
Journal of Informatics and Web Engineering

Vol. 4 No. 2 (June 2025)

eISSN: 2821-370X

Sentiment Analysis of the 2024 General Election Through Twitter using Long-Short- Term Memory Algorithm

Angga Wahyu W¹, Haidar Hilmy Andana¹, Junta Zeniarja^{2*}, Aris Febriyanto³

¹ Department of Computer Science, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji-shi, Tokyo 192-0397, Japan

² Department of Informatics Engineering, Faculty of Computer Science, Dian Nuswantoro University, Imam Bonjol Street No. 207, Semarang, Indonesia

³ Faculty of Engineering, Dian Nuswantoro University, Imam Bonjol Street No. 207, Semarang, Indonesia

*corresponding author: (junta@dsn.dinus.ac.id; ORCID: 0000-0001-5009-666X)

Abstract - This study analyses sentiment related to the 2024 Indonesian Presidential Election using the Long Short-Term Memory (LSTM) algorithm. A total of 2,400 tweets in the Indonesian language were gathered, with approximately 400 tweets sampled per week. In the data preparation, lexicon-based sentiment tagging, oversampling for class balance, and the creation and training of an LSTM model are all included in the study approach. The built model consists of embedding layers, Conv1D, and two LSTM layers. The LSTM model was selected due to its ability to capture long-range contextual dependencies in sequential text data like tweets, facilitated by its gate mechanisms (input, forget, output) that regulate information flow. The model achieved 84.3% accuracy in classifying sentiments (positive, neutral, negative), demonstrating its potential for real-time public opinion monitoring. The results provide actionable insights for election organisers and political analysts. For further study, using a wider spectrum of data to supplement model performance will help development. Tweaking hyperparameters and playing with other architectural models like GRUs or Transformers could improve model accuracy. Improved sentiment tagging calls for a more thorough and relevant sentiment vocabulary. The proposed model can be further developed into a real-time sentiment analysis tool to provide insights into public opinion on elections and other concerns.

Keywords—Sentiment Analysis, 2024 General Election, Twitter, Long Short-Term Memory, Preprocessing, Data Crawling, Natural Language Processing

Received: 1 March 2025; Accepted: 23 May 2025; Published: 15 June 2025

This is an open access article under the [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



1. INTRODUCTION

The race of the President and bad habit President could be an essential occasion in a country that maintains equitable standards. The results of this constituent prepare set up the nation's vital pioneer and significantly impact approaches and authority direction. Consequently, comprehending popular opinion, attitudes, and sentiments towards presidential and vice-presidential candidates is crucial. When it comes to politics, in this day and age of digital communication, social media has grown into a powerful voice in the public sphere [1].



Journal of Informatics and Web Engineering

<https://doi.org/10.33093/jiwe.2025.4.2.25>

© Universiti Telekom Sdn Bhd.

Published by MMU Press. URL: journals.mmupress.com/jiwe

In this period, society is observing a situation where social media has emerged as the primary platform for discourse and exchanges among netizens regarding political concerns and potential leaders[2]-[3]. These internet platforms embody the varied perspectives and ambitions of the community, which can profoundly influence the trajectory of politics and national leadership. In light of the 2024 Indonesian Race, it is basic to analyse well known disposition on social media concerning candidates and political issues. This examination presents points of view, assumptions, and feelings which will influence voters' choices. Thus, a progressed and effective assumption investigation strategy is required.

Opinion mining, sometimes known as sentiment analysis, is an automated method for understanding, extracting, and processing textual material to produce insights into sentiment buried inside opinion statements [4]-[5]. Sentiment analysis classifies text inside sentences or papers according to its polarity to determine whether the stated opinion is positive, negative, or neutral. This sentiment analysis aims to determine whether comments or opinions about a problem are good or negative, so providing a standard for improving product or service quality. Their great impact and advantages are driving fast growth of sentiment analysis-based research and applications [6].

From unstructured data, sentiment analysis is a powerful tool for analysing user happiness. This data processing can produce results for sentiment analysis. Deep Learning is one approach used in sentiment analysis. Deep Learning is an artificial neural network that extracts insights from data; the model improves its performance by changing the error value depending on its errors [7]. One deep learning method for text categorisation is Long Short-Term Memory (LSTM) [8]. Applicable to Natural Language Processing (NLP), which comprises speech recognition, text translation, and sentiment analysis, long and LSTM is a deep learning method. LSTM was created as a betterment on the original recurrent neural network (RNN) model to solve the missing step issue RNNs frequently experience [9].

The principle of Augmented Intelligence, which emphasises the synergy between AI technologies and human expertise [10], helps to augment the use of AI in evaluating public mood about the 2024 General Election. Using LSTM algorithms, artificial intelligence effectively examines vast volumes of tweet data to find trends and sentiment patterns over time and catch subtle public opinions about politicians, politics, or events. When combined with human interpretation, this information becomes more potent and allows campaign teams and policy analysts to create evidence-based strategic projects using AI-generated insights. Augmented intelligence guarantees that raw data is converted into usable information and connected to human judgement to enhance its relevance and influence during elections by combining automated data analysis with human decision-making.

2. LITERATURE REVIEW

A well-liked social network, Twitter (X) offers a brief stage for individuals to share their ideas and viewpoints on a range of subjects and concerns. Using NLP, sentiment analysis extracts, processes, and evaluates attitudes in tweets, categorising them as positive, negative, or neutral. Widely used in many applications, including sentiment analysis for cyberbullying detection [11]-[13], political and social analysis [14]-[16], and e-commerce data processing [17]-[18], LSTM is a complex deep learning architecture. Its recursive structure expertly captures sequential dependencies in text input, therefore making it very appropriate for activities requiring contextual understanding, such as sentiment analysis and multi-label classification. Studies show that LSTM models often outperform traditional machine learning models in accuracy.

