

Manuscript received May 27, 2023; revised June 20, 2023; accepted June 21, 2023; date of publication July 30, 2023
Digital Object Identifier (DOI): <https://doi.org/10.35882/jeemi.v5i3.304>

Copyright © 2023 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License ([CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).

How to cite: Leader Lawanga Ontshick, Jean-Christophe Mulangu Sabue, Placide Mbala Kiangebeni, Olivier Tshiani Mbayi, Jean-Michel Nsengi Ntamabyaliro, Jean-jacque Muyembe Tamfumu and Rostin Mabela Makengo Matendo, "Comparison of the Performance of Linear Discriminant Analysis and Binary Logistic Regression Applied to Risk Factors for Mortality in Ebola Virus Disease Patients, vol. 5, no. 3, pp. 143–149, July 2023.

Sentiment Analysis on Satusehat Application Using Support Vector Machine Method

Shahmirul Hafizullah Imanuddin¹, Kusworo Adi², dan Rahmat Gernowo²

¹ Faculty of Graduate School, Diponegoro University, Semarang, Indonesia

² Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

Corresponding author: mirulhafiz@students.undip.ac.id

ABSTRACT The lack of analysis regarding public opinion on the use of the SatuSehat application, as well as the absence of research delving deeper into the performance of the Support Vector Machine (SVM) method in classifying user sentiment of the SatuSehat app, creates an opportunity to identify and interpret public sentiment contained within the app's user reviews. This situation demands an effective and efficient method. Therefore, this research aims to determine SVM's performance in classifying SatuSehat app user reviews into positive and negative sentiments while displaying visualizations to identify user reviews' most frequently occurring words. This study is expected to contribute to identifying sentiment patterns among SatuSehat app users, providing information on how satisfied or dissatisfied the public is with this application. Based on the research results, 25,000 data points were divided into 18,359 negative class data and 6,641 positive class data. In the classification phase, SVM produced 73.4% negative and 26.6% positive sentiments. Furthermore, accuracy testing of SVM yielded 91% with a positive sentiment of 92% precision, 71% recall, and 80% F1-score. In contrast, negative sentiment had 90% precision, 98% recall, and 94% F1-score. Visualization results revealed that topics frequently appearing in positive reviews were good, steady, and excellent. Conversely, the negative reviews commonly mentioned words such as update, difficult, strange, log in, and bug.

INDEX TERMS Sentiment Analysis, Support Vector Machine, SatuSehat, Google Play Store

I. INTRODUCTION

Coronavirus, commonly known as COVID-19 (SARS-CoV-2), is a virus that causes a new respiratory illness [1]–[3]. It emerged in late December 2019 in Wuhan, Hubei Province, China [4]–[6]. The first COVID-19 case in Indonesia appeared on March 2, 2020, in Depok, West Java [7]. As the fourth most populous country in the world with an uneven distribution, the impact of COVID-19 will undoubtedly be considerable for Indonesia [8], [9]. One of the government's efforts to tackle the spread of COVID-19 is vaccination [10]–[12].

In implementing the new normal program, the government developed an Android application downloadable from Google Play Store and Apple's App Store called "SatuSehat," previously named "PeduliLindungi" [13], [14]. This app aims to monitor people vaccinated while engaged in activities that require them to be in potentially crowded places [15]. SatuSehat handles the public's vaccination history and, with the help of the QR code scanning feature, records the locations

visited by the users. SatuSehat also provides registration facilities for individuals who have not been vaccinated or those unaware of where to register for vaccination.

Although SatuSehat was launched to facilitate community monitoring in the new normal era, many issues arose due to the app's immaturity. Some problems include the app frequently crashing due to high user numbers, the requirement for an active GPS 24 hours a day, which rapidly drains phone batteries, errors in vaccine recipient data, prolonged certifying issuance, users forced to log in again, and OTPs failing to be sent [16].

Sentiment analysis could be a solution to these problems. Sentiment analysis, also known as opinion mining, is a data analysis process that reflects a person's opinions, behaviours, and emotions towards an entity [17], [18]. This entity could be an individual, event, or a topic currently happening around them. Two types of sentiment analysis commonly used are Fine-grained Sentiment Analysis, which uses specific ratings,

and Emotion Detection, which refers to people's emotions about the subject [19].

In the context of the SatuSehat app, the Support Vector Machine (SVM) method can be used to analyze sentiment. The SVM method is a machine-learning technique for classifying data based on predefined labels. Using SVM, it is possible to identify app user sentiments and emotions regarding various issues they face. SatuSehat developers can better understand user problems and take necessary corrective actions through sentiment analysis and SVM method application. The result is expected to improve app quality and performance so the community can reap more optimal benefits during this new normal era.

