

User Reviews on the NU Online Application Using the Long Short-Term Memory (LSTM) Model

Atik Muthmainnah¹, Wahab Sya'roni², Ratri Enggar Pawening³

^{1,2,3}, Teknik Informatika, Universitas Nurul Jadid, Probolinggo, Indonesia

Article Info

Article history:

Received July 17, 2025

Revised Sept 19, 2025

Accepted Okt 2, 2025

Keywords:

Sentiment Analysis

NU Online

LSTM

Gradio

User Reviews

ABSTRACT

This study aims to analyze user sentiment toward the NU Online application by applying a Long Short-Term Memory (LSTM) model. The research was based on 13,576 user reviews collected from the Google Play Store, which underwent preprocessing, sentiment labeling, and Exploratory Data Analysis (EDA). To ensure balanced classification, undersampling was used, resulting in 6,612 reviews equally divided into positive and negative classes. The text data was processed using tokenization and padding before being input into the LSTM model. Model training involved the use of binary crossentropy, Adam optimizer, EarlyStopping, and ReduceLROnPlateau techniques. The model achieved 93% precision, recall, and F1-score, with low error rates and strong generalization ability. EDA results showed that positive feedback mainly focused on worship features like salat schedules, while negative reviews addressed technical issues such as the azan sound. User review peaks occurred during religious periods and major updates. A Gradio-based web interface was also developed to display results and enable user-friendly access to visual sentiment insights. This implementation proves the practical potential of integrating LSTM with an interactive platform for effective sentiment analysis

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Wahab Syaroni,

Universitas Nurul Jadid, Karanganyar Paiton, Probolinggo 67291, Indonesia

Email: wahab@unuja.ac.id

1. INTRODUCTION

NU Online is one of the most popular Islamic-based digital applications in Indonesia, with more than one million downloads on the Google Play Store by early 2025. This application provides various Islamic services such as a digital Quran, prayer schedules, Hijri calendar, daily prayers, zakat calculator, and other religious content [1]. However, NU Online's rating of only 3+ does not fully represent the level of user satisfaction accurately [2]. This is due to the discrepancy between the star rating and the content of the reviews. For example, there are five-star reviews that still contain criticism or suggestions for improvement, as well as one-star reviews that actually contain positive comments and only highlight minor aspects that need improvement [3]. This discrepancy makes the numerical rating less reliable as an indicator of user satisfaction, making it difficult for developers to determine priorities for feature improvements and application quality enhancements. Therefore, sentiment analysis of user reviews is an important step in understanding public opinion in a systematic, comprehensive, and objective manner.

Several previous studies have analyzed user sentiment toward religious applications using various machine learning approaches, but they still have limitations. Hairani and Nurpalah applied Naive Bayes to the Muslim Pro application with 90% accuracy, but it was less effective for long reviews because it did not consider word order [4]. accuracy, but it was less effective for long reviews because it did not consider word order [4]. Lusia et al. used Artificial Neural Networks (ANN) on NU Online, but had difficulty capturing the semantic relationships between words [5]. Meanwhile, the SVM used by Muzayyanah et al. achieved an accuracy of 95.46%, but ignored temporal order, making it less than optimal for modeling contextual meaning [6] These limitations emphasize the need for methods that are capable of maintaining context, understanding word order,

and capturing long-term patterns more accurately. Therefore, this study adopts the Long Short-Term Memory (LSTM) algorithm, a type of Recurrent Neural Network (RNN) designed to recognize patterns in sequential data. LSTM excels at overcoming context loss in long reviews, capturing semantic relationships between words, and is equipped with an internal memory cell mechanism that allows the model to remember important information during the training process [7] [8].

This study analyzed 13,576 user reviews of the NU Online application from the Google Play Store collected through scraping techniques between February 9, 2021 and January 21, 2025. The reviews underwent preprocessing and sentiment labeling, followed by exploratory analysis to identify major patterns in user opinions. For sentiment classification, this study employs a Long Short-Term Memory (LSTM) model, which is effective for processing sequential text data and capturing contextual relationships between words [9] [10]. Model performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score, to ensure a reliable representation of user sentiment.

To facilitate the interpretation of sentiment analysis results for non-technical users and developers, this study uses a web-based visual interface with Gradio [11], which allows rapid integration with machine learning models [12]. This interface displays sentiment classification results in the form of word clouds, bar charts, and pie charts based on positive, negative, and neutral categories, and provides a text input feature for direct sentiment prediction using the LSTM model. Therefore, this study aims to apply and analyze the Long Short-Term Memory (LSTM) method for sentiment classification of NU Online user reviews and to generate deeper insights such as identifying dominant positive and negative words and periods with the highest number of reviews, which can be used to understand user satisfaction more comprehensively and provide actionable recommendations for improving the application's service quality.

2. METHOD

The sentiment analysis process for NU Online application reviews on the Google Play Store was carried out through a number of stages, which are presented systematically in Figure 1.

