Moestopo International Review on Societies, Humanities, and Sciences (MIRSHuS)
Vol. 6, No. 1 (2026) pp. 68-80
ISSN: 2775-9601

Transformer-Based Sentiment Analysis of DOKU E-Wallet User Reviews

Wenny TY. Situmorang*
Universitas Bina Nusantara, Jakarta, Indonesia
Tuga Mauritsius
Universitas Bina Nusantara, Jakarta, Indonesia

*Correspondence: wenny.situmorang@binus.ac.id

ARTICLE INFO

Article History:
received: 26/01/2026
revised: 15/02/2026
accepted: 30/03/2026

Keywords:
Sentiment Analysis; Digital Wallet; User Reviews; Natural Language Processing; Transformer Models

DOI:
10.32509/mirshus.v6i1.174

ABSTRACT

The rapid advancement of digital payment technologies has accelerated the widespread adoption of mobile wallet applications, making it increasingly important for service providers to understand user perceptions and experiences. User reviews published on mobile application platforms represent valuable sources of feedback that reflect satisfaction, complaints, and expectations regarding service performance. However, the large volume of textual reviews makes manual analysis inefficient and difficult to manage. This study aims to analyze user sentiment toward the DOKU e-wallet application by applying transformer-based natural language processing techniques. A total of 11,685 user reviews collected from mobile application platforms were analyzed using two transformer-based models. The analytical process followed a structured data mining approach, including data collection, preprocessing, model training, and evaluation using accuracy, precision, recall, and F1-score metrics. The results show that the IndoBERT model achieved an accuracy of 93.1%, while the GPT-3.5 Turbo model achieved 93.2%, indicating strong performance in sentiment classification tasks. In addition, the analysis identified several recurring issues reported by users, including account access problems, verification difficulties, transaction errors, and customer service responsiveness. This study contributes to the literature by providing a comparative evaluation of transformer-based models in the context of digital payment platforms, particularly within the Indonesian ecosystem.

INTRODUCTION

Rapid technological advancement and the acceleration of digital transformation have significantly reshaped how financial transactions are conducted in modern society. One of the most prominent innovations resulting from this transformation is the emergence of digital payment systems, widely known as electronic wallets or e-wallets. An e-wallet is a digital financial service that enables users to store monetary value electronically and perform various financial transactions through mobile devices. Through mobile wallet applications, users can conveniently conduct activities such as purchasing goods or services, transferring funds, paying bills, and managing financial transactions efficiently using smartphones (El Azzouzy et al., 2025). Digital wallets function as virtual payment instruments that securely store financial information and facilitate transactions using technologies such as QR codes, near-field communication (NFC), and mobile applications, thereby enabling faster, more efficient, and convenient cashless payment experiences (Shalihah et al., 2023).

Among the various digital wallet services available in Indonesia, the DOKU e-wallet has emerged as one of the platforms providing integrated digital financial transaction services. Developed by PT Nusa Inti Satu Artha, a company operating in the electronic payment and risk management industry, the DOKU e-wallet offers a wide range of transaction features, including purchasing mobile credits or data packages, paying monthly bills, conducting QRIS-based payments, and transferring balances either to fellow users or to bank accounts. As a mobile-based platform, the application is accessible through major distribution channels such as the Google Play Store and the Apple App Store. Based on available platform statistics, the application has been downloaded more than one million times and has accumulated thousands of user reviews, reflecting active user engagement in providing feedback related to application performance and service quality. The increasing adoption of mobile wallet applications reflects the rapid evolution of digital payment technologies that enable users to conduct financial transactions in a more efficient, secure, and convenient manner through mobile devices (El Azzouzy et al., 2025). Furthermore, digital wallet platforms contribute to the expansion of cashless financial ecosystems by integrating advanced payment technologies that enhance transaction efficiency and security (Chellappan et al., 2025). From a theoretical perspective, the adoption and continued use of digital wallet applications can be explained through frameworks such as the Technology Acceptance Model (TAM) and user experience theory, which emphasize that perceived usefulness, perceived ease of use, and service quality are critical determinants of user satisfaction and behavioral intention. These dimensions are often reflected in user-generated reviews, making such data a valuable source for evaluating the performance and quality of digital financial services.

