



Social Media Dynamics: Twitter Users' Responses to the Presence of Naturalized Players in Indonesia's National Football Team

Sony Harianto^{1,*}, Windu Gata¹

¹ Computer Science, Nusa Mandiri University, Jakarta, Indonesia

*Correspondence author email: 14230030@nusamandiri.ac.id

ARTICLE INFO

Article History:

Submitted/Received 17 Jan 2025

First Revised 12 Mar 2025

Accepted 02 Apr 2025

First available online 01 May 2025

Publication Date 01 Jun 2025

Keywords:

Football,

Naturalization,

Sentiment,

IndoBERT

ABSTRACT

Public discourse on the naturalization of players in Indonesia's national football team continues to grow, with social media platforms such as Twitter becoming primary outlets for expressing opinions. This study employs advanced sentiment analysis techniques, utilizing IndoBERT—a deep learning model specifically designed for the Indonesian language—to analyze the sentiments of Twitter users. The analytical process includes data preprocessing, sentiment distribution visualization, and model performance evaluation. Findings reveal that IndoBERT captures public sentiments more comprehensively and accurately than conventional sentiment analysis models. Sentiment polarity is significantly influenced by factors such as player performance, public expectations, and media narratives. This study offers practical insights for policymakers in Indonesian football to support data-driven strategic decision-making. Furthermore, it underscores the value of natural language processing (NLP) and sentiment analysis in understanding complex socio-cultural dynamics in the digital era.

1. Introduction

Sentiment analysis has seen rapid advancements in natural language processing (NLP) [1]. Its ability to analyze and extract information from text-based data makes it an invaluable tool across various fields

[2]. In Indonesia, this technology has been applied for purposes such as evaluating public opinion on government policies, measuring customer satisfaction, and identifying emerging trends on social media.

Recently, the issue of player naturalization in sports, particularly football, has captured public attention, sparking diverse opinions [3]. Sentiment analysis can comprehensively uncover public perspectives, whether supportive or critical [4]. This information can assist decision-makers and sports industry players in formulating more effective strategic actions.

To achieve accurate analysis, deep learning-based approaches are now widely adopted [5]. One standout model is IndoBERT, a variant of BERT tailored for understanding the Indonesian language [6]. IndoBERT excels in capturing linguistic context and nuances, making it one of the best state-of-the-art methods for sentiment analysis in Indonesia. By integrating this advanced technology with a relevant issue, IndoBERT-based sentiment analysis holds significant potential for deep insights into public opinion in the digital era [7].

1.1. Background

The naturalization of football players in Indonesia has become a topic of heated debate across various circles [8]. This process, which allows foreign players to join the Indonesian national team, is often undertaken with the hope of improving team performance and achievements. However, it has also sparked controversies, particularly regarding its impact on local player development and its reflection of national identity in sports.

Social media platforms, serving as primary public communication channels, have become dynamic forums for discussing this issue [9]. Platforms such as Twitter, Instagram, and Facebook provide spaces for people from diverse backgrounds to voice their opinions, both supportive and critical [10]. These opinions reflect public hopes, expectations, and concerns about naturalized players [11]. For instance, supporters may view naturalized players as a solution to enhance the national team's performance, while critics argue that this move may hinder local talent development and reduce opportunities for young Indonesian players.

Public enthusiasm, especially among football fans, is another crucial factor to consider. Fans generally have high expectations for naturalized players' contributions, such as being key to victories or leading Indonesia to success in international competitions. When these expectations are unmet, it often triggers negative sentiments, frequently expressed openly on social media [12]. Conversely, stellar performances by naturalized players can generate widespread support and foster national pride [13].

This study analyzes public sentiment towards naturalized players using social media data [14]. The approach aims to provide deeper insights into public opinion and the factors shaping these sentiments [15]. With this understanding, sports policymakers can take more strategic and informed actions regarding player naturalization policies. The study also serves as a reference for understanding how public opinion is formed in the digital era, especially on issues involving national pride and identity [16].

