

ตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์ สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์

พิศิษฐ์ บวรเลิศสุธี¹ วรภัทร ไพรเกรง^{2*}

^{1,2*}สาขาเทคโนโลยีสารสนเทศ วิทยาลัยนวัตกรรมการเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์,
กรุงเทพมหานคร, ประเทศไทย

*ผู้ประพันธ์บรรณกิจ อีเมล : worapat.png@dpu.ac.th

รับต้นฉบับ : 1 พฤศจิกายน 2564; รับบทความฉบับแก้ไข : 21 มกราคม 2565; ตอบรับบทความ : 28 กุมภาพันธ์ 2565
เผยแพร่ออนไลน์ : 27 มิถุนายน 2565

บทคัดย่อ

การตัดสินใจซื้อสินค้าและบริการออนไลน์ ผู้บริโภคส่วนใหญ่จะทำการสืบค้นข้อมูลเกี่ยวกับสินค้าและบริการรวมทั้งพิจารณาจากการแสดงความคิดเห็น (Comments) และบทวิจารณ์ (Reviews) จากผู้ที่เคยซื้อสินค้าไปแล้ว เพื่อเป็นแนวทางในการตัดสินใจเลือกซื้อ อย่างไรก็ตาม ข้อมูลการแสดงความคิดเห็นและบทวิจารณ์ในปัจจุบันนี้มีปริมาณมหาศาลและเกิดขึ้นใหม่อยู่ตลอดเวลา ซึ่งทำให้ผู้บริโภคต้องเสียเวลาในการวิเคราะห์ข้อดีข้อเสียของผลิตภัณฑ์เหล่านี้ด้วยตนเองเพื่อใช้ในการตัดสินใจเลือกซื้อสินค้าและบริการ ซึ่งถ้าหากมีการสร้างตัวแบบเพื่อแจ้งให้ผู้ใช้งานทราบถึงการวิเคราะห์ความรู้สึกได้จะช่วยให้ผู้ซื้อตัดสินใจได้รวดเร็วมากยิ่งขึ้น ดังนั้นงานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอ ตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์โดยใช้เทคนิคส่วนการเรียนรู้ของเครื่อง (Machine Learning) ร่วมกับเทคนิคการตัดคำ (Word Tokenization) และการสร้างคลังคำศัพท์ (Bag of Words) จากนั้นนำเข้ากระบวนการจำแนกประเภทผลการวิเคราะห์ 4 เทคนิค ได้แก่ LSTM, SGD, Logistic Regression และ Support Vector Machines ในการสร้างตัวแบบมีกระบวนการทดลอง 5 ขั้นตอน ได้แก่ 1) การเตรียมข้อมูล 2) การตัดคำ 3) การฝึกอบรมข้อมูล 4) ขั้นตอนการแยกประเภท และ 5) การประเมินตัวแบบ ทำการทดลองโดยใช้ข้อมูลตัวอย่างการแสดงความคิดเห็นต่อสินค้าและบริการออนไลน์ภาษาไทย จำนวน 12,900 ข้อมูล ใช้เป็นข้อมูลในการสร้างตัวแบบ เพื่อช่วยให้ผู้บริโภคใช้ในการตัดสินใจซื้อสินค้าและบริการ และช่วยให้ผู้ประกอบการมีข้อมูลสำหรับการพัฒนาสินค้าและบริการต่อไปในอนาคต จากผลการทดลองสร้างตัวแบบการวิเคราะห์ระดับความรู้สึกทางอารมณ์ 3 ระดับ คือ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) โดยแต่ละตัวแบบให้ค่าความถูกต้องในการทำนายผล ดังนี้ LSTM 81.27%, Logistic Regression 69%, SGD 66%, และ Support Vector Machines 65% สรุปได้ว่าการสร้างตัวแบบโดยใช้เทคนิคการจำแนกประเภทแบบ LSTM ให้คะแนนความถูกต้องสูงที่สุดในการทำนายผลด้วยการเรียนรู้เชิงลึก จึงเหมาะสมแก่การนำไปใช้สร้างตัวแบบในการวิเคราะห์ระดับความรู้สึกกับข้อความที่เป็นภาษาไทยโดยพิจารณาจากค่า F1

คำสำคัญ : การวิเคราะห์อารมณ์ความรู้สึก การเรียนรู้ของเครื่อง การซื้อสินค้าออนไลน์ การตัดคำ

The Model of Sentiment Analysis for Classifying the Online Shopping Reviews

Pisit Bowornlertsutee¹ Worapat Paireekreng^{2*}

^{1,2*}*Information Technology, College of Innovative Technology and Engineering, Dhurakij Pundit University,
Bangkok, Thailand*

*Corresponding Author. E-mail Address: worapat.png@dpu.ac.th

Received: 1 November 2021; Revised: 21 January 2022; Accepted: 28 February 2022

Published online: 27 June 2022

Abstract

Nowadays, users can make a decision to order online goods and services from searching information related to goods and services. These may be based on opinions and reviews from previous purchasers as a guideline for purchasing decisions. Moreover, the current opinions information and reviews are enormous and increased all the time. consumers have to spend time for information analytics. Therefore, the model of sentiment analysis regarding goods and services reviews is needed. This research aims to build a model of Sentiment analysis with 3-level of emotion. they are positive neutral and negative, regarding previous user's reviews and opinions towards online products and services. The techniques used in this research are Machine Learning including Word Segmentation and Bag of Words which compared four categories of sentiment analysis methods: LSTM, SGD, Logistic Regression and Support Vector Machines. There are 5 steps for model building as following: 1) Data Preparation Phase 2) Word Tokenization Phase 3) Training & Streaming Phase 4) Classification phase and 5) Model Evaluation Phase. The consumers' opinions were gathered the datasets from open data with number of 12,900 comments. The model can help consumers to make a decision for purchasing of goods and services, and help entrepreneurs gain the information. This is to improve products and services in the future. This proposed method can classify the opinions into 3 scales which are positive, neutral and negative opinions. In summary, the proposed sentiment analysis model can perform LSTM accuracy is at 81.27%, Logistic Regression accuracy is at 69%, SGD accuracy is at 66% and Support Vector Machines accuracy is at 65%. However, the LSTM shows better performance on the classification compared to other techniques with deep learning approach. It also found that the F1-score can be implemented for Thai text appropriately.

