Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3788~3795
e-ISSN: 2548-6861

TikTok Sentiment Analysis on Koperasi Merah Putih Using SVM and ANN

Meliysa Pasa Bagna Aprilia Said [1], Kholifatus Sholihah [2], Ifnu Wisma Dwi Prastya [3], Afril Efan Pajri [4]
Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
bagnasaid4@gmail.com [1], olalap125@gmail.com [2], ifnuprastya@unugiri.ac.id [3], afril@unugiri.ac.id [4]

Article Info

Article history:

Received 2026-06-19
Revised 2026-07-24
Accepted 2026-08-12

Keyword:

ANN,
Koperasi Merah Putih,
Lexicon-based labeling,
SVM,
TikTok.

ABSTRACT

TikTok has become a relevant social media source for observing public responses to public issues, including the Koperasi Merah Putih program. This study compares Support Vector Machine (SVM) and Artificial Neural Network (ANN) for classifying sentiment in TikTok comments. The dataset was obtained through TikTok comment scraping and consisted of 25,669 raw comments. After removing empty comments and applying preprocessing stages consisting of cleaning, case folding, normalization, tokenization, stopword removal, and stemming, 21,026 comments were used for sentiment analysis. Sentiment labels were generated automatically using a lexicon-based sentiment labeling approach and grouped into three classes: positive, negative, and neutral. TF-IDF was used for feature extraction with a maximum of 5,000 features and unigram-bigram representation. The dataset was split into training and testing sets with an 80:20 ratio, while Stratified K-Fold Cross Validation and SMOTE were applied to strengthen evaluation and address class imbalance. The results show that SVM achieved the best overall performance before SMOTE with an accuracy of 86.66% and an F1-score of 86.76%. ANN achieved an accuracy of 85.31% before SMOTE and improved slightly after SMOTE to 85.47%. These findings indicate that SVM is more stable for TF-IDF-based TikTok comment classification, while SMOTE can improve ANN performance slightly but does not always increase all models equally.

I. INTRODUCTION

Social media has become an important source for monitoring public opinion because users respond to public issues through comments, reposts, short videos, and informal discussions. TikTok is particularly relevant because its comment section reflects immediate, expressive, and spontaneous responses toward social and economic issues [1], [2]. Compared with longer-text platforms, TikTok comments are generally shorter, more informal, and often contain slang, abbreviations, repeated characters, humor, and sarcasm. These characteristics make TikTok comments challenging for sentiment analysis because the meaning of a comment is not always explicitly represented by formal words.

One public issue that has attracted attention is the Koperasi Merah Putih program. Public responses to this issue are diverse, ranging from support and optimism to criticism, doubt, satire, and distrust. Analyzing these responses is important because social media comments can provide an early picture of public perception toward policy-related issues. However, sentiment analysis on TikTok comments related to public policy issues is still relatively limited. Previous studies have mostly focused on application reviews, marketplace reviews, political discussions, general social media sentiment, or other popular topics, while studies that specifically examine short TikTok comments on the Koperasi Merah Putih issue remain uncommon.

Sentiment analysis is a Natural Language Processing (NLP) approach used to classify textual opinion into categories such as positive, negative, and neutral [3], [4]. Social media comments are challenging because they are short, informal, and often contain non-standard language. Therefore, preprocessing and feature extraction are essential stages before classification [5], [6]. Previous studies have shown that Support Vector Machine (SVM) is effective for sentiment classification because it performs well in high-dimensional TF-IDF feature spaces [7], [8]. Artificial Neural Network (ANN) is also widely used because it can learn nonlinear patterns from numerical representations [3], [9]. Nevertheless, the effectiveness of

ANN depends strongly on the quality and richness of the input representation.

Several previous studies have compared machine learning models for sentiment analysis in Indonesian contexts. Nurfirdaus et al. compared SVM and Random Forest for TikTok E10 fuel comments and used lexicon-based labeling, TF-IDF, SMOTE, and classification metrics [1]. Other studies have applied SVM, ANN, and SMOTE for sentiment classification on application reviews and social media data [3], [8], [10], [11]. These studies show that model performance is influenced by preprocessing quality, class distribution, and feature representation. However, most of these studies do not specifically discuss how short, informal, and satire-prone TikTok comments affect model performance on public-policy-related issues. In addition, the impact of SMOTE on linear models such as SVM and neural-network-based models such as ANN still needs to be examined further because oversampling does not always improve every algorithm equally.

