Published online on the journal's webpage: **http://journal.ittelkom-pwt.ac.id/index.php/dinda**

Journal of Dinda

Data Science, Information Technology, and Data Analytics

Vol. 4 No. 2 (2024) 91 - 97 E-ISSN: 2809-8064

Sentiment Classification of User Reviews for KAI Access Application Using Naive Bayes Method

Rafi Andi Hidayah[1*], Rokhmatul Insani[2], Berlian Rahmy Lidiawaty[3]

[1,2,3*]Information System, Faculty of Industrial Engineering, Telkom University Surabaya

[1*]raprapandi@student.telkomuniversity.ac.id , [2]rokhmatul@telkomuniversity.ac.id , [3]berlianerel@tellkomuniversity.ac.id

Abstract

KAI Access is a train ticket booking application that offers convenience and various features for its users. However, the app has received a low rating of 2.4 out of 5 stars on the Play Store, indicating user dissatisfaction. This study conducts a quantitative sentiment analysis of the KAI Access application based on user sentiments expressed on Twitter. Using the CRISP-DM method, data were collected from Twitter with the Tweepy tool, amassing around 4,000 tweets from June to August 2023. The data underwent a preprocessing stage to ensure the quality and accuracy of the analysis. This stage involved removing duplicate tweets, eliminating retweets, and filtering out emoticons and other non-text elements. In the modeling stage, the Multinomial Naive Bayes Classifier algorithm was employed, achieving an accuracy rate of 84.6%. The model performed better at identifying negative reviews, with a precision of 0.96, recall of 0.86, and an F1-score of 0.91. In contrast, the identification of positive reviews was less effective, with a precision of 0.41, recall of 0.75, and an F1-score of 0.53. These findings shed light on the low ratings for KAI Access, particularly in the context of user reviews. The results of this study provide further understanding regarding the low rating given to KAI Access, particularly in the context of user reviews. By using this classification system, it is hoped that developers can design more specific improvements to enhance the user experience, especially in handling positive reviews which have the potential for performance improvement.

Keywords: *Sentiment Analysis, KAI Access, Naive Bayes, CRISP-DM*

1. Introduction

PT Kereta Api Indonesia has developed and launched a train ticket booking application known as KAI Access. This official app aims to meet the needs of passengers traveling on long-distance, medium-distance, and local/commuter trains. Despite the convenience it offers, KAI Access has received a low rating of 2.4 out of 5 stars on the Play Store as of June 21, 2023, indicating user dissatisfaction. This low rating necessitates an investigation into user experiences and perceptions of the app [1].

Previous studies have explored user experiences with similar train ticket booking applications. Research suggests that users with limited experience with information systems tend to focus on ease of implementation, whereas experienced users emphasize the benefits derived from the system [2], [3]. Thus, user experience levels significantly influence service evaluations. Additionally, KAI Access users have reported issues related to electronic payment processes, delays in train searches, and a cumbersome interface, which may contribute to the low user satisfaction.

Given the array of features KAI Access offers including online ticket purchasing, schedule changes, local train ticket purchases, ticket cancellations, additional ticket options, news updates, and promotional information the app is designed to provide convenience and flexibility. However, user reviews on the Play Store highlight several issues, such as slow page loading, delayed transaction responses, and frequent crashes.

Additionally, a major update to the KAI Access app's user interface (UI) has caused difficulties for users in navigating and using the app's features. This update introduced new design elements that, while visually appealing, have led to confusion among users accustomed to the previous layout. As a result, users have reported challenges in accessing essential features and completing transactions efficiently.

Received: 18-07-2024 | Accepted: 01-08-2024 | Published: 04-08-2024

These problems disrupt user experiences, prompting a need for a more detailed analysis of user sentiments. Understanding these sentiments can help developers address these issues and enhance the app's overall usability.

The objective of this study is to analyze user sentiments that may contribute to the low rating of KAI Access and explore user perceptions of its features. Utilizing sentiment analysis can identify key issues frequently mentioned in user reviews, such as app features, ease of use, and service quality [4]. By employing classification methods like Naïve Bayes, this study aims to provide clear and significant insights into user sentiments, guiding developers in making targeted improvements to enhance user satisfaction. The findings will serve as a valuable reference for developers in addressing user needs and improving the overall user experience.

