decomposition by using Minitab, to determine the Mean Absolute Percent Error (MAPE). The MAPE is used to evaluate performance of models by comparing the efficiency of model's forecasting, the lower MAPE, the more effective of the model. Sixty months of actual sales data was collected and divided into two sets; the first set of 48 months from January 2016 to December 2019 will be tested with 3 methods of forecasting models, and second sets of 12 months from January to December 2020 will be verified the correctness of the model. The results showed that product sales forecasted by using first model, Moving Average, gave the highest efficiently results with the MAPE values equal to 31.46 percentages.

**Keywords:** Time Series Analysis, Moving Average, Double Exponential Smoothing, Decomposition.



## 1. บทนำ

เมื่อยอดขายสินค้าในปัจจุบันเพิ่มขึ้น ลดลง ไม่มีรูปแบบการเติบโต และปัญหาการแพร่ระบาดของเชื้อไวรัสโคโรนา (COVID-19) ที่มีความรุนแรงและขยายวงกว้างมากขึ้น จึงทำให้องค์กรต่างๆ ต้องเตรียมวางแผนการจัดเก็บสินค้าคงคลังให้มีประสิทธิภาพมากที่สุด จึงเป็นที่มาของการพยากรณ์ยอดขายสินค้าที่เป็นอันดับต้น ๆ ขององค์กรคือ Tapioca Alpha Starch เพื่อวางแผน และจัดเตรียมสินค้าให้พร้อมกับยอดขายที่เพิ่มขึ้นในอนาคต

เพื่อให้ธุรกิจดำเนินไปได้ ต้องมีการคาดการณ์หรือพยากรณ์ยอดขายที่มีประสิทธิภาพ ผู้วิจัยจึงได้ไปศึกษา งานวิจัยที่ผ่านมาพบว่า มีเทคนิคการพยากรณ์ยอดขายหลายรูปแบบ เช่น วิธีค่าเฉลี่ยเคลื่อนที่ (Moving Average: MA) วิธีปรับเรียบเอกซ์โปเนนเชียลกำลังสอง (Double Exponential Smoothing: DES) และแบบจำลองการแยกส่วนประกอบ (Decomposition) นอกจากนี้ได้มีการพยากรณ์ความต้องการแวนดา กรณีศึกษา : ร้านรักแวน [1] การจัดการคลังสินค้าผ้าที่เหมาะสมสำหรับอุตสาหกรรมสิ่งทอ [2] การพยากรณ์ปริมาณน้ำฝนรายเดือนในจังหวัดนครสวรรค์โดยใช้เทคนิคการพยากรณ์ทางสถิติ [3] ตัวแบบพยากรณ์ราคามังคุดคละ [4] การพยากรณ์ความต้องการวัตถุดิบสำหรับการวางแผนการผลิตป๊อปคอร์น กรณีศึกษา โรงภาพยนตร์ ABC [5] เป็นต้น

จากปัญหาและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยจึงมีแนวคิดในการนำเทคนิคการวิเคราะห์อนุกรมเวลา มาช่วยในการพยากรณ์ยอดขาย โดยเปรียบเทียบการพยากรณ์ 3 วิธี ได้แก่ วิธีการค่าเฉลี่ยเคลื่อนที่ วิธีการปรับเรียบแบบเอกซ์โปเนนเชียลกำลังสอง และวิธีแยกส่วนประกอบ เพื่อพยากรณ์สินค้าที่มียอดขายสูงที่สุดในทุก ๆ เดือนขององค์กรคือ Tapioca Alpha Starch เพื่อนำวิธีที่เหมาะสมมาประยุกต์ใช้กับองค์กร ช่วยในการพยากรณ์ยอดขายสินค้า การเตรียมสินค้าเข้าคลังได้อย่างแม่นยำที่สุด พร้อมปรับปรุงและพัฒนายอดขายต่อไป

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้รวบรวมข้อมูลทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยแบ่งเป็น 5 หัวข้อหลัก ดังนี้

### 2.1 การพยากรณ์แบบอนุกรมเวลา

การพยากรณ์แบบอนุกรมเวลา เป็นการพยากรณ์เชิงปริมาณที่ต้องอาศัยข้อมูลหลักการทางสถิติในอดีต เพื่อช่วยในการพยากรณ์สิ่งที่จะเกิดขึ้นในอนาคต การพยากรณ์แบบอนุกรมเวลาเป็นการวิเคราะห์ที่ได้มีการยอมรับโดยทั่วไปว่ามีความถูกต้องมากที่สุด สามารถนำข้อมูลที่มีค่าหลากหลายจำนวนมากมาคำนวณหาค่าพยากรณ์ได้ [7]

### 2.2 วิธีแบบจำลองค่าเฉลี่ยเคลื่อนที่ (Moving Average)

วิธีค่าเฉลี่ยเคลื่อนที่ที่เหมาะสมสำหรับการพยากรณ์ในช่วงเวลาระยะสั้น ๆ และข้อมูลที่มีลักษณะค่อนข้างแน่นอนเป็นเส้นตรงและคงที่ตามแนวโน้ม ไม่เหมาะสมกับข้อมูลที่มีลักษณะเป็นแนวโน้ม ถูกลด หรือข้อมูลที่มีลักษณะเป็นการเปลี่ยนแปลงเป็นขั้นบันได เทคนิคนี้ใช้หลักการในการหาค่าเฉลี่ยคือ ใช้ค่าเฉลี่ยจากการสังเกตข้อมูลในอดีตมาคำนวณหาค่าเฉลี่ยแล้วใช้ค่าเฉลี่ยที่ได้นี้เป็นค่าพยากรณ์สำหรับช่วงเวลาถัดไป มีรูปแบบสมการดังนี้ [5]

$$F_{t+1} = \frac{(Y_t + Y_{t-1} + Y_{t-2} + \dots + Y_{t-n+1})}{N} \quad (1)$$

โดย  $F_t$  = ค่าพยากรณ์ในช่วงเวลา  $t$

$t$  = เวลาที่ทำการพยากรณ์

$A_t$  = ค่าความต้องการในงวดที่  $t$

$N$  = จำนวนข้อมูลที่น่ามาเฉลี่ย

### 2.3 วิธีปรับเรียบแบบเอกซ์โปเนนเชียลกำลังสอง (Double Exponential Smoothing Model: DES)

เป็นเทคนิคการพยากรณ์ที่เหมาะสมกับข้อมูลที่มีแนวโน้มแบบเส้นตรง แต่ไม่มีความเป็นฤดูกาล และยังเหมาะกับการพยากรณ์ในระยะสั้น จนถึงการพยากรณ์ในระยะปานกลาง ข้อมูลที่ใช้ในการคำนวณควรจะมีอย่างน้อย 5 ชุด จะช่วยปรับค่าพยากรณ์ให้ขึ้นหรือลง เป็นไปตามแนวทิศทาง มีรูปแบบสมการดังนี้ [10]

$$F_t = \alpha Y_{t-1} + (1-\alpha)F_{t-1} \quad (2)$$

$$F'_t = \alpha F_t + (1-\alpha)F'_{t-1}C \quad (3)$$

โดย  $F_t$  = ค่าพยากรณ์ความต้องการจากวิธี SES

$Y_t$  = ค่าจริงที่เกิดขึ้น ณ เวลา t

$F'$  = ค่าพยากรณ์จากวิธีการ DES สำหรับเวลา t

### 2.1.3 วิธีแยกส่วนประกอบ (Decomposition)

เป็นวิธีที่แยกส่วนประกอบต่าง ๆ ของอนุกรมเวลาออกจากกัน โดยที่จะทำให้พบการเคลื่อนไหวของอนุกรมเวลาและนำไปสร้างตัวแบบพยากรณ์ได้ ส่วนประกอบของอนุกรมเวลาได้แก่ แนวโน้ม ความผันแปรของฤดูกาล ความผันแปรตามวัฏจักร และส่วนประกอบที่ผิดปกติหรือความคลาดเคลื่อน วิธีแยกส่วนประกอบสามารถแบ่งเป็น 2 แบบ ได้แก่ วิธีแยกส่วนประกอบแบบบวก ซึ่งเหมาะกับลักษณะอนุกรมเวลาที่ความผันแปรของฤดูกาลคงที่ และวิธีแยกส่วนประกอบแบบคูณ ซึ่งเหมาะกับลักษณะอนุกรมเวลาที่มีความผันแปรของฤดูกาลไม่คงที่ ซึ่งในวิจัยนี้ได้ใช้วิธีแยกส่วนประกอบแบบคูณ โดยสมการของวิธีแยกส่วนประกอบแบบคูณมีดังนี้ [7]

$$F_t = T_t \times S_t \quad (4)$$

โดย  $F_t$  = ค่าพยากรณ์ของช่วงเวลาปัจจุบัน

$Y_t$  = ค่าของข้อมูลจริงในช่วงเวลาปัจจุบัน

$T_t$  = ค่าประมาณของแนวโน้ม

$S_t$  = ค่าประมาณของการผันแปรตามฤดูกาล

$C_t$  = ค่าประมาณของการแปรผันตามวัฏจักร

$I_t$  = ค่าประมาณของความไม่แน่นอน

$N$  = จำนวนข้อมูล

### 2.2 งานวิจัยที่เกี่ยวข้อง

สรนันท์ และสุภาวดี [6] ได้ทำการศึกษาความต้องการใช้วัตถุดิบเพื่อวางแผนการผลิตปีอบคอร์น โดยได้ใช้วิธีการพยากรณ์ทั้งหมด 7 วิธี ได้แก่ วิธี Moving Average วิธี Double Moving Average วิธี Weighted Moving Average วิธี Exponential Smoothing วิธี Holt-Winter's Method for Additive Seasonal Effects

วิธี Double Exponential Smoothing และวิธี Holt-Winter's Method for Multiplicative Seasonal Effects ผลการวิจัยพบว่าวิธี Moving Average ให้ค่าความผิดพลาดกำลังสองน้อยที่สุด

ธีระพงษ์ ทับพร และคณะ [8] ได้ทำการศึกษายอดขายสินค้าคงปลาทูหมักยักซ์แช่แข็ง เพื่อนำมาหาวิธีพยากรณ์ที่เหมาะสม จำนวน 6 ได้แก่ วิธีค่าเฉลี่ยเคลื่อนที่ วิธีค่าเฉลี่ยถ่วงน้ำหนัก วิธีเอ็กโพเนนเชียล วิธีปรับให้เรียบดับเบิลเอ็กโพเนนเชียล วิธีการปรับให้เรียบแบบฤดูกาล โซลต์วินเทอร์ และวิธีการปรับให้เรียบแบบแนวโน้มของฤดูกาลของโซลต์วินเทอร์ พบว่า วิธีวิธีค่าเฉลี่ยถ่วงน้ำหนักมีค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยน้อยที่สุดร้อยละ 9.48

Parasian, Hidayatullah and Romanus [9] ทำการพยากรณ์การรับนักศึกษาใหม่ใช้ข้อมูลการรับนักศึกษาใหม่เป็นเวลา 6 ปีย้อนหลังในทุกหลักสูตร โดยใช้เทคนิค Double Exponential Smoothing เป็นการพยากรณ์เพื่อทำนายการรับนักศึกษาใหม่ในมหาวิทยาลัย ผลการวิจัยพบว่าใช้ค่าพารามิเตอร์  $\alpha = 0.8$  และ  $\beta = 1$  ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ที่ร้อยละ 11.72

### 3. วิธีการดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อทำการศึกษา คัดเลือกเปรียบเทียบแบบจำลองที่เหมาะสมในการพยากรณ์ยอดขายด้วยเทคนิคการวิเคราะห์อนุกรมเวลา โดยมีรายละเอียดและขั้นตอนการดำเนินงานดังนี้

#### 3.1 การเก็บรวบรวมข้อมูล

การเก็บข้อมูลยอดขายสินค้า Tapioca Alpha Starch ตั้งแต่เดือนมกราคม พ.ศ. 2559 ถึง เดือนธันวาคม พ.ศ. 2563 รวม 60 เดือน เป็นข้อมูลแบบรายเดือน แสดงหน่วยนับเป็น “กิโลกรัม” จัดการข้อมูลให้อยู่ในรูปแบบเดียวกัน โดย แบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ชุดข้อมูล Train Set เป็นข้อมูลยอดขายสินค้าตั้งแต่เดือนมกราคม พ.ศ. 2559 ถึงเดือนธันวาคม พ.ศ. 2562 รวม 48 เดือน เพื่อใช้ในการสร้างแบบจำลองและเปรียบเทียบประสิทธิภาพในการพยากรณ์ของแบบจำลอง และชุดข้อมูล Test Set เป็นข้อมูลยอดขายสินค้า ตั้งแต่เดือน

มกราคม พ.ศ. 2563 ถึงเดือนธันวาคม พ.ศ. 2563 รวม 12 เดือน เพื่อใช้เป็นข้อมูลสำหรับทวนสอบความแม่นยำของการพยากรณ์

**3.2 การสร้างแบบจำลองพยากรณ์**

การสร้างแบบจำลองการพยากรณ์ใช้โปรแกรม Minitab เป็นเครื่องมือ สำหรับการพยากรณ์ใช้เทคนิค Time Series Analysis 3 วิธี ได้แก่ วิธี Moving Average โดยทำการกำหนดค่า MA Length 3, 6, 9, 12 วิธี Double Exponential Smoothing กำหนดพารามิเตอร์ค่า  $\alpha, \gamma$  ตั้งแต่ 0.05, 0.1 ถึง 0.9 และวิธี Decomposition โดยทำการกำหนดค่า Length 3, 6, 9, 12

**3.3 การคัดเลือกแบบจำลองในการพยากรณ์**

พิจารณาประสิทธิภาพของแบบจำลองพยากรณ์ทั้ง 3 วิธี ด้วยค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) และค่าเบี่ยงเบนสัมบูรณ์เฉลี่ย (MAD) ที่ต่ำที่สุด

**3.4 การทวนสอบความถูกต้องของแบบจำลอง**

ทวนสอบความถูกต้องของแบบจำลองพยากรณ์ นำข้อมูลการพยากรณ์ที่ได้จากแบบจำลองมาทวนสอบความถูกต้องและแม่นยำกับข้อมูลชุดที่ 2 (Test Set) จำนวน 12 เดือนพิจารณาค่าความคลาดเคลื่อนในการพยากรณ์ ด้วยค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE)

**3.5 นำเสนอข้อมูลการพยากรณ์**

นำเสนอข้อมูลการพยากรณ์ขอคขายในรูปแบบ Dashboard ด้วย โปรแกรม Power BI

**4. ผลการดำเนินงาน**

ผลการดำเนินงานวิจัยแบ่งออกเป็น 4 หัวข้อ แสดงรายละเอียดได้ดังนี้

**4.1 ผลการดำเนินงานสร้างแบบจำลองพยากรณ์**

ผลการสร้างแบบจำลองพยากรณ์ขอคขายสินค้า โดยเทคนิคการวิเคราะห์แบบอนุกรมเวลา จำนวน 3 วิธี ได้แก่ วิธี Moving Average วิธี Double Exponential Smoothing และวิธี Decomposition วิเคราะห์ค่าสถิติ

ด้วยโปรแกรม Minitab ได้ผลการวิจัยมีรายละเอียดดังตารางที่ 1 ถึง 3

**ตารางที่ 1:** การพยากรณ์ด้วยวิธี Moving Average

วิธีพยากรณ์	พารามิเตอร์	MAPE	MAD
Moving Average	<b>MA Length 3</b>	<b>18</b>	<b>910</b>
	MA Length 6	20	1002
	MA Length 9	20	1018
	MA Length 12	21	1075

จากตารางที่ 1 เมื่อพิจารณาค่าความคลาดเคลื่อนพบว่า MA Length 3 ให้ประสิทธิภาพที่ดีที่สุด โดยค่า MAPE เท่ากับ 18 และค่า MAD เท่ากับ 910

**ตารางที่ 2:** การพยากรณ์ด้วยวิธี Double Exponential Smoothing

วิธีพยากรณ์	พารามิเตอร์	MAPE	MAD
Double Exponential Smoothing	$\alpha = 0.1, \gamma = 0.5$	26	1236
	<b><math>\alpha = 0.2, \gamma = 0.1</math></b>	<b>25</b>	<b>1177</b>
	$\alpha = 0.2, \gamma = 0.2$	26	1195

จากตารางที่ 2 เมื่อพิจารณาค่าความคลาดเคลื่อนพบว่า  $\alpha = 0.2$  และ  $\gamma = 0.1$  ให้ประสิทธิภาพที่ดีที่สุด โดยค่า MAPE เท่ากับ 20 และค่า MAD เท่ากับ 1177

**ตารางที่ 3:** การพยากรณ์ด้วยวิธี Decomposition

วิธีพยากรณ์	พารามิเตอร์	MAPE	MAD
Decomposition	Seasonal Length 3	31	1448
	Seasonal Length 6	31	1431
	<b>Seasonal Length 9</b>	<b>28</b>	<b>1282</b>
	Seasonal Length 12	30	1418

จากตารางที่ 3 เมื่อพิจารณาค่าความคลาดเคลื่อนพบว่าค่า Seasonal Length 9 ให้ประสิทธิภาพดีที่สุด โดยมีค่า MAPE เท่ากับ 28 และค่า MAD เท่ากับ 1282

**4.2 การเปรียบเทียบและคัดเลือกแบบจำลอง**

การเปรียบเทียบและคัดเลือกแบบจำลองนำค่าความผิดพลาดของตัวแบบการพยากรณ์ ได้แก่ ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percent Error : MAPE) และค่าเบี่ยงเบนสัมบูรณ์เฉลี่ย (Mean Absolute Deviation : MAD) ของวิธี Moving Average วิธี Double Exponential Smoothing และวิธี

Decomposition มาเปรียบเทียบกับเพื่อหาแบบจำลองพยากรณ์ที่เหมาะสม และมีความแม่นยำกับข้อมูลยอดขายสินค้า Tapioca Alpha Starch โดยพิจารณาจากวิธีพยากรณ์ที่ให้ค่าความผิดพลาดของแบบจำลองพยากรณ์ต่ำที่สุด และพบว่าวิธีแบบจำลองค่าเฉลี่ยเคลื่อนที่ เป็นวิธีที่ค่าความผิดพลาดต่ำที่สุด ซึ่งเหมาะสมที่จะใช้ในการพยากรณ์ยอดขายสินค้า Tapioca Alpha Starch รายละเอียดแสดงดังตารางที่ 4

ตารางที่ 4: การเปรียบเทียบความแม่นยำของแบบจำลองพยากรณ์

วิธีพยากรณ์	พารามิเตอร์	MAPE	MAD
<b>Moving average</b>	<b>MA Length 3</b>	<b>18</b>	<b>910</b>
Double Exponential Smoothing	$\alpha = 0.2$ , $\gamma = 0.1$	25	1177
Decomposition	Seasonal Length 9	28	1282

จากตารางที่ 4 แสดงการเปรียบเทียบผลการพยากรณ์ยอดขายสินค้าทั้ง 3 วิธี พบว่าแบบจำลองพยากรณ์วิธีแบบจำลองค่าเฉลี่ยเคลื่อนที่ ให้ประสิทธิภาพในการพยากรณ์สูงสุด โดยมีค่า MAPE เท่ากับ 18 และค่า MAD เท่ากับ 910

4.3 การทวนสอบความถูกต้องของแบบจำลอง

ผลการทวนสอบความถูกต้องของแบบจำลองจากการพิจารณาค่าความคลาดเคลื่อนของแบบจำลองพยากรณ์ที่ให้ผลลัพธ์ที่ดีที่สุด ได้แก่ วิธี Moving Average รองลงมาคือวิธี Double Exponential Smoothing และวิธี Decomposition นำข้อมูลการพยากรณ์ที่ได้มาทำการทวนสอบความถูกต้อง และความแม่นยำกับข้อมูลยอดขายสินค้า จำนวน 12 เดือน พบว่า แบบจำลองพยากรณ์วิธีแบบจำลองค่าเฉลี่ยเคลื่อนที่ มีความคลาดเคลื่อนต่ำที่สุดที่ร้อยละ 31.46 รายละเอียดแสดงดังตารางที่ 5

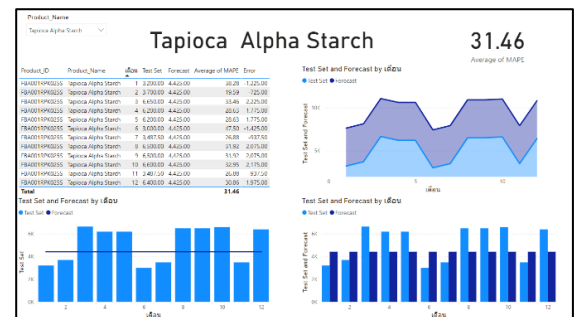
ตารางที่ 5: ผลการทวนสอบความถูกต้องของแบบจำลอง

เดือน	ยอดขายสินค้าต่อกิโลกรัม	Moving Average		
		Test Set	Forecast	Error
JAN	3200.00	4425.00	-1225.00	38.28
FEB	3700.00	4425.00	-725.00	19.59

MAR	6650.00	4425.00	2225.00	33.46
APR	6200.00	4425.00	1775.00	28.63
MAY	6200.00	4425.00	1775.00	28.63
JUN	3000.00	4425.00	-1425.00	47.50
JUL	3487.50	4425.00	-937.50	26.88
AUG	6500.00	4425.00	2075.00	31.92
SEP	6500.00	4425.00	2075.00	31.92
OCT	6600.00	4425.00	2175.00	32.95
NOV	3487.50	4425.00	-937.50	26.88
DEC	6400.00	4425.00	1975.00	30.86
<b>ร้อยละ</b>				<b>31.46</b>

4.4 การนำเสนอข้อมูล

ผู้วิจัยนำเสนอผลการพยากรณ์ โดยใช้วิธีแบบจำลองค่าเฉลี่ยเคลื่อนที่ เป็นตัวแบบในการพยากรณ์ยอดขายสินค้าในอนาคต โดยใช้โปรแกรม Power BI Desktop เป็นเครื่องมือช่วยในการแสดงผลพยากรณ์ในรูปแบบของรายงาน แสดงดังภาพที่ 1



ภาพที่ 1: แผนภูมิภาพแสดงข้อมูลยอดขายสินค้า

จากภาพที่ 1 แสดงการเปรียบเทียบข้อมูลจำนวนยอดขายสินค้า Tapioca Alpha Starch ที่เกิดขึ้นจริงจำนวน 12 เดือน ตั้งแต่เดือนมกราคม ถึงเดือนธันวาคม พ.ศ. 2563 กับข้อมูลยอดขายที่มาจากกรสร้างแบบจำลองพยากรณ์ด้วยวิธีแบบจำลองค่าเฉลี่ยเคลื่อนที่พบว่าจำนวนยอดขายสินค้า Tapioca Alpha Starch มีความสอดคล้องและใกล้เคียงกัน

5. สรุป

งานวิจัยฉบับนี้ผู้วิจัยได้ศึกษาการพยากรณ์ด้วยเทคนิคอนุกรมเวลาเพื่อให้ได้วิธีการพยากรณ์ที่ดีและ

เหมาะสมที่สุดสำหรับการพยากรณ์ยอดขายสินค้า Tapioca Alpha Starch ผู้วิจัยได้คัดเลือกวิธีการพยากรณ์ที่ได้รับความนิยม จำนวน 3 วิธี ได้แก่ วิธี Moving Average วิธี Double Exponential Smoothing และวิธี Decomposition ทดสอบประสิทธิภาพของแบบจำลองการพยากรณ์ด้วยค่า MAPE และค่า MAD จากผลการวิจัยพบว่า วิธีแบบจำลองค่าเฉลี่ยเคลื่อนที่มีความเหมาะสมที่สุด ค่า MAPE เท่ากับ 18 ค่า MAD เท่ากับ 910 และหลังจากนำข้อมูลมาทวนสอบกับข้อมูลจริงจำนวน 12 เดือน สรุปได้ว่าการพยากรณ์ยอดขายสินค้า Tapioca Alpha Starch โดยสร้างแบบจำลองการพยากรณ์ด้วยวิธีแบบจำลองค่าเฉลี่ยเคลื่อนที่มีความแม่นยำ พบความคลาดเคลื่อนในการพยากรณ์ที่ร้อยละ 31.46

### เอกสารอ้างอิง

- [1] อนุสรณ์ บุญสง่า, การพยากรณ์ความต้องการแวนดา กรณีศึกษา ร้านรักแวน. การค้นคว้าอิสระวิทยาศาสตร์ มหาวิทยาลัย สาขาการจัดการ โลจิสติกส์และโซ่อุปทาน วิทยาลัยโลจิสติกส์และซัพพลายเชน มหาวิทยาลัยศรีปทุม.
- [2] นิตยา วงศ์ระวีง, การจัดการคลังสินค้าที่เหมาะสมสำหรับ อีคอมเมิร์ซสิ่งทอ การค้นคว้าอิสระวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาการจัดการวิศวกรรม ภาควิชา วิศวกรรมอุตสาหกรรมและการจัดการ มหาวิทยาลัยศิลปากร.
- [3] ชม ปานตา และยุภาวดี สารัญญฤทธิ์, “การพยากรณ์ปริมาณน้ำฝนรายเดือนในจังหวัดนครสวรรค์โดยใช้เทคนิคการพยากรณ์ทางสถิติ” วารสารวิชาการ วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครสวรรค์ ปีที่ 9 ฉบับที่ 10 กรกฎาคม-ธันวาคม พ.ศ. 2560 หน้า 127-142.
- [4] วรางคณา เรียนสุทธิ์, “ตัวแบบพยากรณ์ราคามังคุดคละ” วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัย อุบลราชธานี ปีที่ 19 ฉบับที่ 2 พฤษภาคม-สิงหาคม 2560 หน้า 31-42.
- [5] วณิกเกียรติ์ ยิ่งพันธ์ บัววรรณ ศรีชัยกุล จตุพร เหลืองอุบล และ พุทธิไกร ประมวล, “ตัวแบบทำนายอุบัติเหตุการเกิดโรคไข้เลือดออกของประชาชน จังหวัดอุบลราชธานี” วารสารวิจัยสาธารณสุขศาสตร์ มหาวิทยาลัยราชภัฏ อุบลราชธานี ปีที่ 9 ฉบับที่ 2 กรกฎาคม-ธันวาคม พ.ศ. 2563 หน้า 174-184.
- [6] สรนนท์ ทัพนันท์ และสุภาวดี สายสนิท, “การพยากรณ์ความต้องการวัตถุดิบสำหรับการวางแผนการผลิตปีอปคอร์น กรณีศึกษา โรงภาพยนตร์ ABC” การประชุมวิชาการ มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ครั้งที่ 4 และการประชุมระดับนานาชาติ มหาวิทยาลัยเทคโนโลยี ราชมงคลรัตนโกสินทร์ ครั้งที่ 1 26-28 มิถุนายน 2562
- [7] อภิชัย พรหมอ่อน, การศึกษาการพยากรณ์แบบอนุกรมเวลา (Time Series) เพื่อการวางแผนการสั่งซื้อวัตถุดิบ กรณีศึกษา บริษัทผลิตชิ้นส่วนท่อยางรถยนต์. สารนิพนธ์ปริญญา บริหารธุรกิจมหาบัณฑิต บัณฑิตวิทยาลัย สาขาวิชาการจัดการอุตสาหกรรม สถาบันเทคโนโลยีไทย-ญี่ปุ่น.
- [8] ชีระพงษ์ ทับพร, ยอดคนภา เกษเมือง, เอกพล ทับพร และ พัชรดิษฐ์ แปงจิตต์, “การพยากรณ์ยอดขายและการบริหารสินค้าคงคลัง ของสินค้ากวางหมักยักษ์แช่แข็ง : บริษัท สยาม แม็คโคร จำกัด มหาชน” วารสารวิชาการมหาวิทยาลัย ธนบุรี(วิทยาศาสตร์และเทคโนโลยี) ปีที่ 2 ฉบับที่ 2 เดือน กรกฎาคม – ธันวาคม 2561 หน้า 28-41
- [9] Parasian DP., S., Hidayatullah H., and Romanus D., “Forecasting Acceptance of New Students using Double Exponential Smoothing Method.” *Journal of Critical Reviews* Vol 7 Issue 1, 2020, Page 300-305.
- [10] อภิชัย พรหมอ่อน, การพยากรณ์ความต้องการในการสั่งซื้อ และจัดการวัตถุดิบคงคลัง กรณีศึกษาบริษัทผลิตเครื่องฟอกอากาศ. สารนิพนธ์วิทยาศาสตรมหาบัณฑิตบัณฑิตวิทยาลัย สาขาวิชาการจัดการโลจิสติกส์และโซ่อุปทาน คณะ โลจิสติกส์ มหาวิทยาลัยบูรพา.

# ระบบการจัดการเวลาเข้าทำงานในพื้นที่โดยประยุกต์ใช้ 2FA และตำแหน่งที่ตั้ง

## Application Attendance System Base 2FA and Location

ณัฐกมล ศิริภาณุรักษ์ (Natkamon Siripanurak)<sup>1</sup> และชนพล เจนสุทธิเวชกุล (Tanapon Jensuttiwetachakul)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858236@email.kmutnb.ac.th, <sup>2</sup>tanapon.j@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอการพัฒนา ระบบการจัดการเวลาเข้าทำงาน ที่นำการยืนยันตัวตนที่มีความปลอดภัยที่สูงขึ้นมาใช้งานร่วมกับระบบการตรวจสอบรูปแบบเดิมแต่ทำการพัฒนาในส่วนของการตรวจสอบตำแหน่ง GPS ในขณะที่ทำการบันทึกเวลา โดยการกำหนดขอบเขตการบันทึกเวลาให้ชัดเจนขึ้น เพื่อให้ระบบนั้นแม่นยำและตรวจสอบง่าย โดยใช้ระบบยืนยันตัวตนแบบสองชั้น หรือ Two Factor Authentication (2FA) หลังจากล็อกอินด้วยรหัสผ่านตามปกติ จะมีการยืนยันผ่าน OTP (One-Time Password) ที่ส่งให้ผ่าน Email พนักงานเพื่อป้อนรหัสให้ตรงกับระบบเพื่อยืนยันตัวตน

จากการทดสอบสามารถวัดผลประสิทธิภาพของระบบได้ค่าความถูกต้อง (Accuracy) มากกว่า 90 เปอร์เซ็นต์ และผ่านทุกการทดสอบที่จำลองขึ้น ดังนั้นสามารถสรุปได้ว่าระบบที่ถูกพัฒนามานั้นมีประสิทธิภาพในการบันทึกเวลาเข้างานได้อย่างปลอดภัย และระบบรับข้อมูลที่น่าเชื่อถือ สะดวกต่อการตรวจสอบข้อมูลมากยิ่งขึ้น

**คำสำคัญ:** ระบบจัดการเวลาเข้างาน การกำหนดขอบเขตพื้นที่ ระบบยืนยันตัวตนสองชั้น

### Abstract

This research aims to develop time attendance management system. Security of the system was improved by using GPS location monitoring while recording time and using two factor Authentication by sending OTP via email to identify user.

The result indicated that the performance of the system with an accuracy (Accuracy) of 90 percent and passed every simulated test. Therefore, it can be concluded that the developed system is effective in recording time attendance securely, and the system will receive reliable and more convenient data verification.

**Keywords:** Time Attendance Management System, Area Scope Determination, Two Factor Authentication System.

### 1. บทนำ

ปัจจุบันการทำงานในบริษัทมีการบันทึกเวลาเข้าทำงานทุกแห่งในประเทศไทย เพื่อให้บริษัทได้รับรู้ถึงการเข้าทำงานของพนักงานเพื่อนำไปวิเคราะห์และประเมินการจ่ายค่าแรงที่เหมาะสม ซึ่งเป็นพื้นฐานในการคำนวณค่าแรงรายเดือนของพนักงาน โดยทั่วไปบริษัทมีระบบการจัดการเวลาดั้งเดิมที่เข้ามาเป็นเวลานาน เช่น การบริหารเวลาโดยจดบันทึกลงหนังสือเก็บเวลา การบริหารเวลาโดยใช้บัตรเวลาหรือการตอกบัตร การบริหารเวลาโดยบันทึกลง Excel และในปัจจุบันนิยมใช้กันมากที่สุดคือ การจัดการบริหารเวลาโดยใช้ระบบการจัดการเวลา ระบบจะช่วยให้การบันทึกและจัดการข้อมูลของพนักงาน โดยใช้บัตรหรือการเก็บลายนิ้วมือแทนบัตรลงเวลา ข้อมูลจะถูกรวบรวมและเก็บบันทึกโดยอัตโนมัติ ข้อมูลสามารถยืนยันได้แบบเรียลไทม์และสามารถมั่นใจข้อมูลได้มากขึ้นในความถูกต้องของเวลาที่บันทึก เนื่องจากเป็นระบบคอมพิวเตอร์

จากข้อมูลของการจัดการบริหารเวลาที่ผ่านมามีการพัฒนาของการจัดการให้ดีขึ้น โดยนำส่วนของเทคโนโลยีเข้ามาใช้ปรับปรุงเพื่ออำนวยความสะดวกและการจัดการที่รวดเร็วและแม่นยำกว่าวิธีการจัดการแบบเดิม ซึ่งในปัจจุบันการทำเอาสมาร์ตโฟนมาใช้ในการจัดการบริหารเวลามากขึ้น เพราะสมาร์ตโฟนเป็นสิ่งอำนวยความสะดวกที่เข้ามามีสำคัญมากในการใช้ชีวิต การติดต่อสื่อสาร สามารถติดตัวไปทุกที่ และมีหลายรุ่นให้เลือกใช้ตามจุดประสงค์ของผู้ใช้ เช่น เลือกตามคุณสมบัติหรือเลือกตามราคา เป็นต้น ทำให้สมาร์ตโฟนเข้าถึงประชากรทุกระดับ จึงทำให้การสร้างระบบบันทึกเวลาเข้างานเป็นสิ่งที่ทำได้ โดยระบบที่จะนำมาใช้งานบนสมาร์ตโฟนได้แก่ Application Attendance System แต่ละที่ก็จะมีลักษณะการทำงานและการออกแบบที่แตกต่างกันออกไป แต่จุดมุ่งหมายที่แท้จริงของระบบนี้คือการบันทึกเวลาเข้างานและตรวจสอบว่าการเข้างานนั้นถูกต้องระดับไหน โดยใช้ความสามารถบางประการของสมาร์ตโฟนในการตรวจสอบ เช่น การบันทึกตำแหน่ง GPS ของอุปกรณ์ การดึงเวลาในเขตเวลาของโลกตามตำแหน่งและผู้ใช้บริการเครือข่าย รวมถึงการใช้อินเทอร์เน็ตสำหรับส่งข้อมูลการบันทึกเวลา จากที่กล่าวมา Application Attendance System บนสมาร์ตโฟนนั้นมีประโยชน์มากมายแต่พบข้อสงสัยว่าสามารถตรวจสอบการตั้งใจทุจริตสำหรับกลุ่มพนักงานบางกลุ่มที่ใช้จุดผิดพลาดของระบบบันทึกเข้าทำงานได้ จึงเกิดปัญหาเดิมขึ้นอีกครั้ง เช่น ทำให้ยากต่อการตรวจสอบเนื่องจากพนักงานมีจำนวนมากทำให้เจ้าหน้าที่ตรวจสอบผิดหรือพลาดการตรวจสอบได้

จึงเกิดเป็นแนวคิดในการพัฒนา Application Attendance System ที่นำระบบยืนยันตัวตนที่มีความปลอดภัยที่สูงขึ้นมาใช้ร่วมกับระบบจัดการเวลาแบบเดิมและทำการพัฒนาในส่วนของการตรวจสอบตำแหน่งในขณะที่ทำการบันทึกเวลาโดยการกำหนดขอบเขตในการบันทึกเวลาให้ชัดเจนขึ้น เพื่อให้ระบบนั้นแม่นยำและตรวจสอบง่าย ส่วนระบบยืนยันตัวตนที่มีความปลอดภัยในปัจจุบันที่หลายโครงการนำไปใช้ คือ 2

Factor Authentication (2FA) เป็นการยืนยันตัวตนขั้นที่ 2 หลังจากเข้าสู่ระบบด้วยรหัสผ่านตามปกติ จะมีการยืนยันผ่าน OTP (One-Time Password) ที่ส่งให้เราผ่านข้อความ SMS บนโทรศัพท์สมาร์ตโฟน หรือแอป Authentication เพื่อป้อนรหัสให้ตรงกัน และยืนยันตัวตน โดยรหัส OTP จะเป็นรหัสผ่านที่ใช้งานได้เพียงครั้งเดียว และมีอายุการใช้งานสั้นเพียงไม่กี่วินาที หรืออาจยาวนานถึง 15 นาที หากไม่ใช่เจ้าของบัญชี ไม่มี OTP หรือป้อนรหัสไม่ถูกต้องจะไม่สามารถเข้าสู่ระบบ ดังนั้น Application Attendance System ที่พัฒนาขึ้นจะมีความสามารถในการยืนยันตัวตน 2FA และการพัฒนาการกำหนดขอบเขตให้กับแผนที่ในการบันทึกเวลาทำงาน ซึ่งจะทำให้ระบบนั้นมีความแม่นยำปลอดภัยและตรวจสอบง่ายมากขึ้นแน่นอน

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ระบบลงเวลาทำงาน

ระบบลงเวลาทำงาน คือ ระบบที่ช่วยบริหารและตรวจสอบเวลาทำงานของพนักงาน เพื่อจุดประสงค์ในการบริหารค่าจ้างของพนักงาน และส่งผลต่อการตรวจสอบประสิทธิภาพในการทำงานของพนักงานแต่ละคนในองค์กร ระบบนี้อาจเป็นระบบเดี่ยวที่ถูกผนวกเข้ากับระบบจัดการเงินเดือนขององค์กรได้ นอกจากนี้ยังถูกผนวกเข้ากับระบบตรวจสอบคนงาน การบริหารการเข้าออกอาคาร และการจัดสรรเวลาสำหรับพนักงานได้ด้วย [1][2]

### 2.2 หลักการในการบันทึกข้อมูลเวลาเข้างาน

หลักการในการบันทึกเวลาบนระบบบันทึกเวลาเข้าทำงานที่พัฒนาขึ้นให้เกิดความถูกต้อง เนื่องจากเวลาเข้าทำงานเป็นเรื่องที่ละเอียดอ่อน สามารถทำให้พนักงานถูกโทษในการเข้าทำงานสายเพียงหลักนาที ตามกฎที่ไม่ยืดหยุ่นของชั่วโมงการทำงานในบริษัท และการนำระบบที่พัฒนาเข้ามาตรวจสอบไม่เป็นผลดีในเรื่องความสะดวกสบายหรือรวดเร็ว แต่ช่วยในเรื่องของการตรวจสอบและการจัดการของฝ่ายบุคคลมากกว่า ซึ่งการใช้งานระบบทำให้เวลานั้นเพิ่มขึ้นสำหรับการ Check In

เพราะผู้ทดสอบนั้นต้องตรวจสอบว่าอยู่ในพื้นที่หรือทำการรอ One Time Password ซึ่งอาจจะใช้เวลาเกิน 1-3 นาที ทำให้เมื่อบันทึกการเข้าทำงานสำเร็จ เวลาเข้างานนั้นมีโอกาสที่จะเกินกำหนดไปแล้ว จากการสังเกตถึงปัญหาจึงทำการปรับปรุงขั้นตอนการทำงานของแอปพลิเคชันใหม่ โดยระบบจะทำการบันทึกเวลาการเข้างานไปเก็บในระบบฐานข้อมูลก่อน ในจังหวัดกนครเริ่มต้นการตรวจสอบ โดยยังไม่ผ่านการตรวจสอบใด ๆ แต่หลังจากนั้นในการตรวจสอบพื้นที่และการยืนยันตัวตนจะเป็นการปรับปรุงสถานะการทำงาน หากทำการตรวจสอบไม่สำเร็จ การทำงานในรายการนี้ถูกปรับเป็นสถานะล้มเหลว แต่ถ้าบันทึกตรวจสอบสำเร็จ รายการก็จะมีสถานะเข้าทำงานที่สำเร็จ แสดงออกมาทางรายงาน โดยไม่มีผลกระทบในเรื่องของการเสียเวลาในการตรวจสอบในเวลาเร่งรีบ [2][5]

### 2.3 การนำเข้าข้อมูลแผนที่

การนำเข้าข้อมูล (Input data) เป็นกระบวนการบันทึกข้อมูลเข้าสู่ระบบดิจิทัล การสร้างฐานข้อมูลที่ละเอียด ถูกต้อง เป็นสิ่งสำคัญอย่างยิ่งในการปฏิบัติงานด้วยระบบสารสนเทศภูมิศาสตร์ซึ่งจำเป็นต้องมีการประเมินคุณภาพข้อมูล ที่นำเข้าระบบในเรื่องแหล่งที่มาของข้อมูล วิธีการสำรวจข้อมูล มาตรฐานของแผนที่ ความถูกต้อง ความละเอียด พื้นที่ที่ข้อมูลครอบคลุมและปีที่จัดทำข้อมูล เพื่อประเมินคุณภาพ และคัดเลือกข้อมูลที่จะนำเข้าสู่ระบบฐานข้อมูล [1][3]

### 2.4 การเลือกแสดงขอบเขตพื้นที่

การเลือกแสดงขอบเขตพื้นที่เพื่อลดการทำงานให้กับแอปพลิเคชันเมื่อผู้ใช้งานได้ทำการเพิ่มขอบเขตพื้นที่เข้าไปในระบบฐานข้อมูลเป็นจำนวนมาก เนื่องจากมีความต้องการใช้งานที่มากขึ้น ทำให้จำนวนรูป Polygon ที่แสดงบนแอปพลิเคชันนั้นมีมากส่งผลโดยตรงกับประสิทธิภาพของสมาร์ตโฟน เนื่องจากแอปพลิเคชันรันบนระบบปฏิบัติการ Android ที่มีสมาร์ตโฟน จำนวนมากที่ใช้ระบบปฏิบัติการดังกล่าว จึงมีความหลากหลายทั้งตัวเครื่อง ความสามารถ จึงเกิดปัญหาความไม่เท่าเทียมของประสิทธิภาพสมาร์ตโฟน กล่าวคือหากสมาร์ต

โฟนรุ่นไหนมีประสิทธิภาพที่ดี ก็จะสามารถประมวลผลได้มากและรวดเร็ว แต่เครื่องที่มีประสิทธิภาพต่ำกว่าจะประมวลผลได้น้อยและช้ากว่า เกิดข้อผิดพลาดในการทำงาน เนื่องจากทรัพยากรที่ใช้ในการประมวลผลนั้นไม่เพียงพอ ส่วนมากจะพบได้ในรุ่นราคาประหยัด หน่วยประมวลผลน้อย ดังนั้นการทำการคัดเลือกแสดงขอบเขตพื้นที่ที่ต้องการ จึงช่วยให้เกิดการประหยัดทรัพยากรในการประมวลผล



ภาพที่ 1: The coordinates grid in google maps [3]

โดยหลักการคัดเลือกรุ่นจะใช้หลักการของ Google Maps Grid จากระบบหลังบ้านของ Google Maps นั้นมีการแบ่งพื้นที่ออกเป็นส่วนเล็ก ๆ หรือที่เรียกว่าระบบ Grid ซึ่งได้จากการคำนวณจะได้ค่าออกมาเป็นตัวเลขที่อ้างอิงถึงตำแหน่ง และนำมาเทียบกับตำแหน่งที่ผู้ทดสอบระบบอยู่ เมื่อนำค่าตำแหน่งดังกล่าวไปผ่านการคำนวณและเปรียบเทียบจะสามารถบอกได้ว่า ตำแหน่งที่ผู้ทดสอบอยู่ที่ Grid ไหนและแสดงขอบเขตพื้นที่ที่อยู่ในบริเวณ Grid เดียวกันขึ้นมา

พื้นที่ที่อยู่นอกเหนือจาก Grid ที่ผู้ทดสอบอยู่จะไม่ถูกวาดลงบนระบบช่วยในเรื่องของการเพิ่มประสิทธิภาพได้ในระดับหนึ่ง หากผู้ทดสอบมีการเคลื่อนไหวหรือย้ายตำแหน่งไปยัง Grid ข้างเคียง Polygon ที่อยู่ใน Grid ข้างเคียงก็จะถูกวาดขึ้นอย่าง Realtime เป็นผลมาจากความสามารถของ Realtime Database NoSQL ที่มาใช้ในการเก็บข้อมูลในส่วนข้อมูลแผนที่ [3]

### 2.5 Assisted Global Positioning System (A-GPS)

Assisted Global Positioning System หรือ A-GPS เป็นระบบที่ช่วยในการเพิ่มสมรรถภาพของ GPS ในช่วงการเริ่มต้นการใช้งาน A-GPS นิยมพบได้ในสมาร์ตโฟนที่มีความสามารถ GPS โดยทั่วไป GPS จะใช้เวลาประมาณ 30-40 วินาทีในการเริ่มต้นระบุตำแหน่ง ใน



ระบบ A-GPS จึงถูกนำมาใช้ โดยการทำงานผ่านทาง เซิร์ฟเวอร์ของระบบ โดยข้อมูลเป็นข้อมูลจากวงโคจรที่ถูกนำมาเก็บไว้ล่วงหน้าในฐานข้อมูล และเมื่อเครื่องมือสามารถดาวน์โหลดข้อมูลเหล่านี้โดยตรงผ่านทางคลื่นสัญญาณโทรศัพท์มือถือ เช่น GSM, CDMA, LTE หรือแม้แต่สัญญาณวิทยุอย่าง Wi-Fi ซึ่งข้อมูลเหล่านี้ปกติจะมีการดาวน์โหลดด้วยความเร็วที่สูงกว่า ทำให้ใช้เวลาในการรับข้อมูลได้เร็วขึ้น [3]

## 2.6 วิธีการตรวจสอบการปลอมตำแหน่งขณะตรวจสอบ

ในการพัฒนาระบบต้องมีการคำนึงถึงการตรวจสอบตำแหน่งจากผู้ใช้งานว่าเป็นข้อมูลจริง โดยหลักการในการตรวจสอบ จะตรวจจับการเคลื่อนที่ด้วยความเร็วเกินความจริง หรือมีการข้ามตำแหน่งที่อยู่ใกล้กันมาก โดยความสามารถในการเก็บตำแหน่งการเคลื่อนที่ในขณะที่ใช้งานระบบของฐานข้อมูล ทำให้สามารถตรวจสอบการเคลื่อนที่จากตำแหน่งต้นทางและปลายทางว่าปกติหรือไม่ เนื่องจากหากเกิดการปลอมตำแหน่งนั้น การเคลื่อนที่ของตำแหน่งจะผิดปกติเมื่อระบบพบสิ่งผิดปกติเกิดขึ้นจะแจ้งเตือนให้ผู้ใช้งานทราบ และเก็บบันทึกความผิดปกติไว้เป็นสถิติ [3]

## 2.7 Realtime Database

Firebase Realtime Database (NoSQL) ที่เก็บข้อมูลในรูปแบบของ JSON ในลักษณะ Key และ Value และมีการ sync ข้อมูลแบบ Realtime กับทุกอุปกรณ์ที่เชื่อมต่อแบบอัตโนมัติในเสี้ยววินาที รองรับการทำงานเมื่อออฟไลน์ (ข้อมูลจะถูกเก็บไว้ใน local จนกระทั่งกลับมาออนไลน์ก็จะทำการ sync ข้อมูลให้อัตโนมัติ) รวมถึงมี Security Rules สามารถออกแบบเงื่อนไขการเข้าถึงข้อมูลทั้งการอ่านและเขียน สำหรับโครงการนี้ใช้ Realtime Database ในการเก็บข้อมูลเกี่ยวกับแผนที่สำหรับใช้ประมวลผลใน Google Maps APIs [6]

## 2.8 Two Factor Authentication (2FA)

Two Factor Authentication (2FA) คือการยืนยันรหัสผ่านการเข้าสู่ระบบอีกครั้งโดยส่วนใหญ่แล้วจะใช้กับธุรกรรมออนไลน์ (Online Transaction) เป็นส่วนมาก โดยจุดประสงค์ของการใช้งานจะเน้นเรื่อง

ความปลอดภัยของบัญชีเป็นหลัก กล่าวคือเป็นการยืนยันตัวตนขั้นที่ 2 หลังจากเข้าสู่ระบบด้วยรหัสผ่านตามปกติ จะมีการยืนยันผ่าน OTP (One-Time Password) เพื่อป้องกันการให้ตรงกัน และยืนยันตัวตน โดยรหัส OTP จะเป็นรหัสผ่านที่ใช้งานได้เพียงครั้งเดียว และมีอายุการใช้งานสั้นมาก หากไม่ใช่เจ้าของบัญชีจะไม่ได้รับ OTP ก็ไม่สามารถเข้าสู่ระบบเพื่อทำงานได้ [4][7]

## 3. วิธีดำเนินการวิจัย

ในการพัฒนาระบบการจัดการเวลาเข้าทำงานในพื้นที่ที่สามารถทำการตรวจสอบพื้นที่และการยืนยันตัวตนนั้น มีหลายส่วนที่ต้องพัฒนาขึ้นมาทำงานร่วมกันเพื่อให้เป็นระบบที่ใช้งานได้สมบูรณ์ ซึ่งจะแบ่งการพัฒนาระบบออกเป็น 4 ส่วนได้แก่ ฐานข้อมูล แอปพลิเคชัน เว็บไซต์ และ API โดยแต่ละส่วนมีรายละเอียดในการพัฒนาระบบดังนี้

### 3.1 การสร้างฐานข้อมูล

ฐานข้อมูลของระบบจะมีอยู่ 2 ส่วนคือส่วนที่เป็น Relational database และ Non-relational data แบ่งข้อมูลออกเป็นสองส่วน ส่วนของการเก็บข้อมูลทั่วไปจะเก็บอยู่บน SQL และข้อมูลในส่วนของแผนที่จะถูกเก็บอยู่บน Firebase Realtime Database โดยเก็บข้อมูลในรูปแบบของ JSON และเมื่อนำไปใช้งานร่วมกับการพัฒนาแอปพลิเคชันในระบบปฏิบัติการ Android จะมีฟังก์ชัน onChange ที่คอยอัปเดตแบบเรียลไทม์ ทุกครั้งที่ทำการอัปเดตพื้นที่ทำงาน ทางแอปพลิเคชันก็จะแสดงผลพื้นที่ทำงานขึ้นมาทันทีโดยไม่ต้องปิดแอปพลิเคชันและเริ่มต้นทำงานใหม่ ซึ่งประเภทของฐานข้อมูลแต่ละอันมีข้อดีและข้อเสียที่แตกต่างกันออกไปตามโครงสร้างและเทคโนโลยี ดังนั้นการเลือกประเภทของฐานข้อมูลที่เหมาะสมกับการใช้งานจึงเป็นเรื่องที่สำคัญ

### 3.2 การพัฒนา Mobile Application

ทำการพัฒนาแอปพลิเคชันลงบนระบบปฏิบัติการ Android โดยใช้ภาษา Java ในการพัฒนา โดยพัฒนาให้มีความสามารถในการบันทึกเวลาเข้าทำงาน โดยจะผ่าน

การตรวจสอบเงื่อนไข คือ ตรวจสอบตำแหน่งปัจจุบันที่ผู้ใช้งานนั้นทำการบันทึกอยู่ในขอบเขตที่ทำการที่ตั้งไว้หรือไม่ และจะต้องยืนยันตัวตนด้วยรหัส One Time Password ของระบบ เพื่อตรวจสอบว่าเป็นผู้ใช้งานจริง และหลังจากการตรวจสอบจนครบเงื่อนไขที่กำหนดแล้วสามารถเข้าถึงการบันทึกเวลาของระบบ ซึ่งระบบจะทำการเก็บ Location ที่เป็น Latitude และ Longitude เวลาตามระบบและข้อมูลพนักงานที่จำเป็นในการใช้อ้างอิงการบันทึกเข้างาน ลงในฐานข้อมูลที่เตรียมไว้ข้างต้นโดยผ่าน APIs

### 3.3 การพัฒนา Web Application

ในส่วนของการพัฒนาเว็บไซต์ของผู้ดูแลระบบจะพัฒนาโดยใช้ภาษา HTML5 สำหรับพัฒนา และใช้ภาษา PHP ในการติดต่อกับฐานข้อมูล เว็บไซต์จะมีเมนูในการเปิดเข้าไปยัง Google Maps โดยใช้บริการของ Google Cloud Platform เพื่อทำการกำหนดขอบเขตสถานที่ทำงานในรูปแบบของ Polygon ซึ่งปัจจุบัน Google Map APIs นั้นสามารถแสดง Polygon ลงบนแผนที่โดยใช้ภาษา JavaScript ในการควบคุมการสร้างอ้างอิงการวาดจากพิกัดจุด X,Y (LAT, LNG) พื้นที่ปฏิบัติงานที่ทำการศึกษาและบันทึกลงในระบบพร้อมกับถ่ายโอนข้อมูลไปยังระบบบันทึกเวลาเข้างานทันทีโดยเรียกวิธีการนำเข้าสู่ข้อมูลของแผนที่ตรงนี้ว่าการ Digitization พร้อมทั้งพัฒนาให้ระบบใช้ในการตรวจสอบรายชื่อพนักงาน ดูรายการบันทึกทั้งหมดเพื่อนำออกเป็นรายงานหรือตรวจสอบในแต่ละรายการที่มีการบันทึกเวลา โดยเว็บไซต์จะนำขึ้นไปวางบน Server ของค่าย Heroku: Cloud Application Platform

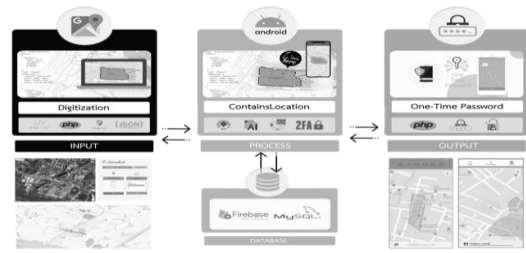
### 3.4 การพัฒนา RESTful API และการส่ง OTP

จากงานทั้งสามส่วนที่ถูกพัฒนาขึ้นมา จำเป็นต้องมีตัวกลางในการเชื่อมต่อข้อมูลทั้งหมดเข้าด้วยกัน จึงพัฒนา API โดยใช้ NodeJS ขึ้นมาใช้ในการเชื่อมต่อนฐานข้อมูลกับระบบในทุกส่วน

ในการส่ง OTP ไปยัง Email ของพนักงาน จะใช้บริการของ SendGrid สำหรับส่ง Email โดยการส่ง

Email เริ่มต้นจากเรียกฟังก์ชันและทำการดึงค่า OTP ที่ถูกจัดเก็บอยู่ที่ฐานข้อมูลส่งไปยังปลายทาง

### 3.5 กรอบแนวคิดของงานวิจัย



ภาพที่ 3: ภาพแสดงกรอบแนวคิดของงานวิจัย

หลังจากพัฒนาระบบทุกส่วนจะสามารถนำมาจัดวางเป็นกรอบแนวคิดโดยการใช้งานจะเริ่มจากส่วนของผู้ดูแลระบบทำการนำเข้าสู่พื้นที่ทำงานผ่านทางเว็บไซต์ หลังจากนั้นเป็นส่วนของผู้ใช้แอปพลิเคชันในการบันทึกเวลาเข้างาน โดยผ่านการตรวจสอบพื้นที่ จากนั้นทำการยืนยันตัวตน ซึ่งระบบทั้งหมดจะใช้ข้อมูลพื้นฐานข้อมูลจุดเดียวกันที่แบ่งฐานข้อมูลออกเป็นสองส่วนใช้งานผสมกัน โดยมีหน้าที่และจุดประสงค์ที่แตกต่างกันไป พร้อมทั้งมีการพัฒนา APIs เป็นเส้นทางในการเชื่อมต่อเพื่อส่งผ่านข้อมูลจากฐานข้อมูลไปยังทุกส่วนของระบบ

## 4. ผลการทดสอบ

จากการพัฒนาระบบลงเวลาเข้าทำงานที่นำเอาการคัดแยกพื้นที่ และการใช้ 2FA เข้ามาประยุกต์นั้นจะใช้ค่าความแม่นยำในการทดสอบระบบ โดยการสร้างพื้นที่จำลองในการสร้างพื้นที่ที่เปรียบเสมือนสถานที่ทำงานเพื่อทำให้ง่ายต่อการทดสอบระบบ ดังนั้นการทดสอบระบบจึงทำการทดสอบได้ 2 ส่วน คือ ส่วนของเจ้าหน้าที่ที่จัดการเรื่องพื้นที่ของระบบและส่วนของพนักงานที่ใช้ระบบในการบันทึกเวลาเข้าทำงาน ซึ่ง ระบบลงเวลาเข้าทำงานที่พัฒนาขึ้นจะมีการวัดประสิทธิภาพการทำงานด้วยค่าความถูกต้อง (Accuracy)

ตารางที่ 1 : ตารางบันทึกผลการทดสอบการบันทึก

CASE	ผลการทดสอบ	เวลา OTP	เวลารวม
ทดสอบ 1	สำเร็จ	< 5s	1m10s
ทดสอบ 2	สำเร็จ	< 5s	40s
ทดสอบ 3	สำเร็จ	< 5s	55s
ทดสอบ 4	สำเร็จ	< 5s	1m20s

CASE	ผลการทดสอบ	เวลา OTP	เวลารวม
ทดสอบ 5	สำเร็จ	< 5s	35s
ทดสอบ 6	สำเร็จ	< 5s	1m30s
ทดสอบ 7	สำเร็จ	< 5s	50s
ทดสอบ 8	สำเร็จ	< 5s	55s
ทดสอบ 9	ไม่สำเร็จ	> 60s	2m
ทดสอบ 10	สำเร็จ	< 5s	1m10s
ค่าเฉลี่ย		<10.5s	66.5s

ในการทดสอบระบบ ได้ทำการทดสอบจำนวน 10 ครั้ง โดยแต่ละครั้งเป็นคนละสถานที่ มีปัจจัยที่ต่างกัน เช่น ความเร็วของอินเทอร์เน็ตและสัญญาณโทรศัพท์ สถานที่ การรับสัญญาณ GPS ซึ่งเป็นการจำลองที่อิงการใช้งานจริงเป็นหลัก

ผลการทดสอบพบว่าในการทดสอบระบบบันทึกเวลาสามารถทำงานสำเร็จได้ 9 จาก 10 ครั้งหรือคิดเป็น 90 เปอร์เซ็นต์ แต่เวลาที่ใช้ในการทดสอบจะแตกต่างกันออกไปโดยเวลาเฉลี่ยในการได้รับ OTP ประมาณ 5 วินาที เนื่องจากในการส่ง OTP ไปยัง Email นั้นเกิดการหน่วงเวลาในบางกรณีหรือเหตุการณ์ที่อุปกรณ์หาตำแหน่งในสถานที่อับสัญญาณ จึงใช้เวลามากกว่าสถานที่โล่งแจ้ง ทำให้มีผลทำให้เวลาในการทำงานนั้นแตกต่างกันออกไป และเวลารวมเฉลี่ยที่ทำได้ออกคือ 66.5s หรือประมาณ 1 นาที 7 วินาที แต่อย่างไรก็ตามเวลาที่ใช้ในการลงบันทึกนั้นยังอยู่ในขอบเขตที่กำหนดไว้จึงทำให้ไม่มีผลกระทบต่อการใช้งานและยังใช้งานได้อย่างมีประสิทธิภาพ

จากการทดสอบระบบในส่วนของผู้ใช้พบว่าระบบสามารถจัดการขอบเขตของพื้นที่ได้อย่างถูกต้องและแม่นยำ โดยทำการทดสอบหรือใช้งานผ่านทาง Web Application บน Google Maps APIs ที่ถูกพัฒนาขึ้นตรงตามจุดประสงค์

## 5. ผลสรุปและข้อเสนอแนะ

จากงานวิจัยระบบการจัดการเวลาเข้าทำงานในพื้นที่พัฒนาโดยการนำขั้นตอน 2FA และความสามารถในการกำหนดขอบเขตพื้นที่ ซึ่งจากการทดสอบจำนวน 10 ครั้ง

ที่มีปัจจัยที่ต่างกัน ได้ค่าความถูกต้องของผลลัพธ์ที่ 90% ดังนั้น งานวิจัยนี้สามารถนำไปใช้งานเป็นระบบการบันทึกเวลาเข้างานที่มีความปลอดภัยและได้ผลการเก็บบันทึกที่มีความเชื่อถือได้มากขึ้นเมื่อเทียบกับระบบเดิม อย่างไรก็ตามระบบนี้ไม่อาจช่วยในเรื่องของความสะดวกสบายหรือความรวดเร็วในการบันทึกเวลาเข้างาน แต่เป็นการช่วยในเรื่องของการตรวจสอบ และการจัดการข้อมูลของฝ่ายบุคคลมากกว่า โดยการนำระบบมาใช้งาน ควรมีแผนการเปลี่ยนแปลงควบคู่ไปกับการใช้ระบบเดิมที่ใช้งานอยู่ในปัจจุบันเพราะหากนำระบบนี้มาทดแทนทันทีอาจเกิดการปรับตัวที่มีมากเกินไปสำหรับพนักงาน และส่งผลให้เกิดข้อผิดพลาดจากผู้ใช้งาน (User Error) ได้มากยิ่งขึ้น และเมื่อถึงเวลาที่เหมาะสมระบบจะถูกแทนที่อย่างสมบูรณ์

## เอกสารอ้างอิง

- [1] Shermin Sultana, Asma Enayet and Ishrat Jahan Mouri (2558) "A Smart, Location Based Time And Attendance Tracking System Using Android Application"
- [2] Mohammad Salah Uddin, Member, IACSIT, S. M. Allayear, N. C. Das, and F. A. Talukder (2557) "A Location Based Time and Attendance System"
- [3] J.Chandramohan, R.Nagarajan, M. Ashok kumar, T.Dineshkumar, G.Kannan, R.Prakash (2560) "Attendance Monitoring System of Students Based on Biometric and GPS Tracking System"
- [4] Yew Kwang Hooi, Khairul Shafee Kalid and Serdarmammet Tachmammedov (2561) "Multi-Factor Attendance Authentication System"
- [5] Benfano Soewito, Ford Lumban Gaol, Echo Simanjuntak, Fergyanto E. Gunawan (2558) "Attendance system on Android smartphone"
- [6] Md. Milon Islam, Md. Kamrul Hasan, Md Masum Billah, Md. Manik Uddin (2561) "Development of smartphone-based student attendance system"
- [7] Tinthid Jaikla, Sasakorn Pichetjamroen, Chalee Vorakulpipat, Achara Pichetjamroen (2563) "A Secure Four-factor Attendance System for Smartphone Devic."

# การพัฒนาหุ่นยนต์อัตโนมัติเพื่อสนับสนุนการทำงานในส่วนงานจัดซื้อ และส่วนงานบัญชีเจ้าหนี้

## The Development of Robotic Process Automation System for Support Process of Purchasing and Accounting Payable

กนต์ธร อินทร์ประสิทธิ์ (Kontorn Inprasith)<sup>1</sup> ศักดิ์ชาย ตังวรณวิทย์ (Sakchai Tangwannawit)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

S6307021858295@email.kmutnb.ac.th, sakchai.t@itd.kmutnb.ac.th

### บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาหุ่นยนต์อัตโนมัติเพื่อสนับสนุนการทำงานในส่วนงานจัดซื้อและส่วนงานบัญชีเจ้าหนี้ งานวิจัยนี้ประกอบด้วย 1) เพื่อวิเคราะห์และสังเคราะห์กระบวนการทำงานขององค์กร โดยวัดผลจากการนำข้อมูลกระบวนการทำงานขององค์กรมาวิเคราะห์ว่าจุดไหนคือจุดที่ทำให้กระบวนการทำงานล่าช้าและต้องใช้ทรัพยากรบุคคลเป็นจำนวนมาก และ 2) เพื่อเปรียบเทียบประสิทธิภาพการทำงานของระบบหุ่นยนต์อัตโนมัติและการทำงานของพนักงานในองค์กร ระบบหุ่นยนต์อัตโนมัติช่วยลดการทำงานที่ซ้ำซากและกระบวนการที่เป็นกิจวัตรให้น้อยลงทำให้พนักงานในองค์กรสามารถนำเวลาไปพัฒนาและดำเนินงานในส่วนอื่นๆได้มากขึ้น โดยวัดผลจากการนำ Transaction และ เวลาในการดำเนินการในแต่ละกระบวนการมาทำการวิเคราะห์และเปรียบเทียบประสิทธิภาพของหุ่นยนต์กับพนักงานในองค์กร ดังนั้นผู้วิจัยจึงมีแนวคิดที่จะนำเทคโนโลยี Robotic Process Automation มาช่วยแก้ปัญหาดังกล่าว โดยขั้นแรกเริ่มจากการรวบรวมข้อมูลและวิเคราะห์หลักการการทำงานของพนักงานในส่วนงานจัดซื้อและส่วนงานบัญชีเจ้าหนี้ขององค์กร ขั้นที่สองเพื่อวิเคราะห์ ปัญหาในการทำงานเหตุผลที่ทำให้การทำงานไม่ได้ประสิทธิภาพและเกิดข้อผิดพลาดในการทำงานจึงพบว่าการทำงานในส่วนงานของหน่วยงานจัดซื้อในส่วนของการเปิดเอกสารใบคำขอซื้อจากใบคำสั่งซื้อนั้นเป็นระบบงาน ที่มีความเป็นกิจวัตรมีขั้นตอนในระบบ

ERP ที่เหมือนเดิม และ ในส่วนงานของหน่วยงานบัญชีเจ้าหนี้ที่มี การดำเนินการตั้งหนี้หลังจากหน่วยงาน Receiving รับของแล้ว ซึ่งวิธีการดำเนินงานของทั้งสองส่วนเป็นขั้นตอนที่ Robotic Process Automation สามารถเข้ามาทำงานทดแทนได้ และยังสามารถทำให้ระบบห่วงโซ่อุปทานขององค์กรสามารถทำงานอย่างอัตโนมัติ ซึ่งผลลัพธ์จากงานวิจัยช่วยแก้ปัญหาในส่วนนี้ได้ นอกจากจะเพิ่มประสิทธิภาพการทำงานได้แล้วยังช่วยลดต้นทุนและทรัพยากรขององค์กรได้อย่างมาก

**คำสำคัญ:** หุ่นยนต์อัตโนมัติ จัดซื้อ บัญชีเจ้าหนี้

การวางแผนทรัพยากรองค์กร

### Abstract

The purpose of this research was to develop robotic process automation (RPA) system to support the process of purchasing and accounting payable. This research consisted of 1) analysis and synthesis of the process in the organization, 2) comparing the performance of RPA and employee work. The RPA can reduce repetitive work and routine processes to help employees in the organization take more time to develop and operate in other works. Therefore, the researcher came up with an idea to use RPA technology to help solve the problem. The first step to data collection and analyzing the working principles of employees in the purchasing division and account payable division. The second step was to analyze problems in the work for the reason they

*work is not effective and there is an error in the work, The results of the analysis found that 1) the purchase requisition document from the purchase order has a routine with the same steps in the ERP system, and 2) the department of accounts payable that has indebtedness operation after the RCV has received the goods. The two parts of the method of operation are the steps that the RPA can take over. It can also make the organization's supply chain system become fully automated as well. The results of this research found that solving this problem, in addition to increasing the efficiency of work, can greatly reduce the cost and resources of the organization.*

**Keywords:** Robotic Process Automation, Purchasing, Accounting Payable, Enterprise Resource Planning.

## 1. บทนำ

ในยุคปัจจุบัน การใช้เทคโนโลยีเข้ามามีส่วนร่วมกับการดำเนินงานในทุกๆ ส่วนงานมีความสำคัญไม่แพ้การดำเนินธุรกิจในส่วนอื่นๆ โดยเฉพาะอย่างยิ่งในอุตสาหกรรมการผลิต และ อุตสาหกรรมบริการขนส่งซึ่งได้นำเทคโนโลยีมาใช้ในการเพิ่มประสิทธิภาพในการทำงานหรือวัดค่าความถูกต้องความคุ้มค่า เพื่อเข้ามาช่วยลดความเสียหาย และ ลดต้นทุนในการผลิตสินค้า ดังนั้นจะเห็นได้ว่า บริษัททั้งในและต่างประเทศมีการเริ่มใช้งานระบบต่างๆ และ อุปกรณ์เทคโนโลยีต่างๆ เข้ามามีบทบาทในการดำเนินงานในปัจจุบันถ้าจะมองย้อนหลังไปในช่วงปี 1990 เป็นต้นมาอุตสาหกรรมการผลิตทั่วโลก ได้มีการวางแผนการใช้ทรัพยากรที่มีอยู่ขององค์กร ซึ่งเรียกว่า ERP(Enterprise Resource Planning) โดยแนวคิดมาจากการพัฒนาระบบการบริหารการผลิตรวม (Material Requirement Resource Planning / Manufacturing Resource Planning, MRP System) ของอุตสาหกรรมผลิตในอเมริกา โดยคำว่า ERP และแนวคิดของ ERP นั้นก็พัฒนามาจาก MRP นั่นเอง และได้มีการวิวัฒนาการมาเรื่อยๆ [1]

โดยปัจจุบัน การบริหารทรัพยากรขององค์กร ERP(Enterprise Resource Planning) ต้องมีการพัฒนา

ระบบงานให้สามารถตอบสนองความต้องการของผู้ใช้งาน และตอบสนองอุตสาหกรรมยุค 4.0[7][14] ได้มากยิ่งขึ้น โดยแต่ละองค์กร ก็เลือก ERP ที่ใช้งานแตกต่างกันออกไป แต่ในงานวิจัยนี้ ได้เลือกระบบ ERP(SAP) มาใช้งานเนื่องจากเป็นระบบ ERP ที่มีขนาดใหญ่มีผู้ใช้งานทั่วโลก และมีขีดความสามารถและประสิทธิภาพที่ได้มาตรฐานในระดับสากล ซึ่งในปัจจุบันมีระบบที่มีความน่าสนใจและช่วยเพิ่มขีดจำกัดความสามารถของ ERP ขึ้นไปอีกขั้น ก็คือ RPA ซึ่งเป็นเครื่องมือที่ใช้งานในงานวิจัยที่ถูกพัฒนาจากทางบริษัท SAP เองเลยทำให้มีความสามารถในการเชื่อมต่อและพัฒนาได้สะดวกและรวดเร็วขึ้นเนื่องจากอยู่ในแพลตฟอร์มเดียวกัน ซึ่งทั้งนี้ระบบ RPA ก็จะมาช่วยให้การทำงานของโรงงานอุตสาหกรรมมีความต่อเนื่องของกระบวนการทำงานและลดข้อผิดพลาดในการทำงานลงได้ ทำให้ระบบห่วงโซ่อุปทาน (Supply Chain) ขององค์กรมีความเป็นอัตโนมัติมากขึ้นจากที่เป็นอยู่และจากการศึกษา

## 2. วรรณกรรมที่เกี่ยวข้อง

### 2.1 ERP (Enterprise Resource Planning)

ERP คือ การวางแผนทรัพยากรทางธุรกิจขององค์กรโดยรวม เพื่อให้เกิดการใช้ประโยชน์ได้อย่างสูงสุดของทรัพยากรทางธุรกิจขององค์กร คือระบบที่ใช้ในการจัดการและวางแผนการใช้ทรัพยากรต่างๆ ขององค์กร โดยเป็นระบบที่เชื่อมโยงระบบงานต่าง ๆ ขององค์กรเข้าด้วยกันเช่น หากเป็น ERP ของบริษัทจะหมายรวมถึงตั้งแต่ระบบงานทางด้านบัญชี และการเงิน ระบบงานทรัพยากรบุคคล ระบบบริหารการผลิตรวมถึงระบบการกระจายสินค้าเพื่อช่วยให้การวางแผนและบริหารทรัพยากรของบริษัทนั้นเป็นไปอย่างมีประสิทธิภาพทั้งยังช่วยลดเวลาและขั้นตอน (Algorithm) การทำงานได้อีกด้วย [2]

### 2.2 RPA (Robotic Process Automation)

RPA คือ โปรแกรมที่ช่วยให้ธุรกิจต่างๆ สามารถสร้างหุ่นยนต์หรือ Bot ขึ้นมาทำงานต่างๆ ตาม Pattern ที่กำหนดเอาไว้ได้ โดย Bot แต่ละตัวที่สร้างขึ้นมานั้นก็จะ

ทำงานได้ตาม Pattern ที่ถูกกำหนดเอาไว้แตกต่างกันไป เพื่อนำไปใช้ในการทำงานซ้ำๆ ในรูปแบบที่แตกต่างกัน



ภาพที่ 1: Robotic Process Automation

### 2.3 ระบบบัญชีเจ้าหนี้ (Account Payable: AP)

AP คือกระบวนการทางบัญชีหลังมีการซื้อ สินค้าหรือ บริการและเมื่อหนี้ครบกำหนด ส่งรายการหนี้ให้ฝ่าย การเงิน เพื่อดำเนินการขั้นตอนต่างๆทางการเงิน ตั้งแต่รับ วางบิลเตรียมจ่ายให้เจ้าหนี้ จ่ายโดยเงินสด เช็คและการ โอนเงินผ่านธนาคารควบคุมการจ่ายเงินชำระหนี้ให้เจ้าหนี้ อย่างถูกต้องรวมถึงช่วยในการจัดการเกี่ยวกับ ข้อมูลต่างๆ (Data) ของเจ้าหนี้ เช่น วงเงินเครดิต (Credit Limit), เงื่อนไขการชำระเงิน (Terms of Payment) ข้อกำหนด เกี่ยวกับการซื้อสินค้าต่างๆช่วยให้ทราบถึงสถานะของ หนี้สิน (Aging) ที่มีอยู่เป็นต้น [3]

### 2.4 การจัดซื้อ (Purchasing: PUR)

PUR คือกระบวนการที่มีหน้าที่รับผิดชอบในการจัดหา วัตถุดิบต่างๆ ตามความต้องการของหน่วยงานต่างๆ ของ องค์กรและเป็นกระบวนการที่เชื่อมโยงระหว่างองค์กร ต่างๆ ใน Supply Chain และทำหน้าที่ในการประสานงาน ในด้านการไหลของข้อมูลระหว่างองค์กรและผู้ขาย วัตถุดิบ ดังนั้นกระบวนการจัดซื้อ (Procurement) จึงมี ความสำคัญต่อระดับบริการลูกค้าทางการขนส่ง (Logistics customer service) รวมถึงหน้าที่ในการออก เอกสาร ใบคำขอซื้อ(Purchase Requirement : PR) และใบ คำสั่งซื้อ(Purchase Order : PO) เป็นต้น[4]

### 2.5 แนวคิดเกี่ยวกับระบบซอฟต์แวร์อัตโนมัติ

(Robotic Process Automation :RPA)นิยามและความหมาย ของ RPA

Decker et al. (2018) กล่าวถึงระบบซอฟต์แวร์อัตโนมัติ (Robotic Process Automation: RPA) ว่าเป็นเทคโนโลยีขั้น พื้นฐานที่เกี่ยวข้องกับกระบวนการทางธุรกิจแบบใช้ พนักงานเป็นผู้ดำเนินการในระดับหน้าที่มากที่สุด

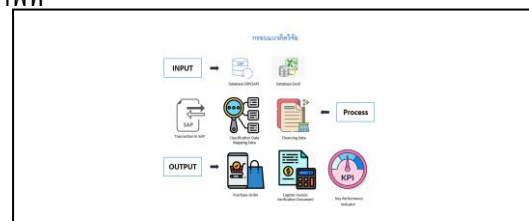
Bruno et al. (2017) กล่าวว่า ระบบซอฟต์แวร์อัตโนมัติ หรือ RPA (Robotic Process Automation) คือ ระบบ ซอฟต์แวร์ที่มีความสามารถในการลอกเลียนแบบ พฤติกรรมการใช้งานแพลตฟอร์มทางเทคโนโลยีของ มนุษย์[5]

### 2.6 ตัวอย่างและกรณีศึกษาของการนำ RPA ไปใช้ใน บริษัทต่างๆ

2.6.1 ตัวอย่างการนำ RPA ไปใช้ในบริษัทด้าน พลังงานแห่งหนึ่ง UTILITY (Lacity et al., 2015) บริษัท UTILITY เป็นบริษัทที่ประกอบธุรกิจด้านการจัดหาและ ผลิตกระแสไฟฟ้าต้องการที่จะลดต้นทุนแทนที่การขึ้น ราคาค่าไฟฟ้า และต้องการที่จะปรับปรุงคุณภาพการ บริการจึงนำ RPA เข้ามาใช้ตั้งแต่กลางปี ค.ศ.2000 ตัวอย่าง ของการนำ RPA ไปใช้ได้ประสบความสำเร็จคือ กระบวนการอ่านค่ามิเตอร์การใช้ไฟฟ้าและตรวจสอบ ความถูกต้อง ซึ่งมีปริมาณรายการมากกว่า 1,000 รายการ กระบวนการนี้จะต้องใช้พนักงานจำนวนมาก และมี ข้อผิดพลาดอยู่เป็นประจำหลังจากใช้ RPA บริษัทสามารถ ลดจำนวนพนักงานได้ถึง 25-30 คนลดต้นทุนในส่วน ของ พนักงานประจำได้ถึง 60% และกลายเป็นส่วนหนึ่งของกล ยุทธ์การใช้เทคโนโลยีช่วยเพิ่มความพึงพอใจให้กับลูกค้า [6]

### 3. วิธีดำเนินงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบอัตโนมัติให้ เข้ากับการทำงานขององค์กรและเพื่อเปรียบเทียบ ประสิทธิภาพการทำงานของ ระบบอัตโนมัติและการ ทำงานของพนักงานในองค์กรได้กำหนดแนวทางในการ วิจัยโดยมีรายละเอียดและขั้นตอนการดำเนินงานแสดงดัง ภาพที่



ภาพที่ 2: กระบวนการวิจัย

จากภาพที่ 2 แสดงกรอบแนวคิดของการวิจัย กระบวนการทำงานแบ่งออกเป็น 3 ส่วน ส่วนที่ 1 เป็นการ

รวบรวมข้อมูลและเตรียมข้อมูลก่อนประมวลผล เป็นข้อมูลเริ่มต้นจากฐานข้อมูล (PR Pending , Receive Billing) ส่วนที่ 2 การสร้างระบบเพื่อทำการตรวจสอบค่าความถูกต้องจำแนกหมวดหมู่และทำการเดินรายการในระบบ SAP ให้เกิดค่าเฉลี่ยของความผิดพลาดน้อยที่สุดและได้ค่าประสิทธิภาพสูงสุด ส่วนที่ 3 เป็นการสร้างเอกสาร Purchase Order และ Liv Document บนระบบ SAP [7]

**3.1 การเตรียมข้อมูล**

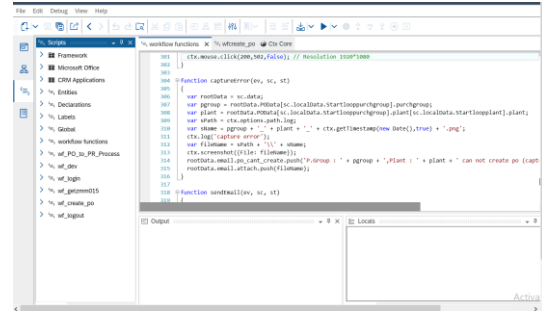
ในการศึกษาค้นคว้าได้ทำการใช้ข้อมูลในการทดสอบของบริษัทแห่งหนึ่ง โดยเป็นข้อมูล ในปี พ.ศ.2564 ในช่วงเดือน พฤศจิกายน-ธันวาคม ข้อมูลที่ใช้งานประกอบด้วย 1.เอกสารใบคำสั่งซื้อ 2.เอกสารการตั้งหนี้ โดยใช้โปรแกรม Microsoft Excel ในการจัดเก็บข้อมูลตรวจสอบความสมบูรณ์ของข้อมูลหลังจากนำข้อมูลออกมาจากระบบ SAP โดยข้อมูลที่เลือกมาใช้คือข้อมูลทั้งหมดของเดือนดังกล่าวโดยจะเอาข้อมูลมาทำการวิเคราะห์และทำการเปรียบเทียบระหว่างการทำงานของ RPA และพนักงานขององค์กร[8]และวิเคราะห์ปัญหาที่เกี่ยวข้องกับงานวิจัยศึกษาปัญหาและความต้องการของระบบ

Attribute
1.เอกสารใบคำสั่งซื้อ
2.เอกสารการตั้ง
3.กระดานราคาของจัดซื้อ

ภาพที่ 3: แอตทริบิวต์ที่ใช้ในการวิเคราะห์ประสิทธิภาพ

**3.2 การพัฒนาระบบ**

ทำการพัฒนาระบบ โดยใช้โปรแกรมที่ชื่อว่า Desktop Agent และ Studio โดยเป็นการเขียน โปรแกรมโดยใช้ภาษา JavaScript และ ABAP ในส่วนของ ERP(SAP) เพื่อให้ระบบอัตโนมัติมีประสิทธิภาพสูงสุดและลดความผิดพลาดให้เกิดขึ้นน้อยที่สุด



ภาพที่ 4: หน้าต่างในการพัฒนาระบบ(2)

**3.3 การเปรียบเทียบประสิทธิภาพ**

เมื่อได้ผลการวิเคราะห์การทำงานของระบบอัตโนมัติที่ต้องการแล้ว ก่อนที่จะนำผลลัพธ์ไปใช้งานจะต้องมีการเปรียบเทียบ ประสิทธิภาพ ของผลลัพธ์ที่ได้ว่ามีประสิทธิภาพมากน้อยเพียงใด โดยพิจารณา จากจำนวนการสร้างเอกสารที่เท่ากันและเปรียบเทียบจากระยะเวลาและความถูกต้องของเอกสารที่ถูกสร้างขึ้นในระบบ SAP

**3.4 การทดสอบความถูกต้องของระบบอัตโนมัติ**

ทวนสอบความถูกต้องของระบบอัตโนมัติ โดยนำระบบอัตโนมัติมาประเมินความสามารถในการนำไปใช้กับข้อมูลจริงและหาค่าประสิทธิภาพของการใช้งานว่าอยู่ในเกณฑ์ที่คาดหวังได้

**3.5 แผนการดำเนินงาน**

เริ่มต้นตั้งแต่วันที่ 1 กรกฎาคม 2021 ถึง 31 มีนาคม 2022

แผนการดำเนินงาน  
แผนการดำเนินงาน ตั้งแต่ 01 July 2021 ถึง 31 March 2022

ช่วงเวลา	Jul 2021- Aug 2021	Sep 2021- Oct 2021	Nov 2021- Dec 2021	Jan 2022- Feb 2022	Mar 2022
1 ศึกษาและวิเคราะห์ปัญหา	✓				
2 ศึกษาขั้นตอนการพัฒนา	✓				
3 ออกแบบรูปแบบการทำงานและแนวคิด		✓			
4 วางแผนและพัฒนาระบบ		✓	✓		
5 ทดสอบระบบ			✓		
6 นำข้อมูลที่ยอมรับหรือใช้งานมาวิเคราะห์				✓	
7 สรุปผล				✓	
8 จัดทำรูปเล่ม				✓	✓
9 ปรับปรุงแก้ไขตามข้อเสนอแนะ					✓

ภาพที่ 5: แผนการดำเนินงาน

#### 4. ผลการดำเนินงาน

จากการศึกษาเรื่อง

“การพัฒนาระบบอัตโนมัติเพื่อสนับสนุนการทำงานในส่วนงานจัดซื้อและส่วนงานบัญชีเจ้าหนี้”

สามารถสรุปผล การศึกษาได้ดังนี้

4.1 ผลการเปรียบเทียบประสิทธิภาพ RPAและคน

4.2 ผลการนำระบบอัตโนมัติไปใช้

4.3 ความได้เปรียบทางการแข่งขันที่คาดว่าจะได้รับ

จากการนำระบบซอฟต์แวร์อัตโนมัติมาใช้งานองค์กร ดังตารางที่ 1 โดยเป็นข้อมูล Transaction การเปิดเอกสาร Purchase Order ,LIV ของหน่วยงานจัดซื้อจัดจ้างและบัญชีขององค์กรหนึ่ง ที่มีบริษัทในเครืออยู่ 12 องค์กรย่อย ทำให้มีจำนวน Transaction ของเอกสารในจำนวนที่สูง โดยจะเห็นว่าใน 2 เดือน มีการออกเอกสาร

ทั้งหมด 2,027 (PO) และ 3,510 (LIV)

Transaction 694 (PO) และ 3,510 (LIV) Document การเปิด Manual ของพนักงานในองค์กรจะเฉลี่ยการเปิดเอกสารหนึ่งใบ ในระบบ SAP ใช้เวลาประมาณ 10 และ 7 นาทีต่อฉบับ แต่หากใช้งาน Robot จะสามารถทำได้ในเวลาเฉลี่ยโดยประมาณ 3 และ 2 นาทีต่อฉบับ ทำให้ ยอดรวม ตามตาราง

จะเห็นว่า หากพนักงาน ในองค์กรเปิดเอกสารเอง จะทำให้กินเวลาไปทั้งสิ้น 6,940 (PO) และ 24,570 (LIV) นาที หรือประมาณ 115 (PO) และ 409 (LIV) ชั่วโมงจากเวลาทำงานทั้งหมดทำให้เวลาในช่วงนี้ทางผู้วิจัยเห็นว่า หากพัฒนาตัวระบบอัตโนมัติเข้ามาช่วยจะทำให้พนักงานในองค์กรสามารถนำเวลาประมาณ 115 และ 409 ชั่วโมงต่อเดือนไปพัฒนางานให้องค์กรในส่วนอื่น ๆ ได้

โดยจากข้อมูลที่น่ามาวิเคราะห์ ตัว Robot สามารถทำงานได้โดยใช้เวลาแค่ 2,082 (PO) และ 7,020 (LIV) นาที หรือประมาณ 34 (PO) และ 117 (LIV) ชั่วโมงเท่านั้น จะเห็นว่าสามารถประหยัดเวลาได้ประมาณ 4 เท่า [9]

4.1 ผลการเปรียบเทียบประสิทธิภาพของRPAและคน

จากการทดสอบประสิทธิภาพพบว่า ระบบ RPA สามารถทำงานได้มีประสิทธิภาพกว่าคน และมีค่าของความคิดพลาดน้อยกว่าอีกด้วย

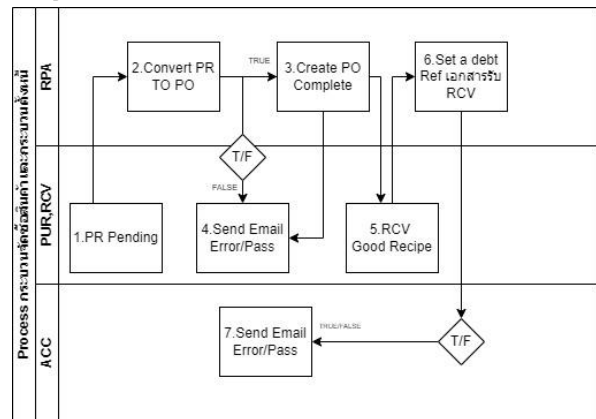
ตารางที่ 1: ผลวัดประสิทธิภาพของระบบอัตโนมัติในส่วน  
ของหน่วยงานจัดซื้อ (Process of Purchasing)และระบบ  
อัตโนมัติในส่วนของหน่วยงานบัญชีเจ้าหนี้ (Process of  
Accounting Payable)

Process	PR TO PO Automation	LIV Automation
Transaction	2,027	3,510
Document	694	3,510
Human Error(Document)	9	17
Robot Error(Document)	0	0
Time Human	10	7
Time RPA	3	2
SummaryHuman(min)	6,940	24,570
SummaryRPA(min)	2,082	7,020

4.2 ผลการนำระบบอัตโนมัติไปใช้

หลังจากนำระบบอัตโนมัติ ที่ทำการพัฒนาขึ้นมาไปใช้ พบ ว่าค่าความผิดพลาดในการเดินรายการไม่เกิดขึ้นเลยในระยะเวลา 2 เดือนที่นำข้อมูลมาวิเคราะห์และตัว RPA สามารถใช้ระยะเวลาที่น้อยกว่าพนักงานในองค์กรได้ถึง 4 เท่าตัว (วิเคราะห์จากตารางที่ 1)

4.3 รูปภาพในส่วนของฝั่งงานแสดงกระบวนการ



ภาพที่ 6: Swim Lane แสดงกระบวนการ

#### 5. สรุปผลและอภิปราย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบอัตโนมัติให้เข้ากับการทำงานขององค์กรและเพื่อเปรียบเทียบประสิทธิภาพการทำงาน ของระบบอัตโนมัติและการทำงานของพนักงานในองค์กร ซึ่งพิจารณาจากข้อมูลการ



เปรียบเทียบการทำงานระหว่างบุคลากรกับตัว RPA พบว่า การดำเนินการทำงานและ ข้อผิดพลาดและระยะเวลาในการปฏิบัติการของ RPA มีค่าเฉลี่ยของระยะเวลาและความผิดพลาดน้อยที่สุด

ดังนั้นการวิเคราะห์ และเปรียบเทียบประสิทธิภาพดังกล่าวรวมถึงความพึงพอใจของผู้ใช้งาน สามารถนำไปประยุกต์เพื่อพัฒนาระบบอัตโนมัติอื่นๆที่จะสามารถมาช่วยพัฒนาและลดค่าใช้จ่ายรวมถึงระยะเวลาการทำงานให้แก่องค์กรได้ในอนาคตอีกด้วย

**5.1 เพื่ออธิบายถึงประสิทธิภาพการทำงานของระบบและวัตถุประสงค์ โดยทางผู้วิจัยได้นำแบบประเมินส่งให้ทางผู้เชี่ยวชาญ จำนวน 5 ท่านในการประเมินระบบอัตโนมัติที่ถูกพัฒนาขึ้น โดยแบ่งได้ 3 หัวข้อใหญ่ๆดังนี้**

5.1.1 ด้านกระบวนการให้บริการ

5.1.2 ด้านวัตถุประสงค์และประโยชน์ที่ได้รับ

5.1.3 ด้านคุณภาพของระบบซึ่งจากผลการประเมินระบบ ทั้ง 5 ฉบับ มีเกณฑ์คะแนนอยู่ที่ช่วง 4 และ 5 คะแนน หรือคือช่วงมากและมากที่สุดทำให้ออกมาได้ว่า อยู่ในเกณฑ์ที่ผ่านและสามารถปรับปรุงและพัฒนาการทำงานขององค์กรได้อย่างมีประสิทธิภาพ

