Scientific Journal of Informatics
Vol. 12, No. 3, August 2025

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Implementation of IndoBERT for Sentiment Analysis of the Constitutional Court's Decision Regarding the Minimum Age of Vice Presidential Candidates

Very Dwi Setiawan[1*], Dwi Utari Iswavigra[2], Endang Anggiratih[3]
[1, 3]Department of Informatics, Universitas Pignatelli Triputra, Indonesia
[2]Department of Computer Science, Universitas Sugeng Hartono, Indonesia

Abstract.
Purpose: This study aims to analyze the effectiveness of the IndoBERT model for sentiment analysis of Indonesian anguage YouTube comments related to the legal Court's ruling on the minimum age of vice presidential candidates for 2024. While previous research applied conventional machine learning methods, this study addresses the challenge of understanding nuanced public opinion using a language-specific transformer model.
Methods: A dataset of 23,796 YouTube comments was collected using the YouTube Data API in January 2025. The comments underwent extensive preprocessing including normalization, case folding, text cleansing, symbol removal, stopword elimination, and stemming. Sentiment labels (positive, negative, neutral) were assigned through a lexicon based approach. Three models IndoBERT, BERT, Support Vector Machine (SVM), and Random Forest were trained and tested using an 80% and 20% split. Model result was evaluated with accuracy, precision, recall, and F1-score metrics.
Result: IndoBERT achieved the maximum result with 95% accuracy, outperforming BERT 92%, SVM 88%, and Random Forest 85%. This confirms IndoBERT's superior ability to capture contextual nuances in Indonesian sentiment analysis compared to other models.
Novelty: This research demonstrates the advantage of transformer based models, particularly IndoBERT, in analyzing complex Indonesian social media texts. The findings support the use of IndoBERT for automated sentiment monitoring to inform government and media responses. Future work could extend to broader discourse analysis across diverse public sectors.

Keywords: IndoBERT, Sentiment analysis, Transformer model, Constitutional court's, Vice presidential candidates
Received June 2025 / **Revised** July 2025 / **Accepted** July 2025

INTRODUCTION

Social media has now become the main space for people to voice their opinions on political and legal issues [1]. Public opinion that develops on digital platforms can influence the perception and legitimacy of decisions by state institutions. One of the issues that has been widely discussed is the legal Court's decision concerning the minimum age boundary for vice presidential candidates, which has triggered various emotional responses on social media. This situation shows the need for sentiment analysis to scientifically examine public responses, particularly by policy analysts, legal researchers, and government institutions seeking to gauge citizen perception on legal issues [2],[3]. This approach is important to support policy making, digital democracy studies, and understanding the social impact of legal decisions. The stakeholders who need to examine public responses include policy makers, government institutions, political researchers, and civil society organizations [4].

Although sentiment outputs such as positive, negative, and neutral may appear general, they provide essential aggregate signals for decision makers [5]. A predominantly negative public sentiment, for example, can prompt policymakers to reconsider or further explain a controversial decision. Sentiment analysis research in politics and law has been conducted to examine public responses to issues such as general elections [6], the ratification of the Job Creation Law [7], and court decisions [8]. Various machine

[*]Corresponding author.
Email addresses: ferystiawan54@gmail.com (Setiawan)
DOI: 10.15294/sji.v12i3.26360

learning techniques such as SVM, Naïve Bayes, and Long Short-Term Memory (LSTM). For example, a study by Setiawan et al. [9] analyzed sentiment towards the Job Creation Law using Naïve Bayes and obtained an accuracy of 56%, but had difficulty distinguishing nuances of sarcasm and ambiguity in Twitter text. This shows that traditional methods are often less effective in capturing contextual meaning, especially in complex and emotional political discourse.

To overcome these limitations, the Bidirectional Encoder Representations from Transformers (BERT) model was introduced as a transformer-based approach that is able to understand context in two directions [10],[11],[12], making it much more effective in capturing the meaning of complex sentences compared to traditional machine learning techniques such as SVM or LSTM. Research by Smairi et al. [13] showed that the BERT model managed to achieve an accuracy of up to 91% in sentiment analysis of movie reviews, surpassing the performance of the SVM model which only reached 88%, and the Random Forest model 86% on the same dataset. In addition, research by Koto et al. [14] explained that IndoBERT gave better results with an F1-Score of 84%, evaluated to the Naïve Bayes techniques which obtained an F1-Score of 71% and Logistic Regression of 72% on the Indonesian language sentiment analysis dataset. Simanjuntak et al, [15] on fake news detection using IndoBERT showed that hyperparameter tuning with Bayesian optimization provided the best performance in optimizing the IndoBERT-base-p1 model. Specifically, Bayesian optimization achieved a precision of 88.79%, a recall of 94.5%, and an F1-score of 91.56% for the "fake" label. However, the application of IndoBERT to sentiment analysis, especially on legal and political issues in Indonesia, is still limited and has yet to be studied in depth.

