TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 4, August 2026, pp. 1157~1167
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i4.27419

NLP-driven hate speech detection on TikTok: a case study from UIN Sunan Ampel Surabaya

Achmad Teguh Wibowo[1], Aris Fanani[1], Mujib Ridwan[1], Bramasta Kurnia Aji[2]
[1]Faculty of Science and Technology, UIN Sunan Ampel Surabaya, Surabaya, Indonesia
[2]Information Technology Center and Database, UIN Sunan Ampel Surabaya, Surabaya, Indonesia

Article Info

Article history:

Received Jul 21, 2025
Revised Apr 10, 2026
Accepted May 25, 2026

Keywords:

Hate speech detection
Lexicon-based sentiment analysis
Natural language processing
Sentiment analysis
TikTok

ABSTRACT

This study examines hate speech detection in TikTok comments using natural language processing (NLP) techniques within the student community of UIN Sunan Ampel Surabaya. A dataset of 10,000 comments associated with the hashtag #PBAKUINSA2023 was analyzed using a lexicon-based sentiment analysis approach implemented through the TextBlob library, combined with Indonesian text preprocessing techniques, including tokenization, normalization, stopword removal, and stemming using the Sastrawi library. The results indicate that the proposed approach achieved an accuracy of 0.85, with precision of 0.88, recall of 0.83, and an F1-score of 0.854. Most comments were classified as neutral, while 31.8% were positive, and only a small proportion were negative. These findings suggest that discussions related to campus activities tend to be neutral or supportive. However, the findings also reveal that sentiment polarity does not always directly correspond to hate speech, as certain harmful expressions may appear neutral in lexicon-based analysis. This limitation highlights the need for more context-aware approaches. Overall, the proposed method provides an efficient solution for monitoring online discourse in academic environments.

Corresponding Author:

Bramasta Kurnia Aji
Information Technology Center and Database, UIN Sunan Ampel Surabaya
Ahmad Yani Street No. 117, Jemur Wonosari, Wonocolo District, Surabaya City, Indonesia
Email: ajibramasta@uinsa.ac.id

1. INTRODUCTION

The rapid growth of social media has significantly transformed communication and generated large volumes of unstructured data that are difficult to analyze and manage. Platforms such as TikTok have become widely used communication channels, particularly among students at UIN Sunan Ampel Surabaya. Along with this growth, social media has also facilitated the spread of harmful content, including hate speech, which may negatively affect online interactions and digital environments [1]. Therefore, systematic approaches are required to effectively process and analyze social media data in order to detect potentially harmful content [2], [3].

Social media has also changed the role of users from passive consumers of information to active producers who continuously generate digital content and interactions [4]. This shift results in increasingly complex data that requires modern analytical techniques for effective analysis [5]. In this context, natural language processing (NLP) provides powerful tools for extracting meaningful insights from large-scale textual data generated through social media platforms [6], [7].

Recent studies highlight that NLP techniques, including sentiment analysis approaches, have been widely applied to detect hate speech and analyze sentiment patterns in online communication [8], [9].

Sentiment analysis enables the classification of textual data into positive, neutral, or negative sentiments, which can help identify hostile or offensive language targeting individuals or groups based on attributes such as race, religion, gender, or nationality [10]. In general, sentiment analysis approaches can be categorized into lexicon-based methods and machine learning-based methods [11]. Lexicon-based approaches rely on predefined sentiment dictionaries, while machine learning approaches learn sentiment patterns from labeled datasets [12], [13].

This study applies NLP techniques to analyze TikTok comments in order to identify hate speech and sentiment patterns within the academic community of UIN Sunan Ampel Surabaya [14]. By utilizing a lexicon-based sentiment analysis approach combined with Indonesian text preprocessing techniques using the sastrawi library, the research develops an automated system capable of classifying sentiment in social media comments [15]. The system is implemented through a prototype application designed to monitor and analyze discussions related to campus activities on TikTok [16].

Although numerous studies have investigated hate speech detection on social media platforms, limited research has specifically examined TikTok interactions within Indonesian academic communities. This study addresses this gap by analyzing TikTok comments associated with campus-related hashtags and implementing a sentiment analysis system for automated monitoring. The findings are expected to provide insights into student interaction patterns on social media and support efforts to promote safer and more responsible digital communication within academic environments.

2. METHOD

This study applies NLP techniques to analyze TikTok comments and identify sentiment patterns related to potential hate speech within the academic community of UIN Sunan Ampel Surabaya. The proposed approach follows a structured workflow consisting of data collection, text preprocessing, sentiment analysis, classification, and performance evaluation. The overall pipeline of the proposed system is illustrated in Figure 1.

