Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 283-294
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

# Performance Analysis of Random Forest and Support Vector Machine in TikTok Shop Review Classification

**Nailah Hafsah[1,*], Avrillia Andhini[1], Shanda Nurhalizah[1], Agus Perdana Widarto[2], Anjar Wanto[2]**

[1] Information Systems Study Program , STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
Jl. Kartini, Proklamasi, Kec. Siantar Barat, Sumatera Utara, 21143, Indonesia
[2] Master of Informatics Study Program, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
Jl. Kartini, Proklamasi, Kec. Siantar Barat, Sumatera Utara, 21143, Indonesia
Email:[1,*] naylahafsah72@gmail.com, [2]avrilliaandhini494@gmail.com, [3]shandanurhalizah99@gmail.com, [4]gusperdana@amiktunasbangsa.ac.id, [5]anjarwanto@gmail.com
Correspondence Author Email: naylahafsah72@gmail.com
Submitted: **01/06/2026**; Accepted: **12/06/2026**; Published: **30/06/2026**

**Abstract**−TikTok Shop generates a large volume of user reviews containing valuable information regarding customer satisfaction, service quality, and user experience. However, manually analyzing large amounts of review data is inefficient and time-consuming, making sentiment classification an important approach for automatically extracting user opinions. This study aims to analyze and compare the performance of Random Forest and Support Vector Machine (SVM) algorithms in classifying TikTok Shop reviews. The dataset was obtained from Kaggle and consisted of 3,145 review records. Data preprocessing was conducted through cleaning, case folding, tokenization, stopword removal, and stemming, resulting in 2,376 valid review records. Feature extraction was performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The dataset was then divided into training and testing sets using an 80:20 ratio before classification using Random Forest and SVM algorithms. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. The results showed that Random Forest achieved an accuracy of 89.71%, precision of 79.50%, recall of 87.00%, and F1-score of 83.00%, while SVM achieved an accuracy of 91.60%, precision of 87.70%, recall of 82.60%, and F1-score of 85.10%. These findings indicate that SVM outperformed Random Forest and provided better overall performance for sentiment classification of TikTok Shop reviews.

**Keywords**: Sentiment Analysis; TikTok Shop; Random Forest; Support Vector Machine; TF-IDF; Text Classification

## 1. INTRODUCTION

The development of digital technology and the internet has significantly driven the growth of e-commerce and social e-commerce in recent years, thereby transforming the way people interact in digital trade activities (Saepudin et al., n.d.). Social commerce is a form of electronic commerce that integrates social media with online transaction activities, enabling users to interact, promote, and purchase products within a single digital platform(Meridha Isnaeni et al., 2026; Rissa Nurfitriana Handayani, 2021). One of the fastest-growing social e-commerce platforms is TikTok Shop, which integrates short-form videos, live streaming, affiliate marketing, and online transaction features into a single digital application(Eva Lasfiana1, 2023; Sopi Nurdini & Dadang Munandar, 2025). The emergence of TikTok Shop has not only transformed companies' digital marketing strategies but has also influenced consumer behavior in seeking product information and making online purchasing decisions(Leo Edward 1, 2025; Tyas Febiani1, 2023).

In the purchase decision-making process, online customer reviews serve as an important source of information that consumers use to evaluate product quality, service quality, and user experiences related to a product or e-commerce platform(Azza Amira et al., 2020; William Chandra, 2024). Customer reviews are also known to influence consumer trust, customer loyalty, and purchase decisions, as reviews reflect the actual experiences of previous users with a product or digital service(Apriana et al., 2025; Nurul et al., 2024; Rahmat Yuliawan, 2025). The high level of user activity in providing opinions and reviews on TikTok Shop generates a large volume of textual data that can be utilized as a valuable source of information for understanding users' perceptions and sentiments toward the products and services available on the social commerce platform(Raden Sigit Surya Pratama, 2025; Rissa Nurfitriana Handayani, 2021).

The increasing number of customer reviews on digital platforms has made manual opinion analysis less effective and more time-consuming(Ambika Hapsari & Dwi Indriyanti, 2023; Chehal et al., 2021). Therefore, sentiment analysis has become an important approach in the fields of text mining and Natural Language Processing (NLP) for automatically identifying positive, negative, and neutral sentiments from unstructured textual data(Arif Bijaksana Putra Negara, 2023; Harnelia, 2024). Sentiment analysis is widely used to understand user opinions on social media, marketplaces, and e-commerce platforms, as it can assist companies in evaluating service quality, customer satisfaction, and consumer perceptions of a product(Ambika Hapsari & Dwi Indriyanti, 2023; Larasati et al., 2022). With the growing volume of customer review data, various machine learning techniques have been widely adopted to conduct automated text classification, offering greater accuracy and efficiency compared to manual analysis(E. Fitri et al., 2020; Rohanah et al., 2021).

Previous research has demonstrated that machine learning algorithms, including Support Vector Machine (SVM), Random Forest, Naïve Bayes, Decision Tree, and Logistic Regression, are widely employed in sentiment classification because of their effectiveness in handling high-dimensional text data(E. Fitri et al., 2020; Rohanah et al., 2021; Saepudin et al., n.d.). In text classification, preprocessing plays a crucial role, as customer review data often contain noise, misspellings, symbols, emojis, abbreviations, and non-standard language that can negatively impact the performance of classification models(Jessica Angelina et al., 2023). Therefore, preprocessing techniques such as cleaning, case folding, tokenization, stopword removal, normalization, and stemming are essential for generating a more structured representation of textual data prior to the classification stage(Azza Amira et al., 2020). In addition to preprocessing, feature extraction methods such as Term Frequency–Inverse Document Frequency (TF-IDF) are also widely used in sentiment analysis research because they can assign weights to important words, thereby improving the quality of feature representation in textual data(Y. Setiawan et al., 2023).

Various machine learning algorithms have been employed in sentiment classification research to analyze user opinions on social media and e-commerce platforms(E. Fitri et al., 2020; Larasati et al., 2022). Algorithms such as Naïve Bayes, Decision Tree, Logistic Regression, Support Vector Machine (SVM), and Random Forest are widely applied because they can perform automated text classification with relatively high accuracy on high-dimensional data(Homepage et al., 2026; Oktavianto et al., 2024; Rohanah et al., 2021). Previous studies have shown that Support Vector Machine (SVM) is frequently used in sentiment analysis because it has strong capabilities in handling sparse matrices and text data generated through TF-IDF feature extraction(Cristina Angelika & Suhari, 2026; Fathirachman Mahing et al., 2024). Meanwhile, Random Forest is also widely used in text classification because it can reduce overfitting through ensemble learning and bagging approaches applied to decision trees (R. R. Fitri et al., 2025). Furthermore, several previous studies have demonstrated that the combination of text preprocessing, feature extraction, and machine learning algorithms can improve the performance of sentiment classification on social media and digital marketplace data(Aldean et al., 2022; Khab Sulaiman Dalam et al., n.d.).

Although various machine learning algorithms have been widely used in sentiment analysis research, each method still exhibits different performance characteristics in the text classification process (Alqaraleh et al., 2024; Dai et al., 2024). Support Vector Machine is known for its strong ability to handle high-dimensional data and sparse matrices, which often results in high classification performance in text-based sentiment analysis studies(Das et al., 2021; J. C. Setiawan et al., 2023). Furthermore, SVM is considered effective in the classification process because it can construct an optimal hyperplane to separate data classes more accurately (Madjid et al., 2023; Pakpahan et al., 2023). Other studies have shown that the use of appropriate kernels and parameters can improve the classification performance of Support Vector Machine on unstructured text data(Ahmed Khan et al., 2024; Başarslan & Bal, 2025). On the other hand, Random Forest has the advantage of reducing overfitting through ensemble learning and bagging approaches applied to decision trees(Dwi Utami Putra, 2025; Rahmawati et al., 2024).

