JCSIS - Journal of Computer Science and Intelligent Systems

Available Online: https://eduscienta.com/index.php/jcsis

JOURNAL OF JCSIS

Vol.X, No.X, February XXXX

Comparing SVM, Random Forest, Logistic Regression, and Decision Tree for TikTok Sentiment Analysis of *Koperasi Desa Merah Putih*

Salman Alfarisi[1)*], Lidya Lunardi[2)], Fauzi[3)]

[1)]Department of Informatics Engineering, Faculty of Engineering and Computer Science, Indraprasta PGRI University
Nangka Street No.58 (TB. Simatupang), South Jakarta, Indonesia
[1)]salman.alfarisi@unindra.ac.id

[2)]Department of Information Systems, Faculty of Science and Technology, Buddhi Dharma University
41 Imam Bonjol Street, Tangerang, Indonesia
[2)]lidya.lunardi@ubd.ac.id

[3)]Department of Digital Business, Faculty of Economics and Business, Sheikh Yusuf Islamic University
Maulana Yusuf Street No.10, Tangerang, Indonesia
[3)]fauzi@unis.ac.id

Article History:

Received 27 June 2026;
Accepted 26 July 2026;
Available Online 29 July 2026

Keywords:

Koperasi Desa Merah Putih
Machine Learning
Sentiment Analysis
Support Vector Machine
TikTok

Abstract

The Koperasi Desa Merah Putih program is a strategic Indonesian government initiative designed to strengthen village economies through integrated cooperative services. This study analyzes sentiment expressed in Indonesian-language TikTok comments and compares Support Vector Machine, Random Forest, Logistic Regression, and Decision Tree under identical experimental conditions. The dataset comprised 49,805 labeled comments obtained from a public Kaggle repository. Data preparation included quality screening, case folding, slang normalization, tokenization, stop-word removal while retaining negation terms, and Indonesian stemming. The processed comments were represented using Term Frequency Inverse Document Frequency features and divided through a stratified 80:20 hold-out scheme, producing 39,844 training instances and 9,961 testing instances. Because Neutral comments represented 72.89% of the test set, evaluation emphasized macro-averaged metrics alongside accuracy. Support Vector Machine achieved the best performance, with 93.35% accuracy, 91.32% macro precision, 84.78% macro recall, and 87.70% macro F1-score, providing the most balanced overall classification performance among the evaluated models.

ISSN XXXX-XXXX (online) XXXX-XXXX (print) DOI:

Published by EduScienta

I. INTRODUCTION

Cooperatives play an important role in Indonesia's economy because they are based on the principles of togetherness, mutual cooperation, and equitable distribution of economic benefits. At the village level, cooperatives support local economic development by providing financial services, facilitating MSME growth, strengthening local supply chains, and improving community welfare through collective economic participation [1].

The Indonesian Government introduced the Koperasi Desa Merah Putih program as an effort to strengthen the village economy. This program is supported by Presidential Instruction of the Republic of Indonesia Number 9 of 2025, which mandates the establishment, development, and revitalization of approximately 80,000 village and urban-village cooperatives throughout Indonesia [2]. These cooperatives are expected to provide integrated services, including the provision of basic necessities, savings and loan services, pharmacies, warehousing, logistics, and cold storage facilities, with the objective of strengthening local economies, expanding employment opportunities, and shortening distribution chains.

The success of the program is determined not only by the number of cooperatives established but also by public acceptance and participation. Previous studies have shown that public trust, community participation, and transparent governance are essential factors influencing the sustainability and effectiveness of village cooperative programs [1]. Koperasi Desa Merah Putih may therefore receive diverse public responses, ranging from support regarding its potential economic benefits to criticism related to funding mechanisms, governance, and transparency.

TikTok has become one of the fastest-growing social media platforms for public discussion and opinion sharing regarding government policies and social issues. User-generated comments on TikTok provide valuable textual data that reflect public perceptions, attitudes, and reactions in real time, making the platform a valuable source for public opinion analysis. Recent studies have increasingly utilized social media comments as a reliable source for sentiment analysis because they capture spontaneous public responses to contemporary issues [3].

To analyze a large volume of comments efficiently, this study applies sentiment analysis using a machine learning approach. The comments are classified into three sentiment categories: positive, negative, and neutral. This study compares four widely used machine learning algorithms, namely Support Vector Machine (SVM), Random Forest, Logistic Regression, and Decision Tree, which have demonstrated competitive performance in Indonesian-language sentiment classification tasks [3].

The collected comments are processed through several text preprocessing stages, including cleaning, case folding, tokenization, normalization, stop-word removal, and stemming. Subsequently, the processed text is transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) technique, which remains one of the most effective feature extraction methods for traditional machine learning-based sentiment analysis [4]. The performance of each algorithm is evaluated using accuracy, precision, recall, F1-score, and confusion matrix.

This study aims to compare the performance of the four machine learning algorithms and identify the most effective model for classifying public sentiment toward Koperasi Desa Merah Putih. The findings are expected to provide an overview of public opinion regarding the program and serve as a practical reference for selecting appropriate machine learning algorithms in Indonesian social media sentiment analysis.

ISSN XXXX-XXXX (online) XXXX-XXXX (print) © The Authors. DOI:

Published by EduScienta

II. LITERATURE

The use of social media data to examine public responses to government policies has been demonstrated in several previous studies. Muhandhis and Ritonga [5] analyzed 5,036 TikTok comments concerning Indonesia's Public Housing Savings or *Tabungan Perumahan Rakyat* policy using the Random Forest algorithm. Their findings showed that negative sentiment dominated the public responses, while the classification model achieved an accuracy of 89%. The study demonstrated that TikTok comments can provide empirical information regarding public acceptance and criticism of government policies. However, the researchers employed only one classification algorithm, making it impossible to determine whether Random Forest was the most appropriate model for the dataset.

Rihastuti and Rosyidi [6] also applied Random Forest to analyze public sentiment expressed in TikTok comments regarding the development of Indonesia's new capital city. The study used 1,472 Indonesian-language comments categorized into positive and negative sentiments. The model achieved an accuracy of 77%, with precision of 78%, recall of 77%, and an F1-score of 77%. These relatively consistent results indicate that Random Forest can provide stable classification performance for public-policy discussions. Nevertheless, the absence of comparative algorithms limited the study's ability to assess the relative superiority of Random Forest.

