ISSN 2085-4552

Sentiment Analysis of User Satisfaction with Access by KAI Application Using Support Vector Machine and Random Forest Algorithms

Muhammad Ilham Nas[1], Aji Darmawan[2], Ahmad Awaludin Amri[3], Kurnia Gusti Ayu[4]
[1,2,3,4] Faculty of Computer Science, Mercu Buana University, Jakarta, Indonesia
ilhamnas29@gmail.com[1], ajidarmawan.id@gmail.com[2],
ahmadawaludinamri18@gmail.com[3], kurnia.gusti@mercubuana.ac.id[4]

Accepted 19 June 2026
Approved 28 June 2026

***Abstract*— PT Kereta Api Indonesia (Persero), a State-Owned Enterprise (SOE) in the railway sector, holds an important responsibility in providing efficient services. To enhance their services, they introduced the application "Access by KAI." Although this application represents a technological innovation, it often falls short of users' expectations due to existing shortcomings. Therefore, this study proposes the use of sentiment analysis to gauge users' perspectives on the "Access by KAI" application. By applying Support Vector Machine (SVM) and Random Forest algorithms to user review datasets, this study classifies user sentiments into three categories: positive, neutral and negative, and compares the performance of these two algorithms. The results of this study offer valuable insights into users' attitudes towards the application, which can assist PT Kereta Api Indonesia in improving service quality. The classification results using the Random Forest and Support Vector Machine (SVM) algorithms, based on the factors used in this study, show varying performance. Random Forest achieved an accuracy of 86% for an 80:20 and 90:10 data ratios, and 85% for a 70:30 ratio, while SVM achieved 83% accuracy across different data ratios.**

***Index Terms*— Access by KAI; Sentiment Analysis; Support Vector Machine; Random Forest.**

I. INTRODUCTION

The development of information and communication technology plays an important role in supporting daily activities, including transportation. PT Kereta Api Indonesia (Persero), the only State-Owned Enterprise (SOE) under the Ministry of Transportation operating in the railway sector, introduced Access by KAI to facilitate ticket booking and related service[1]. On July 7, 2023, the application, formerly known as KAI Access, was updated with features such as ticket booking, rescheduling, cancellation, and e-boarding passes. These features are expected to improve the train travel experience. As one of Indonesia's preferred public transportation modes, trains are widely used. Transactions through KAI Access reached 9.1 million (61.77%), followed by B2B partners with 4 million transactions (27.10%), counters with 1.2 million transactions (8.47%), the KAI website with 374 thousand transactions (2.52%), vending machines with 14 thousand transactions (0.10%), and Contact Center 121 with 6 thousand transactions (0.04%).

Access by KAI is available on both the App Store and Google Play Store. However, the application has not fully met user expectations because several limitations still need to be improved. User reviews are valuable for application development, yet the large number and variety of opinions make systematic evaluation difficult. Alawaji and Aloraini stated that sentiment analysis of digital service application reviews can be used to assess user satisfaction and provide insights for service improvement strategies[2]. Currently, there is no standardized method for classifying Access by KAI user reviews into positive, neutral, or negative sentiment. Therefore, sentiment analysis is proposed as an approach to evaluate user opinions on the Access by KAI application[3]. The neutral class is important because it may represent users who do not express strong dissatisfaction but still show uncertainty, hesitation, or moderate experience with the application. In the context of user satisfaction, neutral reviews may serve as early indicators of potential service issues before they develop into negative perceptions. Therefore, identifying neutral sentiment can assist PT KAI in designing preventive service improvement strategies. Sentiment analysis is a method used to extract opinions, attitudes, and sentiments about a topic, product, or service from unstructured text data[4]. In this study, sentiment classification was conducted using classification-based machine learning methods to systematically categorize user opinions toward the Access by KAI application.

Previous studies have examined sentiment analysis and classification algorithm comparison. Rahayu and

Rahmat compared Naïve Bayes and SVM on Spotify user reviews from the Play Store, with SVM achieving 84% accuracy and Naïve Bayes achieving 86.4% [5]. Another study by Wandani, Fauziah, and Andrianingsih compared K-NN, Random Forest, and Naïve Bayes for sentiment analysis of Twitter users on flash sale events using "flash sale" and "Shopee flash sale" datasets. Their results showed accuracies of 83.53% and 81.48% for Naïve Bayes, 82.94% and 77.78% for K-NN, and 80.59% and 74.07% for Random Forest, respectively [6]. Lustiansyah, Widiyanto, and Wahyono applied Long Short-Term Memory (LSTM) to analyze sentiment toward the PeduliLindungi application based on Play Store reviews, achieving 82.44% accuracy, 78.66% precision, and 87.03% recall [7].