Keywords: Sentiment analysis, Machine learning, Online shopping, Word tokenization

1) บทนำ

ในปัจจุบันประเทศไทยถือเป็นแหล่งซื้อขายสินค้าออนไลน์ที่สำคัญแห่งหนึ่งของโลก สะท้อนได้จากปริมาณการครอบครองโทรศัพท์มือถือและการเข้าถึงอินเทอร์เน็ต ซึ่งมีแนวโน้มขยายตัวอย่างต่อเนื่อง แสดงถึงการตอบรับที่ดีในด้านการซื้อขายสินค้าออนไลน์ในประเทศไทย อีกทั้งยังมีเว็บไซต์ที่เปิดให้บริการในการซื้อขายสินค้าออนไลน์ชื่อดังอยู่เป็นจำนวนมาก เช่น LAZADA.COM SHOPEE.COM เป็นต้น [1] ในการตัดสินใจซื้อสินค้าและบริการผู้บริโภคส่วนใหญ่จะทำการสืบค้นข้อมูลเกี่ยวกับตัวสินค้าและบริการ เช่น การอ่านข้อมูลการแสดงความคิดเห็น และบทวิจารณ์จากผู้ที่เคยซื้อสินค้าไปแล้ว เป็นต้น เนื่องจากเว็บไซต์ที่เปิดให้บริการซื้อขายออนไลน์ส่วนใหญ่จะเปิดให้มีการแสดงความคิดเห็น (Opinion) ต่อตัวสินค้าและบริการดังกล่าว ทำให้ผู้ที่สนใจในผลิตภัณฑ์เข้าไปศึกษาเพื่อใช้เป็นแนวทางในการตัดสินใจเลือกซื้อผลิตภัณฑ์ ซึ่งการแสดงความคิดเห็นเป็นการแสดงถึงความรู้สึกทางอารมณ์ในรูปแบบข้อความที่มีต่อสินค้าและบริการ อีกทั้งข้อมูลการแสดงความคิดเห็นในปัจจุบันนั้นมีปริมาณมหาศาลและเกิดขึ้นใหม่อยู่ตลอดเวลา ซึ่งทำให้ผู้บริโภคต้องเสียเวลาในการวิเคราะห์ข้อมูลข้อดีข้อเสียของผลิตภัณฑ์เหล่านี้ [2] เพื่อใช้ในการตัดสินใจเลือกซื้อสินค้าและบริการ และเนื่องจากยังไม่มีตัวช่วยในการวิเคราะห์การแสดงความคิดเห็นต่อสินค้าและบริการแบบอัตโนมัติที่จะช่วยสรุปข้อดีข้อเสียเพื่อช่วยในการตัดสินใจในการเลือกซื้อสินค้าและบริการได้อย่างสะดวกและรวดเร็ว

การวิเคราะห์ความรู้สึก (Sentiment Analysis) เป็นการนำเอาเทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) ที่สามารถใช้วิเคราะห์ความรู้สึกทั้งในระดับเอกสาร ระดับประโยค และในระดับคุณลักษณะของวัตถุ รวมถึงการใช้วิเคราะห์ความคิดเห็นของผู้บริโภคที่มีความรู้สึกต่อสินค้าและบริการต่าง ๆ รวมถึงการทำเหมืองความคิดเห็น (Opinion Mining) โดยมีวัตถุประสงค์ในการสกัดข้อความที่เป็นความคิดเห็นที่มีต่อสิ่งใดสิ่งหนึ่งออกมาจากข้อเท็จจริงเพื่อนำไปสกัดคุณลักษณะและส่วนประกอบของวัตถุที่แสดงความคิดเห็นออกมา เพื่อพิจารณาว่าความคิดเห็นนั้นแสดงความรู้สึกไปในเชิงบวก เชิงลบ หรือเป็นกลาง [3] และเทคนิคการทำเหมืองข้อมูล (Text mining) มาประยุกต์ใช้เพื่อมุ่งเน้นการวิเคราะห์และตรวจสอบความรู้สึก (Opinion) ได้อย่างอัตโนมัติ [4] รวมถึงการแทนค่าในความคิดเห็นให้อยู่ในรูปแบบ

โครงสร้าง เช่น การแทนด้วยถุงคำ (Bag of Words) และค่าความถี่ของคำที่เกิดขึ้น (TF-IDF) [2], [5] และเทคนิคการสกัดคำ (Word Extraction) ซึ่งเป็นกระบวนการของการทำเหมืองข้อความเพื่อใช้ในการวิเคราะห์คำออกจากเอกสาร ข่าวสาร ข้อความ และสารสนเทศต่าง ๆ ที่เป็นตัวอักษร โดยสามารถนำไปทำการแบ่งกลุ่ม (Clustering) จำแนกกลุ่ม และหาความสัมพันธ์ขั้นตอนการแบ่งกลุ่มโดยใช้เทคนิคต่าง ๆ เช่น 1) การตัดคำ (Word Segmentation) เป็นการแยกแต่ละคำจากเอกสารออกจากกัน โดยยังคงมีความหมายที่ถูกต้องสมบูรณ์อยู่ โดยการตัดคำนั้นใช้ฐานข้อมูลพจนานุกรมคำศัพท์ในการแบ่งคำออกมา ในการตัดคำภาษาไทยจะนิยมใช้อัลกอริทึมจาก PyThaiNLP ในการตัดคำภาษาไทยโดยเฉพาะ เช่น อัลกอริทึม NewMM Engine พัฒนาจากเทคนิคการแบ่งส่วนคำศัพท์ภาษาไทยด้วยพจนานุกรม (Dictionary-based) โดยใช้อัลกอริทึมการจับคู่สูงสุดและการจัดกลุ่มอักขระภาษาไทย อัลกอริทึม Longest Engine ซึ่งเป็นอัลกอริทึมพัฒนาจากเทคนิคการแยกคำภาษาไทยที่ยาวที่สุดที่จับคู่กับพจนานุกรม (Dictionary-based) และ อัลกอริทึม AttaCut Engine ซึ่งเป็นอัลกอริทึมที่สามารถตัดคำภาษาไทยได้อย่างรวดเร็วและแม่นยำ โดยใช้อัลกอริทึมการจัดกลุ่ม (Clustering) อักขระภาษาไทย (Rules purposed) [6], [7] 2) การกำจัดคำหยุด (Stop word) เป็นการตัดคำที่ไม่มีความหมายออกจากเอกสารโดยการกำจัดคำหยุดนั้นใช้ฐานข้อมูลคำศัพท์ที่เป็นคำที่ไม่มีความหมายในการนำมาประมวลผล [7], [8] เพื่อช่วยให้การจำแนกประเภท (Classification) มีความถูกต้องมากยิ่งขึ้น [9] ในส่วนของการวิเคราะห์ความรู้สึกและการประมวลผลในภาษาไทยยังมีความยุ่งยากมากกว่าภาษาอังกฤษ เพราะโครงสร้างประโยคและไวยากรณ์มีความซับซ้อน ไม่มีการแบ่งส่วน แบ่งคำ ประโยคติดกัน อักขระที่ไม่มีขอบเขตคำ และไม่มีโครงสร้างที่ชัดเจนทำให้เกิดความกำกวม [3], [10], [11] นอกจากนี้ยังมีการพัฒนาเฟรมเวิร์ค (Framework) ที่เรียกว่า S-Sense (Social Media Sensing) สำหรับวิเคราะห์ความรู้สึกบนโซเชียลมีเดียสำหรับภาษาไทย ตัวแบบจะเปรียบเทียบประสิทธิภาพระหว่างชุดคำศัพท์ที่แตกต่างกัน 2 ชุด คือ คำศัพท์ทั่วไปและคำใบ้ ซึ่งการรวมคำใบ้เข้ากับเวกเตอร์คุณลักษณะสำหรับการสร้างแบบจำลองการจำแนกประเภททำให้มีการปรับปรุงอย่างมีนัยสำคัญในแง่ของความถูกต้อง [12] และในปัจจุบันการใช้เทคนิคการเรียนรู้ด้วยเครื่อง (Machine Learning) จะเป็นที่

