

LATENT DIRICHLET ALLOCATION (LDA) METHOD ANALYSIS ABOUT COVID-19 VACCINE ON TWITTER SOCIAL MEDIA

Happy Alyzhya Haay^{1*}, Adi Setiawan²

^{1,2} Study Programme of Master of Data Science, Department of Mathematics and Data Science, Faculty of Science and Mathematics, Universitas Kristen Satya Wacana
Jl. Diponegoro No.52-60, Salatiga, Kec. Sidorejo, Kota Salatiga, Jawa Tengah 50711, Indonesia

Corresponding author e-mail: ^{1*} 632020002@student.uksw.edu

Abstract. Twitter is one social media that often provides much information for its users, one of which is information regarding the COVID-19 vaccination. This study aimed to explore and find out what topics are often discussed on Twitter social media. One of which is the topic of COVID-19 vaccination using the Latent Dirichlet Allocation (LDA) method and analysis of the frequency of keywords that often appear with this topic. The Tweet data used in this study was taken from Twitter users worldwide in November 2021. In this study, the results of sentiment analysis were obtained from the tweet data taken, which was divided into positive sentiment and negative sentiment, namely "vaccination" with 40 words and "Covid19" with 35 words

Keywords: COVID-19, Latent Dirichlet Allocation (LDA), Twitter

Article info:

Submitted: 30th December 2021

Accepted: 20th February 2022

How to cite this article:

H. A. Haay and A. Setiawan, "LATENT DIRICHLET ALLOCATION (LDA) METHOD ANALYSIS ABOUT COVID-19 VACCINE ON TWITTER SOCIAL MEDIA", *BAREKENG: J. Il. Mat. & Ter.*, vol. 16, iss. 1, pp. 189-198, Mar. 2022.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).
Copyright © 2022 Happy Alyzhya Haay, Adi Setiawan.

1. INTRODUCTION

The outbreak of a new disease that first appeared in Wuhan, China, at the end of 2019 caused the coronavirus or often known as COVID-19. It has caused thousands of patients to die from the coronavirus. It quickly spreads throughout the world, resulting in a global pandemic. Therefore, the health protocol was enforced in all aspects of community activities worldwide, from social restrictions to several countries implementing Lockdown to break the chain of the spread of COVID-19 [1]. However, this has severely hampered the activities of people worldwide. It potentially brings significant challenges to the world's health systems. It has consequences on the global economy if the spread of this virus is not controlled. Therefore, one way to prevent the spread of this virus from increasing is by developing a vaccine. Over time, researchers worldwide are working together to develop a vaccine against the COVID-19 virus rapidly.

In general, a vaccine is a liquid injected into the body to increase the body's immunity to fight disease and prevent disease [2]. Giving vaccines to the entire community is one of the most effective measures to overcome the COVID-19 pandemic, which continues today [3]. It is a topic that is always discussed through Twitter and social media. Twitter social media users who have received and have not received the COVID-19 Vaccination share their experiences via tweets on their respective Twitter accounts. Therefore, this study aims to determine what topics are being discussed by Twitter social media users related to the COVID-19 vaccination discourse using the Latent Dirichlet Allocation (LDA) method.

The previous research was conducted using the Latent Dirichlet Allocation method by [4]. The study was conducted on Sentiment Analysis of the Pros and Cons of the Indonesian Society about the COVID-19 Vaccine on Twitter Social Media. The sentiment analysis results were that the Indonesian people gave a lot of positive responses to the COVID-19 vaccine compared to those with a negative response. The words that are spoken tend to have positive sentiments. Using the LDA model, the topics of conversation of the Indonesian people on Twitter vary widely from the price of vaccines, *Merah-Putih* vaccines, and certification of vaccine-worthy tests to general discussions such as the function of vaccines. Another study [5] examined the classification of final project report documents for final year students of the Department of Computer Engineering based on topic modeling using the Latent Dirichlet Allocation (LDA) method. The analysis results obtained that the LDA method is sensitive to word composition. When the data contains many common words, the LDA method can group documents with specific topics.

Another related research is conducted by [6] regarding the implementation of the LDA method to detect topics regarding fashion that are being discussed a lot on social media. The tweets that have been obtained will then be classified using the Naïve Bayes method, and data cleaning is carried out to obtain the best iteration results. Other related research is research conducted by [7] regarding the classification of complaints handling on Twitter using Text Mining, messages on Twitter, or some people sending tweets containing complaints and tax complaints. A tax complaint classification is carried out with text mining to handle it. Classifications such as Naïve Bayes, Support Vector Machine (SVM), and Decision Tree are carried out. The best accuracy results are obtained for the SVM method.

