JIEM : JOURNAL INFORMATIC, EDUCATION AND MANAGEMENT
Vol 8 No 2 (2026): March - August 2026, pp. 925~932
ISSN: 2716-0696, DOI: 10.61992/jiem.v8i2.420

SENTIMENT ANALYSIS OF JKN MOBILE APPLICATION REVIEWS ON PLAYSTORE

Okky Agung Sunata [1], Hendra Gunawan [1]

[1] Informatics Engineering Study Program , STMIK IM, Bandung, Indonesia

Article Info

Article history:
Received June 06 , 2026
Revised July 15 , 2026
Accepted July 28 , 2026

Keywords:
Sentiment Analysis, Mobile JKN, Google Play Store, Naive Bayes, Support Vector Machine, Text Mining.

ABSTRACT

The JKN Mobile application developed by BPJS Kesehatan aims to facilitate digital access to healthcare services for the public. However, user reviews on the Google Play Store show various opinions, ranging from satisfaction to complaints about technical issues. This study aims to analyze user sentiment, evaluate the performance of the Naive Bayes algorithm, and identify the main technical obstacles. The methods used include data scraping from the Google Play Store, text preprocessing (case folding, cleaning, tokenization, normalization, stopword removal, stemming), word weighting using TF-IDF, and classification using Naive Bayes. The results show that negative sentiment dominates (50.7%) compared to positive sentiment (45.9%). The Naive Bayes algorithm achieved an accuracy of 85.61% with an F1-Score of 85.8%, but the low Cross Validation value indicates model instability. The main complaints from users were login and OTP issues (1,632 complaints), complicated interface, as well as errors and bugs. From this analysis, it is concluded that aspects of system stability and verification still require significant improvement. These classification results can serve as a reference for BPJS Kesehatan in prioritizing feature development and technical improvements.

Corresponding Author:
Okky Agung Sunata | STMIK IM, Bandung, Indonesia
Email: okkyagung5@gmail.com1

INTRODUCTION

In today's digital age , the public service sector must adapt to technology to increase efficiency and transparency . One major step in healthcare services in Indonesia was the launch of the Mobile JKN application by BPJS Kesehatan. This application helps participants access services more easily , such as registering for health facility queues , changing data , and checking hospital bed availability without having to visit a branch office . The Google Play Store provides a platform for the public to provide direct feedback about apps . User reviews are rich in information about their experiences . However , due to the large and unstructured volume of reviews , manual monitoring is impractical . Therefore , an automated method for processing this data is needed , namely sentiment analysis . Using natural language processing and machine learning techniques , user opinions can be classified as positive , neutral , or negative . This is crucial for mapping whether apps have met public expectations or have become new barriers to accessing healthcare .

RESEARCH METHODS

This study uses a quantitative approach , applying text mining and machine learning techniques to analyze the sentiment of users of the Mobile JKN application . Broadly speaking , the methodology consists of several main stages : data collection , data preprocessing , word weighting , sentiment classification , and model evaluation . The following is a detailed explanation of each stage .

Data Collection

Data collection was carried out using web scraping techniques from the Google Play Store platform using the google-play-scraper library in the Python programming language . The extracted data included username , review content , rating score (1-5), and review date . The scraping process was carried out on the Mobile JKN application review page with the ID id.bpjs.jknmobile . After the data was successfully collected , sentiment labeling was carried out automatically based on the rating scores given by users . This classification refers to the rules commonly used in sentiment analysis research , namely ratings 4 and 5 are categorized as positive sentiment , rating 3 as neutral sentiment , and ratings 1 and 2 as negative sentiment .

Preprocessing Data

Data preprocessing is a crucial stage in text mining, aiming to clean and convert raw text data into a machine - readable format . The preprocessing steps taken in this study are as follows :

a. Case Folding

This initial stage changes all letters in the review text to lowercase . The goal is to equalize the representation of the same word even if it is written with a capital letter at the beginning or middle of a sentence . For example , the words " Easy ", "EASY", and " easy " will be changed to " easy " so that they are considered identical words .

b. Cleaning

At this stage , various elements that are irrelevant or disruptive to the analysis process are removed from the text . These elements include numeric digits (0-9), punctuation marks such as periods , commas , question marks , and exclamation points , URL links , emojis, special characters such as currency symbols and mathematical notation , and unnecessary spaces .

c. Tokenization

Tokenization is the process of breaking a sentence or paragraph into smaller pieces called tokens, which are generally individual words . For example , the sentence " This application is very helpful " would be broken down into tokens: [" application " , " this ", "very", " helpful.

