Jurnal Teknik Informatika (JUTIF) Vol. 7, No. 3, June 2026, Page. 2798-2811
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

Optimizing Naive Bayes for Sentiment Analysis of M-Passport Reviews Using N-Gram and Synthetic Minority Over-sampling Technique

Devia Kartika*[1], Sarjon Defit[2]

[1,2,3]Universitas Putra Indonesia YPTK Padang, Indonesia

Email: [1]deviakartika@upiyptk.ac.id

Received : Apr 15, 2026; Revised : May 7, 2026; Accepted : May 12, 2026; Published : Jun 15, 2026

Abstract

The diverse user perceptions and increasing number of negative reviews of the M-Passport application indicate the need for sentiment analysis-based evaluation to more accurately measure the quality of digital immigration services. This study aims to analyze user sentiment towards the M-Passport application using an optimized Naïve Bayes classification model. Review data was obtained through web scraping from various digital platforms and processed using text preprocessing, TF-IDF feature extraction, N-Gram representation, and the Synthetic Minority Over-sampling Technique (SMOTE) technique to address data representativeness. The proposed model classifies user reviews into positive, neutral, and negative sentiment categories. Test results show that optimization using N-Gram and SMOTE successfully improved model performance, with accuracy increasing from 61% to 77.51%, precision from 0.75 to 0.78, recall from 0.53 to 0.78, and F1-score from 0.50 to 0.77. These results demonstrate that the combination of feature engineering and data balancing can improve text context representation and sentiment classification stability across multiple classes. Furthermore, sentiment analysis successfully identified key factors contributing to user dissatisfaction, such as technical constraints, feature limitations, and application difficulty. These results demonstrate that the proposed approach is effective in supporting data-driven evaluation to improve the quality of digital immigration services.

Keywords : *Digital Public Services, M-Passport, N-Gram, Naïve Bayes, Sentiment Analysis, SMOTE, TF-IDF.*

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License

1. INTRODUCTION

As time progresses, technological developments continue to accelerate and become an inseparable part of human life. Technological advancement evolves alongside scientific development and increasingly influences various aspects of society, particularly in improving communication, information exchange, and public services [1].

The development of internet-based technology also shows significant growth in Indonesia, as reflected by the increasing number of internet users each year. In 2017, Indonesia ranked fifth in the world in terms of internet users, with an internet penetration rate of 50.4%. Along with population growth and increasing technological demands, Indonesia later rose to fourth place with an internet penetration rate of 77.75% [2]. This rapid technological growth has encouraged transformation in various sectors, including economics, social services, tourism, and government administration.

The government strongly supports digitalization as part of efforts to improve the effectiveness, efficiency, and transparency of public services. Digitalization refers to the process of converting information or data from physical forms into digital forms that can be processed through information technology. The concept of digitalization has been widely implemented across various sectors, such as business, education, and government services. The implementation of digital concepts aims to improve and encourage the creation of effective, efficient, and transparent public services in order to achieve

Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

better governance (good governance) [3]. One implementation of digital transformation in the Indonesian immigration sector is the development of an online-based application called M-Passport, which aims to facilitate online passport registration services [4].

M-Passport is an online application developed by the Directorate General of Immigration of Indonesia. The primary objective of this application is to reduce long queues and lengthy service times in passport application processes at immigration offices in Indonesia. Although the M-Passport application offers various conveniences in digital immigration services, user reviews on multiple digital platforms indicate varying perceptions regarding the quality and usability of the application. Some users provide positive feedback related to service accessibility and convenience, while others express dissatisfaction caused by technical problems, application instability, limited features, and difficulties during online service processes. The large number of negative reviews indicates inconsistencies between expected service quality and actual user experience. Therefore, an evaluation based on user reviews is necessary to identify public perceptions and improve the quality of digital immigration services. Sentiment analysis is a branch of Natural Language Processing (NLP) that aims to identify and classify opinions, emotions, or user perceptions expressed in textual data into positive, neutral, and negative sentiment categories. Various machine learning methods have been widely applied in sentiment analysis studies, including Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes [5], [6], [7].

Among these methods, Naïve Bayes is considered one of the most effective classification algorithms because of its simple structure, computational efficiency, and strong performance in text classification tasks [6]. Research conducted by Putri [5] compared Naïve Bayes and Support Vector Machine optimized using Particle Swarm Optimization (PSO) for sentiment analysis of OVO e-wallet reviews and demonstrated that Naïve Bayes achieved higher accuracy than SVM. In addition, research by Quadri and Selvakumar [6] reported that Naïve Bayes achieved an average accuracy of 85.92% across various sentiment analysis datasets. However, several other studies showed that classification performance may vary depending on dataset characteristics and feature extraction techniques [7].

In sentiment analysis, preprocessing and feature extraction stages play important roles in improving classification performance. Techniques such as TF-IDF and N-Gram are widely used to capture contextual relationships between words and enrich textual feature representations [[8], [9]. Several studies reported that the N-Gram technique significantly improves sentiment classification performance because it preserves contextual relationships between adjacent words [10], [11]. In addition, preprocessing stages such as cleansing, tokenizing, stopword removal, and stemming help reduce noise and improve text consistency before classification [12] - [13] . Another challenge commonly encountered in sentiment analysis is class imbalance, where the number of samples in each sentiment category is uneven. Imbalanced datasets may cause classification models to become biased toward majority classes and reduce the ability of models to recognize minority classes accurately. To overcome this issue, data balancing techniques such as Synthetic Minority Over-sampling Technique (SMOTE) are widely used to improve model stability and classification performance [14].

