Journal of Social and Political Sciences

The Asian Institute of Research *Social and Political Sciences* is a peer-reviewed International Journal. The journal covers scholarly articles in the fields of Social and Political Sciences, which include, but are not limited to, Anthropology, Government Studies, Political Sciences, Sociology, International Relations, Public Administration, History, Philosophy, Arts, Education, Linguistics, and Cultural Studies. As the journal is Open Access, it ensures high visibility and the increase of citations for all research articles published. The *Journal of Social and Political Sciences* aims to facilitate scholarly work on recent theoretical and practical aspects of Social and Political Sciences.

ASIAN INSTITUTE OF RESEARCH
Connecting Scholars Worldwide

Text Mining Algorithm Naive Bayes Classifier to Improve Quality Sentiment Analysis Passport Mobile Application

Wilonotomo[1], Budy Mulyawan[2], M. Ryanindityo[3], Muhammad Alvi Syahrin[4], Feni Yuli Triana[5]

[1,2,3,4] Immigration Polytechnic, Ministry of Law and Human Rights, Indonesia
[5] Immigration Analyst Directorate General of Immigration, Ministry of Law and Human Rights, Indonesia

Correspondence: Wilonotomo, Immigration Polytechnic, Ministry of Law and Human Rights, Indonesia.
Email: wilo@poltekim.ac.id

Abstract
Mobile Passport is an application that can be used as a digital service for people in Indonesia to apply for a new passport and an official online passport replacement from the Directorate General of Immigration replacing APAPO (Online Passport Service Application). User reviews of the Mobile Passport application are the output of big data generated as a result of the Internet of Things. The problem formulation in this research is how the implementation of the Naive Bayes text mining classifier algorithm can analyze the reviews contained in the Mobile Passport application as well as the accuracy, precision and recall values. This research uses the KDD (Knowledge Discovery and database) method which consists of data selection, data preprocessing, transformation, data mining, and evaluation using the R Studio tool. The resulting knowledge and information from this process is used as a useful knowledge base in decision making. The Naive Bayes classifier algorithm method in this research is used because of its reliability in handling data quickly and accurate predictions based on class probabilities, thus enabling research to obtain consistent and reliable results.

Keywords: Text Mining, Algorithm, Naive Bayes Classifier, Sentiment Analysis, KDD, Mobile Passport

1. Introduction

The development of information and communication technology encourages society to enter the era of Society 5.0, this era is a condition of society that can overcome various obstacles and social problems by utilizing various technologies that always make innovations born in the era of the Industrial Revolution 4.0 such as large amounts of data (Big Data), artificial intelligence (Artificial Intelligence), internet of everything (Internet on Things) and robots to improve the quality of human life (Kantar & Kilimci, 2023). This condition is directly proportional to the increase in internet users in Indonesia. The results of the 2022-2023 survey show that the number of Internet users in Indonesia in 2023 recorded by the Indonesian Internet Service Provider Association (APJII) is more than 196,700,000 people out of a total population of approximately 266,910,000 people. Thus, the percentage of internet users in Indonesia is 73.7%, this figure is greater than the survey in 2018 of 64.8% and has increased over time. This is evidenced by the survey period, internet users grew 8.9% or a total of 25,537,353 people (APJII, n.d.).

Google Play and Apple apps are platforms that provide the Mobile Passport Application, the Google Play website states that "Mobile Passport is an application that can be used by the public to submit new passport applications and replace passports online receipts from the Directorate General of Immigration, Ministry of Law and Human Rights of the Republic of Indonesia." This application is here to improve the quality of service (McConkey & Olukoya, 2023). The Mobile Passport application replaces the function of the Online Passport Service (APAPO) application by adding more complex application features and functions. Based on evaluations and observations on the social media of the directorate general of immigration in November 2023, the use of the Mobile Passport application has several obstacles and problems that occur both during the trial and at the time of its current use. Both of these mobile-based applications can be downloaded on Google Play and App Store (Chembakottu et al., 2023).

