

Sentiment Analysis of Pospay Application Reviews Using The Bert Deep Learning Method

Erna Mulyati^{1*}, Muhammad Ibnu Choldun Rachmatullah²,
Adri Sapta Firmansyah³

^{1,2} Universitas Logistik dan Bisnis Internasional,
³ Universitas Pendidikan Indonesia
^{1,2}Jl. Sariosih 54, Bandung, Indonesia
³Jl. Dr. Setiabudhi No. 229, Bandung, Indonesia

ABSTRACT

Article:

Accepted: August 17, 2025
Revised: April 13, 2025
Issued: October 30, 2025

© Mulyati et al, (2025).



This is an open-access article
under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

*Correspondence Address:
ernamulyati@ulbi.ac.id

E-money usage in Indonesia has grown significantly due to increasing internet penetration and smartphone adoption. Digital transactions are becoming more common, with platforms like GoPay, OVO, and Dana leading the market. The government and financial institutions actively support this shift through regulations and initiatives. This study analyzes user sentiment on the Pospay application using the BERT deep learning method, based on 16,760 Google Play Store reviews. To the best of our knowledge, this is the first study to apply BERT for sentiment analysis of Pospay user reviews in Indonesia. The goal is to understand user perceptions and satisfaction. BERT helps capture subtle nuances in reviews, including informal expressions and abbreviations like "gk" for negative sentiment. The model achieves high accuracy, with precision scores of 0.82 (negative) and 0.93 (positive), and recall scores of 0.92 (negative) and 0.93 (positive). Findings suggest PT Pos should enhance application stability, security, transaction processing, and customer service. Regular updates are recommended to improve performance and user satisfaction.

Keywords : *sentiment analysis; e-money application; pospay, machine learning; pos ind.*

1. INTRODUCTION

The development of electronic money (e-money) usage in Indonesia has shown significant growth in recent years. Supported by the increasing penetration of the internet and the use of smartphones, people are increasingly shifting to more practical and efficient digital transactions. The government and various financial institutions are also actively encouraging the use of e-money through various initiatives and regulations that support the digital payment ecosystem. Platforms like GoPay, OVO, and Dana have become an essential part of daily life, enabling quick and easy payments for various needs, from transportation and shopping to bill payments.

According to Bank Indonesia, the total value of electronic money transactions in Indonesia reached IDR 835.84 trillion in 2023, marking a 43.45% year-on-year increase. This significant growth reflects the rapid adoption of digital payment methods across the country.

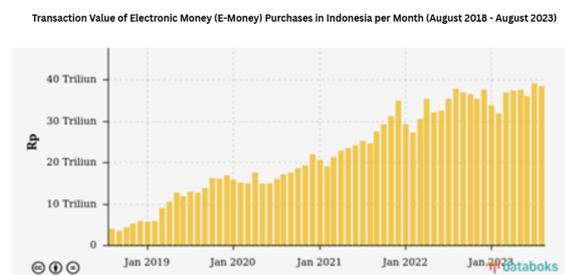


Figure 1. Trend of electronic money usage in Indonesia (source : <https://databoks.katadata.co.id>)

The use of e-wallet applications in Indonesia has increased along with changes in people's transaction behavior. Supported by advancements in information technology, internet access, and the widespread use of smartphones, people are shifting from conventional to digital payment methods. Based on download numbers, e-wallets such as DANA (100 million+), OVO (50 million+), and GoPay, ShopeePay, and LinkAja (10 million+) dominate the market. Meanwhile, Pospay has been downloaded over 5 million times, ranking 7th, alongside apps like Octo Mobile.

Position	Application Name	Total Downloads	Standing in App Store
1	DANA	100 Million++	2
2	OVO	50 Million++	12
3	GoPay	10 Million++	3
4	LinkAja	10 Million++	55
5	ShopeePay	10 Million++	1
6	Octo Mobile	5 Million++	72
7	Pospay	5 Million++	131
8	i.saku	1 Million++	110
9	DOKU	1 Million++	175
10	Sakuku	1 Million++	115

Figure 2. Leaderboard of e-money applications from download total by Indonesian people & Pospay total download (source : <https://rankia.id/> & <https://play.google.com/>)

As a result, e-wallet applications are now not only used for daily transactions such as shopping and bill payments, but also for various other services such as investments and ticket purchases, making them an integral part of the digital lifestyle of the Indonesian people.

