EDUMATIC Jurnal Pendidikan Informatika

e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Vol. 9 No. 1, April, 2025, Hal. 89-98
DOI: 10.29408/edumatic.v9i1.29267

Understanding Public Sentiments on the 2024 Presidential Election through BERT-Powered Analysis

Abiyyu Daffa Haidar Setiawan [1,*] , Warih Maharani [1]

[1] Department of Informatics, Telkom University, Indonesia.
[*] Correspondence: abiyyudaffa@student.telkomuniversity.ac.id

Received: 9 January 2025 | Revised: 14 January 2025 | Accepted: 1 February 2025 | Published: 11 April 2025

Abstract
Social media platforms serve as dynamic communication across boundaries, with X serving as a platform for opinion exchange. This research examines public sentiment on the 2024 Indonesian Presidential Election to understand voter sentiments based on what happened during the pre-election. Using the Twitter API, 2,146 tweets were collected based on election-related keywords and hashtags, focusing on Indonesian-language tweets with direct opinions. The method that we use is data crawling using the Twitter API. Preprocessing steps included case folding (converting text to lowercase), cleansing (removing noise like URLs and emojis), tokenization, stemming (reducing words to base forms), and stop word removal (e.g., "yang," "dan"). Slang was standardized with a custom dictionary to ensure consistency and accurate interpretation. Leveraging BERT for sentiment analysis, the model achieved 99% accuracy; results indicate that 93.1% of analyzed tweets expressed negative sentiment, highlighting public dissatisfaction about the 2024 presidential election. Hyperparameters are also tested to optimize model performances. With the best result accuracy in 99% using an 80:20 split ratio, with a batch size of 16 and a learning rate of 0.00001. This research underlines the importance of sentiment analysis in elections, demonstrating BERT's capability to handle linguistic complexities and providing a methodological framework for analyzing social media data in political contexts.

Keywords: bert; sentiment analysis; social media; x

INTRODUCTION

Social media has become a platform where people can freely express their opinions on various topic, including politic. In democratic systems, it is expected that citizens stay updated on current events, participate in political discussions, and take an active role in the processes that shape governance and policy (Alaparthi & Mishra, 2021). In Indonesia, the utilization of social media has experienced a high growth. In early 2024, approximately 139 million individuals in the nation were active social media users. Data Reportal shows that Indonesia has 139 million social media users in January 2024 representing 49.9% of the total population of Indonesia. There is also a presentation of the number of active users of X or in Indonesia as much as 24.69 million and X's advertising planning tool recorded an increase of 693 thousand in the span of early 2023 to early 2024. Sentiments refer to an individual's feelings, emotions, opinions, or an event (Wankhade et al., 2022). Comments across social media platforms can be converted into data for sentiment analysis. Sentiment analysis can play a critical role for strategic planning, decision-making, and content filtering (Zharifa & Ujianto, 2024). By using advanced data analysis techniques, including Natural Language Processing (NLP) (Sabilillah, et al., 2024; Taherdoost & Madanchian, 2023), we can analyze the opinions, emotions, and reactions of individuals as expressed in online environments.

Despite significant progress in sentiment analysis, challenges remain. Particularly in understanding complex linguistic phenomena such as sarcasm, sentiment shifts and the use of slang. In Indonesian context, regional dialects, mixed languages used per tweets, and cultural expressions add complexity. For instance, like "gua", "lo" and many other examples to complicate interpretation. Lexicon-based methods, such as SentiWordNet (Tran et al., 2021), TextBlob (Aljedaani et al., 2022), and VADER (Lee et al., 2022), have proven effective for basic sentiment analysis tasks, However, these methods struggle with nuanced language and require substantial manual labelling efforts, limiting their scalability and accuracy in handling diverse datasets. This highlights the need for advanced machine learning approaches.

