DECODE: JURNAL PENDIDIKAN TEKNOLOGI INFORMASI
P-ISSN: 2775-2984 | E-ISSN: 2775-1813

Vol. 5, No 3, November 2025, Hal: 941-952
Doi: http://dx.doi.org/10.51454/decode.v5i3.1257
https://journal.umkendari.ac.id/index.php/decode

Sentiment Analysis of User Reviews on Maxim Application Using the Long Short-Term Memory (LSTM) Methods

Maria Ilona Junide Bria[1], Anik Vega Vitianingsih[1*], Anastasia Lidya Maukar[2], SY. Yuliani[3], Pamudi[1]

[1]Department Informatic Engineering, Universitas Dr. Soetomo, Indonesia.
[2]Department Industrial Engineering, Universitas Presiden, Indonesia.
[3]Department Informatic, Universitas Multimedia Nusantara Jakarta, Indonesia

Artikel Info

Keywords:

Machine Learning;
Maxim Apps;
Natural Language Processing;
Sentiment Analysis;
Service Quality.

Article History :

Submitted: 17 July, 2025
Accepted: 23 October, 2025
Published: 27 October, 2025

Abstrak: The technological developments have encouraged the emergence of app-based transportation services that are increasingly popular with the public, one of which is the Maxim app. Despite offering convenience in booking transportation and other services, this app still receives various reviews from users regarding service quality. User feedback is provided through the Maxim app review section available on the Google Play Store platform. Sentiment analysis is applied in this study to identify shortcomings in the Maxim app, to help developers improve service quality and understand user satisfaction. The research procedure it comprises several phases, including data collection, text preprocessing, determining sentiment labels, assigning weights to terms, and a classification process using the Long Short-Term Memory (LSTM) algorithm. This studi unlike previous studies that commonly used classical machine learning techniques including Naïve Bayes and SVM, or BiLSTM, this research applies an LSTM model with lexicon-based sentiment labeling to improve consistency and contextual understanding in sentiment classification. A confusion matrix was utilized to evaluate the model's performance. Overall, 1,200 user reviews were gathered through web scraping techniques from June 2024 to June 2025. The sentiment classification process uses a lexicon-based method to categorize user reviews grouped into three sentiment classes: positive, neutral, and negative. The findings suggest that 762 reviews are labelled as positive, 157 as neutral, and 281 as negative. The LSTM method testing demonstrated excellent performance, achieved 95.21% accuracy, 97.22% precision, 84.02% recall, and an F1-score of 88.84%.

Corresponding Author:
Anik Vega Vitianingsih
Email: vega@unitomo.ac.id

INTRODUCTION

The development of technology has made public transportation an essential aspect of daily life. These services are now easily accessible through mobile applications connected to the internet, enabling users to order transportation conveniently via the Google Play Store (Iqrom et al., 2025). One such platform is Maxim, an application-based transportation service operating in Indonesia since 2018 under PT Teknologi Perdana Indonesia. Currently available in over 200 cities, Maxim provides two-wheeled and four-wheeled transport, as well as goods, cargo, and food delivery services at competitive prices, attracting a broad user base.

Maxim application provides convenience for users in accessing transportation services. However, behind this convenience, various responses in the form of criticism and suggestions indicate that service quality still requires further enhancement to align with user expectations (Damayanti et al., 2024). Some common complaints include features that do not run optimally, inaccuracies in the pickup and delivery system, and slow application performance. The review section on the Google Play Store serves as a medium for users to convey various opinions, ranging from praise to suggestions and complaints (M. Saputra & Wahyuni, 2024). The total count of reviews available presents a significant challenge in processing and categorizing data. These problems also affect the application's assessment and impact the level of user trust. Sentiment analysis is a relevant approach in exploring user opinions more systematically (Putri et al., 2024). This technique allows comments to be grouped under positive, negative, and neutral sentiment labels. The outcome of this analysis can serve as evaluation material for developers to improve service quality (Gaafar et al., 2022). Previous studies on the Maxim application generally applied classical machine learning techniques such as Naïve Bayes and SVM, or BiLSTM. This research differs by applying the LSTM algorithm with lexicon-based sentiment labeling, which enables more consistent classification and deeper contextual understanding of textual data (Kokab et al., 2022).

Several previous studies have explored sentiment analysis of customer feedback on the Maxim application, which is listed on the Google Play Store. Researchers (Wewengkang et al., 2025) applied the Bidirectional LSTM (BiLSTM) technique in conjunction with FastText for word vector representation, by utilizing data of 10,000 user reviews obtained through a web scraping process from the Google Play Store, which showed excellent model performance, achieved an accuracy level of 94%, precision of 96%, recall of 95%, and an F1 score of 95%. Research (Rizki & Sanjaya, 2024) evaluated user satisfaction with the Maxim app in Palembang using the UEQ method on 206 respondents. Results showed positive scores for attractiveness (1.04), clarity (1.42), and efficiency (1.03), while novelty (0.21) and stimulation (0.86) were lower, suggesting a need for greater innovation to boost user engagement.