The switch to the use of the APAPO application to Mobile Passport has several obstacles and problems that occur both during the trial and at the time of its current use. Both of these mobile-based applications can be downloaded on Google Play and App Store (Lee et al., 2023). Users of the Mobile Passport application can provide reviews and ratings on the application. The provision of this review is one of the data outputs generated from the use of the internet in this online-based passport service. With the increasing amount of text data from Mobile Passport Application reviews. Then the data can be processed into information in detail and intact. Google Play itself is a service operated by Google. This website has stores or media used to sell products such as apps, games, music or songs, and books. Users can access Google Play Services from an Android smartphone (Play Store), website, or Google TV (Sinclair et al., 2024). The Apple Apps site is a service operated by Apple, just like Google Play, Apple Apps also provides application products, games, and music/songs, users can also access Apple Apps services from iOS smartphones (Apps Store) (Nokkaew et al., 2024). The review feature from its users is interesting to these two sites because it is useful as a benchmark for the responses of application users downloaded on the platform. Application reviews that contain positive feedback suggestions or negative complaints can affect prospective passport applicants who will download the application. Getting review data is not an easy thing because the number of reviews is quite large and makes it difficult if data is taken manually. So a method is needed that can filter and retrieve review data automatically and quickly in classifying negative reviews and positive reviews (Nilashi et al., 2023).

Users of the Mobile Passport application can provide reviews and ratings on the application. The provision of this review is one of the data outputs generated from the use of the internet in this online-based passport service. With the increasing amount of text data from Mobile Passport Application reviews. Then the data can be processed into information in detail and intact (Sadare et al., 2023). Google Play itself is a service operated by Google. This website has stores or media used to sell products such as apps, games, music or songs, and books. Users can access Google Play Services from an Android smartphone (Play Store), website, or Google TV. The Apple Apps site is a service operated by Apple, just like Google Play, Apple Apps also provides application products, games, and music/songs, users can also access Apple Apps services from iOS smartphones (Apps Store). The review feature from its users is interesting to these two sites because it is useful as a benchmark for the responses of application users downloaded on the platform. Application reviews that contain positive feedback suggestions or negative complaints can affect prospective passport applicants who will download the application (Qureshi et al., 2023). Getting review data is not an easy thing because the number of reviews is quite large and makes it difficult if data is taken manually. So a method is needed that can filter and retrieve review data automatically and quickly in classifying negative reviews and positive reviews.

In this study, the implementation of text mining was carried out which is part of data mining, where the data is in text format, all data for text mining is in text format (Perea-Khalifi et al., 2024). Text mining is a fairly broad term that describes various technologies for analyzing and processing semistructured and unstructured text data. In the process, technology is needed to meet the need to "convert text into numbers" so that powerful algorithms can be applied to process large databases in a document to convert text into structured data, in the form of numbers, and apply analytical algorithms certainly requires knowledge of how to use and combine these techniques. The initial challenge was to convert this text data into numerical format for subsequent analysis. One of the criteria of text mining is that the review data of the Mobile Passport application that enters the Google Play and Apple Apps sites continues to grow every day, this makes it difficult for related agencies to obtain comprehensive information data

from all existing reviews (Grotenhuis et al., 2024). So a special method is needed to collect large amounts of data and is not done manually. The right special technique to collect this information data is the web scraping technique. In terms, scraping is a technique in the process of data collection, analysis, and visualization. In other words web scraping is the process of extracting data sourced from websites (Xi et al., 2022).

The Sentiment Classification of Mobile Passport application reviews is carried out using the Naive Bayes Classifier Algorithm. That's because, according to Gary Miner in the chapter on the history of text mining, this method is a simple "probabilistic" classifier based on Bayes' theorem, i.e. the probability of an event equals the intrinsic probability (data available now), likely to occur again in the future (based on knowledge of past events), this idea is used in Naive Bayes classification only to the extent that the proportion of each class in a data set reflects an "unclear" probability. That each new object will fit in one class or another (Han et al., 2024). This classification makes a very simple (naïve) assumption that all objects to be classified are completely independent of each other in the terms used to characterize them. Despite the simple design and naïve assumptions that rarely occur in the real world, this classification can be very efficient and accurate, especially when the number of variables is high. The resulting classification is in the form of positive sentiment and negative sentiment, the negative sentiment obtained will be followed up to find out what factors must be done to improve the results of the negative reviews obtained (Thiengburanathum & Charoenkwan, 2023). This Mobile Passport Application Review is projected to be a reference for improvements to this Application to reduce problems that are obstacles in immigration service activities and increase applicant satisfaction to get effective and efficient immigration services in all Immigration Technical Implementation Units in Indonesia.

Figure 1: The wedge between sentiment mobile app analytics and text mining

Furthermore, judging from the picture above, it can be described that Text mining and sentiment analysis contained in mobile applications have a very close relationship and intersect in the context of scientific research (Hamzah et al., 2024). By using the text mining method, text data contained in various sources, such as reviews in the Mobile Passport application that have been documented in Google Play online, can be analyzed in depth to extract important information or relevant patterns. Then, by utilizing advances in mobile technology, the results of the analysis can be implemented into a very sharp and responsive mobile application (Lee et al., 2023). The app can

provide users with instant access to the latest information, customized recommendations, or even personalized solutions, creating an environment where scientific research and mobile technology collaborate effectively to support users in exploring and utilizing information efficiently.