This study utilizes IndoBERT, a transformer-based model specifically designed for Indonesian text, to address sentiment analysis challenges (Pamungkas, et al., 2024). IndoBERT was developed by Koto et al. (2020). Unlike the original BERT (Kenton & Toutanova et al., 2019), IndoBERT was trained on Indonesian linguistic datasets, making it far more effective to capture the nuances of the language. Built using Hugging Face framework (Ait et al., 2024) this model is a powerful tool for analyzing sentiments in Indonesian text. To enhance its performance, this study also incorporates advanced text preprocessing techniques, such as stemming, and normalization, using the Sastrawi library (Rosid et al., 2020). These steps ensure that text is cleaned and standardized for better analysis. By combining IndoBERT's language-specific training with careful preprocessing, this research aims to deliver a more accurate and reliable approach to sentiment analysis for Indonesian text, addressing common challenges that arise in working with this language. This study improves the understanding of slang, dialects, and mixed-language usage, this study provides practical benefits for businesses, policymakers, and researchers who rely on accurate sentiment analysis for strategic decision-making. Furthermore, the development of a voting-based labelling system enhances classification accuracy, reducing biases, inherent in individual methods. These contributions are not only relevant to NLP research but also demonstrate the practical application of IndoBERT in understanding public sentiment (Catelli et al., 2023).

The adoption of transformer-based models such as BERT and IndoBERT in NLP (Jim et al., 2024) tasks is supported by their ability to generalize across diverse datasets while maintaining high performance. Studies by Mozafari et al. (2020) and Talaat (2023) validate BERT's robustness in sentiment analysis, demonstrating its ability to outperform traditional and deep learning models like LSTM (Mahadevaswamy & Swathi, 2022) and CNN (Khan et al., 2022). However, its effectiveness heavily depends on language-specific customization and preprocessing, as highlighted in (Salnikova, 2021). IndoBERT, specifically fine-tuned for Indonesian text, builds upon these theories by addressing the limitations of generic BERT models in non-English contexts.

While previous study has focused on sentiment analysis using English datasets or general-purpose models like BERT, these approaches often fail to account for the unique linguistic features of languages like Indonesian. For example, research by Salnikova (2021) explored the customization of BERT models for Russian and Ukrainian, while Geni et al. (2023) compared different labelling algorithms. However, these studies lack a focus on voting-based labelling systems or the complexities of mixed-language and dialectal data. IndoBERT overcomes these gaps by capturing the nuances of Indonesian text and providing a scalable solution for sentiment analysis. This study further addresses the need for improved classification accuracy through a voting-based labelling system that reduces biases inherent in individual methods.

This research is crucial for advancing sentiment analysis methodologies in linguistically diverse contexts like Bahasa Indonesia. Indonesia's unique linguistic landscape, with slang, regional dialects, and mixed-language expressions, poses challenges for traditional sentiment analysis methods. Accurate analysis in this context is essential for businesses to improve customer engagement, tailor marketing strategies, and enhance brand perception. Policymakers can use sentiment analysis to monitor public opinion and address societal concerns effectively.

Researchers benefit from improved tools and datasets, enabling deeper exploration of social trends and public sentiment. By leveraging IndoBERT, this study provides a robust solution for handling these complexities, ensuring more accurate sentiment analysis. Ultimately, this research bridges linguistic diversity and technical innovation, offering practical benefits for decision-making in both public and private sectors.

This research aims to analyze public sentiment on the 2024 Presidential Election using BERT-based models, examining trends, opinions, and discourse surrounding the event. It seeks to enhance sentiment classification accuracy by addressing linguistic challenges such as slang, dialect variations, and mixed-language usage commonly found in Indonesian social media. Additionally, the study focuses on improving sentiment detection through advanced preprocessing techniques and a voting-based labelling system to ensure more reliable classification. By leveraging BERT's contextual understanding, this research contributes to NLP methodologies by developing scalable tools for sentiment analysis, supporting applications in social media monitoring and data-driven decision-making. Ultimately, the findings provide valuable insights for policymakers, researchers, and the public, reinforcing the role of transformer-based models in analyzing political sentiment in Bahasa Indonesia.

