Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1028−1036
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7777

Sentiment Analysis of SiKasep Application Reviews on the Play Store Using the Naïve Bayes Approach

Ariiq Afrahtama*, Putu Harry Gunawan

School of Computing, Telkom University, Bandung, Indonesia
Email: [1,*]afrahtama@student.telkomuniversity.ac.id., [2]phgunawan@telkomuniversity.ac.id
Correspondence Author Email: afrahtama@student.telkomuniversity.ac.id
Submitted: **26/06/2025**; Accepted: **01/09/2025**; Published**: 02/09/2025**

Abstract−The Ministry of Public Works and Public Housing (PUPR) launched the SiKasep application (Subsidized Housing Mortgage Information System) to streamline subsidized housing loan applications. This research analyzes user sentiment toward SiKasep using 3,416 Google Play Store reviews through Naïve Bayes classification to provide actionable insights for government digital service improvement. The methodology encompasses data scraping, comprehensive preprocessing addressing Indonesian language challenges (slang normalization and morphological complexity), TF-IDF feature extraction, and Complement Naïve Bayes classification with hyperparameter optimization. The preprocessing pipeline reduced vocabulary sparsity by 47%, while RandomOverSampler addressed significant class imbalance. The Complement Naïve Bayes classifier achieved 75.98% accuracy with balanced performance across sentiment classes (precision: 79%, recall: 76%, F1-score: 76%). Analysis revealed predominantly negative sentiment (52.4%), primarily related to registration and authentication difficulties, including document verification, login functionality, and KTP integration issues. Positive sentiment highlighted user appreciation for core housing services when technical barriers were absent. The findings emphasize the importance of streamlined registration processes and robust technical infrastructure for government digital services. This research contributes to understanding Indonesian e-government user experiences and provides a replicable sentiment analysis framework supporting evidence-based policy development for enhanced digital service delivery.

Keywords: Sentiment Analysis; Naïve Bayes Classification; Indonesian Government Applications; E-Government Services; Google Play Store Reviews; Digital Service Improvement; Indonesian Language Processing; Text Mining

1. INTRODUCTION

Housing is one of the basic needs that is very important for everyone[1]. To ensure that everyone in Indonesia has access to decent housing, the Ministry of Public Works and Public Housing (PUPR) has launched subsidized housing loans managed by the Housing Finance Liquidity Facility (FLPP) program [1].

In the development of information technology, the Ministry of Public Works and Public Housing (PUPR) launched the SiKasep application (Subsidized Housing Mortgage Information System) as an innovation to support the process of applying for subsidized Housing Selection Loans (KPR). SiKasep is a technology-based application designed to make it easier for the public. This application offers convenience in searching for subsidized housing, applying for mortgages, and also facilitates the data verification process through integration with various institutions such as Dukcapil of the Ministry of Home Affairs. With the presence of SiKasep, the government aims to enhance efficiency and transparency in the distribution of subsidized housing to the public, supporting the achievement of decent housing goals for all Indonesian people.

In the context of the SiKasep application, sentiment analysis can be applied to analyze user reviews of this application, especially on the Google Playstore platform. Reviews reflect various opinions regarding the application's performance, which can provide insights for application development. Sentiment analysis is an approach in natural language processing aimed at extracting and understanding the opinions or feelings contained in text. This method is used to assess whether a statement or document is positive, negative, or neutral [2].

In sentiment analysis, the use of natural language processing techniques, machine learning, and classification algorithms is typically involved to identify keywords, phrases, or contexts that indicate a particular sentiment [3] In general, there are five stages in conducting this analysis, namely data collection, pre-processing, feature selection, classification, and evaluation. There are commonly used data classification methods including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naïve Bayes [4].

Naïve Bayes, as one of the renowned methods in the domain of machine learning studies, particularly in classification, stands out due to its effectiveness, especially in text processing [5]. Despite its simplicity, this method can quickly handle classification problems and is easy to implement [6]. By calculating the posterior probability of each category or label based on observed features from training data, Naïve Bayes can be used in various applications such as sentiment analysis, text classification, and pattern recognition [7].

This research employs the Naïve Bayes method to analyze sentiment in SiKasep application reviews. The analyzed data was obtained from user reviews on Google Playstore through scraping techniques using Google Colaboratory, totaling 3,416 reviews [8]. Subsequently, the data will be stored in CSV format and then categorized into positive, neutral, and negative categories [8]. Following this process, the dataset will be processed using Google Colaboratory [8].