Support Vector Machine has also been widely applied to sentiment classification. Fide et al. [7] analyzed user reviews of the TikTok application obtained from the Google Play Store using Support Vector Machine with a radial basis function kernel. Their research involved data scraping, text preprocessing, sentiment scoring, TF-IDF feature weighting, SVM classification, and text-association analysis. The results demonstrated that SVM could distinguish positive and negative opinions in high-dimensional textual data. However, the dataset consisted of application reviews rather than comments posted directly on TikTok, and the study evaluated only one principal classification algorithm.

Tangke et al. [8] compared Support Vector Machine and Random Forest for analyzing user sentiment toward the TikTok application. The study employed TF-IDF for feature weighting and Synthetic Minority Over-sampling Technique to address class imbalance. Random Forest achieved an accuracy of 93%, while Support Vector Machine achieved 90%. These results suggest that an ensemble-based classifier may outperform a margin-based classifier under certain dataset conditions. Nevertheless, the study focused on perceptions of the TikTok application itself rather than public responses to a national socioeconomic program.

Comparative classification approaches have been applied to evaluate the performance of C4.5, Naïve Bayes, and Support Vector Machine in predicting customer behavior. C4.5 generates interpretable decision rules through hierarchical feature splitting, Naïve Bayes estimates class probabilities based on conditional independence assumptions, while Support Vector Machine identifies an optimal separating boundary between classes. The comparison demonstrates that classification performance is strongly influenced by the structure of the dataset, feature characteristics, class distribution, and the learning mechanism of each algorithm. Therefore, no single classifier can be assumed to provide consistently superior performance across different application domains [9].

Different findings were reported by Mariwy et al. [10], who compared Support Vector Machine, Naïve Bayes, and K-Nearest Neighbors for classifying 7,900 TikTok comments related to cyberbullying. The dataset was divided into training and testing subsets using an 80:20 ratio. Support Vector Machine achieved the highest accuracy of 91.5%, followed by Naïve Bayes at 82.8% and K-Nearest Neighbors at 80.8%. These findings indicate that SVM is effective for

ISSN XXXX-XXXX (online) XXXX-XXXX (print) DOI:

Published by EduScienta

classifying sparse and high-dimensional social media text. However, the linguistic characteristics of cyberbullying comments may differ considerably from public-policy discussions involving institutional, economic, and governance-related terminology.

Fadhillah et al. [11] conducted a broader comparison involving Naïve Bayes, Support Vector Machine, Logistic Regression, and Random Forest in analyzing public sentiment toward TikTok Shop. The study used data collected from the X platform and applied lexicon-based sentiment labelling. Support Vector Machine achieved the highest accuracy of 81%, followed by Random Forest at 80%, Logistic Regression at 79%, and Naïve Bayes at 75%. The relatively small differences among the algorithms indicate that model performance may be strongly influenced by the characteristics of the dataset and feature representation. However, the study did not collect comments directly from TikTok and did not include Decision Tree as an independent classification model.

Data mining techniques have also been applied in cooperative financial management to identify members who are at risk of experiencing bad credit. Classification models such as Naïve Bayes and C4.5 can process historical borrower attributes, including payment behavior, loan characteristics, and member profiles, to distinguish between low-risk and high-risk credit cases. Naïve Bayes performs classification through probabilistic estimation, whereas C4.5 constructs hierarchical decision rules based on the attributes that provide the greatest information gain. Such comparative applications demonstrate that algorithm effectiveness depends on data structure, attribute quality, class distribution, and the interpretability required by decision-makers [9].

Nurfirdaus et al. [12] compared Support Vector Machine and Random Forest in analyzing TikTok comments regarding the E10 bioethanol fuel policy. Before class balancing, both algorithms achieved an accuracy of 82.19%. After the application of SMOTE, the accuracy increased to 89.55% for SVM and 89.59% for Random Forest. The minimal performance difference between the two classifiers demonstrates that preprocessing and class-balancing techniques can substantially affect classification results. The findings also indicate that selecting a model based exclusively on accuracy may be insufficient, particularly when the sentiment classes are unevenly distributed.

Deep learning approaches have been developed to improve the accuracy of text classification through the use of Convolutional Neural Networks (CNNs) combined with Particle Swarm Optimization (PSO). CNNs are employed to automatically extract local patterns and important relationships among words, while PSO is used to identify more optimal hyperparameter configurations, such as the number of filters, kernel size, and learning rate. This combination demonstrates that text classification performance is influenced not only by the model architecture but also by the effectiveness of the parameter optimization process. Although this approach offers more complex representational capabilities than TF-IDF-based methods, it generally requires greater computational resources and longer training processes [13].

The composition and contextual characteristics of a dataset can substantially influence the performance of deep learning models. In image classification, background elements may provide useful contextual cues or introduce irrelevant visual noise, depending on the relationship between the target object and its surrounding environment. Consequently, removing the background does not automatically improve classification accuracy because potentially informative features may also be eliminated. This indicates that dataset construction and preprocessing decisions should be evaluated empirically according to the characteristics of the data and the architecture of the classification model [14].

ISSN XXXX-XXXX (online) XXXX-XXXX (print) © The Authors. DOI:

Published by EduScienta

Deep learning classification performance is influenced not only by the selected model architecture but also by the composition and contextual characteristics of the dataset. In image-based classification, background elements may contain contextual information that supports feature extraction and helps the model distinguish among classes. Removing the background does not necessarily improve performance because it may eliminate relevant visual cues, although it can reduce input complexity and computational requirements. These findings emphasize that preprocessing decisions should be evaluated empirically, as their effects may vary according to the dataset characteristics and the architecture used [14].

The results of these studies reveal that no single algorithm consistently produces the best performance across all sentiment-analysis datasets. Random Forest outperformed Support Vector Machine in the analysis of TikTok application reviews [5], whereas SVM achieved the highest performance in studies concerning cyberbullying and TikTok Shop [6], [7]. In the analysis of the E10 fuel policy, SVM and Random Forest produced nearly identical accuracies after class balancing [8]. These inconsistencies suggest that algorithmic performance depends on the research topic, sample size, feature representation, preprocessing methods, class distribution, and linguistic characteristics of the dataset.

Several research gaps can be identified from the existing literature. First, previous studies have examined sentiment toward TikTok applications, cyberbullying, housing policy, capital-city development, digital commerce, and energy policy, whereas public sentiment concerning *Koperasi Desa Merah Putih* remains insufficiently investigated. The program contains distinctive socioeconomic, institutional, and governance dimensions that may produce different sentiment patterns from those found in commercial application reviews or other public-policy discussions.