Several previous studies on sentiment analysis, such as the research conducted by Mustopa in 2020 on sentiment analysis of the PeduliLindungi app, used two methods: Naïve Bayes and Support Vector Machine (SVM). This research collected data from PeduliLindungi app user reviews on the Google Play Store. This study used the Particle Swarm Optimization (PSO) method combined with both classification methods to obtain higher accuracy. Naïve Bayes and SVM each produced an accuracy of 69% for Naïve Bayes, an AUC value of 0.659, and 93% for SVM, with an AUC value of 0.977. The researcher concluded that SVM was superior at sentiment analysis compared to the NB method [20].

Another study analyzed apps using the SVM and Lexicon-Based approaches [21]. The results showed that implementing Support Vector Machine in Caring Protection classification achieved the highest accuracy with a training and test data distribution of 70:30 at 84.34%, a Precision of 83.72%, a Recall of 93.75%, a Specificity of 67.59%, and an F1 Score of 91.14%. Further research is still being conducted on community monitoring apps during the COVID-19 pandemic, such as PeduliLindungi, with apps from outside Indonesia also being studied. 2021 Ahmad examined Naïve Bayes, Support Vector Machine (SVM), and the Random Forest method. Using 34,534 review data from 46 areas, an average F1-Score of 94.8% indicated that the app is feasible [22]. From the various studies, the most commonly used method for

sentiment analysis processing is the Support Vector Machine (SVM) algorithm. SVM is chosen for its high accuracy, about 80 to 90 per cent, relatively easy implementation, and flexibility, allowing the algorithm to be combined with other methods [23].

In a previous study, sentiment analysis was conducted on the SatuSehat app, focusing on app rating data found on Google Play Store and Apple App Store. Sentiment analysis was performed using Naïve Bayes and Support Vector Machine (SVM) methods. This study concluded that sentiment analysis using SVM was superior, with an accuracy of 93% and an area under the curve (AUC) score of 0.977, compared to Naïve Bayes, with an accuracy of 69% and AUC value of 0.659 [20]. The results were limited due to the use of app rating data from Google Play Store and Apple App Store being non-continuous, and users could only rate once, leaving it unknown whether the SatuSehat app gradually improved or worsened.

Therefore, there is a problem, which is the lack of analysis related to public opinion using the SatuSehat application, as well as the absence of research that delves deeper into the performance of the Support Vector Machine (SVM) method in classifying sentiments of the SatuSehat application users. This situation creates a challenge in identifying and interpreting public sentiment contained in user reviews and requires a more effective and efficient method. Thus, this study aims to determine the performance of the SVM method in assessing public opinions on the use of the SatuSehat application taken from the Google Play Store reviews by classifying the SatuSehat application user reviews into positive and negative sentiments, as well as displaying visualization to determine the most frequently mentioned words from user reviews.

Hence, this research is expected to contribute to identifying user sentiment patterns in the SatuSehat app, providing information on how the community evaluates the app's usefulness based on positive and negative sentiment classification. Furthermore, this research can help SatuSehat app developers with improvements and enhancements to

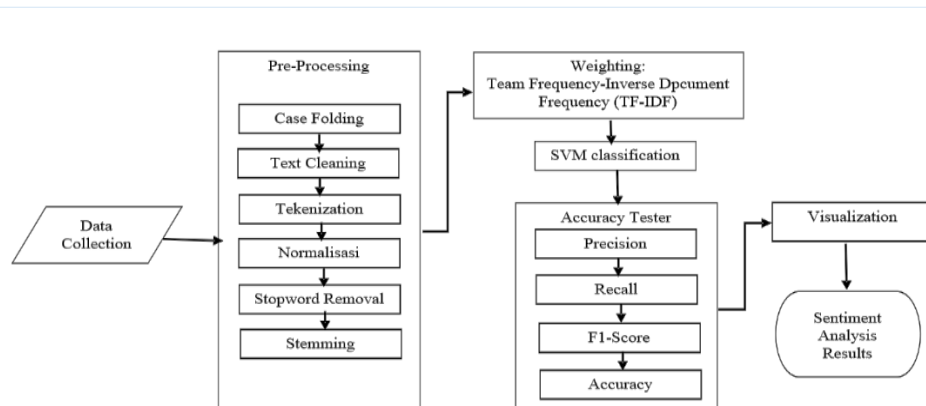


FIGURE 1. Research Framework

achieve optimal user satisfaction. Lastly, the research results may serve as a reference for other researchers who wish to explore further sentiment analysis using the Support Vector Machine or other methods on similar applications.