d. Normalization

Normalization aims to change non- standard words , abbreviations , or slang frequently used by users into standard forms according to the Big Indonesian Dictionary (KBBI). For example , the words "gak", "ga", " gak ", and " gk " are normalized to " tidak ", " bgt " becomes " banget

", " sy " becomes " saya ", and " login " becomes " masuk ". This process is important because user reviews in digital media often use informal language .

e. Stopword Removal

Stopwords are common words that frequently appear in a document but do not significantly contribute to the meaning or sentiment of the text . Examples of stopwords in Indonesian include "yang", "dan", "di", " dari ", " ke ", " ini ", " itu ", " saja ", " sudah ", as well as words such as " sih ", " lah ", " kok ", and "dong". The removal of these words aims to reduce the feature dimensionality and improve computational efficiency .

f. Stemming (Basic Word Search)

Stemming is the process of converting affixed words into their root words by removing prefixes , suffixes , and infixes . This research uses the Sastrawi library , which is specifically designed for Indonesian . For example , the words " mengerak " , " lestari " , and " mengerak " will be stemmed into the root word " baik ". This process simplifies word representation so that words with the same root can be grouped together .

Word Weighting

After the text data has gone through the preprocessing stage , the next step is to convert the text data into a numeric representation that can be processed by a machine learning algorithm . This study uses the TF-IDF (Term Frequency - Inverse Document Frequency) method to weight words . This method calculates how important a word is in a document relative to the entire corpus . Mathematically , the TF-IDF weight is calculated using the following formula :

$$W_{d,t} = tf_{d,t} \times log\left(\frac{N}{df_t}\right)$$

Information :

$W_{d,t}$: The weight of word t in document d
$tf_{d,t}$: Frequency of occurrence of word t in document d (Term Frequency)
N: Total number of documents
df_t: Number of documents containing the word t (Document Frequency)

Sentiment Classification

To classify user review sentiment into three categories (positive , neutral , and negative), this study uses a machine learning algorithm that has been proven effective in the field of text classification , namely the Naive Bayes Classifier (NBC).

Naive Bayes is a probability- based algorithm that applies Bayes' Theorem with a strong (naive) assumption of independence between features . This assumption assumes that the presence of a word in a document is independent of the presence of other words . Although this assumption is often not entirely true in the real world , the Naive Bayes algorithm has proven to be very effective and efficient for text classification , especially with large data sizes . The basic formula used is :

$$P(C_k \mid X) = \frac{P(X \mid C_k) \cdot P(C_k)}{P(X)}$$

where $P(C_k \mid X)$ C_k $P(C_k \mid X)$ $C_k P(C_k)$ C_k

Model Evaluation

Model evaluation aims to measure the performance of the Naive Bayes algorithm in classifying user review sentiment . This study uses a Confusion Matrix as the basis for calculating evaluation metrics , including Accuracy , Precision, Recall, and F1-Score as the main metrics .

a. Accuracy

Measure the percentage of correct predictions as a whole :

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

b. Precision

Measuring the accuracy of the model's positive predictions :

$$\text{Precision} = \frac{TP}{TP + FP}$$

c. Recall

Measuring the model's ability to find all positive data :

$$\text{Recall} = \frac{TP}{TP + FN}$$

d. F1-Score (Key Metric)

The harmonic mean of precision and recall. Chosen as the primary metric because it provides a balanced picture , especially when the data distribution is imbalanced .

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Information :

TP : True Positive (positive data predicted positive)
TN : True Negative (negative data is predicted to be negative)
FP : False Positive (negative data is predicted to be positive)
FN: False Negative (data positive predicted negative)

e. F1-Score for Multiclass Classification

Since there are three sentiment classes (positive , neutral , negative), this study uses macro-averaging F1-Score to calculate the average F1 of the three classes :

$$F1_{\text{macro}} = \frac{F1_{\text{positif}} + F1_{\text{netral}} + F1_{\text{negatif}}}{3}$$

This is done so that each class gets the same weight , especially because the data distribution is unbalanced .

RESULTS AND DISCUSSION

Based on the results of scraping reviews of the Mobile JKN application from the Google Play Store with the ID app.bpjs.mobile , 3,370 valid reviews were collected after going through the preprocessing stage . The distribution of data based on sentiment is presented in Table 1 below :

Table 1. Shows That Negative Sentiment

Sentiment	Number of Reviews	Percentage
Positive	1546	45.9%
Neutral	117	3.5%
Negative	3370	50.7%
Total	3370	100%

Table 1 shows that negative sentiment dominates at 50.7 %, followed by positive sentiment at 45.9 %, and neutral sentiment at only 3.5%. This indicates that more than half of users experienced difficulties using the Mobile JKN application .