Although previous studies have demonstrated the effectiveness of Naïve Bayes in sentiment analysis across various domains, most studies mainly focused on product reviews, social media platforms, and e-commerce applications. In addition, previous studies generally emphasized algorithm comparison without specifically addressing class imbalance problems in public service application reviews. Research related to sentiment analysis on the M-Passport application is still limited and has not comprehensively integrated TF-IDF feature extraction, N-Gram representation, and SMOTE data balancing within a single classification framework.

Therefore, there remains a research gap regarding the optimization of sentiment classification performance for digital public service applications through the simultaneous integration of contextual feature extraction and data balancing techniques.

Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

This study collected user review data from the M-Passport application using web scraping techniques. Web scraping refers to the process of reading, retrieving, collecting, or crawling data and documents from websites automatically [8]. The purpose of this study is to analyze user sentiment toward the M-Passport application, particularly regarding aspects such as ease of use, service features, speed, and application security. From a methodological perspective, this study proposes an optimized Naïve Bayes-based sentiment analysis model by integrating TF-IDF feature extraction, N-Gram representation, and the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance problems in the dataset. The proposed approach aims to improve contextual text representation and enhance sentiment classification performance for positive, neutral, and negative reviews. The novelty of this study lies in the integration of contextual feature extraction and data balancing techniques within the domain of digital immigration services. In addition, this study contributes to providing data-driven insights regarding user perceptions to support the improvement of digital public service quality in the immigration sector.

2. METHOD

This research methodology consists of several systematic stages designed to obtain accurate sentiment analysis results. The stages include data collection, data preprocessing, feature extraction, data balancing, data analysis, and sentiment classification using the Naive Bayes algorithm. The complete process flow of the proposed method is illustrated in Figure 1.

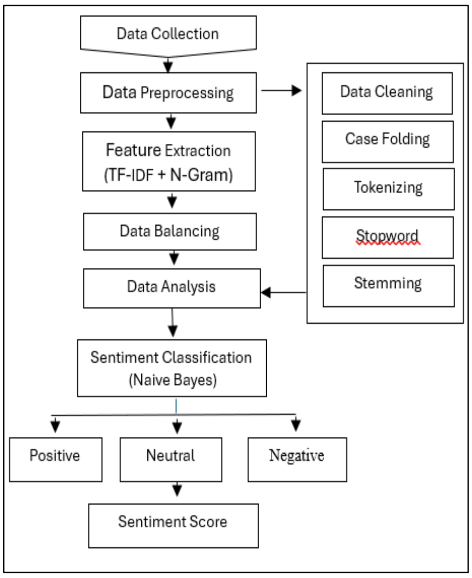


Figure 1. Research Stages

Figure 1 illustrates the stages of the sentiment analysis process carried out in this study. The process begins with the data collection stage, which functions to collect review or opinion data from relevant sources. Next, data preprocessing is carried out, consisting of several steps, namely data cleaning, case folding, tokenizing, stopword removal, and stemming to obtain clean text that is ready for processing. The next stage is feature extraction using the TF-IDF and N-Gram methods to convert text data into numeric representations that can be processed by the algorithm. After that, data balancing is performed to address class imbalance in the data. The process continues with data analysis to evaluate the results of the data representation before classification. The final stage is sentiment classification using the Naive Bayes algorithm, which groups data into three sentiment categories: positive, neutral, and negative. The classification results are then used to determine a sentiment score that represents the opinion tendencies of the analyzed data.

Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

2.1 Naïve Bayes Classifier Algorithm

The Naïve Bayes classification method is a technique used in sentiment analysis. This approach has theoretical advantages in terms of data consistency and classification computation. Naïve Bayes is commonly used in classification techniques, especially in social media platforms, using variations such as Unigram Naïve Bayes, Multinomial Naïve Bayes, and Maximum Entropy Classification [15]. This algorithm has a simpler structure than other algorithms and often produces quite high accuracy. Naïve Bayes operates on the assumption that each input attribute is conditionally independent of the others. This algorithm is very flexible and can handle large-scale data. Based on the principle of "learning by doing," Naïve Bayes uses preprocessed data with relevant features as input for the training process. The main feature of Naïve Bayes classification is its ability to generate strong hypotheses about various conditions or events. The calculation of group probabilities in Naïve Bayes relies on the Bayesian algorithm, which utilizes special equations [16]. Once trained, this algorithm is able to analyze the polarity of a text, such as determining whether the text is positive or negative in tone. Bayes' theorem states that the posterior probability, P(A|B) is calculated from P(A), P(B), and P(B|A) as defined in the following equation, [6] ;

$$P(A \mid B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

Where:

P(A|B) is the posterior probability, P(B|A) is the likelihood probability, P(A) is the prior probability, P(B) is the evidence probability.

