



Sentiment Analysis on PeduliLindungi Application Using TextBlob and VADER Library

F Illia, M P Eugenia, S A Rutba

Politeknik Statistika STIS, Jalan Otto Iskandardinata No. 64C East Jakarta, DKI Jakarta

*Corresponding author's e-mail: 221911038@stis.ac.id, 221910737@stis.ac.id, 221910717@stis.ac.id

Abstract. The Covid-19 virus has become a global pandemic, including Indonesia. Various efforts have been made by the Government to reduce the negative impact by this pandemic, one of which is through the PeduliLindungi application. The research was conducted to obtain public sentiment towards the application by using twitter data. The data collection period is from August 31 to September 7, 2021, this period was chosen due to the emergence of news regarding vaccine data leaks associated with data leaks in the PeduliLindungi application. Sentiment analysis is carried out using the TextBlob and VADER libraries. The results of this sentiment analysis are sufficient to display public opinion and it is hoped that decision makers can improve applications based on these opinions. Then, it was found that the VADER library can be said to be better in conducting sentiment analysis in research because the lexicon approach used is based on social media.

1. Introduction

The Covid-19 virus has become a pandemic in the whole world, including Indonesia. The World Health Organization (WHO) urges the public to implement health protocols to avoid the spread of Covid-19 by maintaining a minimum distance of one meter, wearing masks, washing hands, limiting mobility, and staying away from crowds. The Indonesia Government released policies such as social distancing which were carried out in the year 2020. In addition, the government through the Covid-19 Task Force has launched four strategies to reinforce the physical distancing policy. The first is the use of a mask, the second is the activity to search contacts (tracing) who identified infected Covid-19 with the rapid test, the third is education and preparation of self-isolation, and the latter is the isolation in the hospital when cases who infected Covid-19 requires hospital clinic services [1].

Activity to search contacts according to WHO is the activity of monitoring the identification of people that have a possibility of exposure to Covid-19. Contact is defined as someone who interacts within a distance of 1 meter in a time less than 15 minutes with someone who is infected Covid-19. Contacts are required to perform self-isolation in a period of time of about 14 days to reduce the possibility of contact spreading the Covid-19 virus to others. According to WHO, electronic devices and information technology are able to assist contact tracing activities on a large scale [2]. Based on that statement, the government through the collaboration between the Ministry of Communication and Information, Health, Agency Enterprises Owned, and the National Board for Disaster Management, as well as operators of telecommunications that exist in Indonesia makes an application called PeduliLindungi. This application can help people do the tracing, tracking, and fencing against people



who are infected Covid-19. The PeduliLindungi application will identify users who are in the range that is close to people who are infected and who are identified to be infected to Covid-19 [3].

To use the PeduliLindungi application, name, cell phone number, and geographic location data are needed [4]. After the Covid-19 vaccination, the public will be given a vaccine certificate with a QR code listed that can be downloaded in the PeduliLindungi application. Vaccine certificates store data such as population registration numbers [20]. Meanwhile, the QR Code on the vaccine certificate stores the vaccine registration number [19]. These data must be kept confidential and should not be disseminated for fear of being misused by irresponsible people [20,21]. However, on September 3, 2021, public panic arose because there was news about vaccine data being leaked [6]. Regardless of the emergence of the news, from the field of law, there is the term called the principles of data personal as contained in the General Data Protection Regulation (GDPR). According to research from [7], the PeduliLindungi application violates the principles of personal data which include the principles of transparency, goal limitation, and data minimization. While The National Cyber and Crypto Agency stated that the PeduliLindungi application was in accordance with the principles of personal data protection and security and was in accordance with Government Regulation No. 71 of 2019 Article 14 Paragraph 1 and Article 15 concerning the Implementation of Electronic Systems and Transactions [8]. Outside the negative news, research that analyzes the usability of the PeduliLindungi application indicates that the application provides information that is in accordance with the users' expectations [9]. In addition, based on research on public acceptance of PeduliLindungi application stated that the ease of use and usability of the application is a factor that is important for the users [10]. The PeduliLindungi application has its negative and positive sides in the eyes of the public. After seeing the incident that occurred on September 3, it is necessary to analyze the public response about the PeduliLindungi application before and after the news came out. So, the PeduliLindungi application can be improved based on those responses.

