การเปรียบเทียบประสิทธิภาพของอัลกอริทึมตัดคำภาษาไทย ในการวิเคราะห์ความรู้สึกของผู้ใช้แอปพลิเคชันติ๊กต๊อก

COMPARISON OF THAI WORD SEGMENTATION ALGORITHMS FOR SENTIMENT ANALYSIS OF TIKTOK APPLICATION USERS

ปฏิวัติ พินิจสุวรรณ[1]* และ ศรันย์ นาคถนอม[2]
Patiwat Phinitsuwan[1] and Sarun Nakthanom [2]

สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช[1,2]
School of Science and Technology Sukhothai Thammathirat Open University[1,2]
Author email: patiwat.pinitsuwan@gmail.com[1]*, sarun.nak@stou.ac.th[2]

Received: August 13, 2025
Revised: September 21, 2025
Accepted: September 29, 2025

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) เปรียบเทียบประสิทธิภาพของอัลกอริทึมการตัดคำภาษาไทยของไลบรารีไพไทยเอ็นแอลพี (PyThaiNLP) ได้แก่ แม็กซิมัมแมตชิง (MM), นิวแม็กซิมัมแมตชิง (NewMM) และลองเกสต์แมตชิง (Longest) ร่วมกับการลบคำหยุด (Stopword) จาก PyThaiNLP และคำหยุดของผู้วิจัย และ 2) ประเมินประสิทธิภาพของแบบจำลอง ได้แก่ นาอีฟเบย์ (NB), รีเกรสชันโลจิสติก (LR), ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และหน่วยความจำระยะยาว-ระยะสั้น (LSTM) ในการจำแนกความคิดเห็นของผู้ใช้แอปพลิเคชันติ๊กต๊อก สถิติที่ใช้ในการวิเคราะห์ได้แก่ ความถูกต้อง (Accuracy), ความแม่นยำ (Precision), ความไว (Recall), คะแนนเอฟวัน (F1-score) และเวลา (Training Time) กลุ่มตัวอย่างได้แก่ข้อมูลความคิดเห็นจำนวน 5,519 รายการที่จาก Google Play Store ผลการวิจัยพบว่า 1) อัลกอริทึมการตัดคำแบบ NewMM ให้ผลลัพธ์ดีที่สุด โดยมีค่า Accuracy 83.52% และ ค่า F1-score 83.06% 2) แบบจำลอง NB ให้ผลลัพธ์สูงสุดในทุกด้าน โดยมีค่า Accuracy 83.73% และ F1-score 83.43% 3) การไม่ลบคำหยุด (Stopword) ให้ผลลัพธ์ที่ดีที่สุดในทุกกรณี แสดงให้เห็นว่าอัลกอริทึมตัดคำและการลบคำหยุดมีผลต่อประสิทธิภาพของการวิเคราะห์ความคิดเห็นอย่างมีนัยสำคัญ

คำสำคัญ: การประมวลผลภาษาธรรมชาติ, การวิเคราะห์ความคิดเห็น, อัลกอริทึมตัดคำภาษาไทย

Abstract

The purposes of this research were to: 1) compare the performance of Thai word segmentation algorithms provided by the PyThaiNLP library—namely Maximum Matching (MM), New Maximum Matching (NewMM), and Longest Matching (Longest)—in combination with stopword removal using both PyThaiNLP stopwords and researcher-defined stopwords; and 2) evaluate the performance of classification models,

including Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM), for sentiment classification of TikTok application user reviews. The performance metrics used in this study were accuracy, precision, recall, F1-score, and training time. The dataset consisted of 5,519 user review comments collected from the Google Play Store.

The results indicated that: 1) the NewMM word segmentation algorithm achieved the best performance, with an accuracy of 83.52% and an F1-score of 83.06%; 2) the Naïve Bayes model demonstrated the highest overall performance, yielding an accuracy of 83.73% and an F1-score of 83.43%; and (3) retaining stopwords (i.e., not performing stopword removal) consistently produced the best results across all experiments. These findings highlight that both word segmentation algorithms and stopword handling strategies significantly affect the performance of Thai sentiment analysis.

Keywords: Natural language processing, Sentiment analysis, Thai word segmentation

บทนำ

ในยุคดิจิทัล การสื่อสารของมนุษย์ได้เปลี่ยนแปลงไปจากการพบปะโดยตรงมาเป็นการติดต่อผ่านเทคโนโลยีและแพลตฟอร์มออนไลน์ต่าง ๆ โดยเฉพาะในประเทศไทยที่มีอัตราการใช้งานอินเทอร์เน็ตผ่านสมาร์ตโฟนอยู่ในระดับสูง ส่งผลให้สมาร์ตโฟนกลายเป็นเครื่องมือหลักในการเข้าถึงข้อมูล ข่าวสาร และปฏิสัมพันธ์กับผู้อื่น หนึ่งในแอปพลิเคชันที่มีความนิยมสูง คือ TikTok ซึ่งเป็นแพลตฟอร์มวิดีโอสั้นที่เปิดโอกาสให้ผู้ใช้สร้างสรรค์เนื้อหาพร้อมใส่เสียงดนตรี เอฟเฟกต์ภาพ และฟีเจอร์หลากหลาย เพื่อความน่าสนใจและการเข้าถึงกลุ่มเป้าหมายได้อย่างมีประสิทธิภาพ ปัจจุบัน TikTok ได้กลายเป็นหนึ่งในแอปพลิเคชันที่มีผู้ใช้งานมากที่สุดในโลก [1], [2] โดยมีระบบอัลกอริทึมแนะนำเนื้อหาที่สามารถเรียนรู้ความชอบของผู้ใช้และนำเสนอวิดีโอที่ตรงกับความสนใจได้อย่างต่อเนื่อง [3]

ความคิดเห็นจากผู้ใช้งานที่แสดงผ่านการรีวิวแอปพลิเคชัน เช่น Google Play Store และ App Store ถือเป็นแหล่งข้อมูลที่มีคุณค่าแสดงถึงประสบการณ์การใช้งาน ความพึงพอใจ ข้อเสนอแนะ และปัญหาที่พบเจอ โดยข้อมูลเหล่านี้สามารถนำมาวิเคราะห์ในการพัฒนาแอปพลิเคชัน ทางการตลาด และการตัดสินใจของผู้ใช้งานรายใหม่ได้อย่างมีประสิทธิภาพ อย่างไรก็ตามหากข้อมูลที่มีจำนวนมาก การวิเคราะห์ความคิดเห็นในลักษณะอัตโนมัติจึงมีความจำเป็น โดยเฉพาะการประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) [4] และการวิเคราะห์ความคิดเห็น (Sentiment Analysis) เพื่อจำแนกความคิดเห็นออกเป็นเชิงบวกหรือเชิงลบอย่างถูกต้องแม่นยำ [5], [6]

