ILKOM Jurnal Ilmiah Vol. 17, No. 2, August 2025, pp. 150-161
Accredited 2nd by KEMDIKTISAINTEK No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716

Research Article

Sentiment Analysis towards Jokowi Post-Presidential Term Using CNN-BiLSTM with Multi-head Attention on Platform X

Muhammad Rizki Setyawan [a,1] ; **Fajar Rahardika Bahari Putra** [a,2,*] ; **Ardhina Ramadhani** [a,3]
[a,] *Informatics Engineering Study Program, Muhammadiyah University of Sorong, Southwest Papua, 98416, Indonesia*
[1] *Rizki@um-sorong.ac.id;* [2] *fajar_rbp@um-sorong.ac.id;* [3] *ardhina012@gmail.com*
* *Corresponding author*

Article history: Received June 25, 2025; Revised July 08, 2025; Accepted August 14, 2025; Available online August 19, 2025

Abstract

The development of social media has changed the way the public expresses political opinions, especially regarding the evaluation of President Joko Widodo's (Jokowi) leadership after his term. Platform X (formerly Twitter) has become the primary source of public opinion data, but the use of informal language and sarcasm makes accurate sentiment analysis challenging. This study creates a sentiment analysis model that uses deep learning with a CNN-BiLSTM structure and a multi-head attention mechanism. The dataset consists of 52,643 tweets that have been labeled and embedded using IndoBERT. To address class imbalance, the SMOTE method was applied to the training data, enabling the model to better learn from minority classes. The results indicate that the model achieves a high accuracy of 98.78%, with an average precision, recall, and F1-score of 0.98. These findings indicate that the model is not only accurate but also reliable in distinguishing each sentiment class. A comparison with other model variants suggests that the complete combination of CNN-BiLSTM and Multi-Head Attention delivers the best performance, although the improvement is relatively small. This model can support policymakers by providing timely insights into public opinion, helping to improve communication and policy decisions.

Keywords: CNN-BiLSTM; Jokowi; Multi-Head Attention; Sentiment Analysis; Platform X.

Introduction

The development of social media has reconstructed the dynamics of political communication within society, particularly in expressing opinions and responding to public policy. As a primary platform for political discourse in Indonesia, Platform X (formerly Twitter) has become a critical space for the public to evaluate the leadership of President Joko Widodo (Jokowi), even after the end of his term in office [1], [2]. Nevertheless, the linguistic characteristics of social media—typically informal, rich in sarcasm, and often ambiguous—present unique challenges in accurately measuring public sentiment. This condition points to the need for analytical approaches capable of decoding the complexity of human language in a contextual manner [3], [4].

In this context, artificial intelligence through the field of natural language processing (NLP) offers an innovative solution for objectively measuring public sentiment [5]. AI-based sentiment analysis not only enables the processing of data at massive scales but also improves interpretive accuracy when compared to conventional survey methods, which are prone to subjective biases [6], [7]. Unlike traditional political surveys limited by small samples and delays, sentiment analysis on social media offers real-time, large-scale insights into public opinion. From a computational political science view, this supports social listening and e-government feedback, enabling policymakers to monitor public mood and political engagement continuously. Such data-driven approaches improve transparency, responsiveness, and foster more adaptive, participatory governance [8].

While earlier studies have employed classical machine learning techniques such as Naïve Bayes, Support Vector Machine, and Random Forest have long been used [9], [10], [11], their reliance on static features like TF-IDF or n-grams limits their capacity to understand contextual word relationships, especially in the Indonesian language, which is characterized by high variation and informality [12], [13]. Consequently, these limitations reduce the effectiveness of classical methods in accurately interpreting sentiment in this domain

The advancement of deep learning has opened new opportunities to overcome these limitations. Convolutional Neural Networks (CNNs), for instance, are effective at extracting local patterns through convolution operations, yet they are less optimal at interpreting long-range word dependencies [14], [15]. On the other hand, Long Short-Term

http://dx.doi.org/10.33096/ilkom.v17i2.2843.150-161

Memory (LSTM) and Bidirectional LSTM (BiLSTM) networks are specifically designed to capture sequential context from both past and future directions within a sentence [16], [17]. A combination of CNN and BiLSTM has thus been adopted to integrate the advantages of spatial feature extraction with temporal context understanding, resulting in a more adaptive model [18], [19], [20].

However, challenges still arise when texts contain dispersed or implicit keywords, necessitating an additional mechanism to filter relevant information. The attention mechanism was introduced to allow the model to assign varying weights to each word, focusing on phrases that most influence sentiment [21], [22]. Further development through multi-head attention enhances this ability by analyzing text from multiple subspace perspectives simultaneously, thereby increasing the model's capacity to capture multidimensional nuances of meaning [23], [24]. Integrating these techniques with the CNN-BiLSTM architecture holds the potential to produce a more robust system, especially for dynamic languages like Indonesian.

Optimization efforts have also been carried out through the use of pre-trained models such as IndoBERT, which was specifically developed to capture the linguistic characteristics of the Indonesian language. Based on the BERT architecture, IndoBERT can interpret the meaning of words based on entire sentence contexts, including idioms, sarcasm, or informal abbreviations that frequently appear on social media [25]. Unfortunately, the exploration of combining CNN-BiLSTM, multi-head attention, and IndoBERT remains limited, particularly in sentiment analysis for the Indonesian language [26], [27].