2.2 SMOTE Data Balancing

Synthetic Minority Over-sampling Technique (SMOTE) is a data balancing technique used to overcome class imbalance problems by generating synthetic samples from minority classes. SMOTE works by selecting one sample from the minority class and identifying its k-nearest neighbors[17], [18]. New synthetic samples are generated between the selected sample and its neighboring samples using interpolation. The synthetic sample generation process can be formulated as follows:

$$X_{\text{new}} = X_i + \delta(X_{zi} - X_i) \quad (2)$$

Where:

- X_i : represents the minority class sample
- X_{zi} : represents one of the nearest neighbors
- δ : is a random value between 0 and 1
- X_{new} : is the generated synthetic sample

By applying SMOTE, the dataset distribution becomes more balanced, thereby improving the classification model's ability to recognize minority sentiment classes more accurately.

2.3. Confusion matrix

Confusion matrix is a tabular representation that describes the results of the classification of test data, including the number of correct and incorrect predictions [12], [19].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{F1-Score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (6)$$

Where TP represents True Positive, TN represents True Negative, FP represents False Positive, and FN represents False Negative.

2.4 Sentiment Analysis

Sentiment analysis is a branch of natural language processing (NLP) that studies a person's opinions, evaluations, and emotions towards a product, service, organization, or other topic expressed in text form. Sentiment analysis interprets and classifies user emotions (positive, negative, or neutral) about a subject in text data using text analysis. The steps commonly found in sentiment analysis text classification are (1) Determining the dataset domain, (2) Pre-processing, and (3) Transformation, [5].

3. RESULT

This study uses the Naïve Bayes Classifier algorithm to analyze user reviews of the M-Passport application published on the Google Play Store. The purpose of this study is to classify user reviews of the M-Passport application into positive, neutral, and negative sentiment categories, and to evaluate the accuracy, precision, and recall values of the Naïve Bayes Classifier algorithm.

3.1 Data Selection

The data used in this study comes from reviews published by M-Passport app users on the Google Play Store. Researchers collected 1000 reviews using the "most relevant" sorting method, scraping them using the Google-Play-scraper library.

3.2 Preprocessing

The text preprocessing process aims to clean and prepare data for easier processing by machine learning models. At this stage, several key processes are carried out, namely removing irrelevant characters (cleansing), converting all letters to lowercase (case folding), separating text into tokens (tokenizing), removing common words that have no important meaning (stopword removal or filtering), and converting words to their basic form (stemming). These stages significantly contribute to improving the accuracy and stability of sentiment classification models [6]. According to research [20], the comprehensive application of preprocessing stages such as case folding, tokenizing, stopword removal, and stemming has been proven to improve the performance of deep learning models in analyzing Indonesian language texts. Research [21] also shows that optimization of stemming and stopword removal techniques helps produce more consistent root word forms, thus supporting more accurate classification results. In addition, research [22] emphasized that the effective use of stemming can improve the classification accuracy of non-formal Indonesian conversational texts. Research [23] explained that the combination of preprocessing techniques such as cleaning, stemming, stopword removal, and N-Gram-based feature formation was proven to be able to improve the performance of sentiment analysis models on social media review data. Research [24] which shows that the consistent application of case folding and stemming can improve classification results in word embedding-based deep learning models. Therefore, the proper implementation of preprocessing stages, including the integration of N-Gram techniques, is a crucial component in the sentiment analysis pipeline because it can reduce noise, normalize text, enrich the context between words, and improve the model's ability to understand natural language more accurately.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

3.3 Dataset

In this study, the dataset was obtained from user reviews of the M-Passport application collected from the Google Play Store. Researchers collected 1,000 review data using web scraping, an automated technique for extracting data from web pages. The data collection process was carried out using the app_store_scraper library in the Python programming language with Google Colab as the development environment. The collected review data were stored in CSV (Comma Separated Values) format. Before the classification process, the dataset underwent preprocessing stages, including cleansing, case folding, tokenizing, stopword removal, and stemming. Furthermore, the dataset was divided into training and testing data using an 80:20 ratio, where 80% of the data were used for model training and 20% for model evaluation.

3.4 Cleansing

The cleaning stage is the process of removing words that do not affect the sentiment classification results. In a tweet document, there are various attributes that do not affect sentiment because almost all tweets contain these attributes. Examples of attributes that are considered unimportant include mentions that begin with '@', hashtags that begin with '#', links that begin with 'http' or 'bit.ly', and symbol characters such as ~@#$%^&*()+?<>,.?:{}[]|. These irrelevant attributes will be removed from the document and replaced with spaces [7].

3.5 Case Folding

The process of standardizing letterforms by converting all characters in a text to lowercase, so that the system does not distinguish between uppercase and lowercase letters. This step helps reduce redundancy in words that have different forms but the same meaning [24], [25].

3.6 Tokenizing

This is the process of breaking down sentences into word units or tokens. This stage serves to facilitate linguistic analysis at the word level [13] .

3.7 Stopwords

After the tokenization process, stopword removal is carried out, namely the removal of common words such as "to," "from," "want," and "ini" which do not make a significant contribution to the meaning of the document [14].

3.8 Stemming

The process of returning a word to its base form by removing affixes or endings. This process aims to unify various derived word forms into a single base form, thereby increasing consistency in text analysis [22].

3.9 Balancing Data

Balancing data is a preprocessing technique used to equalize the number of samples in each class so that the dataset distribution becomes more balanced[26]. This process is important because imbalanced datasets tend to cause classification models to favor the majority class, thereby reducing the ability of the model to correctly identify minority classes. Techniques such as under-sampling and over-sampling are commonly applied to improve classification performance and model generalization [27]. The stages of text preprocessing applied in this study are presented in Table 1 The preprocessing pipeline consists of dataset preparation, cleansing, case folding, tokenizing, stopword removal, and stemming.

Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

Table 1. Preprocessing Results

Preprocessing	Results
Dataset	To: Widodo Ekatjahjana, Director General of Immigration Agato P.P. Simamora, Director of Immigration Information Systems and Technology Your immigration m-passport application is incompetent. Please improve its quality until it receives at least the same 3-star review rating as the one on your shoulders. Please like this review so it rises to the top. So everyone knows the name of the person running this careless application.
Cleansing	To Widodo Ekatjahjana, Director General of Immigration, Agato PP Simamora, Director of Immigration Information Systems and Technology, Your immigration m-passport application is incompetent. Please improve the quality at least until the star review value here is the same as the one on your shoulders. Please like this review so that it rises to the top. So that all the people know the name of the person who leads this careless application.
Case Folding	to widodo ekatjahjana, director general of immigration, agato pp simamora, director of immigration information systems and technology, your immigration m-passport application is incompetent. please improve the quality at least until the star review value here is the same as the one on your shoulders. please like this review so that it rises to the top so that all the people know the name of the person who leads this careless application.
Tokenizing	[to, widodo, ekatjahjana, director general, immigration, agato, pp, simamora, director, system, and, information, technology, immigration, application, m-passport, immigration, you, this, is, incompetent, please, improve, the quality, at least, to, the, star, review, value, here, the same, as, the, one, on, your, shoulders, please, like, guys, review, this, so that, it, rises, to, the, top, so, all, the, people, know, who, the, name, of, the, person, who, leads, this, careless, application]
Stopwords	widodo ekatjahjana, director general of immigration, agato pp simamora, director of immigration information technology systems, m-passport immigration application is good, please improve the quality, at least a star review rating, please like the review at the top so that people know the name of the person leading the application, careless
Stemming	widodo ekatjahjana, director general of immigration, agato pp simamora, director of immigration information technology systems, m-passport immigration application, becus, please have a minimum quality level, star review score, please like, guys, review the most so people know the name of the person leading the careless application

Table 1 shows an example of the transformation process performed on user review data from the immigration m-passport application. The cleansing stage removes unnecessary symbols and adjusts sentence structures, while case folding converts all text into lowercase characters to maintain consistency. Tokenizing divides sentences into individual terms, followed by stopword removal to eliminate insignificant words. Finally, stemming converts words into their root forms so that the resulting features become more effective for machine learning processing.

3.10. N-Gram in Text Feature Extraction

N-grams are a method in Natural Language Processing (NLP) used to represent sequences of words or characters in text. This concept forms combinations of N consecutive items, such as unigrams (one word), bigrams (two words), and trigrams (three words), allowing the model to better understand the context between words [8]. This approach is widely used to improve the capabilities of text classification, feature extraction, and sentiment analysis. According to research [28], The application of the N-Gram feature in the 2024 Election sentiment analysis was able to increase the model accuracy to 85% with the Support Vector Machine (SVM) and Random Foresight algorithms. *Research* [10] also proved that the combination of N-Gram and Term Frequency–Inverse Document Frequency (TF-IDF)

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

is effective in analyzing product reviews in e-commerce, because it enriches the representation of word features. Meanwhile, research [11] shows that the use of N-Gram features can significantly improve the performance of various machine learning algorithms in financial news classification. In addition, research by [29]**,** found that the application of N-Gram in a Gaussian Naïve Bayes model to student app reviews improved the accuracy of sentiment prediction. In this study, [30] also confirmed the effectiveness of N-Gram in mapping public perceptions of the digital auction system in Indonesia. Meanwhile, and research [9]**,** Using the Automated Lexicon Senti N-Gram approach, it was demonstrated that the combination of the lexicon method with *N-Gram* features resulted in more precise sentiment classification of national policy texts. The application of N-Gram plays a crucial role in machine learning-based sentiment analysis pipelines, as it captures linguistic context, enriches text representation, and improves model accuracy across various analysis domains.

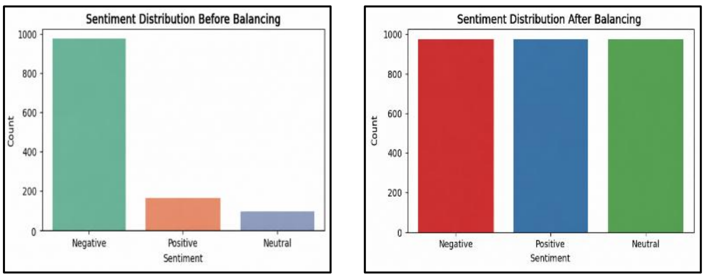
3.11 Evaluation

Model performance evaluation was conducted using accuracy, precision, recall, F1-score, and confusion matrix metrics. Accuracy is used to measure the percentage of correct predictions across the entire test data, while precision, recall, and F1-score are used to assess the model's performance on each sentiment class in more depth. Based on the test results, the model's accuracy increased from **61%** to **77.51%** after optimization through a combination of N-Gram and the SMOTE data balancing technique. Confusion matrix visualization shows that before optimization, the model tended to be biased towards the negative class, while after optimization the prediction distribution became more balanced across the three classes (negative, neutral, and positive). This improvement proves that the applied optimization strategy successfully improved the representation of text features and data balance, thereby enhancing the model's ability to classify user review sentiment more accurately and reliably.