PT. Pos Ind is a state-owned company engaged in postal and logistics services, with a long history since 1746 and a vast network covering the entire region of Indonesia. Besides traditional postal services such as mail and package delivery, PT. Pos Ind offers integrated logistics services, financial services such as bill payments and money transfers, as well as retail services and partnerships with the government and corporations for delivery and logistics solutions. One of the innovations created by the company is an application called Pospay, an e-wallet application that facilitates digital financial transactions. Pospay enables bill payments, money transfers, digital product purchases, QR code transactions, and various other financial services, demonstrating PT. Pos Ind adaptation to technological advancements and modern consumer needs.

Pospay was chosen as the research object because, although it is one of the digital payment applications managed by PT. Pos Ind, a company with a vast network and a large user base in Indonesia, based on image 1.2, this application still cannot occupy the top e-wallet positions like GoPay, OVO, Dana, and other e-wallets. Although Pospay is an e-wallet application similar to other e-wallets in general, unlike similar applications, Pospay offers the integration of postal services with financial services, making it unique and different from other payment applications that generally only focus on financial transactions. This research is important to identify and address the challenges

faced by the Pospay application, as well as to provide in-depth insights into the acceptance and use of this service integration by users, in order to enhance Pospay's competitiveness in the e-wallet market.

This research introduces novelty by applying the BERT model for sentiment analysis in evaluating the performance of the Pospay application, an approach that has not been widely explored. Previous studies have demonstrated BERT's superiority over traditional models like SVM in sentiment analysis. For instance, research by UIN Sunan Ampel found that IndoBERT achieved 78% accuracy in sentiment classification, significantly outperforming SVM, which only reached 65% (Putri & Ramadhani, 2023). Similarly, a study from UPN Veteran Jawa Timur confirmed that BERT's bidirectional contextual understanding enhances sentiment classification accuracy (Sujana Yudianto 2024)

By leveraging BERT's advanced capabilities, this study aims to provide a deeper understanding of user sentiments towards Pospay, evaluate its performance based on sentiment analysis, and offer strategic recommendations for improvement. The insights from sentiment analysis can help PT Pos identify user pain points, address common issues, and enhance service quality. For instance, frequent complaints about transaction failures or app performance can guide technical improvements, while positive feedback on promotional offers can shape more effective marketing strategies. By integrating these insights into its strategic decision-making, PT Pos can optimize Pospay's services, improve customer experience, and strengthen its competitive position in Indonesia's digital payment market.

There are several theoretical studies that serve as references as well as supporting theories for this research, including the following :

1.1. Sentiment Analysis

Sentiment analysis is the process of understanding and classifying opinions expressed in text, particularly in the context of social media, product reviews, and customer service. This analysis is important for companies because it allows them to understand customer feelings toward their products or

services, which can then be used to improve customer experience and marketing strategies (Liu, 2012).

1.2. Customer Satisfaction Theory

Customer satisfaction is a measure of how well a product or service meets or exceeds customer expectations. Factors influencing customer satisfaction include product quality, customer service, price, and overall experience. The relationship between customer satisfaction and online reviews is closely linked, where positive reviews can enhance satisfaction among other customers and vice versa (Kotler et al., 2016).

In line with this, Titing et al. (2023) highlighted that the completeness of features and quality of information significantly impact user satisfaction in the Pospay application, emphasizing how well-designed features directly influence perceived satisfaction. Similarly, (Hayati and Mahmudah 2016) analyzed Gronroos's service quality dimensions and found that these dimensions strongly correlate with customer satisfaction in the context of Pospay services. Furthermore, (Halimah et al. 2004) demonstrated that perceived usefulness plays a crucial role in shaping user satisfaction and drives electronic word-of-mouth (e-WOM), which in turn can enhance repurchase intention among Pospay users.

1.3. Application Reviews and Their Role in E-commerce

Application reviews play a crucial role in e-commerce by providing valuable feedback for developers and potential users. These reviews can influence purchasing decisions and overall perceptions of products or services. Studies have shown that better reviews tend to increase sales and customer loyalty (Chevalier & Mayzlin, 2006).

In the context of service applications, (Pujiati et al. 2024) found that differentiation of Pospay kiosk services significantly affects user decisions to adopt the application, illustrating how user feedback and unique service features are intertwined. Although not directly related to e-commerce, (Setiawan and Ratri 2022) explored user experience in educational applications and showed how design and usability aspects reflected in user reviews shape

users' adoption decisions. These findings reinforce the role of application reviews as a key component in influencing user perceptions and engagement, similar to how reviews function in the e-commerce ecosystem. Additionally, (Halimah et al. 2024) emphasized that positive user perceptions, often expressed through reviews, are pivotal in fostering e-WOM, further strengthening the impact of reviews in digital service adoption.

