Journal of Informatics and Vocational Education (JOIVE)
Vol.9, No.3, November 2026, pp. 114~124
ISSN: 2746-7813

Performance Evaluation of Support Vector Machine and Naïve Bayes Methods in Mining Sentiment from Shopee Application Reviews

Dandi Azaidane[1], Muhammad Masrur Aji Dorojatun[2], Bisma Satrio Bimantoro[3], Anggraini Puspita Sari*[4], Addien Haniefardy[5]
[1,2,3,4,5] Department of Informatics, Universitas Pembangunan Nasional "VETERAN" Jawa Timur

Article Info

Article history:

Received Jun 26, 2026
Revised Jul 15, 2026
Accepted Nov 1, 2026

Corresponding Author:

Anggraini Puspita sari, Department of Informatics, Universitas Pembangunan Nasional "VETERAN" Jawa Timur, Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294
Email: anggraini.puspita.if@upnjatim.ac.id

ABSTRACT

Shopee is one of the top e-commerce platforms in Indonesia, which millions of consumers share their experience through an online review. These reviews provide a good insight into customer satisfaction and the quality of services provided by the platform. This research is a study to classify the general opinion of Shopee customer reviews by utilising Multinomial Naïve Bayes and Support Vector Machine (SVM) approach. The dataset was collected from Kaggle, initially containing 99,999 Shopee user reviews. As our work focuses on binary sentiment classification, we deleted reviews with a 3-star rating, leaving 65,910 tagged reviews. The research process included text preprocessing (case folding, cleaning, tokenization, stopword removal and stemming), sentiment labelling based on user ratings, feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF), splitting the dataset into 52,728 training and 13,182 testing reviews with an 80:20 ratio and data balancing on the training dataset using Synthetic Minority Over-sampling Technique (SMOTE). The classification performance was examined based on accuracy, precision, recall, F1-score and confusion matrix. The experimental findings indicated that the Multinomial Naïve Bayes classifier obtained 82.89% accuracy, 96.55% precision, 79.45% recall and 87.17% F1-score. In contrast, the Support Vector Machine classifier attained an accuracy of 83.23%, precision of 94.67%, recall of 81.66% and F1-score of 87.69%. Both the models performed similar but SVM classifier got slightly better accuracy, recall and F1-score with less classification errors than Multinomial Naïve Bayes. The results imply that both techniques are applicable for binary sentiment classification of Shopee user evaluations, and that SVM provides somewhat better overall classification performance.

Keywords: Sentiment Analysis, *Naïve Bayes*, *Support Vector Machine*, TF-IDF, SMOTE.

1. INTRODUCTION

In the last many years, the development of digital technology and e-commerce users in Indonesia has grown quickly. Shopee is one of the e-commerce platforms with the most subscribers in Indonesia (Asih, 2024). The platform is commonly used by the public because of the various benefits supplied by the platform such as easy transactions, promotion programs and fast delivery services (Yasmine et al., 2025). The number of users has also increased as the number of reviews and comments submitted by the individuals. Such reviews can provide useful information on the user enjoyment, quality of the service and problems found when using the program (Octaviana et al., 2026).

Sentiment analysis (SA) is a part of Natural Language Processing (NLP) that attempts to find out views, sentiments or judgements written by humans about any specific object using text input (Lin, 2020). Sentiment analysis can be used in e-commerce systems to gauge the impression of users about the services offered (Rana et al., 2025). Such information can assist companies in evaluating their services, improving application quality, and understanding user needs more effectively (Saber Ismail et al., 2024). However, the large volume of review

data makes manual analysis inefficient and time-consuming (Maulida et al., 2026). Therefore, an automated classification method is needed to accurately categorize user sentiments into positive and negative classes. Several machine learning algorithms commonly used for sentiment classification are *Naïve Bayes* and *Support Vector Machine* (SVM) (Rahmat et al., 2025). The *Naïve Bayes* algorithm is known for its simple and fast computational process in text data processing, while SVM has strong capabilities in finding the optimal separating boundary (hyperplane) between classes, often resulting in high classification accuracy (Lase et al., 2026).

Numerous studies have previously been conducted on sentiment classification using *Naïve Bayes* and *Support Vector Machine* algorithms (Jelita & Saad, 2025). Several studies have shown that Naïve Bayes performs well in text classification due to its simple computational process and effectiveness in handling high-dimensional data (Aisah et al., 2023). On the other hand, *Support Vector Machine* is widely used because it can provide stable classification performance when dealing with complex textual data (Prasetyo et al., 2021). Nevertheless, the performance of these two algorithms may vary depending on dataset characteristics, preprocessing techniques, and word-weighting methods employed (Hate et al., 2026). Hence, more research on the performance comparison of both algorithms on Shopee user review data is required.

This study seeks to classify sentiment of customer reviews on Shopee using Naïve Bayes and Support Vector Machine techniques. The classification method will be done after the review data has been preprocessed in multiple levels such as case folding, stopword removal and stemming. The performance of both algorithms will also be assessed and compared in terms of accuracy, precision, recall and F1-score measures. The conclusion of this study is expected to provide suggestions on the best appropriate algorithm for sentiment classification of Shopee user reviews and a reference for further research on sentiment analysis based on machine learning.

