การจำแนกประเภทความรู้สึกของรีวิวผู้ใช้บริการแอพสปอติฟายด้วยวิธีการเรียนรู้ด้วยเครื่อง

SPOTIFY REVIEWS SENTIMENT CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

ศรัณย์พร กาญจนสุคนธ์

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2566

การจำแนกประเภทความรู้สึกของรีวิวผู้ใช้บริการแอพสปอติฟายด้วยวิธีการเรียนรู้ด้วยเครื่อง

ศรัณย์พร กาญจนสุคนธ์

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล
คณะคณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2566
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

SPOTIFY REVIEWS SENTIMENT CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

SARANPORN KANJANASUKHON

A Master's Project Submitted in Partial Fulfillment of the Requirements
for the Degree of MASTER OF SCIENCE
(Data Science)
Faculty of Science, Srinakharinwirot University
2023

Copyright of Srinakharinwirot University

สารนิพนธ์เรื่อง

การจำแนกประเภทความรู้สึกของรีวิวผู้ใช้บริการแอพสปอติฟายด้วยวิธีการเรียนรู้ด้วยเครื่อง

ของ

ศรัณย์พร กาญจนสุคนธ์

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.....

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญากุล)
คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ)

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.นุวีย์ วิวัฒนวัฒนา)

.....กรรมการภายนอก
(รองศาสตราจารย์ ดร.ดวงดาว วิชาดากุล)

ชื่อเรื่อง	การจำแนกประเภทความรู้สึกของรีวิวผู้ใช้บริการแอพสปอติฟาย ด้วยวิธีการเรียนรู้ด้วยเครื่อง
ผู้วิจัย	ศรัณย์พร กาญจนสุคนธ์
ปริญญา	วิทยาศาสตรมหาบัณฑิต
ปีการศึกษา	2566
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ

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), Na‘ve 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

Title	SPOTIFY REVIEWS SENTIMENT CLASSIFICATION USING MACHINE LEARNING TECHNIQUES
Author	SARANPORN KANJANASUKHON
Degree	MASTER OF SCIENCE
Academic Year	2023
Advisor	Assistant Professor Dr. Sirisup Laohakiat

Currently, social media played 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 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

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ไม่สามารถสำเร็จลุล่วงได้ด้วยผู้วิจัยเพียงคนเดียว ยังมีบุคคลที่ให้คำแนะนำความช่วยเหลือ และให้การสนับสนุน

ขอขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ อาจารย์ที่ปรึกษาที่ได้ให้คำปรึกษาและความช่วยเหลือ รวมถึงเป็นแรงผลักดันชี้แนะแนวทางต่าง ๆ อันเป็นประโยชน์อย่างยิ่งในการทำการวิจัยนี้

ขอขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ซึ่งประกอบไปด้วย รศ.ดร.ดวงดาว วิชาดากุล และ ผศ.ดร.นุวีย์ วิวัฒนวัฒนา วิชาดากุล ที่กรุณาให้เกียรติเป็นคณะกรรมการ รวมทั้งให้คำปรึกษาและข้อเสนอแนะอันเป็นประโยชน์อย่างมากต่อการทำวิจัยและวิทยานิพนธ์ฉบับนี้

ขอขอบคุณ คณาจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาศรินครินทรวิโรฒ ผู้ถ่ายทอดความรู้และคำแนะนำอันเป็นประโยชน์ในการประยุกต์ใช้ในวิทยานิพนธ์ฉบับนี้

และสุดท้ายนี้ ขอขอบพระคุณครอบครัว เพื่อน ซึ่งเป็นแรงขับเคลื่อนสำคัญที่คอยสนับสนุนให้กำลังใจให้คำชี้แนะ จนวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

ศรัณย์พร กาญจนสุคนธ์

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ญ

สารบัญรูปภาพ ..... ฎ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ 2

1.3 ขอบเขตการวิจัย 2

1.4 ขั้นตอนการดำเนินงาน 3

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1 ทฤษฎีที่เกี่ยวข้อง 4

2.1.1 Machine Learning 4

2.1.2 Sentiment Classification 4

2.1.3 Term Frequency-Inverse Document Frequency (TF-IDF) 5

2.1.4 Word2Vec 6

2.1.5 Random Forest 7

2.1.6 Naïve Bayes 8

2.1.7 Logistic Regression 9
2.1.8 Support Vector Machine 9
2.1.9 eXtreme Gradient Boosting (XGBoost) 10
2.1.10 DistilBERT Base Model 11
2.1.12 การวัดประสิทธิภาพโมเดล 12
2.2 งานวิจัยที่เกี่ยวข้อง 14
2.2.1 Business Reviews Classification Using Sentiment Analysis 18
2.2.2 Sentiment Analysis using supervised classification algorithms 19
2.2.3 A survey of sentiment analysis techniques 20
2.2.4 Sentiment Analysis on Twitter Data using KNN and SVM 22
2.2.5 Comparative Study on Machine Learning Algorithms for Sentiment Classification 23
2.2.6 Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning 24
2.2.7 A comparison of features extraction methods for Arabic sentiment analysis 26
2.2.8 Emotion Analysis and Classification of Movie Reviews Using Data Mining 27
2.2.9 Tourist Place Reviews Sentiment Classification Using Machine Learning Techniques 27
2.2.10 Sentiment analysis and classification of Indian farmers' protest using twitter data 28
บทที่ 3 การสร้างโมเดลการจำแนกประเภทความรู้สึก 30
3.1 แนวคิดและภาพรวมวิธีการวิจัย 30
3.2 ขั้นตอนการรวบรวมข้อมูล 31

3.3 ขั้นตอนการเตรียมข้อมูลก่อนการวิเคราะห์ 31
3.3.1 Data Labeling 31
3.3.2 Data Cleaning 32
3.3.3 Word Tokenization 33
3.3.4 ขั้นตอนการการจัดเตรียมชุดข้อมูล (Preparing the Dataset) 35
3.4 ขั้นตอนการสร้างโมเดล 35
3.4.1 Feature Extraction 35
3.4.2 การสร้างโมเดล (Modeling) 37
3.5 การประเมินประสิทธิภาพโมเดล 40
3.6 การหาคุณลักษณะที่สำคัญ (Feature Importance) 40
บทที่ 4 ผลการศึกษา 42
4.1 ผลการสร้างและปรับปรุงประสิทธิภาพโมเดล 42
4.2 ผลคุณลักษณะทีสำคัญต่อการจำแนกประเภท (Feature Importance) 45
บทที่ 5 บทสรุป 53
5.1 สรุปผลการวิจัย 53
5.2 ข้อจำกัดในวิทยานิพนธ์ 54
5.3 ข้อเสนอแนะ 55
บรรณานุกรม 56

สารบัญตาราง

หน้า

ตาราง 1 เมทริกซ์ความสับสน....12
ตาราง 2 สรุปงานวิจัยที่ทำการศึกษา....15
ตาราง 3 สรุปงานวิจัยที่ทำการศึกษา (ต่อ)....16
ตาราง 4 สรุปงานวิจัยที่ทำการศึกษา (ต่อ)....17
ตาราง 5 ผลการเปรียบเทียบอัลกอริทึมที่ใช้วิธีการสกัดข้อมูลแตกต่างกันของ Business Reviews....18
ตาราง 6 ผลการเปรียบเทียบอัลกอริทึมของ SMS [10]....20
ตาราง 7 สรุปบทความที่ได้ทำการสำรวจ [3]....21
ตาราง 8 ผลการเปรียบเทียบประสิทธิภาพอัลกอริทึม [22]....22
ตาราง 9 สรุปผลการเปรียบเทียบประสิทธิภาพโมเดลของชุดข้อมูล IMDB และ Amazon[23]....24
ตาราง 10 ผลการวิเคราะห์ความรู้สึกชุดข้อมูลรีวิวร้านอาหารในการาจี [24]....25
ตาราง 11 ผลการจำแนกประเภทตามกลุ่ม [24]....25
ตารางที่ 12 ผลการวิเคราะห์ความรู้สึกของภาษาอาหรับ[25]....26
ตาราง 13 ผลลัพธ์เปรียบเทียบประสิทธิภาพการจำแนกประเภทความคิดเห็นของชุดข้อมูลรีวิวสถานที่ท่องเที่ยว [27]....28
ตาราง 14 ผลลัพธ์ประสิทธิภาพการอัลกอริทึมจำแนกประเภทร่วมกับวิธี Bag-of-Word[28]....29
ตาราง 15 ผลลัพธ์เปรียบเทียบประสิทธิภาพอัลกอริทึมการจำแนกประเภทความรู้สึกร่วมกับวิธี TF-IDF [28]....29
ตารางที่ 16 ผลเปรียบเทียบการประเมินประสิทธิภาพโดยรวมในแต่ละอัลกอริทึมซึ่งได้จากการทำการสกัดคุณลักษณะด้วยวิธี TF-IDF....42
ตารางที่ 17 ผลเปรียบเทียบการประเมินประสิทธิภาพโดยรวมในแต่ละอัลกอริทึมซึ่งได้จากการทำการสกัดคุณลักษณะด้วยวิธี Word2Vec....43
ตารางที่ 18 ผลการประเมินประสิทธิภาพอัลกอริทึม DistilBERT....44
ตารางที่ 19 สรุปคุณลักษณะที่สำคัญเชิงบวก....48
ตารางที่ 20 สรุปคุณลักษณะที่สำคัญเชิงบวก (ต่อ)....49
ตารางที่ 21 สรุปคุณลักษณะที่สำคัญเชิงลบ....51

สารบัญรูปภาพ

หน้า

ภาพประกอบ 1 กระบวนการการจำแนกประเภทความรู้สึก 5
ภาพประกอบ 2 โครงสร้างโมเดล CBOW และ Skip-gram 7
ภาพประกอบ 3 โครงสร้าง Random Forest 7
ภาพประกอบ 4 กระบวนการทำงาน Naïve Bayes 8
ภาพประกอบ 5 โครงสร้างอัลกอริทึม SVM 10
ภาพประกอบ 6 โครงสร้างของ XGBoost 11
ภาพประกอบ 7 กระบวนการเรียนรู้ของ DistillBERT 12
ภาพประกอบ 8 Receiver Operating Characteristic (ROC CUVE) 14
ภาพประกอบ 9 ผลการเปรียบเทียบอัลกอริทึมของ Business Review 19
ภาพประกอบ 10 กระบวนการวิเคราะห์ความรู้สึก [3] 21
ภาพประกอบ 11 ขั้นตอนและวิธีการดำเนินงาน 30
ภาพประกอบ 12 ตัวอย่างชุดข้อมูล 31
ภาพประกอบ 13 สัดส่วนการระบุประเภทความคิดเห็น 32
ภาพประกอบ 14 ตัวอย่างข้อความที่ผ่านกระบวนการ Word Tokenization 33
ภาพประกอบ 15 Word Cloud ของประเภทความรู้สึกเชิงบวก 34
ภาพประกอบ 16 Word Cloud ของประเภทความรู้สึกเชิงลบ 34
ภาพประกอบ 17 ตัวอย่างผลลัพธ์การสกัดคุณลักษณะด้วยวิธี TF-IDF 36
ภาพประกอบ 18 ตัวอย่างผลลัพธ์การสกัดคุณลักษณะด้วยวิธี Word2Vec 37
ภาพประกอบ 19 ผลลัพธ์ ROC Curve ของแต่ละอัลกอริทึมจาการสกัดคุณลักษณะด้วย TF-IDF 43
ภาพประกอบ 20 ผลลัพธ์ ROC Curve ของแต่ละอัลกอริทึมจาการสกัดคุณลักษณะด้วย Word2Vec 44
ภาพประกอบ 21 ผลลัพธ์ ROC Curve ของอัลกอริทึม DistilBERT 45
ภาพประกอบ 22 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงบวกด้วยวิธี TF-IDF ร่วมกับ Logistic Regression 47
ภาพประกอบ 23 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงบวกด้วยวิธี SHAP 47
ภาพประกอบ 24 กราฟแสดงค่า coefficients คุณลักษณะที่สำคัญเชิงบวกตามคำที่กำหนด 49

ภาพประกอบ 25 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงลบด้วยวิธี TF-IDF ร่วมกับ Logistic 50
ภาพประกอบ 26 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงลบด้วยวิธี SHAP.....50
ภาพประกอบ 27 กราฟแสดงค่า coefficients คุณลักษณะที่สำคัญเชิงลบตามคำที่กำหนด 52

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ความคิดเห็นเป็นหัวใจสำคัญของกิจกรรมต่าง ๆ ของมนุษย์เกือบทั้งหมด เมื่อใดก็ตามที่จำเป็นต้องตัดสินใจก็มักต้องการทราบความคิดเห็นของผู้อื่นร่วมด้วย รวมไปถึงธุรกิจและองค์กรต่างๆ ก็ต้องการความคิดเห็นของผู้บริโภคเกี่ยวกับผลิตภัณฑ์และบริการของตนเอง และผู้บริโภคแต่ละรายก็ต้องการทราบความคิดเห็นของผู้ใช้ผลิตภัณฑ์ที่มีอยู่ก่อนที่จะซื้อ ซึ่งการเก็บรวบรวมความคิดเห็นของผู้บริโภคเป็นสิ่งที่หลายธุรกิจทำมาอย่างยาวนาน เพราะเป็นสิ่งที่สามารถนำมาพัฒนาด้านการตลาด ประชาสัมพันธ์ หรือปรับปรุงสินค้าและบริการของตนเองได้ ในอดีตเมื่อธุรกิจต้องการความคิดเห็นของผู้บริโภค จะมีการจัดทำแบบสำรวจเพื่อสำรวจความคิดเห็นของกลุ่มตัวอย่าง แต่ในปัจจุบันโซเชียลมีเดียมีการเติบโตขึ้นอย่างก้าวกระโดดทำให้ผู้บริโภคสามารถวิจารณ์หรือแสดงความคิดเห็นบนแพตฟอร์มออนไลน์ได้อย่างสาธารณะ และหากใครต้องการซื้อผลิตภัณฑ์อุปโภคบริโภคก็สามารถอ่านบทวิจารณ์หรือความคิดเห็นเกี่ยวกับผลิตภัณฑ์หรือบริการต่าง ๆ ได้อย่างสาธารณะบนโลกออนไลน์เช่นเดียวกัน [1] เช่นเดียวกับสปอติฟาย (Spotify) ซึ่งเป็นธุรกิจให้บริการสตรีมเพลง พอดแคสต์ และวิดีโอ ที่สามารถใช้งานได้ผ่านทางแอพพลิเคชั่นหรือทางเว็บเบราว์เซอร์ มีเพลงให้เลือกฟังมากกว่า 90 ล้านเพลงจากค่ายเพลงต่างๆ ทั่วโลก พร้อมระบบแนะนำเพลงตามรสนิยมของผู้ฟัง ผู้ใช้สามารถเลือกฟังเพลงแบบเรียงลำดับเอง หรือเลือกฟังเพลยลิสต์ที่จัดทำไว้แล้ว นอกจากนี้สปอติฟาย ยังมีฟีเจอร์เพื่อการแชร์และไลฟ์สไตล์ เช่น การแชร์เพลยลิสต์กับเพื่อน, อ่านเนื้อเพลงคู่ขนานไปกับการฟัง, ซิงก์พร้อมกับเพื่อน ๆ ด้วย Group Session เป็นต้น [2] สปอติฟายมีระบบรีวิวและให้คะแนนความพึงพอใจจากผู้ใช้งานจริง ซึ่งผู้ใช้สามารถแสดงความคิดเห็นเกี่ยวกับประสบการณ์การใช้งานและคุณภาพของบริการได้อย่างเปิดเผย บนแพลตฟอร์มออนไลน์ต่าง ๆ เช่น Google Play Store, App Store ความคิดเห็นเหล่านี้ถือเป็นข้อมูลที่มีค่ามากสำหรับทาง Spotify ในการนำไปวิเคราะห์และปรับปรุงบริการ ทั้งในด้านคุณภาพเสียง ประสบการณ์ผู้ใช้งาน ระบบแนะนำเพลง ปัญหาข้อผิดพลาด การออกแบบและฟีเจอร์ต่างๆ เพื่อพัฒนาแอปพลิเคชันให้ตอบโจทย์และตรงใจผู้ใช้งานมากที่สุด และความคิดเห็นเหล่านี้สามารถนำไปใช้ประโยชน์ในเชิงการวิเคราะห์ความรู้สึกของผู้บริโภคได้ซึ่งเป็นสิ่งที่กำลังนิยมในปัจจุบันเพราะสามารถทำให้เข้าใจความคิดเห็นของผู้บริโภคได้มากยิ่งขึ้น

การวิเคราะห์ความรู้สึก (Sentiment Analysis) หรือการจำแนกประเภทความรู้สึก (Sentiment Classification) เป็นกระบวนการดึงความคิดเห็นหรืออารมณ์ออกมาจากข้อความโดยการประยุกต์ใช้การประมวลผลธรรมชาติ (Natural language processing : NLP) โดยการทำวิเคราะห์ความรู้สึกต้องผ่านกระบวนการเตรียม (Data Preprocessing) ก่อนนำเข้าไปสู่การ

จำแนกประเภท ซึ่งแนวคิดพื้นฐานคือค้นหาขั้วประเด็นของข้อความแล้วระบุประเภทออกเป็น 2 ประเภท ได้แก่ เชิงบวก (Positive) หรือเชิงลบ (Negative) [3]

จากความสำคัญที่กล่าวมาข้างต้น ผู้วิจัยจึงมีแนวความคิดที่จะนำเอาเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) และหลักการการจำแนกประเภทความรู้สึก (Sentiment Classification) มาใช้ในการสร้างโมเดลเพื่อใช้ในการจำแนกประเภทความรู้สึกความคิดเห็นของผู้ที่มาใช้บริการสปอติฟาย (Spotify) จากแหล่งข้อมูล Kaggle โดยจะจำแนกความรู้สึกออกเป็น 2 ประเภท ได้แก่ ความคิดเห็นเชิงบวก และความคิดเห็นเชิงลบ โดยจะใช้คะแนนเป็นตัวจำแนกประเภทความรู้สึกเริ่มต้นเพื่อให้โมเดลใช้ในการเรียนรู้ โดยงานวิจัยนี้จะมีการใช้โมเดลที่เหมาะกับการจำแนกประเภทหลายแบบเพื่อนำมาเปรียบเทียบหาโมเดลที่มีประสิทธิภาพมากที่สุดที่สามารถจำแนกประเภทความสรู้สึกจากข้อมูลชุดนี้ได้

1.2 วัตถุประสงค์

1) เพื่อสร้างและปรับปรุงประสิทธิภาพโมเดลให้สามารถจำแนกประเภทความรู้สึกจากความคิดเห็นภาษาอังกฤษของผู้ใช้บริการสปอติฟาย

2) เพื่อศึกษาเปรียบเทียบโมเดลที่สามารถจำแนกประเภทความรู้สึกจากความคิดเห็นภาษาอังกฤษของผู้ใช้บริการสปอติฟาย

1.3 ขอบเขตการวิจัย

1) ข้อมูลที่ใช้ในการจำแนกประเภทความรู้สึกจะใช้ข้อมูลความคิดเห็นที่เป็นภาษาอังกฤษของผู้ใช้บริการแอพสปอติฟายได้จากแหล่งข้อมูลบนเว็บไซต์ Kaggle ที่ถูกรวบรวมความคิดเห็นมาจาก Google Play Store

2) ข้อมูลที่ได้มีจำนวน 54,708 ความคิดเห็น โดยจะจำแนกความรู้สึกออกเป็น 2 ประเภท ได้แก่ ความคิดเห็นเชิงบวก และความคิดเห็นเชิงลบ โดยจะใช้คะแนนเป็นตัวแบ่งประเภทความรู้สึกเริ่มต้น ดังนี้ หากความคิดเห็นที่ให้คะแนนมากกว่า 3 จัดให้เป็นประเภทความคิดเห็นเชิงบวก และความคิดเห็นที่ให้คะแนนน้อยกว่า 3 จัดให้เป็นประเภทความคิดเห็นเชิงลบ

3) สร้างโมเดลเพื่อใช้ในการจำแนกประเภทความรู้สึกความคิดเห็นที่เป็นภาษาอังกฤษของผู้ใช้บริการสปอติฟายที่เป็นภาษาอังกฤษ ด้วยหลักการการจำแนกประเภทความรู้สึกโดยใช้การสกัดคุณลักษณะ (Feature Extraction) ด้วยวิธี TF-IDF และ Word2vec และเทคนิคการเรียนรู้ของเครื่องด้วยอัลกอริทึม Random Forest, Naïve Bayes, Logistic Regression, Support Vector Machine, XGBoost รวมถึงมีการใช้โมเดล DistilBERT ซึ่งเป็นโมเดล Transformer ร่วมด้วย

4) เปรียบเทียบประสิทธิภาพหาโมเดลที่ประสิทธิภาพดีที่สุดมาใช้จำแนกประเภทความรู้สึกความคิดเห็นภาษาอังกฤษของผู้ใช้บริการสปอติฟาย และมีการเปรียบเทียบดูตัวแปรที่สำคัญ (Feature Importance) ต่อประสิทธิภาพในการทำนายความรู้สึกของโมเดล

1.4 ขั้นตอนการดำเนินงาน

1) ศึกษางานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกหรือการจำแนกประเภทความรู้สึกด้วยเทคนิคการเรียนรู้ของเครื่องที่เป็นภาษาอังกฤษ

2) รวบรวมข้อมูลความคิดเห็นที่เป็นภาษาอังกฤษของผู้ใช้บริการสปอติฟายได้จากแหล่งข้อมูลบนเว็บไซต์ Kaggle ที่ถูกรวบรวมความเห็นมาจาก Google Play Store

3) ทำการคัดเลือกความคิดเห็นที่จะนำไปใช้ในการสร้างโมเดลจากความคิดเห็นทั้งหมด

นำความคิดเห็นที่คัดเลือกมาแล้วทำการระบุประเภทความรู้สึกจำแนกเป็น 2 ประเภท โดยใช้คะแนนเป็นตัวแบ่ง

4) สร้างโมเดลจากความคิดเห็นที่มีการระบุประเภทความรู้สึกโดยใช้เทคนิคการเรียนรู้ของเครื่อง และการจำแนกประเภทความรู้สึก ร่วมถึงมีการหาตัวแปรที่สำคัญต่อการทำนาย

5) เปรียบเทียบประสิทธิภาพของโมเดลที่ได้

6) เปรียบเทียบคุณลักษณะหรือตัวแปรที่สำคัญต่อการทำนายความรู้สึกระหว่างเชิงลบและเชิงบวก

