Jurnal Manajemen Pelayanan Publik
Vol. 10 No. 01, February 2026
http://dx.doi.org/10.24198/jmpp.v10i1.68645

Evaluation of the M-Paspor App as a Public Service: A Sentiment Analysis

[a] Ignatius Novianto Hariwibowo; [b] Wimpie Yustino Setiawan
[a b] Accounting Department, Universitas Atma Jaya Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia

ABSTRAK

Tujuan penelitian ini adalah untuk mengevaluasi kinerja layanan publik Kantor Imigrasi melalui aplikasi M-Paspor, menggunakan analisis sentimen. Penelitian ini menggunakan data komentar atau review dari pengguna yang ada di Google Play Store. Pengambilan data menggunakan teknik web scraping dengan menyasar komentar pengguna mendapatkan data sebanyak 12.138 komentar. Analisis data dalam penelitian ini menggunakan analisis sentimen untuk mengidentifikasi dan labeling pernyataan positif atau negatif. Untuk memastikan hasil labeling, maka penelitian ini juga melakukan uji model untuk memastikan tingkat akurasinya. Hasil penelitian ini menunjukkan bahwa nilai komentar negatif adalah 69,1% dan positif sebanyak 30,9%. Hasil uji model menunjukkan tingkat akurasi di atas 80% pada algoritma Naive Bayes, Random Forest, Logistic Regression, dan Decision Tree. Hasil ini menunjukkan bahwa mayoritas masyarakat pengguna M-Paspor tidak puas dengan layanan tersebut. Hal ini mengindikasikan kinerja sistem dalam M-Paspor yang tidak dapat memenuhi harapan pengguna.

ABSTRACT

This study is purposed to evaluate the public service performance of the Immigration Office's M-Passport application using sentiment analysis. This study implemented the data of comments or customer reviews from the Google Play Store. A web scraping technique was employed to collect data, with a total number of 12,138 comments collected. Sentiment analysis was applied to identify and label positive and negative comments. In order to ensure the labeling results, a model test was carried out to confirm the accuracy level. The findings show that 69.1 % comments are negative and 30.9% are positive. The model testing results indicated a more than 80% accuracy level based on the Naive Bayes algorithm, Random Forest, Logistic Regression, and Decision Tree. The results showed that the majority of users are not satisfied with the service, indicating that the system performance of M-Passport has not satisfied user expectations.

ARTICLE HISTORY
Submitted: 09 12 2025
Revised: 27 01 2026
Accepted: 30 01 2026
Published: 28 02 2026

KATA KUNCI
Layanan Publik; M-Paspor; Analisis Sentimen

KEYWORDS
Public Service; M-Paspor; Sentiment Analysis

INTRODUCTION

The digitalization of public services is a central pillar of the Indonesian government's strategy to provide transparent, efficient, and accessible services. In the immigration sector, this is embodied by the M-Paspor app, which allows citizens to schedule passport services online. While such innovations align with the national e-Government mandate in Law No. 25 of 2009, their success is ultimately measured by their ability to meet public expectations and generate satisfaction as a key performance indicator for government services (Reddick & Demir, 2024; Shafritz et al., 2016; Van Ryzin, 2013).

The depth, novelty, and relevance of this study lie in the synthesis of this established theory with this cutting-edge methodology in a new and critical context. While previous research has applied sentiment analysis methods in sectors like transportation (Saputra & Kumar, 2025), social factors (Abouelela et al., 2025), and public safety (Săvescu et al., 2025), their application to evaluate Indonesia's digital immigration services is unprecedented. Moreover, research related to the success of M-Paspor has also not used sentiment analysis. Hence, this study

Correspondence Ignatius Novianto Hariwibowo Email : novianto.wibowo@uajy.ac.id

proposes that machine learning-based sentiment analysis is not merely appropriate but is the most suitable method to fill the existing gap for three reasons: (1) it leverages the vast, untapped resource of organic user reviews as real-time feedback; (2) it provides a quantitative, objective, and scalable measure of the "perceived performance" construct in EDT; and (3) it enables a dynamic analysis that can track changes in sentiment over time, offering insights far beyond a static snapshot. Therefore, the substantial contribution of this study is to bridge this theoretical and methodological gap. This study proposes a novel, integrated framework that applies machine learning-based sentiment analysis to extensive user review data to quantitatively assess the disconfirmation of expectations and the resultant public satisfaction with the M-Paspor application.

The significance of this work is twofold. First, it provides policymakers and immigration authorities with actionable, evidence-based insights into the application's strengths and weaknesses, directly informing future system improvements. Second, it contributes to academic discourse by demonstrating how established Expectancy Disconfirmation Theory (EDT) can be effectively operationalized with modern computational methods to create a robust and transferable framework for evaluating digital public services, not just in immigration but across the broader e-Government landscape.