2. RESEARCH METHOD

This study employs a machine learning approach by implementing two classification algorithms, namely Naïve Bayes and Support Vector Machine (SVM), to classify sentiments in Shopee user reviews. The dataset used in this research was obtained from the Kaggle platform and processed using the Python programming language. The research process begins with data collection and concludes with conclusions and recommendations. The overall research workflow is illustrated in Figure 1.

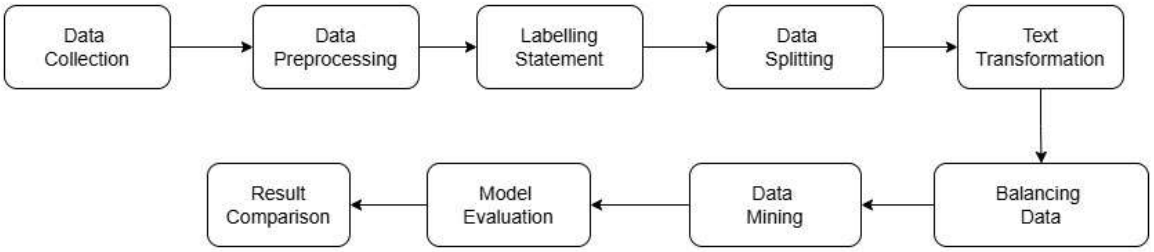

Figure 1. Research Flow

2.1. DATA COLLECTION

The data used in this study were taken from the Kaggle platform in the form of Microsoft Excel (.xlsx) consisting of as many as 99.999 users' review of the Shopee application. The dataset includes four attributes: userName, score, at and content. Then, the data went through many rounds of processing such data preprocessing, sentiment labelling, data balancing with SMOTE, feature extraction using TF-IDF, and Naïve Bayes and SVM classification methods. The initial verification was made by the Python code to verify the number of rows and columns and the names of the attributes. The verification results showed that the dataset contains 99.999 records and 4 attributes: userName, score, at, and content.

2.2. DATA PREPROCESSING

The textual data has been pre-processed to clean and prepare the data and the sentiment classification approach has been utilised on the textual data (Mumtaz et al., 2025). The first step is case folding, all characters are changed to lower case letters so that we have a homogeneous data structure (Moeslim et al., 2025). The second step involved the cleaning procedure to remove URLs, numbers, punctuation marks, symbols, special characters and other unnecessary sections of the text (Wijaya et al., 2025). Then, the sentences were tokenised to split the phrases into words or tokens (Astuti et al., 2024). Next stopword removal This is the method of deleting the most used words that do not convey any relevant sentiment for sentiment analysis (Pratama & Ghozi, 2025). Finally, the inflected words were translated to its root form using Sastrawi library for stemming. The result of the preprocessing phase is a cleaner, structured text collection ready for feature extraction.

In the implementation, the preparation step is implemented from the case folding stage where each review is changed to the lower case using the case_folding (text) function. This is to normalise the data and remove discrepancies caused by upper case letters . Then the cleaning(text) technique removes URLs, user mentions, non-alphabetic letters and redundant white spaces to keep only the relevant textual data. The sentences are tokenised with the word_tokenize function that separates the sentences into tokens. These tokens are then joined into a string for the subsequent preparation procedures .

The stopword removal method addresses the removal of frequent terms that contribute little to the sentiment analysis, and focuses on keywords with high semantic relevance using a specific stopword list. The next step of preparation is stemming which will change the words into root form using the cached_stem function of the Sastrawi library. A caching scheme is provided for improving the efficiency of processing. All steps of the preparation are included in the preprocess(text) function. It is applied to each review with the progress_apply method utilising the tqdm library. This technique gives us a new column clean_review with cleaned tokenised stopword free and stemmed text data ready for feature extraction and sentiment classification.

Table 1. Text Preprocessing Process

Stage	Description	Example
Case Folding	Convert all text to lowercase	"BAGUS" → "bagus"
Cleansing	Remove URLs, user mentions, numbers, symbols, and irrelevant characters	"@shopee 123!" → "shopee"
Stopword Removal	Remove common words that do not contribute meaningful information to sentiment analysis	"dan", "atau", "yang" dihapus
Tokenisasi	Split sentences into individual word tokens	"produk bagus" → ["produk", "bagus"]
Stemming (Sastrawi)	Convert inflected words into their root forms using the Sastrawi library	"pengiriman" → "kirim"

2.3. LABELLING STATEMENT

Sentiment labeling is the task of giving a category to a textual data so as to capture the underlying opinion trend represented in the text . In binary sentiment analysis, each review is classified into one of the two main classes, positive sentiment and negative sentiment. This technique is necessary to obtain a clear objective for the classification algorithm to learn the word patterns from the reviews (Sari et al., 2024). In this study, sentiment labels were assigned based on the user ratings available in the Shopee review dataset. Reviews with ratings of 4 and 5 were labeled as positive sentiment (label 1) because they generally indicate user satisfaction with the application, products, or services. Conversely, reviews with ratings of 1 and 2 were labeled as negative sentiment (label 0) since they typically represent user dissatisfaction regarding product quality, pricing, shipping services, customer support, or overall shopping experience.

