Journal of Applied Statistics and Information Technology

Vol 7 | No 1 | 2022

NIDA WISDOM for Change
คณะสถิติประยุกต์
Graduate School of Applied Statistics

Online ISSN : 2465-4949

บทบรรณาธิการ

วารสารสถิติประยุกต์และเทคโนโลยีสารสนเทศ เป็นวารสารวิชาการที่จัดทำโดยคณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ มีจุดประสงค์เพื่อส่งเสริมงานวิจัยในด้านสถิติประยุกต์ เทคโนโลยีสารสนเทศและศาสตร์ที่เกี่ยวข้อง โดยการเผยแพร่ผลงานวิจัยที่มีคุณภาพจากนักวิชาการ นักวิจัยหรือนิสิตนักศึกษาโดยทั่วไปในรูปของวารสารแบบออนไลน์ สำหรับฉบับที่ 1 ปี 2565 ประกอบด้วยบทความวิจัยจำนวน 5 บทความ กองบรรณาธิการวารสารสถิติประยุกต์และเทคโนโลยีสารสนเทศหวังเป็นอย่างยิ่งว่าบทความวิจัยที่ตีพิมพ์ในวารสารฉบับนี้จะเป็นประโยชน์ต่อผู้อ่านที่จะนำไปประยุกต์ใช้หรือทำวิจัยต่อไป

กองบรรณาธิการ

รองศาสตราจารย์ ดร.สุรพงค์ เอื้อวัฒนามงคล **บรรณาธิการ**

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

ผู้ช่วยศาสตราจารย์ ดร. ปรีชา วิจิตรธรรมรส **รองบรรณาธิการ**

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

ศาสตราจารย์ ดร.สำรวม จงเจริญ **กรรมการ**

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

ศาสตราจารย์ ดร. ชิดชนก เหลือสินทรัพย์ **กรรมการ**

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา **กรรมการ**

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

รองศาสตราจารย์ ดร.เยาวดี เต็มธนาภัทร์ **กรรมการ**

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

รองศาสตราจารย์ ดร. ฐิติวดี ชัยวัฒน์ **กรรมการ**

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

รองศาสตราจารย์ ดร.วราภรณ์ จิรชีพพัฒนา **กรรมการ**

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

คณะกรรมการบริหารวารสาร

ผู้ช่วยศาสตราจารย์ ดร. ปราโมทย์ ลือนาม	ประธานคณะกรรมการ
อาจารย์ ดร.ศิวิกา ดุษฎีโหนด	กรรมการ
ผู้ช่วยศาสตราจารย์ ภัทราวดี ธนวงศ์สุวรรณ	กรรมการ
รองศาสตราจารย์ ดร.สุรพงศ์ เอื้อวัฒนามงคล	กรรมการ
ผู้ช่วยศาสตราจารย์ ดร.ธนาสัย สุคนธ์พันธุ์	กรรมการ
อาจารย์ ดร.ธนชาตย์ ฤทธิ์บำรุง	กรรมการ
นางสุภาพร หงษาวงษ์	กรรมการและเลขานุการ

สารบัญ

บทความรับเชิญ

ใคร Vote ให้ ใคร?
ศึกชิงผู้ว่ากรุงเทพมหานคร ปี 2565

พาชิตชนัต ศิริพานิช[1] และ เดือนเพ็ญ ธีรวรรณวิวัฒน์[2]

การเลือกตั้งเป็นช่องทางที่สำคัญในการที่ประชาชนจะเข้าไปมีอิทธิพลต่อนโยบายสาธารณะ ซึ่งถือได้ว่าเป็นรูปแบบหนึ่งของการมีส่วนร่วมทางการเมือง (Wolfinger and Rosenstone, 1980) การเลือกตั้งไม่ว่าในระดับประเทศหรือระดับท้องถิ่นก็ตาม นอกจากประชาชนผู้มีสิทธิ์ออกเสียงเลือกตั้งแล้ว ยังมีกลุ่มผู้มีส่วนได้เสียที่สำคัญอีก 2 กลุ่ม คือ ผู้สมัครรับเลือกตั้ง และหน่วยงานที่จัดการเลือกตั้ง แต่กลุ่มประชาชนผู้มีสิทธิ์ออกเสียงเลือกตั้งถือว่ามีความสำคัญมาก เพราะผลของการเลือกตั้งจะออกมาเช่นใด ผู้สมัครคนใดชนะการเลือกตั้งก็ขึ้นอยู่กับการเลือกของผู้มีส่วนได้เสียกลุ่มนี้ อย่างไรก็ตามมีประเด็นปัญหามากมายที่เกี่ยวข้องกับพฤติกรรมการเลือกตั้งของประชาชนผู้มีสิทธิ์ออกเสียงเลือกตั้ง อาทิเช่น แหล่งและวิธีการหาข้อมูลเกี่ยวกับผู้สมัคร รวมทั้งลักษณะของข้อมูลที่มี (เช่นเป็นแบบ fully หรือ แบบ shortcut หรือแบบอื่น ๆ) ประเด็นเหล่านี้ส่งผลต่อการตัดสินใจในการลงคะแนนเสียงอย่างไร (Banerjee and Others, 2011; Ryan, 2010; Bartels, 1996) การมีหรือไม่มีข้อมูลที่เกี่ยวกับการเลือกตั้งและตัวผู้สมัครรับเลือกตั้งเกี่ยวข้องกับการไปใช้สิทธิ์เลือกตั้งหรือไม่ (Crowder-Meyer and Gadarian, 2020; McMurray, 2015; Matsusaka, 1995) การเลือกผู้สมัครเป็น rational choice หรือไม่ อะไรเป็นแรงจูงใจในการไปลงคะแนนเสียงเลือกตั้ง และเหตุใดจึงเลือกผู้สมัครคนนั้นคนนี้ (Lovernier and Barilla, 2006; Ebeid and Rodden, 2006) สิ่งเหล่านี้เป็นคำถามที่ซับซ้อนและยังไม่มีคำตอบที่ชัดเจน แต่ทั้งหมดนี้อยู่บนพื้นฐานของคำถามง่ายๆว่า “ใคร Vote ให้ใคร?”

ด้วยเหตุผลดังกล่าวข้างต้น การศึกษานี้จึงมีวัตถุประสงค์เพื่อศึกษาแบบแผนการเลือกผู้สมัครผู้ว่าราชการกรุงเทพมหานครในปี 2565 ของผู้มีสิทธิ์ลงคะแนนเสียงเลือกตั้งที่มีลักษณะทางประชากร สังคม และเศรษฐกิจแตกต่างกัน โดยใช้ข้อมูลจากการสำรวจตัวอย่าง (sample survey) ของนิด้าโพล ผลการศึกษานี้จะเป็นข้อมูลเชิงประจักษ์ที่จะทำให้เราทราบว่า “ใคร Vote ให้ใคร?” หรือกล่าวอีกนัยหนึ่ง คือผู้ที่ตั้งใจจะไปลงคะแนนเสียงให้กับผู้สมัครผู้ว่าราชการกรุงเทพมหานครแต่ละคน เป็นประชาชนชาว กทม. กลุ่มใด มีอายุ เพศ การศึกษา รายได้ และอาชีพเป็นอย่างไร คำตอบของคำถามนี้เป็นทั้งผลลัพธ์ปลายทางที่สะท้อนให้เห็นถึงผลของยุทธศาสตร์/กลยุทธ์และวิธีการหาเสียงของผู้สมัครชิงตำแหน่งผู้ว่าราชการกรุงเทพมหานคร และเป็นจุดเริ่มต้นที่สำคัญของการศึกษาในเชิงวิชาการในประเด็นปัญหาต่างๆที่เกี่ยวข้องต่อไปด้วย

[1, 2] รองศาสตราจารย์ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ (นิด้า)

1. ระเบียบวิธีศึกษา

การรับทราบลักษณะส่วนบุคคลของผู้ที่ไปใช้สิทธิ์เลือกตั้ง เพื่อทราบว่า “ใคร Vote ให้ใคร” จำเป็นต้องใช้ข้อมูลจากการสำรวจตัวอย่าง (Sample Survey) การศึกษานี้ใช้ข้อมูลทุติยภูมิ ซึ่งได้มาจากการสำรวจของศูนย์สำรวจความคิดเห็น “นิด้าโพล” สถาบันบัณฑิตพัฒนบริหารศาสตร์ โดยข้อมูลหลักที่ใช้ ได้แก่ ข้อมูลจากการสำรวจตัวอย่างขนาด 2,500 คน เรื่อง “22 พ.ค. 2565 คนกรุงเทพฯ เลือกใครเป็นผู้ว่าฯ กทม.” (https://nidapoll.nida.ac.th/survey_detail?survey_id=573 ค้นเมื่อ 25 พฤษภาคม 2565) ซึ่งดำเนินการสำรวจระหว่างวันที่ 18 – 21 พฤษภาคม 2565 (เผยแพร่ผลการสำรวจวันที่ 22 พฤษภาคม 2565 หลังเวลา 17.00 น.)

ในการสำรวจครั้งนี้ นิด้าโพลใช้แผนการเลือกตัวอย่างสุ่มแบบง่าย (Simple Random Sampling) จากตัวอย่างหลัก (Master Sample) ของศูนย์ฯ โดยพิจารณาขนาดตัวอย่าง (n) ที่ทำให้ความคลาดเคลื่อน (e) ของการประมาณค่าสัดส่วนประชากร (p) ด้วยสัดส่วนตัวอย่าง ($\hat{p}$) มีค่าไม่เกิน 0.02 ด้วยความเชื่อมั่น 95%

ในการเลือกตั้งนั้น สิ่งที่ใช้ในการพิจารณาคือสัดส่วนหรือร้อยละ ดังนั้น การนำเสนอและการวิเคราะห์ข้อมูลในการศึกษานี้ จึงขอนำเสนอในรูปร้อยละ ประกอบกับการใช้การทดสอบ Z – Test และ Chi-square Test สำหรับทดสอบสมมุติฐานเกี่ยวกับสัดส่วนตามความเหมาะสม กล่าวคือ ใช้การทดสอบ Z – Test สำหรับทดสอบสัดส่วนประชากร 1 ประชากรเพื่อเปรียบเทียบสัดส่วนตัวอย่าง ($\hat{p}$) กับสัดส่วนที่ได้จากผลการลงคะแนนเสียงจริง (p) และสำหรับทดสอบสัดส่วนประชากร 2 ประชากรเพื่อเปรียบเทียบระหว่างสัดส่วนตัวอย่างของคุณลักษณะ 2 อย่าง เช่น สัดส่วนของผู้หญิง ($\hat{p}_F$) และผู้ชาย ($\hat{p}_M$) ที่เลือกผู้สมัครคนหนึ่ง และใช้การทดสอบ Chi-square Test สำหรับทดสอบสัดส่วนประชากรมากกว่า 1 ประชากร

2. ผลการศึกษา

ข้อมูลจากผลการสำรวจตัวอย่างจำนวน 2,500 คน เรื่อง “22 พ.ค. 2565 คนกรุงเทพฯ เลือกใครเป็นผู้ว่าฯ กทม.” (https://nidapoll.nida.ac.th/survey_detail?survey_id=573 ค้นเมื่อ 25 พฤษภาคม 2565) พบว่า ตัวอย่างที่ใช้ในการศึกษาประกอบด้วยประชาชนอายุ 18 – 93 ปี (ค่าเฉลี่ยของอายุ = 46.67 ปีและค่าเบี่ยงเบนมาตรฐาน = 16.34 ปี) จากทั้ง 50 เขตของกรุงเทพมหานคร โดยร้อยละ 45.44 เป็นผู้ชาย และร้อยละ 54.56 เป็นผู้หญิง นอกจากนี้ตัวอย่างยังครอบคลุมทุกกลุ่มอาชีพ ระดับรายได้ และอื่น ๆ

ผลการสำรวจของนิด้าโพลดังกล่าว พบว่า ผู้สมัครที่ได้คะแนนสูงสุด 5 อันดับแรก ได้แก่ (1) ดร.ชัชชาติ สิทธิพันธุ์ ผู้สมัครอิสระ (2) พล.ต.อ.อัศวิน ขวัญเมือง (3) ดร.สุชัชวีร์ สุวรรณสวัสดิ์ (4) ดร.วิโรจน์ ลักขณาอดิศร และ (5) นายสกลธี ภัททิยกุล ซึ่งตรงกับผลการเลือกตั้งผู้ว่าฯ กทม. เมื่อวันที่ 22 พฤษภาคม 2565 กล่าวคือ ลำดับที่ 1 ได้ผลเหมือนกันทั้งตัวบุคคลและคะแนนที่ได้รับเสียงสนับสนุนก็มีค่าใกล้กัน และรายชื่อของลำดับที่ 2 ถึงลำดับที่ 5 ก็เป็นชุดเดียวกัน นอกจากนี้ ผลการสำรวจของนิด้าโพลยังพบว่า สัดส่วนของคะแนนสนับสนุนสำหรับผู้สมัคร 5 คนนี้แตกต่างจากผู้สมัครคนอื่น ๆ อยู่พอสมควร ซึ่งตรงกับผลการเลือกตั้งผู้ว่าฯ กทม. กล่าวคือ ในการเลือกตั้งจริง ผู้สมัคร 5 คนดังกล่าวเป็นผู้ที่ได้คะแนนเสียงสนับสนุนมากกว่า 200,000 คะแนน ส่วนผู้สมัครลำดับรองลงไปอื่น ๆ ได้คะแนนเสียงสนับสนุนไม่เกิน 100,000 คน (www.thaipost.net/hi-light/146948/ ค้นเมื่อ 25 พฤษภาคม 2565) ดังนั้น การนำเสนอผลการศึกษาในครั้งนี้ จะนำเสนอผลเฉพาะที่เกี่ยวกับผู้สมัครผู้ว่าฯ กทม. 5 คนดังกล่าวนี้เท่านั้น

2.1 ผลการศึกษาในภาพรวม

ในการเลือกตั้งผู้ว่าฯ กทม. เมื่อวันที่ 22 พฤษภาคม 2565 ที่ผ่านมามีผู้ไปใช้สิทธิ์ 2,673,696 คนจากผู้มีสิทธิ์ทั้งหมด 4,402,948 คนคิดเป็นร้อยละ 60.73 (https://workpointtoday.com/wptodaybkk2022-5/ ค้นเมื่อ 25 พฤษภาคม 2565) และผู้ที่ได้รับเสียงสนับสนุนสูงสุด คือ ดร.ชัชชาติ สิทธิพันธุ์ ผู้สมัครอิสระ ได้เสียงสนับสนุนถึง 1,386,215 เสียง/คน คิดเป็นร้อยละ 51.85 เมื่อเปรียบเทียบกับผลการสำรวจของนิด้าโพลแล้ว พบว่า ผลการสำรวจของ

นิด้าโพลมีค่าต่ำกว่าผลการเลือกตั้งจริงอยู่ร้อยละ 1.13 (ดูตารางที่ 1) แต่ความแตกต่างดังกล่าวไม่มีนัยสำคัญทางสถิติ (Z = -1.127 และ p – Value = 0.13) ในทำนองเดียวกัน ผลการสำรวจของนิด้าโพลสำหรับ ดร.สุชัชวีร์ สุวรรณสวัสดิ์ และ ดร.วิโรจน์ ลักขณาอดิศร กับผลผลการเลือกตั้งจริงแม้จะต่างกันอยู่บ้างแต่ความแตกต่างก็ไม่มีนัยสำคัญทางสถิติ (Z = 1.419, 1.334 และ p – Value = 0.078, 0091 ตามลำดับ)

อย่างไรก็ตาม ผลการสำรวจของนิด้าโพลสำหรับ พล.ต.อ.อัศวิน ขวัญเมือง และนายสกลธี ภัททิยกุล พบว่า สัดส่วนคะแนนเสียงของผู้สมัครทั้ง 2 คนที่ได้รับจากประชาชนในการเลือกตั้งจริง มีความแตกต่างจากผลโพลของนิด้าอย่างมีนัยสำคัญทางสถิติที่ระดับสูงมาก (Z = 5.162, -3.246 และ p – Value = 0.000, 0001 ตามลำดับ) นอกจากนี้ ยังพบว่าลำดับที่ของ พล.ต.อ.อัศวิน ขวัญเมือง คลาดเคลื่อนไปพอสมควร ดังรายละเอียดแสดงไว้ในตารางที่ 1

ตารางที่ 1 ร้อยละของผู้มีสิทธิ์เลือกตั้ง จำแนกตามผู้สมัครรับเลือกตั้งผู้ว่าฯ กทม. 5 คนแรกที่ได้มีค่าสูงสุด จากผลการเลือกตั้งจริงเมื่อวันที่ 22 พฤษภาคม 2522 และจากผลการสำรวจของนิด้าโพล พร้อมทั้งค่าสถิติ Z และค่าพี (p-Value) สำหรับการทดสอบ Z – Test

ผู้สมัครรับเลือกตั้งผู้ว่า กทม.	ผลการเลือกตั้ง[1] (22 พ.ค. 2565)	ผลการสำรวจโดยนิด้าโพล	ค่าสถิติ Z (Z Statistic)	ค่าพี (p – Value)
ดร.ชัชชาติ (อิสระ)	51.85	50.72	-1.127	0.130
พล.ต.อ.อัศวิน (อิสระ)	8.03	10.84	5.162	0.000**
ดร.สุชัชวีร์ (พรรคประชาธิปัตย์)	9.53	10.36	1.419	0.078
ดร.วิโรจน์ (พรรคก้าวไกล)	9.50	10.28	1.334	0.091
นายสกลธี (อิสระ)	8.62	6.80	-3.246	0.001**

[1]ที่มา: https://www.thaipost.net/hi-light/146948/ ค้นเมื่อวันที่ 25 พฤษภาคม 2565

หมายเหตุ ** หมายถึง ความแตกต่างของสัดส่วนมีนัยสำคัญทางสถิติที่ระดับ 1%

2.2 ผลการศึกษาจำแนกตามลักษณะส่วนบุคคล

ข้อมูลจากการสำรวจ แม้จะเป็นเพียงข้อมูลจากตัวอย่างสุ่ม แต่ก็มีข้อดีในแง่ที่ว่า เป็นข้อมูลที่ทราบลักษณะส่วนบุคคลของผู้ใช้สิทธิ์เลือกตั้ง โดยผลการเลือกตั้งที่ประกาศเป็นทางการไม่สามารถให้รายละเอียดเกี่ยวกับลักษณะส่วนบุคคลของประชาชนที่ลงคะแนนเลือกผู้สมัครแต่ละคนได้ด้วยเหตุผลหลายประการ โดยเฉพาะอย่าง ด้านที่เกี่ยวข้องกับการซื้อสิทธิ์ขายเสียง ในกรณีที่ใช้ผลการสำรวจของนิด้าโพลนี้ จะช่วยให้เห็นรายละเอียดว่า ผู้สมัครผู้ว่าฯ กรุงเทพมหานคร ที่ได้คะแนนสูงสุด 5 อันดับแรกนั้น “ใคร Vote ให้ใคร?” หรือผู้สมัครคนใด ได้คะแนนเสียงจากประชาชนกลุ่มใดมาก-น้อยกว่ากัน

2.3 ผลการศึกษาจำแนกตามอายุ

เมื่อพิจารณาผลการสำรวจของนิด้าโพลจำแนกตามอายุ พบว่า ดร.ชัชชาติ สิทธิพันธุ์ ผู้ได้รับคะแนนเสียงเลือกตั้งเป็นอันดับที่ 1 มีสัดส่วนสูงมากทุกช่วงอายุ โดยเฉพาะในกลุ่มประชาชนที่มีอายุไม่เกิน 35 ปี พบว่า มีมากกว่าร้อยละ 60 ที่สนับสนุน ดร.ชัชชาติ ดังรายละเอียดแสดงในตารางที่ 2 แม้ว่าสัดส่วนดังกล่าวนี้ จะมีแนวโน้มลดลงตามช่วงอายุที่เพิ่มขึ้น แต่ก็ยังนับว่ามีสัดส่วนของผู้สนับสนุนมากกว่าผู้สมัครคนอื่น ๆ อย่างมาก โดยอายุเฉลี่ยของผู้สนับสนุน ดร.ชัชชาติ เท่ากับ 44.29 ปี ในทำนองเดียวกัน สัดส่วนของผู้สนับสนุน ดร.วิโรจน์ ลักขณาอดิศร ก็ลดลงตามอายุที่เพิ่มขึ้นเช่นเดียวกัน ดังเห็นได้ชัดเจนในแผนภูมิที่ 1 โดยสัดส่วนของประชาชนในกลุ่มอายุไม่เกิน 35 ปีที่

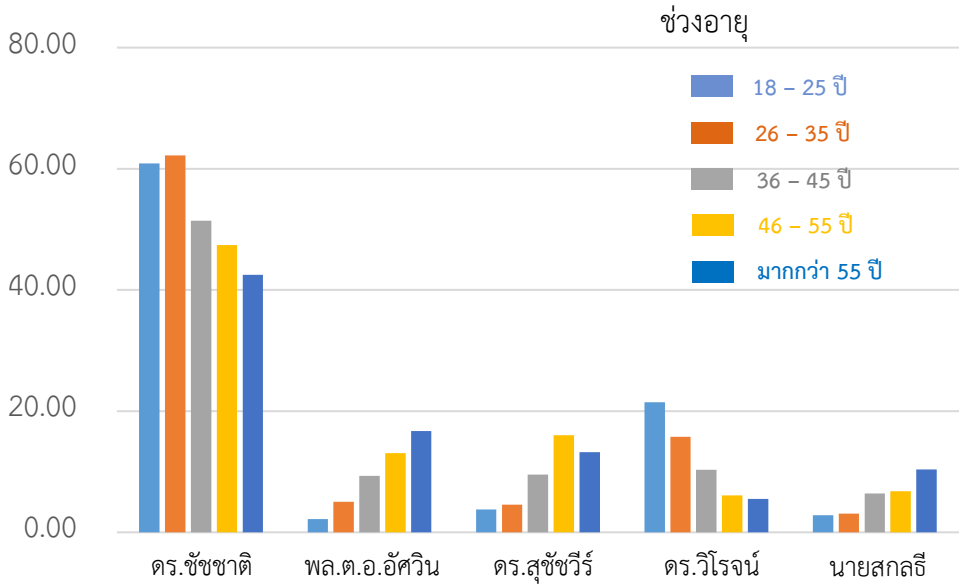
สนับสนุน ดร.วิโรจน์ แม้ว่าจะมีสัดส่วนเป็นอันดับสองที่ไม่มากเท่าของ ดร.ชัชชาติ แต่เมื่อเทียบกับผู้สมัครคนอื่น ๆ แล้ว เห็นได้ชัดเจนว่า ดร.วิโรจน์ได้รับการสนับสนุนจากกลุ่มผู้มีอายุน้อย (ไม่เกิน 35 ปี) โดยอายุเฉลี่ยของผู้ให้การสนับสนุน ดร.วิโรจน์ มีค่าเท่ากับ 38.23 ปี

ในทางกลับกัน พบว่า สัดส่วนของประชาชนที่สนับสนุน พล.ต.อ.อัศวิน ขวัญเมือง และนายสกลธี ภัททิยกุล เพิ่มขึ้นตามช่วงอายุที่เพิ่มขึ้น ในขณะที่ ประชาชนที่สนับสนุน ดร.สุชัชวีร์ สุวรรณสวัสดิ์ ก็เพิ่มขึ้นตามช่วงอายุที่เพิ่มขึ้นจนถึงอายุไม่เกิน 55 ปี แล้วสัดส่วนของประชาชนผู้มีอายุมากกว่า 55 ปีที่สนับสนุน ดร.สุชัชวีร์ ก็ลดลงเล็กน้อยเมื่อเปรียบเทียบกับผู้มีอายุ 46 – 55 ปี อายุเฉลี่ยของผู้สนับสนุนผู้สมัคร 3 คนนี้เท่ากับ 54.36 ปี 53.81 ปี และ 53.06 ปี ตามลำดับ

ตารางที่ 2 ร้อยละของผู้มีสิทธิ์เลือกตั้งในแต่ละช่วงอายุ จำแนกตามผู้สมัครรับเลือกตั้งผู้ว่าฯ กทม. 5 คนแรกที่ได้มีค่าสูงสุด จากผลการสำรวจของนิด้าโพล พร้อมทั้งค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของอายุ

ผู้สมัครรับเลือกตั้งผู้ว่าฯ กทม.	ช่วงอายุ (ปี)					รวม	**ค่าเฉลี่ยของอายุ**	ค่าเบี่ยงเบนมาตรฐาน
	18 – 25	26 – 35	36 – 45	46 – 55	> 55			
ดร.ชัชชาติ	60.88	62.20	51.45	47.40	42.48	50.72**	**44.29**	(16.27)
พล.ต.อ.อัศวิน	2.21	5.02	9.30	13.09	16.71	10.84**	**54.36**	(13.95)
ดร.สุชัชวีร์	3.79	4.55	9.50	16.03	13.25	10.36**	**53.06**	(14.79)
ดร.วิโรจน์	21.45	15.79	10.33	6.09	5.49	10.28**	**38.23**	(15.44)
นายสกลธี	2.84	3.11	6.40	6.77	10.38	6.80**	**53.81**	(15.03)
อื่น ๆ	8.83	9.33	13.02	10.61	11.69	11.00	**47.49**	(15.15)
รวม	12.68	16.72	19.36	17.72	33.52	100.00	**46.67**	(16.34)

หมายเหตุ ** หมายถึง ความแตกต่างของสัดส่วนมีนัยสำคัญทางสถิติที่ระดับ 1%
(ค่าสถิติไคกำลังสองมีค่ามากกว่า 33 และค่าพีมีค่าน้อยกว่า 0.000 ทุกกรณี)



ภาพที่ 1 ร้อยละของประชาชนในช่วงอายุต่าง ๆ จำแนกตามผู้สมัครรับเลือกตั้งผู้ว่า กทม. 5 คนแรกที่ได้มีค่าสูงสุด จากผลการสำรวจของนิด้าโพล

2.4 ผลการศึกษาจำแนกตามเพศ

ผลการสำรวจของนิด้าโพล พบว่า ดร.ชัชชาติ สิทธิพันธุ์ และดร.วิโรจน์ ลักขณาอดิศร ได้รับการสนับสนุนจากผู้ชายในสัดส่วนที่สูงกว่าผู้หญิง ในทางกลับกัน พบว่า พล.ต.อ.อัศวิน ขวัญเมือง ดร.สุชัชวีร์ สุวรรณสวัสดิ์ และนายสกลธี ภัททิยกุล ได้รับการสนับสนุนจากผู้หญิงในสัดส่วนสูงกว่าผู้ชาย อย่างไรก็ตาม พบว่า ความแตกต่างของสัดส่วนดังกล่าว ไม่มีนัยสำคัญทางสถิติ ผู้ชายที่สนับสนุน ดร.ชัชชาติ มีสัดส่วนสูงกว่าผู้หญิงอย่างมีนัยสำคัญที่ระดับ 5% (Z Statistic = 1.827 และ p-Value = 0.034) รายละเอียดแสดงไว้ในตารางที่ 3

ตารางที่ 3 ร้อยละของผู้มีสิทธิ์เลือกตั้งชาย/หญิง และประชาชนในแต่ละระดับการศึกษา จำแนกตามผู้สมัครรับเลือกตั้งผู้ว่าฯ กทม. 5 คนแรกที่ได้มีค่าสูงสุด จากผลการสำรวจของนิด้าโพล

สมัครรับเลือกตั้ง	เพศ		ระดับการศึกษา			รวม
ผู้ว่าฯ กทม.	ชาย	หญิง	< ป.ตรี	ป.ตรี	> ป.ตรี	
ดร.ชัชชาติ	53.52	48.39*	46.89	55.26	48.01	50.72**
พล.ต.อ.อัศวิน	9.33	12.10	13.71	9.27	5.78	10.84**
ดร.สุชัชวีร์	9.15	11.36	10.15	9.63	13.72	10.36
ดร.วิโรจน์	12.06	8.80	11.70	9.36	8.66	10.28
นายสกลธี	5.37	7.99	5.48	7.22	10.47	6.80*
อื่น ๆ	10.56	11.36	12.07	9.27	13.36	11.00
รวม	45.44	54.56	43.76	44.88	11.08	100.00

หมายเหตุ * หมายถึง ความแตกต่างของสัดส่วนมีนัยสำคัญทางสถิติที่ระดับ 5% และ
** หมายถึง ความแตกต่างของสัดส่วนมีนัยสำคัญทางสถิติที่ระดับ 1%

2.5 ผลการศึกษาจำแนกตามระดับการศึกษา

ในด้านการศึกษาของผู้มีสิทธิ์ลงคะแนนเลือกตั้งผู้ว่าฯ กทม. ขอพิจารณาโดยแบ่งระดับการศึกษาออกเป็น 3 ระดับคือ ต่ำกว่าปริญญาตรี ปริญญาตรี และสูงกว่าปริญญาตรี ผลการสำรวจของนิด้าโพล แสดงไว้ในตารางที่ 3 ซึ่งเห็นได้ว่า สัดส่วนของผู้สำเร็จการศึกษาระดับต่ำกว่าปริญญาตรีสนับสนุน พล.ต.อ.อัศวิน ขวัญเมือง และ ดร.วิโรจน์ ลักขณาอดิศร มากกว่าระดับการศึกษาอื่น โดยสัดส่วนดังกล่าวของผู้สมัครทั้ง 2 คนนี้ มีค่าลดลงเมื่อมีระดับการศึกษาสูงขึ้น เมื่อทดสอบ ความแตกต่างของสัดส่วนดังกล่าวโดยใช้การทดสอบไคกำลังสอง พบว่า สัดส่วนของผู้สำเร็จการศึกษาระดับต่าง ๆ ที่สนับสนุน พล.ต.อ.อัศวิน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับสูง (χ^2 = 19.56 และ p-Value = 0.000) แต่ความแตกต่างของสัดส่วนดังกล่าวสำหรับ ดร.วิโรจน์ ไม่มีนัยสำคัญทางสถิติ

สัดส่วนของผู้สำเร็จการศึกษาระดับปริญญาตรีสนับสนุน ดร.ชัชชาติ สิทธิพันธุ์ เป็นสัดส่วนสูงกว่าระดับการศึกษาอื่น กล่าวคือ ผู้สำเร็จการศึกษาระดับปริญญาตรีร้อยละ 55.26 สนับสนุน ดร.ชัชชาติ และผู้สำเร็จการศึกษาระดับปริญญาต่ำกว่าตรี และสูงกว่าปริญญาตรีมีร้อยละ 46.89 และ 48.01 ตามลำดับ ในกรณีนี้ เมื่อทดสอบสมมุติฐานโดยใช้การทดสอบไคกำลังสอง พบว่า สัดส่วนของผู้สำเร็จการศึกษาระดับต่าง ๆ ที่สนับสนุน ดร.ชัชชาติ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับสูง (χ^2 = 16.47 และ p-Value = 0.000)

ในกรณีของ ดร.สุชัชวีร์ สุวรรณสวัสดิ์ และนายสกลธี ภัททิยกุล พบว่า ผู้สำเร็จการศึกษาระดับสูงกว่าปริญญาตรีสนับสนุนผู้สมัครทั้ง 2 คนนี้ในสัดส่วนที่สูงกว่าระดับการศึกษาอื่น โดยสัดส่วนดังกล่าวของผู้สมัครทั้ง 2 คนนี้ มีค่าเพิ่มขึ้นเมื่อมีระดับการศึกษาสูงขึ้น และเมื่อทดสอบ ความแตกต่างของสัดส่วนดังกล่าวโดยใช้การทดสอบไคกำลังสอง พบว่า ความแตกต่างระหว่างสัดส่วนของผู้สำเร็จการศึกษาระดับต่าง ๆ ที่สนับสนุน ดร.สุชัชวีร์ ไม่มีนัยสำคัญทางสถิติ แต่ความแตกต่างของสัดส่วนดังกล่าวสำหรับ นายสกลธี แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 5% (χ^2 = 9.16 และ p-Value = 0.010)

2.6 ผลการศึกษาจำแนกตามระดับรายได้

เมื่อพิจารณาสัดส่วนของผู้มีสิทธิ์เลือกตั้งในแต่ละระดับรายได้ พบว่า สัดส่วนของผู้มีรายได้ระดับปานกลางถึงค่อนข้างสูง (10,001 – 30,000 และ 30,001 – 50,000 บาท/เดือน) ส่วนมาก (ร้อยละ 54.57 และ 55.17 ตามลำดับ) สนับสนุน ดร.ชัชชาติ สิทธิพันธุ์ และผู้มีรายได้ระดับอื่นที่สนับสนุน ดร.ชัชชาติ มีร้อยละ 43 – 46 ซึ่งยังนับว่าเป็นสัดส่วนที่สูงมากเมื่อเทียบกับสัดส่วนที่สนับสนุนผู้สมัครคนอื่น ๆ และเมื่อทดสอบสมมุติฐานเกี่ยวกับสัดส่วนของคะแนนสนับสนุนที่ ดร.ชัชชาติ ได้รับจากผู้มีสิทธิ์เลือกตั้งที่มีรายได้ระดับต่าง ๆ แล้ว พบว่า สัดส่วนดังกล่าว มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับสูง (χ^2 = 20.20 และ p-Value = 0.000)

ในกลุ่มผู้มีรายได้ระดับปานกลางถึงค่อนข้างสูงนี้ พบว่า ร้อยละ 11.19 ของผู้มีรายได้ปานกลาง (10,001 – 30,000 บาท/เดือน) สนับสนุน ดร.วิโรจน์ ลักขณาอดิศร มากเป็นอันดับ 2 รองจาก ดร.ชัชชาติ และร้อยละ 10.34 ของผู้มีรายได้ค่อนข้างสูง (30,001 – 50,000 บาท/เดือน) สนับสนุน ดร.สุชัชวีร์ สุวรรณสวัสดิ์ มากเป็นอันดับ 2 รองจาก ดร.ชัชชาติ นอกจากนี้ยังพบว่า ร้อยละ 14.57 ของผู้มีรายได้สูง (มากกว่า 50,000 บาท/เดือน) สนับสนุน ดร.สุชัชวีร์ นับว่ามากเป็นอันดับ 2 รองจาก ดร.ชัชชาติ ด้วยเช่นกัน ดังรายละเอียดแสดงไว้ในตารางที่ 4 อย่างไรก็ตาม จากการทดสอบสมมุติฐาน พบว่า ความแตกต่างของสัดส่วนของคะแนนสนับสนุนที่ ดร.วิโรจน์ และ ดร.สุชัชวีร์ แต่ละคนได้รับจากผู้มีสิทธิ์เลือกตั้งที่มีรายได้ระดับต่าง ๆ ดังกล่าว ไม่มีนัยสำคัญทางสถิติ (χ^2 = 3.37, 6.18 และ p-Value = 0.498, 0.189 ตามลำดับ)

ตารางที่ 4 ร้อยละของผู้มีสิทธิ์เลือกตั้งในแต่ละกลุ่มรายได้ จำแนกตามผู้สมัครรับเลือกตั้งผู้ว่าฯ กทม. 5 คนแรกที่ได้มีค่าสูงสุด จากผลการสำรวจของนิด้าโพล

ผู้สมัครรับเลือกตั้งผู้ว่าฯ กทม.	ระดับรายได้ (บาท/เดือน)					รวม
	ไม่มีรายได้	≤ 10,000	10,001 – 30,000	30,001 – 50,000	> 50,000	
ดร.ชัชชาติ	45.49	43.18	54.57	55.17	45.70	50.72**
พล.ต.อ.อัศวิน	14.48	19.70	8.83	8.91	8.61	10.84**
ดร.สุชัชวีร์	11.23	9.85	8.72	10.34	14.57	10.36
ดร.วิโรจน์	9.90	9.85	11.19	7.76	9.93	10.28
นายสกลธี	7.68	2.27	6.67	6.61	9.27	6.80
อื่น ๆ	11.23	15.15	10.01	11.21	11.92	11.00
รวม	27.08	5.28	37.16	13.92	6.04	100.00

