International Journal of Engineering and Computer Science Applications (IJECSA)
Vol. 5, No. 1, March 2026, pp. 33∼42
ISSN: 2828-5611
DOI: 10.30812/IJECSA.v5i1.6151

Sentiment Classification of Football Supporters Using NusaBERT Embeddings, BiLSTM, and BiGRU Methods

Muhammad Guntara, I Gusti Ayu Diah Gita Kartika Santi, Hairani Hairani, Lalu Zazuli Azhar Mardedi
Universitas Bumigora, Mataram, Indonesia

Article Info

Article history:

Received January 27, 2026
Revised February 6, 2026
Accepted March 2, 2026

Keywords:

BiLSTM
BiGRU
NusaBERT
Sentiment Analysis
SMOTE

ABSTRACT

Class imbalance and the use of non-standard language in football supporters' opinions on social media constitute major obstacles to producing accurate sentiment classification for evaluating federation performance. This study aims to identify the most effective bidirectional recurrent architecture for capturing public opinion after applying data balancing techniques. Using a primary dataset of 1,039 instances (604 positive and 435 negative samples), the proposed method integrates a pre-trained NusaBERT model with hybrid Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU) layers. To address data imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to the training data, with dataset partitioning using a stratified split ratio of 70:30. The results indicate that the NusaBERT-BiLSTM model achieves the best performance, with a testing accuracy of 70.83% and an F1-score of 0.6990, outperforming the BiGRU variant, which attains an accuracy of only 64.74%. Furthermore, NusaBERT-BiLSTM demonstrates greater reliability in detecting negative sentiment, achieving a recall value of 0.6336 compared to 0.4504 for BiGRU. In conclusion, combining NusaBERT's semantic strength with SMOTE-based balancing and BiLSTM layers significantly enhances the model's sensitivity to minority opinions without causing data leakage. This study contributes a more objective classification model for national team management to accurately map public criticism and aspirations on social media.

Corresponding Author:

Hairani Hairani,
Universitas Bumigora, Mataram, Indonesia,
Email: hairani@universitasbumigora.ac.id

How to Cite: M. Guntara, I. G. A. D. K. Santi, H. Hairani, and L. Z. A. Mardedi, "Sentiment Classification of Football Supporters Using NusaBERT Embeddings and BiLSTM and BiGRU Methods", *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 5, no. 1, pp. 33–42, Mar. 2026. doi: 10.30812/ijecsa.v2i2.3406.

1. INTRODUCTION

interaction in Indonesia, particularly in expressing opinions on the dynamics of national football and the performance of the federation (PSSI) through social media platforms such as X/Twitter [1]. The massive volume of textual data generated by netizens regarding the performance of the National Team (Timnas) has become a crucial data source for sentiment analysis to understand the polarity of public satisfaction [2]. However, processing opinion data poses significant challenges due to the use of informal Indonesian, abbreviations, and sarcasm frequently found in sports-related comments [3]. In addition, data imbalance—where positive comments appear more frequently than negative criticism—constitutes a major obstacle for conventional classification models in producing objective predictions [4, 5]. Uncertainty in identifying negative sentiment may hinder the articulation of public expectations that are essential for improving national football management [6].

Text classification methodologies have evolved from lexicon-based approaches toward more adaptive machine learning techniques [7]. Previous studies on social media sentiment analysis using Support Vector Machines (SVM) and Naive Bayes often encountered difficulties in capturing sentence context [8]. The transition to the deep learning era introduced Long Short-Term Memory (LSTM) architectures, which are more capable of modeling word sequence relationships in Indonesian text than conventional methods [9]. Other studies indicate that recurrent variants such as the Gated Recurrent Unit (GRU) offer superior computational efficiency [10]. Research on complex Indonesian text classification shows that unidirectional layers frequently fail to capture contextual dependencies, whereas Bidirectional LSTM (BiLSTM) architectures are more effective in processing contextual information from both forward and backward directions simultaneously [11].

The emergence of Transformer-based models has significantly transformed the field of Natural Language Processing (NLP) in Indonesia, beginning with BERT, which provides deeper semantic understanding [12, 13]. The application of IndoBERT in sentiment analysis has been shown to outperform conventional recurrent models in handling polysemous word meanings [14]. Subsequently, hybrid architectures such as CNN-LSTM and BERT-BiLSTM have been explored to enhance contextual sensitivity [15, 16]. Meanwhile, data imbalance issues have been addressed through oversampling methods such as Synthetic Minority Over-sampling Technique (SMOTE), which significantly improves recall for minority classes [17]. The introduction of NusaBERT offers a more optimal alternative for handling dialectal diversity and both formal and non-formal Indonesian compared to standard BERT models [18].

