

การวิเคราะห์ความรู้สึกทางอารมณ์ในบทความแนะนำสินค้าออนไลน์

Sentiment Analysis Techniques of Online Product Reviews

พิศิษฐ์ บวรเลิศสุธี และ วรภัทร ไพรีเกรง *

Pisit Bowornlertsutee and Worapat Paireekreng *

Received: 16 July 2020, Revised: 19 April 2021, Accepted: 1 October 2021

บทคัดย่อ

ในงานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอการสร้างตัวแบบสำหรับการวิเคราะห์ระดับความรู้สึกของการแสดงความคิดเห็นต่อสินค้าและบริการออนไลน์ โดยใช้เทคนิคส่วนการเรียนรู้ของเครื่อง (Machine Learning) ด้านการทำเหมืองข้อความ (Text Mining) ร่วมกับเทคนิคการตัดคำ (Word Segmentation) 3 รูปแบบ และการสร้างคลังคำศัพท์ (Bag of words) จากนั้นนำเข้ากระบวนการแยกประเภทผลการวิเคราะห์ 4 เทคนิค ได้แก่ K-Nearest Neighbors, Random Forest, Logistic Regression และ Support Vector Machines (SVM) ในการสร้างตัวแบบมีกระบวนการทดลอง 5 ขั้นตอน ได้แก่ 1) การเก็บรวบรวมและการเตรียมข้อมูล (Data Preparation Phase) 2) การประมวลผลข้อความ (Text File Processing Phase) 3) การฝึกอบรมข้อมูล (Training & Streaming Phase) 4) การแยกประเภท (Classification Phase) และ 5) การประเมินตัวแบบ (Model Evaluation Phase) ทำการทดลองโดยใช้ข้อมูลตัวอย่างการแสดงความคิดเห็นต่อสินค้าและบริการออนไลน์บน Social Blog จำนวน 252 บทความ การแสดงความคิดเห็นจำนวน 1,412 ความเห็น จำนวน 83,670 คำ ใช้เป็นข้อมูลในการสร้างตัวแบบ เพื่อช่วยให้ผู้บริโภคตัดสินใจในการซื้อสินค้าและบริการ และช่วยให้ผู้ประกอบการมีข้อมูลสำหรับการพัฒนาสินค้าและบริการในอนาคต ผลการทดลองสร้างตัวแบบการวิเคราะห์ระดับความรู้สึกทางอารมณ์ 3 ระดับ คือ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) โดยแต่ละตัวแบบให้ความแม่นยำในการทำนายผล ดังนี้ K-Nearest Neighbors 52.6% Random Forest 71.6% Logistic Regression 77.2% และ Support Vector Machines (SVM) 79% สรุป การสร้างตัวแบบใช้เทคนิคการแยกประเภทแบบ Support Vector Machines (SVM) ร่วมกับเทคนิคการตัดคำแบบ Longest ให้คะแนนความแม่นยำดีที่สุดเมื่อเปรียบเทียบกับเทคนิคอื่นๆ เหมาะแก่การนำไปใช้ในการสร้างตัวแบบในการวิเคราะห์ระดับความรู้สึกสำหรับข้อความที่เป็นภาษาไทย

* สาขาเทคโนโลยีสารสนเทศ วิทยาลัยนวัตกรรมการเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์ 110/1-4 ถนนประชาชื่นหลักสี่ กรุงเทพมหานคร 10210

* Information Technology, College of Innovative Technology and Engineering, Dhurakij Pundit University, 110/1-4 Pracha Chuen Road, Lak Si, Bangkok 10210, Thailand.

* ผู้รับผิดชอบประสานงาน ไปรษณีย์อิเล็กทรอนิกส์ (Corresponding author, e-mail): worapat.png@dpu.ac.th Tel: 0 2954 7300 ext. 602

คำสำคัญ: การวิเคราะห์อารมณ์ความรู้สึก, การทำเหมืองข้อความ, แนวทางการเรียนรู้ของเครื่อง, การซื้อสินค้าออนไลน์, การตัดคำภาษาไทย, การทำเหมืองความคิดเห็น

ABSTRACT

This research proposes a model for Sentiment analysis of emotional levels of users' opinions towards online products and services. The techniques used in this research are Machine Learning, Text Mining integrated with 3 types of Word Segmentation and Bag of Words. Four sentiment analysis methods: K-nearest neighbors Classifier, Random Forest Classifier, Logistic Regression and Support Vector Machines (SVM) were used for analysis. The model was developed through 5 steps including 1) Data Collection and Preparation Phase, 2) Text Processing Phase, 3) Training & Streaming Phase, 4) Classification Phase, and 5) Model Evaluation Phase. Consumers' opinions were gathered from social blogs related to online products and services. 252 articles with 1,412 comments were collected making up 83,670 words in the dataset. The model can help consumers make a decision for purchasing of goods and services, and help entrepreneurs gain the information to improve products and services in the future. This proposed method classified opinions into 3 scales: positive, neutral and negative opinions. Each method offered different levels of forecast ability: the K-Nearest Neighbors revealed the accuracy index at 56.2%, Random Forest Classifier at 71.6%, Logistic Regression at 77.2% and Support Vector Machines (SVM) at 79%. However, the combination of SVM and Longest Word Segmentation performed better compared to other techniques and is more appropriate as part of sentimental analysis model for Thai reviews and comments.

Keywords: sentiment analysis, text mining, machine learning-based approach, online shopping, Thai word segmentation, opinion mining

บทนำ

ปัจจุบันการซื้อขายสินค้าและบริการผ่านช่องทางออนไลน์หรือ อีคอมเมิร์ซ (e-Commerce) ในประเทศไทย มีสถิติการใช้อินเทอร์เน็ตเพิ่มขึ้นจากยอด 9.3 ล้านคนในปี พ.ศ. 2551 เป็น 45 ล้านคนในปี พ.ศ. 2560 จากยอดจดทะเบียนโทรศัพท์มือถือกว่า 124 ล้านเลขหมาย นอกจากนี้พบว่า เวลาการใช้อินเทอร์เน็ตโดยเฉลี่ยคิดเป็นเวลา 10.05 ชั่วโมง/วัน/คน ข้อมูลเหล่านี้สะท้อนให้เห็นถึงการเข้าถึงและพฤติกรรมการใช้อินเทอร์เน็ตของคนไทยซึ่งเปลี่ยนไปตามยุคของเทคโนโลยีดิจิทัลที่ได้เข้ามา

บทบาทในชีวิตประจำวันมากขึ้น ซึ่งเป็นปัจจัยสำคัญที่ส่งผลให้มูลค่าอีคอมเมิร์ซในประเทศไทยมีการเติบโตขึ้นอย่างต่อเนื่อง โดยเฉลี่ยเพิ่มขึ้น 10% ต่อปี และการซื้อขายสินค้าและบริการออนไลน์เป็น 1 ใน 5 กิจกรรมยอดนิยมที่มีการใช้งานบนอินเทอร์เน็ตมากที่สุด (Electronic Transactions Development Agency (ETDA), 2019) ซึ่งในประเทศไทยมีเว็บไซต์ที่เปิดให้บริการในการซื้อขายสินค้าออนไลน์อยู่เป็นจำนวนมาก ในการตัดสินใจซื้อสินค้าและบริการ ผู้บริโภคจำนวนมากจะทำการสืบค้นข้อมูลต่างๆ เกี่ยวกับตัวสินค้าและบริการ เช่น การสืบค้น

ข้อคิดเห็นต่อผลิตภัณฑ์ ข้อคิดเห็นต่อสินค้าและบริการ เป็นต้น ซึ่งเป็นช่องทางให้ผู้ที่มีสนใจใช้เป็นแนวทางในการตัดสินใจซื้อสินค้าและบริการ ซึ่งการแสดงความคิดเห็นดังกล่าวเป็นการแสดงถึงความรู้สึกทางอารมณ์ในรูปแบบข้อความที่มีต่อสินค้าและบริการ อีกทั้งข้อมูลการแสดงความคิดเห็นในปัจจุบันนี้มีปริมาณมหาศาลและเกิดขึ้นใหม่อยู่ตลอดเวลา ซึ่งทำให้ผู้บริโภคต้องใช้เวลาในการวิเคราะห์ข้อมูลเหล่านี้เพื่อใช้ในการตัดสินใจซื้อสินค้าและบริการมากขึ้น เนื่องจากยังไม่มีตัวช่วยในการวิเคราะห์คำและข้อความจากการแสดงความคิดเห็นต่อสินค้าและบริการแบบอัตโนมัติที่ช่วยในการตัดสินใจการเลือกซื้อสินค้าและบริการได้อย่างรวดเร็ว