Although transformer-based models such as IndoBERT, RoBERTa, and other contextual embedding approaches have shown strong performance in text classification, this study focuses on TF-IDF-based SVM and ANN as lightweight and reproducible baseline models. This choice was made to evaluate whether conventional sparse text representation can still provide reliable performance for short TikTok comments. SVM was selected because it is suitable for high-dimensional sparse features, while ANN was selected as a nonlinear comparison model to observe whether a neural-based classifier can improve performance under the same TF-IDF representation. Transformer-based models are not included as the main baseline in this study because they require greater computational resources and a different representation framework. However, they are considered important directions for future comparison.

The novelty of this study does not lie solely in comparing SVM and ANN, but in applying and analyzing these models on a large TikTok comment dataset related to the Koperasi Merah Putih public issue. This study combines lexicon-based sentiment labeling, TF-IDF feature extraction with unigram-bigram representation, Stratified K-Fold Cross Validation, SMOTE-based class balancing, and comparative evaluation of SVM and ANN. The study also discusses how class imbalance and TF-IDF representation influence model performance, particularly why SVM can outperform ANN on short and sparse TikTok comment data.

This study aims to compare the performance of Support Vector Machine (SVM) and Artificial Neural Network (ANN) in classifying TikTok comments about Koperasi Merah Putih into positive, negative, and neutral sentiment classes. The contribution of this study is expected to provide a reproducible baseline for TikTok-based public opinion analysis and to clarify the strengths and limitations of TF-IDF-based SVM and ANN for short Indonesian social media comments.

II. METHODS

The research workflow consists of data collection, preprocessing, lexicon-based labeling, feature extraction, data splitting, cross-validation, data balancing using SMOTE, classification using SVM and ANN, and performance evaluation. This structure follows a sequential sentiment-analysis pipeline commonly used in social-media-based classification studies [1], [12], [13].

A. Data Collection

The dataset was obtained from TikTok comments related to the Koperasi Merah Putih issue using an automated scraping process. Data collection was conducted from May to June 2025 by searching TikTok videos using keywords related to the issue, such as “Koperasi Merah Putih”, “Kopdes Merah Putih”, “Koperasi Desa Merah Putih”, and “program Koperasi Merah Putih”. These keywords were used to retrieve videos and comments discussing the program, public responses, criticism, support, and related opinions. The data collection focused on publicly available TikTok comments, and private account information was not used in the analysis.

The raw dataset contained 25,669 comment records collected from multiple TikTok videos. Each record included comment text and metadata such as comment ID, username, video link, number of likes, and number of replies. The metadata was retained only for documentation and duplicate checking, while the comment text was used as the main input for sentiment classification. To improve data quality, comments with empty text, missing values, duplicate entries, and non-relevant content were removed. After this filtering process, 21,026 comments were used as the final dataset for sentiment analysis.

The scraping and filtering procedures were performed to ensure that the dataset represented TikTok users’ responses to the Koperasi Merah Putih issue. However, this study only used comments from TikTok and did not include other platforms such as X, Instagram, Facebook, or YouTube. Therefore, the results represent public responses found in TikTok comments and should not be generalized to all social media users. The dataset summary is presented in Table I.

TABLE I
DATASET SUMMARY

Description	Value
Raw comments	25,669
Non-empty comments	25,646
Final comments	21,026
Classes	Pos., Neg., Neu.

B. Data Preprocessing

The preprocessing stage was performed to reduce noise and standardize the textual form. The stages included

cleaning, case folding, normalization, tokenization, stopword removal, and stemming using the Sastrawi algorithm. Cleaning removed URLs, symbols, numbers, punctuation, mentions, and non-textual elements. Case folding converted all letters to lowercase. Normalization converted non-standard social media terms into conventional forms. Tokenization split comments into word units. Stopword removal removed common words while retaining negation terms such as “tidak”, “bukan”, and “belum”. Stemming transformed inflected Indonesian words into base forms [6], [14]. The preprocessing stages are shown in Table II to illustrate the transformation of raw comments into standardized text data.

