Jurnal Ekonomi Kreatif dan
Manajemen Bisnis Digital
https://transpublika.co.id/ojs/index.php/JEKOMBITAL
Online ISSN 2961-8428

https://doi.org/10.55047/jekombital.v4i1.1011

Original Article

Sentiment Classification of Livin' by Mandiri Reviews in Indonesia Using LSTM for Digital Banking Service Improvement

Sigit Mulyanto[1*], Dwika Lovitasari Yonia[2], Kheylina Lidya Situmorang[3], Bambang Sutejo[4]

[1,4]Department of Management, Faculty of Business, Universitas Darwan Ali, Sampit, Indonesia
[2]Department of Information Systems, Faculty of Computer Science, Universitas Darwan Ali, Sampit, Indonesia
[3]Department of Informatics, Universitas Methodist Indonesia, Medan, Indonesia
Email: [1)] sigitmul@gmail.com, [2)] dwika.lovitasari.yonia@gmail.com, [3)] kheylina172000@gmail.com, [4)] tejosampit@gmail.com

Received : 22 July - 2025
Accepted : 20 August - 2025
Published online : 26 August - 2025

Abstract

The rapid expansion of digital banking services in Indonesia has increased the need for continuous monitoring of user satisfaction, particularly through feedback submitted via app reviews. This study analyses user sentiment toward the Livin' by Mandiri mobile banking application using a deep learning approach. A total of 5,000 user reviews were collected exclusively from the Google Play Store and pre-processed through text cleaning steps such as slang normalization, stemming, and tokenization. Sentiment labels (positive, neutral, negative) were assigned using an Indonesian lexicon-based method, and a Long Short-Term Memory (LSTM) model was trained and evaluated with accuracy, precision, recall, and F1-score metrics. Results indicate negative sentiment dominates (37.6%), with frequent complaints about login failures and slow performance, while the LSTM model achieved 98% accuracy. This study is limited by its single-platform data source, potential linguistic bias in Indonesian user reviews, and the model's limitations in detecting sarcasm or complex emotions. Nonetheless, the findings demonstrate the applicability of sentiment analysis as a real-time monitoring tool to support feature enhancement and user experience improvements in Indonesian mobile banking services.

Keywords: Deep Learning, LSTM, Mobile Banking, Sentiment Analysis, User Reviews.

1. Introduction

The digital transformation of Indonesia's financial sector has significantly accelerated since the COVID-19 pandemic, which reshaped consumer behaviour and increased demand for safe, efficient, and contactless financial services. Digital banking has become a dominant channel, as evidenced by Bank Indonesia's report showing a 46.53% year-on-year increase in digital banking transaction value and a 41.35% rise in electronic money usage by February 2022 (Tanika et al., 2022). According to Bank Indonesia (2023), mobile banking transactions have experienced double-digit growth annually, driven by increasing smartphone penetration and evolving consumer behavior toward cashless payments. This trend is supported by national digital infrastructure and strategic policy frameworks, particularly the Indonesia Payment System Blueprint 2025, which aims to strengthen digital payment ecosystems and promote financial inclusion. Empirical studies further confirm that the pandemic accelerated public trust and reliance on mobile banking platforms, driven by ease of access, perceived security, and usefulness (Sebayang et al., 2024). Banks that swiftly implemented digital strategies, such as mobile-first services and online onboarding, demonstrated higher resilience and customer engagement during the pandemic-induced shift (Farida et al., 2023; Mulyanto et al., 2025). Moreover, research shows that institutions with strong digital offerings were more likely to sustain customer retention and operational stability in crisis periods (Hidayat & Kassim, 2023).

Livin' by Mandiri has emerged as one of the most downloaded financial applications in Indonesia, competing closely with BRImo (by BRI) and BCA Mobile. This dominance not only illustrates the success of Bank Mandiri's digital transformation strategy but also reflects the growing expectations of Indonesian consumers for seamless, secure, and responsive digital banking experiences (Putri et al., 2025). Among various digital banking platforms, Livin' by Mandiri has emerged as one of the leading applications, serving millions of users nationwide. Despite its popularity, user ratings and reviews on digital distribution platforms reveal recurring complaints related to login failures, application crashes, and slow performance—issues that, if unresolved, can erode trust and customer loyalty. As the adoption of digital banking continues to rise, user-generated reviews on platforms like Google Play Store and App Store serve as an essential extension of customer feedback, functioning as a modern "voice of the customer" (Adiningtyas & Auliani, 2024). Consumer reviews serve as a vital source of user feedback that reflects customer satisfaction and directly influences product development and service quality decisions (Purba & Safrin, 2024). Moreover, sentiment analysis has been shown to uncover valuable insights for strategic decision-making, such as identifying feature-specific complaints or positive trends that can inform service design and customer retention initiatives (Sebayang et al., 2023, 2024). Taken together, the rise of Livin' by Mandiri, the growing role of sentiment analysis, and regulatory support from Bank Indonesia highlight the convergence of consumer behaviour, technology, and policy in shaping Indonesia's mobile banking landscape.