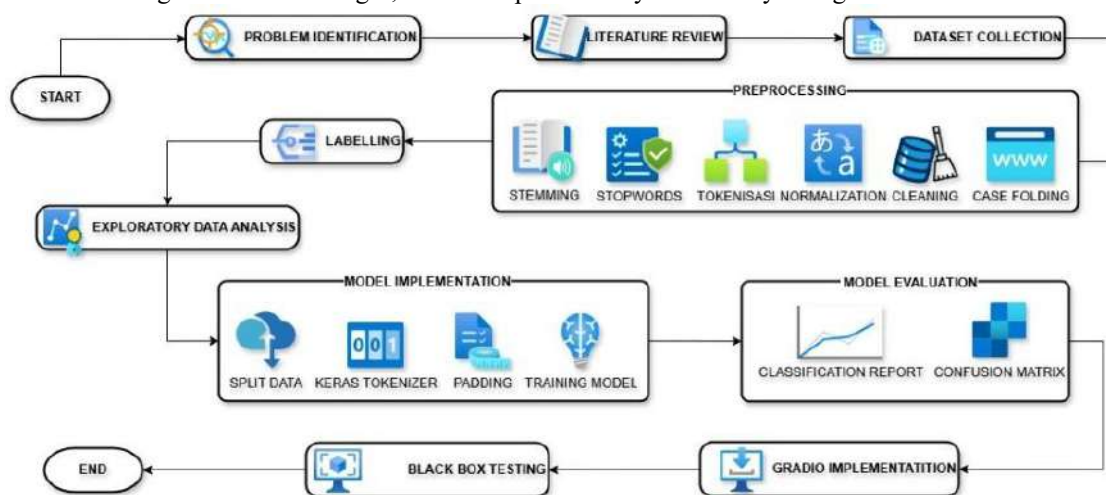


Figure 1. Research Flow of Sentiment Analysis on NU Online Reviews

2.1 Problem Identification

NU Online, a religious application with over one million downloads, faces the challenge of unstructured and biased user reviews. The complexity of opinions and data imbalance make sentiment analysis difficult. This study applies LSTM to improve classification accuracy and support application feature development. LSTM is chosen because it is capable of capturing word order, maintaining long-term context, and recognizing semantic relationships in lengthy user reviews, making it more suitable for understanding nuanced opinions compared to traditional machine learning methods [10] [13].

2.2 Literature Study

Methods such as Naive Bayes, SVM, and ANN have limitations in handling context and imbalanced data [14] [6] [5]. LSTM is superior in understanding word sequences and has been proven effective on various platforms [10] [9]. This study applies LSTM to analyze the sentiment of NU Online application users. Reviews on religious applications often contain loanwords, Arabic terms written in Indonesian, and nuanced expressions

that require a deep understanding of context. LSTM, with its ability to capture long-term dependencies and semantic relationships, is the right choice for analyzing these reviews more accurately.

2.3 Dataset Collection

The data in this study came from user reviews of the NU Online application on the Indonesian language Google Play Store. The collection was carried out using the Google Play Scraper library and stored in CSV format to facilitate the analysis process. The dataset consists of 13,576 records with 11 columns, including information such as review content, rating, upload time, and application version. The data range taken was between February 9, 2021 to January 21, 2025.

2.4 Preprocessing

Text preprocessing was performed to clean and standardize the review data, ensuring consistency in the dataset. The steps included case folding, where all text was converted to lowercase to maintain uniformity [15], cleaning to remove invalid characters such as numbers, symbols, punctuation, and emojis [16], and normalization using a combined dictionary from Kaggle and GitHub which is manually adjusted to correct typos or abbreviations [17]. Furthermore, tokenization is carried out to break sentences into tokens [18], stopwords removal to remove common words such as “yang”, “dan”, “atau” [19], Stemming with the Sastrawi library to change words to their basic form [20].

2.5 Labelling

Data labeling is done to automatically categorize reviews into positive or negative sentiment. This process uses the Indonesian sentiment dictionary from GitHub which contains a list of words with positive and negative nuances. The use of this dictionary facilitates classification and ensures consistency and quality of data for training sentiment analysis models.

2.6 Exploratory Data Analysis (EDA)

The Exploratory Data Analysis (EDA) stage is conducted to understand patterns in the NU Online application review dataset, such as dominant words, score distribution, and peak review periods. EDA helps reveal user satisfaction and is the basis for training LSTM-based sentiment models.

2.7 Model Application

The Long Short-Term Memory (LSTM) model was employed to perform sentiment analysis by partitioning the dataset into training and testing subsets. The textual data were transformed into numerical representations through tokenization, retaining the 5,000 most frequent words, [21] and subsequently standardized using sequence padding up to 100 tokens [22]. Model training was optimized using the binary cross-entropy loss function and the Adam optimizer [23][24], with the addition of EarlyStopping and ReduceLROnPlateau techniques to mitigate overfitting [25] [26]. This approach allows LSTM to recognize text context and improve classification accuracy. This methodological design enables the LSTM to capture contextual dependencies within the text and enhance the accuracy of sentiment classification. Mathematically, the internal operations of the LSTM are expressed as follows [13]:

$$f_t = \sigma (W_f [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma (W_i [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\hat{C}_t = \tanh(W_c [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = (f_t * C_{t-1} + i_t * \hat{C}_t) \quad (4)$$

$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

2.8 Model Evaluation

Performance evaluation is an important stage in machine learning to assess the extent to which the model is able to predict new data accurately and reliably [27]. This study uses the Classification Report and Confusion Matrix methods to measure the performance of the classification model. The Classification Report

presents metrics such as precision, recall, F1-score, and support for each class [28], where precision shows the proportion of correct positive predictions, recall measures the model's ability to detect all positive samples, and F1-score balances the two [29]. The Confusion Matrix provides a detailed overview of the number of correct and incorrect predictions by class, through components such as TP, TN, FP, and FN [30]. These metrics are particularly important because the dataset used in this study is imbalanced, making accuracy alone insufficient to represent performance comprehensively. Therefore, precision, recall, and F1-score complement accuracy by highlighting the model's performance on minority classes. From this matrix, accuracy is calculated using the formula [31]:

$$Accuracy = \frac{TP+TN}{Total\ Samples} \quad (7)$$

and the error rate as

$$Error\ Rate = 1 - Accuracy \quad (8)$$

These two methods complement each other, where the Classification Report is suitable for viewing overall performance, while the Confusion Matrix is useful for analyzing specific classification errors.

2.9 Gradio

Gradio is an open-source Python framework that makes it easy to create interactive interfaces for machine learning models such as LSTMs, without the need for complex server setup [12]. In this study, Gradio was used to display the results of sentiment analysis of the NU Online application. This framework supports platforms such as Colab and Hugging Face, and allows interactive visualizations such as bar charts and word clouds [11], thus accelerating model interpretation and collaboration [32]

2.10 Black Box Testing

Black Box Testing is a functional testing method that evaluates a system based on input and output without looking at the internal structure [33]. Its advantage lies in its ability to detect non-conformities to specifications, although it is limited because it does not access the source code [34]. In this study, this method was applied to a Gradio based user interface to ensure that each element functions as intended.