Most sentiment analysis research in Indonesia still uses a general approach that is not specifically directed at public policy issues [16],[17], including the Constitutional Court's decision. Until now, no study has been found that specifically analyzes public sentiment towards The Constitutional Court's ruling on the minimum age requirement for candidates for vice president. In addition, the application of IndoBERT as a language model specifically designed for Indonesian has not been optimally utilized in the national socio-political context [18]. There are also still very few empirical studies that integrate social media data with legal analysis through a modern Natural Language Processing (NLP) approach[19], thus opening up new research opportunities that are more relevant and contextual.

This study is an initial effort that combines actual national legal-political issues with a modern NLP approach based on IndoBERT [14],[15],[18] ,[19] . This study offers a new perspective in understanding public responses more accurately and contextually. The use of IndoBERT allows for deeper meaning extraction from Indonesian-language texts, thus making a significant contribution to the development of data-based constitutional policy analysis in the digital era.

METHODS

This part provides an elucidation of the processing technique used in this investigation. Dataset preparation is followed by a number of text pretreatment steps to remove any items that could impede the data processing process, data labelling, information conversion into numerical form using the weighting approach, and finally data mining and evaluation. The phases of this investigation are depicted in Figure 1 [20].

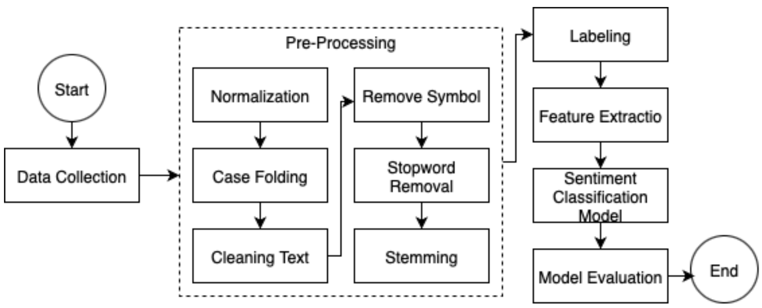

Figure 1. Proposed methods step

Dataset

The information utilised in this research came from YouTube comments collected using the YouTube Data API. To crawl YouTube comment data, the Python program must first be connected to the YouTube API

Key obtained through the Google Developer Console[19]. The YouTube accounts whose comments were taken include: Dirty Vote, TV One, Kompas TV, Metro TV, Kumparan, and Koran Sindo, all of which have content related to the discussion of the Constitutional Court's ruling on the minimum age requirement for vice presidential candidates in 2024. The data crawling process was carried out in January 2025. To ensure relevance to the decision of the Constitutional Court, the dataset was filtered using specific keywords such as “MK,” “Mahkamah Konstitusi,” “batas usia,” “cawapres,” “calon wakil presiden,” “Putusan MK,” “Anwar Usman,” “putusan,” “UU Pemilu,” and “konstitusi.” These keywords were chosen because they are directly associated with discussions surrounding the Court's ruling about the minimum age needed to run for vice president. The inclusion of both general and specific terms ensured that the comments collected were contextually linked to the legal and political discourse targeted in this study.The total data obtained was 23,796 comments, and the data was stored in CSV format. Figure 2 visualizes the stages of crawling YouTube data.

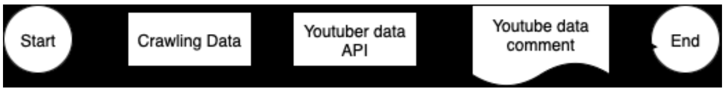

Figure 2. YouTube data crawling stages

Based on the lexicon-based labeling process, the dataset consists of 8,287 positive, 6,951 negative, and 8,558 neutral comments. Below are illustrative examples for each sentiment category Positive: "Putusan MK ini menunjukkan keberpihakan pada regenerasi pemimpin muda." Negative: "MK sekarang sudah tidak independen lagi, hanya alat politik." Neutral: "Putusan MK tentang usia cawapres sudah diumumkan kemarin.".