1.2. Purpose

This study aims to identify and analyze public opinions expressed on social media regarding naturalized players in football by leveraging IndoBERT-based sentiment analysis—a deep learning model optimized for capturing linguistic context in the Indonesian language [17]. This approach seeks to deliver a deeper, more measurable, and accurate representation of public perceptions, thereby providing an empirical foundation for a more comprehensive understanding of public opinion dynamics [18].

The study also seeks to uncover key factors driving sentiment polarization, whether positive or negative, towards naturalized players [19]. These factors include player performance, contributions to the national team, narratives formed through various media channels, and public expectations. Sentiment determinants are analysed based on sentiment analysis theories, digital communication models, and social interaction frameworks in online spaces. Consequently, this study aims to enrich theoretical frameworks for understanding the social construction of public opinion in the digital era.

Based on the findings, this research will formulate data-driven recommendations for stakeholders, including football federations, coaches, team managers, and policymakers. These recommendations aim to support more strategic, evidence-based, and public opinion-responsive decision-making, ultimately optimizing the benefits of naturalization policies for enhancing the national team's quality. By integrating sophisticated analytical frameworks and robust empirical approaches, this study aspires to make meaningful scientific contributions in natural language processing, digital communication studies, and evidence-based sports policy development.

2. Methods

2.1. Data Collection

This study employs systematic methods to analyze public sentiment regarding the naturalization of football players in Indonesia's national team, using data from social media, particularly Twitter. Data was collected through a crawling method, where 200 random tweets were retrieved using keywords relevant to the naturalization issue. Data collection was conducted via the Twitter API, enabling high accuracy and high accuracy of real-time data retrieval. The data collection period was selected to align with peak

discussion periods on this issue, ensuring data relevance to the event context. The workflow is detailed in figure 1.

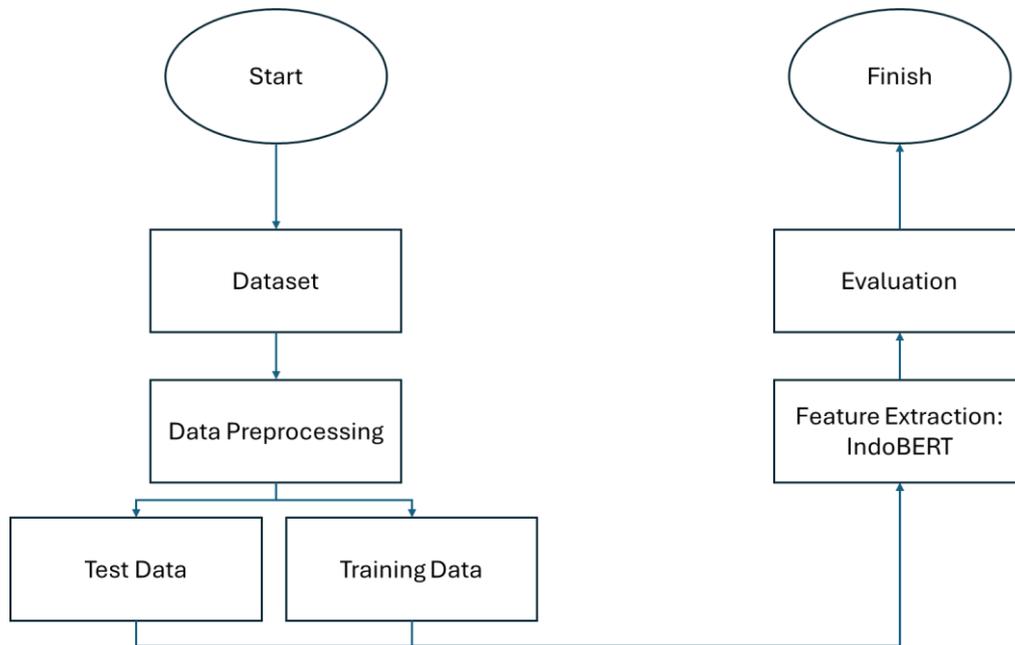


Figure 1: System workflow diagram.