3. RESULTS AND DISCUSSION

3.1 Data Collection

Data collection was done using Google Play Scraper, a Python library to extract NU Online app user reviews from the Google Play Store. This process uses parameters such as the app ID id.or.nu.app, Indonesian language and country, and sorting by the most relevant reviews. The initial target was 15.000 reviews, but only 13.576 were available between February 9, 2021 and January 21, 2025. The data was processed using Pandas and Datetime for timekeeping, then filtered into three main columns: date (at), score (score), and review content (content), which were saved in CSV format for further analysis. The results of data collection can be seen in Table 1:

Table 1 NU Online Review Dataset

No	Date	Score	Content
1	2021-02-10 13:36:51	5	Alhadulillah, superapps NU Online sekarang sudah tersedia di google. apps. Tersedia banyak konten untuk memudahkan warga NU menjalankan amaliah seperti Qur'an. tahlil, wirid, dll. Monggo diunduh dan diberi masukan untuk perbaikan versi selanjutnya. Makasih Ada Hizib Jausan juga. Top kalo ditambahkan Dalailul Khoirot.
2	2023-03-07 2:01:37	1	Bagus lengkap sesuai amalan NU 👍 gk ada iklamn Tapiiiii Sayang seribu sayang Gak ada pilihan suara adzan. 😡😡 Bisa gak sih tiap mencet satu ayat ada pilihan lihat terjemahan. Jadi gk hrus lari ke pengaturan dulu yg saya kritik ama jawabannya gak nyambung blas boooooossss mbok di baca lagi pertanyaannya yg bener. terus dibenerin aplikasinya. jgn asal jawab tp gk nyambung
...	...	...	...

1357 6	2025-01- 18 3:44:23	5	the most beautiful application!! 🍀🍀 sukaa bangett sama apk ini lengkap dan membantuuu banget.., makasii buat developer yg udah buat apk sebgas dan semanfaat iniii sukaaa banget semoga sukses slalu yaaa..
-----------	---------------------------	---	---

3.2 Prerprocessing

This study conducted data preprocessing using NLTK and Sastrawi, starting by removing 2.845 duplicate and empty data from the content column, resulting in 10.731 reviews ready for analysis.

3.2.1 Case Folding

In the initial stage, all review texts are converted to lowercase using the lower() function so that the writing format is more uniform and consistent, thus facilitating the subsequent analysis process. The results of Case Folding can be seen in Table 2.

Table 2 Preprocessing Result – Case Folding

No	Date	Case Folding
1	2021-02-10 13:36:51	alhadulillah, superapps nu online sekarang sudah tersedia di google. apps. tersedia banyak konten untuk memudahkan warga nu menjalankan amaliah seperti qur'an. tahlil, wirid, dll. monggo diunduh dan diberi masukan untuk perbaikan versi selanjutnya. makasih ada hizib jausan juga. top kalo ditambahkan dalailul khoiroh.
2	2023-03-07 2:01:37	bagus lengkap sesuai amalan nu 🙌 gk ada iklann tapiiiii sayang seribu sayang gak ada pilihan suara adzan. 🙄🙄 bisa gak sih tiap mencet satu ayat ada pilihan lihat terjemahan. jadi gk hrus lari ke pengaturan dulu yg saya kritik ama jawabannya gak nyambung blas boooosssss mbok di baca lagi pertanyaannya yg bener. terus dibenerin aplikasinya. jgn asal jawab tp gk nyambung
...	...	...
1073 1	2025-01-18 3:44:23	the most beautiful application!! 🍀🍀 sukaa bangett sama apk ini lengkap dan membantuuu banget.., makasii buat developer yg udah buat apk sebgas dan semanfaat iniii sukaaa banget semoga sukses slalu yaaa..

3.2.2 Cleaning

Next, text cleaning is performed by removing irrelevant characters such as punctuation, numbers, emojis, and special symbols. This process aims to produce cleaner, more structured, and ready-to-process data. The cleaning results can be seen in Table 3:

Table 3 Preprocessing Result – Cleaning

No	Date	Cleaning
1	2021-02-10 13:36:51	alhadulillah superapps nu online sekarang sudah tersedia di google apps tersedia banyak konten untuk memudahkan warga nu menjalankan amaliah seperti quran tahlil wirid dll monggo diunduh dan diberi masukan untuk perbaikan versi selanjutnya makasih ada hizib jausan juga top kalo ditambahkan dalailul khoiroh
2	2023-03-07 2:01:37	bagus lengkap sesuai amalan nu gk ada iklann tapiiiii sayang seribu sayang gak ada pilihan suara adzan bisa gak sih tiap mencet satu ayat ada pilihan lihat terjemahan jadi gk hrus lari ke pengaturan dulu yg saya kritik ama jawabannya gak nyambung blas boooosssss mbok di baca lagi pertanyaannya yg bener terus dibenerin aplikasinya jgn asal jawab tp gk nyambung
...	...	...
1073 1	2025-01-18 3:44:23	the most beautiful application sukaa bangett sama apk ini lengkap dan membantuuu banget makasii buat developer yg udah buat apk sebgas dan semanfaat iniii sukaaa banget semoga sukses slalu yaaa

3.2.3 Normalization

After the text is cleaned, the normalization stage is carried out to change non-standard words or abbreviations into formal forms. This process uses manual normalization dictionaries from various sources, such as changing “apk” to “aplikasi” or “tdk” to “tidak”. The results of normalization can be seen in Table 4:

Table 4 Preprocessing Result – Normalization

No	Date	Normalization
1	2021-02-10 13:36:51	alhamdulillah superapps nu online sekarang sudah tersedia di google aplikasi tersedia banyak konten untuk memudahkan warga nu menjalankan amaliah seperti quran tahlil wirid dan lain lain monggo diunduh dan diberi masukan untuk perbaikan versi selanjutnya terima kasih ada hizib jausan juga top kalau ditambahkan dalailul khoirot
2	2023-03-07 2:01:37	bagus lengkap sesuai amalan nu tidak ada iklan tapi sayang seribu sayang tidak ada pilihan suara azan bisa tidak sih tiap mencet satu ayat ada pilihan lihat terjemahan jadi tidak harus lari ke pengaturan dulu yang saya kritik sama jawabannya tidak menyambung blas boooooossss tolong di baca lagi pertanyaannya yang benar terus dibenerin aplikasinya jangan asal jawab tapi tidak menyambung
...	...	...
1073 1	2025-01-18 3:44:23	the paling beautiful application suka banget sama aplikasi ini lengkap dan membantu banget makasi buat developer yang sudah buat aplikasi sebgus dan semanfaat ini suka banget semoga sukses selalu iya