1.4. Traditional Sentiment Analysis Methods

Traditional sentiment analysis methods include lexicon-based methods and machine learning methods. Lexicon-based methods use sentiment word dictionaries to classify text, while machine learning methods involve training models on labeled data to make predictions. Both methods have their own advantages and disadvantages (Pang & Lee, 2008).

1.5. Deep Learning

Deep learning is a subfield of machine learning that uses artificial neural networks with many layers (deep neural networks) to model complex data. In the context of natural language processing (NLP), deep learning has proven to be highly effective for tasks such as sentiment analysis, machine translation, and speech recognition (LeCun et al., 2015).

1.6. BERT Model (Bidirectional Encoder Representations from Transformers)

BERT is a deep learning model developed by Google that has revolutionized natural language processing. BERT uses a Transformer architecture to process text bidirectionally, allowing the model to understand context from both directions in a sentence. This makes BERT highly effective for various NLP tasks, including sentiment analysis (Devlin et al., 2018)

1.7. Application of BERT in Sentiment Analysis

BERT has been widely used in sentiment analysis with highly satisfactory results. In many studies, BERT has shown a significant improvement in accuracy compared to other models such as LSTM and GRU. These studies indicate that BERT's ability to understand

context more effectively aids in classifying sentiment more accurately (Sun et al., 2019)

1.8. Sentiment Analysis of E-commerce Application Reviews

Sentiment analysis methods for e-commerce application reviews involve collecting review data, preprocessing the data, and applying machine learning or deep learning models to classify sentiment. Numerous studies have been conducted to analyze the sentiment of e-commerce application reviews, demonstrating various techniques and models that can be used to improve the accuracy and efficiency of the analysis.

Furthermore, recent research emphasizes the use of transformer-based models like BERT, which have shown superior performance in understanding context and nuances in user reviews. These advancements are particularly relevant in the context of e-commerce applications, where the volume and diversity of reviews require robust analytical approaches.

In the Indonesian context, (Hayati and Mahmudah 2016) examined service quality dimensions of Pospay services using customer feedback, providing early insights into user sentiments towards digital postal services. Although their approach was qualitative, the study highlights the importance of systematically analyzing user opinions to gauge satisfaction and service improvement areas.

To further strengthen sentiment analysis applications, (Imron et al, 2023) investigated aspect-based sentiment analysis on Indonesian marketplace product reviews using a combination of BERT, LSTM, and CNN models. Their findings demonstrated that incorporating transformer-based contextual understanding with deep learning architectures significantly improved sentiment classification accuracy compared to conventional methods. This study underscores the relevance of BERT in effectively analyzing local user reviews by capturing nuanced sentiments specific to different product aspects, thereby enhancing insights for service and product development.

1.9. Sentiment Analysis of the Pospay Application

According to (Amri & Andarsyah, 2024), sentiment analysis of Pospay application reviews on Google PlayStore was conducted using the Support Vector Machine (SVM) method. This study collected user review data through web scraping techniques and then performed preprocessing to clean and prepare the data. The SVM model was trained using the processed dataset, achieving an accuracy of 88.1% for positive sentiment classification. This research provides valuable insights for the Pospay SuperApp developers in understanding user sentiment and improving the quality and satisfaction of users. (Putri, 2023) explains that sentiment analysis is a text mining process that classifies unstructured data to efficiently generate sentiment information using data mining algorithms, specifically Support Vector Machine (SVM). This study aims to determine whether user sentiment towards the Pospay application on Google Play Store is generally positive or negative and to evaluate the performance of the SVM algorithm in this analysis. Putri also emphasizes the importance of text data preprocessing before applying text mining methods and algorithms to improve the accuracy of sentiment analysis results.

1.10. Implementation of BERT Model for Sentiment Analysis of Pospay Application Reviews

According (Devlin et al., 2018), BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model that uses a transformer architecture to understand bidirectional context in text. BERT has demonstrated superior performance in various natural language processing (NLP) tasks, including sentiment analysis, by learning deep representations of text through training on large datasets.