The original dataset consisted of 99.999 Shopee user reviews. Since this study focuses on binary sentiment classification, reviews with a 3-star rating were considered neutral and excluded from the analysis.

Removing neutral reviews allows the classification models to focus on learning the distinctive characteristics of positive and negative sentiments while reducing ambiguity during the training process. After excluding all neutral reviews, the final dataset contained 65.910 review instances, which were subsequently used for feature extraction, data splitting, model training, and evaluation.

Table 2 presents several examples of Shopee user reviews after the preprocessing and sentiment-labeling stages. The table illustrates how the original review texts were transformed into cleaned textual representations before being assigned sentiment labels based on their ratings. Reviews with ratings of 4 and 5 were labeled as Positive (1), whereas reviews with ratings of 1 and 2 were labeled as Negative (0).

Table 2. Example of Reviews after Preprocessing and Sentiment Labeling

Content	Clean Review	Rating	Label
Sangat membantu..sekali	membantusekali	5	Positive (1)
Tempat belanja online no 1, harga lebih murah daripada di pasaran...	belanja online no harga murah pasar kirim cepat barangnya realpict harga kualitas saran banyakin voucher potong harga gretong	5	Positive (1)
yaudah	yaudah	5	Positive (1)
Ongkir nya mahal...	ongkir mahal masak mahal ongkir dr harga payahhhhhhhhhhhh	1	Negative (0)
bagus	bagus	5	Positive (1)

2.4. DATA SPLITTING

The preprocessed dataset was then tagged with sentiment and features extracted. The dataset was split into training and test with ratio 80:20. The classification models were trained on the training set and evaluated on the testing set to test the models' ability to categorise unseen data. The original dataset contained 99.999 Shopee user reviews. However, this study was limited to binary sentiment classification and the reviews with 3-star ratings (neutral sentiment) were not included in the analysis. Therefore, we used 65,910 review instances with just positive and negative sentiments as the final dataset for model building.

The dataset was then split into 52.728 training occurrences (80%) and 13.182 testing examples (20%). We used the training data to train the models to learn the sentiment patterns in the review texts. The testing data were used just for the evaluation of the classification performance on the unseen data. The data splitting was done using the stratify parameter to maintain the ratio of positive and negative sentiment classes in both training and testing datasets. This method guarantees that every subset has a class distribution that is representative, hence, it provides a fair and unbiased evaluation of the classification models.

2.5. TEXT TRANSFORMATION

The preprocessed data were then converted in numerical using Term Frequency-Inverse Document Frequency (TF-IDF) . This method re-weights words according to their frequency within a document and their relative value in the total dataset. TF-IDF transformation yields a feature matrix which can be used as input for the classification algorithms (Adhiatma & Qoiriah, 2022). The preparation stage has been done effectively like the text data in the form of strings in the clean_review column. Moreover, each review was labelled with an emotion label, which provided a link between the input features and the target classes before feature extraction.

The maximum number of features was 5,000 by limiting the TF-IDF vectorizer to extract features and the model can select the most informative terms from the training data set. The vectorizer was fit exclusively on the training data to prevent any information leaking from the test data. Thus, the training data has been translated into a TF-IDF matrix of 52,728 × 5,000 and the testing data has been transformed into a matrix of 13,182 × 5,000. Each of these matrices has rows for a review document and columns for a unique phrase selected by the TF-IDF vectorizer. The max_features parameter was set to 5.000 to maintain the most informative sentences, while limiting the feature dimensionality and computational complexity. This option is

a trade-off between the preservation of relevant textual information and the effectiveness of the classification models. Then the calculated TF-IDF matrices were used as input to the classification model of Naïve Bayes and the classification model of Support Vector Machine (SVM). This way of counting words, algorithms are able to recognise patterns of word-weight that correspond to pleasant and negative moods. Therefore, TF-IDF is essential to link the text preparation phase and the sentiment classification models, transforming the unstructured textual data into meaningful numerical features utilised in machine learning.

2.6. BALANCING DATA

After the dataset was divided into training and testing sets, data balancing was performed only on the training dataset using the Synthetic Minority Over-sampling Technique (SMOTE). This approach was employed to address the class imbalance problem by generating synthetic samples for the minority class without altering the original testing data. Applying SMOTE exclusively to the training set prevents data leakage and ensures that the evaluation results accurately reflect the models' ability to classify previously unseen data (Aprihartha et al., 2024).