Based on this background, this study aims to develop a sentiment analysis model by integrating CNN-BiLSTM, Multi-Head Attention, and IndoBERT for labeling and embedding the dataset to evaluate public opinion regarding Jokowi after his presidential term on Platform X. The urgency of this study lies in the need to map public opinion in support of responsive policymaking. Sentiment analysis toward political figures such as Jokowi post-presidency not only reflects retrospective evaluation but also serves as a representation of public expectations for future leadership. The resulting data can serve as an empirical basis for political elites to design appropriate communication strategies while also reinforcing legitimacy through policies that align with public aspirations.

The novelty of this study lies in the hybrid integration of CNN-BiLSTM, multi-head attention, and IndoBERT embedding specifically tailored for Indonesian-language political sentiment analysis. Similar to the hybrid deep-learning strategies proposed in recent studies such as [28], which demonstrated improved performance by combining convolutional and recurrent neural networks for Indonesian text classification, our model advances this approach by incorporating multi-head attention and contextual embedding via IndoBERT. This integration enhances the model's ability to capture both local patterns and long-range dependencies, as well as subtle linguistic nuances including sarcasm and informal expressions common in social media. Consequently, this hybrid model delivers superior accuracy and interpretability compared to prior works, providing a robust computational tool that fills a critical gap in analyzing Indonesian political discourse on social media.

Method

This study aims to design and evaluate a CNN-BiLSTM model with a multi-head attention mechanism in the context of sentiment analysis. The approach used is quantitative, with a data-based computational experiment method to assess the performance of the model in conducting sentiment analysis [29]. The research stages include problem identification, data collection, preprocessing, data labeling and splitting, data embedding and resampling, model development, performance evaluation, and result analysis, as shown in **Figure 1**.

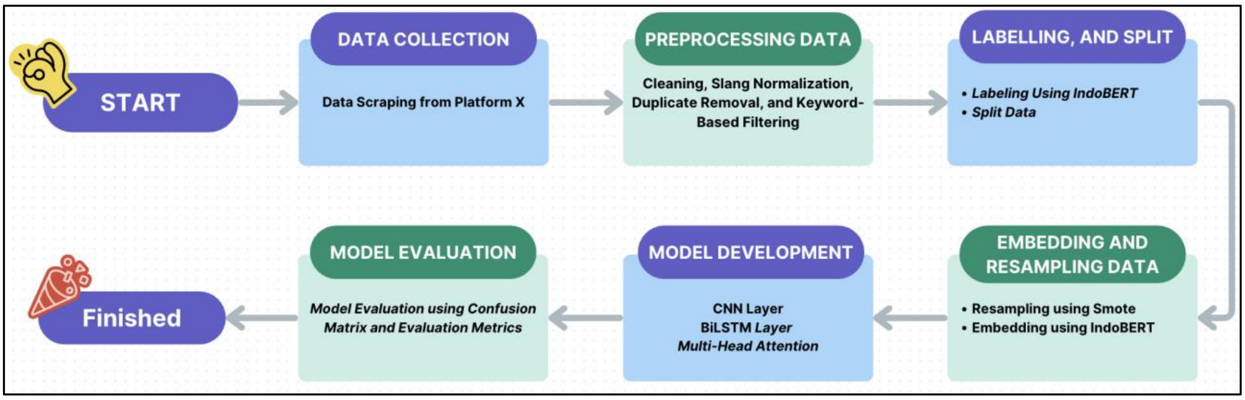


Figure 1. Research Flow

A. Data Collection

Data was collected using scraping methods from Platform X to gather discussions or reviews related to President Jokowi. The collected data consists of tweets containing the keyword "Jokowi" taken from October 2024, after President Joko Widodo officially ended his term, until March 2025.

B. Preprocessing Data

Data preprocessing aims to clean and standardize text for accurate analysis by removing URLs, hashtags, special characters, converting text to lowercase, and normalizing slang into formal words. Duplicate entries are then removed to prevent bias. Additionally, only rows containing keywords such as ‘Jokowi’, ‘Joko Widodo’, ‘Mulyono’, or ‘Presiden’ in the ‘clean_text’ column are retained to ensure dataset relevance .

C. Labelling and Split Data

Data labeling assigns sentiment labels (positive, negative, neutral) using IndoBERT, starting with tokenization to break text into meaningful units. IndoBERT then analyzes context to auto-assign initial labels, which can be manually refined. The dataset is split into training (70%), validation (10%), and testing (20%) sets to enable effective model training and evaluation.

During validation, only automatic inference by IndoBERT is used to assess label consistency and reliability without manual review. This ensures an unbiased evaluation of the labeling quality and helps identify possible misclassifications for further improvement.

D. Embedding and Resampling Data

Embedding uses IndoBERT to convert text into rich contextual vectors, capturing word meanings and relationships for effective deep learning input [30]. To handle label imbalance, Synthetic Minority Oversampling Technique (SMOTE) is applied only to training data, preserving real-world distributions in validation and testing sets [31].

E. Model Development

The model proposed in this study for Indonesian sentiment analysis is a CNN-BiLSTM architecture integrated with multi-head attention. The proposed model architecture diagram is shown in **Figure 2**.