### เอกสารอ้างอิง

- [1] การประยุกต์ใช้ระบบซอฟต์แวร์อัตโนมัติในหน่วยงานสนับสนุน เพื่อความได้เปรียบทางการแข่งขันในธุรกิจธนาคารพาณิชย์ไทย กรณีศึกษา ธนาคารเอปียี สาขาปฏิบัติการ (2563).[ออนไลน์].[สืบค้นเมื่อวันที่ 15 กรกฎาคม 2564]
- [2] ปัจจัยที่มีผลต่อการตัดสินใจใช้ระบบกระบวนการทำงาน แบบอัตโนมัติในงานบัญชี (2557).[สืบค้นเมื่อวันที่ 1 กรกฎาคม 2564]
- [3] การประยุกต์ใช้ระบบการวางแผนทรัพยากรทางธุรกิจขององค์กรโดยรวมเพื่อการพัฒนาธุรกิจ(2559).[สืบค้นเมื่อวันที่ 1 กรกฎาคม 2564]
- [4] ระบบการทำงานอัตโนมัติแบบหุ่นยนต์-การปฏิบัติที่ยิ่งใหญ่ลำดับต่อไปในการปรับปรุงประสิทธิภาพการปฏิบัติงาน (2557).[สืบค้นเมื่อวันที่ 2 กรกฎาคม 2564]
- [5] เกียรติพงษ์ อุดมชนะธีระ. วิวัฒนาการ ระบบการวางแผนจัดการทรัพยากรในองค์กร (ERP). [ออนไลน์]. [สืบค้นวันที่ 3 กรกฎาคม 2564].
- [6] งานพัฒนาระบบพัสดุฝ่ายทรัพย์สินและพัสดุและ IT ฝ่ายสารสนเทศ. ระบบ SAP > MM > PO. (2558). [ออนไลน์]. [สืบค้นวันที่ 5 กรกฎาคม 2564].
- [7] รายงานการศึกษา เรื่อง “อุตสาหกรรมหุ่นยนต์ของประเทศไทย” (2558).[สืบค้นเมื่อวันที่ 5 กรกฎาคม 2564]
- [8] การวัดประสิทธิภาพของ Collaborative robot [สืบค้นเมื่อวันที่ 1 กรกฎาคม 2564]
- [9] Robotic Process Automation- ระบบดำเนินการอัตโนมัติด้วยหุ่นยนต์ [สืบค้นเมื่อวันที่ 5 กรกฎาคม 2564]
- [10] การใช้งาน ระบบ System Application Program (SAP) เพื่อการจัดซื้อ ของกองพัสดุและจัดการทรัพย์สิน มหาวิทยาลัยหอการค้าไทย จาก (2558).[สืบค้นเมื่อวันที่ 11 กรกฎาคม 2564]
- [11] ปัจจัยที่มีผลต่อความสำเร็จในการนำระบบ SAP โมดูล FI มาใช้งานบัญชีในมุมมองของผู้ปฏิบัติงาน จาก (2558).[สืบค้นเมื่อวันที่ 11 กรกฎาคม 2564]
- [12] Blueprint Business Flow SAP Module MM ของบริษัทแห่งหนึ่งสืบค้นเมื่อวันที่ 1 กรกฎาคม 2564]
- [13] Blueprint Business Flow SAP Module FI ของบริษัทแห่งหนึ่ง[สืบค้นเมื่อวันที่ 1 กรกฎาคม 2564]
- [14] ปัจจัยความต้องการซอฟต์แวร์การวางแผนทรัพยากรองค์กร (ERP) ทางการจัดการทรัพยากรมนุษย์ในธุรกิจส่งออกข้าว [สืบค้นเมื่อวันที่ 1 กรกฎาคม 2564]
- [15] ตัวอย่างการนำ RPA ไปใช้ในงานของบริษัท UTILITY (Lacity et al., 2015) และ Telefonica O2 [สืบค้นเมื่อวันที่ 15 กรกฎาคม 2564]
- [16] การประยุกต์ใช้ระบบซอฟต์แวร์อัตโนมัติในหน่วยงานสนับสนุน เพื่อความได้เปรียบทางการแข่งขันในธนาคารเอปียี สาขาปฏิบัติการ ชนันทรศรนุรักษ์ O2 [สืบค้นเมื่อวันที่ 15 กรกฎาคม 2564]

# การสร้างแบบทดสอบและตรวจคำตอบอัตโนมัติในการหาปริมาตรและพื้นที่ผิววัตถุสามมิติ

## Automatic Quiz Creation and Verification for the Volume and Surface Area of Three-dimensional Objects.

รัตนากร แสงอ่อน (Rattanakorn Saeng-Aun)<sup>1</sup> และพงษ์เทพ รักศกาวงศ์ (Phongthep Raxpakawong)<sup>2</sup>

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพิบูลสงคราม

<sup>1</sup>rsaengaun@gmail.com, <sup>2</sup>thep@psru.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ในการพัฒนาระบบสร้างแบบทดสอบและตรวจคำตอบอัตโนมัติในการหาปริมาตรและพื้นที่ผิววัตถุสามมิติโดยใช้คำสั่งจาก กูเกิลแอปสคริปต์ มีการกำหนดค่าสูตรคำนวณของรูปทรงวัตถุสามมิติเพื่อให้ระบบสามารถหาคำตอบและนำมาตรวจคำตอบจากผู้ทำ โดยงานวิจัยนี้ได้คัดเลือกจากกลุ่มตัวอย่างนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงคราม จำนวน 85 คน ซึ่งเป็นกลุ่มตัวอย่างที่ศึกษาในสถานศึกษาเดียวกันกับผู้วิจัย โดยแบ่งการทำแบบทดสอบเป็น 2 ครั้งเพื่อหาข้อผิดพลาดและตรวจสอบการทำงานของระบบ ผลจากการทดลองพบว่าระบบสามารถสร้างแบบทดสอบโดยใช้เวลาเฉลี่ย 30.78 วินาทีต่อ 1 ไฟล์ และใช้เวลาในการตรวจคำตอบเฉลี่ย 35.67 วินาทีต่อ 1 ไฟล์ ซึ่งระบบสามารถสร้างแบบทดสอบและตรวจคำตอบโดยมีค่าความถูกต้อง 99.353%

**คำสำคัญ:** แบบทดสอบออนไลน์ การหาปริมาตรและพื้นที่ผิว การตรวจแบบทดสอบผ่านระบบคอมพิวเตอร์

### Abstract

The objective of this research to develop a system to create a test and check the answer automatically to find the volume and surface area of three-dimensional objects using commands from Google App Script. Formulas for calculating three-dimensional object shapes are configured so that the

system can find answer and check the answer from the test takers. This research was selected from a sample of undergraduate students Pibulsongkram Rajabhat University 85 people, which is a sample group studied at the same educational establishment as the researcher. By dividing the test into 2 times to find errors and check system's operation. The results of the experiment showed that system was able to create a test with an average time of 30.78 seconds per file, and an average time of 35.67 seconds to check answer per file, the system can create quizzes and check answer with an accuracy of 99.353%

**Keywords:** Online Quiz, Finding Volume and Surface Area, Examination of the Test Through a Computer System

### 1. บทนำ

การทำแบบทดสอบถือเป็นการวัดผลสัมฤทธิ์ของผู้เรียน โดยเมื่อผู้เรียนได้เรียนจนครบหลักสูตรจะต้องมีการทดสอบเพื่อวัดผลสัมฤทธิ์ของผู้เรียน [1] การทำแบบทดสอบของครูประจำรายวิชา จะใช้วิธีการออกแบบทดสอบเพียงชุดเดียวเพื่อใช้ในการทดสอบ ซึ่งการทดสอบแต่ละครั้งนั้นจะมีการกำหนดสถานที่และเวลาในการจัดการทดสอบนั้นจะขึ้นอยู่กับสถานศึกษา หรือครูประจำรายวิชาเป็นผู้กำหนด

คณิตศาสตร์มีบทบาทสำคัญยิ่งต่อความสำเร็จในการเรียนรู้ในศวรรษที่ 21 เนื่องจากคณิตศาสตร์ช่วยให้

มนุษย์มีความคิดที่ริเริ่มสร้างสรรค์ คิดอย่างมีเหตุผล เป็นระบบ มีแบบแผนสามารถวิเคราะห์ปัญหาหรือสถานการณ์ได้อย่างรอบคอบและถี่ถ้วน ช่วยให้การคาดการณ์วางแผนตัดสินใจแก้ปัญหา ได้อย่างถูกต้องและเหมาะสม และสามารถนำไปใช้ในชีวิตจริงได้อย่างมีประสิทธิภาพ นอกจากนี้คณิตศาสตร์ยังเป็นพื้นฐานในการศึกษาทางด้านวิทยาศาสตร์และเทคโนโลยี อันเป็นรากฐานในการพัฒนาประเทศให้สามารถทัดเทียมกับนานาชาติ [2]

สถานการณ์การแพร่ระบาดของไวรัสโคโรนา 2019 (Covid-19) ทำให้มีผู้ติดเชื้อและเสียชีวิตเป็นจำนวนมาก สถานศึกษาหลายแห่งจึงต้องงดการเรียนการสอนเพื่อลดการแพร่ระบาดของไวรัสโคโรนา 2019 (Covid-19) และหันมาใช้ในการเรียนการสอนในรูปแบบออนไลน์ [3] และถ้าหาต้องการจัดการทดสอบต้องสอบผ่านระบบออนไลน์อาจเกิดการทุจริตได้เนื่องจากครูผู้สอนไม่สามารถควบคุมการสอบ ประกอบกับการออกแบบทดสอบเพียงแค่ 1-3 ชุดเท่านั้นต่อการสอบ 1 ครั้ง เพื่อป้องกันการทุจริต

จากหลักการและเหตุผลดังกล่าวผู้วิจัยจึงมีแนวคิดในการสร้างแบบทดสอบและตรวจคำตอบอัตโนมัติในการหาปริมาตรและพื้นที่ผิววัตถุสามมิติ โดยให้ระบบสามารถสร้างโจทย์ที่ไม่ซ้ำกัน และสามารถตรวจคำตอบได้อย่างรวดเร็วโดยใช้การทำงานของ Google App Script และผสานการทำงานของ Google Document, Google Slide ในการสั่งการทำงานในการสร้างแบบทดสอบเพื่อลดภาระของครูผู้สอนประจำรายวิชาและเป็นการลดการทุจริตในการสอบ ทั้งนี้ผู้ทดสอบสามารถสอบถามแลกเปลี่ยนความคิดเห็นได้โดยที่แบบทดสอบนั้นจะไม่ซ้ำกัน และช่วยลดภาระในการสร้างและตรวจแบบทดสอบ

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ทฤษฎีที่เกี่ยวข้อง

Google Apps Script คือชุดของคำสั่งที่แสดงขั้นตอนการทำงาน ตามที่ออกแบบและกำหนดไว้การ

เขียน Google Apps Script โดยจะเรียก กลุ่มของการทำงานว่า ฟังก์ชัน (function) (แบบเดียวกับ เมธอด (method) ของภาษา จาวา (Java)) ซึ่งการทำงานของ Google Apps Script จะเกิดจากการสั่งรันฟังก์ชัน (Niw Popkom, 2014) [4]

Google Slide เป็นเครื่องมือสำหรับสร้างงานนำเสนอ (Presentation) ในรูปแบบต่างๆ เช่นเดียวกับโปรแกรม Microsoft PowerPoint โดยผู้ใช้งานสามารถใส่ข้อความ รูปภาพเคลื่อนไหว เอฟเฟกเสียงดนตรี หรือ Template เพื่อเพิ่มความสวยงามและดึงดูดสนใจสำหรับ Google Slides ทำงาน โดยไม่ต้องติดตั้งโปรแกรม และสามารถทำงานร่วมกับผู้อื่นไปพร้อมกันได้อีกด้วย (EST Education Technology Development And Service, 2560) [5]

การหาปริมาตรและพื้นที่ผิววัตถุสามมิติ การหาปริมาตรและพื้นที่ผิวจะแบ่งเป็น พีระมิด ทรงกระบอก ปริซึม กรวย และทรงกลม

พีระมิด รูปเรขาคณิตสามมิติที่มีฐานเป็นรูปเหลี่ยมใดๆ มียอดแหลมที่ไม่อยู่ในระนาบเดียวกับฐาน และหน้าทุกหน้าเป็นรูปสามเหลี่ยมที่มีจุดรวมกันที่ยอดแหลมนั้น โดยมีสูตรคำนวณการหาปริมาตรและพื้นที่ผิวพีระมิด ดังนี้

$$\text{การหาปริมาตร} = \frac{1}{3} \times \text{base area} \times \text{height} \quad (2-1)$$

$$\text{การหาพื้นที่ผิว} = \frac{1}{2} \times \text{slant height} \times \text{base area} \quad (2-2)$$

ทรงกระบอก รูปเรขาคณิตที่มีฐานสองฐานเป็นรูปวงกลมที่เท่ากันทุกประการ และอยู่บนระนาบเดียวกัน เมื่อตัดรูปเรขาคณิตนี้ด้วยระนาบที่ขนานกับฐาน จะได้หน้าที่เป็นวงกลมที่เท่ากันทุกประการเสมอ โดยสูตรคำนวณการหาปริมาตรและพื้นที่ผิวทรงกระบอกมีดังนี้

$$\text{การหาปริมาตร} = \pi r^2 h \quad (2-3)$$

$$\text{การหาพื้นที่ผิว} = 2\pi r h + 2\pi^2 \quad (2-4)$$

ปริซึม รูปเรขาคณิตสามมิติที่มีฐาน ทั้งสองเป็นรูปเหลี่ยมใดๆ เท่ากันทุกประการ ฐานทั้งสองอยู่ในระนาบเดียวกัน และด้านข้างแต่ละด้านเป็นรูปสี่เหลี่ยมด้านขนาน โดยมีสูตรคำนวณการหาปริมาตรและพื้นที่ผิวปริซึมมีดังนี้

$$\text{การหาปริมาตร} = \text{base area} \times \text{height} \quad (2-5)$$

$$\text{การหาพื้นที่ผิว} = 2(\text{base area}) + 4(\text{lateral face}) \quad (2-6)$$

กรวย รูปเรขาคณิตสามมิติที่มีฐานเป็นทรงกลม และมียอดแหลมที่ไม่อยู่ในระนาบเดียวกับฐาน และเส้นที่ต่อระหว่างจุดกับยอดใดๆ บนขอบของฐานเป็นส่วนหนึ่งของเส้นตรง และด้านข้างของกรวยจะเป็นรูปสามเหลี่ยม โดยสูตรคำนวณการหาปริมาตรและพื้นที่ผิวกรวยมีดังนี้

$$\text{การหาปริมาตร} = \frac{1}{3} \pi r^2 h \quad (2-7)$$

$$\text{การหาพื้นที่ผิว} = \pi r l + \pi r^2 \quad (2-8)$$

ทรงกลม รูปเรขาคณิตสามมิติที่มีผิวโค้งเรียบ และจุดทุกจุดบนผิวโค้งอยู่ห่างจากจุดจุดหนึ่งเป็นระยะที่เท่ากัน จุดตรงที่นั้นเรียกว่า จุดศูนย์กลางของทรงกลม ระยะที่ผ่านจุดศูนย์กลางที่เท่ากันนั้นเรียกว่า รัศมี โดยมีสูตรคำนวณการหาปริมาตรและพื้นที่ผิวทรงกลมดังนี้

$$\text{การหาปริมาตร} = \frac{4}{3} \pi r^3 \quad (2-9)$$

$$\text{การหาพื้นที่ผิว} = 4\pi r^2 \quad (2-10)$$

[6]

## 2.2 งานวิจัยที่เกี่ยวข้อง

จิตร โสภิน ทศนสุวรรณ และ พงษ์เทพ รักผกาวงศ์ [7] ได้ทำการศึกษาวิจัยการสร้างโจทย์ปัญหาหลายชั้น และตรวจคำตอบอัตโนมัติทางคณิตศาสตร์ : เป็นการพัฒนาโปรแกรมสร้างไฟล์โจทย์ปัญหาหลายชั้น และตรวจคำตอบอัตโนมัติทางคณิตศาสตร์โดยตัวข้อสอบต้องไม่ซ้ำกัน ระบบจะส่งแบบทดสอบไปให้ผู้ทำแบบทดสอบและเมื่อหมดเวลาทำแบบทดสอบระบบจะตรวจคำตอบและรายงานผลคะแนน โดยผู้ทำแบบทดสอบจะได้แบบทดสอบคนละ 1 ไฟล์ ให้ผู้ทำแบบทดสอบ 3 โจทย์ ซึ่ง 1 ไฟล์ใช้เวลาในการสร้างโจทย์ปัญหา 22.88 วินาที และใช้เวลาในการตรวจ 65.33 วินาที ซึ่งช่วยลดภาระของอาจารย์ได้เป็นอย่างดี ซึ่งคะแนนเฉลี่ยการทำแบบทดสอบ 1 ไฟล์ผู้ทำแบบทดสอบผ่านเกณฑ์เฉลี่ย 87.93%

ศุภราช วรรณทิพภากรณ์ และ พงษ์เทพ รักผกาวงศ์ [8] ได้ทำการศึกษาวิจัย การออกข้อสอบ การเขียนโปรแกรมเชิงวัตถุ ภาษา JavaScript : งานวิจัยนี้มี

วัตถุประสงค์ เพื่อสร้างชุดข้อสอบให้ได้หลายโจทย์เพื่อใช้ฝึกเขียนโปรแกรมภาษาจาวาสคริปต์ และตรวจคำตอบอัตโนมัติบนอินเทอร์เน็ต โดยให้ข้อสอบต้นแบบทำการสุ่มเปลี่ยนค่าเพื่อให้ได้ข้อสอบที่ไม่ซ้ำกัน ผลจากการทดลองงานความเร็วโดยเฉลี่ยของการสร้างชุดข้อสอบ 1 ชุดใช้เวลา 25.90 วินาที และความเร็วในการตรวจ 57.84 วินาที และผลโปรแกรมทำงานถูกต้อง 100%

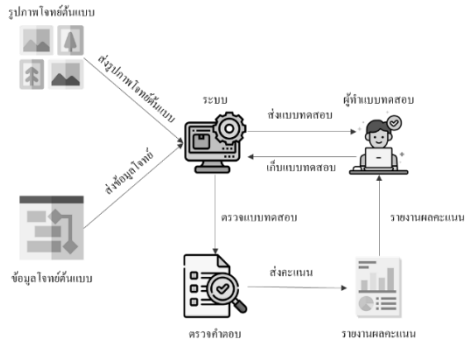
พงษ์เทพ รักผกาวงศ์ และ อุไรวรรณ รักผกาวงศ์ [9] ได้ทำการศึกษาวิจัย การพัฒนาส่วนเสริมในกูเกิลชีทเพื่อการผสานจดหมายและส่งอีเมล : โดยงานวิจัยนี้มีวัตถุประสงค์ที่ต้องการส่งจดหมายเวียนโดยแนบไฟล์ผ่านอีเมลและใช้คำสั่งคอมพิวเตอร์เพื่อส่งจดหมายจากการทดลองเตรียม 470 รายชื่อ และเตรียมเอกสารต้นแบบ 5 ฉบับ ผลจากการทดลองพบว่าสามารถส่งไฟล์แนบที่เป็นไฟล์พีดีเอฟ ไมโครซอฟต์เวิร์ด ไมโครซอฟต์เอกเซล ไมโครซอฟต์พาวเวอร์พ้อยต์ที่มีเนื้อหาเฉพาะบุคคล ใช้เวลาเฉลี่ย 28.21 วินาทีต่อ 1 อีเมล

มนต์เทพ สุดใจ และ สุรศักดิ์ มั่งสิงห์ [10] ได้ทำการศึกษาวิจัย การพัฒนาระบบตรวจให้คะแนนคำตอบข้อสอบอัตโนมัติแบบอัตโนมัติด้วยเทคนิคการค้นหาข้อความแบบเต็มและฟังก์ชันการเปรียบเทียบข้อความของภาษาพีเอชพี : บทความนี้ผู้วิจัยได้นำเสนอการพัฒนาระบบตรวจให้คะแนนคำตอบข้อสอบอัตโนมัติบนระบบเครือข่ายอินเทอร์เน็ต โดยใช้เทคนิคการค้นหาข้อความแบบเต็ม และฟังก์ชันเปรียบเทียบข้อความของภาษาพีเอชพี ผลจากการทดลองนำไปใช้งานกับนักเรียนผู้ทดสอบระบบจำนวน 250 คน ในการทำแบบทดสอบอัตโนมัติผลการทดสอบพบว่าครูผู้สอนมีค่า  $\bar{x} = 4.32$  ความพึงพอใจอยู่ในระดับดี

วีระศักดิ์ วีระวงศ์ [11] ได้ทำการศึกษาวิจัย การบริหารการสอบออนไลน์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยแห่งชาติลาว : โดยงานวิจัยนี้ผู้วิจัยได้ออกแบบระบบจัดการการสอบออนไลน์ โดยใช้โปรแกรม Moodle เพื่อบริหารระบบการจัดการสอบออนไลน์ผ่านเว็บได้และเพื่อหาคุณภาพของระบบที่ทำการ

จัดการบริหารจัดการสอบออนไลน์ ผลจากการวิจัยสรุปว่า ประชากรมีความพึงพอใจอยู่ในระดับมาก สามารถนำไปประยุกต์ใช้เพื่อสนับสนุนการทำงานของปฏิบัติ และสนับสนุนการตัดสินใจของผู้บริหารได้เป็นอย่างดี

### 3. วิธีการดำเนินงานวิจัย



ภาพที่ 1: กรอบแนวคิดของงานวิจัย

จากภาพที่ 1 หลักการทำงานของระบบโดยรวมของการสร้างแบบทดสอบนั้นข้อมูลของโจทย์ต้นแบบจะทำการสุ่มค่าตัวเลขไปใส่ที่ตัวแปรของรูปภาพ โจทย์ต้นแบบที่กำหนดไว้ เมื่อระบบได้สร้างแบบทดสอบเสร็จ ระบบจะสุ่ม โจทย์คำถามและส่งไปให้ผู้ทำแบบทดสอบผ่านอีเมล เมื่อถึงเวลาที่กำหนดจะเก็บแบบทดสอบ รวมถึงคำตอบและตรวจคำตอบโดยอัตโนมัติ จากนั้นระบบจะรายงานผลการทำแบบทดสอบพร้อมทั้งเฉลยและวิธีการคิดให้กับผู้ทำแบบทดสอบเมื่อทำการตรวจแบบทดสอบเสร็จ โดยขั้นตอนการดำเนินงานมีดังนี้

#### 3.1 การสร้างรูปภาพโจทย์และการสร้างโจทย์ปัญหา

การสร้างรูปภาพ โจทย์ผู้วิจัยได้ทำการออกแบบรูปภาพ โจทย์เป็นรูปทรงวัตถุสามมิติโดยใช้โปรแกรม Google Sketch Up ในการออกแบบรูปทรงวัตถุสามมิติ จากนั้นจะนำภาพไปใส่ใน Google Slide และทำการรันคำสั่งใน Google App Script สำหรับการสุ่มตัวเลขและสร้าง โจทย์คำถาม โดยการทำงานของระบบนั้นจะนำค่าที่เตรียมไว้ใน Google App Script มาใส่ที่รูปภาพตามตัวแปรที่กำหนดไว้ตามจุดต่างๆ และคำนวณคำตอบสำหรับการตรวจคำตอบตามสูตรคณิตศาสตร์ที่ได้มีการเขียน คำสั่งไว้ใน Google App Script เมื่อได้

แบบทดสอบที่สมบูรณ์แล้วนั้นจะนำไปเก็บไว้ที่ Google Drive ตามที่ได้มีการเขียนคำสั่งไว้



ภาพที่ 2: การกำหนดตัวแปร

การสร้างประโยคคำถาม ในการสร้างประโยคคำถามนั้นรูปภาพ โจทย์ 1 ภาพจะมีคำถามการหาปริมาตรและการหาพื้นที่ผิว โดยรูปแบบของประโยคคำถามจะมีการกำหนดรูปแบบไว้ คือ ปริมาตรของวัตถุ[ตัวแปร] และพื้นที่ผิววัตถุ[ตัวแปร] ตัวแปรนั้นจะได้อาจมาจากการสร้าง โจทย์รูปภาพ

ในกรณีคำตอบที่เป็นทศนิยมมากกว่า 3 ตำแหน่งให้ผู้ทำแบบทดสอบตอบเพียงแค่ 3 ตำแหน่งเท่านั้น

เมื่อสร้างแบบทดสอบเสร็จแล้วนั้นระบบจะเตรียมการสุ่ม โจทย์สำหรับการส่งแบบทดสอบผ่านอีเมล โดยเก็บข้อมูลของผู้ทำแบบทดสอบไว้ใน Google Spreadsheet เพื่อสะดวกต่อการค้นหารายชื่อผู้ทำแบบทดสอบ และการส่งอีเมลแบบทดสอบโดยมีรายละเอียดของแต่ละบุคคล

ภาพที่ 3: การจัดเก็บข้อมูลของผู้ทำแบบทดสอบ

#### 3.2 การทำแบบทดสอบ

เมื่อถึงเวลาที่กำหนดในการทำแบบทดสอบระบบจะแจ้งเตือนให้ผู้ทำแบบทดสอบนั้นทราบผ่านอีเมลของผู้ทำแบบทดสอบ และผู้ทดสอบจะสามารถเข้าไปแก้ไขไฟล์เพื่อตอบคำถามโดยแบบทดสอบที่เป็นไฟล์ Google Document ดังภาพที่ 2 ผู้ทำแบบทดสอบ

สามารถเข้าไปทบทวนหรือแก้ไขคำตอบได้ถ้าหากยังไม่หมดเวลาการทำแบบทดสอบ การแก้ไขไฟล์เพื่อตอบคำถามนั้นคอมพิวเตอร์หรืออุปกรณ์ที่ใช้ทำแบบทดสอบจะต้องเชื่อมต่ออินเทอร์เน็ตทุกครั้ง และเมื่อหมดเวลาทำแบบทดสอบตามที่กำหนดผู้ทำแบบทดสอบจะไม่สามารถแก้ไขคำตอบได้อีก

ข้อ 1	จากรูป จงตอบคำถาม ข้อ 1.1-1.5	
คำถาม	คำตอบ	
1.1> พื้นที่ผิววงรี F		
1.2> ปริมาตรวงรี F		
1.3> ปริมาตรวงรี G		
1.4> ปริมาตรวงรี Z		
1.5> ปริมาตรวงรี R		

ภาพที่ 4: แบบทดสอบที่สร้างเสร็จแล้ว

### 3.3 การตรวจคำตอบและให้คะแนน

เมื่อหมดเวลาในการทำแบบทดสอบ ระบบจะตรวจคำตอบของผู้ทำแบบทดสอบโดยการนำคำตอบที่ได้ไปเปรียบเทียบกับคำตอบของแต่ละข้อที่ได้คำตอบมาจากรุ่นต้นการสร้างโจทย์คำถาม ถ้าหากผู้ทำแบบทดสอบตอบถูก จะขึ้นเครื่องหมายถูกในช่องคำตอบ แต่ถ้าหากผู้ทำแบบทดสอบตอบผิดจะขึ้นเครื่องหมายกากบาทและคำตอบที่ถูกในช่องคำตอบ พร้อมทั้งแสดงวิธีการคิดและคำตอบที่ถูกต้อง และจะรวบรวมคะแนนและรายงานผลผ่านทางอีเมลล์ของผู้ทำแบบทดสอบ ผู้ทำแบบทดสอบจะไม่สามารถแก้ไขได้อีก สามารถดูได้แค่คะแนนรวมข้อที่ทำถูก ข้อที่ทำผิดวิธีการคิดของแต่ละข้อ

คำถาม	คำตอบ
1.1> ปริมาตรวงรี E = $(1/3) * (6 * (3/4) * 12 * 2) * 16$ = 27648	27648 ✓ (2/2 คะแนน)
1.2> ปริมาตรวงรี G = $4/3 * (3.14 * 11^2 * 3)$ = 5572.453333333333	5572.45 ✗ 5572.453 (0/2 คะแนน)
1.3> พื้นที่ผิววงรี X = $((3.14 * 6) * 19) + (3.14 * 6 * 2)$ = 471	471 ✓ (2/2 คะแนน)

ภาพที่ 5: การตรวจคำตอบ

## 4. ผลการดำเนินงาน

ผลการทดลองการนำแบบทดสอบไปใช้ในการทำแบบทดสอบการหาปริมาตรและพื้นที่ผิววัตถุสามมิติ โดยมีกลุ่มตัวอย่างจากนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงครามจำนวน 85 คน คนละ 2 ไฟล์ ไฟล์ละ 10 ข้อ โดยให้ทำแบบทดสอบทั้งหมดคนละ 2 ครั้ง ได้ค่าความถูกต้องของการตรวจคำตอบดังนี้

ตารางที่ 1: ค่าความถูกต้องของการตรวจคำตอบ

การตรวจ	ครั้งที่	
	1	2
ถูก	839	850
ผิด	11	0
ค่าเฉลี่ย	98.706%	100%
รวม	99.353	

ดังนั้น จึงสรุปผลได้ว่าผลจากการทดลองใช้งานการสร้างแบบทดสอบและตรวจคำตอบอัตโนมัติในการหาปริมาตรและพื้นที่ผิววัตถุสามมิติจากการทดลองใช้แบบทดสอบกับกลุ่มนักศึกษาทั้งหมด 85 คน ผลปรากฏว่าระบบใช้สามารถสร้างแบบทดสอบ 30.78 วินาทีต่อ 1 ไฟล์ และใช้เวลาในการตรวจคำตอบ 35.67 วินาทีต่อ 1 ไฟล์ โดยระบบสามารถตรวจคำตอบได้ค่าความถูกต้องในการตรวจคำตอบครั้งแรก 98.706% และครั้งที่สองค่าความถูกต้องในการตรวจคำตอบอยู่ที่ 100% ค่าเฉลี่ยรวมในการตรวจคำตอบอยู่ที่ 99.353%

## 5. สรุปผลการวิจัยและข้อเสนอแนะ

จากหลักการและเหตุผลที่ผู้วิจัยกล่าวไว้ข้างต้นระบบสามารถส่งแบบทดสอบผ่านอีเมลล์ได้ ระบบสามารถสร้างแบบทดสอบที่ไม่ซ้ำกันได้ โดยสรุปได้ดังนี้

### 1. การส่งแบบทดสอบผ่านอีเมลล์

การส่งแบบทดสอบผ่านอีเมลล์ผู้วิจัยต้องการผสมผสานจดหมายสำหรับการส่งแบบทดสอบซึ่งสอดคล้องกับงานวิจัย [9] ซึ่งระบบสามารถส่งแบบทดสอบได้ และสามารถทำแบบทดสอบผ่านระบบอินเทอร์เน็ตได้เลย

### 2. การสร้างแบบทดสอบที่ไม่ซ้ำกัน

การสร้างแบบทดสอบที่ไม่ซ้ำกันโดยระบบสามารถสร้างแบบทดสอบที่ไม่ซ้ำกัน โดยการสุ่ม โจทย์ซึ่งสอดคล้องกับงานวิจัยของ [7] ซึ่งระบบสามารถสร้าง โจทย์ที่ไม่ซ้ำกัน โดยการสุ่มได้

ข้อเสนอแนะ ถ้าหากต้องการนำระบบนี้ไปต่อยอด ควรจะศึกษาการหาปริมาตรสามมิติรูปทรงอื่นๆ เพิ่มเติม เพื่อพัฒนาให้ระบบมีประสิทธิภาพมากขึ้น

### เอกสารอ้างอิง

- [1] M. Rashad, M. Kandil, A. Hassan, and M. Zaher. "An Arabic Web-Based Exam Management System." *International Journal of Electrical & Computer Sciences IJECS-IJENS*, Vol: 10, No: 1, pp: 35-41, 2010.
- [2] ตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้คณิตศาสตร์ (ฉบับปรับปรุง พ.ศ. ๒๕๖๐) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๖๑ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ. โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย, กรุงเทพมหานคร, 2561.
- [3] มนธิชา ทองหัตถา. "สภาพการจัดการเรียนรู้แบบออนไลน์ ในสถานการณ์การแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา(Covid-19) ของกลุ่มการเรียนรู้ภาษาต่างประเทศ." *วารสารละติจูด มหาวิทยาลัยราชภัฏเทพสตรี*, ปีที่ 5, ฉบับที่ 1, หน้า 43-51, 2564.
- [4] N. Popkorn. [ออนไลน์]. "การเรียกใช้งาน Google App Script." [สืบค้นเมื่อวันที่ 11 พฤศจิกายน 2564]. จาก <https://blog.niwpopkorn.com/2014/02/google-apps-script.html>
- [5] EST Education Technology Development And Service [ออนไลน์]. "Google Slide create and edit presentations online." [สืบค้นเมื่อวันที่ 11 พฤศจิกายน 2564]. จาก [https://sornorjeab.files.wordpress.com/2017/12/manual\\_google-slides.pdf](https://sornorjeab.files.wordpress.com/2017/12/manual_google-slides.pdf).
- [6] วิลาสินี นรินทร์รัตน์. [ออนไลน์]. "ปริมาตรและพื้นที่ผิว." [สืบค้นเมื่อวันที่ 11 พฤศจิกายน 2564]. จาก <http://geometryonlinenow.weebly.com/36123641365735923633360436073635.html>
- [7] จิตร โสภณ หัตสนสุวรรณ, และ พงษ์เทพ รักผกาวงศ์. "การสร้างโจทย์ปัญหาหลายชั้นและตรวจคำตอบอัตโนมัติทางคณิตศาสตร์." *การประชุมวิชาการวิทยาศาสตร์และเทคโนโลยีระดับปริญญาตรี มหาวิทยาลัยราชภัฏพระนครศรีอยุธยา*, ปีที่ 6, หน้า 225-233, 2563.
- [8] สุภราช วรรณทิพย์, และ พงษ์เทพ รักผกาวงศ์. "การออกแบบข้อสอบ การเขียนโปรแกรมเชิงวัตถุ ภาษา Java Script." *การประชุมวิชาการระดับชาติและนานาชาติ มหาวิทยาลัยภาคตะวันออกเฉียงเหนือ*, ครั้งที่ 6, หน้า 1125-1133, 2562.
- [9] พงษ์เทพ รักผกาวงศ์, และ อุไรวรรณ รักผกาวงศ์. "การพัฒนาส่วนเสริมในกูเกิลชีท เพื่อประสานจดหมาย และส่งอีเมลล์." *การนำเสนอผลงานระดับชาติ เครือข่ายบัณฑิตศึกษา มหาวิทยาลัยราชภัฏภาคเหนือ*, ครั้งที่ 17, หน้า 2671-2680, 2560.
- [10] มนต์เทพ สุดใจ, และ สุรศักดิ์ มั่งสิงห์. "การพัฒนาระบบตรวจให้คะแนนคำตอบข้อสอบอัตโนมัติแบบอัตโนมัติด้วยเทคนิคการค้นหาข้อความแบบเต็มและฟังก์ชันการเปรียบเทียบข้อความของภาษาพีเอชพี." *วารสารวิทยาศาสตร์และเทคโนโลยีสารสนเทศ*, ปีที่ 11, ฉบับที่ 1, ม.ค.-มิ.ย, หน้า 8-17, 2564.
- [11] วีระศักดิ์ วีระวงศ์. "การบริหารระบบการจัดการสอบออนไลน์ คณะวิศวกรรมศาสตร์มหาวิทยาลัยแห่งชาติลาว." *วิทยานิพนธ์มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี*, 2554.

# ระบบแจ้งและติดตามปัญหาข้อขัดข้องทางเทคโนโลยีสารสนเทศ

## กรณีศึกษา:ศาลแพ่ง

### An Incident Tracking System: A Case Study of Civil Court

นนท์ จิตรจันทน์ (Non Chitchamnon)<sup>1</sup> และวราพร จิระพันธุ์ทอง (Waraporn Jirapanthong)<sup>2</sup>

วิทยาลัยครีเอทีฟดีไซน์แอนด์เอ็นเตอร์เทนเมนต์เทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิตย์

<sup>1</sup>ds.non.ch@gmail.com, <sup>2</sup>waraporn.jir@dpu.ac.th

#### บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและพัฒนา ระบบแจ้งและติดตามปัญหาข้อขัดข้องทางเทคโนโลยีสารสนเทศ กรณีศึกษา:ศาลแพ่ง การวิจัยนี้พัฒนาขึ้นโดยใช้บัญชีทางการของไลน์ (LINE Official Account) เป็นช่องทางที่เจ้าหน้าที่ศาลแพ่งที่ใช้แจ้งปัญหาฯ ให้แก่ส่วนเทคโนโลยีสารสนเทศทราบ และมีไดอะล็อกโฟล (Dialogflow) ทำหน้าที่ถามตอบกับผู้ใช้และรวบรวมความต้องการของผู้ใช้เพื่อส่งต่อให้ไฟเบสฟังก์ชัน (Firebase functions) ประมวลผลและเก็บข้อมูลลงในฐานข้อมูลไฟสตอร์ (Firestore Database) ต่อไป และพัฒนาโมบายแอปพลิเคชันให้เจ้าหน้าที่ในส่วนเทคโนโลยีสารสนเทศใช้เพื่อรับทราบปัญหาที่ผู้ใช้แจ้งเข้ามา และตอบสนองกับปัญหาที่เกิดขึ้น

**คำสำคัญ:** ไดอะล็อกโฟล ฐานข้อมูลไฟสตอร์ ไฟเบสฟังก์ชัน

#### Abstract

This research aim to study and develop notification system and track incidents about information technology. We have created a case study of the Civil Court. This research is developed by using "Line Official Account" as a way for Civil court officers to submit incidents to Information Technology (IT) department. "Dialogflow" is provided for users to communicate their problems. It includes questions and answers. Then the incidents and user requirements are compiled and collected to Firebase functions in order to process and record them into Firestore Database. We also develop a mobile

application for the IT staffs to receive and response to problems from the user.

**Keyword:** Dialogflow, LINE Official Account, Firebase functions , Firestore Database.

#### 1. บทที่ 1

ในปัจจุบันเทคโนโลยีสารสนเทศเข้ามามีความจำเป็นต่อการทำงานในทุกส่วนขององค์กรตั้งแต่ระดับพื้นฐานถึงระดับสูง มีการส่งข้อมูลระหว่างกันในองค์กรเป็นจำนวนมาก มีการนำข้อมูลเข้าและนำข้อมูลออกจากระบบงานต่างๆ ในปริมาณสูง เมื่อในองค์กรมีการใช้งานฮาร์ดแวร์และซอฟต์แวร์ต่างๆ ในปริมาณมากและหลากหลายรูปแบบ ทำให้มีปัญหาด้านเทคโนโลยีสารสนเทศเกิดขึ้นเป็นจำนวนมาก ปัญหาต่างๆเหล่านี้ต้องการการจัดการที่เป็นรูปแบบ สามารถจัดลำดับความสำคัญและติดตามผลการแก้ไขปัญหาได้

ศาลแพ่ง เป็นศาลยุติธรรมชั้นต้นซึ่งมีอำนาจพิจารณาคดีแพ่งทั้งปวงและคดีอื่นใดที่มีใ้ได้อยู่ในอำนาจของศาลยุติธรรมอื่น โดยในปี 2564 มีจำนวนบุคลากรมากกว่า 400 คน มีเครื่องคอมพิวเตอร์และอุปกรณ์ต่อพ่วงต่างๆเป็นจำนวนมาก ทำให้เกิดปัญหาข้อขัดข้องทางด้านเทคโนโลยีสารสนเทศมากขึ้นตามไปด้วย

Line application เป็นแอปพลิเคชัน ส่งข้อความระหว่างผู้ใช้ด้วยกันและยังมีบริการ line official account ที่สามารถใช้ chatbot และ api ในการติดต่อกับผู้ใช้ได้ และ Line ยังเป็น แอปพลิเคชันที่มีผู้ใช้ในประเทศไทยเป็นจำนวนมาก โดยในปี 2563 มีจำนวนผู้ใช้ 45 ล้านบัญชี



การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาระบบแจ้งและติดตามผลปัญหาข้อขัดข้องทางเทคโนโลยีสารสนเทศของบุคลากรในศาลแพ่ง เพื่อเป็นระบบที่ผู้ใช้สามารถแจ้งปัญหาและติดตามผลข้อขัดข้องทางเทคโนโลยีสารสนเทศได้โดยสะดวก โดยมีกลุ่มเป้าหมายเป็นบุคลากรในศาลแพ่ง ผู้ใช้ต้องเพิ่มเพื่อนผ่าน QR code โดยไม่จำเป็นต้องลงโปรแกรมเพิ่มเติม เป็นการพัฒนาระบบโดยมีส่วนร่วมติดต่อกับผู้ใช้ผ่าน line official account และเขียนการโต้ตอบกับผู้ใช้ด้วย DialogFlow และส่งข้อมูลไปประมวลผลด้วยภาษา Node js บน firebase functions และใช้ firebase firestore เป็นระบบจัดการฐานข้อมูลและพัฒนาโมบาย แอปพลิเคชันสำหรับให้ผู้ใช้ดูแลระบบสามารถบริหารจัดการการแก้ไขปัญหาข้อขัดข้องทางเทคโนโลยีสารสนเทศได้โดยสะดวก

**2. ทบทวนวรรณกรรม**

**2.1 ITIL Incident Management process**

การจัดการข้อขัดข้องของทางเทคโนโลยีสารสนเทศมีความสำคัญต่อการทำงานอย่างต่อเนื่อง ระบบเทคโนโลยีสารสนเทศจำเป็นต้องมีการจัดการข้อขัดข้องเพื่อให้มั่นใจได้ว่าระบบจะสามารถทำงานได้อย่างเต็มประสิทธิภาพ [1]

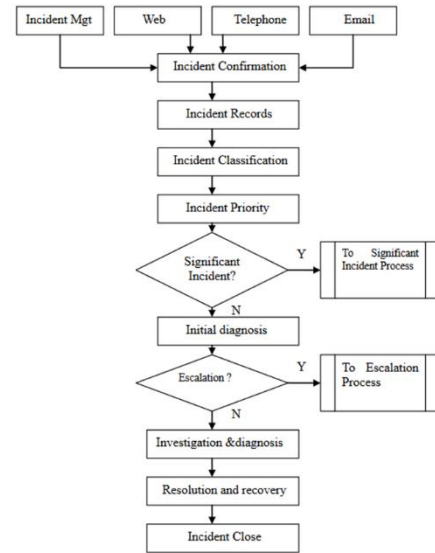
ITIL คือมาตรฐานหนึ่งในระดับนานาชาติสำหรับการดำเนินงานด้าน IT โดย ITIL อธิบายคำว่า incident ไว้ว่าการขัดข้องของการบริการด้าน IT โดยไม่คาดหมาย หรือคุณภาพของการบริการด้าน IT ลดลง และ Incident Management process ก็คือกระบวนการสำหรับการรับมือกับ incident ต่างๆ ที่เกิดขึ้น

ITIL Incident Management process มีขั้นตอนต่างๆ ดังภาพที่ 1 ตั้งแต่ ยืนยัน จดบันทึก จัดกลุ่ม จัดลำดับความสำคัญ เริ่มวิเคราะห์ขั้นต้น จัดระดับความสำคัญอีกครั้ง หากสาเหตุเพิ่มเติมและวินิจฉัย แก้ไขและฟื้นคืน ปิด [2]

**2.2 Line**

ปัจจุบันไลน์แอปพลิเคชัน ได้เข้ามามีอิทธิพลและบทบาทในชีวิตประจำวันเรามากขึ้น ในเรื่องการติดต่อสื่อสาร ตลอดจนแลกเปลี่ยนข้อมูลข่าวสารในการทำงาน เพื่อให้เกิดประโยชน์ต่อการทำงานสูงสุด อย่างในเรื่องของความสะดวกรวดเร็ว และได้รับความนิยมนและ

การตอบรับจากกลุ่มผู้ใช้ได้มากขึ้นและหลากหลาย [3] และกลุ่มผู้สูงอายุยอมรับไลน์จากการที่เห็นว่ามีประโยชน์ในการใช้งานและใช้งานง่าย [4]



ภาพที่ 1: ขั้นตอนของ ITIL

**2.3 Chatbot**

Chatbot สามารถแบ่งเป็นสองรูปแบบคือ แบบผู้ช่วยส่วนบุคคลเสมือน (Virtual Personal Assistants (VPAs)) เช่น Apple Siri, Microsoft Cortana, Google Assistant, Amazon Alexa และแบบเฉพาะเจาะจงกลุ่มผู้ใช้ [5] ทุกวันนี้ก็มีองค์กรจำนวนมากที่กำลังพัฒนาทดสอบ หรือใช้งานแชทบอทแล้ว อย่างไรก็ตามการใช้แชทบอทได้ต้องมีความมั่นใจว่าสามารถเข้าใจความต้องการของผู้ใช้ได้โดยไม่ต้องมีคนเข้ามาเกี่ยวข้อง [6] สามารถเปรียบเทียบแชทบอทกับช่องทางสื่อสารอื่นๆได้ตามตารางที่ 1 [6]

ตารางที่ 1: เปรียบเทียบแชทบอทกับช่องทางสื่อสารอื่นๆ

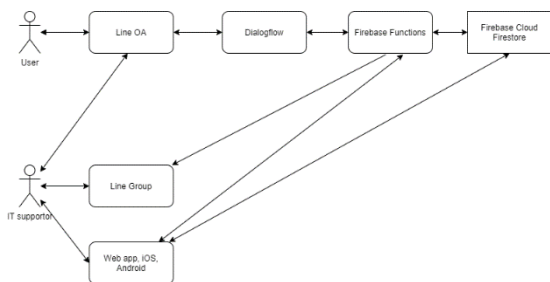
	Chatbot	Website	Social Media	App	Hotline
Contacting by companies	+	--	++	+	+
Contacting by users or customers	++	+	+	+	++
Customer service outside the service hour	++	++	-	-	--
Communication to new customers (acquisition)	-	++	+	-	+
Communication to customers (retention)	++	+	++	+	+
Communication among users	+	-	++	+	--
Machine/deep learning (Artificial Intelligence)	++	-	+	+	--
Personalization / 1-to-1 marketing	++	+	+	+	++
Context-/location-based services	++	-	+	+	--

Legend: ++ strength + possible - difficult -- weakness (not possible)

### 3. ระบบแจ้งและติดตามปัญหาข้อขัดข้องทางเทคโนโลยีสารสนเทศ

#### 3.1 ภาพรวม (conceptual framework สำหรับกระบวนการจัดการ)

ในการออกแบบระบบในส่วนของผู้ใช้นั้นในการสร้างระบบที่สามารถใช้งานได้ง่าย ไม่ต้องลงแอปพลิเคชันเพิ่มเติมและสามารถใช้งานได้บนโทรศัพท์เคลื่อนที่ อีกทั้งไลน์แอปพลิเคชันมีผู้ใช้งานในสัดส่วนที่สูงมากในประเทศไทย ในส่วนของผู้ใช้จึงใช้บัญชีทางการของไลน์ (LINE Official Account) เป็นส่วนที่ใช้ติดต่อกับผู้ใช้ (User interface) โดยบัญชีทางการของไลน์จะส่งผ่านข้อความที่ผู้ใช้พิมพ์เข้ามาไปยังไดอะล็อกโฟล (Dialogflow) โดยไดอะล็อกโฟลทำหน้าที่ประมวลผลข้อความสำหรับตอบโต้กลับไปยังบัญชีทางการของไลน์ หรือรวบรวมความต้องการของผู้ใช้และส่งต่อไปยังเว็บเซิร์ฟเวอร์ที่อยู่ในไฟเบสฟังก์ชันเพื่อประมวลผล และใช้ฐานข้อมูลไฟลด์โตรเป็นฐานข้อมูล ดังภาพที่ 2



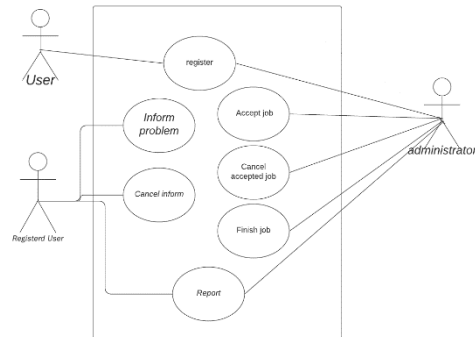
ภาพที่ 2: องค์ประกอบของระบบ

#### 3.2 ข้อมูลที่เกี่ยวข้อง

Use case diagram

- ลงทะเบียน ผู้ใช้จะต้องเพิ่มเพื่อนกับบัญชีทางการของไลน์ แล้วลงทะเบียนเพื่อระบุชื่อที่ต้องการใช้ก่อน
- แจ้งปัญหา เมื่อลงทะเบียนแล้วผู้ใช้สามารถแจ้งปัญหาได้โดยระบุ ปัญหา สถานที่ เบอร์โทรศัพท์
- ยกเลิกการแจ้งปัญหา เมื่อแจ้งปัญหาสำเร็จ สามารถยกเลิกการแจ้งปัญหาได้
- แจ้งรับงาน ผู้ดูแลระบบ สามารถรับงาน/รับแก้ปัญหาก็ได้

- แจ้งยกเลิกการรับงาน ผู้ดูแลระบบ สามารถยกเลิกการรับงาน/รับแก้ปัญหาก็ได้
- แจ้งทำงานสำเร็จ ผู้ดูแลระบบ สามารถแจ้งทำงานสำเร็จพร้อมระบุ เหตุผลได้



ภาพที่ 3: Use case diagram

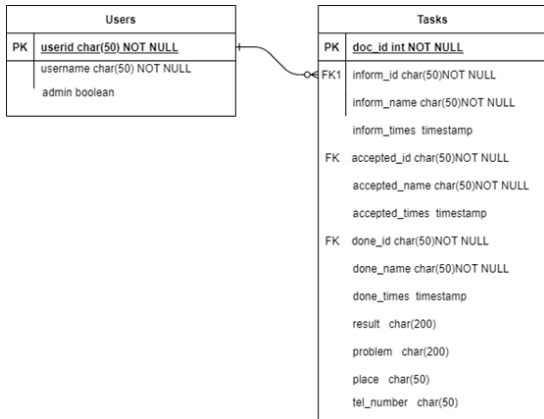
ER diagram ประกอบด้วยตาราง users และตาราง tasks

ตาราง Users มีฟิลด์ต่างๆ ดังนี้

- userid คือค่าที่ได้จากบัญชีทางการของไลน์ (LINE Official Account) โดยเป็นค่าที่ระบบจะต้องใช้ในการส่งข้อความให้แก่ผู้ใช้
- username คือค่าที่ผู้ใช้ลงทะเบียนในบัญชีทางการของไลน์ (LINE Official Account)
- admin ค่าสถานะความเป็นแอดมินของระบบ

ตาราง Tasks มีฟิลด์ต่างๆ ดังนี้

- doc\_id เป็น running number
- problem ปัญหาที่ผู้ใช้แจ้งเข้ามา
- place สถานที่ที่เกิดปัญหา
- tel\_number เบอร์โทรศัพท์ของผู้แจ้งเพื่อให้สามารถติดต่อสามารถเพิ่มเติมได้
- ข้อมูลผู้แจ้ง userid, username ของผู้แจ้งและเวลาที่มีการแจ้ง
- ข้อมูลผู้รับงาน userid, username ของผู้รับงานและเวลาที่มีการรับงาน
- ข้อมูลผู้เสร็จงาน userid, username ของผู้เสร็จงานและเวลาที่มีการเสร็จงาน พร้อมทั้งเหตุผลประกอบการทำงานเสร็จ

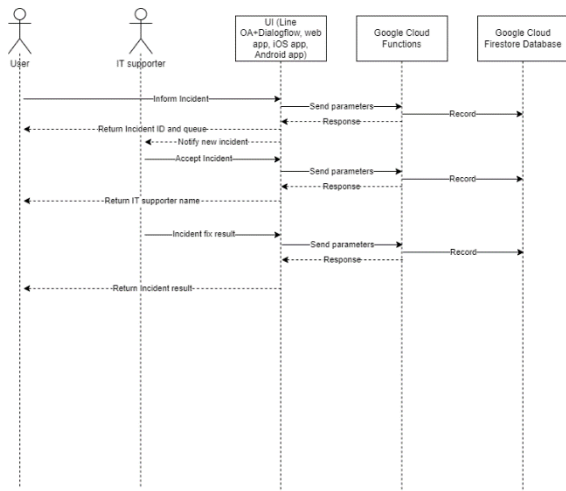


ภาพที่ 4: ER diagram

Sequence diagram

Sequence diagram ของกระบวนการทำงานหลักของระบบคือการที่ผู้ใช้แจ้งปัญหาเข้ามาแล้วผู้ดูแลระบบตอบสนองต่อปัญหานั้น มีขั้นตอนดังนี้

1. โดอะล็อกโพลส่งข้อมูลที่ได้จากผู้ใช้ไปประมวลผลที่ไฟเบสฟังก์ชัน



ภาพที่ 5: Sequence diagram

2. ไฟเบสฟังก์ชัน เขียนข้อมูลเก็บไว้ในฐานข้อมูลไฟสโตร์
3. ไฟเบสฟังก์ชัน ส่งข้อความกลับไปบอกผู้ใช้งานถึงลำดับคิวและส่งข้อความแจ้งเตือนงานใหม่ เข้าไลน์กลุ่ม
4. ผู้ดูแลระบบกด “รับงาน” ผ่านบัญชีทางการของไลน์หรือโมบายแอป ซึ่งจะประมวลผลที่ไฟเบสฟังก์ชันและเขียนข้อมูลเก็บไว้ในฐานข้อมูลไฟสโตร์
5. ไฟเบสฟังก์ชัน ส่งข้อความกลับไปบอกผู้ใช้งานถึงชื่อของผู้รับงาน และส่งข้อความแจ้งเตือนการรับงาน เข้าไลน์กลุ่ม

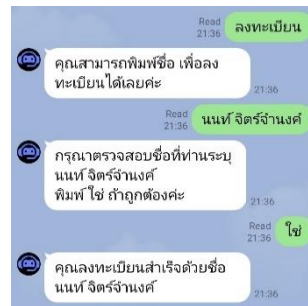
6. ผู้ดูแลระบบกด “เสร็จงาน” พร้อมระบุถึง สิ่งที่ทำ สิ่งที่เป็นสาเหตุ หรือหมายเหตุ ผ่านบัญชีทางการของไลน์หรือโมบายแอป ซึ่งจะประมวลผลที่ไฟเบสฟังก์ชันและเขียนข้อมูลเก็บไว้ในฐานข้อมูลไฟสโตร์

7. ไฟเบสฟังก์ชัน ส่งข้อความกลับไปบอกผู้ใช้งานถึงชื่อของผู้เสร็จงานพร้อมหมายเหตุ และส่งข้อความแจ้งเตือนการเสร็จงาน เข้าไลน์กลุ่ม

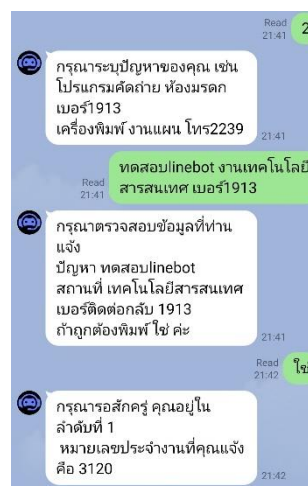
3.3 การพัฒนาระบบต้นแบบสำหรับระบบแจ้งและติดตามปัญหาข้อขัดข้องทางเทคโนโลยีสารสนเทศ

ส่วนที่ 1 การทำงานผ่านบัญชีทางการของไลน์ของผู้ใช้และผู้ดูแลระบบ

-เมื่อผู้ใช้ทำการเพิ่มบัญชีทางการของไลน์เป็นเพื่อน จะต้องทำการลงทะเบียนการใช้งานเพื่อระบุชื่อที่ใช้ตามภาพที่ 6



ภาพที่ 6: การลงทะเบียน



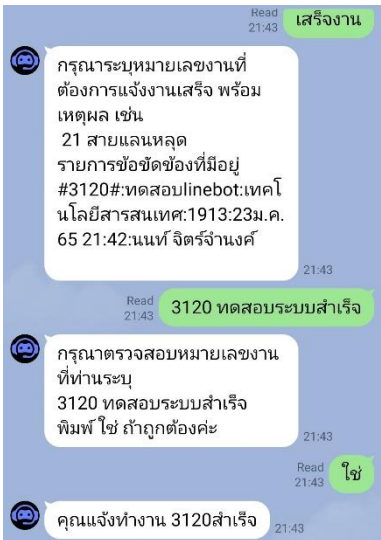
ภาพที่ 7: การแจ้งปัญหา

-ผู้ใช้สามารถพิมพ์คำว่า “เมนู” เพื่อดูรายการที่สามารถทำได้

- ผู้ใช้สามารถพิมพ์คำว่า “แจ้งปัญหา” หรือ “2” เพื่อแจ้งปัญหาหรือแจ้งสิ่งที่ต้องการได้ ตามภาพที่ 7
- ผู้ดูแลระบบสามารถรับงานได้ ตามภาพที่ 8
- ผู้ดูแลระบบสามารถแจ้งทำงานเสร็จได้ ตามภาพที่ 9

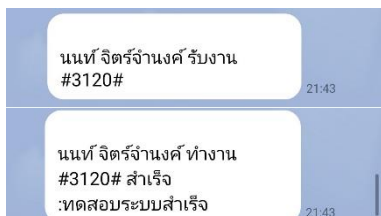


ภาพที่ 8: การแจ้งรับงาน



ภาพที่ 9: การแจ้งทำงานเสร็จ

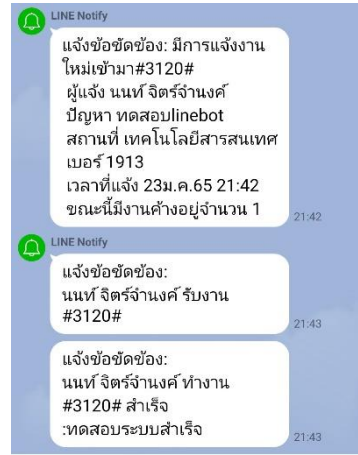
- เมื่อมีการรับงานหรือแจ้งทำงานเสร็จจะมีการแจ้งเตือนไปยังผู้แจ้งและกลุ่มของผู้ดูแลระบบได้ ตามภาพที่ 10 และ 11



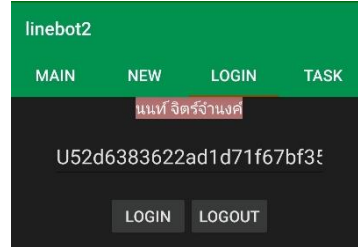
ภาพที่ 10: การแจ้งเตือนกลับไปยังผู้ใช้

ส่วนที่ 2 การใช้งานของผู้ดูแลระบบผ่านโมบายแอป

- ผู้ดูแลระบบสามารถลงชื่อเข้าใช้ด้วย userid ของบัญชีทางการของไลน์และเมื่อเข้าระบบสำเร็จ จะปรากฏชื่อที่ลงทะเบียนตามภาพที่ 12



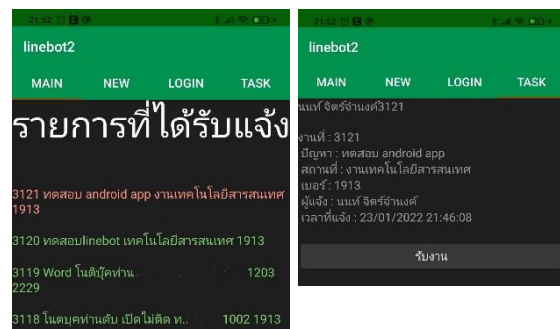
ภาพที่ 11: การแจ้งเตือนไปยังไลน์กลุ่มผู้ดูแลระบบ



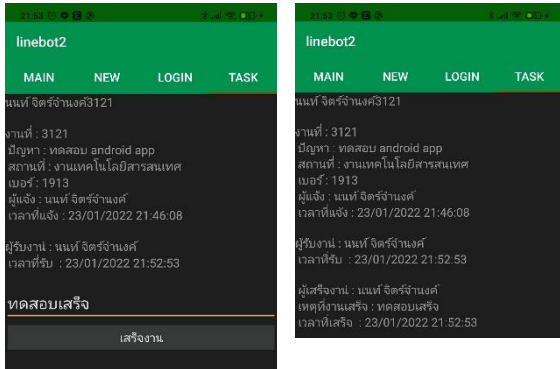
ภาพที่ 12: การเข้าระบบ

- ผู้ดูแลระบบสามารถดูรายการที่ได้รับแจ้งมาทั้งหมดโดยสีแดงหมายถึงงานที่ยังไม่เสร็จและสีเขียวหมายถึงงานที่เสร็จแล้วตามภาพที่ 13 ก

- สามารถกดดูรายละเอียดของงานแต่ละชิ้นโดยงานที่ยังไม่มีผู้รับงานจะปรากฏปุ่มรับงาน และงานที่มีผู้รับแล้วแต่ยังไม่เสร็จงานจะปรากฏปุ่มเสร็จงาน และงานที่เสร็จแล้วจะปรากฏรายละเอียดของงานทั้งหมดตามภาพที่ 13 ข – 14



ภาพที่ 13: ก) รายการที่ได้รับแจ้ง ข) รายการที่ยังไม่ได้รับ



ภาพที่ 14: ก) รายการที่ยังไม่ได้รับ ข) รายการที่เสร็จเรียบร้อยแล้ว

#### 4. ผลการดำเนินงาน

งานวิจัยนี้ได้ทำการทดลองใช้ระบบเป็นเวลา 6 เดือน มีผู้ใช้งานในระบบจำนวน 228 คน มีข้อขัดข้องแจ้งเข้ามาจำนวน 1394 ครั้ง โดยเวลาที่ไฟเบสฟังก์ชันใช้ประมวลผลเฉลี่ยเท่ากับ 0.07 วินาที มีการตอบสนองกลับมายังบัญชีทางการของไลน์ได้อย่างรวดเร็ว

การวัดความพึงพอใจด้วยแบบสอบถามของกลุ่มตัวอย่างผู้ใช้ระบบซึ่งเป็นเจ้าหน้าที่ภายในศาลแพ่งจำนวน 22 คน ได้ผลตามตารางที่ 2

ตารางที่ 2: ความพึงพอใจของระบบ

หัวข้อ	เฉลี่ย	S.D.
ความสะดวกในการใช้งาน (ติดตั้งระบบ)	4.27	1.08
ความง่ายในการแจ้งปัญหา	4.32	1.32
การติดตามผลและการแจ้งลำดับคิว	4.23	1.23
ความพึงพอใจรวมของระบบ	4.27	1.20

#### 5. สรุป แผนการต่อยอด

สรุประบบแจ้งและติดตามปัญหาข้อขัดข้องทางเทคโนโลยีสารสนเทศสามารถทำงานตามที่ออกแบบไว้ ผู้ใช้สามารถแจ้งปัญหาผ่านระบบได้โดยง่ายไม่ต้องติดตั้งแอปพลิเคชันเพิ่มเติม ผู้แจ้งสามารถรับทราบถึงผู้ที่มาแก้ไขปัญหาให้ ทั้งทราบถึงลำดับคิวของงานที่แจ้งไว้ด้วย ในส่วนของผู้ดูแลระบบ ยังช่วยลดการตกหล่นของการแก้ไขปัญหาข้อขัดข้อง มองเห็นภาพรวมของปัญหาที่

เกิดขึ้นในช่วงเวลานั้น ทำให้สามารถเลือกแก้ไขปัญหาที่มีความสำคัญก่อนได้

แผนการต่อยอด เพิ่มฟังก์ชันการกำหนดความสำคัญของข้อขัดข้องที่ได้รับแจ้ง รวมทั้งสามารถมอบหมายให้ผู้ดูแลระบบคนใดทำงานที่กำหนดได้

#### เอกสารอ้างอิง

- [1] V. R. Palilingan, and J. R. Batmetan. "Incident Management in Academic Information System using ITIL Framework." *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 306, no. 1, Feb, 2018.
- [2] J. Cao, and S. Zhang. "ITIL Incident Management Process Reengineering in Industry 4.0 Environments." *atlantis-press.com*, 2016.
- [3] ขวัญฤดี ฮวดหุ่น. "อิทธิพลของแอปพลิเคชันไลน์ในการสื่อสารยุคปัจจุบัน." *วารสารศิลปการจัดการ*, ปีที่ 1, ฉบับที่ 2, พฤษภาคม – สิงหาคม, หน้า 75-88, 2560.
- [4] วิศปภัช ชัยช่วย. "การใช้LINEของผู้สูงอายุ: การศึกษาเชิงปรากฏการณ์วิทยา." *Veridian E-Journal, Silpakorn Univ.* ฉบับภาษาไทย สาขามนุษยศาสตร์ สังคมศาสตร์ และศิลปะ, ปีที่ 10, ฉบับที่ 1, เดือนมกราคม – เมษายน, หน้า 905-918, 2560.
- [5] A. Hutapea. "CHATBOT: Architecture, Design, & Development." University of Pennsylvania, 2017.
- [6] D. Zumstein, and S. Hundertmark. "Chatbots-An Interactive Technology for Personalized Communication, Transactions and Services Chatbot View project." *IADIS Int. J. WWW/Internet*, Vol. 15, No. 1, pp. 96–109, 2017.

# การทดสอบประสิทธิภาพเครื่องมือแปลงไฟล์พีดีเอฟเป็นซีเอสวีเท็กซ์ไฟล์

## PERFORMANCE TESTING OF CONVERTING TOOLS FROM PDF FILE TO CSV TEXT FILE

ภูธรเศ สุวรรณรัตน์ (Phutares Suwannarat)<sup>1</sup> และพรฤดี เนติโสภาคกุล (Ponrudee Netisopakul)<sup>2</sup>

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

<sup>1</sup>phutares.pax@gmail.com, <sup>2</sup>ponrudee@it.kmitl.ac.th

### บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการตรวจจับตาราง ความถูกต้องของการแปลงข้อมูลตัวอักษรของเครื่องมือจากซอฟต์แวร์โอเพนซอร์ส ข้อมูลที่นำมาใช้เป็นพีดีเอฟที่มีข้อมูลภายในเป็นภาษาอังกฤษ ใบแสดงราคาสินค้า ใบแจ้งหนี้ และอื่นๆ ในรูปแบบตารางแปลงพีดีเอฟให้เป็นข้อมูลซีเอสวีเท็กซ์ไฟล์อย่างเหมาะสมเพื่อลดระยะเวลาในการนำข้อมูลจากพีดีเอฟใช้งานและเพิ่มความถูกต้องของการนำไปใช้ข้อมูล การเปรียบเทียบครั้งนี้ได้ครอบคลุมถึงกระบวนการเขียนโปรแกรมด้วยภาษาไพทอนในการเรียกใช้เครื่องมือจากซอฟต์แวร์โอเพนซอร์ส มีการวิจัยเครื่องมือที่นำมาแก้ปัญหาทั้งหมดด้วยวิธีที่คุ้มค่าและเชื่อถือได้ เพื่อให้ได้แก้ปัญหาของการแปลงเอกสารพีดีเอฟแต่ละประเภท กระบวนการวิจัยอธิบายเปรียบเทียบความถูกต้องข้อมูล และผลลัพธ์ของเครื่องมือ

**คำสำคัญ:** โอเพนซอร์ส พีดีเอฟ ไพทอน ซีเอสวี

### Abstract

Research aims to compare performance, table detection and accuracy of data extraction in character by software open source. Data are containing English as quotation invoice and others, those are tabular structure. Then converting PDF data into CSV text files automation appropriately, to reduce time spending for insertion of PDF data and increasing accuracy. In the comparison it covers the process of coding by Python programming language

for executing reputable software open source. To get a solution to the problem of converting, thus this research for each converting tools that can solve all problems in a cost-effective and the most reliable tools. The research process has described about accuracy comparing of data, followed by the results of each converting tools.

**Keywords:** Open-source, PDF, Python, CSV.

### 1. บทนำ

ในปัจจุบันทางบริษัทมีการนำเอกสารพีดีเอฟของลูกค้าผู้ขายรถยนต์ มีข้อมูลภายในรูปแบบของตาราง มีประเภทพีดีเอฟแบบเท็กซ์เบสค์เป็นรูปแบบที่สามารถพิมพ์ค้นหาข้อความได้ และอิมเมจเบสค์เป็นรูปแบบที่ไม่สามารถพิมพ์ค้นหาข้อความได้ แล้วนำข้อมูลกรอกเข้าระบบของบริษัทที่เป็นระบบบริหารจัดการขายรถยนต์ด้วยพนักงานในบริษัท ซึ่งในปัจจุบันการกรอกข้อมูลจำนวนมากนั้นใช้เวลานานและยังมีความผิดพลาด

เนื่องจากเอกสารพีดีเอฟของลูกค้าผู้ขายรถยนต์มีสองรูปแบบเท็กซ์เบสค์และอิมเมจเบสค์ทำให้ วัตถุประสงค์ในการวิจัยนี้ได้นำเสนอการศึกษาและการทดสอบประสิทธิภาพเครื่องมือแปลงไฟล์พีดีเอฟทั้งประเภทเท็กซ์เบสค์และอิมเมจเบสค์เป็นซีเอสวีเท็กซ์ไฟล์ด้วยซอฟต์แวร์โอเพนซอร์สในด้านความถูกต้องและความเร็วในการประมวลผล โดยมีทั้งในรูปแบบแพ็คเกจที่ใช้งานในภาษาไพทอนและเอพีไอ รวมถึงนำเครื่องมือที่ใช้งานง่ายและมีราคาถูกหรือไม่มีค่าใช้จ่ายหรือต่ำ

## 2. ทฤษฎีที่เกี่ยวข้อง

### 2.1 โอซีอาร์ (OCR)

โอซีอาร์ (OCR) หรือ Optical Character Recognition หรือการรู้จำตัวอักษรเป็นเทคโนโลยีการจำข้อความจากรูปภาพโดยอัตโนมัติ ใช้ประเภทรูปภาพได้หลากหลาย เช่น JPG, PNG และเอกสาร PDF โดยทำการวิเคราะห์ภาพ แล้วแปลงภาพตัวอักษรให้เป็นตัวอักษร ผลลัพธ์จะออกมาแตกต่างกันขึ้นอยู่กับรูปภาพที่นำไปประมวลผล การได้ความถูกต้องของข้อมูลทั้งหมดเป็นไปได้ยาก แต่สามารถปรับปรุงได้ด้วยการประยุกต์ใช้งานกับข้อมูลพจนานุกรม ด้วยการจดจำอักษร เข้ากระบวนการเรียนรู้ และจดจำคำทั้งหมดในชุดข้อมูลของพจนานุกรมที่ใช้ และมีปัจจัยหลายอย่างที่มีผลต่อการประมวลผลของเทคโนโลยี OCR เช่น ขนาดตัวอักษร รูปแบบตัวอักษร สัญลักษณ์ สีและความเข้มของพื้นหลัง รวมถึงเงาและแสงบนรูปภาพ แต่หากมีเอกสารหรือรูปภาพที่มีความละเอียดสูง เทคโนโลยี OCR จะทำงานได้อย่างมีประสิทธิภาพ และได้ผลลัพธ์ออกมาดีที่สุด [1]

ขั้นตอนการทำงานของเทคโนโลยี OCR ทั้งหมดมี 5 ขั้นตอน

#### 2.1.1 เตรียม การก่อน การประมวลผล (Pre Processing)

การปรับขนาดรูปภาพ การหมุนแกนภาพ การแปลงรูปภาพให้เป็นขาวดำ การจัดการกับเงาหรือแสงบนรูปภาพ การจัดการกับสิ่งรบกวนอื่นๆบนรูปภาพ ในภาษาไพทอนสามารถใช้เครื่องมืออย่าง OpenCV ได้

#### 2.1.2 การแบ่งส่วน (Segmentation)

การจัดแบ่งกลุ่มข้อความที่พบเจอบนเอกสารเพื่อนำไปประมวลผลเพิ่มเติม โดยมีหลายระดับที่สามารถทำได้ ตั้งแต่ การแบ่งระดับบรรทัด การแบ่งระดับคำ การแบ่งระดับตัวอักษร เช่นการวัดการแบ่งกลุ่มคำด้วยแนวอนของตัวอักษรด้วย Horizontally Histogram Projection การค้นหาเส้นระหว่างบรรทัดของคำ การวัดการแบ่งกลุ่มตัวอักษรตามแนวตั้งด้วย Vertical Histogram Projection หรือหาความกว้างของรูป เทคนิคนี้สามารถช่วยค้นหาแบ่งกลุ่มคำได้ (Word Segmentation) ดังนั้นเมื่อนำสอง

เทคนิคนี้รวมเข้าด้วยกันจะได้ชุดข้อความ แต่ละบรรทัด และแต่ละกลุ่มคำ

#### 2.1.3 การสกัดคุณลักษณะ (Feature Extraction)

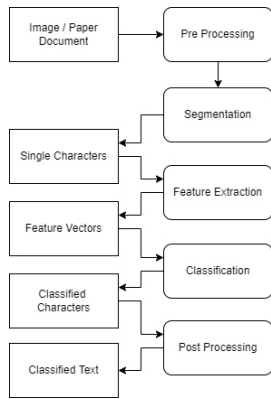
การดึงโครงสร้างพื้นฐานที่สำคัญของตัวอักษรที่พยายามจำแนกคุณลักษณะต่างๆ ของอักขระในภาพให้อยู่ในรูปแบบเวกเตอร์ซึ่งเป็นเทคนิคที่นิยมในปัจจุบัน โดยจะต้องมีการกำหนด เส้นตรง เส้นเอียง หัว ส่วนโค้ง ส่วนเว้า จุดแตกกิ่ง จุดตัด ไว้เมื่อแยกองค์ประกอบตัวอักษรออกมาได้ จะนำเสนอรูปภาพของตัวอักษรในรูปแบบที่มีองค์ประกอบพื้นฐานต่างๆ แทน ซึ่งจะนำไปเป็นข้อมูลนำเข้าในกระบวนการแยกประเภทหรือการรู้จำต่อไป

#### 2.1.4 การแยกประเภทหรือการรู้จำ (Classification / Recognition)

ในขั้นตอนนี้เป็นขั้นตอนการนำเวกเตอร์จากคุณลักษณะนำมาแยกประเภท โดยต้องมีคลังข้อมูลเพื่อเปรียบเทียบกับเวกเตอร์คุณลักษณะต่างๆ ขั้นตอนนี้อยู่กับการฝึกอบรมตัวแบบจำลองและจำนวนพารามิเตอร์ ในการฝึกอบรมแบบจำลองมีเทคนิคหลากหลาย เช่น ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) และโครงข่ายประสาทเทียม (Neural Network) ที่เป็นที่ยอมรับในปัจจุบัน

#### 2.1.5 หลังการประมวลผล (Post processing)

การตรวจจับและแก้ไขการสะกดผิดของไวยากรณ์จากผลลัพธ์ที่ได้จากเทคโนโลยี OCR การตรวจจับนั้นสามารถชุดข้อมูลพจนานุกรมมาเปรียบเทียบกับผลลัพธ์ที่ได้รับออกมา และการแก้ไขการสะกดผิดของไวยากรณ์สามารถนำผลลัพธ์ที่เป็นชุดคำอย่างเช่น bigram เข้ามาเปรียบเทียบกับแบบจำลองที่สร้างไว้สำหรับ bigram หรือข้อความบางส่วนที่หายไปสามารถใช้ Noisy Channel Model เพื่อแนะนำตัวอักษรบางตัวที่อยู่ข้างๆ ที่ขาดหายไป [2][3]



ภาพที่ 1: ขั้นตอนของกระบวนการรู้จำตัวอักษร [2]

## 2.2 ประเภทพีดีเอฟ

เอกสาร PDF สามารถแบ่งออกได้เป็น 3 ประเภทขึ้นอยู่กับวิธีการสร้างไฟล์

### 2.2.1 True or Digitally Created PDFs (Text Based)

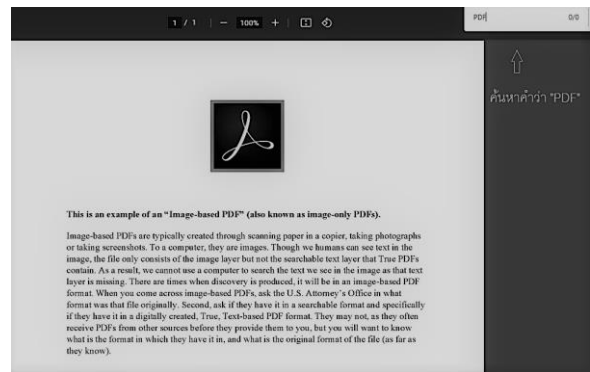
ถูกสร้างขึ้นโดยใช้ซอฟต์แวร์ เช่น Microsoft Word, Excel หรือผ่าน "พิมพ์" แปลงให้เป็น PDF ด้วยซอฟต์แวร์ประกอบด้วยข้อความและรูปภาพ ใส่ข้อความเพิ่มเติม ค้นหา คัดลอก วางข้อความและรูปภาพได้ โดยชั้นของข้อมูลประกอบด้วยชั้นรูปภาพและชั้นข้อความโดยชั้นข้อความเหมือนกับเอกสารดั้งเดิมที่ได้ถูกแปลงมา [4]



ภาพที่ 2: เอกสารพีดีเอฟประเภท Text-based [4]

### 2.2.2 Image-only or Scanned PDFs (Imaged Based)

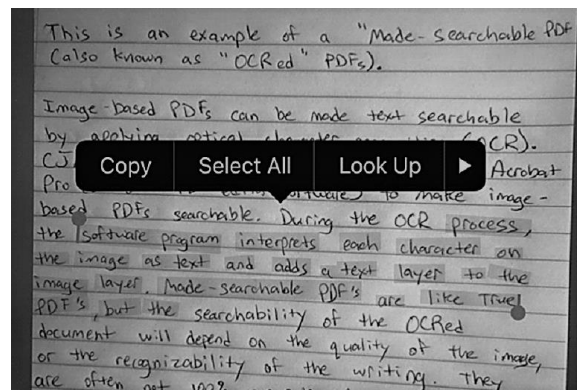
PDF ที่ถูกสแกนเป็นรูปภาพจากเอกสาร เนื้อหาถูกล็อกไว้ในภาพ คล้ายกับการแปลงภาพจากกล้องเป็น PDF เอกสารประเภทนี้ไม่สามารถค้นหาข้อความ แก้ไข คัดลอกวางข้อความได้ จึงไม่มีชั้นของข้อความอยู่ เช่นการสแกนเอกสารผ่านเครื่องถ่ายเอกสาร [4]



ภาพที่ 3: เอกสารพีดีเอฟประเภท Image-based [4]

### 2.2.3 Made-Searchable PDFs

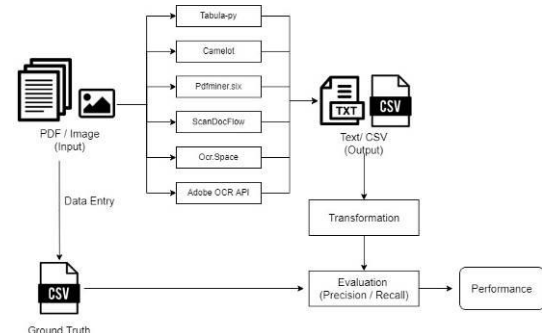
PDF ที่ได้ผ่านกระบวนการของโอซีอาร์แล้วส่วนใหญ่เกิดจากการภาพถ่ายเอกสารแล้วนำข้อมูลเข้าเทคโนโลยีโอซีอาร์เพื่อได้ชั้นของข้อความเพิ่มเข้าไปในเอกสาร ทำให้สามารถค้นหา คัดลอก วางข้อความได้ แต่ความถูกต้องจะไม่ได้ 100% เช่น Google Translate API [4]



ภาพที่ 4: เอกสารพีดีเอฟประเภท Made-Searchable [4]

## 3. วิธีการดำเนินการวิจัย

การวิจัยในครั้งนี้คือการทดสอบประสิทธิภาพเครื่องมือ OCR กับเอกสาร PDF ที่แปลงให้ได้ผลลัพธ์ในรูปแบบข้อความตัวอักษรอย่างอัตโนมัติ



ภาพที่ 5: กระบวนการทดสอบประสิทธิภาพ



กระบวนการที่ใช้ในการทดสอบประสิทธิภาพในครั้งนี้แบ่งออกเป็น 5 ขั้นตอน

### 3.1 ข้อมูลนำเข้า

นำเอกสารข้อมูลเกี่ยวกับใบแสดงราคาสินค้ารถยนต์และอื่นๆ มีรูปแบบ Text-based และ Image-based ในขั้นตอนนี้จะมีการแปลงเอกสารพีดีเอฟไปเป็นรูปภาพและปรับขนาดรูปภาพ ขึ้นอยู่กับการรองรับของเครื่องมือ

### 3.2 ข้อมูลต้นฉบับ

กรอกข้อมูลที่อยู่ในเอกสารพีดีเอฟที่นำไปเป็นข้อมูลต้นฉบับเก็บไว้ในรูปแบบตารางเพื่อเตรียมไว้สำหรับการประเมินผลลัพธ์ มีการเข้ารหัส Base64 แปลงไปดัดที่มีข้อมูลไบนารีหรือข้อความอักขระ ASCII การเข้ารหัสข้อมูลของเราช่วยเพิ่มโอกาสในการประมวลผลข้อมูลได้อย่างถูกต้องอย่างเครื่องมือ ScanDocFlow จำเป็นต้องใช้การเข้ารหัส Base64 encoding

### 3.3 การเรียกใช้เครื่องมือโอซีอาร์

การนำข้อมูลเข้าเครื่องมือ โอซีอาร์สำหรับแพ็คเกจของเครื่องมือที่สามารถติดตั้งลงเครื่องคอมพิวเตอร์แล้วเรียกใช้ หรือทำการขอคำร้องไปที่ API ของเครื่องมือด้วย POST method ด้วยภาษาไพทอน

### 3.4 การแปลงข้อมูล

การเตรียมข้อมูลเพื่อการประเมิน เพื่อให้สอดคล้องกับการประเมินกับข้อมูลต้นฉบับ เป็นรูปแบบ string หรือ list ในตัวแปร

### 3.5 การประเมินประสิทธิภาพ

เนื่องจากการวัดความแม่นยำโดยปกติจะใช้การจับคู่ตรงกันหรือไม่ตรงกัน ไม่สามารถประเมินประสิทธิภาพได้เพียงพอ การใช้อัตราความผิดพลาด (Error Rate) และ Distance Levenshtein เป็นขั้นตอนวิธีการวัดหาค่าความต่างกันของอักขระสองชุด โดยมีข้อความจากเครื่องมือโอซีอาร์และข้อความต้นฉบับ (Ground Truth) และเนื่องจากผลลัพธ์ยังมีลักษณะข้อผิดพลาดของคำที่ต้องพิจารณาได้ เช่น ข้อผิดพลาดในการแทนที่ (Substitution error) จะเป็นตัวอักษรหรือคำที่สะกดผิด ข้อผิดพลาดในการลบ (Deletion error) ตัวอักษรหรือคำที่หายไป ข้อผิดพลาดในการแทรก (Insertion error) การแทรกอักขระหรือคำที่ไม่

ถูกต้องเพิ่มเข้าไปเกิน ดังนั้นการวัดข้อผิดพลาดของระหว่างข้อความสองชุดสามารถทำได้ด้วย Distance Levenshtein โดยมีสองตัวชี้วัดที่ใช้ในการประเมินประสิทธิภาพคือ 1) อัตราข้อผิดพลาดของตัวอักษร (Character Error Rate) [5]

$$CER = \frac{S + D + I}{N}$$

ภาพที่ 6: สมการอัตราข้อผิดพลาดของตัวอักษร [5]

S คือจำนวนการแทนที่

D คือจำนวนการลบ

I คือจำนวนการแทรก

N คือจำนวนของตัวอักษรต้นฉบับ (Ground Truth)

โดย N สามารถคำนวณได้ด้วย  $N = S + D + C$  โดย C คือจำนวนอักขระที่ถูกต้อง ผลลัพธ์คือการทำนายที่ไม่ถูกต้องจากเครื่องมือ โอซีอาร์ยังคงคะแนนของสมการนี้ต่ำหรือใกล้ 0 หมายถึงเครื่องมือ โอซีอาร์มีประสิทธิภาพสูง และมีปัญหาหนึ่งสำหรับสมการนี้คือถ้าคะแนนของสมการนี้เกิน 100 เปอร์เซ็นต์สามารถเกิดขึ้นได้จากการแทรกข้อความเข้าไปจำนวนมากกว่าต้นฉบับ 2) อัตราข้อผิดพลาดของคำ (Word Error Rate)

$$WER = \frac{S_w + D_w + I_w}{N_w}$$

ภาพที่ 8: สมการอัตราข้อผิดพลาดของคำ

เป็นการวัดข้อความที่เกี่ยวกับข้อความที่มาเป็นชุดเช่น รหัสสินค้า หมายเลขโทรศัพท์ โดยการวัดข้อผิดพลาดของคำเหมือนกับการวัดข้อผิดพลาดของตัวอักษรแต่ใช้ระดับของคำ คือจำนวนของคำที่เกิดการแทนที่ การลบ หรือการแทรกที่สามารถเปลี่ยนความหมายของประโยคได้ โดยคะแนนของสมการของการวัดข้อผิดพลาดของคำสามารถมีคะแนนที่สูงได้กว่าการวัดข้อผิดพลาดของตัวอักษร [5] 3) อัตราข้อผิดพลาดของตัวอักขระพิเศษ (Special Symbols Error Rate) เช่น (=, +, \*, -, /, ^, %, #, |, ?, !, @) [1] และ 4) อัตราข้อผิดพลาดของตัวเลข (Numerical Error Rate) เช่น 0, 10, 100, 9999 ใช้สมการเดียวกันกับการวัดข้อผิดพลาดของตัวอักษรแต่ใช้ของอักขระพิเศษและของตัวเลข

### 3.6 การทดลองเบื้องต้น

ทดสอบประสิทธิภาพเครื่องมือโอซีอาร์กับเอกสารใบแสดงราคาสินค้าและอื่นๆ เฉพาะพีดีเอฟประเภท Text-based จำนวน 5 เอกสาร และประเภท Image-based จำนวน 5 เอกสารที่เป็นภาษาอังกฤษ ใช้การประเมินผลแบบ CER WER SER และ NER การทดลองเบื้องต้นพัฒนาบนภาษาไพทอน โดยใช้แพ็คเกจที่สำคัญในการประเมินผล CER และ WER คือ Asrtoolkit เพื่อลดความซับซ้อนของขั้นสำหรับการสร้าง ปรับแต่ง วิเคราะห์ และเป็นเครื่องมือที่ช่วยประเมินผลของเทคโนโลยีรู้จำเสียงพูด (Automatic Speech Recognition) ส่วนของ SER และ NER ทำการเขียน โปรแกรมด้วยภาษาไพทอนในการคำนวณขึ้นมา โดยนำจำนวนตัวเลขและตัวอักษรพิเศษที่เจอและถูกต้องมาหารด้วยจำนวนทั้งของต้นฉบับแล้วลบด้วย 100 ซึ่งจะได้ค่า SER และ NER ตามสมการ

**ตารางที่ 1:** ค่าเฉลี่ยอัตราความผิดพลาดของตัวอักษร อัตราความผิดพลาดของคำ อัตราความผิดพลาดของตัวระบุพิเศษและอัตราความผิดพลาดของตัวเลขในแต่ละเครื่องมือกับเอกสารประเภท Text-based

เครื่องมือ	ผลการประเมิน CER	ผลการประเมิน WER	ผลการประเมิน SER	ผลการประเมิน NER
Tabula-py	7.67%	8.67%	2.5%	8.15%
Camelot	22.2%	24.43%	16.5%	21.3%
Pdfminer.six	5.25%	16.7%	2%	21.54%
ScanDocFlow	15.63%	21.53%	-50.22%	26.95%
OCR.Space	54.87%	58.11%	17.16%	60.31%
Adobe OCR API	6.102%	10.82%	4.5%	11.6%

จากผลการทดลองเบื้องต้นผลลัพธ์ของเอกสารประเภท Text-based ได้เครื่องมือที่ใช้งานที่ดีที่สุดเป็นลำดับที่หนึ่งคือ Tabula-py เนื่องจากมีคะแนน CER WER SER และ NER รวมกันได้น้อยที่สุด หมายความว่าเครื่องมือนี้มีประสิทธิภาพการแปลงเอกสารประเภทพีดีเอฟไปเป็นซีเอสวีเท็กซ์ไฟล์จะมีข้อผิดพลาดทางด้านตัวอักษร คำอักขระพิเศษ และตัวเลขน้อย ลำดับที่ 2. Adobe OCR API 3. Pdfminer.six 4. Camelot 5. ScanDocFlow และลำดับสุดท้ายคือ OCR.Space แม้ว่า ScanDocFlow จะมีค่าอัตราความผิดพลาดของตัวอักขระพิเศษอยู่ติดลบ หมายความว่าจำนวนตัวอักขระพิเศษมีมากกว่าจำนวน

ต้นฉบับ แต่อัตราความผิดพลาดของตัวอักษรและคำอยู่ในเกณฑ์ที่ดีกว่ามาก

**ตารางที่ 2:** ค่าเฉลี่ยอัตราความผิดพลาดของตัวอักษร อัตราความผิดพลาดของคำ อัตราความผิดพลาดของตัวระบุพิเศษและอัตราความผิดพลาดของตัวเลขในแต่ละเครื่องมือกับเอกสารประเภท Image-based

เครื่องมือ	ผลการประเมิน CER	ผลการประเมิน WER	ผลการประเมิน SER	ผลการประเมิน NER
Tabula-py	97.15%	100%	52.5%	94.87%
Camelot	97.15%	100%	52.5%	94.87%
Pdfminer.six	100%	100%	100%	100%
ScanDocFlow	28.39%	32.79%	-29.83%	27.71%
OCR.Space	46.17%	42.19%	-8.66%	54.95%
Adobe OCR API	41.896%	48.47%	-39.66%	37.23%

จากผลการทดลองเบื้องต้นผลลัพธ์ของเอกสารประเภท Image-based ได้เครื่องมือที่ใช้งานที่ดีที่สุดเป็นลำดับที่หนึ่งคือ ScanDocFlow เนื่องจากมีคะแนน CER WER SER และ NER รวมกันได้น้อยที่สุด หมายความว่าเครื่องมือนี้มีประสิทธิภาพการแปลงเอกสารประเภทพีดีเอฟไปเป็นซีเอสวีเท็กซ์ไฟล์จะมีข้อผิดพลาดทางด้านตัวอักษร คำอักขระพิเศษ และตัวเลขน้อย ถึงแม้จะมีพบตัวอักขระพิเศษจำนวนมากกว่าในข้อมูลต้นฉบับก็ตาม ลำดับที่ 2. Adobe OCR API 3. OCR.Space ในลำดับที่ 4 และ 5 คือ Tabula-py และ Camelo ซึ่งมีคะแนนเท่ากัน ลำดับสุดท้ายคือ Pdfminer.six ซึ่งไม่สามารถค้นหาข้อความตัวอักษร คำ ตัวอักขระพิเศษและตัวเลขในเอกสารประเภท Image-based ได้เลย

### 4. บทสรุป

การทดสอบนี้เป็นการทดสอบประสิทธิภาพเครื่องมือแปลงเอกสารพีดีเอฟเป็นเท็กซ์ไฟล์หรือผลลัพธ์อยู่ในรูปแบบข้อความตัวอักษรได้อย่างอัตโนมัติซึ่งประกอบไปด้วย 5 ขั้นตอนหลัก ขั้นตอนแรกการเตรียมข้อมูลนำเข้าทั้งรูปแบบ Text-based และ Image-based ขั้นตอนการเตรียมข้อมูลต้นฉบับไว้สำหรับประเมินประสิทธิภาพกับผลลัพธ์จากเครื่องมือโอซีอาร์ ขั้นตอนที่มีการเรียกใช้เครื่องมือแต่ละเครื่องมือโอซีอาร์ในแต่ละรูปแบบทั้งรูปแบบ APIs ที่จำเป็นต้องมี Access Key หรือไม่จำเป็นต้องมี หรือรูปแบบติดตั้ง Package ลงบนอุปกรณ์

ขั้นตอนแปลงข้อมูลทั้งข้อมูลต้นฉบับและผลลัพธ์ที่ได้จาก เครื่องมือโอซีอาร์ให้อยู่ในรูปแบบ Text แบบตามลำดับ และแยกคำที่ถูกต้องเพื่อนำไปประเมินประสิทธิภาพ ขั้นตอนสุดท้ายการประเมินประสิทธิภาพเครื่องมือโอซีอาร์ การวัดด้วยอัตราความผิดพลาดของตัวอักษร อัตราความผิดพลาดของคำ อัตราความผิดพลาดของตัวอักษร พิเศษและอัตราความผิดพลาดของตัวเลข โดยผลการเปรียบเทียบระหว่างเอกสารประเภท Text-based กับ Image-based กับแต่ละเครื่องมือ

จากการทดลองเอกสารใบแสดงราคาสินค้า ที่เป็นเอกสารพีดีเอฟประเภท Text-based มีเครื่องมือที่สะดวกต่อการใช้งานคือ Tabula-py เนื่องจากมีฟังก์ชันที่สามารถแปลงข้อมูลในตารางให้อยู่ในรูปแบบโครงสร้างได้แบบอัตโนมัติและมีผลการประเมินประสิทธิภาพที่อยู่ในเกณฑ์สูง และเอกสารพีดีเอฟประเภท Image-based มีเครื่องมือที่สะดวกต่อการใช้งานคือ ScanDocFlow และมีความจำเป็นที่ต้องตรวจสอบข้อมูลก่อนนำไปใช้ด้วยมนุษย์อีกครั้ง

สำหรับงานวิจัยนี้การวัดด้วยอัตราความผิดพลาดของตัวอักษร อัตราความผิดพลาดของคำ อัตราความผิดพลาดของตัวอักษรพิเศษและอัตราความผิดพลาดของตัวเลขกับเอกสารพีดีเอฟประเภท Text-based จำนวน 5 ชุดและ Image-based จำนวน 5 ชุด ซึ่งสังเกตเห็นได้ว่าเครื่อง Tabula-py Camelot และ Pdfminer.six ไม่สามารถทำงานกับเอกสารพีดีเอฟประเภท Image-based ได้ แต่ทำงานได้ดีมากกับการแปลงเอกสารประเภท Text-based เป็นซีเอสวีเท็กซ์ไฟล์ รวมถึงสามารถใช้งานได้ง่ายต่อการใช้งาน และเครื่องมืออย่าง ScanDocFlow Adobe OCR API และ OCR.Space สามารถทำงานได้กับพีดีเอฟทั้งสองประเภทแต่ก็จะมีค่าใช้จ่ายจำนวนหนึ่งแต่ไม่สูงมาก สำหรับ ScanDocFlow และ OCR.Space แต่ Adobe OCR API สามารถใช้งานได้ฟรี ง่ายต่อการใช้งานและมีข้อจำกัด

สังเกตเห็นได้ว่าความเหมาะสมกับเอกสารในแต่ละด้าน ตัวอักษรหรือคำ อักษรพิเศษและตัวเลข ในประเภท Text-based เครื่องมือ Pdfminer.six เหมาะสมกับเอกสาร

ประเภทที่มีอักษรพิเศษที่สุด เครื่องมือ Tabula-py เหมาะสมกับงานด้านตัวอักษรหรือคำ และตัวเลขมากที่สุด ในความเหมาะสมกับเอกสารในแต่ละด้าน ตัวอักษรหรือคำ อักษรพิเศษและตัวเลข ในประเภท Image-based เครื่องมือ OCR.Space เหมาะสมกับเอกสารประเภทที่มีอักษรพิเศษที่สุด เครื่องมือ ScanDocFlow เหมาะสมกับงานด้านตัวอักษรหรือคำ และตัวเลขมากที่สุด

### เอกสารอ้างอิง

- [1] S.Vijayarani, and A.Sakila. "Performance Comparison of OCR Tools." *International Journal of UbiComp (IJU)*, Vol.6, No.3, July, 2015.
- [2] G. Y. Tawde, and J. M. Kundargi. "An Overview of Feature Extraction Techniques in OCR for Indian Scripts Focused on Offline Handwriting." *International Journal of Engineering Research and Applications (IJERA)*, Vol. 3, No. 1, Jan – Feb, pp. 919-926, 2013.
- [3] P. Singh, and S. Budhiraja. "Feature Extraction and Classification Techniques in O.C.R. Systems for Handwritten Gurmukhi Script – A Survey." *International Journal of Engineering Research and Applications (IJERA)*, Vol. 1, No. 4, September, pp. 1736-1739, 2017.
- [4] S. Broderick. Three Types of PDFs. Available Online at <https://nlsblog.org/2020/06/12/three-types-of-pdfs/>, accessed on 12 June 2020.
- [5] K. Leung. Evaluate OCR Output Quality with Character Error Rate (CER) and Word Error Rate (WER). Available Online at <https://towardsdatascience.com/evaluating-ocr-output-quality-with-character-error-rate-cer-and-word-error-rate-wer-853175297510>, accessed on 24 June 2021.

