ตัวแบบพยากรณ์การตรวจจับข่าวปลอมด้วยเทคนิคการเรียนรู้แบบรวมกลุ่ม

A predictive model for fake news detection using ensemble learning techniques

ธงชัย แก้วกิริยา[1] และ กาญจนา ศิลาวราเวทย์[2]

Thongchai Kaewkiriya[1] and Kanchana Silawarawet[2]

Received: 6 August 2024 ; **Revised**: 30 September 2024 ; **Accepted**: 11 November 2024

บทคัดย่อ

ในปัจจุบันนี้ สื่อสังคมออนไลน์เป็นหนึ่งในกิจกรรมสำคัญของประชาชนทั่วไป เพื่อวัตถุประสงค์ในการแบ่งปันข่าวสารและเรื่องราวต่าง ๆ แก่ผู้คน ด้วยเหตุนี้ จึงก่อให้เกิดกลุ่มผู้ไม่ประสงค์ดีที่ต้องการหาผลประโยชน์แก่ตนเองและก่อความวุ่นวายในสังคมด้วยการสร้างข่าวลวงและเผยแพร่แก่บุคคลทั่วไป ในช่วงเวลาที่ผ่านมาได้มีความพยายามในการสร้างระบบสำหรับตรวจจับข่าวลวงบนสื่อสังคมออนไลน์ได้โดยอัตโนมัติ เช่น การตรวจจับโดยใช้การเรียนรู้ของเครื่องจักรเรียนรู้ข้อมูลข่าวสาร หรือ การใช้ระบบการเรียนรู้พฤติกรรมของผู้คนบนสื่อสังคมออนไลน์ เป็นต้น ซึ่งจากวิธีการที่ได้กล่าวมา จึงก่อให้เกิดแนวคิดของการบูรณาการการวิเคราะห์ข้อมูลแบบหลายปัจจัย เพื่อเสริมความถูกต้องของการทำนายระบบตรวจจับข่าวลวงให้มีประสิทธิภาพมากขึ้น ในงานวิจัยฉบับนี้จึงมีวัตถุประสงค์เพื่อนำเสนอตัวแบบพยากรณ์ในการตรวจจับข่าวปลอม ด้วยเทคนิคการเรียนรู้แบบรวมกลุ่มด้วยวิธีการโหวต โดยประกอบไปด้วยชุดข้อมูล 20,000 ชุดข้อมูล ซึ่งพิจารณาปัจจัยแหล่งที่มาของสื่อประกอบและปฏิสัมพันธ์ของผู้ใช้งานต่าง จากผลการทดสอบร่วมกับฐานข้อมูลข่าวลวงซึ่งบันทึกจากสื่อสังคมออนไลน์อย่าง Fakeddit.com พบว่าแบบจำลองที่นำเสนอมีประสิทธิภาพความถูกต้องในการพยากรณ์เท่ากับ 96.97% ซึ่งมากกว่าเทคนิคการเรียนรู้อื่นที่นำมาใช้ประกอบการเรียนรู้

คำสำคัญ: การพยากรณ์ข้อมูล, การตรวจจับข่าวปลอม, การเรียนรู้ของเครื่อง, การเรียนรู้แบบรวมกลุ่ม

Abstract

Nowadays, social media is one of the important activities for the general public for the purpose of sharing news and various stories with people. As a result, there are ill-intentioned individuals who seek to benefit themselves and cause chaos in society by creating and spreading fake news to the public. Recently, efforts have been made to develop systems for automatically detecting fake news on social media, such as using machine learning to analyze news data or using systems that learn the behavior of people on social media. From these methods, the concept of integrating multi-factor data analysis has emerged to enhance the accuracy and efficiency of fake news detection systems. This research aims to present a predictive model for detecting fake news using ensemble learning techniques with voting methods. It includes a dataset of 20,000 data points, considering factors such as media sources and user interactions. The test results, in conjunction with the fake news database recorded from social media like Fakeddit.com, show that the proposed model has a prediction accuracy of 96.97%, which is higher than other learning techniques used for comparison.

Keywords: Predictive model, fake news detection, machine learning, ensemble learning

[1-2] สาขาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยธรรมศาสตร์ ปทุมธานี 12120

[1-2] Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University, Pathumthani, 12120

Corresponding Email: tkaewkiriya@gmail.com

บทนำ

ด้วยการเติบโตทางเทคโนโลยีอินเทอร์เน็ตที่มีความก้าวหน้าอย่างรวดเร็ว ทำให้เกิดสื่อสังคมออนไลน์ (Social Media) จำนวนมากที่ประชาชนทั่วไปสามารถเข้าถึงได้ไม่ยาก การติดตามและแบ่งปันข่าวสารกับคนใกล้ชิดหรือแม้แต่สู่สาธารณะกลายเป็นหนึ่งในกิจวัตรประจำวันของบุคคลทั่วไป และเมื่อมีข่าวสารเกิดขึ้น ย่อมมีบุคคลไม่หวังดีบางกลุ่มที่พยายามสร้างและเผยแพร่ข่าวลวง (Fake News) สู่สื่อสังคมออนไลน์ โดยมีจุดประสงค์ที่ไม่ดี เช่น ความรู้เท่าไม่ถึงการณ์ของบุคคลบางกลุ่มในการเผยแพร่เนื้อหาหลอกลวงเพื่อกลั่นแกล้งผู้คน การมุ่งสร้างผลประโยชน์โดยกลุ่มมิจฉาชีพโดยแอบอ้างข่าวลวง การก่อความวุ่นวายของกลุ่มหัวรุนแรงทางการเมือง เป็นต้น ซึ่งข่าวลวงเหล่านี้อาจก่อความเสียหายต่อบุคคลและสังคมได้ ทั้งการทำให้เสื่อมเสียชื่อเสียง ไปจนถึงความขัดแย้งภายภายในสังคม เป็นต้น

