Journal of Computer Science

Research Article

# Toward Intelligent Evaluation of Digital Public Services: Evidence from Indonesia's SIM Online Platform

**Evaristus Didik Madyatmadja, Ricky Kosasih, Najla Aurelia Evanthe, Rudy and Betley Heru Susanto**

*Information Systems Department, School of Information Systems, Bina Nusantara University, Indonesia*

*Article history*
Received: 21-08-2025
Revised: 07-04-2026
Accepted: 23-04-2026

**Corresponding Author:**
Evaristus Didik Madyatmadja
Information Systems Department, School of Information Systems, Bina Nusantara University,Indonesia
Email: emadyatmadja@binus.edu

**Abstract:** This investigation examines more than 65,000 user evaluations from the Google Play Store concerning sentiment toward the Digital Korlantas POLRI application. Reviews were systematically processed and categorized through Naïve Bayes, Support Vector Machine, and Random Forest classifiers, thereby highlighting the perceived strengths and weaknesses of the platform. Among the tested models, Naïve Bayes outperformed the others, yielding the greatest accuracy, precision, and recall. The analysis therefore provides quantifiable evidence of the application's functional efficacy, the nuances of user experience, and the quality-of-service delivery. Such findings constitute actionable feedback for iterative enhancements aimed at optimizing application performance and augmenting user satisfaction. Future investigations may expand this foundational work by broadening the dataset to encompass diverse feedback channels and employing aspect-based sentiment analysis to isolate and scrutinize particular features and localized user concerns.

**Keywords:** Digital Traffic Corps of the Indonesian National Police, Sentiment Analysis, Naïve Bayes Algorithm, Digital Korlantas Police Application, User Feedback, Machine Learning, Support Vector Machine Algorithm, Random Forest Algorithm, Electronic Government, User Experience, Public Service Optimization

## Introduction

Gusman (2024) explains that the incorporation of Information and Communication Technology (ICT) into public administration has significantly transformed service delivery systems. Indonesia's e-government initiatives aim to improve transparency, efficiency, and citizen participation. Digital Korlantas POLRI application (also referred to as SINAR) is one of the applications developed by Indonesian National Police (POLRI) which enables the issuance and renewal of drivers' licenses (SIM), registration of vehicles, and provision of information on traffic violations (Fig. 1). Andriansyah, et al. (2024) note that digitization removes the need for citizens to physically visit police offices for services.

Jusal (2024) emphasizes that user satisfaction and user experience are critical indicators of the effectiveness of digital public service platforms. Puspasari et al. (2024) state that online reviews and user feedback from the Google Play Store provide valuable insights into public perceptions of digital applications. The low rating of 3.7 stars out of 5 in April 2025, along with numerous user complaints indicates disappointing services. (Kurniawati et al., 2019) found that user ratings and feedback often reveal service dissatisfaction and highlight areas requiring improvement.

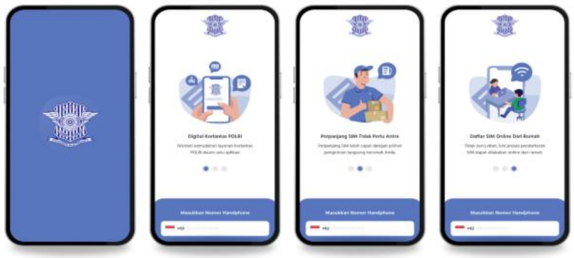

**Fig. 1:** Digital Korlantas POLRI Application

The increasing user base of the Digital Korlantas POLRI has led to evaluations that enhance its resource capacity, leading to better the quality of digital public services for law enforcement. User reviews, particularly research analyzing reviews, clearly discusses certain aspects of the overall status of user experience, functionality-related difficulties, and general user satisfaction. Purba and Yadi (2023); Utami and Erfina (2021) indicate that application store reviews are a valuable source of user-generated feedback for evaluating digital services. Thus, the goal of this research is to identify focus areas that need intervention through the application of critical sentiment analysis methods along with the evaluation of trends in user sentiment across these areas. The research will take into account the evassluation of the feedback coming from users concerning the Digital Korlantas POLRI application which has been acquired across the one-year duration.

Sentiment analysis is used as a text-mining tool measures user perspectives based on both positive and negative reviews about Digital Korlantas POLRI application (Avlach and Yamasari, 2025; Nurcahyawati and Mustaffa, 2023). The public reviews are fundamental in determining application development strategies and digital government service innovations (Thakkar and Varma, 2020). The research's main objective is to collect and analyze user appraisals on the Digital Korlantas POLRI application from the Google Play Store. (Wankhade and Patil, 2020) explain that web scraping combined with sentiment analysis enables systematic evaluation of user feedback. This helps enhance understanding of users' expectations. Their pain points and the satisfaction factor lead to purposeful evidence application into practical application to improve strategy.

Sentiment analysis, in NLP parlance, is the computational study of the emotions, sentiments, and opinions expressed through textual data. Sentiment analysis has been used in the area of public service delivery in understanding user reviews for a range of e-government applications, forming data-informed insights to improve services (Rizkiansyah et al., 2022). Sentiment analysis has been used in studies to evaluate the user's view on some regional e-government applications in Indonesia, such as Sapawarga and Jaki. This could show the important factors that shape user satisfaction such as application performance, usability, and data security (Nurmalasari et al., 2023).

The Digital Korlantas POLRI application can be utilized for the analysis of sentiments as to the experiences and attitudes of users with the app in order to make recommendations for its innovations in service delivery thereby enhancing the image of POLRI to the public (Handika et al., 2024). Naïve Bayes is a classification method utilizing Bayes' theorem and is statistical by using probability. Random Forest, which employs internal division on the predictor variable to predict the outcome and is simple in its model, and Support Vector Machine, which is based on statistical learning theory, are some algorithms that can be analyzed for user sentiment and find inside the most common themes from user-generated evaluations on the Digital Korlantas POLRI platform (Chen et al., 2021; Choudhury, 2012; O'Connell et al., 2025).

In this study, Naïve Bayes, Random Forest, and Support Vector Machine algorithms are applied to the collection of reviews written by Indonesian users over the span of one year. These models will be further tested to determine their classification effectiveness in the attempt to extract meaningful and reliable methodologies for analyzing Indonesian user reviews. The insights gleaned could be instrumental in gathering information that could be utilized for the proactive enhancement of user satisfaction, the robustness of the application, and the overall user experience.