หมายเหตุ ** หมายถึง ความแตกต่างของสัดส่วนมีนัยสำคัญทางสถิติที่ระดับ 1%

ในส่วนของ พล.ต.อ.อัศวิน ขวัญเมือง พบว่า ร้อยละ 19.70 ของผู้มีรายได้น้อย (ไม่เกิน 10,000 บาท/เดือน) และร้อยละ 14.48 ของผู้ไม่มีมีรายได้ ลงคะแนนเสียงสนับสนุน พล.ต.อ.อัศวิน มากเป็นอันดับ 2 รองจาก ดร.ชัชชาติ เมื่อพิจารณาสัดส่วนของคะแนนสนับสนุนที่ พล.ต.อ.อัศวิน ได้รับจากผู้มีสิทธิ์เลือกตั้งที่มีรายได้ระดับต่าง ๆ แล้ว พบว่า มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับสูง (χ^2 = 25.05 และ p-Value = 0.000)

เมื่อเปรียบเทียบสัดส่วนของผู้มีสิทธิ์เลือกตั้งที่มีรายได้ระดับต่าง ๆ พบว่า ผู้มีรายได้สูง (มากกว่า 50,000 บาท/เดือน) มีสัดส่วนที่สนับสนุนสนับสนุนนายสกลธี ภัททิยกุล มากกว่ากลุ่มระดับรายได้อื่น แต่สัดส่วนดังกล่าวก็ยังไม่เกินร้อยละ 10 (ดูตารางที่ 4) และความแตกต่างของสัดส่วนของคะแนนสนับสนุนที่นายสกลธี ได้รับจากผู้มีสิทธิ์เลือกตั้งที่มีรายได้ระดับต่าง ๆ ดังกล่าว พบว่า ไม่มีนัยสำคัญทางสถิติ (χ^2 = 6.50 และ p-Value = 0.165)

2.7 ผลการศึกษาจำแนกตามอาชีพ

ผลการศึกษาเห็นได้ชัดเจนประการหนึ่ง คือ นักเรียน/นักศึกษามีสัดส่วนสูงถึงร้อยละ 62.72 และ 22.49 ที่สนับสนุน ดร.ชัชชาติ สิทธิพันธุ์ และ ดร.วิโรจน์ ลักขณาอดิศร เป็นอันดับ 1 และ 2 ตามลำดับ และผู้สมัครผู้ว่าฯ กทม. คนอื่น ๆ แต่ละคน ได้เสียงสนับสนุนจากนักเรียน/นักศึกษาไม่ถึงร้อยละ 5 โดยมีรายละเอียดได้ในตารางที่ 5

ตารางที่ 5 ร้อยละของผู้มีสิทธิ์เลือกตั้งในแต่ละกลุ่มอาชีพ จำแนกตามผู้สมัครรับเลือกตั้งผู้ว่าฯ กทม. 5 คนแรกที่ได้มีค่าสูงสุด จากผลการสำรวจของนิด้าโพล

ผู้สมัครรับเลือกตั้งผู้ว่าฯ กทม.	อาชีพ						รวม
	ราชการ/รัฐวิสาหกิจ	พนักงานเอกชน	ส่วนตัว/อิสระ	แรงงาน/เกษตร	แม่บ้าน/ว่างงาน/เกษียณ	นักเรียน/นักศึกษา	
ดร.ชัชชาติ	52.66	58.53	50.32	48.31	40.03	62.72	50.72**
พล.ต.อ.อัศวิน	7.98	6.50	10.00	16.85	17.45	2.37	10.84**
ดร.สุชัชวีร์	9.04	9.10	10.81	7.30	14.17	3.55	10.36**
ดร.วิโรจน์	6.91	11.42	10.81	12.92	5.61	22.49	10.28**
นายสกลธี	5.32	4.48	7.74	5.06	10.90	1.18	6.80**
อื่น ๆ	18.09	9.97	10.32	9.55	11.84	7.69	11.00
รวม	7.52	27.68	24.80	7.12	25.68	6.76	100.00

หมายเหตุ ** หมายถึง ความแตกต่างของสัดส่วนมีนัยสำคัญทางสถิติที่ระดับ 1%

แม้ว่าสัดส่วนของผู้ที่สนับสนุน ดร.ชัชชาติ มีค่าสูงมากทุกกลุ่มอาชีพ แต่มีกลุ่มที่น่าสนใจ คือ กลุ่มแม่บ้าน พ่อบ้าน คนว่างงาน และเกษียณอายุการทำงาน มีร้อยละ 40.03 ที่สนับสนุน ดร.ชัชชาติ แม้จะมีค่าสูง แต่ก็นับเป็นสัดส่วนที่ต่ำที่สุดเมื่อเทียบกับกลุ่มอื่น ๆ ไม่ว่าจะเป็นกลุ่มหญิง/ชาย กลุ่มอายุ กลุ่มระดับการศึกษา กลุ่มรายได้ หรือกลุ่มอาชีพอื่น ๆ ในทางกลับกัน ในบรรดากลุ่มต่าง ๆ ที่สนับสนุน นายสกลธี ภัททิยกุล กลุ่มอาชีพนี้มีมากถึงร้อยละ 10.90 ซึ่งนับว่าสูงกว่ากลุ่มอื่น ๆ ทุกกลุ่ม

ในส่วนที่เกี่ยวกับ พล.ต.อ.อัศวิน ขวัญเมือง พบว่า ร้อยละ 16.85 และ 17.45 ของกลุ่มแรงงาน/เกษตร และกลุ่มแม่บ้าน/ว่างงาน/เกษียณ ตามลำดับ สนับสนุนพล.ต.อ.อัศวิน นับเป็นสัดส่วนสูงเป็นอันดับ 2 รองจาก ดร.ชัชชาติ เมื่อทดสอบสมมุติฐานเพื่อเปรียบเทียบสัดส่วนของกลุ่มอาชีพต่าง ๆ ที่ผู้สมัครแต่ละคนได้รับการสนับสนุน พบว่า มีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับสูงทุกกรณี (χ^2 > 22 และ p-Value > 0.001)

3. สรุปและอภิปรายผล

การศึกษานี้ใช้ข้อมูลตัวอย่างของผู้มีสิทธิ์ลงคะแนนเสียงเลือกตั้งผู้ว่าราชการกรุงเทพมหานคร ขนาด 2,500 คน ซึ่งสุ่มมาจากตัวอย่างหลัก (Master Sample) ของศูนย์สำรวจความคิดเห็น “นิด้าโพล” สถาบันบัณฑิตพัฒนบริหารศาสตร์ ผลการศึกษาในภาพรวม พบว่า ผลการสำรวจของนิด้าโพลค่อนข้างใกล้เคียงกับผลการเลือกตั้งผู้ว่าฯ กทม. เมื่อวันที่ 22 พฤษภาคม 2565 ทั้งในด้านผู้สมัคร 5 คนแรกที่ได้รับคะแนนเสียงสนับสนุนสูงสุด และสัดส่วน (ร้อยละ) ของคะแนนที่แต่ละคนได้รับ

ผลของการศึกษาสำหรับผู้สมัคร 5 คนที่แรกที่ได้รับคะแนนสนับสนุนสูงสุด สรุปเป็นรายบุคคลได้ ดังนี้

ดร.ชัชชาติ สิทธิพันธุ์ ได้รับคะแนนเสียงสนับสนุนสูงมากจากทุกกลุ่ม ไม่ว่าจะเป็นกลุ่มหญิง/ชาย กลุ่มอายุ กลุ่มระดับการศึกษา กลุ่มรายได้ หรือกลุ่มอาชีพ โดยเฉพาะอย่างยิ่ง กลุ่มอายุไม่เกิน 35 ปี และกลุ่มนักเรียน/นักศึกษา มีมากกว่าร้อยละ 60 ที่สนับสนุน ดร.ชัชชาติ ในทางตรงกันข้าม กลุ่มผู้มีอายุมากกว่า 55 ปี และกลุ่มแม่บ้าน/ว่างงาน/เกษียณ มีน้อยกว่าร้อยละ 45 ที่สนับสนุน ดร.ชัชชาติ

พล.ต.อ.อัศวิน ขวัญเมือง ได้รับการสนับสนุนจากกลุ่มต่าง ๆ อยู่ระหว่างร้อยละ 2 – 20 โดยประมาณ กลุ่มที่สนับสนุนมากกว่าร้อยละ 15 มี 4 กลุ่ม ได้แก่ กลุ่มอายุมากกว่า 55 ปี กลุ่มเงินเดือนไม่เกิน 10,000 บาท/เดือน กลุ่มอาชีพแรงงาน/เกษตร และกลุ่มแม่บ้าน/ว่างงาน/เกษียณ ส่วนกลุ่มที่สนับสนุนน้อยกว่าร้อยละ 5 มี คือ กลุ่มอายุ 18 – 25 ปี เห็นได้ว่า พล.ต.อ.อัศวิน ได้รับการสนับสนุนจากผู้สูงอายุ และผู้มีฐานะทางเศรษฐกิจ/สังคมระดับล่าง

ดร.สุชัชวีร์ สุวรรณสวัสดิ์ ได้รับการสนับสนุนจากกลุ่มต่าง ๆ อยู่ระหว่างร้อยละ 3 – 16 โดยประมาณ กลุ่มที่สนับสนุนมากกว่าร้อยละ 15 มี ได้แก่ กลุ่มอายุ 46 – 55 ปี ส่วนกลุ่มที่สนับสนุนน้อยกว่าร้อยละ 5 มี คือ กลุ่มอายุไม่เกิน 35 ปี เห็นได้ว่า ดร.สุชัชวีร์ ได้รับการสนับสนุนจากคนวัยทำงานช่วงกลางถึงปลาย และที่น่าสังเกตคือ ผู้ที่มีการศึกษาระดับสูงกว่าปริญญาตรี สนับสนุน ดร.สุชัชวีร์ เป็นอันดับ 2 รองจาก ดร.ชัชชาติ

ดร.วิโรจน์ ลักขณาอดิศร ได้รับการสนับสนุนจากกลุ่มต่าง ๆ อยู่ระหว่างร้อยละ 5.5 – 22.5 โดยประมาณ กลุ่มที่สนับสนุนมากกว่าร้อยละ 15 มี 4 กลุ่ม ได้แก่ กลุ่มอายุไม่เกิน 35 ปี และกลุ่มนักเรียน/นักศึกษา นอกจากนี้ ไม่พบว่ามีกลุ่มใดที่สนับสนุนน้อยกว่าร้อยละ 5 มี เห็นได้ชัดเจนว่า ดร.วิโรจน์ ได้รับการสนับสนุนจากคนรุ่นใหม่

ในการศึกษาครั้งนี้ นายสกลธี ภัททิยกุล ได้รับการสนับสนุนอยู่ในอันดับที่ 5 (ผลการเลือกตั้งอยู่อันดับที่ 4) โดยได้รับการสนับสนุนจากกลุ่มต่าง ๆ ไม่เกินร้อยละ 11 และมีหลายกลุ่มที่สนับสนุนน้อยกว่าร้อยละ 5 ได้แก่ กลุ่มอายุไม่เกิน 35 ปี กลุ่มรายได้ไม่เกิน 10,000 บาท/เดือน กลุ่มพนักงานบริษัทเอกชน และกลุ่มนักเรียน/นักศึกษา เห็นได้ชัดว่า นายสกลธี ไม่ได้อยู่ในใจของคนรุ่นใหม่ แต่มีแนวโน้มว่าจะได้รับการสนับสนุนจากกลุ่มผู้มีอายุมากกว่า 55 ปี มีการศึกษาระดับสูงกว่าปริญญาตรี และกลุ่มแม่บ้าน/ว่างงาน/เกษียณ เพราะมีการสนับสนุนจากแต่ละกลุ่มดังกล่าวมากกว่าร้อยละ 10

แม้นโยบายที่ผู้สมัครแต่ละคนประกาศไว้ในช่วงหาเสียงเป็นสิ่งสำคัญที่ทำให้ผู้สมัครได้/ไม่ได้รับการสนับสนุนจากประชาชน แต่ขอบเขตของการศึกษาครั้งนี้ไม่ได้ครอบคลุมในเรื่องดังกล่าวโดยตรง แต่นโยบายที่มากมายและครอบคลุมแทบทุกเรื่อง ตลอดจนการเปิดตัวเร็ว มีระยะเวลาเผยแพร่นโยบายให้ประชาชนได้รับทราบเร็วและทั่วถึงก็มีผลทำให้เป็นที่รู้จักอย่างกว้างขวาง สิ่งต่าง ๆ เหล่านี้ย่อมเป็นผลดีแก่ผู้สมัคร

สิ่งสำคัญอีกประการหนึ่ง คือ ภาพลักษณ์หรือภูมิหลัง รวมทั้งชื่อเสียงหรือสังกัด ของผู้สมัครเป็นสิ่งที่ช่วยเรียกคะแนนสียงสำหรับคนบางกลุ่ม ในขณะเดียวกันก็อาจเป็นสิ่งที่ฉุดรั้งสำหรับคนอีกกลุ่มหนึ่ง ผู้สมัครที่สามารถลบภาพที่ไม่พึงปรารถนาได้ ก็จะสามารถเรียกคะแนนกลับคืนได้ จากผลการศึกษานี้ เห็นได้ว่า สัดส่วนของคนบางอาชีพ หรือบางระดับการศึกษา สนับสนุนผู้สมัครบางคนมากกว่า

จากข้อมูลสถิติประชากรทางการทะเบียนราษฎร สำนักบริหารการทะเบียน กรมการปกครอง กระทรวงมหาดไทย พบว่า ในการเลือกตั้งผู้ว่าฯ กทม. ครั้งนี้ผู้มีสิทธิ์เลือกตั้งที่มีอายุ 18 – 27 ปี มีร้อยละ 15.56 เมื่อรวมกับผู้มีสิทธิ์ที่มีอายุ 28 – 40 ปี อีกร้อยละ 22.59 รวมเป็นร้อยละ 38.15 ของผู้มีสิทธิ์เลือกตั้งทั้งหมด ดังรายละเอียดแสดงไว้ในภาพที่ 2 เห็นได้ว่า นโยบายและภาพลักษณ์/ภูมิหลังที่ “โดนใจ” คนกลุ่มนี้ ก็สามาถเพิ่มคะแนนเสียงได้พอสมควร ซึ่งสอดคล้องกับผลการศึกษาในครั้งนี้ กล่าวคือ ผู้สมัครบางคนมีสัดส่วนของผู้มีอายุ 18 – 25 ปี หรือสัดส่วนนักเรียน/นักศึกษา ที่สนับสนุนสูงกว่าช่วงอายุอื่น ๆ และน่าจะทำให้ได้คะแนนเสียงสนับสนุนมากในการเลือกตั้งจริง

ภาพที่ 2 จำนวนผู้มีสิทธิ์เลือกตั้งผู้ว่าฯ กทม. พ.ศ. 2565

ที่มา : https://www.bangkokbiznews.com/politics/1001230

ค้นเมื่อ 27 พฤษภาคม 2565

4. กิตติกรรมประกาศ

ขอบคุณศูนย์สำรวจความคิดเห็น “นิด้าโพล” ที่ยินยอมให้ใช้ข้อมูล และขอขอบคุณบุคลากรของศูนย์ฯ ที่ช่วยประสานงานและอำนวยความสะดวกในการจัดส่งข้อมูลและสารสนเทศต่าง ๆ อย่างรวดเร็วและครบถ้วน

5. เอกสารอ้างอิง

ไทยโพสต์. 2022. **สรุปผลการเลือกตั้ง 'ส.ก.-ผู้ว่าฯกทม.' อย่างไม่เป็นทางการ.** สืบค้น 25 พฤษภาคม 2565. จาก https://www.thaipost.net/hi-light/146948.

กรุงเทพธุรกิจ. 2022. **จำนวนผู้มีสิทธิ์เลือกตั้งผู้ว่าฯ กทม. พ.ศ. 2565**. สืบค้น 27 พฤษภาคม 2565. จาก https://www.bangkokbiznews.com/politics/1001230.

นิด้าโพล.2022. **22 พ.ค. 2565 คนกรุงเทพฯ เลือกใครเป็นผู้ว่าฯ กทม**. สืบค้น 25 พฤษภาคม 2565 จาก https://nidapoll.nida.ac.th/ survey_detail?survey_id=573

Banerjee, Abhijit V., Selvan Kumar, Rohini Pande and Felix Su. 2011. **Do Informed Voters Make Better Choices? Experimental Evidence from Urban India**. Unpublished manuscript.

Bartels, Larry M. 1996. Uninformed Votes: Information Effects in Presidential Elections. **American Journal of Political Science**. 40(1): 194-230.

Crowder-Meyer, Melody, Shana K. Gadarian and Jessica Trounstine. 2020. Voting Can Be Hard, Information Helps. **Urban Affairs Review**. 56(1): 124-153.

Ebeid, Michael and Jonathan Rodden. 2006. Economic Geography and Economic Voting: Evidence from the US States. **British Journal of Political science**. 36: 527-547.

Levernier, William and Anthony G. Barilla. 2006. The Effect of Regional, Demographics, and Economic Characteristics on County-Level Voting Patterns in 2000 Presidential Election. **The Review of Regional Studies**. 36(3): 427-447.

Matsusaka, John G. 1995. Explaining voter turnout patterns: An information theory. **Public Choice**. 84: 91-117.

McMurray, Joseph. 2015. The paradox of information and voter turnout. **Public Choice**. 165: 13-23.

Ryan, John B. 2010. The Effects of Network Expertise and Biases on Vote Choice. **Political Communication**. 27(1): 44-58.

Wolfinger, Raymond E. and Stephen J. Rosenstone. 1980. **Who Votes?**. New Haven: Yale University Press.

การพัฒนาแอปพลิเคชันบนอุปกรณ์เคลื่อนที่ระบบคลาวด์ – เนทีฟ สำหรับการเพิ่มประสิทธิภาพและจัดการสินค้าคงคลัง

สุนันท์ ธาติ[1], อลงกรณ์ เมืองไหว[2], ธณิดา โขนงนุช[3] ,วิชิต ธรรมพิทักษ์[4] และ เอกชัย สัจจอังกูร[5]

Received: 10 November 2020; Revised: 20 October 2021; Accepted: 7 March 2022

บทคัดย่อ

ปัจจุบันการจัดการโซ่อุปทาน (Supply Chain Management) ถูกขับเคลื่อนด้วยข้อมูลแทนการใช้ประสบการณ์หรือการคาดการณ์จากเหตุการณ์ เนื่องด้วยความต้องการของลูกค้าเปลี่ยนแปลงอยู่เสมอซึ่งค่อนข้างยากที่สถานประกอบการจะสามารถระบุปริมาณความต้องการเพื่อกำหนดทิศทางในการผลิต การประยุกต์ใช้เทคโนโลยีเพื่อช่วยให้ผู้ประกอบการทราบปริมาณความต้องการและนำไปสู่การวางแผนการผลิตที่มีประสิทธิภาพนั้นมีประโยชน์ทั้งในแง่ของคุณภาพของผลิตภัณฑ์และการควบคุมปริมาณการผลิต ผู้วิจัยศึกษากระบวนการจัดการสินค้าคงคลังจากกรณีศึกษาของบริษัท ศรีแก้วหล่มเก่า จำกัด พบข้อจำกัดหลายประการเนื่องด้วยการบันทึกข้อมูลลงในสมุดหรือโปรแกรมสำเร็จรูปพื้นฐานก่อให้เกิดความยุ่งยากและซ้ำซ้อนในการทำงานอย่างมาก ทั้งนี้เพื่อลดข้อผิดพลาด และเพิ่มประสิทธิภาพในกระบวนการผลิตและจัดการสินค้าคงคลังและช่วยให้บริษัทสามารถวิเคราะห์ความต้องการของลูกค้าเพื่อวางแผนการผลิตได้อย่างมีประสิทธิภาพมากขึ้น ผู้วิจัยจึงพัฒนาแอปพลิเคชันบนอุปกรณ์เคลื่อนที่ระบบคลาวด์ – เนทีฟ สำหรับติดตามสินค้าคงคลังเพื่อเก็บข้อมูลการคงอยู่ของสินค้าแบบเวลาจริง (Real Time) ลงในฐานข้อมูลและนำร่องใช้งานกับสถานประกอบการและพบว่าแอปพลิเคชันที่พัฒนาขึ้นสามารถช่วยลดต้นทุนแรงงานถึงร้อยละ 98.34 และลดเวลาในการทำงานคิดเป็นร้อยละ 98.33

คำสำคัญ: การจัดการโซ่อุปทาน, การจัดการสินค้าคงคลัง, แอปพลิเคชันบนอุปกรณ์เคลื่อนที่, คลาวด์ – เนทีฟ

*Corresponding email: *sununt@nu.ac.th*

[1] *คณะโลจิสติกส์และดิจิทัลซัพพลายเชน มหาวิทยาลัยนเรศวร*

[2,3] *สาขาวิชาวิศวกรรมโลจิสติกส์ คณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏพิบูลสงคราม*

[4,5] *ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร*

Cloud-Native Mobile Application Development for Optimization and Inventory Management

Sunun Tati[1], Alongkorn Muangwai[2], Thanida Khanongnuch[3], Wichit Thampitak[4] and Akachai Sajjaanggoon[5]

Received: 10 November 2020; Revised: 20 October 2021; Accepted: 7 March 2022

ABSTRACT

Supply chain management is now based on data instead of experience forecasting. Due to constantly changing customer demands, it is quite difficult for an establishment to determine the amount of demand and set the direction and execution of production activities. Technology can help operators know the quantity of demand and production planning. Effective production planning is beneficial in terms of both product quality and production quantity control. The researcher studied the inventory management process from the business case study Sri Kaew Lom Kao Co., Ltd. We found limitations due to the traditional process of data collection Data is collect to a book or program was very difficult and redundant in the job. To reduce errors and increase the efficiency of production processes, we developed application to manage inventory and help companies analyze customer needs to plan production more efficiently. A developed cloud-native mobile application store real-time inventory data in a database. The developed application reduced labor costs by 98.34% and working time by 98.33%

Keyword: Supply Chain Management, Inventory Management, Mobile Application, Cloud-Native

*Corresponding email: *sununt@nu.ac.th*

[1] *Faculty of Logistics and Digital Supply Chain, Naresuan University*

[2,3] *Department of Logistics Engineering, Faculty of Industrial Technology, Pibulsongkram Rajabhat University*

[4,5] *Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University*

1. บทนำ

การแข่งขันทางธุรกิจในปัจจุบันเป็นการวางแผน กลยุทธ์ที่สร้างความได้เปรียบให้กับธุรกิจของตน และกลยุทธ์สำคัญที่อุตสาหกรรมไม่อาจปฏิเสธได้ในยุคนี้ คือการจัดการโซ่อุปทาน ซึ่งการจัดการโซ่อุปทานดังกล่าวได้เข้ามามีบทบาทสำคัญที่จะขับเคลื่อนธุรกิจให้ประสบความสำเร็จ หรือเรียกอีกอย่างหนึ่งว่าธุรกิจต้องการพัฒนารูปแบบในการผลิตสินค้า หรือการบริการที่เป็นรูปแบบใหม่เพื่อก้าวให้ทันต่อการเปลี่ยนแปลงของโลก ตลอดจนการก้าวกระโดดของความทันสมัยในด้านเทคโนโลยีที่รวดเร็ว ก็เป็นอีกปัจจัยหนึ่งที่ทำให้ธุรกิจต้องให้ความสำคัญเป็นอย่างมาก เนื่องจากเทคโนโลยีเหล่านี้จะช่วยอำนวยความสะดวกในการบริหารจัดการงานด้านต่าง ๆ ให้กับธุรกิจได้เป็นอย่างดี และส่งผลทำให้เกิดการลดต้นทุนโดยรวมของธุรกิจในระยะยาว ยกตัวอย่างเช่น การวางแผนการผลิตจากการวิเคราะห์ข้อมูลการขาย อุตสาหกรรมขนาดใหญ่ใช้การวิเคราะห์ข้อมูลแทนการการคาดเดาในการตัดสินใจระหว่างกระบวนการผลิต เป็นต้น โดยทั่วไปแล้วโซ่อุปทานจะเป็นงานด้าน การบริหารการดำเนินงานทั้งระบบในภาพรวม และยังมองถึงรายละเอียดด้านการผลิตด้วย เช่น การจัดซื้อ การควบคุมสินค้าคงคลังและคลังสินค้า การขนส่ง การเคลื่อนย้ายวัสดุการบรรจุภัณฑ์ การติดต่อสื่อสารกับผู้ให้การสนับสนุน เป็นต้น (Bowersox et al., 2007) ซึ่งในงานวิจัยนี้จะมุ่งเน้นที่การจัดการสินค้าคงคลังและคลังสินค้าให้กับธุรกิจกรณีศึกษา เพื่อให้เกิดประสิทธิภาพในการบริหารงานที่เหมาะสมที่สุด นอกจากนี้ยังเป็นการช่วยยกระดับให้ธุรกิจมีการตอบสนองต่อลูกค้าที่ถูกต้องแม่นยำและทำให้ภาพลักษณ์ของธุรกิจดีขึ้นจนทำให้ลูกค้าเกิดความเชื่อมั่นในองค์กร ตลอดจนเกิดความพึงพอใจในการใช้บริการเพิ่มมากขึ้น

โครงการปฏิรูปอุตสาหกรรมด้วยนวัตกรรมและเทคโนโลยี (INNO-TECH Transformation) กิจกรรมการยกระดับและพัฒนาผลิตภัณฑ์ด้านต่างๆ หรือประยุกต์ใช้งานวิจัยของสถาบันการศึกษาเครือข่ายภายในพื้นที่เพื่อพัฒนาสินค้าต้นแบบของสถานประกอบการ บริษัท ศรีแก้วหล่มเก่า จำกัด 20 หมู่ 11 ต.วังบาล อ.หล่มเก่า จ.เพชรบูรณ์ ทางสถานประกอบการดำเนินการผลิต และจัดจำหน่ายผลิตภัณฑ์แปรรูปจากมะขาม โดยแบ่งผลิตภัณฑ์แปรรูปหลายชนิด อาทิเช่น มะขามคลุกน้ำตาล มะขามจี๊ดจ้าด มะขามกวน มะขามอบน้ำผึ้ง มะขามแก้ว เป็นต้น โดยเริ่มดำเนินกิจการตั้งแต่ปี พ.ศ. 2558 โดยเป็นบริษัทผลิตผลิตภัณฑ์แปรรูปจากมะขามทั้งขายปลีกหน้าร้านและขายส่งให้กับร้านค้า ผลิตภัณฑ์ของสถานประกอบการมีด้วยกันหลายชนิดอาทิ มะขามจี๊ดจ้าด มะขามแช่อิ่มอบน้ำผึ้ง มะขามคลุกไร้เมล็ด มะขามแก้วรสบ๊วย มะขามกวนตัด มะขามแช่อิ่มน้ำผึ้ง เป็นต้น ผลิตภัณฑ์ถูกจัดจำหน่ายภายใต้แบรนด์ของตนเอง และขอรับการรับรองระบบคุณภาพเพื่อยกระดับคุณภาพของผลิตภัณฑ์ของตนได้แก่ มาตรฐาน GMP และ อย. โดยทางสถานประกบการยังมีการพัฒนาทั้งระบบการผลิตและผลิตภัณฑ์อย่างต่อเนื่อง เพื่อตอบสนองให้กับลูกค้าที่เพิ่มมากขึ้น อีกทั้งยังเพิ่มช่องทางการขายโดยใช้ระบบเว็บไซต์อีกทางหนึ่งด้วย

การจัดการคลังสินค้าที่มีความคลาดเคลื่อนกับข้อมูลจริงและการทำงานมีหลายขั้นตอนที่ซ้ำซ้อนและไม่ตรงกันระหว่างฝ่ายผลิตและฝ่ายคลังสินค้า ไม่สามารถตรวจเช็คจำนวนที่แน่นอนของยอดขายในแต่ละวันที่ตรงกับจำนวนของที่อยู่ในคลัง ดังนั้นการปรับปรุงและพัฒนาการจัดการคลังสินค้า (สินค้าคงคลัง) ในคลังสินค้าให้มีประสิทธิภาพมากขึ้นจึงมีความสำคัญอย่างมาก แอปพลิเคชันเบื้องต้นในการจัดการและบริหารคลังสินค้าให้เหมาะสมกับการใช้งานในสภาพปัจจุบันเพื่อจัดการและบริหารคลังสินค้ามีประสิทธิภาพมากขึ้น สำหรับกรณีศึกษาบริษัท ศรีแก้วหล่มเก่า จำกัดนั้นผู้วิจัยพบข้อจำกัดหลายประการ เนื่องด้วยการบันทึกข้อมูลลงในสมุดหรือการบันทึกลงโปรแกรม Microsoft Excel ก่อให้เกิดความยุ่งยากและซ้ำซ้อนในการทำงานอย่างมาก บริษัทต้องการจะนำเทคโนโลยีเข้ามาช่วยในการทำงานเพื่อลด ข้อผิดพลาด และเพิ่มประสิทธิภาพในกระบวนการผลิตและจัดการสินค้าคงคลังที่มีหลายประเภทและมีจำนวนมาก อีกทั้งเพื่อให้บริษัทสามารถวิเคราะห์ความต้องการของลูกค้าเพื่อวางแผนการผลิตได้อย่างมีประสิทธิภาพมากขึ้น

2. การทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง

ปัจจุบันการจัดการโซ่อุปทาน (Supply Chain Management) ถูกขับเคลื่อนด้วยข้อมูลแทนการใช้ประสบการณ์หรือการคาดการณ์จากเหตุการณ์ ทั้งนี้ก็เพื่อเพิ่มประสิทธิภาพการบริหารจัดการ (Li & Liu, 2019) ประสบการณ์ที่น่า

ผิดหวังของผู้ซื้อคือการมีความต้องการต่อสินค้าชิ้นใดชิ้นหนึ่งแต่ไม่ทราบข้อมูลที่เกี่ยวข้องกับความสามารถในการสั่งซื้อ อาทิ ข้อมูลการคงอยู่ของสินค้า ข้อมูลจำนวนแยกตามลักษณะเฉพาะ (“Big Data in Logistics-A DHL Perspective on how to move beyond the hype”, 2013)

ความต้องการของลูกค้าเปลี่ยนแปลงอยู่เสมอซึ่งค่อนข้างยากที่สถานประกอบการจะสามารถระบุปริมาณความต้องการเพื่อกำหนดทิศทางในการผลิต (Tahiduzzaman et al., 2017) การประยุกต์ใช้ฐานข้อมูลขนาดใหญ่สำหรับการจัดการโซ่อุปทาน (Supply Chain Management) ช่วยให้ผู้ประกอบการทราบปริมาณความต้องการซึ่งนำไปสู่การวางแผนการผลิตที่ดีทั้งในแง่ของคุณภาพและปริมาณ การรวบรวมข้อมูลที่มีขนาดใหญ่และซับซ้อนช่วยให้ผู้ประกอบการทราบข้อมูลจริงในเวลาปัจจุบัน การเก็บข้อมูลในระบบฐานข้อมูลแบบเวลาจริง (Real Time) ช่วยให้สามารถลดต้นทุนการผลิตลง ลดเวลาในการทำงาน อีกทั้งเพิ่มความพร้อมของสินค้าเพื่อการส่งต่อสู่ตลาด

การพัฒนาระบบสารสนเทศ (Information System Development) อาทิ ระบบสารสนเทศเพื่อการบริหารจัดการ ระบบสนับสนุนการตัดสินใจ ระบบประมวลผลข้อมูล หรือระบบผู้เชี่ยวชาญ เป็นต้น ทั้งนี้ซึ่งระบบสารสนเทศแต่ละชนิดจะมีความแตกต่างกัน แต่ไม่ว่าระบบสารสนเทศชนิดใด องค์ประกอบสำคัญของระบบสารสนเทศคือฐานข้อมูล การออกแบบฐานข้อมูลสำหรับระบบสารสนเทศจึงถือเป็นปัจจัยสำคัญต่อการบรรลุตามวัตถุประสงค์ของการพัฒนาระบบเพื่อตอบโจทย์ผู้ใช้งาน (Rob & Semaan, 2000; Toj, 2011; Keng, 2007) การออกแบบฐานข้อมูลมี 7 ขั้นตอน ดังต่อไปนี้

(1) วางแผน (Planning) เป็นขั้นตอนเพื่อระบุจุดประสงค์และเป้าหมายของงานโดยการวางแผนนี้จะพิจารณาหลายประเด็น อาทิ โครงสร้างการเก็บข้อมูล, บูรณภาพของข้อมูล (Consistency), การเพิ่ม - ขยาย และความเปลี่ยนแปลงของข้อมูล

(2) วิเคราะห์ความต้องการ (Data and Requirement Analysis) เป็นขั้นตอนการรวบรวมข้อมูลความต้องการผู้ใช้งานซึ่งวิธีการเก็บข้อมูลนั้นสามารถทำได้หลายวิธี อาทิ การสัมภาษณ์ผู้ที่จะใช้ระบบ ศึกษาระบบสารสนเทศที่เกี่ยวข้อง การรวบรวมเอกสารกระดาษที่เกี่ยวกับการพัฒนาระบบ เป็นต้น ซึ่งการเก็บข้อมูลความต้องการของผู้ใช้ต้องคำนึงถึงนโยบายและกฎระเบียบขององค์กรด้วย ผู้พัฒนาระบบจะสำรวจหาข้อมูลในประเด็นต่าง ๆ เกี่ยวกับระบบงาน ได้แก่ ปัญหาที่เกิดขึ้นในปัจจุบัน ความเป็นไปได้ของการพัฒนาระบบที่ต้องการ สิ่งที่จะช่วยเพิ่มประสิทธิภาพของกลยุทธ์ในการดำเนินงาน และประมาณการของค่าใช้จ่ายที่ต้องใช้

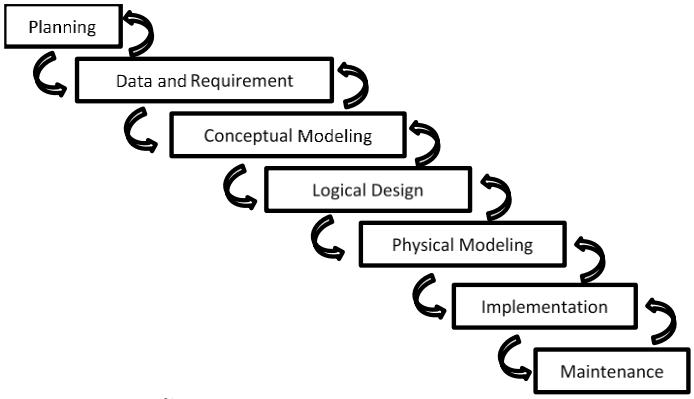
(3) การออกแบบแนวคิด (Conceptual Modeling) เป็นขั้นตอนการออกแบบเพื่อร่างแผนภูมิเพื่อการออกแบบโครงสร้างฐานข้อมูลประกอบด้วยตารางคอลัมภ์ ความสัมพันธ์ รวมถึงการทำนอร์มัลไลซ์ตารางข้อมูลด้วย ซึ่งแผนภูมิที่นิยมใช้ในการออกแบบที่สุดคือ แผนภูมิความสัมพันธ์ (Entity-Relationship Diagram)

(4) การออกแบบเชิงตรรกะ (Logical Design) เป็นขั้นตอนการเขียน DDL (Data Definition Language) เพื่อสร้างตารางในฐานข้อมูลตาม ER-Diagram ที่ออกแบบไว้

(5) การออกแบบทางกายภาพ (Physical Modeling) เป็นขั้นตอนการปรับคิวรี่เพื่อสนับสนุนลักษณะทางกายภาพ บางอย่าง อาทิ การพิจารณาความจำเป็นที่จะต้องเก็บอ็อปเจ็กต์ (Object) ที่เป็นไบนารีขนาดใหญ่ไว้เป็นไฟล์แยกต่างหากจากตารางมาตรฐาน

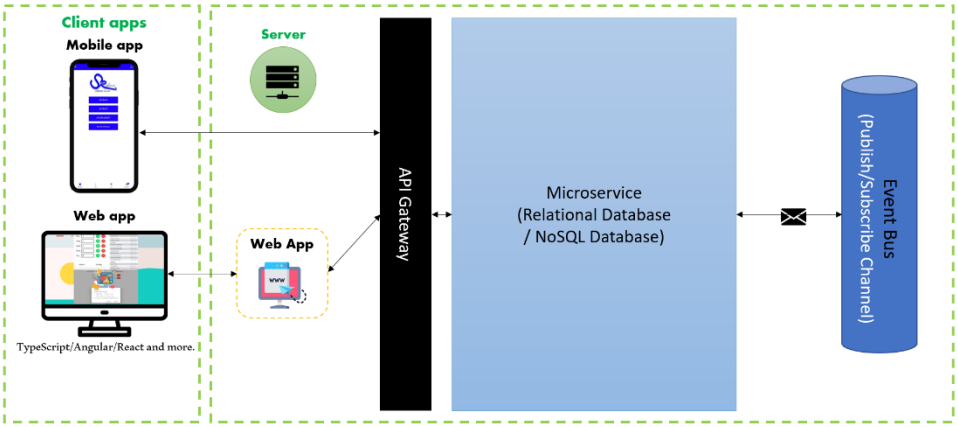
(6) การสร้างระบบ (Implementation) เป็นขั้นตอนเขียนโปรแกรมและชุดคำสั่งจัดการฐานข้อมูล

