

## Sentiment Analysis of M-Paspor App Reviews Using Multinomial Naive Bayes

Muhammad Luthfi Fajri Martanto, Wirawan Istiono

Universitas Multimedia Nusantara

*wirawan.istiono@umn.ac.id*

**Abstract.** This study conducts sentiment analysis on user reviews of the M-Paspor application, an Indonesian mobile application for passport services, available on the Google Play Store in 2023. By leveraging the Multinomial Naive Bayes algorithm, a machine learning technique suitable for text classification tasks, the research aims to evaluate the sentiment polarity (positive, negative, or neutral) of user reviews. The analysis is based on a dataset of 6,443 reviews scraped from the Google Play Store, spanning the period from January 1, 2023, to October 18, 2023. The study employs a comprehensive methodology, including data preprocessing techniques such as case folding, tokenization, and normalization, as well as feature extraction using TF-IDF (Term Frequency-Inverse Document Frequency). The performance of the Multinomial Naive Bayes model is rigorously evaluated using various metrics, including accuracy, F1-score, precision, and recall, calculated through a confusion matrix. The impact of different train-test data split ratios (60:40, 70:30, and 80:20) on model performance is also investigated. The results indicate that the 80:20 data split yields the highest performance, with an accuracy of 66.45%, an F1-score of 64.79%, a precision of 72.24%, and a recall of 64.12%. This study contributes to the understanding of user sentiments and experiences with the M-Paspor application, providing valuable insights for enhancing user satisfaction and improving the quality of digital government services.

**Keywords:** Confusion Matrix, M-paspor, Multinomial Naive Bayes, Sentiment Analysis, Textblob

## 1. Introduction

A passport is an ID card or original identity card that is used when traveling overseas for the purpose of doing business, studying science, or simply visiting the target country (King et al., 2023). According to the Immigration.go.id website, the Directorate-General of Immigration successfully issued as many as 1,595.096 passport books in the first quarter of 2023, with an average of 13,292 successful passports issued every day. When compared to the average passport issuance prior to the epidemic in 2019, the figure has risen to 38%. With such a significant increase now, the Immigration Directorate-General is working to improve the passport-making process. The Immigration Directorate-General's activities include the improvement of the M-paspor application (Imigrasi, 2023; Indra, 2022; Zaenuddin, 2023). Sentiment analysis is a technique that is carried out automatically to comprehend, extract, or process text data in order to detect information about sentiments included in a sentence (Basarlan & Kayaalp, 2020). In general, sentiment analysis is used to examine product reviews, social media posts, and app reviews. Understanding the sentiment included in a user's review allows application developers to respond to such reviews and improve the product's quality (Kabir et al., 2021; OSMANOĞLU et al., 2020). Based on the text or sentence, sentimental analysis is split into three categories: positive, negative, and neutral (Elbagir & Yang, 2018; Singh et al., 2017). The Multinomial Naive Bayes algorithm is one method used in sentiment analysis.

Similar research was carried out by Nur Lickha Lavenia and Reisa Permatasari in Twitter sentiment analysis about depressive disorders using the Naive Bayes classifier algorithm. From the results of this research it was found that the Multinomial Naive Bayes algorithm got an accuracy of 90.13%, the Bernoulli Naive Bayes algorithm got an accuracy of 90.13%. 85.36%, the Gaussian Naive Bayes algorithm gets an accuracy of 88.37% (Lavenia & Permatasari, 2023). Research using the multinomial naive Bayes method was carried out by Sartika Mandasari and his friends to carry out sentiment analysis on grab services. From the results of that research, the Multinomial Naive Bayes method produced an accuracy of 86.57% and in that research it was stated that the multinomial naive Bayes method had several advantages. among other things, it is simple, fast and has good accuracy in processing data in text form (Mandasari et al., 2022). After reviewing multiple studies, it can be inferred that the multinomial naïve Bayes method is more effective in handling textual data. Therefore, this research has selected the Multinomial Naive Bayes method for processing text data in order to conduct a review of the Google Play Store, beside that Multinomial Naive Bayes method was chosen in this research because Multinomial Naive Bayes is a frequency-based machine learning model in text analysis represented by a series of words in data or document (Hamzah, 2021; Resti et al., 2022). The naive bayes approach has various advantages, including being simple to apply on both continuous and discrete data, being able to handle big data sets, being able to classify or group data with multiple labels (Permana & Putra, 2023), and being best suited for training natural language processing models (Damanik & Setyohadi, 2021; Pham et al., 2020). In this research, a sentiment analysis program was created to understand customer feelings towards products or services, improve customer service, design effective marketing strategies, and make data-based business decisions. This allows companies to monitor reputation, respond quickly to complaints, customize services, and identify market trends and predictions. In addition, sentiment analysis helps in product innovation, risk management and early crisis detection, so that companies can act preventatively and make more appropriate investment decisions (Wankhade et al., 2022).

