



การปรับปรุงการตรวจจับข่าวปลอมด้วยซอฟต์แวร์แมชชีนโดยวิธีการคัดเลือกคุณลักษณะที่
เหมาะสม

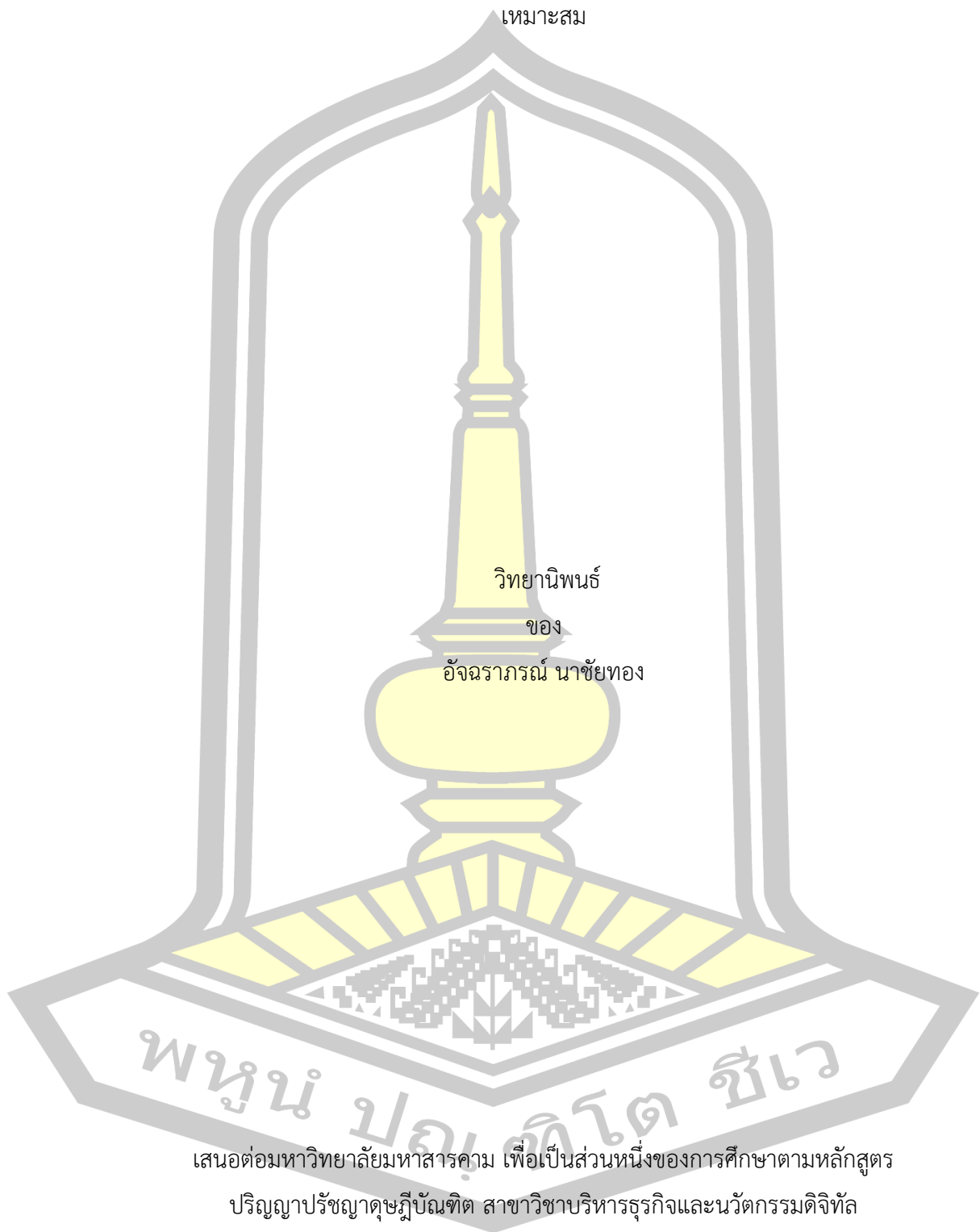
วิทยานิพนธ์
ของ
อัจฉราภรณ์ นาชัยทอง

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล
สิงหาคม 2567

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การปรับปรุงการตรวจจับข่าวปลอมด้วยซอฟต์แวร์แมชชีนโดยวิธีการคัดเลือกคุณลักษณะที่

เหมาะสม



เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

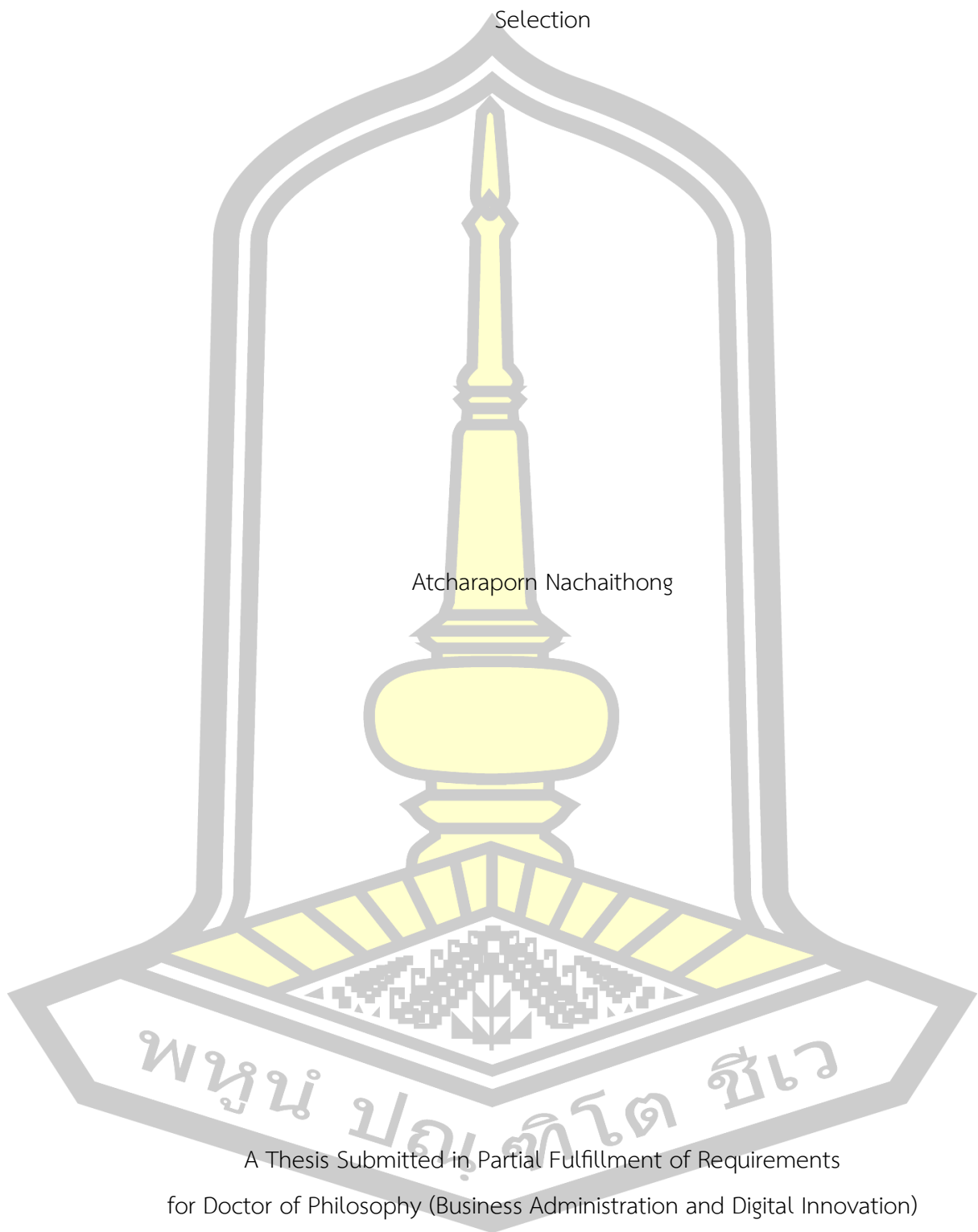
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล

สิงหาคม 2567

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Improving Fake News Detection with Support Vector Machines Using Optimal Feature
Selection

Atcharaporn Nachaithong



A Thesis Submitted in Partial Fulfillment of Requirements
for Doctor of Philosophy (Business Administration and Digital Innovation)

August 2024

Copyright of Mahasarakham University



คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนางสาวอัจฉราภรณ์ นาชัย
ทอง แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต
สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผศ. ดร. ธงชัย แก้วกิริยา)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รศ. ดร. กิตติพล วิแสง)

.....กรรมการ

(ดร. พงศธร ตันตระกูลบัณฑิตย์)

.....กรรมการ

(ดร. กาญจนา หินเระว้)

.....กรรมการ

(ดร. การินต์ กิจระการ)

มหาวิทยาลัยขอนแก่นให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญา ปรัชญาดุษฎีบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล ของมหาวิทยาลัย
มหาสารคาม

.....
(ดร. ชลธิชา ธรรมวิญญู)

คณบดีคณะการบัญชีและการจัดการ

.....
(รศ. ดร. กริสน์ ชัยมูล)

คณบดีบัณฑิตวิทยาลัย

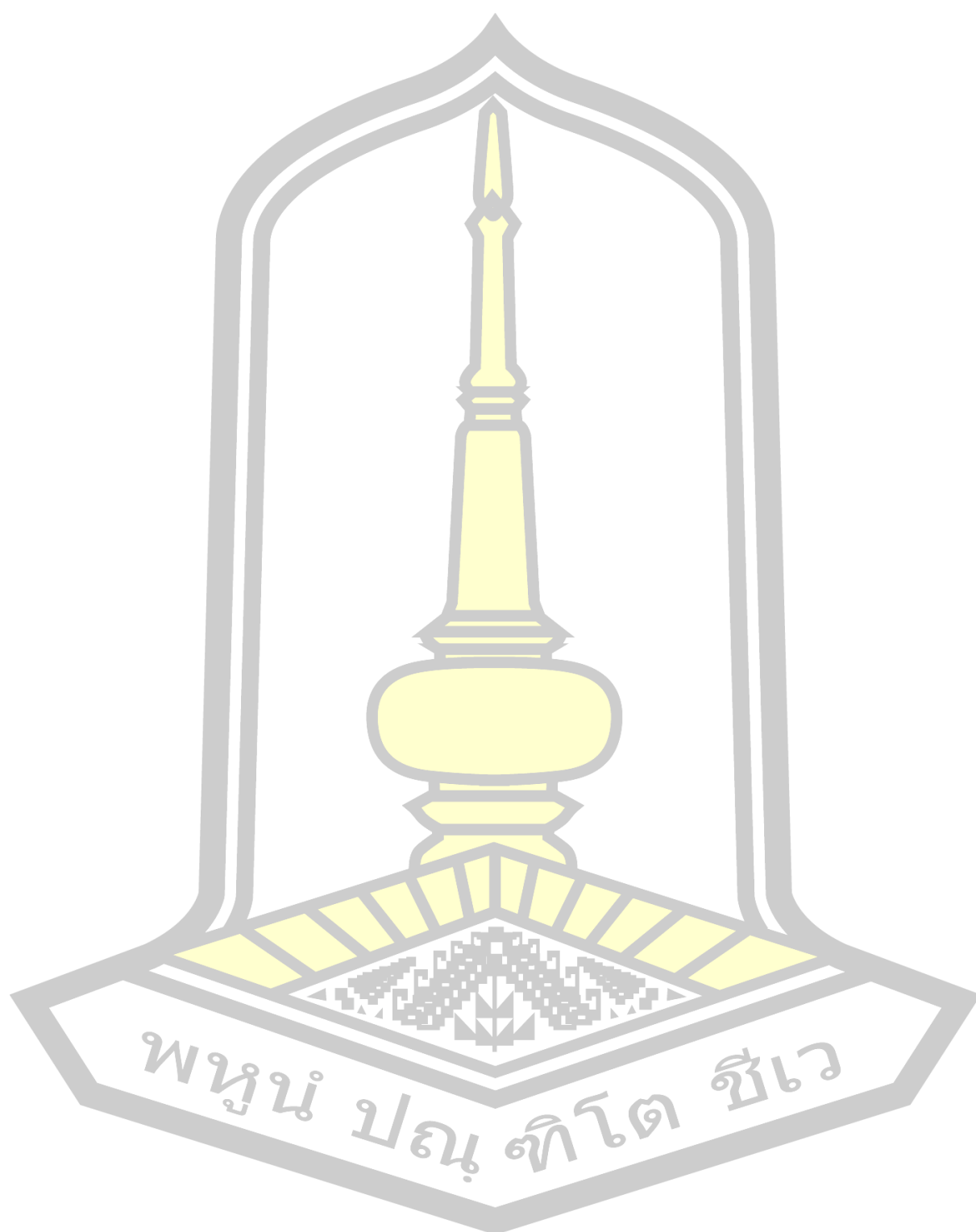
ชื่อเรื่อง	การปรับปรุงการตรวจจับข่าวปลอมด้วยซอฟต์แวร์แมชชีนโดยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม		
ผู้วิจัย	อัจฉราภรณ์ นาชัยทอง		
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. กิตติพล วิแสง		
ปริญญา	ปรัชญาดุษฎีบัณฑิต	สาขาวิชา	บริหารธุรกิจและนวัตกรรมดิจิทัล
มหาวิทยาลัย	มหาวิทยาลัยมหาสารคาม	ปีที่พิมพ์	2567

บทคัดย่อ

การวิจัยนี้เน้นการปรับปรุงการตรวจจับข่าวปลอมโดยใช้ซอฟต์แวร์แมชชีน ผ่านการคัดเลือกคุณลักษณะที่เหมาะสม เมื่อทดสอบและเปรียบเทียบประสิทธิภาพของอัลกอริทึมต่างด้วยค่าความถูกต้อง ความแม่นยำ ความระลึก และค่าถ่วงดุล ผลลัพธ์แสดงให้เห็นถึงความเพิ่มประสิทธิภาพของระบบในการตรวจจับข่าวปลอม โดยมีฐานข้อมูล Kaggle ที่มีประสิทธิภาพสูงที่สุดในการตรวจจับข่าวปลอมและการคัดเลือกคุณลักษณะที่เหมาะสมโดยวิธีการปรับปรุงอัลกอริทึมเพื่อเพิ่มประสิทธิภาพความถูกต้องของระบบตรวจจับข่าวปลอมให้มีประสิทธิภาพมากขึ้น

สรุปผลการวิจัย ผลการวิจัยแสดงให้เห็นถึงประสิทธิภาพของการปรับปรุงอัลกอริทึมตรวจจับข่าวปลอมด้วยซอฟต์แวร์แมชชีนผ่านการคัดเลือกคุณลักษณะที่เหมาะสม การทดสอบและเปรียบเทียบระหว่างอัลกอริทึมจะแสดงให้เห็นถึงความถูกต้องและความแม่นยำที่สูง โดยนาอ์ฟเบย์ให้ค่าความถูกต้องเท่ากับ 91.29% มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 94.89% วิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องเท่ากับ 97.35% มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 97.06% วิธีต้นไม้แห่งการทํานายให้ค่าความถูกต้องเท่ากับ 98.05% มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 97.74% วิธีการถดถอยโลจิสติกให้ค่าความถูกต้องเท่ากับ 94.49% มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 95.36% และวิธีซอฟต์แวร์แมชชีนให้ค่าความถูกต้องเท่ากับ 96.67% มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 99.61% จากผลการทดสอบพบว่า วิธีวิธีซอฟต์แวร์แมชชีนให้ค่าความถูกต้องในการตรวจจับข่าวปลอมเมื่อปรับปรุงคุณลักษณะแล้วมากที่สุดเท่ากับ 96.67%

คำสำคัญ : ข่าวปลอม, สื่อสังคมออนไลน์, ซอฟต์แวร์แมชชีน, การคัดเลือกคุณลักษณะ



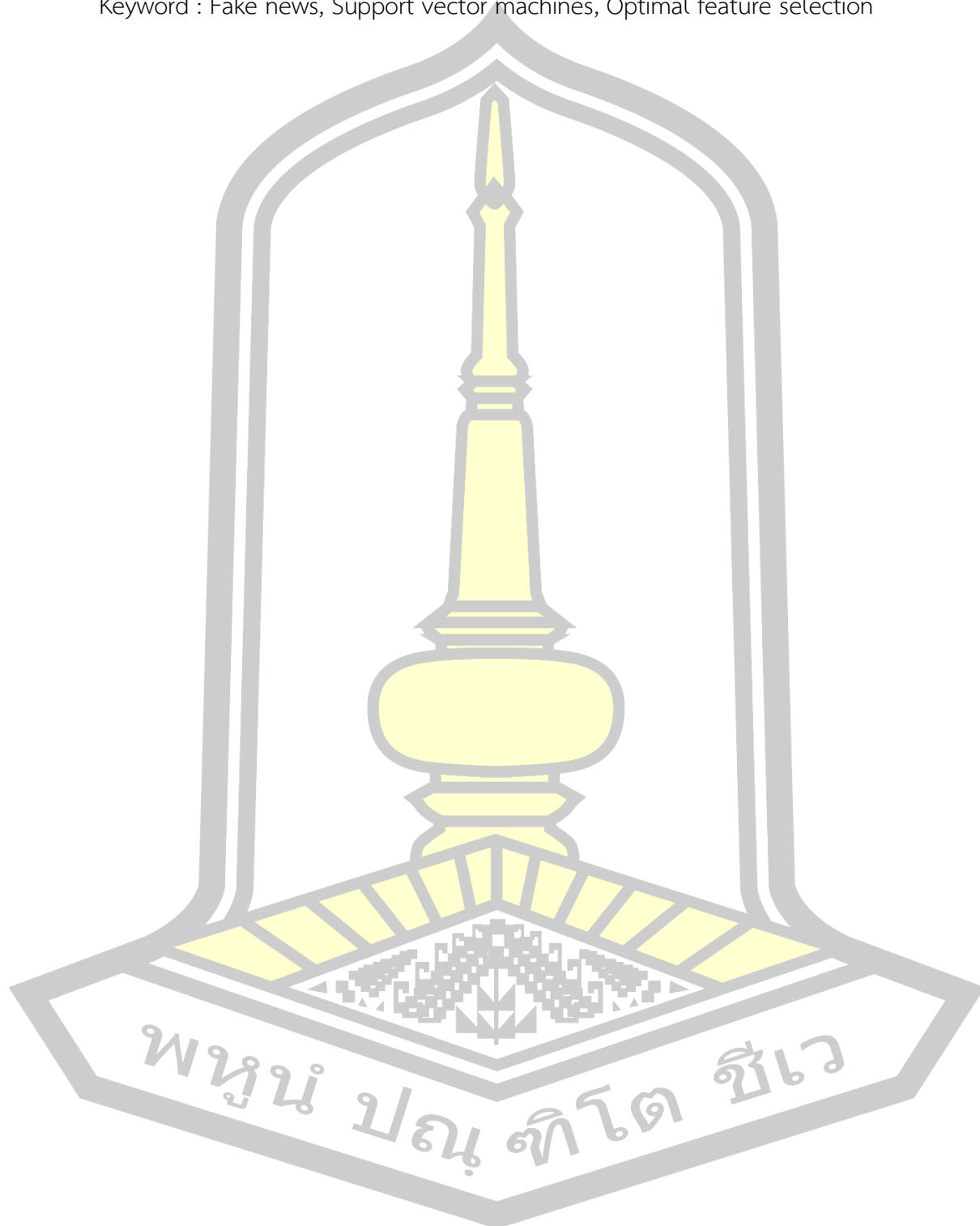
TITLE	Improving Fake News Detection with Support Vector MachinesUsing Optimal Feature Selection		
AUTHOR	Atcharaporn Nachaithong		
ADVISORS	Associate Professor Kittipol Wisaeng , Ph.D.		
DEGREE	Doctor of Philosophy	MAJOR	Business Administration and Digital Innovation
UNIVERSITY	Maharakham University	YEAR	2024

ABSTRACT

This research focuses on improving fake news detection using Support Vector Machine (SVM) through appropriate feature selection. When testing and comparing the performance of different algorithms using accuracy, precision, recall, and F1 score, the results show an enhancement in the system's ability to detect fake news. The Kaggle dataset, with its high efficiency in detecting fake news, and the feature selection method used for improving algorithm performance, contributed to a more accurate fake news detection system.

Conclusion The research demonstrates the effectiveness of improving fake news detection algorithms using Support Vector Machine through appropriate feature selection. Testing and comparing different algorithms reveal high accuracy and precision. Naive Bayes achieved an accuracy of 91.29% with a precision of 94.89% in detecting fake news. Decision Tree provided an accuracy of 97.35% with a precision of 97.06%. Random Forest had an accuracy of 98.05% and a precision of 97.74%. Logistic Regression yielded an accuracy of 94.49% with a precision of 95.36%. Support Vector Machine achieved the highest accuracy of 96.67% and a precision of 99.61% in detecting fake news. The results indicate that the Support Vector Machine, when optimized with appropriate feature selection, provides the highest accuracy of 96.67% in fake news detection.

Keyword : Fake news, Support vector machines, Optimal feature selection



กิตติกรรมประกาศ

การทำวิทยานิพนธ์เรื่อง "การปรับปรุงการตรวจจับชาวปลอมด้วยซอฟต์แวร์แมชชีน โดยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม" ฉบับนี้สำเร็จลุล่วงไปได้ด้วยการสนับสนุนและความช่วยเหลือจากหลายฝ่าย ข้าพเจ้าขอแสดงความขอบคุณและกตเวทิตาต่อบุคคลและหน่วยงานดังต่อไปนี้

ก่อนอื่น ข้าพเจ้าขอขอบพระคุณอาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.กิตติพล วิแสง ผู้ให้คำแนะนำ ชี้แนะ และตรวจสอบวิทยานิพนธ์อย่างใกล้ชิดตลอดระยะเวลาที่ผ่านมา ทั้งนี้ยังให้กำลังใจและสนับสนุนให้ข้าพเจ้าไม่ละทิ้งความพยายามจนกระทั่งวิทยานิพนธ์นี้สำเร็จลงได้

ข้าพเจ้าขอขอบคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่าน ที่ได้ให้ข้อเสนอแนะและคำวิจารณ์ อันมีคุณค่าในการปรับปรุงวิทยานิพนธ์นี้ให้มีความสมบูรณ์และมีคุณภาพยิ่งขึ้น

นอกจากนี้ ข้าพเจ้าขอขอบคุณเพื่อนนักศึกษาและเพื่อนร่วมงานทุกคน ที่ให้การสนับสนุนด้าน ข้อมูล ความรู้ และการให้กำลังใจที่สำคัญในการทำวิจัยนี้

สุดท้ายนี้ ข้าพเจ้าขอขอบคุณครอบครัวและคนรัก ที่ได้ให้กำลังใจ สนับสนุน และอยู่เคียงข้าง ข้าพเจ้าเสมอในทุกขั้นตอนของการทำวิจัยและการศึกษานี้ ข้าพเจ้าขอขอบคุณทุกท่านจากใจจริง

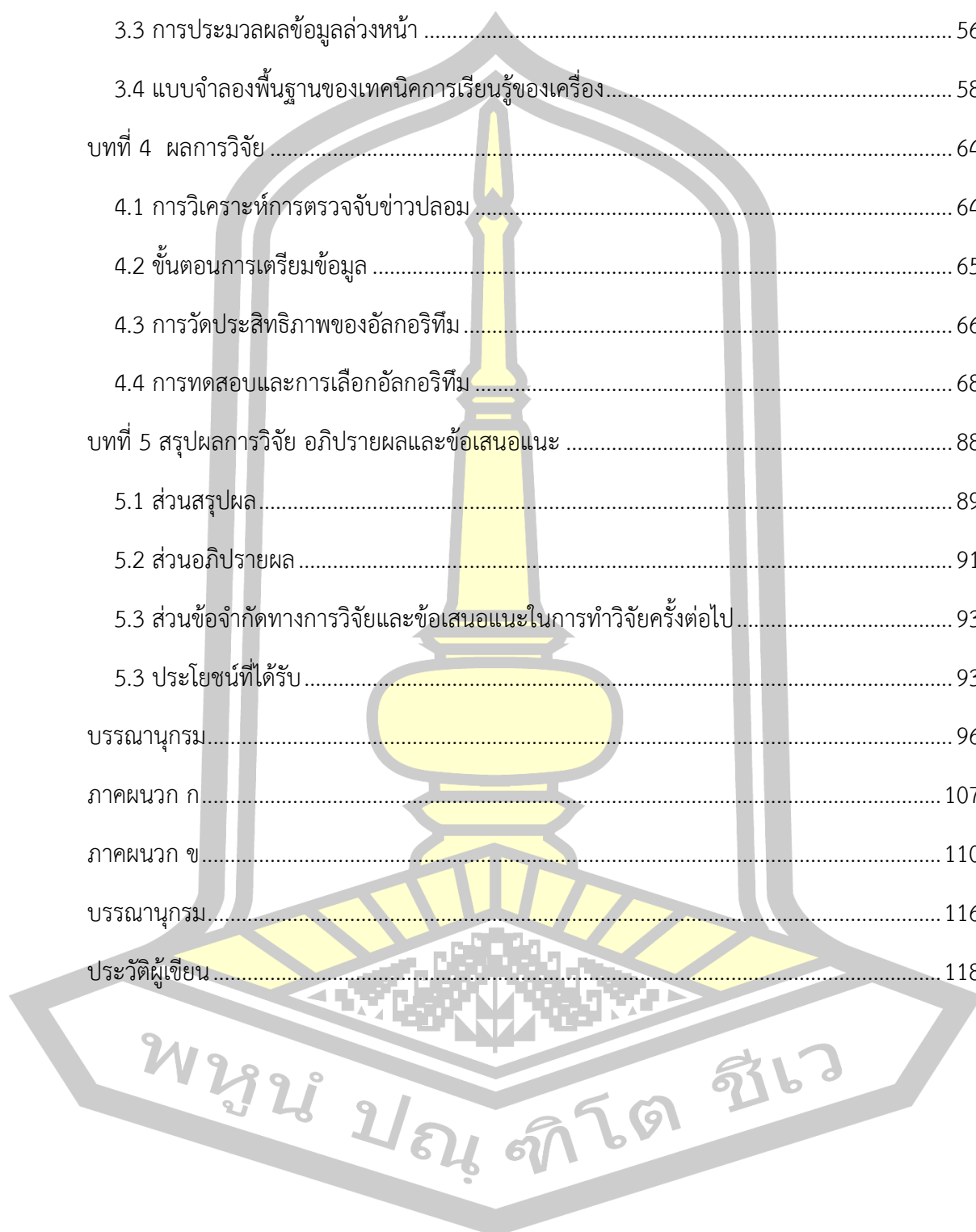
อัจฉราภรณ์ นาชัยทอง



สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	ฉ
กิตติกรรมประกาศ.....	ช
สารบัญ.....	ฌ
สารบัญตาราง.....	ฉ
สารบัญรูปภาพ.....	ฐ
บทที่ 1 บทนำ.....	1
1.1 ภูมิหลัง.....	1
1.2 วัตถุประสงค์ของการวิจัย	9
1.3 ขอบเขตของการวิจัย	9
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	12
1.5 นิยามศัพท์เฉพาะ	12
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	13
2.1 ข่าวปลอม (Fake news).....	13
2.2 สื่อสังคมออนไลน์ (Social media).....	26
2.3 ซัพพอร์ทเวกเตอร์แมชชีน (Support vector machine).....	33
2.4 การเตรียมข้อมูล (Data preprocessing).....	41
2.5 การวัดประสิทธิภาพของโมเดล (Evaluation measure).....	46
2.6 งานวิจัยที่เกี่ยวข้อง.....	48
บทที่ 3 วิธีการดำเนินงานวิจัย.....	54
3.1 ขั้นตอนและวิธีการ	54

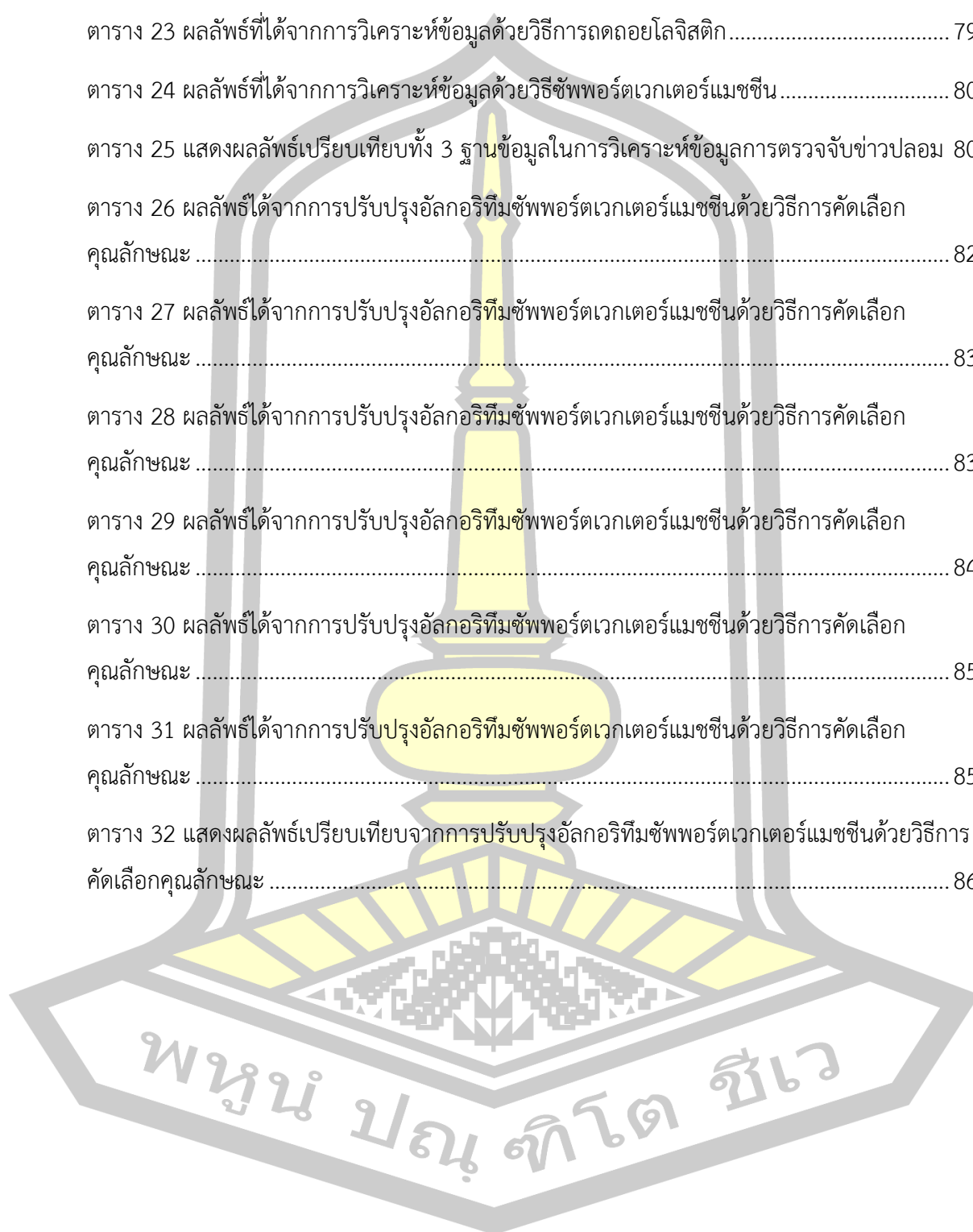
3.2 ชุดข้อมูล	55
3.3 การประมวลผลข้อมูลล่วงหน้า	56
3.4 แบบจำลองพื้นฐานของเทคนิคการเรียนรู้ของเครื่อง	58
บทที่ 4 ผลการวิจัย	64
4.1 การวิเคราะห์การตรวจจับข่าวปลอม	64
4.2 ขั้นตอนการเตรียมข้อมูล	65
4.3 การวัดประสิทธิภาพของอัลกอริทึม	66
4.4 การทดสอบและการเลือกอัลกอริทึม	68
บทที่ 5 สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ	88
5.1 ส่วนสรุปผล	89
5.2 ส่วนอภิปรายผล	91
5.3 ส่วนข้อจำกัดทางการวิจัยและข้อเสนอแนะในการทำวิจัยครั้งต่อไป	93
5.3 ประโยชน์ที่ได้รับ	93
บรรณานุกรม	96
ภาคผนวก ก	107
ภาคผนวก ข	110
บรรณานุกรม	116
ประวัติผู้เขียน	118



สารบัญตาราง

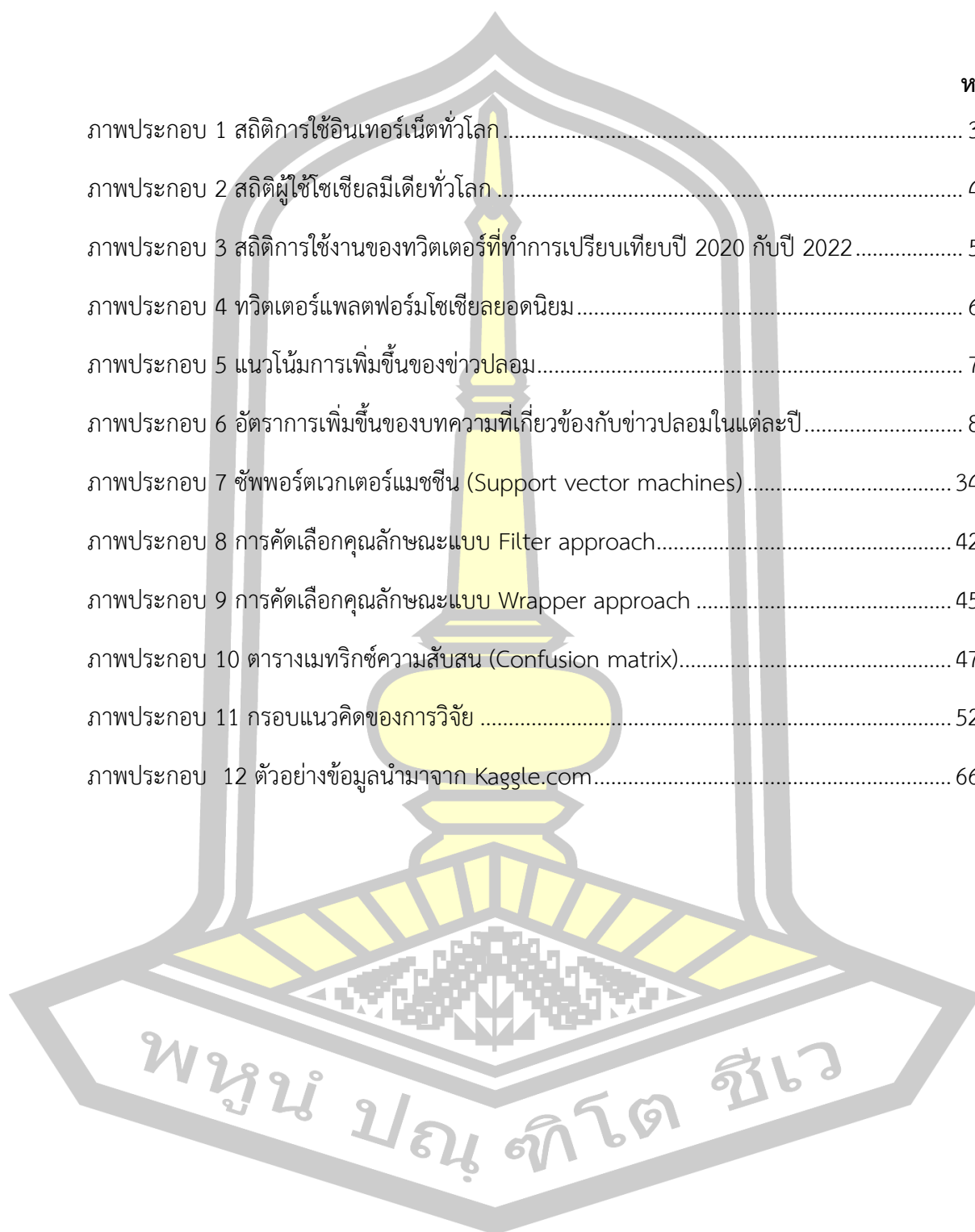
	หน้า
ตาราง 1 สรุปประเด็นที่เกี่ยวข้องกับข่าวปลอม	22
ตาราง 2 ชนิดของสื่อสังคมออนไลน์	28
ตาราง 3 ตัวอย่างงานวิจัยที่นำมาเทคนิคมาเพื่อใช้ในการวิจัยระบุข่าวปลอม	35
ตาราง 4 การเปรียบเทียบวิธีการนำอัลกอริทึมมาใช้ในการตรวจสอบข่าวปลอม	37
ตาราง 5 การเปรียบเทียบการวัดประสิทธิภาพของอัลกอริทึมกับงานวิจัยที่เกี่ยวข้อง	40
ตาราง 6 สถิติของชุดข้อมูลที่พัฒนาแล้ว	56
ตาราง 7 ฐานข้อมูลที่นำมาใช้ในการพยากรณ์ตรวจจับข่าวปลอม	65
ตาราง 8 ตารางเมทริกซ์ความสับสน (Confusion matrix)	66
ตาราง 9 ผลลัพธ์ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอ์ฟเบย์	69
ตาราง 10 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ	70
ตาราง 11 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้แห่งการทํานาย	70
ตาราง 12 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก	71
ตาราง 13 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน	72
ตาราง 14 แสดงผลลัพธ์เปรียบเทียบทั้ง 5 อัลกอริทึมในการวิเคราะห์ข้อมูลการตรวจจับข่าวปลอม	73
ตาราง 15 ผลลัพธ์ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอ์ฟเบย์	74
ตาราง 16 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ	74
ตาราง 17 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้แห่งการทํานาย	75
ตาราง 18 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก	76
ตาราง 19 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน	76
ตาราง 20 ผลลัพธ์ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอ์ฟเบย์	77
ตาราง 21 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ	78

ตาราง 22 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้แห่งการทํานาย.....	78
ตาราง 23 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก.....	79
ตาราง 24 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน.....	80
ตาราง 25 แสดงผลลัพธ์เปรียบเทียบทั้ง 3 ฐานข้อมูลในการวิเคราะห์ข้อมูลการตรวจจับชาวปลอม	80
ตาราง 26 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือก คุณลักษณะ	82
ตาราง 27 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือก คุณลักษณะ	83
ตาราง 28 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือก คุณลักษณะ	83
ตาราง 29 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือก คุณลักษณะ	84
ตาราง 30 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือก คุณลักษณะ	85
ตาราง 31 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือก คุณลักษณะ	85
ตาราง 32 แสดงผลลัพธ์เปรียบเทียบจากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการ คัดเลือกคุณลักษณะ	86



สารบัญรูปภาพ

	หน้า
ภาพประกอบ 1 สถิติการใช้อินเทอร์เน็ตทั่วโลก.....	3
ภาพประกอบ 2 สถิติผู้ใช้โซเชียลมีเดียทั่วโลก.....	4
ภาพประกอบ 3 สถิติการใช้งานของทวีตเตอร์ที่ทำการเปรียบเทียบปี 2020 กับปี 2022.....	5
ภาพประกอบ 4 ทวิตเตอร์แพลตฟอร์มโซเชียลยอดนิยม.....	6
ภาพประกอบ 5 แนวโน้มการเพิ่มขึ้นของชาวปาลอม.....	7
ภาพประกอบ 6 อัตราการเพิ่มขึ้นของบทความที่เกี่ยวข้องกับชาวปาลอมในแต่ละปี.....	8
ภาพประกอบ 7 ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines).....	34
ภาพประกอบ 8 การคัดเลือกคุณลักษณะแบบ Filter approach.....	42
ภาพประกอบ 9 การคัดเลือกคุณลักษณะแบบ Wrapper approach.....	45
ภาพประกอบ 10 ตารางเมทริกซ์ความสับสน (Confusion matrix).....	47
ภาพประกอบ 11 กรอบแนวคิดของการวิจัย.....	52
ภาพประกอบ 12 ตัวอย่างข้อมูลนำมาจาก Kaggle.com.....	66



บทที่ 1

บทนำ

1.1 ภูมิหลัง

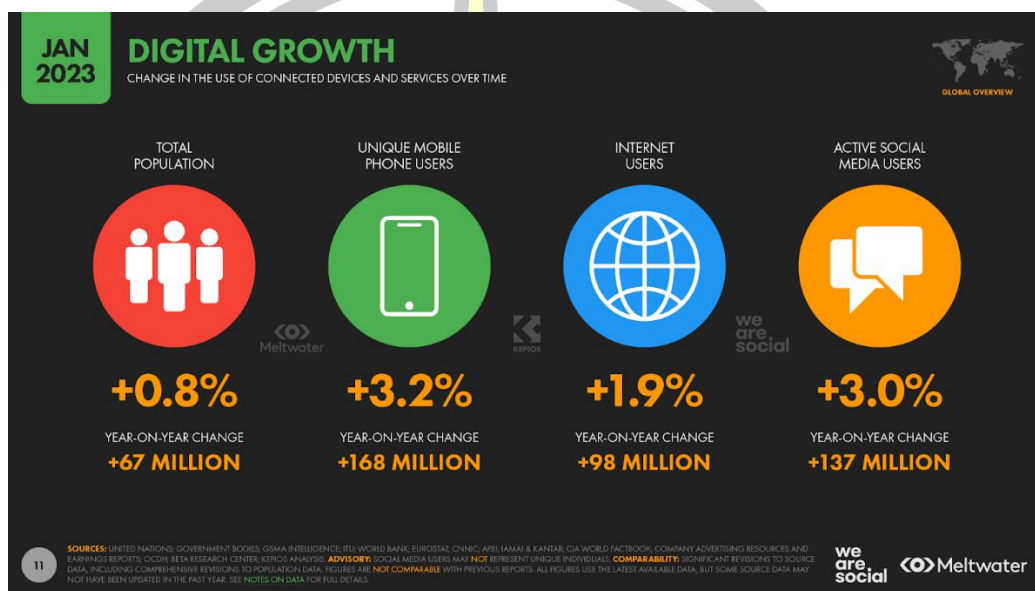
ปัจจุบันข่าวปลอมถูกมองว่าเป็นหนึ่งในภัยคุกคามที่สำคัญต่อประเทศชาติ ประชาธิปไตย และสื่อสารมวลชนมีเหตุการณ์มากมายที่บันทึกไว้ในปี 2565 ที่เผยแพร่ข่าวปลอมผ่านสื่อที่มีชื่อเสียง และแพลตฟอร์มเว็บในระหว่างการเลือกตั้ง ซึ่งข่าวปลอมนั้นจะมาจากปฏิกิริยาความคิดเห็น และการแชร์ (Zhou et al., 2019),(Zabludovsky, 2023) ข่าวปลอมเป็นข้อมูลที่ถูกดัดแปลงให้ดูเหมือนเนื้อหาของสื่อข่าวโดยธรรมชาติ แต่ไม่อยู่ในโครงสร้างการจัดการ หรือเจตนา มีการเผยแพร่อย่างต่อเนื่องผ่านโซเชียลมีเดีย หนังสือพิมพ์ บล็อกออนไลน์ ฟอรัม และนิตยสาร ทำให้ยากที่จะระบุแหล่งข่าวที่เชื่อถือได้ ข่าวปลอมที่แพร่ระบาดอย่างต่อเนื่อง (Lazer et al., 2018) การแพร่กระจายข่าวปลอมนั้นอันตรายพอ ๆ กับการแพร่กระจายไวรัส ผู้คนกำลังเผชิญกับข่าวปลอมทุกวัน ข้อมูลของข่าวปลอมนี้ทำให้เกิดความกลัวและความตื่นตระหนกในหมู่ผู้คน นอกจากนี้ข่าวปลอมที่ทำการเผยแพร่ออกไปยังทำให้เศรษฐกิจอ่อนแอ และที่สำคัญการแพร่กระจายของข่าวปลอมนั้นยังส่งผลให้เกิดภาวะเครียด และสุขภาพจิตเสื่อมโทรม ข่าวปลอมสามารถแพร่กระจายในรูปแบบของข้อความ วิดีโอ รูปภาพและเสียงผ่านทางเครือข่ายโซเชียลมีเดีย เช่น เฟซบุ๊ก และ ทวิตเตอร์ ปัญหาข่าวปลอมมีมาช้านาน คนเคยเชื่อข่าวดังกล่าวแม้ว่าจะเป็นเท็จก็ตาม

ข่าวปลอมจึงภัยคุกคามที่ยิ่งใหญ่ที่สุดต่อมนุษยชาติ ซึ่งส่งผลกระทบต่อทุกด้านของสังคม มนุษย์ มุ่งเน้นไปที่ปัญหาของสังคม ส่งผลให้ประชาชนที่รับรู้ข่าวปลอมเกิดความเสียหาย และได้รับผลกระทบจากมูลเหตุอันเป็นเท็จ ข่าวปลอมมุ่งเน้นไปที่การจำแนกความน่าเชื่อถือของโพสต์ทวิต แต่ผู้ใช้เครือข่ายสังคมออนไลน์ก็ตระหนักดีถึงประเด็นนี้มากกว่า ผู้ใช้เครือข่ายสังคมออนไลน์จึงมีความพยายามในการควบคุมข่าวปลอม เพื่อไม่ให้แพร่กระจายไปในวงกว้าง ทั้งนี้ได้มีกฎหมายควบคุมการเผยแพร่ข้อมูลอันเป็นเท็จเพื่อตรวจสอบข้อเท็จจริงก่อนที่จะมีการเผยแพร่กันไปในวงกว้าง หากพิจารณาผู้ใช้เครือข่ายสังคมออนไลน์ต้องพิจารณาว่าข่าวที่ได้รับบนโซเชียลมีเดียมีความน่าเชื่อถือได้หรือไม่ และข่าวนั้นมีความสำคัญอย่างไร เช่นเดียวกับการเลือกปฏิบัติผู้ใช้ที่ไม่น่าเชื่อถือได้ จากผู้ที่เผยแพร่เนื้อหาข่าวปลอม หรือข่าวไม่ได้รับการยืนยันที่ชัดเจน ทั้งนี้ข่าวปลอมที่เผยแพร่ออกไปบนสื่อสังคมออนไลน์นั้น ผู้ใช้จะมองข้ามถึงพฤติกรรมในการรับรู้ และเชื่อข่าวโดยไม่มีการพิจารณาหรือความรู้ที่เท่าทันข่าว (Atodiresei et al., 2018a),(Mladenova & Valova, 2022), (Sansonet et al., 2021),(Jose et al., 2021), (Swapna & Soniya, 2022), (Bodaghi & Oliveira, 2022)

จากปัญหาของข่าวปลอมที่กล่าวถึงข้างต้นจะเห็นว่าข่าวปลอมสามารถส่งผลกระทบหรือส่งผลกระทบต่อสังคมโดยรวม จึงเป็นภัยอย่างหนึ่งที่ร้ายแรงโดยส่งผล และมีความเกี่ยวเนื่องในเรื่องต่างๆ ตามมา เช่น ความมั่นคงและความปลอดภัยทางด้านการเมือง ความปลอดภัยและความมั่นคงของชาติ ภัยพิบัติต่าง ๆ ที่ผู้แพร่กระจายข่าวปลอมอาจเป็นเจตนาร้ายจงใจเผยแพร่เพื่อสร้างสถานการณ์ เพื่อให้สังคมเกิดความวุ่นวาย เป็นต้น ข่าวปลอมจึงมีผลกระทบอย่างมากต่อสังคมยุคปัจจุบัน ทำให้เกิดการสื่อสารข้อมูลต่าง ๆ ที่สามารถทำได้อย่างรวดเร็ว ผ่านระบบเครือข่ายอินเทอร์เน็ต การส่งเนื้อหาข่าวออกไปนั้น สามารถส่งได้ทั้งข้อความ ภาพหรือเสียงได้ในเวลาเดียวกัน อีกทั้งยังสามารถกระจายข่าวสารไปถึงผู้รับปริมาณได้ครั้งละมาก ๆ ในคราวเดียวกัน เกิดการเปลี่ยนแปลงข้อมูลให้อยู่ในรูปอิเล็กทรอนิกส์ทำให้สามารถส่งต่อได้อย่างง่ายดาย ทำให้เนื้อหา ข่าวสารสามารถแพร่กระจายออกไปได้อย่างรวดเร็ว โดยที่ผู้ส่งไม่อาจทราบเลยว่าข่าวสารที่ตนเองได้ส่งออกไปถูกรับไปโดยผู้ใด ในมุมมองอีกด้านหนึ่งทางฝั่งผู้รับข่าวสารเอง เมื่อได้รับข่าวสารใด ๆ จะทราบได้อย่างไรว่า เนื้อหาข้อมูลข่าวสารที่ได้รับมาเป็นความจริง หรือเป็นข้อมูลที่ผ่านการเปลี่ยนแปลงแก้ไขที่อาจถูกบิดเบือนข้อเท็จจริงไปก่อนหน้าหรือไม่ การระบุข่าวปลอมจึงเป็นขั้นตอนสำคัญในประเทศไทย และได้มีนักวิจัยหลายท่านได้ทำอัลกอริทึมหรือวิธีการที่มีประสิทธิภาพ สามารถนำมาใช้ในการตรวจสอบข่าวปลอม อัลกอริทึมที่ใช้ เช่น naïve Bayes (NB), logistic regression (LR), Random forest (RF), Support vector machine (SVM) (Nomesb & Saravanan, 2022b), (Keya et al., 2021), (Ahmad et al., 2020), (Sansonetti G, Gasparetti F, D'Aniello G. et al., 2021), (Monti F, Frasca F, Eynard D et al., 2019) เป็นต้น

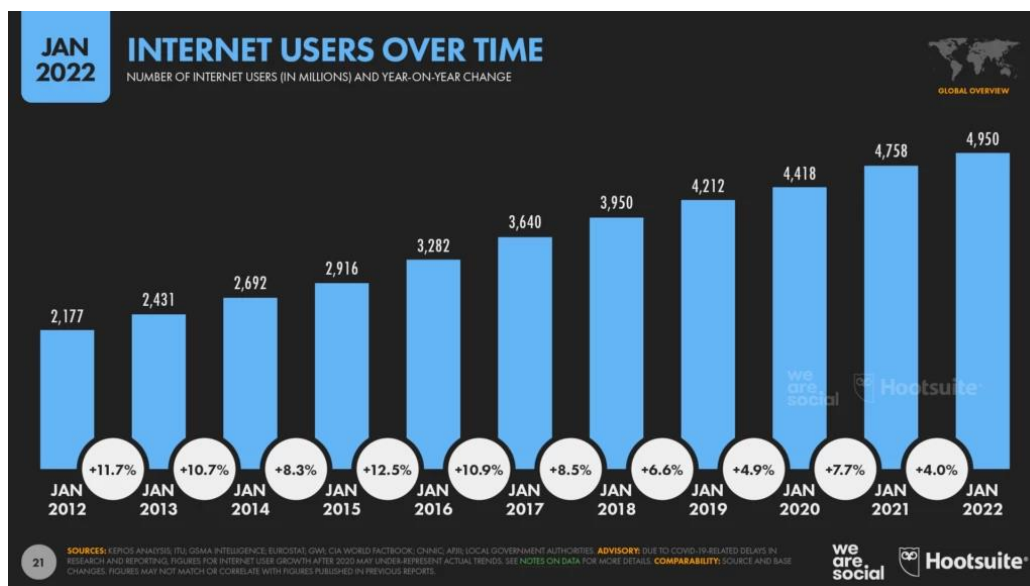
ในยุคของเทคโนโลยีดิจิทัลและเครือข่ายสังคมออนไลน์ มีความก้าวหน้าเป็นแรงขับเคลื่อนให้เกิดกระแสการเปลี่ยนแปลงทางสังคมในด้านต่าง ๆ ที่เห็นได้อย่างชัดเจน ได้เข้ามามีบทบาทและมีการแทรกซึม ไม่ว่าจะเป็นการเสริมสร้างข้อมูล การเดินทาง ความบันเทิง และการแบ่งปันข้อมูล (Vasist & Sebastian, 2022), (Weidinger et al., 2021), (Zhang et al., 2023) เครือข่ายสังคมออนไลน์เปรียบเสมือนประตูไร้พรมแดนเพื่อเชื่อมโยงผู้คนที่มีความสนใจ ชื่นชอบในสิ่งที่คล้ายกันเข้าร่วมเป็นกลุ่มสังคมออนไลน์ เพื่อติดต่อสื่อสาร แลกเปลี่ยนข้อมูล และสร้างปฏิสัมพันธ์ระหว่างกัน ซึ่งใช้แพลตฟอร์มโซเชียลมีเดียที่มีการเพิ่มขึ้น และการแพร่กระจายอย่างรวดเร็ว เช่น เฟซบุ๊ก, วอทแอป, ทวิตเตอร์, และเทเลแกรม เป็นต้น ทำให้พฤติกรรมการใช้งานบนโลกออนไลน์ของคนยุคดิจิทัลมีการปรับเปลี่ยนไป จนกลายเป็นส่วนหนึ่งในชีวิตประจำวันของคนทั่วโลก (Rohera et al., 2022), (Nomesb & Saravanan, 2022a), (Victor & Raza, 2022) จากการรายงานวีอาร์โซเชียล ร่วมมือกับฮูทสวีท พบว่า จำนวนประชากรโลกอยู่ที่ 8.01 พันล้านคน มีผู้ใช้งานอินเทอร์เน็ตผ่านโทรศัพท์มือถือ 5.44 พันล้านคน คิดเป็น 68% โดยเพิ่มขึ้น 3.2% จากปีที่ผ่านมา จำนวนผู้ใช้งานอินเทอร์เน็ตปัจจุบันอยู่ที่ 5.16 พันล้านคน คิดเป็น 64.4% ซึ่งเพิ่มขึ้นจากปีที่แล้ว 1.9% ผู้ใช้งาน

โซเชียลมีเดียทั่วโลกมี 4.76 พันล้านคน คิดเป็น 59.4% ของประชากรโลก เป็นการเพิ่มขึ้นอย่างช้า ๆ โดยในปีนี้มีผู้ใช้งานโซเชียลมีเดียเพิ่มขึ้นเพียง 137 ล้านคนเท่านั้นซึ่งข้อมูลดังที่กล่าวมาแล้วข้างต้นสามารถแสดงได้ดังภาพประกอบ 1



ภาพประกอบ 1 สถิติการใช้อินเทอร์เน็ตทั่วโลก
ที่มา : The growth master (2023 : Online)