### *Data Mining*

Data mining is an application of machine learning that employs machine learning processes to determine important trends or patterns in large databases. Essentially, the process generates "knowledge" by analyzing the data in a systematic manner from amongst a large volume of data. Through this, it could search for the underlying patterns found on the applications advantages and disadvantages using such techniques to analyze the users' comments regarding the Digital Korlantas POLRI application. Murphy (2012) describes data mining as a process of extracting meaningful knowledge from large datasets.

### *Sentiment Analysis*

Sentiment analysis refers to a collection of natural language processing techniques that interpret user emotions based on textual feedback to provide a data-centric view on user satisfaction and experience. It analyzes and identifies linguistic and psychological understandings from comments and app reviews, dividing them based on sentiments to discover the features in the application that strike a chord with users or aggravate them. Rahman et al. (2020) suggest that sentiment analysis helps developers understand user emotions and improve application design.

### *Aspect-Based Sentiment Analysis (ABSA)*

ABSA (Aspect-Based Sentiment Analysis) is an advanced form of sentiment analysis focused on the particular aspects of a product and the feelings corresponding to them by the users (Witten et al., 2011). It thus allows intricate analysis of user feedback about different aspects of the application, such as interface design and system performance, when applied in the Digital Korlantas POLRI application.

### *Classification*

Classification includes a set of systematic processes to assign data elements to pre-existing categorical groups in machine learning. Mikolov et al. (2013) explain that classification techniques assign data elements into predefined categories. This classification technique helps in understanding user sentiments by putting unstructured textual feedback into structured, analysable data capable of spotting patterns in user experiences and their satisfaction levels.

### *Confusion Matrix*

Manning et al. (2008) described the confusion matrix as a common technique for evaluating classification accuracy.

### *Related Work*

Recent studies into the sentiment analysis around the Digital Korlantas POLRI program demonstrate the efficacy of the machine learning models in user-review scenarios. It has been observed from the work assessing the public sentiment about the performance of POLRI that the implementation of Naïve Bayes serves quite well when large-scale sentiment data are involved (Handika et al., 2024). A number of comparative studies carried out on different machine learning algorithms emphasized the need for carefully selecting an appropriate one to increase the accuracy. It has been established that Naïve Bayes is an excellent choice for sentiment analysis of application reviews, given the high level of accuracy with which it can determine the sentiment of users (Puspasari et al., 2024). Furthermore, Random Forest serves to analyze very complex sentiment data for public services like the Digital Korlantas POLRI application because it is recognized as a good algorithm for handling complex data sets while some-how maintaining the stability of the models (Handika et al., 2024). Continuous development of these models gives insight into the further enhancement of digital public service platforms like POLRI's (Puspasari et al., 2024).

## Methods

Analysis of sentiment starts from implementing the Google Play Scraper library using Python in Jupyter Notebook to execute web scraping. The data was retrieved, went through the necessary steps including pre-processing, feature extraction, and dividing the data into training and testing data subsets. Afterward, classification of sentiments was performed for analysis and results were compared and evaluated. Ultimately, the information gained throughout the process is visualized in order to gain greater understanding.

### *Data Collection*

The entirety of the data collection process is outlined in Figure 2. For this particular case, the data under consideration was gathered from users' ratings and reviews of the Digital Korlantas POLRI application on Google Play Store. This process involves web scraping and data exporting. With regard to this web scraping project, we employed the "Google Play Scraper" package in Python. The complete dataset comprises approximately 65,629 reviews.

Following this step, the results are stored in an XLSX file. The collected dataset comprises several columns containing attributes and functions which are shown in Table 1.

The dataset supporting this research is publicly available through Zenodo: https://doi.org/10.5281/zenodo.16877219

### *Text Preprocessing*

Text preprocessing refers to the methods of the procedure defined as transforming unprocessed text into an organised, clean format that can be analysed. Kannan and Gurusamy (2015) explain that text preprocessing converts raw textual data into a structured format suitable for analysis.

### *Data Labeling*

Data labeling is the process of attaching a particular category or sentiment to a given piece of text, marking the beginning of supervised learning. Proper labeling guarantees that models learn the linkage between the features and output classes to be accurate (Sambasivan et al., 2021). In this study, reviews are labeled into categories based on sentiment to enable the application of classification algorithms.

### *Case Folding*

Case folding is a type of text normalization which makes all letters of a given text into the same case, usually lower case, for uniformity across the text. This step reduces the effect of capitalization differences, treating words like 'Good', 'GOOD', and 'good' as the same for analysis purposes. Case folding makes the text uniform and enhances the consistency of results in later steps like tokenization, filtering, and classification. Literature shows that incorporating case folding in the preprocessing pipeline improves the performance of text classification tasks by lower the perplexity and increasing accuracy on models (Uysal and Gunal, 2014).

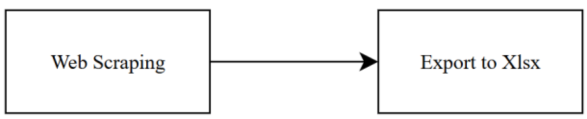


**Fig. 2:** Data collection method

**Table 1:** Data Attribute and Data Type

| Attribute | Description | Type |
|---|---|---|
| userName | User's name used in the Google Play Store | String |
| score | User's rating | Integer |
| at | Date when the review was posted | Datetime |
| content | User's review | String |

### *Data Filtering*

Data filtering enhances the accuracy and effectiveness of any subsequent analyses or modelling procedures through the selection of relevant data from a larger dataset by removing redundancy, irrelevant data, or noisy information (Radha et al., 2016).

### *Tokenizing Data*

Tokenizing refers to the splitting of words into smaller parts known as tokens. Camacho-Collados and Pilehvar (2018) state that tokenization facilitates the discovery of certain linguistic patterns within the corpus. For data processing, tokenization is vital as it enables the computer system to process and comprehend the data by treating each token as an individual entity.

### *Stop Word Removal*

Stop words include words such as "the," "is," "and," "in," and "of." Within the domain of Natural Language Processing (NLP), such words are often omitted as they add unwanted noise to the data and increase its dimensionality. Eliminating stop words yields a well-structured dataset that is less complex and more streamlined for analysis.

It has been noted in previous works that the elimination of stop words is a fundamental step in tasks such as information retrieval, indexing, and even in more advanced operations like topic modeling and text classification (Sarica and Luo, 2021). Removing these words improves the text's signal-to-noise ratio, thus increasing its statistical relevance in relation to the words that are essential for the operation at hand. For this research, a list of custom dataset-specific stop words was created by merging the provided Indonesian stop words list from the NLTK database. This approach helps ensure the generic as well as context-specific stop words are removed which improves the quality of textual data for deeper analysis.

