Jurnal Jaminan Kesehatan Nasional,
Volume 4 Number 2, (December, 2024). Page 176 - 188
DOI : 10.53756/jjkn.v4i2.269
ISSN : 2798-6705 (online)
ISSN : 2798-7183 (print)

Sentiment Analysis of National Health Security Mobile Application Review Using Machine Learning

Deta Novian Anantika Putra

Direktorat Jenderal Kekayaan Negara, Kementerian Keuangan, *e-mail: deta.novian90@gmail.com*

Abstract: It is widely acknowledged that digital transformation provides the opportunity for business process improvements. We can see many businesses, specifically in Indonesia's private and public sectors, leveraging technology to provide better service to their stakeholders. This research seeks insight into stakeholders' engagement in digital transformation in the Indonesian healthcare system, namely the National Health Security Mobile application (Mobile JKN). This study employs a quantitative method to analyze user sentiment from Google Play reviews. Firstly, user reviews are extracted, preprocessing steps are applied, and machine learning-based sentiment labeling is employed afterward to categorize them into positive, neutral, and negative sentiments. The machine labeling process is carried out in document-level analysis, meaning user reviews represent their overall sentiment. Subsequently, we study the most prominent words using Word Cloud to determine the topics mainly discussed in positive and negative sentiments. The result is that 56.10% of reviews from 7 June 2016 to 14 July 2024 contain positive sentiments, while 43.17% contain negative sentiments. Neutral reviews contribute the most minor proportion, making up only 0.73%. The most prominent words in positive sentiment reviews, such as *easy*, *sound*, and *helpful,* suggest the user perception of the National Health Security Mobile as being easy to use and successfully accommodating the user's needs. In contrast, the words application, update, complex, login, code, verification, register, open, use, and error dominate the negative reviews, indicating users had difficulties logging in and registering user accounts, mainly related to frequent updates and Time Password errors.

Keywords: sentiment analysis; logistic models; machine learning; application review

INTRODUCTION

In 2004, Indonesia formed a social security system, namely Jaminan Sosial (Social Security), to ensure that the people of Indonesia could meet their basic needs (Indonesian Regulation No. 14 on National Social Security System, 2004). This social security consists of a) health insurance, b) work accident insurance, c) old-age insurance, d) pension insurance, and e) life insurance. This system includes *Badan Penyelenggara Jaminan Sosial* (BPJS), or Health Social Security Agency, which administers national health insurance programs, from managing social security funds to providing medical care coverage for its participants. This agency has a significant role in Indonesia's healthcare sector, providing coverage for over 273.5 million individuals, which accounts for approximately 97.01% of the country's total population (BPJS Kesehatan, 2024).

Digital transformation has disrupted industries in recent years by breaking down barriers and enabling innovation in products, services, and business processes. (Schwertner, 2017). We can see that the public and private sectors in Indonesia leverage technology to improve stakeholder engagement, particularly in healthcare. This change has improved operational efficiency, telemedicine access, data analysis for personalized treatment, and patient safety. (Syahwali et al., 2023). In 2017, BPJS launched the National Health Security Mobile application (Mobile JKN) to provide a better service.

Mobile JKN is an innovation from BPJS that provides an online platform for healthcare participants to access healthcare services. People can download the application from Google Play or the App Store on their smartphone. Since its launch in 2017, this application has been downloaded over 50 million times. We can use BPJS services, such as registration, data updates, billing information, submitting complaints, and other inquiries related to BPJS services, more quickly than before (Suhadi et al., 2022). This finding is in line with a study from Sari et al. (2024). They have discovered the impact of Mobile JKN in improving the health care service from BPJS. According to them, the application provides functionality that eases the healthcare participants to contribute. Moreover, the reduced number of participants visiting the BPJS office indicates a transition from a traditional site-visit service to a digital service.