II. METHODS

The [FIGURE 1](#) presents the research framework flow with the following explanations:

A. DATA COLLECTION

Data collection in this research is taken from user reviews of the SatuSehat application on the Google Play Store conducted within three months, from January 2023 to March 2023. The acquired data totalled 25,000 reviews, with 18,359 negative and 6,641 positive classes. This dataset was then imported to Jupyter Notebook software and underwent several preprocessing steps, including case folding, text cleaning, tokenization, normalization, stopword removal, and stemming. After preprocessing, the dataset was used in the weighting process using the Term Frequency - Inverse Document Frequency (TF-IDF) method to weigh the relationship between words and documents before classifying them into positive or negative sentiment using the Support Vector Machine (SVM) method. The dataset's characteristics include variations in user ratings, ranging from scales 1 to 5, where negative sentiment is associated with scales 1 to 3 and positive sentiment with scales 4 to 5. Additionally, this dataset covers various topics users discuss, such as updates, difficulties, odd entries, bugs, and others, which can be analyzed using Wordcloud. Using this dataset, the research will analyze the accuracy rate in classifying SatuSehat user reviews into positive and negative sentiments.

B. PRE-PROCESSING

Preprocessing is the initial stage to normalize data used in the sentiment analysis process by discarding unnecessary words. Some preprocessing steps are case folding, text cleaning, tokenization, normalization, stopword removal, and stemming:

1) CASE FOLDING

In this step, all letters in the text are converted to lowercase. The purpose is to reduce complexity in processing the text, so uppercase and lowercase letters are not considered different entities.

2) TEXT CLEANING

This process involves the removal of URLs, punctuation, symbols, and usernames (marked as '@username'). The purpose is to eliminate elements that do not contribute to sentiment analysis and reduce the noise level in the data.

3) TOKENIZATION

After cleaning the text, each sentence is broken down into words or tokens. This process separates words into sentences so each word can be individually analyzed.

4) NORMALIZATION

In this stage, slang words or abbreviations are returned to their standard form. This is important because some slang words may have unclear sentiments, and normalization will help clarify the meaning of those words.

5) STOPWORD REMOVAL

Words considered irrelevant or unimportant to sentiment analysis are removed. Stopwords, such as personal pronouns, conjunctions, and question words, have no sentiment value.

6) STEMMING

This stage involves transforming words with affixes (such as prefixes, suffixes, and infixes) into their base form. The stemming process helps reduce complexity in sentiment analysis and identify the underlying sentiment meaning in the text.

C. WEIGHTING

This process calculates how many words appear in a document. The method used is TF-IDF for the weighting process in the research to be carried out.

D. SVM CLASSIFICATION

This process is carried out to group labels obtained from the weighting process. This grouping is divided into two classes, positive and negative.

E. ACCURACY TESTING

In this process, accuracy testing is used to generate how accurate research results are with precision, recall, f1-score, and accuracy.

F. SENTIMENT ANALYSIS RESULTS

This process is the final stage, the result of the fundamental research conducted.

G. VISUALIZATION

This process extracts information frequently discussed by SatuSehat application users in the Google Play Store review column. Therefore, in the context of training and testing, this research goes through several crucial steps. First, the collected data will be divided into two parts: training data and testing data. Training data are labelled data with "positive" and "negative" labels, while the testing data do not have such labels. After obtaining the training and testing data, the next step is to perform preprocessing on that data. This preprocessing process involves case folding, text cleaning, tokenization, normalization, stopword removal, and stemming from preparing the data before term weighting.

Then, the data obtained through the preprocessing stage will be used for term weighting using the TF-IDF method. After term weighting is completed, the next step is to train the classification model. This research uses Support Vector Machine (SVM) as the classification algorithm, which will involve learning from the training data. The trained model is then tested on the testing data to evaluate the extent to which the model can classify new data. Subsequently, the classification results are evaluated using a confusion matrix and calculating evaluation metrics such as precision, recall, f1-score, and accuracy. The evaluation results analysis will be carried out to draw conclusions and suggestions related to the classification model's performance. The equations used related to this research are: the equation for linear SVM kernel:

$$K(x,y) = x \cdot y \quad (1)$$

This equation is used in linear SVM classification to calculate the kernel function between two vectors, x and y . The linear kernel measures the similarity between two input vectors in

the same feature space. The equation for non-linear SVM kernel:

$$K(x,y) = \exp(-\|x-y\|^2 / (2\sigma^2)) \quad (2)$$

This equation is used in non-linear SVM classification to calculate the kernel function between two vectors, x and y using the exponential (Gaussian) kernel function. The non-linear kernel enables SVM to classify data that cannot be linearly separated by transforming it into a higher-dimensional feature space. The parameter σ in this equation controls the width of the Gaussian kernel and can be adjusted to produce an optimal classification.