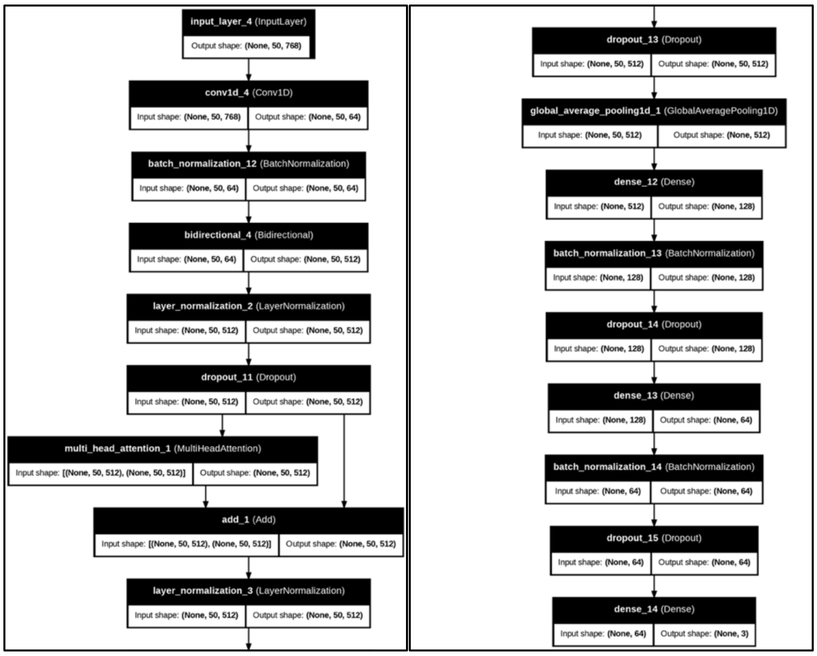


Figure 2. The proposed model architecture diagram

Figure 2 shows the architecture of the proposed model, showing the arrangement of its layers. Below is a detailed description of the model's structural components:

- Input layer: The model receives input sequences with shape (MAX_SEQ_LENGTH=50, EMBEDDING_DIM=768), representing embedded text data with a fixed sequence length and embedding dimension.

- Conv1D layer: Applies 64 filters with a kernel size of 5 and ReLU activation to extract local features from the sequence. Padding is set to ‘same’ to preserve the sequence length.
- Batch Normalization: Normalizes the Conv1D output to speed up convergence and improve training stability.
- Bidirectional LSTM layer: Processes the sequence bidirectionally with 256 units per direction, capturing dependencies from both past and future tokens, outputting a sequence of 512 features per timestep.
- Layer Normalization: Normalizes the BiLSTM output to stabilize training and improve gradient flow.
- Dropout: Applies a dropout rate of 30% to reduce overfitting by randomly deactivating neurons in the BiLSTM output.
- Multi-Head Attention layer: Uses 8 attention heads with key dimension 256 to model complex relationships between tokens across the sequence in parallel.
- Add (Residual Connection): Adds the attention output back to the BiLSTM output to maintain gradient flow and model stability.
- Layer Normalization: Normalizes the combined output from the residual connection.
- Dropout: Applies a 30% dropout rate after attention and normalization to prevent overfitting.
- Global Average Pooling 1D: Aggregates sequence features by averaging across the time dimension, producing a fixed-length vector.
- Dense layer (128 units): Fully connected layer with 128 neurons and ReLU activation for further feature transformation, including L2 regularization to reduce overfitting.
- Batch Normalization: Normalizes the dense layer output for training stability.
- Dropout: Applies a 30% dropout rate to prevent overfitting in the dense layer.
- Dense layer (64 units): Second fully connected layer with 64 neurons and ReLU activation for deeper feature extraction, also with L2 regularization.
- Batch Normalization: Normalizes the output of the second dense layer.
- Dropout: Applies 30% dropout for additional regularization.
- Output Dense layer: Final layer with neurons equal to the number of classes, using softmax activation to output label probabilities for multi-class classification.

In the development of the CNN + BiLSTM model with a Multi-Head Attention mechanism, several key hyperparameters have been determined to optimize training performance and model generalization, as shown in **Table 1**.

Table 1. Parameter setup CNN-BiLSTM

Parameter	Value
LSTM Units	256
Conv1D Kernel Size	5
Conv1D Filters (Kernels)	64
Dropout Rate (global)	0.3
LSTM Dropout	0.4
LSTM Recurrent Dropout	0.4
Attention Heads	8
Dense Units (1st layer)	128
Dense Units (2nd layer)	64
Regularization (L2)	1e-4
Optimizer	AdamW
Learning Rate	0.001
Clip Norm	1.0
Epochs	10
Batch Size	64

F. Model Evaluation

Model evaluation was conducted to assess the overall performance of sentiment classification on the test dataset. This process employed a confusion matrix as the primary tool to visualize the relationship between true labels and the model’s predictions, enabling the identification of classification errors across the sentiment classes (positive, negative, and neutral). Based on the confusion matrix, the evaluation metrics including accuracy, precision, recall, and F1-score were calculated [32]. The following mathematical equations were used for the calculations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = \frac{2(Precision)(Recall)}{(Precision)(Recall)} \quad (4)$$

Where:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

Results and Discussion

a) Result

1. Data Collection

The data collected using scraping methods on Platform X, a total of 96,317 tweets, were successfully gathered, focusing on the keyword "Jokowi" as shown in **Figure 1**. This dataset represents public opinion during the leadership transition phase and serves as the primary foundation for labeling and sentiment analysis.

```
Dataset tweet shape: (96317, 1)
                                           full_text
0  @jokowi Bener banget nih harusnya kita manfaat...
1  @jokowi Setuju banget harusnya kita bisa lebih...
2  @jokowi Mantap nih tingkat kandungan komponen ...
3  @jokowi Keren! Semoga semakin banyak startup I...
4  @jokowi Wah keren banget nih prestasinya Ibu A...
```

Figure 3. Data Collection

Figure 3 shows the collected tweet data, where at this stage the author retained only the column containing the tweet text, while the other columns were removed as they were considered irrelevant for further analysis.

a. Preprocessing Data

After applying the data preprocessing steps—cleaning, normalization, duplicate removal, and keyword-based filtering—the dataset was refined to include only relevant and high-quality entries, resulting in a total of 52,643 tweets. An example of the results from this preprocessing stage is shown in **Table 2**.