# การจับคู่ข้อความโดยใช้ค่าคะแนน BLEU สำหรับระบบสั่งอาหารด้วยเสียงพูด

## Text Matching using BLEU Score for Voice-based Food Ordering System

กฤตเมธ มนเทียรกุล (Kittamet Montheankul)<sup>1</sup> วณิชพร มาสมบูรณ์ (Vanitchaporn Masomboon)<sup>2</sup>

ศุภสุดา สดประเสริฐ (Supasuda Sudprasert)<sup>3</sup> และ โสภณ ผู้มีจรรยา (Sopon Phumeechanya)<sup>4</sup>

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์และเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยศิลปากร

<sup>1</sup>montheankul\_k@su.ac.th, <sup>2</sup>masomboon\_v@su.ac.th, <sup>3</sup>sudprasert\_s3@su.ac.th, <sup>4</sup>phumeechanya\_s@su.ac.th

### บทคัดย่อ

บทความนี้นำเสนอวิธีการจับคู่ข้อความเพื่อใช้ในการรับรู้เจตนาของผู้ใช้งานสำหรับนำมาประยุกต์ใช้ในการสร้างระบบสั่งอาหารด้วยเสียงพูด โดยการนำค่าคะแนน BLEU มาใช้เป็นตัววัดความคล้ายคลึงกันของข้อความระหว่างข้อความอินพุตที่ได้มาจากการแปลงเสียงพูดเป็นข้อความ เทียบกับข้อความในคลังข้อความอ้างอิงที่ได้เตรียมไว้ ซึ่งประกอบไปด้วยข้อความการทักทาย การสนทนาทั่วไป และเมนูอาหาร โดยระบบสั่งอาหารด้วยเสียงพูดที่ได้พัฒนาขึ้นมานี้มีความสามารถในการตรวจจับใบหน้าของผู้ใช้งานเพื่อเริ่มต้นการทักทาย มีความสามารถในการโต้ตอบการสนทนาทั่วไป การสั่งอาหารพร้อมระบุจำนวนที่ต้องการ การยกเลิกเมนูอาหารที่สั่ง การเรียกชำระเงิน การบันทึกคะแนนและข้อเสนอแนะจากการใช้บริการ โดยจากการทดลองพบว่า การจับคู่ข้อความในส่วนของ การทักทายและการสนทนาทั่วไป มีความถูกต้องที่ 92.50% และการสั่งอาหารมีความถูกต้องที่ 81.90%

**คำสำคัญ:** การจับคู่ข้อความ การรับรู้เจตนา ระบบสั่งอาหารด้วยเสียงพูด

### Abstract

This paper presents a text matching method for recognizing the intents of a user for developing a voice-based food ordering system. The BLEU score is used as a measure of the similarity between input texts that are converted from speech to text and reference texts in the prepared text bank which includes greeting messages, general conversations, and menus. Our food ordering system has the ability

to detect the user's face to start greetings, interact with common conversations, order food with the required amount, cancel some orders, call for payment and record the user's rating and feedback. The experiment results show that the accuracy of the text matching in the greetings and general conversations is 92.50%. Meanwhile, the accuracy of the text matching in the food ordering process is 81.90%.

**Keywords:** Text Matching, Intent Recognition, Voice-Based Food Ordering System.

### 1. บทนำ

การสั่งงานด้วยเสียงพูดถูกนำมาใช้อย่างแพร่หลายในปัจจุบัน เช่น การใช้เสียงพูดควบคุมแสงสว่างภายในบ้านสำหรับผู้สูงอายุ [1] การใช้เสียงพูดควบคุมอุปกรณ์ไฟฟ้า [2] การควบคุมหุ่นยนต์ด้วยเสียงพูด [3] การใช้เสียงพูดควบคุมการทำงานของคอมพิวเตอร์ [4] และการโต้ตอบด้วยเสียงพูดกับผู้ใช้ระบบ เป็นต้น

ในปัจจุบันการสั่งอาหารในร้านอาหาร ลูกค้าสามารถสั่งอาหารได้ในหลายช่องทาง เช่น สั่งอาหารด้วยปากเปล่ากับผู้ให้บริการ สั่งอาหารด้วยตนเองผ่านหน้าจอทัชสกรีนสั่งอาหารด้วยตนเองผ่านแอปพลิเคชันในสมาร์ทโฟน เป็นต้น โดยในบทความนี้มีความสนใจในการพัฒนาระบบสั่งอาหารด้วยเสียงพูด ที่มีความสามารถในการโต้ตอบกับผู้ใช้ การทักทาย การสนทนาทั่วไป และการสั่งเมนูอาหาร เพื่อเป็นอีกทางเลือกหนึ่งให้ผู้ใช้มีความสะดวกสบายในการสั่งอาหารมากขึ้น และเป็นการเพิ่มความน่าสนใจให้กับร้านอาหารในการดึงดูดลูกค้า อีกด้วย

ส่วนสำคัญของระบบการโต้ตอบด้วยเสียงพูดในร้านอาหาร คือ การที่ระบบจะต้องทราบให้ได้ว่าผู้ใช้งานที่กำลังพูดอยู่นั้น ต้องการสิ่งใด มีเจตนาที่กำลังจะสื่อสารในประเด็นใด เช่น ทักทาย สอบถามข้อมูล โพรโมชัน ตั้งเมนูอาหาร หรือ เรียกชำระเงิน เป็นต้น ดังนั้น ในบทความนี้จะได้นำเสนอวิธีการจับคู่ข้อความ (Text Matching) เพื่อใช้ในการรับรู้เจตนา (Intent Recognition) ของผู้ใช้งานว่าต้องการสิ่งใด เพื่อประมวลผลให้ได้มาซึ่งการโต้ตอบจากระบบกลับไปเป็นเสียงพูดให้ผู้ใช้งานได้ยิน

## 2. BLEU

BLEU ย่อมาจาก Bilingual Evaluation Understudy [5] คือ ค่าคะแนนที่ใช้ในการวัดความคล้ายคลึงกันของข้อความสองข้อความ ได้แก่ ข้อความอินพุต (Input Text) และข้อความอ้างอิง (Reference Text) ซึ่งค่าคะแนน BLEU นี้มักถูกนำมาใช้ในการคำนวณค่าความถูกต้องของการแปลภาษา (Translation) [6] ใช้ในการวัดประสิทธิผลของวิธีการสร้างประโยคอัตโนมัติ [7] และใช้ในการคำนวณค่าความถูกต้องของการสร้างคำบรรยายภาพจากภาพ (Image Captioning) [8]

ค่าคะแนน BLEU มีค่าอยู่ในช่วง 0 ถึง 1 โดยที่ 1 คือค่าคะแนนที่บ่งบอกว่าข้อความอินพุตกับข้อความอ้างอิงเหมือนกันมากที่สุด (ข้อความอินพุตกับข้อความอ้างอิงเป็นข้อความเดียวกัน) โดยสามารถคำนวณค่าคะแนน BLEU ได้จากสมการที่ (1)

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right) \quad (1)$$

โดยที่

$N$  คือ จำนวนแกรม (gram) ที่ใช้ในการคำนวณค่าคะแนน BLEU โดยทั่วไปจะใช้เป็น 4 ได้แก่ unigram bigram 3-gram และ 4-gram

$p_n$  คือ ค่า Precision ของแต่ละแกรม

$w_n$  คือ ค่าถ่วงน้ำหนักของแต่ละแกรม

$BP$  คำนวณได้จากสมการที่ (2)

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases} \quad (2)$$

โดยที่

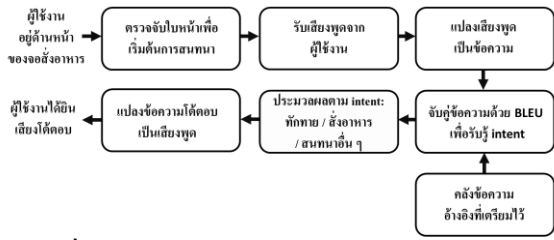
$r$  คือ จำนวนคำในข้อความอ้างอิง

$c$  คือ จำนวนคำในข้อความอินพุต

โดยในบทความนี้ได้ประยุกต์ใช้ค่าคะแนน BLEU ในการจับคู่ข้อความเพื่อการรับรู้เจตนาของผู้ใช้งาน โดยข้อความอินพุตที่ได้จากการแปลงเสียงพูดมาเป็นข้อความ (Speech-to-Text) จะถูกจับคู่กับข้อความอ้างอิงในคลังข้อความ (Text Bank) ที่ได้เตรียมไว้ โดยดูจากค่าคะแนน BLEU ที่มีค่ามากที่สุด ซึ่งข้อความอ้างอิงที่อยู่ในคลังข้อความ เช่น ข้อความทักทาย ข้อความสนทนาทั่วไป และรายชื่อเมนูอาหาร เป็นต้น

## 3. ระบบสั่งอาหารด้วยเสียงพูด

ภาพรวมของระบบสั่งอาหารด้วยเสียงพูดที่นำเสนอในบทความนี้ แสดงดังภาพที่ 1 โดยมีการทำงานเริ่มต้นจากผู้ใช้งานหรือลูกค้า เข้ามานั่งในร้านอาหารแล้วนั่งอยู่ด้านหน้าจอของระบบสั่งอาหารที่อยู่บนโต๊ะอาหาร โดยในบทความนี้ ขอเรียกระบบสั่งอาหารด้วยเสียงพูดนี้ว่า “สมศรี” เมื่อสมศรีตรวจพบใบหน้าของผู้ใช้งานจะมีการทักทายผู้ใช้งานเพื่อเริ่มต้นการสนทนา โดยเมื่อผู้ใช้งานเริ่มพูดกับสมศรี เสียงพูดของผู้ใช้งานจะพูดแปลงเป็นข้อความ (Speech-to-Text) [9] ข้อความอินพุตจากผู้ใช้งาน จะถูกนำไปจับคู่กับข้อความอ้างอิงในคลังข้อความโดยใช้การคำนวณค่าคะแนน BLEU ตามสมการที่ (1) และเมื่อสมศรีทำการจับคู่ข้อความเรียบร้อยแล้ว ก็จะทราบได้ว่าผู้ใช้งานต้องการสิ่งใด เช่น กำลังทักทายกับสมศรี หรือ กำลังสั่งอาหาร เป็นต้น ตัวอย่างในกรณีที่ต้องการสั่งอาหาร เช่น ขอเสิร์ฟหมูพริกไทยดำ 2 จาน สมศรีจะดำเนินการนำรายชื่อเมนูอาหารนี้พร้อมราคาต่อหน่วยและราคารวม แสดงให้ผู้ใช้งานทราบที่หน้าจอ และจะมีข้อความโต้ตอบตามที่ได้กำหนดไว้ ถูกแปลงไปเป็นเสียงพูด (Text-to-Speech) [10] เพื่อโต้ตอบกับผู้ใช้งาน



ภาพที่ 1: ภาพรวมการทำงานของระบบสั่งอาหารด้วยเสียงพูด

ระบบสั่งอาหารด้วยเสียงพูดที่นำเสนอ มีความสามารถดังต่อไปนี้

### 3.1 การตรวจจับใบหน้าผู้ใช้งานเพื่อเริ่มต้นการสนทนา

วิธีการตรวจจับใบหน้าที่ใช้ในบทความนี้ ได้แก่ วิธีการ Haar Cascades [11] โดยเมื่อผู้ใช้งานเริ่มนั่งที่โต๊ะอาหาร สมศรีจะทำการตรวจจับใบหน้าและเริ่มต้นการทักทายผู้ใช้งานด้วยข้อความแบบสุ่ม เช่น ทักทายว่า “สวัสดี สมศรียินดีต้อนรับค่ะ ลูกค้ายกที่ท่านคะ” หรือ “สวัสดีจ้า หนูชื่อสมศรีนะคะ ลูกค้ายกที่ท่านคะ” เป็นต้น

### 3.2 การรับรู้เจตนา

ในบทความนี้ได้แบ่งกลุ่มของเจตนา (Intents) ออกเป็น 2 หมวด ได้แก่ หมวดที่ 1. การทักทายและการสนทนาทั่วไป และ หมวดที่ 2. เมนูอาหารสำหรับการสั่งอาหาร

ในหมวดที่ 1 การทักทายและการสนทนาทั่วไป ประกอบไปด้วย ข้อความทักทาย เรียกเก็บเงิน สอบถามวันที่และเวลา สอบถามส่วนลดและโปรโมชั่น แจ้งจำนวนลูกค้า และสอบถามเมนูแนะนำ ดังแสดงในตารางที่ 1

ตารางที่ 1: ตัวอย่างข้อความอ้างอิงที่ใช้ในการจับคู่กับข้อความอินพุต เพื่อใช้ในการทักทายและการสนทนาทั่วไป

ตัวอย่างเจตนา (Intents)	คลังข้อความอ้างอิงที่ใช้ในการจับคู่	
	จำนวนข้อความ	ตัวอย่างข้อความอ้างอิง
การทักทาย	9	“สวัสดีครับ”, “หวัดดี”, “สวัสดีสมศรี”, ...
โต๊ะสำหรับ 4 คน	29	“มา 1 คนจ้า”, “3 คน”, “มา 4 คนครับ”, ...
ชำระเงิน	22	“เก็บตังด้วยครับ”, “คิดเงินด้วยค่ะ”, ...

ในหมวดที่ 2 เป็นส่วนของการสั่งอาหาร ประกอบด้วยชื่อเมนูอาหารต่างๆ ดังแสดงในตารางที่ 2

ตารางที่ 2: ตัวอย่างข้อความอ้างอิงที่ใช้ในการจับคู่กับข้อความอินพุต เพื่อใช้ในการสั่งอาหาร

ตัวอย่างชื่อเมนูอาหาร (Intents)	คลังข้อความอ้างอิงที่ใช้ในการจับคู่		ราคา (บาท)
	จำนวนข้อความ	ตัวอย่างข้อความอ้างอิง	
สเต็กหมูพริกไทยดำ	59	“สเต็กหมูพริกไทยดำ”, “หมูพริกไทยดำ”, “หมูพริกไทย”, ...	60
สปาเก็ตตี้คาโบนาร่า	54	“สปาเก็ตตี้คาโบนาร่า”, “คาโบนาร่า”, “ซอคาโบนาร่า”, ...	60
ข้าวเหนียวถั่วดำ	31	“ข้าวเหนียวถั่วดำน้ำกะทิ”, “เอาเป็นข้าวเหนียวถั่วดำ”, “รับเป็นข้าวเหนียวถั่วดำ”, ...	30

ในการจับคู่ข้อความโดยใช้ค่าคะแนน BLEU มีขั้นตอนดังนี้ เริ่มต้นโดยข้อความอินพุตและข้อความอ้างอิงจะถูกตัดคำ โดยในบทความนี้ได้ใช้ deepcut [12] ในการทำหน้าที่ตัดคำ ก่อนที่จะส่งต่อไปยังการคำนวณค่าคะแนน BLEU ดังสมการที่ (1) โดยในการคำนวณได้กำหนดให้การถ่วงน้ำหนักเพื่อพิจารณา จากคำเดียว และ สองคำติดกัน ที่เท่ากัน ส่วนในกรณีของสามคำติดกัน และสี่คำติดกันไม่ต้องนำมาพิจารณา จึงได้มีการกำหนดค่าถ่วงน้ำหนักไว้ ได้แก่  $w_1 = 0.5, w_2 = 0.5, w_3 = 0, w_4 = 0$  ตัวอย่างการคำนวณค่าคะแนน BLEU แสดงในตารางที่ 3

ตารางที่ 3: ตัวอย่างค่าคะแนน BLEU ที่คำนวณได้จากกรณีจับคู่ข้อความ

ตัวอย่างข้อความอินพุต	ค่าคะแนน BLEU ที่คำนวณได้
“ขอสั่งสเต็กหมูพริกไทยดำ 2 จาน” เทียบกับ	
ข้อความอ้างอิงที่ 1: “อยากกินสเต็กหมูเกาหลี 2 จาน”	0.3303
ข้อความอ้างอิงที่ 2: “เอาผักกาด”	0.0000
ข้อความอ้างอิงที่ 3: “ขอบัวลอย 3 ถ้วย”	0.0000
ข้อความอ้างอิงที่ 4: “เอาสเต็กหมูพริกไทยดำ 3 จาน”	0.4671
ข้อความอ้างอิงที่ 5: “ขอเป็นน้ำมะพร้าว 2 แก้ว”	0.0000

จากตารางที่ 3 จะเห็นได้ว่า ค่าคะแนน BLEU ที่มากที่สุด คือ 0.4671 ซึ่งเป็นค่าคะแนน BLEU ระหว่างข้อความอินพุตเทียบกับข้อความอ้างอิงที่ 4 ซึ่งแสดงถึงการที่ข้อความอินพุตมีค่าใกล้เคียงกับข้อความอ้างอิงที่ 4 มากที่สุดเมื่อเทียบกับข้อความอ้างอิงอื่น ๆ ดังนั้น ผลการจับคู่

ที่ได้จึงได้เจตนา (Intents) ที่ต้องการสั่งเป็น สเต็กหมู พริกไทยดำ โดยจำนวนจานที่ต้องการสั่ง จะสกัดตัวเลขมาจากข้อความอินพุต นั่นคือ 2 จาน

**3.3 การระบุจำนวนของเมนูอาหารที่ต้องการ**

ผู้ใช้งานสามารถสั่งอาหารพร้อมระบุจำนวนจานที่ต้องการสั่งได้ เช่น “ขอสั่งสปาเก็ตตี้คาโบนาร่า 3 จาน” เป็นต้น โดยสมศรีจะทำการค้นหาตัวเลขในข้อความ เพื่อนำมาใช้ในการคำนวณราคาอาหารที่สั่งให้ผู้ใช้งานเห็นที่หน้าจอ แต่ในกรณีที่ไม่มีตัวเลขอยู่ในข้อความ หรือผู้ใช้งานไม่ได้ระบุจำนวนจาน สมศรีจะกำหนดให้เป็น 1 จาน

**3.4 การยกเลิกเมนูอาหารที่ได้สั่งไปแล้ว**

ผู้ใช้งานสามารถสั่งยกเลิกเมนูอาหารที่สั่งไปแล้วได้ โดยสมศรีจะดำเนินการตัดรายการเมนูอาหารนั้นออก พร้อมคำนวณราคารวมใหม่ให้ผู้ใช้งานทราบ แต่หากผู้ใช้งานพุดยกเลิกเมนูอาหารที่ไม่มีในรายการอาหารที่สั่งไปแล้ว สมศรีจะมีการแจ้งเตือนว่าไม่พบเมนูดังกล่าวให้ผู้ใช้งานทราบ

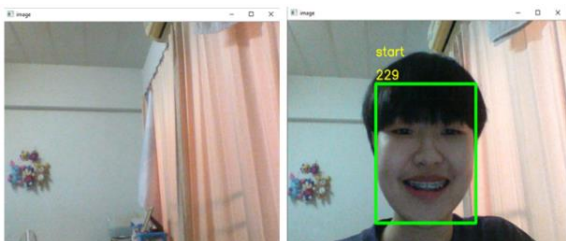
**3.5 การจัดเก็บคะแนนความพึงพอใจและข้อเสนอแนะ**

เมื่อผู้ใช้งานเรียกชำระเงิน สมศรีจะสอบถามคะแนนความพึงพอใจและข้อเสนอแนะจากผู้ใช้งาน โดยผู้ใช้งานสามารถพุดเลขคะแนนที่ต้องการให้ และพุดข้อเสนอแนะคำติชมต่าง ๆ ได้ สมศรีจะดำเนินการเก็บค่าคะแนนและข้อเสนอแนะเป็นข้อความลงไฟล์ข้อมูลต่อไป

**4. การทดลอง**

**4.1 การตรวจจับใบหน้าเพื่อเริ่มต้นการสนทนา**

เมื่อผู้ใช้งานนั่งอยู่หน้าจอเพื่อสั่งอาหาร สมศรีจะทำการตรวจจับใบหน้า แสดงตัวอย่างดังภาพที่ 2 และจะเริ่มต้นทักทายกับผู้ใช้งาน



**ภาพที่ 2:** การตรวจจับใบหน้าเพื่อเริ่มต้นการสนทนา (ซ้าย) ไม่พบใบหน้า (ขวา) พบใบหน้า

**4.2 การรับรู้เจตนาในหมวดของการทักทายและการสนทนาทั่วไป**

ในการทดลองนี้ เป็นการที่ผู้ใช้งานพุดข้อความต่าง ๆ หลากหลายรูปแบบเพื่อทดสอบการรับรู้เจตนา (Intents) ในหมวดของการทักทายและการสนทนาทั่วไป โดยผลที่ได้จะเป็นจำนวนครั้งของความถูกต้องในการจับคู่ข้อความที่ผู้ใช้งานพุด เทียบกับข้อความอ้างอิงในคลังข้อความ โดยมีผลการทดลองดังแสดงในตารางที่ 4 ซึ่งได้ค่าความถูกต้องเฉลี่ยคิดเป็น 92.50%

**ตารางที่ 4:** ผลการทดลองการรับรู้เจตนา (Intents) ในหมวดของการทักทายและการสนทนาทั่วไป

เจตนา (Intents)	จำนวนครั้งของข้อความอินพุตจากผู้ใช้งานที่มีการพุดบริบทของข้อความที่แตกต่างกัน	ความถูกต้อง (ครั้ง)	ร้อยละความถูกต้อง
การทักทาย	10	8	80%
โต๊ะสำหรับ1_4คน	10	10	100%
โต๊ะสำหรับ5_8คน	10	10	100%
ชำระเงิน	10	9	90%
วันที่	10	9	90%
เวลา	10	10	100%
โทรโมชั่น	10	8	80%
เมนูแนะนำ	10	10	100%
ค่าเฉลี่ยความถูกต้อง	92.50%		

**4.3 การรับรู้เจตนาในหมวดของการสั่งอาหาร**

ในการทดลองนี้ เป็นการที่ผู้ใช้งานพุดข้อความต่าง ๆ หลากหลายรูปแบบเพื่อทดสอบการรับรู้เจตนา (Intents) ในหมวดของการสั่งอาหาร โดยมีผลการทดลองดังแสดงในตารางที่ 5 ซึ่งได้ค่าความถูกต้องเฉลี่ยคิดเป็น 81.90% โดยเมนูอาหารที่มีความผิดพลาดมากที่สุด ได้แก่ เฟรนฟราย ซึ่งมีสาเหตุมาจากการออกเสียงที่ไม่ชัดเจนหรือคลุมเครือ ส่งผลให้ขั้นตอนในการแปลงเสียงพุดเป็นข้อความนั้น แปลงข้อความออกมาเป็นคำอื่น ซึ่งส่งผลกระทบต่อค่าคะแนน BLEU ที่คำนวณได้ จึงไม่สามารถจับคู่ได้ตรงตามความเป็นจริง

ตารางที่ 5: ผลการทดลองการรับรู้เจตนา (Intents) ในหมวดของการสั่งอาหาร

เมนูอาหารเจตนา (Intents)	จำนวนครั้งของข้อความอินพุตจากผู้ใช้งานที่มีการพูดบริบทของข้อความที่แตกต่างกัน	ความถูกต้อง (ครั้ง)	ร้อยละความถูกต้อง
สเต็กหมูพริกไทยดำ	10	9	90%
สเต็กหมูเกาหลี	10	10	100%
สเต็กแฮม	10	9	90%
สเต็กเนื้อ	10	9	90%
สเต็กปลาแซลมอน	10	8	80%
สปาเก็ตตี้คาโบนาร่า	10	6	60%
สปาเก็ตตี้เบคอนพริกแห้ง	10	8	80%
แฮมเบอร์เกอร์หมู	10	8	80%
แฮมเบอร์เกอร์ไก่	10	9	90%
นักร้อง	10	8	80%
เฟรนฟราย	10	4	40%
ชีสบอล	10	8	80%
มันบด	10	9	90%
ขนมปังกระเทียม	10	10	100%
ลาซานญ่าหมู	10	8	80%
กุ้งโตน้ำ	10	6	60%
สาเกุแคนคาจูป	10	9	90%
บัวลอย	10	10	100%
ข้าวเหนียวถั่วดำ	10	8	80%
น้ำส้ม	10	9	90%
น้ำมะพร้าว	10	7	70%
ค่าเฉลี่ยความถูกต้อง		81.90%	

#### 4.4 การสั่งอาหารและการยกเลิกเมนูอาหารที่สั่งไปแล้ว

ในการทดลองนี้ผู้ใช้งานสามารถเพิ่ม ลด จำนวนเมนูอาหารที่สั่งได้ โดยในการเพิ่มเมนูอาหาร ผู้ใช้งานสามารถสั่งด้วยเสียงพูดเพื่อแปลงไปเป็นข้อความอินพุตที่ถูกนำไปเปรียบเทียบกับข้อความอ้างอิง เพื่อทำการเพิ่มเมนูอาหารตามที่ตั้งที่หน้าจอ เช่น “เอาสเต็กหมูพริกไทยดำ 3 จาน” “เอาเป็นสปาเก็ตตี้คาโบนาร่า 2 จาน” และ “ขอเป็นแฮมเบอร์เกอร์ไก่” เป็นต้น โดยในกรณีนี้ถ้าผู้ใช้งานสั่งอาหารโดยมีการระบุจำนวนจาน สมศรีจะมีการบันทึกจำนวนจานอาหารตามที่ผู้ใช้งานสั่งแล้วแสดงผลที่หน้าจอดังแสดงในภาพที่ 3 แต่ถ้าผู้ใช้งานไม่ได้ระบุจำนวนจาน สมศรีจะกำหนดให้เป็น 1 จาน นอกจากนี้ จะมีการอัปเดตราคาต่อจานของแต่ละเมนูอาหาร ราคา รวมของแต่ละเมนูอาหาร และราคา รวมทั้งหมด ให้ผู้ใช้งานทราบ

รายการอาหาร	จำนวน	ราคา	จำนวนเงิน
สเต็กหมูพริกไทยดำ	3	60	180
สปาเก็ตตี้คาโบนาร่า	2	60	120
แฮมเบอร์เกอร์ไก่	1	50	50
บัวลอย	2	30	60
สเต็กแฮม	2	60	120
น้ำส้ม	2	30	60
ขนมปังกระเทียม	1	50	50
ลาซานญ่าหมู	1	60	60
มันบด	3	40	120
ราคารวม			820 บาท

ภาพที่ 3: ตัวอย่างหน้าจอแสดงรายการอาหารที่สั่ง

ในส่วนของการยกเลิกเมนูอาหารที่สั่งไปแล้วนั้น ผู้ใช้งานสามารถทำได้โดยพูดคำว่า “ขอยกเลิก” แล้วตามด้วยชื่อเมนูอาหารที่ต้องการยกเลิก เช่น “ขอยกเลิกแฮมเบอร์เกอร์ไก่” สมศรีจะทำการยกเลิกเมนูอาหารนี้ให้ดังแสดงในภาพที่ 4

รายการอาหาร	จำนวน	ราคา	จำนวนเงิน
สเต็กหมูพริกไทยดำ	3	60	180
สปาเก็ตตี้คาโบนาร่า	2	60	120
บัวลอย	2	30	60
สเต็กแฮม	2	60	120
น้ำส้ม	2	30	60
ลาซานญ่าหมู	1	60	60
ราคารวม			600 บาท

ภาพที่ 4: การยกเลิกเมนูอาหารบางรายการ

นอกจากนี้ ผู้ใช้งานยังสามารถสั่งเมนูอาหารอื่น ๆ เพิ่มเติม หลังจากทำการยกเลิกเมนูอาหารบางรายการไปแล้วได้ ส่วนในกรณีที่ผู้ใช้งานพูดคำว่า “ขอยกเลิก” แล้วตามด้วยเมนูอาหารที่ไม่มีในรายการที่ได้สั่งไปแล้ว สมศรีจะพูดคำว่า “ไม่พบรายการที่คุณขอยกเลิก กรุณาพูดอีกครั้งค่ะ” เนื่องจากไม่พบเมนูอาหารนั้น

#### 4.5 การรับคะแนนความพึงพอใจและข้อเสนอแนะจากผู้ใช้งาน

เมื่อผู้ใช้งานรับประทานอาหารเสร็จแล้ว และมีการเรียกชำระเงิน สมศรีจะมีการถามถึงคะแนนความพึงพอใจในครั้งนี้ ซึ่งผู้ใช้งานสามารถให้คะแนนได้ตั้งแต่ 1 คะแนนถึง 5 คะแนน จากนั้น จะมีการรับข้อเสนอแนะจากผู้ใช้งานดังแสดงในภาพที่ 5 เพื่อทำการจัดเก็บคะแนนและข้อเสนอแนะที่ได้ไว้ในไฟล์ข้อมูล ดังแสดงในภาพที่ 6 โดยมีการจัดเก็บข้อมูล ได้แก่ ข้อเสนอแนะ คะแนน วันที่ และ เวลา เพื่อนำไปใช้ในการปรับปรุงการให้บริการต่อไป



<---ฉันกำลังฟังคุณอยู่--->  
 คุณพูดว่า --> เช็คบิลด้วยคะ  
 หมวดคำถาม : ชำระเงิน  
 ค่าอาหารทั้งหมด930บาทคะ  
 ขอทราบความพึงพอใจของลูกค้าสำหรับอาหารมือนี่ด้วยคะ  
 ลูกค้าสามารถให้คะแนนได้ตั้งแต่ 1คะแนน ถึง 5คะแนนคะ

<---ฉันกำลังฟังคะแนนจากคุณอยู่--->  
 คุณพูดว่า --> 5 คะแนนคะ  
 รบกวนลูกค้าคอมเมนต์ให้กับทางร้านด้วยคะ

<---ฉันกำลังฟังคอมเมนต์คุณอยู่--->  
 คุณพูดว่า --> กุ้งโดนัทกรอบมากคะ  
 ขอขอบคุณสำหรับความคิดเห็นคะ

ภาพที่ 5: ตัวอย่างการรับคะแนนความพึงพอใจและข้อเสนอแนะจากผู้ใช้งานเมื่อเรียกชำระเงิน

A	B	C	D
comment	score	date	time
บิวลอยกลมกล่อมมากคะ	5	23/02/2022	17:26:54
รสชาติเฉยๆคะ	3	23/02/2022	17:32:27
สเด็กหมูปริกไทยตำร่อยมากคะ	5	23/02/2022	22:34:35
น้ำส้มร่อยคะ	5	23/02/2022	22:35:25
กุ้งโดนัทกรอบมากคะ	5	23/02/2022	22:46:25

ภาพที่ 6: ตัวอย่างการจัดเก็บคะแนนความพึงพอใจและข้อเสนอแนะจากผู้ใช้งานลงในไฟล์ข้อมูล

### 5. บทสรุป

บทความนี้ประยุกต์ใช้ค่าคะแนน BLEU มาทำหน้าที่เป็นตัวชี้วัดความคล้ายคลึงกันของข้อความ เพื่อใช้ในการจับคู่ข้อความอินพุตที่แปลงจากเสียงพูดของผู้ใช้งานมาเป็นข้อความ เทียบกับข้อความในคลังข้อความอ้างอิง ซึ่งเป็นส่วนสำคัญในการนำมาสร้างเป็นระบบสั่งอาหารด้วยเสียงพูด โดยในบทความนี้ได้พัฒนาให้ระบบสั่งอาหารด้วยเสียงพูดมีความสามารถในการทักทาย สนทนาทั่วไป สั่งอาหาร ยกเลิกเมนูอาหาร เรียกชำระเงิน ให้คะแนนความพึงพอใจและข้อเสนอแนะกับร้านอาหาร และจากการใช้ค่าคะแนน BLEU ในการจับคู่ข้อความ ซึ่งมีการคำนวณที่ไม่ซับซ้อน จึงทำให้ระบบการสั่งอาหารด้วยเสียงพูดที่พัฒนาขึ้นมานี้ สามารถประมวลผลต่อคำสั่งเสียงพูดและได้ตอบกับผู้ใช้งานได้อย่างราบรื่น

### เอกสารอ้างอิง

[1] Pitiphat Pinbut, Kairat Jaroenrat, Kittipong Pooputwibul and Natchamol Srichumroenrattana, "The application of voice-controlled smart home lighting for the elderly," *Journal of Management Science Nakhon Pathom Rajabhat University*, Vol.7 No.1, pp. 306-319, January-June 2020.

[2] Tanin Muangpool and Ouychai Intharasombat, "Voiced-based Controlled Electrical Devices using Android and Microcontroller," *International Computer Science and Engineering Conference (ICSEC)*, pp. 20-30, December 2016.

[3] Lin Wei, Liqing He and Yelu Liu, "Study of Artificial Intelligence Flight Co-Pilot Speech Recognition Technology," *IEEE 2<sup>nd</sup> International Conference Civil Aviation Safety and Information Technology (ICCASIT)*, pp. 681-685, 2020.

[4] Abhinav Sharma, Anshu Sharma, PK Juneja and Vrinda Jain, "Spectral Features based Speech Recognition for Speech Interfacing to Control PC Windows," *IEEE 2<sup>nd</sup> International Conference Advances in Computing, Communication & Materials (ICACCM)*, pp. 341-345, 2020.

[5] Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation," *Proceeding of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, July 2002, pp. 311-318.

[6] K. Nemkul and S. Shakya, "English to Nepali Sentence Translation Using Recurrent Neural Network with Attention," *International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2021, pp. 607-611.

[7] D. Russell, L. Li and F. Tian, "Generating Text using Generative Adversarial Networks and Quick-Thought Vectors," *IEEE 2<sup>nd</sup> International Conference on Computer and Communication Engineering Technology (CCET)*, 2019, pp. 129-133.

[8] Zhou, Congcong, Lei, Zhou, Chen, Shengbo, Huang, Yiyong and Liu, Xianrui, "A Sparse Transformer-Based Approach for Image Captioning," *IEEE Access*, 2020. pp. 1-10.

[9] Speech Recognition Python library. Available online at <https://pypi.org/project/SpeechRecognition>

[10] gTTS (Google Text-to-Speech) Python library. Available online at <https://pypi.org/project/gTTS>

[11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[12] deepcut Python library. Available online at <https://pypi.org/project/deepcut/>

# การพัฒนาโมเดลการพยากรณ์การอนุมัติสินเชื่อธุรกิจด้วยเทคนิคเหมืองข้อมูล

## Development of Model for Forecasting Business Credit Approvals Using Data Mining Techniques

กาศิญา กลิ่นถาวร (Karsiya Klintharvorn)<sup>1</sup> และมมเกียรติ รัตน์ศิริวงศ์วุฒิ (Montean Rattanasiriwongwut)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ<sup>1,2</sup>

S6307021858015@email.kmutnb.ac.th<sup>1</sup>, montean.r@itd.kmutnb.ac.th<sup>2</sup>

### บทคัดย่อ

การวิจัยครั้งนี้มีจุดประสงค์เพื่อทำการพยากรณ์อนุมัติสินเชื่อ และทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึมเพื่อนำมาใช้ในการสนับสนุนการตัดสินใจการพยากรณ์อนุมัติสินเชื่อธุรกิจด้วยเทคนิคเหมืองข้อมูล นำข้อมูลสินเชื่อธุรกิจจำนวน 3,000 รายการในอดีตจากสถาบันการเงินแห่งหนึ่งมาทำการวิเคราะห์ข้อมูล โดยเปรียบเทียบระหว่างอัลกอริทึม 5 ตัว ได้แก่ ต้นไม้การตัดสินใจ, ป่าสุ่ม, นาอิวเบย์, โครงข่ายประสาทเทียม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน จากนั้นนำข้อมูลมาแสดงผลรายงานในรูปแบบ Power BI โดยผลการเปรียบเทียบพบว่าอัลกอริทึม Random Forest มีประสิทธิภาพสูงสุด มีค่าความถูกต้องร้อยละ 98.03 และค่าความแม่นยำร้อยละ 98.26 จากนั้นนำข้อมูลและอัลกอริทึมที่ได้มาแสดงผลรายงานในรูปแบบ Dash Board ผ่านโปรแกรม Power BI

**คำสำคัญ:** การพยากรณ์, สินเชื่อธุรกิจ, เทคนิคเหมืองข้อมูล

### Abstract

The purpose of this research was to Forecasting Business Credit Approvals. And to compare the performance of the algorithms for use to support business credit approval forecasting decision making used 3,000 historical business credit data from a financial institution to analyze. There 5 algorithms used for analysis and forecasting: Decision Tree, Random Forest, Naïve Bayes, Artificial Neural Network and Support Vector

Machine. The source of the report in Power BI Desktop format. In this performance test, it was found that the most efficient algorithm was Random Forest Algorithm with 98.03% Accuracy and 98.26% Precision

**Keywords:** Forecasting, Business loan, Data Mining.

### 1. บทนำ

จากสถิติสถาบันการเงินมีรายงานรายได้ค่าใช้จ่ายของธนาคารพาณิชย์ทั้งระบบ โดยรายได้จากดอกเบี้ยเงินให้สินเชื่อมีปริมาณสูงสุดของรายได้ดอกเบี้ยเมื่อสถาบันการเงินมีความต้องการเพิ่มปริมาณสินเชื่อสิ่งที่จะต้องคำนึงถึงเป็นสิ่งสำคัญคือการประเมินและวิเคราะห์ความเสี่ยงของการให้สินเชื่อเพื่อป้องกันไม่ให้เกิดสินเชื่อที่ไม่ก่อให้เกิดรายได้ (Non-performing Loan) ในปริมาณที่สูงซึ่งในปี 2563 สินเชื่อขยายตัวเพิ่มขึ้น 5.1% จากปีก่อนอยู่ที่ 2% โดยส่วนใหญ่เติบโตมาจากสินเชื่อธุรกิจขยายตัวที่ 5.4% เทียบกับปีก่อนที่หดตัว-0.8 (สินเชื่อธุรกิจคิดเป็นร้อยละ 64.2 ของสินเชื่อรวม) ปัจจัยหลักจากการเติบโตของสินเชื่อธุรกิจขนาดใหญ่ที่ส่วนหนึ่งกลับมาใช้สินเชื่อแทนการออกตราสารหนี้ในช่วงไตรมาสที่ 2 ของปี 2563 ขณะที่สินเชื่อธุรกิจขนาดกลางและขนาดย่อม (SMEs) หดตัว-2.8% ถือเป็นอัตราที่ลดลงจากผลของมาตรการสินเชื่อดอกเบี้ยต่ำ (soft loan)

ในสถานการณ์โรคติดเชื้อไวรัสโคโรนา (โควิด-19) นโยบายของภาครัฐที่ให้สถาบันการเงินปล่อยสินเชื่อมากขึ้น กระตุ้นเศรษฐกิจและเพื่อเป็นการช่วยเหลือ

ผู้ประกอบการที่ได้รับผลกระทบแต่จากสถานการณ์ปัจจุบันสถาบันการเงินยังคงต้องระมัดระวังการปล่อยสินเชื่อให้ SMEs จากความเสี่ยงของธุรกิจและภาวะเศรษฐกิจที่ยังไม่แน่นอนสูงและมาตรการช่วยเหลือเพิ่มเติมของภาครัฐอาทิมาตรการสนับสนุนการให้สินเชื่อแก่ผู้ประกอบการตาม พ.ร.ก. ฟื้นฟูจะช่วยให้มาตรฐานการให้สินเชื่อสำหรับธุรกิจขนาดใหญ่และ SMEs มีแนวโน้มผ่อนคลายลงเล็กน้อยแม้สถาบันการเงินยังมีแนวโน้มเพิ่มความระมัดระวังการปล่อยสินเชื่อให้ SMEs สำหรับกลุ่มลูกค้าใหม่และธุรกิจที่ยังไม่ฟื้นตัวจากการระบาดโดยสถาบันมีแนวโน้มปรับเพิ่ม Margin และเพิ่มเงื่อนไขประกอบสัญญาผู้รวมถึงเพิ่มเงื่อนไขหลักประกัน โดยเฉพาะกลุ่มลูกค้าเสี่ยง

จากภาวะเศรษฐกิจสถานการณ์ในปัจจุบันและแนวโน้มความต้องการสินเชื่อของธุรกิจในขณะนี้ทำให้สถาบันการเงินมีความกังวลเกี่ยวกับสินเชื่อที่ไม่ก่อให้เกิดรายได้ (Non-performing Loan) ซึ่งในขณะเดียวกันการแข่งขันของสถาบันการเงินในการปล่อยสินเชื่อก็มีการแข่งขันที่สูงและจากปัญหาดังกล่าวผู้วิจัยจึงมีแนวความคิดที่จะพัฒนาแบบจำลองสำหรับการพยากรณ์การอนุมัติสินเชื่อธุรกิจด้วยเทคนิคเหมืองข้อมูลให้เกิดความแม่นยำเกิดมาตรฐานในการวิเคราะห์สินเชื่อและจากการศึกษางานวิจัยที่เกี่ยวข้องมีอัลกอริทึมที่น่าสนใจ ดังนี้ Decision Tree, Random Forest, Naïve Bayes, Artificial Neural Network และ Support Vector Machine เพื่อเป็นการประเมินเบื้องต้นในการประเมินความเสี่ยงในการอนุมัติสินเชื่อธุรกิจและลดความเสี่ยงสินเชื่อที่ไม่ก่อให้เกิดรายได้ (Non-performing Loan) และในอนาคตจากการให้สินเชื่อแก่ลูกค้าแต่ละรายโดยการพัฒนาแบบจำลองสำหรับการพยากรณ์การอนุมัติสินเชื่อธุรกิจด้วยเทคนิคเหมืองข้อมูลกับการอนุมัติสินเชื่อธุรกิจในอดีต

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 เกณฑ์ในการประเมินคุณภาพผู้ขอสินเชื่อ 5C's

เกณฑ์ในการประเมินคุณภาพผู้ขอสินเชื่อ 5C's เป็น เป็นวิธีหนึ่งที่นิยมใช้ โดยทั่วไปจะช่วยในการกำหนดความสามารถในการชำระหนี้ของนิติบุคคลหรือความสามารถในการชำระหนี้ของบุคคลธรรมดาซึ่งประกอบไปด้วย Character, Capacity, Capital, Collateral และ Conditions

### 2.2 เทคนิคเหมืองข้อมูล (Data Mining)

เทคนิคเหมืองข้อมูล (Data Mining) คือ กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหา รูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้นในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภททั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหารในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคมโดยมีตอนการทำเหมืองข้อมูลวิธี Cross-Industry Standard Process for Data Mining (CRISP-DM) ประกอบด้วย 6 ขั้นตอนคือ

2.1.1 Business Understanding การทำความเข้าใจในงาน

2.1.2 Data Understanding ทำความเข้าใจข้อมูล

2.1.3 Data Preparation ทำการแปลงข้อมูล (Raw Data) ให้กลายเป็นข้อมูลที่สามารถนำมาช่วยในการวิเคราะห์ต่อไปได้

2.1.4 Modeling การสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูล

2.1.5 Evaluation การประเมินผลลัพธ์

2.1.6 Deployment การนำเอาข้อมูลที่เป็นผลลัพธ์

### 2.2 ต้นไม้การตัดสินใจ (Decision Tree)

เป็นการเรียนรู้โดยการจำแนกประเภท (Classification) ข้อมูลออกเป็นกลุ่ม (class) ต่างๆ โดยใช้คุณลักษณะ (attribute) ข้อมูลในการจำแนกประเภท ต้นไม้ตัดสินใจที่ได้จากการเรียนรู้ทำให้ทราบว่าคุณลักษณะใดเป็นตัวกำหนดการจำแนกประเภท และคุณลักษณะแต่ละตัวมีความสำคัญมากน้อยต่างกันอย่างไร

$$E(S) = - \sum_{j=1}^n p_S(j) \log_2 p_S(j)$$

S คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้น และตัวแปรตามหลายๆกรณี

$P_S(j)$  คืออัตราส่วนของกรณีใน S ที่ตัวแปรตาม หรือผลลัพธ์มีค่า j

### 2.3 ป่าสุ่ม (Random Forest)

เป็นหนึ่งในกลุ่มของ โมเดลที่เรียกว่า Ensemble learning ที่มีหลักการคือการเทรนโมเดลที่เหมือนกันหลายๆครั้ง (หลาย Instance) บนข้อมูลชุดเดียวกันโดยแต่ละครั้งของการเทรนจะเลือกส่วนของข้อมูลที่เทรนไม่เหมือนกันแล้วเอาการตัดสินใจของโมเดลเหล่านั้นมาโหวตกันว่า Class ไหนถูกเลือกมากที่สุด

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2$$

### 2.4 นาอิวเบย์ (Naïve Bayes)

การทำเหมืองข้อมูลในแบบ classifier ที่ถูกสร้างขึ้น โดยหลักความน่าจะเป็น Naïve Bayesian Classification จะใช้วิเคราะห์หาความน่าจะเป็นของสิ่งที่ยังไม่เคยเกิดขึ้น โดยการคาดเดาจากสิ่งที่เคยเกิดขึ้นมาก่อน

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability  
Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

### 2.5 โครงข่ายประสาทเทียม (Artificial Neural Network)

ระบบคอมพิวเตอร์จากโมเดลทางคณิตศาสตร์เพื่อจำลองการทำงาน โครงข่ายประสาทชีวภาพที่อยู่ในสมองของสัตว์ โครงข่ายประสาทเทียมสามารถเรียนรู้ที่จะทำงานที่มอบหมายได้จากการเรียนรู้ผ่านตัวอย่างโดยไม่ถูกโปรแกรมด้วยกฎเกณฑ์ตายตัวแบบระบบอัตโนมัติ

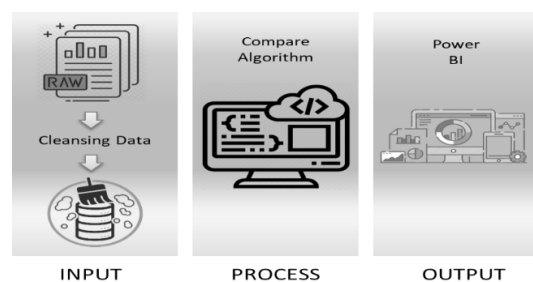
### 2.6 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

อัลกอริทึมที่สามารถนำมาช่วยแก้ปัญหาการจำแนกข้อมูลใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูลโดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้โดยเน้นไปยังเส้นแบ่งแยกแยกกลุ่มข้อมูลได้ดีที่สุด

$$h_{\theta}(x) = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$$

$$= w^T x + b$$

### 3. วิธีดำเนินการวิจัย



ภาพที่ 1 : กรอบแนวคิดการพัฒนาแบบจำลองการพยากรณ์อนุมัติสินเชื่อธุรกิจด้วยเทคนิคเหมืองข้อมูล

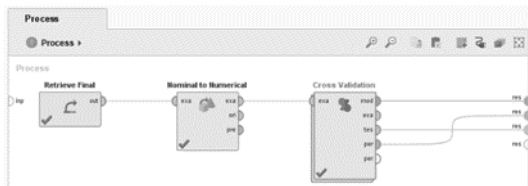
จากภาพที่ 1 การวิจัยครั้งนี้เป็นการวิจัยแบบประยุกต์โดยใช้เทคนิค Decision Tree, Random Forest, Naïve Bayes, Artificial Neural Network และ Support Vector Machine จากนั้นดำเนินการเปรียบเทียบเพื่อเลือกอัลกอริทึมที่เหมาะสมในการเรียนรู้ข้อมูลในอดีตโดยใช้เทคนิคการจำแนกประเภทข้อมูล (Data Classification) และใช้โปรแกรม Rabbit Miner Studio ในการทดลองสำหรับพัฒนาโมเดลดังกล่าวมีขั้นตอนการวิจัยดังนี้

3.1 ศึกษากระบวนการ CSISP-DM ซึ่งเป็นกระบวนการในการวิเคราะห์ข้อมูลด้านเหมืองข้อมูล

3.2 รวบรวมข้อมูลที่ต้องการนำมาใช้ ซึ่งข้อมูลที่นำมาใช้ในงานวิจัยครั้งนี้เป็นข้อมูลการอนุมัติสินเชื่อธุรกิจจากสถาบันการเงินแห่งหนึ่ง ช่วงระหว่าง ตั้งแต่ ค.ศ. 2561 ถึง ค.ศ. 2563 จำนวน 3,000 รายการ ข้อมูลที่จากนั้นคัดเลือกข้อมูลที่ใช้ในการพัฒนาโมเดลการ

พยากรณ์ให้สอดคล้องกับหลักทฤษฎีเกณฑ์ในการประเมินคุณภาพผู้ขอสินเชื่อ 5C's

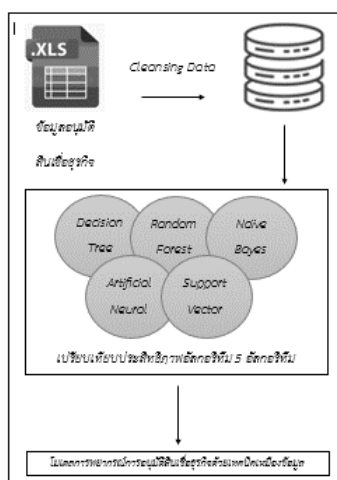
3.3.1 คำเนินการเปรียบเทียบประสิทธิภาพแบบจำลองที่มีประสิทธิภาพสูงสุดด้วยโปรแกรม Rapid Miner Studio โดยขั้นตอนประกอบด้วย 1) นำเข้าไฟล์ข้อมูลการอนุมัติสินเชื่อธุรกิจที่มีการทำความสะอาดข้อมูลเรียบร้อยแล้ว 2) ใช้เครื่องมือ Cross Validation เพื่อประเมินประสิทธิภาพของแบบจำลองทั้ง 5 เทคนิคในการเปรียบเทียบดังนี้ Decision Tree, Random Forest, Naïve Bayes, Artificial Neural Network และ Support Vector Machine กระบวนการภายในเครื่องมือ Cross Validation มีการแบ่งการทำงานออกเป็น 2 ส่วน คือ Training และ Testing ในส่วนของการ Training นำเครื่องมือ Neural Network มาใช้เพื่อเป็นการเรียนรู้ ในส่วนของ Testing นำเครื่องมือ Apply Model และใช้เครื่องมือ Performance (Classification) เพื่อแสดงประสิทธิภาพของแบบจำลอง หลังจากนั้นทำการเลือกเทคนิคที่ถูกต้องและแม่นยำที่สุด



ภาพที่ 2: เปรียบเทียบอัลกอริทึม (1)



ภาพที่ 3: เปรียบเทียบอัลกอริทึม (2)



ภาพที่ 4: การเปรียบเทียบประสิทธิภาพแบบจำลอง

3.4 ขั้นตอนการพัฒนาหน้าจอการแสดงผล เป็นการนำภาพรวมของข้อมูลที่ได้จากการสร้างโมเดลมาแสดงผลรายงานในรูปแบบ Dashboard ผ่านโปรแกรม Power BI

#### 4. ผลการวิจัย

##### 4.1 ผลเปรียบเทียบอัลกอริทึม

เปรียบเทียบอัลกอริทึมโดยนำชุดข้อมูลตัวอย่างมาเพื่อทดสอบประสิทธิภาพของอัลกอริทึมตัวอย่างชุดข้อมูลแสดงดังตารางที่ 1

ตารางที่ 1 : ผลการเปรียบเทียบค่าความถูกต้อง และ แม่นยำ

Algorithms	Accuracy	Precision
Decision Tree	97.77%	98.04%
Random Forest	98.03%	98.26%
Naïve Bayes	64.03%	60.08%
Artificial Neural Network	76.67%	71.24%
Support Vector Machine	80.03%	80.17%

จากตารางที่ 1 สามารถสรุปผลได้ดังนี้

แบบจำลองต้นไม้ตัดสินใจ (Decision Tree) มีความถูกต้องเท่ากับ 97.77% และค่าความแม่นยำ เท่ากับ 98.04%

แบบจำลองแรนคอมฟอเรสต์ (Random forest) มีความถูกต้องเท่ากับ 98.03% และค่าความแม่นยำ เท่ากับ 98.26%

แบบจำลองนาอิวเบย์ (Naïve Bayes) มีความถูกต้องเท่ากับ 64.03% และค่าความแม่นยำเท่ากับ 60.08%

แบบจำลองโครงข่ายประสาทเทียม (Artificial Neural Network) มีความถูกต้องเท่ากับ 76.67% และค่าความแม่นยำเท่ากับ 71.24%

แบบจำลองซัพพอร์ตเวกเตอร์แมทชีน (Support Vector Machine:SVM) มีความถูกต้องเท่ากับ 80.03% และค่าความแม่นยำเท่ากับ 80.17%

ในการเปรียบเทียบประสิทธิภาพแบบจำลองพบว่าเมื่อนำข้อปัจจัยต่างๆมาสร้าง โมเดลและปรับค่าพารามิเตอร์ในการวัดประสิทธิภาพด้วย แบบจำลอง แรนดอมฟอเรสต์ (Random forest) สามารถให้ประสิทธิภาพสูงที่สุด โดยอธิบายได้จากจำนวน 3,000 รายการ ที่มีการแบ่งข้อมูลสำหรับการเรียนรู้และข้อมูลสำหรับการทดสอบ จำนวนข้อมูลที่พยากรณ์ทำนายถูกว่าได้รับสินเชื่อ 2,153 รายการ จำนวนข้อมูลที่พยากรณ์ทำนายถูกว่าไม่ได้รับสินเชื่อ 788 รายการ จำนวนข้อมูลที่พยากรณ์ทำนายคิดว่าได้รับสินเชื่อ 49 รายการ และจำนวนข้อมูลที่พยากรณ์ทำนายคิดว่าไม่ได้รับสินเชื่อ 10 รายการ

4.2 หน้าแสดงผลรายงาน



ภาพที่ 5 : หน้าแสดงรายงานภาพรวม

จากภาพที่ 5 เป็นตัวอย่างการแสดงผลภาพรวมผลการอนุมัติสินเชื่อ โดยจะแสดงผลการอนุมัติสินเชื่อ และไม่อนุมัติสินเชื่อ แยกตามประเภทธุรกิจ และ ประเภทกิจการ



ภาพที่ 6: หน้าแสดงการพยากรณ์การอนุมัติสินเชื่อ



ภาพที่ 7: หน้าแสดงการพยากรณ์การอนุมัติสินเชื่อ (ต่อ)

จากภาพที่ 6 - 7 เป็นการแสดงผลการพยากรณ์จากอัลกอริทึมที่นำมาใช้ในการพยากรณ์ในครั้งนี้ คือ เทคนิค Random Forest มีประสิทธิภาพในการทำนายสูงกว่าเทคนิคอื่นๆ ซึ่งแสดงผลการอนุมัติสินเชื่อ หรือ ไม่อนุมัติเปรียบเทียบกับผลการพยากรณ์ที่พยากรณ์ไว้ โดยสามารถรายงานเลือกตามประเภทธุรกิจ และ ประเภทกิจการ

5. สรุปผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองสำหรับการพยากรณ์การอนุมัติสินเชื่อธุรกิจด้วยเทคนิคเหมืองข้อมูลแบบการจำแนกประเภท และเพื่อประเมินผลการพยากรณ์ของแบบจำลองสำหรับการพยากรณ์การอนุมัติสินเชื่อธุรกิจที่มีการตั้งสมมติฐานไว้ว่าผลการของแบบจำลองสำหรับการพยากรณ์การอนุมัติสินเชื่อธุรกิจมีค่าความถูกต้องมากกว่า 80% โดยการนำข้อมูลการอนุมัติสินเชื่อธุรกิจจากสถาบันการเงินแห่งหนึ่งมาทำให้ข้อมูลมีความถูกต้อง (Data Cleansing) และแบ่งชุดข้อมูลด้วยการวัดผลด้วยวิธีการวัดความถูกต้องแบบไขว้ (Cross Validation) มาพัฒนาแบบจำลองด้วย 5 เทคนิค เพื่อเปรียบเทียบประสิทธิภาพ ได้แก่ Decision Tree, Random Forest, Naïve Bayes, Artificial Neural Network และ Support Vector Machine โดยการประเมินประสิทธิภาพด้วยค่าความถูกต้อง (Accuracy) และ ค่าความแม่นยำ (Precision) ในการเลือกเทคนิคที่มีประสิทธิภาพมากที่สุด ผลลัพธ์ในการเปรียบเทียบพบว่าเทคนิค Random Forest มีประสิทธิภาพในการทำนายสูงกว่าเทคนิคอื่นๆ และการพัฒนาแบบจำลองสำหรับการพยากรณ์อนุมัติสินเชื่อธุรกิจในครั้งนี้พบว่าประสิทธิภาพแบบจำลองสำหรับหารอนุมัติสินเชื่อเพื่อธุรกิจมีค่าความถูกต้องอยู่ที่ 98.03% และค่าความแม่นยำเท่ากับ 98.26% หลังจากสร้างตัวแบบการพยากรณ์เสร็จแล้ว จึงได้นำข้อมูลมาแสดงผลในรูปแบบ Dash Board ผ่าน โปรแกรม Power BI

## 6. ข้อเสนอแนะ

6.1 การศึกษาและเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ครั้งนี้ยังไม่ได้ใช้งานจริง ดังนั้นเมื่อหากต้องการนำไปใช้งานจริงควรนำไปลองทำการประเมินครั้งซ้ำอีกครั้งและความเหมาะสมของแบบจำลองมีความแตกต่างกันขึ้นอยู่กับลักษณะของข้อมูลที่น่ามาพยากรณ์

6.2 ชุดข้อมูลที่นำมาพัฒนาแบบจำลองมีปริมาณที่จำกัดทั้งในเรื่องของปริมาณ และ จำนวนวงเงินที่อนุมัติการใช้ข้อมูลที่มากขึ้นหรือวงเงินที่แตกต่างอาจส่งผลให้มีประสิทธิภาพที่มากขึ้น

วารสารวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยอุบลราชธานี. ปีที่ 20 ฉบับที่ 1 : 44-58.

- [5] Trilok, N. P., et al. (2017). "Credit Risk Analysis using Machine Learning Classifiers." International Conference on Energy, Communication, Data Analytics and Soft Computing. Chennai: The Institute of Electrical and Electrical Engineers (IEEE), (1850-1854).

## เอกสารอ้างอิง

- [1] กอบกาญจน์วัชรจิร โสภณ. (2562). *การพัฒนาแบบจำลองสำหรับการพยากรณ์การอนุมัติสินเชื่อที่อยู่อาศัย*. ปัญหาพิเศษวิทยาศาสตรมหาบัณฑิต สาขาวิชาการบริหารสารสนเทศเพื่อการจัดการภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- [2] Amanda, F., et al. (2020). "Use of machine learning techniques in bank credit risk analysis." *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*. Vol.36, Issue 3, 2020
- [3] ขนิษฐาผลเกตุ. (2562). *การพัฒนาต้นแบบการให้คะแนนสินเชื่อเพื่อการพิจารณาอนุมัติสินเชื่อด้วยเทคนิคเหมืองข้อมูล*. ปัญหาพิเศษวิทยาศาสตรมหาบัณฑิต สาขาวิชาการบริหารสารสนเทศเพื่อการจัดการภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- [4] ณัฐดิษฐ์บุญมีและพงศรินทร์สรุ่ง. (2561). "การประยุกต์ใช้เทคนิคจำแนกข้อมูลแบบต้นไม้ตัดสินใจเพื่อการวินิจฉัยโรคในเบื้องต้นบนโทรศัพท์มือถือ."

# การจำแนกประเภทของอาวุธโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน

## Classification of Weapons using Convolutional Neural Networks

เอกราช นินทรา (Eakarach Nintra)<sup>1</sup> สุรพันธุ์ เอื้อ ไพบุลย์ (Surapan Airphaiboon)<sup>2</sup>

และสมชาติ จิริวิภากร (Somchat Jiriwibhakorn)<sup>3</sup>

<sup>1</sup>สาขาวิชาวิศวกรรมป้องกันประเทศ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เขตลาดกระบัง กรุงเทพฯ 10520

<sup>1</sup>59601154@kmitl.ac.th , <sup>2</sup>surapan.ai@kmitl.ac.th , <sup>3</sup>somchat.ji@kmitl.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนากระบวนการสำหรับการจำแนกอาวุธจากข้อมูลภาพ จำนวน 8 ชนิด โดยการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึก พร้อมทั้งได้พัฒนาโมเดลในการจำแนกย่อยเฉพาะในคลาส เครื่องบินที่สามารถติดยุทโธปกรณ์ ซึ่งมีความยากและซับซ้อนในการจำแนก เนื่องจากมีลักษณะทางกายภาพที่ใกล้เคียงกัน เช่น สี ขนาด และ รูปร่าง ดังนั้นงานวิจัยจึงใช้ TensorFlow และ Keras ซึ่งเป็นไลบรารีสำหรับพัฒนาการเรียนรู้ของเครื่อง ด้วยอัลกอริทึมโครงข่ายประสาทเทียมแบบคอนโวลูชัน อีกทั้งยังเพิ่มประสิทธิภาพของโมเดลการเรียนรู้เชิงลึก พร้อมทั้งเปรียบเทียบ โมเดลพร้อมใช้งานที่ได้รับการฝึกฝนมาก่อน 8 แบบ ผลลัพธ์ที่ได้จากงานวิจัยนี้ โมเดล EfficientNet-B0 มีความถูกต้องในการจำแนกอาวุธจากข้อมูลภาพ เท่ากับ 94.01%

**คำสำคัญ:** โครงข่ายประสาทเทียมแบบคอนโวลูชัน โมเดลการเรียนรู้เชิงลึก โมเดลพร้อมใช้งานที่ได้รับการฝึกฝนมาก่อน

### Abstract

This research aims to develop a procedure for weapons classification from image data of 8 types by using deep learning technique. This paper uses the image data of 8 types of weapons. In addition to this, there are developed model to classify of aircraft that can be armed with equipment. These are aircraft which are difficult to classify because of their similar physical characteristics, such as color, dimension and shape.

Therefore, the research uses TensorFlow and Keras, libraries for developing machine learning with convolutional neural network algorithms. Furthermore, there was also an improvement the efficient of the deep learning model along with the comparison of 8 pre-prepared training models. According to the results of this research, the EfficientNet-B0 model has accuracy of weapons classification from image data 94.01%

**Keywords:** Convolutional Neural Networks, deep learning models, Pre-Trained Model.

### 1. บทนำ

ในบทความนี้ใช้เทคนิคในการประมวลผลภาพโดยใช้หลักการของโครงข่ายประสาทเทียมมาวิเคราะห์ประเภทและลักษณะของอาวุธจากข้อมูลภาพ 8 ชนิด คือ รถถัง เครื่องบินรบ เรือรบ อาวุธปืน จีปนาวุธ เครื่องยิงจรวด มิด และ เฮลิคอปเตอร์ พร้อมกันนี้ได้นำไปปรับใช้กับโมเดลการเรียนรู้ในการจำแนกวัตถุสามารถแยกความแตกต่างระหว่างเครื่องบินทางทหารที่ใช้ในการรบสามารถติดอาวุธหรือยุทโธปกรณ์ในการรบ กับเครื่องบินพลเรือนที่ใช้ในการขนส่ง โดยสาร และเครื่องบินขนาดเล็กส่วนบุคคลได้ โดยใช้ตัวอย่างข้อมูลภาพเครื่องบิน 10 แบบ แตกต่างกันเป็นข้อมูลในการทดสอบและเรียนรู้

### 2. วรรณกรรมหรืองานวิจัยที่เกี่ยวข้อง

#### 2.1 การนำเทคโนโลยีไปใช้ในการพัฒนาอาวุธ

การพัฒนาเทคโนโลยีอุตสาหกรรมป้องกันประเทศ [1] เป้าหมายเพื่อยกระดับไปสู่อุตสาหกรรมที่ใช้เทคโนโลยี



ให้มีขีดความสามารถในการผลิตอาวุธยุทธโปกรณ์ เพื่อต่อต้านก่อการร้ายและสงครามนอกแบบ ด้วยเครื่องมือตรวจจับ เครื่องมืออุปกรณ์สนับสนุนทางยุทธวิธี อาวุธพิเศษ และการใช้ประโยชน์จาก เทคโนโลยีอื่นๆ เป็นต้น

## 2.2 การใช้โครงข่ายประสาทเทียมกับการจำแนกอาวุธ

Dwivedi et al. [2] กล่าวถึงการใช้เทคโนโลยีโครงข่ายประสาทเทียมในการเรียนรู้เชิงลึก เพื่อการจำแนกอาวุธในการก่ออาชญากรรม ช่วยในการระบุความเป็นไปได้ของการเกิดอาชญากรรม โดยใช้สถาปัตยกรรม VGGNet ซึ่งเป็นโมเดลพร้อมใช้งานที่ได้รับการฝึกฝนมาก่อนและเป็นที่ยอมรับในการแข่งขัน ImageNet 2014 การทดลองใช้ภาพ 3 คลาส คือ มิด ปืน และไม่มีอาวุธ ผลการทดลองมีระดับความแม่นยำ 98.41%

Asrith et al. [3] ใช้เทคโนโลยีโครงข่ายประสาทเทียม การจดจำใบหน้าและการตรวจจับอาวุธจากภาพที่มีความละเอียดต่ำมาก ใช้อัลกอริทึม Haar Cascades ในการตรวจจับวัตถุ เพื่อแยกความแตกต่างระหว่างใบหน้าและอาวุธ โดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันในการเรียนรู้ ซึ่งการประมวลผลภาพให้ผลลัพธ์ที่ดีกว่าเมื่อเปรียบเทียบกับอัลกอริทึมการจำแนกรูปภาพแบบอื่น

## 2.3 ไลบรารี TensorFlow และ Keras

การเรียนรู้เชิงลึกเป็นสาขาหนึ่งของปัญญาประดิษฐ์ โดยใช้สถาปัตยกรรมโครงข่ายประสาทเทียมบนแพลตฟอร์มโอเพ่นซอร์สแบบ End-to-End สำหรับการเรียนรู้ของเครื่อง มีระบบสถาปัตยกรรมที่ครอบคลุมและยืดหยุ่นของเครื่องมือ ไลบรารี และทรัพยากร เพื่อพัฒนาและฝึกโมเดล โดย รวมถึงปรับใช้ในระบบคลาวด์ สามารถทำงานได้หลายแพลตฟอร์ม ครอบคลุมหลายภาษา

## 2.4 Transfer Learning

Lili et al. [4] กล่าวถึงการพัฒนาในการเรียนรู้ของเครื่องทำให้เกิดอัลกอริทึมที่สามารถแก้ปัญหาปัญหาที่ซับซ้อน โดยการใช้คุณสมบัติการดึงข้อมูลภาพออกมาประมวลผล บทความนี้ทำการทดลอง และประเมินการเรียนรู้เชิงลึก แบบถ่ายทอดวิธีการเรียนรู้เพื่อเปรียบเทียบคุณลักษณะและกำหนดรูปแบบที่เหมาะสมที่สุดในการปรับใช้คุณสมบัติของโมเดล สามารถตรวจจับประเภทของผีเสื้อด้วยการถ่ายภาพแบบเรียลไทม์ หรือเลือกหนึ่งภาพได้

โดยใช้วิธีการเรียนรู้แบบถ่ายทอดจากโมเดลพร้อมใช้งานที่ได้รับการฝึกฝนมาก่อนผลลัพธ์ในการใช้ VGG19 ให้ค่าความถูกต้อง 98.4%

Mohd et al. [5] การศึกษาการจำแนกรูปภาพตามการเรียนรู้เชิงลึก และใช้ TensorFlow ในการจำแนกข้อมูลนำเข้าคลาสดอกไม้ 5 ประเภท ขนาด 224x224 พิกเซล จำนวน 3,670 ภาพ ด้วยวิธีการ Transfer Learning จาก MobileNet ซึ่งเป็นโมเดลพร้อมใช้งานที่ได้รับการฝึกฝนมาก่อน ผลลัพธ์ที่ได้ความถูกต้องของการจำแนกภาพถูกลบได้ 90.585% และชนิดของดอกไม้ที่เหมือนกัน ผลเฉลี่ยความถูกต้องสูงถึง 90%

ณัฐนันท์ และคณะ [6] แสดงการจำแนกภาพเอกซเรย์คอมพิวเตอร์ โรคหลอดเลือดสมองด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน ภาพที่ผ่านการวินิจฉัยของแพทย์แบ่งเป็น 3 กลุ่ม ข้อมูลภาพขนาด 512x512 พิกเซล จำนวน 450 ภาพ ใช้วิธีการปรับปรุงข้อมูลโดยการประมวลผลภาพก่อนทำการเรียนรู้ร่วมกับโครงข่ายประสาทเทียมแบบคอนโวลูชัน และทดลองด้วย ไลบรารี TensorFlow และ Keras เรียนรู้จำนวน Epochs 500 รอบ พบว่ามีความแม่นยำในการเรียนรู้ Train 99.96%, และ Test 92.60%

## 2.5 โมเดล VGG

Abitya et al. [7] นำเสนอการประเมินโมเดล VGG-16 และ VGG-19 โดยใช้สถาปัตยกรรม การเรียนรู้เชิงลึกเพื่อจำแนกกลุ่มผู้ป่วยโรคสมองเสื่อม ซึ่ง VGG เป็นโครงข่ายประสาทเทียมที่มีความลึก 16 และ 19 ชั้นตามลำดับ ใช้ภาพขนาด 176x208 พิกเซล จำนวน 6,400 ภาพ ค่าความแม่นยำสูงสุดถูกบันทึกที่ Train 99.68% สำหรับการฝึกอบรมและ Validation 99.38% สำหรับการตรวจสอบ

## 2.6 โมเดล MobileNet

Debjoyoti al. [8] นำเสนอการใช้ Thin MobileNet ที่ปรับปรุงสถาปัตยกรรมใหม่ ในการทดสอบข้อมูลภาพ CIFAR-10 dataset ซึ่งการใช้งานเหมาะกับเครื่องมือที่มีประสิทธิภาพในการประมวลผลน้อย ซึ่งใช้ภาพขนาด 32x32 พิกเซล ในการเรียนรู้จำนวน 6,000 ภาพ และใช้ทดสอบ 1,000 ภาพ ค่าความแม่นยำ 85.61% ขนาดของโมเดล 9.9 MB ใช้เวลาในการคำนวณต่อรอบ Epoch 14

วินาที MobileNet เป็นอีกรูปแบบหนึ่งที่ได้รับการฝึกฝนมาแล้ว มีจุดประสงค์ออกแบบให้ใช้งานได้ในทรัพยากรอย่างจำกัด แต่ต้องแลกกับประสิทธิภาพที่น้อยลง

**2.7 โมเดล ResNet**

การใช้งานแบบจำลอง ResNet หรือ Deep Residual Network ได้รับการศึกษาอย่างแพร่หลายและถูกนำมาใช้กับหลากหลายงานวิจัย Arief et al. [9] ได้ศึกษาการจำแนกประเภทมะเร็งผิวหนังเมลาโนมา เป็นมะเร็งที่ตรวจจับเพื่อรักษา ยกโดยใช้โมเดล ResNet50 และการเสริมข้อมูลภาพเพื่อเพิ่มค่าความถูกต้องแม่นยำด้วย Data Augmentation ผลลัพธ์ค่าความถูกต้องแม่นยำ 83 %

**2.8 โมเดล EfficientNet**

Van-Thanh et al. [10] ได้เสนอการวิเคราะห์เชิงปฏิบัติเกี่ยวกับสถาปัตยกรรมของ EfficientNet-B0 ทำการทดสอบประสิทธิภาพโมเดล โดยใช้ข้อมูลภาพ จาก ฐานข้อมูล ILSVRC 2012 ขนาด 224x224 พิกเซล จำนวน 1,200,000 ภาพ และ 50,000 ภาพ แบ่งเป็น 1,000 คลาส เสนอแนวคิดของการตัดในส่วน Squeeze-and-Excitation modules (SE modules) ที่มีความสำคัญต่อการเพิ่มประสิทธิภาพของโมเดล ซึ่งถูกใช้ใน ResNets, VGG และ Inception ในการทดลองได้นำเสนอการตัดออกไปแต่ EfficientNet-B0 ยังสามารถคงประสิทธิภาพการทำงานได้ดี

Chousak et al. [11] ได้นำเสนอ ระบบการเรียนรู้เชิงลึกเพื่อการจดจำและการรู้จำ หมายเลขซีเรียลแถบบนเครื่องบินเพื่อนในฮาร์ดดิสก์ ซึ่งยากต่อการอ่านและแปลความตัวอักษร กระบวนการทดสอบประสิทธิภาพด้วย EfficientNet-B0 แสดงให้เห็นประสิทธิภาพการจัดหมวดหมู่ที่เหนือกว่าการอ่านของมนุษย์ การทดลองกับภาพถ่าย 15,000 ภาพให้ความแม่นยำมากกว่า 99%

**3. วิธีการดำเนินการวิจัย**

**3.1 การเก็บรวบรวมข้อมูล**

การนำข้อมูลภาพจาก แหล่งข้อมูลเว็บไซต์ของ Kaggle หรือแหล่งข้อมูลอื่นๆ เช่น จากภาพถ่าย นิตยสารเป็นต้น เข้าไปเตรียมไว้ ขนาดไม่เกิน 640x480 พิกเซล

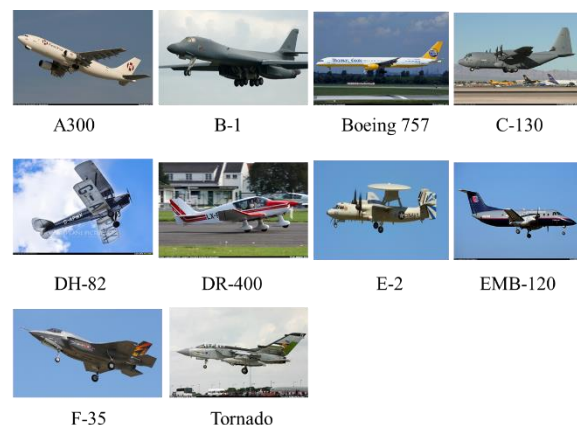
**3.2 การเตรียมชุดข้อมูล**

ในบทความนี้แบ่งข้อมูลชุดข้อมูลสำหรับฝึกฝนโมเดล จากกลุ่มที่ 1 จำนวนภาพ 10,789 รูปภาพ ชุดข้อมูลรูปภาพ อาวุธ รถถัง (Tank), เครื่องบินรบ (Military Aircraft), อาวุธปืน (Gun), เรือรบ (Military ship), จีปนาวุธ (Missile rocket), เครื่องยิงจรวด (Missile launcher), มีด(Knife) และ เฮลิคอปเตอร์ (Helicopter) กลุ่มที่ 2 จำนวนภาพ 21,211 รูปภาพ ชุดข้อมูลรูปภาพเครื่องบิน 10 แบบ เครื่องบินทางทหารที่สามารถติดขีปนาวุธ (B-1, C-130, E-2, EM B-120, F-35, Tornado) เครื่องบินพลเรือนขนส่งโดยสาร (A300, Boeing 757) และเครื่องบินพลเรือนขนาดเล็กใช้ส่วนบุคคล (DH-82, DR-400)

ในบทความนี้การแบ่งข้อมูลออกเป็น 3 ส่วนคือ Train 70%, Validation 15% และ Test 15% ใช้สำหรับการฝึกฝนข้อมูล การปรับพารามิเตอร์ และการทดสอบโมเดล เพื่อจำแนกข้อมูลชุดภาพ



ภาพที่ 1: ตัวอย่างชุดข้อมูลรูปภาพอาวุธ 8 ชนิด

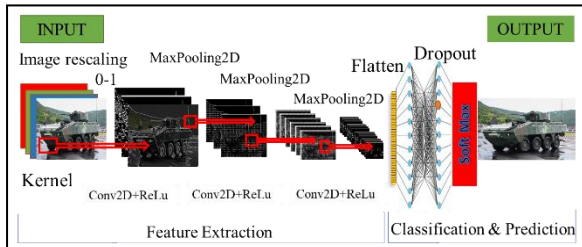


ภาพที่ 2: ตัวอย่างชุดข้อมูลรูปภาพเครื่องบิน 10 แบบ

**3.3 โมเดลที่ใช้ในการจำแนกรูปภาพ**

ในงานวิจัยนี้เลือกทดลองใช้โมเดลจำนวน 9 แบบ เพื่อใช้ในการจำแนกรูปภาพ คือโมเดลที่ผู้วิจัยสร้างแบบจำลองขึ้น โมเดลมาตรฐาน Standard จำนวน 1 โมเดล และ

เปรียบเทียบกับ โมเดลพร้อมใช้งานที่ได้รับการฝึกฝนมาก่อนแล้วล่วงหน้า 8 แบบ ได้แก่ VGG19, ResNet50V2, MobileNetV2, EfficientNet-B0, B1, B2, B5 และ B7 ที่ถูกรวมไว้ใน Keras API โดยนำเข้าชุดข้อมูลรูปภาพในการเรียนรู้ ซึ่งกำหนดขนาดภาพและจำนวนภาพเท่ากันในแต่ละชุดข้อมูลของการทดลอง



ภาพที่ 3: แบบจำลองโมเดลมาตรฐาน Standard

### 3.4 วัดประสิทธิภาพโมเดลด้วย Confusion Matrix

งานวิจัยนี้ประเมินประสิทธิภาพของตัวโมเดลประเภทอาวุธ ด้วยภาพค่าความถูกต้อง (Accuracy), ค่าความแม่นยำ (Precision), ค่าความถูกต้องระลึกพิจารณาแยกที่ละกลุ่ม (Recall), และ ค่าอฟสกออร์ (F1-Score) ดังสมการที่ 1, 2, 3 และ 4

ตารางที่ 1: วัดประสิทธิภาพด้วย Confusion Matrix

Confusion Matrix	Actual Positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

$$Accuracy = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$Precision = TP/(TP+FP) \quad (2)$$

$$Sensitivity/Recall = TP/(TP+FN) \quad (3)$$

$$F1 = 2 * [(precision * recall) / (precision + recall)] \quad (4)$$

TP (True Positive) คือ โมเดลทำนายว่าประเภทอาวุธ ถูกต้องตามความเป็นจริงในคลาสที่กำลังพิจารณา

FP (False Positive) คือ โมเดลทำนายผิดว่าประเภทอาวุธถูกต้องแต่ไม่ตรงกับเป็นความจริง

FN (False Negative) คือ โมเดลทำนายผิดว่าประเภทอาวุธ ไม่ถูกต้อง แต่ประเภทอาวุธตรงกับความเป็นจริง

TN (True Negative) คือ โมเดลทำนายถูกต้องว่าประเภทอาวุธผิด และประเภทอาวุธไม่ถูกต้องจริงในคลาสที่กำลังพิจารณา

### 3.5 เครื่องมือที่ใช้ในการพัฒนาระบบ

การเตรียมอุปกรณ์และสภาพแวดล้อมการทำงานวิจัยนี้ ใช้เครื่องมือที่มีการพัฒนารองรับมาตรฐานเพื่อสร้างสภาพแวดล้อมในการทำงานบน Anaconda และได้ติดตั้งไลบรารีรุ่นที่เหมาะสมเพิ่มเติม เช่น Python, TensorFlow GPU, Keras API, Jupyter Notebook, OpenCV, numpy etc. อุปกรณ์ที่ใช้ในการพัฒนาระบบ CPU AMD Ryzen7 2.3GHz, RAM 24576MB, GPU NVIDIA GeForce GTX 1660 Ti

### 4. ผลการดำเนินงาน

ทำการทดลองโดยนำข้อมูลเข้าสู่โมเดลมาตรฐาน Standard เพื่อทำการเรียนรู้ใช้ชุดข้อมูลตามภาพที่ 1 ขนาด 128x128 พิกเซล จำนวน 10,789 ภาพ ใช้ในการเรียนรู้ 8,651 ภาพ ใช้สำหรับตรวจสอบ 1,069 ภาพ และใช้สำหรับทดสอบ 1,069 ภาพ ทำการเรียนรู้ Epochs 100 รอบ เพื่อทดสอบในข้อมูลพื้นฐานแบบเดียวกัน

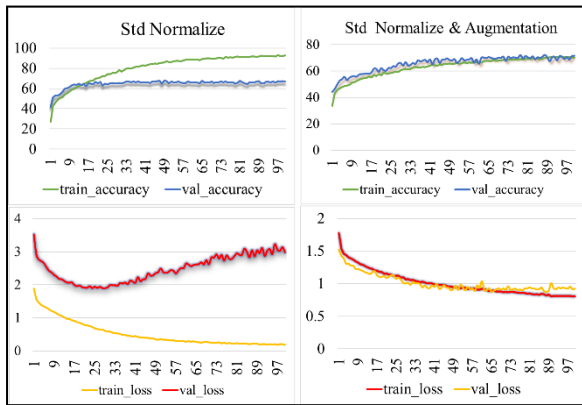
จากการทดลองมีการปรับค่าการคำนวณทางตัวเลขโดยการปรับค่าพารามิเตอร์ด้วยเทคนิค Batch Normalization เพื่อทำการปรับค่าข้อมูลให้อยู่ในช่วง 0-1 ก่อนนำเข้าไปเรียนรู้ ผลลัพธ์ที่ได้จากตารางที่ 2 โมเดลมาตรฐาน Standard มีค่าความแม่นยำ 59.40% เมื่อพิจารณาจากกราฟ ในภาพที่ 4 การทดลองเกิดปัญหา Overfitting ซึ่งพิจารณาจาก ค่า Loss เป็นหลัก พบว่ายิ่งมีการ Train มากขึ้น ค่า Training Loss จะลดลงอย่างต่อเนื่อง ขณะที่ Validation Loss จะลดลงถึงจุดหนึ่งแล้วหลังจากนั้นกลับมีการเพิ่มค่าขึ้นเรื่อยๆ จากผลเปรียบเทียบประสิทธิภาพระหว่างโมเดลมาตรฐาน Standard และ การเสริมข้อมูล ภาพด้วยการทำ Data Augmentation เพิ่มประสิทธิภาพด้วยการใช้เทคนิคในการปรับมุมมองภาพก่อนนำเข้าสู่การโมเดล และการ Dropout ผลลัพธ์ที่ได้ค่าความแม่นยำเพิ่มขึ้นเป็น 66.79%

การทดลองทำต่อไปใน โมเดลพร้อมใช้งานที่ได้รับการฝึกฝนมาก่อน (Pre-Trained) โดยการการทำ Transfer Learning เพิ่มประสิทธิภาพการเรียนรู้ของโมเดล พร้อมกับการเปรียบเทียบโมเดลพร้อมใช้งานที่ได้รับการฝึกฝนมาก่อน 8 โมเดล เป็นตัวแยกคุณลักษณะคงที่ Fixed Feature Extractor จากนั้นนำมาลบส่วนของ Dense Layer สุดท้าย

ออกไปจะได้ Feature Extractor ที่สามารถสร้างส่วนหัว Custom Head Layer ตามความเหมาะสมกับงานวิจัย

ตารางที่ 2: ผลการประเมิน โมเดลมาตรฐาน Standard

Standard	Accuracy (%)	Weighted avg		
		Precision	Recall	F1-score
Normal	59.40	0.61	0.59	0.60
Aug & Drop	66.79	0.67	0.67	0.67



ภาพที่ 4: เปรียบเทียบประสิทธิภาพ โมเดลมาตรฐาน Standard

ทำการทดลองการเรียนรู้โดยปิดตัวแยกคุณลักษณะของ โมเดล (Freeze model) การทดลองจากตารางที่ 3 ผลการ ประเมิน โมเดล Pre-Trained พบว่า EfficientNet-B0 ให้ ผลลัพธ์ความแม่นยำเพิ่มขึ้นเป็น 92.33% เพื่อให้ได้ผล การทดลองที่ดีที่สุด จึงพัฒนาสู่การจำแนกภาพที่ซับซ้อน มากขึ้น ซึ่งสามารถเพิ่มประสิทธิภาพของ โมเดลด้วยการ ปรับค่าพารามิเตอร์และเทคนิค Data Augmentation ใน การปรับมุมมองภาพก่อนนำเข้าสู่การ โมเดล และการทำ Fine-Tuning ปิดการเรียนรู้บางส่วน of โมเดล (Freeze Model) และเปิดการเรียนรู้เพียงส่วนส่วน (Unfreeze Model) เพื่อไม่ให้มีการอัปเดตค่าของ Weight จากนั้น Train เฉพาะส่วน หัว (Custom Head Layer) เพื่อปรับ Weight ตามความเหมาะสม

จากการทดลองในตารางที่ 4 ทำการทดลองปรับค่าใน EfficientNet-B0 ซึ่งมีจำนวนชั้นแยกคุณลักษณะแบบคอน โวลูชัน 238 Layers ทำการเปิดการเรียนรู้ 50, 100 และ 200 Layers ตามลำดับให้ผลลัพธ์ความแม่นยำมากที่สุดใน การเปิดการเรียนรู้ 100 Layers เพิ่มขึ้นเป็น 94.01% ใช้ รูปภาพในการทดสอบโมเดล Support 1069 ภาพ นำมา คำนวณหาค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก Weighted

Average ในกรณีที่ข้อมูลภาพในแต่ละคลาสไม่เท่ากัน ผล การประเมิน Precision 0.94, Recall 0.94, F1-score 0.94 ตามภาพที่ 5 แสดงค่าของการประเมิน Confusion Matrix ของรูปภาพ 8 ชนิด ซึ่งใช้ในการแสดงค่าความถูกต้องของ การทำนายผลของโมเดล

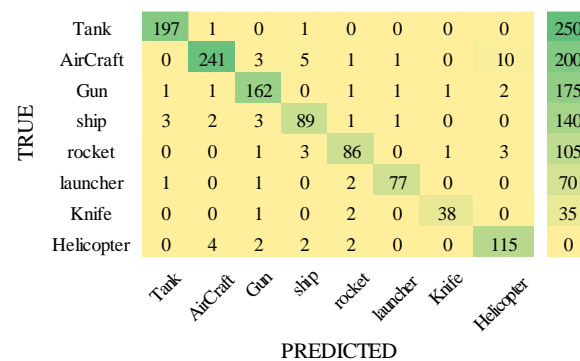
การประยุกต์ใช้งานของโมเดลด้วยการนำไปทดสอบ กับชุดข้อมูลรูปภาพเครื่องบินทางทหารที่สามารถติด ยุทธโประกรณ์ ซึ่งเป็นอันตรายและภัยคุกคามต่อความมั่นคง ในจำนวนเครื่องบิน 10 แบบ เพื่อทดสอบความสามารถ ของโมเดลในการจำแนกภาพที่มีความซับซ้อนมากขึ้น จาก ตารางที่ 4 ผลลัพธ์ความแม่นยำอยู่ในเกณฑ์ที่สูง 93.67% ผลการประเมิน Precision 0.94, Recall 0.94, F1-score 0.94 แสดงให้เห็นว่าโมเดลมีความยืดหยุ่นและทำงานได้ อย่างมีประสิทธิภาพ แม้จะมีการเพิ่มจำนวนข้อมูล

ตารางที่ 3: ผลการประเมิน โมเดลจากการ Transfer Learning

Pre-Trained	Accuracy (%)	Weighted avg		
		Precision	Recall	F1-score
VGG19	85.22	0.86	0.85	0.85
ResNet50V2	58.75	0.69	0.59	0.62
MobileNetV2	88.59	0.89	0.89	0.89
EfficientNetB0	92.33	0.92	0.92	0.92
EfficientNetB1	89.52	0.90	0.89	0.89
EfficientNetB2	90.36	0.91	0.90	0.90
EfficientNetB5	86.81	0.87	0.87	0.87
EfficientNetB7	88.96	0.89	0.89	0.89

ตารางที่ 4: ผลการประเมิน โมเดลจากการ Fine-Tuning

EfficientNet-B0 Fine-Tuning	Accuracy (%)	Weighted avg		
		Precision	Recall	F1-score
Unfreeze 50 Layer	91.67	0.92	0.92	0.92
Unfreeze 100 Layer	94.01	0.94	0.94	0.94
Unfreeze 200 Layer	93.45	0.94	0.93	0.93
10 type of Aircraft	93.67	0.94	0.94	0.94



ภาพที่ 5: Confusion Matrix ของการจำแนกอาวุธ 8 ชนิด

## 5. สรุป

งานวิจัยนี้ได้ทำการทดลองประสิทธิภาพของโมเดลในการจำแนกประเภทของอาวุธโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน พร้อมกับการเปรียบเทียบโมเดลพร้อมใช้งานที่ได้รับการฝึกฝนมาก่อน 8 โมเดล พร้อมกับพัฒนาผู้การจำแนกภาพที่มีข้อมูลซับซ้อนมากขึ้น วิธีการปรับเสริมข้อมูลภาพมาใช้ร่วมกับโครงข่ายประสาทเทียมด้วยวิธีการ Transfer Learning และ Fine-Tuning เพิ่มประสิทธิภาพในการจำแนกหมวดหมู่อาวุธ จำนวน 8 ชนิด และการจำแนกความแตกต่างในคลาสอาวุธที่เป็นเครื่องบินรบสามารถคิดชุดโปรแกรม และเครื่องบินพลเรือนได้อย่างมีประสิทธิภาพ ผลลัพธ์จากการทดลองในงานวิจัยนี้ EfficientNet-B0 มีความเหมาะสมในการนำไปใช้งานที่สุด สามารถทำงานได้อย่างถูกต้องแม่นยำสูงสุด 94.01% จากผลการวิจัยชี้ให้เห็นว่าแนวทางการพัฒนาเพิ่มประสิทธิภาพของโมเดลสามารถนำไปใช้ในการจำแนกกลุ่มของข้อมูลภาพที่มีลักษณะทางกายภาพที่คล้ายคลึงกันซึ่งมีประโยชน์ในการพัฒนาใช้กับอุปกรณ์ติดตั้งกับอากาศยานไร้คนขับในการสำรวจ ค้นหา ใฝ่ระวัง ติดตามเป้าหมาย หรือการลาดตระเวนทางอากาศ เพื่อใช้ในงานอุตสาหกรรมป้องกันประเทศในอนาคต

## 6. กิตติกรรมประกาศ

ขอขอบคุณสถาบันเทคโนโลยีป้องกันประเทศ (องค์การมหาชน) กระทรวงกลาโหม ผู้สนับสนุนในการทำวิจัยในครั้งนี้จนสำเร็จลุล่วง

### เอกสารอ้างอิง

- [1] สถาบันเทคโนโลยีป้องกันประเทศ (องค์การมหาชน) กระทรวงกลาโหม, “แผนที่ นำทางเทคโนโลยีป้องกันประเทศ,” [ออนไลน์]. [สืบค้นเมื่อ 28 กุมภาพันธ์ 2565], จาก <https://www.dti.or.th/download/file/e7dd8350.pdf>
- [2] N. Dwivedi, D. K. Singh and D. S. Kushwaha, "Weapon Classification using Deep Convolutional Neural Network," *2019 IEEE Conference on Information and Communication Technology*, 2019, pp. 1-5, doi: 10.1109/CICT48419.2019.9066227.
- [3] M. J. N. V. S. K. Asrith, K. P. Reddy and Sujihelen, "Face Recognition and Weapon Detection from Very Low Resolution Image," *2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR)*, 2018, pp. 1-5, doi: 10.1109/ICETIETR.2018.8529108.
- [4] L. Zhu and P. Spachos, "Towards Image Classification with Machine Learning Methodologies for Smart phones," *Machine Learning and Knowledge Extraction*, Licensee MDPI, Basel, Switzerland, 14 Feb, 2019, pp.1039-1057.
- [5] M. A. Abu, N. H. Indra, A. H. A. Rahman, N. A. Sapiee and I. Ahmad, "A study on Image Classification based on Deep Learning and TensorFlow," *International Journal of Engineering Research and Technology*, Vol. 12, No. 4, pp. 563-569, 2019.
- [6] Nattanon Lamsamut and Sompong Valuvanathorn, "Stroke Disease Classification on Computerized Tomography Scan Images Using Convolutional Neural Network," *The 17th National Conference on Computing and Information Technology*, May 13- 14, 2021, pp. 43-48.
- [7] A. Bagaskara and M. Suryanegara, "Evaluation of VGG-16 and VGG-19 Deep Learning Architecture for Classifying Dementia People," *4th International Conference on Computer and Informatics Engineering (IC2IE)*, DOI:10.1109/IC2IE53219.2021.9649132, 2021.
- [8] D. Sinha and M. El-Sharkawy, "Thin MobileNet: An Enhanced MobileNet Architecture," *IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, DOI: 10.1109/UEMCON47517.2019.8993089, 2020.
- [9] A. Budhiman, S. Suyanto and A. Arifianto, "Melanoma Cancer Classification Using ResNet with Data Augmentation," *International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, DOI: 10.1109/ISRITI48646.2019.9034624, 2020.
- [10] V. T. Hoang and K.H. Jo, "Practical Analysis on Architecture of EfficientNet," *14th International Conference on Human System Interaction (HSI)*, DOI: 10.1109/HSI52170.2021.9538782, 2021.
- [11] C. Chousangsuntorn, T. Tongloy, S. Chuwongin and S. Boonsang, "A Deep Learning System for Recognizing and Recovering Contaminated Slider Serial Numbers in Hard Disk Manufacturing Processes," *Sensors*. 2021; 21(18) :6261. <https://doi.org/10.3390/s21186261>, 2021.