In the context of sentiment analysis of Pospay application reviews, BERT offers several significant advantages over traditional methods. With its ability to understand context from both directions in a text, BERT can provide more accurate interpretations of the sentiment expressed in user reviews. This is particularly important given that user reviews often contain unstructured language and complex emotional expressions.

(Amri & Andarsyah, 2024) explain that the Support Vector Machine (SVM) method has been used for sentiment analysis of Pospay application reviews with satisfactory results. However, BERT can offer further improvements in accuracy and precision of sentiment classification by leveraging its more advanced transformer capabilities. BERT not only recognizes key words in reviews but also understands how the context and sequence of those words affect overall sentiment.

(Putri, 2023) also shows that sentiment analysis of Pospay application reviews using traditional algorithms like SVM yields good results but can still be enhanced by using more advanced models like BERT. Utilizing BERT in sentiment analysis can help capture finer and more complex nuances of sentiment that often arise in user reviews.

Overall, implementing BERT in sentiment analysis of Pospay application reviews can provide deeper and more accurate insights into user perceptions and satisfaction. This can assist application developers in better understanding user needs and issues, as well as formulating more effective strategies to enhance user experience.

2. METHODS

This research was conducted on April 24, 2024, on user reviews of the Pospay application on Google Play Store using the Python programming language through the Google Colaboratory platform. This study employs the Knowledge Discovery in Database (KDD) method, which is one of the various methods used in research, particularly related to data mining. KDD involves five main stages: (1) Data Selection, (2) Data Preprocessing, (3) Data Transformation, (4) Data Mining, and (5) Evaluation. Below is the research workflow that will be followed:

All these stages refer to the implementation of KDD (Knowledge Discovery in Databases) using the BERT model, conducted with the Python programming language through the Google Colaboratory platform, with the following details for each stage:

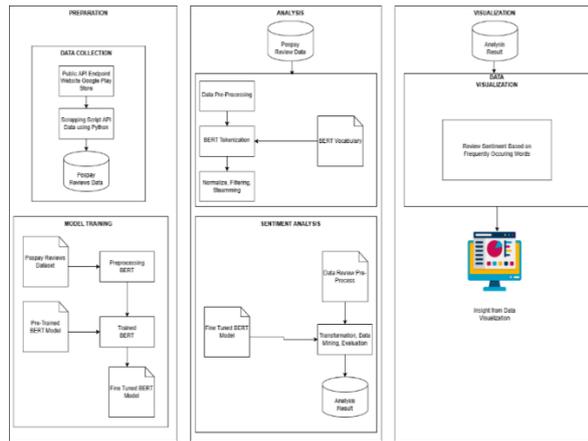


Figure 3. Overview of the research

2.1. Data Selection

At this stage, data is collected using the data scraping method. Scraping techniques are used to gather user reviews from various sources, such as Google Play Store and Apple App Store. This process involves programming specific scripts that automatically extract important information from review pages, including review text, ratings, publication dates, and user identities (if available). The collected data is then stored in a structured format for further preprocessing, which enables comprehensive sentiment analysis and evaluation of the application's performance.

In total, 16,759 user reviews were collected for analysis. Based on the rating scores, sentiment labels were assigned using the following rule:

1. Ratings ≤ 2 are labeled as Negative.
2. Ratings ≥ 3 are labeled as Positive.

The Neutral category was not used explicitly; reviews with a score of 3 were grouped into the Positive class to simplify classification. As a result, the final labeled dataset consists of:

1. 6,848 Negative reviews.
2. 9,911 Positive reviews.

One limitation of the dataset is the imbalance in sentiment distribution. Since sentiment labels are assigned based on user ratings (e.g., scores ≤ 2 as negative, score 3 as neutral, and scores ≥ 4 as positive), the neutral class tends to be underrepresented. This imbalance may lead to a bias in model predictions, making it difficult to accurately classify neutral sentiments.

To mitigate this bias, several techniques can be applied:

- a. **Class balancing:** Implementing oversampling or undersampling techniques to ensure a more even distribution of sentiment classes.
- b. **Weight adjustment:** Assigning higher weights to underrepresented classes during training to prevent bias toward the dominant classes.
- c. **Alternative labeling approaches:** Using more advanced heuristics or human-annotated datasets to refine sentiment classification beyond numerical ratings.

By addressing these limitations, the model can achieve better generalization and more accurate sentiment classification across all sentiment categories.