To maintain service quality and strengthen user satisfaction, digital payment service providers must continuously monitor and evaluate user experiences. One effective way to understand user perceptions is through the analysis of user-generated content, particularly online reviews published on digital platforms. User reviews often contain valuable information related to satisfaction levels, complaints, expectations, and overall experiences when interacting with a digital service. This is further reinforced by previous research indicating that the utilization of information technology significantly influences service effectiveness and user satisfaction in digital environments (Pujiastuti, 2025). Consequently, analyzing such feedback can provide strategic insights that help organizations improve service performance, optimize system functionality, and develop more user-centered innovations. Previous studies indicate that user perceptions and experiences significantly influence the adoption and continued use of mobile payment platforms, as users tend to evaluate usefulness, convenience, and security based on their interaction with the system (Khasawneh & AlBahsh, 2024). Similarly, research on digital wallet usage shows that user satisfaction, perceived risks, and user responses reflected in feedback are important indicators that help organizations understand consumer behavior and enhance digital financial services (Ajina et al., 2023).

Sentiment analysis, a subfield of Natural Language Processing (NLP), has become a widely used approach for automatically identifying and classifying opinions expressed in textual data. By analyzing linguistic patterns, contextual expressions, and semantic relationships between words, sentiment analysis enables textual information to be categorized into positive, negative, or neutral sentiments (Mao et al., 2024). In recent years, sentiment analysis has been increasingly applied to evaluate public perceptions of products, services, and organizations across multiple domains, including digital finance, social media, and e-commerce environments (Ferdous et al., 2024). Within the context of digital payment services, understanding user sentiment toward e-wallet applications is particularly important because it allows service providers to identify user expectations, detect potential service issues, and develop strategies to improve customer experience. Techniques that analyze user-generated content such as application reviews enable organizations to extract valuable insights related to customer satisfaction, service quality, and product perception from large volumes of textual data (Ounacer et al., 2023).

Despite the growing number of studies exploring sentiment analysis across various digital platforms, several important gaps remain insufficiently addressed. In particular, limited research has focused on analyzing user sentiment toward digital wallet applications using advanced transformer-based language models, especially within the Indonesian digital payment context. This limitation is critical given the rapid expansion of mobile payment adoption in Indonesia, where user-generated reviews represent a valuable yet underexplored source of insights into user experience and service quality. Furthermore, the increasing volume of user reviews available on application platforms such as the Google Play Store and the App Store presents significant analytical challenges, as conventional approaches often struggle to efficiently process and interpret large-scale and context-rich textual data. Although recent developments in transformer-based language models, including Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT), have significantly improved the performance of sentiment analysis tasks by enabling models to capture contextual relationships within textual data more effectively (Li et al., 2023), there is still a lack of comparative empirical studies that evaluate their effectiveness specifically in the domain of digital wallet applications. Transformer-based architectures utilize attention mechanisms to understand semantic relationships between words within a sentence, allowing sentiment classification models to achieve higher accuracy and deeper contextual understanding compared with traditional machine learning approaches (Areshey & Mathkour, 2024). Therefore, a systematic comparative analysis of these models is necessary to better understand their performance and applicability in analyzing user sentiment within digital financial services.

Therefore, this study aims to analyze user sentiment toward the DOKU e-wallet application by utilizing transformer-based language models. Specifically, this research compares the performance of the IndoBERT model and the GPT-3.5 Turbo model in classifying sentiments expressed in user reviews collected from the Google Play Store and the App Store. Through this comparative approach, the study seeks to evaluate the effectiveness of both models in capturing user sentiment and interpreting user-generated feedback within the context of digital financial services.

METHOD

This study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) as a structured methodological framework for conducting sentiment analysis. CRISP-DM provides a systematic approach that guides data mining and machine learning processes from problem identification to model evaluation, ensuring transparency and reproducibility (Ma'aly et al., 2024; Shameti et al., 2026). In this study, the framework is operationalized into four main stages: data collection, data preparation, modeling, and evaluation. This approach enables the transformation of large-scale user-generated textual data into meaningful insights related to user perceptions of digital financial services.

The dataset consists of 11,685 user reviews of the DOKU e-wallet application collected from the Google Play Store and the Apple App Store between 2023 and March 2024. Data were obtained using automated web scraping techniques implemented in Python, utilizing publicly available APIs and scraping libraries to extract review text, timestamps, and user ratings. The data collection process applied several filtering criteria, including the removal of duplicate entries, non-relevant content, and non-Indonesian language reviews to ensure data quality and consistency. The collected textual data were then preprocessed through data cleaning, case folding, and slang word normalization to standardize informal expressions commonly found in user reviews (Chen et al., 2022; Mao et al., 2024). Furthermore, the dataset was manually labeled into three sentiment categories positive, negative, and neutral by two independent annotators based on predefined annotation guidelines. To ensure labeling reliability, the annotation results were cross validated, achieving a substantial level of agreement between annotators.

