A Comprehensive Performance Analysis of Supervised Machine Learning Techniques for Sentiment Analysis

Korakot Matarat
Faculty of Science and Technology
Sakon Nakhon Rajabhat University
Sakon Nakhon, Thailand 47000

Chaidan Mingmuang
Faculty of Science and Technology
Sakon Nakhon Rajabhat University
Sakon Nakhon, Thailand 47000

Weerasak Charoenrat
Faculty of Science and Technology
Sakon Nakhon Rajabhat University
Sakon Nakhon, Thailand 47000

ABSTRACT
Sentiment analysis plays a crucial role in deciphering opinions and emotions expressed in textual data, with wide-ranging applications in business such as customer feedback analysis and social media monitoring. This paper conducts a thorough performance analysis of supervised machine learning algorithms in sentiment analysis, utilising the Wongnai reviews dataset, which comprises 40,000 reviews. By utilising a sophisticated preprocessing pipeline and conducting a comparative analysis of feature extraction methods, the research improves sentiment analysis by eliminating stop words (e.g., < > □% I < / # + -;- * & @ $). Subsequently, it will eradicate words that are meaningless for processing the text, for example, มี, เฉยๆ, เช่นใด, เพียงแต่, น้อยๆ, ข้างเคียง and hashtag removal, POS tagging, sentiment score computation, and TF-IDF analysis.

The research introduces a novel approach to dominant feature extraction, surpassing traditional bag-of-words methods. By applying six algorithms Logistic Regression (LR), Multinomial Naïve Bayes (NB), Decision Tree Classifier (DT), Neural Network (NN), Gradient Descent (SGD), and Support Vector Machine (SVC), the study compares their accuracy, precision, and recall values, revealing notable insights within the context of Wongnai reviews.

In conclusion, this paper not only contributes to understanding sentiment analysis performance but also serves as a valuable resource for optimising models in diverse domains. SVC emerges as the top-performing algorithm by achieving a 0.73 accuracy score, outclassing LR, NB, NN, and SGD with identical performances by achieving a 0.72 accuracy score, while DT exhibits the lowest performance. Further analysis combining TF-IDF with BoW shows improved performance by SGD and SVC by achieving a 0.74 accuracy score, reinforcing the superior performance of SVC in this experiment. This concise summary provides a foundation for practitioners and researchers engaged in sentiment analysis, aiding informed decision-making and paving the way for future exploration with advanced machine learning algorithms.

Keywords
Performance analysis, Supervised learning, Bag-of-words, TF-IDF analysis, Thai language data analysis, Sentiment analysis.

1. INTRODUCTION
The intricate tapestry of human communication is woven with threads of emotion and sentiment, making verbal expressions a distinctive feature that sets individuals apart. At the heart of understanding and deciphering these expressions lies the field of sentiment analysis, a domain dedicated to scrutinising the nuanced sentiments conveyed through language [1]. In the pursuit of unravelling, this research study places a spotlight on supervised learning, a machine learning methodology renowned for its prowess in conducting experiments across a diverse spectrum of queries. This method involves the initial training of a model with labelled sample data, allowing it to learn and categorise information effectively. The subsequent testing phase evaluates the model's ability to classify new, unseen data based on its acquired training. The accuracy of the classification serves as a metric for the model's performance. Within the realm of sentiment analysis, leverage the power of Term Frequency–Inverse Document Frequency (TF-IDF) and regression analysis to glean insights into users' thoughts. By combining TF-IDF and traditional bag-of-words (BoW) algorithms, this study comes up with a new way to dominate feature extraction that goes beyond the limitations of current methods [2]. Comparative Analysis of Machine Learning Algorithms: The initiative embarks on a comparative analysis of six machine learning algorithms. Logistic Regression (LR), Multinomial Naïve Bayes (NB), Decision Tree Classifier (DT), Neural Network (NN), Gradient Descent (SGD), and Support Vector Machine (SVC). Through a meticulous examination of accuracy, precision, and recall values, this study sheds light on significant findings in the context of Wongnai reviews. The allure of machine learning techniques and the intricate nature of sentiment analysis have attracted scholars to these burgeoning fields. This research envisions the construction of a high-performance intelligent system, showcasing the proficiency of sentiment analysis and machine learning in artificial intelligence. Yet, the complexity of selecting the most suitable machine learning technique poses a challenge for researchers, potentially leading to suboptimal results. The motivation for the investigation is to analyse the performance of available machine learning techniques for sentiment analysis in Thai language. Focused solely on supervised machine learning methods, this comparative study aims to guide researchers in choosing the most fitting technique for their specific requirements, ensuring optimal accuracy and model performance.

The subsequent sections of this paper are structured to provide a comprehensive exploration of this research. Section II reviews related works in the literature, offering context and insights. Section III elucidates the various supervised machine learning methods employed in this study. Section IV provides a detailed walkthrough of the methods applied. Section V presents the experimental data and discussions, while Section VI encapsulates this paper conclusions.

