International Journal of Engineering and Computer Science Applications (IJECSA)
Vol. 4, No. 1, March 2025, pp. 59∼70
ISSN: 2828-5611
DOI: doi.org/10.30812/ijecsa.v4i1.4981

Bidirectional Encoder Representations from Transformers Fine-Tuning for Sentiment Classification of Cek Bansos Reviews

Erna Haerani[1], Alam Rahmatulloh[1], Souhayla Elmeftahi[2]
[1]Universitas Siliwangi, Kota Tasikmalaya, Indonesia
[2]École Nationale des Sciences Appliquées d'Al Hoceima, Morocco

Article Info

Article history:

Received March 12, 2025
Revised March 21, 2025
Accepted April 18, 2025

Keywords:

Cek Application
BERT Sentiment Analysis
Streamlit
User Reviews

ABSTRACT

Social assistance programs are essential government initiatives aimed at supporting underprivileged communities. One such program is facilitated through the Cek Bansos application, which enables users to check their eligibility for social aid. However, user experiences with the application vary, leading to various sentiments in their reviews. Understanding these sentiments is crucial for improving the application's functionality and user satisfaction. This study focuses on sentiment analysis of user reviews of the Cek Bansos application by leveraging a fine-tuned Indonesian-language Bidirectional Encoder Representations from Transformers (BERT) model. **This research aims** to evaluate the BERT model's effectiveness in classifying sentiments in user reviews and provide insights that could improve the Cek Bansos application. **This research method** is the BERT model was fine-tuned using hyperparameters such as a learning rate of 3e-6, batch size of 16, and 9 epochs. The dataset consisted of 8,000 reviews, divided into training (70%), validation (20.1%), and test (9.9%) sets. Review scores were manually categorized, where ratings of 1 to 2 were classified as negative sentiment, 3 as neutral, and 4 to 5 as positive. **The results of this research** are as follows: the fine-tuned model achieved an accuracy of 77%, with additional evaluation metrics such as precision, recall, and F1 score, demonstrating the model's effectiveness in identifying positive, negative, and neutral sentiments separately. **This study concludes** that the BERT model provides a reliable method for sentiment classification of user reviews, which could support developers and policymakers in refining the Cek Bansos application to enhance user experience. Additionally, a web-based application developed using Streamlit allows government officials to visualize sentiment trends in real time, improving their understanding of user feedback. Future research could further explore alternative machine learning models and additional linguistic features to improve sentiment classification accuracy and the overall user experience.

Corresponding Author:

Alam Rahmatulloh
Universitas Siliwangi, Kota Tasikmalaya, Indonesia.
Email: alam@unsil.ac.id

How to Cite: E. Haerani, A. Rahmatulloh, and S. Elmeftahi, "Bidirectional Encoder Representations from Transformers Fine-Tuning for Sentiment Classification of Cek Bansos Reviews," *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 4, no. 1, pp. 59-70, Mar. 2025. doi: 10.30812/ijecsa.v4i1.4981.

1. INTRODUCTION

Mobile phone applications are becoming increasingly important in everyday life in the era of modern technology, including in government work and public services. One important application is the Social Assistance Check application, which is designed to help the public access information about social assistance. Reviews from users of this application are a valuable source of information that can provide insight into user satisfaction and areas for improvement. However, manual analysis of these reviews is highly inefficient, given the large and diverse volume of data [1]. This study addresses a gap in sentiment analysis research by using a fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model, specifically tailored for the Indonesian language, to analyze user feedback and improve application functionality. Previous studies have shown the effectiveness of BERT models in sentiment classification, particularly for Indonesian-language text [2], but limited research focuses on social assistance apps like Cek Bansos. Unlike traditional machine learning models such as Naïve Bayes [3–8] or SVM [9–11], BERT excels in understanding the bidirectional context of words [12–14], which improves its performance in sentiment classification tasks. This study aims to fill the gap by using BERT fine-tuning to analyze user sentiments in the Cek Bansos application. By comparing BERT with traditional sentiment analysis models, this research provides insights into how BERT can enhance sentiment classification in the context of social assistance apps, which have received limited attention in the existing literature.