The modeling stage employs two transformer-based language models, namely IndoBERT and GPT-3.5 Turbo, due to their ability to capture contextual relationships within textual data (Areshey & Mathkour, 2024). The dataset was divided into training (70%), validation (20%), and testing (10%) sets to ensure robust model evaluation. The IndoBERT model was fine-tuned using the Adam optimizer with a learning rate of 2e-5, a batch size of 16, and training conducted for five epochs. The training process was implemented in a Python-based environment using deep learning libraries and executed on a GPU-enabled platform. In parallel, the GPT-3.5 Turbo model was fine-tuned using OpenAI's API with parameters including three training epochs, a batch size of 3, and a learning rate multiplier of 0.3, allowing the model to learn contextual sentiment patterns from large-scale textual data (Nasiopoulos et al., 2025).

Model performance was evaluated using a confusion matrix to compare predicted and actual sentiment labels. Based on this matrix, evaluation metrics including accuracy, precision, recall, and F1-score were calculated to provide a comprehensive assessment of classification performance. These metrics are widely used in sentiment analysis to measure model effectiveness in handling imbalanced and context-dependent textual data (Ma'aly et al., 2024). Through this evaluation framework, the study ensures that the performance of both models is systematically assessed and comparable in analyzing user sentiment within digital financial platforms.

RESULT AND DISCUSSION

This section presents the findings of the sentiment analysis conducted on user reviews of the DOKU e-wallet application collected from the Google Play Store and the Apple App

Store. The analysis was carried out using two transformer-based language models, namely IndoBERT and GPT-3.5 Turbo, following the analytical workflow defined by the CRISP-DM framework. The objective of this analysis is to evaluate the performance of both models in classifying user sentiment and to identify recurring issues experienced by users based on their reviews.

Data Collection

Apple App Store. These reviews represent user experiences, opinions, and evaluations regarding the functionality and performance of the DOKU e-wallet application. Data collection was performed using web scraping techniques to retrieve publicly available reviews posted between 2023 and March 2024.

User-generated reviews provide valuable insights into user perceptions and experiences when interacting with digital financial services. These reviews reflect various aspects of application usability, system reliability, and service quality. An example of the raw data obtained during the crawling process is presented in **Figure 1**, which illustrates the structure of user-generated textual feedback collected from the application platforms.

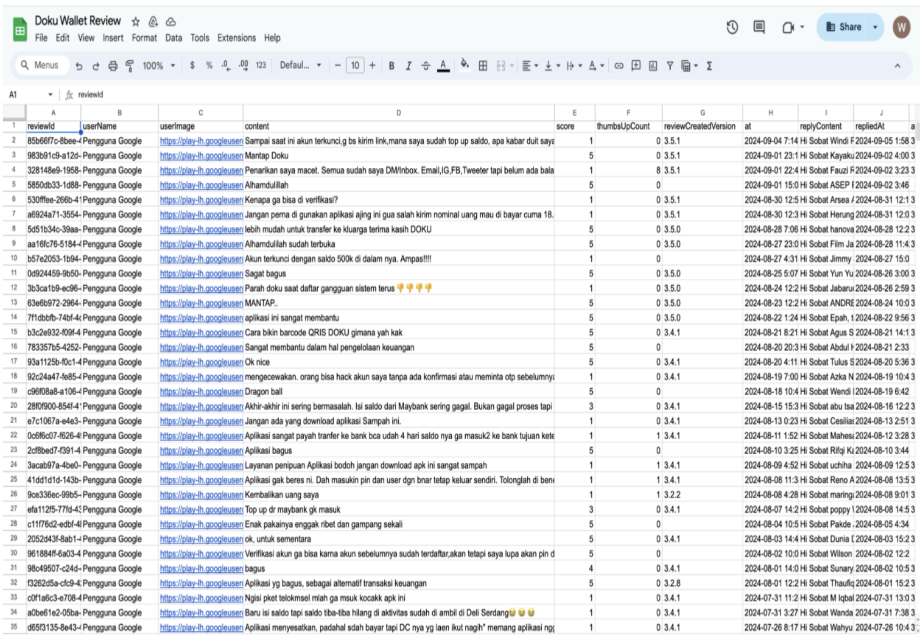

Figure 1. Example of Crawled Review Data from Application Platforms

The collected reviews were then prepared for further analysis through several preprocessing procedures aimed at improving the quality and consistency of the dataset.

Data Preprocessing

The preprocessing stage was conducted to transform raw textual data into a structured format suitable for machine learning analysis. Reviews collected from digital platforms often contain informal expressions, abbreviations, typographical errors, and inconsistent writing patterns. Therefore, several preprocessing techniques were applied to improve the quality of the dataset prior to model training.