Table 2. Example of Preprocessing Results

original	cleaned	normalized	clean_text
@maman1965 @jokowi Iya nih semoga Pak Jokowi bisa mempertimbangkan untuk mengadakan program pengiriman gratis buku ke taman bacaan masyarakat secara rutin. Bisa banget nih bantu para pegiat literasi.	*iya nih semoga pak jokowi bisa mempertimbangkan untuk mengadakan program pengiriman gratis buku ke taman bacaan masyarakat secara rutin bisa banget nih bantu para pegiat literasi*	*iya nih semoga pak jokowi bisa mempertimbangkan untuk mengadakan program pengiriman gratis buku ke taman bacaan masyarakat secara rutin bisa banget nih bantu para pegiat literasi*	*iya nih semoga pak jokowi bisa mempertimbangkan untuk mengadakan program pengiriman gratis buku ke taman bacaan masyarakat secara rutin bisa banget nih bantu para pegiat literasi*

Table 1 shows a sample of the data preprocessing results. It is important to note that the processes of duplicate removal and keyword-based filtering are not illustrated in this table, as these steps involve the elimination of entire records rather than modification of the tweet content itself.

2. Labelling and Split Data

The data labeling results show that IndoBERT successfully categorized the texts into three sentiment classes: positive, negative, and neutral. The label distribution in the dataset is shown in **Figure 4**.

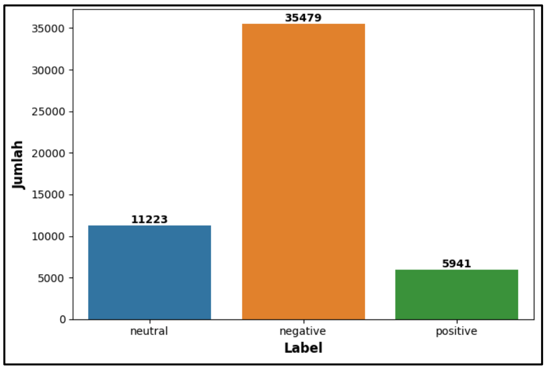


Figure 4. Label Distribution

Figure 4 shows the distribution of sentiment labels in the dataset, with the negative label being the most dominant with 35,479 samples, followed by neutral with 11,223 samples and positive with 5,941 samples. The chart highlights the imbalance in data across the labels.

After labeling, the dataset was split into three subsets consisting of 70% for training, 10% for validation, and 20% for testing. The resulting distribution of samples across these subsets is shown in **Table 3**.

Table 3. Dataset Split Results

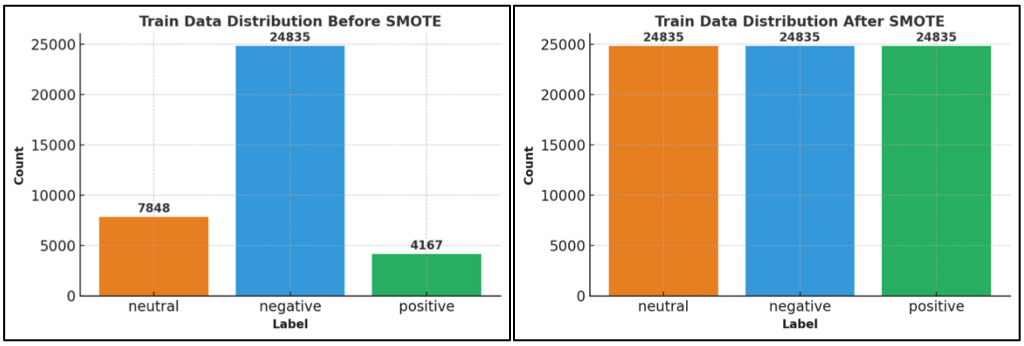
Label	Training (70%)	Validation (10%)	Testing (20%)	Total
Negative	24,835	7,096	3,548	35,479
Positive	4,167	1,205	569	5,941
Neutral	7,848	2,227	1,148	11,223
Total	**36,850**	**10,528**	**5,265**	**52,643**

Table 3 shows the distribution of the dataset across different sentiment labels, with 36,850 samples allocated to training, 10,528 samples to validation, and 5,265 samples to testing.

3. Embedding and Resampling Data

The text embedding process for the training data was conducted using the IndoBERT model, where each sentence was converted into a 50×768 matrix representing 50 tokens, with each token encoded as a 768-dimensional vector. After converting all training data into a three-dimensional format, it was reshaped into two dimensions to enable the application of the Synthetic Minority Oversampling Technique (SMOTE), which was used to address the issue of label imbalance. The oversampled data was then reshaped back into its original three-dimensional form to match the input requirements of the deep learning model. The distribution of the training data before and after applying SMOTE is shown in **Figure 5**.

Figure 5 (a) shows the label distribution in the training dataset before applying SMOTE, where the neutral label contains 7,848 samples, the negative label has 24,835 samples, and the positive label has 4,167 samples. A significant label imbalance is evident, with the negative label being the majority. After applying SMOTE, as shown in **Figure 5 (b)**, the number of samples in each label becomes balanced, with all three classes (neutral, negative, and positive) containing 24,835 samples each. This process aims to improve data balance, enabling the model to learn more fairly across all classes.