2. RELATEDWORK
The initial investigation focuses on rapid sentiment analysis methods applied to product recommendations. Two strategies, bag-of-words and bag-of-concepts, are employed, demonstrating precision and time efficiency for online implementation [3]. The study introduces Contextual Analysis

(CA), utilising a Hierarchical Knowledge Tree (HKT) structure. The suggested Tree Similarity Index (TSI) and Tree Differences Index (TDI) use supervised machine learning to analyse sentiment. They show that accuracy and the Tree Similarity Index are related in a good way [4]. Addressing the effectiveness of supervised machine learning, the study compares various algorithms and identifies Decision Trees (DT) as yielding the highest precision (approximately 90%) [5]. Exploring word embeddings as an enhancement to bag-of-words classifiers, the study demonstrates their superior accuracy in predicting sentiment challenges and improved representation of class distribution [6]. A systematic study uses feature extraction and preprocessing techniques on a tweet dataset, employing TextBlob for sentiment analysis. Among the five supervised machine learning models tested, the Extra Trees Classifier (ETC) demonstrates superior performance [7]. The study employs Apache Spark for sentiment analysis on big datasets of Amazon Fine Food reviews, implementing Linear SVC, Logistic Regression, and Naïve Bayes. Linear SVC outperforms in terms of efficiency, with over 80% accuracy [8]. This paper undertakes an analysis of customer evaluations pertaining to a variety of restaurants located throughout Karachi. He implemented a variety of classification algorithms, including the Naive Bayes Classifier, Logistic Regression, Support Vector Machine (SVM), and Random Forest. A comparison of the efficacy of each of these algorithms is provided. The implementation of a random forest algorithm [9]. Achieved optimal performance with 95% accuracy. This work introduces aspect-based sentiment analysis using two models with different approaches (TF-IDF and POS labeling). The models employ decision trees, KNN, Naïve Bayes, and Support Vector Machines to classify aspects based on sentiment polarity [10]. This research presents and covers the concept-sentiment analysis and its application in full. By referring to the procedure for carrying out sentiment analysis as well as its numerous real-world applications. A brief description of the recent approaches employed in the analysis has been provided, along with the suitable performance indicators that have been utilized. The study clarifies the necessity and significance of sentiment analysis [11]. This study presents a logistic regression methodology integrated with term and inverse document frequency (TF-IDF) to classify the sentiment of Arabic-language evaluations of services in Lebanon. Experiments reveal three fundamental results: 1) The confidence level of the classifier is high when predicting positive reviews. 2) The model's predictive accuracy skews towards negative sentiment evaluations. 3) The small proportion of negative reviews in the corpus exacerbates the uncertainty of the logistic regression model [12]. This study compares various text representation techniques for sentiment analysis in Turkish texts, revealing that Random Forest is the most efficient and pertinent method, particularly noteworthy for comparing word vectors with Bag of Words [13]. Using SentiWordNet and SVM, the study employs a rule-based sentiment analysis method on Indonesian texts. The accuracy varies based on dataset balancing, achieving 89% and 52% for SentiWordNet and SVM, respectively, in balanced datasets [14]. The article explores the application of ontology in sentiment analysis, emphasising the significance of domain-specific training and test data for successful sentiment classification [15]. A new ensemble learning model that combines Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and a serial model shows better accuracy and F1 score in analysing the sentiment of Arabic-language reviews [16]. The results of an examination using various Machine Learning (ML) and Lexicon investigation techniques are presented in this article. Analyzing the outcomes facilitates conducting an evaluation study and verifying the estimation of the current composition. This will facilitate subsequent researchers in comprehending current developments in the configuration of possibility examination [17]. A comparative analysis of three standard datasets and various supervised Machine Learning (ML) algorithms demonstrates their effectiveness in predicting sentiment values, serving as a foundation for further investigation [18]. Utilizing ML and Deep Learning (DL) algorithms, the study predicts potential pathogens or diseases affecting crops based on meteorological conditions, with the Artificial Neural Network achieving the highest accuracy [19]. The landscape of sentiment analysis on Twitter data reveals the Multilayer Perceptron (MLP) as a formidable performer, exhibiting optimal results in the face of extensive datasets [20]. The article looks at supervised learning and soft computing methods in stress diagnosis and suggests that problems could be solved by using a hybrid model that combines the best parts of each method [21]. The study focuses on opinion mining and sentimental analysis using a Kaggle hotel review dataset, highlighting the importance of quality control based on guest feedback for attracting customers to hotels [22]. Efforts to identify fraudulent reviews on e-commerce sites using TF-IDF and classification techniques (NB, SVM, AB, and RF) reveal high accuracy scores, with RF outperforming others [23]. A comparison of ML classification methods (SVM, NB, Logistic Regression, and Hybrid Algorithm) suggests the hybrid algorithm's superior accuracy in categorising text sentiments [24]. The proposed ensemble learning and neural network integrated method, utilizing four fundamental classifiers, shows effective sentiment polarity forecasting and enhanced precision in text sentiment analysis [25]. The method that has been presented consists of three discrete phases. Preprocessing is executed in the initial phase to refine and filter the text data. During the subsequent phase, the TF-IDF (Term Frequency and Inverse Document Frequency) method is employed to extract features. The investigations are conducted using a Twitter dataset for US Airlines that is accessible to the public. Various performance indicators, including accuracy, precision, recall, and F1 score, document the outcomes. In the end, the results produced by the support vector machine were the most pertinent [26]. The paper summarises current research and related works, emphasising the need for improved sentiment analysis performance and acknowledging the limitations of SVM algorithms [27]. The study compares sentiment lexicons from domain-specific and generic documents and finds that using a specific lexicon improves classification performance, especially for sentiments that are not neutral [28]. The proposed system utilises sentiment analysis and classification techniques (NB and RF) for recommending movies to users, with the Random Forest Algorithm demonstrating superior performance in terms of time and memory [29]. The study suggests that unlabeled data can enhance performance but should be handled carefully when comparing semi-supervised and supervised deep neural networks [30]. His investigated a variety of experiments in order to compare the efficacy of classifiers constructed with and without SGD learning, as well as those constructed with SGD learning and hyper-parameter discovery. SGD learning optimises the accuracy of the selected classifiers while decreasing the execution time and Perceptron classifiers, according to the findings of his research [31]. The research demonstrates that supervised techniques trained on labeled product reviews from different websites achieve comparable performance to unsupervised lexicon-based approaches [32].