3.2.4 Tokenization

Then, the normalized text is broken down into word units (tokens) to make it more structured. For example, the sentence [“Saya”, “sangat”, “menyukai”, “aplikasi”, “ini”]. The Tokenization Results can be seen in Table 5:

Table 5 Preprocessing Result – Tokenization

No	Date	Tokenization
1	2021-02-10 13:36:51	['alhadulillah', 'superapps', 'nu', 'online', 'sekarang', 'sudah', 'tersedia', 'di', 'google', 'aplikasi', 'tersedia', 'banyak', 'konten', 'untuk', 'memudahkan', 'warga', 'nu', 'menjalankan', 'amaliah', 'seperti', 'quran', 'tahlil', 'wirid', 'dan', 'lain', 'lain', 'monggo', 'diunduh', 'dan', 'diberi', 'masukan', 'untuk', 'perbaikan', 'versi', 'selanjutnya', 'terima', 'kasih', 'ada', 'hizib', 'jausan', 'juga', 'top', 'kalau', 'ditambahkan', 'dalailul', 'khoirot']
2	2023-03-07 2:01:37	['bagus', 'lengkap', 'sesuai', 'amalan', 'nu', 'tidak', 'ada', 'iklan', 'tapi', 'sayang', 'seribu', 'sayang', 'tidak', 'ada', 'pilihan', 'suara', 'azan', 'bisa', 'tidak', 'sih', 'tiap', 'mencet', 'satu', 'ayat', 'ada', 'pilihan', 'lihat', 'terjemahan', 'jadi', 'tidak', 'harus', 'lari', 'ke', 'pengaturan', 'dulu', 'yang', 'saya', 'kritik', 'sama', 'jawabannya', 'tidak', 'menyambung', 'blas', 'boooooossss', 'tolong', 'di', 'baca', 'lagi', 'pertanyaannya', 'yang', 'benar', 'terus', 'dibenerin', 'aplikasinya', 'jangan', 'asal', 'jawab', 'tapi', 'tidak', 'menyambung']
...	...	...
1073 1	2025-01-18 3:44:23	['the', 'beautiful', 'application', 'suka', 'banget', 'aplikasi', 'lengkap', 'membantu', 'banget', 'makasi', 'developer', 'aplikasi', 'sebagus', 'semanfaat', 'suka', 'banget', 'semoga', 'sukses', 'iya']

3.2.5 Stopwords Removal

The next step is the removal of stopwords, which are common words such as “dan”, “yang”, “atau”, and “dari” which are considered not to contribute significant information in the analysis. By removing these words, the text becomes more concise and meaningful. The Stopword results can be seen in Table 6:

Table 6 Preprocessing Result – Stopwords Removal

No	Date	Stopwords Removal
1	2021-02-10 13:36:51	['alhadulillah', 'superapps', 'nu', 'online', 'tersedia', 'google', 'aplikasi', 'tersedia', 'konten', 'memudahkan', 'warga', 'nu', 'menjalankan', 'amaliah', 'quran', 'tahlil', 'wirid', 'monggo', 'diunduh', 'masukan', 'perbaikan', 'versi', 'terima', 'kasih', 'hizib', 'jausan', 'top', 'dalailul', 'khoirot']

2	2023-03-07 2:01:37	['bagus', 'lengkap', 'sesuai', 'amalan', 'nu', 'iklan', 'sayang', 'seribu', 'sayang', 'pilihan', 'suara', 'azan', 'sih', 'mencet', 'ayat', 'pilihan', 'lihat', 'terjemahan', 'lari', 'pengaturan', 'kritik', 'jawabannya', 'menyambung', 'blas', 'boooooossss', 'tolong', 'baca', 'pertanyaannya', 'dibenerin', 'aplikasinya', 'menyambung']
...	...	...
1073 1	2025-01-18 3:44:23	['the', 'beautiful', 'application', 'suka', 'banget', 'aplikasi', 'lengkap', 'membantu', 'banget', 'makasi', 'developer', 'aplikasi', 'sebagus', 'semanfaat', 'suka', 'banget', 'semoga', 'sukses', 'iya']

3.2.6 Stemming

Finally, stemming is performed to change words into their basic form using the Sastrawi library. For example, the word “menjalankan” is changed to “jalan”, and “terbaik” to “baik”. This process aims to reduce word variations and increase analysis efficiency. The stemming results can be seen in Table 7:

Table 7 Preprocessing Result – Stemming

No	Date	Stemming
1	2021-02-10 13:36:51	alhadulillah superapps nu online sedia google aplikasi sedia konten mudah warga nu jalan amaliah quran tahlil wirid monggo unduh masuk baik versi terima kasih hizib jausan top dalailul khoiro
2	2023-03-07 2:01:37	bagus lengkap sesuai amal nu iklan sayang ribu sayang pilih suara azan sih mencet ayat pilih lihat terjemah lari atur kritik jawab sambung blas boooooossss tolong baca tanya dibenerin aplikasi sambung
...	...	...
1073 1	2025-01-18 3:44:23	the beautiful application suka banget aplikasi lengkap bantu banget makas developer aplikasi bagus manfaat suka banget moga sukses iya

3.3 Labelling

Before the labeling process, 401 duplicate data and 203 empty rows were checked and removed in the stemming column, leaving 8.936 reviews to be analyzed. Sentiment labeling was carried out using a lexicon-based method using an Indonesian sentiment dictionary containing 3.609 positive words and 6.609 negative words, each with a value weight between -5 and +5. These values were added up in each review to determine sentiment; if the result is >0 then it is labeled positive, <0 negative, and $=0$ is considered neutral although only two categories, namely positive and negative, are used in this analysis stage. The labeling process is shown in Table 8, while the distribution of labeling results based on visualization can be seen in Figure 2.