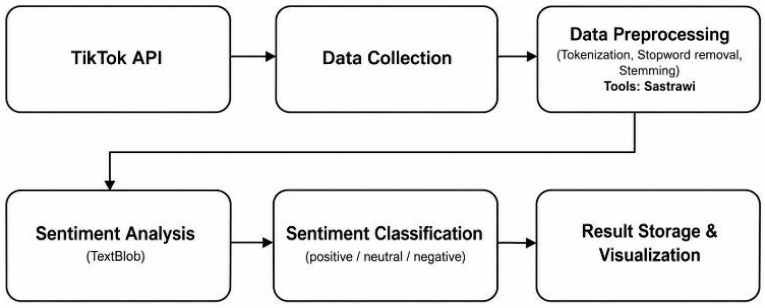


Figure 1. Overall workflow of the proposed TikTok sentiment analysis system, including data collection, preprocessing, sentiment analysis using TextBlob, and result storage and visualization

2.1. Data collection

The dataset used in this study consists of 10,000 TikTok comments collected from videos associated with the hashtag #PBAKUINSA2023, which relates to campus orientation activities at UIN Sunan Ampel Surabaya. The data were retrieved using the TikTokApi Python library, focusing on publicly available comments posted by users on videos containing the selected hashtag. The collected comments are primarily written in Indonesian and often contain informal expressions, abbreviations, slang, and mixed-language usage commonly found in social media communication. These characteristics make social media text particularly challenging for automated analysis. From the collected dataset, 8,000 comments were used for training purposes and 2,000 comments were used for testing in order to evaluate the performance of the sentiment classification system. This data partition allows the proposed approach to be assessed on unseen data and ensures a more reliable evaluation of classification performance.

2.2. Data preprocessing

Prior to performing sentiment analysis, the textual data undergo several preprocessing stages in order to improve data quality and reduce noise in the dataset. Text obtained from social media platforms often contains irregular sentence structures, informal expressions, abbreviations, slang, and unnecessary

symbols that may negatively affect the accuracy of sentiment analysis. Therefore, several preprocessing techniques are applied to standardize the textual data before analysis [17].

These steps include tokenization, which splits sentences into individual tokens or words; stopword removal, which eliminates common words that do not significantly contribute to sentiment interpretation; normalization, which converts informal or abbreviated words into their standard forms; and stemming, which reduces words to their base or root form. In this study, stemming is performed using the Sastrawi library, which is specifically designed for Indonesian language processing.

2.3. Sentiment analysis

Sentiment classification in this study is performed using the TextBlob library, which implements a lexicon-based sentiment analysis approach. In this method, each word in the text is assigned a sentiment score based on a predefined sentiment lexicon. Negative sentiment is considered a potential indicator of hate speech-related expressions within the analyzed comments. TextBlob computes two primary sentiment indicators for each text input, namely polarity and subjectivity scores [18]. The polarity score represents the sentiment orientation of a text and ranges between −1 and +1. A polarity value greater than zero indicates positive sentiment, a value equal to zero represents neutral sentiment, and a value less than zero indicates negative sentiment [19]. Mathematically, the polarity value P of a sentence can be represented as the average sentiment score of all words contained in the text, expressed as:

$$P = \frac{\sum_{i=1}^{n} s_i}{n} \quad (1)$$

where:

- P represents the polarity score of a sentence, indicating the overall sentiment orientation.
- $\sum_{i=1}^{n}$ denotes the summation operator, which adds the sentiment scores of all words in the sentence.
- s_i represents the sentiment score of the i -th word in the text.
- n denotes the total number of words in the sentence.

In addition to polarity, TextBlob also calculates a subjectivity score S, which measures the degree of personal opinion contained in the text. The subjectivity score ranges between 0 and 1, where values closer to 1 indicate highly subjective expressions, while values closer to 0 indicate more objective statements.

To evaluate the performance of the proposed sentiment analysis system, the model was tested using 2,000 TikTok comments that were not included in the training dataset. The evaluation aims to measure the effectiveness of the sentiment classification approach in identifying sentiment polarity in social media comments. Several standard performance metrics were used to assess the classification results, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive evaluation of the model's ability to correctly classify sentiment polarity and detect hate speech-related expressions within the dataset. Accuracy measures the proportion of correctly classified instances among all evaluated samples [20]. It reflects the overall correctness of the classification system and is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

where true positive (TP) represents correctly classified positive instances, true negative (TN) represents correctly classified negative instances, false positive (FP) represents instances incorrectly classified as positive, and false negative (FN) represents instances incorrectly classified as negative. Precision measures the proportion of correctly predicted positive instances among all predicted positive instances. It indicates the reliability of positive sentiment predictions and is defined as:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall measures the proportion of actual positive instances that are correctly identified by the system. It evaluates the model's ability to detect relevant positive cases and is defined as:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

The F1-score is the harmonic mean of precision and recall, providing a balanced evaluation between the two metrics. It is expressed as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (5)$$

In addition to these metrics, a confusion matrix was used to provide a more detailed analysis of the classification performance. The confusion matrix presents the distribution of correctly and incorrectly classified instances by showing the number of TP, TN, FP, and FN [21]. This representation allows for a deeper understanding of the model's strengths and weaknesses, particularly in identifying misclassification patterns. By analyzing the confusion matrix, it becomes possible to observe how well the model distinguishes between hate speech and non-hate speech comments and to identify potential sources of classification errors.