4. DISCUSSIONS

The evaluation was conducted to assess the performance of the Naïve Bayes algorithm in classifying the sentiment of M-Passport application user reviews, before and after optimization using a combination of N-Gram and the SMOTE (Synthetic Minority Over-sampling Technique) data balancing technique. Before conducting model evaluation, the distribution of sentiment classes was analyzed to identify potential class imbalance in the dataset. Figure 2 illustrates the sentiment distribution before and after applying the SMOTE balancing technique.

Figure 2(a) shows that before balancing, the dataset was highly imbalanced, with the negative sentiment class dominating the dataset, while the positive and neutral classes contained significantly fewer samples. Such imbalance can cause the classification model to become biased toward the majority class and reduce its ability to recognize minority sentiment categories accurately. After applying SMOTE, as shown in Figure 2(b), the number of samples in each sentiment category became relatively balanced. This balancing process contributed to improving classification stability and reducing prediction bias in the sentiment analysis model.



(a) (b)

Figure 2. Sentiment Distribution (a)Before and (b)After SMOTE Balancing

Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

Model performance was measured using four main metrics: accuracy, precision, recall, and F1-score. The model testing results are presented in Table 2, which shows a comparison of model performance before and after the optimization process.

Table 2. Naive Bayes Evaluation Results before and after optimization

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes (Initial)	0.61	0.75	0.53	0.50
Naive Bayes + N-Gram + Balancing	0.7751	0.78	0.78	0.77

Based on the evaluation results in Table 2, the initial Naïve Bayes model achieved an accuracy of 61%, with a precision value of 0.75, recall of 0.53, and F1-score of 0.50. These results indicate that the initial model had difficulty recognizing a balanced distribution of data across sentiment classes. In particular, the low recall and F1-score values show that the model was less effective in identifying minority sentiment categories, especially positive reviews. This condition suggests that the initial model tended to be biased toward the majority class distribution.

To provide a more detailed analysis of model performance for each sentiment category, the classification report before optimization is presented in Table 3.

Table 3. Classification Report Before Optimization

Class	Precision	Recall	F1-Score
Negative	0.68	0.66	0.67
Neutral	0.56	0.82	0.67
Positive	1.00	0.09	0.17
Macro Avg	0.75	0.53	0.50

Based on Table 3, the positive sentiment category was the most difficult class for the model to recognize, with a recall value of only 0.09 and an F1-score of 0.17. This indicates that most positive reviews were incorrectly classified into other sentiment categories, particularly the neutral class. Although the precision value for the positive class reached 1.00, the very low recall indicates that the model detected only a small portion of the actual positive data. Meanwhile, the neutral class showed the highest recall value of 0.82, indicating that the initial model tended to predict data toward the neutral sentiment category.

After optimization through the application of N-Gram feature extraction and SMOTE data balancing, the model performance improved significantly. Accuracy increased from 61% to 77.51%, while recall improved from 0.53 to 0.78 and the F1-score increased from 0.50 to 0.77. These improvements indicate that the proposed optimization strategy successfully enhanced contextual feature representation and reduced classification bias toward the majority class. Furthermore, the more balanced precision, recall, and F1-score values demonstrate that the optimized model achieved more stable classification performance across all sentiment categories. The detailed classification results after optimization are presented in Table 4.

Table 4. Classification Report After Optimization

Class	Precision	Recall	F1-Score
Negative	0.75	0.84	0.79
Neutral	0.83	0.54	0.65
Positive	0.76	0.97	0.85
Macro Avg	0.78	0.78	0.77

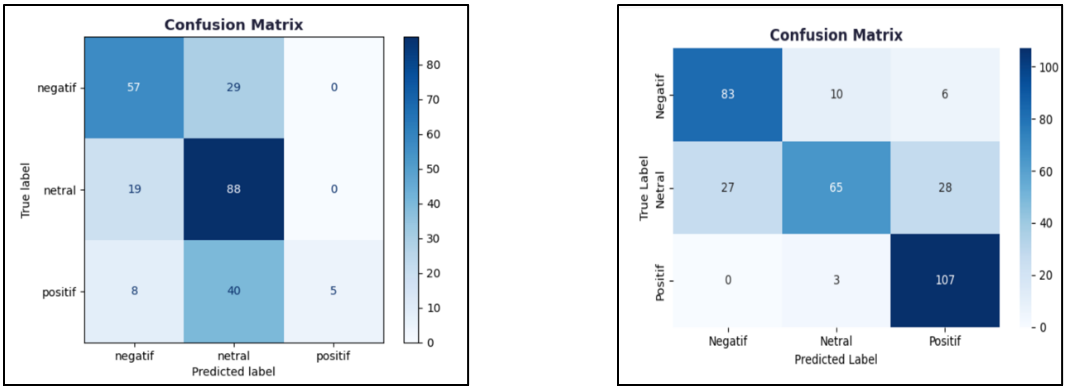
Based on Table 4, the optimized model achieved more balanced classification performance across all sentiment categories. The positive sentiment class experienced the most significant improvement,

Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

with recall increasing from 0.09 to 0.97 and F1-score increasing from 0.17 to 0.85. This result indicates that the combination of N-Gram and SMOTE successfully improved the model's ability to recognize minority sentiment classes more accurately. In addition, the negative sentiment category also achieved stable performance with an F1-score of 0.79, while the neutral class maintained relatively balanced precision and recall values.