นิยมมากกว่า โดยจะอาศัยการสร้างชุดข้อมูลฝึกฝน (Training Set) และนำชุดข้อมูลทดสอบ (Test Set) [8], [13] เข้ากระบวนการจัดกลุ่มหรือจำแนกประเภท (Classification) ในส่วนของการเลือกใช้ตัวจำแนกประเภทที่มีความนิยมในการพัฒนา Sentiment Analysis ภาษาไทย ได้แก่ Logistic Regression, Random Forest, Support Vector Machine, Naive Bayes, K-Nearest Neighbors, และ decision tree เป็นต้น [8], [14], [15] และในการประเมินประสิทธิภาพของตัวแบบที่ได้โดยสรุปจาก Accuracy, Precision, Recall และ F1 Measure [3], [14] นอกจากนี้ การวิเคราะห์ข้อความความรู้สึกส่วนใหญ่จะเป็นการตัดสินใจเพียง 2 ระดับ คือ เชิงบวกและเชิงลบ อย่างไรก็ตาม จะมีความคิดเห็นที่เป็นกลาง ซึ่งถ้าหากถูกตัดสินเพียง 2 ระดับ อาจจะทำให้ข้อคิดเห็นโดยรวมเกิดอคติได้ ดังนั้น ถ้าหากมีการวิเคราะห์เป็น 3 ระดับ คือ มีระดับเป็นกลางเพิ่มเข้ามาจะทำให้หน้าหน้ของการตัดสินใจข้อความรู้สึกมีความถูกต้องมากขึ้น [6] อีกทั้งการตัดคำ (Word Segmentation) จะต้องเลือกโมดูล (Module) ที่เหมาะสม ถูกพัฒนาขึ้นเพื่อใช้งานกับภาษาไทยโดยเฉพาะทั้งนี้เพื่อให้การตัดคำถูกต้องและพร้อมใช้งานในการฝึกฝนข้อมูล

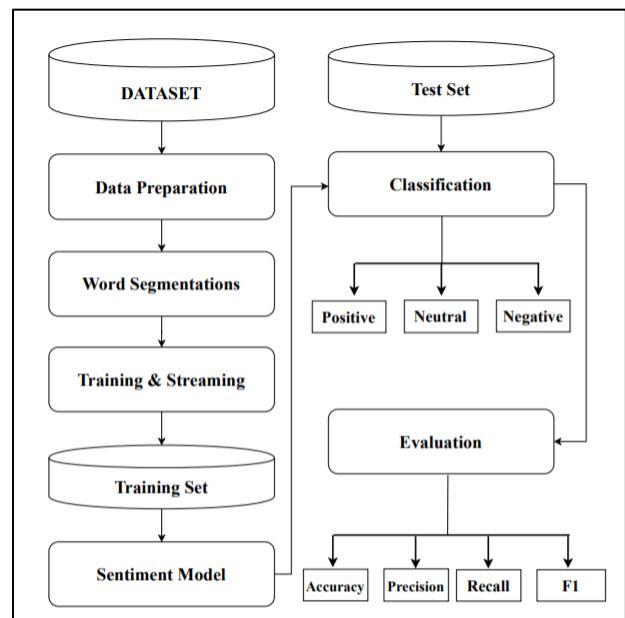
การศึกษาค้นคว้าครั้งนี้ผู้วิจัยมีแนวคิดที่จะสร้างตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์ สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์ (Sentiment Analysis Model) โดยใช้เทคนิคส่วนการเรียนรู้ของเครื่อง (Machine Learning) แบบมีผู้สอน (Supervised Learning) ร่วมกับเทคนิคการทำเหมืองข้อความ (Text Mining) เพื่อให้ตัวแบบสามารถแก้ไขปัญหาในการแบ่งแยกคำภาษาไทยและวิเคราะห์คำจากการแสดงความคิดเห็นได้อย่างมีประสิทธิภาพ นอกจากนั้นจะมีการใช้เทคนิคการจำแนกประเภทโดยมีการเปรียบเทียบผลการวิเคราะห์ 4 รูปแบบ ได้แก่ Long Short-Term Memory (LSTM), Stochastic Gradient Descent (SGD), Logistic Regression (LR) และ Support Vector Machines (SVM) โดยทำการทดลองวิเคราะห์ประโยคเนื้อหาที่เป็นภาษาไทย 3 ระดับ ได้แก่ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) กับชุดข้อมูลจำนวน 12,900 ข้อมูล เพื่อหาตัวแบบที่ดีที่สุดที่สามารถใช้เป็นแนวทางในการพัฒนาตัวแบบการวิเคราะห์ระดับความรู้สึกทางอารมณ์สำหรับบทความแนะนำสินค้าออนไลน์ภาษาไทย ทั้งนี้เพื่อช่วยให้ผู้บริโภคสามารถตัดสินใจเลือกซื้อสินค้าและบริการได้อย่างรวดเร็ว

2) วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์ภาษาไทย
2. เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคการจำแนกประเภทที่เหมาะสมในการทำ Sentiment Analysis

3) วิธีดำเนินการวิจัย

ขั้นตอนการดำเนินการวิจัย และการออกแบบการทดลอง (Experimental Design) ประกอบด้วย 5 ขั้นตอน โดยแต่ละขั้นตอนใช้เครื่องมือการพัฒนาโปรแกรมภาษาไพทอน (Python) ร่วมกับโมดูล (Module) ต่าง ๆ แสดงดังรูปที่ 1



รูปที่ 1 : แสดงขั้นตอนการดำเนินงานวิจัย และการออกแบบการทดลอง 5 ขั้นตอน

จากรูปที่ 1 สามารถอธิบายขั้นตอนการดำเนินการวิจัย และการออกแบบการทดลองได้ดังต่อไปนี้

3.1) การเตรียมข้อมูล (Data Preparation)

การเก็บข้อมูลเก็บจากชุดข้อมูลที่ให้เข้าถึงสาธารณะจากเว็บไซต์ <https://pythainlp.github.io/> (Wisensight Sentiment Corpus) เข้าถึงเมื่อวันที่ 10 ธันวาคม 2564 จำนวน 12,900 ข้อมูล 257,834 คำ ซึ่งเป็นข้อมูลการแสดงความคิดเห็นต่อสินค้าและบริการภาษาไทยที่มีการติดป้ายกำกับ (Label) ความรู้สึก 3 ระดับ ได้แก่ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral)

และเชิงลบ (Negative) โดยสามารถแสดงจำนวนข้อมูลแยกตามระดับความรู้สึกได้ดังตารางที่ 1

ตารางที่ 1 : แสดงจำนวนข้อมูลแยกตามระดับความรู้สึก

ความรู้สึก	จำนวนข้อมูล (Rows)
เชิงบวก	4,300
เป็นกลาง	13,105
เชิงลบ	6,140
รวม	23,545

จากตารางที่ 1 พบว่าชุดข้อมูลจะมีการติดป้ายกำกับความคิดเห็น 3 ระดับเรียบร้อยแล้ว จากชุดข้อมูลที่รวบรวมมามีขนาดข้อมูลที่เป็นประเภทของการแสดงความคิดเห็นในแต่ละประเภทไม่เท่ากัน โดยจะมีความคิดเห็นเป็นกลางมากที่สุด ส่วนความคิดเห็นเชิงบวกและเชิงลบมีจำนวนน้อยกว่า ดังนั้นจึงได้มีการปรับขนาดของชุดข้อมูลในแต่ละประเภทความคิดเห็นโดยการทำการสุ่มเลือกความคิดเห็นแบบเป็นระบบ โดยนับเป็นความคิดเห็นเป็นกลาง เชิงบวก และเชิงลบ ให้มีจำนวนเท่า ๆ กัน เพื่อนำไปจัดการการเรียนรู้ข้อมูลเพื่อสร้างตัวแบบ แสดงดังตารางที่ 2

ตารางที่ 2 : แสดงจำนวนข้อมูลที่จะนำไปสร้างตัวแบบ

ความรู้สึก	จำนวนข้อมูล (Rows)
เชิงบวก	4,300
เป็นกลาง	4,300
เชิงลบ	4,300
รวม	12,900

จากตารางที่ 2 จะมีการสุ่มเลือกข้อมูล (Dataset) สำหรับสร้างตัวแบบให้มีขนาดเท่ากันทั้ง 3 ระดับความรู้สึก โดยใช้วิธีการทำ Down sampling เป็นการลดขนาดข้อมูลที่มีจำนวนมาก ทั้งนี้เพื่อป้องกันการเกิดอคติในการสร้างตัวแบบในการวิเคราะห์ผล และสามารถแสดงตัวอย่างการแสดงความคิดเห็นแสดงแยกตามความรู้สึกได้ดังตารางที่ 3

ตารางที่ 3 : ตัวอย่างการแสดงความคิดเห็นแยกตามระดับความรู้สึก

การแสดงความคิดเห็น	ความรู้สึก
บอกเลยว่า Nissan Navara ตัวนี้เข้าใช้มากครับ	เชิงบวก
Nissan Note. ดาวน์เท่าไรคะ	เป็นกลาง
Nissan รถห่วยไม่ทนเลยขายไม่ดี	เชิงลบ

ในขั้นตอนต่อมาคือการนำชุดข้อมูลเข้าสู่กระบวนการทำความสะอาดข้อมูล (Data Cleaning) เพื่อแก้ไข ปรับแต่ง และตัดส่วนที่ไม่ต้องการออกจากชุดข้อมูล เช่น HTML Tag, URL หรือเครื่องหมายคำพูด (Punctuation) เป็นต้น และมีการตรวจสอบและแก้ไขคำที่สะกดผิดโดยใช้ PyThaiNLP Spellchecker เพื่อให้ข้อมูลมีคุณภาพ มีความถูกต้องและเหมาะสมที่จะนำไปใช้ในการสร้างตัวแบบ

3.2) การตัดคำ (Word Segmentation)