สำหรับข้อความภาษาไทยนั้นไม่มีตัวเว้นวรรคระหว่างคำเหมือนภาษาอังกฤษ ทำให้การตัดคำ (Word Segmentation) [7] กลายเป็นขั้นตอนที่มีอิทธิพลโดยตรงต่อความแม่นยำของโมเดลวิเคราะห์ความคิดเห็น ไลบรารี PyThaiNLP ซึ่งเป็นเครื่องมือโอเพนซอร์สสำหรับประมวลผลภาษาธรรมชาติภาษาไทย ได้จัดเตรียมอัลกอริทึมตัดคำไว้หลายแบบ เช่น Maximum Matching (MM), New Maximum Matching (NewMM) และ Longest Matching (Longest) ซึ่งแต่ละอัลกอริทึมมีหลักการและผลลัพธ์ที่แตกต่างกัน [8] โดยงานวิจัยอื่นๆเกี่ยวกับการตัดคำภาษาไทยมักประเมินจากชุดข้อมูลทั่วไป เช่น ข่าว บทความหรือรีวิวสินค้า ยังไม่มีงานที่ศึกษาเปรียบเทียบอัลกอริทึมตัดคำหลายแบบร่วมกับโมเดลที่หลากหลาย บนข้อมูลความคิดเห็นผู้ใช้ TikTok โดยตรง

ดังนั้น งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมตัดคำทั้งสามแบบในการวิเคราะห์ความคิดเห็นของผู้ใช้ TikTok บน Google Play Store โดยใช้แบบจำลองการเรียนรู้ของเครื่อง ได้แก่ Naïve Bayes (NB),

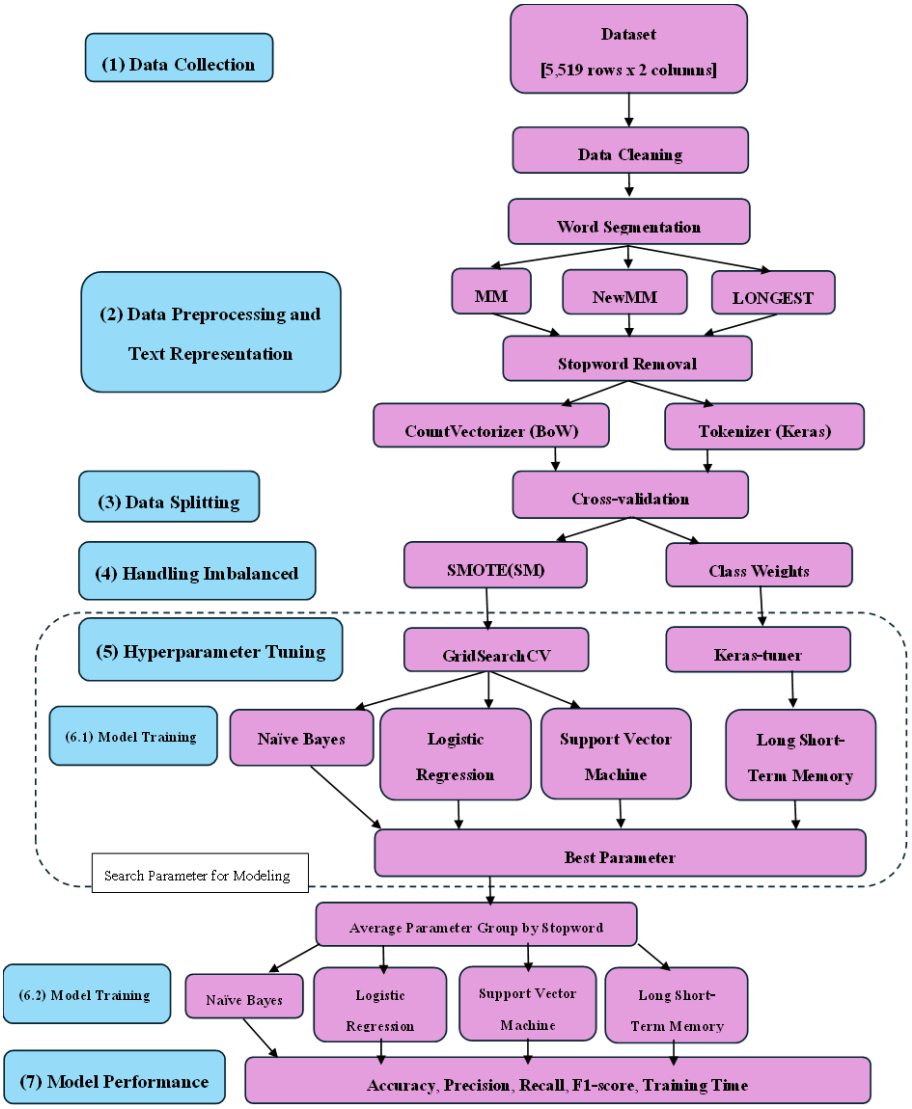
Logistic Regression (LR), Support Vector Machine (SVM) และ Long Short-Term Memory (LSTM) เพื่อจำแนกความคิดเห็นเชิงบวกและเชิงลบ โดยประเมินผลลัพธ์ด้วยค่าชี้วัด Accuracy, Precision, Recall, F1-score และระยะเวลา เพื่อพิจารณาว่าอัลกอริทึมตัดคำรูปแบบใดส่งผลต่อความถูกต้อง ความแม่นยำของแบบจำลองมากที่สุด และสามารถประยุกต์ใช้กับงานวิเคราะห์ข้อความภาษาไทยในบริบทอื่น ๆ ได้อย่างมีประสิทธิภาพ

วัตถุประสงค์การวิจัย

1. เปรียบเทียบประสิทธิภาพของอัลกอริทึมตัดคำภาษาไทย 3 แบบ ได้แก่ MM, NewMM และ Longest
2. ประเมินประสิทธิภาพของแบบจำลอง ได้แก่ NB, LR, SVM และ LSTM

วิธีดำเนินการวิจัย

ขั้นตอนการดำเนินการวิจัยแบ่งออกเป็น 7 ขั้นตอน โดยอ้างอิงมาจากกระบวนการเรียนรู้ของเครื่อง (Machine Learning Process) [9] ดังนี้



รูปที่ 1 ขั้นตอนการดำเนินงานวิจัย

1. การเก็บรวมรวมข้อมูล (Data collection)

ข้อมูลที่ใช้ในการวิจัยถูกดึงมาจาก Google Play Store โดยใช้ Google Play Scraper API [10] ข้อมูลประกอบไปด้วย ข้อความคิดเห็นและคะแนนรีวิว จำนวน 5,519 ข้อความ ระหว่าง ธันวาคม พ.ศ. 2566 - พฤษภาคม พ.ศ. 2567