Machine learning-based sentiment analysis methods, especially the BERT model, can be used to address this problem. offers an effective solution. BERT has been shown to excel in understanding natural language context due to its transformer-based architecture and its ability to consider the bidirectional context of words in a sentence [15]. Compared to Term Frequency - Inverse Document Frequency (TF-IDF)-based approaches [16] or other machine learning techniques, the BERT model can perform sentiment classification better because of this adjustment [2]. Research [17] the comparison of the Random Forest, Naïve Bayes, and BERT Algorithms for Multi-Class Classification on the Cable News Network (CNN) article is the basis for selecting this model. In that study, BERT had the highest accuracy score of 92 percent, with an average F1 score of 92 percent. Many studies have been conducted previously on sentiment classification, such as this study [18]. This study explores using BERT for multi-label sentiment classification in imbalanced code-switching text. Developing a BERT fine-tuning method with a data augmentation and ensemble learning approach to handle unbalanced emotion distribution shows better results than previous models. This research has weaknesses such as imbalance, potential dataset quality issues, and the risk of overfitting on small datasets. The research [19] applies transfer learning from BERT to the CNN-BiLSTM model to improve sentiment analysis performance. The results show that the BERT model provides superior binary classification performance compared to traditional embedding methods such as Word2Vec and GloVe.

This research has weaknesses such as **the limitations** of model generalization to other domains, the risk of overfitting on small datasets, interpretability challenges, and the need for high computing resources due to the complexity of BERT. The study [20] used the BERT method for multi-aspect sentiment analysis on reviews of the Bibit investment application. The analysis results show that the service aspect has the highest accuracy, while the system aspect shows the lowest accuracy, thus providing an improvement input for application developers. The shortcoming of this study is the low accuracy in the system aspect, which shows the challenge of the BERT model in handling complex data. The study [21] used transfer learning with the BERT model for sentiment classification of reviews. The results showed that the BERT model achieved superior accuracy compared to traditional classification methods. Comparative testing on positive and negative reviews on Yelp showed that BERT fine-tuning could improve performance in sentiment classification. This study has weaknesses, such as the risk of overfitting in small datasets and the need for high computing due to the complexity of the model. Research [22] shows that the Naive Bayes method has weaknesses in handling too many features, which can lead to errors in sentiment classification. This model is also sensitive to unbalanced data distributions and often cannot identify complex relationships between features. The research [23] applies a transformer-based sentiment analysis methodology to examine temporal models of attitudes towards famous politicians from 2001-2021. The study used target-BERT for automatic attitude analysis, with results showing an F1-score quality of 0.799 for the Ukrainian model and 0.741 for the Russian model. These studies show that BERT fine-tuning can be effectively applied in various sentiment analysis contexts, including application reviews. This provides a strong foundation for this study to apply a similar method in analyzing reviews of the Cek Bansos application.

The study [21] used transfer learning with the BERT model for sentiment classification of reviews. The results showed that the BERT model achieved superior accuracy compared to traditional classification methods. Comparative testing on positive and negative reviews on Yelp showed that BERT fine-tuning could improve performance in sentiment classification. This study has weaknesses, such as the risk of overfitting in small datasets and the need for high computing due to the complexity of the model. Research [22] shows that the Naive Bayes method has weaknesses in handling too many features, which can lead to errors in sentiment classification. This model is also sensitive to unbalanced data distributions and often cannot identify complex relationships between features. The research [23] applies a transformer-based sentiment analysis methodology to examine temporal models of attitudes towards famous politicians from 2001 to 2021. The study used target-BERT for automatic attitude analysis, with results showing an F1-score quality of 0.799 for the Ukrainian model and 0.741 for the Russian model. **There are gaps** that have not been resolved by previous research, namely: the risk of overfitting when using small datasets, which could affect the generalization capability of the model; the computational demand of BERT models, which requires high computational resources, making it less accessible for certain applications; and the application of BERT fine-tuning in more specific contexts, such as application reviews, which remains underexplored. **The difference between** this research and the previous ones is that this study focuses on applying BERT fine-tuning specifically to sentiment analysis of Cek Bansos reviews, addressing the gap in applying this method to such reviews. Moreover, this study aims to optimize the model's performance by reducing the risk of overfitting and exploring strategies to mitigate the high computational costs of using the BERT model. This study demonstrates that BERT fine-tuning can effectively classify sentiments in application reviews, such as those for Cek Bansos. By addressing the limitations of previous research, such as overfitting and high computational costs, the findings suggest that BERT can be a viable solution for sentiment analysis in diverse review contexts.

Therefore, **this study aims** to apply BERT fine-tuning to classify sentiment in reviews of the Cek Bansos application. By utilizing a dataset of user reviews taken from the Google Play Store platform, this study will explore the effectiveness of the BERT model in identifying positive, negative, and neutral sentiment. Through this approach, it is hoped that a better understanding of public perceptions of the application can be obtained, as well as recommendations for improvements that can be applied to improve the quality of service. This study's results are expected to significantly contribute to application developers and government officials in improving the user experience and the effectiveness of social assistance programs.