# การศึกษาและพัฒนาอัลกอริทึม Bitmap Intersection Lookup สำหรับ Packet Classification

## A Study and Development of Bitmap Intersection Lookup Algorithm for Packet Classification

วรเชษฐ์ นุ่มเจริญ (Worachet Nimcharoen)<sup>1</sup>, สุภักวี สุโพธิ์ (Supakkavee Supho)<sup>2</sup>

และอักรินทร์ คุณกิตติ (Akharin Khunkitti)<sup>3</sup>

ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

<sup>1</sup>61070196@it.kmitl.ac.th, <sup>2</sup>61070251@it.kmitl.ac.th, <sup>3</sup>akharin@it.kmitl.ac.th

### บทคัดย่อ

โครงการเรื่อง การศึกษาและพัฒนาอัลกอริทึม Bitmap Intersection Lookup สำหรับ Packet Classification มีจุดประสงค์ในการจัดทำเพื่อลดเวลาในการค้นหาข้อมูลในระบบเครือข่าย โดยใช้ Algorithm โดยมุ่งเน้น ในด้านการสร้าง Bitmap Intersection Lookup Algorithm โดยนำมาทดลองใช้กับ Firewall Rules เพื่อวัดประสิทธิภาพ โดยเทียบประสิทธิภาพกับ Search Algorithm ต่างๆ โดยมีการทดสอบ 3 รูปแบบ ได้แก่ วัดความถูกต้องในการค้นหา วัดเวลาที่ใช้ วัดความจุที่ใช้ โดยจะมีการทดสอบโดยการแบ่ง Segment เพื่อลดความจุ ลดเวลาในการสร้างตารางเป็นต้น โดยจะหาว่าการแบ่งรูปแบบใดมีประสิทธิภาพมากที่สุด จากการวิจัย Linear Search นั้นจะขึ้นอยู่กับจำนวนกฎ ซึ่งต่างจาก Bitmap Intersection Lookup ที่ไม่ขึ้นอยู่กับจำนวน Rules โดยในด้านความถูกต้องนั้นการค้นหาของ Bitmap Intersection Lookup เมื่อเทียบกับ Linear Search แล้ว มีการค้นหาที่ถูกต้องเหมือนกัน ยิ่งแบ่ง Segment มากจะลดเวลาที่ใช้ในการทำงานและความจุน้อยลง แต่เวลาในการค้นหาจะเพิ่มขึ้นตามจำนวน Segment

**คำสำคัญ:** Bitmap Intersection Lookup(BIL) การค้นหาแบบเส้นตรง เซ็กเมนต์ อัลกอริทึม

### Abstract

*Project name: A Study and Development of Bitmap Intersection Lookup Algorithm for Packet Classification. It is intended to reduce the time spent searching for information in the network. Focus on creating Bitmap Intersection Lookup Algorithm to be tested with Firewall Rules to measure efficiency. It compares its performance to various search algorithms. There are three types of tests: 1. correct search 2. measure the time spent 3. Measure the used capacity It will be tested by dividing the segments into smaller segments to reduce the capacity, reduce the time to create tables, etc. To find out which form of division is most effective. According to research, Linear Search is a number of rules, unlike Bitmap Intersection Lookup is independent of rule count. In terms of correct, Bitmap Intersection Lookup searches, compared to Linear Search, have the same result. Segmentation takes less time to operate and less capacity. But the search time increases with the number of segments.*

**Keywords:** Bitmap Intersection Lookup(BIL), Linear Search, Segment, Algorithm.

## 1. บทนำ

ปัจจุบันมีผู้ใช้งานระบบเครือข่ายจำนวนมาก ส่งผลให้มีข้อมูลในระบบเครือข่ายเพิ่มมากขึ้น ระบบเครือข่ายมีความล่าช้า ตัวอย่างที่ทำให้ระบบเครือข่ายล่าช้า เช่น การส่งข้อมูลเข้าไปในเครือข่ายที่มีความปลอดภัยสูง ซึ่งมี Firewall Rules จำนวนมากที่ต้องคัดกรองข้อมูล จึงทำให้เกิดปัญหาในการค้นหา Firewall Rules ที่ใช้ในการคัดกรองข้อมูลที่เข้ามา ในกรณีที่เกิดที่ต้องการค้นหาข้อมูลอยู่ในลำดับสุดท้าย จะใช้ระยะเวลาในการค้นหา และในกรณีที่มีข้อมูลจำนวนมากก็ยิ่งส่งผลให้เวลาในการค้นหาเพิ่มมากขึ้น ทำให้เสียเวลา มาก จึงมีจุดประสงค์เพื่อพัฒนา Algorithm ที่ช่วยในการจัดการปัญหาดังกล่าวเพื่อจัดการปัญหา Packet Classification และยังสามารถใช้ในการจัดการปัญหาอื่นๆ เช่น ปัญหาที่ Firewall Ids หรือ Ips เป็นต้น โดยเลือกใช้ Bitmap Intersection Lookup ในพัฒนา Algorithm เพื่อให้สามารถทำงานได้เร็วขึ้น โดยจะให้สามารถทำงานได้ทั้งบน Software และ Hardware โดยมีวัตถุประสงค์เพื่อนำเทคนิค Bitmap Intersection Lookup มาช่วยลดเวลาในการทำ Packet classification โดยมุ่งเน้นให้เวลาในการค้นหาไม่ขึ้นอยู่กับจำนวน Rules สามารถนำความรู้และเทคนิค Bitmap Intersection Lookup นำไปประยุกต์ใช้ได้ในอนาคต สามารถนำ Algorithm ที่พัฒนามาใช้งานเพื่อใช้ในการแก้ปัญหาที่มีลักษณะคล้ายกัน สามารถนำความรู้ที่ได้มาช่วยในการจัดการระบบเครือข่ายล่าช้าที่มีสาเหตุมาจาก Firewall Rules ที่มี จำนวนมาก

## 2. การทบทวนเอกสารวรรณกรรม

ผู้วิจัยได้รวบรวมศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องที่เป็นความรู้พื้นฐาน ในด้านเนื้อหาและเทคโนโลยี เพื่อเป็นข้อมูลในการพัฒนา ส่วนประกอบในบทนี้ประกอบไปด้วย

### 2.1 Search Algorithm

การค้นหา คือหนึ่งในการดำเนินงาน ของ Data Structure ใช้ในการค้นหาข้อมูลจาก List ของข้อมูล การค้นหานี้ สามารถเป็นได้ทั้งข้อมูลที่เรียงลำดับ และเป็น

ข้อมูลที่ไม่เรียงลำดับ ทั้งนี้ขึ้นอยู่กับประเภทของ Search Algorithm ที่ใช้ ตัวอย่างของ Search Algorithm

#### 2.1.1 Linear Search

Linear Search เป็น Search Algorithm ที่ไม่มีความซับซ้อน เป็นการ ค้นหาข้อมูลที่ละข้อมูลตามลำดับ ตัวอย่างของ Linear Search ถ้าต้องการหาค่า 53 โดยมีการเก็บข้อมูล 11 ตัว Linear Search จะทำการตรวจสอบที่ละตัวจนพบข้อมูลที่ต้องการ โดยวิธีนี้ใช้ขั้นตอนในการค้นหา 10 ขั้นตอนจึงพบข้อมูลที่ต้องการ

#### 2.1.2 Binary search

Binary Search เป็น Search Algorithm ที่มีความเร็วและไม่ซับซ้อนเหมาะกับข้อมูลที่มีปริมาณไม่มาก การค้นหาข้อมูลในรูปแบบนี้จะสามารถใช้ค้นหาได้เฉพาะข้อมูลที่ถูกระบุค่าแล้ว โดยเริ่มจากการหาข้อมูลที่อยู่กึ่งกลาง โดยมีสูตรว่า ข้อมูลตำแหน่งกึ่งกลาง =  $[(ตำแหน่งข้อมูลตัวแรก + ตำแหน่งข้อมูลตัวสุดท้าย) / 2]$  ขั้นตอนที่ 2 เปรียบเทียบข้อมูลที่ ต้องการหา ในกรณีนี้คือ 53 ถ้าข้อมูลที่ ต้องการหามีค่าน้อยกว่าค่าแห่งข้อมูลกึ่งกลาง จะตัดตำแหน่งข้อมูลทั้งหมดในส่วนที่มากกว่าค่าแห่งข้อมูลกึ่งกลาง หรือถ้าข้อมูลที่ ต้องการหามีค่ามากกว่าค่าแห่งข้อมูลกึ่งกลาง จะตัดตำแหน่งข้อมูลที่ค่าน้อยกว่าค่าแห่งข้อมูลกึ่งกลางดังรูป (2.2) ทำขั้นตอนเดิมซ้ำๆจนพบค่าที่ต้องการค้นหา โดย Binary Search ใช้ขั้นตอนในการค้นหา 7 ขั้นตอน

### 2.2 แนวคิดการทำงานของ Bitmap Intersection Lookup

กลไก Bitmap Intersection Lookup (BIL) เป็นการ ใช้ Bitmap Table ในการ ค้นหา โดยในข้อมูลแต่ละ Field จะมีกฎ จะถูกแทนที่ด้วย BIT แล้วใส่ในตาราง ถ้ามีกฎอยู่ 100 กฎ ก็จะมี BIT 100 BIT โดยที่ BIT ตัวที่ 1 จะหมายถึงกฎข้อที่ 1 โดยจะแทนจนครบจำนวนกฎที่มี และสามารถเพิ่มกฎได้ ถ้าตำแหน่ง ในแต่ละ BIT มีค่าเป็น 1 แสดงว่า ค่านั้นตรงกับกฎข้อนั้นๆ ถ้าเป็น 0 แสดงว่าไม่ตรงกับกฎข้อนั้นๆ เมื่อ ได้ค่ามาแล้วค่ามาเข้าตรรกะ AND กันจนครบทุกตัวจะได้ค่าของกฎที่ต้องการ 2.3.1 ประเภทของเทคโนโลยีความเป็นจริงเสริม

หลักการทำงานของเทคโนโลยีความเป็นจริงเสริมนี้ สามารถแบ่งประเภทตามส่วนวิเคราะห์ภาพ (Image

Analysis) ได้เป็น 2 ประเภท โดยอ้างอิงจากการค้นหาภาพจาก Marker จากภาพที่ได้จากกล้องแล้วสืบค้นจากฐานข้อมูล (Marker Database)

วิธีการทำงานของ Bitmap Intersection Lookup 1 สร้างตาราง BIL เริ่มจากการแบ่ง Segment ในการสร้างตาราง เช่น แบ่ง 2 Segment ขนาด 8 ต่อ 8 จะได้ 2 ตารางขนาด 2 ยกกำลัง 8 จำนวน 2 ตาราง เมื่อสร้างตาราง นำ Index ไป match กับ Rules ถ้าถูกต้องก็อัปเดตตารางด้วยค่า 1 ถ้าไม่ถูกต้องก็อัปเดตตารางด้วยค่า 0 และไป Rules ต่อไปจนครบทุก Rules

2 ทำขั้นตอนที่ 1 ในทุก Index จนสร้างตารางเสร็จ

3 การค้นหาให้นำ Input มาตรวจสอบกับ Bil Table ที่สร้างขึ้น ถ้ามีหลาย Segment ให้นำ Input มาแบ่งตาม Segment แล้วเอาไป Search ในแต่ละตารางตามลำดับแล้วนำค่ามา AND กัน

4 ถ้า Rules ที่ Search เจอเป็น 1 ให้มาเช็ค Rule ID เช่น 00101 เจอตำแหน่งที่ 3 Rule ID = 3 แล้วจบการทำงาน (ทำงานแบบ first come first serve)

### 2.3 แนวคิดการจัดประเภทแพ็คเกจในระบบเครือข่าย

การจัดประเภทแพ็คเกจที่มีจุดประสงค์เพื่อให้สามารถใช้งานระบบเครือข่ายได้อย่างมีประสิทธิภาพและมีความปลอดภัย และเพื่อให้ระบบเครือข่ายมีความเร็วที่เพิ่มมากขึ้น จึงต้องมีการพัฒนาอุปกรณ์ต่างๆมาช่วยเพิ่มประสิทธิภาพ ซึ่งจำเป็นต้องมีการจัดการแพ็คเกจเพื่อแยกแยะให้แพ็คเกจแต่ละตัวที่เข้ามาถูกจัดประเภทอย่างเหมาะสม โดยการจัดประเภทแพ็คเกจมีความเร็วไม่มากพอสำหรับความเร็วในการเชื่อมต่อที่พัฒนาไป

ในปัจจุบัน การพัฒนาอุปกรณ์เครือข่ายให้มีความเร็วสูงขึ้น โดยการนำเทคนิคการประมวลผลและการจัดประเภทแพ็คเกจมาพัฒนาด้วยอุปกรณ์ Hardware เพื่อให้สามารถทำงานได้ด้วยความเร็วที่สูงได้เต็มที่

### 2.4 เทคโนโลยีและเครื่องมือที่ใช้ในการพัฒนา

#### 2.4.1 ฮาร์ดแวร์ (Hardware)

อุปกรณ์ Field-Programmable Gate Array (FPGA) เป็นชิปจำนวนเล็กน้อยมาก สามารถโปรแกรมให้สามารถเป็นวงจรถิจิตลหลากหลายรูปแบบได้

#### 2.4.2 ซอฟต์แวร์และเครื่องมือ (Software and Tools)

Visual Studio Code เป็นโปรแกรมสำหรับการเขียนโปรแกรมหรือซอฟต์แวร์ต่างๆ เป็นซอฟต์แวร์ของไมโครซอฟท์ที่ไม่เสียค่าใช้จ่าย สามารถรองรับการทำงานได้หลายระบบปฏิบัติการ เช่น Windows, macOS, Linux เป็นต้น Visual Studio Code รองรับหลายภาษาคอมพิวเตอร์ Visual Studio Code สามารถลงส่วนเสริมที่สามารถช่วยให้การทำงานสะดวกสบายมากยิ่งขึ้น

#### 2.4.3 ภาษาโปรแกรม (Programming Languages)

ภาษา C ถูกพัฒนาขึ้นโดย เดนิส ริทซ์ พัฒนามาจากภาษาบี ภาษาซีถูกนำไปพัฒนาระบบปฏิบัติการ การยูนิกซ์ เป็นภาษา

## 3. วิธีการดำเนินงาน

การวิจัยครั้งนี้เป็นการวิจัยเพื่อตรวจสอบความถูกต้องและประสิทธิภาพของ Algorithm BIL ดังนั้นผู้จัดทำโครงการมีวิธีดำเนินงาน ดังนี้

### 3.1 เครื่องมือใช้ดำเนิน

- โปรแกรม Visual studio code
- ติดตั้งภาษา C ในเครื่อง
- Microsoft Excel
- Microsoft Word

### 3.2 ขั้นตอนการทำงานของ Bitmap Intersection Lookup

1 Create Table สร้างตาราง Bitmap Intersection Lookup เช่น มีข้อมูลขนาด 4 Bits สร้าง Bitmap Table 1 field มีค่า index 2 กำลัง 4 ค่า

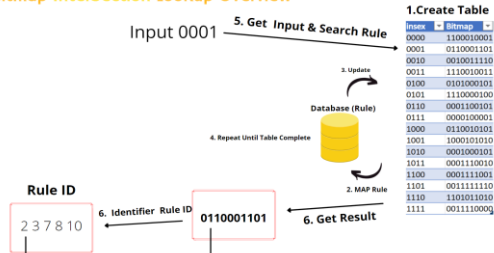
2 Map Rule นำค่า index ไป match ใน rules ซึ่งใน Rules จะมีค่า Value กับ Mask โดยตรวจสอบที่ละ Rules ตั้งแต่ กฎที่ 1 ไปจนถึง Rules สุดท้าย

3 Update Table จากขั้นตอนที่ 2 ในกรณีที่ Rules Match กับ Index ให้นำค่า 1 ไป Update ใน Table ในกรณีที่ Rules ไม่ Match กับ Index ให้นำค่า 0 ไป Update ใน Table



- 4 Repeat Until Table Complete ทำซ้ำจนครบทุก Index (Row ใน Bitmap Table คือ ตำแหน่งของ Rules)
- 5 Get Input & Search Rule นำค่า Input ที่รับมาไป ค้นหาใน Bitmap Table
- 6 Identifier Rule ID นำผลลัพธ์ที่ได้มา ระบุ Rule ID

**Bitmap InterSection Lookup Overview**

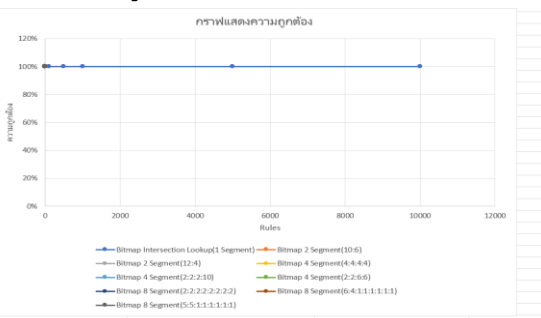


รูปที่ 1: ขั้นตอนการทำงานของ Bitmap InterSection Lookup

**4. ผลการทดลอง**

การทดลองของ Bitmap InterSection Lookup แบ่งออกเป็น 3 ประเภท โดยเริ่มต้นที่ทดสอบความถูกต้องก่อน เพื่อดูว่าสามารถนำมาใช้ได้หรือไม่ จากนั้นคือ ทดสอบประสิทธิภาพ โดยวัดจากความเร็วการทำงาน และสุดท้ายความจุ เพื่อบอกว่า BIL สามารถนำมาใช้แทน Linear search ได้หรือไม่ โดยจะทดลองที่ 1 Field 16 Bits และ 10, 100, 500, 1000, 5000, 10000 Rules

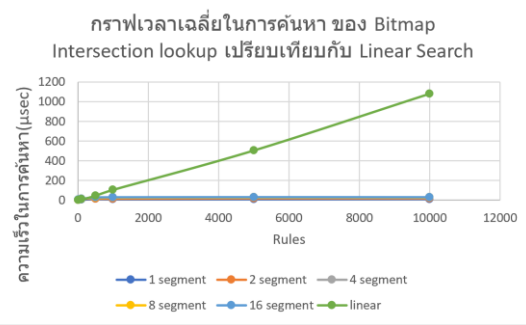
**4.1 ความถูกต้อง**



รูปที่ 2: กราฟแสดงความถูกต้อง

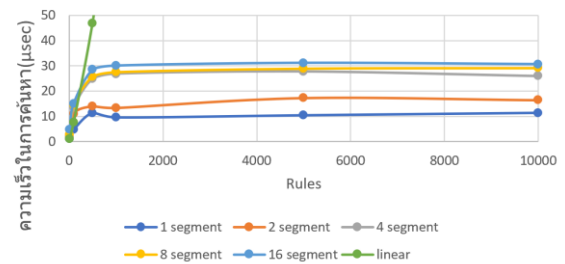
จากผลการทดลอง 4.1 Bitmap InterSection Lookup มีความถูกต้อง 100% เมื่อเปรียบเทียบกับ Linear search จึงสรุปว่า สามารถนำ Bitmap InterSection Lookup ใช้งานได้

**4.2.1 เวลาเฉลี่ยในการค้นหา**



รูปที่ 3: กราฟแสดงเวลาเฉลี่ยในการค้นหา

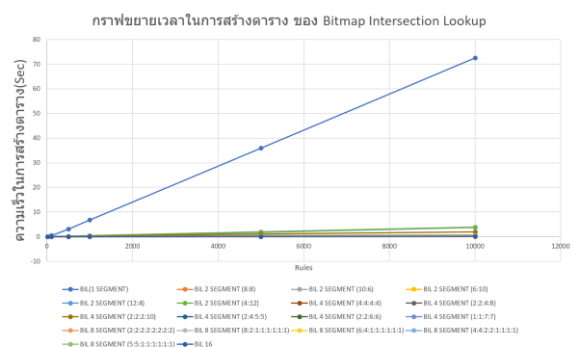
**ภาพขยาย เวลาเฉลี่ยในการค้นหา ของ Bitmap Intersection lookup เปรียบเทียบกับ Linear Search**



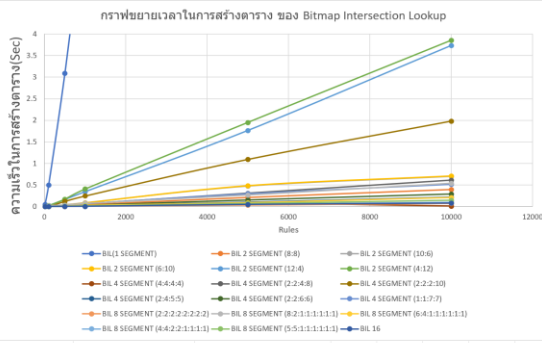
รูปที่ 4: ภาพขยายกราฟเวลาเฉลี่ยในการค้นหา

จากผลการทดลอง 4.2.1 สรุปได้ว่า Linear search ใช้เวลาในการค้นหามากที่สุดยิ่ง rules มากก็ยิ่งใช้เวลาเยอะ แต่ BIL นั้นจะใช้นเวลาน้อยกว่าและถึงจุดๆ หนึ่งเวลาจะคงที่ โดยความเร็ว นั้นจะเร็วที่สุดเมื่อมี 1 Segment และจะช้าลงเมื่อมีการแบ่ง Segment เพิ่ม เพราะการค้นหาตั้งแต่ 2 Segment ขึ้นไปต้องนำค่ามา AND กันจึงจะได้คำตอบเวลาเลยช้ากว่า 1 Segment มาก ขึ้นเรื่อยๆตาม Segment ที่ต้อง AND กัน

**4.2.2 เวลาในการสร้างตาราง**



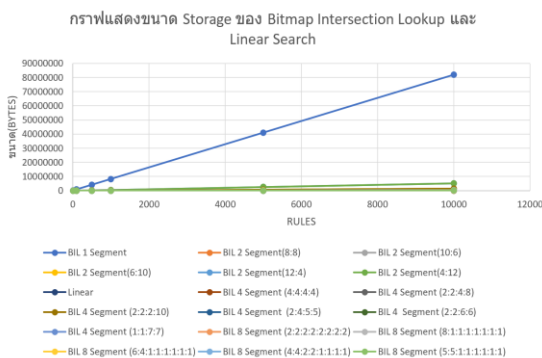
รูปที่ 5: กราฟแสดงเวลาในการสร้างตาราง



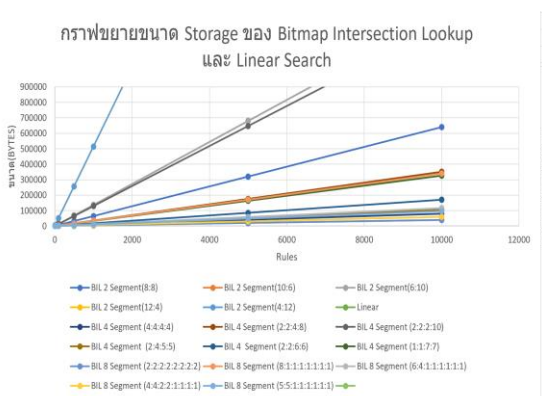
รูปที่ 6: กราฟขยายของ เวลาในการสร้างตาราง

จากผลการทดลอง 4.2.2 จะเห็นว่า BIL 1 Segment จะใช้เวลาเยอะสุดเพราะสร้างตารางมี  $2^{16} = 65,536$  Index หรือ rows และ 16 Segment จะเร็วสุด เพราะการแบ่ง Segment ช่วยลด Index จึงทำให้เวลาในการทำงานลดลง เช่น 4 Segment แบ่ง แบบ (4:4:4:4) สร้างตารางที่มี  $(2^4)^4$  ตาราง = 64 Index หรือ rows จะให้ว่ามี 16 Bits เหมือนกันแต่ 1 Segment จะใช้เยอะกว่า 4 Segment มาก และยิ่ง Rules มาก เวลาใช้จะมากตามไปด้วย

4.3 ความจุ



รูปที่ 7: กราฟแสดงขนาดความจุของ BIL และ Linear Search



รูปที่ 8: กราฟขยายขนาดความจุของ BIL และ Linear Search

จากผลการทดลอง 4.3 จากการคำนวณและเก็บค่าความจุ BIL 1 Segment ก็จะมีขนาดมากที่สุดและ 16 Segment จะมีขนาดน้อยสุด เหตุผลก็เหมือนกับเวลาสร้างตาราง ยิ่ง Index และ Rules มาก ก็ใช้ความจุมากตามไปด้วย ดังนั้นยิ่งแบ่ง Segment มากเท่าไรความจุที่ใช้ก็จะน้อยลงเท่านั้น สรุปได้ว่าจำนวน Segment กับ Rules มีผลต่อขนาดความจุ

5. สรุปผล

จากผลการทดลองทั้งหมดสรุปได้ว่า BIL สามารถใช้ได้เพราะผลลัพธ์ถูกต้อง 100% เมื่อเทียบกับ Linear Search แต่จะเหมาะกับการทำงานที่มากกว่า 1 ครั้งขึ้นไป เพราะการทำงานครั้งแรกจะต้องสร้างตารางทำให้ช้ากว่า Linear แต่หลังจากนั้น จะเร็วกว่าเพราะเก็บค่าตารางไว้แล้วเพราะจะทำการดึงค่าจากตารางมาเลย ถ้ามีหลาย Segment (ตาราง) จะนำค่ามา AND กันเวลาอาจเพิ่มขึ้นเล็กน้อย แต่ Linear Search จะตรวจสอบไปที่ละค่าในการค้นหาที่ Rules มีจำนวนมาก เวลาที่ใช้ยิ่งมากตาม แต่กลับกัน BIL ที่สร้างตารางแล้ว จะดึงค่าจากตารางมาความเร็วจึงไม่ขึ้นอยู่กับจำนวน Rules ไม่ต้องทำการ Update Table เหมือนขั้นตอนการสร้างตารางจึงมีความเร็วมากกว่า Linear Search

ถ้าจะนำไปใช้งานแนะนำให้ดูที่ความจุเป็นอันดับแรก เพราะความจุมากเวลาที่ใช้ในการค้นหาจะน้อย ความจุน้อยเวลาที่ใช้ในการค้นหาจะมาก โดยจะกำหนดความจุที่ต้องการ หรือเหมาะสมกับขนาดความจำของอุปกรณ์ ถ้าความจำที่ใช้ในการทำ BIL มีขนาดมากกว่าความจำที่กำหนดให้ทำการแบ่ง Segment ให้เหมาะสม

เอกสารอ้างอิง

[1] A. Khunkitti, N. Promrit. "Bitmap Intersection Lookup (BIL) : A Packet Classification 's Algorithm with Rules Updating." Available Online at <https://www.koreascience.or.kr/article /CFKO20053 3239340120.pdf>, accessed on 6 July 2021.

[2] "Linear search." Available Online at <http://cs.uok.edu.in/ Files/7 9 7 5 5 f0 7 - 9 5 5 0 - 4 aeb-bd6 f->

5 d8 0 2 d5 6 b4 6 d/ Custom/unit4 a\_autumn\_data%20structures.pdf, accessed on 10 July 2021.

- [3] “การค้นหาข้อมูลแบบไบนารี (Binary Search).” [online]. available : <http://agritech.pcu.ac.th/new/page/e-learningdata/9.5 Binary.php>, สืบค้นเมื่อวันที่ 12 กรกฎาคม 2564.
- [4] “ระบบเครือข่ายคอมพิวเตอร์.” [online].available :<https://sites.google.com/site/computernetwork2470/bth-thi-1-rabb-kherux-khay-khxmphiwtextx>, สืบค้นเมื่อวันที่ 10 กรกฎาคม 2564.
- [5] รวมภาพขยายของโครงงาน : [online].available: <https://docs.google.com/document/d/12Dn0BBYaXt8gKb2Mp7IIDCwxuxzSm-xEbJPALbES140/edit?usp=sharing> , สืบค้นเมื่อวันที่ 11 มีนาคม 2565.

# ระบบธุรกิจอัจฉริยะสำหรับการพยากรณ์ราคาสินค้าเกษตร

## Business Intelligence System for Price Forecasting of Agricultural Product

พรพรรณ เจริญศิลป์ (Phonpan Charoensin)<sup>1</sup> ศักดิ์ชาย ตั้งวรรณวิทย์ (Sakchai Tangwannawit)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858368@email.kmutnb.ac.th, <sup>2</sup>sakchai.t@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้ทำการพัฒนาระบบธุรกิจอัจฉริยะเพื่อสนับสนุนการพยากรณ์ราคาสินค้าเกษตร โดยมีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์ราคาสินค้าเกษตร ซึ่งได้นำข้อมูลราคาสินค้าเกษตร ศึกษาเปรียบเทียบประสิทธิภาพแบบจำลองในการพยากรณ์ ใช้โปรแกรม RapidMiner Studio โดยเป็นการเปรียบเทียบอัลกอริทึมจำนวน 3 อัลกอริทึม ได้แก่ 1) อัลกอริทึมการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) 2) อัลกอริทึมโครงข่ายประสาทเทียม (Neural Network) 3) อัลกอริทึมการเรียนรู้เชิงลึก (Deep Learning) โดยศึกษาจากราคาสินค้าเกษตรจำนวน 6 ชนิด ระหว่างปี พ.ศ.2559-2564 เปรียบเทียบแบบจำลองด้วยค่าเฉลี่ยของความผิดพลาด RMSE (Root Mean Square Error: RMSE) ที่ต่ำที่สุด จากผลวิจัยพบว่าการพยากรณ์ราคาสินค้าเกษตรประเภทผักที่มีอายุการเพาะปลูก 45-60 วัน ประกอบด้วย ผักกวางตุ้ง แดงกวาง มะเขือเจ้าพระยา มะระจีน กะนํ้า ผักกาดหอม อัลกอริทึมการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) มีความแม่นยำมากที่สุด เนื่องจากมีค่าเฉลี่ยของความผิดพลาด RMSE (Root Mean Square Error: RMSE) ที่ต่ำที่สุด

**คำสำคัญ:** ราคาสินค้าเกษตร การถดถอยเชิงเส้น การพยากรณ์ค่าเฉลี่ยของความผิดพลาด

### Abstract

This research presents the development of a business intelligence system to support agricultural price forecasting. The research's objective was to compare models for forecasting the price of

agriculture using RapidMiner studio. The methodology for analyzing data uses three algorithms consisting of 1) linear regression algorithm, 2) neural network algorithm, and 3) deep learning algorithm. The data was collected between 2016 and 2021 from six agricultures and evaluated the models by root mean square error (RMSE). The result shows that the forecast of agricultures with 45-60 days of cultivation, consisting of Chinese cabbage, cucumber, eggplant, bitter melon, kale, and lettuce found that the linear regression algorithm has the highest performance because RMSE was at the lowest. The research result can be applied to forecast the prices of other agricultural products.

**Keywords:** Agricultural price, Linear regression, Neural Network, Deep Learning.

### 1. บทนำ

ประเทศไทยเป็นประเทศเกษตรกรรมมีการประกอบอาชีพเกษตรกรรมมากกว่าครึ่งประเทศภาคเกษตรนับว่ามีความสำคัญต่อเศรษฐกิจและสังคมของประเทศไทยอย่างมาก เพราะมีการจ้างงานสูงถึงกว่าร้อยละ 30 ของกำลังแรงงานทั้งประเทศ ครอบคลุมถึง 6.4 ล้านครัวเรือนและที่ดินทำการเกษตรครอบคลุมถึงร้อยละ 40 ของพื้นที่ทั่วประเทศแต่ภาคเกษตรกลับมีส่วนในมูลค่าผลิตภัณฑ์มวลรวมภายในประเทศเพียงร้อยละ 10 มีอัตราการเติบโตช้า และยังเติบโตในอัตราที่ช้ากว่าประเทศผู้ผลิตสินค้าเกษตรชั้นนำของโลก หรือแม้แต่ประเทศเพื่อนบ้านในเอเชียหลายประเทศ [1] ดังนั้น ราคาสินค้าเกษตรนอกจาก

จะเป็นปัจจัยหนึ่งในการกำหนดค่าใช้จ่ายในการบริโภค และต้นทุนการผลิตสินค้าต่อเนื่องในธุรกิจการเกษตรแล้ว ยังเป็นปัจจัยสำคัญในการกำหนดรายได้ของประชากร ส่วนใหญ่ของประเทศอีกด้วย หากเราสามารถทำการพยากรณ์ราคาสินค้าเกษตรหลายชนิดในประเทศที่จะเกิดขึ้นในอนาคตได้อย่างแม่นยำจะส่งผลให้ผู้ผลิตและเกษตรกรสามารถวางแผนการผลิตได้อย่างเหมาะสม [2] สามารถช่วยเกษตรกรตอบปัญหาที่สำคัญในการดำเนินการผลิตได้ว่า จะผลิตสินค้าอะไร จึงจะได้รับผลประโยชน์เป็นที่น่าพอใจ

ในปัจจุบันมีการนำเทคโนโลยีเข้ามามีบทบาทในชีวิต จึงต้องมีการนำเทคโนโลยีมาประยุกต์ใช้ให้เหมาะสม การพยากรณ์เป็นเครื่องมือที่ช่วยให้ทราบถึงข้อมูลที่จะเกิดขึ้น โดยถ้าค่าพยากรณ์ที่ได้มีความแม่นยำสูงจะช่วยให้ผู้วางแผนสามารถดำเนินการได้อย่างมีประสิทธิภาพ โดยมีงานวิจัยในอดีตที่มีการวิจัยในการพยากรณ์ต่างๆ เช่น ระบบพยากรณ์ราคาสินค้าเกษตรอัตโนมัติโดยใช้วิธีทางเศรษฐมิติ [3] แบบจำลองสำหรับการพยากรณ์ยอดขายการจำหน่ายสินค้าด้วยเทคนิคการวิเคราะห์อนุกรมเวลา [4] พยากรณ์ยอดขายปลีกแก่สรถยนต์ ด้วยวิธีการวิเคราะห์อนุกรมเวลาและโครงข่ายประสาทเทียม [5] แบบจำลองการพยากรณ์ราคามันสำปะหลังโดยใช้โครงข่ายประสาทเทียมแบบหลายชั้น [6] เมื่อเวลาผ่านไปมีปัจจัยต่าง ๆ ที่มา มีผลกระทบต่อการค้าจริงชีวิต เช่น ประการ เช่น สภาพดินฟ้าอากาศ ฤดูกาล การเกิดโรคระบาด อุทกภัย ภัยแล้ง สภาพเศรษฐกิจ จำนวนประชากร ความนิยม เป็นต้น ซึ่งอาจจะปัจจัยที่มีผลกระทบกับราคาสินค้าเกษตรที่จะเกิดขึ้นได้

จากการศึกษาข้างต้น จึงทำให้ผู้วิจัยมีแนวคิดในการศึกษาการสร้างแบบพยากรณ์และคัดเลือกตัวแบบพยากรณ์ที่เหมาะสมที่สุดสำหรับการพยากรณ์ราคาสินค้าเกษตร รวมถึงปัจจัยต่าง ๆ มาวิเคราะห์ เพิ่มเติม จากในอดีต เพื่อให้ผู้ที่สนใจในเรื่องการผลิตสินค้าเกษตรสามารถทราบผลหรือการพยากรณ์ราคาสินค้าเกษตร เพื่อให้สามารถวางแผนการผลิตสินค้าเกษตรได้ และต่อ

ขอในการพัฒนาเครื่องมือสำหรับทำการวิเคราะห์เพื่อช่วยสนับสนุนการตัดสินใจในอนาคตได้

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การเกษตร

เป็นการทำงานเพื่อควบคุมธรรมชาติในอันที่จะผลิตพืชและสัตว์ให้ได้ ตามความต้องการของมนุษย์ โดยอาศัยการเจริญเติบโตของพืชและสัตว์เป็นพื้นฐาน มีมนุษย์เป็นผู้ควบคุมดำเนินการอย่างมีระบบแบบแผน มีการวางแผนปฏิบัติงานล่วงหน้า คิดคำนวณรายได้รายจ่ายในการดำเนินการทั้งหมด ดังนั้นการเกษตรจึงมีขอบเขตครอบคลุมการปลูกพืช การเลี้ยงสัตว์และการประมง

### 2.2 เหมืองข้อมูล (Data Mining)

เป็นกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหารในด้านวิทยาศาสตร์และการแพทย์ในทั้งในด้านเศรษฐกิจและสังคมการทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บตีความหมายข้อมูลจากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล

### 2.3 การวิเคราะห์การถดถอย (Linear Regression)

คือวิธีการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ เทคนิคการวิเคราะห์การถดถอยนี้ อาศัยความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรเพื่อใช้ในการทำนายลักษณะของสมการ Linear Regression แสดงให้เห็นถึงค่าความสัมพันธ์ของตัวแปรตามกับตัวแปรอิสระในลักษณะของข้อมูลเชิงปริมาณสามารถเขียนเป็น สมการที่ 1

$$Y = a + bx_1 + Cx_2 + \dots \quad (1)$$

### 2.4 โครงข่ายประสาทเทียม (Neural Network)

เทคนิคการใช้โมเดลคณิตศาสตร์ที่จำลองการทำงานของเซลล์ประสาทของเซลล์สมองมนุษย์ ซึ่งประกอบขึ้นจากแบบจำลองของ โหนดจำนวนมากมาเชื่อมต่อกันเป็น

โครงข่าย บางโหนดเชื่อมต่อกับสิ่งแวดล้อมภายนอก สามารถทำหน้าที่เป็นหน่วย Input หรือ Output ได้ แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) และจุดประสานประสาท (Synapses)

### 2.5 การเรียนรู้เชิงลึก (Deep Learning)

Deep Learning คือชุดคำสั่ง (Algorithm) ที่ถูกสร้างขึ้นมาเพื่อการเรียนรู้ของเครื่องจักรหรือเครื่องคอมพิวเตอร์ โดยชุดคำสั่งนี้จะทำให้ตัวเครื่องจักรสามารถประมวลผลข้อมูลจำนวนมากด้วยการจำลองเครือข่ายประสาทแบบเดียวกับในสมองของมนุษย์เป็นวิธีการหนึ่งของการเรียนรู้ของเครื่องที่พยายามเรียนรู้วิธีการแทนข้อมูลอย่างมีประสิทธิภาพเช่น รูปภาพหนึ่งสามารถแทนได้เป็นเวกเตอร์ของความสว่างต่อจุดพิกเซล หรือมองในระดับสูงขึ้นเป็นเซตของขอบของวัตถุต่างๆ หรือมองว่าเป็นพื้นที่ของรูปร่างใดๆ ก็ได้ การแทนความหมายดังกล่าวจะทำให้การเรียนรู้ที่จะทำงานต่างๆ ทำได้ง่ายขึ้น ไม่ว่าจะเป็นการเรียนรู้จดจำ ใบหน้าหรือการรู้จำการแสดงออกทางสีหน้า

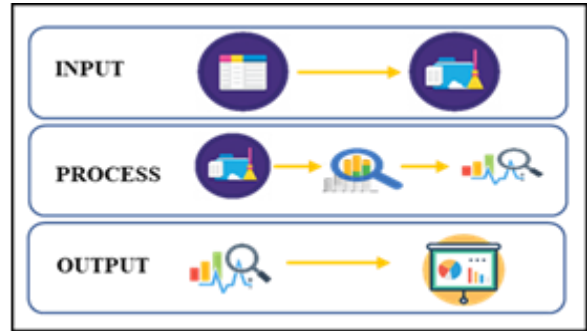
### 2.6 งานวิจัยที่เกี่ยวข้อง

พัชรี ศรีสุวรรณ [7] ได้ทำการศึกษาวิจัยเรื่อง การพยากรณ์ราคาเหล็กในโครงการก่อสร้าง โดยใช้โครงข่ายประสาทเทียมการเรียนรู้เชิงลึก ผลการวิจัยพบว่า จากการพัฒนาแบบจำลอง เปรียบเทียบและประเมิน การพยากรณ์ดัชนีราคาเหล็กด้วยเทคนิคโครงข่ายประสาทเทียมและโครงข่ายประสาทเทียมการเรียนรู้เชิงลึก (Deep Learning) มีผลการพยากรณ์ที่มีความคลาดเคลื่อนน้อยกว่าเทคนิคโครงข่ายประสาทเทียม (Neural Net)

ณัฐนารี สุขเสกสรรค์ และคณะ[8] ได้การพยากรณ์ปริมาณอาหารไก่โดยการประยุกต์ใช้สมการถดถอยเชิงเส้น กรณีสึกษา ฟาร์มไก่ จังหวัดปราจีนบุรี ผลการดำเนินงานพบว่า ปัจจัยที่มีผลต่อปริมาณการกินของไก่เนื้อคืออายุของไก่ จึงนำปัจจัยที่ได้มาหาสมการ ถดถอยเชิง ซึ่งมิต่ำ  $R^2$  มากกว่า 95% แสดงว่า สมการดังกล่าวสามารถใช้พยากรณ์ได้ ก่อนข้างแม่นยำ

### 3. วิธีดำเนินงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบธุรกิจอัจฉริยะ เพื่อสนับสนุนการพยากรณ์ราคาสินค้าเกษตร ได้ กำหนดแนวทางในการวิจัยโดยมีรายละเอียดและขั้นตอนการดำเนินงานแสดงดังภาพที่ 1



ภาพที่ 1: กระบวนการวิจัย

จากภาพที่ 1 แสดงกรอบแนวคิดของการวิจัย กระบวนการทำงานแบ่งออกเป็น 3 ส่วน ส่วนที่ 1 เป็นการรวบรวมข้อมูลและเตรียมข้อมูลก่อนประมวลผล เริ่มต้นเก็บรวบรวมข้อมูลจากเว็บไซต์กรมการค้าภายใน และกรมอุตุนิยมวิทยา และรายงานราคาน้ำมัน ส่วนที่ 2 การสร้างแบบจำลองและทดสอบประสิทธิภาพ เลือกค่าเฉลี่ยของความผิดพลาดน้อยที่สุด ส่วนที่ 3 เป็นการนำเสนอการพยากรณ์

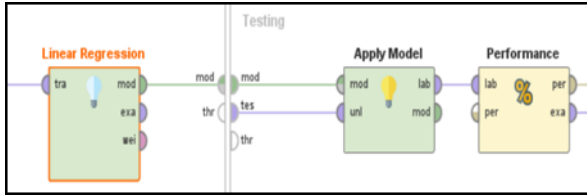
#### 3.1 การรวบรวมข้อมูลและเตรียมข้อมูล

ในการศึกษาค้นคว้าได้ทำการเก็บรวบรวมข้อมูลราคาสินค้าเกษตรรายวัน ในสินค้าประเภทผักที่มีอายุการเพาะปลูก 45-60 วัน จากเว็บไซต์กรมการค้าภายใน ตั้งแต่ พ.ศ.2559-2564 ข้อมูลในการพยากรณ์ประกอบด้วย ข้อมูลปริมาณ น้ำฝน และข้อมูลอุณหภูมิจากกรมอุตุนิยมวิทยา ข้อมูลราคาน้ำมัน โดยใช้โปรแกรม Microsoft Excel ในการจัดเก็บข้อมูลตรวจสอบความสมบูรณ์ของข้อมูล หากข้อมูลมีค่าว่างจะทำการตัดรายการข้อมูลนั้นออกไปซึ่งถือว่าเป็นลักษณะของข้อมูลที่นำเข้ามาวิเคราะห์แล้วอาจจะได้ค่าพยากรณ์ที่ไม่ถูกต้อง จากนั้นปรับเปลี่ยนรูปแบบและแปลงข้อมูลเพื่อให้ข้อมูลเหมาะสมกับแบบจำลองที่นำมาทดสอบ

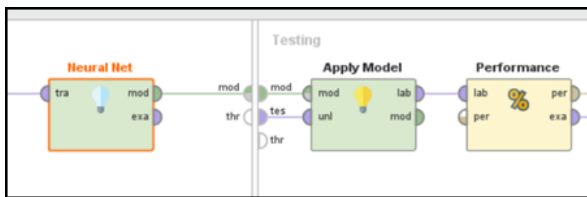
#### 3.2 การสร้างแบบจำลองหรือโมเดล

นำข้อมูลมาสร้างแบบจำลอง ด้วยโปรแกรม RapidMiner Studio โดยเป็นการเปรียบเทียบ อัลกอริทึม

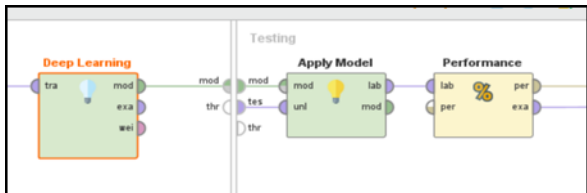
จำนวน 3 อัลกอริทึม ได้แก่ 1) อัลกอริทึมการวิเคราะห์การถดถอยเชิงเส้น 2) อัลกอริทึมโครงข่ายประสาทเทียม 3) อัลกอริทึมการเรียนรู้เชิงลึก เพื่อหาประสิทธิภาพและหาค่าเฉลี่ยของความผิดพลาดน้อยที่สุดของการพยากรณ์



ภาพที่ 2: การจำแนกด้วยอัลกอริทึมการถดถอยเชิงเส้น



ภาพที่ 3: การจำแนกด้วยอัลกอริทึมโครงข่ายประสาทเทียม



ภาพที่ 4: การจำแนกด้วยอัลกอริทึมการเรียนรู้เชิงลึก

### 3.3 การเปรียบเทียบประสิทธิภาพและเลือกแบบจำลอง

เมื่อได้ผลการวิเคราะห์ข้อมูลที่ต้องการแล้ว ก่อนที่จะนำผลลัพธ์ไปใช้งานจะต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่ามีประสิทธิภาพมากน้อยเพียงใด พิจารณาค่าเฉลี่ยของความผิดพลาด RMSE (Root Mean Square Error: RMSE) ที่ต่ำที่สุด

### 3.4 การทวนสอบความถูกต้องของแบบจำลองและการนำเสนอการพยากรณ์

ทวนสอบความถูกต้องของแบบจำลอง โดยนำแบบจำลองมาประเมินความสามารถในการนำไปใช้พยากรณ์ราคาสินค้าเกษตร และหาค่าเฉลี่ยของความผิดพลาด RMSE (Root Mean Square Error: RMSE) ที่ต่ำที่สุด

## 4. ผลการดำเนินงาน

หลังจากที่ได้ดำเนินการตามขั้นตอน จะได้แบบจำลองที่เหมาะสมในการพยากรณ์ราคาสินค้าเกษตร สามารถแสดงผลรายละเอียดการวิจัยดังนี้

### 4.1 ผลการรวบรวมข้อมูลและเตรียมข้อมูล

ผลการเก็บข้อมูลสินค้าประเภทผักที่มีอายุการเพาะปลูก 45-60 วัน จากเว็บไซต์กรมการค้าภายใน ตั้งแต่ พ.ศ.2559-2564 ข้อมูลปริมาณน้ำฝน และอุณหภูมิ จากกรมอุตุนิยมวิทยา ข้อมูลราคาน้ำมัน โดยใช้โปรแกรม Microsoft Excel ในการจัดเก็บข้อมูลจะได้ข้อมูลของผักแต่ละชนิดจำนวน 1,346 รายการ

### 4.2 ผลการสร้างแบบจำลองพยากรณ์

แบบจำลองถูกสร้างขึ้นโดยใช้อัลกอริทึม 3 แบบ ผลการวิจัยพบว่าค่า RMSE (Root Mean Square Error: RMSE) มีดังนี้

4.2.1 การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) จะมีค่า RMSE (Root Mean Square Error: RMSE) ของผัก แสดงดังตารางที่ 1

ตารางที่ 1: ผลประสิทธิภาพอัลกอริทึมการถดถอยเชิงเส้น

รายการ	Root mean square error
ผักกวางตุ้ง	1.734
แตงกวา	1.705
มะเขือเจ้าพระยา	1.658
มะระขี้นก	2.157
คะน้า	2.596
ผักกาดหอม	3.614

4.2.2 การวิเคราะห์อัลกอริทึมโครงข่ายประสาทเทียม (Neural Network) โดยกำหนดค่าพารามิเตอร์ learning rate อยู่ในช่วง 0 ถึง 1 ค่า momentum อยู่ในช่วง 0 ถึง 1 และค่า training cycle ที่ 200, 400, 600, 800 จากการพัฒนาแบบจำลองการพยากรณ์พบว่าแบบจำลองการพยากรณ์ที่พัฒนาขึ้น กำหนดค่า learning rate เท่ากับ 0.01 ค่า momentum เท่ากับ 0.9 และค่า training cycle เท่ากับ 200 ให้ค่า RMSE ของผักน้อยที่สุด แสดงดังตารางที่ 2

ตารางที่ 2: ผลประสิทธิภาพอัลกอริทึมโครงข่ายประสาทเทียม

รายการ	Root mean square error
ผักกวางตุ้ง	1.860
แตงกวา	4.493
มะเขือเจ้าพระยา	1.666
มะระขี้นก	2.253
คะน้า	2.616
ผักกาดหอม	3.699

**4.2.3 การวิเคราะห์อัลกอริทึมการเรียนรู้เชิงลึก (Deep Learning) จะมีค่า RMSE ของผัก แสดงดังตารางที่ 3**

**ตารางที่ 3:** ผลประสิทธิภาพอัลกอริทึมการเรียนรู้เชิงลึก

รายการ	Root mean square error
ผักกวางตุ้ง	1.966
แตงกวา	4.487
มะเขือเจ้าพระยา	1.750
มะระขี้นก	2.502
คะน้า	2.871
ผักกาดหอม	4.173

**4.3 ผลการเปรียบเทียบประสิทธิภาพและเลือกแบบจำลอง**

จากผลการวิจัยจึงนำการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) เนื่องจากได้ค่าต่ำสุด แสดงดังตารางที่ 4 และได้สมการ แสดงดังตารางที่ 5

**ตารางที่ 4:** ผลการเปรียบเทียบอัลกอริทึม

รายการ	การถดถอยเชิงเส้น	โครงข่ายประสาทเทียม	การเรียนรู้เชิงลึก
ผักกวางตุ้ง	1.734	1.860	1.966
แตงกวา	1.705	4.493	4.487
มะเขือเจ้าพระยา	1.658	1.666	1.750
มะระขี้นก	2.157	2.253	2.502
คะน้า	2.596	2.616	2.871
ผักกาดหอม	3.614	3.699	4.173

**ตารางที่ 5:** ผลแสดงสมการอัลกอริทึมการถดถอยเชิงเส้น

รายการ	สมการ
ผักกวางตุ้ง	$Y = -0.142 * \text{Diesel fuel} + 0.033 * \text{Rain} + 1.062 * \text{Wholesale Price} + 13.861$
แตงกวา	$Y = 0.160 * \text{Diesel fuel} + 0.072 * \text{temperature} + 0.969 * \text{Wholesale Price} + 5.628$
มะเขือเจ้าพระยา	$Y = -0.118 * \text{Diesel fuel} + 0.023 * \text{Rain} + 1.020 * \text{Wholesale Price} + 8.355$
มะระขี้นก	$Y = 0.122 * \text{Diesel fuel} - 0.030 * \text{Rain} + 0.0712 * \text{temperature} + 0.984 * \text{Wholesale Price} + 8.355$
คะน้า	$Y = -0.183 * \text{Diesel fuel} + 0.094 * \text{temperature} + 1.085 * \text{Wholesale Price} + 12.380$
ผักกาดหอม	$Y = -0.297 * \text{Diesel fuel} + 1.028 * \text{Wholesale Price} + 18.980$

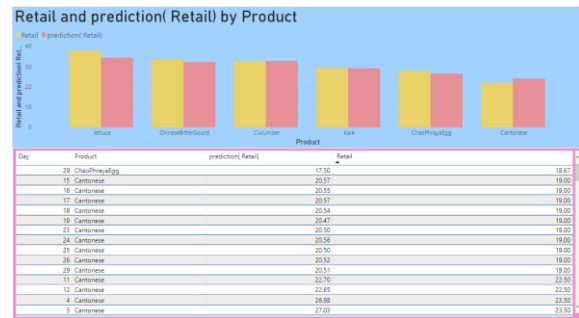
**4.4 ผลการทวนสอบความถูกต้องของแบบจำลองและการนำเสนอการพยากรณ์**

จากการทวนสอบความแม่นยำ ของแบบจำลองการพยากรณ์ โดยใช้ข้อมูลราคาสินค้าเกษตรในปีพ.ศ. 2564 เป็นข้อมูลชุดทดสอบ เปรียบเทียบกับค่าพยากรณ์ ได้ผลการทดสอบ ดังตารางที่ 6

**ตารางที่ 6:** ผลวัดประสิทธิภาพการทวนสอบความแม่นยำของการพยากรณ์

รายการ	Root mean square error
ผักกวางตุ้ง	1.325
แตงกวา	1.353
มะเขือเจ้าพระยา	1.235
มะระขี้นก	0.874
คะน้า	1.721
ผักกาดหอม	2.809

เมื่อได้แบบจำลองที่เหมาะสมแล้วผู้วิจัยจึงทำการนำค่าพยากรณ์มาเปรียบเทียบและแสดงผลกับราคาสินค้าและนำมาจัดทำระบบธุรกิจอัจฉริยะด้วยโปรแกรม Microsoft Power BI Desktop โดยมีตัวอย่างดังภาพที่ 5 ภาพที่ 6 และภาพที่ 7

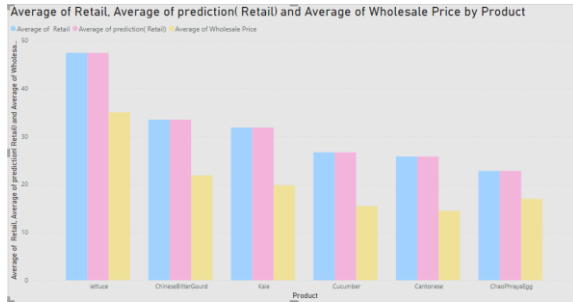


**ภาพที่ 5:** รายงานแสดงค่าพยากรณ์



**ภาพที่ 6:** รายงานแสดงค่าพยากรณ์





ภาพที่ 7: รายงานแสดงค่าพยากรณ์

## 5. สรุปผลและอภิปรายและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อทำการศึกษาคัดเลือกโดยเป็นการเปรียบเทียบ อัลกอริทึมจำนวน 3 อัลกอริทึม ได้แก่ 1) อัลกอริทึมการวิเคราะห์การถดถอยเชิงเส้น 2) อัลกอริทึมโครงข่ายประสาทเทียม 3) อัลกอริทึมการเรียนรู้เชิงลึก โดยศึกษาจากราคาสินค้าเกษตรจำนวน 6 ชนิด ระหว่างปี พ.ศ.2559-2564 เปรียบเทียบแบบจำลองด้วยค่าเฉลี่ยของความผิดพลาด RMSE (Root Mean Square. Error: RMSE) ที่ต่ำที่สุด จากผลวิจัยพบว่าพยากรณ์ราคาสินค้าเกษตร ประเภทผักที่มีอายุการเพาะปลูก 45-60 วัน ประกอบด้วย ผักกวางตุ้ง แดงกวาง มะเขือเจ้าพระยา มะระจีน กะน้า ผักกาดหอม อัลกอริทึมการวิเคราะห์การถดถอยเชิงเส้น มีความแม่นยำมากที่สุด เนื่องจากมีค่าเฉลี่ยของความผิดพลาด RMSE (Root Mean Square. Error: RMSE) ที่ต่ำที่สุด ดังนั้นการวิเคราะห์การถดถอยเชิงเส้น ให้ประสิทธิภาพและค่าเฉลี่ยของความผิดพลาดการพยากรณ์ดีกว่าอัลกอริทึมโครงข่ายประสาทเทียม และอัลกอริทึมการเรียนรู้เชิงลึก ผลการวิจัยที่ได้ศึกษาค้นคว้านี้ สามารถนำไปเป็นแนวทางเพื่อให้สามารถวางแผนการผลิตสินค้าเกษตรได้ และต่อยอดในการพัฒนาเครื่องมือสำหรับทำการวิเคราะห์และช่วยสนับสนุนการตัดสินใจในอนาคตได้

อย่างไรก็ตามสินค้านี้เป็นผักที่มีอายุการเก็บเกี่ยวระยะสั้นและ ระยะยาวในการวิจัยในอนาคต อาจจะต้องปรับปัจจัยที่นำมาวิจัยร่วม เช่น ราคาอุปกรณ์ ราคาเกษตร รวมทั้งปัจจัยที่มีผลกระทบต่อการค้าจริงชีวิต เช่น ภาวะโรคระบาด เป็นต้น

## เอกสารอ้างอิง

- [1] ภูมิทัศน์ภาคเกษตรไทย จะพลิกโฉมอย่างไรสู่การพัฒนาที่ยั่งยืน พ.ศ. 2560. [ออนไลน์] จาก [https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article\\_26Sep2019.aspx](https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article_26Sep2019.aspx)
- [2] กัญญาวิริ คำพิมาย. (2563). การพัฒนาธุรกิจอัจฉริยะเพื่อสนับสนุนการพยากรณ์ยอดขายสินค้าด้วยเทคนิคอนุกรมเวลา. ปัญหาพิเศษ. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, วิทยาศาสตร์มหาบัณฑิต.
- [3] กมลรัตน์ พ่วงแพ. (2547). ระบบพยากรณ์ราคาสินค้าเกษตรอัตโนมัติโดยใช้วิธีทางเศรษฐมิติ. วิทยานิพนธ์. จุฬาลงกรณ์มหาวิทยาลัย, วิทยาศาสตร์มหาบัณฑิต.
- [4] ฎิยาภรณ์ ปั้นดี. (2561). แบบจำลองสำหรับการพยากรณ์ยอดขายการจำหน่ายสินค้าด้วยเทคนิคการวิเคราะห์อนุกรมเวลา. ปัญหาพิเศษ. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, วิทยาศาสตร์มหาบัณฑิต.
- [5] ไววิทย์ พานิชย์ศร. (2560). พยากรณ์ยอดขายปลีกแก่สรณนต์ ด้วยวิธีการวิเคราะห์อนุกรมเวลาและโครงข่ายประสาทเทียม. ปัญหาพิเศษ. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, วิทยาศาสตร์มหาบัณฑิต.
- [6] ศุภศิ บุญรอดและกรวัฒน์ พลเยี่ยม. (2560). แบบจำลองการพยากรณ์ราคามันสำปะหลังโดยใช้โครงข่ายประสาทเทียมแบบหลายชั้น. ปัญหาพิเศษ. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, วิทยาศาสตร์มหาบัณฑิต.
- [7] พัชร ศรีสุวรรณ. (2560). การพยากรณ์ราคาเหล็กในโครงการก่อสร้าง โดยใช้โครงข่ายประสาทเทียมการเรียนรู้เชิงลึก. ปัญหาพิเศษ. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, วิทยาศาสตร์มหาบัณฑิต.
- [8] ฉัฐนารี สุขเสกสรร และคณะ “การพยากรณ์ปริมาณอาหารไก่โดยการประยุกต์ใช้สมการถดถอยเชิงเส้น กรณีศึกษาฟาร์มไก่ จังหวัดปราจีนบุรี” *วารสารวิศวกรรมธรรมชาติ ปีที่ 5 ฉบับที่ 1*
- [9] Andualem Yihun1, Manzoor Ahmed Kirmani, Meseret Molla (2020). “Characterization of Marketing and Management System of indigenous Chicken Ecotypes in Awi Zone, Ethiopia”. *International Journal of Agricultural Economics* 2020, 5(6): 264-278
- [10] Godlove Shu, Jules René Minkoua Nzie , Ernest L. Molua (2018). “Optimal Cropping Sequence in Pluriactive Non-specialised Vegetable Farms in the Northwest Region of Cameroon”, *International Journal of Agricultural Economics*: 112-117.

## แชปปี : แชทบอทผู้ช่วยตามอารมณ์ SAPPY : Emotion-Based Assistant Chatbot

ณัฐพัชร กองทอง (Natthaphat Kongthong)<sup>1</sup> วชรวรรษ เกษบรรเทิง (Vacharavat Kesbuntherng)<sup>1</sup>

และ สุพัฒน์ดา โชติพันธ์ (Supannada Chotipant)<sup>1</sup>

<sup>1</sup>คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
61070051@kmitl.ac.th 61070192@kmitl.ac.th supannada@it.kmitl.ac.th

### บทคัดย่อ

อารมณ์ความรู้สึกของมนุษย์เป็นปัจจัยสำคัญที่ส่งผลต่อการดำเนินชีวิตประจำวัน ไม่ว่าจะเป็นการทำงาน การเรียน หรือความสัมพันธ์ และยังส่งผลต่อสุขภาพอีกด้วย การได้พูดคุยหรือระบายความรู้สึกออกมาเป็นอีกวิธีหนึ่ง ที่ช่วยให้คลายเครียดได้ แชทบอทนับเป็นทางเลือกหนึ่งที่ผู้คนเลือกที่จะมาพูดคุยหรือระบายความรู้สึก ในปัจจุบันมีแอปพลิเคชันแชทบอทเกิดขึ้นมากมายที่สามารถเป็นเพื่อนคุย เตือนนัดหมายสำคัญ และบอกข่าวสารได้ แต่ยังไม่มียแอปพลิเคชันใดที่สามารถวัดค่าความรู้สึกและพูดคุยตามความรู้สึกของผู้ใช้ได้ ในงานวิจัยนี้ได้นำเสนอแอปพลิเคชันแชทบอทแชปปีที่สามารถรับความรู้สึกของผู้ใช้ และเปลี่ยนรูปแบบการสนทนาเพื่อความเหมาะสมกับอารมณ์ของผู้ใช้ โดยการตรวจจับความรู้สึก ผู้ใช้จำเป็นต้องสวมใส่สมาร์ทวอตช์เพื่อตรวจจับอัตราการเต้นของหัวใจ จากนั้นแอปพลิเคชันจะจำแนกความรู้สึกของผู้ใช้ และปรับรูปแบบการสนทนาให้เหมาะสมกับอารมณ์ของผู้ใช้ โดยงานวิจัยฉบับนี้มุ่งเน้นที่การออกแบบแอปพลิเคชันและประยุกต์ใช้อัลกอริทึมการเรียนรู้ของเครื่องเพื่อจำแนกอารมณ์ของผู้ใช้จากอัตราการเต้นของหัวใจ โดยการทดลองแสดงให้เห็นว่าการใช้โมเดล SVM ดีที่สุด สังเกตจากการจำแนกอารมณ์โกรธที่มีค่าอัตราความถูกต้อง 88.28 เปอร์เซ็นต์ ซึ่งมากกว่าวิธีอื่นอย่าง Decision tree หรือ KNN

**คำสำคัญ:** อัตราการเต้นของหัวใจ ตรวจจับความรู้สึก

แชทบอท การเรียนรู้ของเครื่อง

### Abstract

Emotion is an important factor affecting people's daily life relating to their work, study, relationships, and health. Talking or expressing feelings is a way to help relieve their stress. Consequently, talking with chatbots is an option that people choose for solving this issue. At present, many chatbot applications can be people's friends, remind people's appointments, and inform the news. Unfortunately, an application that can talk based on people's feelings does not exist. This paper proposes a chatbot application, called Sappy, that can recognize a user's feelings and change the conversation to suit the user's emotion. To classify the user's emotion, the user needs to wear a smartwatch to detect his/her heart rate. This paper focuses on the proposed application design and applying machine learning algorithms to classify the user's feelings based on a given heart rate. The experiments presented that the SVM model performed the best overall accuracy in all emotions. Moreover, it can classify the anger emotion with an accuracy rate of 88.28 percent, which outperforms the other algorithms, such as Decision tree and KNN.

**Keywords:** Heart Rate, Emotion Recognition, Chatbot, Machine Learning.

### 1. บทนำ

สังคมในปัจจุบัน สามารถเกิดเหตุการณ์ที่ไม่คาดคิดกันได้ตลอดเวลา ยกตัวอย่างเหตุการณ์ปัจจุบัน คือการ

แพร่ระบาดของไวรัสโควิด 19 และปัญหาเศรษฐกิจที่ตกต่ำ อีกทั้งในปัจจุบันสังคมมีการแข่งขันที่สูงขึ้น ไม่ว่าจะเป็นการแข่งขันทางการศึกษา การแข่งขันทางอาชีพ โดยสิ่งเหล่านี้สามารถทำให้ผู้คนเกิดความเครียดได้

จากสถิติของกรมสุขภาพจิต [1] ในการประเมินสุขภาพจิตของคนไทย ตั้งแต่วันที่ 1 มกราคม พ.ศ.2563 จนถึงปัจจุบัน มีผู้ตอบแบบประเมินทั้งหมด 2,895,145 คน พบว่า มีภาวะเครียดสูงถึง 7.98 เปอร์เซนต์ และเสี่ยงซึมเศร้าถึง 9.41 เปอร์เซนต์

จากสถิตินี้ทำให้เห็นว่าความเครียดมีอิทธิพลต่อการดำเนินชีวิต ดังนั้นการบรรเทาความเครียดจะส่งผลให้ใช้ชีวิตในแต่ละวันง่ายขึ้น ซึ่งในการจัดการความรู้สึกของแต่ละคนจะแตกต่างกันออกไป บางคนทำสิ่งที่ชอบ บางคนใช้วิธีการพูดคุยบอกเล่าหรือระบายความรู้สึกด้วยวิธีต่างๆ โดยแชทบอตเป็นหนึ่งในตัวเลือกที่ดีที่สุดสำหรับคนที่อยากเล่าระบายความรู้สึกของตน โดยที่ไม่ต้องการคำแนะนำอะไร แค่ต้องการพื้นที่สำหรับระบายความรู้สึกเท่านั้น ในปัจจุบันมีแชทบอตเกิดขึ้นมากมายที่มีไว้สนทนากับผู้คน โดยแต่ละตัวมีความสามารถที่แตกต่างกัน แต่ยังไม่แชทบอตใดที่สามารถรับรู้ความรู้สึกของมนุษย์ และตอบโต้ตามความรู้สึกของผู้ใช้ได้เลย และถ้าหากแชทบอตสามารถจับความรู้สึกของผู้ใช้ในช่วงการสนทนาได้ จะทำให้การสนทนามีชีวิตชีวามากขึ้น ผู้ใช้จะรู้สึกเหมือนคุยกับเพื่อนมากกว่าคุยกับบอตที่มีบทสนทนาแบบเดิมตามที่ตั้งเอาไว้

ดังนั้นในงานวิจัยนี้จะเสนอการออกแบบแอปพลิเคชันแชปปี้ (Sappy: Emotion-Based Assistant Chatbot) ซึ่งเป็นแอปพลิเคชันแชทบอตที่สามารถสนทนาได้ตามอารมณ์ของผู้ใช้ ที่ทำงานร่วมกับอุปกรณ์สมาร์ทวอตซ์ โดยสมาร์ทวอตซ์จะเป็นเครื่องมือที่วัดค่าอัตราการเต้นของหัวใจ และนำค่านี้ให้แชปปี้คำนวณหาอารมณ์ เมื่อรับรู้อารมณ์แล้ว แชทบอตจะเปลี่ยนรูปแบบการสนทนาให้เหมาะสมกับอารมณ์ของผู้ใช้ เช่น อารมณ์ในประเภทของความสุข แชทบอตจะสนทนาในรูปแบบที่มีความเฮฮาไปกับผู้ใช้งาน และถ้าหากผู้ใช้มีอารมณ์เศร้า แชทบอตจะรับฟังมากขึ้นและให้กำลังใจผู้ใช้เพื่อช่วยบรรเทาความเศร้า

## 2. ทฤษฎีและเทคโนโลยีที่เกี่ยวข้อง

### 2.1 ทฤษฎีที่เกี่ยวข้องกับอารมณ์

#### 1. การจำแนกอารมณ์ของมนุษย์ด้วยเซนเซอร์

งานวิจัย [2] การจำแนกอารมณ์ของมนุษย์โดยการใช้อัตราการเต้นของหัวใจ และสามารถจำแนกอารมณ์ออกเป็น 4 ประเภทหลัก คือ Neutral (ปกติ), Happy (มีความสุข), Sad (เศร้า) และ Angry (โกรธ) โดยอัตราการเต้นของหัวใจที่แตกต่างกันจะช่วยแยกอารมณ์แต่ละสถานะที่แตกต่างกันได้ เช่น อารมณ์ปกติ อัตราการเต้นของหัวใจจะอยู่ระหว่าง 60 ถึง 80 bpm อารมณ์มีความสุขจะอยู่ระหว่าง 60 ถึง 140 bpm อารมณ์ความเศร้าอยู่ระหว่าง 80 ถึง 100 bpm และอารมณ์โกรธจะอยู่ระหว่าง 100 ถึง 120 bpm

#### 2. งานวิจัยโปรแกรมเล่นเพลงตามอารมณ์

งานวิจัยนี้ [3] กล่าวถึงแอปพลิเคชันเล่นเพลงที่สามารถเปลี่ยนเพลงตามอารมณ์ของผู้ใช้ โดยมีวิธีการจำแนกอารมณ์ 2 วิธี คือ การจำแนกอารมณ์ตามอัตราการเต้นหัวใจจริง การจัดประเภทอารมณ์ตามช่วง และการจำแนกอารมณ์ตามอัตราการเต้นหัวใจเฉลี่ย การจัดประเภทอารมณ์ตามค่าเฉลี่ย โดยทั้ง 2 วิธีจะใช้ค่าอัตราการเต้นของหัวใจในการจำแนกอารมณ์

### 2.2 เทคโนโลยี และอุปกรณ์ที่ใช้ในการพัฒนาระบบ

TicWatch E2 [4] เป็นสมาร์ทวอตซ์ที่สามารถวัดอัตราการเต้นหัวใจตลอด 24 ชั่วโมง ทำให้สามารถเก็บข้อมูลอัตราการเต้นหัวใจในช่วงวันได้ โดย TicWatch E2 ทำงานด้วยระบบปฏิบัติการ แวร์โอเอสของ กูเกิล ทำให้รองรับการแสดงผลการแจ้งเตือน และทำงานร่วมกับแอปพลิเคชันต่างๆ บนสมาร์ทโฟนได้อย่างยืดหยุ่น

## 3. การวิเคราะห์และออกแบบระบบ

### 3.1 การศึกษาระบบเดิม

ผู้วิจัยได้รวบรวมข้อมูลเกี่ยวกับแอปพลิเคชันแชทบอตที่มีในปัจจุบัน จำนวน 4 แอปพลิเคชัน ได้แก่

#### 3.1.1 บอทน้อย (BotNoi)

บอทน้อย [5] เป็นระบบเอไอแชทบอตที่พัฒนาขึ้นโดยคนไทย บอทน้อยเป็นแชทบอตที่คุยเก่งเหมือนเพื่อน ซึ่งแตกต่างจากสิริ หรือแชทบอตตัวอื่นที่เน้นช่วยงานมากกว่า

**3.1.2 ซิมซิมิ (SimSimi)**

ซิมซิมิ [6] เป็นแอปพลิเคชันสัญชาติเกาหลี โดยซิมซิมิสามารถโต้ตอบได้เหมือนคนจริงๆ หากคำไหนที่ผู้ใช้งานไปแล้ว แต่ซิมซิมิไม่ตอบกลับมานั้นแสดงว่าซิมซิมิไม่เข้าใจคำถามผู้ใช้ ผู้ใช้สามารถสอนให้ซิมซิมิตอบคำถามได้

**3.1.3 ฟ้ำใส เลขส่วนตัว**

ฟ้ำใส [7] คือระบบเลขส่วนตัวที่ถูกสร้างขึ้นเพื่ออำนวยความสะดวกให้แก่ผู้ใช้ในด้านต่างๆ ฟ้ำใสสามารถเป็นเพื่อนคุยยามเหงาได้ และฟ้ำใสจะมีระดับความสัมพันธ์กับผู้ใช้งานด้วย

**3.1.4 ใใส่ใจ (Psyjai)**

แชทบอตใใส่ใจ ให้บริการผ่านทางเฟสบุค โดยใช้เทคโนโลยีปัญญาประดิษฐ์ วิเคราะห์อารมณ์จากเนื้อหาการพูดคุย และมีระบบจัดการสนทนาสำหรับส่งข้อความตอบให้สอดคล้องกับอารมณ์และปัญหาที่ปัญญาประดิษฐ์วิเคราะห์หรือตรวจจับได้

ผู้วิจัยได้ทำการศึกษาระบบการทำงานของแอปพลิเคชันแชทบอตในปัจจุบันและแบ่งการทำงานเป็นหัวข้อเพื่อเปรียบเทียบการทำงานของแอปพลิเคชันที่มีและแอปพลิเคชันแชทปี่ ดังตารางที่ 1

**ตารางที่ 1:** ตารางเปรียบเทียบการทำงานของแอปพลิเคชันในปัจจุบัน

	บอทน้อย	ซิมซิมิ	ฟ้ำใส	ใใส่ใจ
ข่าวสารน่ารู้	✓		✓	✓
ไปรษณีย์ของบอตและผู้ใช้	✓			
ระดับความสัมพันธ์ระหว่างผู้ใช้และบอต			✓	
สอนบอตตอบข้อความ		✓		
เปลี่ยนรูปแบบการสนทนาตามความรู้สึก				✓

จากตารางที่ 1 จะเห็นว่าแชทบอตแต่ละตัวนั้นมีฟังก์ชันที่ไม่เหมือนกัน และบางฟังก์ชันที่เหมือนกัน แชทบอตใใส่ใจ จะเป็นแชทบอตที่เกี่ยวข้องในการช่วยบรรเทาความเครียดของผู้ใช้ และเป็นแชทบอตเพียงตัวเดียวที่สามารถเปลี่ยนรูปแบบการสนทนากับผู้ใช้ได้

**3.2 การออกแบบระบบ**

**3.2.1 ภาพรวมของระบบ**

ภาพรวมการทำงานของแอปพลิเคชันแชทปี่ประกอบไปด้วย 3 ส่วนได้แก่ 1) ส่วนรับข้อมูลอัตราการเต้นหัวใจจากนาฬิกา 2) ส่วนจำแนกอารมณ์ 3) ส่วนการเลือกแชทบอตเพื่อสนทนากับผู้ใช้ตามอารมณ์

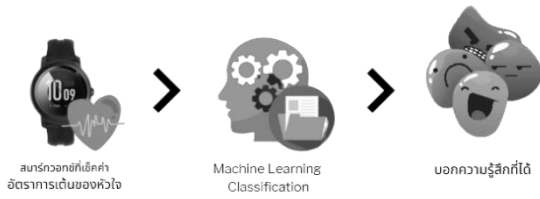
โดยการทำงาน โดยรวมของระบบดังที่แสดงในภาพที่ 1 ผู้ใช้จะต้องสวมใส่อุปกรณ์สมาร์ทวอทซ์เพื่อวัดอัตราการเต้นของหัวใจ หลังจากนั้นระบบจะนำค่าที่ได้ไปคำนวณหาอารมณ์ของผู้ใช้ เมื่อระบบได้อารมณ์ที่ถูกต้อง ระบบจะส่งค่าอารมณ์ไปยังแชทบอตเพื่อเลือกบทสนทนาที่เหมาะสมกับอารมณ์ของผู้ใช้



**ภาพที่ 1:** ภาพรวมการทำงานของระบบ

**3.2.2 การตรวจจับอารมณ์**

ในการทำงานโดยรวมของการตรวจจับอารมณ์ที่แสดงในภาพที่ 2 นาฬิกาสมาร์ทวอทซ์จะส่งค่าอัตราการเต้นของหัวใจ และนำผลลัพธ์ที่ได้ไปทำการจำแนกคลาสด้วยแมชชีนเลิร์นนิ่งจากชุดข้อมูลอัตราการเต้นของหัวใจ เพื่อให้ได้อารมณ์ของผู้ใช้ในตอนนั้น โดยอารมณ์ที่สามารถบอกได้มีทั้งหมด 4 อารมณ์ คือ อารมณ์ความสุข (Happy) อารมณ์ความเศร้า (Sad) อารมณ์ปกติ (Neutral) และอารมณ์โกรธ (Angry)



ภาพที่ 2: การตรวจจับความรู้สึก

### 3.2.3 การทำงานของแฮตบอต

ในการทำงานโดยรวมของแฮตบอตที่แสดงในภาพที่ 3 หลังจากทีระบบรู้ผลลัพธ์อารมณ์ที่ได้จากการตรวจจับอารมณ์ ระบบจะส่งผลลัพธ์ไปให้แฮตบอตเลือกบทสนทนาที่เหมาะสมกับอารมณ์ของผู้ใช้ โดยแฮตบอตจะมี 4 รูปแบบตาม 4 อารมณ์ คือ อารมณ์ปกติ อารมณ์สุข อารมณ์เศร้า และอารมณ์โกรธ สำหรับรายละเอียดของบทสนทนา ยังไม่ได้แสดงในงานวิจัยนี้เนื่องจากอยู่ในขั้นตอนการออกแบบ และต้องมีการศึกษาและกำหนดบทสนทนา



ภาพที่ 3: การทำงานแฮตบอต

### 3.2.4 การทำงานของแอปพลิเคชัน

แอปพลิเคชันจะมีการทำงานของหลายฟังก์ชัน โดยฟังก์ชันหลักของแอปพลิเคชัน คือการทำงานร่วมกับสมาร์ทวอตช์เพื่อหาค่าของอารมณ์ และการทำงานในส่วนของแฮตบอต เพื่อโต้ตอบกับผู้ใช้ นอกจากนี้ แอปพลิเคชันยังถูกออกแบบให้มีฟังก์ชันเพิ่มเติม ดังนี้

1. บัญชีผู้ใช้งาน ผู้ใช้งานสามารถบัญชีผู้ใช้และเข้าสู่ระบบได้
2. โปรไฟล์ของผู้ใช้งาน และ โปรไฟล์ของบอต ผู้ใช้งานสามารถตั้งชื่อของผู้ใช้ และชื่อของบอต รวมถึงรูปภาพโปรไฟล์ของผู้ใช้และบอต
3. การจัดการกับการตรวจจับอารมณ์รายวัน ผู้ใช้งานสามารถตั้งค่าในการตรวจจับอารมณ์รายวันได้ โดยรูปแบบจะเหมือนกับการตั้งเวลาปลุก

4. สอนบอตตอบทสนทนา ผู้ใช้งานสามารถสอนบทสนทนา คำเพิ่มเติมที่บอตไม่เข้าใจ โดยผู้ใช้จะต้องเลือกหมวดหมู่อารมณ์ให้เหมาะสมกับคำที่ต้องการจะสอนเพิ่มเติม

5. คูประวัติอารมณ์ย้อนหลัง ผู้ใช้สามารถดูอารมณ์ย้อนหลังของตนเองได้ โดยรูปแบบจะเป็นปฏิทินให้ผู้ใช้เลือกวัน เดือน ปี ที่ต้องการจะดูอารมณ์ของตนเอง

6. ข่าวสารน่ารู้ประจำวัน ผู้ใช้สามารถอ่านข่าวที่น่าสนใจในแต่ละวัน เพื่อให้ผู้ใช้ไม่พลาดกระแส และเหตุการณ์ต่างๆ

7. การตั้งค่าการแจ้งเตือน ผู้ใช้สามารถปิดการแจ้งเตือนของแฮตบอต และการแจ้งเตือนการตรวจจับอารมณ์รายวันได้

## 4. การออกแบบส่วนติดต่อผู้ใช้

### 4.1 หน้าหลักของผู้ใช้งาน

หน้าหลักของผู้ใช้งาน จะเป็นหน้าโปรไฟล์ ที่แสดงโปรไฟล์ของผู้ใช้ และโปรไฟล์ของบอต โดยข้างล่างจะมีเมนูให้ผู้ใช้เลือกฟังก์ชันอื่นที่ต้องการใช้งาน ดังแสดงในภาพที่ 4

### 4.2 หน้าต่างการแชทระหว่างผู้ใช้และบอต

หน้าต่างนี้จะแสดงหน้าแชทที่ผู้ใช้สามารถคุยกับบอตได้ จะมีช่องสำหรับพิมพ์แชทและมีปุ่มที่สามารถเลือกส่งอีโมจิได้ บอตสามารถคุยกับผู้ใช้ได้ตามอารมณ์ของผู้ใช้ขณะนั้นและจะเปลี่ยนรูปแบบการสนทนาได้ตามอารมณ์ของผู้ใช้

### 4.3 หน้าต่างแสดงผลอารมณ์และการเช็คอารมณ์

หน้าต่างนี้จะแสดงผลอารมณ์ของผู้ใช้โดยจะมีปุ่มให้กดเช็คเพื่อตรวจสอบว่าผู้ใช้มีความรู้สึกอย่างไรในขณะนั้น หน้าต่างนี้จะแสดงวันที่ เวลา และแสดงอารมณ์ของผู้ใช้ที่ได้กดเช็คล่าสุด และมีปุ่ม “ Emotion History ” เป็นหน้าต่างที่จะแสดงประวัติผลอารมณ์ในแต่ละวันของผู้ใช้ หากกดปุ่มเลือกวันที่ในปฏิทินจะแสดงผลอารมณ์และสแตตัสของผู้ใช้ในวันนั้นๆ



ภาพที่ 4: หน้าหลักของผู้ใช้งาน หน้าการแชตระหว่างผู้ใช้และบอทหน้าการแสดงผลอารมณ์และการเช็คอารมณ์

## 5. การทดลองการจำแนกอารมณ์ด้วยอัตราการเต้นของหัวใจ

### 5.1 การรวบรวมชุดข้อมูล

ทางผู้วิจัยได้ทำการทดลองการจำแนกอารมณ์ด้วยอัตราการเต้นของหัวใจ โดยมีการเก็บข้อมูลจากอาสาสมัครจำนวน 5 คน ที่มีช่วงอายุ 21 ถึง 22 ปี จำนวน 4 คน ซึ่งเป็นช่วงอายุเป้าหมายที่คาดว่าจะใช้แชตบอท และอายุ 52 ปี จำนวน 1 คน เพื่อให้ได้ข้อมูลที่หลากหลายครอบคลุมผู้ใช้ในหลายช่วงวัย โดยมีวิธีการเก็บข้อมูลดังนี้

1. ก่อนเริ่มผู้วิจัยจะถามข้อมูลส่วนตัวของอาสาสมัคร เช่น เพศ อายุ ส่วน สูง น้ำหนัก และโรคประจำตัว จากนั้นจะให้อาสาสมัครสวมสมาร์ทวอตช์เพื่อเตรียมเก็บข้อมูล

2. การเก็บข้อมูลอารมณ์ปกติ (Neutral) โดยอาสาสมัครจะได้นั่งผ่อนคลายเป็นเวลา 5 นาที ซึ่งผู้วิจัยจะจดบันทึกอัตราการเต้นของหัวใจทุก 5 วินาที หลังจากครบ 5 นาที อาสาสมัครจะได้พักเพื่อปรับอารมณ์และเตรียมเก็บอารมณ์เป้าหมายต่อไป

3. การเก็บข้อมูลอารมณ์เศร้า (Sad) และการเก็บข้อมูลอารมณ์สุข (Happy) อาสาสมัครจะได้รับชมคลิปวิดีโอที่เกี่ยวข้องกับอารมณ์เศร้าจำนวน 2 คลิป และคลิปวิดีโอที่เกี่ยวข้องกับอารมณ์สุขจำนวน 2 คลิป คลิปละ 5 นาที ซึ่งผู้วิจัยจะจดบันทึกอัตราการเต้นของหัวใจทุก 5 วินาที หลังจบ 1 คลิปจะพัก 5 นาทีเพื่อปรับอารมณ์และดำเนินการซ้ำแบบเดิมจนครบ หลังจากนั้นอาสาสมัครจะได้พักเพื่อปรับอารมณ์และเตรียมเก็บอารมณ์เป้าหมายต่อไป

4. การเก็บข้อมูลอารมณ์โกรธ (Angry) อาสาสมัครจะได้เล่นเกมที่ผู้วิจัยจัดเตรียมไว้เป็นเวลา 10 นาที ซึ่งผู้วิจัยจะจดบันทึกอัตราการเต้นของหัวใจทุก 5 วินาที

โดยข้อมูลที่ได้อาจจากการเก็บจากอาสาสมัครทั้ง 5 ทำให้ได้ข้อมูลสะท้อนอารมณ์ต่างๆ ทั้งหมด 1,813 ข้อมูล ซึ่งจะนำข้อมูลเหล่านี้ไปจำแนกอารมณ์ของผู้ใช้ด้วยการใช้อัลกอริทึมสำหรับการเรียนรู้เครื่อง

### 5.2 การจำแนกอารมณ์จากอัตราการเต้นของหัวใจ

ทางผู้วิจัยได้ประยุกต์ใช้อัลกอริทึมการเรียนรู้ของเครื่องอย่าง Decision Tree (DT), SVM และ KNN สำหรับจำแนกกลุ่มอารมณ์ของผู้ใช้ และอัตราการเต้นของหัวใจ ณ เวลานั้นๆ โดยข้อมูลที่มีทั้งหมด 1,813 ข้อมูล แบ่งข้อมูลสำหรับการสอน โมเดล 80 เปอร์เซ็นต์ คิดเป็น 1,446 ข้อมูลและแบ่งข้อมูลสำหรับการทดสอบ 20 เปอร์เซ็นต์ คิดเป็น 367 ข้อมูล

โดยฟีเจอร์ที่นำมาใช้ในการจำแนกอารมณ์มีทั้งหมด 4 ฟีเจอร์ ดังนี้

1. อายุ (Age) อายุตรงที่ได้มาจากอาสาสมัคร ที่มีตั้งแต่ช่วง 20 ถึง 52 ปี
2. ดัชนีมวลกาย (BMI) ได้มาจากการนำส่วนสูงและน้ำหนักของอาสาสมัครมาคิดคำนวณ โดยนำน้ำหนักหารด้วยส่วนสูงที่มีหน่วยเป็นเมตร ยกกำลัง 2 จะเขียนด้วยสูตรทางคณิตศาสตร์ดังแสดงใน (1)

$$BMI = \frac{\text{น้ำหนักตัว (กิโลกรัม)}}{\text{ส่วนสูง (เมตร)}^2} \quad (1)$$

3. ค่าอัตราการเต้นของหัวใจ (Heart rate) ได้มาจากการจดบันทึกค่าอัตราการเต้นของหัวใจที่สะท้อนอารมณ์ของอาสาสมัครทุกๆ 5 วินาที
4. ค่าความต่างของอัตราการเต้นของหัวใจ (Heart rate difference) เป็นค่าสัมบูรณ์ของความแตกต่างระหว่างอัตราการเต้นของหัวใจ ณ ข้อมูลอันดับที่  $n$  และข้อมูลอันดับที่  $n-1$  จะเขียนด้วยสูตรทางคณิตศาสตร์ดังแสดงใน (2)

$$\Delta H_n = |H_n - H_{n-1}| \quad (2)$$

โดย  $\Delta H_n$  แทนค่าความต่างของอัตราการเต้นของหัวใจ ณ ข้อมูลอันดับ  $n$  และ  $H_n$  แทนค่าอัตราการเต้นของหัวใจ ณ ข้อมูลอันดับ  $n$

ทั้ง 4 ฟีเจอร์ที่นำมาใช้ในการจำแนกอารมณ์นั้น ทางผู้วิจัยทำการศึกษาจากการทดลองว่ามีผลที่จะช่วยในการเพิ่มความแม่นยำให้กับการจำแนกอารมณ์มากขึ้น

**5.3 การประเมินความถูกต้องของการจำแนกอารมณ์**

ผู้วิจัยประเมินประสิทธิภาพของโมเดลจากค่าอัตราความถูกต้อง โดยการคิดค่าอัตราความถูกต้องจากสัดส่วนของตัวอย่างที่จำแนกอารมณ์ได้ถูกต้องกับจำนวนตัวอย่างที่ใช้ทดสอบทั้งหมด ดังแสดงใน (3)

$$ACCURACY = \left( \frac{\text{ตัวอย่างที่จำแนกอารมณ์ที่ถูกต้อง}}{\text{ตัวอย่างที่ใช้ทดสอบทั้งหมด}} \right) \times 100 \tag{3}$$

**5.4 ผลการทดลอง**

ข้อมูลที่ได้จากการทดลองจะถูกนำไปใช้ในการจำแนกอารมณ์ผ่าน โมเดลทั้ง 3 โมเดลได้แก่ Decision Tree (DT), SVM, KNN และได้ตารางเปรียบเทียบค่าอัตราความถูกต้องสำหรับอารมณ์ต่างๆ ดังตารางที่ 2 ถึง ตารางที่ 5 ตามลำดับ

**ตารางที่ 2:** แสดงค่าอัตราความถูกต้องของการจำแนกอารมณ์ปกติ

Features	Neutral		
	DT	SVM	KNN
Age, BMI, Heart rate	56.95%	60.22%	57.22%
Age, BMI, Heart rate, Heart rate difference	58.04%	59.95%	58.59%

จากตารางที่ 2 จะเห็นได้ว่าโมเดล SVM จะมีค่าอัตราความถูกต้องในส่วนของอารมณ์ปกติ (Neutral) มากกว่า Decision Tree และ KNN

**ตารางที่ 3:** แสดงค่าอัตราความถูกต้องของการจำแนกอารมณ์เศร้า

Features	Sad		
	DT	SVM	KNN
Age, BMI, Heart rate	87.73%	88.83%	89.37%
Age, BMI, Heart rate, Heart rate difference	85.83%	88.83%	87.74%

จากตารางที่ 3 จะเห็นได้ว่าโมเดล KNN จะมีค่าอัตราความถูกต้องในส่วนของอารมณ์เศร้า (Sad) มากกว่า DT

และ SVC หากฟีเจอร์ที่นำมาใช้ใน โมเดลมีเพียง อายุ, ดัชนีมวลกาย และค่าอัตราการเต้นของหัวใจ

**ตารางที่ 4:** แสดงค่าอัตราความถูกต้องของการจำแนกอารมณ์สุข

Features	Happy		
	DT	SVM	KNN
Age, BMI, Heart rate	78.20%	80.38%	78.75%
Age, BMI, Heart rate, Heart rate difference	77.66%	80.38%	78.20%

จากตารางที่ 4 จะเห็นได้ว่าโมเดล SVM จะมีค่าอัตราความถูกต้องในส่วนของอารมณ์สุข (Happy) มากกว่า DT และ KNN

**ตารางที่ 5:** แสดงค่าอัตราความถูกต้องของการจำแนกอารมณ์โกรธ

Features	Angry		
	Decision Tree	SVM	KNN
Age, BMI, Heart rate	85.02%	88.28%	85.56%
Age, BMI, Heart rate, Heart rate difference	84.46%	88.28%	88.01%

จากตารางที่ 5 จะเห็นได้ว่าโมเดล SVM จะมีค่าอัตราความถูกต้องในส่วนของอารมณ์โกรธ (Angry) มากกว่า DT และ KNN

**5.5 สรุปผลการทดลอง**

ในการทดลองนี้สรุปได้ว่าการใช้โมเดล SVM จะมีค่าอัตราความถูกต้องโดยรวมมากกว่าการใช้ DT และ KNN แต่สำหรับอารมณ์เศร้า KNN ให้ผลลัพธ์ที่มากกว่าค่าความต่างของอัตราการเต้นของหัวใจ ( $\Delta H_n$ ) นอกจากนี้ พบว่าฟีเจอร์  $\Delta H_n$  ไม่ส่งผลต่อความถูกต้องของการจำแนกอารมณ์มากนัก อาจเพราะ  $\Delta H_n$  เป็นค่าสัมบูรณ์ ซึ่งไม่สามารถสะท้อนได้ว่าการเต้นของหัวใจเพิ่มขึ้นหรือลดลง ณ เวลานั้นๆ ซึ่งอาจมีผลต่อความถูกต้องของการจำแนกอารมณ์ได้ อีกทั้งค่าอัตราความถูกต้องที่ได้จากอารมณ์ปกติค่อนข้างน้อยเมื่อเทียบกับการจำแนกอารมณ์อื่นๆ เนื่องจากอัตราการเต้นของหัวใจในสภาวะปกติมีความใกล้เคียงหรืออยู่ในขอบเขตของอัตราการเต้นของหัวใจในสภาวะอารมณ์ต่างๆ ซึ่งอาจมีผลต่อความถูกต้องของการจำแนกอารมณ์ได้

## 6. บทสรุป

แอปพลิเคชันแชทปี่ คือแอปพลิเคชันแชทบอตที่สามารถจำแนกอารมณ์ของผู้ใช้จากอัตราการเต้นของหัวใจ และกำหนดรูปแบบการสนทนาที่จะเกิดขึ้นระหว่างผู้ใช้และแชท

บอตให้สอดคล้องตามอารมณ์ของผู้ใช้ ซึ่งภาพรวมงานวิจัยนี้ได้นำเสนอเฉพาะการออกแบบแอปพลิเคชันและการประยุกต์ใช้อัลกอริทึมการเรียนรู้ของเครื่องเพื่อจำแนกอารมณ์ของผู้ใช้จากอัตราการเต้นของหัวใจ โดยมีการทดลองการจำแนกอารมณ์และสรุปได้ว่าการใช้โมเดล SVM ในการจำแนกอารมณ์มีความแม่นยำมากกว่าอัลกอริทึมอย่างเห็นได้ชัด โดยเฉพาะอารมณ์เศร้าและโกรธที่มีค่าความถูกต้องประมาณ 88 เปอร์เซ็นต์ แต่ถึงอย่างไรก็ตาม จำเป็นต้องมีการปรับปรุงโมเดลสำหรับการตรวจจับอารมณ์ปกติให้ถูกต้องมากขึ้น อีกทั้งต้องกำหนดบทสนทนาตามอารมณ์เพื่อให้แชทบอตทำงานได้สมบูรณ์ต่อไป

### เอกสารอ้างอิง

- [1] DMH, “Department of Mental Health”. Available: <https://checkin.dmh.go.th/> , 2563.
- [2] Quazi, M. T. Human emotion recognition using smart sensors. *Palmerston North*, pp 17-19, 2012.
- [3] Chankuptarat, K., Sriwatanaworachai, R., & Chotipant, S. Emotion-Based Music Player. In *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, pp. 1-4. IEEE, 2019.
- [4] Mobvoi, “TicWatch E2”. Available: <https://www.mobvoi.com/>. 2022.
- [5] Botnoi Consulting , “Botnoi”. Available: <https://botnoigroup.com/>. 2022.
- [6] guru sanook, “What is Sisimi”, 26 Nov 2013.
- [7] FAHSAI STUDIO CO., LTD, “Fahsai Assistant”. Available: <https://www.fahsai.in.th/>. 2022
- [8] Andreas C. Müller, Sarah Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists. *O'Reilly Media, Inc.*, 2016.



