บทความวิจัย

การเปรียบเทียบขั้นตอนวิธีแบบเบย์อย่างง่ายและซัพพอร์ตเวกเตอร์แมชชีน เพื่อพัฒนาแบบจำลองการทำนายบทวิจารณ์ร้านอาหาร

ลดาวัลย์ แย้มครวญ จิติมนต์ อั่งสกุล และ ธรา อั่งสกุล*
สำนักวิชาศาสตร์และศิลป์ดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 0 4422 5799 อีเมล: angskun@sut.ac.th
DOI: 10.14416/j.kmutnb.2023.10.002
รับเมื่อ 5 ตุลาคม 2564 แก้ไขเมื่อ 29 พฤศจิกายน 2564 ตอบรับเมื่อ 28 ธันวาคม 2564 เผยแพร่ออนไลน์ 18 ตุลาคม 2566
© 2023 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

บทคัดย่อ

คนไทยให้ความสนใจเรื่องอาหาร เนื่องจากอาหารเป็นหนึ่งในปัจจัยสำคัญสำหรับการดำเนินชีวิตและยังเป็นส่วนหนึ่งของวัฒนธรรมไทย โดยที่ผู้บริโภคนิยมเขียนบทวิจารณ์เกี่ยวกับอาหาร เพื่อแบ่งปันประสบการณ์ของตนให้กับเพื่อนบนเครือข่ายสังคมออนไลน์ ทำให้บทวิจารณ์เกี่ยวกับอาหารและร้านอาหารมีมากมายและอยู่กันอย่างกระจัดกระจาย บทความนี้นำเสนอแบบจำลองแนะนำร้านอาหารจากบทวิจารณ์ในทวิตเตอร์ เรียกว่า ฟู้ดฟายเดอร์ โดยแบ่งระดับของบทวิจารณ์เป็น 3 ประเภท ได้แก่ บทวิจารณ์ด้านบวก เป็นกลาง และลบ แบบจำลองดังกล่าวรวบรวมบทวิจารณ์จากเครือข่ายสังคมออนไลน์ทวิตเตอร์ (เอ็กซ์) จำนวน 1,490 บทวิจารณ์ และวิเคราะห์ข้อมูลโดยใช้ขั้นตอนวิธีแบบนาอีฟเบย์ เปรียบเทียบกับขั้นตอนวิธีแบบซัพพอร์ตเวกเตอร์แมชชีน ผลที่ได้พบว่า ค่าความถูกต้องในการทำนายประเภทของบทวิจารณ์ด้วยวิธีแบบซัพพอร์ตเวกเตอร์แมชชีน ได้รับค่าความถูกต้องร้อยละ 82.89 ซึ่งมากกว่าวิธีนาอีฟเบย์ ที่ได้รับค่าความถูกต้องร้อยละ 80.20 และเมื่อทดสอบกับข้อมูลใหม่ พบว่า แบบจำลองด้วยวิธีแบบซัพพอร์ตเวกเตอร์แมชชีนได้รับค่าความถูกต้อง ร้อยละ 80.87 ส่วนวิธีนาอีฟเบย์ได้รับค่าความถูกต้อง ร้อยละ 78.86 นอกจากนั้น บทความนี้ยังได้นำเสนอการวิเคราะห์ข้อมูลโดยการนับจำนวนคำ แฮชแท็ก และสัญลักษณ์ที่ปรากฏในบทวิจารณ์โดยการนับความถี่ เพื่อเพิ่มความถูกต้องในการทำนายประเภทบทวิจารณ์

คำสำคัญ: เครือข่ายสังคมออนไลน์ ทวิตเตอร์ (เอ็กซ์) แบบจำลองแนะนำร้านอาหาร นาอีฟเบย์ ซัพพอร์ตเวกเตอร์แมชชีน

การอ้างอิงบทความ: ลดาวัลย์ แย้มครวญ, จิติมนต์ อั่งสกุล และ ธรา อั่งสกุล, "การเปรียบเทียบขั้นตอนวิธีแบบเบย์อย่างง่ายและซัพพอร์ตเวกเตอร์แมชชีน เพื่อพัฒนาแบบจำลองการทำนายบทวิจารณ์ร้านอาหาร," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 33, ฉบับที่ 4, หน้า 1–10, เลขที่บทความ 234-205509, ต.ค.–ธ.ค. 2566.

วารสารวิชาการพระจอมเกล้าพระนครเหนือ ปีที่ 33 ฉบับที่ 4 ต.ค.–ธ.ค. 2566
The Journal of KMUTNB., Vol. 33, No. 4, Oct.–Dec. 2023

Research Article

A Comparison of Naive Bayes and Support Vector Machine to Develop a Restaurant Review Prediction Model

Ladawan Yamkuan, Jitimon Angskun and Thara Angskun*
Institute of Digital Arts and Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand

* Corresponding Author, Tel. 0 4422 5799, E–mail: angskun@sut.ac.th DOI: 10.14416/j.kmutnb.2023.10.002
Received 5 October 2021; Revised 29 November 2021; Accepted 28 December 2021; Published online: 18 October 2023
© 2023 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

Abstract

Thai people are interested in food because it is one of the important factors for living. It is also a part of the Thai Culture. Consumers usually write reviews about it to share their experiences with friends on social networks. This has resulted in numerous and sparse reviews about food and restaurants. This article presents a restaurant recommendation model based on reviews in Twitter (X), called FoodFinder. The reviews are divided into three categories i.e. positive, neutral, and negative reviews. The model collected 1,490 reviews from the social network, Twitter. The data were analyzed by Naive Bayes compared with the Support Vector Machine. The results indicated that the prediction accuracy of review category with the Support Vector Machine was 82.89 percent, which was higher than that of Naive Bayes that obtained 80.20 percent. Testing with unseen data revealed that model with Support Vector Machine method obtained 80.87 percent of accuracy, while Naïve Bayes obtained 78.86 percent of accuracy. In addition, this article also presents a data analysis by counting the number of words, hashtags and symbols that are appeared in the reviews with TF-IDF to increase the accuracy of review category prediction.

Keywords: Social Network, Twitter (X), Restaurant Recommendation Model, Naive Bayes, Support Vector Machine

Please cite this article as: L. Yamkuan, J. Angskun, and T. Angskun, "A comparison of naive bayes and support vector machine to develop a restaurant review prediction model," *The Journal of KMUTNB*, vol. 33, no. 4, pp. 1–10, ID. 234-205509, Oct.–Dec. 2023 (in Thai).