2. RESEARCH METHOD

This research method is designed to implement the fine-tuning of the BERT model in the sentiment classification of the Cek Bansos application review. As shown in *Figure 1*, the research process is divided into several stages, from data collection to deployment.

Figure 1. Research Stages

2.1. Data Collection and Pre-processing

The data for this study was obtained from Google Play Store reviews related to the Social Assistance Check application using web scraping techniques. Python provides several libraries that can be used for web scraping, such as BeautifulSoup, Scrapy, and Google Play Scraper [24]. This study used Google Play Scraper to collect raw data. Furthermore, the review text was cleaned of special characters, numbers, and irrelevant punctuation. All text was converted to lowercase for consistency. Finally, the text was divided into tokens that the model could process using the BERT tokenizer.

2.2. Data Labeling

Data labeling is an important process in sentiment classification that allows the model to understand user opinions in reviews [25]. In this study, labeling was done based on the score from the category column taken from the Cek Bansos application review. Each review was categorized into three labels: positive, negative, and neutral.

2.3. Dataset Division

Proper dataset division ensures that the model can learn from training data and be evaluated with previously unseen data. The labeled dataset is divided into three parts: training data (70%), validation data (20.1%), and test data (9.9%). These are used to evaluate the final performance of the model. This division ensures that the model generalizes well on new data and prevents overfitting [26].

2.4. BERT Implementation

The BERT model is used to study Indonesian language corpus data in the pre-training process. The database used consists of four billion words, covering various formal and informal language variations from twelve different corpora in Indonesia. This dataset is then trained using a conventional BERT architecture, which consists of twelve transformer layers [27].

2.5. Fine-Tuning BERT Model

The process of retraining existing machine learning models is known as fine-tuning. This process uses a more specialized dataset relevant to the purpose of the model [28]. The pre-trained BERT model is retrained using the prepared dataset at this stage. Hyperparameters such as learning rate, batch size, and number of epochs are set based on literature and initial experiments. Optimization techniques such as Adam are used to refine the model weights based on prediction errors during training. The dataset was divided into training (70%), validation (20.1%), and test (9.9%) sets. This division ensures the model can generalize well on unseen data and prevents overfitting. Stratified sampling was used to maintain a balanced proportion of labels across the splits.

2.6. Model Evaluation

The model is evaluated using accuracy, precision, recall, and F1-score metrics obtained from the previous fine-tuning stage. Error analysis is also performed to understand where the model might fail. The fine-tuning process used the Adam optimizer with a learning rate of 3e-6, a batch size of 16, and 9 epochs, as recommended in similar studies [15]. Based on initial experiments, this setup was selected to optimize the model's performance. Overfitting was mitigated by using early stopping during training and cross-validation on the validation set to ensure the model generalizes well on unseen data.

2.7. Deployment Using Streamlit

The trained model is used to integrate into an interactive web application. Streamlit is an open-source framework created to facilitate the creation of interactive web applications using the Python programming language [29]. The application allows users to enter review text and get sentiment classification results in real-time [30]. User testing of the application shows that the user interface is simple and easy to use, and sentiment classification results are displayed quickly.

3. RESULT AND ANALYSIS

This section presents the results and discussion of sentiment classification using the BERT model, including evaluation of model performance based on accuracy, precision, recall, and F1-score. The following are the results and discussion of sentiment classification using the BERT model:

3.1. Data Collection and Pre-processing

In the data collection stage, as many as 8,000 reviews of the Social Assistance Checking Application users were successfully retrieved from the Google Play Store using the scraping technique via the package google_play_scraper. The scraper sets the language of the reviews to Indonesian. Data collection was carried out on June 15, 2024, with the most recent reviews. From the data collected, two main columns were taken: the description column, which contains the review text, and the category column, which contains the user rating. An example of the results of data collection is in *Table 1*.

Table 1. Dataset

No	Description	Category
0	Kode OTP jangan Cuma lewat whatsapp tolong diupdate fiturnya...	5
1	Sudah berhasil login dan sudah buat pin juga. Dikeluarin dari aplikasinya.	1
2	Aplikasi buruk, ingin mengganti data individu sertifikat, malah stuck..	1
3	Aplikasinya gak bisa masuk. Masukin email ga ada no verifikasi masuk...	1

After collection, the data is unstructured. Therefore, a preprocessing approach is used that can be used in various natural language processing tasks by utilizing packages from the Natural Language Toolkit. The preprocessing process is carried out to clean the data in the dataset, including removing numbers, punctuation marks, emoticons, and irrelevant words (stopwords). However, stemming is not performed because it can change the sentence's overall meaning.