# การพยากรณ์ผลผลิตอ้อยของประเทศไทยด้วยเทคนิคการเรียนรู้ของเครื่อง

## Forecasting Sugar Cane Yield in Thailand with machine learning

ธีรวัฒน์ ศิลปรัศมี (Teerawat Sinlapratsamee)<sup>1</sup> และ กาญจนา วิริยะพันธ์ (Kanchana Viriyapant)<sup>2</sup>

ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307011858047@email.kmutnb.ac.th, <sup>2</sup>kanchana.v@itd.kmutnb.ac.th

### บทคัดย่อ

อุตสาหกรรมอ้อยและน้ำตาลทรายของไทยมีความก้าวหน้าค่อนข้างมากเมื่อเทียบกับอดีตที่ผ่านมา จึงมีการนำระบบโควตาไร่อ้อยและการจัดการผลผลิตโดยสำนักงานคณะกรรมการอ้อยและน้ำตาลทรายเป็นผู้กำหนดโควตาให้แต่ละโรงงานสามารถมีโควตาในการส่งอ้อยและน้ำตาลเท่าไร่และนำไปจัดสรรคให้ชาวไร่อ้อยในการปลูกผลผลิตและส่งอ้อยเข้าโรงงาน แต่ระบบโควตาวิธีการคำนวณปริมาณอ้อยมีความแม่นยำค่อนข้างน้อยจึงเกิดปัญหาการจัดสรรคที่ไม่ลงตัว บทความนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่ส่งผลกระทบต่อพยากรณ์ผลผลิตอ้อยและนำปัจจัยที่ศึกษาพยากรณ์ผลผลิตอ้อยงานวิจัยนี้ใช้ข้อมูลย้อนหลัง 5 ปีจำนวน 200 ตัวอย่างตั้งแต่ปี พ.ศ.2559 – 2563 ปัจจัยที่ศึกษาทั้งหมด 10 ปัจจัยประกอบด้วยปริมาณน้ำฝนเฉลี่ย, ความชื้นสัมพัทธ์, ผลผลิตอ้อยต้นต่อไร่, ปริมาณอ้อยจากการคำนวณพื้นที่, อุณหภูมิค่าสุด, อุณหภูมิเฉลี่ย, จำนวนวันที่ฝนตก, อุณหภูมิสูงสุด, พื้นที่ปลูกอ้อยและพื้นที่ที่เก็บเกี่ยวอ้อย(อ้อยบางส่วนปลูกไว้เพื่อขยายพันธุ์) การศึกษาความสัมพันธ์ของปัจจัยใช้วิธี การสุ่มป่าไม้ จากไลบรารีในภาษาไพธอน ค่าความสัมพันธ์ของปัจจัยที่ศึกษามีผลใกล้เคียงกันและทั้ง 10 ปัจจัยมีผลต่อการพยากรณ์ผลผลิต การพยากรณ์ผลผลิตใช้วิธี การวิเคราะห์การถดถอยเชิงพหุคูณ จากการผลทดสอบการพยากรณ์ตามจำนวนชุดข้อมูล พบว่าผลการเปรียบเทียบค่าความผิดพลาดเฉลี่ยน้อยที่สุดเท่ากับร้อยละ 2.05 เมื่อใช้ข้อมูลเรียนรู้ร้อยละ 90 และข้อมูลชุดทดสอบร้อยละ 10 แสดงให้เห็นว่าเมื่อใช้ข้อมูลเรียนรู้จำนวนมากขึ้นส่งผลให้ความแม่นยำในการพยากรณ์เพิ่มมากขึ้น

สรุปได้ว่าการพยากรณ์ผลผลิตอ้อยด้วยวิธี การวิเคราะห์การถดถอยเชิงพหุคูณมีค่าความผิดพลาดที่สามารถยอมรับและสามารถนำอัลกอริทึมที่ได้จากการทดลองไปใช้ในการพัฒนาระบบพยากรณ์ผลผลิตอ้อยเพื่อให้สามารถพยากรณ์ผลผลิตอ้อยที่แม่นยำและกำหนดโควตาให้แต่ละโรงงานถูกต้องมากยิ่งขึ้น

**คำสำคัญ:** การเรียนรู้ของเครื่อง, การวิเคราะห์การถดถอยเชิงพหุคูณ, การสุ่มป่าไม้

### Abstract

Thailand's sugar cane and sugar industry have developed significantly compared to the past. Although the sugarcane quota system and production management were adopted and control by the Office of the Cane and Sugar Board to determines the quota for sending cane and sugar each factory and allocating the amount of growing sugarcane for farmers to pass to industry, calculating the amount of sugar cane is low accuracy, affected inappropriate allocation. This research aims to study factors which effect sugarcane forecasting and apply these factors to develop a forecasting model used historical data from 2016 to 2020 200 sample which have 10 factors as average rainfall, relative humidity, yield, amount of sugarcane from area calculation, lowest temperature, average temperature, highest temperature, rainy days, Sugarcane mill (Some of the sugarcane is planted for propagation.) and sugarcane planting area using random forest. The result found that 10 factors are related to forecasting

*model which using multiple regression and the system performance testing from historical data show that the mean absolute percentage error is 2.05 percent. It can be concluded that the system is effective in prediction in Thailand sugar cane yield.*

**Keywords:** Machine Learning, Multiple Linear Regression, Random Forest

## 1. บทนำ

อ้อยนับว่าเป็นพืชเศรษฐกิจหลักชนิดหนึ่งที่มีความสำคัญของประเทศไทย ซึ่งใช้เป็นวัตถุดิบในอุตสาหกรรมอ้อยและน้ำตาลทรายที่มีความสำคัญต่อการพัฒนาระบบเศรษฐกิจของไทยเป็นอย่างมาก ผลผลิตอ้อยสามารถนำมาผลิตน้ำตาลและผลิตภัณฑ์อุตสาหกรรมต่อเนื่องต่าง ๆ เช่น อุตสาหกรรมอาหาร พลังงานทดแทน อ้อยเป็นพืชที่เติบโตได้ดีในภูมิอากาศร้อนและชุ่มชื้น ประเทศไทยสามารถปลูกอ้อยแล้วได้ผลผลิตดีและสามารถปลูกได้ทุกพื้นที่ยกเว้นภาคใต้เพราะภาคใต้สำนักงานคณะกรรมการอ้อยและน้ำตาลทราย กระทรวงอุตสาหกรรมได้แบ่งบริเวณการปลูกอ้อยเป็น 4 ภาค ได้แก่ ภาคเหนือ ภาคกลาง ภาคตะวันออก และภาคตะวันออกเฉียงเหนือ จากข้อมูลปีการผลิตปี 63/64 ทั้งประเทศมีพื้นที่ปลูกอ้อย 10,862,610 ไร่ แบ่งเป็นภาคตะวันออกเฉียงเหนือ 4,594,444 ไร่ ภาคกลาง 2,939,319 ไร่ ภาคเหนือ 2,668,136 ไร่ และภาคตะวันออก 660,711 ไร่ [1]

อุตสาหกรรมอ้อยและน้ำตาลทรายของประเทศไทยมีความก้าวหน้าค่อนข้างมากเมื่อเทียบกับอดีตที่ผ่านมาจึงมีการนำระบบโควตาไร้อ้อยและการจัดการผลผลิตโดยสำนักงานคณะกรรมการอ้อยและน้ำตาลทรายเป็นผู้กำหนดโควตาให้แต่ละโรงงานสามารถมีโควตาในการส่งออกอ้อยและน้ำตาลเท่าไร้และนำไปจัดสรรคให้ชาวไร้อ้อยในการปลูกผลผลิตและส่งออกอ้อยเข้าโรงงาน แต่ระบบโควตาวิธีการคำนวณปริมาณอ้อยมีความแม่นยำค่อนข้างน้อยจึงเกิดปัญหาการจัดสรรคที่ไม่ลงตัว ด้วยสาเหตุดังกล่าวผู้วิจัยได้เห็นถึงความจำเป็นและความสำคัญที่จะต้องดำเนินการวิจัยการพยากรณ์ผลผลิตอ้อยของประเทศไทย เพื่อนำ

แบบจำลองที่แม่นยำพัฒนาเป็นระบบระบบการพยากรณ์ผลผลิตอ้อยของประเทศไทย

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การเรียนรู้ของเครื่อง (Machine Learning)

การพัฒนาให้คอมพิวเตอร์สามารถเรียนรู้ประสบการณ์ที่ผ่านมา ซึ่งแนวคิดลักษณะนี้ก็เป็นแนวคิดลักษณะเดียวกันกับการเรียนรู้ของมนุษย์ ตั้งแต่แรกเกิดจนถึงเป็นผู้ใหญ่ เราล้วนผ่านกระบวนการการเรียนรู้ต่างๆ ทั้ง คลาน นั่ง ยืน เดิน วิ่ง รวมไปถึงกิจกรรมอื่นๆ ที่เราพัฒนาขึ้นในสมองของเรา นักพัฒนาการเรียนรู้ของเครื่องจึงนำแนวคิดนี้มาใช้ด้วยการนำเอาข้อมูลจำนวนที่มากพอมานำมาสร้างประสบการณ์ เพื่อให้คอมพิวเตอร์ได้เรียนรู้และการเรียนรู้เหล่านี้ก็เป็นเสมือนบทเรียนพื้นฐานสำคัญในการสร้างการสร้างความสามารถในการคาดการณ์เหตุการณ์ในอนาคตของคอมพิวเตอร์

### 2.2 ป่าสุ่ม (Random Forest)

ป่าสุ่มถูกพัฒนาขึ้นจาก ต้นไม้ตัดสินใจ ต่างกันที่ป่าสุ่ม เป็นการเพิ่มจำนวน ต้นไม้เป็น ต้นไม้หลาย ๆ ต้น ทำให้ประสิทธิภาพในการทำงานสูงขึ้น แม่นยำมากขึ้น โดยการนำเอาที่พุ่มของ ต้นไม้ตัดสินใจ หลายอันเพื่อสร้างผลลัพธ์สุดท้าย [3]

### 2.3 การวิเคราะห์การถดถอยเชิงเส้น (linear regression analysis, LRA)

เป็นวิธีที่ใช้เทคนิคทาง สถิติในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระ ซึ่งเป็นตัวแปรที่มีการกำหนดค่าที่แน่นอนไว้ล่วงหน้า และตัวแปรตามซึ่งเป็นตัวแปรที่เปลี่ยนแปลงไปตามตัวผลของตัวแปรอิสระ และมีความสัมพันธ์ของสัมประสิทธิ์ในสมการการถดถอยเป็นแบบเชิงเส้น ซึ่งมีรูปแบบสมการโดยที่ ตัวแปรตามคือ ตัวแปรอิสระ คือ สัมประสิทธิ์ของสมการถดถอย และ ค่าความคลาดเคลื่อน สัมประสิทธิ์ของสมการ ถดถอยเป็นพารามิเตอร์ที่ไม่ทราบค่าจึงต้องทำการประมาณค่า [4]

### 2.4 ปัจจัยที่ส่งผลต่อการพยากรณ์ผลผลิตอ้อย

จากการทบทวนงานวิจัย พบว่ามีปัจจัยหลายปัจจัยที่ส่งผลต่อการพยากรณ์ผลผลิต ดังนี้

การพยากรณ์ผลผลิตอ้อยในบริเวณภาคตะวันออกเฉียงเหนือของประเทศไทยใช้ข้อมูลจากการรวบรวมจาก 6 จังหวัดภาคตะวันออกเฉียงเหนือตั้งแต่ปี 2545 – 2557 ชุดข้อมูลประกอบด้วย ขนาดของพื้นที่ปลูกอ้อย, ปริมาณอ้อยส่งเข้าหีบ, ราคาอ้อยเฉลี่ย, อุณหภูมิสูงสุด, อุณหภูมิต่ำสุด, ปริมาณน้ำฝนโดยรวม, จำนวนวันที่ฝนตก และปริมาณน้ำฝนสูงสุด โดยศึกษาความสัมพันธ์รายฤดูระหว่างตัวอิสระทั้ง 8 ตัวด้วยสัมประสิทธิ์สหสัมพันธ์-เพียร์สัน ก่อนที่นำมาข้อมูลไปพยากรณ์และผลการพยากรณ์โดยใช้เทคนิคโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้น ก็มีความถูกต้องใกล้เคียงผลจริง [5]

ทำนายผลผลิตข้าวสาลีโดยใช้ปัญญาประดิษฐ์ใช้เทคนิคป่าสุ่มเปรียบเทียบกับ โครงข่ายประสาทเทียมชุดข้อมูลที่ใช้ทดสอบคือชุดข้อมูลข้าวสาลีฤดูหนาว จำนวน 182,549 แถวและมีคุณสมบัติ 26 คุณสมบัติแต่เลือกใช้แค่ 7 คุณสมบัติคือ ละติจูด, ลองจิจูด, อุณหภูมิสูงสุดที่รู้สึกได้, อุณหภูมิต่ำสุดที่รู้สึกได้, อุณหภูมิสูงสุด, อุณหภูมิต่ำสุด, ความชื้นสรุปผลการทดสอบป่าสุ่มมีค่า ความผิดพลาดเฉลี่ยอยู่ที่ร้อยละ 0.999 ดีกว่าโครงข่ายประสาทเทียมมีค่าความผิดพลาดเฉลี่ยอยู่ที่ร้อยละ 11.1782 [6]

โครงข่ายประสาทเทียมและระบบสารสนเทศภูมิศาสตร์ในการพยากรณ์ผลผลิตอ้อย โดยใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Backpropagation Algorithm) ชุดข้อมูลที่ใช้คือ ประเภทอ้อย, พันธุ์อ้อย, ลักษณะดิน, ปริมาณน้ำฝน, อายุอ้อย, ปริมาณอ้อย, อุณหภูมิ, ความชื้นสัมพัทธ์ตรวจสอบโดยใช้โปรแกรมแมตแล็บในการทดลอง สรุปผลการทดสอบค่าความผิดพลาดเฉลี่ยอยู่ที่ร้อยละ 18.99 [7]

การใช้เทคนิคโครงข่ายประสาทเทียมเพื่อพยากรณ์ผลผลิตถั่วเหลืองในฟอร์ตลอเดอร์เดล จังหวัดฟลอริดา ประเทศสหรัฐอเมริกา ชุดข้อมูลที่ใช้ถูกเก็บรวบรวมตั้งแต่ปี 2546 ถึง 2559 ประกอบไปด้วย ผลผลิตถั่วเหลือง, พื้นที่ปลูกจากภาพถ่ายดาวเทียม, อุณหภูมิพื้นผิว, อุณหภูมิพื้นผิวที่คืนเฉลี่ย สรุปผลการทดสอบค่า ความผิดพลาด เฉลี่ยอยู่ที่ร้อยละ 14.21[8]

พยากรณ์ผลผลิตอ้อยเพื่อป้อนเข้าสู่โรงงานผลิตน้ำตาล โดยรวบรวมข้อมูลเกษตรกรแต่ละรายที่ทำการปลูกอ้อยและส่งขายให้กับโรงงานอุตสาหกรรมผลิตน้ำตาล จาก 24 จังหวัดตั้งแต่ปี 2553 - 2557รวบรวมข้อมูลปัจจัยต่างๆที่อาจส่งผลต่อผลผลิตอ้อยจำนวน 9 ปัจจัยประกอบด้วย พื้นที่เพาะปลูกอ้อยใหม่, พื้นที่เพาะปลูกอ้อยต่อ, พื้นที่เพาะปลูกอ้อยต่อ 2 พื้นที่เพาะปลูกอ้อยต่อ 3, ปริมาณโควตาอ้อยที่เกษตรกรทาสัญญาไว้กับโรงงาน, พันธุ์อ้อย, ปริมาณน้ำฝนเฉลี่ยตลอดปีของพื้นที่เพาะปลูก, จำนวนวันที่ฝนตกตลอดปีของพื้นที่เพาะปลูก และปริมาณฝนสูงสุดของพื้นที่เพาะปลูก โดยนำอัลกอริทึม back propagation neural network (BPNN) เปรียบเทียบกับ genetic algorithm สรุปผลการทดสอบ ค่าความคลาดเคลื่อนของ genetic algorithm อยู่ในช่วงร้อยละ 9.64–11.65 ในขณะที่ค่าความคลาดเคลื่อนของ โครงข่ายประสาทเทียม เมื่อทดสอบกับข้อมูลชุดเดียวกันมีค่าอยู่ในช่วงร้อยละ 10.27 – 84.31 [9]

การพยากรณ์ผลผลิตมันสำปะหลังด้วยข้อมูลปัจจัยการผลิตมันสำปะหลังจากสำนักงานเกษตรจังหวัดกำแพงเพชร ประกอบด้วยทั้งหมด 27 ปัจจัยแต่หลังจากทำการวิเคราะห์สหสัมพันธ์ของข้อมูลแล้วได้ข้อมูลปัจจัยที่ส่งผลต่อการพยากรณ์ผลผลิต 9 ปัจจัยได้แก่ พันธุ์มันสำปะหลัง, อายุปลูก, เนื้อที่ปลูก, เนื้อที่เก็บเกี่ยว, เนื้อที่เสียหาย, ปริมาณปุ๋ยเคมี, ปริมาณปุ๋ยคอก, ปริมาณปุ๋ยชีวภาพ, ปริมาณปุ๋ยอินทรีย์ และใช้เทคนิคการจำแนกประเภทข้อมูล(Classification) ด้วยวิธีต้นไม้ตัดสินใจ (DecisionTree) โดยจะใช้อัลกอริทึม จำนวน 5 ตัว ได้แก่ J48, RandomTree, SimpleCart, NaïveBayes และ LADTree แล้วทำการทดสอบด้วยแบบการพยากรณ์ด้วยวิธี Cross-validation Test พบว่ายังให้ค่าแม่นยำเพียง 70.96% ซึ่งผู้วิจัยเห็นว่ายังไม่ดีพอจึงได้ปรับปรุงวิธีการทดสอบโดยการแบ่งชุดข้อมูลออกเป็น 2 ส่วนได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) จำนวน 5 ชุด แล้วทำการสร้างแบบการพยากรณ์อีกครั้งพบว่า อัลกอริทึม J48 ให้ค่าความแม่นยำสูงสุดที่ร้อยละ 75.64 อัลกอริทึม SimpleCart ให้ค่าความแม่นยำสูงสุดที่ร้อยละ

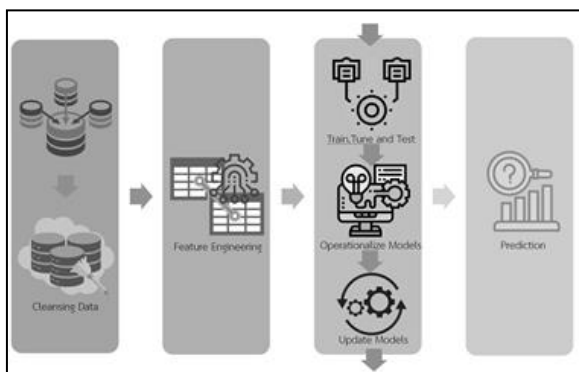
80.12 และอัลกอริทึม LADTree ให้ค่าความแม่นยำสูงสุดที่ร้อยละ 89.55 [3]

การพยากรณ์ผลผลิตพืชโดยใช้เทคนิค Linear Regression, Decision Tree Regression, K- Nearest Neighbors Regression, Support Vector Regression, และ Multilayer Perceptron Network. เปรียบเทียบกันโดยวันค่าความผิดพลาดจากค่าสถิติ สัมประสิทธิ์ของการหาพหุคูณ (R2) และ Root Mean Square ความผิดพลาด (RMSE) ชุดข้อมูลที่ใช้คือ พื้นที่ปลูกและผลผลิตลิ้นจี่ในประเทศไทย ในช่วง 2547-2561 และพยากรณ์ผลผลิตในปี 2562 ชุดข้อมูลประกอบด้วยปัจจัยดังนี้ ผลผลิตของลิ้นจี่, ปริมาณต้นลิ้นจี่ต่อผลผลิต, ปริมาณน้ำ, อุณหภูมิ, การไหลของอากาศ และคุณภาพดินแบ่งชุดข้อมูลเพื่อใช้สอนร้อยละ 80 ชุดข้อมูลเพื่อทดสอบร้อยละ 20 สรุปผลการทดสอบ Support Vector Regression มีความแม่นยำสูงที่สุดในการพยากรณ์ ตามด้วย Multilayer Perceptron Network, Linear Regression, Decision Tree Regression และ K- Nearest Neighbors Regression ตามลำดับ [11]

### 3. วิธีดำเนินการวิจัย

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องพบว่า multilinear linear regression ให้ผลแม่นยำและเหมาะสมกับข้อมูล ดังนั้นผู้วิจัยจึงประยุกต์งานวิจัย [3,11] multilinear linear regression และส่วนของ Input ข้อมูล ผู้วิจัยประยุกต์จาก [5,6,7,8,9] ซึ่งมีการทดสอบปัจจัยที่ส่งผลกระทบต่อพยากรณ์ก่อนจะนำปัจจัยต่างๆ ก่อนจะนำปัจจัยต่างๆ เข้ามาช่วยในการพยากรณ์

#### 3.1 กรอบแนวคิดของงานวิจัย



ภาพที่ 1 : กรอบแนวคิดของงานวิจัย

### 3.2 การรวบรวมข้อมูล (Data Collection)

รวบรวมข้อมูลปัจจัยที่มาจากหลายแหล่ง ประกอบด้วย พื้นที่ปลูกอ้อย, ปริมาณจากการคำนวณพื้นที่, พื้นที่เก็บเกี่ยวอ้อย, ผลผลิตอ้อยต้นต่อไร่, อุณหภูมิสูงสุด, อุณหภูมิต่ำสุด, อุณหภูมิเฉลี่ย, ปริมาณน้ำฝนเฉลี่ย, จำนวนวันที่ฝนตก และความชื้นสัมพัทธ์

#### 3.3 การเตรียมและทำความสะอาดข้อมูล

ทำความสะอาดข้อมูล (Cleansing Data) ข้อมูลที่ทำการรวบรวมอาจจะมีข้อมูลบางส่วนที่ไม่ครบ หรือคุณสมบัติบางอย่างขาดหายไป และข้อมูลอาจมีการซ้ำกันจึงจำเป็นต้องทำความสะอาดข้อมูลและแก้ไขก่อนนำมาใช้งาน

#### 3.4 หาค่าความสัมพันธ์ของปัจจัย

ทำ Feature Engineering เพื่อวิเคราะห์และศึกษาปัจจัยที่มีความสัมพันธ์กันของชุดข้อมูลที่จะนำไปพยากรณ์

### 3.5 การสร้างแบบจำลอง (Model Building)

#### 3.5.1 การแบ่งชุดข้อมูล

แบ่งชุดข้อมูลออกเป็นร้อยละ 70:30, 80:20, 90:10 สำหรับการเรียนรู้และทดสอบหากผลการทดสอบได้ผลการทดสอบไม่เป็นที่น่าพอใจจะทำการปรับแต่งแบบจำลองให้มีความผิดพลาดน้อยที่สุด

### 3.6 การประเมินแบบจำลอง (Performance Evaluation)

เพื่อทดสอบประสิทธิภาพของแบบจำลองว่ามีความเหมาะสมในการนำมาใช้หรือไม่ โดยใช้การคำนวณหาค่าความผิดพลาด โดยใช้ Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{(O_i - F_i)}{O_i} * 100$$

$O_i$  คือ ปริมาณอ้อยตามจริง

$F_i$  คือ ปริมาณอ้อยที่ได้จากการพยากรณ์

$n$  คือ จำนวนข้อมูลทั้งหมด

### 4. ผลการทดสอบ

#### 4.1 ผลการศึกษาปัจจัยที่มีความสัมพันธ์กันของชุดข้อมูล

นำข้อมูลทั้งหมดที่ผ่านการทำความสะอาดข้อมูล (Cleansing Data) เรียบร้อยแล้วมาศึกษาปัจจัยที่มี

ความสัมพันธ์กันของชุดข้อมูล ซึ่งข้อมูลทั้งหมด ประกอบด้วยพื้นที่ปลูกอ้อย, พื้นที่เก็บเกี่ยวอ้อย, ผลผลิตอ้อยต้นต่อไร่, อุณหภูมิสูงสุด, อุณหภูมิต่ำสุด, อุณหภูมิเฉลี่ย, ปริมาณน้ำฝนเฉลี่ย, ปริมาณอ้อยจากการคำนวณพื้นที่, จำนวนวันที่ฝนตก และความสัมพันธ์ ข้อมูลทั้งตั้งแต่ปี พ.ศ. 2559 – 2563 ทั้งหมด จำนวน 200 ตัวอย่าง

ตารางที่ 1 : ปัจจัยที่มีความสัมพันธ์กันของชุดข้อมูล

ลำดับ	ปัจจัย	ค่าความสัมพันธ์กับตัวแปรอิสระ
1	พื้นที่เก็บเกี่ยวอ้อย	0.597
2	พื้นที่ปลูกอ้อย	0.505
3	ผลผลิตอ้อยต้นต่อไร่	0.201
4	อุณหภูมิสูงสุด	0.097
5	ปริมาณน้ำฝนเฉลี่ย	0.088
6	ความสัมพันธ์	0.035
7	จำนวนวันที่ฝนตก	0.027
8	อุณหภูมิต่ำสุด	0.008
9	อุณหภูมิเฉลี่ย	0.007

จากตารางที่ 1 ปัจจัยที่มีความสัมพันธ์กันของชุดข้อมูล จากการหาความสัมพันธ์ระหว่างปัจจัยที่ศึกษาทั้งหมดด้วยวิธี การตัดแยกด้วยวิธีป่าสุ่มจากไลบรารีในภาษาไพธอน พบว่าปัจจัยทั้งหมดมีค่าความสัมพันธ์กับตัวแปรอิสระดังนี้ พื้นที่เก็บเกี่ยวอ้อย 0.597, พื้นที่ปลูกอ้อย 0.505, ผลผลิตอ้อยต้นต่อไร่ 0.201, อุณหภูมิสูงสุด 0.097, ปริมาณน้ำฝนเฉลี่ย 0.088, ความสัมพันธ์ 0.035, จำนวนวันที่ฝนตก 0.027, อุณหภูมิต่ำสุด 0.008, อุณหภูมิเฉลี่ย 0.007 และ ปริมาณ อ้อยจากการ คำนวณ พื้นที่ที่มีความสัมพันธ์กัน 0.75 จึงไม่นำมาพิจารณา

4.2 ผลการพยากรณ์ผลผลิตอ้อย

ตารางที่ 2 : ตารางเปรียบเทียบร้อยละความผิดพลาดของผลผลิตจากการพยากรณ์ตามจำนวนชุดข้อมูล

ลำดับ	การแบ่งชุดข้อมูลสำหรับ Train และ Test (%)	MAPE (%)
1	70: 30	7.824686
1	80: 20	6.486999
2	90: 10	2.052512

จากตารางที่ 2 ตารางเปรียบเทียบร้อยละความผิดพลาดของผลผลิตจากการพยากรณ์ตามจำนวนชุดข้อมูล พบว่าผลการเปรียบเทียบค่าความผิดพลาดเฉลี่ยน้อยที่สุดเท่ากับร้อยละ 2.05 เมื่อใช้ข้อมูลเรียนรู้ร้อยละ 90 และข้อมูลชุดทดสอบ 10

5.ผลสรุปและข้อเสนอแนะ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อพยากรณ์ผลผลิตอ้อยของประเทศไทย จากการรวบรวมปัจจัย 10 ปัจจัย ประกอบด้วย พื้นที่ปลูกอ้อย, ปริมาณจากการคำนวณพื้นที่, พื้นที่เก็บเกี่ยวอ้อย, ผลผลิตอ้อยต้นต่อไร่, อุณหภูมิสูงสุด, อุณหภูมิต่ำสุด, อุณหภูมิเฉลี่ย, ปริมาณน้ำฝนเฉลี่ย, จำนวนวันที่ฝนตก และความสัมพันธ์ ซึ่งเป็นข้อมูลจากปี พ.ศ. 2559 – 2563 จำนวน 200 ตัวอย่าง และข้อมูล 47 จังหวัดที่มีการปลูกอ้อยและวิเคราะห์ความสัมพันธ์ของปัจจัย 10 ปัจจัยด้วยวิธีการตัดแยกด้วยวิธีป่าสุ่มจากไลบรารีในภาษาไพธอน พบว่าปัจจัยที่มีผลต่อการพยากรณ์ผลผลิตอ้อยพื้นที่เก็บเกี่ยวอ้อยมีผลต่อการพยากรณ์ผลผลิตอ้อยที่ 0.597, พื้นที่ปลูกอ้อยมีผลต่อการพยากรณ์ผลผลิตอ้อยที่ 0.505, ผลผลิตอ้อยต้นต่อไร่มีผลต่อการพยากรณ์ผลผลิตอ้อยที่ 0.201, อุณหภูมิเฉลี่ยมีผลต่อการพยากรณ์ผลผลิตอ้อยที่ 0.097, ปริมาณน้ำฝนเฉลี่ยมีผลต่อการพยากรณ์ผลผลิตอ้อยที่ 0.088, ความสัมพันธ์มีผลต่อการพยากรณ์ผลผลิตอ้อยที่ 0.035, จำนวนวันที่ฝนตกมีผลต่อการพยากรณ์ผลผลิตอ้อยที่ 0.027, อุณหภูมิต่ำสุดมีผลต่อการพยากรณ์ผลผลิตอ้อยที่ 0.008, อุณหภูมิเฉลี่ยมีผลต่อการพยากรณ์ผลผลิตอ้อยที่ 0.007 และค่าปริมาณอ้อยจากการคำนวณพื้นที่ที่มีค่ามากกว่า 1 จึงไม่นำมาพิจารณาด้วยปัจจัย 9 ปัจจัยนี้มีความสัมพันธ์กันจึงนำปัจจัยทั้งหมดไปใช้ในการพยากรณ์ผลผลิตอ้อยด้วยวิธี การวิเคราะห์การถดถอยเชิงพหุคูณจากไลบรารีในภาษาไพธอนซึ่งมีการแบ่งชุดข้อมูลสำหรับเรียนรู้และทดสอบเป็น ละ 70:30, 80:20, 90:10 ตามลำดับ จากการพยากรณ์ตามจำนวนชุดข้อมูล พบว่าผลการเปรียบเทียบค่าความผิดพลาดเฉลี่ยน้อยที่สุดเท่ากับร้อยละ 2.05 เมื่อใช้ข้อมูลเรียนรู้ร้อยละ 90 และข้อมูลชุดทดสอบ 10 แสดงให้เห็นว่าเมื่อใช้ข้อมูลเรียนรู้

จำนวนมากขึ้นส่งผลให้ความแม่นยำในการพยากรณ์เพิ่มมากขึ้น สรุปได้ว่าการพยากรณ์ผลผลิตอ้อยด้วยวิธีการวิเคราะห์การถดถอยเชิงพหุคูณมีค่าความผิดพลาดที่สามารถยอมรับและสามารถนำอัลกอริทึมที่ได้จากการทดลองไปใช้ในการพัฒนาระบบพยากรณ์ผลผลิตอ้อยเพื่อให้สามารถพยากรณ์ผลผลิตอ้อยที่แม่นยำและกำหนดโควตาให้แต่ละโรงงานถูกต้องมากยิ่งขึ้น จากการศึกษาครั้งนี้สามารถนำงานวิจัยนี้เป็นข้อเสนอแนะในการพัฒนาการพยากรณ์ผลผลิตอ้อยซึ่งปัจจัยที่มีความสัมพันธ์ต่อผลผลิตอ้อยอาจจะมีปัจจัยอื่น ๆ ที่มีค่าสัมพันธ์ต่อผลผลิตอ้อยอีกหลายปัจจัยและวิธีในการพยากรณ์ผลผลิตอ้อยใช้วิธีอื่น ๆ ซึ่งอาจจะมีค่าความผิดพลาดที่น้อยกว่าการวิเคราะห์การถดถอยเชิงพหุคูณ

#### เอกสารอ้างอิง

- [1] สำนักงานคณะกรรมการอ้อยและน้ำตาลทราย “[สืบค้นวันที่ 21 ตุลาคม 2564] <http://www.ocsb.go.th/th/cms/detail.php?ID=923&SystemModuleKey=journal>
- [2] กรมส่งเสริมอุตสาหกรรม. *Machine Learning*. (2564). [ออนไลน์]. [สืบค้นวันที่ 1 มีนาคม 2565]. จาก <https://dip360.dip.go.th/machine-learning-คืออะไร/>
- [3] Preecha Limtrakul, Vipa Jaroenpuntaruk and Wittaya Pornpatcharapong “Development of a Model to Predict Cassava Yield Using Data Mining” Veridian E- Journal, Science and Technology Silpakorn University ISSN 2408 – 1248 ,Bangkok, 2016
- [4] Norawat Luangtong and Nantachai Kantanatha “Selection of the Appropriate Agricultural Yield Forecasting Models” ,Bangkok, 2015
- [5] จตุภัทร เมฆพ่าย “การพยากรณ์ผลผลิตอ้อยในบริเวณภาคตะวันออกเฉียงเหนือของประเทศไทยด้วยเทคนิค โครงข่ายประสาทเทียม” วิจัยประเภทงบประมาณเงินรายได้จากเงินอุดหนุนรัฐบาล (งบประมาณแผ่นดิน) มหาวิทยาลัยบูรพา, 2560
- [6] Mullangi Ramu and Javvadi Teja Sri “Wheat yield prediction using Artificial Intelligence models and its comparative analysis for better Prediction” 2021 International Conference on Advance Computing and Innovative Technologies in Engineering )ICACITE(, India, 2021
- [7] สุรเชษฐ์ เกื้อนแก้วสิงห์ “การใช้โครงข่ายประสาทเทียมและระบบสารสนเทศภูมิศาสตร์ในการพยากรณ์ผลผลิตอ้อย”, ขอนแก่น. มหาวิทยาลัยขอนแก่น, 2552
- [8] Anil Suat Terliksiz and D. Turgay Altılar “Use Of Deep Neural Networks For Crop Yield Prediction: A Case Study Of Soybean Yield in Lauderdale County, Alabama, USA”, Turkey,2019
- [9] ณัฐนนท์ ถีลาตระกูล, สุณิสา रिमเจริญ และ สุภาวดี ศรีภาติ “การพยากรณ์ผลผลิตอ้อยเพื่อป้อน เข้าสู่โรงงานอุตสาหกรรม”, The 2nd International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (ISMSI-2018), Phuket, 2018
- [10] Nantinee Soodtoetong, Eakbodin Gedkhaw and Montean Rattanasiriwongwut “The Performance of Crop Yield Forecasting Model based on Artificial Intelligence” 2020 17th International Conference on Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology )ECTI-CON(, Bangkok, 2020

# ระบบธุรกิจอัจฉริยะจากการทำนายพฤติกรรมกรรมการซื้อสินค้าออนไลน์ เพื่อสนับสนุนแผนการตลาด

## Business Intelligence System from Predicts Online Purchase Behavior for Support Marketing Plan

ปนัดดา ยีมาชา (Panatda Yeemasa)<sup>1</sup> และศักดิ์ชาย ตั้งวรรณวิทย์ (Sakchai Tangwannawit)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021856110@email.kmutnb.ac.th, <sup>2</sup>sakchai.t@itd.kmutnb.ac.th

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบธุรกิจอัจฉริยะสำหรับการสนับสนุนการตลาดจากการทำนายพฤติกรรมกรรมการซื้อสินค้าออนไลน์ วิจัยประกอบด้วย 1) การวิเคราะห์พฤติกรรมกรรมการซื้อสินค้าของลูกค้าด้วยอัลกอริทึม F-P Growth และ 2) พัฒนาระบบธุรกิจอัจฉริยะจากความสัมพันธ์จากอัลกอริทึม F-P growth โดยข้อมูลจัดเก็บรวบรวมการซื้อสินค้าออนไลน์ผ่านเว็บไซต์ topvalue.com ของบริษัท ทีโอพีแวลู คอร์ปอเรท จำกัด ตั้งแต่เดือน เมษายน 2564 - กรกฎาคม 2565 นำมาใช้เพื่อวิเคราะห์หากฎความสัมพันธ์ ด้วยโปรแกรม Rapid Miner Studio

การวิเคราะห์หากฎความสัมพันธ์กำหนดค่าสนับสนุน (Support) 30% ขึ้นไป ค่าความเชื่อมั่น (Confidence) 80% ขึ้นไป ผลจากกฎความสัมพันธ์จะแสดงความสัมพันธ์ของสินค้าในแต่ละเดือนผ่านระบบธุรกิจอัจฉริยะในการเสนอรายงานแบบกราฟฝ่ายการตลาดสามารถนำระบบธุรกิจอัจฉริยะเพื่อสนับสนุนโครงการด้านการตลาดและปรับกลยุทธ์ให้เข้าใจพฤติกรรมกรรมการซื้อของลูกค้า เพื่อเสนอความต้องการของลูกค้า สร้างความพึงพอใจ และรักษาลูกค้า

**คำสำคัญ:** กฎความสัมพันธ์, ระบบธุรกิจอัจฉริยะ, พฤติกรรมกรรมการซื้อออนไลน์

### Abstract

The objective of this research was to the development of business intelligence system from predicts online purchase behavior for supporting

marketing. The research methodology consists of 1) analyzing customer purchase behavior with the FP-Growth algorithm and 2) developing a business intelligence system with association rules from the F-P growth algorithm. The data was collected from online shopping through the topvalue.com website of Top Value Corporation Co., Ltd., between April 2021 - July 2022. The analysis of the data using association rules with rapid miner studio.

The association rule was to set values of support was above 30%, and confidence was above 80%. The results of the association rules showed the relation of items in each month via the business intelligence system by graph reports. The marketing department can use a business intelligence system to support marketing projects and adjust the strategy to understand the behavior of customers to offer the customer's needs, creating customer satisfaction and retain the customer.

**Keywords:** Association rule, Business intelligence, Online purchase behavior

### 1. บทนำ

ในปัจจุบันการแข่งขันทางธุรกิจทางการตลาดสูงมาก จึงส่งผลให้ธุรกิจต่าง ๆ ต้องหากกลยุทธ์ทางการตลาด คือ ธุรกิจพาณิชย์อิเล็กทรอนิกส์ (E-Commerce) ซึ่งเป็นช่องทางในการทำธุรกิจที่ทำให้ผู้ประกอบการสามารถเข้าถึงกลุ่มผู้บริโภคได้อย่างรวดเร็วมากขึ้น ความก้าวหน้าของ

เทคโนโลยี สร้างความสะดวกสบายในการค้นหาหรือการสั่งซื้อสินค้าผ่านอุปกรณ์สื่อสารเป็นสื่อกลาง ส่งผลให้พฤติกรรมของผู้บริโภคในปัจจุบันใช้บริการช้อปปิ้งออนไลน์ผ่านเว็บไซต์หรือแอปพลิเคชัน (Application) ทำให้ในแต่ละวันมีข้อมูลการซื้อขายสินค้าเกิดขึ้นมากมาย แนวโน้มการทำการตลาดออนไลน์จึงเติบโตขึ้นอย่างต่อเนื่อง มีการวางแผนการตลาดด้วยการใช้เทคนิคเหมืองข้อมูลวิเคราะห์ข้อมูล ยกตัวอย่างเช่น การจัดทำแผนการตลาด ออกผลิตภัณฑ์หรือโปรโมชั่นใหม่ๆ ที่ดึงดูดลูกค้าและเพิ่มทางเลือกให้แก่ผู้บริโภคมากขึ้น โดยอาศัยความเข้าใจในรูปแบบของพฤติกรรมผู้บริโภคออนไลน์เพื่อตอบสนองความต้องการของผู้บริโภค ทำให้ธุรกิจตลาดออนไลน์เพิ่มกลยุทธ์ในการแข่งขันกันอย่างรุนแรง ทั้งด้านการสร้างเว็บไซต์หรือแอปพลิเคชัน ที่เข้าถึงกลุ่มเป้าหมายมากขึ้น

ตามความสำคัญข้างต้น การดำเนินธุรกิจพาณิชย์อิเล็กทรอนิกส์ มีการใช้เทคโนโลยีสารสนเทศเข้ามาช่วยจัดเก็บ บริหารจัดการข้อมูล รวมถึงการนำข้อมูลรายการทางการค้ามาวิเคราะห์เพื่อใช้วางแผนการตลาดในการดำเนินธุรกิจ ส่วนใหญ่ข้อมูลจะมีปริมาณมากถูกจัดเก็บไว้แต่ยังไม่ได้มีการนำข้อมูลดังกล่าวมาวิเคราะห์เพื่อให้เกิดประสิทธิภาพในงานบริหารจัดการ เช่น ข้อมูลด้านประชากรศาสตร์ของลูกค้า ข้อมูลการซื้อขาย การจัดส่งสินค้า เป็นต้น โดยสามารถนำข้อมูลพฤติกรรมและแนวโน้มต่าง ๆ สร้างความได้เปรียบทางการตลาดได้ การเพิ่มยอดขายหรือการจัดโปรโมชั่นสินค้า จึงเป็นกลยุทธ์ทางการตลาดที่สามารถเพิ่มรายได้ให้กับธุรกิจ โดยปัจจุบันบริษัท ทีโอพี แอวู คอร์ปอเรท จำกัด ผู้จัดทำและให้บริการ topvalue.com เว็บไซต์ช้อปปิ้งออนไลน์ ที่รวบรวมสินค้ากลุ่มเครื่องใช้ไฟฟ้าและอุปกรณ์อิเล็กทรอนิกส์จากแบรนด์ชั้นนำมาจัดจำหน่ายให้ผู้บริโภคผ่านเว็บไซต์ ด้วยเทคโนโลยีที่ทันสมัย (ERP) ซึ่งจัดเก็บข้อมูลการซื้อขายสินค้าที่ลูกค้าซื้อในแต่ละครั้งไว้อย่างเป็นระบบ ส่วนด้านส่งเสริมการตลาดทางบริษัทมีการวางแผนการตลาดเป็นลักษณะนำเสนอข่าวสารต่าง ๆ ตามช่องทางที่ผู้บริโภคยินยอมรับข่าวสาร แต่ยังคงขาดความเข้าใจพฤติกรรมผู้บริโภค

สินค้าและความต้องการซื้อที่จะเป็นตัวขับเคลื่อนธุรกิจ และสามารถปรับกลยุทธ์ทางการตลาดให้เหมาะสม เพื่อเสนอความต้องการของลูกค้าได้อย่างตรงจุดมากที่สุด

ดังนั้นผู้วิจัยจึงมีแนวคิดที่จะนำเทคโนโลยีการวิเคราะห์และพยากรณ์ข้อมูล มาช่วยแก้ปัญหาดังกล่าว โดยใช้ข้อมูลการซื้อขายสินค้าออนไลน์ของลูกค้า หากมีความสัมพันธ์จากพฤติกรรมการซื้อสินค้าออนไลน์ ด้วยเทคนิค FP-Growth เนื่องจากวิธีการนี้จะทำให้ได้โมเดลที่จะนำมาใช้ให้เกิดความได้เปรียบเชิงกลยุทธ์ทางการตลาด และยังสามารถนำข้อมูลพฤติกรรมผู้บริโภคที่ได้จากการวิเคราะห์มาประยุกต์ใช้ร่วมกับการพัฒนาระบบธุรกิจอัจฉริยะ เพื่อทำให้ธุรกิจเข้าใจพฤติกรรมมาซื้อของลูกค้า ช่วยในการบริหารจัดการวางแผนการตลาด หรือจัดกิจกรรมส่งเสริมการขายที่มีความสอดคล้องกับความต้องการของลูกค้า และเหมาะสมกับพฤติกรรมการซื้อสินค้าออนไลน์ ผ่านช่องทางเว็บไซต์ topvalue.com เพื่อให้ลูกค้าเกิดความพึงพอใจและกลับมาใช้บริการซ้ำ

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

**2.1 การหาความสัมพันธ์ (Association Rules Discovery)** เป็นเทคนิคของ Data Mining หมายถึง การค้นหาความสัมพันธ์ของข้อมูลที่มีอยู่ เพื่อนำไปหารูปแบบที่เกิดขึ้นบ่อย ๆ (Frequent Pattern) และใช้ในการวิเคราะห์ความสัมพันธ์ทำนายปรากฏการณ์ต่าง ๆ ซึ่งฐานข้อมูลที่ใช้ในการทำเหมืองความสัมพันธ์ (Association Mining) มักเป็นฐานข้อมูลประเภท Transaction Database โดยผลลัพธ์ที่ได้จะเป็นกฎความสัมพันธ์ (Association Rule) สามารถเขียนได้ในรูปเซตของรายการที่เป็นเหตุ ไปสู่เซตที่เป็นผล ซึ่งมีรากฐานมาจากการวิเคราะห์ตะกร้าตลาด (Market Basket Analysis) เช่น ลูกค้าที่ซื้อผ้าอ้อมส่วนใหญ่จะซื้อเบียร์ด้วย

**2.2 เทคนิคเอพี-กโรธ (FP-Growth Algorithms)** การหาความสัมพันธ์ด้วยเทคนิคอปริโอริ (Apriori) มีการอ่านข้อมูลหลายครั้งจากฐานข้อมูล และยังคงสร้างกลุ่มแข่งขัน (Candidate) จำนวนมาก ทำให้มีการคิดค้นวิธีการแก้ไขข้อบกพร่องนี้ Han, J.P. and Yin, (2000) ได้



พัฒนาอัลกอริทึมใหม่ขึ้นมา เพื่อลดจำนวนการอ่านข้อมูล จากฐานข้อมูล พร้อมนำเสนอโครงสร้างข้อมูลแบบใหม่ ขึ้นมา ชื่อว่า “FP-Tree” โดยใช้ชื่อว่า อัลกอริทึม FP-Growth เป็นอัลกอริทึมที่อ่าน ข้อมูลจากฐานข้อมูลเพียง 2 ครั้ง

**2.3 แผนภาพข้อมูล (Data Visualization)** การนำข้อมูลดิบ มาแปลงเป็นรูปภาพเพื่อทำให้คนอ่านเข้าใจง่าย เห็นภาพรวมของข้อมูลดิบนั้น ได้รวดเร็วภายในระยะเวลาที่จำกัด ซึ่งจะช่วยในการแสดงผลข้อมูลเบื้องต้น ช่วยเพิ่มความเร็วในการสำรวจข้อมูลว่ามีอะไรที่น่าสนใจบ้าง เพื่อเป็นแนวทางก่อนการวิเคราะห์ เจาะลึกข้อมูลที่ซับซ้อนมากยิ่งขึ้น [1]

#### 2.4 งานวิจัยที่เกี่ยวข้อง

สุภาพรณ คงฉวีพรรณ (2562) งานวิจัยนี้ได้ศึกษาเพื่อ ค้นหากฎความสัมพันธ์ ด้วยอัลกอริทึม FP-Growth และจัด กลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้า ด้วยเทคนิค RFM รวมทั้งมีการค้นหากฎความสัมพันธ์ของการซื้อผลิตภัณฑ์เสริมอาหารยี่ห้อ MYHEALTH ของแต่ละกลุ่มลูกค้า โดยใช้วิธีการทำเหมืองข้อมูล ด้วยข้อมูลของลูกค้าที่ซื้อสินค้า MYHEALTH ภายใต้กรอบ CRISP-DM ผลการวิจัยแสดงให้เห็นกฎความสัมพันธ์ของการซื้อผลิตภัณฑ์เสริมอาหารยี่ห้อ MY HEALTH จำนวน 7 กฎ ซึ่งความสัมพันธ์เหล่านี้ สามารถช่วยในการจัดทำแผนการตลาด เพื่อตอบสนองความต้องการของลูกค้า และส่งเสริมให้เกิดการซื้ออย่างต่อเนื่องของลูกค้าได้ [2]

สุพิชฌาย์ แสงชาติ (2559) งานวิจัยนี้เพื่อศึกษาพฤติกรรมของผู้บริโภคที่ซื้อสินค้าคู่แข่ง ด้วยการ ใช้เหมืองข้อมูลในการค้นหากฎความสัมพันธ์ด้วยเทคนิค FP-Growth และการแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อด้วยเทคนิค RFM ผลการหาความสัมพันธ์ด้วยเทคนิค FP-Growth แสดงให้เห็นว่าลูกค้าที่ซื้อผลิตภัณฑ์เสริมอาหารยี่ห้อ BLACK จะซื้อผลิตภัณฑ์อาหารเสริม ส่วนผลการแบ่งกลุ่มลูกค้าตามพฤติกรรมซื้อด้วยเทคนิค RFM พบว่า สามารถแบ่งกลุ่มลูกค้าเป็น 3 กลุ่ม ซึ่งผลการวิเคราะห์ ข้อมูลดังกล่าวสามารถนำมาประยุกต์ใช้กับธุรกิจร้านขายยาสาขาอื่น ๆ ในบริษัท Pharmacy Chain AA หรือเพื่อเป็น

แนวทางให้กับธุรกิจและผู้ประกอบการอื่นในการดำเนินกลยุทธ์ทางการตลาดต่อไป [3]

ณัญญาพร ชื่นมัจฉา (2559) งานวิจัยนี้เพื่อหาความสัมพันธ์การซื้อสินค้าของซูเปอร์มาเก็ตขนาดใหญ่ โดยการวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองความสัมพันธ์ของการซื้อสินค้าโดยใช้กฎความสัมพันธ์ ด้วยเทคนิค FP-Growth โดยวิเคราะห์จากการซื้อสินค้าของลูกค้าในแต่ละรายบุคคล ซึ่งผลลัพธ์ที่ได้จากงานวิจัยนี้ สามารถพยากรณ์ความต้องการของลูกค้าออกเป็นกฎความสัมพันธ์ได้ 4 รูปแบบ ซึ่งสามารถนำไปใช้ในการสนับสนุนการตัดสินใจ การนำเสนอสินค้าได้ตามความต้องการของลูกค้าในแต่ละราย เป็นการสร้างความพึงพอใจให้กับลูกค้า ซึ่งเป็นกลยุทธ์หนึ่งของการวางแผนการตลาดได้อย่างเหมาะสม และมีประสิทธิภาพ [4]

### 3. วิธีการดำเนินงาน

**3.1 กรอบการดำเนินการวิจัย** งานวิจัยนี้มีขั้นตอนการดำเนินการตามแบบ (Cross-industry standard process for data mining : CRISP-DM) ดังนี้

1) Business Understanding ในการศึกษาคั้งนี้ ได้ศึกษาธุรกิจของบริษัท ที่อปแวลู คอร์ปอเรท จำกัด เปิดให้บริการเมื่อพฤศจิกายน ปี 2014 จัดทำและให้บริการ topvalue.com เว็บไซต์ช้อปปิ้งออนไลน์ ที่รวบรวมสินค้ากลุ่มเครื่องใช้ไฟฟ้าและอุปกรณ์อิเล็กทรอนิกส์จากแบรนด์ชั้นนำมาจัดจำหน่าย โดยดำเนินธุรกิจในรูปแบบ B2C มีการวางแผนกิจกรรมส่งเสริมด้านการตลาดเป็นการตั้งเอง มักเกิดจากความต้องการผลิตภัณฑ์สินค้า หรือคาดการณ์จากประสบการณ์จากเจ้าหน้าที่การตลาดเอง

2) Data Understanding ข้อมูลที่ใช้ในการวิจัยนี้ได้มาจากข้อมูลลูกค้าที่ทำการซื้อขายสินค้าออนไลน์ (Transaction Data) ผ่านเว็บไซต์ topvalue.com ของ บริษัท ที่อปแวลู คอร์ปอเรท จำกัด ที่ได้จัดเก็บไว้ในฐานข้อมูล เป็นระยะเวลาทั้งสิ้น 16 เดือน ตั้งแต่ เมษายน ปี พ.ศ. 2563 จนถึง กรกฎาคม ปี พ.ศ. 2564 และบันทึกข้อมูลการขายสินค้าเป็นไฟล์ Excel

3) Data Preparation การเตรียมข้อมูลเป็นการจัดข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปหากฎความสัมพันธ์ด้วยโปรแกรม Rapid Miner Studio โดยการคัดเลือกข้อมูลสำหรับการหากฎความสัมพันธ์ จะต้องใช้ข้อมูลจากการซื้อสินค้าแต่ละครั้งนำมาแปลงข้อมูลจัดกลุ่ม จากนั้นนำข้อมูลที่ได้ออกไปหารูปแบบของการซื้อสินค้าที่เกิดขึ้นร่วมกันบ่อย ๆ และนำไปสร้างกฎความสัมพันธ์ โดยข้อมูลหลักที่นำมาใช้สำหรับหากฎความสัมพันธ์ สำหรับงานวิจัยนี้คือวันที่ทำการซื้อขาย (Order Date) และ รหัสสินค้า (SKU)

4) Modeling นำข้อมูลมาสร้างแบบจำลองด้วยเทคนิคเหมืองข้อมูล เลือกใช้โปรแกรม Rapid Miner Studio โดยวิธีการค้นหาความสัมพันธ์ระหว่างข้อมูล ด้วยเทคนิค FP-Growth เพื่อใช้ในการตัดสินใจด้านการวางแผนสำหรับการจัดโปรโมชันให้ตรงตามความต้องการลูกค้า โดยการหากฎความสัมพันธ์ของสินค้าที่มีการซื้อตรงกันในแต่ละเดือน

5) Evaluation งานวิจัยนี้ทำการค้นหาความสัมพันธ์ของการซื้อสินค้าผ่านเว็บไซต์ออนไลน์ เป็นตัวชี้วัดประสิทธิภาพของกฎความสัมพันธ์ จะตรวจสอบได้จากค่าสนับสนุน 30% ขึ้นไป ค่าความเชื่อมั่น 80% ขึ้นไป และค่าความสอดคล้อง (Lift) มากกว่า 1

6) Deployment การนำเอาเทคนิค FP-Growth มาใช้ค้นหาความสัมพันธ์ของการซื้อสินค้าผ่านเว็บไซต์ออนไลน์ เมื่อได้แบบจำลองที่ผ่านการวิเคราะห์และสามารถนำข้อมูลไปพัฒนาระบบธุรกิจอัจฉริยะ ในรูปแบบการเสนอรายงาน โดยสามารถเรียกดูข้อมูล Web Browser ใช้เครื่องมือที่ชื่อว่า Power BI เป็นเครื่องมือในการพัฒนาสร้างรายงานให้ออกมาในรูปแบบ Dashboard สามารถนำข้อมูลในองค์กรมาทำการวิเคราะห์เชิงการตลาด ช่วยให้ผู้บริหารและเจ้าหน้าที่ทางการตลาดมีข้อมูลเชิงลึกเพิ่มความสามารถในการตัดสินใจ และวางแผนได้อย่างชาญฉลาดในรูปแบบ Data Visualization

**3.2 ข้อมูลที่ใช้ในการวิจัย** ในงานวิจัยนี้หากฎความสัมพันธ์จากข้อมูลการซื้อสินค้าออนไลน์ผ่านเว็บไซต์ topvalue.com ที่ได้จัดเก็บไว้ในฐานข้อมูล เป็นระยะเวลาตั้งแต่ เมษายน ปี พ.ศ. 2563 จนถึง กรกฎาคม ปี พ.ศ. 2564

ซึ่งสามารถแบ่งข้อมูลสำหรับค้นหาความสัมพันธ์เป็นรายเดือน ทั้งสิ้น 16 เดือน

### 3.3 เครื่องมือที่ใช้ในการวิจัย

1) โปรแกรม Microsoft Excel 2019 เป็นโปรแกรมสำหรับการสร้างตารางงานในลักษณะที่ต้องการนำข้อมูลมาวิเคราะห์ และจัดรูปแบบ รวมไปถึงการนำข้อมูลจากตารางงานที่สร้างขึ้นไปทำการคำนวณ หรือ นำไปสร้างเป็น Chart เพื่อนำเสนอข้อมูล รวมถึงการบริหาร และจัดการข้อมูลในลักษณะที่มีปริมาณมากได้ [5]

2) โปรแกรม Rapid Minor Studio เป็นซอฟต์แวร์ Data Science ใช้สำหรับการเตรียมข้อมูล การเรียนรู้เครื่อง การเรียนรู้ลึก การทำเหมืองข้อความ และการวิเคราะห์การทำนาย (Predictive analysis) เป็นซอฟต์แวร์ที่ช่วยในการจัดส่งข้อมูล และลดข้อผิดพลาด RapidMiner มีขั้นตอนพร้อมสำหรับการทำ Data mining และ Machine learning ซึ่งรวมไปถึงการโหลดและการแปลงข้อมูล (ETL) การประมวลผลล่วงหน้าและการวาดภาพจากข้อมูล การวิเคราะห์เชิงพยากรณ์ และการสร้างแบบจำลองทางสถิติ การประเมินผล และการปรับใช้ต่าง ๆ

3) โปรแกรม Power BI เป็นโปรแกรมที่ใช้สำหรับการประมวลผลข้อมูลจำนวนมาก ให้อยู่ในรูปแบบของชาร์ต (Chart) หรือ ตาราง (Table) เพื่อให้สามารถอ่านข้อมูลและนำข้อมูลไปวิเคราะห์ประโยชน์ในทางธุรกิจต่อไป คำว่า BI ย่อมาจาก Business intelligence เป็น Software มีทั้งรูปแบบ Desktop และแบบ Website เพื่อสามารถใช้ประโยชน์จากข้อมูลได้ลึกขึ้น [6]

## 4. ผลการดำเนินงาน

งานวิจัยนี้ นักวิจัยได้ประยุกต์ใช้กระบวนการทำเหมืองข้อมูลแบบ CRISP-DM มาเป็นแนวทางในการวิจัย เนื่องจากเป็นมาตรฐานที่ทั่วโลกยอมรับ โดยงานวิจัยนี้ใช้วิธีการความสัมพันธ์การซื้อสินค้าออนไลน์ ผ่านเว็บไซต์ topvalue.com ของบริษัท ที่อปแวล คอร์ปอเรท จำกัด เป็นรายเดือน จำนวนทั้งสิ้น 16 เดือน เพื่อนำไปสู่การเพิ่มยอดขายเพื่อที่จะตอบสนองและเสนอโปรโมชั่นได้ตรงจุด โดยการ

กำหนดค่า minimum support 30 % หมายถึง การขายสินค้า นั้น ๆ ได้เท่ากับหรือมากกว่า 9 วันขึ้นไปในแต่ละเดือน

ผลสรุปความสัมพันธ์สินค้าแต่ละเดือน สามารถขายสินค้าออนไลน์ได้กี่วันในแต่ละเดือน โดยมีระยะเวลา ตั้งแต่เดือนเมษายน 2563 ถึง กรกฎาคม 2564 มีผลลัพธ์ ดังนี้

**ตารางที่ 1:** แสดงผลยอดขายสินค้าที่มีค่า Support มากกว่า หรือเท่ากับ 30% ขึ้นไป

ลำดับ	เดือน ปี 2020	จำนวน สินค้า	ลำดับ	เดือน ปี 2021	จำนวน สินค้า
1	เมษายน	41	10	มกราคม	1
2	พฤษภาคม	6	11	กุมภาพันธ์	-
3	มิถุนายน	8	12	มีนาคม	1
4	กรกฎาคม	5	13	เมษายน	2
5	สิงหาคม	6	14	พฤษภาคม	2
6	กันยายน	5	15	มิถุนายน	1
7	ตุลาคม	2	16	กรกฎาคม	1
8	พฤศจิกายน	3			
9	ธันวาคม	2			

จากตารางที่ 1 พบว่า รายการสินค้าที่ขายได้มากกว่า 9 วัน ในแต่ละเดือน (minimum support >= 30%) ตัวอย่าง เช่น เดือนเมษายนที่เป็นเดือนแรกของการขายสินค้าออนไลน์ผ่านเว็บไซต์จึงทำให้มีรายการสินค้า 41 รายการที่ขายได้มากกว่า 9 วันขึ้นไปเป็นต้น สำหรับการหาความสัมพันธ์ของสินค้าแต่ละเดือนพบว่า เดือนเมษายน 2563 พบ 794 กฏ และเดือนสิงหาคม 2563 พบ 1 กฏ โดยมีค่า Support มากกว่าหรือเท่ากับ 30% ขึ้นไป และค่า Confidence มากกว่าหรือเท่ากับ 80% ขึ้นไป ดังตารางที่ 2 ซึ่งทั้ง 2 เดือนมีรายการขายสินค้า ที่มากกว่าเดือนอื่น ส่งผลให้ยอดขายสินค้าที่ขายดี และพบกฎจำนวนดังกล่าว นอกจากนี้พบว่าทุกความสัมพันธ์มีค่า lift สูงกว่า 1 ซึ่งสามารถตีความได้ว่าสินค้าในแต่ละชุดความสัมพันธ์นั้น ลูกค้านักซื้อพร้อมกัน จากกฎความสัมพันธ์นี้ เจ้าหน้าที่ทางการตลาดสามารถนำไปวางแผนการจัดโปรโมชันทางการตลาด และสามารถใช้อ้างอิงข้อมูลเหล่านี้เพื่อสนับสนุนการตัดสินใจ

**ตารางที่ 2:** แสดงกฎความสัมพันธ์ สินค้าที่มีค่า Support มากกว่าหรือเท่ากับ 30% และค่า Confidence มากกว่าหรือเท่ากับ 80% ขึ้นไป

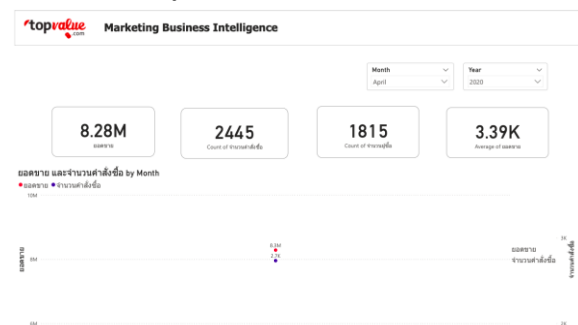
ลำดับ	เดือน ปี 2020	กฏ
1	เมษายน	794
2	พฤษภาคม	1

จากผลการสร้างกฎความสัมพันธ์ของการซื้อสินค้าออนไลน์ โดยยกกฎที่มีค่า Support เท่ากับ 45% ขึ้นไป (จับคู่กันแบบ 1:1) ดังนี้

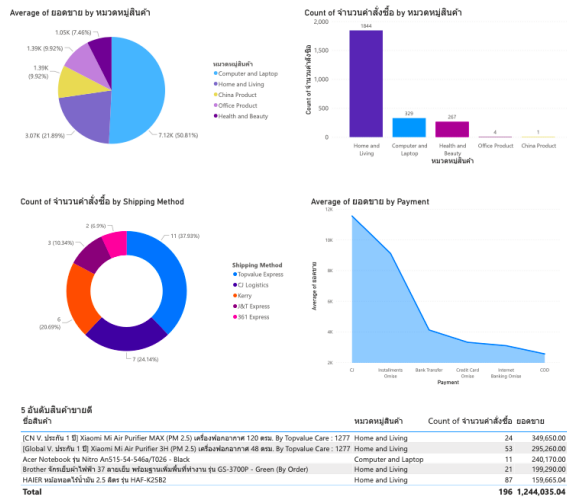
1) หากลูกค้าซื้อสินค้า 130014715 MASTERKOOL พัดลมไอเย็น 9.3 ลิตร รุ่น MIK-08EC จะซื้อ 130014067 HAIER หม้อทอดไร้น้ำมัน 2.5 ลิตร รุ่น HAF-K25B2 ด้วยค่า Support 46.7 % และค่า Confidence 82.4 % หมายความว่า เมื่อซื้อสินค้า 130014715 แล้วพบว่ามีการซื้อ 130014067 ด้วย 14 วันในเดือนเดียวกัน

2) หากลูกค้าซื้อสินค้า 130008385 OTTO หม้อทอดไร้น้ำมัน 2.8 ลิตร รุ่น CO-725 จะซื้อ 130014067 HAIER หม้อทอดไร้น้ำมัน 2.5 ลิตร รุ่น HAF-K25B2 ด้วยค่า Support 46.7 % และค่า Confidence 93.3 % หมายความว่า เมื่อซื้อสินค้า 130008385 แล้วพบว่ามีการซื้อ 130014067 ด้วย 14 วันในเดือนเดียวกัน

นอกจากนี้ได้พัฒนาสร้างรายงานให้ออกมาในรูปแบบ Dashboard สามารถนำข้อมูลในองค์กรมาทำการวิเคราะห์เชิงการตลาดในรูปแบบ Data Visualization



**ภาพที่ 1:** รายงานสรุปภาพรวมยอดขาย ยอดขายและจำนวนคำสั่งซื้อรายเดือน



ภาพที่ 2: รายงานสรุปยอดขายเฉลี่ยต่อหมวดหมู่สินค้า จำนวนคำสั่งซื้อต่อหมวดหมู่สินค้า จำนวนคำสั่งซื้อต่อประเภทการจัดส่ง ยอดขายเฉลี่ยต่อประเภทการชำระเงิน และ 5 อันดับสินค้าขายดี

**ความสัมพันธ์ของตะกร้าสินค้า**

Item1	Item2	Item3	Item4	Item5	เปอร์เซ็นต์การซื้อต่อเนื่อง
120020233					0.33
120024848					0.40
120024929					0.53
120024929	130002484				0.40
120024929	130002484	130008385			0.30
120024929	130002484	130008385	220000017		0.30
120024929	130002484	130008509			0.30
120024929	130002484	130008509	220000017		0.30
120024929	130002484	220000017			0.33
120024929	130004176				0.33
120024929	130004176	130008385			0.30
120024929	130008385				0.40

**พฤติกรรมกรการซื้อสินค้า**

ลูกค้าเลือกซื้อ	ลูกค้าที่จะซื้อตาม	เปอร์เซ็นต์การซื้อต่อเนื่อง	เปอร์เซ็นต์ความน่าเชื่อถือ
120020233	130008385	0.33	1.00
120020233	130008385, 220000017	0.30	0.90
120020233	130014067, 220000127	0.30	0.90
120020233	130014067	0.30	0.90
120020233	130014067, 130008385	0.30	0.90
120020233	220000017	0.30	0.90
120020233	220000127	0.30	0.90
120024929, 130002484	130014067	0.33	0.83
120024929, 130002484	220000017	0.33	0.83
120024929, 130002484, 130008385	220000017	0.30	1.00
120024929, 130002484, 130008509	220000017	0.30	1.00
120024929, 130002484, 220000017	130008385	0.30	0.90

ภาพที่ 3: รายงานสรุปความสัมพันธ์ของตะกร้าสินค้า และ พฤติกรรมกรการซื้อสินค้าออนไลน์ ของเดือนเมษายน 2020

**5. ข้อเสนอแนะ**

การศึกษาครั้งนี้มีข้อมูลเพียง 16 เดือน และเป็นข้อมูลที่อยู่ในช่วงสถานการณ์โควิดซึ่งส่งผลต่อการซื้อขายสินค้าออนไลน์ ควรมีการนำข้อมูลใหม่ที่ถูกจัดเก็บไว้จำนวน 5

ปีขึ้นไป มาวิเคราะห์เปรียบเทียบความสัมพันธ์รายเดือนต่อรายปี

**เอกสารอ้างอิง**

- [1] Data Visualization คืออะไร?. (2564). สืบค้น 18 กุมภาพันธ์ 2565, จาก <https://blog.skooldio.com/data-visualization-google-data-studio/>
- [2] สุภาพรชน คงฉวีพรหม. (2562). การหาความสัมพันธ์จากฐานข้อมูลการซื้อผลิตภัณฑ์เสริมอาหารยี่ห้อมายเฮลท์ของกลุ่มลูกค้าโดยใช้อัลกอริทึม เอพพีโกรีท และการแบ่งกลุ่มลูกค้าตามพฤติกรรมกรการซื้อสินค้ามายเฮลท์ โดยเทคนิคอาร์เอพเอ็มด้วยโปรแกรม แรพพิคไมเนอร์ : กรณีศึกษา ของร้านยาชนแห่งหนึ่ง. วารสารระบบสารสนเทศด้านธุรกิจ (JISB), 5 (4), 21-39.
- [3] พิมพ์พนิต สิทธิชนวิธ. (2563). การใช้เหมืองข้อมูลสำหรับผลิตภัณฑ์เสริมอาหารเบสลิค เพื่อ กำหนดแนวทางในการเพิ่มส่วนแบ่งทางการตลาดของบริษัท ฟาร์มมาซี เซน เออ โดยศึกษาพฤติกรรมของผู้บริโภคที่ซื้อสินค้าคู่แข่งด้วยโปรแกรม แรพพิคไมเนอร์. วารสารระบบสารสนเทศด้านธุรกิจ (JISB), 6 (3), 56-67.
- [4] ัญญาพร ชื่นมัจฉา. (2559). การสร้างแบบจำลองความสัมพันธ์สำหรับฐานข้อมูล การสั่งซื้อสินค้าโดยใช้เทคนิค เอพพี-โกรีท (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). กรุงเทพฯ. มหาวิทยาลัยศรีปทุม. สืบค้นจาก <http://www.dspace.spu.ac.th/handle/123456789/5165>.
- [5] Microsoft Excel 2019. (2562). สืบค้น 18 กุมภาพันธ์ 2565, จาก <https://www.arit.co.th/training/57/microsoft-excel-2019#>.
- [6] ทำไมคนส่วนใหญ่เลือกใช้งาน โปรแกรม PowerBI. (2562). สืบค้น 18 กุมภาพันธ์ 2565, จาก <https://www.arit.co.th/topic/5/#:~:text=%>

# การจำแนกข้อความสำหรับระบบการตอบกลับสนทนาผ่านข้อความอัตโนมัติ

## Sentiment Analysis for Service Chatbot

ธิตี ทองอำพัน (Thiti Tongumpun)<sup>1</sup> และ กาญจนา วิริยะพันธ์ (Kanchana Viriyapant)<sup>2</sup>

ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307011858209@email.kmutnb.ac.th , <sup>2</sup>kanchana.v@itd.kmutnb.ac.th

### บทคัดย่อ

ปัจจุบันปริมาณข้อมูลข้อความจากโซเชียลมีเดียเพิ่มขึ้นอย่างมาก ด้วยความก้าวหน้าทางเครือข่ายสังคมออนไลน์ เว็บไซต์ บล็อกต่าง ๆ และระบบตอบกลับข้อความอัตโนมัติ ในปัจจุบันระบบตอบกลับการสนทนาอัตโนมัติเป็นที่นิยมอย่างแพร่หลายแต่ไม่สามารถจำแนกคำหรือประเภทเองได้ จากปัญหาดังกล่าวผู้วิจัยสนใจที่จะพัฒนาระบบวิเคราะห์ความรู้สึกจากข้อความด้วยการเรียนรู้ของเครื่อง เพื่อให้สามารถตอบได้ตรงวัตถุประสงค์ของผู้ใช้งาน และสามารถนำข้อเสนอแนะไปปรับปรุงแก้ไขให้ดีขึ้น

จากการทดสอบประสิทธิภาพของระบบจากข้อมูลด้านการบริการได้ค่าความถูกต้อง (Accuracy) มากกว่า 80 เปอร์เซ็นต์ในแต่ละแบบจำลอง ดังนั้นสามารถสรุปได้ว่าระบบที่พัฒนาขึ้นนั้นมีประสิทธิภาพในการจำแนกข้อความภาษาไทยในด้านบริการและสามารถนำไปใช้งานได้อย่างถูกต้อง

**คำสำคัญ :** การวิเคราะห์ความรู้สึก, การจำแนกข้อความภาษาไทย, การเรียนรู้ของเครื่อง

### Abstract

*In recent years, the volume of text messages sent via social media has increased exponentially with the development of social networks, websites, blogs, and chatbot. Chatbots are very popular nowadays, but they are unable to recognize words or sentences on their own. For solving such problem sentiment analysis for service chatbot using machine learning has been developed. In order for the system to*

*answer exactly the purpose of the user and can improve correctly from suggestion.*

*The system performance testing from services data show that the accuracy is more than 80 percent. It can be concluded that the system is effective in Thai classification in term of services.*

**Keywords:** Sentiment Analysis, Text Classification, Machine Learning.

### 1. บทนำ

ปัจจุบันปริมาณข้อมูลข้อความจากโซเชียลมีเดียเพิ่มขึ้นอย่างมาก ด้วยความก้าวหน้าทางเครือข่ายสังคมออนไลน์ เว็บไซต์ หรือบล็อกต่าง ๆ ในปัจจุบัน การวิเคราะห์ความรู้สึกถือเป็นส่วนหนึ่งของการประมวลภาษาธรรมชาติ (Natural Language Processing: NLP) ในการระบุประเภทของความรู้สึก (Sentiment Polarity) ของข้อความที่ถูกจำแนกออกเป็น 2 ประเภทได้แก่ เชิงบวกและเชิงลบ หลายบริษัทในประเทศไทยเริ่มเห็นความสำคัญของการวิเคราะห์โซเชียลมีเดียเพื่อให้ได้ข้อมูลเชิงลึก การที่มีระบบที่ช่วยวิเคราะห์ข้อความจากโซเชียลมีเดีย ทำให้ธุรกิจนั้นสามารถปรับปรุงข้อเสนอแนะ และเพิ่มช่องทางธุรกิจมากขึ้น [1]

ระบบตอบกลับการสนทนาผ่านข้อความอัตโนมัติเป็นที่นิยมอย่างแพร่หลายในปัจจุบัน สามารถช่วยให้ลดการประสบปัญหาตอบลูกค้าไม่เร็วทันใจ ระบบตอบกลับการสนทนาอัตโนมัติกลับแบ่งเป็น 2 ประเภทในการตอบกลับ 1. ประโยคทั่วไป 2. ประโยคด้านลบ แต่ระบบตอบกลับการสนทนาอัตโนมัติไม่สามารถจำแนกคำหรือประโยคเองได้

จากปัญหาดังกล่าวผู้วิจัยสนใจที่จะพัฒนาระบบวิเคราะห์ความรู้สึกจากข้อความ เพื่อให้ระบบตอบกลับการสนทนาผ่านข้อความอัตโนมัติสามารถตอบได้ตรงจุดประสงค์ของผู้ใช้งาน หากในกรณีเป็นประโยคคำถาม ระบบตอบกลับการสนทนาผ่านข้อความอัตโนมัติสามารถสลับไปเป็นคนเข้ามาแก้ไขปัญหาแทนได้ และสามารถนำข้อเสนอแนะไปปรับปรุงแก้ไขได้ตรงจุดประสงค์

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 กระประมวลผลภาษาตามธรรมชาติ (Natural Language Processing)

NLP มีจุดมุ่งหมายให้คอมพิวเตอร์สามารถเข้าใจภาษามนุษย์ได้ดียิ่งขึ้น เนื่องจากคอมพิวเตอร์ถูกออกแบบมาให้เหมาะกับการเข้าใจข้อมูลของตัวเลขหรือรหัส ซึ่งไม่ตรงกับวิธีการสื่อสารของมนุษย์ NLP จึงเกิดขึ้นเพื่อลดช่องว่างในการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์ [2]

#### 2.2 พื้นฐานของพจนานุกรม (Lexicon Based)

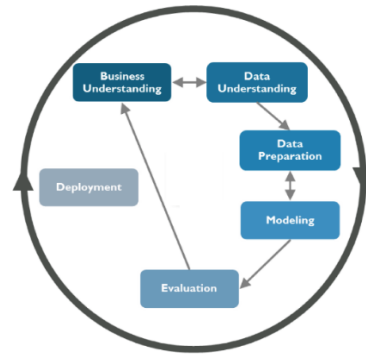
เป็นการวิเคราะห์โดยอ้างอิงจากคลังคำศัพท์ที่บ่งบอกถึงความรู้สึกและรวบรวมคำที่บ่งบอกถึงความรู้สึกออกมาเป็นเชิงบวกหรือเชิงลบ การสร้างคลังศัพท์ คือการใช้ข้อมูลความสัมพันธ์ระหว่างคำในพจนานุกรมเพื่อรวบรวมคำ บอกความรู้สึก หรือสกัดเอาคำบอกความรู้สึกจากข้อมูลการใช้ภาษาจริงในคลังข้อมูลภาษา [3]

#### 2.3 การเรียนรู้ของเครื่อง (Machine Learning)

การทำให้ระบบคอมพิวเตอร์สามารถเรียนรู้ได้จากตัวอย่างด้วยตัวเองโดยปราศจากการป้อนคำสั่งของโปรแกรมเมอร์และใช้ ประกอบได้ด้วยข้อมูลและเครื่องมือทางสถิติเพื่อทำนายผลลัพธ์ออกมา ผลลัพธ์ที่ออกมาถูกใช้เพื่อทำประโยชน์ในเชิงลึกต่อไปและเกี่ยวข้องกับเหมืองข้อมูล (Data Mining)

### 2.3.1 เหมืองข้อมูล (Data Mining)

เป็นกระบวนการค้นหาความรู้ที่ซ่อนอยู่ในข้อมูลจำนวนมากที่ถูกจัดเก็บไว้ในคลังข้อมูลด้วยหลักการทางคณิตศาสตร์และสถิติ โดยขั้นตอนประกอบด้วย 6 ขั้นตอน คือ การเข้าใจปัญหา, การทำความเข้าใจข้อมูล, การเตรียมข้อมูล, การสร้างแบบจำลอง, การประเมินผล และการนำผลลัพธ์ที่ได้ไปใช้งาน



ภาพที่ 1: กระบวนการในการทำเหมืองข้อมูล [4]

### 2.3.2 เหมืองข้อความ (Text Mining)

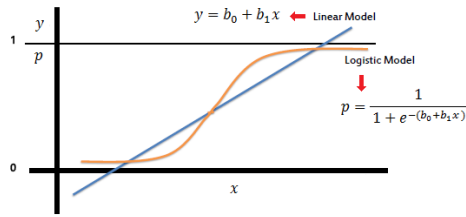
การค้นหาข้อมูลสารสนเทศที่มีประโยชน์จากภาษาธรรมชาติ (Natural Language) เป็นการสกัดความรู้เอาข้อมูลที่ไม่เคยเรียนรู้มาก่อนมาใช้ประโยชน์ได้ตามวัตถุประสงค์ [4]

### 2.4 เทคนิคการจำแนกประเภทและการเรียนรู้เชิงลึก (Deep Learning)

#### 2.4.1 Logistic Regression

Logistic Regression เป็น Supervised Learning สำหรับใช้ในการ Classification การวิเคราะห์ความสัมพันธ์ของตัวแปรในรูปแบบของการทำนายหรือโอกาสความน่าจะเป็นของการเกิดขึ้นหรือไม่เกิดเหตุการณ์สนใจ ใ้กับปัญหา Binary Classification มีตัวแปรตามเป็นแบบ 1 / 0 หรือ Yes / No กับตัวแปรอิสระที่เป็นได้ทั้งตัวแปรเชิงปริมาณหรือตัวแปรเชิงกลุ่ม

โดย กฎของความน่าจะเป็น ค่าจะเป็นตัวเลขค่าไม่เกิน 1 โดยใช้ Sigmoid function แกน Y คือความน่าจะเป็นนั้น ๆ อยู่ในช่วง 0-1 และแกน X คือค่า Regression ที่ทำนายออกมาได้ [5]



ภาพที่ 2: แบบจำลอง Logistic Regression [5]

### 2.4.2 Naïve Bayes Classification

เป็นการแบ่งกลุ่มโดยใช้หลักการความน่าจะเป็น (Probability) ทฤษฎีความเป็นเบย์ (Bayes Theorem) คือความน่าจะเป็นที่เกิดเหตุการณ์หนึ่ง ก็ต่อเมื่อเกิดเหตุการณ์หนึ่งได้เกิดไปแล้ว โดย C คือ Class, X คือ Attribute, P คือ Probability มาใช้หาคำตอบ และเลือกผลลัพธ์ (X) ที่ดีที่สุดมาเป็นคำตอบ [4]

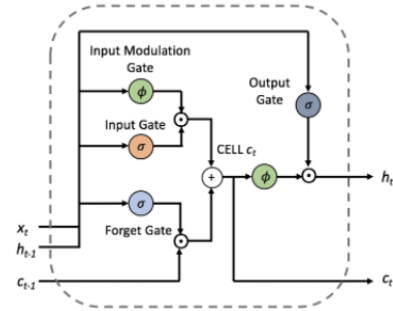
$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Labels in the diagram:  
 - Likelihood of the Evidence given that the Hypothesis is True:  $P(E|H)$   
 - Prior Probability of the Hypothesis:  $P(H)$   
 - Posterior Probability of the Hypothesis given that the Evidence is True:  $P(H|E)$   
 - Prior Probability that the evidence is True:  $P(E)$