Random Forest is also known to provide stable classification performance in text classification and sentiment analysis studies involving social media and digital marketplace data(Arifin et al., 2021; Aufa et al., 2023). Furthermore, Random Forest is capable of aggregating multiple decision trees, thereby improving classification accuracy on complex and diverse datasets (Megawan et al., 2025; Setyabudi & Aryanny, 2025). However, Random Forest models tend to be more difficult to interpret and require greater computational resources compared to other simpler classification methods(Lestari et al., 2025; Zainottah et al., 2025). In addition to algorithmic factors, text preprocessing, dataset balancing, and feature extraction techniques such as TF-IDF are also known to significantly influence the performance of sentiment classification in various machine learning studies(Armaeni et al., 2024; Maharani & Yell, 2025; Muawanah et al., 2023).

Previous studies on sentiment analysis in digital platforms have generally focused on the use of a single classification algorithm without conducting a comprehensive comparative analysis of the performance of multiple machine learning models(Alqaraleh et al., 2024; Das et al., 2021). Most previous studies have also primarily utilized review data from Twitter, YouTube, Shopee, Tokopedia, and other general marketplaces rather than Indonesian-language data from TikTok Shop(Dai et al., 2024; Rahmawati et al., 2024). Studies that specifically compare the performance of Support Vector Machine and Random Forest in the classification of TikTok Shop reviews remain relatively limited, making it unclear which algorithm performs best for the characteristics of social commerce data(Nur Khaliza et al., 2022; Setyabudi & Aryanny, 2025). Furthermore, differences in text preprocessing, feature extraction techniques, and the amount of training data used have led to varying model accuracy results across previous studies(Lestari et al., 2025; Zainottah et al., 2025).

The characteristics of textual data on social commerce platforms also tend to be more dynamic, as they contain informal language, abbreviations, emojis, typographical errors, and non-standard language that can affect the sentiment classification process(Başarslan & Bal, 2025; J. C. Setiawan et al., 2023). Several previous studies have also exhibited dataset imbalance and disproportionate sentiment class distributions, which may affect the stability of classification model performance(Arifin et al., 2021; Megawan et al., 2025). Furthermore, some studies have focused solely on accuracy values without considering precision, recall, and F1-score as evaluation metrics for a more comprehensive assessment of classification model performance(Indrayuni & Acmad Nurhadi, 2025; Pakpahan et al., 2023). Therefore, further research is needed to analyze the performance of classification algorithms that can provide optimal results on TikTok Shop review data, which are characterized by dynamic and unstructured textual content(Aufa et al., 2023; Dwi Utami Putra, 2025).

Based on these issues, this study proposes a performance analysis of the Random Forest and Support Vector Machine algorithms for the classification of TikTok Shop reviews using a Natural Language Processing approach and TF-IDF feature extraction(Ebrahim et al., 2023; Gasparetto et al., 2022). This study applies several preprocessing steps, including cleaning, case folding, tokenization, stopword removal, normalization, and stemming, to improve the quality of textual data before the classification process is performed(Ahmad Farhan AlShammari, 2023; Kalaivani &

Kuppuswami, 2019). In addition, the use of TF-IDF is known to provide a more optimal representation of word weights, thereby improving feature quality in machine learning-based text classification processes(Mohammad Alkharboush, 2021; Sri, 2025). This study also compares the performance of Support Vector Machine and Random Forest using evaluation metrics such as accuracy, precision, recall, and F1-score to identify the best classification model for TikTok Shop review data(Bellar et al., 2024; Palanivinayagam & Damaševičius, 2023).

The use of these evaluation metrics is considered capable of providing a more comprehensive assessment of model performance than relying solely on accuracy(Chinnalagu & Durairaj, 2021; Qasim et al., 2022). In addition, TikTok Shop review data exhibit unstructured textual characteristics, such as typographical errors, emojis, abbreviations, and informal language, which require optimal preprocessing before sentiment classification can be performed(Atia Nevrada & Syaputra, 2025; Puh & Bagić Babac, 2023). Previous studies have also shown that the quality of preprocessing and feature extraction has a significant influence on the performance of machine learning-based sentiment classification models(Deski Manalu et al., 2025; Saraswati & Diatri Indradewi, 2022). Therefore, this study is expected to contribute to the development of sentiment analysis methods on social commerce platforms and serve as a reference for future research on sentiment classification using digital review data(Alshehri & Algarni, 2023; Zhou et al., 2024).

# 2. RESEARCH METHODS

## 2.1 Dataset Collection and Preprocessing

The dataset used in this study was obtained from Kaggle under the title “TikTok Tokopedia Seller Center Reviews”. The dataset contains user reviews related to experiences, complaints, and satisfaction with TikTok Shop services and features. A total of 3,145 review records were collected and categorized into positive and negative sentiment classes. These reviews were used as the primary data source for sentiment classification using Random Forest and Support Vector Machine algorithms. Examples of the original review dataset used in this study are presented in Table 1.

**Table 1**. Examples of TikTok Shop Review Dataset

| No | Review Content | Sentiment |
|---|---|---|
| 1 | tolong keuangan transfer rekening berbeda rekening daftarkan buka toko orang rekening ter isi rekening tolong pendaftaran toko batasi akun daftarkan hilang lupa sandi kasih minim kali | Negative |
| 2 | sumpah ya aplikasi seller buruk tuh duitnya langsung masuk saldo nunggu berhari-hari butuh modal belanja pakai kecewa aduh payah | Negative |
| 3 | susah ribet | Negative |
| 4 | capai bikin dekripsi tolak kacau | Negative |
| 5 | pelanggaran upload barang | Negative |
| ... | ... | ... |
| 3145 | kendala masuk tiktikshop seller verifikasi dokumen dokumen terdaftar akun uang ketarik tolong solusinya pengambilan dana | Negative |

As shown in Table 1 several examples of the original review dataset used in this study. The collected reviews contain various writing styles, inconsistent word forms, and textual noise that may affect the performance of sentiment classification. Therefore, a preprocessing stage was conducted to improve the quality of the textual data before feature extraction and classification. The preprocessing procedures included cleaning, case folding, tokenization, stopword removal, and stemming. Cleaning was performed to remove unnecessary characters, symbols, punctuation marks, and irrelevant text elements from the review data. Case folding was applied to convert all characters into lowercase letters to ensure consistency in text representation. Tokenization was then used to split review sentences into individual words or tokens. Stopword removal was conducted to eliminate commonly used words that do not contribute significantly to sentiment classification. Finally, stemming was applied to transform words into their root forms, reducing variations in word representation. These processes were applied to transform the review texts into a cleaner and more structured format suitable for machine learning analysis. Examples of the preprocessing results are presented in Table 2.