Based on these studies, this research compares SVM and Random Forest in analyzing sentiment toward the Access by KAI application. The dataset was collected from the Google Play Store using a web scraping technique. Unlike previous research that focused on aspect-based sentiment analysis of KAI Access reviews, this study compares SVM and Random Forest across three data-splitting scenarios and evaluates their performance at the class level for positive, neutral, and negative sentiments. This study also emphasizes the neutral class as an early indicator of potential dissatisfaction and provides practical insights for improving the Access by KAI application.

II. METHOD

This study follows a series of stages required to analyze the sentiment of Access by KAI reviews, as illustrated in the flowchart in Figure 1 below.

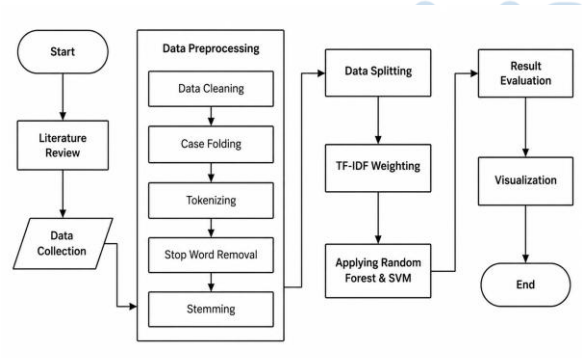


Fig. 1. Research Flowchart

A. Data Scraping

Data scraping, or web scraping, is a method for retrieving data from online sources, particularly from web pages that use markup languages such as HTML or XHTML. This process involves parsing documents to obtain the desired data, regardless of its size [8]. Web scraping automates the retrieval and extraction of data from websites, eliminating the need for manual copying[9].

B. Data Preprocessing

The data preprocessing stage is carried out to convert raw data into a higher-quality form, ensuring it can be effectively processed in subsequent steps[10]. This step addresses problematic or inconsistent data, such as data containing noise or error[11][12]. The preprocessing stage consists of five main steps: data cleaning, case folding, tokenization, stop word removal, and stemming.

1) Data Cleaning

Data cleaning involves preparing text documents by eliminating or correcting irrelevant elements, such as numbers, URLs, usernames, hashtags, and other nonessential components. The goal is to identify and fix inaccurate or incomplete data[13].

2) Case Folding

Case Folding involves converting all uppercase letters in a document to lowercase to ensure uniformity. For example, "PT Astra" is transformed to "pt astra," so that both uppercase and lowercase letters are treated the same [14]

3) Tokenization

Tokenization is a method by segmenting textual data into smaller units, known as tokens, to support individual word-level analysis.

4) Stop Word Removal

At this stage, commonly used words, including "what," "who," "and," and "to," are removed because they have limited relevance to sentiment analysis and may introduce noise into the dataset [15].

5) Stemming

Stemming aims to convert words into their base forms, thereby simplifying the text analysis process[16]. For Indonesian datasets, the Sastrawi library in the Python programming language is employed for this purpose. For instance, "*makanan*" is stemmed to "*makan*".

C. Data Splitting

In machine learning, dataset partitioning is a necessary step to support model development and evaluation. The dataset is separated into two subsets: one for training the model and another for testing its ability to make predictions on data that were not used during training. The split ratio is important because an unsuitable proportion may negatively affect the model's precision and accuracy[17].

D. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF represents textual data as weighted numerical features, making it suitable for classification models such as SVM and Random Forest that require numerical input[16]. This technique also assigns greater weight to terms that are more informative,

thereby supporting improved model performance in the classification stage.

The formula for calculating TF-IDF is:

$$tf_{td}\, id_{ft} = tf_{td} * \log\left(\frac{N}{df_t}\right) \quad (1)$$

Where:

$tftd\ \ idft$: The weighted score of a word in a document.

