GENERATION JOURNAL
Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

Sentiment Analysis of The Incident of The Downing of an Indian Rafale Fighter Jet by a Pakistani J-10CE Fighter Jet Using a Deep Learning Model

Arifathur al Hafidz*[1], Estu Sinduningrum[2]
[1,2] Universitas Muhammadiyah Prof. Dr. Hamka, Jalan Tanah Merdeka No.6, Jakarta Timur, Indonesia
Email: fathurrrzx@gmail.com, estu.ningrum@uhamka.ac.id
*corresponding author : fathurrrzx@gmail.com

Article Info

Received	2026-01-01
Revised	2026-01-31
Accepted	2026-02-02

Keywords:* *Sentiment Analysis, YouTube, India–Pakistan Conflict, LSTM, Deep Learning

Abstract

The rapid growth of digital technology and social media has significantly influenced the dissemination of information and public opinion worldwide. YouTube, as one of the largest social media platforms, is widely used by users to express opinions on international issues, including geopolitical conflicts. One event that attracted substantial public attention was the reported downing of an Indian Rafale fighter jet by a Pakistani J-10CE within the context of the India–Pakistan conflict. This study aims to analyze public sentiment expressed in YouTube comments related to this incident using a Deep Learning approach based on the Long Short-Term Memory (LSTM) algorithm.

A total of 1,336 English-language YouTube comments were collected using the YouTube Data API v3. The data were automatically labeled into three sentiment categories: positive (38.32%), negative (31.21%), and neutral (30.46%). The research process includes text preprocessing, sentiment labeling using VADER, LSTM model training, and performance evaluation using accuracy, precision, recall, and F1-score metrics. Experimental results show that the proposed model achieved an accuracy of 61% with a macro-averaged F1 score of 0.61 on the test set. These findings indicate that the model provides moderate and stable performance in analyzing sentiment within conflict-driven geopolitical discussions on social media.

INTRODUCTION

The rapid development of digital technology and social media platforms has significantly transformed how individuals express opinions and respond to global events. Among these platforms, YouTube functions not only as a video-sharing service but also as a public forum where users actively engage in discussions related to political, social, and international issues. Previous research has demonstrated that YouTube comment sections reflect diverse public perceptions toward public policy and social issues [1]. Deep learning approaches for text classification began gaining attention with the introduction of Convolutional Neural Networks (CNN) for sentence classification [2]. In parallel, foundational theoretical frameworks in sentiment analysis were established, defining polarity detection and feature-based opinion mining as core tasks in opinion mining research [3]. Social media platforms were later validated as rich corpora for sentiment analysis, although their informal and noisy linguistic characteristics increase classification complexity [4]. Lexicon-based sentiment analysis methods such as VADER were developed to efficiently analyze social media text by incorporating rule-based heuristics for polarity scoring [5]. Subsequently, transformer-based contextual models such as BERT demonstrated superior bidirectional contextual understanding in various NLP tasks [6].

Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

Recurrent neural networks further advanced contextual modeling. Neural document modeling using gated recurrent architectures showed improved sentiment classification by preserving semantic dependencies [7]. Long Short-Term Memory (LSTM), introduced to address the vanishing gradient problem, enables effective modeling of long-range dependencies in sequential text [8]. A comprehensive survey confirmed that neural architectures consistently outperform traditional machine learning models in complex sentiment analysis tasks [9]. Empirical applications of sentiment analysis on YouTube have been conducted in various domains. YouTube has been validated as a relevant platform for public opinion mining in technology-related discourse [10]. Comparative studies demonstrated that LSTM outperforms SVM and Naïve Bayes in sentiment classification tasks [11]. Further empirical evidence showed that applying LSTM to YouTube discussions improves classification performance compared to classical machine learning approaches [12]. Other studies combined LSTM with Word2Vec embeddings to enhance contextual representation in public policy sentiment analysis [13]. Comparative evaluation between VADER and contextual models confirmed that transformer-based architectures achieve higher accuracy, although lexicon-based approaches remain computationally efficient for large-scale labeling [14]. Additionally, VADER-based sentiment analysis has been applied to YouTube discussions in environmental contexts [15].