Table 8 Labelling Result

No	Date	Stemming	Word Scores	Score Total	Label
1	2021-02-10 13:36:51	alhadulillah superapps nu online sedia google aplikasi sedia konten mudah warga nu jalan amaliah quran tahlil wirid monggo unduh masuk baik versi terima kasih hizib jausan top dalailul khoiro	['sedia (+2.0)', 'aplikasi (-4.0)', 'sedia (+2.0)', 'mudah (-1.0)', 'quran (+3.0)', 'wirid (+4.0)', 'monggo (+3.0)', 'masuk (-3.0)', 'baik (-1.0)', 'terima (+2.0)', 'kasih (-1.0)', 'top (+5.0)']	11	Positive
2	2023-03-07 2:01:37	bagus lengkap sesuai amal nu iklan sayang ribu sayang pilih suara azan sih mencet ayat pilih lihat terjemah lari atur kritik jawab sambung blas boooooossss tolong baca tanya dibenerin aplikasi sambung	['bagus (-4.0)', 'lengkap (-3.0)', 'sesuai (+3.0)', 'amal (+4.0)', 'sayang (-3.0)', 'sayang (-3.0)', 'pilih (-2.0)', 'suara (-1.0)', 'ayat (-4.0)', 'pilih (-2.0)', 'lihat (+3.0)', 'lari (-2.0)', 'atur (-4.0)', 'kritik (-5.0)', 'jawab (-5.0)', 'sambung (-3.0)', 'tolong (-2.0)', 'baca (-1.0)', 'tanya (-2.0)', 'aplikasi (-4.0)', 'sambung (-3.0)']	-43	Negative
...	...	...	...	...	...
893 6	2025-01-18	the beautiful application suka banget aplikasi lengkap bantu	['suka (-1.0)', 'banget (+1.0)', 'aplikasi (-4.0)', 'lengkap (-3.0)',	-9	Negative

3:44:23	banget makas developer aplikasi bagus manfaat suka banget moga sukses iya	'bantu (-4.0)', 'banget (+1.0)', 'aplikasi (-4.0)', 'bagus (-4.0)', 'manfaat (+5.0)', 'suka (-1.0)', 'banget (+1.0)', 'moga (+1.0)', 'sukses (+3.0)']
---------	---	---

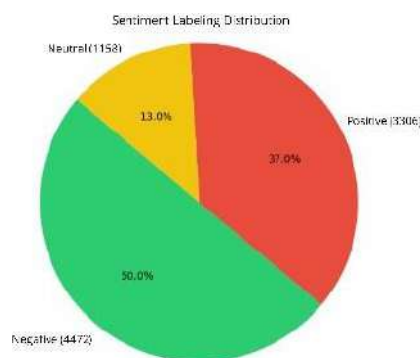


Figure 2 Sentiment Labeling Distribution

Of the total data, 4,472 reviews (50%) were negative, 3,306 reviews (37%) were positive, and 1,158 reviews (13%) were neutral however, neutral data was not used in the further analysis stage because it did not show a clear trend of opinion.

3.4 Exploratory Data Analysis (EDA)

Further analysis was conducted to gain insights from the LSTM-based sentiment classification results on NU Online application reviews. The main focus includes identifying the most frequently occurring words in each sentiment category, as well as the period with the highest number of reviews. The purpose of this analysis is to understand the tendency of user opinions and provide relevant input in application development. Visualization of dominant words from the classification results is shown in Figure 3 for positive sentiment and Figure 4 for negative sentiment. In the positive category, words such as add, steady, and pray reflect satisfaction with the worship features, especially the prayer schedule feature. For example, one user wrote: "nyaman pakai aplikasi ini terutama jadwal salat" (August 10, 2023). Conversely, in negative sentiment, words such as complete, azan, and read often appear as complaints, especially regarding the sound of the call to prayer that does not work optimally. This is reinforced by a user review on October 24, 2023: "tolong benerin suara azan soalnya suka henti pas notifikasi, lihat notifikasi pas besar suara kecil, ya henti, tolong benerin, dengar azan selesai enak dengar, terima kasih ya oke." Analysis of these words provides an overview of the aspects of the application that get the most attention from users, both positively and negatively.

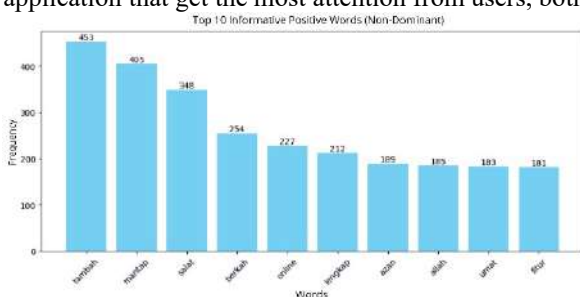


Figure 3 Top Positive Words(Non Dominant)

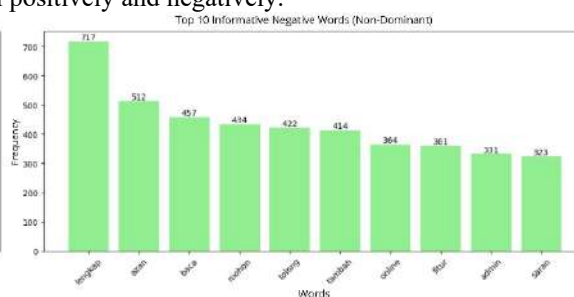


Figure 4 Top Negative Words(Non Dominant)

In addition, an analysis of the highest review period is shown in Figure 5. The graph shows a seasonal pattern from 2021 to early 2025, with the peak of reviews occurring in February and March especially around the month of Ramadan and a decline in the number of reviews in the middle of the year. A spike was also recorded in October, along with the commemoration of National Santri Day. The most reviews were recorded in February 2021 during the initial launch of the application, while the peak of negative reviews occurred in February 2022, coinciding with the introduction of new features. This pattern indicates that users are more active in providing reviews when there is an application update or coincides with important religious moments.