Meanwhile, research utilizing the Support Vector Machine (SVM) method on 1,200 review data from the Google Play Store yielded an accuracy result of 79%. Researchers (Widyaningrum & Kamayani, 2023) conducted sentiment examination of user sentiments regarding the Maxim service through the utilization of the Naïve Bayes classification algorithm. The study utilized 1,170 tweets collected from Twitter, which were labeled into positive and negative sentiments using an Indonesian lexicon-based approach. The model reached an overall accuracy of 82.5%. For negative sentiments it scored in the case of positive sentiments, the model reached a precision rate of 80%, a recall of 90%, and an F1-score of 85%, while overall performance reached a precision of 87%, a recall of 74%, and an F1-score of 80% for positive sentiment classification. The results showed that public perception of Maxim was predominantly positive, indicating a generally favorable response toward the service. Another study (Pohan et al., 2024) employed the Random Forest method for sentiment classification on 2000 review data from the Google Play Store and Twitter, achieving an accuracy result of 63.6%. Afterward, the dataset was partitioned into two subsets training and testing using a 1175 to 825 ratio.

Meanwhile, researchers (Hasanah & Sari, 2024) employed the Naïve Bayes classification approach was utilized in conjunction with TF-IDF as the feature extraction technique weighting, utilizing 1,000 review data points. The evaluation indicated that the developed model was able to group user reviews effectively, achieving 84%, a precision of 83%, a recall of 93%, and an F1-score of 88%. The studies demonstrate that classification techniques, techniques like the Naïve Bayes algorithm, SVM, Random Forest, and BiLSTM demonstrate differences their performance in analyzing user sentiments toward the Maxim application, with the combination of BiLSTM and FastText yielding the most optimal results. However, there has been no research that explicitly employs the LSTM (Long Short-Term Memory) approach, without a two-sided direction (BiLSTM), to extract important features from review texts, as well as the use of a lexicon-based approach in labelling the Maxim application reviews data.

The objective of this research is to evaluate users' opinions of the Maxim apps by categorizing reviews into categories of positive, negative, and neutral sentiments. This LSTM method is employed due to its ability to handle long text data, maintain word context, and produce accurate sentiment classification (Labuguen, 2025). In addition, sentiment labelling utilizes the Lexicon-Based method, which automatically groups data based on sentiment polarity (positive, negative, or neutral), enabling large-scale data analysis (Aripiyanto et al., 2022). The data was obtained by automatically fetching user feedback data collected from the Google Play Store. Following this, the dataset underwent pre-processing and sentiment labelling using the Lexicon-Based approach.

Furthermore, the data was processed using the LSTM method for sentiment classification. Performance evaluation was the evaluation was the evaluation was performed with a confusion matrix was employed to measure the accuracy, precision, recall, and F1-score obtained by the model applied model (Shyahrin et al., 2023). This research responds to the importance of understanding user perceptions of the quality of application services. Previous studies have shown that user reviews significantly contribute to improving the quality of service. The research emphasizes the need for a more in-depth sentiment analysis approach to help developers improve the quality and level concerning users’ satisfaction with the Maxim app.

This outcomes of this research are expected to provide constructive contributions by enhancing developers' awareness of user sentiment, thus offering meaningful perspectives. These insights assist developers in identifying user reactions to the application, thereby providing valuable input. These details helps developers understand how users interact with the app and find aspects that need improvement, and describe the advantages and disadvantages of the services offered. This improvement in service quality has a direct impact on customer satisfaction and a considerable effect on the organization's overall reputation.

METHOD

To enhance the quality of application-based services, this research examines user feedback on the Maxim application with the objective of achieving optimal classification accuracy by implementing the LSTM model combined with Dictionary-Based Labeling. Figure 1 illustrates the sequence of stages conducted in this study. The process starts with data collection using web scraping, followed by several preprocessing steps, including data cleaning, case folding, tokenization, stop word removal, and stemming. Subsequently, sentiment labeling is carried out using a lexicon-based approach to classify user reviews into three sentiment categories: positive, neutral, and negative. To handle class imbalance among the sentiment categories, the SMOTE technique is employed prior to the weighting phase. The next stage involves applying the TF-IDF method for word weighting, followed by the classification process using the LSTM model to identify textual patterns. The dataset is then partitioned into 80% training data and 20% testing data to evaluate performance through a confusion matrix that measures accuracy, precision, recall, and F1-score.