2. Research Method

The method used in this study is text mining with word cloud formation and sentiment analysis. Text mining is the process of finding information hidden from text-shaped data by extracting information from a written source. Word cloud is a visual representation of text. Word cloud makes it easy for readers to see the frequency of occurrence of a text word. The larger the font size in word cloud, the more often the word appears. Sentiment analysis is the process of understanding, extracting, and processing data in the form of text to obtain sentiment information contained in it.

This study used data derived from Indonesian-language tweets with the keyword PeduliLindungi in the period 31 August 2021 to 7 September 2021. The data before preprocessing was obtained amounting to 19,649 tweets. The data is collected using *tweepy* integrated with the Twitter API. After the data is collected, then preprocessing is done before the data analysis is done.

The research steps conducted by researchers consist of several stages as seen in **Figure 1**.

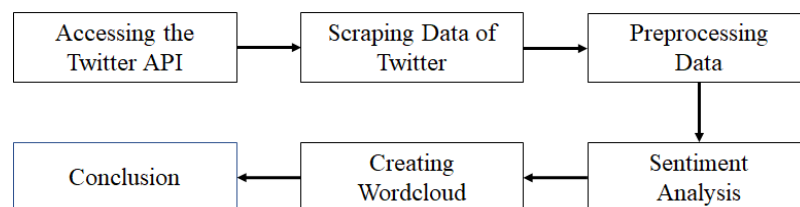


Figure 1. Research Framework Flow Chart

The preprocessing stage is the process of extracting interesting and important knowledge from unstructured text data [22]. Before preprocessing, the retweet will be removed first. The stages of text preprocessing are case folding, normalization, tokenizing, stopword removal, and stemming. Case folding is the process of converting letters in text into lowercase letters [14]. After that, words that are not standard and are not in the KBBI will be normalized manually into words that have the closest meaning. Tokenizing is the process of cutting text based on each of its constituent words. Then,



filtering is done by doing stop words removal. Stop words removal aims to remove common words that do not have a significant influence in tweets[14]. Finally, stemming is done to remove the word add-on in the tweet.

The sentiment analysis process is carried out with a lexical approach. In this study, two sentiment analysis methods were used in python, namely TextBlob and VADER (Valence Aware Dictionary and Sentiment Reasoner). TextBlob is a python library for processing textual data[16]. Sentiment analysis using TextBlob and VADER is only available in English so researchers translate data from preprocessing results into English before sentiment analysis.

TextBlob offers APIs for natural language processing (NLP) such as part-of-speech tagging, noun extraction, sentiment analysis, classification, translation, and more. Sentiment analysis in TextBlob is done by restoring two properties, namely polarity, and subjectivity[19]. Polarity is a float value that ranges from -1 to 1. A polarity value of -1 indicates negative sentiment, while a polarity value of +1 indicates positive sentiment. Subjectivity is described with values ranging from 0 to 1. If subjectivity is less than 0.5, then sentences tend to be subjective rather than objective.

VADER is a lexicon and rule-based sentiment analysis tool. VADER is open-source under an MIT license developed by George Berry, Ewan Klein, and Pier Paolo. VADER provides positive, negative, and neutral labeling based on text sentiment calculations according to its semantic orientation[18]. VADER maintains the benefits of traditional lexicon sentiments such as linguistic inquiry and word (LIWC)[16]. VADER returns the probability value of the given input and classifies it into positive, neutral, and negative sentiments.

3. Result

The research began by crawling Twitter data using the keyword PeduliLindungi on August 31, 2021 to September 7, 2021. From the results of crawling the data and preprocessing data, there are 9,820 tweets. The following **Figure 2** shows the distribution of the tweets.

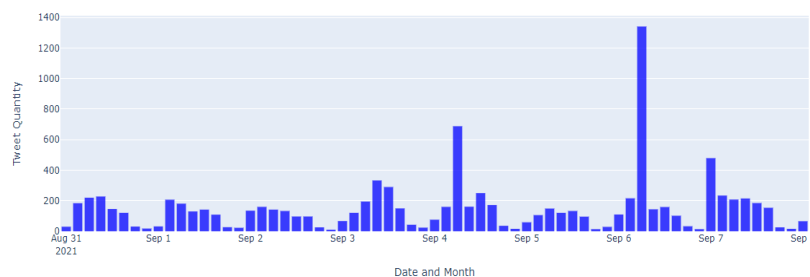


Figure 2. Tweet Distribution.