Related research on sentiment analysis has been extensively conducted previously , Gilbert et al. (2023) conducted sentiment analysis of the Mypertamina application reviews on Google Play Store using the Naïve Bayes method with 3,948 review data, achieving 91% accuracy, 92% precision, and 100% recall [9]. Another comparative

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1028−1036
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7777

study by Ramadhani et al. (2022) analyzed sentiment analysis performance comparing Naïve Bayes method with K-Nearest Neighbors (KNN), demonstrating 90.77% accuracy for Naïve Bayes and 74.77% accuracy for KNN method [10]. Additionally, Alviani et al. (2023) conducted sentiment analysis on WeTV application reviews from Google Play Store using Support Vector Machine with 4,024 review data, achieving 89% accuracy, 87% precision, and 83% recall, focusing on streaming application user satisfaction analysis [3]. Similarly, Lestandy et al. (2021) analyzed sentiment of COVID-19 vaccine tweets using both Recurrent Neural Network and Naïve Bayes with TF-IDF technique on 5,000 tweet dataset, where RNN (TF-IDF) achieved 97.77% accuracy compared to Naïve Bayes (TF-IDF) with 80% accuracy [5].

However, previous studies have primarily focused on commercial applications such as streaming services, e-commerce, and social media platforms, with limited attention to government digital services. Most existing research has concentrated on binary sentiment classification (positive/negative) rather than comprehensive analysis of government service applications that serve specific demographics with unique needs and expectations. Furthermore, there is a gap in sentiment analysis research specifically targeting subsidized housing applications like SiKasep, which addresses critical social needs and involves different user demographics compared to commercial applications. This research gap necessitates targeted analysis to understand user sentiment patterns and satisfaction levels in government housing digital services, particularly in developing countries like Indonesia where digital government services are rapidly expanding.

Based on the existing problems, this research is conducted to understand user responses to the SiKasep application through reviews they provide on Google Playstore [11]. The main focus of the research is on sentiment analysis of user reviews using the Naïve Bayes algorithm, which has proven effective as demonstrated by previous research outlined above. This algorithm was chosen because it has high accuracy and is superior compared to other classification algorithms, such as KNN. Through the use of this algorithm, the research aims to explore and understand user perspectives toward the SiKasep application, including positive, neutral, and negative aspects that can be identified from their reviews. This application functions not only as a mere technical tool but also as a platform that meets the expectations and needs of Indonesian society in obtaining access to adequate housing.

This study specifically aims to provide actionable insights for government agencies to improve digital service delivery and enhance user satisfaction. The findings will contribute to evidence-based policy development in e-government initiatives.

2. RESEARCH METHODOLOGY

2.1 Research Stages

To illustrate the research flow, a flowchart is needed. The flowchart consists of several steps to complete the research in order to achieve efficient results. The steps are data scraping, data labelling, data preprocessing, splitting data into training and test data, feature extraction using TF-IDF, modelling using Naïve Bayes Classifier (NBC) and performance evaluation[12]. In the preprocessing stage, it consists of cleaning, case folding, normalization, stopword removal, and stemming. The research methodology is designed to address unique challenges in analyzing Indonesian government digital services where traditional commercial application analysis approaches may not be sufficient. The methodology emphasizes reproducibility and transparency, enabling validation and potential replication in similar government digital service contexts across different domains and geographical regions. Ethical considerations regarding user privacy and data protection are carefully addressed in accordance with Indonesian data governance regulations and international best practices. The study adopts an interdisciplinary research framework that integrates computational linguistics, public policy analysis, and human-computer interaction principles. This multidisciplinary approach enhances the depth and breadth of analysis, ensuring that technical findings are contextualized within broader policy and social implications. The methodology incorporates domain-specific adaptations to handle bureaucratic terminology and formal language patterns commonly found in government application reviews. Advanced preprocessing techniques are specifically calibrated for Indonesian language characteristics, including morphological complexity and regional dialectical variations. The research framework accommodates the unique linguistic challenges posed by informal user-generated content in government service contexts. Theoretical foundations draw from both machine learning theory and public administration research to ensure methodological rigor and practical relevance. The research design enables longitudinal analysis capabilities for tracking sentiment trends over time as government services evolve. Quality assurance protocols are implemented at each stage to maintain data integrity and analytical validity. Cultural and linguistic considerations specific to Indonesian society are embedded throughout the analytical framework to ensure culturally sensitive interpretation of user feedback. Comprehensive documentation standards are maintained throughout the research process to facilitate peer review and enable future researchers to validate and extend the findings. Risk mitigation strategies are implemented to address potential challenges in data collection, model training, and result interpretation within the dynamic government service environment. The flowchart sequence from the beginning to the end is shown in Figure 1.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1028−1036
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7777