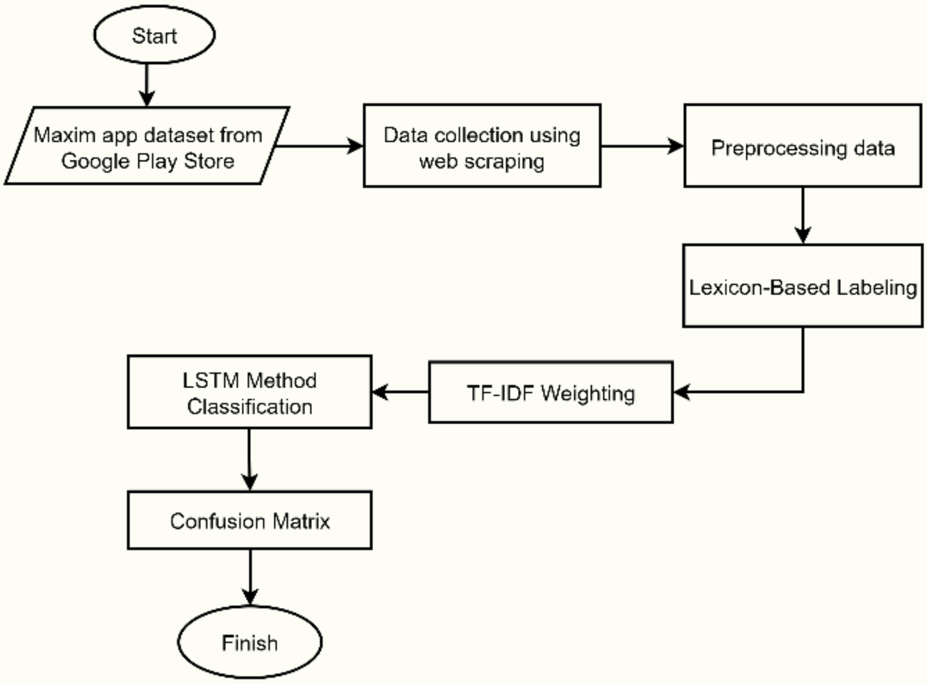

Figure 1. Research Phase

Data Collection

The process from collecting Maxim app review data available through the Google Play Store is carried out by applying web scraping methods supported by the Python programming language. This process utilizes the available Application Programming Interface (API) and focuses on relevant categories. The dataset obtained is raw data that has not been processed, with an estimated data collection timeframe from June 2024 to June 2025.

Data Pre-processing

Data preprocessing seeks to refine and simplify the data, maintain language consistency, and remove elements that are not relevant to the research needs (Husein et al., 2025). Pre-processing refers to a method employed to transform unstructured data into meaningful data through a sequence of steps, such as text case folding, cleaning, stopword removal, stemming, and tokenization (Hakim, 2021).

1. Cleaning: Data cleaning is a vital step in data processing, focusing on detecting and correcting, or eliminating, inaccurate, damaged, redundant, or irrelevant entries (Hakim, 2021).
2. Case Folding: A stage in text processing where all characters are transformed into lowercase to maintain consistency and avoid differences resulting from the use of upper and lowercase letters (Hakim, 2021).
3. Tokenization: A process in NLP that breaks divided into smaller units known as tokens, which may consist of words, phrases, symbols, or characters. This process is crucial in helping computers comprehend the structure, meaning, and context of sentences more effectively in text analysis (Mandar et al., 2020).
4. Stopword Removal: This stage is crucial for enhancing data quality by removing irrelevant words, allowing algorithms to focus more on key elements. By removing words such as conjunctions or common words that do not convey additional meaning, the data becomes more efficient, making it easier to recognize patterns in text (Mandar et al., 2020).
5. Stemming: Stemming is a technique for removing prefixes, suffixes, or infixes from words, regardless of their meaning. Although the results are not always consistent with language rules, this technique is effective in reducing word variations, thereby making text analysis more efficient (Aulia et al., 2021).

Data Labeling

The processed text will be labelled according to the set of words determined in the lexicon, based on existing words (Wirayani et al., 2025). Each word in the text will be matched with the words in the lexicon, then the text will be categorized into positive sentiment (+1), negative (-1), or neutral (0), after matching the word with the lexicon, the system will calculate the sentiment score of each matching word, then add up the scores to get the total sentiment of the text, the calculation is done using Equation (1).

$$Sentimen\ (S^i) = \sum_{i=1}^{n} Sentimen(wi) \tag{1}$$

The lexicon labelling approach plays a crucial role because it offers several advantages. Since this approach doesn't rely on training data, it becomes highly beneficial in situations where labelled data is scarce, unavailable, or limited in quantity. By utilizing the available sentiment dictionary, sentiment analysis can be carried out directly without the need for time-consuming manual labelling (A. Saputra et al., 2024). This approach also provides a high level of interpretation and clarity, as every sentiment score the label given to every a word or sentence can be directly mapped to its respective entry in the lexicon database, which allows researchers to perform a more accurate sentiment interpretation to easily interpret and present the analytical results outcomes in greater depth (Otter et al., 2021). Equation (2) is used to compute the TF-IDF weight. In this context, $tf_{t,d}$ denotes how often the term t occurs within the symbol d represents an individual document, whereas N indicates the total count of documents included in the corpus, whereas idf_t indicates how many documents contain the term t.