Another study was also conducted by [8] regarding Twitter social media sentiment analysis on the case study of the Anti-LGBT campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest algorithms. This case study was taken because Indonesian people on Twitter and social media widely discuss the Anti-LGBT campaign. Therefore, sentiment analysis is carried out to know the tendency of public comments about the anti-LGBT campaign in Indonesia, whether the sentiment is positive, negative, or neutral. Sentiment analysis results indicate that Twitter users in Indonesia tend to provide more neutral comments with an accuracy of 86.43%. It is obtained from the results of data testing using Naïve Bayes. It has higher accuracy than other algorithms, namely Decision Tree and Random Forest, which obtain an accuracy of 82.91%.

Another related research was conducted by [9] by analyzing online news articles and disease data about the coronavirus. Using the LDA model, we combine COVID-19 case data and news articles into a generic LDA to derive a new feature set. The features are generated as additional features in the Machine Learning algorithm and then used by Gibbs Sampling as the primary technique for parameter inference. The results show that the PAN-LDA (Pandemic-Latent Dirichlet Allocation) features result in more identifiable topics and empirically add value to the results.

Another related study was [10], who investigated the descriptive analysis of near-miss cycling incidents reported on the crowdsourcing and BikeMaps.org platform. This analytical method obtains helpful

information from structured text data sources. Text mining, topic modeling, and machine learning are used to gain insights from textual descriptions. The results and methodology used in this study can help engineers, policymakers, and law enforcement officers proactively reduce the potential for cyclists collisions and prioritize areas where cyclists' safety is improved.

Another study was conducted by [11] on the importance of text mining in service management as access to big data increases across digital platforms. It uses sentiment analysis, topic modeling, and Natural Language Processing (NLP) in business management journals. Then, it applied visualization tools for text mining and topic associations to understand the dominant themes and relationships. The results show that the most dominant themes are social media analysis, market analysis, and competitive intelligence. At the same time, other themes such as risk management and detection of fake content are also explored. This analysis becomes the agenda for further research in text mining in service management.

Another study was conducted by [12] regarding the grouping of opinions from the users of the Ruangguru application as a material for evaluation of the application services. The Ruangguru application is an application to help students learn. This application helps students learn during a pandemic by studying at home. Many users have an opinion about the application on Twitter. Another study was also carried out by [5]. The Latent Dirichlet Allocation model obtained news data that is often discussed on social media and Twitter. Then, the Bayesian method was applied with Gibbs Sampling estimation to obtain the highest log-likelihood value. It indicates that the topic is most dominantly discussed in the community and published on news accounts on Twitter. Based on previous studies that have been carried out, this study was carried out with the title "Analysis of the Latent Dirichlet Allocation (LDA) Method on COVID-19 Vaccination on Twitter Social Media" to be able to find out the community's response to COVID-19 Vaccination.

2. RESEARCH METHOD

2.1 Theoretical Foundation

Latent Dirichlet Allocation (LDA)

The Latent Dirichlet Allocation (LDA) method is one of the most widely used topics modeling methods today. LDA is used to analyze extensive documents. LDA is a generative probabilistic model of a collection of writings or what is often referred to as Corpus. The main idea behind this method is that each document is represented as a random mixture of hidden themes. Each of which has a character assigned based on the distribution of the words in it [13]. As an unsupervised method, LDA requires defining the number of topics to be generated by the model. One of the main techniques used to determine the best number of topics that will be used to build an LDA model is to look at topic coherence. Topic coherence is a technique based on the ease of interpreting the output on the resulting topic.

The LDA model is one of the generative probabilistic models from a collection of several documents that will be represented as a mixed model of various topics where each topic is characterized by words [14]-[15]. The LDA model assumes the following generative process for every w document in the corpus D :

1. Choose $N \sim \text{Poisson}(\xi)$
2. Choose $\theta \sim \text{Dir}(\alpha).W$
3. For every N word w_n :
 - a. Choose topic $z_n \sim \text{Multinomial}(\theta)$.
 - b. Choose a word w_n from $p(w_n | z_n, \beta)$

There are some simple assumptions in the LDA model. First, the distribution of the known topics follows the k Dirichlet distribution. The second probability word is a matrix of size $k \times V$ which $\beta_{ij} = p(w^j = 1 | z^j = 1)$ while the Dirichlet k distribution has the following density function:

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1-1} \dots \theta_k^{\alpha_k-1}. \quad (1)$$

Joint distribution form of mixture topic θ from N topic z and N word w conditioned α and β is:

$$p(\theta, z, w | \alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta) p(w_n|z_n, \beta). \quad (2)$$

The LDA model if depicted in a diagram can be illustrated in Figure 1 below.