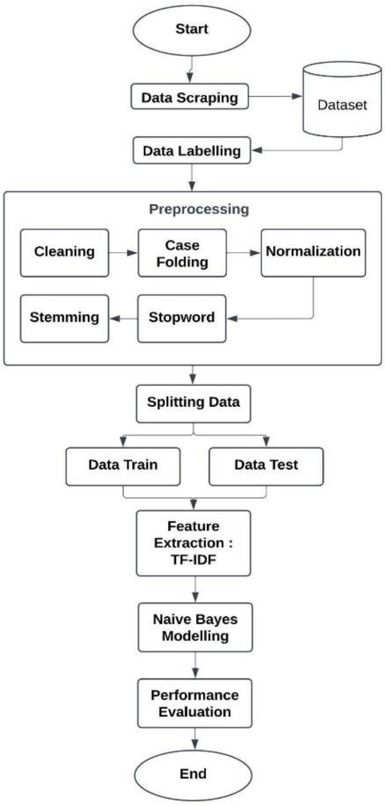

Figure 1. Methodology workflow

Figure 1 Illustrates the complete methodology workflow for sentiment analysis of SiKasep application reviews, consisting of sequential processes starting from data scraping to collect user reviews from Google Play Store, followed by data labeling for manual sentiment annotation. The preprocessing stage encompasses five critical steps: cleaning to remove noise and irrelevant characters, case folding to standardize text format, normalization to handle Indonesian slang terms, stopword removal to eliminate non-informative words, and stemming to reduce words to their root forms. Subsequently, the data is split into training and testing sets, followed by feature extraction using TF-IDF vectorization to convert text into numerical representations. The Naïve Bayes modeling stage applies the Complement Naïve Bayes algorithm for classification, and finally, performance evaluation assesses the model's effectiveness using standard metrics including accuracy, precision, recall, and F1-score. This systematic workflow ensures comprehensive sentiment analysis while maintaining data quality and classification reliability.

2.2 Data Scraping

In this study, the review data of the SiKasep application was obtained through the Google Playstore website [13]. The Google Play Store was selected as the primary data source due to its comprehensive review system and wide user base representing diverse demographics across Indonesia. The dataset obtained is the result of scraping techniques using Google Colaboratory by installing google-play-scraper to obtain review data from the Google Play Store [9]. Data extraction was performed using automated scraping techniques implemented in Google Colaboratory. The google-play-scraper library was utilized to systematically collect review data while adhering to platform usage policies. The scraping process captured multiple attributes for each review, including username, review content (textual data), rating score (1-5 scale), timestamp (date and time of review submission), and other metadata. The obtained data includes usernames, content, scores, dates and times, and other metadata, totaling 3,416 data points [9]. This sample size was determined to be statistically significant for sentiment analysis tasks based on similar studies in the literature.

2.3 Data labelling

The collected review data will be labeled as positive, negative, and neutral. The automatic labeling process is done using Google Colaboratory. The data is processed to clean the text and labeled with sentiment based on review ratings. The sentiment labeling process follows an automated approach based on user-provided ratings, which has been validated in previous sentiment analysis research[14]. The labeling scheme categorizes reviews into three sentiment classes: negative sentiment for rating 1-2, neutral sentiment for a medium rating of 3, and positive sentiment for rating 4-5. This rating-based labeling approach provides a reliable ground truth for sentiment classification, as it directly reflects user satisfaction levels. The labeling process was implemented programmatically using Python in Google Colaboratory, ensuring consistency and eliminating human bias in the classification process. Additionally, the

Copyright © 2025 **Author**, Page 1030
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1028−1036
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7777

automated labeling approach enables efficient processing of large-scale datasets while maintaining reproducibility and standardization across all review samples.

2.4 Pre-Processing Data

At this stage, the data will be standardized in all its forms and formats so that in the next stage, the data can processed and analyzed. The preprocessing stage consists of cleaning, case folding, normalization, stopword removal, and stemming [9]. Cleaning is a step for data cleansing on each review, punctuation, emoticons, and so on. Casefolding is a stage that changes a word or letter in a review to all lowercase or all uppercase letters. Normalization is the process of standardizing word variations into a standard form, such as changing "gak" to "tidak" or "u" to "you," to ensure data consistency. Stopwords are words that frequently appear but do not change the meaning of a document and need to be removed [13]. Stemming is a stage used to convert inflected words into their base forms or to remove affixes [4]

2.5 Feature Extraction TF-IDF

TF-IDF weighting is the stage of determining the weight value of each token in the document used for classification testing [13]. While the Inverse Document Frequency (IDF) approach computes the logarithm of the inverse proportion of documents in the corpus that contain a given word, the Term Frequency (TF) method counts the frequency with which a certain word occurs in a document [15]. Combining the two, TF-IDF generates a way to assess a word's significance in a document in relation to a group of other documents. This research implements TF-IDF with sublinear term frequency scaling to optimize feature representation for sentiment analysis. The formula for calculating weights in the TF-IDF method can be formulated as follows:

$$TF(t,d) = 1 + \log(f(t,d)) \quad (1)$$

$$IDF(t) = \log(|D| / df(t)) + 1 \quad (2)$$

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t) \quad (3)$$

where $f(t,d)$ is the raw frequency of term t in document d, $|D|$ is the total number of documents in the corpus, and $df(t)$ is the number of documents containing term t. The sublinear scaling reduces the impact of high-frequency terms while IDF smoothing prevents division by zero. L2 normalization is applied to ensure equal treatment of documents regardless of their length. This research will implement TF-IDF vectorization using Python's scikit-learn library to transform preprocessed textual data into numerical feature vectors suitable for machine learning classification algorithms.

