SENTIMENT ANALYSIS OF JAKLINGKO APP REVIEWS USING MACHINE LEARNING AND LSTM

Maghfiroh Maulani[1*]; Windu Gata[2]

Computer Science[1,2]
Universitas Nusa Mandiri, Jakarta, Indonesia[1,2]
https://www.nusamandiri.ac.id/[1,2]
14230031@nusamandiri.ac.id[1*], windu@nusamandiri.ac.id[2]

(*) Corresponding Author

Abstract— *Application-based transportation services have rapidly developed in recent years, with various studies indicating that service quality and user experience play a crucial role in the adoption of this technology. Previous research has analyzed user satisfaction with digital transportation applications, highlighting factors such as ease of use, service reliability, and the effectiveness of fare systems. This study aims to analyze user sentiment toward the JakLingko application to assess satisfaction levels and identify aspects that need improvement. Utilizing a dataset of 200 user reviews, this research applies data preprocessing techniques to clean and organize the information before performing sentiment classification. The machine learning models used include Naïve Bayes, Random Forest, Support Vector Machine, Logistic Regression, Decision Tree, and Long Short-Term Memory (LSTM), categorizing sentiment into positive, negative, and neutral. The analysis results indicate a dominance of negative sentiment in user reviews, reflecting a significant level of dissatisfaction with the application. This highlights major challenges in the implementation of transportation applications, potentially affecting public adoption and trust in the service. Therefore, besides providing insights into user perceptions, this study also proposes improvement strategies aimed at enhancing features and the overall user experience. Given the high proportion of negative sentiment, this research emphasizes the importance of improving the accuracy of sentiment analysis models to generate deeper and more precise insights. These findings can serve as a foundation for designing policies and strategies to improve application-based transportation services, ultimately enhancing service quality and expanding user adoption.*

Keywords: *JakLingko, Machine Learning, Sentiment Analysis, Text Mining, User Reviews.*

Intisari— *Layanan transportasi berbasis aplikasi telah berkembang pesat dalam beberapa tahun terakhir, dengan berbagai penelitian menunjukkan bahwa kualitas layanan dan pengalaman pengguna memainkan peran penting dalam adopsi teknologi ini. Beberapa studi sebelumnya telah menganalisis kepuasan pengguna terhadap aplikasi transportasi digital, dengan menyoroti faktor-faktor seperti kemudahan penggunaan, keandalan layanan, dan efektivitas sistem tarif. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi JakLingko guna menilai tingkat kepuasan serta mengidentifikasi aspek yang perlu ditingkatkan. Dengan memanfaatkan dataset yang terdiri dari 200 ulasan pengguna, penelitian ini menerapkan teknik prapemrosesan data untuk membersihkan dan mengorganisasi informasi sebelum dilakukan klasifikasi sentimen. Model machine learning yang digunakan meliputi Naïve Bayes, Random Forest, Support Vector Machine, Logistic Regression, Decision Tree, dan Long Short-Term Memory (LSTM), dengan kategori sentimen yang diklasifikasikan menjadi positif, negatif, dan netral. Hasil analisis menunjukkan dominasi sentimen negatif dalam ulasan pengguna, yang mencerminkan tingkat ketidakpuasan yang cukup tinggi terhadap aplikasi. Hal ini menyoroti tantangan signifikan dalam penerapan aplikasi transportasi, yang berpotensi memengaruhi tingkat penerimaan serta kepercayaan masyarakat terhadap layanan tersebut. Oleh karena itu, selain menggambarkan persepsi pengguna, penelitian ini juga mengusulkan strategi perbaikan yang bertujuan untuk meningkatkan fitur serta pengalaman pengguna secara keseluruhan. Mengingat tingginya proporsi sentimen negatif, penelitian ini menekankan pentingnya meningkatkan akurasi model analisis sentimen agar dapat menghasilkan wawasan yang lebih mendalam dan akurat. Temuan ini dapat dijadikan dasar dalam merancang kebijakan serta strategi peningkatan* *layanan transportasi berbasis aplikasi, yang pada akhirnya diharapkan dapat meningkatkan kualitas layanan dan memperluas adopsi aplikasi oleh masyarakat.*

Kata Kunci*: JakLingko, Pembelajaran Mesin, Analisis Sentimen, Penambangan Teks, Ulasan Pengguna.*

INTRODUCTION

The rapid growth of mobile applications necessitates a comprehensive understanding of user feedback to enhance satisfaction. JakLingko, a transportation app in Indonesia, has received diverse feedback regarding its performance and usability. Analyzing user sentiment toward the application provides valuable insights into areas that require improvement.

The JakLingko Program integrates public transportation in Jakarta, enhancing accessibility and facilities. While prior research has discussed the role of collaborative governance in increasing public interest in public transport and reducing traffic congestion, its direct impact on user sentiment toward the JakLingko app remains underexplored (Hanifa Aulia Rahma, Slamet Rosyadi, Guntur Gunarto, 2024). The JakLingko app facilitates an integrated fare system for public transportation in Jabodetabek, enabling users to conveniently access various modes such as KRL, Transjakarta, MRT, and LRT through a unified ticketing system. This integration is expected to enhance user experience and encourage greater adoption of public transport (Putranto, 2023). Additionally, the JakLingko app is part of the Mikrotrans program, which integrates feeder transport into Jakarta's Bus Rapid Transit system to provide affordable and high-quality public transportation while improving operational patterns in small bus services (Gary Ekatama Bangun, n.d.). Understanding user sentiment toward these services can help in refining the app's functionality to meet user expectations.

