SAINSTECH
INNOVATION JOURNAL

SIJ, Volume 9, Nomor 2, May 2026, Hal. 621-633

Sentiment Analysis of Positive and Negative User Reviews for TikTok and YouTube Applications on the Google Play Store Using Naïve Bayes and Support Vector Machine (SVM) Algorithms

[Rakyatol Hasanah][1)*], [Syahrani Lonang][2)], [Ahmad Fatoni Dwi Putra][1)]

[1)] [Program Studi Ilmu Komputer, Universitas Qamarul Huda Badaruddin Bagu, Lombok Tengah, Nusa Tenggara Barat, Indonesia)]
[2)] [Program Studi Teknologi Informasi, Universitas Qamarul Huda Badaruddin Bagu, Lombok Tengah, Nusa Tenggara Barat, Indonesia)]

**Corresponding Author: rokyatulhasanah77@gmail.com, Tel: +62 877-5549-1163*

Received on July 01, 2026, first revision on July 30, 2026, approved on July 31, 2026, published online on July 30, 2026

Abstract: *The development of digital technology has increased the use of social media applications such as TikTok and YouTube, resulting in a large number of user reviews on the Google Play Store. These reviews can be utilized to determine user satisfaction through sentiment analysis. This study aims to analyze the sentiment of user reviews on TikTok and YouTube applications using the Naïve Bayes and Support Vector Machine (SVM) algorithms, as well as to compare the performance of both algorithms. The research data were obtained through a web scraping process consisting of 20,000 reviews, including 10,000 TikTok reviews and 10,000 YouTube reviews. The data then underwent preprocessing, sentiment labeling, splitting into training and testing datasets, and classification using the Naïve Bayes and Support Vector Machine (SVM) algorithms. Model evaluation was conducted using a confusion matrix with accuracy, precision, recall, and F1-score as evaluation metrics. The results showed that the Naïve Bayes algorithm outperformed SVM. For the TikTok application, Naïve Bayes achieved an accuracy of 80.30%, precision of 80.20%, recall of 80.30%, and F1-score of 80.20%, while SVM achieved an accuracy of 78.70%, precision of 78.60%, recall of 78.70%, and F1-score of 78.50%. For the YouTube application, Naïve Bayes achieved an accuracy of 78.40%, precision of 78.00%, recall of 78.40%, and F1-score of 77.90%, while SVM achieved an accuracy of 77.50%, precision of 77.20%, recall of 77.50%, and F1-score of 76.70%. Based on these results, the Naïve Bayes algorithm demonstrated better performance in classifying user review sentiments on TikTok and YouTube applications.*

Keywords: *Sentiment Analysis, Naïve Bayes, Support Vector Machine (SVM).*

1. INTRODUCTION

The rapid advancement of information and communication technology has significantly increased the use of digital media in various aspects of daily life, including communication, information seeking, and entertainment. Among the most widely used video-sharing platforms are TikTok and YouTube. According to the Digital 2025 Indonesia report, YouTube had approximately 143 million users in Indonesia, while TikTok reached approximately 108 million adult users in early 2025 [1]. In addition to serving as entertainment platforms, both platforms are also utilized for education, promotion, and communication, making the quality of the services they provide an important aspect for their users [2].

One indicator that can be used to evaluate the quality of an application is user reviews on the Google Play Store. These reviews contain various opinions, criticisms, and suggestions regarding users' experiences while using the application. However, the large number of reviews in the form of unstructured text data makes the manual analysis process less effective [3]. Therefore, a method capable of processing review data automatically and accurately is needed. One of the most widely used approaches is machine learning-based sentiment analysis, a technique used to identify the tendency of positive and negative opinions expressed in a text [4].

Various studies have applied the Naïve Bayes and Support Vector Machine (SVM) algorithms in sentiment analysis and have demonstrated good performance. Naïve Bayes is known as a simple, fast, and effective probabilistic classification algorithm for processing text data, whereas SVM offers high classification capability by constructing a hyperplane with a maximum margin [5]. Nevertheless, most previous studies have focused on a single application or employed only one classification algorithm. As a result, comparisons of the performance of both algorithms across two digital media platforms remain relatively limited [6][7].

Based on these conditions, this study compares the performance of the Naïve Bayes and Support Vector Machine (SVM) algorithms in analyzing the sentiment of user reviews for the TikTok and YouTube applications obtained from the Google Play Store. This study uses 20,000 reviews, consisting of 10,000 TikTok application reviews and 10,000 YouTube application reviews, thereby providing a more representative dataset. The novelty of this study lies in the comparison of two classification algorithms across two different digital media applications using publicly available, up-to-date, and dynamic review data. The findings are expected to serve as a reference for selecting appropriate sentiment analysis algorithms and to contribute to the development of Natural Language Processing (NLP) methods for evaluating application quality based on user reviews. [8].