1. บทนำ

ธุรกิจอาหารและเครื่องดื่มในประเทศไทยมีมูลค่าทางการตลาดสูงถึง 2.3 ล้านล้านบาทต่อปี ซึ่งคาดว่าจะยังคงเติบโตมากขึ้น ด้วยแนวคิดการให้บริการที่สอดรับกับพฤติกรรมการใช้เวลากับเครือข่ายสังคมออนไลน์ อาทิ การแจ้งข้อมูลการปรับปรุงสินค้าและบริการ ข้อมูลโปรโมชันผ่านเครือข่ายสังคมออนไลน์ เนื่องจากในปัจจุบันมีจำนวนผู้ใช้เครือข่ายสังคมออนไลน์เพิ่มมากขึ้น จึงเกิดกลุ่มผู้บริโภคที่รู้และเข้าใจเทคโนโลยี และมีพฤติกรรมการใช้สื่อสังคมออนไลน์แทบทุกประเภท [1]

ในประเทศไทยนั้น คนไทยให้ความสนใจเรื่องอาหาร ไม่ว่าจะเป็นในชีวิตประจำวันหรือในขณะท่องเที่ยว เนื่องจากอาหารเป็นหนึ่งในปัจจัยสำคัญสำหรับการดำเนินชีวิตและวัฒนธรรมไทย โดยจากการวิเคราะห์พฤติกรรมผู้บริโภคออนไลน์พบว่า ร้อยละ 58 ของคนไทยให้ความสนใจเรื่องอาหาร ร้านอาหาร และการทำอาหาร [2] โดยข้อมูลการค้นหาผ่านเครื่องมือค้นหากูเกิ้ล (Google) เกี่ยวกับการเลือกซื้อสินค้าพบว่า สินค้าประเภทอาหารอยู่ในอันดับที่ 3 [3]

จากการศึกษาปัจจัยที่ส่งผลต่อการเลือกรับประทานอาหารของผู้บริโภคพบว่า ปัจจัยที่มีอิทธิพลต่อการเลือกรับประทานอาหารของผู้บริโภค คือ การแนะนำ การเขียนบทวิจารณ์เกี่ยวกับอาหาร [4] และปัจจัยที่มีผลต่อการตัดสินใจเลือกร้านอาหาร ได้แก่ คนรู้จักหรือใกล้ชิดแนะนำร้านอาหาร ที่ตั้งร้านอาหาร และการเดินทาง ซึ่งก่อนและระหว่างมื้ออาหาร ผู้บริโภคร้อยละ 49 ได้ศึกษาข้อมูลเกี่ยวกับอาหารและร้านอาหารจากเครือข่ายสังคมออนไลน์ [5]

การค้นหาข้อมูลจากการเขียนบทวิจารณ์ต่างๆ ได้รับความนิยมเช่นเดียวกับการเขียนบทวิจารณ์ โดยจากการสำรวจการค้นหาข้อมูลเพื่ออ่านบทวิจารณ์ของผู้บริโภคใน พ.ศ. 2559 และ 2560 พบว่า ธุรกิจร้านอาหารหรือคาเฟ่ได้รับความนิยมในการค้นหาเพื่ออ่านบทวิจารณ์มากที่สุด คิดเป็นร้อยละ 60 [6] จะเห็นได้ว่าผู้บริโภคให้ความสำคัญกับการอ่านบทวิจารณ์เกี่ยวกับอาหารมาเป็นอันดับต้นๆ โดยเหตุผลในการอ่านบทวิจารณ์ร้านอาหาร คือเพื่อเปรียบเทียบหรือหาทางเลือกก่อนการตัดสินใจใช้บริการ คิดเป็นร้อยละ 54.3 ซึ่งวัตถุประสงค์ในการค้นหาเกี่ยวกับร้านอาหารผ่านเครือข่ายสังคมออนไลน์ คือ เพื่ออ่านข้อความบทวิจารณ์มากที่สุด รองลงมาคือเพื่อดูภาพบรรยากาศภายในร้านอาหาร และดูภาพอาหารตามลำดับ [4]

การที่เครือข่ายสังคมออนไลน์ได้รับความนิยมในการแบ่งปันเกี่ยวกับ ทำให้ข้อมูลเกี่ยวกับอาหาร และร้านอาหารมีจำนวนมากและกระจัดกระจาย ซึ่งยากต่อการค้นหาข้อมูลได้ตรงกับความต้องการ ในการรับประทานอาหารนอกบ้าน 1 ครั้ง ผู้บริโภคใช้เครือข่ายสังคมออนไลน์ 1–2 ประเภทในการค้นหาบทวิจารณ์ร้านอาหาร คิดเป็นร้อยละ 61.10 และใช้ระยะเวลาเพื่อค้นหาบทวิจารณ์ร้านอาหาร 1–2 ชั่วโมง คิดเป็นร้อยละ 57.20 [4]

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบงานวิจัยที่เกี่ยวข้องกับการแนะนำร้านอาหาร ซึ่งสามารถแบ่งออกเป็น 3 กลุ่ม คือ กลุ่มงานวิจัยที่แนะนำร้านอาหารด้วยวิธีการพิจารณาเนื้อหา กลุ่มงานวิจัยที่แนะนำร้านอาหารด้วยวิธีแบบพึ่งพาผู้ใช้ร่วม และกลุ่มงานวิจัยที่แนะนำร้านอาหารด้วยวิธีการแนะนำแบบผสม