$tftd$: The frequency of term *t* in document *d*.

N: Refers to the total documents in the dataset.

dft: Number of documents that contain word *t*.

A high TF-IDF value reflects a term that is frequent in a specific document but uncommon in the overall corpus. In contrast, a low TF-IDF value indicates that the term appears widely across documents, reducing its distinctiveness in the analysis.

E. Support Vector Machine

Support Vector Machine is a machine learning method that applies the Structural Risk Minimization (SRM) principle. This algorithm aims to identify the most optimal hyperplane for separating two classes within the input space[18]. In classification, the hyperplane is chosen to divide the classes by maximizing the margin, defined as the distance between the hyperplane and the nearest data points from each class. SVM is particularly effective in handling data that may not have a linear relationship in high-dimensional spaces[19]. By employing kernel functions, SVM can implement non-linear and high-dimensional input separators, making it a versatile method[20][21]. In sentiment analysis, SVM is used to classify texts as positive or negative. The following is the SVM formula:

$$f(x) = w.x + b \quad (2)$$

Explanation of the formula:

f = The target value

(x) = Vector containing the input data

w = Vector of weight values

b = Bias term

F. Random Forest

Random Forest is an ensemble-based machine learning algorithm that integrates multiple decision trees to perform classification tasks[22]. The final prediction is generated through a majority voting mechanism. This method offers several advantages, including low error rates, strong performance on large datasets, and robustness against overfitting due to its ensemble learning approach[23]. Moreover, Random Forest can be used for both classification and regression tasks and is able to identify the most influential features in the training data, making it useful for feature selection [24].

$$l(y) = argmax_c\ (\sum I_{\text{h}}\text{n}(y)=c) \quad (3)$$

G. Confusion Matrix

A confusion matrix is used to evaluate a model's performance in classifying the target classes. It summarizes the model's ability to classify data correctly and highlights any errors that occur during the process[25].

TABLE I. CONFUSION MATRIX

Actual Class	Predicted Class	
	Positive	**Negative**
Positive	*True Positive* (TP)	*False Negative* (FN)
Negative	*False Positive* (FP)	*True Negative* (TN)
Positive	*True Positive* (TP)	*False Negative* (FN)

A confusion matrix contains four key components used to evaluate the performance of a classification model:

1. True Positive (TP) refers to positive instances that are correctly classified as positive
2. True Negative (TN represents negative instances that are correctly classified as negative
3. False Positive (FP) occurs when negative instances are incorrectly predicted as positive
4. False Negative (FN) occurs when positive instances are incorrectly predicted as negative.

The confusion matrix can also be used to derive several evaluation metrics:

- Accuracy measures the proportion of correct predictions made by the model, both positive and negative, out of the total predictions. The formula for accuracy is[26]:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \times 100\% \quad (4)$$

- Precision measures how well the model identifies true positives out of all the predicted positive results. The formula for precision is[27] :

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \times 100\% \quad (5)$$

- Recall (or Sensitivity) measures the model's effectiveness in identifying true positive cases from all actual positives. The formula for recall is [28]:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \times 100\% \quad (6)$$

- F1 Score is the harmonic mean of precision and recall. It is useful when the dataset has an imbalance between false negatives and false

positives. The formula for the F1 score is [29]:

$$\text{F1 Score} = \frac{2 \text{ x recall x precision}}{\text{recall+precision}} \times 100\% \quad (7)$$

SVM and Random Forest were selected because both methods have strong characteristics for text classification. SVM is effective for high-dimensional TF-IDF features because it maximizes the margin between classes, while Random Forest improves classification stability through ensemble learning and reduces overfitting compared with a single decision tree. Therefore, comparing these two methods provides a more comprehensive understanding of their suitability for sentiment classification in application review data.

III. RESULT AND DISCUSSIONS

A. Data Scraping

Review data were collected from the Access by KAI application on the Google Play Store using the google-play-scraper library. The scraping process retrieved 7,341 review records consisting of four attributes: userName, score, date, and content. A sample of the scraped review data is presented in Table 2.