Pre-processing

This study applies several stages of data pre-processing, namely normalization, case folding, text cleaning, symbol removal, stopword removal, and stemming. Changing non-standard words into standard ones is the goal of normalization so that they are consistently recognized by the model [21]. Case folding converts capital letters to lowercase letters to avoid differences in analysis due to capitalization. Text cleaning is done by removing duplicate data, correcting errors, and completing missing values. The symbol removal stage aims to remove symbols, hashtags, and numbers to standardize the data format [22]. To exclude common terms that are not crucial to the study, stopword removal is utilised. Finally, stemming is done to change affixed words to their basic form. Table 1 below explains the pre-processing process used in this study.

Table 1. Pre-processing steps

Steps	Text Result
Original Text	Gw benci bgt sm org2 yg suka nyebar hoax! Plis deh, jgn bodohin rakyat!!
Normalization	saya benci banget sama orang-orang yang suka menyebar hoax! tolong deh, jangan membodohi rakyat!!
Case Folding	saya benci banget sama orang-orang yang suka menyebar hoax! tolong deh, jangan membodohi rakyat!!
Text Cleansing	saya benci banget sama orang orang yang suka menyebar hoax tolong deh jangan membodohi rakyat
Remove Symbol	saya benci banget sama orang orang yang suka menyebar hoax tolong deh jangan membodohi rakyat
Stop Word Removal	benci banget orang orang suka menyebar hoax tolong jangan membodohi rakyat
Stemming	benci banget orang orang suka sebar hoax tolong jangan bodoh rakyat

Labeling

The lexicon-based method was used to create the sentiment labels (positive, negative, and neutral) [23],[24]. In this method, each word in the cleaned text is compared with a predefined lexicon containing sentiment scores. Words indicating support, justice, or optimism are classified as positive; words showing criticism, disappointment, or dissatisfaction are marked as negative; and neutral words without emotional context are categorized accordingly. To ensure reliability, a subset of 1,500 randomly selected comments was manually annotated by two independent annotators as ground truth. This manually labeled data was then used to evaluate the model's classification performance through accuracy, precision, recall, and F1-score metrics.
The lexicon used in this study was adapted from the translated and extended version of the NRC Emotion Lexicon and manually refined to fit the Indonesian socio political context. Words were categorized into

three sentiment classes positive, negative, and neutral, based on their emotional polarity and contextual usage. Before sentiment scoring, all texts were preprocessed using standard steps normalization, stopword removal, and stemming, to align token forms with entries in the lexicon [25]. Each token in a comment was assigned a sentiment score of +1 (positive), –1 (negative), or 0 (neutral). The sum of all token scores was used to determine a comment's overall sentiment score. A final sentiment label was assigned to the comment as follows: if the total score > 0, the comment was labeled positive; if < 0, labeled negative; and if the total score = 0 or had no recognized sentiment tokens, it was labeled neutral.

Modeling

At this stage, researchers use the data that has been divided to build a model, which is then fed into IndoBERT [26],[27],[28]. IndoBERT is one of the leading models in understanding Indonesian text, as shown in Figure 3. Although the sentiment labels were generated through a lexicon-based approach, this labeling serves only as ground truth for model evaluation. The primary goal of this research is to train and evaluate the IndoBERT model in predicting sentiment labels automatically. IndoBERT is not used for labeling, but as a predictive model whose outputs are compared against the lexicon-based annotations using accuracy, precision, recall, and F1-Score.This model is developed based on the BERT transformer architecture[29],[30], which is often used in English. As seen in the diagram, the IndoBERT structure consists of twelve hidden layers, each measuring 786 dimensions and equipped with twelve attention heads. In this study, a model of the type "indobenchmark/indobert-base-p1" was used which has around 124.5 million parameters. The researchers trained the model for three epochs using the AdamW optimizer with a learning rate of 2e-5. In addition, several variations of batch sizes were used, namely 16, 32, and 128, and the input length was limited to 128 tokens.