Based on the study problems, the multinomial naive bayes method was used to analyze the sentiment of M-PASPOR app reviews on the Google Play store in 2023. This meant that the multinomial naive bayes method was used to measure the sentiment of M-PASPOR app reviews on the Google Play store. A method called sentiment analysis is used to automatically understand, extract, or process text data in order to find out about the feelings that are expressed in a sentence. (Pratmanto et al., 2020; Santoso et al., 2024). Language mining includes sentiment analysis study, which is computer research based on sentiments, emotions, opinions, reviews, comments, and any expressions communicated

through language (Putri et al., 2020), Multinomial Naive Bayes algorithm is used to analysing the M-PASPOR app review sentiment on the Google Play store using the confusion matrix and the data used is a review throughout the year 2023. As for the related research on the Analysis of User Satisfaction of the M-Passport Application Service Using the Technology Acceptance Model (TAM) at TPI Class II Singkawang Immigration Office by Pardi, the study was more focused on evaluating user satisfaction with regard to the immigration service (Pardi, 2024), compared to the previous research in this study focused to analyze the sentiment towards user experience in using M-Paspor application and data used throughout the year 2023. The objective of this study is to apply the Multinomial Naive Bayes algorithm for sentiment analysis of reviews for the M-Passport application on the Google Play Store. Additionally, it aims to evaluate the algorithm's performance by using a confusion matrix to analyse the sentiment of reviews for the M-Passport application in 2023 on the Google Play Store.

## **2. Literature Study**

### **2.1. Sentiment Analysis**

Sentiment analysis is a technique that is performed mechanically to comprehend, extract, or process text data to determine information about sentiments included in a sentence (AlBadani et al., 2022; Singh et al., 2017). language mining includes sentiment analysis study, which is computer research based on sentiments, emotions, opinions, reviews, comments, and any expressions communicated through language, sentiment analysis receives numerically shaped data in which the sentence data must be weighted by extracting data with Term Frequency - Inverse Document Frequency (TF-IDF) (Mehta & Pandya, 2020; Zahoor et al., 2020).

### **2.2. Multinomial Naïve Bayes**

Multinomial Naive Bayes is a variation of the Naive bayes algorithm. Multinomial naive bayes are a frequency-based text classification model represented by a set of words that appear in data or document (Olanrewaju et al., 2022). The multinomial naive bayes have some advantages, namely, they are easy to use on data that is continuous and discreet, can handle large data sets, can be used to classify or group data with multiple labels and are best used to train natural language processing models (Hossain et al., 2021; Nurhidayah, 2020). Multinomial Naïve Bayes (MNB) is a widely used algorithm in the field of machine learning, known for its effectiveness in tasks such as text categorization and natural language processing. The following are the Below are the advantages, disadvantages and limitations of MNB in comparison to other methodologies:

a. Comparison with Logistic Regression:

Logistic Regression is a widely used method for solving binary and multiclass classification issues. Similar to MNB, it offers probabilistic results and is reasonably simple to implement and understand. Unlike Multinomial Naive Bayes (MNB), logistic regression does not depend on the strong assumption of independence. This enables logistic regression to successfully simulate interactions between features. Logistic regression tends to be more accurate when the assumption of independence in Multinomial Naive Bayes (MNB) is not met. However, logistic regression may experience slower training times when dealing with extremely big datasets and may necessitate the use of regularization techniques to avoid overfitting, hence increasing its complexity.

b. Comparison with Support Vector Machines (SVM):

Support Vector Machines (SVM) provide a strong alternative, especially in situations with high-dimensional spaces and intricate relationships between features and target variables. Support Vector Machines (SVMs) employ kernel functions to effectively handle non-linear interactions, resulting in a high degree of flexibility and often superior accuracy compared to Multinomial Naive Bayes (MNB) in these scenarios. Nevertheless, Support Vector Machines

(SVMs) generally require more processing resources and are less easily interpretable compared to Multinomial Naive Bayes (MNB). The process of training a Support Vector Machine (SVM) can be considerably time-consuming, particularly when dealing with huge datasets. Additionally, the resulting models are generally perceived as opaque, offering limited understanding of the decision-making process.

c. Comparison with Decision Trees and Random Forests:

Decision Trees and Random Forests possess distinct advantages and disadvantages. Decision trees are easily understood and offer explicit decision rules, making them highly interpretable. They have the ability to capture complex linkages and interactions between features without presuming that they are independent of each other. Nevertheless, they have a tendency to overfit, particularly when dealing with tiny datasets. Random forests address this problem by taking the average of numerous decision trees, which improves the reliability and ability to apply the model to new data. Although these methods offer flexibility and effectiveness, they can become computationally demanding and sacrifice interpretability as the ensemble's complexity grows.

d. Comparison with Neural Networks:

Neural networks, particularly deep learning models, provide the highest level of flexibility compared to the other methods outlined. Their multi-layered structure allows them to accurately represent intricate and non-linear relationships. These characteristics render them well-suited for applications such as picture and speech recognition, in which alternative approaches may encounter difficulties. Nevertheless, neural networks necessitate substantial quantities of data and computer resources for training, and their opaque nature renders them less interpretable compared to MNB. Moreover, they frequently require meticulous adjustment of multiple hyper parameters, a task that can be both time-consuming and technically demanding.

To summarize, Multinomial Naïve Bayes performs exceptionally well in text classification problems, especially when the condition of independence is reasonably satisfied. It stands out for its simplicity, speed, and overall performance. Logistic regression provides a balanced approach by making fewer assumptions and allowing for clear interpretation. However, it may become more complex when dealing with regularization. Support Vector Machines (SVMs) offer robust non-linear modeling capabilities, however this comes at the cost of computational efficiency and interpretability. Decision trees and random forests provide versatility and resilience, but they can be susceptible to overfitting and heightened intricacy. Neural networks excel at processing intricate data kinds and interactions, but their management necessitates substantial resources and skill. The selection of these methods ultimately relies on the unique criteria of the problem, such as the characteristics of the data, the significance of interpretability, and the computational resources that are accessible.