METHOD

This research utilizes the Bidirectional Encoder Representations from Transformers (BERT) method to classify tweet data into two distinct sentiment categories: positive and negative. The process involves a series of structured and detailed steps, as illustrated in Figure 1. The first stage of the process is data crawling, we crawled 2,146 data during presidential candidate debate from 10 December 2023 until 12 February 2024, we extract data from X using keywords "Anies Baswedan", "Prabowo Subianto", "Ganjar Pranowo", and "pilpres2024". We only extract data from tweets not retweets. So, we can make sure of its sentiments per tweet. This step ensures the dataset is rich in context and contains diverse opinions and sentiments from social media users. Each extracted tweet inherently carries sentiment, either explicitly or implicitly, which will later be classified using a sentiment labelling method. The labelling process ensures that the sentiment of each tweet is accurately identified and categorized to enable further analysis. This approach provides a robust foundation for sentiment classification, offering insights into public opinion and the prevailing mood regarding the presidential candidates during the election period.

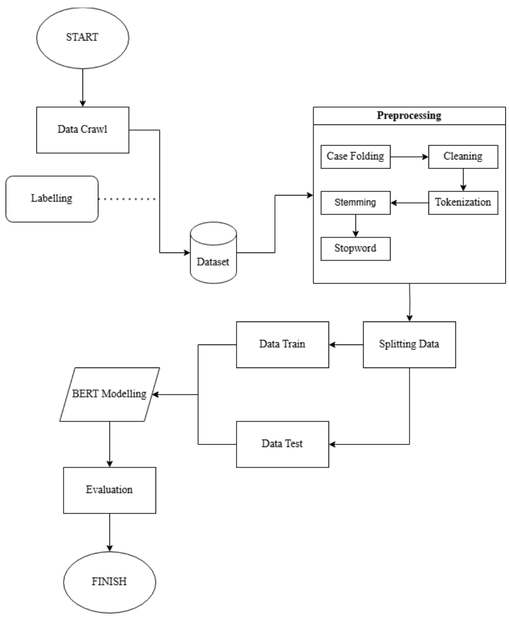

Figure 1. Flowchart

Labelling is conducted using a robust voting mechanism that combines the strengths of three sentiment analysis algorithms: VADER, SentiWordNet, and TextBlob. Each algorithm independently analyzes the sentiment of the crawled data, assigning it either a "positive" or "negative" label based on its unique methodology. VADER excels in handling short, informal texts like social media posts using a lexicon-based approach, while SentiWordNet use lexical resources to assign positive and negative scores to words within WordNet synsets. TextBlob, employs rule-based parsing and Naive Bayes classification to determine sentiment polarity. By combining these methods, the labelling process achieves greater consistency and reliability, ensuring a high-quality dataset for further analysis.

After that preprocess stage occurred. This stage involves the data will go through various process such as case folding is applied to convert all letters into lowercase, ensuring uniformity across the dataset. Then, data cleaning removes unnecessary symbols and punctuation marks, which could introduce noise. The text is further processed through tokenization, breaking it into individual words or tokens and stemming, which reduces words to their based form (e.g., *"menamatkan"* into *"tamat"*) to enhance generalization. Additionally, stopword removal eliminates common conjunctions and filler words, such as *"cuma", "yang"* and many other conjunctions.

The next step is data splitting and modelling, where the dataset is divided for training and testing with various ratios such as 90:10, 80:20, and 70:30. BERT is chosen for sentiment analysis due to its ability to efficiently handle complex linguistic patterns. This research uses IndoBERT, a variant of BERT tailored for Bahasa Indonesia, as it is more effective in addressing the language's unique linguistic challenges. And finally, the evaluation is performed using a ROC-AUC metric to assess the model's ability to distinguish between positive and negative sentiment. The AUC score helps determine the model's effectiveness in classification tasks, ensuring its reliability and suitability for sentiment analysis in this research.