Lately, industries have considered sentiment a key measure of success during product launches. (Yildiz et al., 2020). According to *Kamus Besar Bahasa Indonesia* (Sentimen, 2016), sentiment is an opinion or viewpoint that arises from intense emotions. Therefore, Sentiment analysis is “the field of study that analyses people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.” (Liu, 2022). In other words, by analyzing people’s opinions towards products, we can obtain information on how well users perceive and receive the products. In this context, we aim to seek insight into how users’ sentiment towards the Mobile JKN android application, by extracting opinions from application reviews in Google Play.

There are several techniques to delve into the user’s opinion to classify their sentiment: machine learning, lexicon-based, and hybrid (Aqlan et al., 2019). This study will employ one of the methods from the machine learning approach, supervised learning, to detect user sentiment. According to the authors, supervised learning is a technique in machine learning that uses a dataset known as the training set to make predictions. In this dataset, both input data and corresponding response values are present. Supervised learning methods typically use various training documents to learn and improve predictions.

Studies regarding sentiment analysis of Mobile JKN on user reviews have already been done in previous years, resulting in varied conclusions. Kustanto et al. (2021) They utilized the Naïve Bayes classifier algorithm on user reviews in Google Play from 2020 to 3 October 2021 to obtain user perception of the application's service quality. Their research indicates that Mobile JKN version 3 above tends to receive negative sentiment (64.6%), with most complaints related to frequent updates and difficulties when logging into the application.

In contrast to previous studies, Hokijuliandy et al. (2023) Suggest the opposite result. Employing a Support Vector Machine and Chi-Square Feature Selection with a hyperparameter tunning technique, the authors analyzed user review data from Google Play from 1 February 2023 to 20 March 2023, achieving an accuracy of 96.82%. Their result concluded that 69.74% of reviews convey positive sentiments while 30.25% convey negative sentiments. In line with this research, Roiqoh et al. (2023) They also reported comparable results. Their investigation focused on user reviews of Mobile JKN version 4.2.3 and version 4.3.0 taken on 1 February 2023. Utilizing the Indonesia Sentiment (InSet) Lexicon and Naïve Bayes classification algorithm, they discovered that 75.69% of reviews contain positive sentiment, 1.93% contain neutral sentiment, and 22.38% contain negative sentiment. Both studies highlight the most frequent terms discovered regarding difficulties when logging in and registering user accounts.

Compared to the previous research, this paper will consider more algorithms, training-test scenario compositions, and recent user reviews from 7 June 2016 to 14 July 2024. This paper is structured as follows: In the Method section, we discuss the methodology, including data collection, preprocessing, and analysis. The Result section provides the analysis process result, including sentiment classification and user sentiment over time. The Discussion section highlights the dominant topics or keywords in each classification. Lastly, the Conclusion section concludes the paper and suggests the direction for future work.

METHOD

This research uses quantitative methods with steps illustrated in Figure 1. We began by extracting reviews from Google Play using the Google Play-Scraper Python library. The Python package was configured to download reviews from Indonesia, precisely in Bahasa Indonesia, and sorted by the most recent reviews. Two hundred fifty-seven thousand three hundred twenty-three rows of data from 7 June 2016 to 14 July 2024 were successfully collected.

