การจำแนกประเภทความรู้สึกของรีวิวผู้ใช้บริการแอพสปอติฟายด้วยวิธีการเรียนรู้ด้วยเครื่อง

ศรัณย์พร กาญจนสุคนธ์[1*], ศิริสรรพ เหล่าหะเกียรติ[2]

บทคัดย่อ

ปัจจุบันโซเชียลมีเดียมีบทบาทมากขึ้นในการเป็นช่องทางให้ผู้บริโภคแสดงความคิดเห็นต่อสินค้าและบริการต่างๆ การวิเคราะห์ความรู้สึก (Sentiment Analysis) จึงเป็นเครื่องมือสำคัญในการทำความเข้าใจความรู้สึกของผู้บริโภค งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบโมเดลในการจำแนกประเภทความรู้สึกจากความคิดเห็นภาษาอังกฤษของผู้ใช้บริการแอพ Spotify โดยใช้ข้อมูลความคิดเห็น 54,708 รายการจากแหล่งข้อมูล Kaggle จำแนกออกเป็นความคิดเห็นเชิงบวก (Positive) และเชิงลบ (Negative) ตามคะแนนที่ให้ไว้ โดยแบ่งชุดข้อมูลออกเป็นชุดข้อมูลฝึก (Train Set) 75% และข้อมูลทดสอบ (Test Set) 25% และนำข้อมูลเข้าสู่การสกัดคุณลักษณะ (Feature Extraction) ด้วยวิธี TF-IDF และ Word2Vec จากนั้นทำการสร้างโมเดลด้วยเทคนิคการเรียนรู้ของเครื่องหลายอัลกอริทึม ได้แก่ Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost (XGB) และ Transformer DistilBERT (DB) ผลการศึกษาพบว่า DistilBERT มีประสิทธิภาพสูงสุดในการจำแนกประเภทความรู้สึก โดยมีค่าความเที่ยง (Precision) เป็น 92.53% ค่าเรียกคืน (Recall) เป็น 89.62% ค่าเอฟวัน (F1-score) เป็น 91.05% ค่า ROC เป็น 90.46% และค่าความแม่น (Accuracy) เป็น 90.39% นอกจากนี้ยังมีการศึกษาหาคุณลักษณะที่สำคัญ (Feature Importance) ที่ส่งผลต่อการจำแนกประเภทความรู้สึกทั้งเชิงบวกและเชิงลบด้วยการวัดค่า Coefficients และ SHAP Value เพื่ออธิบายผลการทำนายของโมเดล ซึ่งจะช่วยให้เข้าใจปัจจัยสำคัญในการจำแนกประเภทและนำไปสู่การปรับปรุงประสิทธิภาพของโมเดลต่อไป โมเดลที่ได้สามารถนำไปใช้เป็นเครื่องมือในการวิเคราะห์ความรู้สึกของผู้ใช้บริการ เพื่อนำข้อมูลไปพัฒนาและปรับปรุงผลิตภัณฑ์และบริการให้ตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น

คำสำคัญ : การเรียนรู้ของเครื่อง, การจำแนกประเภทความรู้สึก, การสกัดคุณลักษณะ, การประมวลผลภาษาธรรมชาติ

---

[1] หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

[2] คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

* Corresponding author:Tel.: 083-9793655 E-mail address: saranporn.knk@g.swu.ac.th

Spotify Reviews Sentiment Classification Using Machine Learning Techniques

Saranporn Kanjanasukhon[1*], Sirisup Laohakiat[2]

Abstract

Currently, social media plays an increasingly important role as a channel for consumers to express their opinions about various products and services. Sentiment Analysis is thus a crucial tool in understanding consumer sentiment. The objective of this research is to create and compare models for sentiment classification from English language opinions of users of the Spotify app, using data from 54,708 reviews sourced from Kaggle. These reviews are categorized into positive and negative sentiments based on the given scores. The data is divided into a training set (75%) and a test set (25%), and then subjected to feature extraction using TF-IDF and Word2Vec methods. Subsequently, models were then built using various machine learning techniques including Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost (XGB), and Transformer DistilBERT (DB). The study finds that DistilBERT performs most effectively in sentiment classification, with Precision at 92.53%, Recall at 89.62%, F1-score at 91.05%, ROC at 90.46%, and Accuracy at 90.39%. Additionally, feature importance is studied to understand significant factors affecting sentiment classification, both positive and negative, by measuring Coefficients and SHAP Value. This explanation of model predictions helps to understand important factors in classification and leads to further improvement of model efficiency. The developed models can be utilized as tools for analyzing user sentiment, enabling data-driven development and enhancement of products and services to better meet user needs.

Keywords: Machine Learning, Sentiment Classification, Feature Extraction, Natural language processing

---

[1] Data Science, Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

[2] Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

* Corresponding author:Tel.: 083-9793655 E-mail address: saranporn.knk@g.swu.ac.th