TABLE II
PREPROCESSING STAGES

Stage	Result
Original	kapan BUMN dikelola dikonoha untung
Clean	kapan bumn dikelola dikonoha untung
Norm.	kapan bumn dikelola dikonoha untung
Stem	kapan bumn kelola dikonoha untung

C. Lexicon-Based Sentiment Labeling

Sentiment labeling was conducted automatically using a lexicon-based sentiment labeling approach. This method was selected because it provides a reproducible rule-based mechanism for assigning sentiment labels to a large number of TikTok comments without requiring full manual annotation. The InSet lexicon was used as the main sentiment dictionary because it provides Indonesian positive and negative words with sentiment weights [12]. Each preprocessed comment was compared with the positive and negative lexicons to calculate its sentiment score.

The labeling rule was determined based on the comparison between positive and negative scores. A comment was classified as positive when the positive score was higher than the negative score. A comment was classified as negative when the negative score was higher than the positive score. Meanwhile, a comment was classified as neutral when the positive and negative scores were equal or when no dominant sentiment word was detected. The lexicon-based labels were then used as the target variable for the SVM and ANN models.

To reduce potential labeling errors, a manual validation process was conducted on a sample of 50 comments. The validation was performed by comparing the automatically generated labels with manually checked sentiment labels. This step was used as a control mechanism to observe whether the lexicon-based labels were generally consistent with the actual meaning of the comments. However, because the labeling process was still mainly based on lexicon scores, some limitations remain, especially for comments containing sarcasm, irony, slang, abbreviations, or context-dependent expressions.

Before model training, the sentiment class distribution was analyzed to identify possible imbalance among positive, negative, and neutral categories. The labeling results produced 12,370 negative comments, 4,948 positive comments, and 3,708 neutral comments. This distribution shows that the dataset was imbalanced because negative comments dominated 58.83% of the total data, while positive and neutral comments accounted for 23.53% and 17.64%, respectively. This imbalance motivated the use of SMOTE in the training phase [13], [15]. The distribution of sentiment labels generated using the lexicon-based approach is presented in Table III.

TABLE III
LEXICON LABEL DISTRIBUTION

Label	Count	Percent
Negative	12,370	58.83%
Positive	4,948	23.53%
Neutral	3,708	17.64%
Total	21,026	100%

The sentiment label distribution is also visualized in Figure 1 to provide a clearer overview of the class composition.

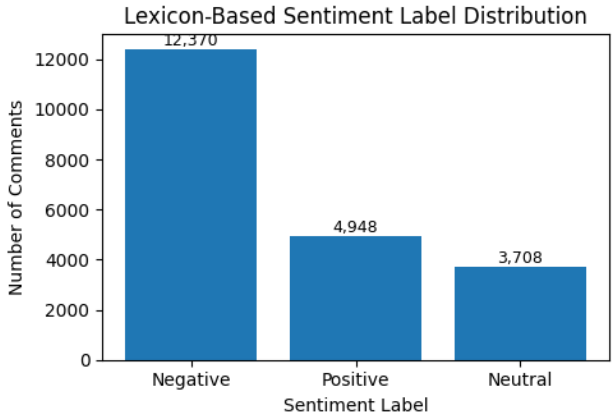

Figure 1. Distribution of Sentiment Labels

The manual validation of 50 sampled comments showed that the lexicon-based labeling results were generally consistent with the manually checked sentiment categories. However, several mismatches were still found in comments containing sarcasm, informal expressions, and context-dependent meanings. This finding indicates that lexicon-based labeling can be used as an initial automatic labeling approach, but it still has limitations in capturing implicit sentiment in TikTok comments.

D. Feature Extraction and Data Splitting

The cleaned and labeled comments were transformed into numerical features using Term Frequency-Inverse Document Frequency (TF-IDF). The vectorizer used a maximum of 5,000 features with unigram and bigram representation. The dataset was divided into 80% training data and 20% testing data using stratified splitting to preserve class proportions. The resulting training set consisted of 16,820 comments, while the testing set consisted of 4,206 comments.

TABLE IV
DATA SPLITTING

Partition	Count	Percent
Training	16,820	80%
Testing	4,206	20%
Total	21,026	100%

E. K-Fold Cross Validation

Stratified K-Fold Cross Validation with five folds was applied to evaluate model stability. Stratification ensures that each fold retains the class distribution of the original dataset. This process reduces dependence on a single train-test split and provides a more robust estimate of model generalization [1], [10]. The K-Fold cross-validation results for SVM and ANN are presented in Table V.