2. RESEARCH METHODOLOGY

2.1 Research Object

The object of this study consists of user reviews of the TikTok and YouTube applications obtained from the Google Play Store. The data were collected using a web scraping technique with the google_play_scraper library in Google Colaboratory. The research dataset consists of 20,000 reviews, including 10,000 reviews of the TikTok application and 10,000 reviews of the YouTube application. Subsequently, the data were used as input for the preprocessing stage, sentiment labeling, classification using the Naïve Bayes and Support Vector Machine (SVM) algorithms, and model evaluation.

2.2 Research Stages

This study was conducted through several stages, beginning with data collection, preprocessing, sentiment labeling, dataset splitting, classification using the Naïve Bayes and Support Vector Machine (SVM) algorithms, and ending with model evaluation. The research workflow is illustrated in Figure 2.1.

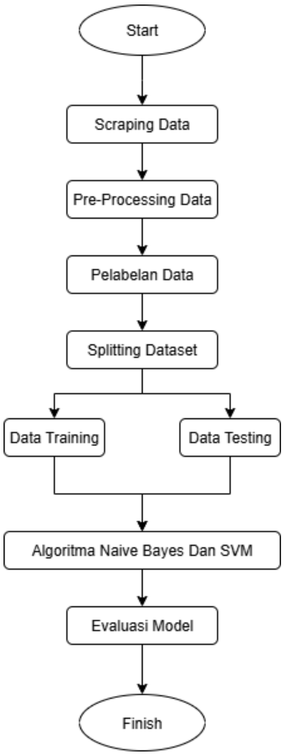

Figure 2.1 Research Stages

Figure 1 illustrates the stages of the research, starting from the data collection process to model evaluation in order to determine the algorithm with the best performance.

2.2.1 Data Scraping

Data were collected using the web scraping method from the Google Play Store through the Google Colaboratory platform by utilizing the google_play_scraper library. The collected data consisted of review texts and user ratings from the TikTok and YouTube applications [9]. A total of 10,000 reviews were collected from each application, resulting in a dataset of 20,000 reviews. The dataset was then stored in CSV format for subsequent analysis.

2.2.2 Preprocessing

At this stage, data preprocessing was performed to transform the data into a more structured form, thereby facilitating the computation process and subsequent stages of the research [10]. The data preprocessing workflow used in this study is illustrated in Figure 2.2.

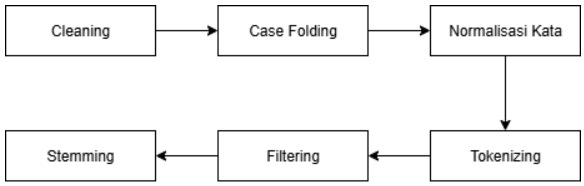

Figure 2.2. Data Preprocessing Stages

a. Cleaning

The cleaning process was performed to remove unnecessary elements from the text, such as numbers, punctuation marks, special characters, emojis, symbols, and excessive whitespace, resulting in cleaner data [11].

b. Case Folding

Case folding converts all letters into lowercase to standardize the writing format and ensure that identical words are not treated as different by the system [12].

c. Normalization

Normalization aims to convert non-standard words, abbreviations, and slang into standard words according to the rules of the Indonesian language, thereby facilitating the analysis process [13].

d. Tokenizing

Tokenizing is the process of splitting a sentence into individual tokens or words, which are subsequently used as the units of analysis in the classification process [14].

e. Stopword Removal / Filtering

Stopword removal is performed to eliminate common words that do not provide important information for the sentiment classification process, allowing the model to focus on words that carry meaningful information [15].

f. Stemming

Stemming aims to convert each word into its root form using the *Sastrawi* library, thereby reducing word variations and making the classification process more effective [16].

2.2.3 Pelabelan Data

After the preprocessing stage, each review was assigned a sentiment label to classify the data into positive and negative categories. In this study, the labeling process was based on the ratings provided by users on the Google Play Store. Reviews with ratings of 4 and 5 were classified as positive sentiment, whereas reviews with ratings of 1 and 2 were classified as negative sentiment. Reviews with a rating of 3 were excluded because they were considered neutral. An example of the data labeling process used in this study is presented in Table 2.1

Table 2.1. Example of the Data Labeling Process

No	Platform	Komentar	Label
1.	TikTok	"Lemot banget pas dibuka, sering keluar sendiri. Tolong diperbaiki!"	Negatif
2.	TikTok	"Fitur editnya keren, gampang dipakai dan hasilnya bagus."	Positif
3.	YouTube	"Terlalu banyak iklan, jadi malas nonton."	Negatif
4.	YouTube	"Kontennya informatif dan bermanfaat, suka sekali channel ini."	Positif