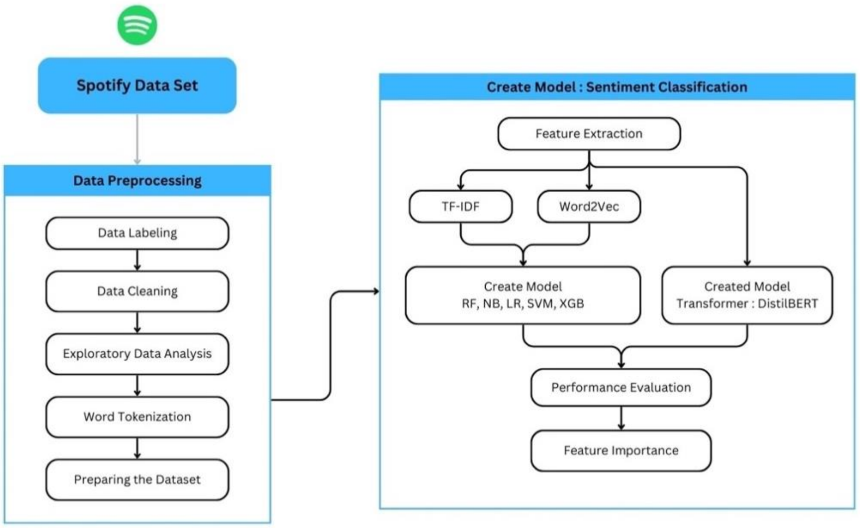
บทนำ

ความคิดเห็นเป็นสิ่งสำคัญสำหรับกิจกรรมต่างๆ ของมนุษย์และธุรกิจ ทั้งในแง่การตัดสินใจและการพัฒนาผลิตภัณฑ์และบริการ ในอดีตการเก็บรวบรวมความคิดเห็นจากผู้บริโภคจำเป็นต้องใช้วิธีการสำรวจแบบเก่า เช่น การส่งแบบสอบถามหรือสัมภาษณ์กลุ่มตัวอย่าง ซึ่งมีข้อจำกัดด้านต้นทุน ระยะเวลา และขนาดของกลุ่มตัวอย่าง [1] อย่างไรก็ตามในปัจจุบันโซเชียลมีเดียได้เติบโตขึ้นอย่างรวดเร็ว ทำให้ผู้บริโภคสามารถแสดงความคิดเห็นได้อย่างเปิดเผยและในวงกว้างบนแพลตฟอร์มออนไลน์ต่างๆ หนึ่งในนั้นคือแพลตฟอร์มสตรีมมิ่งเพลงอย่างสปอติฟาย [2] ที่มีการเปิดให้ผู้ใช้สามารถแสดงความคิดเห็นและให้คะแนนความพึงพอใจหลังจากใช้บริการ ความคิดเห็นเหล่านี้ถือเป็นข้อมูลที่มีคุณค่าสำหรับธุรกิจ เนื่องจากสามารถนำมาวิเคราะห์ความรู้สึกและทัศนคติของผู้บริโภคที่มีต่อผลิตภัณฑ์หรือบริการนั้นๆ ได้

การวิเคราะห์ความรู้สึก (Sentiment Analysis) หรือการจำแนกประเภทความรู้สึก (Sentiment Classification) เป็นกระบวนการทางปัญญาประดิษฐ์ที่ดึงความคิดเห็นหรืออารมณ์ออกมาจากข้อความด้วยเทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ซึ่งต้องผ่านกระบวนการเตรียม (Data Preprocessing) ก่อนนำไปสู่จำแนกข้อความออกเป็น 2 ประเภทหลัก คือ เชิงบวก (Positive) และเชิงลบ (Negative) [3]

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างโมเดลจำแนกประเภทความรู้สึกความคิดเห็นของผู้ใช้บริการสปอติฟาย จากแหล่งข้อมูลจาก Kaggle โดยจะใช้คะแนนความพึงพอใจเป็นตัวกำหนดประเภทความรู้สึกเบื้องต้นเพื่อฝึกสอนโมเดล จากนั้นจะทำการทดสอบประสิทธิภาพของโมเดลการเรียนรู้ของเครื่องหลายแบบเพื่อนำมาเปรียบเทียบหาโมเดลที่มีประสิทธิภาพมากที่สุดที่สามารถจำแนกประเภทความสรู้สึกจากข้อมูลชุดนี้ได้

วิธีดำเนินการ



ภาพที่ 1 ขั้นตอนและวิธีการดำเนินงานวิจัย

ขั้นตอนที่ 1 : การเก็บรวบรวมข้อมูล

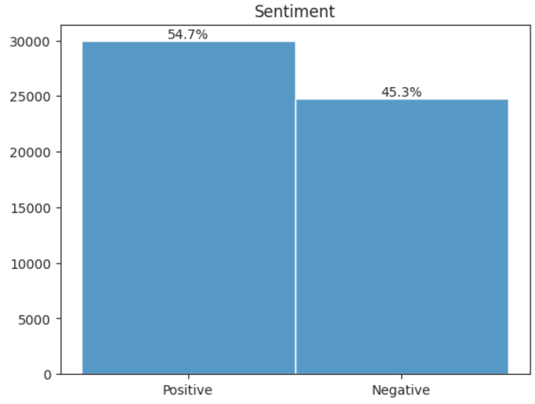
งานวิจัยใช้ชุดข้อมูลที่มีบุคลรวบรวมไว้จากเว็บไซต์ Kaggle ชื่อชุดข้อมูล Spotify App Reviews [4] มีความคิดเห็นทั้งหมด 61,594 ข้อความ ซึ่งชุดข้อมูลจะประกอบไปด้วยคอลัมน์ ดังนี้ ช่วงเวลาในการแสดงความคิดเห็น (Time Submitted) ความคิดเห็น (Review) การระบุคะแนน (Rating) จำนวนถูกใจ (Total Thumbsup) และข้อความตอบกลับ (Reply) ตัวอย่างชุดข้อมูลแสดงดังภาพ

	Time_submitted	Review	Rating	Total_thumbsup	Reply
0	2022-07-09 15:00:00	Great music service, the audio is high quality...	5	2	NaN
1	2022-07-09 14:21:22	Please ignore previous negative rating. This a...	5	1	NaN
2	2022-07-09 13:27:32	This pop-up "Get the best Spotify experience o...	4	0	NaN
3	2022-07-09 13:26:45	Really buggy and terrible to use as of recently	1	1	NaN
4	2022-07-09 13:20:49	Dear Spotify why do I get songs that I didn't ...	1	1	NaN
5	2022-07-09 13:20:20	The player controls sometimes disappear for no...	3	7	NaN
6	2022-07-09 13:19:21	I love the selection and the lyrics are provid...	5	0	NaN
7	2022-07-09 13:17:22	Still extremely slow when changing storage to ...	3	16	NaN
8	2022-07-09 13:16:49	It's a great app and the best mp3 music app I ...	5	0	NaN
9	2022-07-09 13:11:32	I'm deleting this app, for the following reaso...	1	318	NaN

ภาพที่ 2 ตัวอย่างชุดข้อมูล

ขั้นตอนที่ 2 : ขั้นตอนการเตรียมข้อมูล

1) การระบุประเภทความรู้สึกของความคิดเห็น (Data Labeling) เริ่มจากการนำความคิดเห็นทั้งหมด 61,594 ข้อความจากแหล่งข้อมูล Kaggle มาระบุประเภทความรู้สึกออกเป็นบวกหรือลบตามเกณฑ์การให้คะแนน โดยหากมีการให้คะแนนมากกว่า 3 จะถือเป็นความรู้สึกเชิงบวก และน้อยกว่า 3 ถือเป็นความรู้สึกเชิงลบ จากการจำแนกตามเกณฑ์นี้ทำให้ได้ข้อมูลที่ใช้งานได้ 54,708 ข้อความ ประกอบด้วยความคิดเห็นเชิงบวก 29,937 ข้อความ และเชิงลบ 24,771 ข้อความ แสดงสัดส่วนได้ดังภาพ