TABLE II. SAMPLE OF DATA SCRAPING RESULTS FROM GOOGLE PLAY STORE REVIEWS

No.	User Name	Rating	Date	Review Content
1	ahmad fajri	1	2024-04-25 13:20:57	Reduksi tidak bisa dipakai pada aplikasi ini, ...
2	Fyanhie 03	1	2024-04-25 13:03:29	Mau daftar aja susah gimana si
3	Gilang Rizki	5	2024-04-25 12:56:27	Bagus. Aplikasi sangat baik, Bintang 5. Semoga...
4	Marmut Merah jambu	1	2024-04-25 12:55:13	Daftar member MAPCLUB ngga bisa ngisi provinsi...
5	wong mendem	1	2024-04-25 12:18:58	Tiket KA jarak jauh tidak bisa dibatalkan, tida...

B. Data Preprocessing

Data preprocessing was conducted to transform raw review data into a cleaner and more structured format for sentiment classification. This stage included sentiment labeling, data cleaning, case folding, tokenization, stop word removal, and stemming. Sentiment labeling was performed based on the rating scores obtained from Google Play Store reviews. Ratings higher than 3 were categorized as positive sentiment (1), ratings equal to 3 were categorized as neutral sentiment (0), and ratings lower than 3 were categorized as negative sentiment (-1). The threshold value of 3 was used because it represents the midpoint of the 1–5 rating scale. Therefore, a rating of 3 indicates a neutral user evaluation, while ratings above and below this value indicate positive and negative evaluations, respectively.

After the labeling process, text preprocessing was applied to reduce noise and improve data quality. Data cleaning was performed to remove URLs, numbers, punctuation marks, symbols, and irrelevant characters. Case folding was used to convert all text into lowercase. Tokenization was applied to split the text into individual terms. Stop word removal was conducted to eliminate common words that do not provide significant semantic contribution to sentiment classification. Finally, stemming was applied to convert Indonesian words into their root forms. Table 3 presents examples of review text before and after preprocessing. The results show that punctuation marks, irrelevant characters, and less meaningful terms were removed, while the remaining words were converted into a cleaner textual form for sentiment classification.

TABLE III. EXAMPLE OF REVIEW TEXT BEFORE AND AFTER PREPROCESSING

No.	Before Preprocessing	After Preprocessing
1	Reduksi tidak bisa dipakai pada aplikasi ini, ...	reduksi tidak bisa pakai aplikasi
2	Mau daftar aja susah gimana si	daftar susah gimana
3	Bagus. Aplikasi sangat baik, Bintang 5. Semoga...	bagus aplikasi sangat baik bintang semoga
4	Daftar member MAPCLUB ngga bisa ngisi provinsi...	daftar member mapclub ngga bisa ngisi provinsi
5	Tiket KA jarak jauh tidak bisa dibatalkan, tida...	tiket ka jarak jauh tidak bisa batal

C. Data Splitting

The dataset was divided into training and testing sets using three data-splitting scenarios: 90:10, 80:20, and 70:30. The training set was used to build the classification models, while the testing set was used to evaluate model performance on unseen data. These scenarios were applied to compare the consistency of the SVM and Random Forest models across different proportions of training and testing data. The number of training and testing data for each scenario is presented in Table 4.

TABLE IV. NUMBER OF TRAINING AND TESTING DATA IN EACH DATA SPLIT

Data Split	Training Data	Testing Data
90:10	6,606	735
80:20	5,872	1,469
70:30	5,138	2,203

D. TF-IDF Weighting

TF-IDF weighting was applied to convert the preprocessed review text into numerical features that could be processed by machine learning algorithms. Term Frequency (TF) represents the frequency of a term in a document, while Inverse Document Frequency (IDF) indicates the importance of a term across the entire document collection. Words that

frequently appear in a specific review but rarely appear in other reviews receive higher weights. The resulting TF-IDF feature matrix was used as input for the SVM and Random Forest classification models.

E. Model Application

The TF-IDF feature matrix was used as input for the SVM and Random Forest models. Both models were trained using the training set and evaluated using the testing set under three data-splitting scenarios: 90:10, 80:20, and 70:30. Model performance was assessed using accuracy, precision, recall, and F1-score for each sentiment class.

1) Support Vector Machine

The SVM model was applied to classify user reviews into positive, neutral, and negative sentiment classes. SVM was selected because it is effective for high-dimensional text data represented by TF-IDF features. The model works by finding an optimal separating hyperplane that maximizes the margin between sentiment classes. In this study, the SVM model achieved consistent accuracy of 83% across all data-splitting scenarios.