Previous studies on sentiment analysis in Indonesian financial technology contexts have largely focused on e-wallets, digital payment gateways, or generalized banking services (Bimantara & Zufria, 2024). While these studies demonstrate the utility of sentiment analysis for capturing user perceptions, they rarely address mobile banking applications within the Indonesian banking sector using domain-specific datasets. Furthermore, research applying deep learning models such as Long Short-Term Memory (LSTM) to Indonesian-language banking reviews remains scarce, despite LSTM's proven effectiveness in handling sequential linguistic patterns (Yonia et al., 2024). For instance, (Adiningtyas & Auliani, 2024) analysed Google Play reviews of BCA Mobile, Livin', and BNI Mobile and found that issues such as login errors, service downtime, and data privacy concerns frequently appeared in negative reviews, signalling key service quality gaps. Similarly, (Lubis, 2024) explored sentiment in Islamic banking apps and revealed that trust, interface functionality, and service responsiveness were dominant drivers of user perception. These findings align with the growing consensus that public sentiment expressed through app reviews functions as a real-time, unsolicited feedback mechanism that can support customer-centric innovation.

A significant research gap exists in the application of sentiment analysis to Indonesian mobile banking, despite its widespread use in domains such as e-commerce, digital wallets, and social media. Most prior studies have focused on general satisfaction metrics or technical performance, without specifically addressing the classification of user sentiment using deep learning models adapted for financial reviews in Bahasa Indonesia. This gap is twofold: practically, there is an urgent need for automated, real-time, and domain-specific sentiment monitoring to capture the unique linguistic and service characteristics of Indonesian mobile banking users; theoretically, there is an opportunity to advance the literature on user behavior, digital banking adoption, and natural language processing (NLP) within Indonesia's linguistic and cultural context. To address this, the present study constructs a domain-specific lexicon-labeled dataset from 5,000 Livin' by Mandiri user reviews sourced from the Google Play Store and trains a Long Short-Term Memory (LSTM) model for sentiment classification. This approach enables rigorous evaluation of model performance while generating actionable insights for enhancing service quality and user experience in Indonesia's digital banking sector.

The contributions of this study are as follows. First, development of a new sentiment dataset consisting of curated user reviews from the Livin' by Mandiri mobile application, which can serve as a valuable resource for future research in financial sentiment analysis (Adiningtyas & Auliani, 2024). Second, implementation of an automatic labelling procedure using an Indonesian lexicon-based approach, allowing for scalable and reproducible dataset annotation without the need for extensive manual labelling (Lubis, 2024). Third, evaluation of the LSTM model's performance in classifying Indonesian-language sentiment data from the financial domain (Gowandi et al., 2021)

This study focuses on analysing user sentiment toward the Livin' by Mandiri mobile banking application using a deep learning approach. A curated dataset and lexicon-based labelling method are employed to support sentiment classification in the Indonesian language context. The structure of this paper includes an introduction and related work, research methodology, results and discussion, and conclusions.

2. Methods

This chapter describes the research methods used to analyse user sentiment about the Livin' application, employing a quantitative exploratory approach that emphasizes reproducibility. The workflow follows the CRISP-DM framework adapted for natural language processing and consists of six sequential stages business understanding, data understanding, data preparation, modelling, evaluation, and deployment as shown in Figure 1. Each stage is explained in detail to align with the problem statement and research objectives and to ensure transparency and replicability of the entire process.

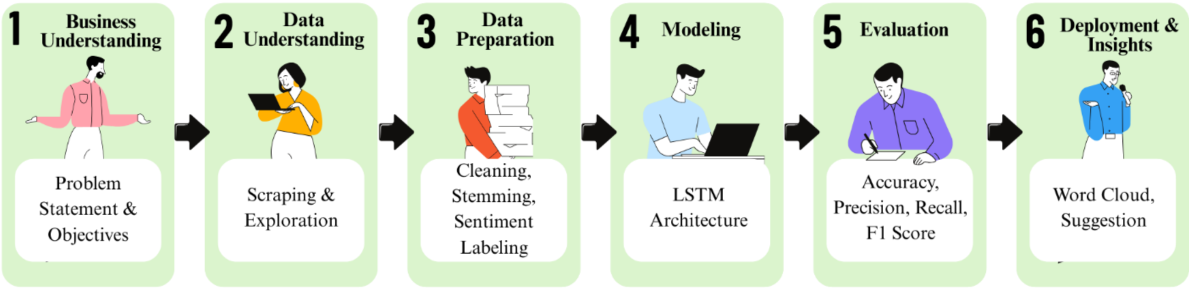

Figure 1. CRISP-DM Research Design for Livin' Sentiment Analysis

2.1. Business Understanding

This study supports Bank Mandiri in understanding user perceptions of its Livin' mobile banking application, as shown in Figure 2. Despite continued growth in downloads, users have reported inconsistent experiences, particularly related to slow login times and failed transactions (Sulistiyani et al., 2024). Such mixed feedback makes it difficult for product teams to determine which features promote user trust and which one's trigger dissatisfaction (Rosman et al., 2022). In Indonesia's competitive digital banking sector, relying solely on intuition without empirical insight may lead to misplaced development efforts and declining customer loyalty (Wati et al., 2024). Therefore, this study formulates a clear problem statement and outlines objectives to identify interface elements that affect user confidence and measure their impact on sentiment. The research converts visual and functional components into quantifiable experience indicators to analyze user feedback patterns across different groups (Rahmi & Handayani, 2024). These findings are translated into actionable recommendations that can improve satisfaction and retention, reinforcing Bank Mandiri's position in the mobile banking industry (Sebayang et al., 2024).