Figure 5 Monthly Distribution of User Reviews by Sentiment Class

3.5 Model Application

The LSTM model is built using data that has gone through the tokenization and padding process so that each input has a uniform length. The model architecture and parameters are summarized in Figure 6, which shows the layer structure, number of neurons, and total parameters trained. The training process is carried out for a maximum of 10 epochs with 10% of the training data allocated as validation data. To avoid overfitting and improve training efficiency, two callbacks are used, namely EarlyStopping to stop training when the validation loss does not improve, and ReduceLROnPlateau to adjust the learning rate dynamically. The training results show that the highest validation accuracy of 94.90% is achieved at the 9th epoch, and training is automatically stopped at the 10th epoch. The accuracy and loss graphs shown in Figure 7 and Figure 8 show a stable trend: the accuracy continues to increase, while the loss value decreases, indicating that the model has successfully learned the data pattern well without symptoms of overfitting, and has strong generalization ability to new data.

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 100, 128)	640,000
dropout_6 (Dropout)	(None, 100, 128)	0
lstm_4 (LSTM)	(None, 100, 128)	131,584
dropout_7 (Dropout)	(None, 100, 128)	0
lstm_5 (LSTM)	(None, 64)	49,408
dense_4 (Dense)	(None, 64)	4,160
dropout_8 (Dropout)	(None, 64)	0
dense_5 (Dense)	(None, 1)	65

Total params: 825,217 (3.15 MB)
Trainable params: 825,217 (3.15 MB)
Non-trainable params: 0 (0.00 B)

Figure 6 LSTM Model Architecture Summary

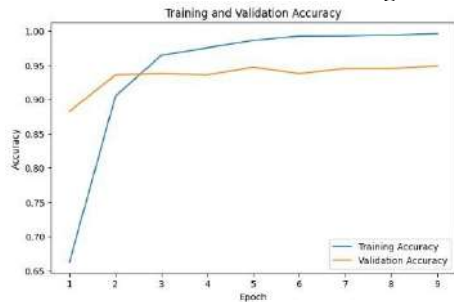


Figure 7 Training and Validation Accuracy Curve

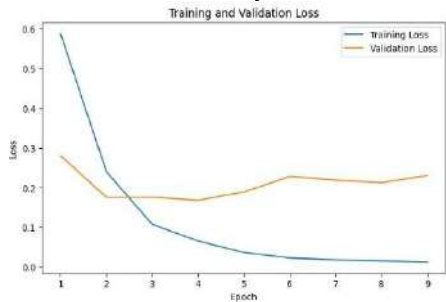


Figure 8 Training and Validation Loss Curve

3.6 Model Evaluation

Evaluation of the LSTM model was conducted to assess its ability to classify the sentiment of NU Online application user reviews. The evaluation results are shown in Figure 9 (Classification Report) and Figure 10 (Confusion Matrix). Based on the Classification Report, the model shows balanced performance in both classes (positive and negative), with precision, recall, and F1-score values of 0.93 each. The Confusion Matrix shows that 626 negative data and 605 positive data were correctly classified, while 44 negative data and 48 positive data were misclassified. This distribution shows that the model is not biased towards one class and is able to learn proportionally after undersampling techniques are performed.

Based on the evaluation metrics, the model achieved a precision of 93.22%, recall of 92.65%, F1-score of 92.93%, and accuracy of 93.04%, with an error rate of 6.96%. These values indicate that the developed

LSTM model has high accuracy, low error, and good generalization ability to new data. Thus, this model is effective in recognizing user opinions in a balanced manner and can be a reliable tool in sentiment analysis of the NU Online application.

Classification Report:				
	precision	recall	f1-score	support
Negatif	0.93	0.93	0.93	670
Positif	0.93	0.93	0.93	653
accuracy			0.93	1323
macro avg	0.93	0.93	0.93	1323
weighted avg	0.93	0.93	0.93	1323

Figure 9 Classification Report Results

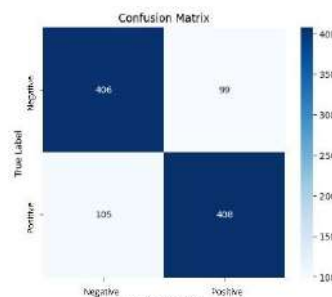


Figure 10 Confusion Matrix Results

3.7 Gradio Implementation

As part of the development of the sentiment analysis system, the author built a web interface using Gradio to facilitate access to the analysis results, especially for non-technical users. Users can upload a CSV file containing the published and comment columns, then the system automatically displays the sentiment classification results in the form of a table containing the date, comment content, sentiment label, and prediction accuracy value. The results can be downloaded again for further analysis. In addition, the system also provides interactive visualizations such as pie charts for sentiment distribution, bar charts for the most words, and line charts for review trends based on time, both overall and per sentiment class. The application has been hosted on Hugging Face Spaces and can be accessed via the link:

<https://huggingface.co/spaces/atikmtmainah/analisisnuonline>

3.8 Black Box Testing

System testing was conducted using the Black Box method by providing input according to the scenario and matching the output results with the expected output. The test results are shown in Table 9. All features show results according to the scenario and are declared acceptable without the need for improvement, which proves that the system is running well, stable, and easy to use. The Gradio based interface has proven to be efficient in processing data automatically, making it easier for users to perform sentiment analysis without having to understand complex technical aspects.

Table 9 Black Box Testing Results

Features/Functions Tested	Test Scenario	Expected results	Test Results (Match/Non-Match)	Conclusion	
				Accepted	Repair
Upload File CSV	Upload a CSV file with Published and Comment columns	The file was successfully uploaded and is displayed in the upload column.	Match	✓	
Invalid File Upload	Upload a file other than CSV or CSV without required columns, then click "Start Analysis"	An error message appears and the file is not received	Match	✓	
"Start Sentiment Analysis" button	Press the button after uploading a valid file	The analysis process runs and the results are displayed.	Match	✓	
All Reviews	Select the "All Reviews" tab once	Displays table of all reviews, bar chart, pie	Match	✓	

	the analysis is complete.	chart, line chart, sentiment pie, wordcloud		
Positive Reviews	Select the “Positive Sentiment” tab after analysis	Displaying tables, bar charts, line charts, and word clouds of positive reviews.	Match	✓
Negative Reviews	Select the “Negative Sentiment” tab after analysis	Displaying tables, bar charts, line charts, and word clouds of negative reviews.	Match	✓
Single Review	Fill in the single review column and click the “Analyze Now” button.	The analysis results (Positive/Negative) appear in the results column.	Match	✓
Download All Results	Click the download button to download all analysis results	CSV/ZIP file successfully downloaded according to analysis result data	Match	✓
Download Positive/Negative Results	Click the download button based on a specific sentiment	CSV/ZIP file successfully downloaded according to filter (positive/negative)	Match	✓
Click Analyze without Uploading File	Pressing the “Start Analysis” button without uploading a file	An error and warning appears that the file has not been uploaded.	Match	✓

4. CONCLUSION

This study demonstrates that the Long Short-Term Memory (LSTM) algorithm is capable of accurately classifying sentiment in user reviews of the NU Online application, achieving precision, recall, and F1-score values of 93%. The analysis also revealed that negative reviews are generally related to technical issues such as the azan sound feature, while positive responses are often directed at worship-related features like the prayer schedule. Review spikes were observed during app launches and religious periods, indicating strong user engagement. Additionally, the Gradio based interface successfully enhances the accessibility of sentiment analysis results for non-technical users and shows potential for broader application in other sectors.