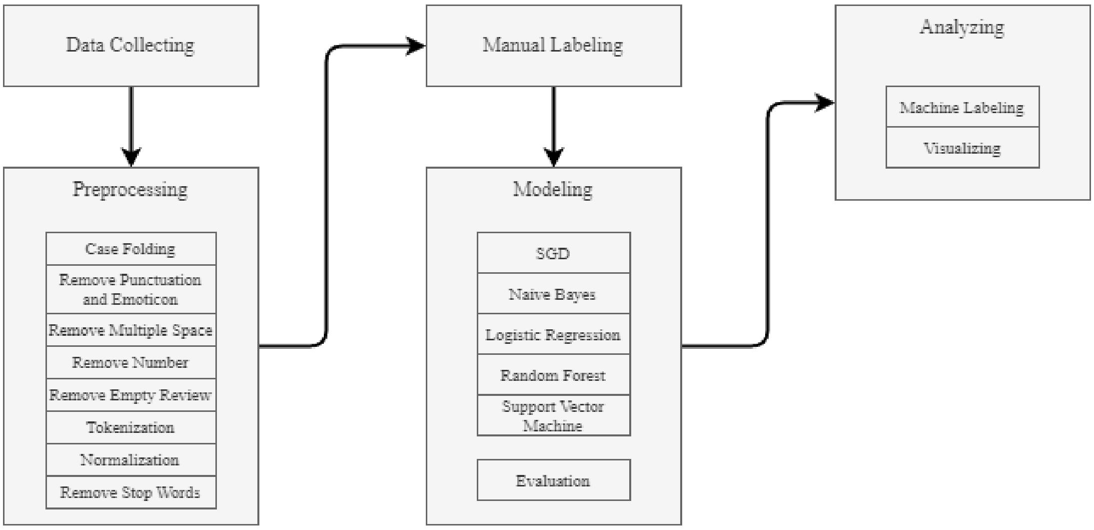

Figure 1 Research Diagram

After the reviews were extracted, several preprocessing techniques were employed to enhance the machine learning algorithm's performance. Preprocessing techniques are commonly applied in natural language processing to prepare text for classification. (Duong & Nguyen-Thi, 2021). Firstly, we tokenized the content and removed the stop words. Subsequently, we converted the content to lowercase and removed punctuations, numbers, emoticons, and multiple whitespaces. The normalization technique was then applied using Peter Norvig's approach. This technique suggests using probability based on Bayes' Theorem to find the correct spelling of all possible candidates. (Norvig, 2016). Any empty data resulting from the preprocessing step was omitted, returning 248,558 clean unlabeled data.

The next step is manual labeling. In this step, we randomly sampled and split 3,000 rows of data, then assigned a class to indicate the sentiment. Three classes were used: "-1" to indicate negative sentiment, "0" to indicate neutral sentiment, and "1" to indicate positive sentiment. The composition of the assigned class is detailed in Table 1.

Table 1 Assigned Class Composition

Assigned Class	Count
-1	1,212
0	159
1	1,629

Before training the algorithm, we applied a word weighting vectoriser technique to determine the importance of keywords within specific documents called the Term Frequency - Inverse

Document Frequency (TF-IDF) method (Qaiser & Ali, 2018). TF measures the frequency of a term's occurrence within a document, while IDF gives less weight to commonly occurring words and assigns more weight to less frequent words. To summarise, the more often the words appear in a document, the higher its weight value. In this research, TF-IDF is calculated using the sci-kit-learn Python library.

After applying the TF-IDF vectorizer technique, we train the dataset using five algorithms: Stochastic Gradient Descent (SGD), Logistic Regression (LG), Random Forest (RF), Naïve Bayes (NB), and Support Vector Machine (SVM). Each algorithm was trained in four distinct scenarios using a labeled dataset from the previous step. Table 2**Error! Reference source not found.** Shows four scenarios employed during the training step to see which algorithms perform best in each scenario.

Table 2 Training and Test Composition Scenario

Scenario	Training	Test
1	90%	10%
2	80%	20%
3	70%	30%
4	60%	40%

RESULT

The algorithm training concludes that the Logistic Regression and Support Vector Machine in Scenario 1 are identical in performance, achieving 91.67% accuracy. Both algorithms also excel in predicting class -1 and class 1. However, when predicting class 0, both algorithms perform very poorly. One possible explanation for this poor performance is the imbalance in the manually labeled dataset. The percentage of positive and negative sentiment in the dataset accounted for 54.3% and 40.4%, respectively, while the neutral sentiment contributes only to a slight proportion, accounting for 5.3%. We justify this imbalance because the primary concern of this research is to study the positive and negative sentiments.