2.2. Pre-Processing Data

In this stage, text processing is performed on the content attribute (reviews), which includes:

- a. **Case Folding:** This involves converting all uppercase letters to lowercase.
- b. **Cleaning Text:** This stage uses the regex library to remove punctuation symbols, numeric digits, correct consecutive character duplication, and excessive whitespace.
- c. **Tokenize:** Tokenization is carried out using BERT's tokenizer on the Pospay application review content. This process transforms the review text into a series of tokens that can be interpreted by the BERT model, including breaking down the text into words or sub-words and adding special tokens such as [CLS] and [SEP] to mark the beginning and end of each review. The results of tokenization are then used as input for the BERT model for sentiment analysis and a deeper understanding of user perceptions of the Pospay application.
- d. **Normalize:** Normalizing words that do not conform to Indonesian spelling.
- e. **Filtering:** This process selects meaningful words and removes those without significance.
- f. **Stemming:** This process reduces inflected words to their root form.

2.3. Data Transformation

The tokenization process is carried out using BERT's tokenizer with the transformers library. Before applying BERT tokenization, the data is cleaned of missing values to enhance model performance. This tokenization process converts review text into tokens that match the BERT input format, adds special tokens [CLS] and [SEP], and adjusts text length with padding or truncation. The vector representations generated from this tokenization are then ready for training and inference with the BERT model, enabling in-depth sentiment analysis of Pospay application reviews.

2.4. Data Mining

In the data mining stage, the fine-tuned BERT-base-uncased model is applied to analyze Pospay application reviews. This model is used to extract key features from the review text and classify it based on sentiment—positive, negative, or neutral. The model is tested using a data split scenario as follows:

Training Set: Approximately 80% of the total data.

Validation Data: 20% of the total data.

To achieve optimal model performance, the BERT model is fine-tuned using the following hyperparameters:

Learning rate: Default value in Trainer (explicitly stated if modified).

Batch size: 4 for training, 8 for evaluation.

Number of epochs: 1.

Optimizer: AdamW (default in Trainer).

Warmup steps: 500.

Weight decay: 0.01.

These hyperparameter settings are designed to ensure stable training while preventing overfitting.

2.5. Model Evaluation

In this stage, the results from the model will be evaluated using a Confusion Matrix with the help of libraries such as Sklearn, Matplotlib, and Seaborn to identify the model with the best performance. The model with the highest performance will be exported in a specialized format.

2.6. Data Visualization

A Bar Plot will be used to visualize the data by filtering and displaying the top classes of frequently occurring words associated with negative and positive sentiments.

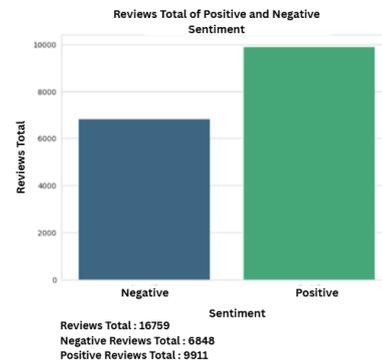


Figure 4. Data visualization of positive and negative sentiment counts

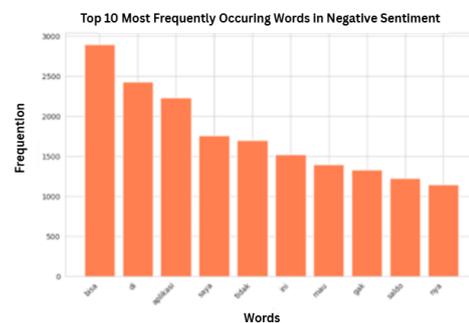


Figure 5. Visualization of negative sentiment based on frequently occurring words

3. RESULTS AND DISCUSSION

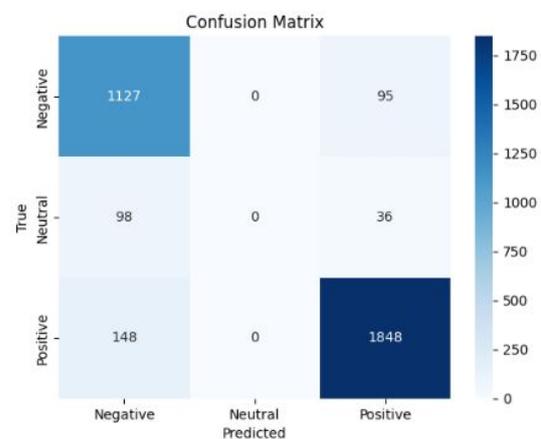


Figure 6. Results of the BERT model displayed through the confusion matrix

The results from the confusion matrix show that the BERT model performs exceptionally well in sentiment classification

for positive and negative categories. The model accurately predicted 1,127 negative reviews and 1,848 positive reviews, demonstrating a high level of accuracy in identifying these sentiments. This success reflects BERT's ability to capture nuanced language patterns and understand context within user reviews.