Figure 2 shows an increasing number of tweets that were previously constant. This change began to occur on 3-4 September 2021, this change was allegedly because on that date, news related to the alleged leak of Indonesian people's personal data related to vaccinations, especially those that happened to high-ranking officials in Indonesia became a trending topic on twitter, so the PeduliLindungi platform became a trending topic. one of the applications in the vaccination program that was carried out became the most talked about topic on twitter. This is due to the alleged data leak originating from the PeduliLindungi application because it contains the same information. Therefore, twitter data collection was carried out on 31-7 September 2021 to capture how the public reacted and assessed.

This study divides the data into two, namely tweets on 31 August-3 September 2021 and 4-7 September 2021. From the results of the division of the period, it was found that the number of tweets was 3,884 and 5,936 respectively. Then, sentiment analysis was carried out on tweets generated in both periods by using two algorithms respectively in the TextBlob and VADER libraries.



Figure 3. Positive Sentiment Wordcloud on TextBlob Library on 31 August-3 September 2021



Figure 4. Positive Sentiment Wordcloud on TextBlob library on 4-7 September 2021

Figure 3 and **Figure 4** are visualizations of the words that appear most often in tweets classified as positive sentiment on the algorithm in the TextBlob library. In the first period, words such as #benar, #wajib, #aman were some of the words that most often emerged from the results of people's tweets that actually had positive meanings. Then for the second period, #benar, #guna, #kontrol were some of the many words that most often appear from the results of people's tweets that really have a positive meaning.



Figure 5. Positive Sentiment Wordcloud on VADER Library on 31 August-3 September 2021



Figure 8. Negative Sentiment Wordcloud on TextBlob Library on 4-7 September 2021

In **Figure 7** and **Figure 8** are visualizations of words that appear most often in tweets classified as negative sentiment on the algorithm in the TextBlob library. In the first period, words such as #bobol, #palsu, #bocor, dan #ilegal were some of the many words that most often emerged from the results of people's tweets that actually had negative meanings. Then for the second period, #susah, #bingung, and #bocor were some of the many words that most often appear from the results of negative public tweets.



Figure 9. Negative Sentiment Wordcloud on VADER Library on 31 August-3 September 2021



Figure 10. Negative Sentiment Wordcloud on VADER Library on 4-7 September 2021

Figure 9 and **Figure 10** are visualizations of words that appear most often in tweets classified as negative sentiment by using the algorithm in the VADER library. In the first period, words such as #salah, #bocor, and #bencana were one of the many words that most often emerged from the results of people's tweets that actually had negative meanings. Then for the second period, #bocor, #susah and #panik were some of the many words that most often appear from the results of negative public tweets. The following is **Table 2** which contains some quotes from tweets that are classified as positive sentiments.



Table 2. Quoted Tweets with Negative Sentiment

No	Quotes Tweet	Library	Date Range Tweet Created
1	aplikasi pedulilindungi susah nunggu kode verifikasi	TextBlob	31 August-3 September 2021
2	aneh ni app pedulilindungi mw login otp g msk gjlss	TextBlob	4-7 September 2021
3	sudah dari kemaren2 akses pedulilindungi susah karena selalu minta location	VADER	31 August-3 September 2021
4	Sumpah cape bgitu make app pedulilindungi	VADER	4-7 September 2021

In contrast to the previous sentiment analysis, the quote of the tweets above is basically about the experience and criticism of the PeduliLindungi application as the wrong one of the policies of the Indonesian Government in dealing with the COVID-19 pandemic in Indonesia, especially with regard to vaccination. The experience of the community in using the application and the negative comments given can be useful as an evaluation for the Government of Indonesia to improve the application so that this policy can run according to its objectives, namely as an effort to minimize and overcome the negative impact in the era of disruption caused by the COVID-19 pandemic.

In **Figure 11** and **Figure 12** the distribution of sentiment on twitter data obtained by using the algorithm in the TextBlob and Vader libraries. For the TextBlob library, both in the first and second periods, the percentage of tweets was dominated by neutral, positive, and negative sentiments, respectively. In addition, there is a significant difference in the percentage in the two periods. For negative and neutral sentiment, only a percentage decrease. Meanwhile, positive sentiment has increased both in percentage and the actual number of tweets.

In contrast to the sentiment analysis generated by the TextBlob library, the Vader tweet library is dominated by positive, neutral, and negative sentiments, respectively. In this algorithm, the percentage difference in sentiment in the two periods is quite large, namely there is a difference that reaches up to 10%. For negative and neutral sentiment, only a percentage decreases but the actual number of tweets continues to increase. Meanwhile, positive sentiment has increased both in percentage and the actual number of tweets.