The first step involved data labeling, in which each review was manually categorized into one of three sentiment classes: positive, negative, or neutral. To ensure the reliability of the annotation process, the labeling procedure was validated by two independent reviewers.

The second preprocessing step consisted of data cleansing and case folding, which involved removing punctuation marks, numbers, and irrelevant symbols while converting all text into lowercase characters to ensure consistency across the dataset.

The third step involved slang word normalization, where informal expressions commonly used in user reviews were converted into standardized language forms. This process was implemented using a custom-built slang dictionary developed during the dataset exploration stage. The dictionary was implemented in Python and contains pairs of slang expressions and their standardized equivalents.

Through these preprocessing procedures, the dataset became cleaner, more consistent, and more suitable for the subsequent modeling stage.

Modeling

The modeling stage involved training transformer-based language models to classify user sentiments based on the processed review dataset. Two models were implemented in this study: IndoBERT and GPT-3.5 Turbo.

The dataset consisting of 11,685 reviews was divided into three subsets:

1. 70% training data (8,179 reviews)
2. 20% validation data (2,349 reviews)
3. 10% testing data (1,157 reviews)

This partitioning strategy was applied to ensure reliable model training and evaluation.

1. Fine-Tuning IndoBERT

   In the first experiment, the IndoBERT model was fine-tuned using the labeled review dataset. The model used in this study was the IndoBERT-base-p1 architecture, which consists of 12 transformer encoder layers with 12 attention heads in each layer.

   The fine-tuning process was conducted for five training epochs to adapt the pre-trained model to the sentiment classification task. The training performance across the five epochs is presented in Table 1.

Table 1. Training Performance of the IndoBERT Model

Epoch	Train Loss	Validation Loss	Accuracy	F1 Score	Recall	Precision
1	0.3921	0.2383	87%	0.64	0.63	0.66
2	0.2191	0.2215	93%	0.81	0.79	0.84
3	0.1861	0.2209	93%	0.80	0.79	0.83
4	0.1630	0.2260	93%	0.80	0.79	0.81
5	0.1428	0.2228	93%	0.80	0.79	0.83

   Based on the results presented in Table 1, the best model performance was achieved at epoch 2, which produced the lowest validation loss (0.2215) and the highest F1-score (0.81). Therefore, the model obtained at epoch 2 was selected as the optimal model for further evaluation.

Moestopo International Review on Societies, Humanities, and Sciences (MIRSHuS)
Vol. 6, No. 1 (2026) pp. 68-80
ISSN: 2775-9601

2. Fine-Tuning GPT-3.5 Turbo

The second experiment utilized the GPT-3.5 Turbo model, a transformer-based language model developed by OpenAI. The labeled dataset was converted into JSONL format, which is required for OpenAI fine-tuning pipelines.

The fine-tuning process used the GPT-3.5-turbo-0125 model with the following parameters:

- Epochs: 3
- Batch size: 3
- Learning rate multiplier: 0.3
- Random seed: 1685996073

During training, the model processed approximately 1,819,476 tokens. The fine-tuning workflow of the GPT model is illustrated in Figure 2, which describes the model training process using the prepared dataset.