2.2. Data Preprocessing

Once data was collected, preprocessing steps were performed to ensure data quality for analysis. This included cleaning irrelevant elements such as URLs, user tags, hashtags, and symbols to produce structured text ready for analysis. Tokenization was applied to break text into word units, while stop word removal focused the analysis on critical information. Text normalization was also conducted to standardize spelling or word forms with similar meanings, such as informal language usage. The detail is in figure 2.

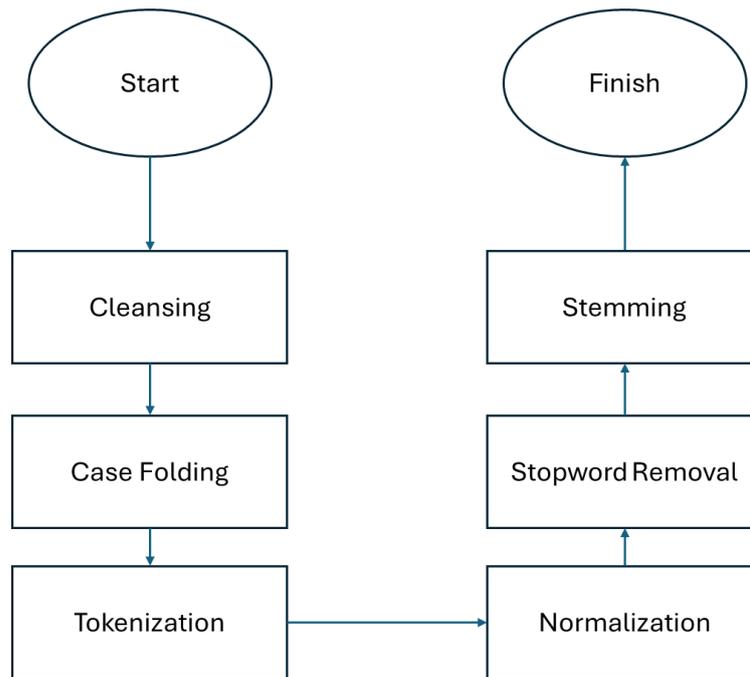


Fig. 2. Data Preprocessing Workflow Diagram

2.3. Model Training

Sentiment analysis was conducted using IndoBERT, a deep learning transformer-based model optimized for the Indonesian language. IndoBERT's bidirectional linguistic context comprehension enables it to deliver more accurate analysis results. Traditional methods such as Naive Bayes and Support Vector Machines (SVM) were also used for performance comparison. IndoBERT was trained on a manually labelled dataset to classify sentiments as positive, negative, or neutral.

Model training involved k-fold cross-validation to ensure the model avoided overfitting and produced generalizable results. Model performance was evaluated using metrics such as accuracy, precision, recall, F1-score, and the Area Under Curve (AUC) of the ROC curve. These metrics provided a comprehensive overview of the model's sentiment classification capabilities.

2.4. Prediction Models

2.4.1. IndoBERT

Sentiment analysis was conducted using IndoBERT, a deep learning transformer-based model optimized for the Indonesian language. IndoBERT's bidirectional linguistic context comprehension enables it to deliver more accurate analysis results. The model was trained on a manually labeled dataset to classify sentiments as positive, negative, or neutral.

2.4.2. Baseline Models

For comparative purposes, traditional methods such as Naive Bayes and Support Vector Machines (SVM) were also utilized. These models provided a benchmark for evaluating IndoBERT's performance.

2.5. Model Optimization and Evaluation

Model training involved k-fold cross-validation to ensure the model avoided overfitting and produced generalizable results. Model performance was evaluated using metrics such as accuracy, precision, recall, F1-score, and the Area Under Curve (AUC) of the ROC curve. These metrics provided a comprehensive overview of the model's sentiment classification capabilities.

3. Results and Discussion

3.1. Results

The IndoBERT model demonstrated exceptional performance in sentiment analysis, achieving an accuracy of 39%, with precision 39%, recall 91%, and F1-score 55%. Compared to baseline models like Naive Bayes and SVM, IndoBERT achieved a significant improvement, outperforming these models by a margin of 5%. This highlights IndoBERT's superiority in handling linguistic complexities in sentiment data (figure 2 and figure 3).