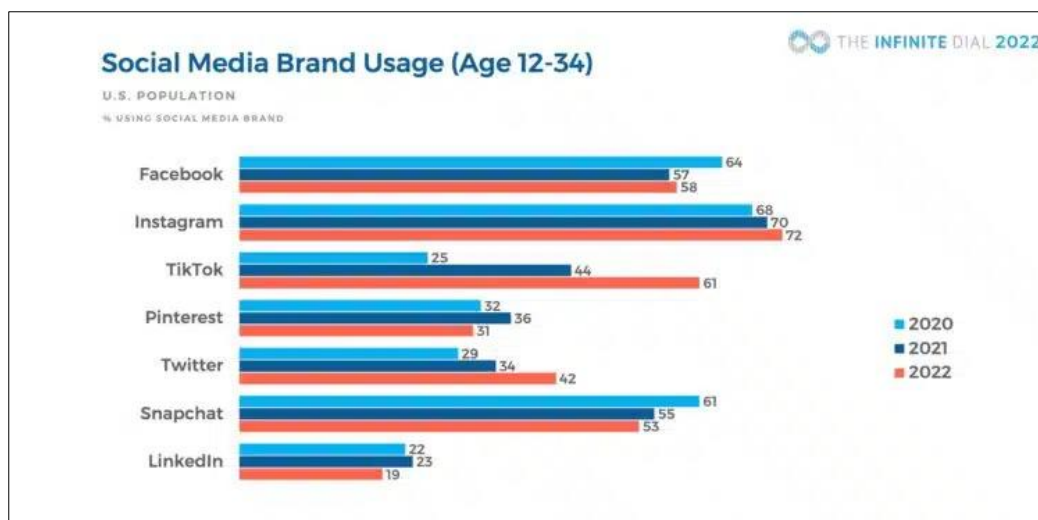
มีผู้ใช้อินเทอร์เน็ตทั่วโลกเพิ่มขึ้นเป็น 4.95 พันล้านคนเมื่อต้นปี 2565 โดยขณะนี้มีการเข้าถึงอินเทอร์เน็ตอยู่ที่ 62.5% ของประชากรทั้งหมดทั่วโลก ข้อมูลแสดงให้เห็นว่าผู้ใช้โซเชียลมีเดียเติบโตขึ้น 192 ล้านคน (+4.0%) ในปีที่ผ่านมา แต่ข้อจำกัดอย่างต่อเนื่องในการวิจัยและการรายงานเนื่องจาก COVID-19 หมายความว่าแนวโน้มการเติบโตที่แท้จริงอาจสูงกว่าตัวเลขนี้มาก และมีผู้ใช้โซเชียลมีเดียทั่วโลก 4.62 พันล้านคนทั่วโลกในเดือนมกราคม 2565 คิดเป็นร้อยละ 58.4% ของประชากรทั้งหมดของโลก โดยมีผู้ใช้โซเชียลมีเดียทั่วโลกเติบโตขึ้นมากกว่า 10% ในช่วง 12 เดือนที่ผ่านมา โดยมีผู้ใช้ใหม่ 424 ล้านคนเริ่มต้นเส้นทางโซเชียลมีเดียในช่วงปี 2564 ข้อมูลดังกล่าวสามารถแสดงได้ดังภาพประกอบ 2



ภาพประกอบ 2 สถิติผู้ใช้โซเชียลมีเดียทั่วโลก
ที่มา : กรุงเทพธุรกิจ (2023 : ออนไลน์)

ช่วงหลายปีที่ผ่านมาแพลตฟอร์มโซเชียลมีเดียได้เปลี่ยนแปลงวิถีของผู้คนมากมาย ทั้งนี้ในไตรมาสแรกของทวิตเตอร์ได้รายงานผู้ใช้งานทวิตเตอร์ ซึ่งปัจจุบัน 42% ของชาวอเมริกันอายุ 12-34 ปีใช้ทวิตเตอร์เทียบกับเพียง 29% ในปี 2020 ซึ่งเพิ่มขึ้น 36.6% ในช่วง 2 ปี เมื่อเปรียบเทียบกับบน อินสตาแกรมเพิ่มขึ้นเพียง 5.7% ในช่วงเวลาเดียวกัน แต่ในขณะเดียวกัน ถ้าดูจากภาพประกอบที่ 2 จะเห็นว่าความสนใจของ Gen Z ส่วนใหญ่นิยม ตึกตอกและอินสตาแกรมมากที่สุด (Pizarro, 2020),(Rainmaker, 2023 : Online), (Edison Research, 2023 : Online) ข้อมูลดังกล่าวสามารถแสดงได้ดังภาพประกอบ 3

พูน ปณ จิต ชีเว

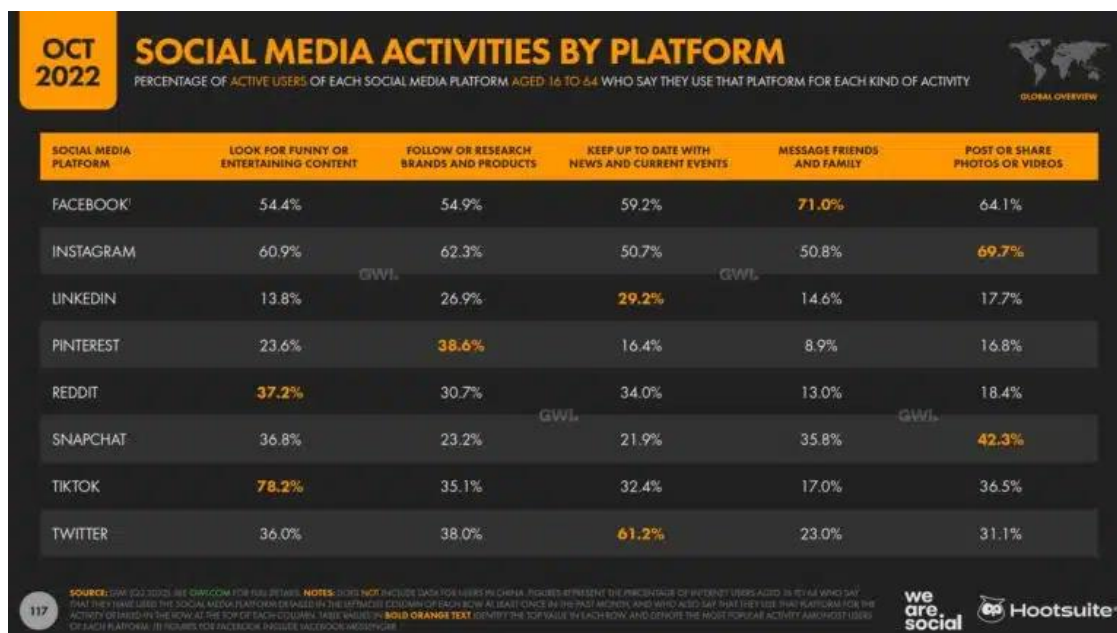


ภาพประกอบ 3 สถิติการใช้งานของทวิตเตอร์ที่ทำการเปรียบเทียบปี 2020 กับปี 2022

ที่มา : Edison Research (2023 : Online)

สัดส่วนผู้ใช้งานทวิตเตอร์ที่เป็นผู้ใช้ช่วงอายุ 18-34 ปี ชอบทวิตเตอร์มากที่สุดในบรรดาช่วงอายุทั้งหมด คิดเป็น 41% ที่ชอบทวิตเตอร์มากที่สุด โดยมองว่าทวิตเตอร์ ‘ดีมาก’ คิดเป็น 16% และทวิตเตอร์ ‘ค่อนข้างดี’ คิดเป็น 25% ซึ่งตามช่วงอายุนี้มันก็นับเป็น Gen Z ไม่ได้ ส่วนกลุ่มอายุที่มียอดรวมสูงเป็นอันดับ 2 คือผู้ใช้อายุ 35-44 ปี รวมกันทั้งหมด 38% ที่ชอบทวิตเตอร์ มากที่สุด และทวิตเตอร์เป็นแพลตฟอร์มโซเชียล ยอดนิยมสำหรับข่าวสารและเหตุการณ์ปัจจุบัน ซึ่งจะเห็นว่าการพูดคุยบนทวิตเตอร์มักจะเป็นข้อความสั้น ๆ แบบเรียลไทม์ และยังคงเป็นอยู่ในปัจจุบัน โดย 61.2% ของผู้คนกล่าวว่า Twitter คือที่ที่พวกเขาไปเพื่อติดตามข่าวสารและกิจกรรมล่าสุด แปรนด์สามารถใช้ประโยชน์จากสิ่งนี้ได้โดยการทวิตสดจากการตั้งประเด็นและสร้างแฮชแท็กกิจกรรมเพื่อรีทวีตจากและเริ่มการสนทนากับผู้เข้าร่วม ข้อมูลดังกล่าวสามารถแสดงได้ดังภาพประกอบ 4

พูน ปณ ทิโต ชเว



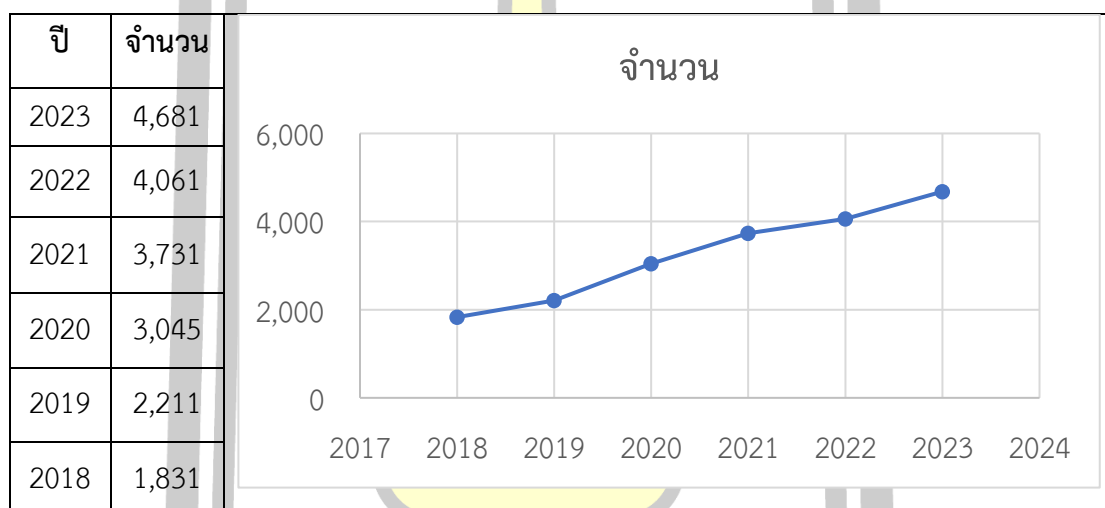
ภาพประกอบ 4 ทวิตเตอร์แพลตฟอร์มโซเชียลยอดนิยม

ที่มา : Edison Research (2023 : Online)

ทวิตเตอร์ (Twitter) เป็นแพลตฟอร์มโซเชียลมีเดียสำหรับการสื่อสารออนไลน์ที่ได้รับความนิยมมากที่สุด โดยมีผู้ใช้งานจำนวนมากที่ใช้ข้อมูลที่เผยแพร่ผ่านแพลตฟอร์มทวิตเตอร์ทุกวัน โดยในช่วงแรกเปิดให้มีการใช้งานที่ผู้ใช้สามารถส่งเฉพาะข้อความที่เป็นตัวอักษรไม่เกิน 140 ตัวอักษร ต่อมาได้มีการปรับความยาวในการส่งข้อความเป็นไม่เกิน 280 ตัวอักษร ลักษณะคล้ายกับการส่งข้อความผ่านโทรศัพท์เคลื่อนที่หรือโปรแกรมสนทนาออนไลน์ ทำให้ผู้ใช้งานคุ้นชินกับการใช้งานทวิตเตอร์เพื่อส่งข่าวสารสั้น ๆ กระชับได้ใจความ ทวิตเตอร์เป็นเครือข่ายสังคมออนไลน์ที่ได้รับความนิยมมากขึ้นเรื่อย ๆ จนถึงปัจจุบันมีผู้ใช้งานจำนวนมาก และในช่วงหลายปีที่ผ่านมาผู้ใช้งานทวิตเตอร์ 353 ล้านคนที่ใช้งานในทวิตเตอร์ต่อเดือน มีการใช้ข้อมูลที่เผยแพร่ผ่านแพลตฟอร์มเครือข่ายสังคมออนไลน์ทวิตเตอร์เป็นประจำทุกวัน การส่งข้อมูลผ่านทวิตเตอร์ยังเป็นการส่งข้อมูลที่มีประสิทธิภาพ ถึงแม้ว่าผู้ใช้จะได้รับอนุญาตให้ส่งข้อมูลในจำนวนอักขระที่จำกัด แต่ผู้ใช้มีอิสระในการเผยแพร่เนื้อหาเป็นอะไรก็ได้ตามแต่ความต้องการ (Castillo et al., 2013), (Karami et al., 2020), (Nugroho et al., 2020), (Luo et al., 2021)

จากที่กล่าวมา จะเห็นได้ว่าแพลตฟอร์มโซเชียลมีเดีย มีอิทธิพลต่อไลฟ์สไตล์และส่งผลกระทบต่อทัศนคติของผู้คน จนกลายเป็นเครื่องมือที่มีอิทธิพลโดยตรงต่อการรับรู้ถึงเหตุการณ์ ผู้คน หรือผลิตภัณฑ์ บุคคล และองค์กรบางแห่ง โดยมีการใช้ทวิตเตอร์สร้างสถานการณ์ด้วยการแพร่กระจาย

ข่าวปลอม ข้อมูล ข่าวลือ และข้อมูลอันเป็นเท็จที่เป็นสาเหตุของความกังวลเพื่อหวังผลประโยชน์ต่อตนเองในทางที่ผิด เพื่อดำเนินการหลายอย่างที่ไม่พึงประสงค์ และที่มีความประสงค์ร้าย เช่น สร้างสแปม ข่าวปลอม ข้อความปลอม และบัญชีปลอม ที่สามารถแพร่กระจายข้อมูล ข่าวสารที่เป็นอันตราย และส่งผลกระทบต่อความสงบสุขของสังคม สร้างความวุ่นวาย ความปลอดภัย หรือทำให้ความมั่นคงของมาตุภูมิสั่นคลอน การแพร่กระจายข่าวลือหรือข่าวปลอมต่าง ๆ ที่เกิดขึ้นนั้น ไม่เพียงแต่ส่งผลกระทบในทางลบ แต่ยังรวมถึงชีวิตของผู้คนที่ได้รับผลกระทบจากข่าวปลอมอีกด้วย ทั้งนี้ข่าวปลอมมีการเพิ่มขึ้นทุก ๆ ปี และในแต่ละปีข่าวปลอมมีแนวโน้มเพิ่มขึ้นเรื่อย ๆ (Shanathi, 2022), (Alrubaian et al., 2019) จากข้อมูลที่กล่าวมาสามารถแสดงได้ดังภาพประกอบ 5



ภาพประกอบ 5 แนวโน้มการเพิ่มขึ้นของข่าวปลอม
ที่มา : ScienceDirect (2024 : Online)

จากภาพประกอบที่ 5 จะพบว่าจากปี 2019 จนถึงปี 2024 มีนักวิจัยหลายท่าน ได้วิจัยเกี่ยวกับการตรวจสอบข่าวปลอม ทั้งนี้จะเห็นว่าข่าวปลอมมีการวิจัยหรือศึกษามานานแล้ว แต่ในปัจจุบันที่อยู่ในยุคของแพลตฟอร์มโซเชียลมีเดียที่เป็นแหล่งกระจายข้อมูลที่สำคัญ ดังนั้นจึงมีนักวิจัยได้ศึกษาเกี่ยวกับเทคนิคหรือวิธีการตรวจสอบข่าวปลอมเพิ่มขึ้นเรื่อย ๆ และผู้วิจัยได้ทำการเปรียบเทียบอัตราการเพิ่มขึ้นของบทความที่เกี่ยวข้องกับข่าวปลอมในแต่ละปี ดังที่กล่าวมาแล้วข้างต้นสามารถแสดงได้ดังภาพประกอบ 6



ภาพประกอบ 6 อัตราการเพิ่มขึ้นของบทความที่เกี่ยวข้องกับข่าวปลอมในแต่ละปี

ที่มา : ScienceDirect (2024 : Online)

จากภาพประกอบ 6 จะเห็นว่าอัตราการเพิ่มขึ้นของบทความวิจัยที่เกี่ยวข้องกับข่าวปลอมนั้นมีนักวิจัยได้ให้ความสำคัญเกี่ยวกับการศึกษาวิธี หรือเทคนิคที่จะนำมาใช้สำหรับการตรวจสอบข่าวปลอมที่มีการเพิ่มขึ้นทุกปี

จากงานวิจัยจะเห็นว่า การแพร่กระจายข่าวปลอมเกิดขึ้นอย่างรวดเร็วในยุคปัจจุบัน จากการตรวจสอบการแพร่กระจายข่าวปลอมที่ได้รับการยืนยันทั้งหมดที่เผยแพร่บนทวิตเตอร์ ตั้งแต่ปี 2549 ถึง 2560 ข้อมูลประกอบด้วย 126,000 เรื่องที่ทวีตโดย 3 ล้านคนมากกว่า 4.5 ล้านครั้ง การจัดประเภทข่าวว่าจริงหรือเท็จโดยใช้ข้อมูลจากองค์กรตรวจสอบข้อเท็จจริง หกแห่งที่แสดงข้อตกลง 95 ถึง 98% เกี่ยวกับการจำแนกประเภท ข่าวปลอมแพร่กระจายไปไกลกว่า เร็วกว่า ลึกกว่า และกว้างกว่าความจริง ในข้อมูลทุกประเภทอย่างมีนัยสำคัญ และผลกระทบที่เด่นชัดกว่าสำหรับข่าวปลอม คือ ข่าวทางการเมืองมีข้อเท็จจริงมากกว่าข่าวเกี่ยวกับการก่อการร้าย ภัยธรรมชาติ วิทยาศาสตร์ ด่านเมือง หรือข้อมูลทางการเงิน (Vosoughi et al., 2018), (Bazmi et al., 2023), (Allein et al., 2023), (Yu et al., 2022) อีกทั้งยังพบว่าข่าวปลอมเป็นเรื่องแปลกใหม่มากกว่าข่าวจริง ซึ่งแสดงให้เห็นว่าผู้คนมีแนวโน้มที่จะแบ่งปันข้อมูลข่าวปลอมมากกว่า และในขณะเดียวกันข่าวปลอมเป็นแรงบันดาลใจให้เกิดความกลัว ความรังเกียจ ความคาดหวัง ความเศร้า ความสุข ความไม่หวังดี ประสงค์ร้ายและความไว้วางใจ เป็นต้น แพลตฟอร์มโซเชียลมีเดียเร่งการแพร่กระจายของข่าวปลอมและข่าวจริงในอัตราที่แตกต่างกัน ซึ่งหมายความว่าข่าวปลอมมีแนวโน้มที่จะแพร่กระจายมากกว่าความจริง

และได้มีนักวิจัยหลาย ๆ ท่าน ได้ศึกษาค้นคว้าที่เกี่ยวข้องกับการตรวจสอบข่าวปลอม โดยงานวิจัยจะมีลักษณะของการนำเครื่องมือมาใช้ในการตรวจสอบข่าวปลอม โดยเฉพาะซอฟต์แวร์คอมพิวเตอร์แมชชีน เป็นเครื่องมือที่อยู่ในส่วนหนึ่งของวิธีการเรียนรู้ด้วยเครื่องที่ได้รับความนิยมที่ให้ความแม่นยำสูง สำหรับการตรวจสอบข่าวปลอม เป็นวิธีการเรียนรู้แบบมีผู้สอนที่ใช้สำหรับการวิเคราะห์การจำแนกประเภทข้อมูล สามารถนำมาใช้ในการแก้ปัญหาการจำแนกข้อมูลใช้ในการวิเคราะห์ข้อมูล และจำแนกข้อมูล และคุณสมบัติอีกประการหนึ่งคือยังเป็นเทคนิคหนึ่งที่ได้รับการนิยมนำมาใช้แพร่หลายในงานวิจัยที่เกี่ยวข้องกับการจัดจํารูปแบบตลอดจนการแก้ปัญหาการจัดกลุ่ม โดยอาศัยหลักการของการหาสัมประสิทธิ์สมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกต้องเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกกลุ่มข้อมูลได้ดีที่สุด (Zubiaga et al., 2018), (Nagaraja et al., 2021), (Ehsanfar & Mansouri, 2017), (Praseed et al., 2023)

ดังที่กล่าวมาข้างต้น จะเห็นว่าการพัฒนาอย่างรวดเร็วและแพร่หลายของเทคโนโลยีอินเทอร์เน็ต การแพร่กระจายและการแพร่กระจายของข้อมูลกลายเป็นเรื่องง่ายและรวดเร็วยิ่งขึ้น ในขณะที่ทำให้ชีวิตสะดวกสบายยิ่งขึ้น อินเทอร์เน็ตยังส่งเสริมการแพร่กระจายข่าวปลอมในวงกว้าง ซึ่งจะส่งผลกระทบอย่างมากต่อประเทศ สังคม บุคคล และยังส่งผลกระทบต่อทุกๆ ด้านของสังคมมนุษย์ ดังนั้นจึงมีความพยายามวิจัยมากมายเพื่อต่อสู้กับข่าวปลอม โดยทั่วไปการตรวจจับข่าวปลอมเป็นปัญหาการจำแนกประเภทที่มุ่งตรวจสอบความถูกต้องของเนื้อหาข่าว ซึ่งอาจรวมถึงข้อความ รูปภาพ และวิดีโอ บทความนี้ ผู้วิจัยจึงมีแนวคิดที่จะเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม พร้อมทั้งนำอัลกอริทึมไปประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย และปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม

1.2 วัตถุประสงค์ของการวิจัย

- 1.2.1 เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม
- 1.2.2 เพื่อนำอัลกอริทึมไปประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย
- 1.2.3 เพื่อปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม

1.3 ขอบเขตของการวิจัย

การวิจัย เรื่อง การปรับปรุงการตรวจจับข่าวปลอมด้วยซอฟต์แวร์คอมพิวเตอร์แมชชีนโดยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม ผู้วิจัยได้ใช้กรอบแนวคิดในการวิจัย ดังนี้

- 1.3.1 ฐานข้อมูล (Data base) งานวิจัยนี้ใช้ฐานข้อมูลที่มีการเก็บรวบรวมข้อมูลเกี่ยวกับชุดข้อมูลของข่าวปลอม ซึ่งมีการประยุกต์จากแนวคิดของ (Kaggle, Online, Data word and UCI)

(Sansonetti et al., 2021), (Atodiresei et al., 2018b) ประกอบด้วย จำนวน 20,717 แถว 5 แอททริบิวต์ มีเนื้อหาดังนี้

- 1) หัวเรื่อง : ชื่อบทความ.
- 2) news_url : URL ของบทความ
- 3) โดเมนต้นทาง : โดเมนเว็บไซต์ที่โพสต์บทความ
- 4) tweet_num : จำนวนการรีทวีตสำหรับบทความนี้
- 5) Real : คอลัมน์ป้ายกำกับโดยที่ 1 เป็นของจริงและ 0 เป็นของปลอม

1.3.2 การเตรียมข้อมูล (Data preprocessing) ประยุกต์จากแนวคิดของ

(Aphiwongsophon, S., & Chongstitvatana, P., 2018) (Atodiresei et al., 2018a), (Buntain & Golbeck, 2017) ประกอบด้วย

1. การทำความสะอาดข้อมูล (Data cleaning) เป็นวิธีการทำความสะอาดข้อมูล หรือ การทำข้อมูลให้สมบูรณ์ เป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล เพราะหมายถึง ความไม่สมบูรณ์ ความไม่ถูกต้อง ความไม่สัมพันธ์กับข้อมูลอื่นๆ เป็นต้น จึงต้องมีการแทนที่ การปรับปรุง หรือการลบข้อมูลที่ไม่ถูกต้องเหล่านั้นออกไป เพื่อให้ข้อมูลมีคุณภาพ มีขั้นตอน

1.1 Tokenize เป็นขั้นตอนการตัดข้อความหรือประโยคออกเป็นคำ (word)

1.2 Normalization เป็นขั้นตอนการปรับรูปแบบของคำให้เหมือนกัน เช่น การแปลงให้เป็นตัวอักษรตัวเล็กทั้งหมด

1.3 Stemming เป็นขั้นตอนการตัดส่วนท้ายของคำที่เพิ่มเข้าไปออก

1.4 Remove Stop Words เป็นขั้นตอนการลบคำที่เป็น stop word เช่น a, and, the ทิ้งไป

1.5 Part Of Speech Tagging เป็นขั้นตอนการหาประเภทของคำ เช่น คำนาม (noun) คำกริยา (verb)

2. การคัดเลือกคุณลักษณะ (Feature selection) เป็นวิธีการลดขนาดหรือมิติของข้อมูลเพื่อคงลักษณะที่สำคัญของข้อมูล และวิเคราะห์ข้อมูลที่มีมิติจำนวนมาก ที่มีความซับซ้อนสูง ซึ่งขั้นตอนวิธีสำหรับการเรียนรู้ที่จะนำมาใช้ในการจัดการกับข้อมูลเหล่านี้ จะต้องใช้ทรัพยากรทั้งในด้านการคำนวณและการใช้หน่วยความจำจำนวนมาก ในขณะที่ประสิทธิภาพของกระบวนการเรียนรู้อาจลดลงเนื่องจากอาจมีสัญญาณรบกวน (noise) ในข้อมูลที่เกิดจากมิติที่ไม่เกี่ยวข้องกับการวิเคราะห์ข้อมูลนั้น ๆ ดังนั้น เพื่อแก้ปัญหาดังกล่าว มิติของข้อมูลเหล่านั้นจะต้องถูกลดจำนวนลงด้วยวิธีการเลือกคุณลักษณะที่สำคัญสำหรับการวิเคราะห์ข้อมูล (ไกรรุ่ง เสงพระพรหม และคณะ, 2557)

1.3.3 การเรียนรู้ของเครื่อง (Machine learning) ประยุกต์จากแนวคิดของ (Altheneyan & Alhadlaq, 2023), (Aphiwongsophon & Chongstitvatana, 2018), (Joju & Kammath, 2021), (Mishra et al., 2022), (Srinivasa Rao et al., 2021), (Victor & Raza, 2022)

1. นาอิวเบย์ (Naïve bayes) อัลกอริทึมนาอิวเบย์เป็นการเรียนรู้ที่อาศัยหลักการความน่าจะเป็น ตามทฤษฎีของเบย์ (Bayes Theorem) ซึ่งมีอัลกอริทึมที่ไม่ซับซ้อนเป็นขั้นตอนวิธีการจำแนกข้อมูล โดยการเรียนรู้ปัญหาที่เกิดขึ้น เพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ หลักการของนาอิวเบย์ใช้การคำนวณน่าจะเป็นในการทำนายผลเป็นเทคนิคในการแก้ปัญหาแบบจำลองประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ จะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมาก และคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน

2. ต้นไม้ตัดสินใจ (Decision tree) เป็นโมเดลทางคณิตศาสตร์ที่ใช้ในการทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ โดยมีลักษณะการตัดสินใจเป็นแบบแผนผังต้นไม้ ประกอบด้วย ส่วนที่เป็นบัพภายใน (Inner node) หรือที่เรียกว่าโหนดของต้นไม้ใช้แทนตัวแปรต่าง ๆ ที่ใช้ในการตัดสินใจ ส่วนที่กิ่ง (Branch) ใช้แทนค่าที่เป็นได้ของตัวแปร และส่วนที่เป็นบัพใบ (Leaf node) หรือที่เรียกว่าใบโหนด ใช้แทนค่าคำตอบของการตัดสินใจหรือการจำแนก

3. ต้นไม้สุ่ม (Random forest) เป็นหนึ่งในกลุ่มของโมเดลที่เรียกว่า Ensemble learning ที่มีหลักการคือการเทรนโมเดลที่เหมือนกันหลายๆ ครั้ง (หลาย Instance) บนข้อมูลชุดเดียวกัน โดยแต่ละครั้งของการเทรนจะเลือกส่วนของข้อมูลที่เทรนไม่เหมือนกัน แล้วเอาการตัดสินใจของโมเดลเหล่านั้นมาโหวตกันว่า Class ไหนถูกเลือกมากที่สุด

4. การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression) เป็นเทคนิคสถิติที่ใช้ในการพยากรณ์ความน่าจะเป็นที่จะเกิดเหตุการณ์หรือไม่เกิดเหตุการณ์ของเหตุการณ์หนึ่งๆ ที่สนใจ จะใช้เมื่อตัวแปรตาม (Dependent Variable) เป็นตัวแปรเชิงคุณภาพ ส่วนตัวแปรอิสระ (Independent Variable) หรือตัวแปร พยากรณ์ (Predictor) เป็นตัวแปรที่สนใจหรือเป็นปัจจัยที่สามารถพยากรณ์การเกิดเหตุการณ์หนึ่งๆ ที่สนใจ หรือเป็นปัจจัยที่มีอิทธิพลต่อการเกิดเหตุการณ์หนึ่งๆ ที่สนใจได้ ซึ่งเป็นได้ทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพ สามารถมีได้มากกว่า 1 ตัวแปร

5. ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines) เป็นเทคนิคหนึ่งที่ได้รับ ความนิยมอย่างแพร่หลายในงานวิจัยที่เกี่ยวข้องกับการจัดจำรูปแบบตลอดจนการแก้ปัญหาการจัดกลุ่ม โดยอาศัยหลักการของการหาสมประสิทธิผลการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกต้องเข้าสู่กระบวนการสอนให้ระบบเรียนรู้โดยเน้นไปยังเส้นแบ่งแยกกลุ่มข้อมูลได้ดีที่สุด

1.3.4 การวัดประสิทธิภาพ (Evaluation Measure) ในการประเมินโมเดลผู้วิจัยได้ใช้เมตริกการประเมินผลหลายรายการ ซึ่งผู้วิจัยได้นำมาใช้ในการตรวจสอบพารามิเตอร์สำหรับการตรวจจับ

ข่าวปลอม โดยมีการประยุกต์จากแนวคิดของ (Rohera et al., 2022), (Meesad, 2021), (Nagaraja, N, et al., 2021), (Aldwairi & Alwahedi, 2018), (Carlos Pereira-Kohatsu et al., 2019)

1. ค่าความถูกต้อง (Accuracy) คือ จำนวนของข้อมูลที่ทำนายถูกทุกคลาส เป็นการวัดความถูกต้องของโมเดล โดยพิจารณาทุกคลาส
2. ค่าความแม่นยำ (Precision) คือ ค่าที่ดูจากสิ่งทำนายออกมาแล้วทายถูกได้กี่เปอร์เซ็นต์
3. ค่าความระลึก (Recall) คือ ค่าของจำนวนที่ทำนายถูกที่ตัว เป็นการวัดความถูกต้องของโมเดล
4. ค่าความถ่วงดุล (F – measure) คือ ค่าเฉลี่ยของค่าความแม่นยำและค่าระลึก

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1 สามารถทราบถึงการเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม
- 1.4.2 สามารถนำอัลกอริทึมมาประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย
- 1.4.3 สามารถทราบถึงวิธีการปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม
- 1.4.4 สามารถนำผลที่ได้ไปใช้ในการวางแผนการตรวจสอบข่าวปลอมสำหรับองค์กรธุรกิจได้
- 1.4.5 สามารถนำผลที่ได้ไปใช้ในการวางแผนการตรวจสอบข่าวปลอมสำหรับบุคคลประชาชนทั่วไปได้

1.5 นิยามศัพท์เฉพาะ

1.5.1 ข่าวปลอม (Fake news) หมายถึง ข่าวที่นำไปสู่ความเข้าใจผิดและไม่ตั้งอยู่บนพื้นฐานความจริงแตกต่างจากข่าวเสียดสี (satire news) ซึ่งดัดแปลงข้อมูลเพื่อบ่มสร้างอารมณ์ขันให้กับผู้อ่าน ข่าวปลอมจงใจให้สารสนเทศที่ผิดพลาด ไม่ใช่เพื่อความขบขัน แต่หวังผลประโยชน์จากความเข้าใจผิดนั้น ผู้ขาดทักษะในการจำแนกข้อเท็จจริง มีโอกาสถูกยั่วให้หลงเชื่อข่าวปลอมได้ง่าย

1.5.2 สื่อสังคมออนไลน์ หมายถึง เครือข่ายสังคมออนไลน์เป็นรูปแบบของเว็บไซต์ ที่มีเครือข่ายความสัมพันธ์เสมือนที่โยงใยให้บุคคลที่มีเรื่องที่สนใจเหมือนกันมาพูดคุยกันสามารถเชื่อมโยงระหว่างบุคคลอีกบุคคลหนึ่งเป็นการสร้างเครือข่ายสังคม

1.5.3 ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines) หมายถึง อัลกอริทึมที่อยู่ในกลุ่มของวิธีการเรียนรู้ด้วยเครื่องแบบมีผู้สอน ที่นำมาช่วยแก้ปัญหาการจำแนกข้อมูล

1.5.4 การคัดเลือกคุณลักษณะ (Feature selection) หมายถึง การลดขนาด หรือ มิติของข้อมูลและมีการคงลักษณะสำคัญของข้อมูล

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาวิจัยครั้งนี้ ผู้วิจัยได้ทำการศึกษาแนวความคิด ทฤษฎีที่เกี่ยวข้อง และศึกษา งานวิจัยที่เกี่ยวข้อง ก่อนการทำงานวิจัยเพื่อใช้ประกอบการศึกษางานวิจัยเรื่อง การปรับปรุงการ ตรวจจับข่าวปลอมด้วยซอฟต์แวร์แมชชีนโดยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม ซึ่งมี ประเด็นที่เกี่ยวข้องในงานวิจัย ดังนี้

1. ข่าวปลอม (Fake news)
2. สื่อสังคมออนไลน์ (Social media)
3. ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine)
4. การเตรียมข้อมูล (Data preprocessing)
5. การวัดประสิทธิภาพ (Evaluation measure)
6. งานวิจัยที่เกี่ยวข้อง

2.1 ข่าวปลอม (Fake news)

2.1.1 ความหมายของข่าวปลอม (Fake news)

สุปัญญา อภิวงศ์โสภณ (2561) กล่าวว่า ข่าวปลอม หรือ ข่าวลวง เป็นคำบอกเล่า เรื่องราวที่เป็นเรื่องที่เกิดใหม่ หรือเป็นเรื่องที่ถูกทำให้มีความสนใจ อาจเป็นคำบอกกล่าว คำเล่าลือที่ยัง ไม่มีหลักฐานยืนยันว่าเป็นเรื่องจริงที่น่าเชื่อถือ หรืออาจเป็นเหตุการณ์ที่มีผู้ที่เกี่ยวข้องออกมา แถลงการณ์ยืนยันแล้วว่าเนื้อหาข่าวถูกบิดเบือนไป หรือไม่ได้เป็นความจริงตามที่เนื้อข่าวนั้นได้กล่าว อ้าง

รัตนภรณ์ จอมมูล (2020) กล่าวว่า ข่าวปลอมหรือข่าวลวง เป็นลักษณะของข้อมูลที่มี เนื้อหาอันเป็นเท็จ ถูกประดิษฐ์ขึ้นด้วยถ้อยคำให้เป็นเป็นเรื่องราวโดยที่ไม่มีข้อเท็จจริงอยู่ในประโยค นั้น ๆ โดยไม่สามารถตรวจสอบแหล่งที่มาของประโยคหรือคำพูดนั้นได้ ในบางลักษณะของข่าวปลอม อาจมีบางส่วนของข้อมูลข่าวที่สามารถตรวจสอบข้อมูลได้ แต่อาจมีบางข้อมูลที่ถูกสร้างขึ้นไม่สามารถ ตรวจสอบได้

Lazer et al (2018) กล่าวว่า ข่าวลวงเป็นลักษณะของข่าวที่มีข้อความเนื้อหาอันเป็น เท็จ หรืออาจจะเป็นการสร้างสถานการณ์ เพื่อมีเจตนาบิดเบือน รวมถึงการเรียบเรียงข้อมูลข่าวที่ ได้รับการสนับสนุนอย่างปิดบัง แอบแฝงซึ่งนำเสนอในสื่อสังคม แพลตฟอร์มออนไลน์อื่น ๆ

European Commission (2017) กล่าวว่า ข้อมูลหรือข้อความชุดหนึ่งที่สูงขึ้นมาจากมีเจตนาปิดบัง บิดเบือนและมีเจตนาโจมตีบุคคล กลุ่มคนหรือองค์กร และบางครั้งข่าวลวงหมายความถึงการสร้างเขียนข่าวหรือเขียนข่าวในแง่เชิงลบ รวมถึงการโฆษณาชวนเชื่อ และมีรูปแบบการโฆษณาที่ชวนให้ผู้ได้รับข่าวเข้าใจผิด จากทัศนคติของผู้รับสาร รวมถึงการรายงานข่าวในรูปแบบของการเสียดสี และมีการใส่ข้อมูลที่ผิดไป โดยมีเจตนาโจมตีทางโลกออนไลน์

Molek (2561) กล่าวว่า ข่าวปลอมหรือ Fake news คือกระบวนการสร้างข้อมูล ความจริงอีกทางขึ้น หรือสร้างเรื่องราวเป็นเท็จให้เป็นเรื่องราวจนทำให้เกิดความเชื่อใหม่ขึ้นมาในข่าวเหล่านั้น อาจส่วนจริงอยู่บ้างหรือบางทีก็อาจจะไม่มีเลย เพียงแต่เล่นกับจิตวิทยาของมนุษย์ในการสร้างเรื่องราวที่ตรงกับความลำเอียงของกลุ่มเป้าหมายขึ้นมา หรือสร้างเรื่องที่กลุ่มเป้าหมายอยาก ฟังขึ้นมาทำให้เกิดการคล้อยตามได้ง่ายพร้อมใช้กระแสสังคมในการกระพือว่าสิ่งที่ได้เห็น ได้ยิน ได้ฟังเหล่านั้นเป็นเรื่องจริง ทั้งที่อาจจะไม่จริงเลยก็ตาม

Hu et al., (2024) กล่าวว่าข่าวปลอมถูกสร้างขึ้นโดยเจตนาและยืนยันว่าเป็นเท็จ ลักษณะเด่นที่สำคัญของข่าวปลอมคือความถูกต้องและความตั้งใจ ง่ายต่อการแยกแยะข่าวปลอมจากแนวคิดอื่นๆ ที่เกี่ยวข้อง

สุวัฒน์ เอมสัมฤทธิ์ (2561) กล่าวว่า ข่าวปลอมคือข้อความหรือเนื้อหาที่ถูกทำให้เข้าใจว่าเป็นข่าวจริง มี เป้าหมายจงใจให้ผู้ฟังเชื่อ แสดงความคิดเห็นและส่งผลต่อ เพื่อหวังผลประโยชน์จากการส่งต่อแบ่งปัน และถูกแพร่กระจายผ่านสื่อออนไลน์โดยผู้สร้างข่าวปลอมและกลุ่มเป้าหมายที่หลงเชื่อข่าวนั้น

สำนักข่าวเอพี (2562) กล่าวว่า ข่าวปลอม คือข่าวสารเท็จที่ถูกสร้างขึ้นเพื่อความบันเทิงสำหรับการเผยแพร่บนอินเทอร์เน็ต

สรานนท์ อินทนนท์ (2562) กล่าวว่า ข่าวปลอม คือข่าวที่ไม่มูลของความจริงหรืออาจเป็นข่าวที่มีข้อเท็จจริงเพียงบางส่วนและไม่สามารถตรวจสอบแหล่งข่าวได้ รวมถึงข่าวที่เขียนด้วยอคติจงใจให้ร้าย โฆษณาชวนเชื่อ ขึ้นมาความคิดของคนด้วยมุมมองเพียงด้านเดียว

สรุป ข่าวปลอม คือข้อความที่มีการเผยแพร่ออกไปโดยไม่ตั้งใจและโดยเจตนาของการเล่าเรื่องหรือข้อเท็จจริงที่ทำให้เข้าใจผิดหรือผิดพลาด มีการใช้คำจำกัดความที่คล้ายกันหรือหัวข้อเดียวกัน

2.1.2 ประเภทข่าวปลอม

การแบ่งประเภทของข่าวปลอมหรือข่าวลวง มีหลายรูปแบบและใช้ในวัตถุประสงค์หรือบริบทที่ต่างกัน การแบ่งประเภทข่าวปลอมนั้น นำมาเพื่อประโยชน์ในการวิเคราะห์และจัดการปัญหาของข่าวปลอมที่มีลักษณะแตกต่างกัน เช่น

เฉลิมชัย กีกเกียรติกุล (2561) ได้จำแนกประเภทและการหาแนวทางการจัดการข่าวลวงในยุคอินเทอร์เน็ต ดังนี้

1. เนื้อหาที่อยู่ในกรอบจำกัด (Content bubbles) เกิดขึ้นเมื่อบุคคลมีปฏิสัมพันธ์กับแหล่งข่าวเดียวแต่กลไกของอัลกอริทึม (Algorithm) ทำหน้าที่ป้อนข่าวให้ผู้อ่านคล้ายตามมุมมองหรือเข้าใจที่ผู้อ่านชอบหรือสนใจ ซึ่งทำให้เกิดสิ่งที่เรียกว่า อคติเชิงพฤติกรรมหรือเชื่อ ต่อย้ายในสิ่งที่ตนเองเชื่ออยู่ (Confirmation bias) และรวมถึงการใช้อัลกอริทึมสำหรับอัลกอริทึมค้นหาข้อมูลบนอินเทอร์เน็ต (Search engine) ที่อิงการค้นหาในอดีตและเนื้อหาที่มีการเลือกอ่านที่ล่าสุดประกอบทำให้กลุ่มของข้อมูลนำเสนอผู้อ่านแคบจากเฉพาะบางแหล่งข้อมูลเท่านั้น

2. ข่าวลวงที่มีการเผยแพร่อย่างไม่ได้ตั้งใจให้เกิดผลให้เกิดผลร้ายหรือเข้าใจผิด (Misinformation) เป็นการนำเสนอความคิดเห็น แต่อาจผิดพลาดหรือไม่จริง ซึ่งสามารถเพิ่มหรือกระตุ้นให้ข่าวลวงแพร่กระจายและน่าเชื่อถือมากขึ้น

3. ข่าวลวงที่มีการเผยแพร่อย่างจงใจ (Disinformation) เป็นกรณีที่มีการใช้อินเทอร์เน็ตหรือสื่อสังคมแพลตฟอร์มด้วยมีเจตนาในการจัดการข้อมูลเพื่อบิดเบือนหรือสร้างอิทธิพลทางความคิดเห็นของผู้คนทั่วไป

สรานนท์ อินทนนท์ (2563) ข่าวปลอมอยู่ภายในระบบนิเวศขนาดใหญ่ของข้อมูลที่ผิดพลาดและมีการบิดเบือน การทำให้ผู้อ่านเข้าใจผิด อาจเกิดโดยตั้งใจและไม่ตั้งใจก็ได้ แต่การบิดเบือนข้อมูลเป็นเท็จที่สร้างขึ้นโดยเจตนา และแพร่กระจายเพื่อสร้างอิทธิพลต่อความคิดเห็นของสาธารณะชนหรือบิดเบือนความจริง ซึ่งแบ่งประเภทข่าวปลอมดังนี้

1. ข่าวพาดหัว ยั่วให้คลิก หรือคลิกเบต (Click bait) ข่าวที่ใช้ถ้อยคำหรือรูปภาพพาดหัวที่ทำให้ดูชวนสงสัยหรือดึงดูดให้ผู้ใช้งานอินเทอร์เน็ตทั่วไปคลิกเข้าไปอ่าน ผู้สร้างข่าวในลักษณะนี้จะอาศัยประโยชน์จากข้อมูลเล็ก ๆ น้อยๆ พอชวนให้ผู้อ่านสงสัย แต่ไม่พอจะขจัดความสงสัยนั้นจนต้องคลิกเข้าไปดูเนื้อหาอื่น ๆ ทั้งที่เนื้อหาอาจไม่คำนึงถึงคุณภาพ หรือความถูกต้องของข้อมูล แต่การพาดหัวทำให้คนหลงกลคลิกเข้าไปเพื่อเรียกยอดวิว

2. โฆษณาชวนเชื่อ (Propaganda) เป็นการนำเสนอข้อมูลข่าวสารที่มีการชักจูงทัศนคติของผู้รับสารต่อเหตุการณ์หรือมุมมองบางอย่างโดยการนำเสนอการให้เหตุผลเพียงข้างเดียว การโฆษณาชวนเชื่อ มักจะทำซ้ำ และกระจายในสื่อหลายชนิดเพื่อหวังผลให้ผู้รับสารเชื่อและคล้อยตาม

3. ข่าวแฝงการโฆษณา (Sponsored content, Native advertising) รูปแบบการโฆษณาที่เนื้อหาแนบเนียนกับเนื้อหาปกติในเว็บไซต์นั้น ๆ หรือแนบเนียนไปกับสิ่งแวดล้อมของแพลตฟอร์มของสื่อ นั้น ที่เป็นอยู่ พร้อมทำหน้าที่ในเนื้อหาที่คนต้องการรับรู้ หรือรับชม โดยไม่ทราบว่าเป็นการโฆษณาจนกว่าจะได้อ่านหรือดูจนจบ ข่าวแฝงในการโฆษณานี้จะทำการแฝง (Tie – in) เรื่องราวของแบรนด์และสินค้าไม่มากเกินไป ทำให้คนอ่านหรือคนเสพสื่อนั้นอาจจะเป็นการผลิตโดยผู้ลงโฆษณา

4. ข่าวล้อเลียนและเสียดสี (Satire and hoax) ข่าวที่ดัดแปลงข้อมูลเพื่อสร้างอารมณ์ขันให้กับผู้อ่านใช้เนื้อที่ตลกขบขัน เพื่อแสดงความคิดเห็นเกี่ยวกับเหตุการณ์ข่าวในโลกแห่งความเป็นจริงผ่านการล้อเลียนหรือเสียดสี

5. ข่าวที่ผิดพลาด (Error) บางครั้งแม้แต่ข่าวที่เผยแพร่จากสำนักข่าวออนไลน์ที่เชื่อถือได้ อาจมีความผิดพลาดได้เช่นกัน เช่น การเขียนข้อความที่ผิด ชื่อบุคคลหรือรูปภาพผิดจากเนื้อหาข่าวจริง ๆ ซึ่งทำให้ผู้รับสารเข้าใจไปในทิศทางอื่น หรือไม่เข้าใจในข่าวนั้น

6. ข่าวเอนเอียงเลือกข้าง (Partisan) เป็นลักษณะของข่าวที่มีการบิดเบือน ซึ่งผู้เสนอข่าวสารมักจะเลือกข้าง โดยนำเสนอข่าววิพากษ์วิจารณ์ในทางที่ไม่ดีต่อฝ่ายที่ไม่ใช่ฝ่ายของตนเอง และในขณะอีกฝ่ายเป็นฝ่ายที่ตนเองชื่นชอบ และในปัจจุบันจะเป็นลักษณะของข่าวด้านการเมือง

7. ทฤษฎีสมคบคิด (Conspiracy theory) (พระมหาสาทรและคณะ, 2564) เป็นลักษณะเรื่องเล่า หรือข้อความที่สร้างขึ้นจากความคิดของคน หรือกลุ่มบุคคลมีการนำเอาเหตุการณ์ต่าง ๆ มาปะติดปะต่อเข้าด้วยกัน ทฤษฎีสมคบคิดจะอธิบายถึงเหตุการณ์ที่เกิดขึ้นโดยอาศัยข้อมูลที่ไม่มีความสัมพันธ์กันแล้วนำมาเชื่อมโยงกัน

8. วิทยาศาสตร์ลวงโลก (Pseudoscience) เป็นลักษณะของข้อเขียนที่มีการอ้างว่าเป็นวิทยาศาสตร์และเป็นข้อเท็จจริง แต่แล้วมีความขัดแย้งกัน ที่ทำให้ไม่สามารถใช้ได้กับกระบวนการทางวิทยาศาสตร์ หรือไม่มีหลักฐานหรือความเป็นไปได้มาสนับสนุน และไม่สามารถทำการตรวจสอบหรือขาดฐานความคิดเป็นวิทยาศาสตร์ ในวงการข่าววิทยาศาสตร์ลวงโลกจะมาในรูปแบบของบทความทางการแพทย์หรือบทความทางด้านสุขภาพที่แฝงการโฆษณาขายรักษาหรืออุปกรณ์เพื่อสุขภาพ โดยใช้การแอบอ้างว่าได้ผ่านการวิจัยทางวิทยาศาสตร์แล้ว มีการสร้างภาพผู้เชี่ยวชาญขึ้นมาเพื่อให้ดูน่าเชื่อถือ

9. ข่าวที่ให้ข้อมูลผิด ๆ (Misinformation) คือข่าวที่ไม่ได้ตรวจสอบให้แน่ชัดก่อน ข้อมูลอาจจะทั้งจริงและเท็จผสมกัน ผู้ส่งสารตั้งใจจะส่งออกไป แต่อาจไม่ได้ตระหนักข่าวนั้นมีข้อมูลที่ผิดพลาดอยู่เช่น ข่าวลือ

10. ข่าวหลอกลวง (Bogus) เป็นข่าวปลอมที่เจตนาในการสร้างขึ้นมาเพื่อจงใจให้แพร่กระจาย โดยมีวัตถุประสงค์เพื่อหลอกลวงอาจจะมีเนื้อหาของเรื่อง ภาพ หรือข้อมูลต่าง ๆ ที่เป็นเท็จมาประกอบกัน อาจรวมถึงการแอบอ้างเป็นแหล่งข่าวหรือบุคคลที่อยู่ในเหตุการณ์ทุกวิธีการที่จะทำให้ข่าวนั้นปลอมที่สมบูรณ์มากขึ้น

2.1.3 รูปแบบเนื้อหาของข่าวปลอม

สรานนท์ อินทน์ (2563) ลักษณะเนื้อหาของข่าวปลอม (first draft news) ซึ่งเป็นองค์กรที่ตั้งขึ้นมาเพื่อต่อสู้กับข่าวปลอม ร่วมกับโซเซียลมีเดีย รวมถึงเฟซบุ๊ก ทวิตเตอร์ ได้จัดรูปแบบเนื้อหาของข่าวปลอมไว้ 7 รูปแบบโดยเรียงตามระดับความรุนแรงจากน้อยไปหามาก ดังนี้

1. เนื้อหาล้อเลียนเสียดสี เป็นข่าวล้อเลียน ไม่ได้มีเจตนาในการสร้างความเข้าใจผิดหรือต้องการให้ผู้อ่านหลงเชื่อ แต่ต้องการล้อเลียน หรือทำให้ขบขัน มักเป็นการล้อเลียนเหตุการณ์บ้านเมืองในปัจจุบัน หรือล้อเลียนคนมีชื่อเสียง โดยมักมีการจัดหน้าเลียนแบบหรือรูปแบบให้เหมือนข่าวจริง จนบางครั้ง ผู้อ่านหลงเชื่อว่าเป็นข่าวจริงได้ เช่น บทความใน “ผู้จัดกวน” หรือในเว็บไซต์ “ข่าวปด” จริงแล้วข่าวล้อเลียนไม่ใช่ข่าวปลอม แต่การที่ผู้อ่านอาจขาดความรู้ความเข้าใจ จึงทำให้ข่าวล้อเลียนมีคุณลักษณะถูกจัดว่าเป็นข่าวปลอมได้เช่นกัน ข่าวประเภทนี้มีระดับความรุนแรงน้อยที่สุด

2. เนื้อหาไม่ตรงพาดหัว คือข่าวที่มีรูปแบบการเชื่อมโยงเนื้อหาที่ผิด พาดหัวข่าว รูปภาพ หรือคำบรรยาย ไม่ได้เชื่อมโยงกับเนื้อหาข่าวจริง เป็นการโยงสองสิ่งไม่ได้เกี่ยวข้องกันเลยแต่ถูกนำมากล่าวถึงในข่าวเดียวกันหรือทำให้มาเชื่อมโยงกัน โดยพาดหัวมักจะเป็นการเร้าอารมณ์ดึงดูดให้คนเข้ามาอ่าน เนื่องจากปัจจุบัน ผู้อ่านมักจะมีช่วงความสนใจที่สั้นลงจึงทำให้สำนักข่าวออนไลน์เน้นพาดหัวข่าวให้หวือหวา หรือใช้รูปที่ไม่ได้เกี่ยวข้องโดยตรงกับข่าวเพื่อดึงดูดความสนใจให้คนกดเข้ามาอ่าน

3. เนื้อหาชี้นำ เป็นข่าวที่มีเนื้อหาข้อเท็จจริงแต่จงใจบิดเบือนเรื่องราวหรือใส่ร้ายผู้อื่นให้เข้าใจผิดโดยการชี้นำไปในทางใดทางหนึ่ง เป็นการเขียนข่าวโดยใช้อคติของผู้เขียน เช่น ข่าวยุทธพลดลือคักกัญชาเสรีที่นักข่าวเจตนาชี้นำให้คนอ่านเข้าใจผิดว่ารัฐบาลจะทำให้การเสพกัญชาถูกกฎหมายทั้งที่ในความจริงเป็นการเปิดโอกาสให้สามารถนำกัญชาไปใช้ประโยชน์ทางการแพทย์

4. เนื้อหาที่ผิดบริบท เป็นข่าวที่มีเนื้อหาข้อมูลจริงแต่นำบริบทอื่นที่ไม่เกี่ยวกับเนื้อหานั้นมาเชื่อมโยงทำให้คนตีความผิด เช่น เนื้อหาของข่าวเป็นเรื่องจริงแต่นำภาพประกอบจากแหล่งอื่นมาประกอบ เช่น ข่าวกัญชาจับงูเห่าลื้อเข้าบ้าน เนื้อข่าวเป็นเรื่องจริง แต่นำภาพประกอบการจับงูนาคอนดาจากภาพยนตร์มาประกอบ ซึ่งทำให้คนเข้าใจผิดว่างูเห่าลื้อที่ถูกจับนั้นตัวใหญ่มาก

5. เนื้อหาแอบอ้าง เป็นข่าวที่มีการแอบอ้างบุคคล แหล่งข้อมูลหรือแหล่งข่าวที่ไม่จริง หรืออ้างตัวเป็นแหล่งข่าวที่น่าเชื่อถือ ข่าวปลอมประเภทนี้มักเป็นข่าวออนไลน์ที่สร้างรูปแบบให้

เหมือนสำนักข่าวจริง ๆ หรือแอบอ้างชื่อสำนักข่าวที่มีชื่อเสียง ซึ่งสร้างความสับสนให้ผู้อ่านอย่างมาก เช่นแอบอ้างเป็นสำนักข่าว CNN โดยใช้รูปแบบและชื่อโดเมนที่ใกล้เคียงกับของเว็บไซต์ของ CNN