(7) การดูแลรักษาระบบ (Maintenance) เป็นขั้นตอนการดูแลรักษาระบบหลังจากการสร้าง ฐานข้อมูลเรียบร้อยแล้ว



รูปที่ 1 ขั้นตอนการออกแบบฐานข้อมูลสำหรับระบบสารสนเทศ

แอปพลิเคชันคลาวด์ – เนทีฟ (Cloud Native Application) (Arora et al., 2019) เป็นโปรแกรมที่ออกแบบมาสำหรับสถาปัตยกรรมการประมวลผลแบบคลาวด์ แอปพลิเคชันเหล่านี้ทำงานบนในระบบจัดเก็บข้อมูลคลาวด์ และได้รับการออกแบบเพื่อใช้ประโยชน์จากคุณลักษณะของคลาวด์คอมพิวติ้ง แอปพลิเคชันคลาวด์ – เนทีฟมีสถาปัตยกรรมแบบไมโครเซอร์วิสที่จัดสรรทรัพยากรอย่างมีประสิทธิภาพให้กับแต่ละบริการที่แอปพลิเคชันใช้ ทำให้แอปพลิเคชันมีความยืดหยุ่นและปรับให้เข้ากับระบบจัดเก็บข้อมูลคลาวด์ได้



รูปที่ 2 แอปพลิเคชันคลาวด์ – เนทีฟ (Cloud Native Application)

แอปพลิเคชันคลาวด์ – เนทีฟ (Cloud Native Application) ได้รับการออกแบบมาเพื่อใช้ประโยชน์จากความเร็วและประสิทธิภาพของระบบคลาวด์

- ประหยัดเวลา ทรัพยากร และเงิน เนื่องจากทรัพยากรการประมวลผลและการจัดเก็บสามารถปรับขนาดได้ตามต้องการ สามารถเพิ่มขนาดเซิร์ฟเวอร์เสมือนได้ การทดสอบและแอปพลิเคชันบนคลาวด์สามารถเปิดใช้งานได้อย่างรวดเร็ว คอนเทนเนอร์ยังสามารถใช้เพื่อเพิ่มจำนวนไมโครเซอร์วิสที่ทำงานบนโฮสต์ ไมโครเซอร์วิสสามารถปรับขนาดได้อิสระ แต่ละไมโครเซอร์วิสถูกแยกออกจากกัน หากไมโครเซอร์วิสหนึ่งถูกปรับขนาด ไมโครเซอร์วิสอื่นๆ จะไม่ได้รับผลกระทบ
- เชื่อถือได้ หากเกิดความล้มเหลวในไมโครเซอร์วิส จะไม่มีผลกระทบต่อบริการที่อยู่ติดกัน เนื่องจากแอปพลิเคชันบนคลาวด์เหล่านี้ใช้คอนเทนเนอร์
- ง่ายต่อการจัดการ แอปที่มาพร้อมระบบคลาวด์ใช้ระบบอัตโนมัติในการปรับใช้ฟีเจอร์และการอัปเดตของแอป นักพัฒนาสามารถติดตามไมโครเซอร์วิสและส่วนประกอบทั้งหมดในขณะที่อัปเดต เนื่องจากแอปพลิเคชันแบ่ง

ออกเป็นส่วยย่อยเล็กๆ นักพัฒนาจึงสามารถจัดการที่ไมโครเซอร์วิสเฉพาะและการแก้ไขหรืออัปเดตไม่มีผลกระทบกับไมโครเซอร์วิสอื่นๆ

- อิสระจากผู้ขายและใช้คอนเทนเนอร์เพื่อย้ายไมโครเซอร์วิสระหว่างโครงสร้างพื้นฐานของผู้ขายที่แตกต่างกัน ช่วยหลีกเลี่ยงการล็อคอินของผู้ขาย

3. การศึกษากระบวนการจัดการคลังสินค้ากรณีศึกษาเพื่อนำมาออกแบบแอปพลิเคชัน

กระบวนการจัดการคลังสินค้าเดิมมีขั้นตอนดังต่อไปนี้

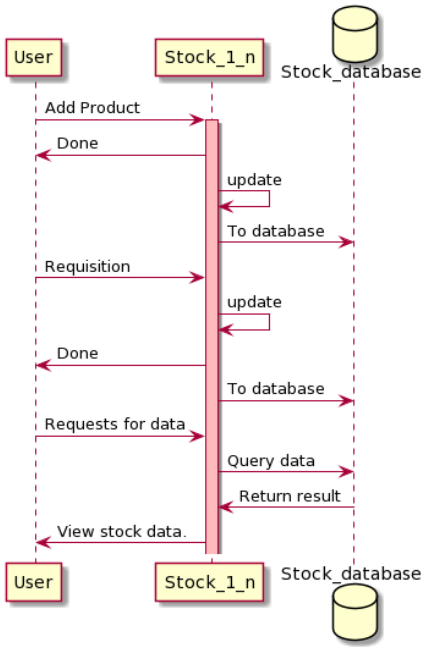
1) พนักงานรับคำสั่งซื้อจากลูกค้าทางโทรศัพท์ และจดรายการสินค้าที่ลูกค้าสั่งลงในกระดาษ
2) พนักงานเดินเข้าไปตรวจเช็คสินค้าตามที่ลูกค้าสั่งทั้ง 3 คลัง โดยใช้เวลาประมาณ 5 นาที
3) พนักงานโทรศัพท์แจ้งสถานะการมีอยู่ของสินค้าที่สั่ง (มีอยู่/ไม่มีอยู่ในคลังสินค้า) และความพร้อมในการจัดส่งแก่ลูกค้า
4) ถ้าสินค้ามีพร้อมส่ง พนักงานจะปริ้นใบสั่งซื้อเพื่อส่งให้แผนกจัดสินค้าเตรียมส่งต่อไป แต่ถ้าสินค้าไม่เพียงพอสำหรับคำสั่งซื้อนั้น ก็จะทำการสั่งผลิตและโทรศัพท์กลับไปแจ้งวันที่พร้อมส่งสินค้าให้ลูกค้าทราบอีกครั้ง

4.การออกแบบและพัฒนาแอปพลิเคชัน

ระบบที่ผู้วิจัยออกแบบขึ้นหลังวิเคราะห์ความต้องการ (Data and Requirement Analysis) โดยวิธีการเก็บข้อมูลในหลายวิธี เช่น การสัมภาษณ์ผู้ที่จะใช้ระบบ การรวบรวมเอกสารกระดาษที่เกี่ยวข้อง ศึกษากระบวนการจัดการคลังสินค้าของสถานประกอบการ เป็นต้น รวมถึงการทบทวนวรรณกรรมเกี่ยวกับระบบสารสนเทศที่เกี่ยวข้อง ผู้วิจัยนำข้อมูลที่ได้มาวิเคราะห์เพื่อออกแบบระบบ โดยระบบประกอบด้วย 3 ส่วนการทำงานดังต่อไปนี้

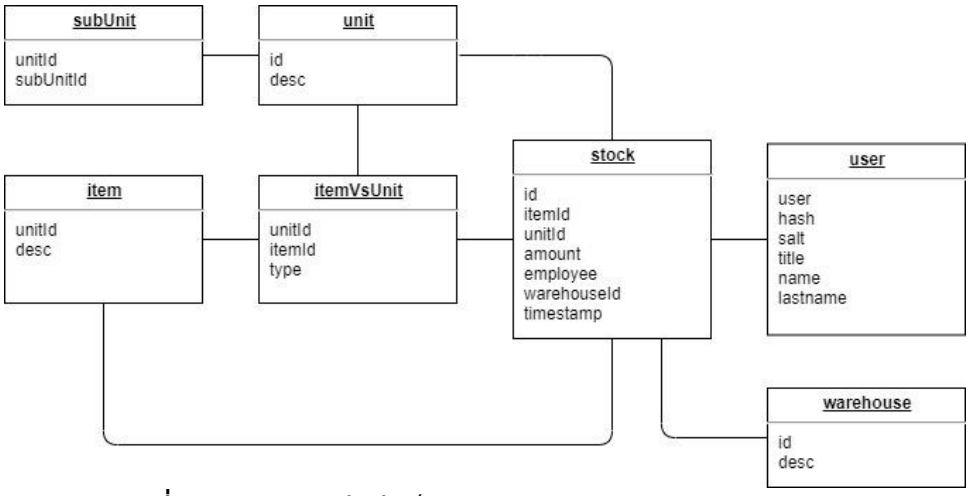
1) การเพิ่มสินค้าในคลังสินค้า
2) การเบิกสินค้า
3) การแจ้งเตือนข้อมูลสินค้า (เรียกดูรายละเอียดสินค้าในคลังสินค้า)

แผนภาพกิจกรรมของระบบที่ออกแบบขึ้นดังแสดงในรูปที่ 3



รูปที่ 3 แผนภาพกิจกรรม (Activity Diagram) ของระบบ

ผู้วิจัยออกแบบแนวคิด (Conceptual Modeling) โดยร่างแผนภูมิเพื่อออกแบบโครงสร้างฐานข้อมูลซึ่งประกอบด้วยคอลัมน์และความสัมพันธ์ระหว่างตารางต่างๆ รายละเอียดดังแผนภูมิความสัมพันธ์(Entity-Relationship Diagram) ในรูปที่ 4

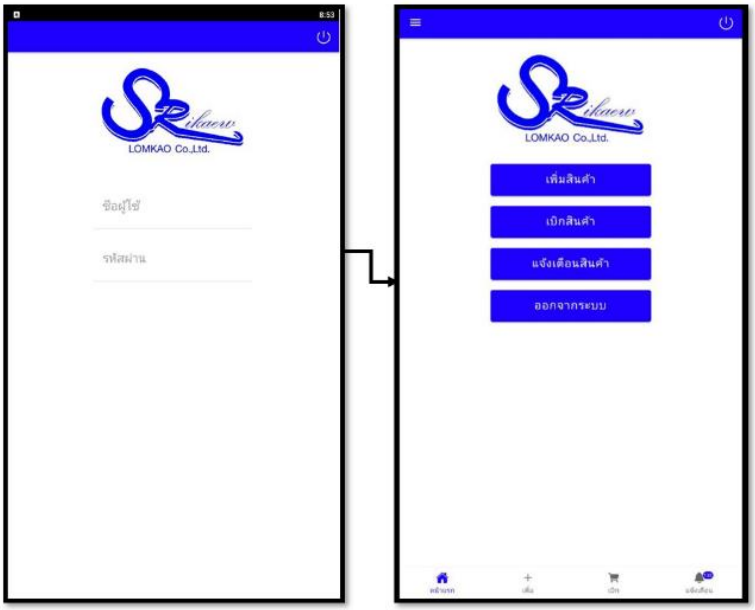


รูปที่ 4 แผนภูมิความสัมพันธ์(Entity-Relationship Diagram) ของระบบ

5. การพัฒนาแอปพลิเคชัน

ผู้วิจัยเลือกใช้ Ionic framework ในการพัฒนาระบบ Ionic framework เป็นเฟรมเวิร์คสำหรับการพัฒนาแอปพลิเคชันบนอุปกรณ์เคลื่อนที่และคอมพิวเตอร์ที่ประยุกต์ใช้พื้นฐานการพัฒนาเว็บ อาทิ HTML CSS และ JavaScript ซึ่งแตกต่างจาก Native Mobile Application ดั่งเดิมที่ต้องต้องเรียกใช้ผ่านระบบปฏิบัติการอย่างเฉพาะเจาะจงเพียงหนึ่งระบบปฏิบัติการเท่านั้น ข้อดีของ Ionic Framework คือแอปพลิเคชันที่พัฒนาขึ้นเป็นแอปพลิเคชันบนอุปกรณ์เคลื่อนที่แบบครอสแพลตฟอร์ม (Cross Platform) ซึ่งถูกสร้างให้เรียกใช้บนระบบปฏิบัติการอย่างอิสระ ซึ่งได้เปรียบ Native Mobile Application ซึ่งต้องพัฒนาระบบแยกกันหากต้องการได้แอปพลิเคชันที่ต่างระบบปฏิบัติการ

ผู้วิจัยได้พัฒนาแอปพลิเคชันบนอุปกรณ์เคลื่อนที่สำหรับการจัดการคลังสินค้า ผู้ใช้ต้องเข้าสู่ระบบ (login) เนื่องจากการจัดการข้อมูลแต่ละครั้งจะถูกการจัดเก็บพร้อมทั้งระบุตัวตนของผู้ดำเนินการในทุกขั้นตอน เมื่อเข้าสู่ระบบสำเร็จระบบจะแสดงเมนูการทำงานทั้งหมดบนหน้าจอ ระบบจะแสดงผลดังรูปที่ 5



รูปที่ 5 หน้าเข้าสู่ระบบและเมนูการทำงานหลัก

การเพิ่มสินค้าเข้าสู่คลังสินค้า ผู้ใช้ต้องระบุประเภทสินค้าและเลือกคลังสินค้าที่ต้องการเพิ่มสินค้า หน้าจอระบบจะแสดงจำนวนหน่วยของสินค้าที่สามารถเพิ่มได้ จากนั้นผู้ใช้สามารถเลือกเพิ่มจำนวนของสินค้าได้ตามที่ต้องการ ระบบจะแสดงผลดังรูปที่ 6

รูปที่ 6 หน้าเพิ่มสินค้าเข้าสู่คลังสินค้าของระบบ

การเบิกสินค้าจากคลังสินค้าเพื่อนำมาขายหรือส่งต่อนั้นมีขั้นตอนการทำงานเช่นเดียวกับการเพิ่มสินค้าในคลังสินค้า ซึ่งก็คือการเลือกสินค้าเลือกคลังที่เก็บสินค้าหลังจากนั้นระบุจำนวนที่ต้องการ ระบบจะแสดงผลดังรูปที่ 7

รูปที่ 7 หน้าเบิกสินค้าออกจากคลังสินค้าของระบบ

หลังจากมีการบันทึกข้อมูลการเพิ่มสินค้าและการเบิกสินค้า ข้อมูลทั้งหมดที่เกี่ยวข้องกับปริมาณสินค้าในคลังสินค้าจะถูกบันทึกแบบเวลาจริง(Real Time) และเพื่อการติดตามข้อมูลเพื่อลดขั้นตอนที่พนักงานต้องเดินเข้าไปตรวจเช็คสินค้าตามที่ลูกค้าสั่ง ผู้พัฒนาระบบได้ออกแบบส่วนการแจ้งเตือนจำนวนสินค้า ในเมนูดังกล่าวสามารถดูจำนวนคงเหลือของแต่ละคลังสินค้าพร้อมทั้งเลือกดูรายการสินค้าเฉพาะอย่างได้ ระบบจะแสดงผลดังรูปที่ 8

รูปที่ 8 หน้าดูรายการสินค้าเพื่อการติดตามจำนวนสินค้าคงคลัง

6. อภิปรายผลการวิจัย

6.1 การลดต้นทุนเชิงเวลา

แอปพลิเคชันที่ใช้ในการจัดการและบริหารคลังสินค้าเมื่อมีการสั่งสินค้าที่มีลักษณะพิเศษ (ลูกค้าไม่ประจำ) จากเดิมใช้เวลาในการตรวจเช็คสินค้าในคลังประมาณ 5 นาที ต่อการสั่ง 1 ครั้ง โดยมีรายละเอียดดังนี้

(1) พนักงานรับคำสั่งซื้อจากจากลูกค้า ประมาณ 0.5 นาที

(2) พนักงานเดินเข้าไปตรวจเช็คคลังสินค้าทั้ง 3 คลัง ประมาณ 5 นาที

(3) พนักงานกลับมาแจ้งลูกค้าว่าสินค้าที่สั่งมีเพียงหรือไม่ 1 นาที

(4) พนักงานพิมพ์ใบสั่งซื้อ 0.5 นาที

ในแต่ละวันมีการสั่งสินค้าที่มีลักษณะพิเศษ (ลูกค้าไม่ประจำ) ประมาณ 4 ครั้งต่อวัน ดังนั้นใน 1 วันจะใช้เวลาในการตรวจนับสินค้าทั้งสิ้น 20 นาที แต่หลังจากที่มีการใช้แอปพลิเคชันเบื้องต้นจะใช้เวลาในการตรวจนับคลังในคลังสินค้าประมาณ 5 วินาที เพราะฉะนั้นใน 1 วัน จะใช้เวลาในการตรวจสินค้ารวม 20 วินาที หรือลดลงคิดเป็นร้อยละ 98.33

6.2 การลดต้นทุนแรงงาน

พนักงานมีค่าจ้าง 300 บาทต่อวัน คิดเป็น 37.5 บาทต่อชั่วโมง คิดเป็น 3.125 บาท ต่อการสั่งซื้อ 1 ครั้ง แต่เมื่อใช้แอปพลิเคชันมาช่วยในการปฏิบัติงานจะคิดเป็นต้นทุนแรงงานเพียงแค่ 0.052 บาทต่อการสั่งซื้อ 1 ครั้ง หรือคิดเป็นร้อยละ 98.34

ตารางที่ 1 ผลวิจัยหลังการใช้งานระบบที่พัฒนาขึ้น

ตัวชี้วัดเชิงปริมาณ	การประเมินผล		
	ก่อน	หลัง	ผลลัพธ์
1.เวลาในการปฏิบัติงาน (วินาทีต่อการสั่งซื้อ 1 ครั้ง)	300	5	ระยะเวลาลดลง ร้อยละ 98.33
2.ต้นทุนแรงงาน (บาทต่อการสั่งซื้อ 1 ครั้ง)	3.125	0.052	ต้นทุนแรงงานลดลงร้อยละ 98.34

7. กิตติกรรมประกาศ

คณะผู้วิจัยในโครงการวิจัยนี้ ขอขอบพระคุณ โรงงานมะขามศรีแก้ว อ.หล่มเก่า จ.เพชรบูรณ์ และพนักงานที่เกี่ยวข้อง ที่ได้ให้ข้อมูลที่เป็นประโยชน์ต่อการทำวิจัยและขอบคุณคณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏพิบูลสงคราม จ.พิษณุโลก ที่ให้ความอนุเคราะห์อุปกรณ์เครื่องมือ และสถานที่ในการทำวิจัย รวมไปถึงหน่วยงานที่ให้การสนับสนุนอันเป็นประโยชน์อย่างยิ่งต่อการทำวิจัยครั้งนี้

8. เอกสารอ้างอิง

Bowersox, D. J., Closs, D. J., & M Bixcooper. (2007). **Supply chain logistics management**. Mcgraw-Hill.

Li, Q., & Liu, A. (2019). **Big Data Driven Supply Chain Management**. Procedia CIRP, 81, 1089–1094. https://doi.org/10.1016/j.procir.2019.03.258

White Paper of DHL. (2013a). **Big Data in Logistics-A DHL Perspective on how to move beyond the hype.** DHL Customer Solutions & Innovation.

Tahiduzzaman, Md & Rahman, Mustafizur & Rahman, Soyed & Dey, Samrat & Akash, Sadi. (2017). **Big data and its impact on digitized supply chain management**.. 3. 196-208.

Rob, P., & Semaan, E. (2000). **Databases : design, development and deployment : using Microsoft Access.** Irwin/Mcgraw-Hill.

Toj Teorey. (2011). **Database modeling and design : logical design**. Morgan Kaufmann Publishers.

Keng Siau. (2007). **Contemporary issues in database design and information systems development**. Igi Pub.

Arora, K., Farr, E., Gilbert, J., & Zonooz, P. (2019). **Architecting Cloud Native Applications: Design High-Performing and Cost-effective Applications for the Cloud**. Birmingham: Packt Publishing, Limited.

การหาตำแหน่งตราสินค้านมโคพาสเจอร์ไรส์จากความเห็นผู้บริโภคบนสื่อสังคมออนไลน์

พัชรพล อ่วมโอพาร [1*] และธนชาตย์ ฤทธิ์บำรุง [2]

Received: 19 July 2021; Revised: 20 November 2021; Accepted: 1 December 2021

บทคัดย่อ

ในปัจจุบัน การแข่งขันทางธุรกิจที่เข้มข้น และเทคโนโลยีที่พัฒนาอย่างรวดเร็ว ทำให้ทุกอุตสาหกรรมพิจารณาการสร้างสินค้าใหม่ให้แตกต่างจากคู่แข่ง เพื่อให้สินค้านั้นมีความโดดเด่นกว่าสินค้าอื่นในตลาด และทำให้ผู้บริโภคสามารถจดจำสินค้านั้นได้ เมื่อองค์กรต้องการผลิตและขายสินค้าใหม่ องค์กรมักเริ่มจากการแบ่งกลุ่มตลาด การกำหนดเป้าหมาย และการกำหนดตำแหน่งของตลาด เพื่อให้สินค้านั้นสามารถครองใจผู้บริโภคและอยู่รอดในตลาดได้ การศึกษานี้มีวัตถุประสงค์ในการหาตำแหน่งตราสินค้าของนมโคพร้อมดื่มพาสเจอร์ไรส์ในประเทศไทยจำนวน 7 ตราสินค้าที่วางขายในตลาดด้วยข้อมูลความเห็นผู้บริโภคบนสื่อสังคมออนไลน์ Pantip.com นำความเห็นมาตัดคำ คัดเลือกคำสำคัญ 500 คำ โดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ ก่อนจัดตำแหน่งตราสินค้าโดยใช้เทคนิคการแบ่งสเกลหลายมิติตามมุมมองปัจจัยที่ผู้บริโภคพิจารณาในการบริโภคนมโคพาสเจอร์ไรส์ 8 ด้าน พร้อมเทียบผลกับการรับรู้ตำแหน่งตราสินค้าจากแบบสอบถามที่สร้างขึ้นจากมุมมองปัจจัย 8 ด้านเช่นกัน จากนั้นจึงสร้างกราฟโครงข่ายเชิงความหมายเพื่อพิจารณาความเชื่อมโยงระหว่างคำสำคัญ รวมถึงค้นหาข้อมูลเชิงลึกโดยใช้เทคนิคการตรวจหาเครือข่ายชุมชน พบว่าการจัดตำแหน่งตราสินค้าด้วยการแบ่งสเกลหลายมิติ เมื่อเปรียบเทียบความสัมพันธ์ระหว่างข้อมูลจากแบบสอบถามและข้อมูลจาก Pantip.com มีค่าสหสัมพันธ์ QAP เท่ากับ 0.56 และ pseudo p-value เท่ากับ 0.01

คำสำคัญ: เทคนิคการแบ่งสเกลหลายมิติ, การประมวลผลภาษาธรรมชาติ , การวิเคราะห์โครงข่ายเชิงความหมาย, การตรวจหาเครือข่ายชุมชน

** Corresponding E-mail: pacharapol.oum@stu.nida.ac.th*

[1,2]สาขาการวิเคราะห์ธุรกิจและวิทยาการข้อมูล คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

Pasteurized milk brand positioning from online social media reviews

Pacharapol Oumolarn [1*] and Thanachart Ritbumroong [2]

Received: 19 July 2021; Revised: 20 November 2021; Accepted: 1 December 2021

Abstract

The technological advancement has brought a new level of intense market competition encouraging most players in any industry to differentiate themselves from competititors in order to capture consumers' attention. As a results, various new products and services are being offered to a market. Segmentation, targeting, and positioning is a marketing technique generally utilized to sell new products and services. With the differentiation strategy employed to compete in this intense competition, choosing a unique position for new products and services is pivotal for business. This research explored the market positions of seven pasteurized cow's milk brands in Thailand. Data were collected from a famous social network forum in Thailand, Pantip.com. Comments from the forum were tokenized. The results of tokenization revealed 500 keywords. These keywords were categorized into eight influential factors affecint milk consumption. Brand perceptual maps by the influential factors were constructed using Multidimensional Scaling. Sementic Network and Community Detection techniques were performed on the dataset. From a discussion with industry experts, the results revealed new insights. To validate the results of brand perceptual maps, self-administered questionnaire survey was conducted asking respondents about brand position. There were no significant discrepancies between the results between Multidimensional Scaling and survey methods with QAP Correlation = 0.56 with pseudo p-value = 0.01.

Keywords: Multidimensional Scaling, Natural Language Processing, Semantic Network Analysis, Community Detection

** Corresponding E-mail: pacharapol.oum@stu.nida.ac.th*

[1,2]Business Analytics and Data Science, Graduate School of Applied Statistics, National Institute of Development and Administration

1. บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบัน เมื่อองค์กรต้องการผลิตและวางขายสินค้าใหม่ จะต้องมีการวิเคราะห์ตลาดเสียก่อน ด้วยการแบ่งกลุ่มตลาด การกำหนดเป้าหมาย และกำหนดตำแหน่งทางการตลาด เพื่อให้ผู้บริโภครับรู้การมีอยู่ของสินค้า และตัวสินค้าสามารถอยู่รอดในตลาดได้ (รังสรรค์ เลิศในสัตย์, 2549)

การหาตำแหน่งทางการตลาดตราสินค้าและผลิตภัณฑ์ในประเทศไทยในช่วงที่ผ่านมา ใช้วิธีสำรวจ สัมภาษณ์ก่อนนำคำตอบมาสร้างแผนภาพการรับรู้ของผู้บริโภค (Perceptual map) โดยใช้เทคนิคการแบ่งสเกลหลายมิติ หรือ Multidimensional Scaling - MDS (ชื่นสุมล และ บุษราภรณ์, 2557) วิธีนี้แม้เป็นวิธีที่มีประสิทธิภาพ แต่ก็เสียเวลาและงบประมาณในการทำแบบสอบถามและการสัมภาษณ์ อีกทั้งอาจไม่ได้ความเห็นที่แท้จริงของผู้บริโภคเนื่องจากอคติของผู้ถูกสัมภาษณ์ (ชูชัย สมิทธิไกร, 2562) หากผู้บริโภคไม่มีความสนใจในตราสินค้านั้น หรือมีประสบการณ์ไม่ดีเกี่ยวกับผลิตภัณฑ์ จะไม่เต็มใจให้ความเห็นกับผู้สัมภาษณ์ (Malhotra, 2019)

ทุกวันนี้ การพัฒนาแพลตฟอร์มออนไลน์ในประเทศไทยได้เติบโตแบบก้าวกระโดด ผู้คนเลือกรับข้อมูลข่าวสารผ่านสื่อใหม่ (New Media) เช่น สื่อสังคม (Social Media) และแพลตฟอร์มออนไลน์อื่นๆ มากขึ้น โดยคนไทยใช้งานสื่อสังคมคิดเป็น 91.2 % ของผู้ใช้อินเทอร์เน็ตในประเทศ (สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์, 2563, น. 39) ปัจจุบันแพลตฟอร์มออนไลน์นั้นเต็มไปด้วยข้อมูลที่เกี่ยวกับการบริโภคผลิตภัณฑ์ เกือบทั้งหมดอยู่ในรูปแบบข้อมูลไม่มีโครงสร้าง และด้วยปริมาณที่มากสามารถมองให้เป็นข้อมูลจากการสัมภาษณ์กลุ่มแบบเจาะจงขนาดใหญ่ได้ (Netzer, Feldman, Goldenberg & Fresko, 2012) หากสามารถนำข้อมูลดังกล่าวมาประมวลให้อยู่ในรูปแบบที่ประยุกต์ใช้งานได้ องค์กรย่อมสามารถวางแผน ออกแบบ กำหนดคุณลักษณะผลิตภัณฑ์และกลุ่มผู้บริโภค เพื่อให้ผู้บริโภคสามารถจดจำผลิตภัณฑ์ได้ และตัดสินใจเลือกซื้อต่อเนื่องอย่างเต็มใจ

1.2 วัตถุประสงค์ของงานวิจัย

งานวิจัยนี้มีแนวคิดในการระบุตำแหน่งทางการตลาดตราสินค้านมโคพร้อมดื่มพาสเจอร์ไรส์ จำนวน 7 ตราสินค้าที่วางขายในประเทศไทย ประกอบด้วย เมจิ ดัชมิลล์ ไทยเดนมาร์ค โฟร์โมสต์ แดรี่โฮม เอ็มมิลค์ และโชคชัย โดยใช้เทคนิคการแบ่งสเกลหลายมิติ (MDS) พิจารณาตำแหน่งทางการตลาดตราสินค้าในมิติปัจจัยที่มีผลต่อการบริโภคนมแตกต่างกัน 8 ด้าน และใช้เทคนิคการวิเคราะห์โครงข่ายเชิงความหมาย (Semantic Network Analysis - SNA) พิจารณาองค์ประกอบที่เกี่ยวข้องกับตราสินค้า ทั้งหมดใช้ข้อมูลความเห็นของการบริโภคสินค้าจากกระทู้ในสื่อสังคมออนไลน์ Pantip.com เสมือนการสำรวจตลาด เพื่อนำเสนอข้อมูลเชิงลึก ลดการทำแบบสอบถามหรือลงพื้นที่สำรวจความเห็น ก่อนนำข้อมูลไปใช้วางแผนในการจัดกลุ่มตลาดหรือวางสินค้าใหม่ต่อไป

1.3 ประโยชน์ที่คาดว่าจะได้รับ

การใช้สื่อสังคมออนไลน์ทำให้ทั้งองค์ประกอบสำคัญที่เคยนำมาเป็นข้อคำถามในแบบสอบถามและองค์ประกอบที่ไม่เคยทราบมาก่อน ถูกนำมาพิจารณาในการกำหนดตำแหน่งตราสินค้า อีกทั้งการสร้างภาพนิทัศน์ความเชื่อมโยงระหว่างองค์ประกอบ ทำให้ผู้บริหารหรือผู้ใช้งานทั่วไปสามารถมองเห็นข้อมูลเชิงลึกที่ไม่เคยถูกนำมาเป็นข้อคำถามในแบบสอบถาม หรือเป็นปัจจัยใหม่ที่มีผลต่อการซื้อสินค้าของผู้บริโภคที่องค์กรไม่เคยรับรู้มาก่อน

ดังที่กล่าวมาทั้งหมดนี้ องค์กรสามารถระบุความแตกต่างระหว่างตราสินค้าในมิติต่างๆ ก่อนนำไปสู่การคิดค้นสินค้าหรือบริการที่โดดเด่นจากคู่แข่งได้ ทั้งยังตรงความต้องการของผู้บริโภคมากขึ้น กำหนดกลยุทธ์การทำตลาดได้เหมาะสม ลดเวลาและงบประมาณในการสัมภาษณ์ เนื่องจากหากต้องใช้ข้อคำถามและจำนวนตัวอย่างจำนวนมากเพื่อให้มีข้อมูลมากพอจนสะท้อนภาพรวมประชากรในสื่อสังคมออนไลน์ ย่อมหมายถึงจำนวนทรัพยากรที่ต้องใช้มากขึ้นและข้อคำถามที่มีจำนวนมากนั้นสร้างอคติต่อแบบสอบถามเพิ่มมากขึ้นเช่นกัน

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ปัจจัยที่มีผลต่อการบริโภคนม

ผู้วิจัยได้ทบทวนวรรณกรรมเกี่ยวกับปัจจัยที่มีผลต่อการบริโภคนมทั้งในประเทศไทยและต่างประเทศ สามารถสรุปได้ 8 กลุ่ม ดังต่อไปนี้

1. แม่และเด็ก Kurajdova, Petrovicova & Kascakova (2015) ได้ระบุว่าเป็นปัจจัยหนึ่งที่มีผลต่อการบริโภคจากผลการศึกษา และ Hsu & Lin (2006) ยังได้ระบุว่า เด็กมักจะชอบนมที่มีรสอร่อย จึงทำให้เป็นที่นิยม รวมถึงในประเทศไทย ปิยฉัตร ช่างเหล็ก, ยอดมนี เทพานนท์ และณัฐพล พันธุ์ภักดี (2561) ได้มีการกล่าวถึงประเด็นเกี่ยวกับนมโรงเรียน และความสำคัญในประเด็นเกี่ยวกับเด็ก

2. การส่งเสริมการขาย หรือการจัดโปรโมชัน De Alwis, Edirisinghe, & Athauda (2011) และสุภชาติ ชัยณรงค์สิงห์ (2539) ได้ระบุตรงกันว่าเป็นปัจจัยที่มีความสำคัญในการพิจารณาบริโภคนม เช่น การลดราคา ของแถม และการจัดโปรโมชันต่างๆ

3. คุณภาพ ความปลอดภัย ความสะอาด และมาตรฐานการผลิต จากการศึกษาของ Yin et al. (2016) และ Hsu & Lin (2006) ได้กล่าวถึงอย่างชัดเจนจากกรณีศึกษาในประเทศจีน ผู้บริโภคยอมจ่ายมากขึ้นเพื่อซื้อนมที่ผ่านการรับรองด้านคุณภาพจากทางการ เพื่อหลีกเลี่ยงสารปรุงแต่งที่ไม่มีประโยชน์หรือเป็นโทษต่อสุขภาพ

4. ความยากง่ายในการหาซื้อ De Alwis et al. (2011) และสุภชาติ ชัยณรงค์สิงห์ (2539) ได้ระบุตรงกันว่าเป็นปัจจัยที่มีความสำคัญในการพิจารณาบริโภคนม เช่น ประเภทห้างสรรพสินค้า ช่องทางการขายที่แตกต่างกัน

5. การทำกาแฟ ทำอาหาร และขนม Kurajdova et al. (2015) ได้ระบุว่าเป็นปัจจัยที่มีผลต่อการจูงใจให้บริโภคนม โดยผู้บริโภคนำนมมาผสมกับกาแฟ ขนม หรืออาหาร Kurajdova et al. ได้กล่าวเพิ่มเติมว่าบางวัฒนธรรมถือว่านมเป็นส่วนหนึ่งของอาหารที่ต้องมีในการบริโภคแต่ละมื้อ

6. ปัจจัยด้านสุขภาพ (Kurajdova et al., 2015; De Alwis et al., 2011; สุภชาติ ชัยณรงค์สิงห์, 2539) สำหรับปัจจัยในข้อนี้ มีการกล่าวถึงเหมือนกันในหลายการศึกษา ถือเป็นปัจจัยหลักในการพิจารณาบริโภคนม เช่น ความสูง การขับถ่าย คุณค่าทางอาหาร โรคภัย และการออกกำลังกาย

7. ส่วนผสม และสารอาหาร Hsu & Lin (2006) และ De Alwis et al. (2011) ได้ระบุตรงกันว่าเป็นปัจจัยที่มีความสำคัญเชิงบวกในการพิจารณาบริโภคนม เช่น ไขมัน แคลเซียม โดย Hsu & Lin กล่าวเพิ่มเติมอีกว่าผู้บริโภคยอมจ่ายมากขึ้นเพื่อสารอาหารที่มีคุณประโยชน์มากกว่า

8. ราคาและความคุ้มค่า ปิยฉัตร ช่างเหล็ก และคณะ (2561) และ De Alwis et al. (2011) จากผลการศึกษาได้ระบุว่าปัจจัยทางเศรษฐกิจและราคามีผลต่อการบริโภค เช่น รายได้ การตั้งราคา ความคุ้มค่า

2.2 การวิเคราะห์โครงข่ายเชิงความหมาย (Semantic Networks Analysis)

โครงข่ายเชิงความหมาย หรือ Semantic Networks เป็นเทคนิคหนึ่งในการแสดงความสัมพันธ์และความเชื่อมโยงระหว่างคู่ของวัตถุ การรับรู้ หรือพฤติกรรม โดยใช้กราฟโครงข่าย (Doerfel, 1998) เช่น องค์ประกอบทั่วไปของตราสินค้า การมีส่วนร่วมของผู้บริโภคต่อสินค้า และภาพลักษณ์ในสังคม (Henderson, Iacobucci & Calder, 1998)