**Table 2.** Examples of Text Preprocessing Results

| No | Original Review | Cleaning | Tokenizing | Stopword Removal | Stemming |
|---|---|---|---|---|---|
| 1 | tolong keuangan transfer rekening berbeda rekening daftarkan buka toko orang rekening ter isi rekening tolong pendaftaran toko | tolong keuangan transfer rekening berbeda rekening daftarkan buka toko orang rekening ter isi rekening tolong pendaftaran toko | ['tolong', 'keuangan', 'transfer', 'rekening', 'berbeda', 'rekening', 'daftarkan', 'buka', 'toko', 'orang', 'rekening', 'ter', 'isi', 'rekening', 'tolong', 'pendaftaran', 'toko', | ['tolong', 'keuangan', 'transfer', 'rekening', 'berbeda', 'rekening', 'daftarkan', 'buka', 'toko', 'orang', 'rekening', 'ter', 'isi', 'rekening', 'tolong', 'pendaftaran', 'toko', | tolong keuangan transfer rekening berbeda rekening daftarkan buka toko orang rekening ter isi rekening tolong pendaftaran toko |

| No | Original Review | Cleaning | Tokenizing | Stopword Removal | Stemming |
|---|---|---|---|---|---|
| | batasi akun daftarkan hilang lupa sandi kasih minim kali | batasi akun daftarkan hilang lupa sandi kasih minim kali | 'batasi', 'akun', 'daftarkan', 'hilang', 'lupa', 'sandi', 'kasih', 'minim', 'kali'] | 'batasi', 'akun', 'daftarkan', 'hilang', 'lupa', 'sandi', 'kasih', 'minim', 'kali'] | batasi akun daftarkan hilang lupa sandi kasih minim kali |
| 2 | sumpah ya aplikasi seller buruk tuh duitnya langsung masuk saldo nunggu berhari butuh modal belanja pakai kecewa aduh payah | sumpah ya aplikasi seller buruk tuh duitnya langsung masuk saldo nunggu berhari butuh modal belanja pakai kecewa aduh payah | ['sumpah', 'ya', 'aplikasi', 'seller', 'buruk', 'tuh', 'duitnya', 'langsung', 'masuk', 'saldo', 'nunggu', 'berhari', 'butuh', 'modal', 'belanja', 'pakai', 'kecewa', 'aduh', 'payah'] | ['sumpah', 'ya', 'aplikasi', 'seller', 'buruk', 'tuh', 'duitnya', 'langsung', 'masuk', 'saldo', 'nunggu', 'berhari', 'butuh', 'modal', 'belanja', 'pakai', 'kecewa', 'aduh', 'payah'] | sumpah ya aplikasi seller buruk tuh duitnya langsung masuk saldo nunggu berhari butuh modal belanja pakai kecewa aduh payah |
| 3 | susah ribet | susah ribet | ['susah', 'ribet'] | ['susah', 'ribet'] | susah ribet |
| 4 | capai bikin dekripsi tolak kacau | capai bikin dekripsi tolak kacau | ['capai', 'bikin', 'dekripsi', 'tolak', 'kacau'] | ['capai', 'bikin', 'dekripsi', 'tolak', 'kacau'] | capai bikin dekripsi tolak kacau |
| 5 | pelanggaran upload barang | pelanggaran upload barang | ['pelanggaran', 'upload', 'barang'] | ['pelanggaran', 'upload', 'barang'] | pelanggaran upload barang |
| … | … | … | … | … | … |
| 2376 | kendala masuk tiktikshop seller verifikasi dokumen dokumen terdaftar akun uang ketarik tolong solusinya pengambilan dana | kendala masuk tiktikshop seller verifikasi dokumen dokumen terdaftar akun uang ketarik tolong solusinya pengambilan dana | ['kendala', 'masuk', 'tiktikshop', 'seller', 'verifikasi', 'dokumen', 'dokumen', 'terdaftar', 'akun', 'uang', 'ketarik', 'tolong', 'solusinya', 'pengambilan', 'dana'] | ['kendala', 'masuk', 'tiktikshop', 'seller', 'verifikasi', 'dokumen', 'dokumen', 'terdaftar', 'akun', 'uang', 'ketarik', 'tolong', 'solusinya', 'pengambilan', 'dana'] | kendala masuk tiktikshop seller verifikasi dokumen dokumen terdaftar akun uang ketarik tolong solusinya pengambilan dana |

Based on the preprocessing results shown in Table 2, the review texts became cleaner and more structured through the application of cleaning, case folding, tokenization, stopword removal, and stemming. After preprocessing, a total of 2,376 review records were retained and used for feature extraction and sentiment classification. The preprocessed data were then transformed into numerical representations using the TF-IDF method before being classified using the Random Forest and Support Vector Machine algorithms. To ensure the quality and consistency of the dataset used in this study, a duplicate removal process was conducted before the feature extraction stage. This process aimed to eliminate redundant review records that could potentially affect the performance and reliability of the classification models.

a. Duplicate Removal

Before the preprocessing stage, duplicate reviews were identified and removed to improve data quality and avoid redundancy in the classification process. The original dataset contained 3,145 review records, of which 769 records were identified as duplicates. After the duplicate removal process, 2,376 unique review records remained and were used for subsequent preprocessing, feature extraction, and classification stages.

**Table 3.** Dataset Summary Afer Duplicate Removel

| Description | Records |
|---|---|
| Original Dataset | 3,145 |
| Duplicate Reviews Removed | 769 |
| Final Dataset | 2,376 |

As shown in Table 3, the original dataset consisted of 3,145 review records. After the duplicate removal process, 769 duplicated reviews were eliminated, resulting in a final dataset of 2,376 unique review records. The resulting dataset was subsequently used for TF-IDF feature extraction and sentiment classification using Random Forest and Support Vector Machine algorithms.

b. Feature Extraction Using TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) is a feature extraction method used to transform textual data into numerical representations. TF-IDF assigns weights to terms based on their importance within a document and across the dataset. Several mathematical formulations of TF-IDF are presented in Equations (1) - (3).

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 283-294
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

$$TF(t,d) = \frac{f(t,d)}{\sum_k f(k,d)} \tag{1}$$

$$(IDF(t) = \log \frac{N}{DF(t)} \tag{2}$$

$$TFIDF(t,d) = TF(t,d) \times IDF(t) \tag{3}$$

Equation (1) calculates the term frequency of a word within a document. Equation (2) computes the inverse document frequency to measure the uniqueness of a term across all documents. Equation (3) combines TF and IDF to generate the final TF-IDF weight used as input features for Random Forest and Support Vector Machine classification.

c. Train-Test Split

After the TF-IDF transformation process, the dataset was divided into training and testing sets using the train-test split technique. The data were partitioned using an 80:20 ratio, where 80% of the data were used for training and 20% were used for testing. This partitioning was applied before the classification process using Random Forest and Support Vector Machine algorithms.

**Table 4.** Dataset Partition Using Train-Test Split

| Dataset | Number of Records | Percentage |
|---|---|---|
| Training Data | 1900 | 80% |
| Testing Data | 476 | 20% |
| Total | 2376 | 100% |

As shown in Table 4, the dataset was divided into 1,900 training records and 476 testing records. The training data were used to build the classification models, while the testing data were utilized to evaluate and compare the performance of Random Forest and Support Vector Machine models.

### 2.2 Research Framework

This study employed a research framework to systematically analyze and compare the performance of Random Forest and Support Vector Machine algorithms in classifying TikTok Shop reviews. The framework consists of several stages, including dataset collection, duplicate removal, text preprocessing, TF-IDF feature extraction, train-test split, classification using Random Forest and Support Vector Machine, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The overall research process is illustrated in Figure 1.