คือการแยกคำภาษาไทยที่เป็นประโยคยาว ๆ ออกมาเป็นแต่ละคำ เนื่องจากภาษานั้นไม่มีการใช้ตัวอักษรหรือสัญลักษณ์ที่นำมาใช้คั่นระหว่างคำและการเว้นวรรคเหมือนกับภาษาอังกฤษในการศึกษาวิจัยนี้ใช้กระบวนการประมวลผลภาษาธรรมชาติ (Natural language processing : NLP) หรือโมดูล PyThaiNLP โดยใช้ NewMM Engine พัฒนาจากเทคนิคการแบ่งส่วนคำศัพท์ภาษาไทยด้วยพจนานุกรม (Dictionary-based) โดยใช้อัลกอริทึมการจับคู่สูงสุดและการจัดกลุ่ม (Clustering) ถูกพัฒนาขึ้นเพื่อตัดคำภาษาไทยบนภาษา Python เพื่อหาความสัมพันธ์ระหว่าง "คำ" และ "เอกสาร" ที่มีอยู่ในรูปแบบของเวกเตอร์ (Vector) เพื่อให้คอมพิวเตอร์สามารถแปลผลได้ ทั้งนี้เพื่อให้รูปแบบของคำเป็นแบบมีโครงสร้างและสามารถทำการวิเคราะห์ระดับของคำได้ จากนั้นทำการลบ Stop words คือการตัดคำที่ไม่จำเป็นสำหรับการวิเคราะห์ออก ทั้งนี้เพื่อลดขนาดของข้อมูลที่ใช้ในการประมวลผล [2] และ Punctuation (เครื่องหมายวรรคตอน) ออกจากข้อความและเปลี่ยนข้อความให้มีช่องว่างระหว่างคำเพื่อให้ง่ายต่อการนำไปสร้าง Bag of Words (Bow) ตัวอย่างการตัดคำด้วย NewMM Engine แสดงดังตารางที่ 4

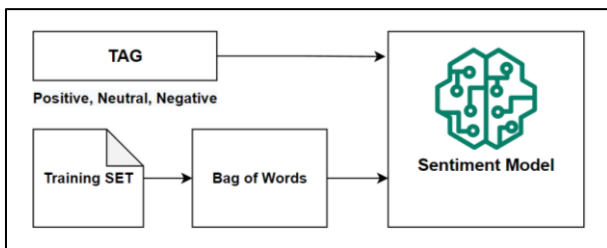
ตารางที่ 4 : ตัวอย่างการตัดคำ

ข้อความต้นฉบับ
บอกเลยว่า Nissan Navara ตัวนี้เข้าใช้มากครับ
ข้อความที่ผ่านการตัดคำโดยใช้ NewMM Engine
['บอก', 'เลย', 'ว่า', 'Nissan', 'Navara', 'ตัว', 'นี้', 'เข้า', 'ใช้', 'มาก', 'ครับ']

3.3) การฝึกอบรมข้อมูล (Training & Streaming)

ในขั้นตอนการสร้างตัวแบบสำหรับการเรียนรู้แบบมีผู้สอน (Supervised Learning) จะทำการแบ่งชุดข้อมูลออกเป็น 2 ชุด (Split Test) คือ ชุดข้อมูลฝึกฝน (Training Set) 10,320 ข้อมูล

หรือ 80% และชุดข้อมูลทดสอบ (Testing Set) 2,580 ข้อมูล หรือ 20% ซึ่งการสร้างแบบจำลองวิเคราะห์ความรู้สึกจากข้อความ จะมีการเพิ่มขั้นตอนการแปลงข้อมูลให้อยู่ในรูปแบบที่มีจำนวนของตัวแปรที่เท่ากัน โดยใช้เทคนิคการสร้างถุงคำ (Bag of Words) เพื่อให้ง่ายต่อการนำไปประมวลผล จากนั้นนำเข้ากระบวนการการเรียนรู้โดยอาศัยข้อมูลที่ป้อนเข้าไปเก็บไว้เป็นตัวอย่างเพื่อให้คอมพิวเตอร์ใช้ในการเปรียบเทียบกับข้อมูลใหม่ที่เข้ามาแล้วทำนายหรือจัดหมวดหมู่ที่มีความใกล้เคียงกันมากที่สุดให้อยู่ในกลุ่มเดียวกัน แสดงขั้นตอนการทำงานได้ดังรูปที่ 2



รูปที่ 2 : แสดงขั้นตอนการฝึกอบรมข้อมูล

3.4) ขั้นตอนการแยกประเภท (Classification)

คือกระบวนการสกัดความคิดเห็นจากการเรียนรู้ของเครื่อง (Machine Learning) แบบมีผู้สอน (Supervise Learning) โดยทดสอบกับชุดข้อมูล (Testing Set) 20% นำเข้ากระบวนการทำนายผล (Prediction) และแยกผลการวิเคราะห์ระดับความรู้สึก (Sentiment Analysis) โดยใช้เทคนิคการแยกประเภทผลการวิเคราะห์ 4 เทคนิค ได้แก่

3.4.1) Long Short-Term Memory (LSTM) คืออัลกอริทึมเรียนรู้เชิงลึก (Deep Learning) ที่ใช้เทคนิคบนพื้นฐานของโครงข่ายประสาทเทียมแบบเกิดซ้ำ (Neural Network) แบบหน่วยความจำระยะสั้น โดยมุ่งเน้นไปที่การเรียนรู้การเป็นตัวแทนที่ซ่อนอยู่ [10]

3.4.2) Stochastic Gradient Descent (SGD) คืออัลกอริทึมที่มีการปรับปรุง (Update) ค่าพารามิเตอร์ในทุก ๆ ชุดข้อมูลฝึกฝน ซึ่งเป็นอัลกอริทึมที่สามารถประมวลผลได้ค่อนข้างไว จะมีการปรับปรุงแค่ครั้งเดียวต่อการฝึกฝน 1 รอบ

3.4.3) Logistic Regression (LR) คืออัลกอริทึมที่ใช้เทคนิคการวิเคราะห์ความสัมพันธ์ของตัวแปร เพื่อทำนายโอกาสความน่าจะเป็น (Probability) ที่จะเกิดเหตุการณ์ที่น่าสนใจ [6], [16]

3.4.4.) Support Vector Machines (SVM) คืออัลกอริทึมใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูลโดยอาศัยหลักการของการหาสมมติฐานของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกต้องเข้าสู่กระบวนการฝึกฝนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกและกลุ่มข้อมูล [6]

3.5) การประเมินตัวแบบ (Model Evaluation)

เมื่อผ่านขั้นตอนการแยกประเภท (Classification Phase) เรียบร้อยแล้ว จะนำผลการทำนายเข้ากระบวนการประเมินประสิทธิภาพของตัวแบบโดยพิจารณาจาก ความถูกต้อง (Accuracy) ดังสมการที่ 1 ความแม่นยำ (Precision) ดังสมการที่ 2 ค่าความระลึก (Recall) ดังสมการที่ 3 และค่าเฉลี่ย (F1 Score) ของ Precision และ Recall [13] ดังสมการที่ 4 ซึ่งผลการประเมินตัวแบบแสดงดังตารางที่ 4