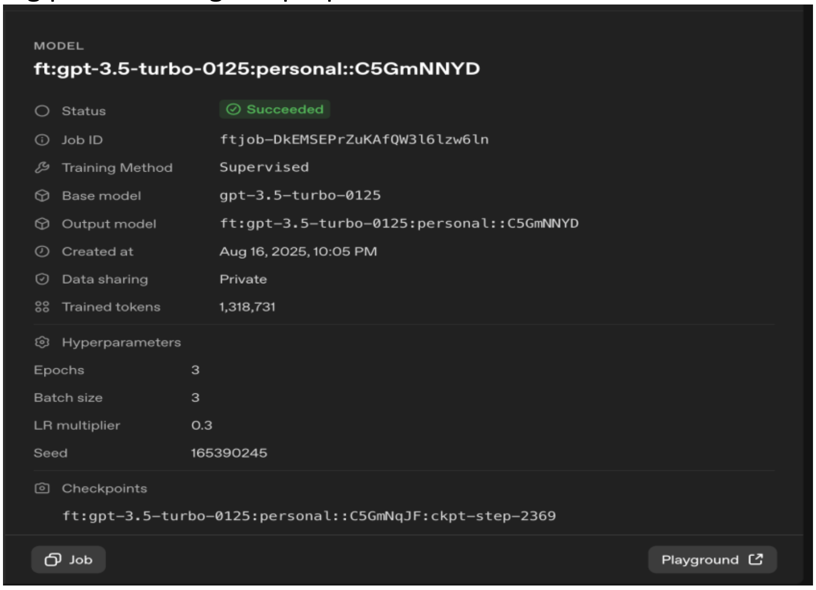

Figure 2. Fine-Tuning Process of the GPT-3.5 Turbo Model

The model training process generated a loss curve that reflects how the model gradually improved during training. As shown in **Figure 3**, the loss value initially started at a relatively high level (around 3.5) and rapidly decreased during the early training stage before stabilizing between **0.2 and 0.5**. This trend indicates that the model progressively learned meaningful patterns from the dataset and improved its prediction performance during training.

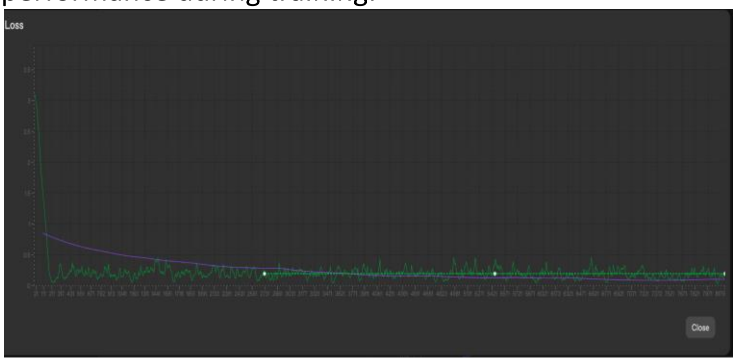

Figure 3. Training Loss During GPT-3.5 Turbo Training

Model Evaluation

Model performance was evaluated using a confusion matrix, which compares predicted sentiment labels with the actual labels to assess classification accuracy and identify error patterns. This evaluation approach provides a comprehensive understanding of how each model performs across different sentiment categories.

1. Evaluation of IndoBERT

   The classification performance of the IndoBERT model is illustrated in *Figure 4*, which presents the confusion matrix representing the distribution of predicted sentiment classes against the actual labels. As shown in *Figure 4*, the model successfully classified 971 positive reviews, with only a small number of instances misclassified as neutral or negative. In addition, the model demonstrated strong capability in identifying negative sentiment, correctly classifying 1,159 reviews. These findings indicate that IndoBERT is highly effective in capturing dominant sentiment patterns, particularly when user opinions are expressed clearly.

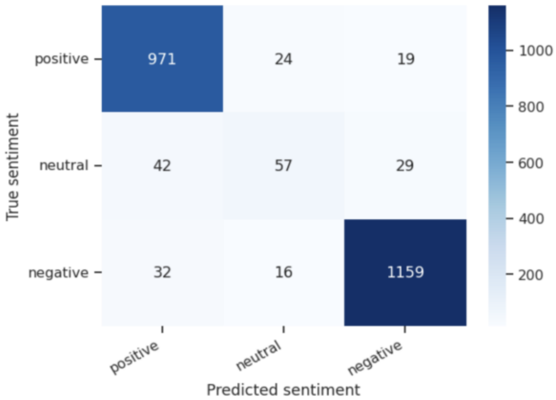

Figure 4. Confusion Matrix of the IndoBERT Model

However, the model exhibited relatively lower performance in classifying neutral sentiment. As observed in *Figure 4*, only 57 neutral reviews were correctly identified, while a substantial number of instances were misclassified as either positive or negative. This limitation can be attributed to the inherent ambiguity of neutral expressions, which often contain overlapping contextual cues and less explicit emotional polarity. Consequently, distinguishing neutral sentiment remains a challenging task for the model.

The overall evaluation metrics of the IndoBERT model are summarized in *Figure 5*, which presents the classification report. As shown in *Figure 5*, the model achieved precision values of 0.93 for positive sentiment, 0.59 for neutral sentiment, and 0.96 for negative sentiment. These results further confirm that IndoBERT performs consistently well in dominant sentiment classes while highlighting its limitations in handling neutral sentiment.

Moestopo International Review on Societies, Humanities, and Sciences (MIRSHuS)
Vol. 6, No. 1 (2026) pp. 68-80
ISSN: 2775-9601

	precision	recall	f1-score	support
positive	0.93	0.96	0.94	1014
neutral	0.59	0.45	0.51	128
negative	0.96	0.96	0.96	1207
accuracy			0.93	2349
macro avg	0.83	0.79	0.80	2349
weighted avg	0.93	0.93	0.93	2349

Figure 5. Classification Report of the IndoBERT Model

2. Evaluation of GPT-3.5 Turbo

The classification performance of the GPT-3.5 Turbo model is presented in *Figure 6*, which illustrates the confusion matrix of predicted and actual sentiment labels. As depicted in *Figure 6*, the model successfully classified 988 positive reviews with minimal misclassification, demonstrating its strong capability in identifying positive sentiment patterns. Similar to IndoBERT, the model also achieved high accuracy in classifying negative sentiment, indicating its effectiveness in capturing clearly expressed opinions.