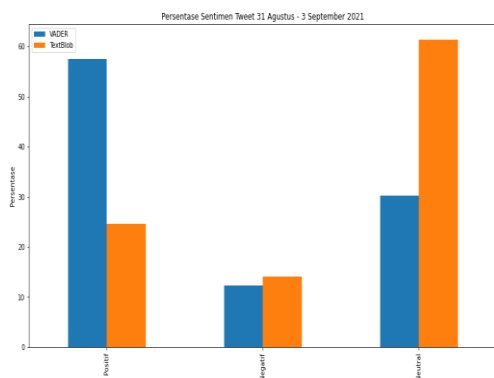


Figure 11. Tweet Percentage Comparison in 31 August-3 September 2021

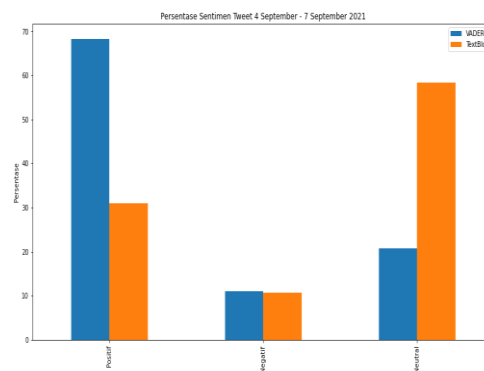


Figure 12. Tweet Percentage Comparison in 4-7 September 2021



The following is **Table 3** and **Table 4** which contain tweets classified as positive and negative sentiments respectively on the algorithm using the VADER library but not on TextBlob.

Table 3. Tweet Quotes with Positive Sentiments for the period August 31-September 2021 and September 4-7 2021, respectively.

No	Quotes Tweet
1	Aplikasi keren banget nih
2	Baru tau kalau sekarang hasil PCR langsung masuk ke PeduliLindungi. Kita bisa ngecek sendiri via web untuk kelayakan

Table 4. Quotes for Tweets with Negative Sentiments for the period August 31- September 2021 and September 4-7 2021 respectively

No	Quotes Tweet
1	sudah dari kemaren2 akses pedulilindungi susah karena selalu minta location
2	eror mulu aplikasi

In addition, **Figure 11** and **Figure 12** also show that the percentage of positive sentiment in the VADER library is always greater than the algorithm in the TextBlob library in both periods. Meanwhile, the percentage of negative sentiment produces a value that is not much different. This high difference is allegedly because there are differences in the approach of the two. At VADER is a lexicon and sentiment analysis tool that is specifically tailored to the sentiments expressed on social media [18]. Meanwhile, TextBlob uses movie reviews as its corpus [17]. Therefore, the percentage of results obtained is much higher in the VADER library than TextBlob because of its compatibility with the Twitter data in the study. In addition, in **Table 3** and **Table 4** there are several examples of tweets that are not classified as negative sentiment and positive sentiment respectively in the TextBlob library, but in VADER the opposite is true. Thus, in this study the VADER library can be said to be superior to TextBlob. The use of the lexicon on social media in the construction of the VADER classifier enables VADER to classify tweets more precisely based on their sentiments.

4. Conclusion

After analyzing sentiment with the TextBlob and VADER algorithms for tweets before and after the news of the vaccine certificate data leak, the following conclusions were obtained

1. There was an increase of sentiment from September 3 to September 4, 2021. The sentiment increment occurred at the same time when there was news related to the alleged leak of vaccine certificate data became a trending topic on Twitter.
2. Tweets before the news of vaccine certificate data leak identified on August 31 - September 3, 2021 and after the news of vaccine certificate data leak identified on September 4 - September 7, 2021.
3. With the TextBlob library, before the news was released, words such as #benar, #wajib, #aman were some of the words that have positive meanings, and words such as #susah, #sulit, #bocor, #negatif have a negative meaning. After the news was released, #benar, #guna, #kontrol is one of the many words that have a positive meaning, and #masalah, #tuntut, #kesal is one of the many words that have a negative meaning.



4. With VADER library, before the news was released, words such as, #aman, #peduli are tweets that are truly positive and words such as #salah, #bocor, #paksa are the tweets that have a negative meaning. After the news was released, #kontrol, #aktif, #peduli, #dukung are words that have a positive meaning and #bocor, #panik are words that have a negative meaning.
5. The VADER library is specifically intended to analyze sentiment on social media. It also shows that the VADER tweet library is dominated by positive, neutral, and negative sentiments, respectively and the sentiment's differences in the two periods are not too large.