Literature Review

Expectancy Disconfirmation Theory (EDT)

The foundational framework for understanding service satisfaction can be explained by Expectancy Disconfirmation Theory (EDT) (Oliver, 1980). EDT posits that satisfaction is a function of the gap between user expectations and their perceived performance of a service. This theory is particularly relevant for digital services such as M-Paspor, where high expectations of efficiency often clash with the user's experience, leading to dissatisfaction (Bressolles et al., 2014; Guy & Ely, 2018; Morgeson, 2013; Priyanto, 2024).

Prior Studies

To evaluate the success of the M-Paspor app, prior research has identified significant technical and functional issues, such as data resynchronization and usability challenges, indicating that its implementation requires critical evaluation (Bahasoan, 2024; R. A. Pratama & Utami, 2023; Rabbani & Fajar, 2024; Stahl, 2022). However, a thorough review of existing literature reveals a critical methodological gap. Previous studies have predominantly relied on conventional methods such as surveys, interviews, and descriptive qualitative analysis, which are often limited by sample size, geographic scope, and temporal constraints (Claudia & Lestari, 2024; Liang & Yuan, 2016; Yaziji et al., 2023). While these approaches provide valuable initial insights, they are inherently constrained. They cannot capture the massive, dynamic, and real-time avalanche of unstructured user feedback spread across digital platforms like the Google Play Store, Apple App Store, and social media. This inability to process large-scale, organic feedback represents a significant gap in achieving a comprehensive and objective evaluation of the service.

To operationalize this theory at scale and address the identified methodological gap, this study offers advanced analytical approaches validated in other domains. A robust body of literature demonstrates the potent efficacy of machine learning-based sentiment analysis for public service evaluation. For instance, a study by Zharif Mustaqim et al. (2024) successfully applied this method to assess broad public satisfaction with digital government service platforms, demonstrating its capacity to handle large volumes of data. Similarly, studies on public safety services (Giri Van Transco & Asti Herliana, 2025) and policy responses (I. R. Pratama et al., 2025) have effectively utilized Natural Language Processing (NLP) and classification algorithms such as Support Vector Machines (SVM) to extract actionable insights from social media and user reviews. These studies confirm that machine learning is not just a technical tool but a robust methodological paradigm for systematically classifying public opinion, identifying prevalent complaints, and measuring satisfaction levels from unstructured text data.

RESEARCH METHODS

Research Approach

This research uses a quantitative approach with a descriptive and exploratory research design. The quantitative approach was chosen because this research aims to analyze large amounts of data collected from user reviews of the M-Paspor application. By using the sentiment analysis method, this research aims to provide an objective picture of the level of user satisfaction and identify problems in the M-Paspor application service. In addition, this approach is also exploratory as it aims to explore the factors that influence user satisfaction, including expectations, perceived performance, and differences in expectations and experiences, analyzed using Expectancy Disconfirmation Theory (EDT).

Research Data

The type of data used in this study is secondary data obtained from user reviews of the M-Paspor app published on platforms such as the Google Play Store. This review data contains direct feedback from users, which includes opinions, complaints, and praise related to the use of the application. Sentiment data collected will be analyzed to assess the positive and negative sentiments reflected in the reviews. This research will be limited to the evaluation of the M-Paspor developed by the Indonesian Immigration Office. The sample size for the study was determined randomly using the Slovin approach. The number of reviews on the Google Play Store was 41,400. With an alpha level of 1%, the minimum number of reviews that could be used as a sample was 3,901. The data used are only public reviews that can be accessed by anyone. In addition, this study will follow the ethical guidelines applicable in research using data from public platforms.

Data Collection Technique

The data collection techniques used in this study are web scraping and downloading all review data from the Play Store platforms. Web scraping will be used to extract text data from reviews that include user opinions about the M-Paspor app. These reviews will be collected randomly over a period of time to get a more complete picture of the user experience over time. The data obtained will be processed using the Python language on Google Colab.

Data Analysis Technique

To analyze the collected data, this study uses machine learning-based sentiment analysis. This analysis process includes the three steps applied by Jin et al. (2023). The first step is text preprocessing. This stage is carried out with several steps, namely: cleaning (removing punctuation marks), case folding (making text into lowercase letters), normalization (converting into standard words), stop-word removal (removing conjunctions), tokenization (separating words in sentences), and stemming (converting words into basic words). The second step is labeling. At this stage, the results of the cleaned sentences will be identified as positive or negative. In this labeling process, we will use a transformation with the Term Frequency-Inverse Document Frequency (TF-IDF) method. This method is used for word weighting. The third step is model evaluation. At this stage, the results of the labels obtained will be evaluated for accuracy. This step is to ensure the labeling results are acceptable. In this research, model evaluation is carried out with Naive Bayes, Random Forest, Logistic Regression, and Decision Tree.

RESULTS AND DISCUSSIONS

Text Data Preprocessing

The data in this study were obtained by using web scraping. Web scraping is done with a library, which is Google Play Scraper. Figure 1 shows an example of a screen capture of the code used in web scraping.