III. RESULTS

A. DATA COLLECTION

Data collection involves gathering data from user reviews of the SatuSehat app on the Google Play Store. The collected data will then be imported into the Jupyter Notebook software; In this research process, the rating data is also converted into sentiments that can be seen in TABLE 1.

TABLE 1
Results of changing ratings into sentiments

Text	Score	Label
Cannot Enter OTP Code While Logging In Due to Continuous Loading.	1	Negative
This application should be simpler, but instead, it's getting more complicated. The process of downloading the vaccination proof should be made easier, not like an amateur application.	2	Negative
The OTP code still cannot be sent via SMS, it must be sent through WhatsApp. Please make the OTP code available through SMS. It's a pity for those who don't have the WhatsApp application. Thank you and wish you continued success.	3	Negative
Great, very helpful.	4	Positive
Thank you, the vaccination certificate can now be downloaded in the latest version of the app.	5	Positive

TABLE 1 shows that sentiment on a scale of one to three is considered negative, while positive sentiment is on a scale of four to five.

B. PRE-PROCESSING

Stages in pre-processing involve further processing of the raw data collected to clean the data, which includes removing noise, clarifying features, and modifying the raw data to suit research needs. The pre-processing steps are illustrated in FIGURE 2.

1) CASE FOLDING

The purpose of the case folding step is to change the letters previously mixed between uppercase and lowercase letters to lowercase letters only. The case folding stage in this research can be seen in TABLE 2.

TABLE 1
Case folding

Before	@SATUSEHAT I'm confused!! just a picture of the application. There are no commands or directions from the application. https://t.co/vskqzhujpn
After	@satusihat I'm confused!! just a picture of the application. There are no commands or directions from the application. https://t.co/vskqzhujpn.

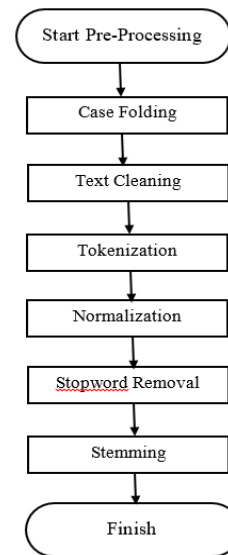


FIGURE 2. Preprocessing Stages

2) TEXT CLEANING

This stage aims to clean sentences from panels or hyperlinks, punctuation marks, mentions of usernames (@username), URLs, and numbers that should not be present in the dataset. The text-cleaning process in this study can be seen in Table 3.

TABLE 3
Text cleaning

Before	@satusihat I'm confused!! just a picture of the application. There are no commands or directions from the application. https://t.co/vskqzhujpn
After	satusihat, I'm confused. Only the picture of the application. There are no instructions or directions from the application.

3) TOKENIZATION

At this stage, all collected words will have their punctuation removed, such as symbols, characters, and anything that is not a letter. The aim is to separate sentences into words. The tokenizing process in this research can be seen in Table 4.

TABLE 4
Proses tokenizing

Before	SatuSehat, I'm confused. Only the picture of the application. There are no instructions or directions from the application.
After	['satusihat', 'I am', 'confused', 'only', 'the picture', 'of the application', 'itself', 'there is', 'no', 'any', 'instruction', 'or', 'direction', 'from', 'the application', 'itself']

4) NORMALIZATION

The normalization process converts words that were initially abbreviations or slang into standard words. The stages of normalization can be seen in Table 5.

TABLE 5
Normalization

Before	['satusihat', 'I am', 'confused', 'only', 'picture', 'application', 'itself', 'only', 'there is', 'no', 'any', 'instruction', 'or', 'direction', 'from', 'application', 'itself']
After	['satusihat', 'I am', 'confused', 'only', 'picture', 'application', 'itself', 'only', 'there is', 'no', 'any', 'instruction', 'or', 'direction', 'from', 'application', 'itself']

5) STOPWORD REMOVAL

At this stage, conjunctions or irrelevant words, such as first, second, and third-person pronouns, names, conjunctions, and question words, will be removed. These words do not have meaning if separated from other words and are not related to the adjectives associated with the sentiment, as seen in Table 6.