ในช่วงหลายปีที่ผ่านมานี้ บริษัท องค์กร หน่วยงานภาครัฐ และนักวิจัยจำนวนมาก ได้มีความพยายามที่จะกำจัดและยับยั้งข่าวลวงเหล่านี้ด้วยวิธีการที่หลากหลาย หนึ่งในวิธีการแรกเริ่มที่มีง่ายที่สุด นั่นคือการใช้บุคคลากรในการเฝ้าสังเกตเนื้อหาข่าวที่ได้รับรายงานมาจากผู้ใช้ในระบบ โดยเมื่อทางบุคคลากรได้รับรายงาน ก็จะทำการตรวจสอบข้อเท็จจริง โดยหากพบว่าเนื้อหาข่าวนั้นมีการนำเสนอข้อมูลเท็จ ก็จะทำการนำเนื้อหาข่าวนั้น ๆ ออกจากระบบไป อย่างไรก็ตามวิธีการดังกล่าวนั้นจำเป็นต้องใช้บุคลากรในการเฝ้าสังเกตข่าวในระบบเป็นจำนวนมากและใช้เวลาในการพิจารณาที่ยาวนาน เนื่องจากสื่อสังคมออนไลน์มีผู้ใช้งานจำนวนมหาศาล และจำนวนข่าวเผยแพร่ในแต่ละวันนั้นก็มีปริมาณสูงเช่นกัน อีกทั้งอาจมีบางกรณีที่เนื้อหาข่าวดังกล่าวเป็นข่าวปลอม แต่ผู้ใช้งานเข้าใจผิดว่าเป็นข่าวจริง จึงไม่ทำการรายงานข่าวนั้น ด้วยเหตุนี้จึงมีความพยายามให้วิธีการอื่น ๆ ที่พึงพาแรงงานมนุษย์น้อยลง และใช้เวลาในการพิจารณาที่ลดลงเช่นกัน

อย่างไรก็ดี เมื่อเวลาผ่านไป ผู้ไม่หวังดีได้มีการปรับปรุงรูปแบบการนำเสนอข่าวปลอม ทำให้เนื้อหาข่าวหรือพาดหัวข่าวมีความน่าเชื่อถือมากยิ่งขึ้น และด้วยเหตุนี้การพิจารณาข้อมูลผ่านทางพาดหัวข่าวและเนื้อหาข่าวอาจจะยังไม่เพียงพอ ต่อมามีนักวิจัยทำการเพิ่มข้อมูลประกอบการพิจารณาและคัดกรองข่าวลวง โดยเพิ่มการวิเคราะห์รูปภาพประกอบข่าว เพื่อดูความเชื่อมโยงของเนื้อหาข่าว และความน่าเชื่อถือของข่าว ซึ่งผลการทดลองชี้ให้เห็นว่าการเพิ่มคุณลักษณะประกอบการตัดสินใจ ช่วยเหลือการคัดกรองข่าวลวงได้เป็นอย่างดี

จากประเด็นที่กล่าวมานี้ทำให้เห็นโอกาสในการพัฒนาและปรับปรุงวิธีการคัดกรองข่าวลวง โดยมีข้อสันนิษฐานว่าหากมีการเพิ่มคุณลักษณะของข่าว โดยพิจารณาข้อมูลการมีส่วนร่วมของผู้ใช้งาน เช่น การกดไลก์ การแสดงความคิดเห็นเข้าร่วมในการตัดสินใจ จะช่วยในการคัดกรองเพิ่มเติมได้ว่าข่าวดังกล่าวนั้นเป็นข่าวลวงหรือไม่

ในงานวิจัยฉบับนี้จึงได้นำเสนอแบบจำลองรูปแบบในการตรวจจับและคัดแยกข่าวลวงบนสื่อสังคมออนไลน์อย่างมีประสิทธิภาพ โดยแบบจำลองนี้ใช้เทคนิคการเรียนรู้ของเครื่องจักร (Machine Learning)เพื่อพิจารณาข้อมูลปะกอบไปด้วย 2 ส่วนหลักๆ ได้แก่ ข้อมูลรูปภาพข่าว และข้อมูลการมีส่วนร่วมของผู้ใช้งาน โดยใช้ฐานข้อมูลข่าวลวงของสื่อสังคมออนไลน์เรดดิต (Fakeddit) กว่า 20,000 ชุดข้อมูล

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. ข่าวลวงและผลกระทบที่อาจเกิดขึ้น

ข่าวลวง หมายถึง ข่าวที่มีความพยายามปลอมแปลงให้มีความน่าเชื่อถือและมีความเป็นผู้ชำนาญการ เพื่อให้ผู้คนทั่วไปไม่สามารถแยกออกได้ว่าข่าวดังกล่าวเป็นข่าวลวงหรือไม่ (Mustafaraj & Metaxas, 2017) โดยจากการศึกษาพบว่าข่าวลวงนั้นเกิดมาตั้งแต่สมัยโรมันโบราณ โดยในยุคนั้นมีจุดมุ่งหมายของการเผยแพร่ข่าวลวงเพื่อการสงคราม (Shu *et al*., 2017) และด้วยการเติบโตของสื่อต่าง ๆ ในศตวรรษที่ 20 เช่น โทรทัศน์ หนังสือพิมพ์ และสื่อประเภทหนึ่งสู่จำนวนมาก (one-to-many) ทำการมีการเติบโตของข่าวลวงเพิ่มสูงขึ้น (Allcott & Gentzkow, 2017) และในศตวรรษที่ 21 การมาถึงของเทคโนโลยีอินเทอร์เน็ต ทำให้เกิดสื่อสังคมออนไลน์ (social media) เช่น Facebook, Twitter, Reddit เป็นต้น ทำให้การเผยแพร่ข่าวลวงเกิดขึ้นหลากหลายรูปแบบ และมีการแพร่กระจายของข่าวลวงเกิดขึ้นเป็นทวีคูณ