Figure 1.
Web Scraping with google_play_scraper

```
# Mengimpor pustaka google_play_scraper untuk mengakses ulasan dan informasi aplikasi dari Google Play Store.
from google_play_scraper import app, reviews_all, Sort


scrapreview = reviews_all(
    'id.go.imigrasi.paspor_online',          # ID aplikasi
    lang='id',              # Bahasa ulasan (default: 'en')
    country='id',  # Negara (default: 'us')
    sort=Sort.MOST_RELEVANT,  # Urutan ulasan (default: Sort.MOST_RELEVANT)
    count=12000)  # Jumlah maksimum ulasan yang ingin diambil
```

Source: Processed data (2025)

In the code, the app address, country, and number of reviews are specified for download. The result of the code is data in the form of a data table in CSV format that is ready to be processed. Although the minimum sample size is 3,901, this study took 12,000 reviews, which is the number of the system that can be efficiently retrieved. The total amount of data obtained by scraper algorithm is 12,138 comments with 11 columns of information. However, for the data to be used, the data needs to be prepared by cleaning it of empty data values. Data cleaning is done by eliminating the command: clean_df = data.dropna (axis=1, how ='any'). The result of the data after cleaning is that there are 2 columns that are deleted, so there are 7 columns that are used. The result is shown in Figure 2.