6. เนื้อหาหลอกลวง เป็นข่าวตัดต่อ หรือข่าวที่มีเนื้อหาข้อมูลหรือภาพข่าวจริง ๆ แต่ถูกดัดแปลงด้วยการปลอมหรือตัดต่อ เพื่อสร้างเรื่องหลอกลวง เช่น ภาพของอดีตประธานาธิบดีสหรัฐ George W. Bush อ่านนิทานกับเด็กในโรงเรียนแห่งหนึ่ง แต่เขากลับถือหนังสือกลับหัว แต่ความจริงแล้วมันเป็นภาพตัดต่อที่สร้างขึ้น และในภาพจริงนั้นเขาถือหนังสือถูกต้อง

7. เนื้อหาที่ก่อกวน เป็นข่าวที่ก่อกวนขึ้นมาทั้งหมดเป็นข้อมูลเท็จ 100% มีเจตนาที่จะหลอกลวงหรือใส่ร้ายข่าวถูกสร้างขึ้นโดยผู้ไม่หวังดี โดยอาจจะทำเองหรือจ่ายเงินจ้างให้ผู้อื่นทำ เพื่อหวังผลให้เกิดความเข้าใจผิดในวงกว้าง เช่น การนำเสนอข่าวว่า ผู้มีชื่อเสียงบางคนได้เสียชีวิตแล้ว ทั้งที่เจ้าตัวยังมีชีวิตอยู่ ส่วนใหญ่ข่าวมักมีเนื้อหาเกี่ยวกับการเมือง เช่น ข้าราชการมนตรีตมกาแฟแก้วละหมื่นสองพันบาท ข่าวประเภทนี้มีระดับความรุนแรงมากที่สุด

2.1.4 สาเหตุที่ทำให้เกิดข่าวปลอม

สุปัญญา อภิวงศ์โสภณ (2561) ได้กล่าวถึงสาเหตุที่ผู้อ่านให้ความสนใจข่าวปลอม ข่าวลวง เกิดจากประเด็นดังนี้

1. ผู้คนสนใจอยากรู้อยากเห็นและสงสัยในข่าวที่เป็นประเด็น อาจเป็นเรื่องที่อยู่บนพื้นฐานข้อเท็จจริงบางอย่าง หรืออาจเป็นเรื่องลอบที่แต่งขึ้นก็ได้ ซึ่งความไม่แน่นอนที่ปรากฏในเนื้อหาข่าว จะกระตุ้นให้ผู้คนสนใจติดตามอย่างไม่รู้ตัว

2. ผู้คนส่วนใหญ่มีความอยากรู้อยากเห็นในเรื่องคนอื่น เนื่องจากมนุษย์เป็นสัตว์สังคม จึงมีความอยากรู้อยากเห็นในเรื่องราวต่าง ๆ ที่เกิดขึ้นในสังคม ที่แม้จะไม่ใช่วิจัยของตนเองก็ตาม

3. ผู้คนชอบเรื่องที่เห็นคนอื่นทำอะไรผิดไปจากทำนองครองธรรม ที่ควรจะเป็น เนื่องจากจะทำให้รู้สึกสะใจ เหมือนกำลังได้ติดตามนิยายเรื่องหนึ่ง

4. ผู้อ่านชอบอ่านเรื่องราวจะเดาตอนจบของเรื่องได้ ไม่ว่าเรื่องราวนั้นจะเป็นจริงหรือไม่ก็ตาม หรือจะมีความเป็นไปได้อย่างไร ซึ่งเรื่องราวที่ได้อ่าน ทำให้เกิดความตื่นเต้น สนใจติดตามตอนจบนั้นจริง ๆ

5. ผู้อ่านชอบการวิพากษ์วิจารณ์ ซึ่งในข่าวฉาวส่วนใหญ่จะเป็นเรื่องราวการกระทำความผิดของคน ซึ่งเมื่อเรื่องราวถูกเปิดเผยสู่สาธารณชนแล้ว จะมีการวิเคราะห์ที่โจ่งแจ้งในลักษณะว่ากล่าวตึง นับว่าเป็นการลงโทษจากสังคมออนไลน์ โดยผู้คนส่วนใหญ่จะรู้สึกพอใจกับผลการกระทำจากสังคมออนไลน์ในลักษณะที่พึงพอใจกับความยุติธรรมที่เกิดขึ้น และมีความพึงพอใจกับการลงโทษคนผิดในประเด็นข่าว

สรานนท์ อินทนนท์ (2563) กล่าวว่า ข่าวปลอมถูกสร้างโดยใครก็ได้ที่สามารถเข้าถึงอินเทอร์เน็ตเนื่องจากเทคโนโลยีปัจจุบันทำให้การสร้างข่าวที่ดูเหมือนจริงทำได้ง่ายยิ่งขึ้น ไม่ว่าจะเป็นการตัดต่อภาพ ทำวิดีโอ สร้างเว็บไซต์ หรือแพลตฟอร์มที่ดูเหมือนข่าวจริง ๆ และข่าวเหล่านี้ไม่จำเป็นต้องผ่านกระบวนการตรวจสอบความถูกต้องหรือกลั่นกรองจากกองบรรณาธิการ ยิ่งไปกว่านั้น ความเป็นนิรนามของสื่อออนไลน์ทำให้คนเหล่านี้ปกปิดตัวตนที่แท้จริงได้ และเทคโนโลยีอินเทอร์เน็ตยังทำให้ข่าวปลอมง่ายต่อการเข้าถึงและแพร่สะพัดได้อย่างรวดเร็วและเป็นวงกว้าง เรื่องที่สำคัญไม่น้อยไปกว่าใครเป็นคนสร้างข่าว คือ แรงจูงใจของพวกเขา การสร้างข่าวปลอมของนักสร้างข่าวเกิดจากเหตุผลและแรงจูงใจได้มากมาย เช่น ความเกลียดชัง คับแค้นใจ เผยแพร่อุดมการณ์ ใส่ร้ายคู่แข่ง ขยายสินค้าและบริการอย่างไรก็ตาม แรงจูงใจและสาเหตุหลักของการสร้างข่าวปลอมมี 3 ประการดังนี้

1. ล้อเลียนเสียดสี ผู้สร้างข่าวปลอมต้องการให้เกิดความขบขัน โดยการล้อเลียนหรือเสียดสีผู้มีอำนาจ คนที่มีชื่อเสียง หรือเหตุการณ์ปัจจุบัน เนื่องจากเป็นการง่ายกว่าที่ทำให้ผู้อ่านสนใจที่ตัวบุคคล แทนที่จะเป็นการเสนอความเห็นหรืออภิปรายเรื่องนโยบายที่ซับซ้อน เช่น ข่าวล้อเลียนในหนังสือพิมพ์ผู้จัดกวน

2. สร้างอิทธิพลต่อความคิดความเชื่อ ผู้สร้างข่าวปลอมอาจมีอคติ หรือมีทัศนคติที่เอนเอียงเลือกข้าง จึงต้องการชักนำผู้อ่านให้คล้อยตามโดนการบิดเบือนข้อมูล เนื้อหาข่าวอาจจะเป็นการชื่นชมบุคคลหรือฝ่ายที่ตนเองชอบเกินจริง หรือใส่ร้ายฝ่ายตรงข้าม เช่น การพาดหัวข่าวที่ใส่ “ความเห็นส่วนตัว” ลงไปในลักษณะชักนำผู้อ่าน ถึงแม้เนื้อหาของข่าวนั้นจะมีความจริงอยู่บ้าง

3. การสร้างรายได้ ข่าวปลอมสามารถสร้างรายได้ให้คนทำได้ในการแข่งขันทางการเมืองหรือทางธุรกิจ อาจมีผู้ว่าจ้างให้คนทำข่าวปลอมเพื่อใส่ร้ายฝ่ายตรงข้ามด้วยการบิดเบือนข้อมูลและเผยแพร่ในโลกออนไลน์ ในโลกอินเทอร์เน็ต ข่าวที่มีคนเข้าไปอ่านมากจะทำรายได้จากโฆษณา ยิ่งคนเข้าไปอ่านข่าวมากเท่าไร ก็ยิ่งเพิ่มโอกาสที่โฆษณาในหน้านั้น จะถูกเห็น และทำให้ผู้ดูแลเว็บไซต์นั้น ๆ มีรายได้ ผู้ดูแลจึงมักใช้พาดหัวข่าวในลักษณะคลิกเบท ในบางกรณี ผู้สร้างข่าวปลอมหลอกให้ผู้อ่านชมคลิปข่าวปลอมเป็นจำนวนหลักหมื่น แล้วเว็บไซต์ประเภทนี้ได้ควบคุมบัญชีเฟสบุ๊คของผู้หลงเข้าไปเพื่อไปใช้ประโยชน์ทางการค้าพอกดดูคลิป ก็จะต้องลงชื่อเข้าใช้ เฟสบุ๊คพอกดอนุญาตในเฟสบุ๊ค คนสร้างเว็บปลอมนั้นก็ใช้เฟสบุ๊คของเหยื่อเป็นบอท (หุ่นยนต์) และนำไปเป็นส่วนหนึ่งของกองทัพโลก เพื่อขายโลกอีกต่อหนึ่ง

2.1.5 ปัญหาและผลกระทบที่เกิดจากข่าวปลอม

สถาบันอิศรา (2565) การเข้ามาของสื่อสังคมออนไลน์ ประกอบกับความก้าวหน้าทางเทคโนโลยีการสื่อสารและระบบอินเทอร์เน็ต ส่งผลกระทบอย่างหนักต่อพฤติกรรมและการเสพติดข่าวสารของผู้บริโภคในยุคปัจจุบันที่เปลี่ยนแปลงไป ผู้คนส่วนใหญ่หันมาเสพข่าวสารจากสื่อโซเชียลมีเดียมาก

ขึ้น สิ่งที่มาคือบุคคลเหล่านั้นมีการส่งต่อข้อความที่มีลักษณะเนื้อหาของข่าวปลอม (Fake news) เป็นจำนวนมาก มานิจ สุขสมจิตร (2565) โดยเฉพาะที่มีเนื้อหาเกี่ยวกับข้อมูลสุขภาพ ซึ่งพบว่า มีการส่งต่อกันแบบผิด ๆ สูงกว่าร้อยละ 80 จากบรรดาข่าวปลอมทั้งหมด และในระยะหลังยังพบเห็นข่าวปลอมเกี่ยวกับประเด็นทางการเมืองหรือข่าวต่างประเทศอีกด้วย นอกจากนี้ยังพบข้อมูลที่น่าสนใจอีกว่า ข่าวปลอมในสื่อสังคมออนไลน์ อย่างเช่น ทวิตเตอร์ มีแนวโน้มที่จะถูกทวีตมากกว่าข่าวที่น่าเชื่อถืออย่างถูกต้องและเป็นความจริง ผลที่เกิดจากการศึกษาของผลกระทบของข่าวปลอม พบว่า มีผลกระทบต่อผู้บริโภคสื่อ ทั้งในแง่ทัศนคติความเชื่อ เช่น เรื่องการเมือง การสร้างความรู้ความเข้าใจที่ผิดเพี้ยน บางเรื่องยังส่งผลกระทบต่อการใช้ชีวิต เช่น เรื่องยา ข้อมูลสมุนไพรต่าง ๆ บางเรื่องนำไปสู่การเอาผิดเอาเปรียบผู้บริโภค เช่น เรื่องผลิตภัณฑ์สุขภาพ นอกจากนี้หากผู้บริโภคสื่อ มีการส่งต่อเรื่องราวที่เป็นเท็จ หมิ่นประมาท อาจเข้าข่ายความผิดทางกฎหมายได้ด้วย

สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (2565) ในปัจจุบันในขณะที่บุคคลได้รับข้อมูลข่าวสารมากมาย ทั้งเรื่องจริงเรื่องปลอม โดยที่ไม่สามารถแยกแยะได้เรื่องราวของข่าวไม่ใช่เรื่องจริง อาจส่งผลให้เกิดผลเสียตามมา ดังนี้

1. ผู้รับได้ข้อมูลไม่ถูกต้อง ทำให้ตัดสินใจผิดพลาด อาจส่งผลเสียต่อชีวิตและทรัพย์สิน เช่น แอร์ข้อมูลว่าดื่มน้ำมะนาวช่วยรักษาโรคมะเร็งได้ ผู้ป่วยอาจเลิกไปรักษาด้วยวิธีเคมีบำบัดกับหมอ ทำให้มะเร็งลุกลามถึงขั้นเสียชีวิต

2. ผู้รับเกิดความตระหนกตกใจ เช่น ข้อมูลเท็จเกี่ยวกับภัยพิบัติ หรือโรคระบาด (Narumon Pinto, 2020) อาจทำให้ผู้คนแตกตื่น แห่กักตุนของกินของใช้ หรือไปเข้าคิวฉีดวัคซีนป้องกันโรคซึ่งไม่เกิดขึ้นจริง ข่าวการเมืองหรือนโยบายของรัฐที่อาจทำให้หุ้นขึ้นหรือลง นักลงทุนเทขายหุ้นหรือซื้อเพื่อเก็งกำไร ผู้ถูกแอบอ้างได้รับความเสียหาย เช่น ถูกล้อเลียน ดูหมิ่น กลั่นแกล้งรังแก (bully) เพราะข้อมูลเท็จที่เกิดจากการตัดต่อให้ดูตลกขบขัน ถูกเกลียดชังจากข้อมูลเท็จเชิงใส่ร้ายป้ายสี หรือตัวอย่างข่าวดาราดังป่วยหนักใกล้เสียชีวิต ทำให้ประชาชนสงสาร มีฉาชีพดกฉวยโอกาสเรียกรับเงินช่วยเหลือครอบครัวดาราดัง

3. ข้อมูลที่ทำให้เกิดความขัดแย้งในสังคม เช่น ข้อมูลเท็จทางการเมือง ข่าวสถานการณ์ระหว่างประเทศ อาจนำไปสู่ความไม่สงบสุขในสังคม สร้างปัญหาระหว่างประเทศได้

สรานนท์ อินทนนท์ (2563) ได้สรุปผลกระทบที่เกิดขึ้นหลังจากการแชร์ข่าวปลอมไว้ดังนี้

1. ผลกระทบต่อความคิดและความเชื่อ เป็นผลกระทบที่ชัดเจนและสำคัญที่สุด เพราะข้อมูลข่าวสารย่อมมีอิทธิพลต่อความคิด ทัศนคติของผู้อ่าน โดยผู้อ่านมักจะตัดสินใจบนพื้นฐานของข้อมูลที่ได้รับ หากข้อมูลที่เราได้รับเป็นข่าวปลอม ข่าวบิดเบือนหรือข่าวที่ก่อกวน พวกเขาก็ไม่สามารถใช้การวิจารณ์ญาณได้อย่างถูกต้อง เนื่องจากขาดข้อเท็จจริง และส่งผลกระทบต่อตัดสินใจของพวกเขา ไม่ว่าจะเป็นการตัดสินใจระดับชาติ

2. ผลกระทบด้านการเงินและสุขภาพ เป็นข่าวปลอมที่ถูกสร้างขึ้นโดยมีรายได้เป็นแรงจูงใจจะชักจูงให้ผู้รับสารจ่ายเงินเพื่อสินค้าและบริการ และหากสินค้าและบริการนั้นไม่ได้คุณภาพก็อาจส่งผลเสียต่อร่างกายและสุขภาพของผู้หลงเชื่อด้วย เช่น ข่าวปลอมที่ชักชวนให้คนอ่านมาลงทุนเพื่อจะได้รับผลตอบแทนที่ฟังดูเกินจริง หรือโฆษณาผลิตภัณฑ์ลดความอ้วนที่นำเสนอในรูปแบบของวิทยาศาสตร์ลวงโลกให้คนคล้อยตาม

3. ผลกระทบด้านอารมณ์ความรู้สึก ข่าวคลิกเบทอาจสร้างความหงุดหงิดให้ผู้อ่านเมื่อพบว่าเนื้อหาข่าวไม่ได้มีอะไรน่าสนใจอย่างที่รูปภาพหรือพาดหัวดึงดูดให้เข้ามาอ่าน และข่าวปลอมที่สร้างขึ้นมาด้วยความคะนองอาจทำให้ผู้อ่านเกิดความกลัวและวิตกกังวล เช่น ข่าวพยากรณ์อากาศภัยพิบัติที่ไม่มีมูลความจริง หรือข่าวกลั่นแกล้งว่ามีชาตกรโรคจิตกำลังออกอาละวาดละแวกบ้าน

4. ผลกระทบด้านทัศนคติ เป็นข่าวที่ขึ้นอาจสร้างอคติและทัศนคติเชิงลบแก่บุคคลหรือกลุ่มที่ถูกใส่ร้ายอย่างไม่เป็นธรรม เช่น โฆษณาชวนเชื่อทางการเมือง ข่าวต่อต้านรัฐบาล หรือข่าวที่สร้างความเกลียดชังต่อคนที่มีอัตลักษณ์ทางสังคมที่แตกต่างจากคนส่วนใหญ่ เช่น กลุ่มเกย์ ชาวมุสลิม หรือคนต่างด้าว

5. ผลกระทบด้านความรุนแรง เป็นข่าวลือที่เร้าอารมณ์และโจมตีใส่ร้ายบุคคลอาจก่อให้เกิดความรุนแรง ผู้ที่ถูกกล่าวหาอาจจะถูกกลั่นแกล้งทางออนไลน์เพราะความเข้าใจผิด หรืออาจถูกข่มขู่ คุกคาม และทำร้ายในชีวิตจริงได้

6. ผลกระทบต่อสื่อวิชาชีพ ข่าวปลอมสร้างความเสียหายต่อชื่อเสียงและทำลายความน่าเชื่อถือของสำนักข่าวได้ สำนักข่าวหลายแห่งทั้ง ไทยรัฐ ข่าวสด และมติชน ต่างถูกหลอกเลียนแบบเว็บไซต์เพื่อหลอกลวงผู้อ่าน ซึ่งกว่าที่องค์กรสื่อจะรู้ตัว เว็บไซต์ปลอมเหล่านั้นได้ถูกส่งต่อและสร้างความสับสนกับผู้อ่านเรียบร้อยแล้ว

7. ผลกระทบด้านสังคม ข่าวลือที่เกี่ยวข้องกับประเด็นสำคัญในสังคมแพร่ระบาดจะทำให้คนขาดข้อเท็จจริงที่ใช้ในการวิเคราะห์ ส่งผลให้เกิดความสับสน การแบ่งฝ่ายและสร้างความขัดแย้งในสังคมเพราะแต่ละฝ่ายรับรู้ชุดข้อมูลที่ต่างกัน ข่าวปลอมยังส่งผลต่อการตัดสินใจเลือกตั้งของผู้มีสิทธิโหวตเมื่อข้อมูลผิดได้ถูกเผยแพร่ ข่าวโฆษณาชวนเชื่อของรัฐบาลเองก็มีผลกระทบต่อความคิดของสังคมเช่นกัน เช่น ข่าวชี้นำชักชวนให้ลงประชามติรับร่างรัฐธรรมนูญไปก่อนหากต้องการให้ประเทศมีการเลือกตั้งโดยเร็วและทำให้ประชาชนลดความสนใจที่จะอ่านเนื้อหาของร่างรัฐธรรมนูญและสนใจการเลือกตั้งมากกว่า

จากผลกระทบของข่าวปลอมยังทำให้เกิดความขัดแย้งและสับสนในสังคม การจัดการกับข้อมูลที่ไม่เป็นความจริงและข่าวปลอมเป็นปัญหาที่ซับซ้อนและจำเป็นต้องบูรณาการความร่วมมือกันระหว่าง ภาคประชาสังคมและรัฐบาล นักวิชาการรวมถึงองค์กรด้านเทคโนโลยีและสื่อมวลชนรวมถึง

ประชาชนจะต้องมีทักษะในการตรวจสอบข่าวปลอมได้ด้วยตนเองเพื่อเป็นการป้องกันผลกระทบที่อาจจะเกิดขึ้น ผู้วิจัยจึงสรุปประเด็นที่เกี่ยวข้องกับข่าวปลอม สามารถแสดงข้อมูลดังตาราง 1

ตาราง 1 สรุปประเด็นที่เกี่ยวข้องกับข่าวปลอม

ชื่อผู้แต่ง	ชื่อวิจัย	สรุป	ประเด็นที่สนใจ
สุนันทา แยม ทัพและ อรัญญา ศิริ ผล (2564)	ข่าวปลอมกับ การทำงานของ นักข่าวในยุค ดิจิทัล	ท่ามกลางความก้าวหน้าทางเทคโนโลยี สื่อสารยุคดิจิทัลที่เปิดโอกาสให้ ผู้ใช้งาน อินเทอร์เน็ตสามารถสร้างสรรค์เนื้อหา สื่อสารด้วยตนเอง และเปิดให้ แบ่งปัน เนื้อหาแก่ผู้อื่นได้โดยง่าย ด้านหนึ่งเป็น การเปิดให้โลกแห่งข่าวสารไหลเวียนได้ ง่าย แต่อีกด้านกลับ ประสบปัญหาที่ผู้รับ สารไม่สามารถแยกแยะ ได้ชัดว่า อะไรคือ ข้อเท็จจริง (fact) อะไรคือข่าวปลอม (fake) ในบทความนี้มีวัตถุประสงค์ ตรวจสอบความหมายของข่าวปลอมใน บริบทยุคดิจิทัล และ เพื่อเสนอ กระบวนการตรวจสอบข่าวปลอมให้แก่ นักข่าวและผู้ใช้สื่อ	1. ตรวจสอบ ความหมายของ ข่าวปลอม 2. เสนอ กระบวนการใหม่ ในการตรวจสอบ ข่าวปลอม
Atodiresei et al., (2018)	Identifying Fake News and Fake Users on Twitter	เครือข่ายสังคม เช่น Facebook หรือ Twitter เป็นเครือข่ายถูกทำเป็นบัญชี ปลอมและเข้าซุ่มขึ้นมา และเผยแพร่ ข้อมูลที่เป็นเท็จ สนับสนุนหรือโจมตี แนวคิด พลิตภัณฑ์ หรือ ข่าวและการไลค์ ต่าง ๆ ไปในทิศทางที่ผิด ซึ่งส่งผลและมี อิทธิพลต่อผู้ใช้เครือข่ายจริงในการ ตัดสินใจ	1. นำเสนอการ สร้างระบบโดยมี จุดประสงค์เพื่อ ระบุผู้ใช้ปลอม และข่าวปลอมใน เครือข่าย โซ เชียล Twitter

ตาราง 1 สรุปประเด็นที่เกี่ยวข้องกับข่าวปลอม (ต่อ)

ชื่อผู้แต่ง	ชื่อวิจัย	สรุป	ประเด็นที่สนใจ
Aphiwongsophon & Chongstitvatana, (2018)	Detecting Fake News with Machine Learning Method	ข้อมูลจำนวนมากถูกสร้างขึ้นบนเครือข่ายสังคมออนไลน์ด้วยรูปแบบโซเชียลมีเดียต่างๆ มีโพสต์จำนวนมาก แหล่งข้อมูลที่น่าเชื่อถือทำให้ผู้คนเห็นข่าวปลอม เรื่องลวง ข่าวลือ ทฤษฎีสมคบคิด และข่าวที่ทำให้เข้าใจผิด ข่าวปลอมที่ออกอากาศซึ่งสร้างความสับสนตามลักษณะของเหตุการณ์	1. เทคนิคการเรียนรู้ของเครื่องเพื่อตรวจจับข่าวปลอม วิธีที่นิยมใช้ในการทดลอง ได้แก่ Naïve Bayes, Neural Network และ Support Vector Machine (SVM)
Joju & Kammath, (2021)	Analysis on Fake news Detection using machine learning	จากการระบาดของ COVID-19 ปัญหาด้านสุขภาพและเศรษฐกิจหลายเกิดขึ้น ทำให้เกิดข้อมูลที่ผิดและความสับสนทั่วโลก ประเด็นข่าวปลอมได้รับความนิยมเพิ่มขึ้นในการแพร่กระจายของการสร้างข่าว โซเชียลมีเดียกลายเป็นแหล่งข่าวหลักสำหรับผู้คน นับล้านเนื่องจากเข้าถึงได้ง่าย ราคาถูก น่าดึงดูดยิ่งขึ้น และเผยแพร่เนื้อหาข่าวบนโซเชียลมีเดีย ผู้ที่แพร่หวังผลประโยชน์การเงินและการเมือง ข่าวปลอมดังกล่าวมีผลกระทบด้านลบต่อสังคม	1. วิธีการตรวจจับข่าวปลอม 2. ทดสอบอัลกอริทึมการเรียนรู้ของเครื่องเพื่อวิเคราะห์ข่าวปลอม

พูน ปณ จิต ชเว

ตาราง 1 สรุปประเด็นที่เกี่ยวข้องกับข่าวปลอม (ต่อ)

ชื่อผู้แต่ง	ชื่อวิจัย	สรุป	ประเด็นที่สนใจ
Srinivasa Rao et al., (2021)	Fake News Detection Using Machine Learning Technique	จากปี 1605 มีลักษณะของข่าวหนังสือพิมพ์ แต่มาในปี 2021 ในยุคดิจิทัลได้มีการพัฒนาข่าวไปในลักษณะของโซเชียลมีเดีย ข่าวปลอมก็มีวิวัฒนาการเช่นกัน และผู้คนมักจะเชื่อโดยข่าวปลอมที่มีอยู่ซึ่งถูกแชร์โดยโปรไฟล์ปลอมในสื่อดิจิทัล ทำให้ผู้บริโภคข่าวหลงเชื่อ และทำให้เข้าใจในข้อมูลต่าง ๆ ที่ผิดไป	1. สร้างแบบจำลองเพื่อตรวจจับข่าวปลอมในเว็บไซต์ต่างๆ 2. การจำแนกประเภทของข่าวปลอมจากการใช้วิธีการรวมกลุ่มการเรียนรู้ของเครื่อง
Meesad et al., (2022)	Artificial Intelligent Techniques for Thai Fake News Detection	การเผยแพร่ข่าวปลอมที่มีจำนวนมากซึ่งเป็นปัญหาที่ทำให้มีผู้หลงเชื่อข่าวปลอมที่เป็นข่าวจริงและนำไปแชร์ต่อทำให้เกิดความเสียหาย และทำให้เกิดปัญหาอื่นๆตามมา ด้วยในปัจจุบันความนิยมของสาธารณชนในโซเชียลมีเดีย และเครือข่ายสังคม ทำให้เกิดการแพร่กระจายของข่าวปลอม	1. วิธีทางปัญญาประดิษฐ์สำหรับการตรวจข่าวปลอมภาษาไทย 2. เสนอกรอบการทำงานเพื่อสร้างระบบตรวจจับข่าวปลอมออนไลน์ของไทยโดยอัตโนมัติ
Aldwairi & Alwahedi (2018)	Detecting Fake News in Social Media Networks	โซเชียลมีเดียกลายเป็นหนึ่งในข้อมูลหลักสำหรับผู้คนทั่วโลก การใช้โซเชียลมีเดียมีต้นทุนต่ำ เข้าถึงได้ง่าย และเผยแพร่อย่างรวดเร็ว มันมาพร้อมกับการเปิดรับ 'ข่าวปลอม' ที่โดยเจตนาโดยมุ่งเป้าไปที่การหลอกลวงและจัดการผู้อ่านเพื่อติดตามวาระทางการเมืองหรือเศรษฐกิจ	1. การตรวจจับข่าวปลอมโดยอัตโนมัติ

ตาราง 1 สรุปประเด็นที่เกี่ยวข้องกับข่าวปลอม (ต่อ)

ชื่อผู้แต่ง	ชื่อวิจัย	สรุป	ประเด็นที่สนใจ
Yang et al., (2019)	Unsupervised Fake News Detection on Social Media: A Generative Approach	โซเชียลมีเดียได้กลายเป็นหนึ่งในช่องทางหลักสำหรับผู้คนในการเข้าถึงและบริโภคข่าวสาร เนื่องจากความรวดเร็วและต้นทุนต่ำในการเผยแพร่ข่าว โซเชียลมีเดียยังทำให้เป็นแหล่งของการเผยแพร่ข่าวปลอมซึ่งส่งผลกระทบในทางลบต่อบุคคลและสังคมอื่นๆ <i>ข่าวปลอม</i> ซึ่งประกอบด้วยบทความที่เขียนขึ้นโดยเจตนาเพื่อถ่ายทอดข้อมูลเท็จ	1. เสนอวิธีการตรวจจับข่าวปลอม 2. สร้างแบบจำลองการจัดหมวดหมู่โดยพิจารณาจากชุดคุณลักษณะต่างๆ รวมถึงเนื้อหาข่าว
Monti et al., (2019)	Fake News Detection on Social Media using Geometric Deep Learning	โซเชียลมีเดียเป็นหนึ่งในแหล่งข่าวหลักสำหรับผู้คนหลายล้านคนทั่วโลก มีต้นทุนต่ำ เข้าถึงได้ง่าย และเผยแพร่อย่างรวดเร็ว ' <i>ข่าวปลอม</i> ' ซึ่งเขียนขึ้นโดยเจตนา สร้างขึ้นมาเพื่อให้ผู้อ่านเข้าใจผิด	1. รูปแบบการตรวจจับข่าวปลอมอัตโนมัติแบบใหม่โดยอิงจากการเรียนรู้เชิงลึกทางเรขาคณิต 2. อัลกอริทึมหลักคือโครงข่ายประสาทเทียมแบบคลาสสิกให้เป็นกราฟ

จากตาราง 1 จะเห็นว่าข่าวปลอมได้รับความนิยมอย่างมากควบคู่ไปกับการนำสื่อดิจิทัลมาใช้อย่างรวดเร็ว โดยเฉพาะสื่อสังคมออนไลน์ที่เครือข่ายกระจายอำนาจหรือการเผยแพร่ของเนื้อหาข่าวที่ไม่มีการกรอง ข่าวปลอมก่อให้เกิดความกังวลอย่างกว้างขวางซึ่งเป็นปัญหาที่ทำให้มีผู้หลงเชื่อข่าวปลอมที่เป็นข่าวจริงและนำไปแชร์ต่อทำให้เกิดความเสียหาย และทำให้เกิดปัญหาอื่นๆ ตามมา จึงทำให้นักวิจัยหลายๆ ท่าน ได้หาเทคนิคและวิธีการที่จะนำมาใช้ในการจำแนกหรือตรวจสอบข่าวปลอม

ขึ้น จากการศึกษางานวิจัยดังกล่าว ผู้วิจัยจึงได้สนใจหาวิธีการและเทคนิคต่างๆ ที่เกี่ยวข้อง เพื่อนำมาใช้ในการปรับปรุงประสิทธิภาพวิธีการเรียนรู้ของเครื่องสำหรับการจำแนกข่าวปลอม

2.2 สื่อสังคมออนไลน์ (Social media)

ปัจจุบันการสื่อสารของมนุษย์ได้รับการพัฒนาให้สื่อสารถึงกันได้อย่างรวดเร็ว แพลตฟอร์มโซเชียลมีเดียจึงเป็นอัลกอริทึมที่เปรียบเสมือนการเชื่อมโยงกัน เพื่อสร้างเครือข่าย ในการตอบสนองความต้องการทางสังคมอันเป็นผลมาจากความก้าวหน้าด้านเทคโนโลยีคอมพิวเตอร์ที่ได้เข้ามามีบทบาทในชีวิตประจำวันของมนุษย์มากขึ้น ซึ่งเป็นสื่อชนิดหนึ่งที่ผู้ใช้สามารถมีส่วนร่วม สร้าง และแลกเปลี่ยนความคิดเห็นต่าง ๆ ผ่านระบบอินเทอร์เน็ตได้ คือ โซเชียลมีเดีย (Social media)

2.2.1 ความหมายของสื่อสังคมออนไลน์ (Social media)

กฤษฎาญจน์ เสนแก้ว (2560) กล่าวว่า สื่อสังคมออนไลน์ คือ เครือข่ายสังคมออนไลน์เป็นรูปแบบของเว็บไซต์ ที่มีเครือข่ายความสัมพันธ์เสมือนที่โยงใยให้บุคคลที่มีเรื่องที่สนใจเหมือนกันมาพูดคุยกันสามารถเชื่อมโยงระหว่างบุคคลอีกบุคคลหนึ่งเป็นการสร้างเครือข่ายสังคม (นครินทร์ ชานะมัย, 2563) สำหรับผู้ใช้งานในอินเทอร์เน็ต เขียนและอธิบายความสนใจ กิจกรรมที่ได้ทำ และเชื่อมโยงกับความสนใจและกิจกรรมของผู้อื่น ในบริการเครือข่ายสังคมมักจะประกอบไปด้วย การแชท ส่งข้อความ ส่งอีเมลวิดีโอ เพลง อับโหลดรูป บล็อก โดยแอปพลิเคชันที่ได้รับความนิยมได้แก่ เฟสบุ๊ก, ไลน์, ทวิตเตอร์ และ อินสตาแกรมและยังสามารถเชื่อมโยงการสื่อสารภายในองค์กร และภายนอกองค์กรเข้าด้วยกันได้อย่างมีประสิทธิภาพ ไม่ต้องประสบปัญหาการบิดเบือนข้อความและยังรวดเร็ว วิดีโอ เพลง อับโหลดรูป บล็อก

สุชาลักษณ์ ธรรมดวงดี (2562) กล่าวว่า สื่อสังคมออนไลน์ คือ รูปแบบของเว็บไซต์หรือโปรแกรมประยุกต์ (Application) ที่มีเครือข่ายความสัมพันธ์เสมือนที่โยงใยให้บุคคลที่มีเรื่องที่สนใจเหมือนกันมาพูดคุยกันสามารถเชื่อมโยงระหว่างบุคคลอีกบุคคลหนึ่งได้ (วรเชษฐ์ โพธิ์อิน, 2562) เป็นการสร้างเครือข่ายสังคม สำหรับผู้ใช้งานในอินเทอร์เน็ต เขียนและอธิบายความสนใจ กิจกรรมที่ได้ทำ การเชื่อมโยงกับความสนใจและกิจกรรมของผู้อื่น ในบริการเครือข่ายสังคมมักจะประกอบไปด้วย การแชท ส่งข้อความ ส่งอีเมล วิดีโอ เพลง อับโหลดรูป บล็อก โดยแอปพลิเคชันที่ได้รับความนิยมได้แก่ เฟซบุ๊ก, ไลน์, ทวิตเตอร์ และ อินสตาแกรม

สกุลศรี ศรีสารคามและคณะ (2559) กล่าวว่า สื่อสังคมออนไลน์ คือ การรวมกันของแพลตฟอร์มทุกประเภทที่มีการเชื่อมต่ออินเทอร์เน็ต ทั้งแอปพลิเคชัน เว็บไซต์ สื่อสังคม และได้ให้ความหมายของสื่อสังคม คือ โซเชียลมีเดีย (Social media) ซึ่งเป็นเครือข่ายสังคมออนไลน์ เช่น เฟซบุ๊ก, ยูทูบเว็บไซต์ต่าง ๆ บล็อกต่าง ๆ เป็นต้น

จุฑารัตน์ ศราวรรณวงศ์ (2558) กล่าวว่า สื่อสังคมออนไลน์ คือ รูปแบบการสื่อสารทางอิเล็กทรอนิกส์ผ่านระบบอินเทอร์เน็ตผ่านทางผู้ใช้งานในการสร้างชุมชนออนไลน์เพื่อแบ่งปันความคิด แลกเปลี่ยนข้อมูล หรือเนื้อหาอื่น ๆ เป็นช่องทางที่ผู้ใช้งานสามารถเชื่อมโยงกับผู้ใช้งานอื่น ๆ ได้อย่างรวดเร็ว

ครองรัตน์ ดุลลาพันธ์ (2561) กล่าวว่า สื่อสังคมออนไลน์ คือ รูปแบบของเว็บไซต์ที่มีเครือข่ายความสัมพันธ์ที่เชื่อมโยงให้ผู้ให้บริการที่มีความสนใจเหมือนกันและพูดคุยระหว่างกันและกันผ่านอินเทอร์เน็ต

สรุป สื่อสังคมออนไลน์ (Social media) คือ รูปแบบของเว็บไซต์ที่มีการสื่อสารทางอิเล็กทรอนิกส์ผ่านทางอินเทอร์เน็ต โดยผู้ใช้เป็นผู้สื่อสารเรื่องราว รูปภาพ และวิดีโอ เชื่อมโยงกับความสนใจและกิจกรรมของผู้อื่น แบ่งปันให้ผู้อื่นที่อยู่ในเครือข่ายของตนได้รับรู้ในเรื่องดังกล่าว เครือข่ายสังคมออนไลน์ที่นิยมในปัจจุบันคือ เฟซบุ๊ก ไลน์ อิน스타그램 ทวิตเตอร์ เป็นต้น

2.2.2 ประเภทของโซเชียลมีเดีย

โซเชียลมีเดียเป็นอัลกอริทึมหรือแพลตฟอร์มดิจิทัลที่ช่วยให้ผู้ใช้สามารถสร้างและแชร์เนื้อหาเกี่ยวกับเพื่อน หรือสาธารณชนทั่วไปได้อย่างรวดเร็ว โดยที่เนื้อหาบนโซเชียลมีเดียนั้นค่อนข้างจะเปิดกว้าง และไม่มีเส้นแบ่งจำกัดว่าอะไรบ้างที่โพสต์ได้หรือไม่ได้ (เฉพาะบางแพลตฟอร์ม) การแชร์ข่าวหรือเนื้อหาเป็นไปอย่างรวดเร็ว และกว้างขวาง แต่ก็มีเรื่องต้องระวัง เช่น เนื้อหาที่ล่อแหลมไม่เหมาะสม หรือการแชร์ข่าวที่เป็นเท็จ (Fake news) ในปัจจุบันสามารถแบ่งประเภทของ โซเชียลมีเดียไว้ 5 ประเภทดังนี้ (Noria, 2022)

1. โซเชียลมีเดียแบบสังคมออนไลน์ โซเชียลมีเดียประเภทนี้ เป็นโซเชียลที่เชื่อมโยงผู้คนเอาไว้ด้วยการแลกเปลี่ยนความคิดเห็น ไอเดียต่างๆ หรือการติดตามผู้ที่มีความสนใจเดียวกัน เช่น เฟซบุ๊ก และไลน์ เป็นต้น
2. โซเชียลมีเดียแบบเครือข่ายโซเชียลมีเดียประเภทนี้อาจได้แก่คอนเทนต์ที่เป็นมีเดียมากกว่า เช่น แชร์รูปภาพ แชร์วิดีโอ อะไรแบบนี้ต้น เน้นอนว่าเรามักจะติดตามผู้แชร์เนื้อหาที่ถูกต้องเราเช่นกัน เช่น อินสตาแกรม, วิดีโอ และยูทูบ เป็นต้น
3. โซเชียลมีเดียแบบเน้นการสนทนา โซเชียลมีเดียประเภทนี้ เป็นโซเชียลมีเดียที่เน้นการตั้งคำถามและพูดคุยกันเป็นหลัก ไม่ได้เน้นการติดตามกันเหมือนสองประเภทแรก แต่เน้นการแลกเปลี่ยนทัศนคติที่มีต่อหัวข้อเรื่องนั้นๆ เช่น เรดดิต, ควอรา และพันทิพย์ เป็นต้น
4. โซเชียลมีเดียแบบเน้นบทวิจารณ์ โซเชียลมีเดียประเภทนี้ เป็นโซเชียลมีเดียที่เน้นการรีวิวเป็นหลัก ไม่ว่าจะเป็นการรีวิวโรงแรม ร้านอาหาร สถานที่ท่องเที่ยว เพื่อให้ผู้ที่วางแผนจะไปยังสถานที่ดังกล่าวได้ไต่ไลน์จากการรีวิวประกอบการตัดสินใจ เช่น การวิจารณ์การเดินทาง การวิจารณ์เกี่ยวกับธุรกิจ และการวิจารณ์เกี่ยวกับโรงแรม เป็นต้น

5. โซเชียลมีเดียแบบไมโครโซเชียลมีเดียประเภทนี้ เป็นโซเชียลมีเดียที่เน้นการการแชร์ข้อความสั้นๆ เน้นความฉับไวของการเลือนดูเนื้อหา เนื้อหาคอนข้างฟรีสไตร์จะมีสื่อด้วยก็ได้ หรือมีข้อความอย่างเดียวกก็ได้ เช่น ทวิตเตอร์, ทัมเบลอร์ หรือลูกผสมอย่างตึกตอก เป็นต้น

2.2.3 ชนิดของสื่อสังคมออนไลน์

สื่อสังคมออนไลน์มีหลากหลายรูปแบบ มีทั้งประเภทที่เป็นอัลกอริทึม และการประยุกต์ใช้สามารถแบ่งชนิดของสื่อสังคมออนไลน์ได้ สามารถแสดงข้อมูลดังตาราง 2 (วรารพร คำจัน, 2562)

ตาราง 2 ชนิดของสื่อสังคมออนไลน์

ชนิดของสื่อสังคมออนไลน์	คำอธิบาย	ตัวอย่าง
เครือข่ายสังคมออนไลน์ (Social networking site)	เป็นเว็บไซต์ที่บุคคลหรือหน่วยงานสามารถ สร้าง ข้อมูลและเปลี่ยนข้อมูล (สถานะของ ตน) เผยแพร่ รูปภาพภาพเคลื่อนไหว โดยที่ บุคคลอื่นสามารถ เข้ามาแสดงความชอบ หรือส่งต่อ หรือเผยแพร่ หรือแสดงความ เห็น ได้ตอบการสนทนา หรือ แสดงความ คิดเห็นเพิ่มเติมได้	- เฟซบุ๊ก - บาดู - กูเกิ้ลพลัส - ลิงค์อิน - ออเก็ต
ไมโครบล็อก (Micro blog)	เป็นเว็บไซต์ที่ใช้เผยแพร่ข้อมูลหรือข้อความ สั้น ใน เรื่องที่สนใจเฉพาะด้าน รวมทั้ง สามารถใช้ เครื่องหมาย # (hashtag) เพื่อ เชื่อมต่อกับกลุ่มคน ที่มีความสนใจในเรื่อง เดียวกันได้	- ทวิตเตอร์ - เว่ยป้อ - เทาส์ - ทัมเบอร์
เว็บไซต์ที่ให้บริการ แบ่งปันสื่อ ออนไลน์ (Video and photo sharing website)	เป็นเว็บไซต์ที่ให้ผู้ใช้งานสามารถฝากหรือนำสื่อ ข้อมูล รูปภาพ วิดีโอขึ้นเว็บไซต์เพื่อ แบ่งปัน แก่ผู้อื่น	- ฟิคเกอร์ - วีมิได้ - ยูทูบ - อินสตาแกรม - ฟินเทอร์เรสต์
บล็อก ส่วนบุคคลและองค์กร (Personal and corporate blogs)	เป็นเว็บไซต์ที่ผู้เขียนบันทึกเรื่องราว ต่าง ๆ เสมือน เป็นบันทึกไดอารี่ออนไลน์ สามารถเขียนใน ลักษณะไม่เป็นทางการและแก้ไข ได้บ่อย	- บล็อกเกอร์ - เวิร์ดเพส - บล็อกแก๊งค์ - เอ็กทีน

จากตาราง 2 แพลตฟอร์มโซเชียลมีเดียที่ถูกนำมาใช้เป็นอัลกอริทึมในการสื่อสารแลกเปลี่ยนความคิดเห็นและแบ่งปันเรื่องราวกับสาธารณชนผ่านอินเทอร์เน็ต โดยผู้ให้บริการในปัจจุบันสามารถจำแนกออกเป็น 5 ประเภท ดังต่อไปนี้ (Digimusketeers, 2021; Online)

1. เฟซบุ๊ก (Facebook) เป็นแพลตฟอร์มที่มีผู้ใช้งานมากที่สุดในโลกและเหมาะสำหรับธุรกิจที่ต้องการอัลกอริทึมในการสื่อสารและเชื่อมต่อกับลูกค้าและกลุ่มเป้าหมาย ในขณะที่ทำการสำรวจมีจำนวนผู้ใช้งานเฟซบุ๊กทั่วโลกอยู่ 2.85 พันล้านบัญชี และมีจำนวนผู้ใช้งานในประเทศไทย 49.95 ล้านบัญชี สำหรับในไทยนั้นผู้ใช้เฟซบุ๊กส่วนใหญ่มีอายุระหว่าง 18 – 34 ปี และเพจที่ได้รับความนิยมในบ้านเรามีทั้งเพจที่มีเนื้อหาเป็นโพสต์จำพวกรูปภาพ ไลฟ์วิดีโอ ข่าวและเรื่องราวเกี่ยวกับคนดัง และยังพบว่าโพสต์รูปภาพได้รับการตอบรับจากผู้ชมชาวไทยสูงที่สุด ข้อดีของเฟซบุ๊กในทางธุรกิจคือสมัครใช้งานง่าย สามารถใช้สร้างปฏิสัมพันธ์กับลูกค้าได้หลายรูปแบบ และฟีเจอร์เฟซบุ๊กยังเหมาะสำหรับใช้ขายสินค้าและบริการให้กับผู้บริโภคด้วย

2. ยูทูบ (YouTube) เป็นแพลตฟอร์มที่ได้รับความนิยมเป็นอันดับที่สองของโลกด้วยจำนวนผู้ใช้งาน 2.29 พันล้านบัญชี แต่สำหรับในประเทศไทยนั้น จากสถิติในไตรมาส 3 ของปี 2020 พบว่า ยูทูบขึ้นแท่นเป็นโซเชียลมีเดียอันดับหนึ่ง โดยจำนวนผู้ใช้งานยูทูบชาวไทยคิดเป็น 94.2% ของผู้ใช้งานโซเชียลมีเดียทั้งหมดในประเทศ (รองลงมาคือ เฟซบุ๊ก 93.3% และ ไลน์ 86.2%) ซึ่งวัตถุประสงค์หลักของการใช้งานยูทูบ ในทางธุรกิจ คือการแชร์คอนเทนต์ในรูปแบบวิดีโอเพื่อสร้างภาพลักษณ์ให้กับแบรนด์และเชื่อมต่อกับลูกค้า และยูทูบ ดังนั้นการมีช่องยูทูบจึงช่วยเพิ่มโอกาสในการเข้าถึงแบรนด์ได้เป็นอย่างดี

3. อินสตาแกรม (Instagram) เป็นแพลตฟอร์มที่แชร์รูปภาพที่ได้รับความนิยมมากที่สุดในโลก มีจำนวนผู้ใช้งานทั่วโลก 1.38 พันล้านบัญชี และผู้ใช้โอจีส่วนใหญ่มีอายุต่ำกว่า 35 ปี กลุ่มใหญ่ที่สุดคือช่วงอายุ 25 – 34 ปี รองลงมาคือช่วงอายุ 18 – 24 ปี สำหรับประเทศไทยเองมีผู้ใช้ 18.44 ล้านบัญชี 3 และ 63.3% เป็นเพศหญิง สำหรับนักการตลาดดิจิทัลแล้ว อินสตาแกรมเป็นชุมชนบนโลกโซเชียลที่มีสีสันมาก และมีประโยชน์ในการเข้าถึงตัวผู้บริโภคที่เป็นกลุ่มคนรุ่นใหม่ เพื่อสร้างการรับรู้แบรนด์ผ่านรูปภาพสวย ๆ วิดีโอ สตอรี่ ฯลฯ นอกจากนี้ อินสตาแกรมยังมีส่วนช่วยสร้างความน่าเชื่อถือให้กับธุรกิจ และมีการติดตามผลวิเคราะห์ข้อมูลของผู้ติดตามและโพสต์ได้ด้วย

4. ดิกด็อก (TikTok) เป็นแพลตฟอร์มที่ใช้สำหรับแบ่งปันวิดีโอ ซึ่งเป็นแพลตฟอร์มน้องใหม่ที่ธุรกิจไม่ควรมองข้าม ผู้ใช้งานดิกด็อกทั่วโลกมีทั้งสิ้น 732 ล้านบัญชี ในไทยเมื่อเดือนมกราคม 2021 พบว่ามียอดดาวน์โหลดดิกด็อกเพิ่มขึ้น 44% และผู้ใช้งานในไทยยังใช้เวลาบนแพลตฟอร์มนี้เพิ่มขึ้นถึง 71% คลิปดิกด็อกที่เน้นความบันเทิง และการเล่าเรื่องอย่างสร้างสรรค์ในความยาวไม่เกิน 3 นาทีนี้มีเอกลักษณ์โดนใจวัยรุ่น รวมถึงเหล่าเซเลบคนดังและแบรนด์สินค้าต่าง ๆ เป็นอย่างมาก ทำให้มียอดดาวน์โหลดเกินคานา และประสบความสำเร็จเป็นอย่างมาก ล่าสุดก็ได้ออก ดิกด็อกสำหรับธุรกิจ

5. ทวิตเตอร์ (Twitter) เป็นแพลตฟอร์มที่มีผู้ใช้งานทั่วโลก 397 ล้านบัญชี สำหรับในประเทศไทยมีผู้ใช้ 9.45 ล้านบัญชี แต่ที่สำคัญคือทวิตเตอร์กลายเป็นแหล่งติดตามข้อมูลข่าวสารที่สำคัญ ทั้งยังเป็นกระบอกเสียงแบบเรียลไทม์ที่เติบโตเร็วมาก และมีส่วนผลักดันพฤติกรรม การซื้อออนไลน์ของบริโภคด้วย โดยทางทวิตเตอร์เปิดเผยว่าในช่วงเกือบหนึ่งปีที่ผ่านมาคนไทยมีการใช้แฮชแท็กที่เกี่ยวข้องกับโควิดมากกว่า 73 ล้านทวีต 5 เพื่อแชร์ประสบการณ์และให้การช่วยเหลือผู้อื่น และทวีตข้อความที่เกี่ยวข้องกับสินค้ากลุ่มอุปโภคบริโภคก็ได้รับความนิยมมากโดยเฉพาะในกลุ่มผู้ใช้ที่เป็นผู้หญิง ด้านอาหารเครื่องดื่มและธุรกิจความงามหันมาเลือกใช้ทวิตเตอร์เป็นแพลตฟอร์มเพื่อเชื่อมต่อกับผู้บริโภคกันมากขึ้น

จากแพลตฟอร์มโซเชียลมีเดียที่นำมาใช้เป็นอัลกอริทึมในการสื่อสารแลกเปลี่ยนความคิดเห็น และแบ่งปันเรื่องราวที่กล่าวมาข้างต้น จะพบว่า ทวิตเตอร์

2.2.4 ทวิตเตอร์ (Twitter) มีต้นกำเนิดจากการระดมความคิดที่ถูกจัดขึ้นโดยบริษัทโอดีโอ ซึ่งเป็นบริษัทพอดแคสติง โดยดอร์ซี เป็นนักศึกษาจากมหาวิทยาลัยนิวยอร์ก ได้นำเอาความคิดการส่งบริการข้อความสั้น หรือ เอสเอ็มเอส (SMS) ที่ใช้ในการพูดคุย และสนทนากันภายในกลุ่มเล็ก ๆ ช่วงแรก ๆ โค้ดเนมของบริการนี้มีชื่อ “twtttr” ซึ่งวิลเลียมส์ได้แนะนำให้กับโนอาร์ กลาส โดยเป็นชื่อในรูปแบบนี้ และมีแรงบันดาลใจมาจากฟลิคเกอร์ (Flickr) ให้กำหนดเพียง 5 ตัวอักษรคล้ายกับการส่งข้อความสั้นแบบ นักพัฒนาได้กำหนดหมายเลข “10958” แต่ภายหลังได้เปลี่ยนเป็น “40404” เพื่อความสะดวกในการใช้งาน

คุณลักษณะเด่นของทวิตเตอร์

Zortout (2022) จากรายงานของวีอาร์โซเชียล พบว่า มีผู้ใช้ทวิตเตอร์ในประเทศไทยมากถึง 7.35 ล้านคน ซึ่งเราสามารถใช้อัปโหลดทวีตเพื่ออัปเดตสิ่งที่พบเจอมาในแต่ละวัน ไม่ว่าจะเป็นอาหารที่กิน สถานที่ท่องเที่ยว บ่นเรื่องงาน เรื่องแฟน หรือวิจารณ์เหตุการณ์ต่าง ๆ ที่เกิดขึ้นในบ้านเมืองก็ได้ รวมถึงยังสามารถตอบกลับ ทวิตของคนอื่นก็ได้เช่นเดียวกัน และการเล่นทวิตเตอร์เหมาะกับใครก็สามารถเล่นได้ เพราะนอกจากจะใช้พูดคุยสร้างสัมพันธ์กับคนอื่นแล้ว ยังสามารถใช้พูดคุยกับศิลปินที่คุณชื่นชอบ หรือขอความช่วยเหลือได้ ก็สามารถใช้ประโยชน์จากทวิตเตอร์เพื่อประชาสัมพันธ์สินค้า รวมถึงสร้างความสัมพันธ์กับลูกค้าได้ ซึ่งทวิตเตอร์จะมีคุณลักษณะเด่นที่สำคัญดังนี้ สุปัญญา อภิวงศ์โสภณ (2563) ได้กล่าวว่า

1. ทวิตเตอร์เน้นการแสดงข้อความและมีรูปภาพแนบขยายความ
2. การใช้งานของทวิตเตอร์ ผู้ใช้บางคนมีการใช้งานเสมือนเป็นการประโยคบอกเล่า ไม่มีระบุ เป็นลักษณะของการบอกกล่าวอย่างไรเหตุผล การใช้คำพูดพรวดพราด การแสดงความคิดเห็น หรือการเขียนข้อความใด ๆ แบบไม่เป็นทางการ

3. ไม่มีระบบเพื่อน กลุ่ม หรือหน้าเว็บเพจ จึงไม่มีระบบคัดกรองคนดู ข้อความที่มีการแสดงออกมาทั้งหมด จะถูกตั้งค่าการมองเห็นเป็นแบบสาธารณะทั้งหมด ใครๆ ก็สามารถดูข้อความได้ทั้งหมด ไม่ว่าจะเป็นผู้ใช้งานที่มีบัญชีผู้ใช้งานหรือไม่ และจะเป็นผู้ติดตามบัญชีนั้นหรือไม่ก็ตาม

4. ระบบการติดตาม เมื่อผู้ใช้ติดตาม (Follow) บัญชีที่ตนเองสนใจแล้ว บัญชีผู้ใช้ที่ต้องการติดตามได้ตั้งค่าความเป็นส่วนเป็นแบบสาธารณะ (Public) เมื่อผู้ใช้ติดตามบัญชีนั้นแล้วจะเห็นข้อความของบัญชีนั้นในส่วนการรับข่าวสารได้ทั้งหมด