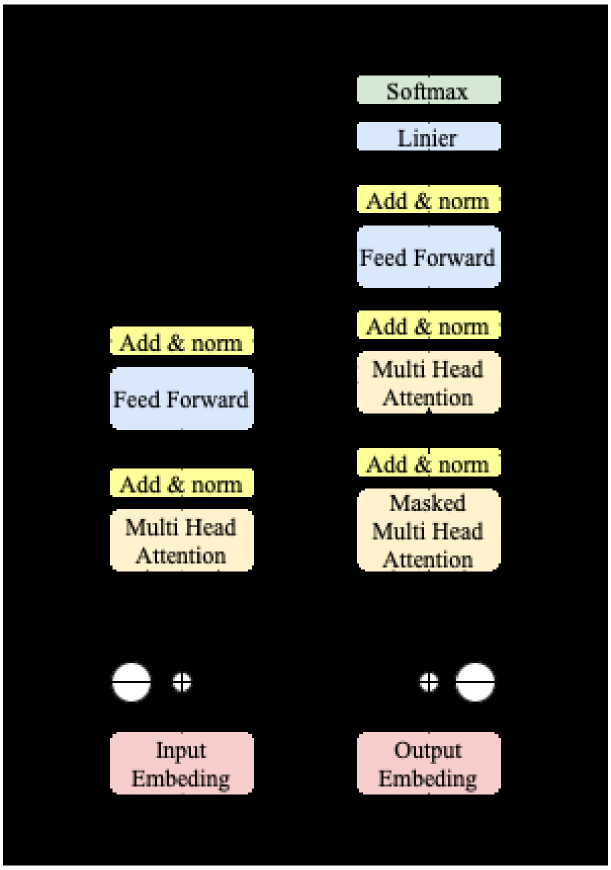

Figure 3. IndoBERT Architecture [26]

Evaluation Model

Following the modelling phase, researchers utilise a matrix of confusion to evaluate the model and a classification report to determine the test data's accuracy, precision, recall, and F1-Score values. A tool for assessing model correctness is the matrix of confusion, which produces values for precision, recall, and f1-score [31], [32]. Confusion matrix is shown in Figure 4.

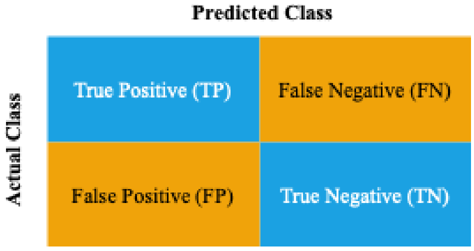

Figure 4. Confusion matrix

The True Positive (TP) is the number of data points with positive score that are anticipatorily positive. The True Negative (TN) is the proportion of data with negative score that are estimated to be negative. FP stands for False Positives, which are the proportion of data with negative score that were estimated to be positive. FN stands for False Negatives, which are the proportion of data with positive score that were estimated to be negative. The accuracy, precision, recall, and F1-Score of the model can be computed using the following equation based on Figure 4.

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

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

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

$$F1\,Score = 2 \times \frac{(Precision \times Recall)}{(Precision \times Recall)} \quad (4)$$

Accuracy is defined as the percentage of all data that is correctly classified, demonstrating how closely the model's predictions match the labels [33]. When positive prediction mistakes need to be reduced, precision which measures the proportion of positive predictions that are truly accurate is crucial. The ability of the model to capture all really positive data is demonstrated by recall, also known as sensitivity, which is highly helpful when we wish to prevent missing positive data. The F1 score, on the other hand, is the harmonic mean of precision and recall and offers a thorough picture of how well the two are balanced, particularly when the data being used is unbalanced [34].

RESULTS AND DISCUSSIONS

The objective of this study is to analyze the results of the IndoBERT model in conducting sentiment analysis on public opinion regarding The Constitutional Court's decision regarding the minimum age for candidates running for vice president. Two conventional machine learning algorithms, namely SVM and Random Forest, are applied to the same dataset for comparative evaluation. In addition, a general BERTModel pretrained on English corpora is also included as a benchmark to observe how language-specific models like IndoBERT compare to multilingual or non-Indonesian pretraining approaches. This study investigates how well the three models identify sentiment into three classes positive, negative, and neutral, using a dataset of YouTube comments written in Indonesian. 23,796 tweets that were gathered from the YouTube site within a week of the MK ruling's announcement make up the dataset that was used. Case folding, tokenisation, link and symbol removal, and stopword removal are all steps in the preprocessing process. Following this procedure, 20% of the data is used for testing and 80% is used for training. According to the data distribution, 36.02% of comments are neutral, 29.19% are negative, and 34.78% of comments are favourable. The sentiment label distribution is displayed in Figure 5.