This study seeks to develop a sentiment classification model using the Naïve Bayes method to offer valuable insights for further development and enhancement of the KAI Access application. By understanding user sentiments, developers can identify areas for improvement and effectively respond to user expectations, ultimately aiming to improve the app's rating and user satisfaction.

2. Research Methods

This study employs a quantitative approach using primary data collected through Twitter data scraping. The methodology used is the Cross Industry Standard Process for Data Mining (CRISP-DM). By following the CRISP-DM method, PT Kereta Api Indonesia (Persero) can start with initial stages to understand the business context and the data related to the usage of the KAI Access application [5].

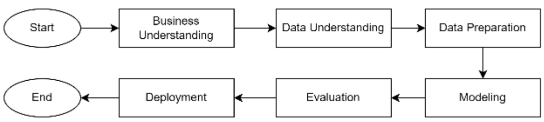

Figure 1. CRISP-DM Flowchart

2.1. *Business Understanding*

Business Understanding focuses on understanding the business problems, goals, and requirements, as well as how data mining can help address them [6]. For the KAI Access application, the main identified issue is the low rating (2.4 out of 5) on the Play Store, indicating user dissatisfaction. By analyzing this data, specific patterns such as login issues, user interface problems, or difficulties in the ticket booking process can be identified and addressed in the development and improvement of the application. The primary objective is to enhance user satisfaction and improve the app's rating. This involves sentiment analysis of user reviews to pinpoint specific areas for improvement, such as performance, speed, and features of the app.

2.2. *Data Understanding*

Data Understanding in the CRISP-DM methodology focuses on collecting and evaluating data for the analysis process [6]. Data was collected using the Twitter scraping method with the Tweepy library, gathering data using the keyword 'KAI Access', which is the mobile application of Kereta Api Indonesia. From July 2, 2023, to August 23, 2023, Tweepy scraping yielded 4,753 tweets. Data collection was carried out in 2023 because a significant update to the application was released at that time, leading to user sentiments that could provide future options for KAI Access developers. This research was completed in 2024, allowing ample time for a thorough analysis of the user feedback and sentiment trends, which are critical for making informed decisions and improvements to the application moving forward.

2.3. *Data Preparation*

In the data preparation stage, data is cleaned, transformed, and organized to ensure optimal quality and usability before analysis [7]. Data selection focused on retweets and avoiding duplicate tweets to reduce bias in the analysis. During the data cleaning stage, usernames, links, and punctuation were removed to ensure the tweet text was free of irrelevant elements. This cleaning process aims to enhance data quality and prevent potential distortion of analysis results. The next step is stemming, where words in the tweet text are simplified to their base form to improve consistency and reduce data dimensions. Following this, automatic labeling using a lexicon assigns sentiment to each tweet based on the words found in the lexicon, allowing researchers to automatically label each tweet as positive or negative.

2.4. *Modeling*

This study uses the Naïve Bayes algorithm to classify sentiments. Naive Bayes is a simple and effective classification method for many sentiment analysis tasks. It does not require large data storage and often provides clear results. The strengths of Naive Bayes lie in its simplicity, speed, and its ability to handle uncertainty while providing a probabilistic approach to classification. There are three commonly used types of Naïve Bayes:

Bernoulli Naïve Bayes, suitable for binary classification with binary attributes like spam detection; Multinomial Naïve Bayes, suitable for text data with attributes that can take on more than two values, often used in document classification; and Gaussian Naïve Bayes, effective for continuous data that follows a normal distribution [8]. By testing different models, it can be determined if there is a model that is more suitable or performs better for the specific characteristics of the dataset used [9], [10]. Therefore, Naive Bayes is chosen for sentiment analysis due to its ability to deliver clear and valuable insights for understanding and exploring sentiment data.

2.5. *Evaluation*

Model performance evaluation is conducted after model testing to measure the results of the chosen method [5]. The model testing stage uses a 2x2 Confusion Matrix. Confusion Matrix summarizes classification performance by tallying the number of instances correctly classified for each category based on the true value of the predicted class of objects [11].