2) Random Forest

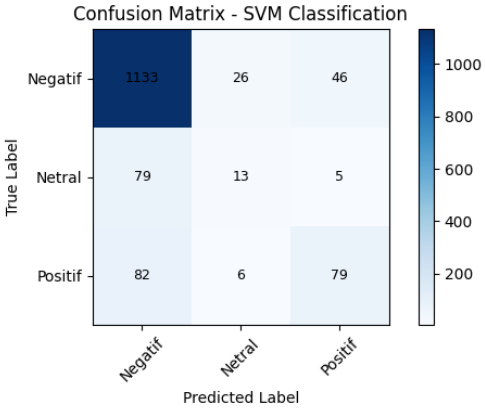
The Random Forest model was applied using the same TF-IDF features and data-splitting scenarios. Random Forest was selected because it applies ensemble learning by combining multiple decision trees, which improves classification stability and reduces the risk of overfitting compared with a single decision tree. In this study, Random Forest achieved 86% accuracy in the 90:10 and 80:20 splits and 85% accuracy in the 70:30 split.

F. Evaluation (Confusion Matrix)

Model evaluation was performed using the confusion matrix and classification report to assess accuracy, precision, recall, and F1-score for each sentiment class. For readability, only the confusion matrix results for the 80:20 data split are presented as the representative scenario, while the complete performance comparison across all data splits is summarized in Table 5.

1) Support Vector Machine

Figure 2 presents the SVM confusion matrix for the 80:20 data split. Based on the classification report, the model achieved an overall accuracy of 83%. The model performed well in detecting negative sentiment, with 88% precision, 94% recall, and 91% F1-score. However, the neutral class showed lower performance, with 29% precision, 13% recall, and 18% F1-score. For the positive class, the model achieved 61% precision, 47% recall, and 53% F1-score.

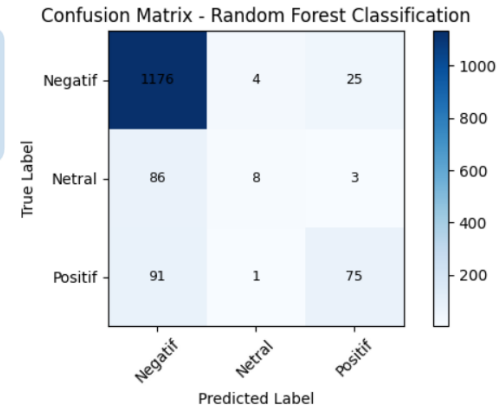


	precision	recall	f1-score	support
-1	0.88	0.94	0.91	1205
0	0.29	0.13	0.18	97
1	0.61	0.47	0.53	167
accuracy			0.83	1469
macro avg	0.59	0.52	0.54	1469
weighted avg	0.81	0.83	0.82	1469

Fig. 2. SVM Confusion Matrix Results on Test Data (20%)

2) Random Forest

Figure 3 presents the Random Forest confusion matrix for the 80:20 data split. Based on the classification report, the model achieved an overall accuracy of 86%. The model also performed strongly in detecting negative sentiment, with 87% precision, 98% recall, and 92% F1-score. However, the neutral class remained difficult to classify, with 62% precision, 8% recall, and 15% F1-score. For the positive class, the model achieved 73% precision, 45% recall, and 56% F1-score.



	precision	recall	f1-score	support
-1	0.87	0.98	0.92	1205
0	0.62	0.08	0.15	97
1	0.73	0.45	0.56	167
accuracy			0.86	1469
macro avg	0.74	0.50	0.54	1469
weighted avg	0.84	0.86	0.83	1469

Fig 3. Random Forest Confusion Matrix Results on Test Data.

IV. DISCUSSION

A. Discussion of Model Performance Results

Table 5 summarizes the performance of SVM and Random Forest across three data-splitting scenarios: 90:10, 80:20, and 70:30. The confusion matrix results are presented in a summarized form to improve readability and reduce redundancy. Overall, Random Forest achieved higher accuracy than SVM, with 86% accuracy for the 90:10 and 80:20 splits and 85% for the 70:30 split, while SVM maintained 83% accuracy across all splits. Both models performed well in detecting negative sentiment. SVM achieved recall values between 94% and 95% for the negative class, while Random Forest achieved higher recall values between 97% and 98%. However, the neutral class showed the weakest performance, with recall values below 15% in all scenarios. This indicates that many neutral reviews were misclassified as either positive or negative.