Political and social research uses sentiment analysis to understand public opinion. It helps to evaluate sentiment on social trends, policies, and political candidates[14]. The prior study [19] looks at how well a deep learning model combining Bidirectional Encoder Representations from Transformers (BERT)-based data representation with the LSTM technique performs. This model is applied to YouTube comment data on political videos about Indonesia's 2024 presidential election. The effectiveness of the model is evaluated using performance indicators such as accuracy, precision, and recall. With an accuracy of 0.8783, the findings show the BERT-LSTM model outperforms the independent BERT model.

A different paper [20] looks at how well different deep learning architectures work for multi-label sentiment classification of YouTube comments about the 2024 Indonesian presidential election. Comments from discussion films starring Anies Baswedan, Prabowo Subianto, and Ganjar Pranowo make up the dataset. The work assesses a hybrid CNN-BiLSTM model, Bidirectional Long Short-Term Memory (Bi-LSTM), and Convolutional Neural Networks (CNN). Normalisation, removal of superfluous characters, case folding, stop word removal, and text

augmentation made up the preprocessing steps. Including class weights into the loss function helped to offset class imbalance. With an average accuracy of 98%, the Bi-LSTM model performed best among the models evaluated on accuracy, Area Under the Curve (AUC), and Hamming Loss.

Daffa et al. [21] looked at how BERT and CNN were used for sentiment analysis for the 2024 Indonesian presidential election. Particularly with regard to political concerns before the election, the paper underlines the advantages of combining BERT and CNN to increase the accuracy of sentiment analysis on X. The study found that the BERT model had an average accuracy of 90.2%, outperforming CNN's 88.19%. With a projected support of 43.82%, Prabowo Subianto led the field for the 2024 Indonesian presidential election, followed by Ganjar Pranowo at 33.83% and Anies Baswedan at 22.35%, according to sentiment forecasts using BERT. When compared to the real election outcomes, there were discrepancies: Anies Baswedan got 24.95% (a difference of 2.60%), Ganjar Pranowo 16.47% (a change of 17.36%), and Prabowo Subianto 58.58% (a difference of 14.76%). The results indicate that the BERT model reliably forecasts election results and correctly reflects public opinion. The research on Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Analysis of Twitter data using IndoBERT language models reveals that 83.7% of Indonesians are neutral regarding the 2024 elections. With an accuracy of 83.5%, the experimental results show that IndoBERT large-p1 is the most successful model. This outperforms baseline systems, improving the accuracy and F1-score of TextBlob by 48.5% and 46.49%, Multinomial Naïve Bayes by 2.5% and 14.49%, and Support Vector Machine by 3.5% and 13.49%, respectively[19].

Using Word2Vec to give weights to particular words in tweets, the study on the Sentiment Analysis of Cyberbullying with Bidirectional LSTM on Twitter The study uses a dataset of 47,692 tweets gathered from Kaggle. Using a Confusion Matrix, the assessment produces classification statistics of 82.29% accuracy, 82.04% precision, 81.95% recall, and an F1-Score of 81.89%. Dealing with and preventing cyberbullying on Twitter depends on this sentiment analysis tool. The relative comparison with accepted reference algorithms shows that the proposed categorisation method works well [22].

3. RESEARCH METHODOLOGY

3.1 Proposed Method

This study initiates by gathering unprocessed sentiment data from Twitter through the crawling method. The raw data undergoes preprocessing, which entails several processes, including noise removal, format correction, and tokenization to isolate words. The purpose of these preprocessing stages is to generate a more organised dataset suitable for subsequent analytical processing.

In the conceptual modelling framework presented in this study as depicted in Figure 1, the process begins with the extraction of tweets from Platform X using a crawling technique implemented in Python, leveraging the Tweet Harvest package. Once preprocessing is completed, sentiment annotation is performed on the dataset, followed by techniques to balance the distribution of sentiment classes to address any data imbalance issues.

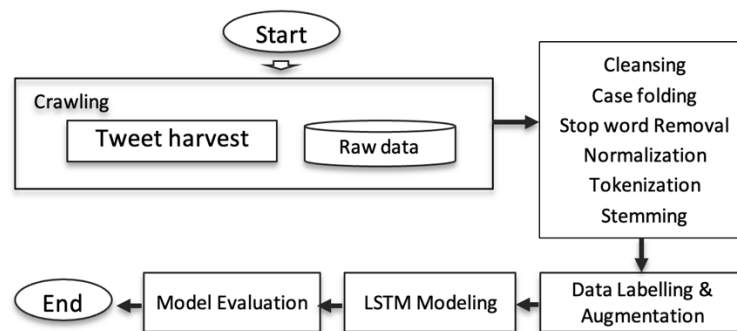


Figure 1. Proposed Method

This phase targets content related to harsh weather in Indonesia, using keywords such as "Election" to filter relevant discussions. The next stage is data preprocessing, which encompasses several steps including cleansing, case folding, stop word removal, normalization, tokenization, and stemming to ensure the text is clean and standardised for analysis.

Subsequently, LSTM model is trained using the cleaned and labelled dataset to capture sequential dependencies and contextual meaning within the text. In the model evaluation phase, the performance of the LSTM model is assessed through standard classification metrics such as accuracy, precision, recall, and F1-score, ensuring a robust evaluation of its ability to classify sentiments effectively. Finally, the result analysis phase involves interpreting and forecasting the sentiment classification outcomes, as well as analysing the aspect-based elements embedded in the generated tweets, providing meaningful insights into public sentiment patterns.