ภาพที่ 3: แบบจำลอง Naïve Bayes [4]

### 2.4.3 Long Short-Term Memory (LSTM)

เป็นการพัฒนาต่อมาจาก Recurrent Neural Network (RNN) โดย RNN เป็นวิธีการที่ออกมาแก้ปัญหาสำหรับงานที่มีข้อมูลมีลำดับ (Sequence) โดยใช้หลักการส่งผลลัพธ์ย้อนกลับไปเป็นชุดข้อมูลใหม่ เพื่อให้โมเดลจำอินพุตของลำดับได้ และ LSTM เป็นการพัฒนาต่อยอดมาจาก RNN ที่มีกรต่อของลำดับ (Sequence) ยาว ๆ เลย มีการนำมาพัฒนาการใช้ Long Short-Term Memory ขึ้นมาโดยใช้หลักของคอมพิวเตอร์ที่มีการ Read, Write, Forget (Delete) เพื่อควบคุม memory [6]



ภาพที่ 4: แบบจำลอง LSTM [6]

## 2.5 Pre-processing

### 2.5.1 การทำความสะอาดข้อความ (Text Cleaning)

เป็นการจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ได้อย่างถูกต้อง ถือว่าเป็นหนึ่งในขั้นตอนสำคัญของการจัดการชุดข้อมูลที่จะนำไปใช้วิเคราะห์ เพื่อทำการสร้างประสิทธิภาพในการวิเคราะห์ที่ดี โดยผู้วิจัยเลือกการทำความสะอาดข้อมูลโดยมีพื้นฐานจากภาษาอังกฤษ [7]

### 2.5.2 การตัดคำ (Word Segmentation)

หลังจากการทำความสะอาดข้อมูลแล้ว Word Segmentation เป็นขั้นตอนแรกเริ่มในการนำข้อมูลมาใช้ในการวิเคราะห์ โดยทำการแยกคำออกจากประโยคให้อยู่ในรูปแบบคำเดี่ยวหรือกลุ่มคำที่มีความหมาย โดยเรียกว่า Token จากนั้นทำการรวมกลุ่มคำที่แตกต่างกัน และทำการนับจำนวนคำเหล่านั้นที่ปรากฏอยู่ในชุดข้อมูล

### 2.5.3 การสร้างถุงคำ (Bag of Words)

Bag of Words เป็นวิธีการ NLP ในการจัดเตรียมข้อมูลเพื่อนำไปใช้ในการวิเคราะห์บนระบบคอมพิวเตอร์ โดยเป็นวิธีการสร้างคุณลักษณะของข้อความขึ้นมาโดยใช้หลักการ One-Hot Encoding ในการเข้ารหัสข้อมูลในทุกคำของชุดข้อมูล โดยเข้ารหัสผ่าน Token ที่ได้จากการสร้างเอาไว้แล้ว โดยแทนค่า 1 เมื่อคำเหล่านั้นปรากฏในชุดข้อมูลและ 0 เมื่อคำเหล่านั้นไม่ปรากฏในชุดข้อมูล

**2.5.4 Term Frequency – Inverse Document Frequency (TF-IDF)**

TF-IDF ได้มีการพัฒนามาจาก Bag of Words ที่ถูกออกแบบขึ้นมาสำหรับใช้งานในส่วนของการแปลงข้อมูลที่อยู่ในรูปแบบของข้อความ ให้อยู่ในรูปแบบตัวเลขเพื่อให้คอมพิวเตอร์สามารถทำความเข้าใจและนำไปใช้ประโยชน์ต่อได้ โดย TF-IDF ใช้แก้ปัญหาในส่วนของการที่ปรากฏเป็นจำนวนมากในข้อมูลเพียงแค่ 1 ระเบียบที่ เกิดขึ้นจากการนำ Bag of Words ไปใช้งาน TF-IDF ใช้หลักการคล้ายคลึงกับ Bag of Words แต่มีจุดเด่น เหนือกว่า Bag of Words ไม่เพียงแต่นับจำนวนคำศัพท์ที่พบมากที่สุดในช่วงข้อมูลเท่านั้น

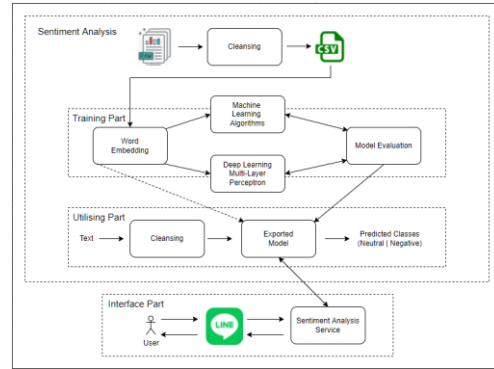
**2.5.5 การฝังคำ (Word Embedding)**

การฝังคำ คือการสร้างเวกเตอร์คุณลักษณะขึ้นมา จาก Token ที่สร้างไว้ โดยทำการสร้างขึ้นมาจากประโยคหรือเอกสารที่มีอยู่ในข้อมูลเพื่อทำการสร้างคุณลักษณะที่อยู่ในรูปของตัวเองทำให้สามารถนำไปใช้ในการคำนวณต่อได้ การทำงานของการฝังคำเริ่มจากการเข้ารหัสคำแต่ละคำให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถเข้าใจได้ด้วยนำประโยคในชุดข้อมูลหรือเอกสารที่ได้ทำการกำหนดไว้ในชุดข้อมูลมาเข้ารหัส [8]

**3. วิธีดำเนินการวิจัย**

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องพบว่าแบบจำลองในการจำแนกประเภทที่ให้ผลแม่นยำและเหมาะสมได้แก่ Logistic Regression, Naïve Bayes และ LSTM โดยใช้แบบ Bidirectional LSTM เพื่อเพิ่มการเรียนรู้สองทิศทางทำให้รับรู้บริบทของข้อความได้มากยิ่งขึ้น ดังนั้นผู้วิจัยจึงประยุกต์แบบจำลองทั้ง 3 แบบในการพัฒนาแบบจำลอง

**3.1 กรอบแนวคิดของงานวิจัย**



**ภาพที่ 5 :กรอบแนวคิดของงานวิจัย**

แนวคิดหลักของงานวิจัยนี้คือการนำเทคนิคการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกมาประยุกต์ใช้กับเทคนิคการวิเคราะห์ความรู้สึกในการจำแนกข้อความสนทนาผ่านระบบตอบกลับอัตโนมัติ และนำไปใช้พัฒนาระบบจำแนกความรู้สึก โดยระบบทำการโหลดแบบจำลองและแปลงอินพุต เมื่อระบบได้รับประโยคอินพุต ระบบจะทำการแปลงเป็น Token และทำนายความรู้สึก ด้วยค่าที่ป้อนจากข้อมูลเคน

**3.2 การรวบรวมข้อมูล (Data Collection)**

ชุดข้อมูล (Dataset) ที่นำมาใช้ในงานวิจัยและสร้างแบบจำลองนำมาจากแพลตฟอร์มการตอบกลับสนทนาด้านการให้บริการจากบริษัทขนส่งบริษัทหนึ่ง ทั้งทาง Line, Mobile, Web โดยเก็บข้อมูลตั้งแต่ 1 กรกฎาคม - กันยายน 2564 และ Wisersight Corpus จำนวน 36,066 ข้อความ โดยแบ่งเป็นด้านปกติจำนวน 25,285 ข้อความ และด้านลบจำนวน 10,781 ข้อความ นำออกเป็นไฟล์นามสกุล .csv และนำมาแบ่งประเภทข้อความออกเป็น 2 ประเภทได้แก่ Neutral และ Negative [7]

**3.3 การเตรียมและทำความสะอาดข้อมูล**

หลังจากเก็บรวบรวมข้อมูล ในการ Label ข้อความแบ่งออกเป็นการ Label ความรู้สึกว่าเป็นข้อความที่มีความรู้สึกอย่างไร โดยจะดูจากภาพรวมของประโยค หากคำเช่น แย่มาก จะ Label ให้เป็นความรู้สึกเชิงลบ หรือถ้าไม่ได้กล่าวถึงความรู้สึกใดๆ จะ Label เป็นกลาง และเนื่องจากข้อความด้านปกติมีจำนวนมากกว่าด้านลบ ข้อมูลจึงเป็นข้อมูลที่ไม่สมดุล (Imbalanced) ทำให้ส่งผลต่อความแม่นยำการพยากรณ์ของแบบจำลอง (bias) เพื่อสร้างสมดุลให้ข้อมูลที่นำไปสร้างแบบจำลอง จึงใช้



เทคนิคการสุ่มตัวอย่างชุดข้อมูลที่ไม่สมดุล (Random Resample) ด้วยการการสุ่มข้อมูลซ้ำเพื่อเพิ่มตัวอย่างจำนวนข้อมูลน้อย (Random Over Sampling) ให้เท่ากับข้อมูลที่มีจำนวนมาก การสุ่มในกรณีนี้คือ ข้อความด้านลบให้ใกล้เคียงกับข้อความด้านปกติ จากนั้นทำความเข้าใจข้อความและเก็บในรูปแบบของ .csv [9]

**3.3.1 การตัดคำ (Tokenize)**

ในการใช้ข้อมูลเพื่อสร้างแบบจำลอง ข้อมูลจะต้องอยู่ในรูปแบบของตัวเลขหรือ Token อย่างไรก็ตามมีปัญหาในภาษาไทย เพราะภาษาไทยไม่ใช่อักขระช่องว่าง (White Space) เพื่อป้องกันขอบเขตของแต่ละคำเหมือนภาษาอังกฤษ อีกทั้งคอมพิวเตอร์ยังแยกคำภาษาไทยแต่ละคำเป็นประโยคไม่ได้เพราะไม่รู้จักแต่ละคำเหมือนมนุษย์ ดังนั้นจึงจำเป็นต้องใช้อัลกอริทึมเพื่อแบ่งสัดส่วนหรือ Token คำ

**3.3.2 การฝังคำ (Word Embedding)**

ในแบบจำลองการเรียนรู้ของเครื่อง Logistic Regression, Naïve Bayes นำ Token ที่ได้จากการตัดคำไปทำ Feature Extraction ในขั้นตอนต่อไป แต่แบบจำลองการเรียนรู้เชิงลึก LSTM ได้ใช้อัลกอริทึม Word Embedding ในการแปลงข้อความเพื่อใช้ในการสร้างแบบจำลอง การฝังคำเป็นกระบวนการที่แปลงแต่ละคำเป็นประโยคในรูปแบบเวกเตอร์ ซึ่งแทนความหมายของคำที่ข้อมูลตัวเลขในหลายมิติ โดยกำหนดให้ Token มีความยาวเท่ากันทั้งหมดในเอกสาร โดยหาความยาวที่สูงสุดของเอกสารในชุดข้อมูล

**3.3.3 การสกัดคุณลักษณะ (Feature Extraction)**

ทำการหาความถี่ของคำในเอกสาร เพื่อแสดงคำในเอกสารปรากฏบ่อย ๆ ด้วย TfIdf Vectorizer จาก Scikit-learn โดยวิเคราะห์ข้อมูลจาก Matrix ของ Count Vectorize จาก Bag of Words

**3.4 การสร้างแบบจำลอง (Model Building)**

**3.4.2 การแบ่งชุดข้อมูล**

ข้อมูลพร้อมใช้งานสำหรับสร้างแบบจำลอง ชุดข้อมูลจำเป็นต้องแยกเป็นชุด Train และ Test เพื่อให้แบบจำลองมีแนวโน้มที่จะทำนายผลลัพธ์ได้อย่างถูกต้องโดยที่แบบจำลอง

ไม่ได้เรียนรู้ข้อมูลมาก่อน มีการแบ่งอัตราส่วนชุดข้อมูลสำหรับฝึกอบรม 80% และชุดข้อมูลสำหรับทดสอบ 20%

**3.4.2 อัลกอริทึม**

จากการศึกษาพบว่าการใช้ Logistic Regression, Naïve Bayes และ LSTM เป็นที่นิยมในการใช้จำแนกความรู้สึกและได้ประสิทธิภาพ ดังนั้นแบบจำลองที่จะนำมาใช้เปรียบเทียบจะประกอบด้วย Logistic Regression, Naïve Bayes, LSTM แบบเรียนรู้แบบสองทิศทาง (Bidirectional LSTM) ซึ่งทำการเปรียบเทียบแบบจำลองทั้งสาม โดยใช้ชุดข้อมูลและตัวชี้วัดเดียวกัน

**4.ผลการทดสอบ**

การทดสอบประสิทธิภาพแบบภาพแบบจำลองการจำแนกอารมณ์ของข้อความ จะทดสอบความแม่นยำโดยใช้การวัด F-measure (F1) ค่าความเที่ยงตรง (Precision) และค่าความไวหรือค่าระลึก (Recall) เป็นเกณฑ์ในการกำหนดประสิทธิภาพของแบบจำลอง โดยเปรียบเทียบการจำแนกของแบบจำลองที่ถูกฝึกสอนด้วยคุณลักษณะในรูปแบบต่าง ๆ ตามที่ผู้วิจัยได้ทำการออกแบบไว้ ประกอบด้วยผลการทดสอบประสิทธิภาพแบบจำลองของอัลกอริทึม Machine Learning และ Deep Learning ได้แก่ Logistic Regression, Naïve Bayes และ LSTM

**ตารางที่ 1 :** ตารางความแม่นยำ Recall และ F1 Score ของแต่ละแบบจำลอง

Algorithm (Model)	Precision	Recall	F1	Training Time (second)
Logistic Regression	0.8351	0.8332	0.8341	51.51
Naïve Bayes	0.8020	0.8049	0.8010	48.44
LSTM	0.8641	0.9052	0.8841	1299.57

ในแบบจำลองทั้ง 3 แบบจำลองที่ได้ทำการฝึกอบรม ได้ค้นพบว่าการสร้างแบบจำลองในการจำแนกความรู้สึกให้ผลลัพธ์ขึ้นอยู่กับความยาวในข้อความ ความขัดแย้งกันในข้อความและความหมายของแต่ละคำ ในประโยค ในด้านการทำนายผลลัพธ์แบบจำลองการเรียนรู้เชิงลึก LSTM สามารถสร้างแบบจำลองที่มีความแม่นยำในการจำแนกความรู้สึกภาษาไทยมากที่สุดที่ 88 % แต่จะใช้เวลาในการฝึกอบรมมากที่สุด

ตารางที่ 2 : ตารางบันทึกผลการทดสอบจำแนกประโยค

ประโยค	ผลลัพธ์	Logistic Regression	Naïve Bayes	LSTM
ในข้อปืบอกกำลังนำส่ง แต่ในนี้หาไม่เจอ นำเมื่อมาก	Neg <sup>1</sup>	Neu <sup>2</sup>	Neu <sup>2</sup>	Neg <sup>1</sup>
แกะของค่อหน้าพนักงานก่อนจ่ายเงินจะได้ไหมครับ	Neu <sup>2</sup>	Neu <sup>2</sup>	Neu <sup>2</sup>	Neu <sup>2</sup>
ส่งของสาขาใกล้บ้านขึ้นว่าไม่รองรับเก็บเงินปลายทางทุกสาขาเลยคะในตัวเมืองราชบุรี	Neg <sup>1</sup>	Neu <sup>2</sup>	Neu <sup>2</sup>	Neg <sup>1</sup>

Neg<sup>1</sup> = Negative, Neu<sup>2</sup> = Neutral



ภาพที่ 6: ผลการพัฒนาแบบจำลองร่วมกับแอปพลิเคชัน Line

### 5.ผลสรุปและข้อเสนอแนะ

จากการทดสอบทั้ง 3 แบบจำลอง แบบจำลองการเรียนรู้เชิงลึก LSTM แบบเรียนรู้สองทิศทางได้คะแนน

F1 ดีที่สุด นอกจากนั้นการเรียนรู้แบบสองทิศทางทำให้ LSTM ทำงานได้ดีในการจำแนกประเภทของข้อความ และตีความหมายทั้งหมดของเอกสารเมื่อมีความขัดแย้งกันในประโยค ดังนั้นงานวิจัยนี้สามารถนำมาใช้งานเป็นระบบตอบกลับการสนทนาผ่านข้อความอัตโนมัติในการให้บริการที่มีประสิทธิภาพ อย่างไรก็ตาม ความสำคัญในการทดลองนี้คือชุดข้อมูลที่สร้างขึ้นควรสำรวจอย่างละเอียดก่อนดำเนินการ การที่แบบจำลองจะให้ประสิทธิภาพในการทำงานที่สูงนั้น ไม่เพียงแต่การเลือกอัลกอริทึมที่เหมาะสมสำหรับการสร้างแบบจำลอง แต่รวมถึงการคัดเลือกและทำการสร้างคุณลักษณะของข้อมูลที่น่าไปวิเคราะห์ให้เหมาะสมกับประเภทงานที่จะวิเคราะห์ ดังนั้นการที่จะนำไปใช้งานกับชุดข้อมูลประเภทอื่น ๆ จะต้องทำการวิเคราะห์ข้อมูลและสร้างคุณลักษณะที่เหมาะสมสำหรับการใช้งานด้วย รวมไปถึงข้อมูลมีผลต่อการทำนายควรทำความสะอาดข้อมูลและวิเคราะห์อย่างละเอียดก่อนนำมาใช้งาน

### เอกสารอ้างอิง

- [1] Kim, K. and J. Lee, *Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction*. Pattern Recognition, 2014. 47(2): p. 758-768.
- [2] Choochart Haruechaiyasak, et al., *S-Sense: A Sentiment Analysis Framework for Social Media Monitoring Applications*. S-Sense: A Sentiment Analysis Framework for Social Media Monitoring Applications, 2018. 14(1).
- [3] Kusriani and M. Mashuri. *Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication*. in *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT)*. 2019.
- [4] Rajput, R. and A. Solanki, *Real time sentiment analysis of tweets using machine learning and semantic analysis*. 2016. 687-692.
- [5] Imamah and F.H. Rachman. *Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Resgresion*. in *2020 6th Information Technology International Seminar (ITIS)*. 2020.
- [6] Pal, S., S. Ghosh, and A. Nag. *Sentiment Analysis in the Light of LSTM Recurrent Neural Networks*. International Journal of Synthetic Emotions, 2018. 9: p. 33-39.
- [7] Suriyawongkul Arthit, et al., *PyThaiNLP/wisesight-sentiment: First release*. Zenodo, 2019. 1.0.
- [8] Hvitfeldt, E. and J. Silge, *Word Embeddings*. 2021. p. 73-98.
- [9] Das, S. and M. el Ghzaoui, *Performance Analysis of Over sampled OFDM over PLC Network*. Journal of Nano- and Electronic Physics, 2021. 13.

# การพัฒนาสื่อการเรียนรู้ในรูปแบบเกมมือถือ เรื่องอนุกรมวิธาน

หัวข้อ อาณาจักรสัตว์

## Development of Learning Media Topic: Taxonomy Animal Kingdoms

ณัฐภูมิ บุญเรืองโรจน์ (Natthapoom Boonruangroj)<sup>1</sup> และณัฐวี ฤกษ์ (Nattavee Utakrit)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021858333@email.kmutnb.ac.th <sup>2</sup>nattavee.u@itd.kmutnb.ac.th

### บทคัดย่อ

วิชาชีววิทยา ในหัวข้ออาณาจักรสัตว์ เป็นวิชาที่จำเป็นต้องใช้การจดจำเนื้อหาค่อนข้างมาก งานวิจัยนี้เป็นการนำเสนอสื่อการเรียนรู้ในรูปแบบเกมบนมือถือ สมาร์ทโฟน ซึ่งช่วยทำให้นักเรียนสามารถเข้าใจเนื้อหาได้ดียิ่งขึ้น ซึ่งผลการทดสอบประสิทธิภาพสื่อการเรียนรู้พบว่ามีประสิทธิภาพที่ค่า 82/86.5 ในส่วนผลการประเมินความพึงพอใจของผู้เรียนอยู่ในระดับพึงพอใจมากที่สุด โดยมีค่าเฉลี่ยอยู่ที่ 4.58

**คำสำคัญ:** สื่อการเรียนรู้ การเรียนรู้ด้วยเกม เกมอาณาจักรสัตว์

### Abstract

*Biology subject on the Taxonomy in Animal Kingdoms topic requires considerable memorization of content. This study proposes a game-based learning on smartphone. The game helps students better understand the contents. The results of this study show that the game's efficiency is effective at 82/86.5. The student satisfaction assessment result was at the highest level of satisfaction with an average of 4.58.*

**Keywords:** Learning Materials, Game-based learning, Animal Kingdom Game.

### 1. บทนำ

ทั่วโลกได้เข้าสู่ยุคดิจิทัลอย่างสมบูรณ์ ทำให้เกิดการเปลี่ยนแปลงอย่างรวดเร็วแบบไร้ขีดจำกัดในด้านเทคโนโลยี ส่งผลให้สังคมโลกเกิดการเปลี่ยนแปลงแบบ

ก้าวกระโดด และเทคโนโลยีได้เข้ามามีบทบาทในชีวิตประจำวันโดยเฉพาะในปัจจุบัน จากสถานการณ์การแพร่ระบาดของโรคติดต่อเชื้อไวรัสโคโรนา 2019 (COVID-19) ที่ทำให้ต้องมีมาตรการป้องกันโดยการให้ผู้คนกักตัวอยู่กับบ้าน ทำให้เป็นตัวเร่งให้ผู้คนต้องใช้เทคโนโลยีในการติดต่อสื่อสารกันหรือช่วยในการทำงานที่บ้านมากยิ่งขึ้น รวมถึงนักเรียนที่ต้องเรียนออนไลน์ ซึ่งทำให้เกิดความยุ่งยากในการมีส่วนร่วมในการเรียนหรือทำให้ความเข้าใจในเนื้อหาบทเรียนน้อยลง เพราะมีผลกระทบจากสภาพแวดล้อมที่บ้านที่จะทำให้ไม่เหมาะสมกับการเรียน เป็นต้น จากผลวิจัยชี้ว่าการเรียนออนไลน์ทำให้นักเรียนกว่าร้อยละ 20 ไม่เข้าร่วมชั้นเรียนเนื่องจากความเครียด [1] และจากงานวิจัย พบว่าการเรียนออนไลน์เป็นทางเลือกที่ใช้งานได้ภายใต้สถานการณ์ปัจจุบัน อย่างไรก็ตามการศึกษายังพบว่าควรให้อาจารย์ พยายามทำให้การศึกษาด้านออนไลน์เป็นที่ยอมรับในหมู่นักศึกษา โดยอาจมีการนำเทคนิคการเกมพีเลชั่นมาประยุกต์ใช้ในการศึกษา [2] โดยหนึ่งในวิชาที่มีเนื้อหาที่มีเนื้อหาเยอะและต้องใช้ความจำมากก็คือวิชาชีววิทยาซึ่งเป็นแขนงหนึ่งของวิชาวิทยาศาสตร์ธรรมชาติ เป็นการศึกษาเกี่ยวกับสิ่งมีชีวิตในธรรมชาติทั้งในอดีตจนถึงปัจจุบัน โดยจุดประสงค์ของวิชานี้คือการศึกษาเกี่ยวกับการจำแนกหรือการจัดหมวดหมู่ของสิ่งมีชีวิต ซึ่งต้องศึกษาในด้านต่าง ๆ ได้แก่ การศึกษาอาณาจักรสิ่งมีชีวิตนั้นมีอยู่หลากหลายแต่ผู้วิจัยได้เลือกศึกษาอาณาจักรสัตว์ โดยการศึกษาเรื่องความหลากหลายทาง

ชีวภาพมี เนื้อหาและหัวข้อในการเรียนที่ค่อนข้างซับซ้อนและเข้าใจยากและต้องใช้ความจำและทำความเข้าใจ เข้าใจ ทำให้ผู้เรียนรู้สึกไม่อยากเรียน และในสถานการณ์ในปัจจุบันทำให้ยิ่งต้องใส่ใจกับการเพิ่มความสนใจ การเรียนรู้ของนักเรียน ดังนั้น เพื่อแก้ปัญหาในเรื่องนี้ จึงทำให้เกิดแนวคิดในการนำมาเสริมการเรียนรู้ของนักเรียน ในปัจจุบันในการการพัฒนาสื่อใช้หลักการ ADDIE MODEL ในการออกแบบกระบวนการเรียนรู้และพัฒนาอย่างเป็นระบบที่ได้รับการยอมรับทั่วโลก โดยสื่อจะถูกออกแบบมาเพื่อความสนุกสนานและสอดแทรกความรู้ อีกทั้งยังปิดช่องโหว่ของข้อจำกัด การเรียนรู้ของนักเรียนที่เบี่ยงเบนจากการเรียน [3] อีกทั้งในการเรียนรู้ด้วยสื่อที่นั้นยังเหมาะสมกับการใช้ในยุคดิจิทัล เพราะปัจจุบันนักเรียนส่วนใหญ่สามารถเข้าถึงสื่อมัลติมีเดียหรือการนำเสนอสื่อผ่านทางแอปพลิเคชัน ได้ง่ายยิ่งขึ้นผ่านโทรศัพท์มือถือ จากการสำรวจพบว่า ประชาชนอายุ 15 – 24 ปี ใช้อุปกรณ์โทรศัพท์มือถือ ร้อยละ 98.8 [4] ซึ่งเป็นช่วงอายุในการเรียนมัธยมศึกษาตอนปลายตรงกับกลุ่มเป้าหมายทำให้ผู้วิจัยเลือกพัฒนาสื่อในรูปแบบเกมมือถือและในการทดลองครั้งนี้ จะเน้น การพัฒนา สื่อ และ ทดสอบ ประสิทธิภาพของสื่อทั้งด้านเนื้อหาและประสิทธิภาพของสื่อให้ได้ประสิทธิภาพของสื่อ

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

**2.1 ทฤษฎีการเรียนรู้** ทฤษฎีการเรียนรู้ของ Bloom (Bloom's taxonomy) – ความรู้ที่เกิดจากการจดจำ – ความเข้าใจ แปลความหมาย – การประยุกต์นำความรู้ไปใช้ – การวิเคราะห์ แยกแยะเรื่องราว หาความแตกต่างกันได้ – การสังเคราะห์ การนำความรู้ไปผสมผสานเรื่องราวให้เกิดสิ่งใหม่สมบูรณ์กว่าเดิม – การประเมินค่ามีความสามารถในการตัดสินใจได้อย่างเหมาะสม [5]

**2.2 เนื้อหาของวิชาชีววิทยา** ซึ่งเป็นเนื้อหาหลักในการพัฒนาสื่อ โดยใช้เฉพาะ เรื่องอนุกรมวิธานในหัวข้ออาณาจักรสัตว์ โดยมีเนื้อหาทั้งอาณาจักรสัตว์ยัง

แบ่งออกเป็น 9 ไฟล์ ได้แก่ 1) ไฟล์ พอริเฟอร่า ได้แก่ ฟองน้ำชนิดต่าง 2) ไฟล์ ซีเลนเทอรอลา ได้แก่ ไฮดรา แมงกะพรุน ปะการัง 3) ไฟล์ แพลทีเฮลมีนเทส ได้แก่ พยาธิใบไม้ พยาธิตัวดีด 4) ไฟล์ เนมาโทดา ได้แก่ พยาธิไส้เดือนตัวกลม พยาธิตัวจิ๋ว 5) ไฟล์ มอลลัสกา ได้แก่ หอย หมึก แมงเพรียง 6) ไฟล์ แอนเนลิดา ได้แก่ ไส้เดือนดิน แมงเพรียง 7) ไฟล์ อาร์โทรพอดา ได้แก่ กุ้ง ปู แมลง แมงมุม ตะขาบ 8) ไฟล์ เอไลโนเดอรัมาตา เช่น ปลิงทะเล เม่นทะเล 9) ไฟล์ คอร์ดาตา ได้แก่ ปลา กบ เต่า นก แมว คน [6]

**2.3 หลักการออกแบบกระบวนการเรียนรู้และพัฒนารูปแบบเป็นระบบ ADDIE MODEL** ประกอบด้วย 5 ขั้นตอน คือ 1) การวิเคราะห์ คือการกำหนดเป้าหมายของสื่อ ระดับการเรียนรู้ที่ต้องการ 2) ขั้นตอนการออกแบบ เป็นขั้นตอนการออกแบบการเรียนรู้ ออกแบบรูปแบบกระบวนการทั้งหมด ตามเป้าหมายที่กำหนดไว้ 3) การพัฒนา เป็นการสร้างโมเดลจากที่ได้ออกแบบไว้ให้เป็นไปตามที่กำหนด 4) การดำเนินการ เป็นการดำเนินการเรียนรู้ตามที่ออกแบบไว้ หรือนำบทเรียนหรือสื่อไปใช้ในการทดสอบ 5) การประเมินผล เป็นการประเมินการเรียนรู้ตามเป้าหมายที่กำหนดไว้ [3]

**2.4 การทดสอบประสิทธิภาพ นวัตกรรมหรือสื่อการสอน** ทดสอบประสิทธิภาพ (Developmental Testing) ใช้วิธีการทดสอบประสิทธิภาพ โดยการใช้สูตร E1/E2 สำหรับการทดสอบประสิทธิภาพของกระบวนการ (Process-E1) และทดสอบประสิทธิภาพของผลลัพธ์ (Product-E2) โดยการทดลองในการหาประสิทธิภาพของสื่อ นั้นเน้นความสัมพันธ์ของกระบวนการและผลลัพธ์ เป็นสำคัญ [7]

**2.5 งานวิจัยที่เกี่ยวข้อง** ด้านการพัฒนาสื่อส่งเสริมการเรียนรู้ นักวิจัยหลากหลายท่านที่ศึกษา

และนำไปประยุกต์ใช้ เช่น วินัยนันท์[8] ได้พัฒนาแอปพลิเคชันสำหรับการเรียนรู้เรื่องเพศศึกษาและทักษะชีวิตของนักเรียนชั้นมัธยมศึกษาปีที่ 3 โดยมีการนำสื่อมาพัฒนาโดยการออกแบบด้วยหลักการ ADDIE MODEL มาออกแบบ สื่อและนำไปทดสอบกับนักเรียน โดยแบ่งกลุ่มควบคุมกับกลุ่มทดลองพบว่านักเรียนทั้งสองกลุ่มมีผลการเรียนรู้ไม่แตกต่างกัน [9] ได้ใช้บทเรียนคอมพิวเตอร์ช่วยสอนแบบฝึกทักษะการใช้ “Verb to be” สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 โดยบทเรียนที่พัฒนาและแบบทดสอบวัดผลทดสอบระหว่างเรียนและหลังเรียนพบว่าประสิทธิภาพของบทเรียนที่พัฒนาขึ้นมีประสิทธิภาพ 84.44/83.11 สูงกว่าเกณฑ์มาตรฐาน 80/80 [10] การพัฒนาสื่อการเรียนการสอนทักษะแห่งศตวรรษที่ 21 สาระเศรษฐศาสตร์ โดยศึกษาการพัฒนาสื่อให้ได้ตามเกณฑ์มาตรฐาน 80/80 โดยพบว่าร้อยละมีค่า E1/E2 = 82.07/93.70 สูงกว่าเกณฑ์มาตรฐานที่กำหนดไว้ [11] การพัฒนาสื่อการสอนมัลติมีเดีย เรื่องอาเซียน ตามแนวคิดทฤษฎีพฤติกรรมนิยม โดยใช้หลักการออกแบบด้วย ADDIE MODEL ในการพัฒนาสื่อและประเมินคุณภาพสื่อการเรียนการสอนมัลติมีเดียและแบบวัดผลสัมฤทธิ์ทางการเรียน

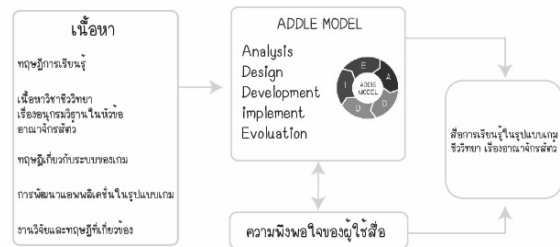
### 3. วิธีดำเนินการวิจัย

#### 3.1 ศึกษาและรวบรวมข้อมูล

ศึกษารายละเอียดของเนื้อหาของวิชาชีววิทยาเฉพาะเรื่องอนุกรมวิธานหัวข้ออาณาจักรสัตว์ และนำทฤษฎีการเรียนรู้ของ Bloom Taxonomy มาประยุกต์ใช้กับการจัดการเนื้อหาที่นำเสนอ และใช้หลักการ ADDIE MODEL ในการพัฒนาสื่อการเรียนรู้ ศึกษาหลักการออกแบบและพัฒนาสื่อ ต่างๆ และศึกษาการทดสอบประสิทธิภาพของสื่อ

### 3.2 การออกแบบกระบวนการเรียนรู้ ADDIE

#### MODEL



ภาพที่ 1: กรอบแนวคิดการพัฒนาสื่อ

โดยนำเนื้อหาที่ศึกษามาออกแบบตามขั้นตอน ADDIE MODEL ทั้ง 5 ดังนี้ 1) การวิเคราะห์ (Analysis) หลังจากการรวบรวมข้อมูล กำหนดความรู้ที่ได้รับของนักเรียน โดยออกแบบสื่อเป็นรูปแบบของเกม โดยใช้ทฤษฎีการเรียนรู้ของบลูม Bloom Taxonomy ในการกำหนดความรู้ตามระดับการเรียนรู้สำหรับนักเรียนชั้นมัธยมปลาย สายสามัญ วิทยาศาสตร์และคณิตศาสตร์ โดยกำหนดการเรียนรู้ให้ผู้เรียนสามารถจดจำ เข้าใจ นำข้อมูลไปใช้ได้จริง และสามารถวิเคราะห์ จัดหมวดหมู่ ของสิ่งมีชีวิตประเภทสัตว์ได้ โดยสื่อการเรียนรู้ ออกแบบให้ผู้เรียนวิเคราะห์ว่าสัตว์ที่มีลักษณะจำเพาะแบบนี้เป็นสัตว์ชนิดไหน เพื่อให้ผู้เรียนใช้ทักษะการวิเคราะห์และจำแนกจากข้อมูลที่มีและสามารถนำความรู้หรือข้อมูลไปศึกษากับสิ่งมีชีวิตจริงได้ อีกทั้งยังทำให้ผู้เรียนสามารถจดจำลักษณะจำเพาะของสัตว์สัตว์ตัวนั้นได้ และสามารถจำแนกสัตว์ชนิดนั้นได้ 2) การออกแบบ (Design) โดยได้ออกแบบให้เกมมีรูปแบบการเล่นแบบการ์ดเกม โดยการกำหนดชนิดสัตว์โดยเลือกจากทั้ง 9 ไฟล์และคลาสในไฟล์หลักที่นักเรียนเรียนมาเป็นตัวอย่าง 60 ตัว และกำหนดลักษณะจำเพาะของสัตว์ที่ครอบคลุมสัตว์ทั้งหมดเช่น ลักษณะการสืบพันธุ์ ลักษณะการเคลื่อนที่ ลักษณะการหายใจ จำนวนของขาหรืออวัยวะภายในและภายนอก และอื่น ๆ รวมทั้งสิ้น 30 ลักษณะ เพื่อให้ผู้เรียนสามารถใช้จำแนกความแตกต่างสัตว์ได้

ครอบคลุมทุกชนิด โดยใช้สัตว์แต่ละชนิดมาออกแบบเป็นการ์ดคำถาม และใช้ลักษณะจำเพาะของสัตว์แต่ละชนิดมาเป็นการ์ดคำใบ้ให้กับผู้เล่น ซึ่งรูปแบบการเล่นจะมีความเรียบง่าย โดยเกมจะเริ่มจากให้การ์ดคำถามมาหนึ่งใบ ซึ่งก็คือการ์ดสัตว์ที่เป็นปริศนาไว้ และผู้เล่นเพียงเลือกการ์ดคำใบ้หรือลักษณะจำเพาะของสัตว์ที่ต้องการถาม แล้วระบบจะทำการแสดงลักษณะที่ถูกต้องมาให้ผู้เล่น โดยเมื่อผู้เล่นรวบรวมลักษณะจำเพาะของสัตว์และวิเคราะห์ได้แล้วว่าเป็นสัตว์ชนิดใด ผู้เล่นก็สามารถเลือกภาพสัตว์ที่ถูกต้องให้กับระบบ หากถูกต้องผู้เล่นก็จะได้รับคะแนนในรูปแบบดาวไป หากตอบผิดผู้เล่นก็จะโดนหักคะแนนหรือดาวไป 3) **การพัฒนา (Development)** การพัฒนาสื่อด้วยโปรแกรมสำเร็จรูป Unity โดยใช้สร้างโครงสร้างระบบของเกมหลักที่ออกแบบมาจากขั้นตอนออกแบบ (Design) และจัดทำกราฟิกภายในเกมทั้งการวาดภาพสัตว์ที่ประกอบภายในเกมและออกแบบสร้าง UI ของเกม ด้วยโปรแกรมสำเร็จรูป Krita, illustrator รวมถึงการประเมินประสิทธิภาพสื่อของผู้เชี่ยวชาญก่อนนำไปทดสอบกับกลุ่มเป้าหมายด้วยเพื่อวัดว่าสื่อสามารถนำไปทดสอบในขั้นตอนต่อไปได้หรือไม่ 4) **ขั้นตอนการนำไปใช้ (Implement)** โดยวิธีการทดสอบตามหลักการทดสอบประสิทธิภาพของ นวัตกรรมหรือสื่อการสอน โดยมีการจัดทำแบบทดสอบในรูปแบบออนไลน์ด้วยเว็บแอปพลิเคชันสำเร็จรูป Quizizz โดยรูปแบบคำถามจะอ้างอิงจะหนังสือชีววิทยา ซึ่งชุดแบบทดสอบจัดทำมาทั้งหมด 20 ข้อ และทำการแบ่งชุดแบบทดสอบเป็น 2 ช่วง ดังนี้ 1.) เป็นการทดสอบระหว่างเล่นเพื่อวัดประสิทธิภาพของกระบวนการ (E1 Efficiency of process) ให้นำ 1 แบบทดสอบ 20 ข้อมาแบ่งเป็น 2 ช่วงโดยใช้เกณฑ์ 10 ข้อ 15 คะแนน โดยจะมี 5 ข้อในคำถามที่ยากให้ 2 คะแนน รวมทั้ง 2 ช่วงจะรวมคะแนนเต็ม 30 คะแนน

และ 2) การทดสอบหลังใช้สื่อเพื่อวัดประสิทธิภาพของผลลัพธ์ (E2 Efficiency of Product) โดยใช้แบบทดสอบเดิมแต่รวมทั้งหมด 20 ข้อ แล้วทำการสลับข้อ โดยจะให้คะแนนข้อละ 1 คะแนน ดังนั้นการทดสอบประสิทธิภาพระหว่างเล่นจะกำหนดให้เป็น ซึ่งการทดสอบจะทดสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 10 คน โดยวิธีการเลือกแบบเจาะจง 5) **การประเมินผล (Evaluation)** จากการทดสอบประสิทธิภาพของสื่อ ในรอบสุดท้ายเป็นการทดสอบหลังจากใช้สื่อว่าผู้เล่นจดจำ ความรู้ที่ได้ใช้สื่อมาน้อยแค่ไหนโดยมีการกำหนดมาตรฐานประสิทธิภาพสื่อไว้ที่ 80/80 หมายถึงคะแนนระหว่างใช้สื่อและคะแนนหลังใช้สื่อมาเปรียบเทียบกัน ซึ่งหากคะแนนได้ 78/79 แสดงว่าประสิทธิภาพสื่อต่ำกว่าเกณฑ์ หากได้ 80.5/81.5 แสดงว่าสื่อมีประสิทธิภาพตามเกณฑ์ที่กำหนด และหากคะแนน 95/96 แสดงว่าสื่อมีประสิทธิภาพสูงกว่าเกณฑ์ที่กำหนด

#### 4. ผลการดำเนินงาน

จากการศึกษาและพัฒนาสื่อการเรียนรู้ในรูปแบบเกมมือถือ เรื่องอนุกรมวิธาน หัวข้ออาณาจักรสัตว์ ได้พัฒนาตัวสื่อในรูปแบบเกมในมือถือ ซึ่งประกอบไปด้วยส่วนการทำงานอยู่ 3 อย่าง ดังนี้ 1) ฟังก์ชัน Game เป็นฟังก์ชันในการเล่นเกมน โดยเริ่มแรก ระบบจะแสดงการ์ดคำถามสัตว์เป็นรูปแบบเครื่องหมายคำถาม และจะแสดงคำใบ้ลักษณะเฉพาะของสัตว์ตัวนั้นให้ 3 ข้อ หลังจากนั้นผู้เล่นสามารถเลือกการ์ดคำใบ้ที่เป็นลักษณะของสัตว์ที่ผู้เล่นต้องการรู้ โดยการลากการ์ดคำใบ้ไปใส่ที่การ์ดคำถาม ระบบก็จะแสดงคำใบ้ต่อไปมาในช่องที่แสดงคำใบ้ เมื่อผู้เล่นตัดสินใจว่าจะตอบว่าสัตว์คืออะไร ผู้เล่นสามารถปิดแถบสีเทาด้านขวาของจอ หน้าต่างจะแสดงภาพสัตว์ให้เลือกตอบ โดยผู้เล่นสามารถลากภาพสัตว์ไปใส่การ์ดคำถามได้ เมื่อถูกผู้เล่นจะได้รับคะแนนเป็นรูปแบบของดาว แล้วระบบจะแสดงการ์ดสัตว์ต่อไป 2)

Encyclopedia เป็นฟังก์ชันการดูเนื้อหาสารานุกรม โดยจะแสดงข้อมูลลักษณะจำเพาะของสัตว์เพิ่มเติมจากเกม รวมถึงข้อมูลไฟล์หลักของสัตว์ 9 ไฟล์หลัก และ 3) Trophy เป็นฟังก์ชันผลคะแนนจากการเล่นเกมที่สะสมมาในรูปแบบดาว โดยคะแนนสูงสุดเท่ากับ 3 ดาว น้อยที่สุดจะไม่มีดาว



ภาพที่ 2: ตัวอย่างหน้าจอการเรียนรู้

ผลประเมิน ประสิทธิภาพ ของสื่อการเรียนรู้ โดยผู้เชี่ยวชาญ 3 ท่าน โดยมีการประเมินอยู่ 2 ด้าน ได้แก่ ด้านเนื้อหาของสื่อและคุณภาพของสื่อ พบว่าสื่อมีประสิทธิภาพอยู่ในระดับมาก ที่ค่าเฉลี่ย 4.22 ค่าเบี่ยงเบนมาตรฐานที่ระดับ 0.60

ตารางที่ 1: ผลการประเมินประสิทธิภาพสื่อการเรียนรู้ของผู้เชี่ยวชาญ 3 ท่าน

รายการประเมิน	$\bar{X}$	S.D.	ระดับคุณภาพ
1. ด้านเนื้อหา	4.30	0.42	มาก
2. ด้านคุณภาพของสื่อ	4.15	0.78	มาก
รวม	4.22	0.60	มาก

ผลการทดสอบประสิทธิภาพสื่อ โดยกลุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 10 คน โดยผลการทดสอบตั้งมาตรฐาน (E1/E2) ไว้ที่ 80/80 โดยที่ E1 คือคะแนนระหว่างใช้สื่อที่แบ่งเป็นการทดสอบ 2 ช่วง โดยใช้การทำแบบทดสอบความจำจากการเล่นเกมโดยตอบคำถามสัตว์จำนวน 10 ตัว แรกในเกม และรอบที่สอง อีก 10 ตัวต่อมาในเกม โดยผลคะแนนระหว่างใช้สื่อจะมีคะแนนเต็มรวม ที่ 30 คะแนน ช่วงแรก 15 คะแนน ช่วงสอง 15 คะแนน โดยการทดสอบนี้ใช้ผู้ทดสอบจำนวน 10 คน รวมทั้งสิ้น 300 คะแนน และ E2 เป็นการเก็บคะแนนหลังจากเล่นเกม โดยการรวมแบบทดสอบใน

ช่วงแรกแล้วมาสลับข้อคำถามเพื่อวัดประสิทธิภาพหลังใช้สื่อ โดยกำหนดคะแนนเต็มไว้ที่ 20 คะแนน ทดสอบการกลุ่มทดสอบเดิม 10 คน รวมทั้งสิ้น 200 คะแนน ซึ่งผลคะแนนทำแบบทดสอบออกมาที่คะแนน 82/86.5 สรุปได้ว่า การทดสอบประสิทธิภาพของสื่อในการเรียนรู้ในรูปแบบเกมมือถือ เรื่องอนุกรมวิธานในหัวข้ออาณาจักรสัตว์ สื่อมีประสิทธิภาพ

ตารางที่ 2: ผลการทดสอบประสิทธิภาพของสื่อ โดยกลุ่มเป้าหมาย

นักเรียน	คะแนนระหว่างทดสอบ (E1)	คะแนนหลังทดสอบ
10 (คน)		
รวม	248	173

ผลการประเมินความพึงพอใจของผู้ใช้สื่อจากกลุ่มตัวอย่างพบว่า ความพึงพอใจของผู้ใช้สื่ออยู่ในระดับมากที่สุดที่ค่าเฉลี่ย 4.58 ค่าเบี่ยงเบนมาตรฐานที่ระดับ 0.54

ตารางที่ 3: ผลประเมินความพึงพอใจ

รายการประเมิน	$\bar{X}$	S.D.
1. ด้านเนื้อหา	4.72	0.45
2. ด้านคุณภาพของสื่อ	4.49	0.58
3. ด้านประโยชน์จากสื่อ	4.55	0.50
รวม	4.58	0.54

### 5. สรุปและอภิปรายผล

การพัฒนาสื่อการเรียนรู้ในรูปแบบเกมมือถือในหัวข้ออนุกรมวิธาน เรื่องอาณาจักรสัตว์ เป็นการพัฒนาเพื่อศึกษาการพัฒนาสื่อการเรียนรู้และทดสอบประสิทธิภาพของสื่อว่ามีคุณภาพตามเกณฑ์หรือไม่ โดยใช้หลักการออกแบบกระบวนการเรียนรู้และพัฒนาสื่อการสอน ADDIE MODEL ตัวสื่อจะมีฟังก์ชัน 3 อย่าง ได้แก่ ฟังก์ชันเกม หรือการเข้าเล่นเกมหลักโดยระบบเกมจะแสดงการ์ดสัตว์ปริศนาและแสดงลักษณะจำเพาะของสัตว์ในเชิงคำใบ้ โดยผู้เล่นมีหน้าที่จำแนกข้อมูลตามคำใบ้เพื่อหาคำตอบที่ถูกต้อง

ของสัตว์ตัวนั้น ๆ หากตอบถูกระบบก็จะทยอยตัวต่อไปจนครบ 60 ตัว ฟังก์ชันสารานุกรมและฟังก์ชันรางวัลมีไว้สำหรับข้อมูลเพิ่มเติมของสื่อเท่านั้น

โดยจากการวิจัยและพัฒนาสื่อได้นำไปประเมินประสิทธิภาพกับผู้เชี่ยวชาญ โดยประสิทธิภาพของสื่ออยู่ในระดับที่ดีและได้รับคำติชมจากผู้เชี่ยวชาญว่า ภาพรวมของสื่อการเรียนรู้ที่เหมาะสมและน่าสนใจที่จะนำไปใช้ในการเรียนการสอนต่อไปได้ โดยเฉพาะรูปภาพของสัตว์ซึ่งเป็นสื่อหลักออกแบบมาได้เป็นอย่างดีได้มีการทดสอบประสิทธิภาพของสื่อซึ่งมี ประสิทธิภาพของสื่ออยู่ในระบบมาตรฐานตามเกณฑ์ที่กำหนด สอดคล้องกับงานวิจัยของ อิกัลส[9] สรรสยา[10] และยุทธพงษ์[11] ที่ได้ผลลัพธ์ในทิศทางเดียวกัน การประเมินความพึงพอใจของผู้ใช้ทั้งหมดอยู่ในระดับมากที่สุด ข้อเสนอแนะในการศึกษาครั้งต่อไปในการพัฒนาต่อไปในอนาคต สามารถนำสื่อการเรียนรู้ไปประยุกต์ใช้กับการเรียนการสอนได้เพื่อเพิ่มความสนใจให้นักเรียนได้ อีกทั้งยังสามารถนำสื่อไปต่อยอดเป็นผลิตภัณฑ์เพื่อพัฒนาความรู้ของประชาชนที่สนใจในเรื่องชีววิทยา หัวข้อเกี่ยวกับอาณาจักรสัตว์ได้อีกด้วย

### เอกสารอ้างอิง

- [1] สยามรัฐ. ผลวิจัยชี้เรียนออนไลน์ทำเด็กเครียดจนโดดเรียน 20% การเรียนถดถอย 50%. (2564). [ออนไลน์]. [สืบค้นวันที่ 23 ธันวาคม 2564]. จาก <https://siamrath.co.th/n/269568>
- [2] Pinaki Chakraborty, Prabhat Mittal, Manu Sheel Gupta, Savita Yadav and Anshika Arora], “Opinion of students on online education during to COVID-19 Pandemic”, *Human Behavior and Emerging Technologies*, vol.3, Issue 3, pp. 357-365, 2020.
- [3] PEOPLE VALUE. ADDIE MODEL คืออะไร. (2564). [ออนไลน์]. [สืบค้นวันที่ 23 ธันวาคม 2564]. จาก <https://www.peoplevalue.co.th/content/9119/addie-model-คืออะไร>.
- [4] สำนักงานสถิติแห่งชาติ. *สรุปผลที่สำคัญสำรวจการมีการใช้เทคโนโลยีสารสนเทศและการสื่อสารในครัวเรือน พ.ศ. 2563*. สำนักงานสถิติแห่งชาติ กรุงเทพมหานคร, 2564.
- [5] Anyflip. ทฤษฎีการเรียนรู้ Bloom's taxonomy. (2564) [ออนไลน์]. [สืบค้นวันที่ 30 ธันวาคม 2564]. จาก <https://anyflip.com/tkhwo/dvyw/basic>
- [6] ประสงค์ หล้าสะอาด และ จิตเกษม หล้าสะอาด *ชีววิทยา ม.6 เล่ม 6* สำนักพิมพ์ พ.ศ. พัฒนาจำกัด กรุงเทพมหานคร. 2560.
- [7] ชัยยงค์ พรหมวงศ์ “การทดสอบประสิทธิภาพ สื่อหรือชุดการสอน”,วารสารศิลปการศึกษาศาสตร์วิจัย ปีที่ 5 ฉบับที่ 1, 2556.
- [8] วินชนันท์ นวลคา และ พิศนันท ชูโตศรี “การพัฒนาแอปพลิเคชัน ส่งเสริมการเรียนรู้เรื่องเพศศึกษาและทักษะชีวิต ในนักเรียนแบบบูรณาการ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3”, การประชุมวิชาการนำเสนอผลงานวิจัยระดับชาติของนักศึกษาด้านมนุษยศาสตร์และสังคมศาสตร์ ครั้งที่ 2 วันที่ 19 มกราคม 2562 ณ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา, หน้า 1782-1795, 2562.
- [9] อิกัลส อูมา. *การใช้บทเรียนคอมพิวเตอร์ช่วยสอนแบบฝึกทักษะการใช้ “Verb to be” ในวิชาภาษาอังกฤษสำหรับนักเรียนชั้นมัธยมศึกษาปีที่1*, สาขาวิชาเทคโนโลยีและสื่อสารการศึกษามหาวิทยาลัยสงขลานครินทร์, 2554.
- [10] สรรสยา ชันวิเศษ, ผศ.ดร. ประยูร แสงใส “การพัฒนาสื่อการเรียนการสอนทักษะแห่งศตวรรษที่ 21 สาขาสหศาสตร์ โรงเรียนเขาวไร่ศึกษา อำเภอโกสัมพินัย จังหวัดม.ห.ส.ร.ค.ม.” *Journal of Buddhist Education and Research : JBER*, Vol.7 No.2, pp. 160-170, 2564.
- [11] ยุทธพงษ์ สีลาขวา ฐิติชญา หมูสี และ นิพล สังสุทธิ การพัฒนาสื่อการเรียนการสอนมัลติมีเดีย เรื่องสัตว์สี่ขาเขี้ยวตามแนวคิดทฤษฎีพฤติกรรมนิยม. วารสารโครงการงานวิชาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏอุดรธานี, 2558



# การออกแบบประสบการณ์ผู้ใช้สื่อการสอนประเภทความจริงเสริม วิชาภาษาญี่ปุ่นในระดับเริ่มต้น เพื่อเพิ่มผลสัมฤทธิ์ทางการเรียนรู้ User Experience design of Augmented Reality teaching materials for basic level Japanese subjects to Increase Learning Achievement

นิรัชพร อุ่นนิมิตร (Niratchaporn Ounnimit)<sup>1</sup> และอภิชญา นิมคุ้มภัย (Apichaya Nimkoompai)<sup>2</sup>  
วิทยาศาสตร์มหาบัณฑิต (สาขาเทคโนโลยีสารสนเทศ) คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยี ไทย – ญี่ปุ่น  
<sup>1</sup>ou.niratchaporn\_st@tmi.ac.th, <sup>2</sup>apichaya@tmi.ac.th

## บทคัดย่อ

ปัจจุบันคนไทยเริ่มเข้ามาทำธุรกิจในประเทศไทยใน  
ด้านอุตสาหกรรมต่าง ๆ จำนวนมาก คนที่มีความสามารถ  
ในการสื่อสารภาษาญี่ปุ่นจึงมีโอกาสเติบโตและโอกาสที่  
จะได้รับการคัดเลือกเข้าทำงานสูงกว่าแต่มีคนจำนวนมาก  
ที่มีปัญหาในการเรียนรู้ภาษาญี่ปุ่น เนื่องจากสถานการณ์  
COVID-19 ทำให้สื่อการสอนรูปแบบเดิมใช้งานได้ไม่เต็ม  
ประสิทธิภาพและทำให้ความน่าสนใจในการเรียนลดลง  
ผู้วิจัยจึงนำเทคโนโลยีความจริงเสริมร่วมกับหลักการ  
วิเคราะห์ประสบการณ์ผู้ใช้มาพัฒนาสื่อการสอน  
ภาษาญี่ปุ่นในระดับเริ่มต้น เพื่อเพิ่มผลสัมฤทธิ์ทางการ  
เรียนให้มากขึ้น ผลการทดลองพบว่า กลุ่มที่ใช้สื่อการสอน  
ประเภทความจริงเสริมในการเรียนมีอัตราการเพิ่มขึ้นของ  
คะแนน 73.68% ส่วนกลุ่มที่เรียนแบบปกติมีอัตราการ  
เพิ่มขึ้นของคะแนน 62.27% แสดงให้เห็นว่าการใช้สื่อการ  
สอนประเภทความจริงเสริมในการเรียนภาษาญี่ปุ่นช่วยให้  
ผู้เรียนเข้าใจบทเรียนและมีผลการเรียนที่ดีขึ้น

**คำสำคัญ:** สื่อการสอน วิชาภาษาญี่ปุ่น ความจริงเสริม  
ประสบการณ์ผู้ใช้

## Abstract

At present, Japanese people are active in  
Thailand in many industries. People who can  
communicate in Japanese have a chance to grow and  
have a higher chance of being recruited for a job.  
But many people have trouble learning Japanese.  
Due to the COVID-19 situation, traditional teaching  
materials are ineffective and decrease their interest

in learning. Therefore, the researcher has applied  
augmented reality technology together with the  
principle of user experience analysis to develop  
Japanese language teaching materials at the  
beginner level to increase academic achievement  
more. The results showed that The group using  
augmented reality teaching materials in learning  
Japanese had a score of 73.68%, while the group  
that did not use augmented reality teaching material  
had a score increase of 62.27%. It has been shown  
that the use of augmented reality teaching materials  
in learning Japanese helps learners understand the  
lesson and achieve better results.

**Keywords:** Teaching materials, Japanese subjects,  
Augmented Reality, User Experience.

## 1. บทนำ

ในปัจจุบันนักลงทุนญี่ปุ่นที่เข้ามาลงทุนในประเทศไทยมีค่อนข้างมาก สำนักงานคณะกรรมการส่งเสริมการ  
ลงทุน (BOI) เปิดเผยว่า มูลค่าขอรับการส่งเสริมการ  
ต่างประเทศ (FDI) ในช่วงเดือน ม.ค. - ก.ย. ปี 2564  
ประเทศที่มีมูลค่าลงทุนสูงสุดได้แก่ ประเทศญี่ปุ่น มูลค่า  
ลงทุน 67,816 ล้านดอลลาร์ คิดเป็น 18% ของมูลค่าเงินลงทุน  
ทั้งหมด [1] ซึ่งทำให้คนไทยญี่ปุ่นได้เข้ามาทำธุรกิจใน  
ประเทศไทยในอุตสาหกรรมต่าง ๆ โดยการติดต่อทาง  
ธุรกิจ การค้า หรือการทำงานร่วมกับองค์กรญี่ปุ่น คนที่มี  
ความสามารถในการสื่อสารภาษาญี่ปุ่นจะได้รับความ  
ไว้วางใจมากกว่า มีโอกาสเติบโตมากกว่า และมีโอกาสที่

จะรับการคัดเลือกเข้าทำงานสูงกว่า เพราะในปัจจุบันมีงานที่จำเป็นต้องใช้ความสามารถในการสื่อสารภาษาญี่ปุ่นค่อนข้างมาก แต่บุคคลากรที่สามารถสื่อสารภาษาญี่ปุ่นได้ในตอนนี้ถือว่ายังมีน้อยกว่าความต้องการของตลาดอยู่มาก ดังนั้นคนที่สามารถสื่อสารภาษาญี่ปุ่นได้จึงมีโอกาสตกงานน้อยมาก

อย่างไรก็ตามมีคนจำนวนมากที่เรียนภาษาญี่ปุ่นแล้วรู้สึกว่าภาษาญี่ปุ่นเป็นภาษาที่ยาก ไม่สามารถสื่อสารภาษาญี่ปุ่นได้อย่างถูกต้อง จากการสำรวจปัญหาในการใช้ภาษาญี่ปุ่นที่พบในการฝึกงานพบว่า กลุ่มเป้าหมายประสบปัญหาด้านคำศัพท์มากที่สุด ทั้งในเรื่องของคำศัพท์เฉพาะทางที่ใช้ในการปฏิบัติงาน และการเลือกใช้คำศัพท์ให้เหมาะสมกับสถานการณ์ต่าง ๆ [2] และนอกจากนี้สื่อการสอนในรูปแบบเดิมอาจจะไม่ช่วยสร้างความน่าสนใจในการเรียนที่เพียงพอให้กับผู้เรียนได้ ประกอบกับสถานการณ์การแพร่ระบาดของเชื้อไวรัสโคโรนา 2019 (COVID-19) ในปัจจุบัน ทำให้สื่อการสอนในรูปแบบเดิมไม่สามารถใช้งานได้เต็มประสิทธิภาพ

จากปัญหาที่กล่าวมาข้างต้น ผู้วิจัยจึงมีแนวคิดในการนำเทคโนโลยีความจริงเสริม (Augmented Reality : AR) เข้ามาประยุกต์ใช้ในการพัฒนาสื่อการสอนภาษาญี่ปุ่นระดับเริ่มต้น เพื่อใช้เป็นเครื่องมือให้กับผู้ที่เริ่มต้นศึกษาภาษาญี่ปุ่น เนื่องด้วยสื่อประเภท AR สามารถใช้งานได้บนอุปกรณ์สมาร์ตโฟน ซึ่งเหมาะกับการเรียนรู้ได้ทุกที่ทุกเวลา อีกทั้งยังช่วยสร้างประสบการณ์ที่แปลกใหม่ให้กับผู้เรียน และเป็นเทคโนโลยีที่ถูกนำมาใช้กันอย่างแพร่หลายในปัจจุบัน เพราะการใช้งานที่ค่อนข้างสะดวกและใช้อุปกรณ์ไม่เยอะ ทำให้ผู้เรียนสามารถเรียนรู้ได้สะดวกมากขึ้นและได้รับประสบการณ์ในการฝึกทักษะภาษาญี่ปุ่นในรูปแบบใหม่ ซึ่งจะช่วยสร้างความสนใจให้กับผู้เรียนมากขึ้น อีกทั้งยังส่งผลต่อผลสัมฤทธิ์ในการเรียนอีกด้วย

## 2. วรรณกรรมที่เกี่ยวข้อง

### 2.1 ความจริงเสริม (Augmented Reality : AR)

Augmented Reality (AR) หรือความจริงเสริม เป็นเทคโนโลยีที่มีผู้ให้นิยามความหมายไว้หลากหลายนิยาม

ด้วยกัน โดยงานวิจัยของ Ronald ได้ให้นิยามว่า AR เป็นเทคโนโลยีที่ผสมผสานโลกเสมือนเข้ากับโลกความจริง โดยใช้วัตถุเสมือนมาวางอยู่บนสภาพแวดล้อมจริง [3] ในขณะที่ Grubert และคณะได้ให้ความหมายของ AR ว่าเป็นเทคนิคที่ช่วยให้ผู้ใช้งานสามารถโต้ตอบกับสภาพแวดล้อมทางกายภาพผ่านการซ้อนทับข้อมูลดิจิทัล [4] และงานวิจัยของ Feng Zhou ได้อธิบายความหมายว่า AR เป็นเทคโนโลยีที่ช่วยให้คอมพิวเตอร์สร้างภาพเสมือนจริงขึ้นมา ซึ่งภาพที่สร้างขึ้นจะซ้อนทับกับวัตถุจริงในเวลาจริง โดยผู้ใช้จะเข้าไปอยู่ในสภาพแวดล้อมเสมือนจริงและสามารถโต้ตอบกับภาพเสมือนที่สร้างขึ้นได้ [5] ซึ่งสอดคล้องกับงานวิจัยของพนิดา ที่ได้อธิบายว่า AR เป็นการนำระบบความจริงเสมือนมาผนวกกับเทคโนโลยีภาพเพื่อสร้างสิ่งเสมือนจริงให้กับผู้ใช้ [6]

โดยสรุปแล้ว เทคโนโลยีความจริงเสริมหมายถึงเทคโนโลยีที่ผสมผสานระหว่างโลกของความจริงกับโลกเสมือนจริงให้เสมือนกับว่ามีวัตถุหรือมีเหตุการณ์เกิดขึ้นจริงบนสภาพแวดล้อมปัจจุบัน โดยผ่านซอฟต์แวร์และอุปกรณ์เชื่อมต่อต่าง เพื่อให้ผู้ใช้งานโต้ตอบกับภาพเสมือนที่ปรากฏอยู่ได้

เทคโนโลยี AR แบ่งออกเป็น 2 ประเภท ได้แก่ 1. แบบที่ใช้ภาพสัญลักษณ์ และ 2. แบบที่ใช้ระบบพิกัด ในการวิเคราะห์ข้อมูลเพื่อสร้างข้อมูลบนโลกเสมือนจริง ซึ่งภาพสัญลักษณ์จะนิยมเรียกว่า Maker หรือ AR Code โดยใช้ง็ล้องเว็บแคมในการรับภาพ เมื่อซอฟต์แวร์ที่ใช้งานประมวลผลรูปภาพพบสัญลักษณ์ที่กำหนดไว้ จะแสดงข้อมูลภาพสามมิติที่ตั้งค่าไว้ใน โปรแกรมให้เห็น ทำให้ผู้ใช้งานสามารถที่จะหมุนดูภาพที่ปรากฏได้ทุกทิศทาง

ปัจจุบัน AR มีความนิยมมากขึ้นเรื่อย ๆ จะเห็นว่ามีการนำเทคโนโลยีนี้ไปประยุกต์ใช้ในด้านต่าง ๆ เช่น อุตสาหกรรม การแพทย์ การตลาด รวมทั้งด้านการศึกษาด้วย [7] มีการศึกษาเกี่ยวกับการนำ AR มาใช้ในการสร้างสื่อการสอนในรูปแบบต่าง ๆ จากงานวิจัยของ Dita เขาได้นำ AR มาใช้ในการพัฒนาแอปพลิเคชันสื่อการเรียนรู้อภาษาต่างประเทศ โดยใช้มือถือในการสแกนอักษรแล้วประมวลผลออกมาเป็นรูปภาพ [8] และงานวิจัยของ Geng

และ Yamada ได้นำเทคโนโลยี AR มาพัฒนาระบบที่ช่วยทำความเข้าใจและเรียนรู้ศัพท์กริยาประสมในภาษาญี่ปุ่นที่มีความซับซ้อนมาก โดยจะใช้มือถือสแกนไปที่บัตรคำศัพท์ จะมีตัวละครนำคำศัพท์ไปประสมให้แล้วประมวลผลออกทางหน้าจอ [9]

จากงานวิจัยของ Plecher และคณะได้พัฒนาเกมแนวผจญภัยเพื่อเรียนรู้เกี่ยวกับตัวอักษรคันจิในภาษาญี่ปุ่น โดยเกมหลักจะเป็นการต่อสู้แบบผลัดกันเล่นและมีเกมย่อย ๆ เพื่อฝึกฝนเกี่ยวกับตัวอักษรคันจิในด้านต่าง ๆ ซึ่งผลลัพธ์โดยรวมมีแนวโน้มไปในทางที่ดี แต่ยังมีข้อวิพากษ์วิจารณ์อยู่ คือตัวอักษรคันจิยังมีจำนวนที่จำกัดอยู่ และยังไม่สามารถที่จะเรียนรู้ภาษาญี่ปุ่นตั้งแต่เริ่มต้นได้ [10]

โดยส่วนมากงานวิจัยที่เกี่ยวข้องกับ AR จะมุ่งเน้นไปที่การสร้างแอปพลิเคชันที่ใช้มือถือสแกนตัวอักษรหรือรูปภาพ แล้วแสดงผลออกทางหน้าจอ เพื่อเสริมสร้างความน่าสนใจให้ผู้ใช้งานและเป็นตัวช่วยในการพัฒนาการเรียนการสอนให้ดียิ่งขึ้น

**2.2 ประสบการณ์ของผู้ใช้ (User Experience : UX)**

User Experience (UX) หรือประสบการณ์ของผู้ใช้ คือประสบการณ์ของผู้ใช้งาน ในด้านความรู้สึกของผู้ใช้งาน (User) ที่มีต่อการใช้งานผลิตภัณฑ์ หรือระบบต่าง ๆ เช่น ความสะดวกสบาย ใช้งานง่าย ความสนุกสนาน จนเกิดเป็นความพึงพอใจสูงสุดหรือเกิดประสบการณ์ที่ดีของผู้ใช้งาน ในอีกแง่หนึ่ง UX เป็นผลจากวิวัฒนาการของ User Interface (UI) เมื่อผู้ใช้งานมีการปฏิสัมพันธ์กับ UI สิ่งหนึ่งที่ได้รับคือประสบการณ์ของผู้ใช้งาน อาจได้รับความรู้สึกในแง่บวก หรือได้รับความรู้สึกแง่ลบ นักพัฒนาจึงจำเป็นต้องศึกษาความต้องการของผู้ใช้งาน เพื่อให้ตอบโต้กับผู้ใช้งานมากที่สุด ฉะนั้นอาจเปรียบได้ว่า UX คือ “ศาสตร์แห่งความพยายามเข้าใจผู้อื่นเพื่อประโยชน์อันสูงสุด” [11] [12]

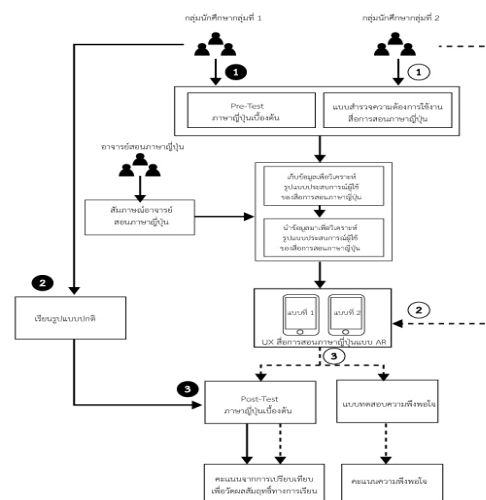
ประสบการณ์ของผู้ใช้จะช่วยสร้างความพึงพอใจในการใช้งานให้กับผู้ใช้งาน มีการศึกษาเกี่ยวกับประสบการณ์ของผู้ใช้ที่มีต่อเว็บไซต์ของแบรนด์ต่าง ๆ โดย Sillence และคณะระบุว่าการออกแบบส่งผลต่อความประทับใจแรกของผู้ใช้งานสูงถึง 94% [13] ซึ่งสอดคล้อง

กับงานวิจัยของ Phillips และ Chaparro ที่บอกว่าการออกแบบเว็บไซต์ให้มีความสวยงาม จะสามารถดึงดูดและสร้างความประทับใจแรกของผู้ใช้งานได้มากกว่าการใช้งานที่ง่าย [14] ซึ่งถ้าความประทับใจแรกมีแนวโน้มไปในแง่บวก ก็สามารถเพิ่มความพึงพอใจของผู้ใช้งานได้

นอกจากนี้ยังมีการศึกษาการออกแบบประสบการณ์ผู้ใช้ เพื่อนำไปเป็นกลยุทธ์การออกแบบประสบการณ์ผู้ใช้พื้นฐานในการพัฒนาแอปพลิเคชันบนมือถือโดย Yazid และ Jantan ได้ศึกษางานที่เกี่ยวข้องและครอบคลุมในด้านประสบการณ์ผู้ใช้เกี่ยวกับแอปพลิเคชันบนมือถือประเภทต่าง ๆ และศึกษาการออกแบบประสบการณ์ผู้ใช้สำหรับแอปพลิเคชันจองตั๋วของเที่ยวบินบนมือถือ เพื่อหาแนวทางที่ดีที่สุดของการออกแบบประสบการณ์ผู้ใช้ ซึ่งกลยุทธ์พื้นฐานในการพัฒนามีองค์ประกอบทั้งหมด 7 องค์ประกอบ ได้แก่ ความสะดวกในการใช้งาน, ความสามารถในการเรียนรู้, ส่วนต่อประสานกับผู้ใช้, ความพึงพอใจของผู้ใช้, ความปลอดภัย, ความตั้งใจเชิงพฤติกรรม และสิ่งแวดล้อม [15]

งานวิจัยที่เกี่ยวข้องกับประสบการณ์ของผู้ใช้ ส่วนใหญ่จะมุ่งเน้นไปที่การหารูปแบบที่ดีที่สุดในการพัฒนาเว็บไซต์หรือแอปพลิเคชันบนมือถือโดยใช้ประสบการณ์ของผู้ใช้เป็นปัจจัยหลักในการพัฒนา เพื่อให้ผู้ใช้งานเกิดความพึงพอใจต่อผลิตภัณฑ์ที่ใช้งาน

**3. วิธีการดำเนินการวิจัย**



ภาพที่ 1: วิธีการดำเนินการวิจัย

จากภาพข้างต้น แสดงให้เห็นถึงขั้นตอนในการดำเนินการวิจัย โดยมีรายละเอียดดังต่อไปนี้

**3.1 เก็บข้อมูลเพื่อวิเคราะห์รูปแบบประสบการณ์ผู้ใช้**

การเก็บข้อมูลการสอนภาษาญี่ปุ่นจากการสัมภาษณ์อาจารย์สอนภาษาญี่ปุ่น และสำรวจความต้องการของนักศึกษาระดับปริญญาตรี ชั้นปีที่ 1 คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีไทย-ญี่ปุ่น จำนวน 54 คน เพื่อนำข้อมูลมาวิเคราะห์รูปแบบประสบการณ์ผู้ใช้

**3.2 แบบทดสอบวัดความรู้ของผู้เรียนก่อนเรียน**

ให้กลุ่มตัวอย่างทำแบบทดสอบวัดความรู้ภาษาญี่ปุ่นเบื้องต้น เพื่อใช้ในการเปรียบเทียบผลก่อนและหลังจากเรียนภาษาญี่ปุ่น

**3.3 ออกแบบและพัฒนาสื่อการสอน**

นำข้อมูลจากการเก็บข้อมูลมาพัฒนาสื่อการสอน โดยเลือกรูปแบบที่กลุ่มตัวอย่างนิยม 2 รูปแบบ เพื่อวิเคราะห์รูปแบบที่เหมาะสมและความพึงพอใจของผู้ใช้งาน

**3.4 ทดลองใช้งานและวิเคราะห์ความพึงพอใจ**

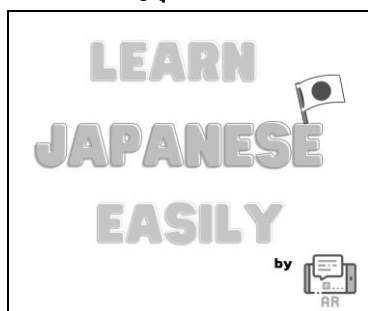
แบ่งกลุ่มตัวอย่างเป็น 2 กลุ่ม โดยกลุ่มที่ 1 ให้เรียนภาษาญี่ปุ่นในรูปแบบปกติจำนวน 27 คน กลุ่มที่ 2 ให้เรียนภาษาญี่ปุ่นโดยใช้สื่อการสอนที่พัฒนาขึ้นจำนวน 27 คน และทำแบบทดสอบประเมินความพึงพอใจหลังจากใช้งาน เพื่อวิเคราะห์ความพึงพอใจของผู้ใช้

**3.5 แบบทดสอบวัดความรู้ของผู้เรียนหลังเรียน**

ให้กลุ่มตัวอย่างทั้ง 2 กลุ่มทำแบบทดสอบหลังจากเรียนภาษาญี่ปุ่น นำมาเปรียบเทียบกับผลที่ได้จากการทำแบบทดสอบก่อนเรียน เพื่อวัดผลสัมฤทธิ์ทางการเรียนรู้

**4. ผลการดำเนินงาน**

**4.1 สื่อการสอนภาษาญี่ปุ่น**



ภาพที่ 2: แสดงหน้าหลักของสื่อการสอน

จากภาพที่ 2 แสดงให้เห็นถึงหน้าจอหลักของสื่อการสอนภาษาญี่ปุ่นที่พัฒนาขึ้นจากใช้ข้อมูลการสำรวจกลุ่มตัวอย่าง ทั้งนี้ผู้วิจัยได้เลือกจากรูปแบบที่กลุ่มตัวอย่างนิยมที่สุด 2 อันดับแรก นั่นคือแอนิเมชันและประโยคสนทนา มาพัฒนาสื่อการสอนภาษาญี่ปุ่น 2 รูปแบบโดยใช้บทสนทนาเป็นเนื้อหาหลักของสื่อการสอนคือรูปแบบแอนิเมชัน และรูปแบบประโยคสนทนา



ภาพที่ 3: ตัวอย่างหน้าจอแสดงผลสำหรับสื่อการสอนรูปแบบแอนิเมชัน



ภาพที่ 4: ตัวอย่างหน้าจอแสดงผลสำหรับสื่อการสอนรูปแบบประโยคสนทนา

จากภาพที่ 3 และ 4 แสดงตัวอย่างหน้าจอแสดงผลของสื่อการสอนที่พัฒนาขึ้น โดยภาพที่ 3 จะแสดงตัวอย่างหน้าจอแสดงผลสำหรับสื่อการสอนรูปแบบแอนิเมชัน และภาพที่ 4 จะแสดงตัวอย่างหน้าจอแสดงผลสำหรับสื่อการสอนประเภทประโยคสนทนา



ภาพที่ 5: แสดงวิธีการใช้งานสื่อการสอน

จากภาพที่ 5 แสดงวิธีการใช้งานสื่อการสอนทั้ง 2 รูปแบบ โดยจะต้องติดตั้งแอปพลิเคชันสื่อการสอนลงบน

มือถือของกลุ่มตัวอย่าง และสแกนคำศัพท์ผ่านแอปพลิเคชัน จากนั้นจะปรากฏแอนิเมชันขึ้นมา สำหรับการสอนรูปแบบแอนิเมชัน และจะปรากฏประโยคสนทนา สำหรับการสอนรูปแบบประโยคสนทนาพร้อมการออกเสียง

#### 4.2 คะแนนเฉลี่ยก่อนและหลังเรียนภาษาญี่ปุ่น

**ตารางที่ 1:** ตารางเปรียบเทียบคะแนนเฉลี่ยก่อนและหลังเรียนภาษาญี่ปุ่นของกลุ่มตัวอย่างทั้ง 2 กลุ่ม

	คะแนนเต็ม (100 คะแนน)	
	กลุ่มที่ 1	กลุ่มที่ 2
แบบทดสอบก่อนเรียน	13.91	12.06
แบบทดสอบหลังเรียน	76.18	85.74
อัตราการเพิ่มขึ้น (%)	62.27%	73.68%

จากตารางที่ 1 แสดงคะแนนการทำแบบทดสอบวัดความรู้ภาษาญี่ปุ่น ซึ่งแบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่ม กลุ่มที่ 1 คือกลุ่มตัวอย่างที่เรียนภาษาญี่ปุ่นในรูปแบบปกติ และกลุ่มที่ 2 คือกลุ่มตัวอย่างที่ทดลองใช้สื่อการสอนที่พัฒนาขึ้น โดยกลุ่มที่ 1 มีอัตราการเพิ่มขึ้นของคะแนน 62.27% และกลุ่มที่ 2 มีอัตราการเพิ่มขึ้นของคะแนน 73.68% จะเห็นได้ว่ากลุ่มที่ทดลองใช้สื่อการสอนภาษาญี่ปุ่นประเภทความจริงเสริมมีอัตราการเพิ่มขึ้นของคะแนนที่สูงกว่ากลุ่มที่เรียนภาษาญี่ปุ่นแบบปกติ

#### 4.3 คะแนนความพึงพอใจ

**ตารางที่ 2:** ตารางคะแนนความพึงพอใจเฉลี่ยหลังจากใช้งานสื่อการสอนภาษาญี่ปุ่นประเภทความจริงเสริมทั้ง 2 รูปแบบ

หัวข้อการประเมิน	คะแนนเฉลี่ย (5 คะแนน)	
	รูปแบบที่ 1	รูปแบบที่ 2
สร้างความสนใจในการเรียน	4.79	4.27
สร้างความสนุกสนานในการเรียน	4.57	4.23
ความสะดวกในการใช้งาน	4.21	4.19
ช่วยให้เข้าใจภาษาญี่ปุ่นมากขึ้น	4.73	4.64
การติดตั้งแอปพลิเคชัน	3.59	3.61
คะแนนรวม	4.38	4.19

จากตารางที่ 2 แสดงคะแนนความพึงพอใจเฉลี่ยหลังจากใช้งานสื่อการสอนภาษาญี่ปุ่นประเภทความจริงเสริมทั้ง 2 รูปแบบ โดยรูปแบบที่ 1 คือรูปแบบแอนิเมชัน

และรูปแบบที่ 2 คือรูปแบบประโยคสนทนา จะเห็นว่าสื่อการสอนรูปแบบที่ 1 มีคะแนนเฉลี่ย 4.38 คะแนน ซึ่งสูงกว่ารูปแบบที่ 2 ที่มีคะแนนเฉลี่ย 4.19 คะแนน และหัวข้อประเมินที่มีคะแนนน้อยที่สุดคือการติดตั้งแอปพลิเคชัน เนื่องจากมือถือบางรุ่นไม่รองรับการใช้งาน AR ทำให้มีปัญหาในการติดตั้งและมีปัญหาด้านการใช้งาน

### 5. สรุป

เนื่องจากการเรียนภาษาญี่ปุ่นรูปแบบเดิมอาจจะไม่ช่วยสร้างความน่าสนใจในการเรียนที่มากพอ และจากสถานการณ์ COVID-19 ในปัจจุบัน ทำให้สื่อการสอนรูปแบบเดิมใช้งานได้ไม่เต็มประสิทธิภาพ ผู้วิจัยจึงนำเอาความจริงเสริมร่วมกับการวิเคราะห์ประสบการณ์ผู้ใช้มาพัฒนาสื่อการสอนภาษาญี่ปุ่นในระดับเริ่มต้น เพื่อให้ผู้ใช้งานสามารถเรียนรู้อาษาญี่ปุ่นผ่านมือถือได้ทุกที่ทุกเวลา ช่วยเพิ่มความสนใจในการเรียนและเพิ่มผลสัมฤทธิ์ทางการเรียนรู้อาษาญี่ปุ่น จากผลการทดลองแบ่งกลุ่มตัวอย่างในการเรียนภาษาญี่ปุ่นออกเป็น 2 กลุ่ม พบว่ากลุ่มที่ใช้สื่อการสอนประเภทความจริงเสริมในการเรียนภาษาญี่ปุ่นมีอัตราการเพิ่มขึ้นของคะแนน 73.68% และกลุ่มที่เรียนแบบปกติมีอัตราการเพิ่มขึ้นของคะแนน 62.27% แสดงให้เห็นว่าการใช้สื่อการสอนประเภทความจริงเสริมในการเรียนภาษาญี่ปุ่นช่วยให้ผู้เรียนเข้าใจบทเรียนและมีผลการเรียนที่ดีขึ้น และจากการวิเคราะห์ความพึงพอใจของกลุ่มที่ใช้สื่อการสอนประเภทความจริงเสริมในการเรียนภาษาญี่ปุ่นเพื่อหารูปแบบของสื่อการสอนที่เหมาะสมในการเรียนภาษาญี่ปุ่น พบว่าคะแนนความพึงพอใจเฉลี่ยหลังจากใช้งานสื่อการสอนแบบแอนิเมชัน 4.38 คะแนน (87.6 %) และสื่อการสอนแบบประโยคสนทนา 4.19 คะแนน (83.8 %) จะเห็นว่าสื่อการสอนแบบแอนิเมชันมีคะแนนความพึงพอใจที่สูงกว่า เนื่องจากสื่อการสอนแบบแอนิเมชันทำให้ได้รู้วิธีการนำคำไปใช้จริง จึงทำให้จดจำได้มากกว่าสื่อการสอนแบบประโยคสนทนา และจะเห็นว่าหัวข้อประเมินที่มีคะแนนน้อยที่สุดคือการติดตั้งแอปพลิเคชัน เนื่องจากมือถือบาง

รุ่นไม่รองรับการใช้งาน AR ทำให้มีปัญหาในการติดตั้ง และมีปัญหาด้านการใช้งาน

นอกจากนี้ยังสามารถนำสื่อการสอนประเภทความจริงเสริมนี้ไปเป็นต้นแบบในการพัฒนาสื่อการสอนภาษาญี่ปุ่นให้มีประสิทธิภาพมากขึ้น และในอนาคตผู้วิจัยจะนำการทดลองนี้ไปประยุกต์ใช้กับการเรียนการสอนในวิชาอื่น ๆ ที่ใกล้เคียงกัน เพื่อช่วยเพิ่มความน่าสนใจในการเรียนและเพิ่มผลสัมฤทธิ์ทางการเรียนรู้ให้กับวิชาอื่น ๆ

### เอกสารอ้างอิง

- [1] Thailand Board of Investment. "Thailand BOI Approves Measures to Promote Industry 4.0 Transformation; Re-ports USD15 Billion in Jan-Sep Investment Applications." Available Online at [https://www.boi.go.th/index.php?page=press\\_releases\\_detail&topic\\_id=129511&language=en](https://www.boi.go.th/index.php?page=press_releases_detail&topic_id=129511&language=en), accessed on 20 Oct 2021.
- [2] วรลธิ จันทโร และธนภัส สนธิรักษ์. "การวิเคราะห์ปัญหาด้านการใช้ภาษาญี่ปุ่นที่พบในการฝึกงาน:กรณีศึกษาผู้เรียนวิชาเอกภาษาญี่ปุ่น." *วารสารเครือข่ายผู้ปุ่นศึกษา*, ปีที่ 10, ฉบับที่ 1, หน้า 65-82, 2020.
- [3] R. Azuma. "A Survey of Augmented Reality." *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355-385, 1997.
- [4] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht. "Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality." *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 6, pp. 1706-1724, 2017.
- [5] Feng Zhou, H. Duh, and M. Billinghurst. "Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR." *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2008.
- [6] พนิดา ตันศิริ. "โลกเสมือนผสมผสานโลกจริง Augmented Reality." *วารสารนักบริหาร มหาวิทยาลัยกรุงเทพ*, ปีที่ 30, ฉบับที่ 2, หน้า 169-173, 2010.
- [7] จิราภรณ์ ปกรณ์. "AR (Augmented Reality) เทคโนโลยีโลกเสมือนผสมผสานโลกแห่งความจริง." Available Online at <https://www.scimath.org/article-technology/item/7755-ar-augmented-reality>, accessed on 7 Oct 2021.
- [8] F. DITA. "A Foreign Language Learning Application using Mobile Augmented Reality." *Informatica Economica*, vol. 20, no. 42016, pp. 76-87, 2016.
- [9] X. Geng, and M. Yamada. "An augmented reality learning system for Japanese compound verbs: study of learning performance and cognitive load." *Smart Learning Environments*, vol. 7, no. 1, 2020.
- [10] D. Plecher, C. Eichhorn, J. Kindl, S. Kreisig, M. Wintergerst and G. Klinker. "Dragon Tale - A Serious Game for Learning Japanese Kanji." *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, pp. 577-583, Oct, 2004.
- [11] DEMETER ICT, "UX / UI คืออะไร?." Available Online at <https://www.dmit.co.th/th/ข่าวสาร/ux-vs-ui>, accessed on 7 Oct 2021.
- [12] moonlightkz. "User Interface (UI) กับ User Experience (UX) คืออะไร? และ แตกต่างกันอย่างไรร?." Available Online at <https://tips.thaiware.com/1437.html>, accessed on 7 Oct 2021.
- [13] E. Sillence, P. Briggs, L. Fishwick, and P. Harris. "Trust and mistrust of online health sites." *CHI 2004 Conference on Human Factors in Computing Systems (CHI '04)*, pp. 663-670, Jan, 2004.
- [14] C. Phillips, and B. Chaparro. "Visual appeal vs. usability: which one influences user perceptions of a website more." *Usability News*, vol. 11, no. 2, pp. 1-9, 2009.
- [15] M. A. Yazid, and A. H. Jantan. "User Experience Design (UXD) of Mobile Application: An Implementation of a Case Study." *JTEC*, vol. 9, no. 3-3, pp. 197-200, Oct, 2017.

## การพัฒนาเกมหันทภัยไวรัสด้วยเทคโนโลยีเสมือนจริง Threatening Virus Game Development Using Virtual Reality

ภูริภัทร ธีระวุฒิ (Phuriphat Theerawut)<sup>1</sup> สุชาสินี โทวุฒิกุล (Sutasinee Thovuttikul)<sup>2</sup>

และวัชรีย์ จำปามูล (Watcharee Jumpamule)<sup>3</sup>

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

<sup>1</sup>Phuriphat\_th@elearning.cmu.ac.th, <sup>2</sup>sutasinee.th@cmu.ac.th, <sup>3</sup>watcharee.j@cmu.ac.th

### บทคัดย่อ

ปัจจุบันยังคงมีการระบาดของไวรัสโคโรนาสายพันธุ์ใหม่ 2019 อยู่อย่างต่อเนื่อง ทำให้ผู้คนใช้ชีวิตอยู่ในที่พักอาศัยของตนเองมากขึ้น กิจกรรมนอกบ้านหลาย ๆ อย่าง ถูกยกเลิกหรือปรับเปลี่ยนเป็นรูปแบบออนไลน์ เพื่อลดโอกาสเสี่ยงในการรับหรือแพร่เชื้อ รวมถึงการสื่อสารและการเรียนรู้ส่วนใหญ่ถูกปรับเปลี่ยนไปในรูปแบบออนไลน์แล้วไม่สามารถกระตุ้นการเรียนรู้ได้เหมือนกับการออกไปทำกิจกรรมจริง ๆ ผู้พัฒนาจึงสนใจพัฒนาสื่อการเรียนรู้ผ่านเกมที่จะกระตุ้นให้ผู้เล่นตระหนักถึงข้อปฏิบัติในการดูแลตนเองในช่วงสถานการณ์การแพร่ระบาดของไวรัสโคโรนา ผู้วิจัยได้ศึกษาแนวทางการดูแลและป้องกันตนเองด้วยการหมั่นล้างมือเพื่อลดอัตราการติดเชื้อแล้วนำลักษณะกิจกรรมมาเป็นต้นแบบและพัฒนาเกมจำลองการใช้ชีวิตในสถานการณ์การแพร่ระบาดนี้

โดยเกม "มหันตภัยไวรัสเสมือนจริง" นี้ถูกพัฒนาในรูปแบบเกมสามมิติในมุมมองบุคคลที่หนึ่ง สามารถเล่นผ่านชุดอุปกรณ์สวมศีรษะสำหรับภาพเสมือนจริงเพื่อเรียนรู้กิจกรรมต่าง ๆ เช่น การระงับอัตราการติดเชื้อที่เพิ่มขึ้น การล้างมือด้วยสเปรย์แอลกอฮอล์ และการพกอุปกรณ์ป้องกันตัวเองเพื่อลดโอกาสการติดเชื้อ เป็นต้น เกมนี้ถูกพัฒนาด้วยเอ็นจินคอร์ ยูทิลิตี้ ในโปรแกรมยูนิตี้จากการพัฒนาพบว่าผู้เล่นสามารถเข้าใจและตระหนักถึงการพกพาสเปรย์แอลกอฮอล์และการหมั่นล้างมือเพื่อไม่เพิ่มโอกาสติดเชื้อได้

**คำสำคัญ:** ความเป็นจริงเสมือน เกมจำลอง อุปกรณ์สวมศีรษะสำหรับภาพเสมือนจริง สภาพแวดล้อมเสมือนจริง

### Abstract

Nowadays, there is still an outbreak of the new coronavirus 2019, so people still stay at home. The outdoor activities had been changed to an online event for preventing the coronavirus. An online event was not suitable for learning like the real situation. We are interested in the virtual reality technique and learning through play for designing and developing learning systems that help people aware of the coronavirus pandemic. We research the guidelines to prevent the virus then design the gameplay and activities in the game.