Second, many previous studies employed only one or two algorithms, thereby limiting the comparative evidence available for model selection. Studies simultaneously comparing Support Vector Machine, Random Forest, Logistic Regression, and Decision Tree using Indonesian-language TikTok comments remain limited. The inclusion of Decision Tree is particularly relevant because its performance as a single-tree classifier can be compared directly with Random Forest as an ensemble of multiple trees.

Third, several previous studies emphasized accuracy as the principal indicator of model performance. Accuracy alone may provide an incomplete or misleading evaluation when positive, negative, and neutral comments are unequally distributed. A classifier may achieve high overall accuracy by correctly predicting the majority category while performing poorly on minority classes. Consequently, a comprehensive comparison should incorporate precision, recall, F1-score, and confusion matrices in addition to accuracy.

Based on these gaps, the present study compares Support Vector Machine, Random Forest, Logistic Regression, and Decision Tree under identical experimental conditions for classifying TikTok comments concerning *Koperasi Desa Merah Putih*. The study differs from previous research through its specific policy context, use of Indonesian-language TikTok comments, inclusion of four classifiers representing different learning mechanisms, and comprehensive evaluation using multiple performance metrics. This comparative approach is expected to provide a more robust understanding of algorithm suitability for sentiment analysis in the context of village-based economic policy.

ISSN XXXX-XXXX (online) XXXX-XXXX (print) © The Authors. DOI:

Published by EduScienta

III. RESEARCH METHOD

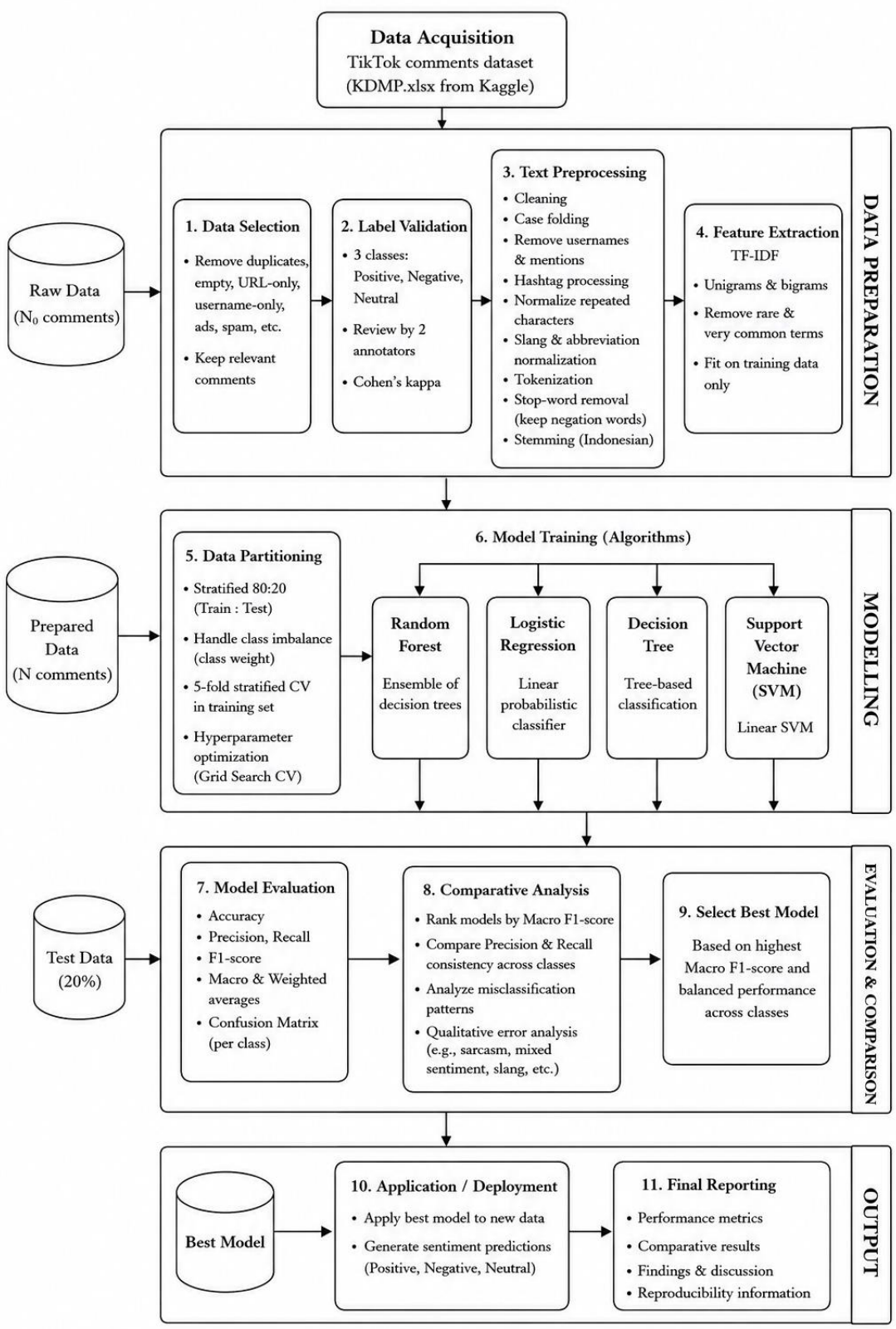

Figure 1. Study Methodology

ISSN XXXX-XXXX (online) XXXX-XXXX (print)
Published by EduScienta

The dataset used in this study was obtained from the public Kaggle repository entitled *Dataset Komentar TikTok Koperasi Desa Merah Putih*. The selected KDMP.xlsx file contains public TikTok comments discussing the *Koperasi Desa Merah Putih* program. Because the dataset was obtained from an existing public repository, it was treated as secondary data. The comment text and sentiment label were used as the primary variables, while usernames and other personal identifiers were excluded from the modelling process.

The data-understanding stage was conducted to examine the structure and quality of the dataset before analysis. This stage included identifying the available attributes, calculating the number of comments, examining missing values, detecting duplicate records, and evaluating the distribution of positive, negative, and neutral sentiment classes. Several comment samples were also reviewed to understand the linguistic characteristics of TikTok data, including slang, abbreviations, repeated characters, hashtags, account mentions, and non-standard Indonesian words. The class distribution was examined to determine whether the dataset was balanced and whether class-weight adjustment was required during model training.