### **2.3. Confusion Matrix**

The confusion matrix is a performance measurement tool for classifying machine learning models, which allows classifications to be multiple classes. In a confusion matrix, there are four combinations of actual or actual values and predictive values that are the result of classification using algorithms in machine learning (Chicco et al., 2021; Xu et al., 2020). The Confusion Matrix uses four terms used to represent the results of the classification process: true positive(tp), true negative(tn), false positive(fp) and false negative(fn) as shown as on figure 1 (Heydarian et al., 2022).

		True Class	
		Positif	Negatif
Predict Class	Positif	TP	FP
	Negatif	FN	TN

Fig. 1: Confusion Matrix

### 3. Methodology

The research methodology that used in this study is the Sentiment Analysis method based on Multinomial Naive Bayes Algorithm. The steps taken in this research are in six steps, namely, literature study, data scrapping, system design, algorithm implementation and development, testing, evaluation and documentation.

The study of literature is an early stage in research. This phase aims to study and understand information from the subject being studied. The information collected comes from journals, articles, writings, books and websites. Relevant theories in this study include; Sentiment Analysis, Multinomial Naive Bayes, TF-IDF, M-Paspor and Confusion Matrix.

Then in the next stage, data is collected using the data scraping method via the google-play-scraper library. In this study, the data to be taken is a review of the M-Paspor application throughout the year 2023 starting from January 1, 2023 until when the research was carried out on October 18, 2023 and obtained as much as 6443 data review of M-Paspor application.

The next step is system design phase begins with the creation of a research stream or flowchart, perform labelling with text blob, perform text pre-processing process on the data, perform TF-IDF extraction features and perform the division of training data and test data. After that, the next step is algorithm implementation, where at this stage the multinomial naive Bayes algorithm, TF-IDF and others algorithm were applied in this research.

The next procedure is testing phase is a phase that has the function to perform tests on code that has been built to find out if there are errors and bugs of the system that was designed to be built and will also perform the performance testing of modeling that had been applied to the system. It's done to make sure that the created system can run properly. In the (Braiek & Khomh, 2020) are divided into 60:40, 70:30 and 80:20. To calculate the accuracy, precision, recall and f1-score levels of the confusion matrix by comparing classification labels prior to modeling and classifying labels after modeling.

on, most of the data (60%, 70%, or 80%) is allocated for model training, while the rest (40%, 30%, or 20%) is used to test the performance of the trained model. This approach allows the model to learn from known data (training data) and be tested on data it has never seen before (test data), thus providing a good estimate of how the model will behave on new data. It is common practice in machine learning model evaluation to ensure that the model has a good ability to generalize patterns from the data (Xie et al., 2011).

The final step in this methodology is documentation of all processes from the beginning of the research, design, and modeling processes to the modeling of classifications. It's done gradually from introduction to conclusion and suggestion.

Data Scraping is the process of collecting and storing data from web pages automatically. In the data scraping process, use a tool or program to retrieve information contained in the web and store it in a used format such as csv. Although scraping data has technical challenges such as rate limiting, Captcha, and bot detection, which can slow down or halt the data collection process. The data obtained

may be inconsistent or noisy, requiring significant time for cleaning and processing. The scraping process also demands ongoing maintenance and complex debugging, along with scalability and performance issues when collecting data on a large scale but data scrapping has been chosen because it can gather large volumes of data quickly and efficiently from various online sources, enabling organizations to access valuable information for analysis and decision-making. In this study the data collection was done by scraping data from the review of the M-Paspor application in the google play store using the google-play-scraper library (Amir Latif et al., 2019). Data taken throughout the year 2023 from January 1, 2023 until the time of the study on October 18, 2023 and obtained as much as 6443 data review of the M-Paspor application.

Text-preprocessing is a useful step for selecting text to be structured. Figure 2 is a text-preprocessing phase starting with Case Folding, Data Cleaning, Tokenizing, Stemming, Normalization, Stopwords and Stemming. Where at the Case folding stage, the text is changed to lowercase, Meanwhile, in the Data Cleaning stage, data is selected and cleaned from symbols, emoji, links, single characters, numbers, white space and punctuation that do not have sentiment in a text. Then in the Tokenizing stage, the sentence is splitting sentence into a list of words using the nltk library (Yao, 2019). In the normalization stage, a CSV file is created containing the wrong words in the writing along with the corrections. At the Stopwords stage, words are deleted that do not have a special meaning, for example conjunctions. At the Stemming stage, words that have affixes are changed or extracted to become the basis of the word using a sastrawi library (Rosid et al., 2020).

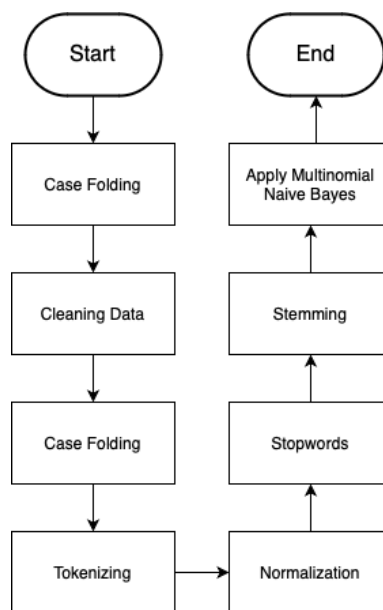


Fig. 2: Text Pre-processing Flowchart

After the text pre-processing is done then the data will be labelled, the data labelling is a useful stage as the giving of a sentiment classification label on the text data obtained through the data scraping process and has passed the stage of text-pre processing. In this study the labelling stage will be done automatically using the Library Textblob (Lorla dan Steven, 2020) which will be described in figure 3.