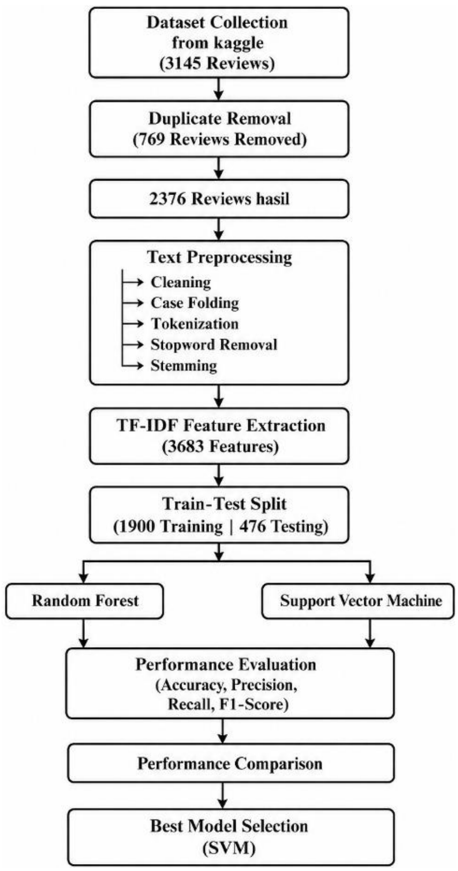


**Figure 1.** Research Framework of the Proposed Study

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 283-294
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

As shown in Figure 1, the research process began with collecting the TikTok Tokopedia Seller Center Reviews dataset consisting of 3,145 review records. Duplicate reviews were removed to improve data quality and eliminate redundant information, resulting in 2,376 review records used for further analysis. The data were subsequently preprocessed through cleaning, case folding, tokenization, stopword removal, and stemming. After preprocessing, TF-IDF was applied to transform textual data into numerical features, generating 3,683 features. The resulting features were divided into training and testing datasets using an 80:20 ratio, consisting of 1,900 training records and 476 testing records. The classification process was then performed using Random Forest and Support Vector Machine algorithms. Finally, both models were evaluated using accuracy, precision, recall, and F1-score metrics, followed by a performance comparison to determine the best-performing model.

### 2.3 Random Forest

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve classification performance and reduce overfitting. The algorithm constructs several decision trees using different subsets of training data and determines the final prediction through a majority voting mechanism. Several mathematical formulations used in the Random Forest algorithm are presented in Equations (4) - (9).

$$D_b \subset D \tag{4}$$

$$Gini(D) = 1 - \sum_{i=1}^{c} p_i^2 \tag{5}$$

$$Gini_{split}(D) = \sum_{j=1}^{n} \frac{|D_j|}{|D|} Gini(D_j) \tag{6}$$

$$Gain(D, A) = Gini(D) - Gini_{split}(D) \tag{7}$$

$$\hat{y} = \mathrm{mode}(h_1(x), h_2(x), \dots, h_n(x)) \tag{8}$$

$$FI_j = \sum_{t=1}^{T} \Delta I(t, j) \tag{9}$$

Equation (4) represents the bootstrap sampling process used to generate different training subsets for each decision tree. Equations (5) and (6) describe the Gini Index and Gini Split, which are used to measure node impurity and determine the best data partition. Equation (7) calculates Information Gain to evaluate the contribution of each feature during tree construction. Equation (8) represents the majority voting mechanism used to produce the final classification result from multiple decision trees. Finally, Equation (9) measures feature importance, which indicates the contribution of individual features to the sentiment classification process.

### 2.4 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm widely used for text classification and sentiment analysis. The main objective of SVM is to find an optimal hyperplane that maximizes the margin between different classes. By maximizing the separation boundary, SVM can achieve high classification performance, particularly when handling high-dimensional data generated through TF-IDF feature extraction. Several mathematical formulations used in the SVM algorithm are presented in Equations (10) - (16).

$$f(x) = w^T x + b \tag{10}$$

$$w^T x + b = 0 \tag{11}$$

$$w^T x + b \geq 1 \tag{12}$$

$$w^T x + b \leq - \tag{13}$$

$$Margin = \frac{2}{\|w\|} \tag{14}$$

$$\min \frac{1}{2} \| w \|^2 \tag{15}$$

$$K(x_i, x_j) = x_i^T x_j \tag{16}$$

Equation (10) represents the SVM decision function used to classify input data. Equations (11), (12), and (13) define the hyperplane and class boundaries used to separate different sentiment classes. Equation (14) calculates the margin between classes, where a larger margin generally leads to better generalization performance. Equation (15) represents the optimization objective used to determine the optimal hyperplane by minimizing the weight vector. Equation (16) shows the linear kernel function used to calculate the similarity between feature vectors in the transformed feature space.

### 2.5 Performance Evaluation

Performance evaluation was conducted to assess and compare the classification performance of Random Forest and Support Vector Machine models. The evaluation process was based on the confusion matrix and measured using

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 283-294
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

accuracy, precision, recall, and F1-score metrics. Several evaluation metrics used in this study are presented in Equations (17) - (20).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{17}$$

$$Precision = \frac{TP}{TP+FP} \tag{18}$$

$$Recall = \frac{TP}{TP+FN} \tag{19}$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{20}$$

Equation (17) measures the overall classification accuracy of the model. Equation (18) evaluates the proportion of correctly predicted positive instances among all predicted positives. Equation (19) measures the ability of the model to identify actual positive instances, while Equation (20) combines precision and recall into a single metric to provide a balanced evaluation of classification performance.

# 3. RESULTS AND DISCUSSION

## 3.1 Random Forest Classification Results

The Random Forest model was trained using 1,900 training records and evaluated using 476 testing records. The model performance was assessed using a confusion matrix and classification metrics, including accuracy, precision, recall, and F1-score. The confusion matrix of the Random Forest model is presented in Figure 2.

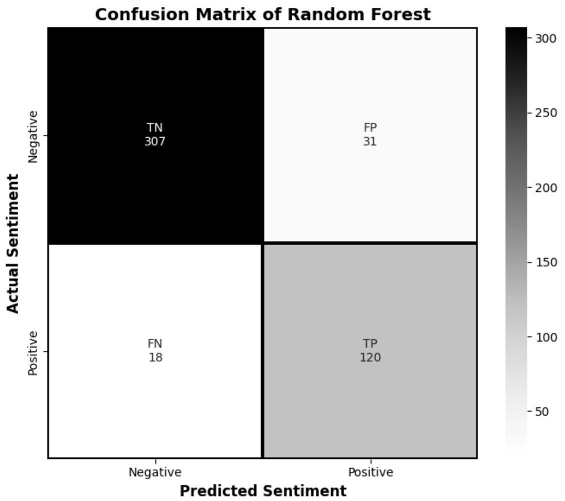


**Figure 2**. Confusion Matrix of Random Forest

As shown in Figure 2, the Random Forest model correctly classified 307 negative reviews and 120 positive reviews. However, 31 negative reviews were incorrectly classified as positive, while 18 positive reviews were incorrectly classified as negative. The results indicate that the Random Forest model was able to distinguish between positive and negative sentiments effectively, as evidenced by the relatively high number of correct predictions compared to misclassified instances. Overall, the confusion matrix demonstrates that Random Forest achieved good performance in classifying TikTok Shop reviews.

**Table 5.** Performance of Random Forest Model

| Metric | Value |
|---|---|
| Accuracy | 89.71% |
| Precision | 79.50% |
| Metric | Value |
| Recall | 87.00% |
| F1-Score | 83.00% |

As shown in Table 5, the performance evaluation results of the Random Forest model. The model achieved an accuracy of 89.71%, indicating that the majority of review sentiments were classified correctly. Furthermore, the model obtained a precision of 79.50%, recall of 87.00%, and F1-score of 83.00%. The relatively high recall value indicates that Random Forest was effective in identifying sentiment instances, while the overall results demonstrate that the model provided good classification performance for TikTok Shop review data.

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 283-294
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

### 3.2 Support Vector Machine Classification Results

To provide a fair comparison, the Support Vector Machine model was trained and evaluated using the same training and testing datasets as the Random Forest model. The classification performance of the Support Vector Machine model is presented through the confusion matrix shown in Figure 3.