2.2.4 Dataset Splitting

a. Training Data

Training data is the portion of the dataset used to build the classification models using the Naïve Bayes and Support Vector Machine (SVM) algorithms. In this study, 80% of the total dataset was used as training data to learn the relationship patterns between text features and sentiment labels.

b. Testing Data

Testing data is the portion of the dataset used to evaluate the performance of the classification model. In this study, 20% of the total dataset was used as testing data to measure the model's ability to classify sentiment in previously unseen data. The testing results were then used in the model evaluation stage using a confusion matrix [17].

2.3 Naïve Bayes and Support Vector Machine (SVM) Algorithms

At the classification stage, the dataset that had undergone preprocessing and dataset splitting was classified using two algorithms, namely Naïve Bayes and Support Vector Machine (SVM). Both algorithms were applied to the same dataset so that their classification results could be compared based on their performance in identifying positive and negative sentiments.

2.3.1 Naïve Bayes

The Naïve Bayes algorithm was used to classify sentiment based on the probability of word occurrence in each class. The model was built using the training data and then used to predict the sentiment of the testing data based on the highest probability among the classes [18][19].

2.3.2 Support Vector Machine (SVM)

The Support Vector Machine (SVM) algorithm was used to classify sentiment by constructing an optimal hyperplane that separates the positive and negative classes. The model was then used to predict the sentiment class of the testing data based on the position of the data relative to the constructed hyperplane [20].

2.4 Model Evaluation

Model evaluation was performed using a confusion matrix with the accuracy, precision, recall, and F1-score metrics to measure the performance of the Naïve Bayes and Support Vector Machine (SVM) algorithms. In addition, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) were used to evaluate the model's ability to distinguish between positive and negative sentiment classes. The evaluation results were then used as the basis for comparing the performance of the two algorithms [21].

3. RESULTS AND DISCUSSION

This section presents the results of the study, including the stages of data collection, preprocessing, sentiment labeling, and classification using the Naïve Bayes and Support Vector Machine (SVM) algorithms. This study was conducted to identify positive and negative sentiments in user reviews of the TikTok and YouTube applications on the Google Play Store. In addition, the study aimed to implement and evaluate the performance of the Naïve Bayes and Support Vector Machine (SVM) algorithms in the sentiment analysis process. The classification results obtained from both algorithms were subsequently compared to determine the differences in their accuracy in classifying user sentiments. The data used in this study were secondary data derived from user reviews of the applications, which were collected through a web scraping process using the Python programming language.

3.1 Data Scraping Results

The web scraping process successfully collected 20,000 user reviews from the Google Play Store, consisting of 10,000 reviews of the TikTok application and 10,000 reviews of the YouTube application. The collected data included user names, review dates, ratings, and review texts. All data were then stored in CSV format as the research dataset before proceeding to the preprocessing stage. The results of the data collection from the TikTok and YouTube applications are presented in Tables 3.1 and 3.2.

Table 3.1 TikTok Application Data Scraping Results

No	Date	Username	Rating	Raview text
1.	2026-04-27 08:28:13	Pengguna Google	5	bermanfaat sekali
2.	2026-05-02 03:33:50	Pengguna Google	5	suka bgt 😭
3.	2026-04-27 08:31:30	Pengguna Google	1	kenapa tik tok GX bisa di gunakan
4.	2026-04-27 08:31:24	Pengguna Google	1	masalah live terus ada gangguan
5.	2026-04-27 08:30:00	Pengguna Google	1	saya loading tidak bisa masuk.

Tabel 3.2 YouTube Application Data Scraping Results

No	Date	Username	Rating	Raview text
1.	2026-05-02 03:45:16	Pengguna Google	5	Ok semua bagus
2.	2026-05-02 03:37:27	Pengguna Google	1	erorr gk bisa di buka
3.	2026-05-02 03:20:55	Pengguna Google	1	kenapa aplikasi ini tidak dapat di update
4.	2026-05-02 03:32:36	Pengguna Google	2	knpa g bisa bukak
5.	2026-04-27 08:30:00	Pengguna Google	5	menonton vidio para yutuber

Based on Tables 3.1 and 3.2, the collected data exhibit the characteristics of free-text reviews with varying comment lengths. In addition, each review is associated with a rating, which was subsequently used as the basis for sentiment labeling.