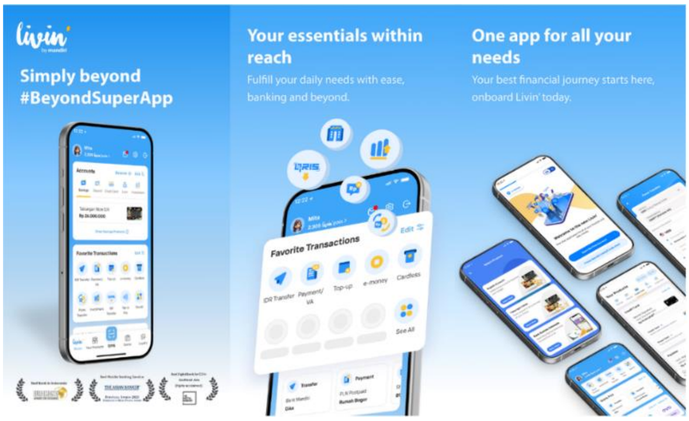

Figure 2. User Interface of the Livin' by Mandiri Mobile Banking Application

2.2. Dataset Understanding

To build the sentiment analysis corpus, this study collected 5,000 Indonesian-language user reviews from the Livin' by Mandiri application on Google Play between 1 May and 30 June 2025. The primary data source for this study is the official Google Play Store page of the Livin' by Mandiri mobile banking application. Each entry includes the full review text, star rating, and timestamp, resulting in a temporally rich and diverse dataset that reflects real-world user interactions with mobile banking services (Asri et al., 2025). These early explorations revealed critical signals, such as spikes in negative reviews following system updates or outages, offering contextual clues for temporal sentiment shifts (Aris et al., 2024). Together, this stage provided foundational insight into user behaviour and informed the next steps in sentiment classification modelling.

2.3. Data Preparation

All user reviews were processed through a structured text-cleaning pipeline to ensure consistency and accuracy for sentiment analysis. This included trimming whitespace, lowercasing, and removing URLs, emojis, and repeated punctuation. An Indonesian-specific stop-word list was applied to eliminate low-value terms, and the Sastrawi stemmer was used to normalize word forms by reducing inflected variants to their roots (e.g., *mentransfer*, *ditransfer* → *transfer*)(Fahmi et al., 2020). Finally, non-Indonesian texts were filtered out using the langid tool, retaining only those with a language probability score above 0.80. (Setiabudi et al., 2021).

Each review's sentiment was scored using the INAfeel lexicon, a manually curated Indonesian sentiment dictionary that assigns a polarity value ωi to each cleaned token. The sentiment score for a review r is computed using the formula (1):

$$(r) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \omega i \qquad (1)$$

where N is the number of tokens after preprocessing. Based on empirical calibration, reviews with $Score(r)$ ≥0.05 are labelled positive, those with $Score(r)$ ≤−0.05 as negative, and those in between as neutral (Fide et al., 2021). To assess the reliability of this lexicon-based annotation, a manual validation was conducted on 300 randomly selected reviews, coded by bilingual annotators. Inter-rater agreement was calculated using Cohen's Kappa (2), yielding:

$$\kappa = \frac{P_0 - P_e}{1 - P_e} = 0{,}82 \tag{2}$$

The calculation above indicating almost perfect agreement. This validation confirms that combining INAfeel for polarity and Sastrawi stemming offers a dependable foundation for training sentiment classifiers in Indonesian (Hidayatullah, 2024).

2.4. Modelling

The sentiment classification model in this study begins with a 128-dimensional embedding layer that converts tokens into dense vectors, optimized during training to capture semantic meaning. These embeddings feed into two stacked LSTM layers with 128 and 64 hidden units, respectively, to model both local and global dependencies in the text (Af'idah et al., 2020). Dropout (0.3) is applied after each LSTM to prevent overfitting, followed by a dense ReLU-activated layer with 32 neurons and a softmax output that classifies reviews into positive, neutral, or negative sentiments. The model is trained shown in Table 1 using the Adam optimizer with a learning rate of 1×10^{-3}, batch size of 32, for 15 epochs. Gradient clipping (norm = 5), early stopping (patience = 3), and learning rate scheduling help stabilize training and prevent overfitting (Abhisikta et al., 2024). The model uses categorical cross-entropy as its objective function, mathematically expressed as (3):

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}.\sum_{c=1}^{3} y_{ic}\,\log(\hat{y}_{ic}) \tag{3}$$

where y_{ic} is the one-hot ground truth for review i and class c. Gradient norms are clipped at 5 to stabilise updates, a scheduler halves the learning rate when validation loss stagnates for two epochs, and early stopping with patience 3 monitors macro F1 to prevent overfitting (Singh et al., 2022).