ตัวอย่างกรณีที่การเติบโตของการเผยแพร่ข่าวลวงที่น่าสนใจ นั่นคือ ช่วงการเลือกตั้งของสหรัฐอเมริกาในปี ค.ศ. 2016 นั้น พบว่ามีการเติบโตของข่าวลวงบนสื่อต่าง ๆ อย่างเห็นได้ชัด โดยกระทำไปเพื่อผลประโยชน์ในการเลือกตั้งและการโจมตีทางการเมือง (Gimpel *et al*., 2021)

ผลกระทบของข่าวลวงนั้นสามารถสร้างผลกระทบได้ตั้งแต่ความเกิดเข้าใจผิดในระดับบุคคล เช่น การเสียชีวิตของบุคคลที่มีชื่อเสียง (Paschen, 2020) ไปจนถึงความสูญเสียและความตื่นตระหนกเป็นวงกว้างในสังคม เช่น การเผยแพร่ข่าวลวงเกี่ยวข้องที่เกี่ยวข้องกับโรคระบาด (Al Asaad & Erascu, 2018) เป็นต้น

2. สื่อสังคมออนไลน์ Reddit.com

เพื่อพิสูจน์สมมุติฐานดังกล่าว จึงจำเป็นจำต้องมีกรณีตัวอย่างที่น่าเชื่อถือและเป็นกรณีที่เคยเกิดขึ้นในโลกความเป็นจริง เพื่อเป็นการจำลองเหตุการณ์ที่อาจเกิดขึ้นเมื่อนำโครงประกอบไปประยุกต์ใช้ โดยในกรณีนี้ทางผู้วิจัยได้จำกัดขอบเขตการศึกษาที่เว็บไซต์ Reddit.com

สำหรับ Reddit.com นั้น เป็นสื่อสังคมออนไลน์ที่มีชื่อเสียงของสหรัฐอเมริกา โดยมีการรายงานว่ามีผู้ใช้งานรวมมากกว่า 52 ล้านคนต่อวัน และยังมีกระทู้ย่อย (Subreddit) สำหรับให้ผู้ใช้งานเข้าพูดคุยรวมกันมากกว่า 138,000 กระทู้ ซึ่งเว็บไซต์นี้เคยมีส่วนสำคัญในประเด็นอันเป็นที่ถกเถียงมากมาย เช่น การค้นหามือระเบิดจากเหตุระเบิดเมืองบอสตัน รวมไปถึงการพูดคุยในประเด็นการเหยียดเชื้อชาติและการเหยียดหยามเพศ (Proferes, 2021) อีกนัย Reddit.com เป็นไซต์ข่าวสังคม ซึ่งเนื้อหาและการกลั่นกรองเป็นลักษณะของการทำงานแบบมีส่วนร่วมโดยผู้ใช้ ตรงข้ามกับไซต์อื่น ๆ ที่เนื้อหาได้รับการจัดการโดยทีมบรรณาธิการหรืออัลกอริทึม ในทางปฏิบัติ ผู้ใช้ Reddit เลือกที่จะส่งลิงก์ไปยังไซต์อื่นหรือแบ่งปันมุมมองของตนเองในโพสต์ Redditors รายอื่นที่สมัครเป็นสมาชิกดังกล่าวจะพิจารณาแยกกันว่าโพสต์นั้นต้องการการโหวตขึ้นหรือลง ซึ่งจะถูกรวมเข้าด้วยกันเพื่อให้คะแนนของโพสต์นั้น โดย The Reddiquette ซึ่งเป็นชุดกฎที่เป็นลายลักษณ์อักษรและค่านิยมที่ Redditors สนับสนุนในการปฏิบัติตาม โดยเน้นย้ำถึงความจำเป็นในการลงคะแนนโดยพิจารณาว่าโพสต์หรือความคิดเห็นมีส่วนในการสนทนาหรือไม่ การโหวตขึ้นและลงเหล่านี้ทำหน้าที่เป็นการวิเคราะห์พฤติกรรมเชิงตัวเลขของการยอมรับทางสังคม (Yu el al., 2022) อัลกอริทึมจะจัดเรียงและจัดลำดับโพสต์และเนื้อหาอื่น ๆ ตามคะแนนที่ได้รับจากการโหวตและเวลาตั้งแต่เผยแพร่ ซึ่งแตกต่างจากการกดไลค์ทั่วไปบน Facebook หรือ Twitter ภายใต้แต่ละโพสต์และลิงก์ที่แบ่งปัน Redditors ได้รับการสนับสนุนให้พูดคุยผ่านความคิดเห็นที่ได้รับการประเมินและกลั่นกรองด้วยการโหวตขึ้นและลงและเรียงลำดับตามนั้น (Duguay, 2022)

จากเหตุผลการใช้งานปริมาณมหาศาลที่เกิดขึ้น และความเสี่ยงต่อการแพร่กระจายข่าวที่ไม่ถูกต้องนี้ จึงนำไปสู่ความพยายามในการนำโครงประกอบที่นำเสนอมาทดสอบกับข้อมูลที่ได้รับจากเว็บไซต์นี้ เพื่อพิสูจน์ประสิทธิภาพของโครงประกอบที่จะนำเสนอในบทถัดไป

3. การเรียบรู้แบบรวมกลุ่ม

เทคนิคการเรียนรู้แบบกลุ่มเป็นเทคนิคของการเรียนรู้ของเครื่องการเข้นสูงที่ช่วยเพิ่มประสิทธิภาพการพยากรณ์ผล โดยการใช้เทคนิคที่ใช้แบบจำลองประเภทการจำแนกข้องมูลหรือเทคนิคการเรียนรู้แบบมีผู้สอนหลาย ๆ แบบจำลอง มาช่วยในการหาคำตอบ ซึ่งเป็นเทคนิคที่มีประสิทธิภาพสูง (Dietterich, 2000) โดยอัลกอริทึมที่พัฒนาบนแนวคิดเทคนิคแบบรวมกลุ่มที่นำมาใช้ในงานวิจัยครั้งนี้คือเทคนิคแบบโหวต (Vote Ensemble) เป็นการใช้ชุดข้อมูลการเรียนรู้ชุดเดียวกันแต่สร้างแบบจำลองด้วยเทคนิคหลากหลายต่าง ๆ โดยหลังจากได้แบบจำลองมาชุดหนึ่งแล้วจะทำการนำไปพยากรณ์ข้อมูลและนำคำตอบมารวมกันเพื่อดูว่าคำตอบไหนเหมาะสมที่สุดด้วยใช้วิธีการโหวตและทำการเลือกคำตอบที่ตอบตรงกันมากที่สุด แสดงตาม Figure 1