Before applying SMOTE, the training dataset consisted of 38.561 positive reviews (label 1) and 14.167 negative reviews (label 0), indicating a substantial class imbalance in which the positive class was the majority. To overcome this issue, SMOTE generated synthetic samples for the minority class until both sentiment classes contained an equal number of instances. After applying SMOTE, the training dataset became perfectly balanced, with 38.561 instances in the positive class and 38.561 instances in the negative class, resulting in a total of 77.122 training instances. The testing dataset remained unchanged to preserve its original data distribution and provide an unbiased evaluation of the classification models. Balancing the training data reduced the tendency of the classification algorithms to favor the majority class and enabled the models to learn representative patterns from both positive and negative sentiments more effectively. Consequently, this approach improved the robustness of the sentiment classification models while maintaining the validity of the evaluation process.

The balanced training dataset enabled both the Multinomial Naïve Bayes and Support Vector Machine (LinearSVC) models to learn representative patterns from both positive and negative sentiment classes more effectively. Consequently, this balancing strategy reduced the tendency of the models to favor the majority class, improved their capability to recognize minority-class instances, and enhanced the reliability of the sentiment classification results.

Table 3. Data Distribution Comparison

	Before SMOTE	After SMOTE
0	14.167	38.561
1	38.561	38.561

2.7. DATA MAINING

This work uses two supervised machine learning methods, namely Multinomial Naïve Bayes (MNB) and Support Vector Machine (SVM) to classify the Shopee customer reviews into positive and negative sentiment classes. The whole process consists of text preparation, sentiment labelling, data division, feature extraction, data balancing, model training and model evaluation. The data set was pre-processed and sentiment labelled, and then divided into training and testing sets at a ratio of 80:20. It produced 52.728 training reviews and 13.182 testing reviews. The split was done using the stratify parameter to maintain the proportion of sentiment classes in both subsets. Then, the text data is converted into numerical feature vectors by the use of Term Frequency - Inverse Document Frequency (TF-IDF) technique with maximum number of 5.000 features. The TF-IDF vectorizer was fitted on the training data only and then used to transform the test data with the learned vocabulary. This strategy prevents the leakage of information from the test data during feature extraction .

Since the dataset is imbalanced in classes Synthetic Minority Over-sampling Technique (SMOTE) was applied on the training data only after TF-IDF transformation. The SMOTE generated synthetic samples for the minority class until both sentiment classes had 38.561 samples, leading to a balanced training set of 77.122 reviews. The testing dataset was not deliberately manipulated, therefore its original distribution was retained and an unbiased evaluation of the classification models could be undertaken. The first classifier was

Multinomial Naïve Bayes (MultinomialNB) which is well suited for the text classification since it predicts the posterior probability of each sentiment class based on the word-frequency data given by TF-IDF. The model is built on conditional independence across features and was implemented using default parameters in the Scikit-learn module. The second model was Support Vector Machine (SVM) built with LinearSVC classifier and random_state=42. The linear kernel was chosen because the textual data described by TF-IDF are often high dimensional sparse feature vectors for which linear SVM has shown excellent classification performance while retaining the computational efficiency. The random_state was set to 42 in an effort to ensure the reproducibility of the experimental results. Finally, both models were tested on the original testing data set which was not involved in the SMOTE procedure and the TF-IDF fit process. The model was tested using Accuracy, Precision, Recall, F1-score and Confusion matrix. Then the comparative results are examined to discover the most effective algorithm for sentiment categorisation of Shopee customer reviews.

Table 4. Model Configuration

Parameter	Configuration
Initial dataset	99.999 reviews
Final dataset	65.910 reviews
Training data	52.728
Testing data	13.182
TF-IDF	max_features = 5000
Data balancing	SMOTE (Training only)
Naïve Bayes	MultinomialNB()
SVM	LinearSVC(random_state=42)
Evaluation	Accuracy, Precision, Recall, F1-score, Confusion Matrix

2.8. MODEL EVALUATION

Model evaluation was performed using metrics including Confusion Matrix, Accuracy, Precision, Recall, and F1-Score. The Confusion Matrix was used to examine the number of correct and incorrect predictions of each sentiment class. This evaluation approach gives precise information about the classification performance of the model by detecting true positives, true negatives, false positives and false negatives.

The model's overall correctness was measured using accuracy. While, Precision, Recall and F1-Score were used to evaluate the model's performance in the classification of positive and negative feelings. Precision measures the fraction of true positives out of all positive predictions. Recall assesses the model's ability to detect all real positive cases, whereas F1-Score gives a balanced assessment by integrating Precision and Recall into a single score. The findings produced from the Naïve Bayes and Support Vector Machine (SVM) algorithms were compared to determine which method performed the best for the sentiment analysis of Shopee application user reviews.

2.9. RESULT COMPARISION

The performance of Naïve Bayes and Support Vector Machine (SVM) approach were compared with evaluation metrics from testing dataset. To analyze the performance of each algorithm in classifying the sentiments of Shopee user reviews into positive and negative classes, we compared the Accuracy, Precision, Recall and F1-Score.