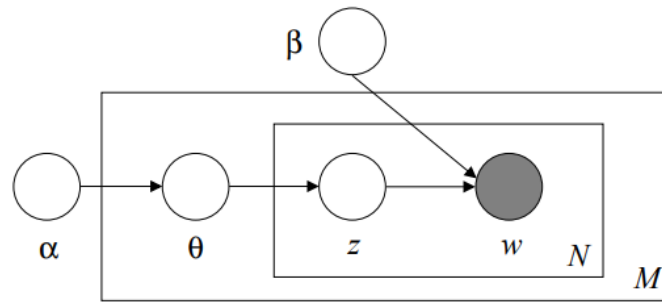


Figure 1. LDA Graphical Model Representation

Where

- β is the Dirichlet parameter of the distribution of words to the topic,
- W is word,
- N is a collection of words,
- M is a collection of documents,
- Z is the topic of the index assignment,
- Θ is a document,
- α is the Dirichlet parameter on the topic distribution of the LDA document formulated in equation (2).

Sentiment Analysis

Sentiment analysis is one of the methods used to determine human emotions. Sentiment analysis is related to text mining which aims to determine public perceptions of a topic of discussion, event or problem. The primary task of sentiment analysis is to classify a particular text and determine whether the emotion of the text or document is positive and negative, or neutral [16]. Each word in the sentence will be given a value, namely one or (1) for a word with a positive sentiment and a negative value for a negative 1 or (-1) for a word with a negative sentiment. Giving this value/score aims to determine a sentence with a positive or negative sentiment. [17].

Text and Hashtag Frequency Analysis

This analysis is to see the frequency of occurrence of specific text or words and the use of hashtags. The hashtag in question is symbolized by "#" on Twitter social media. The function of this symbol is to make it easier for users to easily follow the topics they like. Hashtags are used to associate tweets with specific topics discussed the most by Twitter users who are often related to an event.

The word cloud is used to visualize the frequency with which words appear. Word cloud is one of the most frequently used data visualization techniques [18]. In word cloud visualization, the letter size of words will be directly proportional to the frequency of word occurrences, so the more often the topic appears, the larger the font size. When extracting keywords, word position information, such as text titles, abstracts, and paragraphs' first and last sentences, contain essential information. [19].

2.2 Methodology

This study uses data from Twitter regarding the responses and opinions of Twitter users about the COVID-19 vaccination in Indonesia. The data were taken using the web scraping technique on Twitter social media using the Application Programming Interface (API) key from the researchers' Twitter account using a package designed to collect and manage Twitter data using the Twitter API from the R application.

The keywords used to make it easier to capture tweets from Twitter users in the web scraping process were to use two keywords, namely "COVID-19 Vaccination". The tweets data taken were tweets posted between 26-28 November 2021 due to limitations in data collection.

The following are the steps in the analysis:

1. Tweet taken using the Twitter API and enter the keyword "COVID19 Vaccination."
2. Performing text preprocessing which includes:
3. Perform data cleaning: URLs, excessive spaces, blank lines, punctuation, Retweet, at (@), and alphanumeric words.
4. Convert all characters to lowercase.

From the word cloud visualization that has been shown previously, these are topics that are often discussed on Twitter. From the largest font size, it is a word that is often talked about on Twitter. Lots of Twitter users tweet about the 'COVID-19 vaccination'. More than two hundred million Indonesian have been injected with the COVID-19 vaccine in the last few months. Therefore, the Indonesian government is very optimistic about pursuing the vaccination target by the end of the year.

Figure 3 is a visualization of the word cloud. It has been divided into positive and negative sentiments. The words in Figure 2 are words that often appear on Twitter social media. However, in this case, most Indonesian people who use Twitter social media mostly give positive tweets about vaccination. Therefore indirectly through Twitter social media, Twitter users have helped the government to ensure that vaccinations are safe and halal, as well as provide information related to this COVID-19 vaccination for people in all places that have not received the vaccine.