3.2 Model Evaluation

Using a test dataset made up of tweets and retweets, this paper assesses the LSTM model's ability to identify text sentiment about the 2024 Election in Indonesia. Testing employs evaluation standards including perplexity network, accuracy, review, and F1-score. The perplexity lattice promotes the comparison of show expectations with actual values, advertising a thorough appraisal of demonstrate execution in estimation categorisation, so ensuring that the results fit with the enquire about goals of assessing open temperament with respect to the 2024 Decision.

The evaluation findings will clearly show how well the LSTM model can identify and classify positive, negative, and neutral sentiments from texts related to the 2024 Election. This assessment is crucial to guarantee the model produces pertinent and correct results according the research goals. The ponder points to fully assess the model's ability to classify content into estimation classes correctly using measures including precision, exactness, review, and F1-score. The results of this study will provide useful insights into the performance of the model in examining public opinion in relation to the 2024 Race.

3.3 Result Analysis

The final stage in this analytical procedure is to evaluate the predictive outcomes of the trained model. The objective is to implement the model on new tweets and evaluate its efficacy in predicting sentiment from the text. The primary objective of this thesis is to predict sentiment based on tweets, making this phase essential for evaluating the efficacy of the developed model.

4. RESULTS AND DISCUSSIONS

4.1 Data Collection

The data collection for this study is conducted by a crawling procedure on Twitter pertaining to the 2024 Indonesian Election, which will subsequently yield training and testing data. The crawling is conducted with tweet-harvest. Data collection was conducted using the phrase "Elections" to gather Indonesian tweets sent between January 17, 2024, to February 28, 2024, so encompassing tweets both prior to and following the election day of February 14, 2024. The quantity of tweets collected by tweet-harvest can be modified as required; in this instance, 400 data points are gathered weekly. Data was collected via Twitter crawling using the keyword "Election" from January 17 to February 28, 2024. The data summarised in Table 1 was collected in accordance with Twitter's Application Programming Interface (API) compliance guidelines.

Table 1. Tweet Dataset Information

Characteristic	Detail
Collection Period	17 January – 28 February 2024
Keyword	"Election"
Initial Data Volume	400 tweets/week (total 2,400 tweets)
Language	Indonesian (includes colloquialisms/typos)
Initial Sentiment Distribution	Neutral (60%), Positive (20%), Negative (20%)

The collection period spanned from January 17 to February 28, 2024, strategically encompassing the national election day on February 14. Tweets were retrieved using relevant keywords such as “Pemilu 2024” (Election 2024), “Pilpres” (Presidential Election), and the names of presidential candidates. A total of 2,400 tweets in the Indonesian language were gathered, with approximately 400 tweets sampled per week. The dataset included informal language features such as slang and typographical variations to reflect authentic social media discourse. To ensure ethical integrity, all usernames and personal identifiers were anonymised during the preprocessing stage, adhering to responsible data handling practices.

4.2 Data Preprocessing

Upon collection of the raw data from Twitter, the preprocessing phase removes extraneous aspects from the tweet data, rendering it appropriate for processing and analysis. This stage envelops various basic methods, counting cleansing, case collapsing, stop word end, normalization, tokenization, and stemming. These steps are basic for moving forward the quality of the content information earlier to its examination with the LSTM demonstrate. The information is upgraded in cleanliness and consistency by the expulsion of pointless pieces, standardization of wording, end of stop words, and usage of stemming. These strategies improve the model's capacity to comprehend and assess opinion, coming about in more tried and true expository results. Detailed preprocessing steps:

1. **Cleansing:** Removed usernames (@user), URLs, hashtags (#), and emojis.
Example: "Election 2024 @user hope it's fair #CandidateX" → "Election 2024 hope it's fair"
2. **Case Folding:** Converted text to lowercase.
Example: "ELECTION" → "election"
3. **Stop word Removal:** Used a custom stop word list (e.g., informal abbreviations).
4. **Stemming:** Applied Porter Stemmer for Indonesian.
Example: "government's" → "govern"
5. **Tokenization:** Split text into word units using regex.

The preprocessing pipeline:

1. **Cleansing:**
Removed metadata (@mentions, URLs, hashtags).
Example: "Pemilu 2024 @user semoga adil #Jokowi" → "Pemilu 2024 semoga adil"
2. **Normalization:**
Converted colloquial terms to standard Indonesian (e.g., "gmn" → "bagaimana").
3. **Stemming:**
Applied Sastrawi stemmer for Indonesian (e.g., "pemerintahan" → "perintah").

4.3 Lexicon-Based Labelling and Oversampling

Lexicon-based sentiment labelling was performed using the InSet sentiment dictionary, which contains 3,609 positive and 6,609 negative Indonesian words, each assigned a polarity score ranging from -5 to +5. For example, the word "adil" (fair) carries a strong positive sentiment with a score of +4, while "korupsi" (corruption) reflects a strong negative sentiment with a score of -5. These scores were used to determine the overall sentiment of each tweet based on the aggregation of word-level polarity. Prior to oversampling, the dataset exhibited a class imbalance, with neutral sentiments dominating at 60%, and both positive and negative sentiments each representing 20% of the data. To address this imbalance, random oversampling was applied to equalise the class distribution, resulting in a balanced dataset with 33% for each sentiment class (positive, neutral, and negative), thereby improving model fairness and classification reliability.

4.4 LSTM Architecture

The LSTM model comprises many layers: an embedding layer for word representation, an LSTM layer for sequence analysis, and an output layer for sentiment categorization. The show is prepared employing a curated dataset. Amid preparing, it learns to recognise designs within the content arrangement and alters its inner settings to make strides

expectation precision. At each age, the information passes through the LSTM unit, where cell and covered up states hold key data from past information. This strategy revives the model's energetic information representation.