กราฟโครงข่ายนั้นสามารถสร้างขึ้นได้โดยใช้ Adjacency Matrix ที่มีลักษณะเป็น Matrix จัตุรัส ในบางการศึกษา การสร้าง Semantic Networks ของคำสำคัญจากแหล่งที่สนใจ สามารถทำได้โดยนำคำเหล่านั้นนับความถี่การปรากฏร่วมกันระหว่างคำ 2 คำในเอกสารเดียวกัน นำมาสร้าง Co-occurrence Matrix หรือ Co-word Analysis ซึ่งมีลักษณะคล้ายกับ Adjacency Matrix ก่อนแปลงองค์ประกอบคำสำคัญระหว่างแถวและคอลัมน์ใน Matrix นั้นให้กลายเป็น Node หรือ Vertex (จุดยอด) และเส้นเชื่อมความสัมพันธ์ระหว่างกัน โดยแปลงค่าความถี่ให้กลายเป็นน้ำหนักของเส้น เช่น ความถี่ของการติดต่อสื่อสารผ่านอุปกรณ์ระหว่าง Actor 2 คน (Tabassum, Pereira, Fernandes & Gama, 2018) หรืออาจสร้างเป็น Distance Matrix เพื่อเชื่อมโยงความสัมพันธ์ระหว่างแถวและคอลัมน์ด้วยค่า

Interpersonal Tie ซึ่งสามารถแปลงให้เป็นความยาวของเส้นได้ (Vemprala, Xiong, Liu & Choo, 2019)

สำหรับผลลัพธ์เชิงประจักษ์ของการเชื่อมโยงความสัมพันธ์ระหว่างตราสินค้าและองค์ประกอบสินค้า Aaker (1996, น.94), Henderson et al. (1998) และ Peter & Olson (2010, น.55) ได้แสดงกราฟโครงข่ายซึ่งประกอบด้วยตราสินค้าร่วมกับองค์ประกอบต่างๆ Netzer et al. (2012) ได้ให้ความเห็นที่เกี่ยวข้องกับการเชื่อมโยงความสัมพันธ์ไว้ว่า “ตราสินค้า 2 ตัวที่เชื่อมโยงใกล้กันมีแนวโน้มว่าเกิดจากความจำระยะยาวที่ผู้คนนึกถึงบ่อย ทำนองเดียวกันองค์ประกอบของสินค้าใดๆ ที่เชื่อมโยงใกล้กับตราสินค้า มักปรากฏขึ้นในประโยคร่วมกันกับชื่อตราสินค้าบ่อยครั้ง”

2.3 Similarity Measure

ในการศึกษาเกี่ยวกับ Semantic Network Analysis ของตราสินค้า มีการใช้ Similarity Measure หรือมาตรวัดค่าความคล้ายคลึงสำหรับการ Normalize ความสัมพันธ์ระหว่างองค์ประกอบหรือคำสำคัญของตราสินค้าใน Co-occurrence Matrix เพื่อป้องกันผลกระทบจากอคติที่เกิดจากขนาดตลาด (Market size bias) เนื่องจากแต่ละตราสินค้ามีจำนวนการกล่าวถึงอย่างไม่สมดุลกัน (Malhotra & Bhattacharyya, 2019) จากการทบทวนวรรณกรรม งานวิจัยมีการใช้ Similarity Measure สำหรับการสร้าง Semantic Network ดังต่อไปนี้

2.3.1 Jaccard Index

เป็น Similarity Measure บนพื้นฐานของเซต โดยพิจารณา Cardinality ของเซตองค์ประกอบหรือคำสำคัญ 2 ตัวที่นำมาเทียบกัน ดังแสดงในสมการที่ (1)

$$J(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \qquad (1)$$

จากสมการที่ (1) *A* และ *B* คือ องค์ประกอบหรือคำสำคัญ A และ B ที่ปรากฏในทุกเอกสาร ค่า $A \cap B$ คือ องค์ประกอบหรือคำสำคัญ A และ B ปรากฏพร้อมกันในเอกสารเดียวกัน

2.3.2 Lift

เป็น Similarity Measure บนพื้นฐานของความน่าจะเป็น ซึ่งมีสมการแตกต่างจาก Lift Ratio ที่ใช้ในทฤษฎี Association Rules ดังแสดงในสมการที่ (2)

$$Lift\ (A,B) = \frac{P(A \cap B)}{P(A)\ P(B)'} \qquad (2)$$

โดย $P(A \cap B)$ หมายถึงความน่าจะเป็นที่องค์ประกอบหรือคำสำคัญ A และ B ปรากฏพร้อมกันในเอกสารเดียวกัน $P(A)$ และ $P(B)$ คือความน่าจะเป็นที่องค์ประกอบหรือคำสำคัญ A และ B ปรากฏขึ้นจากทุกเอกสาร และค่า $P(B)'$ เท่ากับ 1-$P(B)$ ถ้าหากว่าตัวส่วน $P(A)$ เป็น 0 ค่า Similarity Measure ที่ได้จะเป็นค่าอนันต์ ซึ่งการเกิด 0 นั้นสามารถเกิดจากกรณีองค์ประกอบ A ไม่มีปรากฏขึ้นจริงในทุกรายการเอกสาร

2.3.3 Inclusion Index

เป็น Similarity Measure ที่คิดค้นโดย Qin He ในปี 1999 เพื่อนำมาทำ Co-word Analysis ระหว่างคำสำคัญ ดังแสดงในสมการที่ (3)

$$I\ (A,B) = \frac{C_{AB}}{min(C_A, C_B)} \qquad (3)$$

C_{AB} = จำนวนเอกสารที่องค์ประกอบหรือคำสำคัญ A และ B ปรากฏพร้อมกันในเอกสารเดียวกัน และ $min(C_A, C_B)$ เป็นฟังก์ชันเลือกค่าต่ำสุดของจำนวนเอกสารที่องค์ประกอบหรือคำสำคัญ A และ B ปรากฏทั้งหมด อย่างไรก็ตาม Similarity Measure นี้ยังคงมีปัญหาการเกิดค่าอนันต์ลักษณะคล้ายกันกับ Lift ถ้าค่าในฟังก์ชัน min = 0

2.3.4 Cosine Similarity

เป็น Similarity Measure ที่พิจารณาองศาบน Inner Product Space ระหว่างองค์ประกอบหรือคำสำคัญ A และ B นิยมประยุกต์ใช้กับงานหลายประเภท เช่น ตัวแบบช่วยแนะนำสินค้า (Product Recommendation) และการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ดังแสดงในสมการที่ (4)

$$cos(\theta) = \frac{AB}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i}\sqrt{\sum_{i=1}^{n} B_i}} \quad (4)$$

2.3.5 Association Strength

เป็น Similarity Measure อีกรูปแบบที่คิดค้นและปรับปรุงโดย Van Eck & Waltman (2010) ใช้กับการแสดงผลในเทคนิค Visualization of Similarity (VOS) สามารถใช้จำนวนการปรากฏขององค์ประกอบหรือคำสำคัญ i และ j ในเอกสารทั้งหมด หรือ ผลรวมจำนวนการปรากฏร่วมกันระหว่างองค์ประกอบหรือคำสำคัญที่ได้คำนวณลงใน Matrix มาแล้ว ดังแสดงในสมการที่ (5)

$$S_{ij} = \frac{C_{ij}}{w_i, w_j} \quad (5)$$

โดยกำหนดให้ S_{ij} คือ Similarity Measure, C_{ij} คือจำนวนการปรากฏพร้อมกันขององค์ประกอบหรือคำสำคัญ i กับ j, w_i คือ จำนวนการปรากฏองค์ประกอบหรือคำสำคัญ i และ w_j คือจำนวนการปรากฏองค์ประกอบหรือคำสำคัญ j

2.4 Community Detection แบบ Modularity Base

เป็นวิธีการแบ่งกลุ่มขององค์ประกอบในกราฟโครงข่ายที่คิดค้นโดย Newman (2004) เพื่อเป็นวิธีทางเลือกนอกเหนือจาก Community Detection แบบ Girvan-Newman ที่ Newman เองได้ร่วมคิดค้นขึ้นในปี 2002 ต่อมาได้มีนักวิจัยหลายคนได้นำแนวคิดนี้ไปประยุกต์เป็นอัลกอริทึมใหม่ เช่น Louvain Community Detection (Blondel et al., 2008) หรือ Visualization of Similarity Clustering (Waltman, Van Eck & Noyons, 2010)

Modularity เป็นกระบวนการหาค่าเหมาะสมที่สุดอย่างหนึ่ง ตัว Modularity เป็นค่า Scalar ที่มีค่าระหว่าง -1 ถึง 1 หาโดยวัดความหนาแน่นของจำนวนจุดเชื่อมต่อภายใน Community แล้วนำไปเปรียบเทียบกับจำนวนจุดเชื่อมต่อภายในของ Community อื่น เพื่อสรุปผลการแบ่งกลุ่มดังสมการที่ (6)

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \quad (6)$$

กำหนดให้ Q คือ Modularity, A_{ij} คือน้ำหนักของเส้นระหว่าง Vertex i และ j, k_i และ k_j คือผลรวมของน้ำหนักเส้นที่เชื่อมต่อกับ Vertex i, j ในกราฟ, (c_i, c_j) คือ Community ที่กำหนดไว้ที่ Vertex i และ j, δ เป็นฟังก์ชัน δ (u, v) ที่มีค่าเท่ากับ 1 ถ้า u = v แต่ถ้าไม่ใช่ δ มีค่าเท่ากับ 0 และ m มีค่าดังสมการที่ (7)

$$m = \frac{1}{2}\sum_{j} A_{ij} \quad (7)$$

เมื่อหาค่า Modularity ได้แล้ว ค่า Q จะถูกใช้เป็นตัวเทียบคุณภาพของ Community Detection โดยใช้เทคนิคการหาค่าเหมาะสมที่สุด เพื่อให้ได้ค่า Q ที่ต่ำที่สุด วนรอบคำนวณค่า Q ซ้ำจนครบทุก Vertex ซึ่งวิธีการประเมินค่า Q นั้นแตกต่างกันไปตามวิธีของนักวิจัยที่คิดค้นอัลกอริทึม อาจถือได้ว่า Modularity เป็นฟังก์ชันวัตถุประสงค์ (Objective function) ในเทคนิคการหาค่าเหมาะสมที่สุดเพื่อใช้หา Community

2.5 Visualization of Similarity (VOS)

Van Eck & Waltman (2006) ได้คิดค้นวิธีการนำเสนอกราฟโครงข่ายขนาดใหญ่ ชื่อว่า Visualization of Similarities (VOS) มีวัตถุประสงค์ในการแก้ไขปัญหาการซ้อนทับกันขององค์ประกอบในกราฟโครงข่ายที่มีจำนวนมาก แต่แรกเดิมนั้น VOS ใช้สำหรับสร้างกราฟโครงข่ายความสัมพันธ์การอ้างอิงบรรณานุกรมงานวิจัย ซึ่งมีข้อมูลจำนวนมหาศาล โดยผู้วิจัยแต่ละท่าน มีจำนวนงานวิจัยและจำนวนการอ้างอิงไม่เท่ากัน ทั้งนี้ VOS ยังคงใช้ Co-occurrence Matrix เป็นข้อมูลตั้งต้นในการสร้างกราฟโครงข่าย เหมือนกับการสร้างกราฟโครงข่ายทั่วไป แต่จะใช้ Similarity Measure แบบ Association Strength ดังที่ได้กล่าวไปในข้อที่ 2.3.5 แล้ว

การสร้างกราฟโครงข่ายจาก Co-occurrence Matrix มี 3 ขั้นตอน

1. สร้าง Co-occurrence Matrix คำนวณ Similarity Measure โดยใช้ Association Strength

2. Mapping แต่ละ Vertex องค์ประกอบ โดยให้ค่า Similarity สูงอยู่ใกล้กัน และ Similarity น้อยอยู่ห่างกัน ดำเนินการหาค่าเหมาะสมที่สุดโดยให้ผลรวมของ Square Euclidean Distance มีค่าน้อยที่สุดระหว่างคู่องค์ประกอบหรือคำสำคัญ

3. ดำเนินการ Translation, Rotation และ Reflection

เมื่อสร้างกราฟโครงข่ายเชิงความหมายเรียบร้อยแล้ว VOS จะทำ Community Detection องค์ประกอบภายในต่อ โดยใช้วิธีแบบ Modularity Base ดังที่ได้กล่าวไปใน 2.4 เพียงแต่มีการปรับเล็กน้อยตามการศึกษาของ Waltman et al. (2010) โดยมีการใช้พารามิเตอร์พิเศษเพิ่มเติม ด้วยการเพิ่มส่วนสำหรับหาค่า Resolution γ ที่ได้จากส่วนกลับของค่า Distance ระหว่าง Vertex 2 จุด และนำค่า γ ไปแทนค่าในสมการวัตถุประสงค์ ดังสมการที่ (8)

$$V = \frac{1}{2m}\sum_{i<j}\delta(x,x_j)w_{ij}\left[c_{ij}-\gamma\frac{c_i c_j}{2m}\right] \qquad (8)$$

โดย $w_{ij}=\frac{2m}{c_i c_j}$ และ $distance_{ij}=\begin{cases}0 & if\ x_i = x_j\\ \frac{1}{\gamma} & if\ x_i \neq x_j\end{cases}$

ค่า Resolution ของสมการวัตถุประสงค์นี้ มีไว้เพื่อป้องกันปัญหาการเกิด Community ที่เล็กเกินไปจนทำให้ไม่สื่อความหมาย ซึ่ง Community ที่มีขนาดเล็กนั้นจะถูกรวบเข้าไปใน Community ที่ใหญ่กว่า

2.6 การแบ่งสเกลหลายมิติ (Multidimensional Scaling)

การแบ่งสเกลหลายมิติ หรือ Multidimensional Scaling (MDS) เป็นกระบวนการนำเสนอการรับรู้ ความชอบ ความสัมพันธ์ทางจิตวิทยาระหว่างสิ่งเร้า (Stimuli) ด้วยกราฟพิกัดเรขาคณิต การคำนวณตำแหน่งวัตถุที่สนใจนั้นใช้ตาราง Proximity Matrix สรุป Dissimilarity Measure ระหว่างองค์ประกอบ หรือใช้เป็นค่า Distance ระหว่างองค์ประกอบที่สนใจ ค่าดังกล่าวสามารถแปลงค่ามาจาก Similarity Measure ที่กล่าวถึงในข้อที่ 2.3 ได้ จากนั้นนำเสนอข้อมูลผ่านกราฟพิกัดเรขาคณิตที่เรียกว่า Spatial map หรือ Perceptual map

Malhotra (2019) ได้ระบุว่าปัจจุบันการแบ่งสเกลหลายมิตินั้นนิยมนำมาสร้างตำแหน่งทางการตลาดระหว่างวัตถุที่สนใจในมิติเฉพาะด้านสำหรับการวิเคราะห์ตลาด เช่น การวัดระดับภาพลักษณ์องค์กร การแบ่งกลุ่มตลาด การประเมินประสิทธิผลของการโฆษณา การเลือกช่องทางการขาย และการวิเคราะห์ราคา ตำแหน่งทางการตลาดสามารถช่วยระบุความแตกต่างของการกำหนดราคาในตลาดของตราสินค้าคู่แข่งได้ และทราบว่าสินค้าของตนเองหากเข้าตลาดไปแล้วจะต้องแข่งขันกับใครในมิติดังกล่าว

กระบวนการสร้างตัวแบบ MDS มีการประเมิน Goodness-of-fit ของตัวแบบ โดยใช้ค่า Stress ในส่วนนี้นิยมใช้ Kruscal's Stress ดังสมการที่ (9)

$$Stress = \sqrt{\frac{\sum\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum d_{ij}^{\ 2}}} \qquad (9)$$

โดย d_{ij} คือค่า Distance ระหว่างสิ่งเร้า 2 ตัวใน Proximity Matrix และ $\hat{d}_{ij}$ คือค่า Distance ระหว่างสิ่งเร้า 2 ตัว ที่ตัวแบบ MDS คำนวณตำแหน่งออกมา เกณฑ์การประเมิน Goodness-of-fit นั้น ถ้าค่ามากกว่า 0.2 ขึ้นไป = ไม่ดี, 0.1-0.2 = พอใช้, 0.05-0.099 = ดี, 0.025 = ดีมาก และ 0.000 สมบูรณ์แบบ

ในด้านจำนวนตัวอย่างกับการแบ่งสเกลหลายมิติ ไม่ได้มีการระบุจำนวนตัวอย่าง (n) ที่เหมาะสมสำหรับการทำแบบสอบถามการรับรู้ แต่เมื่อทบทวนวรรณกรรมแล้ว พบว่ามีงานวิจัยของ Rodger (1991) ได้ทำการทดลองหาว่าควรใช้ n เท่าใดจะเหมาะสมกับการทำ MDS โดยประเมิน Goodness-of-fit ด้วยค่า Stress ดำเนินการสร้างแบบจำลอง Monte-Carlo เพื่อหา n ที่มีผล Stress ดีที่สุด พบว่า n ระหว่าง 6-15 คน จะให้ผล Stress ได้อยู่ในระดับพอใช้จนถึงดี n ที่มากกว่านั้นไม่ได้ทำให้ Stress ลดลง หรือ เพิ่มขึ้น แต่กลับมีแนวโน้มคงที่ นอกจากนี้ Howard (1998) ได้อ้างอิงผลการศึกษาของ Rodger เช่นกัน แต่ระบุเพิ่มเติมว่าควรเลือกตัวอย่างให้มีความหลากหลายครอบคลุมทุกมิติ

2.7 Quadratic Assignment Procedure (QAP)

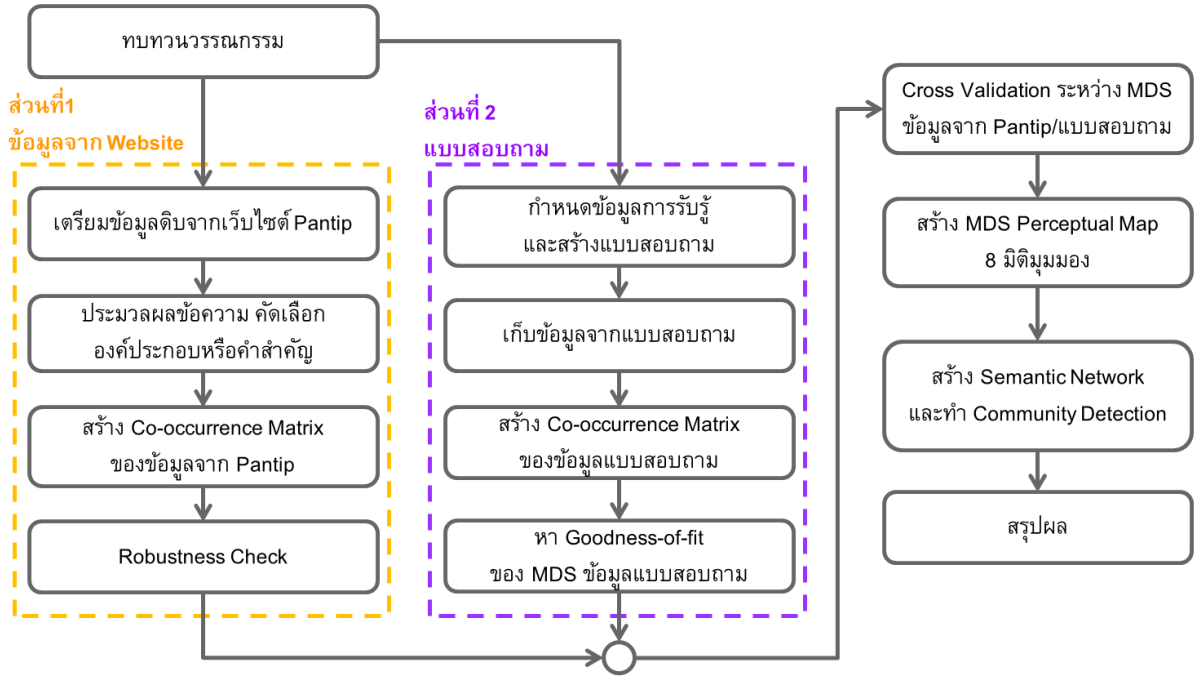
เป็นวิธีทดสอบสหสัมพันธ์แบบหนึ่งที่ใช้ใน Semantic Network Analysis และ Social Network Analysis Simpson (2001) ได้นิยามวิธีการนี้ไว้ว่า Quadratic Assignment Procedure (QAP) ใช้เพื่อวิเคราะห์ความสัมพันธ์ระหว่างแถวและคอลัมน์ของตารางข้อมูลแบบ Dyadic (ตารางข้อมูลที่แสดงความสัมพันธ์คู่ต่อคู่) เช่น Proximity Matrix ด้วยการใช้วิธีจำลองการเรียงสับเปลี่ยนข้อมูลในตาราง (QAP Permutations) เพื่อทดสอบค่าสหสัมพันธ์ระหว่างแถวและคอลัมน์ของตาราง Dyadic อย่างไรก็ตาม ตัวองค์ประกอบในแถวกับคอลัมน์นั้นอาจจะไม่มีความสัมพันธ์ระหว่างกันจริง จึงเป็นข้อที่ต้องระวังในการตีความ การทดสอบสหสัมพันธ์ ใช้ค่า Pseudo p-value ในการทดสอบสมมติฐานค่าสหสัมพันธ์ สามารถกำหนดสมมติฐานได้ดังนี้

H_0: No relationship between 2 matrices

H_1: Otherwise

3. ขั้นตอนวิธีดำเนินการวิจัย

การศึกษานี้อิงขั้นตอนวิธีดำเนินการวิจัยที่ใช้ร่วมกันโดยหลักจาก Netzer et al. (2012), ชื่นสุมล และ บุษราภรณ์ (2557), Ringel & Skiera (2016), Vemprala et al. (2019) และ Malhotra & Bhattacharyya (2019) เน้นการปรับปรุงและต่อยอดการศึกษาของ ชื่นสุมล และ บุษราภรณ์ และ Netzer et al. เป็นหลัก ผู้วิจัยเห็นว่าแนวทางการสร้าง Percetual Map ของ ชื่นสุมล และ บุษราภรณ์ มีความเหมาะสมเนื่องจากสามารถเลือกสรุปการรับรู้ตามกลุ่มปัจจัยหรือสิ่งเร้าที่สนใจได้ แต่จำนวนสิ่งเร้านั้นมีไม่มากเนื่องจากสร้าง Percetual Map จากแบบสอบถาม ผู้วิจัยจึงประยุกต์ใช้ข้อมูลสิ่งเร้าจากคำสำคัญที่เลือกมาจาก Pantip.com แทน และแสดงความสัมพันธ์ระหว่างตราสินค้าและองค์ประกอบหรือสิ่งเร้าเพิ่มเติมนอกเหนือการใช้ Percetual Map โดยผสมผสานขั้นตอนดำเนินการและจุดแข็งของงานวิจัยทั้ง 5 งานที่ได้กล่าวถึงในข้างต้น ทั้งนี้ เพื่อให้เข้าใจภาพรวมได้ง่าย ผู้วิจัยได้นำเสนอ Flow Chart ขั้นตอนวิธีดำเนินการวิจัยโดยรวม ดังในรูปที่ 1



รูปที่ 1 Flow Chart ภาพรวมของขั้นตอนวิธีดำเนินการวิจัย

3.1 ขั้นตอนการเตรียมข้อมูลดิบ

ผู้วิจัยได้สร้างโปรแกรมสำหรับการเก็บ URL ของกระทู้ที่มีการกล่าวถึงตราสินค้าด้วยโปรแกรมภาษา Python โดยใช้ URL ของการค้นหากระทู้และข้อความของ Pantip ที่มี Prefix เป็น https://pantip.com/search?q=xxx โดย xxx คือคำค้นที่ต้องการค้นหา เช่น https://pantip.com/search?q=นมเมจิ ดำเนินการลักษณะนี้ให้ครบทั้ง 7 ตราสินค้า โดยสามารถปรับคำค้นให้เหมาะสมได้กับบริบทของชื่อตราสินค้านั้น หากการค้นหาในแต่ละครั้งพบ URL กระทู้ที่เหมือนกันให้ลบ URL กระทู้ที่ปรากฏซ้ำออก เมื่อได้ URL กระทู้ที่ต้องการทั้งหมดแล้ว ทำการ Scraping ทีละ URL โดยเก็บเนื้อหาทั้งหัวกระทู้ และความเห็นแยกกัน ไม่ได้เก็บข้อมูลผู้เขียนข้อความ จากนั้นบันทึกข้อมูลทั้งหมดลงฐานข้อมูล MongoDB เมื่อทำการสำรวจข้อมูลพบว่า มีจำนวนกระทู้รวม 4,995 กระทู้ และจำนวนความเห็นทั้งหมด 225,360 ความเห็น ตั้งแต่วันที่ 3 มีนาคม 2556 ถึงกระทู้ใหม่ที่สุด 20 มีนาคม 2564

3.2 ขั้นตอนการประมวลผลข้อความ และคัดเลือกองค์ประกอบสำคัญ

ผู้วิจัยทำการลบกระทู้และความเห็นที่ถูกระงับการแสดงผล หรือ ถูกลบออกเพื่อรักษาบรรยากาศการสนทนาด้วยโปรแกรมภาษา Python เนื่องจาก Pantip ไม่ลบ URL ให้หายไปแต่ใช้ข้อความระบุว่ากระทู้หรือความเห็นนั้นถูกลบตามด้วยเหตุผลในการลบ รวมถึงลบความเห็นที่ถูกเผยแพร่ซ้ำโดยไม่ได้ตั้งใจ และความเห็นที่ไม่มีผู้ใช้เข้าไปกรอกความเห็น

เมื่อลบกระทู้และความเห็นเสร็จ ใช้ PyThaiNLP (Phatthiyaphaibun et al., 2016) ในการตัดคำ (Tokenized) โดยใช้อัลกอริทึม Newmm เพื่อให้เหมาะสมกับบริบทของ Pantip และใช้ระยะเวลาในการตัดข้อความที่ไม่มากจนเกินไป การประมวลผลใช้คำภาษาไทยเป็นหลัก ถ้ามีภาษาอังกฤษให้แปลงเป็นภาษาไทยให้มากที่สุด ยกเว้นชื่อห้างสรรพสินค้าและชื่อสารอาหารบางตัว ร่วมกับการสร้าง Customized Dictionary เฉพาะ Domain ของอุตสาหกรรมนม ประกอบด้วย ชื่อตราสินค้า ชื่อผลิตภัณฑ์ และองค์ประกอบที่สำคัญอื่น เมื่อตัดคำแล้ว ดำเนินการสร้าง Label ตราสินค้า และ Domain ของข้อความจากข้อความที่ถูกตัด เพื่อระบุว่ากระทู้หรือความเห็นนั้นกล่าวถึงชนิดของนมและชื่อตราสินค้าใดบ้าง ดัดแปลงตามวิธีของ Lee & Bradlow (2011) เพื่อนำไปใช้เป็นตัวกรองเอกสารที่ไม่ต้องการออก ทั้งนี้การระบุ Label เป็นการระบุโดยตรงจากคำสำคัญที่ปรากฏแล้วตรงตามเงื่อนไขแบบตรงไปตรงมา พิจารณาจากข้อความในหัวกระทู้ หรือในความเห็น เมื่อกรองเอกสารตามเงื่อนไขแล้ว เหลือกระทู้เพียง 1,077 กระทู้ และ 12,546

ความเห็น ด้วยเงื่อนไขของ Domain นมพาสเจอร์ไรส์เท่านั้น

หลังจากกรองเอกสารออก สร้าง Bag of Words ของคำทั้งหมดพร้อมสร้าง LDA Model ด้วย LDAvis (Sievert & Shirley, 2014) เพื่อช่วยคัดเลือกชุดคำที่จะนำมาเป็นองค์ประกอบสำหรับการสร้าง Co-occurrence Matrix โดยไม่ได้มีเป้าหมายในการสร้างตัวแบบหัวข้อ (Topic Modeling) แต่เพื่อลดจำนวนคำที่ถูกตัดให้เหลือแต่คำที่มีความน่าจะเป็นของการปรากฏในทุกเอกสารไม่มากและไม่น้อยเกินไป ลดโอกาสการเกิดค่าอนันต์ของ Similarity Measure ระหว่างการสร้าง Co-occurrence Matrix ก่อนนำคำทั้งหมดที่ถูกเลือกมาจัดหัวข้อของคำด้วยมือ พิจารณาตามปัจจัยที่มีผลต่อการบริโภคนม 8 กลุ่มปัจจัย หากความหมายกำกวมหรือไม่ทราบ ให้จัดเข้ากลุ่มที่ 9 คือกลุ่มที่ความหมายไม่เข้าพวก เพื่อนำไปหาข้อมูลเชิงลึกในขั้นตอนการทำ Community Detection ต่อไป สำหรับคำสำคัญทั้งหมดที่ได้คัดเลือกมา มีทั้งหมด 500 คำ ใน 9 กลุ่มหัวข้อ เหตุผลที่จัดหัวข้อของคำด้วยมือ เนื่องจากต้องการให้ข้อมูลบางส่วนของการหาตำแหน่งตราสินค้าอยู่ในรูปแบบเปรียบเทียบกันได้กับสิ่งเร้าในแบบสอบถามที่จำเป็นต้องกำหนดกลุ่มของคำถามการรับรู้ไว้ล่วงหน้าก่อน (A Priori) ในขณะที่ข้อความที่มาจาก Pantip เป็นประสบการณ์ของตนหรือคนอื่นที่ไม่ได้กำหนดหัวข้อตายตัว (A posteriori)

3.3 ขั้นตอนการสร้างแบบสอบถาม และการเตรียมข้อมูลการรับรู้จากแบบสอบถาม

ถึงแม้ว่าการใช้ข้อมูลจากสื่อสังคมออนไลน์นั้นมีเป้าหมายเพื่อลดการทำแบบสอบถาม แต่ในการศึกษานี้ยังคงใช้เพื่อแสดงกระบวนการ Cross Validation ของผลลัพธ์ระหว่างตำแหน่งทางการตลาดตราสินค้าจาก Pantip.com กับแบบสอบถามให้เป็นที่ประจักษ์ โดยเริ่มจากกำหนดชุดสิ่งเร้าจากปัจจัยที่มีผลต่อการบริโภคนม 8 กลุ่ม ก่อนเริ่มสร้างคำถาม แบบสอบถามนั้นแบ่งออกเป็น 3 ส่วน ประกอบด้วย รูปภาพประกอบซึ่งเป็นรูปภาพตัวอย่างสินค้าที่วางขายในตลาด ตัวอย่างการกรอกแบบสอบถาม และส่วนสุดท้ายคือช่องสำหรับกรอกข้อมูลการรับรู้ตราสินค้าจากสิ่งเร้า โดยใช้ Scale แบบ Attribute Rating พร้อมระบุความหมายของ Scale อย่างชัดเจนในทิศทางเดียวกัน จำนวนตัวอย่างที่ใช้ในการสำรวจ ใช้เพียง 40 คน เป็นพนักงานในโรงพยาบาลรามาธิบดี เขตราชเทวี จังหวัดกรุงเทพมหานคร เหตุผลที่ใช้ตัวอย่างเพียง 40 คนนั้น เนื่องจากพบว่า Rodger (1991) สรุปไว้ว่าจำนวนตัวอย่างระหว่าง 6-15 คนจะให้ค่า Stress ซึ่งเป็นค่า Goodness-of-fit ของตัวแบบ MDS อยู่ในเกณฑ์ดี ทำให้ถึงแม้แบบสอบถามที่ตอบกลับมาไม่สมบูรณ์และถูกตัดออกถึง 50% จำนวนตัวอย่างนั้นยังคงสอดคล้องกับการศึกษาของ Rodger ผู้วิจัยได้จัดทำแบบสอบถามและดำเนินการทดสอบกับกลุ่มตัวอย่างที่ไม่ได้อยู่ในกลุ่มสำรวจ จำนวน 5 คน เพื่อตรวจสอบคำผิดและความเข้าใจของตัวแบบสอบถาม จากนั้นจึงดำเนินการสุ่มแบบตามสะดวก ร่วมกับการสุ่มแบบเจาะจง เมื่อได้ข้อมูลกลับมา ทำ Feature Scaling กับผลลัพธ์ ด้วยวิธี Min-Max Normalization

3.4 ขั้นตอนการสร้าง Co-occurrence Similarity Matrix

หลังจากรวบรวมข้อมูลจาก Pantip และแบบสอบถามแล้ว ดำเนินการสร้าง Co-occurrence Matrix จากคำสำคัญที่เลือกมาด้วยโปรแกรมภาษา Python เพื่อนำไปสร้างตัวแบบ MDS และกราฟโครงข่าย โดยใช้ Similarity Measure แบบ Jaccard Index เหตุผลที่ใช้ Jaccard นั้นเพราะต้องการให้ตัวแบบยืดหยุ่นต่อการสำรวจองค์ประกอบใหม่ในตลาด ลดโอกาสการเกิดค่าอนันต์หากคำสำคัญคำใดคำหนึ่งไม่ปรากฏจริงในเอกสาร ทั้งนี้แต่ละ Matrix มีกระบวนการในการคำนวณที่แตกต่างกัน แบ่งเป็น 2 กลุ่ม ประกอบด้วย Matrix ขนาดใหญ่ที่นำไปใช้สร้างกราฟโครงข่าย และ Matrix ขนาดเล็กที่นำไปใช้ทำ Robustness check, MDS และ Cross Validation ดังนี้

กลุ่มที่ 1 คำนวณค่า Jaccard Index เพื่อสร้าง Co-occurrence Matrix จากการเทียบคู่การปรากฏคำสำคัญพร้อมกันในเอกสาร ขนาด 500x500 ทำการ Normalized ด้วย Jaccard Index

กลุ่มที่ 2 สร้าง Co-occurrence Matrix ขนาด 7x7 ตราสินค้า ด้วย Jaccard Index ด้วยวิธีเทียบการปรากฏร่วมกัน 3 ส่วน ระหว่างตราสินค้า A คำสำคัญ X และตราสินค้า B เทียบคู่ตราสินค้ากับคำสำคัญจนครบทั้ง 500 คำ และครบทุกคู่ตราสินค้า จากนั้นหาค่าเฉลี่ยของ Jaccard Index จากคู่ตราสินค้าเทียบกับคำสำคัญ เพื่อให้เหลือค่า Jaccard

Index เฉพาะระหว่างคู่ตราสินค้า 7 ตราเท่านั้น ดัดแปลงสอดคล้องกับวิธีของ Lee (2007) ที่ได้ดำเนินการในลักษณะคล้ายกันในการหา Jaccard Index เฉลี่ย ในตาราง Collaborative Filtering ของการให้คะแนน Rating ภาพยนตร์ใน MovieLens และ Netflix ระหว่างผู้ใช้มากกว่า 1 คนที่สนใจภาพยนตร์เรื่องเดียวกัน

3.5 ขั้นตอนการทดสอบความแข็งแกร่งทางสถิติ (Robustness check) และ Cross Validation

ผู้วิจัยได้เลือกวิธีการทดสอบความแข็งแกร่งทางสถิติ (Robustness check) และ Cross Validation ตามแนวทางของ Netzer et al. (2012) และ Malhotra & Bhattacharyya (2019) แต่ไม่ได้เลือกใช้ทั้งหมดเนื่องจากบางวิธีไม่สามารถปรับใช้กับข้อมูลจาก Pantip.com ได้ เพื่อให้แน่ใจว่ากระบวนการสร้าง Co-occurrence Matrix มีความถูกต้อง โดยสร้าง Matrix เพื่อทำการทดสอบด้วยวิธีการแบบกลุ่มที่ 2 ในข้อที่ 3.4 ด้วยโปรแกรมภาษา Python ดังนี้

1. Robustness check โดยเทียบค่าสหสัมพันธ์ QAP ระหว่าง Co-occurrence Matrix ขนาดมิติ 7x7 ที่สร้างในรูปแบบทางเลือก (Alternative form) จำนวน 5 Matrix ด้วย Similarity Measure ที่แตกต่างกัน 5 แบบ ประกอบด้วย Jaccard Index, Cosine Similarity, Lift, Inclusion Index และ Association Strength

2. Robustness check โดยเทียบค่าสหสัมพันธ์ QAP ระหว่าง Co-occurrence Matrix ขนาดมิติ 7x7 จำนวน 5 Matrix ที่สร้างด้วย Jaccard Index โดย Matrix ทั้งหมดถูกสร้างจากจำนวนเอกสารตั้งต้นที่สุ่มมาจากเอกสารทั้งหมด 12,546 เอกสาร ในสัดส่วน 1/1, 1/2, 1/4, 1/8 และ 1/16 ตามลำดับ การสุ่มแต่ละครั้งเป็นแบบสุ่มแล้วใส่คืน