2. การเตรียมข้อมูลและการแปลงข้อความเป็นเวกเตอร์ (Data Preprocessing and Text Representation)

2.1 การทำความสะอาดข้อมูล (Data Cleaning) ทำการลบข้อมูลที่ไม่จำเป็นออกเช่น ข้อความภาษาอังกฤษ, คำศัพท์ภาษาอังกฤษ, อักขระพิเศษ “[?.;:!\"\'\/#$%&÷(),ๆ 🙄]”, ตัวเลข และ อิโมจิ (Emoji) เพื่อลดสัญญาณรบกวน (Noise)

2.2 การใส่ป้ายกำกับ (Data Labelling) ให้คะแนนรีวิวเป็นเกณฑ์ในการกำหนด 1-3 เชิงลบ มีจำนวน 2,161 ข้อความ และ 4-5 เชิงบวก มีจำนวน 3,358 ข้อความ

2.3 การตัดคำภาษาไทย (Thai Word Segmentation) ใช้ 3 อัลกอริทึมตัดคำ ได้แก่ แม็กซิมัมแมตชิง (MM), นิวแม็กซิมัมแมตชิง (NewMM) และลองเกสต์แมตชิง (Longest) [8]

2.4 การลบคำหยุด (Stopword Removal) ใช้ PyThaiNLP Stopword จำนวน 1,030 คำ และชุดคำหยุดที่กำหนดโดยผู้วิจัย จำนวน 194 คำ ในการลบคำที่ไม่สื่อความหมายออกจากข้อความ เนื่องจากคำหยุดมีแนวโน้มไม่ช่วยในการจำแนกความหมายเชิงบวก-ลบ จึงทดสอบทั้งการลบและไม่ลบเพื่อวิเคราะห์ผลกระทบต่อโมเดล [7]

2.5 การแปลงข้อความเป็นเวกเตอร์ (Text Vectorization) ข้อความที่ผ่านการตัดคำแล้วจะถูกแปลงเป็นเวกเตอร์เพื่อใช้ในโมเดลการเรียนรู้ของเครื่อง โดยการทำ Bag of Words (BoW) ด้วย CountVectorizer สำหรับแบบจำลอง NB, LR และ SVM ส่วนแบบจำลอง LSTM ใช้ Keras Tokenizer [7]

3. การแบ่งชุดข้อมูล (Data Splitting)

ใช้วิธีการแบ่งแบบ Stratified K-Fold โดยกำหนดจำนวนชุดพับ (Fold) เท่ากับ 5 เพื่อให้ทุกชุดข้อมูลมีการกระจายของคลาสที่สมดุลกันในการฝึกสอนและทดสอบ [11]

4. จัดการปัญหาข้อมูลไม่สมดุล (Imbalanced Data Handling)

เป็นการจัดการกับข้อมูลที่ไม่สมดุลระหว่างคลาสผลลัพธ์ โดยใช้ Synthetic Minority Oversampling Technique (SMOTE) [12] เพื่อเพิ่มจำนวนความคิดเห็นเชิงลบให้สมดุลกับความคิดเห็นเชิงบวกกับแบบจำลอง NB, LR และ SVM ร่วมกับการกำหนดค่าน้ำหนักคลาส (Class Weight) [13] เพื่อให้โมเดลให้ความสำคัญกับคลาสที่มีจำนวนน้อยมากขึ้นกับ LSTM

5. การปรับจูนค่าพารามิเตอร์ (Hyperparameter Tuning)

ใช้ GridSearchCV เพื่อค้นหาค่าพารามิเตอร์ที่เหมาะสมที่สุดสำหรับแบบจำลอง NB, LR และ SVM ใช้ Keras tuner สำหรับแบบจำลอง LSTM โดยหาพารามิเตอร์ที่ดีที่สุดแยกตามประเภทของคำหยุดโดยการนำค่ามาเฉลี่ยและใช้เป็นตัวแทนของโมเดลนั้นๆแยกตามประเภทของคำหยุด

6. การสร้างและฝึกอบรมโมเดล (Model Training)

ฝึกฝนแบบจำลอง NB, LR, SVM และ LSTM โดยใช้พารามิเตอร์ตัวแทนแยกตามประเภทของคำหยุด

7. การประเมินผลและเปรียบเทียบประสิทธิภาพ (Model Evaluation and Performance Comparison)

ใช้ตัวชี้วัดทางสถิติในการวิเคราะห์ผลลัพธ์ ได้แก่ Accuracy (ค่าความถูกต้อง), Precision (ค่าความแม่นยำ), Recall (ค่าความไว), F1-score (คะแนนเอฟวัน) และ Training Time (ระยะเวลา) โดยเอาผลลัพธ์ของแต่ละแบบจำลองแยกตามอัลกอริทึมการตัดคำและคำหยุด

ผลการวิจัย

1. ผลการวิจัยเปรียบเทียบประสิทธิภาพของอัลกอริทึมตัดคำไทย

อัลกอริทึมตัดคำโดยใช้ Maximum Matching (MM), New Maximum Matching (NewMM) และ Longest Matching (Longest) โดยพิจารณาผลลัพธ์จาก Accuracy, Precision, Recall, F1-Score และ Training Time เพื่อวิเคราะห์ว่าอัลกอริทึมตัดคำใดมีประสิทธิภาพดีที่สุดในการจำแนกความคิดเห็นของผู้ใช้ TikTok บน Google Play Store โดยทำการเฉลี่ยและเปรียบเทียบข้อมูลดังต่อไปนี้