RESULT AND DISCUSSION

Result

The dataset obtained from social media crawling comprises a total of 2,146 tweets, which were used for analysis in this research. Among these, the distribution of sentiments reveals a significant imbalance. Out of the total tweets, only 148 are classified as having positive sentiments, while the vast majority, 1,998 tweets, are categorized as having negative sentiments. This difference in sentiment distribution highlights the negative public perception regarding the 2024 Presidential Election on social media. The imbalance in the dataset posed challenges in ensuring that the model could accurately classify sentiments without being biased toward the majority class.

The potential impact of class imbalance on the model's performance is significant, as it can lead to biased predictions and reduced generalization ability. Models trained on imbalanced datasets tend to favor the majority class, which in this case is negative sentiment, while struggling to correctly classify the minority class, resulting in lower recall and F1-score for positive sentiment. This imbalance can cause misclassification, ultimately affecting the model's reliability in real-world applications where accurate sentiment detection is crucial. To mitigate these issues, various resampling techniques, including oversampling and undersampling, were employed alongside data augmentation to improve the model's learning capacity and reduce bias. Oversampling artificially increases the representation of the minority class, while undersampling reduces the dominance of the majority class, ensuring a more balanced training process. Several experiments were conducted to address these challenges and improve the model's ability to generalize across both sentiment classes. Below is an example of our data distribution in table 1.

Table 1. Before and after preprocessing

Before	After
Ayo Indonesia semangat!!!! Kita masih punya Peluang satu Langit dan bumi untuk bisa MENANG PILPRES 2024, Jangan mau terkecoh dengan Qouik count ayam sayur itu! Kita akan mempunyai Presiden berprestasi, mencintai Rakyat, dan tentunya dicintai Seluruh penduduk langit dan Bumi 👏👏	*ayo indonesia semangat kita masih punya peluang satu langit dan bumi untuk bisa menang pilpres 2024 jangan terkecoh dengan qouik count ayam sayur itu kita akan punya presiden prestasi cinta rakyat dan tentu cinta seluruh penduduk langit dan bumi*

The second step focuses on the preprocessing of data that has already been collected. The process involves several steps, such as case folding to convert text to lowercase. This step is to standardize the text, so words like *"Sebelum"* and *"sebelum"* are treated as the same. Next step is removing emojis and eliminating symbols. This step is to remove digits and non-alphanumerical characters so it can eliminate noise that does not contribute to sentiment. Next step is tokenizing the text. We use IndoBERT tokenizer to break down text into individual token. After that we use normalization for converting synonyms to a standard form. Next is a stop word. Stop word purposes is to removes common words like *"dan", "si"* which carry little values to sentiment, and the last step is stemming. Stemming is to return words to their base form for example *"berlari"* to *"lari"* These steps are crucial for standardizing the text data and preparing it for further analysis. The preprocessing workflow can be customized based on the specific characteristics and requirements of the dataset. Below are the illustrations of data preprocessing