Recent Indonesian studies strengthened the application of deep learning in politically sensitive and geopolitical discourse. Bidirectional LSTM improved macro-F1 performance in political sentiment analysis on Indonesian YouTube channels [16], while hybrid deep learning architectures integrating embeddings and recurrent layers enhanced classification stability in geopolitical sentiment analysis tasks [17]. Despite these methodological advancements, most existing studies focus on entertainment, domestic political discourse, or general public policy issues [12], [13], [16]. Although previous studies demonstrated the effectiveness of LSTM and transformer-based models, most research contexts remain limited to entertainment or domestic political discourse. These contexts differ significantly from international military conflict discussions, which are often emotionally polarized and geopolitically sensitive. Consequently, model behavior in such high-tension discourse environments remains insufficiently explored.

Empirical investigation of public sentiment toward international military conflicts remains limited [17]. Therefore, this study focuses on analyzing sentiment toward the reported downing of an Indian Rafale fighter jet by a Pakistani J-10CE within the India–Pakistan conflict. The novelty of this study lies in its integration of automatic pseudo-labeling with sequential deep learning modeling in the context of a real-world international air combat incident. Unlike prior research focusing on domestic or entertainment contexts, this study evaluates model robustness in geopolitically sensitive discourse environments. This study differentiates itself by integrating VADER-based pseudo-labeling [5] with supervised LSTM modeling [8], [12] and providing direct comparison with a Naïve Bayes baseline [1], [11].

RESEARCH METHOD

A. Theory Study

The basic science and supporting theories in this study include:

1. Theory of Sentiment Analysis

   Sentiment analysis is a subfield of NLP that focuses on identifying and classifying subjective information expressed in textual data [3]. Social media platforms serve as valuable corpora for sentiment analysis but present challenges due to informal and noisy language [4].

2. Classical Machine Learning and CNN-Based Approaches

   Traditional sentiment classification relies on feature engineering techniques such as bag-of-words and TF-IDF combined with classifiers like Naïve Bayes and SVM [1]. CNN-based text classification was introduced to capture local textual features [2]. Comparative studies showed that LSTM outperforms classical models due to its sequential modeling capability [11].

GENERATION JOURNAL

Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

3. Lexicon-Based Sentiment Theory (VADER)

VADER was introduced as a rule-based sentiment analysis tool optimized for social media text [5]. Comparative studies between VADER and transformer-based models such as BERT demonstrated that contextual architectures achieve higher accuracy [6], [14]. VADER has been applied in large-scale YouTube sentiment labeling [15].

4. Word Embedding and Neural Document Modeling

Gated recurrent neural networks improve document-level sentiment classification by preserving contextual semantics [7]. Deep learning surveys emphasize the superiority of embedding-based models over classical representations [9].

5. Long Short-Term Memory (LSTM)

LSTM addresses the vanishing gradient problem and enables long-term dependency modeling [8]. Empirical studies applying LSTM to YouTube sentiment analysis demonstrated improved classification performance [12]. YouTube has also been validated as a relevant platform for public opinion mining [10]. Although transformer-based models such as BERT [6] achieve strong contextual understanding, LSTM remains computationally efficient for medium-scale datasets.