Data preparation was conducted to reduce textual noise and transform the TikTok comments into a consistent format suitable for machine-learning classification [15], [16]. The preprocessing stages were as follows. Data preparation began with data cleaning to remove duplicate records, empty comments, hyperlinks, HTML characters, unnecessary punctuation, irrelevant symbols, advertisements, and repeated spam. All alphabetical characters were then converted to lowercase through case folding so that words with different capitalization were treated as the same feature. Usernames and account mentions beginning with the "@" symbol were removed because they generally did not contribute to sentiment classification. During hashtag processing, the hashtag symbol was eliminated while meaningful terms were retained; for example, *#TolakKoperasi* was converted into textual terms that preserved its semantic meaning. Excessive character repetitions were also normalized into standard forms, such as converting *baguuuus* into *bagus*. Informal expressions, abbreviations, and social media vocabulary were subsequently converted into their standard Indonesian equivalents; for instance, *gk*, *nggak*, and *ga* were normalized to *tidak*.

The cleaned comments were then tokenized into individual words or tokens to support computational processing. Common stop words with limited discriminative value were removed, while negation terms such as *tidak*, *bukan*, *belum*, and *jangan* were retained because they could alter the sentiment orientation of a sentence. Stemming was subsequently applied to reduce inflected Indonesian words to their root forms, such as transforming *membantu*, *dibantu*, and *bantuan* into *bantu*. After completing these preprocessing stages, the comments were transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency method.

The prepared dataset was divided into training and testing subsets using a stratified ratio of 80:20. Eighty percent of the data were used to train the classification models, while the remaining 20% were used for final evaluation. Stratification was applied to maintain approximately the same proportions of positive, negative, and neutral comments in both subsets. The same training and testing partitions were used for all algorithms to ensure a fair comparison.

Support Vector Machine classifies data by identifying an optimal hyperplane that maximizes the distance between different sentiment classes. The linear decision function is expressed as:

$$f(x) = w^T x + b \quad (1)$$

where (x) is the TF-IDF feature vector, (w) is the weight vector, and (b) is the bias.

Random Forest is an ensemble algorithm that combines multiple Decision Trees. Each tree is trained using a bootstrap sample of the training data and a randomly selected subset of features. The final prediction is determined through majority voting:

ISSN XXXX-XXXX (online) XXXX-XXXX (print) © The Authors. DOI:

Published by EduScienta

$$\hat{y} = mode\{(h_1(x), h_2(x), \ldots, h_B(x)\} \quad (2)$$

where $h_B(x)$ represents the prediction generated by the b-th tree and B represents the total number of trees. Combining multiple trees reduces the variance and overfitting risk associated with a single Decision Tree.

Logistic Regression estimates the probability that a comment belongs to a particular sentiment category. For binary classification, the probability is calculated using the sigmoid function:

$$P(y = k \mid x) = \frac{e^{w_k^t x + b_k}}{\sum_{j=i}^{K} e^{w_j^T x + b_j}} \quad (3)$$

where K represents the number of sentiment classes and $P(y=k|x)$ is the probability that comment x belongs to class k. The class with the highest probability is selected as the predicted sentiment.

Decision Tree classifies comments by recursively dividing the feature space into increasingly homogeneous subsets. The best feature and split point can be selected using the Gini impurity:

$$Gini(S) = 1 - \sum_{k=1}^{K} p_k^2 \quad (4)$$

$$Gini(S) = 1 - \left(p_{positive}^2 + p_{negative}^2 + p_{neutral}^2\right) \quad (5)$$

where p_k represents the proportion of observations belonging to class k within node S. A lower Gini value indicates that the observations within the node are more homogeneous. The splitting process continues until a stopping criterion, such as maximum tree depth or minimum number of samples, is reached.

The performance of the four classification algorithms was evaluated using accuracy, precision, recall, and F1-score [17], [18]. These metrics were calculated based on true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values [19], [20].

Accuracy measures the proportion of correctly classified comments among all observations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision measures the proportion of comments predicted as a particular sentiment that were correctly classified:

$$Precicion = \frac{TP}{TP + FP} \quad (7)$$

Recall measures the proportion of actual comments belonging to a particular sentiment that were correctly identified:

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

F1-score represents the harmonic mean of precision and recall:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

Because the study involved three sentiment categories, the metrics were calculated for each class and summarized using macro averaging. Macro averaging assigns equal importance to positive, negative, and neutral classes regardless of the number of comments in each category. A confusion matrix was also used to identify the distribution of correct and incorrect predictions across the three sentiment classes.

Comparative analysis was conducted by comparing the accuracy, macro precision, macro recall, and macro F1-score produced by Support Vector Machine, Random Forest, Logistic Regression,

ISSN XXXX-XXXX (online) XXXX-XXXX (print)
Published by EduScienta

and Decision Tree on the same testing dataset. Macro F1-score was used as the primary indicator because it provides a balanced assessment of precision and recall for all sentiment categories, particularly when the dataset is imbalanced.

The algorithm with the highest and most consistent evaluation results was identified as the best-performing model. The analysis also examined the confusion matrix of each algorithm to determine which sentiment categories were most frequently misclassified. In addition to numerical performance, the comparison considered the characteristics of each algorithm, including its ability to process high-dimensional TF-IDF features, resistance to overfitting, classification stability, and capacity to identify minority sentiment classes.

IV. Results and Discussion

Results

The dataset used in this study was classified into three sentiment classes: Neutral, Positive, and Negative. The classification process was conducted using four machine learning algorithms: Support Vector Machine (SVM), Random Forest, Logistic Regression, and Decision Tree. Each algorithm was applied to the same dataset to ensure that the evaluation results could be compared objectively.

The performance of each algorithm was evaluated based on the confusion matrix. The confusion matrix provides information on the number of data instances that were correctly and incorrectly classified within each sentiment class. These results were then used to calculate accuracy, precision, recall, and F1-score as performance evaluation metrics for each algorithm.

Support Vector Machine

The first experiment employed the Support Vector Machine (SVM) algorithm to classify the sentiment data. The classification results are presented in Table 1.

Table 1. SVM Confusion Matrix

	True Neutral	True Positive	True Negative
Pred. Neutral	7203	194	227
Pred. Positive	28	1111	109
Pred. Negative	30	74	985

The confusion matrix indicates the distribution of correct and incorrect predictions across the three sentiment classes. These classification outcomes were used to determine the overall performance metrics presented in Table 2.