Due to the imbalanced dataset, we use the weighted average of the F-1 score to assist in selecting the algorithm, as the F1-Score metric is considered a better choice for the imbalanced dataset. Although SVM achieves the highest F-1 score in scenario 1 (89.64%), its performance noticeably fluctuates in another scenario in contrast to LR. Since the F-1 score gap in scenario one between LR and SVM is not substantial, we ultimately decided to use Logistic Regression because of its stable performance between scenarios, achieving 89.62% F-1 score and 91.67% accuracy in scenario 1. The comparison of the F-1 score between Logistic Regression and Support Vector Machine is illustrated by Figure 2.

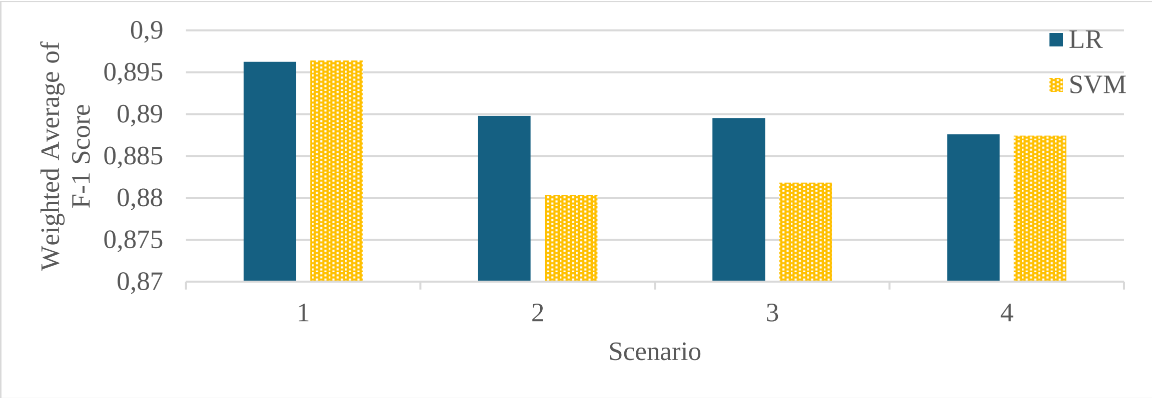

Figure 2 Weighted Average of F1-Score Comparison

Figure 3 Illustrates the performance comparison of our classifier model using a confusion matrix. The matrix reveals 163 true positives, 14 false positives, and four false negatives for class 1. As for class -1's matrix reveals 112 true positives, 11 false positives, and eight false negatives. However, when predicting class 0, the model struggles as no prediction was made. The poor performance when predicting the 0 class indicates the data imbalance on neutral sentiment. Additionally, the model performs very well when predicting class 1 with 0.921 precision, 0.976 recall, and 0.948 F1-Score. Moreover, the model also accurately predicted class -1 with 0.911 precision, 0.933 recall, and 0.922 F1-Score.