Table 2. Pre-processing Results

Pre-Processing	Result
Original text	Apk nya aneh, udah masukin email dengan baik. tetep ga muncul^{2} kode otp nya, eh sekalinya muncul telat waktunya udah abis 😭
Cleaning Data	apk nya aneh udah masukin email dengan baik tetep ga muncul kode otp nya eh sekalinya muncul telat waktunya udah abis
Normalization Word	apk nya aneh udah masukin email dengan baik tetep ga muncul kode otp nya eh sekalinya muncul telat waktunya udah abis
Stopwords Removal	apk aneh masukin email baik ga muncul kode otp sekalinya muncul telat abis

Table 2 shows the stages of text data processing for sentiment analysis starting from the original text, which is a user review about problems in the application that includes informal words and emoticons, for example: "Apk nya aneh, udah masukin email dengan baik. tetep ga muncul^{2} kode otp nya, eh sekalinya muncul telat waktunya udah abis 😭." In the cleaning stage, the text is cleaned of unnecessary characters, such as emoticons and special symbols, resulting in cleaner text: "The APK is weird, I entered the email correctly, but the otp code still does not appear, oh, once it appears too late, the time is up." The next step is normalization, where informal words are changed to a more formal form, but the word changes are insignificant in this example. After that, in the stopwords removal stage, common words that are not important for sentiment analysis such as "his," "with," and "and" are removed, resulting in a more concise text that focuses on important keywords, namely: "Strange Apk, enter email, good, no otp code appears, once it appears, it's too late."

The model achieved an accuracy of 77%, with precision, recall, and F1-score values of 75%, 74%, and 74%, respectively. This demonstrates the model's ability to effectively classify positive and negative sentiments, but there is room for improvement in classifying neutral sentiment. A confusion matrix was used to analyze the model's performance on the test set, revealing that the model struggles more with neutral sentiment. Future improvements could incorporate additional features, such as linguistic analysis, to better identify neutral sentiment. Error analysis revealed that the model performs well on clear positive and negative reviews but misclassifies neutral reviews due to the subtle language used. This suggests the need for more sophisticated techniques to handle neutral sentiment effectively.

3.2. Data Labeling

The processed reviews are then labeled based on their sentiment. Labeling is done using the score from the category column: a rating ≥ 4 is given a positive label, a rating < 2 is given a negative label, and a rating of 3 is considered neutral. After labeling, the dataset produces 2000 positive sentiments, 2000 negative sentiments, and 2000 neutral sentiments. This process is done manually to ensure label accuracy. The following is the proportion of sentiment class target variables after labeling in Figure 2.

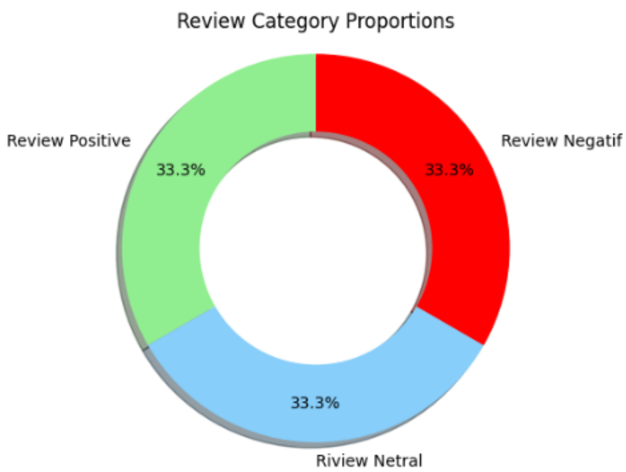

Figure 2. Proportion of target variables

3.3. Dataset Division

Of the total 6,000 available data points, the data is divided into three parts to ensure the model can be properly trained and evaluated. The composition is 70%, or as many as 4200 training data points used to train the model, 20.1% validation data, or as many as 1206 used to monitor model performance during training, and 9.9% test data, or as many as 594 used to evaluate the final performance of the model.

3.4. BERT Implementation

The Pre-Trained Model dictionary from indobencmark/indobert-base-p1 will be used for the fine-tuning stage. One of the contents of the Transformer library, namely BertTokenizer, will be used for fine-tuning. For fine-tuning, the BERT model is trained with the Optimizer Adam for each task with a learning rate = 3e-6 and an epoch number = 9.