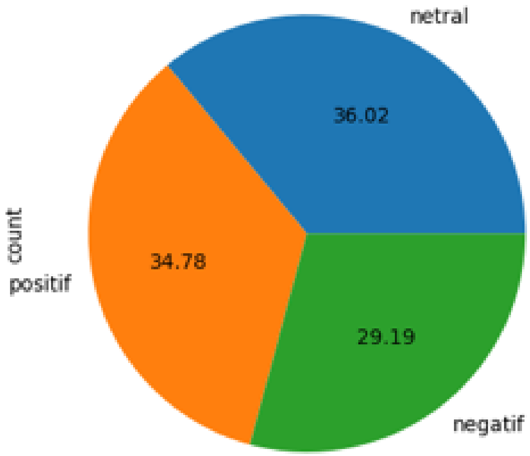

Figure 5. Distribution labeling sentiment

Figure 5 shows a balanced sentiment distribution, so there is no need for a process to add data sensibly. Evaluation such as accuracy, precision, recall, and F1-score were used to score the model. According to the findings, the IndoBERT model yields the best outcomes with an accuracy of 95%, precision of 95%, recall of 93%, and F1-score of 95%. Meanwhile, the BERT model recorded an accuracy of 92%, the SVM model recorded an accuracy of 88% and Random Forest only reached 85%. The performance of the metric evaluation comparison can be seen in Figure 6.