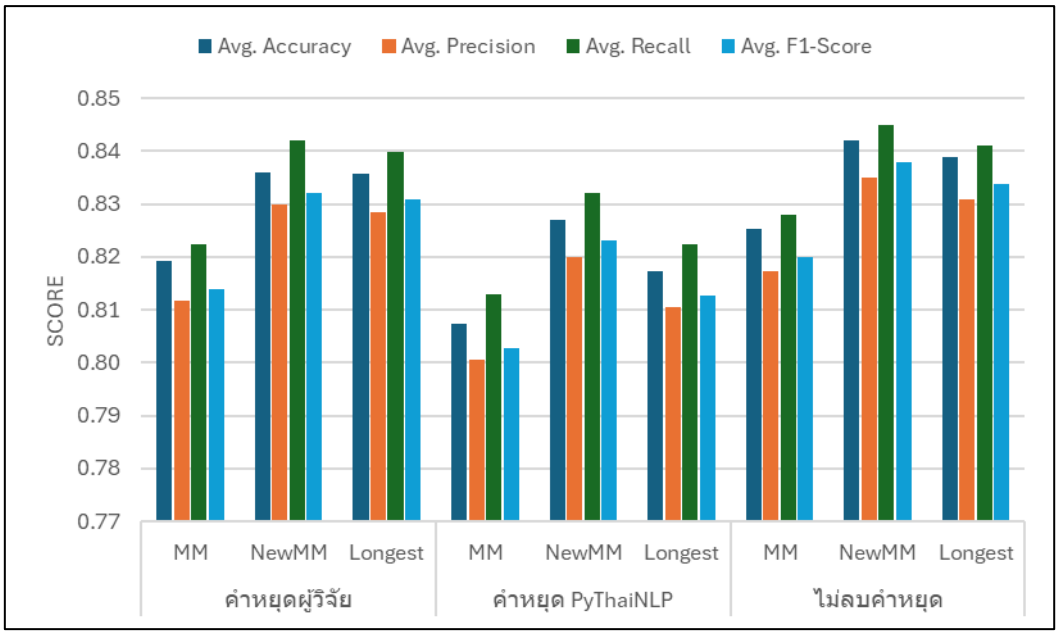
ตารางที่ 1 ประสิทธิภาพการตัดคำในแต่ละอัลกอริทึมกรณีที่ใช้คำหยุดเดียวกัน

ที่	คำหยุด	อัลกอริทึม	Avg. Accuracy	Avg. Precision	Avg. Recall	Avg. F1-score	Avg. Time (seconds)
1	คำหยุดผู้วิจัย	MM	0.8191	0.8118	0.8224	0.8139	72.2440
		NewMM	**0.8360**	**0.8300**	**0.8420**	**0.8320**	**61.6500**
		Longest	0.8356	0.8284	0.8397	0.8309	64.2720
2	คำหยุด PyThaiNLP	MM	0.8073	0.8005	0.8128	0.8027	85.3780
		NewMM	**0.8270**	**0.8200**	**0.8320**	**0.8230**	**83.1900**
		Longest	0.8173	0.8106	0.8224	0.8127	105.3000
3	ไม่ลบคำหยุด	MM	0.8252	0.8173	0.8279	0.8200	89.0820
		NewMM	**0.8420**	**0.8350**	**0.8450**	**0.8380**	**82.6800**
		Longest	0.8388	0.8309	0.8410	0.8337	79.6900

จากตารางที่ 1 พบว่าในกลุ่มที่ใช้คำหยุดของผู้วิจัย พบว่าอัลกอริทึม NewMM และ Longest มีประสิทธิภาพค่อนข้างใกล้เคียงกัน โดย NewMM ให้ค่า Accuracy เฉลี่ยอยู่ที่ 0.8360 ส่วน Longest อยู่ที่ 0.8356 ซึ่งทั้งคู่ดีกว่า MM อย่างชัดเจน (MM = 0.8191) สำหรับค่า F1-score ก็สอดคล้องกัน โดย NewMM อยู่ที่ 0.8320 และ Longest อยู่ที่ 0.8309 แสดงให้เห็นว่าทั้งสองอัลกอริทึมสามารถจัดการกับการตัดคำในกลุ่มคำหยุดนี้ได้อย่างสมดุล สำหรับเวลาในการประมวลผล MM ใช้เวลามากที่สุดในกลุ่ม (72.2440 วินาที) ในขณะที่ NewMM ใช้เวลาน้อยที่สุด (61.6500 วินาที)

สำหรับกลุ่ม คำหยุดของ PyThaiNLP พบว่าผลโดยรวมต่ำกว่ากลุ่มอื่นในทุกอัลกอริทึม โดยค่า Accuracy เฉลี่ยของ MM, NewMM และ Longest อยู่ที่ 0.8073, 0.8270 และ 0.8173 ตามลำดับ ส่วนค่า F1-score ที่ดีที่สุดในกลุ่มนี้ คือ NewMM ซึ่งอยู่ที่ 0.8230 แต่ก็ยังต่ำกว่าที่ได้จากคำหยุดของผู้วิจัยหรือการไม่ลบคำหยุด ทำให้เห็นว่าชุดคำหยุดจาก PyThaiNLP อาจตัดคำที่มีความสำคัญทางอารมณ์ออกไปมากเกินไป

ในกรณีที่ไม่ลบคำหยุดพบว่าผลโดยรวมดีที่สุดในทั้งสามกลุ่ม โดยเฉพาะเมื่อใช้กับอัลกอริทึม NewMM และ Longest ซึ่งให้ค่า Accuracy ที่ 0.8420 และ 0.8388 ตามลำดับ ส่วนค่า F1-score ของอัลกอริทึม NewMM อยู่ที่ 0.8380 และ Longest อยู่ที่ 0.8337



รูปที่ 2 ประสิทธิภาพการตัดคำในแต่ละอัลกอริทึมในกรณีที่ใช้คำหยุดเดียวกัน

จากรูปที่ 2 แสดงให้เห็นว่าในกรณีไม่ลบคำหยุด พบว่าผลโดยรวมดีที่สุดในทั้งสามกลุ่ม โดยเฉพาะเมื่อใช้กับอัลกอริทึม NewMM และ Longest ทำให้เห็นว่าการไม่ลบคำหยุดอาจช่วยให้โมเดลจับบริบทหรือคำสำคัญบางอย่างได้ดีขึ้น โดยเฉพาะในกรณีของความคิดเห็นที่มีการใช้คำเชื่อมหรือคำเสริมอารมณ์

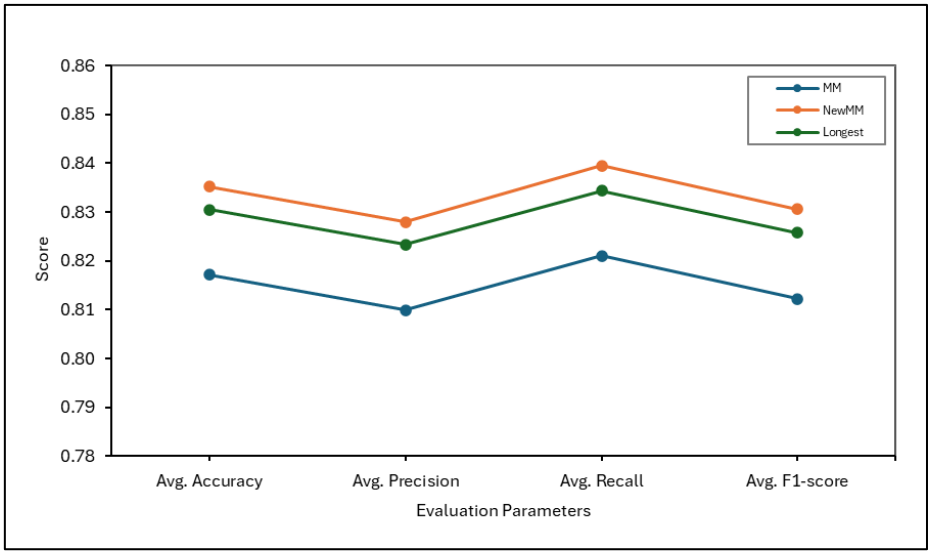
ตารางที่ 2 เปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึมตัดคำ

อัลกอริทึม	Avg. Accuracy	Avg. Precision	Avg. Recall	Avg. F1-score	Avg. Time (seconds)
MM	0.8172	0.8099	0.8210	0.8122	82.2347
NewMM	**0.8352**	**0.8279**	**0.8395**	**0.8306**	**75.8384**
Longest	0.8305	0.8233	0.8344	0.8258	83.0867

จากตารางที่ 2 พบว่าอัลกอริทึม MM มีค่าเฉลี่ย Accuracy อยู่ที่ 0.8172 และ F1-score ที่ 0.8122 ซึ่งต่ำที่สุดในกลุ่มอัลกอริทึมทั้งสามแบบ แสดงให้เห็นว่าการตัดคำโดยใช้หลักการตรงไปตรงมาอาจไม่เหมาะกับข้อมูลที่มีความซับซ้อน เช่น ข้อความจากแพลตฟอร์มโซเชียลที่มักมีการสะกดคำผิด ใช้คำไม่เป็นทางการ หรือคำที่ไม่อยู่ในพจนานุกรมแบบดั้งเดิม นอกจากนี้ MM ยังใช้เวลาในการประมวลผลเฉลี่ยสูงที่สุด (82.2347 วินาที)

สำหรับอัลกอริทึม NewMM ซึ่งเป็นอัลกอริทึมที่มีการปรับปรุงจาก MM โดยมักใช้การประเมินหลายทิศทางและเลือกผลลัพธ์ที่เหมาะสมที่สุด แสดงให้เห็นถึงประสิทธิภาพที่ดีที่สุดในทุกตัวชี้วัด โดยมี Accuracy เท่ากับ 0.8352 และ F1-score อยู่ที่ 0.8306 ซึ่งเป็นค่าสูงที่สุดในกลุ่ม และใช้เวลาเฉลี่ยในการประมวลผลเฉลี่ยเพียง 75.8384 วินาที ซึ่งเร็วกว่าการใช้ MM ซึ่งแสดงให้เห็นถึงความสามารถของ NewMM ในการตัดคำได้อย่างยืดหยุ่นและสอดคล้องกับลักษณะของภาษาธรรมชาติที่หลากหลาย

สำหรับอัลกอริทึม Longest ซึ่งเลือกคำที่ยาวที่สุดในพจนานุกรมก่อน มีค่าประสิทธิภาพโดยรวมอยู่ในระดับดี โดย Accuracy อยู่ที่ 0.8305 และ F1-score เท่ากับ 0.8258 ซึ่งต่ำกว่า NewMM เล็กน้อย แต่ก็ยังสูงกว่า MM ข้อดีของ Longest คือความสามารถในการตัดคำยาวหรือคำรวมได้แม่นยำ



รูปที่ 3 เปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึมตัดคำ

จากรูปที่ 3 พบว่าอัลกอริทึมตัดคำ NewMM ให้ประสิทธิภาพดีที่สุดในทุกตัวชี้วัดแสดงให้เห็นถึงความสามารถของ NewMM ในการตัดคำภาษาไทยกับชุดข้อมูลในงานวิจัยได้เป็นอย่างดีและสอดคล้องกับลักษณะของภาษาธรรมชาติที่หลากหลาย

2. ผลการประเมินประสิทธิภาพของแบบจำลอง

การประเมินประสิทธิภาพของแบบจำลองใช้ตัวชี้วัด ได้แก่ Accuracy, Precision, Recall, F1-Score และ Time

ตารางที่ 3 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองจำแนกความคิดเห็น

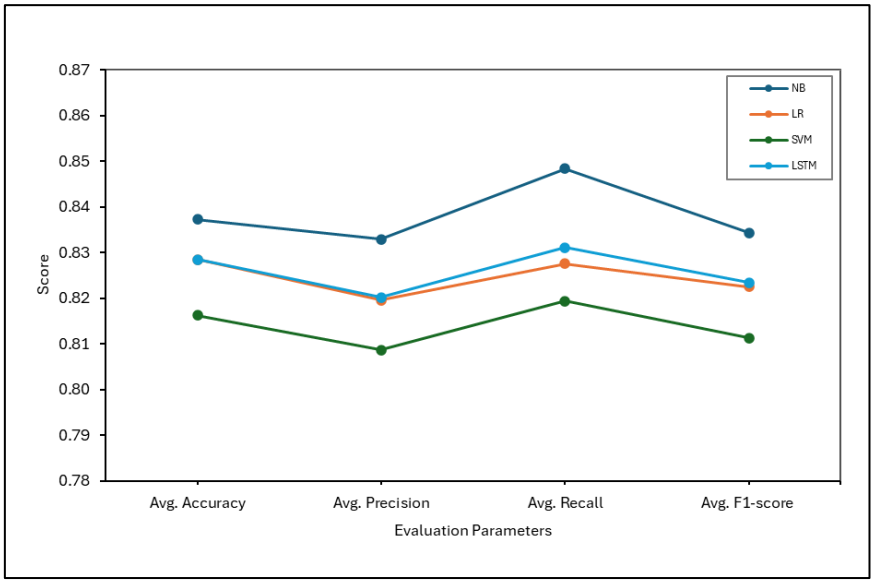
แบบจำลอง	Avg. Accuracy	Avg. Precision	Avg. Recall	Avg. F1-score	Avg. Time (seconds)
NB	**0.8373**	**0.8329**	**0.8484**	**0.8343**	**0.0111**
LR	0.8285	0.8196	0.8276	0.8225	0.9491
SVM	0.8163	0.8087	0.8194	0.8113	1.9806
LSTM	0.8285	0.8202	0.8311	0.8234	318.6100

จากตารางที่ 3 พบว่า NB มีจุดเด่นในด้านความเร็วในการฝึกโมเดล ใช้เวลาเฉลี่ยเพียง 0.0111 วินาที ต่อรอบการฝึก ใช้เวลาน้อยที่สุดเมื่อเทียบกับโมเดลอื่นๆ ด้านประสิทธิภาพ NB มีค่า Accuracy เฉลี่ย 0.8373 และมีค่า F1-score ที่ 0.8343 สูงที่สุดในทุกโมเดล ส่วนค่า Precision = 0.8329 และ Recall = 0.8484 ซึ่งถือว่าสูงมากเช่นกันแสดงให้เห็นว่า NB สามารถจำแนกความคิดเห็นได้ทั้งสองคลาสได้แม่นยำและรวดเร็ว เหมาะสมกับระบบที่ต้องการความเร็วสูง และมีข้อจำกัดด้านทรัพยากรในการประมวลผล