5. กรณีสรางบัญชีแบบปิด ซึ่งเป็นบัญชีเฉพาะที่สามารถกำหนดให้บุคคลที่อนุญาตให้ติดตามเท่านั้น จึงจะสามารถดูข้อความที่แสดงได้ทั้งหมด

6. ทุกคนมีสิทธิ์สมัครบัญชีผู้ใช้งานทวิตเตอร์ได้มากกว่า 1 บัญชี ดังนั้นแต่ละผู้ใช้จึงสามารถเปิดหลายบัญชีแยกกันสำหรับใช้งานตามวัตถุประสงค์ส่วนตัวได้

7. ผู้ใช้ทวิตเตอร์เน้นการอ่านข้อความ ซึ่งอาจเป็นข้อความหรือข้อความที่มีคนกล่าว ถึง หากหัวข้อข่าว (Topic) เป็นที่น่าสนใจ จะมีการใช้แฮชแท็ก (# Hashtag) ปริมาณมากจนขึ้นเป็น เทรนด์

8. การค้นหาเรื่องที่สนใจในทวิตเตอร์สามารถค้นหาผ่านแฮชแท็ก หากต้องการให้ข้อความไปอยู่ในหมวดใด สามารถเพิ่มแฮชแท็กได้ เพื่อให้คนใจในเรื่องเดียวกันสามารถค้นหาข้อความที่มีการแสดงออกไป

Daily Gizmo (2022) กล่าวว่า ทวิตเตอร์เป็นแพลตฟอร์มโซเชียลมีเดียที่หลายคนรู้จัก และมักเป็นที่นิยมของกลุ่มวัยรุ่นเป็นอย่างมาก เช่น การตามติดไอดอลเกาหลีจนบางทีขึ้นติดเทรนด์ทวิตเตอร์ให้ใครๆ ก็ต้องตามไปดูเช่นกัน แต่ก็มีคนมักตั้งคำถามว่าทำไมผู้คนถึงหันไปสนใจกับทวิตเตอร์เป็นอย่างมาก และทวิตเตอร์มีคุณลักษณะที่เด่น ๆ หลายประการ ดังนี้

1. รวดเร็ว ทันเหตุการณ์ผ่านแฮชแท็ก เป็นสิ่งที่เป็เอกลักษณ์ของทวิตเตอร์หนีไม่พ้น เรื่องของแฮชแท็กที่มีมาก่อนและไม่เหมือนใคร ทำให้ความรวดเร็วในเรื่องที่เป็นกระแสติดเทรนด์เป็นอันดับต้น ๆ ได้ที่สำคัญเป็นข้อดีที่ทำให้เรารู้ได้ว่าในแต่ละวันเกิดเหตุการณ์อะไรขึ้นบ้าง มีต้นตอที่มาอย่างไร หรือใครที่ต้องการติดตามเรื่องราวต่างๆ อย่างละเอียด ก็สามารถตามดูได้ในแฮชแท็กนั้นๆ ทันที

2. ข้อความสั้น กระชับ ไม่เว้นเว้อ เนื่องด้วยทวิตเตอร์เป็นแพลตฟอร์มที่สามารถโพสต์ข้อความได้ในจำนวนจำกัดเพียงแค่ 280 ตัวอักษรเท่านั้น จึงทำให้ข้อความต่างๆ ที่โพสต์ในทวิตเตอร์เป็นข้อความที่สั้น กระชับ เข้าใจง่าย ตอบสนองพฤติกรรมคนรุ่นใหม่ที่ไม่ต้องการอ่านข้อมูลเยอะๆ แต่อยากรู้เรื่องราวที่มาเป็นอย่างไร ตั้งแต่ต้นจนจบได้ทีเดียวทีเดียว

3. ค้นหาสิ่งที่อยากรู้ได้ผ่านคีย์เวิร์ดสั้น ๆ แฮชแท็กเป็นสิ่งที่ทำให้ผู้ใช้งานสามารถค้นหาสิ่งที่ต้องการทราบได้เป็นอย่างดี เช่นเดียวกับการค้นหาสิ่งที่เราอยากรู้ผ่านคีย์เวิร์ดสั้นๆ เช่น “รถติด

ลาดพร้าว” ก็จะมีข้อความต่างๆ ที่เกี่ยวข้องปรากฏขึ้นมา ทั้งจากผู้คนทั่วไปที่ใช้แพลตฟอร์ม และสื่อต่างๆ

4. แบ่งปันสิ่งที่มีประโยชน์ได้ ทวิตเตอร์ ถือเป็นแพลตฟอร์มที่ใครก็สามารถเป็นสื่อมวลชนได้ รายงานข่าวได้ รายงานสิ่งที่พบเจอได้ เนื่องด้วยทวิตเตอร์เป็นแพลตฟอร์มที่สามารถโพสต์ข้อความได้เพียง 280 ตัวอักษรเท่านั้น ดังนั้นการแชร์เรื่องราวสั้นๆ ที่บ่งบอกว่า “ใครทำอะไรที่ไหน อย่างไร หรือเกิดอะไรขึ้นที่ไหนเมื่อไหร่” ผ่านทวิตเตอร์ก็ถือเป็นเรื่องราวที่ดี ที่ทุกคนสามารถแบ่งปันให้ผู้อื่นได้ อย่างเช่น รถติดจุดไหน ควรเลี่ยงเส้นทางใด หรือ เกิดเหตุอะไรที่ทำให้จราจรติดขัด เป็นต้น เผลอๆ บางทีสำนักข่าวก็เอาข้อมูลที่แชร์ๆ เนี่ยแหละไปเขียนข่าวเพิ่ม

5. แสดงความคิดเห็นได้อย่างอิสระ เป็นการแสดงความคิดเห็นที่แตกต่าง แต่ก็สามารถแสดงออกไปได้ตามความรู้สึกของผู้ใช้ทวิตเตอร์ แต่บางครั้งผู้ใช้ควรต้องดูแลตัวเอง ว่าเป็นการแสดงความคิดเห็นลักษณะไหนที่เหมาะสม และลักษณะไหนที่ไม่เหมาะสม

6. ติดตามคนที่มีชื่อเสียงได้อย่างใกล้ชิด ทวิตเตอร์มีข้อดีที่แตกต่างจากโซเชียลมีเดียในเรื่องของแอคเคาท์ เช่น ในเฟซบุ๊กอาจจะไม่ได้บ่งบอกว่า แอคเคาท์ไหนที่เป็นตัวจริง ซึ่งต่างกับทวิตเตอร์ที่ผู้ใช้งานจริงๆ จะเป็นเจ้าตัว ไม่ว่าจะเป็นดารานักแสดง หรือ นักการเมือง คนมีชื่อเสียงในต่างประเทศ ทำให้สามารถติดตามบุคคลเหล่านี้ได้อย่างใกล้ชิด และเผลอตกทนายอะไรไปอาจจะได้รับการตอบรับกลับมา

ประโยชน์ของทวิตเตอร์

1. แพลตฟอร์มโซเชียลมีเดีย มีอิทธิพลต่อไลฟ์สไตล์ของผู้คน
2. ส่งผลต่อการตัดสินใจของผู้คนจนกลายเป็นอัลกอริทึมที่มีอิทธิพลโดยตรงต่อการรับรู้ถึงเหตุการณ์ต่าง ๆ
3. ผู้คนใช้ทวิตเตอร์ในสร้างทวิตหรือรีทวีตด้วยการแพร่กระจายข่าวปลอมข่าวลือ หรือ ข้อมูลต่าง ๆ
4. ทวิตเตอร์เป็นแพลตฟอร์มที่คนทั่วโลกนิยมใช้ในการปั่นกระแสด้านต่าง ๆ
5. ทวิตเตอร์เป็นแพลตฟอร์มที่สามารถตอบกลับทวิตของผู้อื่นได้อย่างอิสระ
6. ทวิตเตอร์สามารถรายงานสถานการณ์สดของข่าวสาร หรือเหตุการณ์ที่พบเจอมาในแต่ละวันได้

จากที่กล่าวมาข้างต้น สามารถสรุปได้ว่า สื่อสังคมออนไลน์ในปัจจุบันที่อยู่ในรูปแบบแพลตฟอร์มที่ส่งข้อมูล ข่าวสารต่าง ๆ ผ่านทางอินเทอร์เน็ต โดยสามารถส่งเรื่องราว รูปภาพ และวิดีโอ เชื่อมโยงกับความสนใจและกิจกรรมของผู้อื่น แบ่งปันให้ผู้อื่นที่อยู่ในเครือข่ายของตนได้รับรู้ในเรื่องดังกล่าว นั้น ทวิตเตอร์จึงเป็นอีกแพลตฟอร์มหนึ่งที่มีความนิยมเป็นอย่างมากในหมู่ผู้คน เพราะทวิตเตอร์สามารถส่งข้อมูลแบบเรียลไทม์ได้ และสามารถส่งข้อมูลแบบนาที่ต่อนาที่ การส่ง

ข้อมูลหรือข่าวปลอมจึงไม่แปลกที่แพลตฟอร์มทวิตเตอร์จึงกลายเป็นอัลกอริทึมในการแพร่กระจายข่าวปลอมได้อย่างรวดเร็วในปัจจุบัน

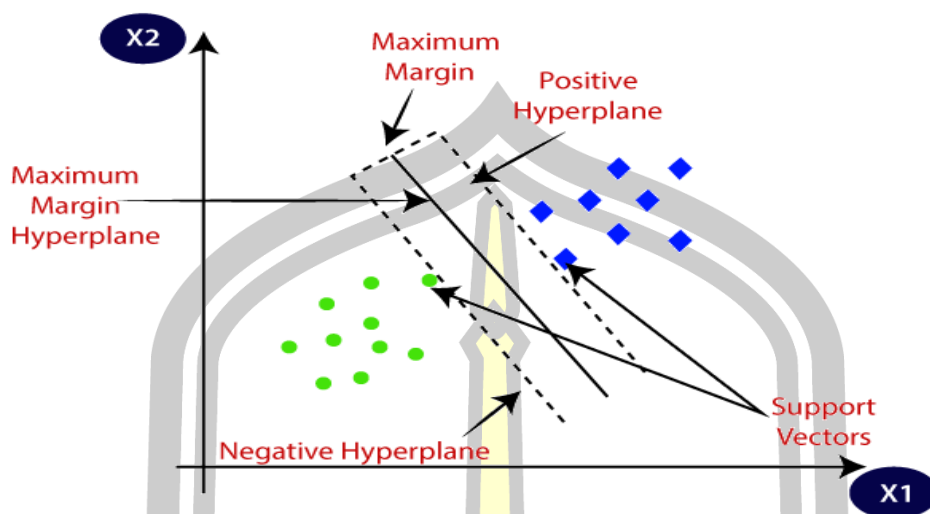
2.3 ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine)

ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine) หรือ SVM เป็นอีกหนึ่งอัลกอริทึมในกลุ่มวิธีการเรียนรู้ด้วยเครื่องแบบมีผู้สอน ที่สามารถนำมาช่วยแก้ปัญหาการจำแนกข้อมูลได้ โดยเฉพาะกับปัญหาที่มีขนาดของข้อมูลไม่ใหญ่มาก แต่คุณลักษณะ (Features) ของข้อมูลมีจำนวนมาก SVM จะถือได้ว่าเป็นอัลกอริทึมที่ทำงานได้ง่าย ค่อนข้างมีประสิทธิภาพมาก หลักการทำงานของ SVM จะอาศัยการใช้การสร้างเส้นแบ่ง หรือไฮเปอร์เพลน (Hyperplane) ในการแบ่งแยกคลาส ของข้อมูลออกจากกัน จากนั้นจะทำการหาไฮเปอร์เพลน จากนั้นจะทำการหาไฮเปอร์เพลนใดเป็นเส้นที่ใช้แยกคลาสของข้อมูลได้ดีที่สุด โดยจะใช้ฟังก์ชันแมปข้อมูลจาก Input space ไปยัง Feature space และสร้างฟังก์ชันวัดความคล้ายที่เรียกว่า เคอร์เนล ฟังก์ชัน (Kernel function) บน Feature space เหมาะใช้สำหรับข้อมูลที่มีลักษณะมิติของข้อมูลที่มีปริมาณมาก โดยกำหนดให้ $(x_i, y_i), \dots, (x_n, y_n)$ เป็นตัวอย่างที่ใช้สำหรับการสอน n คือ จำนวนข้อมูลตัวอย่าง m คือ จำนวนมิติข้อมูลเข้า และ y คือ ผลลัพธ์มีค่า $+1$ หรือ -1 ดังสมการ $(x_i, y_i), \dots, (x_n, y_n)$ เมื่อ $x \in \mathbb{R}^m, y \in \{-1, +1\}$

สำหรับปัญหาเชิงเส้น มิติข้อมูลขนาดสูงได้ถูกแบ่งเป็น 2 กลุ่ม โดยระบอบตัดสินใจซึ่งคำนวณได้ดังสมการ และเมื่อ w คือ น้ำหนักและ b คือค่า bias สมการใช้สำหรับการจำแนกประเภทของข้อมูล $(w \cdot x) + b > 0$ ถ้า $y_i = -1$ และ $(w \cdot x) + b < 0$ ถ้า $y_i = 1$

อย่างไรก็ตาม ซัพพอร์ตเวกเตอร์แมชชีน มีเคอร์เนล ฟังก์ชัน (Kernel Function) ผู้ที่สามารถประยุกต์ใช้ในการแก้ปัญหาได้หลายวิธี โดยผู้วิจัยต้องเลือกเคอร์เนลให้เหมาะสมกับลักษณะข้อมูล สามารถแสดงข้อมูลดังภาพประกอบ 7

พูน ปณ จิต ชีเว



ภาพประกอบ 7 ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines)

ที่มา : Easa et al., 2023a

วิวัฒน์ กิตตินารพรและคณะ (2566) ได้ทำการการประยุกต์ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคโครงข่ายประสาทเทียมในการพยากรณ์ปริมาณงานก่อสร้างในประเทศไทย ปัจจัยที่เลือกนำมาใช้ในการพัฒนาแบบพยากรณ์ซัพพอร์ตเวกเตอร์แมชชีน และโครงข่ายประสาทเทียมเพื่อพยากรณ์ปริมาณการก่อสร้างอื่น ๆ ของรัฐบาลในประเทศไทยประกอบด้วย 4 ปัจจัย ดังนี้ T35 ค่า Disposal of Current Revenue, T72 ค่า Private, T109 ค่า Non-residential และ T122 ค่า Total C.i.f. Value ตัวแบบพยากรณ์ที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนใช้เคอร์เนลฟังก์ชันแบบ Radial Basis Function (RBF) ทำการแมปข้อมูลและตัวแบบพยากรณ์ที่ใช้เทคนิคโครงข่ายประสาทเทียมใช้ฟังก์ชันกระตุ้นชนิดเกาส์เซียน (Gaussian function) ในการแปลงค่า ใช้วิธีการตรวจสอบความถูกต้องด้วยข้อมูลแบบไขว้กันโดยแบ่งข้อมูลเป็น 10 กลุ่ม (10 fold cross validation) ทำการสอน และทดสอบข้อมูลเพื่อวัดประสิทธิภาพของแบบจำลอง ผลลัพธ์ของแบบจำลองจะถูกระบุในรูปแบบของ Mean Error Percentage Error (MAPE) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ผลลัพธ์ของค่า MAPE คือ 1.31 และค่าสหสัมพันธ์ยกกำลังสอง (Coefficient of determination, R^2) เท่ากับ 0.998 ส่วนเทคนิคโครงข่ายประสาทเทียม ผลลัพธ์ของค่า MAPE คือ 6.41 ค่าสัมประสิทธิ์และค่าสหสัมพันธ์ยกกำลังสอง (Coefficient of determination, R^2) เท่ากับ 0.982 มีค่าเข้าใกล้ 1 แสดงว่าแบบจำลองมีความเหมาะสม จากผลการวิจัยแสดงให้เห็นอย่างชัดเจนว่าสามารถใช้เทคนิคการสร้างแบบจำลองซัพพอร์ตเวกเตอร์แมชชีน และโครงข่ายประสาทเทียมในการพยากรณ์อนุกรมเวลาของปริมาณการก่อสร้างเมื่อเปรียบเทียบกับ การประมาณงานก่อสร้างของสำนักงานคณะกรรมการ และการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ พบว่าได้ผลใกล้เคียงกัน มีประโยชน์สำหรับองค์กร

เนื่องจากกระบวนการพยากรณ์ค่อนข้างง่ายเป็นสิ่งจำเป็นสำหรับการวางแผนการตัดสินใจและการจัดการ

2.3.1 เทคนิคการใช้อัลกอริทึม

เทคนิคกระบวนการเรียนรู้ของเครื่องเป็นที่ยอมรับในด้านการประยุกต์ใช้งานกับการแก้ไขปัญหาหลายด้าน หรือใช้สำหรับโปรแกรมประยุกต์มากมาย โดยแต่ละกระบวนการเรียนรู้มีความเหมาะสมกับข้อมูลรูปแบบเฉพาะ กระบวนการเรียนรู้ด้วยเครื่องที่แก้ปัญหด้วยการสอนโดยใช้ข้อมูลที่มีคำตอบที่แท้จริงของปัญหาด้วยการสอนโดยใช้ข้อมูลที่มีคำตอบที่แท้จริงของปัญหา กระบวนการเรียนรู้ของเครื่องโดยทั่วไปแล้วจะเป็นลักษณะของการสนับสนุนด้านการแนะนำ (provide a recommendation) หรือการพยากรณ์เพื่อปรับปรุงแก้ไข (Predictive Maintenance) การเพิ่มประสิทธิภาพในการถือหุ้น (Portfolio optimization) การทำงานอัตโนมัติ และอื่น ๆ ที่เกี่ยวข้อง การเรียนรู้ของเครื่อง เรียกอีกอย่างหนึ่งว่า การเรียนรู้แบบมีผู้สอน (Supervised Learning) ดังนั้น ผู้วิจัยได้สรุปงานวิจัยที่เกี่ยวข้องพร้อมสรุปเทคนิคอัลกอริทึมที่นำมาใช้ในงานวิจัย สามารถแสดงข้อมูลดังตาราง 3

ตาราง 3 ตัวอย่างงานวิจัยที่นำมาเทคนิคมาเพื่อใช้ในงานวิจัยระบุข่าวปลอม

ปีที่พิมพ์	ผู้วิจัย	ชื่อวิจัย	เทคนิค
2023	M. Park and S. Chai	Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques	VM, RF, LR, CART, and NNET,
2022	P. Jain, S. Sharma, Monica and P. K. Aggarwal,	Classifying Fake News Detection Using SVM, Naive Bayes and LSTM,	SVM, Naive Bayes and LSTM
2022	N. L. Siva Rama Krishna and M. Adimoolam	Fake News Detection System using Logistic Regression and Compare Textual Property with Support Vector Machine Algorithm	LR SVM
2022	N. L. S. R. Krishna and M. Adimoolam	Fake News Detection system using Decision Tree algorithm and compare textual property with Support Vector Machine algorithm	DT,SVM

ตาราง 2 ตัวอย่างงานวิจัยที่นำมาเทคนิคมาเพื่อใช้ในการวิจัยระบุข่าวปลอม (ต่อ)

ปีที่พิมพ์	ผู้วิจัย	ชื่อวิจัย	เทคนิค
2021	Nagaraja A, N S, Naik P. et al.,	Fake News Detection Using Machine Learning Methods	NB SVM
2020	Vogel I, Meghana M.,	Detecting fake news spreaders on twitter from a multilingual perspective	CNN SVM
2019	Hiramath, Chaitra K. Deshpande, G. C.,et al.	Fake News Detection Using Deep Learning Techniques	LR, NB RF, DNN SVM
2018	Aphiwongsophon S, & Chongstitvatana P,	Detecting Fake News with Machine Learning Method	NB NN SVM
2019	Monti F, Frasca F, Eynard D et al.,	Fake News Detection on Social Media using Geometric Learning	CNN Deep
2018	Aldwairi M, Alwahedi A.,	Detecting Fake News in Social Media Networks	NB
2021	H. Saleh, A. Alharbi and S. H. Alsamhi,	OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection	RNN, LSTM,DT LR, KNN, RF, SVM, NB
2021	Sansonetti G, Gasparetti F, D'Aniello G. et al.,	Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection	NN, SVC, SVM- SGD, KNN

จากตาราง 3 จะพบว่าเทคนิคที่นิยมนำมาประยุกต์ใช้ในการงานที่เกี่ยวกับการระบุข่าวปลอม ประกอบด้วย นาอ์ฟเบย์ (Naïve Bayes) โครงข่ายประสาทแบบคอนโวลูชัน (โครงข่ายประสาทแบบคอนโวลูชัน) ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine) วิธีโครงข่ายประสาท (Neural network) เป็นต้น

งานวิจัยส่วนใหญ่ที่ผู้วิจัยได้ศึกษาค้นคว้าที่เกี่ยวข้องกับการตรวจสอบข่าวปลอม จะเป็นงานวิจัยในภาษาต่างประเทศ เช่น ภาษาอังกฤษ ภาษาอาหรับ ภาษาฮินดู โดยงานวิจัยดังกล่าวถึงมีข้างต้นจะมีลักษณะเป็นการวิเคราะห์ข่าวเพื่อจำแนกข่าวปลอม หรือเป็นการวิเคราะห์เนื้อหาข่าวในการวิเคราะห์อารมณ์ความรู้สึกที่ได้จากการอ่านข่าวหรือเนื้อหาข่าว เมื่อมีการเปลี่ยนแปลงเนื้อหาข่าวเป็นภาษาอื่น ๆ บริบททางด้านภาษา ไวยากรณ์หรือความหมายของคำในแต่ละภาษาจะเปลี่ยนแปลง ทำให้ไม่สามารถใช้ผลการวิเคราะห์ที่ได้กับเนื้อหาข่าวในภาษาที่เปลี่ยนแปลงไป (Zubiaga et al., 2018), (Nagaraja et al., 2021), (Ehsanfar & Mansouri, 2017), (Praseed et al., 2023) จากที่กล่าวมาแล้วข้างต้นผู้วิจัยได้ทำการศึกษาและการเปรียบเทียบวิธีการนำอัลกอริทึมมาใช้ในการตรวจสอบข่าวปลอมสามารถแสดงได้ดังตาราง 4

ตาราง 4 การเปรียบเทียบวิธีการนำอัลกอริทึมมาใช้ในการตรวจสอบข่าวปลอม

ชื่อผู้วิจัย	วิธีการ	จุดเด่น	จุดด้อย
Li and Lei (2022)	1. Recurrent neural network; RNN 2. Long short term memory network; LSTM 3. Graph neural networks; GNN	เป็นวิธีการใช้งานที่ง่ายที่สุดในการจำแนกข้อมูล RNN และ LSTM เป็นตัวแยกคุณลักษณะที่สามารถเรียนรู้ข้อมูล	Deep Neural Networks; CNN สามารถจัดการได้เฉพาะข้อมูลโครงสร้างกริดเท่านั้น
Tiemtud and Leesutthipornchai (2023)	Support Vector Machine, Naive Bayes, and Neural Network.	SVM ประสบความสำเร็จและมีความแม่นยำ และมีประสิทธิภาพความแม่นยำของโมเดล	ต้องใช้ชุดข้อมูลที่มีความน่าเชื่อถือ ซึ่งมีจะซับซ้อน ใช้เวลานาน เสียค่าใช้จ่ายสูงในการจัดหา หรือไม่พร้อมใช้งาน
Easa et al (2023)	Support vector machine (SVM)	SVM ทำงานได้และมีประสิทธิภาพในการให้ผลลัพธ์ที่ดีมาก	เนื้อหาของข้อมูลจะต้องมีความครบถ้วน

ตาราง 4 การเปรียบเทียบวิธีการนำอัลกอริทึมมาใช้ในการตรวจสอบข่าวปลอม (ต่อ)

ชื่อผู้วิจัย	วิธีการ	จุดเด่น	จุดด้อย
Jing et al (2023)	1.CNN 2. SVM-TS 3. GRU 4. Text GCN	เทคนิคการเรียนรู้เชิงลึกมีความสามารถในการเรียนรู้การเป็นตัวแทนที่ทรงพลัง และมีประสิทธิภาพ	1) การแยกคุณสมบัติออกจากข้อความด้วยตนเองนั้นไม่เพียงแต่ยากและใช้เวลานานในการออกแบบ 2) ไม่สามารถใช้ประโยชน์จากเนื้อหาของข้อความได้อย่างเต็มที่
Khanam et al (2021)	1. LR 2. RF 3. NB 4. KNN 5. SVM 6. Decision Tree; DT 7. Extreme Gradient Boosting; XGBoot,	การแยกข่าวปลอมมีความแม่นยำสูง	ผลลัพธ์ไม่ถูกต้องหากแหล่งที่มาของข่าวไม่ถูกต้อง
Sansonetti et al (2020)	1. NN 2. SVM using stochastic gradient descent; SVM-SGD 3. K-Nearest Neighbour Algorithm; KNN	เป็นวิธีการตรวจสอบความแม่นยำและเป็นโมเดลที่ดีที่สุด	ต้องการข้อมูลที่มีความสมบูรณ์มากที่สุด
Ahmad et al (2020)	1. K-NN 2. SVM 3. LR 4. LSVM 5. DT 6. SGD	อัลกอริทึมที่นำมาใช้นี้มีความแม่นยำสูง	การนำมาใช้ในการค้นหาแบบกริดในพารามิเตอร์มีค่าใช้จ่ายสูง

ตาราง 4 การเปรียบเทียบวิธีการนำอัลกอริทึมมาใช้ในการตรวจสอบข้าวปลอม (ต่อ)

ชื่อผู้วิจัย	วิธีการ	จุดเด่น	จุดด้อย
Aphiwongsophon (2018)	1. Naïve bayes; NB 2. Neural network; NN 3. Support vector machine; SVM	สามารถตรวจสอบข้าวปลอมด้วยความแม่นยำ	ข้อมูลที่ไม่รู้จักจำนวนมากแบบจำลองจะไม่สามารถจำแนกข้อมูลได้อย่างถูกต้อง หากไม่มีข้อมูลการ เทรินนิ่งในจำนวนมากเพียงพอ

จากตาราง 4 การเปรียบเทียบวิธีการนำอัลกอริทึมมาใช้ในการตรวจสอบข้าวปลอมที่กล่าวถึงนั้น ได้มีนักวิจัยจำนวนมากมุ่งเน้นไปที่การศึกษาค้นหาเทคนิค หรือวิธีการแก้ปัญหาสำหรับการตรวจสอบข้าวปลอมมากมาย และจะเห็นว่าข้าวปลอมนั้นมีความหลากหลายมากขึ้นการแก้ปัญหาของข้าวปลอมจึงทำให้เกิดความต้องการที่สูงขึ้นและมีความท้าทายมากขึ้นด้วย ทั้งนี้จะพบว่า การตรวจจับข้าวปลอมนิยมใช้โมเดลการเรียนรู้ของเครื่อง และที่ได้รับความนิยม ได้แก่ Support Vector Machine, Naive Bayes , Decision tree , Random forest , Logistic regression ซึ่งเป็นเทคนิคที่นำมาใช้ในการตรวจสอบข้าวปลอมที่มีประสิทธิภาพและความเสถียรที่ดียิ่งขึ้นของการฝึกโมเดล

ดังนั้นความสำคัญของงานวิจัยชิ้นนี้ ผู้วิจัยจึงมีแนวคิดที่จะศึกษาอัลกอริทึมที่มีอยู่เดิมในงานวิจัยที่เกี่ยวข้อง พัฒนาอัลกอริทึมใหม่เพื่อตรวจจับข้าวปลอม เปรียบเทียบประสิทธิภาพของอัลกอริทึมและคัดเลือกอัลกอริทึมที่เหมาะสมในการตรวจจับข้าวปลอม พร้อมทั้งนำเสนออัลกอริทึมใหม่ประยุกต์ใช้กับฐานข้อมูลที่มีความหลากหลายเพื่อปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม

ซึ่งสอดคล้องกับงานวิจัยของ Meesad (2021) ได้กล่าวถึงเทคนิคที่นำมาประยุกต์ใช้ในการระบุข้าวปลอม ซึ่งงานวิจัยดังกล่าวอยู่ในฐานข้อมูล SN Computer Science และได้สังเคราะห์เทคนิคที่นำมาใช้ในการระบุข้าวปลอมสรุปได้ตามผลสอดคล้องกับอัลกอริทึมทั้ง 4 ประเภทข้างต้นสามารถแสดงข้อมูลดังตาราง 5

ตาราง 5 การเปรียบเทียบการวัดประสิทธิภาพของอัลกอริทึมกับงานวิจัยที่เกี่ยวข้อง

ชื่องานวิจัย	อัลกอริทึม	ค่าความถูกต้อง (Accuracy)	ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่าความถ่วงดุล (F – Measure)
Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection	NN SVC SVM-SGD KNN	91.47% 90.02% 89.26% 81.54%	-	-	-
Detecting fake news spreaders on twitter from a multilingual perspective	CNN SVM	78% 78%	-	-	-
Detecting Fake News with Machine Learning Method	NB NN SVM	96.08% - 99.90% 99.90%	-	-	-
Fake News Detection Using Deep Learning Techniques	LR RF SVM NB DNN	75% 77% 79% 89% 91%	-	-	-
Fake News Detection Using Machine Learning Methods	NB SVM	63% 75%	68% 74%	63% 74%	-

จากตาราง 5 พบว่าผลงานวิจัยของ (Billones et al., 2022), (Krishna & Adimoolam, 2022) มีค่าความถูกต้องมากที่สุดเมื่อเทียบกับอีกหกผลงาน ซึ่งสอดคล้องกับงานวิจัยของ (Kiahwar & Zafar, 2022) ได้ทำการวิจัยเรื่องการตรวจจับข่าวปลอมในข่าวปากีสถานโดยใช้แมชชีนเลิร์นนิง

และการเรียนรู้เชิงลึก (ธนาภัทร ภัทรวิจิ, 2563) ด้วยอัลกอริทึมได้แก่ Support voter machine, Logistic Regression, Decision Trees, Naïve Bayes และ Random forest

จากการศึกษาการทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้องทั้งทางด้านและแมชชีนเลิร์นนิงที่ได้กลายเป็นความต้องการในเกือบทุกสาขาหรือทุกศาสตร์ ทำให้การเรียนรู้และใช้งานเพื่อตอบวัตถุประสงค์อย่างมากมายจนกลายเป็นเรื่องธรรมดาและมีอัลกอริทึมต่าง ๆ มากมายที่มีบทบาทเป็นอย่างมากที่ส่งผลต่อประสิทธิภาพในการระบุข่าวปลอม ซึ่งอัลกอริทึมที่มีประสิทธิภาพในการตรวจสอบข่าวปลอมนั้น ประกอบด้วย 5 อัลกอริทึมด้วยกัน คือ นาอ์ฟเบย์ (Naïve bayes), ต้นไม้ตัดสินใจ (Decision tree), ต้นไม้แห่งการทนาย (Random forest), การถดถอยโลจิสติก (Logistic Regression) และ ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines)

2.4 การเตรียมข้อมูล (Data preprocessing)

การเตรียมข้อมูลเป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้จากการเก็บรวบรวมมาให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ โดยการกระบวนการอาจเกี่ยวข้องกับการทำเหมืองข้อมูล (Data mining) และ โมเดลการทำนายของ Bayes (Bayesian predictive models) (Capuano et al., 2023), (Pal et al., 2023) เครื่องคอมพิวเตอร์จะรับข้อมูลเข้ามาและใช้อัลกอริทึมเพื่อหาคำตอบขึ้น และ เทคโนโลยีดิจิทัลที่ทันสมัยเพิ่มความสะดวกแก่ผู้ใช้งานในการเผยแพร่ข่าวสารบนสื่อสังคมออนไลน์ โดยส่วนหนึ่งได้เผยแพร่ข่าวปลอมที่มีจำนวนมากซึ่งเป็นปัญหาที่ทำให้ผู้คนหลงเชื่อข่าวปลอมว่าเป็นข่าวจริงและนำไปแชร์ต่อทำให้เกิดปัญหาอื่น ๆ ต่อเนื่องตามมา และปัจจุบันมีนักวิจัยได้นำขั้นตอนมาช่วยในสนับสนุนการตรวจสอบการระบุข่าวปลอม ดังนี้

2.4.1 กระบวนการประมวลผลข้อมูล (Data processing) เป็นเทคโนโลยีการประมวลผลซึ่งเป็นกระบวนการหนึ่ง ที่นำมาใช้ในการจัดการกับข้อมูลจำนวนมากให้อยู่ในรูปของข้อมูลที่ต้องการประกอบด้วย

1. กระบวนการทำความสะอาดข้อมูล (Cleansing data) (Rahm et al., 2012) เป็นกระบวนการทำความสะอาดข้อมูล การตรวจสอบและการแก้ไขรายการข้อมูลที่ไม่ถูกต้องออกจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล การทำความสะอาดข้อมูลเกิดขึ้นเนื่องจากมีความมาสอดคล้องของข้อมูล ที่เกิดขึ้นจากข้อผิดพลาดของการบันทึกข้อมูล การส่งข้อมูลหรือการให้ความหมายของข้อมูล ที่จัดเก็บแตกต่างกัน จึงต้องมีการแทนที่ ปรับปรุง หรือการลบข้อมูลที่ไม่ถูกต้องนี้ออกไป เพื่อให้ข้อมูลมีคุณภาพมากขึ้น และเหมาะสมกับการนำข้อมูลดังกล่าวไปใช้

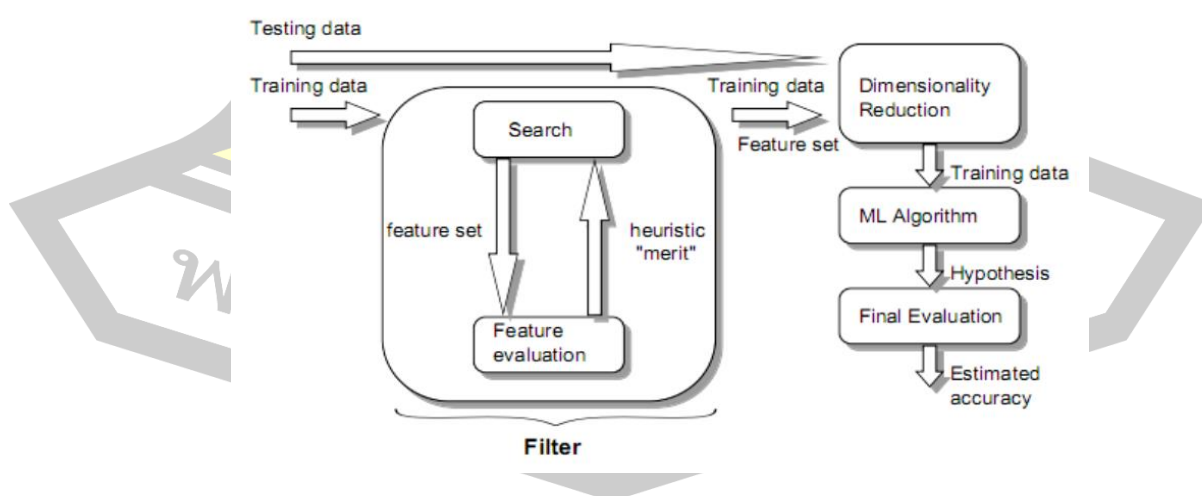
2.4.2 การคัดเลือกคุณลักษณะ (Feature selection)

การคัดเลือกคุณสมบัติ (Feature selection) เป็นขั้นตอนการสำคัญในการเตรียมข้อมูลก่อนการทำเหมืองข้อมูล เป็นเทคนิคการลดขนาดของมิติ ลดจำนวนคุณลักษณะ การคัดเลือกคุณลักษณะที่มีความสำคัญต่อการพยากรณ์เหมืองข้อมูล และอาจจะกล่าวได้ว่าประสิทธิภาพของการจำแนกประเภทเหมืองข้อมูลขึ้นอยู่กับคุณลักษณะ (Feature) ที่นำมาใช้ ซึ่งนอกจากจะช่วยลดระยะเวลาในการสร้างตัวแบบพยากรณ์ให้เร็วขึ้นแล้ว ยังช่วยตัดคุณลักษณะที่ไม่จำเป็นต่อการสร้างแบบพยากรณ์ได้อีกด้วย การคัดเลือกคุณลักษณะเป็นการเลือกเซตย่อยของคุณลักษณะที่มีขนาดเล็กที่สุดและดีที่สุด โดยสอดคล้องกับเงื่อนไขดังนี้

1. ความแม่นยำของการจำแนกประเภทจะไม่ลดลงอย่างมีนัยสำคัญ
2. การกระจายของกลุ่ม (Class distribution) เมื่อใช้เฉพาะคุณลักษณะที่ถูกเลือกมีลักษณะใกล้เคียงกับการกระจายของกลุ่มเริ่มต้นเมื่อมีคุณลักษณะครบ

อัจฉิมา มณฑาพันธ์ (2562) การคัดเลือกคุณลักษณะเป็นสิ่งที่มีความสำคัญ เป็นการคัดเลือกเซตของคุณลักษณะเฉพาะใหม่จากเซตของคุณลักษณะเดิม โดยจะเป็นเซตของเซตคุณลักษณะเดิม เพื่อเพิ่มประสิทธิภาพในการทำนาย และการทำงานให้เร็วขึ้น การคัดเลือกคุณลักษณะที่สำคัญสามารถแบ่งออกเป็น 2 ประเภท ดังนี้

1. Filter Approach เป็นวิธีที่พยายามคัดเลือกคุณลักษณะที่สำคัญ โดยการคำนวณค่าน้ำหนักหรือค่าความสัมพันธ์ของแต่ละคุณลักษณะ โดยการเลือกคุณลักษณะที่มีความสำคัญเก็บเอาไว้ เช่น เทคนิค Correlation based feature selection (CFS), เทคนิค Information gain (IG), เทคนิค Gain ratio (GR) และ เทคนิค Chi square สามารถแสดงข้อมูลดังภาพประกอบ 8



ภาพประกอบ 8 การคัดเลือกคุณลักษณะแบบ Filter approach

ที่มา : Mark A. Hall (1999)

จากภาพประกอบ 8 การคัดเลือกคุณลักษณะแบบ Filter approach ประกอบด้วยเทคนิคดังต่อไปนี้

1.1 เทคนิค Correlation based feature selection (CFS) เป็นเทคนิคในการเลือกคุณสมบัติของคุณลักษณะ โดยใช้การพิจารณาบนพื้นฐานความสัมพันธ์ ของกลุ่มคุณลักษณะที่ได้จากการประเมินค่า จากความสามารถในการคาดการณ์ คุณลักษณะที่ถูกคัดเลือกใช้สำหรับการจำแนกประเภทข้อมูล และยังสามารถจัดการกับคุณลักษณะที่ไม่เกี่ยวข้อง CFS จะจัดอันดับกลุ่มย่อยของมิติข้อมูล ทำการคัดเลือกกลุ่มย่อยของมิติข้อมูลที่มีความสัมพันธ์สูงกับคลาส และไม่มีความสัมพันธ์กับคลาสอื่น ๆ สำหรับมิติข้อมูลที่ไม่เกี่ยวข้องหรือมีความสัมพันธ์ต่ำกับคลาสจะถูกลบทิ้ง มิติข้อมูลที่ซ้ำซ้อนจะถูกจัดออกไปจากกลุ่มมิติข้อมูลที่มีความสัมพันธ์สูง สมการประเมินกลุ่มย่อยของมิติข้อมูลแบบ CFS1 สามารถแสดงในสมการที่ 1 ภัทรารุณิ แสงศิริ (2553)

$$M_s = \frac{\overline{kr_{cf}}}{\sqrt{k+k(k-1)r_{ff}}} \quad (1)$$

โดยที่

M_s คือ ค่าที่ค้นหาได้ของมิติข้อมูลกลุ่มย่อย S ซึ่งประกอบด้วยมิติข้อมูล K

$\overline{r_{ff}}$ คือ ค่าเฉลี่ยความสัมพันธ์ของตัวแปรกับคลาส ($f \in s$)

$\overline{r_{cf}}$ คือ ค่าเฉลี่ยความสัมพันธ์ระหว่างมิติข้อมูล

1.2 เทคนิค Information gain (IG) เป็นเทคนิคที่ใช้ในการชี้วัดเพื่อเลือกคุณลักษณะข้อมูลที่ต้องการที่น้อยที่สุดที่เหมาะสมในการนำไปใช้ระบุ โดยการคำนวณค่า Gain สำหรับแต่ละมิติข้อมูล ซึ่งมิติข้อมูลใดมีค่า Gain สูงสุดจะถูกคัดเลือกเพื่อนำมาใช้ระบุ สามารถแสดงในสมการที่ 2 แสดงการคำนวณค่า IG หรือค่า Entropy ของชุดข้อมูลทั้งหมด สามารถแสดงในสมการที่ 3 แสดงการคำนวณค่า Entropy ของชุดข้อมูลในแต่ละคุณลักษณะ ดังสมการที่ 4 เป็นการคำนวณค่า IG สำหรับการพิจารณามิติข้อมูลคุณลักษณะ A

$$E(D) = -\sum_i^n = 1P_i \log_2(P_i) \quad (2)$$

$$E_A(D) = -\sum_{j=1}^m \frac{|D_j|}{D} \times E(D_j) \quad (3)$$

$$\text{Gain}(A) = E(D) - E_A(D) \quad (4)$$

โดยที่ P_i คือ ค่าความน่าจะเป็นที่เรคคอร์ดหนึ่ง ๆ จะมีหมวดหมู่ของข้อมูล

หากการคำนวณค่า IG คือการวัดค่า Entropy ก่อนที่จะมีการแบ่งข้อมูลออกตามมิติข้อมูลและหลังการแบ่งว่ามีประสิทธิภาพดีขึ้นหรือไม่ ถ้ามีประสิทธิภาพดีขึ้น ค่า IG จะมีค่าสูงขึ้น

1.3 เทคนิค Gain ratio (GR) เป็นตัวชี้วัดการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลย่อยที่พัฒนามาจาก IG เนื่องจากการใช้ IG ในการแบ่งชุดข้อมูลจะมีโอกาสทำให้เกิดความเอนเอียงขึ้น เมื่อคุณลักษณะที่ทำการพิจารณาได้ค่า gain ที่สูงเป็นจำนวนมาก ทำให้คุณลักษณะที่ถูกคัดเลือกไม่ถูกต้อง จากความเอนเอียง ทำให้มีการพัฒนาตัวชี้วัดการแบ่งข้อมูลใหม่ที่เรียกว่า Gain ratio โดยการประยุกต์ใช้การคำนวณอัตราส่วนค่า Information gain ด้วยการใช้ค่า Split information ซึ่งสามารถแสดงในสมการที่ 5

$$\text{SplitInfo}_A(D) = -\sum_{j=1}^m \frac{|D_j|}{|D|} \times \log_2 \frac{|D_j|}{|D|} \quad (5)$$

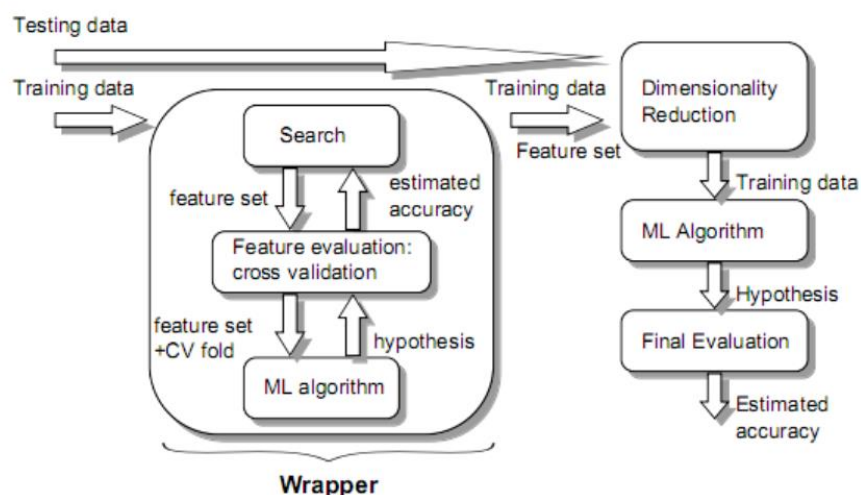
โดยค่า $\text{SplitInfo}_A(D)$ หมายถึง ปริมาณข้อมูลที่ถูกพิจารณาโดยการแบ่งข้อมูลในชุดข้อมูล D ออกเป็น m ชุดข้อมูลย่อยตามค่าคุณลักษณะ A โดยหลังจากการทำการคำนวณหาค่า $\text{SplitInfo}_A(D)$ แล้ว จะสามารถคำนวณหาค่า GR ได้ สามารถแสดงในสมการที่ 6

$$\text{Gain Ratio}(A) = \frac{\text{Gain}(A)}{\text{SplitInfo}_A(A)} \quad (6)$$

1.4 เทคนิค Chi-Square เป็นเทคนิค Chi-Square (χ^2) เป็นการประเมินค่าของคุณลักษณะโดยใช้การคำนวณค่า Chi-Square ทางสถิติ เพื่อศึกษาว่าการแจกแจงความถี่ของตัวแปรคุณลักษณะเป็นไปตามรูปแบบที่กำหนดไว้หรือไม่ สามารถแสดงในสมการที่ 7

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (7)$$

2. Wrapper approach เป็นวิธีการคัดเลือกคุณลักษณะที่สำคัญ โดยการคำนวณค่าน้ำหนักการวัดค่าความถูกต้องในการแบ่งกลุ่มข้อมูล มาสร้างเซตของคุณลักษณะใหม่ โดยการเพิ่มหรือลดจำนวนคุณลักษณะจากเซตเดิม เช่น เทคนิค Forward selection, เทคนิค Backward Elimination และเทคนิค Evolutionary selection สามารถแสดงข้อมูลดังภาพประกอบ 9



ภาพประกอบ 9 การคัดเลือกคุณลักษณะแบบ Wrapper approach

ที่มา : Mark A. Hall (1999)

จากภาพประกอบ 9 การคัดเลือกคุณลักษณะแบบ Wrapper approach ประกอบด้วยเทคนิค ดังต่อไปนี้

2.1 เทคนิค Forward selection เป็นเทคนิคที่ใช้วิธีการคัดเลือกคุณลักษณะซึ่งเป็นตัวแปรอิสระเข้ามาในสมการทีละตัว ขั้นแรกพิจารณาจากค่าสัมบูรณ์ของสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ระหว่างตัวแปรตามกับตัวแปรอิสระที่ให้ค่าสูงสุด หากตัวแปรอิสระตัวนั้นมีคุณสมบัติตามเกณฑ์การนำเข้า ก็จะถูกนำเข้าสมการ หลังจากนั้นเป็นการพิจารณาสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรตามกับตัวแปรอิสระที่ไม่อยู่ในสมการถดถอยทีละตัว ตัวแปรอิสระที่มีค่าสัมบูรณ์ของสัมประสิทธิ์สหสัมพันธ์บางส่วนสูงสุดจะถูกนำเข้าสมการ หากตัวแปรนั้นมีคุณสมบัติตามเกณฑ์การนำเข้า ขั้นตอนจะทำซ้ำจนกระทั่งพบว่าไม่สามารถนำตัวแปรอิสระเข้าสมการได้จึงหยุด

2.2 เทคนิค Backward elimination เป็นเทคนิคที่พยายามคัดเลือกตัวแปรที่ดีที่สุด และได้โมเดลประหยัดในการพยากรณ์เช่นเดียวกัน โดยในขั้นตอนแรกจะนำตัวแปรพยากรณ์ทุกตัวเข้ามาในสมการและดำเนินการพิจารณาตัวแปรพยากรณ์ที่มีค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนกับตัวแปรเกนซ์ และควบคุมอิทธิพลของตัวแปรพยากรณ์อื่นๆ ซึ่งมีค่าต่ำที่สุดออกจากการสมการ แล้วจึงดำเนินการทดสอบสอบว่า ค่า R^2 ลดลงอย่างมีนัยสำคัญทางสถิติหรือไม่ ถ้าพบว่าลดลงอย่างไม่มีนัยสำคัญทางสถิติแสดงว่าตัวแปรนั้นไม่ได้มีส่วนทำให้การพยากรณ์ตัวแปรเกนซ์เพิ่มขึ้นเลย

2.3 เทคนิค Evolutionary selection เป็นเทคนิคการคัดเลือกคุณลักษณะโดยการนำเข้าข้อมูลคุณลักษณะมาหาประสิทธิภาพในการคาดการณ์คำตอบ จากนั้นนำคุณลักษณะมาจับคู่แล้วหาค่าประสิทธิภาพอีกครั้ง หากค่าประสิทธิภาพในการคาดการณ์สูงขึ้นจะเก็บข้อมูลนั้นไว้ แล้วทำการเลือกคุณลักษณะอื่นเข้าไปเพิ่ม หากค่าประสิทธิภาพต่ำจะถอดคุณลักษณะนั้นออก แล้วทำการเลือกคุณลักษณะอื่นเข้าไป (เอกสิทธิ์ พชรวงศ์ศักดิ์, 2557)

2.5 การวัดประสิทธิภาพของโมเดล (Evaluation measure)

การวัดประสิทธิภาพของโมเดล ถือเป็นปัจจัยหนึ่งที่ใช้ในการพิจารณาผลการดำเนินงาน ซึ่งนำมาประเมินความสามารถของระบบประเภทนั้น เน้นความสามารถในการตัดสินใจ หรือจำแนกหมวดหมู่ที่ถูกต้องของวิธีการเพื่อสร้างแบบจำลอง วิธีการวัดประสิทธิภาพของแบบจำลองสามารถวัดค่าประสิทธิภาพ ประกอบด้วย 6 ประเภท ดังนี้ ค่าความถูกต้อง (Accuracy), ค่าความแม่นยำ (Precision), ค่าความระลึก (Recall), ค่าความถ่วงดุล (F – Measure), ค่าความไว (Sensitivity) และ ค่าจำเพาะ (Specificity) ซึ่งการวัดประเมินผลมีความสำคัญ เนื่องจากนำมาใช้ในการวิเคราะห์ประสิทธิภาพการทำงานของอัลกอริทึม ตัวชี้วัดในการประเมินผลสามารถคำนวณได้จากเมทริกซ์ความสับสน (Confusion matrix) สามารถแสดงข้อมูลดังภาพประกอบ 10 ดังนี้ (Rohera et al., 2022), (Meesad, 2021), (เพชรรัตน์ ม่วงน้อยและคณะ, 2563), (Altheneyan & Alhadlaq, 2023)

พหุ ประสิทธิภาพ

		ผลการจำแนก	
		คำตอบเป็นบวก (P)	คำตอบเป็นลบ (N)
ค่าที่แท้จริง	คำตอบเป็นบวก (P)	True positives (TP)	False negative (FN)
	คำตอบเป็นลบ (N)	False positives (FP)	True negative (TN)

ภาพประกอบ 10 ตารางเมทริกซ์ความสับสน (Confusion matrix)

ที่มา : สุรวัชร ศรีเปารยะและสายชล สันสมบูรณ์ทอง (2560)

โดย True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่า “จริง” และ มีค่าเป็น “จริง”

True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่า “ไม่จริง” และ มีค่า “ไม่จริง”

False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่า “จริง” แต่ มีค่าเป็น “ไม่จริง”

False Negative (FN) คือ สิ่งที่โปรแกรมทำนายว่า “ไม่จริง” แต่ มีค่าเป็น “จริง”

สามารถใช้ Confusion Matrix มาคำนวณการประเมินประสิทธิภาพของการทำนายด้วย Model ของเราในรูปแบบค่าต่าง ๆ ได้ดังนี้

1. ค่าความถูกต้อง (Accuracy) คือ เป็นการวัดค่าของความถูกต้องของโมเดล จำนวนของข้อมูลที่ทำนายถูกทุกคลาส โดยพิจารณาทุกคลาส สามารถแสดงในสมการที่ 8

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

2. ค่าความแม่นยำ (Precision) คือ เป็นการวัดค่าความแม่นยำของโมเดล ที่ดูจากสิ่งทำนายออกมาแล้วทายถูกได้กี่เปอร์เซ็นต์ สามารถแสดงในสมการที่ 9

$$\text{Precision} = \frac{TP}{TP + FN} \quad (9)$$

3. ค่าความระลึก (Recall) คือ เป็นการวัดค่าความระลึกของโมเดล โดยพิจารณาจำนวนที่ทำนายถูกก็ตัว เป็นการวัดความถูกต้องของโมเดล สามารถแสดงในสมการที่ 10

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

4. ค่าความถ่วงดุล (F – Measure) คือ เป็นการวัดค่าของ Precision และ Recall พร้อมกันของโมเดล จากนั้นพิจารณาค่าเฉลี่ยของค่าความแม่นยำและค่าระลึก สามารถแสดงในสมการที่ 11

$$\text{F-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (11)$$

2.6 งานวิจัยที่เกี่ยวข้อง

สุนันทา แยมทัพ และอรัญญา ศิริพล (2563) กล่าวถึง ข่าวดปลอมกับการทำงานของนักข่าวในยุคดิจิทัลที่มีการเสนอกระบวนการตรวจสอบข่าวปลอมให้แก่ักข่าวและผู้ใช้สื่อ ซึ่งลักษณะของข่าวปลอมที่เกิดขึ้นในชุมชนออนไลน์ปัจจุบันมีอย่างน้อย 3 รูปแบบ คือ (1) โฆษณาประชาสัมพันธ์ให้ดูเหมือนข่าวจริง (2) การประชดประชันเสียดสีล้อเลียนที่เนื้อหาผิดไปจากข้อเท็จจริง (3) ตั้งใจสร้างข่าวเพื่อหวังผลทางการเมืองทำให้คนเกิดอารมณ์ร่วมแบ่งฝ่าย ทั้ง 3 รูปแบบนี้มีมิติร่วมในการผลิตข่าวปลอม คือ การไม่นำเสนอข้อเท็จจริงอย่างถึงที่สุด (Facticity) และการลวง (Deception) เพื่อบรรลุเป้าหมายของตน