การวิเคราะห์ความรู้สึกต่อความคิดเห็น (Sentiment Analysis) คือ กระบวนการวิเคราะห์และประเมินความรู้สึกต่อการแสดงความคิดเห็นบนสื่อสังคมออนไลน์ (Social Network) เช่น twitter.com, facebook.com, Instagram.com หรือ pantip.com เป็นต้น โดยมีเป้าหมายเพื่อแสดงทัศนคติทางอารมณ์ที่มีต่อข้อมูล บทวิจารณ์ โพสต์ กระทู้ ที่เกี่ยวข้องกับ การตลาด การประชาสัมพันธ์ การแสดงความคิดเห็นต่อผลิตภัณฑ์และบริการของลูกค้าบนโซเชียลมีเดีย (Sepandar and Jonathan, 2011) จากงานวิจัยที่ผ่านมา การวิเคราะห์ระดับความรู้สึก (Sentiment Analysis) ใช้กระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) และการทำเหมืองข้อความ (Text Mining) เป็นส่วนใหญ่ (Trupthi et al., 2017; Anurag and Vijay, 2015) อีกทั้งยังมีการใช้เทคนิค Bag-of-Words (CBOW) และ Skip-gram Model ร่วมด้วยเพื่อสร้างคลังคำศัพท์ในการวิเคราะห์ความรู้สึกที่เป็นเชิงบวกและเชิงลบ (Fan et al., 2016; Chumwatana, 2015) ในกรณีชุดข้อมูลที่เป็นภาษาไทย จะมีข้อจำกัดคือ การแสดงความคิดเห็นที่เป็น

ภาษาไทยจะเป็นข้อความที่ไม่มีการแบ่งส่วน แบ่งคำ ประโยคติดกัน อักษรที่ไม่มีขอบเขตคำ (Chumwatana, 2015; Netisopakul and Thong-iad, 2018) อีกทั้งคำศัพท์ในภาษาไทยหนึ่งคำยังสามารถสื่อความรู้สึกได้มากกว่าหนึ่งคำ (Netisopakul and Thong-iad, 2018) และในปัจจุบันการวิจัยเรื่องดังกล่าวในภาษาไทยยังมีไม่มากนัก อีกทั้งการประมวลผลภาษาไทยยังมีความยุ่งยากกว่าภาษาอังกฤษ เพราะโครงสร้างประโยคและไวยากรณ์มีความซับซ้อนมากกว่า (Somjapo, 2019) และขาดแคลนแหล่งข้อมูลและเครื่องมือต่างๆ เช่น โปรแกรมแท็ก (Tag) ชนิดคำ เป็นต้น รวมไปถึงแนวโน้มของการเลือกใช้แบบจำลองในการวิเคราะห์ข้อมูล ซึ่งในปัจจุบันวิธีการเรียนรู้ด้วยเครื่อง (Machine Learning) เป็นที่นิยมมากกว่า โดยอาศัยการสร้างชุดข้อมูลฝึกฝน (Training Set) (Lorattanachaiyong, 2017) และนำชุดข้อมูลทดสอบ (Test Set) เข้ากระบวนการจัดกลุ่มหรือแยกประเภท (Classification) และการประเมินคะแนนความแม่นยำของตัวแบบที่ได้โดยสรุปจากค่าความแม่นยำจาก Precision และ Recall นอกจากนี้ การทำการวิเคราะห์ระดับความรู้สึกทางอารมณ์ส่วนใหญ่จะเป็นการตัดสินใจแบบ 2 ทิศทาง คือ แบบบวกและแบบลบ อย่างไรก็ตาม จะมีความคิดเห็นที่เป็นกลาง ซึ่งถ้าหากถูกตัดสินเพียง 2 ระดับ อาจจะทำให้ข้อคิดเห็นโดยรวมเกิดอคติได้ ดังนั้น ถ้าหากมีการวิเคราะห์เป็นแบบ 3 ระดับ คือ มีระดับกลางเพิ่มขึ้นมา จะทำให้น้ำหนักของการตัดสินใจแบบบวกและแบบลบ มีความถูกต้องมากขึ้น

จากปัญหาดังกล่าวผู้วิจัยจึงมีแนวคิดที่จะสร้างตัวแบบการวิเคราะห์ระดับความรู้สึกทางอารมณ์สำหรับบทความแนะนำสินค้าออนไลน์ (Sentiment Analysis Model) โดยใช้เทคนิคส่วนการเรียนรู้ของเครื่อง (Machine Learning) แบบมีผู้สอน (Supervised Learning) ด้านการทำเหมืองข้อความ

(Text Mining) ร่วมกับเทคนิคการตัดคำ (Word Segmentation) 3 เทคนิค และการสร้างคลังคำศัพท์ (Bag of Words) เพื่อให้ตัวแบบสามารถแก้ไขปัญหาในการแบ่งแยกคำภาษาไทยและวิเคราะห์คำจากการแสดงความคิดเห็นได้อย่างมีประสิทธิภาพ นอกจากนั้นจะมีการใช้เทคนิคการแยกประเภทโดยมีการเปรียบเทียบผลการวิเคราะห์ 4 เทคนิค ได้แก่ K-Nearest Neighbors, Random Forest, Logistic Regression และ Support Vector Machines (SVM) ซึ่ง 4 เทคนิคดังกล่าวสามารถแยกประเภทได้อย่างแม่นยำ ทำนายผลได้สูง น่าเชื่อถือ และมีการใช้กันอย่างแพร่หลายในงานดังกล่าว (Haseena and Ahmad, 2014) ในการวิเคราะห์ประโยคเนื้อหาที่เป็นภาษาไทยซึ่งสามารถแยกประเภทของระดับความรู้สึกทางอารมณ์ได้ 3 ระดับ ได้แก่ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) เพื่อหาตัวแบบที่ดีที่สุดที่สามารถใช้เป็นแนวทางในการพัฒนาตัวแบบการวิเคราะห์ระดับความรู้สึกทางอารมณ์สำหรับบทความแนะนำสินค้าออนไลน์ภาษาไทย โดยทั่วไปเทคนิคการตัดคำ (Word Segmentation) มักจะใช้อัลกอริทึมแบบที่แยกเป็นคำ ซึ่งไม่เหมาะกับการเขียนที่เป็นประโยคยาวต่อเนื่อง เช่น ภาษาไทย ดังนั้นในงานวิจัยนี้จะใช้การตัดคำรูปแบบ Longest Engine ซึ่งเป็นอัลกอริทึมที่สามารถแยกคำภาษาไทยที่ยาวที่สุดและจับคู่กับการใช้พจนานุกรม (Dictionary-based) ร่วมกับการสร้างคลังคำศัพท์ (Bag of Words) เพื่อให้การตัดคำมีความเหมาะสม ได้ตัวแบบที่มีประสิทธิภาพดียิ่งขึ้น นอกจากนี้ตัวแบบยังสามารถระบุความคิดเห็นแบบเป็นกลาง (Neutral) ได้ และทำให้การตัดสินใจความรู้สึกเชิงบวก (Positive) และเชิงลบ (Negative) มีความถูกต้องมากยิ่งขึ้น

วิธีดำเนินการวิจัย

การออกแบบการทดลอง (Experimental Design) ประกอบด้วย 5 ขั้นตอน โดยแต่ละขั้นตอนใช้เครื่องมือการพัฒนาโปรแกรมภาษาไพทอน (Python) ร่วมกับดูล (Module) ต่างๆ โดยมีรายละเอียดดังนี้

1) การเก็บรวบรวมและการเตรียมข้อมูล (Data Preparation Phase) รวบรวมข้อมูลจากสื่อสังคมออนไลน์แบบบล็อก โดยใช้ตัวอย่างข้อมูลบทความแนะนำสินค้าและบริการที่เป็นภาษาไทยบนเว็บไซต์ www.blognone.com จำนวน 252 บทความ 1,412 ความเห็น และจำนวน 83,670 คำ ช่วงเวลาที่เก็บข้อมูล เดือนมกราคม พ.ศ. 2562 ถึง เดือนเมษายน พ.ศ. 2562 เนื่องจากเว็บไซต์ดังกล่าวเป็นเว็บไซต์ที่มีผู้เข้าชมเป็นจำนวนมาก มีข้อมูลและบทความแนะนำสินค้าทางด้านเทคโนโลยีที่หลากหลาย มีข้อมูลและบทความใหม่อยู่ตลอดเวลา และสามารถแสดงความคิดเห็นภายในบทความได้ ต่อมาในขั้นตอนการดึงข้อมูลการแสดงความคิดเห็นจะใช้เทคนิค Web Scraping สามารถแสดงขั้นตอนการทำงานได้ดังภาพที่ 1 จากนั้นนำชุดข้อมูลเข้าสู่กระบวนการทำความสะอาดข้อมูล (Data Cleaning) โดยใช้ Regular Expression เพื่อแก้ไข ปรับแต่ง และตัดส่วนที่ไม่ต้องการออกจากชุดข้อมูล เช่น HTML Tag, URL หรือเครื่องหมายคำพูด (Punctuation) เป็นต้น เพื่อให้ข้อมูลมีคุณภาพ มีความถูกต้องเหมาะสมที่จะนำไปใช้วิเคราะห์และสร้างตัวแบบสำหรับการเรียนรู้ด้วยเครื่องจักร (Machine Learning) ต่อไป