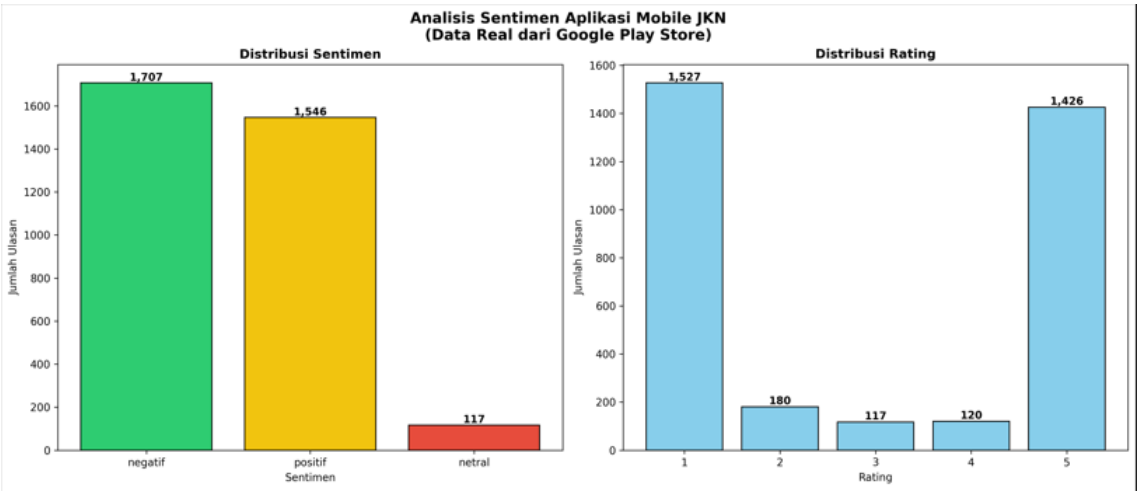

Figure 1Sentiment Distribution

World Cloud Data Visualization

The analysis of the most frequently appearing words in user reviews is visualized through a Word Cloud in Figure 1. This Word Cloud represents the frequency of word appearance after going through the preprocessing stage (case folding, cleaning, stopword removal, and stemming).

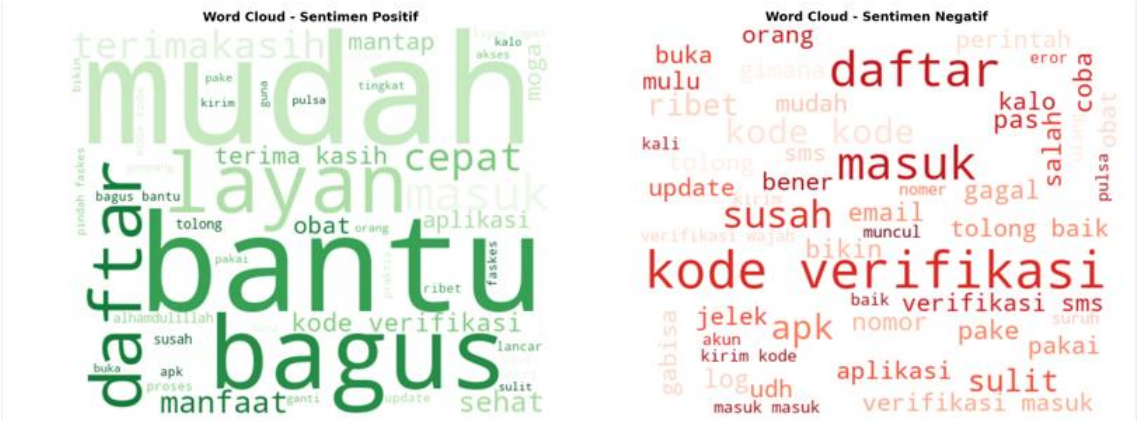

Figure 2. 2and Negative Sentiment Word Cloud

Based on Figure 3.2, there is a significant difference between the dominant words in positive and negative sentiments :

a. Positive Sentiment:

Dominant words include " easy ," "register," " good ," " benefit ," " help ," and " fast ." This indicates that users who provided positive reviews felt helped by the ease of online queue registration and experienced the benefits and speed of service provided by the Mobile JKN application .

b. Negative Sentiment:

The dominant words include " difficult ", " complicated ", "error", " failed ", "update", "login", " enter ", and " number ".