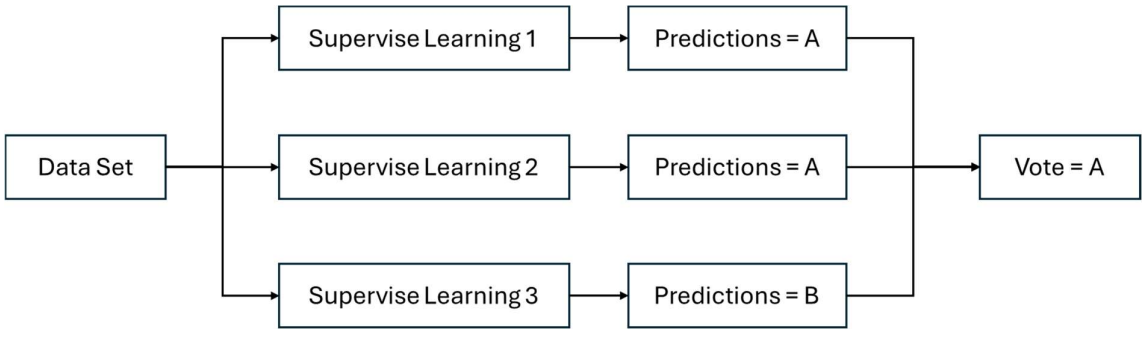

Figure 1 Ensemble learning techniques with voting

4. การแบ่งชุดข้อมูลและประเมินประสิทธิภาพตัวแบบพยากรณ์

ภายในงานวิจัยนี้ได้ใช้วิธีการแบ่งข้อมูลที่เรียกว่า Cross-validation โดยเป็นเป็นวิธีที่นิยมในการทำงานวิจัยเพื่อใช้ในการทดสอบประสิทธิภาพของโมเดลเนื่องจากผลที่ได้มีความน่าเชื่อถือ เป็นการแบ่งข้อมูลออกเป็นหลายส่วน มักจะแสดงด้วยค่า k เช่น 5-fold cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน และมีข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล โดยทำวนจนครบจำนวนที่แบ่งไว้

การวัดประสิทธิภาพโมเดลประเภทการจำแนกข้อมูล (Classification) จะมุ่งไปที่การหาค่าความถูกต้องโดยจะต้องทำความใจ โดย

1. True Positive (TP) คือสิ่งที่โปรแกรมทำนายว่าเป็นข่าวจริงและมีค่าเป็นจริง

2. True Negative (TN) คือสิ่งที่โปรแกรมทำนายว่าเป็นข่าวปลอมและมีค่าเป็นข่าวปลอมจริง

3. False Positive (FP) คือสิ่งที่โปรแกรมทำนายว่าเป็นข่าวจริงแต่มีค่าเป็นข่าวปลอม

4. False Negative (FN) คือสิ่งที่โปรแกรมทำนายว่าเป็นข่าวปลอมแต่มีค่าเป็นข่าวจริง

ซึ่งค่าดังกล่าวจะถูกนำไปแทนในสูตรเพื่อหาค่าวัดผลต่างๆ ได้ โดยจะได้มาจากตาราง Confusion Matrix ที่ให้ผลลัพธ์หลักๆ อยู่ 3 ค่า คือ Precision เป็นการวัดความถูกต้องของข้อมูล โดยพิจารณาแยกทีละคลาส Recall เป็นการวัดความถูกต้องของโมเดลโดยพิจารณาแยกทีละคลาส และ Accuracy เป็นการวัดความถูกต้องของตัวแบบ และเป็นค่าที่ใช้นำเสนอภายในงานวิจัย โดยจะพิจารณารวมทุกคลาส ด้วยสมการ (1)

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

5. งานวิจัยการคัดแยกข่าว

จากการศึกษาข่าวปลอมและผลกระทบที่อาจเกิดขึ้น จะเห็นได้อย่างชัดเจนถึงผลกระทบร้ายแรงที่อาจเกิดขึ้นเป็นวงกว้าง หากมีข่าวปลอมเกิดขึ้นในสื่อต่างๆ และมีการเผยแพร่โดยไม่มีการหาช่องทางตรวจจับหรือระงับไว้ ด้วยเหตุนี้จึงมีนักวิจัยจำนวนมากได้ทำการวิจัย และคิดค้นเทคโนโลยีในการตรวจจับข่าวปลอมที่อาจเกิดขึ้นเป็นสื่อสังคมออนไลน์ เพื่อให้ระบบคัดแยกข่าวปลอมมีประสิทธิภาพ จำเป็นจะต้องมีการใช้ข้อมูลที่มีความสำคัญและเป็นประโยชน์ในการตัดสินใจของระบบ หนึ่งในแนวทางที่นักวิจัยจำนวนมากมุ่งเน้น นั่นคือการพิจารณาจากเนื้อหาของข่าวที่ปรากฏ เช่น หัวข้อ บทบรรยาย เป็นต้น โดยงานวิจัยลักษณะดังกล่าวนั้นมีนักวิจัยจำนวนมากได้นำเสนอแนวทางการพิจารณา โดยมีความแตกต่างคือ การใช้เทคนิคการเรียนรู้ของเครื่องจักรที่แตกต่างกัน หรือการใช้เทคนิคการประมวลผลภาษา (Przybyla, 2020)