กลุ่มงานวิจัยที่แนะนำร้านอาหารด้วยวิธีการพิจารณาเนื้อหา เป็นการแนะนำที่มาจากการรวบรวม การวิจัย และการกรองข้อมูล โดยเจาะจงไปที่สินค้าที่แนะนำ ซึ่งพิจารณาจากข้อมูลความชอบ ความพึงพอใจและความสนใจของผู้ใช้ที่มีต่อสินค้าเป็นอันดับแรก โดยอ้างอิงจากสิ่งที่ผู้ใช้ชอบในอดีต เป็นการแนะนำที่บ่งบอกความชอบของผู้ใช้อย่างชัดเจน [7] โดยการทำนายร้านอาหารที่ผู้ใช้ชอบหรือเลือกใช้บริการ ด้วยการจำแนกบทวิจารณ์และคะแนนที่ได้รับจากผู้ใช้ พร้อมทั้งเปรียบเทียบระหว่างวิธีการถดถอยเชิงเส้น และวิธีการถดถอยเชิงตรรกะ [8] พบว่า วิธีการถดถอยเชิงเส้นใช้เวลาในการแนะนำน้อยกว่าวิธีการถดถอยเชิงตรรกะ แต่วิธีการถดถอยเชิงตรรกะให้ค่าความถูกต้องในการแนะนำมากกว่าวิธีการถดถอยเชิงเส้น อยู่ที่ร้อยละ 52.3 ในขณะที่วิธีการถดถอยเชิงตรรกะ ให้ค่าความถูกต้องในการแนะนำอยู่ที่ร้อยละ 50.0

ส่วนการแนะนำร้านอาหารที่ตรงกับความต้องการของผู้ใช้ โดยพิจารณาจากข้อมูลการแบ่งปัน การแสดงความคิดเห็น

การเยี่ยมชมเว็บไซต์ต่างๆ ของผู้ใช้ร่วมด้วย เพื่อหาปัจจัยที่ส่งผลต่อการรับประทานอาหารของผู้ใช้ [9] ผลที่ได้พบว่าปัจจัยที่ส่งผลต่อการรับประทานอาหารของผู้ใช้คือ ปัจจัยด้านอารมณ์ และปัจจัยด้านส่วนบุคคล และยังมีการใช้เทคนิคการทำเหมืองความคิดเห็นในการจำแนกบทวิจารณ์ที่ผู้ใช้เผยแพร่ผ่านเครือข่ายสังคมออนไลน์โดยจำแนกประเภทความคิดเห็นเป็นบวก กลาง และลบ [10] นอกจากนี้ ยังมีการวิเคราะห์ความรู้สึกของผู้ใช้ที่มีต่อร้านอาหารจากการให้คะแนนบนเว็บไซต์ต่างๆ โดยใช้ข้อความบทวิจารณ์เพื่อวิเคราะห์ความเชื่อมั่นของคำในพจนานุกรมเกี่ยวกับอาหารและร้านอาหารในภาษาพม่า [11]

กลุ่มงานวิจัยที่แนะนำร้านอาหารด้วยวิธีแบบพึ่งพาผู้ใช้ร่วม เป็นการใช้ข้อมูลของผู้ที่มีความชอบ หรือรสนิยมใกล้เคียงกับผู้ใช้เป้าหมาย อาจกล่าวได้ว่าเป็นการใช้ความสัมพันธ์ระหว่างคนกับคน [12] กลุ่มงานวิจัยนี้พิจารณาข้อมูลของผู้ใช้จากเครือข่ายสังคมออนไลน์แล้วนำมาวิเคราะห์ความรู้สึกด้านบวกและลบของผู้ใช้โดยได้พัฒนาระบบแนะนำร้านอาหารจากบทวิจารณ์และการให้คะแนนจากผู้ใช้ ด้วยวิธีแบบพึ่งพาผู้ใช้ร่วม ซึ่งจำแนกประเภทของบทวิจารณ์โดยใช้วิธีการจำแนกแบบเบย์อย่างง่าย และวิเคราะห์ความรู้สึกด้านบวกและลบจากบทวิจารณ์และคะแนนที่ได้รับจากผู้ใช้ [13]

ส่วนการวิเคราะห์ความรู้สึกของผู้ใช้ที่มีต่อร้านอาหารจากบทวิจารณ์ออนไลน์จากการให้คะแนนจากผู้ใช้ โดยแบ่งระดับคะแนนเป็น 5 ระดับ [14] ผลการวิเคราะห์พบว่า บทวิจารณ์ที่มีคะแนนมากกว่า 3 ถูกทำนายว่าเป็นบทวิจารณ์ด้านบวก ส่วนบทวิจารณ์ที่มีคะแนนตั้งแต่ 3 ลงไป ถูกทำนายว่าเป็นบทวิจารณ์ด้านลบ ส่วนการพัฒนาระบบแนะนำร้านอาหารเพื่อจำแนกประเภทร้านอาหาร จากบทวิจารณ์อาหารบนเว็บไซต์ ด้วยวิธีการจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนเพื่อวิเคราะห์ความรู้สึกด้านบวกและลบของผู้ใช้ [15] พบว่ามีความถูกต้องในการจำแนกความรู้สึกด้านบวกและด้านลบร้อยละ 88.90 และมีความถูกต้องในการจำแนกร้านอาหารร้อยละ 80 ทั้งนี้ การพิจารณาคุณลักษณะของร้านอาหาร 43 คุณลักษณะ เช่น ประเภทร้านอาหาร ที่ตั้งร้านอาหาร ข้อมูลทางการเงิน วันเปิดร้านอาหาร เป็นต้น เพื่อทำนายรายได้ของร้านอาหาร [16] ผลที่ได้พบว่า วิธีการจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนมีความถูกต้องในการทำนายมากกว่าวิธีแบบเบย์อย่างง่าย

กลุ่มงานวิจัยที่แนะนำร้านอาหารด้วยวิธีการแนะนำแบบผสมนั้น เป็นกลุ่มงานวิจัยซึ่งพิจารณาจากบทวิจารณ์และภาพ โดยการพัฒนาระบบแนะนำร้านอาหารจากบทวิจารณ์และภาพ [17] โดยพิจารณาจากความชอบของผู้ใช้ จำแนกประเภทของภาพที่ผู้ใช้อัปโหลดลงบนเว็บไซต์ โดยแบ่งออกเป็น 4 ประเภท ได้แก่ ภาพอาหาร ภาพเครื่องดื่ม ภาพร้านอาหารกลางแจ้ง และภาพร้านอาหารในร่ม ด้วยวิธีการจำแนกแบบโครงข่ายประสาทเทียมแบบสังวัตนาการ และจำแนกบทวิจารณ์ประเภทข้อความด้วยวิธีการจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีน ผลที่ได้พบว่า ความถูกต้องในการแนะนำร้านอาหารมีความถูกต้องร้อยละ 70.78