ภาพที่ 3 สัดส่วนการระบุประเภทความคิดเห็น

2) การทำความสะอาดข้อมูล (Data Cleaning) เนื่องจากข้อความความคิดเห็นมักเขียนในลักษณะที่ไม่เป็นทางการ อาจมีเครื่องหมายวรรคตอน Emoji หรือสัญลักษณ์พิเศษปะปนอยู่ จึงมีการนำข้อความมาทำความสะอาด โดยเปลี่ยนให้เป็นตัวพิมพ์เล็กทั้งหมด และลบเครื่องหมาย Emoji หรือเครื่องหมายวรรคตอนออก เพื่อเก็บไว้เฉพาะตัวอักษรภาษาอังกฤษและตัวเลข
3) การตัดคำ (Word Tokenization) โดยใช้ไลบรารี NLTK มาตัดแยกคำในประโยค ลบคำที่ไม่สื่อความหมาย (Stop Words) เช่น the, a, is ออกไป และแปลงคำที่เหลืออยู่ให้อยู่ในรูปแบบรากศัพท์อ้างอิงตามพจนานุกรม เพื่อลดความซ้ำซ้อนของคำ

หลังผ่านขั้นตอนการเตรียมข้อมูลดังกล่าวแล้ว จึงนำชุดข้อมูลมาแบ่งเป็นชุดข้อมูลสำหรับฝึกสอนโมเดล (Train Set) 75% หรือ 41,031 ข้อความ และชุดข้อมูลทดสอบ (Test Set) 25% หรือ 13,677 ข้อความ เพื่อนำไปสู่ขั้นตอนการสร้างและประเมินประสิทธิภาพโมเดลจำแนกประเภทความรู้สึกต่อไป

ขั้นตอนที่ 3 : การสร้างโมเดล

1) การสกัดคุณลักษณะ (Feature Extraction)

สำหรับอัลกอริทึม Random Forest, Naive Bayes, Logistic Regression, SVM และ XGBoost แปลงข้อความเป็นเวกเตอร์ด้วยเทคนิค TF-IDF (Term Frequency-Inverse Document Frequency) และ Word2Vec

- TF-IDF เป็นเทคนิคการให้น้ำหนักความสำคัญของคำโดย Term Frequency (TF) เป็นความถี่ที่คำปรากฏในข้อความ ส่วน Inverse Document Frequency (IDF) เป็นการวัดความสำคัญสากลของคำหรือคือการหาค่าน้ำหนัก (Weight) แนวคิดหลัก คือ หากคำปรากฏบ่อยขึ้นในข้อความหนึ่ง (ค่า TF สูง) และปรากฏน้อยมากในข้อความอื่น ๆ (ค่า TF ต่ำ ค่า IDF สูง) จะถือว่าคำหรือวลีนั้นสามารถแสดงถึงบทความได้ดี [5]

- Word2Vec ใช้เครือข่ายประสาทเทียมในการแปลงคำเป็นเวกเตอร์แทนความหมาย โดยมี 2 เทคนิคหลักในการเรียนรู้เวกเตอร์คำ คือ CBOW (Continuous Bag-of-Words) ทำนายคำปัจจุบันจากคำรอบข้าง และ Skip-gram ทำนายคำรอบข้างจากคำปัจจุบัน [6]

สำหรับ DistilBERT ใช้ Embedding Layer ของโมเดล DistilBERT ชื่อว่า distilbert-base-uncased ที่ผ่านการฝึกสอน (Pretrained) มาก่อนแล้วในการแปลงข้อความเป็นเวกเตอร์แทนความหมาย [7]

2) การสร้างโมเดล (Modeling)

เมื่อทำการแปลงชุดข้อมูลจากข้อความให้อยู่ในรูปแบบของตัวเลขด้วยวิธีที่กล่าวไปข้างต้นแล้วจะนำเวกเตอร์ที่ได้มาทำการสร้างโมเดลด้วยด้วยการใช้ค่าเริ่มต้น (Default) ดังต่อไปนี้

- Random Forest (RF)

เป็นอัลกอริทึมแบบต้นไม้ตัดสินใจแบบกลุ่ม ทำงานโดยสร้างต้นไม้ตัดสินใจ (Decision Tree) หลายต้นจากชุดข้อมูลเดียวกัน จากนั้นใช้โหวตของต้นไม้แต่ละต้นเพื่อทำการคาดการณ์ ต้นไม้แต่ละต้นสร้างขึ้นโดยใช้วิธีการที่เรียกว่า Bagging ซึ่งเป็นการสุ่มตัวอย่างข้อมูลและคุณสมบัติบางส่วนในการสร้างต้นไม้แต่ละต้น [8]

- Naïve Bayes (NB)

เป็นอัลกอริทึมที่เรียบง่ายและมีประสิทธิภาพสูงในการจำแนกประเภทข้อมูล ทำงานโดยใช้ทฤษฎีความน่าจะเป็นของ Bayes ในการคำนวณความน่าจะเป็นของข้อมูลตัวอย่างในแต่ละคลาส จากนั้นเลือกคลาสที่มีความน่าจะเป็นสูงที่สุดเป็นคำตอบ [9]

- Logistic regression (LR)

ใช้ในการจำแนกประเภทข้อมูลแบบสองทาง (Binary) โดยใช้สมการเชิงเส้นเพื่อคำนวณความน่าจะเป็นของข้อมูลว่าจัดอยู่ในกลุ่ม (Class) ใด สมการเชิงเส้นที่ใช้เรียกว่าฟังก์ชัน Logit จะแปลงความน่าจะเป็นให้เป็นค่าเชิงเส้น ค่าความน่าจะเป็นที่คำนวณได้จากฟังก์ชันสามารถใช้เป็นเกณฑ์ในการจำแนกประเภทข้อมูลได้ [10]

- Support Vector Machine (SVM)

แนวคิดพื้นฐานคือการค้นหาเส้นแบ่งข้อมูล (Hyperplane) ในพื้นที่ที่สามารถแยกจุดข้อมูลออกเป็นสองประเภทด้วยระยะขอบสูงสุด (Maximum Margin) ระยะขอบมาร์จิน (Margin) คือระยะห่างระหว่างเส้นแบ่งข้อมูลและจุดข้อมูลที่ใกล้ที่สุดของแต่ละประเภท โดยใช้ฟังก์ชั่น Kernel [11]