3.5. Fine-Tuning BERT Model

The dataset that has been previously divided into training data, validation data, and test data is used to fine-tune the BERT model. The training data is loaded according to the predefined hyperparameter settings, such as a learning rate of 3e-6. In the training process, a forward pass is used to make predictions and loss calculations, and a backward pass is used to update the model weights.

3.6. Model Evaluation

At this stage, the confusion matrix is used to show the results of model training and experimentation on test data. This evaluates how the model correctly predicts sentiment on the test data. *Figure 3* shows the confusion matrix. From *Figure 3*, the model tends to classify negatively and positively labeled data better. However, the model still often misclassifies neutrally labeled data. After the trial is conducted per epoch, the results will be displayed as a learning curve, as seen in *Figure 4*.

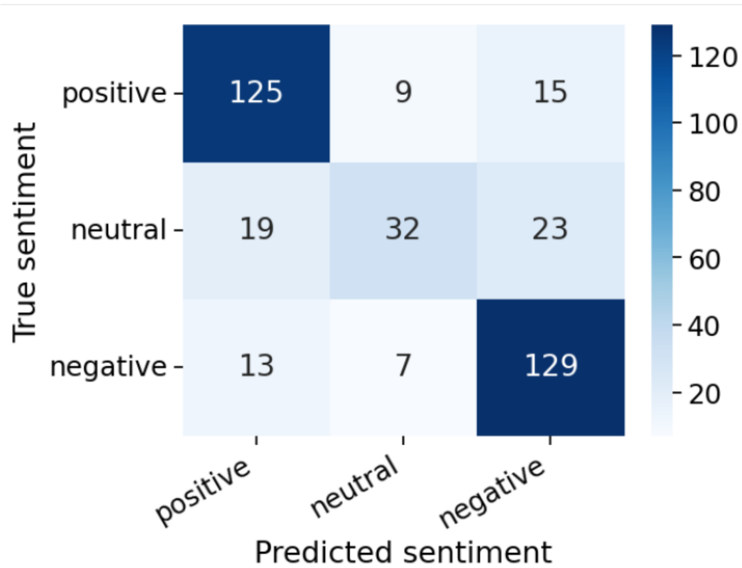

Figure 3. Confusion Matrix

In *Figure 4*, the train acc result states that the train acc value in the training history is 0.9. Meanwhile, the validation acc result states that the value in the training history is 0.7. *Figure 5* shows the last step. Here, the classification_report function from sklearn is used to calculate accuracy, precision, recall, and F1 score. This research demonstrates that fine-tuning the BERT model can significantly improve sentiment classification accuracy for Cek Bansos application reviews. The model achieved 77% accuracy, with further improvements possible by addressing the neutral sentiment classification. The study also highlights the potential of Streamlit for deploying sentiment analysis models in real-time applications. Future research could explore alternative models, such as RoBERTa or DistilBERT, to improve classification accuracy and identify nuanced sentiments more effectively. It can be seen from *Figure 5* that this sentiment classification model performs quite well, especially in identifying positive and negative data. However, the model still needs improvement in identifying neutral data.

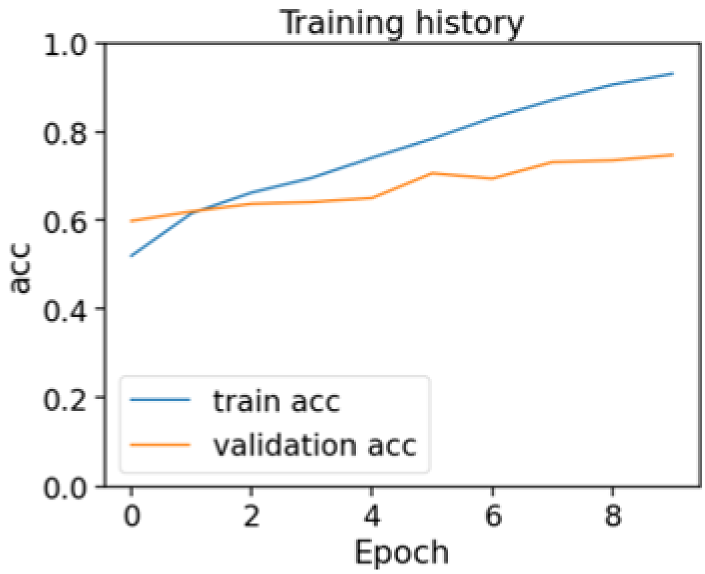

Figure 4. Learning Curve

	precision	recall	f1-score	support
positive	0.80	0.84	0.82	149
neutral	0.67	0.43	0.52	74
negative	0.77	0.87	0.82	149
accuracy			0.77	372
macro avg	0.75	0.71	0.72	372
weighted avg	0.76	0.77	0.76	372