นันทิกา หนูสม และวิโรจน์ สุทธิสีมา (2562) กล่าวถึง ลักษณะของข่าวปลอมในประเทศไทย และระดับความรู้เท่าทันข่าวปลอมบนเฟซบุ๊กของผู้รับสารในเขตกรุงเทพมหานคร มีการจำแนกและรูปแบบเนื้อหาข่าวปลอมของผู้รับสาร โดย 1) ประเภทของข่าวปลอมที่พบมากที่สุดระหว่าง การเก็บข้อมูลคือข่าวปลอมที่ถูก สร้างขึ้นอย่างสมบูร์น (Bogus) และรูปแบบเนื้อหาข่าวปลอมที่พบมากที่สุดคือ เนื้อหาที่ประดิษฐ์ขึ้น (Fabricated content) โดยมีเจตนาในการสร้างขึ้นมาเพื่อให้เป็นข่าวปลอมที่สมบูร์น แบนเนียนที่สุด มี เนื้อหาที่ประดิษฐ์ขึ้นใหม่ทั้งหมด ทั้งภาพ พาดหัวข่าว เนื้อหา และข้อมูลที่ปรากฏในข่าว สร้างขึ้นมาเป็นข่าว ปลอมทั้งสิ้น 2) ระดับความรู้เท่าทันข่าวปลอมของผู้รับสารในเขตกรุงเทพมหานครมีค่าเฉลี่ยอยู่ในระดับปาน กลาง โดยกลุ่มตัวอย่างมีความสามารถในการรู้เท่าทันสื่อในระดับการเข้าถึง การวิเคราะห์ และการประเมิน ในระดับมาก ส่วนในระดับการสร้างสรค่นั้นกลุ่มตัวอย่างมีความสามารถในระดับน้อยที่สุด 3) ส่วนประกอบ ของข่าวที่ทำให้กลุ่มตัวอย่างที่เชื่อถือในระดับมาก คือข่าวที่มีแหล่งข่าวเป็นบุคคลที่มีชื่อเสียง ส่วน องค์ประกอบอื่นๆ บน

สื่อสังคมออนไลน์ที่ทำให้เชื่อถือข่าวนั้นในระดับมาก เพราะเป็นข่าวที่ถูกนำเสนอบน หน้านิวส์ฟีดของ เพชบุรีบ่อยๆและผู้รับสารเชื่อว่าที่สามารถหาเหตุผลมาสนับสนุนความเชื่อของตนเอง

สุปัญญา อภิวงศ์โสภณ (2561) กล่าวถึงการตรวจสอบข่าวปลอมด้วยวิธีการเรียนรู้ด้วยเครื่อง และนำเสนอวิธีการตรวจจับข่าวบนเครือข่ายสังคมออนไลน์ทวิตเตอร์ด้วยวิธีการเรียนรู้ด้วยเครื่อง โดยใช้การเรียนรู้ด้วยเครื่อง 3 วิธี ได้แก่ Naïve bayes, Neural network และ Support vector machine ซึ่งทั้ง 3 วิธีนี้ สามารถตรวจจับข่าวปลอมในชุดข้อมูลได้อย่างถูกต้อง ทั้งนี้ร้อยละความ ถูกต้องของวิธี Naïve Bayes คือ 96.08 เปอร์เซ็นต์ Neural network 99.89 เปอร์เซ็นต์ และ Support vector machine 99.89 เปอร์เซ็นต์

พยุ่ง มีสัจและคณะ (2565) กล่าวถึง เทคนิคทางปัญญาประดิษฐ์สำหรับการตรวจจับข่าวปลอม ภาษาไทย ซึ่งเสนอวิธีการตรวจสอบข่าวปลอมภาษาไทยด้วยเทคนิคปัญหาประติรูป โดยใช้ 3 เทคนิค หลักในการสร้างแบบจำลองการตรวจจับข่าวปลอมได้แก่ การสืบค้นข้อมูลสารสนเทศ การ ประมวลผลภาษาธรรมชาติและการเรียนรู้ของเครื่อง ลักษณะของการสืบค้นข้อมูลสารสนเทศเป็น กลไกในการดึงข้อมูลจากเว็บไซต์ข่าวออนไลน์และโซเชียลมีเดีย การประมวลผลภาษาธรรมชาติจะ วิเคราะห์ข่าวที่ได้กลับคืนมาเพื่อสกัดข้อมูลที่มีคุณลักษณะและมีความโดดเด่นเป็นอย่างดี การเรียนรู้ ของเครื่องทำการรับข้อมูลฟีดเจอร์และจัดประเภทบทความ ข่าว ออกเป็น 3 ประเภทได้แก่ ข่าวจริง ข่าวปลอม และข่าวน่าสงสัย

Kiahwar and Zafar, (2022) กล่าวถึง การศึกษาวิธีการตรวจจับข่าวปลอมในข่าวปากีสถาน โดยใช้การเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก ซึ่งเป็นการใช้วิธีแมชชีนเลิร์นนิงสำหรับการจำแนก บทความข่าวโดยอัตโนมัติ การศึกษาของเราสำรวจคุณสมบัติที่เป็นข้อความต่างๆ ที่สามารถใช้ แยกแยะเนื้อหาปลอมออกจากของจริงได้ ด้วยการใช้คุณสมบัติเหล่านั้น เราฝึกอัลกอริทึมการเรียนรู้ ของเครื่องต่างๆ โดยใช้วิธีการทั้งมวลต่างๆ และประเมินประสิทธิภาพในชุดข้อมูลในโลกแห่งความ เป็นจริง 4 ชุด การประเมินเชิงทดลองยืนยันประสิทธิภาพที่เหนือกว่าของแนวทางผู้เรียนทั้งมวลที่เรา เสนอเมื่อเปรียบเทียบกับผู้เรียนแต่ละคน

Cantarella et al (2023) กล่าวถึง ข่าวปลอมส่งผลต่อพฤติกรรมกรรมการลงคะแนนเสียงหรือไม่ ซึ่งเป็นการใช้ข้อมูลเพื่อเปรียบเทียบผลกระทบของการเปิดเผยข่าวปลอมต่อการลงคะแนนให้พรรค ประชาชนิยมในการเลือกตั้งปี 2556 และ 2561 ซึ่งข้อบ่งชี้ใหม่ของประชาชนนิยมโดยใช้การขุดข้อความ บนโพสต์ Facebook ของพรรคิิตาก่อนการเลือกตั้ง ผลการวิจัยของเราสนับสนุนมุมมองที่ว่า การ สัมผัสกับข่าวปลอมสนับสนุนพรรคประชาชนนิยมโดยไม่คำนึงถึงการสนับสนุนล่วงหน้าสำหรับพรรค ประชาชนิยม แต่ยังรวมถึงข่าวปลอมเพียงอย่างเดียวไม่สามารถอธิบายการเติบโตของประชาชนนิยมส่วน ใหญ่

Bazmi et al., (2023) กล่าวถึง การศึกษาเครือข่ายความสนใจร่วมหลายมุมมองสำหรับการตรวจจับข่าวปลอมโดยการสร้างแบบจำลองผู้ใช้เฉพาะหัวข้อและความน่าเชื่อถือของแหล่งข่าว ซึ่งผลการศึกษาเป็นวิธีการขอเสนอ Multi-View Co-Attention Network (MVCAN) ที่ร่วมกันสร้างแบบจำลองความน่าเชื่อถือเฉพาะหัวข้อแฝงของผู้ใช้และแหล่งข่าวสำหรับการตรวจจับข่าวปลอม แนวคิดหลักคือการแสดงบทความข่าว ผู้ใช้ และแหล่งข่าวในลักษณะที่มุมมองเฉพาะของบทความข่าว อดีต SC ของผู้ใช้ที่กำหนดมุมมองของผู้ใช้ในการแบ่งปันข่าว และอดีตของพรรคพวกของแหล่งข่าวถูกเข้ารหัสเป็นเวกเตอร์ จากนั้นจึงเสนอตัวแปรใหม่ของกลไก Multi-Head Co-Attention (MHCA) เพื่อเข้ารหัสการโต้ตอบร่วมกันจากมุมมองที่แตกต่างกัน รวมถึง news-source และ news-user เพื่อจำลองความน่าเชื่อถือของผู้ใช้และแหล่งข่าวโดยปริยายโดยพิจารณาจากการโต้ตอบของพวกเขาในข่าวจริงและข่าวปลอมที่เผยแพร่ในหัวข้อข่าว เราทำการทดลองอย่างกว้างขวางกับชุดข้อมูลสาธารณะสองชุด ผลลัพธ์แสดงให้เห็นว่า MVCAN มีประสิทธิภาพเหนือกว่าวิธีการ state-of-the-art อื่น ๆ อย่างมีนัยสำคัญและมีประสิทธิภาพเหนือกว่าเส้นฐานที่ดีที่สุดโดยเฉลี่ย 3% ในแง่ของ F1 และความแม่นยำ

Jing et al (2023) กล่าวถึง การตรวจจับข่าวปลอมต่อเนื่องหลายรูปแบบผ่านเครือข่ายพิวชันโปรเกรสซีฟ การตรวจจับการบิดเบือนข้อมูลต่อเนื่องหลายรูปแบบ ซึ่งรวบรวมข้อมูลการแสดงของแต่ละกิริยาในระดับที่แตกต่างกัน และบรรลุการหลอมรวมระหว่างกิริยาในระดับเดียวกันและในระดับต่างๆ โดยใช้มิกเซอร์เพื่อสร้างการเชื่อมต่อที่แข็งแกร่งระหว่างกิริยา โดยเฉพาะอย่างยิ่ง เราใช้โครงสร้างหม้อแปลงซึ่งมีประสิทธิภาพในงานวิทัศน์ของคอมพิวเตอร์ เป็นตัวแยกคุณลักษณะภาพเพื่อค่อยๆ สุ่มตัวอย่างคุณสมบัติในระดับต่างๆ และรวมคุณสมบัติที่ได้จากตัวแยกคุณลักษณะข้อความและข้อมูลโดเมนความถี่ของภาพในระดับต่างๆ สำหรับการสร้างแบบจำลองแบบละเอียด นอกจากนี้เรายังออกแบบแนวทางการหลอมรวมคุณลักษณะเพื่อสร้างการเชื่อมต่อระหว่างรูปแบบต่างๆ ได้ดีขึ้นซึ่งสามารถปรับปรุงประสิทธิภาพให้ดียิ่งขึ้นและเหนือกว่าโครงสร้างเครือข่ายอื่นๆ ในวรรณคดี เราทำการทดลองอย่างกว้างขวางกับชุดข้อมูลจริงสองชุดคือ Weibo และ Twitter ซึ่งวิธีการของเรามีความแม่นยำ 83.3% ในชุดข้อมูล Twitter ซึ่งเพิ่มขึ้นอย่างน้อย 4.3% เมื่อเทียบกับวิธีการ state-of-the-art อื่นๆ สิ่งนี้แสดงให้เห็นถึงประสิทธิภาพของ MPFN ในการระบุข่าวปลอม และวิธีการนี้ถึงระดับที่ค่อนข้างสูงโดยการรวมระดับข้อมูลที่แตกต่างกันจากแต่ละกิริยาและวิธีการหลอมรวมกิริยาที่ทรงพลัง

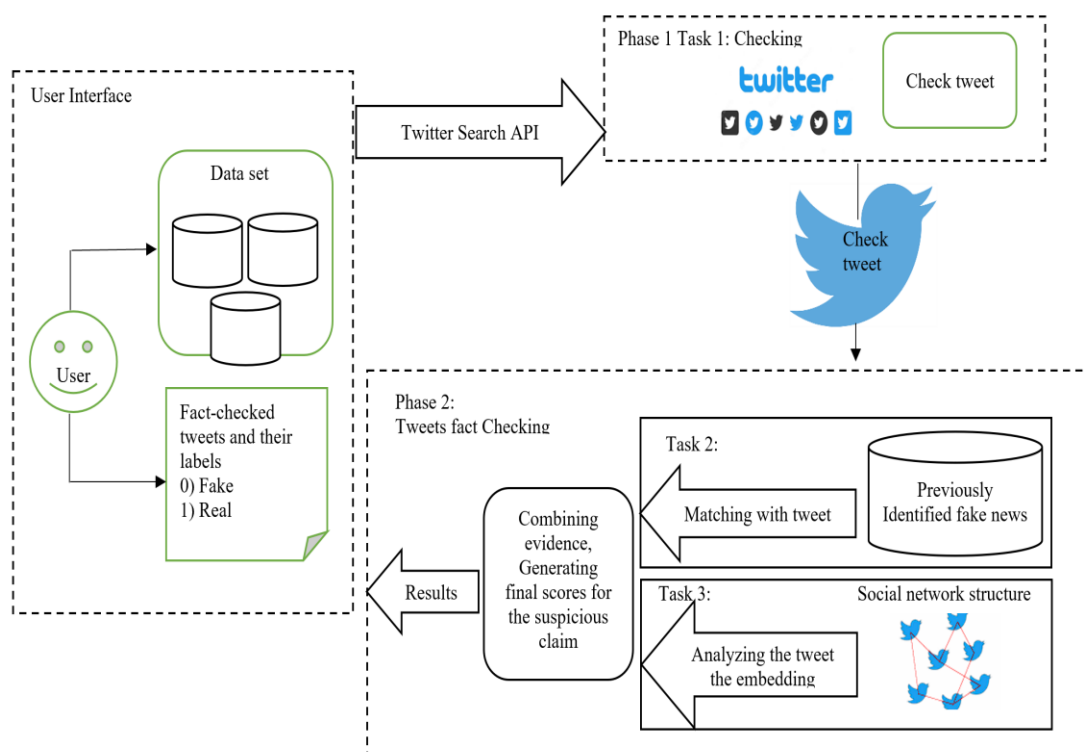
Sreedhar et al (2022) กล่าวถึง เทคนิคที่ได้รับการปรับปรุงเพื่อระบุข่าวปลอมบนเครือข่ายโซเชียลมีเดียโดยใช้แนวคิดการเรียนรู้ของเครื่องภายใต้การดูแล ซึ่งเสนอระบบแยกข่าวโซเชียลมีเดียที่ได้รับการปรับปรุงซึ่งเรียกว่าระบบ Fake news detection (UFND) ที่ไม่มีโครงสร้าง และมีจุดมุ่งหมายเพื่อระบุข้อมูลข่าวโซเชียลมีเดียที่ไม่มีโครงสร้างซึ่งเป็นของชั้นเรียนปลอมหรือจริงตามความน่าจะเป็นและได้รับการปรับปรุง UFND ที่เสนอประกอบด้วยสี่ขั้นตอนที่มีแนวโน้มว่า Pre-

processing การฝึกอบรม การจับคู่ และการตรวจสอบตามลำดับ เทคนิคนี้ระบุสิ่งที่ตรงกันบนองค์ประกอบข้อมูลแต่ละรายการตามแบบจำลองคำสำคัญที่กำหนด และคำที่ไม่เกี่ยวข้องในเอกสารระบบที่เสนอได้ฝึกอบรมชุดข้อมูล Pre-processed ผ่านกระบวนการแยกชุดข้อมูลออกเป็นสองคลาสคือของปลอมและของจริงตามเทคนิคความน่าจะเป็น และระบบได้ระบุเอกสารข่าวการทดสอบที่กำหนดเป็นของป้ายกำกับชั้นเรียนที่มีอยู่เหนือชุดข้อมูลการฝึกอบรมตามเทคนิค Naive Bayes ที่ได้รับการปรับปรุง ระบบ UFND ประเมินประสิทธิภาพมากกว่าผลลัพธ์ของขั้นตอนก่อนหน้า ผลการทดลองของ UFND ได้แสดงให้เห็นถึงผลลัพธ์ที่มีประสิทธิภาพเหนือกว่าในการแยกและระบุรูปแบบข่าวปลอมเหนือชุดข้อมูลข่าวโซเชียลมีเดียที่ไม่มีโครงสร้างด้วยความแม่นยำที่ดีตามวิธีการความน่าจะเป็นภายใต้การดูแล

Khalil et al (2021) กล่าวถึง การตรวจจับข่าวปลอมภาษาอาหรับโดยใช้แมชชีนเลิร์นนิง มีการพัฒนาโมเดลแมชชีนเลิร์นนิง (ML) เพื่อตรวจจับเนื้อหาปลอมและเตือนผู้อ่าน ซึ่งมีชุดข้อมูลข่าวปลอมภาษาอาหรับจำนวนจำกัดในแง่ของบทความและแหล่งข่าว คลังข่าวปลอมภาษาอาหรับขนาดใหญ่แห่งแรกซึ่งประกอบด้วยบทความ 606912 บทความที่รวบรวมจากแหล่งข่าวออนไลน์ภาษาอาหรับ 134 แห่ง แพลตฟอร์ม Fact-check ของภาษาอาหรับใช้เพื่อใส่คำอธิบายประกอบแหล่งข่าวว่าน่าเชื่อถือ ไม่น่าเชื่อถือ และตัดสินใจไม่ได้ นอกจากนี้ยังมีการใช้อัลกอริทึม ML ที่แตกต่างกันสำหรับงานตรวจจับ การทดลองแสดงให้เห็นว่าโมเดลการเรียนรู้เชิงลึกทำงานได้ดีกว่าโมเดล ML แบบเดิม การฝึกอบรมแบบจำลองแสดงให้เห็นปัญหาที่ไม่เหมาะสมและไม่เหมาะสมซึ่งบ่งชี้ว่าคลังข้อมูลมีทำทนาย

Baarir and Djefal (2021) ได้นำเสนอวิธีการตรวจจับข่าวปลอมโดยใช้เครื่องเวกเตอร์สนับสนุน โดยพยายามกำหนดคุณสมบัติและเทคนิคที่ดีที่สุดในการตรวจจับข่าวปลอม โดยเริ่มต้นด้วยการศึกษาด้านข่าวปลอม ผลกระทบ และวิธีการตรวจจับ จากนั้นได้ทำการออกแบบและปรับใช้โซลูชันที่ใช้ชุดข้อมูลข่าวที่ผ่านการประมวลผลล่วงหน้าโดยใช้เทคนิคการทำความสะอาด การสตีม การเข้ารหัส N-gram ถูกรวบรวม และ TF-IDF เพื่อแยกชุดคุณลักษณะที่ช่วยให้ตรวจจับข่าวปลอมได้ และใช้อัลกอริทึม Support Vector Machine ในชุดข้อมูลคุณลักษณะของเราเพื่อสร้างแบบจำลองที่ช่วยให้สามารถจำแนกข้อมูลใหม่ได้

ผู้วิจัยได้ศึกษาและวิเคราะห์ขั้นตอนจากวรรณกรรมที่เกี่ยวข้องและทำการสรุปกรอบแนวคิดเพื่อใช้ในการปรับปรุงการตรวจจับข่าวปลอมด้วยซอฟต์แวร์แมชชีนโดยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม ซึ่งมีลำดับขั้นตอนสำหรับการเก็บรวบรวมข้อมูล และนำเข้าข้อมูลก่อนเข้าสู่การประมวลผลด้วยเทคนิคการเรียนรู้ของเครื่อง และทำการเปรียบเทียบประสิทธิภาพของแต่ละเทคนิค ประกอบด้วยขั้นตอนต่าง ๆ สามารถแสดงข้อมูลดังภาพประกอบ 11



ภาพประกอบ 11 กรอบแนวคิดของการวิจัย

จากประกอบ 11 สามารถอธิบายขั้นตอนของกรอบแนวคิดเป็นการนำเสนอการตรวจจับข่าวปลอมแบบสองขั้นตอน (Two-Phased Fake News Detection Framework หรือ FNDF) ที่สามารถระบุข้อมูลที่ไม่เป็นความจริงบนทวีตเตอร์ได้อย่างมีประสิทธิภาพ ซึ่งประกอบด้วย

ขั้นตอนที่ 1 การประมวลผลข้อมูลล่วงหน้า

กระบวนการที่เกี่ยวข้องกับตรวจสอบข้อมูล การทำความสะอาดข้อมูล หรือจัดการค่าที่ว่าง และทำการเลือกพารามิเตอร์ที่จำเป็นในการตรวจสอบเนื้อหาของข่าวในทวีต ขั้นตอนนี้มีเป้าหมายเพื่อเพิ่มความแม่นยำในการระบุข่าวปลอมจากเนื้อหาในทวีต

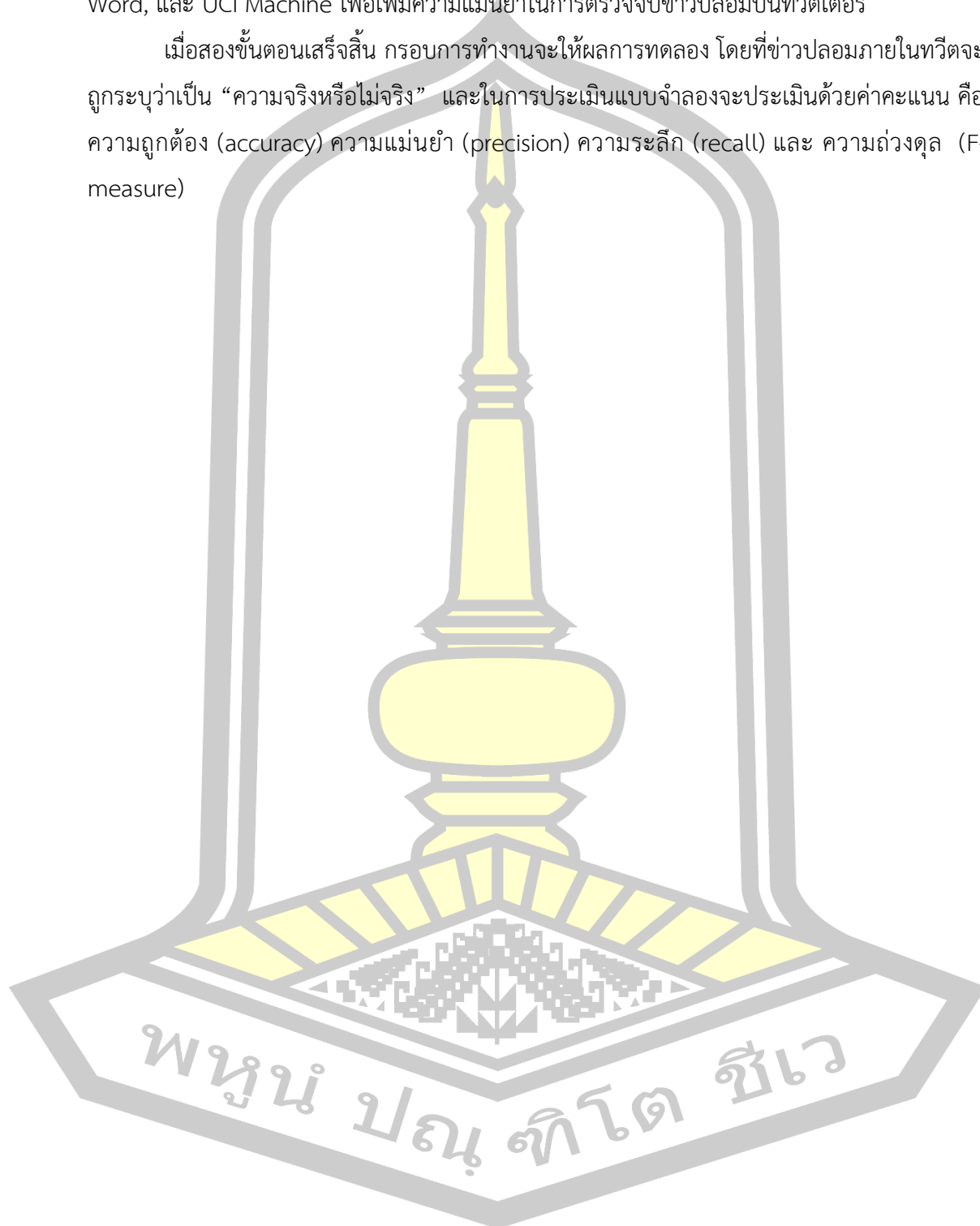
ขั้นตอนที่ 2 การตรวจจับข่าวปลอม

กรอบการทำงานนี้จะกำหนดสถานะความเป็นจริงของข่าวปลอมภายในทวีต โดยที่

1. ใช้ประโยชน์จากชุดข้อมูลข่าวปลอมที่มีอยู่ ซึ่งอิงตามวิธีการเรียนรู้ของเครื่อง ประกอบด้วย นาอ์ฟเบย์, ซัพพอร์ตเวกเตอร์แมชชีน, ต้นไม้ตัดสินใจ , ต้นไม้แห่งการทำนาย และ การถดถอยโลจิสติก

2. การรวบรวมข้อมูลจากแพลตฟอร์มทวิตเตอร์ ประกอบด้วย 3 ฐานข้อมูล Kaggle, Data Word, และ UCI Machine เพื่อเพิ่มความแม่นยำในการตรวจจับข่าวปลอมบนทวิตเตอร์

เมื่อสองขั้นตอนเสร็จสิ้น กรอบการทำงานจะให้ผลการทดลอง โดยที่ข่าวปลอมภายในทวิตจะถูกระบุว่าเป็น “ความจริงหรือไม่จริง” และในการประเมินแบบจำลองจะประเมินด้วยค่าคะแนน คือ ความถูกต้อง (accuracy) ความแม่นยำ (precision) ความระลึก (recall) และ ความถ่วงดุล (F-measure)



บทที่ 3

วิธีการดำเนินงานวิจัย

จากการศึกษาวิจัยเรื่องการปรับปรุงการตรวจจับข่าวปลอมด้วยซอฟต์แวร์แมชชีนโดยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม มีวัตถุประสงค์ในการวิจัยเพื่อ 1) เปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม 2) นำอัลกอริทึมมาประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย 3) ปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม

ทั้งนี้ผู้วิจัยใช้ข้อมูลจากฐานข้อมูลข่าวจากสื่อโซเชียลมีเดียทวิตเตอร์ และนำเข้าผ่านโปรแกรมเพื่อทำการวิเคราะห์ข้อมูล ซึ่งดำเนินการตามขั้นตอนวิธีการดำเนินการวิจัยดังนี้

1. ขั้นตอนและวิธีการ
2. ชุดข้อมูล
3. การประมวลผลข้อมูลล่วงหน้า
4. แบบจำลองพื้นฐานของเทคนิคการเรียนรู้ของเครื่อง

3.1 ขั้นตอนและวิธีการ

การศึกษานี้เป็นการนำเสนอกรอบการตรวจจับข่าวปลอมซึ่งเป็นลักษณะแบบสองขั้นตอนที่เป็นนวัตกรรมและมีประสิทธิภาพ (FNDF) สามารถระบุข้อมูลที่ไม่ใช่ข้อเท็จจริงบนทวิตเตอร์ได้อย่างมีประสิทธิภาพ ในกรอบการทำงานประกอบด้วยสองขั้นตอนหลักๆ ได้แก่ การประมวลผลข้อมูลล่วงหน้า และการตรวจจับข่าวปลอม โดยประกอบด้วย 2 ระยะดังนี้

ระยะแรก การประมวลผลข้อมูลล่วงหน้าเกี่ยวข้องกับการล้างข้อมูลและเลือกพารามิเตอร์ให้กับอัลกอริทึมในการตรวจสอบเนื้อหาของข่าวบนทวิตเตอร์

ขั้นตอนที่ 1 ขั้นตอนนี้มีจุดมุ่งหมายเพื่อเพิ่มความแม่นยำในการระบุข่าวปลอมจากเนื้อหาของทวิตเตอร์

ระยะที่ 2 ระยะตรวจจับข่าวปลอม กรอบการทำงานจะกำหนดสถานะข้อเท็จจริงของการอ้างข่าวปลอมภายในทวิต

ขั้นตอนที่ 2 ศึกษาเอกสาร งานวิจัยที่เกี่ยวข้องเกี่ยวกับอัลกอริทึมซึ่งโดยการใช้ประโยชน์จากชุดข้อมูลข่าวปลอมที่มีอยู่ตามวิธีการเรียนรู้ของเครื่อง เช่น นาอ์ฟเบย์ (Naïve bayes), ต้นไม้ตัดสินใจ (Decision tree), ต้นไม้แห่งการทํานาย (Random forest), การถดถอยโลจิสติก (Logistic Regression) และ ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines)

ขั้นตอนที่ 3 การรวบรวมข้อมูลข่าวปลอมจากแพลตฟอร์มทวิตเตอร์ ซึ่งประกอบด้วย ชุดข้อมูล Kaggle, Data Word และ UCI เพื่อปรับปรุงความแม่นยำในการตรวจจับข่าวปลอมบน ทวิตเตอร์

เมื่อทั้งสองระยะเสร็จสิ้น กรอบการทำงานจะส่งกลับไปยังผลการทดลอง โดยที่การ ตรวจสอบข่าวปลอมในทวิตจะถูกระบุว่าเป็นข้อเท็จจริงหรือไม่ใช่ข้อเท็จจริง เพื่อประเมินแบบจำลอง ด้วยความถูกต้อง ความแม่นยำ ความระลึก และการเรียกคืน ซึ่งการศึกษานี้ตั้งสมมติฐานว่าสามารถ เปรียบเทียบประสิทธิภาพของโมเดลแมชชีนเลิร์นนิง กับชุดข้อมูลต่างๆ เพื่อกำหนดพารามิเตอร์ของ ซัพพอร์ตเวกเตอร์แมชชีน ด้วยเคอร์เนลพหุนามที่ตรวจจับข่าวปลอมได้อย่างแม่นยำมากกว่าวิธีอื่นๆ นอกจากนี้ การศึกษายังตั้งสมมติฐานว่าซัพพอร์ตเวกเตอร์แมชชีนพร้อมเคอร์เนลพหุนามที่ได้รับการ ฝึกอบรมและทดสอบกับชุดข้อมูล Kaggle มีประสิทธิภาพที่ดีที่สุดในการระบุข่าวปลอมบนทวิตเตอร์

ขั้นตอนที่ 4 สุดท้าย จะประกอบด้วยการรวมส่วนประกอบทั้งหมดเข้ากับการประมวลผล ล่วงหน้า และการเลือกอัลกอริทึมที่เหมาะสม การศึกษายังตั้งสมมติฐานว่าประสิทธิภาพการตรวจจับที่ดี สามารถทำได้เหนือกว่าเทคนิคที่ล้ำสมัยในรูปแบบ end-to-end

3.2 ชุดข้อมูล

ปัจจุบันมีฐานข้อมูลข่าวปลอมโอเพ่นซอร์สจำนวนมาก โดยแต่ละฐานข้อมูลมีระบบการ จำแนกประเภท ฐานข้อมูลยอดนิยมบางส่วน ได้แก่ UCI Machine Learning Repository, Data World, Subreddit /Datasets R Package – DS Labs และ Kaggle โดยทั่วไปนักวิจัยจะใช้ ฐานข้อมูลเหล่านี้เพื่อสร้างแบบจำลองเพื่อช่วยระบุข่าวปลอม นักวิจัยมักจะดึงข้อมูลจากเว็บไซต์และ แหล่งอื่นๆ เพื่อสร้างแบบจำลองเหล่านี้ หนึ่งในฐานข้อมูลข่าวปลอมที่ใช้กันอย่างแพร่หลายที่สุดคือชุด ข้อมูล Internet News Liar, Liar Pants on Fire ซึ่งมีรายการข่าวสั้น 12.8 K รายการที่เรียกว่า LIAR ชุดข้อมูลนี้ประกอบด้วยข่าวปลอมจากแหล่งต่างๆ เช่น Twitter, Buzz Feed News และ Weir-do ชุดข้อมูลเปรียบเทียบต่างๆ ได้รับการพัฒนาเพื่อตรวจจับข่าวปลอม และชุดข้อมูลดังกล่าว หนึ่งชุดคือปลอมหรือจริง ชุดข้อมูลเหล่านี้ประกอบด้วยข้อมูลข่าวที่มีป้ายกำกับซึ่งครอบคลุมหลาย หัวข้อโดยเน้นไปที่หัวข้อทางสังคม ธุรกิจ และการเมือง นอกจากนี้ Kaggle ยังเป็นแหล่งชุดข้อมูลข่าว อันทรกคุณค่าที่ใช้ในงานวิจัยต่างๆ เพื่อกำหนดข้อกำหนดเฉพาะของข่าว อย่างไรก็ตาม การเข้าถึงชุด ข้อมูลชุดแรกที่สูงขึ้นโดยเฉพาะสำหรับข่าวในบริบททางสังคมนั้นถูกจำกัด และมีบทความข่าวที่มี ป้ายกำกับเพียง 344 บทความที่ดึงมาจากอินเทอร์เน็ต มีชุดข้อมูลอื่นๆ ซึ่งสามารถช่วยในการพัฒนา และประเมินโปรแกรม AI (ปัญญาประดิษฐ์) ต่างๆ สำหรับการศึกษา มีการใช้แหล่งข้อมูลหลาย แหล่งเพื่อรวบรวมข้อมูลข่าวเกี่ยวกับข่าวปลอม โดยใช้คำศัพท์สำคัญที่เกี่ยวข้องกับข่าวบริบททาง สังคม ข่าวการเมือง ข่าวทางธุรกิจ และข่าวทั่วไป

กระบวนการรวบรวมข้อมูลเพื่อระบุข่าวปลอมเกี่ยวข้องกับการใช้แหล่งข้อมูลต่างๆ Kaggle หนึ่งในชุมชนวิทยาศาสตร์ข้อมูลที่ได้รับการยอมรับอย่างกว้างขวางที่สุดคือแหล่งข้อมูลที่สำคัญ Kaggle เป็นที่รู้จักในด้านการสนับสนุนวัตถุประสงค์ด้านวิทยาศาสตร์ข้อมูลหลายประการ รวมถึงการระบุข่าวปลอม แพลตฟอร์มดังกล่าวนำเสนอปัญหามากมายและชุดข้อมูลที่ทันสมัยสำหรับการตรวจจับข่าวปลอม ซึ่งเป็นทรัพยากรที่มีคุณค่าสำหรับนักวิจัยในสาขานี้ ในเรื่องนี้ ชุดข้อมูลที่พัฒนาแล้วซึ่งมีรายการข่าวที่เกี่ยวข้องได้ถูกสร้างขึ้น ชุดข้อมูลนี้รวบรวมข้อมูลข่าวสารทั้งประเภทข่าวจริงและข่าวปลอม การศึกษานี้ใช้ชุดข้อมูลจำนวน 3 ชุด เพื่อประเมินประสิทธิภาพของตัวแยกประเภทในการระบุข่าวปลอม ชุดข้อมูลเหล่านี้ได้มาจากไลบรารี Kaggle, Data World และ Machine Learning Repository (UCI) ตาราง 6 รวบรวมชุดข้อมูลที่ใช้ในการศึกษา ความแม่นยำของโมเดลประเมินโดยใช้การแยกการทดสอบ 70:30 โดยที่ 70% ของชุดข้อมูลถูกใช้สำหรับการฝึก และ 30% สำหรับการทดสอบ ข้อมูลการทดสอบถูกใช้เพื่อปรับไฮเปอร์พารามิเตอร์ของโมเดล นอกจากนี้ ทวิตข่าวปลอมจากผู้ใช้น Twitter ยังถูกรวบรวมในช่วงสามปี ตั้งแต่ปี 2018 ถึง 2020 ดังแสดงในตาราง 6 ชุดข้อมูลเหล่านี้ให้ข้อมูลเชิงลึกที่มีคุณค่าเกี่ยวกับประสิทธิภาพของตัวแยกประเภทในการระบุข่าวปลอม

ตาราง 6 สถิติของชุดข้อมูลที่พัฒนาแล้ว

ฐานข้อมูล	ขนาด	ข้อมูลการฝึกอบรม	ข้อมูลการทดสอบ
	ตัวอย่าง	(70%)	(30%)
Kaggle	20,717	14,502	6,215
Data word	19,530	13,671	5,859
UCI Machine Learning Repository	21,248	14,875	6,373

3.3 การประมวลผลข้อมูลล่วงหน้า

ก่อนที่จะดำเนินการวิเคราะห์ข้อมูลใดๆ ชุดของขั้นตอนการประมวลผลข้อมูลล่วงหน้าเป็นขั้นตอนที่มีความสำคัญ และจะต้องได้รับการดำเนินการอย่างพิถีพิถันเพื่อให้มั่นใจถึงความถูกต้องและความน่าเชื่อถือของผลลัพธ์หรือข้อมูล ผู้วิจัยจะต้องมีการเอาใจใส่เป็นพิเศษในการกำจัดจุดข้อมูลที่ขาดหายไป ซึ่งเกี่ยวข้องกับการตรวจสอบบันทึกข้อมูลทั้งหมดอย่างรอบคอบ และการลบอินสแตนซ์ใดๆ ที่มีข้อมูลที่ไม่สมบูรณ์หรือไม่สอดคล้องกันออกจากชุดข้อมูล นอกจากนี้ ข้อมูลยังถูกแบ่งอย่างถูกต้องเป็นชุดการฝึกอบรมและการทดสอบแยกกัน ซึ่งช่วยให้สามารถพัฒนาและตรวจสอบความ

ถูกต้องของแบบจำลองการคาดการณ์บนชุดข้อมูลที่เป็นอิสระได้ ขั้นตอนเหล่านี้ดำเนินการด้วยความเอาใจใส่และใส่ใจในรายละเอียดเป็นอย่างยิ่ง เพื่อให้แน่ใจว่าการวิเคราะห์และแบบจำลองที่ได้จะมีประสิทธิภาพ และเชื่อถือได้ ซึ่งมีการดำเนินงานดังนี้

3.3.1 การทำความสะอาดข้อมูล

การทำความสะอาดข้อมูลเป็นกระบวนการล้างข้อมูลที่สำคัญในการจัดการข้อมูลที่เกี่ยวข้องกับการระบุและลบข้อมูลที่ไม่สมบูรณ์ ไม่เกี่ยวข้อง และไม่ถูกต้องออก กระบวนการนี้มีความสำคัญเพื่อให้แน่ใจว่าข้อมูลมีความถูกต้อง สม่าเสมอ และปราศจากข้อผิดพลาด เพื่อให้บรรลุเป้าหมายในการวิเคราะห์ข้อมูล ผู้วิจัยมีการดำเนินการหลายวิธีในระหว่างการล้างข้อมูล หนึ่งในกระบวนการแรกๆ ในการล้างข้อมูลคือการกำจัดบันทึกที่ซ้ำกันในหลายกรณี รายการที่ซ้ำกันซึ่งมีข้อความและป้ายกำกับซ้ำตรงกันทุกประการสามารถพบได้ในข้อมูลที่รวบรวมจากแหล่งต่างๆ เนื่องจากค่าย่อซ้ำจากแต่ละแหล่งเหมือนกัน จึงไม่มีข้อขัดแย้งในรายการที่ซ้ำกัน ดังนั้น ในขั้นตอนแรกของการล้างข้อมูล บันทึกที่ซ้ำกันทั้งหมดจะถูกตัดออก เพื่อให้แน่ใจว่าชุดข้อมูลจะมีเฉพาะบันทึกข้อมูลใหม่ที่ไม่ซ้ำกัน กระบวนการที่จำเป็นอีกประการหนึ่งในการล้างข้อมูลคือการลบข้อมูลที่ขาดหายไป เนื้อหาข่าวเป็นองค์ประกอบที่สำคัญที่สุดในข้อมูลในการตรวจจับข่าวปลอม ด้วยเหตุนี้ บันทึกข่าวทั้งหมดที่ข้อความข่าวหายไปจะถูกละทิ้งเป็นขั้นตอนต่อไปในกระบวนการล้างข้อมูล เพื่อให้แน่ใจว่าชุดข้อมูลจะมีเฉพาะข้อมูลข่าวข้อความต้นฉบับที่สมบูรณ์เท่านั้น

โดยสรุปการทำความสะอาดข้อมูลถือเป็นสิ่งสำคัญเพื่อให้แน่ใจว่าข้อมูลมีความถูกต้อง สม่าเสมอ และปราศจากข้อผิดพลาด ชุดข้อมูลจะมีความน่าเชื่อถือและเหมาะสำหรับการวิเคราะห์ และการตัดสินใจโดยการกำจัดบันทึกที่ซ้ำกันและลบข้อมูลที่ขาดหายไป

3.3.2 การเลือกคุณสมบัติ

หลังจากประมวลผลล่วงหน้า และการทำความสะอาดข้อมูลแล้ว ขั้นตอนต่อไปคือการแยกคุณลักษณะ สิ่งนี้เกี่ยวข้องกับการลดขนาดของข้อมูลโดยการกำจัดคุณสมบัติพิเศษ ซึ่งทำให้ง่ายต่อการจัดประเภทข้อความ ผู้วิจัยใช้เทคนิคการเลือกคุณลักษณะที่ได้รับความนิยมสองเทคนิคในการวิจัยนี้ ได้แก่ การเลือกคุณลักษณะที่อิงความสัมพันธ์และวิธีการเลือกคุณลักษณะตามความสัมพันธ์ ซึ่งเกี่ยวข้องกับการเลือกคุณลักษณะที่มีความสัมพันธ์สูงกับตัวแปรคลาส วิธีนี้มีข้อได้เปรียบเนื่องจากมีประสิทธิภาพในการคำนวณและสามารถจัดการกับคุณลักษณะต่างๆ ได้มากมาย ในทางกลับกัน อีกวิธีที่ใช้ในการวิจัยนี้ที่ควรกล่าวถึงในข้อความต้นฉบับคือวิธีการเลือกคุณลักษณะข้อมูลร่วมกัน วิธีการนี้จะเลือกคุณลักษณะที่มีคะแนนข้อมูลร่วมกันสูงพร้อมกับตัวแปรคลาส โดยรวมแล้ว เทคนิคการเลือกคุณลักษณะเหล่านี้มีความสำคัญในการลดมิติของข้อมูล ซึ่งช่วยปรับปรุงประสิทธิภาพและความแม่นยำของแบบจำลองการจำแนกประเภท

3.4 แบบจำลองพื้นฐานของเทคนิคการเรียนรู้ของเครื่อง

กระบวนการเรียนรู้ของเครื่องเป็นที่ยอมรับในด้านการประยุกต์ใช้งานกับการแก้ไขปัญหาหลายด้าน หรือใช้สำหรับโปรแกรมประยุกต์มากมาย โดยแต่ละกระบวนการเรียนรู้มีความเหมาะสมกับข้อมูลรูปแบบเฉพาะ กระบวนการเรียนรู้ด้วยเครื่องที่แก้ปัญหาด้วยการสอนโดยใช้ข้อมูลที่มีคำตอบที่แท้จริงของปัญหาด้วยการสอนโดยใช้ข้อมูลที่มีคำตอบที่แท้จริงของปัญหา กระบวนการของการเรียนรู้ของเครื่องโดยทั่วไปจะเป็นการสนับสนุนด้านการแนะนำ หรือการพยากรณ์เพื่อปรับปรุงแก้ไข การเพิ่มประสิทธิภาพในการถือหุ้น การทำงานอัตโนมัติ และอื่น ๆ ที่เกี่ยวข้อง การเรียนรู้ของเครื่อง เรียกอีกอย่างหนึ่งว่า การเรียนรู้แบบมีผู้สอน (Supervised Learning) ทั้งนี้การศึกษาวิจัยครั้งนี้ ผู้วิจัยได้นำอัลกอริทึมมาใช้ในการตรวจสอบข่าวปลอมทั้งหมด 5 อัลกอริทึม ได้แก่

1. นาอิวเบย์ (Naïve bayes)
2. ต้นไม้ตัดสินใจ (Decision tree)
3. ต้นไม้แห่งการทำนาย (Random forest)
4. การถดถอยโลจิสติก (Logistic Regression)
5. ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines)

1. นาอิวเบย์ (Naïve bayes) อัลกอริทึมนาอิวเบย์เป็นเครื่องจักรเรียนรู้ที่อาศัยหลักการความน่าจะเป็น ตามทฤษฎีของเบย์ (Bayes Theorem) ซึ่งเป็นอัลกอริทึมที่ไม่ซับซ้อน เป็นขั้นตอนวิธีการจำแนกข้อมูล โดยการเรียนรู้ปัญหาที่เกิดขึ้น เพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ หลักการของนาอิวเบย์ใช้การคำนวณหาความน่าจะเป็นในการทำนายผล เป็นเทคนิคในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ จะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์เหมาะสมกับกรณีของเซตตัวอย่างที่มีจำแนกมาก และคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน

วิธีการเรียนรู้เบย์อย่างง่าย (Naïve Bayesian learning) เป็นวิธีการจำแนกประเภทข้อมูลที่มีประสิทธิภาพวิธีหนึ่ง โดยที่ใช้งานได้ดีเหมาะสมกับกรณีของเซตตัวอย่างมีจำนวนมาก และคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกันมีการจำแนกประเภทเบย์อย่างงานไปประยุกต์ใช้งานในด้านการจำแนกประเภทข้อความ การวินิจฉัย ผู้วิจัยเลือกวิธีการนี้มาใช้งานวิจัย เนื่องจากเป็นวิธีการจำแนกข้อมูลได้อย่างมีประสิทธิภาพ และมีอัลกอริทึมในการทำงานที่ไม่ซับซ้อน

การกำหนดให้ความน่าจะเป็นของข้อมูลที่จะเป็นกลุ่ม V_j สำหรับข้อมูลที่มีคุณสมบัติ n ตัว $X=\{a_1, a_2, \dots, a_n\}$ หรือใช้สัญลักษณ์ว่า $P(a_1, a_2, \dots, a_n|V_j)$ สามารถแสดงในสมการที่ 12

$$P(a_1, a_2, \dots, a_n | v_j) = \prod_{i=1}^n P(a_i | v_j) \quad (12)$$

โดย $\prod$ หมายถึง ผลคูณของค่า $P(a_i | v_j)$ ทั้งหมด
 $i = 1, 2, 3, \dots, n$ และ $j = 1, 2, 3, \dots, n$

การนำวิธีการนาอ์ฟเบย์ไปใช้มีวิธีการดังต่อไปนี้

1. หาค่าความน่าจะเป็นของค่าที่พบในแต่ละกลุ่มโดยนำค่า $P(a_1, a_2, \dots, a_n | v_j)$ จากสมการที่ 12 มาคูณกับค่าความน่าจะเป็นของกลุ่มนั้น ๆ คือ $P(v_j)$ ได้เท่ากับ v_{NB}
2. นำค่าที่ได้เปรียบเทียบกับ กลุ่มที่มีค่าความน่าจะเป็นสูงสุด คือ คำตอบ ดังนั้นจะได้วิธีการจำแนกประเภทของนาอ์ฟเบย์ สามารถแสดงในสมการที่ 13

$$v_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \times \prod_{i=1}^n P(a_i | v_j) \quad (13)$$

จากการสมการที่ 13 สามารถเขียนเป็นอัลกอริทึมนาอ์ฟเบย์ได้ดังต่อไปนี้

Naïve_Bayes_Learn(examples)

FOR EACH target value v DO

$\bar{P}(v_j) \leftarrow \text{estimate } P(V_j)$

FOR EACH attribute value a of each attribute DO

$\bar{P}(a_i | v_j) \leftarrow \text{estimate } P(a_i | v_j)$

Classify_New_Example(x)

$$v_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \times \prod_{i=1}^n \bar{P}(a_i | v_j)$$

2. ต้นไม้ตัดสินใจ (Decision tree) เป็นกระบวนการทางด้านเหมืองข้อมูล (Data mining) ได้นำการตัดสินใจแบบโครงสร้างต้นไม้มาช่วยในการทำงานด้านการตัดสินใจต่าง ๆ ไม่ว่าจะเป็นทางด้านระบบธุรกิจหรือด้านอื่น ๆ โดยปกติประกอบด้วยกฎในรูปแบบ “ถ้าเงื่อนไข แล้ว ผลลัพธ์” เช่น “If Income = High and Married = No THEN Risk = Poor” โดยลักษณะของการตัดสินใจแบบโครงสร้างต้นไม้จะมีลักษณะคล้ายกับต้นไม้กลับหัวโดยโหนดแรกสุดจะเป็นรากของต้นไม้ (Root node) แต่ละโหนดแสดงคุณลักษณะ (Attribute) แต่ละกิ่งจะแสดงค่าผลในการทดสอบ และโหนดใบ (Leaf node) การแสดงคลาสที่กำหนด

ในปัจจุบันนั้นมีการพัฒนาขั้นตอนวิธีในการสอน (training) ต้นไม้การตัดสินใจมากมาย ซึ่งส่วนมากมาจากวิธีพื้นฐานวิธีหนึ่งซึ่งเป็นการค้นหาแบบจากบนลงล่าง (top-down) ชื่อว่า ID3 การสร้างต้นไม้การตัดสินใจจากบนลงล่าง ควรจะเป็นรากของต้นไม้การตัดสินใจต้น และถามซ้ำ ๆ ไปเรื่อย ๆ เพื่อหาต้นไม้ทั้งต้นด้วยการเขียนโปรแกรมด้วยความสัมพันธ์ โดยในการเลือกว่าลักษณะใดดีที่สุดนั้นดูจากค่าของลักษณะ เรียกว่าเกนความรู้ (Information gain) สามารถแสดงในสมการที่ 14

$$E(s) = \sum_{i=1}^n P(v_i) \log_2 P(v_i) \quad (14)$$

โดย $E(S)$ = ค่า Entropy ของเซต (S) ข้อมูลทั้งหมด, $S=P(V_1), P(V_2), P(V_3), \dots, P(V_n)$

$$P(V_1) = \text{ค่าความน่าจะเป็นของข้อมูลที่สนใจ}$$

จากสมการที่ 19 สามารถคำนวณได้ดังในสมการที่ 15

$$\text{Gain}(S,A) = E(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} E(S_v) \quad (15)$$

3. ต้นไม้แห่งการทำนาย (Random forest) เทคนิคต้นไม้แห่งการทำนายเป็นชุดของการจำแนกประเภทแบบไม่ตัดแต่งกิ่ง หรือต้นไม้ถลออย ซึ่งถูกสร้างจากการนำข้อมูลฝึกสอนไปสุ่มตัวอย่าง ข้อมูลและคุณลักษณะข้อมูลแล้วนำมาสร้างเป็น ต้นไม้ตัดสินใจ ซึ่งมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือก เรียกข้อมูลส่วนนี้ว่า Out-of-Bag จะถูกนำมาในการทดสอบต้นไม้ตัดสินใจ อัลกอริทึมของต้นไม้แห่งการทำนายที่เสนอโดย L. Breiman มีการกำหนดลักษณะความแม่นยำของต้นไม้แห่งการทำนายดังนี้

3.1 ต้นไม้แห่งการทำนายแบบบรรจุ เป็นการกำหนดกลุ่มของตัวแยกประเภท $h_1(x), h_2(x), \dots, h_k(x)$ และด้วยชุดการฝึกवादโดยการสุ่มจากการแจกแจงของเวกเตอร์สุ่ม Y, X , กำหนดฟังก์ชันระยะขอบ สามารถแสดงในสมการที่ 16

$$Mg(X,Y) = \text{avg}_k |h_k(X) = Y| = \max_{j \neq Y} \text{avg}_k |h_k(X)=j| \quad (16)$$

โดยที่ $|(\cdot)|$ คือฟังก์ชันตัวบ่งชี้ ส่วนต่างจะวัดขอบเขตที่จำนวนคะแนนเสียงเฉลี่ยที่ X, Y สำหรับกลุ่มที่ถูกต้องเกินกว่าคะแนนเฉลี่ยสำหรับกลุ่มอื่นๆ ยิ่งมีระยะขอบมากเท่าใดก็ยิ่งมีความมั่นใจมากขึ้นในการจัดประเภท ข้อผิดพลาดทั่วไปถูกกำหนดโดยสามารถแสดงในสมการที่ 17

$$PE^* = P_{x,Y} (mg(X,Y)<0) \quad (17)$$

โดยที่ตัวห้อย X,Y ระบุว่าความน่าจะเป็นอยู่เหนือช่องว่าง X,Y

ในต้นไม้แห่งการทำนาย $h_k(X) = h(X, \theta_k)$ สำหรับต้นไม้จำนวนมาก เป็นไปตามกฎที่แข็งแกร่งของตัวเลขจำนวนมากและโครงสร้างต้นไม้ที่:

3.2 ความแข็งแกร่งและความสัมพันธ์ สำหรับต้นไม้แห่งการทำนาย ขอบเขตบนสามารถรับข้อผิดพลาดในการทำให้เป็นข้อมูลทั่วไปในแง่ของพารามิเตอร์สองตัวที่ใช้วัดว่าตัวแยกประเภทแต่ละรายการมีความแม่นยำเพียงใดและการพึ่งพาหะร่วมกัน การมีปฏิสัมพันธ์ระหว่างทั้งสองเป็นรากฐานสำหรับการทำความเข้าใจการทำงานของป่าสุ่ม ดังนี้

1) ฟังก์ชันระยะขอบสำหรับต้นไม้แห่งการทำนาย สามารถแสดงในสมการที่ 18

$$Mr(X,Y) = P_{\theta}(h(x, \theta)=Y) - \frac{\max_{j \neq Y}}{j \neq Y} P_{\theta}(h(x, \theta)=j) \quad (18)$$

และกำลังของชุดลักษณะนาม $\{h(x, \theta)\}$ สามารถแสดงในสมการที่ 19

$$s = E_{X,Y} mr(X, Y) \quad (19)$$

สมมติว่า $s \geq 0$ ความไม่เท่าเทียมกัน สามารถแสดงในสมการที่ 20

$$PE^* \leq \text{var}(mr)/s^2 \quad (20)$$

4. การถดถอยโลจิสติก (Logistic Regression) การวิเคราะห์การถดถอยโลจิสติกเป็นเทคนิคการวิเคราะห์ตัวแปรเชิงพหุที่มีวัตถุประสงค์เพื่อประมาณค่าหรือทำนายเหตุการณ์ที่สนใจว่าจะเกิดหรือไม่เกิดเหตุการณ์นั้น ภายใต้อิทธิพลของตัวปัจจัย แบบจำลองโลจิสติกประกอบด้วยตัวแปรตาม (หรือตัวแปรเกณฑ์) ที่ต้องเป็นตัวแปรแบบทวินาม (Dichotomous Variable) กล่าวคือมีสองค่า เช่น เกิด กับ ไม่เกิด หรือ จริง กับไม่จริง เป็นต้น และตัวแปรอิสระ (หรือตัวแปรทำนาย) ที่อาจมีตัวเดียวหรือหลายตัวที่เป็นได้ทั้งตัวแปรเชิงกลุ่ม (Categorical Variable) หรือตัวแปรแบบต่อเนื่อง (Continuous Variable) การวิเคราะห์การถดถอยโลจิสติกเกี่ยวข้องกับทฤษฎีความน่าจะเป็นทวินาม ถูกเรียกว่า Binomial Logistic Regression ถ้าตัวแปรตามเป็นพหุนามจะเรียกว่า Multinomial Logistic Regression การถดถอยโลจิสติกจัดเป็นเครื่องมือที่นำมาวิเคราะห์ข้อมูลในการศึกษาการวิจัยที่มีวัตถุประสงค์เพื่อทำนายเหตุการณ์ หรือประเมินความเสี่ยง