Table 1. Hyperparameters And Their Constraints

Hyperparameters	Description & Constraints
Embedding size	128 dimensions, uniform initialisation, trainable
LSTM layers	Two layers; first 128 units, second 64 units
Dropout rate	0.3 after each LSTM output; tuned in {0.2, 0.3, 0.4}
Dense layer	32 neurons, ReLU activation
Output layer	Softmax over 3 sentiment classes
Batch size	32, constrained by GPU memory (≈ 3 GB)
Epochs	15 maximum with early stopping patience 3
Optimiser	Adam, learning rate 1×10^{-3}; $\beta_1 = 0.9$, $\beta_2 = 0.999$
Gradient clip	$\|\nabla\theta\|_2 \leq 5$
Seed	42 for NumPy, PyTorch, and CUDA

2.5. Evaluation

The labelled dataset is divided into 80% for training and 20% for testing using stratified sampling to preserve the original distribution of positive, neutral, and negative sentiments across both splits. This prevents minority classes, especially neutral, from being underrepresented in the evaluation set. A fixed random seed (42) is used to ensure reproducibility across experiments, and no oversampling or under sampling is applied to keep the class distribution reflective of real-world conditions. Model performance is evaluated using four macro averaged metrics: accuracy (4), precision (5), recall (6), and F1 score (7), with each metric computed per class using the one versus all method and averaged without class weighting to give equal importance to all sentiment categories.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = 2 \bullet \frac{Precision \bullet Recall}{Precision + Recall} \tag{7}$$

Here N is the total number of test instances, while TP, FP, and FN denote true positives, false positives, and false negatives for class c. In this study a true positive (TP) occurs when a review that is genuinely labelled *positive* is also predicted *positive* by the model, whereas a false positive (FP) arises when the model predicts *positive* for a review that is in fact *neutral* or *negative*. Conversely, a false negative (FN) is a genuinely *positive* review that the model misclassifies as *neutral* or *negative*, and a true negative (TN) represents any review that is truly *neutral* or *negative* and is likewise predicted as either *neutral* or *negative*, indicating a correct rejection of the positive class. Intuitively, high precision means the model seldom mis-labels other sentiments as class c; high recall means it rarely overlooks genuine examples of class c. The F1 score balances these two aspects, penalising systems that favour one at the expense of the other (Yonia et al., 2024). Overall accuracy gives a quick snapshot of correctness but can mask class-specific weaknesses, hence the additional per-class metrics.

2.6. Deployment & Insight

In the Deployment and Insight stage, each review that has received a sentiment label flows into two automatic modules. First, the system creates word clouds and n gram tables for bigrams and trigrams within each sentiment class by counting the most frequent terms and phrases; these visuals refresh on a set schedule, for example once a day, and appear on an interactive dashboard so product teams can quickly observe dominant themes of praise or complaint. Second, a recommendation module scans the list of recurring negative n grams, maps them to specific Livin interface features, and compiles a prioritised list of UX improvements together with response templates for customer service; the output forms a structured backlog that links every feature to its priority and estimated sprint effort and provides agents with draft replies, ensuring that sentiment insights immediately guide application development and service policy.

3. Results and Discussion

The implementation of web scraping, text preprocessing, lexicon-based sentiment labelling, and LSTM model training used in this study is available at the following GitHub repository:https://github.com/kheylina/Analisa-Sentimen-dan-Webscraping-aplikasi-Livin. The repository includes all relevant scripts, annotated datasets, and Jupyter notebooks used throughout the modelling pipeline. This section presents the experimental results, covering sentiment distribution, model evaluation, keyword analysis, and practical implications.

3.1. Text Preprocessing

Before training the sentiment classification model, all 5,000 Livin' by Mandiri user reviews were processed through a comprehensive text preprocessing pipeline to ensure input consistency and clarity. The scraping process was executed using automated Python scripts based on the google-play-scraper library, which retrieved structured metadata including review content, posting date, and star ratings (Nugroho et al., 2021). The result of the web scraping process, which contains extracted user reviews, is illustrated in Figure 3. Review entries were filtered to remove short or duplicate comments and ensure data relevance across user groups.

	ulasan
0	sangat membantu
1	sering matikan hp dulu.kemudiandihidupkan .bar...
2	sangat bagus dan membantu sekali
3	sangat puas dengan aplikasi ini
4	sangat membantu

Figure 3. Visualization of the User Review

Exploratory analysis was conducted to examine sentiment distribution, detect outliers, and identify recurring complaints. After preprocessing steps such as lowercasing, punctuation removal, slang normalization, stop word removal, and stemming using the Sastrawi stemmer, the reviews were tokenized and padded to 50 tokens is shown in Figure 4. Sentiment labelling was then performed using a lexicon-based method by comparing the number of positive and negative words in each review: a review is labelled positive if positive words dominate, negative if negative words dominate, and neutral if the counts are equal. An example of the resulting dataset after preprocessing and lexicon-based labelling is shown in Figure 5.

	ulasan	clean_ulasan
0	sangat membantu	sangat bantu
1	sering matikan hp dulu.kemudiandihidupkan .bar...	sering mati hp dulukemudiandihidupkan baru bis...
2	sangat bagus dan membantu sekali	sangat bagus dan bantu sekali
3	sangat puas dengan aplikasi ini	sangat puas dengan aplikasi ini
4	sangat membantu	sangat bantu