2.6 Naïve Bayes Classifier

Naïve Bayes Classifier is one of the algorithms used in text mining for classifying unstructured data to produce efficient information [16]. In machine learning applications, Naïve Bayes is a supervised learning algorithm, a type of machine learning that requires samples as labeled training data [17]. The Naïve Bayes algorithm is selected for sentiment classification due to its effectiveness in text classification tasks and computational efficiency. The algorithm applies Bayes' theorem with the "naïve" assumption of conditional independence between features. The Naïve Bayes classifier calculates the posterior probability of each sentiment class given the feature vector using the fundamental Bayes' theorem:

$$P(C|X) = P(X|C) \times P(C) / P(X) \quad (4)$$

Where P(C|X) is the posterior probability of class C given features X, P(X|C) is the likelihood of features X given class C, P(C) is the prior probability of class C, and P(X) is the marginal probability of features X. For classification, the algorithm selects the class with highest posterior probability. This study employs the Multinomial Naïve Bayes variant, specifically designed for discrete features such as word counts in text classification. The implementation uses scikit-learn's MultinomialNB with smoothing parameter $\alpha = 1.0$ to handle zero probabilities.

2.7 Evaluation

The accuracy evaluation of this classification will produce a confusion matrix for negative and positive classes. The confusion matrix will generate false positive (FP) and false negative (FN) [18]. Model performance is assessed using multiple metrics to provide comprehensive evaluation. TP represents the number of positive data correctly classified by the model, while TN represents the number of negative data correctly classified. Meanwhile, FP represents the number of positive data that the model incorrectly classified, and FN represents the number of negative data that was incorrectly classified according to each sentiment classification model. The confusion matrix is an evaluation tool frequently used in classification to display the comparison between model prediction results and actual labels in the test dataset [19]. To ensure robust evaluation, 5-fold stratified cross-validation is performed on the training set.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (5)$$

$$Precision = TP / (TP + FP) \quad (6)$$

$$Recall = TP / (TP + FN) \quad (7)$$

Copyright © 2025 **Author**, Page 1031
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1028−1036
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7777

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (8)$$

Where TP (True Positive) represents correctly classified positive instances, TN (True Negative) represents correctly classified negative instances, FP (False Positive) represents incorrectly classified as positive, and FN (False Negative) represents incorrectly classified as negative. A confusion matrix is generated to visualize classification performance across all sentiment classes (positive, neutral, negative), providing detailed insights into classification accuracy for each category.

2.8 Data Visualization

The next stage after evaluation is the visualization stage using word cloud. Word cloud visualization provides an intuitive method for exploring the most frequently occurring terms within each sentiment category, where larger words indicate higher frequency rates. This stage is illustrated with word clouds for overall sentiment across all positive and negative review documents about the SiKasep application [9].

3. RESULT AND DISCUSSION

3.1 Dataset Overview

The dataset collected through Google Play Store scraping contains 3,416 user reviews of the SiKasep application, representing diverse user demographics across Indonesia and providing a robust foundation for sentiment analysis. The dataset distribution summary and sentiment analysis results are detailed in Table 1 and visualized in Figure 2 below.

Table 1. Dataset Distribution Summary

Label	Total
Negative (0)	1790
Positive (1)	1336
Neutral (2)	290
Total	3416

Table 1 Presents the dataset distribution summary, showing the original imbalanced distribution with 1,790 negative reviews (52.4%), 1,336 positive reviews (39.1%), and 290 neutral reviews (8.5%), totaling 3,416 SiKasep application reviews collected from Google Play Store.