หนึ่งในตัวอย่างงานวิจัยที่น่าสนใจ คือ การใช้เทคนิคการเรียนรู้ของเครื่องจักรจำพวกนาอีฟ เบย์ (Naïve Bayes) และเอสวีเอ็ม (SVM) ร่วมกับเทคนิคการประมวลผลภาษา ซึ่งจากการทดลองพบว่าการใช้เทคนิคร่วมกันดังกล่าวสามารถเพิ่มประสิทธิภาพการคัดแยกข่าวปลอมได้ โดยมีค่าความถูกต้องสูงถึง 83.50% (Jain *et al*., 2019) จะเห็นได้ว่าการพิจารณาข้อมูลจากเนื้อหาของข่าว สามารถช่วยระบบในการตรวจจับข่าวปลอมได้มีประสิทธิภาพในระดับหนึ่ง แต่อย่างไรก็ดี ยังมีนักวิจัยอีกส่วนหนึ่งที่เล็งเห็นโอกาสในการตรวจจับข่าวปลอมโดยอาศัยปัจจัยอื่นๆ ที่มิใช่จากเนื้อหาของข่าว ประกอบกับการเติบโตของเทคโนโลยีด้านสื่อสังคมออนไลน์ จึงเกิดเป็นอีกหนึ่งแนวทางที่เริ่มได้รับความสนใจ นั่นคือ การพิจารณาปัจจัยทางสื่อสังคมออนไลน์เพื่อการคัดแยกข่าวปลอม

ในการศึกษาเชิงสำรวจ (Tandoc, 2020) พยายามทำความเข้าใจการแพร่กระจายของข้อมูลที่บิดเบือนโดยการตรวจสอบว่าผู้ใช้โซเชียลมีเดียตอบสนองต่อข่าวปลอมอย่างไรและเพราะเหตุใด โดยรวมผลลัพธ์จากการสำรวจกับผู้ตอบแบบสอบถาม 2,501 คนเข้ากับชุดการสัมภาษณ์เชิงลึกกับผู้เข้าร่วม 20 คนจากประเทศสิงคโปร์ที่มีขนาดเล็กแต่มีเศรษฐกิจและเทคโนโลยี พบว่าผู้ใช้โซเชียลมีเดียส่วนใหญ่ในสิงคโปร์เพิกเฉยต่อโพสต์ข่าวปลอมที่พบบนโซเชียลมีเดีย พวกเขาจะเสนอการแก้ไขเฉพาะเมื่อปัญหานั้นเกี่ยวข้องกับผู้ตอบแบบสอบถามอย่างมากและกับคนที่ผู้ตอบแบบสอบถามมีความสัมพันธ์ระหว่างบุคคลหรือมีความใกล้ชิด ซึ่งจากการเกิดขึ้นของสื่อสังคมออนไลน์ มักได้พบเห็นเสียงสะท้อนท่ามกลางการแพร่กระจายของข่าวปลอม ซึ่งทำให้ผู้คนลังเลที่จะมีส่วนร่วมในการแชร์ข่าวจริงเพราะกลัวว่าข้อมูลดังกล่าวจะเป็นเท็จ ทำให้มีความจำเป็นที่จะต้องตรวจพบและนำเนื้อหาปลอมออกจากโซเชียลมีเดีย โดยจากการศึกษานี้สำรวจวิธีการต่าง ๆ อย่าง Natural Language Processing หรือ Hybrid model ระบุว่าการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องแบบผสมผสานและความพยายามร่วมกันของมนุษย์อาจมีโอกาสสูงในการต่อสู้กับข้อมูลที่ผิดบนโซเชียลมีเดีย (Collins, 2021) จากการค้นคว้า พบว่ามีนักวิจัยส่วนหนึ่งที่ได้ทำการศึกษาและทดลองเกี่ยวกับการพัฒนาระบบเพื่อการตรวจจับข่าวปลอมผ่านความน่าเชื่อถือของผู้ใช้งานระบบ กล่าวคือ หากผู้ใช้ส่วนใดมีแนวโน้มจะเผยแพร่ข่าวปลอม ก็เป็นไปได้ที่จะสรุปว่าข่าวต่อไปที่ผู้ใช้คนดังกล่าวจะเผยแพร่ก็อาจเป็นข่าวปลอมเช่นเดียวกัน (Pizarro, 2019) นอกจากนี้ยังมีการพิจารณาปัจจัยความสัมพันธ์ของแหล่งที่มาข่าวและความการมีปฏิสัมพันธ์ของผู้ใช้ระบบ (Shu, 2019) ซึ่งผลการทดลองชี้ให้เห็นว่าการพิจารณาปัจจัยดังกล่าวเพิ่มประสิทธิภาพในการตรวจจับข่าวปลอมได้เป็นอย่างดี

เพิ่มเติมจากงานวิจัยรูปแบบการตรวจจับข่าวปลอมบนโซเชียลมีเดียด้วยการใช้ประโยชน์การวิเคราะห์ความ

รู้สึกของเนื้อหาข่าวและการวิเคราะห์อารมณ์ของความคิดเห็นของผู้ใช้ (Hamed *et al*., 2023) ให้ความสำคัญกับข่าวปลอมที่มักมีความรู้สึกเชิงลบ โดยแยกคุณลักษณะต่างๆ จากการวิเคราะห์ความรู้สึกของบทความข่าวและการวิเคราะห์อารมณ์ของความคิดเห็นของผู้ใช้ จากชุดข้อมูล Fakeddit จำนวน 22,788 รายการมาตรฐานที่มีชื่อข่าวและความคิดเห็นที่โพสต์เกี่ยวกับข้อมูลเหล่านั้นเพื่อฝึกอบรมและทดสอบโมเดลด้วยเทคนิค Long Short-Term Memory เป็นโครงข่ายประสาทเทียมแบบหนึ่ง ให้ความถูกต้องในการตรวจจับสูงที่ 96.77% เช่นเดียวกันกับงานวิจัยการประเมินความน่าเชื่อถือของแหล่งข่าว: กรณีศึกษาของ Reddit (Amini, 2024) ได้นำเสนอ CREDiBERT การประเมินความน่าเชื่อถือโดยใช้การนำเสนอตัวเข้ารหัสแบบสองทิศทางจาก Transformers ที่เน้นวาทกรรมทางการเมืองเป็นการสนับสนุนหลัก ด้วยการเข้ารหัสเนื้อหาที่ส่งโดยใช้ CREDiBERT และรวมเข้ากับเครือข่ายประสาทเทียม โดยได้รับคะแนน F1 เพิ่มขึ้น 9% เมื่อเทียบกับวิธีการที่มีอยู่