ภาพที่ 1 แสดงขั้นตอนการดึงข้อมูลจากเว็บเพจ

จากภาพที่ 1 แสดงขั้นตอนการดึงข้อมูลการแสดงความคิดเห็นจากเว็บเพจ (Web Scraping) โดยใช้ภาษาไพทอน (Python) และไลบรารี (Libraries) BeautifulSoup มี 3 ขั้นตอนดังนี้ ขั้นตอนที่ 1 ทำการหา Webpage ตามที่ต้องการ โดยคัดเลือกเว็บเพจที่มีการแสดงความคิดเห็น 3 รายการขึ้นไป ขั้นตอนที่ 2 กำหนด URL และ Class เพื่อดึงข้อมูลการแสดงความคิดเห็น และในขั้นตอนที่ 3 เมื่อได้ข้อมูลแล้วจะทำการบันทึกข้อมูลเก็บในรูปแบบ CSV File เพื่อนำข้อมูลไปทำตามสะดวกในขั้นตอนต่อไป

2) การประมวลผลข้อความ (Text File Processing Phase) ทำการประมวลผลข้อความจากชุดข้อมูลที่เตรียมไว้สำหรับฝึกฝนข้อมูล (Training Set) ให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ประเภทของระดับความรู้สึก กระบวนการสร้าง Sentiment Analysis จะมีการนำข้อความมาแบ่งคำ แล้วหาความสัมพันธ์ระหว่าง "คำ" และ "เอกสาร" ที่มีอยู่ในรูปแบบของเวกเตอร์ (Vector) ซึ่งจะต้องใช้คลังข้อมูลความรู้สึกดี (Positive) และความรู้สึกที่ไม่ดี (Negative) และเป็นกลาง (Neutral) ในภาษาไทย โดยใช้ Library จาก PyThaiNLP (Python library for Thai Natural Language Processing : PyThaiNLP) เพื่อวิเคราะห์ระดับความรู้สึก 3 ระดับได้แก่ ความรู้สึกเชิงบวก เป็นกลาง และเชิงลบ โดยนำเข้ากระบวนการตัดคำ (Word Segmentation) โดยใช้ 3 อัลกอริทึมโดยมีคุณสมบัติดังนี้

1) NewMM Engine ซึ่งเป็นอัลกอริทึมจาก PyThaiNLP พัฒนาจากเทคนิคการแบ่งส่วนคำศัพท์ภาษาไทยด้วยพจนานุกรม (Dictionary-based) โดยใช้อัลกอริทึมการจับคู่สูงสุดและการจัดกลุ่ม (Cluster) อักษรภาษาไทย (TCC) ถูกพัฒนาขึ้นโดย Korakot Chaovavanich ในปี 2018 (PyThaiNLP, 2018)

2) Longest Engine ซึ่งเป็นอัลกอริทึมพัฒนาจากเทคนิคการแยกคำภาษาไทยที่ยาวที่สุดที่จับคู่กับพจนานุกรม (Dictionary-based) ถูกพัฒนาขึ้นโดย Patorn Utenpattannun ในปี 2018 (PyThaiNLP, 2018)

3) AttaCut Engine ซึ่งเป็นอัลกอริทึมที่สามารถตัดคำภาษาไทยได้อย่างรวดเร็วและแม่นยำ โดยใช้ อัลกอริทึมการจัดกลุ่ม (Clustering) อักษรภาษาไทย (Rules purposed) ถูกพัฒนาขึ้นโดย Pattarawat Chormai ในปี 2019 (Chormai *et al.*, 2019)

ในการเลือกใช้การตัดคำ (Word Segmentation) ทั้ง 3 แบบที่มีคุณสมบัติแตกต่างกัน ทั้งนี้เพื่อเป็นการเพิ่มประสิทธิภาพในการนำไปวิเคราะห์การแยกประเภทเพื่อให้สามารถวิเคราะห์คำที่เป็นภาษาไทยได้อย่างถูกต้องและรวดเร็ว อีกทั้งยังเป็นการเพิ่มประสิทธิภาพในการสร้างชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบที่จะนำไปใช้ในการสร้างตัวแบบให้มีความแม่นยำมากยิ่งขึ้นอีกด้วย ในขั้นตอนต่อไปจะนำเข้ากระบวนการสร้างคลังคำศัพท์ (Bag of Words) ซึ่งเป็นโมเดลที่ใช้ในงานจัดแบ่งประเภทข้อความ (Text Classification) กลุ่มของคำจะถูกอธิบายด้วยคลังคำศัพท์หรือการกลุ่มรวมของคำ โดยไม่ได้

คำนึงถึงหลักไวยากรณ์ ความถี่ที่พบ และลำดับของคำ ซึ่งหากมีคำเกิดขึ้นในข้อความที่นำมาวิเคราะห์จะแทนค่าด้วย 1 แต่ถ้าในข้อความไม่มีคำที่กำหนดไว้จะแทนค่าด้วย 0 โดยนำมาใช้เป็นคุณสมบัติ (Feature) ในการฝึกอบรม (Training) ตัวจัดแบ่งข้อความ (Classifier) จากนั้นทำการประมวลผลให้แต่ละข้อมูลติดป้ายกำกับ

(Label) ระดับของอารมณ์ 3 ระดับ และแปลงชุดข้อมูลให้อยู่ในรูปแบบไฟล์ CSV เพื่อนำไปใช้เป็นชุดข้อมูลสำหรับฝึกอบรมแบบมีผู้สอน (Supervised Learning) ต่อไป ซึ่งผลการวิเคราะห์สามารถแสดงในตารางที่ 1 ได้ดังนี้

ตารางที่ 1 แสดงผลการติดป้ายกำกับ (Label) ระดับของอารมณ์ 3 ระดับ

ระดับความรู้สึก/อารมณ์	เปอร์เซ็นต์
เชิงบวก (Positive)	20%
เป็นกลาง (Neutral)	44%
เชิงลบ (Negative)	36%

จากตารางที่ 1 แสดงผลการวิเคราะห์ระดับความรู้สึกจากการประมวลผลข้อความและการติดป้ายกำกับ (Label) ระดับของอารมณ์ 3 ระดับจากชุดข้อมูลฝึกฝน (Training Set) เมื่อพิจารณาจากจำนวนข้อมูลพบว่า กลุ่มข้อมูลความรู้สึกเป็นกลาง (Neutral)

มีจำนวนข้อมูลมากที่สุด 44% กลุ่มข้อมูลความรู้สึกเชิงลบ (Negative) 36% และกลุ่มข้อมูลความรู้สึกเชิงบวก (Positive) 20% ซึ่งสามารถแสดงตัวอย่างการแสดงความเห็นแยกตามระดับความรู้สึกจากชุดข้อมูล 3 ระดับ ได้ดังตารางที่ 2