3.2 Data Preprocessing Results

After the data collection process was completed, all user reviews were processed through the preprocessing stage to improve the quality of the data before the classification process. This stage aimed to remove unnecessary elements, standardize the text format, and transform the data into a more structured form so that it could be processed by the Naïve Bayes and Support Vector Machine (SVM) algorithms. The preprocessing stages applied in this study included cleaning, case folding, normalization, tokenizing, stopword removal, and stemming. The preprocessing results for the TikTok and YouTube application data are presented in Tables 3.3 and 3.4.

Table 3.3 TikTok Application Data Scraping Results

Tahapan	Hasil
Original	suka bgt 😭
Cleaning	suka bgt
Case Folding	suka bgt
Normalization	suka banget
Tokenizing	[suka, banget]
Stopword Removal	[suka, banget]
Stemming	suka banget

Table 3.4 YouTube Application Data Scraping Results

Tahapan	Hasil
Original	Ok semua bagus
Cleaning	Ok semua bagus
Case Folding	ok semua bagus
Normalization	ok semua bagus
Tokenizing	[ok, semua, bagus]
Stopword Removal	[ok, bagus]
Stemming	ok bagus

Based on Tables 3.3 and 3.4, each preprocessing stage resulted in changes to the text data according to its purpose. The cleaning process successfully removed unnecessary characters, while case folding standardized all letters into lowercase. The normalization stage converted non-standard words into standard words, after which tokenizing split the sentences into individual word tokens. Subsequently, stopword removal eliminated words that did not have a significant impact on the classification process, and stemming converted each word into its root form. These preprocessing results produced cleaner and more structured data, making them ready for the classification process using the Naïve Bayes and Support Vector Machine (SVM) algorithms.

3.3 Data Labeling Results

After the preprocessing stage, all review data were assigned sentiment labels to classify user opinions into positive and negative categories. The labeling process was based on the ratings provided by users on the Google Play Store. Reviews with high ratings were classified as positive sentiment, whereas reviews with low ratings were classified as negative sentiment. The labeling process resulted in a sentiment data distribution that was used in the classification process using the Naïve Bayes and Support Vector Machine (SVM) algorithms. The data labeling results for the TikTok and YouTube applications are presented in Figures 3.1 and 3.2.

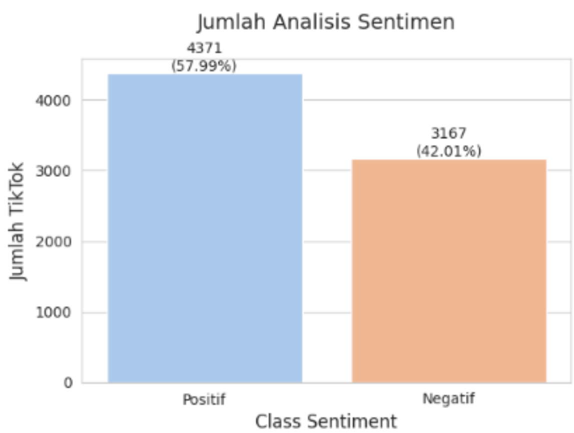

Figure 3.1. Sentiment Labeling Graph of the TikTok Application

Figure 3.1 shows the sentiment labeling results for the TikTok application, indicating that 4,371 reviews (57.99%) were categorized as positive sentiment and 3,167 reviews (42.01%) were categorized as negative sentiment. These results indicate that the majority of users provided positive responses toward the TikTok application based on the reviews collected from the Google Play Store.

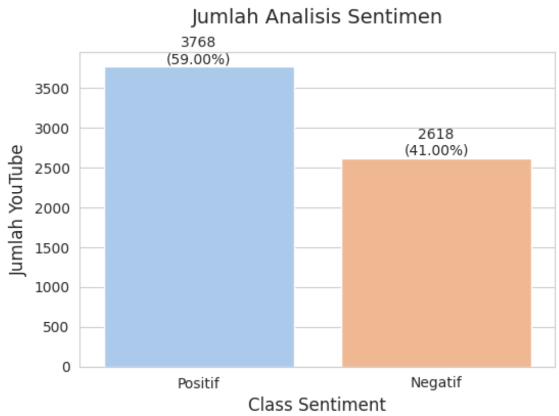

Figure 3.2. Sentiment Labeling Graph of the YouTube Application

Figure 3.2 shows the sentiment labeling results for the YouTube application, indicating that 3,768 reviews (59.00%) were categorized as positive sentiment and 2,618 reviews (41.00%) were categorized as negative sentiment. These percentages indicate that positive reviews also dominated over negative reviews, suggesting that users generally provided good evaluations of the YouTube application.