- XGBoost

เป็นอัลกอริทึมแบบบูสต์ (Boosting) หลักการพื้นฐานคือการนำต้นไม้ตัดสินใจ (Decision Tree) หลายๆ ต้นที่มีความแม่นยำต่ำ (Low Accuracy) มารวมกันเป็นโมเดลที่มีความแม่นยำสูงขึ้น จุดเด่นคือการใช้แนวคิดของการไล่ระดับความชัน (Gradient Descent) ในการสร้างต้นไม้แต่ละต้น โดยอาศัยต้นไม้ที่สร้างในรอบก่อนหน้าเป็นฐานแล้วปรับไปในทิศทางที่ลดค่าฟังก์ชันเป้าหมาย (Objective Function) ให้ต่ำสุด เมื่อผ่านการเรียนรู้ของต้นไม้ตัดสินใจหลายๆ ต้น ทำให้ค่าความผิดพลาด (Loss) ลดลงอย่างต่อเนื่อง และได้โมเดลการทำนายมีความลึกมากพอหรือไม่มีค่าความผิดพลาดเกิดขึ้นในต้นไม้ตัดสินใจก่อนหน้าแล้ว [12]

- DistilBERT

เป็นโมเดลประเภท Transformer ซึ่งมีขนาดเล็กและทำงานได้เร็วกว่า BERT โดยอาศัย BERT base model เป็นโมเดลต้นแบบ [7] โดยงานวิจัยนี้มีขั้นตอนการสร้างโมเดล DistilBERT ดังนี้ โหลด Pretrained DistilBERT Model

และ Tokenizer จาก Hugging Face Transformers ที่ชื่อว่า distilbert-base-uncased จากนั้นสร้าง Tokenize Function เพื่อแปลงข้อความในชุดข้อมูลเป็นรหัสตัวเลข แล้วสร้าง Dataset จากชุดข้อมูลที่แปลงเป็นเวกเตอร์แล้วเพื่อนำไปฝึกโมเดลโดยมีการระบุการจำแนกจากชุดข้อมูลทดสอบเข้าไปด้วย ตั้งค่าพารามิเตอร์สำหรับการฝึกโมเดล epoch=8, batch size=16 ใน training_args ฝึกโมเดลโดยใช้ Trainer จาก Transformers library พร้อมชุดข้อมูลฝึกและพารามิเตอร์ training_args บันทึกโมเดล DistilBERT ที่ผ่านการฝึกแล้ว จากนั้นสร้าง Pipeline จากไลบรารี Transformers เพื่อนำโมเดลที่ฝึกแล้วไปทำนายผลกับชุดข้อมูลทดสอบต่อไป

ขั้นตอนที่ 4 : การประเมินประสิทธิภาพโมเดล

โมเดลที่สร้างขึ้นจะถูกนำมาทำนายกับข้อมูลทดสอบเพื่อจำแนกความรู้สึกทั้ง 2 ประเภท คือ ความคิดเห็นเชิงบวก และความคิดเห็นเชิงลบ จากนั้นทำการวัดประสิทธิภาพของแต่ละโมเดลโดยดูที่ค่า Accuracy, Precision, Recall, F1-Scorec และ ROC ซึ่งถูกคำนวณมาจากการใช้ Confusion matrix [13] ซึ่งมีรายละเอียดดังตาราง

ตาราง 1 Confusion Matrix

	Actual Positive Class	Actual Negative Class
Predicted Positive Class	True Positives (TP)	False Negatives (FN)
Predicted Negative Class	False Positives (FP)	True Negatives (TN)

จากตาราง Confusion Matric จะสามารถคำนวณค่าวัดประสิทธิภาพต่าง ๆ ได้ ดังนี้

1) ความแม่นยำ (Accuracy) วัดสัดส่วนของการทำนายที่ถูกต้องต่อจำนวนการทำนายทั้งหมด สูตรในการคำนวณ คือ (TP+TN) / (TP+FP+TN+FN)
2) ความเที่ยง (Precision) วัดความถูกต้องของการจำแนกข้อมูลเป็นบวก สูตรในการคำนวณ คือ TP / (TP+FP)
3) เรียกคืน (Recall) วัดสัดส่วนของข้อมูลบวกที่ถูกจำแนกอย่างถูกต้อง สูตรในการคำนวณ คือ TP / (TP+FN)
4) เอฟวัน (F1-score) เป็นค่าเฉลี่ยถ่วงน้ำหนักระหว่างความเที่ยงและเรียกคืน สูตรในการคำนวณ คือ 2 x (Precision x Recall) / (Precision + Recall)
5) ROC Curve แสดงความสัมพันธ์ระหว่าง TP (Sensitivity) และ FP (1–Specificity) เพื่อประเมินความสามารถในการจำแนกประเภทของโมเดล โดยโมเดลที่ดีจะมีเส้นโค้ง ROC เข้าใกล้มุมบนซ้ายของกราฟ สูตรในการคำนวณ คือ Sensitivity = Recall และ Specificity = TN/(FP+TN)

ขั้นตอนที่ 5 : การหาคุณลักษณะที่สำคัญ (Feature Importance)

งานวิจัยนี้จะมีการหาตัวแปรหรือคุณลักษณะที่สำคัญซึ่งในชุดข้อมูลที่เป็นข้อความตัวแปรในที่นี่คือคำ (Word) เพื่อดูว่าตัวแปรใดส่งผลต่อการจำแนกประเภทว่าความคิดเห็นว่าเป็นเชิงบวก หรือเชิงลบ โดยจะใช้วิธีการหาคุณลักษณะที่สำคัญ 2 วิธีด้วยกันดังนี้

1) วิธี TF-IDF ร่วมกับโมเดล Logistic Regression
ทำการสร้างโมเดล Logistic Regression โดยใช้ข้อมูลที่ถูกสกัดคุณลักษณะด้วย TF-IDF จากนั้นฝึกโมเดลด้วยชุดข้อมูลฝึก แล้วคำนวณค่าความสำคัญของคำ (coefficients) จากโมเดลที่ฝึกแล้ว นำค่า coefficients มาแสดงผลเป็นกราฟเพื่อดูว่าคุณลักษณะใดที่มีความสำคัญมากที่สุด

2) วิธี SHAP (SHapley Additive exPlanations)
คำนวณค่า SHAP Values ด้วยไลบารี่ explaine กับชุดข้อมูลทดสอบ ที่ผ่านการทดสอบจากโมเดล DistilBERT ที่ฝึกแล้ว จากนั้นหาค่าเฉลี่ย SHAP Values สำหรับแต่ละคำ แยกระหว่างความคิดเห็นบวกและลบ นำค่า SHAP Values เฉลี่ยมาแสดงผลเป็นกราฟ เพื่อแสดงคำที่มีความสำคัญในการจำแนกประเภทมากที่สุด