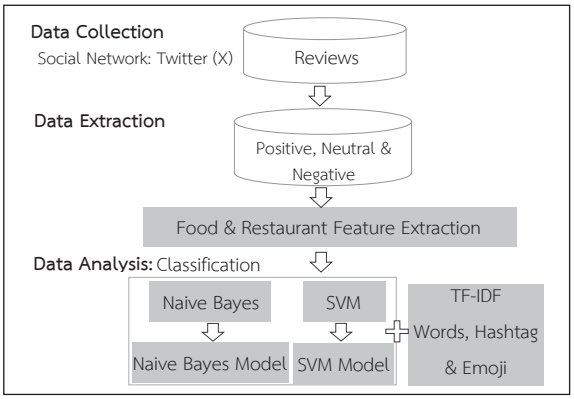
นอกจากนี้ ในการวิเคราะห์บทวิจารณ์ร้านอาหารด้านกายภาพ คุณภาพอาหาร คุณภาพการให้บริการ และราคาอาหาร ด้วยวิธี Hybrid ELMoWikipedia; HEW และ Hybrid Expanded Opinion Lexicon-SentiCircle; HEOLS [18] ผลการวิเคราะห์ที่ได้ พบว่า วิธี HEW และ HEOLS ให้ค่าประสิทธิภาพโดยรวมเพิ่มขึ้นเฉลี่ยร้อยละ 6 เมื่อเปรียบเทียบกับวิธีอื่นที่คล้ายคลึงกัน

งานวิจัยนี้ใช้ข้อมูลจากเครือข่ายสังคมออนไลน์ทวิตเตอร์ มีชื่อเรียกว่า ฟู้ดฟายเดอร์ สำหรับสร้างแบบจำลองการทำนายบทวิจารณ์ร้านอาหารด้านบวก เป็นกลาง และด้านลบ เพื่อเปรียบเทียบผลการจำแนกบทวิจารณ์ระหว่างขั้นตอนวิธีแบบเบย์อย่างง่าย และซัพพอร์ตเวกเตอร์แมชชีน นอกจากนี้ยังใช้ข้อมูลอื่นๆ ที่ปรากฏอยู่ในบทวิจารณ์ ได้แก่ คำ แฮชแท็ก และสัญลักษณ์ต่างๆ เพื่อใช้ในการวิเคราะห์ความถี่ (TF-IDF) ที่ปรากฏในบทวิจารณ์

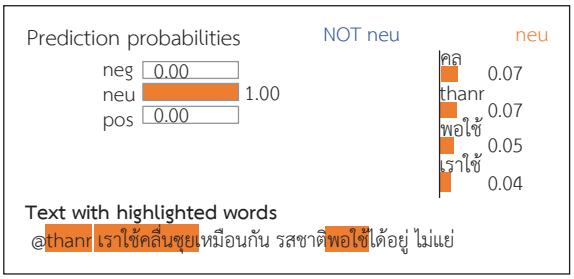
2. วัสดุ อุปกรณ์และวิธีการวิจัย

2.1 สถาปัตยกรรมของแบบจำลอง

ในการสร้างแบบจำลองการทำนายบทวิจารณ์ด้านบวก เป็นกลาง และด้านลบของบทวิจารณ์ งานวิจัยนี้มีสถาปัตยกรรมของแบบจำลองดังรูปที่ 1



รูปที่ 1 สถาปัตยกรรมของฟู้ดฟายเดอร์



รูปที่ 2 ตัวอย่างการตัดคำของบทวิจารณ์

2.1.1 Data Collection

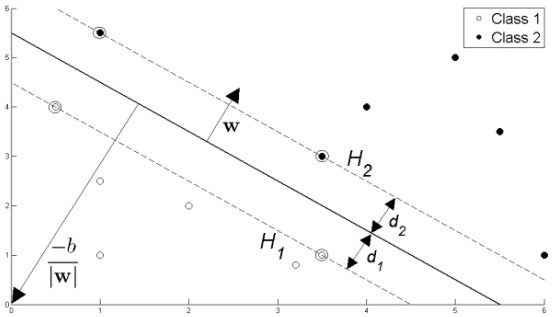
งานวิจัยนี้รวบรวมบทวิจารณ์จากทวิตเตอร์ (เอ็กซ์) เป็นข้อมูลที่รวบรวมในเดือนสิงหาคม–ตุลาคม พ.ศ. 2562 โดยรวบรวมข้อมูลแบบออนไลน์ แล้วนำมาวิเคราะห์ข้อมูลแบบออฟไลน์ และพัฒนาวิธีการรวบรวมบทวิจารณ์ ด้วยภาษาไพธอน

2.1.2 Data Extraction

การสกัดข้อมูลโดยจำแนกคุณลักษณะ (Attributes) ต่างๆ ของบทวิจารณ์ ซึ่งจำแนกบทวิจารณ์ออกเป็น 3 ประเภท จำแนกคำที่เกี่ยวข้องกับความคิดเห็นด้านบวก เป็นกลาง และลบ และกำหนดคุณลักษณะเป้าหมาย (Label Attribute) โดยผู้เชี่ยวชาญจำนวน 3 ท่าน แล้วนำไปวิเคราะห์โดยตัวอย่างการตัดคำ แสดงดังรูปที่ 2

2.1.3 Data Analysis

ในการวิเคราะห์ข้อมูลใช้จำแนกข้อมูล (Classification) ด้วยขั้นตอนวิธี Naive Bayes ซึ่งเป็นขั้นตอนวิธีที่จำแนกข้อมูลตามหลักความน่าจะเป็น เปรียบเทียบกับขั้นตอนวิธี



รูปที่ 3 การจำแนก SVM บนข้อมูลขนาด 2 มิติ

Support Vector Machine; SVM ซึ่งเป็นขั้นตอนวิธีที่มีการจำแนกข้อมูลในรูปแบบเชิงเส้น โดยมีจุดมุ่งหมายในการเปรียบเทียบขั้นตอนวิธีทั้งสองที่มีคุณลักษณะต่างกัน เพื่อพิจารณาผลลัพธ์และศึกษาแนวโน้มที่จะนำไปใช้ในการวิจัยในอนาคตต่อไป โดยแต่ละขั้นตอนวิธีมีรายละเอียดดังนี้