These studies collectively advance sentiment analysis, underscoring the importance of diverse approaches and applications across various domains. Comparing different algorithms and methodologies provides valuable insights into the strengths and limitations of sentiment extraction techniques from textual data. This notably emphasises distinctively featuring reviews for reliable analysis. While the majority of research focuses on English text, this research delves into the Thai language, specifically leveraging Wongnai's review data.

The goal is to identify reliable methods for classifying review characteristics, enhancing the efficiency of evaluation data gathering. By employing Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF), contributing not only to methodological advancements but also to exploring innovative ways to present sentiment insights.

3. METHODOLOGY AND DATASET

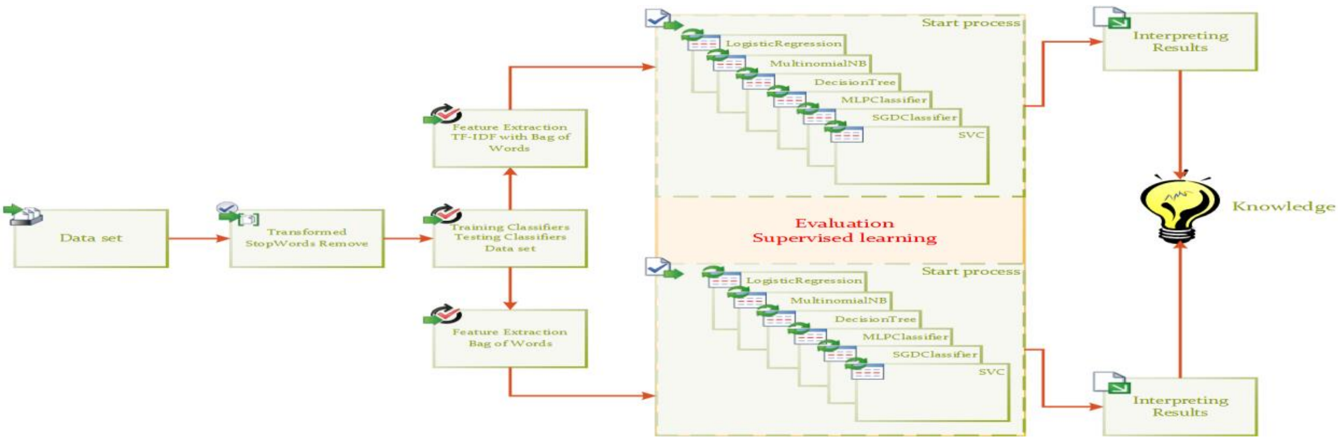

Fig 1: Workflow of the Comprehensive Performance Analysis Process

Several key factors justify the decision to employ the Wongnai reviews dataset as the research's primary data source. Firstly, the platform boasts a substantial user base in Thailand, as per Wongnai, providing a representative sample of user opinions and experiences. A crucial consideration in favour of selecting the Wongnai review dataset is its exclusive focus on textual content. While other social media platforms incorporate diverse content types such as videos, pictures, and web links, Wongnai exclusively features text-based reviews, offering a unique perspective [33].

However, it is essential to acknowledge both the positive and negative aspects of utilising the Wongnai dataset. On the positive side, the text-centric nature of the platform simplifies the data-gathering process. On the flip side, the dataset introduces challenges, including the presence of hashtags, slangs, and URL links within the textual content. As depicted in Figure 1, the next step in the research process involves leveraging the acquired information from the Wongnai review dataset. A more in-depth analysis of the dataset allows for a comprehensive understanding of user sentiments and opinions in the context of the study, providing further details and insights.

3.1 Preprocessing

Pre-processing enhances the effectiveness of sentiment analysis. To reduce complexity and improve data categorisation, it is therefore necessary to perform a cleansing process on the dataset. The initial step involves classifying datasets as positive or negative based on their scores, with scores of five or less categorised as negative and scores above five as positive. Afterwards, eliminate irrelevant words that do not contribute to the text processing, including "มี," "เฉยๆ," "เช่นใด," "เพียงแต่," "น้อยๆ," and "ข้างเคียง." Subsequent to the categorisation, the dataset undergoes tokenization, breaking down the text into tokens, followed by stemming to reduce words to their root form. The stemming process is crucial for minimising unnecessary word variations within a document. For this purpose, the research utilises the Natural Language Toolkit in Python, which includes a list of stop words to further refine the preprocessing phase. Figure 2 shows the details reviewed with tags as positive or negative.