Figure 5. Classification Report

3.7. Deployment Using Streamlit

The BERT model has been successfully fine-tuned and performs well in sentiment classification of reviews. The next step is to integrate it into a web application using Streamlit. This process involves several steps: first, the model is imported using the PyTorch framework to make predictions; second, a simple user interface is created so that users can enter review text and get sentiment classification results (positive, negative, or neutral); third, the application processes reviews through the BERT model and displays prediction results in real-time. In *Figure 6*, the researcher entered the sentence "The application is good and very helpful, I like it." A positive prediction appears with a green background and a smiling emoji with 99.26% confidence. In *Figure 7*, the researcher entered the sentence "very late, the application is bad. Every time it is opened, there is an error," and a negative prediction appears with a pink background and a disappointed emoji with a confidence of 92.81%. In *Figure 8*, the researcher entered the sentence "Already not used for anything, but just in case I want an update, I'll log in again." A neutral prediction appears with a blue background and a flat-faced emoji with a 73.32% confidence level.

Figure 6. Positive Prediction Sample

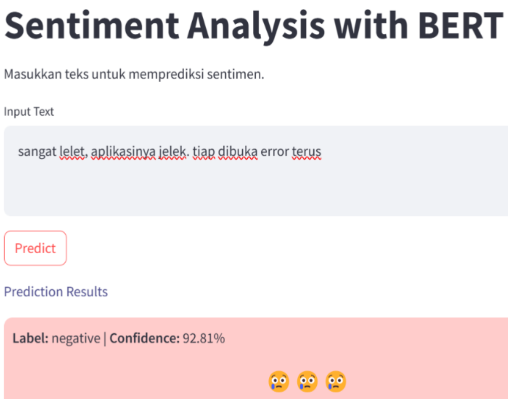

Figure 7. Negative Prediction Sample

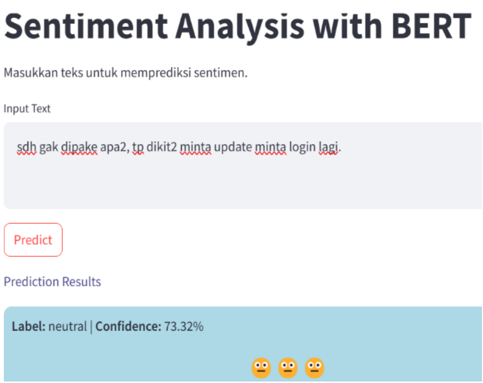

Figure 8. Neutral Prediction Sample

3.8. Findings of the Research

The findings of this research are as follows: The fine-tuned BERT model achieved an accuracy of 77%, with precision, recall, and F1-score values of 75%, 74%, and 74%, respectively. These results demonstrate the model's capability to accurately identify positive and negative sentiments in the Cek Bansos reviews. However, the model's performance on neutral sentiment classification was less accurate, suggesting room for improvement in handling this category. The results of this research **are in line** with or supported by previous research. Specifically, BERT has been shown in prior studies to outperform traditional machine learning models like Naïve Bayes and SVM in sentiment analysis tasks. For instance, previous research, such as a study [17] on BERT's performance for multi-class classification demonstrated similar high performance with BERT, achieving accuracy levels comparable to the results in this study. The findings are also supported by the study on sentiment analysis in Indonesian reviews, where BERT's contextual understanding was highlighted as a key factor in its superior performance. As seen in *Table 3*, this research shows a competitive accuracy of 77%, comparable to other studies such research [20], but with room for improvement compared to research [17]. The lower performance on neutral sentiment classification in this study aligns with challenges seen in Setyani's work, where complex sentiment expressions were harder for BERT to classify accurately.

Table 3. Comparison BERT Model

Study	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
This Research	BERT	77	75	74	74
[17]	BERT	92	91	90	91
[20]	BERT	80	78	79	78

4. CONCLUSION

This research shows that fine-tuning the BERT model can be used effectively for sentiment classification of reviews of the Cek Bansos application. With an accuracy of 77%, the model predicts positive and negative sentiment well, but the model still needs improvement in detecting data with neutral sentiment. This model performs well in classifying sentiment, as proven by the simple deployment that has been made. Deployment using streamlit allows users to easily carry out sentiment analysis in real time and provide feedback for Cek Bansos application developers. For future research, several areas of development that need attention include improving the ability to identify neutral sentiment. In addition, research can compare BERT with other transformer models, such as RoBERTa or DistilBERT, to find more effective models in classifying sentiment in Indonesian.