Figure 4. Example of Raw and Cleaned User Reviews

	ulasan	clean_ulasan	sentimen
0	sangat membantu	sangat bantu	positif
1	sering matikan hp dulu.kemudiandihidupkan .bar...	sering mati hp dulukemudiandihidupkan baru bis...	netral
2	sangat bagus dan membantu sekali	sangat bagus dan bantu sekali	positif
3	sangat puas dengan aplikasi ini	sangat puas dengan aplikasi ini	positif
4	sangat membantu	sangat bantu	positif
5	Sangat bagus tolong dung akhir2 ini apk sering...	sangat bagus tolong dung akhir ini apk sering ...	positif
6	sangat baik	sangat baik	positif
7	Ok	ok	netral
8	Kenapa yaa klo mau transfer daftar nama si pen...	kenapa yaa klo mau transfer daftar nama si ter...	netral
9	ok	ok	netral
10	Baik untuk mempermudah nasabah tp ni troble	baik untuk mudah nasabah tapi ni troble	positif
11	sangat baek sya gunakan licin mandiri	sangat baek saya guna licin mandiri	netral
12	tdk bisa Di update	tidak bisa di update	negatif

Figure 5. Sample of Sentiment-Labelled User Reviews using Lexicon-Based Classification

3.2. Sentiment Distribution

A total of 5,000 user reviews of the Livin' by Mandiri application were collected using the google-play-scraper tool, covering a broad range of user feedback from May to June 2025. Each review was then automatically labelled using a lexicon-based approach and categorized into one of three sentiment classes: negative, neutral, or positive. As detailed in Table 2 and illustrated in Figure 5, negative sentiment dominates the dataset, with 1,882 reviews (37.6%), followed closely by neutral sentiment with 1,807 reviews (36.1%). Positive sentiment accounts for the smallest proportion, totalling 1,311 reviews or 26.2% of the corpus. These results reveal that over a third of the users expressed dissatisfaction, especially regarding the application's technical issues such as login delays, transaction failures, or slow responsiveness. The high prevalence of negative and neutral feedback presents a strong indicator that the app's user experience still requires significant improvement. This distribution pattern serves as an actionable insight for developers and product strategists, guiding efforts to enhance user satisfaction and retention through targeted design and performance optimizations, as visualized in Table 2 and Figure 5.

Table 2. Sentiment Distribution

Sentiment	Number of Reviews	Percentage (%)
Negative	1,882	37.6 %
Neutral	1,807	36.1 %
Positive	1,311	26.2 %
Total	5,000	100 %

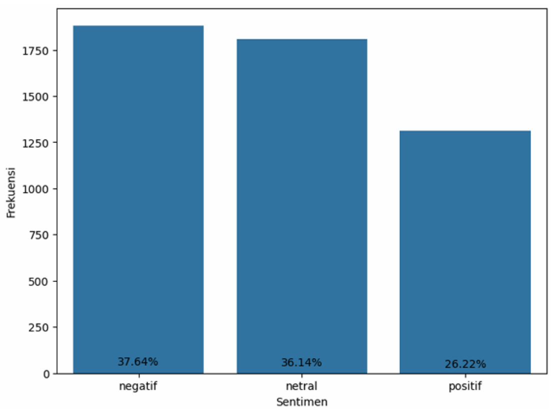

Figure 5. Sentiment Distribution in Livin' By Mandiri Dataset

3.3. Model Evaluation

The LSTM model was trained using 80% of the data for training and 20% for testing over 5 epochs with a batch size of 16. As shown in Table 3, the model achieved high performance across all sentiment classes. Specifically, it reached 99% precision and 98% recall for positive sentiment, 97% precision and 98% recall for neutral sentiment, and 99% precision and 99% recall for negative sentiment. The macro-averaged accuracy across all classes was 98%, demonstrating the model's excellent generalizability and robustness for real-world sentiment classification in Indonesian-language mobile banking reviews. These metrics suggest that the model not only correctly identifies sentiment polarity with high fidelity, but also maintains strong recall for minority classes, avoiding bias toward the dominant sentiment. The high F1-scores across all classes further confirm that the model balances precision and recall effectively, ensuring both accuracy and coverage in its predictions.

Table 3. Model Evaluation Metrics

Sentiment Class	Precision	Recall	F1-Score	Accuracy
Positive	0.99	0.98	0.99	0.98
Neutral	0.97	0.98	0.98	0.98
Negative	0.99	0.99	0.99	0.99
Macro Average	0.98	0.98	0.98	0.98

3.4. Keyword Analysis

To gain deeper insights into user perspectives across sentiment categories, a keyword frequency analysis was conducted using Word Cloud visualizations, which provide an intuitive view of dominant vocabulary patterns within user feedback. After undergoing thorough preprocessing, the reviews were grouped by sentiment class: negative, neutral, and positive. Each group was then analysed independently to extract the most frequently occurring words, allowing the identification of recurring expressions, concerns, or praises specific to each sentiment type. This granular approach helps highlight sentiment-specific themes such as technical complaints in negative reviews, generic usage comments in neutral feedback, and expressions of satisfaction or ease in positive reviews. These findings not only enrich the interpretation of user sentiment but also offer valuable direction for targeted product improvements and service communication strategies.