2. Method

In an era where mobile apps have become an integral part of everyday life, this study highlights the importance of careful data analysis to improve the functionality and usability of apps. The use of Naïve Bayes methods in data analysis from mobile applications allows researchers to identify users' behavior patterns, preferences, and their needs. By utilizing this classification algorithm, applications can be better tailored to the individual needs of users, improving the overall user experience. In the scientific description, the study provides valuable insights into how information technology, particularly in the context of mobile applications, can be maximized to better meet user expectations and demands. Using data from a wide span of time, the research builds a solid foundation for the development of applications that are more adaptive and responsive to future user needs. Mobile application research and research using text mining using the Naïve bayes Classifier algorithm show a close relationship in the time span of 2017 to 2023 in the context of information technology development reflected in the figure below.
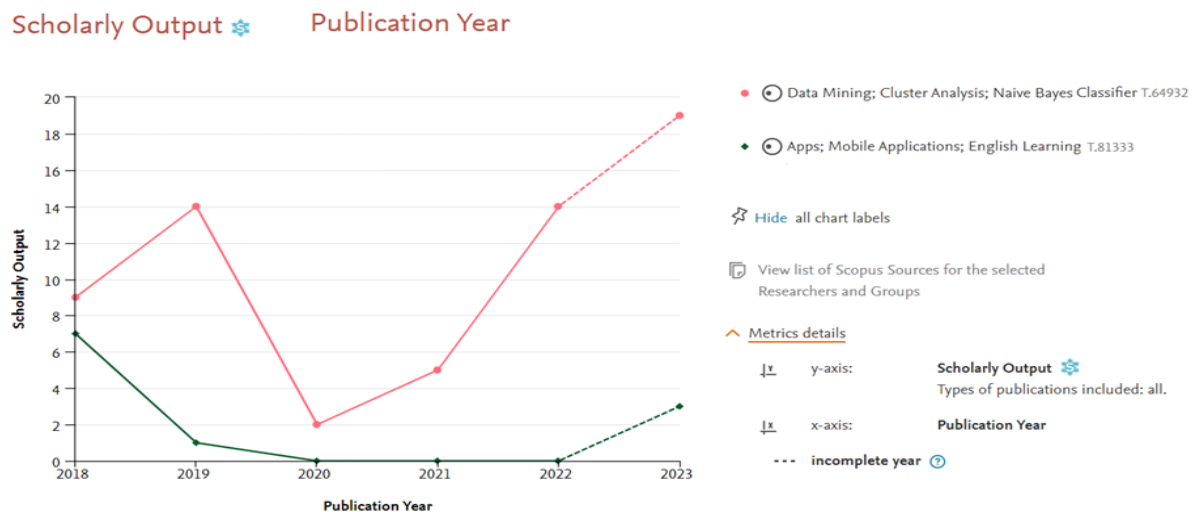
Figure 2: Text Mining and Mobile Application Research Data

The application of text mining is driven by trial-and-error experiments based on the experience of previous researchers (Huo et al., 2024). While relative data mining methodologies have long been used (e.g. CRISP-DM, SEMMA, KDD), Available data measures encourage early sampling and simplification activities. As the older brother of text mining, data mining through the same process has been around since the early 1990s developing several well-known methodologies. In this study, researchers conducted text mining, then researchers determined the right method using the Knowledge Discovery in Databases (KDD) method (Alazab et al., 2024). Understanding Knowledge Discovery in Databases is a method of obtaining knowledge from existing databases. There are relationships in existing databases, the results of knowledge and information from the process are used as a knowledge base that is useful in decision-making. KDD in data mining is often used to dig up information hidden in databases in very large quantities (Yu et al., 2023). This research design refers to the Knowledge Discovery in Database (KDD) model which has 5 stages, namely data selection, data preprocessing, transformation, data mining, and evaluation. Data Selection, Preprocessing, Transformation, Data Mining, and Evaluation.
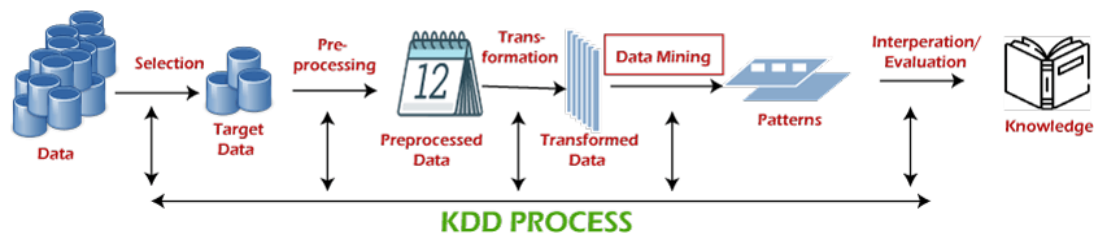
Figure 3: Model Desain Knowledge Discovery in Database (KDD)