Table 2. Support Vector Machine Evaluation Results

Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
93,35	91,32	84,78	87,70

The evaluation results show that the SVM algorithm achieved an accuracy of 93,35%, precision of 91,32%, recall of 84,78%, and F1-score of 87,70%. These values indicate the ability of the SVM algorithm to distinguish among Neutral, Positive, and Negative sentiments in the research dataset.

Random Forest

Random Forest was also applied to classify the sentiment data using the same experimental setting. Its classification outcomes are shown in Table 3.

Table 3. Random Forest Confusion Matrix

	True Neutral	True Positive	True Negative
Pred. Neutral	7219	311	412
Pred. Positive	23	1016	100
Pred. Negative	19	52	809

ISSN XXXX-XXXX (online) XXXX-XXXX (print) © The Authors. DOI:

Published by EduScienta

The confusion matrix illustrates the distribution of predicted classes relative to their actual labels. The diagonal entries represent correct classifications, while the remaining entries indicate classification errors.

Table 4. Random Forest Evaluation Results

Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
90,79	90,68	78,11	83,06

Based on the evaluation results, the Random Forest algorithm achieved an accuracy of 90,79%, precision of 90,68%, recall of 78,11%, and F1-score of 83,06%. These results demonstrate the ability of Random Forest to perform sentiment classification and serve as a basis for comparison with the other algorithms.

Logistic Regression

Logistic Regression was employed as another classification approach for identifying Neutral, Positive, and Negative sentiments. The resulting confusion matrix is presented in Table 5.

Table 5. Logistic Regression Confusion Matrix

	True Neutral	True Positive	True Negative
Pred. Neutral	7216	468	541
Pred. Positive	29	861	93
Pred. Negative	16	50	687

The confusion matrix reflects how the predicted sentiment labels correspond to the actual classes and provides the basis for calculating the model's performance measures.

Table 6. Logistic Regression Evaluation Results

Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
87,98	88,85	71,27	77,45

The evaluation results show that Logistic Regression achieved an accuracy of 87,98%, precision of 88,85%, recall of 71,27%, and F1-score of 77,45%. These values represent the model's ability to classify all sentiment classes within the research dataset.

Decision Tree

The Decision Tree algorithm was used as the final classification model in the experiment. Its prediction outcomes for the three sentiment categories are summarized in Figure 2.

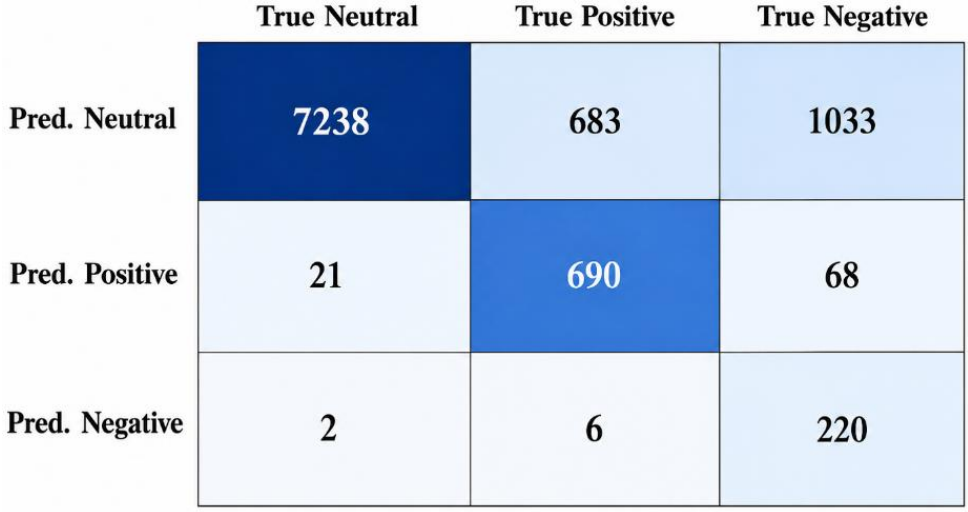

Figure 2. Decision Tree Confusion Matrix

The confusion matrix shows the distribution of correct and incorrect predictions for each sentiment category. The results were further evaluated using the four performance metrics presented in Figure 3.

ISSN XXXX-XXXX (online) XXXX-XXXX (print) © The Authors. DOI:

Published by EduScienta

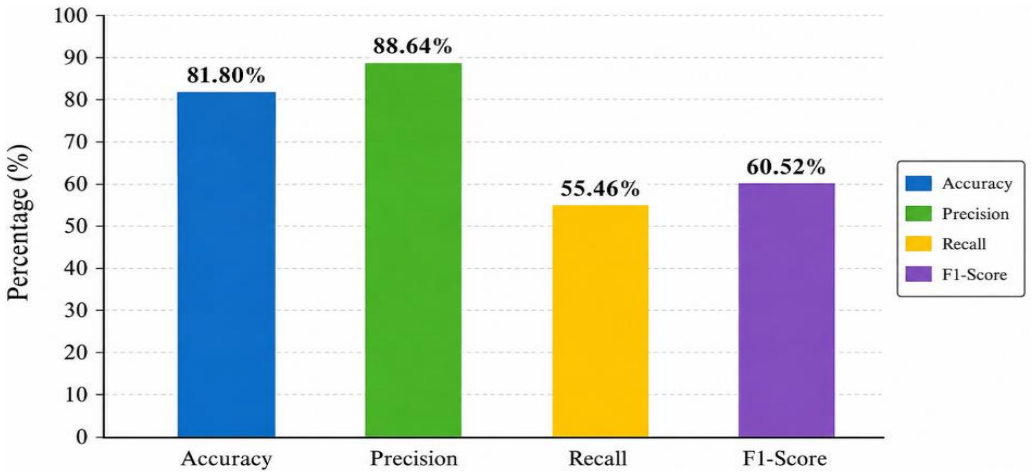

Figure 3. Decision Tree Evaluation Results

The evaluation results show that the Decision Tree algorithm achieved an accuracy of 81.80%, precision of 88.63%, recall of 55.46%, and F1-score of 60.54%. The model demonstrated a very high ability to identify the Neutral class, achieving a recall of 99.68%.