การจำแนกข้อมูลด้วย Naive Bayes เป็นขั้นตอนวิธีที่ทำนายสิ่งที่ไม่เคยเกิดขึ้นจากสิ่งที่เคยเกิดขึ้นมาก่อน [19] โดยมีวิธีการคำนวณหาความน่าจะเป็นดังสมการที่ (1)

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

การจำแนกข้อมูลด้วย SVM เป็นการจำแนกที่มีประสิทธิภาพในการจำแนกข้อมูลที่มีมิติมาก โดยการหาเส้นตรงที่มีเส้นขอบที่มากที่สุด (Maximum Margin) โดยแบ่งข้อมูลออกเป็น 2 คลาส ทำให้แยกข้อมูลได้โดยมีความผิดพลาดน้อยที่สุด โดยมี Support Vector เป็นตัวกำหนดขนาดเส้นขอบ ในการหาเส้นขอบที่มากที่สุด ข้อมูล × จะถูกแบ่งเป็นระนาบบวก และระนาบลบ [20] ดังรูปที่ 3

2.2 การทดสอบประสิทธิภาพของแบบจำลอง

งานวิจัยนี้รวบรวมข้อมูลทั้งหมดจำนวน 1,490 บทวิจารณ์ แบ่งข้อมูลที่ใช้ในการสร้างและทดสอบแบบจำลองเป็นร้อยละ 80 (1,192 บทวิจารณ์) และ 20 (298 บทวิจารณ์) ข้อมูลที่ใช้ทดสอบแบ่งเป็นด้านบวก เป็นกลาง และด้านลบ จำนวน 212, 50 และ 36 บทวิจารณ์ตามลำดับ และวิเคราะห์

ข้อมูลแบบ 10-folds Cross Validation ทั้งนี้ จากการแบ่งข้อมูลที่ใช้ในการสร้างและทดสอบแบบจำลองเป็น 60 : 40 70 : 30 และ 80 : 20 พบว่า การแบ่งแบบ 80 : 20 มีค่าความถูกต้องในการวิเคราะห์มากกว่า นอกจากนี้ ได้นำแบบจำลองที่ได้ไปทดสอบกับข้อมูลใหม่ (Unseen Data) จำนวน 298 บทวิจารณ์ แบ่งเป็นบทวิจารณ์ด้านบวก เป็นกลาง และด้านลบ จำนวน 209, 55 และ 34 บทวิจารณ์

ในการทดสอบประสิทธิภาพของแบบจำลอง นำเสนอด้วยการแสดงเวลาในการประมวลผลบนคอมพิวเตอร์หน่วยประมวลผลกลางรุ่น Intel® Core™ i7-8700K หน่วยความจำ 32 กิกะไบต์ โดยรันผลทั้งหมด 10 ครั้ง แล้วนำมาหาค่าเฉลี่ย และคำนวณหาค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความแม่นยำเฉลี่ยถ่วงน้ำหนัก (Weighted Average Precision) ค่าความระลึกเฉลี่ยถ่วงน้ำหนัก (Weighted Average Recall) ค่าวัดประสิทธิภาพ (F-Measure) และค่าความถูกต้อง (Accuracy) [21] ดังสมการที่ (2)–(7)

การคำนวณหาค่าความแม่นยำ

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

การคำนวณหาค่าความระลึก

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

ค่าความแม่นยำเฉลี่ยถ่วงน้ำหนัก

$$Weighted\ Average_P = \frac{\sum(P_i \times N_i)}{N} \quad (4)$$

ค่าความระลึกเฉลี่ยถ่วงน้ำหนัก

$$Weighted\ Average_R = \frac{\sum(P_i \times N_i)}{N} \quad (5)$$

การคำนวณหาค่าวัดประสิทธิภาพ

$$F\text{-}Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

การคำนวณหาค่าความถูกต้อง

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

3. ผลการทดลอง

จากการวิเคราะห์ข้อมูลจากเครือข่ายสังคมออนไลน์ Twitter (X) โดยจำแนกข้อมูลด้วยขั้นตอนวิธีแบบ Naive Bayes เปรียบเทียบกับขั้นตอนวิธีแบบ SVM โดยผลการทำนายประเภทของบทวิจารณ์ด้วยขั้นตอนวิธีแบบ Naive Bayes แสดงดังตารางที่ 1

ตารางที่ 1 ผลการวิเคราะห์ด้วยขั้นตอนวิธี Naive Bayes

ค่าที่ทำนาย/ค่าจริง	ด้านบวก	เป็นกลาง	ด้านลบ
ด้านบวก	204	33	11
เป็นกลาง	6	14	4
ด้านลบ	2	3	21
	Precision	**Recall**	**F1-Score**
ด้านบวก	0.82	0.96	0.89
เป็นกลาง	0.58	0.28	0.38
ด้านลบ	0.81	0.58	0.68
Weighted Avg	0.78	0.80	0.78

ผลการวิเคราะห์ข้อมูลด้วยขั้นตอนวิธี Naive Bayes พบค่า ความถูกต้องในการทำนายร้อยละ 80.20 ค่าความแม่นยำเฉลี่ยถ่วงน้ำหนักร้อยละ 78.00 ค่าความระลึกเฉลี่ยถ่วงน้ำหนักร้อยละ 80.00 และค่าวัดประสิทธิภาพร้อยละ 78.00 ซึ่งทำนายถูกต้องว่าเป็นบทวิจารณ์ด้านบวก เป็นกลาง และด้านลบจำนวน 204, 14 และ 21 บทวิจารณ์ ตามลำดับ โดยระยะเวลาในการประมวลผลคือ 1.51 วินาที ทั้งนี้ เมื่อทดสอบกับข้อมูลใหม่ ผลที่ได้แสดงดังตารางที่ 2

ตารางที่ 2 ผลการวิเคราะห์ด้วย Naive Bayes กับข้อมูลใหม่

ค่าที่ทำนาย/ค่าจริง	ด้านบวก	เป็นกลาง	ด้านลบ
ด้านบวก	200	35	11
เป็นกลาง	9	16	4
ด้านลบ	0	4	19
	Precision	**Recall**	**F1-Score**
ด้านบวก	0.81	0.96	0.88
เป็นกลาง	0.55	0.29	0.38
ด้านลบ	0.83	0.56	0.67
Weighted Avg	0.77	0.79	0.76

