
CommniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Sentiment Analysis of Halodoc Application User Satisfaction Using The Naïve Bayes Method, SVM and LSTM

Parintan Manurung[1] , Guntoro*[2]

[12]Informatics of Engineering, Faculty of Computer Science, Universitas Lancang Kuning
*Corresponding Author
Email: guntoro@unilak.ac.id

Received dd/mm/yyyy, Revised 14/06/2025, Accepted 01/08/2025, Published Online First 03/08/2025, Published 03/08/2025

Abstract

The growth of digital technology in the health sector has encouraged the emergence of various health service applications, one of which is Halodoc. This study aims to analyze the sentiment of user satisfaction with the Halodoc application through reviews left on the Google Play Store. The method used in this research is Naïve Bayes, with stages including data collection, preprocessing (case folding, cleansing, tokenizing, stopword removal, stemming), weighting using TF-IDF, and sentiment classification. This research uses 350 review data as a dataset. The evaluation results showed that the Naïve Bayes-based sentiment classification model achieved 84% accuracy, with the majority of user sentiments being positive. The findings illustrate that the Halodoc application is generally well received by its users, but still needs improvement in some aspects of the service.

Keywords: Sentiment Analysis, Halodoc, Naïve Bayes, User Satisfaction, Text Mining

Introduction

The rapid development of information technology has profoundly influenced various sectors, particularly healthcare. The emergence of digital platforms such as Halodoc plays a crucial role in advancing healthcare delivery in Indonesia. This application enables users to access medical services conveniently and efficiently without needing to visit healthcare facilities physically. Halodoc's features, which include virtual consultations with doctors, pharmaceutical purchases, and appointment scheduling, exemplify how telemedicine can transform healthcare service access (Anthony, 2020).

Halodoc has positioned itself as a leading telemedicine solution in Indonesia, boasting a user base that accounts for a significant portion of telemedicine users, although exact figures may vary between sources (Taha et al., 2022). This significant adoption underscores the application's potential to extend healthcare access, particularly to those in remote locations, while also reducing associated costs and time for users. However, despite its advantages, Halodoc is not without challenges. User reviews on platforms like Google Play reflect various grievances, including unsatisfactory consultation durations and delivery errors for medications. Understanding and addressing these concerns is essential for further enhancing user satisfaction and expanding the application's reach (Wernhart et al., 2019).

To gauge user satisfaction accurately, a user-centric approach is proposed, leveraging sentiment analysis to interpret public opinion systematically. This technique allows researchers to classify thousands of user reviews as having positive, negative, or neutral sentiments, thus capturing the users' voices effectively. The study employs three machine learning classification methods—Naïve Bayes, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM)—to evaluate user sentiment regarding Halodoc's services. Naïve Bayes is known for its straightforward yet effective probabilistic method, while SVM is adept at isolating data through optimal hyperplanes, and LSTM is particularly suited for handling complex sequential data such as extensive text reviews (Alghamdi et al., 2024).

The comparative analysis of these algorithms aims to elucidate the best method for accurately classifying user sentiment, thereby providing valuable insights into user satisfaction with Halodoc. Such findings are anticipated to offer data-driven recommendations for enhancing the application, thus contributing meaningfully to the general improvement of digital healthcare services in Indonesia. Insights gained from this research could significantly influence future healthcare delivery models, emphasizing the need for continuous adaptation in the evolving landscape of telemedicine (Aldekhyyel et al., 2021).

Materials and Methods

This study was conducted to analyze user satisfaction sentiment towards the Halodoc application using three classification algorithms, namely Naïve Bayes, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM). The data used in this study were obtained from user reviews of the Halodoc application on the Google Play Store platform. The stages in this study consist of data collection, preprocessing, labeling, classification, and model evaluation. The following is an explanation of each stage:

Data Collection

The data collection process was meticulously carried out through a web scraping process using Google Colab with the Python programming language. A total of 2,500 user reviews of the Halodoc application available on the Google Play Store, specifically those marked as 'Most Relevant' and in Indonesian, were downloaded. These reviews provide a comprehensive reflection of various user experiences with the main features of the Halodoc application, such as doctor consultations, purchasing medicine, and hospital scheduling.