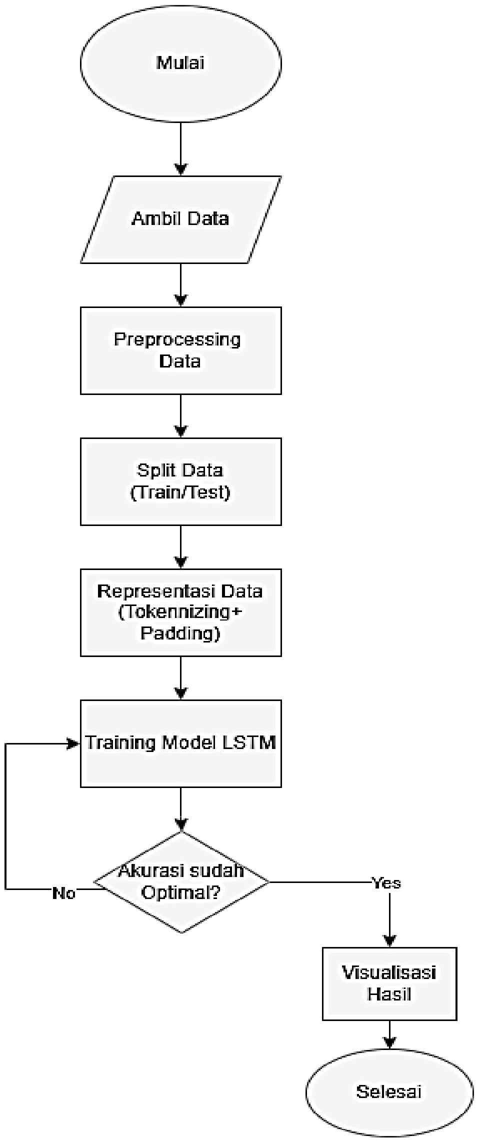

Figure 1. Research Workflow Flowchart

GENERATION JOURNAL
Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

Long Short-Term Memory (LSTM)

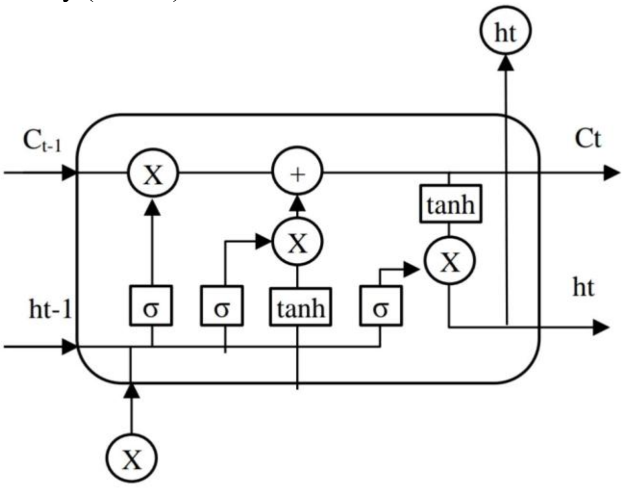

Figure 2. LSTM Architecture

1. Input gate:
   $it = \sigma(Wi \times Xi + Ui \times ht{-}1 + bi)$ ……………………………. (1)
2. Forgate:
   $ft = \sigma(Wf \times Xt + Uf \times ht{-}1 + bf)$ ……………………………. (2)
3. Cell State:
   $ct = ft \odot ct{-}1 + it \odot c't$ ……………………………. (3)
4. New Candidate:
   $c't = tanh(Wc \times Xc + Uc \times ht{-}1 + bc)$ ……………………………. (4)
5. Output gate:
   $ot = \sigma(Wo \times Xi + Uo \times ht{-}1 + bo)$ ……………………………. (5)
6. Hidden State:
   $ht = ot \odot \sigma(ct)$ ……………………………. (6)

B. Data Collection

The dataset consists of English-language user comments collected from several YouTube videos discussing the India–Pakistan air conflict incident. Data were retrieved using the YouTube Data API v3 through Python in the Google Colab environment, following procedures adopted in previous YouTube-based sentiment studies [10], [12], [15]. Prior research demonstrated that the YouTube Data API enables systematic and reproducible extraction of user-generated comments for opinion mining purposes [10], while LSTM-based YouTube sentiment studies confirmed the feasibility of using API-collected datasets for deep learning experimentation [12]. Additionally, VADER-based sentiment labeling on YouTube data has been successfully implemented using similar extraction techniques [15].

The collected data were stored in CSV format and included metadata attributes such as username, comment text, publication time, and like count. However, only the comment text was used as the primary feature for sentiment classification. In total, 1,336 comments were obtained. Although this dataset size is considered moderate for deep learning training, prior comparative studies indicate that LSTM can achieve stable performance on medium-scale datasets when proper regularization techniques are applied [11], [12].

It is important to acknowledge that YouTube comments may contain spam, duplicated content, sarcasm, or multilingual expressions. Such characteristics necessitate rigorous preprocessing before model training to ensure data consistency and reliability [4].

Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

1. Data Preprocessing

Due to the informal and unstructured nature of social media text, preprocessing plays a crucial role in improving model performance. Social media corpora often contain slang, abbreviations, emojis, irregular grammar, and non-standard punctuation, which increase classification difficulty [4].