For future development, it is recommended to adopt transformer-based labeling methods such as IndoBERT to improve contextual understanding. The dataset should also be expanded to include reviews from other platforms to capture more diverse user opinions. Exploring newer deep learning architectures like Transformer is also encouraged for improved performance. On the application side, NU Online developers are advised to focus on enhancing features that received negative feedback and to implement regular feedback monitoring to ensure continuous service improvement.

ACKNOWLEDGEMENTS

The author wishes to express deep gratitude to all parties who provided support, direction, and assistance throughout the writing of this research journal. Special thanks go to the academic advisor for their valuable guidance and insights. The author also extends heartfelt appreciation to their family and peers for their unwavering encouragement during the completion of this work.

REFERENCES

- [1] M. Khamim, “NU Online SuperApp and Strengthening Religious Moderation of Nahdlatul Ulama Based on Literacy in the New Media Era 4.0,” *JNUS: Journal of Nahdlatul Ulama Studies*, vol. 3, no. 2, pp. 92–113, Jul. 2022, doi: 10.35672/jnus.v3i2.92-113.
- [2] M. K. Khoirul Insan, U. Hayati, and O. Nurdyawan, “ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 478–483, Mar. 2023, doi: 10.36040/jati.v7i1.6373.

- [3] M. Musfiroh, A. Tholib, and Z. Arifin, "Analisis Sentimen Terhadap Ulasan Aplikasi Shopee di Google Play Store Menggunakan Metode TF-IDF dan Long Short-Term Memory)," *Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 2, pp. 371–381, Oct. 2024, doi: 10.33650/jeeecom.v6i2.8713.
- [4] E. Hairani and I. Nurpalah, "Sentiment Analysis of Muslim Pro 'A1-Quran Adzan' on Google Play Using Multinomial Naïve Bayes," in *2024 12th International Conference on Cyber and IT Service Management (CITSM)*, IEEE, Oct. 2024, pp. 1–5. doi: 10.1109/CITSM64103.2024.10775862.
- [5] D. A. Lusia, G. Anuraga, and F. Rahman, "Sentiment Analysis of NU Online Applications Using Artificial Neural Network," *Southeast Asian Journal of Islamic Education*, vol. 06, no. 02, pp. 197–208, 2024, doi: 10.21093/sajie.v6i2.8822.
- [6] A. B. Muzayyanah, R. E. Pawening, and Z. Arifin, "ANALISIS SENTIMEN PADA ULASAN APLIKASI EHADRAH DI GOOGLE PLAYSTORE MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)," *IDEALIS : InDonEsiA journal Information System*, vol. 7, no. 2, pp. 258–266, Jul. 2024, doi: 10.36080/idealis.v7i2.3250.
- [7] M. Rizki, S. Basuki, and Y. Azhar, "Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory Untuk Prediksi Curah Hujan Kota Malang," *REPOSITOR*, vol. 2, no. 3, pp. 331–338, 2020.
- [8] M. Khadapi and V. Maruli Pakpahan, "Analisis Sentimen Berbasis Jaringan LSTM dan BERT terhadap Diskusi Twitter tentang Pemilu 2024," *JUKI: Jurnal Komputer dan Informatika*, vol. 6(2), p. 130, Nov. 2024.
- [9] B. Wicaksono, V. Rahmayanti, and S. Nastiti, "Analisis Sentimen dalam Opini Publik di Chanel Youtube Indonesia Lawyers Club Tentang Isu Populer dengan Menggunakan Metode LSTM dan Bi-LSTM," *Jurnal Algoritma*, vol. 22, pp. 241–251, Dec. 2024, doi: 10.33364/algoritma/v.21-2.1696.
- [10] D. Wahyuni, N. Fadhillah, and W. widya Ariestyas, "Long Short-Term Memory dan Lexicon Based Untuk Analisis Sentimen Ulasan Aplikasi TikTok," *Jurnal Ilmiah Komputasi*, vol. 23, no. 2, pp. 173–180, Jun. 2024, doi: 10.32409/jikstik.23.2.3579.
- [11] A. Sawant, S. Phadol, S. Mehre, P. Divekar, S. Khedekar, and R. Dound, "ChatWhiz: Chatbot Built Using Transformers and Gradio Framework," in *Proceedings - 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2024*, Tirunelveli India: Institute of Electrical and Electronics Engineers Inc., May 2024, pp. 454–461. doi: 10.1109/ICICV62344.2024.00077.
- [12] R. Ferreira, M. Canesche, P. Jamieson, O. P. V. Neto, and J. A. M. Nacif, "Examples and tutorials on using Google Colab and Gradio to create online interactive student-learning modules," *Computer Applications in Engineering Education*, vol. 32, no. 4, pp. 1–17, Feb. 2024, doi: 10.1002/cae.22729.
- [13] A. Khumaidi and I. A. Nirmala, *Algoritma Long Short Term Memory dengan Hyperparameter Tuning: Prediksi Penjualan Produk*. Sleman: Deepublish, 2022.
- [14] L. Yusuf and S. Masripah, "SENTIMEN ANALISIS CHATGPT DENGAN ALGORITMA NAÏVE BAYES DAN OPTIMASI PSO," *INTI Nusa Mandiri*, vol. 18, no. 1, pp. 59–64, Aug. 2023, doi: 10.33480/inti.v18i1.4230.
- [15] J. Supriyanto, D. Alita, and A. R. Isnain, "Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 74–80, Mar. 2023, doi: 10.33365/jatika.v4i1.2468.
- [16] W. Ningsih, Rahmadden, S. Adrianto, F. Kurniawan, and B. Alfiana, "Analisis Sentimen Pengguna Aplikasi Gen Ai dari Appstore dan Googleplay Menggunakan Algoritma C45," *The Indonesian Journal of Computer Science*, vol. 13, no. 5, Sep. 2024, doi: 10.33022/ijcs.v13i5.4327.
- [17] H. Hidayatullah and Y. Umaidah, "PENERAPAN NAÏVE BAYES DENGAN OPTIMASI INFORMATION GAIN DAN SMOTE UNTUK ANALISIS SENTIMEN PENGGUNA APLIKASI CHATGPT," 2023.
- [18] M. Ilmar Rifaldi, Y. Raymond Ramadhan, and I. Jaelani, "Analisis Sentimen Terhadap Aplikasi Chatgpt Pada Twitter Menggunakan Algoritma Naïve Bayes," 2023.
- [19] R. Zulfiqri, B. N. Sari, and T. N. Padilah, "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MEDIA SOSIAL INSTAGRAM PADA SITUS GOOGLE PLAY STORE MENGGUNAKAN NAÏVE BAYES CLASSIFIER," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4995.
- [20] M. R. Firdaus, N. Rahaningsih, and R. D. Dana, "Analisis Sentimen Aplikasi Shopee di Goole Play Store Menggunakan Klasifikasi Algoritma Naïve Bayes," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, pp. 228–237, Mar. 2024.
- [21] M. Y. Baihaqi, E. Halawa, R. A. S. Syah, A. Nurrahma, and W. Wijaya, "Emotion Classification in Indonesian Language: A CNN Approach with Hyperband Tuning," *Jurnal Buana Informatika*, vol. 14, no. 02, pp. 137–146, Oct. 2023, doi: 10.24002/jbi.v14i02.7558.