Figure 2 depicts the engineering of a Repetitive Neural Organize (RNN) joining LSTM, commonly utilised for preparing successive information like content or time arrangement. The method commences with input in successive information delineated as an implating network, wherein the numerical information is changed into a vector representation. The LSTM layer hence forms this representation, capturing transient relationships or designs within the information by holding data from earlier time steps through inner memory. The yield of the LSTM layer is transmitted to the covered-up layer, which serves as an progressed include to distinguish non-linear connections inside the information. The comes about from the covered-up layer are eventually utilised to supply yield, which may be a classification, expectation, or nonstop esteem, unexpected upon the particular assignment. This engineering empowers the demonstrate to observe complex designs in successive information.

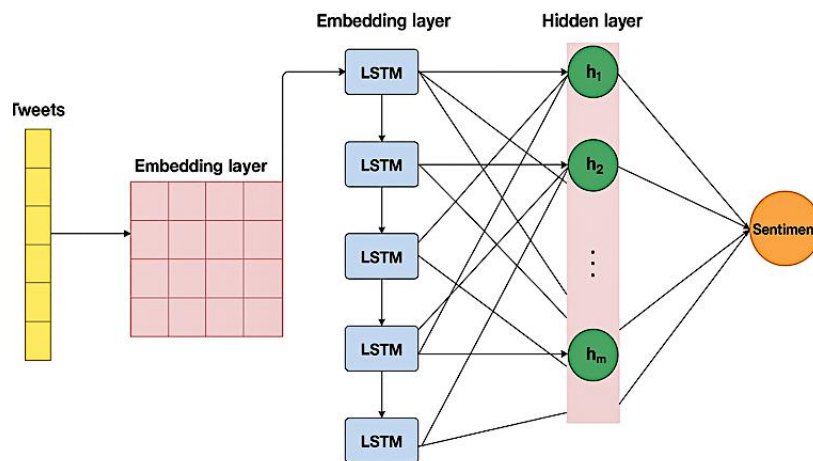


Figure 2. LSTM Model Architecture for Sentiment Analysis of Election Tweets

The sentiment classification model begins with an input layer that is passed through an embedding layer consisting of a vocabulary size of 8,000 words and 120-dimensional embeddings. The architecture includes two LSTM layers: the first layer contains 128 memory units and applies a 20% dropout rate to prevent overfitting, while the second layer comprises 64 memory units to further extract temporal features. The output layer utilises a softmax activation function to classify the sentiment into three categories: positive, neutral, and negative. For training, the model runs for 20 epochs using the Adam optimiser, which ensures efficient convergence. The dataset is divided using an 80:20 split for training and testing, respectively, to evaluate the model's generalization ability. This setup is designed to effectively capture the sequential nature of tweet data while managing class imbalances. The dimensional settings and layer design aim to strike a balance between performance and complexity, particularly for handling informal and diverse linguistic expressions in Indonesian tweets. This architecture supports robust learning from context-dependent sentiment signals. The model evaluation includes accuracy, precision, recall, and F1-score to measure predictive reliability across classes.

4.5 Model Evaluation

The LSTM model demonstrated several key strengths in the sentiment classification task as it compared to Naïve Bayes and Support Vector Machine (SVM), presented in Table 2. First, its ability to retain contextual information over long sequences allowed it to effectively capture crucial expressions such as "pemilu jujur" (fair election) in lengthy tweets, preserving semantic meaning. Second, the model's forget gate proved beneficial in managing noisy input, such as typographical errors—for example, correcting "pemlu" to "pemilu"—thereby enhancing robustness. Third, the incorporation of dropout layers helped mitigate overfitting, improving the model's generalization to unseen data. These architectural features collectively contributed to the model's strong performance.

Table 2. Performance Comparison

Model	Accuracy	F1-Score
LSTM (Proposed)	84.3%	0.84
Naïve Bayes	75.1%	0.72
SVM	78.9%	0.76

As illustrated in the confusion matrix (Figure 3), the model achieved its highest precision in classifying the positive sentiment class, reaching 93%. This highlights the model’s strength in correctly identifying tweets with favourable tones. Overall, the LSTM architecture not only maintained semantic continuity but also showed resilience to informal and erroneous language commonly found in social media data.

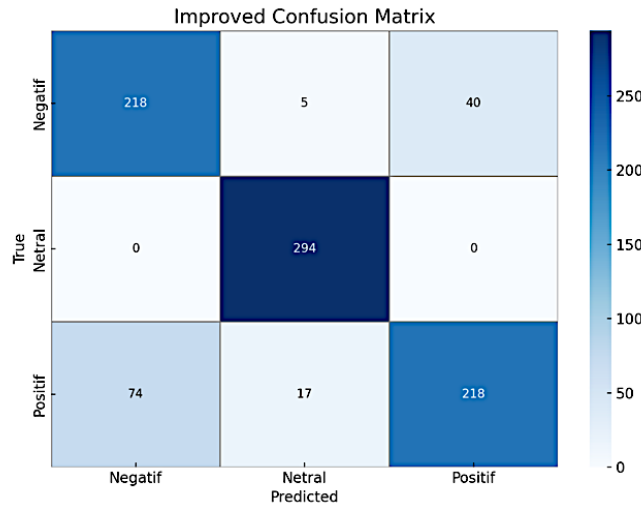


Figure 3. The Confusion Matrix

4.6 Model Evaluation

Table 3 presents a comparative analysis of model performance before and after applying oversampling techniques. After oversampling, the overall accuracy improved from 75% to 84%, indicating a substantial enhancement in the model’s predictive performance. Additionally, the recall for the Negative class increased markedly from 52% to 83%, highlighting the effectiveness of oversampling in addressing class imbalance and improving detection of minority class instances.