The preprocessing pipeline in this study includes:

- Case folding (converting text to lowercase)
- Removal of punctuation, numbers, URLs, and emojis
- Tokenization
- Stopword removal using the NLTK stopword list

Preprocessing decisions significantly influence sentiment classification accuracy in deep learning models [9]. Unlike traditional NLP pipelines, no stemming or lemmatization was applied in this study. This decision is supported by findings indicating that preserving original word forms can maintain contextual nuances in social media sentiment tasks [14], [15].

Sentiment labels were automatically generated using the VADER sentiment analyzer [5]. While lexicon-based labeling improves scalability and eliminates the need for manual annotation, it may introduce labeling noise, particularly in sarcastic or context-dependent expressions [14]. Nevertheless, VADER remains computationally efficient and widely adopted for large-scale pseudo-label generation in social media research [15].

2. Dataset Splitting

The dataset of 1,336 comments was randomly divided into training and testing sets using an 80:20 ratio, resulting in 1,068 training samples and 268 testing samples. This ratio is commonly adopted in sentiment classification research to balance training sufficiency and evaluation reliability [11], [12]. Maintaining proportional class distribution during splitting is essential to prevent biased performance evaluation in multi-class sentiment classification.

3. Text Representation

Before being processed by neural networks, textual data must be converted into numerical representations. Traditional machine learning approaches such as TF-IDF rely on frequency-based representations, which ignore contextual semantics and word order [1]. Such limitations reduce their effectiveness in modeling emotionally nuanced and sequential text.

To address this limitation, tokenization was applied to convert words into integer indices, followed by sequence padding to ensure uniform input length. An embedding layer was then employed to transform word indices into dense vector representations. Distributed word representations have been shown to significantly enhance sentiment classification performance by preserving semantic similarity and contextual relationships [7]. Comprehensive reviews confirm that embedding-based deep learning models outperform classical bag-of-words approaches in complex sentiment analysis tasks [9].

C. LSTM Model Training

The LSTM architecture, introduced to overcome the vanishing gradient problem in recurrent neural networks, enables modeling of long-term contextual dependencies through gated memory mechanisms [8]. Compared to classical machine learning classifiers, LSTM captures sequential patterns and emotional transitions within textual sequences.

Empirical evidence supports the effectiveness of LSTM for YouTube sentiment classification. Comparative studies demonstrated that LSTM outperforms SVM and Naïve Bayes due to its ability to model sequential dependencies [11]. Further studies applying LSTM to YouTube discussions reported improved accuracy compared to traditional classifiers [12]. Indonesian research also highlighted the effectiveness of recurrent architectures in politically sensitive sentiment analysis tasks [16], [17].

In this study, the model architecture consists of:

- An embedding layer
- An LSTM layer with 64–128 units

GENERATION JOURNAL
Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

- A dropout layer to mitigate overfitting
- A dense softmax output layer for three-class classification

The model was trained using the Adam optimizer and categorical cross-entropy loss for five epochs with a batch size of 64. Although transformer-based models such as BERT often achieve higher contextual accuracy [6], LSTM remains computationally efficient and suitable for medium-scale datasets. However, deep learning models are susceptible to overfitting, especially with limited data, which justifies the inclusion of dropout regularization [8].

D. Baseline Model: Naïve Bayes with TF-IDF

To provide empirical comparison, a Multinomial Naïve Bayes classifier with TF-IDF vectorization was implemented as a baseline model. Frequency-based classifiers remain competitive in structured sentiment tasks [1]; however, they assume feature independence and fail to capture sequential dependencies.

Comparative studies confirmed that while Naïve Bayes is computationally efficient, it often underperforms compared to LSTM in sequential modeling tasks [11]. Therefore, including Naïve Bayes as a baseline enables quantitative evaluation of whether deep sequential modeling provides measurable improvement over classical approaches in conflict-oriented discourse.

E. Model Evaluation and Visualization

Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. These metrics are widely adopted in multi-class sentiment classification research to assess both overall performance and class-level prediction reliability [9], [11]. A confusion matrix was generated to analyze misclassification patterns across positive, neutral, and negative sentiment categories.

Training and validation accuracy-loss curves were visualized to detect potential overfitting behavior, which is a known challenge in deep neural network training [8]. Word cloud visualizations were also generated to complement quantitative findings by illustrating frequently occurring sentiment-associated terms within the dataset.