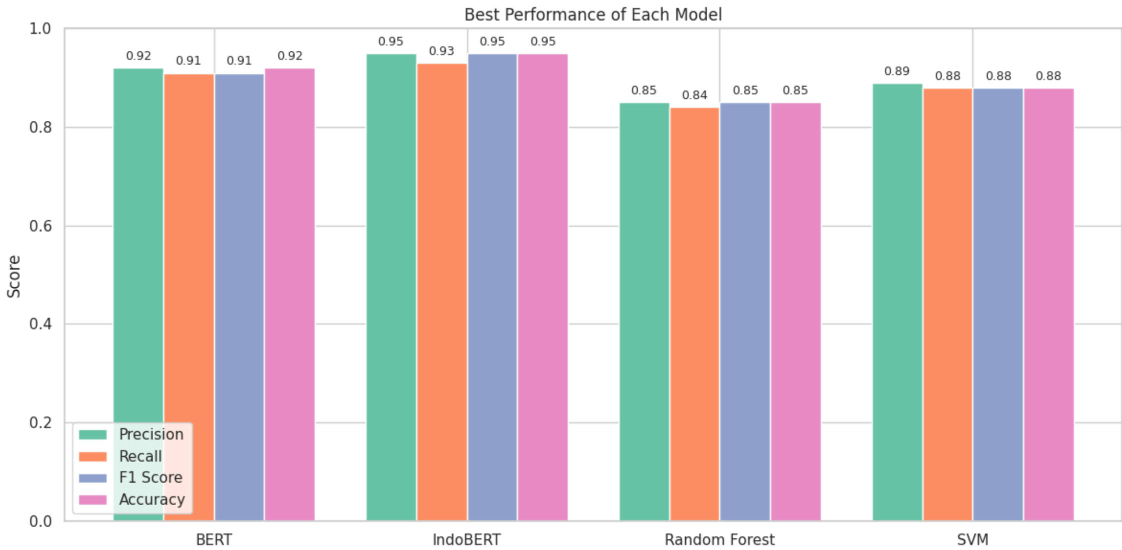

Figure 6. Best performance of each model

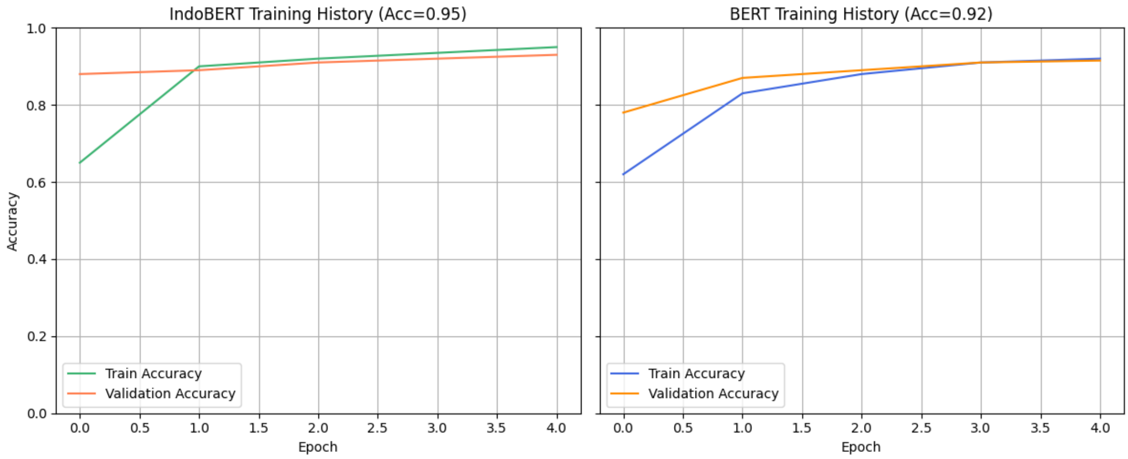

Figure 7. Training accuracy per epoch for IndoBERT and BERT Models

Figure 6 shows that IndoBERT has the best performance because it is specifically trained on the Indonesian language corpus, making it more accurate in understanding context and emotion in text, with all evaluation metrics above 0.93%. BERT also shows high performance thanks to its robust architecture, although slightly below IndoBERT due to limitations in language specifications. The figure 7 illustrates the learning dynamics of both models over 5 epochs, showing consistently increasing training and validation accuracy without signs of overfitting. Both models were fine-tuned using the AdamW optimizer, a learning rate of 2e-5, and batch sizes of 16, 32, and 128, with a maximum input length of 128 tokens and an 80:20 train-test split. IndoBERT required 42 minutes and English BERT 38 minutes of training on a single T4 GPU in gooogle colab. SVM achieves intermediate results due to its ability to classify based on numeric features, but is less than optimal in understanding language context. Meanwhile, Random Forest shows the lowest performance due to its limitations in capturing the order and contextual meaning in natural language sentences. To see which words are dominant from sentiment analysis that affect the evaluation matrix, see Figure 8.

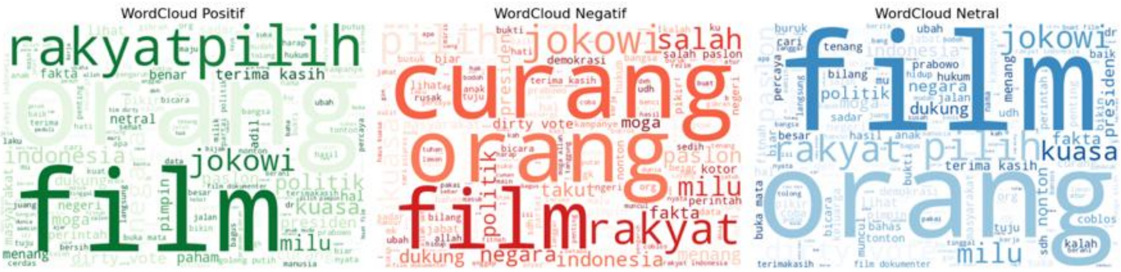

Figure 8. WordCloud

The WordCloud visualization on Gigure 7 shows the most frequently occurring words in each sentiment category, namely positive, negative, and neutral. This WordCloud provides a contextual overview of the review data, where it can be seen that the words "film", "orang", and "rakyat" dominate the three categories. However, words such as "pilih", "menang", and "terima kasih" are more prominent in positive sentiment; while the words "curang", "salah", and "takut" appear more frequently in negative sentiment, reflecting a sharper expression of emotion. Meanwhile, neutral sentiment is dominated by descriptive and informative words such as "negara", "politik", and "kekuasaan", indicating a tendency towards an impartial narrative.

IndoBERT outperformed English BERT and traditional models such as SVM and Random Forest in both training and validation accuracy, reaching 95% and 93% respectively. Its superior performance is attributed to its pretraining on Indonesian text, which enhances its contextual understanding. While IndoBERT demands higher computational resources, it offers significantly better results than classical models, which are more lightweight but suffer from lower accuracy. These findings suggest that transformer-based models like IndoBERT are well-suited for analyzing complex Indonesian sentiments. Moreover, this study highlights the potential of NLP to automatically monitor public opinion, providing valuable insights for policymakers and researchers.

CONCLUSION

This research demonstrates that the IndoBERT model is the best choice for sentiment analysis of Indonesian-language public opinion, especially in the situation of the Constitutional Ruling ruling on the age limit for vice presidential candidates. Compared to BERT, SVM, and Random Forest, IndoBERT is significantly superior with accuracy and other evaluation metrics exceeding 90%. WordCloud visualization also strengthens the evaluation results by displaying the dominance of typical words from each sentiment category. Although it requires greater computing resources, IndoBERT's superiority in recognising the subtleties and contexts of the Indonesian language makes it very relevant for application in automatic public opinion monitoring. This finding opens up great opportunities in the use of NLP to support decision-making in the government, media, and social research sectors based on digital data.

REFERENCES

[1] G. Gosztonyi, J. Bálint, and G. F. Lendvai, “Public Figures and Social Media from a Freedom of Expression Viewpoint in the Recent U.S. and EU Jurisdiction,” *Journal. Media*, vol. 6, no. 1, 2025, doi: 10.3390/journalmedia6010026.