3. Results

Text mining using the Naive Bayes Classifier algorithm has been widely done, but each study has a different update (novelty) and object of research. The research to be carried out uses the object of research on the Mobile Passport application officially issued by the Directorate General of Immigration of Indonesia. Contributions from each journal above are used as material related to the collection of theories, methods, and techniques of text mining carried out. Several journals are listed so that the research carried out is stronger and has a reference basis in its implementation. Then a process flow is carried out to parse the text mining process following the Knowledge Discovery and Database (KDD) method consisting of data selection, data preprocessing, transformation, data mining, and evaluation (Liu et al., 2024).

3.1 Data Selection

The data taken is primary data sourced from Google Play and Apple Apps sites, data collection is carried out using web scrapping techniques using Python programming languages through colab.research.google.com website. The next step, open the website that will be scrapping the review data, namely Mobile Passport on the Google Play site to see the App ID in the address. Based on the version history of the Mobile Passport application, limited data is retrieved since the last version and the version before the last to get a more accurate sentiment analysis based on the latest application conditions in version 5.0.3 and version 5.2.1. Review data obtained from both sources (Google Play, Apple Apps) is put together in the form of file.csv with a total of 1,598 reviews from October 1, 2023, to December 30, 2023, consisting of 1,358 reviews sourced from the Google Play site and 240 reviews sourced from the Apple Apps site.

3.2 Preprocessing Data

Data processing is done using descriptive analysis and machine learning methods. The descriptive analysis method is used to describe in general the Mobile Passport Application reviews contained in the Google Play and Apple Apps websites then data processing with machine learning methods, in this case, the data that has been obtained will be preprocessed data which includes translating foreign language reviews, spelling normalization, case folding, deleting numbers and punctuation, filtering and tokenizing (Ashar & Maharani, 2022).

3.3 Transformation

The transformation stage of the KDD method can be run as a model representation, this model representation is made from data that is still in the form of sentences that are still arranged in the form of words into numbers so that they can be processed and calculated (Farooqi et al., 2023). Each review will be assigned a sentiment label based on the sentiment score that has been obtained. The sentiment score of each word is obtained based on the number of words detected in the lexicon dictionary, the lexicon dictionary is commonly called the dictionary of opinion words which usually take the form of positive and negative. Output files with the name format 'pelabelan.csv' are the results of labeling sentiment scores with 3 kinds of classification results including negative, neutral, and positive reviews. Based on the calculation of sentiment class scores and labeling, the number of sentiments with a score smaller than zero (<0) was 495 reviews or 31%, reviews with a sentiment score equal to

zero (=0) were 827 reviews or 52%, and 17% of reviews had a sentiment score of more than zero (>0) with a total of 276 reviews. Reviews with positive labels obtained contain thanks, positive impressions that users get during the process of making passports, and other compliments, while reviews with negative labels contain insults, insults and negative impressions in the process of making passports usually also as a form of user dissatisfaction, while reviews with neutral labels can be caused by several things including the number of negative words and positive words in 1 sentence the same amount, It can be caused by negative and positive words that are not found in the Lexicon dictionary.

3.4 Data Mining

The process of data mining or data mining at the KDD stage can also be called text mining or text mining, because the data that researchers use is in the form of text, then at this stage, the Naive Bayes classifier algorithm is implemented for sentiment analysis of Mobile Passport application reviews (Karthikeyan et al., 2024). This stage is carried out after the data has passed the sentiment class labeling, this process is carried out by classification analysis, namely by making training data and testing data, then classifying with the Naive Bayes classifier algorithm and visualizing barplots and word clouds of the positive and negative words produced.