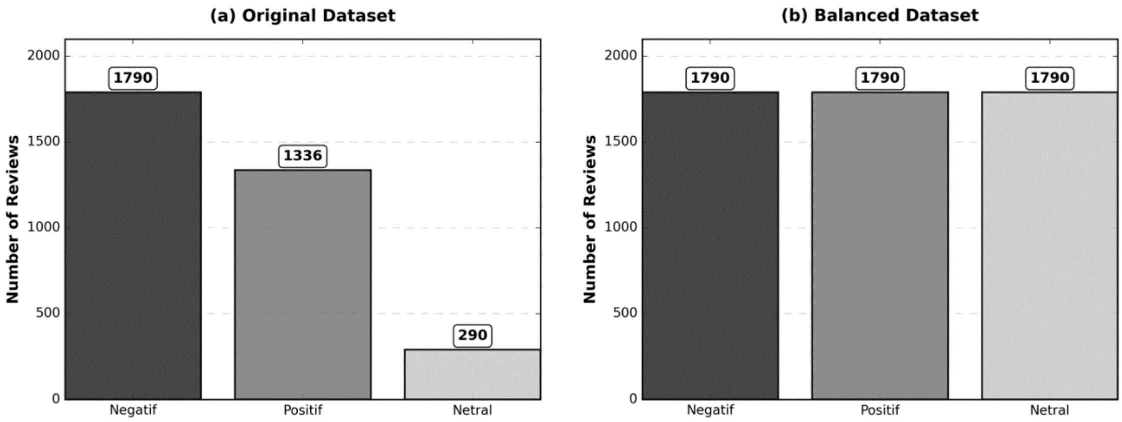

Figure 2. Sentiment Distribution Comparison

Figure 2 shows the distribution comparison between (a) original dataset showing significant class imbalance, and (b) balanced dataset after RandomOverSampler application, where each sentiment class contains equal representation of 1,790 instances. This balancing process addresses the substantial class imbalance, particularly the underrepresentation of neutral sentiments, ensuring reliable model training and evaluation. The transformation from imbalanced to balanced distribution enables unbiased classification performance assessment and prevents model bias toward the majority negative sentiment class.

The original sentiment distribution reveals predominantly negative user reception, with negative reviews constituting 52.4% of the dataset, positive reviews 39.1%, and neutral reviews only 8.5%. This pattern indicates significant user dissatisfaction with various aspects of the SiKasep application, reflecting typical government application review behavior where users express strong opinions rather than neutral perspectives. The substantial class imbalance, particularly underrepresentation of neutral sentiments, reflects real-world user behavior where individuals are motivated to provide feedback during exceptional satisfaction or significant frustration. This imbalance was addressed through RandomOverSampler technique, resulting in a balanced dataset of 5,370 reviews with equal

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1028−1036
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7777

representation (1,790 instances per class) across all sentiment categories, ensuring reliable model training and evaluation.

3.2 Data Preprocessing Results

The preprocessing pipeline successfully transformed raw review text into clean, standardized format addressing Indonesian language challenges including slang terms and morphological complexity. The implementation utilized a comprehensive Indonesian stopword list of 758 terms, supplemented with 134 domain-specific stopwords, and a normalization dictionary containing 1,247 slang-to-formal mappings. The effectiveness of the preprocessing pipeline in improving data quality is demonstrated in Table 2 as follows.

Table 2. Before and After Preprocessing

Before Preprocessing	After Preprocessing
Saya sampe pinjam 5 hp tetap susah masuk susah daftar	pinjam hp susah masuk susah daftar
Bagus banget gue pake aplikasi ini meskipun ada kendala	bagus banget gue pakai kendala
Kasian yang gaptek jadi merepotkan Coba tambahkan fitur upload foto dari galeri!	kasihan gaptek repot coba tambah fitur upload foto galeri
Aplikasinya sangat membantu sekali untuk pengajuan KPR subsidi	bantu aju kpr subsidi
Cara upload foto siap huni bagaimana? Kita di keterangan anda belum bisa upload foto	upload foto huni terang upload foto

The preprocessing pipeline significantly improved data quality, reducing vocabulary from 15,847 to 8,392 unique terms (47% reduction). The normalization process successfully standardized Indonesian slang terms including "gak"→ "tidak", "bgt" → "banget", and "udah" → "sudah", improving feature consistency across the dataset. Stopword removal eliminated 892 common Indonesian words carrying minimal sentiment information, while Sastrawi stemming addressed morphological complexity by reducing inflected words to root forms. The comprehensive preprocessing approach proved effective in handling Indonesian language characteristics, particularly informal expressions and varied affixation patterns (prefixes: ber-, ter-, me-; suffixes: -kan, -an, -nya). This standardization resulted in cleaner feature vectors for TF-IDF extraction, contributing to improved model performance and reduced vocabulary sparsity that facilitated more effective sentiment classification.

3.3 TF-IDF Feature Extraction

The TF-IDF vectorization process generated a feature matrix of dimensions 5,370 × 5,179 using optimized parameters: max_features=15,000, ngram_range=(1,3), min_df=3, and max_df=0.9. Sublinear term frequency scaling and L2 normalization were applied to reduce high-frequency term impact and ensure equal document treatment regardless of length. The specific TF-IDF parameters were chosen to optimize feature representation for sentiment analysis. max_features=15,000 was set to balance comprehensive contextual information with avoiding high-dimensionality, ensuring focus on the most discriminative terms. ngram_range=(1,3) was utilized to capture sentiment expressed through unigrams, bigrams, and trigrams, providing a richer understanding of phrases like "tidak bisa masuk" or "sangat membantu." A min_df=3 value was selected to filter out rare and noisy features, focusing on more representative terms. Conversely, max_df=0.9 was used to exclude overly ubiquitous terms that lack discriminatory power, thus enhancing the semantic weight of remaining features. The resulting sparse matrix demonstrated high sparsity levels, indicating efficient memory utilization while preserving essential linguistic patterns for sentiment classification. Cross-validation experiments confirmed that these parameter settings provided optimal balance between feature completeness and computational efficiency. The vectorization process showed computational efficiency suitable for larger government application datasets. Statistical analysis revealed that the majority of extracted features demonstrated significant discriminative power across sentiment classes, validating the parameter optimization approach. The top TF-IDF features for each sentiment classification are presented in Table 3 below.