Sentiment analysis involves the automatic identification of sentiment expressed in textual data, utilizing machine learning and deep learning techniques for improved accuracy across various applications, including social media monitoring and product reviews (Suryawanshi, 2024). Prior research highlights the effectiveness of sentiment analysis in evaluating public perception of digital services (Gary Ekatama Bangun, n.d.; Putranto, 2023), highlighting its relevance in assessing user feedback on the JakLingko app. By leveraging sentiment analysis, developers can systematically address critical user concerns and enhance the overall application experience, ensuring better alignment with user needs and expectations.

This study applies machine learning and deep learning techniques to classify user sentiments and extract meaningful insights for improving the JakLingko application. The selection of machine learning algorithms, including Naïve Bayes, Random Forest, Support Vector Machine, Logistic Regression, and Decision Tree, is based on their effectiveness in text classification tasks. Naïve Bayes is widely used for probabilistic classification, Random Forest and Decision Tree provide robustness and interpretability, while Support Vector Machine and Logistic Regression are known for their strong performance in high-dimensional data.

Additionally, this study employs Long Short-Term Memory (LSTM), a deep learning algorithm, due to its superior ability to capture long-range dependencies and contextual relationships in text data. The utilization of LSTM in this research aligns with prior findings on its effectiveness in handling natural language processing tasks, ensuring robust sentiment analysis for user feedback interpretation. By integrating both traditional machine learning and deep learning techniques, this study aims to enhance the accuracy and reliability of sentiment classification in the JakLingko application.

By integrating these methodologies, this study aims to provide a structured framework for assessing user sentiment, enabling targeted improvements to the JakLingko app, and enhancing its usability and adoption among the public.

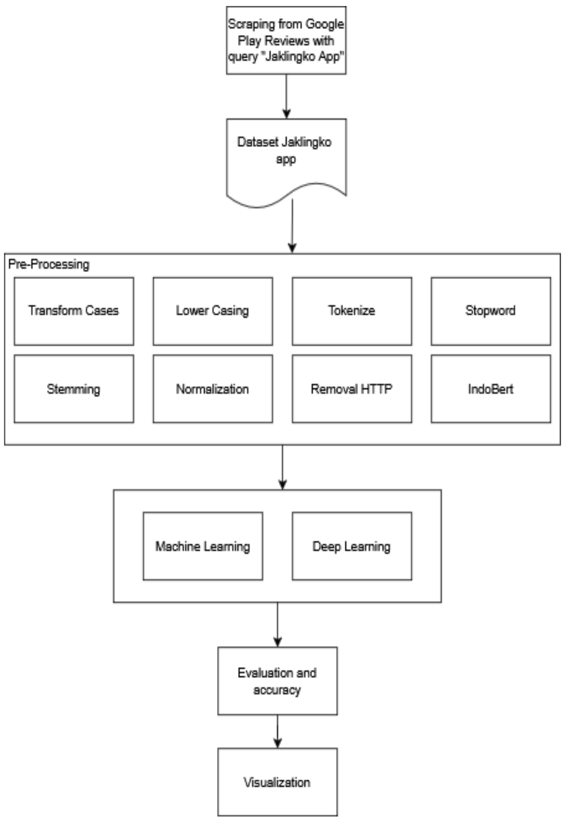
MATERIALS AND METHODS

This study employed a dataset of 200 user reviews collected from the JakLingko app. The data preprocessing steps included text cleaning, eliminating duplicates, and addressing missing values. Sentiment analysis was conducted using machine learning models like *Naïve Bayes, Random Forest, Support Vector Machine, Logistic Regression, Decision Tree, and Long Short-Term Memory (LSTM)*. The analysis was carried out using Python libraries such as The *Natural Language Toolkit* (NLTK), Sastrawi, and Scikit-learn.

The study focuses on sentiment analysis of Google Play Store app reviews, utilizing ensemble approaches like Random Forest and Boosting, which outperform individual algorithms. It highlights the importance of understanding user feedback for application developers to enhance their products (Bikbov et al., 2020)(Idris, 2024). The Jak Lingko app is part of Jakarta's transportation integration initiative, facilitating collaboration between the government and transportation service providers. It aims to streamline various modes of transport, addressing traffic issues and enhancing public transportation efficiency in the city (Adinegoro, 2022).

The paper discusses using deep learning techniques, specifically CNN and LSTM models, for Indonesian fake news classification. It incorporates the pre-trained IndoBERT language model to enhance performance, achieving a notable accuracy increase in identifying fake news (Nugraheni, 2024). Besides the BERT-based model, IndoBERT is an adaptation of BERT for the Indonesian language, designed for sentiment analysis. It demonstrates good performance but is affected by label uncertainty, which can be improved using methods like Confident Learning to enhance model accuracy and reliability (Liebenlito, 2024).