TABLE 6
Stopword removal process

Before	['satusihat', 'I am', 'confused', 'only', 'picture', 'application', 'itself', 'only', 'there is', 'no', 'any', 'instruction', 'or', 'direction', 'from', 'application', 'itself']
After	['I am', 'confused', 'only', 'picture', 'application', 'itself', 'only', 'there is', 'no', 'any', 'instruction', 'or', 'direction', 'from', 'application', 'itself']

6) STEMMING

At this stage, the goal is to change the word form into its basic form by the structure of the Indonesian Dictionary, such as words with initial and final affixes like saya, mem, meny, meng, di, per, ber, an, kan, i, nya, etc. These affixed words will be transformed into their basic form using Python, as shown in TABLE 7.

TABLE 7
Stemming process

Before	['confused', 'only', 'picture', 'application', 'its', 'only', 'instruction', 'direction', 'application', 'its']
After	['confuse', 'only', 'picture', 'app', 'only', 'instruct', 'direct', 'app']

C. WEIGHTING

In this phase, the weighting will be conducted using the Term Frequency - Inverse Document Frequency (TF-IDF). The purpose of TF-IDF is to weigh the relationship between a word (Term) and a document before calculating the algorithm and overcoming the problem of classifying data into positive or negative sentiments. The stages of TF-IDF that have been carried out can be seen in TABLE 8.

TABLE 8
TF-IDF

Document	TF*IDF	Label
['download', 'give', 'accept', 'vaccine', 'certificate', 'application']	[0.5681, 0.4553, 0.4814, 0.2989, 0.3206, 0.2147]	Positive
['good', 'help']	[0.6432, 0.7657]	
['pity', 'success', 'whatsapp', 'SMS', 'please', 'OTP', 'code', 'use', 'give', 'receive', 'application']	[0.3049, 0.2898, 0.4608, 0.4337, 0.1684, 0.4064, 0.3676, 0.1468, 0.1759, 0.1860, 0.0829]	Negative
['amateur', 'evidence', 'process', 'complicated', 'simple', 'rich', 'download', 'open', 'vaccine', 'easy', 'application']	[0.4422, 0.3259, 0.3326, 0.3672, 0.3758, 0.3000, 0.2897, 0.1213, 0.1524, 0.2277, 0.2190]	
['otp', 'code', 'login', 'constantly', 'loading', 'enter']	[0.4576, 0.4138, 0.2714, 0.2714, 0.5323, 0.2626]	

D. SVM CLASSIFICATION

In this process, the data that has gone through the weighting stage will be classified into two classes, namely the positive and negative classes, with the condition that if the weight > 0, it will be included in the positive class. In contrast, a weight > 0 will be included in the negative class. FIGURE 3 in the percentage graph shows SVM classification in this study.

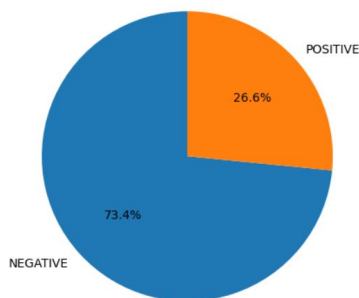


FIGURE 3. Percentage Chart

FIGURE 3 shows the results of the SVM classification presentation. From the research findings, the sentiment analysis produced more negative sentiments, amounting to 73.4%, compared to positive sentiments, which were only 26.6%.

E. ACCURACY TESTING

Based on the testing results obtained, this study produces accuracy, precision, recall, and f1-score, which can be seen in TABLE 9 below:

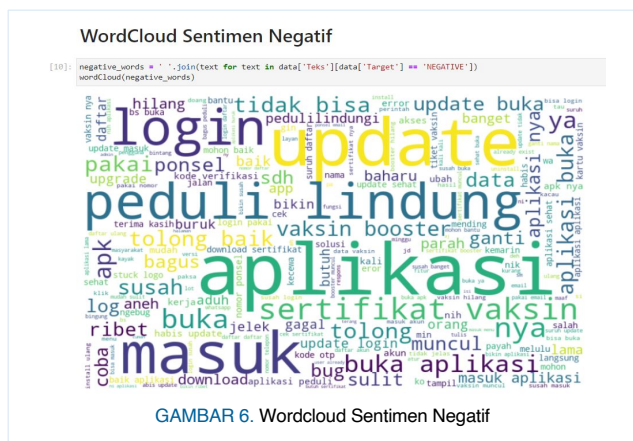
TABLE 9
Results Accuracy, Recall, Precision And F1-Score