The 3D game "Threatening Virus VR game", a game from the first-person view using a headset. This game was designed to help players practice and realize activities such as handwashing self-protecting from the virus. This game was developed by using XR Core Utilities in the Unity3D. We found that players be aware to wash hands for preventing themselves from the virus

**Keyword:** virtual reality, simulation game, VR headset, virtual environment.

### 1. บทนำ

ในสถานการณ์ปัจจุบันที่มีการระบาดของไวรัสสายพันธุ์ใหม่ ไวรัสโคโรนา 2019 (COVID-19) มาตั้งแต่ปลายปี พ.ศ. 2562 จึงทำให้ต้องป้องกันตัวเองเพื่อไม่ให้ติดเชื้อ เช่น หลีกเลี่ยงการพบเจอกัน รักษาระยะห่าง อยู่แต่ในที่ที่ปกปิด เพื่อไม่เกิดการระบาดที่หนักขึ้น แต่การระบาดของเชื้อโรคยังคงเกิดขึ้นอย่างต่อเนื่องและยาวนาน กิจกรรมนอกบ้านหลาย ๆ อย่างถูกปรับเปลี่ยนรูปแบบเป็น

รูปแบบออนไลน์ แต่การสื่อสารหรือการเรียนรู้ผ่านจอคอมพิวเตอร์นั้น ไม่สามารถกระตุ้นการรับรู้ได้ เหมือนกับการออกไปเจอกันเพื่อทำกิจกรรมร่วมกันจริงๆ และผู้จัดทำได้สังเกตเห็นปัญหาการสื่อสารและการเรียนรู้ด้วยการนำเทคโนโลยีเสมือนจริงมาช่วยให้ได้รับประสบการณ์เหมือนจริง เหมือนได้ออกไปเรียนรู้ข้างนอกแต่จะยังไม่ปลอดภัยจากโรคโควิดได้ในสถานการณ์ปัจจุบัน

บทความนี้จะนำเสนอหลักการ/ทฤษฎีและงานวิจัยที่เกี่ยวข้องในส่วนถัดไป ตามด้วยวิธีดำเนินงาน ระบบที่นำเสนอและปิดท้ายด้วยส่วนสรุปและอภิปรายผล

## 2. ทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง

เทคโนโลยีวีอาร์ (Virtual reality: VR) [1-2] นั้นเป็นเทคโนโลยีความจริงเสมือน ควบคู่กับการใช้งานอุปกรณ์ VR โดยผ่านการรับรู้ เช่น การมองเห็นและการได้ยินเสียง ทำให้เราสามารถตอบสนองต่อสิ่งที่จำลองนั้นได้ ด้วยการเพิ่มมุมมองการเรียนรู้ 360 องศา โดยจำลองสถานที่ สถานการณ์ และประสบการณ์ต่าง ๆ

เทคโนโลยี VR เป็นเทคโนโลยีที่นิยมในการนำมาพัฒนาสื่อต่าง ๆ [3-4] ในสมัยใหม่ที่จะให้ความรู้สึกตามกับโลกเสมือนที่สร้างขึ้นให้เหมือนโลกจริง และตัวเทคโนโลยียังสามารถจำลองสถานที่ต่าง ๆ ให้ผู้คนที่มีความรู้สึกเข้าไปสถานที่นั้นจริง ยกตัวอย่างเช่น Google Earth VR ที่จะเป็นที่จำลองขึ้นตามสถานที่จริงภายในโลกที่ให้ผู้เล่น ได้รู้สึกเหมือนไปเที่ยวจริง ๆ เทคโนโลยี VR นั้นจึงเหมาะสมมากในการพัฒนาตัวโปรแกรมหรือการเล่นเกมจำลองในรูปแบบต่าง ๆ โดยในการพัฒนานี้ ผู้พัฒนาเลือกเกมการจำลองชีวิตในสถานการณ์การระบาดของไวรัส ให้ผู้เล่นได้สำรวจและจับกลไกการดำเนินการในสภาพแวดล้อมเสมือนจริง ทำให้ผู้เล่นตระหนักถึงการป้องกันการติดเชื้อ ความสะอาด ระวังการใกล้ชิดผู้คน เป็นต้น ภูมิคุ้มกันไข่เลข [5] ถูกออกแบบเพื่อให้เรียนรู้ระบบภูมิคุ้มกันของร่างกายและกลไกการทำงานของวัคซีนนั้นที่ข้อมูลและการหมุนดูเซลล์ หลังจากเรียนรู้แล้ว พบว่าโปรแกรมตอบสนองความต้องการของผู้ใช้ได้ดี และในงานวิจัยนี้ ผู้วิจัยเห็นว่า การนำ VR มาต่อยอดเพื่อใช้ในการ

จำลองการใช้ชีวิตแบบปกติใหม่ (new normal) น่าจะเกิดประโยชน์ให้เกิดการตระหนักในระว่างการติดเชื้อในชีวิตประจำวันได้เช่นกัน

ในปัจจุบันเทคโนโลยีเสมือนจริงนั้นได้รับความนิยมในการนำมาใช้ในรูปแบบที่หลากหลาย โดยมีนิยามดังนี้

### 2.1 นิยามของความเป็นจริงเสมือน [1]

การนิยามอยู่ 4 ปัจจัยหลัก ดังต่อไปนี้

2.1.1 โลกเสมือน (Virtual World) คือ โลกที่ถูกสร้างหรือถูกจำลองขึ้นโดยความคิดของผู้สร้างสรรค์ มีเนื้อเรื่องมีเรื่องราว เช่น ภาพยนตร์ ละคร และเพลง เป็นต้น

2.1.2 ความรู้สึกท่วม (Immersion) คือ ความรู้สึกท่วมและประสบการณ์ร่วมที่ผู้ใช้ได้รับทั้งทางจิตใจและทางร่างกาย เช่น ภาพยนตร์ที่มีเนื้อเรื่องตื่นเต้นเร้าใจ จนผู้ชมรู้สึกตื่นเต้นเสมือนเป็นตัวละครในภาพยนตร์ เป็นต้น

2.1.3 การรับรู้ที่เสมือนจริง (Sensory Feedback) คือ การที่ผู้ใช้ได้รับเสมือนว่าได้เข้าไปอยู่ในสถานที่เสมือน และมีตัวตนจริง ๆ ในโลกเสมือนแห่งนั้นจริง ๆ เป็นต้น

2.1.4 การมีปฏิสัมพันธ์ (Interactivity) คือ ผู้ใช้สามารถตอบโต้กับสิ่งของและสภาพแวดล้อมในโลกเสมือนได้ เช่น สภาพแวดล้อมที่สามารถหยิบหรือจับได้ เป็นต้น

สรุปนิยามของ VR ได้ว่า โลกเสมือนที่ถูกสร้างหรือจำลองขึ้นด้วยคอมพิวเตอร์ โดยที่ผู้ใช้จะได้รับรู้เรื่องราวความรู้สึกและประสบการณ์ที่โลกเสมือนสร้างขึ้นได้

### 2.2 ประเภทของ VR [1]

2.2.1 VR แบบบางส่วน (Low and Partial VR) ผู้ใช้ได้รับประสบการณ์หรือเรื่องราวจากโลกเสมือนได้น้อย จะแสดงผลด้วยจอแสดงภาพ (จอคอมพิวเตอร์หรือจอโทรทัศน์)

2.2.2 VR แบบเต็มรูปแบบ (Full VR) ผู้ใช้สามารถได้รับประสบการณ์หรือเรื่องราวจากโลกเสมือนได้เต็มรูปแบบ ที่มีภาพ เสียงหรือการมีปฏิสัมพันธ์กับสิ่งแวดล้อมทุกอย่าง จะแสดงผลด้วยอุปกรณ์ VR ขนาดใหญ่ เช่น HTC VIVE, Oculus Rift เป็นต้น



### 2.3 โปรแกรมที่ใช้พัฒนา [6]

โปรแกรมยูนิตี้ (Unity) เป็นเครื่องมือที่ใช้ในการพัฒนาเกม (Game Engine) ในรูปแบบ 2 มิติ และ 3 มิติรวมทั้งการสร้างแบบจำลองสถานการณ์ต่าง ๆ ได้

### 2.4 ภาษาที่ใช้ในการพัฒนาเกม [6]

ในการพัฒนาเกม ทางผู้พัฒนาได้นำภาษาซีชาร์ป (C-Shape : C#) มาพัฒนาเกม รูปแบบการเขียนสคริปต์ (Script) โดยเป็นการเขียนโปรแกรมเชิงวัตถุ

### 2.5 อุปกรณ์ที่ใช้พัฒนา [7]

อุปกรณ์ที่นำมาพัฒนาในงานวิจัยนี้คืออุปกรณ์ VR เอชทีซี ไวฟ์ (HTC Vive Original) เพื่อในการทดสอบการทำงานของตัวเกมที่พัฒนา เพราะว่ามีสถานีฐาน (Base-Station) ที่ตรวจจับการเคลื่อนไหวและตำแหน่งของผู้ใช้ได้แม่นยำ และมีอุปกรณ์ที่แสดงภาพความละเอียดสูง



ภาพที่ 1: อุปกรณ์ชุดสร้างภาพเสมือนจริง HTC VIVE

## 3. ขั้นตอนการดำเนินการพัฒนา

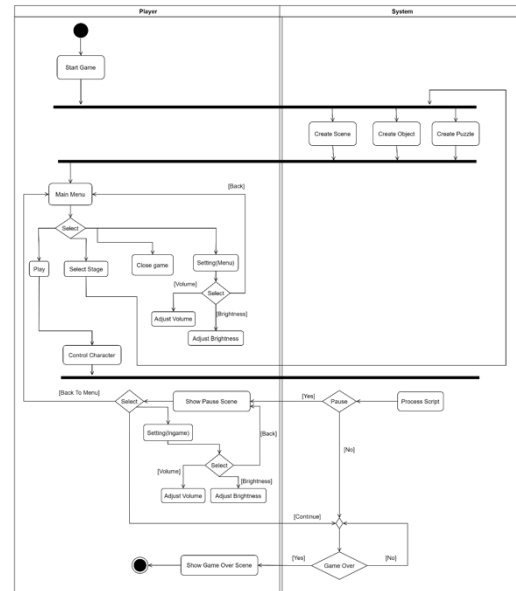
ในการพัฒนาเกม เพื่อจำลองการเรียนรู้การใช้ชีวิตในสถานการณ์โรคโควิด 19 มีจุดประสงค์เพื่อให้ผู้เล่นได้รับประสบการณ์ในชีวิตประจำวันด้วยความตระหนักถึงการป้องกันและรักษาความสะอาด เพื่อบรรลุวัตถุประสงค์ดังกล่าวจึงได้มีขั้นตอนการออกแบบดังนี้

### 3.1 การออกแบบและวิเคราะห์ระบบ

แผนภาพกิจกรรม (Activity) เป็นแผนภาพที่แสดงถึงภาพรวมของระบบและลำดับในการทำงานของกิจกรรมต่าง ๆ ภายในระบบทำให้ผู้พัฒนาเข้าใจถึงการขั้นตอนการทำงานจากระบบ ดังภาพที่ 2

เมื่อผู้เล่นเริ่มเล่นเกมการทำงานของระบบจะสร้างฉากพื้นฐาน (Create Scene) สิ่งของภายในฉาก (Create Object) และ ปริศนา (Create Puzzle) ภายในฉาก หลังจากนั้นผู้เล่นเลือกเมนู คือ เล่นเกม (Play) ตั้งค่า (Setting Menu) เลือกฉาก (Select Stage) และปิดเกม (Close game) หากผู้เล่นเลือกเล่นเกม ผู้เล่นจะสามารถควบคุมตัวละครได้ (Control

Character) หากเลือกตั้งค่า ผู้เล่นจะทำการเลือกเพิ่ม/ลดแสงและเพิ่ม/ลดเสียง หากเลือกฉาก ระบบสร้างฉากจะสร้างตามที่ได้เลือก และหากผู้เล่นเลือกปิดเกม(Close game) เกมก็จะจบทันที



ภาพที่ 2: แผนภาพกิจกรรมของเกมหัตถ์กับไวรัส VR

หลังจากเริ่มเล่นเกมจะหยุดเกม (Pause) ได้ ขณะหยุดเกมจะตั้งค่าเกมได้ (Setting In game) และ หากจบเกม (Show Game Over Scene) จะเป็นการจบการทำงานระบบกิจกรรมในเกมถูกออกแบบและอ้างอิงจากข้อกำหนดในสถานการณ์การระบาด [8-9] ต้องหมั่นล้างมือด้วยน้ำหรือสเปรย์แอลกอฮอล์บ่อย ๆ กิจกรรมในเกมจึงมีการล้างมือด้วยสเปรย์แอลกอฮอล์อย่างสม่ำเสมอภายในระยะเวลาที่กำหนด เพื่อป้องกันการติดเชื้อเมื่อออกมาอยู่นอกบ้าน

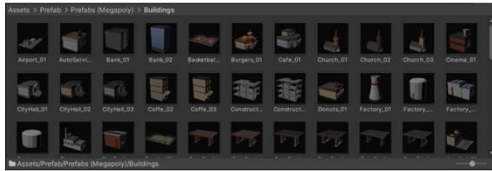
### 3.2 การพัฒนาระบบตัวเกมโดยใช้ Unity

องค์ประกอบเพื่อสร้างเป็นเกม [5] โดยมีส่วนประกอบหลักอยู่ 4 ส่วน ดังนี้

**3.2.1 กลุ่มก้อนวัตถุ (Game Object)** คือ วัตถุที่อยู่ในฉาก อย่างเช่น พื้นดิน ท้องฟ้า ผู้คน เป็นต้น

**3.2.2 คุณสมบัติ (Component)** คือคุณสมบัติของวัตถุในการเคลื่อนไหว เช่น ผู้คนที่สามารถเดินได้เสมือนจริง ท้องฟ้าที่มีเมฆลอยอยู่ และมีพระอาทิตย์ที่ส่องแสงอยู่

**3.2.3 รูปแบบสำเร็จรูป (Asset)** คือ โมเดล 3 มิติ เช่น แก้วน้ำ รถยนต์ รวมไปถึงจนถึงวัตถุทุกอย่างภายในเกม



ภาพที่ 3: รูปแบบสำเร็จรูป (Asset) ในรูปแบบ 3 มิติ

3.2.4 ฉากภายในเกม (Scene) คือสภาพแวดล้อมที่ถูกสร้างขึ้น เช่น ฉากภายในบ้าน ฉากในเมือง เป็นต้น

ในหัวข้อถัดไปอธิบายถึงการนำองค์ประกอบทั้งสี่มาเพื่อออกแบบและสร้างเกมตามวัตถุประสงค์ที่ได้กล่าวข้างต้น

#### 4. การพัฒนาระบบ

การพัฒนาระบบตัวเกมโดยใช้ Unity องค์ประกอบทุกส่วนจะถูกมองเป็นวัตถุแล้วนำมารวมกันเพื่อสร้างเป็นเกม [5] โดยมีส่วนประกอบหลักอยู่ 4 ส่วน ดังนี้

ดังนั้นหากนำองค์ประกอบ 4 อย่าง ข้างต้นมารวมกันจะสามารถสร้างเกม ขึ้นมาได้ ผู้พัฒนาได้ออกแบบเกมและเลือกใช้ Asset ต่าง ๆ ในการพัฒนาจาก Unity Asset store มีดังนี้

##### 4.1 แผนที่ในเกมและฉากในเกม (Game Scene)

4.1.1 ฉากบ้าน (Home) เป็นการจำลองสถานการณ์ใช้ชีวิต ให้สมจริงอ้างอิงจากบ้านจริงมีห้องนอน ห้องนั่งเล่น



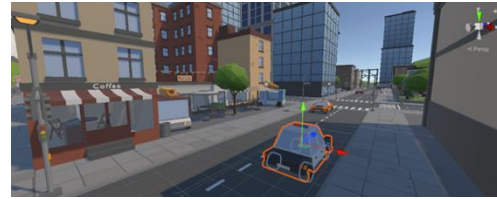
ภาพที่ 4: ฉากบ้าน – บริเวณภายนอกบ้าน



ภาพที่ 5: ฉากบ้าน - ห้องนั่งเล่นภายในบ้าน



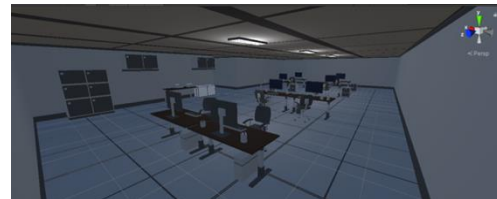
ภาพที่ 6: ฉากตัวเมือง – เมืองมุมบนท้องฟ้า



ภาพที่ 7: ฉากตัวเมือง – ร้านกาแฟ

4.1.2 ฉากตัวเมือง (City) เป็นฉากภายในเมืองให้ผู้เล่นได้ออกมาเดินเพื่อไปทำงาน ฉากนี้ถูกออกแบบให้มีขนาดใหญ่ การป้องกันตัวควรจะรอบคอบมากขึ้น

4.1.3 ฉากสถานที่ทำงาน (Office) ถูกออกแบบให้ผู้เล่นรู้สึกที่สามารถเข้าไปทำงานจริงด้วยฉากสำนักงาน



ภาพที่ 8: ฉากที่ทำงาน

##### 4.2 ระบบการโต้ตอบในเกม (Interactive System)

4.2.1 การควบคุมและการเคลื่อนไหว (Player Control) ระบบ Plug-in มีการรองรับ VR ด้วย Component และ Library ที่เกี่ยวข้อง ในการควบคุมตัวละครจะมีการสร้าง Game object XR Origin ขึ้น ประกอบไปด้วย มุมกล้องหลัก (Main Camera) คอนโทรลเลอร์ด้านซ้ายและขวา (Left & Right Controller) แสดงผลด้วย Asset Animation มือดังภาพที่ 9 ให้ผู้เล่นรู้สึกที่สามารถควบคุมตัวละครได้



ภาพที่ 9: การเคลื่อนไหวร่างกาย – มือภายในเกม

ต่อมาคือการพัฒนา Component : Locomotion system และ Component : Teleportation area เพื่อให้เคลื่อนที่ได้ด้วยการ เทเลพอร์ต (Teleportation) ดังภาพที่ 10



ภาพที่ 10: การเคลื่อนที่แบบ Teleport

#### 4.2.2 การหยิบจับสิ่งของ (Interactive)

การหยิบจับสิ่งของภายในเกม ทำได้โดยการติดตั้ง Component: XR Grab Interactable เพื่อให้สิ่งของภายในฉากนั้นสามารถจับต้องได้ โดยแบ่งการหยิบจับเป็น 2 ประเภท เพื่อให้การใช้งานของผู้ใช้สะดวกยิ่งขึ้น

1) หยิบสิ่งของในระยะใกล้ คือการหยิบจับสิ่งของได้โดยตรง โดยการเพิ่ม Component: XR Direct Interactor เพื่อให้มือทั้งสองข้างสามารถหยิบจับโดยตรงกับสิ่งของนั้นได้ดังภาพที่ 11



ภาพที่ 11: การหยิบจับสิ่งของระยะใกล้ภายในเกม

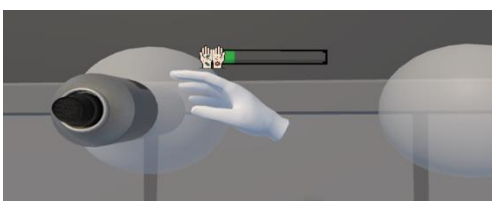
2) หยิบสิ่งของในระยะไกล คือการหยิบจับสิ่งของทางอ้อม โดยการเพิ่ม Component : XR Ray Interactor เพื่อให้มือทั้งสองข้างสามารถหยิบจับสิ่งของได้แทนการเอื้อมมือ เพื่อให้ผู้ใช้งานมีความสะดวกในการเล่นและสามารถ กดเมนูภายในเกมได้ ดังภาพที่ 12



ภาพที่ 12: เส้นชี้การจับในระยะไกลสามารถจับได้

#### 4.2.3 ระบบช่องเก็บของ (Inventory)

ระบบช่องเก็บของเป็นการออกแบบให้ตัวละครผู้เล่นสามารถเก็บของภายในฉากได้ ด้วยการสร้าง Game Object : Sphere เป็นพื้นที่ทรงกลมแล้วเพิ่ม Component : XR Socket Interactor เพื่อให้สามารถเก็บวัตถุที่เข้าไปอยู่ในพื้นที่นั้นได้ ดังภาพที่ 13



ภาพที่ 13: ช่องเก็บของ – เก็บสเปรย์แอลกอฮอล์อยู่

#### 4.2.4 การล้างมือ โดยใช้สเปรย์แอลกอฮอล์ (Washing)

ระบบการติดเชื้อภายในเกม และระบบการล้างมือภายในเกมนั้น ได้มีการออกแบบให้ผู้เล่นรู้สึกว่าการออกเดินทางไปนอกบ้าน จะมีโอกาสติดเชื้อมากขึ้น ผู้เล่นสามารถดูโอกาสติดเชื้อได้จากหลอดการติดเชื้อ (Infection Bar) ที่จะเพิ่มขึ้นตามระยะเวลา หากผู้เล่นไปจับสิ่งของขึ้นมา โอกาสการติดเชื้อก็จะเพิ่มมากขึ้นอีก แต่ถ้าผู้เล่นหยิบสเปรย์แอลกอฮอล์นำมาพ่นที่มือเพื่อฆ่าเชื้อไวรัส โอกาสการติดเชื้อก็จะลดลงดังภาพที่ 14 โดยการออกแบบท่าทางการฉีดสเปรย์อ้างอิงมาจากท่าทางการใช้งานจริงสเปรย์แอลกอฮอล์ ดังภาพที่ 15



ภาพที่ 14: การใช้สเปรย์แอลกอฮอล์ล้างมือภายในเกม



ภาพที่ 15: การใช้สเปรย์แอลกอฮอล์ล้างมือในชีวิตจริง

#### 4.3 ระบบเสียงในเกม

เพลงประกอบในเกมจะถูกปรับแต่งเพื่อให้การเล่นนั้นมีความเพลิดเพลิน โดยจะติดตั้งที่ Component : Audio Source และเพลงนำมาใช้เป็นเพลงที่ทำให้ความรู้สึกตื่นเต้น เมื่อขณะเล่นเกมว่าจะให้ประสบการณ์ที่แปลกใหม่อย่างไรบ้าง ชื่อเพลง Motivational Upbeat Corporate Background โดย Rinkevich Music ที่ไม่มีลิขสิทธิ์ในการนำมาใช้

#### 4.4 การเปรียบเทียบประสบการณ์จากการเล่นเกม VR

เกมนี้เป็นการจำลองชีวิตในสถานการณ์การระบาดของไวรัส เพื่อให้ผู้เล่นได้สำรวจในสภาพแวดล้อมเสมือนจริง เพื่อให้ได้รับประสบการณ์ที่จะทำให้ผู้เล่นตระหนักถึงการป้องกันการติดเชื้อ ความสะอาด โดยเมื่อเปรียบเทียบกับการแสดงในตารางที่ 1

**ตารางที่ 1:** ประสบการณ์ที่ผู้เล่นจะได้รับระหว่างการเล่นเกม

ประสบการณ์ที่ผู้เล่นจะได้รับระหว่างการเล่นเกม	โลกจริง	โลกเสมือน VR
1. ความสะอาดของมือเมื่อสัมผัสสิ่งของต่าง ๆ	- ต้องตระหนักด้วยตนเอง	- มีหลอดติดเชื้อแสดงให้สังเกตได้ชัดเจนตั้งแต่แสดงในรูป 9-14
2. เมื่อเวลาผ่านไปต้องระวังอัตราการติดเชื้อที่เพิ่มขึ้น	- ไม่ทราบความเปลี่ยนแปลงเมื่อจับสิ่งของต่าง ๆ	- มีหลอดติดเชื้อในรูป 10 - เมื่อจับสิ่งของหลอดจะเปลี่ยนจากสีเขียวเป็นสีเหลือง และสีแดงในที่สุด
3. การติดเชื้อหากไม่ระมัดระวัง	- สามารถป่วยและติดเชื้อได้จริง	- หากติดเชื้อสามารถลองเล่นใหม่ได้ จนกว่าจะเข้าใจข้อปฏิบัติที่ถูกต้อง

### 5. บทสรุป

เกมมหันตภัยไวรัส VR ถูกพัฒนาขึ้นโดยมีวัตถุประสงค์เพื่อประยุกต์ใช้เทคโนโลยี VR ในการพัฒนาเกมและเพื่อให้มีสื่อที่ทำให้ผู้เล่นได้รับประสบการณ์ในชีวิตประจำวันด้วยความตระหนักถึงการป้องกันตัวเองจากเชื้อโรคและการรักษาความสะอาดนั้นเป็นเรื่องใกล้ตัวและต้องใส่ใจปฏิบัติ โดยผู้เล่นจะได้รับความรู้ที่สมจริงจากการเล่นเกมมหันตภัยไวรัส VR

จากการพัฒนาเกม VR สามารถให้ความสมจริงเพื่อให้ผู้ใช้งานรู้สึกเหมือนเข้าไปทำกิจกรรมนั้นจริง โดยจากการสอบถามผู้เล่นพบว่าหลังจากเล่นเกมแล้วพบว่าผู้เล่นเข้าใจและตระหนักถึงการล้างมือด้วยสเปรย์ แอลกอฮอล์ ดังข้อมูลที่แสดงตารางที่ 1 และเมื่อทำการสัมภาษณ์ผู้เล่นแล้วได้ข้อมูลเพิ่มเติมว่า สามารถดูโอกาสหรือสถานะการติดเชื้อได้จากหลอดการติดเชื้อจึงทำให้รู้ว่ามีกิจกรรมใดบ้างที่มีโอกาสในการติดเชื้อก็อยากหลีกเลี่ยงไม่ทำกิจกรรมนั้นและเมื่อหลอดติดเชื้อมีสีส้มหรือแดงก็ต้องหมั่นใช้สเปรย์แอลกอฮอล์เพื่อลดโอกาสติดเชื่อนั่นเอง

นอกจากนี้ หากกล่าวถึงการพัฒนาแล้วยังพบข้อจำกัดของอุปกรณ์และเกมดังนี้

1. ตัวเกมนั้นจำเป็นต้องมีอุปกรณ์ VR เท่านั้นหากไม่มีอุปกรณ์จะไม่สามารถเล่นได้

2. ต้องมีพื้นที่ที่เหมาะสมกับการเล่นเนื่องจากหากพื้นที่ในการเล่นน้อยอาจทำให้ผู้เล่นได้รับความเสียหายและผู้เล่นอาจได้รับบาดเจ็บจากผนังหรือสิ่งกีดขวางได้

4. ตัวอุปกรณ์มีสายต่อพ่วงดังนั้น ต้องระมัดระวังในการเคลื่อนไหวหรือหมุนตัว หากสายพันกันอาจทำให้ได้รับบาดเจ็บได้

5. ผู้ใช้งานไม่สามารถใช้เป็นเวลานานได้ อาจจะทำให้ผู้ใช้งานมีอาการ วิงเวียนศีรษะได้

เกมมหันตภัยไวรัส VR นี้จะถูกนำไปพัฒนาต่อให้มีความเสมือนจริงมากยิ่งขึ้น โดยจะเพิ่มความฉลาดให้กับเกมด้วยปัญญาประดิษฐ์ที่สามารถสื่อสารหรือโต้ตอบกับผู้เล่นได้อีก 2 แบบ ดังนี้ (1) การให้คำแนะนำเพื่อการปฏิบัติตัวในการป้องกันตัวเองจากเชื้อโรคและการรักษาความสะอาด (2) ตัวละครที่ผู้เล่นไม่ได้ควบคุม (Non player character) ที่สามารถโต้ตอบกับผู้เล่นได้

### เอกสารอ้างอิง

- [1] W. R. Sherman, A. B. Craig, Understanding virtual reality. San Francisco, CA: Morgan Kauffman, 2003.
- [2] A. Menin, T. Rafael, and N. Luciana, "An analysis of VR technology used in immersive simulations with a serious game perspective." IEEE computer graphics and applications 38.2, pp. 57-73, 2018.
- [3] J. V. Christensen, et al. "Player experience in a VR and non-VR multiplayer game." Proceedings of the Virtual Reality International Conference-Laval Virtual. 2018.
- [4] M. Katzourin, et al. "Swordplay: Innovating game development through VR." IEEE Computer Graphics and Applications 26.6, 15-19, 2006.
- [5] Unity Real-Time Development Platform | 3D, 2D VR & AR Engine. <https://unity.com>. วันที่สืบค้น 27 กันยายน 2564.
- [6] STEAM®. (2564). STEAM®VR SUPPORT. [https://support.steampowered.com/steamvr/HTC\\_Vive/](https://support.steampowered.com/steamvr/HTC_Vive/). วันที่สืบค้น 27 กันยายน 2564.
- [7] พีรวัส วงศ์สวัสดิ์ กยัคเคช คำมูล และชนพัฒน์ ดวงพิกุล, "ภูมิคุ้มกันใช้เลย" รายงานฉบับสมบูรณ์ โครงการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย ครั้งที่ 24
- [8] 7 steps to clean to hands. <https://covid19.anamai.moph.go.th> กรมอนามัย กระทรวงสาธารณสุข. วันที่สืบค้น 27 กันยายน 2564.
- [9] คำแนะนำการใช้แอลกอฮอล์ในการทำความสะอาดและฆ่าเชื้อโรคในสถานการณ่การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019. <https://covid19.anamai.moph.go.th> กรมอนามัย กระทรวงสาธารณสุข. วันที่สืบค้น 27 กันยายน 2564.

## แอนเดอร์เกรด : เกมจำลองชีวิตนักศึกษาไอที สจล.

### Undergrad: IT KMITL Student Life Simulation Game

สุนิสา จันทร์บาง (Sunisa Chanbang)<sup>1</sup> สุภกิจ น้าชม (Supakit Nachom)<sup>2</sup>

และสุพัฒน์ดา โชติพันธ์ (Supannada Chotipant)<sup>3</sup>

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

<sup>1</sup>61070249@kmitl.ac.th, <sup>2</sup>61070250@kmitl.ac.th, <sup>3</sup>supannada@it.kmitl.ac.th

#### บทคัดย่อ

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง มีการเรียนการสอนหลายหลักสูตร ทั้งนี้เพื่อตอบสนองความต้องการของตลาดในปัจจุบัน และเพื่อให้เข้าถึงผู้ที่สนใจหลักสูตรได้อย่างมีประสิทธิภาพ คณะจึงต้องมีการประชาสัมพันธ์หลักสูตรผ่านช่องทางต่าง ๆ อาทิ เว็บไซต์ สื่อสังคมออนไลน์ หรือกิจกรรมทางวิชาการต่าง ๆ อย่างไรก็ตามรูปแบบและเนื้อหาที่ยังไม่น่าสนใจเท่าที่ควร จากปัญหาดังกล่าวจึงริเริ่มในการพัฒนา “แอนเดอร์เกรด เกมจำลองชีวิตนักศึกษาไอที สจล.” เพื่อให้ข้อมูลน่าสนใจและเข้าใจง่าย จึงนำเสนอผ่านรูปแบบของเกมส์สามมิติแบบผู้เล่นหลายคน ซึ่งผู้เล่นจะได้สวมบทบาทสมมุติเป็นนักศึกษาในการทำภารกิจต่าง ๆ ที่เกี่ยวข้องกับแขนงวิชาของคณะ นอกจากนี้วิจัยฉบับนี้จะช่วยในการประชาสัมพันธ์คณะแล้วผู้เล่นยังได้รับความรู้ เกี่ยวกับรายวิชา และประสบการณ์การเรียนในคณะ ซึ่งสามารถใช้ประกอบการตัดสินใจในการเลือกเรียนได้ในอนาคต

**คำสำคัญ:** เกมจำลองสถานการณ์ การบูรณาการของกลศาสตร์เกม คณะไอที สจล.

#### Abstract

Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang has several courses to meet current market demands and effectively make course information accessible to interested parties. The course information publishes on several channels such as the website, social media, various academic activities. However, the descriptive

content seems to be unattractive. Therefore, this paper proposes "Undergrad"; a student life simulation game, to make the information more engaging and intelligible for teenagers. Undergrad is a 3D multiplayer game. A player assumes the role of a student to complete missions relating to three subject areas. The game not only promotes necessary information about the course but also helps the players to gain knowledge and experience of studying in the faculty, which is advantageous for decision making in choosing to study in the future.

**Keywords:** Simulation game, Gamification, IT KMITL.

#### 1. บทนำ

คณะเทคโนโลยีสารสนเทศ สจล. มีการจัดกิจกรรมทางวิชาการต่าง ๆ ขึ้นเป็นประจำ ทั้งนี้เพื่อประชาสัมพันธ์ให้นักเรียนหรือผู้ที่สนใจได้เข้ามาเยี่ยมชมนิทรรศการหรือกิจกรรมต่าง ๆ ที่คณะจัดขึ้น โดยมีรูปแบบที่หลากหลาย เช่น กิจกรรมเปิดบ้าน โพสต์บทความ หรือการถ่ายทอดสด ซึ่งเนื้อหาและรูปแบบมักมีลักษณะคล้ายเดิมในแต่ละครั้ง จึงอาจจะไม่สามารถดึงดูดนักเรียนหรือผู้ที่สนใจให้เข้ามาเยี่ยมชมได้เท่าที่ควร รวมทั้งวันและเวลาในการจัดกิจกรรม หรือความสะดวกในการเดินทางเพื่อมาเข้าร่วมที่คณะก็เป็นอีกหนึ่งปัญหาสำคัญ ถึงแม้ในปัจจุบันกิจกรรมต่าง ๆ ของคณะได้มีการปรับเปลี่ยนเป็นรูปแบบออนไลน์มากขึ้นเนื่องด้วยสถานการณ์โรคระบาด แต่ยังคงมีข้อจำกัดในการรับชม ส่งผลให้นักเรียนหรือผู้ที่สนใจเข้ามาเยี่ยมชมคณะได้รับประสบการณ์อย่างจำกัด

ด้วยปัญหาดังกล่าว ผู้พัฒนาจึงมีความคิดริเริ่มในการพัฒนาเกม “อันเดอร์แกรนด์ เกมจำลองชีวิตนักศึกษาไอที สจล.” ที่นำการบูรณาการของกลศาสตร์เกมมาช่วยในการประชาสัมพันธ์คณะ ให้นักเรียนหรือผู้ที่สนใจเข้ามาเยี่ยมชมคณะ เพื่อสร้างแรงจูงใจและความน่าตื่นเต้นในการเรียนรู้ โดยผู้เล่นจะได้สวมบทบาทสมมุติเป็นนักศึกษา คณะเทคโนโลยีสารสนเทศ สจล. ในการใช้ชีวิตภายในสถาบันและได้ความรู้และความเข้าใจเกี่ยวกับรายวิชาต่างๆ ในแต่ละแขนงวิชาผ่านการเล่นเกมในรูปแบบที่เข้าใจได้ง่ายและสนุกสนานไปพร้อมกัน

## 2. ทฤษฎีและเทคโนโลยีที่เกี่ยวข้อง

### 2.1 Gamification

เกมมิฟิเคชัน [1] เป็นรูปแบบการเรียนรู้ที่ได้นำเอาหลักการพื้นฐานในการออกแบบเกม กลไกการเล่นเกม มาใช้เพื่อสร้างแรงจูงใจให้ผู้เล่นเพื่อให้เล่นเกมต่อไปเรื่อยๆ โดยจะมอบ “รางวัล” ให้แก่ผู้เล่น หลังจากทำภารกิจสำเร็จ ผู้เล่นจะได้รับรางวัลนั้น เช่น การได้ไอเท็มใหม่ การปลดล็อกด่านใหม่ ๆ เป็นต้น เมื่อได้รับรางวัลนั้นผู้เล่นจึงอยากเล่นต่อเพื่อหวังรางวัลชิ้นถัดไป ส่งเสริมให้ผู้เล่นเกิดพฤติกรรมมีส่วนร่วมกับการฝึกในเกมส์ได้และทำให้เนื้อหาที่สอดแทรกภายในเกมส์มีความน่าสนใจ

### 2.2 Multiplayer video game

Multiplayer video game [2] เป็นการเล่นที่ผู้เล่นหลายคนสามารถเล่นได้ ในสภาพแวดล้อมเดียวกันภายในเกมส์ได้มากกว่าหนึ่งคน

### 2.3 Stylized Art

Stylized Art คือการนำเสนอรูปแบบของงานศิลป์ภายในเกมส์ โดยจะเน้นเกี่ยวกับฉาก สิ่งของ บุคคล รวมถึงสิ่งต่าง ๆ ภายในเกมส์ให้อยู่ในรูปแบบเกินกว่าความเป็นจริง เพื่อถ่ายทอดถึงจินตนาการของผู้พัฒนาให้ได้มากที่สุด

### 2.4 หลักสูตรปริญญาตรีของคณะเทคโนโลยีสารสนเทศ

ในระดับปริญญาตรี คณะเทคโนโลยีสารสนเทศ สจล. [3] มีการเปิดการเรียนการสอนอยู่ด้วยกันทั้งหมด 3 หลักสูตรได้แก่ หลักสูตรเทคโนโลยีสารสนเทศ หลักสูตรวิทยาการข้อมูลและการวิเคราะห์เชิงธุรกิจ และหลักสูตรเทคโนโลยีสารสนเทศทางธุรกิจ (หลักสูตรนานาชาติ)

โดยเนื้อหาและภารกิจภายในเกมส์จะอ้างอิงจากหลักสูตรเทคโนโลยีสารสนเทศเท่านั้น ซึ่งประกอบด้วยวิชาเฉพาะแขนง 3 แขนงวิชา ได้แก่ แขนงวิชาวิศวกรรมซอฟต์แวร์ แขนงวิชาเทคโนโลยีเครือข่ายและระบบ และ แขนงวิชาการพัฒนาสื่อประสมและเกม

### 2.5 โปรแกรมที่ใช้ในการพัฒนาเกม

1) Unreal Engine [4] คือ ชุดเครื่องมือพัฒนาที่สมบูรณ์ที่ทำงานด้วยเทคโนโลยีแบบเรียลไทม์ ใช้สำหรับพัฒนางานด้านสามมิติในอุตสาหกรรมต่าง ๆ

2) Autodesk Maya [5] คือ ซอฟต์แวร์เพื่องานด้านงานกราฟิกสามมิติแบบครบวงจร

3) Substance Painter คือ ซอฟต์แวร์ที่ใช้ในการปรับแต่งพื้นผิววัตถุในงานสามมิติ

4) ZBrush [6] คือ ซอฟต์แวร์สำหรับงานประติมากรรมดิจิทัล ระบายสีแบบดิจิทัล งานแบบนูนต่ำ การสร้างพื้นผิว ซึ่งนิยมใช้ในการทำโมเดลที่มีความละเอียดสูง

5) Adobe Illustrator คือ ซอฟต์แวร์ที่ใช้ออกแบบงานด้านกราฟิก โดยลักษณะของการทำงานของโปรแกรมจะแสดงภาพเป็นเวกเตอร์ (Vector)

### 2.6 ส่วนเสริมที่ใช้ในการพัฒนาเกม

Advanced Sessions Plugin คือ เครื่องมือเสริมสำหรับโปรแกรม Unreal Engine เพื่อเชื่อมต่อระบบโหมดผู้เล่นหลายคนระหว่างตัวเกมส์และระบบสตรีม

## 3. การวิเคราะห์และออกแบบระบบ

### 3.1 การเก็บข้อมูลและวิเคราะห์ความต้องการของผู้ใช้

#### 3.1.1 ผู้ตอบแบบสอบถาม

ผู้ที่ร่วมตอบแบบสอบถาม เป็นกลุ่มตัวอย่างจากกลุ่มเป้าหมายที่ต้องการประชาสัมพันธ์หลักสูตร โดยมีเกณฑ์คุณสมบัติผู้เข้าร่วมการทำแบบสอบถาม ได้แก่ ไม่จำกัดเพศ อายุ 18 ปีขึ้นไป และมีระดับการศึกษามัธยมศึกษาตอนปลายขึ้นไป เป็นผู้ที่สนใจเข้าศึกษาต่อหรือผู้ที่ต้องการศึกษาหาข้อมูลเกี่ยวกับหลักสูตรของคณะเทคโนโลยีสารสนเทศ สจล.

#### 3.1.2 เครื่องมือและเทคนิคที่ใช้ในการเก็บข้อมูล

เป็นการเก็บรวบรวมความต้องการของผู้ใช้ภายใต้หัวข้อ “แบบสอบถามเพื่อพัฒนาเกม เรื่อง อันเดอร์เกรด : เกมจำลองชีวิตนักศึกษาไอที สจล.” เพื่อใช้ในการวิเคราะห์ระบบภายในเกม ซึ่งแบบสอบถามจะอยู่ในรูปแบบของกูเกิลฟอร์ม (Google Forms)

3.1.3 วิเคราะห์ความต้องการของผู้ใช้

ผู้วิจัยได้ทำการวิเคราะห์ข้อมูลความสนใจและความต้องการจากผู้ใช้งานแบบสอบถามซึ่งใช้ระยะเวลาในการรวบรวมความต้องการทั้งสิ้น 7 วันและมีผู้ตอบแบบสอบถามทั้งสิ้น 63 คน ซึ่งมีผลลัพธ์ดังนี้

1) อันดับความสนใจข้อมูลด้านต่าง ๆ ตามลำดับ

- ข้อมูลด้านสภาพแวดล้อมภายในคณะ (87.3%)
- ข้อมูลด้านหลักสูตร (84.1%)
- ข้อมูลด้านการใช้ชีวิต และอัตลักษณ์ของคณะ (82.5%)
- ข้อมูลกิจกรรม และสันตนาการ (74.6%)

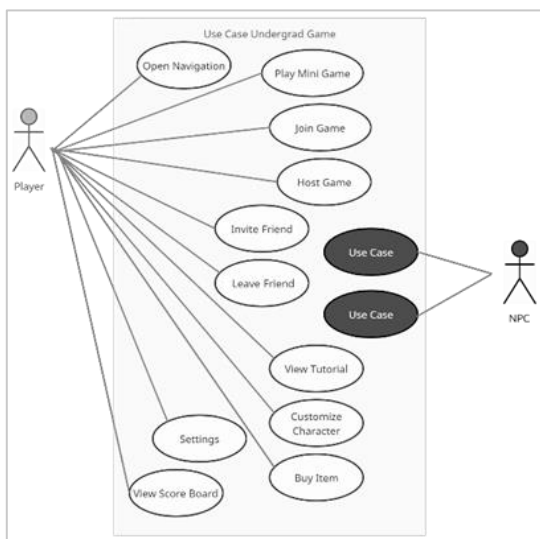
2) ความต้องการเกี่ยวกับการพัฒนาเกม

ต้องการฟีเจอร์เสริมเกี่ยวกับ ระบบแผนที่, ระบบนำทาง, ระบบสอนการเล่น, ระบบแต่งตัว, ระบบร้านค้า, ระบบคำใบ้ และข้อมูลแนะนำการเดินทางมาที่คณะ รวมถึงบอกถึงการประกอบอาชีพของสายงานนี้ในอนาคต

3.2 การออกแบบระบบเกม

3.2.1 แผนภาพยูสเคส

ผู้วิจัยได้ออกแบบระบบเกม ซึ่งมีฟังก์ชันการทำงานตามแผนภาพยูสเคส ดังภาพที่ 1



ภาพที่ 1: แสดงแผนภาพยูสเคส (Use Case Diagram)

ยูสเคสของเกม Undergrad ประกอบด้วยแอกเตอร์ 2 แอกเตอร์ และมีฟังก์ชันที่เกี่ยวข้องดังนี้

1) ยูสเคสของผู้เล่น

- Play Mini Games คือ การกิกต่างๆ ภายในเกม
- Join Game คือ ยอมรับหรือปฏิเสธคำเชิญ เข้าร่วมเล่นเกมกับเพื่อนบนสตีม
- Invite Friend คือ ชวนเพื่อนบนสตีมเล่นเกมได้
- Leave Friend คือ ผู้เล่นที่เป็นเจ้าของห้องสามารถลบผู้เล่นอื่นออกจากห้องได้
- Buy Item คือ ซื้อไอเท็มบนร้านค้าได้
- Customize Character คือ เลือกและตกแต่งตัวละครได้
- View Score Board คือ ดูผลสรุปของเกรดเฉลี่ยหลังจบเกมได้
- View Tutorial คือ ดูบทช่วยสอนเล่นเกมได้
- Settings คือ ระบบตั้งค่าเกมได้
- Open Navigation คือ ใช้ระบบนำทาง เพื่อนำทางไปยังตำแหน่งภารกิจได้

2) ยูสเคสของ NPC

- Move คือ ตัวละครที่ไม่ใช่ผู้เล่นจะสามารถเดินรอบคณะได้
- Interact คือ เมื่อผู้เล่นสนทนากับตัวละครที่ไม่ใช่ผู้เล่น จะทำการสุ่มบทสนทนาสั้น ๆ ขึ้นมา

3.3 การออกแบบภารกิจเกม

ส่วนของการออกแบบภารกิจต่าง ๆ ในเกมเป็นส่วนสำคัญมากที่สุดเพื่อสื่อให้ผู้เล่นเข้าใจภาพรวมการเรียนการสอนในแต่ละแขนงวิชาของสาขาเทคโนโลยีสารสนเทศ จึงนำเอาเนื้อหาความรู้เบื้องต้นของแต่ละวิชาเฉพาะแขนงมาใช้ออกแบบภารกิจต่าง ๆ นอกจากนี้ ยังมีภารกิจทั่วไปเกี่ยวกับชีวิตประจำวัน ซึ่งภารกิจจะไม่ยากหรือง่ายจนเกินไป จึงเหมาะสำหรับผู้เล่นระดับเบื้องต้น

ภารกิจในเกมมีจำนวนทั้งหมด 10 ภารกิจ โดยแบ่งออกเป็น 5 หมวด โดยแบ่งหมวดละ 2 ภารกิจ ซึ่งแสดงรายละเอียด ดังต่อไปนี้

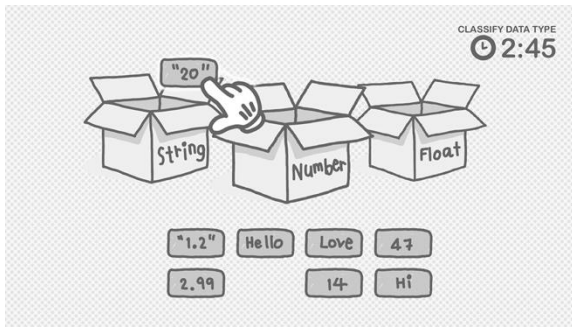
3.3.1 หมวดวิศวกรรมซอฟต์แวร์

1) ภารกิจแยกประเภทตัวแปร (Classify Data Type)

ผู้เล่นจะต้องทำการเลือกตัวแปรต่าง ๆ ที่กำหนด แล้วนำไปวางในกล่องให้ถูกต้องตามชนิดของตัวแปร ภายในระยะเวลาที่กำหนด

2) การฝึกงานทดสอบระบบ (Software Tester)

ผู้เล่นจะต้องทำการตรวจสอบคุณสมบัติของระบบซอฟต์แวร์ ว่าทำงานถูกต้องหรือไม่ ภายในระยะเวลาที่กำหนด



ภาพที่ 2: แสดงหน้าภารกิจแยกประเภทตัวแปร

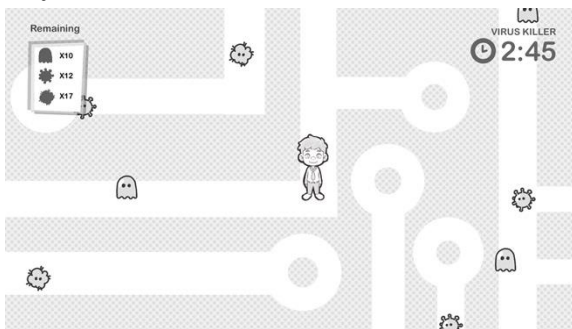
3.3.2 หมวดเทคโนโลยีเครือข่ายและระบบ

1) การฝึกตามล่าหาไวรัส (Virus Killers)

ผู้เล่นจะต้องใช้เครื่องดูดฝุ่นไปทำการเก็บไวรัสในแผนที่ที่มีการการเกิดไวรัส โดยจะมีการกำหนดจำนวนและประเภทของไวรัสที่ต้องเก็บ เช่น เก็บ Worms ให้ครบ 10 ตัว, เก็บ Spy ให้ครบ 5 ตัว เป็นต้น ซึ่งผู้เล่นจะต้องทำให้สำเร็จภายในระยะเวลาที่กำหนด

2) การฝึกถอดรหัสแฮช (Hashing Password)

ผู้เล่นจะต้องทำการเข้าสู่ระบบบนเว็บไซต์ ซึ่งสิ่งที่ผู้เล่นจะต้องทำคือจะต้องหารหัสผ่านให้พบ โดยจะมีเพียงข้อความแฮชเท่านั้น และให้ผู้เล่นไปทำการถอดรหัสเอง ซึ่งผู้เล่นจะต้องทำให้สำเร็จภายในระยะเวลาที่กำหนด



ภาพที่ 3: แสดงหน้าภารกิจตามล่าหาไวรัส

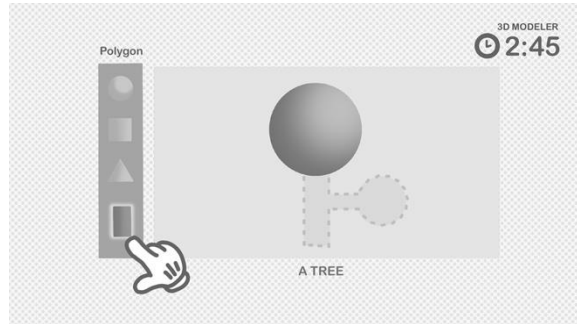
3.3.3 หมวดการพัฒนาสื่อประสมและเกม

1) การฝึกนักปั้น โมเดลสามมิติ (3D Modeler)

ผู้เล่นต้องประกอบโพลีกอนพื้นฐานของโมเดลสามมิติ ให้มีรูปร่างตามโจทย์ที่ได้รับมอบหมาย ภายในระยะเวลาที่กำหนด ซึ่งภารกิจจะมีลักษณะคล้ายกับการต่อจิ๊กซอว์

2) การฝึกสร้างเว็บไซต์ (WEB Programmer)

ผู้เล่นจะต้องทำการลากวางคำสั่งที่กำหนด เพื่อเปลี่ยนสี, ขนาด และรูปแบบเพื่อสร้างเว็บไซต์ให้สมบูรณ์ ภายในระยะเวลาที่กำหนด ซึ่งภารกิจจะมีลักษณะคล้ายกับการเติมคำศัพท์ลงในช่องว่าง



ภาพที่ 4: แสดงหน้าภารกิจนักปั้นโมเดลสามมิติ

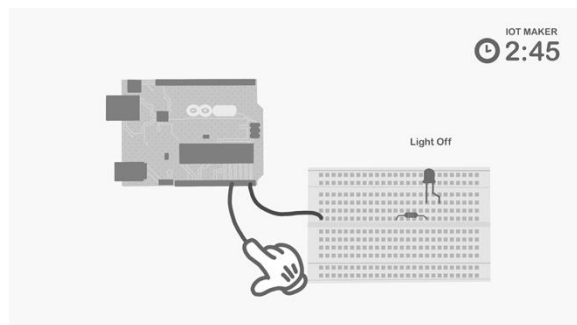
3.3.4 หมวดเทคโนโลยีสารสนเทศ

1) การฝึกต่อสายไฟไอโอที (IOT Maker)

ผู้เล่นจะต้องทำการต่อสายไฟเพื่อเปิดปิดไฟ ซึ่งจะต้องทำการต่อสายไฟผ่านบอร์ดอาณูโนภายในระยะเวลาที่กำหนด

2) กระบวนการคิดเชิงคำนวณ (Computer Thinking)

ผู้เล่นจะต้องทำการเติมคำตอบในช่องว่างให้ถูกต้องภายในระยะเวลาที่กำหนด ซึ่งจะเป็นโจทย์ในลักษณะของแนวคิดเชิงคำนวณ



ภาพที่ 5: แสดงหน้าภารกิจต่อสายไฟไอโอที

3.3.5 หมวดเกี่ยวกับชีวิตประจำวัน

1) การฝึกปลุกเพื่อนจากเก้าอี้ดูวิญญาณ (Wake Friends Up)

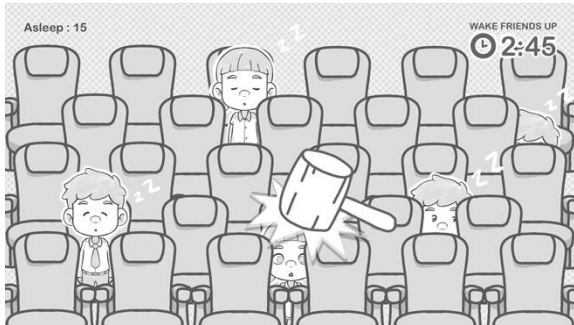
ผู้เล่นจะได้เข้าไปนั่งเรียนที่ห้องประชุมออกติ-ทอเรียน และได้รับการฝึกจากอาจารย์ โดยจะต้องปลุกเพื่อนที่กำลัง



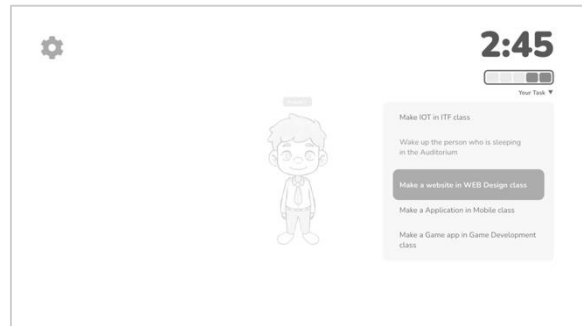
หลับอยู่ให้ตื่น ซึ่งภารกิจจะมีลักษณะคล้ายกับเกมทบทวน

2) ภารกิจบอร์ดเกมนี้ฉันจอง (Reserved This Board)

ผู้เล่นจะได้รับภาพเงาปริศนา ซึ่งเป็นรูปภาพโลโก้ของบอร์ดเกมต่าง ๆ ที่ถูกวางไว้ภายในห้องคอมมอน ซึ่งผู้เล่นจะต้องตามหาให้ครบภายในระยะเวลาที่กำหนด



ภาพที่ 6: แสดงหน้าภารกิจปลุกเพื่อนจากเก้าอี้ดูคู่มือ

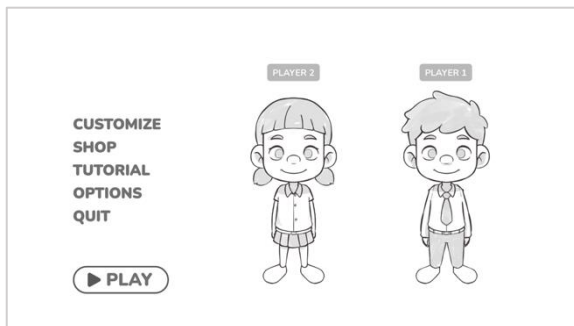


ภาพที่ 8: แสดงหน้าการเล่นเกม

4. ระบบต้นแบบ

4.1 หน้าเมนูหลัก (Main menu)

ในหน้าเมนูหลักจะแสดงเมนูต่าง ๆ คือ Customize, Shop, Tutorial, Options, Quit และ Play รวมทั้งแสดงตัวละครของผู้เล่นทั้งหมดในหน้านี้ด้วยเช่นกัน ดังภาพที่ 7



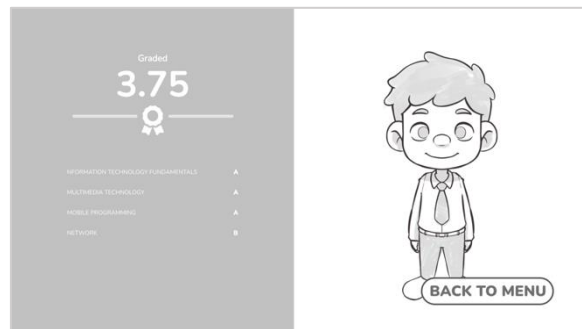
ภาพที่ 7: แสดงหน้าเมนูหลัก

4.2 หน้าเล่นเกม (Game Player)

โดยจะมีองค์ประกอบที่แสดงเวลาในการเล่น และแถบความคืบหน้าของภารกิจที่ทำ สามารถกดเพื่อดูรายการภารกิจได้ นอกจากนี้ยังมีปุ่มตั้งค่า สำหรับตั้งค่าระหว่างเกม หรือออกจากเกมอยู่บริเวณมุมซ้ายบนของหน้าจอ ดังภาพที่ 8

4.3 หน้าสรุปผลคะแนน (Summary)

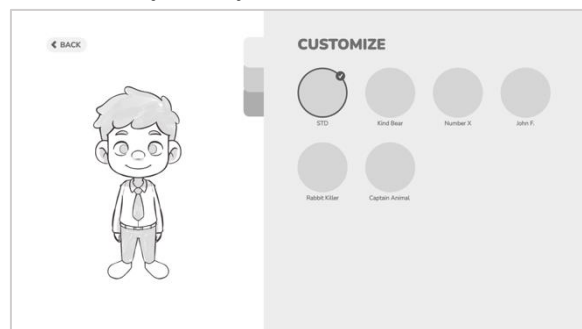
โดยจะแสดงคะแนนที่ผู้เล่นได้จากการเล่นเกมเป็นเกรด A – F และสรุปคะแนนเป็นเกรดเฉลี่ย เมื่อกดปุ่ม Back to menu จะนำผู้เล่นกลับสู่หน้าเมนูหลัก ดังภาพที่ 9



ภาพที่ 9: แสดงหน้าสรุปผลคะแนน

4.4 หน้าปรับแต่งตัวละคร (Customize)

โดยหน้าปรับแต่งตัวละครจะแบ่งตามหมวดหมู่สิ่งของต่าง ๆ และแสดงตัวละครเพื่อดูผลการปรับแต่ง เมื่อกดปุ่ม Back จะกลับสู่หน้าเมนูหลัก ดังภาพที่ 10



ภาพที่ 10: แสดงหน้าปรับแต่งตัวละคร

5. ผลการดำเนินงาน

ผู้วิจัยได้ดำเนินการออกแบบและพัฒนาทั้งในส่วนจของรายละเอียดเกม และระบบเกม โดยได้มีการรวบรวมสรุป

ข้อมูล และจัดทำแบบสอบถามเพื่อประเมินความพึงพอใจ เกี่ยวกับการออกแบบภารกิจเกมกับนักศึกษาชั้นปีที่ 4 ใน สาขาเทคโนโลยีสารสนเทศ ซึ่งมีผู้สนใจเข้าร่วมตอบ แบบสอบถามจำนวน 35 คน ซึ่งผู้พัฒนาได้กำหนดระดับ ความพึงพอใจจาก น้อยที่สุดถึงมากที่สุด เท่ากับ 1 – 5 คะแนน โดยสามารถสรุปผลได้ดังตารางที่ 1

ตารางที่ 1: สรุปผลประเมินความพึงพอใจ

สรุปผลประเมินความพึงพอใจ	คะแนน
1. ภาพรวมระบบของเกมมีความเหมาะสม และน่าสนใจ	4.2
2. เนื้อหาภารกิจเกมมีความเหมาะสมกับ กลุ่มเป้าหมาย	4.37
3. เนื้อหาภารกิจเกมสะท้อน/สื่อถึงการเรียน/ วัตถุประสงค์/การใช้ชีวิตในคณะไอที	4.34
4. ความสามารถในการประชาสัมพันธ์ หลักสูตรของคณะไอที	4.43

จากตารางที่ 1 สามารถสรุปได้ว่าส่วนใหญ่ผู้เข้าร่วม ตอบแบบสอบถามมีความพึงพอใจต่อการออกแบบ และ นอกจากนี้ยังมีความคิดเห็นเพิ่มเติมที่น่าสนใจ เช่น คาดหวังถึงประสบการณ์การเล่นที่สามารถเป็นส่วนช่วย ในการตัดสินใจในการเลือกเรียนต่อได้ในอนาคต และ อยากให้มีเนื้อหาที่เข้าใจง่ายเป็นต้น ซึ่งผู้พัฒนาจะได้นำมา ประยุกต์ใช้ในการพัฒนาระบบต่อไป

## 6. สรุป

บทความนี้นำเสนอการพัฒนาเกม “แอนเดอร์แกรด เกม จำลองชีวิตนักศึกษาไอที สจล.” เพื่อช่วยในการ ประชาสัมพันธ์หลักสูตรของคณะเทคโนโลยีสารสนเทศ ในรูปแบบของการเล่นเกมคอมพิวเตอร์ โดยภารกิจต่าง ๆ ภายในเกมได้สะท้อนเนื้อหาของแต่ละแขนงวิชาใน หลักสูตร เพื่อให้ผู้เล่นได้เข้าใจภาพรวมการเรียนการสอน ในแต่ละแขนงวิชาผ่านภารกิจที่มีวิธีเล่นที่หลากหลาย

จากแบบสำรวจความพึงพอใจ แสดงให้เห็นว่า เนื้อหา ภารกิจเกมแอนเดอร์แกรดเป็นที่พึงพอใจของผู้เข้าร่วมตอบ

แบบสอบถามส่วนใหญ่ ด้วยมองว่าเป็นวิธีที่มี ประสิทธิภาพในการสื่อถึงการเรียน การใช้ชีวิตและอัต ลักษณ์ของคณะไอทีได้เป็นอย่างดี แต่ถึงอย่างไรก็ตาม ผล การสำรวจนี้ มาจากนักศึกษาที่เรียนในสาขาเทคโนโลยี สารสนเทศเท่านั้น แต่ยังคงการสำรวจจากอาจารย์ผู้สอน ที่เกี่ยวข้องในแขนงนั้น ๆ

นอกจากนี้ ผู้วิจัยได้พัฒนาสภาพแวดล้อมภายในเกม ให้ใกล้เคียงกับสภาพแวดล้อมจริงมากที่สุด จึงสามารถ พัฒนาต่อเพื่อนำไปใช้ในกิจกรรมเปิดบ้านออนไลน์ โดยผู้ เล่นสามารถเข้ามาเยี่ยมชมนิทรรศการต่าง ๆ ได้ภายในเกม นี้ อีกทั้งยังสามารถพัฒนาต่อเพื่อบูรณาการเข้ากับการเรียน การสอนของคณะ โดยเพิ่มพีเจอร์ให้กับเกมเพื่อใช้เป็น แพลตฟอร์มในการจัดการเรียนการสอนออนไลน์ได้

## เอกสารอ้างอิง

- [1] ชนัตต์ พูนเดช และ ธนิตา เลิศพรกุลรัตน์, “แนวทางการ จัดการเรียนรู้อด้วยแนวความคิดเกมมิฟิเคชัน”, *วารสารศึกษาศาสตร์ มหาวิทยาลัยนครสวรรค์* ปีที่ 18 ฉบับที่ 3 กันยายน พ.ศ. 2559 หน้า 331-339.
- [2] Mary Keo, “Graphical Style in Video Games” [Online]. เข้าถึงจาก: [https://www.theseus.fi/bitstream/handle/10024/133067/Keo\\_Mary.pdf](https://www.theseus.fi/bitstream/handle/10024/133067/Keo_Mary.pdf). สืบค้นเมื่อ 3 สิงหาคม 2564.
- [3] คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง, “ข้อมูลหลักสูตรคณะเทคโนโลยี สารสนเทศ สจล.” [Online]. เข้าถึงจาก : <https://www.it.kmitl.ac.th/th/program/>. สืบค้นเมื่อ 3 สิงหาคม 2564.
- [4] Epic Games, “Why Unreal Engine?” [Online]. เข้าถึง จาก : <https://www.unrealengine.com/en-US/unreal>. สืบค้นเมื่อ 6 สิงหาคม 2564.
- [5] “สร้างงาน 3D และแอนิเมชันด้วย MAYA 2018 ฉบับ สมบูรณ์”, กรุงเทพฯ: ริโวว่า, พ.ศ. 2563 หน้า 7.
- [6] “สร้างโมเดล 3 มิติด้วย ZBrush 2018”, กรุงเทพฯ: ริโวว่า, ตุลาคม พ.ศ. 2562 หน้า 5.

# การประยุกต์ใช้ปัญญาประดิษฐ์กับหุ่นยนต์สนทนาบนระบบไลน์ Applied an Artificial Intelligence with Chatbot's on Line Office Account

ธีรวัฒน์ เพ็ชรน้ำ (Theerawat Patcham)<sup>1</sup> และมณฑิธร รัตนศิริวงศ์วุฒิ (Montean Rattanasiriwongwut)<sup>2</sup>

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>1</sup>s6307021856080@email.kmutnb.ac.th, <sup>2</sup>montean.r@itd.kmutnb.ac.th

## บทคัดย่อ

บทความนี้เป็นการศึกษาและเปรียบเทียบอัลกอริทึมจากไลบรารี PyThaiNLP ที่ใช้สำหรับการตัดคำ การสะกดคำ และการเพิ่มข้อความหรือแต่งเติมข้อความด้วยภาษาไพธอน เพื่อประเมินผลหาค่าความถูกต้องและเปรียบเทียบอัลกอริทึมที่มีประสิทธิภาพและเหมาะสมสำหรับการพัฒนาระบบหุ่นยนต์สนทนา เพื่อนำมาสร้างแบบจำลองหุ่นยนต์สนทนาบนระบบไลน์ของภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ โดยมีวัตถุประสงค์เพื่อช่วยให้นักศึกษาของภาควิชาฯ สามารถเข้าถึงสารสนเทศต่าง ๆ ของภาควิชาฯ ได้สะดวกยิ่งขึ้น

**คำสำคัญ:** ปัญญาประดิษฐ์ หุ่นยนต์สนทนา การประมวลผลภาษาธรรมชาติ

## Abstract

*This article explores and compares algorithms from PyThaiNLP library for word segmentation, spelling, and augmentation with Python programming. To evaluate the accuracy and compare efficient algorithm and suitable for development chatbot systems to create a model of chatbot on the Line system for Department of Information Technology Management Faculty of Information Technology and Digital Innovation at King Mongkut's University of Technology North Bangkok. The purpose is to help students in the*

*department can access information of the department easier.*

**Keyword:** Artificial Intelligence (AI), Chatbot, Natural Language Processing (NLP)

## 1. บทนำ

เทคโนโลยีเข้ามามีบทบาทในการใช้ชีวิตประจำวันของมนุษย์เป็นอย่างมาก เทคโนโลยีที่ได้รับความนิยมคือเทคโนโลยีทางด้านปัญญาประดิษฐ์ หรือ Artificial Intelligence (AI) ซึ่งเป็นวิธีการที่ทำให้คอมพิวเตอร์มีความสามารถใกล้เคียงกับมนุษย์หรือเลียนแบบพฤติกรรมของมนุษย์ เช่น การรับรู้ การเรียนรู้ การให้เหตุผล และการแก้ไขปัญหาต่าง ๆ [1] และอีกหนึ่งเทคโนโลยีที่กำลังได้รับความนิยมในปัจจุบันคือเทคโนโลยีทางการติดต่อสื่อสารผ่านสื่อสังคมออนไลน์ เนื่องจากมีความสะดวกรวดเร็วในการใช้งานและครอบคลุมทุกการติดต่อสื่อสาร

ในปัจจุบันผู้วิจัยได้ศึกษาอยู่ในภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ได้เล็งเห็นถึงปัญหาในการเข้าถึงข้อมูลสารสนเทศต่าง ๆ ของภาควิชาการจัดการเทคโนโลยีสารสนเทศ ไม่ว่าจะเป็นผลการเรียนของนักศึกษา ตารางเรียนหรือตารางสอบของนักศึกษา และคู่มือเอกสารต่าง ๆ ผู้วิจัยจึงได้นำเอาแนวคิดเกี่ยวกับเทคโนโลยีปัญญาประดิษฐ์กับหุ่นยนต์สนทนามาประยุกต์ใช้กับภาควิชาการจัดการเทคโนโลยีสารสนเทศ เพื่ออำนวยความสะดวก

สะดวกกับนักศึกษาให้สามารถเข้าถึงสารสนเทศต่าง ๆ ของภาควิชาได้สะดวกรวดเร็วยิ่งขึ้น

ในการนำเทคโนโลยีปัญญาประดิษฐ์กับหุ่นยนต์ สนทนามาประยุกต์ใช้งานกับภาควิชาฯ นั้น จำเป็นต้องมี เทคนิคหรืออัลกอริทึมที่ดีและเหมาะสมในการพัฒนา ระบบหุ่นยนต์สนทนา ผู้วิจัยจึงได้ทำการศึกษาค้นคว้า เกี่ยวกับการประมวลผลภาษาธรรมชาติ หรือ Natural Language Processing (NLP) ซึ่ง เป็น การ ทำ ให้ คอมพิวเตอร์สามารถสื่อสารได้ตอบด้วยภาษาของมนุษย์ และทำให้คอมพิวเตอร์เข้าใจภาษามนุษย์มากขึ้น ผู้วิจัยจึง ได้เลือกใช้ไลบรารี PyThaiNLP ในการพัฒนาระบบ

ผู้วิจัยจึงได้ทำการศึกษาเปรียบเทียบเทคนิคหรือ อัลกอริทึมจากไลบรารี PyThaiNLP ที่ใช้สำหรับการตัด คำ การสะกดคำ และการเพิ่มข้อความหรือแต่งเติมข้อความ เพื่อ ประเมิน ผล หา ค่า ความ ถูก ต้อง และ วิเคราะห์ เปรียบเทียบหาเทคนิคอัลกอริทึมที่มีประสิทธิภาพและ เหมาะสำหรับการพัฒนาระบบหุ่นยนต์สนทนา เพื่อนำมา สร้างแบบจำลองหุ่นยนต์สนทนาบนระบบออนไลน์

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ปัญญาประดิษฐ์ หรือ Artificial Intelligence (AI)

ปัญญาประดิษฐ์หรือ AI คือระบบประมวลผลที่มี ต้นแบบมาจาก โครงข่ายประสาทของมนุษย์ สามารถ เรียนรู้และเพิ่มประสิทธิภาพการประมวลผลได้ตาม จำนวนข้อมูลที่เพิ่มขึ้นผ่านกระบวนการเรียนรู้ด้วยตนเอง ซึ่งสามารถจดจำ คิด วิเคราะห์ เรียนรู้และเชื่อมโยงข้อมูล ต่าง ๆ ที่ซับซ้อนได้อย่างรวดเร็วเสมือนระบบสมองของ มนุษย์ ดังนั้น ปัญญาประดิษฐ์จึงถือเป็นเทคโนโลยีที่กำลัง ถูกจับตามองที่สุดในปัจจุบันและเข้ามามีบทบาทสำคัญต่อ การใช้ชีวิตการทำงาน รวมถึงการนำมาใช้ในการเสริม สักยภาพทางธุรกิจและอุตสาหกรรม ซึ่งจะส่งผล ต่อการเจริญเติบโตทางด้านเศรษฐกิจ

### 2.2 การประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาติ หรือ Natural Language Processing (NLP) คือวิทยาการแขนงหนึ่งใน หมวดหมู่ของเทคโนโลยีปัญญาประดิษฐ์ ซึ่งช่วยให้

คอมพิวเตอร์สามารถเข้าใจ ตลอดจนถึงความและใช้งาน ภาษาปกติที่มนุษย์ใช้สื่อสารได้ โดยเทคโนโลยี NLP มี รากฐานจากวิทยาการหลากหลายสาขาด้วยกัน โดยเฉพาะ ด้านวิทยาการคอมพิวเตอร์ (Computer Science) และ ภาษาศาสตร์เชิงคำนวณ (Computational Linguistics) เพื่อวัตถุประสงค์ในการปิดช่องว่างทางการสื่อสารระหว่าง มนุษย์และระบบคอมพิวเตอร์[2]

### 2.3 ไลบรารี PyThaiNLP

PyThaiNLP คือไลบรารี Python สำหรับงานทางด้านการประมวลผลข้อมูลภาษาไทย ถูกพัฒนาขึ้นมาโดยคนไทย มีฟังก์ชันที่มีประโยชน์มากมายสำหรับการ ประมวลผลภาษาไทย ใช้สำหรับประมวลผลข้อความและ การวิเคราะห์ทางภาษา เช่นเดียวกับ NLTK[3] แต่ใช้กับ ภาษาไทยโดยเฉพาะ มีฟังก์ชันการทำงานที่หลากหลาย เช่น Character Set อักษรไทย คำไทย, เรียงคำภาษาไทย, Stop Words ภาษาไทย, ตัดคำภาษาไทย, วิเคราะห์ชนิด ของคำทางไวยากรณ์, ตรวจสอบตัวสะกด แก่คำผิด เป็นต้น[4]

### 2.5 งานวิจัยที่เกี่ยวข้อง

Pattarawat Chormai [5] นำเสนอ งาน วิจัย เรื่อง AttaCut: A Fast and Accurate Neural Thai Word Segmenter การแบ่งคำนั้นเป็นขั้นตอนพื้นฐานหนึ่งสำหรับการ ประมวลผลภาษาธรรมชาติของไทย ผู้วิจัยจึงได้ทำการ เปรียบเทียบประสิทธิภาพในการตัดคำภาษาไทยทั้งใน เรื่องของความเร็วและความแม่นยำในการตัดคำ โดย เทคนิคที่จะนำมาเปรียบเทียบหรือวัดประสิทธิภาพ คือ Attacut และ Deepcut ผลการวิจัยพบว่า อัลกอริทึม Attacut มีการตัดคำที่เร็วกว่า Deepcut ถึง 5.6 เท่าเป็นอย่างน้อย และประสิทธิภาพการตัดคำสามารถเทียบเคียง Deepcut ได้ และมีบางกลุ่มข้อความที่สามารถทำได้ดีกว่า

แทนพงศ์ วัชรธาดา [6] นำเสนอ งาน วิจัย เรื่อง การ พัฒนาหุ่นยนต์สนทนาบนแอปพลิเคชันไลน์เพื่อ สนับสนุนการทำงานของแผนกทรัพยากรบุคคลแบบ อีเล็คทรอนิกส์ ซึ่งประยุกต์ใช้ปัญญาประดิษฐ์กับการ พัฒนาหุ่นยนต์สนทนา เพื่อที่จะช่วยลดการทำงานของ แผนกทรัพยากรบุคคล โดยนำวิธีการเรียนรู้ของเครื่องมา ใช้โดยวิธีการเรียนรู้แบบมีผู้สอน และใช้เทคนิคการตัดคำ

โดยใช้พจนานุกรม (Lexto) จากนั้นกำจัดคำหยุดและใช้ TF-IDF (Term Frequency-Inverse Document Frequency) เพื่อหาคำที่สำคัญในเอกสาร โดยนับความถี่ของคำที่ปรากฏในเอกสารและความถี่ผกผันกับเอกสารทั้งหมด ผลการวิจัยพบว่า จากข้อความจำนวน 20 ข้อความ ให้ ความถูกต้องร้อยละ 78 ซึ่งสามารถนำไปใช้ช่วยแผนก บุคคลตอบคำถามและให้ข้อมูลการลาแก่พนักงานใน องค์กร ได้

### 3. วิธีการดำเนินการวิจัย

วิธีการดำเนินงานวิจัยในการประยุกต์ใช้ ปัญญาประดิษฐ์กับหุ่นยนต์สนทนาบนระบบไลน์ ผู้วิจัยได้ วางแผนการดำเนินงานวิจัยและกำหนดขั้นตอนการ ดำเนินงานเพื่อเป็นแนวทางในการดำเนินงาน ดังนี้

#### 3.1 การรวบรวมข้อมูล

ผู้วิจัยได้รวบรวมชุดคำถามที่พบบ่อยจากภาควิชาฯ และนำมาแบ่งหมวดหมู่ออกเป็น 8 หมวดหมู่ ตัวอย่าง ข้อมูลแสดงดังตารางที่ 1

ตารางที่ 1 ตัวอย่างข้อมูลคำถามที่พบบ่อยจากภาควิชาฯ

หมวดหมู่	คำถาม
การสมัครเรียน	เปิดรับสมัครเรียนวันไหนคะ
หลักสูตร	ภาคเปิดหลักสูตรไหนบ้างครับ
การตรวจสอบผล การเรียน	เกรดออกเมื่อไรหะคะ
การตรวจสอบตาราง เรียนและตารางสอบ	ดูตารางเรียนจากได้ไหนได้บ้าง ค่ะ
เอกสาร	ดูตัวอย่างวิทยานิพนธ์ได้จาก ไหน
การลงทะเบียน	เปิดให้ลงทะเบียนเรียนวันไหน ครับ
การผ่อนผันการ ลงทะเบียน	ขอผ่อนผันการลงทะเบียนเรียน ครับ
การขออนุมัติ ผลสำเร็จการศึกษา	ขอจบทำเรื่องยัง ไงบ้างครับ

#### 3.2 การเปรียบเทียบอัลกอริทึม

ผู้วิจัยเลือกใช้อัลกอริทึมจากไลบรารี PyThaiNLP พัฒนาด้วยภาษา Python นำมาวัดประสิทธิภาพของแต่ละ อัลกอริทึมจากค่าความถูกต้องและค่าความเร็วในการ ประมวลผล ในการเปรียบเทียบอัลกอริทึมผู้วิจัยได้แบ่ง อัลกอริทึมออกเป็น 3 ประเภท ได้แก่

##### 3.2.1 การตัดคำ ประกอบด้วยอัลกอริทึม ดังนี้

- NEWMM
- LONGEST
- NERCUT
- DEEPCUT
- ATTACUT

```

# Compare Segmentation Algorithms
import csv
from pythainlp.tokenize import word_tokenize, subword_tokenize
from ast import literal_eval

with open('compare1.csv', encoding='utf-8') as f:
    reader = csv.reader(f, delimiter=',')
    line_count = 0

    for row in reader:
        if line_count == 0:
            line_count += 1
        else:
            spell_word = ''
            segment = word_tokenize(row[1].strip(), engine='attacut')
            print(segment)

# return longest nercut deapcut attacut
['ณั', 'ณ', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั']
['ณั', 'ณ', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั']
['ณั', 'ณ', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั']
['ณั', 'ณ', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั']
['ณั', 'ณ', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั', 'ณั']

```

ภาพที่ 1 : โปรแกรมสำหรับประมวลผลอัลกอริทึมการตัด คำ

##### 3.2.2 การสะกดคำ ประกอบด้วยอัลกอริทึม ดังนี้

- Correct(engine=PHUNSPELL)
- Correct (engine=SYMSPELLPY)
- Spell(engine=PHUNSPELL)
- Spell (engine=SYMSPELLPY)

```

# Spell Compare
import csv
from pythainlp.spell import correct, spell
from ast import literal_eval

with open('compare1.csv', encoding='utf-8') as f:
    reader = csv.reader(f, delimiter=',')
    line_count = 0

    for row in reader:
        if line_count == 0:
            line_count += 1
        else:
            for spellword in literal_eval(row[1].strip()):
                try:
                    spell_word = correct(spellword, engine='phunspell')
                    # print(spell_word)
                    if spellword.strip() != '' and len(spell_word) > 0 and spellword != spellword :
                        print('Incorrect word "' + spellword + '"', spell_word)
                except:
                    print("An exception occurred")

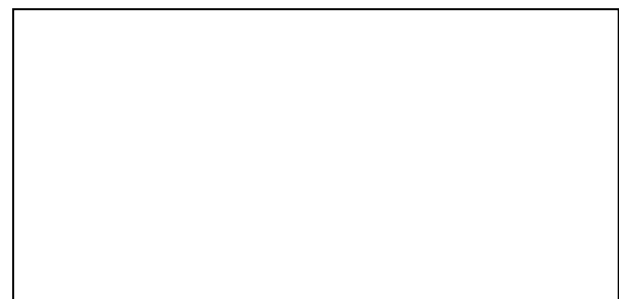
Incorrect word ณั => ณั
Incorrect word ณั => ณั
Incorrect word ณั => ณั
Incorrect word ณั => ณั
Incorrect word ณั => ณั
Incorrect word ณั => ณั
Incorrect word ณั => ณั
Incorrect word ณั => ณั
Incorrect word ณั => ณั

```

ภาพที่ 2 : โปรแกรมสำหรับประมวลผลอัลกอริทึมการ สะกดคำ

3.2.3 การเพิ่ม ข้อความ หรือ แต่งเติม ข้อความ ประกอบด้วยอัลกอริทึม ดังนี้

- WordNetAug
- Thai2fitAug
- LTW2V Aug
- BPEmbAug
- Thai2transformersAug



```

(9) # Augmentation Compare
import cv
from pythainlp.augment import Morphology
from pythainlp.augment.com2vec import Thai2FitAug, LSTMAug
from pythainlp.augment.com2vec.bpe2vec import BPEAug
from pythainlp.augment.le import Thai2TransformersAug

aug = Thai2TransformersAug()

with open('compare.csv', mode='w') as f:
    reader = csv.reader(f, delimiter=',')
    line_count = 0

    for row in reader:
        if line_count == 0:
            line_count += 1
        else:
            print(aug.augment(row[1].strip().replace("\n", "")))
            line_count += 1
    
```

**ภาพที่ 3 :** โปรแกรมสำหรับประมวลผลอัลกอริทึมการเพิ่มข้อความหรือแต่งเติมข้อความ

**3.3 การพัฒนาระบบ**  
ขั้นตอนในการพัฒนาระบบหุ่นยนต์สนทนาบนระบบ

ไลน์ สามารถแบ่งออกเป็น 5 ขั้นตอน ดังนี้

- 3.3.1 ขั้นตอนการพัฒนาฐานข้อมูล เพื่อใช้สำหรับเก็บข้อมูลนักศึกษา ข้อมูลผลการเรียน ข้อมูลรายวิชา และข้อมูลการตั้งค่าระบบต่าง ๆ
- 3.3.2 ขั้นตอนการพัฒนาส่วนติดต่อกับผู้ใช้
- 3.3.3 ขั้นตอนการพัฒนาหน้าเว็บไซต์สำหรับลงทะเบียน โดยใช้ LIFP ที่มีบนระบบไลน์ แสดงเป็นหน้าเว็บไซต์สำหรับลงทะเบียนในห้องแชท
- 3.3.4 ขั้นตอนการพัฒนา API สำหรับลงทะเบียน
- 3.3.5 ขั้นตอนการพัฒนาแบบจำลอง

**4. ผลการดำเนินงาน**

ผลการดำเนินการเปรียบเทียบอัลกอริทึมและการพัฒนาหุ่นยนต์สนทนาบนระบบไลน์ ผู้วิจัยได้ทำการแบ่งออกเป็นแต่ละประเภท ดังนี้

**4.1 ผลการเปรียบเทียบอัลกอริทึม**

ในการดำเนินการเปรียบเทียบอัลกอริทึม ผู้วิจัยได้แบ่งการเปรียบเทียบออกเป็น 3 ประเภทคือ การเปรียบเทียบอัลกอริทึมการตัดคำ การเปรียบเทียบอัลกอริทึมการสะกดคำ และการเปรียบเทียบอัลกอริทึมการเพิ่มข้อความหรือแต่งเติมข้อความ

**4.1.1 การเปรียบเทียบอัลกอริทึมการตัดคำ**

ทำการเปรียบเทียบ ค่าความถูกต้องแม่นยำและความเร็วในการประมวลผลของแต่ละอัลกอริทึมจากชุดข้อมูลเดียวกัน ความเร็วในการประมวลผลนั้นผู้วิจัยได้เลือกใช้โมดูล timeit ที่มีมาในภาษา python ในการวัด

ความเร็วในการประมวลผล โดยใช้พารามิเตอร์ number=10 เพื่อทดสอบความเร็วในการประมวลผลของแต่ละอัลกอริทึม

**ตารางที่ 2 :** ผลการเปรียบเทียบอัลกอริทึมการตัดคำ

Algorithms	Accuracy (%)	Time (Second)
NEWM	83.02119	0.01192
LONGEST	82.05819	0.03004
NERCUT	79.67000	0.06957
DEEPCUT	97.19349	6.97932
ATTACUT	96.34002	1.24406

จากตารางที่ 2 แสดงผลการเปรียบเทียบอัลกอริทึมที่ใช้ในการตัดคำ อัลกอริทึม NEWM LONGEST และ NERCUT นั้นมีค่าความถูกต้องที่ค่อนข้างต่ำผู้วิจัยจึงไม่เลือกใช้อัลกอริทึมทั้ง 3 อัลกอริทึมนี้ ส่วนอัลกอริทึม DEEPCUT นั้นให้ค่าความถูกต้องในการตัดคำมากที่สุด แต่ใช้เวลาในการประมวลผลค่อนข้างมาก ผู้วิจัยจึงเลือกใช้อัลกอริทึม ATTACUT ในการตัดคำเนื่องจากให้ค่าความถูกต้องที่ใกล้เคียงกับ DEEPCUT แต่ใช้ระยะเวลาในการประมวลผลน้อยกว่าอัลกอริทึม DEEPCUT ค่อนข้างมาก ผู้วิจัยจึงตัดสินใจเลือกใช้อัลกอริทึม ATTACUT ในการตัดคำ

**4.1.2 การเปรียบเทียบอัลกอริทึมการสะกดคำ**

ทำการเปรียบเทียบ ค่าความถูกต้องแม่นยำและความเร็วในการประมวลผลของแต่ละอัลกอริทึมจากชุดข้อมูลเดียวกัน ความเร็วในการประมวลผลนั้นผู้วิจัยได้เลือกใช้โมดูล timeit ที่มีมาในภาษา python ในการวัดความเร็วในการประมวลผลเช่นเดียวกับการเปรียบเทียบอัลกอริทึมการตัดคำ โดยใช้พารามิเตอร์ number=10 เพื่อทดสอบความเร็วในการประมวลผลของแต่ละอัลกอริทึม

**ตารางที่ 3 :** ผลการเปรียบเทียบอัลกอริทึมการสะกดคำ

Algorithms	Accuracy (%)	Time (Second)
Correct(PHUNSPELL)	46.15384	2.09795
Correct(SYMSPELLPY)	23.07692	0.00432
Spell(PHUNSPELL)	46.15384	2.02693
Spell(SYMSPELLPY)	19.23076	0.00559

จากตารางที่ 3 แสดงผลการเปรียบเทียบอัลกอริทึมการสะกดคำ ซึ่งจากตารางพบว่าแต่ละอัลกอริทึมให้ค่าความถูกต้องที่ค่อนข้างต่ำ เนื่องจากชุดข้อมูลที่นำมาเข้ากระบวนการเปรียบเทียบนั้นเป็นชุดข้อมูลที่มาจากการตัดคำที่ผิดพลาด เช่น คำว่า “ประเมียบอาจารย์” เมื่อผ่านอัลกอริทึมตัดคำแล้วจะได้คำว่า “ประเมียบอาจารย์” ผลลัพธ์ที่ถูกต้องควรเป็น ประเมียบ | อาจารย์ เมื่อนำคำว่า “ประเมียบอาจารย์” มาผ่านอัลกอริทึมสะกดคำแล้วจึงทำให้มีการสะกดคำที่ไม่ถูกต้องเกิดขึ้น ทำให้ค่าความถูกต้องที่ได้้นั้นค่อนข้างต่ำ

สรุปอัลกอริทึม Correct(engine=SYMSPELLPY) และอัลกอริทึม Spell(engine=SYMSPELLPY) นั้นให้ค่าความถูกต้องที่ค่อนข้างต่ำผู้วิจัยจึงไม่นำทั้งสองอัลกอริทึมนี้มาพิจารณา ส่วนอัลกอริทึม Correct (engine=PHUN-SPELL) และ อัล ก อ ริ ทึ ม Spell(engine=PHUNSPELL) นั้นให้ค่าความถูกต้องและเวลาในการประมวลผลที่ใกล้เคียงกัน ผู้วิจัยจึงตัดสินใจเลือกใช้อัลกอริทึม Correct(engine=PHUNSPELL) เนื่องจากอัลกอริทึมนี้สามารถนำไปประยุกต์ใช้ได้ง่ายและสะดวกกว่าอัลกอริทึม Spell(engine= PHUNSPELL) ที่ได้ให้ผลลัพธ์เป็นชุดข้อมูลแบบ array ทำให้ต้องเพิ่มขึ้นตอนการใช้งานมากขึ้น ผู้วิจัยจึงไม่เลือกใช้อัลกอริทึมนี้ในการพัฒนาแบบจำลอง

**4.1.3 การเปรียบเทียบอัลกอริทึมการเพิ่มข้อความหรือแต่งเติมข้อความ**

ทำการเปรียบเทียบความเร็วในการประมวลผลของแต่ละอัลกอริทึมจากชุดข้อมูลเดียวกัน ความเร็วในการประมวลผลนั้นผู้วิจัยได้เลือกใช้โมดูล timeit ที่มีมาในภาษา python ในการวัดความเร็วในการประมวลผล เช่นเดียวกับการเปรียบเทียบอัลกอริทึมการตัดคำ โดยใช้พารามิเตอร์ number=10 เพื่อทดสอบความเร็วในการประมวลผลของแต่ละอัลกอริทึม

อัลกอริทึมการเพิ่มข้อความหรือการแต่งเติมข้อความนั้นไม่สามารถเปรียบเทียบค่าความถูกต้องแม่นยำได้เนื่องจากผลลัพธ์ที่ได้จากการประมวลผลนั้นมีความ

แตกต่างกันของแต่ละอัลกอริทึม ผู้วิจัยจึงเลือกเปรียบเทียบแต่ละอัลกอริทึมจากระยะเวลาในการประมวลผลเพียงอย่างเดียว

**ตารางที่ 4 :** ผลการเปรียบเทียบอัลกอริทึมการเพิ่มข้อความหรือแต่งเติมข้อความ

Algorithms	Time (Second)
WordNetAug	0.00586
Thai2fitAug	0.31930
LTW2VAug	5.75606
BPEmbAug	0.66799
Thai2transformersAug	2.97457

จากตารางที่ 4 แสดงผลการเปรียบเทียบอัลกอริทึมการเพิ่มข้อความหรือแต่งเติมข้อความ ซึ่งอัลกอริทึม WordNetAug และ Thai2fitAug ใช้เวลาในการประมวลผลน้อย แต่ผลลัพธ์ที่ได้้นั้นไม่ครอบคลุมและมีความคลาดเคลื่อนในความหมายของรูปประโยค อัลกอริทึม LTW2VAug ให้ผลลัพธ์ของรูปประโยคได้ดี แต่ใช้เวลาในการประมวลผลค่อนข้างมาก อัลกอริทึม BPEmbAug ใช้เวลาในการประมวลผลค่อนข้างน้อย แต่ผลลัพธ์ที่ได้เหมือนกับประโยคต้นฉบับ อัลกอริทึม Thai2transformersAug ใช้เวลาในการประมวลผลที่ระดับปานกลาง ผลลัพธ์ที่ได้มีความสมบูรณ์ของประโยค ผู้วิจัยจึงเลือกอัลกอริทึม Thai2transformersAug เป็นอัลกอริทึมที่ดีที่สุดในการเพิ่มข้อความหรือแต่งเติมข้อความ

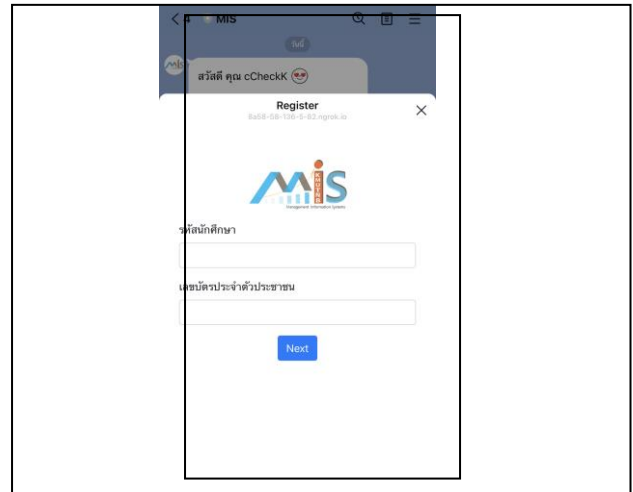
**4.2 ผลการพัฒนาหุ่นยนต์สนทนาบนระบบไลน์**

ในการพัฒนาหุ่นยนต์สนทนาบนระบบไลน์ ผู้วิจัยเลือกใช้ภาษา Python ในการพัฒนา Webhook เพื่อให้สอดคล้องกับแบบจำลองที่พัฒนาด้วยภาษา Python เช่นเดียวกัน ซึ่งระบบไลน์นั้นมีไลบรารีที่รองรับสำหรับนักพัฒนาที่ต้องการพัฒนาด้วยภาษา Python ซึ่งทำให้ง่ายต่อการพัฒนาระบบ



ภาพที่ 4 : แสดง MIS Line Official Account ก่อนลงทะเบียนและหลังลงทะเบียน

จากภาพที่ 4 แสดงหน้าจอหลังเพิ่มเพื่อนกับ MIS Line Official Account โดยจะมีข้อความต้อนรับและแสดงเมนูก่อนลงทะเบียนและหลังลงทะเบียน รูปทางด้านซ้ายแสดงเมนูก่อนลงทะเบียนประกอบไปด้วยเมนูลงทะเบียน ข่าวสารและติดต่อเรา รูปทางด้านขวาแสดงเมนูหลังลงทะเบียนประกอบไปด้วยเมนูผลการเรียน ตารางเรียน ตารางสอบ ขอกเอกสาร ข่าวสารและติดต่อเรา



ภาพที่ 5 : แสดงหน้าจอลงทะเบียนสำหรับนักศึกษา

ภาพที่ 5 เมื่อกดที่เมนูลงทะเบียนจะแสดงหน้าจอลงทะเบียนสำหรับนักศึกษา โดยระบุรหัสนักศึกษาและเลขบัตรประจำตัวประชาชน เพื่อเป็นการยืนยันตัวตนว่าไลน์นี้เป็นนักศึกษาท่านใดของภาควิชาฯ เมื่อทำการลงทะเบียนสำเร็จนักศึกษาจะสามารถใช้งานเมนูต่าง ๆ ของไลน์ได้

ผลการทำงานของหุ่นยนต์สนทนาบนระบบไลน์ของภาควิชาฯ นั้น ระบบสามารถทำงานได้ถูกต้องตามฟังก์ชันการใช้งาน แต่หากผู้ใช้มีการพิมพ์ข้อความที่ไม่ชัดเจนหรือพิมพ์บางประโยคไม่ถูกต้อง ระบบไม่สามารถประมวลคำตอบได้จึงอาจต้องมีการฝึกสอนระบบด้วยข้อมูลเพิ่มเติม

## 5. สรุปอภิปรายผล

บทความวิจัยนี้เป็นการประยุกต์ใช้เทคโนโลยีปัญญาประดิษฐ์ในการพัฒนาหุ่นยนต์สนทนาหรือแชทบอท เพื่อช่วยให้นักศึกษาสามารถเข้าถึงสารสนเทศต่างๆ ของภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ ได้สะดวกรวดเร็วยิ่งขึ้น โดยกระบวนการในการพัฒนาระบบผู้วิจัยได้ทำการศึกษาค้นคว้าอัลกอริทึมจากไลบรารี PyThaiNLP เพื่อศึกษาเทคนิคทางด้านการตัดคำ (Segmentation) การสะกดคำ (Spelling) และการเพิ่มข้อความหรือแต่งเติมข้อความ (Augmentation) โดยนำอัลกอริทึมแต่ละด้านมาเปรียบเทียบประสิทธิภาพในการประมวลผลความถูกต้อง



แม่นยำของข้อมูล และนำมาสร้างแบบจำลองโมเดลในการพัฒนาระบบหุ่นยนต์สนทนาหรือแชทบอทต่อไป

จากผลการวิจัยพบว่าอัลกอริทึมทางด้านการตัดคำ (Segmentation) ที่มีความถูกต้องแม่นยำและมีประสิทธิภาพในการประมวลผลมากที่สุดคืออัลกอริทึม ATTACUT โดยมีค่าเฉลี่ยความถูกต้องแม่นยำที่ 96.34% และใช้เวลาในการประมวลผลเฉลี่ยที่ 1.244 วินาที อัลกอริทึมทางด้านการสะกดคำ (Spelling) ที่มีความถูกต้องแม่นยำและมีประสิทธิภาพในการประมวลผลมากที่สุดคืออัลกอริทึม correct(engine= phunspell) โดยมีค่าเฉลี่ยความถูกต้องแม่นยำที่ 42.86% และใช้เวลาในการประมวลผลเฉลี่ยที่ 2.097 วินาที อัลกอริทึมการเพิ่มความหรือแต่งเติมข้อความ (Augmentation) ที่มีประสิทธิภาพในการประมวลผลมากที่สุดคืออัลกอริทึม Thai2transformersAug โดยใช้เวลาในการประมวลผลเฉลี่ยที่ 2.974 วินาที

ดังนั้นผู้วิจัยจึงเลือกใช้ใช้อัลกอริทึม ATTACUT ในการตัดคำ เลือกใช้อัลกอริทึม correct(engine=phunspell) ในการสะกดคำ และเลือกใช้อัลกอริทึม Thai2transformersAug ในการเพิ่มความหรือแต่งเติมข้อความ และนำแต่ละอัลกอริทึมมาพัฒนาแบบจำลองโมเดลในการพัฒนาระบบหุ่นยนต์สนทนาหรือแชทบอทของภาควิชาฯ

### เอกสารอ้างอิง

- [1] ขวัญชนก พุทธจันทร์, “ปัญญาประดิษฐ์ (AI : Artificial Intelligence)”.2563 [อินเทอร์เน็ต][สืบค้นเมื่อ 10 สิงหาคม 2564]. แหล่งข้อมูล : <https://www.lib.ku.ac.th/2019/index.php/covid-19/1045-artificialintelligence>
- [2] THE STANDARD TEAM, “ทำความเข้าใจทุกความรู้สึกที่เราบอกดียิ่งกว่าแฟน”.2563 [อินเทอร์เน็ต][สืบค้นเมื่อ 10 สิงหาคม 2564]. แหล่งข้อมูล : <https://thestandard.co/kbank-thai-nlp/>
- [3] ณัฐพงศ์ ทรัพย์ทวีพงศ์ “การใช้งาน Library NLTK เบื้องต้น”.2564 [อินเทอร์เน็ต][สืบค้นเมื่อ 10 สิงหาคม 2564]. แหล่งข้อมูล : t.ly/y268
- [4] ลีรสิทธิ์ บุญกลาง, “มาทำความรู้จักกับ PyThaiNLP”.2563 [อินเทอร์เน็ต][สืบค้นเมื่อ 10 สิงหาคม 2564]. แหล่งข้อมูล : t.ly/h99z
- [5] Pattarawat Chormai (2562), “AttaCut: A Fast and Accurate Neural Thai Word Segmenter” Max Planck School of Cognition Leipzig, Germany, 13 หน้า
- [6] แทนพงศ์ วิชาธาดา. (2561) “การพัฒนาหุ่นยนต์สนทนาบนระบบส่งข้อความของแอปพลิเคชันไลน์เพื่อสนับสนุนการทำงานของแผนกทรัพยากรบุคคล.” กรุงเทพมหานคร: คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ; 48 หน้า

# ระบบจัดการข้อมูลส่วนบุคคลบนคลาวด์ตามกฎหมายคุ้มครองข้อมูลส่วนบุคคล Data Privacy Management System on Cloud According to Personal Data Protection Act

กันตภณ เศษ โชติ (Kantapon Seshajoti)<sup>1</sup> และเทพฤทธิ์ บัณฑิตวัฒนาวงศ์ (Thepparit Banditwattanawong)<sup>2</sup>

<sup>1</sup>ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์

มหาวิทยาลัยเกษตรศาสตร์ ถ.พหลโยธิน จตุจักร กรุงเทพฯ 10900

<sup>1</sup>kantapon.se@ku.th, <sup>2</sup>fsctrb@ku.ac.th

## บทคัดย่อ

พ.ร.บ. คุ้มครองข้อมูลส่วนบุคคล มีผลบังคับให้ผู้ควบคุมข้อมูลส่วนบุคคลจำเป็นต้องขอความยินยอมการใช้ข้อมูลจากเจ้าของข้อมูลส่วนบุคคลเพื่อให้สามารถใช้งานได้ตามวัตถุประสงค์ ในอีกด้านหนึ่งภาคธุรกิจได้เปลี่ยนมาใช้บริการคลาวด์แทนที่ระบบแบบดั้งเดิมเพื่อลดค่าใช้จ่ายในการดูแลรักษาระบบและใช้ประโยชน์จากตัวแบบธุรกิจจ่ายตามปริมาณการใช้งานจริง งานวิจัยนี้จึงเสนอระบบจัดเก็บและให้สิทธิ์ข้อมูลส่วนบุคคลที่อาศัยบริการคลาวด์แอมะซอนเว็บเซอร์วิสโดยเฉพาะบริการป้องกันการโจมตีเว็บไซต์ การพิสูจน์ตัวตน และการเข้ารหัสลับข้อมูล จากผลการประเมินโดยผู้ใช้งานจำนวน 21 คนพบว่าระบบสามารถช่วยสร้างความมั่นใจด้านความมั่นคงปลอดภัยให้แก่เจ้าของข้อมูล ผู้ประมวลผลข้อมูล และผู้ควบคุมข้อมูลส่วนบุคคล โดยมีความพึงพอใจในภาพรวมเท่ากับ 4.35

**คำสำคัญ:** พ.ร.บ. คุ้มครองข้อมูลส่วนบุคคล โปรแกรมประยุกต์บนเว็บ ความเป็นส่วนตัวข้อมูล การคำนวณแบบคลาวด์

## Abstract

Personal Data Protection Act As is now so effective that data controller must request the consents of personal data usage from data owners to meet usage purposes. On the other hand, business sectors have been adopting cloud services to replace traditional systems to reduce maintenance costs and to take benefit of pay-per-use business model. This paper proposes a system for storing personal data

along with its usage authorization based on Amazon Web Services, especially web application firewall, user authentication and data encryption services. Evaluation results by 21 users showed that the proposed system could have data owners, data processor, and data controller confident in the system's security with overall satisfaction score 4.35.