3.5 Evaluation

The process of implementing the text mining algorithm Naive Bayes classifier for sentiment analysis of Mobile Passport reviews is carried out by forming training data and testing data. Previously, this process was carried out based on the KDD method with stages of data selection, preprocessing, transformation, data mining, and evaluation. Based on Table 4.26, the average total value of accuracy, precision, and recall in this study was 84.99%, 87.37%, and 91.31%. Accuracy is the sum of the proportions of predictions that are appropriate or correct. Accuracy is also a parameter of the degree of similarity and accuracy between the actual value and the predicted value. Then, Precision is the proportion of the number of relevant and controlled text documents among all documents selected by the system. Precision is used as the level of accuracy between the information requested and the answer issued by the system (Behiry & Aly, 2023). Recall is the proportion of the number of relevant text documents under control among all relevant text documents in a collection or data set. Recall is used as a measure of the success of the system in recovering information (Manning et al., 2009). Thus, the magnitude of the confusion matrix value is more than 80%, so it can be said that the Naive Bayes classifier algorithm is well used in this study.

4. Discussion

The process of data mining or data mining at the KDD stage can also be called text mining or text mining, because the data that researchers use is in the form of text, then at this stage, the Naive Bayes classifier algorithm is implemented for sentiment analysis of Mobile Passport application reviews (Amarudin et al., 2024). This stage is carried out after the data has passed the sentiment class labeling, this process is carried out by classification analysis, namely by making training data and testing data, then classifying with the Naive Bayes classifier algorithm and visualizing barplots and word clouds of the positive and negative words produced. Training data is used to form models in classification, the model will later become a representation of knowledge that will be used to predict new data classes. The new data used will be known as data testing. The amount of training data in this text-mining process affects the accuracy value. As much training data as possible, the model will learn more so that its accuracy will be better, Comparison of the value of training and testing data obtained in the formation of the classification model there are 3 data components, including a comparison of 70% training data and 30% testing data from a total of 1598 reviews as many as 1119 reviews as training data and 479 reviews as testing data, comparison of 80% training data and 20% testing data from a total of 1598 reviews as many as 1278 reviews as training data and 320 reviews as testing data and comparison of 90% training data and 10% testing data from a total of 1598 reviews as many as 1438 reviews as training data and 160 reviews as testing data.

From the three comparisons above, each data set ratio (training data and data testing) was carried out 5 times by making pseudorandom numbers many times using the set.seed(n) formula in the R Studio software used. Pseudorandom numbers are numbers that are used as references in pseudorandomization, pseudo-meaning is that

the number is not a number that is generated randomly, but through a process that has a material or seed (seed). The selected seed will be used as a parameter of the random function. Furthermore, classification is carried out by implementing the Naive Bayes classifier algorithm for positive sentiment and negative sentiment, in the process of text mining, of course using the confusion matrix method in the evaluation process to find out the accuracy value of the algorithm used (Shahpurkar et al., 2024). Confusion matrix is one of the important tools in evaluation methods used in machine learning which usually contains two or more categories. From the confusion matrix table, it can be seen that the value of 306 is a TP (True Positive) value which means that there is a lot of data whose actual class is positive and the prediction class is positive. The value of 31 is an FN (False Negative) value which means that there is a lot of data whose actual class is positive and the prediction class is negative. Then, the value 49 is said to be FP (False Positive) means that there is a lot of data whose actual class is negative and the prediction class is positive. The value 93 called TN (True Negative) means that there is a lot of data whose actual class is negative and the prediction class is negative.

In the dataset with a ratio of 70:30, the accuracy, precision, and recall values for experiment 1 were 83.3%, 86.2%, and 90.8%. As for experiment 2, the confusion matrix values obtained successively for accuracy, precision, and recall values were 84.55%, 85.95%, and 93.13%, then for experiment 3 were 85.8%, 87.3%, and 92.66%, then for experiment 4 each the accusation, precision and recall values obtained were 82.05%, 84.02%, and 90.43%. From the results of research on the 70:30 dataset, the highest accuracy value was occupied by experiment 5 with set. seed(10) with an accuracy value of 86.64%, then the highest precision value of 88.27% was obtained in experiment 5, and the highest recall value in experiment 5 with a recall value of 93.49%. It can be concluded that in the 70:30 dataset, the fifth experiment with a randomization value of 10 (set. seed(10)) has a large confusion matrix value compared to randomization values of 2, 4, 6, and 8. Table 4.26 above shows that with a ratio of 80:20 in experiment 1 the accuracy, precision, and recall values obtained were 83.75%, 87.39%, and 89.73% respectively. Then in the second experiment, 85.94%, 90.87%, and 88.84% were obtained for accuracy, precision, and recall values, respectively. Furthermore, in the third experiment, it was 86.25%, 87.89%, and 92.02% and in the fourth experiment, the confusion matrix values obtained were 83.44%, 86.90%, and 89.64% then in the fifth experiment the accuracy, precision, and recall values obtained were 86.88%, 87.34% and 94.52%. Thus, the highest accuracy value in the 80:20 dataset was obtained in experiment 5 with a value of 86.88%, while the highest precision value was found in experiment 2 with a value of 90.87%, and the highest recall value in experiment 5 with a value of 94.52%.