3.4 Naïve Bayes and SVM Classification Results on TikTok

The classification process on the TikTok application dataset was performed using the Naïve Bayes and Support Vector Machine (SVM) algorithms. Both algorithms were applied to the testing data that had undergone preprocessing and sentiment labeling stages. The classification results were evaluated using a confusion matrix along with the accuracy, precision, recall, and F1-score metrics.

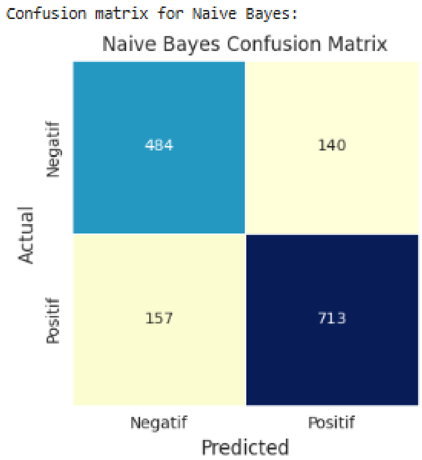

Figure 3.3. Naïve Bayes Confusion Matrix for the TikTok Application

Based on Figure 3.3, the Naïve Bayes algorithm also demonstrated good classification performance on the TikTok application sentiment data. Compared to SVM, the Naïve Bayes classification results produced a higher number of correct predictions, resulting in better classification performance on the research dataset.

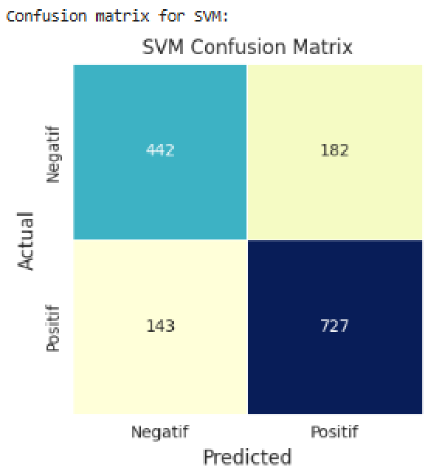

Figure 3.4 Confusion Matrix (SVM) of the TikTok Application

Based on Figure 3.4, the Support Vector Machine (SVM) algorithm was able to classify most of the sentiment data accurately. Although there were still some misclassifications in both positive and negative classes, the confusion matrix results indicate that the model was able to distinguish between the two sentiment classes reasonably well.

Based on the confusion matrix, both algorithms were able to classify positive and negative sentiments effectively. Furthermore, the evaluation results of each algorithm are summarized in Tables 3.5 and 3.6.

Table 3.5. Naïve Bayes Classification Report Results for the TikTok Application

	precision	recall	F1-score	support
Negatif	76,80	75,30	76,00	619.000
Positif	82,70	83,80	83,20	871.000
Accuracy	80,30	80,30	80,30	80.30
Macro avg	79,70	79,50	79,60	1490.000
Weighted avg	80,20	80,30	80,20	1490.000

Based on Table 3.5, the Naïve Bayes algorithm achieved an accuracy of 80.30%, which was higher than the SVM algorithm. The precision, recall, and F1-score values for the positive class reached 82.70%, 83.80%, and 83.20%, respectively, while the macro average and weighted average values were 79.60% and 80.20%, respectively. These results indicate that Naïve Bayes was able to provide better classification performance on the TikTok application dataset.

Table 3.6. SVM Classification Report Results for the TikTok Application

	precision	recall	F1-score	support
Negatif	77,10	69,50	73,10	619.000
Positif	79,70	85,30	82,40	871.000
Accuracy	78,70	78,70	78,70	78.70
Macro avg	78,40	77,40	77,70	1490.000
Weighted avg	78,60	78,70	78,50	1490.000

Based on Table 3.6, the Support Vector Machine (SVM) algorithm achieved an accuracy of 78.70%. The precision, recall, and F1-score values for the positive class were 79.70%, 85.30%, and 82.40%, respectively, indicating that the model performed better in identifying positive sentiment compared to negative sentiment. Overall, the macro average value of 77.70% and the weighted average value of 78.50% indicate that SVM had good classification performance on the TikTok application dataset.

3.5 Naïve Bayes and SVM Classification Results on YouTube

The classification process on the YouTube application dataset was performed using the Support Vector Machine (SVM) and Naïve Bayes algorithms. The classification results were evaluated using a confusion matrix and classification report to determine the ability of each algorithm in classifying positive and negative sentiments.