Discussion

The experimental results demonstrate clear differences in the performance of the four machine learning algorithms used for sentiment classification. Among the evaluated models, Support Vector Machine (SVM) achieved the best overall performance, with an accuracy of 93.35%, precision of 91.32%, recall of 84.78%, and F1-score of 87.70%. Random Forest ranked second with an accuracy of 90.79%, followed by Logistic Regression at 87.98% and Decision Tree at 81.80%. The comparison of the four algorithms is presented in Figure 4.

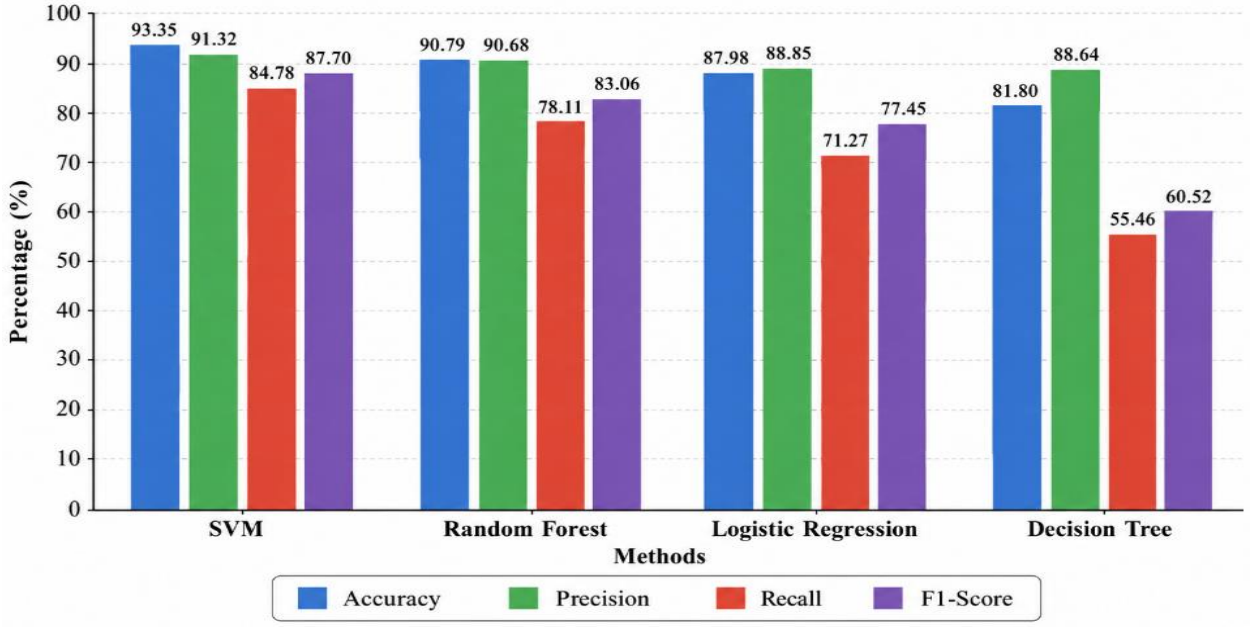

Figure 4. Comparison of Classification Model Performance

The superior performance of SVM in this study is consistent with recent studies showing its effectiveness for text-based classification. Khan et al. [21] compared SVM and Random Forest for sentiment analysis and demonstrated the competitive performance of SVM for textual sentiment classification. Similarly, Ruziqiana et al. [22] compared SVM, Naïve Bayes, and Random Forest for classifying Instagram comments into Positive, Negative, and Neutral categories. Their results

ISSN XXXX-XXXX (online) XXXX-XXXX (print) © The Authors. DOI:

Published by EduScienta

showed that SVM initially achieved 75.47% accuracy compared with 69.88% for Random Forest, while hyperparameter optimization increased SVM accuracy to 97.79%. These findings support the results of the present study, where SVM also outperformed Random Forest and achieved an accuracy of 93.35%.

Further evidence is provided by Hidayat et al. [23], who compared SVM, Logistic Regression, and Decision Tree for sentiment classification. Their experimental results showed that SVM consistently outperformed Logistic Regression and Decision Tree, achieving an average accuracy of 88.18% and reaching 92.69% in K-Fold testing. The authors also reported that Logistic Regression and Decision Tree showed lower accuracy and less stable performance when processing high-dimensional textual data. This pattern is consistent with the present study, where SVM achieved 93.35% accuracy, while Logistic Regression and Decision Tree obtained 87.98% and 81.80%, respectively.

The effectiveness of SVM and Random Forest is also supported by Subandono et al. [24], who examined feature-selection strategies for sentiment analysis. Their study found that SVM combined with Chi-Square feature selection achieved an accuracy of 93%, while Random Forest achieved 91%. Interestingly, these results are close to the findings of the present study, where SVM obtained an accuracy of 93.35% and Random Forest obtained 90.79%. This similarity provides additional evidence that both algorithms are effective for sentiment classification, although SVM demonstrated a slight advantage over Random Forest.

The results should also be interpreted by considering the distribution of sentiment classes. Of the 9,961 evaluated instances in the present study, 7,261 belonged to the Neutral class, 1,379 to the Positive class, and 1,321 to the Negative class. Therefore, approximately 72.89% of the dataset consisted of Neutral instances, indicating a substantial class imbalance. Recent research has emphasized that class distribution can affect the evaluation of sentiment classification models. A 2025 comparative study evaluating SVM, Naïve Bayes, KNN, Decision Tree, and Random Forest specifically examined model performance on both imbalanced and balanced sentiment datasets using accuracy, precision, recall, and F1-score. This reinforces the importance of considering multiple evaluation metrics rather than relying exclusively on accuracy [25].

The influence of class imbalance was particularly apparent in the Decision Tree results. Although Decision Tree correctly classified 7,238 of 7,261 Neutral instances, resulting in a recall of 99.68% for the Neutral class, only 690 of 1,379 Positive instances and 220 of 1,321 Negative instances were correctly identified. A considerable proportion of Positive and Negative instances were instead classified as Neutral. Consequently, Decision Tree obtained an overall recall of only 55.46% and an F1-score of 60.52%. This indicates that strong performance on the dominant class does not necessarily represent balanced classification performance across all sentiment categories.

In contrast, SVM demonstrated a more balanced ability to distinguish the three sentiment categories. It correctly classified 7,203 Neutral, 1,111 Positive, and 985 Negative instances. Random Forest correctly classified 7,219 Neutral, 1,016 Positive, and 809 Negative instances, whereas Logistic Regression correctly classified 7,216 Neutral, 861 Positive, and 687 Negative instances. Thus, SVM was better able to recognize the Positive and Negative classes while maintaining high classification performance for the dominant Neutral class. This explains why SVM obtained not only the highest accuracy but also the highest recall and F1-score.