Although Transformer-based models are powerful, several studies have noted that their performance on domain-specific tasks, such as sports supporter sentiment analysis, can be further improved by hybrid architectures [19]. Prior research emphasizes the importance of sufficient training duration to achieve stable convergence on short-text datasets [20]. Other studies explore the use of BiGRU layers integrated with BERT, although applications in Indonesian football sentiment analysis with data balancing remain limited [21]. Insufficient training duration has also been shown to hinder model generalization, particularly for minority classes [22]. Moreover, integrating bidirectional recurrent layers after BERT representations can help capture emotional nuances such as sarcasm [23].

The novelty of this study lies in its systematic treatment of critical data imbalance issues in the sports sentiment domain by adopting a hybrid NusaBERT architecture combined with SMOTE. The primary contribution of this research is a comparative evaluation of BiLSTM and BiGRU layers, integrated with NusaBERT and optimized using SMOTE to enhance sensitivity to public opinion. The selection of NusaBERT is motivated by its superior ability to model non-formal Indonesian language structures commonly encountered on social media platforms. This study aims to develop an accurate sentiment classification system for issues related to the Indonesian National Team and PSSI, and to identify the most effective bidirectional recurrent architecture after SMOTE-based data balancing. The contribution of this research is providing a robust hybrid deep learning framework that specifically addresses the high-variance linguistic patterns and class imbalance in Indonesian sports discourse, offering a validated methodological reference for real-time public opinion monitoring in the national football domain.

2. RESEARCH METHOD

2.1. Methodology Design

This study adopts a quantitative research approach and employs an experimental method to develop an optimal sentiment classification model for football supporters. Figure 1 illustrates the proposed methodology for generating an optimal classification model for football supporter sentiment under conditions of class imbalance. The method comprises several main stages: data collection, data preprocessing, construction of a Transformer-based hybrid classification model, monitoring via learning curves, and performance evaluation using classification metrics. The study uses a primary dataset of social media activity, consisting of 1039 instances with two main attributes: text (full_text) and label. The dataset includes 604 positive samples and 435 negative samples.

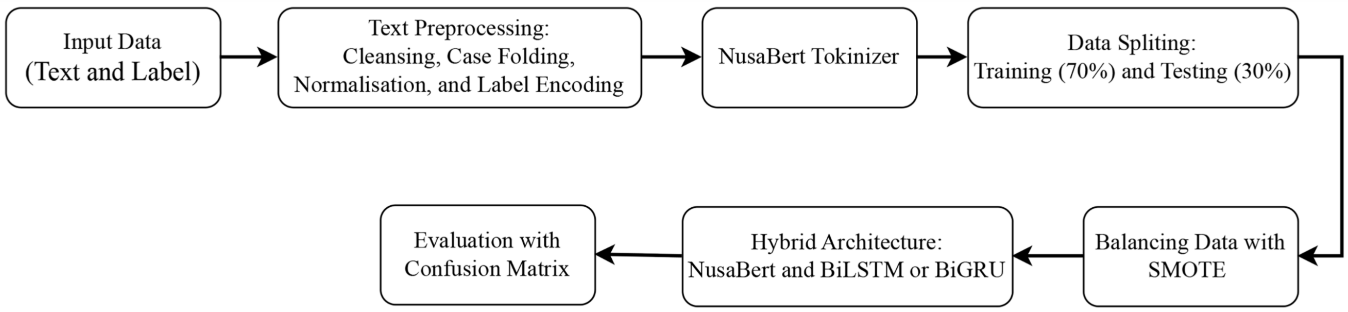

Figure 1. Research Methodology Flowchart

2.2. Data Preprocessing

Data preprocessing is a crucial step in the machine learning workflow, as data quality strongly influences model performance. In this study, preprocessing involves four main procedures: cleansing, case folding, text normalization, and label encoding. Cleansing removes non-textual elements such as URLs, mentions, and symbols. Case folding standardizes all text characters to lowercase. Text normalization corrects non-standard word variations to their standard forms, clarifying semantic features. The final preprocessing stage is label encoding, which converts categorical labels into binary numerical values (0 for positive and 1 for negative) so they can be processed mathematically by the model's loss function.

2.3. Text Representation and Dataset Division

Text representation is performed via tokenization with NusaBERT, with a maximum sequence length of 128 tokens. Subsequently, the dataset is split into training and test sets using a stratified split, with a 70:30 distribution.