This research utilizes an approach that involves comparing machine learning and deep learning methods to assess sentiment regarding the Jaklingko app through reviews on Google Play Store. The purpose of this approach is to examine the relative strengths of both traditional and modern methods in text-based sentiment analysis. As a result, the findings from this study are anticipated to aid in the advancement of more efficient and accurate sentiment analysis techniques in the future. The research method for this study, focusing on sentiment analysis of the Jaklingko app on Google Play Store, comparing machine learning and deep learning, is illustrated in Figure 1.



Source : (Research result, 2025)

Figure 1. Research Methods

Scrapping Data

Data scraping, commonly known as web scraping, involves extracting relevant data from vast information pools on the web. It is valuable across various fields, with distinct phases and procedures, while also facing technical, legal, and ethical challenges (Sabri, 2024).

Pre-Processing

Pre-processing of training data is crucial for ensuring proper formatting before training topic models. The chapter various tools and strategies necessary for effective pre-processing, highlighting its significant role in the overall success of topic modeling practices (P. Mishra, 2022).

This research employs several preprocessing techniques, including tokenization, stopword removal, stemming, and lemmatization, which are essential in refining text data for analysis. Additionally, Transform Cases is utilized to standardize text representation, significantly impacting classification performance. The choice of preprocessing strategy is tailored to the specific task and model, affecting outcomes in text classification tasks (Cascia, 2023).

Modelling

Machine learning algorithms are a subset of artificial intelligence that enable computers to learn from data, identify patterns, and make predictions or decisions. They are categorized into supervised, unsupervised, and reinforcement learning, with applications across various industries, including agriculture (A. S. and G. D. and A. J. and A. Mishra, 2024). This study specifically employs Naïve Bayes, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Decision Tree for classification tasks. These algorithms are selected based on their proven effectiveness in sentiment analysis, with Naïve Bayes offering probabilistic classification, Random Forest and Decision Tree providing robustness and interpretability, while SVM and Logistic Regression excel in handling high-dimensional data.

In addition to machine learning, this study also utilizes deep learning algorithms. Deep learning is a powerful machine learning technique with multiple neural network layers, excelling in various fields like cyber security, medical data analysis, automation, and linguistic processing (Vikkram, 2024). In this research, the Long Short-Term Memory (LSTM) network is implemented due to its effectiveness in processing sequential data and improving classification accuracy.

Evaluation and Accuracy

The paper discusses evaluating diagnostic tools using the ROC curve and AUC, emphasizing

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 22, No. 1 March 2025 | DOI https://doi.org/10.33480/techno.v22i1.6375

that AUC pseudovalues can be derived from jackknife sampling. It addresses potential inaccuracies due to reader interpretation and proposes tests to improve AUC estimate reliability (Su, 2024). Accuracy is the ratio of correctly identified cases to the total number of cases. The effectiveness of different machine learning methods was evaluated by comparing their accuracies using all features. The formula for accuracy calculation is as follows (Abbas Nawar Khalifa, Hadi Raheem Ali & Jebur, 2024):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Precision is the ratio of correctly predicted positive outcomes to all positive events (Abbas Nawar Khalifa, Hadi Raheem Ali & Jebur, 2024).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall indicates how well a model identifies positive instances, calculated as the ratio of correctly predicted positives to the total number of actual positives (Abbas Nawar Khalifa, Hadi Raheem Ali & Jebur, 2024).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

The precision and recall weighted average is known as the F1-Score (Abbas Nawar Khalifa, Hadi Raheem Ali & Jebur, 2024).

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

Visualization

During the visualization stage, a word cloud is utilized to display the most frequently occurring words in Google Play reviews, categorized by positive, negative, and neutral sentiments. WordCloud is a visualization technique used in the paper to present significant keywords from Google Custom Search results. It effectively displays keyword trends and topics, enhancing user understanding of search queries and improving content development strategies (Yusuf, 2024).

RESULTS AND DISCUSSION

The analysis showed that the sentiment distribution consisted of 45% negative, 35% neutral, and 20% positive reviews. Frequent issues mentioned were app crashes, delays in ticket activation, and a subpar user interface design. Comparisons with other studies underline similar problems faced by transportation apps, stressing the importance of ongoing improvements in user experience.

Data Collection

Data collection in this study involves two main stages: data scraping and cleaning. The first stage, scraping, is carried out by extracting data from Google Play Store reviews using the keyword 'Jaklingko App' through the Google Play Developer API scraping library. The collected data is then exported into the desired file format. The second stage, cleaning, is aimed at removing irrelevant elements from the data to ensure its relevance and quality. The final dataset consists of 21 positive reviews, 20 neutral reviews, and 159 negative reviews, categorized based on sentiment analysis. This distribution ensures a balanced representation for training and evaluation purposes.

Pre-Processing Process

Once the data is scraped from the Google Play store and cleaned, the next step is pre-processing. Since the data cannot be directly applied for sentiment analysis, it proceeds to the pre-processing stage.

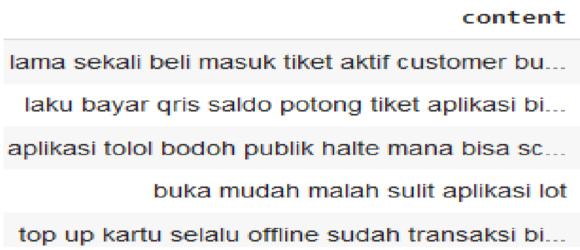
The operators involved in document processing include case transformation, lowercasing, tokenization, stopword removal, stemming, normalization, HTTP removal, and IndoBERT. The steps are as follows:

1. Transform Cases

In this study, the lowercase feature is used to convert any uppercase letters in the text into lowercase.