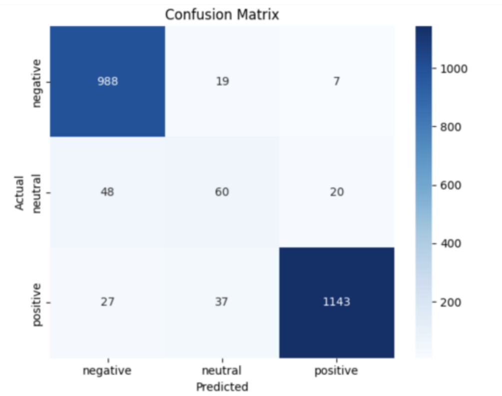

Figure 6. *Confusion Matrix of the GPT-3.5 Turbo Model*

Nevertheless, the GPT-based model exhibited similar challenges in identifying neutral sentiment. As shown in *Figure 6*, the number of correctly classified neutral reviews remains relatively low, suggesting that contextual ambiguity in user-generated text affects the model's ability to distinguish neutral expressions from other sentiment categories. This finding indicates that even advanced transformer-based models encounter limitations when dealing with subtle or mixed sentiment signals.

=== Classification Report ===

	precision	recall	f1-score	support
negative	0.93	0.97	0.95	1014
neutral	0.52	0.47	0.49	128
positive	0.98	0.95	0.96	1207
accuracy			0.93	2349
macro avg	0.81	0.80	0.80	2349
weighted avg	0.93	0.93	0.93	2349

Figure 7. Classification Report of the GPT-3.5 Turbo Model

The classification report presented in *Figure 7* provides a summary of the quantitative evaluation results. As illustrated in *Figure 7*, the GPT-3.5 Turbo model achieved precision values of 0.93 for positive sentiment, 0.59 for neutral sentiment, and 0.96 for negative sentiment. These results demonstrate that the model delivers consistent and reliable performance across dominant sentiment categories, while also reflecting similar limitations observed in the IndoBERT model for neutral sentiment classification.

Overall, both models demonstrate strong capability in sentiment classification, particularly for positive and negative categories. However, the consistent difficulty in accurately identifying neutral sentiment across both models highlights an important challenge in sentiment analysis tasks involving user-generated content, where contextual ambiguity and mixed expressions are prevalent.

Discussion and Topic Analysis

Based on the sentiment classification results, both models demonstrated strong capability in identifying user sentiment toward the DOKU e-wallet application. The IndoBERT model achieved an accuracy of 93.2%, while the GPT-3.5 Turbo model achieved an accuracy of 93.1%, indicating comparable performance in sentiment prediction. Despite this similarity, the results suggest that IndoBERT, as a language model specifically trained on Indonesian text, is slightly more effective in capturing linguistic nuances and contextual expressions within local user reviews. In contrast, GPT-3.5 Turbo demonstrates strong generalization ability due to its large-scale training, allowing it to perform consistently across diverse textual inputs. This indicates that while both models are highly effective, the choice of model may depend on the specific linguistic and contextual characteristics of the dataset.

Beyond model performance, the sentiment analysis revealed several recurring issues frequently mentioned in negative user reviews. The most dominant topics include account access problems, customer service responsiveness, OTP and PIN verification issues, and account recovery difficulties. Among these, account access and verification-related issues appear to be the most frequently reported, indicating that authentication processes remain a critical pain point for users. In addition, complaints related to transaction failures and balance discrepancies suggest concerns regarding system reliability and financial security. These patterns highlight that technical stability and efficient customer support are key determinants of user satisfaction in digital financial services. This finding is consistent with previous studies emphasizing that digital service systems must continuously adapt and innovate in response to user feedback and criticism, particularly in rapidly evolving digital environments where user expectations are highly dynamic (Syukri, 2025).