การวิเคราะห์การถดถอยโลจิสติก เป็นเทคนิคการวิเคราะห์ตัวแปรเชิงพหุ เป็นการทำนายความน่าจะเป็นของการเกิดเหตุการณ์ หรือไม่เกิดเหตุการณ์ ความสัมพันธ์จะอยู่ในรูปแบบของสมการเส้นถดถอย (Regression Equation) มีตัวแปรตามที่มีค่าเพียง 2 ค่าคือ เป็น 0 และ 1 ตัวแปรอิสระ

อาจจะมีตัวเดียวหรือหลายตัวก็ได้ค่าพยากรณ์หรือค่าตัวแปรจะมีค่าอยู่ที่ 0 กับ 1 โดยที่ถ้าค่าความน่าจะเป็น >0.5 ถือว่าค่าพยากรณ์คือ 1 ถ้าค่าความน่าจะเป็น <0.5 ถือว่าค่าพยากรณ์คือ 0 สามารถแสดงในสมการที่ 21

$$L(Y) = b_0 + b_1 X + \dots + b_k X^k \quad (21)$$

โดย	L(Y)	คือ ค่าพยากรณ์ของตัวแปรตาม
	Y	คือ ตัวแปรตาม
	b_0	คือ ค่าคงที่
	b_1	คือ ค่าสัมประสิทธิ์การถดถอยสำหรับตัวแปร
	b_k	คือ ค่าสัมประสิทธิ์การถดถอยสำหรับตัวแปร
	X	คือ ค่าของตัวแปรอิสระ

5. ซัพพอร์ทเวกเตอร์แมชชีน (Support vector machines) หรือ SVM เป็นอีกหนึ่งอัลกอริทึมในกลุ่มวิธีการเรียนรู้ของเครื่องแบบมีผู้สอน ที่สามารถนำมาช่วยแก้ปัญหาการจำแนกข้อมูลได้ โดยเฉพาะกับปัญหาที่มีขนาดของข้อมูลไม่ใหญ่มาก แต่คุณลักษณะ (Features) ของข้อมูลมีเป็นจำนวนมาก SVM จะถือได้ว่าเป็นอัลกอริทึมที่ทำงานได้ค่อนข้างจะมีประสิทธิภาพ ๆ อัลกอริทึมหนึ่งหลักการทำงานของ SVM อาศัยใช้การสร้างเส้นแบ่ง หรือไฮเปอร์เพลน (Hyperplane) ในการแบ่งแยกคลาสของข้อมูลออกจากกัน จากนั้นจะทำการหาไฮเปอร์เพลนใดเป็นเส้นที่ใช้แยกคลาสของข้อมูลได้ดีที่สุด

ซัพพอร์ทเวกเตอร์แมชชีน เป็นเทคนิคที่ช่วยในการเรียนรู้ ซึ่งมีลักษณะการทำงานคล้ายกับโครงข่ายประสาทเทียม แต่ส่วนแตกต่างกัน คือ ในเทคนิคโครงข่ายประสาทเทียมใช้หลักการลดความเสี่ยงเชิงการทดสอบให้มีค่าต่ำที่สุด ส่วนซัพพอร์ทเวกเตอร์แมชชีนใช้หลักการลดความเสี่ยงเชิงโครงสร้างให้ต่ำที่สุด โดยซัพพอร์ทเวกเตอร์แมชชีนประยุกต์การใช้งานได้ 2 รูปแบบ คือ การวิเคราะห์การถดถอย หรือการประมาณค่าของฟังก์ชัน และการจำแนกประเภท อีกทั้งซัพพอร์ทเวกเตอร์แมชชีนยังถูกนำมาใช้ในการจำแนกข้อมูลในลักษณะหลายมิติ โดยการใช้ Hyper plane ในการแบ่งแยกคุณสมบัติออกจากกัน สำหรับการจำแนกในลักษณะไบนารี (Binary) นั้นมี Class A และ B โดย W คือน้ำหนัก (Weight) , x คือ คุณสมบัติต่าง ๆ และ b คือ Bias สามารถแสดงในสมการที่ 22 – 25

$$W^T X + b = 0 \quad (22)$$

โดยที่

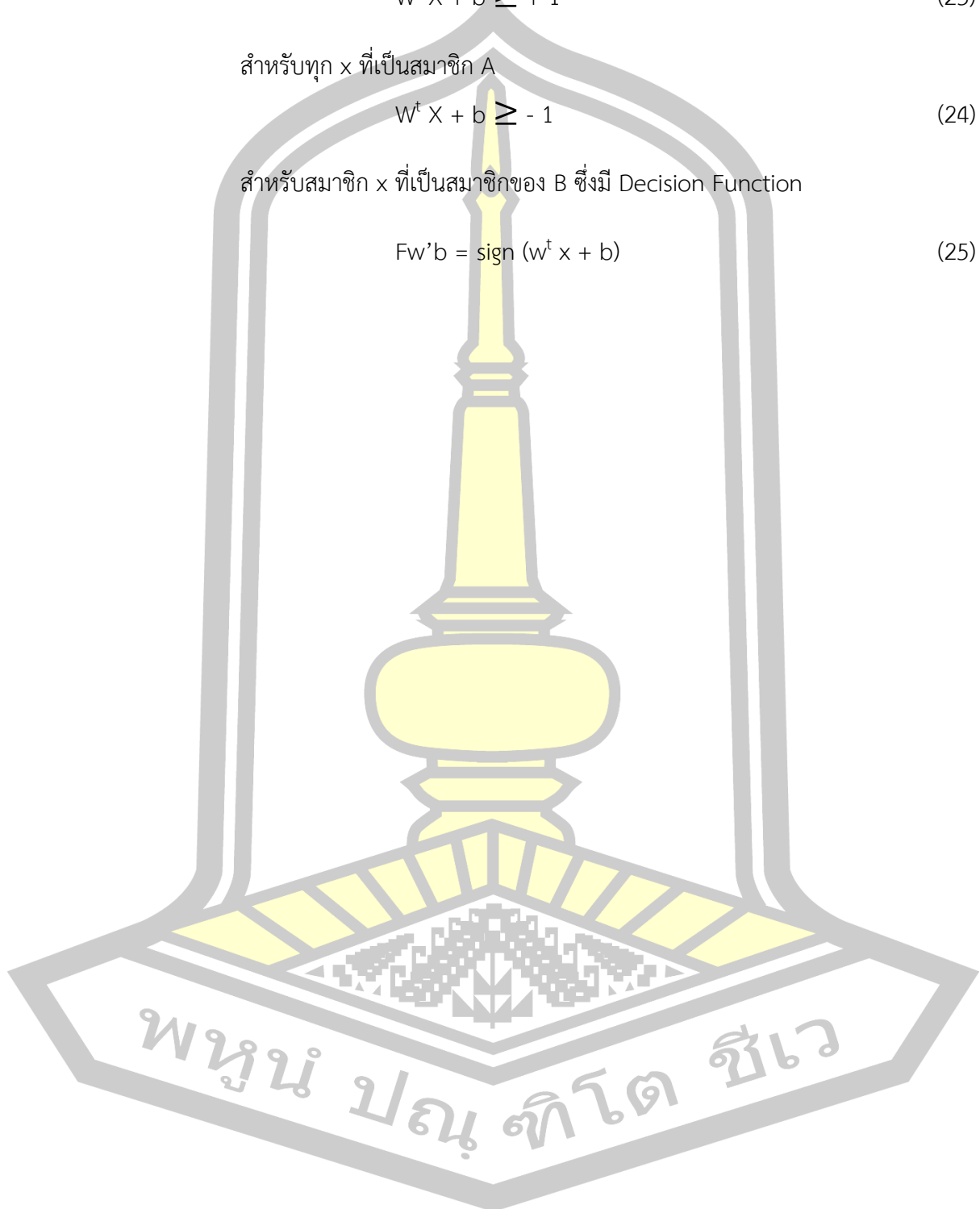
$$w^t x + b \geq +1 \quad (23)$$

สำหรับทุก x ที่เป็นสมาชิก A

$$w^t x + b \geq -1 \quad (24)$$

สำหรับสมาชิก x ที่เป็นสมาชิกของ B ซึ่งมี Decision Function

$$f_{w,b} = \text{sign}(w^t x + b) \quad (25)$$



บทที่ 4

ผลการวิจัย

งานวิจัยเรื่อง การปรับปรุงการตรวจจับข่าวปลอมด้วยซัพพอร์ทเวกเตอร์แมชชีนโดยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม มีวัตถุประสงค์เพื่อ 1) เปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม 2) นำอัลกอริทึมไปประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย และ 3) ปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม เมื่อผู้วิจัยได้รวบรวมข้อมูลตามกระบวนการที่ได้ออกแบบไว้ สามารถสรุปผลการวิจัยดังนี้

4.1 การวิเคราะห์การตรวจจับข่าวปลอม

การวัดประสิทธิภาพของระบบจะทดสอบและทำการประเมินแบบจำลองด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F – Measure) ในการพยากรณ์การตรวจจับข่าวปลอม โดยทำการเปรียบเทียบอัลกอริทึมในการจำการจำแนกทั้งหมด 5 อัลกอริทึม คือ นาอิวเบย์ (Naïve bayes) ต้นไม้ตัดสินใจ (Decision tree) ต้นไม้แห่งการทำนาย (Random forest) การถดถอยโลจิสติก (Logistic Regression) และ ซัพพอร์ทเวกเตอร์แมชชีน (Support vector machines) ซึ่งเนื้อหาในบทนี้จะประกอบด้วยข้อมูลที่ใช้ในการทดสอบประสิทธิภาพการตรวจจับข่าวปลอม โดยแบ่งเป็นการทดสอบประสิทธิภาพการตรวจจับข่าวปลอมด้วย ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F – Measure) ซึ่งทั้งหมดจะถูกนำมาอธิบายผลลัพธ์ที่ได้ตามข้อมูลในส่วนที่สอง

ในการกระบวนการเตรียมข้อมูลเพื่อนำมาทดสอบการตรวจจับข่าวปลอมนั้น ข้อมูลที่นำมาจากฐานข้อมูลที่อยู่ในเว็บไซต์ Kaggle.com , Data Word และ UCI ซึ่งเป็นฐานข้อมูลที่ทำให้การเก็บรวบรวมข้อมูลทางด้านวิทยาศาสตร์ที่มีขนาดใหญ่และมีชื่อเสียงที่สุดในโลก ซึ่งช่วยให้นักวิจัยบรรลุเป้าหมายด้านวิทยาศาสตร์ ทั้งนี้ข้อมูลหลายประการ มีความท้าทายหลายประการที่เกี่ยวข้องกับการตรวจจับข่าวปลอม ความท้าทายทั้งหมดเหล่านี้มีชุดข้อมูล เพื่อทำการตรวจจับข่าวปลอม ข้อมูลข่าวที่รวบรวมโดยชุดข้อมูลเหล่านี้ มีการรวบรวมบันทึกข่าวทั้งหมด (WELFake) เป็นชุดข้อมูลประกอบด้วยจำนวน 5 แอปทริบิวท์ ผู้วิจัยได้รวมชุดข้อมูลข่าวยอดนิยม เช่น Kaggle, McIntire, Reuters, BuzzFeed Political: ดังนี้

- 1) หัวเรื่อง : ชื่อบทความ.
- 2) news_url : URL ของบทความ
- 3) โดเมนต้นทาง : โดเมนเว็บไซต์บทความ
- 4) tweet_num : จำนวนการทวีตสำหรับบทความนี้
- 5) Real : คอลัมน์ป้ายกำกับโดยที่ 1 เป็นของจริงและ 0 เป็นของปลอม

4.2 ขั้นตอนการเตรียมข้อมูล

การเตรียมข้อมูลเพื่อใช้ในการตรวจจับข่าวปลอม ซึ่งสามารถจำแนกได้เป็น 2 ขั้นตอน ซึ่งจะแสดงรายละเอียดในแต่ละขั้นตอน ดังนี้

4.2.1 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลได้จากฐานข้อมูล จำนวน 3 ฐานข้อมูล ผู้วิจัยนำข้อมูลดังกล่าวมาผ่านการวิเคราะห์และตรวจสอบความถูกต้อง สมบูรณ์ของข้อมูล พบว่าข้อมูลทั้ง 3 ฐานข้อมูลนั้นมีความถูกต้อง ครบถ้วน และสามารถนำไปใช้ในการตรวจจับข่าวปลอมได้จริง ดังตารางประกอบ 7

ตาราง 7 ฐานข้อมูลที่นำมาใช้ในการพยากรณ์ตรวจจับข่าวปลอม

ฐานข้อมูล	ขนาดของข้อมูล	การฝึก (70%)	การทดสอบ (30%)
Kaggle	20,717	14,502	6,215
Data word	19,530	13,671	5,859
UCI Machine Learning Repository	21,248	14,875	6,373

4.2.2 การเตรียมข้อมูล

ผู้วิจัยได้ทำการปรับข้อมูลตามกระบวนการตรวจสอบและแก้ไขรายงานข้อมูลที่ไม่ถูกต้องออกจากชุดข้อมูล (Data Cleansing) เพื่อให้ข้อมูลมีความถูกต้องมากที่สุด ซึ่งตัวอย่างข้อมูลที่ได้จากแบบสอบถามสามารถแสดงได้ดังภาพประกอบ 12

	A	B	C	D	E
1	title	news_url	source_domain	tweet_num	label
2	Kandi Burruss Explodes Over Rape Accusation on 'Real Housewives of Atlanta'	http://toofab.com/2017/05/08/real-housewives-of-atlanta-kandi-burruss-rape-accusation/	toofab.com	42	1
3	People's Choice Awards 2018: The best red carpet looks	https://www.today.com/style/see-people-choose-awards-2018-the-best-red-carpet-looks	www.today.com	0	1
4	Sophia Bush Sends Sweet Birthday Message to 'One Tree Hill' Co-Star Hilary Duff	https://www.etonline.com/news/220806_sophia-bush-sends-sweet-birthday-message-to-one-tree-hill-co-star-hilary-duff	www.etonline.com	63	1
5	Colombian singer Maluma sparks rumours of inappropriate relationship with Jennifer Lopez	https://www.dailymail.co.uk/news/article-5311111/Colombian-singer-Maluma-sparks-rumours-of-inappropriate-relationship-with-Jennifer-Lopez.html	www.dailymail.co.uk	20	1
6	Gossip Girl 10 Years Later: How Upper East Siders Shocked the World and Changed the Game	https://www.zerchoo.com/entertainment/gossip-girl-10-years-later-how-upper-east-siders-shocked-the-world-and-changed-the-game	www.zerchoo.com	38	1
7	Gwen Stefani Got Dumped by Blake Shelton Over "Jealousy and Drama"	www.intouchweekly.com/posts/gwen-stefani-got-dumped-by-blake-shelton-over-jealousy-and-drama	www.intouchweekly.com	45	0
8	Broward County Sheriff Fired For Lying About Parkland Shooting	https://yournewswire.com/broward-county-sheriff-fired-for-lying-about-parkland-shooting	yournewswire.com	124	0
9	Amber Rose Shuts Down French Montana Dating Rumors, Calls Rapper Her Friend	www.etonline.com/news/214798_amber-rose-shuts-down-french-montana-dating-rumors-calls-rapper-her-friend	www.etonline.com	4	0
10	Mindy Kaling makes first post-baby appearance at Disneyland with her 1-year-old son	https://www.aol.com/article/entertainment/celebrity-news/mindy-kaling-makes-first-post-baby-appearance-at-disneyland-with-her-1-year-old-son	www.aol.com	59	1
11	Katharine McPhee Butchers Tony Nominations: 'I Feel Like I Have Not Been Dressed'	https://www.98online.com/2018/05/02/katharine-mcphée-butchers-tony-nominations-i-feel-like-i-have-not-been-dressed	www.98online.com	10	1
12	'WAGS Miami' Stars Ashley Nicole Roberts and Philip Wheeler Are Engaged	https://www.usmagazine.com/celebrity-news/wags-miami-stars-ashley-nicole-roberts-and-philip-wheeler-are-engaged	www.usmagazine.com	18	1
13	Mel Gibson: Hollywood Pedophiles Have Nowhere Left To Hide	yournewswire.com/mel-gibson-hollywood-pedophiles-have-nowhere-left-to-hide	yournewswire.com	95	0
14	Medium Tyler Henry Addresses The 'Chilling' Messages Kristin Cavallari Sent Him	https://perezhillton.com/tyler-henry-says-kr-perezhillton.com	perezhillton.com	4	1
15	DWTS Season 27 Results: Week 5 - Disney Night	https://abc.go.com/shows/dancing-with-the-stars/dwts-season-27-results-week-5-disney-night	abc.go.com	0	1
16	5 Reasons Why Tarek El Moussa Will Overcome His Latest Back Injury	www.eonline.com/news/932478/5-reasons-why-tarek-el-moussa-will-overcome-his-latest-back-injury	www.eonline.com	63	0
17	David Cassidy cuts estranged daughter Katie completely out of his will	https://www.thesun.co.uk/news/5086439/david-cassidy-cuts-estranged-daughter-katie-completely-out-of-his-will	www.thesun.co.uk	54	1
18	Actress Brigitte Nielsen, 54, has given birth to her fifth child	https://www.usatoday.com/story/life/people/2018/05/02/brigitte-nielsen-gives-birth-fifth-child	www.usatoday.com	64	1
19	The Kardashians donate whopping sum to Las Vegas mass shooting victims	https://web.archive.org/web/20171228144400/http://www.abcnews.com/2017/12/28/kardashians-donate-whopping-sum-to-las-vegas-mass-shooting-victims/	web.archive.org	1	0
20	Is Brad Pitt Open To Getting Back Together With Angelina After Rumors of Divorce?	hollywoodlife.com/2017/08/27/would-brad-pitt-get-back-together-with-angelina-after-rumors-of-divorce/	hollywoodlife.com	19	0
21	Pink Announces 2018 Tour and Drops New Single	https://celebmix.com/pink-announces-2018-tour-and-drops-new-single	celebmix.com	120	1
22	A glimpse at the relationship of Meghan Markle, Kate Middleton, Prince Harry and William	www.sacbee.com/news/nation-world/world/article148484841.html	www.sacbee.com	8	0
23	Johnny Depp, Charlie Sheen and Kevin Dillon reunite for 'The Contender'	https://www.dailymail.co.uk/tvshowbiz/art-4648484.html	www.dailymail.co.uk	91	1
24	Are Katie Holmes & Jamie Foxx Expecting A Baby? New Report Claims They Are	hollywoodlife.com/2018/02/12/jamie-foxx-katie-holmes-expecting-a-baby-new-report-claims-they-are/	hollywoodlife.com	14	0
25	Ryan Phillippe shows off insanely chiseled bod shirtless in the snow	https://www.msn.com/en-us/tv/celebrity/ryan-phillippe-shows-off-insanely-chiseled-bod-shirtless-in-the-snow	www.msn.com	47	1
26	10 Pretty Celebrity Books That Would Look Good on Your Coffee Table	https://stylecaster.com/celeb-coffee-table/	stylecaster.com	49	1
27	Pink Panics Before Hanging Off L.A. Hotel for AMAs 2017 Performance	https://www.usmagazine.com/celebrity-news/pink-panics-before-hanging-off-la-hotel-for-amas-2017-performance	www.usmagazine.com	1	1

ภาพประกอบ 12 ตัวอย่างข้อมูลนำมาจาก Kaggle.com

4.3 การวัดประสิทธิภาพของอัลกอริทึม

การทดสอบและการคัดเลือกอัลกอริทึม จะสามารถวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F – Measure) ในการตรวจจับข่าวปลอม โดยมีรายละเอียดของวิธีการวัดประสิทธิภาพ ดังนี้

การวัดประเมิณผลมีความสำคัญเนื่องจกนำมาใช้ในการวิเคราะห์ประสิทธิภาพการทำงานของอัลกอริทึม สำหรับการวัดผลของอัลกอริทึมที่ใช้ในการจำแนกข้อมูล ตัวชี้วัดการประเมินผลที่สามารถที่จะพัฒนาจากเมทริกซ์ความสับสน (Confusion matrix) ซึ่งเป็นตาราง crosstabs ขนาด 2*2 โดยมีการกำหนดให้แนวนอน คือค่า Actual result และแนวตั้งคือ Prediction result ดังตาราง 8

ตาราง 8 ตารางเมทริกซ์ความสับสน (Confusion matrix)

		ผลการจำแนก	
		คำตอบเป็นบวก (P)	คำตอบเป็นลบ (N)
ค่าที่แท้จริง	คำตอบเป็นบวก (P)	True positives (TP)	False negative (FN)
	คำตอบเป็นลบ (N)	False positives (FP)	True negative (TN)

โดยมีคำจำกัดความดังนี้

1. True Positive (TP) คือ ผลของการตรวจจับข่าวปลอม พบว่า ข่าวที่โปรแกรมทำนายว่า “จริง” และผลที่เกิดขึ้นจริง พบว่า “จริง”
2. True Negative (TN) คือ ผลของการตรวจจับข่าวปลอม พบว่า ข่าวที่โปรแกรมทำนายว่า “ไม่จริง” และผลที่เกิดขึ้นพบว่า “ไม่จริง”
3. False Positive (FP) คือ ผลของการตรวจจับข่าวปลอม พบว่า ข่าวที่โปรแกรมทำนายว่า “จริง” และผลที่เกิดขึ้นพบว่า “ไม่จริง”
4. False Negative (FN) คือ และผลที่เกิดขึ้นพบว่า “ไม่จริง” และผลที่เกิดขึ้นพบว่า เป็น “จริง”
5. Condition positive (P) คือ จำนวนของผลที่แท้จริงที่มีความจริงต่อการตรวจจับข่าวปลอมทั้งหมด ที่อยู่ในข้อมูล = TP + FN
6. Condition negative (N) คือ จำนวนของผลที่แท้จริงที่ไม่จริงต่อการตรวจจับข่าวปลอมทั้งหมด ที่อยู่ในข้อมูล = FP + TN

จากตาราง 8 สามารถนำมาคำนวณค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึก และค่าความถ่วงดุล ได้ดังนี้

1. ค่าความถูกต้อง (Accuracy) คือ เป็นการวัดค่าของความถูกต้องของโมเดล จำนวนของข้อมูลที่ทำนายถูกทุกคลาส โดยพิจารณาทุกคลาส สามารถแสดงในสมการที่ 26

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (26)$$

2. ค่าความแม่นยำ (Precision) คือ เป็นการวัดค่าความแม่นยำของโมเดล ที่ดูจากสิ่งทำนายออกมาแล้วหาได้กี่เปอร์เซ็นต์ สามารถแสดงในสมการที่ 27

$$\text{Precision} = \frac{TP}{TP + FN} \quad (27)$$

3. ค่าความระลึก (Recall) คือ เป็นการวัดค่าความระลึกของโมเดล โดยพิจารณาจำนวนที่ทำนายถูกกี่ตัว เป็นการวัดความถูกต้องของโมเดล สามารถแสดงในสมการที่ 28

$$\text{Recall} = \frac{TP}{TP + FN} \quad (28)$$

4. ค่าความถ่วงดุล (F – Measure) คือ เป็นการวัดค่าของ Precision และ Recall พร้อมกันของโมเดล จากนั้นพิจารณาค่าเฉลี่ยของค่าความแม่นยำและค่าระลึก สามารถแสดงในสมการที่ 29

$$F\text{-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (29)$$

4.4 การทดสอบและการเลือกอัลกอริทึม

ขั้นตอนในการดำเนินการวิจัยในการทดสอบประสิทธิภาพการตรวจจับข่าวปลอมของทั้ง 5 อัลกอริทึม จะสามารถทำได้โดยการนำชุดข้อมูล จำนวน 3 ชุด เข้าตัวแบบทดสอบเพื่อให้สามารถตอบวัตถุประสงค์ของงานวิจัย การนำข้อมูลที่ได้จากฐานข้อมูล Kaggle เพื่อป้อนเข้าสู่ระบบผ่านขั้นตอนการทำความสะอาดข้อมูลเรียบร้อยแล้วเข้าทดสอบกับ 5 อัลกอริทึม ดังนี้

1. นาอีฟเบย์ (Naïve bayes)
2. ต้นไม้ตัดสินใจ (Decision tree)
3. ต้นไม้แห่งการทำนาย (Random forest)
4. การถดถอยโลจิสติก (Logistic Regression)
5. ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines)

4.4.1 ทดสอบเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม

โดยกำหนดการทดสอบแต่ละอัลกอริทึม เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม และคัดเลือกอัลกอริทึมที่เหมาะสมในการตรวจจับข่าวปลอมที่เหมาะสมที่สุด โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F – Measure)

- 1) นาอีฟเบย์ (Naïve bayes)

เป็นหนึ่งในอัลกอริทึมที่อาศัยหลักการความน่าจะเป็นตามทฤษฎีของเบย์ (Bayes theorem) ซึ่งมีอัลกอริทึมที่ไม่ซับซ้อน โดยใช้ขั้นตอนในการจำแนกเพื่อการเรียนรู้ปัญหาที่เกิดขึ้น แล้วนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ โดยใช้หลักการในการคำนวณหาความน่าจะเป็นในการทำนายผลเป็นอัลกอริทึมในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ และทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละ

ความสัมพันธ์ เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 20,716 รายการ และทำการวิเคราะห์ ด้วยวิธีนาอ์ฟเบย์ ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 9

ตาราง 9 ผลลัพธ์ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอ์ฟเบย์

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	91.29	94.89	94.79	94.84
True	2332	912				
Fake	893	16579				
Class recall	72.31	94.79				

จากตาราง 9 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอ์ฟเบย์ ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 91.29% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 94.89% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 94.79% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 94.84% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดีมาก

2) ต้นไม้ตัดสินใจ (Decision tree)

เป็นหนึ่งในอัลกอริทึมที่ช่วยให้วิเคราะห์เหตุการณ์ หรือสถานการณ์เพื่อการตัดสินใจ ได้อย่างเป็นระบบและรวดเร็ว ต้นไม้การตัดสินใจมีลักษณะเป็นกราฟรูปลูกไม้ ซึ่งแสดงที่ตั้งต้นที่มีราก และแขนงต่างๆ แยกออกมาจากต้นไม้ออกไปในทิศทางเดียว จนกระทั่งนำไปสู่ข้อสรุปสำหรับการตัดสินใจได้ ต้นไม้การตัดสินใจมีประโยชน์ในการสรุปการตัดสินใจที่มีความซับซ้อนให้ง่ายต่อความเข้าใจ เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 20,716 รายการ และทำการวิเคราะห์ด้วยวิธีต้นไม้ตัดสินใจ ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 10

พูน ปณ จิต ชเว

ตาราง 10 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	97.35	97.06	99.57	98.45
True	2695	20				
Fake	530	17471				
Class recall	83.57	99.89				

จากตาราง 10 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 97.35% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 97.06% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 99.57% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 98.45% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดีมาก

3) ต้นไม้แห่งการทำนาย (Random Forest)

เป็นหนึ่งในอัลกอริทึมที่อยู่ในกลุ่มของการสร้างโมเดลที่เรียกว่า Ensemble learning ที่มีหลักการคือการเทรนโมเดลที่เหมือนกันหลายๆ ครั้ง (หลาย Instance) บนข้อมูลชุดเดียวกัน โดยแต่ละครั้งของการเทรนจะเลือกส่วนของข้อมูลที่เทรนไม่เหมือนกัน แล้วเอาการตัดสินใจของโมเดลเหล่านั้นมาโหวตกันว่า Class ไหนถูกเลือกมากที่สุด เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 20,716 รายการ และทำการวิเคราะห์ด้วยวิธีต้นไม้แห่งการทำนาย ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 11

ตาราง 11 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้แห่งการทำนาย

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	98.05	97.74	100.00	98.86
True	2820	0				
Fake	405	17491				
Class recall	87.44	100.00				

จากตาราง 11 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้แห่งการทํานาย ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 98.05% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 97.74% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 100% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 98.86% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดีมาก

4) การถดถอยโลจิสติก (Logistic Regression)

เป็นหนึ่งในอัลกอริทึมที่นำมาวิเคราะห์ตัวแปรเชิงพหุที่มีวัตถุประสงค์เพื่อประมาณค่าหรือทำนายเหตุการณ์ที่สนใจว่าจะเกิดหรือไม่เกิดเหตุการณ์นั้นภายใต้อิทธิของตัวปัจจัยแบบจำลองโลจิสติกประกอบด้วยตัวแปรตาม (หรือตัวแปรเกณท์) ที่ต้องเป็นตัวแปรแบบทวินาม (Dichotomous Variable) เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 20,716 รายการ และทำการวิเคราะห์ด้วยวิธีการถดถอยโลจิสติก ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 12

ตาราง 12 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	94.49	95.36	98.26	96.78
True	2388	305				
Fake	837	17186				
Class recall	74.05	98.26				

จากตาราง 12 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 94.49% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 95.36% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 98.26% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 96.78% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดีมาก

5) ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines)

เป็นหนึ่งในอัลกอริทึมของวิธีการเรียนรู้ของเครื่อง (Machine learning) ประเภทแบบจำลองการเรียนรู้ภายใต้การกำกับ (Supervised learning) ด้วยขั้นตอนวิธี (Algorithm) ที่ใช้ในการวิเคราะห์ข้อมูลเพื่อจำแนกประเภท (Classification) เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 20,716 รายการ และทำการวิเคราะห์ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 13

ตาราง 13 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	99.67	99.61	100.00	99.81
True	3157	0				
Fake	68	17491				
Class recall	97.89	100.00				

จากตาราง 13 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 99.67% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 99.61% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 100.00% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 99.81% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดีมาก จากที่กล่าวมาข้างต้นจะพบว่าการทดสอบแต่ละอัลกอริทึมทั้ง 5 อัลกอริทึมนั้นจะได้ผลลัพธ์ดังสรุปในตาราง 14

พูน ปณ จิต ชีเว

ตาราง 14 แสดงผลลัพธ์เปรียบเทียบทั้ง 5 อัลกอริทึมในการวิเคราะห์ข้อมูลการตรวจจับข่าวปลอม

อัลกอริทึม	ประสิทธิภาพ			
	ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
นาอ์ฟเบย์	91.29	94.89	94.79	94.84
ต้นไม้ตัดสินใจ	97.35	97.06	99.57	98.45
ต้นไม้แห่งการทำนาย	98.05	97.74	100.00	98.86
การถดถอยโลจิสติก	94.49	95.36	98.26	96.78
ซัพพอร์ตเวกเตอร์แมชชีน	99.67	99.61	100.00	99.81

จากตาราง 14 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลการตรวจจับข่าวปลอม จากทั้ง 5 อัลกอริทึม พบว่า ซัพพอร์ตเวกเตอร์แมชชีน มีค่าความถูกต้องในการตรวจจับข่าวปลอมมากที่สุด ที่ 99.67% ค่าความแม่นยำที่ 99.61% ค่าระลึกที่ 100.00% และค่าถ่วงดุลที่ 99.81% จึงสรุปได้ว่า ซัพพอร์ตเวกเตอร์แมชชีน มีประสิทธิภาพในการตรวจจับข่าวปลอมดีที่สุดสำหรับการนำข้อมูลข่าวปลอม จำนวน 20,716 รายการ

4.4.2 ทดสอบอัลกอริทึมเพื่อนำไปประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย

กำหนดการทดสอบแต่ละอัลกอริทึม เพื่อนำอัลกอริทึมไปประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F – Measure) นำข้อมูลที่ได้จากฐานข้อมูล Data Word เพื่อป้อนเข้าสู่ระบบผ่านขั้นตอนการทำความสะอาดข้อมูลเรียบร้อยแล้วทำการทดสอบ ดังนี้

1) นาอ์ฟเบย์ (Naïve bayes)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 19,530 รายการ และทำการวิเคราะห์ ด้วยวิธี นาอ์ฟเบย์ ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 15

ตาราง 15 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอ์ฟเบย์

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	71.84	75.26	87.12	80.75
True	2491	1706				
Fake	3794	11539				
Class recall	39.63	87.12				

จากตาราง 15 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอ์ฟเบย์ ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 71.84% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 75.26% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 87.12% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 80.75% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดี

2) ต้นไม้ตัดสินใจ (Decision tree)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 19,530 รายการ และทำการวิเคราะห์ด้วยวิธีต้นไม้ตัดสินใจ ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 16

ตาราง 16 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	75.52	73.87	98.91	84.57
True	1650	145				
Fake	4635	13100				
Class recall	26.25	98.91				

จากตาราง 16 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 75.52% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 73.87% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 98.91% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 84.57% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดี

3) ต้นไม้แห่งการทำนาย (Random Forest)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 19,530 รายการ และทำการวิเคราะห์ด้วยวิธีต้นไม้แห่งการทำนาย ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 17

ตาราง 17 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้แห่งการทำนาย

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	74.13	74.87	93.10	82.99
True	2146	914				
Fake	4139	12331				
Class recall	34.14	93.10				

จากตาราง 17 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้แห่งการทำนาย ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 74.13% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 74.87% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 93.10% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 82.99% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดีมาก

4) การถดถอยโลจิสติก (Logistic Regression)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 19,530 รายการ และทำการวิเคราะห์ด้วยวิธีการถดถอยโลจิสติก ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 18

ตาราง 18 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	74.13	74.87	93.10	82.99
True	2146	914				
Fake	4139	12331				
Class recall	34.14	93.10				

จากตาราง 18 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 74.13% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 74.87% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 93.10% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 82.99% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดี

5) ซัพพอร์ทเวกเตอร์แมชชีน (Support vector machines)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 19,530 รายการ และทำการวิเคราะห์ด้วยวิธีซัพพอร์ทเวกเตอร์แมชชีน ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 19

ตาราง 19 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ทเวกเตอร์แมชชีน

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	73.11	74.23	92.47	82.35
True	2023	998				
Fake	4253	12247				
Class recall	32.33	92.47				

จากตาราง 19 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ทเวกเตอร์แมชชีน ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 73.11% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอม

ถูกต้องเท่ากับ 74.23% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 92.47% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 82.35% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดีมาก

กำหนดการทดสอบแต่ละอัลกอริทึม เพื่อนำอัลกอริทึมไปประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F – Measure) นำข้อมูลที่ได้จากฐานข้อมูล UCI Machine Learning Repository เพื่อป้อนเข้าสู่ระบบผ่านขั้นตอนการทำความสะอาดข้อมูลเรียบร้อยแล้วทำการทดสอบ ดังนี้

1) นาอ์ฟเบย์ (Naïve bayes)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 21,248 รายการ และทำการวิเคราะห์ ด้วยวิธีนาอ์ฟเบย์ ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 20

ตาราง 20 ผลลัพธ์ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอ์ฟเบย์

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	81.10	86.20	89.70	87.91
True	2627	1677				
Fake	2339	14605				
Class recall	52.90	89.70				

จากตาราง 20 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอ์ฟเบย์ ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 81.10% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 86.20% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 89.70% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 87.91% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดี

2) ต้นไม้ตัดสินใจ (Decision tree)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 21,248 รายการ และทำการวิเคราะห์ด้วยวิธีต้นไม้ตัดสินใจ ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 21

ตาราง 21 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	88.74	87.96	98.85	93.08
True	2762	188				
Fake	2204	16094				
Class recall	55.62	98.85				

จากตาราง 21 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 88.74% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 87.96% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 98.85% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 93.08% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดี

3) ต้นไม้แห่งการทำนาย (Random Forest)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 21,248 รายการ และทำการวิเคราะห์ด้วยวิธีต้นไม้แห่งการทำนาย ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 22

ตาราง 22 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้แห่งการทำนาย

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	89.22	88.35	98.99	93.36
True	2840	165				
Fake	2126	16117				
Class recall	57.19	98.99				

จากตาราง 22 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้แห่งการทำนาย ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 89.22% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้อง

เท่ากับ 88.35% ให้ค่าระลีกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 98.99% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 93.36% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดีมาก

4) การถดถอยโลจิสติก (Logistic Regression)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 21,248 รายการ และทำการวิเคราะห์ด้วยวิธีการถดถอยโลจิสติก ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 23

ตาราง 23 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก

ตัวแปร			ประสิทธิภาพ			
			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลีก (%)	ค่าถ่วงดุล (%)
	True	Fake	84.29	85.31	96.03	90.35
True	2273	646				
Fake	2693	15636				
Class recall	45.77	96.03				

จากตาราง 23 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 84.29% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 85.31% ให้ค่าระลีกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 96.03% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 90.35% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดี

5) ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines)

เมื่อทำการนำข้อมูลป้อนเข้า จำนวน 21,248 รายการ และทำการวิเคราะห์ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน ทำให้ได้ผลลัพธ์ดังที่แสดงในตาราง 24

พูน ปณ จิต ชเว

ตาราง 24 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

ตัวแปร			ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	82.17	84.62	93.81	88.96
True	2186	1008				
Fake	2780	15274				
Class recall	44.02	93.81				

จากตาราง 24 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 82.17% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 84.62% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 93.81% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 88.96% แสดงประสิทธิภาพการทดสอบที่มีความถูกต้องแม่นยำที่ดีมากจากการทดสอบข้อมูลการตรวจจับข่าวปลอมที่ส่งผลต่อการตรวจจับข่าวปลอม เมื่อทำการเปรียบเทียบจากฐานข้อมูลทั้ง 3 ฐานข้อมูล ดังสามารถแสดงได้ในตาราง 25

ตาราง 25 แสดงผลลัพธ์เปรียบเทียบทั้ง 3 ฐานข้อมูลในการวิเคราะห์ข้อมูลการตรวจจับข่าวปลอม

อัลกอริทึม	ประสิทธิภาพ			
	ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
Kaggle dataset				
นาอูฟเบย์	91.29	94.89	94.79	94.84
ต้นไม้ตัดสินใจ	97.35	97.06	99.57	98.45
ต้นไม้แห่งการทำนาย	98.05	97.74	100.00	98.86
การถดถอยโลจิสติก	94.49	95.36	98.26	96.78
ซัพพอร์ตเวกเตอร์แมชชีน	99.67	99.61	100.00	99.81

ตาราง 25 แสดงผลลัพธ์เปรียบเทียบทั้ง 3 ฐานข้อมูล (ต่อ)

อัลกอริทึม	ประสิทธิภาพ			
	ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
Data Word dataset				
นาอ์ฟเบย์	71.84	75.26	87.12	80.75
ต้นไม้ตัดสินใจ	75.52	73.87	98.91	84.57
ต้นไม้แห่งการทำนาย	74.13	74.87	93.10	82.99
การถดถอยโลจิสติก	74.13	74.87	93.10	82.99
ซัพพอร์ตเวกเตอร์แมชชีน	73.11	74.23	92.47	82.35
ซัพพอร์ตเวกเตอร์แมชชีน	84.29	85.31	96.03	90.35
UCI Machine Learning Repository				
นาอ์ฟเบย์	81.10	86.20	98.70	87.91
ต้นไม้ตัดสินใจ	82.17	84.62	93.81	88.96
ต้นไม้แห่งการทำนาย	88.74	87.96	98.85	93.08
การถดถอยโลจิสติก	89.22	88.35	98.99	93.36

จากตาราง 25 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลการตรวจจับขบวนการปลอม จำนวน 3 ฐานข้อมูลพบว่า ฐานข้อมูลของ Kaggle ซึ่งประกอบด้วย 5 อัลกอริทึม ประกอบด้วย นาอ์ฟเบย์ มีความถูกต้องที่ 91.29 ต้นไม้ตัดสินใจ มีความถูกต้องที่ 97.35 ต้นไม้แห่งการทำนาย มีความถูกต้องที่ 98.05 การถดถอยโลจิสติก มีความถูกต้องที่ 94.49 และซัพพอร์ตเวกเตอร์แมชชีน มีความถูกต้องที่ 99.67 จึงสรุปได้ว่า ฐานข้อมูลของ Kaggle มีประสิทธิภาพในการตรวจจับขบวนการปลอมดีที่สุดสำหรับการนำเข้าสู่ข้อมูลของขบวนการปลอม

4.4.3 ปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม

กำหนดการทดสอบอัลกอริทึมจากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน ผู้วิจัยได้ทำการปรับปรุงและทำการคัดเลือกปัจจัยหรือการคัดเลือกคุณสมบัติ ที่ช่วยลดจำนวนตัวแปรในการพยากรณ์ หรือเป็นการเลือกตัวแปรที่ดีที่สุดเพียงตัวเดียว เพื่อให้การสร้างตัวแบบพยากรณ์มีประสิทธิภาพ ลดมิติของข้อมูล และช่วยให้การวิเคราะห์ดำเนินการได้เร็วขึ้น และมีประสิทธิภาพมากขึ้น โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F – Measure) นำข้อมูลที่ได้จากฐานข้อมูลเพื่อป้อนเข้าสู่ระบบแล้วทำการทดสอบ ดังตาราง 26

ตาราง 26 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ

ตัวแปร			การปรับปรุงพารามิเตอร์	ประสิทธิภาพ			
				ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	- Optimize Parameter	99.66	99.60	100	99.80
True	3155	0	- SVM.kernel				
Fake	70	17491	_type-Radial				
Class recall	97.83	100					

จากตาราง 26 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ โดยที่ผู้วิจัยได้ทำการ Optimize Parameter โดยเลือก SVM.kernel_type กำหนด Kernel เป็น Radial หลังจากทำการปรับปรุงอัลกอริทึม ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 99.66% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 99.60% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 100% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 99.80% แสดงประสิทธิภาพจากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนที่มีความถูกต้องแม่นยำเพิ่มขึ้นที่ดีมาก

ตาราง 27 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ

ตัวแปร			การปรับปรุงพารามิเตอร์	ประสิทธิภาพ			
				ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	- Optimize	91.63	91.27	99.67	95.27
True	1549	57	Parameter				
Fake	1676	17434	- SVM.kernel				
Class recall	48.03	99.67	_type -polynomial				

จากตาราง 27 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ โดยที่ผู้วิจัยได้ทำการ Optimize Parameter โดยเลือก SVM.kernel_type กำหนด Kernel เป็น polynomial หลังจากทำการปรับปรุงอัลกอริทึม ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 91.63% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 91.27% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 99.67% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 95.27% แสดงประสิทธิภาพจากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนที่มีความถูกต้องแม่นยำเพิ่มขึ้นที่ดีมาก

ตาราง 28 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ

ตัวแปร			การปรับปรุงพารามิเตอร์	ประสิทธิภาพ			
				ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	- Optimize	70.48	86.56	76.89	80.01
True	1152	4042	Parameter				
Fake	2073	13449	- SVM.kernel				
Class recall	35.72	76.89	_type - neural				

จากตาราง 28 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ โดยที่ผู้วิจัยได้ทำการ Optimize Parameter โดยเลือก SVM.kernel_type กำหนด Kernel เป็น neural หลังจากทำการปรับปรุงอัลกอริทึม ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 70.48% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 86.56% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 76.89% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 80.01% แสดงประสิทธิภาพจากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนที่มีความถูกต้องแม่นยำเพิ่มขึ้นที่ดีมาก

ตาราง 29 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ

ตัวแปร			การปรับปรุงพารามิเตอร์	ประสิทธิภาพ			
				ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	- Optimize Parameter	98.92	99.39	99.33	99.36
True	3119	117	- SVM.kernel				
Fake	106	17374	_type				
Class recall	96.71	99.33	- anova				

จากตาราง 29 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ โดยที่ผู้วิจัยได้ทำการ Optimize Parameter โดยเลือก SVM.kernel_type กำหนด Kernel เป็น anova หลังจากทำการปรับปรุงอัลกอริทึม ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 98.92% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 99.39% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 99.33% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 99.36% แสดงประสิทธิภาพจากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนที่มีความถูกต้องแม่นยำเพิ่มขึ้นที่ดีมาก

ตาราง 30 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ

ตัวแปร			การปรับปรุงพารามิเตอร์	ประสิทธิภาพ			
				ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	- Optimize Parameter	99.70	99.61	100	99.81
True	3157	0	- SVM.kernel				
Fake	68	17491	_type				
Class recall	97.89	100	- epachnenikov				

จากตาราง 30 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ โดยที่ผู้วิจัยได้ทำการ Optimize Parameter โดยเลือก SVM.kernel_type กำหนด Kernel เป็น epachnenikov หลังจากทำการปรับปรุงอัลกอริทึม ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 99.70% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 99.61% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 100% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 99.81% แสดงประสิทธิภาพจากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนที่มีความถูกต้องแม่นยำเพิ่มขึ้นที่ดีมาก

ตาราง 31 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ

ตัวแปร			การปรับปรุงพารามิเตอร์	ประสิทธิภาพ			
				ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
	True	Fake	- Optimize Parameter	84.43	84.43	100.00	91.56
True	3157	0	- SVM.kernel				
Fake	68	17491	_type				
Class recall	97.89	100	multiquadric				

จากตาราง 31 ผลลัพธ์ที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ทเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ โดยที่ผู้วิจัยได้ทำการ Optimize Parameter โดยเลือก SVM.kernel_type กำหนด Kernel เป็น multiquadric หลังจากทำการปรับปรุงอัลกอริทึม ให้ค่าความถูกต้องต่อการตรวจจับข่าวปลอมเท่ากับ 84.43% ให้ค่าความแม่นยำในการตรวจจับข่าวปลอมถูกต้องเท่ากับ 84.43% ให้ค่าระลึกในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 100.00% และให้ค่าความถ่วงดุลถูกต้องเท่ากับ 91.56% แสดงประสิทธิภาพจากการปรับปรุงอัลกอริทึมซัพพอร์ทเวกเตอร์แมชชีนที่มีความถูกต้องแม่นยำเพิ่มขึ้นที่ดีมาก

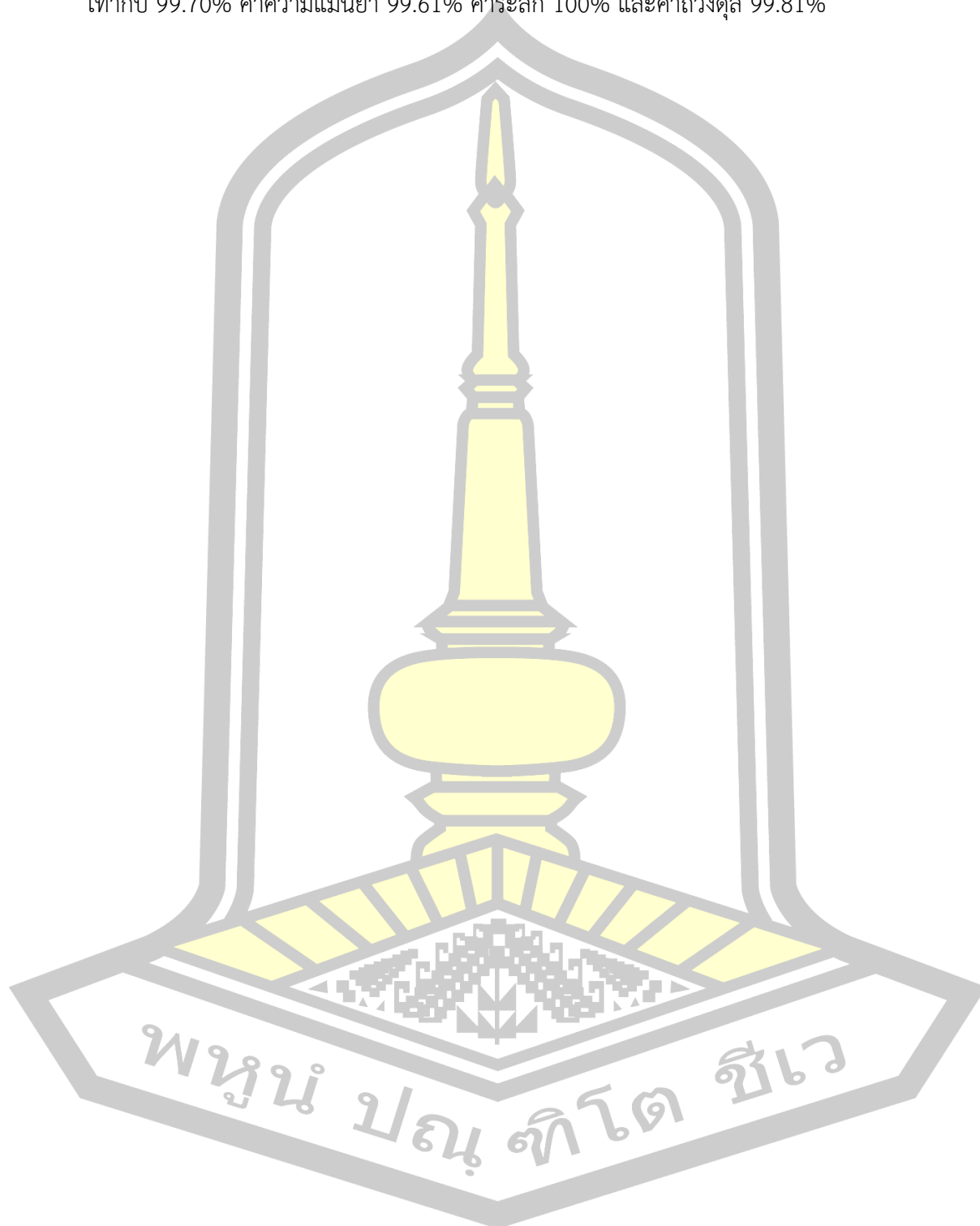
จากการทดสอบข้อมูลการตรวจจับข่าวปลอมของอัลกอริทึมซัพพอร์ทเวกเตอร์แมชชีนที่ส่งผลต่อการตรวจจับข่าวปลอม เมื่อทำการปรับปรุงอัลกอริทึมซัพพอร์ทเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ ดังสามารถแสดงได้ในตาราง 32

ตาราง 32 แสดงผลลัพธ์เปรียบเทียบจากการปรับปรุงอัลกอริทึมซัพพอร์ทเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ

อัลกอริทึมและการปรับปรุงพารามิเตอร์	ประสิทธิภาพ			
	ความถูกต้อง (%)	ความแม่นยำ (%)	ค่าระลึก (%)	ค่าถ่วงดุล (%)
SVM.kernel_type (radial)	99.66	99.60	100	99.80
SVM.kernel_type (polynomial)	91.63	91.27	99.67	95.27
SVM.kernel_type (neural)	70.48	86.56	76.89	80.01
SVM.kernel_type (anova)	98.92	99.39	99.33	99.36
SVM.kernel_type (epachnenikov)	99.70	99.61	100	99.81
SVM.kernel_type (multiquadric)	84.43	84.43	100	91.56

จากตาราง 32 จากผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลจากการเปรียบเทียบที่ได้จากการปรับปรุงอัลกอริทึมซัพพอร์ทเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ โดยที่ผู้วิจัยได้ทำการ Optimize Parameter หลังจากทำการปรับปรุงอัลกอริทึมแล้ว พบว่า ค่าความถูกต้องต่อการตรวจจับข่าวปลอมของอัลกอริทึมซัพพอร์ทเวกเตอร์ แมชชีน เพิ่มขึ้นเพิ่มมากขึ้นทุก parameter แสดงถึงประสิทธิภาพจากการปรับปรุงอัลกอริทึมซัพพอร์ทเวกเตอร์แมชชีนที่มีความถูกต้องแม่นยำเพิ่มขึ้นที่ดีมาก จึงสรุปได้ว่าหลังจากที่มีการปรับปรุงอัลกอริทึมของ SVM.kernel_type (epachnenikov) มี

ประสิทธิภาพในการตรวจจับข่าวปลอมที่ดีที่สุดสำหรับการนำเข้าสู่ข้อมูลของข่าวปลอม มีค่าความถูกต้องเท่ากับ 99.70% ค่าความแม่นยำ 99.61% ค่าระลึก 100% และค่าถ่วงดุล 99.81%



บทที่ 5

สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ

ในปัจจุบันมีการนำเทคนิคของตัววิธีการต่าง ๆ เพื่อเพิ่มความแม่นยำในการตรวจจับข่าวปลอม ผู้วิจัยใช้ข้อมูลข่าวบนทวิตเตอร์มีการดำเนินการจัดเก็บข้อมูลเกี่ยวกับข่าวปลอมจากฐานข้อมูลที่อยู่บนเว็บไซต์ Kaggle.com ซึ่งฐานข้อมูล Kaggle เป็นหนึ่งของฐานข้อมูลในชุมชนวิทยาศาสตร์ที่มีขนาดใหญ่และมีชื่อเสียงที่สุดในโลก ซึ่งช่วยให้นักวิจัยบรรลุเป้าหมายด้านวิทยาศาสตร์ ทั้งนี้มีข้อมูลหลากหลายประการ มีความท้าทายเป็นอย่างมากที่นำมาใช้ในการวิเคราะห์การตรวจจับข่าวปลอม

วัตถุประสงค์ของงานวิจัยวิทยานิพนธ์นี้คือ เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม เพื่อนำอัลกอริทึมมาประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย เพื่อปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม ซึ่งในงานวิจัยในครั้งนี้ใช้เทคนิคของแมชชีนเลิร์นนิง โดยใช้อัลกอริทึมของซัพพอร์ตเวกเตอร์แมชชีนมาประยุกต์ใช้เพื่อนำความสามารถของซัพพอร์ตเวกเตอร์แมชชีนเข้ามาช่วยในการวิเคราะห์ข้อมูลสำหรับการตรวจจับข่าวปลอมที่เกิดขึ้น ผลงานวิจัยนี้สามารถพัฒนากระบวนการการสร้างโมเดลของซัพพอร์ตเวกเตอร์แมชชีนที่เลือกพารามิเตอร์ที่เหมาะสม และสามารถตรวจจับข่าวปลอมบนทวิตเตอร์ได้อย่างมีประสิทธิภาพมากกว่าโมเดลแบบทั่วไป และสามารถนำโมเดลที่ได้ไปประยุกต์ใช้กับฐานข้อมูลที่มีความหลากหลาย และท้ายสุดจะแสดงให้เห็นว่า ในบรรดาอัลกอริทึมในแมชชีนเลิร์นนิงต่าง เช่น นาอ์ฟเบย์, ต้นไม้ตัดสินใจ, ต้นไม้แห่งการทํานาย, การถดถอยโลจิสติก และ ซัพพอร์ตเวกเตอร์แมชชีน ที่มีเคอร์เนล epachnenikov นำไปสู่ผลลัพธ์ที่ดีที่สุด

ในการสรุปผลการศึกษาระหว่างการสรุปผลเป็นลำดับ ดังนี้

5.1 ส่วนสรุปผล

5.1.1 ผลการศึกษาตามวัตถุประสงค์ของงานวิจัย

5.1.1.1 เปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม

5.1.1.2 เพื่อนำอัลกอริทึมมาประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย

5.1.1.3 ปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม

5.2 ส่วนอภิปรายผล

5.3 ส่วนข้อจำกัดทางการวิจัยและข้อเสนอแนะในการทำวิจัยครั้งต่อไป

5.1 ส่วนสรุปผล

ผู้วิจัยขอเสนอผลการศึกษา เรื่องการปรับปรุงการตรวจจับข่าวปลอมด้วยซอฟต์แวร์แมชชีนโดยวิธีการคัดเลือกคุณลักษณะที่เหมาะสมตามลำดับความสำคัญของวัตถุประสงค์ ดังนี้