- [22] M. R. Faisal *et al.*, “A Social Community Sensor for Natural Disaster Monitoring in Indonesia Using Hybrid 2D CNN LSTM,” in *Proceedings of the 8th International Conference on Sustainable Information Engineering and Technology*, New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 250–258. doi: 10.1145/3626641.3626932.
- [23] R. Waddar, V. Rathod, N. H. S. Chikkamath, N. S. R., and S. V Budihal, “A CNN-based Stutter Detection Using MFCC Features with Binary Cross-Entropy Loss Function,” *IEEE International Conference on Contemporary Computing and Communications (InC4)*, pp. 1–6, Mar. 2024, doi: 10.1109/InC460750.2024.10649122.
- [24] M. Fathur, R. Khairul, R. Setya Perdana, and P. Korespondensi, “ARSITEKTUR SISTEM PERCAKAPAN OTOMATIS BERBAHASA INDONESIA DENGAN NORMALISASI BAHASA INFORMAL MENJADI BAKU,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 11, pp. 1009–1016, Oct. 2024, doi: 10.25126/jtiik.2024117984.
- [25] R. Kurniawan, I. Ramadhan, and R. Hartono, “IMPLEMENTASI SISTEM QUESTION ANSWERING MENGGUNAKAN METODE LONG SHORT TERM MEMORY (LSTM) PADA STUDI KASUS BAHASA SUNDA,” Aug. 2024.
- [26] T. R. Mahesh *et al.*, “Transformative Breast Cancer Diagnosis using CNNs with Optimized ReduceLRonPlateau and Early Stopping Enhancements,” *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Dec. 2024, doi: 10.1007/s44196-023-00397-1.
- [27] Z. A. Dwiyantri and C. Prianto, “Prediksi Cuaca Kota Jakarta Menggunakan Metode Random Forest,” *Jurnal Tekno Insentif*, vol. 17, no. 2, pp. 127–137, Oct. 2023, doi: 10.36787/jti.v17i2.1136.
- [28] R. B. Widodo, *MACHINE LEARNING METODE k-NEAREST NEIGHBORS - KLASIFIKASI ANGKA BAHASA ISYARAT*. Malang: Media Nusa Creative, 2022.
- [29] A. B. A. Aulia and Alamsyah, “Peningkatan Hiperparameter Framework Deep Learning VGG-16 untuk Pendeteksian Tumor Otak pada Teknologi MRI,” *Indonesian Journal of Mathematics and Natural Sciences*, vol. 47, no. 2, pp. 99–107, Oct. 2024, [Online]. Available: <https://journal.unnes.ac.id/journals/JM/index>
- [30] R. Safitri, I. Ali, and N. Rahaningsih, “ANALISIS SENTIMEN TERHADAP TREN FASHION DI MEDIA SOSIAL DENGAN METODE SUPPORT VECTOR MACHINE (SVM),” 2024.
- [31] M. Haris, A. Suharso, and E. Haodudin Nurkifli, “ANALISIS SENTIMEN PADA GAME EFOOTBALL DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA INDOBERT,” 2024.
- [32] E. Nandhini and G. Vadivu, “Convolutional Neural Network-Based Multi-Fruit Classification and Quality Grading with a Gradio Interface,” in *Proceedings of the 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2024*, Chennai, India: Institute of Electrical and Electronics Engineers Inc., Dec. 2024, pp. 1–17. doi: 10.1109/ICSES63760.2024.10910350.
- [33] M. Ar Rachman *et al.*, “PENERAPAN BLACK BOX TESTING UNTUK EVALUASI FUNGSIONALITAS WEBSITE MAGGOPLAST,” *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 1, 2025, doi: <https://orcid.org/0000-0002-4061-9241>.
- [34] A. C. Praniffa, A. Syahri, F. Sandes, U. Fariha, and Q. A. Giansyah, “PENGUJIAN BLACK BOX DAN WHITE BOX SISTEM INFORMASI PARKIR BERBASIS WEB BLACK BOX AND WHITE BOX TESTING OF WEB-BASED PARKING INFORMATION SYSTEM,” *Jurnal Testing dan Implementasi Sistem Informasi*, vol. 1, no. 1, pp. 1–16, 2023.