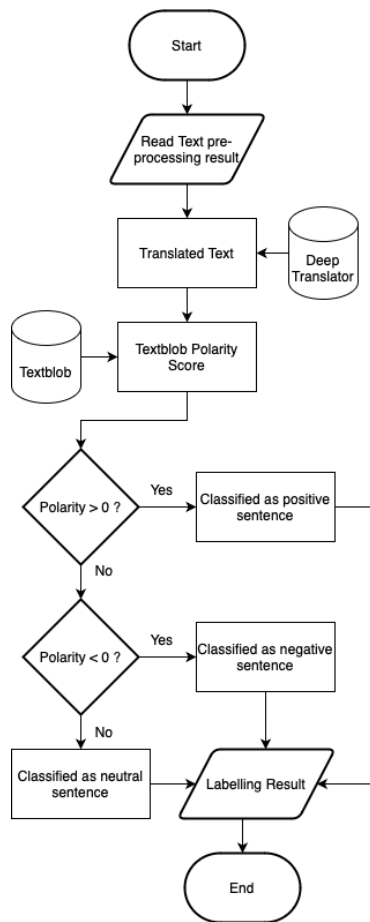


Fig. 3: Text Blob Flowchart

Feature Extraction TF-IDF is a phase that has the function to give weight values to the text data contained in the sentence. This weighing is done with the aim of enabling the multinomial naive bayes algorithm to understand the text input. TF-IDF was selected due to its simplicity and ease of implementation, allowing for quick analysis without the need for pre-trained models or substantial computational resources (Chen et al., 2016). TF-IDF is effective for keyword extraction and provides easily interpretable results, as the scores directly reflect the importance of words within the document. Additionally, TF-IDF scales well for large datasets and automatically reduces the influence of common words that frequently appear across many documents (Zhou, 2022). Although it does not account for word order or semantic relationships, TF-IDF remains a good baseline method and is often used as an initial step before applying more complex techniques (Kamath et al., 2018).

After the Feature Extraction is done, the next step is to divide the data using the train-test split data. At the stage, the data will be divided into two parts, namely, data train and data test. The data train is used as data to train the modeling made using the multinomial naive bayes algorithm. This, done to evaluate and know performance in modeling.

After the data has been divided into two steps, the next step is to implement the multinomial naive bayes algorithm and classify the data. Starting with the creation of a modeling of the classification of multinomial Naive Bayes then performing modeling with the data train that has been performed feature extraction and the results of the modelling that have been made using the multinomial naive bayes algorithm will be used to predict the test data as shown as figure 4.

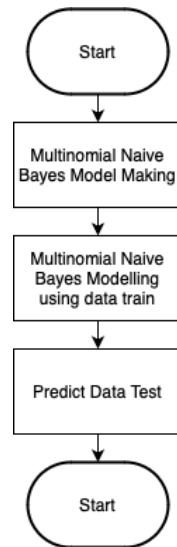


Fig. 4: Apply Multinomial Naive Bayes Flowchart

The evaluation phase will measure the performance of the modeling that has been successfully made using the Multinomial Naive Bayes algorithm. In this study, the evaluation is done using a confusion matrix with three labels comparisons consisting of positive, negative and neutral labels.

#### 4. Result and Discussion

Figure 5 shows the implementation of data scrapping, the data collected in the research was done automatically using the data scrapping method. In the process of scrapping data application M-Paspor on google play store using a tool such as library google-play-scraper and pandas library (Stepanek, 2020), NumPy (Harris et al., 2020) to store data in the form of csv and found 6,443 review data.

	userName	score	at	content
0	Dena Aditya Setiawan	1	2023-10-17 10:31:49	email tidak valid terusssss
1	Putra Antara	1	2023-10-17 07:14:26	Ribet
2	Me	1	2023-10-17 06:52:42	Bad aps
3	Ahmad Alamudi	5	2023-10-17 06:06:07	Sangat membantu dan ngga lemot
4	Hwang Joe	1	2023-10-16 20:28:35	Apk eek Ngestak di yang ke 5 Udh isi semuanya ...
...	...	...	...	...
6438	Firman Ahmad	1	2023-01-02 07:53:48	astaghfirullah.. begini bener ya ASN nyari dui...
6439	ANITA SARLIM	1	2023-01-01 16:35:55	kata sandi email sy dianggap salah, padahal sd...
6440	Stefanus Deo Alfian	1	2023-01-01 12:20:34	Tidak mungkin tim IT pemerintah gajinya kecil....
6441	Indah Nuraeni	5	2023-01-01 11:28:39	Alhamdulillah dimudahkan. Lancar2 saja sih say...
6442	Shinta deni Yuliasih	5	2023-01-01 10:58:25	Sangat membantu

6443 rows x 4 columns

Fig. 5: Scrapping Data Result

After successfully carrying out the text-preprocessing step, the next step is to automatically label the text data using the textblob library. However, before the labeling process, the data will be translated from Indonesian to English because the textblob library is a library that contains an English data dictionary. To translate the data, the deep translator library will be used (MIT, 2020).

After the language translation is carried out, the next step is to calculate the polarity using the textblob library. The initial step starts from importing the textblob library then calculating the polarity of the text data which has been translated from Indonesian to English. If the result of polarity > 0 is

positive, polarity < 0 is negative and score = 0 is neutral. Table 1 shows examples of results from polarity calculations using textblobs and labeling them positive, negative or neutral.