### *Data Stemming*

According to Jivani (2011), stemming is an integral part of an NLP project and is defined as the process that involves reduction of inflected or derived words to their base or root form. Text normalization is the process of transforming a vast number of diverse word forms into a small number of standardized and uniform representations. Stemming helps to improve the efficiency of text normalization and is also a useful step for a number of other text-related tasks including text classification, information retrieval, and even sentiment analysis.

In this research, the Sastrawi stemming library for the Indonesian language has been used to apply the stemming technique to the Indonesian text data set. The stemming process reduces the redundancy of various word forms (by removing affixes) and improves the uniformity of the words. This leads to an increase in the effectiveness and accuracy of the text analyses that will be performed subsequently.

### *Data Splitting*

Data splitting is an essential step in a machine learning workflow and occurs during the preprocessing phase in which a dataset is divided into multiple subsets. At the very least, one of the subsets is used for training and the other is used for testing. The evaluation method utilizes the principle of determining how accurate a model is that has been trained on one subset of data and produces predictions on a different subset of data (which was not used for training) and, therefore, enables the evaluation of the model's ability to generalize. The importance of data splitting for the evaluation of machine learning models is discussed by Joseph and Vakayil (2022).

In this research study, the pre-processed dataset was split into training and testing subsets following an 80:20 split. An 80:20 data split is commonly used for this purpose because it aids in achieving a model-learning/ evaluation data sample balance (Alakwaa et al., 2017; Joseph and Vakayil, 2022; Gholamy et al., 2018) indicate that a data split allocation of around 70–80% is recommended for training the model and obtaining an evaluable model performance. Using this split, 9,021 samples were used for training and 2,256 for testing, allowing the model to be trained on a sufficiently large data sample, while an unbiased data sample was retained for evaluating the model performance.

### *Classification of Sentiment Analysis*

After completing the processes concerning data preprocessing, the next step of the investigation will be sentiment analysis implementation. In this step, the training of models which employs specific data mining techniques will be performed. The study performs comparative analysis using three algorithms for classification, namely: Naive Bayes, Support Vector Machine, and Random Forest.

### *Naïve Bayes*

Naive Bayes is a probabilistic classifier which considers that the features are strongly independent, calculated through Bayes' theorem. It is widely employed in sentiment analysis because of its effectiveness and ease, especially with large datasets. With all its shortcomings, it still greatly excels in classifying text (Sumantiawan, et al., 2023).

### *Support Vector Machine (SVM)*

Support Vector Machine (SVM) is a supervised machine learning algorithm used at large for classification activities like sentiment analysis. SVM does this by determining the hyperplane that has the maximum distance from the nearest data points of any class, also known as support vectors. In sentiment analysis, SVM is proven to work well in multidimensional feature space due to their considering each word as a feature. Since SVM is robust to overfitting and accommodates numerous sophisticated decision boundaries, it is well-suited to differentiating positive and negative sentiments in the text data (Ahmad et al., 2018).

### *Random Forest*

Random Forest is an ensemble learning method that builds large numbers of decision trees and averages their predictions to increase accuracy and stability. It is particularly helpful in sentiment analysis due to the large datasets where they tend to overfit (Pervan and Keleş, 2017).

### *Sentiment Analysis Algorithm*

The next step after the preprocessing phases is the execution of the implementation of the processes of sentiment classification. An 11,277-record dataset is split, using a division of 80 to 20 for training versus testing. This study applies 3 classification algorithms, which include Naive Bayes, Support Vector Machine (SVM), and Random Forest, using the scikit-learn library. These algorithms allow data to be classified in a systematic way, thus enabling data to be classified for in-depth analytics.

### *Evaluation and Comparison*

The Confusion Matrix for comparison and evaluation, which is used in this research, allows for evaluation across all evaluation and classification methods. Each step of the process receives a score, as all classification methods are compared on how well each instance is classified. This leads to one of the four outcomes: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Once classification outcomes are obtained, the effectiveness of each classification method is assessed based on the measures of Precision, Recall, F1 Score, and accuracy.

## Results and Discussion

This study utilized a dataset of 65,629 user reviews collected from the Google Play Store within the timeframe of *22*□□ April 2024 to *22*□□ April 2025. During the data cleaning phase, a substantial portion of the raw data was excluded to ensure analytical reliability and data quality. Entries were filtered out if they had duplicates, missing or blank text fields, non-Indonesian language content, irrelevant characters, spam-like behavior, too short, semantically empty comments, or reviews with only emojis or punctuation. Using these validation and preprocessing parameters, 11,277 samples remained and were kept for analysis. The dataset was preprocessed and split into training and testing subsets in an 80:20 ratio, resulting in 9,021 samples for training and 2,256 samples for testing. In addition, the distribution analysis showed that among the reviews in the processed dataset, negative reviews were more than those that were positive.

When it comes to model construction, the same training set is employed to implement each method, and a different testing set is used to determine the validity of the model. Each of the processes, including the classification, was done in the Python programming environment and using the Scikit-Learn library. Below are the results for model classification.

### *Classification Model*

This study uses three classification models: Naive Bayes, Support Vector Machine (SVM), and Random Forest. The comparison aims to identify which model performs best in classifying user sentiment.

### *Naïve Bayes*

Along with customized module implementations for Scikit-Learn and NLTK, the Naive Bayes classification model was employed for keyword-based classification of data. Evaluation of the Naive Bayes method (see Figure 3) was done using the confusion matrix which indicated 1114 True Positives (TP), 132 False Positives (FP), 948 True Negatives (TN), and 62 False Negatives (FN).

The Naïve Bayes model has an accuracy of 0.91 on the testing set, which Table 2 shows as one of the better classification performances. This model has a precision of 0.89 and a recall of 0.95 for negative reviews, which shows it was good at pinpointing negative sentiment and had very few false negatives. This is a relevant class for public service reviews, as problematic negative feedback will have a complaint or a signal of a broken system that requires attention.