[2] O. Somantri, R. H. Maharrani, and S. Purwaningrum, “Coastal Sentiment Review Using Naïve Bayes with Feature Selection Genetic Algorithm,” *Sci. J. Informatics*, vol. 10, no. 3, pp. 229–238, 2023, doi: 10.15294/sji.v10i3.43988.

[3] V. S. G. S. Phaneendra Bottu and K. Ragavan, “Emotion-Based Music Recommendation System Integrating Facial Expression Recognition and Lyrics Sentiment Analysis,” *IEEE Access*, vol. 13, pp. 87740–87752, 2025, doi: 10.1109/ACCESS.2025.3570011.

[4] G. Asimakopoulos, H. Antonopoulou, K. Giotopoulos, and C. Halkiopoulos, “Impact of Information and Communication Technologies on Democratic Processes and Citizen Participation,” *Societies*, vol. 15, no. 2, 2025, doi: 10.3390/soc15020040.

[5] Y. Sun, H. Yuan, and F. Xu, “Financial sentiment analysis for pre-trained language models incorporating dictionary knowledge and neutral features,” *Nat. Lang. Process. J.*, vol. 11, p. 100148, 2025, doi: https://doi.org/10.1016/j.nlp.2025.100148.

[6] F. A. Wenando, R. Hayami, Bakaruddin, and A. Y. Novermahakim, “Tweet Sentiment Analysis for 2019 Indonesia Presidential Election Results using Various Classification Algorithms,” in *2020 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering (ICITAMEE)*, 2020, pp. 279–282. doi: 10.1109/ICITAMEE50454.2020.9398513.

[7] F. Guo, Z. Liu, Q. Lu, S. Ji, and C. Zhang, “Public Opinion About COVID-19 on a Microblog Platform in China: Topic Modeling and Multidimensional Sentiment Analysis of Social Media,” *J. Med. Internet Res.*, vol. 26, 2024, doi: https://doi.org/10.2196/47508.

[8] G. A. Sandag, E. H. E. Soegiarto, L. Laoh, A. Gunawan, and D. Sondakh, “Sentiment Analysis of Government Policy Regarding PPKM on Twitter Using LSTM,” in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, 2022, pp. 1–6. doi: 10.1109/ICORIS56080.2022.10031474.

[9] A. S. Setiawan, H. H. Nuha, and M. W. A. Bawono, “Classification of Sentiment Analysis Against Omnibus Law on Twitter Social Media and News Websites Using the Naïve Bayes Method,” in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2022, pp. 141–147. doi: 10.1109/ISRITI56927.2022.10052796.

[10] T. Mudiyanselage Anupama Udayangani Gunathilaka, J. Zhang, and Y. Li, “Fine-Grained Feature Extraction in Key Sentence Selection for Explainable Sentiment Classification Using BERT and CNN,” *IEEE Access*, vol. 13, pp. 68462–68480, 2025, doi: 10.1109/ACCESS.2025.3560482.

[11] K. I. Roumeliotis, N. D. Tselikas, and D. K. Nasiopoulos, “Optimizing Airline Review Sentiment Analysis: A Comparative Analysis of LLaMA and BERT Models through Fine-Tuning and Few-Shot Learning,” *Comput. Mater. Contin.*, vol. 82, no. 2, pp. 2769–2792, 2025, doi: https://doi.org/10.32604/cmc.2025.059567.

[12] S. S. Malik *et al.*, “Multi-Modal Emotion Detection and Sentiment Analysis,” *IEEE Access*, vol. 13, pp. 59790–59810, 2025, doi: 10.1109/ACCESS.2025.3552475.

[13] N. Smairi, H. Abadlia, H. Brahim, and W. L. Chaari, “Fine-tune BERT based on Machine Learning Models For Sentiment Analysis,” *Procedia Comput. Sci.*, vol. 246, no. C, pp. 2390–2399, 2024, doi: 10.1016/j.procs.2024.09.531.