Preprocessing

The preprocessing stage aims to clean and prepare data before being classified. The preprocessing steps include:

a. Data Cleaning: Removes unnecessary non-alphabetic characters, punctuation, numbers, and symbols.
b. Case Folding: Converts all text to lowercase for consistency.
c. Tokenizing: Breaks text into words (tokens).
d. Stopword Removal: Removes common words that have no significant meaning in the analysis.
e. Stemming: Changes words to their basic form to have a consistent meaning.

Data Labeling

After the data is cleaned, the labeling process is carried out automatically using the Python library, TextBlob. TextBlob was chosen for its simplicity and efficiency in sentiment analysis tasks. It determines the sentiment polarity of each review, which is then used as a dataset ready for the classification process.

Classification

This study uses three classification methods, namely:

a. The first classification method used in this study is Naïve Bayes, a probabilistic-based algorithm that assumes independence between features. This algorithm is particularly suitable for text classification due to its efficiency and speed, providing a robust foundation for our analysis.
b. The second classification method employed is the Support Vector Machine (SVM), an algorithm that searches for the optimal hyperplane to maximally separate sentiment classes. Its adaptability and robustness make it a valuable addition to our study.
c. Long Short-Term Memory (LSTM): A variant of Recurrent Neural Network (RNN) that can capture long-term dependencies in text data.

Each model was trained and tested using a data split into 70% training data and 30% test data. Additionally, the models were tested with an 80:20 split scenario to assess the stability of the model's performance under different data distribution conditions. This step was crucial to ensure the robustness of the models and their ability to generalize to unseen data.

Model Evaluation

To evaluate the performance of the three algorithms, the following evaluation metrics were used:

a. Accuracy: Percentage of correct classification.
b. Precision: The model's ability to correctly classify positive data.
c. Recall: The model's ability to detect all positive data.
d. F1-Score: The harmonic mean of precision and recall.

The evaluation was carried out using the confusion matrix of each algorithm to determine the classification performance of positive, negative, and neutral sentiments.

Results and Discussion

Data Collection and Preparation

The data in this study were obtained through a web scraping process from the Google Play Store using Google Colab tools and the Python programming language, demonstrating our expertise in these technical aspects. A total of 2,500 Halodoc application user reviews were collected, which were then focused on Indonesian language reviews with the most relevant category.

Following the data collection, a meticulous preprocessing process was undertaken to clean and prepare the text. The preprocessing stages, including data cleaning, case folding, tokenizing, stopword removal, and stemming, were executed with precision. This process was designed to eliminate noise and transform the data into a more structured and easily processed form.

Pre-Processing Stages

a. Data Cleaning

Table 1 below shows the results obtained from data cleaning at this stage; characters other than letters will be deleted, including punctuation, numbers, and emoticon symbols.

Tabel 1. Data Cleaning Results

Initial Review Data	Data Cleaning Results
Terimakasih halodoc. Aplikasi sangat memudahkan pasien yg memang sedang terkendala sakit, tempat konsul dr yg jauh, maupun sgt cocok untuk org yg krg suka keramaian dan mageran👍.	Terimakasih halodoc Aplikasi sangat memudahkan pasien yg memang sedang terkendala sakit tempat konsul dr yg jauh maupun sgt cocok untuk org yg krg suka keramaian dan mageran

b. Case Folding

After Data Cleaning is done, Case Folding will be done. This stage aims to normalize words into lowercase letters (Lower Text) but still in the same sentence form.

Tabel 2. Case Folding Data Results

Data Cleaning Results	Case Folding Data Results
Terimakasih halodoc Aplikasi sangat memudahkan pasien yg memang sedang terkendala sakit tempat konsul dr yg jauh maupun sgt cocok untuk org yg krg suka keramaian dan mageran	terimakasih halodoc aplikasi sangat memudahkan pasien yg memang sedang terkendala sakit tempat konsul dr yg jauh maupun sgt cocok untuk org yg krg suka keramaian dan mageran

c. Text Normalization

Table 3 below shows the results obtained after the Text Normalization stage. At this stage, the initial review that has short words will be changed into standard word form. This aims to make it easier when carrying out the data labeling and classification process.

Tabel 3. Text Normalization Data Results

Data Case Folding Results	Text Normalization Data Results
terimakasih halodoc aplikasi sangat memudahkan pasien yg memang sedang terkendala sakit tempat konsul dr yg jauh maupun sgt cocok untuk org yg krg suka keramaian dan mageran	terimakasih halodoc aplikasi sangat memudahkan pasien yang memang sedang terkendala sakit tempat konsultasi dari yang jauh maupun sangat cocok untuk orang yang kurang suka keramaian dan mageran

d. Tokenization

In Table 4 below, this aims to cut words into a single form, to make it easier to identify a sentence.