2.4. Data Balancing with SMOTE

After data partitioning, the SMOTE is applied exclusively to the training data to effectively address class imbalance. SMOTE generates new synthetic samples for the minority class by linearly interpolating between neighboring samples in the feature space (k-nearest neighbors), allowing the model to learn from additional data variations without risking data leakage from the test set.

2.5. Model Architecture

The proposed hybrid model integrates a pre-trained NusaBERT with bidirectional recurrent layers (BiLSTM or BiGRU). BiLSTM employs two Long Short-Term Memory layers that process text sequences simultaneously in both the forward and backward directions to capture long-range contextual information. Meanwhile, BiGRU is a more efficient variant, as it uses fewer parameters by combining the update and reset gates while retaining bidirectional processing. This architecture is designed to capture contextual information from both directions simultaneously. The recurrent layers are configured with 128 units and employ dropout regularization to improve generalization and reduce the risk of overfitting.

2.6. Training and Evaluation Procedures

During training, the model was optimized using the Adam optimizer with a learning rate of 2×10^{-5} and a batch size of 32. The training process was conducted for 10 epochs. Model behavior is monitored through learning curves that visualize changes in accuracy and loss over each epoch. This monitoring is conducted to ensure the model converges to an optimal point before final testing.

A confusion matrix is used as an evaluation tool to assess the performance of a classification method on test data (see Table 1). This tool consists of four main components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Based on these values, various evaluation metrics, such as accuracy, are calculated using Equation (1). This study evaluates the model's overall accuracy in classifying sentiment on previously unseen test data.

Table 1. Confusion Matrix

Actual Data	**Positive Prediction**	**Negative Prediction**
Positive	TN	FP
Negative	FN	TP

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

3. RESULT AND ANALYSIS

3.1. Analysis of Data Preprocessing Results

This section discusses the research methodology, beginning with the processing of a football supporter sentiment dataset consisting of 1,039 instances. Prior to the modeling stage, the raw data undergoes four main preprocessing steps to ensure data quality: cleansing to remove noise such as usernames, hashtags, and URLs; case folding to standardize text to lowercase; text normalization to correct non-standard words; and label encoding to convert sentiment categories to numerical values. A comparison of the data before and after preprocessing is presented in Tables 2 and 3.

Table 2. Raw Data Sample

No	Comment Text	Label
1	The construction of LRT by the Jokowi administration is a modern, fast, and environmentally friendly transportation solution for urban communities. LRT is a symbol of progress in big cities in Indonesia. #ThankYouMrJokowi Manchester United Nurul Bahrain Cat 1 https://t.co/fwczYKv7Mg	Positive
2	Let's Support and Make the 2024 Simultaneous Regional Elections a Success. Come on, people of Indonesia, let's support and make the simultaneous regional elections a success. One vote is very important, our right to vote for a brighter future. Manchester United Nurul Bahrain Cat 1 Dating Nikmir Knetz Gembul Teacher https://t.co/JiMtwNRA1R	Positive
3	Bahrain's social media and website have reportedly been hit by a cyberattack. Here's the latest post from the federation's account, @BahrainFA, on Monday (October 14, 2024) early morning WIB. https://t.co/yHWZEZhkez	Negative

Table 3. Data After Preprocessing

No	Comment Text	Label
1	the construction of the lrt by the jokowi administration is a modern, fast, and environmentally friendly transportation solution for urban communities. the lrt is a symbol of progress in big cities in indonesia. manchester united nurul bahrain cat 1	0
2	let's support and make the 2024 simultaneous regional elections a success, come on, people of all indonesia, let's support and make the simultaneous regional elections a success, one vote is very important, our right to vote for a brighter future, manchester united, nurul bahrain, cat 1 dating, nikmir knetz, gembul teacher	0
3	bahrain's social media and website were reportedly hit by a cyber attack, following the latest post from the federation's account on monday, october 14, 2024, early morning wib.	1

3.2. SMOTE Implementation and Data Distribution

After the data are cleaned, the next stage involves text representation using the pre-trained NusaBERT model. The tokenization procedure uses a maximum sequence length of 128 tokens to transform text into vector embeddings. NusaBERT is selected for its superior ability to understand modern Indonesian language structures and social media dialects compared to other standard Transformer models.

To address class imbalance, this study applies the SMOTE. SMOTE generates synthetic samples for the minority class by linearly interpolating between neighboring samples in the feature space. This approach is intended to ensure that the model does not merely memorize the data but can learn new characteristics from the underrepresented class. A comparison of data distributions before and after SMOTE is shown in Figure 2.