Table 1: Example of textblob labeling results

Text	TextEng	Polarity	Label
Email valid terus	Email continues to be valid	0.000000	Neutral
Ribet	Complicated	-0.500000	Negative
Bad apps	Bad apps	-0.700000	Negative

Figure 6 is a pie chart diagram of the total result of labelling using textblob. It can be seen on the image of 6392 data of 39.50% or 2525 data with negative labels, 28.32% or 1810 data with positive labels and 32.18% or 2057 data with neutral labels.

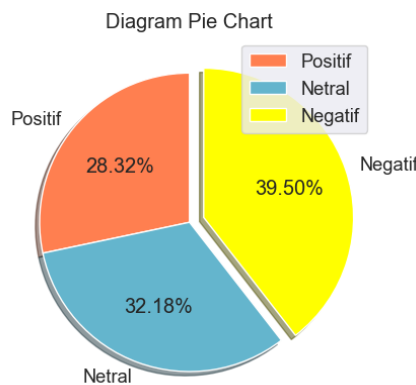


Fig. 6: Text Blob Pie Chart Diagram

Once the data label is generated, the visible word cloud is a visualization of the origin of the word from the classification text. Larger font sizes for words with high frequency appearances in text and smaller fonts for word with small frequencies in text. Figure 7 is the world cloud of every word in the text that is in the positive class group. See the occurrence of words that have the highest frequency or occurrence on positive labels such as “bayar”, “mudah” and “tolong”.



Fig. 7: Positive Word Cloud

Figure 8 is the world cloud of every word in the text that is in the negative class group. On the picture of the title, see the occurrence of words that have the highest frequency or occurrence on negative labels such as error, “susah”, and “gagal”.



This is done to measure the performance behavior of multinomial naive bayes modeling and how much the comparison of train data and test is influenced in the performance of the modeling.

In this test, 6398 data will be divided into 60% training data and 40% test data, 70% training data 30% test data and 80% training data 20% test data. Figure 10a is a confusion matrix from testing results using a comparison of 60% training data and 40% test data. It can be seen in Figure 10b that the results of the modeling performance of the confusion matrix are; Accuracy 64.29%, F1-Score 62.11%, Precision 72.12% and Recall 61.43%.

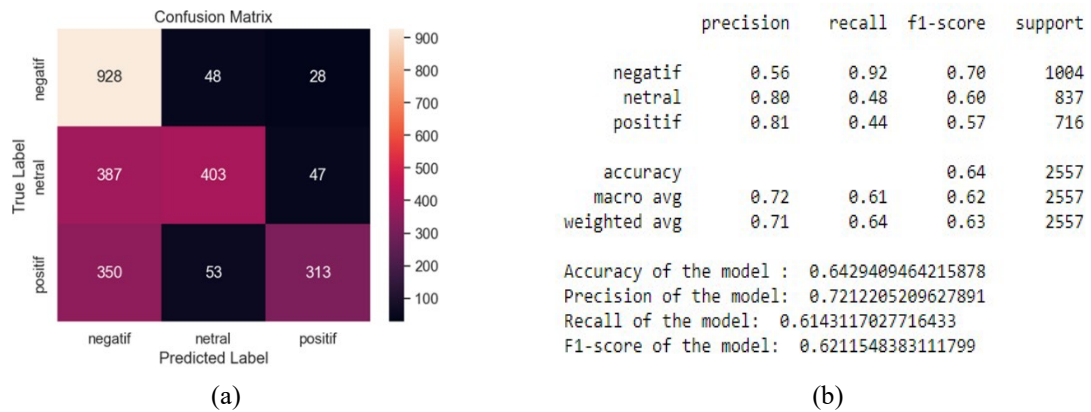


Figure 10. Confusion Matrix 60:40 and Classification Report 60:40

Figure 11a is a confusion matrix from test results using a comparison of 70% training data and 30% test data. Figure 11b shows the modeling performance results of the confusion matrix, namely; Accuracy 64.70%, F1-Score 62.84%, Precision 71.77% and Recall 62.23%.

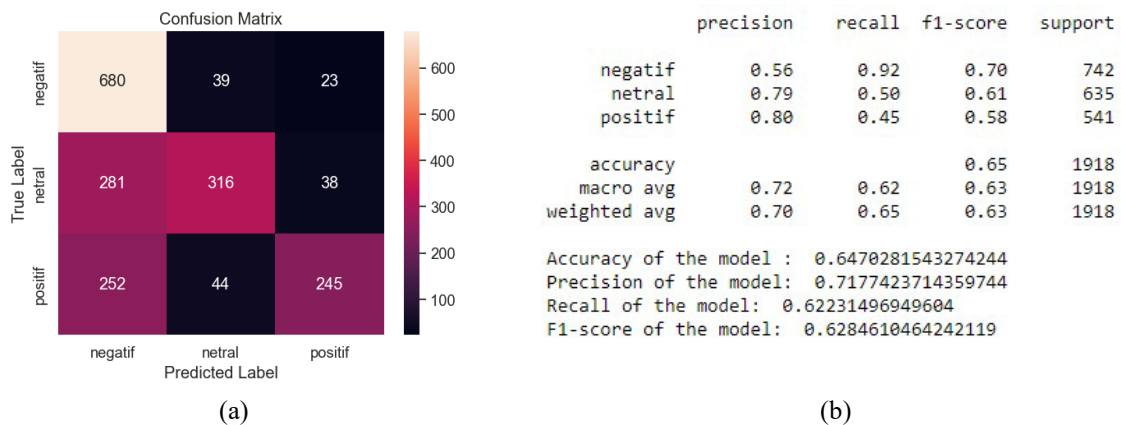


Figure 11. Confusion Matrix 70:30 and Classification Report 70:30

Figure 12a is a confusion matrix from test results using a comparison of 70% training data and 30% test data. Figure 12b shows the modeling performance results of the confusion matrix, namely; Accuracy 66.45%, F1-Score 64.79%, Precision 72.24% and Recall 64.12%.