2.4. Sentiment classification

After obtaining the polarity score from the sentiment analysis process, each comment is classified into one of three sentiment categories: positive, neutral, or negative. The classification is determined based on the polarity value obtained from the TextBlob analysis. The sentiment classification function can be defined as:

$$C(x) = \begin{Bmatrix} Positive, P > 0 \\ Neutral. P = 0 \\ Negative, P < 0 \end{Bmatrix} \quad (6)$$

where P represents the polarity score of the analyzed comment.

Through this classification mechanism, each TikTok comment is assigned a sentiment label. The classification results are then aggregated to analyze the distribution of sentiments within discussions related to the hashtag #PBAKUINSA2023. This enables the identification of sentiment patterns within the dataset.

2.5. Result storage and visualization

After the sentiment classification process is completed, the results are stored and visualized to facilitate interpretation of the sentiment distribution. Each analyzed comment is stored in a CSV file containing the comment text, polarity score, subjectivity score, and sentiment classification label. To present the sentiment distribution in a more interpretable form, the results are visualized using graphical representations. In this study, a donut chart generated using the Matplotlib library is used to display the proportion of positive, neutral, and negative comments associated with the analyzed hashtag.

This visualization helps provide a clear overview of sentiment trends in TikTok discussions and enables easier interpretation of the results by researchers and stakeholders. Furthermore, graphical visualization allows the identification of dominant sentiment patterns and potential indicators of harmful or negative online interactions within the academic digital environment.

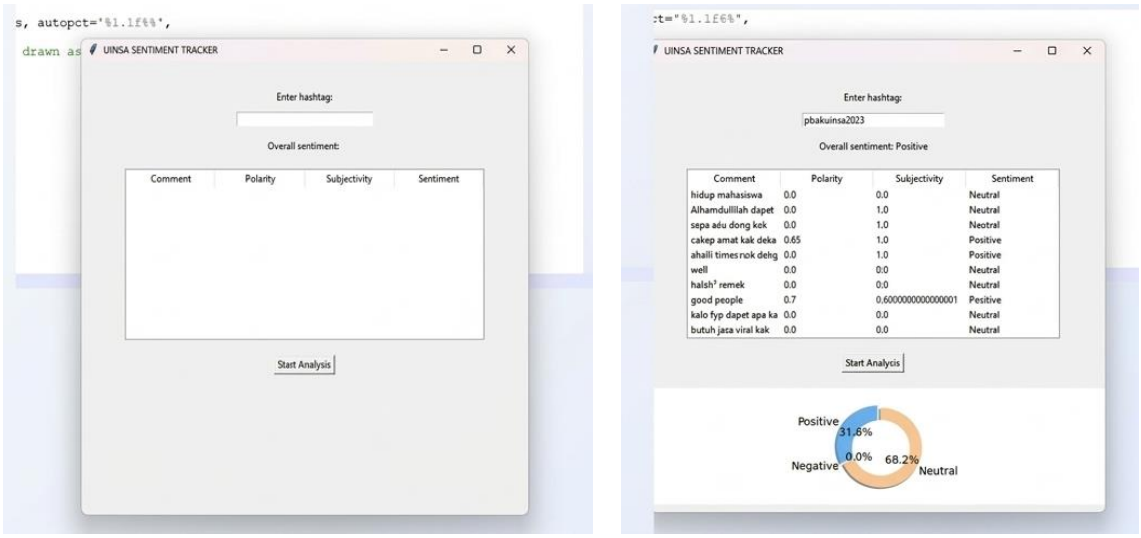
3. RESULTS AND DISCUSSION

This section presents the results obtained from the implementation of the proposed sentiment analysis system and discusses the performance of the model in analyzing TikTok comments. The results include the application implementation, sentiment classification performance, sentiment analysis outcomes, and interpretation of sentiment patterns found in the dataset.

3.1. Application implementation and key features

The UINSA Sentiment Tracker application was developed to analyze sentiment from TikTok comments associated with specific hashtags. The application was implemented using the Python programming language with the Tkinter framework to provide a user-friendly graphical interface for social media sentiment analysis. Through this interface, users can input a hashtag to retrieve related TikTok videos and comments. The system then processes the collected comments, performs sentiment analysis, and presents the results in both numerical and graphical formats. Figure 2 illustrates the user interface of the UINSA Sentiment Tracker application, including the main dashboard as in Figure 2(a) and the sentiment analysis processing interface as in Figure 2(b).

The application utilizes the TextBlob library to perform lexicon-based sentiment analysis, while the Sastrawi library is used for Indonesian text preprocessing, particularly for stemming operations. Comments extracted from TikTok videos are first processed through the preprocessing pipeline before sentiment polarity and subjectivity scores are calculated. The processed data are subsequently stored in a CSV file for further analysis and documentation. To facilitate interpretation of the results, the sentiment distribution is visualized using a donut chart generated with the Matplotlib library, which displays the proportion of positive, neutral, and negative sentiments. The Tkinter-based interface provides an intuitive layout that allows users to easily interact with the system and observe the sentiment analysis results.