2. Lower Casing

The next process is lowercasing, which involves converting all characters in the text to lowercase. This step ensures consistency and eliminates the impact of case differences when analyzing the text.



content
lama sekali beli masuk tiket aktif customer bu...
laku bayar qris saldo potong tiket aplikasi bi...
aplikasi tolol bodoh publik halte mana bisa sc...
buka mudah malah sulit aplikasi lot
top up kartu selalu offline sudah transaksi bi...

Source : (Research result, 2025)

Figure 2. Lower Casing

3. Tokenization

Tokenization is the process of breaking down the text into individual words or tokens. This step helps in isolating the key elements of the text, making it easier to analyze and process for further tasks such as sentiment analysis.

	content
0	[lama, sekali, pembelian, masuk, tiket, aktifn...
1	[melakukan, pembayaran, qris, saldo, terpotong...
2	[aplikasi, tolol, pembodohan, publik, halte, m...
3	[bukanya, mempermudah, malah, mempersulit, apl...
5	[top, up, kartu, selalu, offline, sudah, trans...

Source : (Research result, 2025)
Figure 3. Tokenization

4. Stopword

The stopword stage is the process aimed at removing common words (such as "the," "and," "in") that don't contribute significantly to sentiment analysis. These words are often filtered out to focus on the more meaningful terms in the text.

	content	userName	repliedAt
0	lama sekali pembelian masuk tiket aktifnya cus...	Inki Riska	2024-11-11 21:56:55
1	melakukan pembayaran qris saldo terpotong tike...	Ardyan Suwanhara	2024-10-26 20:15:17
2	aplikasi tolol pembodohan publik halte manapun...	Pendi Rusadi	2024-10-25 22:05:46
3	bukanya mempermudah malah mempersulit aplikasi...	bagus satrio	2024-11-08 20:48:50
5	top up kartu selalu offline sudah transaksi bi...	Nomo (Nomo)	2024-11-02 21:16:44
...	...	...	...
192	aneh waktu buka qris jak lingko muncul peringa...	Kevin Marsion	2024-05-20 21:50:34
194	didownload berfungsi servernya bermasalah teru...	adit dit	2024-03-25 22:42:29
195	wah aplikasinya payah 2 hari tidak akses dikat...	syarifah aulia	2024-04-01 20:48:00
197	gua kira rute jaklingko nya pencarian suruh na...	Baim	2024-04-01 20:52:15
198	aplikasi sialan guna sama sekali buat pesen ti...	Merlinda Widianingrum	2024-04-01 20:52:56

Source : (Research result, 2025)
Figure 4. Stopword

5. Stemming

At this stage, stemming is performed, which involves reducing words to their root form (e.g., changing "running" to "run"). This helps in standardizing variations of words, making them easier to analyze in their basic form.

	Unnamed: 0	content
0	0	lama sekali beli masuk tiket aktif customer bu...
1	1	laku bayar qris saldo potong tiket aplikasi bi...
2	2	aplikasi tolol bodoh publik halte mana bisa sc...
3	3	buka mudah malah sulit aplikasi lot
4	5	top up kartu selalu offline sudah transaksi bi...

Source : (Research result, 2025)
Figure 5. Stemming

6. Normalization

The normalization stage involves standardizing the text by correcting spelling errors and converting abbreviations. This ensures that the text is consistent and that variations in spelling or shortened forms do not affect the analysis.

	content	userName	repliedAt
0	lama sekali dari pembelian ke masuk tiket akti...	Inki Riska	2024-11-11 21:56:55
1	sudah melakukan pembayaran dengan qris dan sal...	Ardyan Suwanhara	2024-10-26 20:15:17
2	aplikasi tolol pembodohan publik di halte mana...	Pendi Rusadi	2024-10-25 22:05:46
3	bukanya mempermudah malah mempersulit aplikasi...	bagus satrio	2024-11-08 20:48:50
5	top up kartu selalu offline ketika sudah trans...	Nomo (Nomo)	2024-11-02 21:16:44

Source : (Research result, 2025)
Figure 6. Normalization

7. HTTP Removal

The HTTP removal stage involves removing any URLs or web addresses found in the text. This step eliminates irrelevant information that may interfere with the analysis.

	content	userName	repliedAt
0	lama sekali pembelian masuk tiket aktifnya cus...	Inki Riska	2024-11-11 21:56:55
1	melakukan pembayaran qris saldo terpotong tike...	Ardyan Suwanhara	2024-10-26 20:15:17
2	aplikasi tolol pembodohan publik halte manapun...	Pendi Rusadi	2024-10-25 22:05:46
3	bukanya mempermudah malah mempersulit aplikasi...	bagus satrio	2024-11-08 20:48:50
5	top up kartu selalu offline sudah transaksi bi...	Nomo (Nomo)	2024-11-02 21:16:44

Source : (Research result, 2025)
Figure 7. HTTP Removal

8. IndoBert

The IndoBERT stage involves using the IndoBERT model for processing Indonesian language data in sentiment analysis. IndoBERT is a pre-trained language model specifically designed for Indonesian text, which helps in understanding the context and nuances of the language for more accurate sentiment analysis.