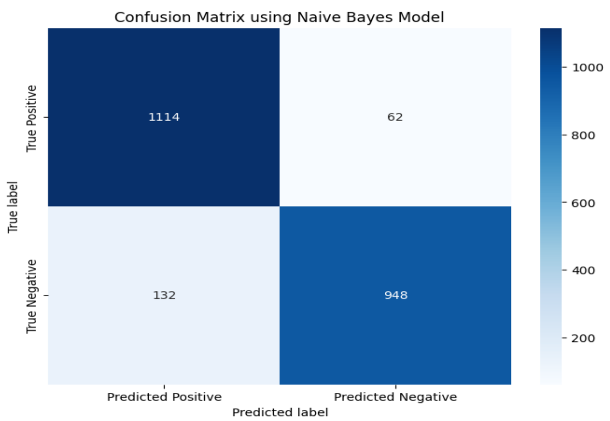


**Fig. 3:** Confusion Matrix of *Naïve* Bayes Model

**Table 2:** Classification Report of Naïve Bayes model

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Negative | 0.89 | 0.95 | 0.92 | 1176 |
| Positive | 0.94 | 0.88 | 0.91 | 1080 |
| Accuracy | - | - | 0.91 | 2256 |
| Macro avg | 0.92 | 0.91 | 0.91 | 2256 |
| Weighted avg | 0.92 | 0.91 | 0.91 | 2256 |

**DOI:** 10.3844/jcssp.2026.1811.1822

For positive reviews, she had a better precision of 0.94, but a worse recall of 0.88, meaning that when the model predicted a review as positive, she was almost correct, but a handful of positive reviews remained misclassified. The negative and positive sentiment F1-scores of 0.92 and 0.91 tell us that the model was able to sustain a similar performance across both classes. All of the above demonstrates that Naïve Bayes, despite criticism for a lack of sensitivity for negative user feedback, can actually be invaluable for pinpointing service problems for the dataset presented and the public digital services domain due to the consistency of performance it has in the majority of the classes it predicts

### *Support Vector Machine (SVM)*

Support Vector Classifier (SVC) makes use of the linear kernel from the Python Scikit-Learn module. To assess the performance of the Support Vector Machine (SVM) method a confusion matrix is used which is shown in Figure 4. In the data there were 1122 True Positives (TP), 141 False Positives (FP), 939 True Negatives (TN), and 54 False Negatives (FN).

Table 3 shows that the Support Vector Machine Model has an overall accuracy of 0.91 which is on par with the Naïve Bayes model. In terms of negative reviews, the model scored a precision of 0.89 and a recall of 0.95. This is the same as Naïve Bayes and shows that the model is good at identifying the negative sentiments. On the other hand, SVM scored a positive review precision of 0.95 which is a bit better than Naïve Bayes, but the recall was 0.87 which is lower. This indicates that SVM was good at identifying positive reviews, but Naïve Bayes were able to find more positives that SVM missed. The F1 Scores were good as well with SVM obtaining 0.92 for negative and 0.91 for positive sentiments, thus proving the model to be well performing overall. The slightly lower recall for positive reviews shows that SVM may be more careful when labeling the positive class. Overall, this makes SVM a good choice to be an alternative model. When compared with Naïve Bayes which is much less complex and more efficient, SVM does have less to offer.

### *Random Forest*

Random Forest is also evaluated using classification metrics such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). In this study, the model achieved a TP of 1099, FP of 171, TN of 909, and FN of 77, which is shown in Figure 5.

Random Forest outperformed all three models in Table 4 by achieving a 0.89 accuracy score. Although this achievement did not match the accuracies of Naive Bayes nor the SVM models, the score is still a testimony to the model's potential. For negative sentiments, the model achieved a precision of 0.87, and a recall of 0.93 meaning the model could still capture most negative reviews even when it was less precise than the two other models. For the positive sentiment class, the precision was 0.92, however the recall at 0.84 was the lowest score, the lowest score when compared to the other two models as well. this indicates a relative struggle by the model to identify all positive reviews resulting in a large number of false negatives in the positive class. The negative class and positive class F1-scores of 0.90 and 0.88 respectively restate the model's performance as weaker and less balanced in comparison to SVM and Naive Bayes. These issues are symptoms of tree-based ensemble model methods. Although the Random Forest model achieved results that can be textual data, it is clear, the model fell short on the task of classifying sentiments.

**Table 3:** Classification Report of Support Vector Machine model

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Negative | 0.89 | 0.95 | 0.92 | 1176 |
| Positive | 0.95 | 0.87 | 0.91 | 1080 |
| Accuracy | - | - | 0.91 | 2256 |
| Macro avg | 0.92 | 0.91 | 0.91 | 2256 |
| Weighted avg | 0.92 | 0.91 | 0.91 | 2256 |

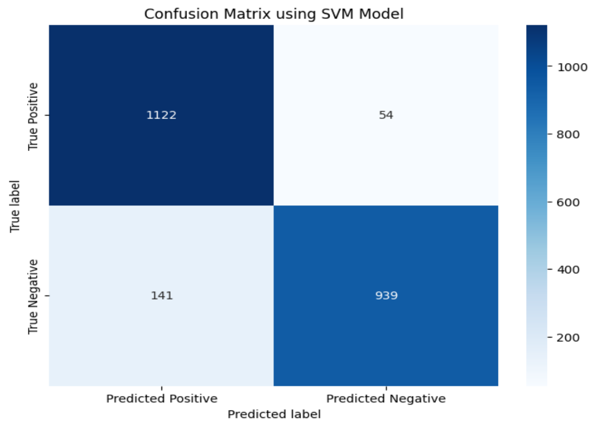


**Fig. 4:** Confusion Matrix of Support Vector Machine Model

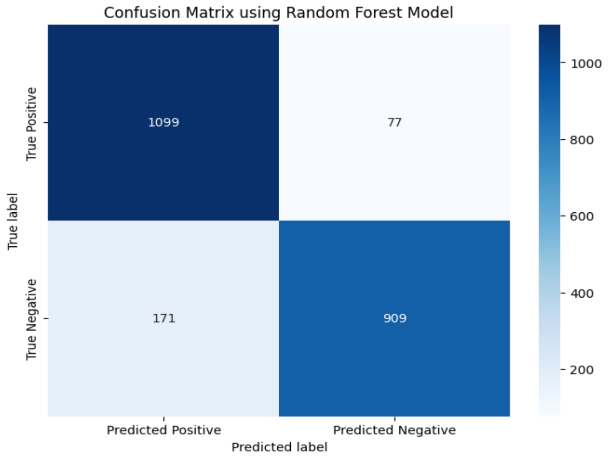


**Fig. 5:** Confusion Matrix of Random Forest Model

**DOI:** 10.3844/jcssp.2026.1811.1822

**Table 4:** Classification Report of Random Forest model

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Negative | 0.87 | 0.93 | 0.90 | 1176 |
| Positive | 0.92 | 0.84 | 0.88 | 1080 |
| Accuracy | - | - | 0.89 | 2256 |
| Macro avg | 0.89 | 0.89 | 0.89 | 2256 |
| Weighted avg | 0.89 | 0.89 | 0.89 | 2256 |

*SSentiment Analysis Classification Comparison*

The performance evaluation for three models of classification such as Naïve Bayes, Support Vector Machine (SVM), and Random Forest was carried out through four standard criteria: Precision, Recall, F1-Score, and Accuracy. These metrics outline the capability of each model in classifying the polarity of sentiment for the provided reviews regarding the Digital Korlantas POLRI app.