(a) Positive Sentiment (b) Neutral Sentiment (c) Negative Sentiment

Figure 6. Word Cloud of User Reviews Categorized by Positive, Neutral, and Negative Sentiment

The Word Cloud results are illustrated in Figure 6, with the following observations:

1) Positive Sentiment

Reviews classified as positive frequently include words such as “bantu” (helpful), “mudah” (easy), “praktis” (practical), and “cepat” (fast). These keywords indicate user appreciation toward the application’s convenience and functionality, especially in terms of ease of use, transaction speed, and reliability. Words like “bagus”, “baik”, and “puas” also appear prominently, reflecting a generally satisfying user experience.

2) Neutral Sentiment

Neutral reviews are often descriptive or factual. Common words include “atm”, “bank”, “mandiri”, “akun”, and “debit”, suggesting reviews that mention banking activities without clear emotional expression. These reviews may contain feedback, inquiries, or brief remarks unrelated to satisfaction or dissatisfaction.

3) Negative Sentiment

Negative reviews predominantly contain words such as “tidak” (not), “error”, “gagal” (fail), “macet” (lag/crash), and “lemot” (slow). These words reveal user frustration with technical issues like failed login, app crashes, or slow performance. The word “kecewa” (disappointed) is also a strong indicator of unmet expectations.

The keyword visualizations reinforce the sentiment classification results by illustrating the dominant themes within each sentiment group. Positive sentiment is associated with words that express satisfaction and ease of use, whereas negative reviews prominently feature technical complaints. The prevalence of these negative terms, which corresponds to 37.6% of the dataset, underscores the urgency of addressing issues related to system stability, login processes, and overall app performance.

3.5. Discussion

The high proportion of negative sentiment (37.6%), dominated by login failures, application crashes, and slow response times, aligns with prior studies on Indonesian mobile banking apps such as BCA Mobile and BRImo (Adiningtyas & Auliani, 2024), where similar technical reliability issues were the primary sources of dissatisfaction. This alignment reinforces established studies (Purba & Safrin, 2024) and (Sebayang et al., 2024*)*, which highlight system performance and reliability as key determinants of user satisfaction and continued use. The findings not only confirm these earlier reports but also extend the evidence base by quantifying sentiment distribution and linking it to specific technical pain points in the context of one of Indonesia’s most widely used banking applications.

The LSTM classifier in this study achieved 98% accuracy, outperforms typical results from earlier Indonesian sentiment studies commonly used in prior Indonesian-language sentiment studies using TF-IDF + SVM (e.g., 93–95% in (Bimantara & Zufria, 2024)) or IndoBERT fine-tuning (96% in (Nugroho et al., 2021)). This superior performance supports findings from NLP (Vishnuprabha et al., 2021) that recurrent architectures like LSTM are more effective than traditional classifiers in capturing sequential dependencies and contextual meaning in noisy, short-text reviews. In the context of mobile banking, this higher accuracy enables more reliable real-time sentiment monitoring, directly addressing the need for automated, domain-specific analysis identified in previous research.

From a practical perspective, these findings provide targeted, data-driven insights that Bank Mandiri can apply to enhance its digital banking services, in direct response to previously unaddressed gaps in sentiment modelling for Indonesian financial applications:

1) Real Time Sentiment Monitoring

The LSTM based sentiment classifier, trained on a curated and automatically labelled dataset of Livin’ user reviews, can be embedded into Bank Mandiri’s internal dashboards. This enables continuous real time monitoring of user sentiment trends. It provides early warnings of negative sentiment spikes, especially after system updates or outages, allowing teams to act quickly before issues escalate.

2) Feature Improvement Roadmap

The dominance of complaints about app crashes, failed logins, and loading delays, reflected in both the sentiment distribution and keyword analysis, underscores the urgent need for technical enhancements. These insights can help set development priorities, ensuring stability, reliability, and responsiveness in future releases.

3) User Experience Optimization

By associating sentiment polarity with frequently mentioned terms, product teams can identify recurring friction points and use these insights to inform UX and UI design changes. For instance, optimizing transaction workflows or clarifying error messages can alleviate user confusion and improve satisfaction.

4) Strategic Communication Support

The ability to detect shifts in sentiment in real time also equips customer service and communication teams to tailor their messaging more effectively. During service disruptions, empathetic and timely responses grounded in actual user feedback can help preserve trust and reduce churn.

This study contributes both a robust sentiment classification model and an evidence-based framework that bridges theory and practice in digital banking sentiment analysis. The framework replicates the methodological rigor of lexicon-based labelling combined with deep learning, while extending prior literature by applying it to a large, domain specific dataset in Bahasa Indonesia. By integrating insights from empirical results with established models of user behaviour and digital banking adoption, this research provides a blueprint for translating unstructured user feedback into targeted service enhancements thereby advancing both scholarly understanding and operational strategies in the Indonesian banking sector.