ACKNOWLEDGEMENTS

We sincerely thank Universitas Siliwangi, particularly the Program Studi Information Systems and Informatika, for providing the necessary resources and support for this research. We also appreciate the contributions of our colleagues and fellow researchers whose insights have helped refine this work. Lastly, we thank our families and friends for their unwavering encouragement throughout this research journey. Thank you all for your support.

REFERENCES

[1] E. D. Madyatmadja *et al.*, "Sentiment Analysis on User Reviews of Threads Applications in Indonesia," *Journal Européen des Systèmes Automatisés*, vol. 57, no. 4, pp. 1165–1171, Aug. 2024. DOI: 10.18280/jesa.570423.

[2] K. S. Nugroho *et al.*, "BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews," in *6th International Conference on Sustainable Information Engineering and Technology 2021*, Malang Indonesia: ACM, Sep. 2021, pp. 258–264. DOI: 10.1145/3479645.3479679.

[3] S. N. Asvia *et al.*, "Comparison of Naïve Bayes and Random Forest on Sentiment Analysis of Brand Reputation Provider Live. On," in *2024 18th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, Bali, Indonesia: IEEE, Oct. 2024, pp. 1–5. DOI: 10.1109/TSSA63730.2024.10863962.

[4] A. Rahmatulloh *et al.*, "Sentiment Analysis of Ojek Online User Satisfaction Based on the Naïve Bayes and Net Brand Reputation Method," in *2021 9th International Conference on Information and Communication Technology (ICoICT)*, Yogyakarta, Indonesia: IEEE, Aug. 2021, pp. 337–341. DOI: 10.1109/ICoICT52021.2021.9527466.

[5] F. F. Sabiq *et al.*, "Performance Comparison of Multinomial and Bernoulli Naïve Bayes Algorithms with Laplace Smoothing Optimization in Fake News Classification," in *2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)*, Indonesia: IEEE, Aug. 2024, pp. 19–24. DOI: 10.1109/ICoABCD63526.2024.10704399.

[6] S. Elmeftahi, M. D. Rakhman, and A. Rahmatulloh, "Comparison of Machine Learning Algorithms in Detecting Contaminants in Drinkable Water," *Innovation in Research of Informatics (INNOVATICS)*, vol. 6, no. 1, pp. 7–14, Mar. 2024. DOI: 10.37058/innovatics.v6i1.10385.

[7] R. Ardianto *et al.*, "Sentiment Analysis on E-sports for Education Curriculum Using Naive Bayes and Support Vector Machine," *Jurnal Ilmu Komputer dan Informasi*, vol. 13, no. 2, pp. 109–122, Jul. 2020. DOI: 10.21609/jiki.v13i2.885.

[8] P. Pristiyono *et al.*, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm," *IOP Conference Series: Materials Science and Engineering*, vol. 1088, no. 1, p. 012 045, Feb. 2021. DOI: 10.1088/1757-899X/1088/1/012045.

[9] W. Leong, R. Kelani, and Z. Ahmad, "Prediction of air pollution index (API) using support vector machine (SVM)," *Journal of Environmental Chemical Engineering*, vol. 8, no. 3, p. 103 208, Jun. 2020. DOI: 10.1016/j.jece.2019.103208.

[10] A. Kurani *et al.*, "A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting," *Annals of Data Science*, vol. 10, no. 1, pp. 183–208, Feb. 2023. DOI: 10.1007/s40745-021-00344-x.

[11] D. Mustafa Abdullah and A. Mohsin Abdulazeez, "Machine Learning Applications based on SVM Classification A Review," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 81–90, Apr. 2021. DOI: 10.48161/qaj.v1n2a50.

[12] A. K. Kalusivalingam *et al.*, "Leveraging BERT and LSTM for Enhanced Natural Language Processing in Clinical Data Analysis," *International Journal of AI and ML*, vol. 2, no. 3, pp. 1–24, Feb. 2021.

[13] N. M. Gardazi *et al.*, "BERT applications in natural language processing: A review," *Artificial Intelligence Review*, vol. 58, no. 6, p. 166, Mar. 2025. DOI: 10.1007/s10462-025-11162-5.

[14] B. K. Mandal, P. Majumder, and B. P. Tewari, "Role of BERT Model for Sequential Text Classification in Biomedical Abstracts," in *Real-World Applications and Implementations of IoT*, A. Acharyya, P. Dey, and S. Biswas, Eds., Singapore: Springer Nature Singapore, 2025, pp. 67–82.