Table 3. Performance Before/After Oversampling

Metric	Before Oversampling	After Oversampling
Accuracy	75%	84%
Negative Recall	52%	83%

Key advantages of LSTM before and after oversampling presented in Table 3:

- Gate Mechanism: Forget gate eliminated noise (e.g., typos); input gate preserved critical context (e.g., "corruption" for negative sentiment).
- Sequential Processing: Captured long-range dependencies in tweets (e.g., "This year’s election is better than 2019").
- Dropout Layer: Mitigated overfitting on limited datasets.

The classification report presented in Table 4 shows that the model performs best on the Positive class, achieving a perfect recall of 1.00 and a high precision of 0.93, resulting in an excellent F1-score of 0.96. This indicates that almost all positive instances are correctly identified with very few false positives. The Negative class has a lower precision

of 0.75 and a recall of 0.83, suggesting that while most negative cases are detected, some positive or neutral samples may be misclassified as negative.

Table 4. Classification Report

Class	Precision	Recall	F1-Score	Support
Negative	0.75	0.83	0.79	263
Positive	0.93	1.00	0.96	294
Neutral	0.84	0.71	0.77	309

The Neutral class shows a decent precision of 0.84 but a lower recall of 0.71, which means that the model struggles to detect all neutral instances, possibly confusing them with negative or positive classes. The F1-scores reflect this trend, with the highest score in the Positive class and the lowest in the Neutral class. Overall, the model performs well, especially in identifying positive sentiment, but requires improvement in distinguishing neutral sentiment accurately.

4.7 Data Labelling and Data Oversampling

The next stage after the preparation phase is data annotation. Data labelling in this work is done using a lexicon-based sentiment analysis tool. This process calls for looking at a text file's presentation of a lexicon of positive and negative emotions. The lexicon calculates the sentiment score for every tweet collected from Twitter about the 2024 election. Based on the keyword frequency in the positive and negative dictionaries, the sentiment calculation function scores every tweet. Results of this calculation depicted in Figure 4 are used to assign sentiment labels to every tweet: tweets with positive scores are labelled 'Positive', tweets with zero scores are labelled 'Neutral', and tweets with negative scores are labelled 'Negative'.

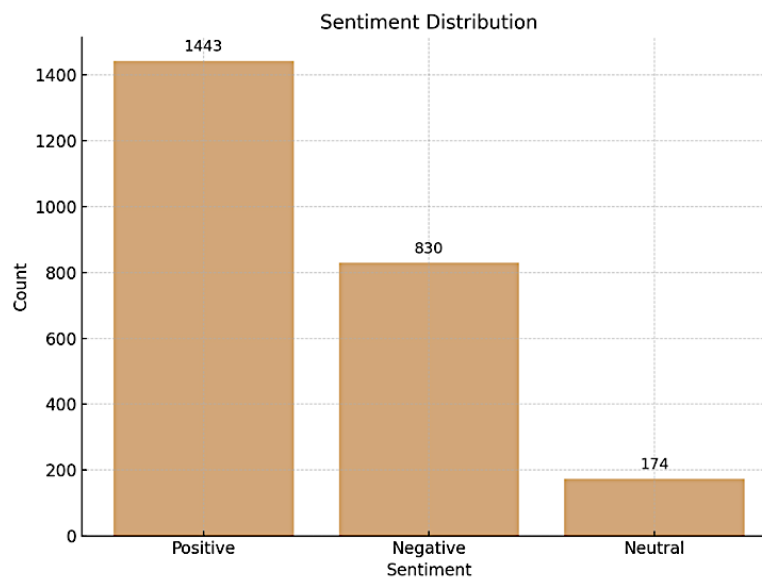


Figure 4. Distribution of Sentiment Data After Labelling

The lexicon used in this work comes from Fajri Koto and Gemala Rahmanyngtyas's study, InSet 'Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs'. This work introduced InSet; a sentiment lexicon especially meant for the examination of microblogs in Indonesia. InSet is an Indonesian sentiment dictionary with 3,609 positive words and 6,609 negative ones. Every word in this lexicon is given a weight based on its polarity, ranging from -5 to +5 [23].

Classifying the data reveals that neutral sentiment data far outnumbers negative and positive sentiment data. LSTM model accuracy in data prediction could suffer from this imbalance. Therefore, by duplicating occurrences of these

emotions, the researcher employs the random oversampling approach to increase the dataset of both negative and positive emotions, so balancing the three sentiment categories. This approach ensures that every category has the same amount of data, therefore improving the performance of the model in sentiment analysis. Improve the balance of the dataset and get it ready for the model training process.

4.8 LSTM model

The LSTM model development for Twitter sentiment analysis in this work includes data partitioning, label encoding, sequence padding, and tokenization. Often limited to 8000 words, the Hard Tokenizer converts text into a numerical sequence; padding then preserves consistent sequence length of 100 words. To allow categorisation, sentiment labels are encoded into numerical values. The dataset was split into training and test data in an 80:20 ratio. This separation ensures the model is evaluated on data it has not previously seen during training. Comprising several layers, the LSTM model features an Embedding Layer with 8000 inputs and 120 outputs, a Convolutional Layer with 200 filters and 9 kernels, two LSTM Layers with 128 and 64 units using a dropout rate of 0.8, and a Dense Layer using softmax activation for multi-class classification. The model is set up using the 'sparse_categorical_crossentropy' loss function and the 'adam' optimiser, including model checkpointing callbacks to maintain the best model. Over 20 epochs, training recorded loss and accuracy for every epoch.

The LSTM model shows strong performance shown in Figure 5 by attaining high training accuracy and keeping steady testing accuracy. Though some ages showed modest instability, the performance showed interesting generalisation to new information. This suggests the programme has completed the opinion investigation task on the used Twitter dataset.