(a) Before SMOTE (b) After SMOTE

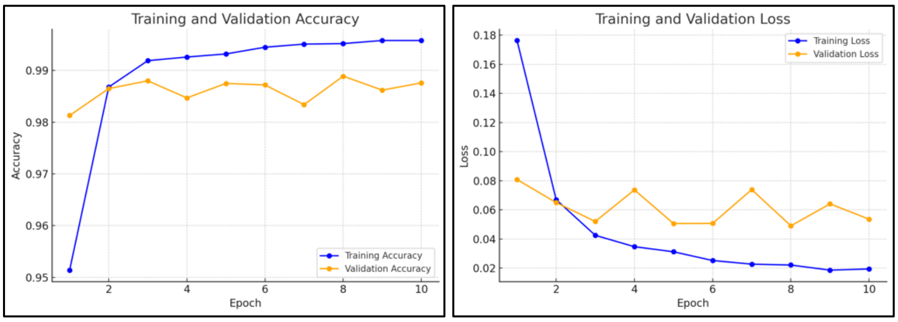
Figure 5. Resampling Dataset

Meanwhile, the validation and testing datasets were also processed through the IndoBERT embedding without applying SMOTE in order to preserve the original label distribution. This entire process produced consistent and meaningful numerical representations, which were subsequently used for validation and evaluating the model.

4. Model Development

The proposed model was trained using the SMOTE-resampled training data, namely X_train_resampled and y_train_resampled. The training process was conducted for 10 epochs with a batch size of 64. During training, the model was also validated using the validation data X_valid_seq and labels y_valid to monitor its performance periodically.

Additionally, the ReduceLROnPlateau callback was employed to automatically reduce the learning rate by a factor of 0.5 if the validation loss (val_loss) did not improve for 3 consecutive epochs. This callback also restores the best model weights based on the validation loss to prevent overfitting and ensure optimal model performance. The training process is displayed in detail with verbose=1. The curves depicting accuracy and loss during both the training and validation stages are shown in **Figure 6**.



(a) Accuracy (b) Loss

Figure 6. Training and Validation Graph

Figure 6 (a) shows the training and validation accuracy over 10 epochs. Training accuracy steadily rises, exceeding 99% by the tenth epoch. Validation accuracy also trends upward, peaking at 98.89% in the eighth epoch, indicating effective learning and good generalization. **Figure 6 (b)** shows the training and validation loss during the same period. Training loss steadily decreases, while validation loss generally declines with minor increases at epochs four and seven. The lowest validation loss (0.04905) occurs at epoch eight, matching the highest validation accuracy. These results suggest the model performs well without significant overfitting.

5. Model Evaluation

Model evaluation was conducted to assess the sentiment classification performance on the testing set. The confusion matrix is shown in **Figure 7**.

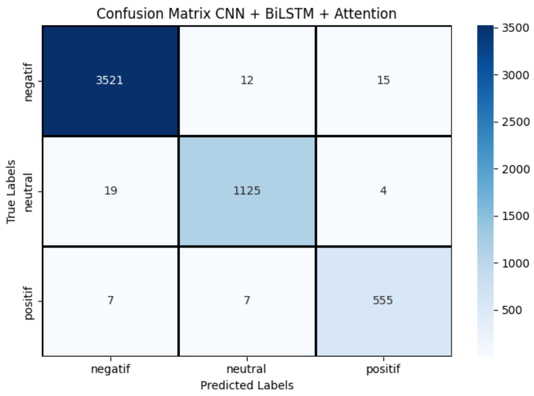


Figure 7. Confusion Matrix

Figure 7 shows the confusion matrix of the proposed model, which is capable of categorizing text into three sentiment classes: negative, neutral, and positive. The model demonstrates very good performance, particularly in the negative class with 3,521 correct predictions. The model also demonstrates good accuracy for the neutral class with 1,125 correct and the positive class with 555 correct, with relatively few misclassifications between classes. This indicates that the proposed model effectively captures important patterns in textual data. The evaluation metrics, including accuracy, precision, recall, and F1 score, derived from the confusion matrix, are shown in **Figure 8**.

```
Akurasi Model CNN + BiLSTM + Attention pada Data Testing: 0.9878

Classification Report CNN + BiLSTM + Attention:
              precision    recall  f1-score   support

     negatif       0.99      0.99      0.99      3548
     neutral       0.98      0.98      0.98      1148
     positif       0.97      0.98      0.97       569

    accuracy                           0.99      5265
   macro avg       0.98      0.98      0.98      5265
weighted avg       0.99      0.99      0.99      5265
```

Figure 8. Evaluation Metrics

Figure 8 shows the performance evaluation of the proposed model in a classification report. The model achieved an accuracy of 98.78% on the test data, with precision, recall, and F1 scores around 0.97 to 0.99 across negative, neutral, and positive classes. Macro and weighted averages are also high, at 0.98 and 0.99, respectively, indicating consistent and accurate sentiment classification.

6. Discussion

This study develops a sentiment classification model using CNN-BiLSTM with multi-head attention and applies SMOTE to balance the dataset. Starting with 96,317 tweets about "Jokowi" collected from Platform X, the data was preprocessed—cleaned, normalized, duplicates removed, and filtered—yielding 52,643 tweets. Sentiment labeling with IndoBERT categorized tweets as positive, negative, or neutral. The data was then split into training, validation, and testing sets. IndoBERT embeddings were generated for the training data, followed by SMOTE to address label imbalance. The proposed model achieved training accuracy exceeding 99%, while validation accuracy peaked at 98.89%, demonstrating strong learning capability and good generalization. On the test set, the model attained an overall accuracy of 98.78%, accurately classifying negative, neutral, and positive sentiments with minimal misclassifications. The average precision, recall, and F1 scores were 0.98 across all classes, confirming the model's high and consistent accuracy.