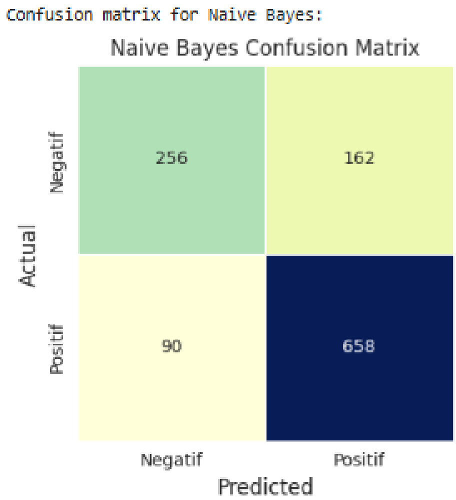

Figure 3.5. Naïve Bayes Confusion Matrix for the YouTube Application

Based on Figure 3.5, the Naïve Bayes algorithm was also able to classify most of the sentiment data effectively. The confusion matrix results show that the model achieved a higher number of correct predictions compared to SVM, resulting in better classification performance on the research dataset.

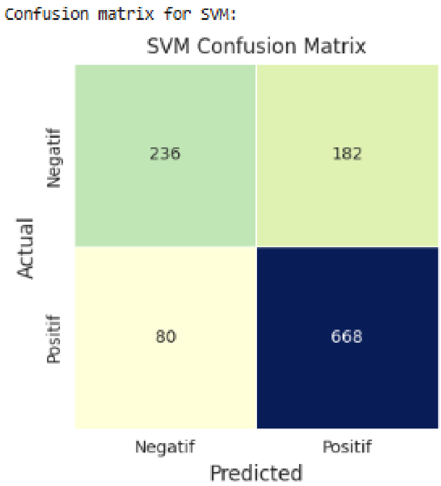

Figure 3.6. SVM Confusion Matrix for the YouTube Application

Based on Figure 3.6, the Support Vector Machine (SVM) algorithm was able to classify most of the sentiment data effectively. Although there were still some misclassifications in both sentiment classes, the model demonstrated a good ability to distinguish between positive and negative reviews.

Table 3.7. Naïve Bayes Classification Report Results for the YouTube Application

	precision	recall	F1-score	support
Negatif	74,00	61,20	67,00	418.000
Positif	80,20	88,00	83,90	748.000
Accuracy	78,40	78,40	78,40	78.40
Macro avg	77,10	74,60	75,50	1166.000
Weighted avg	78,00	78,40	77,90	1166.000

Based on Table 3.7, the Naïve Bayes algorithm achieved an accuracy of 78.40% on the YouTube application dataset. The precision, recall, and F1-score values for the positive class were 80.20%, 88.00%, and 83.90%, respectively, indicating that the model had a good ability to identify reviews with positive sentiment. Meanwhile, the macro average value of 75.50% and the weighted average value of 77.90% indicate that, overall, the model was able to perform sentiment classification effectively on the YouTube application dataset.

Table 3.8. SVM Classification Report Results for the YouTube Application

	precision	recall	F1-score	support
Negatif	74,70	56,50	64,30	418.000
Positif	78,60	89,30	83,60	748.000
Accuracy	77,50	77,50	77,50	77.50
Macro avg	76,60	72,90	74,00	1166.000
Weighted avg	77,20	77,50	76,70	1166.000

Based on Table 3.8, the Support Vector Machine (SVM) algorithm achieved an accuracy of 77.50% on the YouTube application dataset. The precision, recall, and F1-score values for the positive class were 78.60%, 89.30%, and 83.60%, respectively, indicating that the model had a good ability to identify reviews with positive sentiment. Meanwhile, the macro average value of 74.00% and the weighted average value of 76.70% indicate that, overall, the model had fairly good classification performance on the YouTube application dataset.

3.6 Performance Comparison of the Two Algorithms

To determine the algorithm that provides the best performance, a comparison of the classification results of Naïve Bayes and Support Vector Machine (SVM) was conducted based on the accuracy, precision, recall, and F1-score values on the TikTok and YouTube application datasets. The comparison results are presented in Table 3.9.

Table 3.9. Performance Comparison of Naïve Bayes and Support Vector Machine (SVM)

Dataset	Algoritma	Accuracy	Precision	Recall	F1-score
TikTok	Naïve Bayes	80,30 %	83,58 %	81,95 %	82,75 %

TikTok	SVM	78,70 %	79,97 %	83,56 %	81,74 %
YouTube	Naïve Bayes	78,40 %	80,24 %	87,97 %	83,93 %
YouTube	SVM	77,50 %	78,59 %	89,30 %	83,60 %

Based on Table 3.8, both algorithms demonstrated good performance in classifying user review sentiments for the TikTok and YouTube applications. However, Naïve Bayes consistently achieved higher accuracy values compared to Support Vector Machine (SVM) on both datasets. For the TikTok application, Naïve Bayes achieved an accuracy of 80.30%, while SVM achieved 78.70%. Meanwhile, for the YouTube application, Naïve Bayes achieved an accuracy of 78.40%, which was slightly higher than SVM, which achieved 77.50%.