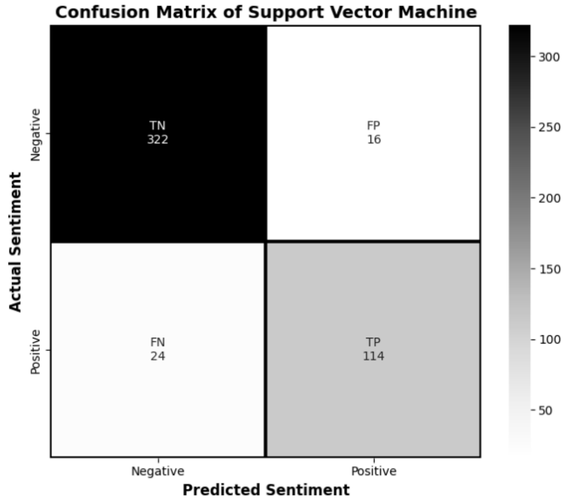


**Figure 3.** Confusion Matrix of Support Vector Machine

As shown in Figure 3, the Support Vector Machine model correctly classified 322 negative reviews and 114 positive reviews. However, 16 negative reviews were incorrectly classified as positive, while 24 positive reviews were incorrectly classified as negative. Compared to the Random Forest model, SVM produced fewer false positive predictions, indicating a better ability to distinguish negative sentiment reviews. Overall, the confusion matrix demonstrates that the Support Vector Machine model achieved strong classification performance on the TikTok Shop review dataset.

**Table 6.** Performance of Support Vector Machine Model

| Metric | Value |
|---|---|
| Accuracy | 91.60% |
| Precision | 87.70% |
| Recall | 82.60% |
| F1-Score | 85.10% |

As shown in Table 6, the performance evaluation results of the Support Vector Machine model. The model achieved an accuracy of 91.60%, which was higher than the accuracy obtained by Random Forest. In addition, SVM achieved a precision of 87.70%, recall of 82.60%, and F1-score of 85.10%. These results indicate that Support Vector Machine provided strong and balanced classification performance, making it effective for sentiment classification of TikTok Shop reviews represented using TF-IDF features.

### 3.3 Performance Comparison of Random Forest and Support Vector Machine

To compare the performance of Random Forest and Support Vector Machine, several evaluation metrics were analyzed, including accuracy, precision, recall, and F1-score. The comparison results are presented in Figure 4.

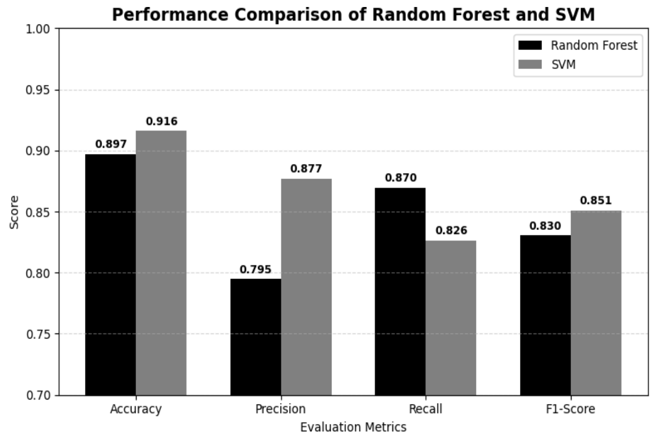


**Figure 4**. Performance Comparison of Random Forest and Support Vector Machine

Vol 1, No 1, June 2026, Page 283-294
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

Figure 4 shows that Support Vector Machine achieved higher accuracy, precision, and F1-score values than Random Forest. SVM obtained an accuracy of 91.60%, precision of 87.70%, and F1-score of 85.10%, while Random Forest achieved 89.71%, 79.50%, and 83.00%, respectively. However, Random Forest produced a higher recall value of 87.00% compared to 82.60% achieved by SVM. These results indicate that SVM generally provided better classification performance for the TikTok Shop review dataset.

**Table 7.** Performance Comparison of Random Forest and Support Vector Machine

| Metric | Random Forest | SVM |
|---|---|---|
| Accuracy | 89.71% | 91.60% |
| Precision | 79.50% | 87.70% |
| Recall | 87.00% | 82.60% |
| F1-Score | 83.00% | 85.10% |

As shown in Table 7, Support Vector Machine outperformed Random Forest in most evaluation metrics. The higher accuracy and F1-score achieved by SVM indicate a better balance between precision and recall, suggesting that SVM was more effective for sentiment classification of TikTok Shop reviews in this study.

### 3.4 Discussion of Random Forest and Support Vector Machine

Figure 5 shows that the Support Vector Machine model achieved a higher accuracy score than the Random Forest model. SVM obtained an accuracy of 91.60%, while Random Forest achieved an accuracy of 89.71%. The difference indicates that SVM was more effective in separating positive and negative sentiment classes within the TikTok Shop review dataset. This superior performance can be attributed to the ability of SVM to handle high-dimensional feature spaces generated by TF-IDF representation. Consequently, SVM demonstrated better overall classification capability and can be considered the most suitable algorithm for sentiment classification in this study.

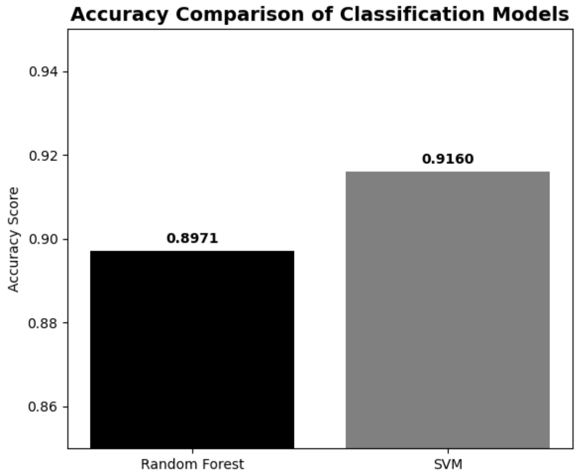


**Figure 5.** Accuracy Comparison of Classification Models

Figure 5 This study analyzed and compared the performance of Random Forest and Support Vector Machine algorithms for sentiment classification of TikTok Shop reviews using TF-IDF feature extraction. A total of 2,376 review records were obtained after duplicate removal and preprocessing, and subsequently transformed into 3,683 textual features using TF-IDF. The experimental results showed that both algorithms achieved good classification performance. Random Forest obtained an accuracy of 89.71%, precision of 79.50%, recall of 87.00%, and F1-score of 83.00%. Meanwhile, Support Vector Machine achieved an accuracy of 91.60%, precision of 87.70%, recall of 82.60%, and F1-score of 85.10%. Overall, Support Vector Machine outperformed Random Forest in terms of accuracy, precision, and F1-score, indicating its superior capability in classifying TikTok Shop review sentiments. Therefore, Support Vector Machine can be considered the most suitable classification model for the dataset used in this study.