The research by [33], Applying the SMOTE method to the Support Vector Machine algorithm successfully increased accuracy to 88.51%, while for the Naïve Bayes algorithm, SMOTE improved accuracy to 71.51%. However, the use of SMOTE does not always lead to improved accuracy. Another research by [34] showed that the model without SMOTE actually achieved higher accuracy compared to the one using SMOTE. Meanwhile, research by [35], which implemented a multi-head attention mechanism on a transformer model, managed to improve accuracy to 97.22%. However, the research by [36] also stated that the use of multi-head attention can degrade performance if the number of attention layers is increased beyond a certain point.

Therefore, to evaluate the effectiveness of the proposed model, the researchers also conducted three comparative experiments: (1) a CNN-BiLSTM model without SMOTE and without Multi-Head Attention, (2) a CNN-BiLSTM model with SMOTE but without Multi-Head Attention, and (3) a CNN-BiLSTM model with Multi-Head Attention but without SMOTE. The three models did not undergo significant architectural changes, except for the removal of the multi-head attention component in the versions that did not utilize it. These experiments aimed to evaluate the contribution of each component to the overall performance of the model. The accuracy comparison results of the four configurations are shown in **Table 4**.

Table 4. Comparison of Accuracy Score

	SMOTE	Accuracy	Precision	Recall	F1-Score
CNN-BiLSTM	No	0.9861	0.98	0.98	0.98
CNN-BiLSTM	Yes	0.9875	0.98	0.98	0.98
CNN-BiLSTM Multi-head Attention	No	0.9858	0.98	0.98	0.98
CNN-BiLSTM Multi-head Attention	**Yes**	**0.9878**	**0.98**	**0.98**	**0.98**

Table 4 shows the evaluation results of the CNN-BiLSTM model and the CNN-BiLSTM with multi-head attention, both without and with the application of the SMOTE technique. All models show consistent accuracy, precision, recall, and F1-score values of around 98%. The use of SMOTE shows a slight improvement across all evaluation metrics in both types of models. The proposed model, which is CNN-BiLSTM with multi-head attention and SMOTE, shows the highest accuracy of 98.78%, slightly higher than the other configurations. These results show that the implementation of multi-head attention and SMOTE contributes positively to the model's performance, although the improvements are relatively small.

The addition of multi-head attention in the CNN-BiLSTM model improves its ability to capture complex context in text, such as irony, sarcasm, and informal expressions common on social media. By focusing on multiple parts of the input simultaneously, the model gains richer sentiment representations beyond simple word cues.

However, this added complexity may increase overfitting risks, especially with limited or less diverse training data. Domain-specific language and rare slang can also challenge the model if underrepresented in the dataset. Thus, careful tuning and sufficient data are essential to fully benefit from multi-head attention while maintaining generalization.

To further illustrate the model's capability in capturing complex sentiment expressions, including ambiguity and sarcasm common in social media texts, several example tweets are shown in **Table 5**.

Table 5. Example Tweets Illustrating Model's Handling of Ambiguity and Sarcasm

Tweet	Original Label	Predicted Label	Explanation
infone pas mlebu got ketemu jokowi lorrr	Neutral	Neutral	Model correctly labels this as neutral, capturing the casual and ambiguous tone.
jokowi pun ditahan gibran pun di tahan dan dimakzulkan luhut pun ditahan salahnya dimana ya	Neutral	Neutral	Model accurately classifies this as neutral despite the ironic questioning style.
bahas korupsi jadi ke jet pribadi lawak kamu pantesan jokowi dan anak nya tenang aja toh yg mau ngejatohin orang idiot sdm rendah cuma modal ngetik di x dan terindikasi anak abah	Negative	Negative	Model correctly identifies the negative sentiment in this sarcastic and critical statement.
jokowi tidak takut dosa masih tetap berbohong dan berdusta	Negative	Negative	Model properly classifies this clear negative accusation.
bumil kudu tenang sabar dalam menghadapi netizen dan semoga kelahiran anakmu lancar kalo cowok bisa seganteng dan secerdas mas gibran atau pak jokowi	Positive	Positive	Model successfully labels this positive and supportive message.

Political Interpretation of Findings

The high level of negative sentiment in tweets following the end of the presidency may reflect ongoing public dissatisfaction with the former leader's legacy, decisions, or policies. This sustained negativity suggests citizens continue to critically assess past governance, likely influenced by unresolved socio-economic issues, political controversies, or unmet expectations during and after the presidential term.

Furthermore, the presence of nuanced and sarcastic sentiments indicates a politically charged and polarized environment where opinions are complex and ambivalent rather than simply supportive or oppositional. This reflects an engaged society actively using social media as a platform to express dissatisfaction, skepticism, and calls for accountability. These insights can assist political analysts and policymakers in understanding public sentiment dynamics, guiding communication strategies, and policy decisions to address citizen concerns and rebuild trust.

Conclusion

This study successfully developed a sentiment classification model for tweets about "Jokowi" by combining the CNN-BiLSTM architecture, multi-head attention mechanism, and SMOTE oversampling technique. The model can classify sentiments into positive, negative, and neutral categories with high accuracy, reaching 98.78%, and average precision, recall, and F1-score values of 0.98. These results show that the model is not only accurate but also reliable in distinguishing each sentiment class. Comparison with other model variants indicates that the full combination of CNN-BiLSTM and Multi-Head Attention delivers the best performance, although the improvement is relatively small. Overall, this model is effective for sentiment analysis in the Indonesian language on social media, which features diverse expressions.