The results of the study in the 90:10 data set presented in Table 4.26 above showed the results of accuracy, precision, and recall in experiment 1 were 79.38%, 81.82%, and 87.38%. Then in experiment 2 of 84.38%, 88.18%, and 88.99%, the accuracy value in experiment 3 was 87.5%, precision was 90% and recall was 91.67%. As for experiment 4 88.12%, 89.74%, and 93.75% for the accuracy value of precision and recall. In experiment 5 obtained confusion matrix values of 86.88%, 88.5%, and 92.59%. So it can be concluded in the dataset 90:10 the highest accuracy value was obtained in experiment 4 at 88.12%, the highest precision value in experiment 3 with a value of 90%, and the highest recall value of 93.75% in experiment 4 with the set.seed(8) / randomization 8.

The results of the three data set experiments above found that the average accuracy value of precision and recall in the 70:30 data set was 84.47%, 86.38%, and 92.10%. As for the 80:20 dataset, the average values of accuracy, precision, and recall were 85.25%, 88.08%, and 90.56%. The average value of accuracy, precision, and recall in the 90:10 dataset was 85.25%, 87.65%, and 90.88%. It can be concluded that the highest accuracy value is with an average of 85.25% on the 80:20 dataset and 90:10 dataset, then the highest precision value is with a value of 88.08% on the 80:20 dataset, and the highest recall value is on the 70:30 dataset with a value of 92.10%. Thus, the 80:20 data set is the best dataset for classifying existing reviews because it has the highest scores for both confusion matrix values (accuracy and precision). These results are under the Pareto Principle, namely 80: 20, and states that the 80: 20 concept can be applied in all aspects of human life ranging from socio-cultural, socio-economic, socio-political, and others. The mathematical spirit of Pareto's law is that 80% of reactions result from 20% of actions performed. but it does not rule out the possibility that in a study not only using this ratio.
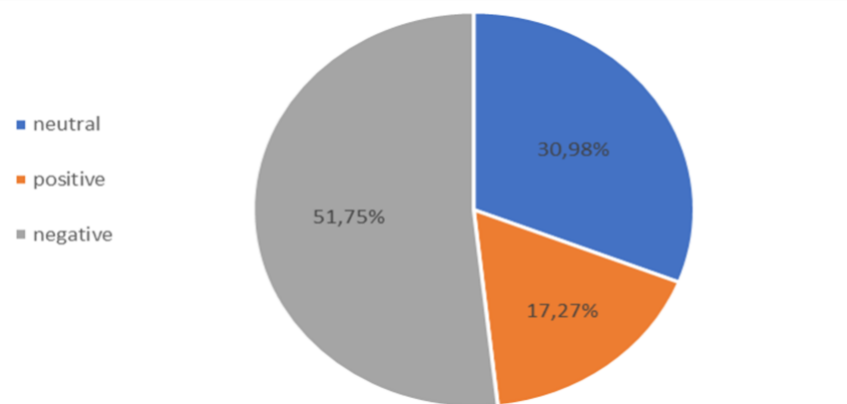
Figure 4: Results of Sentiment Analysis using Knowledge Discovery in Database (KDD) Design Model

The Mobile Passport application is an application issued officially by the Directorate General of Immigration, this application can be downloaded on Google Play and Apple Apps, since it was inaugurated on January 26, 2022, this application has been downloaded by more than 1,000,000 users. The rating of this application only reaches 2.7 out of 5 stars on the Google Play platform and a rating of 2.1 out of 5 stars on the Apple Apps platform. This value shows that the services provided by the Directorate General of Immigration are not perfect and have not met user expectations. The results of the sentiment analysis showed that of the 1,598 reviews taken since the application was updated in version 5.03 and version 5.2.1, 495 of them obtained a sentiment score smaller than zero (<0), and 276 of them obtained a positive sentiment score or greater than zero (>0) and 827 other reviews or equivalent to 52% of reviews had a sentiment score equal to zero or in this case categorized as neutral. These results show that service errors in passport-making services are still quite high, reaching 31% negative reviews.