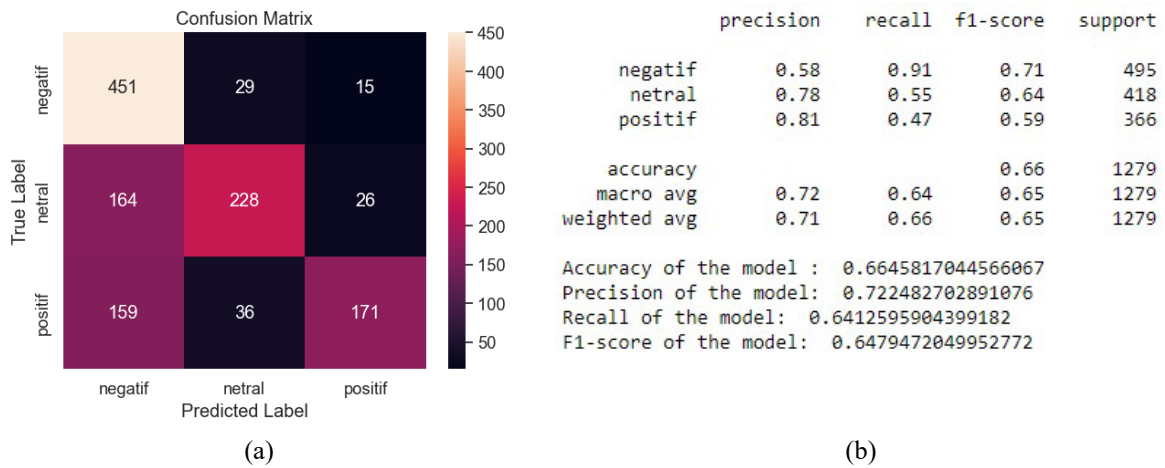


Figure 12. Confusion Matrix 80:20 and Classification Report 80:20

As shown as on Table 2 is 60:40, 70:30 and 80:20 data set, the performance of the modeling performance of confusion matrices namely; accuracy, F1-Score, precision and recall.

Table 3: Data Set Comparison

Matrix %	Dataset 60 : 40	Dataset 70 : 30	Dataset 80 : 20
Accuracy	64.29%	64.70%	66.45%
F1 – Score	62.11%	62.84%	64.79%
Precision	72.12%	71.77%	72.24%
Recall	61.43%	62.23%	64.12%

After testing with dataset comparisons as shown in Table 3, the results showed that the 80:20 dataset division had better performance than the other comparisons with results of Accuracy 66.45%, F1-Score 64.79%, Precision 72.24% and Recall 64.12%. Comparison of the 80:20 dataset gets better results than 70:30 and 60:40, this is because more training data is studied by modeling than the others, resulting in better performance.

## 5. Conclusion

This study successfully implemented sentiment analysis on user reviews of the M-Paspor application, an Indonesian mobile application for passport services, using the Multinomial Naive Bayes algorithm. By analyzing a dataset of 6,443 reviews scraped from the Google Play Store in 2023, the research evaluated the sentiment polarity (positive, negative, or neutral) expressed by users towards the application.

The study followed a rigorous methodology, including data scraping, text preprocessing steps (case folding, tokenization, normalization), feature extraction using TF-IDF, and the implementation of the Multinomial Naive Bayes algorithm. The performance of the sentiment analysis model was assessed through various metrics, such as accuracy, F1-score, precision, and recall, calculated using a confusion matrix.

The findings revealed that the 80:20 data split ratio, allocating 80% of the data for training and 20% for testing, yielded the highest performance among the evaluated scenarios. Specifically, this configuration achieved an accuracy of 66.45%, an F1-score of 64.79%, a precision of 72.24%, and a recall of 64.12%.

From the results of this sentiment analysis research, it is hoped that it will be useful for M-pasport application developers so that it can become input in developing or improving the application to make it better in the future based on the sentiment classification of reviews from users. And from this research

it is hoped that the performance of the Multinomial Naive Bayes method in sentiment analysis will also be known so that this sentiment analysis research is expected to be a reference for researchers who want to develop related topics. While the study provides valuable insights into user sentiments towards the M-Paspor application, several limitations should be acknowledged. First, the analysis focused solely on reviews from the Google Play Store, potentially excluding user feedback from other platforms or channels. Additionally, the study did not explore the potential impact of demographic factors, such as age, gender, or location, on user sentiments. Future research could address these limitations by incorporating data from multiple sources and investigating the influence of user characteristics on sentiment patterns.

After conducting the automatic labeling process using the textblob library in the research, it was observed that there were still instances where the sentiment of the data did not align with what was expected. Therefore, it is advisable to explore alternative automated labeling methods such as lexicon-based VADER or SentiwordID, as well as manual labeling, in order to potentially achieve more accurate results. Furthermore, as sentiment analysis techniques continue to evolve, future studies could explore the integration of more advanced algorithms, such as deep learning models or ensemble methods, to potentially enhance the accuracy and robustness of the sentiment analysis process. Additionally, incorporating qualitative analysis or user interviews could provide deeper insights into the reasons behind positive or negative sentiments, complementing the quantitative findings.