7) สรุปผลการดำเนินงานและข้อเสนอแนะ

8) จัดทำบทความวิชาการ

9) จัดทำรูปเล่มวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1) สามารถนำโมเดลไปใช้เป็นเครื่องในการจำแนกประเภทความรู้สึกความคิดเห็นภาษาอังกฤษของผู้ใช้บริการสปอติฟายได้ โดยโมเดลที่มีประสิทธิภาพที่ดีจะช่วยให้สามารถเข้าใจความคิดเห็นของผู้ใช้บริการได้อย่างแม่นยำมากขึ้น ซึ่งจะนำไปสู่การพัฒนาบริการและผลิตภัณฑ์ที่ตรงกับความต้องการของผู้ใช้บริการได้ดีขึ้น

2) ผู้วิจัยได้ศึกษาเปรียบเทียบโมเดลที่สามารถจำแนกประเภทความรู้สึกจากความคิดเห็นภาษาอังกฤษ ซึ่งจะช่วยพัฒนาองค์ความรู้และเทคนิคใหม่ ๆ ในการจำแนกประเภทความรู้สึกจากข้อความได้ในอนาคต

3) สามารถนำความรู้และเทคนิคมาประยุกต์ใช้เพื่อวิเคราะห์กับชุดข้อมูลอื่นได้ในอนาคต

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

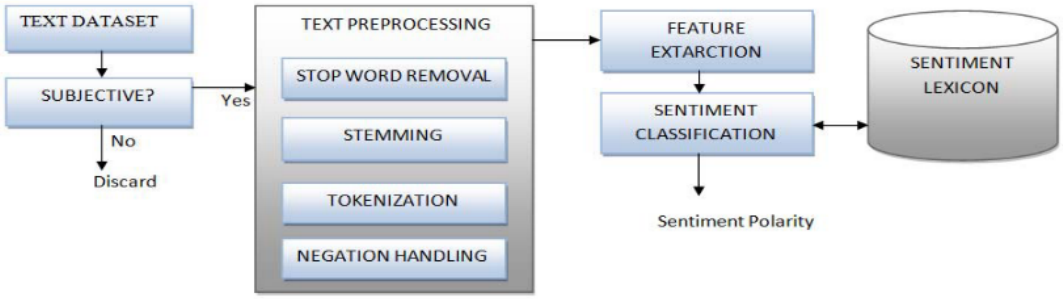
2.1.1 Machine Learning

การเรียนรู้ของเครื่อง (Machine Learning) เป็นองค์ประกอบหนึ่งของปัญญาประดิษฐ์ (Artificial intelligence) ที่ใช้ตัวอย่างในอดีตหรือก่อนหน้าช่วยในการแก้ปัญหา การเรียนรู้ของเครื่องแตกต่างปัญญาประดิษฐ์ตรงที่เกี่ยวข้องกับการเรียนรู้รูปแบบที่ซ่อนอยู่ภายในข้อมูล และต่อมาใช้รูปแบบเพื่อจำแนกหรือทำนายเหตุการณ์ที่เกี่ยวข้องกับปัญหา การเรียนรู้ของเครื่องสามารถเป็นได้ 2 รูปแบบ คือ การเรียนรู้แบบมีผู้สอน (Supervised Learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) การเรียนรู้แบบมีผู้สอนจะเกี่ยวข้องกับแอตทริบิวต์เอาต์พุตที่กำหนดไว้ล่วงหน้า นอกเหนือจากการใช้แอตทริบิวต์อินพุต อัลกอริทึมจะพยายามทำนายและจัดประเภทแอตทริบิวต์ที่กำหนดไว้ล่วงหน้า ซึ่งความถูกต้องและความผิดพลาดในการจัดประเภทของอัลกอริทึมจะขึ้นอยู่กับจำนวนของแอตทริบิวต์ที่กำหนดไว้ล่วงหน้าทีทำนายหรือจัดประเภทได้ว่าถูกต้องหรือไม่ การเรียนรู้แบบมีผู้สอนสามารถแบ่งออกเป็น 2 ประเภท ได้แก่ อัลกอริทึมการจำแนกประเภท (Classification) และอัลกอริทึมการถดถอย (Regression) อัลกอริทึมการจัดประเภทใช้เพื่อทำนายประเภทของชุดข้อมูลใหม่ ในขณะที่อัลกอริทึมการถดถอยใช้เพื่อทำนายค่าตัวเลขของชุดข้อมูลใหม่ การเรียนรู้แบบไม่มีผู้สอนจะเกี่ยวข้องกับการจดจำรูปแบบโดยไม่เกี่ยวข้องกับแอตทริบิวต์เป้าหมาย นั่นคือตัวแปรทั้งหมดที่ใช้ในการวิเคราะห์จะถูกใช้เป็นอินพุต ดังนั้นวิธีการนี้จึงเหมาะสำหรับทำการแบ่งกลุ่ม (Clustering) และการค้นหาความสัมพันธ์ (Association Mining) อัลกอริธึมการแบ่งกลุ่มใช้เพื่อจัดกลุ่มชุดข้อมูลที่คล้ายกันไว้ด้วยกัน ในขณะที่อัลกอริธึมการค้นหาความสัมพันธ์ ใช้เพื่อระบุความสัมพันธ์ระหว่างชุดข้อมูลต่างๆ [4]

2.1.2 Sentiment Classification

การจำแนกประเภทความรู้สึก (Sentiment Classification) ส่วนใหญ่มักถูกกำหนดให้เป็นปัญหาการจำแนกประเภทที่มี 2 คลาส คือเชิงบวกและเชิงลบ โดยข้อมูลการฝึก (Training Data) และข้อมูลทดสอบ (Testing Data) ที่ใช้โดยปกติจะเป็นเกี่ยวกับการรีวิวผลิตภัณฑ์ เนื่องจากรีวิวผลิตภัณฑ์หรือบริการทางออนไลน์จะมีคะแนนกำหนดไว้ให้สำหรับผู้รีวิว เช่น 1-5 ดาว เป็นต้น ซึ่งคะแนนดังกล่าวจะสามารถนำมาแบ่งประเภทความรู้สึกเชิงบวกและเชิงลบได้ ตัวอย่างเช่น ความคิดเห็นที่มี 4 หรือ 5 ดาวจะถือเป็นเชิงบวก และความคิดเห็นที่มี 1 ถึง 2

ดาวถือเป็นเชิงลบ การจำแนกความรู้สึกเป็นปัญหาการจำแนกข้อความแบบหนึ่งโดยพื้นฐานแล้ว การจัดประเภทข้อความแบบดั้งเดิมส่วนใหญ่มักจะจัดประเภทข้อความตามหัวข้อต่างๆ เช่น การเมือง วิทยาศาสตร์ และกีฬา ซึ่งในการจำแนกประเภทดังกล่าวคำที่เกี่ยวข้องกับหัวข้อถือเป็นลักษณะสำคัญที่ใช้ในการจำแนกประเภท แต่ในการจำแนกประเภทความรู้สึกจะให้ความสำคัญกับคำที่แสดงความรู้สึกหรือความคิดเห็นที่บ่งบอกถึงความคิดเห็นเชิงบวกหรือเชิงลบ เช่น เยี่ยม (great), ยอดเยี่ยม (excellent), น่ากลัว (horrible), แย่ (bad) เป็นต้น [1] โดยทั่วไปขั้นตอนกระบวนการในการจำแนกประเภทความรู้สึกจะมีขั้นตอนดังภาพ [3]



ภาพประกอบ 1 กระบวนการการจำแนกประเภทความรู้สึก

2.1.3 Term Frequency-Inverse Document Frequency (TF-IDF)

เทคนิคการแยกคำตามความสำคัญ (Term Frequency-Inverse Document Frequency : TF-IDF) เป็นวิธีการทางสถิติแบบถ่วงน้ำหนักใช้เพื่อประเมินความสำคัญของคำที่มีต่อข้อความหรือคลังข้อความ (Corpus) ความสำคัญของคำจะเพิ่มขึ้นตามความถี่ที่ปรากฏในข้อความ แต่จะลดลงตามสัดส่วนผกผันกับความถี่ที่ปรากฏในคลังข้อความ (Corpus) โดย Term Frequency (TF) เป็นค่าที่บอกความถี่ของคำที่ปรากฏในข้อความ ค่า TF มักถูกทำให้เป็นมาตรฐาน (normalize) เพื่อป้องกันไม่ให้คำเดียวกันมีความสำคัญมากกว่าในข้อความยาวเมื่อเทียบกับข้อความสั้น กล่าวคือคำเดียวกันอาจมีความถี่สูงกว่าในข้อความยาวเมื่อเทียบกับข้อความสั้น โดยไม่คำนึงว่าคำนั้นมีความสำคัญหรือไม่ ส่วน Inverse Document Frequency (IDF) เป็นการวัดความสำคัญสากลของคำหรือคือการหาค่าน้ำหนัก (Weight) แนวคิดหลักคือคำที่ปรากฏในทุกข้อความ ไม่สามารถแสดงถึงเนื้อหาหลักของข้อความได้ แต่ถ้ามีข้อความอื่นๆ ที่มีคำนี้อยู่น้อยกว่า แสดงว่าคำนั้นมีความสามารถในการจำแนกประเภทที่ดี ดังนั้นแนวคิดหลักของเทคนิคTF-IDF คือ หากคำหรือวลีปรากฏบ่อยขึ้นในบทความหนึ่ง (ค่า TF สูง) และปรากฏน้อยมากในบทความอื่นๆ (ค่า TF ต่ำ ค่า IDF สูง) จะถือว่าคำหรือวลีนั้นสามารถแสดงถึงบทความได้ดีและ

สามารถใช้เพื่อการจำแนกประเภทได้ สามารถคำนวณค่า TF-IDF ตามสมการที่ 1,2 และ 3 ได้ดังนี้ [5]

$$TF_{ij} = \frac{f_{ij}}{n_j}$$

โดยที่

f_{ij} แทนจำนวนความถี่ของคำ (term) i ในข้อความ j (document)

n_j แทนจำนวนคำทั้งหมดในข้อความ j (document)

$$IDF_i = 1 + log\frac{\mathrm{N}}{c_i}$$

โดยที่

N แทนจำนวนข้อความทั้งหมด

c_j แทนจำนวนข้อความที่มีคำ i ปรากฏอยู่ในข้อความ

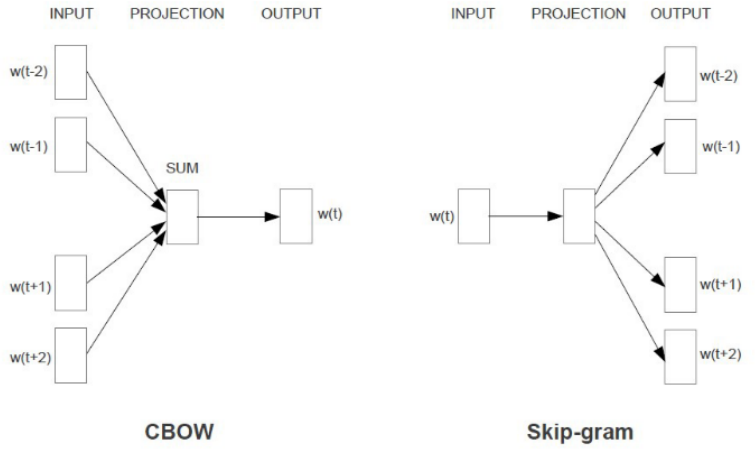
$$TF - IDF_{ij} = TF_{ij} \times IDF_i$$

โดยที่

$TF - IDF_{ij}$ แทนค่าน้ำหนักของคำที่แยกตามความสำคัญ

2.1.4 Word2Vec

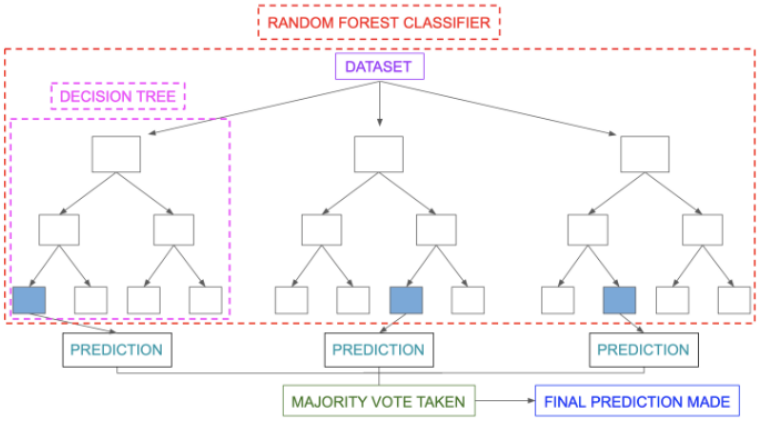
เทคนิค word2vec เป็นเทคนิคที่ใช้เครือข่ายประสาทเทียม (Neural network) ในการเรียนรู้เพื่อแทนคำเป็นเวกเตอร์ (Vector) เวกเตอร์ความหมายของคำคำหนึ่งในประโยคนั้นมีความสัมพันธ์กับความหมายของคำที่อยู่รอบข้าง และสามารถนำไปใช้สำหรับงานประมวลผลภาษาธรรมชาติที่หลากหลาย เทคนิค word2vec โดยสามารถใช้ 2 เทคนิคในการเรียนรู้เวกเตอร์คำ ได้แก่ เทคนิค continuous bag-of-words (CBOW) จะทำนายคำปัจจุบันโดยคำนึงถึงบริบทของคำที่ปรากฏก่อนและหลังคำปัจจุบัน และเทคนิค Skip-gram สามารถทำนายคำบริบทรอบข้างได้จากคำปัจจุบัน ซึ่งแสดงโครงสร้างของทั้งสองเทคนิคได้ ดังภาพ [6]



ภาพประกอบ 2 โครงสร้างโมเดล CBOW และ Skip-gram

2.1.5 Random Forest

Random Forest เป็นอัลกอริทึมการเรียนรู้ของเครื่องแบบต้นไม้ตัดสินใจแบบกลุ่ม ทำงานโดยสร้างต้นไม้ตัดสินใจ (Decision Tree) หลายต้นจากชุดข้อมูลเดียวกัน จากนั้นใช้โหวตของต้นไม้แต่ละต้นเพื่อทำการคาดการณ์ ต้นไม้แต่ละต้นสร้างขึ้นโดยใช้วิธีการที่เรียกว่า Bagging ซึ่งเป็นการสุ่มตัวอย่างข้อมูลและคุณสมบัติบางส่วนในการสร้างต้นไม้แต่ละต้น ประสิทธิภาพของ Random Forest สามารถปรับปรุงได้โดยใช้เทคนิคต่างๆ ดังนี้ การปรับขนาดคุณสมบัติ (Feature Scaling), การลดขนาดข้อมูล (Data Reduction) และ การคัดเลือกคุณสมบัติ (Feature Selection) [7] โครงสร้างการทำ Random Forest แสดงได้ดังภาพ [8]



ภาพประกอบ 3 โครงสร้าง Random Forest

2.1.6 Naïve Bayes

Naïve Bayes เป็นอัลกอริทึมที่เรียบง่ายและมีประสิทธิภาพสูงในการจำแนกประเภทข้อมูล สามารถนำไปประยุกต์ใช้กับงานต่างๆ ได้หลากหลาย รวมถึงการจำแนกประเภทข้อความ การประมวลผลภาพ และการประมวลผลเสียง Naïve Bayes ทำงานโดยใช้ทฤษฎีความน่าจะเป็นของ Bayes ในการคำนวณความน่าจะเป็นของข้อมูลตัวอย่างในแต่ละคลาส จากนั้นเลือกคลาสที่มีความน่าจะเป็นสูงที่สุดเป็นคำตอบ ทฤษฎีความน่าจะเป็นของ Bayes สามารถเขียนเป็นสมการได้ดังนี้

$$P(H|E) = \frac{P(H) \times P(E|H)}{P(E)}$$

โดยที่ H และ E คือเหตุการณ์ และ $P(E) \neq 0$

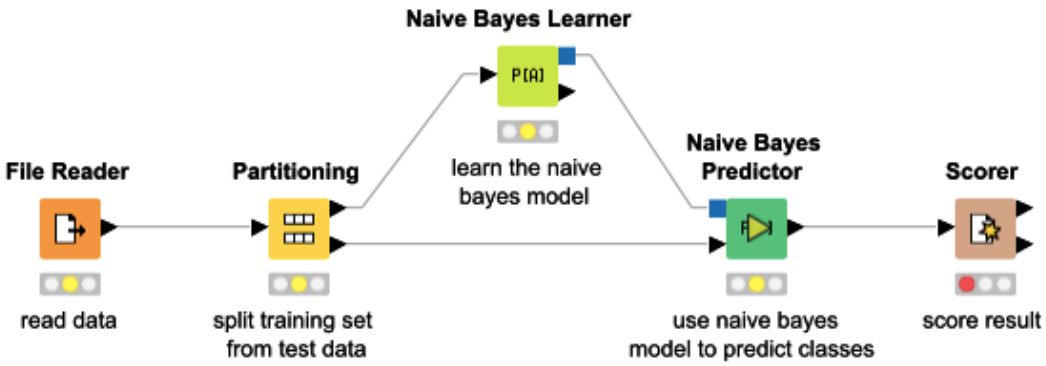
$P(H|E)$ คือ ความน่าจะเป็นของเหตุการณ์ H เกิดขึ้นเมื่อเหตุการณ์ E เกิดขึ้นแล้ว

$P(H)$ คือ ความน่าจะเป็นของเหตุการณ์ H เกิดขึ้นโดยไม่คำนึงถึงเหตุการณ์ E

$P(E|H)$ คือ ความน่าจะเป็นของเหตุการณ์ E เกิดขึ้นเมื่อเหตุการณ์ H เกิดขึ้นแล้ว

$P(E)$ คือ ความน่าจะเป็นของเหตุการณ์ E เกิดขึ้นโดยไม่คำนึงถึงเหตุการณ์ H

อย่างไรก็ตาม Naïve Bayes มีข้อสันนิษฐานสำคัญประการหนึ่งคือ คุณสมบัติต่างๆ ของข้อมูลตัวอย่างเป็นอิสระต่อกัน หมายความว่า คุณสมบัติหนึ่งไม่ได้ส่งผลต่อคุณสมบัติอื่น สมมติฐานนี้อาจไม่ถูกต้องเสมอไปสำหรับข้อมูลจริง อย่างไรก็ตาม Naïve Bayes ก็ยังมีประสิทธิภาพสูงสำหรับข้อมูลจำนวนมาก กระบวนการทำงานแสดงได้ดังภาพ [9], [10]



ภาพประกอบ 4 กระบวนการทำงาน Naïve Bayes

2.1.7 Logistic Regression

Logistic regression (LR) เป็นเทคนิคการเรียนรู้ของเครื่องที่ใช้ในการจำแนกประเภทข้อมูลแบบสองทาง (Binary) โดยที่ผลลัพธ์สามารถเป็นหนึ่งในสองค่าที่เป็นไปได้ เช่น ใช่/ไม่ หรือ จริง/เท็จ เป็นต้น อัลกอริทึม Logistic regression นี้ใช้สมการเชิงเส้นเพื่อคำนวณความน่าจะเป็นของข้อมูลว่าจัดอยู่ในกลุ่ม (Class) ใด สมการเชิงเส้นที่ใช้เรียกว่าฟังก์ชัน Logit จะแปลงความน่าจะเป็นให้เป็นค่าเชิงเส้นสามารถคำนวณได้ ดังสมการ

$$logit(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

โดยที่

p คือ ความน่าจะเป็นของเหตุการณ์ (เช่น ความคิดเห็นเชิงบวก, สำเร็จ)

β_0 คือ ค่าคงที่

$\beta_1, \beta_2, \ldots, \beta_k$ คือ ค่าสัมประสิทธิ์ของตัวแปรอิสระ

$X_1, X_2, \ldots, X_k$ คือ ค่าของตัวแปรอิสระ

ค่าสัมประสิทธิ์ของตัวแปรอิสระสามารถตีความได้ ดังนี้

1) ค่าสัมประสิทธิ์ที่เป็นบวกบ่งชี้ว่าตัวแปรอิสระมีความสัมพันธ์เชิงบวกกับเหตุการณ์หนึ่ง
2) ค่าสัมประสิทธิ์ที่เป็นลบบ่งชี้ว่าตัวแปรอิสระมีความสัมพันธ์เชิงลบกับเหตุการณ์หนึ่ง

ความน่าจะเป็นที่คำนวณได้จากฟังก์ชัน Logit สามารถแปลงกลับเป็นค่าความน่าจะเป็นได้ดังนี้

$$P(event) = \frac{1}{1 + e^{-logit(p)}}$$

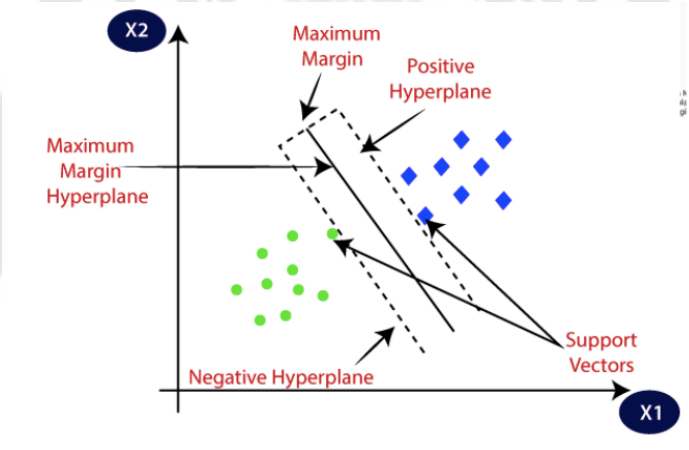
โดยที่ e คือ ลอการิทึมธรรมชาติ (มีค่าประมาณ 2.71828....)