TABLE V
K-FOLD ACCURACY

Model	F1	F2	F3	F4	F5	Avg.
SVM	0.8739	0.8613	0.8596	0.8706	0.8632	0.8658
ANN	0.8704	0.8573	0.8523	0.8671	0.8590	0.8612

A paired t-test was conducted using the five-fold cross-validation scores to examine whether the performance difference between SVM and ANN was statistically significant. The SVM model obtained a mean score of 0.8657, while the ANN model obtained a mean score of 0.8612. The paired t-test produced a t-statistic of 6.3074 and a p-value of 0.0032. Since the p-value was lower than 0.05, the difference between SVM and ANN was statistically significant. This result indicates that the higher performance of SVM was not only observed from the average score but was also supported by statistical testing across the five folds.

F. SMOTE and Classification Models

The class distribution in the training set was imbalanced, with 9,896 negative comments, 3,958 positive comments, and 2,966 neutral comments. This imbalance shows that the negative class dominated the training data, while the positive and neutral classes had fewer samples. To address this condition, Synthetic Minority Over-sampling Technique (SMOTE) was applied only to the training set. SMOTE was not applied to the testing set to avoid data leakage and to ensure that model evaluation was performed on the original data distribution. After applying SMOTE, each sentiment class in the training set contained 9,896 instances, resulting in 29,688 training samples. The distribution of training data before applying SMOTE is shown in Figure 2.

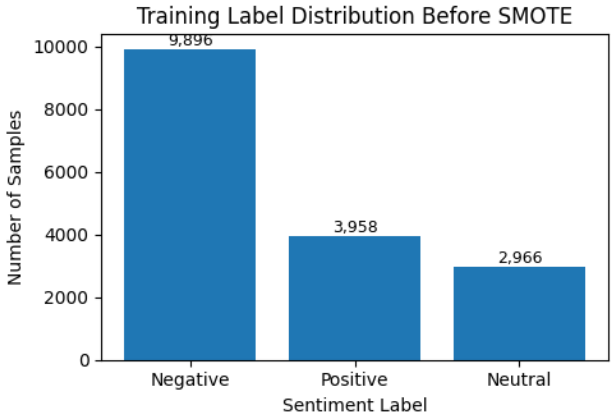

Figure 2. Training Label Distribution Before SMOTE

After applying SMOTE, the training data became more balanced across sentiment classes, as shown in Figure 3. This balanced distribution was used to evaluate whether oversampling could improve the ability of SVM and ANN to recognize minority sentiment classes. The testing set remained unchanged so that the evaluation results represented the model performance on real imbalanced TikTok comment data [13], [15].

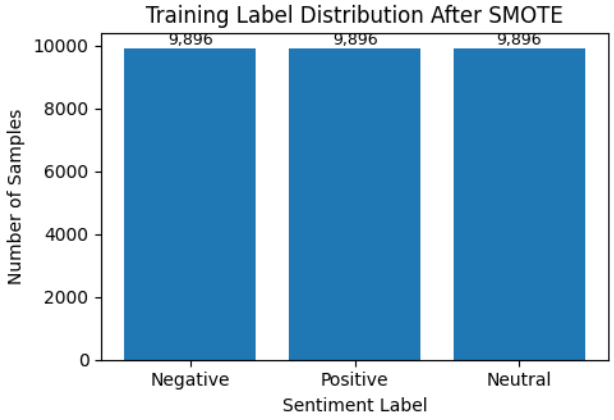

Figure 3. Training Label Distribution After SMOTE

Two classification models were used in this study, namely Support Vector Machine (SVM) and Artificial Neural Network (ANN). The SVM model was implemented using LinearSVC because TF-IDF produces high-dimensional sparse features that are generally suitable for linear separation. Before applying SMOTE, the SVM model used a balanced class weight to reduce the effect of class imbalance. After applying SMOTE, the model used the default class weight because the training data had been balanced. The main SVM configuration used a linear decision function with TF-IDF features as input.