ผลการวิเคราะห์ด้วยขั้นตอนวิธี Naive Bayes กับข้อมูลใหม่ พบค่า ความถูกต้องในการทำนายร้อยละ 78.86 ค่าความแม่นยำเฉลี่ยถ่วงน้ำหนักร้อยละ 77.00 ค่าความระลึกเฉลี่ยถ่วงน้ำหนักร้อยละ 79.00 และค่าวัดประสิทธิภาพร้อยละ 76.00 ซึ่งทำนายบทวิจารณ์ด้านบวก เป็นกลาง และด้านลบถูกต้องจำนวน 200, 16 และ 19 บทวิจารณ์ มีระยะเวลาในการประมวลผล 1.54 วินาที

ผลการทำนายความรู้สึกด้วย SVM แสดงดังตารางที่ 3

ตารางที่ 3 ผลการวิเคราะห์ด้วยขั้นตอนวิธี SVM

ค่าที่ทำนาย/ค่าจริง	ด้านบวก	เป็นกลาง	ด้านลบ
ด้านบวก	202	23	9
เป็นกลาง	9	24	6
ด้านลบ	1	3	21
	Precision	Recall	F1-Score
ด้านบวก	0.86	0.95	0.91
เป็นกลาง	0.62	0.48	0.54
ด้านลบ	0.84	0.58	0.69
Weighted Avg	0.82	0.83	0.82

จากผลการวิเคราะห์ด้วย SVM มีค่าความถูกต้องในการทำนายร้อยละ 82.89 ค่าความแม่นยำเฉลี่ยถ่วงน้ำหนักร้อยละ 82.00 ค่าความระลึกเฉลี่ยถ่วงน้ำหนัก ร้อยละ 83.00 และค่าวัดประสิทธิภาพ ร้อยละ 82.00 ซึ่งทำนายบทวิจารณ์ด้านบวก ปานกลาง และด้านลบถูกต้องจำนวน 202, 24 และ 21 บทวิจารณ์ โดยระยะเวลาในการประมวลผล คือ 1.38 วินาที เมื่อทดสอบกับข้อมูลใหม่ ผลที่ได้แสดงดังตารางที่ 4

ตารางที่ 4 ผลการวิเคราะห์ด้วย SVM กับข้อมูลใหม่

ค่าที่ทำนาย/ค่าจริง	ด้านบวก	เป็นกลาง	ด้านลบ
ด้านบวก	191	24	9
เป็นกลาง	17	28	3
ด้านลบ	1	3	22
	Precision	Recall	F1-Score
ด้านบวก	0.85	0.91	0.88
เป็นกลาง	0.58	0.51	0.54
ด้านลบ	0.85	0.65	0.73
Weighted Avg	0.80	0.81	0.80

จากผลการวิเคราะห์ข้อมูลด้วยขั้นตอนวิธี SVM พบค่าความถูกต้องในการทำนายร้อยละ 80.87 ค่าความแม่นยำเฉลี่ยถ่วงน้ำหนัก ร้อยละ 80.00 ค่าความระลึกเฉลี่ยถ่วงน้ำหนัก ร้อยละ 81.00 และค่าวัดประสิทธิภาพ ร้อยละ 80.00 ซึ่งทำนายบทวิจารณ์ด้านบวก เป็นกลาง และด้านลบถูกต้องจำนวน 191, 28 และ 22 บทวิจารณ์ตามลำดับ โดยระยะเวลาในการประมวลผลคือ 1.43 วินาที

นอกจากนี้ เมื่อวิเคราะห์ข้อมูลโดยการนับความถี่ (TF-IDF) กับคำ แฮชแท็ก และสัญลักษณ์ต่างๆ 3 อันดับแรก ผลที่ได้จากการนับความถี่คำ แสดงดังตารางที่ 5

ตารางที่ 5 ผลการวิเคราะห์ TF-IDF ของคำ

ลำดับ	ด้านบวก	ความถี่
1	อร่อย	0.018524
2	อร่อยมาก	0.013394
3	ดี	0.010429
ลำดับ	**เป็นกลาง**	**ความถี่**
1	กินได้	0.017096
2	อยากกิน	0.073787
3	พอได้	0.009799
ลำดับ	**ด้านลบ**	**ความถี่**
1	เหี้ย	0.036148
2	ชาติหมา	0.025803
3	อย่าแดก	0.024391

จากผลการวิเคราะห์ความถี่คำพบว่า อันดับแรกของคำในบทวิจารณ์ด้านบวก คือ อร่อย มีความถี่ร้อยละ 1.85 ส่วนบทวิจารณ์ที่เป็นกลาง คำที่ปรากฏเป็นอันดับแรก คือ กินได้ มีความถี่ร้อยละ 1.71 และคำในบทวิจารณ์ด้านลบเป็นอันดับแรก คือ เหี้ย ซึ่งเป็นคำสบถ มีความถี่ร้อยละ 3.61

ผลการวิเคราะห์ความถี่แฮชแท็ก แสดงดังตารางที่ 6

ตารางที่ 6 ผลการวิเคราะห์ TF-IDF ของแฮชแท็ก

ลำดับ	ด้านบวก	ความถี่
1	#อร่อยไปแดก	0.125069
2	#อร่อยบอกต่อ	0.087849

ตารางที่ 6 ผลการวิเคราะห์ TF-IDF ของแฮชแท็ก (ต่อ)

ลำดับ	ด้านบวก	ความถี่
3	#รีวิวคาเฟ่	0.059043
ลำดับ	**เป็นกลาง**	**ความถี่**
1	#อร่อยไปแดก	0.028296
2	#รีวิวเชเว่น	0.023350
3	#รีวิวเกาหลี	0.022144
ลำดับ	**ด้านลบ**	**ความถี่**
1	#ไม่อร่อยอย่าแดก	0.021010
2	#รีวิวเชเว่น	0.019869
3	#ชิมช้อปใช้	0.016575

จากผลการวิเคราะห์ความถี่ของแฮชแท็กในบทวิจารณ์ด้านบวกและบทวิจารณ์ที่เป็นกลางมีความถี่เป็นอันดับแรก คือ #อร่อยไปแดก มีความถี่ร้อยละ 12.51 และ 2.83 ส่วนแฮชแท็กที่ปรากฏในบทวิจารณ์ที่ด้านลบที่มีความถี่เป็นอันดับแรกคือ #ไม่อร่อยอย่าแดก มีความถี่ร้อยละ 2.10