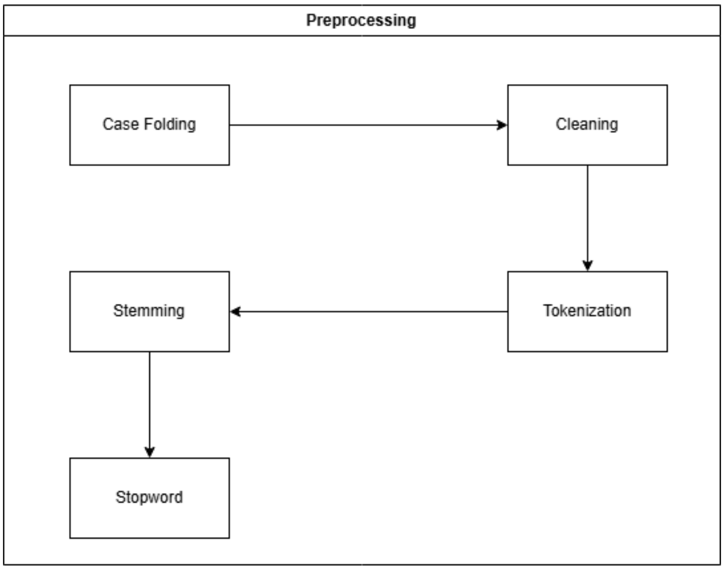

Figure 2. Preprocessing workflow

The third step is fine-tuning. Fine-tuning is a key technique in deep learning technique used to customize pre-trained models for a specific task. Fine-tuning works by training BERT model on a smaller and task-specific dataset using a low learning rate. This helps the model to keep the general knowledge it gained during pre-training while learning new features for the specific task. Another strong feature is it minimizes the chance of overfitting. In this research, fine-tune focuses on learning specific pattern with encoder and decoder BERT model. This will ensure optimal performance for sentiment analysis task in this research

The fourth step is modelling with BERT. BERT process starts from the steps above will be divided into training data and test data with varying split ratios. Training data will use BERT model to be classified based on parameters that have been adjusted to batch size of 16 and learning rate of 0.00001. In this research, encoder and decoder concepts are encapsulated within BERT pre-trained model for sentiment analysis. This eliminates the need for a traditional decoder, as the self-attention mechanism in the encoder effectively captures contextual relationships within the text.

Based on figure 3 below, our dataset is heavily tilted on negative sentiments. This can potentially influence the performance on machine learning that will be conducted. Below are our experiments to test the performance of this dataset using BERT model. In this experiment, we use IndoBERT for our model and the configuration is dropout rate set to 0.1 and we configured positive and negative labels for binary classification, after that we use encoding to tokenizes with padding, truncation, and a maximum length of 64 tokens then we convert array label to PyTorch tensors. For data training we utilize Adamw optimizer with weight decay for regularization and we set our learning rate to 1e-5, then we loop over a maximum of 7 epochs with early stopping to prevent overfitting data and then we track training loss by using cross entropy loss.

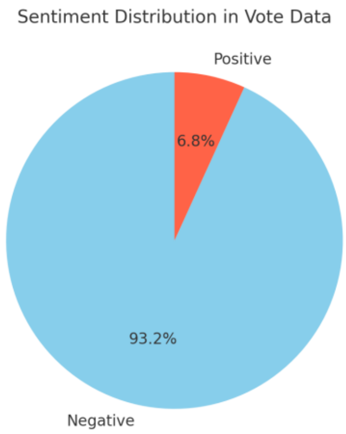

Figure 3. Sentiment distribution label

Our experiment used two primary stages to optimize and evaluate the performance of the BERT model in sentiment analysis. Then we use One-Way ANOVA to evaluate differences among Precision, Recall and F1-Score simultaneously. First, the dataset was split into a data train and data test with varying ratios, enabling the evaluation of the model's performance on unseen data and providing a robust measure of its capabilities. By combining these structured stages with the rigorous statistical analysis provided by One-Way ANOVA, our experiment delivered a comprehensive framework for optimizing and evaluating the BERT model. This methodology not only enhanced the reliability of our findings but also demonstrated the versatility of BERT in handling complex sentiment analysis tasks.

In table 2 summarizes the performance metrics from Precision, Recall, and F1-Score across three dataset split ratios. Overall, the table highlights the trade-offs between split ratios. The 80:20 split provides the best precision for positive and negative class. The 70:30 ratios in recall for the positive class but suffers in precision and F1-Score for the negative class. On the other hand, 90:10 ratios achieve the highest performance for the negative class, making it suitable for datasets where the negative class is higher importance. These insights emphasize the importance of selecting an appropriate split ratio based on specific requirements for sentiment analysis tasks.