	content	Sentiment	Label
0	lama sekali beli masuk tiket aktif customer bu...	Negative	0
1	laku bayar qris saldo potong tiket aplikasi bi...	Neutral	1
2	aplikasi tolol bodoh publik halte mana bisa sc...	Positive	2
3	buka mudah malah sulit aplikasi lot	Negative	0
4	top up kartu selalu offline sudah transaksi bi...	Neutral	1

Source : (Research result, 2025)
Figure 8. IndoBert

Machine Learning

This study evaluates the effectiveness of five machine learning algorithms—Naive Bayes, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Decision Tree—for sentiment analysis of the Jaklingko App on the Google Play Store. The dataset was processed through text cleaning and feature extraction using TF-IDF. The experimental results indicate that Naive Bayes had the lowest accuracy at 50%, followed by Random Forest and SVM, each achieving 72% accuracy. Decision Tree and Logistic

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 22, No. 1 March 2025 | DOI https://doi.org/10.33480/techno.v22i1.6375

Regression demonstrated the highest accuracy at 78%. Although Naive Bayes is efficient in training, its performance is less effective compared to the more complex algorithms. Overall, Decision Tree and Logistic Regression proved to be the most effective for sentiment analysis, while Random Forest and SVM also delivered strong results on the relatively complex Google Play Store dataset.

Table 1. Naive Bayes

	Precision	Recall	F1-Score	Support
Negative	0.25	0.50	0.33	2
Neutral	0.78	0.54	0.64	13
Positive	0.20	0.33	0.25	3
Accuracy			0.50	18
Macro Avg	0.41	0.46	0.41	18
Weighted Avg	0.62	0.50	0.54	18

Source : (Research result, 2025)

Table 1 presents the performance metrics of the Naive Bayes algorithm in sentiment analysis. The model demonstrates the weakest performance in identifying neutral sentiments, with an F1 score of 0.64, and performs even worse in classifying negative sentiments, achieving an F1 score of only 0.33. Overall, the algorithm achieves an accuracy of 0.50, highlighting its limitations in effectively managing complex sentiments, particularly within the negative sentiment category.

Table 2. Random Forest

	Precision	Recall	F1-Score	Support
Negative	0.00	0.00	0.00	2
Neutral	0.72	1.00	0.84	13
Positive	0.00	0.00	0.00	3
Accuracy			0.72	18
Macro Avg	0.24	0.33	0.28	18
Weighted Avg	0.52	0.72	0.61	18

Source : (Research result, 2025)

Table 2 illustrates the performance of the Random Forest algorithm in sentiment analysis. The model performs relatively well in classifying neutral sentiments, achieving an F1 score of 0.84. However, it fails to classify both negative and positive sentiments, recording an F1 score of 0.00 across all metrics. With an overall accuracy of 0.72, these results highlight the limitations of the Random Forest algorithm in dealing with complex sentiments, particularly in the negative and positive categories.

Table 3. Support Vector Machine

	Precision	Recall	F1-Score	Support
Negative	1.00	0.50	0.67	2
Neutral	0.72	1.00	0.84	13
Positive	0.50	0.33	0.40	3
Accuracy			0.72	18
Macro Avg	0.24	0.33	0.28	18
Weighted Avg	0.52	0.72	0.61	18

Source : (Research result, 2025)

Table 3 illustrates the performance of the Support Vector Machine algorithm in sentiment analysis. The model performs relatively well in classifying neutral sentiments, achieving an F1 score of 0.84. However, it fails to classify both negative and positive sentiments, recording an F1 score of 0.00 across all metrics. With an overall accuracy of 0.72, these results highlight the limitations of the Support Vector Machine algorithm in dealing with complex sentiments, particularly in the negative and positive categories.

Table 4. Logistic Regression

	Precision	Recall	F1-Score	Support
Negative	1.00	0.50	0.67	2
Neutral	0.76	1.00	0.87	13
Positive	0.50	0.33	0.40	3
Accuracy			0.78	18
Macro Avg	0.59	0.50	0.51	18
Weighted Avg	0.66	0.78	0.70	18

Source : (Research result, 2025)

Table 4 highlights the performance of the Logistic Regression algorithm in sentiment analysis. The model demonstrates strong performance in classifying neutral sentiments, achieving an F1 score of 0.87. However, it struggles to accurately classify both negative and positive sentiments, with an F1 score of 0.00 in these categories. With an overall accuracy of 0.78, the results reveal the limitations of the Logistic Regression algorithm in addressing more complex sentiments, particularly in the negative and positive classes.

Table 5. Decision Tree

	Precision	Recall	F1-Score	Support
Negative	1.00	0.50	0.67	2
Neutral	0.80	0.92	0.86	13
Positive	0.50	0.33	0.40	3
Accuracy			0.78	18
Macro Avg	0.77	0.59	0.64	18
Weighted Avg	0.77	0.78	0.76	18

Source : (Research result, 2025)

Table 5 outlines the performance of the Decision Tree algorithm in sentiment analysis. The model performs well in identifying neutral sentiments, achieving an F1 score of 0.86, with high precision (0.80) and recall (0.92). In contrast, it performs moderately on negative sentiments, with an F1 score of 0.67, and struggles significantly with positive sentiments, achieving an F1 score of only 0.40 due to low recall (0.33). With an overall accuracy of 0.78, these results emphasize the algorithm's effectiveness in neutral sentiment classification but also highlight its challenges in accurately handling negative and positive sentiments.