## 4. CONCLUSION

This study analyzed and compared the performance of Random Forest and Support Vector Machine (SVM) algorithms for sentiment classification of TikTok Shop reviews using TF-IDF feature extraction. The dataset consisted of 3,145 review records obtained from Kaggle. After the duplicate removal process, 2,376 unique review records remained and were used for preprocessing, feature extraction, and classification. The preprocessed reviews were transformed into 3,683 numerical features using the TF-IDF method before the classification process was performed. The experimental results demonstrated that both Random Forest and SVM achieved good classification performance. Random Forest obtained an accuracy of 89.71%, precision of 79.50%, recall of 87.00%, and F1-score of 83.00%. Meanwhile, SVM

Vol 1, No 1, June 2026, Page 283-294
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

achieved an accuracy of 91.60%, precision of 87.70%, recall of 82.60%, and F1-score of 85.10%. The comparison results showed that SVM achieved higher accuracy, precision, and F1-score values than Random Forest. Although Random Forest produced a slightly higher recall value, SVM provided better overall classification performance. The ability of SVM to handle high-dimensional TF-IDF feature representations contributed to its superior performance in classifying TikTok Shop review sentiments. Therefore, Support Vector Machine can be considered the most suitable classification algorithm for the dataset used in this study. The findings of this study may serve as a reference for future sentiment analysis research and assist in selecting appropriate machine learning models for social commerce review classification. Future studies may explore larger datasets, additional feature extraction techniques, and advanced machine learning or deep learning approaches to further improve classification performance.

## REFERENCES

Ahmad Farhan Alshammari. (2023). Implementation Of Keyword Extraction Using Term Frequency-Inverse Document Frequency (Tf-Idf) In Python. In *International Journal Of Computer Applications* (Vol. 185, Number 35).

Ahmed Khan, T., Sadiq, R., Shahid, Z., Mansoor Alam, M., Bin Mohd Su, M., Rd, S., & Faisal Cantonment, K. (2024). Sentiment Analysis Using Support Vector Machine And Random Forest. *Journal Of Informatics And Web Engineering*, *3*(Sentiment Analysis Using Support Vector Machine And Random Forest), 1–9. Https://Doi.Org/10.33093/Jiwe.2023.3.1.5

Aldean, M. Y., Paradise, & Setya Nugraha, N. A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac). *Informatics, Information System, Software Engineering And Application*, *4*(Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Di Twitter Menggunakan Metode Random Forest Classifier), 1–9.

Alqaraleh, M., Alzboon, M. S., Al-Batah, M. S., Wahed, M. A., Abuashour, A., & Alsmadi, F. H. (2024). Harnessing Machine Learning For Quantifying Vesicoureteral Reflux: A Promising Approach For Objective Assessment. *International Journal Of Online And Biomedical Engineering*, *20*(11), 1–23. Https://Doi.Org/10.3991/Ijoe.V20i11.49673

Alshehri, A., & Algarni, A. (2023). Tf-Tda: A Novel Supervised Term Weighting Scheme For Sentiment Analysis. *Electronics (Switzerland)*, *12*(7), 1–25. Https://Doi.Org/10.3390/Electronics12071632

Ambika Hapsari, N., & Dwi Indriyanti, A. (2023). *Analisis Sentimen Pada Aplikasi Dompet Digital Menggunakan Algoritma Random Forest*.

Apriana, A., Suhud, U., & Bara Berutu, M. (2025). Peran Gaya Hidup Berbelanja, Influencer Marketing, Ulasan Online Dalam Keputusan Pembelian Dan Loyalitas Gen Z Di Marketplace Tiktok Shop. *Jurnal Ilmiah Multidisiplin*, *4*(9), 1–10.

Arif Bijaksana Putra Negara. (2023). The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, *14*(3), 1–4. Https://Doi.Org/10.24843/Lkjiti.2023.V14.I03.P05

Arifin, N., Enri, U., & Sulistiyowati, N. (2021). *String (Satuan Tulisan Riset Dan Inovasi Teknologi) Penerapan Algoritma Support Vector Machine (Svm) Dengan Tf-Idf N-Gram Untuk Text Classification*.

Armaeni, P. P., Wiguna, I. K. A. G., & Parwita, W. G. S. (2024). Sentiment Analysis Of Youtube Comments On The Closure Of Tiktok Shop Using Naïve Bayes And Decision Tree Method Comparison. *Jurnal Galaksi*, *1*(2), 1–11. Https://Doi.Org/10.70103/Galaksi.V1i2.15

Atia Nevrada, N., & Syaputra, M. A. (2025). Sentiment Analysis Of Telegram App Reviews On Google Play Store Using The Support Vector Machine (Svm) Algorithm. In *Journal Of Applied Informatics And Computing (Jaic)* (Vol. 9, Number 1). Http://Jurnal.Polibatam.Ac.Id/Index.Php/Jaic

Aufa, R. N., Prasetiyowati, S. S., & Sibaroni, Y. (2023). The Effect Of Feature Weighting On Sentiment Analysis Tiktok Application Using The Rnn Classification. *Building Of Informatics, Technology And Science (Bits)*, *5*(1), 1. Https://Doi.Org/10.47065/Bits.V5i1.3597

Azza Amira, S., Utama, S., & Muhammad Hanif Fahmi, Dan. (2020). Edu Komputika Journal Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pada Review Pelanggan Hotel. *Edu Komputika*, *7*(2), 1–9. Http://Journal.Unnes.Ac.Id/Sju/Index.Php/Edukom

Başarslan, M. S., & Bal, F. (2025). The Effect Of Text Representation And Model Selection On Classification Performance: A Comprehensive Comparison Of Tf-Idf, Bow And Transformer-Based Methods On The Covid19-Fnir Dataset. *Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, *14*(4), 1–15. Https://Doi.Org/10.28948/Ngumuh.1694988

Bellar, O., Baina, A., & Ballafkih, M. (2024). Sentiment Analysis: Predicting Product Reviews For E-Commerce Recommendations Using Deep Learning And Transformers. *Mathematics*, *12*(12), 1. Https://Doi.Org/10.3390/Math12152403

Che Mohd Safawi, N. U., & Shafie, A. N. (2024). View Of Performance Of Tf-Idf For Text Classification Reviews On Google Play Store_ Shopee. *Of Computing Research And Innovation*, *9*, 1–10.

Chehal, D., Gupta, P., & Gulati, P. (2021). Evaluating Annotated Dataset Of Customer Reviews For Aspect Based Sentiment Analysis. In *Journal Of Web Engineering* (Vol. 21, Number 2, Pp. 1–34). River Publishers. Https://Doi.Org/10.13052/Jwe1540-9589.2122

Chinnalagu, A., & Durairaj, A. K. (2021). Context-Based Sentiment Analysis On Customer Reviews Using Machine Learning Linear Models. *Peerj Computer Science*, *7*, 1. Https://Doi.Org/10.7717/Peerj-Cs.813

Cristina Angelika, R. S., & Suhari, Y. (2026). Analisis Sentimen Ulasan Pelanggan Restoran Menggunakan Support Vector Machine. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 10, Number 1). Www.Llauk.Com

Dai, S., Li, K., Luo, Z., Zhao, P., Hong, B., Zhu, A., & Liu, J. (2024). Ai-Based Nlp Section Discusses The Application And Effect Of Words Models And Tf Based Nlp Section Discusses The Application And Effect Of Words Models And Tf-Idf In Nlp Tasks Based Nlp Section Discusses The Application And Effect Of Bag-Of. In *Journal Of Artificial Intelligence General Science Jaigs* (Vol. 5). Http://Creativecommons.Org/Lijournalofartificialintelligencegeneralsciencejaigshttps://Ojs.Boulibrary.Com/Index.Php/Jaigs//Ojs.Boulibrary.Com/Index.Php/Jaigshttp://Creativecommons.Org/Licenses/By/4.0

Das, M., Kamalanathan, S., & Alphonse, P. (2021). *A Comparative Study On Tf-Idf Feature Weighting Method And Its Analysis Using Unstructured Dataset*.