From a practical perspective, these findings provide important implications for fintech service providers. First, improving authentication mechanisms, such as OTP and account recovery processes, can significantly enhance user trust and reduce user frustration. Second, strengthening customer service responsiveness is essential to address user complaints in a timely manner, particularly for issues related to financial transactions. Third, ensuring system reliability and minimizing transaction errors are critical to maintaining user confidence in digital payment platforms. By addressing these issues, service providers can improve overall user experience and foster long-term customer loyalty. This perspective aligns with prior research suggesting that user trust and perception are significantly influenced by corporate strategies, including branding and social responsibility initiatives, which play a critical role in sustaining user engagement in digital services (Kusuma, 2024).

Furthermore, the findings of this study are consistent with prior research that highlights the effectiveness of sentiment analysis in extracting actionable insights from user-generated content. (Rahman & Maryani, 2024) demonstrated that sentiment analysis can support organizations in enhancing service quality by systematically identifying customer concerns and feedback patterns. In addition, recent studies emphasize that transformer-based approaches significantly improve the capability of sentiment analysis models in capturing contextual and semantic relationships within textual data (Duru & Sunar, 2025). In line with these findings, the present study confirms that sentiment analysis serves as a valuable analytical tool for interpreting user feedback and understanding user experiences. Moreover, this research extends existing literature by providing a comparative evaluation of transformer-based models within the context of digital wallet applications, particularly in the Indonesian digital payment ecosystem, thereby offering more context-specific empirical insights into the application of advanced NLP techniques in financial technology services.

Overall, the integration of sentiment analysis with transformer-based models not only enables accurate sentiment classification but also facilitates a deeper understanding of user experiences and service challenges. These insights can support organizations in developing data-driven strategies to enhance service quality, improve user satisfaction, and strengthen trust in digital financial platforms.

CONCLUSION

This study examined user sentiment toward the DOKU e-wallet application by applying transformer-based language models to analyze user reviews collected from major mobile application platforms. The findings indicate that both IndoBERT and GPT-3.5 Turbo achieved strong performance in sentiment classification tasks, with accuracy values of 93.1% and 93.2%, respectively. These results demonstrate the effectiveness of transformer-based models in capturing contextual patterns within user-generated textual data and accurately identifying sentiment polarity. In addition to model performance, the analysis revealed recurring user concerns, particularly related to account access, OTP verification, customer service responsiveness, and transaction reliability, highlighting critical aspects that influence user experience in digital financial services.

From a scientific perspective, this study contributes to the existing literature by providing a comparative evaluation of transformer-based models in the context of digital wallet applications, particularly within the Indonesian digital payment ecosystem. The findings extend prior research by demonstrating that both domain-specific models, such as IndoBERT, and general large-scale models, such as GPT-3.5 Turbo, can achieve comparable performance in sentiment analysis tasks, while offering different strengths in handling linguistic and contextual variations. From a practical standpoint, the results offer valuable insights for fintech service providers, emphasizing the importance of improving authentication processes, enhancing customer service responsiveness, and ensuring system reliability to increase user satisfaction and trust. By leveraging sentiment analysis, organizations can adopt a data-driven approach to continuously monitor user feedback and optimize service quality.

Despite these contributions, this study has several limitations that should be acknowledged. The analysis relies on user reviews collected from specific application platforms within a limited time frame, which may not fully represent all user experiences. In addition, the relatively lower performance in classifying neutral sentiment highlights challenges associated with ambiguous textual expressions. Future research is encouraged to incorporate larger and more diverse datasets, explore additional transformer-based architectures, and develop real-time sentiment analysis systems to provide more comprehensive and dynamic insights into user behavior and digital service performance.

REFERENCE

Ajina, A. S., Joudeh, J. M. M., Ali, N. N., Zamil, A. M., & Hashem, T. N. (2023). The effect of mobile-wallet service dimensions on customer satisfaction and loyalty: An empirical study. *Cogent Business & Management*, *10*(2), 2229544. https://doi.org/10.1080/23311975.2023.2229544

Areshey, A., & Mathkour, H. (2024). Exploring transformer models for sentiment classification: A comparison of BERT , RoBERTa , ALBERT , DistilBERT , and XLNet. *Expert Systems*, *41*(11), e13701. https://doi.org/10.1111/exsy.13701

Chellappan, K., Elanchselvan, T., Department of Electrical, Electronics & Systems Engineering, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, Malaysia, Abu-Samah, A., & Department of Electrical, Electronics & Systems Engineering, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, Malaysia. (2025). E-wallet Delivery Technology Architecture Adoption: A Review. *Jurnal Kejuruteraan*, *37*(1), 219–232. https://doi.org/10.17576/jkukm-2025-37(1)-14

Chen, M., Ubul, K., Xu, X., Aysa, A., & Muhammat, M. (2022). Connecting Text Classification with Image Classification: A New Preprocessing Method for Implicit Sentiment Text Classification. *Sensors*, *22*(5), 1899. https://doi.org/10.3390/s22051899