Table 1. Table software and supporting hardware

Confusion Matrix	Actual Positive	Actual Negative
Predict Positive	TP	FN
Predict Negative	FN	TN

TP = Number of correctly classified positive instances

TN = Number of correctly classified negative instances

FP = Number of incorrectly classified positive instances

FN = Number of incorrectly classified negative instances

2.6. *Deployment*

In the Deployment stage, several steps need to be prepared. First, feature classification results obtained from various features of KAI Access are compiled. Feature classification is derived from Document Frequency obtained during the preprocessing phase.

The next step is visualization using Looker Studio, employing diagrams to represent data results, such as donut charts, time series diagrams, total tweet counts, positive and negative counts, and detailed tweet texts. Finally, a dashboard is created as a visual platform to facilitate easier and more effective access and analysis of data for users[12]. By hosting the dashboard on a web platform, it can be accessed from anywhere and at any time, as long as users have an internet connection, greatly enhancing accessibility for those who want to view the sentiment analysis results of KAI Access. The web allows for the development of a user-friendly interface, making it easy to interact with and navigate through various features of the dashboard, enabling users to filter, search, and view data visualizations without needing to download additional applications.

3. Results and Discussion

The series of research results is based on a logical sequence / arrangement to form a cherry. The content shows facts / data and do not discuss the results. Can use Tables and Numbers but does not decipher repeatedly against the same data in images, tables, and text. To further clarify the description, you can use subtitles.

Data collection was conducted by scraping Twitter, structuring the data into key columns: 'tweet', 'tweet id', 'date', 'location', 'username', 'source', 'language', 'likes', and 'retweet'. This process involves organizing and cleaning the data to ensure the quality and relevance of the information used for further analysis. The next step is to clean the data by ensuring that all tweets are in Indonesian. This is crucial to maintain language consistency in the analysis and to avoid distortion due to language differences [11]. The author cleaned the data using the filter feature on a spreadsheet.

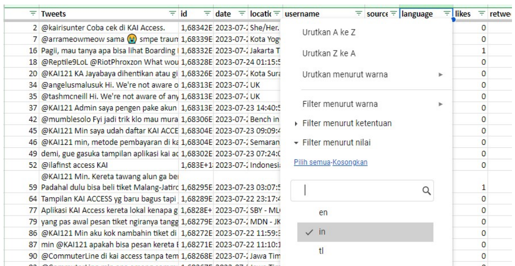

Figure 2. Language Filter Process

In Figure 2, the process shown is the language filtering using a spreadsheet. This was done to facilitate the analysis process and ensure language consistency. The filtered data resulted in 2,964 entries. Additionally, retweets were eliminated to ensure the dataset contained only original content. This helps reduce duplication and maintain the uniqueness of each data point. Tweets in the form of 'menfess' (mention confession), which contain personal opinions or confessions not directly related to 'KAI Access', were also removed.

Figure 3. Language Filter Process

In Figure 3, the process shown is tweet filtering by specifying keywords to be removed, such as RT (retweet) and menfess. This ensures the dataset does not contain duplicates and focuses on tweets related to KAI Access. The filtered data, excluding retweets, menfess, and unrelated tweets, resulted in 1,037 entries. Through this process, a clean and relevant dataset is provided, which is crucial for the subsequent analysis and modeling stages in CRISP-DM. The filtering process resulted in 1,037 data points out of the initial 4,753. This careful data cleaning and filtering ensure that the information analyzed accurately reflects the perceptions and experiences of KAI Access app users.

Afterward, punctuation marks, links, and emojis were removed to ensure clarity in the analysis. Additionally, the entire text dataset was converted to lowercase to maintain uniformity and prevent redundancy. These steps are effective in providing cleaner data for further analysis, minimizing distortion and unnecessary variability. The process of removing mentions was carried out using a Python library called re (Regular Expression Operation). The regular expression used can be seen in Table 2.