Overall, the optimization process improved both model stability and generalization performance in sentiment classification tasks.

The classification performance before and after optimization is further visualized using confusion matrices, as shown in Figure 3.



(a) (b)

Figure 3. Confusion Matrix: (a) before optimization and (b) after optimization

Figure 3(a) shows that the initial model tended to predict data toward the neutral class, resulting in many misclassifications in the positive sentiment category. In contrast, Figure 3(b) demonstrates that after optimization, the prediction distribution became more balanced across all classes, as indicated by the increase in correctly classified samples (True Positives). This result confirms that the optimized model achieved better classification stability and generalization performance on the testing data.

The improvement achieved in this study supports previous findings that Naïve Bayes is effective for sentiment analysis tasks and can achieve competitive performance compared to other classification methods such as Support Vector Machine (SVM) [5]. The integration of N-Gram feature extraction and SMOTE balancing improved contextual text representation and reduced classification bias toward majority classes, resulting in more stable classification performance across positive, neutral, and negative sentiment categories.

In addition to the N-Gram-based sentiment classification approach, word visualization was also performed using WordCloud analysis, as shown in Figure 4. This visualization displays the most frequently occurring words in M-Passport user reviews, making it easier to identify dominant opinion patterns and overall user perception trends regarding the application.

Figure 4. WordCloud Analysis of M-Passport Reviews

The WordCloud visualization shows that the dominant words associated with the app's user experience, both positive and negative, are present. This provides a more intuitive understanding of

Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

users' focus and perceptions of the service, while also strengthening the sentiment analysis results obtained.

5. ANALYSIS SYSTEM

This section describes the results of the implementation of the Sentiment Analysis system developed as an application of the optimized Naïve Bayes model using N-Gram and SMOTE. This dashboard is designed using the Gradio framework, which allows users to input review text, predict the sentiment category (positive, neutral, or negative), and generate a WordCloud visualization to illustrate the distribution of dominant words from the input text.

Figure 5. Sentiment Analysis Dashboard for M-Passport Reviews

Below, we show the performance evaluation results of the Naïve Bayes model optimized using N-Gram and SMOTE. The evaluation was conducted to measure accuracy, precision, recall, and F1-score, as well as to observe the distribution of predictions using a confusion matrix.

Figure 6. Model Evaluation

The display above is the result of the Naïve Bayes model evaluation on the Sentiment Analysis Dashboard. The results show an increase in accuracy to 77.51%, with balanced precision, recall, and F1-score. The confusion matrix shows a more stable distribution of predictions among positive, neutral, and negative classes after optimization. This application can automatically analyze user sentiment towards the M-Passport application service, so it can help related agencies identify the level of satisfaction and service problems as a basis for improving the quality of digital public systems and services.

Vol. 7, No. 3, June 2026, Page. 2798-2811
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5853

6. CONCLUSION

Based on the analysis of user reviews from various digital platforms, perceptions toward the M-Passport application show considerable variation. Some reviews reflect positive responses regarding service accessibility and convenience, while others indicate dissatisfaction related to technical issues, limited features, and difficulties during the online service process. These findings indicate that the M-Passport application has significant potential to support the digitalization of public services, although several aspects still require improvement to enhance user experience and service quality. The application developed in this study successfully implemented a sentiment analysis system for M-Passport user reviews using the Naïve Bayes algorithm optimized with TF-IDF feature extraction, N-Gram representation, and SMOTE data balancing techniques. The evaluation results showed that the initial model achieved an accuracy of 61%, with a precision of 0.75, recall of 0.53, and F1-score of 0.50. After optimization, the model performance increased significantly, achieving an accuracy of 77.51% with more balanced precision, recall, and F1-score values across positive, neutral, and negative sentiment categories. These results demonstrate that the integration of contextual feature extraction and data balancing techniques can improve text representation quality, reduce classification bias, and increase sentiment classification stability on imbalanced datasets. The novelty of this study lies in the integration of N-Gram and SMOTE optimization techniques within a Naïve Bayes framework for sentiment analysis in the domain of digital immigration services, specifically the M-Passport application. In addition, this study contributes to the development of data-driven evaluation systems that can assist developers and related institutions in understanding user perceptions and improving the quality of digital public services.

For future work, further research can explore the implementation of deep learning and transformer-based methods, such as Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT), using larger and more diverse datasets to improve sentiment classification performance and contextual understanding in digital public service applications.

ACKNOWLEDGEMENT

The author would like to express his sincere gratitude to Universitas Putra Indonesia YPTK Padang for the academic and financial support provided during the research and study period. He also expresses his gratitude to all parties who contributed to the completion of this research, particularly in facilitating data collection and providing valuable insights.