ผลการวิเคราะห์ความถี่สัญลักษณ์ แสดงดังตารางที่ 7

ตารางที่ 7 ผลการวิเคราะห์ TF-IDF ของสัญลักษณ์

ลำดับ	ด้านบวก	ความถี่
1	😂	0.023207
2	😋	0.015781
3	😍	0.013755
ลำดับ	**เป็นกลาง**	**ความถี่**
1	😋	0.015972
2	😂	0.013215
3	😭	0.009011
ลำดับ	**ด้านลบ**	**ความถี่**
1	😂	0.025229
2	😢	0.020167
3	😭	0.016575

จากตารางที่ 7 สัญลักษณ์ที่อันดับแรกที่ปรากฏในบทวิจารณ์ด้านบวกและด้านลบคือ 😂 มีความถี่ร้อยละ 2.32 และ 2.52 ส่วนบทวิจารณ์ที่เป็นกลางมีสัญลักษณ์ 😋 อยู่ในอันดับแรก โดยมีความถี่ร้อยละ 1.60 ทั้งนี้ ยังพบสัญลักษณ์อื่นๆ ที่ใช้ในการแสดงอารมณ์ เช่น 🤩 😻 😑 😆 😬 🙄 และ 👍 ปรากฏในบทวิจารณ์ทั้ง 3 ประเภทอีกด้วย

4. อภิปรายผลและสรุปผล

จากผลการวิเคราะห์โดยการจำแนกบทวิจารณ์ด้วยขั้นตอนวิธี Naive Bayes เปรียบเทียบกับขั้นตอนวิธี SVM ผลที่ได้พบว่า ขั้นตอนวิธี SVM ให้ค่าความถูกต้องในการทำนายร้อยละ 82.89 มากกว่าขั้นตอนวิธี Naive Bayes ที่ให้ค่าความถูกต้องในการทำนายร้อยละ 80.20 และเมื่อทดสอบกับข้อมูลใหม่ ขั้นตอนวิธี SVM ให้ค่าความถูกต้องในการทำนายร้อยละ 80.87 ส่วน Naive Bayes ให้ค่าความถูกต้องในการทำนายร้อยละ 78.86 ซึ่งสอดคล้องกับงานวิจัยก่อนหน้านี้ที่มีการเปรียบเทียบระหว่าง 2 ขั้นตอน วิธีดังกล่าว โดยการจำแนกบทวิจารณ์ด้วยขั้นตอนวิธี Naive Bayes และ SVM ทั้งสองขั้นตอนวิธีทำนายบทวิจารณ์ที่เป็นบวกได้ถูกต้องมากที่สุด โดยระยะเวลาในการประมวลผลขั้นตอนวิธี SVM ใช้เวลาน้อยกว่า Naive Bayes

จากผลการทำนายที่ขั้นตอนวิธี SVM ทำนายถูกต้องมากกว่าขั้นตอนวิธี Naive Bayes อาจสืบเนื่องมาจากคุณลักษณะของแต่ละขั้นตอนวิธี โดย SVM มีการจำแนกข้อมูลออกเป็นมิติ ซึ่งเหมาะสมกับการจำแนกข้อมูลที่มีมิติมากกว่า Naive Bayes ที่เป็นขั้นตอนวิธีซึ่งใช้ความน่าจะเป็นในการทำนาย อีกทั้ง ข้อมูลที่ใช้ในทดสอบมีจำนวนน้อย จึงทำให้การวิเคราะห์ด้วย Naive Bayes ซึ่งเป็นตอนวิธีที่ทำนายสิ่งที่ไม่เคยเกิดขึ้นจากสิ่งที่เคยเกิดขึ้นมาก่อน ไม่ได้มีการทำนายในจำนวนที่เหมาะสม จึงได้ค่าความถูกต้องในการทำนายน้อยกว่า SVM

ทั้งนี้ จากผลการทำนายที่ผิดพลาด อาจเกิดจากจำนวนของบทวิจารณ์ โดยในอนาคตผู้วิจัยอาจเพิ่มจำนวนของข้อมูลและศึกษาขั้นตอนวิธีอื่นๆ เพิ่มเติมมาเปรียบเทียบกัน เช่น ขั้นตอนวิธี Neural Network หรือการพัฒนาขั้นตอนวิธีในการตัดคำ เพื่อให้ได้ผลลัพธ์ในการทำนายที่ถูกต้องและตรงกับ

บทวิจารณ์ที่เป็นจริงมากขึ้น อีกทั้ง การนำผลที่ได้จากการวิเคราะห์ไปใช้สำหรับสร้างแบบจำลองในการแนะนำร้านอาหารแก่ผู้ใช้ในงานวิจัยครั้งต่อไป

นอกจากนี้ จากผลที่ได้พบว่า บทวิจารณ์ที่เป็นกลางถูกทำนายว่าเป็นบทวิจารณ์ด้านบวกเป็นจำนวนมาก อาจเป็นผลมาจากบทวิจารณ์ด้านบวกและบทวิจารณ์ที่เป็นกลางมีการใช้คำที่บ่งบอกความรู้สึกคล้ายคลึงกัน อีกทั้งบทวิจารณ์ที่รวบรวมมาส่วนใหญ่เป็นบทวิจารณ์ด้านบวก จึงทำให้ผลการวิเคราะห์บทวิจารณ์ด้านบวกมีความแม่นยำในการทำนายมากกว่าบทวิจารณ์ด้านอื่นๆ ซึ่งหากมีอัตราส่วนของบทวิจารณ์แต่ละด้านที่ใกล้เคียงกัน อาจทำให้ผลการวิเคราะห์ถูกต้องและแม่นยำมากยิ่งขึ้น