นอกจากนี้จากงานวิจัยที่มีการใช้เทคนิคในการตรวจจับข้อมูลข่าวปลอมบนโซเชียลมีเดีย (Reshmi, 2022) โดยใช้เทคนิคแบบรวมกลุ่ม ที่มีการจำแนกประเภทการเรียนรู้ของเครื่อง 3 แบบได้แก่ Decision Tree, Naïve Bays และ K-NN จากฐานข้อมูล Politifact.com โดยไม่ระบุจำนวนข้อมูลที่ใช้ มีค่าความถูกต้องเฉลี่ยเท่ากับ 98.54% และนอกจากนี้ยังมีงานวิจัยที่ใช้เทคนิค (Kaushik ,2024) Batch Genetic Algorithm และแบบจำลองการเรียนรู้เชิงลึกแบบรวมกลุ่มบนชุดข้อมูล Gossipcop จำนวน 22,152 ชุดข้อมูลข่าว ด้วยอัลกอริทึม LSTM, BiLSTM, CNNLSTM และ CNNBiLSTM ซึ่งให้ความแม่นยำที่ 85.99% และอีกงานวิจัยเสนอวิธีการที่ใช้การจำแนก โดยคำนึงถึงการประมวลผลภาษาธรรมชาติและแนวทางการเรียนรู้ของเครื่องสำหรับการประมวลผลภาษาธรรมชาติ ฟีเจอร์ TFIDF และ Word2vec และปรับให้เหมาะสมโดยใช้วิธี Wrapper ผ่านอัลกอริทึม RFE บนชุดข้อมูลจาก Kaggle กว่า 45,676 ชุดข้อมูล โดยมีค่าความถูกต้องเฉลี่ยเท่ากับ 98.29%

จากประเด็นนี้ ผู้วิจัยพิจารณาเห็นถึงโอกาสในการพัฒนาโครงประกอบที่พิจารณาทั้งเนื้อหาของข่าวและปัจจัยทางสื่อสังคมออนไลน์ โดยมีสมมติฐานว่าการพิจารณาปัจจัยทั้งสองพร้อมกัน จะช่วยเพิ่มประสิทธิภาพและความถูกต้องในการทำนายและคัดแยกข่าวปลอมจากแหล่งข่าวได้

การทดลอง

1. ชุดข้อมูล

ชุดข้อมูลที่ใช้ เป็นการรวบรวมข้อมูลทั้งหมดที่ได้จากแหล่งข่าวที่ได้มาจากเว็บไซต์ Reddit.com ซึ่งเก็บมาจากโพสต์ของผู้ใช้ต่างๆ โดยข้อมูลขาเข้าจะประกอบด้วยปัจจัยอันได้แก่ แหล่งของรูปภาพ จำนวนความคิดเห็น ผลคะแนนที่เกี่ยวข้อง จำนวนการติดตาม อัตราส่วนระหว่างการโหวตเห็นด้วยและผลลัพธ์ประเภทข่าวโดยประกอบไปด้วยชุดข้อมูล 20,000 ชุดข้อมูล ซึ่งเป็นข้อมูลที่เปิดให้ทดลองใช้ โดยเลือกเฉพาะผลลัพธ์ ข่าวปลอม หรือข่าวจริง เท่านั้น (2-Ways) และเลือกด้วยวิธีการสุ่มที่เน้นผลลัพธ์ข่าวปลอมร้อยละ 60 และข่าวจริงร้อยละ 40 มีรายละเอียดทางสถิตินำเสนอดังนี้

1. Domain ของสื่อส่วนใหญ่คือ i.redd.it (6346) i.imgur.com (2642) imgur.com (1071) ตามลำดับ

2. จำนวนข้อคิดเห็นเท่ากับ 0 มี 9,286 ข้อความ ข้อคิดเห็นเท่ากับ 1 มี 2,230 ข้อความ ข้อคิดเห็นเท่ากับ 2 มี 1,585 ข้อความ และข้อคิดเห็นเท่ากับ 3 มี 1,154 ข้อความ

3. หมวดหมู่จัดโดยผู้ใช้ มากที่สุดคือ psbattle_artwork จำนวน 5,906mildlyinteresting จำนวน 3,091 photoshopbattles จำนวน 2,025 และ pareidolia จำนวน 1,620 ตามลำดับ

4. ค่าคะแนน น้อยสุดคือติดลบ 52 มากสุดคือ 82,135 โดยมมีค่าเฉลี่ยเท่ากับ 409.26 คะแนน และค่าเบี่ยงเบนเท่ากับ 3,232.85 จากข้อมูลทั้งหมด 20,000 ข้อมูล

5. ค่า upvote_ratio เฉลี่ยคือ 60%

6. จำนวนข่าวปลอมทั้งหมดคือ 12,018 ข้อมูล และข่าวจริงจำนวน 7,982 ข้อมูล

2. การเตรียมข้อมูล

เนื่องจากข้อมูลในแต่ละโพสต์นั้นมีองค์ประกอบจำนวนมาก อีกทั้งยังอาจมีบางส่วนที่ไม่พร้อมต่อการนำไปดำเนินการต่อในการเรียนรู้ในขั้นตอนถัดไป ซึ่งในขั้นตอนนี้จึงเป็นการทำความสะอาดข้อมูลแต่ละส่วน เช่น ข้อมูลตัวเลขจะต้องทำการแปลงจากรูปแบบตัวอักษรเป็นตัวเลข การคัดกรองข้อมูลที่ขาดหายข้อมูลบางส่วนออกจากการพิจารณาและการเรียนรู้ การคัดกรองข้อมูลคุณภาพต่ำและข้อมูลที่ไม่มีความจำเป็นออกจากการพิจารณาและการเรียนรู้ ข้อมูลตัวอักษรจะต้องทำการนำอักขระที่ไม่มีความจำเป็นออก เช่น เครื่องหมายอีโมจิ เป็นต้น