4. Conclusion

This study concludes that user sentiment toward the Livin' by Mandiri mobile banking application derived from 5,000 Google Play Store reviews is predominantly negative (37.6%), with recurring technical issues such as login failures, slow response, and application crashes that, if left unaddressed, risk eroding customer trust and loyalty. The LSTM-based model, trained on a domain-specific lexicon-labelled dataset, achieved 98.2% accuracy and outperformed traditional classifiers in prior Indonesian sentiment studies, making it a strong candidate for integration into real-time sentiment monitoring systems. These findings highlight the strategic value of embedding sentiment analysis into service improvement cycles, enabling real-time detection of sentiment shifts after updates or outages, a data-driven roadmap for addressing high-impact technical issues, user experience optimization through targeted design changes, and strategic communication to maintain customer confidence during disruptions. Future research should extend this work by incorporating multi-platform and longitudinal datasets, developing transformer-based models tailored for Bahasa Indonesia in mobile banking contexts, and applying aspect-based sentiment analysis to link sentiment directly to specific features such as transfers or bill payments providing both higher analytical precision and actionable intelligence for sustaining competitive advantage in Indonesia's digital banking sector.

5. References

Abhisikta, A., Agarwal, M., Mallick, P. K., & Padhy, N. (2024). Beyond Words: Contextual Sarcasm Detection in News Texts using Advanced Models. *ESIC 2024 - 4th International Conference on Emerging Systems and Intelligent Computing, Proceedings*, 566–571. https://doi.org/10.1109/ESIC60604.2024.10481639

Adiningtyas, H., & Auliani, A. S. (2024). Sentiment analysis for mobile banking service quality measurement. *Procedia Computer Science*, *234*, 40–50.

https://doi.org/10.1016/J.PROCS.2024.02.150

Af'idah, D. I., Kusumaningrum, R., & Surarso, B. (2020). Long short term memory convolutional neural network for Indonesian sentiment analysis towards touristic destination reviews. *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 630–637. https://doi.org/10.1109/ISEMANTIC50169.2020.9234210

Aris, V., Ashdaq, M., & Taufik, M. (2024). Data Mining and Social Media Analysis to Identify Key Dimension of Service Quality for BTN Mobile App in Indonesia. *Jurnal Ekonomi Dan Bisnis Digital*, *3*(1), 143–154. https://doi.org/10.55927/MINISTAL.V3I1.7565

Asri, Y., Kuswardani, D., Suliyanti, W. N., Manullang, Y. O., & Ansyari, A. R. (2025). Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN mobile application. *Indonesian Journal of Electrical Engineering and Computer Science*, *38*(1), 677–688. https://doi.org/10.11591/IJEECS.V38.I1.PP677-688

Bimantara, M. D., & Zufria, I. (2024). Text Mining Sentiment Analysis on Mobile Banking Application Reviews using {TF-IDF} Method with Natural Language Processing Approach. *JINAV: Journal of Information and Visualization*, *5*(1), 115–123. https://doi.org/10.35877/454RI.JINAV2772

Fahmi, S., Purnamawati, L., Shidik, G. F., Muljono, M., & Fanani, A. Z. (2020). Sentiment analysis of student review in learning management system based on sastrawi stemmer and SVM-PSO. *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 643–648. https://doi.org/10.1109/ISEMANTIC50169.2020.9234291

Farida, F., Binastuti, S., Silfianti, W., & Hermana, B. (2023). The impact of Covid-19 on electronic payment system: Lesson learned from Indonesian banking. *Technium Social Sciences Journal*, *40*, 200–212. https://doi.org/10.47577/TSSJ.V40I1.8295

Fide, S., Suparti, S., & Sudarno, S. (2021). Analisis Sentimen Ulasan Aplikasi Tiktok di Google Play Menggunakan Metode Support Vector Machine (SVM) dan Asosiasi. *Jurnal Gaussian*, *10*(3), 346–358. https://doi.org/10.14710/J.GAUSS.10.3.346-358

Gowandi, T., Murfi, H., & Nurrohmah, S. (2021). Performance Analysis of Hybrid Architectures of Deep Learning for Indonesian Sentiment Analysis. *Communications in Computer and Information Science*, *1489 CCIS*, 18–27. https://doi.org/10.1007/978-981-16-7334-4_2

Hidayat, A., & Kassim, S. (2023). The Determinants of Digital Banking Adoption Among Banks Offering Islamic Banking Services. *Journal of Islamic Monetary Economics and Finance*, *9*(4), 559–588. https://doi.org/10.21098/JIMF.V9I4.1688

Hidayatullah, A. F. (2024). Code-Mixed Sentiment Analysis on Indonesian-Javanese-English Text Using Transformer Models. *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2024*, 340–345. https://doi.org/10.1109/ICITISEE63424.2024.10730138

Lovitasari Yonia, D., Irham, A., Oranova Siahaan, D., & Mahfud, I. (2024). Learning Models for Software Feature Extraction from Disaster Tweets: A Comparative Study. *ICECOS 2024 - 4th International Conference on Electrical Engineering and Computer Science, Proceeding*, 83–88. https://doi.org/10.1109/ICECOS63900.2024.10791279