However, the model encountered a significant challenge in classifying neutral sentiments. The evaluation results indicate that no neutral reviews were correctly predicted, resulting in an F1-score of 0.00 for the neutral class. This outcome highlights a critical limitation in the model's performance when dealing with neutral sentiment detection.

Several factors contribute to this failure:

- a. **Class Imbalance:** The dataset used for model training is heavily skewed towards positive (9,911 reviews) and negative (6,848 reviews) sentiments, while the neutral category is either underrepresented or merged with positive labels due to labeling simplifications. This imbalance leads the model to focus more on distinguishing between positive and negative sentiments, neglecting the neutral class.
- b. **Ambiguity in Neutral Labels:** Reviews that are intended to be neutral often contain language that can be interpreted as mildly positive or negative. Without clear distinguishing features, the model struggles to differentiate neutral sentiments from other categories.
- c. **Labeling Methodology:** The sentiment labeling in this study is based on rating scores only, without additional semantic analysis. Reviews with a score of 3 were grouped into the positive class, further reducing the distinct representation of neutral sentiments. This approach limits the model's exposure to explicitly neutral data during training.

Despite these challenges, the model achieved an overall accuracy of 89%, with strong performance metrics in other areas:

1. Precision: 0.82 (negative), 0.93 (positive)
2. Recall: 0.92 (negative), 0.93 (positive)
3. F1-score: 0.87 (negative), 0.93 (positive)

The macro average for precision, recall, and F1-score were lower at 0.58, 0.62, and 0.60 respectively, primarily due to the poor performance on neutral classification. In contrast, the weighted average remains high

(precision: 0.86, recall: 0.89, F1-score: 0.87), reflecting the model's effectiveness in handling the dominant sentiment classes.

	precision	recall	f1-score	support
Negative	0.82	0.92	0.87	1222
Neutral	0.00	0.00	0.00	134
Positive	0.93	0.93	0.93	1996
accuracy			0.89	3352
macro avg	0.58	0.62	0.60	3352
weighted avg	0.86	0.89	0.87	3352

Figure 7. Evaluation results of the BERT model predictions

The evaluation results of the BERT model demonstrate excellent performance in sentiment classification with an overall accuracy of 89%. The model exhibits a precision of 0.82 for negative sentiment and 0.93 for positive sentiment, indicating that most negative and positive predictions are correct. The recall for negative sentiment is 0.92 and for positive sentiment is 0.93, showing that the model is capable of identifying most truly negative and positive reviews.

The F1-score, which reflects the balance between precision and recall, is 0.87 for negative sentiment and 0.93 for positive sentiment. However, an issue arises with neutral sentiment classification, as the model failed to correctly identify any neutral reviews, leading to an F1-score of 0.00 for the neutral category. This suggests that the model struggles to differentiate between neutral and other sentiment classes.

The macro average for precision, recall, and F1-score are 0.58, 0.62, and 0.60, respectively, while the weighted average shows higher values with precision of 0.86, recall of 0.89, and F1-score of 0.87. Overall, the BERT model demonstrates outstanding performance in classifying negative and positive sentiments, providing accurate and reliable results in sentiment analysis.

To address the limitations in neutral sentiment classification, several improvements are recommended for future studies:

- a. **Data Augmentation for Neutral Class:** Increasing the quantity and diversity of neutral sentiment data through augmentation techniques can help balance the dataset. Synthetic generation of neutral reviews or semi-supervised labeling methods may provide richer training data for this class.
- b. **Multi-Label Classification Approach:** Instead of forcing reviews into discrete

positive, negative, or neutral categories, adopting a multi-label classification framework allows the model to recognize mixed or overlapping sentiments within a single review, improving flexibility and accuracy.

- c. **Refined Sentiment Annotation:** Incorporating manual labeling or using advanced sentiment lexicons to distinguish neutral expressions more precisely could enhance model training. Leveraging context-aware annotation guidelines will ensure clearer separation between sentiment classes.
- d. **Class Weighting and Sampling Strategies:** Applying class weighting adjustments during model training or employing resampling techniques (such as SMOTE) can help mitigate the impact of class imbalance and improve the model's sensitivity to minority classes.