หลังจากที่ได้ทำการทำความสะอาดข้อมูลแล้ว จากนั้นจะทำการแบ่งข้อมูลที่ผ่านขั้นตอนที่กล่าวมาออกเป็น 3 ส่วนหลักประกอบด้วย ข้อมูลรูปภาพข่าว ข้อมูลพาดหัวข่าว และข้อมูลปัจจัยทางสื่อสังคมออนไลน์ เพื่อนำไปให้หน่วยการเรียนรู้ได้ทำการเรียนรู้ถัดไป

3. สังเคราะห์ตัวแบบพยากรณ์

เป็นการสังเคราะห์ตัวแบบพยากรณ์ สำหรับนำมาเปรียบเทียบค่าความถูกต้องของตัวแบบ โดยนำเสนอกรอบการทำงานของแบบจำลองตาม Figure 2

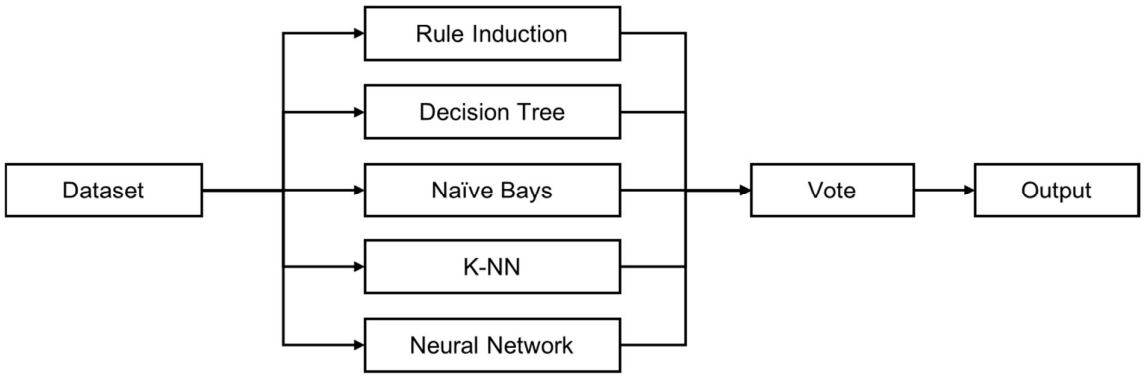

Figure 2 Various techniques are used in the voting technique

เป็นการนำชุดข้อมูลมาใช้ในการเรียนรู้โดยผู้วิจัยได้คัดเลือกผ่านอัลกอริทึมแบบมีผู้สอน ประเภทการจำแนกข้อมูล เพื่อให้สามารถสร้างตัวแบบพยากรณ์ ได้แก่ กฎการอุปนัย ต้นไม้ตัดสินใจ นาอีฟเบย์ K-NN และโครงข่ายประสาทเทียม โดยกำหนดค่าพารามิเตอร์ที่สำคัญ ได้แก่ กฎการอุปนัย กำหนดเกณฑ์สำหรับการเลือกแอตทริบิวต์และการแบ่งตัวเลขแบบ Information Gain และอัตราส่วนตัวอย่างที่ 90% ต้นไม้ตัดสินใจ กำหนดเกณฑ์สำหรับการเลือกแอตทริบิวต์และการแบ่งตัวเลขแบบ Information Gain ความลึกของต้นไม้ = 10 ความเชื่อมั่นที่ใช้ในการคำนวณข้อผิดพลาดเชิงลบของการตัดแต่งกิ่งเท่ากับ 10% และ K-NN กำหนดค่า K เท่ากับ 5 และโครงข่ายประสาทเทียม กำหนด Activation เป็น Rectifier ให้มี 2 เลเยอร์ซ่อนโดยแต่ละเลเยอร์จะมี 5 โหนด

หลังจากนั้นจะนำผลลัพธ์ที่ได้จากการเรียนรู้ทั้ง 5 อัลกอริทึมมาโหวตหาผลลัพธ์ที่มีผลมากที่สุด ในลักษณะนับผลลัพธ์การทำนายหรือเสียงข้างมาก ผ่านโปรแกรม Rapid-Miner และหาค่าความถูกต้องด้วยตัวจัดการในการทดสอบโมเดล 5-Fold Cross Validation หรือการแบ่งข้อมูลออกเป็น 5 ชุด ที่ทำการแบ่งข้อมูลเผื่อวัดประสิทธิภาพที่ 5 ชุดข้อมูล โดยภายในจะทำการฝึกโมเดลด้วยเทคนิคการรวมกลุ่มแบบโหวตที่ประกอบไปด้วยอัลกอริทึมทั้ง 5 ที่ได้กล่าวในข้างต้น จากนั้นนำโมเดลที่มาหาค่าประสิทธิภาพแล้วนำผลลัพธ์ส่งไปยัง Output ถัดไป

ผลการทดลองและอภิปรายผล

ผลการทดลองแบบกลุ่มด้วยวิธีการโหวตจาก 5 อัลกอริทึม แสดงค่าความถูกต้องของแต่ละอัลกอริทึมและค่าความถูกต้องรวมได้ตาม Table 1

Table 1 Accuracy values of various techniques

Techniques	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Rule Induction	91.39	91.43	91.41	91.40
Decision Tree	89.90	89.94	89.92	89.92
Naïve Bays	78.90	67.33	72.66	72.65
K-NN	92.55	92.28	92.41	92.70
Neural Network	90.15	90.46	90.30	90.42
Ensemble Learning	97.06	96.91	96.89	96.97