REFERENCES

[1] A. Pujianto, A. Mulyati, And R. Novaria, “Pemanfaatan Big Data Dan Perlindungan Privasi Konsumen Di Era Ekonomi Digital Agung Pujianto 1) , Awin Mulyati 2) , Rachmawati Novaria 3),” *Majalah Ilmiah Bijak*, Vol. 127, No. 2, Pp. 127–137, 2018, [Online]. Available: Http://Ojs.Stiami.Ac.Id | https://doi.org/10.31334/bijak.v15i2.201

[2] S. Mufti Prasetiyo, R. Gustiawan, And F. Rizzel Albani, “Biikma : Buletin Ilmiah Ilmu Komputer Dan Multimedia Analisis Pertumbuhan Pengguna Internet Di Indonesia,” Vol. 2, No. 1, 2024, [Online]. Available: Https://Jurnalmahasiswa.Com/Index.Php/Biikma| | https://doi.org/10.31454/biikma.v15i2.321

[3] Deandlles Christover, Aji Syarif Hidayattulah, And Indah Mawarni, “Penerapan Konsep-Konsep Digitalisasi Dalam Pelayanan Publik Di Kantor Desa Manunggal Jaya Kecamatan Tenggarong Seberang Kabupaten Kutai Kartanegara,” *Journal Of Research And Development On Public Policy*, Vol. 2, No. 2, Pp. 199–214, 2023, https://doi.org/10.58684/jarvic.v2i2.73

[4] A. Zahra, R. Mayasari, And I. Pernamasari, "Analisis Sentimen Pada Aplikasi M-Paspor Menggunakan Algoritma Naïve Bayes Classifier," *Action Research Literate*, Vol. 8, No. 8, Pp. 2365–2371, 2024, https://doi.org/10.46799/arl.v8i8.466

[5] D. A. Putri, D. A. Kristiyanti, E. Indrayuni, A. Nurhadi, And D. R. Hadinata, "Comparison Of Naive Bayes Algorithm And Support Vector Machine Using Pso Feature Selection For Sentiment Analysis On E-Wallet Review," *J. Phys. Conf. Ser.*, Vol. 1641, No. 1, 2020, https://doi.org/10.1088/1742-6596/1641/1/012085

[6] M. H. S. Quadri* And Dr. R. K. Selvakumar, "Performance Of Naïve Bayes In Sentiment Analysis Of User Reviews Online," *International Journal Of Innovative Technology And Exploring Engineering*, Vol. 10, No. 2, Pp. 64–68, 2020, https://doi.org/10.35940/ijitee.a8198.1210220

[7] P. D. P. Silitonga, M. Hasibuan, Z. Situmorang, And D. Purba, "Comparison Of Tiktok User Sentiment Analysis Accuracy With Naïve Bayes And Support Vector Machine," *International Journal Of Advanced Trends In Computer Science And Engineering*, Vol. 12, No. 1, Pp. 11–15, 2023, https://doi.org/10.30534/ijatcse/2023/031212023

[8] M. Horvat, G. Gledec, And F. Leontić, "Hybrid Natural Language Processing Model For Sentiment Analysis During Natural Crisis," *Electronics (Switzerland)*, Vol. 13, No. 10, May 2024, https://doi.org/10.3390/electronics13101991

[9] R. A. Siregar, Y. A. Sari, And I. Indriati, "Analisis Sentimen Kebijakan New Normal Dengan Menggunakan Automated Lexicon Senti N-Gram," *Jurnal Teknologi Informasi Dan Ilmu Komputer*, Vol. 10, No. 1, Pp. 29–34, Feb. 2023, https://doi.org/.25126/jtiik.2023105006

[10] M. E. Purbaya, D. P. Rakhmadani, M. P. Arum, And L. Z. Nasifah, "Implementation Of N-Gram Methodology To Analyze Sentiment Reviews For Indonesian Chips Purchases In Shopee E-Marketplace," *Jurnal Resti*, Vol. 7, No. 3, Pp. 609–617, Jun. 2023, https://doi.org/10.29207/resti.v7i3.4726

[11] A. M. Priyatno And F. I. Firmananda, "N-Gram Feature For Comparison Of Machine Learning Methods On Sentiment In Financial News Headlines," *Riggs: Journal Of Artificial Intelligence And Digital Business*, Vol. 1, No. 1, Pp. 01–06, Jul. 2022, https://doi.org/10.31004/riggs.v1i1.4

[12] D. Normawati And S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-Sakti*, Vol. 5, No. 2, Pp. 697–711, 2021. http://dx.doi.org/10.30645/j-sakti.v5i2.369

[13] L. Hermawan, M. B. Ismiati, J. Bangau, N. 60, And M. Charitas, "Pembelajaran Text Preprocessing Berbasis Simulator Untuk Mata Kuliah Information Retrieval," *Transformatika*, Vol. 17, No. 2, Pp. 188–199, 2020. https://doi.org/10.26623/transformatika.v17i2.1705

[14] H. Ma'rifah, A. Prasetya Wibawa, And M. I. Akbar, "Sains, Aplikasi, Komputasi Dan Teknologi Informasi Klasifikasi Artikel Ilmiah Dengan Berbagai Skenario Preprocessing," Vol. 2, No. 2, P. 70, 2020. http://dx.doi.org/10.30872/jsakti.v2i2

[15] S. A. H. Bahtiar, C. K. Dewa, And A. Luthfi, "Comparison Of Naïve Bayes And Logistic Regression In Sentiment Analysis On Marketplace Reviews Using Rating-Based Labeling," *Journal Of Information Systems And Informatics*, Vol. 5, No. 3, Pp. 915–927, 2023, https://doi.org/10.51519/journalisi.v5i3.539