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

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

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

- โดยที่ - TP คือ ข้อมูลที่ทำนายถูกต้องเมื่อเทียบกับเฉลย
- FP คือ ข้อมูลที่ทำนายแล้วไม่ถูกต้องเมื่อเทียบกับเฉลย
- FN คือ ข้อมูลที่อยู่ในเฉลยแต่ไม่มีการทำนาย (ตรงข้ามกับ FN)
- Precision คือ ค่าความแม่นยำ เกิดจากการนำค่า TP มาเทียบกับ FP
- Recall คือ ค่าความระลึก เกิดจากการนำค่า TP มาเทียบกับ FN
- F1 Score คือ ค่าเฉลี่ยของ Precision และ Recall

$$F_1 = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) \quad (4)$$

4) ผลการวิจัย

จากการทดลองสร้างตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์ โดยใช้ชุดข้อมูล จำนวน 12,900 ข้อมูล ซึ่งเป็นข้อมูลการแสดง

ความคิดเห็นต่อสินค้าและบริการภาษาไทยที่มีการติดป้ายกำกับ (Label) ความรู้สึก 3 ระดับ ได้แก่ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) ผ่านกระบวนการพัฒนาตัวแบบ 5 ขั้นตอน ได้แก่ 1) การเตรียมข้อมูล 2) การตัดคำ 3) การฝึกอบรมข้อมูล 4) ขั้นตอนการแยกประเภท 5) การประเมินตัวแบบ และมีการเปรียบเทียบผลการวิเคราะห์จากเทคนิคการจำแนกประเภท 4 รูปแบบ ได้แก่ Long Short-Term Memory (LSTM), Stochastic Gradient Descent (SGD), Logistic Regression (LR) และ Support Vector Machines (SVM) สามารถสรุปผลการทดลองได้ดังตารางที่ 5 ดังนี้

ตารางที่ 5 : แสดงการประเมินประสิทธิภาพของตัวแบบ

Classification	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
LSTM	81.27	81.27	81.27	81.27
LR	69.00	69.00	69.00	69.00
SGD	66.00	66.00	66.00	66.00
SVM	65.00	65.00	65.00	65.00

จากตารางที่ 5 แสดงการประเมินประสิทธิภาพของตัว จากผลการทดลอง เมื่อพิจารณาจากความถูกต้องสามารถสรุปได้ดังนี้

อันดับที่ 1 การใช้เทคนิคการจำแนกประเภทแบบ Long Short-Term Memory (LSTM) คะแนนความแม่นยำ (Precision) 81.27% ค่าความระลึก (Recall) 81.27% และค่าเฉลี่ย (F1 Score) 81.27% ตัวแบบให้ค่าความถูกต้องในการวิเคราะห์ผล (Accuracy) 81.27% ในการทดลองนี้ได้จำลองการทำงานของเทคนิคการสร้างตัวแบบความรู้สึกทางอารมณ์โดยใช้การจำแนกประเภทแบบ LSTM [17] ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 2 การใช้เทคนิคการจำแนกประเภทแบบ Logistic Regression (LR) คะแนนความแม่นยำ (Precision) 69% ค่าความระลึก (Recall) 69% และค่าเฉลี่ย (F1 Score) 69% ตัวแบบให้ค่าความถูกต้องในการวิเคราะห์ผล (Accuracy) 69% ในการทดลองนี้ได้จำลองการทำงานของเทคนิคการสร้างตัวแบบความรู้สึกทางอารมณ์โดยใช้การจำแนกประเภทแบบ Logistic Regression [6] ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 3 การใช้เทคนิคการจำแนกประเภทแบบ Stochastic Gradient Descent (SGD) คะแนนความแม่นยำ (Precision) 66% ค่าความระลึก (Recall) 66% และค่าเฉลี่ย

(F1 Score) 66% ตัวแบบให้ค่าความถูกต้องในการวิเคราะห์ผล (Accuracy) 66% ในการทดลองนี้ได้จำลองการทำงานของเทคนิคการสร้างตัวแบบความรู้สึกทางอารมณ์โดยใช้การจำแนกประเภทแบบ Stochastic Gradient Descent [18] ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 4 การใช้เทคนิคการจำแนกประเภทแบบ Support Vector Machines (SVM) คะแนนความแม่นยำ (Precision) 65% ค่าความระลึก (Recall) 65% และค่าเฉลี่ย (F1 Score) 65% ตัวแบบให้ค่าความถูกต้องในการวิเคราะห์ผล (Accuracy) 65% ในการทดลองนี้ได้จำลองการทำงานของเทคนิคการสร้างตัวแบบความรู้สึกทางอารมณ์โดยใช้การจำแนกประเภทแบบ Support Vector Machines [4], [6], [17], [19] ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

และเมื่อพิจารณาจากค่าความถูกต้องของตัวแบบ (Accuracy) โดยเปรียบเทียบผลการวิเคราะห์จากเทคนิคการจำแนกประเภท 4 รูปแบบ การสร้างตัวแบบโดยใช้เทคนิคการจำแนกประเภทแบบ LSTM ตัวแบบให้ค่าความถูกต้องมากที่สุด 81.27% เมื่อเทียบกับการจำแนกประเภทแบบ Logistic Regression (LR), Stochastic Gradient Descent (SDG) และ Support Vector Machines (SVM) ทำให้เห็นว่าการนำ Deep Learning มาใช้ในการจำแนกความรู้สึกจะช่วยปรับปรุงความถูกต้องของตัวแบบได้ดียิ่งขึ้น ซึ่งสอดคล้องกับการศึกษาของ [17] ที่มีการนำเสนอตัวแบบไฮบริดโดยใช้ Manhattan LSTM (MaLSTM) วิเคราะห์ข้อมูลการรีวิว (Review) ภาพยนตร์จาก IMDB ภาษาอังกฤษเพื่อเปรียบเทียบกับวิธีการจำแนกประเภทแบบ Recurrent Neural Networks (RNN), Support Vector Machines (SVM) และ [19] ที่มีการนำเสนอตัวแบบการวิเคราะห์ความรู้สึกในโพสต์ Twitter ภาษาไทย เพื่อเปรียบเทียบกับวิธีการจำแนกประเภทแบบ Support Vector Machines และ Naïve Bayes ซึ่งผู้วิจัยนำมาพัฒนาต่อยอดเพื่อสร้างตัวแบบความรู้สึกทางอารมณ์สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์ด้วยเทคนิคการเรียนรู้เชิงลึก (Deep Learning) บนพื้นฐานของโครงข่ายประสาทเทียมแบบเกิดซ้ำรูป (Neural Network) แบบหน่วยความจำระยะสั้น โดยมุ่งเน้นไปที่การเรียนรู้การเป็นตัวแทนที่ซ่อนอยู่ลึกหลายชั้น (hidden Layers) กับชุดข้อมูลที่เป็นภาษาไทย ร่วมกับการตัดคำ (Word Segmentation) โดยใช้ NewMM ที่พัฒนาเพื่อใช้กับภาษาไทย โดยเฉพาะเพื่อตัดคำที่เป็นประโยคยาว ๆ ออกมาเป็นแต่ละคำให้