(a) (b)

Figure 2. User interface of the proposed system: (a) application dashboard and (b) sentiment analysis processing interface

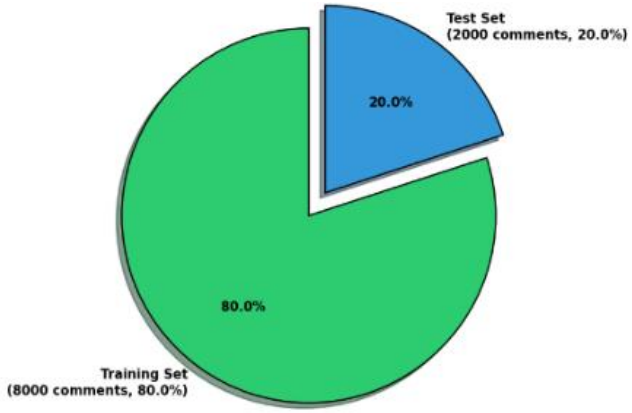
3.2. Algorithm components

The developed application integrates several components to perform automated sentiment analysis on TikTok comments. The system utilizes Tkinter to provide a graphical user interface (GUI), allowing users to input hashtags and view analysis results. TikTok comments associated with the specified hashtag are retrieved using the TikTokApi Python library. After the comments are collected, the system performs text preprocessing including cleaning and stemming using the Sastrawi library to standardize Indonesian-language text. Sentiment analysis is then conducted using the TextBlob library, which calculates polarity and subjectivity scores for each comment. Based on the polarity value, comments are classified into positive, neutral, or negative sentiment categories. The analysis results are stored in a CSV file and visualized using a donut chart generated with the Matplotlib library. This visualization helps illustrate the distribution of sentiment categories within the analyzed dataset and enables users to easily interpret the sentiment patterns of TikTok comments.

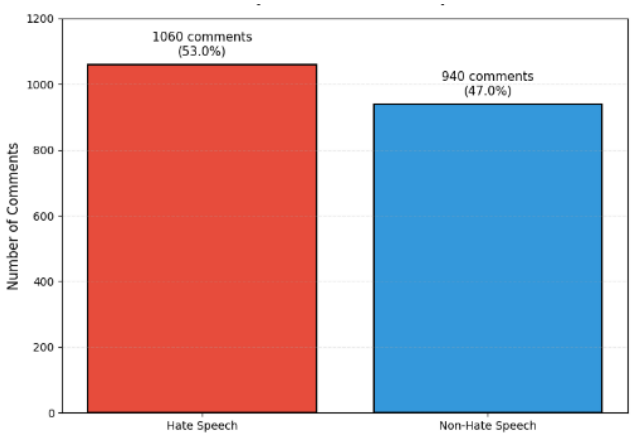
3.3. Application testing

To evaluate the performance of the proposed NLP-based approach, several evaluation metrics were used, including accuracy, precision, recall, F1-score, and confusion matrix analysis. The dataset consisted of 10,000 TikTok comments, with 8,000 comments used for training and 2,000 comments used for testing. The model achieved an accuracy of 0.85, with precision of 0.88, recall of 0.83, and an F1-score of 0.854, indicating effective performance in classifying sentiment in informal social media text. This performance can be attributed to the preprocessing techniques applied, which help normalize slang, abbreviations, and non-standard expressions commonly found in TikTok comments. Further analysis of the classification results is presented using confusion matrix and precision–recall evaluation. The distribution of the dataset used for model training and evaluation is illustrated in Figure 3.

Figure 3(a) shows the distribution of data for training and testing, while Figure 3(b) presents the class distribution within the testing dataset, illustrating the proportion of hate speech and non-hate speech comments used for model evaluation. The test dataset contains 1,060 hate speech comments (53%) and 940 non-hate speech comments (47%), indicating that the dataset is relatively balanced. A balanced dataset is important for ensuring that the classification model does not become biased toward one particular class and can effectively detect both hate speech and non-hate speech instances. This data split ensures that the model can be trained effectively and evaluated using unseen data to measure its generalization capability. Figure 4 shows the visualization of the evaluation metrics of the hate speech classification model based on the selected performance metrics, including accuracy, precision, recall, and F1-score.



(a)



(b)

Figure 3. Dataset distribution used for model development: (a) distribution of data for training and testing and (b) class distribution in the testing dataset