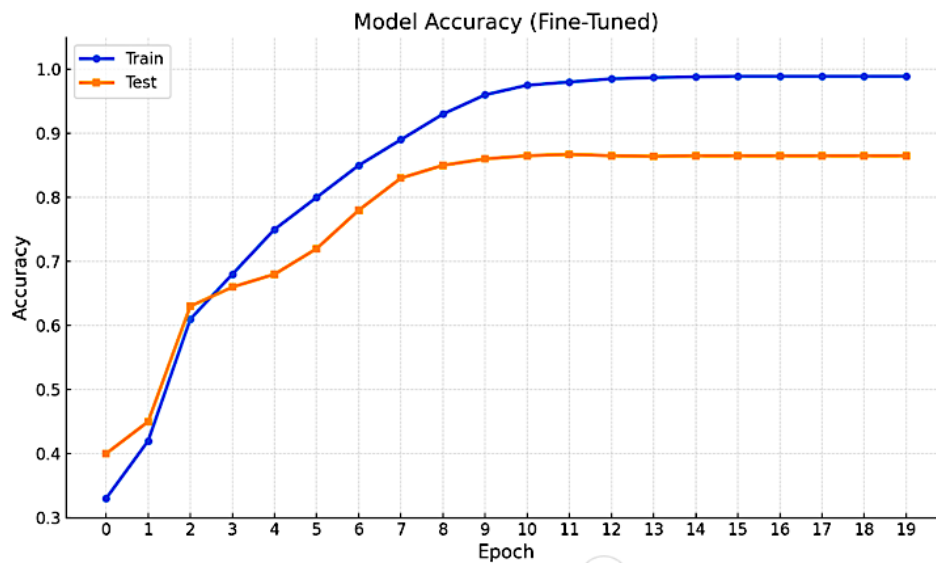


Figure 5. Model Accuracy Graph

4.9 Model Evaluation

The next step is to assess the LSTM model for sentiment analysis to verify its best performance after training. Forecasting sentiment labels for the test data, examining the prediction results via classification reports and confusion matrices, and evaluating the model's general correctness are all included in this evaluation. The results of this study presented in Table 5 offer a comprehensive picture of the model's performance in sentiment classification. Every category is given by the categorisation report accuracy, recall, and F1-score based on The Confusion Matrix depicted in Figure 3. The next table shows the categorisation findings based on the model assessment.

Table 5. Model Evaluation Report

Class	Precision	Recall	F1-Score	Support
Negative	0.75	0.83	0.79	263
Positive	0.93	1.00	0.96	294
Neutral	0.84	0.71	0.77	309
Macro avg	0.84	0.84	0.84	866
Weighted avg	0.84	0.84	0.84	866

The perplexity network offers a graphical delineation of the model's course expectations on the test information. The consequent perplexity framework is delivered from the show evaluation.

Indicating good performance in sentiment categorisation, the LSTM model used for sentiment analysis on the test data attained an accuracy of 84%. The classification report shows that the neutral class has the greatest F1-score, recall, and accuracy, therefore proving very accurate forecasts. Though still within an acceptable range, negative and positive classes show poorer performance defined by lower precision and recall relative to neutral classes. The evaluation using manual metric computations and the confusion matrix shows that this approach is consistent for sentiment analysis on Twitter regarding the 2024 Election.

4.10 Result Analysis

The last stage of this analytical process is to evaluate the trained model's forecast results. The goal is to apply the model on fresh tweets and assess its effectiveness in sentiment prediction depending on the text. Forecasting sentiment generated from tweets is the main goal of this work; thus, this stage is vital for judging the performance of the constructed model. The results of the experiment assessing new tweets using the trained model depicted in Figure 6 are as follows:

```
Text: menurut gw paslon pemilu kali ini lebih bersahabat, merakyat, dan profesional, semoga aja pemilu kali transparan dengan hasilnya
Predicted Sentiment: Positif
```

Figure 6. New Tweet Prediction Results

The results above allow the programme to identify the positive mood of the tweet. The algorithm not only found phrases like "friendly," "popular," and "professional," all of which have good meanings, but also the expectation of "transparency," which also has good connotations.

4.11 Deployment

In this study, Streamlit was employed, as illustrated in Figure 7, to enhance user experience and facilitate practical utilization of the sentiment analysis model. By using Streamlit, users are not required to initiate the process from scratch or retrain the model, thereby making the system more accessible to non-technical users. The Streamlit interface was developed using Python, enabling seamless integration with the trained LSTM model and allowing real-time sentiment predictions to be performed through an intuitive web-based application. A new tweet can be submitted using the sidebar input, which will be transformed into a data frame presented on the main page. Upon submitting a new tweet, the user merely needs to click the predict button to categorise the tweet as neutral, positive, or negative sentiment. The model implemented via Streamlit effectively classified new tweets as positive sentiment. Specifically, it analysed the tweet containing the phrase "Semoga KPU dapat memberikan kinerja maksimal agar tidak terjadi kerusuhan seperti pemilu sebelumnya" (hopefully the 2024 election will run smoothly no matter who the winner is, and hopefully the committee can provide maximum performance so that there will be no riots like the previous elections) categorising it as positive sentiment.