By employing this meticulous pre-processing strategy, to ensure that the dataset is optimally prepared for sentiment analysis, thereby enhancing the accuracy and relevance of the subsequent analysis of user sentiments.

	review	tag
0	ร้านอาหารใหญ่มากกกกกกก \nเลี้ยวเข้ามาเจอห้องน้...	neg
1	อาหารที่นี่เป็นอาหารจีนแคะที่หากินยากในบ้านเรา...	pos
2	ปอเปี๊ยะสด ทุกวันนี้รู้สึกว่าหากินยาก (ร้านที่...	neg
3	ร้านคัพเค้กในเมืองไทยมีไม่มาก หลายๆคนอาจจะสงสั...	pos
4	อร่อย!!! เดินผ่านDigital gatewayทุกวัน ไม่ยักร...	pos
...	...	...
39995	รู้จักร้านนี้จากวงใน ร้านอยู่ในอาคารพาณิชย์ตรง...	pos
39996	ร้านซูชิอาหารญี่ปุ่น อยู่ตรงสะพานลอย เกษตรประต...	pos
39997	"Cantina Wine Bar & Italian Kitchen" ร้านเล็กๆ...	pos
39998	ร้านเค้กน่ารักๆ ตรงชั้นล่างของห้างเซ็นทรัลลาดพ...	neg
39999	วันนี้มากินกันไกลถึงแม่สอดจังหวัดตากติดกับชายแ...	neg

40000 rows × 2 columns

Fig 2: Review with tag as positive or negative

3.2 Training

To ensure the effectiveness of machine learning models that rely on learning and experience, it is crucial to divide the dataset into two distinct sets: the training set and the testing set. In this paper, It have carefully allocated the dataset into two subsets, maintaining an optimal proportion of 70% for training and 30% for testing. By exposing the model to a diverse range of examples during training, it can learn robust patterns. The testing set, separate from the training data, serves as an unbiased measure of the model's proficiency, assessing how well it can apply its acquired knowledge to new, unseen data. This methodology aims to strike a balance, fostering a well-trained and reliable machine learning model.

3.3 Feature Extraction

To leverage machine learning models for review analysis, employ feature extraction techniques to distil meaningful information. Two widely used methods for this purpose are Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF). BoW simplifies the text by representing it as an unordered set of words, disregarding grammar and word order. On the other hand, TF-IDF evaluates the importance of words based on their frequency in a document relative to their occurrence across the entire dataset, emphasising unique and relevant terms. These techniques play a crucial role in transforming raw review data into structured features, facilitating the effective training and utilisation of machine learning models.

TF-IDF, a feature extraction technique in text analysis, enhances performance by assigning weights to features. This method calculates the weight of each document feature by dividing the inverse document frequency (IDF) by the term frequency (TF). Term frequency (TF) reflects the prevalence of a feature within a document and is proportional to the document's length. The inverse document frequency (IDF) factor ensures that common terms across documents receive lower weights, emphasising the significance of unique terms in characterising the content. In summary, TF-IDF is instrumental in capturing the importance of features in text data, enabling more nuanced and effective analysis [2]. It can be defined as:

$$tf_{t,d} = \frac{count_{t,d}}{totalcount_d}, \tag{1}$$

Where count (t, d) is the number of term t in the document d and totalcountd is the total number of all terms in the document d. IDF measures the extent of a term t being informative in a document for model training [7]. It can be computed as:

$$idf = N/Df_{t,} \tag{2}$$

Where *N* is the number of documents in the corpus and *Dft* is the number of documents that contain the term *t*. IDF measures the weight of a term t low when term t occurs frequently in many documents. For instance, stopwords have a low IDF value. Finally, TF-IDF can be defined as:

$$tf - idf = tf_{t,d} * \log(idf). \tag{3}$$

Bag of Words (BoW) stands out as a fundamental feature extraction technique extensively used in information retrieval and Natural Language Processing (NLP) tasks. Primarily employed in text classification, BoW trains models based on word frequency within documents, creating a vocabulary from the occurrence frequencies of unique words. To boost the effectiveness of machine learning models, a suggested approach involves the concatenation of BoW and Term Frequency-Inverse Document Frequency (TF-IDF) features. This combination, illustrated in Figure 3, proves advantageous for enhancing model precision. By incorporating both techniques, this harnesses the strengths of BoW's simplicity and TF-IDF's nuanced weighting, contributing to a more robust and accurate representation of text data for machine learning tasks.

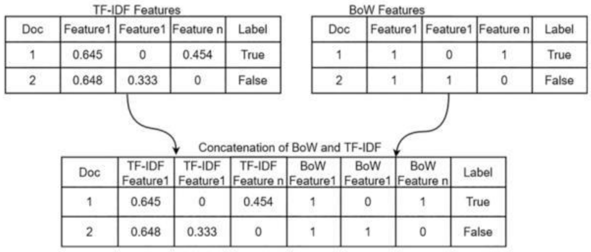

Fig 3: Our proposed approach based on concatenation of BoW and TF-IDF [7]

3.4 Evaluation Supervised learning

To assess the performance of the machine learning models, this employed four evaluation parameters: the accuracy score, the precision score, the recall score, and the F1 score. The accuracy metric quantifies the proportion of accurate predictions. The accuracy score has the potential to reach a maximum of 1 and a minimum of 0. The accuracy score is given as:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

Four basic notations are explained as follows.