Deski Manalu, P., Simanjuntak, M., Umri, C., Tinggi, S., Ekonomi, I., Tinggi, P., & Profesional, M. (2025). Implementasi Algoritma Klasifikasi Untuk Analisis Sentimen Media Sosial Tiktok Tahun 2025. *Jurnal Teknik Informatika Dan Teknologi Informasi*, *5*(1), 1–17. Https://Doi.Org/10.55606/Jutiti.V5i1.5644

Dwi Utami Putra, D. M. (2025). Analisis Sentimen Terhadap Fenomena Artis Berjualan Di Tiktok Shop Pada Media Sosial X Menggunakan Model Naïve Bayes. *Sistem & Teknologi Informasi Komunikasi*, *8*(2), 1–6.

Ebrahim, M., Sedky, A. A. H., & Mesbah, S. (2023). Accuracy Assessment Of Machine Learning Algorithms Used To Predict Breast Cancer. *Data*, *8*(2). Https://Doi.Org/10.3390/Data8020035

Eva Lasfiana1, E. E. (2023). The Influence Of Price, Trust, And Product Reviews On Online Purchase Decisions On Tiktok Shop. In *Management Studies And Entrepreneurship Journal* (Vol. 4, Number 5). Http://Journal.Yrpipku.Com/Index.Php/Msej

Fathirachman Mahing, N., Gunawan, A. L., Foresta, A., Zen, A., Abdurrachman Bachtiar, F., Wicaksono, S. A., Brawijaya, U., & Korespondensi, P. (2024). Klasifikasi Tingkat Stres Dari Data Berbentuk Teks Dengan Menggunakan Algoritma Support Vector Machine (Svm) Dan Random Forest. *Teknologi Informasi Dan Ilmu Komputer (Jtiik)*, (Klasifikasi Tingkat Stres Dari Data Berbentuk Teks Dengan Menggunakan Algoritma Support Vector Machine (Svm) Dan Random Forest), 1–10. Https://Doi.Org/10.25126/Jtiik2024118010

Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *Transformtika*, *18*(Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine), 1–10. Www.Nusamandiri.Ac.Id,

Fitri, R. R., -, A., & Apriandari, W. (2025). Penggunaan Random Forest Dalam Sistem Klasifikasi Kecemasan Pada Generasi Z. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(Penggunaan Random Forest Dalam Sistem Klasifikasi Kecemasan Pada Generasi Z), 1–11. Https://Doi.Org/10.23960/Jitet.V13i3.6905

Gasparetto, A., Marcuzzo, M., Zangari, A., & Albarelli, A. (2022). A Survey On Text Classification Algorithms: From Text To Predictions. *Information (Switzerland)*, *13*(2), 1–39. Https://Doi.Org/10.3390/Info13020083

Harnelia, H. (2024). Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (Svm). *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2), 1–9. Https://Doi.Org/10.23960/Jitet.V12i2.4095

Homepage, J., Nurmadewi, D., Fahmi Jailani, Z., Rafi, H., Anggoro, A., & Setiowati, D. (2026). Malcom: Indonesian Journal Of Machine Learning And Computer Science Comparison Of Support Vector Machine And Naïve Bayes For E-Commerce Review Sentiment Classification Perbandingan Support Vector Machine Dan Naïve Bayes Untuk Klasifikasi Sentimen Ulasan E-Commerce. *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, *6*(2), 923–933. Https://Doi.Org/10.57152/Malcom.V6i2.2648

Huang, H., Asemi, A., & Mustafa, M. B. (2023). Sentiment Analysis In E-Commerce Platforms: A Review Of Current Techniques And Future Directions. *Iee Access*, *16*, 1–16. Https://Doi.Org/10.1109/Access.2023.0322000

Indrayuni, E., & Acmad Nurhadi. (2025). Analisis Sentimen Aplikasi Tiktok Shop Seller Center Menggunakan Naive Bayes, Svm Dan Logistic Regression. *Inti Nusa Mandiri*, *20*(1), 1–9. Https://Doi.Org/10.33480/Inti.V20i1.6851

Jessica Angelina, S., Bijaksana Putra Negara, A., Muhardi, H., Nawawi, J. H., & Barat, K. (2023). Analisis Pengaruh Penerapan Stopword Removal Pada Performa Klasifikasi Sentimen Tweet Bahasa Indonesia Analyzing The Impact Of Applying Stopword Removal On Indonesian Tweet Sentiment Classification. *Aplikasi Dan Riset Informatika*, *02*(1), 1–9. Https://Doi.Org/10.26418/Juara.V2i1.69680

Kalaivani, K. S., & Kuppuswami, S. (2019). Exploring The Use Of Syntactic Dependency Features For Document-Level Sentiment Classification. In *Bulletin Of The Polish Academy Of Sciences: Technical Sciences* (Vol. 67, Number 2, Pp. 1–9). Polish Academy Of Sciences. Https://Doi.Org/10.24425/Bpas.2019.128608

Khab Sulaiman Dalam, J., Oktavia Praneswara, A., Cahyono, N., & Amikom Yogyakarta, U. (N.D.). Analisis Sentimen Ulasan Aplikasi Tiktok Shop Seller Center Di Google Playstore Menggunakan Algoritma Naive Bayes. *Indonesian Journal Of Computer Science Attribution*, *12*(6).

Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). *Analisis Sentimen Ulasan Aplikasi Dana Dengan Metode Random Forest* (Vol. 6, Number 9). Http://J-Ptiik.Ub.Ac.Id

Leo Edward 1. (2025). The Effect Of Income, Past Experience, Customer Reviews, Store Ratings On Tiktok Shop Customer Buying Decisions. *Jurnal Akutansi Manajemen Dan Ekonomi*, (The Effect Of Income, Past Experience, Customer Reviews, Store Ratings On Tiktok Shop Customer Buying Decisions), 1–9.

Lestari, A., Irma Purnamasari, A., Bahtiar, A., & Tohidi, E. (2025). *Sentiment Analysis To Classify Tiktok Shop Users On Twitter With Naïve Bayes Classifier Algorithm* (Vol. 4, Number 2). Https://Ioinformatic.Org/

Madjid, M. F., Ratnawati, D. E., & Rahayudi, B. (2023). Sentiment Analysis On App Reviews Using Support Vector Machine And Naïve Bayes Classification. *Sinkron*, *8*(1), 1–7. Https://Doi.Org/10.33395/Sinkron.V8i1.12161

Maharani, D., & Yell, M. B. (2025). Customer Review Sentiment Analysis Of Alisa Batik Solo E-Commerce On Tiktok Using Naive Bayes Algorithm. *Journal Innovations Computer Science*, *4*(2), 298–13. Https://Doi.Org/10.56347/Jics.V4i2.334

Megawan, S., Gohzali, H., Halim, F., Ramadhan, H., & Sitepu, D. O. (2025). Sentiment Analysis Of Triv Application Reviews Using Support Vector Machine Algorithm. *Journal Of Computer Science And Informatics Engineering*, *4*(3), 1–13. Https://Doi.Org/10.55537/Cosie.V4i3.1223

Meridha Isnaeni, H., Stiawan, E., & Arisandi, D. (2026). *Analisis Pengaruh Ulasan Produk, Persepsi Harga, Dan Live Streaming Marketing Terhadap Minat Beli Produk Halal Di Tiktok Shop* (Vol. 12, Number 2).

Mohammad Alkharboush, F. (2021). Features Extraction Effect On The Accuracy Of Sentiment Classification Using Ensemble Models. *International Journal Of Science And Research (Ijsr)*, *10*(3), 1–4. Https://Doi.Org/10.21275/Sr21303123511

Muawanah, S., Muzayanah, U., Pandin, M. G. R., Alam, M. D. S., & Trisnaningtyas, J. P. N. (2023). Machine Learning Applications Based On Svm Classification: A Review. *Qubahan Academic Journal*, *3*(4), 1–10. Https://Doi.Org/10.48161/Issn.2709-8206

Nur Khaliza, A., Herry Chrisnanto, Y., Nurul Sabrina, P., & Id, A. (2022). *Sentiment Analysis Online Shop Social Media On Tik-Tok App With Naïve Bayes Classifier ( Nbc )Method*.