TF-IDF Weighting

The Term Frequency–Inverse Document Frequency (TF-IDF) model serves as a statistical approach to evaluate the relevance of a word in a specific document by comparing its occurrence within that document to its frequency in the overall dataset. assigning it a specific weight context (Sujadi, 2022). It integrates two core components in its calculation: term frequency represents how often a word occurs in a particular document, while inverse document frequency measures how rare or unique that word is across the entire collection of documents (IDF) weight, which indicates how rare or typical the word is among the entire set of documents. In other words, the rate at which a term appears in a single document serves as an indicator of its significance within that particular document.

$$W_{t,d} = tf_{t,d} \,.\, log\ (\frac{N}{df_t}) \tag{2}$$

Classification Model LSTM

The classification model in this research was carried out using the LSTM method. LSTM represents a more sophisticated version of the Recurrent Neural Network (RNN) architecture, created to be more efficient in processing long sequential data (DiPietro & Hager, 2020). One of the notable benefits of LSTM is that it can analyze data even when there is a significant time gap between elements in the sequence (Selle et al., 2022). This is possible because LSTM consists which consists of a forget gate, an input gate, an output gate, and a memory cell, which function to process and produce output values as hidden layers for the next processing stage (Alghifari et al., 2022). This process begins with forgetting gate stage, which determines whether previous information will be retained or ignored previous information will be retained or deleted. This process is shown through Equation (3).

$$ft = \sigma(Wf \times [xt + ht - 1] + bf) \tag{3}$$

Subsequently, at this gate, previous output is combined with the current input and then processed through two activation functions: a sigmoid for input values and a tanh for memory cell candidates. This process uses Equations (4) and (5).

$$it = \sigma(Wi \times [xt + ht - 1] + bi) \tag{4}$$

$$C\ t = \tanh\ (WC\ \times [xt + ht - 1]\ \ + bC) \quad (5)$$

Next, at this stage, two values are combined, namely the forget gate value multiplied by the previous cell stage and the input gate value multiplied by the memory cell candidate. This process uses Equation (6).

$$Ct = ft\ \times Ct - 1 + it\ \times C\ t \quad (6)$$

Then, this gate produces an output obtained from combining the previous value and the current value that has been processed through the sigmoid activity function. This calculation uses Equation (7).

$$ot = \ \sigma(Wo\ \times [xt\ + ht - 1]\ \ + bo) \quad (7)$$

That last process involves hidden layer, which is influential in the further processing of values. The hidden layer value is generated by multiplying the output by the cell stage or memory cell that has been activated, using the tanh function. This process uses Equation (8).

$$ht = ot\ \times \tanh\ (Ct) \quad (8)$$

Once the values of the output and hidden states have been obtained, sentiment classification is performed using the sigmoid or softmax activation function to generate probabilities according to the sentiment categories.

Evaluation Model

The evaluation model in this study utilizes the Confusion Matrix, a classification model in machine learning that compares actual values with model prediction results (Valero-Carreras et al., 2023). This model is employed to calculate evaluation metrics for example performance metrics including metrics like accuracy, precision, and recall (Tamami et al., 2025). Meanwhile, the F1 Score offers a summary that reflects the equilibrium between combining precision and recall into a single metric (Putri et al., 2025). In this research, model performance evaluation is a crucial aspect, so the confusion matrix is used as an analysis tool. Accuracy is calculated based on the ratio produce accurate classifications for both positive and negative cases sentiment when compared to overall dataset, as described in Equation (9). Precision is described as the ratio of true positive predictions to the total number of instances classified as positive, as shown in Equation (10). Meanwhile, recall measures the extent to which actual positive events were correctly predicted compared to the total number of correct predictions for positive and negative events, as calculated through Equation (11).

$$accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (9)$$

$$precision = TP/(TP + FP) \quad (10)$$

$$recall = TP/(TP + FM) \quad (11)$$

A True Positive (TP) denotes the number of instances belonging to the positive class that are correctly classified as positive. Conversely, a True Negative (TN) indicates the number of negative instances accurately categorized as negative. A False Positive (FP) arises when samples from the negative class are misclassified as positive, whereas a False Negative (FN) occurs when positive samples are incorrectly labeled as negative.