Based on the results of sentiment analysis, the results can be used as a reference to evaluate the performance of the PeduliLindungi application. The word that have negative meaning such as #bocor appears frequently. Based on those opinions, the government can improve the security system on the PeduliLindungi application to relieve public unrest. The word that have positive meaning such as #benar, #peduli, and #kontrol appear frequently. Based on those opinions, the government can maintain the PeduliLindungi application's usefulness to always do tracking to control the spread of the Covid-19 virus. Advice for the government for the optimization of PeduliLindungi applications in the future, the verification process of PeduliLindungi application should be further tightened by imposing two factor authentication either with biometrics or digital signatures. In addition, some suggestions from the study [23] related to data security, among others, changing the terms of Use of Protection Care to comply with PP Nomor 71 Tahun 2019 pasal 3 on Electronic System and Transaction Operators (PSE), adopting best practices from ISO 27001 for information technology security and ISO 27701 for the protection of personal data, and PeduliLindungi data must be encrypted and can only be decrypted by the PeduliLindungi application.

5. Limitations and Suggestions

These are limitations and suggestions for improvement in future studies

1. This study does normalization manually. So, for further studies, normalization can be done automatically and preprocessing activities can be done even better, such as stopwords. It is also possible to use stopwords that are not only limited to certain stoplist dictionaries.
2. This study doesn't do evaluation for the model and it is possible to evaluate the sentiment analysis model to determine the accuracy of the model made.
3. The process of translating data into English before doing sentiment analysis is still manually with the help of google translate.

References

- [1] Satuan Tugas Penanganan COVID-19. (2020, April 8). Empat Strategi Pemerintah Atasi COVID-19. Empat Strategi Pemerintah Atasi COVID-19. <https://covid19.go.id/p/berita/empat-strategi-pemerintah-atasi-covid-19>
- [2] World Health Organization (WHO). (2020, July 28). Pertanyaan dan jawaban: Pelacakan kontak untuk COVID-19. pertanyaan jawaban terkait COVID-19, HIV, antiretroviral di Indonesia. <https://www.who.int/indonesia/news/novel-coronavirus/qa/qa-contact-tracing>
- [3] MediaIndonesia. (2020, April 15). Tiga Kementerian, BNPB, dan Operator Buat Aplikasi PeduliLindungi. Tiga Kementerian, BNPB, dan Operator Buat Aplikasi PeduliLindungi. <https://mediaindonesia.com/teknologi/304217/tiga-kementerian-bnpb-dan-operator-buat-aplikasi-pedulilindungi>
- [4] PeduliLindungi. (n.d.). Kebijakan Privasi - PeduliLindungi. Kebijakan Privasi - PeduliLindungi. <https://www.pedulilindungi.id/kebijakan-privasi-data#2-perolehan-dan-pengumpulan-data>
- [5] Indonesiabaik.id. (2021, April). Jangan Sebar Sertifikat Vaksinasi Covid-19. <http://indonesiabaik.id/infografis/jangan-sebar-sertifikat-vaksinasi-covid-19>
- [6] Bestari, N. P. (2021, September 3). Sertifikat Vaksin Jokowi Bocor, Netizen: Apalagi Data Kita. Sertifikat Vaksin Jokowi Bocor, Netizen: Apalagi Data Kita. <https://www.cnbcindonesia.com/tech/20210903114615-37-273446/sertifikat-vaksin-jokowi-bocor-netizen-apalagi-data-kita>