Overall, the experimental results establish the following performance order: SVM (93.35%) > Random Forest (90.79%) > Logistic Regression (87.98%) > Decision Tree (81.80%). More importantly, SVM consistently achieved the highest performance across accuracy, precision, recall, and F1-score. These findings are broadly consistent with recent sentiment-classification

ISSN XXXX-XXXX (online) XXXX-XXXX (print) © The Authors. DOI:

Published by EduScienta

research in which SVM frequently demonstrated competitive or superior performance compared with other conventional machine learning algorithms.

The primary contribution of this study lies in providing an empirical comparison of four machine learning algorithms under the same dataset and experimental conditions. The findings demonstrate that model evaluation on sentiment data should not be based solely on overall accuracy, particularly when the dataset contains an unequal distribution of sentiment classes. In this study, SVM showed the strongest ability to maintain a balance between precision and recall, as reflected by its F1-score of 87.70%, making it the most effective model among the four evaluated algorithms.

Nevertheless, the dominance of Neutral instances represents an important limitation of the current experiment. Future studies should investigate data-balancing techniques, such as oversampling, undersampling, SMOTE, or class weighting, and evaluate whether these approaches improve the recognition of Positive and Negative sentiments. Hyperparameter optimization and feature-selection methods may also be explored, particularly because recent studies have demonstrated that parameter tuning and feature selection can substantially influence SVM and Random Forest performance.

V. Conclusion

This study addressed three questions concerning sentiment distribution, comparative model performance, and class-specific classification difficulty. Under the dataset, preprocessing procedure, TF-IDF representation, and evaluation design used in this study, Support Vector Machine achieved the strongest hold-out performance, with an accuracy of 93.35%, macro precision of 91.32%, macro recall of 84.78%, and macro F1-score of 87.70%. Its macro F1-score exceeded those of Random Forest, Logistic Regression, and Decision Tree by 4.64, 10.25, and 27.16 percentage points, respectively. Thus, SVM was the best-performing model among the four conventional classifiers evaluated in this experiment.

Class-specific evaluation showed that the Negative category was the most difficult to identify. SVM achieved a Negative-class recall of 74.56%, whereas Random Forest, Logistic Regression, and Decision Tree achieved 61.24%, 52.01%, and 16.65%, respectively. The dominance of Neutral comments, which represented approximately 72.89% of the test set, encouraged the models, particularly Decision Tree, to classify minority-class instances as Neutral. This result demonstrates that accuracy should be interpreted together with macro-averaged and class-specific metrics when evaluating imbalanced sentiment datasets.

From a technical perspective, SVM can be recommended as a relatively efficient baseline for Indonesian-language sentiment monitoring using sparse TF-IDF features. However, its superiority should be understood within the specific dataset, feature representation, and experimental design used in this study. Operational implementation should therefore include human review of ambiguous predictions, periodic model updating, bias auditing, and concept-drift monitoring to maintain classification reliability as public discourse and vocabulary change over time.

From a substantive perspective, the findings describe sentiment expressed in the sampled TikTok comments and should not be interpreted as a representative estimate of Indonesian public opinion as a whole. Future studies should validate these findings using group-based or temporal data splitting, independently annotated datasets, external test data, and class-balancing strategies. Comparative evaluation with contextual language models, including IndoBERT and XLM-

ISSN XXXX-XXXX (online) XXXX-XXXX (print) © The Authors. DOI:

Published by EduScienta

RoBERTa, is also recommended to determine whether deeper contextual representations can improve minority-class recognition and overall model robustness..

VI. Acknowledgments

The authors would like to express their sincere gratitude to all parties who contributed to the completion of this research. Appreciation is extended to the institution for providing academic support, research facilities, and administrative assistance throughout the study. The authors also thank the respondents, participants, or informants who voluntarily provided valuable data and information for this research. In addition, the authors acknowledge the constructive input and academic suggestions from colleagues and reviewers that helped improve the quality of this manuscript. Any errors, interpretations, and conclusions presented in this article remain the sole responsibility of the authors.

VII. Supporting Information

All data used in this study are provided through a publicly accessible repository as part of the authors' commitment to transparency, data openness, and scientific accountability. The primary dataset consists of TikTok comments related to the *Koperasi Desa Merah Putih* (KDMP) and was obtained from the publicly available Kaggle repository entitled Dataset Komentar TikTok Koperasi Desa Merah Putih. The original dataset was subsequently processed through several text-preprocessing stages to improve data quality and ensure its suitability for sentiment classification. The processed data were classified into three sentiment categories: Neutral, Positive, and Negative. The resulting dataset was used as the primary input for the training and testing processes involving Support Vector Machine (SVM), Random Forest, Logistic Regression, and Decision Tree. Detailed information regarding the dataset source, attribute structure, class distribution, inclusion criteria, and sentiment-label specifications is provided in Appendix.

References

[1] D. Mulyadi and A. Rahmati, "Revitalizing Merah Putih Village Cooperatives through Maqasid Sharia : A qualitative study of rural economic recovery in Indonesia," *J. Islam. Econ. Lariba*, vol. 12, no. 1, pp. 673–694, 2026, doi: 10.20885/jielariba.vol12.iss1.art24.

[2] "Inilah Inpres 9/2025 tentang Percepatan Pembentukan Koperasi Desa/Kelurahan Merah Putih | Sekretariat Negara."

[3] H. S. Ramadhan, A. S. Akbar, K. Y. Sinaga, L. Muthoharoh, A. Satria, and M. C. T. Manullang, "Sentiment Analysis of AI Adoption in Indonesian Higher Education Using Machine Learning and Transformer-Based Models," Apr. 2026.

[4] A. N. I. Pasha, E. F. Putri, L. Muthoharoh, A. Satria, and M. C. T. Manullang, "Hybrid TF--IDF Logistic Regression and MLP Neural Baseline for Indonesian Three-Class Sentiment Analysis on Social Media Text," May 2026.

[5] I. Muhandhis and A. S. Ritonga, "Public Sentiment Analysis on TikTok about Tapera Policy using Random Forest Classifier," *SISTEMASI*, vol. 14, no. 1, p. 354, Jan. 2025, doi: 10.32520/stmsi.v14i1.4878.