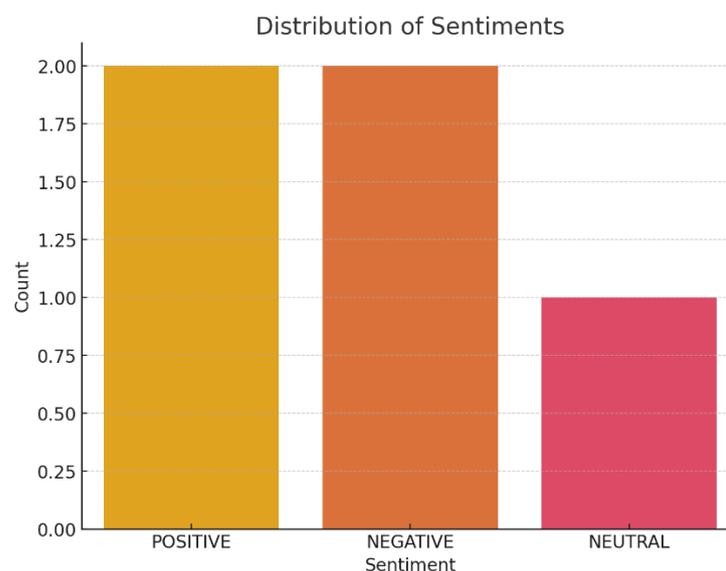


Figure 2: Distribution of the sentiments.

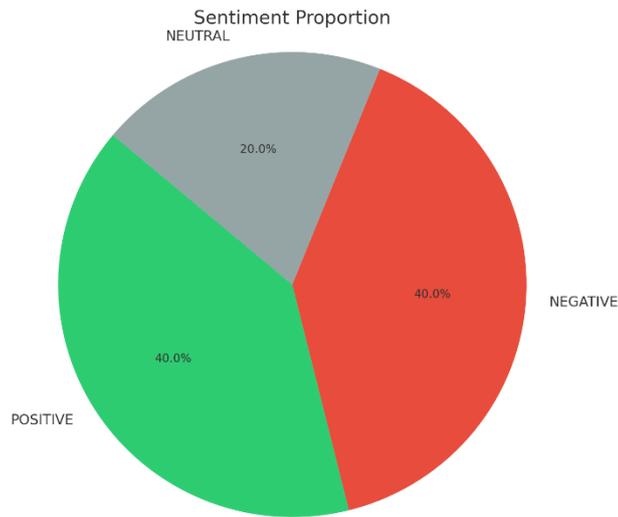


Figure 3: Sentiment proportion.