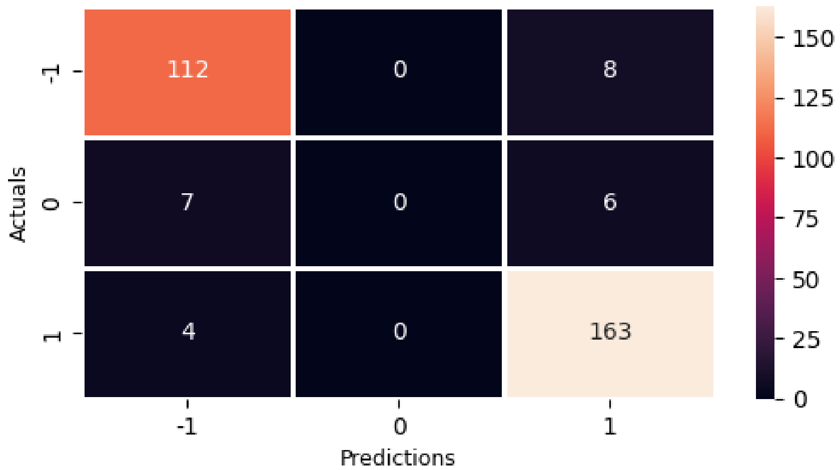

Figure 3 Confusion Matrix of Logistic Regression Classifier Model

After choosing the best-performing algorithm, we automatically employed the model to predict sentiment on 248,558 clean unlabeled data. Table 3 shows the sample of predicted sentiment or class generated from the model. Subsequently, we summarised and analyzed the predicted sentiments to gain insight into their composition and movement over time. We also observed how predicted sentiment correlated with the length of reviews and star ratings to explore user behavior in writing reviews. Word Cloud analysis was then employed to find the most frequent topic mentioned by application users.

Table 3 Predicted Class Sample

No.	Review (Indonesia)	Review (English)	Predicted Class
1.	gak bisa verifikasi nmr tlp udah dicoba berkali kali tetap gak bisa	I cannot verify my phone number. I have tried many times but still can't	-1
2.	aplikasi yang sangat membantu	Very helpful application.	1
3.	mobile jkn sangat membatu pidah alamat maupun pidah pase atu pun yg lainnya poke layanan ini pertahankan siiiiiiiip	Mobile JKN is a helpful application to change address and first medical facility. Keep up the good works!	1
4.	sulit mendaftar antrian selalu muncul tidak bisa memproses	It is hard to register in the queue. It always shows "cannot process".	-1
5.	aksesnya mudah dan fiturnya komplit	Easy to access and it has complete features.	1

The review collected from Mobile JKN Google Play amounted to 257,323 entries from 7 June 2016 to 14 July 2024. After preprocessing and manual labeling, the data was refined to 245,558 entries. Subsequently, the selected algorithm assigned these data a predicted class. Figure 4 It illustrates the composition of all classes predicted by the algorithm. We observed that more than half of the reviews indicate a positive sentiment toward the Mobile JKN application, 56.10% or 137,758 reviews. On the other hand, negative sentiment accounts for 43.17% or 105,596 reviews. Neutral reviews contribute the most minor proportion, only 0.73% or 1,804 reviews.

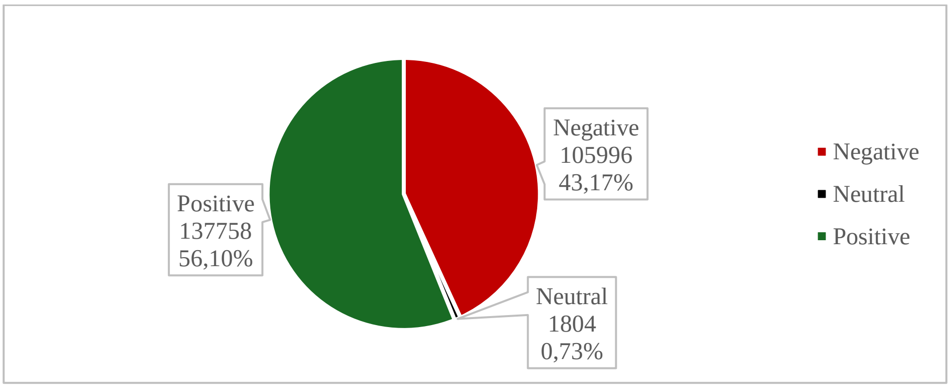

Figure 4 Sentiment Class Composition of User Reviews

We observed the trend of user sentiment during our research period. Figure 5 illustrates the Mobile JKN Android user reviews, categorized by predicted sentiment over time, from 2016 to 2024. The data exhibit variability across the timeline. The first noticeable positive spike was observed around November 2018, followed by a second spike between May and July 2019. The highest positive spike occurred in December 2019, coinciding with the first noticeable negative spike at the beginning of 2020. The negative sentiment continued to dominate until August 2022, after which it began to decline to the present.