ค่าความน่าจะเป็นที่คำนวณได้จากฟังก์ชันสามารถใช้เป็นเกณฑ์ในการจำแนกประเภทข้อมูลได้ เช่น หากค่าความน่าจะเป็นมากกว่า 0.5 ข้อมูลจะถูกจำแนกประเภทเป็นบวก หากค่าความน่าจะเป็นน้อยกว่า 0.5 ข้อมูลจะถูกจำแนกประเภทเป็นลบ เป็นต้น [11]

2.1.8 Support Vector Machine

Support Vector Machine (SVM) เป็นอัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised) ที่สามารถใช้สำหรับงานการจำแนกประเภท (Classification) และการวิเคราะห์การถดถอย (Regression) ได้ โดย SVM เหมาะสำหรับข้อมูลมิติสูงและปัญหาที่ข้อมูลไม่

สามารถแบ่งกลุ่มด้วยการใช้เส้นตรงได้ แนวคิดพื้นฐานของ SVM คือการค้นหาเส้นแบ่งข้อมูล (Hyperplane) ในพื้นที่ที่สามารถแยกจุดข้อมูลออกเป็นสองประเภทด้วยระยะขอบสูงสุด (Maximum Margin) ระยะขอบมาร์จิน (Margin) คือระยะห่างระหว่างเส้นแบ่งข้อมูลและจุดข้อมูลที่ใกล้ที่สุดของแต่ละประเภท อัลกอริทึม SVM ใช้เทคนิคที่หลากหลายเพื่อค้นหาเส้นแบ่งที่มีระยะขอบสูงสุด รวมถึงฟังก์ชั่น Kernel และ Soft margin หากข้อมูลที่ไม่สามารถแบ่งกลุ่มได้ด้วยเส้นตรง (Linear) สามารถเพิ่มวิธีการ Kernel หลักการคือจะปรับให้จุดข้อมูลเดิมไปยังพื้นที่มิติสูงขึ้นซึ่งสามารถทำให้แยกจุดข้อมูลออกได้ในเชิงเส้นมากขึ้น Soft margin เป็นตัวจัดการกับข้อมูลที่มีสัญญาณรบกวน (Noisy) และค่าผิดปกติ (Outliers) โดยอนุญาตให้จุดข้อมูลจำนวนเล็กน้อยตกอยู่ด้านที่ผิดของเส้นแบ่งได้ ซึ่งจะใช้พารามิเตอร์ค่า C ในการปรับขนาดความกว้างของเส้นแบ่งเพื่อให้ได้ระยะขอบที่เหมาะสมได้ [12] โครงสร้างของอัลกอริทึมSVMสามารถแสดงได้ดังภาพ [13]

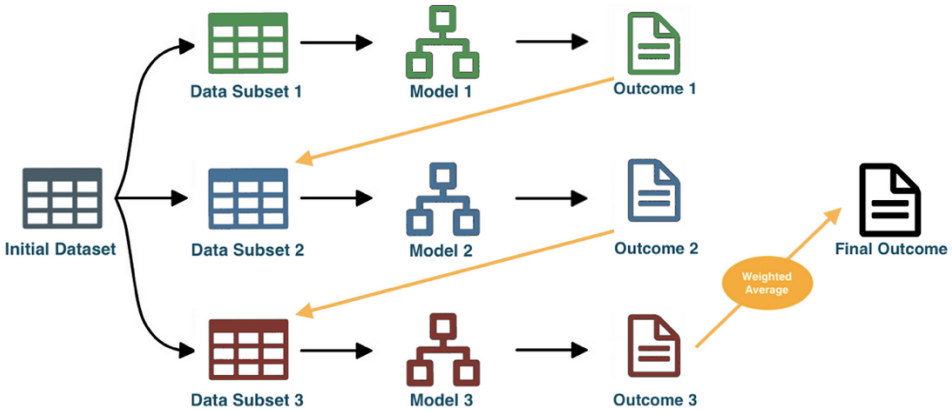


ภาพประกอบ 5 โครงสร้างอัลกอริทึม SVM

2.1.9 eXtreme Gradient Boosting (XGBoost)

XGBoost เป็นอัลกอริทึมการเรียนรู้แบบบูสต์ (Boosting) ที่พัฒนาต่อยอดมาจากอัลกอริทึม GBDT (Gradient Boosting Decision Tree) หลักการพื้นฐานคือการนำต้นไม้ตัดสินใจ (Decision Tree) หลายๆ ต้นที่มีความแม่นยำต่ำ (Low Accuracy) มารวมกันเป็นโมเดลที่มีความแม่นยำสูงขึ้น จุดเด่นของ XGboost คือการใช้แนวคิดของการไล่ระดับความชัน (Gradient Descent) ในการสร้างต้นไม้แต่ละต้น โดยอาศัยต้นไม้ที่สร้างในรอบก่อนหน้าเป็นฐาน แล้วปรับไปในทิศทางที่ลดค่าฟังก์ชันเป้าหมาย (Objective Function) ให้ต่ำสุด เมื่อผ่านการเรียนรู้ของต้นไม้ตัดสินใจหลายๆ ต้น ทำให้ค่าความผิดพลาด (Loss) ลดลงอย่างต่อเนื่อง และได้โมเดลการทำนาย

มีความลึกมากพอหรือไม่มีค่าความผิดพลาดเกิดขึ้นในต้นไม้ตัดสินใจก่อนหน้าแล้ว [14] สามารถแสดงโครงสร้างได้ดังภาพ [15]



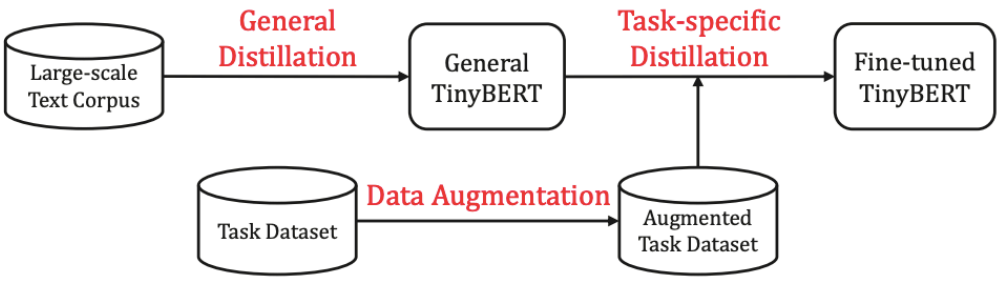
ภาพประกอบ 6 โครงสร้างของ XGBoost

2.1.10 DistilBERT Base Model

DistilBERT เป็นโมเดลประเภท Transformer ซึ่งมีขนาดเล็กและทำงานได้เร็วกว่า BERT โดยผ่านการเรียนรู้แบบไม่ต้องมีผู้สอน (self-supervised learning) จากคลังข้อมูลเดียวกันกับ BERT base model โดยอาศัย BERT base model เป็นโมเดลต้นแบบ (teacher) หมายความว่า DistilBERT ได้รับการฝึกฝนเบื้องต้นจากข้อความดิบเท่านั้น โดยไม่ต้องอาศัยการติดป้ายกำกับ (labels) ข้อมูลกระบวนการดังกล่าวใช้การสร้างข้อมูลป้อนเข้าและข้อมูลติดป้ายกำกับโดยอัตโนมัติจากข้อความด้วย BERT base model การเรียนรู้แบบกลั่นกรอง (Distillation) ได้รับการฝึกฝนเบื้องต้นด้วยวัตถุประสงค์ 3 ประการ ได้แก่

1) Distillation loss: ฝึกโมเดลให้ส่งผลลัพธ์เป็นความน่าจะเป็น (probabilities) ที่ตรงกับ BERT base model
2) Masked language modeling: MLM: เป็นส่วนหนึ่งของการคำนวณ Loss ในการฝึกฝนเบื้องต้นของ BERT base model โดยกระบวนการนี้จะสุ่มปิดบังคำศัพท์ 15% ในประโยคที่ป้อนเข้าสู่โมเดล จากนั้นประมวลผลประโยคทั้งประโยคที่ถูกปิดบังนั้นผ่านโมเดล และทำนายคำศัพท์ที่ถูกปิดบัง กระบวนการ MLM ช่วยให้โมเดลเรียนรู้การแทนค่าประโยคแบบทิศทางสองทาง (Bidirectional Representation)
3) Cosine embedding loss: ฝึกโมเดลให้สร้างสถานะที่ซ่อนอยู่ (hidden states) ให้ใกล้เคียงกับ BERT base model มากที่สุด

4) ด้วยวิธีการดังกล่าว DistilBERT จึงเรียนรู้การแทนค่าภายในของภาษาอังกฤษได้เช่นเดียวกับโมเดลต้นแบบแต่ใช้เวลาน้อยกว่าในการประมวลผล [16] สามารถแสดงโครงสร้างได้ดังภาพ [17]



ภาพประกอบ 7 กระบวนการเรียนรู้ของ DistillBERT

2.1.12 การวัดประสิทธิภาพโมเดล

การวัดประสิทธิภาพของโมเดล (Model Performance Evaluation) และแสดงผลด้วยเมทริกซ์ (Metrix) เป็นกระบวนการที่สำคัญในการประเมินความสามารถของโมเดลหรือการเรียนรู้ของเครื่อง การทดสอบประสิทธิภาพโมเดลช่วยให้เราเข้าใจว่าโมเดลของเราทำงานอย่างไรในการทำนายหรือการจำแนกประเภทข้อมูลซึ่งสามารใช้ตัววัดประสิทธิภาพโมเดลได้ดังนี้

2.1.12.1) เมทริกซ์ความสับสน (Confusion Matrix) เป็นตารางที่ใช้ในการประเมินประสิทธิภาพของโมเดลการจำแนกประเภทที่แสดงอยู่ในรูปแบบของตาราง แถวของเมทริกซ์ความสับสนแสดงถึงคลาสทำนาย และคอลัมน์แสดงถึงคลาสที่จริง สามารถแสดงได้ ดังตาราง [18]

ตาราง 1 เมทริกซ์ความสับสน

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

โดยที่

True Positives (TP) คือ จำนวนข้อมูลที่ทำนายว่าเป็นบวก และมีค่าเป็นบวก

True Negatives (TN) คือ จำนวนข้อมูลที่ทำนายว่าเป็นลบ และมีค่าเป็นลบ

False Positives (FP) คือ จำนวนข้อมูลที่ทำนายว่าเป็นจริง แต่มีค่าเป็นลบ

False Negatives (FN) คือ จำนวนข้อมูลที่ทำนายว่าเป็นลบ แต่มีค่าเป็นบวก

จากค่าดังกล่าวสามารถนำมาใช้คำนวณมาตรวัดต่าง ๆ เพื่อประเมินประสิทธิภาพของโมเดลการจำแนกประเภท ซึ่งมาตรวัดที่ใช้ในงานวิจัยมี ดังนี้ [18]

1) ค่าความแม่นยำ (Accuracy) วัดความถูกต้องโดยรวมของโมเดล โดยทั่วไปตัววัดความแม่นยำจะวัดอัตราส่วนของการทำนายที่ถูกต้องต่อจำนวนการทำนายทั้งหมด คำนวณได้ดังสมการ

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2) ค่าความเที่ยง (Precision) วัดความถูกต้องของโมเดล โดยใช้การวัดรูปแบบเชิงบวกที่ทำนายได้อย่างถูกต้องจากรูปแบบที่ทำนายทั้งหมดในระดับบวก คำนวณได้ดังสมการ

$$Percision = \frac{TP}{TP + FP}$$

3) ค่าเรียกคืน (Recall) ใช้เพื่อวัดส่วนของรูปแบบเชิงบวกที่จำแนกได้อย่างถูกต้อง คำนวณได้ดังสมการ

$$Recall = \frac{TP}{TP + TN}$$

4) ค่าเอฟวัน (F1-score) หรือ F-Measure วัดประสิทธิภาพโดยรวมระหว่างค่าความเที่ยงและค่าเรียกคืน คำนวณได้ดังสมการ

$$F1 - score = \frac{2 \times Percision \times Recall}{Percision + Recall}$$

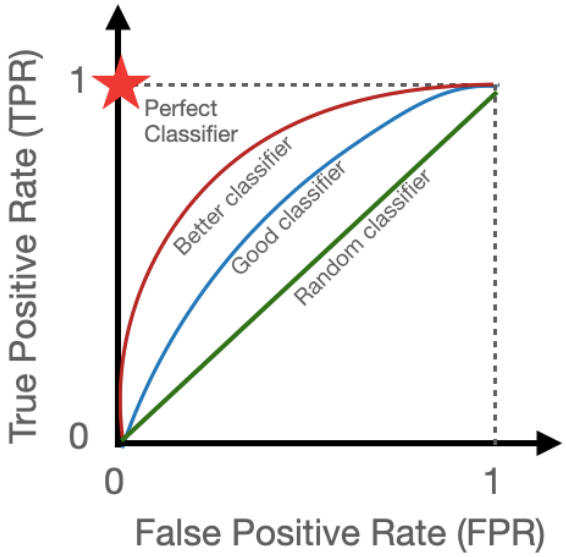
2.2.12.2) Receiver Operating Characteristic (ROC CUVE) เป็นเส้นที่ใช้วัดประสิทธิภาพของโมเดลแบบ Classification ว่าสามารถทำนายได้อย่างแม่นยำมากน้อยแค่ไหน

โดยทั่วไปนิยมวัดประสิทธิภาพของโมเดลแบบสองทาง (Binary) โดยสร้างกราฟความสัมพันธ์ระหว่าง true positive rate (Sensitivity) กับ false positive rate (1–Specificity) จากสมการ

$$\text{Sensitivity} = Recall$$

$$\text{Specificity} = \frac{TN}{FP + TN}$$

โดยการแปรค่าจุดตัด (cut-off point) ที่ใช้ต่าง ๆ ได้ดังภาพ [19]



ภาพประกอบ 8 Receiver Operating Characteristic (ROC CUVE)

เส้นทแยงมุม (Diagonal) จุดบนเส้นทแยงมุมหมายถึงตัวจำแนกประเภทที่ประสิทธิภาพไม่ดีไปกว่าการเดาสุ่มจุดที่ (0, 1) แทนการจำแนกประเภทที่สามารถจำแนกประเภทได้อย่างสมบูรณ์แบบ ซึ่งโดยทั่วไปจุดที่อยู่สูงบนแกน Y (TPR สูง) และต่ำบนแกน X (FPR ต่ำ) ถือว่าอยู่ระดับที่ดี [20]

2.2 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่ใช้วิธีการจำแนกประเภทข้อความเกี่ยวข้องกับความรู้สึกด้วยวิธีการเรียนรู้ด้วยเครื่อง สามารถสรุปผลได้ ดังตาราง

ตาราง 2 สรุปงานวิจัยที่ทำการศึกษา

No	Paper Topic	Year	Objective	Label	Data Source	Model Techniques
1	Business Review Classification Using Sentiment Analysis	2016	Comparative study on the effectiveness of the ensemble methods for reviews sentiment classification.	2 Class: - Positive - Negative	Yelp: Yelp Challenge	TF-IDF, Bag of Words, Naïve Bayes, Linear SVC, Logistic Regression and SGD
2	Sentiment Analysis using supervised classification algorithms	2017	Comparison of five supervised classification algorithms	2 Class: - Positive - Negative	SMS Spam	Decision Tree, SVM, Naïve Bayes, PART and Logistic Regression
3	A survey of sentiment analysis techniques	2017	Survey of main approaches used for sentiment classification	-	-	-
4	Sentiment Analysis on Twitter Data using KNN and SVM	2017	Develop such a functional classifier which can correctly and automatically classify the sentiment of an unknow tweet	2 Class: - Positive - Negative	Twitter	K-nearest neighbor and Support Vector Machine

ตาราง 3 สรุปงานวิจัยที่ทำการศึกษา (ต่อ)

No	Paper Topic	Year	Objective	Label	Data Source	Model Techniques
5	Comparative Study on Machine Learning Algorithms for Sentiment Classification	2018	Investigates the different ways of sentiment analysis from customer' review using machine learning algorithms	2 Class: - Positive - Negative	- IMDB movie review - Amazon Book review	Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Logistic Regression, Random Forest, SGD, SNM
6	A comparison of features extraction methods for Arabic sentiment analysis	2019	Analyze the impact of feature extraction methods on the performance of sentiment analysis	3 Class: - Positive - Negative - Neutral	ElecMorocco2017	Bag-of-Word, TF-IDF, AraVec, Logistic Regression, Extra Trees and Support Vector Machine
7	Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning	2020	Analyze the reviews and classify them with respect to category	2 Class: - Positive - Negative	Facebook: KARACHI's Restaurants Cafes Dhabas HBFE & Takeouts	Naïve Bayes Classification (NBC), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest
8	Emotion Analysis and Classification of Movie Reviews Using Data Mining	2020	Model for Classification of movie reviews	2 Class: - Positive - Negative	Movie reviews	TF-IDF, Naïve Bayes, Random Forest and J48

ตาราง 4 สรุปงานวิจัยที่ทำการศึกษา (ต่อ)

No	Paper Topic	Year	Objective	Label	Data Source	Model Techniques
9	Tourist Place Reviews Sentiment Classification Using Machine Learning Techniques	2020	Compared of performance for combinations of feature extraction and classification algorithms.	3 Class: - Positive - Negative - Neutral	Tourism reviews	CountVectorization, TFIDFVectorization, Naïve Bayes, Support Vector Machine and Random Forest
10	Sentiment analysis and classification of Indian famers' protest using twitter data	2021	Model to categorize and analyze the sentiments-based tweets on the protest	2 Class: - Positive - Negative	Twitter: famers' protest	Bag of Word, TF-IDF, Naïve Bayes, Decision Trees, Random Forests and Support Vector Machine

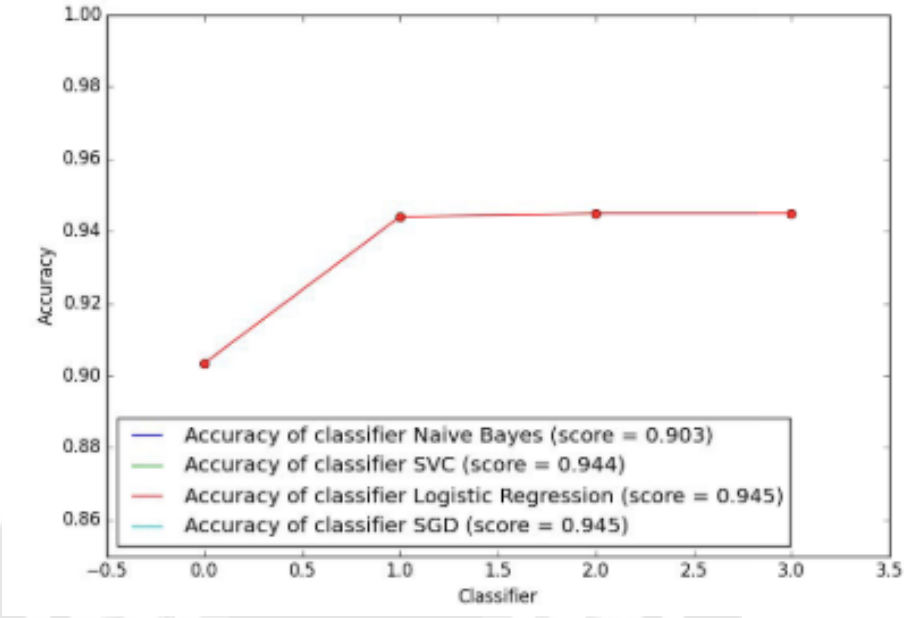
2.2.1 Business Reviews Classification Using Sentiment Analysis

งานวิจัยนี้นำเสนอแนวทางการวิเคราะห์ความรู้สึกต่อการจัดประเภทรีวิวธุรกิจโดยใช้ชุดข้อมูลรีวิวขนาดใหญ่ที่จัดทำโดย Yelp: ชุดข้อมูล Yelp Challenge โดยใช้วิธีการแยกคุณลักษณะ 2 วิธี ได้แก่ TF-IDF (Term frequenc-inverse document frequency) และ bag of Words และจำแนกประเภทความรู้สึกด้วยวิธีการเรียนรู้ของเครื่อง 4 อัลกอริทึม ได้แก่ Naïve Bayes, Support vector machines (SVM), Linear Support Vector Classification (SVC), Logistic regression และ Stochastic Gradient Descent Classifier (SGD) จากนั้นศึกษาเปรียบเทียบประสิทธิภาพของวิธีการทั้งหมดสำหรับการจำแนกประเภทความคิดเห็น โดยใช้การให้ดาวเป็นตัวแบ่งประเภทความคิดเห็นซึ่ง คะแนนเฉลี่ยของรีวิว Yelp อยู่ที่ประมาณ 3.7 การให้ดาวของการรีวิวธุรกิจเป็นจำนวนเต็มตั้งแต่ 1 ถึง 5 ผู้วิจัยได้ตัดสินใจไม่ใช้ความคิดเห็นที่คะแนนดาวเท่ากับ 3 และให้คะแนนทั้งหมดที่สูงกว่า 4 ถือเป็นความรู้สึกเชิงบวก ส่วนคะแนนทั้งหมดที่ต่ำกว่า 3 ถือเป็นความรู้สึกเชิงลบ เราได้รับชุดข้อมูลที่มีบทวิจารณ์ 1,346,545 ความคิดเห็น โดยผู้วิจัยมีการเปรียบเทียบค่าความแม่นยำ (Accuracy) ของแต่ละโมเดลที่มีการใช้วิธีการจำแนกคุณลักษณะแตกต่างกันซึ่งวิธีการจำแนกคุณลักษณะ TF-IDF จะใช้ร่วมกับอัลกอริทึม Linear SVC, SGD และ Logistic regression ส่วน Naïve Bayes จะใช้ร่วมกับวิธี Bag of Word พบว่าวิธีการจำแนกคุณลักษณะด้วยการทำ Unigrams, Stop words, Removeing punctuations และ Handling Negations รวมกันทำให้โมเดล SGD มีค่าความแม่นยำถึง 92% แสดงผลลัพธ์ได้ดังตาราง

ตาราง 5 ผลการเปรียบเทียบอัลกอริทึมที่ใช้วิธีการสกัดข้อมูลแตกต่างกันของ Business Reviews

Features	Naïve Bayes	Linear SVC	Logistic Regression	SGD
Unigrams + Stop words + without removing punctuations + Without handing Negations	0.881	0.917	0.893	0.914
Unigrams + Stop words + without removing punctuations + Handing Negation	0.883	0.919	0.898	0.911
Unigram + Stop words + removing punctuations + Handing Negations	0.897	0.923	0.900	0.926
Unigrams + POS + WSD	0.787	0.786	0.787	0.786

หลังจากนำวิธีการจำแนกคุณลักษณะที่ดีที่สุดมาใช้และทำการทดสอบกับชุดข้อมูลทดสอบของ Yelp Challenge โดยวิธี 3-Fold Cross Validatin ร่วมกับวิธีการจำแนกคุณลักษณะพบว่า Linear SVC และ SGD มีความแม่นยำ (Accuracy) สูงถึง 94.4% ดังภาพ



ภาพประกอบ 9 ผลการเปรียบเทียบอัลกอริทึมของ Business Review

ผู้วิจัยคิดว่าความแม่นยำของระบบยังสามารถปรับปรุงได้โดยการใช้ Bigrams หรือ Trigrams word chunks หรือแม้กระทั่ง Part-of-Speech (POS) มีคุณสมบัติเพื่อแยกความแตกต่างระหว่างคุณสมบัติคำเดียวกันที่ใช้เป็น POS ที่แตกต่างกัน [21]