Deep Learning

This study evaluates the application of the *Long Short-Term Memory* (LSTM) deep learning algorithm for sentiment analysis of the Jaklingko App reviews on the Google Play Store. The LSTM model is used to classify sentiments into positive, negative, and neutral categories after undergoing a data preprocessing stage. Experimental results show that LSTM achieved an accuracy of 72%, with strong performance in precision, recall, and F1-score metrics. The model has proven effective in handling temporal context and understanding word relationships in reviews, including more complex sentiments such as irony or sarcasm.

Table 6. Long Short-Term Memory

	Precision	Recall	F1-Score	Support
Negative	0.67	1.00	0.80	2
Neutral	0.83	0.77	0.80	13
Positive	0.33	0.33	0.33	3
Accuracy			0.72	18
Macro Avg	0.61	0.70	0.64	18
Weighted Avg	0.73	0.72	0.72	18

Source : (Research result, 2025)

Table 6 presents the performance of the Long Short-Term Memory algorithm in sentiment analysis. The model performs well in classifying negative sentiments, achieving a perfect recall of 1.00 and an F1 score of 0.80. Similarly, for neutral sentiments, it maintains a strong performance with an F1 score of 0.80, reflecting high precision (0.83) and recall (0.77). However, the model struggles significantly with positive sentiments, achieving an F1 score of just 0.33 due to low precision and recall. With an overall accuracy of 0.72, the results underscore the algorithm's strength in identifying negative and neutral sentiments while highlighting its challenges in handling positive sentiments effectively.

Evaluation

The evaluation results are summarized in a comparison table, as shown in Table 7, which provides an overview of the confusion matrix outcomes for each test scenario. This table facilitates a deeper analysis of the model's performance in classifying data, highlighting its strengths and weaknesses. By examining the value distribution, researchers can detect error patterns and identify areas requiring improvement, enabling further optimization of the model to achieve more accurate predictions in the future.

Tabel 7. Evaluation Matrix Machine Learning dan Deep Learning

Model	Accuracy	Precision	Recall	F-Measure
Naive Bayes	0.50	0.62	0.50	0.53
Random Forest	0.72	0.52	0.61	0.60
Support Vector Machine	0.72	0.52	0.72	0.60
Logistic Regression	0.78	0.78	0.94	0.92
Decision Tree	0.78	0.77	0.78	0.76
LSTM	0.72	0.73	0.72	0.72

Source : (Research result, 2025)

The evaluation results are summarized in a comparison table, as shown in Table 7, which provides an overview of the confusion matrix outcomes for each test scenario. This table facilitates a deeper analysis of the model's performance in classifying data, highlighting its strengths and weaknesses. By examining the value distribution, researchers can detect error patterns and identify areas requiring improvement, enabling further optimization of the model to achieve more accurate predictions in the future.

Tabel 8. The Accuracy Result of Machine Learning dan Deep Learning

Model	Accuracy
Naive Bayes	0.50
Random Forest	0.72
Support Vector Machine	0.72
Logistic Regression	0.78
Decision Tree	0.78
LSTM	0.72

Source : (Research result, 2025)

The results of this study are compared with previous research on sentiment analysis for transportation applications. For example, (Agoestanto, 2024) achieved an accuracy of 0.84 using a Naïve Bayes and Decision Tree model, while (Ladayya, Faroh and Siregar, Dania and Pranoto, Wiligis Eka and Muchtar, 2022) reported an accuracy of 0.70 with Support Vector Machine. Compared to these studies, our models, particularly Logistic Regression and Decision Tree, demonstrate competitive accuracy levels. However, the LSTM model in this study achieves an accuracy of only

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 22, No. 1 March 2025 | DOI https://doi.org/10.33480/techno.v22i1.6375

0.72, which is lower than previous deep learning approaches.
The variation in accuracy results may be attributed to several factors, including dataset size, preprocessing techniques, and feature selection methods. Previous studies often utilized larger and more balanced datasets, while this study relies on Google Play Store reviews, which may contain noisy or imbalanced data. Additionally, hyperparameter tuning and the use of advanced word embeddings (such as BERT or Word2Vec) in previous research may have contributed to higher accuracy levels.

Table 8 presents the accuracy levels of various machine learning and deep learning models used in sentiment analysis. Decision Tree and Logistic Regression demonstrate the highest accuracy at 0.78, followed by Random Forest and Support Vector Machine, both achieving an accuracy of 0.72. In contrast, Naive Bayes records the lowest accuracy at 0.50. These findings emphasize the superior performance of Decision Tree and Logistic Regression in sentiment analysis tasks.