True Positives (TP): The number of correct positive predictions of a class.

True Negatives (TN): The number of correct negative predictions of a class.

False Positives (FP): The number of incorrect positive predictions of a class.

False Negatives (FN): The number of incorrect negative predictions of a class.

Precision indicates the exactness of the classifiers. It lies in [0, 1] and calculated as

$$precision = \frac{TP}{TP+FP} \tag{5}$$

Recall indicates about the completeness of a classifier. It lies in [0, 1] and calculated as

$$recall = \frac{TP}{TP+FN} \tag{6}$$

$F1$ score is a harmonic mean of precision and recall scores. It lies in [0, 1] and calculated as

$$F_1 = 2 * \frac{(precision*recall)}{precision+recall} \tag{7}$$

3.4.1 Logistic regression

Scientists widely employ logistic regression, an algorithm for machine learning. Logistic regression is a technique that analyses the vectors of variables, calculates coefficients for input expressions, and subsequently determines the class of text using a word vector. The logistic regression function determines multiple linear functions, denoted as:

$$Logit(p) = \beta 0 + \beta 1X1 + \beta 2X2 \ldots \beta kXk \tag{8}$$

P represents the probability of the occurrence of the feature. X1, X2…. Xk represents the value of predictor and B1β2 …. Bk represents the model’s intercept [8].

3.4.2 Multinomial Naïve Bayes

The supervised machine learning technique used for classification tasks is the basis of the Bayesian theorem. The Bayesian theorem is explained in [8].

$$P\left(\frac{a}{b}\right) = \frac{P(a)P(b/a)}{P(b)} \quad (9)$$

3.4.3 Decision Trees

Decision tree learning is a supervised learning approach used in statistics, data mining, and machine learning. In this formalism, a classification or regression decision tree is used as a predictive model to draw conclusions about a set of observations. The theorem is explained following [5].

$$(x, Y) = (x_1, x_2 x_3, \ldots, x_k, Y) \quad (10)$$

The dependent variable, *Y*, is the target variable that are trying to understand, classify or generalize. The vector *X* is composed of the features, *x1, x2, x3 etc*., that are used for that task.

3.4.4 MLPClassifier

An MLP model, the most fundamental type of deep neural network, consists of a sequence of completely connected layers. These techniques have the capability to surmount the substantial computational power demands that contemporary deep learning architectures impose. MLP is highly suitable for problems involving classification and regression prediction [20]. Perform the calculation as follows:

$$y = \varphi(\sum_{i=1}^{n} w_i x_i + b) = \varphi(w^T X + b) \quad (11)$$

Where w is weight vector, x is input vector, b is bias and phi is non-linear activation function respectively.

3.4.5 SGDClassifier

The optimization algorithm Gradient Descent (GD) is widely used to find the optimal coefficients in a given condition, minimizing the cost associated with inaccurate predictions. The cost function, a crucial element in GD, evaluates the discrepancy between predicted and actual values for each sample in the training set. It is calculated as represented by the text contents for each Dk by extracting the numeric feature vectors. Therefore, feature vector extractor (φ) computes each vector feature for each input φ (Dk) = {φ1(Dk),…, φd(Dk) }, where φ(D) ∈ Rd is a point in the dimensional space. Moreover, the parameter vector that specifies the contributions of each feature vector to the prediction process is P = {P1… Pd}, where P ∈ Rd. Consequently, this can be mathematically calculated (f) by compiling both φ (D) and P [30].

$$f = \varphi(D).P \quad (12)$$

3.4.6 Support Vector Machine

Support Vector Machine (SVM) is the methodology of a vector-oriented model that is related to the transformation of text documents into feature vectors, which are then processed for classification. The primary goal is to find the separating hyperplane with the help of support vectors. Currently, SVM is considered to be the most accurate supervised classification approach. The performance of SVM is independent of the size of the training data. The equation is:

$$W^T x = -b \quad (13)$$

The support vectors are defined by the sigmoidal function:

$$f(x) = sign(w^t x + b) \quad (14)$$

Where, the classes are represented in a bipolar manner i.e. +1 and -1. Here one class can be represented by a value of -1, and other class can be represented by a value of +1 [18].

4. RESULTS AND ANALYSIS

The implementation of this research was carried out using the Python Programming Language. The dataset comprised a total of 40,000 records, with 25,000 instances in the positive class and 15,000 instances in the negative class. This split the dataset into training and testing sets, following an optimal proportion of 70:30, to effectively train and evaluate the models. Figure 4 provides a visual representation of the classification report for both the positive (pos) and negative (neg) classes. This report offers insights into the performance of the machine learning models in distinguishing between positive and negative instances. The strategic distribution of the dataset and the subsequent classification report contribute to a comprehensive understanding of the model's effectiveness in handling the given data.