In addition to achieving higher accuracy, the Naïve Bayes algorithm also produced better precision and F1-score values on both datasets. However, the SVM algorithm achieved slightly higher recall values, particularly in identifying reviews with positive sentiment. Overall, the results of this study indicate that Naïve Bayes provided better classification performance and was therefore more suitable for sentiment analysis of TikTok and YouTube application reviews in this study.

3.7 Model Evaluation

Model evaluation was conducted using the Receiver Operating Characteristic (ROC) Curve and the Area Under the Curve (AUC) to measure the ability of the algorithms to distinguish between positive and negative sentiment classes. The closer the AUC value is to 1, the better the model's classification performance.

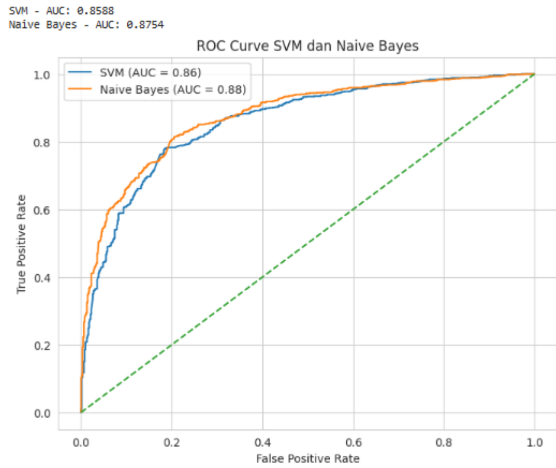

Figure 3.7. ROC Curves of the SVM and Naïve Bayes Models for the TikTok Application

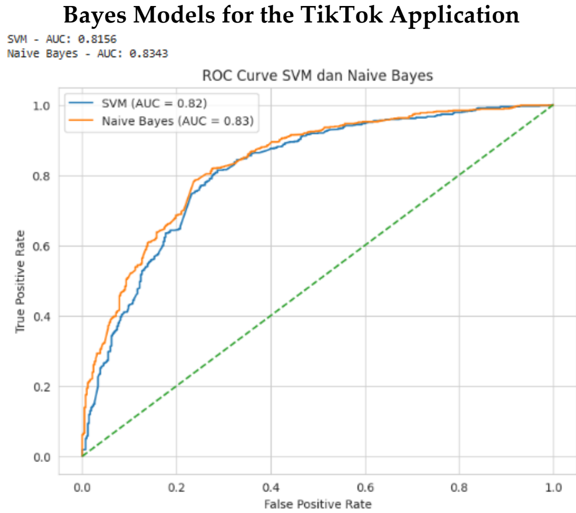

Figure 3.8. ROC Curves of the SVM and Naïve Bayes Models for the YouTube Application

Based on the evaluation results, both algorithms demonstrated good classification performance on both datasets. For the TikTok application, the Naïve Bayes algorithm achieved an AUC value of 0.88, while the Support Vector Machine (SVM) algorithm achieved an AUC value of 0.86. Meanwhile, for the YouTube application, the Naïve Bayes algorithm achieved an AUC value of 0.87, whereas the SVM algorithm achieved an AUC value of 0.85. Since all AUC values were above 0.80, the results indicate that both algorithms have a good ability to distinguish between positive and negative sentiment classes.

Overall, the Naïve Bayes algorithm achieved slightly higher AUC values than the Support Vector Machine (SVM) algorithm on both datasets. This indicates that Naïve Bayes has better discriminative ability in classifying the sentiment of user reviews for the TikTok and YouTube applications.

3.8 Discussion

The results of this study indicate that the Naïve Bayes algorithm performed slightly better than the Support Vector Machine (SVM) algorithm in classifying the sentiment of user reviews for the TikTok and YouTube applications. This is demonstrated by the higher accuracy, precision, and F1-score values achieved on both datasets. These findings indicate that the probabilistic approach employed by Naïve Bayes was able to effectively utilize the word distribution in the review data after the preprocessing stage.

Nevertheless, the Support Vector Machine (SVM) algorithm also demonstrated good performance, with only a relatively small difference in the evaluation results. This indicates that both algorithms are effective for sentiment analysis. The dominance of positive sentiment in the TikTok and YouTube reviews also indicates that most users provided favorable evaluations of both applications, although there were still a number of negative reviews related to application performance and user experience.