## Acknowledgment

Thank you to the Universitas Multimedia Nusantara, Indonesia which has become a place for researchers to develop this journal research. Hopefully, this research can make a major contribution to the advancement of technology in Indonesia

## References

- AlBadani, B., Shi, R., & Dong, J. (2022). A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating the Universal Language Model Fine-Tuning and SVM. *Applied System Innovation*, 5(1). <https://doi.org/10.3390/asi5010013>
- Amir Latif, R. M., Talha Abdullah, M., Aslam Shah, S. U., Farhan, M., Ijaz, F., & Karim, A. (2019). Data scraping from google play store and visualization of its content for analytics. *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies, ICoMET 2019, July 2020*. <https://doi.org/10.1109/ICOMET.2019.8673523>
- Basarslan, M. S., & Kayaalp, F. (2020). Sentiment Analysis with Machine Learning Methods on Social Media. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 9(3), 5–15. <https://doi.org/10.14201/adcaij202093515>
- Braiek, H. Ben, & Khomh, F. (2020). On testing machine learning programs. *Journal of Systems and Software*, 164. <https://doi.org/10.1016/j.jss.2020.110542>
- Chen, K., Zhang, Z., Long, J., & Zhang, H. (2016). Turning from TF-IDF to TF-IGM for term weighting in text classification. *Expert Systems with Applications*, 66, 1339–1351. <https://doi.org/10.1016/j.eswa.2016.09.009>
- Chicco, D., Tötsch, N., & Jurman, G. (2021). The matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, 14, 1–22. <https://doi.org/10.1186/s13040-021-00244-z>
- Damanik, F. J., & Setyohadi, D. B. (2021). Analysis of public sentiment about COVID-19 in Indonesia on Twitter using multinomial naive bayes and support vector machine. *IOP Conference Series: Earth and Environmental Science*, 704(1). <https://doi.org/10.1088/1755-1315/704/1/012027>

Elbagir, S., & Yang, J. (2018). Sentiment analysis of twitter data using machine learning techniques and scikit-learn. *ACM International Conference Proceeding Series*, 0–5. <https://doi.org/10.1145/3302425.3302492>

Hamzah, M. B. (2021). Classification of Movie Review Sentiment Analysis Using Chi-Square and Multinomial Naïve Bayes with Adaptive Boosting. *Journal of Advances in Information Systems and Technology*, 3(1), 67–74. <https://doi.org/10.15294/jaist.v3i1.49098>

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>

Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. *IEEE Access*, 10, 19083–19095. <https://doi.org/10.1109/ACCESS.2022.3151048>

Hossain, E., Sharif, O., & Moshiul Hoque, M. (2021). Sentiment Polarity Detection on Bengali Book Reviews Using Multinomial Naïve Bayes. *Advances in Intelligent Systems and Computing*, 1299 AISC, 281–292. [https://doi.org/10.1007/978-981-33-4299-6\\_23](https://doi.org/10.1007/978-981-33-4299-6_23)

Imigrasi, D. J. (2023). *Kuartal I 2023 Imigrasi Terbitkan 13.292 Paspor Setiap Hari: Naik 38% Sejak 2019*. Imigrasi.Go.Id. <https://www.imigrasi.go.id/id/2023/05/22/kuartal-i-%0A2023-imigrasi-terbitkan-13-292-paspor-setiap-hari-naik-38-sejak-2019/>

Indra. (2022). *Mari Kita Bahas Aplikasi Sejuta Umat! Google Play Store*. Idmetafora.Com. <https://idmetafora.com/news/read/898/Mari-Kita-Bahas-%0AAplikasi-Sejuta-Umat-Google-Play-Store.html>

Kabir, M., Kabir, M. M. J., Xu, S., & Badhon, B. (2021). An empirical research on sentiment analysis using machine learning approaches. *International Journal of Computers and Applications*, 43(10), 1011–1019. <https://doi.org/10.1080/1206212X.2019.1643584>

Kamath, C. N., Bukhari, S. S., & Dengel, A. (2018). Comparative study between traditional machine learning and deep learning approaches for text classification. *Proceedings of the ACM Symposium on Document Engineering 2018, DocEng 2018*. <https://doi.org/10.1145/3209280.3209526>

King, M. R. N., Timms, P. D., & Mountney, S. (2023). A proposed universal definition of a Digital Product Passport Ecosystem (DPPE): Worldviews, discrete capabilities, stakeholder requirements and concerns. *Journal of Cleaner Production*, 384(December 2022), 135538. <https://doi.org/10.1016/j.jclepro.2022.135538>

Lavenia, N. L., & Permatasari, R. (2023). Sentiment Analysis on Twitter Social Media Regarding Depression Disorder Using the Naive Bayes Method. *CoreID Journal*, 1(2), 66–74. <https://doi.org/10.60005/coreid.v1i2.14>

Lorla dan Steven. (2020). TextBlob Documentation Release 0.16.0. *TextBlob*. <https://textblob.readthedocs.io/en/dev/>

Mandasari, S., Hayadi, B. H., & Gunawan, R. (2022). Sentiment Analysis of Online Transportation Users Towards Grab Indonesia Services Using Multinomial Naive Bayes Classifier. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 5(2), 118. <https://ojs.trigunadharma.ac.id/index.php/jsk/article/view/5635>

Mehta, P., & Pandya, S. (2020). A review on sentiment analysis methodologies, practices and applications. *International Journal of Scientific and Technology Research*, 9(2), 601–609.