Tabel 4. Data Tokenization Results

Text Normalization Data Results	Data Tokenization Results
terimakasih halodoc aplikasi sangat memudahkan pasien yang memang sedang terkendala sakit tempat konsultasi dari yang jauh maupun sangat cocok untuk orang yang kurang suka keramaian dan mageran	['terimakasih', 'halodoc', 'aplikasi', 'sangat', 'memudahkan', 'pasien', 'yang', 'memang', 'sedang', 'terkendala', 'sakit', 'tempat', 'konsultasi', 'dari', 'yang', 'jauh', 'maupun', 'sangat', 'cocok', 'untuk', 'orang', 'yang', 'kurang', 'suka', 'keramaian', 'dan', 'mageran']

e. Stopword Removal

After Tokenization is performed, the data obtained previously will undergo the Stopword Removal stage, namely to remove conjunctions that have no meaning. Examples of words that will be removed are "dan", "yang", "lalu", "gak", "aja", etc.

Tabel 5. Data Stopword Removal Results

Data Tokenization Results	Data Stopword Removal Results
['terimakasih', 'halodoc', 'aplikasi', 'sangat', 'memudahkan', 'pasien', 'yang', 'memang', 'sedang', 'terkendala', 'sakit', 'tempat', 'konsultasi', 'dari', 'yang', 'jauh', 'maupun', 'sangat', 'cocok', 'untuk', 'orang', 'yang', 'kurang', 'suka', 'keramaian', 'dan', 'mageran']	terimakasih halodoc aplikasi memudahkan pasien terkendala sakit konsultasi cocok orang suka keramaian mageran

f. Stemming

Table 6 below is the result of the Stemming process, where sentences that previously still had affixes have been deleted, and the latest results are obtained.

Tabel 6. Data Stemming Results

Data Stopword Removal Results	Data Stemming Results
terimakasih halodoc aplikasi memudahkan pasien terkendala sakit konsultasi cocok orang suka keramaian mageran	terimakasih halodoc aplikasi mudah pasien kendala sakit konsultasi cocok orang suka ramai mageran

Data Labeling

The distribution of review sentiments towards the Halodoc application shows that the majority of users gave positive responses, with a total of 1,095 data points (43.8%). Meanwhile, neutral sentiment is in second place with 765 data (30.6%), followed by negative sentiment with 640 data (25.6%). These results indicate that Halodoc is generally well-received by its users, as indicated by the percentage of positive reviews that dominate. However, the presence of a significant proportion of neutral and negative reviews also suggests that there are certain aspects that still need to be improved, both in terms of features, interface, and customer service. Thus, although the image of this application tends to be good, feedback from users can be valuable evaluation material for further development.

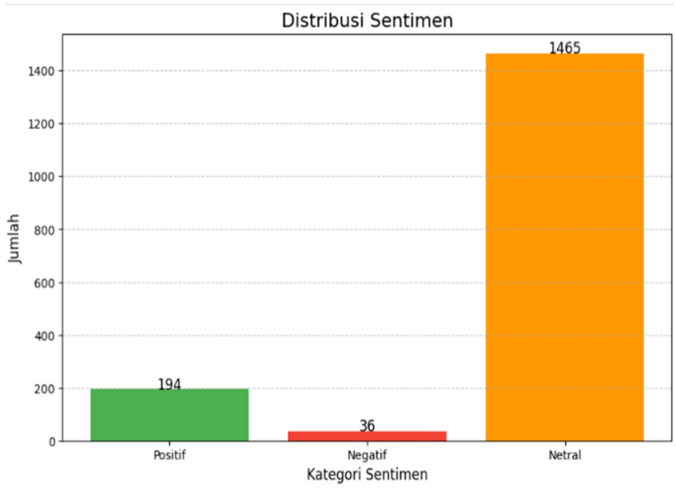

Figure 1. Data Labeling

TF-IDF Weighting

After the data labeling stage, the robust TF-IDF weighting process is initiated. This process identifies the top ten words that frequently appear, such as application, doctor, help, halodoc, medicine, consultation, fast, doctor, really, and good. These frequently occurring words are then presented in a diagram, highlighting the robustness of the TF-IDF weighting process.