คอมพิวเตอร์สามารถนำไปประมวลผลได้ และการสร้างคลังคำศัพท์ (Bag of Words) เพื่ออธิบายกลุ่มรวมของคำโดยไม่ได้คำนึงถึงหลักไวยากรณ์ ความถี่ที่พบ และลำดับของคำ โดยนำมาใช้เป็น Feature ในการแทนตัวจัดแบ่งข้อความ Classifier ซึ่งจะช่วยให้ตัวแบบสามารถทำนายผลความรู้สึกเชิงบวก เป็นกลาง และเชิงลบได้มีประสิทธิภาพมากกว่าการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ร่วมกับเทคนิคการจำแนกประเภท (Classification) ทั่ว ๆ ไป ที่ทำการศึกษา

5) สรุปและอภิปรายผล

ข้อมูลความคิดเห็นการเลือกซื้อสินค้าและบริการออนไลน์มีความหลากหลาย และมีข้อมูลใหม่เกิดขึ้นอยู่ตลอดเวลา ผู้บริโภคต้องเสียเวลาตีความการแสดงความคิดเห็นต่าง ๆ ที่มีอยู่เป็นจำนวนมาก ถ้าหากมีกลไกที่สามารถช่วยในการสรุปความคิดเห็นที่มีอยู่ในตัวสินค้าและบริการจะช่วยประหยัดเวลาและทำให้สามารถนำไปใช้ในตัดสินใจเลือกซื้อผลิตภัณฑ์ได้อย่างรวดเร็ว ดังนั้นงานวิจัยนี้นำเสนอวิธีการสร้างตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์ โดยใช้เทคนิคส่วนการเรียนรู้ของเครื่อง (Machine Learning) เพื่อตัดสินใจความคิดเห็นต่อสินค้าและบริการ โดยนำเสนอเทคนิคการตัดคำ (Word Segmentation) ที่มีการใช้ไลบรารี (Library) ที่พัฒนาเพื่อใช้กับภาษาไทย โดยเฉพาะ และการลบ Stop Words เพื่อลบคำที่เกิดซ้ำออกซึ่งยังช่วยให้ขนาดของข้อมูลที่จะนำไปฝึกฝนมีขนาดเล็กลงอีกด้วย อีกทั้งยังมีการใช้เทคนิคการสร้างคลังคำศัพท์ (Bag of Words) โดยมีการวิเคราะห์ถึงรูปแบบและเทคนิคที่เหมาะสมที่จะนำมาใช้กับความคิดเห็นที่เป็นภาษาไทย เพื่อช่วยกรองและแบ่งประเภทคำที่เป็นเชิงบวกและเชิงลบ รวมถึงคำกริยา คำนามต่าง ๆ ทำให้เกิดประสิทธิภาพในการตัดคำที่ดีขึ้น ซึ่งผลลัพธ์โดยรวมจะเห็นได้ว่า การใช้เทคนิคที่นำเสนอจะได้ผลลัพธ์ที่ดีกว่าเทคนิคที่ใช้โดยทั่วไป ขั้นตอนการทดลองสำหรับการทดสอบตัวแบบ มีการเปรียบเทียบเทคนิคการแยกประเภท 4 เทคนิค ได้แก่ Long Short-Term Memory (LSTM), Stochastic Gradient Descent (SGD), Logistic Regression (LR) และ Support Vector Machines (SVM) ซึ่งมีกระบวนการสร้างตัวแบบ 5 ขั้นตอน ได้แก่ 1) การเตรียมข้อมูล 2) การตัดคำ 3) การฝึกอบรมข้อมูล 4) ขั้นตอนการแยกประเภท และ 5) การประเมินตัวแบบ เพื่อสร้างตัวแบบในการแยกประเภทระดับความรู้สึกแบบอัตโนมัติ

ออกเป็น 3 ระดับคือ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) จากการทดลองสร้างตัวแบบการสร้างตัวแบบโดยใช้เทคนิคการจำแนกประเภทแบบ LSTM ซึ่งสูงกว่าเทคนิคทั้งหมดที่ทำการศึกษา และเหมาะสมที่สุดที่จะนำไปประยุกต์ใช้ในการสร้างตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์ที่เป็นภาษาไทย และจากกระบวนการสร้างตัวแบบที่นำเสนอข้างต้นยังสามารถนำไปประยุกต์ใช้ในการวิเคราะห์ความเห็นโดเมนอื่น ๆ ได้อีกด้วย เช่น ข้อมูลความคิดเห็นโรงแรมและที่พัก ข้อมูลความคิดเห็นการซื้อสินค้าและบริการทั่ว ๆ ไป และข้อมูลความคิดเห็นร้านอาหาร เป็นต้น

6) ข้อจำกัด

ข้อจำกัดของการศึกษาวิจัยในครั้งนี้ ได้แก่ ตัวแบบยังไม่สามารถวิเคราะห์คำประชดประชัน (Sarcasm) ได้เนื่องจากชุดข้อมูลที่นำมาใช้ยังไม่มีคำประเภดังกล่าว ซึ่งต้องอาศัยตัวแบบในการเรียนรู้แยกต่างหากและอาศัยเหมือนข้อความ (Text mining) แบบใช้กฎ (Rule-based) เข้ามาช่วย แต่ถ้าสามารถนำแนวทางของงานวิจัยมาเชื่อมต่อได้ในอนาคต