The ANN model was designed as a feed-forward neural network using TF-IDF features as the input layer. The network consisted of two hidden layers with 128 and 64

neurons, ReLU activation functions, dropout of 0.3 to reduce overfitting, and a softmax output layer for three sentiment classes. The model was trained using categorical cross-entropy as the loss function and Adam optimizer. Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix [3], [9].

III. RESULTS AND DISCUSSION

The results are discussed based on class distribution, model performance before and after SMOTE, confusion matrix analysis, and overall model comparison. The sentiment distribution shows that the dataset was imbalanced, with negative sentiment dominating the data. This condition is important because imbalanced data can affect the model's ability to recognize minority classes, especially positive and neutral comments. Therefore, SMOTE was applied only to the training data to evaluate whether class balancing could improve model performance without changing the original testing distribution.

A. *Classification Results Before SMOTE*

Before applying SMOTE, the SVM model achieved an accuracy of 86.66%, precision of 86.91%, recall of 86.66%, and F1-score of 86.76%. The ANN model achieved an accuracy of 85.31%, precision of 85.07%, recall of 85.31%, and F1-score of 85.10%. These results indicate that SVM produced better overall performance than ANN on the original imbalanced training data. The classification performance before applying SMOTE is presented in Table VI.

TABLE VI
BEFORE SMOTE RESULTS

Model	Acc.	Prec.	Rec.	F1
SVM	0.8666	0.8691	0.8666	0.8676
ANN	0.8531	0.8507	0.8531	0.8510

The confusion matrix of SVM before SMOTE is shown in Figure 4.

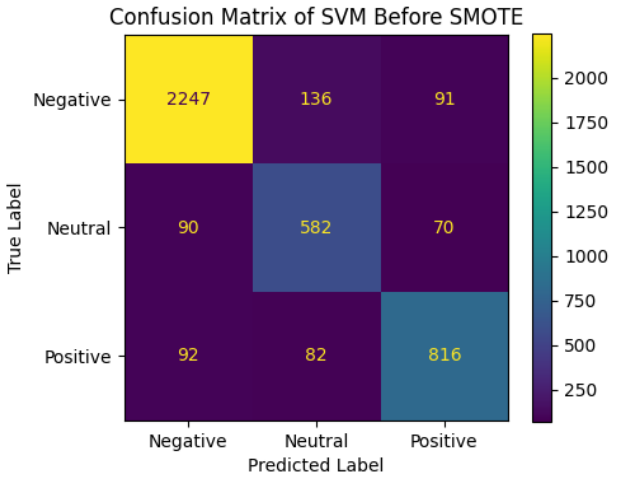

Figure 4. Confusion Matrix of SVM Before SMOTE

The confusion matrix of ANN before SMOTE is shown in Figure 5.

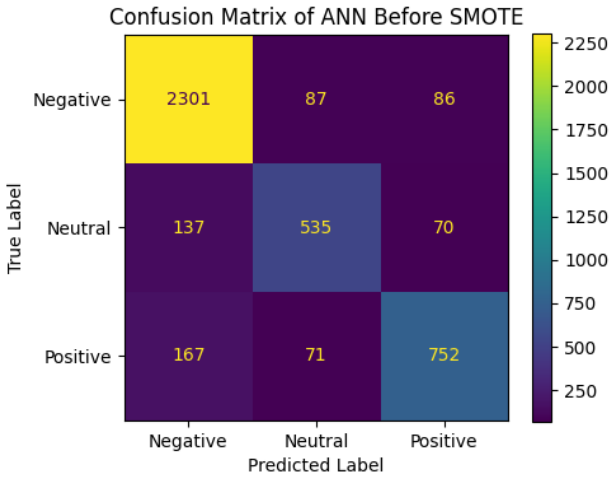

Figure 5. Confusion Matrix of ANN Before SMOTE

The SVM confusion matrix shows that the negative class was classified most accurately, with 2,247 correct predictions from 2,474 negative test instances. The neutral class was more difficult to distinguish because neutral comments often have ambiguous expressions and overlap with positive or negative words. ANN also performed well but showed lower recall for the positive class than SVM. This indicates that SVM is more stable for TF-IDF-based features on short TikTok comments.