Classification of Naïve Bayes, SVM, and LSTM

In this study, we investigated the performance of three classification algorithms, Naïve Bayes, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM), on a dataset. The dataset was divided into two different schemes, with a 70% training data and 30% testing data division. The results were significant, with SVM recording the highest accuracy of 93.7%, followed by LSTM with 91.0%, and Naïve Bayes with 88.4%. These findings underscore the superiority of SVM in handling data with a 70:30 division, a crucial insight for data scientists and machine learning practitioners.

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

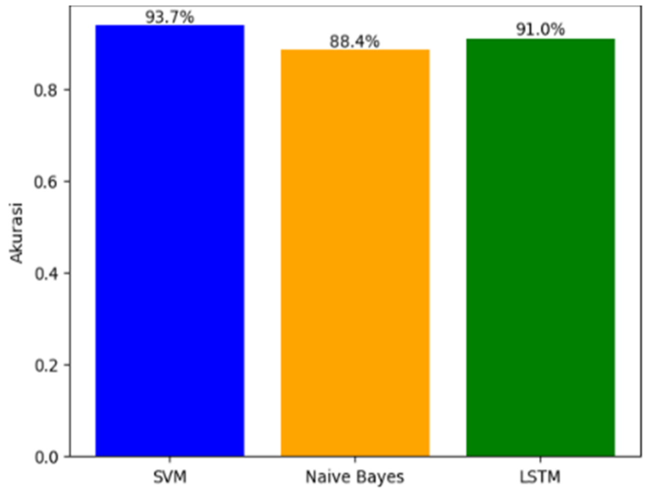

Figure 2. Comparison Results

Furthermore, when the dataset was divided into 80% training data and 20% testing data, SVM once again proved its adaptability, achieving an accuracy of 93.8%. Naïve Bayes followed with 88.2%, and LSTM, despite its previous high accuracy, experienced a decrease to 87.3%. This comparison highlights the adaptability of SVM, consistently delivering high performance regardless of the data division ratio. These findings can guide the selection of an algorithm for sentiment analysis, particularly when stability and accuracy are paramount.

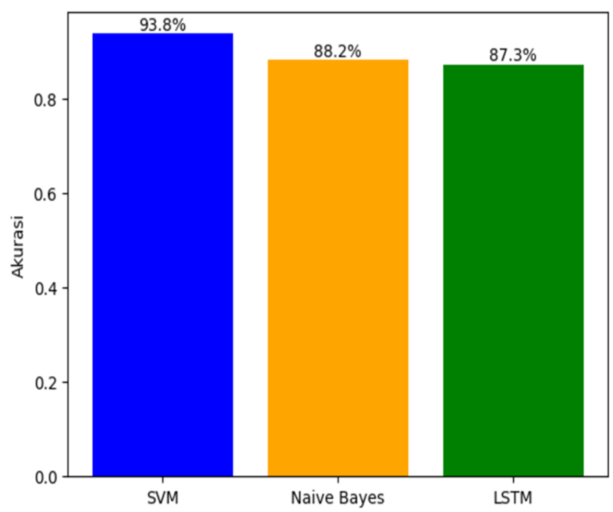

Figure 3. Comparison Results

Sentiment Classification

The sentiment analysis in this study is categorized into three main groups to understand user perceptions comprehensively:

CommiTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Positive Sentiment

Positive sentiment reflects responses that support or provide good assessments of the object being studied. This category is identified through keywords such as "satisfied", "happy", "good", and "impressive", which indicate user appreciation of the service or product. Visualization through the Positive Sentiment Wordcloud (Figure 4) strengthens this finding by displaying dominant words that are optimistic, indicating that most respondents had a pleasant experience.

Figure 4. *Wordcloud Positive Sentiment*

Negative Sentiment

Negative sentiment is characterized by expressions of dissatisfaction or criticism, such as "disappointed", "bad", "unsatisfied", and "disappointing". These words reflect complaints or unfulfilled expectations from users. The Negative Sentiment Wordcloud (Figure 5) displays these critical phrases visually, helping to identify specific aspects that need attention for improvement.