3. Cross Validation เทียบค่าสหสัมพันธ์ QAP ระหว่างข้อมูลการรับรู้จาก Pantip.com และข้อมูลการรับรู้จากแบบสอบถาม โดยใช้ Co-occurrence Matrix 2 ตัว สร้างด้วย Jaccard Index ขนาดมิติ 7x7 ตราสินค้า โดย Matrix ข้อมูลจาก Pantip.com สร้างจากชุดคำสำคัญที่ถูกจัดเข้ากลุ่มปัจจัย 8 กลุ่มตามวิธีในข้อที่ 3.2 โดยเอาคำในกลุ่มทั้งหมดมารวมกัน แต่ไม่รวมกลุ่มที่ความหมายไม่เข้าพวก เพื่อให้สามารถเปรียบเทียบกันได้กับข้อมูลจากแบบสอบถาม

3.6 ขั้นตอนการสร้าง MDS Perceptual map

สร้าง Co-occurrence Matrix 9 ชุด ด้วยโปรแกรมภาษา Python แบ่งตามมิติปัจจัยที่มีผลต่อการบริโภคนมจากคำสำคัญที่ความหมายสอดคล้องในแต่ละปัจจัย จำนวน 8 ชุด และ Matrix ที่สร้างจากคำสำคัญทุกคำรวมกัน 1 ชุด คำนวณค่า Co-occurrence ด้วยวิธีการแบบกลุ่มที่ 2 ในข้อที่ 3.4 แต่เปลี่ยนจาก Jaccard Index เป็น Jaccard Distance เมื่อได้ Matrix แล้ว นำเข้าข้อมูลสู่ตัวแบบ MDS เพื่อกำหนดพิกัดตราสินค้าบน Perceptual map ในแต่ละมิติปัจจัย โดยมิติของแกนฉายที่ใช้เป็นแบบ 2 มิติ ผู้วิจัยได้เพิ่มฟังก์ชันการทำ K-Mean Clustering เข้าไปด้วย ใช้วิธีหาจำนวนกลุ่มที่เหมาะสมแบบ Silhouette Score เพื่อช่วยจัดกลุ่มตราสินค้าใน Perceptual Map เพิ่มเติม สาเหตุที่ใช้ Silhouette Score เนื่องจากง่ายต่อการสร้างตัวแบบ เพราะว่าค่า Score ที่ต้องการเป็นค่าสูงสุด แตกต่างจากการใช้ Elbow Method

3.7 ขั้นตอนการสร้างโครงข่ายเชิงความหมาย (Semantic Network)

นำ Co-occurrence Similarity Matrix ขนาดมิติ 500x500 ที่ใช้วิธีคำนวณค่า Jaccard Distance ด้วยวิธีการแบบกลุ่มที่ 1 ในข้อที่ 3.4 จำนวน 1 ชุด มาเป็นข้อมูลตั้งต้นในการสร้าง Semantic Network โดยใช้โปรแกรม R Studio ร่วมกับ Package ชื่อ Bibliometrix และ VOS Viewer ที่พัฒนาขึ้นตามแนวทางของ Visualization of Similarity เหตุผลที่ผู้วิจัยเลือก VOS Viewer เพราะว่าวิธีการแสดงผลดังกล่าวสามารถแก้ไขปัญหาความหนาแน่นของ Vertex และเส้นในกราฟโครงข่ายเมื่อต้องการแสดงผลองค์ประกอบและคำสำคัญจำนวนมาก นอกจากนี้ VOS Viewer ใช้วิธีการทำ Community Detection แบบ Modularity Base ในตัว ซึ่งมีประสิทธิภาพสำหรับกราฟที่มีขนาดใหญ่

4. ผลการวิจัย

4.1 ผลการตอบแบบสอบถาม และ Goodness-of-fit ของตัวแบบ MDS

จากแบบสอบถามที่ได้รับกลับมา มีจำนวน 32 ตัวอย่าง จาก 40 ตัวอย่างที่ได้ส่งไป มีรายละเอียดของ เพศผู้ตอบแบบสอบถาม เป็นชาย 6 คน หญิง 26 คน อยู่ในช่วงอายุ 21-30 ปี มากที่สุด 19 คน ช่วงอายุ 31-40 ปี จำนวน 10 คน ช่วงอายุ 41-50 ปี 2 คน และมากกว่า 51 ปี 1 คน จากตัวอย่างทั้งหมด มีการกรอกข้อมูลในแต่ละตราสินค้าจำนวนแตกต่างกันตามการรับรู้ หากตราสินค้าเป็นที่รู้จักดีจะมีผู้กรอกข้อมูลมาก อย่างไรก็ตาม เนื่องจากข้อมูลบางส่วนไม่สมบูรณ์ จึงตัดตัวอย่างออกไป 2 ชุด เหลือเพียง 30 ชุด เพื่อใช้สร้างตัวแบบ MDS หลังจากได้ตัวแบบแล้วคำนวณค่า Goodness-of-fit ของตัวแบบ ด้วย Kruskal's Stress ได้ค่าเท่ากับ 0.14 ซึ่งอยู่ในเกณฑ์พอใช้ การทดสอบส่วนนี้ดำเนินการด้วยโปรแกรมภาษา Python

4.2 ผลการทำ Robustness check และ Cross Validation

Matrix ที่ได้สร้างขึ้นสำหรับการทำ Robustness check ทั้งหมด ถูกนำมาหาค่าสหสัมพันธ์ QAP ด้วยวิธีการที่กำหนดไว้เพื่อพิจารณา Reliability การทดสอบค่าทั้งหมดดำเนินการในโปรแกรม R Studio เริ่มจากตารางที่ 1 แสดงผลค่าสหสัมพันธ์ระหว่าง Similarity Measure แบบ Jaccard Index และอีก 4 ตัวที่แตกต่างกันเพื่อพิจารณารูปแบบทางเลือก (Alternative form) พบว่า Jaccard Index มีความสัมพันธ์กับรูปแบบทางเลือกในระดับสูงเกือบทั้งหมดที่ระดับนัยสำคัญน้อยกว่า 0.01

สำหรับในตารางที่ 2 เป็นการคำนวณค่าสหสัมพันธ์ระหว่าง Jaccard Index สร้างจากเอกสารที่ถูกสุ่มแล้วใส่คืนมาในสัดส่วนแตกต่างกัน 4 แบบ เทียบกับ Matrix ที่ถูกสร้างจากเอกสารทั้งหมด พบว่ามีความสัมพันธ์กันในระดับสูงเกือบทั้งหมดที่ระดับนัยสำคัญน้อยกว่า 0.01

ตารางที่ 1 ผลสหสัมพันธ์ QAP แบบ Alternative form ของ Similarity Measure ที่แตกต่างกัน

	Jaccard	Lift	Inclusion	Cosine	Association
Jaccard Index	1.0000				
Lift	0.5777 *	1.0000			
Inclusion Index	0.7114 *	0.6883 *	1.0000		
Cosine Similarity	0.64 (R) *	0.2989 (R) ***	0.3299 (R) ***	1.0000	
Association Strength	0.9264 *	0.6366 *	0.842 *	0.4598 (R) **	1.0000

ตารางที่ 2 ผลสหสัมพันธ์ QAP ระหว่าง Similarity Matrix ที่ถูกสร้างจากเอกสารในสัดส่วนแตกต่างกัน

	1/16 (สุ่ม)	1/8 (สุ่ม)	1/4 (สุ่ม)	1/2 (สุ่ม)
1/1 (ทุกคำสำคัญ)	0.7531 *	0.5618 *	0.7722 *	0.9642 *

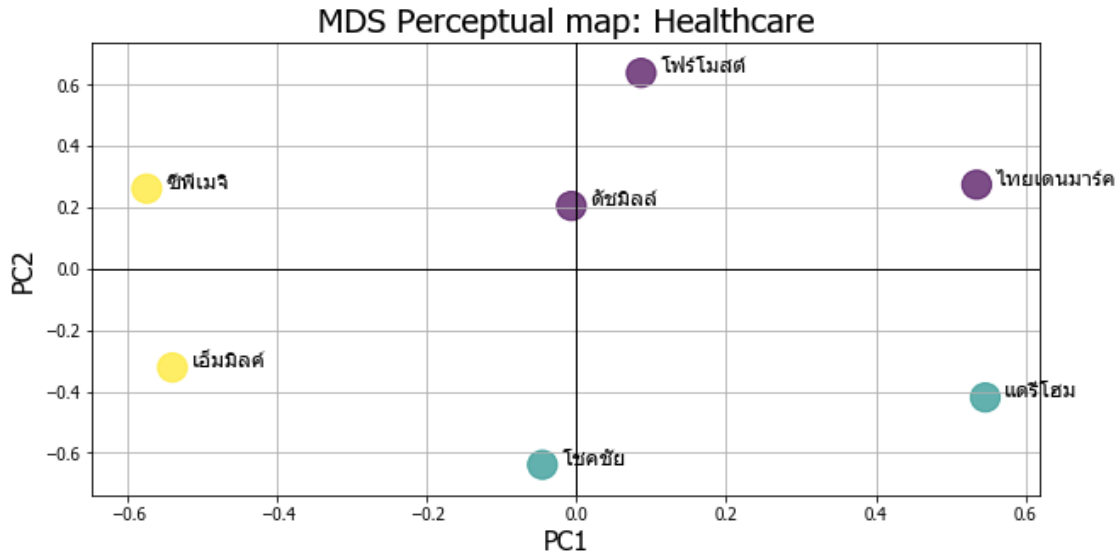
* p-value sig ที่ <= 0.01, ** p-value sig ที่ <= 0.05, *** p-value sig ที่ <= 0.10

ในส่วนการทำ Cross Validation ระหว่าง Matrix จากแบบสอบถามและ Pantip.com ที่ได้คัดเลือกชุดคำสำคัญจัดกลุ่มตามปัจจัย เพื่อให้สามารถเปรียบเทียบกันได้ ได้ค่าสหสัมพันธ์ QAP = 0.5642 ค่า p-value = 0.011 ซึ่งมีความสัมพันธ์ระหว่างกันในระดับกลางค่อนข้างสูงอย่างมีนัยสำคัญ แต่ถ้ารวมกลุ่มที่ความหมายไม่เข้าพวกเข้าไปใน Matrix ข้อที่ 3.5.3 ส่วนที่สร้างด้วยข้อมูลการรับรู้จาก Pantip.com เมื่อเทียบกับแบบสอบถามแล้ว มีผลสหสัมพันธ์ QAP สูงขึ้นอีกอย่างมีนัยสำคัญ เท่ากับ 0.6681 และค่า p-value = 0.005

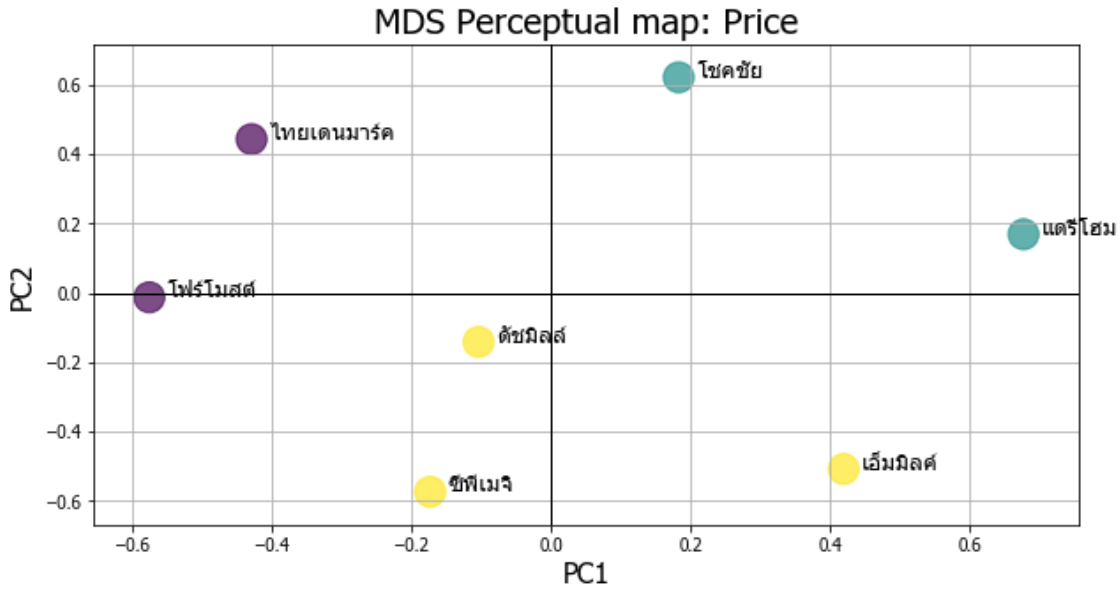
4.3 MDS Perceptual map

การสร้าง Perceptual map ตำแหน่งทางการตลาด 8 มิติจาก Dissimilarity Matrix ของข้อมูลจาก Pantip.com ร่วมกับการใช้ K-Mean Clustering ในการจัดกลุ่มตราสินค้า ผู้วิจัยใช้โปรแกรมภาษา Python สร้างตัวแบบขึ้น ได้

ตัวอย่างผลลัพธ์ดังรูปที่ 2 และรูปที่ 3



รูปที่ 2 MDS Perceptual map จาก Pantip.com ในมุมมองปัจจัยด้านสุขภาพ

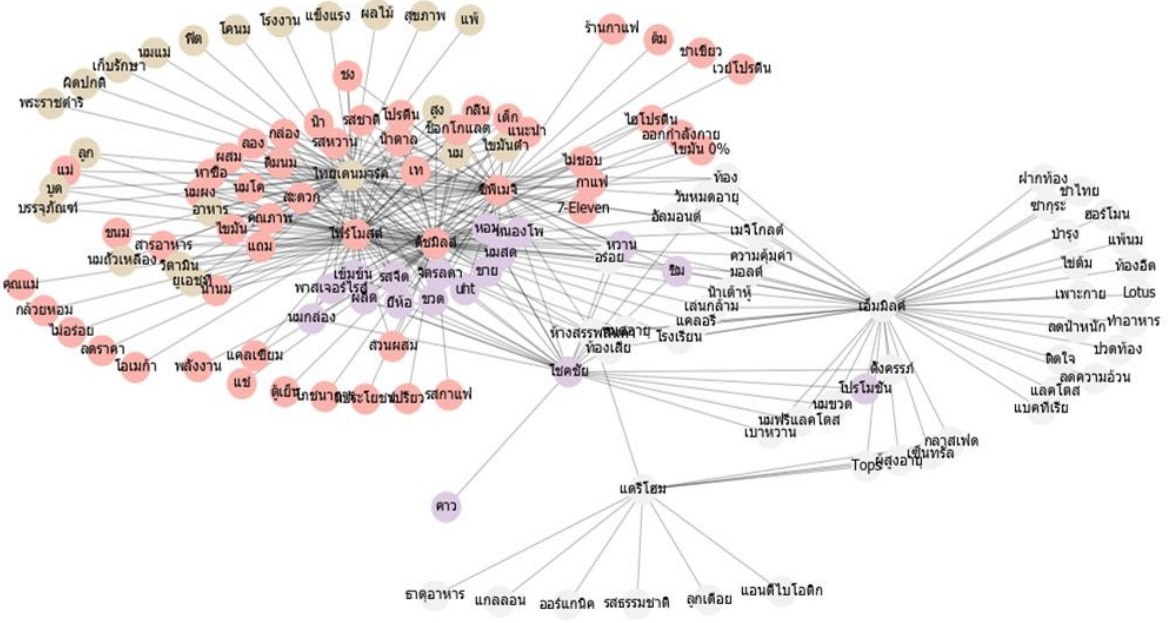


รูปที่ 3 MDS Perceptual map จาก Pantip.com ในมุมมองปัจจัยด้านราคา

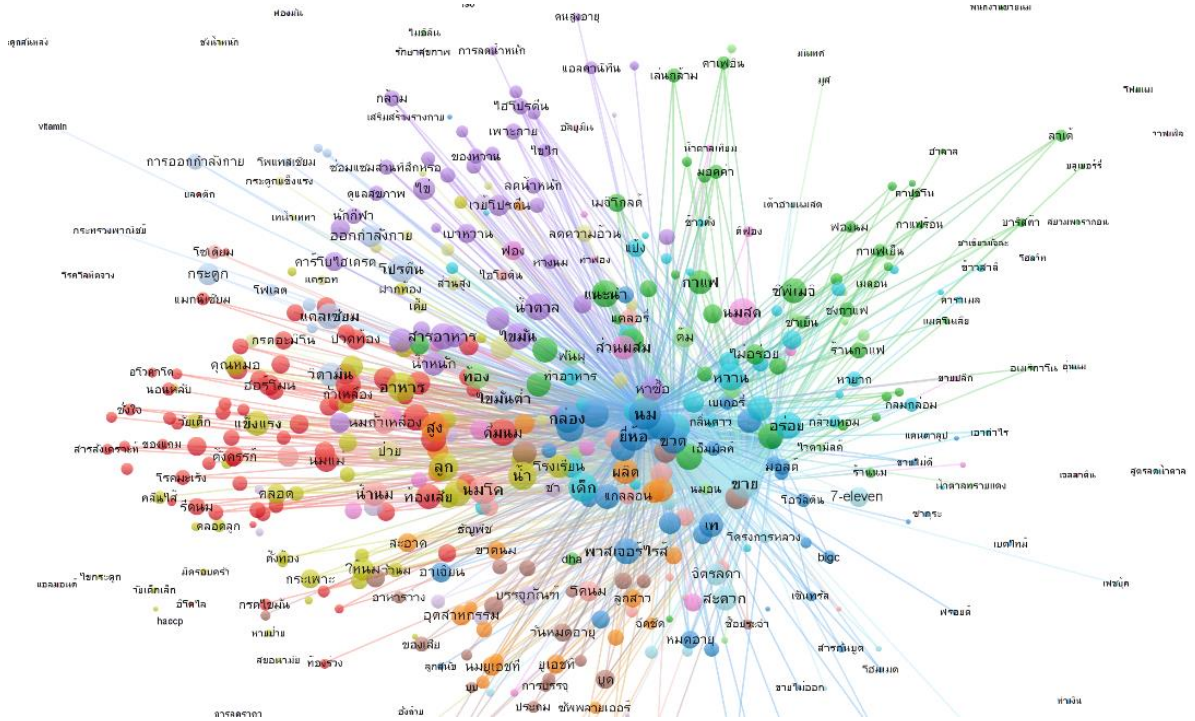
4.4 Semantic Network และ Community Detection

การสร้าง Semantic Network ของข้อมูลจาก Pantip.com ใช้ Similarity Matrix ขนาด 500x500 นำมาสร้างกราฟโครงข่าย 2 แบบ คือ แบบ Zoom-in เจาะจงเฉพาะตราสินค้า และแบบแสดงผลทุกคำสำคัญ ในแบบที่ 1 Zoom-in เจาะจงเฉพาะตราสินค้า มีผลลัพธ์ดังรูปที่ 4 ซึ่งสร้างจาก Library ชื่อว่า NetworkX ในโปรแกรมภาษา Python เพื่อแสดงผลองค์ประกอบร่วมแบบเจาะจงตราสินค้า โดยใช้ Layout แบบ Kamada-Kawai

อย่างไรก็ตาม การแสดงผลแบบ Zoom-in นั้นมีปัญหาที่สำคัญ หากต้องการแสดงผลกราฟโครงข่ายให้ครบทุกเส้นเพื่อให้ได้ความเชื่อมโยงที่สมบูรณ์ จำนวน Vertex และเส้นที่ปรากฏนั้นมีจำนวนมากซ้อนทับกันอย่างหนาแน่นจนไม่สามารถเข้าใจข้อมูลที่ปรากฏได้ ผู้วิจัยจึงได้เลือกใช้ VOS Viewer ซึ่งเป็นเครื่องมือที่สามารถแสดงผลกราฟโครงข่ายขนาดใหญ่ ดังตัวอย่างในรูปที่ 5



รูปที่ 4 Zoom-in Network Graph



รูปที่ 5 VOS Viewer Network Graph

จากรูปที่ 5 เป็นการแสดงผลบน VOS Viewer ที่มีการแบ่งสีตาม Community สามารถเลือกแสดงชนิดของเส้น (เส้นโค้ง-ตรง) หรือกรองจำนวนเส้นที่ปรากฏบน Canvas ได้ หากมองไม่ชัดว่าองค์ประกอบที่มีขนาดเล็กหมายถึงอะไร ผู้ใช้สามารถขยายมุมมองเพื่อดูได้ หรือ นำเมาส์ไปชี้ที่ Vertex ที่ต้องการให้ VOS Viewer แสดงเฉพาะเส้นและ Vertex อื่นที่เชื่อมต่อด้วย นอกเหนือจากความสัมพันธ์ในโครงข่าย ผู้วิจัยได้แจกแจงรายละเอียดองค์ประกอบที่ถูกจัดกลุ่มด้วยวิธี Community Detection เพื่อพิจารณาความสอดคล้องกับปัจจัยที่ได้จากการทบทวนวรรณกรรม และเปิดเผยข้อมูลเชิงลึกที่ซ่อนอยู่ในกลุ่มที่ความหมายไม่เข้าพวก โดยมีรายละเอียดดังตารางที่ 3

ตารางที่ 3 สรุป Community ที่ค้นพบ

ลำดับที่	ชื่อ Community	จำนวนสมาชิก
1	สุขภาพและโรคภัย	70
2	ผสมชาและกาแฟ	53
3	การขายในห้างสรรพสินค้า	52
4	แม่และเด็ก	46
5	เพาะกาย และโภชนาการของการเพาะกาย	45
6	รสชาติและส่วนผสม	43
7	นมกับการเป็นอาหารเสริมสำหรับเด็ก	30
8	คุณภาพการผลิต กับการเก็บรักษา	30
9	ราคาถูก ความสะอาด ออร์แกนิค	21
10	แหล่งโปรตีนอื่นที่นอกจากนม	21
11	ร้านค้าส่ง & ร้านของชำ	20
12	การบำรุงกระดูก	19
13	ลดความอ้วน	18
14	การส่งเสริมการขาย	15
15	ร้านสะดวกซื้อ	15

VOS Viewer ได้จัด Community จำนวน 15 กลุ่ม โดยตัดองค์ประกอบที่มีความสัมพันธ์น้อยผิดปกติ (Outlier) 2 คำสำคัญออก เหลือ 498 ชุดคำ ปัจจัยที่ได้จากการทบทวนวรรณกรรมนั้นปรากฏครบถ้วน แต่มีการแบ่งกลุ่มย่อยเพิ่มเติมอีก ซึ่งเป็นประโยชน์ในทางธุรกิจ ขอยกตัวอย่างคำสำคัญในบางกลุ่มดังต่อไปนี้

1. ตัวอย่างคำใน Community แม่และเด็ก เช่น โรงเรียน, ให้นม, คุณแม่, คุณหมอ, คลอด, ลูกชาย, กระเพาะ, เครียด, น้ำอัดลม, มีลูก, เลี้ยงลูก, ความสูง, วัยเด็ก, วัวนม, โรงพยาบาล, เด็กโต, ทำกิจกรรม, คลื่นไส้, มื้อเช้า, ตั้งท้อง, คลอดลูก, ฝากท้อง, เป็นหวัด, กลางดึก, ระบบขับถ่าย, ไข่ดาว, แพ้ท้อง, ฝากครรภ์, ตัวเตี้ย, มีครอบครัว, หายป่วย, ประกอบอาหาร, วัยเด็กเล็ก, ฟื้นฟูสภาพ, สุขอนามัย, ไขกระดูก, แม่ลูกอ่อน, เด็กอ่อน

2. ตัวอย่างคำใน Community เพาะกาย และโภชนาการ เช่น สารอาหาร, ไขมัน, โภชนาการ, พลังงาน, น้ำหนัก, ไข่, ไขมัน 0%, เวย์โปรตีน, แบคทีเรีย, คาร์โบไฮเดรต, เบาหวาน, กล้ามเนื้อ, เพาะกาย, ความหวาน, นักกีฬา, ลดน้ำหนัก, ไฮโปรตีน, กล้าม, ความคุ้มค่า, ของหวาน, ดูแลสุขภาพ, หางนม, ซ่อมแซมส่วนที่สึกหรอ, กลูโคส, ออกกำลัง, เล่นเวท, ปริมาตร, ไข่ขาว, ฟิตเนส, ไข่ไก่, แอลคานีทีน, อกไก่, เพิ่มน้ำหนัก

5. การอภิปรายผลและสรุป

งานวิจัยนี้ได้นำเสนอวิธีการสำรวจตลาดผ่านสื่อสังคมออนไลน์อย่าง Pantip.com โดยนำเสนอกราฟโครงข่ายและแผนผังการรับรู้ซึ่งแสดงตำแหน่งการตลาดของตราสินค้า สำหรับตำแหน่งทางการตลาดที่สร้างด้วย MDS ในขั้นตอน Cross Validation ระหว่างข้อมูลจากกระทู้ในเว็บไซต์กับข้อมูลจากแบบสอบถาม พบว่าอาจมีมุมมองหรือปัจจัยอื่นเพิ่มเติมที่ไม่ปรากฏในการทบทวนวรรณกรรมอีก ซึ่งอาจเกิดจากกิจกรรมตลาดที่เปลี่ยนแปลงตลอดเวลา หรือเลือกคำสำคัญไม่ครบถ้วนพอ เพราะว่าเมื่อทดลองเพิ่มชุดคำสำคัญที่ความหมายไม่เข้าพวกเข้าไปรวมกับชุดคำสำคัญ 8 กลุ่มที่มีอยู่เดิมเพื่อเป็นข้อมูลตั้งต้นในการทดสอบ ค่าสหสัมพันธ์นั้นสูงขึ้นอย่างมีนัยสำคัญ นอกจากนี้ยังพบว่าตำแหน่งทางการตลาดตราสินค้าในปัจจัยด้านส่วนผสม ด้านแม่และเด็ก และด้านสุขภาพ ตราสินค้าทั้งหมดถูกแบ่งกลุ่มได้เป็นมิติละ 3 กลุ่ม มีตราสินค้าในแต่ละกลุ่มเหมือนกันทั้งสามมิติ สามารถตีความได้ว่า สมาชิกในกลุ่มกำลังแข่งขันกัน

ภายในมิติมุมมองดังกล่าว ซึ่งสอดคล้องกับข้อมูลเชิงลึกของผู้เชี่ยวชาญจากบริษัทนม บริษัทนมต้องกำหนดกลยุทธ์อย่างระมัดระวัง เช่น การอบรมการแนะนำสินค้าให้กับพนักงานแนะนำสินค้าในมิติที่กำลังแข่งขันกันอยู่อย่างเข้มข้น การป้องกันการซื้อตัวพนักงานแนะนำสินค้าโดยบริษัทที่กำลังแข่งขันอยู่ หรือการแย่งชิงพื้นที่ขายเพื่อเข้าถึงกลุ่มเป้าหมายตามมิติตำแหน่งตราสินค้าที่ได้สรุปไว้

สำหรับ Semantic Network และกระบวนการ Community Detection หลังจากอภิปรายผลร่วมกับผู้เชี่ยวชาญ พบว่าข้อมูลเชิงลึกใหม่ที่ปรากฏ เผยให้เห็นเหตุการณ์ในตลาด กับองค์ประกอบคำสำคัญหลายคำที่บางประเด็นเป็นสิ่งที่ไม่คาดคิดว่าจะพบเห็น ทั้งหมดดำเนินการโดยไม่ต้องใช้ข้อมูลจากการทำแบบสอบถาม

เมื่อเปรียบเทียบงานวิจัยในประเทศไทยก่อนหน้านี้ งานวิจัยของ ชื่นสุมล และ บุษราภรณ์ (2557) ดำเนินการสร้างตำแหน่งทางการตลาดตราสินค้าโดยใช้ MDS ด้วยแบบสอบถาม พบว่าเมื่อใช้ข้อมูลจากสื่อสังคมออนไลน์ได้ข้อมูลที่มีจำนวนมิติมุมมองมากกว่าการทำแบบสอบถาม เพราะการใช้แบบสอบถามทำให้แนวทางการกำหนดองค์ประกอบในตลาดมีโอกาสกำหนดได้ไม่ครบถ้วน เนื่องจากแต่ละบริษัทมีการคิดค้นสินค้าใหม่หรือกิจกรรมส่งเสริมการขายใหม่อยู่เสมอ เมื่อเทียบงานของ Netzer et al. (2012) ที่ได้ดำเนินการวิจัยในลักษณะคล้ายกันกับงานนี้ โดยมีกระบวนการทำ Cross Validation จากแบบสอบถามเหมือนกัน มีค่าสหสัมพันธ์ QAP เท่ากับ 0.43 พบว่าผลลัพธ์ของงานวิจัยอยู่ในเกณฑ์ดีกว่าของ Netzer et al. ไม่มากนัก ซึ่งอาจเป็นผลมาจากการเลือกชุดคำเพื่อสร้าง Matrix เปรียบเทียบนั้นไม่ครบถ้วน แต่ว่าการศึกษานี้สามารถแก้ปัญหาการแสดงผลกราฟที่หนาแน่น ยากที่จะตีความได้ และปรับปรุงประสิทธิภาพของกระบวนการ Community Detection แบบ Girvan-Newman ในการศึกษาของ Netzer et al. ที่กินเวลาและทรัพยากรประมวลผลมากกว่าเมื่อกราฟมีขนาดใหญ่และหนาแน่นด้วยการใช้ VOS Viewer ร่วมกับกระบวนการ Community Detection แบบ Modularity Base

6. ข้อจำกัดและข้อเสนอแนะ

อย่างไรก็ตาม งานวิจัยยังมีข้อจำกัด การใช้ข้อมูลจาก Pantip.com มีความเห็นในเชิงเปรียบเทียบค่อนข้างน้อย ผู้ใช้ระบบมักไม่ค่อยใส่ชื่อตราสินค้าลงในช่องความเห็น ส่วนมากจะใส่ในหัวกระทู้ ทำให้แยกชื่อตราสินค้าได้ยาก นอกจากนี้ การแยกประเภทของ Domain นมพาสเจอร์ไรส์และตราสินค้า ทำเพียงกรองข้อความด้วยเงื่อนไข If-Else แบบตรงไปตรงมา ไม่ได้ใช้เทคนิคการเรียนรู้ของเครื่องจักร (Machine Learning) เช่นเดียวกับการคัดเลือกคำเข้าแต่ละกลุ่มปัจจัยที่ทำด้วยมือ ไม่ได้ใช้ตัวแบบหัวข้อ (Topic Modeling) รวมถึงการสุ่มตัวอย่างด้วยวิธีสุ่มแบบตามสะดวกร่วมกับการสุ่มแบบเจาะจง เนื่องจากข้อจำกัดของการเข้าถึงกลุ่มตัวอย่างเพื่อสัมภาษณ์ในพื้นที่โรงพยาบาลช่วงโรค Covid-19 ระบาด ทำให้ตัวอย่างยังมีความหลากหลายไม่เพียงพอ มีอคติ (Bias) ของการสุ่ม และการ Normalized ความสัมพันธ์ระหว่างองค์ประกอบคำสำคัญ ไม่ได้ควบคุมอคติอื่นนอกจากขนาดตลาด

สำหรับการวิจัยในอนาคต ในการหาตำแหน่งตราสินค้า สามารถพิจารณาเพิ่มช่องทางสื่อสังคมออนไลน์อื่นเพิ่มตราสินค้าให้ครอบคลุมทั้งตลาด เพิ่มมิติมุมมอง และเพิ่มประเภทสินค้านมให้ครอบคลุมถึงนมเปรี้ยว โยเกิร์ต วิปครีม นมถั่วเหลือง และนมแพะ หรืออาจพิจารณาสร้างตำแหน่งลงลึกถึงหน่วยผลิตภัณฑ์ในตลาดซึ่งต้องใช้เทคนิคการเรียนรู้ของเครื่องจักรช่วยในการจำแนกข้อความ เพื่อให้ข้อมูลมีความถูกต้องต่อการหาตำแหน่งตราสินค้าต่อไป

7. กิตติกรรมประกาศ

ขอขอบพระคุณ คุณนิธิมา พลชัย ผู้จัดการทั่วไปบริษัทซีพี-เมจิ จำกัด ที่ได้เอื้อเฟื้อข้อมูลและแนวคิดที่เกี่ยวข้อง อีกทั้งยังสนับสนุนกิจกรรมต่างๆ เป็นอย่างดีมาตลอดจนงานวิจัยชิ้นนี้สำเร็จลุล่วงไปด้วยดี

8. เอกสารอ้างอิง

ชูชัย สมิทธิไกร. (2562). **พฤติกรรมผู้บริโภค.** กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ชื่นสุมล บุนนาค, บุษราภรณ์ อารีย์. (2557). การวิเคราะห์ตำแหน่งตราสินค้าของผลิตภัณฑ์กระดาษเช็ดหน้าโดยใช้เทคนิคการแบ่งสเกลแบบหลายมิติ. **Journal of Business, Economics and Communications,** 9(2), 27-38.

ปิยฉัตร ช่างเหล็ก, ยอดมนี เทพานนท์, ณัฐพล พันธุ์ภักดี. (2561). ปัจจัยที่มีผลกับทัศนคติต่อการซื้อนมพาสเจอไรส์ ของผู้บริโภคในเขตกรุงเทพมหานคร. **การประชุมวิชาการเสนอผลงานวิจัย ระดับชาติและนานาชาติ,** 9(1), 935-946.

รังสรรค์ เลิศในสัตย์. (2549). **การตลาดเชิงกลยุทธ์เพื่อความสำเร็จสำหรับผู้บริหาร SME.** กรุงเทพฯ: สมาคมส่งเสริมเทคโนโลยี (ไทย-ญี่ปุ่น).

สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์. (2563). **รายงานผลสำรวจพฤติกรรมผู้ใช้อินเทอร์เน็ตในประเทศไทย ปี 2562.** กรุงเทพฯ: ผู้แต่ง.

สุภชาติ ชัยณรงค์สิงห์. (2539). **การศึกษาปัจจัยที่ส่งผลการบริโภคนมเปรี้ยวพร้อมดื่ม.** วิทยานิพนธ์ปริญญาบริหารธุรกิจมหาบัณฑิต สาขาวิชาบริหารธุรกิจ มหาวิทยาลัยเกษมบัณฑิต.

Aaker, D. A. (1996). **Building Strong Brands.** Free Press, New York.

Blondel, V., Guillaume, J. L., Lambiotte, R., Lefebvre, E. (2008). Fast Unfolding of Communities in Large Networks. **Journal of Statistical Mechanics: Theory and Experiment,** 2008(10). doi:10.1088/1742-5468/2008/10/P10008

De Alwis, A., Edirisinghe, J., Athauda, A. (2011). Analysis of Factors Affecting Fresh Milk Consumption among the Mid-Country Consumers. **Tropical Agricultural Research and Extension,** 12(2), 103-109. doi:10.4038/tare.v12i2.2799

Doerfel, M. (1998). What Constitutes Semantic Network Analysis? A Comparison of Research and Methodologies. **Connections,** 21(2). 16-26.

Henderson, G. R., Iacobucci, D., Calder, B. J. (1998). Brand diagnostics: Mapping branding effects using consumer associative networks. **European Journal of Operational Research,** 111(2), 306-327.

Qin, H. (1999). Knowledge Discovery through Co-Word Analysis. **Library Trends,** 48(1). 133-59.

Howard, L. W. (1998). Validating the Competing Values Model as a Representation of Organization Cultures. **The International Journal of Organization Analysis,** 6(3), 231-250.

Hsu, J. L., & Lin, Y. T. (2006). Consumption and attribute perception of fluid milk in Taiwan. **Nutrition & Food Science,** 36(3), 177-182.

Kurajdova, K., Petrovicova, J. T., Kascakova, A. (2015). Factors Influencing Milk Consumption and Purchase Behavior - Evidence from Slovakia. **Procedia Economics and Finance,** 34, 573-580. doi:10.1016/S2212-5671(15)01670-6

Lee, S. J. (2017). Improving Jaccard Index for Measuring Similarity in Collaborative Filtering. **International Conference on Information Science and Applications,** 2017, 799-806.

Lee, T., Bradlow, E. (2011). Automated Marketing Research Using Online Customer Reviews. **Journal of Marketing Research,** 48(5), 881-894.

Malhotra, N. K. (2019). **Marketing research: An applied orientation** (7th ed.). New York, NY: Pearson.

Malhotra, P., Bhattacharyya, S. (2019). Online Brand Networks: A New Approach to Brand Positioning. **SSRN Electronic Journal,** doi:10.2139/ssrn.3484520

Netzer, O., Feldman, R., Goldenberg, J., Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. **Marketing Science,** 31(3), 521-543.