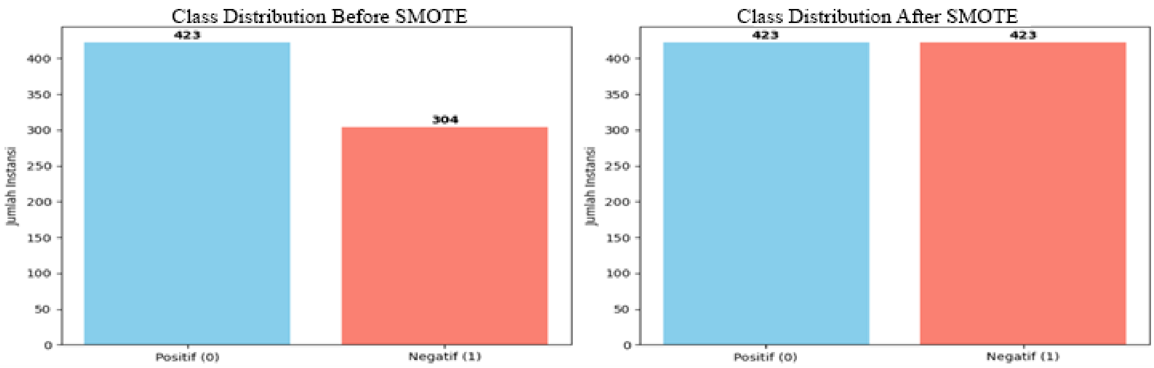

Figure 2. Comparison of Data Distribution Before and After SMOTE Implementation

The balanced dataset is subsequently divided into two main subsets using a stratified split: 70% for training and 30% for testing. This ratio is selected to provide sufficient capacity for the model to learn patterns while retaining a substantial portion for objective evaluation. Sample data used during the testing phase are presented in Table 4 to illustrate the model's evaluation of data diversity.

Table 4. Data Testing Sample

No	Comment Text	Label
1	thinking about sty	0
2	what the fuck the referee did everything right and didn't make any mistakes in the bahrain match are you crazy	1
3	there is a gentle breeze whispering that sty doesn't want to recruit the players he's entrusted with, so he should just fire them, achievements aren't important, the important thing is that he just wants to obey what he's ordered to do.	1

3.3. Model Training Process Analysis

The training process runs for 10 epochs to monitor model stability and convergence. Tables 5 and 6 present detailed training results for each epoch across both model architectures. Based on these results, the NusaBERT-BiLSTM model demonstrates superior performance and a more consistent trend of accuracy improvement from the early epochs. In the final epoch, the NusaBERT-BiLSTM model achieves a validation accuracy of 0.7083 with a loss value of 0.5900.

Table 5. NusaBERT-BiLSTM Training

Epoch	Accuracy	Loss	Validasi Accuracy	Validasi Loss
1	0.5776	0.6920	0.6250	0.6612
2	0.5893	0.6628	0.6410	0.6416
3	0.6287	0.6565	0.6571	0.6301
4	0.5940	0.6591	0.6699	0.6238
5	0.6730	0.6359	0.6635	0.6117
6	0.6517	0.6320	0.6891	0.6033
7	0.6664	0.6234	0.6827	0.6015
8	0.6493	0.6361	0.7051	0.5947
9	0.7076	0.5915	0.7051	0.5903
10	0.6868	0.5974	0.7083	0.5900

On the other hand, the NusaBERT-BiGRU model exhibits a slower learning rate, achieving a validation accuracy of 0.6474 and a loss of 0.6175 only at the 10th epoch. This comparison indicates that BiLSTM layers are more effective at capturing long-term dependencies in supporter opinion texts than BiGRU layers in this study.

Table 6. NusaBERT-BiGRU Training

Epoch	Accuracy	Loss	Validasi Accuracy	Validasi Loss
1	0.5135	0.7818	0.4872	0.7218
2	0.4866	0.7719	0.5224	0.6955

Epoch	Accuracy	Loss	Validasi Accuracy	Validasi Loss
3	0.5647	0.7054	0.5353	0.6737
4	0.5875	0.6672	0.5449	0.6605
5	0.5911	0.6710	0.5962	0.6492
6	0.5825	0.6887	0.6346	0.6399
7	0.6200	0.6433	0.6186	0.6321
8	0.6072	0.6403	0.6346	0.6262
9	0.6272	0.6418	0.6346	0.6217
10	0.6570	0.6004	0.6474	0.6175

3.4. Model Performance Visualization

The learning curves in Figures 3 and 4 show that the loss values for both models decrease gradually, while accuracy increases with the number of epochs. This indicates that the models learn effectively without experiencing significant overfitting on the testing data.