The evaluation results reveal that both systems can classify the sentiment appropriately. However, performance differences were observed in evaluation indicators. These differences reflect the different characteristics of each algorithm in learning and recognizing sentiment patterns through textual input. Naïve Bayes is a probabilistic classifier based on word features probability distribution whereas SVM determines an ideal decision boundary separating sentiment groups based on feature representations. Accuracy was employed to measure the overall correct classification results. The precision was calculated as the ratio of correctly predicted cases to the total number of instances predicted as a particular class. Recall measured the model's ability to find all relevant items in a class, while F1-Score was a balanced statistic that combined Precision and Recall into a single score.

Thus, the comparison of different evaluation metrics can give us a comprehensive picture of the advantages and disadvantages of each method. The better the algorithm's score in the assessment ratings, the more accurate and consistent it is in classifying user moods. The comparison analysis obtained the basis for the most suitable algorithm for sentiment categorization of Shopee's customer reviews. The higher performing

algorithm can be determined by examining all the assessment metrics in concert. The evaluation outcomes were shown in the previous section. The outputs of this comparison provide us with significant information about the performance of machine learning algorithms for sentiment analysis and can be utilized as a reference for future studies on e-commerce review classification.

3. RESULT AND ANALYSIS

The *Multinomial Naïve Bayes* and *Support Vector Machine* (SVM) models were tested on the test dataset of 13.182 reviews of which 20% was used for testing after removing the neutral (3-star) reviews from the labelled dataset . The labelled dataset had 65,910 instances of reviews; 52,728 reviews were used for training and 13,182 reviews were held out for testing. The testing dataset was not subjected to the SMOTE balancing method and preserved its original class distribution. To correct class imbalance without data leaking during model evaluation, SMOTE was only used on the training dataset. The same testing dataset was used to both classification models under the same experimental parameters to enable a fair comparison of their abilities to classify positive and negative feelings .

3.1. NAÏVE BAYES

The Multinomial Naïve Bayes classifier was performed using a testing dataset of 13,182 Shopee user reviews that were not used in the training procedure. The testing dataset was not subject to the SMOTE balancing process and hence preserved its original class distribution, providing an unbiased evaluation of the classification model. The evaluation results show that the Naïve Bayes model has 83% accuracy, which indicates that the classifier could properly predict the sentiment in most review situations.

The classification result shows that the negative sentiment class (label 0) has obtained precision of 0.62, recall of 0.92 and F1-score of 0.74. The high recall value shows that the model was good at detecting most of the unfavourable reviews. But the somewhat lower precision shows that some reviews anticipated to be unfavourable were really in the positive class.

For the positive sentiment class (label 1), the model attained precision of 0.97, recall of 0.79 and F1-score of 0.87. The high precision suggests that the positive reviews were typically classified properly, while the low recall shows that some positive reviews were labelled as negative. The whole macro-average precision was 0.79, macro-average recall was 0.86, and macro-average F1-score was 0.81. In the meantime, the weighted-average precision, recall, and F1-score were 0.87, 0.83, and 0.84, respectively. These results show that Naïve Bayes classifier did rather well in differentiating positive and negative attitudes, while the performance was affected by the imbalance of the original testing dataset.

	precision	recall	f1-score	support
0	0.62	0.92	0.74	3542
1	0.97	0.79	0.87	9640
accuracy			0.83	13182
macro avg	0.79	0.86	0.81	13182
weighted avg	0.87	0.83	0.84	13182

Figure 2. Classification Report Naïve Bayes

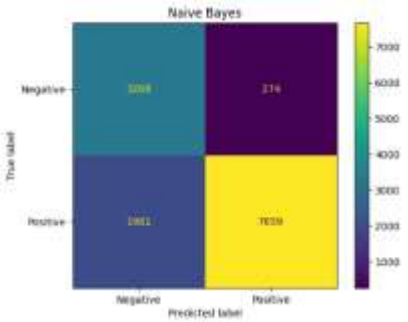

Figure 3. Confusion Matrix Naïve Bayes

The confusion matrix provides a more detailed view of how well the classification performed. 3.542 unfavourable reviews of which 3.268 were accurately identified as negative (True unfavourable) 274 were wrongly tagged as positive (False positive). The same situation 9640 positive reviews. 7659 were accurately identified as positive (True Positive) while 1981 were misclassified as negative (False Negative).

The model properly classified 10.927 reviews of 13.182 testing reviews whereas it wrongly classified 2.255 reviews in total. The high recall of the negative class indicates that the Naïve Bayes classifier is useful for detection of consumer complaints and unfavourable attitudes. However, the increase in the number of false negatives implies that some positive decisions were forecasted as negative. Thus, the Naïve Bayes approach is good in the binary sentiment classification, although the optimisation can increase its performance to find the positive sentiment reviews.

