JOURNAL INFORMATIC, EDUCATION AND MANAGEMENT (JIEM)
Vol 8 No 1 (2026): September 2025 - February 2026, pp. 643 ~ 649
ISSN: 2716-0696, DOI: 10.61992/jiem.v8i1.250

Sentiment Analysis of JMO Application Reviews on the Google Play Store Using BERT

Hendi Putra Wijaya [1*], Adhityah Anugrah [1], Mira Afrina [1], Ali Ibrahim [1]
[1] Universitas Sriwijaya

Article Info
Article history:
Received 1 January 2026
Revised 4 January 2026
Accepted 7 January 2026

Keywords:
Sentiment Analysis, BERT, Natural Language Processing

ABSTRACT
The development of digital technology has encouraged increased use of online-based public service applications, including the JMO (Jamsostek Mobile) application developed by BPJS Ketenagakerjaan to provide easy access for its participants. This application has received many user reviews on the Google Play Store, reflecting the level of satisfaction and public perception of service quality. However, the large and unstructured volume of comments makes manual analysis difficult. This study aims to conduct sentiment analysis on user comments about the JMO application on the Play Store using the Bidirectional Encoder Representations from Transformers (BERT) model. The research method involves collecting comments through web scraping, text preprocessing (such as data cleaning, normalization, and tokenization), and sentiment labeling (positive, negative, and neutral). Evaluation using precision, recall, and F1-score is employed to describe the results. The study is expected to identify patterns of user sentiment and public perceptions of the JMO application. It is also expected to serve as an evaluation material and input for developers to improve service quality and user experience.

Corresponding Author:
Hendi Putra Wijaya | Universitas Sriwijaya
Email: hendiputrawijaya@unsri.ac.id

1. Introduction

In the context of employment social security services, digital transformation has become crucial due to the high volume of participant interactions, especially through application-based services such as JMO, which is used by millions of active workers. One form of this digital transformation is the Jamsostek Mobile (JMO) application developed by BPJS Ketenagakerjaan as an online platform for employment social security services. This application allows participants to check account balances, submit claims, and access membership information without the need for in-person visits (BPJS Ketenagakerjaan, 2023).

As the use of the JMO application has increased, it has received thousands of user reviews on the Google Play Store, including praise, complaints, and suggestions for improvement. These reviews directly reflect users' perceptions and experiences regarding the quality of the services provided (Google Play Console, 2024). However, the unstructured nature of textual comments makes manual analysis inefficient and prone to subjectivity. Therefore, an Artificial Intelligence (AI)–based approach is required to automatically extract opinions and emotions from user-generated text (Liu, 2015; Medhat et al., 2014).

One of the most widely used methods for processing public opinion is sentiment analysis, which is applied to identify and classify opinions into positive, negative, or neutral categories based on textual expressions (Pang & Lee, 2008). Traditional approaches based on lexicons and machine learning algorithms such as Naïve Bayes and Support Vector Machines (SVM) have long been used, but their performance is limited in capturing complex semantic contexts. Consequently, deep learning–based approaches have emerged, enabling a deeper understanding of language context through vector representations.

One of the most influential deep learning models in the field of Natural Language Processing (NLP) is BERT, developed by Google (Devlin et al., 2018). BERT employs a Transformer architecture that can capture the full contextual meaning of words within a sentence (Vaswani et al., 2017). This approach has demonstrated strong performance in various NLP tasks such as text classification, sentiment analysis, and named entity recognition (Sun et al., 2019). For the Indonesian language, IndoBERT has been developed as an adaptation of BERT to better fit local linguistic structures (Koto et al., 2020).

Several studies have shown the superiority of IndoBERT in analyzing Indonesian-language texts compared to conventional models. IndoBERT has achieved higher accuracy than LSTM and Naïve Bayes in e-commerce review sentiment analysis (Putri et al., 2021). Other studies have demonstrated that BERT can classify social media sentiment with accuracy exceeding 90% (Rahmawati & Pratama, 2022). Furthermore, Oktaviani and Wahyudi (2022) applied BERT to analyze government application reviews on the Play Store and obtained similar results. These findings indicate that Transformer-based models are highly suitable for application in public service platforms such as JMO.

Therefore, this study aims to apply the IndoBERT model to identify user sentiment toward the JMO application on the Play Store. The analysis is expected to provide empirical insights into public perceptions of the performance and quality of the JMO application. In addition, the results are intended to serve as an evaluation tool and contribute to improvements in digital service quality and user experience in the future.

2. Research Methodology

The stages conducted to determine the sentiment of user comments on the JMO application on the Play Store using the BERT model include: (1) data collection, (2) data preprocessing, (3) data labeling, (4) BERT model training, and (5) model evaluation. This study adopts a descriptive quantitative approach with a computational experimental method to analyze user sentiment toward the Jamsostek Mobile (JMO) application on the Google Play Store. The

focus of the study is the application of the BERT model, particularly the IndoBERT variant, to classify sentiments into three categories: positive, negative, and neutral (Koto et al., 2020).