ผลการวิจัยและอภิปรายผลการวิจัย

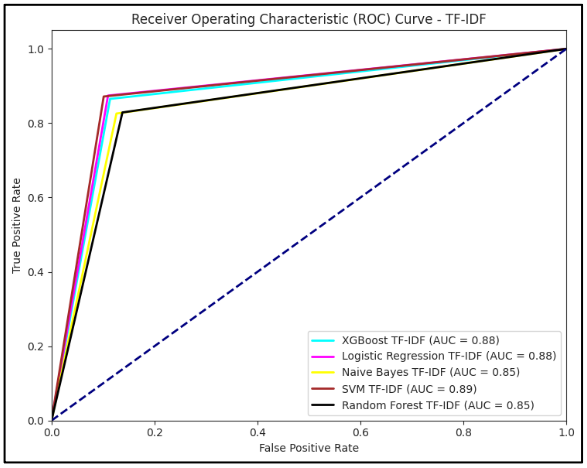
1) ผลการสร้างและปรับปรุงประสิทธิภาพโมเดล

หลังจากผ่านกระบวนการสร้างโมเดลเพื่อจำแนกประเภทความรู้สึกจากข้อความที่เป็นภาษาอังกฤษจากชุดข้อมูลความคิดเห็นจากการใช้บริการ Spotify ที่รวบรวมจากในเว็บ Kaggle โดยใช้การสกัดคุณลักษณะด้วยวิธี TF-IDF และ Word2vec และเทคนิคการเรียนรู้ของเครื่องด้วยอัลกอริทึม Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost (XGB) และ DistilBERT นำไปจำแนกประเภทความรู้สึกออกเป็น 2 ประเภท ได้แก่ ความคิดเห็นเชิงบวก (Positive) และความคิดเห็นเชิงลบ (Negative) กับชุดข้อมูลทดสอบ (Test Set) แล้วทำการวัดประสิทธิภาพของแต่ละโมเดลด้วยการวัดจากค่าความแม่น (Accuracy) ค่าความเที่ยง (Precision) ค่าเรียกคืน (Recall) ค่าเอฟวัน (F1-Score) ค่า ROC ผลลัพธ์ที่ได้จากการประเมินประสิทธิภาพเปรียบเทียบในแต่ละอัลกอริทึม แสดงรายละเอียดดังตารางต่อไปนี้

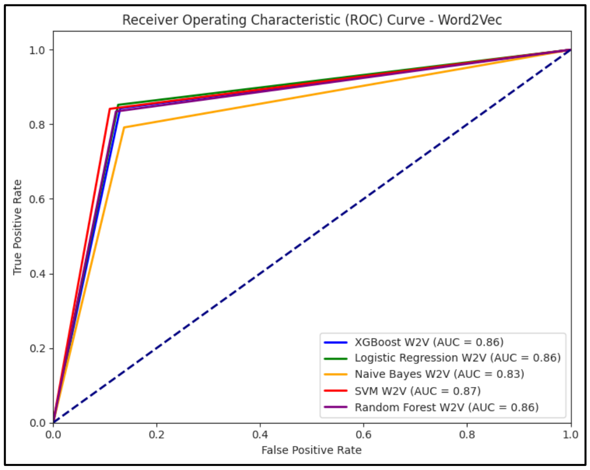
ตาราง 2 ผลการประเมินประสิทธิภาพของแต่ละโมเดล

Feature Extraction	Algorithm	F1-Score	Precision	Recall	ROC	Accuracy
TF-IDF	RF	0.8530	0.8785	0.8289	0.8456	0.8440
	NB	0.8555	0.8872	0.8260	0.8499	0.8477
	LR	0.8896	0.9054	0.8743	0.8823	0.8816
	SVM	0.8913	0.9117	0.8718	0.8852	0.8840
	XGB	0.8830	0.9013	0.8654	0.8758	0.8748
Word2Vec	RF	0.8623	0.8920	0.8345	0.8566	0.8546
	NB	0.8308	0.8740	0.7917	0.8273	0.8240
	LR	0.8711	0.8908	0.8522	0.8633	0.8623
	SVM	0.8660	0.9023	0.8415	0.8660	0.8638
	XGB	0.8642	0.8866	0.8430	0.8567	0.8555
Distilbert		0.9105	0.9253	0.8962	0.9046	0.9039

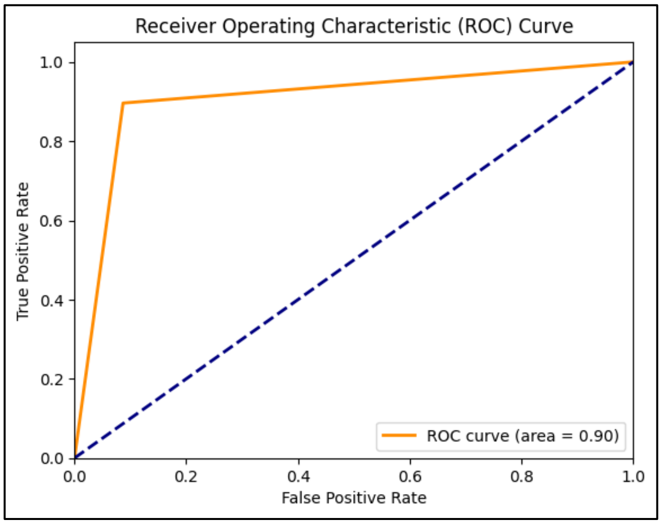
และสามารถแสดงผลลัพธ์ ROC ของแต่ละอัลกอริทึมได้ดังภาพ