	Precision	Recall	F1-Score	Support
Negative	0,90	0,98	0,94	14608
Positive	0,92	0,71	0,80	5382
Accuracy			0,91	19990
Macro avg	0,91	0,84	0,87	19990
Weighted avg	0,91	0,91	0,90	19990

The table indicates that the test results of the Support Vector Machine (SVM) classification method achieved an accuracy of 91%, with positive sentiment having a precision rate of 92%, a recall rate of 71%, and an F1-score of 80%. Meanwhile, negative sentiment has a test precision of 90%, a recall rate of 98%, and an F1-score of 94%.

F. VISUALIZATION

The visualization process is a process that aims to extract information in the form of topics frequently discussed by users of the SatuSehat application. Therefore, the most important information will be taken from the numerous texts, namely the existing reviews. In this study, the visualization of the classification results will be displayed using Wordcloud in



GAMBAR 6. Wordcloud Sentimen Negatif

FIGURES 5 and FIGURE 6. The word cloud visualization in FIGURE 5 clearly shows an overview of the topics and positive words frequently used by the users of the SatuSehat app to provide reviews. In Wordcloud, the larger the size of the word displayed, the more often the word is used by SatuSehat app users as a conversation topic, among others. The word cloud visualization in FIGURE 6 clearly shows the topics and negative words frequently used by SatuSehat app users to provide reviews. In the word cloud, the larger the size of the word displayed, the more often the word is used by SatuSehat app users as a topic of conversation. SatuSehat app users often discuss updates, difficulties, strange entries, bugs, cannot, and so on.

IV. DISCUSSION

In this study, the classification of the SVM method achieved an accuracy of 91%. This indicates that the accuracy of this method can be considered good. On the other hand, the results of this study show that negative sentiment from users of the SatuSehat app is higher (73.4%) than positive sentiment (26.6%). The classification results are then visualized using a word cloud. With the word cloud, the most frequently appearing words from user reviews of the SatuSehat app on the Google Play Store were found. In this study, the words application, certificate, and vaccine appear in positive and negative reviews. Positive User Ratings include the words 'easy' and 'access,' meaning vaccine certificates are easily accessible. In addition, there are also words like 'good,' 'great,'

'nice,' and 'benefits.' Therefore, the positive evaluation of the SatuSehat app should be optimally maintained and its quality improved.

On the other hand, there is negative feedback from users, such as 'login,' 'difficult,' and 'sign in,' meaning that some users have difficulty accessing the SatuSehat app. Also, the words 'code,' 'OTP,' and 'loading' were mentioned, causing users to be unable to enter the OTP code while logging in. Users also used the words 'error' and 'bug' to describe their experiences using the SatuSehat app, which often crashes. In the negative evaluations, there is also the word 'not appearing,' meaning the vaccine certificate does not appear, as well as other words like difficult, complicated, and strange.

Therefore, an evaluation and improvement plan can be carried out to enhance the quality of the SatuSehat app. The evaluation should consider the problems expressed by users, such as difficulties during login, loading issues, and the non-appearance of vaccine certificates. The necessary improvements include increasing the app's stability, fixing frequently appearing bugs and errors, and improving the responsiveness and speed of the app. In this regard, word clouds can be useful for analyzing the most frequently appearing words in user reviews. By analyzing these words, app developers can focus on aspects that require improvement, such as ease of access, quality of service, and the overall reliability of the app.

In addition, referring to previous studies, the current study achieved a classification accuracy of 91% using the Support Vector Machine (SVM) method to analyze user sentiment of the SatuSehat app, which is considered good and higher compared to studies using the Naïve Bayes method, with an accuracy of 69% [20]. This study also used user text review data, providing more detailed information about their experiences using the app, unlike previous studies that only used non-continuous and limited app rating data. Compared to previous studies, the advantage of this study is the use of the SVM method with higher accuracy and the use of text review data that is more informative in obtaining user feedback. Analyzing frequently occurring words in the reviews using word cloud is also an advantage that can help app developers better understand user needs and problems, making improvement efforts more targeted and accurate.

By optimizing positive ratings and addressing user issues in negative reviews, the SatuSehat app can improve and satisfy users in the future. App developers should actively involve users in the improvement and development process. It is also crucial to take steps to improve the security of the SatuSehat app. Maintaining user privacy and data security will be key to building trust and user satisfaction. In this context, developers should ensure that any personal information collected through the app is protected and not misused [24], [25].