RESULTS AND DISCUSSIONS

The data for this study was collected by web scraping feedback from users of the Maxim app available on the Google Play Store. This dataset collected includes user reviews from June 2024 to June 2025, comprising approximately 1,200 review data points. The scraping results are stored in a CSV file format and are shown in Table 1.

Table 1. Web Scraping Results

Score	At	Content
2	12/24/2024	*Aplikasi ojol termurah, tp sayang susah bgt cari titik di maps nya. Dari dulu kaya gini mulu*
3	12/22/2024	*agar di perbaiki mengenai mapsnya kurang jelas pada saat di baca*
5	11/25/2024	*aplikasi sangat bagus dan murahhh, selalu pakai maxim untuk Pengiriman barang. sangat cepat, dan respon sangat baik. sukses selalu!*
1	10/29/2024	*Login susah lebih dr 3x malah ga bisa. Aplikasi ga bagus log in aja dipersulit. Mendingan si ijo dr pda ini*
5	6/1/2024	*Aplikasinya bagus, harga murah dan juga petanya juga oke. Memuaskan sih pake maxim*
1	5/16/2025	*aplikasi yang sangat ngebug, kurang dari maps nya, pelayanan nya chatting nya sangan buruk*
2	5/17/2025	*Apk nya jelek, setelah nitikin lokasi gada peta yang muncul, pengisian dana ke dompet kaspro nya lemot!*

After data collection is finalized, the data cannot be analyzed directly because it still contains a significant amount of noise. Therefore, Data preprocessing represents a crucial phase in the data mining workflow. Its primary objective is to convert raw data into a more organized and structured form that is ready for analysis, while eliminating irrelevant attributes. In the next stage, several pre-processing steps were performed, including data case folding, cleaning, tokenization, stemming, and removal of stop words. The outcomes of the preprocessing steps are presented that Table 2.

Table 2. Pre-processing Results

Raw Dataset	
Aplikasi nya sangat bagus 👍 👍 tampilan nya keren dan mudah di mengerti 👍 biaya ongkos perjalanan dengan pesan makanan nya juga murah banget di banding ojek online lain nya, pokok nya mantap deh 👌 👌 😉	
Text praproses	
Cleaning	*Aplikasi nya sangat bagus tampilan nya keren dan mudah di mengerti biaya ongkos perjalanan dengan pesan makanan nya juga murah banget di banding ojek online lain nya pokok nya mantap deh*
Case Folding	*aplikasi nya sangat bagus tampilan nya keren dan mudah di mengerti biaya ongkos perjalanan dengan pesan makanan nya juga murah banget di banding ojek online lain nya pokok nya mantap deh*
Tokenisasi	*Aplikasi, nya, sangat, bagus, tampilan, nya, keren, dan, mudah, di, mengerti, biaya, ongkos, perjalanan, dengan, pesan, makanan, nya, juga, murah, banget, di, banding, ojek, online, lain, nya, pokok, nya, mantap, deh*
Stopword Removal	*aplikasi, nya, sangat, bagus, tampilan, nya, keren, mudah, mengerti, biaya, ongkos, perjalanan, pesan, makanan, nya, murah, banget, banding, ojek, online, nya, pokok, nya, mantap, deh*
Stemming	*Aplikasi, nya, sangat, bagus, tampil, nya, keren, mudah, ngerti, biaya, ongkos, jalan, pesan, makan, nya, murah, banget, banding, ojek, online, nya, pokok, nya, mantap, deh*

The data that has undergone the initial pre-processing stage is then labelled using a lexicon-based approach with a predefined sentiment dictionary. The labelling process involves calculating the sentiment score of each review based on the occurrence of words included in a list of words grouped into categories of neutral, positive, and negative sentiment. The labeling outcomes indicate that 762 reviews fell into the positive sentiment category, 157 reviews into the neutral category, and 281 reviews into the negative sentiment category. The results of the labelling process are shown in Table 3.

Table 3. Lexicon-Based Labeling Results

Clean_text	Score	Sentiment
thanks maxim aplikasi yg murah dan terbaik antar penumpang sesuai dan harga terjangkau	2	Positive
bagus sekali apk nya sangat bermanfaat dan membantu tapi ada kekurangan nya di inbox nya gk ada kirim foto	0	Neutral
aplikasi ga jelas setiap mau order harus adu jotos dulu sama driver nya karena titik ga sesuai suka minta cancel susah masuk	-4	Negative

Once the sentiment labelling process is complete, the review data is then grouped by sentiment category to identify the words that users use most frequently. The frequency of these words is then visualized in a word cloud. Every review contains a unique combination of words and classified as neutral, positive, or negative. "*maxim*", "*pengemudi* (driver)", "*aplikasi* (app)", "*baik* (good)", "*murah* (cheap)", and "*nyaman* (comforTable)" are the most the most frequently occurring words in the positive reviews as shown in Figure 2.