2.2.2 Sentiment Analysis using supervised classification algorithms

งานวิจัยนี้ได้ทำการวิเคราะห์ความรู้สึกที่เกิดจากชุดข้อมูล SMS Spam Collection จำนวนทั้งหมด 5,574 SMS แต่นำมาใช้ในงานวิจัยเพียง 200 SMS เท่านั้นเพื่อทำการจำแนกความรู้สึกออกเป็นสองประเภท ได้แก่ เชิงบวกโดยระบุเป็น Ham และเชิงลบระบุเป็น Spam จากนั้นทำการเปรียบเทียบอัลกอริทึมด้วย Supervised Classification Algorithms ทั้ง 5 อัลกอริทึม ได้แก่ PART, Support Vector Machine, Decision Tree, Naive Bayes และ Logistic Regression จากผลการศึกกางานวิจัยนี้พบว่าว่าอัลกอริทึม Logistic Regression ได้ค่าความแม่นยำ (Accuracy) ที่จำแนกความรู้สึกได้ถูกต้องสูงที่สุดอยู่ที่ 96.5% ผลลัพธ์เปรียบเทียบของอัลกอริทึมแสดงได้ดังตาราง [10]

ตาราง 6 ผลการเปรียบเทียบอัลกอริทึมของ SMS [10]

	C.C.M	I.C.M	A (%)
PART	181	19	90.5
D.T	178	22	89
N.B	190	10	95
L.R	193	7	96.5
SVM	192	8	96

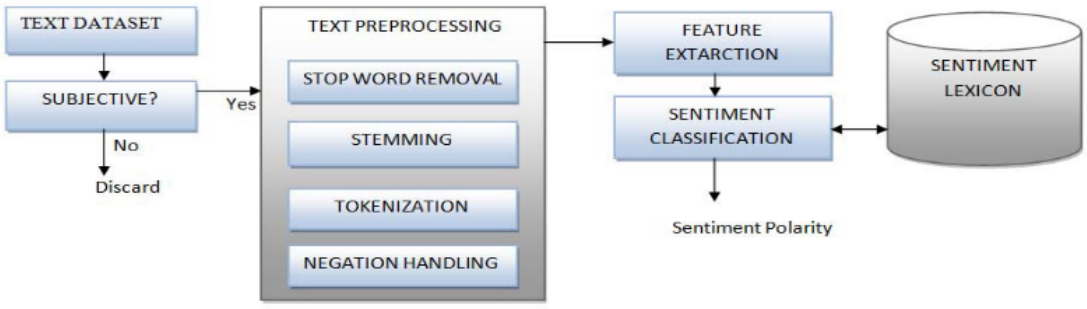
C.C.M.: Correctly Classified Messages

I.C.M.: Incorrectly Classified Messages

A.: Accuracy

2.2.3 A survey of sentiment analysis techniques

งานวิจัยนี้นำเสนอแบบสำรวจการวิเคราะห์ความรู้สึกและอัลกอริทึมการจำแนกประเภท จากการสำรวจนี้สรุปได้ว่าการวิเคราะห์ความรู้สึกสามารถนำไปประยุกต์ใช้ในธุรกิจเพื่อให้บริษัทได้รับข้อเสนอของลูกค้าและสามารถปรับปรุงผลิตภัณฑ์ได้ตามความต้องการ ช่วยสนับสนุนในการตัดสินใจ จากความคิดเห็นที่ถูกสกัดออกมาสามารถช่วยให้เราตัดสินใจได้ในหลายประการ และยังสามารถนำไปพยากรณ์และวิเคราะห์แนวโน้มได้ กระบวนการวิเคราะห์ความรู้สึกจะเริ่มต้นจากการทำการเลือกคุณลักษณะ (Feature Selection) ด้วยการสกัดคุณลักษณะออกมาจากข้อความด้วยวิธีการต่าง ๆ เช่น การทำ N-grams, POS tagging หรือ Stop words เป็นต้น จากนั้นมาทำการจำแนกประเภทความรู้สึกซึ่งมีสองวิธีหลักได้แก่ การจำแนกประเภทความรู้สึกเป็นวิธีการที่ใช้ในการแยกแยะอารมณ์หรือความคิดเห็นจากข้อความ ซึ่งมีสองวิธีหลักที่ใช้ในการจำแนกประเภทความรู้สึก ได้แก่ subjective lexicon คือ คือคำศัพท์ที่มีการกำหนดคะแนนบ่งชี้ถึงความรู้สึกเชิงบวก เชิงลบ เป็นกลาง และไม่ชัดเจน วิธีการนี้จะรวมคะแนนของคำศัพท์เชิงอารมณ์ในข้อความเพื่อกำหนดความรู้สึกโดยรวมของข้อความ และ machine learning คือ คือเทคนิคการจำแนกประเภทโดยอัตโนมัติ โดยจะใช้คุณสมบัติของข้อความเพื่อจำแนกประเภท ซึ่งแบ่งออกเป็นสองประเภท ได้แก่ Supervised learning และ Unsupervised learning โดย machine learning มีประสิทธิภาพมากกว่า subjective lexicon สามารถแสดงกระบวนการวิเคราะห์ความรู้สึกได้ดังภาพ



ภาพประกอบ 10 กระบวนการวิเคราะห์ความรู้สึก [3]

จากการสำรวจผู้วิจัยพบว่า SVM และ naïve bayes เป็นอัลกอริทึมที่ได้รับความนิยมมากที่สุดใช้สำหรับการจำแนกประเภทความรู้สึก และการวิเคราะห์ความรู้สึกของทวิตเตอร์ (Twitter) เป็นที่นิยมมาก รวมถึงชุดข้อมูลจากเว็บไซต์ต่างๆ เช่น Amazon, IMDB, Flipkart ถูกนำมาใช้กันอย่างแพร่หลายสำหรับการวิเคราะห์ความรู้สึกผู้วิจัยได้สรุปการสำรวจบทความงานวิจัยต่าง ๆ ไว้ดังตาราง [3]

ตาราง 7 สรุปบทความที่ได้ทำการสำรวจ [3]

Year	Task	Data set	Algorithm
2011	Sentiment analysis	Digital camera reviews	Multi class SVM
2011	Sentiment analysis	Training data in Chinese	Sentiment
2011	Sentiment Classification	Movie reviews	Lexicon based, semantic
2011	Sentiment analysis	Product reviews	Statistical (Machine learning), semantic
2012	Feature selection	Movie reviews	Statistical, maximum entropy
2012	Sentiment Classification	Restaurant reviews	Naïve bayes, SVM
2012	Sentiment analysis	News	Lexicon based
2012	Emotion detection	Blogs data	Corpus based
2012	Emotion detection	Emotions corpus	Lexicon based, SVM
2013	Sentiment Classification	Movie, camera, book, GPS reviews	Artificial neural network, SVM
2013	Sentiment Classification	Tweets and movie review	SVM, Naïve bayes
2013	Sentiment analysis	Facebook data	Lexicon based, Machine learning
2015	Sentiment analysis	Tweets	Hybrid (lexicon + learning algorithm)
2015	Sentiment analysis	Movie, book, product reviews	SVM
2016	Sentiment analysis	Tweets	Lexicon based
2016	Sentiment analysis	Starbucks twitter dataset	Dynamic architectural artificial neural networks

2.2.4 Sentiment Analysis on Twitter Data using KNN and SVM

งานวิจัยนี้มีจุดมุ่งหมายของคือการพัฒนาตัวแยกประเภทการทำงานซึ่งสามารถจำแนกประเภทของทวีตที่ไม่รู้จักได้อย่างถูกต้องและอัตโนมัติ เนื่องจากผู้ใช้หลายล้านคนแบ่งปันความคิดเห็นในหัวข้อต่างๆ โดยใช้ไมโครบล็อกทุกวัน Twitter เป็นไซต์ไมโครบล็อกยอดนิยมที่อนุญาตให้ผู้ใช้จำกัดตัวอักษรได้ 140 ตัว ข้อจำกัดประเภทนี้ทำให้ผู้ใช้มีความกระชับและแสดงออกในเวลาเดียวกัน ด้วยเหตุผลดังกล่าว มันจึงกลายเป็นแหล่งข้อมูลมากมายสำหรับการวิเคราะห์ความรู้สึก ผู้วิจัยได้รวบรวมความคิดเห็นสำหรับใช้ในงานวิจัยมากกว่า 1,000 ทวิต และทำการสกัดคุณลักษณะของข้อความด้วย Word Feature, N-Gram Feature, Pattern Features, Punctuation Feature และ Key-based feature ร่วมกับเทคนิคในการจำแนกประเภทความรู้สึกสองวิธีวิธีหนึ่งเรียกว่า Sentiment Classification Algorithm (SCA) ได้แก่ K-Nearest Neighbor (KNN) และอีกวิธีหนึ่งใช้ Support Vector Machine (SVM) ในทั้งสองอัลกอริทึม ใช้วิธี Five-fold cross validationและ Confusion Matrix เพื่อคำนวณความแม่นยำ ต่อมาเปรียบเทียบเทคนิคทั้งสองเกี่ยวกับระดับความแม่นยำในการตรวจจับความรู้สึก พบว่าอัลกอริทึมตัวแยกประเภทความรู้สึก (SCA) ทำงานได้ดีกว่า SVM แสดงผลการปรียบเทียบประสิทธิภาพอัลกอริทึมกับข้อมูล 1,000 ทวิต ได้ดังตาราง [22]

ตาราง 8 ผลการเปรียบเทียบประสิทธิภาพอัลกอริทึม [22]

Method	Number of tweets	Precision	Recall	F-score	TRP	FPR	Accuracy
Algorithm	1000	0.81	0.76	0.78	0.79	0.13	79.99%
KNN with normalization (4 feature)	1000	0.83	0.75	0.79	0.81	0.14	80.80%
KNN with normalization and key base (5 feature)	1000	0.85	0.81	0.83	0.68	0.17	84.32%
SVM with (4 feature)	1000	0.65	0.69	0.61	0.78	0.49	58.79%
SNM with normalization (4 feature)	1000	0.55	0.73	0.62	0.80	0.52	58.39%
SVM with normalization and keyword base (5 feature)	1000	0.62	0.79	0.70	0.79	0.50	67.03%
SVM with normalization and keyword base (5 feature) with grid search	1000	0.72	0.89	0.80	0.70	0.30	77.97%

2.2.5 Comparative Study on Machine Learning Algorithms for Sentiment Classification

งานวิจัยนี้จะศึกษาแนวทางต่าง ๆ ในการวิเคราะห์ความรู้สึกจากรีวิวของลูกค้าโดยใช้เทคนิคการเรียนรู้ของเครื่อง สำหรับจัดประเภทข้อความตามความรู้สึกโดยรวม โดยผู้วิจัยจะพิจารณาออกเป็น 2 คลาส ได้แก่ การคาดการณ์ว่าความคิดเห็นหรือรีวิวเป็นความคิดเห็นเชิงบวกหรือความคิดเห็นเชิงลบ ในงานวิจัยใช้ชุดข้อมูลสาธารณะที่เป็นที่นิยม 2 ชุดได้แก่ ชุดข้อมูลรีวิวภาพยนตร์ที่เก็บรวบรวมโดยเว็บไซต์ IMDB และชุดข้อมูลรีวิวหนังสือ Amazon ที่เก็บรวบรวมจากเว็บไซต์ Amazon จากนั้นทำการเตรียมข้อมูลด้วยกระบวนการต่าง ๆ เช่น ลบคำหยุด (Stop words) เปลี่ยนตัวพิมพ์ใหญ่เป็นตัวพิมพ์เล็ก และใช้ Stemming หรือ Lemmatizing เพื่อแปลงคำศัพท์ต่างๆ ให้เป็นรูปแบบเดียวกัน และแปลงข้อความเป็นตัวเลข โดยการแบ่งประโยคออกเป็นคำศัพท์และจัดกลุ่มคำศัพท์โดยใช้ unigrams (คำศัพท์เดี่ยว) ในการสร้างโมเดล Bag-of-Word จากนั้นจะคำนวณความถี่ของคำศัพท์แต่ละคำในข้อความ และกำหนดค่าตัวเลขเฉพาะให้กับคำศัพท์แต่ละคำ จากนั้นใช้โมเดลการเรียนรู้ของเครื่อง ได้แก่ Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Logistic Regression, Stochastic Gradient Descent (SGD), Linear Support Vector Machine (Linear SVM) และ Random Forest เพื่อจัดประเภทข้อความว่าเป็นข้อความเชิงบวกหรือเชิงลบ โดยงานวิจัยนี้จะใช้การทดสอบแบบ 5-fold cross-validation เพื่อฝึกอบรมและทดสอบโมเดล และใช้ค่าความแม่นยำ (Accuracy) ค่าความถูกต้องจำเพาะ (Precision) ค่าความสามารถจำแนก (Recall) และค่า F1 score ในการประเมินผลการทำงานของโมเดล จากผลพบว่า Linear SVM มีประสิทธิภาพในการวิเคราะห์ความรู้สึกที่ดีที่สุดเมื่อเทียบกับโมเดลการเรียนรู้ของเครื่องอื่นๆ โดยมีค่าความแม่นยำเฉลี่ยสูงสุดถึง 88.61% และค่า F1 score สูงสุดถึง 88.97% นอกจากนี้ยังพบว่า Logistic Regression และ Stochastic Gradient Descent มีประสิทธิภาพในการวิเคราะห์ความรู้สึกที่ใกล้เคียงกับ Linear SVM แสดงการเปรียบเทียบประสิทธิภาพดังตาราง [23]

ตาราง 9 สรุปผลการเปรียบเทียบประสิทธิภาพโมเดลของชุดข้อมูล IMDB และ Amazon[23]

Classifier	IMDB dataset (at t=5)	Amazon dataset
MNB	Matthew Correlation Coeff: 0.66998	Matthew Correlation Coeff: -0.80846
	Average Precision: 0.79402	Average Precision: 0.86854
	F1 Score: 0.81860	F1 Score: 0.90551
	Accuracy: 83.16%	Accuracy: 90.42%
BNB	Matthew Correlation Coeff: 0.66231	Matthew Correlation Coeff: 0.77882
	Average Precision: 0.78636	Average Precision: 0.84496
	F1 Score: 0.81910	F1 Score: 0.84496
	Accuracy: 82.91%	Accuracy: 88.93%
LR	Matthew Correlation Coeff: 0.76632	Matthew Correlation Coeff: 0.88910
	Average Precision: 0.83826	Average Precision: 0.88910
	F1 Score: 0.88321	F1 Score: 0.84496
	Accuracy: **88.32%**	Accuracy: **92.18%**
Linear SVM with parameter selection	Matthew Correlation Coeff: 0.77258	Matthew Correlation Coeff: 0.84364
	Average Precision: 0.84354	Average Precision: 0.88930
	F1 Score: 0.88582	F1 Score: 0.88930
	Accuracy: **88.63% at c=0.25**	Accuracy: **92.18% at c=2**
SGD	Matthew Correlation Coeff: 0.76891	Matthew Correlation Coeff: 0.84231
	Average Precision: 0.84251	Average Precision: 0.922165
	F1 Score: 0.88341	F1 Score: 0.922165
	Accuracy: **88.44%**	Accuracy: **92.11%**
RF	Matthew Correlation Coeff: 0.67370	Matthew Correlation Coeff: 0.67079
	Average Precision: 0.77716	Average Precision: 0.77732
	F1 Score: 0.83989	F1 Score: 0.84197
	Accuracy: 83.66%	Accuracy: 83.5%

2.2.6 Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning

งานวิจัยนี้ทำการวิเคราะห์ความคิดเห็นของลูกค้าเกี่ยวกับร้านอาหารต่าง ๆ ทั่วการาจี(Karachi) ซึ่งเป็นหนึ่งในเมืองที่ใหญ่ที่สุดของปากีสถาน สำหรับการวิจัยนี้ ความคิดเห็นของลูกค้าถูกรวบรวมจากแพลตฟอร์ม Facebook ชุมชนที่ชื่อ "SWOT's Guide to KARACHI's Restaurants Cafes Dhabas HBFE & Takeouts" ชุดข้อมูลประกอบด้วยความคิดเห็นของลูกค้าประมาณ 4,000 ความคิดเห็นที่เกี่ยวกับร้านอาหารเป็นภาษาอังกฤษ มีการจัดกลุ่มความคิดเห็นเป็น 4 กลุ่ม ได้แก่ รสชาติอาหาร (Food Taste), บริการ (services) คุ้มค่ากับราคา (value for money) และบรรยากาศ (Ambiance) มีการเตรียมข้อมูลด้วยการใช้เทคนิค NLP ด้วยฟังค์ชั่น Natural Language Toolkit (NLTK) ร่วมกับ Term Frequency Inverse Document Frequency (TF-IDF) ในการช่วยทำความสะอาดข้อมูลและสกัดคุณลักษณะ จากนั้นความคิดเห็นถูกจำแนก

ประเภทความรู้สึกออกเป็นเชิงบวกและเชิงลบ โดยพิจารณาจากลักษณะของความคิดเห็น และใช้อัลกอริทึมการเรียนรู้ของเครื่องจำแนกประเภท ดังนี้ Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM) และ Random Forest เพื่อทำการจำแนกประเภทความคิดเห็น ผลการวิจัยพบว่าอัลกอริทึม Random Forest ให้ผลลัพธ์ที่ดีที่สุดโดยมีความแม่นยำสูงที่สุดอยู่ที่ 95% ทั้งในการวิเคราะห์ความรู้สึกและการจำแนกประเภทความคิดเห็น ผลการเปรียบเทียบประสิทธิภาพแต่ละอัลกอริทึมแสดงได้ดังตาราง

ตาราง 10 ผลการวิเคราะห์ความรู้สึกชุดข้อมูลรีวิวร้านอาหารในการาจี [24]

Algorithm	IMDB The ratio of training versus testing data set												
		80:20				75:25				70:30			
	Sentiment	P%	R%	F%	Acc.%	P%	R%	F%	Acc.%	P%	R%	F%	Acc.%
Naïve Bayes	Positive	92	91	92	91	92	93	92	92	90	92	91	91
	Negative	91	92	92		92	92	92		92	90	91	
Logistic Regression	Positive	93	90	91	91	93	91	92	92	93	91	92	92
	Negative	90	92	92		91	93	92		91	93	92	
SVM	Positive	95	90	92	92	93	91	92	93	91	92	91	93
	Negative	90	95	92		91	93	92		93	91	92	
Random Forest	Positive	96	94	95	94	93	91	92	95	91	93	92	95
	Negative	94	95	95		91	93	92		93	91	92	

ตาราง 11 ผลการจำแนกประเภทตามกลุ่ม [24]

	IMDB The ratio of training versus testing data set															
Algorithm	Naïve Bayes				Logistic Regression				SVM				Random Forest			
Categories	P%	R%	F%	Acc. %	P%	R%	F%	Acc. %	P%	R%	F%	Acc. %	P%	R%	F%	Acc. %
Food Taste	92	91	92	84	92	93	92	86	90	92	91	89	95	97	96	94
Value for Money	91	92	92		92	92	92		92	90	91		94	91	92	
Ambiance	93	90	91		93	91	92		93	91	92		94	90	92	
Service	90	92	92		91	93	92		91	93	92		94	90	92	

ซึ่งผู้วิจัยวางแผนที่จะรวบรวมข้อมูลเพิ่มเติมจากเว็บไซต์เครือข่ายสังคมออนไลน์ต่างๆ และใช้การเรียนรู้เชิงลึกและเครือข่ายประสาทเทียมเพื่อให้ได้ประสิทธิภาพที่ดีกว่าในอนาคต [24]

2.2.7 A comparison of features extraction methods for Arabic sentiment analysis

งานวิจัยนี้ทำการวิเคราะห์ผลกระทบของวิธีการสกัดคุณลักษณะ (Feature Extraction) ต่อประสิทธิภาพของการวิเคราะห์ความรู้สึกที่เป็นภาษาอาหรับ ชุดข้อมูลที่ใช้คือ ElecMorocco2017 ซึ่งรวบรวมจากความคิดเห็นใน Facebook โดยใช้ Facebook Graph API ในระหว่างระยะเวลา 70 วันและมีความคิดเห็น 10,254 ความคิดเห็น ชุดข้อมูลเกี่ยวข้องกับการเลือกตั้งสภาผู้แทนสาธารณรัฐโมร็อกโก ที่เกิดขึ้นในวันที่ 7 ตุลาคม 2016 วิธีสกัดคุณลักษณะข้อความถูกสกัดออกโดยใช้เทคนิค Bag-of-Word, TF-IDF และ Aravec ซึ่ง Aravec เป็นชุดของโมเดล Word Embedding ในภาษาอาหรับที่ใช้ในการประมวลผลภาษาธรรมชาติ (NLP) ในภาษาอาหรับ คุณลักษณะที่สกัดออกมาจะถูกนำมาใช้สำหรับการวิเคราะห์ความรู้สึกโดยใช้ Logistic Regression (LR), Random Forest (RF), Extra Trees และ Support Vector Machine (SVM) จากนั้นทำการเปรียบเทียบประสิทธิภาพของโมเดล พบว่าการใช้วิธี TF-IDF ในการสกัดคุณลักษณะเป็นทางเลือกให้ประสิทธิภาพที่ดีที่สุดกับทุกอัลกอริทึมการจำแนกประเภท ส่วนวิธี Bag-of-Word มีประสิทธิภาพที่ดีในกรณีใช้กับตัวจำแนกประเภท SVM ที่ใช้พารามิเตอร์ nu อีกทั้ง Logistic Regression มีความแม่นยำที่ดีที่สุดในทั้งหมดโดยเปรียบเทียบกับตัวจำแนกอื่น ๆ และวิธี AraVec มีความแม่นยำต่ำสุดในวิธีการสกัดคุณลักษณะทั้งหมด ผู้วิจัยคิดว่าความแม่นยำที่ไม่ดีเป็นผลมาจากโมเดล AraVec ถูกฝึกใช้กับภาษาอาหรับมาตรฐานสมัยใหม่ (MSA) และชุดข้อมูลที่นำมาใช้ในงานวิจัยนี้ประกอบด้วยความคิดเห็นบางส่วนที่เขียนในภาษาอาหรับโมร็อกโกและอีกอย่างคือความคิดเห็นไม่ได้มาตราฐานเนื่องจากแต่ละคนมีรูปแบบการเขียนของตัวเอง ผลการเปรียบเทียบประสิทธิภาพโมเดลแสดงได้ ดังตาราง [25]

ตารางที่ 12 ผลการวิเคราะห์ความรู้สึกของภาษาอาหรับ[25]

Feature extraction methods		LR	RF	ETrees	SVM Linear	SVM RBF
BoW		81.27	78.44	79.61	77.91	65.33
TF-IDF		**82.10**	**79.61**	**80.25**	**81.86**	65.23
Aravec	T-CBOW	70.59	68.94	69.42	69.18	71.18
	T-SG	70.16	68.01	69.33	69.23	65.23
	W-CBOW	70.64	68.94	67.86	69.18	**71.74**
	W-SG	70.11	68.21	68.50	69.42	65.23