ตารางที่ 2 การแสดงความเห็นแยกตามระดับความรู้สึก

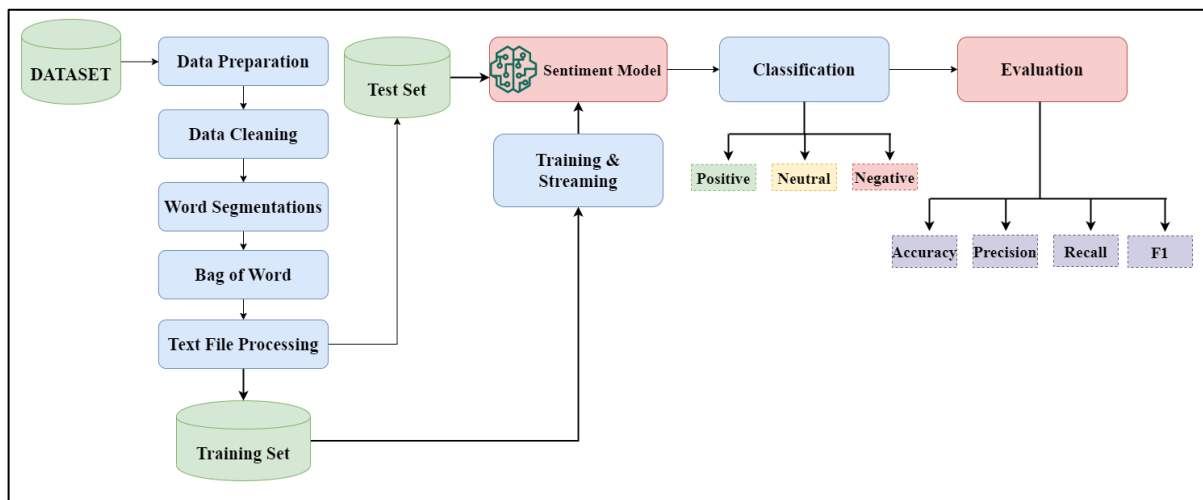
การแสดงความเห็น	ระดับความรู้สึก
มาราคานี้ ดูน่าซื้อขึ้นมาเลย	เชิงบวก
ดีไซน์คุ้นมากๆ	เป็นกลาง
เบตไม่ถึงวัน กล้องไม่มีกันสั่น	เชิงลบ

3) ขั้นตอนการฝึกอบรมข้อมูล (Training & Streaming Phase) ในขั้นตอนการสร้างตัวแบบสำหรับการเรียนรู้ (Supervised Learning) ใช้คุณสมบัติจากชุดข้อมูล (Feature) ทั้งหมด 3 คอลัมน์ เพื่อใช้ในการฝึกฝน (Training) และทดสอบ (Testing) ได้แก่ title tokens, title length, label และแบ่งชุดข้อมูลออกเป็น 2 ชุด (Split Test) คือ ชุดข้อมูลฝึกฝน (Training Set) 80% ในการสร้างตัวแบบการฝึกอบรมข้อมูล จากนั้นทำการนำชุดข้อมูลทดสอบ (Testing Set) 20% เข้า

กระบวนการเพื่อวิเคราะห์ระดับความรู้สึกที่มีต่อการแสดงความเห็นต่อสินค้าและบริการที่เป็นภาษาไทย ซึ่งกระบวนการทดสอบกับชุดข้อมูลจะใช้เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นการเรียนรู้โดยอาศัยข้อมูลที่ป้อนเข้าไปเก็บไว้เป็นตัวอย่างเพื่อให้คอมพิวเตอร์ใช้ในการเปรียบเทียบกับข้อมูลใหม่ที่เข้ามาแล้วทำนายหรือจัดหมวดหมู่ที่มีความใกล้เคียงกันมากที่สุดให้อยู่ในกลุ่มเดียวกัน

4) ขั้นตอนการแยกประเภท (Classification Phase) คือกระบวนการสกัดความคิดเห็นจากการเรียนรู้ของเครื่อง (Machine Learning) แบบมีผู้สอน (Supervise Learning) โดยทดสอบกับชุดข้อมูล (Testing Set) 20% นำเข้ากระบวนการทำนายผล (Prediction) และแยกผลการวิเคราะห์ระดับความรู้สึก (Sentiment Analysis) โดยใช้เทคนิคการแยกประเภทผลการวิเคราะห์ 4 เทคนิค ได้แก่ K-Nearest Neighbors,

Random Forest, Logistic Regression และ Support Vector Machines ในกระบวนการสร้างตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับบทความแนะนำสินค้าออนไลน์ภาษาไทย โดยใช้เทคนิคส่วนการเรียนรู้ของเครื่อง (Machine Learning) ด้านการทำเหมืองข้อความ (Text Mining) ร่วมกับเทคนิคการตัดคำ (Word Segmentation) และการสร้างคลังคำศัพท์ (Bag of words) สามารถแสดงได้ดังภาพที่ 2



ภาพที่ 2 แสดงขั้นตอนกระบวนการสร้างตัวแบบ

จากภาพที่ 2 แสดงขั้นตอนกระบวนการสร้างตัวแบบสำหรับวิเคราะห์ระดับอารมณ์ความรู้สึก โดยเริ่มจากการจัดเตรียมข้อมูลให้พร้อมใช้งาน ตั้งแต่กระบวนการทำความสะอาดข้อมูล การตัดคำ การสร้างคลังคำศัพท์ การจัดชุดข้อมูลให้อยู่ในรูปแบบ text file และแบ่งชุดข้อมูลออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลฝึกฝน (Training Set) 80% และข้อมูลทดสอบ (Testing Set) 20% จากนั้นนำชุดข้อมูลฝึกฝนเข้ากระบวนการฝึกอบรมข้อมูล ขั้นตอนต่อมาคือการนำชุดข้อมูลสำหรับทดสอบเข้ากระบวนการทำนายและแยกประเภท (Classification) ของความรู้สึกอารมณ์ 3

ระดับ ได้แก่ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative)

5) การประเมินตัวแบบ (Model Evaluation Phase) เมื่อผ่านขั้นตอนการแยกประเภท (Classification Phase) เรียบร้อยแล้ว จะนำผลการทำนายเข้ากระบวนการประเมินตัวแบบเพื่อวัดความแม่นยำของตัวแบบโดยพิจารณาจาก ความแม่นยำ (Accuracy) ดังสมการที่ 1 ความถูกต้อง (Precision) ดังสมการที่ 2 ค่าความระลึก (Recall) ดังสมการที่ 3 และค่าเฉลี่ย (F1 Score) ของ Precision และ Recall ดังสมการที่ 4 (Limkonchotiwat, 2018; Promla and Sanrach, 2020)

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

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

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

โดยที่

TP คือ ข้อมูลที่ทำนายถูกต้องเมื่อเทียบกับเฉลย

FT คือ ข้อมูลที่อยู่ในเฉลยแต่ไม่มีการทำนาย
(ตรงข้ามกับ FN)

FN คือ ข้อมูลที่ทำนายแล้วไม่ถูกต้องเมื่อเทียบกับเฉลย

Precision คือ ค่าความแม่นยำ เกิดจากการนำค่า

TP มาเทียบกับ FP

Recall คือ ความถูกต้อง เกิดจากการนำค่า TP

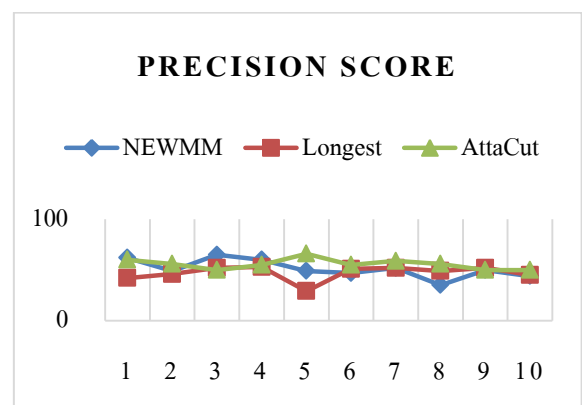
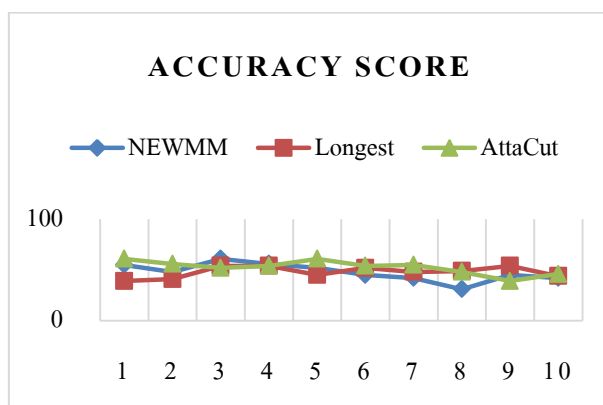
มาเทียบกับ FN