ในการวิเคราะห์หาความถี่ของคำ แฮชแท็ก และสัญลักษณ์พบว่า คำ แฮชแท็ก และสัญลักษณ์อันดับแรกแสดงให้เห็นถึงความรู้สึกของผู้วิจารณ์ที่มีต่ออาหาร หรือร้านอาหารได้ และผลที่ได้จากการวิเคราะห์ดังกล่าว สอดคล้องกับประเภทของบทวิจารณ์ โดยข้อค้นพบนี้ทำให้สามารถนำผลที่ได้ไปใช้ในการวิเคราะห์ข้อความในบทวิจารณ์ในงานวิจัยในอนาคต เพื่อจำแนกประเภทของบทวิจารณ์ว่าเป็นบทวิจารณ์ด้านบวก ด้านลบ หรือเป็นกลาง ซึ่งนอกจากคำที่ปรากฏในบทวิจารณ์แล้ว การวิเคราะห์จากสัญลักษณ์ในบทวิจารณ์เป็นอีกวิธีหนึ่งในการจำแนกประเภทบทวิจารณ์ได้ ซึ่งจะมีประโยชน์อย่างมากสำหรับงานวิจัยที่มีวัตถุประสงค์เพื่อการวิเคราะห์ความรู้สึกของผู้ใช้ที่มีต่ออาหาร และร้านอาหาร

การจำแนกองค์ประกอบอื่นๆ ในบทวิจารณ์ที่นอกเหนือจากการวิจัยในครั้งนี้ อาจถูกนำไปใช้วิเคราะห์ในงานวิจัยครั้งต่อไป เพื่อพัฒนาแบบจำลองสำหรับแนะนำร้านอาหารต่อไป โดยงานวิจัยในอนาคตเป็นการรวบรวมบทวิจารณ์และภาพเกี่ยวกับอาหาร เพื่อสกัดคุณลักษณะ และแนะนำร้านอาหารที่เฉพาะเจาะจงสำหรับผู้บริโภคแต่ละคน โดยพิจารณาจากความชอบของผู้บริโภคผ่านการใช้เครือข่ายสังคมออนไลน์ร่วมกับคุณลักษณะของร้านอาหารที่ทำให้ผู้บริโภคตัดสินใจเลือกร้านรับประทานอาหารที่ร้านอาหาร โดยใช้เทคนิคการเรียนรู้แบบเชิงลึก เพื่อแนะนำร้านอาหารที่มีคุณลักษณะตามความต้องการของผู้บริโภค

เอกสารอ้างอิง

[1] Kasikorn Bank. (2017). 5 Business Trends that surpass of 2017. Kasikorn Bank. Bangkok. Thailand [Online]. (in Thai). Available: https://www.kasikornbank.com/th/business/sme/KSMEKnowledge/article/KSMEAnalysis/Documents/5BusinessTrendsOf2017.pdf.

[2] CleverKid. (2015). Eat share connect: Hot lifestyle in the era of the single app. iReview. Bangkok. Thailand [Online]. (in Thai). Available: https://ireview.in.th/pr-eat-share-connect-lifestyle-force/.

[3] Google Trends. (2017). *See what was trending in 2017 - Thailand*. [Online]. (in Thai). Available: http://trends.google.com/trends/yis/2017/TH.

[4] B. Vutidhammakhun, "The social media restaurant reviews and the decision making in choosing restaurant in Bangkok," B. M. Com. Arts thesis, Independent Study of Graduate School, Digital Marketing Communications, Bangkok University, 2017 (in Thai).

[5] J. Dancy, (2012). *Dine and Dish: Are Social Media and Food the Perfect Pairing*. [Online]. Available: https://www.columnfivemedia.com/work-items/infographic-dine-and-dish-are-social-media-and-food-the-perfect-pairing.

[6] Bright Local. (2017). *Local Consumer Review Survey 2017*. [Online]. Available: https://www.brightlocal.com/learn/local-consumer-review-survey/

[7] R. Burke, "Hybrid web recommender systems," *The Adaptive Web*, vol. 4321, pp. 377–408, 2007.

[8] W.T. Gao, W. Yu, P. Chao, R. Zhang, A. Zhou, and X. Yang, "A restaurant recommendation

system by analyzing ratings and aspects in reviews," *International Conference on Database Systems for Advanced Applications*. vol. 9050, pp. 526–530, 2015.

[9] S. Higgs and J. Thomas, "Social influences on eating," *Current Opinion in Behavioral Sciences*, vol. 9, pp. 1–6, 2016.

[10] I. K. C. U. Perera and H. A. Caldera, "Aspect based opinion mining on restaurant reviews," presented at the 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), Beijing, China, Sep. 8–11, 2017.

[11] Y. M. Aye and S. S. Aung, "Sentiment analysis for reviews of restaurants in Myanmar text," presented at the 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Kanazawa, Japan, Jun. 26–28, 2017.

[12] J. B. Schafer, J. A. Konstan, and J. Riedl, "E-commerce recommendation applications," *Data Mining and Knowledge Discovery*, vol. 5, pp. 115–153, 2001.

[13] N. G. Bhojne, S. Deore, R. Jagtap, G. Jain and C. Kalal, "Collaborative approach based restaurant recommender system using naive bayes," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 6, no. 4, pp. 6–13, 2017.

[14] P. Sasikala and M. I. S. Lourdusamy, "Sentiment analysis of online food reviews using customer ratings," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 15, pp. 3509–3514, 2018.

[15] B. Yu, J. Zhou, Y. Zhang, and Y. Cao. (2017, September). Identifying restaurant features via sentiment analysis on yelp reviews. *arXiv*. [Online]. Available: https://arxiv.org/abs/1709.08698

[16] N. Raul, Y. Shah, and M. Devganiya, "Restaurant revenue prediction using machine learning," *International Journal of Engineering and Science*, vol. 6, no. 4, pp. 2319–6483, 2016.

[17] W. T. Chu and Y. L. Tsai, "A hybrid recommendation system considering visual information for predicting favorite restaurants," *World Wide Web Journal: Internet and Web Information Systems Manuscript*, vol. 20, no. 6, pp. 1313–1331, 2017.

[18] F. Nurifan, R. Sarno, and K. S. Sungkono, "Aspect based sentiment analysis for restaurant reviews using hybrid ELMo-wikipedia and hybrid expanded opinion lexicon-senticircle," *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 6, pp. 47–58, 2019.

[19] J. D. M. Rennie, L. Shih, J. Teevan, and D. R. Karger, "Tackling the poor assumptions of naïve bayes text classifiers," in *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, 2003.

[20] T. Fletcher, *Support Vector Machines Explained*. London's Global University: UK, 2008.

[21] J. Davisand and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings the 23rd International Conference on Machine Learning*, 2006, pp. 233–2