Table 1. Tools and Development Environment

Component	Description
1. Programming Language	Python
2. Main Libraries	TensorFlow, Keras, Pandas, NumPy Sckit-learn, VADER
3. Data Source	YouTube Comments (Collected via YouTube API v3)
4. Data Format	CSV

RESULTS AND DISCUSSION

A. Data Description

The research data were obtained using the YouTube Data API v3 by specifying selected video IDs related to the India–Pakistan air conflict. The collected data consist of publicly available English-language comments. After the data collection process, all comments were stored as a dataset and used as raw data for this study.

Text preprocessing was conducted to improve data quality prior to model training. The preprocessing steps include converting all text to lowercase, removing URLs, symbols, numbers, and punctuation, eliminating emojis and non-alphabetic characters, and removing English stopwords. The output of this process is a cleaned comment dataset ready for sentiment analysis.

GENERATION JOURNAL
Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

B. Sentiment Data Labeling

This study adopts a supervised learning approach; therefore, labeled data are required. Sentiment labeling was performed automatically using the VADER Sentiment Analyzer. Each comment was assigned a sentiment label based on its compound score with the following criteria:

- Negative sentiment if the compound score is less than −0.05
- Neutral sentiment if the compound score is between −0.05 and 0.05
- Positive sentiment if the compound score is greater than 0.05

This labeling process resulted in three sentiment classes negative, neutral, and positive which were used as target labels during the LSTM model training process.

C. LSTM Model Training

The LSTM model was developed using a neural network architecture consisting of an embedding layer, an LSTM layer, a dropout layer, and a dense layer with a softmax activation function. The dataset was divided into training and testing sets using an 80:20 ratio.
Model training was conducted for five epochs with a batch size of 64. As shown in Figure 3, the training accuracy increased significantly up to the fourth epoch. At the fifth epoch, training accuracy continued to increase while validation accuracy decreased, indicating the occurrence of overfitting. The best-performing model was obtained at the fourth epoch, which achieved the highest validation accuracy and was therefore selected as the final model.

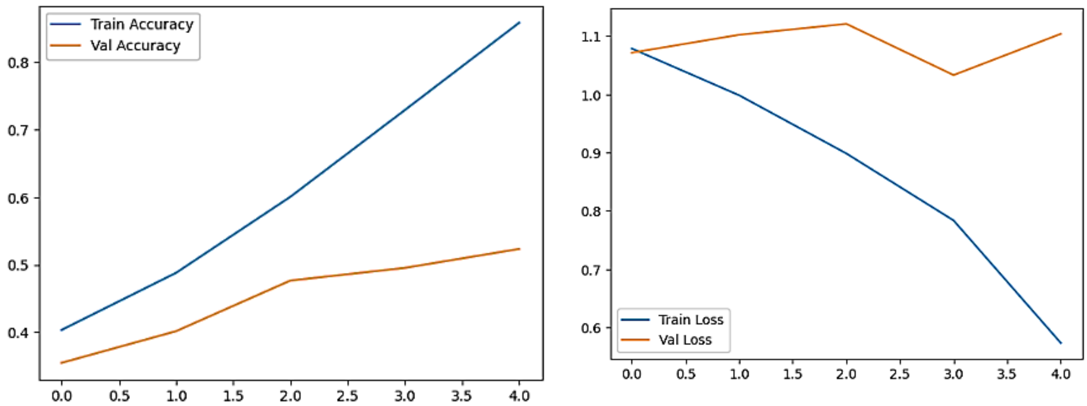

Figure 3. Accuracy and Loss Curves of the LSTM Model During Training

Figure 3 shows the accuracy and loss curves of the LSTM model during the training process over five epochs. Based on the accuracy curve, the training accuracy consistently increases from the first to the fifth epoch, indicating that the model is able to learn patterns from the training data effectively. In contrast, the validation accuracy increases until the fourth epoch and then slightly decreases at the fifth epoch.

This behavior suggests that the model reaches its optimal performance around the fourth epoch. After this point, the improvement in training accuracy is no longer followed by an improvement in validation accuracy, which indicates the beginning of overfitting. This phenomenon is consistent with findings in deep neural network training literature, where prolonged training beyond the optimal epoch may reduce generalization performance [8], [9]. The model starts to fit the training data too closely, resulting in reduced generalization performance on unseen data.