B. *Classification Results After SMOTE*

After applying SMOTE to the training data, SVM achieved an accuracy of 86.40%, precision of 86.62%, recall of 86.40%, and F1-score of 86.49%. ANN achieved an accuracy of 85.47%, precision of 85.38%, recall of 85.47%, and F1-score of 85.42%. SMOTE slightly improved ANN performance but slightly reduced SVM performance. Therefore, SMOTE successfully balanced the training data, but its impact differed across algorithms.

TABLE VII
AFTER SMOTE RESULTS

Model	Acc.	Prec.	Rec.	F1
SVM+SMOTE	0.8640	0.8662	0.8640	0.8649
ANN+SMOTE	0.8547	0.8538	0.8547	0.8542

The confusion matrix of SVM after SMOTE is shown in Figure 6.

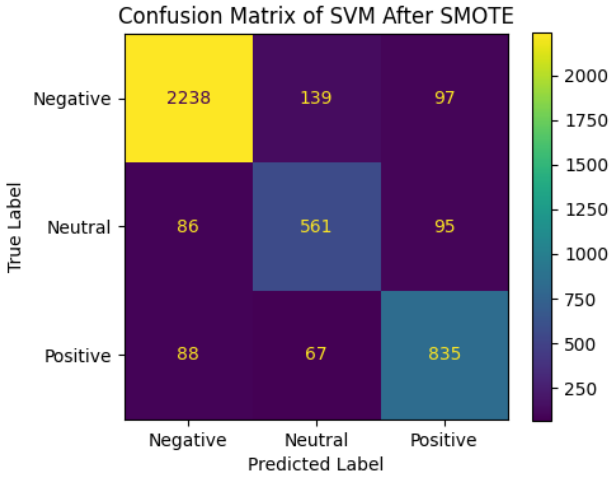

Figure 6. Confusion Matrix of SVM After SMOTE

The confusion matrix of ANN after SMOTE is shown in Figure 7.

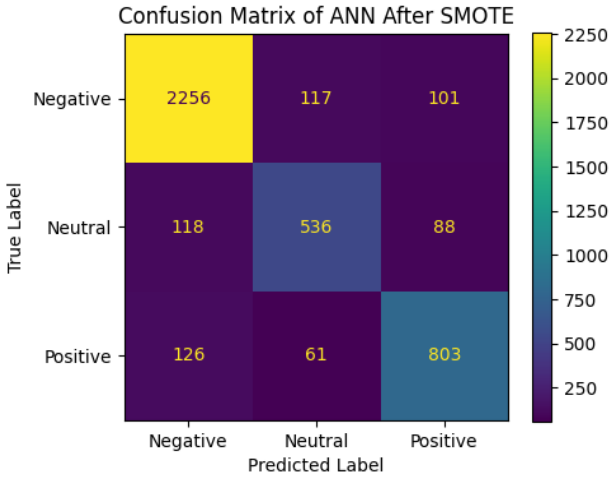

Figure 7. Confusion Matrix of ANN After SMOTE

Based on the confusion matrices, both models classified the negative class more accurately than the positive and neutral classes. This result is consistent with the class distribution, where negative comments formed the largest portion of the dataset. The neutral class was more difficult to classify because many neutral comments contained ambiguous expressions or lacked strong sentiment words. The positive class also showed several misclassifications because supportive comments sometimes used informal expressions that were not fully captured by the TF-IDF representation.

For SVM, the F1-score changed from 0.8676 before SMOTE to 0.8649 after SMOTE. This small decrease suggests that the original SVM with balanced class weight was already able to handle class imbalance effectively. For ANN, the F1-score increased from 0.8510 to 0.8542, showing that ANN benefited slightly from the balanced training distribution. These findings support the view that oversampling is useful for class balancing, but it does not always guarantee improvement for every model [15], [16].

C. Overall Model Comparison

The overall comparison of model performance before and after SMOTE is presented in Table VIII.

TABLE VIII
OVERALL COMPARISON

Model	Acc.	Prec.	Rec.	F1
SVM before	0.8666	0.8691	0.8666	0.8676
ANN before	0.8531	0.8507	0.8531	0.8510
SVM after	0.8640	0.8662	0.8640	0.8649
ANN after	0.8547	0.8538	0.8547	0.8542

The performance comparison of SVM and ANN before and after SMOTE is visualized in Figure 8.