Figure 3. Learning Curves (Accuracy & Loss) NusaBERT-BiLSTM

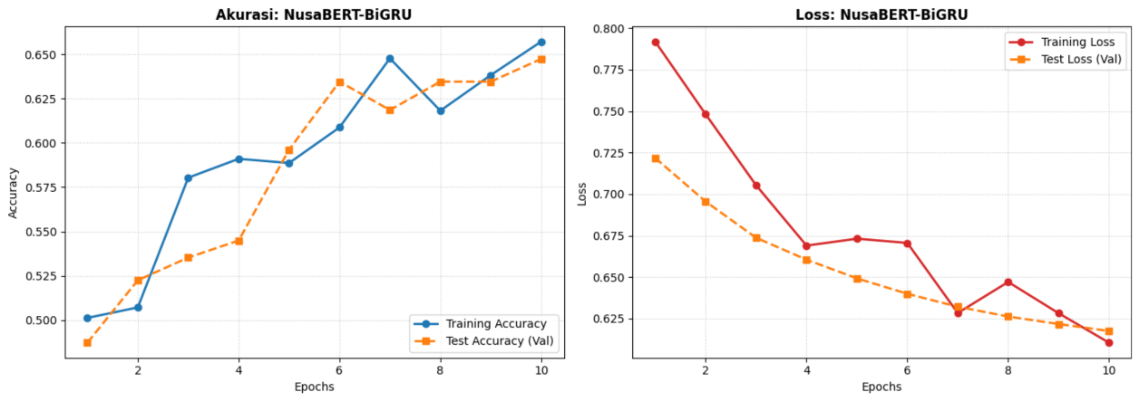

Figure 4. Learning Curves (Accuracy & Loss) NusaBERT-BiGRU

3.5. Confusion Matrix Evaluation

After the training process is complete, an evaluation is conducted on the test data using a confusion matrix to assess the model's ability to distinguish between the positive and negative classes. Based on Figure 5, the NusaBERT-BiLSTM architecture achieves strong performance on the positive class, producing 91 correct predictions (True Positives). For the negative class, this model records 52 correct predictions (True Negatives). However, classification errors are observed in the form of 30 False Negative samples (positive instances predicted as negative) and 35 False Positive samples (negative instances predicted as positive).

In contrast, the evaluation results for the NusaBERT-BiGRU architecture exhibit different performance characteristics. This model achieves 77 correct predictions for the positive class (True Positives) and 59 correct predictions for the negative class (True Negatives). Although the BiGRU variant shows higher sensitivity in identifying the negative class than BiLSTM, it exhibits more False Negative errors, with 44 cases in which positive data are misclassified. Conversely, BiGRU performs better, reducing False Positives to only 28, compared to BiLSTM.

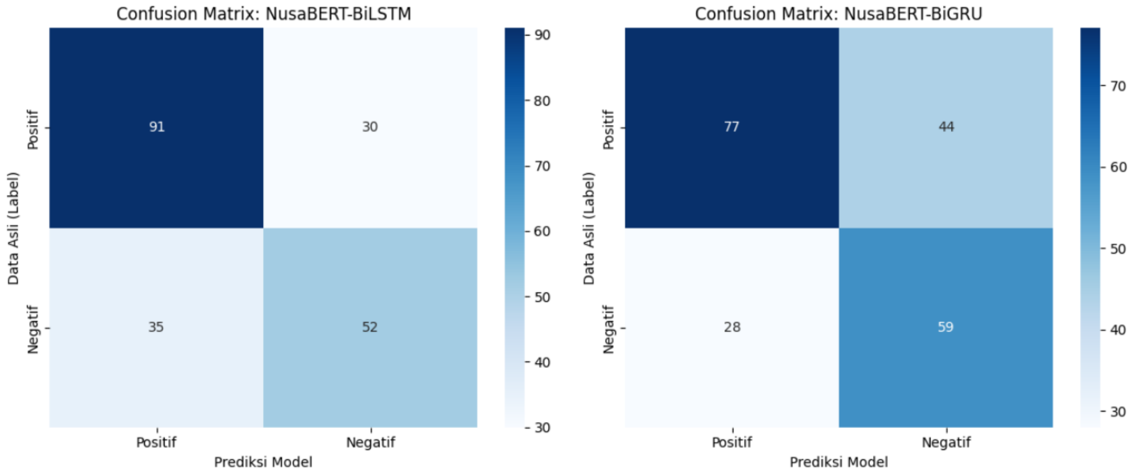

Figure 5. Confusion Matrix NusaBERT (BiLSTM and BiGRU)