**Keywords:** PDPA, web application, data privacy, cloud computing.

## 1. บทนำ

ในปัจจุบันได้มีกฎหมายคุ้มครองข้อมูลส่วนบุคคล (Personal Data Protection Act) มีผลบังคับใช้ เพื่อให้หน่วยงานมีมาตรฐานในการจัดการข้อมูลส่วนบุคคลอย่างมั่นคงปลอดภัย ทั้งการเข้ารหัสซึ่งความลับ การรักษาบูรณภาพข้อมูล และการรักษาภาวะพร้อมใช้งานของข้อมูลส่วนบุคคล [1] นอกจากนี้จำเป็นจะต้องมีการขอสิทธิ์การใช้ข้อมูลส่วนบุคคล เช่น ชื่อ-นามสกุล ที่อยู่ อีเมล เป็นต้น นอกจากนี้ยังมีข้อมูลส่วนบุคคลละเอียดอ่อน ที่จำเป็นจะต้องมีการควบคุมอย่างเข้มงวด เช่น เชื้อชาติ ประวัติอาชญากรรม หรือข้อมูลอื่นใดที่กระทบต่อเจ้าของข้อมูลส่วนบุคคล [2] ซึ่งการจัดเก็บข้อมูลที่ยืนยันตัวตนผ่านคลาวด์นั้น จำเป็นจะต้องมีการเข้ารหัสลับขั้นสูง เมื่อข้อมูลสำคัญเกิดรั่วไหล จะทำให้ไม่สามารถระบุตัวบุคคลได้ ดังนั้นควรเลือกระบบคลาวด์ที่มีความน่าเชื่อถือและมีมาตรฐานเพื่อปกป้องข้อมูลส่วนบุคคล ปัญหาด้านความปลอดภัยของข้อมูลและการปกป้องข้อมูลส่วนบุคคล มีอยู่ในวงจรชีวิตของข้อมูลตั้งแต่การสร้าง การถ่ายโอน การใช้ การแบ่งปัน การจัดเก็บแบบถาวร จนกระทั่งถูกทำลาย [3]

ซึ่งระบบจัดการข้อมูลส่วนบุคคลควรเลือกเป็นประเภทคลาวด์แบบส่วนตัวเพื่อใช้งานระบบที่ต้องการความมั่นคงและความปลอดภัยของข้อมูลที่สูง [4]

บทความนี้จึงเสนอระบบจัดการข้อมูลส่วนบุคคลบนคลาวด์โดยทำการเข้ารหัสลับและจัดเก็บในฐานข้อมูลเชิงสัมพันธ์บนเครือข่ายเสมือนส่วนบุคคลผ่านคลาวด์ (Virtual Private Network)

## 2. เทคโนโลยีและงานวิจัยที่เกี่ยวข้อง

### 2.1 เทคโนโลยีที่เกี่ยวข้อง

บริการแพลตฟอร์มคลาวด์ (Platform as a Service) คือ บริการเพื่อการติดตั้งใช้งาน โปรแกรมประยุกต์ ผู้บริโภคสร้างขึ้นเองหรือได้มาลงบนโครงสร้างพื้นฐานคลาวด์ โดยโปรแกรมประยุกต์สร้างโดยใช้ภาษาหรือเครื่องมือที่ผู้ให้บริการสนับสนุน โดยที่ผู้บริโภคไม่ต้องจัดหาหรือควบคุมโครงสร้างพื้นฐานคลาวด์ แต่ควบคุมโปรแกรมประยุกต์ได้ [5]

แอมะซอนอาร์ดีเอส (Amazon RDS) คือ บริการฐานข้อมูลเชิงสัมพันธ์ [6] ที่ผู้ใช้งานสามารถจัดการฐานข้อมูลได้โดยตรงผ่านเครื่องบริการฐานข้อมูลที่มีความยืดหยุ่น รองรับการระงับการประมวลผลและขนาดข้อมูลที่เพิ่มขึ้นหรือลดลงได้ และยังสามารถสำรองข้อมูลได้อัตโนมัติ

บริการอีซีทู (Amazon EC2) คือ บริการเครื่องเสมือน [7] ที่สามารถเลือกขนาดของ ซีพียู, หน่วยความจำหลัก และความจุของหน่วยเก็บข้อมูล ได้ตามต้องการ โดยมีการแบ่งทรัพยากรบนเครื่องบริการออกเป็นย่อยๆ และค่าใช้จ่ายเสียตามปริมาณทรัพยากรที่ใช้งานจริงเท่านั้น

เอพีไอเกตเวย์ (API Gateway) คือ บริการที่สามารถทำให้เผยแพร่ระบบเพื่อสำหรับการเรียกใช้งาน บำรุงรักษา เฝ้าติดตาม และรักษาความปลอดภัยของเอพีไอ [8] จัดการปริมาณการใช้งาน การอนุมัติจัดการการร้องขอ และการควบคุมการเข้าถึงก่อนส่งต่อไปยังเครื่องบริการอื่น

คลาวด์ฟรอนท์ (Cloudfront) [9] คือ บริการเครือข่ายการจัดส่งเนื้อหา (Content Delivery Network) ที่สามารถทำการรับส่งข้อมูลทั่วโลกอย่างมั่นคงปลอดภัยด้วยเวลาแฝง (Latency) ที่ต่ำ โดยส่งการร้องขอของผู้ใช้งานไปยัง

สถานที่ (Edge Location) และนำข้อมูลที่ถูกแคช (Cache) ไว้ในสถานที่นั้นส่งกลับไปยังผู้ใช้งานแทนการร้องขอข้อมูลจากระบบต้นทางที่มีอยู่

คอกนิโต้ (Cognito) คือ บริการสนับสนุนมาตรฐานการจัดการข้อมูลประจำตัวและการเข้าถึง [10] ที่ปรับขนาดได้สำหรับรองรับผู้ใช้ได้จำนวนมาก รองรับการพิสูจน์ตัวตนแบบหลายปัจจัย การเข้ารหัสลับข้อมูลทั้งระหว่างพักเก็บและระหว่างส่งผ่าน ควบคุมการเข้าถึงทรัพยากร โดยกำหนดบทบาทต่างๆ จับคู่ผู้ใช้งานและกำหนดการเข้าถึงเฉพาะผู้ใช้แต่ละรายได้

### 2.2 งานวิจัยที่เกี่ยวข้อง

งานวิจัย [11] เสนอเครื่องมือช่วยเหลือในการค้นหาข้อมูลส่วนบุคคลที่ระบุตัวตนได้ (Personally Identifiable Information) ค้นหาได้ทั้งเครื่องให้บริการส่วนบุคคลและที่จัดเก็บบนคลาวด์ เพื่อให้องค์กรค้นหาและจัดการข้อมูลส่วนบุคคลได้อัตโนมัติ โดยมีการเพิ่มนโยบายความเป็นส่วนตัวและปฏิบัติตามข้อกำหนดกฎหมายคุ้มครองข้อมูลส่วนบุคคล

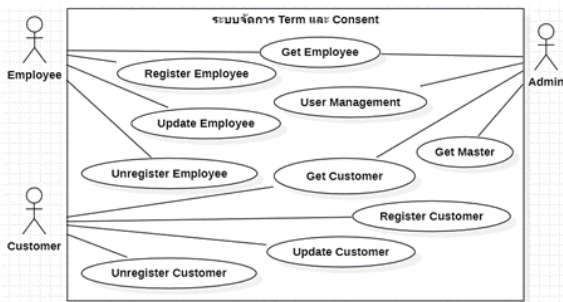
งานวิจัยการเก็บข้อมูลบนคลาวด์และรักษาความเป็นส่วนตัว [12] เสนอให้มีการประมวลผลแบบคลาวด์ โดยนำเทคนิคการเข้ารหัสลับแบบโฮโมมอร์ฟิก (Homomorphic) ไปใช้รักษาความมั่นคงปลอดภัยของข้อมูลผู้ใช้งานที่ทำการใช้ร่วมกันข้อมูลเพื่อการวิเคราะห์กฎเกณฑ์ข้อมูลโดยไม่ต้องแปลงข้อมูลที่ถูกรหัสลับกลับเป็นข้อมูลปกติ กลไกนี้เพิ่มการรบกวนไปในข้อมูล เพื่อช่วยสร้างความเป็นส่วนตัวแก่ข้อมูลส่วนบุคคลในการวิเคราะห์เชิงสถิติ

งานวิจัยตัวแบบความมั่นคงปลอดภัยสำหรับการปรับปรุงความเป็นส่วนตัวในคลาวด์ [13] เสนอการจัดเก็บข้อมูลบนคลาวด์ที่มีขั้นตอนวิธีในการเข้ารหัสลับหลายระดับ โดยใช้ขั้นตอนวิธีอาร์เอสเอ (RSA) ในการเข้ารหัสลับระดับที่ 1 สำหรับข้อความและแฟ้มข้อมูลที่เกี่ยวข้อง ขั้นตอนถัดไปใช้ขั้นตอนวิธี เออีเอส (AES) เพื่อเข้ารหัสลับระดับที่ 2 ก่อนจะทำการจัดเก็บข้อความเข้ารหัสลับลงฐานข้อมูล

งานวิจัยจัดเก็บข้อมูลส่วนตัวบนคลาวด์ให้มีความปลอดภัยโดยการเข้ารหัสลับหลายกุญแจ [14] เสนอขั้นตอนการแบ่งปันข้อมูลและการจัดเก็บข้อมูล ในระบบแบบกระจาย มีการนำขั้นตอนวิธีดีพีเอเค-อาร์อี (DPAK-RE) มาให้ความปลอดภัยในการแบ่งปันข้อมูล โดยผู้ใช้สามารถรวมกลุ่มกุญแจลับได้หลายกลุ่ม กุญแจรวมถูกเข้ารหัสลับใหม่ โดยสร้างเป็นกุญแจส่วนบุคคลเป็นแบบคงที่ ที่ให้ความเป็นส่วนตัวในการแบ่งปันและการจัดเก็บ ซึ่งกุญแจลับที่แยกจากกันจะถูกสร้างแบบพลวัตสำหรับไฟล์ๆ หนึ่งและสร้างกุญแจส่วนตัวใหม่โดยการเข้ารหัสลับกุญแจรวมจะถูกเก็บไว้กับเจ้าของข้อมูล

งานวิจัยเสนอค้นหาคำสำคัญที่รักษาความเป็นส่วนตัวข้อมูลผ่านคลาวด์ที่เข้ารหัสลับด้วยการวิเคราะห์ที่กว้าง [15] เสนอให้อีเมลได้มีการจัดเก็บ ประมวลผล และวิเคราะห์ข้อมูลโดยการคำนวณแบบคลาวด์วิธีการปกป้องข้อมูล เป็นการเข้ารหัสลับข้อมูลก่อนทำการส่งไปยังเครื่องบริการระยะไกล มีการเข้ารหัสลับแบบค้นหาได้ โดยผู้ใช้งานมีการสร้างโทเคน (Token) สำหรับการค้นหา เพื่อให้สามารถเรียกเนื้อหาที่มีการเข้ารหัสลับกลับมาใช้งานได้

3. วิธีการดำเนินงาน



ภาพที่ 1: แผนภาพยูสเคสระบบจัดการข้อมูลส่วนบุคคล

ภาพที่ 1 แผนภาพยูสเคสสำหรับการจัดการหน้าที่และสิทธิ์การเข้าถึงของระบบ ประกอบด้วย 3 ผู้เกี่ยวข้องและ 10 ยูสเคส โดยพนักงาน (Employee) จะมีสิทธิ์เรียกใช้งานฟังก์ชันการทำงานระบบประกอบด้วย เรียกดูข้อมูลพนักงาน (Get Employee) ลงทะเบียนพนักงาน (Register Employee) ปรับปรุงข้อมูลพนักงาน (Update Employee) และยกเลิกลงทะเบียนพนักงาน (Unregister Employee)

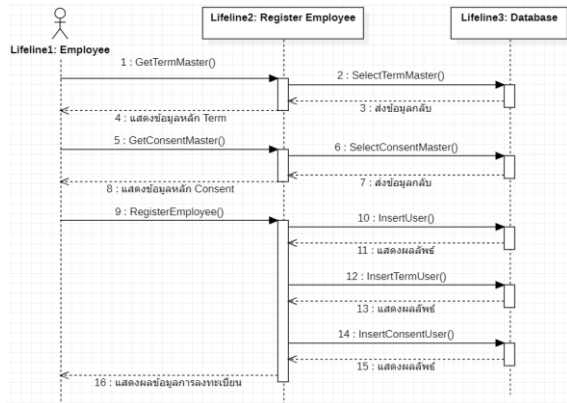
ในส่วนของลูกค้า (Customer) มีสิทธิ์เรียกใช้งานฟังก์ชันการทำงานระบบประกอบด้วย เรียกดูข้อมูลลูกค้า (Get Customer) ลงทะเบียนลูกค้า (Register Customer) ปรับปรุงข้อมูลลูกค้า (Update Customer) และยกเลิกลงทะเบียนลูกค้า (Unregister Customer) และในส่วนผู้ดูแลระบบ (Admin) มีสิทธิ์เรียกใช้งานฟังก์ชันระบบประกอบด้วย เรียกดูข้อมูลพนักงาน เรียกดูข้อมูลลูกค้า เรียกดูข้อมูลหลักที่เกี่ยวกับข้อกำหนด (Term) และการให้ความยินยอม (Consent) และการจัดการข้อมูลผู้ใช้งาน (User Management) โดยฟังก์ชันนี้จะทำงานได้ทั้ง 3 แบบประกอบด้วย ลงทะเบียน ปรับปรุง และยกเลิกลงทะเบียน

<b>Register Employee</b> -SystemId: String -UserId: String -Thaid: String -FirstName: String -LastName: String -MobileNo: String -Term: List -Consent: List +GetTermMaster() +GetConsentMaster() +RegisterEmployee()	<b>User Management</b> -SystemId: String -UserId: String -Thaid: String -FirstName: String1 -LastName: String -MobileNo: String -UserType: Integer -Term: List -Consent: List +SearchUser() +AddUser() +UpdateUser() +DeleteUser()	<b>Register Customer</b> -SystemId: String +UserId: String +Thaid: String +FirstName: String +LastName: String +MobileNo: String +Term: List +Consent: List +GetTermMaster() +GetConsentMaster() +RegisterCustomer()	<b>Database</b> +SelectTermMaster() +SelectConsentMaster() +SelectTermUser() +InsertTermUser() +UpdateTermUser() +DeleteTermUser() +SelectConsentUser() +InsertConsentUser() +UpdateConsentUser() +DeleteConsentUser() +SelectUser() +InsertUser() +UpdateUser() +DeleteUser()
<b>Update Employee</b> -SystemId: String -UserId: String -Thaid: String -FirstName: String -LastName: String -MobileNo: String -Term: List -Consent: List +GetTermUser() +GetConsentUser() +GetProfileEmployee() +UpdateEmployee()	<b>Get Master</b> -SystemId: String +GetTermMaster() +GetConsentMaster()	<b>Update Customer</b> -SystemId: String +UserId: String +Thaid: String +FirstName: String +LastName: String +MobileNo: String +Term: List +Consent: List +GetTermUser() +GetConsentUser() +GetProfileCustomer() +UpdateCustomer()	
<b>Unregister Employee</b> -SystemId: String -UserId: String -Thaid: String +GetProfileEmployee() +UnregisterEmployee()	<b>Get Employee</b> -SystemId: String -Thaid: String +GetTermEmployee() +GetConsentEmployee() +GetDetailEmployee()	<b>Unregister Customer</b> -SystemId: String -UserId: String -Thaid: String +GetProfileCustomer() +UnregisterCustomer()	
<b>Register Customer</b> -SystemId: String -UserId: String -Thaid: String +GetProfileEmployee() +UnregisterEmployee()	<b>Get Customer</b> -SystemId: String -Thaid: String +GetTermCustomer() +GetConsentCustomer() +GetDetailCustomer()	<b>Unregister Customer</b> -SystemId: String -UserId: String -Thaid: String +GetProfileCustomer() +UnregisterCustomer()	

ภาพที่ 2: แผนภาพคลาสระบบจัดการข้อมูลส่วนบุคคล

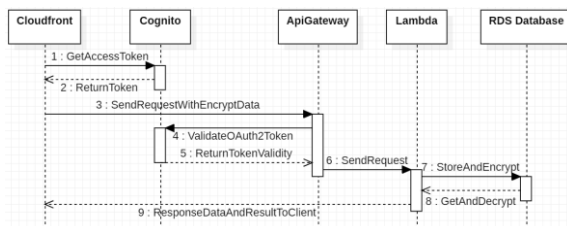
ภาพที่ 2 ระบบที่เสนอประกอบไปด้วย 10 คลาสสำหรับการลงทะเบียนข้อมูลผู้ใช้งาน การยอมรับข้อกำหนดและเงื่อนไข การยินยอมการให้สิทธิ์ข้อตกลงในเรื่องต่างๆ มีการทำงานแต่ละฟังก์ชันจะประกอบไปด้วย ดังนี้ คลาสการลงทะเบียนพนักงานและลูกค้า จะเรียกดูข้อมูลหลักข้อกำหนดและการให้ความยินยอมฉบับล่าสุด และถึงจะทำการลงทะเบียนได้ คลาสการปรับปรุงข้อมูลการให้ความยินยอมพนักงานและลูกค้า มีฟังก์ชันการทำงานประกอบด้วย เรียกดูข้อมูลข้อกำหนดและการให้ความยินยอมและข้อมูลแสดงลักษณะเฉพาะของผู้ใช้งาน ก่อนที่จะปรับปรุงการให้ความยินยอมผู้ใช้งาน คลาสการยกเลิกลงทะเบียนพนักงานและลูกค้าโดยจะมีฟังก์ชันการทำงาน เรียกดูข้อมูลแสดงลักษณะเฉพาะของผู้ใช้งาน โดย

ให้ตรวจสอบข้อมูลก่อนทำการยกเลิกการลงทะเบียน ผู้ใช้งาน คลาสการเรียกดูข้อมูลหลักจะทำการเรียกดูข้อมูล ข้อกำหนดและการให้ความยินยอมฉบับล่าสุด และคลาสการจัดการผู้ใช้งาน โดยจะมีฟังก์ชันการทำงานที่ประกอบไปด้วย การเรียกดูข้อมูลผู้ใช้งาน การเพิ่มผู้ใช้งาน การปรับปรุงผู้ใช้งาน และการลบผู้ใช้งาน และคลาสฐานข้อมูลที่ประกอบด้วยฟังก์ชันการทำงานด้านต่างๆที่เกี่ยวข้องกับการติดต่อฐานข้อมูล



ภาพที่ 3: แผนภาพลำดับการลงทะเบียนผู้ใช้งาน

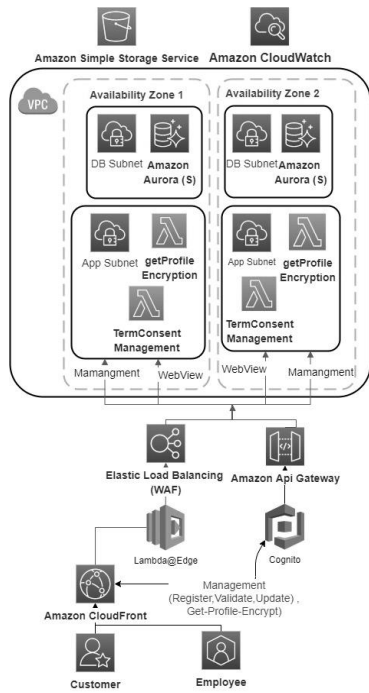
ภาพที่ 3 อธิบายขั้นตอนการทำงานของระบบสำหรับการลงทะเบียนผู้ใช้งาน โดยมีการเรียกดูข้อมูล ข้อกำหนด และข้อตกลงฉบับล่าสุดก่อนให้ผู้ใช้งานลงทะเบียนข้อมูลส่วนบุคคลที่มีการให้ความยินยอมตามเงื่อนไขและข้อตกลง และสุดท้ายจะทำการแสดงผลการลงทะเบียน



ภาพที่ 4 : แผนภาพลำดับขั้นตอนการทำงานของระบบ

ภาพที่ 4 อธิบายขั้นตอนการทำงานของระบบระหว่างส่วนติดต่อผู้ใช้งานและส่วนจัดการระบบ โดยในส่วนที่ 1 ทำการร้องขอโทเค็นจากคอกนิต โดยทำการสร้างโทเค็นใหม่แล้วส่งกลับไปยังส่วนที่ 2 ส่วนที่ 3 ส่งข้อมูลไปยังเฟื่อเกตเวย์ ตรวจสอบโทเค็นในส่วนที่ 4 ส่งผลกลับในส่วนที่ 5 และส่งข้อมูลไปประมวลผลในส่วนที่ 6 ทำการบันทึกเข้ารหัสลับข้อมูลในส่วนที่ 7 เรียกข้อมูลพร้อมถอดรหัสลับในส่วนที่ 8 และส่งผลกลับไปยังส่วนที่ 9

การเข้าใช้งานเว็บไซต์จากคลาวด์ฟรอนท์ผ่านแลมบ์ดา แอทเอดจ์ (Lambda@Edge) ที่ทำการเพิ่มความปลอดภัยส่วนหัว [16] ก่อนส่งกลับไปยังผู้ใช้งานและผ่านการคัดกรองในส่วนบริการป้องกันการโจมตีเว็บไซต์ (Web Application Firewall) ในส่วนถัดไปผ่านการควบคุมเครือข่ายตามกฎที่ระบุ เมื่อติดต่อในส่วนจัดการเว็บไซต์ จะต้องมีการพิสูจน์ตัวตนโดยใช้คอกนิตได้ โดยการสร้างกุญแจแบบพลวัตก่อนการเข้าถึง และทำการติดต่อไปยังส่วนจัดการเว็บไซต์ผ่านเฟื่อเกตเวย์ มีการควบคุมการเข้าถึงก่อนส่งข้อมูลไปยังส่วนถัดไป ข้อมูลที่สำคัญจะเข้ารหัสลับด้วยมาตรฐานขั้นสูงก่อนทำการติดต่อในส่วนจัดการเว็บไซต์เมื่อติดต่อสำเร็จในส่วนการทำงานเบื้องหลังจะทำการถอดรหัสลับข้อมูลที่รับเข้ามาแล้วนำไปทำกระบวนการในส่วนถัดไป โดยมีการจัดการการทำงานในส่วนการเรียกดูเพิ่ม แก้ไข ลบ ข้อมูลจากเจ้าของข้อมูลส่วนบุคคลรวมถึงข้อมูลการให้สิทธิ์ ข้อกำหนดเงื่อนไขและข้อตกลง และส่วนสุดท้ายการติดต่อไปยังฐานข้อมูลอาร์ดีเอส ที่วางอยู่ในเครือข่ายเสมือนแบบส่วนตัวการเข้าถึงฐานข้อมูลจำเป็นต้องเข้าผ่านบริการอีซีทูที่อยู่ในวงเครือข่ายเสมือนวงเดียวกันเท่านั้น แต่การออกแบบครั้งนี้ตีปัญหาการนำแอมะซอนอาร์ดีเอสและแลมบ์ดามาทำการเชื่อมต่อให้อยู่ในวงเครือข่ายเสมือนเดียวกันจะไม่สามารถเชื่อมต่อฐานข้อมูลไดนาโมดีบี (DynamoDB) ได้เนื่องจากตีปัญหาการตั้งค่าบางอย่างที่ไม่สมบูรณ์ จึงเปลี่ยนการออกแบบมาเป็นการใช้ฐานข้อมูลแอมะซอนอาร์ดีเอสในการจัดเก็บแทน ดังภาพที่ 5

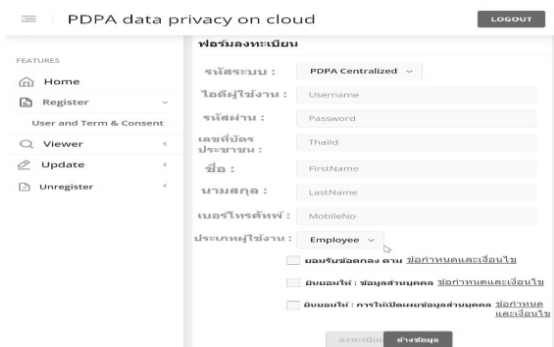


ภาพที่ 5: แผนภาพสถาปัตยกรรมระบบคลาวด์

#### 4. ผลการดำเนินงาน

##### 4.1 โปรแกรมประยุกต์เว็บ

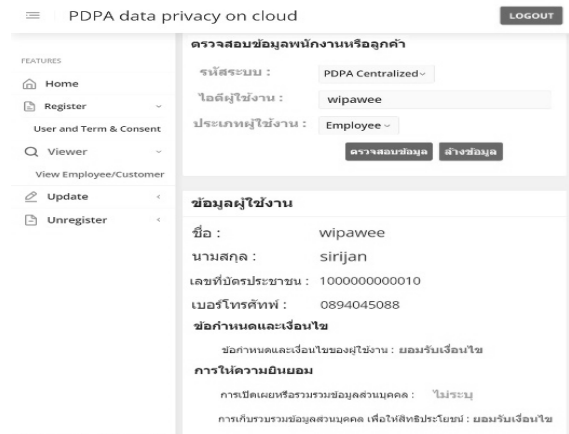
เว็บไซต์จัดการข้อมูลพนักงานและลูกค้าโดยผู้ดูแลระบบสามารถลงทะเบียนแก้ไข เรียกดูข้อมูลให้กับเจ้าของข้อมูลได้หรือเจ้าของข้อมูลส่วนบุคคลสามารถลงทะเบียนด้วยตนเองได้ ในส่วนการลงทะเบียนให้ทำการกรอกรายละเอียดให้ครบถ้วน เลือกประเภทระบบและประเภทผู้ใช้งาน โดยสามารถอ่านข้อกำหนดและเงื่อนไขก่อนทำการยอมรับข้อตกลง โดยในส่วนข้อตกลงของผู้ใช้งาน จะให้ความยินยอมหรือไม่ก็ได้ แสดงดังภาพที่ 1



ภาพที่ 1 : เมนูการลงทะเบียน

เมนูการตรวจสอบข้อมูลส่วนบุคคลและแสดงสิทธิ์ต่างๆ ของผู้ใช้งานตามระบบและประเภทที่ลงทะเบียน ข้อมูลส่วนบุคคล เช่น ชื่อ-นามสกุลและรหัสบัตร

ประชาชน ถ้าเจ้าของข้อมูลส่วนบุคคลไม่ยินยอมให้เปิดเผยข้อมูลส่วนบุคคล ระบบจะไม่ทำการแสดงผลข้อมูลนั้นๆ ถ้าเป็นการเข้าสู่ระบบด้วยระดับผู้ใช้งานทั่วไปจะสามารถเรียกดูข้อมูลได้เฉพาะข้อมูลในส่วนของเจ้าของเท่านั้น เมนูการแก้ไขนั้นสามารถปรับเปลี่ยนความยินยอมการให้สิทธิ์ทางด้านต่างๆ ได้ตลอดเวลา แสดงดังภาพที่ 2



ภาพที่ 2 : เมนูการตรวจสอบข้อมูลและแก้ไขข้อมูล

เมนูการยกเลิกความเป็นสมาชิกที่สามารถยกเลิกได้ตลอดเวลาเมื่อเจ้าของข้อมูลทำการร้องขอ โดยมีการตรวจสอบข้อมูลก่อนทำการยกเลิก การยกเลิกความเป็นสมาชิกนั้นระบบจะทำการลบข้อมูลส่วนบุคคลออกจากระบบทั้งหมด ทั้งข้อมูลการให้ความยินยอมในด้านต่างๆ และจะเก็บเป็นประวัติทำการ โดยที่ไม่สามารถระบุตัวตนของบุคคลนั้นได้ ในส่วนการทำงานของระบบจะมีการทำงานเพิ่มเติมเกี่ยวกับการตรวจสอบช่วงเวลาคงอายุ เมื่ออยู่ในระบบครบตามเวลาที่กำหนดหรือไม่มีกิจกรรมใดๆ เกิดเวลาที่กำหนด ระบบจะทำการออกจากระบบให้โดยอัตโนมัติ สิทธิ์ในการเข้าถึงนั้นถ้าผู้ใช้งานเป็นระดับผู้ดูแลระบบจะเข้าถึงได้ทุกรายการเลือก แต่ถ้าผู้ใช้งานเป็นระดับทั่วไปจะเข้าถึงได้บางรายการเลือกเท่านั้นและเมื่อทำการรายการเสร็จสิ้นจะทำการบันทึกล็อก (log) ลงฐานข้อมูล เมื่อครบกำหนด 90 วัน ระบบจะทำการลบล็อกออกจากระบบ

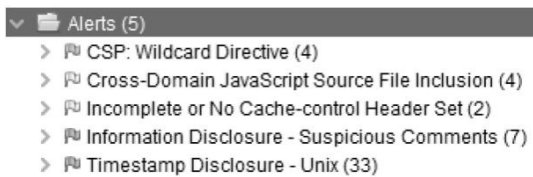
การเก็บข้อมูลของเจ้าของข้อมูลส่วนบุคคล จะทำการเข้ารหัสลับด้วยขั้นตอนวิธีเออีเอส ด้วยขนาดกุญแจ 256 บิต ความยาวตัวอักษรกุญแจ 128 บิต และมีกุญแจลับในการเข้ารหัสลับร่วมด้วย และเมื่อต้องการนำข้อมูลกลับมา

ใช้งานก็จะทำการถอดรหัสลับด้วยการตั้งค่าดังกล่าว แสดงดังภาพที่ 3

SYSTEM_ID	USER_ID	PASSWORD	THAI_ID	FIRST_NAME	LAST_NAME	MOBILE_NO
01021001	adminkan...	Ks990981...	9e835a3de...	02b8e6392e...	5e45183a7e...	32e83e84cb...
01021001	kantapon10	Bg987456*	473feaf19...	33a5658670...	2e45a1c155f...	f82a84ccea2...
01021001	pichayamin	*198745x	062f67498...	e952ce5d09...	ebc18008b3...	7ce6ebe4e4...
01021001	wipawee	G134*dfg	fd5fbd03...	1d8ac04b93...	db176bccf53...	28e84cfcbf2...
01021001	wpmint01	Dcovk*01	2a11de003...	06262ca39a...	2eed05e3af...	1c76c5b33c...
01021002	kantaponso	805Gs*๒๕	abf6e352fc...	1653e0c9ff...	7e1827a327...	c010e4d6c3...

ภาพที่ 3 : การเก็บข้อมูลส่วนบุคคลในฐานข้อมูลมีการเข้ารหัสลับ

ผู้ทำวิจัยได้ทำการทดสอบหาช่องโหว่ระบบด้วยเครื่องมือโอตบเบิลยูเอสพี แซบ (OWASP ZAP) [17] และได้ผลการทดสอบระบบ ประกอบด้วยผลการพบช่องโหว่ระดับปานกลาง ซึ่งเป็นการแจ้งเตือนเกี่ยวกับ นโยบายการรักษาความปลอดภัยของเนื้อหา (Content Security Policy : Wildcard Directive) เกี่ยวกับการกำหนดทรัพยากรที่สามารถดึงออกมาดำเนินการโดยหน้าเว็บไซต์ได้กว้างเกินไป จากการตรวจสอบพบว่าการตั้งค่าปฏิเสธการแสดงผลเฟรม (X-Frame-Options: deny) ซึ่งหมายความว่าถ้ามีเว็บไซต์ไหนนำเว็บไซต์นี้ไปแสดงผลไอเฟรม (iframe) จะไม่สามารถแสดงผลข้อมูลได้ ซึ่งไม่กระทบกับเว็บไซต์หลักเพราะไม่มีส่วนเกี่ยวข้องกับข้อมูลส่วนบุคคล และผลการพบช่องโหว่ระดับต่ำ เป็นส่วนข้อผิดพลาดที่ไม่เกี่ยวข้องกับเว็บไซต์หลักเนื่องจากเป็นผลมาจากคลังโปรแกรมของบุคคลที่สามและผลการพบช่องโหว่ที่เกี่ยวกับการควบคุมแคชพบว่าเว็บไซต์ในส่วนคลาวด์ฟรอนท์ ได้กำหนดไม่ให้เก็บแคชโดยจะทำการดาวน์โหลดหน้าเว็บไซต์ใหม่ทุกครั้งจึงไม่ได้รับผลกระทบดังกล่าว แสดงดังภาพที่ 4



ภาพที่ 4 : ผลทดสอบความปลอดภัยด้วยเครื่องมือ โอตบเบิลยูเอสพี แซบ

ผู้ทำวิจัยวัดผลความพึงพอใจของผู้ใช้งานระบบจำนวน 21 คน ประกอบไปด้วยพนักงานเอกชน 10 คน ข้าราชการ 5 คน ผู้ทำงานอิสระ 4 คน และผู้ประกอบการ 2 คน หรือแบ่งเป็นผู้ควบคุมข้อมูลจำนวน 2 คน ผู้ประมวลผลข้อมูลจำนวน 5 คน และเจ้าของข้อมูลจำนวน 14 คนมีรายละเอียดผลการประเมิน ดังตารางที่ 1 โดยมีคะแนนเต็ม

แต่ละข้อเท่ากับ 5 พบค่าความพึงพอใจเฉลี่ยคิดเป็น 4.35 จัดอยู่ในระดับที่ดี และส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับ 0.60

ตารางที่ 1: ผลการวัดความพึงพอใจของผู้ใช้งาน

เกณฑ์การประเมิน	$\bar{X}$	S.D.
1) ด้านสามารถทำงานได้ตามหน้าที่ (Function)	4.46	0.56
2) ด้านความถูกต้องและน่าเชื่อถือของระบบ (Accuracy)	4.31	0.63
3) ด้านความง่ายต่อการใช้งาน (Usability)	4.25	0.63
4) ด้านสมรรถนะ (Performance)	4.35	0.58
5) ด้านการรักษาความปลอดภัยของข้อมูล (Security)	4.37	0.59
6) ด้านความเข้าใจได้ของระบบ (Understandability)	4.32	0.61
ค่าเฉลี่ย	4.35	0.60

### 5. สรุป

งานวิจัยนี้มีวัตถุประสงค์เพื่อการศึกษาการออกแบบระบบจัดการข้อมูลส่วนบุคคลบนคลาวด์ที่มีการปกป้องข้อมูลก่อนส่งไปยังเครื่องบริการระยะไกล การจัดการสิทธิ์การให้ความยินยอมในการจัดเก็บข้อมูล การขอสิทธิ์ใช้งานด้านต่างๆตามวัตถุประสงค์ การแก้ไขข้อมูล จนกระทั่งการทำลายข้อมูล เมื่อเจ้าของข้อมูลไม่ต้องการให้จัดเก็บข้อมูลส่วนบุคคลนั้นอีกต่อไป ผลการประเมินระบบพบว่าเจ้าของข้อมูลส่วนบุคคลได้จัดเก็บข้อมูลบนคลาวด์ได้อย่างมีความน่าเชื่อถือและปลอดภัย และเนื่องจากขั้นตอนวิธีการเข้ารหัสลับขั้นสูงด้วยขนาดกุญแจ 256 บิต ในปัจจุบันยังคงมีความมั่นคงปลอดภัยที่สูงเพียงพอสำหรับการใช้งาน [18] ซึ่งยังคงไม่สามารถเจาะการเข้ารหัสลับได้ด้วยเทคโนโลยีในปัจจุบัน

แนวทางในการพัฒนาต่อไปโดยการเพิ่มการตรวจสอบการนำข้อมูลไปใช้ให้สอดคล้องตามวัตถุประสงค์ตามที่ได้ความยินยอมหรือไม่ ซึ่งจะมีการตรวจสอบการให้ความยินยอมก่อนการเปิดเผยข้อมูล ว่ามีวัตถุประสงค์ที่ต้องการนำข้อมูลไปใช้ต่อระบบใดบ้างตามประเภทการให้ความยินยอม มีการเพิ่มขึ้นขั้นตอนการยืนยันตัวตนแบบหลายปัจจัย (Multi-factor Authentication) อย่างการสร้างรหัสผ่านแบบใช้แล้วทิ้ง (One-time Password) เพิ่มในการเข้าสู่ระบบ และมีการพัฒนาให้รองรับการ โยคย้าย

(Migrate) ระบบแบบดั้งเดิมไปยังระบบคลาวด์ โดยมีขั้นตอนเบื้องต้นดังนี้ การกำหนดเป้าหมาย การประเมินสภาพแวดล้อม การเตรียมข้อมูลและโปรแกรมประยุกต์ การโอนย้าย การจัดเตรียมโครงสร้างพื้นฐานบนการประมวลผลแบบคลาวด์ และการทดสอบปรับสภาพแวดล้อมให้เหมาะสม เพื่อนำไปประยุกต์และปรับใช้สำหรับองค์กรเพื่อจัดเก็บข้อมูลส่วนบุคคลบนคลาวด์ได้อย่างมีความมั่นคงปลอดภัย

### เอกสารอ้างอิง

- [1] เภษญาภา ช่างประดิษฐ์. *ระบบประมวลผลแบบคลาวด์ (Cloud Computing) : ข้อเสนอแนะเพื่อการยกเว้นพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคลของไทย*. กรุงเทพฯ:มหาวิทยาลัยธรรมศาสตร์, 2561.
- [2] ราชกิจจานุเบกษา. *พระราชบัญญัติ คุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562*. เล่ม 136.ตอนที่ 69.(พฤษภาคม 2562). น.52.
- [3] สุนทรีย์ ส่งเสริม. *พร้อมรู้ PDPA*. กรุงเทพฯ:สำนักงานปลัดกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม, 2564.
- [4] Amazon Web Services. “การใช้ AWS ในบริบทของการรักษาความเป็นส่วนตัวและการคุ้มครองข้อมูลทั่วไป.” บริษัท Amazon Web Services, พฤษภาคม, 2561.
- [5] เทพฤทธิ์ บัณฑิตวัฒนาวงศ์. *พื้นฐานการคำนวณแบบคลาวด์ (Cloud Computing Fundamental)*. โรงพิมพ์ Protext บริษัท แคนเน็กซ์ อินเทอร์เน็ตเซอร์วิส จำกัด, กรุงเทพมหานคร, 2562.
- [6] Amazon Web Services. Amazon Relational Database Service (RDS). Available Online at <https://aws.amazon.com/th/rds/>, accessed on 12 December 2021.
- [7] Amazon Web Services. Amazon Elastic Compute Cloud (EC2). Available Online at <https://aws.amazon.com/th/ec2/>, accessed on 10 December 2021.
- [8] Amazon Web Services. Amazon API Gateway. Available Online at <https://aws.amazon.com/th/api-gateway/features/>, accessed on 9 December 2021.
- [9] Amazon Web Services. Amazon CloudFront. Available Online at <https://aws.amazon.com/th/cloudfront/>, accessed on 9 December 2021.
- [10] Amazon Web Services. Amazon Cognito. Available Online at <https://aws.amazon.com/th/cognito/>, accessed on 9 December 2021.
- [11] S. Fugkeaw, A. Chaturasrivilai, P. Tasungnoen and W. Techaudomthaworn. “AP2I: Adaptive PII Scanning and Consent Discovery System.” *Computer and Communication Technology (ICT)*, Thammasat University, 2564.
- [12] S. Komawar, M. Batwal, S. Shah, S. Shahani, and J. Abraham. “Privacy Preserving Data Aggregation on Secure Cloud.” *Computer Engineering & IT*, College of Engineering Pune, 2561.
- [13] Y. Sharma, H. Gupta, S. K. Khatri. “A Security Model for the Enhancement of Data Privacy in Cloud Computing,” *Institute of Information Technology*, Amity University Campus, 2562.
- [14] M. R. P, S. H. Manjula, and V. K. R. “Secured Privacy Data using Multi Key Encryption in Cloud Storage.” *Department of Computer Science and Engineering*, Bangalore University, 2561.
- [15] P. Sreekumari. “Privacy-Preserving Keyword Search Schemes over Encrypted Cloud Data An Extensive Analysis.” *Department of Computer Science*, Grambling State University, 2561.
- [16] Amazon Web Services. HTTP-Security-Headers. Available Online at <https://aws.amazon.com/th/blogs/networking-and-content-delivery/adding-http-security-headers-using-lambdaedge-and-amazon-cloudfront>, accessed on 2 June 2021.
- [17] ZAP Dev Team. Zed Attack Proxy Getting Started. Available Online at <https://www.zaproxy.org/getting-started/>, accessed on 25 December 2021.
- [18] A. Sousi, D. Yehya, M. Joudi. “AES Encryption: Study & Evaluation.” *CCEE552: Cryptography & Network Security*, Rafik Hariri University, 2563.



# การพัฒนาต้นแบบแชทบอทแอปพลิเคชันสำหรับแนะนำปัญหาด้านสุขภาพทางเพศหญิง

## The Development of Chatbot Prototype on an Application for Problems Recommender of Female's Wellness

อุไรวรรณ อินทร์แหยม (Uraiwan Inyaem)<sup>1</sup> วีณา จันทร์รัชชกุล(Weena Janratchakool)<sup>2</sup> ณัฐวดี สุขเกษม (Nattawadee Sukkasem)<sup>3</sup> และสุวิมล ทวนทอง (Suwimon Tuanthong)<sup>4</sup>

<sup>1,2,3,4</sup>ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

uraiwan.inyaem@rmutt.ac.th, weena\_j@rmutt.ac.th,  
1160109050766@mail.rmutt.ac.th, 1160109051202@mail.rmutt.ac.th

### บทคัดย่อ

ในประเทศไทย กลุ่มเยาวชนและวัยรุ่นสาวยังขาดการดูแลเรื่องสุขภาพของตนเอง ประสบการณ์การควบคุมตนเอง ความรู้และความเข้าใจเกี่ยวกับความเสี่ยงเรื่องสุขภาพที่เพิ่มมากขึ้นและปัญหาด้านสุขภาพต่างๆ ซึ่งอาจส่งผลกระทบต่อการใช้ชีวิต ในยุคดิจิทัลที่ทุกคนไม่ค่อยมีการพูดคุยสื่อสารกับผู้อื่นหรือการไม่มีเวลาพอที่จะไปพบแพทย์ที่โรงพยาบาลด้วยตนเอง การนำเทคโนโลยีสารสนเทศด้านแอปพลิเคชันจึงถูกนำมาประยุกต์ใช้ได้ อย่างง่ายดาย โดยงานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแอปพลิเคชันแชทบอทสำหรับให้คำแนะนำปัญหาสุขภาพทางเพศสำหรับเพศหญิง ในงานวิจัยได้ทำการเก็บรวบรวมข้อมูลเกี่ยวกับความผิดปกติในการมีประจำเดือน ความผิดปกติเกี่ยวกับมดลูกของผู้หญิง ปัญหาทางเพศสัมพันธ์ เพื่อนำมาพัฒนาเป็นแอปพลิเคชันแชทบอทสำหรับให้คำแนะนำผู้ใช้งานเบื้องต้น เมื่อศึกษาผลการวิจัยและนำข้อมูลมาทำเป็นชุดคำถามและคำตอบที่รับรองแล้ว โดยแพทย์เพื่อนำไปใช้ในการพัฒนาแอปพลิเคชันและแชทบอทสำหรับให้คำแนะนำ โดยระบบที่พัฒนาขึ้นมีการประเมินด้วยแบบสอบถาม 2 ส่วน คือ ผู้เชี่ยวชาญและผู้ใช้งานทั่วไป พบว่าผลประเมินจากผู้เชี่ยวชาญมีค่าเฉลี่ยความพึงพอใจเท่ากับ 3.68 อยู่ในระดับมาก และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.59 ระดับส่วนเบี่ยงเบนมาตรฐานมีความแตกต่างไม่มาก และผลประเมินจากผู้ใช้งานทั่วไปมีค่าเฉลี่ยความพึงพอใจเท่ากับ 4.27 อยู่ในระดับมาก และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.63 ระดับ

ส่วนเบี่ยงเบนมาตรฐานมีความแตกต่างไม่มาก ระบบที่พัฒนาขึ้นสามารถนำไปใช้งานได้โดยมีประสิทธิภาพ

**คำสำคัญ:** แชทบอท แอปพลิเคชัน สุขภาพทางเพศหญิง

### Abstract

In Thailand, youth and adolescence lack care for their health. Self-control experience knowledge and understanding of increased health risks and health problems that can affect livelihoods. In the digital age, where not everyone is talking, communicating with others or not having enough time to see a doctor at the hospital in person. Application information technology can be easily applied. This research aims to develop a system called the chatbot applications for problems recommender related female's wellness. The dataset is collected from the sexual health problems of female; for example, abnormal uterine bleeding, uterine diseases, and problems of having sexual relations. These sexual health problems are developed into a chatbot application for introductory recommendations. For the dataset in this research, the researchers approve it from the set of questions and answers certified by doctors and develop to chatbot applications for providing sexual health advice for females. The developed system is evaluated in two questionnaire parts which are questionnaire from experts and general users. The result is found that the average of

satisfactory value from experts is 3.68, in the level “high”, and the standard deviation value is 0.59, in the level “quite different”. The average of satisfactory value from general users is 4.27, in the level “high”, and the standard deviation value is 0.63, in the level “quite different”. The proposed systems can implement effectively.

**Keywords:** Chatbot, Application, Female’s wellness

## 1. บทนำ

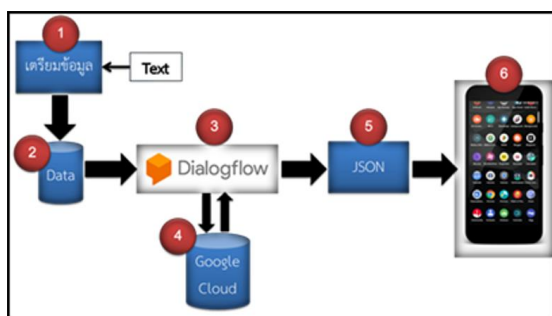
ปัจจุบันกลุ่มเยาวชนและวัยหนุ่มสาว อายุตั้งแต่ (15 - 35 ปี) ในประเทศไทยยังขาดการดูแลเรื่องสุขภาพของตนเอง [1] ประสบการณ์การควบคุมตนเอง ความรู้และความเข้าใจเกี่ยวกับความเสี่ยงเรื่องสุขภาพที่เพิ่มมากขึ้นและปัญหาด้านสุขภาพต่างๆ ได้แก่ การติดเชื้อเอชไอวี (HIV) โรคติดต่อจากการมีเพศสัมพันธ์ การตั้งครรภ์ในสถานการณ์ที่ยังไม่พร้อม [2] ซึ่งอาจส่งผลกระทบต่อดำรงชีวิต เช่น การเรียน การอยู่ร่วมกับผู้อื่นในสังคมเป็นต้น โดยสาเหตุอาจเกิดจากการอับอายที่ต้องให้ผู้อื่นล่วงรู้ถึงปัญหาของตนหรือแม้กระทั่งกลัวการไปพบแพทย์ผู้เชี่ยวชาญเฉพาะทาง ปัจจัยส่วนใหญ่ที่เยาวชนประสบปัญหาดังกล่าว มาจากสภาพแวดล้อม สังคมและความเป็นอยู่ เช่น การอบรมเลี้ยงดูที่ไม่ถูกต้องจากครอบครัว ความสัมพันธ์ระหว่างเพื่อนต่างเพศ หรือสื่อลามกที่สามารถเข้าถึงได้ง่ายขึ้นจากสื่อโซเชียลมากกว่าในอดีต นอกจากนี้ ยังมีปัญหาที่ส่งผลให้เกิดความอับอายต่อผู้อื่น เช่น มีผู้ล่วงรู้ปัญหาพฤติกรรมทางเพศของตนเอง ความอึดอัดใจในการพูดคุยและปรึกษาเรื่องเพศกับผู้อื่น การควบคุมกำกับติดตามของบิดามารดาหรือผู้ปกครองบางคน มีความสัมพันธ์ทางลบให้กับพฤติกรรมเสี่ยงทางเพศ เป็นต้น ทำให้ขาดความรู้และการชี้แนะที่ถูกต้องเกี่ยวกับเรื่องเพศ นอกจากนี้ การตระหนักรู้และเห็นคุณค่าในตนเองจะส่งผลกระทบต่อพฤติกรรมป้องกันการมีเพศสัมพันธ์ในกลุ่มเยาวชนเพิ่มขึ้น [3, 4] ทั้งนี้ไม่ใช่แค่ในกลุ่มเยาวชนที่ประสบกับปัญหาเรื่องสุขภาพทางเพศต่างๆ บางครั้งในกลุ่มผู้ใหญ่ก็ยังคงขาดความรู้และประสบการณ์บางอย่างเกี่ยวกับเรื่องการดูแลป้องกันปัญหาสุขภาพทางเพศ

โรคติดต่อทางเพศสัมพันธ์ หรือแม้กระทั่งการตั้งครรภ์โดยที่ตนเองไม่พร้อม [5, 6] นอกจากนี้ในกลุ่มผู้ใหญ่อาจมีเวลาไม่มากพอที่จะไปขอคำปรึกษาจากคุณหมอเฉพาะทางที่โรงพยาบาล เหตุเพราะทุ่มเทอยู่กับการทำงานอย่างหนักจึงไม่สะดวกที่จะเดินทางไปพบหมอนอกเหนือจากการพูดคุยสื่อสารกับผู้อื่นหรือการไม่มีเวลามากพอที่จะไปพบแพทย์ที่โรงพยาบาลด้วยตนเอง ซึ่งส่วนใหญ่เป็นสิ่งที่ผู้คนในปัจจุบันหลีกเลี่ยงที่จะทำหรือละเลยในการดูแลแก้ปัญหาเมื่อประสบปัญหาทางเพศ ในยุคปัจจุบันที่ข้อมูลข่าวสารสามารถเข้าถึงได้ง่ายด้วยอินเทอร์เน็ตยังมีช่องทางอื่นที่จะให้คำแนะนำแก่บุคคลในกลุ่มนี้ โดยที่ประหยัดเวลาและมีความรวดเร็วในการป้องกันแก้ไขปัญหาเบื้องต้น ทำให้ไม่เกิดความอับอายและมีความเป็นส่วนตัว เช่น การหาข้อมูลจากเว็บไซต์ แอปพลิเคชันบนอุปกรณ์พกพาหรือแชทบอท (Chatbot) [7-9] ซึ่งแชทบอทก็คือระบบตอบโต้อัตโนมัติชนิดหนึ่งที่ถูกพัฒนาขึ้นมาให้มีบทบาทในการตอบกลับการสนทนาด้วยตัวอักษรแบบอัตโนมัติผ่านโปรแกรมส่งข้อความ (Messaging Application) เสมือนการได้พูดคุยโต้ตอบอยู่กับมนุษย์หรืออาจเรียกได้ว่าโปรแกรมตอบกลับอัตโนมัติ (Auto Messaging) ถือเป็นสุดยอดผู้ช่วยอัจฉริยะที่นำมาใช้งานในหลายด้าน และยังสามารถใช้งานผ่านอุปกรณ์พกพาอย่างโทรศัพท์มือถือ ที่ซึ่งอาจเรียกได้ว่าเป็นปัจจัยที่ 5 สำหรับการดำรงชีวิตของวัยรุ่นในปัจจุบันที่มีการพกติดตัวอยู่ตลอดเวลา โดยในบางส่วนหนึ่งมีความสนใจที่จะทำแอปพลิเคชันแชทบอทเกี่ยวกับปัญหาสุขภาพทางเพศของผู้หญิงนี้ขึ้นมาเกิดจากการที่คนรอบข้างเคยประสบกับปัญหาเกี่ยวกับสุขภาพทางเพศและเป็นเรื่องที่ใกล้ตัวมาก และได้ไปเจอเพจหนึ่งเป็นเพจที่มีการเขียนกระทู้เกี่ยวกับปัญหาของผู้หญิงที่น่าสนใจมีรูปภาพ และเนื้อหาอ่านเข้าใจง่าย จึงได้ทำการติดตาม ซึ่งเป็นเพจบนเฟสบุ๊คของ “คุณหมอขวัญชนก หอมแสงประดิษฐ์” [10] เป็นคุณหมอเฉพาะด้านคอยให้คำปรึกษาความรู้เบื้องต้นเกี่ยวกับเรื่องเพศของผู้หญิงจึงยกให้เป็นแหล่งข้อมูลในการนำมาศึกษาเกี่ยวกับปัญหาสุขภาพทางเพศส่วนหนึ่ง และสร้างโปรแกรมตอบกลับอัตโนมัติด้วยโดยการรวบรวมชุด

คำถามและคำตอบมาตรฐาน จากหนังสือต่างๆรวมถึงแหล่งข้อมูลอื่นด้วย จากนั้นนำชุดคำถามและคำตอบที่ได้ไปให้ผู้เชี่ยวชาญเฉพาะด้านทำการตรวจสอบเพื่อความถูกต้องของข้อมูล จากปัญหาที่กล่าวมาข้างต้นผู้วิจัยจึงได้ตระหนักถึงความสำคัญและการมีส่วนร่วมในการช่วยป้องกันและแก้ไขปัญหเกี่ยวกับสุขภาพทางเพศสำหรับผู้หญิง จึงได้สร้างแอปพลิเคชันสำหรับให้ข้อมูลเกี่ยวกับปัญหาทางเพศและมีฟังก์ชันในโปรแกรมตอบกลับอัตโนมัติที่ช่วยแก้ปัญหาในกลุ่มประชากรเพศหญิงที่ต้องการปรึกษาปัญหาสุขภาพทางเพศและต้องการความเป็นส่วนตัวในการขอคำแนะนำซึ่งระบบแอปพลิเคชันและโปรแกรมตอบกลับอัตโนมัติสามารถตอบโต้กับผู้ใช้อย่างสะดวกและสบายใจ ประกอบกับโทรศัพท์มือถือได้เข้ามามีบทบาทสำคัญด้านการส่งผลให้สามารถเข้าถึงแอปพลิเคชันและโปรแกรมตอบกลับอัตโนมัติเพื่อการขอคำแนะนำได้ทุกที่ทุกเวลา

## 2. ขั้นตอนวิธีการดำเนินการวิจัย

บทความนี้นำเสนอการพัฒนาต้นแบบแชทบอทแอปพลิเคชันสำหรับแนะนำปัญหาด้านสุขภาพทางเพศหญิง ดังภาพที่ 1 แสดงกรอบแนวคิดของงานวิจัย



ภาพที่ 1: นำเสนอกรอบแนวคิดของงานวิจัย

โดยมีการเตรียมข้อมูลในรูปแบบข้อความจากแหล่งข้อมูลที่นำเชื่อถือมาเก็บไว้ในฐานข้อมูล นำเทคโนโลยีแชทบอทหรือโปรแกรมโต้ตอบสนทนาอัตโนมัติแทนมนุษย์มาทำหน้าที่ในการตอบสนทนาผ่านตัวอักษรแบบอัตโนมัติ โดยแชทบอทที่ใช้ใน

งานวิจัยนี้ทำงานบนแพลตฟอร์มไดอะล็อกโฟลว์ (Dialogflow) [8] เป็นแบบใช้กฎเป็นตัวตั้งต้นคำสั่งในการสื่อสาร โดยการสร้างกฎหรือคีย์เวิร์ดลงไปในระบบและคำตอบที่ตรงกับคีย์เวิร์ดนั้นถ้าหากคำถามที่ผู้ใช้งานถามตรงกับคีย์เวิร์ดตัวไหนระบบก็จะตอบคำถามตามที่ได้ถูกกำหนดไว้เอง โดยการทำงานผ่านทางหน้าจของผู้ใช้งานแอปพลิเคชันของโทรศัพท์มือถือ ข้อมูลที่ถูกรวบรวมแล้วจากเว็บไซต์จะถูกนำไปผ่านกระบวนการเตรียมข้อมูลโดยมีการทำความสะอาดข้อมูล ตรวจสอบความถูกต้องของข้อมูล และจัดเก็บในรูปแบบข้อความ จากนั้นจะมีการสร้างแชทบอทด้วยโปรแกรมไดอะล็อกโฟลว์ และจัดเก็บไว้บนคลาวด์ของกูเกิลเชิร์ฟเวอร์การนำมาใช้จะใช้ภาษา JSON เป็นตัวดึงข้อมูลจากคลาวด์มาแสดงผลในแอปพลิเคชัน ระหว่างการทำงานต้องมีการเชื่อมต่อกับระบบอินเทอร์เน็ตด้วยจึงจะสามารถใช้งานได้ ข้อมูลที่ใช้ในงานวิจัยได้รวบรวมข้อมูลเกี่ยวกับปัญหาสุขภาพทางเพศของผู้หญิงจากแหล่งข้อมูลต่างๆ เช่น งานวิจัย เว็บไซต์หมอบ หนังสือ หรือสอบถามจากผู้มีประสบการณ์เพื่อรวบรวมคำถามคำตอบ(Q&A) แต่ 80 เปอร์เซ็นต์ในงานวิจัยนี้ รวบรวมมาจากเว็บไซต์ของหมอบ <https://www.pobpad.com/> [11] และเว็บไซต์ของ <https://hd.co.th> [12] จากนั้นได้นำข้อมูลเกี่ยวกับปัญหาสุขภาพทางเพศของผู้หญิงมาวิเคราะห์ทำเป็นชุดคำถาม (Q&A) ในรูปแบบข้อความ โดยสามารถแยกคำถามเป็นหมวดหมู่ ในกระบวนการเตรียมข้อมูลนั้น ข้อมูลที่ผ่านการคัดกรองสำหรับทำชุดคำถามคำตอบจะนำมาผ่านกระบวนการเตรียมความพร้อม ในงานวิจัยนี้ได้มีการทำความสะอาดข้อมูล (Clean Data) ลบข้อมูลส่วนที่ไม่จำเป็นต้องใช้งานออก (Noisy Data) รวมถึงการเติมข้อมูลส่วนที่หายไปให้สมบูรณ์เมื่อเสร็จแล้วก็จัดเก็บในรูปแบบข้อความ (Text File) ไว้ที่คลาวด์กูเกิลเพื่อนำไปใช้ในกระบวนการถัดไป

### 2.1 การพัฒนาต้นแบบแชทบอท

ก่อนอื่นมาทำความเข้าใจกับคำว่าแชทบอทกันก่อน ในงานวิจัยนี้ใช้โปรแกรมไดอะล็อกโฟลว์สำหรับทำแชทบอทเป็นแพลตฟอร์มการสร้างบทสนทนา ด้วย

เทคนิคด้านการเรียนรู้เครื่องจักร (Machine Learning) และการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ที่เชื่อมต่อกับ Google Assistant Dialog flow ทำให้เข้าใจการสนทนาของมนุษย์ การพัฒนาต้นแบบแชทบอทเริ่มจากการนำข้อมูลที่คัดกรองแล้วมาใส่ในเครื่องมือไดอะล็อกโฟลว์ เพื่อทำเป็นแชทบอทสำหรับตอบคำถามผู้ใช้ โดยนำชุดคำถามคำตอบ (Q&A) ไปสร้างเป็นอินเทน (Intents) เพื่อจะเป็นการตรวจสอบว่าข้อความที่ได้รับมา เมื่อผู้ใช้ป้อนข้อมูลที่ใกล้เคียงหรือตรงกับข้อมูลที่กำหนดไว้ล่วงหน้าหรือไม่ (Training Phase) ซึ่งจะเข้าสู่กระบวนการทำงานการคัดกรอง (Intent Classification) โดยจะพยายามหาอินเทนที่ใกล้เคียงกับข้อความที่ได้รับมากที่สุด แต่ถ้าไม่ได้ผลลัพธ์ที่ใกล้เคียงกับข้อความหรืออินเทนใดเลย จะถือว่าเป็นเรื่องอื่นก็จะไปตกที่อินเทนค่าเริ่มต้น (Default Fallback Intent) ที่ได้ตั้งไว้ กล่าวคือคำใดที่เราไม่ได้สอนบอทไว้ หากสิ่งที่ผู้ใช้ถาม สอบถามเข้ามาตรงตามเงื่อนไขของอินเทนที่ตั้งไว้ก็จะจับคู่คำถามคำตอบกัน และส่งคำตอบ (Response) ที่ต้องการกลับไปหาผู้ใช้งานตามลำดับ แชทบอทสามารถทำการฝึกฝน ข้อมูล (Training Data) โดยสามารถตรวจสอบได้ว่า คำตอบที่ส่งไปตรงกับเงื่อนไข (Intents) ที่สร้างไว้หรือไม่ หากคำตอบตรงกับคำถาม ก็จะสามารถอนุมัตินำไปใช้งาน (Approved) ได้เลย แต่ถ้าหากคำตอบไม่ตรงกับคำถาม ก็สามารถแก้ไขคำตอบใหม่ในตัวเงื่อนไข (Intents) ให้ตรงกับที่ผู้ใช้งานสอบถามเข้ามาได้ เพื่อเป็นการฝึกฝนข้อมูลในแชทบอทให้มีประสิทธิภาพมากยิ่งขึ้น ในขั้นตอนรับคำถามจากผู้ใช้งานผ่านตัวแอปพลิเคชันสามารถดูได้ว่าผู้ใช้งานถามคำถามอะไรเข้ามาบ้างและได้คำตอบตรงตามความต้องการนั้นหรือไม่ ตัวแชทบอทบนแอปพลิเคชันมีการตรวจสอบความถูกต้องของการทำงานก่อนนำไปใช้จริง โดยใส่คำถามที่เตรียมไว้ลงไปในตัวบอทที่สร้างขึ้นจำนวน 200 คำถาม จากนั้นนำไปให้ผู้ใช้งานทดลองใช้ ในงานวิจัยนี้ใช้กลุ่มเป้าหมายเป็นบุคลากรในมหาวิทยาลัย จำนวน 30 คน ได้ทดลองใช้หลังจากทำการฝึกอบรม (Training) มาในระยะเวลาหนึ่ง ตัวบอทสามารถตอบตรงประเด็นมากขึ้น สำหรับการนำ

ไดอะล็อกโฟลว์มาใส่ในแอปพลิเคชัน มีขั้นตอนการทำงาน กล่าวคือให้ไปตั้งค่าในส่วนของโปรเจกต์แชทบอทที่คลาวด์กูเกิ้ล (Google Cloud) เพื่อนำ API มาใช้ จากนั้นทำการดาวน์โหลดไฟล์ JSON จากกูเกิ้ลคลาวด์เพื่อนำมาเขียนใส่แอปพลิเคชัน เนื่องจากเครื่องมือไดอะล็อกโฟลว์เป็นเซอร์วิสของกูเกิ้ลคลาวด์อยู่แล้ว จึงสามารถล็อกอินอีเมล (E-Mail) เดียวกับอีเมลที่ใช้ทำไดอะล็อกโฟลว์ได้เลย แล้วตั้งค่าเพื่อทำการดึง API มาใช้ จากนั้น นำมาเขียนใส่ในแอปพลิเคชันและเขียนโค้ดเพื่อให้ปรับระบบการใช้กับแอปพลิเคชันได้ สุดท้ายจะทำการทดสอบตัวแชทบอทบนแอปพลิเคชันที่เสร็จสมบูรณ์แล้ว หลังจากทำการฝึกฝนข้อมูลมาในระยะหนึ่ง ตัวบอทสามารถตอบตรงประเด็นมากขึ้น แต่ก็มีผิดพลาดบ้าง เมื่อนำไดอะล็อกโฟลว์มาใส่ในแอปพลิเคชันเรียบร้อยแล้ว จึงนำไปพัฒนาเป็นแชทบอทแอปพลิเคชันสำหรับให้คำแนะนำด้านสุขภาพภาวะทางเพศหญิง

### 3. ผลการดำเนินการวิจัย

#### 3.1 การวัดประสิทธิภาพตัวต้นแบบ

การวัดประสิทธิภาพออกเป็น 2 ส่วน ได้แก่ ส่วนแรก เป็นการวัดประสิทธิภาพการทำงานของแชทบอท โดยพิจารณาที่ (1) ความถูกต้องในการตอบคำถาม จะคิดเป็นเปอร์เซ็นต์ค่าความถูกต้องและ (2) จำนวนเวลาที่ใช้ในการประมวลผลเพื่อดึงคำตอบมาแสดงผลลัพธ์ให้ผู้ใช้งานได้เห็น ส่วนที่สอง เป็นการวัดประสิทธิภาพตัวระบบแอปพลิเคชัน โดยใช้เกณฑ์การวัดผลแบบมาตราส่วนประมาณค่า 5 ระดับ ได้แก่ มากที่สุด มาก ปานกลาง น้อย และไม่พอใจ ใช้หลักเกณฑ์การแปลความหมายของค่าคะแนนเฉลี่ย (Mean) [13] ดังแสดงในสมการที่ (1)

$$\bar{x} = \frac{\sum x}{n} \quad (1)$$

เมื่อกำหนดให้

- $\bar{x}$  แทน ค่าเฉลี่ยรวมของหัวข้อที่ประเมิน
- $\sum x$  แทน ผลรวมของข้อมูลทั้งหมด
- $n$  แทน จำนวนผู้ประเมินทั้งหมด

และใช้ห้หลักเกณท์การเปลลความหมายของค่าส่วนเบี่ยงเบนมาตรฐาน(Standard Deviation) [14] ดังแสดงในสมการที่ (2)

$$S.D. = \sqrt{\frac{N \sum x^2 - (\sum x)^2}{N(N-1)}} \quad (2)$$

เมื่อกำหนดให้

- S.D. แทน ส่วนเบี่ยงเบนมาตรฐาน
- X แทน คะแนนแต่ละตัว
- N แทน จำนวนคะแนนในกลุ่ม
- Σ แทน ผลรวม

3.2 ผลการศึกษาและอภิปรายผล

ผู้วิจัยได้พัฒนาระบบแอปพลิเคชันแชทบอทที่ให้คำแนะนำปัญหาด้านสุขภาพทางเพศสำหรับเพศหญิงที่ช่วยให้คำแนะนำเบื้องต้นกับผู้หญิงที่ประสบปัญหาเกี่ยวกับสุขภาพทางเพศ โดยนำเทคโนโลยีแชทบอทมาใช้ผ่านช่องทางแอปพลิเคชันบน โทรศัพท์มือถือทำงานบนระบบปฏิบัติการแอนดรอยด์ ดังหน้าจอในภาพที่ 2-3 ใช้ตอบคำถามเบื้องต้นสำหรับสุขภาพทางเพศของผู้หญิง ผู้ใช้งานได้รับความรู้และคำแนะนำเกี่ยวกับเรื่องเพศ ทำให้เพิ่มความสะดวกและมีความเป็นส่วนตัวในการหาคำตอบสำหรับข้อสงสัยสำหรับเพศหญิง โดยพิจารณาความถูกต้องในการตอบคำถามเป็นหลักสำคัญ ซึ่งมีการทดสอบฟังก์ชันทั้งหมด 15 กรณีคำถาม ผลการทดลองพบว่า แชทบอทตอบคำถามถูกจำนวน 14 คำถาม และตอบผิดไป 1 คำถาม คิดเป็นเปอร์เซ็นต์ค่าความถูกต้องได้เป็นจำนวน 9 คำถาม และตอบผิดไป 6 คำถาม คิดเป็นเปอร์เซ็นต์ค่าความถูกต้องได้ เท่ากับ 60 ตรงนี้ ผู้วิจัยวิเคราะห์ว่าจะต้องเพิ่มคำตอบให้มากขึ้นในแต่ละหัวข้อของคำถาม เพื่อให้คำตอบได้ครอบคลุมคำถามและเพิ่มเปอร์เซ็นต์ความถูกต้องให้มีจำนวนเพิ่มมากขึ้นอย่างต่อเนื่อง และควรนำอัลกอริทึมที่เป็นการทำงานแบบอัตโนมัติมาช่วยในการทำงานครั้งนี้ด้วย ส่วนการวัดประสิทธิภาพเวลาที่ใช้ในการประมวลผลเพื่อดึงคำตอบมาแสดงให้ผู้ใช้งานสำหรับการให้คำแนะนำเบื้องต้นอย่างสะดวกและรวดเร็ว ผลการทดสอบโดยตัวแชทบอทใช้เวลาในการประมวลผลและ

ตอบคำถามกลับหาผู้ใช้ ประมาณ 2.65 วินาที ขึ้นไป กรณีไม่สามารถเชื่อมต่ออินเทอร์เน็ต ระบบแชทบอทจะไม่สามารถตอบกลับคำถามได้



ภาพที่ 2: ตัวอย่างหน้าแอปพลิเคชันแชทบอท



ภาพที่ 3: คำถามคำตอบของแชทบอทเกี่ยวกับผิว

จากนั้นทำการประเมินประสิทธิภาพของตัวแชทบอท และทำการทดลองใช้แอปพลิเคชันจริง ผลการทดลองพบว่า ระบบถูกประเมินด้วยแบบสอบถามจากผู้เชี่ยวชาญจำนวน 3 คน และผู้ใช้งาน ทั่วไปจำนวน 30 คน โดยใช้หลักเกณท์การเปลลความหมายของค่าคะแนนเฉลี่ยและหลักเกณท์การเปลลความหมายส่วนเบี่ยงเบนมาตรฐานซึ่งผลลัพธ์ที่ได้จากผู้เชี่ยวชาญมีค่าเฉลี่ยเท่ากับ 3.68 จากระดับคะแนนเต็ม 5 แปลความหมายได้ว่ามีความพึงพอใจอยู่ในระดับมาก และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.59 แปลความหมายได้ว่ามีส่วนเบี่ยงเบนมาตรฐานค่อนข้างมาก และผลลัพธ์ที่ได้จากการประเมินความพึงพอใจของผู้ใช้งานทั่วไป มีค่าเฉลี่ยเท่ากับ 4.27 ซึ่งแปลว่า

ความพึงพอใจอยู่ในระดับมาก และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.63 โดยแปลความหมายได้ว่ามีส่วนเบี่ยงเบนมาตรฐานค่อนข้างมาก

#### 4. สรุป

ผลลัพธ์จากการศึกษาวิจัยแอปพลิเคชันและแชทบอทสำหรับให้คำแนะนำด้านปัญหาสุขภาพทางเพศสำหรับเพศหญิง โดยแอปพลิเคชันทำงานบนระบบปฏิบัติการแอนดรอยด์ผ่านโทรศัพท์มือถือที่มีระบบแชทบอทนี้สามารถตอบคำถามและให้คำแนะนำเบื้องต้นสำหรับผู้หญิงที่มีปัญหา และขาดความรู้ หรือการชี้แนะที่ถูกต้องเกี่ยวกับเรื่องเพศได้ ทั้งนี้รวมไปถึงช่วยอำนวยความสะดวกและประหยัดเวลาให้กับผู้หญิงบางคนที่ยังไม่มีเวลาว่างจะไปขอคำปรึกษาจากคุณหมอเฉพาะทางที่โรงพยาบาลโดยตรง ก็จะสามารถสอบถามอาการและวิธีการป้องกันเบื้องต้นเพื่อรับคำแนะนำจากตัวแอปพลิเคชันและแชทบอทก่อนได้ ในอนาคตจะมีการพัฒนาระบบแชทบอทให้เป็นแบบอัตโนมัติหรือนำเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) เข้ามาช่วยในการพัฒนาเพื่อนำไปต่อยอดได้ในเรื่องของคำถาม การเก็บข้อมูลที่ใช้ส่งคำถามเข้ามาเพื่อนำมาวิเคราะห์และจัดการข้อมูลของคำถามและคำตอบที่คาดว่าผู้ใช้น่าจะถามคำถามที่ต่อเนื่องของเรื่องนั้นๆ และมีการเพิ่มหน้า หรือเนื้อหาที่น่าสนใจเพิ่ม เช่น ลิงก์วิดีโอ รวมถึงเรื่องความปลอดภัยของข้อมูลที่ปรึกษาด้วย

#### เอกสารอ้างอิง

- [1] D. Angkhawanawit, "Problems of violence in children and adolescents". [Internet]. 2014 [updated\2018\May\14]. Availability from: <https://www.rama.mahidol.ac.th>
- [2] J. Kontha, P. Choteamnuay, W. Prasittichok, and U. Intanmas, "Guidelines for the prevention of teenage pregnancy in Thailand: a systematic synthesis," *Ganesha Journal*, Vol.15, No.2, pp. 161-173, 2019.
- [3] A. Pruitworawong, P. Phatthanurak, S. Ritmontri, "Self-esteem of women at Rajvithi home for girls, Bangkok," *Vajira Nursing Journal*, Vol.18, No.2, pp. 51-62, 2016.
- [4] D. Thongyoo, "Guideline for developing adolescent self-esteem based on self-efficacy theory," *Valaya Alongkorn Journal*, Vol.4, No.2, pp. 179-190, 2014.
- [5] S. Rongluen, P. Talengjit, S.Siriborirak, "Unwanted pregnancies in teenagers: a survey of problems needs for health care support," *Siriraj Nursing Journal*, Vol.5, No.1, pp. 14-28, 2012.
- [6] J. Sangsawang, B. Supavititpatana, P. Sriarporn, "Factors related to health promoting behaviors among adolescent mothers," *Nursing Journal*, Vol.46, No.4, pp. 59-69, 2019.
- [7] Modcumram, "The chatbot creating using dialogflow and abdul, mixed content in form of serverless". [Internet]. 2018 [updated \2020\Feb\26]. Availability from: <https://medium.com>.
- [8] P. Kruapanich, "Try making a simple chatbot too, let's flow dialog". [Internet]. 2018 [updated\2018\Apr\19]. Availability from: <https://medium.com>.
- [9] K. Viriyayudhakorn, "Developing a robot system for automating conversations in Thai language with dialog flow". [Internet]. 2018 [updated\2018\Jul\18]. Availability from: <https://kobkrit.com>.
- [10] K. Homsaengpradit, "Page of little sister care". [Internet]. 2019 [updated\2020\Mar\10]. Availability from: <https://www.facebook.com/LittleSisCare>.
- [11] A. Pobpad, "POBPAD". [Internet]. 2019 [updated\2020\Jan\1; cited\2020\Mar\1]. Availability from: <https://www.pobpad.com>.
- [12] A. Hd, "Start here, healthy hd starts here". [Internet]. 2019 [updated\2020\Feb\1; cited\2020\Mar\2]. Availability from: <https://hd.co.th>.
- [13] K. Viriyayudhakorn, "Chatbot Agent". [Internet]. 2019 [updated\2021\Jan\1; cited\2020\Aug\1]. Availability from: <https://kobkrit.com/>.
- [14] C. Sitthimamon, K. Khiaoayu and V. Udomprasertkul, "Selection of statistics in research". [Internet]. 2019 [updated\2007\May\25;cited\2020\Mar\25]. Availability from: <http://cai.md.chula.ac.th>.

## AUTHOR INDEX

	<b>Pages</b>
Adithep Panwanichayakij	394
Akharin Khunkitti	448
Anongwanneeeya Suayroop	266
Anurak Yutthanawa	13
Apichaya Nimkoompai	87, 296, 491
Apisit Kraiyaso	140
Apisit Seelsat	202
Arisa Vaichiaranai	80
Atikhom Siri	111
Awissara Dokauboi	358
Bussayamas Pakkeeptong	13
Chatchadawan Chanprasert	370
Chiraporn Thepthong	248
Chumpol Mokarat	93
Chutimon Pakdeesirote	302
Damrongpol Suwanchatree	235
Eakarach Nintra	442
Eakasit Pacharawongsakda	302
Jamnapa Yodho	260
Janejeera Aksornpim	358
Jidapa Dumrongsombut	105
Jirarat Srichok	134
Kamonchat Deela	388
Kan Khemnak	346
Kanchana Viriyapant	190, 467, 479
Kantapon Seshajoti	516
Kantarakorn Jitharn	302
Kanyarat Charoenpru	117
Karsiya Klintharvorn	436
Kawinthip Thumchon	61
Ketsuda Chainok	358
Kewalin Khamnipat	19
Khamron Sunat	321
Kittamet Montheankul	430
Konglit Hunchangsith	315
Kontorn Inprasith	406
Kornthip Whanchaaim	315

## AUTHOR INDEX

	<b>Pages</b>
Krittamuk Lekseng	290
Maruchai Wattanarungson	43
Montean Rattanasiriwongwut	134, 266, 290, 327, 340, 436, 509
Naris Mingmora	358
Natarda deearom	214
Natkamon Siripanurak	400
Nattaphon Chaiwong	111
Nattapong Laehte	340
Nattavadee Hongboonmee	61
Nattavee Utakrit	67, 99, 220, 346, 485
Nattawadee Sukkasem	523
Natthaphat Kongthong	460
Natthapoom Boonruangroj	485
Nawakarn Ninprawit	254
Neungrutai Deemeewong	321
Niratchaporn Ounnimit	491
Non Chitchamnong	418
Nutsurang Dumlakor	93
On-Uma Pramote	19
Pahron Sanguanbhokai	202
Pakorn Kaewkhampaeng	31
Panatda Yeemasa	473
Passorn Sukmon	308
Pawitra Sariwong	388
Payap Sirinam	49, 196
Peemai Keelakitpaibool	382
Phattaradon Ekthammabordee	87
Phayung Meesad	164, 171
Phithayarat Phochompu	184
Phongthep Raxpakawong	412
Phonpan Charoensin	454
Phumin Duankhan	321
Phuriphath Theerawut	497
Phutares Suwannarat	424
Pichet Wayalun	13
Pongpan Kaeduang	190



## AUTHOR INDEX

	<b>Pages</b>
Pongsarun Boonyopakorn	1, 37, 334
Ponrudee Netisopakul	424
Pramote Luenam	43, 241, 260, 284
Prasong Praneetaphonkrang	196
Prasong Praneetpolgrang	49, 202
Prattana Deeprasertkul	117
Preecha Harnyuenyongsakul	37
Pudsadee Boonrawd	55, 105, 158, 382
Pumarin Tiangpanich	296
Rattanakorn Saeng-Aun	412
Sakchai Tangwannawit	152, 308, 352, 364, 406, 454, 473
Sakolpuck Jearawattanasawadi	302
Saksit Jantaraplin	334
Sappanyou Chukaew	49
Sarin Songkarin	140
Sasima Tanapatpiboon	278
Satit Khwanchum	73
Sirepong Padit	128
Siriporn Mongkolchaiwanich	352
Siwakorn Banluesapy	178
Somboon Udnan	196
Somchat Jiriwibhakorn	442
Sopon Phumeechanya	430
Sudthida Changpan	376
Sujanya Tubtong	146
Sunisa Chanbang	503
Supachanun Wanapu	358
Supakkavee Supho	448
Supannada Chotipant	454, 503
Supaporn Deankasai	321
Supasuda Sudprasert	430
Supawan Pongsupawan	241
Suphakit Awiphan	7
Suphakit Nachom	503
Suphaluck Charoenthum	327
Suphaporn Jamrurai	364

## AUTHOR INDEX

	<b>Pages</b>
Surapan Airphaiboon	442
Sutasinee Thovuttikul	497
Sutasorn Yumak	171
Suthatip Parittapongsachai	164
Suttitee Jathuphornpaserd	7
Suwapat Janmaitree	152
Suwat Pattaramalai	31
Suwimon Tuanthong	523
Taechathat Sirimai	67
Tanapon Jensuttiwelchakul	214, 248, 254, 272, 278, 376, 400
Tanasanee Phienthrakul	315
Tantikorn Poonyakariyakorn	226
Teerawat Sinlaparatsamee	467
Thanachart Ritbumroong	226
Thanakorn Liwveha	220
Thanaphon Tangchoopong	184, 388
Thanaporn Tepkila	272
Thanchanok Rangkaew	302
Thaninthorn Whasphutthisit	122
Thanisa Sungsaeng	158
Thanthicha Nimnuan	55
Thati Phiphatprueksa	358
Theerawat Patcham	509
Thepparit Banditwattanawong	25, 80, 516
Thiti Tongumpun	479
Urachart Kokaew	208
Uraivan Inyaem	523
Vacharavat Kesbuntherng	460
Vanitchaporn Masomboon	430
Wacharaphol Ketwongsa	208
Waraporn Jirapanthong	178, 418
Warunya Wunnasri	140
Watcharakrit Lertthamphaiboon	1
Watcharee Jumpamule	497
Watchareewan Jitsakul	122, 128, 146, 235, 370, 394
Weena Janratchakool	523
Wipada Silarach	184

## **AUTHOR INDEX**

	<b>Pages</b>
Wiroj Tamboonlertchai	93
Wirote Cherdswat	284
Worachet Nimcharoen	448
Wuntana Tientong	352
Wuttachai Chumyen	25
Wuttinun Chaipromnimit	99



# **NCCIT2023**

**The 19<sup>th</sup> National Conference on  
Computing and Information  
Technology**

---

**Organized by**

Faculty of Information Technology and Digital Innovation  
King Mongkut's University of Technology North Bangkok

---

**[WWW.NCCIT.NET](http://WWW.NCCIT.NET)**

## Conference Organizers



ECU



OSU



HNUE



FernUni



KMUTNB



TU Chemnitz



Gesellschaft für Informatik



MSU



KRU



NPRU



PBRU



RMUTK



RMUTT



NIDA



[www.nccit.net](http://www.nccit.net)