Table 2. Cleaning Process

Before	After
@KAI121 hai aplikasi KAI access lg eror kah? Kok nggak ada update tiket hari sabtu udah diriset apknya ttp sma ini gmn ya? https://t.co/uORqf8tZS8	hai aplikasi kai access lg eror kah kok nggak ada update tiket hari sabtu udah diriset apknya ttp sma ini gmn ya
@KAI121 @keretaapikita Apa KAI Access sedang mengalami gangguan? Ada penumpang yang tidak bisa masuk ke KAI Access. Saya bantu untuk hapus cache, hapus data, maupun reinstall apps, tetap tidak bisa. Mohon bantuannya, terima kasih. https://t.co/szSQqsgQZa	apa kai access sedang mengalami gangguan ada penumpang yang tidak bisa masuk ke kai access saya bantu untuk hapus cache hapus data maupun reinstall apps tetap tidak bisa mohon bantuannya terima kasih
@KAI121 min setiap aku masukin NIK di kai access gabisa terus selalu gini tu kenapa yaa☹️ https://t.co/wYMaGOzVqk	min setiap aku masukin nik di kai access gabisa terus selalu gini tu kenapa yaa
@KAI121 min aplikasi ku ga bisa dpt jadwal ya. Masih pake aplikasi KAI access, cm cek d Playstore udh acces by kai https://t.co/RFCgtI956A	min aplikasi ku ga bisa dpt jadwal ya masih pake aplikasi kai access cm cek d playstore udh acces by kai

Next, the model was created by importing the specified modules and then initializing the Naive Bayes model with the defined training data. In this study, a test_size of 0.3 was used, indicating that 30% of the entire dataset was allocated for testing. To achieve the highest accuracy for the created model, the researcher looping on the random state. The random state is a parameter used in several machine learning algorithms to control randomness during the model training process. In this iteration, the researcher performed 200 repetitions, achieving the highest accuracy of 0.84 at the 16th random state and obtaining a test set of 312 data points. Below are the results from the 200 iterations of the random state.

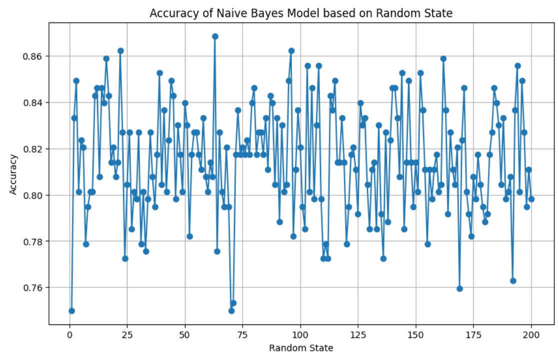

Figure 4. Naive Bayes Model Accuracy Based on Random State

After successfully identifying the Multinomial Naive Bayes classification model, the next step was to test alternative models such as Bernoulli and Gaussian. To ensure a consistent comparison, the tests were conducted using the same test size and random_state value. By evaluating different models, it is possible to determine if a more suitable model exists for handling the specific characteristics of the dataset used.

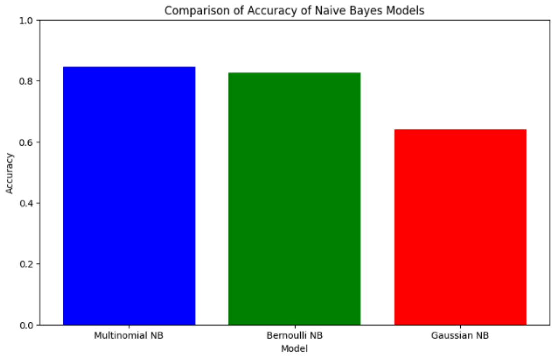

Figure 5. Comparison of Naive Bayes Model Accuracy

In Figure 5, the comparison of Naive Bayes model accuracies shows that the Multinomial Naive Bayes model delivers the best performance with an accuracy of 0.84. Meanwhile, the Bernoulli Naive Bayes model has a slightly lower accuracy of 0.82, and the Gaussian Naive Bayes model exhibits an even lower accuracy of 0.63. Accuracy measures how well a model can correctly predict test data.