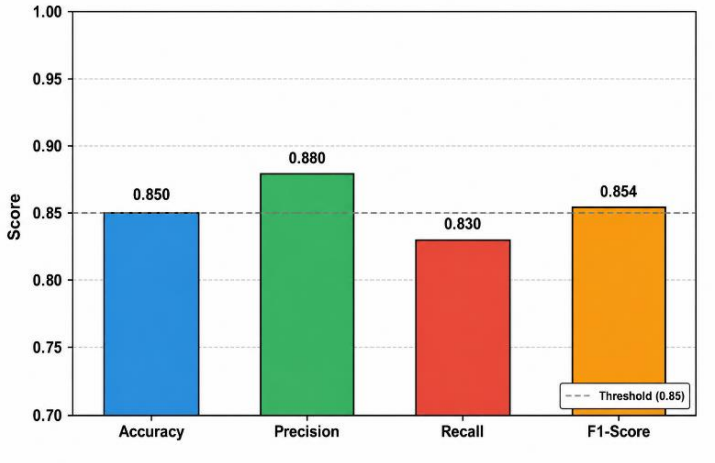


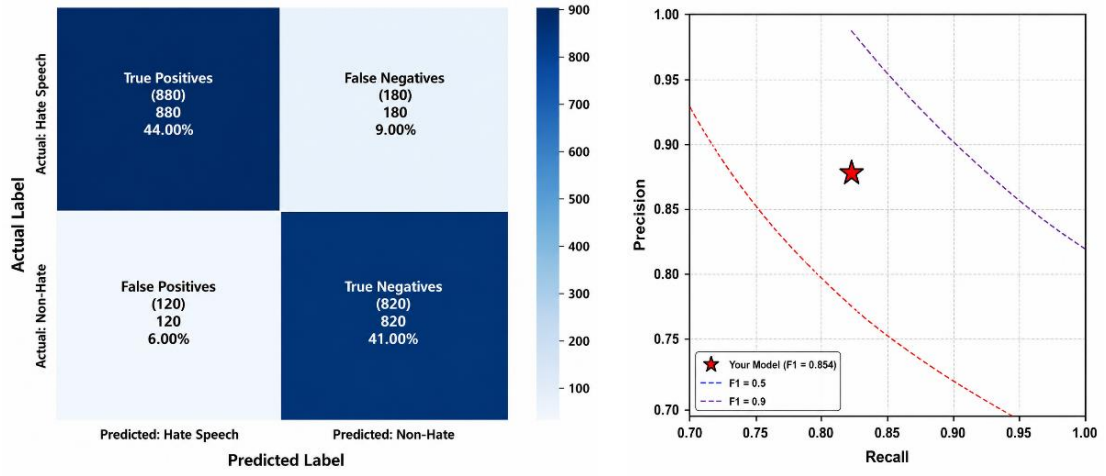
Figure 4. Visualization metrics of hate speech classification model

To further analyze the classification performance, Figure 5 presents the evaluation results of the proposed hate speech classification model using two complementary visualizations. Figure 5(a) illustrates the confusion matrix of the classification results, while Figure 5(b) shows the precision–recall trade-off analysis of the model. Figure 5(a) presents the confusion matrix summarizing the classification results in terms of TP,

TN, FP, and FN. The model correctly identified 880 hate speech comments as TP and 820 non-hate speech comments as TN. However, 180 hate speech comments were misclassified as non-hate speech or FN, while 120 non-hate speech comments were incorrectly classified as hate speech or FP. These results indicate that the proposed model is capable of correctly identifying the majority of instances in the dataset. The occurrence of false negatives suggests that certain hate speech expressions may be subtle or context-dependent, while false positives may arise when neutral comments contain words interpreted as negative by the sentiment lexicon.

Figure 5(b) illustrates the precision–recall trade-off of the classification model, indicating a balanced performance between detecting hate speech and minimizing classification errors. These results confirm the effectiveness of the proposed approach in identifying hate speech in TikTok comments. Further error analysis based on the confusion matrix shows that the model produced 300 misclassified comments (15% of the test dataset), consisting of 180 false negatives and 120 false positives, as illustrated in Figure 6.

The higher number of false negatives indicates that certain hate speech expressions were not detected by the system. This limitation may occur due to the complexity of social media language, including informal expressions, slang, sarcasm, or contextual meanings that are difficult to capture using lexicon-based sentiment analysis methods. Meanwhile, false positives occur when neutral or informal comments contain words that are interpreted as negative by the sentiment lexicon.



(a) (b)

Figure 5. Performance evaluation of the proposed hate speech classification model: (a) confusion matrix showing the classification results of the proposed model and (b) precision–recall trade-off analysis of the proposed hate speech classification model