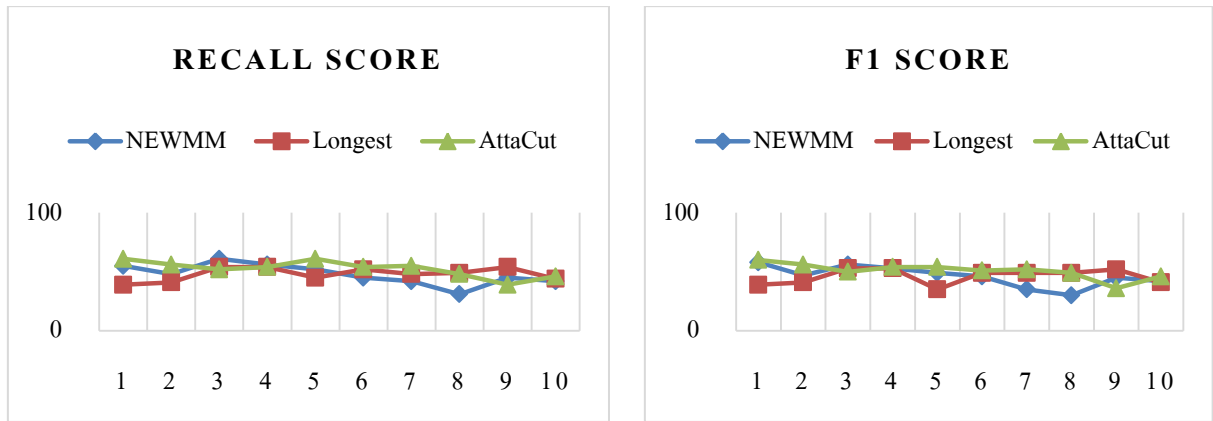
F1 Score คือ ค่าเฉลี่ยของ Precision และ Recall

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

ในลำดับถัดมาจะแสดงผลคะแนนการประเมินตัวแบบ โดยเปรียบเทียบคะแนนความแม่นยำ (Accuracy) ความถูกต้อง (Precision) ค่าความระลึก (Recall) และค่าเฉลี่ย (F1 Score) จากการตัดคำ

ภาษาไทย (Word Segmentation) ทั้ง 3 รูปแบบที่มีคุณสมบัติแตกต่างกัน สามารถแสดงดังภาพที่ 3 ถึงภาพที่ 6 ดังนี้

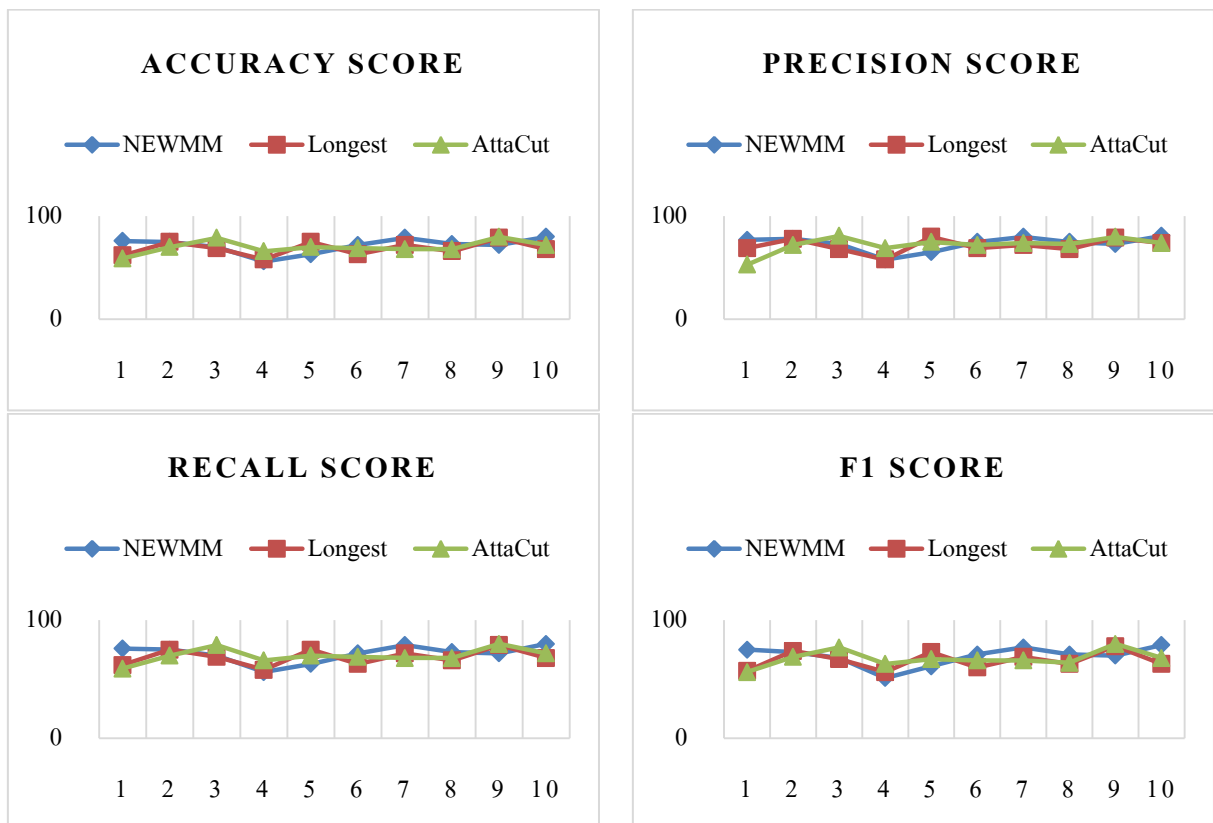




ภาพที่ 3 แสดงผลการประเมินความแม่นยำของตัวแบบโดยใช้เทคนิคการแยกประเภทแบบ K-Nearest Neighbors

จากภาพที่ 3 แสดงผลจากการประเมินความแม่นยำของตัวแบบ โดยใช้เทคนิคการตัดคำ (Word Segmentation) 3 รูปแบบ ร่วมกับเทคนิคกระบวนการสร้างคลังคำศัพท์ (Bag of Words) และการแยกประเภทแบบ K-Nearest Neighbors Classifier ทำการประมวลผลการประเมินตัวแบบแยกตามเทคนิคการตัดคำ 3 เทคนิค เทคนิคละ 10 รอบ ทั้งนี้

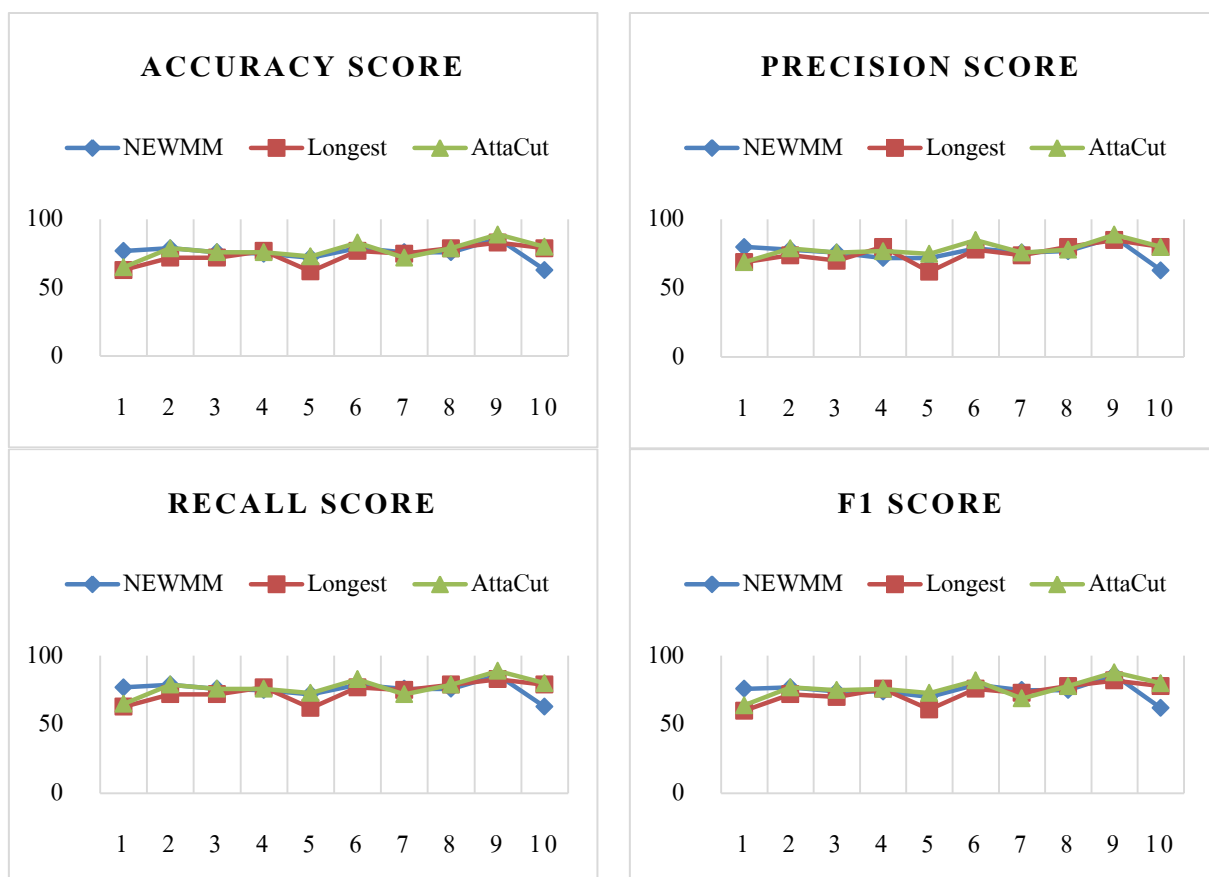
เพื่อเพิ่มความเชื่อมั่นและลดอคติในการประมวลผล เมื่อพิจารณาค่าเฉลี่ยคะแนนความแม่นยำสรุปได้คือ การใช้เทคนิคการตัดคำแบบ AttaCut ให้ความแม่นยำในการทำนายผลมากที่สุด คือ คะแนนความแม่นยำ (Precision) 55.7% คะแนนความถูกต้อง (Recall) 52.6% และค่าเฉลี่ย (F1 Score) 50.1% ตัวแบบมีความแม่นยำในการทำนายผล (Accuracy) 52.6%