The research process consists of several main stages: data collection, preprocessing, labeling, training, and evaluation. Each stage is designed to follow the principles of Indonesian text analysis using deep learning approaches (Devlin et al., 2018; Rahmawati & Pratama, 2022).

Data Collection

The data used in this study consist of user reviews of the JMO application obtained from the Google Play Store. The data were collected using a web scraping technique with the Google Play Scraper library implemented in the Python programming language. The extracted information includes the username, review date, star rating, and the textual content of each comment.

Data Preprocessing

Comments collected from the Play Store are generally in raw text form and contain various irrelevant elements such as excessive punctuation, emoticons, numbers, hyperlinks, and stopwords. Therefore, a preprocessing step is required to prepare the text for model training.

The preprocessing stage focuses on handling the informal language that is dominant in JMO reviews, including the removal of non-linguistic symbols, normalization of non-standard words, and text tokenization using the built-in IndoBERT tokenizer to ensure consistency with the model's representation. Subsequently, during the data labeling stage, each comment is classified into three sentiment categories:

a. Positive: comments containing positive or favorable expressions
b. Negative: comments that include complaints, dissatisfaction, or negative experiences
c. Neutral: comments that are informative, provide suggestions, or do not express a specific emotional tone

BERT Model Implementation

The main model used in this study is BERT, specifically the IndoBERT variant, which has been pre-trained on Indonesian language corpora. The implementation stages include:

a. Pre-trained Model Loading
   The indobenchmark/indobert-base-p1 model from the Hugging Face Transformers library is used as the pre-trained model.
b. Text Tokenization
   The review texts are converted into token IDs according to the BERT input format using the WordPiece tokenizer.
c. Model Training (Fine-tuning)
   The pre-trained BERT model is fine-tuned on the JMO review dataset by adding a classification layer on top of the BERT architecture. The training parameters are as follows:
   - Batch size: 16
   - Learning rate: 2e-5

- Epochs: 2
- Optimizer: AdamW

d. Sentiment Prediction
After the training process is completed, the model is used to predict the sentiment of new user comments.

Model Evaluation

Model evaluation is conducted to measure how well the BERT model can correctly classify sentiments. The dataset is split using an 80:20 ratio to maintain a balance between training and testing, considering that the number of samples in each class has been equalized.

The evaluation metrics used in this study include:

a. Accuracy: the proportion of correct predictions out of the total number of samples
b. Precision: the degree to which the model correctly identifies a particular sentiment class
c. Recall (Sensitivity): the model's ability to detect all instances belonging to a specific sentiment class
d. F1-score: the harmonic mean of precision and recall

3. Results and Discussion

The study employed several Python functions during the experimental process, particularly in the preprocessing stage, including text lowercasing, duplicate removal, and word normalization. The experiments were conducted on a system equipped with an AMD Ryzen 7 5800H processor with Radeon Graphics (16 CPUs, approximately 3.2 GHz), 32 GB DDR4 RAM, running Windows 11 Pro. The programming language used was Python version 3.10.6.

The research dataset consists of 100,000 reviews of the JMO application from BPJS Ketenagakerjaan collected from the Google Play Store. The dataset underwent preprocessing steps such as lowercasing, normalization to remove non-standard words, and the removal of duplicate comments. After preprocessing, 49,901 reviews remained for labeling, from which 3,000 samples were selected for each sentiment category.

Table 1. Data Processing Results

Label	Data After Preprocessing	Data Used
Positive	30,184	3,000
Negative	16,160	3,000
Neutral	3,557	3,000

The dataset used in this study consists of 9,000 user comments, with each sentiment class containing 3,000 samples: positive (3,000), negative (3,000), and neutral (3,000). The data were split into training and testing sets using an 80:20 ratio, with the parameters test_size = 0.2 and random_state = 42 to ensure consistency in data partitioning. Two BERT-based approaches were evaluated in this study:

a. BERT without fine-tuning (baseline), in which the model uses only the pretrained BERT representations for classification without additional training on the application review dataset.
b. BERT with fine-tuning, in which the model is retrained using the review dataset to adjust the representation weights so that they become more specific to the target data domain.

Table 2. Evaluation Results of BERT without Fine-tuning

Label	Precision	Recall	F1-Score
Positive	0.55	0.56	0.55
Negative	0.48	0.51	0.49
Neutral	0.79	0.73	0.75
Accuracy			0.60

In the baseline model, an accuracy of 0.60 was achieved. The positive class showed the highest performance with a precision of 0.79 and an F1-score of 0.75, indicating that the model was more effective in identifying positive comments compared to the other two classes. However, the neutral class still exhibited relatively low performance (F1-score of 0.49), suggesting that the model had difficulty distinguishing neutral comments from positive or negative ones.