By implementing these strategies, future research can achieve a more balanced and accurate sentiment classification, particularly in identifying neutral user feedback.

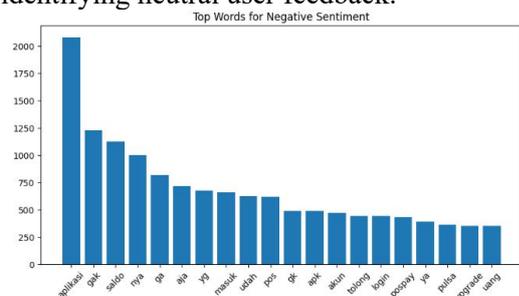


Figure 8. Top 10 words by frequency from negative sentiments

The top words from negative sentiments provide valuable insights into the aspects most frequently criticized by users. The word "aplikasi" appears most often, indicating that many users are discussing issues related to the application in general. Words like "gak" and "ga" also frequently appear, reflecting user dissatisfaction. Specific problems commonly mentioned include "saldo," "masuk," "akun," and "login," suggesting that many users are experiencing difficulties with account access and management. Users also frequently mention "pos" and "pospay," signaling dissatisfaction that may be related to the core services offered.

Additionally, words like "tolong" and "upgrade" indicate requests for help and

improvements, suggesting that users are hoping for enhancements in the application's quality and functionality. Words like "gak," "apk," and "udah" express general disappointment, while "pulsa" and "uang" show that financial transactions are often a source of issues. The relatively high use of words like "ya" and "yg" suggests that many comments may contain specific suggestions or criticisms. Overall, this analysis shows that technical issues and transaction problems are major sources of user dissatisfaction, and improvements in these areas could significantly enhance user perceptions of the application.

CONCLUSION

The use of the BERT model in sentiment analysis shows exceptional performance with an overall accuracy of 89%. The model is capable of classifying negative and positive sentiments with high precision and recall, achieving 0.82 and 0.93 for precision, and 0.92 and 0.93 for recall, respectively. The balanced F1-score, which reflects the equilibrium between precision and recall, further demonstrates that the model is reliable in identifying sentiments in user reviews. With this strong performance, the BERT model proves effective in capturing and understanding sentiment patterns in text, making it highly valuable for in-depth and accurate sentiment analysis.

The negative sentiment analysis reveals that users frequently complain about key issues related to balance, access, accounts, and login problems with the Pospay application, as well as financial transaction issues such as credits and money. To address these negative sentiments, PT. Pos could focus on improving the stability and security of the application, enhancing the transaction system, boosting customer service, and performing regular updates to improve application performance. By concentrating on these improvements, PT. Pos can enhance the user experience and reduce negative sentiments towards the Pospay application, thereby increasing user satisfaction and loyalty.

However, the model exhibits limitations in identifying neutral sentiments, as indicated by the low F1-score for the neutral class. This suggests a need for additional fine-tuning, such as adjusting class weights or incorporating

alternative embeddings to better capture neutral sentiment expressions. Additionally, handling ambiguous statements and sarcasm remains a challenge. Future improvements could include leveraging attention heatmaps to visualize critical words influencing sentiment classification.

To further enhance sentiment analysis, future studies could explore comparative evaluations with other NLP models, such as LSTM or GRU, to validate BERT's effectiveness in this domain. Additionally, implementing a real-time sentiment monitoring API could provide PT. Pos with continuous insights into user feedback, allowing for proactive issue resolution and service improvements.