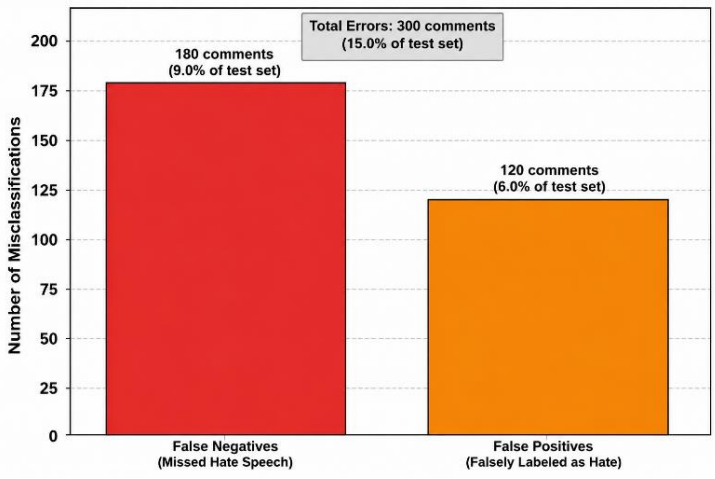


Figure 6. Error analysis misclassification breakdown

3.4. Sentiment analysis result

Table 1 presents a sample of TikTok comments along with their corresponding polarity scores, subjectivity values, and sentiment classification results. The Table 1 illustrates how the proposed sentiment analysis system assigns polarity and subjectivity scores to each comment and subsequently classifies the sentiment into positive, neutral, or negative categories. Based on the overall analysis of the dataset, the majority of comments associated with the hashtag #PBAKUINSA2023 were classified as neutral, while 31.8% of the comments were identified as positive and only a small proportion of comments were classified as negative. The examples shown in Table 1 illustrate how the sentiment classification process works for individual comments. These results indicate that discussions related to the hashtag tend to generate neutral or supportive interactions among users. The predominance of neutral comments suggests that the conversation surrounding the hashtag is largely informational or conversational rather than emotionally polarized. This finding indicates that sentiment polarity alone does not always directly correspond to hate speech, as certain harmful expressions may appear neutral due to the limitations of lexicon-based sentiment analysis.

Table 1. Sample sentiment analysis results from TikTok comments associated with the hashtag #PBAKUINSA2023

Original comment	English translation	Polarity	Subjectivity	Sentiment
farabi1705 adek mu	Farabi1705, is that your younger sibling?	0.0	0.0	Neutral
Alhamdulillah dapet	Thank God, I got it	0.0	0.0	Neutral
sapa aku dong kak	Please say hi to me	0.0	0.0	Neutral
cakep amat kak deka	You look very beautiful, Deka	0.65	1.0	Positive
ahaiii timses e nde ngarep iku wkwkw	Haha, that's the campaign team standing in front	0.0	0.0	Neutral
welll	Well...	0.0	0.0	Neutral
halah² remek	That's just an exaggeration	0.0	0.0	Neutral
good people	Good people	0.7	0.6	Positive
kalo fyp dapet apa kak	What do you get if this reaches the for you page (FYP)?	0.0	0.0	Neutral
butuh jasa viral kak	Do you need a social media promotion service?	0.0	0.0	Neutral
Keamanan tik banyangan ga ketok blas	The security staff in the background are not visible at all	0.5	0.5	Positive
oh iki ibnu	Oh, this is Ibnu	0.0	0.0	Neutral
Gini kan enakkk	This is much better	0.7	0.6	Positive
zahwa_machrunnisa lah siti se fakultas sama kmu	Zahwa_Machrunnisa, Siti is from the same faculty as you	0.0	0.125	Neutral
semangat pbak	Keep up the spirit!	0.0	0.0	Neutral
MasPengggggg emng bowlehh	Wow, that's really impressive!	0.0	0.0	Neutral
izin post sg boleh kak	May I repost this on my story?	0.5	0.5	Positive
kiw kiw	Cheering/flirting expression ("woohoo!")	0.0	0.0	Neutral
lucuu	So cute!	0.25	1.0	Positive
ihhh sok centil banget aku	Haha, I'm being so playful/flirty	0.2	0.2	Positive

3.5. Discussion

The obtained results indicate that lexicon-based sentiment analysis is effective in capturing general sentiment trends in social media data but has limitations in detecting implicit or context-dependent hate speech expressions. The proposed approach achieved an accuracy of 85%, indicating that the model is able to correctly classify the majority of comments. Despite these promising results, several challenges were observed during the analysis of TikTok comments. Social media language often contains ambiguity, slang, and informal expressions, which can reduce the effectiveness of lexicon-based sentiment analysis methods [22]. In addition, hate speech may depend heavily on contextual interpretation, including sarcasm, cultural references, or implicit meanings that are difficult to capture using simple lexical approaches.

TextBlob provides an efficient and practical approach for prototype development by enabling the calculation of polarity and subjectivity scores for sentiment classification. However, lexicon-based approaches have limitations in capturing deeper linguistic nuances. Previous studies have shown that transformer-based models such as bidirectional encoder representations from transformers (BERT) can provide improved performance in sentiment and hate speech detection by capturing contextual relationships within text [23]–[27]. Visualization of sentiment distributions also supports better interpretation of social media data by illustrating sentiment patterns and interaction trends within the dataset [28]. These visualizations help researchers and stakeholders quickly understand the dynamics of user engagement on social media platforms.

Future research may focus on integrating advanced deep learning models, expanding datasets, and improving preprocessing techniques to better handle slang, sarcasm, and contextual language. In addition, multimodal analysis, which combines textual, visual, and audio information, may provide more comprehensive insights into social media communication dynamics [29]. Overall, the proposed system demonstrates the potential of NLP-based approaches for monitoring sentiment patterns in TikTok comments. Although the current system has certain limitations, the findings provide a useful foundation for developing automated tools that support digital safety and responsible social media engagement within academic communities [30].