MIT, L. (2020). *deep-translator*. Pypi.Org. <https://pypi.org/project/deep-translator/>

- Nurhidayah, S. (2020). A MULTINOMIAL NAÏVE BAYES DECISION SUPPORT SYSTEM FOR COVID-19 DETECTION. *FUDMA Journal of Sciences*, 5(1), 55. <https://doi.org/10.33003/fjs-2020-0402-331>
- Olanrewaju, R. O., Olanrewaju, S. A., & Nafiu, L. A. (2022). Multinomial Naïve Bayes Classifier: Bayesian versus Nonparametric Classifier Approach. *European Journal of Statistics*, 2, 8. <https://doi.org/10.28924/ada/stat.2.8>
- OSMANOĞLU, U. Ö., ATAĞ, O. N., ÇAĞLAR, K., KAYHAN, H., & CAN, T. (2020). Sentiment Analysis for Distance Education Course Materials: A Machine Learning Approach. *Journal of Educational Technology and Online Learning*, 3(1), 31–48. <https://doi.org/10.31681/jetol.663733>
- Pardi, P. (2024). Analysis of User Satisfaction of the M-Passport Application Service Using the Technology Acceptance Model (TAM) at TPI Class II Singkawang Immigration Office. *Jurnal Manajemen Pelayanan Publik*, 08(01). <http://journal.unpad.ac.id/jmpp/article/view/51165>
- Permana, A. A., & Putra, P. P. (2023). Sentiment Analysis Public Opinion of Cfw (Citayam Fashion Week) on Social Media Twitter Using Naïve Bayes Classifier. *JIKA (Jurnal Informatika)*, 7(1), 112. <https://doi.org/10.31000/jika.v7i1.7410>
- Pham, B. T., Phong, T. Van, Nguyen, H. D., Qi, C., Al-Ansari, N., Amini, A., Ho, L. S., Tuyen, T. T., Yen, H. P. H., Ly, H.-B., Prakash, I., & Bui, D. T. B. (2020). A Comparative Study of Kernel Logistic Regression, Radial Basis Function Classifier, Multinomial Naïve Bayes, and Logistic Model Tree for Flash Flood Susceptibility Mapping. *Water (Switzerland)*, 12(239). <https://doi.org/10.3390/w12010239>
- Pratmanto, D., Rousyati, R., Wati, F. F., Widodo, A. E., Suleman, S., & Wijianto, R. (2020). App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012043>
- Putri, D. A., Kristiyanti, D. A., Indrayuni, E., Nurhadi, A., & Hadinata, D. R. (2020). Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012085>
- Resti, Y., Irsan, C., Putri, M. T., Yani, I., Anshori, & Suprihatin, B. (2022). Identification of Corn Plant Diseases and Pests Based on Digital Images using Multinomial Naïve Bayes and K-Nearest Neighbor. *Science and Technology Indonesia*, 7(1), 29–35. <https://doi.org/10.26554/sti.2022.7.1.29-35>
- Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, 874(1). <https://doi.org/10.1088/1757-899X/874/1/012017>
- Santoso, D. B., Munna, A., & Untari Ningsih, D. H. (2024). Improved playstore review sentiment classification accuracy with stacking ensemble. *Journal of Soft Computing Exploration*, 5(1), 38–45. <https://doi.org/10.52465/josce.v5i1.247>
- Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Human-Centric Computing and Information Sciences*, 7(1). <https://doi.org/10.1186/s13673-017-0116-3>
- Stepanek, H. (2020). Thinking in pandas: How to use the python data analysis library the right way. In *Thinking in Pandas: How to Use the Python Data Analysis Library the Right Way*. <https://doi.org/10.1007/978-1-4842-5839-2>
- Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. In *Artificial Intelligence Review* (Vol. 55, Issue 7). Springer Netherlands. <https://doi.org/10.1007/s10462-022-10144-1>

Xie, X., Ho, J. W. K., Murphy, C., Kaiser, G., Xu, B., & Chen, T. Y. (2011). Testing and validating machine learning classifiers by metamorphic testing. *Journal of Systems and Software*, 84(4), 544–558. <https://doi.org/10.1016/j.jss.2010.11.920>

Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. *Information Sciences*, 507, 772–794. <https://doi.org/10.1016/j.ins.2019.06.064>

Yao, J. (2019). Automated Sentiment Analysis of Text Data with NLTK. *Journal of Physics: Conference Series*, 1187(5). <https://doi.org/10.1088/1742-6596/1187/5/052020>

Zaenuddin, M. (2023). *Cara Download M-Paspor untuk Mengajukan Permohonan Paspor Online*. Kompas.Com. <https://www.kompas.com/tren/%0Aread/2023/05/14/201500565/cara-download-m-paspor-untuk-mengajukanpermohonan-%0Apaspor-online>

Zahoor, K., Bawany, N. Z., & Hamid, S. (2020). Sentiment analysis and classification of restaurant reviews using machine learning. *Proceedings - 2020 21st International Arab Conference on Information Technology, ACIT 2020, November*. <https://doi.org/10.1109/ACIT50332.2020.9300098>

Zhou, H. (2022). Research of Text Classification Based on TF-IDF and CNN-LSTM. *Journal of Physics: Conference Series*, 2171(1), 218–222. <https://doi.org/10.1088/1742-6596/2171/1/012021>