Lubis, F. A. (2024). User Sentiment Analysis Towards Islamic Banking Applications in Indonesia. *Journal of Islamic Economic and Business Research*, *4*(1), 126–143. https://doi.org/10.18196/JIEBR.V4I1.342

Mulyanto, S., Yonia, D. L., & Sutejo, B. (2025). Finance Loan Risk Assessment Using Machine Learning for Credit Eligibility Prediction and Model Optimization. *IJISTECH (International Journal of Information System and Technology)*, *8*(5), 303–311. https://doi.org/10.30645/IJISTECH.V8I5.376

Nugroho, K. S., Sukmadewa, A. Y., Wuswilahaken Dw, H., Bachtiar, F. A., & Yudistira, N. (2021). BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews. *ACM International Conference Proceeding Series*, 258–264. https://doi.org/10.1145/3479645.3479679

Purba, E. A., & Safrin, F. A. (2024). Pengaruh Online Customer Review, Harga dan Online Customer Ratting Terhadap Keputusan Pembelian Melalui Tokopedia (Studi Pada Konsumen Toko Online Ventela Store Official di Kota Medan). *Jurnal Ekonomi Kreatif dan Manajemen Bisnis Digital*, *2*(3), 362–378. https://doi.org/10.55047/JEKOMBITAL.V2I3.613

Putri, D., Simangunsong, T., Purwana, D., & Musyaffi, A. (2025). Pengaruh E-Service Quality dan Kepuasan Nasabah terhadap Loyalitas Nasabah dalam Menggunakan Livin by Mandiri. *Jurnal Yudistira : Publikasi Riset Ilmu Pendidikan Dan Bahasa*, *3*(2), 07–23. https://doi.org/10.61132/YUDISTIRA.V3I2.1613

Rahmi, R. A., & Handayani, P. W. (2024). The influence of users' perspective factors on mobile banking adoption in Indonesia. *Journal of Science and Technology Policy Management*, *15*(6), 1408–1425. https://doi.org/10.1108/JSTPM-01-2022-0008/FULL/XML

Rosman, M., Setiyawati, N., & Hosanna Bangkalang, D. (2022). The Comparison of Evaluation on User Experience and Usability of Mobile Banking Applications Using User Experience Questionnaire and System Usability Scale. *Proceedings 2022, Vol. 82, Page 87*, *82*(1), 87. https://doi.org/10.3390/PROCEEDINGS2022082087

Sebayang, T. E., Hakim, D. B., Bakhtiar, T., & Indrawan, D. (2023). What Accelerates the Choice of Mobile Banking for Digital Banks in Indonesia? *Journal of Risk and Financial Management 2024, Vol. 17, Page 6*, *17*(1), 6. https://doi.org/10.3390/JRFM17010006

Sebayang, T. E., Hakim, D. B., Bakhtiar, T., & Indrawan, D. (2024). The Investigation of Preference Attributes of Indonesian Mobile Banking Users to Develop a Strategy for Mobile Banking Adoption. *Journal of Risk and Financial Management 2024, Vol. 17, Page 109*, *17*(3), 109. https://doi.org/10.3390/JRFM17030109

Setiabudi, R., Iswari, N. M. S., & Rusli, A. (2021). Enhancing text classification performance by preprocessing misspelled words in Indonesian language. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *19*(4), 1234–1241. https://doi.org/10.12928/TELKOMNIKA.V19I4.20369

Singh, A., Thapliyal, R., Vanave, R., Shedge, R., & Mumbaikar, S. (2022). Analysis of hyperparameters in Sentiment Analysis of Movie Reviews using Bi-LSTM. *ITM Web of Conferences*, *44*, 03012. https://doi.org/10.1051/ITMCONF/20224403012

Sulistiyani, Nurchayati, Nurchayati, & Handani, N. D. (2024). User Experience of Mobile Banking Application in Indonesia: New Technology of Banking. *Global Business and Finance Review*, *29*(2), 127–141. https://doi.org/10.17549/GBFR.2024.29.2.127

Tanika, C., William, Hartanto, M. P., & Tanjung, M. (2022). Factors Influencing Indonesian Consumers to Use Digital Banking. *Proceedings - 4th International Conference on Informatics, Multimedia, Cyber and Information System, ICIMCIS 2022*, 546–551. https://doi.org/10.1109/ICIMCIS56303.2022.10017694

Vishnuprabha, V., Murali, L., & Viswanathan, D. M. (2021). Natural Language Processing. *Cognitive Computing Systems*, 349–400. https://doi.org/10.1201/9781003082033-19

Wati, K. E., Khuzaini, K., & Wijaya, N. Q. (2024). Why Users Choose Mobile Banking: Behavioral Insights using UTAUT. *International Journal of Industrial Engineering, Technology & Operations Management*, *2*(1), 19–25. https://doi.org/10.62157/IJIETOM.V2I1.56

Yonia, D. L., Ariansyah, I., Dina, A. B., & Suciati, N. (2024). Enhancing Image Captioning Performance with VGG16 Feature Extraction and LSTM Sequence Processing. *ICECOS 2024 - 4th International Conference on Electrical Engineering and Computer Science, Proceeding*, 77–82. https://doi.org/10.1109/ICECOS63900.2024.10791213