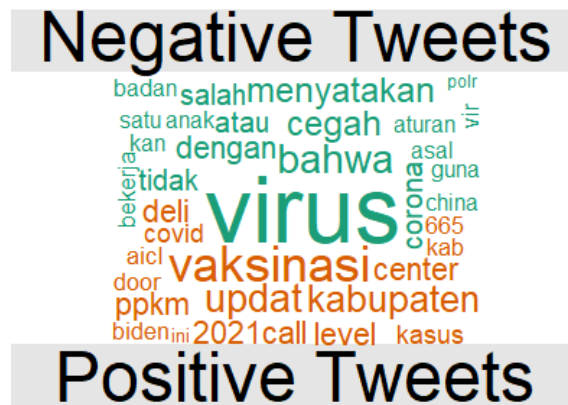


Figure 2. Visualization of Word cloud Positive and Negative Tweets

In addition, the word cloud in Figure 2, the following data shows tweets of twitter users with positive and negative sentiments.

Table 2. Example of Positive Tweets

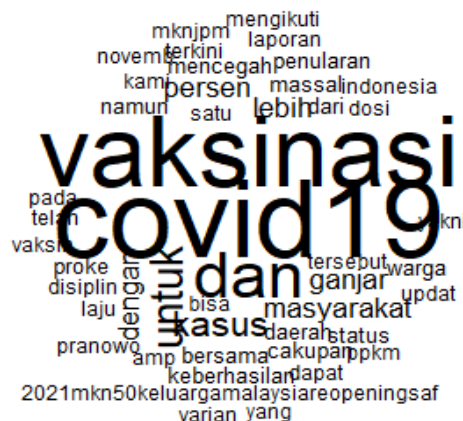
No.	Tweet	Sentiment
1	top news rakyat merdeka Newspaper ahead of the Nataru (Christmas and New Year) holiday, the National Police Chief vaccinates to keep the covid19 under control (top koran rakyat merdeka jelang libur nataru, Kapolri vaksinasi jaga pengendalian Covid-19)	0.28867513
2	you can contact us, click on the whatsapp fast response link for sworn translation of documents (bisa hubungi kami, klik link whatsapp respons cepat untuk penerjemahan tersumpah dokumen)	0.28867513
3	top news rakyat merdeka Newspaper, it is considered successful in suppressing the case of covid 19, Luhut: Don't get conceited of having vaccinated (top berita koran rakyat merdeka nilai berhasil tekan kasus Covid-19, Luhut kita jangan jumawa vaksinasi)	0.26726124
4	Handing out door prizes in the form of groceries for 10 elderly people receiving the first dose of Covid 19 vaccination from the government (penyerahan door prize sembako bagi 10 orang lansia penerima vaksinasi covid-19 dosis pertama dari pemerintah)	0.21380899
5	door prize in the form of groceries for 10 elderly people receiving the first dose of Covid 19 vaccination from the government of Kecenteng Dal (door prize sembako bagi 10 orang lansia penerima vaksinasi covid-19 dosis pertama dari pemerintah kecgenteng dal)	0.20655911

From the sentiment analysis results of tweet data obtained in Table 2 *positive sentiment*, tweet data has a positive sentiment because the government provides door prizes and necessities to people who have received COVID-19 vaccinations, especially the elderly. Based on Table 3, which are negative tweets, it is said to be negative sentiment because some of the tweets made contained some words that were considered negative in the results of this analysis. Like the words "hoax" and "death." Therefore the tweet is included in the analysis of negative sentiment.

Table 3. Example of Negative Tweets

No.	Tweets Data	Sentiment
1.	turn back hoax inaccurate covid 19 vaccination causes someone to get cancer (turn back hoax salah vaksinasi covid19 menyebabkan seseorang terkena kanker)	-0.33333333
2.	after autopsy, the cause of death due to covid 19 vaccination proved pianist 40 suffered brain haemorrhage 'induc (pasca otopsi penyebab kematian karena vaksinasi covid19 terbukti pianist 40 suffered brain haemorrhage 'induc)	-0.27735010
3.	lime moons good afternoon sis for children under 12 years old, they haven't been able to vaccinate against covid19 (limemoons selamat sore kak bagi anak usia dibawah 12 tahun belum dapat melakukan vaksinasi covid19 mengenai sy)	-0.25000000
4.	Actively vaccinating covid 19 can still go hand in hand with the immunization month for school children with PKM (giat vaksinasi covid 19 tetap bisa berjalan seiring sejalan dgn bias bulan imunisasi anak sekolah bersama pkm kem)	-0.18190172
5.	get vaccinated, let's make Papua free from the covid 19 virus 19divisihumaspolri poldapapua polrestolikara (jalani vaksinasi ayo wujudkan papua bebas virus covid 19divisihumaspolri poldapapua polrestolikara)	-0.15075567

Table 2 and Table 3, three-show tweets from Twitter users, and the value of sentiment analysis is obtained. The Indonesian government has also received a lot of positive responses because the government has succeeded in convincing every citizen that the COVID-19 vaccination is safe. By receiving this vaccination injection, it can break the chain of the spread of the COVID-19 virus so that Indonesia is free from the coronavirus.