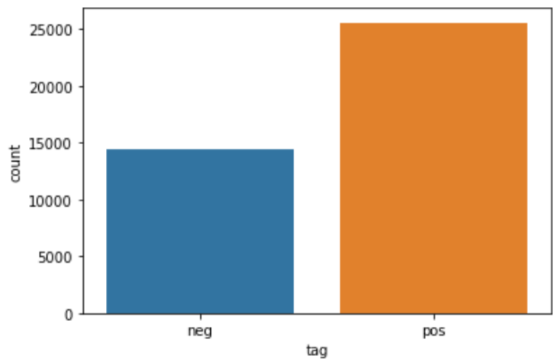

Fig 4: **Detail for positive class and negative**

In this experiment, the initial data set was represented as a Bag of Words. The training sample size was configured to be 70% of the entire dataset, and a random state of 30 was set prior to the experiment's execution. Six different algorithms were implemented for result prediction: Logistic Regression (LR), Multinomial Naive Bayes (MNB), Decision Tree (DT), Multi-Layer Perceptron (MLP), Stochastic Gradient Descent (SGD), and Support Vector Machine (SVC). Upon evaluation, the Support Vector Machine algorithm consistently produced more precise results compared to the alternative algorithms. To determine the top-performing algorithm among the six prediction algorithms, a Wongnai review dataset was used for training. Table 1 presents the classification reports for all algorithms, showcasing their performance metrics and providing insights into their predictive capabilities. The results indicate that the Support Vector Machine algorithm stands out as the most effective in this context.

Table 1. Accuracy of BoW

	LR	**MNB**	**DT**	**MLP**	**SGD**	**SVC**
Accuracy	0.72	0.72	0.61	0.72	0.72	0.73
Precision	0.7	0.72	0.6	0.71	0.72	0.74

	2		1			
Recall	0.7 2	0.72	0.6 2	0.72	0.72	0.74
F1Measur e	0.7 2	0.72	0.6 1	0.72	0.72	0.72

Table 1 details the results of comparing six applied algorithms. The precision and recall values for Logistic Regression (LR), Multinomial Naive Bayes (MNB), Multi-Layer Perceptron (MLP), and Stochastic Gradient Descent (SGD) are each reported as 0.72. In contrast, the Support Vector Machine (SVC) exhibits a higher accuracy of 0.73. To provide a visual representation of the comparison, Figure 5 illustrates the performance metrics of the six algorithms used in this investigation. Furthermore, the figure highlights the superior accuracy achieved by the Support Vector Machine (SVC) in comparison to the other algorithms, providing a comprehensive overview of their effectiveness.

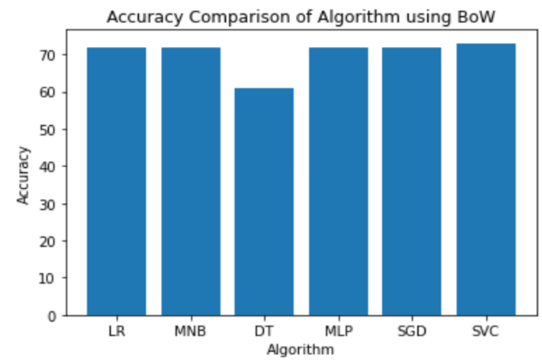

Fig 5: Accuracy comparison of algorithm using BoW

The objective of incorporating both Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) features was to improve the precision of machine learning models. This hypothesises that combining these two feature sets can enhance processing and positively impact the model's accuracy, especially when trained on well-defined data. The results of this study are shown in Table 2, providing more detailed performance metrics for models trained using the combined TF-IDF and BoW features. This table serves as a valuable reference for understanding the effectiveness of the combined feature approach in enhancing model precision. To gain deeper insights into the impact of feature concatenation on the overall accuracy of the machine learning models.

Table 2. Accuracy of TF-IDF and BoW

	LR	**MN B**	**DT**	**ML P**	**SG D**	**SV C**
Accuracy	0.6 9	0.66	0.6 1	0.71	0.72	0.75
Precision	0.6 9	0.73	0.6 1	0.71	0.72	0.74
Recall	0.6 9	0.66	0.6 1	0.71	0.72	0.75
F1Measur e	0.6 9	0.54	0.6 1	0.71	0.72	0.73

The study conducted a comparative analysis of six algorithms leveraging the concatenation of Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) features, and the results are presented in Table 2. Notably, the Support Vector Machine (SVC) stands out with an impressive accuracy score of 0.75, showcasing its effectiveness in leveraging the combined features. To provide a visual representation of the performance contrast among the algorithms, Figure 6 offers a graphical comparison. This figure serves as a concise and accessible overview of the accuracy achievements of the six algorithms employed in the investigation. The superior performance of the Support Vector Machine (SVC) is evident in the results, emphasising its proficiency in handling the concatenated TF-IDF and BoW features. Further detailed analysis may be beneficial to uncover additional insights into the nuances of each algorithm's performance.

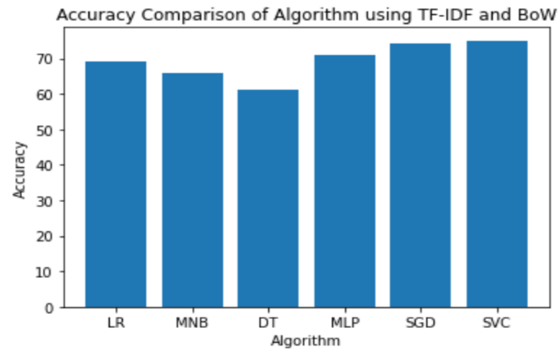

Fig 6: Accuracy comparison of algorithm using TF-IDF and BoW

The effectiveness of each classifier is assessed using the area under the ROC curve (AUC) and the macro-F1 score. In Wongnai review detection, the AUC is particularly significant because the decision threshold regulates the compromise between true and false positive rates. The F1 score is a metric that integrates precision and recall for each class. Among these, the macro F1 score offers the most comprehensive evaluation of the classifier SVC's overall performance. A divide between the training and test sets is used to calculate an interval of 0.69 within the acceptable range. The mean AUC and F1 are based on the accuracy of the TF-IDF and BoW confidence intervals. Figure 7 depicts the empirical results obtained from fitting the models using all previously described features.