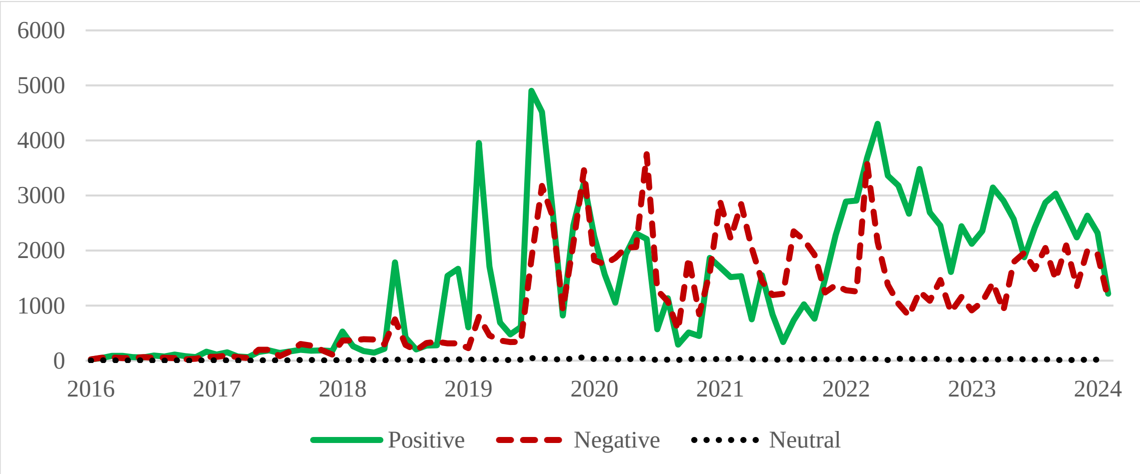

Figure 5 Yearly User Sentiment Trends (2016-2024)

We also used the Pearson correlation coefficient to observe the correlation of user sentiment with application rating and review length. Table 4 The data shows a strong positive correlation between user sentiment and application rating. This suggests that when users give positive reviews, they also provide better application ratings. The data also shows a moderate negative correlation between user sentiment and review length, suggesting users' moderate tendency to write lengthy reviews when feeling unsatisfied with the Mobile JKN.

Table 4 Correlation Matrix Between Predicted Sentiment, Length of Review, and Rating

	Predicted Sentiment	Length of Review	Rating
Predicted Sentiment	1.00	-0.513047	0.816083
Length of Review	-0.513047	1.00	-0.464938
Rating	0.816083	-0.464938	1.00

DISCUSSION

To find the most frequent words from user reviews, we split the machine-labeled dataset by positive and negative sentiment and then utilized a Word Cloud to identify common topics. A Word Cloud is a visual representation of term frequency that displays more common words in larger sizes, making it a simple tool to determine the focus of written material (Atenstaedt, 2017). Each word in the word cloud image shows its significance in the frequency of occurrence. The larger the word, the more common it is.

Figure 6 Shows the most prominent words in positive sentiment reviews, such as mantap, mudah, bagus, and Bantu. These words suggest the user perception of the Mobile JKN as being easy to use and accommodating the user's needs. Some users stressed how easy it is to change the first medical facility using Mobile JKN anytime and anywhere without leaving their houses. In addition, other features are also praised as improvements, such as the ability to add family members, check the invoice, and check user status. This indicates that Mobile JKN has successfully met their expectations.

Figure 6 Word Cloud for Positive Sentiment Review

Our analysis of machine-labeled positive sentiment reviews also revealed some feedback related to further application improvement. Some users emphasized improving the search functionality to locate the nearest medical facility. Despite having their device location turned on, they encountered difficulties using this feature. Additionally, users requested more precise explanations regarding the features of Mobile JKN. Before rolling out updates, some users suggested conducting product reviews to identify issues, such as inaccessible menus or missing membership card displays. Furthermore, interest was expressed in having a version of Mobile JKN optimized for computers or laptops. Lastly, users reported instances where their designated first medical facility was relocated without prior notice.