Figure 2.
Review the Data of Applications Used

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12139 entries, 0 to 12138
Data columns (total 7 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   reviewId       12139 non-null  object
 1   userName       12139 non-null  object
 2   userImage      12139 non-null  object
 3   content        12139 non-null  object
 4   score          12139 non-null  int64
 5   thumbsUpCount  12139 non-null  int64
 6   at             12139 non-null  object
dtypes: int64(2), object(5)
memory usage: 664.0+ KB
```

Source: Processed data (2025)

Table 1 provides an example of the results from the data set that will be used for further analysis.

Table 1.
Research Data Set

	Review ID	User Name	User Image	Content	Score	Thumbs Up Count	Review Created Version	Time
0	429df713-8af6-4381-a01e-7f44bee58706	Pengguna Google	https://play-lh.googleusercontent.com/EGemoI2N...	*LOL. Ga bisa dipake. Baru di Langkah 1, udah e...* (LOL. Can't use it. Just on Step 1, already e...)	1	5	6.1.0	07/07/2024 11:59
1	8255570f-7aa2-435e-a801-884c4d5618d4	Pengguna Google	https://play-lh.googleusercontent.com/EGemoI2N...	*Tolong bantu validasi kalau misalnya sudah pun...* (Please help validate if you have already...)	2	48	6.1.0	04/05/2024 06:30
2	6e71b0e2-31a7-417b-8f09-b5d966f85368	Pengguna Google	https://play-lh.googleusercontent.com/EGemoI2N...	*Aplikasinya perlu terus ditingkatkan, misalnya...* (Its application needs to be continuously improved, for example...)	2	113	6.1.0	28/03/2024 14:36
3	c10ffc98-b1d4-4ee4-b44e-f2562c25ea8d	Pengguna Google	https://play-lh.googleusercontent.com/EGemoI2N...	*Sistem pembayaran kacau sekali. Saat memencet...* (The payment system is a mess. When you press...)	1	93	6.1.0	02/06/2024 05:51
4	ef927aeb-ce03-4590-9b80-21ea5bb114ae	Pengguna Google	https://play-lh.googleusercontent.com/EGemoI2N...	*Saat pendaftaran akun, Email typo, jadi otp ti...* (During account registration, Email typo, so otp ti...)	1	47	6.1.0	11/05/2024 09:32

Source: Processed data (2025)

Data Preprocessing

The next stage is preprocessing. This stage is carried out to clean the sentence from punctuation marks that can interfere with the meaning of the sentence and clean the sentence from the appropriate words, known as stop words. The following is a command in Python that is used to clean the sentence. At this stage, the given comment sentence will be separated into basic words. This step is done to get the basic meaning of the word used in the sentence. The formulation of the functions used to perform tokenization and stemming is shown in Table 2.

Table 2.
Example of Data Preprocessing Results

Content	*LOL. ga bisa dipake. baru di Langkah 1, udah error aja. karena ga ada mekanisme bug report atau kontak CS, mending lewat sini kali. PlatformException(IO_ERROR, A network error occured trying to lookup the supplied coordinates). Mana ga ada aplikasi/cara alternatif lagi. masa ga dikasih CS untuk ini di aplikasinya? gimana mau dengerin feedback? rata-rata bintang dua emang gini kali kerjanya* (LOL. Can't use it. just in Step 1, it's already an error. since there is no bug report mechanism or CS contact, it's better to go here. PlatformException(IO_ERROR, A network error occurred trying to lookup the supplied coordinates). There's no alternative application / way. there's no CS for this in the application? how do you want to listen to feedback? the average two-star works like this.)
text_clean	*LOL ga bisa dipake baru di Langkah udah error aja karena ga ada mekanisme bug report atau kontak CS mending lewat sini kali PlatformExceptionIOERROR A network error occured trying to lookup the supplied coordinates Mana ga ada aplikasicara alternatif lagi masa ga dikasih CS untuk ini di aplikasinya gimana mau dengerin feedback ratarata bintang dua emang gini kali kerjanya* (LOL can't use it just in Step already errored because there is no bug report mechanism or CS contact mending through here times PlatformExceptionIOERROR A network error occured trying to lookup the supplied coordinates Where there is no alternative speaker application again no time given CS for this in the application how do you want to listen to two-star average feedback this is how it works)
text_casefoldingText	*lol ga bisa dipake baru di langkah udah error aja karena ga ada mekanisme bug report atau kontak cs mending lewat sini kali platformexceptionioerror a network error occured trying to lookup the supplied coordinates mana ga ada aplikasicara alternatif lagi masa ga dikasih cs untuk ini di aplikasinya gimana mau dengerin feedback ratarata bintang dua emang gini kali kerjanya*
text_slangwords	*tertawa ga bisa dipake baru di langkah sudah error saja karena ga ada mekanisme bug report atau kontak cs lebih baik lewat sini kali platformexceptionioerror a network error occured trying to lookup the supplied coordinates mana ga ada aplikasicara alternatif lagi masa ga dikasih cs untuk ini di aplikasinya bagaimana mau dengerin feedback ratarata bintang dua emang gini kali kerjanya* (lol can't be used new in step already error just because there is no bug report mechanism or contact cs better through here times platformexceptionioerror a network error occured trying to lookup the supplied coordinates where there is no applicationalternative way again time is not given cs for this in the application how do you want to listen to feedback two-star average is like this times it works laughing can't be used just in the step already error only because there is no bug report mechanism or contact cs better through here times platformexceptionioerror a network error occured trying to lookup the supplied coordinates where there is no applicationalternative speaker again

	no time given cs for this in the application how do you want to listen to feedback two-star average this is how it works)
text_tokenizingText	[*tertawa, ga, bisa, dipake, baru, di, langkah, sudah, error, saja, karena, ga, ada, mekanisme, bug, report, atau, kontak, cs, lebih, baik, lewat, sini, kali, platformexceptionioerror, a, network, error, occured, trying, to, lookup, the, supplied, coordinates, mana, ga, ada, aplikasicara, alternatif, lagi, masa, ga, dikasih, cs, untuk, ini, di, aplikasinya, bagaimana, mau, dengerin, feedback, ratarata, bintang, dua, emang, gini, kali, kerjanya*] [laugh, no, can, use, new, in, step, already, error, only, because, no, mechanism, bug, report, or, contact, cs, better, through, here, times, platformexceptionioerror, a, network, error, occured, trying, to, lookup, the, supplied, coordinates, where, no, there, is, an, application, alternative, again, time, no, given, cs, for, this, in, the, application, how, want, listen, feedback, average, star, two, really, like, this, times, works]
text_stopword	[tertawa, dipake, langkah, error, mekanisme, bug, report, kontak, cs, kali, platformexceptionioerror, network, error, occured, trying, lookup, supplied, coordinates, aplikasicara, alternatif, dikasih, cs, aplikasinya, dengerin, feedback, ratarata, bintang, emang, gini, kali, kerjanya] [laugh, used, step, error, mechanism, bug, report, contact, cs, times, platformexceptionioerror, network, error, occured, trying, lookup, supplied, coordinates, applicationtalk, alternative, given, cs, application, listen, feedback, average, star, really, this, times, works]
text_final	*tertawa dipake langkah error mekanisme bug report kontak cs kali platformexceptionioerror network error occured trying lookup supplied coordinates aplikasicara alternatif dikasih cs aplikasinya dengerin feedback ratarata bintang emang gini kali kerjanya* (laughed at step error mechanism bug report contact cs times platformexceptionioerror network error occured trying lookup supplied coordinates applicationalternative speaker given cs the application listened to feedback average star emang this time it works)

Source: Processed data (2025)

Labeling

This step comes when the data has been cleaned and ready to be processed. This stage is very important to be able to determine positive and negative responses to M-Paspor app's comments. This labeling step is carried out by calculating the frequency of word occurrence (TF), the sum of word occurrences in comments (DF), and IDF calculations, which ultimately determine the weight of the document based on all words. To get the appropriate results, the positive and negative word list groups need to be included in the algorithm first. This group of words will be used as a reference for assessing comments. In this study, positive and negative word lists have been created and identified first into positive lexicon and negative lexicon groups. Figure 3 indicates an example of the command to display the positive lexicon.

Figure 3.
Example of Command for Scoring or Weighting

```
print("Positive Lexicon Sample:", list(lexicon_positive.items())[:5])  # Print first 5 entries