ภาพที่ 4 แสดงผลการประเมินความแม่นยำของตัวแบบ โดยใช้เทคนิคการแยกประเภทแบบ Random Forest

จากภาพที่ 4 แสดงผลจากการประเมินความแม่นยำของตัวแบบ โดยใช้เทคนิคการตัดคำ (Word Segmentation) 3 รูปแบบ ร่วมกับเทคนิคกระบวนการสร้างคลังคำศัพท์ (Bag of Words) และการแยกประเภทแบบ Random Forest Classifier ทำการประมวลผลการประเมินตัวแบบแยกตามเทคนิคการตัดคำ 3 เทคนิค เทคนิคละ 10 รอบ ทั้งนี้เพื่อเพิ่ม

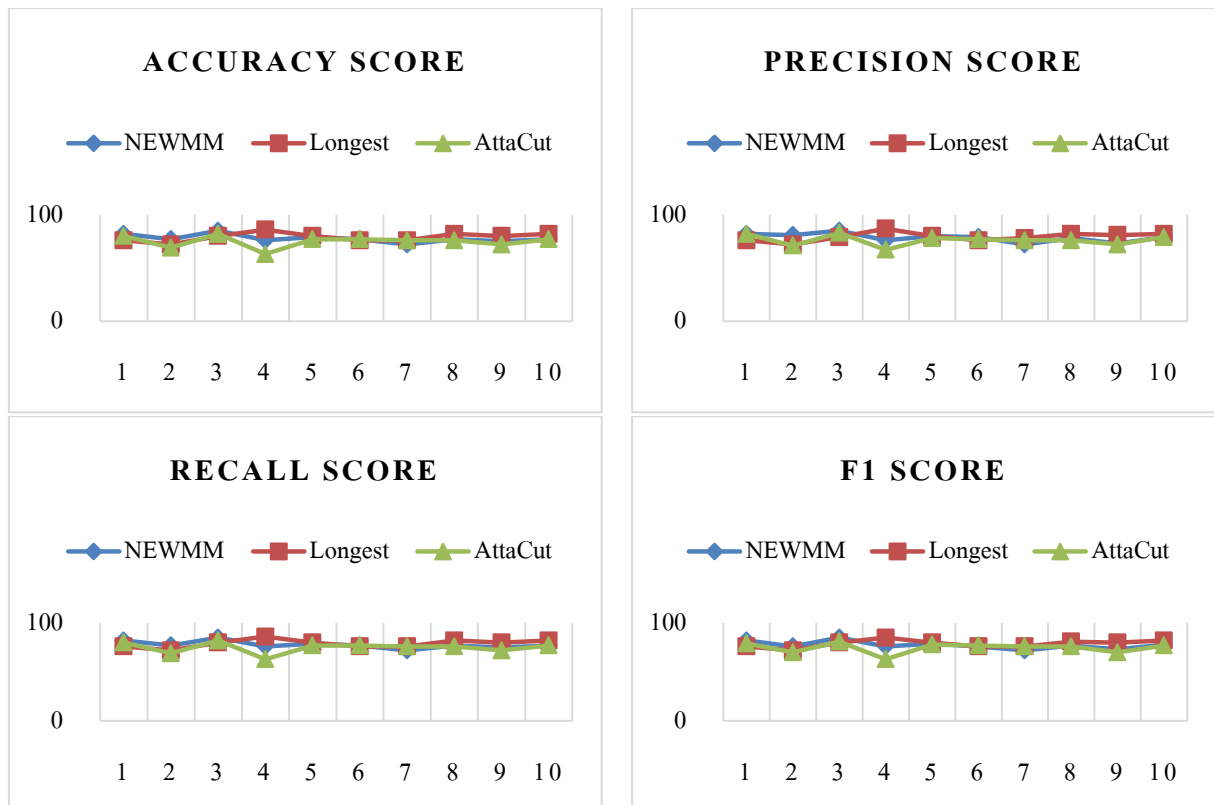
ความเชื่อมั่นและลดอคติในการประมวลผล เมื่อพิจารณาค่าเฉลี่ยคะแนนความแม่นยำสรุปได้คือ การใช้เทคนิคการตัดคำแบบ NewMM ให้ความแม่นยำในการทำนายผลมากที่สุด คือ คะแนนความแม่นยำ (Precision) 73.5% คะแนนความถูกต้อง (Recall) 71.6% และค่าเฉลี่ย (F1 Score) 69.6% ตัวแบบมีความแม่นยำในการทำนายผล (Accuracy) 71.6%



ภาพที่ 5 แสดงผลการประเมินความแม่นยำของตัวแบบ โดยใช้เทคนิคการแยกประเภทแบบ Logistic Regression

จากภาพที่ 5 แสดงผลจากการประเมินความแม่นยำของตัวแบบ โดยใช้เทคนิคการตัดคำ (Word Segmentation) 3 รูปแบบ ร่วมกับเทคนิคกระบวนการสร้างคลังคำศัพท์ (Bag of Words) และการแยกประเภทแบบ Logistic Regression Classifier ทำการประมวลผลการประเมินตัวแบบแยกตามเทคนิคการตัดคำ 3 เทคนิค เทคนิคละ 10 รอบ ทั้งนี้

เพื่อเพิ่มความเชื่อมั่นและลดอคติในการประมวลผล เมื่อพิจารณาค่าเฉลี่ยคะแนนความแม่นยำสรุปได้คือ การใช้เทคนิคการตัดคำแบบ AttaCut ให้ความแม่นยำในการทำนายผลมากที่สุด คือ คะแนนความแม่นยำ (Precision) 78.4% คะแนนความถูกต้อง (Recall) 77.2% และค่าเฉลี่ย (F1 Score) 76.2% ตัวแบบมีความแม่นยำในการทำนายผล (Accuracy) 77.2%



ภาพที่ 6 แสดงผลการประเมินความแม่นยำของตัวแบบ โดยใช้เทคนิคการแยกประเภทแบบ SVM

จากภาพที่ 6 แสดงผลจากการประเมินความแม่นยำของตัวแบบ โดยใช้เทคนิคการตัดคำ (Word Segmentation) 3 รูปแบบ ร่วมกับเทคนิคกระบวนการสร้างคลังคำศัพท์ (Bag of Words) และการแยกประเภทแบบ Support Vector Machines (SVM) Classifier ทำการประมวลผลการประเมินตัวแบบแยกตามเทคนิคการตัดคำ 3 เทคนิค เทคนิคละ 10 รอบ ทั้งนี้เพื่อเพิ่มความเชื่อมั่นและลดอคติในการประมวลผล เมื่อพิจารณาค่าเฉลี่ยคะแนนความแม่นยำสรุปได้คือ การใช้เทคนิคการตัดคำแบบ Longest ให้ความแม่นยำในการทำนายผลมากที่สุด คือ คะแนนความแม่นยำ (Precision) 79.3% คะแนนความถูกต้อง (Recall) 79% และค่าเฉลี่ย (F1 Score) 78.8% ตัวแบบมีความแม่นยำในการทำนายผล (Accuracy) 79%