The highest-performing model is Naive Bayes with Precision 0.915, Recall 0.914, F1-Score 0.914, and Accuracy 0.914. It is also the best model by standard deviation and therefore in addition to the best performance, Naive Bayes has the lowest variability. It indicates that Naive Bayes is very well suited to the text classification task and identifies the sensitive distributions of true and false positives and true and false negatives in classification of the reviews.

Support Vector Machine's Precision 0.916, Recall 0.914, F1-Score 0.913, and Accuracy 0.914 is nearly on par with that of Naive Bayes. While there is a difference in the metrics, it would suggest that SVM is to a greater extent dependent on the particular feature representation or the distribution of the data in this work. However, SVM remains a very competent model to have such a small difference in performance.

Random Forest had the worst performance of all classifiers with the lowest Precision of 0.892, Recall of 0.890, F1-Score of 0.890, and Accuracy of 0.890. Although its performance is still within and acceptable range, it is significantly lower than that of Naïve Bayes and SVM. This indicates that, at least the this instance, Random Forest does not appear to provide a suitable option for text sentiment analysis.

Based on the above results, all three models provide a good performance for sentiment analysis. However, the models that perform the best are Naïve Bayes and Support Vector Machine (SVM), with some 0.91 scores for all of the evaluation criteria. The consistency and performance of Naïve Bayes are very strong, and this is a result of the characteristics of Naive Bayes being very suitable for large and sparse datasets like text analysis. The results for SVM were also very good and demonstrated its strengths for various representations. On the other hand Random Forest had a lower performance than Naive Bayes and SVM, and has therefore led to the conclusion that to have good and reliable sentiments analyses for the hands-on Digital Korlantas POLRI app, Naive Bayes and SVM are the best options. The full results are presented in Table 5.

**Table 5:** Comparison of classification model's result

| Model | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|
| Naïve Bayes | 0.915 | 0.914 | 0.914 | 0.914 |
| Support Vector Machine | 0.916 | 0.914 | 0.913 | 0.914 |
| Random forest | 0.892 | 0.890 | 0.890 | 0.890 |

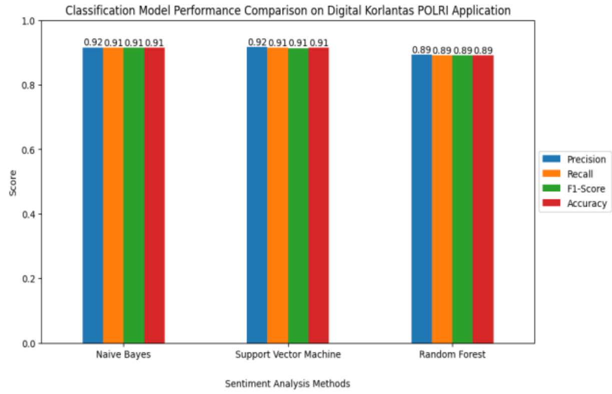


**Fig. 6:** Classification Model Performance Comparison on Digital Korlantas POLRI

Figure 6 represents the performance metrics (precision, recall, F1-score, and accuracy) for all classification algorithms. From the figure, we can see the best performance of the Naïve Bayes classifier and the Support Vector Machine (SVM) classifying algorithms, achieving about 0.92 in precision and 0.91 in recall, F1 and accuracy. This means that both Naïve Bayes and SVM classifiers are better in detecting the intents of users. On the other hand, the Random Forest classifier does a little worse and gets 0.89 in all metrics. This visual evaluation, consistent with the results obtained, confirms that Naïve Bayes and SVM are the best models for sentiment analysis of the Digital Korlantas POLRI application in this study. Such an observation enhances interpretation and corroboration of the conclusions derived from feature analysis.

Figure 6 shows the accuracy, precision, recall and F1 score for the three classification models. With an accuracy of 0.91, Naive Bayes and Support Vector Machine (SVM) have similar performance regarding the three evaluation metrics. This is due to the common traits seen in textual datasets. High dimensional and sparse textual datasets are derived from word representation, and Naive Bayes is predicated to perform well in these situations as it accurately classifies and predicts class probabilities, even when some features are used less frequently. Support Vector Machine (SVM) is also better for environments containing copious amounts of high dimensional text, as it performs better at pinpointing the optimal positive or negative axis or hyperplane that separates the sentiment classes than Naive Bayes. With a slight score of 0.90, Random Forest

(RF) performed the worst, as tree-based ensemble algorithms do poorly in environments that contain a sparse bag of an imprecise collection of words, especially when the relationship between the features are weak and very dispersed. Consequently, the decision boundary of Naive Bayes and the decision boundary of SVM are more stable compared to Random Forest for this classification of sentiment.

### *Sentiment Analysis Results*

This study identifies the dominant sentiment and salient themes among users who left positive feedback after performing sentiment analysis on the cleaned dataset of user reviews for the Digital Korlantas POLRI application. During the preprocessing phase, some words were discarded, which enabled a clearer interpretation of sentiment regarding user feedback. The analysis results were also presented as bar charts generated in Matplotlib in order to illustrate the trends pertaining to the top ten terms most commonly found in positive reviews.

This research has achieved the sentiment analysis goals on the cleaned user reviews dataset on the Digital Korlantas POLRI application by determining the particular opinion and topics expressed by users in positive feedback comments. The preprocessing stage guaranteed the relevant words were the only words included, meaning user sentiment was better interpreted. The outcomes of this analysis were visualized with a bar chart of the top ten most frequent words in positive reviews, using Matplotlib for drawing.

As shown in Figure 7, the most frequently mentioned terms from positive reviews (total 5,178 submissions) are: “sim” (driver’s license, 1,665 mentions), “mudah” (easy, 1,134 mentions), “panjang” (to renew, 993 mentions), “cepat” (fast, 843 mentions), “proses” (process, 672 mentions), “mantap” (great/cool, 600 mentions), “aplikasi” (application/app, 589 mentions), “layan” (service, 577 mentions), “online” (online, 426 mentions), and “terima” (accept, 392 mentions). These keywords capture the common feedback themes from users, indicating the application’s efficiency, user-friendliness, and functionality, particularly in delivering online services.