The low recall of the neutral class may be caused by class imbalance, as negative reviews dominated the dataset, while neutral reviews were relatively limited. Neutral reviews may also contain ambiguous expressions that overlap with other sentiment classes. In addition, TF-IDF focuses on word frequency rather than contextual meaning, which may limit its ability to capture subtle neutral expressions. Overall, Random Forest outperformed SVM in overall accuracy and negative sentiment detection. However, both models still require improvement in detecting neutral sentiment. Future studies may apply SMOTE or class weighting to improve minority-class detection.

TABLE V. SUMMARY OF MODEL PERFORMANCE ACROSS SPLIT DATA

Model	Split	Acc.	Negative P/R/F1	Neutral P/R/F1	Positive P/R/F1
SVM	90:10	83%	87/95/91	42/15/23	54/38/45
SVM	80:20	83%	88/94/91	29/13/18	61/47/53
SVM	70:30	83%	87/94/90	17/8/11	59/41/48
Random Forest	90:10	86%	88/97/92	71/10/17	64/46/54
Random Forest	80:20	86%	87/98/92	62/8/15	73/45/56
Random Forest	70:30	85%	87/97/92	36/6/11	73/42/53

B. Word Visualization Using Word Cloud and Bar Chart

In this study, visualizing the data content is essential for gaining insights. Visualization is a powerful tool for analyzing sentiment and ratings. Bar charts are useful for visualizing sentiment distribution, while word clouds are ideal for understanding text data. Figure 4 shows a visualization of sentiment data classified based on user satisfaction with the Access by KAI application. The data reveals that 6,043 users expressed negative sentiments, 809 users expressed positive sentiments, and 489 expressed neutral sentiments. This indicates that the majority of users are dissatisfied with the service provided by the Access by KAI application.

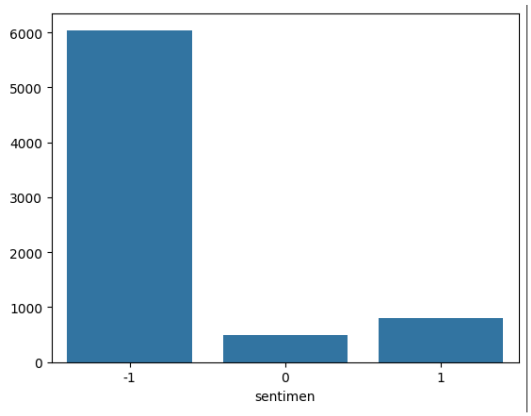


Fig 4. Sentiment Data Visualization

To better understand the content of the review text, word clouds were created for each sentiment class. Three word clouds are displayed:

1) Positive Sentiment Word Cloud

In Figure 5, positive sentiments in the Access by KAI application reviews are visualized. Frequently appearing words such as “application”, “ticket”, “very”, “easy”, “train”, “good”, “help”, “new“, and “more” highlight some of the focal points of positive user feedback.

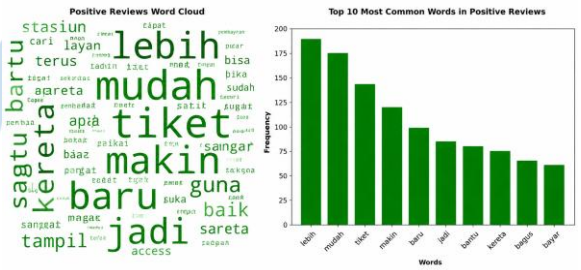


Fig 5. Word Cloud and Bar Chart for Positive Sentiment

2) Neutral Sentiment Word Cloud

Figure 6 presents the word cloud and bar chart for neutral sentiments in Access by KAI reviews. Common words such as "ticket," "application," "old," "more," "new," "good," "pay," "update," and "train" indicate the primary themes within neutral reviews.