ผลการวิจัยและวิจารณ์ผล

จากการทดลองสร้างตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับบทความแนะนำสินค้าออนไลน์ โดยใช้ตัวอย่างข้อมูลบทความแนะนำสินค้าและบริการจากเว็บไซต์ www.blognone.com จำนวน 252 บทความ การแสดงความคิดเห็นจำนวน 1,412 ความเห็น จำนวนคำ 83,670 คำ เพื่อสร้างตัวแบบสำหรับการวิเคราะห์ระดับความรู้สึกต่อการแสดงความคิดเห็นต่อสินค้าและบริการออนไลน์ที่เป็นภาษาไทย โดยใช้เทคนิคส่วนการเรียนรู้ของเครื่อง (Machine Learning) ด้านการทำเหมืองข้อความ (Text Mining) ร่วมกับเทคนิคการตัดคำ (Word Segmentation) 3 รูปแบบ ได้แก่ NewMM , Longest และ AttaCut และการสร้างคลังคำศัพท์ (Bag of words) จากนั้น

นำเข้ากระบวนการแยกประเภทผลการวิเคราะห์ 4 เทคนิค ได้แก่ K-Nearest Neighbors, Random Forest, Logistic Regression และ Support Vector Machines (SVM) และนำเข้ากระบวนการทดลอง 5 ขั้นตอน ได้แก่ 1) การเก็บรวบรวมและการเตรียมข้อมูล (Data Preparation Phase) 2) การประมวลผลข้อความ (Text File Processing Phase) 3) การฝึกอบรมข้อมูล (Training & Streaming Phase) 4) การแยกประเภท (Classification Phase) และ 5) การประเมินตัวแบบ (Model Evaluation Phase) จากการสร้างตัวแบบสามารถแยกประเภทระดับความรู้สึกทางอารมณ์ออกเป็น 3 ระดับคือ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) ซึ่งสอดคล้องกับการศึกษาของ (Anurag and Vijaay, 2015) ได้ศึกษาแนวทางในการวิเคราะห์ความรู้สึกของผู้ใช้โดยใช้ตัวจำแนกการ

ทำเหมืองข้อมูล และการศึกษาของ (Trupthi *et al.*, 2017) ได้ศึกษาวิธีการวิเคราะห์ความรู้สึกโดยการสกัดความคิดเห็นบนทวิตเตอร์ เพื่อเปรียบเทียบสมรรถนะของตัวจำแนกประเภท (Classification) สำหรับการวิเคราะห์ความรู้สึก เพื่อวิเคราะห์ความรู้สึกเชิงบวก ความรู้สึกเป็นกลาง และความรู้สึกเชิงลบ พบว่า ถ้ามีการวิเคราะห์ระดับความรู้สึกออกเป็นแบบ 3 ระดับ ทำให้น้ำหนักของการตัดสินใจระดับของอารมณ์มีความถูกต้องมากขึ้นยิ่งขึ้น ในขั้นตอนการประเมินตัวแบบทำการประเมินแยกตามเทคนิคการตัดคำ 3 เทคนิค และเทคนิคการแยกประเภทผลการวิเคราะห์ 4 เทคนิค โดยเฉลี่ยผลการทำงานเทคนิคทั้งหมดของแต่ละเทคนิค สามารถสรุปผลการประเมินความแม่นยำของตัวแบบได้ดังตารางที่ 3

ตารางที่ 3 แสดงผลการประเมินความแม่นยำของตัวแบบ

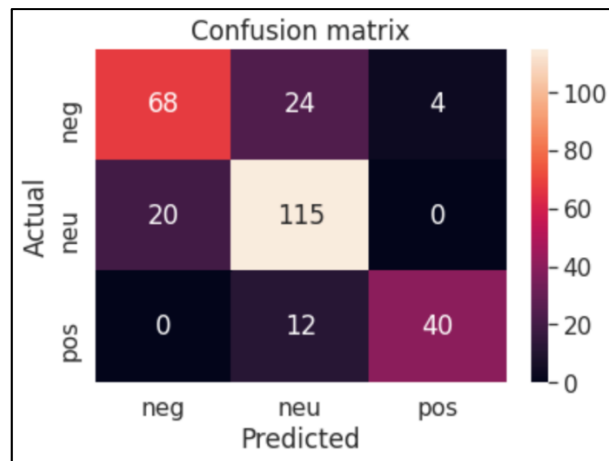
Classifier Model	Word Segmentation Engine											
	NewMM				Longest				AttaCut			
	accuracy	precision	recall	f1	accuracy	precision	recall	f1	accuracy	precision	recall	f1
k-nearest neighbors	47.7	51.3	47.7	46.1	48	47.1	48	46.1	52.6	55.7	52.6	50.8
Random Forest	71.6	73.5	71.6	69.6	68.7	71.5	68.7	66	70.1	72.3	70.1	67.6
Logistic Regression	75.9	76	75.9	74.8	73.9	75.2	73.9	72.6	77.2	78.4	77.2	76.2
SVM	77.7	78.5	77.7	77.4	79	79.3	79	78.8	74.9	76.1	74.9	74.7

จากตารางที่ 3 สามารถอธิบายได้ดังนี้

- 1) การใช้เทคนิคการแยกประเภทแบบ K-Nearest Neighbors ร่วมกับเทคนิคการตัดคำแบบ AttaCut ให้คะแนนความแม่นยำมากที่สุด คือ 52.6%
- 2) การใช้เทคนิคการแยกประเภทแบบ Random Forest ร่วมกับเทคนิคการตัดคำแบบ NewMM ให้คะแนนความแม่นยำมากที่สุด คือ 71.6%

- 3) การใช้เทคนิคการแยกประเภทแบบ Logistic Regression ร่วมกับเทคนิคการตัดคำแบบ AttaCut ให้คะแนนความแม่นยำมากที่สุด คือ 77.2%

- 4) การใช้เทคนิคการแยกประเภทแบบ Support Vector Machines (SVM) ร่วมกับเทคนิคการตัดคำแบบ Longest ให้คะแนนความแม่นยำมากที่สุด คือ 79%



ภาพที่ 7 แสดงตาราง Confusion Matrix การประเมินผลลัพธ์การทำนายตัวแบบ

จากภาพที่ 7 เมื่อพิจารณาจากคะแนนทั้งหมดสรุปได้คือ การใช้เทคนิคการตัดคำ (Word Segmentation) แบบ Longest ซึ่งเหมาะกับการใช้ตัดคำที่ยาวที่สุด ร่วมกับการใช้เทคนิคการสร้างคลังคำศัพท์ (Bag of Words) ซึ่งสอดคล้องกับการศึกษาของ (Chumwatana, 2015) ได้ศึกษาการวิเคราะห์อารมณ์ความคิดเห็น ทศนคติของผู้คน และความรู้สึกต่อการแสดงความคิดเห็นในเชิงบวกหรือเชิงลบจากลูกค้าชาวไทย พบว่า การแสดงความคิดเห็นที่เป็นภาษาไทยเป็นข้อความที่ไม่มีการแบ่งส่วน ประโยคติดกัน อักขระที่ไม่มีขอบเขตคำ เมื่อมีการใช้เทคนิค (Bag of Words) เพื่อสร้างคลังคำศัพท์ร่วมกับการใช้เทคนิคการสกัดคำ (Word Segmentation) เพื่อสร้างตัวแบบทำให้สามารถวิเคราะห์ระดับความรู้สึกทางอารมณ์มีความแม่นยำมากยิ่งขึ้น และเทคนิคการแยกประเภทแบบ Support Vector Machines (SVM) ให้คะแนนความแม่นยำในการทำนายผลการวิเคราะห์ระดับความรู้สึกทางอารมณ์มากที่สุด (Accuracy) คือ 79% ซึ่งสอดคล้องกับการศึกษาของ (Haseena and Ahmad, 2014) ได้ศึกษาการเปรียบเทียบเทคนิคการวิเคราะห์ระดับความรู้สึก พบว่า การจำแนกความรู้สึกโดยใช้เทคนิค Support Vector Machines (SVM) ให้ความแม่นยำสูงกว่าอัลกอริทึมการแยก

ประเภทแบบอื่นเมื่อใช้งานร่วมกับเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) อย่างไรก็ตามในการสร้างตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับบทความแนะนำสินค้าออนไลน์สำหรับบทความภาษาไทย โดยใช้เทคนิคการตัดคำ (Word Segmentation) แบบ Longest Engine ร่วมกับการสร้างคลังคำศัพท์ (Bag of Words) และเทคนิคการแยกประเภทแบบ Support Vector Machines (SVM) เหมาะสมที่สุด