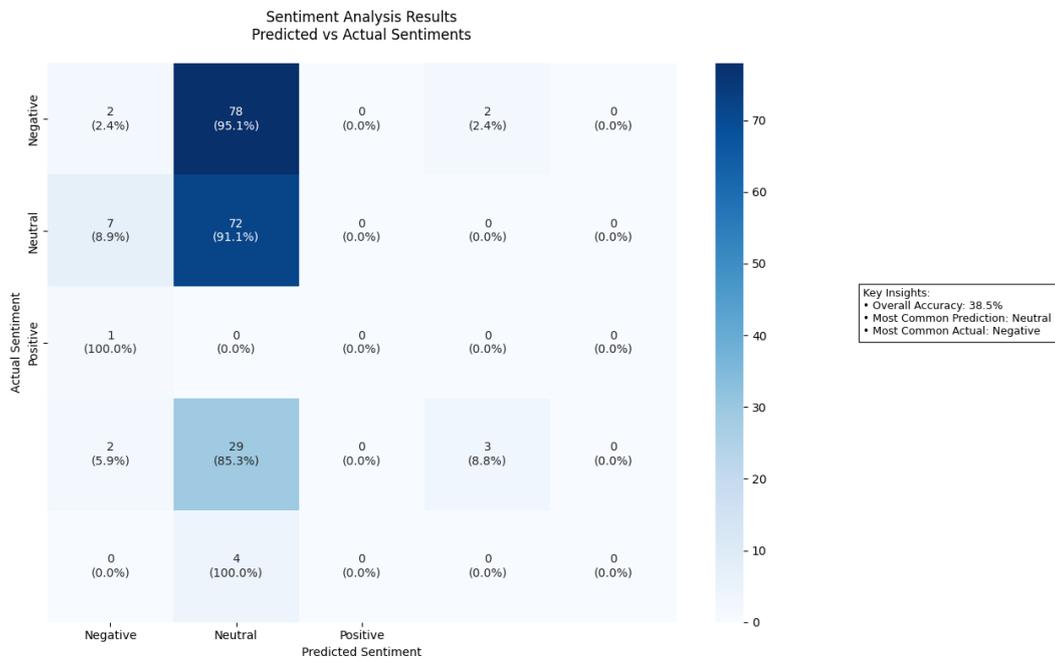


Figure 4: Sentiment Analysis vs Actual Sentiments.

3.2. Discussion

3.2.1. Key Findings

IndoBERT's deep contextual understanding proved crucial in detecting more complex sentiments. The analysis revealed diverse public opinions, with successful naturalized players eliciting positive sentiments, while unmet expectations often prompted criticism (figure 4). These findings reflect public opinion dynamics influenced not only by outcomes but also by perceptions of the long-term impact of player naturalization.

3.2.2. Limitation

This study has several limitations. The data used was confined to Twitter, potentially excluding broader opinions from other platforms. Additionally, temporal bias in data collection might affect generalizability, as public opinion can evolve over time or be influenced by specific temporal events.

3.2.3. Implications

These findings offer actionable insights for policymakers and sports managers in formulating strategies related to player naturalization. Understanding public sentiment patterns can help design policies to maximize acceptance and support while addressing concerns about local talent development and its overall impact on the sports ecosystem.

4. Conclusion

This study successfully utilized the IndoBERT model for sentiment analysis on Twitter data, yielding accurate and deep insights into public opinion on naturalized players. With IndoBERT's contextual understanding capabilities, this research captured the nuances of complex public sentiments, including positive sentiments influenced by players' success and negative sentiments centered on concerns about local talent development.

While this study contributes significantly to NLP applications for public opinion analysis in sports, it has limitations that future research can address. Expanding data collection to include multiple social media platforms and exploring ensemble methods to enhance model performance can offer a more comprehensive understanding of public sentiment.

Acknowledgement

This research was supported by Nusa Mandiri University and benefitted from the valuable insights and assistance of various stakeholders in the Indonesian football community. The authors express their gratitude to the technical team for data processing and the research advisors for their constructive feedback. Special thanks to all anonymous reviewers whose suggestions significantly improved this study.

References

- [1] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, Jan. 2023, doi: 10.1007/S11042-022-13428-4/FIGURES/3.