Positive Lexicon Sample: [('hai', 3), ('merekam', 2), ('ekstensif', 3), ('paripurna', 1), ('detail', 2)]
```

Source: Processed data (2025)

In these examples, the words with positive meanings are “hi”, ‘recording’, “detail”, and others. Meanwhile, an example of a command to display a negative lexicon is shown in Figure 4.

Figure 4.
Example of Command to Display Scoring or Weighting Results

```
print("Negative Lexicon Sample:", list(lexicon_negative.items())[:5])

Negative Lexicon Sample: [('putus tali gantung', -2), ('gelebah', -2), ('gobar hati', -2), ('tersentuh (perasaan)', -1)
```

Source: Processed data (2025)

In this example, the words that are classified as negative words are “breaking the hanging rope”, ‘searched’, “touched”, and others. These positive and negative words have a score level to distinguish the positive and negative content of each word. The use of this lexicon will be combined in sentences so that the tendency can be known to be positive or negative. Table 3 provides an example of a comment that has been cleaned and will be assessed with negative and positive lexicons.

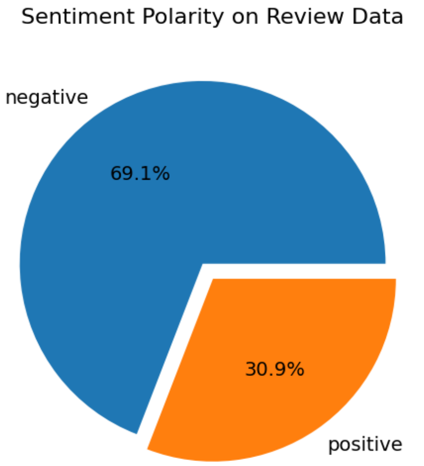
Table 3.
Scoring Statement and Labeling

Final Text	Score	Label
tertawa dipake langkah error mekanisme bug report kontak cs kali platformexceptionioerror network error occured trying lookup supplied coordinates aplikasicara alternatif dikasih cs aplikasinya dengerin feedback ratarata bintang emang gini kali kerjanya (laughed at step error mechanism bug report contact cs times platformexceptionioerror network error occurred trying lookup supplied coordinates applicationalternative speaker given cs the application listened to feedback average star emang this time it works)	2	Positive
tolong bantu validasi passpor salah opsi registrasi pendaftaran munculkan validasi memiliki passpor status pengajuan edit kaum awam rugi doong masuk opsi pembaruan pas pendaftaran pilih munculkan interface opsi pembuatan passpor pergantian passpor (Please help passport validation wrong registration option registration pop up validation have passport application status edit layman loss doong enter registration pass renewal option select pop up interface passport creation option passport change passport)	-1	Negative

Source: Processed data (2025)

The result is that the value for negative sentiment is 8,392 and positive sentiment is 3,747. In percentage terms, the result is summarised in Figure 5.

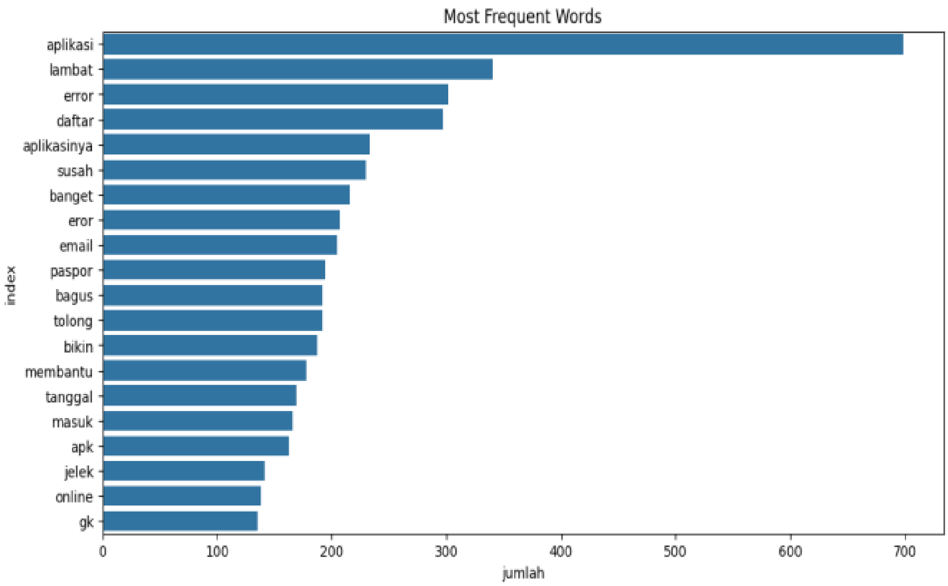
Figure 5.
Sentiment Analysis Results



Source: Processed data (2025)

From these results, it can be seen that, in general, the negative sentiment is greater than the positive sentiment, which is 69.1%. This large number of negative sentiments indicates dissatisfaction with M-Paspor public services. In order to obtain more details about the words tested, Figure 6 provides the list of the words that have the highest frequency of all comments submitted by users.

Figure 6.
List of Most Frequently Occurring Words



Source: Processed data (2025)

From the frequency of these words, it can be seen that there are more negative words than positive words. This condition causes the total accumulation of negative values to be greater, so it can be interpreted that negative sentiment is greater than positive sentiment. The results in Table 2 also show that the dissatisfaction of M-Paspor users is more caused by technical problems. This can be seen from the negative words that often appear: slow, error, and difficult. This condition shows the unpreparedness of the application infrastructure support when serving a large number of consumers, causing technical problems in its use.

Model Evaluation

In order to know and ensure that the results of the labeling process are acceptable, the result evaluation stage needs to be carried out to assess the accuracy of the results through model evaluation. This model evaluation will compare the accuracy of the comment with its label, which is positive or negative. Thus, the step taken is to divide the data into X and Y variables. X variables are comments whose sentences have been cleaned, and Y is the label result of each comment. Figure 7 shows the command to split the data into X and Y.

Figure 7.
Example of Command to Split X and Y Data for Evaluation