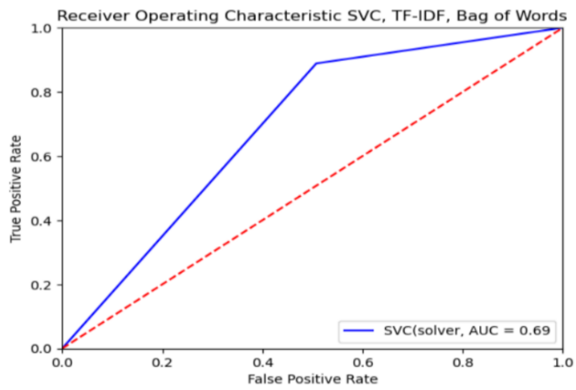

Fig 7: Receiver Operating Characteristic (ROC)

In Figure 8, the evaluation of the show system's categorisation performance relies on text mining-specific criteria. The assessment involves Support Vector Classifier (SVC) and considers both Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) features. The classification outcomes for all Wongnai reviews, categorised into two classes: positive and negative, are systematically captured by a confusion matrix, from which these performance metrics are derived. Using SVC along with feature extraction methods like TF-IDF and BoW lets us get a full picture of how well the system sorts reviews into the right categories using text mining rules.

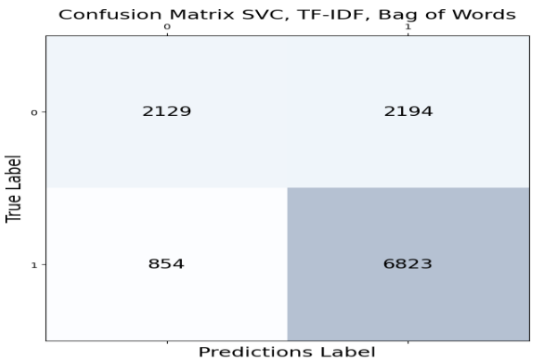

Fig 8: Confusion matrix

5. CONCLUSIONS

The sentiment analysis research was done with the Bag of Words dataset, utilising a 70% training sample size and a random state of 30. Among six algorithms, the Support Vector Machine (SVC) consistently outperformed others, achieving an accuracy of 0.75. This exploration of the TF-IDF and Bag of Words features validates hypotheses, aiming to enhance model precision. The Python version, which used a 40,000-case dataset and split it up, gave useful information about how well the model worked. Figure 7 displays the Receiver operating characteristic curve (ROC) classification reports, and Figure 8 displays the confusion matrix. In summary, this research significantly advances sentiment analysis methodologies by emphasising algorithmic performance and the impact of feature concatenation. Addressing challenges posed by hashtags and slangs highlighted the need for adaptive approaches, crucial for accurately capturing diverse user opinions in online reviews. Ultimately, this work underscores the importance of careful dataset selection, meticulous preprocessing, and ongoing methodological refinement. This study not only contributes to sentiment analysis but also extends implications to natural language processing, emphasising context and dataset-specific considerations in algorithmic development.

6. REFERENCES

[1] Abdul, M., Abdul, K., and Abu, K. (2019). Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset. 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, pp. 266 - 270, Nov. 2019

[2] Alexander, M., Elmina, H., Francisco, C. and Ofer, E. (2021). Sentiment analysis using TF–IDF weighting of UK MPs' tweets on Brexit. Knowledge-Based Systems,Vol**.** 228**,** Sep. 2021

[3] Abdulwahab, A. and Mustafa, A. (2019). Sentiment Analysis of Product Reviews Using Bag of Words and Bag of Concepts. IJEIE, Vol. 11. No.2. pp.49-60, Dec. 2019

[4] Azwa, A. and Andrew, S. (2019). Predicting Supervise Machine Learning Performances for Sentiment Analysis Using Contextual-Based Approaches. in IEEE Access, vol. 8, pp. 17722-17733, Dec. 2020

[5] Devansh, S., Arun, S. and Sudha, P. (2022). Sentimental Analysis Using Supervised Learning Algorithms. ICCAKM, Dubai, United Arab Emirates, pp. 1-6, Dec. 2022

[6] Elena, R., Martin, H., Matthias, W. and Marcelo, J. (2018). More than Bags of Words: Sentiment Analysis with Word Embeddings. Communication Methods and Measures, Vol. 12, No. 2, pp. 140-157, Apr. 2018

[7] Furqan, R., Madiha, W., Vaibhav, R. and Arif, Mehmood. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. PLoS ONE. Vol. 16, No. 2, Feb. 2021

[8] Hafiz, M., et al. (2021). Sentiment Analysis of Online Food Reviews using Big Data Analytics. Elementary Education Online. Vol. 20, No. 2, pp. 827-836, Apr. 2021