ในส่วนของ LR มีความสมดุลระหว่างประสิทธิภาพและความเร็ว โดยมีค่า Accuracy เฉลี่ย 0.8285 และ F1-score ที่ 0.8225 ต่ำกว่า NB เล็กน้อย แต่ก็ถือว่าอยู่ในระดับที่ใช้งานได้ดี โดยเฉพาะในระบบที่ต้องการความเร็วและความแม่นยำปานกลาง ค่า Precision (0.8196) และ Recall (0.8276) อยู่ในระดับดี และใช้เวลาในการฝึกเฉลี่ย 0.9491 วินาที เร็วกว่า SVM และเร็วมากเมื่อเทียบกับ LSTM แสดงให้เห็นว่า LR เป็นอีกทางเลือกที่ดีสำหรับระบบที่ต้องการประสิทธิภาพในระดับหนึ่ง

ในขณะที่ SVM ที่นิยมใช้ในงานจำแนกข้อความด้วยการแยกข้อมูลเชิงเส้นและไม่เชิงเส้น มีค่า Accuracy เฉลี่ยที่ 0.8163, Precision = 0.8087, Recall = 0.8194 และ F1-score = 0.8113 ต่ำกว่า NB และ LR และใช้เวลาในการฝึกนานกว่าทุกแบบจำลอง (เฉลี่ย 1.9806 วินาที) ซึ่งอาจไม่เหมาะกับงานที่ต้องประมวลผลเร็วหรือมีข้อมูลจำนวนมาก เช่น การใช้งานในระบบเรียลไทม์หรือโมบายล์แอปพลิเคชัน

ในขณะที่ LSTM ซึ่งเป็นแบบจำลองลำดับ (sequential model) ที่อยู่ในกลุ่ม Neural Network มีค่า Accuracy เฉลี่ยที่ 0.8285 เท่ากับของ LR และต่ำกว่า NB และ F1-score เฉลี่ยที่ 0.8234 มีค่า Recall เฉลี่ยที่ 0.8311 โดยสูงกว่า SVM และ LR เล็กน้อย แสดงให้เห็นว่ามีความสามารถในการจำแนกคลาสที่ไม่สมดุลได้ดี แต่ใช้เวลาในการฝึกที่สูงมาก เฉลี่ย 318.6100 วินาทีต่อรอบ ซึ่งมากกว่า NB ถึงเกือบ 30,000 เท่า จึงเหมาะสำหรับระบบที่เน้นความแม่นยำสูงและมีทรัพยากรเพียงพอในการประมวลผลระยะยาว



รูปที่ 4 เปรียบเทียบประสิทธิภาพของแบบจำลอง

จากรูปที่ 4 พบว่าแบบจำลอง NB ให้ประสิทธิภาพดีที่สุดในทุกตัวชี้วัดแสดงให้เห็นถึงความสามารถในการจำแนกข้อความคิดเห็นของกลุ่มตัวอย่างในงานวิจัยได้เป็นอย่างดี

จากผลการวิจัยพบว่า วัตถุประสงค์ข้อ 1 บรรลุผลโดยอัลกอริทึม NewMM มีประสิทธิภาพสูงสุดในบรรดาอัลกอริทึมตัดคำและวัตถุประสงค์ข้อ 2 พบว่า NB เป็นแบบจำลองที่ให้ประสิทธิภาพดีที่สุด

สรุปผลและอภิปรายผลการวิจัย

ผลการวิจัยแสดงให้เห็นว่าอัลกอริทึมตัดคำ NewMM ให้ค่าประสิทธิภาพโดยรวมดีที่สุด โดยมีค่าเฉลี่ย Accuracy เฉลี่ย 0.8352 และ F1-score เฉลี่ย 0.8306 ซึ่งสูงกว่าทั้ง MM และ Longest และยังใช้เวลาในการประมวลผลน้อยที่สุดในกลุ่ม 75.8384 วินาที จึงเหมาะกับการใช้งานในระบบที่ต้องการความแม่นยำสูงและประหยัดเวลา ส่วน Longest ให้ผลลัพธ์ใกล้เคียง NewMM มาก ทั้งในด้าน Accuracy (0.8305) และ F1-score (0.8258) แต่อาจใช้เวลามากขึ้นเล็กน้อย (83.0867 วินาที) ส่วนอัลกอริทึม MM มีประสิทธิภาพโดยรวมต่ำที่สุดในกลุ่ม โดยเฉพาะ Accuracy และ F1-score ที่อยู่ที่ 0.8172 และ 0.8122 ตามลำดับ และยังใช้เวลาในการประมวลผลใกล้เคียงกับ Longest จึงอาจไม่เหมาะสำหรับงานที่ต้องการความแม่นยำสูง สำหรับการเปรียบเทียบกลุ่มคำหยุดพบว่าการไม่ใช้คำหยุดให้ผลลัพธ์ที่ดีที่สุด โดยมีค่าเฉลี่ย Accuracy สูงถึง 0.8354 และ F1-score อยู่ที่ 0.8304 แสดงให้เห็นว่าการตัดคำออกน้อยลงอาจช่วยให้แบบจำลองเรียนรู้ข้อมูลได้ครบถ้วนมากขึ้น ขณะที่ คำหยุดของผู้วิจัย ให้ค่าผลลัพธ์ในระดับที่ดี ส่วนชุดคำหยุดจาก PyThaiNLP ให้ค่าประสิทธิภาพต่ำที่สุด โดยมี F1-score เฉลี่ยที่ 0.8127 ซึ่งอาจเนื่องมาจากการที่ชุดคำหยุดมาตรฐานนี้ตัดคำที่มีนัยสำคัญต่อความคิดเห็นออกไป ทำให้โมเดลขาดข้อมูลสำคัญในการวิเคราะห์