- [7] Raila, T. A., Rosadi, S. D., & Permata, R. R. (2020, December). PERLINDUNGAN DATA PRIVASI DI INDONESIA DAN SINGAPURA TERKAIT PENERAPAN DIGITAL CONTACT TRACING SEBAGAI UPAYA PENCEGAHAN COVID-19 SERTA TANGGUNG JAWABNYA. *Jurnal Kependidikan Hukum dan Keadilan*, 2(Volume 2 Nomor 1), 1-16. <https://jurnal.um-palembang.ac.id/KHDK/article/view/3044/2188>
- [8] Cyberthreat.id. (2020, June 29). Ini Hasil Evaluasi BSSN terhadap Aplikasi Covid-19 PeduliLindungi. NEWS: Ini Hasil Evaluasi BSSN terhadap Aplikasi Covid-19 PeduliLindungi. <https://cyberthreat.id/read/7325/Ini-Hasil-Evaluasi-BSSN-terhadap-Aplikasi-Covid-19-PeduliLindungi>
- [9] Sudiarsa, I. W., & Wiraditya, I. G. B. (2020, December). ANALISIS USABILITY PADA APLIKASI PEDULI LINDUNGI SEBAGAI APLIKASI INFORMASI DAN TRACKING COVID-19 DENGAN HEURISTIC EVALUATION. *Journal of Information Technology and Computer Science (INTECOMS)*, 3(Volume 3 Nomor 2), 354-364.
- [10] Kurniawati, & Khadapi, Muamar & Riana, Dwiza & Arfian, A & Rahmawati, E & Heriyanto, (2020). Public Acceptance Of Pedulilindungi Application In The Acceleration Of Corona Virus (Covid-19) Handling. *Journal of Physics: Conference Series*. 1641. 012026. 10.1088/1742-6596/1641/1/012026.
- [11] Karyadi, S., Yasin, H., & Mukid, M. A. (2016). Analisis Kecenderungan Informasi Dengan Menggunakan Metode Text Mining (Studi Kasus: Akun Twitter@ Detikcom). *Jurnal Gaussian*, 5(4), 763-770.
- [12] Maghfira, A., & Ridho, F. (2020). Social Sensing Sensus Penduduk 2020 pada Jejaring Sosial Twitter. In *Seminar Nasional Official Statistics (Vol. 2020, No. 1, pp. 236-246)*.
- [13] Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2013). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS*, 6(1), 37-43.
- [14] Permatasari, N., Yosral, R., & Annisa, C. F. (2020). Analisis Media Sosial Twitter tentang Pendidikan Daring pada Masa Pandemi Covid-19 di Indonesia. In *Seminar Nasional Official Statistics (Vol. 2020, No. 1, pp. 359-369)*.
- [15] Garay, J., Yap, R., & Sabellano, M. J. (2019, February). An Analysis on The Insights of The Anti-Vaccine Movement from Social Media Posts Using K-Means Clustering Algorithm and VADER Sentiment analyzer. In *IOP Conference Series: Materials Science and Engineering (Vol. 482, No. 1, p. 012043)*. IOP Publishing.
- [16] Bonta, V., & Janardhan, N. K. N. (2019). A comprehensive study on lexicon-based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, 8(S2), 1-6.
- [17] Loria, Steven. (2020) TextBlob: Simplified Text Processing. https://textblob.readthedocs.io/en/dev/advanced_usage.html [accessed on 12 September, 2021]
- [18] Analytics Vidhya. (Januari, 2021) <https://www.analyticsvidhya.com/blog/2021/01/sentiment-analysis-vader-or-textblob/> [accessed on 12 September, 2021]
- [19] Kompas.com. (2021, August 20). Viral, Video Barcode Sertifikat Vaksinasi Covid-19 Tidak Dapat Dipindai dengan Aplikasi PeduliLindungi. Viral, Video Barcode Sertifikat Vaksinasi Covid-19 Tidak Dapat Dipindai dengan Aplikasi PeduliLindungi. <https://www.kompas.com/tren/read/2021/08/20/163500065/viral-video-barcode-sertifikat-vaksinasi-covid-19-tidak-dapat-dipindai?page=all>
- [20] Ditjen Aptika. (2021, July 14). Kominfo Minta Masyarakat Cermat Simpan Data Digital Sertifikat Vaksinasi Covid-19. Kominfo Minta Masyarakat Cermat Simpan Data Digital Sertifikat Vaksinasi Covid-19. <https://aptika.kominfo.go.id/2021/07/kominfo-minta-masyarakat-cermat-simpan-data-digital-sertifikat-vaksinasi-covid-19/>
- [21] Detikcom. (2021, August 13). Bahaya Cetak Sertifikat Vaksin COVID-19 di Jasa Pencetak, Siswa Perlu Hati-hati. Bahaya Cetak Sertifikat Vaksin COVID-19 di Jasa Pencetak, Siswa Perlu Hati-hati. <https://www.detik.com/edu/detikpedia/d-5680073/bahaya-cetak-sertifikat-vaksin-covid-19-di-jasa-pencetak-siswa-perlu-hati-hati>



- [22] Hermawan, L., & Ismiati, M. B. (2020). Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval. *Jurnal Transformatika*, 17(2), 188-199.
- [23] Suyudi, Teguh. (2021). Begini Rekomendasi dari ID-IGF untuk Penyempurnaan Aplikasi PeduliLindungi. <https://www.itworks.id/43602/begini-rekomendasi-dari-id-igf-untuk-penyempurnaan-aplikasi-pedulilindungi.html> [accessed on 24 October, 2021]