Naïve Bayes Classification

Confusion Matrix

Testing the Naive Bayes model on the test data produces the following Confusion Matrix :

Table 2. Confusion Matrix of Sentiment Classification

Actual / Prediction	Positive	Neutral	Negative
Positive	248	0	61
Neutral	0	22	0
Negative	12	0	329

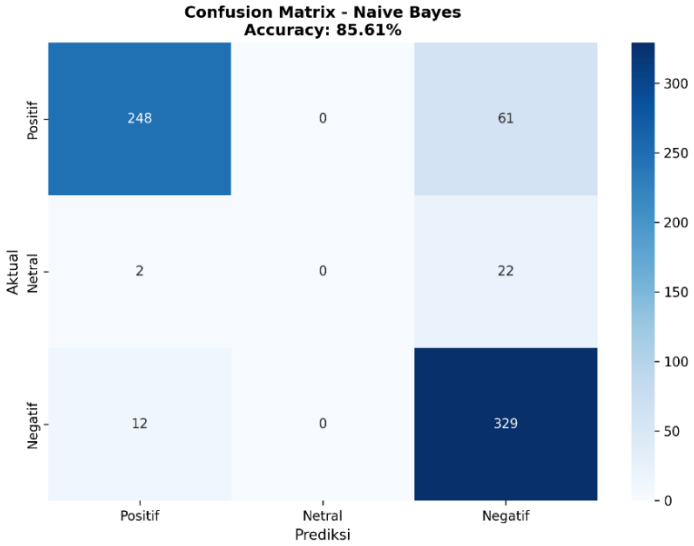

Figure 2. Confusion Matrix

Based on Table 2, from 672 test data tested , the Naive Bayes model successfully predicted 599 reviews correctly (248 positive + 22 neutral + 329 negative) and incorrectly predicted 73 reviews (61 positive incorrectly turned negative + 12 negative incorrectly turned positive). Thus , the model accuracy reached 85.61% .

The model's main weakness lies in the classification of positive sentiment , where 61 positive reviews were incorrectly predicted as negative . This is because positive reviews often contain mild complaint words such as 'error' or ' slow ' which are more dominantly recognized by the model as characteristics of negative sentiment . " Meanwhile , the model is very good at recognizing neutral sentiment because its word patterns are unique and do not overlap with other classes . "

Model Performance

Based on the Confusion Matrix, the Naive Bayes model produces performance as in Table 3.2

Table 3. Model Evaluation Results per Class

Class	Precision	Recall	F1-Score
Positive	95%	80.3%	87.2%
Neutral	100%	100%	100%
Negative	84.4%	96.5%	90.0%

The model achieved an accuracy of 85.61% with a weighted F1-Score of 85.8%. However, the low Cross Validation value (58.16%) indicates the model is unstable , mainly due to the very small amount of neutral sentiment data (only 117 reviews).

Conclusion

Based on the established research objectives , three main conclusions were obtained . First , regarding the level of user satisfaction, negative sentiment dominated by 50.7 % compared to positive sentiment at 45.9 % , which indicates that more than half of the users experienced problems in using the Mobile JKN application . Second , in terms of algorithm performance , Naive Bayes achieved an accuracy of 85.61% with a weighted F1-Score of 85.8%, but the low

Cross Validation value (58.16%) indicated that the model was unstable due to data imbalance , especially because the amount of neutral sentiment data was very small . Third , the main technical obstacles complained about by users were dominated by Login and OTP problems (1,632 complaints), followed by complicated UI/UX interface problems and errors and bugs in the application . Overall , although the Mobile JKN application is useful for users , aspects of system stability and ease of access still require significant improvements .

BIBLIOGRAPHY

Abidin, Z., & Rusli, A. (2022). Sentiment Analysis of JKN Mobile Application Users Using the Support Vector Machine Method. Journal of Informatics and Information Systems , 3(1), 45-53.

BPJS Kesehatan. (2025). Annual Report on Digitalization of Health Services through Mobile JKN . Jakarta: BPJS Kesehatan.

Feldman, R., & Sanger, J. (2007). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann.

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.

Indrayuni , E. (2020). Sentiment Analysis of Online Sales Application Reviews Using the Naive Bayes Algorithm . Journal of Evolution , 8(2), 55-62.

Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Stanford University.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

Nugroho, AS, Witarto , AB, & Handoko, D. (2003). Support Vector Machine: Theory and Application in Bioinformatics . Community of Indonesian Mathematics Scientists and Teachers .

Pratama, AR (2024). Comparison of Classification Algorithms for Sentiment Analysis on Google Play Store Reviews . Journal of Information Technology and Computer Science (JTIIK) , 11(2), 310-318.

Sari, PK, & Asniar , A. (2021). Sentiment Analysis of JKN Mobile Application Users Using Naive Bayes Algorithm and Feature Selection Optimization . Budidarma Informatics Media Journal , 5(3), 820-827.

Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. Information Processing & Management, 45(4), 427-437.

Tala, FZ (2003). A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia . University of Amsterdam.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.