Based on the facts that have been described, it can be seen that the results of sentiment analysis are relevant and have interpreted the rating obtained in the Mobile Passport application. The results of sentiment analysis and rating of the Mobile Passport application illustrate that the error in the application is quite high. Therefore, it is necessary to evaluate and improve the performance of the Mobile Passport application as one of the digital services to the public issued by the Directorate General of Immigration of Indonesia.

4. Conclusion

The process of implementing the Naive Bayes classifier text mining algorithm for sentiment analysis of Mobile Passport reviews is carried out by forming training data and testing data. Previously, this process was carried out based on the KDD method with stages of data selection, preprocessing, transformation, data mining and evaluation. From this research, it can be concluded that the implementation of the text mining algorithm using Naive Bayes to analyze the sentiment of reviews of the Mobile Passport application on the Google Play Store has succeeded in providing very satisfying results. With consistent accuracy, precision and recall values above 80%, this algorithm has proven to be effective in processing and analyzing user reviews well. These results have important implications in improving the quality of mobile passport applications, because they allow developers to be more efficient in understanding and responding to user reviews, thus potentially making a significant contribution to the Indonesian Directorate General of Immigration in making improvements to mobile passport applications.

References

Alazab, M., Abu Khurma, R., Castillo, P. A., Abu-Salih, B., Martín, A., & Camacho, D. (2024). An effective networks intrusion detection approach based on hybrid Harris Hawks and multi-layer perceptron. *Egyptian Informatics Journal*, *25*(November 2023), 100423. https://doi.org/10.1016/j.eij.2023.100423

Amarudin, Ferdiana, R., & Widyawan. (2024). B-DT Model: A Derivative Ensemble Method to Improve Performance of Intrusion Detection System. *Journal of Advances in Information Technology*, *15*(1), 87–103. https://doi.org/10.12720/jait.15.1.87-103

APJII, A. P. J. I. I. (n.d.). *APJII in Indonesia Digital Outlook*. https://apjii.or.id/berita/d/apjii-di-indonesia-digital-outloook-2022_857

Ashar, A. R., & Maharani, C. (2022). Far from Home: Profile of pro-IS Deportees' Mobility throughout 2016-2020 and Prevention Strategy of Indonesian Government. *Journal of ASEAN Studies*, *10*(1), 81–105. https://doi.org/10.21512/jas.v10i1.7732

Behiry, M. H., & Aly, M. (2023). Cyberattack Detection in Wireless Sensor Networks using a Hybrid Feature Reduction Technique with AI and Machine Learning Methods. *Journal of Big Data*. https://doi.org/10.1186/s40537-023-00870-w

Chembakottu, B., Li, H., & Khomh, F. (2023). A large-scale exploratory study of android sports apps in the google play store. *Information and Software Technology*, *164*(Heng Li). https://doi.org/10.1016/j.infsof.2023.107321

Farooqi, A. H., Akhtar, S., Rahman, H., Sadiq, T., & Abbass, W. (2023). Enhancing Network Intrusion Detection Using an Ensemble Voting Classifier for Internet of Things. *Sensors*, *24*(1), 127. https://doi.org/10.3390/s24010127

Grotenhuis, Z., Mosteiro, P. J., & Leeuwenberg, A. M. (2024). Modest performance of text mining to extract health outcomes may be almost sufficient for high-quality prognostic model development. *Computers in Biology and Medicine*, *170*(January), 108014. https://doi.org/10.1016/j.compbiomed.2024.108014

Hamzah, A. S. S. S. A., Saipan saipol, H. F., Nordin, S. Z., Shaffiei, Z. A., & Ohshima, N. (2024). Adopting Text Mining for Patent Analysis to Determine the Attribute and Segment in Automotive Industries. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *37*(2), 94–103. https://doi.org/10.37934/araset.37.2.94103

Han, H., Asif, M., Awwad, E. M., Sarhan, N., Ghadi, Y. Y., & Xu, B. (2024). Innovative deep learning techniques for monitoring aggressive behavior in social media posts. *Journal of Cloud Computing*, *13*(1). https://doi.org/10.1186/s13677-023-00577-6

Huo, H., Wang, C., Han, C., Yang, M., & Shang, W. L. (2024). Risk disclosure and entrepreneurial resource acquisition in crowdfunding digital platforms: Evidence from digital technology ventures. *Information Processing and Management*, *61*(3), 103655. https://doi.org/10.1016/j.ipm.2024.103655