2.2.8 Emotion Analysis and Classification of Movie Reviews Using Data Mining

งานวิจัยนี้นำเสนอโมเดลสำหรับการจำแนกความคิดเห็นเกี่ยวกับภาพยนตร์โดยใช้เหมือข้อมูล (Data Mining) นอกจากนี้ยังเสนอวิธีการสร้างกลุ่มคำ (World Cloud) จากความถี่ของคำที่พบในรีวิวภาพยนตร์ เพื่อสนับสนุนการวิเคราะห์เรื่องที่น่าสนใจและความคิดเห็นของผู้รีวิวบางส่วน งานวิจัยนี้ใช้ข้อมูลรีวิวภาพยนตร์จากเว็บไซต์ Metacritic ข้อมูลรีวิวประกอบด้วยรีวิวจากภาพยนตร์21 เรื่อง ถูกแยกเป็นสองส่วนในการนำไปใช้เป็นชุดสำหรับฝึก (Training set) และชุดทดสอบ (Test set) ชุดฝึกประกอบด้วยรีวิว 462 รีวิวและชุดทดสอบประกอบด้วยรีวิว 238 รีวิว กระบวนการเตรียมข้อมูลเริ่มต้นด้วยการรวบรวมข้อมูลรีวิวโดยการลบสัญลักษณ์พิเศษและเตรียมข้อมูลเข้าสู่โปรแกรม Weka จากนั้นเปลี่ยนข้อความรีวิวเป็นข้อมูลโครงสร้างโดยใช้ตัวกรอง StringToWordVector ในกระบวนการเตรียมข้อมูลนี้จะรวมถึงการลบคำหยุด (Stop Words) ด้วยรายการคำหยุด Rainbow และเปลี่ยนคำให้เป็นคำรากศัพท์โดยใช้อัลกอริทึม Snowball Stemmer และทำการสกัดคุณลักษณะโดยใช้เทคนิค TF-IDF กำหนดค่าน้ำหนักหลังจากนั้นใช้อัลกอริทึม Naïve bayes, Random Forest และ J48 ในการจำแนกข้อมูลรีวิวเป็นกลุ่มบวกและกลุ่มลบ ผลลัพธ์ที่ได้ 80.25%, 79.83% และ 68.06% ตามลำดับ จากการวิจัยพบว่าอัลกอริทึม Naïve Bayes มีค่าความแม่นยำ (Accuracy) สูงสุดที่ 80%, ค่าความเที่ยง (Precision) อยู่ที่ 74% ค่าเรียกคืน (Recall) อยู่ที่ 74% และค่าเอฟ (F-Measure) อยู่ที่ 75.5% งานวิจัยนี้สามารถพัฒนาเป็นระบบที่ช่วยให้ผู้ใช้และนักลงทุนเข้าใจลักษณะเด่นและลักษณะเฉพาะที่มีความสำคัญต่อผู้ชมที่มีแรงจูงใจเพียงพอที่จะวิจารณ์ภาพยนตร์ สิ่งนี้สามารถช่วยในการตัดสินใจว่าจะดูหนังหรือไม่ หรือในด้านของนักลงทุนสามารถปรับปรุงหนังเรื่องต่อไปในด้านใดให้ตรงกับความสนใจของผู้ชมได้ [26]

2.2.9 Tourist Place Reviews Sentiment Classification Using Machine Learning Techniques

จากงานวิจัยได้ทำการวิเคราะห์ความรู้สึกเพื่อที่จะช่วยให้นักท่องเที่ยวสามารถตัดสินใจเลือกสถานที่ท่องเที่ยวที่จะไปท่องเที่ยวได้อย่างง่ายดาย การวิเคราะห์ความรู้สึกดำเนินการโดยใช้การเรียนรู้ของเครื่อง (Machine Learning) มีการรวบรวมรีวิว 3,209 รีวิวบนเว็บไซต์การท่องเที่ยวต่างๆ รีวิวแต่ละรีวิวมีคะแนนที่สอดคล้องกันร่วมมอยู่ด้วย จากการให้คะแนนผู้วิจัยได้นำใช้จำแนกประเภทความรู้สึกเริ่มต้น ดังนี้ หากการให้คะแนนมากกว่า 3 แสดงว่ามีความคิดเห็นเชิงบวก คะแนนเท่ากับ 3 จะเป็นกลาง และคะแนนน้อยกว่า 3 แสดงว่าความคิดเห็นเป็นเชิงลบ หลังจากได้ชุดข้อมูลที่มีระบุประเภทแล้วผู้วิจัยได้แบ่งชุดข้อมูลออกเป็นอัตราส่วน 80:20

สำหรับชุดข้อมูลสอน (train) และทดสอบ (test) ตามลำดับ ทุกข้อความรีวิวได้ดำเนินการประมวลผลข้อมูล (data preprocessing) เช่น การทำโทเค็น (tokenization) การหยุดคำ (stop word removal) การลบเครื่องหมายวรรคตอน (punctuation removal) การแตกคำ (stemming) การลดรูปของคำ (lemmatization) ตามด้วยการสกัดคุณลักษณะ (feature extraction) โดยใช้อัลกอริทึม CountVectorization และ TFIDFVectorization ผลลัพธ์ของการสกัดคุณลักษณะคือเวกเตอร์คุณลักษณะ (feature vector) แล้วนำไปใช้ร่วมกับอัลกอริทึมการจำแนกประเภท Naive Bayes (NB), Support Vector Machine (SVM) และ Random Forest (RF) ประสิทธิภาพของอัลกอริทึมได้รับการเปรียบเทียบโดยใช้พารามิเตอร์ต่างๆ เช่น ความแม่นยำ การเรียกคืน ความแม่นยำ และคะแนนเอฟวัน (F-1 Score) จากงานวิจัยพบว่าอัลกอริทึมการแยกคุณลักษณะ TFIDFVectorization ช่วยปรับปรุงความแม่นยำของอัลกอริทึมการจำแนกได้ดีกว่าเมื่อเปรียบเทียบกับ CountVectorization ในการจำแนกประเภทของความคิดเห็นเกี่ยวกับสถานที่ท่องเที่ยวเทคนิค TFIDFVectorization ร่วมกับอัลกอริทึม Random Forest ให้ความแม่นยำสูงสุด 86% สำหรับชุดข้อมูลการวิจัยนี้ สามารถแสดงผลการเปรียบเทียบประสิทธิภาพการจำแนกได้ดังตาราง

ตาราง 13 ผลลัพธ์เปรียบเทียบประสิทธิภาพการจำแนกประเภทความคิดเห็นของชุดข้อมูลรีวิวสถานที่ท่องเที่ยว [27]

Algorithm	Accuracy	Precision	Recall	F1-Score
CountVectorization + NB	82.71%	82%	83%	82%
TFIDFVectorization + NB	84.11%	83%	84%	83%
CountVectorization + SVM	80.21%	79%	80%	80%
TFIDFVectorization + SVM	85.35%	85%	85%	84%
CountVectorization + RF	85.51%	85%	86%	84%
TFIDFVectorization + RF	86.13%	86%	86%	85%

2.2.10 Sentiment analysis and classification of Indian farmers' protest using twitter data

งานวิจัยนี้ทำการวิเคราะห์ความรู้สึกของประชาชนต่อการประท้วงของเกษตรกรในอินเดีย โดยผู้วิจัยรวบรวมข้อมูลจากเว็บไซต์ Twitter เกี่ยวกับการประท้วงของเกษตรกร ทั้งหมด 18,000 ทวีตภายในช่วงระยะเวลา 4 เดือนที่ทวีตเกี่ยวกับการประท้วง จากนั้นมีการทำความ

สะอาดข้อมูลโดยลบพวกอักขระพิเศษต่าง ๆ ที่ไม่เกี่ยวข้องและใช้วิธี Lexicon เพื่อระบุประเภทของข้อความ วิธีการนี้สามารถคำนวณทิศทางความหมายและใส่ป้ายกำกับข้อความได้ว่าเป็นเชิงกลาง เชิงบวก หรือเชิงลบ จากแง่มุมของนิยาม และทำการสกัดคุณลักษณะเปลี่ยนข้อความเป็นเวกเตอร์โดยใช้ Bag of Words และ TF-IDF ร่วมกับอัลกอริทึมจำแนกประเภท Naive Bayes, Decision Trees, Random Forests และ Support Vector Machines ผลพบว่า Bag of Words ทำงานได้ดีกว่า TF-IDF และ Random Forest มีความแม่นยำในการจำแนกประเภทสูงสุด แสดงผลลัพธ์ประสิทธิภาพได้ดังตาราง

ตาราง 14 ผลลัพธ์ประสิทธิภาพการอัลกอริทึมจำแนกประเภทร่วมกับวิธี Bag-of-Word[28]

	Negative			Neutral			Positive		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Naïve Bayes	0.55	0.65	0.59	0.80	0.78	0.79	0.73	0.68	0.70
Decision Tree	0.73	0.65	0.69	0.83	0.89	0.86	0.78	0.76	0.77
Random Forest	0.99	0.92	0.95	0.95	0.99	0.97	0.98	0.95	0.96
SVC	0.83	0.65	0.73	0.82	0.98	0.89	0.87	0.73	0.79

ตาราง 15 ผลลัพธ์เปรียบเทียบประสิทธิภาพอัลกอริทึมการจำแนกประเภทความรู้สึกร่วมกับวิธี TF-IDF [28]

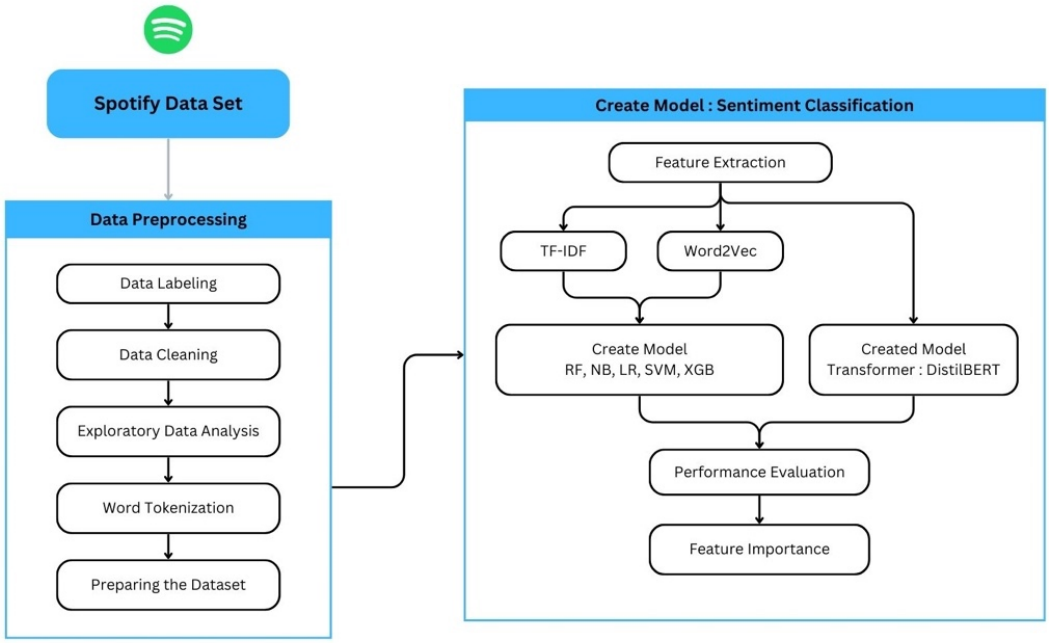
	Negative			Neutral			Positive		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Naïve Bayes	0.88	0.31	0.46	0.71	0.92	0.80	0.72	0.71	0.71
Decision Tree	0.71	0.63	0.67	0.83	0.88	0.86	0.76	0.74	0.75
Random Forest	0.96	0.93	0.94	0.97	0.98	0.97	0.95	0.96	0.96
SVC	0.84	0.60	0.70	0.80	0.97	0.88	0.89	0.76	0.82

ในบทความนี้ผู้วิจัยได้สำรวจวิธีการเพื่อเข้าใจอารมณ์ของผู้คนโดยการสร้างแบบจำลองวิเคราะห์อารมณ์และระบุทิศทางที่การประท้วง พบว่าส่วนใหญ่ของทวีตเป็นความรู้สึกเชิงกลาง ความรู้สึกเชิงบวกมาในอันดับสองและความรู้สึกเชิงลบส่วนน้อย [28]

บทที่ 3
การสร้างโมเดลการจำแนกประเภทความรู้สึก

3.1 แนวคิดและภาพรวมวิธีการวิจัย

งานวิจัยมีแนวคิดที่จะนำเสนอวิธีการจำแนกประเภทความรู้สึกจากข้อความที่เป็นภาษาอังกฤษจากชุดข้อมูลความคิดเห็นจากการใช้บริการ Spotify ที่รวบรวมจากในเว็บ Kaggle โดยถูกรวบรวมความคิดเห็นมาจาก Google Play Store และนำเทคนิคการประมวลผลธรรมชาติ (Natural Language Processing) โดยใช้การสกัดคุณลักษณะ (Feature Extraction) ด้วยวิธี TF-IDF และ Word2vec และเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ด้วยอัลกอริทึม Random Forest, Naïve Bayes, Logistic Regression, Support Vector Machine และ XGBoost รวมไปถึงมีการใช้ DistilBERT ซึ่งเป็นโมเดล Transformer ร่วมด้วย นำไปสร้างโมเดลเพื่อใช้ในการจำแนกประเภทความรู้สึกออกเป็น 2 ประเภท ได้แก่ ความคิดเห็นเชิงบวก (Positive) และความคิดเห็นเชิงลบ (Negative) แล้วทำการเปรียบประสิทธิภาพของผลการทดลองจากการใช้เทคนิคดังกล่าว เพื่อหาโมเดลที่มีประสิทธิภาพดีที่สุด โดยวัดจากค่าความแม่น (Accuracy) ค่าความเที่ยง (Precision) ค่าเรียกคืน (Recall) ค่าเอฟวัน (F1-Score) และค่า ROC รวมถึงมีการทำ Feature Importance เพื่อดูว่าตัวแปรไหนมีความสำคัญในการจำแนกประเภทได้ดีด้วยการใช้ TF-IDF และ SHAP ซึ่งขั้นตอนและวิธีการดำเนินการแสดงได้ดังภาพ



ภาพประกอบ 11 ขั้นตอนและวิธีการดำเนินงาน

3.2 ขั้นตอนการรวบรวมข้อมูล

งานวิจัยใช้ชุดข้อมูลที่มีบุคลรวบรวมไว้จากเว็บไซต์ Kaggle ชื่อชุดข้อมูล Spotify App Reviews ที่ถูกรวบรวมความคิดเห็นมาจาก Google Play Store [29] ซึ่งสามารถดาวห์โหลดมาใช้เป็นไฟล์รูปแบบ .csv 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

ภาพประกอบ 12 ตัวอย่างชุดข้อมูล

3.3 ขั้นตอนการเตรียมข้อมูลก่อนการวิเคราะห์

ในขั้นตอนการเตรียมข้อมูลก่อนการวิเคราะห์ประกอบด้วย 3 ขั้นตอนคือ การระบุประเภทความรู้สึกของความคิดเห็น (Data Labeling) การทำความสะอาดข้อมูล (Data Cleaning) และการตัดคำ (Word Tokenization)

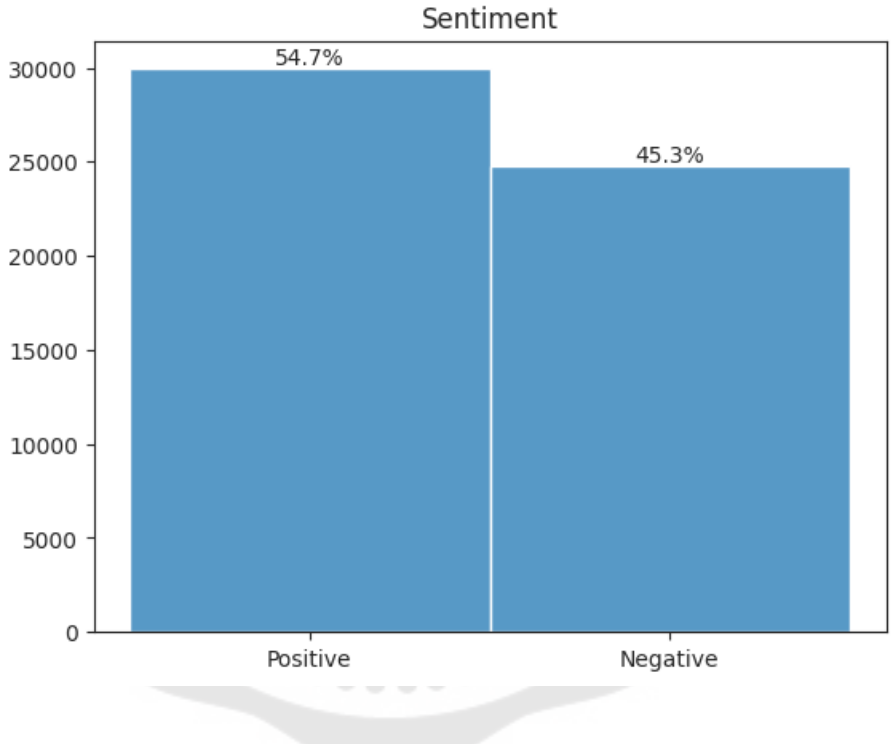
3.3.1 Data Labeling

งานวิจัยนี้นำความคิดเห็นทั้งหมด 61,594 ข้อความ ที่รวบรวมมาจากเว็บไซต์ Kaggle มาทำการระบุประเภทความรู้สึก โดยใช้การให้คะแนน (Rating) ที่อยู่ระหว่าง 1 ถึง 5 เป็นตัวจำแนกประเภทความรู้สึกออกเป็น 2 ประเภท ได้แก่ ความรู้สึกเชิงบวก (Positive) และความรู้สึกเชิงลบ (Negative) ใช้เพียงข้อมูลคอลัมน์ความคิดเห็น (Review) และการระบุคะแนน

(Rating) เกณฑ์การใช้คะแนนจะไม่ใช้ข้อความที่มีการระบุการให้คะแนนเท่ากับ 3 ดังนั้นข้อมูลที่คงเหลืออยู่ 54,708 ข้อความ ซึ่งใช้เกณฑ์ในการให้คะแนนจำแนกประเภทความรู้สึก ดังนี้

- ถ้าระบุคะแนนมากกว่า 3 ให้เป็นความคิดเห็นเชิงบวก (Positive)
- ถ้าระบุคะแนนน้อยกว่า 3 เป็นความคิดเห็นเชิงลบ (Negative)

จากการนำชุดข้อมูล 54,708 ข้อความ มาจำแนกตามเกณฑ์ข้างต้นสามารถจำแนกประเภทความรู้สึกได้ดังนี้ ความคิดเห็นเชิงบวก (Positive) จำนวนทั้งสิ้น 29,937 และความคิดเห็นเชิงลบ (Negative) จำนวนทั้งสิ้น 24,771 แสดงสัดส่วนประเภทความคิดเห็นได้ดังภาพ



ภาพประกอบ 13 สัดส่วนการระบุประเภทความคิดเห็น

3.3.2 Data Cleaning

เนื่องจากข้อมูลความคิดเห็นส่วนมากเขียนด้วยภาษาที่ไม่เป็นทางการ และอาจมีการใส่อักขระพิเศษที่ไม่สามารถนำไปวิเคราะห์ความรู้สึกได้ เช่น เครื่องหมายวรรคตอน (Punctuation) หรือ Emoji เป็นต้น จึงทำความสะอาดข้อมูลและตรวจสอบความถูกต้องของชุดข้อมูลที่อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์และเพิ่มประสิทธิภาพในการเรียนรู้ของโมเดลได้ ซึ่งมีรายละเอียดดังนี้

1) เปลี่ยนข้อความทั้งหมดให้เป็นตัวพิมเล็ก (Lowercase) ตัวอย่างเช่น “Great music server” เป็น “great music server” โดยใช้คำสั่ง .lower

2) ลบ Emoji หรือเครื่องหมายวรรคตอน (Punctuation) เก็บไว้เฉพาะตัวอักษรภาษาอังกฤษและตัวเลข 0-9 ตัวอย่างเช่น “i like it is super🙌!” เป็น “i like it super” โดยใช้ไลบารี่ re [1] คำสั่ง re.sub

3.3.3 Word Tokenization

เนื่องจากความคิดเห็นจะประกอบด้วยคำที่มาเรียงต่อกันเป็นประโยคให้เกิดเป็นข้อความ ซึ่งข้อความมักมีคำที่สามารถบ่งบอกถึงความรู้สึกได้ ผู้วิจัยจึงเลือกใช้ไลบรารี nltk [2] ที่มีความสามารถในการประมวลผลภาษาธรรมชาติที่มีประสิทธิภาพสูงมาใช้ในการช่วยจัดการข้อความภาษาอังกฤษดังนี้

1) ลบคำที่ไม่สื่อความหมาย (Stop Word) เช่น ‘the’, ‘a’, ‘is’ ด้วยคำสั่ง stopwords.words('english')
2) ลดรูปแบบของคำหรือแปลงคำให้อยู่ในรูปแบบรากคำศัพท์อิงตามพจนานุกรม (Dictionary) ด้วยคำสั่ง WordNetLemmatizer() [3]

ความคิดเห็นที่ผ่านกระบวนการทำ Word Tokenization แล้วจะมีลักษณะตัวอย่างดังภาพ

index	Review	sentiment
0	great music service audio high quality app easy use also quick friendly support	Positive
1	please ignore previous negative rating app super great give five star	Positive
2	pop-up get best spotify experience android 12 annoying please let get rid	Positive
3	really buggy terrible use recently	Negative
4	dear spotify get song didnt put playlist shuffle play	Negative
6	love selection lyric provided song youre listening	Positive
8	great app best mp3 music app ever used one problem cant play song find song despite app wonderful recommend best	Positive
9	im deleting app following reason app failing business model whether streaming service like consumer doesnt want pay music cant fully 6 ad successively upon logging single song much closed app ad number 6 im patient way profit youve already peaked thats left decline	Negative
10	love spotify usually app best others stated control button disappear podcasts stop reason fixed deleting reinstalling app first world problem sure highly annoying nonetheless especially paying prime	Negative
11	cant play spotify wifi	Negative

ภาพประกอบ 14 ตัวอย่างข้อความที่ผ่านกระบวนการ Word Tokenization

จากนั้นมีการนำเอาผลลัพธ์ที่ได้มาคำนวณหาคำที่ถูกพบมากที่สุดในแต่ละประเภทความคิดเห็นในรูปแบบของเวิร์ดคลาวด์ (Word Cloud) ได้ดังภาพ