Figure 2. Positive Word Cloud

Figure 3 shows that words, the words "*error*", "*tidak* (no)", "*jelek* (bad)", "*batal* (cancelled)", and "*lama* (long)" appear most frequently in negative reviews, reflecting user dissatisfaction.

Figure 3. Negative Word Cloud

As presented on Figure 4, the words "*harga* (price)", "*pesan* (order)", "*jemput* (pick up)", "*akun* (account)", "*masuk* (login), and "*peta* (maps)" appear frequently in a neutral sentiment review.

Figure 4. Neutral Word Cloud

Text features were extracted using TF-IDF weighting via the Python library. Next, the processed data was tested using LSTM classification. Where each word in the review is converted into a vector representation based on its occurrence weight. After the data processing was complete, a classification model was subsequently developed using the LSTM method. As shown in Figure 5, there were 153 reviews with positive sentiment, 60 reviews with neutral sentiment, and 28 reviews with negative sentiment. The testing process was carried out using 20% of the entire dataset as test data, while the remaining 80% was used for model training. The separation between tiraniku data and test data was carried out thoroughly to ensure that the model performance evaluation was objective and unbiased.

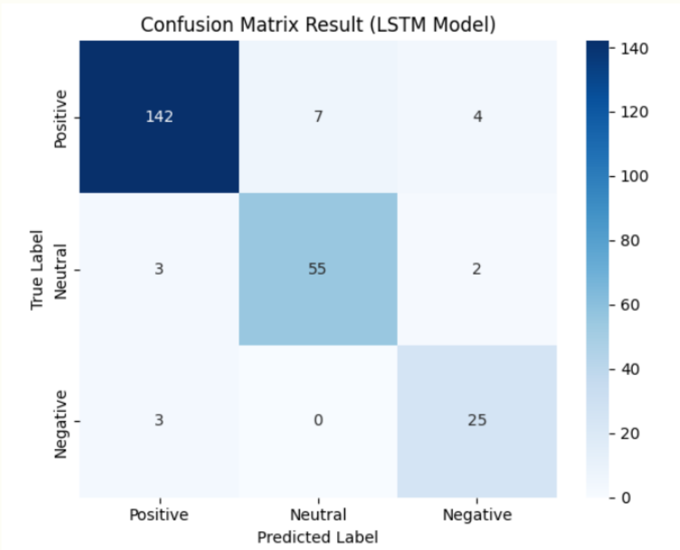

Figure 5. Confusion Matrix of LSTM Model

After conducting a thorough test on the data, the next stage is to assess how effective the developed model is. Performance testing measurements are made through a confusion matrix is employed to determine the performance evaluation metrics include accuracy, precision, recall, and F1-score, the results displayed inside Table 4.

Table 4. Confusion matrix result

Confusion matrix result				
	Accuracy	Precision	Recall	F1-Score
LSTM	95.21%	97.22%	84.02%	88.84%

Based on hhe outcomes of the confusion matrix are presented in Table 4. Performance evaluation from the LSTM model utilized in this research indicates that the LSTM algorithm exhibits very good performance in sentiment classification. LSTM is a capable deep learning technique capture word sequences and contextual information more effectively than traditional classification methods. The LSTM model used in this research obtained an accuracy rate through 95.21%, a precision value of 97.22%, a recall rate of 84.02%, and an F1-score of 88.84%. Although that recall value is slightly lower than precision, this may indicate a minor bias toward the majority class, as positive reviews dominate the dataset, even after applying the SMOTE technique to balance the data distribution. Compared with previous studies that used Naïve Bayes, SVM, and BiLSTM, the LSTM model demonstrates superior results due to its capability to learn and retain long-term relationships and contextual relationships between words, allowing for more accurate and context-aware sentiment classification. This study applies the LSTM model with lexicon-based labeling to enhance the quality and consistency of sentiment annotation before model training, resulting in more reliable classification performance. Figure 6 displays the results of sentiment analysis for Maxim app users, visualized as a word cloud. This word cloud illustrates the words that appear most frequently in feedback collected from users on the Google Play Store. The visualization provides an overview of user perceptions and experiences with the Maxim application.

Figure 6. Word Cloud Data

CONCLUSIONS

According to the findings of this research, it may be inferred that sentiment analysis conducted on Maxim app reviews collected from the Google Play Store between June 2024 and June 2025 yielded a total of 1,200 review data points. Of these, 762 reviews fell into the positive sentiment category, 157 into the neutral category, and 281 into the negative category. Of the total data, 80% is allocated for the training process and 20% for testing to guarantee reliable evaluation results an objective and unbiased model evaluation. This study applied the LSTM method combined with lexicon-based sentiment labeling to analyze user sentiment, which serves as the main contribution and distinguishes this research from previous studies that predominantly used conventional machine learning algorithms. The analysis results showed that LSTM model obtained an accuracy of 95.21%, with a precision of 97.22%, a recall of 84.02%, and an F1-score of 88.84%, indicating excellent classification performance. Further research it is advisable to increase the size of the dataset to enhance model generalization and maintain balanced class distribution using techniques such as oversampling, undersampling, or SMOTE, as well as to explore other deep learning architectures, including GRU, Bidirectional LSTM, and transformer-based models such as BERT.