The loss curves further support this observation. The training loss decreases steadily across all epochs, while the validation loss decreases until the fourth epoch and then increases sharply at the fifth epoch. The increase in validation loss confirms that overfitting occurs when the model is trained beyond the optimal number of epochs. Therefore, the LSTM model trained at the fourth epoch is selected as the final model, as it provides the best balance between training performance and generalization ability on the validation data.

GENERATION JOURNAL

Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

D. Model Evaluation

Model evaluation was performed using the testing dataset to assess the generalization capability of the LSTM model. The evaluation metrics include accuracy, confusion matrix, precision, recall, and F1-score.

The evaluation results indicate that the LSTM model is capable of classifying YouTube comment sentiment with satisfactory performance. The accuracy value on the testing data demonstrates that the model effectively captures sentiment patterns in English-language comments. The confusion matrix, shown in Figure 4, reveals that most data points were correctly classified, although some misclassifications occurred between closely related sentiment classes, particularly between neutral and negative sentiments.

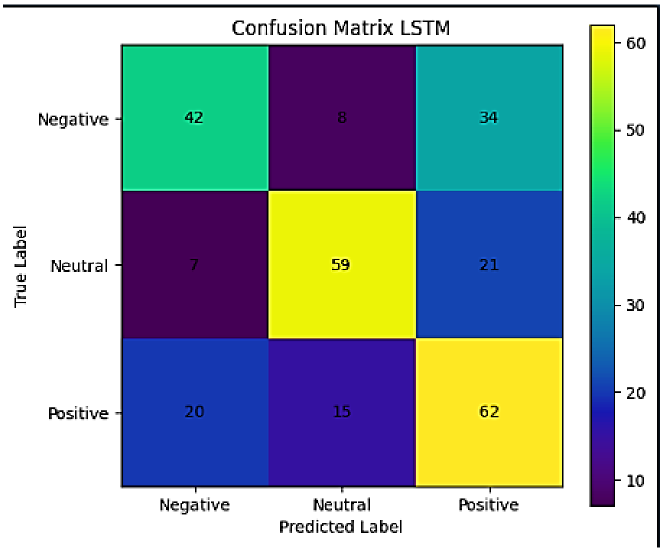

Figure 4. Confusion Matrix of the LSTM Model

1. Sentiment Distribution Analysis

   After model training and evaluation, sentiment classification was applied to the entire dataset to analyze sentiment distribution. The results are presented in **Table 2**.

Table 2. Sentiment Distribution

Sentiment	Percentage
Negative	31,21%
Positive	38,32%
Neutral	30.46%

The dominance of negative sentiment indicates that most users expressed critical responses toward the video content. This tendency is influenced by the topic of the video, which discusses military conflict and defense capability comparisons, issues that often provoke debate and negative opinions.

2. Word Cloud Analysis

   The overall word cloud visualization illustrates the most frequently occurring terms in YouTube comments related to the India–Pakistan air conflict. Dominant words reflect themes of military confrontation, defense capability, and geopolitical tension, indicating that user discussions are strongly centered on the conflict narrative presented in the video. The prominence of these terms supports the sentiment classification results, particularly the dominance of negative sentiment, as conflict-related topics tend to elicit critical and emotionally charged responses. This visualization

Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

provides a concise overview of the discussion context and complements the quantitative findings of the LSTM-based sentiment analysis by highlighting key topics emphasized by users.