Kantar, O., & Kilimci, Z. H. (2023). Deep learning based hybrid gold index (XAU/USD) direction forecast model. *Journal of the Faculty of Engineering and Architecture of Gazi University*, *38*(2), 1117–1128. https://doi.org/10.17341/gazimmfd.888456

Karthikeyan, M., Manimegalai, D., & RajaGopal, K. (2024). Firefly algorithm based WSN-IoT security enhancement with machine learning for intrusion detection. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-023-50554-x

Lee, M., Lin, X., Chai, J. Z. Q., Lee, E. S., Smith, H., & Tudor Car, L. (2023). Smartphone apps for point-of-care information summaries: Systematic assessment of the quality and content. *BMJ Evidence-Based Medicine*, *28*(5), 320–327. https://doi.org/10.1136/bmjebm-2022-112146

Liu, Y., Alias, A. H. bin, Haron, N. A., Bakar, N. A., & Wang, H. (2024). Technology status tracing and trends in construction robotics: A patent analysis. *World Patent Information*, *76*(January), 102259. https://doi.org/10.1016/j.wpi.2023.102259

McConkey, R., & Olukoya, O. (2023). Runtime and Design Time Completeness Checking of Dangerous Android App Permissions Against GDPR. *IEEE Access*, *12*(January), 1–1. https://doi.org/10.1109/access.2023.3347194

Nilashi, M., Abumalloh, R. A., Ahmadi, H., Samad, S., Alrizq, M., Abosaq, H., & Alghamdi, A. (2023). The nexus between quality of customer relationship management systems and customers' satisfaction: Evidence from online customers' reviews. *Heliyon*, *9*(11), e21828. https://doi.org/10.1016/j.heliyon.2023.e21828

Nokkaew, M., Nongpong, K., Yeophantong, T., Ploykitikoon, P., Arjharn, W., Siritaratiwat, A., Narkglom, S., Wongsinlatam, W., Remsungnen, T., Namvong, A., & Surawanitkun, C. (2024). Analyzing online public opinion on Thailand-China high-speed train and Laos-China railway mega-projects using advanced machine learning for sentiment analysis. *Social Network Analysis and Mining*, *14*(1), 1–15. https://doi.org/10.1007/s13278-023-01168-8

Perea-Khalifi, D., Irimia-Diéguez, A. I., & Palos-Sánchez, P. (2024). Exploring the determinants of the user experience in P2P payment systems in Spain: a text mining approach. *Financial Innovation*, *10*(1). https://doi.org/10.1186/s40854-023-00496-0

Qureshi, A. A., Ahmad, M., Ullah, S., Yasir, M. N., Rustam, F., & Ashraf, I. (2023). Performance evaluation of machine learning models on large dataset of android applications reviews. *Multimedia Tools and Applications*, *82*(24), 37197–37219. https://doi.org/10.1007/s11042-023-14713-6

Sadare, O., Melvin, T., Harvey, H., Vollebregt, E., & Gilbert, S. (2023). Can Apple and Google continue as health app gatekeepers as well as distributors and developers? *Npj Digital Medicine*, *6*(1). https://doi.org/10.1038/s41746-023-00754-6

Shahpurkar, A., Patil, R., & Tangod, K. K. (2024). Class imbalance aware drift identification model for detecting diverse attack in streaming environment. *Indonesian Journal of Electrical Engineering and Computer Science*, *33*(2), 981–989. https://doi.org/10.11591/ijeecs.v33.i2.pp981-989

Sinclair, R., Nodi, S., & Kabir, M. A. (2024). Evaluating mobile applications for estimating soil properties: Quality of current apps, limitations and future directions. *Computers and Electronics in Agriculture*, *216*(December 2023), 108527. https://doi.org/10.1016/j.compag.2023.108527

Thiengburanathum, P., & Charoenkwan, P. (2023). SETAR: Stacking Ensemble Learning for Thai Sentiment Analysis Using RoBERTa and Hybrid Feature Representation. *IEEE Access*, *11*(August), 92822–92837. https://doi.org/10.1109/ACCESS.2023.3308951

Xi, X., An, X., Zhang, G., & Liang, S. (2022). Spatial patterns, causes and characteristics of the cultural landscape of the Road of Tang Poetry based on text mining: take the Road of Tang Poetry in Eastern Zhejiang as an example. *Heritage Science*, *10*(1), 1–28. https://doi.org/10.1186/s40494-022-00761-y

Yu, L., Xu, L., & Jiang, X. (2023). A High-Performance Multimodal Deep Learning Model for Detecting Minority Class Sample Attacks. *Symmetry*, *16*(1), 42. https://doi.org/10.3390/sym16010042