โดยค่าความถูกต้อง (Accuracy) ของเทคนิคการเรียนรู้แบบรวมกลุ่มมีค่ามากที่สุดคือ 96.97% โดยถัดมาคือเทคนิค K-NN, Rule Induction, Neural Network, Decision Tree และ Naïve Bays ตามลำดับ รวมถึงค่าอื่นๆ ที่เป็นไปในทิศทางเดียวกัน ซึ่งผลลัพธ์ของแต่ละอัลกอริทึมจะเป็นเพียงค่าประสิทธิภาพที่เรียนรู้จากชุดข้อมูลการทดลองนี้เท่านั้น

อย่างไรก็ตาม เทคนิคการเรียนรู้แบบรวมกลุ่ม เป็นการนำอัลกอริทึมหลายๆ อัลกอริทึมมาทำงานร่วมกันใน

ลักษณะการโหวต หรือก็คือเป็นการให้น้ำหนักกับผลลัพธ์ที่อัลกอริทึมนั้นๆ โดยหากพบว่าอัลกอริทึมตั้งต้น มีความถูกต้องที่ต่ำทั้งหมด จะไม่ได้ทำให้การพยากรณ์แม่นยำเท่าที่ควร

สรุปผลการทดลองและข้อเสนอแนะ

การตรวจจับข่าวปลอมที่อาจพบได้ในสื่อสังคมออนไลน์โดยแบบพยากรณ์ดังกล่าว เมื่อทำการคัดเลือกข้อมูลอันเกี่ยวกับลักษณะทางสังคมออนไลน์ที่สามารถนำมาสังเคราะห์ตัวแบบพยากรณ์ด้วยเทคนิคแบบรวมกลุ่ม พบว่ามีค่าความถูกต้องมากที่สุดเมื่อเทียบกับการใช้เทคนิคแบบมีผู้สอนที่เป็นส่วนประกอบโดยไม่นำมาร่วมกันช่วยโหวตผลลัพธ์พยากรณ์ และเมื่อเปรียบเทียบกับงานวิจัยอื่นๆ ในด้านการตรวจจับข่าวปลอม พบว่า มีค่าความถูกต้องในระดับที่มากกว่า 96% ขึ้นไปใกล้เคียงกันในปริมาณข้อมูลที่ต่างกัน

สำหรับงานในอนาคตได้มีการวางแผนที่จะปรับปรุงแบบจำลอง โดยจะเพิ่มข้อมูลสำหรับการประมวลผลในรูปแบบข้อความต่างๆ ไม่ว่าจะเป็นพาดหัวข่าว รายละเอียดข่าว รวมถึงข้อคิดเห็น การเพิ่มเติมในลักษณะดังกล่าว รวมถึงการทดลองในการนำอัลกอริทึมอื่นๆ มาใช่ร่วมในการทำ Ensemble Voting และการตั้งค่าพารามิเตอร์ของอัลกอริทึมที่หลากหลายและให้ได้มาซึ่งค่าความถูกต้องที่ดีมากยิ่งขึ้น มีสันนิษฐานว่าจะเป็นการเพิ่มประสิทธิภาพและค่าความถูกต้องของแบบจำลองได้เป็นอย่างดีซึ่งจะต้องมีการทดสอบและประเมินประสิทธิภาพต่อไป

เอกสารอ้างอิง

Al Asaad, B., & Erascu, M. (2018, September). A tool for fake news detection. In *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)* (pp. 379–386). IEEE.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 31*(2), 211–236.

Amini, A., Bayiz, Y. E., Ram, A., Marculescu, R., & Topcu, U. (2024). *News source credibility assessment: A Reddit case study*. arXiv preprint arXiv:2402.10938.

Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2021). Trends in combating fake news on social media–a survey. *Journal of Information and Telecommunication, 5*(2), 247–266.

Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems* (pp. 1–15). Springer Berlin Heidelberg.

Duguay, P. A. (2022). Read it on Reddit: Homogeneity and ideological segregation in the age of social news. *Social Science Computer Review, 40*(5), 1186–1202.

Gimpel, H., Heger, S., Olenberger, C., & Utz, L. (2021). The effectiveness of social norms in fighting fake news on social media. *Journal of Management Information Systems, 38*(1), 196–221.

Hamed, S. K., Ab Aziz, M. J., & Yaakub, M. R. (2023). Fake news detection model on social media by leveraging sentiment analysis of news content and emotion analysis of users' comments. *Sensors, 23*(4), 1748.

Jain, A., Shakya, A., Khatter, H., & Gupta, A. K. (2019, September). A smart system for fake news detection using machine learning. In *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)* (Vol. 1, pp. 1–4). IEEE.

Kaushik, D., & Nadeem, M. (2024, May). Fake news detection using evolutionary ensemble deep learning. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)* (pp. 161–166). IEEE.

Mustafaraj, E., & Metaxas, P. T. (2017, June). The fake news' spreading plague: Was it preventable? In *Proceedings of the 2017 ACM on Web Science Conference* (pp. 235–239).

Paschen, J. (2020). Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product & Brand Management, 29*(2), 223–233.

Pizarro, J. (2020, October). Profiling bots and fake news spreaders at PAN'19 and PAN'20: Bots and gender profiling 2019, profiling fake news spreaders on Twitter 2020. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 626–630). IEEE.

Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society, 7*(2), 20563051211019004.

Przybyla, P. (2020, April). Capturing the style of fake news. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 01, pp. 490–497).

Reshmi, T. S. (2022, July). Fake news detection using voting ensemble classifier. In *2022 International Conference on Inventive Computation Technologies (ICICT)* (pp. 1241–1244). IEEE.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter, 19*(1), 22–36.

Shu, K., Wang, S., & Liu, H. (2019, January). Beyond news contents: The role of social context for fake news detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* (pp. 312–320).

Tandoc Jr, E. C., Lim, D., & Ling, R. (2020). Diffusion of disinformation: How social media users respond to fake news and why. *Journalism, 21*(3), 381–398.

Yu, X., Gil-López, T., Shen, C., & Wojcieszak, M. (2022). Engagement with social media posts in experimental and naturalistic settings: How do message incongruence and incivility influence commenting? *International Journal of Communication, 16*, 24.