7) ข้อเสนอแนะ

สำหรับการวิจัยต่อ จะมีกระบวนการเพิ่มประสิทธิภาพของตัวแบบเพื่อให้ผลการทำนายมีประสิทธิภาพขึ้น โดยนำเทคนิคอื่น ๆ รวมในกระบวนการจัดเตรียมข้อมูล เช่น การเพิ่มจำนวนคำศัพท์ เพิ่มเทคนิคการแบ่งชนิดของคำ (Named-Entity Tagging) เพื่อช่วยให้การแยกคำมีความชัดเจนมากขึ้น รวมทั้งการเพิ่มชุดข้อมูลคำแสดงเพื่อทำให้ตัวแบบสามารถวิเคราะห์คำประชดประชันได้ และกลไกการเรียนรู้เป็นไปอย่างมีประสิทธิภาพ รวมถึงเพิ่มข้อมูลของคำ ประโยค ที่เป็นการคัดเลือกคำบอกความรู้สึก บวก ลบ และเป็นกลาง (Sentiment Word level) และสามารถนำเทคนิคการวิเคราะห์ระดับความรู้สึกด้วยการเรียนรู้เชิงลึก (Deep Learning) เข้ามาใช้เพื่อให้ผลลัพธ์การทำงานของตัวแบบดียิ่งขึ้น

กิตติกรรมประกาศ

งานศึกษานี้สำเร็จลุล่วงไปได้ ต้องขอขอบพระคุณผู้เชี่ยวชาญในการระบุและนิยามข้อมูลในแต่ละชุด

REFERENCES

- [1] P. Bowornlertsutee and W. Paireekreng, "The model of sentiment analysis for online shopping review using false detection technique," (in Thai), in *Proc. 14th Nat. Conf. Digit. Technol. Sustain. Wellbeing and Smart Soc.*, Bangkok, Thailand, Jun. 21, 2019, pp. 477–486.
- [2] P. Laowsungsuk, A. Jinda, and S. Sitthisarn, "Sentiment analysis of restaurant reviews on review web sites," (in Thai), *Thaksin Univ. J.*, vol. 20, no. 1, pp. 39–47, 2017.
- [3] M. Masdisornchote, "Constructing Thai opinion mining," (in Thai), *J. Sripatum Rev. Sci. Technol.*, vol. 6, no. 1, pp. 120–128, 2014.
- [4] P. Kasemmetheekarun, P. Meesut, and S. Nuanmeesri, "Sentiment analysis system of Thai video on social media using support vector machine," (in Thai), *CRMA J.*, vol. 14, no. 1, pp. 21–33, 2016.
- [5] B. Chiraratanasopha, "Sentimental analysis and keyword extraction from Thai users of Facebook in COVID-19 period," *Prog. Appl. Sci. Technol.*, vol. 11, no. 1, pp. 66–72, 2021.
- [6] P. Bowornlertsutee and W. Paireekreng, "Thai sentiment analysis techniques for online customer reviews," (in Thai), *RMUTSV Res. J.*, vol. 14, no. 3, 2022, Art. no. 245470.
- [7] M. Manhem, S. Dolah, S. Chunkaew, and S. Mak-on, "Model for the classification of feelings of reviews using techniques decision tree case studies hotel reservation web site," (in Thai), *J. Sci. Technol. Songkhla Rajabhat Univ.*, vol. 1, no. 2, pp. 69–79, 2020.
- [8] S. Srisawakarn and S. Srisuay, "Sentiment analysis opinion mining using text extraction," (in Thai), *J. Educational Technol. Commun. Fac. Educ. Mahasarakham Univ.*, vol. 6, no. 2, pp. 95–104, 2020.
- [9] X. Fan, X. Li, F. Du, X. Li, and M. Wei, "Apply word vectors for sentiment analysis of APP reviews," in *Proc. 3rd Int. Conf. Syst. and Inform. (ICSAI)*, Shanghai, China, Nov. 2016, pp. 1062–1066, doi: 10.1109/ICSAI.2016.7811108.
- [10] T. Chumwatana, "Using sentiment analysis technique for analyzing Thai customer satisfaction from social media," in *Proc. 5th Int. Conf. Comput. and Inform.*, Istanbul, Turkey, Aug. 2015, pp. 659–664.
- [11] N. Sornjapo, "Thai text automatic clustering model using machine learning by unsupervised learning with natural language processing," (in Thai), *Sripatum Chonburi J.*, vol. 14, no. 4, pp. 95–106, 2018.
- [12] C. Haruechaiyasak, A. Kongthon, P. Palingoon, and K. Trakultaweekoon, "S-Sense: A sentiment analysis framework for social media monitoring applications," *Inf. Technol. J.*, vol. 14, no. 1, pp. 11–22, 2018.
- [13] I. Lorattanachaiyong, "Generating Thai sentiment lexicon from online reviews," M.A. thesis, Dept. Arts, Chulalongkorn Univ., Bangkok, Thailand, 2017.
- [14] P. Promla and C. Sanrach, "The comparison of efficiency on the analysis of satisfaction on teaching performance using sentiment analysis by ensemble technique," *KKU Res. J. (Graduate Studies)*, vol. 20, no. 4, pp. 104–149, 2020.
- [15] W. Fongngen and S. Pethan, "Sentiment analysis opinion mining from Facebook comments using data mining technique," (in Thai), *J. Ind. Technol. Ubon Ratchathani Rajabhat Univ.*, vol. 11, no. 1, pp. 121–132, 2021.
- [16] O. Pravatborisut, *Python for data science data visualization and machine learning*. Bangkok, Thailand: Provision (in Thai), 2021.
- [17] S. Hiriyannaiah, S. G. Matt, and S. K. Gopalalyengar, "A hybrid model using MaLSTM based on recurrent neural networks with support vector machines for sentiment analysis," *Eng. Appl. Sci. Res.*, vol. 47, no. 3, pp. 232–240, 2020.
- [18] S. Diab, "Optimizing stochastic gradient descent in text classification base on fine- tuning hyper- parameters approach," *Int. J. Comput. Sci. Inf. Secur.*, vol. 16, no. 12, pp. 155–160, 2018.
- [19] X. Xie, K. Neelora, and S. Laohakiat, "Sentimental analysis using Twitter data," M.S. thesis, Dept. Comput. Sci., Srinakharinwirot Univ., Bangkok, Thailand, 2018.