ภาพที่ 4 ผลลัพธ์ ROC Curve ของแต่ละอัลกอริทึมจากการสกัดคุณลักษณะ TF-IDF



ภาพที่ 5 ผลลัพธ์ ROC Curve ของแต่ละอัลกอริทึมจากการสกัดคุณลักษณะ Word2Vec



ภาพที่ 4 ผลลัพธ์ ROC Curve ของอัลกอริทึม DistilBERT

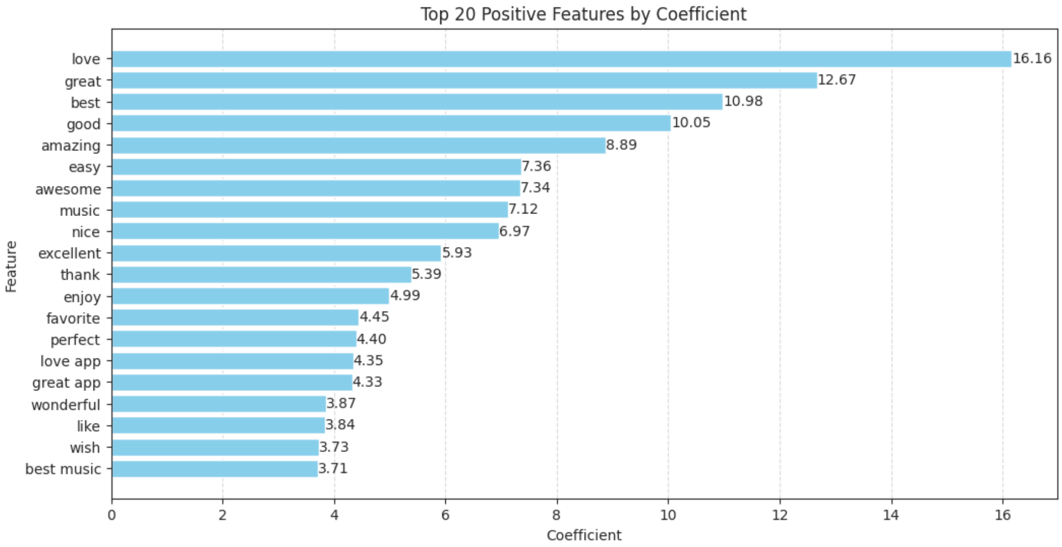
จากผลการประเมินประสิทธิภาพในแต่ละอัลกอริทึมที่ใช้วิธีการสกัดคุณลักษณะด้วยวิธี TF-IDF พบว่าอัลกอริทึม SVM ได้ผลที่มี ประสิทธิภาพโดยรวมสูงที่สุดในทุก ๆ ค่าของตัววัดประสิทธิภาพ โดยมีค่า F1-Score 89.13%, ค่า Precision 91.17%, ค่า Recall 87.18%, ค่า ROC 88.52% และค่า Accuracy 88.40%

2) ผลคุณลักษณะที่สำคัญต่อการจำแนกประเภท (Feature Importance)

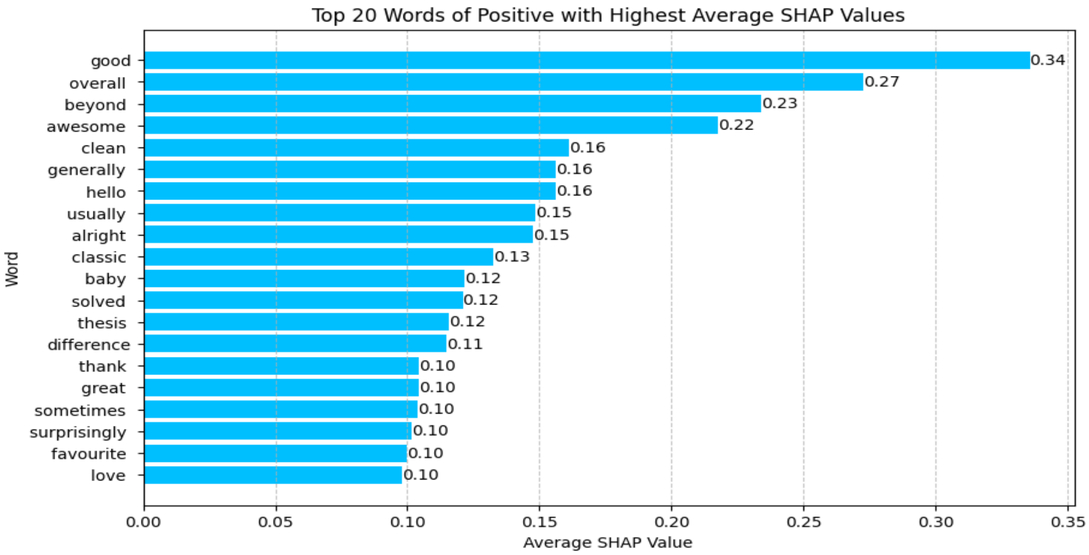
ในงานวิจัยนี้ได้มีการหาคุณลักษณะที่สำคัญซึ่งก็คือคำ (Word) ที่อยู่ในความคิดเห็นเพื่อดูว่าคำใดส่งผลต่อการจำแนกประเภทความคิดเห็นว่าความคิดเห็นนั้นเป็นเชิงบวก หรือเชิงลบ โดยใช้วิธีการหาคุณลักษณะที่สำคัญ 2 วิธี ได้แก่ การสกัดคุณลักษณะด้วยวิธี TF-IDF ร่วมกับอัลกอริทึม Logistic Regression วัดคุณลักษณะที่สำคัญด้วยค่า Coefficients และการใช้อัลกอริทึม SHAP ร่วมกับอัลกอริทึม DistilBERT วัดคุณลักษณะที่สำคัญด้วยค่า SHAP ได้ผลลัพธ์ ดังนี้

- คุณลักษณะจำแนกประเภทที่สำคัญประเภทเชิงบวก

ผลลัพธ์ที่ได้จากการหาคุณลักษณะที่สำคัญที่ถูกจำแนกประเภทเป็นความคิดเห็นเชิงบวก ด้วยวิธีด้วยค่า Coefficients ด้วยวิธี TF-IDF ร่วมกับอัลกอริทึม Logistic Regression แสดงได้ดังภาพที่ 7 และค่าเฉลี่ย SHAP ด้วยวิธี SHAP ดังภาพที่ 8 ตามลำดับ



ภาพที่ 7 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงบวกด้วยวิธี TF-IDF ร่วมกับ Logistic Regression



ภาพที่ 8 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงบวกด้วยวิธี SHAP

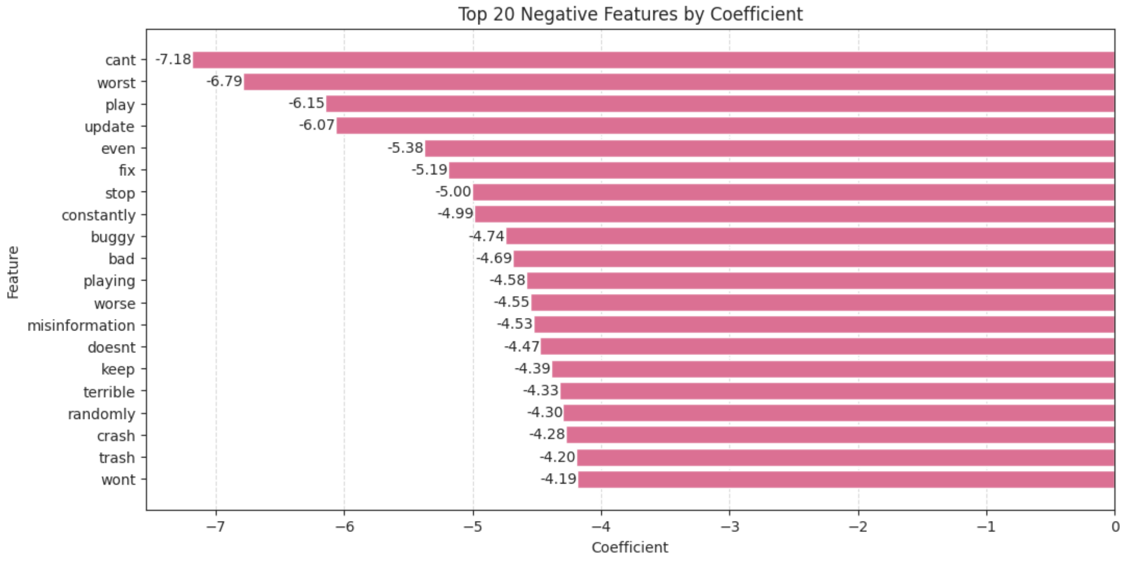
จากภาพการพล็อตกราฟผลลัพธ์คุณลักษณะที่สำคัญ 20 อันดับ แรกเชิงบวกด้วยค่า Coefficients และค่า SHAP คำที่มีร่วมกันระหว่าง 2 วิธี แสดงดังตาราง