Given these findings, the model can be a valuable tool for managing political image in the post-presidency period, aiding policymakers in designing responsive strategies, and monitoring public sentiment toward candidates or political figures. It is recommended to further test and deploy this model in real-time or continuously evolving post-election discourse to capture dynamic public opinion shifts effectively.

Future research may also explore integrating this model with sarcasm detection modules or explainable AI techniques to improve interpretability and handle complex linguistic phenomena more robustly

Acknowledgement

This research is part of the Dikti Research Grant program "Penelitian Dosen Pemula" for the year 2025, funded by Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia.

References

[1] A. N. Ma'aly, D. Pramesti, A. D. Fathurahman, and H. Fakhrurroja, "Exploring Sentiment Analysis for the Indonesian Presidential Election Through Online Reviews Using Multi-Label Classification with a Deep Learning Algorithm," *Inf.*, vol. 15, no. 11, pp. 1–33, 2024, doi: 10.3390/info15110705.

[2] A. A. Firdaus, A. Yudhana, I. Riadi, and Mahsun, "Indonesian presidential election sentiment: Dataset of response public before 2024," *Data Br.*, vol. 52, p. 109993, Feb. 2024, doi: 10.1016/j.dib.2023.109993.

[3] C. G. Kim, Y. J. Hwang, and C. Kamyod, "A Study of Profanity Effect in Sentiment Analysis on Natural Language Processing Using ANN," *J. Web Eng.*, vol. 21, no. 3, pp. 751–766, 2022, doi: 10.13052/jwe1540-9589.2139.

[4] B. A. Pramono and Fitriyani, "Sentiment Analysis of Political Discourse on Platform X Using Graph Neural Network (GNN)," in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, 2024, pp. 302–307. doi: 10.1109/ICICYTA64807.2024.10913222.

[5] S. Gupta *et al.*, "Aspect Based Feature Extraction in Sentiment Analysis using Bi-GRU-LSTM Model," *J. Mob. Multimed.*, vol. 20, no. 4, pp. 935–960, 2024, doi: 10.13052/jmm1550-4646.2048.

[6] R. Olusegun, T. Oladunni, H. Audu, Y. A. O. Houkpati, and S. Bengesi, "Text Mining and Emotion Classification on Monkeypox Twitter Dataset: A Deep Learning-Natural Language Processing (NLP) Approach," *IEEE Access*, vol. 11, no. March, pp. 49882–49894, 2023, doi: 10.1109/ACCESS.2023.3277868.

[7] M. R. Setyawan, F. R. B. Putra, and L. J. Fakhri, "Comparison of Accuracy Level of Certainty Factor Method and Bayes Theorem on Cattle Disease," *Ilk. J. Ilm.*, vol. 16, no. 3, pp. 343–355, 2024, doi: 10.33096/ilkom.v16i3.1943.343-355.

[8] R. H. Ali, G. Pinto, E. Lawrie, and E. J. Linstead, "A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00633-z.

[9] M. Alfreihat, O. S. Almousa, Y. Tashtoush, A. Alsobeh, K. Mansour, and H. Migdady, "Emo-SL Framework: Emoji Sentiment Lexicon Using Text-Based Features and Machine Learning for Sentiment Analysis," *IEEE Access*, vol. 12, no. March, pp. 81793–81812, 2024, doi: 10.1109/ACCESS.2024.3382836.

[10] R. Obiedat *et al.*, "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based

Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.

[11] D. Petrova and V. Bozhikova, "Random forest and recurrent neural network for sentiment analysis on texts in Bulgarian language," in *2021 International Conference on Biomedical Innovations and Applications (BIA)*, 2022, pp. 66–69. doi: 10.1109/BIA52594.2022.9831326.

[12] H. Liu, X. Chen, and X. Liu, "A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis," *IEEE Access*, vol. 10, pp. 32280–32289, 2022, doi: 10.1109/ACCESS.2022.3160172.

[13] C. Lal and Z. Nasir, "Comparative Analysis of Deep Learning Methods in the Realm of Sentiment Analysis," in *2023 International Multi-disciplinary Conference in Emerging Research Trends (IMCERT)*, 2023, pp. 1–3. doi: 10.1109/IMCERT57083.2023.10075107.

[14] F. R. B. Putra, M. R. Setyawan, and L. J. F. Rendra Soekarta, Nabila, "Implementasi Deep Learning Menggunakan Cnn Untuk Klasifikasi Tingkat Kematangan Buah Jeruk Berbasis Android," vol. 8, no. 1, pp. 36–43, 2023, doi: 10.51544/jurnalmi.v9i2.5462.

[15] B. A. Iswenda, Irhamah, and H. Kuswanto, "Convolutional Neural Network for Sentiment Analysis of Indonesian Online News," in *2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C)*, 2024, pp. 693–696. doi: 10.1109/BTS-I2C63534.2024.10941990.

[16] L. Yao, "Sentiment analysis based on CNN - LSTM hotel reviews," *J. Phys. Conf. Ser.*, vol. 2330, no. 1, 2022, doi: 10.1088/1742-6596/2330/1/012018.

[17] T. Ho, H. M. Bui, and T. K. Phung, "A hybrid model for aspect-based sentiment analysis on customer feedback: research on the mobile commerce sector in Vietnam," *Int. J. Adv. Intell. Informatics*, vol. 9, no. 2, pp. 273–285, 2023, doi: 10.26555/ijain.v9i2.976.