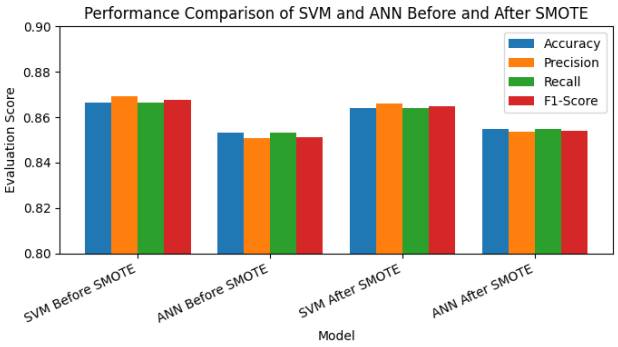

Figure 8. Performance Comparison of SVM and ANN Before and After SMOTE

The overall comparison shows that SVM before SMOTE achieved the highest F1-score of 86.76%. This result indicates that SVM is more suitable for the TF-IDF representation used in this study. TF-IDF produces high-dimensional and sparse features, and SVM works effectively in this condition because it searches for an optimal separating hyperplane between sentiment classes. In contrast, ANN has more trainable parameters and generally requires richer representations or larger training data to learn stable nonlinear patterns. When the input representation is sparse, ANN may not fully capture contextual relationships among words, especially in short TikTok comments. The ANN training curve also indicates a tendency toward overfitting because training accuracy increased faster than validation accuracy. The ANN accuracy curve before SMOTE is shown in Figure 9.

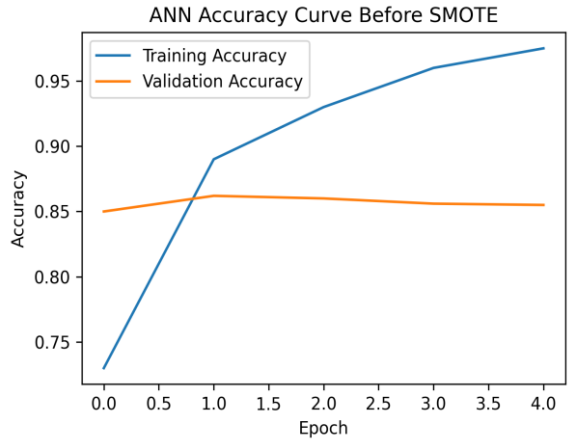

Figure 9. ANN Accuracy Curve Before SMOTE

The findings also show that lexicon-based labeling can be used to generate target labels for supervised classification, as performed in previous TikTok sentiment analysis research [1]. However, lexicon-based labeling has limitations in handling sarcasm, irony, informal expressions, and context-dependent meanings in social media comments. For example, a word with positive polarity may appear in a sarcastic negative sentence. This limitation indicates that lexicon-based labeling should be supported by manual validation, domain-specific lexicon enrichment, or contextual modeling in future studies.

This study also has limitations related to data scope and feature representation. The dataset was collected only from TikTok comments related to the Koperasi Merah Putih issue, so the findings cannot be directly generalized to other social media platforms or other public policy topics. In addition, this study used TF-IDF as the feature representation method. Although TF-IDF is lightweight, interpretable, and suitable for baseline comparison, it does not capture contextual word meanings as effectively as embedding-based or transformer-based models. Therefore, future research should compare TF-IDF-based SVM and ANN with contextual embedding approaches such as FastText, IndoBERT, or RoBERTa to examine whether ANN performance is limited by the classifier architecture or by the sparse TF-IDF representation [2], [17].

IV. CONCLUSION

This study compared SVM and ANN for classifying TikTok comments related to the Koperasi Merah Putih issue. The dataset consisted of 21,026 preprocessed and lexicon-labeled comments collected from TikTok during May to June 2025. Sentiment labels were generated automatically using a lexicon-based approach, while TF-IDF with unigram-bigram representation was used for feature extraction. The dataset was split using an 80:20 ratio, and model evaluation was strengthened using Stratified K-Fold Cross Validation, SMOTE, confusion matrix analysis, and a paired t-test.

The results show that SVM achieved the best overall performance before SMOTE, with an accuracy of 86.66% and an F1-score of 86.76%. The five-fold cross-validation results also showed that SVM obtained a higher mean score than ANN, and the paired t-test indicated that the difference was statistically significant. ANN achieved an accuracy of 85.31% before SMOTE and improved slightly after SMOTE to 85.47%. These results indicate that SVM is more stable than ANN for TF-IDF-based sentiment classification of short TikTok comments. SMOTE successfully balanced the training data but did not improve all models equally.