However, this study also has limitations. Although an accuracy of 91% has been achieved, there is a significant difference between the positive and negative sentiments of the SatuSehat app users. This indicates that evaluation and app improvements still need improvement to improve app quality

and user satisfaction. This limitation suggests that this study could be further developed to be more effective in providing improvement recommendations to app developers.

One limitation of this study is the limited data source, which only comes from the Google Play Store, thus neglecting data from users using the iOS platform. This may affect the conclusions drawn from the sentiment analysis conducted. As the amount of review data used is still limited and related to the SatuSehat app, future research should continuously provide data regularly and expand the data source of various health-related app reviews to help address large-scale pandemics more effectively. Moreover, user review data from other platforms like the Apple App Store can also be included in the analysis to provide a more comprehensive overview of user sentiment towards the SatuSehat app.

IV. CONCLUSIONS

This study uses a dataset of 25,000 records, divided into 18,539 negative and 6,641 positive class instances, obtained from user reviews of the SatuSehat app on the Google Play Store. Sentiment classification is performed using the Support Vector Machine (SVM) method, and the results divide the reviews into two classes: positive and negative. The test results show that public sentiment regarding the SatuSehat app tends to be less favorable, with a negative sentiment percentage of 73.4% and a positive sentiment percentage of 26.6%. The testing accuracy obtained is 91%. These findings highlight the importance of improving the quality of the SatuSehat app for the company to increase user satisfaction in the future. As a recommendation for further research, involving data from other platforms like the Apple App Store and social media can provide more comprehensive sentiment analysis and present a more complete picture of the user experience with the SatuSehat app.

V. REFERENCES

- [1] S. J. Daniel, "Education and the COVID-19 pandemic," *Prospects*, vol. 49, no. 1–2, pp. 91–96, 2020, doi: 10.1007/s11125-020-09464-3.
- [2] S. Hoskins et al., "Relative contribution of essential and non-essential activities to SARS-CoV-2 transmission following the lifting of public health restrictions in England and Wales," *Epidemiol. Infect.*, vol. 151, no. 3, pp. 1–8, 2023, doi: 10.1017/S0950268822001832.
- [3] M. Manica et al., "Estimation of the incubation period and generation time of SARS-CoV-2 Alpha and Delta variants from contact tracing data," *Epidemiol. Infect.*, vol. 151, no. 3, pp. 1–8, 2023, doi: 10.1017/S0950268822001947.
- [4] S. Moslehi, A. Dehghani, G. Masoumi, and F. Barghi Shirazi, "Vulnerability Management of the Elderly During COVID-19 Pandemic: A Systematic Review," *Heal. Emergencies Disasters Q.*, vol. 8, no. 2, pp. 77–86, 2023, doi: 10.32598/hdq.8.2.310.3.
- [5] C. Puchongmart et al., "Clinical factors associated with adverse clinical outcomes in elderly versus non-elderly COVID-19 emergency patients: a multi-center observational study," *Int. J. Emerg. Med.*, vol. 16, no. 1, pp. 1–12, 2023, doi: 10.1186/s12245-023-00482-4.
- [6] K. Yuki, M. Fujioji, and S. Koutsogiannaki, "COVID-19 pathophysiology: A review," *Clin. Immunol.*, vol. 1, no. 1, pp. 1–8, 2020, doi: 10.1016/j.clim.2020.108427.
- [7] V. Philips and T. Y. Wicaksono, "Karakter dan Persebaran Covid-19 di Indonesia," *CSIS Comment.*, vol. 9, no. 4, pp. 1–12, 2020, [Online]. Available: <https://mail.csis.or.id/publications/karakteristik-dan-persebaran-covid-19-di-indonesia-temuan-awal/>
- [8] E. Fadlyana et al., "Immunogenicity and safety in healthy adults of full dose versus half doses of COVID-19 vaccine (ChAdOx1-S or BNT162b2) or full-dose CoronaVac administered as a booster dose after priming with CoronaVac: a randomised, observer-masked,