```
# Pisahkan data menjadi fitur (tweet) dan label (sentimen)
X = clean_df['text_akhir']
y = clean_df['polarity']
```

Source: Processed data (2025)

The results of X and Y values will be compared between the predicted and actual values. The level of accuracy indicates that the predicted results match the actual results. Therefore, the higher the percentage value, the higher the accuracy value. The command to divide the predicted data by the actual data is shown in Figure 8.

Figure 8.
Example of Command to Divide the Predicted Data by the Actual Data

```
# Bagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X_tfidf, y, test_size=0.2, random_state=42)
```

Source: Processed data (2025)

Model evaluation is performed by dividing the data into training and test data from the total data obtained. The data division is 20% test data, while the data for training is 80%. The final results of the evaluation will be seen from four algorithms, namely Logistic Regression, Random Forest, Decision Tree, and Naive Bayes. The comparison of the three after model evaluation is provided in Table 4.

Table 4.
Model Test Results

Model	Accuracy
Logistic Regression	0.897858
Random Forest	0.883031
Decision Tree	0.864086
Naive Bayes	0.857084

Source: Processed data (2025)

The results shown in Table 4 indicate that all show accuracy values above 80%, with the highest value in the logistic regression algorithm. Therefore, Table 4 shows that, overall, the results of the sentiment analysis in this study are accurate.

Discussion

This study is purposed to evaluate the assessment of M-Paspor public services by the Immigration Office using sentiment analysis. The sentiment analysis results show that 69.1% of the collected sentiments are negative, while only 30.9% are positive. The test results of the sentiment analysis model show an accuracy rate of more than 80%, which indicates that the sentiment analysis method can be effectively applied in assessing public perceptions of public services. This result shows that many people are dissatisfied with the services of the M-Paspor app. This result is also shown by research conducted by Zhang et al. (2022) on the acceptance of e-government services, which shows that user dissatisfaction occurs when their expectations do not match the reality. In the context of this study, the results showing 69.1% negative sentiment can be understood as an indication that the M-Paspor service has not met user expectations. Other research conducted by Van Ryzin (2013) also revealed that when technology-based public services do not meet expectations, high levels of dissatisfaction will arise. The research identified that users tend to leave negative reviews when they feel that services are inadequate in terms of speed and transparency (Chatterjee & Suy, 2019; Grimmelikhuijsen & Porumbescu, 2017). This finding supports our research results that show a predominantly negative sentiment towards the M-Paspor service.

These findings indicate that the M-Paspor service still faces major challenges in meeting user expectations in terms of application technicalities (see Figure 6). To increase user satisfaction, the Immigration Office needs to make improvements in aspects that are often complained about, such as service speed, accessibility, and transparency. This is in line with the findings of Yaziji et al. (2023), which suggest that e-government service user satisfaction is strongly influenced by the achievement of user expectations. Therefore, improvement efforts that focus on technical aspects can help reduce dissatisfaction (Claudia & Lestari, 2024; R. A. Pratama & Utami, 2023).

The research results contribute the EDT context to strengthening the role of expectations in building service satisfaction. Through this theory, satisfaction can be understood as a complex cognitive process of assessing the comparison between expectations and reality (Filtenborg et al., 2017). With a model accuracy rate of over 85%, this study provides an understanding that the complexity of public service satisfaction can be multidimensional, namely: technical aspects related to the system, and non-technical aspects related to user psychology. In this case, this study successfully offers a more natural and efficient approach to assessing the complexity of this cognitive process using a sentiment analysis approach. By using public responses to public services, sentiment analysis can assess the performance of public services more broadly. These results can illustrate the broad responses and expectations of the public policies that are made. This approach is a suitable alternative to get theoretical confirmation with a wide response rate and more real-time experience. Thus, the use of sentiment analysis with machine learning can be used to assess the reasons for acceptance