Figure 7 exhibits the most prominent words in negative sentiment. Compared to positive reviews, we discovered the topics in negative reviews are more diverse, such as *aplikasi, update, susah, login, masuk, kode, verifikasi, daftar, buka, pakai,* and *error.* These words convey how the user has perceived the Mobile JKN as problematic. We structure the issues we have discovered into three categories: 1) difficulties when registering user accounts, 2) issues that hamper the service delivery in general, and 3) some difficulties when using the application.

Figure 7 Word Cloud for Negative Sentiment Review

Firstly, our analysis reveals that users struggle to register their accounts. They reported that the phone number verification process should be taking less time. Some users even claimed they did not receive the One Time Password (OTP) sent via Short Message Service (SMS) or email. OTP-related problems were also reported by the user when changing passwords. One possible explanation is that the user requesting OTP has insufficient credit or balance on their mobile number. Some also noted that the registration process is hampered because Mobile JKN detects their phone number as already registered.

Secondly, many users reported that they could not use the application for several reasons, such as device detection errors, frequent application updates, network quality mismatches, missing family membership cards, captcha challenges, and other unspecified system errors. Some users reported that the application falsely detects their device as emulated or rooted, preventing them from using the service. Additionally, many users expressed frustration due to the frequency of application updates, leading to the feeling of inconvenience. Despite having good network coverage, some users found that the application incorrectly identified their network service as poor. This discrepancy puzzled users and highlighted potential issues in network assessment mechanisms. Similarly, some users reported that the family membership card disappeared from the app and that they had to go to the branch office to seek advice.

Lastly, other users reported successfully using the application in general but questioned some functionality they thought might be counterproductive. Some users questioned the unresponsiveness of the "chat with doctor" feature. Some encountered delayed interactions, impacting their confidence in seeking medical advice through the app. Moreover, some users also encountered frequent incorrect captcha validations despite providing accurate responses. This usability issue highlighted the importance of robust security measures without causing unnecessary user friction. Additionally, a subset of users faced system errors while using the application. Even though this reported error remained unspecified, these possible glitches need further investigation.

We also encountered user feedback related to another service from BPJS Kesehatan, Pandawa (*Pelayanan Administrasi Melalui* WhatsApp or Administrative Service via WhatsApp). According to BPJS Kesehatan (2024), Pandawa is a contactless service channel between front-liners and members using WhatsApp messenger to facilitate administration service. We discovered mixed perceptions during our analysis. On the positive side, many users appreciate the feature of registering and managing family members through Pandawa. Additionally, the service's facilitation of health insurance management without needing site visits has been well-received. Some users find it convenient to update data via WhatsApp, even though others question the necessity of using WhatsApp alongside the Mobile JKN app. On the negative side, users stressed Pandawa's unresponsiveness and inconsistent quality. Despite these challenges, some users acknowledge improvements in service quality. We suggest further researching the Pandawa service since this paper focuses on the Mobile JKN android application.

CONCLUSION

This study demonstrates the effectiveness of using machine learning-based sentiment classification to analyze user reviews from Google Play, with Logistic Regression as the most suitable algorithm. The analysis revealed that most reviews contain positive sentiment, indicating that the National Health Security Mobile application is favorable. However, 43.17% of the negative sentiment, characterized by complaints about login and registration issues primarily due to frequent updates and OTP errors, highlights areas for improvement. The practical implication of this finding is substantial for developers to identify the pain point of the current state of mobile applications. Therefore, focusing the development on the right aspect could significantly enhance the user experience and improve the overall application performance. Additionally, this study contributes to the knowledge of leveraging sentiment analysis to improve user satisfaction.

While this study successfully categorized user sentiment on Mobile JKN review, in general, using logistic regression, there are several directions for future research. Firstly, the dataset is limited to user reviews from Google Play. Expanding the dataset to include user reviews from another platform could provide more comprehensive user sentiment. Secondly, the sentiment analysis in this paper is carried out in document-level analysis. User sentiment in a review represents the sentiment as a whole opinion. Future work could explore more detailed methodologies, such as sentence level and aspect level analysis, to gain more comprehensive insight into specific components of the feedback. Lastly, a more granular investigation of a particular feature of app's effect could also be employed to improve application usability and user experience. Additionally, integrating more advanced machine learning techniques could further refine the analysis.