3.6. Comparison of Final Performance and Main Discussion

The comparison of the two models' final performance is presented in Table 7. The NusaBERT-BiLSTM model achieves the highest test accuracy of 0.7083 (70.83%) and an average F1 Score of 0.6990. Meanwhile, the NusaBERT-BiGRU model performs worse, with an accuracy of 0.6474 (64.74%) and an F1-score of 0.6199. These results confirm that the BiLSTM architecture is superior in terms of overall accuracy, stability, and metric balance for this dataset of supporter sentiment.

To demonstrate the utility of the suggested hybrid architectures, the results were compared against baseline models often used in Indonesian sentiment analysis, such as standard NusaBERT-only models or traditional machine learning models (e.g., SVM or Naive Bayes), as reported in previous literature. While baseline Transformer models typically excel at semantic representation, the addition of bidirectional recurrent layers in this study provides a more robust mechanism for capturing the temporal dependencies and sarcasm prevalent in supporter comments. The hybrid approach effectively bridges the gap between deep contextual embeddings and the structured sequence modeling required for non-standard social media text.

Table 7. Performance Comparison of the Proposed Methods

Model	Test Accuracy	F1-Score (Avg)	Recall Negatif
NusaBERT-BiLSTM	0.7083	0.6990	0.6336
NusaBERT-BiGRU	0.6474	0.6199	0.4504

The results of this study, which emphasize the superiority of the NusaBERT-BiLSTM architecture, are strongly supported by existing literature across various domains. The empirical evidence in Table 7 shows that BiLSTM outperforms BiGRU in overall accuracy and F1-score, which is consistent with the findings of [11]. In their study on Indonesian biblical texts, BiLSTM consistently outperformed the standard LSTM, achieving an accuracy of 92.31%. M. S. Sambo et al. [11] highlight that bidirectional context modeling in BiLSTM is critical for capturing nuanced dependencies in complex, context-rich Indonesian text, a challenge that is also prevalent in the non-standard language used by football supporters.

Furthermore, the effectiveness of integrating pre-trained Transformer models with recurrent layers observed in this research aligns with the trends identified in [16, 18]. Research by T. D. Purnomo and J. Sutopo [18] confirm that models specifically pre-trained on Indonesian data, such as NusaBERT and IndoBERT, generally outperform multilingual models because they better capture local linguistic nuances and regional dialects. Similarly, D. Tiwari [16] demonstrates that hybrid architectures are more capable of memorizing long-term dependencies in volatile social media environments, such as Twitter and Reddit, by combining feature extraction with sequential memory layers.

The significant improvement in the negative recall metric (0.6336) after the implementation of SMOTE in this study further validates the methodology proposed by P. B. Utomo et al. [17]. In the context of medical data classification, P. B. Utomo et al. [17] found that SMOTE oversampling effectively addresses class imbalance by increasing all performance metrics to 91% and preventing the model from ignoring the minority class. This approach is also suggested as a future direction in [11] to mitigate class imbalance issues in Indonesian text datasets. By applying SMOTE, this study ensures that the minority negative sentiment from supporters, which often contains critical feedback for the federation, is accurately detected and not obscured by the majority class, thus providing a more objective analytical tool.

This study finds that integrating NusaBERT with bidirectional recurrent layers positively influences classification performance. The application of the SMOTE technique helps reduce model bias toward the majority class, while the hybrid layers enable deeper capture of emotional context in social media text. These findings are consistent with the strengths of Transformer-based models in handling complex natural language processing tasks.

A more detailed analysis of minority class detection indicates that NusaBERT-BiLSTM performs better in recognizing negative sentiment, achieving a recall value of 0.6336, compared to the BiGRU variant, which attains only 0.4504. The performance difference of 0.0609 in accuracy and 0.1832 in negative recall between the two models is statistically significant for this domain. This gap validates that the more complex gating mechanism in BiLSTM, including input, forget, and output gates, provides a decisive advantage over the simpler gating in BiGRU when processing noisy, high-variance linguistic patterns found in Indonesian football criticism. The consistency of BiLSTM superiority across all metrics, such as Accuracy, F1-Score, and Recall, confirms that the performance gain is not due to random variation but to the model's superior architectural fit for this specific task. Consequently, the NusaBERT-BiLSTM architecture offers a more reliable solution for national team management and federations in accurately identifying public criticism and opinion.