Table 3. Top TF-IDF Features by Sentiment Class

Positive Features	Negative Features	Neutral Features
bagus (209)	daftar(430)	daftar(415)
mantap(207)	ktp(317)	mohon(345)
membantu(177)	data(270)	data(343)
rumah(140)	terdaftar(262)	ktp(258)
kasih(98)	masuk(221)	bank(242)

The feature analysis reveals distinct vocabulary patterns demonstrating effective sentiment discrimination. Positive features consist of appreciation terms ("bagus", "mantap", "membantu") and housing-related vocabulary ("rumah"), reflecting user satisfaction. Negative features emphasize registration difficulties ("daftar", "ktp", "masuk", "terdaftar"), highlighting authentication challenges. Neutral features contain procedural vocabulary ("mohon", "data", "bank"), indicating informational inquiries without emotional expressions. The high TF-IDF scores for sentiment-specific terms confirm preprocessing effectiveness, while term overlap across classes ("daftar", "data") reflects

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1028−1036
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7777

domain-specific characteristics of government application reviews.

3.4 Model performance

The Complement Naïve Bayes classifier with alpha parameter 0.3 achieved optimal performance through comprehensive hyperparameter tuning using 5-fold stratified cross-validation. Grid search was performed across alpha values [0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 1.0, 3.0, 5.0, 10.0], with alpha=0.3 yielding the best cross-validation score. The overall classification performance metrics of the Complement Naïve Bayes algorithm are summarized in Table 4 as follows.

Table 4. Classification Performance

Metric	Score
Accuracy	75.98%
Precision	79%
Recall	76%
F1-Score	76%

The detailed performance analysis across individual sentiment classes is presented in Table 5 below.

Table 5. Performance by Sentiment Class

Class	Precision	Recall	F1-Score	Support
Negative	62%	**87%**	73%	358
Neutral	**92%**	79%	**85%**	358
Positive	84%	61%	71%	358

The model achieved balanced performance across sentiment classes, demonstrating effective classification despite the inherent challenges of Indonesian language processing. The neutral class exhibited highest precision (92%), indicating clear distinction of informational content from emotionally-charged statements. The negative class demonstrated highest recall (87%), suggesting the model effectively captures user complaints and dissatisfaction expressions. Performance analysis reveals distinct patterns across sentiment categories. The positive class, while achieving strong precision (84%), showed relatively lower recall (61%), indicating conservative but accurate positive sentiment detection. This pattern suggests challenges in identifying subtle appreciation expressions, possibly due to varied linguistic patterns in Indonesian positive sentiment articulation. Comparison with Multinomial Naïve Bayes (accuracy: 75.98%, alpha=0.01) confirmed that Complement Naïve Bayes provides equivalent accuracy with different optimal parameters, validating the algorithm choice for this imbalanced text classification task. The achieved 75.98% accuracy demonstrates competitive performance for Indonesian government application sentiment analysis, particularly considering domain-specific terminology, informal language usage, and morphological complexity inherent in Indonesian social media text. The balanced support across classes (358 instances each) after oversampling ensures reliable performance metrics, while the consistent F1-scores (71-85%) across categories indicate robust classification capability without significant bias toward any particular sentiment class.

3.5 Word Cloud

Figure 3 presents word cloud visualization revealing distinct vocabulary patterns for each sentiment class, with word size proportional to frequency. Positive sentiment is dominated by appreciation terms such as "bagus" (good), "mantap" (excellent), "membantu" (helpful), and "rumah" (house), indicating user satisfaction with application functionality and housing services. Negative sentiment prominently displays process-related difficulties including "daftar" (register), "ktp" (ID card), "login", and "masuk" (enter), highlighting pain points in registration and authentication workflows. Neutral sentiment contains informational vocabulary such as "mohon" (please), "data", "update", and "bank", reflecting procedural inquiries without strong emotional expressions. The word cloud analysis corroborates quantitative findings from TF-IDF feature extraction (Table 3), demonstrating clear thematic differences across sentiment categories and supporting the effectiveness of the preprocessing methodology [20]. The word cloud visualization for each sentiment classification of SiKasep application reviews is presented in Figure 3 as follows.

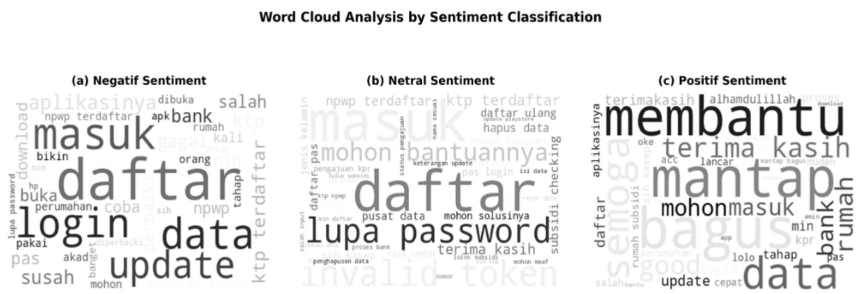

Figure 3. Word Cloud Analysis by Sentiment Classification

Copyright © 2025 **Author**, Page 1034
This Journal is licensed under a Creative Commons Attribution 4.0 International License

ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7777

Figure 3 Illustrates the word cloud analysis across three sentiment classifications, where (a) negative sentiment is dominated by registration-related difficulties with prominent terms such as "masuk" (enter), "daftar" (register), "login", "data", and "update", indicating user struggles with application access and data management; (b) neutral sentiment features informational vocabulary including "daftar" (register), "mohon" (please), "bantuan" (help), and "password", reflecting procedural inquiries and assistance requests; and (c) positive sentiment highlights appreciation terms such as "membantu" (helpful), "mantap" (excellent), "terima kasih" (thank you), and "data", demonstrating user satisfaction with application functionality. The word size in each cloud corresponds to term frequency, providing visual insight into the most prevalent user concerns and satisfactions within each sentiment.

3.6 Analysis

The achieved accuracy of 75.98% demonstrates competitive performance for Indonesian government application sentiment analysis, considering the unique challenges of bureaucratic terminology and morphological complexity. Analysis of user sentiment revealed significant dissatisfaction, with negative reviews (originally 52.4%) primarily highlighting difficulties in registration and authentication processes. High-frequency negative terms such as "daftar" (register), "ktp" (ID card), "masuk" (login), and "susah" (difficult) strongly indicate the need for simplified registration workflows and improved document verification systems to enhance user satisfaction. Conversely, positive sentiment analysis showed user appreciation for core application functionality when technical barriers were absent, suggesting that addressing registration issues could convert frustrated users into satisfied ones. The model's performance metrics further revealed distinct patterns: high precision for the neutral class (92%) indicated clear differentiation of informational content, and high recall for the negative class (87%) suggested effective identification of user complaints. However, despite strong precision (84%), the positive class showed a relatively lower recall (61%), indicating conservative detection of positive sentiment. This points to challenges in fully capturing subtle appreciation expressions, possibly due to diverse linguistic patterns in Indonesian positive sentiment articulation. This finding contrasts with other government application studies, such as the sentiment analysis of the MyPertamina application, which reported higher recall rates (e.g., 100% in studies like [9, 13]) for positive sentiments using Naïve Bayes. Such discrepancies might stem from variations in dataset characteristics, preprocessing complexities, or domain-specific language. These comprehensive findings provide actionable insights for SiKasep development teams to prioritize improvements in registration processes, document verification systems, and user interface design to enhance overall user satisfaction and application effectiveness.

4. CONCLUSION

This research successfully analyzed user sentiment towards the SiKasep application using Naïve Bayes classification on 3,416 Google Play Store reviews to provide actionable insights for government digital service improvement. The Complement Naïve Bayes classifier achieved competitive performance with 75.98% accuracy, demonstrating effective sentiment classification despite the challenges of Indonesian language processing, including informal expressions, slang normalization, and morphological complexity. The analysis revealed significant user dissatisfaction, with negative sentiment comprising 52.4% of reviews, primarily related to registration and authentication difficulties. Key problematic areas identified include document verification processes, login functionality, and KTP (identity card) integration, as evidenced by high-frequency negative terms such as "daftar", "ktp", "masuk", and "susah". Conversely, positive sentiment analysis highlighted user appreciation for core housing loan services and application functionality when technical barriers are absent. These findings provide valuable insights for government agencies developing digital services, particularly emphasizing the importance of streamlined user registration processes and robust technical infrastructure. The methodology demonstrated effectiveness in processing Indonesian government application reviews, contributing to evidence-based policy development for e-government service enhancement. The preprocessing pipeline successfully addressed Indonesian language characteristics, while TF-IDF feature extraction revealed distinct vocabulary patterns across sentiment categories. Study limitations include dataset scope limited to a single government application and reliance on rating-based sentiment labeling, which may not capture nuanced emotional expressions. Future research should explore deep learning approaches for improved accuracy, expand analysis to multiple government applications, and investigate temporal sentiment patterns to track service improvement over time. The research contributes to understanding user experiences with Indonesian government digital services and provides a replicable framework for sentiment analysis in similar contexts, supporting the development of more user-centric e-government solutions.