ภาพประกอบ 15 Word Cloud ของประเภทความรู้สึกเชิงบวก

ภาพประกอบ 16 Word Cloud ของประเภทความรู้สึกเชิงลบ

3.3.4 **ขั้นตอนการการจัดเตรียมชุดข้อมูล** (Preparing the Dataset)

ชุดข้อมูลที่ผ่านการขั้นตอนการเตรียมข้อมูลก่อนการวิเคราะห์นำมาทำการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลฝึก (Train Set) เป็นจำนวน 75% มีความคิดเห็นจำนวน 41,031 ข้อความ และชุดข้อมูลทดสอบ (Test set) เป็นจำนวน 25% มีความคิดเห็นจำนวน 13,677 ข้อความ จากทั้งหมด 54,708 ข้อความ จากนั้นนำข้อมูลที่ผ่านการแบ่งชุดข้อมูลเรียบร้อยเข้าสู่ขั้นตอนการสร้างโมเดลและการประเมินประสิทธิภาพโมเดลต่อไป

3.4 **ขั้นตอนการสร้างโมเดล**

ชุดข้อมูลฝึกที่ผ่านการขั้นตอนการเตรียมข้อมูลก่อนการวิเคราะห์ จะถูกนำมาสร้างโมเดลการจำแนกประเภทความรู้สึกโดยใช้อัลกอริทึม Random Forest, Naïve Bayes, Logistic Regression, Support Vector Machine, XGBoost และ DistilBERT โดยมีรายละเอียดดังนี้

3.4.1 Feature Extraction

ชุดข้อมูลความคิดเห็นที่เป็นข้อความจะถือว่าคำ (word) แต่ละคำในข้อความความคิดเห็นเป็นคุณลักษณะ (feature) และบางครั้งเรียกว่าโทเคน (token) ซึ่งจำเป็นจะต้องทำการแปลงข้อความให้อยู่ในรูปแบบของเวกเตอร์ (Vector) เพื่อให้สามารถนำชุดข้อมูลเข้าสู่การสร้างโมเดลด้วยวิธีการเรียนรู้ของเครื่อง (Machine learning) ในงานวิจัยนี้ผู้วิจัยจะแทนข้อความในชุดข้อมูลฝึกและข้อมูลทดสอบด้วยวิธีการทำ TF-IDF Vector และวิธี Word2Vec กับอัลกอริทึม Random Forest, Naïve Bayes, Logistic Regression, Support Vector Machine และ XGBoost

การสกัดคุณลักษณะด้วยวิธี TF-IDF จะกำหนดพารามิเตอร์ ดังนี้

- analyzer='char': ระบุว่า TfidfVectorizer จะใช้วิธีการแปลงข้อความโดยการสร้างเวกเตอร์จากตัวอักษร (character) ซึ่งหมายความว่าแต่ละคำจะถูกแบ่งเป็นตัวอักษรเพื่อนับความถี่
- ngram_range=(2, 3): ระบุว่า TfidfVectorizer จะใช้ n-grams ที่มีความยาวอยู่ระหว่าง 2 ถึง 3 ตัวอักษรเพื่อสร้างคุณลักษณะ (features) ที่มีความหมายเช่น "ab", "abc", "bc", "bcd" ซึ่งจะช่วยให้โมเดลเรียนรู้เกี่ยวกับลำดับของตัวอักษรที่เกิดขึ้นใน

จากนั้นนำไป fit กับข้อมูลฝึกและข้อมูลทดสอบเพื่อแปลงให้เป็นเวคเตอร์ตัวอย่างผลลัพธ์การทำสกัดคุณลักษณะสามารถแสดงได้ดังภาพ

```
[ ]  # Get the feature names using get_feature_names_out
     feature_names = tfidf.get_feature_names_out()

     # Check the shape of the TF-IDF matrices and display feature names
     X_train_tfidf.shape, X_test_tfidf.shape, feature_names[:10]

     ((41031, 12448),
      (13677, 12448),
      array([' -', ' - ', ' --', ' -0', ' -1', ' -2', ' -3', ' -4', ' -5',
             ' -6'], dtype=object))
```

ภาพประกอบ 17 ตัวอย่างผลลัพธ์การสกัดคุณลักษณะด้วยวิธี TF-IDF

การสกัดคุณลักษณะด้วยวิธี Word2vec จะกำหนดพารามิเตอร์และวิธีการสร้าง ดังนี้

- การกำหนดพารามิเตอร์สำหรับสร้าง Word2Vec

Sentences = X_train.apply(lambda x: x.split()) : การใช้ข้อมูล X_train โดยใช้ช่องว่างเป็นตัวแยกเพื่อแบ่งข้อความเป็นคำ ๆ และใช้ Gensim Library เพื่อสร้างโมเดล Word2Vec ในการเรียนรู้ความสัมพันธ์ระหว่างคำในข้อความที่ให้มา เพื่อสร้าง Word Embeddings ขึ้นมาด้วยชุดข้อมูล X_train

vector_size =100: ขนาดของเวกเตอร์ที่ต้องการสำหรับแต่ละคำ

window=5: ขนาดของช่วงคำที่จะพิจารณารอบคำปัจจุบัน

min_count =1: จำนวนครั้งขั้นต่ำที่คำนั้นต้องปรากฏในข้อมูล เพื่อจะถูกนำมาสร้างเวกเตอร์

workers=4: จำนวนเธรดที่จะใช้ในการคำนวณ

- สร้างฟังก์ชั่น average_word_vectors สำหรับหาค่าเฉลี่ยของเวกเตอร์คำในข้อความนั้นๆ
- สร้างฟังก์ชั่น average_word_vectorizer สำหรับสร้างเวกเตอร์จากข้อความทั้งหมดในชุดข้อมูล ฟังก์ชันนี้จะคืนค่าเป็นอาร์เรย์ของเวกเตอร์ที่ถูกสร้างจากการหาค่าเฉลี่ยของเวกเตอร์คำในแต่ละข้อความ

จากนั้นนำไป fit กับข้อมูลฝึกและข้อมูลทดสอบเพื่อแปลงให้เป็นเวคเตอร์ตัวอย่างผลลัพธ์การทำสกัดคุณลักษณะสามารถแสดงได้ดังภาพ

```
# Check the shape of the Word2Vec matrices
print("Shape of X_train_w2v:", X_train_w2v.shape)
print("Shape of X_test_w2v:", X_test_w2v.shape)

# Get the first 10 sentences from the training set
example_sentences = X_train[:10]

# Tokenize each sentence and display word vectors
for sentence in example_sentences:
    tokenized_sentence = sentence.split()
    for word in tokenized_sentence:
        print("Word:", word)
        print("Vector:", w2v_model.wv[word] if word in w2v_model.wv else "Word not found in vocabulary")
        print()
```

```
Shape of X_train_w2v: (41031, 100)
Shape of X_test_w2v: (13677, 100)
Word: slow
Vector: [ 0.42714342  0.37097347  0.2604563   0.28304887 -0.16577932  0.24225655
 -0.01827581  1.041299    0.0363015  -0.30298704  0.12492549 -0.68920755
  0.04403181  0.0884705  -0.9092903  -1.0592806   0.3316742  -0.9754476
  0.2356209  -1.3883203   0.35283795  0.14413951  0.2772247  -0.59868485
 -0.46967572 -0.01103833  0.4631017   0.14938307 -0.68779135 -0.5244122
  0.61623126  0.84628546  0.29247507 -0.7861966  -0.06058315  0.2883686
 -0.18776618 -0.45860383 -0.489843   -0.29394203  0.5655666  -0.7159091
 -0.73779404  0.49204722 -0.06989697  0.40147445  0.4920853  -0.70960176
  0.39156324 -0.4180342   0.6350755   0.27095222 -0.174554   -0.18964937
 -0.51455516  0.4963303   0.36873654 -0.50265485 -1.2442827  -0.13389926
 -1.0997988  -0.37066078 -0.734721   -0.10193884 -0.26405466 -0.03390601
  0.23471333 -0.5854001  -0.17004561  0.8096407  -0.6421777  -0.6192641
 -0.0052276  -0.2788202  -0.05915216 -0.03189295  0.21251065 -0.5867616
 -1.0310333   1.0622083  -1.0795923  -0.42583662 -0.74073243  1.0582795
 -0.73539823 -0.32832623  0.25617242  0.96920896  0.5672955   0.18295291
 -0.19625017  0.37517145 -0.216905   -0.1416251   0.71211     0.9967223
 -0.7294862   0.6133085  -0.2829913   0.25702566]
```

ภาพประกอบ 18 ตัวอย่างผลลัพธ์การสกัดคุณลักษณะด้วยวิธี Word2Vec

ส่วนของอัลกอริทึม DistilBERT แปลงข้อความแต่ละตัวเป็นเวกเตอร์ (embedding) หรือตัวเลขแทนความหมาย โดยใช้เลเยอร์การฝังตัว (Embedding Layer) ที่ชื่อว่า distilbert-base-uncased ซึ่งเป็นเวอร์ชันพื้นฐานที่ไม่คำนึงถึงพิมพ์ใหญ่-เล็ก (case-insensitive) ของโมเดล DistilBERT ที่ผ่านการฝึกฝนมาแล้ว (Pretrained)

3.4.2 การสร้างโมเดล (Modeling)

เมื่อทำการแปลงชุดข้อมูลจากข้อความให้อยู่ในรูปแบบของตัวเลขแล้วจะนำผลลัพธ์ที่ได้มาทำการสร้างโมเดล โดยจะทดลองใช้อัลกอริทึม Random Forest, Naïve Bayes, Logistic Regression, Support Vector Machine, XGBoost และ DistilBERT โดยแต่ละอัลกอริทึมจะมีการสร้างโมเดลด้วยการใช้ค่าเริ่มต้น (Default) ของแต่ละโมเดล ดังนี้

1) โมเดล Random Forest

โมเดล Random Forest ถูกสร้างโดยใช้ไลบรารี scikit-learn โดยใช้ค่าเริ่มต้นดังนี้

RF = RandomForestClassifier()

- ฝึกโมเดล Random Forest บนชุดข้อมูล TF-IDF (X_train_tfidf) ได้ดังนี้

  RF.fit(X_train_tfidf, y_train)

- ฝึกโมเดล Random Forest บนชุดข้อมูล Word2Vec (X_train_w2v) ได้ดังนี้
  RF.fit(X_train_w2v, y_train)

2) โมเดล Support Vector Machine

โมเดล Support Vector Machine ถูกสร้างโดยใช้ไลบรารี scikit-learn โดยใช้ค่าเริ่มต้นดังนี้ svm = svm.SVC(kernel='linear')

- ฝึกโมเดล Random Forest บนชุดข้อมูล TF-IDF (X_train_tfidf) ได้ดังนี้
  SVM.fit(X_train_tfidf, y_train)
- ฝึกโมเดล Random Forest บนชุดข้อมูล Word2Vec (X_train_w2v) ได้ดังนี้
  SVM.fit(X_train_w2v, y_train)

3) โมเดล Naïve Bayes

โมเดล Naïve Bayes ถูกสร้างโดยใช้ไลบรารี scikit-learn โดยใช้ค่าเริ่มต้นดังนี้
NB = MultinomialNB()

- ฝึกโมเดล Random Forest บนชุดข้อมูล TF-IDF (X_train_tfidf) ได้ดังนี้
  NB.fit(X_train_tfidf, y_train)
- ฝึกโมเดล Random Forest บนชุดข้อมูล Word2Vec (X_train_w2v) ได้ดังนี้
  NB.fit(X_train_w2v, y_train)

4) โมเดล Logistic Regression

โมเดล Logistic Regression ถูกสร้างโดยใช้ไลบรารี scikit-learn โดยใช้ค่าเริ่มต้นดังนี้ LR = LogisticRegression()

- ฝึกโมเดล Random Forest บนชุดข้อมูล TF-IDF (X_train_tfidf) ได้ดังนี้
  LR.fit(X_train_tfidf, y_train)
- ฝึกโมเดล Random Forest บนชุดข้อมูล Word2Vec (X_train_w2v) ได้ดังนี้
  LR.fit(X_train_w2v, y_train)

5) โมเดล XGBoost

โมเดล Logistic Regression ถูกสร้างโดยใช้ไลบรารี xgboost โดยใช้ค่าเริ่มต้นดังนี้ XGB = xgb.XGBClassifier()

- ฝึกโมเดล Random Forest บนชุดข้อมูล TF-IDF (X_train_tfidf) ได้ดังนี้
  XGB.fit(X_train_tfidf, y_train)
- ฝึกโมเดล Random Forest บนชุดข้อมูล Word2Vec (X_train_w2v) ได้ดังนี้
  XGB.fit(X_train_w2v, y_train)

6) โมเดล DistilBERT

โมเดล DistilBERT ในงานวิจัยนี้มีกระบวนการสร้างโมเดล ดังนี้

- Load Model และ Tokenizer

Load Model: ในส่วนของงานวิจัยนี้ได้ใช้โมเดล DistilBERT base uncased โดยโหลดโมเดล DistilBERT ที่ผ่านการฝึกฝนมาแล้ว (Pretrain) สำหรับ งานจำแนกประเภทข้อความ (Sequence Classification) และ Tokenizer สำหรับแปลงข้อความเป็นตัวเลขที่ผ่านการฝึกฝนมาแล้วจาก Hugging Face Transformers ดังนี้

Model = transformers.AutoModelForSequenceClassification.from
_pretrained("distilbert-base-uncased")

Tokenizer = transformers.AutoTokenizer.from_pretrained ("distilbert- base-uncased")

Tokenizer: สร้างฟังก์ชัน tokenize_function ทำหน้าที่แปลงข้อความโดยใช้ tokenizer แปลงข้อความในแต่ละชุดข้อมูล (X_train และ X_test) ให้เป็นข้อมูลตัวเลขที่โมเดลเข้าใจสามารถนำไปประมวลผลได้ โดยมีการกำนดค่า ดังนี้

padding="max_length": กำหนดความยาวสูงสุดของข้อมูล ตัดหรือเติมข้อมูลให้เท่ากัน

truncation=True: ตัดข้อความให้เหลือแค่ความยาวสูงสุดที่กำหนด

- สร้างชุดข้อมูล

กำหนดคลาส SpotifyDataset สร้างชุดข้อมูล PyTorch Dataset ที่ใช้ในการฝึกและทดสอบโมเดล DistilBERT โดยเริ่มจากการเลือกตัวอย่างจากชุดข้อมูลฝึกฝน(train_encodings) จากฟังก์ชั่น tokenize_function และกำหนดประเภทความรู้สึกของตัวอย่างนี้จากชุดข้อมูลฝึกฝน (y_train) เพื่อให้เตรียมข้อมูลในรูปแบบที่ถูกต้องสำหรับการฝึกโมเดล

- ฝึกโมเดล (Train Model)

มีกำหนดพารามิเตอร์สำหรับการฝึกไว้ใน training_args ที่รวมค่าต่างๆ ที่จำเป็นสำหรับการฝึก โดยกำหนดจำนวน epoch เท่ากับ 8 ขนาดของ batchเท่ากับ 16 จากนั้นทำการฝึก

โมเดลโดยใช้ Trainer จากไลบรารี Transformers เพื่อฝึกโมเดลด้วยโมเดล DistilBERT ที่โหลดมาใช้พารามิเตอร์สำหรับการฝึกที่กำหนดไว้ใน training_args ด้วยชุดข้อมูล train_dataset จากนั้นทำการเก็บโมเดลที่ผ่านการฝึกเรียบร้อยเพื่อไปใช้ในการจำแนกประเภทกับชุดข้อมูลทดสอบต่อไป

- นำโมเดลที่ผ่านการฝึกไปทำนาย

โหลดโมเดลที่ผ่านการขั้นตอนการฝึกแล้วนำไปใช้ทำนายผลโดยการสร้าง pipeline จากไลบรารี transformers ได้ดังนี้

```
Classifier=transformers.pipeline("sentimentanalysis",
model=model, tokenizer=tokenizer)
predictions = classifier(X_test.tolist())
```

3.5 การประเมินประสิทธิภาพโมเดล

โมเดลที่ได้จากการฝึกโมเดลของอัลกอริทึม Random Forest, Naïve Bayes, Logistic Regression, Support Vector Machine, XGBoost และ DistilBERT จะถูกนำมาประเมินประสิทธิภาพของโมเดลด้วยการนำแต่ละโมเดลมาทำนายกับข้อมูลทดสอบโดยจำแนกความรู้สึกทั้ง 2 ประเภท คือ ความคิดเห็นเชิงบวก และ ความคิดเห็นเชิงลบ จากนั้นทำการวัดประสิทธิภาพของแต่ละโมเดลด้วยการวัดค่า Accuracy, Precision, Recall, F1-Scorec และ ROC โดยใช้ไลบรารีของ sklearn.metrics และทำการเปรียบเทียบประสิทธิภาพเพื่อหาโมเดลที่สามารถจำแนกประเภทความรู้สึกได้ดีที่สุดกับชุดข้อมูลนี้

3.6 การหาคุณลักษณะที่สำคัญ (Feature Importance)

หลังจากทำการประเมินประสิทธิภาพของโมเดลแล้วในงานวิจัยนี้จะมีการหาตัวแปรหรือคุณลักษณะที่สำคัญซึ่งในชุดข้อมูลที่เป็นข้อความตัวแปรในที่นี่คือคำ (Word) เพื่อดูว่าตัวแปรใดส่งผลต่อการจำแนกประเภทว่าความคิดเห็นว่าเป็นเชิงบวก หรือเชิงลบ โดยจะใช้วิธีการหาคุณลักษณะที่สำคัญ 2 วิธีด้วยกัน ดังนี้

1) TF-IDF ร่วมกับ โมเดล Logistic Regression

- **สร้างโมเดล**: ทำการสร้างโมเดล Logistic Regression โดยใช้ข้อมูลที่ถูกสกัดคุณลักษณะด้วย TF-IDF (Term Frequency-Inverse Document Frequency) ซึ่งเป็นวิธีการที่ใช้ในการแปลงข้อมูลข้อความเป็นเวกเตอร์ของคุณลักษณะ (Features) โดยคำนึงถึงความถี่ของคำแต่ละคำในเอกสาร และความสำคัญของคำนั้นในเอกสารทั้งหมด

- **ฝึกโมเดล**: ทำการฝึกโมเดลที่ได้มาจากการสร้างโมเดลก่อนหน้าด้วยชุดข้อมูลฝึก
- **คำนวณความสำคัญของคุณลักษณะ** (Feature Importance): เมื่อโมเดลที่ถูกฝึกแล้ว สามารถดึงค่า coefficients หรือค่าความสำคัญของคุณลักษณะ (Importance) จากโมเดลได้ ซึ่งในกรณี Logistic Regression ค่า coefficients จะบ่งบอกถึงปริมาณของการเปลี่ยนแปลงในผลลัพธ์ (ค่าของตัวแปรตาม) เมื่อมีการเปลี่ยนแปลงในตัวแปรอิสระ (ค่าของตัวแปรอิสระ) โดยที่ค่า coefficients ที่มากที่สุดจะบ่งบอกถึงคุณลักษณะที่มีผลต่อการทำนายมากที่สุด
- **การแสดงผล**: นำค่า coefficients หรือค่าความสำคัญของคุณลักษณะที่ได้มา แสดงผลเป็นกราฟของคุณลักษณะที่มีความสำคัญที่สุด เพื่อให้สามารถทำความเข้าใจถึงคุณลักษณะที่มีผลต่อผลลัพธ์มากที่สุดได้โดยง่ายและชัดเจน

2) SHAP (SHapley Additive exPlanations)

SHAP เป็นวิธีการอธิบายโมเดล Machine Learning ช่วยวิเคราะห์ว่าคุณลักษณะ (Feature) แต่ละตัวส่งผลต่อผลลัพธ์การทำนายของโมเดลอย่างไร และสามารถวิเคราะห์โมเดลการจำแนกประเภทความคิดเห็นว่าคำหรือวลีใดที่มีอิทธิพลต่อการจำแนกประเภท [30] โดยเริ่มต้นจะคำนวณ SHAP Values ด้วยการใช้ไลบรารี shap.Explainer โดย ต้องการโมเดลที่ได้ถูกฝึก (Classifier) และข้อมูลทดสอบ (X_test) ซึ่งโมเดลและข้อมูล ทดสอบมาจากกระบวนการสร้างโมเดลด้วยอัลกอริทึม DistilBERT เป็นข้อมูลนำเข้า จากนั้น shap.Explainer จะรวบรวมค่า SHAP Values สำหรับแต่ละคำทั้งหมดในชุด ข้อมูลทดสอบ และทำการคำนวณค่าเฉลี่ยของ SHAP values สำหรับแต่ละคำ โดยแยก ระหว่าง SHAP Values ที่เกี่ยวข้องกับความคิดเห็นบวกและลบ

เมื่อได้ค่า SHAP values จะทำการแสดงผลด้วยกราฟ โดยกราฟจะแสดงคำที่มีค่า SHAP Values สูงสุดที่มีส่วนสำคัญในการจำแนกประเภทความคิดเห็นเชิงบวกและลบ

บทที่ 4
ผลการศึกษา

ในการวิจัยการจำแนกประเภทความรู้สึก ซึ่งใช้ข้อมูลของรีวิวผู้ใช้บริการแอพสปอติฟาย ด้วยวิธีการเรียนรู้ด้วยเครื่อง ผู้วิจัยได้ดำเนินการวิจัยโดยการศึกษาตามขบวนการและขั้นตอนต่างๆ จนกระทั่งประเมินประสิทธิภาพของโมเดลที่สร้างขึ้นให้เป็นไปตามวัตถุประสงค์ที่ได้กาหนดไว้ ได้ดังนี้

1) เพื่อสร้างและปรับปรุงประสิทธิภาพโมเดลให้สามารถจำแนกประเภทความรู้สึกจากความคิดเห็นภาษาอังกฤษของผู้ใช้บริการสปอติฟาย

2) เพื่อศึกษาเปรียบเทียบโมเดลที่สามารถจำแนกประเภทความรู้สึกจากความคิดเห็นภาษาอังกฤษของผู้ใช้บริการสปอติฟาย