สรุป

ข้อมูลในการเลือกซื้อสินค้าออนไลน์มีความหลากหลาย ซึ่งผู้ใช้อาจจะมีการอ้างอิงจากความคิดเห็นที่เกิดขึ้นในความคิดเห็นของสินค้านั้นๆ แต่ถ้าหากผู้ใช้อ่านและตีความความคิดเห็นต่างๆ ที่มีจำนวนข้อมูลเป็นจำนวนมาก ถ้าหากมีกลไกช่วยเหลือผู้ใช้ในการสรุปความคิดเห็นที่มีอยู่ในตัวสินค้าและบริการได้ จะช่วยประหยัดเวลาและทำให้ผู้ที่ได้ข้อมูลที่รวดเร็วขึ้นและนำไปใช้ตัดสินใจได้ ดังนั้นงานวิจัยนี้ได้นำเสนอวิธีการสร้างตัวแบบสำหรับวิเคราะห์ระดับความรู้สึกทางอารมณ์สำหรับบทความแนะนำสินค้าออนไลน์ภาษาไทย โดยใช้เทคนิคส่วนการเรียนรู้ของเครื่อง (Machine

Learning) เพื่อตัดสินใจความคิดเห็นรวมต่อสินค้าและบริการ โดยนำเทคนิคเหมืองข้อความ (Text Mining) ร่วมกับเทคนิคการตัดคำ (Word Segmentation) ร่วมกับการสร้างคลังคำศัพท์ (Bag of Words) โดยมีการวิเคราะห์ถึงรูปแบบและเทคนิคที่เหมาะสมที่จะนำมาใช้กับความคิดเห็นที่เป็นภาษาไทย และเป็นรูปแบบประโยคยาว และมีคลังคำศัพท์เพื่อช่วยกรองและแบ่งประเภทคำที่เป็นเชิงบวกและเชิงลบ รวมถึงคำกริยา คำนามต่างๆ ทำให้เกิดประสิทธิภาพในการตัดคำที่ดีขึ้น ซึ่งผลลัพธ์โดยรวมจะเห็นได้ว่า การใช้เทคนิคที่นำเสนอจะได้ผลลัพธ์ที่ดีกว่าเทคนิคที่ใช้โดยทั่วไป ขั้นตอนการทดลองสำหรับการทดสอบตัวแบบ มีการเปรียบเทียบเทคนิคการแยกประเภท 4 เทคนิค ได้แก่ K-Nearest Neighbors, Random Forest, Logistic Regression และ Support Vector Machines ซึ่งมีกระบวนการสร้างตัวแบบ 5 ขั้นตอนได้แก่ 1) การเก็บรวบรวมและการเตรียมข้อมูล (Data Preparation Phase) 2) การประมวลผลข้อความ (Text File Processing Phase) 3) การฝึกอบรมข้อมูล (Training & Streaming Phase) 4) การแยกประเภท (Classification Phase) และ 5) การประเมินตัวแบบ (Model Evaluation Phase) เพื่อสร้างตัวแบบในการแยกประเภทระดับความรู้สึกแบบอัตโนมัติออกเป็น 3 ระดับคือ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) จากการทดลองสร้างตัวแบบ การใช้เทคนิคการตัดคำ (Word Segmentation) แบบ Longest ร่วมกับการสร้างคลังคำศัพท์ (Bag of Words) และเทคนิคการแยกประเภทแบบ Support Vector Machines จะให้ความแม่นยำในการทำนายผลมากที่สุด ที่ระดับความแม่นยำ 79% ซึ่งสูงกว่าเทคนิคอื่นที่เปรียบเทียบ โดยมีความเหมาะสมที่จะนำไปใช้งานร่วมกับภาษาไทยที่เป็นคำที่ยาวที่สุด ในการพัฒนาตัวแบบงานวิจัยนี้เพื่อนำตัวแบบที่ได้ไปช่วยผู้บริโภคให้สามารถตัดสินใจในการ

ซื้อสินค้าและบริการได้อย่างรวดเร็ว และช่วยให้ผู้ประกอบการมีข้อมูลที่จะช่วยในการตัดสินใจปรับปรุงพัฒนาสินค้าและบริการเพื่อตอบสนองต่อความต้องการของผู้บริโภคในอนาคตได้อย่างถูกต้อง และตรงกับความต้องการของผู้บริโภคมากที่สุด

ข้อจำกัดของการศึกษาวิจัยในครั้งนี้ ได้แก่

1) ตัวแบบยังไม่สามารถวิเคราะห์คำประชดประชัน (Sarcasm) ได้เนื่องจากชุดข้อมูลที่นำมาใช้ยังไม่มีคำประเภทดังกล่าว ซึ่งต้องอาศัยตัวแบบในการเรียนรู้แยกต่างหากและอาศัยเหมืองข้อความ (Text mining) แบบใช้กฎ (Rule-based) เข้ามาช่วย 2) ตัวแบบสามารถวิเคราะห์ได้เฉพาะการแสดงความคิดเห็นที่เกี่ยวกับการซื้อสินค้าและบริการออนไลน์เท่านั้น

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงไปได้ ต้องขอขอบพระคุณ ผศ.ดร.วรภัทร ไพรเริง ที่ช่วยให้คำปรึกษาในงานวิจัย และขอขอบพระคุณอาจารย์ ผศ.ดร.อรรถพล ชำรงรัตนฤทธิ์ ที่ช่วยให้คำแนะนำการประเมินผลตัวแบบในการวิจัย รวมถึงผู้เชี่ยวชาญในการระบุและนิยามข้อมูลในแต่ละชุด

เอกสารอ้างอิง

- Anurag, P.J. and Vijay, D.K. 2015. Sentiments Analysis of Twitter Data Using Data Mining, pp. 807-810. *In 2015 International conference on Information Processing (ICIP) Vishwakarma Institute of Technology. Pune, India.*
- Chormai, P., Prasertsom, P. and Rutherford, A. 2019. *AttaCut: A Fast and Accurate Neural Thai Word Segmenter.* Available Source: <https://pythainlp.github.io/attacut/>, June 20, 2020.

- Chumwatana, T. 2015. Using Sentiment Analysis Technique for Analyzing Thai Customer Satisfaction from Social media, pp. 659-664. *In Proceedings of the 5th International Conference on Computing and Informatics (ICOICI) 2015*. Istanbul, Turkey.
- Electronic Transections Development Agency (ETDA). 2019. **mSMEs Scoring Report**. Available Source: <https://www.eta.or.th/publishing-detail/msmes-scoring-report.html>, September 26, 2019. (in Thai)
- Fan, X., Li, X., Du, F., Li, X. and Wei, M. 2016. Apply Word Vectors for Sentiment Analysis of APP Reviews, pp. 1062-1066. *In The 2016 3rd International Conference on Systems and Information (ICSAI 2016) IEEE*. Shanghai, China.
- Haseena, R.P. and Ahmad, T. 2014. Sentiment Analysis Techniques - A Comparative Study. **IJCEM International Journal of Computational Engineering & Management** 17(4): 25-29.
- Limkonchotiwat, P. 2018. **How important is the Evaluate Model?: Machine Learning 101**. Available Source: <https://medium.com/mmp-li/evaluate-model-precision-recall-f1-score-machine-learning-101-89dbbada0c96>, January 20, 2020. (in Thai)
- Lorattanachaiyong, I. 2017. Generating Thai Sentiment Lexicon from Online Reviews. A Thesis for the Degree of Master Thesis of Arts, Chulalongkorn University. (in Thai)
- Netisopakul, N. and Thong-iad, K. 2018. Thai Sentiment Resource Using Thai WordNet, pp. 329-340. *In Conference on Complex, Intelligent, and Software Intensive Systems*. Springer, Cham.
- Promla, P. and Sanrach, C. 2020. The Comparison of Efficiency on The Analysis of Satisfaction on Teaching Performance using Sentiment Analysis by Ensemble Technique. **KU Research Journal (Graduate Studies)** 4(20): 104-149. (in Thai)
- PyThaiNLP. 2018. **PyThaiNLP Tutorials**. Available Source: <http://www.thainlp.org/pythainlp/tutorials/>, June 20, 2020
- Sepandar, D.K. and Jonathan, H. 2011. We feel fine and searching the emotional web, pp. 117-126. *In Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11)*. Association for Computing Machinery, New York, USA.
- Sornjapo, N. 2019. Thai Text Automatic Clustering Model Using Machine Learning by Unsupervised Learning with Natural Language Processing. **Sripatum Chonburi Journal** 14(4): 95-106. (in Thai)
- Trupthi, M., Pabboju, S. and Narasimha, G. 2017. Sentiment Analysis on Twitter Using Streaming API, pp. 915-919. *In 2017 IEEE 7th International Advance Computing Conference (IACC)*. Hyderabad, India.