In addition to comparing with other Naive Bayes models, this research also contrasts Naive Bayes with Decision Tree models to further evaluate their effectiveness, as shown in the following figure.

```
Naive Bayes Classification Report:
              precision    recall  f1-score   support

          -1       0.95      0.86      0.91       272
           1       0.43      0.70      0.53        40

    accuracy                           0.84       312
   macro avg       0.69      0.78      0.72       312
weighted avg       0.88      0.84      0.86       312

Naive Bayes Accuracy: 0.842948717948718

Decision Tree Classification Report:
              precision    recall  f1-score   support

          -1       0.89      0.86      0.87       270
           1       0.25      0.31      0.28        42

    accuracy                           0.78       312
   macro avg       0.57      0.58      0.57       312
weighted avg       0.80      0.78      0.79       312

Decision Tree Accuracy: 0.782051282051282

Naive Bayes Accuracy: 0.842948717948718
Decision Tree Accuracy: 0.782051282051282
```

Figure 6. Classification Report

Figure 6 shows that the Naive Bayes model has an overall accuracy of 84.29%, while the Decision Tree model has an accuracy of 78.20%. In terms of precision, recall, and f1-score for class -1 (negative), the Naive Bayes model demonstrates higher performance (0.95, 0.86, and 0.91) compared to the Decision Tree (0.89, 0.86, and 0.87). For class 1 (positive), the performance of Naive Bayes with precision, recall, and f1-score of 0.43, 0.70, and 0.53, respectively, while the Decision Tree has lower values for precision, recall, and f1-score of 0.25, 0.31, and 0.28. Therefore, selecting the Multinomial Naive Bayes model as the best choice can be considered appropriate for the classification task on the given dataset. Based on the sentiment analysis results using the Naive Bayes model, the obtained accuracy indicates a good category, reflecting the model's ability to classify sentiment with a high degree of accuracy.

After creating the Naive Bayes model, the next step is to evaluate its performance. This evaluation aims to measure the effectiveness of the chosen method. During the model testing phase, a 2x2 Confusion Matrix was generated, as shown in Table 3.

Table 3. Result Confussion Matrix

Confusion Matrix	Actual Positive	Actual Negative
Predict Positive	28	12
Predict Negative	37	235

After the model evaluation is completed, the next step is to analyze the classification results. The classification results, formatted as a CSV file, can be visualized using Looker Studio. The display of the CSV file discussing the KAI Access app login on Looker Studio is shown in the following image.

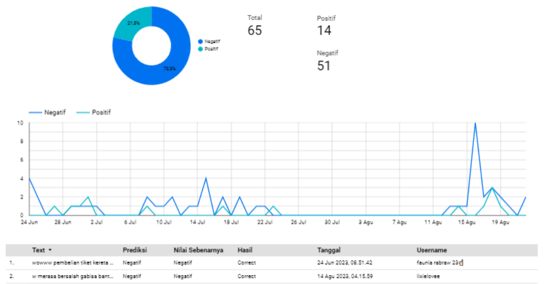

Figure 7. Data Visualization with Looker Studio

Figure 7 illustrates the Looker Studio interface filtered by the keyword "login." The next step involves filtering with different keywords found in the Document Frequency. Once the Looker Studio visualization is created, the next step is to build a dashboard. In this research context, creating a dashboard involves not only organizing data visualizations but also implementing the five feature classifications derived from the study. These classifications include promotions, the login process, rail food services, user interface (UI) and user experience (UX), and the payment system. These findings are based on the Document

Frequency technique, which reveals the frequency and relevance of specific topics within the collected text data.

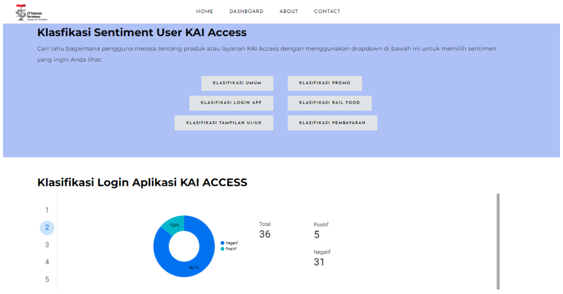

Figure 8. Dashboard

Figure 8 provides one example of the category specification of the previously processed data with a focused analysis of user sentiment related to the KAI Access application login experience. By breaking down the feedback into specific categories, sentiment analysis helps identify and prioritize areas for improvement, ensuring a more user-centric approach to improving the overall application experience. This helps other users with the information they need to make decisions quickly. By looking at the latest and most relevant data, decision making can be done without needing to dive into complicated data details.