[14] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.
[15] Anugerah Simanjuntak *et al.*, “Research and Analysis of IndoBERT Hyperparameter Tuning in Fake News Detection,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 13, no. 1, pp. 60–67, 2024, doi: 10.22146/jnteti.v13i1.8532.
[16] B. Waspodo, Q. Aini, F. R. Singgih, R. H. Kusumaningtyas, and E. Fetrina, “Support Vector Machine and Lexicon based Sentiment Analysis on Kartu Prakerja (Indonesia Pre-Employment Cards Government Initiatives),” in *2022 10th International Conference on Cyber and IT Service Management (CITSM)*, 2022, pp. 1–6. doi: 10.1109/CITSM56380.2022.9935990.
[17] R. H. Astuti, M. Muljono, and S. Sutriawan, “Indonesian News Text Summarization Using MBART Algorithm,” *Sci. J. Informatics*, vol. 11, no. 1, pp. 155–164, 2024, doi: 10.15294/sji.v11i1.49224.
[18] F. Koto, J. H. Lau, and T. Baldwin, “INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization,” in *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, M.-F. Moens, X. Huang, L. Specia, and S. W. Yih, Eds., Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 10660–10668. doi: 10.18653/v1/2021.emnlp-main.833.
[19] G. Z. Nabiilah, S. Y. Prasetyo, Z. N. Izdihar, and A. S. Girsang, “BERT base model for toxic comment analysis on Indonesian social media,” *Procedia Comput. Sci.*, vol. 216, pp. 714–721, 2023, doi: https://doi.org/10.1016/j.procs.2022.12.188.
[20] R. Pramana, M. Jonathan, H. S. Yani, and R. Sutoyo, “A Comparison of BiLSTM, BERT, and Ensemble Method for Emotion Recognition on Indonesian Product Reviews,” *Procedia Comput. Sci.*, vol. 245, pp. 399–408, 2024, doi: https://doi.org/10.1016/j.procs.2024.10.266.
[21] M. A. Palomino and F. Aider, “Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis,” *Appl. Sci.*, vol. 12, no. 17, p. 8765, Aug. 2022, doi: 10.3390/app12178765.
[22] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, “Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021, doi: 10.57152/malcom.v1i1.20.
[23] H. Shim, D. Lowet, S. Luca, and B. Vanrumste, “LETS: A Label-Efficient Training Scheme for Aspect-Based Sentiment Analysis by Using a Pre-Trained Language Model,” *IEEE Access*, vol. 9, pp. 115563–115578, 2021, doi: 10.1109/ACCESS.2021.3101867.
[24] O. Alsemaree, A. S. Alam, S. S. Gill, and S. Uhlig, “An analysis of customer perception using lexicon-based sentiment analysis of Arabic Texts framework,” *Heliyon*, vol. 10, no. 11, p. e30320, 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e30320.
[25] R. Firdaus, I. Asror, and A. Herdiani, “Lexicon-Based Sentiment Analysis of Indonesian Language Student Feedback Evaluation,” *Indones. J. Comput.*, vol. 6, no. 1, pp. 1–12, 2021, doi: 10.34818/indojc.2021.6.1.408.
[26] B. Wilie *et al.*, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” 2020.
[27] K. Purwandari, R. B. Perdana, J. W. C. Sigalingging, R. Rahutomo, and B. Pardamean, “Automatic Smart Crawling on Twitter for Weather Information in Indonesia,” *Procedia Comput. Sci.*, vol. 227, pp. 795–804, 2023, doi: 10.1016/j.procs.2023.10.585.
[28] 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.
[29] G. Takawane, A. Phaltankar, V. Patwardhan, A. Patil, R. Joshi, and M. S. Takalikar, “Language augmentation approach for code-mixed text classification,” *Nat. Lang. Process. J.*, vol. 5, no. November, p. 100042, 2023, doi: 10.1016/j.nlp.2023.100042.
[30] M. Ramaiah, V. Chandrasekaran, V. Ravi, and N. Kumar, “An intrusion detection system using optimized deep neural network architecture,” *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 4, 2021, doi: 10.1002/ett.4221.
[31] S. Lonang, A. Yudhana, and M. K. Biddinika, “Performance Analysis for Classification of Malnourished Toddlers Using K-Nearest Neighbor,” *Sci. J. Informatics*, vol. 10, no. 3, pp. 313–322, 2023, doi: 10.15294/sji.v10i3.45196.
[32] J. A. Aquino, D. J. Liew, and Y.-C. Chang, “Graph-aware pre-trained language model for political sentiment analysis in Filipino social media,” *Eng. Appl. Artif. Intell.*, vol. 146, p. 110317, 2025,

doi: https://doi.org/10.1016/j.engappai.2025.110317.
[33] J. Erbani, P.-É. Portier, E. Egyed-Zsigmond, and D. Nurbakova, "Confusion Matrices: A Unified Theory," *IEEE Access*, vol. 12, pp. 181372–181419, 2024, doi: 10.1109/ACCESS.2024.3507199.
[34] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies*, vol. 12, no. 7, 2024, doi: 10.3390/technologies12070113.