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. **Physical Review E,** 69(6), 066133.

Peter, J. P., Olson, J. C., (2010). **Consumer Behavior and Marketing Strategy** (9th ed.). McGraw-Hill.

Phatthiyaphaibun, W., Chaovavanich, K., Polpanumas, C., Suriyawongkul, A., Lowphansirikul, L. & Chormai, P. (2016). PyThaiNLP: Thai Natural Language Processing in Python. **Zenodo,** doi:10.5281/zenodo. 3519354

Ringel, D., Skiera, B. (2016). Visualizing Asymmetric Competition among More Than 1,000 Products Using Big Search Data. **Marketing Science,** 35. doi:10.1287/mksc.2015.0950

Rodgers, J. L. (1991). Matrix and Stimulus Sample Sizes in the Weighted MDS Model: Empirical Metric Recovery Functions. **Applied Psychological Measurement,** 15(1), 71-77.

Sievert, C., Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. Workshop on Interactive Language Learning. **Visualization, and Interfaces,** 63-70. doi:10.13140/2.1.1394.3043.

Simpson, W. (2001). The Quadratic Assignment Procedure (QAP). **North American Stata Users Group Meetings 2001,** 1.2.

Tabassum, S., Pereira S. F. F., Fernandes S., Gama J. (2018). Social network analysis: An overview. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,** 8(5). doi:10.1002/widm.1256.

Van Eck, N. J., Waltman, L. (2006). VOS: a new method for visualizing similarities between objects. **Advances in Data Analysis: Proceedings of the 30th Annual Conference of the German Classification Society,** 299-306. doi:10.1007/978-3-540-70981-7_34

Van Eck, N. J., Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. **Scientometrics,** 84, 523-538. doi:10.1007/s11192-009-0146-3

Vemprala, N., Xiong, R. R., Liu, C. Z., Choo, K. K. R. (2019). Where Does My Product Stand? A Social Network Perspective on Online Product Reviews. **52nd Hawaii International Conference on System Sciences,** 2345-2354.

Waltman, L., Van Eck, N. J., Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. **Journal of Informetrics,** 4(4), 629-635.

Yin, S., Chen, M., Chen, Y., Xu, Y., Zou, Z., & Wang, Y. (2016). Consumer trust in organic milk of different brands: the role of Chinese organic label. **British Food Journal,** 118(7), 1769-1782.

การวิเคราะห์ผลจากประตูโดยผู้เล่นต่างชาติต่อผลการแข่งขันฟุตบอลด้วยตัวแบบการถดถอยลอจิสติกปัวซงและตัวแบบการถดถอยลอจิสติกอันดับ: กรณีศึกษาการแข่งขันไทยพรีเมียร์ลีก

ณัฐนันท์ วิจิตรอักษร*[1] ศาสตริน วงศ์จีระศักดิ์[2] และ เกศินี ธารีสังข์[3]

Received: 28 August 2021; Revised: 24 November 2021; Accepted: 1 December 2021

บทคัดย่อ

ผู้เล่นฟุตบอลชาวต่างชาติเป็นปัจจัยหนึ่งที่มีผลต่อความสำเร็จในแง่การเพิ่มความแข่งขันภายในทีมและการสนับสนุนเพื่อนร่วมทีมทั้งในระดับชาติและระดับสโมสร การศึกษานี้มีวัตถุประสงค์เพื่อวิเคราะห์ผลจากประตูที่ได้โดยผู้เล่นชาวต่างชาติในการแข่งขันฟุตบอลไทยพรีเมียร์ลีกซึ่งจัดอยู่ในอันดับที่ 8 ของเอเชีย ผลการศึกษาพบว่าจำนวนประตูที่ได้โดยผู้เล่นชาวต่างชาติส่งผลต่อผลการแข่งขันอย่างมีนัยสำคัญทางสถิติทั้งในรูปของจำนวนประตูที่ได้หรือผลการแข่งขันชนะ-เสมอ-แพ้ นอกจากนี้ ความได้เปรียบของการเป็นเจ้าบ้านยังส่งผลต่อผลการแข่งขันอย่างมีนัยสำคัญทางสถิติ ขณะเดียวกัน เมื่อวิเคราะห์จำนวนประตูที่ได้โดยผู้เล่นในประเทศพบว่าจำนวนประตูดังกล่าวเป็นปัจจัยเสริมที่มีผลต่อความสำเร็จในการแข่งขันเช่นกัน

คำสำคัญ: ผลการแข่งขัน, ผู้เล่นต่างชาติ, ไทยพรีเมียร์ลีก, ปัวซงอิสระ, ลอจิสติกอันดับ

* *Corresponding E-mail: nuttanan.wichitaksorn@aut.ac.nz*

[1] *Department of Mathematical Sciences, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand*

[2,3] *สถาบันวิจัยเพื่อการพัฒนาประเทศไทย, กรุงเทพฯ 10310*

Assessing Impact of Goals by Foreign Players on Football Match Outcomes using Poisson and Ordered Logistic Regression Models: An Evidence from Thailand Premier League

Nuttanan Wichitaksorn[*1], Sardtarin Wongjirasak[2], and Kaesinee Tharisung[3]

Received: 28 August 2021; Revised: 24 November 2021; Accepted: 1 December 2021

Abstract

Foreign football players have been widely accepted for their competitive and complementary effects on the success of both national and club levels. This study investigates the contributions from the goals scored by the foreign players in the Thailand premier league, which is a top-eight league in Asia. Using the match- level data, the goals by foreign players have a strong and positive impact on the match outcomes either as the team's total number of goals or the win-draw- lose results. In addition, home advantage also has the considerable positive and negative effects, respectively on the match outcomes. The similar analysis is also performed for the case of goals by local players and reveals the complementary effect of the goals by foreign players.

Keywords: match results, foreign players, Thailand premier league, independent poisson, ordered logit.

[*] *Corresponding E-mail: nuttanan.wichitaksorn@aut.ac.nz*
[1] *Department of Mathematical Sciences, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand*
[2,3] *Thailand Development Research Institute, Bangkok 10310, Thailand*

1. Introduction

The professional football league in Thailand was first introduced in 1996 by the Football Association of Thailand (FAT). Football is one of the most popular sports in Thailand but in the early years its league was not very successful. In 2008, the Asian Football Confederation (AFC) announced "Vision Asia" aiming to raise the standards of Asian football at all levels and link football with the business engagement. As a result, FAT strongly promoted and supported the development of professional football clubs. That helped generate financial flows to the clubs so that they can support the players and build fan bases. This becomes the turning point for the football industry in Thailand where the number of supporters increased dramatically and the clubs can have more flexible strategies to improve and compete in the league. One of the key strategies that might underlie the success of football clubs is recruiting foreign players into their team. In many cases, foreign players are believed to have higher quality and superior to Thai players.

As many football clubs in Thailand realized the benefit of having foreign players in their squad, the number of foreign football players increased dramatically over the years since 2008. However, the policy to limit number of foreign players becomes effective in 2015 where football clubs are allowed to have a quota of five foreign players in their team. This policy aims to provide more chances for Thai players and raises the domestic standards. Football clubs, on the other hand, have to be selective and recruit only quality foreign players in response. It is widely perceived that the foreign players contribute significantly to the team performance, especially, through the match results. This is our main aim to investigate in this study.

The role of foreign players has been investigated in both national and club levels. Royuela & Gásquez (2019) found a majority of prior research focused on the effect of foreign players on the performance of the national team where the benefit from spillover effect was discovered. Hence, they analyzed the influence of foreign players in the club level using the model previously developed and implemented by Bernard & Busse (2004) and Gásquez & Royuela (2016). Their model used the Elo rating score as a proxy for football success that implies the club performance. While controlling the effects from economic, demographic, institutional, and other factors, Royuela & Gásquez (2019) included the proportion of foreign players to investigate their impact and found the foreign players contributed to the success of club performance.

In this study, we analyze the role of foreign players deeper in the game level. Precisely, we quantitatively assess that role through the number of goals scored by the foreign players and its contribution to the match results including the total number of goals and the win-draw-lose status. We also control the effects from the yellow and red cards, the home advantage, and the first-goal status on the match results, which, to our knowledge, were rarely considered previously. To make a fair comparison, we also perform the similar assessment for the number of goals scored by the native or local players. This can help us to determine whether the impact from the goals by foreign players is substitutionary or complementary.

Our study is not only the first that attempts to examine the role of foreign players from the game-level perspective, using the number of goals, on the match results, but it also analyzes the game data from a southeast Asian nation with the flourishing football league. Note that in 2020 Thailand premier league was ranked the fourth in the east region or the eight in Asia by AFC. Using the Poisson and ordered logit regression models, respectively, for the total number of goals and the win-draw-lose status, our results suggest that the number of goals scored by the foreign players and the home advantage have the significantly positive impact on the match results. The two regression models also allow us to obtain the marginal probabilities of total goals with respect to each goal scored by foreign players.

The organization of this paper is as follows: Section 2 reviews the relevant literature and provides the conceptual background. Section 3 discusses the data and methodology. Section 4 presents the results. Section 5 concludes.

2. Literature Review and Conceptual Background

The internationalization of football players has a long history but not until the late twentieth century where the influx of foreign players increased dramatically, especially in England. The growing literature on the effect of foreign players reflected this phenomenon. As shown and reviewed by Royuela & Gásquez (2019) and references therein, the majority of prior works mainly investigated the benefits of importing foreign players on the national team performance.

Complementarity from skills and competition as a better choice are considered to be the major benefits of having foreign players. Many studies concluded that the foreign players had the positive effect on the performance either in the national or in the club level, see Baur & Lehmann (2007), Binder & Findlay (2012),

Berlinschi, Schokkaert, & Swinnen (2013), and Royuela & Gásquez (2019), among others. However, the effect of goals scored by foreign players on the game level or the match results has been rarely investigated.

Analysis on the match results (including goal counts) and the determinants has been well implemented, see Maher (1982), Dixon & Coles (1997), Dixon & Robinson (1998), Karlis & Ntzoufras (2003), Brechot & Flepp (2020), and Pearson, Livingston, & King (2020), among others. Goddard (2005) developed two structural regression models for football forecasting and compare the effectiveness in predicting match outcomes. The first model estimates the number of goals scored or conceded and implies that as the match outcome and team performance. This model treats the number of goals as a count variable that can be predicted using the bivariate Poisson regression. The second model, on the contrary, estimates the match results in the ordinal manner, which is classified as win, draw, and lose. Econometrically, this is a limited dependent model that can be estimated using the ordered regressions such as logit and probit, see e.g. Forrest & Simmons (2000), Goddard & Asimakopoulos (2004), and Forrest, Goddard, & Simmons (2005).

Apart from the ordinal models, there are two other models that are widely mentioned in the field of sport econometrics; the Maher-Type model and the point process model (McHale & Baker, 2014). In the Maher-Type model, the model accounts for the attacking and defending abilities of the team and assumes the number of goals by home and away team is the independent Poisson random variables, which are determined by attacking and defending abilities. The number of goals is predicted based on the ability of the team itself interacting with the opponent's factors, and construct to be a likelihood to score. However, the model entails some shortcomings, which relate to whether the assumptions hold, and the model is only able to capture a static situation. Dixon & Coles (1997) developed a model to overcome these shortcomings by adding a parameter to effectively inflate the probabilities of certain scores that occurred. It gives more dynamic to the estimation by putting more weight in the likelihood of the matches occurring more recently. Developed by Dixon and Robinson (1998), the point process model considers the changing game to predict the match outcome at any point during the game. However, this model is mainly essential for the professional analysis and the betting market as it requires large sets of data including the timing of the events.

There remains an issue regarding the model selection between individual and bivariate Poisson models. The bivariate Poisson model was proposed by Karlis & Ntzoufras (2003) to allow the correlation of scores by the two opposing teams so that the prediction power can be improved. However, Groll, Schauberger, & Tutz (2015) found that if the highly informative covariates of both teams, e.g., home status, are included, the independent Poisson model with the linear predictors might already be sufficient. In addition, Pearson, Livingston, & King (2020) compared three different models including the independent Poisson, Dixon & Coles (1997), and bivariate Weibull count models for the outcome prediction for five leagues from five different countries. They found that the independent Poisson model outperformed the other two models in three leagues.

In this study, we use two approaches from Goddard (2005) to model the match outcomes from the goals and the win-draw-lose results. However, in the first approach, we use the independent Poisson regression model for the number of goals scored and conceded by each team where the home status is included as a covariate. While in the second approach, we use the ordered logit model for the three match results. Since our main objective to investigate the effect or the contribution of foreign players on the match outcomes, we also include the number of goals scored by foreign players as a covariate. In addition, the numbers of yellow and red cards are included as the covariates because they are often quoted as an adverse effect on the match outcomes. First goal is only an additional covariate to assess its effect on the match results in the ordered logit model.

3. Data and Methodology

3.1 Data

The data we used in the analysis are the match outcomes (number of goals and win-draw-lose results) and other relevant information from the Thailand premier league (TPL) in 2015. This is one of the peak years for TPL in which the complete and stable results can be obtained. (For example, in the 2016 season there were some technical issues that caused the suspension of the competition after 31 matches played. After that the TPL popularity has been decreasing, partly due to the national team's worsening performance.) In 2015, the league consists of 18 participating teams where three of them were promoted from the lower division. Throughout the season, the teams played in home-away matches that return 306 matches in total or 612 observations for the analysis. Note that the data were collected from the match fixtures provided by online newspaper websites and were cross-checked with the official results published on Wikipedia.

3.2 Models and Estimation

To assess the game-level effect of goals scored by foreign players, we model the match outcomes using the independent Poisson regression for the number of goals and the ordered logit regression for the win-draw-lose results. In the first model, the contribution of foreign players is examined through the number of goals they score to the total number of goals made by their team. Given the covariates, the (conditional) expected number of goals or the incidence rate (λ_i) for the model is given by

$$\lambda_i = \exp\left(\beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i\right) \quad (1)$$

where i= 1, 2, 3, ..., 612 is the observation index, fgoal is the total number of goals scored by the team's foreign players, lgoal is the total number of goals scored by the team's local players, home is the dummy variable indicating the hosting status where 1 is the home team and 0 otherwise, and yellow and red are the number of the yellow and red cards, respectively, made by the team.

The Poisson regression provides the information on how the goals scored by foreign players have an impact on the team performance in creating goals. The estimated regression coefficients can be interpreted in terms of incidence rate ratios (IRR) that imply the frequency of the incidence. The IRR explains the rate at which the incidence occurs under specific time and conditions. In this case, the IRR is the ratio of the expected number of goals from a unit increase in each covariate to the expected number of goals. We expect the number of goals scored by foreign players and the home advantage have the positive effect on the final score while the yellow and red cards might have the negative effect.

In the second model, we analyze the win-draw-lose results using the ordered logistic (logit) regression where the differences in the ordered results are assumed to be non-linear, i.e., the difference between win and draw is not the same as that of lose and draw. In the analysis, we code the win, draw, and lose results as 2, 1, and 0, respectively. Let y denote the match results and y^* denote the latent or unobserved value. The ordered logit model is then given by

$$y_i^* = \beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i + \beta_6 \text{firstgoal}_i + \epsilon_i \quad (2)$$

where y = 0 if $y^* \leq \mu_1$, y=1, if $\mu_1 \leq y^* \leq \mu_2$, y = 2 if $\mu_2 \leq y^*$, μ_1 and μ_2 are the unobserved levels, firstgoal is the dummy variable taking the value of 1 if the team scores the first goal and 0 otherwise, and ϵ is the error term. The firstgoal variable is included to determine whether scoring the firstgoal has a positive effect on the match results.

The coefficients of the ordered logit regression can be interpreted in terms of changes in log odds of getting a better match outcome. These coefficients might be better interpreted by exponentiating into odds ratios. The odds ratios indicate the times of changes in factors influencing the match results in getting a win compared to draw or lose. Moreover, the specific impact of having goals scored by foreign players on changing outcomes is examined from the marginal probability. The marginal probability indicates the probability of getting certain match results when foreign players score an additional goal for each number of goals scored.

In both models, the estimation is made through the maximization of the corresponding log-likelihood using Stata statistical software package. In the Poisson regression model, the robust estimation is implemented to maintain the assumption that the expectation or the mean and the variance of y are equal. Hence, we obtain the robust standard errors for inference. In addition to the model spefication in equations (1) and (2), we also substitute the number of goals by local players for that of foreign players and put both of them in the models for a fair comparison and to assess whether the impact of goals scored by the foreign players is substitutionary or complementary.

4. Results

4.1 Data Description

Table 1 shows the summary statistics of total goals from all players and the goals from foreign players. In the 2015 season, there were 870 goals scored in total in which 56.67% of them were scored by the foreign players. On average, there were approximately 1.42 goals scored by each team in a particular match. From Figure 1, it can be seen the total goals and the goals by foreign players are not equally dispersed. In addition, the indices of dispersion of total goals from all players and the goals from foreign players in Table 1 are slightly greater than 1. Hence, the corresponding distribution follows a geometric or negative binomial distribution. However, we examined the over-dispersion by testing the likelihood function between the Poisson and the negative binomial regression for each model, as suggested by Cameron & Trivedi (1990) and Hilbe (2007), and found the distribution is not over-dispersed. We can then use Poisson regression for the total goals. The highest recorded number of goals scored by a team in a match was achieved by Muangthong United and Buriram United, in which both of them had the clean-sheet (7-0) win over the opponents. Particularly for Buriram, it is also the match with the highest number of goals scored by the foreign players at 6 goals. See Table 2 for the 2015 season league table.

Table 1 Summary Statistics for Number of Goals

Variable	Sum	Average	S.D.	Min	Max	Index of Dispersion
Total Goals	870	1.4216	1.3042	0	7	1.1959
Goals by Foreign Players	493	0.8056	0.9974	0	6	1.2350

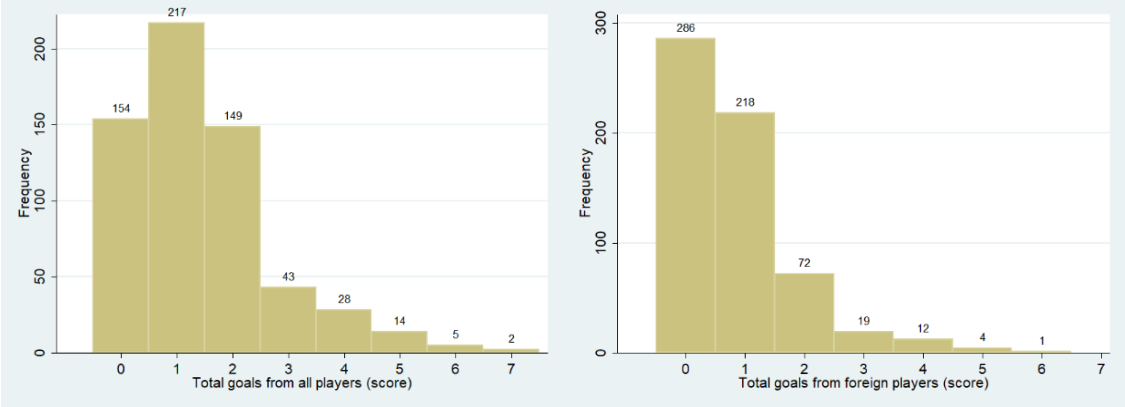


Figure 1 Distribution of Number of Goals

The performance of the foreign players throughout the season seems to be consistent across the game results (Table 3) and the hosting status (Table 4). The remarkable contribution of the goals by foreign players can be seen from the average and the proportion of goals, shown in Table 3, that account for more than half of what their team can achieve in total. The highest proportion of goals came with the draw result. Interestingly, even when the team lost, the foreign players still scored more than their local teammates. Similarly, the hosting status shown in Table 4 didn't deter the foreign players' performance where they scored more in the home games.

In addition to examining the effect of goals scored by foreign players, we also consider the home advantage or the hosting status as a determinant of the match outcome. Table 5 shows the number and proportion of games by match result and hosting status. We found that the home advantage returns the better outcome both in terms of (more) win and (less) lose results. This seems consistent with the higher average number of goals scored by the home team in Table 4.

Table 2 2015 League Table

Position	Team	Played	Won	Drawn	Lost	Goals for	Points
1	Buriram United[a]	34	25	9	0	98	84
2	Muangthong United[b]	34	21	8	5	81	71
3	Suphanburi	34	16	11	7	60	59
4	Chonburi[b]	34	15	12	7	62	57[c]
5	Bangkok United	34	16	9	9	59	57[c]
6	Bangkok Glass	34	15	11	8	47	56
7	Ratchaburi	34	17	4	13	48	55
8	Nakhon Ratchasima	34	13	10	11	37	49
9	Chiangrai United	34	12	8	14	42	44
10	Army United	34	11	8	15	43	41
11	Osotspa Samut Prakan	34	10	9	15	40	39
12	Chainat Hornbill	34	9	10	15	42	37
13	Sisaket	34	9	9	16	30	36
14	Saraburi	34	8	11	15	41	35
15	Navy	34	10	5	19	42	35
16	BEC Tero Sasana	34	7	14	13	42	35
17	Port[d]	34	10	3	21	31	33
18	TOT[d]	34	3	7	24	25	16

Source: Wikipedia
Notes: a = champion and qualified to the AFC champion league group stage
b = qualified to the AFC champion league play-o
c = Chonburi got the better head-to-head results
d = relegated

Table 3 Average and Proportion of Goals by Foreign Players and Match Result

Match Result	Average		Proportion of Goals by Foreign Players
	Total Goals	Goals by Foreign Players	
Win	2.4779	1.4159	57.14%
Draw	1.0253	0.6013	58.64%
Lose	0.6491	0.3421	52.70%
Overall	1.4134	0.8056	56.67%

Table 4 Average and Proportion of Goals by Foreign Players and Hosting Status

Hosting Status	Average		Proportion of Goals by Foreign Players
	Total Goals	Goals by Foreign Players	
Home	1.6340	0.9444	57.80%
Away	1.2092	0.6667	55.14%

Table 5 Number and Percentage of Games by Match Result and Hosting status

Match Result	Hosting Status		Total
	Home	Away	
Win	142 (46.40%)	84 (27.45%)	226 (36.93%)
Draw	79 (25.82%)	79 (25.82%)	158 (25.82%)
Lose	85 (27.78%)	143 (46.73%)	228 (37.25%)
Total	306 (100%)	306 (100%)	612 (100%)

Notes: Number in the parenthesis is the percentage of games.

The other variables taken into consideration are the number of yellow and red cards. From Table 6, it is interesting to see that the teams with 4 or more yellow cards had the higher percentage of winning than losing and drawing. This is probably caused by the attacking strategy of the teams to get the better results. However, the teams with 1-3 yellow cards were likely to lose. This implies that the yellow card is not only a sign of punishment that suppresses the team performance but it can also indicate how the team play, especially, with a more aggressive strategy. Similarly, the teams with red cards were likely to lose. That means the red card is associated with a chance of losing.

Table 6 Number and Percentage of Games by Number of Yellow and Red Cards and Match Result

Match Result	Yellow Card			Red Card		
	0	**1-3**	**≥ 4**	**0**	**1**	**2**
Win	59 (43.38%)	130 (34.03%)	37 (39.36%)	214 (38.56%)	12 (22.22%)	0 (0.00%)
Draw	26 (19.12%)	103 (26.96%)	29 (30.85%)	138 (24.86%)	19 (35.19%)	1 (33.33%)
Lose	51 (37.50%)	149 (39.01%)	28 (29.79%)	203 (36.58%)	23 (42.59%)	2 (66.67%)

Notes: Number in the parenthesis is the percentage of games.

4.2 Poisson Regression on Match Outcome

In estimating the impact of goals scored by foreign players on the total number of goals, we use the Poisson regression that also includes other covariates, i.e., the home advantage and the numbers of yellow and red cards. In addition to assessing the effect on match outcome, these covariates also help mitigate the dependence among the number of goals scored by each team, especially the yellow cards that can indicate the team strategy against the opponent. In the Poisson regression, the model treated the dependent variable as the count variable where only a limited set of countable numbers is specified.

Table 7 shows the results from the estimation of the Poisson regression model in three cases where the number goals by local and foreign players are included and excluded. The regression coefficients represent the expected change in log-count of total goals for a one-unit change in the covariates. The major implications of the regression rely on the sign and the statistical significance of these covariates. Note that the log-count interpretation is generally not useful to some implications, however, its transformation, which is discussed below, is more meaningful. We found that the number of goals scored by either foreign or local players, or both have the strong (in terms of statistical significance) and positive effect on the total number of goals where the coefficient from the goals by foreign players is slightly higher.

The home advantage has also the strong and positive effect on the total number of goals when the numbers of goals by foreign and local players are included separately. However, the home advantage turns to be an insignificant factor when both foreign and local players' goals are jointly included. This is possibly due to the fact that the number of goals by both foreign and local players incorporate and mitigate the effect of home advantage. This can also be confirmed by the information from Tables 4 and 5 where the contributions from the goals by home team and the match results can be obviously observed.

The effect from yellow and red cards are not statistically significant when the number of goals by foreign and local players is not included at the same time. However, the effect from yellow card is positively and strongly significant when the number of goals by foreign and local players is all included. This implies the yellow cards are an effective measure as more than two-third of the games resulting in either winning or drawing status when yellow cards were given (see Table 6). Note that a yellow card might be given following a more defensive strategy from the leading team to pretect their results. The McFadden's adjusted pseudo-R2 of 28.4% indicates the reasonable amount of variation in the dependent variables can be explained by the variation in the covariates when the goals by foreign and local players are both included. This seems to be best-fitted model as its McFadden's adjusted pseudo-R2 is higher than those when the two variables are included separately.

Table 7 Estimated Coefficients of Poisson Regression Model for Number of Goals

Variable	(A)		(B)		(C)	
	Coefficient	**IRR**	**Coefficient**	**IRR**	**Coefficient**	**IRR**
fgoal	0.4458 ***	1.5618 ***	0.4785 ***	1.6136 ***		
	(0.0237)	(0.0371)	(0.0246)	(0.0396)		
lgoal			0.4712 ***	1.6019 ***	0.4255 ***	1.5257 ***
			(0.0292)	(0.0468)	(0.0363)	(0.0554)
home	0.1284 **	1.1370 **	0.0482	1.0494	0.1837 ***	1.2017 ***
	(0.0541)	(0.0615)	(0.0330)	(0.0346)	(0.0608)	(0.0731)
yellow	0.0055	1.0055	0.0293 **	1.0297 **	-0.0159	0.9842
	(0.0173)	(0.0174)	(0.1240)	(0.0128)	(0.0204)	(0.0201)
red	-0.0615	0.9404	0.0552	1.0568	-0.0540	0.9474
	(0.0778)	(0.0732)	(0.0495)	(0.0523)	(0.0899)	(0.0852)
Constant	-0.2132 ***	0.8080 ***	-0.6596 ***	0.5171 ***	-0.0647	0.9373
	(0.0651)	(0.0534)	(0.0538)	(0.0278)	(0.0743)	(0.0697)
McFadden's Adj R2	0.157		0.284		0.117	
Observations	612		612		612	

Notes: (A) $\lambda_i = \exp(\beta_0 + \beta_1 \text{fgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$
(B) $\lambda_i = \exp(\beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$
(C) $\lambda_i = \exp(\beta_0 + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$
*** and ** indicate the statistical significance at 1% and 5% level, respectively.
Number in the parenthesis is the robust standard error.

To analyze further, we consider the transformation of the log-count to the incidence rate ratio (IRR) by exponentiating the estimated equation in (1). The IRR shows the rate at which the incidence occurs and can be interpreted in terms of probability. It is the estimated rate ratio for a one-unit increase in goals scored by a player (foreigner or local) given the other covariates being constant. For example in the model where the goals by local players is excluded, if a foreign player were to score an additional goal for the team, the total number of goals is expected to increase by a factor of 1.5618, while holding other covariates constant. This indicates the value of a unit of goal scored by the foreigner is worth more than 1 goal as this might increase the overall performance and help the team to score more goals. When the goals by foreign and local players are all included, the rates are even higher for both of them with a slightly higher rate by the foreign players. This indicates the complementary effect of the goals by foreign players as the goals by local players can also contribute more. For the statistically significant effect from the yellow cards, the IRR coefficient at 1.0297 implies that on average being given (at least) a yellow card is expected to yield the marginally higher number of goals by 1.0297 times.

The impact of goals scored by foreign and local players can also be illustrated using the predicted number of total goals by the team. According to Table 8, when the foreign player scores 1 goal, on average, the team can score 1.3516 goals in total. Similarly, if the foreign players score 2 goals, on average, the team can score 2.1109 goals in total. This implies the value of goals scored by the foreign players is worth more than a goal itself. That means when a foreign player scores the goals, other teammates are likely to perform better and add more goals on the scoreboards. This is clearly an influential and complementary effect on the other teammates. We also obtain the similar results when the goals by local players are included. Note that the contributions made from the goals by the foreign players are non-linear with the exponential increase as the number of goals increases. That is, when the number of goals by the foreign players reaches 6 goals, the predicted total number of goals that can be made by the team is more than 12. This is also the case with the inclusion of local players' goals.

Table 8 Predicted Total Number of Goals of the Team by the Incremental Change of Foreign Players' Goals

fgoal	(A)			(B)		
	Predicted Total Goal	[95% CI]		Predicted Total Goal	[95% CI]	
0	0.8654	0.7884	0.9424	0.75226	0.7026	0.8026
1	1.3516	1.2720	1.4311	1.2144	1.1710	1.2577
2	2.1109	1.9887	2.2331	1.9595	1.8564	2.0627
3	3.2968	3.0100	3.5836	3.1619	2.8636	3.4603
4	5.1489	4.4952	5.8026	5.1021	4.3866	5.8176

Notes: (A) $\lambda_i = \exp(\beta_0 + \beta_1 \text{fgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$
(B) $\lambda_i = \exp(\beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$
CI = Confident Interval

To make a fair comparison, we also calculate the total predicted goals resulting from the goals by local players (Table 9). We found in both cases that the contribution from the local players' goals is slightly higher than those of foreign players. Note that the total predicted goals from the local players' goals in the case where both variables are included are higher when the local players' goals are four or more. This indicates the complementary effect to both goals by foreign and local players.

Table 9 Predicted Total Number of Goals of the Team by the Incremental Change of Local Players' Goals

lgoal	(B)			(C)		
	Predicted Total Goal	[95% CI]		Predicted Total Goal	[95% CI]	
0	0.8277	0.7768	0.8786	0.9917	0.8993	1.0842
1	1.3260	1.2769	1.3750	1.5131	1.4226	1.6037
2	2.1240	1.9665	2.2815	2.3087	2.0944	2.5229
3	3.4025	2.9697	3.8352	3.5224	2.9793	4.0656
4	5.4504	4.4543	6.4464	5.3743	4.1837	6.5650

Notes: (B) $\lambda_i = \exp(\beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$
(C) $\lambda_i = \exp(\beta_0 + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$
CI = Confident Interval

Note that we want to consider the relationships among the total number of goals, the number of goals by foreign players, and the number of goals by local players through correlation coefficients. We found the higher correlation of 0.75 between the total number of goals and the goals by foreign players whereas the correlation between the total number of goals and the goals by local players is slightly lower but still high at 0.65 while there is almost no correlation between the goals by foreign and local players at -0.028. This is merely a rough estimate of the relationships among variables as we understand the linear correlation we used is not an appropriate measure of the count data. Note, however, that based on the calculated correlations we can deduce the goals by foreign players are rather complementary than substitutionary.

We also assess the impact of the number of goals scored by both foreign and local players and match results where we separate the results into win, draw, and lose, and estimate the Poisson regression for each of them. From Table 10, we found that goals by foreign players still contribute significantly in all results with the increasing scale from win to lose. This reflects the remarkable contributions from the goals by foreign players, especially when the team faces difficult situations. Indicated by the negative coefficient, the yellow cards may lead to lower chances of getting more goals from the limited players' abilities and result in the loss status. On the other hand, the red cards seem to have a statistically and positively significant effect on the total goals when the team wants to equalize the match outcome. Including the yellow and red cards as the covariates in the model clearly justify the use of independent Poisson as the cards reflect the defending strategy.

The similar results were also obtained in the case where the only goals by local players are included. However, the adverse effect from yellow cards is obvious when the team won. This can be caused by the cautious strategy that the team got the yellow cards to protect their results where the local players scored more goals. Note that when both the number of goals by foreign and local players are included, the effects from the goals by foreign players are slightly higher than those of local players in all three match results. Being the host now turns to be disadvantageous as the coefficient in the case of win is negative and statistically significant. This is possibly due to the fact that the home team scored less. Though not statistically significant, this effect can also be observed when the goals by foreign and local players are not jointly analyzed. The positive coefficients from the hosting status when the team lost also confirmed this effect. High pressure from the fans might be a contributing factor to the limited abilities to score for the home team.

The contributions from the goals made by the foreign and local players towards the team's total goals are also evident from the predicted total number of goals in the win-draw-lose results, see Tables 11 and 12, respectively. When the foreign players score no goals, the other teammates contribute differently by the order of the results. It is higher when the team wins and lower when the team, respectively, draws and loses. With the goals by foreign players, it is not only the incremental increase in the predicted total goals but also the contributions that make a substantial difference, especially when the team faces more difficult situations. For example, in the case of win in Model (A) from Table 11, when foreign players score one additional goal from 0 to 1, the predicted total number of goals increases from 1.6772 to 2.1351 indicating the approximate contribution of additional 0.5 goals. These contributions slightly increase when they score more goals. On the other hand, when the team draws or loses, the contributions from foreign players are even higher, e.g., in the case of loss, one additional goal from 2 to 3 by foreign players leads to the additional contribution of 5.1 goals. That means, even when the team loses, the contributions from foreign players are substantial but that is not sufficient to produce the preferred results. However, either in the case of the goals by foreign or local players, the complementary effect from both of them is obvious when both of them are jointly included in the model. Note that the purpose of presenting the predicted total goals by match result is only for comparison and assessing the incremental change. The resulting numbers of predicted goals may seem unrealistic but useful to analyze the players' contributions from their goals.

Table 10 Estimated Coefficients of Poisson Regression Model for Number of Goals by Match Result

Variable	(A)			(B)			(C)		
	Win	**Draw**	**Lose**	**Win**	**Draw**	**Lose**	**Win**	**Draw**	**Lose**
fgoal	0.2414 ***	0.6439 ***	1.0248 ***	0.3331 ***	0.9600 ***	1.1917 ***			
	(0.0179)	(0.0620)	(0.0831)	(0.0129)	(0.0705)	(0.0690)			
lgoal				0.3296 ***	1.0386 ***	1.0868 ***	0.2171 ***	0.5710 ***	0.9832 ***
				(0.0146)	(0.0650)	(0.0808)	(0.0189)	(0.0790)	(0.0775)
home	-0.0560	0.0790	0.1790 **	-0.0334 *	0.0073	0.0445	-0.0165	-0.0063	0.2165 *
	(0.00553)	(0.0902)	(0.1224)	(0.0184)	(0.0545)	(0.0936)	(0.0626)	(0.1017)	(0.1166)
yellow	-0.01384	0.0003	-0.0664 *	0.0070	0.0092	0.0554 *	-0.0462 **	0.0457	-0.0014
	(0.0194)	(0.0240)	(0.03925)	(0.0067)	(0.0129)	(0.0329)	(0.0195)	(0.0306)	(0.0450)
red	-0.0910	0.2359 ***	-0.2445	-0.0069	0.0457	0.0897	-0.1001	0.1607 *	0.2127
	(0.1121)	(0.0884)	(0.1553)	(0.0380)	(0.04466)	(0.0976)	(0.1334)	(0.0965)	(0.1858)
Constant	0.5821 ***	-0.5437 ***	-0.9175 ***	-0.0189	-1.2960 ***	-1.7640 ***	0.7408 ***	-0.3919 ***	-1.0384 ***
	(0.0743)	(0.1246)	(0.1380)	(0.0386)	(0.1025)	(0.1247)	(0.0808)	(0.1266)	(0.1516)
McFadden's Adj R2	0.066	0.077	0.142	0.172	0.195	0.307	0.045	0.028	0.132
Observations	226	158	228	226	158	228	226	158	228

Notes: (A) $\lambda_i = \exp(\beta_0 + \beta_1 fgoal_i + \beta_3 home_i + \beta_4 yellow_i + \beta_5 red_i)$
(B) $\lambda_i = \exp(\beta_0 + \beta_1 fgoal_i + \beta_2 lgoal_i + \beta_3 home_i + \beta_4 yellow_i + \beta_5 red_i)$
(C) $\lambda_i = \exp(\beta_0 + \beta_2 lgoal_i + \beta_3 home_i + \beta_4 yellow_i + \beta_5 red_i)$
***, **and * indicate the statistical significance at 1%, 5% and 10% level, respectively.
Number in the parenthesis is the robust standard error.