4. Conclusion

This research developed a sentiment analysis model using the Multinomial Naive Bayes algorithm by collecting data from Twitter. The model-building process included preprocessing steps such as removing punctuation, usernames, and duplicate data. The subsequent classification using Multinomial Naive Bayes proved to be accurate, achieving an accuracy score of 0.842 or 84.2%. The model performed well in identifying the negative class (class 0), with a precision of 0.95, recall of 0.86, and an f1-score of 0.91, indicating that most of the model's predictions for the negative class were correct. However, for the positive class (class 1), the model's performance was lower, with a precision of 0.43, recall of 0.70, and an f1-score of 0.53. This suggests that the model's ability to identify positive sentiments is less optimal, highlighting potential areas for performance improvement in this class. ability to classify sentiment with a high degree of accuracy.

This analysis provides valuable insights that help other users quickly make informed decisions. By looking at the latest and most relevant data, decision-making can be done efficiently without needing to dive into complicated data details. This streamlined approach ensures that users can access essential information promptly, facilitating better and faster decisions

References

[1] A. and D. J. and G. R. and C. C. A. Aguirre Andrés F. and Villareal-Freire, "Extending the Concept of User Satisfaction from Customer Experience," in *Design, User Experience, and Usability. Design Philosophy and Theory*, W. Marcus Aaron and Wang, Ed., Cham: Springer International Publishing, 2019, pp. 397–407.

[2] G. Radiena and A. Nugroho, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine," 2023.

[3] M. N. Jebril Hend Ali Musbah and Ahmad, "Review of User Satisfaction Models in the Context of Digital Libraries Setting," in *Advances in Visual Informatics*, P. and S. A. F. and D. O. R. L. and J. B. N. and K. S. T. and A. K. R. and M. U. H. and A. M. N. Badioze Zaman Halimah and Robinson, Ed., Singapore: Springer Nature Singapore, 2024, pp. 234–246.

[4] U. Muhammadiyah Jember, M. Izunnahdi, G. Aburrahman, and A. Eko Wardoyo, "Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access Di Google PlayStore Menggunakan Metode Multinomial Naive Bayes," 2023. [Online]. Available: http://jurnal.unmuhjember.ac.id/index.php/JST

[5] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.

[6] Y. A. Singgalen, "Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM," *Journal of Information System Research (JOSH)*, vol. 4, no. 2, pp. 614–625, Jan. 2023.

[7] D. Astuti, A. Rahmat Iskandar, and A. Febrianti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," vol. 1, no. 2, pp. 60–072, 2019.

[8] H. P. Fitrian *et al.*, "Implementasi Metode Naive Bayes Classifier Untuk Aplikasi Filtering Email Spam Dengan Lemmatization Berbasis Web," 2018.

[9] A. T. Zy, "Comparison Algorithm Classification Naive Bayes, Decision Tree, And Neural Network For Analysis Sentiment," 2017.

[10] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, "Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification," in *2019 International Conference on Automation, Computational and Technology Management, ICACTM 2019*, Institute of Electrical and Electronics Engineers Inc., Apr. 2019, pp. 593–596.

[11] N. Putu Gita Naraswati, D. Cindy Rosmilda, D. Desinta, F. Khairi, R. Damaiyanti, and R. Nooraeni, "SISTEMASI: Jurnal Sistem Informasi Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," 2021. [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[12] C. T. Gonçalves, M. J. A. Gonçalves, and M. I. Campante, "Developing Integrated Performance Dashboards Visualisations Using Power BI as a Platform," *Information (Switzerland)*, vol. 14, no. 11, Nov. 2023.