REFERENCES

- [1] Amri, Y., & Andarsyah, R. (2024). *Jurnal Teknik Informatika, Vol. 16, No. 2, April 2024. 16(2), 1–7.*
- [2] Chevalier, J., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research, 43*. <https://doi.org/10.1509/jmkr.43.3.345>
- [3] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.*
- [4] Kotler, P., Keller, Manceau, & Hemonnet-Goujot, A. (2016). Marketing Management 15th. In *Décisions Marketing (Vol. 83).*
- [5] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature, 521, 436–444.* <https://doi.org/10.1038/nature14539>.
- [6] Liu, B. (2012). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition, May, 1–432.* <https://doi.org/10.1017/9781108639286>
- [7] Mrityunjay, S., Jakhar, A., & Pandey, S. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Mining, 11.* <https://doi.org/10.1007/s13278-021-00737-z>
- [8] Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval, 2, 1–135.* <https://doi.org/10.1561/1500000011>
- [9] Putri, A. (2023). Sentiment Analysis of Mobile Provider Application Reviews Using Naive Bayes Algorithm and Support Vector Machine. *Sinkron, 8(2), 824–835.* <https://doi.org/10.33395/sinkron.v8i2.13469>
- [10] Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1, 380–385.*
- [11] Zhao, X., & Sun, Y. (2022). Amazon Fine Food Reviews with BERT Model. *Procedia Computer Science, 208, 401–406.* <https://doi.org/10.1016/j.procs.2022.10.056>
- [12] Putri, A., & Ramadhani, A. (2023). Sentiment Analysis of Mobile Payment Applications Using Deep Learning. *Jurnal Teknologi dan Informasi, 15(2), 120–135.* <https://journal.ummat.ac.id/index.php/fkip/article/download/25651/>
- [13] Y. Sujana, “A Comparative Study of Machine Learning Models for Sentiment Analysis of Dana App Reviews,” *IJIE (Indonesian J. Informatics Educ., vol. 7, no. 2, 2024, doi: 10.20961/ijie.v7i2.93132.Bibliography}*
- [14] Bank Indonesia. (2024). *Monetary Policy Report - Quarter IV 2023.* Retrieved from <https://www.bi.go.id/en/publikasi/laporan/Pages/Laporan-Kebijakan-Moneter-Triwulan-IV-2023.aspx>
- [15] Halimah, H. S., Hurriyati, R., & Hendryati, H. (2024). E-Wom Strategy Through Perceived Usefulness to Increase Repurchase Intention of Pospay PT PI. *Proceedings of the 8th Global*

- Conference on Business, Management, and Entrepreneurship (GCBME 2023)*, 485–486. [https://doi.org/10.2991/978-94-6463-443-3_62:contentReference\[oaicite:0\]{index=0}](https://doi.org/10.2991/978-94-6463-443-3_62:contentReference[oaicite:0]{index=0})
- [16] Setiawan, A., & Ratri, D. (2022). Comparison of Mooc Application User Experience in Indonesia (Case Studies: Ruang Guru and Zenius). *Proceedings of the 2nd ITB Graduate School Conference (IGSC)*, 1–8. <https://seminar.itb.ac.id/igsc/>
- [17] Titing, A. S., Stiadi, M., & Kumalasari, F. (2023). The Effect of Completeness of Features and Quality of Information on User Satisfaction of the Pospay Application in Kolaka Regency. *Jurnal Ekonomi*, 12(2), 1168–1171. [http://ejournal.seaninstitute.or.id/index.php/Ekonomi/article/view/1753:contentReference\[oaicite:2\]{index=2}](http://ejournal.seaninstitute.or.id/index.php/Ekonomi/article/view/1753:contentReference[oaicite:2]{index=2})
- [18] Hayati, N., & Mahmudah, A. S. (2016). Analysis of Gronroos Service Quality Dimensions for Pospay Service and Mail Delivery on Customer Satisfaction (Study in Indonesia). *International Journal of Business Management and Economic Research*, 7(4), 717–725. [http://www.ijbmer.com/docs/volumes/vol7issue4/ijbmer2016070403.pdf:contentReference\[oaicite:0\]{index=0}](http://www.ijbmer.com/docs/volumes/vol7issue4/ijbmer2016070403.pdf:contentReference[oaicite:0]{index=0})
- [19] Pujiati, B., Gunawan, I., Huriyati, R., & Sultan, M. A. (2024). Effect of Differentiation of Pospay Kiosk Services on the Decision to Use an Application as a Consumer Service (Case Study: PT Pos Indonesia (Persero)). *Proceedings of the 8th Global Conference on Business, Management, and Entrepreneurship (GCBME 2023)*, 726–729. [https://doi.org/10.2991/978-94-6463-443-3_95:contentReference\[oaicite:1\]{index=1}](https://doi.org/10.2991/978-94-6463-443-3_95:contentReference[oaicite:1]{index=1})
- [20] Imron, S., Setiawan, E. I., Santoso, J., & Purnomo, M. H. (2023). Aspect Based Sentiment Analysis Marketplace Product Reviews Using BERT, LSTM, and CNN. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(3), 4751. <https://jurnal.iaii.or.id/index.php/RESTI/article/view/4751>