REFERENCES

Alghifari, D. R., Edi, M., & Firmansyah, L. (2022). Implementasi Bidirectional Lstm Untuk Analisis Sentimen Terhadap Layanan Grab Indonesia. *Jurnal Manajemen Informatika (Jamika)*, *12*(2), 89–99. https://doi.org/10.34010/jamika.v12i2.7764

Aripiyanto, S., Tukino, T., Sufyan, A., & Nandaputra, R. (2022). Sentimen Analisis Twitter Ibu Kota Negara Nusantara Menggunakan Long Short-Term Memory Dan Lexicon Based. *Expert: Jurnal Manajemen Sistem Informasi Dan Teknologi*, *12*(2), 119-125. https://dx.doi.org/10.36448/expert.v12i2.2821

Aulia, T. M. P., Arifin, N., & Mayasari, R. (2021). Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19. *Sintech (Science And Information Technology) Journal*, *4*(2), 139–145. https://doi.org/10.31598/sintechjournal.v4i2.762

Damayanti, E., Vitianingsih, A. V., Kacung, S., Suhartoyo, H., & Lidya Maukar, A. (2024). Sentiment Analysis Of Alfagift Application User Reviews Using Long Short-Term Memory (Lstm) And Support Vector Machine (Svm) Methods. *Decode: Jurnal Pendidikan Teknologi Informasi*, *4*(2), 509–521. https://doi.org/10.51454/decode.v4i2.478

Dipietro, R., & Hager, G. D. (2020). Deep Learning: Rnns And Lstm. In *Handbook Of Medical Image Computing And Computer Assisted Intervention*, 503–519. https://doi.org/10.1016/B978-0-12-816176-0.00026-0

Gaafar, A. S., Dahr, J. M., & Hamoud, A. K. (2022). Comparative Analysis Of Performance Of Deep Learning Classification Approach Based On Lstm-Rnn For Textual And Image Datasets. *Informatica (Slovenia)*, *46*(5), 21–28. https://doi.org/10.31449/inf.v46i5.3872

Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning. *Jbase - Journal Of Business And Audit Information Systems*, *4*(2), 16-22. http://dx.doi.org/10.30813/jbase.v4i2.3000

Hasanah, A. N., & Sari, B. N. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Jasa Ojek Online Maxim Pada Google Play Dengan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(1), 60-96. https://doi.org/10.23960/jitet.v12i1.3628

Husein, A. M., Muntaza, L. H., Sinaga, G., Khoirulliza, K., & Sembiring, M. L. B. (2025). Model Cnn-Lstm Untuk Klasifikasi Tingkat Stres Mahasiswa Dalam Menghadapi Ujian Menggunakan Data Elektrokardiogram (Ekg). *Decode: Jurnal Pendidikan Teknologi Informasi*, *5*(1), 191–201. https://doi.org/10.51454/decode.v5i1.1080

Iqrom, M., Afdal, M., Novita, R., Rahmawita, M., & Ahsyar, T. K. (2025). Sentiment Analysis Of Gojek, Grab, And Maxim Applications Using Support Vector Machine Algorithm. *Jurnal Inovtek Polbeng - Seri Iformatika*, *10*(1), 237-248. https://doi.org/10.35314/52fycr56

Kokab, S. T., Asghar, S., & Naz, S. (2022). Transformer-Based Deep Learning Models For The Sentiment Analysis Of Social Media Data. *Computers & Operations Research*, *14, 100157*. https://doi.org/10.1016/j.array.2022.100157

Labuguen, M. Z. (2025). Enhancing Lstm Performance In Sentiment Analysis Through Advanced Data Preprocessing And Model Optimization Techniques. *World Journal Of Advanced Research And Reviews*, *25*(1), 2433–2443. https://doi.org/10.30574/wjarr.2025.25.1.0098

Mandar, G., Muhamamd, A. H., & Sudin, S. (2020). Klasifikasi Berita Indonesia Menggunakan Naïve Bayes Dengan Porter Stemmer. *Jurnal Teknik Informatika (J-Tifa)*, *3*(2), 17–22. https://doi.org/10.52046/j-tifa.v3i2.1121

Otter, D. W., Medina, J. R., & Kalita, J. K. (2021). A Survey Of The Usages Of Deep Learning For Natural Language Processing. *Ieee Transactions On Neural Networks And Learning Systems*, *32*(2), 604–624. https://doi.org/10.1109/TNNLS.2020.2979670

Pohan, S. A., Samsudin, S., & Sibarani, F. H. (2024). Analisis Sentimen Terhadap Aplikasi Maxim Menggunakan Algoritma Random Forest. *Journal Of Science And Social Research*, *7*(3), 1201–1208.