3.2. SUPPORT VECTOR MACHINE (SVM)

The performance of the Support Vector Machine (SVM) classifier was evaluated using a testing dataset of 13,182 Shopee customer reviews that were not used in the training procedure. The testing dataset was kept with its original distribution of classes and was not balanced using SMOTE, to avoid biassing the evaluation of the classification model. The SVM classifier achieved an accuracy of 83% from the assessment findings, which means the model accurately identified most of the review instances. As per the classification report, the negative sentiment class (label 0) had a precision of 0.64, recall of 0.87 and F1-score of 0.74. The reasonably high recall implies that the model was able to identify most unfavourable reviews, however the precision value reveals that some reviews predicted as negative actually belonged to the positive class.

The model performance for the positive sentiment class (label 1) was as follows: precision: 0.95, recall: 0.82, F1-score: 0.88. The high precision implies that most of the reviews labelled as positive were truly positive evaluations, while the recall shows that a section of good reviews were nevertheless misclassified as negative. Furthermore, the macro-average precision, recall and F1-score were 0.79, 0.85 and 0.81 respectively, while the weighted-average precision, recall and F1-score were 0.86, 0.83 and 0.84 respectively. The outcomes suggest that the SVM classifier performed reliably in identifying positive and negative attitudes, even with an uneven class distribution in the original testing dataset.

	precision	recall	f1-score	support
0	0.64	0.87	0.74	3542
1	0.95	0.82	0.88	9640
accuracy			0.83	13182
macro avg	0.79	0.85	0.81	13182
weighted avg	0.86	0.83	0.84	13182

Figure 4. Classification Report SVM

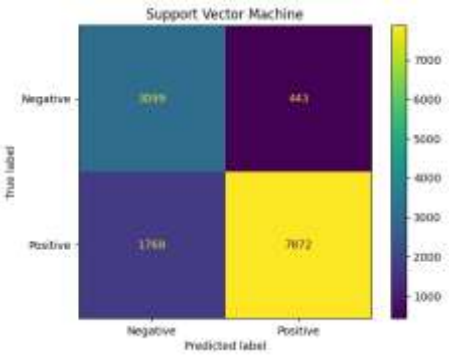

Figure 5. Confusion Matrix SVM

The misclassification matrix gives a more complete summary of the classification results. Out of the 3542 negative assessments, 3099 were accurately classified as negative (True Negative) and 443 were incorrectly classified as positive (False Positive). Similarly , out of 9640 , 7872 positive assessments were correctly predicted positive (True Positive) and 1768 were wrongly predicted negative (False Negative) . The training set included 13,182 reviews and the SVM classifier was able to detect 10,971 reviews and misclassify 2,211 reviews. The Naïve Bayes approach accurately predicted 11,027 reviews and misclassify 2,255 reviews. In SVM model there were 44 more correct predictions and 44 less classification errors. The findings showed that the classification accuracy of SVM classifier was somewhat better than Naïve Bayes model.

Both classifiers had the same overall accuracy (83%), however the SVM model performed better in precision for the negative sentiment class (0.64 vs. 0.62) and had fewer false negative predictions (1,768 vs. 1,981). The results indicated that the SVM classifier performs better in the aspect of balanced classification and the positive and negative opinion of the Shopee customer review dataset is significantly superior.

3.3. ANALYSIS

The Multinomial Naïve Bayes model achieves an accuracy of 82.89% and shows that it classifies the maximum number of the testing reviews correctly. For the class of positive emotion, the model acquired a precision of 96.55%, recall of 79.45% and F1-score of 87.17%. The high precision indicates the majority of anticipated positive reviews were categorised correctly, while the poor recall implies that some positive reviews were still misclassified as negative. The confusion matrix shows that the model successfully identified 10,927 out of 13,182 testing reviews, and misclassified 2,255 evaluations. There were 1981 favourable evaluations categorised as negative (false negatives), which implies that the model was prone to miss some positive feelings. This behaviour is connected to the Naïve Bayes assumption of conditional independence of features . However, in sentiment analysis the meaning of a word is usually dependent on its surrounding context, therefore making this assumption less effective to capture contextual link between terms.

The Support Vector Machine (SVM) model produced a somewhat higher accuracy of 83.23% and F1-score of 87.69%. Further, the model achieved a precision and recall of 94.67% and 81.66% on positive mood class. The SVM model properly identified 10,971 testing reviews and misclassified 2,211 reviews according to the confusion matrix. SVM accurately detected 44 more reviews than Naïve Bayes. The number of classification mistakes was reduced by 44 occurrences compared to Naïve Bayes. The incorrect negative prediction also reduced from 1,981 to 1,768 showing that SVM was better at identifying positive sentiment evaluations. This is because SVM is capable of constructing an optimal decision boundary in the high-dimensional TF-IDF feature space, which enables a greater distinction between positive and negative sentiment classes.