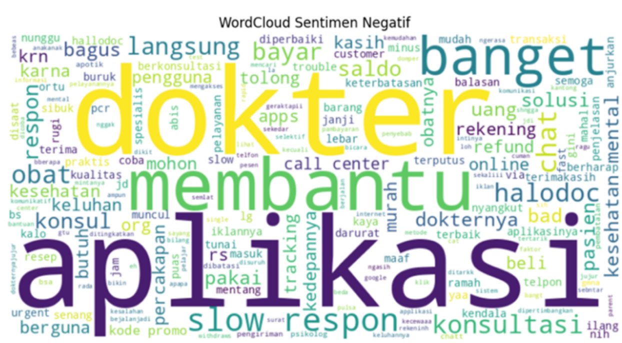

Figure 5. *Wordcloud Negative Sentiment*

Neutral Sentiment

Neutral sentiment includes responses that do not indicate a clear preference, with keywords such as 'normal', 'fair', or 'no opinion'. This category is significant as it indicates that some users do not have a strong enough experience to give a positive or negative assessment. The Neutral Sentiment Wordcloud (Figure 6) shows general words, confirming that there is a segment of users who are ambivalent in evaluating the research object.

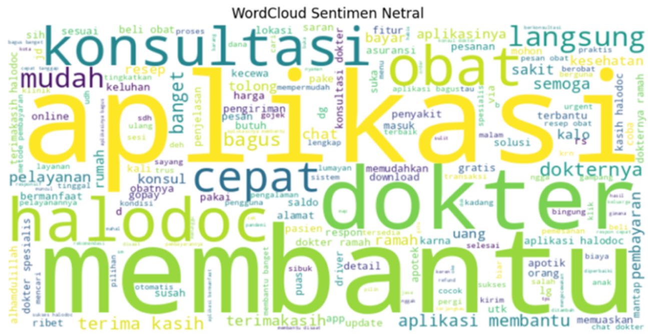

Figure 6. *Wordcloud Neutral Sentiment*

With this division, the analysis not only measures the proportion of each sentiment but also reveals the reasons behind user perceptions through dominant keywords. For instance, if the Negative Sentiment Wordcloud (Figure 5) highlights 'slow' as a dominant word, it suggests that improving the speed of the service could enhance user experience. The results provide a strong basis for recommendations for improvements or strengthening features that are already working well.

Discussion

The classification results show that the Halodoc application gets a dominant positive sentiment, but there are also quite a few negative reviews that indicate user dissatisfaction, especially regarding the consultation and drug delivery features. The high accuracy of the LSTM algorithm shows that this model is most capable of handling complex and lengthy text data, compared to SVM and Naïve Bayes. The selection of the correct algorithm in sentiment analysis significantly affects the validity of the results. The combination of good preprocessing, accurate labeling, and the selection of the right model will produce valuable insights for application developers in improving service quality.

Conclusion

The Naïve Bayes Algorithm, Support Vector Machine (SVM) Algorithm, and LSTM Algorithm can be used in analyzing Halodoc application reviews by producing three classifications, including positive sentiment class, negative sentiment class, and neutral sentiment class. Based on the results of the research that has been done, it can be concluded that in this study, using the Support Vector Machine (SVM) method is more accurate, with higher accuracy, compared to using the Naïve Bayes and LSTM methods.

References

Aldekhyyel, R., Almulhem, J., & Binkheder, S. (2021). Usability of telemedicine mobile applications during covid-19 in saudi arabia: a heuristic evaluation of patient user interfaces. Healthcare, 9(11), 1574. https://doi.org/10.3390/healthcare9111574

Alghamdi, W., Alshaibani, A., & Alotaibi, T. (2024). Study of patients' satisfaction with different types of telemedicine services provided by the ministry of health in saudi arabia. Journal of Advances and Scholarly Researches in Allied Education, 21(1), 11-16. https://doi.org/10.29070/qjc42x21

Anthony, B. (2020). Exploring the adoption of telemedicine and virtual software for care of outpatients during and after covid-19 pandemic. Irish Journal of Medical Science (1971 -), 190(1), 1-10. https://doi.org/10.1007/s11845-020-02299-z

CommiTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Taha, A., Shehadeh, M., Alshehhi, A., Altamimi, T., Housser, E., Simsekler, M., … & Alhajri, N. (2022). The integration of mhealth technologies in telemedicine during the covid-19 era: a cross-sectional study. Plos One, 17(2), e0264436. https://doi.org/10.1371/journal.pone.0264436

Wernhart, A., Gahbauer, S., & Haluza, D. (2019). Ehealth and telemedicine: practices and beliefs among healthcare professionals and medical students at a medical university. Plos One, 14(2), e0213067. https://doi.org/10.1371/journal.pone.0213067