ตาราง 3 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงบวกที่มีร่วมกันระหว่าง 2 วิธี

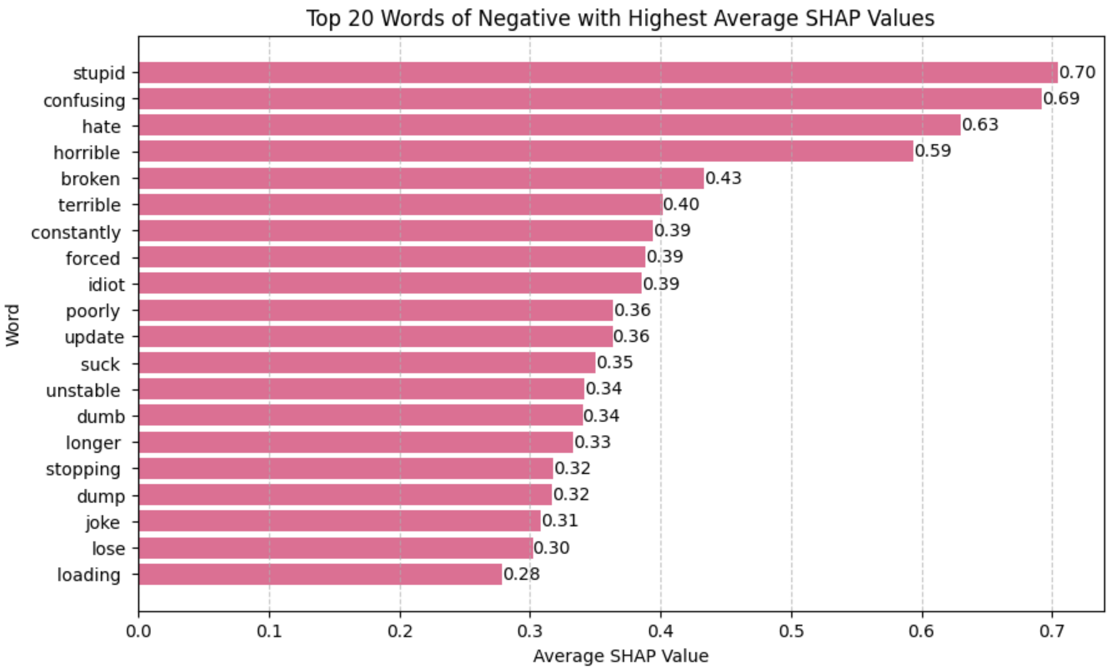
TF-IDF + Logistic Regression		SHAP	
Word	Coef.	Word	SHAP
Love	16.16	Love	0.10
Great	12.67	Great	0.10
Good	10.05	Good	0.34
Awesome	7.34	Awesome	0.22
Thank	5.39	Thank	0.10
Favorite	4.45	Favorite	0.10

- คุณลักษณะจำแนกประเภทที่สำคัญประเภทเชิงลบ

  ผลลัพธ์ที่ได้จากการหาคุณลักษณะที่สำคัญที่ถูกจำแนกประเภทเป็นความคิดเห็นเชิงลบ ด้วยวิธีด้วยค่า Coefficients ด้วยวิธี TF-IDF ร่วมกับอัลกอริทึม Logistic Regression แสดงได้ดังภาพที่ 9 และค่าเฉลี่ย SHAP ด้วยวิธี SHAP ดังภาพที่ 10 ตามลำดับ



ภาพที่ 9 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงลบด้วยวิธี TF-IDF ร่วมกับ Logistic Regression



ภาพที่ 10 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงลบด้วยวิธี SHAP

จากภาพการพล็อตกราฟผลลัพธ์คุณลักษณะที่สำคัญ 20 อันดับ แรกเชิงลบด้วยค่า Coefficients และค่า SHAP คำที่มีร่วมกันระหว่าง 2 วิธี แสดงดังตาราง

ตาราง 4 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงบวกที่มีร่วมกันระหว่าง 2 วิธี

TF-IDF + Logistic Regression		SHAP	
Word	Coef.	Word	SHAP
Update	6.07	Update	0.36
Constantly	4.99	Constantly	0.39
Terrible	4.33	Terrible	0.40

จากการหาคุณลักษณะที่สำคัญด้วยวิธีการวิธีหาค่า Coefficients และการหาค่า SHAP Value ทั้งสองวิธี สามารถวิเคราะห์คุณลักษณะข้อมูลที่สำคัญได้ โดยผลลัพธ์จากวิธีการสกัดคุณลักษณะ TF-IDF ร่วมกับ Logistic Regression เน้นการคำนวณ Coefficients เพื่อหาคุณลักษณะที่สำคัญซึ่งเป็นผลลัพธ์โดยตรงจากโมเดลที่ถูกสร้างขึ้น ในขณะที่ SHAP เน้นการอธิบายผลลัพธ์โดยการให้คำอธิบายว่าแต่ละคุณลักษณะมีผลอย่างไรต่อการทำนายโดยรวมของโมเดล