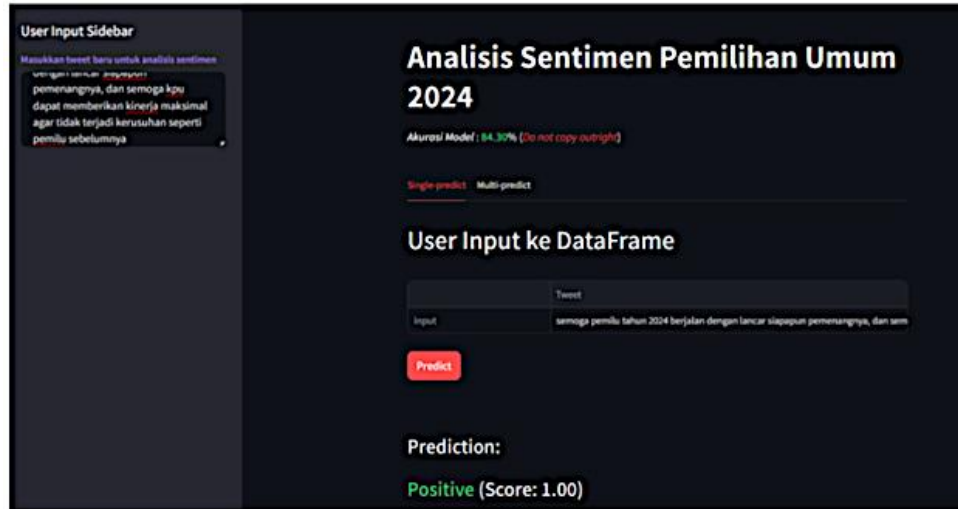


Figure 7. Display of Classification Results Through Streamlit

5. CONCLUSION

This study effectively created LSTM model for sentiment analysis of tweets about the 2024 Election. Data preparation, lexicon-based sentiment tagging, oversampling for class balance, and the creation and training of an LSTM model are all included in the study approach. The built model with embedding layers, Conv1D, and two LSTM layers achieved 84.3% accuracy in sentiment classification. The classification report and confusion matrix evaluation showed sufficient precision, recall, and F1-score values for every category, therefore verifying that this model correctly classified sentiment in election-related tweets. Using a wider spectrum of data to supplement model performance will help development. Furthermore, tweaking hyperparameters and playing with other architectural models like GRUs or Transformers could improve model accuracy. Improved sentiment tagging calls for a more thorough and relevant sentiment vocabulary. This model can be changed into a real-time sentiment analysis tool to provide insights into public opinion on elections and other concerns.

The LSTM model achieved 84% accuracy in classifying election-related tweets. Future research directions include:

- Transformer Models (e.g., BERT) for enhanced contextual analysis.
- Real-Time Monitoring to track sentiment shifts during electoral events.
- Hybrid Architectures (LSTM+CNN) to address class imbalance.

ACKNOWLEDGEMENT

The authors received no funding from any party for the research and publication of this article. The authors would like to thank the anonymous reviewers for their valuable comments.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Angga Wahyu W: Conceptualization, Writing – Original Draft Preparation, Methodology;
Haidar Hilmy Andana: Project Administration, Writing – Review & Editing;
Junta Zeniarja: Validation, Supervision, Writing – Review & Editing;
Aris Febriyanto: Data Curation.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS


Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. <https://publicationethics.org/>. Data anonymization protocols were strictly followed; no personally identifiable information was retained.




REFERENCES

- [1] H. Rosa, "Social Media Filters and Resonances: Democracy and the contemporary public sphere," *Theory, Culture & Society*, vol. 39, no. 4, pp. 17-35, 2022, doi: 10.1177/02632764221103520.
- [2] Y. Theocharis, S. Boulianne, K. Koc-Michalska, and B. Bimber, "Platform affordances and political participation: how social media reshape political engagement," *West European Politics*, vol. 46, no. 4, pp. 788–811, Jun. 2023, doi: 10.1080/01402382.2022.2087410.
- [3] E. Kubin and C. von Sikorski, "The role of (social) media in political polarization: a systematic review," *Annals of the International Communication Association*, vol. 45, no. 3, pp. 188–206, Sep. 2021, doi: 10.1080/23808985.2021.1976070.
- [4] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.
- [5] K. Cortis and B. Davis, "Over a decade of social opinion mining: a systematic review," *Artificial Intelligence Review*, vol. 54, no. 7, pp. 4873–4965, Oct. 2021, doi: 10.1007/s10462-021-10030-2.
- [6] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems*, vol. 226, p. 107134, Aug. 2021, doi: 10.1016/j.knosys.2021.107134.
- [7] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, "An Introductory review of deep learning for prediction models with big data," *Frontiers in Artificial Intelligence*, vol. 3, p. 4, Feb. 2020, doi: 10.3389/frai.2020.00004.
- [8] U. D. Gandhi, P. Malarvizhi Kumar, G. Chandra Babu, and G. Karthick, "Sentiment analysis on Twitter data by using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)," *Wireless Personal Communications*, May 2021, doi: 10.1007/s11277-021-08580-3.
- [9] D. Kumar, S. Bhatia, H. Singh Dhillon, and A. K. Goel, "NLP-based sentiment analysis using deep learning methods," in *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*, Dec. 2024, pp. 1873–1878, doi: 10.1109/ICAC2N63387.2024.10894873.
- [10] N. Mohamed, "Augmented Intelligence: A comprehensive review of the flexibility between human and artificial intelligence," *Journal of The Institution of Engineers (India): Series B*, Mar. 2025, doi: 10.1007/s40031-025-01213-4.
- [11] A. S. Susmitha and P. Pujari, "Sentiment analysis of cyberbullying data in social media," Nov. 08, 2024, *arXiv: arXiv:2411.05958*, doi: 10.48550/arXiv.2411.05958.
- [12] H. N. Irmanda and S. Hartati, "Sentiment analysis of cyberbullying using machine learning," in *2024 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Nov. 2024, pp. 594–600, doi: 10.1109/ICIMCIS63449.2024.10957620.
- [13] L. Bisht and K. Chaudhary, "Exploring machine learning algorithms for sentiment analysis in the sphere of cyberbullying detection," in *2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, Feb. 2025, pp. 446–450, doi: 10.1109/CE2CT64011.2025.10939696.