[18] L. Khan, A. Amjad, K. M. Afaq, and H. T. Chang, "Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media," *Appl. Sci.*, vol. 12, no. 5, 2022, doi: 10.3390/app12052694.

[19] Y. Dong, X. Li, M. He, and J. Li, "DC-BiLSTM-CNN Algorithm for Sentiment Analysis of Chinese Product Reviews," *Appl. Artif. Intell.*, vol. 39, no. 1, 2025, doi: 10.1080/08839514.2025.2461809.

[20] A. S. Mirdan, S. Buyrukoglu, and M. R. Baker, "Advanced deep learning techniques for sentiment analysis: combining Bi-LSTM, CNN, and attention layers," *Int. J. Adv. Intell. Informatics*, vol. 11, no. 1, p. 55, Feb. 2025, doi: 10.26555/ijain.v11i1.1848.

[21] Shivaranjini, A. Srinivasan, S. B. Anjappa, R. V. Selvakumar, and E. Purushotham, "An Evaluation of the CNN-LSTM Model's Efficacy in Sentiment Analysis Using the Bert and Attendance Mechanisms," *Int. Res. J. Multidiscip. Scope*, vol. 5, no. 2, pp. 822–829, 2024, doi: 10.47857/irjms.2024.v05i02.0659.

[22] Z. Huang, H. Liu, C. Duan, and J. Min, "Customer sentiment recognition in conversation based on bidirectional LSTM and self-attention mechanism," *J. Intell. Fuzzy Syst.*, vol. 46, no. 2, pp. 4417–4428, 2024, doi: 10.3233/JIFS-224183.

[23] A. Kumar, V. T. Narapareddy, V. A. Srikanth, A. Malapati, and L. B. M. Neti, "Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM," *IEEE Access*, vol. 8, pp. 6388–6397, 2020, doi: 10.1109/ACCESS.2019.2963630.

[24] H. Qiu, C. Fan, J. Yao, and X. Ye, "Chinese Microblog Sentiment Detection Based on CNN-BiGRU and Multihead Attention Mechanism," *Sci. Program.*, vol. 2020, 2020, doi: 10.1155/2020/8865983.

[25] F. Muftie and M. Haris, "IndoBERT Based Data Augmentation for Indonesian Text Classification," in *2023 International Conference on Information Technology Research and Innovation (ICITRI)*, 2023, pp. 128–132. doi: 10.1109/ICITRI59340.2023.10250061.

[26] L. Khan, A. Qazi, H. T. Chang, M. Alhajlah, and A. Mahmood, "Empowering Urdu sentiment analysis: an attention-based stacked CNN-Bi-LSTM DNN with multilingual BERT," *Complex Intell. Syst.*, vol. 11, no. 1, pp. 1–14, 2025, doi: 10.1007/s40747-024-01631-9.

[27] M. Wnkhade, C. Sekhar, and A. Abraham, "CBMAFM: CNN-BiLSTM Multi-Attention Fusion Mechanism for

sentiment classification," *Multimed. Tools Appl.*, vol. 83, pp. 1–32, Nov. 2023, doi: 10.1007/s11042-023-17437-9.

[28] C. H. Lin and U. Nuha, "Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00782-9.

[29] T. A. Putra, V. Ariandi, and S. Defit, "Enhancing Accuracy by Using Boosting and Stacking Techniques on the Random Forest Algorithm on Data from Social Media X," *Ilk. J. Ilm.*, vol. 16, no. 2, pp. 184–189, Aug. 2024, doi: 10.33096/ilkom.v16i2.2058.184-189.

[30] C. Owen and D. Suhartono, "Artificial Intelligence-Powered Intent Classification for Indonesian E-Wallet Customer Service Chatbots," in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, 2024, pp. 653–658. doi: 10.1109/ICICYTA64807.2024.10913352.

[31] N. Sholihah, F. F. Abdulloh, M. Rahardi, A. Aminuddin, B. P. Asaddulloh, and A. Y. A. Nugraha, "Feature Selection Optimization for Sentiment Analysis of Tax Policy Using SMOTE and PSO," in *2023 3rd International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)*, 2023, pp. 44–48. doi: 10.1109/ICON-SONICS59898.2023.10435269.

[32] P. Durga and D. Godavarthi, "Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN)," *IEEE Access*, vol. 11, no. October, pp. 108433–108447, 2023, doi: 10.1109/ACCESS.2023.3320738.

[33] Y. B. P. Pamukti and M. Rahardi, "Sentiment Analysis of Bandung Tourist Destination Using Support Vector Machine and Naïve Bayes Algorithm," in *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2022, pp. 391–395. doi: 10.1109/ICITISEE57756.2022.10057802.

[34] G. Tamami, W. A. Triyanto, and S. Muzid, "Sentiment Analysis Mobile JKN Reviews Using SMOTE Based LSTM," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 19, no. 1, p. 13, Jan. 2025, doi: 10.22146/ijccs.101910.

[35] F. Meng and C.-A. Wang, "Sentiment analysis with adaptive multi-head attention in Transformer," *Appl. Comput. Eng.*, vol. 50, no. 1, pp. 133–137, 2024, doi: 10.54254/2755-2721/50/20241326.

[36] X. Li, R. Lu, P. Liu, and Z. Zhu, "Graph convolutional networks with hierarchical multi-head attention for aspect-level sentiment classification," *J. Supercomput.*, vol. 78, no. 13, pp. 14846–14865, 2022, doi: 10.1007/s11227-022-04480-w.