Table 2. Split ratio results

Split Ratio		90:10	80:20	70:30
Precision	Positive	0.50	0.70	0.37
	Negative	0.93	0.93	0.94
Recall	Positive	0.13	0.23	0.98
	Negative	0.99	0.99	0.20
F1-Score	Positive	0.21	0.17	0.95
	Negative	0.96	0.96	0.26

Second, resampling techniques were applied to address data imbalance, which is a crucial step in improving the model's ability to learn equitably from both majority and minority classes. By employing methods such as oversampling the minority class and under sampling the majority class, resampling ensured a balanced dataset, leading to more effective and fair learning. The results are shown in table 3. which presents the results of the resampling process. Together, these efforts contributed to a comprehensive and robust framework for enhancing the sentiment analysis task using the BERT model, ensuring the reliability and fairness of the model's predictions across diverse data inputs.

Table 3. Resampling results

Split Ratio		90:10	80:20	70:30
Precision	Positive	0.97	0.93	0.98
	Negative	0.92	0.95	0.87
Recall	Positive	0.92	0.93	0.90
	Negative	0.97	0.98	0.96
F1-Score	Positive	0.94	0.93	0.93
	Negative	0.94	0.97	0.93

Table 3 summarizes how resampling techniques improve model performance across different split ratios, particularly benefiting the minority positive class. The 80:20 split demonstrates the best overall performance, achieving a strong balance between precision, recall, and F1-score for both sentiment classes. Meanwhile, the 90:10 split optimizes the F1-score for the positive class, making it ideal for scenarios where prioritizing positive sentiment detection is crucial. The 70:30 split maintains competitive performance, though its lower precision for the negative class indicates a trade-off in class balance. Additionally, ROC-AUC scores across the split ratios indicate that the 80:20 ratio achieves the highest overall classification reliability, as it maintains a strong balance between sensitivity and specificity. The 90:10 split slightly improves the area under the curve for the positive class, reinforcing the impact of resampling on model generalization, while the 70:30 split shows a decline in negative class precision, affecting overall robustness. These findings emphasize the importance of selecting an appropriate split ratio to optimize both class balance and model generalization.

Discussion

This research presents several key findings that improve sentiment analysis performance, especially on imbalanced datasets. Our model, which incorporates a novel voting system, outperformed previous approaches, such as Geni et al. (2023) achieved 0.81 in F1-Score which provides a useful benchmark for evaluating the performance of various sentiment analysis methods. However, our approach surpasses this baseline by improving upon the performance of all previous models demonstrated in the literature. Our model excels at addressing the imbalance between negative and positive sentiments, a challenge overlooked in earlier studies.

While previous methods focused primarily on traditional classification techniques, our approach adopts a novel voting system to classify each tweet, which enhances the robustness and generalizability of the model. This ensemble-based technique aggregates predictions from multiple classifiers, thereby reducing the risk of overfitting and increasing the model's ability to accurately capture subtle sentiment nuances. The voting mechanism improves baseline performance and ensures a stable, reliable solution for imbalanced sentiment datasets.

In contrast to Geni et al. (2023) and Salnikova (2021) focused on sentiment analysis using the BERT model but reported limited success, largely due to an overemphasis on the model itself while neglecting the critical role of effective text preprocessing. Salnikova's study highlighted that improper preprocessing could lead to suboptimal results, as BERT's contextual understanding capabilities rely heavily on clean and normalized input data. Drawing lessons from Salnikova's findings, we implemented advanced preprocessing techniques tailored to our dataset, including tokenization, stemming, and normalization using the Sastrawi library, to enhance the quality of inputs fed into our BERT-based model. Additionally, Salnikova's findings informed our decision to incorporate resampling techniques to mitigate the impact of class imbalance, which was a significant limitation in their work.

To evaluate the impact between different ratios on model, One-Way ANOVA was conducted to assess variability in confusion matrix (Liu & Wang, 2021). For precision the The F-statistic of 293.18 and p-value of 1.03×10^{6} confirm that precision varies significantly across different split ratios. Post-hoc analysis reveals that the 80:20 split outperformed others, providing a balanced trade-off between minority and majority class precision. Similarly, recall illustrated significant differences with an F-statistic of 10,570.37 and p-value of 2.28×10^{-11}, recall differences are highly pronounced. The 90:10 split ratio achieves the highest recall for the positive class, likely due to increased training samples from oversampling techniques. Finally, for F1 Score, The F-statistic of 11,215.59 and p-value of 1.91×10^{-11} emphasize the sensitivity of F1-score to split configurations. The 80:20 split provides the best overall F1 performance, balancing precision and recall effectively. These results emphasize the critical role of data split ratios in determining model performance and highlight the importance of selecting an appropriate split strategy to optimize sentiment analysis classification outputs. Below is our detailed analysis about this research, from handling overfitting to the usage of resampling from this research.