4. CONCLUSION

This study developed a NLP-based system to analyze sentiment and detect potential hate speech patterns in TikTok comments associated with the hashtag #PBAKUINSA2023. The proposed system integrates TikTok data retrieval, Indonesian text preprocessing using the Sastrawi library, and sentiment analysis using TextBlob to automatically classify user comments. The evaluation results show that the model achieved an accuracy of 0.85, with precision of 0.88, recall of 0.83, and an F1-score of 0.854, indicating reliable performance in sentiment classification. The analysis also reveals that most comments were neutral, while 31.8% were positive, and only a small proportion of comments were classified as negative. Overall, the results demonstrate the potential of NLP-based approaches for analyzing sentiment patterns in social media discussions and supporting safer digital communication within academic communities. Future research may improve classification performance by incorporating advanced transformer-based models and expanding the dataset to better capture linguistic variations in social media content.

FUNDING INFORMATION

This research was supported by a research grant under the National Strategic Study Research Cluster from UIN Sunan Ampel Surabaya, which provided funding and facilities for this study.

REFERENCES

[1] E. Aïmeur, S. Amri, and G. Brassard, "Fake news, disinformation and misinformation in social media: a review," *Social Network Analysis and Mining*, vol. 13, no. 1, 2023, doi: 10.1007/s13278-023-01028-5.
[2] R. Levy, "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment," *American Economic Review*, vol. 111, no. 3, pp. 831–870, 2021, doi: 10.1257/AER.20191777.
[3] A. Ghermandi *et al.*, "Social media data for environmental sustainability: A critical review of opportunities, threats, and ethical use," *One Earth*, vol. 6, no. 3, pp. 236–250, 2023, doi: 10.1016/j.oneear.2023.02.008.
[4] A. Matamoros-Fernández and J. Farkas, "Racism, Hate Speech, and Social Media: A Systematic Review and Critique," *Television and New Media*, vol. 22, no. 2, pp. 205–224, 2021, doi: 10.1177/1527476420982230.
[5] M. Subramanian, V. Easwaramoorthy Sathiskumar, G. Deepalakshmi, J. Cho, and G. Manikandan, "A survey on hate speech detection and sentiment analysis using machine learning and deep learning models," *Alexandria Engineering Journal*, vol. 80, pp. 110–121, 2023, doi: 10.1016/j.aej.2023.08.038.
[6] S. Bazzaz Abkenar, M. Haghi Kashani, E. Mahdipour, and S. M. Jameii, "Big data analytics meets social media: A systematic review of techniques, open issues, and future directions," *Telematics and Informatics*, vol. 57, 2021, doi: 10.1016/j.tele.2020.101517.
[7] N. Hasanati, M. E. Ibrahim, Q. Aini, E. Rustamaji, Nurbojatmiko, and Zulfiandri, "Social Commerce Sentiment Analysis on YouTube Comments with Lexicon-Based and Long Short-Term Memory Methods," in *2024 12th International Conference on Cyber and IT Service Management, CITSM 2024*, 2024. doi: 10.1109/CITSM64103.2024.10775817.
[8] M. Mhalla, J. Yun, and A. Nasiri, "Video-Sharing Apps Business Models: TikTok Case Study," *International Journal of Innovation and Technology Management*, vol. 17, no. 07, Nov. 2020, doi: 10.1142/S0219877020500509.
[9] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text Mining in Big Data Analytics," *Big Data and Cognitive Computing*, vol. 4, no. 1, Jan. 2020, doi: 10.3390/bdcc4010001.
[10] G. K. Wadhwani, P. K. Varshney, A. Gupta, and S. Kumar, "Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models Using Twitter Data on Russia–Ukraine War," *SN Computer Science*, vol. 4, no. 4, 2023, doi: 10.1007/s42979-023-01790-5.
[11] Y. Qi and Z. Shabrina, "Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach," *Social Network Analysis and Mining*, vol. 13, no. 1, 2023, doi: 10.1007/s13278-023-01030-x.
[12] C. Montag, H. Yang, and J. D. Elhai, "On the Psychology of TikTok Use: A First Glimpse From Empirical Findings," *Frontiers in Public Health*, vol. 9, 2021, doi: 10.3389/fpubh.2021.641673.
[13] S. Xu *et al.*, "Rumor detection on social media using hierarchically aggregated feature via graph neural networks," *Applied Intelligence*, vol. 53, no. 3, pp. 3136–3149, 2023, doi: 10.1007/s10489-022-03592-3.
[14] J. Cha, S. Kim, and E. Park, "A lexicon-based approach to examine depression detection in social media: the case of Twitter and university community," *Humanities and Social Sciences Communications*, vol. 9, no. 1, 2022, doi: 10.1057/s41599-022-01313-2.
[15] Y. R. Prayogi, A. R. Barakbah, A. A. Ripai, and A. Fariza, "Automatic Translation of Indonesian Sign Language Based on Multimodal Technology Recognition," in *2025 International Electronics Symposium, IES 2025*, 2025, pp. 922–927. doi: 10.1109/IES67184.2025.11161638.
[16] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics (Switzerland)*, vol. 9, no. 3, 2020, doi: 10.3390/electronics9030483.