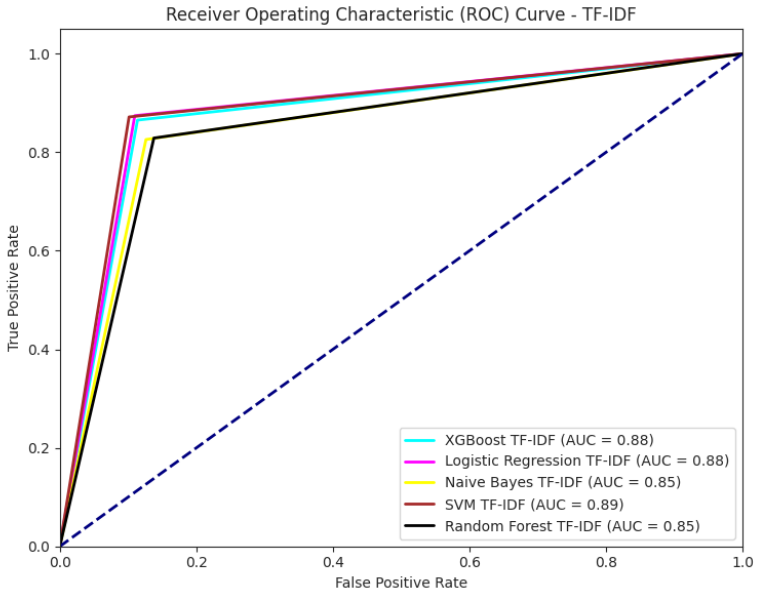
4.1 ผลการสร้างและปรับปรุงประสิทธิภาพโมเดล

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

ตารางที่ 16 ผลเปรียบเทียบการประเมินประสิทธิภาพโดยรวมในแต่ละอัลกอริทึมซึ่งได้จากการทำการสกัดคุณลักษณะด้วยวิธี TF-IDF

Algorithm	F1-Score	Precision	Recall	ROC	Accuracy
Random Forest	0.8530	0.8785	0.8289	0.8456	0.8440
Naïve Bayes	0.8555	0.8872	0.8260	0.8499	0.8477
Logistic Regression	0.8896	0.9054	0.8743	0.8823	0.8816
SVM	0.8913	0.9117	0.8718	0.8852	0.8840
XGBoost	0.8830	0.9013	0.8654	0.8758	0.8748

และสามารถนำการคำนวณค่า ROC มาพล็อตกราฟได้ดังภาพ



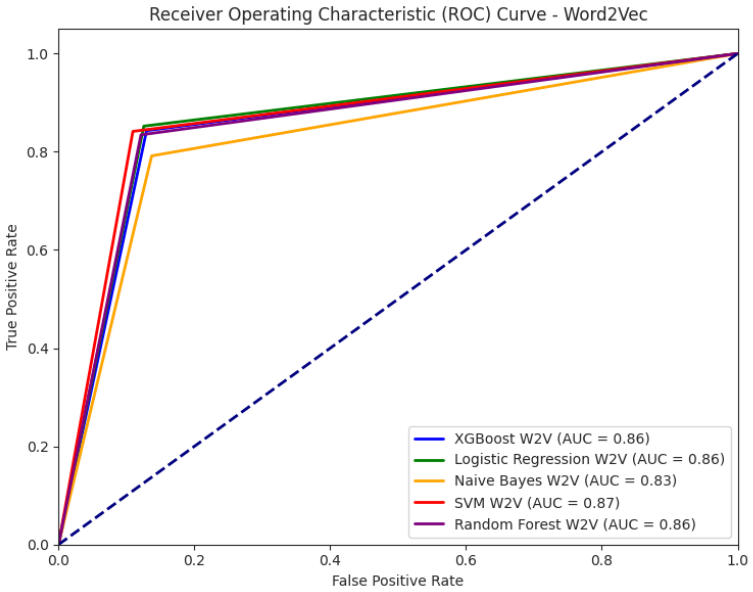
ภาพประกอบ 19 ผลลัพธ์ ROC Curve ของแต่ละอัลกอริทึมจากการสกัดคุณลักษณะด้วย TF-IDF

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%

ตารางที่ 17 ผลเปรียบเทียบการประเมินประสิทธิภาพโดยรวมในแต่ละอัลกอริทึมซึ่งได้จากการทำการสกัดคุณลักษณะด้วยวิธี Word2Vec

Algorithm	F1-Score	Precision	Recall	ROC	Accuracy
Random Forest	0.8623	0.8920	0.8345	0.8566	0.8546
Naïve Bayes	0.8308	0.8740	0.7917	0.8273	0.8240
Logistic Regression	0.8711	0.8908	0.8522	0.8633	0.8623
SVM	0.8660	0.9023	0.8415	0.8660	0.8638
XGBoost	0.8642	0.8866	0.8430	0.8567	0.8555

และสามารถนำการคำนวณค่า ROC มาพล็อตกราฟได้ดังภาพ



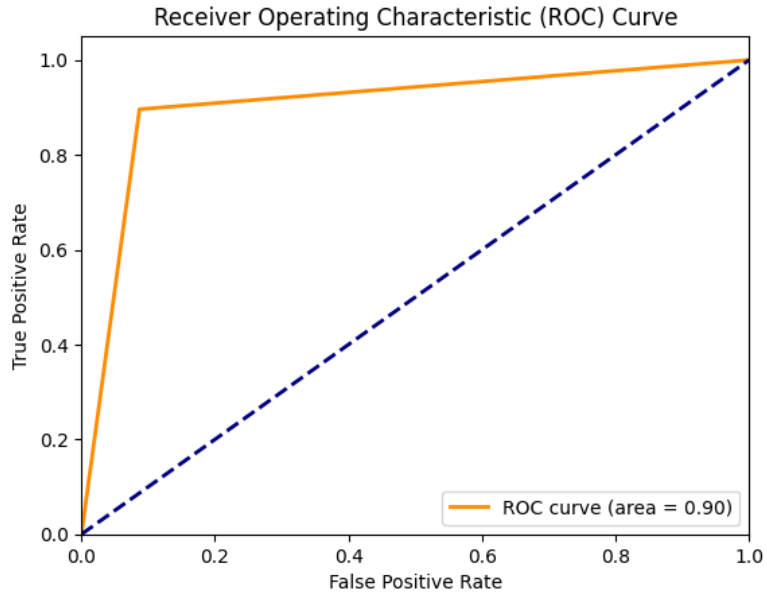
ภาพประกอบ 20 ผลลัพธ์ ROC Curve ของแต่ละอัลกอริทึมจากการสกัดคุณลักษณะด้วย Word2Vec

จากผลการประเมินประสิทธิภาพในแต่ละอัลกอริทึมที่ใช้วิธีการสกัดคุณลักษณะด้วยวิธี Word2Vec พบว่าอัลกอริทึม SVM และ Logistic Regression ได้ผลที่มีประสิทธิภาพใกล้เคียงกัน แต่ SVM มีประสิทธิภาพโดยรวมสูงที่สุดในค่าของตัววัดประสิทธิภาพมากกว่า โดยมีค่า F1-Score 86.60%, ค่า Precision 90.23%, ค่า Recall 84.15%, ค่า ROC 86.60% และค่า Accuracy 86.38%

ตารางที่ 18 ผลการประเมินประสิทธิภาพอัลกอริทึม DistilBERT

Algorithm	F1-Score	Precision	Recall	ROC	Accuracy
Distilbert	0.9105	0.9253	0.8962	0.9046	0.9039

และสามารถนำการคำนวณค่า ROC มาพล็อตกราฟได้ดังภาพ



ภาพประกอบ 21 ผลลัพธ์ ROC Curve ของอัลกอริทึม DistilBERT

จากผลการประเมินประสิทธิภาพของอัลกอริทึม DistilBERT ซึ่งเป็นอัลกอริทึม Transformer ที่เรียนรู้ข้อมูลจาก เรียนรู้จากชุดข้อมูลขนาดใหญ่ (Pre-trained model) ร่วมกับการเรียนรู้จากประเภทความรู้สึกที่เรากำหนดได้ประสิทธิภาพดังนี้ ค่า F1-Score 91.05%, ค่า Precision 92.53%, ค่า Recall 89.62%, ค่า ROC 90.46% และค่า Accuracy 90.39%

ดังนั้นจากการพิจารณาค่าตัววัดในตารางที่ 16, 17 และ 18 เมื่อนำมาเปรียบเทียบกันแล้วพบว่าอัลกอริทึม DistilBERT มีประสิทธิภาพสูงที่สุดในการจำแนกประเภทความรู้สึกว่าเป็นความรู้สึกเชิงบวก (Positive) หรือเชิงลบ (Negative) กับชุดข้อมูลของรีวิวผู้ใช้บริการแอพสปอติฟายในงานวิจัยชุดนี้ ซึ่งประสิทธิภาพโดยรวมสูงที่สุดในทุก ๆ ค่าของตัววัดประสิทธิภาพ โดยมีค่า F1-Score 91.05%, ค่า Precision 92.53%, ค่า Recall 89.62%, ค่า ROC 90.46% และค่า Accuracy 90.39%

4.2 ผลคุณลักษณะที่สำคัญต่อการจำแนกประเภท (Feature Importance)

จากการสร้างโมเดลเพื่อจำแนกประเภทความรู้สึกด้วยอัลกอริทึมต่าง ๆ แล้วผู้วิจัยได้มีการหาคุณลักษณะที่สำคัญซึ่งก็คือคำ (Word) ที่อยู่ในความคิดเห็นเพื่อดูว่าคำใดส่งผลต่อการจำแนกประเภทความคิดเห็นว่าความคิดเห็นนั้นเป็นเชิงบวก หรือเชิงลบ โดยใช้วิธีการหา

คุณลักษณะที่สำคัญ 2 วิธี ได้แก่ วิธีการสกัดคุณลักษณะด้วยวิธี TF-IDF ร่วมกับอัลกอริทึม Logistic Regression วัดคุณลักษณะที่สำคัญด้วยค่า Coefficients และการใช้อัลกอริทึม SHAP (SHapley Additive exPlanations) ร่วมกับอัลกอริทึม DistilBERT วัดคุณลักษณะที่สำคัญด้วยค่า SHAP ซึ่งมีรายละเอียดดังนี้

การหาคุณลักษณะที่สำคัญด้วยวิธี TF-IDF และ Logistic Regressionขั้นตอนแรกทำการสกัดคุณลักษณะจากข้อความความคิดเห็นโดยใช้เทคนิค TF-IDF (Term Frequency-Inverse Document Frequency) ซึ่งช่วยให้สามารถระบุคำที่มีความสำคัญในการจำแนกประเภทความรู้สึกได้ จากนั้นนำคุณลักษณะที่ได้จาก TF-IDF ไปสร้างโมเดล Logistic Regression สำหรับการจำแนกประเภทความรู้สึก เมื่อได้โมเดล Logistic Regression แล้ว สามารถพิจารณาค่า Coefficients ของแต่ละคุณลักษณะ โดยคุณลักษณะที่มีค่า Coefficients สูงเป็นบวกจะเป็นคุณลักษณะที่มีความสำคัญในการจัดอยู่ในกลุ่มความรู้สึกเชิงบวก และคุณลักษณะที่มีค่า Coefficients สูงเป็นลบจะเป็นคุณลักษณะที่มีความสำคัญในการจัดอยู่ในกลุ่มความรู้สึกเชิงล

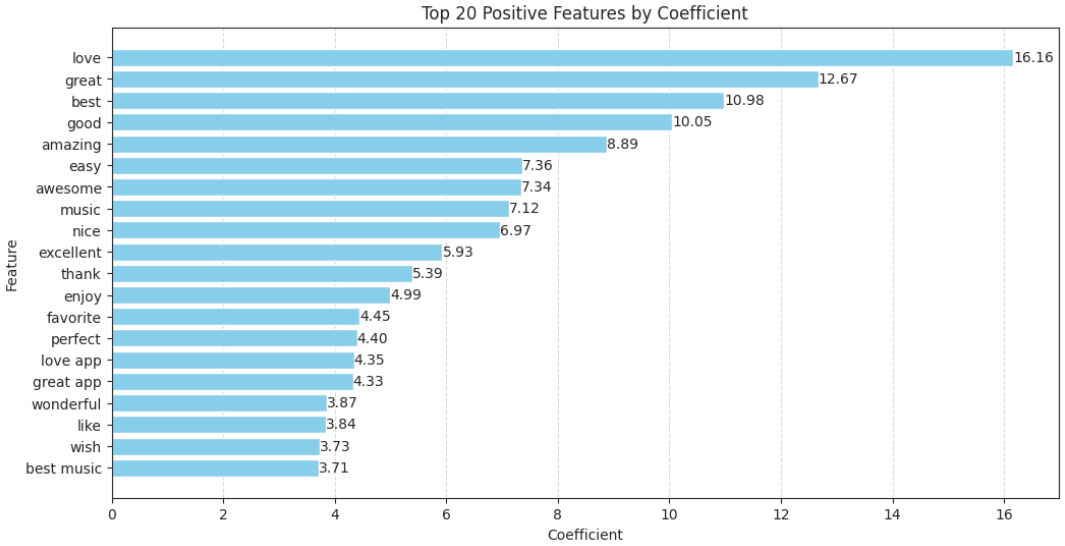
การหาคุณลักษณะที่สำคัญด้วยวิธี SHAP ExplainerSHAP Explainer เป็นเครื่องมือที่ใช้ในการอธิบายผลการทำนายของโมเดลแบบ Black Box โดยคำนวณค่า SHAP Value สำหรับแต่ละคุณลักษณะค่า SHAP Value บวกสูงของคุณลักษณะใดหมายถึงคุณลักษณะนั้นมีผลสำคัญในการเพิ่มโอกาสที่ข้อความจะจัดอยู่ในกลุ่มความรู้สึกเชิงบวก ค่า SHAP Value ลบสูงของคุณลักษณะใดหมายถึงคุณลักษณะนั้นมีผลสำคัญในการเพิ่มโอกาสที่ข้อความจะจัดอยู่ในกลุ่มความรู้สึกเชิงลบ

ความแตกต่างที่สำคัญระหว่าง Coefficients และ SHAP Value คือ Coefficients เป็นค่าที่ได้มาโดยตรงจากการประมาณค่าพารามิเตอร์ของโมเดล ในขณะที่ SHAP Value เป็นการอธิบายผลการทำนายของโมเดลภายหลังจากที่โมเดลถูกสร้างขึ้นแล้ว ทั้งสองวิธีให้มุมมองที่แตกต่างกันในการวิเคราะห์คุณลักษณะที่สำคัญ ซึ่งช่วยให้เข้าใจว่าคำหรือกลุ่มคำใดในความคิดเห็นมีผลต่อการจำแนกความรู้สึกเป็นเชิงบวกหรือเชิงลบมากที่สุด

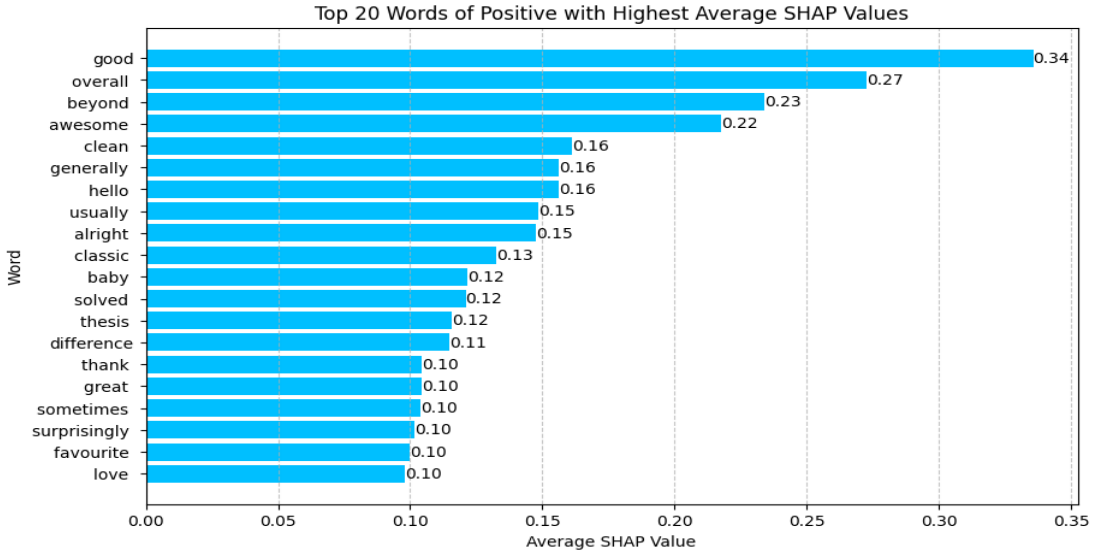
จากนั้นนำเสนอด้วยการพล็อตกราฟเพื่อดูว่าคุณลักษณะที่สำคัญในการพิจารณาว่าเป็นความคิดเห็นเชิงบวกและเชิงลบมีลักษณะใดโดยแสดงคุณลักษณะที่สำคัญ 20 อันดับแรกของทั้ง 2 วิธีการที่กล่าวไปข้างต้น และแยกเป็นคำคุณลักษณะที่สำคัญที่ถูกใช้เป็นตัวพิจารณาว่าเป็นเชิงบวก และเชิงลบ ได้ผลลัพธ์ดังนี้

1) **คุณลักษณะจำแนกประเภทที่สำคัญประเภทเชิงบวก**

ผลลัพธ์ที่ได้จากการหาคุณลักษณะที่สำคัญที่ถูกจำแนกประเภทเป็นความคิดเห็นเชิงบวกด้วยวิธีด้วยวิธี TF-IDF ร่วมกับอัลกอริทึม Logistic Regression แสดงได้ดังภาพที่ 20 และด้วยวิธีการหาค่า SHAP ดังภาพที่ 21 ตามลำดับ



ภาพประกอบ 22 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงบวกด้วยวิธี TF-IDF ร่วมกับ Logistic Regression



ภาพประกอบ 23 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงบวกด้วยวิธี SHAP

จากภาพการพล็อตกราฟผลลัพธ์คุณลักษณะที่สำคัญเชิงบวกด้วยค่า Coefficients ของวิธี TF-IDF ร่วมกับ Logistic Regression 20 อันดับแรกคือคุณลักษณะหรือคำดังนี้ Love, Great, Best, Good, Amazing, Easy, Awesome, Music, Nice, Excellence, Thank, Enjoy, Favorite, Perfect, Love app, Great app, Wonderful, Like, Wish และ Best music ส่วนผลลัพธ์คุณลักษณะที่สำคัญเชิงบวกด้วยค่า SHAP ของวิธี SHAP explainer 20 อันดับแรกคือคุณลักษณะหรือคำดังนี้ Good, Overall, Beyond, awesome, Clean, Generally, Hello, Usually, Alright, Classic, Baby, Solved, Thesis, Difference, Thank, Great, Sometimes, Surprisingly, Favorite และ Love ซึ่งคำที่มีร่วมกันระหว่าง 2 วิธี คือคำว่า Good, Awesome, Thank, Great, Favorite และ Love สามารถสรุปได้ดังรูปแบบตาราง ดังนี้

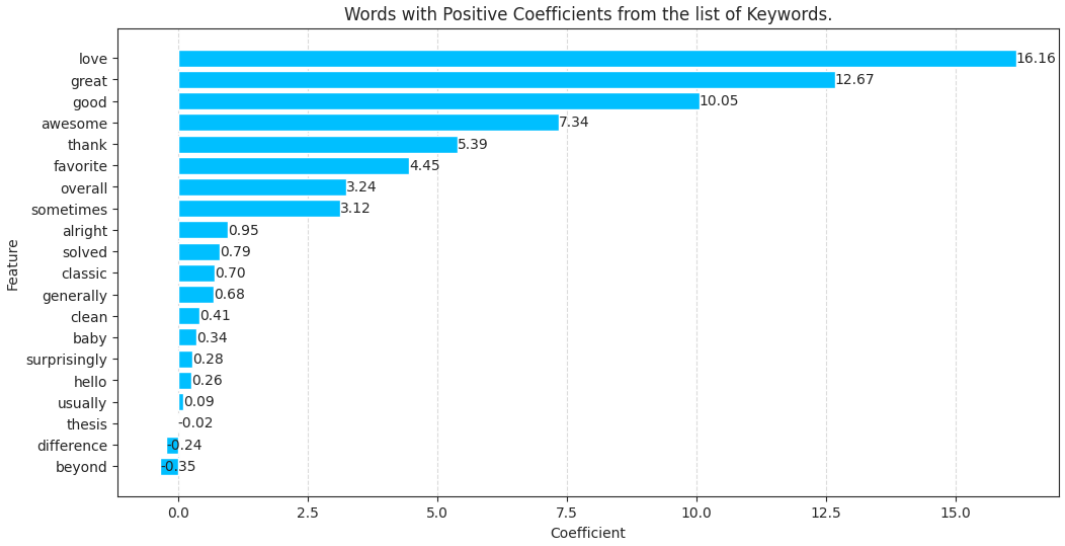
ตารางที่ 19 สรุปคุณลักษณะที่สำคัญเชิงบวก

No	TF-IDF + Logistic Regression		SHAP	
	Word	Coef.	Word	SHAP
1	Love	16.16	Good	0.34
2	Great	12.67	Overall	0.27
3	Best	10.98	Beyond	0.23
4	Good	10.05	Awesome	0.22
5	Amazing	8.89	Clean	0.16
6	Easy	7.36	Generally	0.16
7	Awesome	7.34	Hello	0.16
8	Music	7.12	Usually	0.15
9	Nice	6.97	Alright	0.15
10	Excellence	5.93	Classic	0.13
11	Thank	5.39	Baby	0.12
12	Enjoy	4.99	Solved	0.12
13	Favorite	4.45	Thesis	0.12
14	Perfect	4.40	Difference	0.11
15	Love app	4.35	Thank	0.10
16	Great app	4.33	Great	0.10
17	Wonderful	3.87	Sometimes	0.10
18	Like	3.84	Surprisingly	0.10

ตารางที่ 20 สรุปคุณลักษณะที่สำคัญเชิงบวก (ต่อ)

No	TF-IDF + Logistic Regression		SHAP	
	Word	Coef.	Word	SHAP
19	Wish	3.73	Favorite	0.10
20	Best music	3.71	Love	0.10