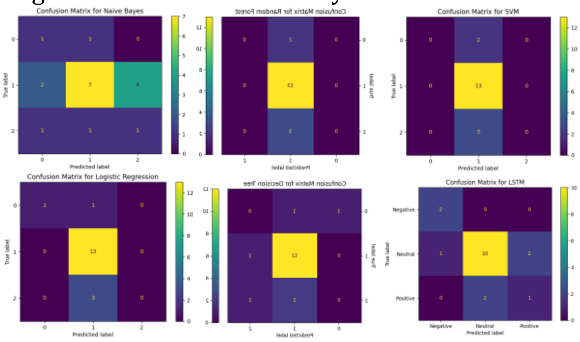
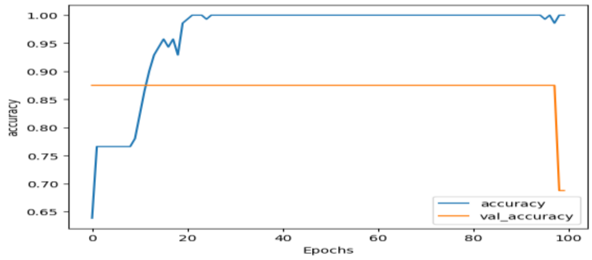


Figure 9. Confusion Matrix Results of Machine Learning and Deep Learning Models

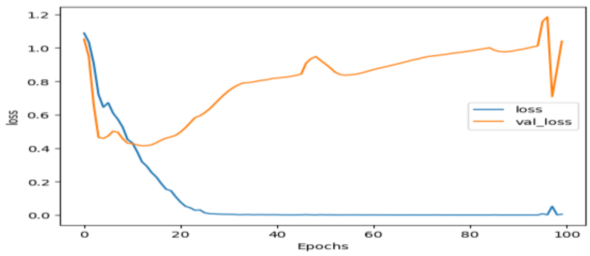
Figure 9 illustrates the confusion matrix results for six models: Naive Bayes, Random Forest, SVM, Logistic Regression, Decision Tree, and LSTM. The confusion matrix provides a detailed breakdown of classification performance by displaying the number of correctly and incorrectly predicted instances for each sentiment category (positive, neutral, and negative).

In this study, Decision Tree and LSTM exhibit the highest classification accuracy, with a balanced distribution of True Positives (TP) across all sentiment categories. In contrast, Naive Bayes struggles, particularly in the negative category, where it frequently misclassifies negative reviews as neutral or positive (high False Negative rate). The inclusion of clear labels in the confusion matrix—such as 'Actual Positive,' 'Predicted Positive,' 'Actual Negative,' and 'Predicted Negative'—enhances the interpretability of the results, making it easier to identify error patterns and areas for model improvement.

This visual representation facilitates a comprehensive evaluation of each model's classification accuracy, allowing for targeted enhancements in future work.



(a) Training Accuracy



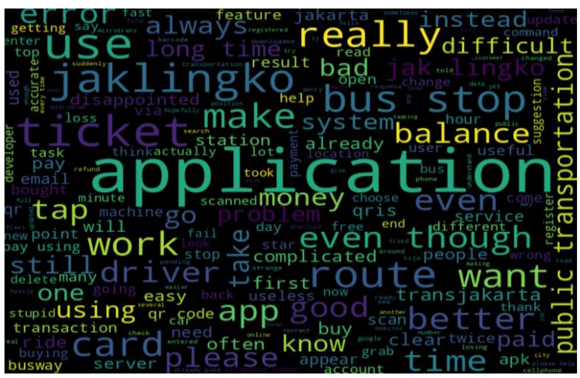
(b) Validation Accuracy

Source : (Research result, 2025)
Figure 10 Training dan Validation Accuracy Deep Learning LSTM

Figure 10 illustrates the training and validation accuracy of the LSTM model. The training accuracy remains stable, approaching 1, while the validation accuracy is consistent, indicating strong model performance without signs of overfitting.

Visualization

After the data was cleaned and classified, sentiment analysis results were visualized using a word cloud, highlighting the most frequently occurring words based on their frequency. Larger words indicate higher occurrences, making it easier to identify dominant keywords in the text data. This visualization was generated using a Python script, as shown in Figure 11.



Source : (Research result, 2025)
Figure 11 Wordcloud Visualization

CONCLUSION

The findings emphasize the urgency of improving JakLingko's functionality to resolve user dissatisfaction. Future studies could investigate the use of advanced natural language processing methods and larger datasets to enhance the accuracy of sentiment analysis. This study compares the performance of various machine learning and deep learning models in sentiment analysis of reviews for the JakLingko App on the Google Play Store. The results indicate that Decision Tree and Logistic Regression achieved the highest accuracy at 78%, followed by LSTM, Random Forest, and SVM at 72%, while Naive Bayes recorded an accuracy of 50% due to its limitations in handling complex sentiments. Decision Tree and Logistic Regression demonstrated superiority in capturing temporal context and word relationships, albeit requiring more computational resources. These findings confirm that machine learning models, particularly Decision Tree and Logistic Regression, are more effective in analyzing dynamic and diverse sentiments compared to deep learning models.

This research provides a foundation for developing more efficient models and applying sentiment analysis in various contexts, such as other social media platforms, to gain deeper and more accurate insights into public opinion. Future research is expected to achieve higher accuracy, and improvements to the JakLingko App itself are encouraged, particularly in terms of user interface, connectivity, and performance, to enhance user comfort. These improvements could contribute to generating more positive sentiment analysis outcomes.