Putri, N. A., Srirahayu, A., & Arif Sudibyo, N. (2025). Analisis Sentimen Terhadap Aplikasi Kitalulus Menggunakan Metode Naive Bayes Dari Ulasan Google Play Store. *Smart Comp: Jurnalnya Orang Pintar Komputer*, *14*(2), 269-279. https://doi.org/10.30591/smartcomp.v14i2.7230

Putri, N. K., Vitianingsih, A. V., Kacung, S., Maukar, A. L., & Yasin, V. (2024). Sentiment Analysis Of Brand Ambassador Influence On Product Buyer Interest Using Knn And Svm. *Indonesian Journal Of Artificial Intelligence And Data Mining*, *7*(2), 327 – 336 http://dx.doi.org/10.24014/ijaidm.v7i2.29469

Rizki, M. R., & Sanjaya, M. R. (2024). Maxim Application User Satisfaction Analysis Using The User Experience Questionnaire Method. *INOVTEK Polbeng - Seri Informatika, 9*(2), 814–825. https://doi.org/10.35314/3p1r9s88

Saputra, A., Hariyono, R. C. S., & Saraswati, N. M. (2024). Analisis Sentimen Pengguna Aplikasi Mypertamina Menggunakan Algoritma Bidirectional Long Short Term Memory. *Jurnal Eksplora Informatika*, *13*(2), 156–163. https://doi.org/10.30864/eksplora.v13i2.973

Saputra, M., & Wahyuni, S. (2024). Analisis Sentimen Pengguna Pada Aplikasi Bank Digital Krom Dengan Algoritma Support Vector Machine. *Infotech Journal*, *10*(2), 327–332. https://doi.org/10.31949/infotech.v10i2.11801

Selle, N., Yudistira, N., & Dewi, C. (2022). Perbandingan Prediksi Penggunaan Listrik Denganmetode Long Short-Term Memory (Lstm) Dan Recurrent Neural Network (RNN). *Jurnal Teknologi Informasi Dan Ilmu Komputer, 9*(1), 155–162. https://doi.org/10.25126/jtiik.2022915585

Shyahrin, M. V., Sibaroni, Y., & Puspandari, D. (2023). Penerapan Metode Long Short-Term Memory Dan Word2vec Dalam Analisis Sentimen Ulasan Pada Aplikasi Ferizy Lstm And Word2vec Application For Sentiment Analysis Of Reviews On Ferizy. *Academic Journal*, *22*(4), 833–842. https://doi.org/10.33633/tc.v22i4.9205

Sujadi, H. (2022). Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine. *Infotech Journal*, *8*(1), 22–27. https://doi.org/10.31949/infotech.v8i1.1883

Tamami, G., Triyanto, W. A., & Muzid, S. (2025). Sentiment Analysis Mobile Jkn Reviews Using Smote Based Lstm. *Ijccs (Indonesian Journal Of Computing And Cybernetics Systems)*, *19*(1), 13-24. https://doi.org/10.22146/ijccs.101910

Valero-Carreras, D., Alcaraz, J., & Landete, M. (2023). Comparing Two Svm Models Through Different Metrics Based On The Confusion Matrix. *Computers & Operations Research*, *152*, 106131. https://doi.org/10.1016/j.cor.2022.106131

Wewengkang, H. F. B., Wungguli, D., Yahya, N. I., Hasan, I. K., & Abdussamad, S. N. (2025). Implementasi Bidirectional Lstm Dengan Word Embedding Fasttext Dalam Analisis Sentimen Ulasan Pengguna Aplikasi Maxim. *Riset Mahasiswa Matematika*, *4*(5), 204–217. https://doi.org/10.18860/jrmm.v4i5.33358

Widyaningrum, I., & Kamayani, M. (2023). Analisis Sentimen Opini Masyarakat Terhadap Penggunaan Layanan Maxim Menggunakan Algoritma Naïve Bayes. *Jurnal Coscitech (Computer Science And Information Technology)*, *4*(3), 651–660. https://doi.org/10.37859/coscitech.v4i3.6194

Wirayani, N. A., Wirastuti, N. M. A. E. D., & Manuaba, I. B. G. (2025). Analisis Sentimen Tanggapan Masyarakat Tentang Garuda Ikn Menggunakan Metode Naive Bayes. *Decode: Jurnal Pendidikan Teknologi Informasi*, *5*(1), 27–40. Https://Doi.Org/10.51454/Decode.V5i1.860