- controlled trial in I," *Lancet Infect. Dis.*, vol. 23, no. 5, pp. 545–555, 2023, doi: 10.1016/S1473-3099(22)00800-3.
- [9] S. Olivia, J. Gibson, and R. Nasrudin, "Indonesia in the Time of Covid-19," *Bull. Indones. Econ. Stud.*, vol. 56, no. 2, pp. 143–174, 2020, doi: 10.1080/00074918.2020.1798581.
- [10] Yuliska, N. F. Najwa, and K. U. Syaliman, "the Implementation of Deep Learning Techniques in Developing Conversational Chatbot As the Source of Vaccination Information," *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 1, pp. 532–538, 2022, doi: 10.37385/jaets.v4i1.1340.
- [11] S. Koesnoe et al., "Using Integrative Behavior Model to Predict COVID-19 Vaccination Intention among Health Care Workers in Indonesia: A Nationwide Survey," *Vaccines*, vol. 10, no. 5, pp. 1–13, 2022, doi: 10.3390/vaccines10050719.
- [12] A. Tavilani, E. Abbasi, F. Kian Ara, A. Darini, and Z. Asefy, "COVID-19 vaccines: Current evidence and considerations," *Metab. Open*, vol. 12, no. 9, pp. 1–5, 2021, doi: 10.1016/j.metop.2021.100124.
- [13] M. I. Nurmansyah, C. Rosidati, Yustiyani, and N. M. Nasir, "Measuring the Success of PeduliLindungi Application Use for Supporting COVID-19 Prevention: A Case Study among College Students in Jakarta, Indonesia," *Kesmas*, vol. 17, no. 1, pp. 11–16, 2022, doi: 10.21109/kesmas.v17isp1.6057.
- [14] M. Shahroz et al., "COVID-19 digital contact tracing applications and techniques: A review post initial deployments," *Transp. Eng. J.*, vol. 5, no. 1, pp. 1–10, 2021, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8132499/pdf/main.pdf>
- [15] R. Chatterjee, S. Bajwa, D. Dwivedi, R. Kanji, M. Ahammed, and R. Shaw, "COVID-19 Risk Assessment Tool: Dual application of risk communication and risk governance," *Prog. Disaster Sci.*, vol. 7, no. 8, pp. 1–11, 2020, doi: 10.1016/j.pdisas.2020.100109.
- [16] G. Priscila and P. Robin, "Pengaisan Big Data & Dunia Kesehatan," *Public Corner*, vol. 16, no. 1, pp. 37–57, 2021, doi: 10.24929/fisip.v16i1.1472.
- [17] S. Huang, X. Wu, X. Wu, and K. Wang, "Sentiment analysis algorithm using contrastive learning and adversarial training for POI recommendation," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, 2023, doi: 10.1007/s13278-023-01076-x.
- [18] M. F. Karim, R. Rahutomo, I. B. K. Manuaba, K. Purwandari, T. N. Mursitama, and B. Pardamean, "Free trade as domestic, economic, and strategic issues: a big data analytics approach," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00722-7.
- [19] A. H. Alamoodi et al., "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review," *Expert Syst. Appl.*, vol. 167, no. 1, pp. 1–13, 2020, doi: <https://doi.org/10.1016/j.eswa.2020.114155>.
- [20] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini, and D. Risdiansyah, "Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm Based on Particle Swarm Optimization," in *International Conference on Informatics and Computing (ICIC)*, Gorontalo, 2020, pp. 1–7. doi: 10.1109/ICIC50835.2020.9288655.
- [21] A. Fathin, "Analisis sentimen terhadap ulasan aplikasi mobile menggunakan metode support vector machine (SVM) dan pendekatan lexicon based," *Universitas Islam Negeri Syarif Hidayatullah*, 2022. [Online]. Available: <https://repository.uinjkt.ac.id/dspace/bitstream/123456789/65009/1/AHMAD FATIHIN-FST.pdf>
- [22] K. Ahmad et al., "Sentiment Analysis of Users' Reviews on COVID-19 Contact Tracing Apps with a Benchmark Dataset," *Comput. Lang.*, pp. 1–11, 2021, [Online]. Available: <http://arxiv.org/abs/2103.01196>
- [23] Y. Al Amrani, M. Lazaar, and K. E. El Kadirp, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Comput. Sci.*, vol. 127, no. 1, pp. 511–520, 2018, doi: 10.1016/j.procs.2018.01.150.
- [24] Y. Mahendradhata, T. Lestari, and R. Djalante, "Strengthening government's response to COVID-19 in Indonesia: A modified Delphi study of medical and health academics," *PLoS One*, vol. 17, no. 9, pp. 1–20, 2022, doi: 10.1371/journal.pone.0275153.
- [25] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini, and D. Risdiansyah, "Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm Based on Particle Swarm Optimization," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 2020, pp. 1–7. doi: 10.1109/ICIC50835.2020.9288655.