สรุปผลการวิจัย

งานวิจัยนี้มุ่งเน้นการจำแนกประเภทความรู้สึกของรีวิวผู้ใช้บริการแอพสปอติฟายด้วยวิธีการเรียนรู้ของเครื่อง โดยมีขั้นตอนหลักๆ ดังนี้ ขั้นตอนแรกเริ่มจากการรวบรวมข้อมูลรีวิวภาษาอังกฤษของแอพสปอติฟายจากเว็บไซต์ Kaggle จำนวน 61,594 รีวิว จากนั้นทำการเตรียมข้อมูลโดยระบุประเภทความรู้สึกออกเป็น 2 ประเภท คือเชิงบวกและเชิงลบ จากการใช้คะแนนรีวิวเป็นตัวแบ่ง ไม่นับรวมรีวิวที่มีคะแนน 3 คงเหลือข้อมูล 54,708 รีวิว แบ่งเป็นความคิดเห็นเชิงบวกที่มีคะแนนมากกว่า 3 จำนวน 29,937 รีวิว และความคิดเห็นเชิงลบที่มีคะแนนน้อยกว่า 3 จำนวน 24,771 รีวิว หลังจากนั้นมีการทำความสะอาดข้อมูลเพื่อเตรียมความพร้อมสำหรับการนำไปสร้างโมเดล ในขั้นตอนการสร้างโมเดล ชุดข้อมูลจะถูกแบ่งออกเป็นชุดข้อมูลสำหรับฝึกฝน 75% และชุดข้อมูลสำหรับทดสอบ 25% จากนั้นนำข้อมูลเข้าสู่กระบวนการสกัดคุณลักษณะด้วยวิธี TF-IDF และ Word2Vec เพื่อแปลงข้อความให้อยู่ในรูปแบบเวกเตอร์ที่เหมาะสมสำหรับการสร้างโมเดล หลังจากนั้นจึงสร้างโมเดลด้วยอัลกอริทึมต่างๆ ได้แก่ Random Forest, Naive Bayes, Logistic Regression, Support Vector Machine, XGBoost และ DistilBERT โดยสำหรับ DistilBERT นั้นได้ใช้วิธีการแปลงข้อความเป็นเวกเตอร์จากตัวแบบที่ผ่านการฝึกสอนมาแล้ว (Pre-trained model) เมื่อได้โมเดลทั้งหมดแล้ว จึงนำไปทดสอบประสิทธิภาพกับชุดข้อมูลทดสอบและวัดผลด้วยค่าต่างๆ ได้แก่ F1-Score, Precision, Recall, ROC และ Accuracy ผลปรากฏว่าอัลกอริทึม DistilBERT ให้ผลการจำแนกประเภทความรู้สึกได้ดีที่สุด โดยมีค่า F1-Score 91.05%, Precision 92.53%, Recall 89.62%, ROC 90.46% และ Accuracy 90.39% นอกเหนือจากนั้น งานวิจัยนี้ยังได้ทำการหาคุณลักษณะที่มีความสำคัญต่อการจำแนกประเภทความรู้สึกเชิงบวกและเชิงลบ 20 อันดับแรก โดยใช้วิธีการคำนวณค่า Coefficients จากการสกัดคุณลักษณะด้วย TF-IDF ร่วมกับอัลกอริทึม Logistic Regression และค่า SHAP Value ซึ่งคุณลักษณะเหล่านี้จะช่วยให้เข้าใจถึงปัจจัยที่ส่งผลต่อการจำแนกประเภทความรู้สึกได้ดียิ่งขึ้น และสามารถนำไปปรับปรุงและพัฒนาโมเดลต่อไปในอนาคต

จากงานวิจัยการจำแนกประเภทความรู้สึกของรีวิวผู้ใช้บริการแอพสปอติฟายด้วยวิธีการเรียนรู้ของเครื่อง มีข้อเสนอแนะเพื่อต่อยอดและพัฒนางานวิจัยในอนาคต ได้แก่ การเพิ่มข้อมูลรีวิวจากแหล่งอื่นๆ ที่มีความหลากหลายทางภาษาและอารมณ์ความรู้สึกเพื่อเพิ่มประสิทธิภาพการจำแนก การใช้ผู้เชี่ยวชาญด้านอารมณ์และความรู้สึกในการติดป้ายข้อมูลแทนการใช้คะแนนรีวิวเพื่อความถูกต้องและแม่นยำมากขึ้น การปรับปรุงโมเดลต่างๆ ให้มีประสิทธิภาพสูงขึ้น การขยายการจำแนกไปสู่อารมณ์ความรู้สึกประเภทอื่นๆ และการทดสอบประเมินโมเดลแบบ Deep Learning หรือ Transformer ต่างๆ เพิ่มเติม เช่น BERT, LSTM หรือ CNN เพื่อเปรียบเทียบและหาวิธีการที่เหมาะสมที่สุดสำหรับงานจำแนกประเภทความรู้สึก

กิตติกรรมประกาศ

การจัดทำวิจัยได้รับการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

เอกสารอ้างอิง

[1] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[2] “What is Spotify? - Spotify.” Accessed: Sep. 10, 2023. [Online]. Available: https://support.spotify.com/us/article/what-is-spotify/

[3] H. Kaur, V. Mangat, and Nidhi, “A survey of sentiment analysis techniques,” *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 921–925, 2017.

[4] “Spotify App Reviews.” Accessed: Mar. 31, 2024. [Online]. Available: https://www.kaggle.com/datasets/mfaaris/spotify-app-reviews-2022

[5] Y. Lu, P. Zhang, and C. Zhang, “Research on News Keyword Extraction Technology Based on TF-IDF and TextRank,” *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, pp. 452–455, 2019.

[6] E. M. Dharma, F. Lumban Gaol, H. Leslie, H. S. Warnars, and B. Soewito, “THE ACCURACY COMPARISON AMONG WORD2VEC, GLOVE, AND FASTTEXT TOWARDS CONVOLUTION NEURAL NETWORK (CNN) TEXT CLASSIFICATION,” *J Theor Appl Inf Technol*, vol. 31, no. 2, 2022, [Online]. Available: www.jatit.org

[7] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” Oct. 2019, [Online]. Available: http://arxiv.org/abs/1910.01108

[8] O. Mbaabu, “Introduction to Random Forest in Machine Learning.” Accessed: Oct. 15, 2023. [Online]. Available: https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

[9] A. D. Gordon, *Classification*, 2nd ed.. Boca Raton : Chapman & Hall/CRC, 1999. [Online]. Available: https://www.taylorfrancis.com/books/9781584888536

[10] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, vol. 398. John Wiley & Sons, 2013.

[11] A. Saini, “Guide on Support Vector Machine (SVM) Algorithm.” Accessed: Oct. 16, 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/#:~:text=SVM%20is%20a%20powerful%20supervised,work%20best%20in%20classification%20problems.

[12] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System”, doi: 10.1145/2939672.2939785.

[13] H. M and S. M.N, “A Review on Evaluation Metrics for Data Classification Evaluations,” *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 01–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201.