Table 11 Predicted Total Number of Goals of the Team by the Incremental Change of Foreign Players' Goals and Match Result

fgoal	(A)			(B)		
	Predicted Total Goal			Predicted Total Goal		
	Win	Draw	Lose	Win	Draw	Lose
0	1.6772	0.6236	0.3674	1.3804	0.4376	0.2720
1	2.1351	1.1873	1.0238	1.9260	1.1429	0.8657
2	2.7180	2.2604	2.8527	2.6872	2.9848	2.9491
3	3.4601	4.3034	7.9492	3.7493	7.7952	9.7102
4	4.4048	8.1930	22.1505	5.2311	20.3582	31.9714

Notes: (A) $\lambda_i = \exp(\beta_0 + \beta_1 \text{fgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$
(B) $\lambda_i = \exp(\beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$

Table 12 Predicted Total Number of Goals of the Team by the Incremental Change of Local Players' Goals and Match Result

lgoal	(B)			(C)		
	Predicted Total Goal			Predicted Total Goal		
	Win	Draw	Lose	Win	Draw	Lose
0	1.9000	0.7580	0.3925	1.5588	0.5018	0.2929
1	2.3606	1.3417	1.0492	2.1675	1.4175	0.8685
2	2.9329	2.3749	2.8045	3.0138	4.0048	2.5750
3	3.6440	4.2035	7.4961	4.1905	11.3145	7.6344
4	4.5276	7.4403	20.0367	5.8266	31.9655	22.6350

Notes: (B) $\lambda_i = \exp(\beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$
(C) $\lambda_i = \exp(\beta_0 + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i)$

4.3 Ordered Logistic Regression on Match Results

In the previous section, we consider the contributions from the number of goals made by foreign players with the comparison to the number of goals by the local players. The contributions were still obvious even we separated the match results into win, draw, and lose. However, scoring goals do not necessarily lead to preferred results. In this section, we pay more attention to the match outcome through the win-draw-lose results using the ordered regression. The key question is whether having foreign players scoring goals raises the probability for the team to win or get a better result and how large the impact of the contribution is. Using the regression coefficients and their odd-ratios (the relative probability of the team getting a win compared to the probability of getting draw and lose) can help answer the question.

The results from Table 13 show from the coefficients that the impact of goals scored by the foreign players is strongly positive and statistically significant. This implies that having foreign players scoring goals induces a higher probability of getting a win or a better result. Having an additional goal scored by the foreign players, there is a 0.8026 increase in the log-odds of getting a better match outcome. However, it is more meaningful to interpret the coefficients in terms of the odd-ratios. That is, the odd ratios of the number of goals scored by foreign players imply that an additional goal by foreign players leads to the odds of getting a win 2.2314 times higher than that of getting a loss or draw, holding other variables constant. This indicates a significant effect of goals scored by foreign players that can help the team to get a win.

Table 13 Estimated Coefficients and Odds Ratio for Ordered Logistic Regression Model on Match Results

Variable	(A)		(B)		(C)	
	Ordered Logit	**Odd Ratios**	**Ordered Logit**	**Odd Ratios**	**Ordered Logit**	**Odd Ratios**
fgoal	0.8026 ***	2.2314 ***	1.0761 ***	2.9334 ***		
	(0.1184)	(0.2643)	(0.1312)	(0.3849)		
lgoal			1.1136 ***	3.0452 ***	0.8263 ***	2.2849 ***
			(0.1354)	(0.4123)	(0.1219)	(0.2784)
home	0.4559 **	1.5776 **	0.4234 **	1.5272 **	0.4671 ***	1.5953 ***
	(0.1767)	(0.2788)	(0.1862)	(0.2844)	(0.1767)	(0.2820)
yellow	0.0274	1.0278	0.0486	1.0498	0.0413	1.0422
	(0.0554)	(0.0570)	(0.0569)	(0.0597)	(0.0548)	(0.0571)
red	-0.6502 **	0.5219 **	-0.5950 **	0.5516 **	-0.5909 **	0.5539 **
	(0.3008)	(0.1570)	(0.2836)	(0.1564)	(0.2638)	(0.1461)
firstgoal	2.5208 ***	12.4390 ***	2.1560 ***	8.6363 ***	2.7693 ***	15.9479 ***
	(0.2087)	(2.5959)	(0.2141)	(1.8492)	(0.2012)	(3.2082)
Constant (cut 1)	1.0025	1.0025	1.5948	1.5948	1.0318	1.0318
	(0.1890)	(0.1890)	(0.2168)	(0.2168)	(0.1935)	(0.1935)
Constant (cut 1)	2.8315	2.8315	3.6295	3.6295	2.8631	2.8631
	(0.2187)	(0.2187)	(0.2429)	(0.2429)	(0.2227)	(0.2227)
McFadden's Adj R2	0.263		0.319		0.261	
Observations	612		612		612	

Notes: (A) $y_i^* = \beta_0 + \beta_1 \text{fgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i + \beta_6 \text{firstgoal}_i + \epsilon_i$
(B) $y_i^* = \beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i + \beta_6 \text{firstgoal}_i + \epsilon_i$
(C) $y_i^* = \beta_0 + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i + \beta_6 \text{firstgoal}_i + \epsilon_i$
*** and ** indicate the statistical significance at 1% and 5% level, respectively.
Number in the parenthesis is the robust standard error.

From both the regression coefficients and odd-ratios, the effect from the dummy variable indicating whether the team scored the first goal can lead to a better match result is positive and statistically significant. That means the team scoring the first goal of the match has a higher probability to end up with a better match outcome. In other words, the advantage of the first goal significantly influences the match outcome. The odd-ratios indicate the influential impact of the team that can score the opening goal with the odds ratios of getting a win being 12.44 times higher than the other results. Home advantage is another dummy variable that indicates the positive and statistically significant effect on the match result. That means the home team is likely to get the better results, regardless of number of goals they score. As expected, the number of red cards leads to the worse result while the effect of yellow cards is not statistically significant.

We also obtain the similar results when the number of goals by local players is solely or jointly included. Note that the coefficients from the number of goals by local players are slightly higher than those of the number of goals by foreign players. However, when they were jointly analyzed, the coefficients are obviously higher than the cases where one of them is excluded. This can confirm the complementary effect of the goals by both foreign and local players.

Considering the marginal probability from an additional goal by foreign players reveals some interesting information. Table 14 shows the marginal probability by match result where the number of goals by foreign players is solely included and jointly analyzed with the number of goals by local players. In the case with only foreign players' goals, when foreign players score no goals, that creates a chance of 19.20% for the team to win, and 40.47% to draw, and 40.33% to lose, respectively. As the number of goals scored by foreign players increases, the probability of winning rises exponentially while the probabilities of getting draw and lose decline dramatically. Among the three match results, the probability of win is higher when foreign players score two or more goals. The results from marginal probabilities confirm the significant contribution to the match outcome from the goals scored by foreign players.

Table 14 Estimated Probability of Match Result by Number of Goals from Foreign Players

fgoal	(A)			(B)		
	Win	**Draw**	**Lose**	**Win**	**Draw**	**Lose**
0	0.1920	0.4047	0.4033	0.1557	0.4295	0.4148
1	0.3464	0.4211	0.2325	0.3511	0.4543	0.1946
2	0.5419	0.3386	0.1195	0.6134	0.3105	0.0761
3	0.7252	0.2174	0.0573	0.8232	0.1495	0.0273
4	0.8549	0.1186	0.0265	0.9318	0.0588	0.0095

Notes: (A) $y_i^* = \beta_0 + \beta_1 \text{fgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i + \beta_6 \text{firstgoal}_i + \epsilon_i$
(B) $y_i^* = \beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i + \beta_6 \text{firstgoal}_i + \epsilon_i$
Number in the table is the marginal probability.

After the number of goals by local players is included, the marginal probability of win is slightly lower when foreign players score no goals. However, the marginal probabilities of win increase significantly if foreign players start scoring the goals. This again confirms the contribution from the goals by foreign players and their complementary effect to the desired winning result. We also perform the similar analysis for the contribution of goals by local players (Table 15) and find the probabilities of win are obviously lower when the only goals by local players is included, especially when the local players score one or no goals where the probabilities of draw are higher.

Table 15 Estimated Probability of Match Result by Number of Goals from Local Players

fgoal	(A)			(B)		
	Win	**Draw**	**Lose**	**Win**	**Draw**	**Lose**
0	0.2116	0.4146	0.3737	0.1810	0.4473	0.3716
1	0.3802	0.4127	0.2071	0.4023	0.4351	0.1626
2	0.5836	0.3138	0.1026	0.6721	0.2680	0.0600
3	0.7621	0.1903	0.0476	0.8619	0.1176	0.0205
4	0.8798	0.0988	0.0214	0.9500	0.0432	0.0068

Notes: (A) $y_i^* = \beta_0 + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i + \beta_6 \text{firstgoal}_i + \epsilon_i$
(B) $y_i^* = \beta_0 + \beta_1 \text{fgoal}_i + \beta_2 \text{lgoal}_i + \beta_3 \text{home}_i + \beta_4 \text{yellow}_i + \beta_5 \text{red}_i + \beta_6 \text{firstgoal}_i + \epsilon_i$
Number in the table is the marginal probability.

In summary, the findings from this study indicate foreign players' goals can make a substantial contribution to the match results in TPL in 2015. However, we only focus on the effect of goals on match results, which is only one aspect of many from the foreign players and TPL. There are still other interesting topics for TPL such as financing, coaching, and influences for national performance but we leave these for future research.

5. Conclusion

Foreign players have been considered as a key factor to many football clubs' success in Thailand. As a result, its football league is flourishing and now ranked a top eight in Asia. This study examines the contribution from the number of goals scored by the foreign players towards the match outcome, both in terms of the total number of goals and the win-draw-lose results. The estimation results from the Poisson regression model on the total number of goals reveals that each goal is scored by foreign players is worth more than a goal in total made by their team. Precisely, an additional goal by the foreign players has an influence on the overall team's performance, which creates an exponential increase in the team's total number of goals. Interestingly, this contribution is highest when their team does not perform well, e.g., when it loses. The similar analysis from the number of goals by local players also reveals the complementary effect from the foreign players' goals

Furthermore, the evidence from the ordered logistic regression model on the match results shows that the goals scored by the foreign players yield a higher probability of getting the better match results. When the number of goals by foreign players increases, the probability of getting a win rises exponentially while the probabilities to lose or draw drop significantly. These probabilities of win are even higher when both of the numbers of goals by foreign and local players are included.

In addition to the contribution by the foreign players, this study also considers the effect from other factors including the red and yellow cards, the home advantage, and the first goal (in the case of match results), on the match outcome. We found that the home advantage has a higher probability to win. While the teams that score the first goal have a higher chance to win, the red cards create the opposite effect. Based on the results from this study, we can conclude that the number of goals by foreign players contribute significantly and complementarily that leads to success in both the club and league levels in Thailand. There are also other interesting topics for TPL such as financing, coaching, and influences for national performance but we leave these for future research.

6. Acknowledgements

We would like to thank the Editor and three anonymous referees for their useful comments and feedback that help improve our manuscript.

7. References

Baur, D. G., & Lehmann, S. (2007). Does the mobility of football players influence the success of the national team?. **IIIS Discussion Paper,** No. 217, Institute for International Integration Studies, Dublin.

Berlinschi, R., Schokkaert, J., & Swinnen, J. (2013). When drains and gains coincide: Migration and international football performance. **Labour Economics,** 21, 1-14.

Bernard, A. B., & Busse, M. R. (2004). Who wins the Olympic Games: Economic resources and medal total. **The Review of Economic and Statistics,** 86, 413-417.

Binder, J. J., & Findlay, M. (2012). The effects of the Bosman ruling on national and club teams in Europe. **Journal of Sports Economics,** 13(2), 107-129.

Brechot, M., & Flepp, R. (2020). Dealing with randomness in match outcomes: How to rethink performance evaluation in European club football using expected goals. **Journal of Sports Economics,** 21(4), 335-362.

Cameron A.C. &Trivedi P.K. (1990). **Regression Analysis of Count Data.** Cambridge University Press. New York.

Dixon, M. J., & Coles, S. C. (1997). Modelling association football scores and inefficiencies in the football betting market. **Applied Statistics,** 46, 265-280.

Dixon, M. J., & Robinson, M. E. (1998). A birth process model for association football matches. **Statistician,** 47, 523-538.

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. **International Journal of Forecasting,** 21, 551-564.

Forrest D. K., & Simmons R. (2000). Forecasting sport: The behaviour and performance of football tipsters. **International Journal of Forecasting,** 16, 317-331.

Gásquez, R., & Royuela, V. (2016). The determinants of international football success: A panel data analysis of the elo rating. **Social Science Quarterly,** 97, 125-141.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. **International Journal of Forecasting,** 21, 331-340.

Goddard, J., & Asimakopoulos, I. (2004). Forecasting football match results and the efficiency of fixed-odds betting. **Journal of Forecasting,** 23, 51-66.

Groll, A., Schauberger, G., & Tutz, G. (2015). Prediction of major international soccer tournaments based on team-specific regularized Poisson regression: An application to the FIFA World Cup 2014. **Journal of Quantitative Analysis in Sports,** 11(2), 97-115.

Hilbe, J.M. (2007). **Negative Binomial Regression.** Cambridge University Press. New York.

Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. **Statistician,** 52, 381-393.

Maher, M. J. (1982). Modelling association football scores. **Statistica Neerlandica,** 36, 109-118.

McHale, I., & Baker, R., (2014). Econometric modelling of match results and scores. In J. Goddard, & P. Sloane (Eds.), **Handbook on the Economics of Professional Football** (130-140). London: Edward Elgar.

Pearson, M., Livingston, G., & King, R. (2020). An exploration of predictive football modelling. **Journal of Quantitative Analysis in Sports,** 16(1): 27-39.

Royuela, V., & Gásquez, R. (2019). On the influence of foreign players on the success of Football Clubs. **Journal of Sports Economics,** 20(5), 718-741.

การศึกษาความสามารถในการทำกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัย

อดิภา เดชคง*[1] และ วรรณฤดี สกุลภักดี[2]

Received: 30 August 2021; Revised: 13 May 2022; Accepted: 6 June 2022

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อวิเคราะห์ความสามารถในการทำกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัย ผู้วิจัยได้รวบรวมข้อมูลทุติยภูมิ เบี้ยประกันที่ถือเป็นรายได้ ค่าสินไหมทดแทนสุทธิ จากเว็บไซต์ของแต่ละบริษัทประกันวินาศภัยที่มีการดำเนินธุรกิจในปัจจุบันเพื่อทำการวิเคราะห์ ในขั้นแรกผู้วิจัยแบ่งกลุ่มบริษัทประกันวินาศภัยออกเป็น 3 กลุ่มจากอัตราส่วนค่าใช้จ่ายอื่นๆ และคำนวณค่าเฉลี่ยของอัตราส่วนค่าใช้จ่ายใน 3 กลุ่มบริษัท ได้แก่ บริษัทที่มีอัตราส่วนค่าใช้จ่ายสูง ปานกลาง และ ต่ำ จากนั้นจึงวิเคราะห์หาฟังก์ชันการแจกแจงความน่าจะเป็นที่เหมาะสมสำหรับอัตราส่วนค่าสินไหมทดแทนในแต่ละกลุ่มพบว่าฟังก์ชันการแจกแจงความน่าจะเป็นที่มีความเหมาะสมกับอัตราส่วนค่าสินไหมทดแทนของกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายปานกลางและต่ำคือฟังก์ชันการแจกแจงปกติแบบผสม และการแจกแจงความน่าจะเป็นที่มีความเหมาะสมกับอัตราส่วนค่าสินไหมทดแทนของกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายสูงคือการแจกแจงปกติ โดยผู้วิจัยใช้การทดสอบภาวะสารูปดี แอนเดอร์สัน – ดาร์ลิง เพื่อทดสอบความเหมาะสมของฟังก์ชันการแจกแจงที่วิเคราะห์มาได้จากขั้นตอนก่อนหน้า จากนั้นจึงนำมาวิเคราะห์หาโอกาสในการทำกำไรจากการรับประกันภัย โดยมีผลการวิจัยดังนี้ บริษัทประกันวินาศภัยที่มีอัตราส่วนค่าใช้จ่ายอยู่ในระดับสูง ปานกลาง และ ต่ำ จะมีความน่าจะเป็นที่จะมีกำไรจากการรับประกันภัยเท่ากับ 0.1211 0.1080 และ 0.8920 ตามลำดับ

คำสำคัญ: อัตราส่วนค่าสินไหมทดแทน อัตราส่วนค่าใช้จ่าย กำไรจากการรับประกันภัย

[1,2] *คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์*

* *E-mail: adipadechkhong@gmail.com*

The study of underwriting profitability for non-life insurance companies

Adipa Dechkhong[*1] and Wanrudee Skulpakdee[2]

Received: 30 August 2021; Revised: 13 May 2022; Accepted: 6 June 2022

Abstract

The objective of this study is to analyze the underwriting profitability of non-life insurance companies. The secondary data of income, insurance premium, and net compensation were collected from the website of each company. Then the expense ratios of the non-life insurance companies were classified into three groups, i.e., the company with high, medium, and low expense ratios. Then, in the next part, the probability distribution function of the loss ratio for each group is analyzed. It was found that the appropriate probability distribution function of loss ratio for the medium and low expense ratio company groups is the normal mixture distribution. The appropriate probability distribution function of the loss ratio for the high expense ratio company group is the normal distribution. In this study, the researcher uses the Anderson-Darling test to fit a sample of data to a specific distribution. Finally, the probabilities of underwriting profit were analyzed and the result is that the probabilities of underwriting profit for non-life insurance companies with high, medium, and low expense ratios are 0.1211, 0.1080, and 0.8920, respectively.

Keywords: Loss Ratio, Expense Ratio, Underwriting Profit

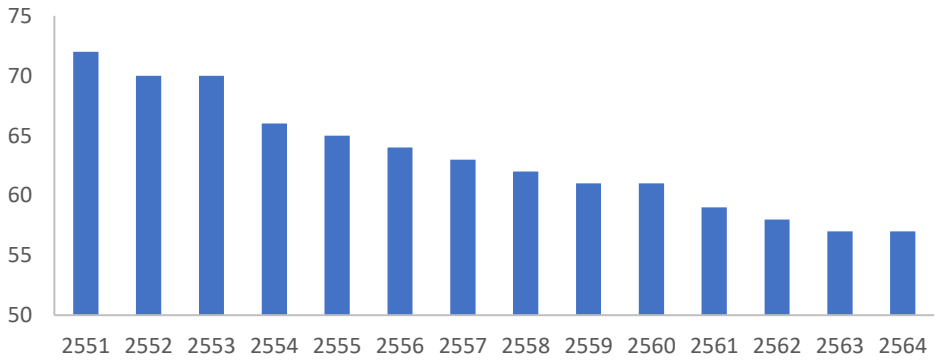
[1,2] *NIDA Applied Statistics School, National Institute of Development Administration*

[*] *E-mail: adipadechkhong@gmail.com*

1. ความเป็นมาและความสำคัญของปัญหา

การประกันภัยเป็นวิธีบริหารจัดการความเสี่ยงที่มีความสำคัญเนื่องจากในปัจจุบันจะเห็นได้ว่าความเสี่ยงภัย ที่เกิดขึ้นมีความหลากหลาย เช่น ความเสี่ยงภัยทั้งต่อร่างกายและความเสี่ยงภัยต่อทรัพย์สิน โดยเมื่อมีความเสียหายเกิดขึ้นอาจมีมูลค่าสูง ดังนั้นการทำประกันภัยสามารถจัดการกับความเสี่ยงและความสูญเสียเหล่านี้ได้ การทำประกันภัยเป็นการถ่ายโอนความเสี่ยงภัยของผู้เอาประกันภัยไปยังบริษัทรับประกันภัย ผู้เอาประกันภัยจะต้องจ่ายเบี้ยประกันภัยให้บริษัทรับประกันภัยตามที่บริษัทกำหนด หากมีความเสียหายเกิดขึ้นบริษัทรับประกันภัยจะเป็นผู้ชดใช้ค่าสินไหมให้นั่นเอง

การประกันภัยแบ่งออกเป็นสองประเภทหลักคือ 1. ประกันชีวิตจะให้ความคุ้มครองชีวิตของผู้เอาประกันภัย และ 2. ประกันวินาศภัยจะให้ความคุ้มครองทรัพย์สินหรือความรับผิดของผู้เอาประกันภัย ปัจจุบันนี้ประเทศไทยมีบริษัทประกันชีวิตและบริษัทประกันวินาศภัยอยู่เป็นจำนวนมากเพื่อเป็นทางเลือกให้กับผู้เอาประกันภัย แต่เมื่อดูจากสถิติจำนวนบริษัทประกันวินาศภัยในประเทศไทยย้อนหลังตั้งแต่ปีพุทธศักราช 2551 ถึงปีพุทธศักราช 2564 จะพบว่าจำนวนบริษัทรับประกันวินาศภัยในประเทศไทยมีจำนวนที่ลดลงอย่างต่อเนื่อง อันเนื่องมาจากทางสำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย (คปภ.) ได้มีคำสั่งให้บริษัทที่ไม่สามารถดำรงเงินกองทุนได้ครบถ้วนตามที่กฎหมายกำหนดจะต้องถูกดำเนินการยกเลิกกิจการ ดังแสดงในภาพที่ 1



ภาพที่ 1 แผนภูมิแท่งแสดงจำนวนบริษัทประกันวินาศภัยในประเทศไทย ปีพุทธศักราช 2551 – 2564
(สำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย, 2564)

หลายครั้งพบว่าการที่บริษัทไม่สามารถดำรงเงินกองทุนได้นั้นเกิดจากหลากหลายสาเหตุด้วยกันเช่น การบริหารงานภายในของบริษัทที่ไม่สามารถแก้ไขฐานะทางการเงินที่ไม่มั่นคง และไม่สามารถหาผู้ร่วมทุนรายใหม่เพื่อนำเงินมาชำระหนี้สินที่มีต่อผู้เอาประกันภัยและเจ้าหนี้อื่นๆของบริษัท รวมถึงการที่บริษัทไม่สามารถบริหารจัดการเบี้ยประกันภัยที่ถือเป็นรายได้หลักของบริษัทให้เกิดกำไรจากการรับประกันภัยได้อีกด้วย

จากที่กล่าวมาข้างต้นผู้ศึกษาจึงมีความสนใจในการศึกษาและวิเคราะห์ความสามารถในการทำกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัยในประเทศไทย เพื่อเป็นแนวทางให้บริษัทประกันวินาศภัยสามารถนำไปปรับใช้ในการดำเนินการจัดการค่าใช้จ่ายที่เกิดขึ้นจากการดำเนินธุรกิจประกันวินาศภัย โดยในการวิจัยครั้งนี้จะทำการศึกษาข้อมูลจากงบการเงินของบริษัทประกันวินาศภัยที่มีการดำเนินธุรกิจในปัจจุบัน ตั้งแต่ปีพุทธศักราช 2559 ถึง ปีพุทธศักราช 2562

การวิจัยในครั้งนี้มีจุดประสงค์เพื่อศึกษาความสามารถในการทำกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัยดังนั้นผู้วิจัยจึงได้ทบทวนวรรณกรรมที่เกี่ยวข้องดังต่อไปนี้

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การวิเคราะห์ประสิทธิภาพในการดำเนินธุรกิจของบริษัทประกันภัย

การวิเคราะห์ประสิทธิภาพในการดำเนินธุรกิจประกันวินาศภัยเกี่ยวข้องกับการสร้างรายได้และการควบคุมค่าใช้จ่ายของบริษัทโดยมีอัตราส่วนที่เกี่ยวข้อง (สุชาดา สถาวรวงศ์ และอธิวัฒน์ ศุภสวัสดิ์วัชร, 2564) ดังนี้

1. อัตราส่วนค่าสินไหมทดแทน (Loss Ratio) เป็นการเปรียบเทียบค่าสินไหมทดแทนที่เกิดขึ้นระหว่างปีกับเบี้ยประกันภัยที่ถือเป็นรายได้ โดยที่อัตราส่วนมาตรฐานคือน้อยกว่าหรือเท่ากับ 60%

$$\text{อัตราส่วนค่าสินไหมทดแทน} = \frac{\text{ค่าสินไหมทดแทนที่เกิดขึ้นสุทธิ}}{\text{เบี้ยประกันภัยที่ถือเป็นรายได้สุทธิ}}$$

2. อัตราส่วนค่าใช้จ่าย (Expense Ratio) เป็นการวัดถึงความสามารถในการควบคุมค่าใช้จ่ายเพื่อสร้างรายได้ โดยที่อัตราส่วนมาตรฐานคือน้อยกว่าหรือเท่ากับ 40%

$$\text{อัตราส่วนค่าใช้จ่าย} = \frac{\begin{array}{c}\text{ค่าใช้จ่ายในการรับประกันภัย} + \text{ค่าใช้จ่ายในการดำเนินงาน} + \\ \text{ค่าจ้างและค่าบำเหน็จสุทธิ}\end{array}}{\text{เบี้ยประกันภัยรับสุทธิ}}$$

3. อัตราส่วนรวมค่าสินไหมทดแทนและค่าใช้จ่ายดำเนินงาน (Combined Ratio) เป็นอัตราส่วนที่ใช้ในการวัดประสิทธิภาพของการรับประกันภัย โดยที่อัตราส่วนมาตรฐานคือน้อยกว่าหรือเท่ากับ 100%

$$\text{อัตราส่วนรวมค่าสินไหมทดแทนและค่าใช้จ่ายดำเนินงาน} = \text{อัตราส่วนค่าสินไหมทดแทน} + \text{อัตราส่วนค่าใช้จ่าย}$$

4. อัตราส่วนกำไรจากการรับประกันภัย (Underwriting Profit Margin) เป็นอัตราส่วนที่ใช้ในการวัดกำไรจากการรับประกันภัยของบริษัทประกันภัย

$$\text{อัตราส่วนกำไรจากการรับประกันภัย} = 1 - \text{อัตราส่วนรวมค่าสินไหมทดแทนและค่าใช้จ่ายดำเนินงาน}$$

2.2 การแจกแจงปกติ (Normal Distribution)

สำหรับการแจกแจงปกติเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีค่าแบบต่อเนื่อง โดยที่ค่าของตัวแปรสุ่มมีแนวโน้มที่จะมีค่าอยู่ใกล้ ๆ กับค่าเฉลี่ย กราฟแสดงค่าฟังก์ชันความหนาแน่นจะเป็นรูปคล้ายระฆังคว่ำ โดยค่าฟังก์ชันความหนาแน่นของการแจกแจงปกติเป็นดังนี้

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0$$

โดยที่ μ คือ ค่าเฉลี่ย และ σ^2 คือ ความแปรปรวน

2.3 การแจกแจงผสม (Mixture Distribution)

กำหนดให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบผสม ซึ่งประกอบด้วยตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นใดๆ จำนวน k องค์ประกอบ กำหนดให้ฟังก์ชันการแจกแจงความน่าจะเป็นของแต่ละองค์ประกอบคือ $f_i(x, \theta_i)$ และสัดส่วนของแต่ละองค์ประกอบเท่ากับ p_i โดยที่ $p_1 + \cdots + p_k = 1$ สำหรับ $i = 1, \ldots, k$ จะได้ฟังก์ชันการแจกแจงความน่าจะเป็นของการแจกแจงแบบผสม (Wilfried Seidel, 2010) ดังนี้

$$f(x, P) = p_1 f_1(x, \theta_1) + \cdots + p_k f_k(x, \theta_k)$$

ซึ่งมีพารามิเตอร์ $P = (p_1, \ldots, p_k, \theta_1, \ldots, \theta_k)$ โดยที่ p_i คือค่าถ่วงน้ำหนัก และ θ_i คือค่าพารามิเตอร์ของแต่ละองค์ประกอบ

2.4 การทดสอบภาวะสารูปดี (Goodness of fit test)

ผู้วิจัยได้ทำการศึกษาวิธีการทดสอบทางสถิติที่สามารถนำมาใช้เทียบการแจกแจงของข้อมูล โดยมีรายละเอียดดังนี้

การทดสอบภาวะสารูปดี (Goodness of fit test) เป็นวิธีทดสอบทางสถิติที่จะปฏิเสธหรือไม่ปฏิเสธการแจกแจงที่เลือกมาว่าเป็นการแจกแจงของตัวแปรสุ่มหรือไม่ สมมติฐานของการทดสอบภาวะสารูปดีคือ $H_0: F(x) = H(x)$ และ $H_1: F(x) \neq H(x)$ ซึ่ง $F(x)$ แทนฟังก์ชันการแจกแจงที่ต้องการทดสอบ และ $H(x)$ แทนฟังก์ชันการแจกแจงของตัวแปรสุ่ม x การทดสอบภาวะสารูปดีมีหลากหลายการทดสอบ เช่น การทดสอบไคสแควร์ แต่ในงานวิจัยนี้เราจะใช้ตัวทดสอบที่ถูกพัฒนาคือการทดสอบแอนเดอร์สัน – ดาร์ลิง ซึ่งมีตัวสถิติทดสอบ (T. W. Anderson; D. A. Darling, 1954) ดังนี้

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[logF(x_i) + \log(1 - F(x_{n-i+1}))]$$

โดยที่ F แทนฟังก์ชันการแจกแจงความน่าจะเป็นแบบสะสมของตัวอย่าง และ n แทนจำนวนตัวอย่าง ค่าวิกฤตของสถิติทดสอบ AD ดูได้จากตาราง Anderson Darling จะปฏิเสธสมมติฐานการแจกแจงความน่าจะเป็นดังกล่าว เมื่อค่าที่คำนวนได้มีค่ามากกว่าค่าวิกฤต

2.5 งานวิจัยที่เกี่ยวข้อง

(มธุรส อรุณรุ่งแสง, 2552) ได้วิเคราะห์ฐานะทางการเงินและสัดส่วนการลงทุนของบริษัทประกันวินาศภัยในประเทศไทย โดยการวิเคราะห์ฐานะทางการเงินและสัดส่วนการลงทุนของบริษัทประกันวินาศภัยไทย ด้วยอัตราส่วนทางการเงินประเภทต่างๆ ผลการวิเคราะห์สามารถสรุปได้ว่า กลุ่มบริษัทประกันวินาศภัยที่เน้นการประกันภัยรถยนต์ มีฐานะทางการเงินในระดับดีเนื่องจาก มีการขยายงานภายใต้ขีดความสามารถทางการเงินของตน และมีแนวโน้มการขยายงานที่สม่ำเสมอ แต่มีสภาพคล่อง และความสามารถในการทำกำไรน้อยกว่ากลุ่มบริษัทประกันวินาศภัยที่ไม่เน้นการประกันภัยรถยนต์ทั้งนี้กลุ่มบริษัทประกันวินาศภัยที่ไม่เน้นการประกันภัยรถยนต์มีความมั่นคงทางการเงินเนื่องจากมีผลกำไรจากการดำเนินงานจากการใช้สินทรัพย์ รวมถึงเงินปันผล และราคาหุ้นของกลุ่มบริษัทดังกล่าวสูงกว่ากลุ่มบริษัทประกันวินาศภัยที่เน้นการประกันภัยรถยนต์ อีกทั้งยังมีสภาพคล่องที่ดีและมีความสามารถในการจ่ายค่าสินไหมที่ดีกว่า

3. วิธีดำเนินการวิจัย

1. รวบรวมข้อมูลทุติยภูมิ เบี้ยประกันที่ถือเป็นรายได้ ค่าสินไหมทดแทนสุทธิ จากเว็ปไซต์ของแต่ละบริษัทประกันวินาศภัยที่มีการดำเนินธุรกิจในปัจจุบันจำนวน 55 บริษัท ตั้งแต่ปีพุทธศักราช 2559 – 2562 เพื่อคำนวณอัตราส่วนค่าสินไหมทดแทน
2. รวบรวมข้อมูลทุติยภูมิ เบี้ยประกันภัยสุทธิ ค่าใช้จ่ายในการรับประกันภัย ค่าใช้จ่ายในการดำเนินงาน และค่าจ้างและค่าบำเหน็จ จากเว็ปไซต์ของแต่ละบริษัทประกันวินาศภัยที่มีการดำเนินธุรกิจในปัจจุบันจำนวน 55 บริษัท ตั้งแต่ปีพุทธศักราช 2559 – 2562 เพื่อคำนวณอัตราส่วนค่าใช้จ่าย
3. จัดกลุ่มบริษัทประกันวินาศภัยจากอัตราส่วนค่าใช้จ่าย โดยเรียงลำดับข้อมูลที่มีอยู่จากสูงไปต่ำ และทำการแบ่งข้อมูลออกเป็น 3 กลุ่ม กลุ่มละเท่ากัน และคำนวณค่าเฉลี่ยของอัตราส่วนค่าใช้จ่าย แต่ละกลุ่ม
4. วิเคราะห์หาฟังก์ชันการแจกแจงความน่าจะเป็นที่เหมาะสมสำหรับอัตราส่วนค่าสินไหมทดแทน ของบริษัทประกันวินาศภัยในแต่ละกลุ่ม และทำการทดสอบภาวะสารูปดีแอนเดอร์สัน – ดาร์ลิง
5. วิเคราะห์ความน่าจะเป็นในการทำกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัยแต่ละกลุ่ม

4. ผลการวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อศึกษาความสามารถในการทำกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัย จากการศึกษาพบว่ากำไรจากการรับประกันภัยนั้นขึ้นกับค่าใช้จ่ายสองส่วนหลัก คือ ค่าสินไหมทดแทน และ ค่าใช้จ่ายอื่น ๆ โดยที่ค่าสินไหมทดแทนเป็นค่าใช้จ่ายที่บริษัทประกันวินาศภัยไม่สามารถกำหนดให้เป็นไปตามต้องการได้ แตกต่างจากค่าใช้จ่ายอื่น ๆ อันประกอบไปด้วย ค่าใช้จ่ายในการรับประกันภัย ค่าใช้จ่ายในการดำเนินงาน และค่าจ้างค่าบำเหน็จ ซึ่งเป็นค่าใช้จ่ายที่บริษัทสามารถกำหนดให้เป็นไปตามนโยบายของบริษัท

ผู้วิจัยจึงมีความสนใจในการพยากรณ์ค่าของอัตราส่วนค่าสินไหมทดแทน โดยได้ทำการหาความสัมพันธ์ระหว่างอัตราส่วนค่าสินไหมทดแทนกับตัวแปรต้นที่ได้จากข้อมูลที่มีอยู่ จากตารางที่ 1 พบว่าความสัมพันธ์ระหว่างอัตราส่วนค่าสินไหมทดแทนกับตัวแปรต้นแต่ละตัวมีความสัมพันธ์กันค่อนข้างต่ำ ผู้วิจัยจึงเห็นว่าตัวแปรต้นที่ได้จากข้อมูลที่มีอยู่ไม่สามารถนำมาใช้ในการพยากรณ์ค่าอัตราส่วนค่าสินไหมทดแทนได้ ดังนั้นผู้วิจัยจึงจะใช้ข้อมูลในอดีตของอัตราส่วนค่าสินไหมทดแทนมาวิเคราะห์หาฟังก์ชันการแจกแจงความน่าจะเป็นที่เหมาะสมเพื่อที่จะใช้อธิบายโอกาสในการเกิดอัตราส่วนค่าสินไหมทดแทน

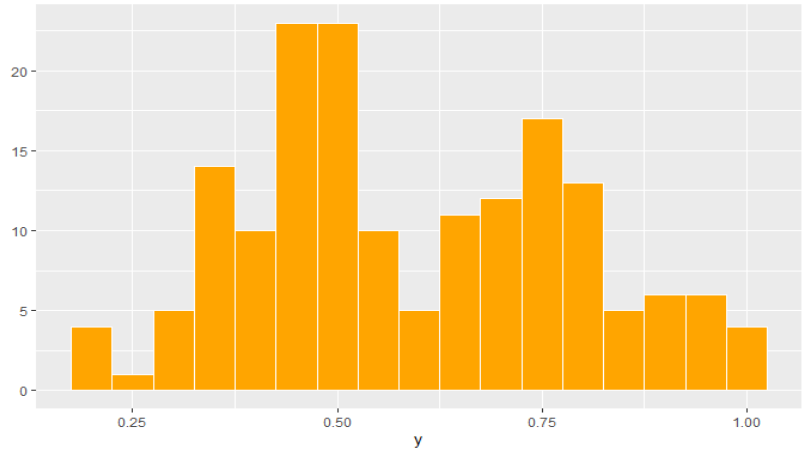
ตารางที่ 1 ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน ระหว่างอัตราส่วนค่าสินไหมทดแทนกับตัวแปรอื่น ๆ