The use of words such as "cepat" (fast) and "mudah" (easy) implies that a majority of the users had a smooth and time-efficient experience. The prominence of terms such as “sim” (driver’s license) and “panjang” (to renew) shows that services related to SIM are among the most utilized and rated positively. Furthermore, phrases such as “mantap” (great/cool) and “terima” (accept) demonstrate user satisfaction and a positive emotional response towards the application. Table 6 provides a summary of the exact word frequencies.

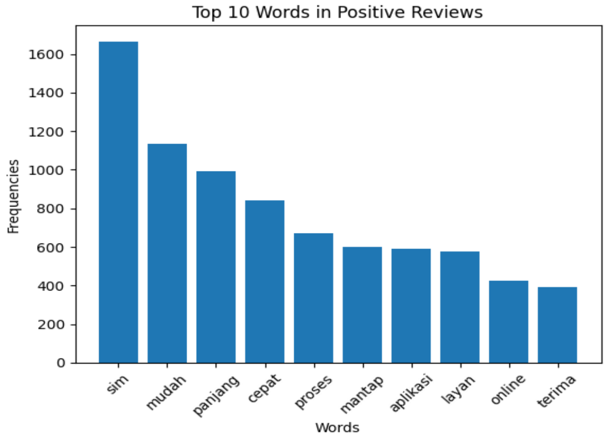


**Fig. 7:** Top 10 Words in Positive Reviews for the Digital Korlantas POLRI Application

**Table 6:** Frequency of the Top 10 Words in Positive Reviews

| No | Word | English Meaning | Frequency |
|---|---|---|---|
| 1 | sim | Driver’s license | 1,665 |
| 2 | mudah | Easy | 1,134 |
| 3 | panjang | To Renew | 993 |
| 4 | cepat | Fast | 843 |
| 5 | proses | Process | 672 |
| 6 | mantap | Great/Cool | 600 |
| 7 | aplikasi | Application | 589 |
| 8 | layan | Service | 577 |
| 9 | online | Online | 426 |
| 10 | terima | Accept | 392 |
| | Total | | 5,178 |

Alongside analyzing the positive sentiment, this study explores the negative aspects as well to understand the dissatisfaction regarding the Digital Korlantas POLRI application. Through the use of preprocessing techniques to filter out irrelevant words, the most ten frequently mentioned words in the negative reviews were found and represented in a bar graph, which is displayed in Figure 8. These insights were drawn from 5693 entries of negative reviews.

The top ten most frequently used words in negative reviews are: “aplikasi” (application, 2,541 mentions), “sim” (driver’s license, 1,947), “gak” (informal no/not, 986), “nya” (possessive suffix, 914), “panjang” (to renew, 901), “ga” (informal no/not, 839), “verifikasi” (verification, 835), “gagal” (failed, 719), “proses” (process, 607), and “online” (online, 576). Collectively, these ten words were recorded 5,639 times, indicating that many reviews contained multiple issues or complaints within a single remark.

These keywords uncover a number of specific issues that users suffer from. The high occurrence of “aplikasi” and “sim” indicate that users’ discontent is mainly towards the functionality of the application and its SIM-related features.

**DOI:** 10.3844/jcssp.2026.1811.1822

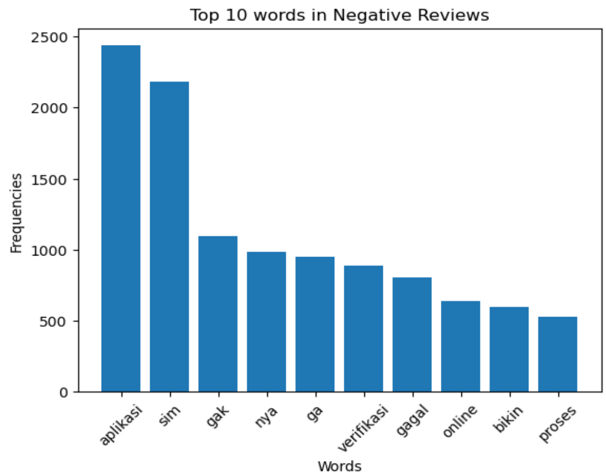


**Fig. 8:** Top 10 Words in Negative Reviews for the Digital Korlantas POLRI

Words “gak”, “ga”, and “gagal” indicate the users’ frustrations regarding expectations not being met due to some system failing to work as intended or due to some bugs. The term “verifikasi” shows users’ annoyance concerning the identity or document verification, which is a very fundamental step toward interacting with services in the digital world. Lastly, the words “panjang”, “proses”, and “online” highlight users’ concerns regarding the SIM renewal process conducted through the application. The term “panjang” may refer both to the act of renewing (perpanjang) a SIM and to the perception that the process takes too long. Combined with “proses” and “online”, these terms suggest that the procedures, especially those carried out online, are often viewed as complicated, time-consuming, or inefficient, potentially hindering users from successfully completing important actions. The detailed frequency of each top word is presented in Table 7.

In addition to frequency analysis, Figures 9 and 10 show their respective positive and negative review word clouds. These figures normalize their most frequent terms to convey to readers that the size of the clouds corresponds to the frequency of words contained in the user feedback, as emotional expressions. The word clouds reinforce the understanding gained from the bar charts by illustrating not just the most common terms, but also the qualitative nature of the sentiment towards the app, thereby deepening the overall sentiment.

### *Discussion and Benchmark Comparison*

Naïve Bayes and Support Vector Machine have the best results for classifying the sentiment of user reviews for the Digital Korlantas POLRI app. Both models achieved an approximate accuracy of 0.91, while Random Forest had an accuracy of 0.89. These results show that both margin-based and probabilistic classifiers are excellent for classifying user reviews in the Indonesian language.

The results are in line with previous research in the field of sentiment analysis of mobile applications and public service platforms. (Ahmad et al., 2018) achieved an accuracy of 0.85 - 0.90 for sentiment classification using Support Vector Machine models. Puspasari et al. (2024) found that Naïve Bayes yielded accurate results in analyzing the sentiment of user feedback in public services as his algorithm was robust in handling features. Therefore, the results of this research, approximately 0.91, are in line with the performance of traditional machine learning models.