**Figure 3. Word cloud visualization with COVID-19 Vaccination Keywords**

It can be seen in Table 4, which shows the number of words that often appear in the tweet. The words "vaccination" and "Covid-19" are often tweeted by Twitter users who give each other news that certain areas have received COVID-19 vaccinations, even the general public and the elderly in certain places. It is proof that the government has succeeded in carrying out the vaccination stage in all regions. It can reduce COVID-19 cases and has prevented the chain of spreading COVID-19. Receiving the vaccine can strengthen the body's immunity from various viruses that threaten.

The Indonesian government also targets that the country can achieve its Herd Immunity target, which means that most of the population is immune to certain infectious diseases. However, experts estimate that Indonesia can achieve the target of 70% to 80% of those who have been vaccinated. However, when Herd Immunity has been achieved, it does not mean that it is free from the COVID-19 virus, but it can slow the spread of the COVID-19 virus. Therefore, the government still urges the entire community to continue to comply with the health protocol.

Table 4. Words that often appear

Words	Word Count
vaccination	40
Covid19	35
and	22
virus	14

Based on the results of the research obtained, it is related to research that has been carried out by [4] from the results of sentiment analysis. The Indonesian people gave more positive sentiments toward the COVID-19 vaccine than negative sentiments. Similarly, another study has been carried out by [20] by text mining 103428 reviews on Google Maps from 64 major airports in the United States to identify public perceptions of COVID-19 policies at airports, namely towards staff, shops, and services. The results obtained indicate that passengers provide positive experiences with staff and shops but negative and neutral experiences for services and spaces, indicating that building design significantly impacts health crisis policies and vulnerabilities. For this reason, it is necessary to design spatial planning, service policies at airports, and building maintenance to improve transportation centers for future health crises.

In addition, some studies have been carried out similar to the results of this study by [21], who conducted a sentiment analysis from a data collection of tweets and news articles about tourist responses to the crisis during the early stages of COVID-19. Therefore, it is necessary to explore what factors influence public sentiment and see how the trends and performance of airlines and the tourism sector have been obtained from public sentiment. The same study results were also conducted by [22], who used LDA for P2P accommodation online platform review data. It was identified that the value given positively impacts consumers repurchase intentions in P2P accommodation. There is also a study conducted by [23] analyzing the sentiment of foursquare tips about accommodation, shopping, and culinary locations in Kupang, NTT. The study uses the Naïve Bayes method for classifying probabilities. An accuracy of 66.22% is obtained, which indicates that several locations, hotels, and culinary delights are the best choices based on sentiment analysis of tips.

4. CONCLUSION

An analysis has been carried out using the LDA method and keyword frequency analysis. Twitter users often tweet about COVID-19 vaccinations. With the LDA method, sentiment analysis is also carried out. However, there are more positive responses based on Twitter users' tweets. The LDA method can take a variety of topics that Twitter users most discuss. One of the topics used by researchers is COVID-19 vaccination. The words that often appear are "vaccination" and "COVID-19".