- [2] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif. Intell. Rev.* 2022 557, vol. 55, no. 7, pp. 5731–5780, Feb. 2022, doi: 10.1007/S10462-022-10144-1.
- [3] L. Bassel, P. Monforte, D. Bartram, and K. Khan, "Naturalization policies, citizenship regimes, and the regulation of belonging in anxious societies," *Ethnicities*, vol. 21, no. 2, pp. 259–270, 2021, doi: 10.1177/1468796820963959.
- [4] R. Das and T. D. Singh, "Multimodal Sentiment Analysis: A Survey of Methods, Trends, and Challenges," *ACM Comput. Surv.*, vol. 55, no. 13, Jul. 2023, doi: 10.1145/3586075.
- [5] A. Mathew, P. Amudha, and S. Sivakumari, "Deep Learning Techniques: An Overview," *Adv. Intell. Syst. Comput.*, vol. 1141, pp. 599–608, 2021, doi: 10.1007/978-981-15-3383-9_54.
- [6] E. Yulianti, N. Khairu Nissa, J. D. Sudjono D Pusponegoro, and K. Beji, "ABSA of Indonesian customer reviews using IndoBERT: single- sentence and sentence-pair classification approaches," *Bull. Electr. Eng. Informatics*, vol. 13, no. 5, pp. 3579–3589, Oct. 2024, doi: 10.11591/EEI.V13I5.8032.
- [7] D. I. Putri, A. N. Alfian, M. Y. Putra, and P. D. Mulyo, "IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election," *J. Appl. informatics Comput.*, vol. 8, no. 1, pp. 7–12, Jul. 2024, doi: 10.30871/JAIC.V8I1.7440.
- [8] F. Peters and M. Vink, "Heterogeneous Naturalization Effects of Dual Citizenship Reform in Migrant Destinations: Quasi-Experimental Evidence from Europe," *Am. Polit. Sci. Rev.*, vol. 118, no. 3, pp. 1541–1548, Aug. 2024, doi: 10.1017/S0003055423001193.
- [9] T. Aichner, M. Grünfelder, O. Maurer, and D. Jegeni, "Twenty-Five Years of Social Media: A Review of Social Media Applications and Definitions from 1994 to 2019," *Cyberpsychology, Behav. Soc. Netw.*, vol. 24, no. 4, pp. 215–222, 2021, doi: 10.1089/cyber.2020.0134.
- [10] A. Bruns, "After the 'APIcalypse': social media platforms and their fight against critical scholarly research," *Disinformation Data Lockdown Soc. Platforms*, pp. 14–36, Dec. 2021, doi: 10.4324/9781003206972-2.
- [11] Y. Gorodnichenko, T. Pham, and O. Talavera, "Social media, sentiment and public opinions: Evidence from #Brexit and #USElection," *Eur. Econ. Rev.*, vol. 136, p. 103772, Jul. 2021, doi: 10.1016/J.EUROECOREV.2021.103772.
- [12] V. Govindan and V. Balakrishnan, "A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection," *J. King Saud Univ. - Comput.*

Inf. Sci., vol. 34, no. 8, pp. 5110–5120, Sep. 2022, doi: 10.1016/J.JKSUCI.2022.01.008.

- [13] A. H. Shapiro, M. Sudhof, and D. J. Wilson, “Measuring news sentiment,” *J. Econom.*, vol. 228, no. 2, pp. 221–243, Jun. 2022, doi: 10.1016/J.JECONOM.2020.07.053.
- [14] M. E. Basiri, S. Nemati, M. Abdar, S. Asadi, and U. R. Acharrya, “A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets,” *Knowledge-Based Syst.*, vol. 228, p. 107242, Sep. 2021, doi: 10.1016/J.KNOSYS.2021.107242.
- [15] N. Jing, Z. Wu, and H. Wang, “A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction,” *Expert Syst. Appl.*, vol. 178, p. 115019, Sep. 2021, doi: 10.1016/J.ESWA.2021.115019.
- [16] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, “A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets,” *IEEE Access*, vol. 11, pp. 11811–11826, 2023, doi: 10.1109/ACCESS.2023.3242290.
- [17] Muhammad Ikram Kaer Sinapoy, Yuliant Sibaroni, and Sri Suryani Prasetyowati, “Comparison of LSTM and IndoBERT Method in Identifying Hoax on Twitter,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 657–662, 2023, doi: 10.29207/resti.v7i3.4830.
- [18] M. A. K. Fata, S. Sumpeno, A. D. Wibawa, and D. A. Feryando, “Evaluating the Sentiment Analysis from Auto-Generated Summary Text Using IndoBERT Fine-Tuning Model in Indonesian News Text,” *Proc. - 2023 15th IEEE Int. Conf. Comput. Intell. Commun. Networks, CICN 2023*, pp. 822–829, 2023, doi: 10.1109/CICN59264.2023.10402345.
- [19] M. Yarchi, C. Baden, and N. Kligler-Vilenchik, “Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media,” *Polit. Commun.*, pp. 1–42, Mar. 2021, doi: 10.1080/10584609.2020.1785067.