NB เป็นแบบจำลองดีที่สุด ใช้เวลาเฉลี่ยเพียง 0.0111 วินาที ต่อรอบ ซึ่งเร็วกว่าทุกโมเดลที่นำมาเปรียบเทียบ และยังให้ค่า F1-score สูงที่สุด ที่ 0.8343 และค่า Accuracy สูงที่สุดเฉลี่ย 0.8373 แสดงให้เห็นว่า NB เหมาะสำหรับระบบที่ต้องการประสิทธิภาพสูงควบคู่กับความเร็วในการประมวลผล ขณะที่ LSTM ให้ค่า Recall ที่ 0.8311 และมีค่า F1-score และ Accuracy เฉลี่ยใกล้เคียงกับ NB ที่ 0.8234 และ 0.8285 ตามลำดับ แม้จะมีจุดเด่นด้านความสามารถในการเรียนรู้ลำดับของข้อความ แต่ต้องแลกมาด้วยเวลาในการฝึกที่นานกว่ามาก (318.61 วินาที) จึงเหมาะกับระบบที่เน้นคุณภาพการจำแนกมากกว่าความเร็ว สำหรับ LR และ SVM ให้ผลลัพธ์กลาง ๆ โดย LR มี F1-score เฉลี่ยที่ 0.8225 และใช้เวลาเพียง 0.9491 วินาที จึงเหมาะกับระบบที่ต้องการความเร็วและความแม่นยำพอประมาณส่วน SVM แม้จะมี F1-score ใกล้เคียงกัน (0.8113) แต่ใช้เวลาฝึกมากที่สุดรองจาก LSTM (1.9806 วินาที) ซึ่งอาจไม่เหมาะสำหรับระบบที่มีข้อจำกัดด้านเวลา

โดยการศึกษานี้ที่พบว่า NewMM ให้ประสิทธิภาพในการตัดคำดีที่สุดสอดคล้องกับการศึกษาจำแนกความรู้สึกของความคิดเห็นโดยใช้ต้นไม้ตัดสินใจกรณีศึกษาเว็บไซต์จองโรงแรม [8] โดยเปรียบเทียบอัลกอริทึมการตัดคำ ได้แก่ MM, NewMM และ Longest ร่วมกับคำหยุด Stopword ของ PyThaiNLP และชุดคำหยุดของผู้วิจัย โดย NewMM มีประสิทธิภาพดีที่สุด มีค่า Accuracy ที่ 0.90 เมื่อใช้ร่วมกับคำหยุดของผู้วิจัย ในส่วนของคำหยุดของ PyThaiNLP ให้ค่าความ Accuracy เพียง 0.81 แสดงให้เห็นว่า NewMM สามารถตัดคำได้ดีกว่า MM และ Longest เนื่องจากมีการพัฒนาให้สามารถแยกคำได้แม่นยำกว่าการตัดคำแบบดั้งเดิม

สอดคล้องกับการศึกษาการวิเคราะห์รูปแบบความคิดเห็นของข้อมูลโรงแรมโดยใช้เทคนิคการตัดคำแบบผสมผสาน โดยประเมินประสิทธิภาพการวิเคราะห์ความคิดเห็นโดย การจำแนกแบบตัดสินใจต้นไม้ แบบนาอีฟ เบย์ แบบเคเนียเรสเนเบอร์ และดำเนินการโดยรวบรวมข้อมูลจากเว็บไซต์อโกด้าเป็นข้อมูลการแสดงความคิดเห็นของผู้ใช้บริการที่เข้าพักโรงแรม นำมาวิเคราะห์ความรู้สึกโดยหลักการวิธีการตัดคำแบบผสมผสาน ซึ่งผลของการวิจัยพบว่าแบบจำลอง NB ให้ค่าความถูกต้องมากที่สุดถึง 99.66% [14] ซึ่งหมายความว่า NB สามารถทำงานได้ดีเมื่อนำมาใช้กับข้อมูลข้อความขนาดใหญ่ และมีลักษณะเป็นข้อมูลไม่สมดุล

สอดคล้องกับการศึกษาที่สร้างแบบจำลองวิเคราะห์ความรู้สึกสำหรับตลาดดิจิทัลโดยใช้ข้อมูลความคิดเห็นจากแพลตฟอร์มช้อปปี้ (Shopee) ซึ่งได้ทำการเปรียบเทียบแบบจำลอง NB, LR และ SVM ผลการวิจัยพบว่าแบบจำลอง NB ให้ประสิทธิภาพดีที่สุดเช่นกัน โดยให้ค่าความแม่นยำสูงสุดถึงร้อยละ 0.64 เมื่อใช้กับข้อมูลความคิดเห็นของสินค้าหมวดหมู่เครื่องใช้ไฟฟ้า [15]

อย่างไรก็ตาม ค่า Accuracy แบบจำลอง NB ของงานวิจัยนี้เท่ากับ 0.83 ซึ่งน้อยกว่าการศึกษาใน [8], [14] ที่มีค่า Accuracy เท่ากับ 0.90 และ 0.99 ตามลำดับ ซึ่งอาจมีสาเหตุมาจากลักษณะชุดข้อมูลที่แตกต่างกัน โดยเฉพาะความคิดเห็นจาก TikTok ที่มักสั้น กระชับ ใช้ภาษาพูดไม่เป็นทางการ มีการใช้สัญลักษณ์ อีโมจิ หรือการสะกดที่ไม่เป็นมาตรฐาน ทำให้ความยากในการตัดคำและการจำแนกความรู้สึกสูงกว่าเมื่อเทียบกับความคิดเห็นในเว็บไซต์จองโรงแรมซึ่งมีเนื้อหายาวและมีโครงสร้างทางภาษาที่ชัดเจนกว่า

นอกจากนี้ เหตุผลทางภาษาศาสตร์ยังชี้ให้เห็นว่า NewMM เหมาะสมกว่าวิธีการแบบ MM และ Longest เนื่องจากอัลกอริทึมถูกออกแบบให้คำนึงถึงความเป็นไปได้ของการเรียงลำดับคำและการจัดการคำที่มีความกำกวมในภาษาไทยได้ดีกว่า ตัวอย่างเช่น คำหลายพยางค์ที่สามารถแบ่งได้หลายแบบ (word segmentation ambiguity) หรือคำที่ไม่อยู่ในพจนานุกรมมาตรฐาน ซึ่ง NewMM สามารถใช้ heuristics และบริบทช่วยตัดสินใจได้แม่นยำกว่าแบบดั้งเดิม ดังนั้น แม้ค่า Accuracy จะต่ำกว่าเมื่อเทียบกับงานที่ใช้ชุดข้อมูลยาวและเป็นทางการ แต่ผลการทดลองยังคงยืนยันว่า NewMM มีความเหมาะสมในการประมวลผลข้อความสั้นที่มีความซับซ้อนเชิงภาษาศาสตร์ มากกว่าวิธีการตัดคำแบบพื้นฐาน

ข้อเสนอแนะ

1. แบบจำลอง Naive Bayes (NB) ร่วมกับอัลกอริทึมตัดคำแบบ NewMM มีประสิทธิภาพสูงในการวิเคราะห์ความคิดเห็นภาษาไทย และเหมาะสำหรับนำไปประยุกต์ใช้ในงานด้านการวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์และข้อเสนอแนะจากลูกค้า