Table 3. Evaluation Results of BERT with Fine-tuning

Label	Precision	Recall	F1-Score
Positive	0.65	0.55	0.59
Negative	0.51	0.66	0.57
Neutral	0.87	0.74	0.80
Accuracy			0.65

The experimental results demonstrate that the BERT model has strong capability in handling sentiment analysis tasks for user reviews on the Google Play Store. However, the model's performance is highly dependent on how well it is optimized for the characteristics of the local dataset.

In the first stage, the baseline BERT model was used without retraining on the application review dataset. The results showed an accuracy of 0.60, with the highest F1-score observed in the positive class (0.75). These results indicate that the general representations learned by BERT from English corpora are still able to capture some sentiment patterns in Indonesian-language review data. However, the relatively low performance on the neutral class (F1-score of 0.49) highlights the model's limitations in recognizing ambiguous or implicit sentiment expressions.

After fine-tuning was performed using the same dataset, the accuracy increased to 0.65, with improvements in F1-scores across all classes, particularly in the neutral class (from 0.49 to

0.57) and the positive class (from 0.75 to 0.80). Fine-tuning had a significant impact on the model's ability to understand the specific linguistic context of Indonesian user comments.

Overall, the fine-tuning process made the model more sensitive to variations in informal language, abbreviations, and writing styles commonly found in Play Store reviews. This is consistent with the theory that a BERT model fine-tuned on domain-specific data can produce more accurate semantic representations (Devlin et al., 2018).

4. Conclusions

This study analyzed the effectiveness of the BERT model in performing sentiment analysis on Play Store application reviews using a labeled dataset of 9,000 samples categorized as positive, negative, and neutral. The model was evaluated under two conditions: without fine-tuning (baseline) and with fine-tuning using 80% of the data for training and 20% for testing. The results show that:

a. The BERT model without fine-tuning achieved an accuracy of 0.60, while the fine-tuned BERT model improved to 0.65.
b. Fine-tuning provided a noticeable performance improvement across all classes, particularly for sentiment categories that were previously difficult to identify.
c. The highest F1-score was achieved in the positive class (0.80) after fine-tuning, indicating the model's strong ability to consistently recognize positive sentiment.

It can be concluded that BERT with fine-tuning is more effective for sentiment analysis on application review data, primarily because it is able to adapt word representations and contextual understanding to the characteristics of local user language.

References

BPJS Ketenagakerjaan. (2022). Transformasi digital layanan BPJAMSOSTEK melalui aplikasi JMO. BPJAMSOSTEK.

BPJS Ketenagakerjaan. (2023). Laporan tahunan BPJS Ketenagakerjaan 2023. BPJS Ketenagakerjaan.

Chen, J., & Zhao, J. (2020). Enhancing sentiment analysis performance using pre-trained transformer models. Journal of Information Systems, 34(3), 251–265.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Fadlilah, S., & Rahardjo, D. (2020). Analisis sentimen aplikasi e-government dengan pendekatan deep learning. Jurnal Ilmu Komputer dan Informatika, 6(2), 85–94.

Google Play Console. (2024). User review and sentiment trends analytics. Google Developer Documentation.

Hidayatullah, A. F., & Kusumasari, R. A. (2021). Sentiment analysis of Google Play Store app reviews using bidirectional LSTM and Word2Vec. International Journal of Advanced Computer Science and Applications, 12(7), 221–228.

Koto, F., Lau, J. H., & Baldwin, T. (2020). IndoBERT and IndoBERT-Lite: Indonesian BERT models for sequence tagging and text classification. In Proceedings of PACLIC 34.

Lailiyah, N., & Rachmawati, R. (2021). Analisis sentimen ulasan aplikasi mobile banking menggunakan CNN dan LSTM. Jurnal Teknologi Informasi dan Komunikasi, 9(2), 114–121.

Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.

Oktaviani, M., & Wahyudi, A. (2022). Implementasi BERT untuk analisis sentimen ulasan aplikasi pemerintah di Play Store. Jurnal Sistem Informasi, 18(3), 295–306.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135.

Putri, D. R., Sari, M., & Hidayat, R. (2021). Analisis sentimen ulasan produk e-commerce menggunakan IndoBERT. Jurnal Informatika dan Komputer, 7(2), 145–154.

Rahmawati, N., & Pratama, A. (2022). Penerapan BERT untuk analisis sentimen komentar media sosial berbahasa Indonesia. Jurnal Teknologi Informasi, 9(1), 22–30.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune BERT for text classification? In Chinese Computational Linguistics (pp. 194–206).

Suryani, D., & Nugroho, D. (2023). Implementasi IndoBERT untuk klasifikasi sentimen berbahasa Indonesia pada komentar media sosial. Jurnal Teknologi Informasi, 12(1), 45–53.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017).