**Table 7:** Frequency of the Top 10 Words in Negative Reviews

| No | Word | English Meaning | Frequency |
|---|---|---|---|
| 1 | aplikasi | Application | 2,541 |
| 2 | sim | Driver’s license | 1,947 |
| 3 | gak | Informal no/not | 986 |
| 4 | nya | Possessive suffix | 914 |
| 5 | panjang | To Renew | 901 |
| 6 | ga | Informal no/not | 839 |
| 7 | verifikasi | Verification | 835 |
| 8 | gagal | Failed | 719 |
| 9 | proses | Process | 607 |
| 10 | online | Online | 576 |
| | Total | | 5,639 |


**Fig. 9:** Word Cloud of Positive Reviews for the Digital Korlantas POLRI Application


**Fig. 10:** Word Cloud of Negative Reviews for the Digital Korlantas POLRI Application

Given the specifics of the dataset and the algorithm, Naïve Bayes has shown noteworthy achievements in the study. Naïve Bayes works well with high dimensional and sparse datasets. These are usually the characteristics of text datasets, such as application reviews. Each word in a text is treated as a new feature. This creates a feature space of large dimensions, many of which will be occupied by extremely few or even no instances. The Naïve Bayes algorithm works well even with sparse data as it has a very sound probabilistic data structure. Naïve Bayes also works with the assumption of feature independence. This not only simplifies the model significantly but also improves generalization in the case of short text documents, such as reviews of mobile applications. Naïve Bayes deals with a fewer number of parameters, and this simplicity, unlike other sophisticated models, reduces the chances of overfitting, especially when there is only a moderate size dataset.

While Support Vector Machine recorded similar accuracy levels, its performance may vary due to feature representation and specific parameter settings. SVM models are based on the premise of optimal hyperplane separation in high dimensional space. The effectiveness of separations hinges on how well features are transformed and how well the right kernel is chosen. In this work, the linear kernel was the best performer, but the simplicity and the way Naïve Bayes models are parameterized and optimized led to more consistency across the board.

The performance of Random Forest models was the lowest of the four. The reason may have to do with how ensemble tree-based methods work with text data. Random Forest models are best suited for data when the features are hierarchically arranged, or when the underlying data has complex nonlinear relations. In bag of words representations of text, the features may not always lend themselves to the kind of partitioning that is utilized in tree-based methods. In text classification tasks, Random Forest models can have less stable boundaries compared to classifiers that rely on class probabilities or margins.

The results of the sentiment analysis gather valuable information regarding the user's experience with the Digital Korlantas POLRI application. Positive feedback praised the application's user-friendly interface, the decreased time it takes to process requests, and the overall convenience of being able to manage administrative tasks related to driving licenses from a mobile device. This app has proven that when utilized properly, a digital public service application has the potential to greatly enhance the accessibility and efficiency of public services.

Negative feedback has pinpointed problems with online verification, system errors, and incomplete transactions. Words like failure, verification, and process highlight the system's weaknesses and the urgent requirement for improved technology and user verification processes. System reliability, user verification, and system feedback to users regarding verified users' needs improvement.

When the outcomes are considered, a sense of practicality ensues. This demonstrates that sentiment analyses can be used to assess the effectiveness of digital public services. Sentiment analyses have helped public administration pinpoint persistent issues users encounter, assess their degree of satisfaction, and prioritize changes based on user needs. Such approaches facilitate the creation of user-centric digital services in public administration.

Overall, the results reaffirm the effectiveness of classical machine learning algorithms like Naïve Bayes and Support Vector Machine for the tasks of sentiment polarity classification and analysis of user reviews in the Indonesian language. Moreover, the results illustrate the possibility of implementing sentiment analysis in the ongoing evaluation of digital public services.

## Conclusion

The research examined reviews of the Digital Korlantas POLRI app from the Google Play Store, focusing on user reviews. After several preprocessing steps, 11,277 reviews were classified and three machine learning algorithms were implemented Naïve Bayes, Support Vector Machine (SVM), and Random Forest. According to the evaluation, Naïve Bayes and SVM were the only models to meet the 91% accuracy threshold while Random Forest was only able to meet 89% accuracy. With a significant margin of over 2%, Naïve Bayes and SVM were the only models to meet the margin criterion and classify sentiments regarding reviews of mobile apps written in Indonesian.

The analysis of user reviews showed a wide variation of sentiments regarding the Digital Korlantas POLRI application. Positive sentiments were centered on quick online services and the convenience of assisting in online driver's license services while negative sentiments in this area often centered on issues regarding verification, system malfunctions, and the online driver's license service. The Digital Korlantas POLRI application offered great progress in providing public services, but issues regarding technical reliability and verification processes created a significant barrier to the Digital Korlantas POLRI application.

Notwithstanding these contributions, some limitations must be acknowledged. First, the data set was limited to reviews on the Google Play Store, miss, in all probablility, the sentiments of users on different platforms or from different demographic perspectives. Second, in sentiment classification, the author limited the reviews to a binary classification of positive and negative. This, undoubtedly, oversimplifies the user sentiments. Third, the author used some of the oldest machine learning methods that arguably do not capture the necessary depth of meaning that the use of the Indonesian language requires.

To some extent, future research can remedy the limitations by collecting data from different social media, feedback systems for governments, app store reviews from other ecosystems, and so on. Additionally, using more advanced and relevant machine learning methods that fall in the category of deep learning like BERT or even DistilBERT, can be a solution to the problems of classification and context of sentiment. It can benefit future researchers to use Aspect-Based Sentiment Analysis (ABSA) to determine which specific features of an application increase user satisfaction, providing more digital of public service funnels for improvement.

## Acknowledgment

The authors express their gratitude to Bina Nusantara University for the support and facilities provided throughout the course of this research. The authors also appreciate the constructive feedback from collaborators that contributed to improving the quality of this manuscript. Minor language and grammar refinements were assisted by automated tools, while the authors remain fully responsible for the content and interpretations presented in this study.

## Funding Information

This study was financially supported by Bina Nusantara University.

## Author's Contributions

**Evaristus Didik Madyatmadja:** Lead research project, coordinated developer, experiment, instructor, and data analysis.

**Ricky Kosasih, Najla Aurelia Evanthe, Rudy and Betley Heru Susanto:** Software programming, validation, formal analysis, investigation, data curation, writing-original draft preparation, writing-review and editing, visualization.

## Ethics

The authors confirm that this manuscript is original, unpublished, and free from any ethical conflicts.

## References

Ahmad, M., Aftab, S., Salman, M., Bashir, M. S., & Hameed, N. (2018). Sentiment Analysis using SVM: A Systematic Literature Review. *International Journal of Advanced Computer Science and Applications*, *9*(2), 182–188. https://doi.org/10.14569/ijacsa.2018.090226

Alakwaa, F. M., Chaudhary, K., & Garmire, L. X. (2017). Deep learning accurately predicts estrogen receptor status in breast cancer metastases. *BMC Bioinformatics*, *17*(1), 337–347. https://doi.org/10.1021/acs.jproteome.7b00595

Andriansyah, F., Arifin, F., & Ali, M. (2024). Analysis of user behavior of the Digital Korlantas Polri application with integrated UTAUT in the community of East Java Province. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima*, *8*(1), 118–126.