2. การวิจัยในอนาคตควรให้ความสำคัญกับการจัดการภาษาวัยรุ่นและคำสะกดผิด โดยอาจนำเทคนิคการแก้คำผิดอัตโนมัติและการทำให้ข้อความเป็นมาตรฐาน (Text Normalization) มาใช้ เพื่อเพิ่มความแม่นยำในการตัดคำและการวิเคราะห์ความคิดเห็น

3. ควรมีการศึกษาวิธีการตรวจสอบความขัดแย้งระหว่างคะแนนและข้อความความคิดเห็น เช่น ความคิดเห็นเชิงลบที่มาพร้อมกับคะแนนความพึงพอใจสูง โดยประยุกต์ใช้เทคนิค Sentiment Conflict Detection เพื่อเพิ่มความน่าเชื่อถือของผลการวิเคราะห์

4. ควรพัฒนาและเปรียบเทียบแบบจำลองที่สามารถเข้าใจบริบทของภาษาไทยได้ดียิ่งขึ้น เช่น โมเดล BERT หรือโมเดลในกลุ่ม Transformer เพื่อรองรับข้อความที่มีความซับซ้อนและหลากหลายรูปแบบมากขึ้น

5. ควรศึกษาการใช้วิธีการแปลงข้อความเป็นเวกเตอร์ในรูปแบบอื่น ๆ เช่น TF-IDF หรือ Word Embedding แทนวิธี Bag of Words เพื่อเพิ่มประสิทธิภาพในการเรียนรู้ของแบบจำลองและความแม่นยำในการจำแนกความคิดเห็น

6. ควรขยายชุดข้อมูลให้ครอบคลุมหลายแพลตฟอร์ม เช่น Twitter, Facebook และ Pantip รวมถึงจัดการปัญหาข้อมูลไม่สมดุลด้วยเทคนิค SMOTE หรือวิธีการสุ่มตัวอย่างอื่น ๆ เพื่อให้แบบจำลองสามารถเรียนรู้ได้อย่างมีประสิทธิภาพ

เอกสารอ้างอิง

[1] N. Phonphong, “Usage behavior and satisfaction of TikTok application of generation Y users in Bangkok,” M. S. thesis, Thammasat Univ., Bangkok, 2019, doi: 10.14457/TU.the.2019.1132. (in Thai)

[2] Spring News, “TikTok usage statistics in 2024: Top countries ranked – Thailand ranks 9th,” Jan. 5, 2024. [Online]. Available: https://www.springnews.co.th/digital-business/digital-marketing/848802. [Accessed: Feb. 10, 2025]. (in Thai)

[3] Predictive, “The key to building a successful application,” Feb. 15, 2024. [Online]. Available: https://predictive.co.th/blog/mobile-development-with-data-strategy. [Accessed: Feb. 2, 2025]. (in Thai)

[4] M. Chali, “Analysis and visualization of tourist behavior of hospitality service in Chiang Mai Province using sentiment analysis method,” M.S. thesis, Chiang Mai Univ., Chiang Mai, 2023. [Online]. Available: https://tdc.thailis.or.th/tdc/browse.php?option=show&browse_type=title&titleid=644515. [Accessed: Feb. 10, 2025]. (in Thai)

[5] T. Polsawat, “Sentiment classification methodology for online consumer reviews using ontology based,” M.S. thesis, Khon Kaen Univ., Khon Kaen, 2020. [Online]. Available: https://tdc.thailis.or.th/tdc/browse.php?option=show&browse_type=title&titleid=584241. [Accessed: Feb. 15, 2025]. (in Thai)

[6] W. Sukcharoen and W. Jitsakul, “The comparison of Thai word segmentation for basic first aid with Deepcut algorithm and Newmm algorithm,” *presented at NCCIT 2021*, King Mongkut’s Univ. of Technology North Bangkok, Bangkok, pp. 264–269, 2021. (in Thai)

[7] N. Lertchanvuth, “Analysis of tourist attractions using machine learning techniques: A case study of Yaowarat road, Thailand,” M.S. thesis, King Mongkut’s Inst. of Technology Ladkrabang, Bangkok, 2021. [Online]. Available: https://tdc.thailis.or.th/tdc/browse.php?option=show&browse_type=title&titleid=593395. [Accessed: Feb. 10, 2025]. (in Thai)

[8] M. Manhem, S. Dolah, S. Chunkaew, and S. Mak-on, “Model for classifying review sentiments using decision tree techniques: A case study of a hotel reservation website,” *Academic Journal of Science and Technology*, vol. 1, no. 2, pp. 69–79, 2020, doi: 10.53845/ajst.v1i2.244056. (in Thai)

[9] W. Romsaiyud, *Machine Learning for Predictive Data Analysis and Applications*. Nonthaburi, Thailand: Sukhothai Thammathirat Open Univ. Press, 2023.

[10] JoMingyu, *google-play-scraper* (version unknown). Python package. Jun. 7, 2024. [Online]. Available: https://github.com/JoMingyu/google-play-scraper. [Accessed: Feb. 5, 2025]. (in Thai)

[11] P. Bowornlertsutee and W. Paireekreng, “The model of sentiment analysis for classifying the online shopping reviews,” *Thai-Nichi Institute of Technology Journal of Applied Science and Technology*, vol. 10, no. 1, pp. 71–79, 2021, doi: 10.61284/tni-jast.2021.246375. (in Thai)

[12] A. Jitpittayaporn, “KMSM: Towards more efficient sampling technique for imbalanced classification,” M.S. thesis, Dhurakij Pundit Univ., Bangkok, 2021. [Online]. Available: https://tdc.thailis.or.th/tdc/browse.php?option=show&browse_type=title&titleid=656328. [Accessed: Feb. 10, 2025]. (in Thai)

[13] T. Vichitsrikamol, “Complaint category recommendation system of state agencies using deep learning model,” M.S. thesis, King Mongkut’s Inst. of Technology Ladkrabang, Bangkok, Thailand, 2024. [Online]. Available: https://tdc.thailis.or.th/tdc/dccheck.php?Int_code=101&RecId=9400&obj_id=64822. [Accessed: Feb. 15, 2025]. (in Thai)

[14] S. Sukkasem, “Analysis of sentiment for hotel data using the hybrid word wrap technique,” M.S. thesis, Thai-Nichi Inst. of Technology, 2020. [Online]. Available: https://tdc.thailis.or.th/tdc/browse.php?option=show&browse_type=title&titleid=534599. [Accessed: Feb. 10, 2025]. (in Thai)

[15] W. Arjamnuaikitja, “Sentiment analysis model for digital marketing marketplace,” M.S. thesis, Thammasat Univ., Bangkok, 2024, doi: 10.14457/TU.the.2023.328. (in Thai)