In this research, the data is shown overfitting because the imbalance data that our dataset have as shown in Figure 3 with a significant majority of negative sentiments compared to positive sentiments. To address this issue, resampling techniques can be employed to balance the dataset. These techniques include oversampling the minority class, under sampling the majority class. By ensuring a more balanced distribution of sentiments, the model can be trained on a dataset that better represents both classes, ultimately improving its performance and fairness in sentiment analysis.

Resampling has proven to be a critical step in addressing data imbalance, significantly enhancing model performance and output quality. It plays a foundational role in building a robust and reliable model. Table 3 highlights the impact of various resampling techniques on model performance across different split ratios, particularly for the minority positive class. While the 80:20 split ratio achieves the best overall performance for both classes, the 90:10 split ratio excels in optimizing the F1-score for the positive class. These results underscore the importance of carefully selecting the appropriate resampling strategy and split ratio to ensure balanced and effective sentiment classification.

Moreover, the integration of BERT methods and advanced preprocessing techniques lays a foundation for future works in tackling imbalanced datasets. Imbalanced datasets often lead to biased models that favour the majority class, reducing the overall effectiveness and fairness of sentiment analysis. Techniques such as Synthetic Minority Oversampling Technique

(SMOTE) can be explored to synthetically generate samples for minority class, improving model, balance and ensuring fairer representation of sentiments. These advancements would contribute significantly to developing robust sentiment analysis models capable of providing accurate and equitable insights.

Conclusion

According to the findings of this research underscore the effectiveness of resampling techniques in mitigating class imbalance in sentiment analysis, thereby enhancing model performance across both sentiment classes. The 80:20 data split emerged as the most optimal configuration, demonstrating consistent and balanced accuracy across key evaluation metrics, including precision, recall, F1-score, and ROC-AUC. These results further validate BERT's capacity to capture contextual nuances in sentiment classification, reinforcing its suitability for such tasks in complex linguistic environments. Moreover, the study highlights the necessity for future advancements, such as the development of customized BERT architectures or the fine-tuning of pre-trained transformer models, to further optimize performance and adaptability. Such enhancements would enable deeper semantic understanding and facilitate more domain-specific applications, ultimately contributing to the broader advancement of sentiment analysis methodologies.

REFERENCES

Ait, A., Izquierdo, J. L. C., & Cabot, J. (2024). HFCommunity: An extraction process and relational database to analyze Hugging Face Hub data. *Science of Computer Programming*, *234*, 1-5. https://doi.org/10.1016/j.scico.2024.103079

Alaparthi, S., & Mishra, M. (2021). BERT: A sentiment analysis odyssey. *Journal of Marketing Analytics*, *9*(2), 118-126. https://doi.org/10.1057/s41270-021-00109-8

Aljedaani, W., Rustam, F., Mkaouer, M. W., Ghallab, A., Rupapara, V., Washington, P. B., ... & Ashraf, I. (2022). Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry. *Knowledge-Based Systems*, *255*, 109780. https://doi.org/10.1016/j.knosys.2022.109780

Batra, H., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2021). Bert-based sentiment analysis: A software engineering perspective. *Database and Expert Systems Applications: 32nd International Conference, DEXA 2021, Virtual Event, September 27–30, 2021, Proceedings, Part I 32*, 138-148. Springer International Publishing. https://doi.org/10.1007/978-3-030-86472-9_13