Table 5. Performance Comparison of Naïve Bayes and SVM

Method	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	0.828933	0.965461	0.794502	0.871678
SVM	0.832271	0.946723	0.816598	0.876859

SVM has a higher recall (81.66% vs 79.45%) and somewhat better F1 score (87.69% vs 87.17%) compared to Naïve Bayes although precision of the positive class for SVM (94.67%) is a bit lower than Naïve Bayes (96.55%). This suggests that the SVM has a better trade-off between correctly identifying positive assessments and reducing the frequency of missing positive cases . The improvement was however quite slight, indicating the two techniques performed similarly in the Shopee review dataset. In general, the performance of both classification models was competitive to predict positive and negative customer reviews of Shopee . The evaluation results demonstrate that Support Vector Machine (SVM) has somewhat better performance than Multinomial Naïve Bayes in the total evaluation metrics, with the best accuracy (83.23%), recall (81.66%), and F1-score (87.69%). However, the two classifiers performed similarly with only 0.43 percentage points increase in accuracy and 0.52 percentage points increase in F1-score . The results indicate that the SVM model provides the maximum performance in general, but both models are practical for binary sentiment classification of Shopee user reviews with SVM having a tiny edge in classification performance in this study .

4. CONCLUTION

This study examined the performance of the algorithm of Multinomial Naïve Bayes and Support Vector Machine (SVM) on binary sentiment classification of Shopee customer reviews. The studies were conducted on 65,910 labelled reviews after deleting neutral (3-star) ratings. The dataset was segmented into 52,728 training reviews and 13,182 testing reviews. The Synthetic Minority Over-sampling Technique (SMOTE) was only applied to the training data to solve class imbalance and prevent data leakage when evaluating the model.The experimental findings demonstrate that Multinomial Naïve Bayes classifier has an accuracy of 82.89%, precision of 96.55%, recall of 79.45% and F1-score of 87.17%. The model showed high precision, which means that the reviews that it predicted as favourable were usually properly classified. However, the somewhat reduced recall shows that some of the positive evaluations were still misclassified as unfavourable. The Support Vector Machine (SVM) classifier obtained an accuracy of 83.23%, precision of 94.67%, recall of 81.66% and F1 score of 87.69%. SVM achieved slightly higher accuracy, recall and F1-score, and fewer misclassified reviews than Multinomial Naïve Bayes. However, the improvement was quite slight, demonstrating that both classifiers performed similarly on the Shopee review dataset.The comparison analysis reveals that Support Vector Machine (SVM) had the greatest overall performance in this study with the highest

accuracy and F1-score among the models tested. However, the performance difference between SVM and Multinomial Naïve Bayes was small, showing that both techniques are appropriate for binary sentiment categorisation of Shopee customer evaluations. While SVM offered a better trade-off between precision and recall, Multinomial Naïve Bayes remained an interesting alternative because to its simplicity, computational economy and competitive classification performance.Moreover, the class imbalance was resolved by applying the SMOTE only on the training dataset; this did not cause any data leakage into the assessment. The classification models were able to learn more representative patterns of sentiment and the original distribution of the testing data was preserved, leading to a more accurate evaluation of the models performance. The results of this study indicate that both machine learning algorithms are able to classify sentiment in Shopee customer reviews correctly. Future work may involve the consideration of larger and more diversified datasets, the use of sophisticated text representation techniques such as contextual word embeddings, or the evaluation of deep learning and transformer-based models to further improve the effectiveness of sentiment classification.

REFERENCES

Adhiatma, F. D., & Qoiriah, A. (2022). Penerapan Metode TF-IDF dan Deep Neural Network untuk Analisa Sentimen pada Data Ulasan Hotel. *JINACS (Journal of Informatics and Computer Science)*, *xx*, 183–193.

Aisah, I. S., Irawan, B., & Suprapti, T. (2023). ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS SENTIMEN ULASAN APLIKASI AL QUR ' AN DIGITAL. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3759–3765.

Aprihartha, M. A., Putrawan, Z., Zulhan, D., & Nurfaizal, F. A. (2024). Algoritma Synthetic Minority Oversampling Technique dan C5.0 dalam Mengatasi Ketidakseimbangan Data pada Klasifikasi Kelulusan Siswa. *UPGRADE: Jurnal Pendidikan Teknologi Informasi*, *2*(1), 1–10. https://doi.org/https://doi.org/10.30812/upgrade.v2i1.4148

Asih, E. M. (2024). Analisis pada Shopee sebagai E-Commerce Terpopuler di Indonesia. *Jurnal Ekonomi Bisnis Antartika*, *2*, 73–79. https://doi.org/https://doi.org/10.70052/jeba.v2i1.299

Astuti, K. C., Firmansyah, A., & Riyadi, A. (2024). Implementasi Text Mining untuk Analisis Sentimen Masyarakat terhadap Ulasan Aplikasi Digital Korlantas Polri pada Google Play Store. *Remik: Riset Dan E-Jurnal Manajemen Informatika Komputer*, *8*(1), 383–394. https://doi.org/http://doi.org/10.33395/remik.v8i1.13421

Hate, I., Detection, S., & Bidirectional, U. (2026). Indonesian Hate Speech Detection Using Bidirectional Long Short-Term. *RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(158), 302–309.