- [14] Y. Matalon, O. Magdaci, A. Almozlino, and D. Yamin, "Using sentiment analysis to predict opinion inversion in Tweets of political communication," *Scientific Reports*, vol. 11, no. 1, p. 7250, Mar. 2021, doi: 10.1038/s41598-021-86510-w.
- [15] P. Chauhan, N. Sharma, and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2601–2627, Feb. 2021, doi: 10.1007/s12652-020-02423-y.
- [16] S. K. Hamed, M. J. Ab Aziz, and M. R. Yaakub, "Fake news detection model on social media by leveraging sentiment analysis of news content and emotion analysis of users' comments," *Sensors*, vol. 23, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/s23041748.
- [17] I. Karabila, N. Darraz, A. EL-Ansari, N. Alami, and M. EL Mallahi, "BERT-enhanced sentiment analysis for personalized e-commerce recommendations," *Multimed Tools and Applications*, vol. 83, no. 19, pp. 56463–56488, Jun. 2024, doi: 10.1007/s11042-023-17689-5.
- [18] P. Rasappan, M. Premkumar, G. Sinha, and K. Chandrasekaran, "Transforming sentiment analysis for e-commerce product reviews: Hybrid deep learning model with an innovative term weighting and feature selection," *Information Processing & Management*, vol. 61, no. 3, p. 103654, May 2024, doi: 10.1016/j.ipm.2024.103654.
- [19] A. Hariz, I. Bin, A. Azrir, P. Naveen, and S.-C. Haw, "Sentiment Analysis using Machine Learning Models on Shopee Reviews," *Journal of System and Management Sciences*, vol. 14, no. 2, Jan. 2024, doi: 10.33168/jsms.2024.0213.
- [20] D. G. Mandhasiya, H. Murfi, A. Bustamam, and P. Anki, "Evaluation of machine learning performance based on BERT data representation with LSTM model to conduct sentiment analysis in Indonesian for predicting voices of social media users in the 2024 Indonesia Presidential Election," in *2022 5th International Conference on Information and Communications Technology (ICOIACT)*, Aug. 2022, pp. 441–446, doi: 10.1109/ICOIACT55506.2022.9972206.
- [21] A. N. Ma'Aly, D. Pramesti, and H. Fakhurroja, "Comparative analysis of deep learning models for multi-label sentiment classification of 2024 Presidential Election comments," in *2024 7th International Conference on Informatics and Computational Sciences (ICICoS)*, Jul. 2024, pp. 502–507, doi: 10.1109/ICICoS62600.2024.10636889.
- [22] D. F. Putra and Y. Sibaroni, "Sentiment analysis for the 2024 Presidential Election (Pilpres) using BERT CNN," *Eduvest - Journal of Universal Studies*, vol. 4, no. 11, Art. no. 11, Nov. 2024, doi: 10.59188/eduvest.v4i11.49961.
- [23] A. I. Safitri and T. B. Sasongko, "Sentiment analysis of cyberbullying using bidirectional Long Short Term Memory algorithm on Twitter," *Jurnal Teknik Informatika* Volume 5, Number 2, April 2024, vol. 5, no. 2, 2024, doi: 10.52436/1.jutif.2024.5.2.1922.
- [24] F. Koto and G. Y. Rahmangtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *2017 International Conference on Asian Language Processing (IALP)*, Dec. 2017, pp. 391–394, doi: 10.1109/IALP.2017.8300625.

BIOGRAPHIES OF AUTHORS

	<p>Angga Wahyu W. hold master degree from South China University of Technology, currently pursuing Ph.D at Graduate School of Systems Design, Department of Computer Science Tokyo Metropolitan University with a focus on artificial intelligence, machine learning, and augmented reality. His research includes real-time facial detection, gesture classification, IoT-based health monitoring, and educational game development. Angga has contributed to practical applications of technology, such as AR for logistics and fuzzy logic in health tracking. His work is widely recognised, with citations reflecting his impact on academia and industry. Through innovative research, he bridges technology and real-world challenges to advance education and societal solutions. He can be contacted at email: angga-wahyu-wibowo@ed.tmu.ac.jp</p>
---	---

	<p>Haidar Hilmy Andana is a graduate in Computer Science from Dian Nuswantoro University in Semarang, Indonesia. With a strong foundation in computer science principles, he specialises in data science, focusing on machine learning, and artificial intelligence. In addition, he has expertise in website development, combining technical skills with creative problem-solving to build user-friendly and efficient web applications. He can be contacted at email: 111202013180@mhs.dinus.ac.id.</p>
	<p>Junta Zeniarja is a scholar who holds a Master of Computer Science (MCS) degree from Universiti Teknikal Malaysia Melaka (UTeM). He is currently affiliated with Universitas Dian Nuswantoro (UDINUS), Indonesia, where he contributes to academic and research activities in the field of computer science. His research interests include data mining, machine learning, sentiment analysis, and recommendation systems. He has published several works in these areas, including applications of K-Means clustering, k-Nearest Neighbour, and hybrid optimization techniques. Junta's studies have been widely cited, reflecting his significant contributions to advancing computational methods and their applications. He can be contacted at email: junta@dsn.dinus.ac.id.</p>
	<p>Aris Febriyanto is an Undergraduate in Electronic Engineering from Dian Nuswantoro University in Semarang, Indonesia. With a strong foundation in computer science principles, he specialises in IoT (Internet of Things), focusing on artificial intelligence. In addition, he has expertise in electrical development. He can be contacted at email: 511202301194@mhs.dinus.ac.id.</p>