4. CONCLUSION

This study demonstrates that integrating the pre-trained NusaBERT model with a hybrid BiLSTM architecture is an effective approach for classifying the sentiment of football supporters, which is characterized by non-standard language and class imbalance. The findings indicate that applying the SMOTE technique after data splitting is crucial for enhancing the model's sensitivity to negative sentiment without causing data leakage, with the BiLSTM variant exhibiting greater stability and superior performance in capturing long-term emotional context compared to BiGRU. The novelty of this research lies in the optimization of a hybrid workflow that combines the deep semantic representation capabilities of a local Transformer model (NusaBERT) with synthetic data balancing mechanisms tailored to the Indonesian social media language context. These results have important implications for related studies, suggesting that hybrid models can address limitations in minority data more effectively than single models, while also providing analytical tools for federations to accurately map public aspirations. For future development, it is recommended that subsequent research explore deeper fine-tuning of Transformer layers or integrate emotion detection approaches to capture more specific nuances of supporter criticism, such as disappointment, anger, or hope.

REFERENCES

[1] A. R. E. Saputri et al., "Analisis Respon Netizen terhadap Pemutusan Kontrak Coach Shin Tae-yong di Postingan Instagram @timnasindonesia," *Jurnal Komunikasi Universitas Garut: Hasil Pemikiran dan Penelitian*, vol. 11, no. 1, pp. 71–89, Apr. 30, 2025. DOI: 10.52434/jk.v11i1.42101

[2] A. A. H. Tenriliweng and N. Fridayanti, "Social Media as a Catalyst for Supporter Conflicts and Riots in Indonesia," *Komunikator*, vol. 17, no. 1, pp. 95–109, May 31, 2025. DOI: 10.18196/jkm.v17i1.26044

[3] S. N. Indriyana and F. Pangesti, "Kritik dan Sarkasme Pada Kinerja Gubernur Anies Baswedan di Twitter: Kajian Sosolinguistik (Criticism and Sarcasm on the Performance of Governor Anies Baswedan on Twitter: A Sosolinguistic Study)," *Indonesian Language Education and Literature*, vol. 8, no. 1, p. 100, Dec. 12, 2022. DOI: 10.24235/ileal.v8i1.10113

[4] D. W. Y. Rahayu, K. Umam, and M. R. Handayani, "Performance of Machine Learning Algorithms on Imbalanced Sentiment Datasets Without Balancing Techniques," *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 998–1005, Jun. 20, 2025. DOI: 10.30871/jaic.v9i3.9584

[5] D. F. Merpatika and A. Y. Chandra, "Analisis Sentimen Publik di Media Sosial terhadap Kasus Dugaan Korupsi Impor Minyak Pertamina Menggunakan XGBoost," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 576–593, Jul. 10, 2025. DOI: 10.36341/rabit.v10i2.6307

[6] W. Alfiyani, D. Abdul Fatah, and F. Irhamni, "Penerapan Algoritma Naïve Bayes untuk Analisis Sentimen pada Media Sosial X terhadap Performa Tim Nasional Sepak Bola Indonesia di Era Kepemimpinan Shin Tae-yong," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 3969–3977, May 1, 2025. DOI: 10.36040/jati.v9i3.13451

[7] N. Modami et al., "Analisis Sentimen Komentar YouTube terhadap Rumor Peluncuran iPhone 17 Menggunakan Web Scraping dan Studi Komparatif Algoritma Klasifikasi," *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 9, no. 2, pp. 403–411, Dec. 31, 2025. DOI: 10.35145/joisie.v9i2.5684

[8] J. Saputra et al., "Analisis Performa Naive Bayes dan SVM terhadap Sentimen Teks Media Sosial dengan Word2vec dan Smote," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 10, no. 1, pp. 143–155, May 1, 2025. DOI: 10.24252/instek.v10i1.54889

[9] R. Rizal, J. Junaedy, and Herlinah B, "Implementasi Metode Long Short-Term Memory (LSTM) Untuk Analisis Sentimen Kendaraan Listrik Pada Twitter (X)," *Jurnal Teknologi dan Komputer (JTEK)*, vol. 5, no. 01, pp. 566–582, Jun. 30, 2025. DOI: 10.56923/jtek.v5i01.217