[6] Siti Rihastuti and Afnan Rosyidi, "ANALISIS SENTIMEN PENGGUNA TIKTOK TENTANG PROGRES PEMBANGUNAN IKN DENGAN METODE RANDOM FOREST," *J. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 19–23, May 2025, doi: 10.54840/jcstech.v5i1.345.

[7] S. Fide, S. Suparti, and S. Sudarno, "No Title," *J. Gaussian; Vol 10, No 3 J. GaussianDO*

- *10.14710/j.gauss.10.3.346-358* , Dec. 2021.
[8] R. Tangke, D. Tineke Salaki, W. Widsli Kalengkongan, and E. Ketaren, "Analisis Sentimen Aplikasi TikTok Menggunakan Algoritma Support Vector Machine (SVM) dan Random Forest," *J. TIMES*, vol. 13, no. 2, pp. 53–62, Dec. 2024, doi: 10.51351/jtm.13.2.2024762.
[9] R. Renaldi and Y. Kurnia, "Alleged Bad Credit at Saving Cooperatives Borrow Flamboyant Assistance PPSW Jakarta With Comparasion the Algorithms Naive Bayes and C4.5," *bit-Tech*, vol. 2, no. 3, pp. 141–147, Nov. 2020, doi: 10.32877/bt.v2i3.163.
[10] C. F. Mariwy, L. Y. Baisa, and A. L. Sumendap, "Comparative Study of Machine Learning Methods for Sentiment Analysis of TikTok Comments Related to Cyberbullying," *Indones. J. Artif. Intell. Data Min.*, vol. 9, no. 1 SE-Articles, pp. 140–152, doi: 10.24014/ijaidm.v9i1.39183.
[11] O. S. D. Fadhillah, J. H. Jaman, and C. Carudin, "Perbandingan Naive Bayes, Support Vector Machine, Logistic Regression dan Random Forest dalam Menganalisis Sentimen Mengenai TikTokShop," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5746.
[12] R. E. Nurfirdaus, M. A. Barata, and I. W. Prastya, "Comparison of SVM and Random Forest for TikTok E10 Fuel Sentiment Analysis," *J. Appl. Informatics Comput.*, vol. 10, no. 2, pp. 1426–1437, Apr. 2026, doi: 10.30871/jaic.v10i2.12395.
[13] A. Hermawan, L. Lunardi, Y. Kurnia, B. Daniawan, and Junaedi, "Optimizing Convolutional Neural Networks with Particle Swarm Optimization for Enhanced Hoax News Detection," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 11, no. 1, pp. 53–64, Mar. 2025, doi: 10.20473/jisebi.11.1.53-64.
[14] N. Azzahra, A. Hermawan, Junaedi, Y. Kurnia, and Edy, "Impact of Dataset Background on Deep Learning-Based Waste Classification," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 10, no. 3, pp. 580–589, Jun. 2026, doi: 10.29207/resti.v10i3.6965.
[15] Y.-W. Mak, H.-N. Goh, and A. H.-L. Lim, "Forum Text Processing and Summarization," *JOIV Int. J. Informatics Vis.*, vol. 8, no. 1, p. 425, Mar. 2024, doi: 10.62527/joiv.8.1.2279.
[16] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," *PLoS One*, vol. 15, no. 5, p. e0232525, May 2020, doi: 10.1371/journal.pone.0232525.
[17] S. M. Nagarajan and U. D. Gandhi, "Classifying streaming of Twitter data based on sentiment analysis using hybridization," *Neural Comput. Appl.*, vol. 31, no. 5, pp. 1425–1433, 2019, doi: 10.1007/s00521-018-3476-3.
[18] D. Michail, N. Kanakaris, and I. Varlamis, "Detection of fake news campaigns using graph convolutional networks," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100104, 2022, doi: 10.1016/j.jjimei.2022.100104.
[19] F. Rahutomo, I. Y. R. Pratiwi, and D. M. Ramadhani, "Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia," *J. Penelit. Komun. Dan Opini Publik*, vol. 23, no. 1, 2019, doi: 10.33299/jpkop.23.1.1805.
[20] D. K. Vishwakarma, D. Varshney, and A. Yadav, "Detection and veracity analysis of fake news via scrapping and authenticating the web search," *Cogn. Syst. Res.*, vol. 58, pp. 217–229, 2019, doi: 10.1016/j.cogsys.2019.07.004.
[21] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *J. Informatics Web Eng.*, vol. 3, no. 1, p. 67, Feb. 2024, doi: 10.33093/jiwe.2024.3.1.5.
[22] M. F. Ruziqiana, L. Hidayah, and M. A. Rasyidi, "Detection of Cyberbullying Using Svm,

ISSN XXXX-XXXX (online) XXXX-XXXX (print)
Published by EduScienta

Naive Bayes, and Random Forest Algorithm," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2830–7062, 2024.

[23] R. Hidayat and F. Aminulhaq, "A Comparative Analysis of Decision Tree, Logistic Regression, and Support Vector Machine Algorithms in Sentiment Analysis of Threads App Reviews," *Intechno J. (Information Technol. Journal)*, vol. 7, no. 2, pp. 45–55, Dec. 2025, doi: 10.24076/intechnojournal.2025v7i2.2497.

[24] A. T. P. Subandono and D. Ariatmanto, "Optimizing Feature Selection in Sentiment Analysis of Bank Saqu: A Comparative Study of SVM and Random Forest using Information Gain and Chi-Square," *SISTEMASI*, vol. 14, no. 3, p. 1205, May 2025, doi: 10.32520/stmsi.v14i3.5106.

[25] S. Kristianto and Y. Kurnia, "Data Mining Implementation on Choosing Potential Customers Using K-Means Algorithm on PT. Koba Metal Indonesia," *Tech-E*, vol. 1, no. 2, p. 1, Feb. 2018, doi: 10.31253/te.v1i2.45.

DISCLAIMER/PUBLISHER'S NOTE

All statements, opinions, and data presented in each publication remain the sole responsibility of the respective authors and/or contributors and do not necessarily reflect the official position of JCSIS or its editorial team. JCSIS and its editors shall not be held liable for any loss, injury, or damage to property resulting from the application or use of any ideas, methods, instructions, products, or other information contained in the published material. In line with the principles of academic accountability and publication ethics, the author confirms that the manuscript submitted to JCSIS complies with the requirements set forth in the Statement of Originality.