5.1 เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการตรวจจับข่าวปลอม

งานวิจัยครั้งนี้ใช้เทคนิคแมชชีนเลิร์นนิงเพื่อตรวจจับข่าวปลอม ผู้วิจัยได้ทบทวนวรรณกรรมและใช้อัลกอริทึมจำนวน 5 อัลกอริทึม ประกอบด้วย นาอ์ฟเบย์ ต้นไม้ตัดสินใจ ต้นไม้แห่งการทํานาย, การถดถอยโลจิสติก และ ซัพพอร์ตเวกเตอร์แมชชีนในการตรวจจับข่าวปลอม พบว่า นาอ์ฟเบย์ให้ค่าความถูกต้องเท่ากับ 91.29% มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 94.89% วิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องเท่ากับ 97.35% มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 97.06% วิธีต้นไม้แห่งการทํานายให้ค่าความถูกต้องเท่ากับ 98.05% มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 97.74% วิธีการถดถอยโลจิสติกให้ค่าความถูกต้องเท่ากับ 94.49 มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 95.36% และวิธีซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องเท่ากับ 96.67 มีค่าความแม่นยำในการตรวจจับข่าวปลอมได้ถูกต้องเท่ากับ 99.61% จากผลการทดสอบพบว่า วิธีวิธีซัพพอร์ตเวกเตอร์ แมชชีนให้ค่าความถูกต้องในการตรวจจับข่าวปลอมมากที่สุดเท่ากับ 96.67 %

5.2 เพื่อนำอัลกอริทึมมาประยุกต์ใช้กับฐานข้อมูลที่หลากหลาย

งานวิจัยครั้งนี้ใช้ฐานข้อมูลที่มีบทความของข่าวปลอม จำนวน 3 ฐานข้อมูล ประกอบด้วย Kaggle, Data word และ UCI Machine Learning Repository มาใช้ในการวิเคราะห์ข้อมูล พบว่า ฐานข้อมูลของ Kaggle ซึ่งประกอบด้วย 5 อัลกอริทึม ประกอบด้วย นาอ์ฟเบย์ มีค่าความถูกต้องที่ 91.29% ต้นไม้ตัดสินใจ มีค่าความถูกต้องที่ 97.35% ต้นไม้แห่งการทํานาย มีค่าความถูกต้องที่ 98.05% การถดถอยโลจิสติก มีความถูกต้องที่ 94.49% และซัพพอร์ตเวกเตอร์แมชชีน มีค่าความถูกต้องที่ 99.67% จึงสรุปได้ว่า ฐานข้อมูลของ Kaggle มีประสิทธิภาพในการตรวจจับข่าวปลอมดีที่สุดสำหรับการนำเข้าสู่ข้อมูลของข่าวปลอม

5.3 เพื่อปรับปรุงอัลกอริทึมด้วยวิธีการคัดเลือกคุณลักษณะที่เหมาะสม

กำหนดการทดสอบอัลกอริทึมจากการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน ทำการปรับปรุงและคัดเลือกปัจจัยหรือการคัดเลือกคุณสมบัติที่ช่วยลดจำนวนตัวแปรในการพยากรณ์ หรือเป็นการเลือกตัวแปรที่ดีที่สุดเพียงตัวเดียว เพื่อให้การสร้างตัวแบบพยากรณ์มีประสิทธิภาพ ลดมิติของข้อมูล และช่วยให้การวิเคราะห์ดำเนินการได้เร็วขึ้น และมีประสิทธิภาพมากขึ้น พบว่า จาก การปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ โดยที่ผู้วิจัยได้ทำการ Optimize Parameter หลังจากที่ทำกรปรับปรุงอัลกอริทึมแล้ว ให้ค่าความถูกต้องต่อการ

ตรวจจับข่าวปลอมของอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน เพิ่มมากขึ้นทุก parameter แสดงถึงประสิทธิภาพจากการปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนที่มีความถูกต้องแม่นยำเพิ่มขึ้นที่ดีมาจากเดิม 99.67 จึงสรุปได้ว่าหลังจากที่มีการปรับปรุงอัลกอริทึมของ SVM.kernel_type (epachnenikov) มีประสิทธิภาพในการตรวจจับข่าวปลอมดีที่สุดสำหรับการนำเข้าสู่ข้อมูลของข่าวปลอม ให้ค่าความถูกต้องเท่ากับ 99.70 % เพิ่มมากขึ้นจากเดิมเมื่อทำการเปรียบเทียบเท่ากับ 0.03 %

จากที่กล่าวมา พบว่า งานวิจัยนี้ได้เปรียบเทียบประสิทธิภาพของห้าอัลกอริทึมในการตรวจจับข่าวปลอม ได้แก่ Naive Bayes, Decision Trees, Random Forest, Logistic Regression, และ Support Vector Machine (SVM) โดยใช้เมตริกต่างๆ เช่น ความถูกต้อง (Accuracy) และความแม่นยำ (Precision) พบว่า SVM มีค่าความถูกต้องสูงสุดในการตรวจจับข่าวปลอมที่ 96.67% ซึ่งแสดงให้เห็นว่า SVM เป็นอัลกอริทึมที่มีประสิทธิภาพสูงสุดในการตรวจจับข่าวปลอมในชุดข้อมูลที่ใช้ การทดลองใช้ฐานข้อมูลสามชุด (Kaggle, Data Word, และ UCI Machine Learning Repository) แสดงให้เห็นว่า Kaggle มีประสิทธิภาพดีที่สุดในการตรวจจับข่าวปลอม โดยมีความถูกต้องสูงสุดที่ 99.67%. การศึกษาแสดงให้เห็นว่าฐานข้อมูล Kaggle เป็นแหล่งข้อมูลที่เหมาะสมที่สุดสำหรับการทดสอบและพัฒนาระบบตรวจจับข่าวปลอม งานวิจัยได้ทำการปรับปรุง SVM โดยการคัดเลือกคุณลักษณะ ซึ่งช่วยลดจำนวนตัวแปรและเพิ่มประสิทธิภาพของโมเดล พบว่าการปรับปรุงนี้ทำให้ค่าความถูกต้องเพิ่มขึ้นจาก 99.67% เป็น 99.70%, แสดงถึงการเพิ่มขึ้นที่สำคัญเพียง 0.03% การปรับปรุงนี้ทำให้การตรวจจับข่าวปลอมมีความแม่นยำสูงขึ้นและมีประสิทธิภาพดีขึ้น

จุดเด่นของงานวิจัยนี้

1. ด้านความแม่นยำสูง

SVM ได้รับการพิสูจน์ว่ามีความแม่นยำสูงที่สุดในการตรวจจับข่าวปลอมในงานวิจัยนี้ โดยมีค่าความถูกต้องถึง 99.67% และเพิ่มขึ้นเป็น 99.70% หลังการปรับปรุงด้วยการคัดเลือกคุณลักษณะ

2. ประสิทธิภาพของฐานข้อมูล

ฐานข้อมูล Kaggle ถูกพิสูจน์ว่ามีประสิทธิภาพดีที่สุดในการตรวจจับข่าวปลอมเมื่อเปรียบเทียบกับฐานข้อมูลอื่นๆ ซึ่งแสดงถึงความสำคัญของการเลือกฐานข้อมูลที่มีคุณภาพสูงสำหรับการฝึกและทดสอบโมเดล

3. การปรับปรุงอัลกอริทึม

การใช้เทคนิคการคัดเลือกคุณลักษณะเพื่อปรับปรุง SVM แสดงถึงการพัฒนาที่สำคัญในการเพิ่มความถูกต้องของระบบ ตรวจจับข่าวปลอม ซึ่งช่วยลดมิติของข้อมูลและทำให้การวิเคราะห์มีประสิทธิภาพมากขึ้น

งานวิจัยนี้ให้ข้อมูลเชิงลึกที่สำคัญเกี่ยวกับการเลือกและปรับปรุงอัลกอริทึมสำหรับการตรวจจับข่าวปลอม และชี้ให้เห็นถึงความสำคัญของการเลือกฐานข้อมูลและเทคนิคการคัดเลือกคุณลักษณะที่เหมาะสมในการพัฒนาโมเดลที่มีประสิทธิภาพสูง

5.2 ส่วนอภิปรายผล

5.2.1 เทคนิคแมชชีนเลิร์นนิงที่ประกอบด้วยอัลกอริทึมที่ใช้งานวิจัยครั้งนี้ จากการทบทวนวรรณกรรม ผู้วิจัยเลือกใช้ 5 อัลกอริทึมที่มีผู้นิยมมาประยุกต์ใช้ในการตรวจจับข่าวปลอม เมื่อนำมาใช้ในงานวิจัยครั้งนี้ พบว่า อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนมีค่าความถูกต้องในการตรวจจับข่าวปลอมดีกว่าอัลกอริทึมอื่น และอัลกอริทึมต้นไม้แห่งการตัดสินใจมีค่าความถูกต้องในการตรวจจับข่าวปลอมในลำดับถัดมา ซึ่งสอดคล้องกับงานวิจัยของ (Divya et al., 2021) ซึ่งทำงานวิจัยเรื่อง การตรวจจับข่าวปลอมบนโซเชียลมีเดีย: มุมมองด้านการทำเหมืองข้อมูล ผลการวิจัยพบว่า การตรวจจับข่าวปลอมในโซเชียลมีเดียจากมุมมองของการทำเหมืองข้อมูล โดยให้ความสำคัญกับการวิเคราะห์และเปรียบเทียบเทคนิคการทำเหมืองข้อมูลที่ใช้ในการตรวจจับข่าวปลอม การศึกษาครอบคลุมการใช้เทคนิคต่างๆ เช่น การเรียนรู้ของเครื่อง (Machine Learning) รวมถึงการใช้ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine, SVM), ต้นไม้ตัดสินใจ (Decision Trees), และนาอีฟเบย์ (Naive Bayes) ในการพัฒนาโมเดลที่สามารถตรวจจับข่าวปลอมได้อย่างมีประสิทธิภาพ และสอดคล้องกับ (Alghamdi et al., 2022) ทำการวิจัยเรื่อง การศึกษาเปรียบเทียบเทคนิคการเรียนรู้ของเครื่องจักรและการเรียนรู้เชิงลึกสำหรับการตรวจจับข่าวปลอม พบว่า ได้ดำเนินการศึกษาเปรียบเทียบโดยใช้อัลกอริทึม ML แบบคลาสสิกที่หลากหลาย เช่น การถดถอยแบบโลจิสติกส์ (LR) เครื่องเวกเตอร์สนับสนุน (SVM) ต้นไม้การตัดสินใจ (DT) เบย์แบบไรเดียงสา (NB) ฟอเรสต์แบบสุ่ม (RF) XGBoost (XGB) และวิธีการเรียนรู้แบบรวมของ อัลกอริทึม ML เช่น เครือข่ายประสาทเทียมแบบคอนโวลูชัน (CNNs), หน่วยความจำระยะยาวระยะสั้นแบบสองทิศทาง (BiLSTM) หน่วยเรียกซ้ำแบบมีประตูแบบสองทิศทาง (BiGRU), CNN-BiLSTM, CNN-BiGRU และแนวทางไฮบริดของเทคนิคดังกล่าว และ โมเดลที่ใช้ตัวแปลง DL เช่น BERT การทดลองดำเนินการโดยใช้การฝังคำที่ผ่านการฝึกอบรมล่วงหน้าหลายวิธีในชุดข้อมูลข่าวปลอมจำนวน 4 ชุด ได้แก่ LIAR PolitiFact GossipCop และ COVID-19 เพื่อตรวจสอบประสิทธิภาพของเทคนิคต่างๆ ในชุดข้อมูล นอกจากนี้ การเปรียบเทียบระหว่างวิธีการฝังที่ไม่ขึ้นอยู่กับการบริบท (เช่น GloVe) กับประสิทธิภาพของ BERT การแสดงภาพตามบริบทในการตรวจจับข่าวปลอม เมื่อเปรียบเทียบกับผลลัพธ์ของเทคโนโลยีขั้นสูงในชุดข้อมูลที่ใช้และได้ผลลัพธ์ที่ดีขึ้นโดยอาศัยข้อความข่าวเพียงอย่างเดียว การศึกษาครั้งนี้จะมอบข้อมูลเชิงลึกที่มีประโยชน์สำหรับนักวิจัยที่ทำงานเกี่ยวกับการตรวจจับข่าวปลอม

5.2.2 ฐานข้อมูลที่ผู้วิจัยนำมาใช้ในการวิเคราะห์ในการตรวจจับข่าวปลอม จำนวน 3 ฐานข้อมูล ประกอบด้วย Kaggle, Data word และ UCI Machine Learning Repository มาใช้ในการวิเคราะห์ข้อมูล พบว่า ฐานข้อมูล Kaggle มีความถูกต้องเมื่อนำมาวิเคราะห์ดีกว่าฐานข้อมูลอื่น และฐานข้อมูล UCI Machine Learning Repository เป็นฐานข้อมูลที่นำมาวิเคราะห์แล้วมีความถูกต้องในลำดับถัดมา ซึ่งสอดคล้องกับงานวิจัยของ (Vilella et al., 2023) ทำงานวิจัยเรื่อง การตรวจจับข่าวปลอม: การทบทวนวรรณกรรมอย่างเป็นระบบเกี่ยวกับอัลกอริทึมการเรียนรู้ของเครื่อง และฐานข้อมูล พบว่า การศึกษาเกี่ยวกับการตรวจจับข่าวปลอมได้มีการนำเสนอการทบทวนอย่างเป็นระบบเกี่ยวกับอัลกอริทึมการเรียนรู้ของเครื่องและฐานข้อมูลที่ใช้ในการฝึกสอนเพื่อการตรวจจับข่าวปลอม งานวิจัยนี้มุ่งเน้นการวิเคราะห์อัลกอริทึม เช่น Stacking Method, Bidirectional Recurrent Neural Network (BiRNN), และ Convolutional Neural Network (CNN) ซึ่งมีความถูกต้องในการตรวจจับข่าวปลอมสูงถึง 99.9% และ 99.8% สำหรับ BiRNN และ CNN ตามลำดับ อย่างไรก็ตาม การวิจัยส่วนใหญ่ยังใช้ฐานข้อมูลในสภาพแวดล้อมที่ควบคุมได้ เช่น Kaggle หรือฐานข้อมูลที่ไม่ได้รับข้อมูลที่อัปเดตตามเวลาจริงจากโซเชียลมีเดีย ซึ่งมีการแพร่กระจายข่าวปลอมอย่างกว้างขวางในอนาคตควรมีการศึกษาเพิ่มเติมในด้านข่าวอื่นๆ นอกเหนือจากการเมืองและการใช้วิธีการแบบผสมเพื่อปรับปรุงความสามารถในการตรวจจับข่าวปลอม และสอดคล้องกับ D'Ulizia et al., 2021 ได้ทำวิจัยเรื่อง การตรวจจับข่าวปลอม: การสำรวจฐานข้อมูลการประเมินผล พบว่า การทบทวนฐานข้อมูลที่เป็นที่นิยม 27 ชุดสำหรับการตรวจจับข่าวปลอม โดยนำเสนอข้อมูลเชิงลึกเกี่ยวกับลักษณะของแต่ละฐานข้อมูลและการวิเคราะห์เปรียบเทียบระหว่างกัน เราได้จัดทำลักษณะการจำแนกประเภทของฐานข้อมูลตรวจจับข่าวปลอม ซึ่งประกอบด้วยลักษณะ 11 ประการที่สกัดมาจากฐานข้อมูลที่สำรวจพร้อมกับชุดข้อกำหนดสำหรับการเปรียบเทียบและการสร้างฐานข้อมูลใหม่ ผลการวิเคราะห์นี้มีความสำคัญสำหรับนักวิจัยจำนวนมากในการเลือกหรือกำหนดฐานข้อมูลที่เหมาะสมสำหรับการประเมินวิธีการตรวจจับข่าวปลอม

5.2.3 การปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนด้วยวิธีการคัดเลือกคุณลักษณะ หลังจากที่ทำการศึกษาปรับปรุงอัลกอริทึม พบว่า อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน มีค่าเพิ่มมากขึ้นทุก parameter โดยการปรับปรุงอัลกอริทึมของ SVM.kernel_type (epachnenikov) มีประสิทธิภาพในการตรวจจับข่าวปลอมดีที่สุด และ SVM.kernel_type (radial) มีค่าถูกต้องในการตรวจจับข่าวปลอมในลำดับถัดมา ซึ่งสอดคล้องกับงานวิจัยของ Chidambaram & Srinivasagan, 2019 พบว่า การคัดเลือกคุณลักษณะและการจำแนกประเภทด้วยซัพพอร์ตเวกเตอร์แมชชีน (SVM) เน้นที่การเพิ่มประสิทธิภาพการจำแนกประเภทข้อมูล โดยการนำเสนอวิธีการที่ประกอบด้วยสามขั้นตอนหลัก การใช้ SVM กับเคอร์เนลต่างๆ การทดลองใช้ SVM กับฟังก์ชันเคอร์เนลสี่แบบ ได้แก่ ฟังก์ชันเชิงเส้น, ฟังก์ชันพหุนาม, ฟังก์ชันฐานเชิงรัศมี (RBF) และฟังก์ชันซิกมอยด์ เพื่อเปรียบเทียบประสิทธิภาพใน

การจำแนกประเภทข้อมูล การใช้การประเมินชุดตัวจำแนกประเภทเพื่อคัดเลือกคุณลักษณะที่ดีที่สุด และเพิ่มประสิทธิภาพของเวกเตอร์คุณลักษณะเพื่อให้ได้ความแม่นยำสูงสุด และการพัฒนาวิธีการ เฟอร์เนลใหม่ที่มีความแม่นยำสูงสุดในการจำแนกประเภทเมื่อเปรียบเทียบกับเฟอร์เนลอื่นๆ และ สอดคล้องกับ Zanasi, 2000 ได้ทำการวิจัยเรื่อง การคัดเลือกคุณลักษณะโดยใช้ซัพพอร์ตเวกเตอร์แม ชชีน พบว่า เสนอวิธีการคัดเลือกคุณลักษณะที่อิงจาก SVM เชิงเส้น โดยการฝึก SVM เชิงเส้นกับชุด ข้อมูลฝึกอบรมย่อยและคงไว้เพียงคุณลักษณะที่สัมพันธ์กับส่วนประกอบที่มีน้ำหนักสูงของระนาบที่ แยกตัวอย่างบวกและลบ วิธีนี้ช่วยลดขนาดของพื้นที่คุณลักษณะและทำให้สามารถฝึกตัวจำแนก ประเภทได้บนชุดข้อมูลฝึกอบรมที่ใหญ่ขึ้น ผลการทดลองแสดงให้เห็นว่าการคัดเลือกคุณลักษณะที่อิง จาก SVM มีประสิทธิภาพดีกว่าวิธีการคัดเลือกแบบดั้งเดิม เช่น อัตราส่วนโอกาส (odds ratio) และ การเพิ่มข้อมูล (information gain) ในการบรรลุความสมดุลระหว่างความกระจัดกระจายของ เวกเตอร์และประสิทธิภาพการจำแนกประเภท

5.3 ส่วนข้อจำกัดทางการวิจัยและข้อเสนอแนะในการทำวิจัยครั้งต่อไป

5.3.1 แมชชีนเลิร์นนิงยังมีข้อจำกัดหนึ่งคือ กรณีที่มีข้อมูลที่ไม่รู้จักจำนวนมากแบบจำลอง จะไม่สามารถจำแนกข้อมูลได้อย่างถูกต้อง หากไม่มีข้อมูลการเทรนนิ่งในจำนวนมากเพียงพอ ดังนั้น จึงจำเป็นต้องมีการรวบรวมข้อมูลจำนวนมากเพื่อรองรับข่าวที่มีความหลากหลายมากขึ้น

5.3.2 ในอนาคตอาจจะมีข้อมูลเพิ่มเติมที่ใช้ในการสร้างแบบจำลองที่ดียิ่งขึ้น เพื่อปรับปรุง ประสิทธิภาพการจำแนกข่าวปลอมให้สามารถรองรับประเด็นข่าวที่มากขึ้น และทันสมัยกับข่าว ปัจจุบัน

5.3 ประโยชน์ที่ได้รับ

1. เชิงวิชาการ

1.1 การพัฒนาทฤษฎีและโมเดลทางวิชาการ: การวิจัยเกี่ยวกับข่าวปลอมช่วยให้เกิดการ พัฒนาและปรับปรุงทฤษฎีด้านภาษาศาสตร์คอมพิวเตอร์ การเรียนรู้ของเครื่อง (Machine Learning) และการประมวลผลภาษาธรรมชาติ (Natural Language Processing) โดยการสร้างและปรับปรุง อัลกอริธึมที่สามารถแยกแยะข่าวปลอมจากข่าวจริงได้อย่างมีประสิทธิภาพมากขึ้น

1.2 การสร้างชุดข้อมูลและมาตรฐาน การวิจัยข่าวปลอมทำให้เกิดการสร้างชุดข้อมูล ขนาดใหญ่และมาตรฐานในการตรวจจับข่าวปลอม ซึ่งเป็นพื้นฐานสำคัญสำหรับการวิจัยและการ พัฒนาเทคโนโลยีในอนาคต ชุดข้อมูลเหล่านี้สามารถใช้ในการฝึกอบรมและทดสอบโมเดลใหม่ ๆ ได้

1.3 การเพิ่มประสิทธิภาพในการสื่อสารมวลชนและการศึกษา การปรับปรุงการตรวจจับ ข่าวปลอมช่วยให้สื่อมวลชนสามารถทำงานได้อย่างมีประสิทธิภาพมากขึ้นในการตรวจสอบและ

นำเสนอข่าวสารที่ถูกต้อง นอกจากนี้ยังสามารถนำไปใช้ในระบบการศึกษาเพื่อสอนนักเรียนและนักศึกษาถึงวิธีการวิเคราะห์และตรวจสอบข้อมูลอย่างถูกต้อง

1.4 การวิจัยด้านจริยธรรมและผลกระทบทางสังคม การศึกษาผลกระทบของข่าวปลอมต่อสังคมและการวิจัยเกี่ยวกับจริยธรรมของการใช้เทคโนโลยีตรวจจับข่าวปลอม ช่วยให้เกิดความเข้าใจในมิติที่ลึกซึ้งเกี่ยวกับบทบาทและความรับผิดชอบของเทคโนโลยีในสังคม

1.5 การพัฒนาความสามารถในการวิเคราะห์ข้อมูลขนาดใหญ่ การตรวจจับข่าวปลอมต้องการการวิเคราะห์ข้อมูลจำนวนมาก การวิจัยในด้านนี้ช่วยพัฒนาทักษะและเทคนิคในการจัดการและวิเคราะห์ข้อมูลขนาดใหญ่ ซึ่งเป็นทักษะที่มีประโยชน์ในหลายๆ ด้านของวิชาการและอุตสาหกรรม

1.6 การเสริมสร้างความรู้และความเข้าใจในเรื่องการสื่อสารดิจิทัล การวิจัยข่าวปลอมช่วยให้นักวิชาการเข้าใจพฤติกรรมและการสื่อสารในยุคดิจิทัล และสามารถนำความรู้ไปปรับใช้ในการพัฒนาเครื่องมือและกลยุทธ์เพื่อป้องกันการแพร่กระจายของข้อมูลที่เป็นเท็จ

2. เชิงสาธารณะ

2.1 การป้องกันการแพร่กระจายของข้อมูลที่ผิดพลาด การตรวจจับข่าวปลอมอย่างมีประสิทธิภาพช่วยลดการแพร่กระจายของข้อมูลที่เป็นเท็จ ทำให้ประชาชนได้รับข้อมูลที่ถูกต้องและเชื่อถือได้มากขึ้น ลดการเข้าใจผิดและการกระทำที่ไม่เหมาะสมอันเกิดจากข้อมูลที่ผิดพลาด

2.2 เสริมสร้างความเชื่อมั่นในสื่อมวลชน เมื่อประชาชนสามารถแยกแยะข่าวปลอมจากข่าวจริงได้ จะส่งผลให้ความเชื่อมั่นในสื่อมวลชนและแหล่งข่าวต่าง ๆ เพิ่มขึ้น ลดความสงสัยและความไม่ไว้วางใจในข้อมูลที่ได้รับ

2.3 การป้องกันผลกระทบทางสังคมและการเมือง ข่าวปลอมสามารถสร้างความขัดแย้งและความแตกแยกในสังคม การตรวจจับข่าวปลอมช่วยลดความเสี่ยงของการเกิดความรุนแรงและการขัดแย้งทางการเมือง โดยการสร้างความเข้าใจที่ถูกต้องในประเด็นที่เป็นที่ถกเถียง

3. เชิงพาณิชย์

3.1 การป้องกันการโจมตีทางไซเบอร์และการฉ้อโกง ข่าวปลอมมักถูกใช้เป็นเครื่องมือในการโจมตีทางไซเบอร์และการฉ้อโกง การปรับปรุงการตรวจจับข่าวปลอมช่วยป้องกันการโจมตีที่อาจเกิดขึ้นกับบริษัท ลดความเสี่ยงและค่าใช้จ่ายที่เกี่ยวข้องกับการแก้ไขปัญหาที่เกิดจากการโจมตีทางไซเบอร์

3.2 การเสริมสร้างความปลอดภัยในการทำธุรกรรมออนไลน์: เมื่อผู้บริโภคได้รับข้อมูลที่ถูกต้องและน่าเชื่อถือ จะช่วยสร้างความมั่นใจในการทำธุรกรรมออนไลน์ ส่งผลให้ยอดขายและการทำธุรกรรมของบริษัทเพิ่มขึ้น

3.3 การเพิ่มประสิทธิภาพในการตลาดและการโฆษณา การตรวจจับข่าวปลอมช่วยให้บริษัทสามารถคัดกรองและตรวจสอบข้อมูลก่อนการเผยแพร่ ทำให้การสื่อสารการตลาดและการโฆษณามีความถูกต้องและน่าเชื่อถือมากขึ้น ลดความเสี่ยงที่ข้อมูลที่ไม่ถูกต้องจะส่งผลกระทบต่อภาพลักษณ์ของบริษัท

3.4 การสร้างความเชื่อมั่นในตลาดหลักทรัพย์ สำหรับบริษัทที่จดทะเบียนในตลาดหลักทรัพย์ การมีระบบตรวจจับข่าวปลอมช่วยลดความเสี่ยงในการเกิดข้อมูลที่อาจทำให้เกิดความผันผวนในราคาหุ้น ทำให้นักลงทุนมีความมั่นใจในการลงทุนมากขึ้น



บรรณานุกรม

- กายกาญจน์ เสนแก้ว. (2560). พฤติกรรมการใช้เครือข่ายสังคมออนไลน์ของ Gen-X ใน กรุงเทพมหานคร. บัณฑิตวิทยาลัย มหาวิทยาลัยกรุงเทพฯ.
- กรุงเทพธุรกิจ (2022). ข้อมูลล่าสุดสถิติใช้ดิจิทัลในไทย จากรายงาน We are social. [ออนไลน์].
ได้จาก <https://www.bangkokbiznews.com/columnist/989552> [สืบค้นเมื่อ วันที่ 15 สิงหาคม 2566].
- ไกรรุ่ง เสงพระพรหม, สุพจน์ เสงพระพรหมและสุวิมล มรรควิบูลย์ชัย. (2557). การค้นหาความรู้
คุณลักษณะสำคัญของนักศึกษาที่มีต่อผลสัมฤทธิ์ทางการเรียนในกลุ่ม โปรแกรมคอมพิวเตอร์
มหาวิทยาลัยราชภัฏนครปฐมด้วยเทคนิคเหมืองข้อมูล. วารสารวิชาการราชภัฏตะวันตก.
9(1), 71 – 80.
- ครองรัตน์ ดุลลาพันธ์. (2561). พฤติกรรมการใช้เครือข่ายสังคมออนไลน์และปัจจัยทางจิตที่พยากรณ์
ภาวะโรคซึมเศร้า. ปริญญานิพนธ์ปริญญาศิลปศาสตรมหาบัณฑิต สาขาวิชานิติศาสตร์และ
นวัตกรรม คณะนิติศาสตร์และนวัตกรรมการจัดการ สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- จุฑารัตน์ ศรวณะวงศ์. (2558). แนวทางการใช้สื่อสังคมออนไลน์กับงานประชาสัมพันธ์
มหาวิทยาลัยเกษตรศาสตร์. วารสารวิจัยสมาคมห้องสมุดแห่งประเทศไทยฯ. 8(2), 55-69.
- เฉลิมชัย กีกเกียรติกุลและธัญญนันทนัฐ ต่านไพบูลย์. (2561). ขวาลง : ปัญหาและความท้าทาย.
วารสารวิชาการ กสทช. ประจำปี 2561. 2(2), 173–192.
- ธนาภัทร ภัทรวินิจ. (2563). การทำนายความผิดพลาดระยะต้นของเครื่องวิเคราะห์อินทรีย์คาร์บอน
โดยการเรียนรู้เชิงลึก. ปริญญานิพนธ์ปริญญาการศึกษามหาบัณฑิต สาขาวิชาวิทยาศาสตร์
คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- นครินทร์ ชานะมัย. (2563). กลยุทธ์การสื่อสารการตลาด Online to Offline to On Ground ของ
โทรศัพท์ดิจิทัลในประเทศไทย. วารสารนิติศาสตร์ มสธ. 10(1). 1 – 12.
- นันทิกา หนูสม และวิโรจน์ สุทธิสีมา. (2562). ลักษณะของชาวพลอมในประเทศไทยและระดับ
ความรู้เท่าทันชาวพลอมบนเฟซบุ๊กของผู้รับสารในเขตกรุงเทพมหานคร. วารสารนิติศาสตร์. 37 (1). 37 – 45.
- พยุ่ง มีสัจ, พนม คล้ายยา, งามอาจ อุ๋นนันต์ และกมลรัตน์ กิจรุ่งไพศาล (2565) เทคนิคทาง
ปัญญาประดิษฐ์สำหรับการตรวจชาวพลอมภาษาไทย. วารสารวิทยาศาสตร์ประยุกต์. 21 (1).
1 – 19.

- พระมหาสาทรและคณะ. (2564). พระพุทธศาสนากับการแก้ปัญหาข่าวปลอมในสังคมไทย.วารสาร มจร พุทธปัญญาปริทรรศน์. 6(2). 182 – 195.
- เพชรรัตน์ ม่วงน้อย, จักรพันธ์ พลาผล และภรณ์ยา ปาลวิสุทธิ. (2563). ตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคเหมืองข้อมูล. วารสารวิชาการการประยุกต์ใช้เทคโนโลยีสารสนเทศ. 7 (1). 54 – 63.
- ภัทรารุณ แสงศิริ. (2553). การคัดแยกประเภทของมะเร็งเม็ดเลือดขาวโดยใช้วิธีการจัดอันดับร่วมกับเทคนิคซ์พอร์ตเวกเตอร์แมชชีน. วารสารวิจัย มช. (บศ.) 10(2). เม.ย.-มิ.ย.2553, 10-7
- มานิจ สุขสมจิตร (2565) บทความพิเศษ Fake news : ข่าวลวง ข่าวปลอม. วารสารสตรี. 15(1). 9 – 39.
- รัตนภรณ์ จอมมูล. (2020). มาตรการทางกฎหมายในการกำกับควบคุมข่าวปลอมทางการเมืองบนโลกไซเบอร์. วารสารนิติสังคมศาสตร์. 13 (1). 1 – 23.
- วรภาพร คำจับ. (2562). สื่อสังคมออนไลน์กับการเรียนรู้ในศตวรรษที่ 21. วารสารศิลปศาสตร์ มหาวิทยาลัยแม่โจ้. 7(2). 143 – 159.
- วรเชษฐ์ โทอิน. (2562). ปัจจัยเชิงสาเหตุและแนวทางป้องกันพฤติกรรมอันไม่เหมาะสมต่อการใช้โทรศัพท์มือถือของนักศึกษาวิทยาลัยศาสนศาสตร์ยโสธรเพื่อการครองตนตามหลักพระพุทธศาสนา. วารสารมหาจุฬานาครทรรศ. 6(10). 4769 – 4785.
- วิวัฒน์ กิตตินารพรและคณะ. (2566). การประยุกต์ใช้เทคนิคซ์พอร์ตเวกเตอร์แมชชีน และโครงข่ายประสาทเทียมพยากรณ์ปริมาณงานก่อสร้างในประเทศไทย. วิศวกรรมสารฉบับวิจัยและพัฒนา. 34(1) 95 – 108.
- สกุลศรี ศรีสารคามและคณะ. (2559). การส่งเสริมแนวทางการใช้ข้อมูลจากสื่อออนไลน์ในกระบวนการรายงานข่าว ในยุคเทคโนโลยีหลอมรวม. สถาบันการจัดการปัญญาภิวัฒน์.
- สถาบันอิสรา. (2565). ผลกระทบของข่าวปลอมและแนวทางการรับมือของผู้บริโภคสื่อ. สืบค้นจาก https://www.isra.or.th/news-and-trainee/item/766-fakenews_766.html.
- สรานนท์ อินทนนท์ (2563). รู้เท่าทันข่าว : News Literacy. ปทุมธานี : วอลค์ ออน คลาวด์.
- สุรวุฑ ศรีเปารยะและสายชล สีนสมบุรณ์ทอง (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง กรณีศึกษา โรงพยาบาลแห่งหนึ่งในประเทศอินเดีย.วารสารวิทยาศาสตร์และเทคโนโลยี. 25(5). กันยายน - ตุลาคม.

สุชาลักษณ์ ธรรมดวงดี (2562) พฤติกรรมการใช้เครือข่ายสังคมออนไลน์ของนักศึกษา.

มหาวิทยาลัยราชภัฏมหาสารคาม.

สุนันทา แยมทัพและอริยญา ศิริผล (2564) ขวาล้อมกับการทำงานของนักข่าวในยุคดิจิทัล.

วารสารศาสตร์. พฤษภาคม - สิงหาคม.

สุปัญญา อภิวงษ์โสภณ. (2561) การตรวจสอบข่าวปลอมด้วยวิธีการเรียนรู้ด้วยเครื่อง. จุฬาลงกรณ์มหาวิทยาลัย

สุวัฒน์ เอสมัทธน์ (2561) การแสดงความเห็นด้วยอคติผ่านข่าวปลอมผ่านสื่อสังคมออนไลน์ : กรณีศึกษาการเลือกตั้งประเทศไทยในปี พ.ศ.2562. มหาวิทยาลัยธรรมศาสตร์

สำนักข่าวเอพี. (2562) Fake News Session by AP. กรุงเทพฯ

สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์. (2565). ข่าวปลอม ข่าวลวงให้เชื่อ หลอกให้แชร์.

สืบค้นจาก <https://www.etda.or.th/th/Useful-Resource/knowledge->

อัจฉิมา มณฑาพันธุ์ (2560) การวิเคราะห์เปรียบเทียบการระบุตัวตน โดยการใช้ลักษณะเฉพาะทางกายภาพของร่างกาย. มหาวิทยาลัยศรีปทุม.

เอกสิทธิ์ พัทรวงศ์ศักดิ์ดา. (2557). การวิเคราะห์ข้อมูล ด้วยเทคนิคดาต้าไมน์นิ่งเบื้องต้น. กรุงเทพฯ: เอเชีย ดิจิตอลการพิมพ์.

Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, 141, 215–222.

Alghamdi, J., Lin, Y., & Luo, S. (2022). A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection. *Information (Switzerland)*, 13(12).

Allein, L., Moens, M. F., & Perrotta, D. (2023). Preventing profiling for ethical fake news detection. *Information Processing & Management*, 60(2), 103-206.

Alrubaian, M., Al-Qurishi, M., Alamri, A., Al-Rakhami, M., Hassan, M. M., & Fortino, G. (2019). Credibility in Online Social Networks: A Survey. *IEEE Access*, 7, 2828–2855.

Altheneyan, A., & Alhadlaq, A. (2023). Big Data ML-Based Fake News Detection using Distributed Learning. *IEEE Access*.

Ahmad, I., Yousaf, M., Yousaf, S., & Ovais Ahmad, M. (2020). Fake News Detection Using Machine Learning Ensemble Methods.

- Aphiwongsophon, S., & Chongstitvatana, P. (2018). Detecting Fake News with Machine Learning Method.
- Atodiresei, C. S., Tănăsescu, A., & Iftene, A. (2018). Identifying Fake News and Fake Users on Twitter. *Procedia Computer Science*, 126, 451–461.
- Baarir, N. F., & Djeflal, A. (2021). Fake News detection Using Machine Learning. 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-Being, IHSH 2020, 125–130.
- Bazmi, P., Asadpour, M., & Shakery, A. (2023). Multi-view co-attention network for fake news detection by modeling topic-specific user and news source credibility. *Information Processing & Management*, 60(1), 103146.
- Billones, P. J. R., MacAsaet, D. D., & Arenas, S. U. (2022). Bilingual Fake News Detection Algorithm Using Naïve Bayes and Support Vector Machine Models. *Proceedings - 2022 IET International Conference on Engineering Technologies and Applications, IET-ICETA 2022*.
- Buntain, C., & Golbeck, J. (2017). Automatically Identifying Fake News in Popular Twitter Threads. *Proceedings - 2nd IEEE International Conference on Smart Cloud, SmartCloud 2017*, 208–215.
- Bodaghi, A., & Oliveira, J. (2022). The theater of fake news spreading, who plays which role? A study on real graphs of spreading on Twitter. *Expert Systems with Applications*, 189, 116110.
- Capuano, N., Fenza, G., Loia, V., & Nota, F. D. (2023). Content-Based Fake News Detection With Machine and Deep Learning: a Systematic Review. *Neurocomputing*, 530, 91–103.
- Cantarella, M., Fraccaroli, N., & Volpe, R. (2023). Does fake news affect voting behaviour?. *Research Policy*, 52(1).
- Castillo, C., Mendoza, M., & Poblete, B. (2013). Predicting information credibility in time-sensitive social media. *Internet Research*, 23(5), 560–588.
- Carlos Pereira-Kohatsu, J., Quijano-Sánchez, L., Liberatore, F., & Camacho-Collados, M. (n.d.). Detecting and Monitoring Hate Speech in Twitter.

- Chidambaram, S., & Srinivasagan, K. G. (2019). Performance evaluation of support vector machine classification approaches in data mining. *Cluster Computing*, 22, 189–196.
- Daily Gizmo. (2022). 6 things that make Twitter a platform that wins the hearts of many people. From <https://www.dailygizmo.tv/2020/02/24/5-use-twitter/>
- Digi musketeers (2021) Include social media that should be used in your business. Retrieved 5 January 2023, from <https://digimusketeers.co.th/blogs/online-marketing/what-is-social-media>
- Divya, T. v, Nagar, A., Barnali Gupta Banik, T., & Professor, A. (2021). An Empirical Study on Fake News Detection System using Deep and Machine Learning Ensemble Techniques. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 12, Issue 12).
- D’Ulizia, A., Caschera, M. C., Ferri, F., & Grifoni, P. (2021). Fake news detection: A survey of evaluation datasets. *PeerJ Computer Science*, 7, 1–34.
- Easa, H. K., Saber, A. A., Hamid, N. K., & Saber, H. A. (2023). Machine learning based approach for detection of fake banknotes using support vector machine. *Indonesian Journal of Electrical Engineering and Computer Science*, 31(2), 1016–1022.
- Edison Research. (2023). *The-Infinite-Dial-2023*. Retrieved 30 April 2023, from <https://www.edisonresearch.com/wp-content/uploads/2023/03/The-Infinite-Dial-2023.pdf>
- Ehsanfar, A., & Mansouri, M. (2017). Incentivizing the dissemination of truth versus fake news in social networks. 2017 12th System of Systems Engineering Conference, SoSE 2017.
- European Commission. (2017). *IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON) : conference proceedings : May 29-June 2, 2017, Kyiv, Ukraine*.
- Hiramath, C. K., & Deshpande, G. C. (2019). Fake news detection using deep learning techniques. In *2019 1st International Conference on Advances in Information Technology (ICAIT)*. 411-415).

- Hu, B., Mao, Z., & Zhang, Y. (2024). An Overview of Fake News Detection: From A New Perspective. Fundamental Research.
- H. Saleh, A. Alharbi and S. H. Alsamhi. (2021). OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection," in IEEE Access, vol. 9, pp. 129471-129489.
- Jain, P., Sharma, S., Monica, & Aggarwal, P. K. (2022). Classifying Fake News Detection Using SVM, Naive Bayes and LSTM. Proceedings of the Confluence 2022 - 12th International Conference on Cloud Computing, Data Science and Engineering, 460–464.
- Jing, J., Wu, H., Sun, J., Fang, X., & Zhang, H. (2023). Multimodal fake news detection via progressive fusion networks. Information Processing & Management, 60(1), 103120.
- Joju, S., & Kammath, Mr. P. S. (2021). Analysis on Fake news Detection Using Machine Learning. <https://www.semanticscholar.org/paper>.
- Jose, X., Madhu Kumar, S. D., & Chandran, P. (2021). Characterization, Classification and Detection of Fake News in Online Social Media Networks. 2021 IEEE Mysore Sub Section International Conference, MysuruCon 2021, 759–765.
- Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and Research: A Systematic Literature Review through Text Mining. IEEE Access, 8, 67698–67717.
- Keya, A. J., Afridi, S., Maria, A. S., Pinki, S. S., Ghosh, J., & Mridha, M. F. (2021). Fake News Detection Based on Deep Learning. 2021 International Conference on Science and Contemporary Technologies, ICST 2021.
- Khalil, A., Jarrah, M., Aldwairi, M., & Jararweh, Y. (2021). Detecting Arabic Fake News Using Machine Learning. 2021 2nd International Conference on Intelligent Data Science Technologies and Applications, IDSTA 2021, 171–177.
- Khanam, Z., Alwasel, B. N., Sirafi, H., & Rashid, M. (2021). Fake News Detection Using Machine Learning Approaches. IOP Conference Series: Materials Science and Engineering, 1099(1), 12-40.

- Kiahwar, A., & Zafar, A (2022). Fake News Detection on Pakistani News Using Machine Learning and Deep Learning. SSRN Electronic Journal.
- Krishna, N. L. S. R., & Adimoolam, M. (2022). Fake News Detection system using Decision Tree algorithm and compare textual property with Support Vector Machine algorithm. 2022 International Conference on Business Analytics for Technology and Security, ICBATS 2022.
- Lazer, D.M.J., Buam, M.A., Bankler, Y., Berinsky, A. J., Greenhill, K.M., Menczer, F., Zittrian, J, L. (2018) The Science of Fake News. Science Magazine, 359 (6380), 1094-1096.
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news: Addressing fake news requires a multidisciplinary effort. Science, 359(6380), 1094–1096.
- Li, J., & Lei, M. (2022). A Brief Survey for Fake News Detection via Deep Learning Models. Procedia Computer Science, 214, 1339–1344. <https://doi.org/10.1016/J.PROCS.2022.11.314>
- Luo, Z., Li, Q., & Zheng, J. (2021). Deep Feature Fusion for Rumor Detection on Twitter. IEEE Access, 9, 126065–126074. <https://doi.org/10.1109/ACCESS.2021.3111790>
- Mark A. Hall. 1999. “Correlation-based Feature Selection for Machine Learning.” Doctor of Philosophy. Department of Computer Science, The University of Waikato, Hamilton, New Zealand
- Meesad, P. (2021). Thai Fake News Detection Based on Information Retrieval, Natural Language Processing and Machine Learning. SN Computer Science, 2(6), 1–17.
- Meesad, P., Kleechaya, P., Aun-a-nan, A., & Kijrungpaisarn, K. (2022). Artificial Intelligent Techniques for Thai Fake News Detection. The Journal of Applied Science, 21(1), 244590–244590.
- M. Park and S. Chai, (2023). Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques," in IEEE Access, vol. 11, pp. 71517-71527

- Mishra, S., Shukla, P., & Agarwal, R. (2022). Review Article Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets.
- Mladenova, T., & Valova, I. (2022). Research on the Ability to Detect Fake News in Users of Social Networks. HORA 2022 - 4th International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings.
- Molek. (2024). Marketingoops : When fake news spreads and hurts brands, how should brands cope.From <https://www.marketingoops.com/exclusive/how-to/fake-news-hurt-brand-how-brand-prepared/>
- Monti, F., Frasca, F., Eynard, D., Mannion, D., Bronstein, M. M., Ai, F., & Lugano, U. (2019). Fake News Detection on Social Media using Geometric Deep Learning.
- N. L. S. R. Krishna & M. Adimoolam. (2022). Fake News Detection system using Decision Tree algorithm and compare textual property with Support Vector Machine algorithm," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 1-6.
- N. L. Siva Rama Krishna and M. Adimoolam. (2022). Fake News Detection System using Logistic Regression and Compare Textual Property with Support Vector Machine Algorithm," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 48-53.
- Nagaraja, A., N. S. K., Naik, P., Sinha, A., & Vinay Rajendrakumar, J. (2021). Fake News Detection Using Machine Learning Methods; Fake News Detection Using Machine Learning Methods.
- Narumon Pinto (2020) A Space of Lying: Between a Small Hole and a Deep Slope. *Journal of Journalism*. 13 (3). September-December Issue: "Questioning Media and Civil Society"
- Nomesh, R., & Saravanan, M. S. (2022). Classification of Fake News on Facebook a Novel Social Network with K-Means Clustering Approach for Against Principal Component Analysis Method for Better Accuracy. 3rd International Conference on Smart Electronics and Communication, ICOSEC 2022 - Proceedings, 722–726.

- Noria (2022). What is Social Media. How many types are there and why do businesses need one. From <https://noria.co.th/th/what-is-social-media-types/>
- Nugroho, R., Nugroho, R., Yang, J., Zhao, W., Paris, C., & Nepal, S. (2020). What and with Whom? Identifying Topics in Twitter through Both Interactions and Text. *IEEE Transactions on Services Computing*, 13(3), 584–596.
- P. Jain, S. Sharma, Monica and P. K. Aggarwal. (2022). Classifying Fake News Detection Using SVM, Naive Bayes and LSTM, 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 460-464,
- Pal, A., Pranav, & Pradhan, M. (2023). Survey of fake news detection using machine intelligence approach. *Data and Knowledge Engineering*, 144.
- Pizarro, J. (2020). Profiling Bots and Fake News Spreaders at PAN'19 and PAN'20: Bots and Gender Profiling 2019, Profiling Fake News Spreaders on Twitter 2020. *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, 626–630.
- Praseed, A., Rodrigues, J., & Thilagam, P. S. (2023). Hindi fake news detection using transformer ensembles. *Engineering Applications of Artificial Intelligence*, 119, 105731.
- Rainmaker. (2023) Open statistics for 6 platforms ready to let everyone know for masters and brands. From <https://www.rainmaker.in.th/social-media-stat-and-trend>.
- Rahm, Erhard and Do, Hong-Hai. (2012) Data Cleaning : Problems and Current Approaches. http://www.witi.cs.uni-magdeburg.de/iti_db/lehre/dw/paper/data_cleaning.pdf. Accessed. 7-12-2012.
- Rohera, D., Shethna, H., Patel, K., Thakker, U., Tanwar, S., Gupta, R., Hong, W. C., & Sharma, R. (2022). A Taxonomy of Fake News Classification Techniques: Survey and Implementation Aspects. *IEEE Access*, 10, 30367–30394.
- Sansonetti, G., Gasparetti, F., & Micarelli, A. (2021). Using Social Media for Personalizing the Cultural Heritage Experience. *UMAP 2021 - Adjunct Publication of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 189–193.

ScienceDirect. (2024). Fake news. Retrieved 30 April 2023, from <https://www.sciencedirect.com/search?q=fake%20new>

Shanthi, C. D. (2022). Comparative Analysis of Fake News Detection Using Different Methodologies. *International Journal of Science and Research*.

Sreedhar, K. S., Ahmed, S. T., & Sreejesh, G. (2022). An Improved Technique to Identify Fake News on Social Media Network using Supervised Machine Learning Concepts. *Proceedings - 2022 IEEE World Conference on Applied Intelligence and Computing, AIC 2022*, 652–658.

Srinivasa Rao, D., Rajasekhar, N., Sowmya, D., Archana, D., Hareesha, T., & Sravya, S. (2021). Fake News Detection Using Machine Learning Technique. *SCRS Conference Proceeding on Intelligent Systems*, 59–69.

Swapna, H., & Soniya, B. (2022). A Review on News-Content Based Fake News Detection Approaches. *Proceedings of International Conference on Computing, Communication, Security and Intelligent Systems, IC3SIS 2022*. <https://doi.org/10.1109/IC3SIS54991.2022.9885447>

Tiemtud, K., & Leesutthipornchai, P. (2023). Thai Fake News Detection Using Machine Learning Model. *Proceeding - 2023 International Electrical Engineering Congress, IEECON 2023*, 346–350.

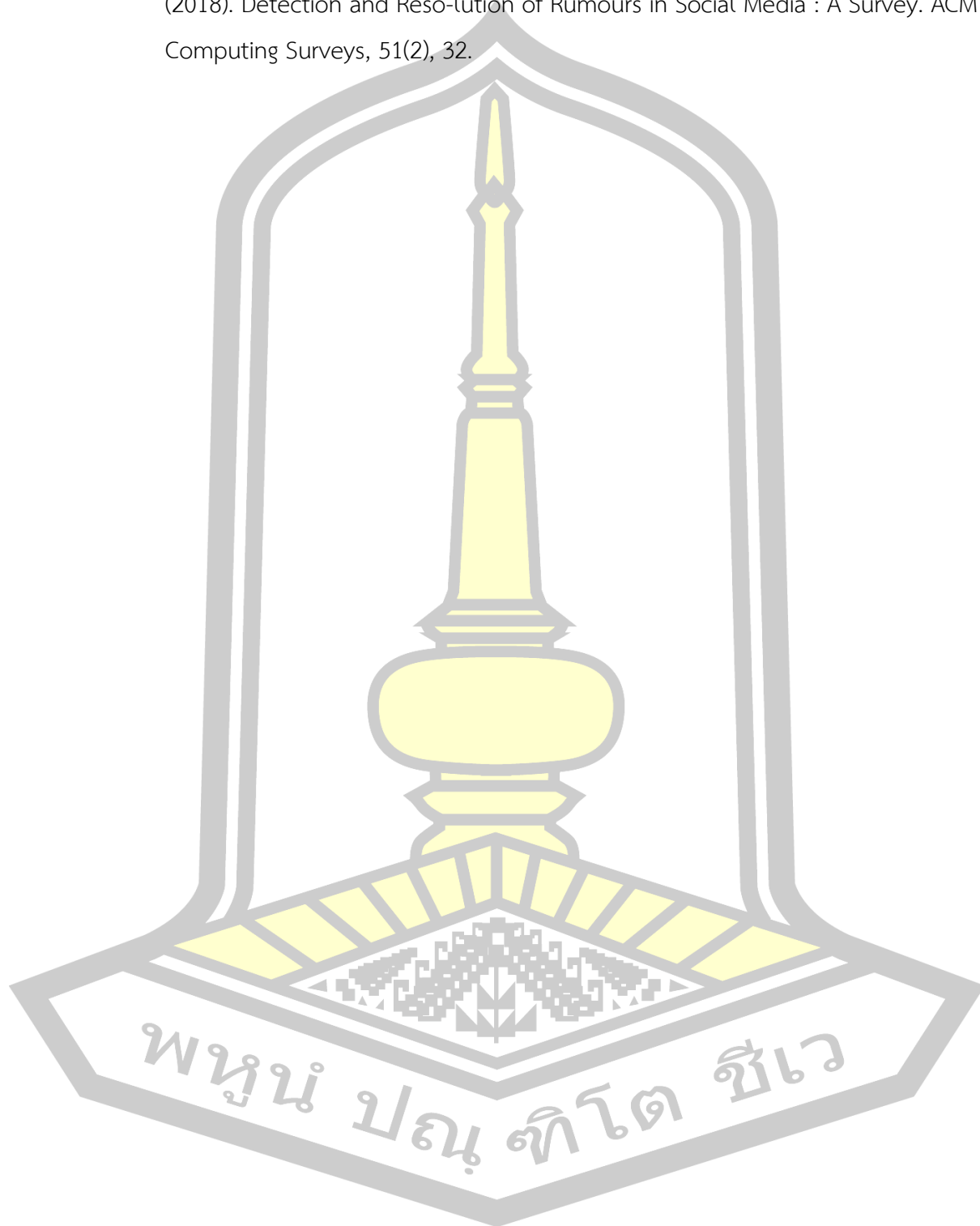
The growth master (2023) Digital 2023 Global Overview report (The essential Guide to the world's connected behaviors. From <https://thegrowthmaster.com/trends/digital-trend-statistic-2023>

Vasist, P. N., & Sebastian, M. P. (2022). Tackling the infodemic during a pandemic: A comparative study on algorithms to deal with thematically heterogeneous fake news. *International Journal of Information Management Data Insights*, 2(2).

Victor, D. B., & Raza, D. M. (2022). Depression Detection from Facebook using Machine Learning Techniques. *2022 6th International Conference on Trends in Electronics and Informatics, ICOEI 2022 - Proceedings*, 1447–1454.