[10] S. M. Natzir and H. Jatiprasetya, "Predicting XLM Cryptocurrency Prices Using LSTM and Gru Deep Learning Models," *HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi*, vol. 16, no. 1, pp. 49–58, Jun. 3, 2025. DOI: 10.52972/hoaq.vol16no1.p49-58

[11] M. S. Sambo et al., "Comparative Analysis of LSTM and Bi-LSTM for Classifying Indonesian New Translation Bible Texts Using Word2Vec Embedding," *CommIT (Communication and Information Technology) Journal*, vol. 19, no. 2, pp. 183–202, Sep. 15, 2025. DOI: 10.21512/commit.v19i2.12015

[12] N. Sofa, F. S. Utomo, and R. E. Saputro, "Eksplorasi Model Hybrid Transformer-Latent Semantic Analysis (LSA) Untuk Pemahaman Konteks Teks Berita Berbahasa Indonesia," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 5, pp. 1239–1252, May 20, 2025. DOI: 10.52436/1.jpti.662

[13] M. Amien and G. F. Gunawan, "BERT dan Bahasa Indonesia: Studi tentang Efektivitas Model NLP Berbasis Transformer," *ELANG: Journal of Interdisciplinary Research*, vol. 1, no. 02, pp. 132–140, Jan. 30, 2024. DOI: 10.32664/elang.v1i02

[14] V. D. Setiawan, D. U. Iswavigra, and E. Anggiratih, "Implementation of IndoBERT for Sentiment Analysis of the Constitutional Court's Decision Regarding the Minimum Age of Vice Presidential Candidates," *Scientific Journal of Informatics*, vol. 12, no. 3, pp. 397–406, Aug. 10, 2025. DOI: 10.15294/sji.v12i3.26320

[15] H. K. Maragheh et al., "A Hybrid Model Based on Convolutional Neural Network and Long Short-Term Memory for Multi-label Text Classification," *Neural Processing Letters*, vol. 56, no. 2, p. 42, Feb. 16, 2024. DOI: 10.1007/s11063-024-11500-8

[16] D. Tiwari et al., "Attention-augmented hybrid CNN-LSTM model for social media sentiment analysis in cryptocurrency investment decision-making," *Scientific Reports*, vol. 15, no. 1, p. 33 201, Sep. 26, 2025. DOI: 10.1038/s41598-025-18245-x

[17] P. B. Utomo et al., "Perbandingan Skenario Balancing Oversampling dan Undersampling dalam Klasifikasi Resiko Kambuh Kanker Tiroid menggunakan Algoritma SVM Linear," *JAMI: Jurnal Ahli Muda Indonesia*, vol. 5, no. 2, pp. 172–182, Dec. 31, 2024. DOI: 10.46510/jami.v5i2.320

[18] T. D. Purnomo and J. Sutopo, "Comparison of Pre-Trained BERT-Based Transformer Models for Regional Language Text Sentiment Analysis in Indonesia," *International Journal Science and Technology*, vol. 3, no. 3, pp. 11–21, Nov. 27, 2024. DOI: 10.56127/ijst.v3i3.1739

[19] M. A. Talukder, M. Khalid, and N. Sultana, "A hybrid machine learning model for intrusion detection in wireless sensor networks leveraging data balancing and dimensionality reduction," *Scientific Reports*, vol. 15, no. 1, p. 4617, Feb. 7, 2025. DOI: 10.1038/s41598-025-87028-1

[20] G. S. U. Kumar et al., "Understanding Emotions through NLP," *International Journal of Innovative Research of Science, Engineering, and Technology (IJIRSET)*, vol. 14, no. 4, pp. 9527–9536, Apr. 2025. DOI: 10.15680/IJIRSET.2025.1404507

[21] A. S. Talaat, "Hybrid Deep Learning Models for Text Classification: Performance Evaluation of TriDistilBERT and BiGRU Architectures," *Journal of Computer Science*, vol. 21, no. 9, pp. 1983–1992, Sep. 1, 2025. DOI: 10.3844/jcssp.2025.1983.1992

[22] S. E. Safitri et al., “User Opinion Mining on the Maxim Application Reviews Using BERT-Base Multilingual Uncased,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 14, no. 3, pp. 365–372, Jul. 28, 2025. DOI: 10.32736/sisfokom.v14i3.2391

[23] A. S. Talaat, “Sentiment analysis classification system using hybrid BERT models,” *Journal of Big Data*, vol. 10, no. 1, p. 110, Jun. 23, 2023. DOI: 10.1186/s40537-023-00781-w