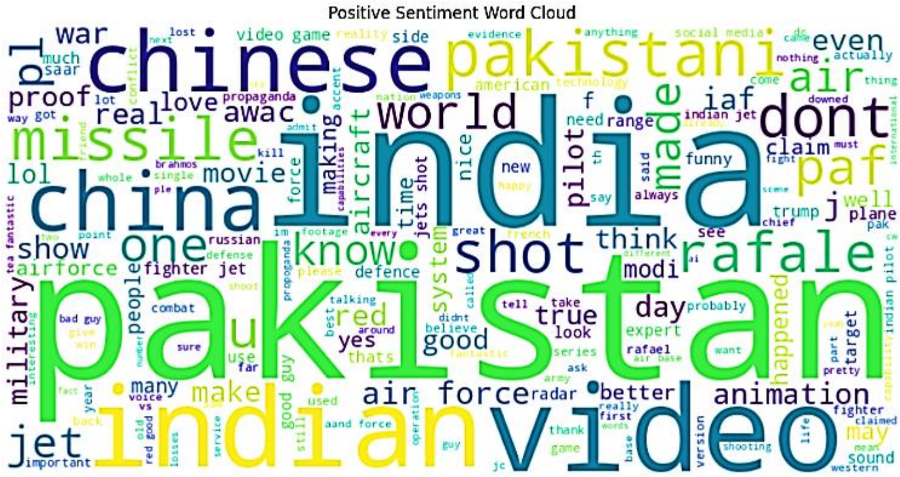

Figure 6. WordCloud Positive of YouTube Comment Data

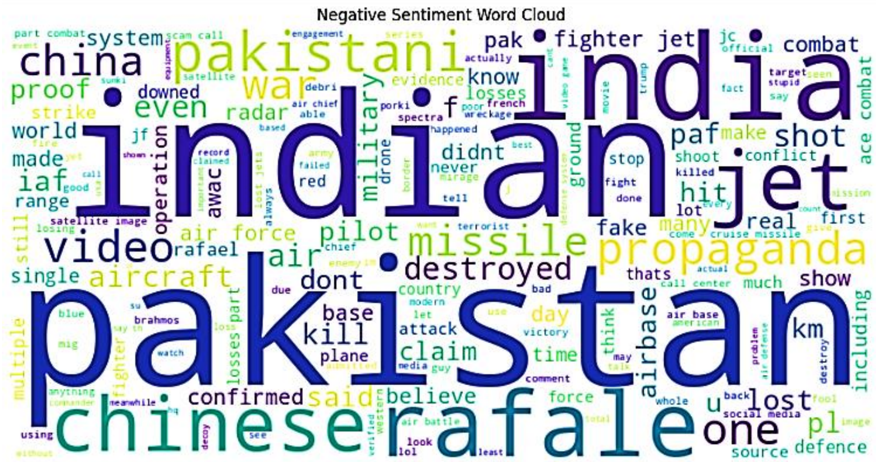

Figure 7. WordCloud Negative of YouTube Comment Data

3. Baseline Comparison: Naïve Bayes vs LSTM

	precision	recall	f1-score	support
0	1.00	0.01	0.02	89
1	1.00	0.05	0.10	77
2	0.37	1.00	0.54	95
accuracy			0.38	261
macro avg	0.79	0.35	0.22	261
weighted avg	0.77	0.38	0.23	261

Figure 8 Evaluation metrics of Naive Bayes with TF-IDF.

Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

	precision	recall	f1-score	support
Negative	0.61	0.50	0.55	84
Neutral	0.72	0.68	0.70	87
Positive	0.53	0.64	0.58	97
accuracy			0.61	268
macro avg	0.62	0.61	0.61	268
weighted avg	0.62	0.61	0.61	268

Figure 9 Precission, Recall, and F1-Score of the LSTM Model.

The Naive Bayes (NB) baseline achieved 38% accuracy. The superiority of LSTM over Naïve Bayes observed in this study aligns with previous comparative findings reported in [11] and [12], where sequential modeling demonstrated improved contextual understanding compared to frequency-based classifiers. However, the overall accuracy achieved in this study (61%) remains lower than results reported in politically controlled or entertainment datasets [12], [16], suggesting that geopolitical conflict discourse presents higher classification complexity. and a macro-F1 of 0.38. While it shows high precision for certain classes, recall is very low for minority classes, indicating that the model often fails to capture neutral or negative sentiments. In contrast, the LSTM model outperforms NB with 61% accuracy and higher macro-F1, demonstrating its ability to capture sequential dependencies and context in YouTube comments more effectively. These results highlight that sequential modeling can extract richer sentiment patterns compared to classical word-frequency-based methods.

CONCLUSION

This study successfully applied the Long Short-Term Memory (LSTM) method for sentiment analysis of English-language YouTube comments. The results show that the LSTM model is capable of processing sequential text data and effectively identifying sentiment patterns in user comments.
Text preprocessing steps, including text cleaning, stopword removal, and word normalization, play an important role in improving data quality and supporting the performance of the LSTM model in sentiment classification. Automatic sentiment labeling using the VADER Sentiment Analyzer enables an efficient supervised learning process without manual annotation, making it suitable for large-scale YouTube comment datasets.