Nurul, A., Abidah, S. El, & Alfredo, R. (2024). Impulsive Buying Behavior In Affiliate Marketing Trends On Tiktok Shop Perilaku Impulsive Buying Dalam Tren Affiliate Marketing Di Tiktok Shop. *Jurnal Ilmu Komunikasi Pattimura*, (Impulsive Buying Behavior In Affiliate Marketing Trends On Tiktok Shop), 1–16. Https://Doi.Org/10.30598/Jikpvol3iss2pp616

Oktavianto, H., Sulistyo, H. W., Wijaya, G., Irawan, D., & Abdurrahman, G. (2024). Analisis Komparasi Kinerja Metode Decision Tree Dan Random Forest Dalam Klasifikasi Teks Data Kesehatan. *Bina Insani Ict Journal*, *11*(Analisis Komparasi Kinerja Metode Decision Tree Dan Random Forest Dalam Klasifikasi Teks Data Kesehatan), 1–10. Https://Www.Kaggle.Com/Datasets/Falgunipatel19/Biomedical-Text-Publication-Classification.

Pakpahan, D., Siallagan, V., & Siregar, S. (2023). Classification Of E-Commerce Product Descriptions With The Tf-Idf And Svm Methods. *Sinkron*, *8*(4), 1–8. Https://Doi.Org/10.33395/Sinkron.V8i4.12779

Palanivinayagam, A., & Damaševičius, R. (2023). Effective Handling Of Missing Values In Datasets For Classification Using Machine Learning Methods. *Information (Switzerland)*, *14*(2), 1–5. Https://Doi.Org/10.3390/Info14020092

Puh, K., & Bagić Babac, M. (2023). Predicting Sentiment And Rating Of Tourist Reviews Using Machine Learning. *Journal Of Hospitality And Tourism Insights*, *6*(3), 1–17. Https://Doi.Org/10.1108/Jhti-02-2022-0078

Qasim, R., Bangyal, W. H., Alqarni, M. A., & Ali Almazroi, A. (2022). A Fine-Tuned Bert-Based Transfer Learning Approach For Text Classification. *Journal Of Healthcare Engineering*, *2022*, 1–17. Https://Doi.Org/10.1155/2022/3498123

Raden Sigit Surya Pratama. (2025). Pengaruh Kegunaan Yang Dirasakan Dari Ulasan Online, Nilai Emosional, Risiko Yang Dirasakan Dan Kepercayaan Terhadap Minat Beli Pada Aplikasi Tiktok Shop. *Jpek (Jurnal Pendidikan Ekonomi Dan Kewirausahaan)*, *9*(1), 1–10. Https://Doi.Org/10.29408/Jpek.V9i1.29544

Rahmat Yuliawan. (2025). Pengaruh Ulasan Influencer Dan Diskon Di Tiktok Dalam Keputusan Pembelian Produk Di Tiktok Shop. *Pendidikan Tata Niaga*, *13*, 1–9.

Rahmawati, Fuadi, W., & Afrillia, Y. (2024). Cosmetic Shop Sentiment Analysis On Tiktok Shop Using The Support Vector Machine Method. *International Journal Of Engineering, Science And Information Technology*, *4*(2), 1–7. Https://Doi.Org/10.52088/Ijesty.V4i1.498

Rissa Nurfitriana Handayani. (2021). Optimasi Algoritma Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Produk Tokopedia Menggunakan Pso. *Media Informatika V*, *20*(Optimasi Algoritma Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Produk Tokopedia Menggunakan Pso), 1–2.

Rohanah, A., Rianti, D. L., Sari, B. N., Informatika, T., & Karawang, U. S. (2021). *String (Satuan Tulisan Riset Dan Inovasi Teknologi) Perbandingan Naïve Bayes Dan Support Vector Machine Untuk Klasifikasi Ulasan Pelanggan Indihome*.

Saepudin, A., Faqih, A., & Dwilestari, G. (N.D.). *Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest Dan Logistic Regression Pada Ulasan Shopee* (Vol. 18, Number 1).

Saraswati, N. W. S., & Diatri Indradewi, I. G. A. A. (2022). Recognize The Polarity Of Hotel Reviews Using Support Vector Machine. *Matrik : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *22*(1), 1–12. Https://Doi.Org/10.30812/Matrik.V22i1.1848

Setiawan, J. C., M. Lhaksmana, K., & Bunyamin, B. (2023). Sentiment Analysis Of Indonesian Tiktok Review Using Lstm And Indobertweet Algorithm. *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *8*(3), 1–7. Https://Doi.Org/10.29100/Jipi.V8i3.3911

Setiawan, Y., Maulidevi, N. U., & Surendro, K. (2023). The Use Of Dynamic N-Gram To Enhance Tf-Idf Features Extraction For Bahasa Indonesia Cyberbullying Classification. *Acm International Conference Proceeding Series*, 1–6. Https://Doi.Org/10.1145/3587828.3587858

Setyabudi, S., & Aryanny, E. (2025). Sentiment Analysis Of Lazada Marketplace User Ratings With Naïve Bayes And Support Vector Machine Methods Analisis Sentimen Penilaian Pengguna Marketplace Lazada Dengan Metode Naïve Bayes Dan Support Vector Machine. *Jurnal Inovtek Polbeng - Seri Informatika*, *10*(1), 1–12.

Sopi Nurdini, & Dadang Munandar. (2025). Pengaruh Video Promosi Tiktok Dan Ulasan Digital Terhadap Pembelian Spontan Produk Kecantikan The Originote Pada Konsumen Wanita Di Tiktok Shop. *Indonesian Journal Of Social Science And Education (Ijosse)*, *1*(3), 1–12. Https://Doi.Org/10.62567/Ijosse.V1i3.1345

Sri, Ch. K. (2025). Feature-Level Rating System Using Customer Reviews And Review Votes. *International Journal For Research In Applied Science And Engineering Technology*, *13*(4), 1–6. Https://Doi.Org/10.22214/Ijraset.2025.68868

Tyas Febiani1, K. D. C. S. S. (2023). Pengaruh Ulasan Pelanggan Daring Dan Berbagi Pengetahuan Konsumen Pada Keputusan Pembelian Di Tiktok Shop. *Brainy*, *4*(Pengaruh Ulasan Pelanggan Daring Dan Berbagi Pengetahuan Konsumen Pada Keputusan Pembelian Di Tiktok Shop), 1–10.

William Chandra. (2024). Peran Kualitas Layanan Dan Ulasan Produk Terhadap Kepuasan Pelanggan Di Tiktok Shop. *Jurnal Manajerial Dan Kewirausahaan*, (Peran Kualitas Layanan Dan Ulasan Produk Terhadap Kepuasan Pelanggan Di Tiktok Shop), 1–7.

Zainottah, M., Saputra, R., Servanda, Y., & Rosita, I. (2025). *Sentiment Analysis Of Tokopedia Electronic Products Using Svm-Logistic & Tf-Idf Ensemble Methods* (Vol. 4, Number 3). Https://Ioinformatic.Org/

Zhou, J., Ye, Z., Zhang, S., Geng, Z., Han, N., & Yang, T. (2024). Investigating Response Behavior Through Tf-Idf And Word2vec Text Analysis: A Case Study Of Pisa 2012 Problem-Solving Process Data. *Heliyon*, *10*(16), 1–22. Https://Doi.Org/10.1016/J.Heliyon.2024.E35945