Catelli, R., Pelosi, S., Comito, C., Pizzuti, C., & Esposito, M. (2023). Lexicon-based sentiment analysis to detect opinions and attitude towards COVID-19 vaccines on Twitter in Italy. *Computers in Biology and Medicine*, *158*, 106876. https://doi.org/10.1016/j.compbiomed.2023.106876

Geni, L., Yulianti, E., & Sensuse, D. I. (2023). Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (JITEKI)*, *9*(3), 746-757. https://doi.org/10.26555/jiteki.v9i3.26490

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of nlp-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, 100059. https://doi.org/10.1016/j.nlp.2024.100059

Kenton, J. D. M. W. C., & Toutanova, L. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of naacL-HLT*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.

Khan, L., Amjad, A., Afaq, K. M., & Chang, H. T. (2022). Deep sentiment analysis using CNN-LSTM architecture of English and Roman Urdu text shared in social media. *Applied Sciences*, *12*(5), 1-18. https://doi.org/10.3390/app12052694

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. *arXiv preprint arXiv:2011.00677*. https://doi.org/10.18653/v1/2020.coling-main.66

Kyrychenko, R. (2021). Sentiment analysis based on the BERT model: attitudes towards politicians using media data. *Advances in Social Science, Education and Humanities Research*, *617*, 39-44.

Lee, S. H., Lee, H., & Kim, J. H. (2022). Enhancing the Prediction of User Satisfaction with Metaverse Service Through Machine Learning. *Computers, Materials & Continua*, *72*(3), 4983 -4997. https://doi.org/10.32604/cmc.2022.027943

Liu, Q., & Wang, L. (2021). t-Test and ANOVA for data with ceiling and/or floor effects. *Behavior Research Methods*, *53*(1), 264-277. https://doi.org/10.3758/s13428-020-01407-2

Mahadevaswamy, U. B., & Swathi, P. (2023). Sentiment analysis using bidirectional LSTM network. *Procedia Computer Science*, *218*, 45-56. https://doi.org/10.1016/j.procs.2022.12.400

Mozafari, M., Farahbakhsh, R., & Crespi, N. (2020). A BERT-based transfer learning approach for hate speech detection in online social media. *Complex Networks and Their Applications VIII: Volume 1 Proceedings of the Eighth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2019 8*, 928-940. Springer International Publishing. https://doi.org/10.1007/978-3-030-36687-2_77

Pamungkas, A. A., Alam, C. N., Atmadja, A. R., & Juliansyah, R. (2024). Integrasi Kamus Multibahasa pada Feed Forward Neural Network dan IndoBERT dalam Pengembangan Chatbot Mobile. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 635-644. https://doi.org/10.29408/edumatic.v8i2.27886

Rosid, M. A., Fitrani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving text preprocessing for student complaint document classification using sastrawi. *IOP Conference Series: Materials Science and Engineering*, *874*, (1), 1-7. IOP Publishing. https://doi.org/10.1088/1757-899X/874/1/012017

Sabilillah, F. T., Winarno, S., & Abiyyi, R. B. (2024). Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing berdasarkan Judul Tugas Akhir. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 585-594. https://doi.org/10.29408/edumatic.v8i2.27791

Taherdoost, H., & Madanchian, M. (2023). Artificial intelligence and sentiment analysis: A review in competitive research. *Computers*, *12*(2), 1-5. https://doi.org/10.3390/computers12020037

Tran, T. A., Duangsuwan, J., & Wettayaprasit, W. (2021). A new approach for extracting and scoring aspect using SentiWordNet. *Indonesian Journal of Electrical Engineering and Computer Science*, *22*(3), 1731-1738. https://doi.org/10.11591/ijeecs.v22.i3.pp1731-1738

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731-5780. https://doi.org/10.1007/s10462-022-10144-1

Zharifa, A. H. A., & Ujianto, E. I. H. (2024). Analisis Sentimen Publik di Twitter Pasca Debat Kelima Pilpres 2024 dengan Naive Bayes. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 754-763. https://doi.org/10.29408/edumatic.v8i2.28048