Fig 6. Word Cloud and Bar Chart for Neutral Sentiment

3) Negative Sentiment Word Cloud

Figure 7 visualizes negative sentiments in Access by KAI reviews. Frequently used words like "application," "ticket," "old," "new," "update," "train," and "good" reflect the predominant issues raised by users in their negative feedback. The word cloud and bar chart provide a visual summary of the main words associated with negative sentiments in the app reviews.

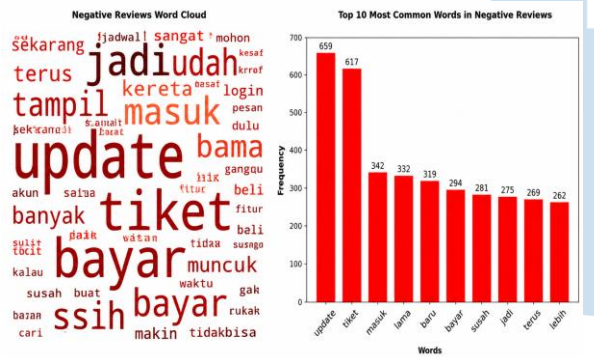


Fig 7. Word Cloud and Bar Chart for Negative Sentiment

V. CONCLUSIONS

Classification using SVM and Random Forest showed that Random Forest achieved higher accuracy, with 86% in the 90:10 and 80:20 splits and 85% in the 70:30 split, while SVM consistently achieved 83% across all splits. Both models performed well in detecting negative sentiment but showed weak performance in identifying neutral sentiment, indicating that many neutral reviews were misclassified as positive or negative. This issue may be addressed in future studies by applying SMOTE or class weighting to improve minority-class detection. The sentiment distribution showed dominant user dissatisfaction, with 6,043 negative reviews compared to 809 positive and 489 neutral reviews. Frequently occurring words indicate that user concerns were mainly related to application usability, ticketing services, payment processes, application updates, and the transition from the old version to the new version. Therefore, PT KAI should prioritize improving payment reliability, simplifying ticket booking, cancellation, and rescheduling procedures, enhancing application stability after updates, and monitoring neutral reviews as early indicators of potential dissatisfaction.

REFERENCES

[1] N. B. Sidauruk and N. Riza, "SENTIMEN ANALISIS DATA PENGGUNA TERHADAP KAI ACCESS Systematic Literature Review," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1297–1303, 2023.

[2] R. Alawaji and A. Aloraini, "Sentiment Analysis of Digital Banking Reviews Using Machine Learning and Large Language Models," *Electron. 2025,* vol. 14, 2025, doi: https://doi.org/ 10.3390/electronics14112125 Copyright:

[3] G. Radiena *et al.*, "APLIKASI KAI ACCESS MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," *J. Pendidik. Teknol. Inf. e-ISSN*, no. April, pp. 1–10, 2023.

[4] A. Dwiki *et al.*, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 2, pp. 636–646, 2021.

[5] A. S. Rahayu and A. Fauzi, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *J. Sist. Komput. dan Inform.*, vol. 4, pp. 349–354, 2022, doi: 10.30865/json.v4i2.5398.

[6] A. Wandani, "Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN , Random Forest , dan Naive Bayes," *J. Sains Komput. Inform.*, vol. 5, no. September, pp. 651–665, 2021.

[7] G. A. Lustiansyah *et al.*, "Analisis klasifikasi sentimen pengguna aplikasi pedulilindungi berdasarkan ulasan dengan menggunakan metode long short term memory," *Jakarta-Indonesia, 20 Agustus 2022*, pp. 327–336, 2022.

[8] R. Sistem, M. Hutomi, and M. Qamal, "Sistem Pengecekan Toko Online Asli atau Dropship pada Shopee," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 1117–1123, 2021.

[9] S. Satriajati and S. Panuntun, Satria Bagus; Pramana, "IMPLEMENTASI WEB SCRAPING DALAM PENGUMPULAN BERITA KRIMINAL PADA MASA PANDEMI COVID-19 (Implementation of Web Scraping in Criminal News Collection during Covid-19 Pandemic)," *Semin. Nas. Off. Stat. 2020 Stat. New Norm. A Chall. Big Data Off. Stat. IMPLEMENTASI*, pp. 300–308, 2020.

[10] S. Dwiasnati and Y. Devianto, "Classification of forest fire areas using machine learning algorithm," *World J. Adv. Eng. Tehnol. Sci.*, vol. 03, no. May, pp. 8–15, 2021.