	Commission Ratio	Underwriting Expense Ratio	Operating Expense Ratio	Company Size Small (Dummy)	Company Size Medium (Dummy)	Company Size Large (Dummy)
Loss Ratio	-0.192	-0.157	-0.256	-0.169	0.205	-0.046

ผู้วิจัยได้แบ่งการวิเคราะห์เป็นสองส่วนคือ อัตราส่วนค่าใช้จ่าย และอัตราส่วนค่าสินไหมทดแทน หลังจากนั้นจะนำผลการวิเคราะห์ที่ได้ในแต่ละส่วนมาหาความสามารถในการทำกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัย

4.1 ผลการวิเคราะห์อัตราส่วนอัตราส่วนค่าใช้จ่าย

เนื่องจากค่าใช้จ่ายในส่วนนี้จะประกอบด้วย ค่าใช้จ่ายในการรับประกันภัย ค่าใช้จ่ายในการดำเนินงาน และค่าจ้างค่าบำเหน็จ ซึ่งเป็นค่าใช้จ่ายที่บริษัทประกันวินาศภัยสามารถกำหนดให้เป็นไปตามต้องการได้ ดังนั้นผู้วิจัยจึงจะใช้ค่าใช้จ่ายส่วนนี้ในการแบ่งบริษัทวินาศภัยออกเป็น 3 กลุ่มได้แก่ 1.บริษัทประกันวินาศภัยที่มีอัตราส่วนค่าใช้จ่ายในระดับสูง 2.บริษัทประกันวินาศภัยที่มีอัตราส่วนค่าใช้จ่ายในระดับกลาง และ 3.บริษัทประกันวินาศภัยที่มีอัตราส่วนค่าใช้จ่ายในระดับต่ำ



ภาพที่ 2 แผนภาพฮิสโทแกรมของอัตราส่วนค่าใช้จ่าย
ของแต่ละบริษัทประกันวินาศภัยที่มีการดำเนินธุรกิจในปัจจุบัน จำนวน 220 ตัวอย่าง

จากภาพที่ 2 จะเห็นได้ว่าข้อมูลในแผนภาพฮิสโทแกรมมีลักษณะคล้ายข้อมูลที่มีฐานนิยม 3 ค่า ดังนั้นในขั้นตอนการแบ่งบริษัทวินาศภัยจากอัตราส่วนค่าใช้จ่าย ผู้วิจัยจึงทำการแบ่งข้อมูลที่มีอยู่ออกเป็น 3 กลุ่ม โดยใช้เปอร์เซ็นไทล์ (Percentile) ในการแบ่งข้อมูล หลังจากนั้นจึงคำนวณค่าเฉลี่ยของแต่ละกลุ่ม ได้ผลลัพธ์ดังนี้

ตารางที่ 2 ค่าเฉลี่ยของอัตราส่วนค่าใช้จ่าย ในแต่ละกลุ่ม

กลุ่มของบริษัทประกันวินาศภัยที่แบ่งจาก อัตราส่วนค่าใช้จ่าย	ค่าเฉลี่ยอัตราส่วนค่าใช้จ่าย
อัตราส่วนค่าใช้จ่ายสูง	0.82
อัตราส่วนค่าใช้จ่ายปานกลาง	0.57
อัตราส่วนค่าใช้จ่ายต่ำ	0.27

4.2 ผลการวิเคราะห์อัตราส่วนค่าสินไหมทดแทน

เนื่องจากค่าสินไหมทดแทนเป็นค่าใช้จ่ายที่บริษัทประกันวินาศภัยไม่สามารถกำหนดหรือควบคุมให้เป็นตามต้องการได้ ดังนั้นค่าสินไหมทดแทนจึงนับเป็นตัวแปรสุ่มตัวหนึ่ง การวิเคราะห์ในส่วนนี้ผู้วิจัยจึงจะทำการวิเคราะห์หาฟังก์ชันการแจกแจงความน่าจะเป็นที่เหมาะสมสำหรับอัตราส่วนค่าสินไหมทดแทน ของบริษัทประกันวินาศภัยในแต่ละกลุ่มที่ได้จากการแบ่งในขั้นตอนก่อนหน้า

ผู้วิจัยใช้โปรแกรม SPSS (Version : Statistics 25) เพื่อทำการทดสอบว่าข้อมูลอัตราส่วนสินไหมทดแทน ในแต่ละกลุ่มมีการแจกแจงความน่าจะเป็นที่มีความเหมือนหรือแตกต่างกัน ด้วยวิธีการทดสอบครัสคัล - วัลลิส (Kruskal-Wallis Test) (Samantha Lomuscio, 2021) พบว่ามีค่า P-value เท่ากับ 0.006 ซึ่งน้อยกว่าระดับนัยสำคัญที่ 0.05 จึงปฏิเสธสมมติฐานที่ว่า ข้อมูลแต่ละกลุ่มมีค่ามัธยฐานที่เท่ากันและมีการแจกแจงที่เหมือนกัน ดังนั้นการแจกแจงความน่าจะเป็นของอัตราส่วนสินไหมค่าทดแทนจึงมีความแตกต่างกันระหว่าง 3 กลุ่มดังแสดงในภาพที่ 3 จากนั้นจึงจับคู่ทดสอบระหว่างกลุ่ม เพื่อทดสอบว่าข้อมูลในแต่ละคู่มีการแจกแจงความน่าจะเป็นหมือนหรือต่างกันด้วยวิธีการทดสอบแมนน์ - วิตนีย์ ยู (Mann-Whitney U test) (Rosie Shier, 2004) ผลการทดสอบพบว่า คู่กลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายในระดับปานกลางและระดับต่ำ มีค่า P-value เท่ากับ 0.063 ซึ่งมากกว่าระดับนัยสำคัญที่ 0.05 จึงไม่ปฏิเสธสมมติฐานที่ว่าข้อมูลทั้งสองกลุ่มมีค่ามัธยฐานที่เท่ากันและมีการแจกแจงที่เหมือนกัน ในขณะที่คู่อื่นมีค่า P-value ที่น้อยกว่าระดับนัยสำคัญที่ 0.05 จึงปฏิเสธสมมติฐานการทดสอบ

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of LR is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	.006	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

ภาพที่ 3 ผลการทดสอบครัสคัล - วัลลิส ระหว่างข้อมูลอัตราส่วนค่าสินไหมทดแทน 3 กลุ่ม

Hypothesis Test Summary

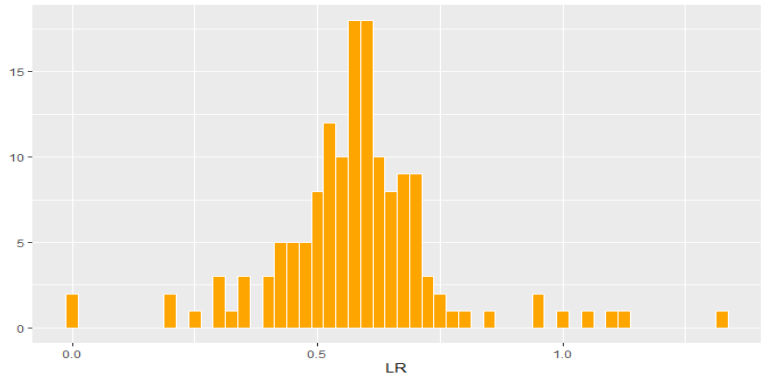
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of LR is the same across categories of GROUP.	Independent-Samples Mann-Whitney U Test	.063	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

ภาพที่ 4 ผลการทดสอบแมนน์ - วิตนีย์ ยู ระหว่างข้อมูลอัตราส่วนค่าสินไหมทดแทน
ของกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายในระดับปานกลางและระดับต่ำ

หลังจากนั้นผู้วิจัยจึงทำการวิเคราะห์หาฟังก์ชันการแจกแจงความน่าจะเป็นที่มีความเหมาะสมกับอัตราส่วนค่าสินไหมทดแทน ในแต่ละกลุ่ม โดยมีผลลัพธ์การวิเคราะห์ของแต่ละกลุ่มดังนี้

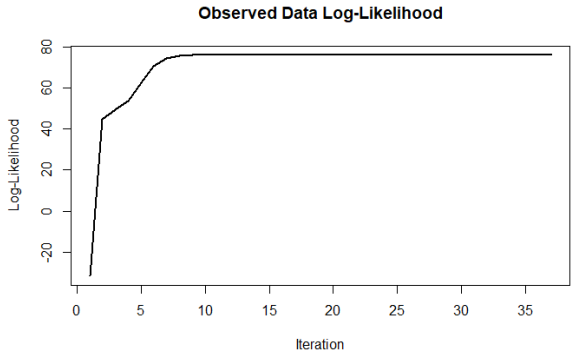
กลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายในระดับปานกลางและต่ำ มีการแจกแจงความน่าเป็นของอัตราส่วนสินไหมทดแทน ที่เหมาะสม คือการแจกแจงปกติแบบผสม (Normal Mixture) โดยผู้วิจัยใช้วิธีอัลกอริทึม EM (Expectation-Maximization) เพื่อประมาณค่าพารามิเตอร์และค่าถ่วงน้ำหนัก และใช้โปรแกรม R ในการวิเคราะห์ข้อมูล ได้ผลลัพธ์การวิเคราะห์ดังนี้



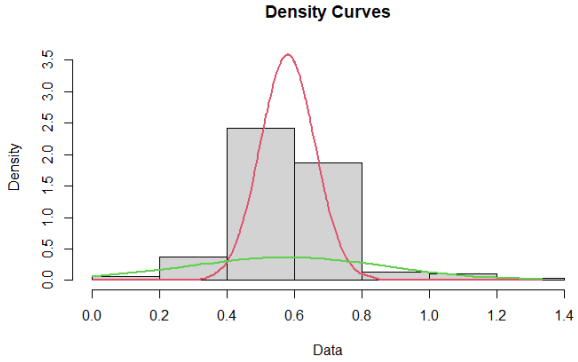
ภาพที่ 5 แผนภาพฮิสโทแกรมของอัตราส่วนค่าสินไหมทดแทน สำหรับกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายในระดับระดับปานกลาง และต่ำ

ตารางที่ 3 ค่าพารามิเตอร์และค่าถ่วงน้ำหนักของการแจกแจงปกติแบบผสมโดยวิธีอัลกอริทึม EM

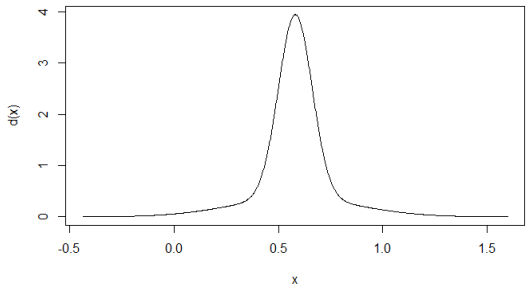
องค์ประกอบ	ค่าถ่วงน้ำหนัก ($\boldsymbol{\lambda}$)	ค่าเฉลี่ย ($\boldsymbol{\mu}$)	ค่าความแปรปรวน ($\boldsymbol{\sigma^2}$)
1	0.7257	0.5808	0.0065
2	0.2743	0.5784	0.0913



ภาพที่ 6 แผนภาพล็อกภาวะน่าจะเป็น (log-likelihood) ของการประมาณค่าพารามิเตอร์ด้วยวิธีอัลกอริทึม EM



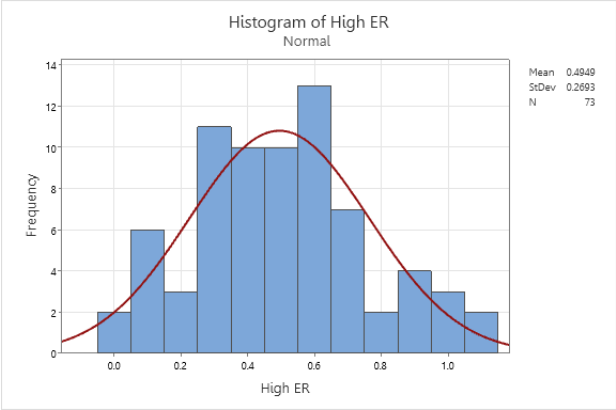
ภาพที่ 7 โค้งความหนาแน่น (Density curve) เปรียบเทียบระหว่างข้อมูลจริงและการแจกแจงที่ได้จากวิธีอัลกอริทึม EM



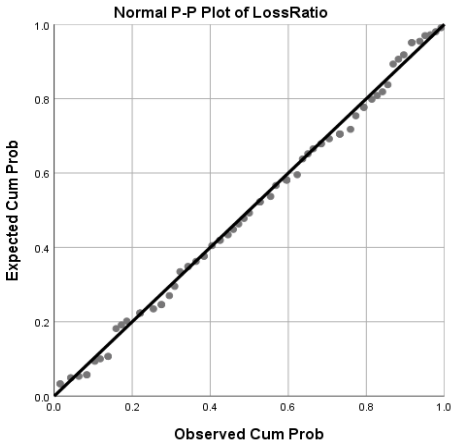
ภาพที่ 8 โค้งความหนาแน่นของการแจกแจงที่ได้จากวิธีอัลกอริทึม EM

หลังจากที่ได้ค่าพารามิเตอร์และค่าถ่วงน้ำหนักดังตารางที่ 3 จึงนำมาทดสอบภาวะสารูปดีแอนเดอร์สัน – ดาร์ลิง ได้ผลการทดสอบ ค่าสถิติแอนเดอร์สัน - ดาร์ลิง เท่ากับ 0.2502 และ P-Value เท่ากับ 0.9703 ซึ่งมากกว่าระดับนัยสำคัญที่ 0.05 ดังนั้นการแจกแจงปกติแบบผสม (Normal Mixture) ที่มีพารามิเตอร์และค่าถ่วงน้ำหนักดังกล่าว จึงมีความเหมาะสมกับข้อมูลอัตราส่วนค่าสินไหมทดแทน สำหรับกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายในระดับปานกลาง และต่ำ

ในส่วนของกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายในระดับสูง ผู้วิจัยได้ทำการวิเคราะห์จากโปรแกรม SPSS พบว่าการแจกแจงความน่าเป็นของอัตราส่วนสินไหมทดแทน ที่เหมาะสม คือการแจกแจงปกติที่มีค่าพารามิเตอร์ ค่าเฉลี่ย (μ) เท่ากับ 0.4949 และ ค่าความแปรปรวน (σ^2) เท่ากับ 0.0725 โดยมีผลการทดสอบภาวะสารูปดีแอนเดอร์สัน – ดาร์ลิง ดังนี้ ค่าสถิติแอนเดอร์สัน - ดาร์ลิง เท่ากับ 0.262 และ P-Value เท่ากับ 0.695 ซึ่งมากกว่าระดับนัยสำคัญที่ 0.05 ดังนั้น การแจกแจงปกติ จึงมีความเหมาะสมกับข้อมูลอัตราส่วนค่าสินไหมทดแทน สำหรับกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายในระดับสูง



ภาพที่ 9 แผนภาพฮิสโทแกรมของอัตราส่วนค่าสินไหมทดแทน
สำหรับกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายในระดับสูง



ภาพที่ 10 แผนภาพ P-P Plot การทดสอบการแจกแจงปกติของอัตราส่วนค่าสินไหมทดแทน สำหรับกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายในระดับสูง

4.3 ผลการวิเคราะห์ความสามารถในการทำกำไรจากการรับประกันภัย (Underwriting Profit Margin)

บริษัทประกันวินาศภัยจะสามารถทำกำไรจากการรับประกันภัยได้ ก็ต่อเมื่อผลรวมของอัตราส่วนค่าสินไหมทดแทน และอัตราส่วนค่าใช้จ่าย รวมกันจะต้องมีค่าไม่เกิน 1 จากผลการแบ่งกลุ่มบริษัทวินาศภัยจากอัตราส่วนค่าใช้จ่ายในตารางที่ 2 จะเห็นได้ว่าความสามารถในการทำกำไรจากการรับประกันภัยของบริษัทในแต่ละกลุ่ม จะขึ้นอยู่กับอัตราส่วนค่าสินไหมทดแทนที่สามารถรองรับได้เพื่อที่บริษัทจะยังคงมีกำไรจากการรรับประกันภัย โดยคำนวณจากนำ 1 ลบออกด้วยค่าเฉลี่ยของอัตราส่วนค่าใช้จ่ายในแต่ละกลุ่ม มีค่าดังนี้

ตารางที่ 4 อัตราส่วนค่าสินไหมทดแทนสูงสุดที่สามารถรองรับได้ของบริษัทในแต่ละกลุ่ม

กลุ่มของบริษัทประกันวินาศภัย ที่แบ่งจากอัตราส่วนค่าใช้จ่าย	ค่าเฉลี่ยอัตราส่วนค่าใช้จ่าย	อัตราส่วนค่าสินไหมทดแทน สูงสุดที่สามารถรองรับได้
อัตราส่วนค่าใช้จ่ายสูง	0.82	0.18
อัตราส่วนค่าใช้จ่ายปานกลาง	0.57	0.43
อัตราส่วนค่าใช้จ่ายต่ำ	0.27	0.73

บริษัทประกันวินาศภัยในแต่ละกลุ่มจะสามารถทำกำไรจากการรับประกันภัยได้ เมื่อบริษัทมีอัตราส่วนค่าสินไหมทดแทนไม่เกินกว่าอัตราส่วนค่าสินไหมทดแทนสูงสุดที่สามารถรองรับได้ ดังนั้นผู้วิจัยจึงพิจารณาความน่าจะเป็นที่บริษัทในแต่ละกลุ่มจะทำกำไรจากการรับประกันภัย จากการหาความน่าจะเป็นที่บริษัทในแต่ละกลุ่มจะมีอัตราส่วนค่าสินไหมทดแทนน้อยกว่าหรือเท่ากับอัตราส่วนค่าสินไหมทดแทนสูงสุดที่สามารถรองรับได้ โดยใช้ฟังก์ชันการแจกแจงความน่าจะเป็นสะสมของอัตราส่วนค่าสินไหมทดแทนในแต่ละกลุ่มที่หามาได้ก่อนหน้า โดยมีผลลัพธ์ในการวิเคราะห์ดังนี้

ตารางที่ 5 ความน่าจะเป็นที่จะทำกำไรจากการรับประกันภัยของบริษัทในแต่ละกลุ่ม

กลุ่มของบริษัทประกันวินาศภัยที่แบ่งจากอัตราส่วนค่าใช้จ่าย	อัตราส่วนค่าสินไหมทดแทนสูงสุดที่สามารถรองรับได้	ความน่าจะเป็นที่จะทำกำไรจากการรับประกันภัย (Underwriting Profit margin)
อัตราส่วนค่าใช้จ่ายสูง	0.18	F(0.18) = 0.1211
อัตราส่วนค่าใช้จ่ายปานกลาง	0.43	F(0.43) = 0.1080
อัตราส่วนค่าใช้จ่ายต่ำ	0.73	F(0.73) = 0.8920

หมายเหตุ: $F(x)$ คือฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่ม x เมื่อ x คืออัตราส่วนค่าสินไหมทดแทน

ตัวอย่างการคำนวณความน่าจะเป็นที่จะทำกำไรจากการรับประกันภัย (Underwriting Profit Margin) ของบริษัทประกันวินาศภัยที่มีอัตราส่วนค่าใช้จ่ายในระดับต่ำ จากตารางที่ 2 จะเห็นว่าบริษัทในกลุ่มนี้มีค่าเฉลี่ยอัตราส่วนค่าใช้จ่าย เท่ากับ 0.27 ประกอบกับบริษัทประกันวินาศภัยจะสามารถทำกำไรจากการรับประกันภัยได้ก็ต่อเมื่ออัตราส่วนรวม มีค่าไม่เกิน 1 ดังนั้นเมื่อพิจารณาแล้วจะพบว่าบริษัทในกลุ่มนี้จะสามารถทำกำไรจากการรับประกันภัยได้เมื่ออัตราส่วนค่าสินไหมทดแทน จะต้องมีค่าไม่เกินกว่า 0.73 ดังแสดงในตารางที่ 5 หลังจากนั้นผู้วิจัยจึงนำค่าดังกล่าวไปใช้ในการหาความน่าจะเป็นที่บริษัทประกันวินาศภัยในกลุ่มนี้จะมีอัตราส่วนค่าสินไหมทดแทนไม่เกินกว่า 0.73 โดยการใช้ฟังก์ชันการแจกแจงสะสมของการแจกแจงผสมแบบปกติที่มีค่าพารามิเตอร์และค่าถ่วงน้ำหนักดังตารางที่ 3 ที่วิเคราะห์มาจากขั้นตอนก่อนหน้า จะได้ว่าความน่าจะเป็นที่จะทำกำไรจากการรับประกันภัยของกลุ่มบริษัทประกันวินาศภัยที่มีอัตราส่วนค่าใช้จ่ายในระดับต่ำ มีค่าดังนี้ F(0.73) = 0.8920

ทั้งนี้ผู้วิจัยได้ทำการวิเคราะห์เพิ่มเติมในส่วนของอัตราส่วนค่าใช้จ่ายของแต่ละกลุ่มบริษัทประกันวินาศภัย โดยคำนวณหาสัดส่วนระหว่าง อัตราส่วนค่าใช้จ่ายในการรับประกันภัย อัตราส่วนค่าใช้จ่ายในการดำเนินงาน และอัตราส่วนค่าจ้างค่าบำเหน็จสุทธิ ต่ออัตราส่วนค่าใช้จ่ายทั้งหมด และพบว่าทั้ง 3 องค์ประกอบของอัตราส่วนค่าใช้จ่ายมีสัดส่วนที่ใกล้เคียงกันในทุกกลุ่มบริษัทประกันวินาศภัย ดังนั้นจึงสามารถสรุปได้ว่า โดยเฉลี่ยแล้วบริษัทประกันวินาศภัยจะบริหารจัดการให้อัตราส่วนค่าใช้จ่ายทั้งสามส่วนนี้อยู่ในระดับที่เท่ากัน

5. สรุปผลการวิจัย

จากการวิเคราะห์ความสามารถในการทำกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัยพบว่ากำไรจากการรับประกันภัยขึ้นอยู่กับค่าใช้จ่ายสองส่วนหลักได้แก่ ค่าใช้จ่ายอื่นๆ และค่าสินไหมทดแทน ในส่วนของค่าใช้จ่ายอื่นๆ เป็นค่าใช้จ่ายที่บริษัทประกันวินาศภัยสามารถควบคุมหรือกำหนดให้เป็นดังต้องการได้ ดังนั้นผู้วิจัยจึงใช้ค่าใช้จ่ายนี้ในการแบ่งบริษัทประกันวินาศภัยออกเป็น 3 กลุ่ม ได้แก่ บริษัทที่มีอัตราส่วนค่าใช้จ่ายสูง ปานกลาง และ ต่ำ ตามลำดับ แล้วจึงหาค่าเฉลี่ยของอัตราส่วนค่าใช้จ่ายในแต่ละกลุ่ม ในขณะที่ค่าสินไหมทดแทนนั้นเป็นค่าใช้จ่ายที่บริษัทประกันวินาศภัยไม่สามารถควบคุมหรือกำหนดให้เป็นไปตามต้องการได้ ดังนั้นค่าใช้จ่ายในส่วนนี้ จึงนับเป็นตัวแปรสุ่มตัวหนึ่ง ผู้วิจัยจึงทำการวิเคราะห์หาฟังก์ชันการแจกแจงความน่าจะเป็นที่เหมาะสมสำหรับอัตราส่วนค่าสินไหมทดแทนของบริษัทในแต่ละกลุ่มพบว่า ฟังก์ชันการแจกแจงความน่าจะเป็นที่มีความเหมาะสมกับอัตราส่วนค่าสินไหมทดแทนของกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายปานกลางและต่ำ คือฟังก์ชันการแจกแจงปกติแบบผสม และฟังก์ชันการแจกแจงความน่าจะเป็นที่มีความเหมาะสมกับอัตราส่วนค่าสินไหมทดแทนของกลุ่มบริษัทที่มีอัตราส่วนค่าใช้จ่ายสูง คือฟังก์ชันการแจกแจงปกติ

บริษัทประกันวินาศภัยจะมีกำไรจากการรับประกันภัยเมื่อผลรวมของอัตราส่วนค่าสินไหมทดแทนและอัตราส่วนค่าใช้จ่ายมีค่าไม่เกิน 1 ดังนั้นเราจึงสามารถหาความน่าจะเป็นในการทำกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัยในแต่ละกลุ่มได้ดังนี้ บริษัทประกันวินาศภัยที่มีอัตราส่วนค่าใช้จ่ายอยู่ในระดับสูงจะมีความน่าจะเป็นที่จะมีกำไรจากการรับประกันภัยเท่ากับ 0.1211 บริษัทประกันวินาศภัยที่มีอัตราส่วนค่าใช้จ่ายอยู่ในระดับปานกลางจะมีความน่าจะเป็นที่จะมีกำไรจากการรับประกันภัยเท่ากับ 0.1080 และบริษัทประกันวินาศภัยที่มีอัตราส่วนค่าใช้จ่ายอยู่ในระดับต่ำจะมีความน่าจะเป็นที่จะมีกำไรจากการรับประกันภัยเท่ากับ 0.8920

6. ข้อเสนอแนะ

บริษัทประกันวินาศภัยสามารถนำผลการวิจัยในครั้งนี้ ไปปรับใช้ในการกำหนดแนวทางการบริหารจัดการอัตราส่วนค่าใช้จ่ายต่างๆ ของบริษัท เพื่อให้บริษัทสามารถทำกำไรจากการรับประกันภัยได้ เช่น การบริหารจัดการอัตราส่วนค่าใช้จ่ายซึ่งเป็นอัตราส่วนที่บริษัทประกันวินาศภัยสามารถควบคุมได้ ให้มีค่าอยู่ในระดับต่ำ จะเห็นได้ว่าจากผลการวิจัย บริษัทที่มีอัตราส่วนค่าใช้จ่ายที่อยู่ในระดับต่ำ จะมีโอกาสในการทำกำไรจากการรับประกันภัยถึงร้อยละ 90 โดยเมื่อเทียบกับบริษัทที่มีอัตราส่วนค่าใช้จ่ายอยู่ในระดับปานกลาง และระดับสูง จะเห็นได้ว่าบริษัทเหล่านี้มีโอกาสในการทำกำไรจากการรับประกันภัยลดลงเป็นอย่างมาก ซึ่งพบว่าโอกาสในการทำไรจากการรับประกันภัยลดลงเหลือเพียงร้อยละ 10 เท่านั้น อย่างไรก็ตามเนื่องจากในการวิจัยครั้งนี้ผู้วิจัยศึกษาเฉพาะกำไรจากการรับประกันภัยของบริษัทประกันวินาศภัยเท่านั้น ซึ่งกำไรจากการรับประกันภัยเป็นกำไรที่มาจากรายได้หลักของบริษัทประกันวินาศภัยซึ่งคือเบี้ยประกันภัยเพียงอย่างเดียว ในความเป็นจริงบางบริษัทประกันวินาศภัยอาจยังมีแหล่งรายได้จากส่วนอื่นๆด้วย เช่น รายได้จากการลงทุน เป็นต้น นอกจากนี้ด้วยข้อจำกัดของข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลภาพรวมของบริษัทประกันวินาศภัย ซึ่งประกอบด้วยการประกันภัยหลากหลายประเภทแตกต่างกันในแต่ละบริษัท ดังนั้นหากนำรายได้ส่วนอื่นๆ มาร่วมวิเคราะห์ หรือสามารถแบ่งการวิเคราะห์แยกตามประเภทของการประกันภัยได้ อาจทำให้สามารถวิเคราะห์ความสามารถในการทำไรของบริษัทประกันวินาศภัยได้ดียิ่งขึ้น ทั้งนี้เพื่อเป็นแนวทางในการพัฒนารูปแบบในการบริหารจัดการค่าใช้จ่ายในส่วนต่างๆของบริษัทประกันวินาศภัยต่อไป

7. เอกสารอ้างอิง

Casella, George; Berger, Roger L. 2001. **Statistical inference**. (2^{nd} ed.). Duxbury. ISBN 0-534-24312-6.

Dempster, A. P., et al. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. **Journal of the Royal Statistical Society,** 1, 1-38.

NIST/SEMATECH. 2013. Anderson-Darling Test. **e-Handbook of Statistical Methods**. สืบค้นวันที่ 1 มีนาคม 2564 จาก https://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm.

Rosie Shier. 2004. **Statistics: 2.3 The Mann-Whitney U Test**. สืบค้นวันที่ 3 มกราคม 2565 จาก https://www.statstutor.ac.uk/resources/uploaded/mannwhitney.pdf.

Samantha Lomuscio. 2021. Getting Started with the Kruskal-Wallis Test. สืบค้นเมื่อวันที่ 28 ธันวาคม 2564 จาก https://data.library.virginia.edu/getting-started-with-the-kruskal-wallis-test.

T. W. Anderson; D. A. Darling. 1954. A Test of Goodness of fit. **Journal of the American Statistical Association**, 49(268), 765-769.

Wilfried Seidel. 2015. Mixture Models. Helmut-Schmidt-University, D-22039 Hamburg, Germany.

มธุรส อรุณรุ่งแสง. 2553. **การวิเคราะห์ฐานะทางการเงินและสัดส่วนการลงทุนของบริษัทประกันวินาศภัย**. สารนิพนธ์มหาบัณฑิต, มหาวิทยาลัยศรีนครินทรวิโรฒ.

สายทอง แจ่มใจ. 2547. **การเปรียบเทียบประสิทธิภาพของตัวสถิติทดสอบในการทดสอบภาวะสารูปสนิทดี**. วิทยานิพนธ์มหาบัณฑิต, มหาวิทยาลัยศิลปากร.

สุชาดา สถาวรวงศ์ และอธิวัฒน์ ศุภสวัสดิ์วัชร. 2564. **การวิเคราะห์ฐานะทางการเงินของบริษัทประกันวินาศภัย**. สืบค้นวันที่ 18 กุมภาพันธ์ 2564 จาก https://www.oic.or.th/sites/default/files/content/90854/04315-bththii-5-kaarwiekhraaahthaanathaangkaarenginkhngbrisathprakanwinaasphay.pdf.

สำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย. 2564. **ข้อมูลทางทะเบียนบริษัทประกันวินาศภัย.** สืบค้นวันที่ 4 เมษายน 2564 จาก https://www.oic.or.th/th/industry/statistic/data/36/2.

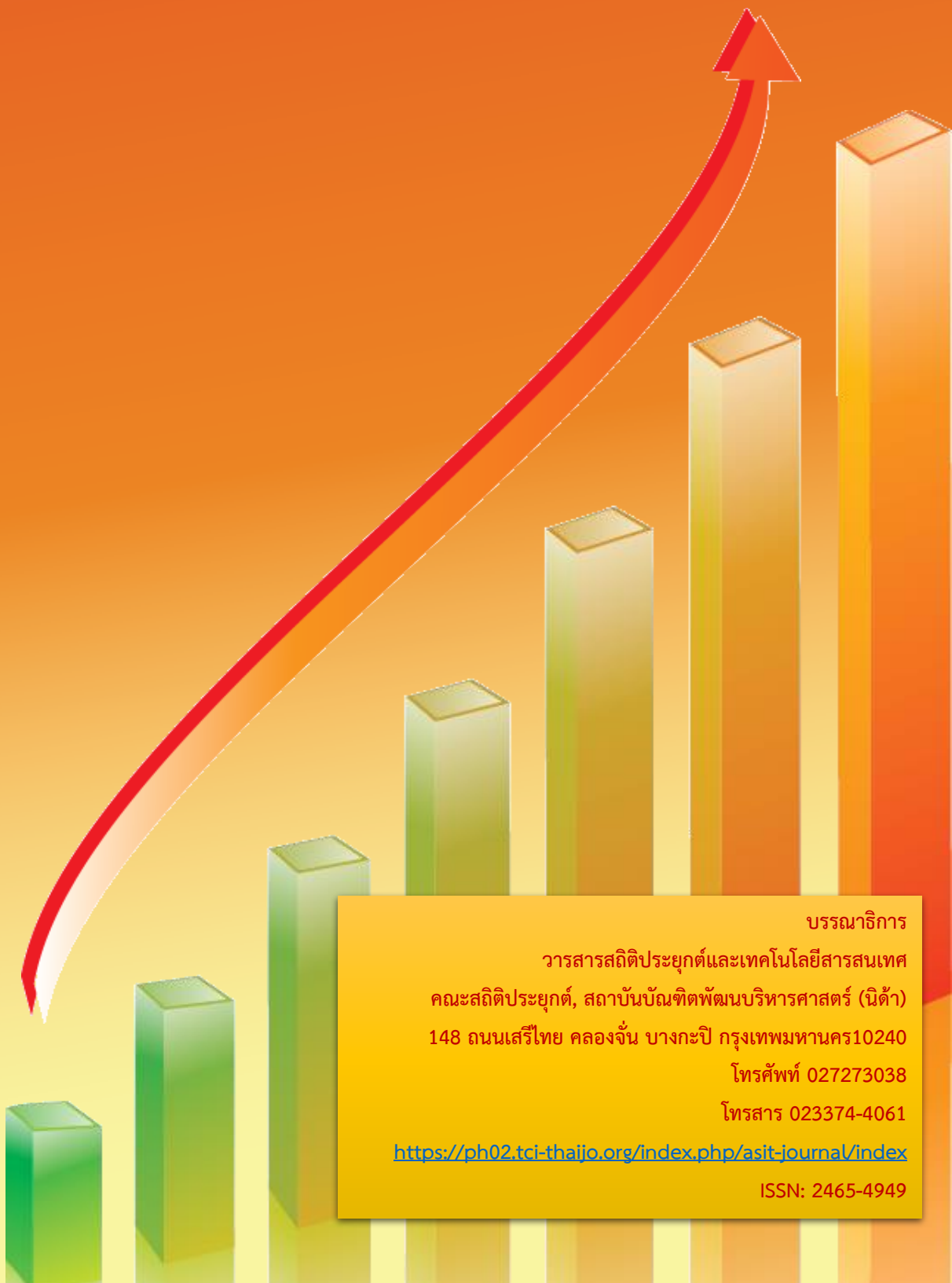
สำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย. 2564. **ผลการดำเนินงานของบริษัทประกันวินาศภัย.** สืบค้นวันที่ 10 กุมภาพันธ์ 2564 จาก https://www.oic.or.th/th/industry/statistic/data/37/2.

อธิวัฒน์ ศุภสวัสดิ์วัชร. 2564. **การจัดการพิจารณารับประกันภัยและการกำหนดอัตราเบี้ยประกันภัย**. สืบค้นวันที่ 18 กุมภาพันธ์ 2564 จาก https://www.oic.or.th/sites/default/files/content/90854/04307-bththii-3-kaarcchadkaarphicchaarnaarabprakanphayaelakaarkamhndatraaebiiyprakanphay.pdf.

ติดต่อหรือส่งบทความ

บรรณาธิการวารสารสถิติประยุกต์และเทคโนโลยีสารสนเทศ
คณะสถิติประยุกต์
สถาบันบัณฑิตพัฒนบริหารศาสตร์
148 ถ. เสรีไทย แขวงคลองจั่น
เขตบางกะปิ กทม. 10240

โทรศัพท์ : 02-727-3038
โทรสาร : 02-374-3061
อีเมลล์ : asit-journal@as.nida.ac.th
Website : https://ph02.tci-thaijo.org/index.php/asit-journal/index



บรรณาธิการ
วารสารสถิติประยุกต์และเทคโนโลยีสารสนเทศ
คณะสถิติประยุกต์, สถาบันบัณฑิตพัฒนบริหารศาสตร์ (นิด้า)
148 ถนนเสรีไทย คลองจั่น บางกะปิ กรุงเทพมหานคร10240
โทรศัพท์ 027273038
โทรสาร 023374-4061
https://ph02.tci-thaijo.org/index.php/asit-journal/index
ISSN: 2465-4949