REFERENCES

Aqlan, A. A., Manjula, B., & Lakshman Naik, R. (2019). A study of sentiment analysis: Concepts, techniques, and challenges. *Proceedings of International Conference on Computational Intelligence and Data Engineering: Proceedings of ICCIDE 2018*, (pp. 147–162).

Atenstaedt, R. (2017). Word cloud analysis of the BJGP: 5 years on. *British Journal of General Practice, 67*, 231–232. https://doi.org/https://doi.org/10.3399/bjgp17X690833

BPJS Kesehatan. (2024, August). *InfoBPJS Kesehatan* (118 ed.). BPJS Kesehatan.

BPJS Kesehatan. (2024). *Layanan Informasi dan Administrasi*. Retrieved August 10, 2024, from Layanan Informasi dan Administrasi. https://www.bpjs-kesehatan.go.id/#/

Duong, H.-T., & Nguyen-Thi, T.-A. (2021). A review: preprocessing techniques and data augmentation for sentiment analysis. *Computational Social Networks, 8*, 1.

Government of Indonesia. (2004). Indonesian Regulation No. 14 on National Social Security System.

Hokijuliandy, E., Napitupulu, H., & Firdaniza. (2023). Application of SVM and Chi-Square Feature Selection for Sentiment Analysis of Indonesia's National Health Insurance Mobile Application. *Mathematics, 11*, 3765. https://doi.org/https://doi.org/10.3390/math11173765

Kustanto, N. S., Yulita, I. N., & Sarathan, I. (2021). Sentiment Analysis of Indonesia's National Health Insurance Mobile Application using Naïve Bayes Algorithm. *2021 International Conference on Artificial Intelligence and Big Data Analytics*, (pp. 38–42).

Liu, B. (2022). *Sentiment analysis and opinion mining.* Springer Nature.

Norvig, P. (2016, 8). *How to Write a Spelling Corrector*. Retrieved 07 30, 2024, from https://norvig.com/spell-correct.html

Qaiser, S., & Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications, 181*, 25–29.

Roiqoh, S., Zaman, B., & Kartono, K. (2023). Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes. *Jurnal Media Informatika Budidarma, 7*, 1582–1592.

Sari, D., Hirdanti, D., Nasution, N. S., & Gurning, F. P. (2024). Syistematic literatur riview: analisis manfaat digital mobile JKN DALAM pembiayaan kesehatan. *ZAHRA: Journal Of Health And Medical Research, 4*, 1–9.

Schwertner, K. (2017). Digital transformation of business. *Trakia Journal of Sciences, 15*, 388–393.

Sentimen. (2016). *Hasil Pencarian: Sentimen*. Retrieved September 21, 2024, from KBBI VI Daring. https://kbbi.kemdikbud.go.id/entri/sentimen

Suhadi, S., Jumakil, J., & Irma, I. (2022). Aplikasi Mobile JKN untuk Memudahkan Akses Pelayanan BPJS. *Jurnal Penelitian Kesehatan" SUARA FORIKES"(Journal of Health Research" Forikes Voice"), 13*, 262–267. https://doi.org/http://dx.doi.org/10.33846/sf13nk148

Syahwali, A. J., Piwari, B., Prabowo, A., & Sutabri, T. (2023). Transformasi Digital Untuk Pengembangan Pelayanan Kesehatan di Rumah Sakit. *IJM: Indonesian Journal of Multidisciplinary, 1*.

Yildiz, E., Safyan, J., & Harper, M. (2020). User Sentiment as a Success Metric: Persistent Biases Under Full Randomization. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (pp. 2891–2899). https://doi.org/https://doi.org/10.1145/3394486.3403340