Overall, this study demonstrates that Naïve Bayes is a more suitable algorithm for sentiment classification on the research dataset. These findings are expected to serve as a reference for selecting sentiment analysis algorithms while also providing insights for application developers in evaluating service quality based on user reviews.

4. CONCLUSION

Based on the results of this study, the Naïve Bayes and Support Vector Machine (SVM) algorithms were able to classify the sentiment of user reviews for the TikTok and YouTube applications effectively. The analysis results indicate that reviews of both applications were dominated by positive sentiment. Furthermore, based on the model evaluation results, the Naïve Bayes algorithm demonstrated slightly better performance than the Support Vector Machine (SVM) algorithm in terms of accuracy, precision, and F1-score. Therefore, Naïve Bayes can be considered a more suitable alternative for classifying application user review sentiment in this study.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Department of Computer Science, Universitas Qamarul Huda Badaruddin Bagu, as well as to all parties who provided support and assistance throughout the research process, making the successful completion of this article possible

REFERENCES

[1] S. Kemp, “No Title,” *DataReportal*, 2025. [Online]. Available: https://datareportal.com/reports/digital-2025-indonesia.

[2] I. R. Ainunnisa and S. Sulastri, “Analisis Sentimen Aplikasi Tiktok dengan Metode Support Vector Machine (SVM), Logistic Regression dan NaÃ¯ve Bayes,” *J. Teknol. Sist. Inf. dan Apl.*, vol. 6, no. 3, pp. 423–430, 2023.

[3] R. Maulana, A. Voutama, and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC,” *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 42–48, 2023.

[4] F. Sandy *et al.*, “Impelentasi Penggunaan Kecerdasan Buatan Dalam Pendidikan Tinggi,” *Semin. Nas. Teknol. Pendidik. UKI Toraja*, pp. 111–117, 2023.

[5] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, “Analisis Sentimen

Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 375–384, 2024.

[6] M. N. Arifin, A. Hamzah, M. A. Huda, and N. Hasanah, "Analysis of Google Play Store User Sentiment Towards Application X Using the SVM Algorithm," vol. 5, no. 1, pp. 249–258, 2025.

[7] Denia Putri Fajrina, Syafriandi Syafriandi, Nonong Amalita, and Admi Salma, "Sentiment Analysis of TikTok Application on Twitter using The Naïve Bayes Classifier Algorithm," *UNP J. Stat. Data Sci.*, vol. 1, no. 5, pp. 392–398, 2023.

[8] D. Ardiansyah, A. Saepudin, R. Aryanti, E. Fitriani, and Royadi, "Analisis Sentimen Review Pada Aplikasi Media Sosial Tiktok Menggunakan Algoritma K-Nn Dan Svm Berbasis Pso," *J. Inform. Kaputama*, vol. 7, no. 2, pp. 233–241, 2023.

[9] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023.

[10] R. Sanjaya, E. Tohidi, E. Wahyudi, and K. Kaslani, "Analisis Sentimen Terhadap Berhentinya Tiktokshop Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 507–514, 2024.

[11] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, 2021.

[12] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023.

[13] M. Isnan, G. N. Elwirehardja, and B. Pardamean, "Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model," *Procedia Comput. Sci.*, vol. 227, pp. 168–175, 2023.

[14] S. A. R. Rizaldi, S. Alam, and I. Kurniawan, "Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 109–117, 2023.

[15] I. Saputra *et al.*, "Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning," *Fakt. Exacta*, vol. 13, no. 4, p. 200, 2021.

[16] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022.

[17] N. A. Nevrada and M. A. Syaputra, "Sentiment Analysis of Telegram App Reviews on Google Play Store Using the Support Vector Machine (SVM) Algorithm," *J. Appl. Informatics Comput.*, vol. 9, no. 1, pp. 96–105, 2025.

[18] T. S. Rambe, M. N. S. Hasibuan, and M. H. Dar, "Sentiment Analysis of Beauty Product Applications using the Naïve Bayes Method," *SinkrOn*, vol. 8, no. 2, pp. 980–989, 2023.

[19] M. A. Java, Mohammad Syafrullah, W. Windarto, and P. Painem, "Analisis Sentimen Ulasan Pengguna Aplikasi Threads pada Google Play Store Menggunakan Multinomial Naive Bayes dan Support Vector Machine," *J. Ticom Technol. Inf. Commun.*, vol. 12, no. 2, pp. 75–80, 2024.

[20] A. Witanti, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm)," Vol. 5, No. 1, Pp. 59–67, 2022.

[21] A. S. Xiao and Q. Liang, "Spam detection for Youtube video comments using machine learning approaches," *Mach. Learn. with Appl.*, vol. 16, no. April, p. 100550, 2024.