[17] D. A. Kristiyanti, S. A. Sanjaya, V. C. Tjokro, and J. Suhali, "Dealing imbalance dataset problem in sentiment analysis of recession in Indonesia," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 2, pp. 2058–2070, 2024, doi: 10.11591/ijai.v13.i2.pp2060-2072.
[18] M. N. Maria, T. Kabir, S. Akter, and R. Khan, "Sentiment Analysis using NLP Libraries and Machine Learning," in *8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2024 - Proceedings*, 2024, pp. 911–916. doi: 10.1109/I-SMAC61858.2024.10714659.
[19] Q. Liu and H. Son, "Methods for aggregating investor sentiment from social media," *Humanities and Social Sciences Communications*, vol. 11, no. 1, 2024, doi: 10.1057/s41599-024-03434-2.
[20] T. Anderson, S. Sarkar, and R. Kelley, "Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques," *Natural Language Processing Journal*, vol. 8, 2024, doi: 10.1016/j.nlp.2024.100097.
[21] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, 2024, doi: 10.53555/ajbr.v27i4s.4345.
[22] B. Tahayna, R. K. Ayyasamy, R. Akbar, N. F. B. Subri, and A. Sangodiah, "Lexicon-based Non-Compositional Multiword Augmentation Enriching Tweet Sentiment Analysis," in *2022 3rd International Conference on Artificial Intelligence and Data Sciences: Championing Innovations in Artificial Intelligence and Data Sciences for Sustainable Future, AiDAS 2022 - Proceedings*, 2022, pp. 19–24. doi: 10.1109/AiDAS56890.2022.9918749.
[23] B. Mathew, P. Saha, S. M. Yimam, C. Biemann, P. Goyal, and A. Mukherjee, "HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection," in *35th AAAI Conference on Artificial Intelligence, AAAI 2021*, 2021, vol. 17A, pp. 14867–14875. doi: 10.1609/aaai.v35i17.17745.
[24] R. Hernandez Urbano, J. Uy Ajero, A. Legaspi Angeles, M. N. Hacar Quintos, J. M. Regalado Imperial, and R. Llabanes Rodriguez, "A BERT-based Hate Speech Classifier from Transcribed Online Short-Form Videos," *ACM International Conference Proceeding Series*, pp. 186–192, 2021, doi: 10.1145/3485768.3485806.
[25] R. Raut and F. Spezzano, "Enhancing hate speech detection with user characteristics," *International Journal of Data Science and Analytics*, vol. 18, no. 4, pp. 445–455, 2024, doi: 10.1007/s41060-023-00437-1.
[26] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. Ali Almazroi, "A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification," *Journal of Healthcare Engineering*, vol. 2022, 2022, doi: 10.1155/2022/3498123.
[27] M. Singh, A. K. Jakhar, and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," *Social Network Analysis and Mining*, vol. 11, no. 1, 2021, doi: 10.1007/s13278-021-00737-z.
[28] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Systems with Applications*, vol. 77, pp. 236–246, 2017, doi: 10.1016/j.eswa.2017.02.002.
[29] H. Beyari, "The Relationship between Social Media and the Increase in Mental Health Problems," *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, 2023, doi: 10.3390/ijerph20032383.
[30] S. A. Habibi and L. Salim, "Static vs. dynamic methods of delivery for science communication: A critical analysis of user engagement with science on social media," *PLoS ONE*, vol. 16, no. 3 March 2021, 2021, doi: 10.1371/journal.pone.0248507.

BIOGRAPHIES OF AUTHORS

Achmad Teguh Wibowo is an assistant professor in the Department of Information Systems, Faculty of Science and Technology, UIN Sunan Ampel Surabaya, Indonesia. He is also involved in managing information technology infrastructure at the university and has contributed to several research projects related to digital transformation and information systems development. His research interests include computer vision, deep learning algorithms, blockchain technology, and information systems. He can be contacted at email: atw@uinsa.ac.id.

Aris Fanani is a lecturer in the Department of Mathematics, Faculty of Science and Technology, UIN Sunan Ampel Surabaya, Indonesia. His research focuses on computational and applied mathematics, particularly in digital image processing, information retrieval, and data mining. He has contributed to several scientific publications related to mathematical modeling and data analysis. He can be contacted at email: arisfa@uinsa.ac.id.

Mujib Ridwan is an assistant professor in the Department of Information Systems, Faculty of Science and Technology, UIN Sunan Ampel Surabaya, Indonesia. His research interests include smart information systems, educational data mining, artificial intelligence applications, and mobile-based surveillance technologies. He has actively contributed to research related to intelligent systems and digital innovation in education. He can be contacted at email: mujibrw@uinsa.ac.id.

Bramasta Kurnia Aji works at the Information Technology Center and Database of UIN Sunan Ampel Surabaya, Indonesia. His research interests include artificial intelligence, natural language processing, data analytics, and digital infrastructure management. His work focuses on developing technology-driven solutions to support digital transformation and information systems within higher education environments. He can be contacted at email: ajibramasta@uinsa.ac.id.