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

[16] R. Rizal *et al.*, "Unveiling the Truth: Detecting Fake News Using SVM and TF-IDF," in *2025 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*, Bandung, Indonesia: IEEE, Feb. 2025, pp. 1–6. DOI: 10.1109/ICADEIS65852.2025.10933324.

[17] N. Husin, "Komparasi Algoritma Random Forest, Naïve Bayes, dan Bert Untuk Multi-Class Classification Pada Artikel Cable News Network (CNN)," *Jurnal Esensi Infokom : Jurnal Esensi Sistem Informasi dan Sistem Komputer*, vol. 7, no. 1, pp. 75–84, May 2023. DOI: 10.55886/infokom.v7i1.608.

[18] T. Tang, X. Tang, and T. Yuan, "Fine-Tuning BERT for Multi-Label Sentiment Analysis in Unbalanced Code-Switching Text," *IEEE Access*, vol. 8, pp. 193 248–193 256, 2020. DOI: 10.1109/ACCESS.2020.3030468.

[19] N. J. Prottasha *et al.*, "Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning," *Sensors*, vol. 22, no. 11, p. 4157, May 2022. DOI: 10.3390/s22114157.

[20] S. Setyani, "Multi Aspect Sentiment Analysis of Mutual Funds Investment App Bibit Using BERT Method," *International Journal on Information and Communication Technology (IJoICT)*, vol. 9, no. 1, pp. 44–56, Jul. 2023. DOI: 10.21108/ijoict.v9i1.718.

[21] A. Areshey and H. Mathkour, "Transfer Learning for Sentiment Classification Using Bidirectional Encoder Representations from Transformers (BERT) Model," *Sensors*, vol. 23, no. 11, p. 5232, May 2023. DOI: 10.3390/s23115232.

[22] I. Wickramasinghe and H. Kalutarage, "Naive Bayes: Applications, variations and vulnerabilities: A review of literature with code snippets for implementation," *Soft Computing*, vol. 25, no. 3, pp. 2277–2293, Feb. 2021. DOI: 10.1007/s00500-020-05297-6.

[23] S. Salnikova and R. Kyrychenko, "Sentiment Analysis Based on the BERT Model: Attitudes Towards Politicians Using Media Data," in *International Conference on Social Science, Psychology and Legal Regulation (SPL 2021)*, Kyiv, Ukraine, 2021, pp. 39–44. DOI: 10.2991/assehr.k.211218.007.

[24] V. A. Flores, P. A. Permatasari, and L. Jasa, "Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword," *Majalah Ilmiah Teknologi Elektro*, vol. 19, no. 2, p. 157, Dec. 2020. DOI: 10.24843/MITE.2020.v19i02.P06.

[25] I. G. T. Isa and B. Junedi, "Hyperparameter Tuning Epoch dalam Meningkatkan Akurasi Data Latih dan Data Validasi pada Citra Pengendara," *Prosiding Sains Nasional dan Teknologi*, vol. 12, no. 1, pp. 231–237, Nov. 2022. DOI: 10.36499/psnst.v12i1.6697.

[26] E. Verianto, "Penerapan LSTM Dengan Regularisasi Untuk Mencegah Overfitting Pada Model Prediksi Tingkat Inflasi di Indonesia," *Jurnal Sistem Informasi dan Sistem Komputer*, vol. 9, no. 2, pp. 195–204, Jul. 2024. DOI: 10.51717/simkom.v9i2.460.

[27] T. Pratama and S. Rjito, "IndoXLNet: Pre-Trained Language Model for Bahasa Indonesia," *International Journal of Engineering Trends and Technology*, vol. 70, no. 5, pp. 367–381, Jun. 2021. DOI: 10.14445/22315381/IJETT-V70I5P240.

[28] H. H. Zain, R. M. Awannga, and W. I. Rahayu, "Perbandingan Model SVM, KNN dan Naïve Bayes untuk Analisis Sentiment pada Data Twitter: Studi Kasus Calon Presiden 2024," *JIM: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, vol. 8, no. 3, pp. 2083–2093, 2023.

[29] M. R. A. Rustiawan and P. T. Prasetyaningrum, "Analisis Sentimen Terhadap Klinik Natasha Skincare di Yogyakarta Dengan Metode Google Review," *Journal of Information Technology Ampera*, vol. 5, no. 1, pp. 75–89, May 2024. DOI: 10.51519/journalita.v5i1.556.

[30] A. Putranto, N. L. Azizah, and I. R. I. Astutik, "Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit," *Kesatria : Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 4, no. 2, pp. 442–452, Apr. 2023. DOI: 10.30645/kesatria.v4i2.180.

[This page intentionally left blank.]