This study is limited to one social media platform, one public issue, and one feature representation method. In addition, lexicon-based labeling may still produce errors in comments containing sarcasm, irony, slang, or context-dependent meanings. Future research may improve labeling quality through larger manual annotation, aspect-based sentiment analysis, and comparison with contextual embedding or transformer-based models such as FastText, IndoBERT, and RoBERTa.

REFERENCES

[1] 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, 2026, [Online]. Available: https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/12395

[2] R. L. Mulianingrum, "Comparative Performance of SVM and BERT-Base Using TikTok Comments for Sentiment Analysis," *J. Appl. Informatics Comput.*, 2025, [Online]. Available: https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/11385

[3] S. R. Putri, A. Asrianda, and L. Rosnita, "Sentiment Analysis of Youtube and Gotube Reviews on Google Play Using the Support Vector Machine (SVM) Method in Indonesia," *J. Appl. Informatics Comput.*, vol. 9, no. 3, pp. 1025–1033, 2025, doi: 10.30871/jaic.v9i3.9461.

[4] B. R. Ermawan, "Optimizing Sentiment Classification Models for TikTok Comments Using Emotion-Based Preprocessing and Hyperparameter Optimization," *J. Appl. Informatics Comput.*, 2026, [Online]. Available: https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/11742

[5] A. P. Widyasari and Muljono, "Sentiment Analysis of YouTube Comments on the 2025 DPR RI Demonstration Using Machine Learning," *J. Appl. Informatics Comput.*, 2025, [Online]. Available: https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/12254

[6] J. Pardede, "Perbandingan Algoritma Stemming Porter, Sastrawi, Idris, dan Enhanced Confix Stripping pada Information Retrieval," *J. Teknol. Inf. dan Ilmu Komput.*, 2025, [Online]. Available: https://jtiik.ub.ac.id/index.php/jtiik/article/view/8860

[7] S. Febriani, V. Wati, Y. Wijayanti, and I. Siswanto, "Twitter Sentiment Analysis on Digital Payment in Indonesia Using Artificial Neural Network," *J. Appl. Informatics Comput.*, vol. 9, no. 2, 2025, [Online]. Available:

https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/8988

[8] A. Widodo, "Optimizing Support Vector Machine (SVM) for Sentiment Analysis Using TF-IDF, SMOTE, and Chi-Square," *J. Appl. Informatics Comput.*, 2025, [Online]. Available: https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/10541

[9] M. Ilmiyah, M. A. Barata, and P. E. Yuwita, "Implementation of ANN Optimization with SMOTE and Backward Elimination for PCOS Prediction," *Sci. J. Informatics*, 2025, [Online]. Available: https://scholar.google.com/citations?hl=en&user=iyoAaIEAAAAJ

[10] A. L. D. Ulhaq and S. Suprayogi, "Comparing Machine Learning Models for Sentiment Analysis of Tokopedia Reviews," *J. Appl. Informatics Comput.*, vol. 9, no. 6, 2025, [Online]. Available: https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/11239

[11] H. Barus, I. N. Fajri, and Y. Pristyanto, "Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models," *J. Appl. Informatics Comput.*, vol. 9, no. 5, 2025.

[12] F. Koto and G. Y. Rahmaningtyas, "InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs," in *in Proc. 21st International Conference on Asian Language Processing (IALP)*, 2017. [Online]. Available: https://github.com/fajri91/InSet

[13] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002, [Online]. Available: https://arxiv.org/abs/1106.1813

[14] D. Musfiroh, "Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, 2021, [Online]. Available: https://journal.irpi.or.id/index.php/malcom/article/download/20/11

[15] M. Mujahid, "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *J. Big Data*, vol. 11, no. 87, 2024, doi: 10.1186/s40537-024-00943-4.

[16] R. T. Aldisa and P. Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.

[17] F. Ansyah and R. R. Suryono, "Sentiment Classification of Indonesian-Language Roblox Reviews Using IndoBERT with SMOTE Optimization," *J. Appl. Informatics Comput.*, vol. 9, no. 4, 2025.