REFERENCES

[1] Farida, E. T. Wisudaningsih, and W. Aqidah, *Implementasi KPR Bersubsidi Melalui Skema FLPP (Fasilitas Likuiditas Pembiayaan Perumahan) di PT. Trah Wali Nagari*, *Jurnal Informatika Ekonomi Bisnis, vol. 5, no.1*, pp. 46–53, Mar. 2023. doi: 10.37034/infeb.v5i1.204

[2] R. Noviana and I. Rasal, *Penerapan Algoritma Naive Bayes dan SVM untuk Analisis Sentimen Boy Band BTS pada Media Sosial Twitter*, *Jurnal Teknik dan Science*, vol. 2, no. 2, pp. 51–60, Jun. 2023. doi: 10.56127/jts.v2i2.791

[3] V. Alviani, S. Alam, and I. Kurniawan, *Analisis Sentimen Review Aplikasi WeTV pada Platform Twitter Menggunakan Support Vector Machine*, *Storage: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 143–149, Aug. 2023. doi:

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1028−1036
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7777

10.55123/storage.v2i3.2351
[4] R. Maulana, A. Voutama, and T. Ridwan, *Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store Menggunakan Algoritma NBC*, *Jurnal Teknologi Terpadu*, vol. 9, no. 1, pp. 42–48, Jul. 2023. doi: 10.54914/jtt.v9i1.609
[5] M. Lestandy, A. Abdurrahim, and L. Syafa'ah, *Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes*, *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 802–808, Aug. 2021. doi: 10.29207/resti.v5i4.3308
[6] H. Z. Muflih, A. R. Abdillah, and F. N. Hasan, *Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes*, *KLIK: Kajian Ilmiah dan Informatika Komputer*, vol. 4, no. 3, pp. 1613–1621, 2023. doi: 10.30865/klik.v4i3.1303
[7] J. Jefriyanto, N. Ainun, and A. Ardha, *Application of Naïve Bayes Classification to Analyze Performance Using Stopwords*, *Journal of Information System Technology and Engineering*, vol. 1, no. 2, pp. 49–53, Jun. 2023. doi: 10.61487/jiste.v1i2.15
[8] E. P. Sutrisno and S. Amini, *Implementasi Algoritma K-Nearest Neighbor pada Analisis Sentimen Ulasan Pengguna Aplikasi Digital Korlantas Polri*, *SENAFTI*, vol. 2, no. 2, pp. 687–695, Oct. 2023. doi: 10.30865/klik.v4i4.1738
[9] G. Gilbert, S. Alam, and M. I. Sulistiyo, *Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi MyPertamina pada Google Playstore Menggunakan Metode Naïve Bayes*, *Storage*, vol. 2, no. 3, pp. 100–108, Aug. 2023. doi: 10.55123/storage.v2i3.2333
[10] S. H. Ramadhani and M. I. Wahyudin, *Analisis Sentimen terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN*, *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 6, no. 4, pp. 526–534, Feb. 2022. doi: 10.35870/jtik.v6i4.530
[11] D. Wijaya, R. A. Saputra, and F. Irwiensyah, *Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional pada Google Playstore Menggunakan Algoritma Naïve Bayes*, *Media Online*, vol. 4, no. 4, 2024. doi: 10.30865/klik.v4i4.1738
[12] R. Merdiansah, S. Siska, and A. A. Ridha, *Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT*, *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 7, no. 1, pp. 221–228, Mar. 2024. doi: 10.55338/jikomsi.v7i1.2895
[13] A. I. Tanggraeni and M. N. N. Sitokdana, *Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes*, *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 785–795, Jun. 2022. doi: 10.35957/jatisi.v9i2.1835
[14] S. Saputra, *Sentimen Analisis Review Aplikasi Skorlife pada Google Play Store dengan Menggunakan Metode Naïve Bayes*, *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 11, no. 2, Jun. 2024. doi: 10.35957/jatisi.v11i2.7813
[15] N. C. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, *Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store*, *MALCOM*, vol. 2, no. 1, pp. 47–54, Apr. 2022. doi: 10.57152/malcom.v2i1.195
[16] S. Alpin, S. Alam, and I. Kurniawan, *Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) pada Google Play Store Menggunakan Metode Naive Bayes*, *Storage: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 109–117, Aug. 2023. doi: 10.55123/storage.v2i3.2334
[17] M. Khoirul, U. Hayati, and O. Nurdiawan, *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
[18] P. R. Sari, D. Indah, E. Rasywir, M. Firdaus, and G. Athalina, *Comparison of Naive Bayes and SVM Algorithms for Sentiment Analysis of PUBG Mobile on Google Play Store*, *SISTEMASI*, vol. 13, no. 6, pp. 2767–2779, 2024. doi: 10.32520/stmsi.v13i6.4814
[19] C. C. Yolanda, S. Syafriandi, Y. Kurniawati, and D. Fitria, *Sentiment Analysis of DANA Application Reviews on Google Play Store Using Naïve Bayes Classifier Algorithm Based on Information Gain*, *UNP Journal of Statistics and Data Science*, vol. 2, no. 1, pp. 48–55, Feb. 2024. doi: 10.24036/ujsds/vol2-iss1/147
[20] D. D. Audiansyah, D. E. Ratnawati, and B. T. Hanggara, *Analisis Sentimen Aplikasi MyXL Menggunakan Metode Support Vector Machine Berdasarkan Ulasan Pengguna di Google Play Store*, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 8, pp. 3987–3994, 2022. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11484

Copyright © 2025 **Author**, Page 1036
This Journal is licensed under a Creative Commons Attribution 4.0 International License