Avlach, M. Z., & Yamasari, Y. (2025). Web-based Construction Project Application Development with ReactJS, NestJS, and Agile Scrum. *Journal of Emerging Information Systems and Business Intelligence (JEISBI)*, *6*(2), 216–229. https://doi.org/10.26740/jeisbi.v6i2.70380

Camacho-Collados, J., & Pilehvar, M. T. (2018). On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis. *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 40–46. https://doi.org/10.18653/v1/w18-5406

Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *EURASIP Journal on Advances in Signal Processing*, *2021*(1), 30. https://doi.org/10.1186/s13634-021-00742-6

Choudhury, S. S. (2012). Overview of support vector machine in modeling machining performances. *Procedia CIRP*, *1*, 1–6.

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). *Why 70/30 or 80/20 relation between training and testing sets*.

Gusman, S. W. (2024). Development of the Indonesian Government's Digital Transformation. *Dinasti International Journal of Education Management And Social Science*, *5*(5), 1128–1141. https://doi.org/10.38035/dijemss.v5i5.2868

Handika, Y., Hanif, I. F., & Hasan, F. N. (2024). Analysis of Public Sentiment Towards POLRI's Performance using Naive Bayes and K-Nearest Neighbors. *International Journal on Informatics for Development*, *13*(1), 386–399. https://doi.org/10.14421/ijid.2024.4500

Jivani, A. G. (2011). A comparative study of stemming algorithms. *International Journal of Computer Technology and Applications*, *2*(6), 1930–1938.

Joseph, V. R., & Vakayil, A. (2022). SPlit: An Optimal Method for Data Splitting. *Technometrics*, *64*(2), 166–176. https://doi.org/10.1080/00401706.2021.1921037

Jusal, M. R. A. (2024). Implementation of E-Government in South Jakarta: A Case Study of Digital Services Through The Digital Application of Polri's Traffic Corps. *Eduvest - Journal of Universal Studies*, *4*(3), 1172–1186. https://doi.org/10.59188/eduvest.v4i3.1123

Kannan, S., & Gurusamy, V. (2015). Preprocessing techniques for text mining. *International Journal of Computer Science and Communication Networks*, *5*(1), 7–16.

Kurniawati, D., Prayitno, E., Sari, D. F., & Putra, S. N. (2019). Sentiment Analysis of Twitter Use on Policy Institution Services using Naïve Bayes Classifier Method. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *2*(1), 14–21. https://doi.org/10.32535/jicp.v2i1.409

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. https://doi.org/10.1017/CBO9780511809071

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *Arxiv*. https://doi.org/10.48550/arXiv.1301.3781

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*.

Nurcahyawati, V., & Mustaffa, Z. (2023). Improving sentiment reviews classification performance using support vector machine-fuzzy matching algorithm. *Bulletin of Electrical Engineering and Informatics*, *12*(3), 1817–1824. https://doi.org/10.11591/eei.v12i3.4830

Nurmalasari, S., Hidayanto, A. N., Huwaida, L. A., & Wulandari, H. (2023). Sentiment Analysis and Topic Modeling of Citizen Satisfaction with the Indonesian Government in Handling a Pandemic. *OPSearch: American Journal of Open Research*, *2*(7), 246–256. https://doi.org/10.58811/opsearch.v2i6.61

O'Connell, N. S., Jaeger, B. C., Bullock, G. S., & Speiser, J. L. (2025). A comparison of random forest variable selection methods for regression modeling of continuous outcomes. *Briefings in Bioinformatics*, *26*(2), 1–12.https://doi.org/10.1093/bib/bbaf096

Pervan, N., & Keleş, H. Y. (2017). Sentiment analysis using a random forest classifier on turkish web comments. *Communications Faculty Of Science University of Ankara*, *59*(2), 69–79.

Purba, M., & Yadi, Y. (2023). Implementation Opinion Mining For Extraction Of Opinion Learning In University. *SinkrOn*, *8*(2), 694–699. https://doi.org/10.33395/sinkron.v8i2.11994

Puspasari, H. M., Mustaqim, I. Z., Utami, A. T., Syalevi, R., & Ruldeviyani, Y. (2024). Evaluation of Indonesia's police public service platforms through sentiment and thematic analysis. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *13*(2), 1596–1607. https://doi.org/10.11591/ijai.v13.i2.pp1596-1607

Radha, R., Devi, M., & Vidhya, R. (2016). Removal of redundant and irrelevant data from dataset using entropy measure. *International Journal of Computer Science and Mobile Computing*, *5*(7), 359–364.

Rahman, M., Rahman, S. S. M. M., Allayear, S. M., Patwary, F. K., & Munna, T. A. (2020). A Sentiment Analysis Based Approach for Understanding the User Satisfaction on Android Application. *Data Engineering and Communication Technology*, *1079*, 397–407. https://doi.org/10.1007/978-981-15-1097-7_33

Rizkiansyah, Herliana, A., Alamsyah, D. P., & Tjoe, T. F. (2022). Comparison of the K-Nearest Neighbor and Decision Tree algorithm to the Sentiment Analysis of Investment Applications Users in Indonesia. *Proceedings of the 2022 Seventh International Conference on Informatics and Computing (ICIC)*, 1–6. https://doi.org/10.1109/icic56845.2022.10006970

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3411764.3445518

Sarica, S., & Luo, J. (2021). Stopwords in technical language processing. *PLOS ONE*, *16*(8), e0254937. https://doi.org/10.1371/journal.pone.0254937

Sumantiawan, D. I., Suseno, J. E., & Syafei, W. A. (2023). Sentiment Analysis of Customer Reviews Using Support Vector Machine and Smote-Tomek Links For Identify Customer Satisfaction. *Jurnal Sistem Informasi Bisnis*, *13*(1), 1–9. https://doi.org/10.21456/vol13iss1pp1-9

Utami, D. S., & Erfina, A. (2022). Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine. *Jurnal Sistem Informasi Dan Informatika (Simika)*, *5*(1), 59–67. https://doi.org/10.47080/simika.v5i1.1411

Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management*, *50*(1), 104–112. https://doi.org/10.1016/j.ipm.2013.08.006

Wankhade, P., & Patil, S. (2020). Sentiment analysis of movie reviews using SVM. *Journal of Data Science and Analytics*, *3*(4), 65–73.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2011). *Data mining: Practical machine learning tools and techniques*. 531–538. https://doi.org/10.1016/C2009-0-19715-5