The experimental results indicate that the LSTM model achieved its best performance at the fourth training epoch. In subsequent epochs, overfitting began to occur, as reflected by an increase in validation loss despite improved training accuracy. Sentiment distribution analysis shows that positive sentiment dominates the dataset (38.32%), followed by negative (31.21%) and neutral (30.46%) sentiments, indicating that user responses tend to be critical. The word cloud visualization further supports these findings, highlighting frequently occurring terms related to military conflict and defense capability comparisons.
The proposed LSTM model outperformed the Naïve Bayes baseline, achieving 61% accuracy and a macro-F1 of 0.61, demonstrating that sequential modelling captures sentiment patterns more effectively.

SUGGESTIONS

Based on the findings of this study, several recommendations for future research are proposed to address existing research gaps:

Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

1. Future studies may expand the dataset by incorporating a larger volume of YouTube comments collected from multiple videos related to similar geopolitical conflicts to improve the robustness and generalizability of sentiment classification results.
2. Further research could explore alternative or hybrid sentiment labeling approaches, such as manual or semi-supervised labeling, to enhance label reliability and reduce potential bias introduced by fully automated labeling methods.
3. Future work may investigate advanced neural architectures, such as Bidirectional LSTM or attention mechanisms, to better capture contextual dependencies and mitigate overfitting issues observed during model training.
4. Subsequent research could focus on aspect-based sentiment analysis to provide more granular insights into public opinions expressed in YouTube comments on geopolitical issues

REFERENCES

[1] R. Gunawan and R. A. Nugroho, "Sentiment Analysis of YouTube Comments on Government Policy Using Lexicon-based and Machine Learning Techniques," *Journal of Intelligent Systems*, 2021.

[2] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *EMNLP*, 2014.

[3] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool, 2012.

[4] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in *Proc. LREC*, 2010.

[5] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," in *Proc. ICWSM*, 2014.

[6] J. Delvin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *NAACL-HLT*, 2019.

[7] D. Tang, B. Qin, and T. Liu, "Document Modeling with Gated Recurrent Neural Network for Sentiment Classification," in *Proc. EMNLP*, 2015.

[8] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in *Neural Computation*, 1997.

[9] Y. Zhang and J. Zhao, "Deep Learning for Sentiment Analysis: A Survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 2020.

[10] T. H. R. Pradita, "Sentiment Analysis of YouTube Comments Toward Chat GPT," *Jurnal Transformatika*, 2023.

[11] E. L. Utari and S. H. Wibowo, "Analisis Komparatif Algoritma SVM, Naive Bayes dan LSTM pada Sentimen Komentar Lagu Labour," *JINTEKS*, 2025.

[12] S. Supriyono, A. P. Wibawa, S. Suyono, and F. Kurniawan, "Analyzing Audience Sentiments in Digital Comedy: A Study of YouTube Comments Using LSTM Models," *Journal of Applied Data Sciences*, 2025.

[13] B. Anderson, M. Irzal, and A. Hendarno, "Sentiment Analysis of Indonesia's Free School Lunch Policy Using LSTM and Word2Vec on YouTube Comments," *J-KOMA: Journal of Computer Science and Applications*, vol. 7, no. 2, pp. 36–44, 2024.

[14] A. A. Bastian and A. Perdana, "Sentiment Analysis of Public Comments on YouTube Regarding the Inaugural Speech of the 8th President of Indonesia Using VADER and BERT Methods," *IJSECS*, vol. 5, no. 1, pp. 126–140, 2025.

[15] H. Yuliansyah, "Sentiment Analysis of the Waste Problem Based on YouTube Comments using VADER and Deep Translator," *Jurnal Media Informatika Budidarma*, vol. 8, no. 1, 2023.

GENERATION JOURNAL
Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
DOI: ihttps://doi.org/10.29407/gj.v10i2.28062

[16] A. Prasetyo, R. Kurniawan, and D. Saputra, “Bidirectional LSTM for Political Sentiment Analysis on Indonesian YouTube Channels,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, pp. 350–358, 2024.

[17] M. Wibisono and F. Hartono, “Hybrid Deep Learning Approaches for Geopolitical Sentiment Analysis on Social Media,” *Jurnal Teknologi dan Sistem Komputer*, vol. 13, pp. 45–54, 2025.