[9] Kanwal, Z., Narmeen, B. and Soomaiya, H. (2020). Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning. ACIT, Giza, Egypt, pp. 1-6, Jan. 2020

[10] Kotagiri, S., and Mary, S. (2019). Aspect Based Sentiment Analysis using POS Tagging and TFIDF. IJEAT, Vol. 8, No. 6, Aug. 2019

[11] Manasee, G. (2015). The Process of Sentiment Analysis: A Study. International Journal of Computer Applications, Vol. 126, No. 7, Sep. 2015

[12] Marwan, O., Moustafa, H., Nacereddine, H. and Amani, S. (2019). Sentiment Classifier: Logistic Regression for Arabic Services' Reviews in Lebanon. International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, pp. 1-5, May. 2019

[13] Metin, B., and Haldun, K. (2019). Sentiment Analysis with Term Weighting and Word Vectors. The International Arab Journal of Information Technology, Vol. 16, No. 5, Sep. 2019

[14] Mohammad, F., and Riyanarto, S. (2019). A comparative study of sentiment analysis using SVM and SentiWordNet. Indonesian Journal of Electrical Engineering and Computer Science, Vol. 13, No. 3, pp. 902-909, Mar. 2019

[15] Mohd, Y., Muhammad, L. and Liyana, Z. (2019). A Review on Sentiment Analysis Techniques and Applications. IOP Conference Series Materials Science and Engineering, Vol. 551, Fab. 2019

[16] Mohamed, C. et al. (2021). LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model. IJACSA, Vol. 12, No. 7, Jul. 2021

[17] Pooja, M. and Sharnil, P. (2020). A Review on Sentiment Analysis Methodologies, Practices and Applications. IJSTR, Vol. 9, No. 2, Feb. 2020

[18] Raj, P. et al. (2016). Comparative Evaluation of Supervised Learning Algorithms for Sentiment Analysis of Movie Reviews. International Journal of Computer Applications, Vol. 142, No. 1, May. 2016

[19] Rutuja, R., Sumit, K. and Ruchi, R. (2022). Comparison of Artificial Intelligence Algorithms in Plant Disease Prediction. Revue d'Intelligence Artificielle, Vol. 36, No. 2, pp. 185-193, Apr. 2022

[20] Manjula, D. et al. (2023). Twitter Sentiment Analysis using Collaborative Multi Layer Perceptron (MLP) Classifier. ICCCI, Coimbatore, India, pp. 1-6, May. 2023

[21] Samriti, S., Gurvinder, S. and Manik, S. (2021). A comprehensive review and analysis of supervisedlearning and soft computing techniques for stress diagnosis in humans. Computers in Biology and Medicine, Vol. 134, Jul. 2021

[22] Samruddhi, K. (2019). Classification Model to Predict the Sentiment of Hotel Review. IRJCS, Vol. 6, No. 6, Jun. 2019

[23] Saleh, N. et al. (2022). Data Analytics for the Identification of Fake Reviews Using Supervised Learning. Computers, Materials & Continua, Vol. 70, No. 2, Sep. 2022

[24] Siva, P. et al. (2019). Feature-Based Opinion Mining for Amazon Product's using MLT. International Journal of Innovative Technology and Exploring Engineering, Vol. 8, No. 11, Sep. 2019

[25] Siyin, L. et al. (2021). Research on Text Sentiment Analysis Based on Neural Network and Ensemble Learning. Revue d'Intelligence Artificielle, Vol. 35, No. 1, pp. 63-70, Feb. 2021

[26] Satyendra, S., Krishan, K. and Brajesh, K. (2022). Sentiment Analysis of Twitter Data Using TF-IDF and Machine Learning Techniques. International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, Faridabad, India, pp. 252-255, May. 2022

[27] Shamsa, U. et al. (2018). Sentiment Analysis Approaches and Applications: A Survey. International Journal of Computer Applications, Vol. 181, No. 1, Jul. 2018

[28] Tanatorn, T., Nuttapong, S. and Udomsak, D. (2020). Sentiment Classification on Thai Social Media Using a Domain-Specific Trained Lexicon, ECTI-CON, Phuket, Thailand, pp. 580-583, Jun. 2020

[29] Tejaswini, M. and Choudhari, G. (2019). Implementation of Sentiment Classification of Movie Reviews by Supervised Machine Learning Approaches. ICCMC, Erode, India, pp. 1197-1200, Mar. 2019

[30] Vivian, L. et al. (2019). Semi-supervised Learning for Sentiment Classification using Small Number of Labeled Data. The Fifth Information Systems International Conference, Vol. 161, pp. 577-584, Jan. 2019

[31] Shadi, D. (2018). Optimizing Stochastic Gradient Descent in Text Classification Based on Fine-Tuning Hyper-Parameters Approach. IJCSIS, Vol. 16, No. 12, Dec. 2018

[32] Waqar, M. et al. (2020). Sentiment analysis of Product Reviews in the Absence of Labelled data using Supervised Learning Approaches. Malaysian Journal of Computer Science, Vol. 32, No. 2, pp. 118-132, Apr. 2020

[33] Korakot, C. (2021). Wongnai corpus. https://github.com/wongnai/wongnai-corpus