Duru, I., & Sunar, A. S. (2025). Transformer and Pre-Transformer Model-Based Sentiment Prediction with Various Embeddings: A Case Study on Amazon Reviews. *Entropy*, *27*(12), 1202. https://doi.org/10.3390/e27121202

El Azzouzy, O., Chanyour, T., & Andaloussi, S. J. (2025). Transformer-based models for sentiment analysis of YouTube video comments. *Scientific African*, *29*, e02836. https://doi.org/10.1016/j.sciaf.2025.e02836

Ferdous, S. M., Newaz, S. N. E., Mugdha, S. B. S., & Uddin, M. (2024). Sentiment Analysis in the Transformative Era of Machine Learning: A Comprehensive Review. *Statistics, Optimization & Information Computing*, *13*(1), 331–346. https://doi.org/10.19139/soic-2310-5070-2113

Khasawneh, O., & AlBahsh, R. (2024). Why do people use a mobile wallet? The case of fintech companies in Jordan. *Investment Management and Financial Innovations*, *21*(2), 89–102. https://doi.org/10.21511/imfi.21(2).2024.07

Kusuma, E., Barus, I., & Pranawukir, I. (2024). The Strategy of Unilever’s Corporate Social Responsibility 6.0 in Building Corporate Personal Branding. Moestopo International Review on Social, Humanities, and Sciences, 4(1), 55-66. https://doi.org/10.32509/mirshus.v4i1.68

Li, Z., Yang, C., & Huang, C. (2023). A Comparative Sentiment Analysis of Airline Customer Reviews Using Bidirectional Encoder Representations from Transformers (BERT) and Its Variants. *Mathematics*, *12*(1), 53. https://doi.org/10.3390/math12010053

Ma’aly, A. N., Pramesti, D., Fathurahman, A. D., & Fakhrurroja, H. (2024). Exploring Sentiment Analysis for the Indonesian Presidential Election Through Online Reviews Using Multi-

Label Classification with a Deep Learning Algorithm. *Information*, *15*(11), 705. https://doi.org/10.3390/info15110705

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, *36*(4), 102048. https://doi.org/10.1016/j.jksuci.2024.102048

Nasiopoulos, D. K., Roumeliotis, K. I., Sakas, D. P., Toudas, K., & Reklitis, P. (2025). Financial Sentiment Analysis and Classification: A Comparative Study of Fine-Tuned Deep Learning Models. *International Journal of Financial Studies*, *13*(2), 75. https://doi.org/10.3390/ijfs13020075

Ounacer, S., Mhamdi, D., Ardchir, S., Daif, A., & Azzouazi, M. (2023). Customer Sentiment Analysis in Hotel Reviews Through Natural Language Processing Techniques. *International Journal of Advanced Computer Science and Applications*, *14*(1). https://doi.org/10.14569/IJACSA.2023.0140162

Pujiastuti, N. R., Triwinata, J., & Maryani, D. (2025). Information Technology Utilization and Its Impact on Service Effectiveness in West Jakarta. Moestopo International Review on Social, Humanities, and Sciences, 5(2), 205-216. https://doi.org/10.32509/mirshus.v5i2.132

Rahman, B. & Maryani. (2024). Optimizing Customer Satisfaction Through Sentiment Analysis: A BERT-Based Machine Learning Approach to Extract Insights. *IEEE Access*, *12*, 151476–151489. https://doi.org/10.1109/ACCESS.2024.3478835

Shalihah, A. H., Ruldeviyani, Y., Saraswati, K., & Ar Rasyiid, A. (2023). An Empirical Investigation of Factors Affecting Personal Information’s Disclosure on Mobile Payment (E-Wallet) Platforms. *International Journal on Advanced Science, Engineering and Information Technology*, *13*(3), 1031–1040. https://doi.org/10.18517/ijaseit.13.3.18054

Shameti, K., Hyseni, D., Manolopoulos, Y., & Çiço, B. (2026). Comparison of CRISP - DM Versus OSEMN Methodologies Using Linear Regression and Statistical Analysis. *Expert Systems*, *43*(2), e70185. https://doi.org/10.1111/exsy.70185

Syukri, Taufiqurokhman, Satispi, E., Andriansyah, & Muhammad Gusti. (2025). Criticism, Innovation, And Adaptation of Governance: From Classic to New Public and Digital Amid Global Challenges. Moestopo International Review on Social, Humanities, and Sciences, 5(2), 337-344. https://doi.org/10.32509/mirshus.v5i2.140