or rejection of a technology or system, as with the Technology Acceptance Model (TAM) model, or an evaluation of the success of a system implementation, as with the system success model from DeLone & Mclean. While it may not be a complete replacement, sentiment analysis with machine learning can provide information on factors that may affect the acceptance or success of system implementation, naturally based on user experience (Mishev et al., 2020; Talaat, 2023). Thus, the theoretical contribution of this study is to provide empirical evidence and logical arguments to enrich the EDT model, making it more relevant and robust for analyzing digital public services in the modern era, where trust and procedural justice are crucial measures of success.

The practical contribution of this study is a call for cultural transformation within the public sector. This transformation shifts the focus from achieving internal technical targets to creating value that is felt by citizens. By adopting user-centered metrics, strengthening regulations on fairness and transparency, and integrating EDT into service development practices, Indonesia can build a digital public service ecosystem that is not only technically efficient but also fair, transparent, and trustworthy. This entire analysis underscores that in the world of electronic governance, user satisfaction is not an end result achieved by chance, but rather the result of intelligent, sustainable, and human-oriented expectation management (Claudia & Lestari, 2024; R. A. Pratama & Utami, 2023).

CONCLUSION

This study evaluates the public service performance of the Immigration Office's M-Passport application using sentiment analysis. From a total of 12,138 comments collected from the Google Play Store, the analysis results show that the majority of users are not satisfied with the service. These findings indicate a less-than-optimal performance of public services by the Immigration Office. Therefore, improvements need to be made to improve service quality and meet community expectations, especially in the technical aspect, which causes a lot of dissatisfaction with the M-Paspor service. In addition, sentiment analysis can be used to confirm the reality of expectations from the implementation of public service systems or policies.

REFERENCES

Abouelela, M., Tirachini, A., Chaniotakis, E., & Antoniou, C. (2025). Why do passengers use pooled-rides services? Social effects and implications for policy making. *Travel Behaviour and Society*, *41*, 101072. https://doi.org/10.1016/j.tbs.2025.101072

Azzami, D. R., Baenil Huda, Agustia Hananto, & Tukino. (2025). Analysis of User Comment Sentiment on the Siwaslu Application Using the Naive Bayes Method. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *4*(3), 2235–2240. https://doi.org/10.59934/jaiea.v4i3.1132

Bahasoan, A. (2024). Public Service Transformation: Innovative Strategies of Ambon City Government Towards Openness and Public Satisfaction. *Jurnal Manajemen Pelayanan Publik, 8*(2), 345-358. https://doi.org/10.24198/jmpp.v8i2.52436

Bressolles, G., Durrieu, F., & Senecal, S. (2014). A consumer typology based on e-service quality and e-satisfaction. *Journal of Retailing and Consumer Services*, *21*(6), 889–896. https://doi.org/10.1016/j.jretconser.2014.07.004

Chatterjee, R., & Suy, R. (2019). An Overview of Citizen Satisfaction with Public Service: Based on the Model of Expectancy Disconfirmation. *Open Journal of Social Sciences*, *07*(04), 243–258. https://doi.org/10.4236/jss.2019.74019

Chen, T., Guo, W., Gao, X., & Liang, Z. (2021). AI-based self-service technology in public service delivery: User experience and influencing factors. *Government Information Quarterly*, *38*(4), 101520. https://doi.org/10.1016/j.giq.2020.101520

Claudia, & Lestari, D. P. (2024). User-Centered Interaction Design for Indonesia's E-Government Passport Application: The M-Paspor Indonesia Case Study. *2024 IEEE International Conference on Data and Software Engineering (ICoDSE)*, 235–240. https://doi.org/10.1109/ICoDSE63307.2024.10829902

Filtenborg, A. F., Gaardboe, F., & Sigsgaard-Rasmussen, J. (2017). Experimental replication: an experimental test of the expectancy disconfirmation theory of citizen satisfaction. *Public Management Review*, *19*(9), 1235–1250. https://doi.org/10.1080/14719037.2017.1295099

Giri Van Transco, & Asti Herliana. (2025). Public Sentiment Analysis on Police Service Satisfaction Using Twitter Dataset Based on NLP and SVM. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *4*(3), 2222–2229. https://doi.org/10.59934/jaiea.v4i3.1139

Grimmelikhuijsen, S., & Porumbescu, G. A. (2017). Reconsidering the expectancy disconfirmation model. Three experimental replications. *Public Management Review*, *19*(9), 1272–1292. https://doi.org/10.1080/14719037.2017.1282000