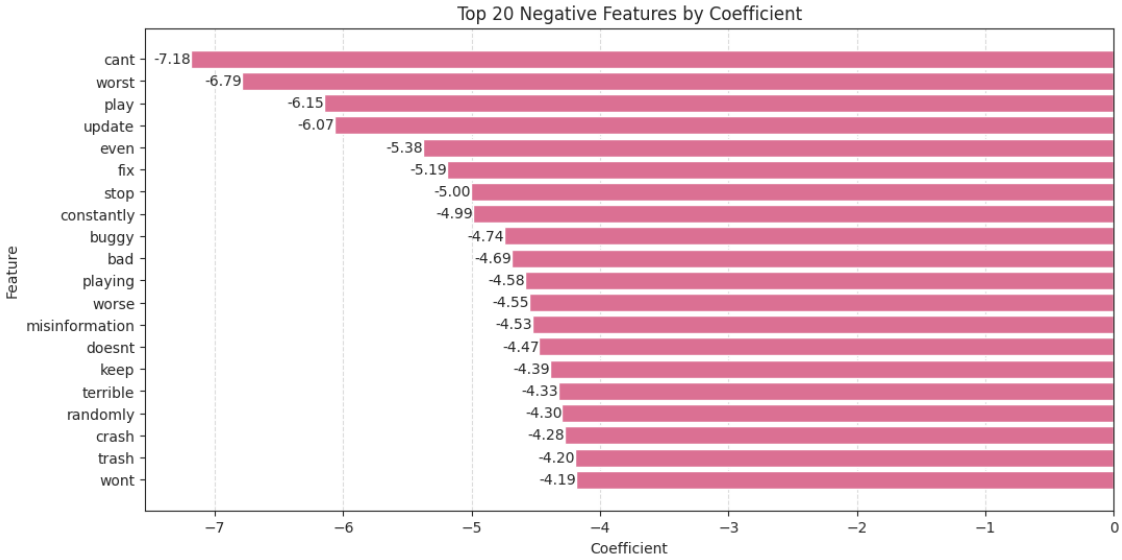
และมีการนำคุณลักษณะเชิงบวก 20 อันดับของวิธี SHAP มาเทียบหาคุณลักษณะเดียวกันในวิธี TF-IDF ร่วมกับ Logistic Regression พบว่ามีบางคุณลักษณะที่ได้ค่า Coefficient พิจารณาเป็นเชิงลบ ได้แก่ Thesis, Difference และ Beyond ซึ่งแสดงผลลัพธ์การพล็อตกราฟได้ดังภาพ



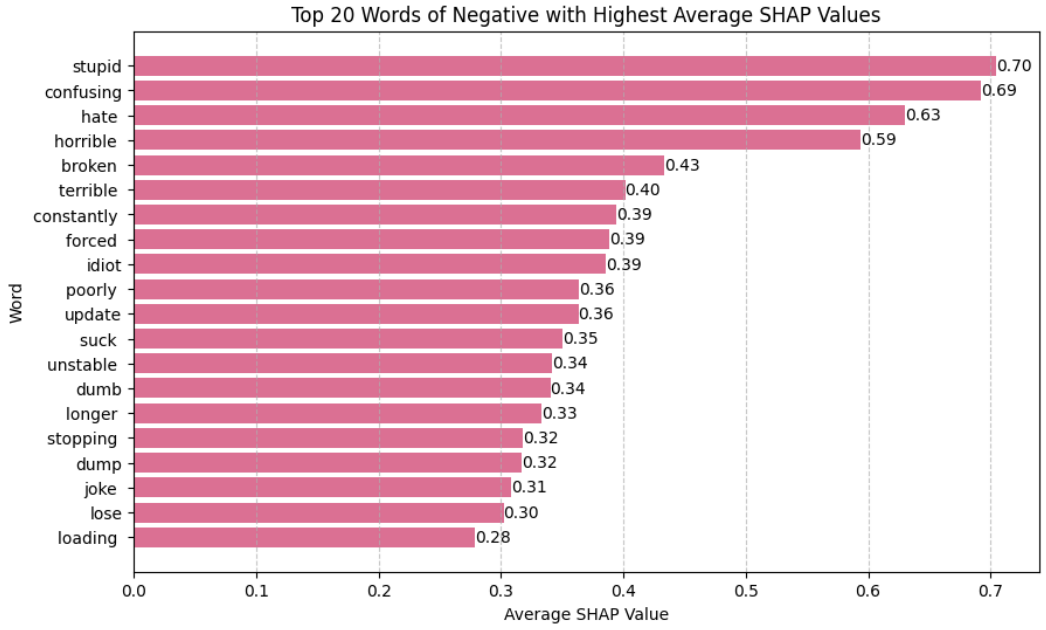
ภาพประกอบ 24 กราฟแสดงค่า coefficients คุณลักษณะที่สำคัญเชิงบวกตามคำที่กำหนด

2) คุณลักษณะจำแนกประเภทที่สำคัญประเภทเชิงลบ

ผลลัพธ์ที่ได้จากการหาคุณลักษณะที่สำคัญที่ถูกจำแนกประเภทเป็นความคิดเห็นเชิงลบด้วยวิธีด้วยวิธี TF-IDF ร่วมกับอัลกอริทึม Logistic Regression แสดงได้ดังภาพที่ 24 และด้วยวิธีการหาค่า SHAP ดังภาพที่ 25 ตามลำดับ



ภาพประกอบ 25 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงลบด้วยวิธี TF-IDF ร่วมกับ Logistic



ภาพประกอบ 26 คุณลักษณะที่สำคัญที่จำแนกประเภทเชิงลบด้วยวิธี SHAP

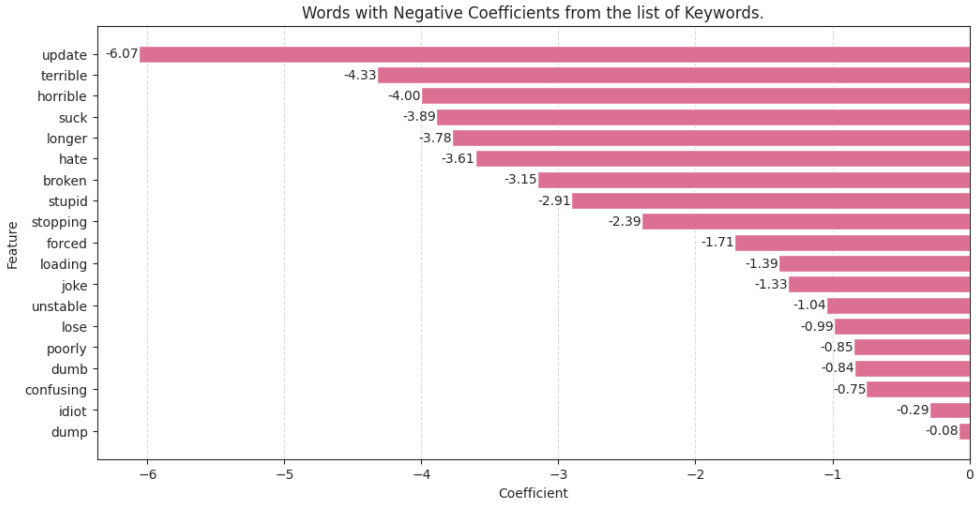
จากภาพการพล็อตกราฟผลลัพธ์คุณลักษณะที่สำคัญเชิงลบด้วยค่า Coefficients ของวิธี TF-IDF ร่วมกับ Logistic Regression 20 อันดับแรกคือคุณลักษณะหรือคำดังนี้ Cant, Worst, Play, Update, Event, Fix, Stop, Constantly, Buggy, Bad, Playing, Worse, Misinformation, Doesn't, Keep, Terrible, Randomly, Crash, Trash และ Wont ส่วนผลลัพธ์คุณลักษณะที่

สำคัญเชิงลบด้วยค่า SHAP ของวิธี SHAP explainer 20 อันดับแรกคือคุณลักษณะหรือคำดังนี้ Stupid, Confusing, Hate, Horrible, Broken, Terrible, Constantly, Forced, Idiot, Poorly, Update, Suck, Unstable, Dumb, Joke, Lose และ Loading ซึ่งคำที่มีร่วมกันระหว่าง 2 วิธี คือ คำว่า Update, Constantly และ Terrible สามารถสรุปได้ดังรูปแบบตาราง ดังนี้

ตารางที่ 21 สรุปคุณลักษณะที่สำคัญเชิงลบ

No	TF-IDF + Logistic Regression		SHAP	
	Word	Coef.	Word	SHAP
1	Cant	7.18	Stupid	0.70
2	Worst	6.79	Confusing	0.69
3	Play	6.15	Hate	0.63
4	Update	6.07	Horrible	0.59
5	Event	5.36	Broken	0.43
6	Fix	5.19	Terrible	0.40
7	Stop	5.00	Constantly	0.39
8	Constantly	4.99	Forced	0.39
9	Buggy	4.74	Idiot	0.39
10	Bad	4.69	Poorly	0.36
11	Playing	4.68	Update	0.36
12	Worse	4.55	Suck	0.35
13	Misinformation	4.53	Unstable	0.34
14	Doesnt	4.47	Dumb	0.34
15	Keep	4.39	Longer	0.33
16	Terrible	4.33	Stopping	0.32
17	Randomly	4.30	Dump	0.32
18	Crash	4.28	Joke	0.31
19	Trash	4.20	Lose	0.30
20	Wont	4.19	Loading	0.28

และมีการนำคุณลักษณะเชิงลบ 20 อันดับของวิธี SHAP มาเทียบหาคุณลักษณะเดียวกันในวิธี TF-IDF ร่วมกับ Logistic Regression พบว่าได้ค่า Coefficient พิจารณาเป็นเชิงลบทั้งหมด แต่ความสำคัญอาจจะน้อยกว่า ซึ่งแสดงผลลัพธ์การพล็อตกราฟได้ ดังภาพ



ภาพประกอบ 27 กราฟแสดงค่า coefficients คุณลักษณะที่สำคัญเชิงลบตามคำที่กำหนด

จากการหาคุณลักษณะที่สำคัญด้วยวิธีการวิธีหาค่า Coefficients จากวิธีการสกัดคุณลักษณะ TF-IDF ร่วมกับอัลกอริทึม Logistic Regression และการหาค่า SHAP Value ด้วยอัลกอริทึม SHAP Explainer ทั้งสองวิธีสามารถวิเคราะห์คุณลักษณะข้อมูลที่สำคัญ (Feature Importance) ได้ โดยผลลัพธ์จากวิธีการสกัดคุณลักษณะ TF-IDF ร่วมกับ Logistic Regression เน้นการคำนวณ Coefficients เพื่อหาคุณลักษณะที่สำคัญซึ่งเป็นผลลัพธ์โดยตรงจากโมเดลที่ถูกสร้างขึ้นผลล ในขณะที่ SHAP เน้นการอธิบายผลลัพธ์โดยการให้คำอธิบายว่าแต่ละคุณลักษณะมีผลอย่างไรต่อการทำนายโดยรวมของโมเดล การหาคุณลักษณะที่สำคัญเป็นส่วนที่ช่วยให้เข้าใจว่าคุณลักษณะหรือคำใดมีผลต่อการตัดสินใจในการจำแนกประเภทความรู้สึกของความคิดเห็นว่าเป็นความรู้สึกเชิงบวกหรือเชิงลบ

บทที่ 5
บทสรุป

5.1 สรุปผลการวิจัย

สารนิพนธ์การจำแนกประเภทความรู้สึกของรีวิวผู้ใช้บริการแอพสปอติฟายด้วยวิธีการเรียนรู้ของเครื่องนี้มีขั้นตอนหลักทั้งหมด 4 ขั้นตอน จากขั้นตอนแรกเป็นขั้นตอนการรวบรวมข้อมูลโดยใช้ข้อมูลที่ถูกรวบรวมไว้จากเว็บไซต์ Kaggle ชื่อชุดข้อมูล Spotify App Reviews ซึ่งเป็นข้อมูลรีวิวภาษาอังกฤษจำนวน 61,594 ข้อความ

ขั้นตอนที่ 2 เป็นขั้นตอนการเตรียมข้อมูลก่อนการวิเคราะห์ โดยจะทำการระบุประเภทความรู้สึกออกเป็น 2 ประเภท โดยใช้คะแนนรีวิวเป็นตัวแบ่งโดยไม่ใช่ความคิดเห็นที่ให้คะแนนเป็น 3 จะคงเหลือข้อมูลจำนวน 54,708 ข้อความ จากนั้นระบุประเภทความคิดเห็นที่มีคะแนนมากกว่า 3 ให้เป็นประเภทความคิดเห็นเชิงบวก มีจำนวน 29,937 ข้อความ และให้ความคิดเห็นที่มีคะแนนน้อยกว่า 3 เป็นประเภทความคิดเห็นเชิงลบ มีจำนวน 24,771 ข้อความ และนำเข้าสู่กระบวนการทำความสะอาดข้อมูลต่าง ๆ เพื่อให้เหมาะสมกับการนำข้อมูลไปใช้ในขั้นตอนต่อไป

ขั้นตอนที่ 3 เป็นขั้นตอนการสร้างโมเดล ซึ่งจะนำข้อมูลจากขั้นตอนที่ 2 มาทำการแบ่งออกเป็นชุดข้อมูลฝึก (Train Set) 75% และข้อมูลทดสอบ (Test Set) 25% และนำข้อมูลเข้าสู่การสกัดคุณลักษณะ (Feature Extraction) ด้วยวิธี TF-IDF และ Word2Vec เพื่อแปลงข้อมูลให้เป็นเวกเตอร์ (Vector) เพื่อให้สามารถนำชุดข้อมูลเข้าสู่การสร้างโมเดลสำหรับด้วยอัลกอริทึม Random Forest, Naïve Bayes, Logistic Regression, Support Vector Machine, XGBoost ในส่วนของ DistilBERT ใช้วิธีการแปลงข้อความเป็นเวกเตอร์และสร้างโมเดลด้วยชุดข้อมูลผ่านการฝึกมาแล้ว (Pretrain) ของ distilbert-base-uncased ซึ่งชุดข้อมูลมีการระบุการจำแนกประเภทด้วยชุดข้อมูลทดสอบ (Train set) ร่วมด้วย หลังจากนั้นนำโมเดลทั้งหมดที่สร้างขึ้นมาทดสอบกับชุดข้อมูลทดสอบ (Test set) และเปรียบเทียบประสิทธิภาพด้วยการวัดค่า F1-Score, Precision, Recall, ROC และ Accuracy พบว่าอัลกอริทึม DistilBERT มีประสิทธิภาพสูงที่สุดในการจำแนกประเภทโดยรวมสูงที่สุด โดยมีค่า F1-Score 91.05%, ค่า Precision 92.53%, ค่า Recall 89.62%, ค่า ROC 90.46% และค่า Accuracy 90.39%

หลังจากนั้นมีการหาคุณลักษณะที่สำคัญ (Feature Importance) ของประเภทความรู้สึกเชิงบวกและเชิงลบ 20 อันดับแรก ด้วยวิธีการหาค่า Coefficients จากวิธีการสกัด

คุณลักษณะ TF-IDF ร่วมกับอัลกอริทึม Logistic Regression และการหาค่า SHAP Value ด้วยอัลกอริทึม SHAP Explainer ผลลัพธ์คุณลักษณะการหาค่า Coefficients เป็นวิธีการที่ให้ผลลัพธ์โดยตรงจากการประมาณค่าพารามิเตอร์ของโมเดล โดยคุณลักษณะที่มีค่า Coefficients สูงเป็นบวกจะเป็นคุณลักษณะที่มีความสำคัญในการจัดอยู่ในกลุ่มความรู้สึกเชิงบวก และคุณลักษณะที่มีค่า Coefficients สูงเป็นลบจะเป็นคุณลักษณะที่มีความสำคัญในการจัดอยู่ในกลุ่มความรู้สึกเชิงลบ

ในขณะผลลัพธ์การหาค่า SHAP Value เป็นการอธิบายผลการทำนายของโมเดลภายหลังจากที่โมเดลถูกสร้างขึ้นแล้ว โดยค่า SHAP Value บวกสูงของคุณลักษณะใดหมายถึงคุณลักษณะนั้นมีผลสำคัญในการเพิ่มโอกาสที่ข้อความจะจัดอยู่ในกลุ่มความรู้สึกเชิงบวก และค่า SHAP Value ลบสูงหมายถึงคุณลักษณะนั้นมีผลสำคัญในการเพิ่มโอกาสที่ข้อความจะจัดอยู่ในกลุ่มความรู้สึกเชิงลบ ความแตกต่างที่สำคัญคือ Coefficients เป็นค่าที่ได้มาโดยตรงจากการประมาณค่าพารามิเตอร์ของโมเดล ในขณะที่ SHAP Value เป็นการอธิบายผลการทำนายของโมเดลภายหลัง ทั้งสองวิธีให้มุมมองที่แตกต่างกันในการวิเคราะห์คุณลักษณะที่สำคัญ ซึ่งเป็นประโยชน์ในการทำความเข้าใจโมเดลและข้อมูลที่ใช้ในการสร้างโมเดลมากยิ่งขึ้น และเป็นส่วนช่วยให้เราปรับปรุงและพัฒนาโมเดลให้มีประสิทธิภาพที่ดีขึ้นได้ในอนาคต

5.2 ข้อจำกัดในวิทยานิพนธ์

1) โมเดลที่สร้างขึ้นจะรองรับเฉพาะข้อความภาษาอังกฤษเท่านั้น เนื่องจากภาษาที่แตกต่างกันมีวิธีการจัดการข้อมูลหรือการตัดแบ่งคำที่ไม่เหมือนกัน

2) ข้อมูลรีวิวทั้งหมดมาจากเว็บไซต์ Kaggle เพียงแห่งเดียว อาจไม่ครอบคลุมความคิดเห็นของผู้ใช้ Spotify ทั้งหมด

3) ข้อมูลที่ใช้ในการสร้างโมเดลถูกติดป้ายความรู้สึกเชิงบวก และเชิงลบ ด้วยคะแนนซึ่งอาจไม่ได้สะท้อนถึงความรู้สึกที่แท้จริงซึ่งอาจส่งผลให้โมเดลมีการจำแนกประเภทความรู้สึกที่แท้จริงคลาดเคลื่อนไปได้

4) ภาษาที่ใช้ในการแสดงความคิดเห็นไม่ใช่ภาษาทางการ อาจมีคำแสลงหรือคำที่ไม่เกี่ยวข้อง อาจส่งผลต่อความคลาดเคลื่อนในการจำแนกประเภทได้

5) งานวิจัยมุ่งเน้นไปที่การจำแนกประเภทความรู้สึกเชิงบวกเชิงลบ ไม่ได้วิเคราะห์อารมณ์ (emotion) หรือความคิดเห็นเชิงลึก

5.3 ข้อเสนอแนะ

1) เพิ่มข้อมูลรีวิวจากแหล่งที่มาอื่น ๆ ที่มีภาษาและความรู้สึกที่แตกต่างกันอาจช่วยให้โมเดลมีประสิทธิภาพที่ดีขึ้นและทำให้มีความสามารถในการจำแนกความรู้สึกได้กว้างขึ้น

2) ติดป้ายด้วยการให้ผู้เชี่ยวชาญทางด้านอารมณ์หรือความรู้สึกแทนการใช้คะแนนอาจเป็นส่วนช่วยให้โมเดลจำแนกประเภทความรู้สึกได้ถูกต้องมากยิ่งขึ้น

3) เพิ่มการปรับปรุงคุณภาพโมเดลแต่ละตัวให้มีประสิทธิภาพเพิ่มมากขึ้น

4) สามารถนำไปพัฒนาต่อยอดในการจำแนกประเภทอารมณ์ (Emotion) ได้

5) นอกจากโมเดล DistilBERT ควรมีการทดสอบและประเมินโมเดล Deep Learning หรือโมเดล Transformer อื่น ๆ เพิ่มเติม เช่น BERT, LSTM, หรือ CNN เพื่อเปรียบเทียบประสิทธิภาพและหาวิธีการที่เหมาะสมที่สุดสำหรับงานจำแนกประเภท

6) ควรมีการแก้ไขคำสะกดผิด เนื่องจากข้อมูลความคิดเห็นมักมีปัญหาเช่น คำสะกดผิดหรือคำพูดไม่เป็นทางการ อาจทำให้โมเดลไม่สามารถจับความหมายของคำได้อย่างถูกต้อง ซึ่งอาจส่งผลกระทบต่อประสิทธิภาพของโมเดล

บรรณานุกรม

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

[2] Spotify AB, “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] M. W. B. Azlinah, M. Bee, and W. Yap, *Supervised and Unsupervised Learning for Data Science Unsupervised and Semi-Supervised Learning Series Editor: M. Emre Celebi*, 1st ed. Springer Cham, 2020. [Online]. Available: http://www.springer.com/series/15892

[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] L. Breiman, “Random Forests,” vol. 45, pp. 5–32, 2001.

[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] Y. Al-Amrani, M. Lazaar, and K. E. Elkadiri, “Sentiment Analysis using supervised classification algorithms,” *BDCA’17: Proceedings of the 2nd international Conference on Big Data, Cloud and ApplicationsMarch 2017Article No.: 61Pages 1–8https://doi.org/10.1145/3090354.3090417*, vol. 17, 2017, doi: 10.1145/3090354.3090417.

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

[12] C. Cortes, V. Vapnik, and L. Saitta, “Support-Vector Networks Editor,” Kluwer Academic Publishers, 1995.

[13] 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.

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

[15] David Martins, “XGBoost: A Complete Guide to Fine-Tune and Optimize your Model.”

[16] 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

[17] 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

[18] 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.

[19] Riccardo Di Sipio, “A Quick Guide to AUC-ROC in Machine Learning Models,” Towards Data Science.

[20] T. Fawcett, “Introduction to ROC analysis,” *Pattern Recognit Lett*, vol. 27, pp. 861–874, Mar. 2006, doi: 10.1016/j.patrec.2005.10.010.

[21] A. Salinca, “Business Reviews Classification Using Sentiment Analysis,” in *Proceedings - 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2015*, Institute of Electrical and Electronics Engineers Inc., Mar. 2016, pp. 247–250. doi: 10.1109/SYNASC.2015.46.

[22] M. R. Huq, A. Ali, and A. Rahman, “Sentiment Analysis on Twitter Data using KNN and SVM,” *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, 2017, [Online]. Available: www.ijacsa.thesai.org

[23] M. Mohaiminul and N. Sultana, “Comparative Study on Machine Learning Algorithms for Sentiment Classification,” *Int J Comput Appl*, vol. 182, no. 21, pp. 1–7, Oct. 2018, doi: 10.5120/ijca2018917961.

[24] K. Zahoor, N. Zakaria Bawany, and S. Hamid, “Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning,” *2020 21st International Arab Conference on Information Technology (ACIT)*, pp. 1–6, 2020.

[25] M. Kasri, E. Jadida Morocco, M. Birjali, and A. Beni-Hssane, “A comparison of features extraction methods for Arabic sentiment analysis,” *BDIoT’19: Proceedings of the 4th International Conference on Big Data and Internet of Things*, vol. 19, pp. 1–6, 2019, doi: 10.1145/3372938.3372998.

[26] K. Moolthaisong and W. Songpan, “Emotion Analysis and Classification of Movie Reviews Using Data Mining,” *2020 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA) : proceedings : Medan, Indonesia, July 16-17, 2020.*, 2020.

[27] A. A. Wadhe and S. S. Suratkar, “Tourist Place Reviews Sentiment Classification Using Machine Learning Techniques,” *2020 International Conference on Industry 4.0 Technology (I4Tech), Pune, India*, pp. 1–6, 2020.

[28] “Sentiment analysis and classification of Indian farmers’ protest using twitter data”, Accessed: Jul. 09, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667096821000124

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

[30] Scott Lundberg, “shap.Explainer.” Accessed: Feb. 01, 2024. [Online]. Available: https://shap.readthedocs.io/en/latest/generated/shap.Explainer.html