REFERENCE

Abbas Nawar Khalifa, Hadi Raheem Ali, A. J., & Jebur, S. A. (2024). Automate Facial Paralysis Detection Using VGG Architectures. In *International Journal of Current Innovations in Advanced Research*. https://doi.org/10.47957/ijciar.v7i1.158

Adinegoro, K. R. R. (2022). Implementasi Sikap Kolaboratif dan Multikultural dalam Kepemimpinan pada Integrasi dan Penataan Transportasi Umum "JAK LINGKO" di DKI Jakarta. *Spirit Publik*, *17*(1). https://doi.org/10.20961/sp.v17i1.57666

Agoestanto, P. A. and A. (2024). Modeling of Naïve Bayes and Decision Tree Algorithms to Analyze Sentiment Related to Jaklingko Public Transportation on Social Media X (Twitter). *Pattimura Proceeding: Conference of Science and Technology*, *5*(1), 67–78. https://doi.org/10.30598/ppcst.knmxxiiv5i1p67-78

Bikbov, B., Purcell, C. A., Levey, A. S., Smith, M., Abdoli, A., Abebe, M., Adebayo, O., Afarideh, M., Agarwal, S. K., Agudelo-Botero, M., Ahmadian, E., Al-Aly, Z., Alipour, V., Almasi-Hashiani, A., Al-Raddadi, R., Alvis-Guzmán, N., Amini, S., Andrei, T., Andrei, C. L., … Murray, C. (2020). Global, Regional, and National Burden of Chronic Kidney Disease, 1990–2017: A Systematic Analysis for the Global Burden of Disease Study 2017. In *The Lancet*. https://doi.org/10.1016/s0140-6736(20)30045-3

Cascia, M. S. and I. T. and M. La. (2023). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*.

Gary Ekatama Bangun, M. S. (n.d.). Evaluasi kebijakan integrasi angkutan pengumpan ke dalam sistem bus rapid transit: Studi pada Mikrotrans Jaklingko. *2024*. https://doi.org/10.30738/sosio.v10i1.16268

Hanifa Aulia Rahma, Slamet Rosyadi, Guntur Gunarto, S. (2024). *Collaborative Governance in Management of the JakLingko Program (Case Study on Public Transportation Management in DKI Jakarta)*. https://doi.org/10.47134/jebmi.v2i1.139

Idris, Y. A. M. and I. S. K. (2024). Ensemble Approach to Sentiment Analysis of Google Play Store App Reviews. *Jambura Journal of Electrical and Electronics Engineering*, *6*(Universitas Negeri Gorontalo), 181–188. https://doi.org/10.37905/jjeee.v6i2.25184

Ladayya, Faroh and Siregar, Dania and Pranoto, Wiligis Eka and Muchtar, H. D. (2022). Analisis Sentimen pada Program Transportasi Publik JakLingko dengan Metode Support Vector Machine. *Jurnal Statistika Dan Aplikasinya*, *6*(2), 381–392.

Liebenlito, D. A. A. and T. E. S. and M. (2024). Confident Learning pada IndoBERT: Peningkatan Kinerja Klasifikasi Sentimen. *Indonesian Journal of Computer Science*, *13*(STMIK Indonesia Padang), 5. https://doi.org/10.33022/ijcs.v13i5.4401

Mishra, A. S. and G. D. and A. J. and A. (2024). Machine Learning Algorithms and Applications. *Auerbach Publications*, 1–31. https://doi.org/10.1201/9781003504900-1

Mishra, P. (2022). Bypassing NIR pre-processing optimization with multiblock pre-processing ensemble approaches. *Nir News*, *33*, 5–8. https://doi.org/10.1177/09603360221139227

Nugraheni, T. C. P. and W. W. and M. (2024).

Indonesian Fake News Classification Using Transfer Learning in CNN and LSTM. *JOIV : International Journal on Informatics Visualization*, *8*(State Polytechnics of Andalas), 1213–1213. https://doi.org/10.62527/joiv.8.2.2126

Putranto, L. S. (2023). Persepsi pengguna transportasi umum di jabodetabek terhadap integrasi tarif pt jaklingko indonesia. *Jurnal Mitra Teknik Sipil*, 71–84. https://doi.org/10.24912/jmts.v6i1.16430

Sabri, S. M. and M. M. and W. A. W. A. B. and M. K. Y. and I. A. A. (2024). *Demystified Overview of Data Scraping*. *6*, 290–296. https://doi.org/10.69511/ijdsaa.v6i6.205

Su, Y. H. and Y.-D. H. and N. (2024). Performance of tests based on the area under the ROC curve for multireader diagnostic data. *Journal of Applied Statistic*, *Taylor and Francis*, 1–23. https://doi.org/10.1080/02664763.2024.2374931

Suryawanshi, N. (2024). Sentiment analysis with machine learning and deep learning: A survey of techniques and applications. *International Journal of Science and Research Archive*, *12*(2), 005–015. https://doi.org/10.30574/ijsra.2024.12.2.1205

Vikkram, V. N. and S. L. and A. S. and R. (2024). Deep Learning. *Auerbach Publications*, *Auerbach Publ.*, 1–8. https://doi.org/10.1201/9781003433309-1

Yusuf, M. I. A. C. and R. (2024). Visualisasi kata kunci pemberitaan pemilu 2024 menggunakan scapy dan wordcloud. *Jurnal TEKNIMEDIA*, *5*(STMIK Syaikh Zainuddin NW Anjani-Lombok Timur), 41–46. https://doi.org/10.46764/teknimedia.v5i1.187