[11] M. A. Riwanto, "ANALISIS DATA KUNJUNGAN WISATAWAN MANCANEGARA KE INDONESIA MENGGUNAKAN MICROSOFT POWER BI M . Akbar Riwanto Sistem Informasi , Fakultas Teknik dan Ilmu Komputer , Universitas Islam Indragiri Email : akbarrwnt@gmail.com Abstrak Pendahuluan Metode," *J. Sist. Inf.*, vol. 2, no. 3, pp. 112–119, 2024.

[12] H. T. Duong, T. Anh, and N. Thi, "A review : preprocessing techniques and data augmentation for sentiment analysis," *Comput. Soc. Networks*, pp. 1–16, 2021, doi: 10.1186/s40649-020-00080-x.

[13] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor," *J. Intell. Syst. Comput.*, pp. 43–49, 2020.

[14] L. Merinda, A. Abdurrahim, and L. Syafa, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 10, pp. 802–808, 2021.

[15] J. David, K. V. I. Saputra, A. M. Panjaitan, F. Victor, and P. Samosir, "Sentiment Analysis of University X Students : Comparing Naive Bayes and BERT Approaches," *Ultim. J.*

Tek. Inform., vol. 17, no. 2, 2026.

[16] D. Musfiroh, U. Khaira, P. Eko, P. Utomo, and T. Suratno, "Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. April, pp. 24–33, 2021.

[17] V. R. Joseph, "Optimal ratio for data splitting," *Wiley*, no. April, pp. 531–538, 2022, doi: 10.1002/sam.11583.

[18] B. Mahesh, "Machine Learning Algorithms - A Review," *Int. J. Sci. Res.*, vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/ART20203995.

[19] P. C. M. F, "Confidence intervals for the random forest generalization error," *Pattern Recognit. Lett.*, vol. 158, pp. 171–175, 2022, doi: 10.1016/j.patrec.2022.04.031.

[20] B. W. Rauf, "Sentimen Analisis Pertambangan Di Konawe Utara Dengan Metode Naïve Bayes," *Pros. Semin. Nas. Pemanfaat. SAINS DAN Teknol. Inf.*, vol. 1, no. 1, pp. 97–102, 2023.

[21] M. T. Anjasmoros, I. Istiadi, and F. Marisa, "Analisis sentimen aplikasi go-jek menggunakan metode svm dan nbc (studi kasus: komentar pada play store)," *Conf. Innov. Appl. Sci. Technol. (CIASTECH 2020)*, no. Ciastech, pp. 489–498, 2020.

[22] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian J. Mach. Learn.*, vol. 2024, pp. 69–79, 2024.

[23] M. P. Little, P. S. Rosenberg, and A. Arsham, "Alternative stopping rules to limit tree expansion for random forest models," *Sci. Rep.*, no. 0123456789, pp. 1–6, 2022, doi: 10.1038/s41598-022-19281-7.

[24] G. A. B. Suryanegara, Adiwijaya, and M. D. P. D. Purbolaksono, "JURNAL RESTI Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 114–122, 2021.

[25] K. G. Ayu, D. W. Sari, I. Farida, D. Harsono, and R. W. Kosaman, "Classification of Employee Competency Assessment Using Naï ve Bayes and K-Nearest Neighbor (KNN) Algorithms," *J. Adv. Inf. Technol.*, vol. 15, no. 7, 2024, doi: 10.12720/jait.15.7.879-885.

[26] U. Purba, Mariana ; Ermatita; Niprison , Handrie; Ayumi, Vina; Salamah, "Effect of Random Splitting and Cross Validation for Indonesian Opinion Mining using Machine Learning Approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, pp. 145–151, 2022.

[27] S. Sathyanarayanan and B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," *African J. Biomed. Res.*, vol. 27, no. 4, 2024.

[28] G. Zeng, "Invariance Properties and Evaluation Metrics Derived from the Confusion Matrix in Multiclass Classification," *Math. J.*, no. 1997, 2025.

[29] K. M. Sujon, R. Hassan, K. Choi, and A. Samad, "empirical evidence from advanced statistics , ML , and XAI for evaluating business predictive models," *J. Big Data*, 2025