Jelita, H. P., & Saad, M. I. (2025). Penerapan Algoritma Naïve Bayes Dalam Analisis Sentimen Masyarakat Terhadap STMIK Widya Cipta Dharma. *Bulletin of Information Technology (BIT)*, *6*(2), 148–160. https://doi.org/10.47065/bit.v5i2.2029

Lase, F. O., S, Y. P., & Lase, K. J. D. (2026). Penerapan Natural Language Processing Dalam Klasifikasi Sentimen Komentar Youtube Tentang Judi Online. *TIN: Terapan Informatika Nusantara*, *6*(9), 1593–1602. https://doi.org/10.47065/tin.v6i9.9256

Lin, X. (2020). Sentiment Analysis of E-commerce Customer Reviews Based on Natural Language Processing. *Proceedings of the 2020 2nd International Conference on Big Data and Artificial Intelligence*, *March*, 32–36. https://doi.org/10.1145/3436286.3436293

Maulida, A., Irawan, B., & Ramdhan, N. A. (2026). Analisis Sentimen Ulasan Wisata Alun-Alun Brebes pada Google Maps Menggunakan Support Vector Machine. *TIN : Terapan Informatika Nusantara*, *6*(8), 1398–1406. https://doi.org/10.47065/tin.v6i8.8883

Moeslim, A. G. S., Firmansyah, E., & Sutara, B. (2025). Perbandingan Kinerja XGBoost dan IndoBERT untuk Klasifikasi Teks Kesehatan Bahasa Indonesia. *DSI : DATA SCIENCES INDONESIA*, *5*(2), 62–72. https://doi.org/https://doi.org/10.47709/dsi.v5i2.7281

Mumtaz, J. A., Komariansyah, K. K., Holik, W., Gumelar, M. G., Pratama, R., & Lestari, H. R. (2025). Analisis Sentimen Ulasan Aplikasi HeyJapan di Google Play Store Menggunakan Algoritma NLP. *Pragmatik : Jurnal Rumpun Ilmu Bahasa Dan Pendidikan*, *3*. https://doi.org/https://doi.org/10.61132/pragmatik.v3i3.1801

Octaviana, N., Yumami, E., & Wahana, D. (2026). Analisis Sentimen Ulasan Dan Komentar Pengguna Pada Aplikasi Toco Menggunakan Algoritma Support Vector Machine. *Universitas Dian Nuswantoro*, *25*(2), 284–294.

Prasetyo, V. R., Benarkah, N., & Chrisintha, V. J. (2021). Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya. *TEKNIKA*, *10*(2), 114–121. https://doi.org/10.34148/teknika.v10i2.370

Pratama, A. Y., & Ghozi, W. (2025). Prediksi Potensi Kinerja Calon Karyawan Customer Service Call Center Menggunakan Model Machine Learning Berbasis Data Rekrutmen. *Building of Informatics, Technology and Science (BITS)*, *7*(1), 201–212. https://doi.org/10.47065/bits.v7i1.7285

Rahmat, Rahim, A., & Arbansyah. (2025). PERBANDINGAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI ALIBABA DI GOOGLE PLAY STORE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(2), 3050–3057.

Rana, M. R. R., Nawaz, A., Rehman, S. U., Abid, M. A., Garayevi, M., & Kajanová, J. (2025). BERT-BiGRU-Senti-GCN: An Advanced NLP Framework for Analyzing Customer Sentiments in E-Commerce. *International Journal of Computational Intelligence Systems*, *18*(1), 1–18. https://doi.org/10.1007/s44196-025-00747-1

Saber Ismail, D. W., Ghareeb, M. M., & Youssry, H. (2024). Enhancing Customer Experience through Sentiment Analysis and Natural Language Processing in E-commerce. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *15*(3), 60–72. https://doi.org/10.58346/JOWUA.2024.I3.005

Sari, A. P., Prasetya, D. A., Aditiawan, F. P., & Haromainy, M. M. Al. (2024). PREDIKSI GANGGUAN KESEHATAN MENTAL PADA KALANGAN MAHASISWA MENGGUNAKAN METODE PSEUDO-LABELING DAN ALGORITMA REGRESI LOGISTIK. *TAMIKA: Jurnal Tugas Akhir Manajemen Informatika & Komputerisasi Akuntansi*, *4*(2), 40–48. https://doi.org/https://doi.org/10.46880/tamika.Vol4No2(SEMNASTIK).pp40-48

Wijaya, A., Dwiyanto, M. A., Muttaqin, Z., & Haq, M. A. F. (2025). Sentiment Analysis ff Netflix Review on Google Play with Machine Learning. *Jurnal Kecerdasan Buatan, Komputasi Dan Teknologi Informasi*, *6*(1), 59–66.

Yasmine, P., Yesada, A., Kusuma, E., Sadewa, W., Prasetyo, D., Ekonomi, F., & Surabaya, U. A. (2025). Peran Fitur - fitur pada Shopee dalam Kecanduan Belanja GEN Z. *Jurnal Akademik Ekonomi Dan Manajemen*, *2*(4), 651–664. https://doi.org/https://doi.org/10.61722/jaem.v2i4.7789