[16] Pristiyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, And F. H. Rambe, "Sentiment Analysis Of Covid-19 Vaccine In Indonesia Using Naïve Bayes Algorithm," *Iop Conf. Ser. Mater. Sci. Eng.*, Vol. 1088, No. 1, P. 012045, 2021, , https://doi.org/10.1088/1757-899x/1088/1/012045

[17] T. Chamidy, "Application Of Smote In Sentiment Analysis Of Myxl User Reviews On Google Play Store," 2025. https://doi.org/10.14421/jiska.2025.10.1.74-86
[18] L. D. Cahya, A. Luthfiarta, J. I. T. Krisna, S. Winarno, And A. Nugraha, "Improving Multi-Label Classification Performance On Imbalanced Datasets Through Smote Technique And Data Augmentation Using Indobert Model," *Jurnal Nasional Teknologi Dan Sistem Informasi*, Vol. 9, No. 3, Pp. 290–298, Jan. 2024, http://dx.doi.org/10.25077/Teknosi.V9i3.2023.290-298
[19] O. Rainio, J. Teuho, And R. Klén, "Evaluation Metrics And Statistical Tests For Machine Learning," *Sci. Rep.*, Vol. 14, No. 1, Dec. 2024, Doi: 10.1038/S41598-024-56706-X.
[20] A. R. Lubis, Y. Y. Lase, D. A. Rahman, And D. Witarsyah, "Improving Spell Checker Performance For Bahasa Indonesia Using Text Preprocessing Techniques With Deep Learning Models," *Ingenierie Des Systemes D'information*, Vol. 28, No. 5, Pp. 1335–1342, 2023, https://doi.org/10.18280/isi.280522
[21] M. U. Albab, Y. K. P., And M. N. Fawaiq, "Optimization Of The Stemming Technique On Text Preprocessing President 3 Periods Topic," *Jurnal Transformatika*, Vol. 20, No. 2, Pp. 1–12, Jan. 2023, https://doi.org/10.26623/Transformatika.V20i2.5374
[22] Rianto, A. B. Mutiara, E. P. Wibowo, And P. I. Santosa, "Improving The Accuracy Of Text Classification Using Stemming Method, A Case Of Non-Formal Indonesian Conversation," *J. Big Data*, Vol. 8, No. 1, Dec. 2021, https://doi.org/10.1186/S40537-021-00413-1
[23] S. Khairunnisa, A. Adiwijaya, And S. Al Faraby, "Pengaruh Text Preprocessing Terhadap Analisis Sentimen Komentar Masyarakat Pada Media Sosial Twitter (Studi Kasus Pandemi Covid-19)," *Jurnal Media Informatika Budidarma*, Vol. 5, No. 2, P. 406, Apr. 2021, https://doi.org/10.30865/Mib.V5i2.2835
[24] R. B. Hadiprakoso, H. Setiawan, R. N. Yasa, And Girinoto, "Text Preprocessing For Optimal Accuracy In Indonesian Sentiment Analysis Using A Deep Learning Model With Word Embedding," *Aip Conf. Proc.*, Vol. 2680, No. 1, P. 020050, Dec. 2023, https://doi.org/10.1063/5.0126116
[25] A. F. Aufar, Mochamad Alfan Rosid, A. Eviyanti, And I. R. I. Astutik, "Optimizing Text Preprocessing For Accurate Sentiment Analysis On E-Wallet Reviews," *Jicte (Journal Of Information And Computer Technology Education)*, Vol. 7, No. 2, Pp. 42–50, Oct. 2023, https://doi.org/10.21070/Jicte.V7i2.1650
[26] D. Wulan Yekti Rahayu *Et Al.*, "Performance Of Machine Learning Algorithms On Imbalanced Sentiment Datasets Without Balancing Techniques," 2025. [Online]. Http://Jurnal.Polibatam.Ac.Id/Index.Php/Jaic | https://doi.org/10.30871/jaic.v9i3.9584
[27] R. M. Prabha And S. Sasikala, "Data Analytics For Imbalanced Dataset," *Journal Of Computer Science*, Vol. 20, No. 2, Pp. 207–217, 2024, Doi: 10.3844/Jcssp.2024.207.217.
[28] M. R. Ramadhan And K. Budiman, "Sentiment Analysis Of Presidential Candidates In 2024: A Comparison Of The Performance Of Support Vector Machine And Random Forest With N-Gram Method," *Recursive Journal Of Informatics*, Vol. 3, No. 1, Pp. 34–42, Mar. 2025, https://doi.org/10.15294/Rji.V3i1.8385
[29] M. B. Alfarazi, M. 'Ariful Furqon, And H. Soepandi, "Sentiment Analysis Of Universitas Jember's Sister For Student Application Using Gaussian Naive Bayes And N-Gram," *Journal Of Innovation Information Technology And Application (Jinita)*, Vol. 6, No. 2, Pp. 101–108, Dec. 2024, https://doi.org/10.35970/Jinita.V6i2.2400
[30] H. Noer Rofiq And G. Mafela Mutiara Sujak, "Analisis Persepsi Masyarakat Terhadap Lelang Indonesia Melalui Analisis N-Gram Dan Sentimen," Vol. 7, 2024, [Online]. Available: https://doi.org/10.31598