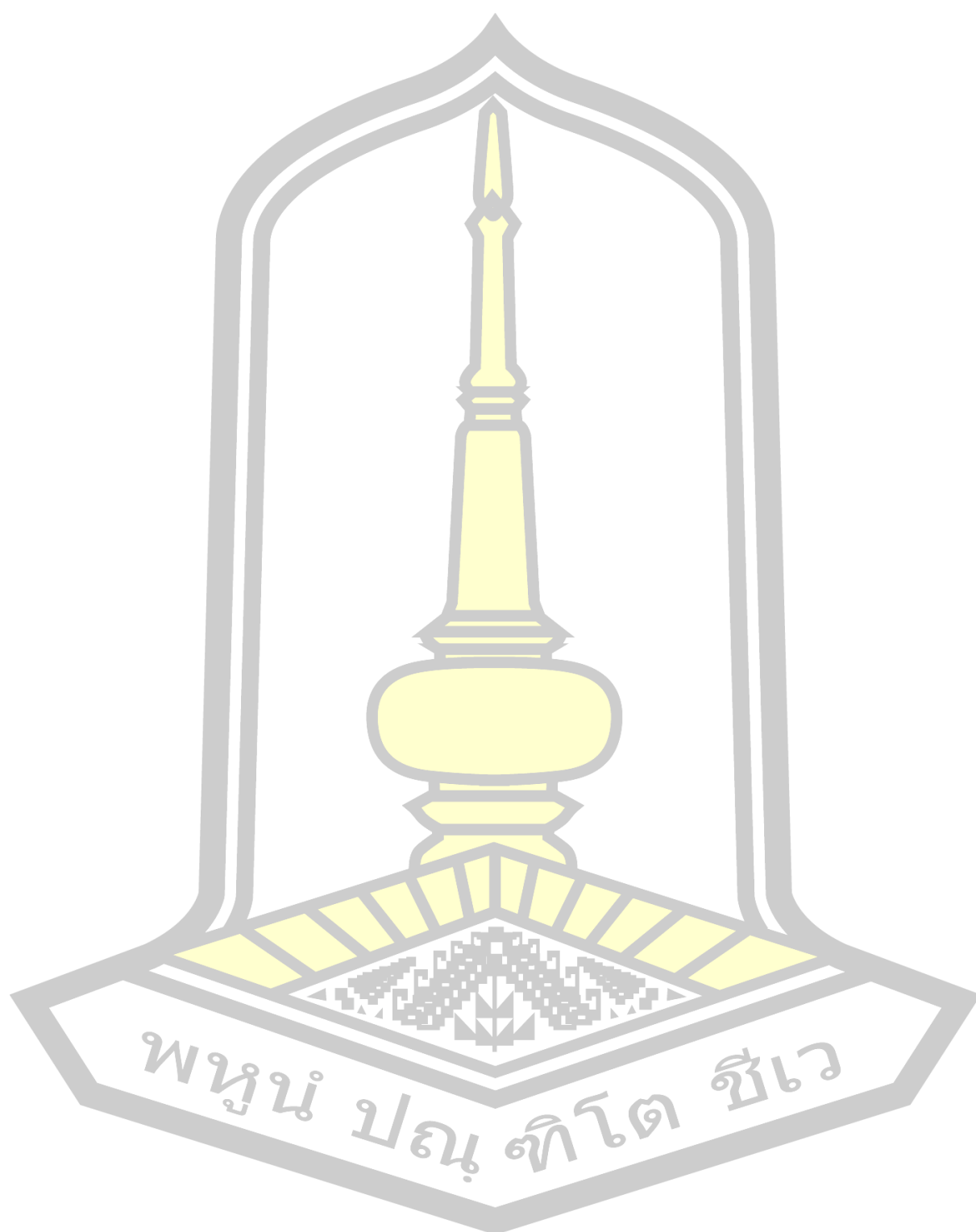
- Villela, H. F., Corrêa, F., Ribeiro, J. S. de A. N., Rabelo, A., & Carvalho, D. B. F. (2023). Fake news detection: a systematic literature review of machine learning algorithms and datasets. *Journal on Interactive Systems*, 14(1), 47–58.
- Vogel, I., & Meghana, M. (2020). Detecting fake news spreaders on twitter from a multilingual perspective. *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, 599–606.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, 359(6380), 1146-1151
- Weidinger, J., Schlauderer, S., & Overhage, S. (2021). Information Technology to the Rescue Explaining the Acceptance of Emergency Response Information Systems by Firefighters. *IEEE Transactions on Engineering Management*.
- Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. (2019). Unsupervised Fake News Detection on Social Media: A Generative Approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 5644–5651.
- Yu, C., Ma, Y., An, L., & Li, G. (2022). BCMF: A bidirectional cross-modal fusion model for fake news detection. *Information Processing & Management*, 59(5), 103063.
- Zabludovsky, K. (2023). buzzfeed. From <https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news>.
- Zanasi, A. (2000). *Data mining III*. WIT Press.
- Zhang, J., Bai, H., Lu, J., & Zheng, L. (2023). Problematic use of social media: The influence of social environmental forces and the mediating role of copresence. *Heliyon*, e12959.
- Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019). Fake News: Fundamental theories, detection strategies and challenges. *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, 836–837.
- Zortout (2022). What is Twitter or Twitter. From <https://zortout.com/blog/what-is-twitter>.

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., Procter, R., Liakata, M., & Procter, R.
 (2018). Detection and Resolution of Rumours in Social Media : A Survey. ACM
 Computing Surveys, 51(2), 32.



ภาคผนวก ก

แบบอนุมัติการพิจารณารับรองจริยธรรมวิจัยในมนุษย์แบบยกเว้น





คณะกรรมการจริยธรรมการวิจัยในคน มหาวิทยาลัยมหาสารคาม

เอกสารรับรองโครงการวิจัย

เลขที่การรับรอง : 529-582/2566

ชื่อโครงการวิจัย (ภาษาไทย) การปรับปรุงการตรวจจับข่าวปลอมด้วยอัลกอริทึมเวกเตอร์แมชชีนโดยวิธีการ
คัดเลือกคุณลักษณะที่เหมาะสม

ชื่อโครงการวิจัย (ภาษาอังกฤษ) Improving Fake News Detection with Support Vector Machines
Using Optimal Feature Selection.

ผู้วิจัย : นางสาวอัจฉราภรณ์ นาชัยทอง

หน่วยงานที่รับผิดชอบ : คณะการบัญชีและการจัดการ

สถานที่ทำการวิจัย : มหาวิทยาลัยมหาสารคาม

ประเภทการพิจารณาแบบ : แบบยกเว้น

วันที่รับรอง : 13 ธันวาคม 2566

วันหมดอายุ : 12 ธันวาคม 2567

ข้อเสนอการวิจัยนี้ ได้รับการพิจารณาและให้ความเห็นชอบจากคณะกรรมการจริยธรรมการวิจัยในคน
มหาวิทยาลัยมหาสารคามแล้ว และอนุมัติในด้านจริยธรรมให้ดำเนินการศึกษาวิจัยเรื่องข้างต้นได้ บนพื้นฐาน
ของโครงร่างงานวิจัยที่คณะกรรมการฯ ได้รับและพิจารณา เมื่อเสร็จสิ้นโครงการแล้วให้ผู้วิจัยส่งแบบฟอร์ม
การปิดโครงการและรายงานผลการดำเนินงานมายังคณะกรรมการจริยธรรมการวิจัยในคน มหาวิทยาลัย
มหาสารคาม หรือหากมีการเปลี่ยนแปลงใดๆ ในโครงการวิจัย ผู้วิจัยจะต้องยื่นขอรับการพิจารณาใหม่

ภคร์ สุทธิจิตร

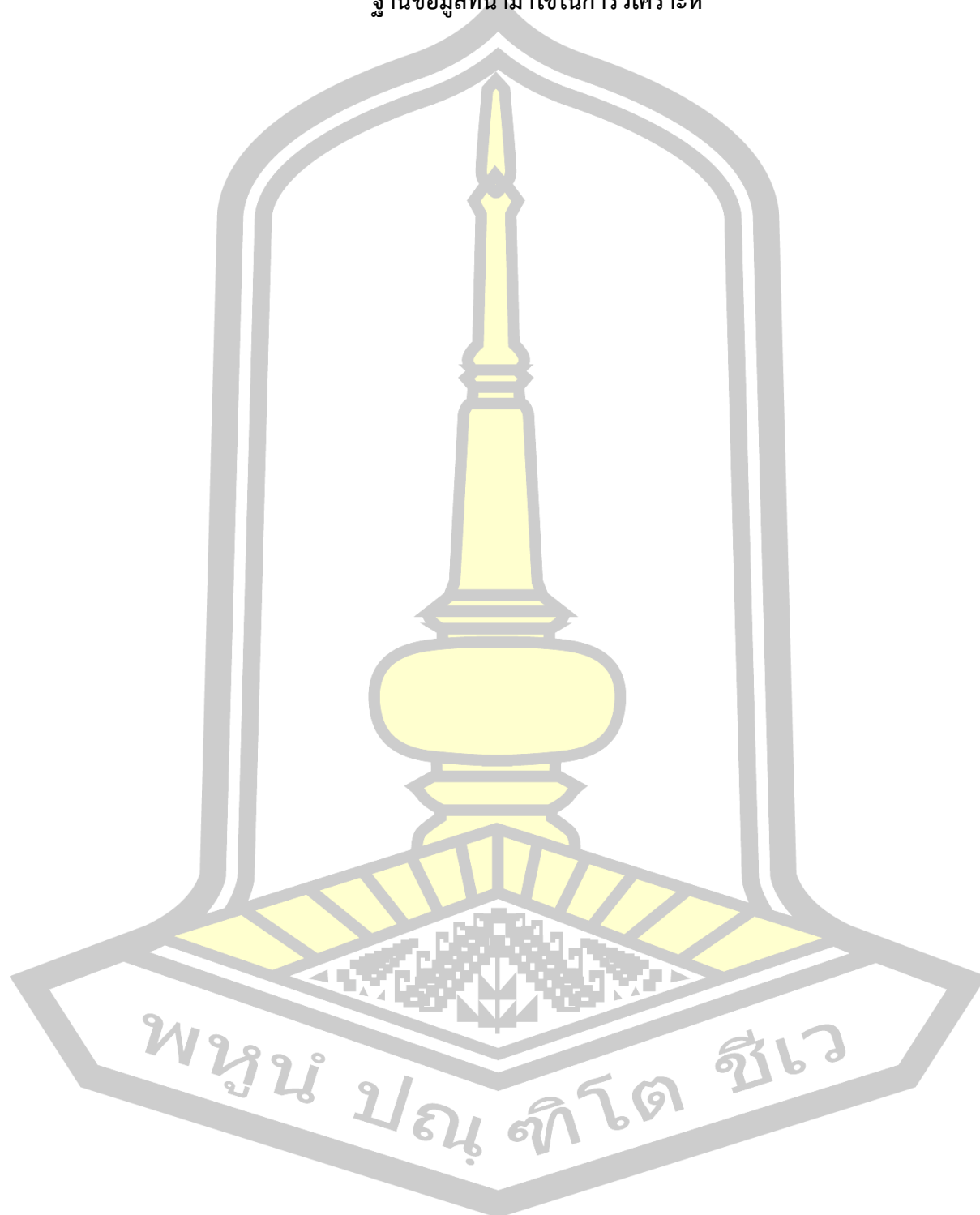
(ผู้ช่วยศาสตราจารย์ เกษียรหญิงรัตรี สว่างจิตร)

ประธานคณะกรรมการจริยธรรมการวิจัยในคน

มหาวิทยาลัยมหาสารคาม

ทั้งนี้ การรับรองนี้มีเงื่อนไขดังที่ระบุไว้ด้านหลังทุกข้อ (ดูด้านหลังของเอกสารรับรองโครงการวิจัย)

ภาคผนวก ข
ฐานข้อมูลที่ใช้ในการวิเคราะห์

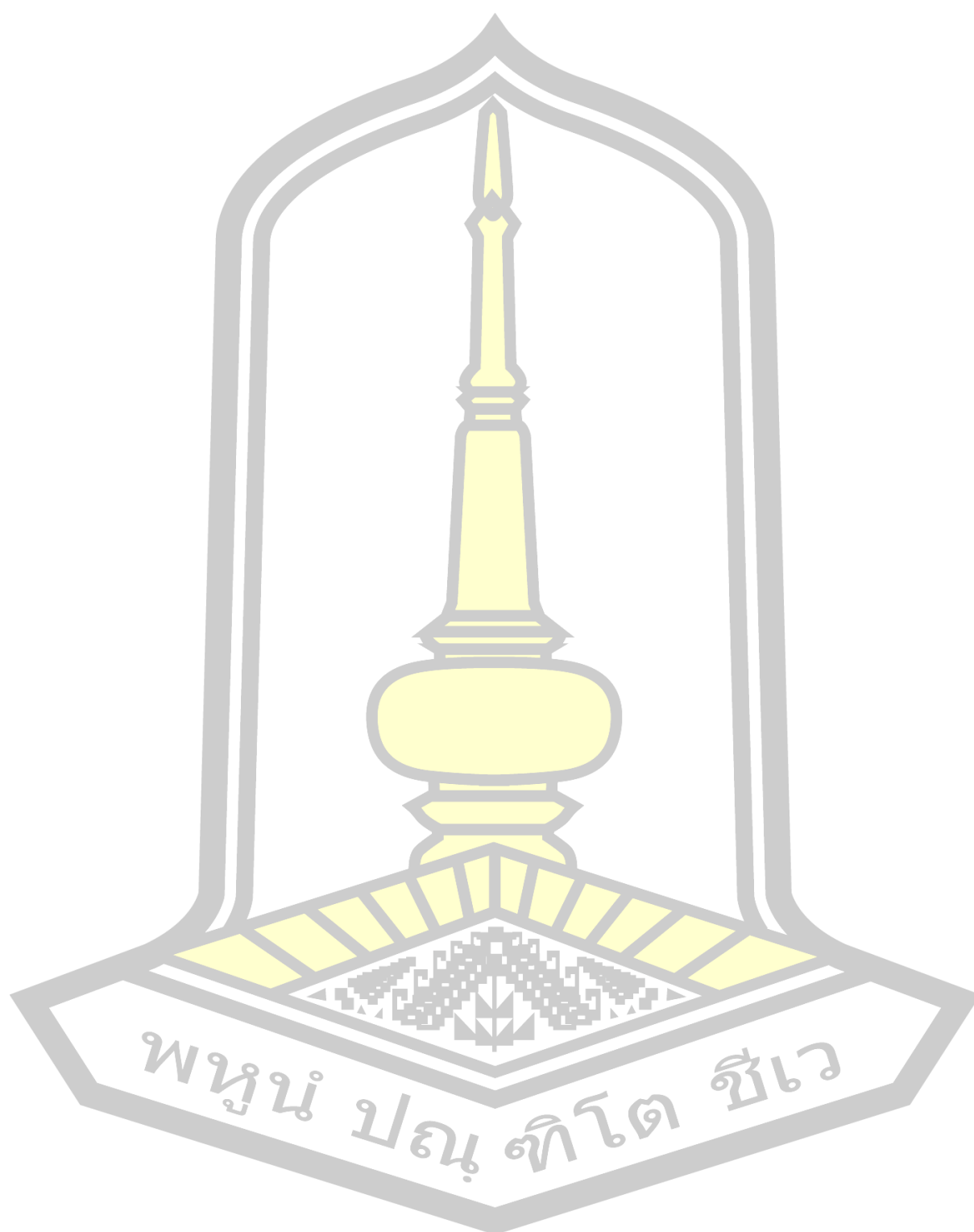


1	title	news_url	source_domain	weet_nu	label
2	Kandi Burruss Explodes Over Rape Acc	http://toofab.com/2017/05/08/real-housewives-atlan	toofab.com	42	1
3	People's Choice Awards 2018: The bes	https://www.today.com/style/see-people-s-choice-aw	www.today.com	0	1
4	Sophia Bush Sends Sweet Birthday Me	https://www.etonline.com/news/220806_sophia_bush	www.etonline.com	63	1
5	Colombian singer Maluma sparks rumo	https://www.dailymail.co.uk/news/article-3365543/C	www.dailymail.co.uk	20	1
6	Gossip Girl 10 Years Later: How Upper	https://www.zerchoo.com/entertainment/gossip-girl-	www.zerchoo.com	38	1
7	Gwen Stefani Got Dumped by Blake Sh	www.intouchweekly.com/posts/gwen-stefani-dumped	www.intouchweekly	45	0
8	Broward County Sheriff Fired For Lying	https://yournewswire.com/broward-county-sheriff-fir	yournewswire.com	124	0
9	Amber Rose Shuts Down French Mont	www.etonline.com/news/214798_amber_rose_shuts	www.etonline.com	4	0
10	Mindy Kaling makes first post-baby ap	https://www.aol.com/article/entertainment/2018/02/	www.aol.com	59	1
11	Katharine McPhee Butchers Tony Nom	https://www.98online.com/2018/05/02/katharine-mc	www.98online.com	10	1
12	â€˜WAGS Miamiâ€™ Stars Ashley Nico	https://www.usmagazine.com/celebrity-news/news/w	www.usmagazine.co	18	1
13	Mel Gibson: Hollywood Pedophiles Ha	yournewswire.com/mel-gibson-hollywood-pedophiles	yournewswire.com	95	0
14	Medium Tyler Henry Addresses The â€	https://perezhilton.com/tyler-henry-says-kristin-cavall	perezhilton.com	4	1
15	DWTS Season 27 Results: Week 5 - Dis	https://abc.go.com/shows/dancing-with-the-stars/nev	abc.go.com	0	1
16	5 Reasons Why Tarek El Moussa Will C	www.eonline.com/news/932478/5-reasons-why-tarek	www.eonline.com	63	0
17	David Cassidy cuts estranged daughter	https://www.thesun.co.uk/news/5086439/david-cassi	www.thesun.co.uk	54	1
18	Actress Brigitte Nielsen, 54, has given	https://www.usatoday.com/story/life/people/2018/0	www.usatoday.com	64	1
19	The Kardashians donate whopping sum	https://web.archive.org/web/20171228144534/http://	web.archive.org	1	0
20	Is Brad Pitt Open To Getting Back Toge	hollywoodlife.com/2017/08/27/would-brad-pitt-get-b	hollywoodlife.com	19	0
21	Pink Announces 2018 Tour and Drops	https://celebmix.com/pink-announces-2018-tour-drop	celebmix.com	120	1
22	A glimpse at the relationship of Megha	www.sacbee.com/news/nation-world/world/article21	www.sacbee.com	8	0
23	Johnny Depp, Charlie Sheen and Kevin	https://www.dailymail.co.uk/tvshowbiz/article-47076	www.dailymail.co.uk	91	1
24	Are Katie Holmes & Jamie Foxx Expect	hollywoodlife.com/2018/02/12/jamie-foxx-katie-holm	hollywoodlife.com	14	0
25	Ryan Phillippe shows off insanely chise	https://www.msn.com/en-us/tv/celebrity/ryan-phillip	www.msn.com	47	1
26	10 Pretty Celebrity Books That Would	https://stylecaster.com/celeb-coffee-table-books/	stylecaster.com	49	1
27	Pink Panics Before Hanging Off L.A. H	https://www.usmagazine.com/celebrity-news/news/p	www.usmagazine.co	1	1
28	Anne Hathaway Preemptively Shut Do	https://www.marieclaire.com/celebrity/a19699818/ar	www.marieclaire.co	83	1
29	WATCH: Kendall Jennerâ€™s first-ever	https://www.channel24.co.za/Gossip/News/watch-ke	www.channel24.co.za	83	1
30	'Bachelor' Alum Vienna Girardi Reveals	https://www.usmagazine.com/celebrity-moms/news/f	www.usmagazine.co	1	1
31	Teen Mom OG's Tyler Baltierra Reveal	https://people.com/tv/teen-mom-og-tyler-baltierra-re	people.com	74	1
32	Kim Kardashian West Says Her Kids Do	https://people.com/parents/kim-kardashian-west-kids	people.com	1	1
33	Why All Ladies Crush on Angelina Jolie	www.eonline.com/news/128009/why-all-ladies-crush-	www.eonline.com	32	0
34	Jennifer Lawrence Blasts Harvey Wein	https://www.msn.com/en-ie/news/other/jennifer-law	www.msn.com	30	1
35	The CW renews 'Supernatural,' 'Arrow	https://ew.com/tv/2018/04/02/cw-renewals-2018-20	ew.com	64	1
36	Harvey Weinstein Hired Investigators	https://variety.com/2017/film/news/harvey-weinstein	variety.com	42	1
37	Mario Lopez admits: 'I cheated on Ali	www.dailymail.co.uk/tvshowbiz/article-1356215/Mari	www.dailymail.co.uk	16	0
38	Watch Married at First Sightâ€™s Jona	https://dailynews.news/watch-married-at-first-sight03	dailynews.news	51	1
39	A Farewell to @midnight: Why the Cor	https://www.longroom.com/discussion/615871/a-fare	www.longroom.com	31	1
40	List of Who Beyonce Knowles Has Dat	https://www.ranker.com/list/beyonce-knowles_loves	www.ranker.com	0	1
41	Katy Perry Says She's Ready for a 'Sou	https://www.etonline.com/katy-perry-says-shes-ready	www.etonline.com	34	1
42	Rose Leslie Made Kit Harington Wear	http://time.com/4986364/rose-leslie-kit-harington-jon	time.com	1	1
43	Kate Middleton wears Diana's tiara at	https://www.dailymail.co.uk/femail/article-6460231/f	www.dailymail.co.uk	27	1
44	Michael Fassbender & Alicia Vikander	hollywoodlife.com/2015/09/17/alicia-vikander-michael	hollywoodlife.com	30	0
45	Ryan Lochte cleared of false-report ch	www.washingtonpost.com/news/early-lead/wp/2017/	www.washingtonpos	115	0
46	â€˜Bacheloretteâ€™ Star Chad Johnso	https://www.inquisitr.com/4336274/bachelorette-cha	www.inquisitr.com	0	1
47	Kourtney Kardashian's home swarming	https://www.dailymail.co.uk/tvshowbiz/article-53696	www.dailymail.co.uk	2	1
48	All About the \$4,200 Crib Kim Kardashi	https://people.com/parents/kim-kardashian-beyonce-	people.com	35	1
49	Kim Kardashian Insisted a Sick Kanye W	people.com/tv/kim-kardashian-insisted-kanye-west-gc	people.com	26	0
50	Audrina Patridge Divorce: Corey Bohar	https://www.usmagazine.com/celebrity-news/news/a	www.usmagazine.co	50	1
51	Mama June Trash Talks Daughter Pum	https://popculture.com/reality-tv/2018/02/02/mama-	popculture.com	19	1
52	"Scott Brought a Girl Out Here!" Disick	http://communitynewsblog.com/2017/05/scott-broug	communitynewsblog	1	1
53	Rachel Brosnahan attends Tony Award	https://www.aol.com/article/entertainment/2018/06/	www.aol.com	79	1
54	Norvell Presents "Dancing With The St	https://smarttan.com/news/index.php/norvell-present	smarttan.com	1	1

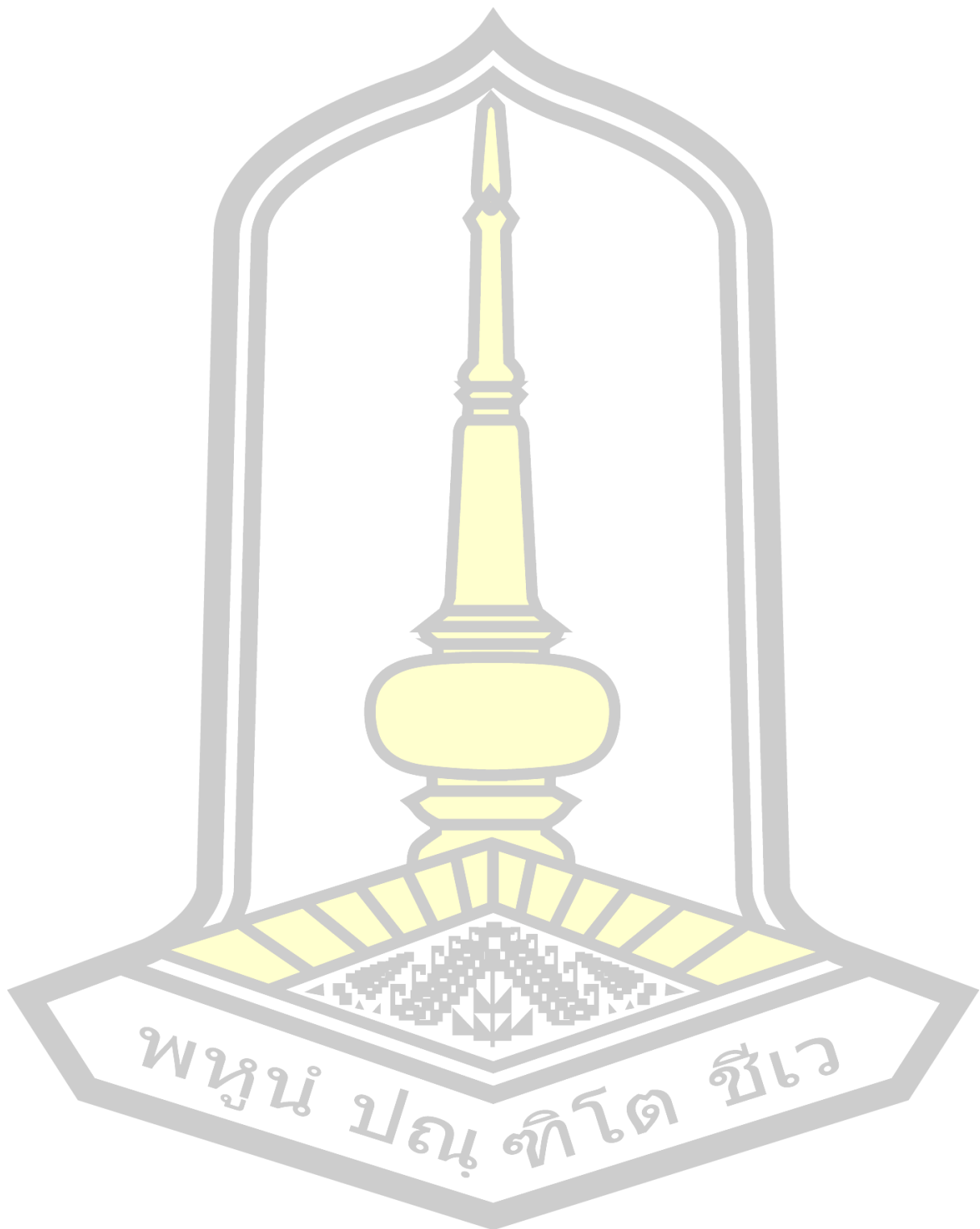
55	Saved By The Bell's Mark-Paul Gossela	https://www.dailymail.co.uk/tvshowbiz/article-3441577/saved-by-the-bells-mark-paul-gossela.html	www.dailymail.co.uk	0	1
56	Sofia Richie to Kourtney Kardashian: S	www.thehollywoodgossip.com/2018/08/sofia-richie-to-kourtney-kardashian/	www.thehollywoodgossip.com	342	0
57	Would you consider a colored lip balm	https://www.quora.com/Would-you-consider-a-colored-lip-balm	www.quora.com	44	1
58	RHOMiami Alum Joanna Krupa & Romain Zago	https://people.com/tv/joanna-krupa-romain-zago-fina	people.com	70	1
59	Kate Middleton Really Wants to Get Pr	www.lifeandstylemag.com/posts/kate-middleton-prince-william	www.lifeandstylemag.com	3	0
60	KJ Apa Roasts Cole Sprouse on Instagram	https://www.teenvogue.com/story/kj-apa-roasts-cole-sprouse	www.teenvogue.com	65	1
61	Less Bounce? Iggy Azalea Says She Los	https://people.com/bodies/iggy-azalea-lost-15-lbs-two	people.com	93	1
62	Inside Matt Lauer's Quiet Life 1 Year A	https://www.msn.com/en-us/tv/celebrity/inside-matt-lauer	www.msn.com	57	1
63	Sean Penn Talks Haiti, Humanitarianism	https://hauteliving.com/2016/11/sean-penn-talks-haiti/	hauteliving.com	39	1
64	J Balvin Lives the Moment With His Dre	https://www.billboard.com/articles/columns/latin/821	www.billboard.com	28	1
65	Anthony Michael Hall Sentenced to Th	www.hollywoodreporter.com/news/anthony-michael-hall-sentenced-to-three-years	www.hollywoodreporter.com	59	1
66	Jada Pinkett Smith Not Leaving Will Sm	www.imdb.com/news/ni61255200	www.imdb.com	4	0
67	Serena loses in exhibition comeback at	https://www.foxsports.com/tennis/story/serena-loses-in-exhibition-comeback-at	www.foxsports.com	62	1
68	Thomas Rhett & Wife Open Up About	https://countryfancast.com/thomas-rhett-adoption-pr	countryfancast.com	15	1
69	Britney Spears Agrees to Pay Kevin Fed	https://people.com/music/britney-spears-child-support	people.com	138	1
70	12 photos of celebrities not wearing th	https://www.thisinsider.com/celebrities-not-wearing-their-clothes	www.thisinsider.com	2	1
71	Luke Bryan Jokes Gwen Stefani Is Too	www.usmagazine.com/celebrity-news/news/luke-bryan-jokes-gwen-stefani-is-too	www.usmagazine.com	12	0
72	Cardi B Has Some Advice for Everyone	www.glamour.com/story/cardi-b-khloe-kardashian	www.glamour.com	65	1
73	Blake Lively Condemns Harvey Weinst	https://www.usmagazine.com/celebrity-news/news/blake-lively-condemns-harvey-weinstein	www.usmagazine.com	12	1
74	See the Hilarious Christmas Gift Jenni	https://people.com/tv/jennifer-lawrence-kris-jenner-p	people.com	74	1
75	Investigating Celebrity Parenting Trend	http://wushack.com/m/feed/view/Investigating-Celebrity-Parenting-Trend	wushack.com	50	1
76	Together Again! Every Photo of Megha	https://people.com/royals/meghan-markle-kate-middleton	people.com	18	1
77	Ben Affleck Gifted A Pet Pooch On Wh	www.x17online.com/2017/10/ben-affleck-last-day-of-his-life	www.x17online.com	6	0
78	Kylie Jenner Says Daughter Stormi Loo	https://www.usmagazine.com/celebrity-moms/news/kylie-jenner-says-daughter-stormi-loo	www.usmagazine.com	1	1
79	No, ZAYN Is Not 'Desperate To Defend	www.953bull.iheart.com/content/2017-09-02-no-zayn-is-not-desperate-to-defend	www.953bull.iheart.com	60	0
80	Angelina Jolie Talks Brad Pitt, Bell's P	www.hollywoodreporter.com/rambling-report/angelina-jolie-talks-brad-pitt-bells-pinkett-smith	www.hollywoodreporter.com	1	1
81	Cara Delevingne claims Harvey Weinst	https://www.thesun.co.uk/news/4664289/cara-delevingne-claims-harvey-weinstein	www.thesun.co.uk	50	1
82	Watch Britney Spears End Las Vegas R	www.billboard.com/articles/columns/pop/808	www.billboard.com	1	1
83	Better Call Saul	https://en.wikipedia.org/wiki/Better_Call_Saul	en.wikipedia.org	48	1
84	Kendra Wilkinson Says She Signed Her	https://people.com/tv/kendra-wilkinson-signed-last-day-of-her-life	people.com	286	1
85	Kim Kardashian Begs Tristan Thompson	www.usmagazine.com/celebrity-news/news/kim-kardashian-begs-tristan-thompson	www.usmagazine.com	139	0
86	Ed Sheeran admits he was able to bed	metro.co.uk/2017/03/07/ed-sheeran-admits-he-was-able-to-bed	metro.co.uk	24	0
87	May 2017 à€“ Page 180 à€“ AMNAYZ	https://amnayzh.wordpress.com/2017/05/page/180/	amnayzh.wordpress.com	27	1
88	Paris Hilton Talks Being a Kim Kardashi	https://www.etonline.com/paris-hilton-talks-being-a-kim-kardashian	www.etonline.com	2	1
89	Selena Gomez And The Weeknd Are Da	www.elle.com/uk/life-and-culture/culture/news/selena-gomez-and-the-weeknd-are-dating	www.elle.com	50	1
90	'What Really Hurt': Jay-Z Opens Up Ab	https://www.nbclosangeles.com/entertainment/entertainment/jay-z-opens-up-about-what-really-hurt	www.nbclosangeles.com	1	1
91	'Fixer Upper' stars Chip and Joanna Ga	https://www.wptv.com/entertainment/-fixer-upper-stars-chip-and-joanna-gale	www.wptv.com	6	1
92	We Don't Know Why Matthew Rhys W	www.bravotv.com/the-daily-dish/we-dont-know-why-matthew-rhys-wrote-the-book	www.bravotv.com	13	1
93	We tried the Starbucks Zombie Frappu	https://www.chicagotribune.com/dining/ct-starbucks-zombie-frappuccino	www.chicagotribune.com	48	1
94	Kim Kardashian slammed for 'ridiculou	www.aol.com/article/entertainment/2018/10/01/kim-kardashian-slammed-for-ridiculous-photos	www.aol.com	1	1
95	Netflix Will Stop Streaming Marvel Mo	https://www.inverse.com/article/47886-marvel-movie-streaming	www.inverse.com	63	1
96	Ariana Grande and Jimmy Fallon Break	https://onehallyu.com/topic/676684-ariana-grande-and-jimmy-fallon-break-up	onehallyu.com	45	1
97	Alex Rodriguez Shares Never-Before-S	https://www.etonline.com/alex-rodriguez-shares-never-before-seen-photos	www.etonline.com	49	1
98	Rob Pattinson Still à€“Cares Deeplyà€”	hollywoodlife.com/2018/06/06/are-robert-pattinson-fully-over-his-teen-idol-life	hollywoodlife.com	3	0
99	Venus Williams, other driver cleared in	https://www.cnn.com/2017/12/21/us/venus-williams-tennis/index.html	www.cnn.com	52	1
100	The Last Black Unicorn: Tiffany Haddis	https://www.amazon.com/Last-Black-Unicorn-Tiffany-Haddis	www.amazon.com	25	1
101	Bachelor in Paradise: Taylor Nolan and	https://people.com/tv/bachelor-in-paradise-taylor-nolan	people.com	32	1
102	Saturday Savings: Jennifer Lopez's Cro	http://nextdivas.com/2018/01/27/saturday-savings-jennifer-lopez	nextdivas.com	147	1
103	Celebrity Marriage: Kendra Wilkinson T	http://cupidspulse.com/123876/celebrity-marriage-kendra-wilkinson	cupidspulse.com	73	1
104	Kim Kardashian Is à€“Over Being Blond	http://celebrityinsider.org/kim-kardashian-is-over-being-blond	celebrityinsider.org	1	1
105	Meryl Streep And Nicole Kidman Feud	www.inquisitr.com/4936378/meryl-streep-nicole-kidman-feud	www.inquisitr.com	1	0
106	Bekah M. Explains Why Her Mother Re	https://www.bustle.com/p/bekah-m-explains-why-her-mother-rejected-her-when-she-was-a-teen	www.bustle.com	37	1
107	Scheana Shay Felt She Was "Harsh" an	http://www.bravotv.com/the-daily-dish/scheana-shay	www.bravotv.com	53	1
108	Hurricane Maria could be a \$95 billion	https://money.cnn.com/2017/09/28/news/economy/hurricane-maria/index.html	money.cnn.com	20	1

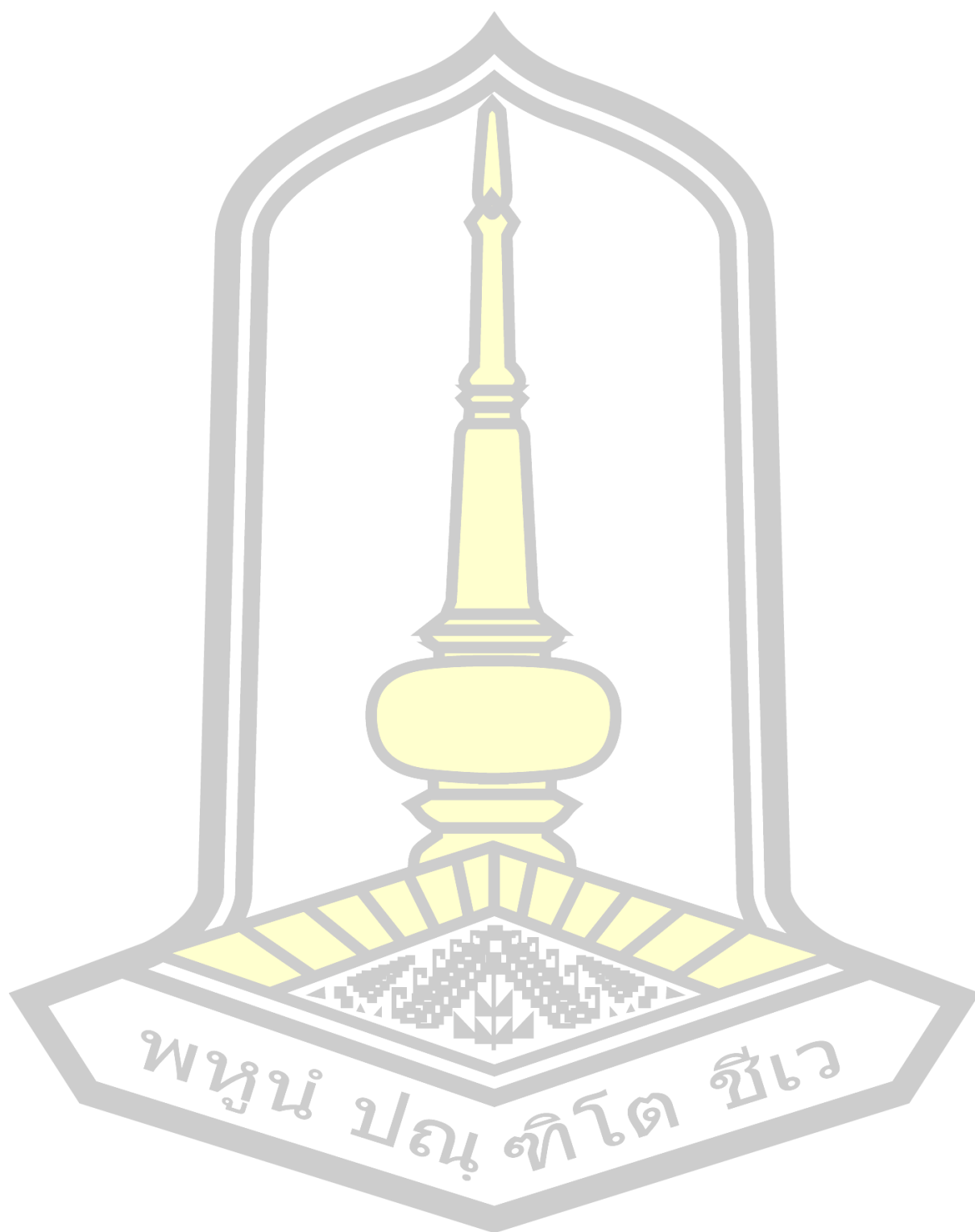
109	9 Ways 'Riverdale' and 'Gossip Girl' Are	https://www.tvinsider.com/659459/riverdale-gossip-g	www.tvinsider.com	36	1
110	The Next 'Bachelor' Won't Be Peter Kr	https://www.bustle.com/p/the-next-bachelor-wont-b	www.bustle.com	63	1
111	Ivanka Trump Us Weekly Cover Meme	www.harpersbazaar.com/celebrity/latest/a9989599/i	www.harpersbazaar.	319	0
112	SparkNotes: The Handmaid's Tale:	https://www.sparknotes.com/lit/handmaid/summary/	www.sparknotes.co	66	1
113	Ed Sheeran says he broke his arm in bil	https://www.cbsnews.com/news/ed-sheeran-breaks-e	www.cbsnews.com	66	1
114	Police recover Elizabeth Smart and arr	https://www.history.com/this-day-in-history/police-re	www.history.com	63	1
115	Kelly Rohrbach cosies up to shirtless be	https://www.dailymail.co.uk/tvshowbiz/article-60204	www.dailymail.co.uk	115	1
116	Tarek and Christina El Moussa Produce	https://people.com/home/tarek-el-moussa-on-produc	people.com	20	1
117	The Walking Dead season 9: The reaso	https://www.independent.co.uk/arts-entertainment/t	www.independent.co	24	1
118	Ellen DeGeneres' wife Portia de Rossi	www.aol.com/article/entertainment/2018/02/01/elle	www.aol.com	28	0
119	â€œTotal Bellasâ€™ Renewed For Seas	https://deadline.com/2018/08/total-bellas-renewed-s	deadline.com	25	1
120	LeBron James And Other Sports Stars	https://news.shaneperrymarketing.com/lebron-james-	news.shaneperrymar	70	1
121	SAG Awards 2018: Millie Bobby Brown	https://www.usmagazine.com/stylish/news/sag-awar	www.usmagazine.co	20	1
122	Halsey Gushing About G-Eazy your hea	https://www.archyworldys.com/halsey-gushing-about	www.archyworldys.c	1	1
123	Brad Pitt And Angelina Jolie's Divor	celebrityinsider.org/brad-pitt-and-angelina-jolies-divor	celebrityinsider.org	6	0
124	Rihanna gives a flash of cleavage at N	https://www.dailymail.co.uk/tvshowbiz/article-48715	www.dailymail.co.uk	67	1
125	10 Times Kim Kardashian Was Absolute	https://people.com/tv/kim-kardashian-savage-sisters-	people.com	30	1
126	How to Wear Gingham This Spring	https://fashionista.com/2015/03/gingham	fashionista.com	25	1
127	Coco Austin and daughter Chanel don	https://www.dailymail.co.uk/tvshowbiz/article-54662	www.dailymail.co.uk	13	1
128	Gwyneth Paltrow Poses Topless in Mu	https://www.usmagazine.com/stylish/news/gwyneth-t	www.usmagazine.co	28	1
129	Media Wall vs Step and Repeat vs Log	www.stepandrepeat.com/blog/2016/04/media-	www.stepandrepeat	55	1
130	Full List of Hip Hop Awards 2017 Winn	https://www.bet.com/shows/hip-hop-awards/2017/p	www.bet.com	122	1
131	WATCH: Ryan Gosling tells how he sav	https://www.dnaindia.com/television/report-watch-ry	www.dnaindia.com	2	1
132	Joanna Gaines on Why Expecting Her F	https://people.com/parents/joanna-gaines-pregnant-f	people.com	52	1
133	Billy Bush Breaks His Silence on Trump	www.hollywoodreporter.com/features/billy-b	www.hollywoodrepc	58	1
134	Darren Aronofsky	https://en.wikipedia.org/wiki/Darren_Aronofsky	en.wikipedia.org	76	1
135	See Meghan Markle (in Navy!) and Prin	https://people.com/royals/princess-eugenie-royal-we	people.com	25	1
136	Ashley Benson Channels the Olsen Twi	www.storiesflow.com/article/14969451709/a	www.storiesflow.co	29	1
137	Michael Strahan, Sara Haines Will Host	variety.com/2018/tv/news/michael-strahan-sara-hain	variety.com	1015	0
138	WATCH LIVE: Tim Kaine and Corey Ste	https://www.youtube.com/watch?v=jC8F3gd5_XY	www.youtube.com	25	1
139	Singer Tina Turner: â€œThis Thanksgiv	https://web.archive.org/web/20171127001649/http://	web.archive.org	35	0
140	Britney Spears	https://en.wikipedia.org/wiki/Britney_Spears	en.wikipedia.org	57	1
141	The Dad Bag	https://thedadbag.com/	thedadbag.com	51	1
142	Gal Gadot Plays 'Box of Lies' On 'Tonig	https://www.thewrap.com/gal-gadot-fallon-lies-video	www.thewrap.com	63	1
143	Gwen Stefani Is Still Smitten With Blak	www.etonline.com/gwen-stefani-still-smitten-blake-sf	www.etonline.com	14	0
144	Brad Pitt 'not in a relationship with Ne	www.dailymail.co.uk/tvshowbiz/article-5891231/Brad	www.dailymail.co.uk	10	0
145	The Estates of Carrie Fisher & Debbie	https://www.apartmenttherapy.com/the-estates-of-c	www.apartmentther	110	1
146	Former reality star Jon Gosselin reveal	https://rare.us/people/former-reality-star-jon-gosselin	rare.us	54	1
147	Did Channing Tatum Cheat on Jenna D	www.newsweek.com/channing-tatum-work-family-je	www.newsweek.com	3	0
148	J Balvin Reveals If He's Related to	https://medium.com/@AndreAguirre25111NTZ/j-balvi	medium.com	69	1
149	Heath Ledger	https://en.wikipedia.org/wiki/Heath_Ledger	en.wikipedia.org	55	1
150	Tyler Henry Speaks with Kristin Cava	https://people.com/tv/tyler-henry-kristin-cavallari-bro	people.com	37	1
151	Serena Williams Channels Hollywood	www.storiesflow.com/article/15433861709/se	www.storiesflow.co	13	1
152	The Latest: 'Three Billboards' wins bes	www.nbc4i.com/news/u-s-world/the-latest-heavy-se	www.nbc4i.com	25	0
153	Are Chloe Lukasiak & Maddie Ziegler S	https://www.bustle.com/articles/108020-are-chloe-lu	www.bustle.com	0	1
154	Gordon Ramsay Says He Dropped Over	www.goodhousekeeping.com/health/diet-nutrition/ne	www.goodhousekee	32	0
155	Erika Girardi Reacts to Eileen Davidso	www.bravotv.com/the-daily-dish/eileen-davids	www.bravotv.com	72	1
156	Taylor Swift teases 2018 tour, creates	https://www.nickiswift.com/82504/taylor-swift-tease	www.nickiswift.com	24	1
157	Mariska Hargitay Shines in Law & Orde	www.trueara.com/mariska-hargitay-shines-in-l	www.trueara.com	1	1
158	Normani Kordei, Dancing With the Star	https://heavy.com/entertainment/2017/05/normani-k	heavy.com	53	1
159	Watch: New â€œMary Poppins Returns	http://collider.com/mary-poppins-returns-featurette/	collider.com	73	1
160	Justin Bieber & Selena Gomez's Rel	bigeye.ug/justin-bieber-selena-gomez-s-rekindled-roma	bigeye.ug	29	0

161	Jennifer Lawrence	https://en.wikipedia.org/wiki/Jennifer_Lawrence	en.wikipedia.org	65	1
162	Serena Williams Shares Her 10 Favorite	https://www.etonline.com/serena-williams-shares-her	www.etonline.com	73	1
163	WWE's Brie Bella Gives Birth: 3 Reasons	https://www.forbes.com/sites/alfredkonuwa/2017/05/	www.forbes.com	0	1
164	Amy Schumer's National Bikini Day 'Gr	https://www.instyle.com/news/amy-schumer-national	www.instyle.com	83	1
165	Celebrities Share Tributes to Victims of	www.instyle.com/news/barcelona-terror-attacks-cele	www.instyle.com	19	0
166	Kate Middleton and Prince William Ma	www.vogue.com/article/kate-middleton-and-prince-w	www.vogue.com	2	0
167	Nicky Jam: El Ganador, Starring The Sir	https://www.forbes.com/sites/veronicavillafane/2018	www.forbes.com	51	1
168	Mike Tyson Offered SERIOUS Ca\$h To	radaronline.com/videos/mike-tyson-offered-cash-to-f	radaronline.com	9	0
169	Brad Pitt Now Moonlighting in the Furr	www.thecut.com/2012/11/pitt-now-moonlighting-in-t	www.thecut.com	29	0
170	Ashton Kutcher has guys' night with 'Th	www.aol.com/article/2015/12/14/ashton-kutcher-has	www.aol.com	34	0
171	Jimmy Kimmel blasts Trump, Graham, C	https://www.businessinsider.com/jimmy-kimmel-graham	www.businessinsider	2	1
172	Even before Jenelle Evans's gun inci	https://durangoherald.com/articles/233325	durangoherald.com	78	1
173	Jessie James Decker Gets Hot and Bot	http://www.trueara.com/jessie-james-decker-gets-hot	www.trueara.com	16	1
174	Selena Gomez Pregnant With The Wee	hollywoodlife.com/2017/02/01/selena-gomez-pregnant	hollywoodlife.com	11	0
175	Khloe Kardashian & Tristan Thompson	hollywoodlife.com/2017/04/30/khloe-kardashian-trist	hollywoodlife.com	26	0
176	Gwen Stefani Call Off Blake Shelton		NA	85	0
177	Kim Kardashian apologizes for her com	http://uk.businessinsider.com/kim-kardashian-aaliyah-uk	uk.businessinsider.co	55	1
178	Inside Danica Patrick's Post-Racing Life	https://www.longroom.com/discussion/1105675/inside	www.longroom.com	44	1
179	Kim On Crumbling Marriage To Kanye:	radaronline.com/videos/kim-kardashian-kanye-west-n	radaronline.com	7	0
180	See Courtney Love Chew Scenery in M	https://www.rollingstone.com/tv/tv-news/see-courtney	www.rollingstone.co	59	1
181	Frances Bean Cobain hopes her dad w	https://www.cnn.com/2018/04/11/entertainment/fran	www.cnn.com	57	1
182	Khloe Kardashian in awe of sister Kim	www.foxnews.com/entertainment/khloe-kard	www.foxnews.com	99	1
183	Rob Kardashian Shows Support for Ex	www.usmagazine.com/celebrity-news/news/ro	www.usmagazine.co	19	1
184	Alicia Keys's 3-Year-Old Son Is an Ar	https://parade.com/651027/lindsaylowe/alicia-keys-3	parade.com	48	1
185	Hottest Celebrity Gossip and Entertain	http://www.qwekusblog.com/2017/06/kim-kardashian	www.qwekusblog.co	40	1
186	Michael Phelps and Wife Nicole Throw	www.etonline.com/michael-phelps-and-wife-n	www.etonline.com	14	1
187	Why Kate Middleton and Prince William	www.travelandleisure.com/travel-tips/celebrity-travel	www.travelandleisur	1	0
188	This GIF of Game of Thrones' Jon Snow	https://www.longroom.com/discussion/638419/this-g	www.longroom.com	24	1
189	JayZ Dazed Confused	www.dazeddigital.com/jay-z	www.dazeddigital.co	1007	0
190	20 Crazy Details Behind The Making Of	screenrant.com/halloween-town-movie-trivia-t	screenrant.com	25	1
191	Ryan Seacrest, 43, packs on the PDA w	https://www.dailymail.co.uk/tvshowbiz/article-60920	www.dailymail.co.uk	1	1
192	Newlyweds Pippa Middleton and Jame	https://www.dailymail.co.uk/news/article-4529482/Pi	www.dailymail.co.uk	74	1
193	Spice Girls	https://en.wikipedia.org/wiki/Spice_Girls	en.wikipedia.org	31	1
194	This Labor Day, we need protests	http://www.pbs.org/wnet/need-to-know/economy/th	www.pbs.org	0	1
195	Shawniece Jackson and Jephthe Pierre V	https://people.com/parents/shawniece-jackson-jephthe	people.com	49	1
196	Scott Disick Is Reportedly Considering	thestir.cafemom.com/celebrities/209022/scott-disick	thestir.cafemom.co	0	0
197	Jennifer Aniston and Justin Theroux: Tl	www.cheatsheet.com/entertainment/jennifer-aniston	www.cheatsheet.co	5	0
198	Simon Cowell rushed to hospital after	www.telegraph.co.uk/news/2017/10/27/simon	www.telegraph.co.u	265	1
199	Watch Jenna Dewan-Tatum Totally Gu	http://bojtv.com/watch-jenna-dewan-tatum-totally-gu	bojtv.com	1	1
200	Ariana Grande and Pete Davidson: a ti	https://www.cosmopolitan.com/uk/entertainment/a2	www.cosmopolitan.	67	1
201	Kerry Washington Loves Being Part of	tlcnaptown.com/2346215/kerry-washington-lo	tlcnaptown.com	0	1
202	Kaskade's Pool Party Dos and Don'ts	www.news9ontime.com/kaskades-pool-party-	www.news9ontime.c	2	1
203	Manchester bomber is linked to Brusse	www.dailymail.co.uk/news/article-4536624/Se	www.dailymail.co.uk	41	1
204	Rumor Bust! Katie Holmes Did NOT Br	extratv.com/2018/04/21/rumor-bust-katie-holmes-did	extratv.com	2	0
205	Chanel Iman and Sterling Shepard Are	www.usmagazine.com/celebrity-moms/news/c	www.usmagazine.co	69	1
206	Paris Hilton thinks she could be like Pri	https://metro.co.uk/2018/12/08/paris-hilton-believes	metro.co.uk	1	1
207	Jay-Z admits his marriage to Beyonc	https://www.usatoday.com/story/life/entertainment/20	www.usatoday.com	32	1
208	Jennifer Lopez, Alex Rodriguez raise \$3	www.upi.com/Jennifer-Lopez-Alex-Rodriguez-r	www.upi.com	2	1
209	US report: "Kim Kardashian announce	www.newidea.com.au/kim-kardashian-kanye-west-an	www.newidea.com.a	1	0
210	Rihanna And Hassan Jameel Romance	snobette.com/2017/07/rihanna-hassan-jameel-relatio	snobette.com	28	0
211	Kate Middleton Wears an Emilia Wicks	www.harpersbazaar.com/celebrity/latest/a25	www.harpersbaza	57	1
212	The "Manus x Machina" Experience	https://www.vogue.com/projects/13433750/met-gala	www.vogue.com	59	1
213	Stephen Colbert To Be Jimmy Kimmel's	https://deadline.com/2017/09/stephen-colbert-jimmy	deadline.com	34	1
214	Chris Harrison on His Next Big Step and	https://www.etonline.com/chris-harrison-on-his-next	www.etonline.com	79	1



บรรณานุกรม





ประวัติผู้เขียน

ชื่อ	นางสาวอัจฉราภรณ์ นาชัยทอง
วันเกิด	9 เมษายน 2526
สถานที่เกิด	จังหวัดกาฬสินธุ์
สถานที่อยู่ปัจจุบัน	504 หมู่ 23 ต.เหนือเมือง อ.เมืองร้อยเอ็ด จ.ร้อยเอ็ด 46110
ตำแหน่งหน้าที่การงาน	อาจารย์ประจำสาขาวิชาคอมพิวเตอร์ธุรกิจดิจิทัล
สถานที่ทำงานปัจจุบัน	คณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด
ประวัติการศึกษา	พ.ศ. 2547 ประกาศนียบัตรวิชาชีพ (ปวช.) สาขาคอมพิวเตอร์ธุรกิจ โรงเรียนพานิชยการกาฬสินธุ์ พ.ศ. 2549 ประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) สาขาคอมพิวเตอร์ ธุรกิจ สถาบันเทคโนโลยีราชมงคลกาฬสินธุ์ พ.ศ. 2552 บริหารธุรกิจบัณฑิต (บธ.บ) สาขาวิชาการจัดการทั่วไป มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตกาฬสินธุ์ พ.ศ. 2554 ปริญญาคอมพิวเตอร์ศึกษา (ค.ม) สาขาวิชาคอมพิวเตอร์ศึกษา มหาวิทยาลัยราชภัฏมหาสารคาม พ.ศ. 2567 ปริญญาบริหารธุรกิจและนวัตกรรมดิจิทัล (ปร.ด.) สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล มหาวิทยาลัยมหาสารคาม
ทุนวิจัย	เป็นผู้ทำผลงานทางวิชาการที่ผลรวมถ่วงน้ำหนักสูงสุด ประจำปีการศึกษา 2565
ผลงานวิจัย	การประยุกต์ใช้เทคโนโลยีดิจิทัลในการส่งเสริมการบริหารจัดการกลุ่ม วิสาหกิจชุมชน กรณีศึกษา : บ้านจั่ว ตำบลธัญธานี อำเภอธวัชบุรี จังหวัด ร้อยเอ็ด.