Guo, Y. (2024). *Digital Government and Public Interaction*. IGI Global. https://doi.org/10.4018/979-8-3693-3665-6

Guy, M.E., & Ely, T.L. (2018). *Essentials of Public Service: An Introduction to Contemporary Public Administration*. Melvin & Leigh, Publishers.

Hadwan, M., Al-Sarem, M., Saeed, F., & Al-Hagery, M. A. (2022). An Improved Sentiment Classification Approach for Measuring User Satisfaction toward Governmental Services' Mobile Apps Using Machine Learning Methods with Feature Engineering and SMOTE Technique. *Applied Sciences*, *12*(11), 5547. https://doi.org/10.3390/app12115547

Jin, Y., Cheng, K., Wang, X., & Cai, L. (2023). A Review of Text Sentiment Analysis Methods and Applications. *Frontiers in Business, Economics and Management*, *10*(1), 58–64. https://doi.org/10.54097/fbem.v10i1.10171

Kowalski, R., Esteve, M., & Jankin Mikhaylov, S. (2020). Improving public services by mining citizen feedback: An application of natural language processing. *Public Administration*, *98*(4), 1011–1026. https://doi.org/10.1111/padm.12656

Morgeson, F. V. (2013). Expectations, Disconfirmation, and Citizen Satisfaction with the US Federal Government: Testing and Expanding the Model. *Journal of Public Administration Research and Theory*, *23*(2), 289–305. https://doi.org/10.1093/jopart/mus012

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research*, *17*(4), 460–469. https://doi.org/10.1177/002224378001700405

Pratama, I. R., Cahyana, Y., Rahmat, & Wahiddin, D. (2025). Model Machine Learning Untuk Analisis Sentimen Masyarakat Terhadap Kenaikan PPN di Media Sosial X. *Bulletin of Computer Science Research*, *5*(4), 277–286.

Pratama, R. A., & Utami, D. A. (2023). Effectiveness of M-Paspor Applications in Aspects of Public Service and E-government. *Jurnal Ilmiah Kajian Keimigrasian*, *6*(1), 109–118. https://doi.org/10.52617/jikk.v6i1.380

Priyanto, H. (2024). Public Service Quality in Banyuwangi District: A Study in Welfare Perspective. *Jurnal Manajemen Pelayanan Publik*, *8*(1), 77-94. https://doi.org/10.24198/jmpp.v8i1.48657

Rabbani, M., & Fajar, A. N. (2024). A Comprehensive Examination of User Satisfaction in Indonesian Passport Services: Insights from the M-Paspor Application. *Article in Journal of Theoretical and Applied Information Technology*, *15*(5). https://www.researchgate.net/publication/379052755

Rahayu, A. I., & Kacung, S. (2025). Sentiment Analysis of Online Lending Services Using Support Vector Machine and Logistic Regression. *Bulletin of Computer Science Research*, *5*(5), 447–455.

Reddick, C.G., & Demir, T. (2024). *Handbook Of Public Service Delivery*. Edward Elgar Publishing.

Saputra, J. P. B., & Kumar, A. (2025). Emotion Detection in Railway Complaints Using Deep Learning and Transformer Models: A Data Mining Approach to Analyzing Public Sentiment on Twitter. *Journal of Digital Society*, *1*(2).

Săvescu, A. R., Mazurencu-Marinescu-Pele, M., Conda, A.-I., Bolovăneanu, V., & Pele, D. T. (2025). Infodemic insights: Mapping COVID-19’s digital discourse in Romania. *Management & Marketing*, *20*(1), 15–34. https://doi.org/10.2478/mmcks-2025-0003

Shafritz, J.M., Russell, E.W., Borick, C.P., & Hyde, A.C. (2016). *Introducing Public Administration*. Taylor & Francis.

Shao, K., & Wang, X. (2023). Do government subsidies promote enterprise innovation?—— Evidence from Chinese listed companies. *Journal of Innovation & Knowledge*, *8*(4), 100436. https://doi.org/https://doi.org/10.1016/j.jik.2023.100436

Stahl, M.M. (2022). *The Promise of Public Service: Ideas and Examples for Effective Service*. Taylor & Francis.

Van Ryzin, G. G. (2013). An Experimental Test of the Expectancy-Disconfirmation Theory of Citizen Satisfaction. *Journal of Policy Analysis and Management*, *32*(3), 597–614. https://doi.org/10.1002/pam.21702

Wang, C. (2025). *Text Mining for Mental Disease Screening for Social Network Users* (pp. 19–27). https://doi.org/10.2991/978-2-38476-422-8_4

Yaziji, W., Aprilia, R., Sensuse, D. I., Lusa, S., & Wibowo, P. A. (2023). Assessing Service Quality in Passport Application Mobile Apps(M-Paspor) and its Influence on User Satisfaction. *Indonesian Journal of Computer Science*, *12*(6). https://doi.org/10.33022/ijcs.v12i6.3545

Zhang, J., Chen, W., Petrovsky, N., & Walker, R. M. (2022). The Expectancy-Disconfirmation Model and Citizen Satisfaction with Public Services: A Meta-analysis and an Agenda for Best Practice. *Public Administration Review*, *82*(1), 147–159. https://doi.org/10.1111/puar.13368

Zharif Mustaqim, I., Melani Puspasari, H., Tri Utami, A., Syalevi, R., & Ruldeviyani, Y. (2024). Assessing public satisfaction of public service application using supervised machine learning. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *13*(2), 1608. https://doi.org/10.11591/ijai.v13.i2.pp1608-1618