REFERENCES

- [1] Levani, Prastya, and Mawaddatunnadila, "Coronavirus Disease 2019 (COVID-19): Patogenesis, Manifestasi Klinis dan Pilihan Terapi," *J. Kedokt. dan Kesehat.*, vol. 17, no. 1, pp. 44–57, 2021, [Online]. Available: <https://jurnal.umj.ac.id/index.php/JKK/article/view/6340>.
- [2] S. Syamaidzar, "Review Vaksin Covid-19," *Res. Gate*, no. July, pp. 1–15, 2020.
- [3] F. Faulin Nur and V. N. Rahman, "Penyuluhan Program Vaksinasi Covid-19 Pada Masyarakat Desa Pakistaji," vol. 03, no. 02, pp. 491–497, 2021.
- [4] F. Rachman, S. P.-I. of Health, and undefined 2020, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," ["Analysis of the Pros and Cons of the Indonesian Society's Sentiment regarding the COVID-19 Vaccine on Twitter Social Media,"] *Inohim.Esaunggul.Ac.Id*, vol. 8, no. 2, pp. 2655–9129, 2020, [Online]. Available: <https://inohim.esaunggul.ac.id/index.php/INO/article/download/223/175>.
- [5] U. T. Setijohatmo, S. Rachmat, T. Susilawati, Y. Rahman, and K. Kunci, "Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik," ["Analysis of Latent Dirichlet Allocation Method for Classification of Final Project Report Documents Based on Topic Modeling,"] *Pros. 11th Ind. Res. Work. Natl. Semin.*, pp. 402–408, 2020.
- [6] Y. U. Al-khairi, Y. Wibisono, and B. L. Putro, "Deteksi Topik Fashion Pada Twitter Dengan Latent Dirichlet Allocation," ["Fashion Topic Detection on Twitter with Latent Dirichlet Allocation,"] *J. Apl. dan Teor. Ilmu Komput.*, vol. 1, no. 1, pp. 1–10, 2017, doi: 10.31227/osf.io/9twmn.
- [7] P. Dellia and A. Tjahyanto, "Tax Complaints Classification on Twitter Using Text Mining," *IPTEK J. Sci.*, vol. 2, no. 1, p. 11, 2017, doi: 10.12962/j23378530.v2i1.a2254.

- [8] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of Twitter social media with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.
- [9] A. Gupta and R. Katarya, "PAN-LDA: A latent Dirichlet allocation based novel feature extraction model for COVID-19 data using machine learning," *Comput. Biol. Med.*, vol. 138, no. July, p. 104920, 2021, doi: 10.1016/j.compbio.2021.104920.
- [10] K. M. Kwayu, V. Kwigizile, K. Lee, J.-S. Oh, and T. Nelson, "Automatic topics extraction from crowdsourced cyclists near-miss and collision reports using text mining and Artificial Neural Networks," *Int. J. Transp. Sci. Technol.*, no. October, 2021, doi: 10.1016/j.ijst.2021.10.005.
- [11] S. Kumar, A. K. Kar, and P. V. Ilavarasan, "Applications of text mining in services management: A systematic literature review," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100008, 2021, doi: 10.1016/j.jjime.2021.100008.
- [12] B. W. Arianto and G. Anuraga, "Topic Modeling for Twitter Users Regarding the 'Ruangguru' Application," *J. ILMU DASAR*, vol. 21, no. 2, p. 149, 2020, doi: 10.19184/jid.v21i2.17112.
- [13] M. Sakiyama, N. Fujii, D. Kokuryo, and T. Kaihara, "Visualization of group discussion using correspondence analysis and LDA in Ideathon," *Procedia CIRP*, vol. 88, pp. 595–599, 2020, doi: 10.1016/j.procir.2020.05.104.
- [14] Zulhanif, "Pemodelan Topik Dengan Latent Dirichlet Allocation," *Semin. Nas. Pendidik. Mat.*, pp. 1–8, 2016.
- [15] T. Williams and J. Betak, "A Comparison of LSA and LDA for the Analysis of Railroad Accident Text," *Procedia Comput. Sci.*, vol. 130, pp. 98–102, 2018, doi: 10.1016/j.procs.2018.04.017.
- [16] C. Liu *et al.*, "Improving sentiment analysis accuracy with emoji embedding," *J. Saf. Sci. Resil.*, vol. 2, no. 4, pp. 246–252, 2021, doi: 10.1016/j.jnlssr.2021.10.003.
- [17] J. Li, D. Lowe, L. Wayment, and Q. Huang, "Text mining datasets of β -hydroxybutyrate (BHB) supplement products' consumer online reviews," *Data Br.*, vol. 30, 2020, doi: 10.1016/j.dib.2020.105385.
- [18] N. Chintalapudi, G. Battineni, M. Di Canio, G. G. Sagaro, and F. Amenta, "Text mining with sentiment analysis on seafarers' medical documents," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100005, 2021, doi: 10.1016/j.jjime.2020.100005.
- [19] Z. Xu and J. Zhang, "Extracting Keywords from Texts based on Word Frequency and Association Features," *Procedia Comput. Sci.*, vol. 187, pp. 77–82, 2021, doi: 10.1016/j.procs.2021.04.035.
- [20] J. Y. Park, E. Mistur, D. Kim, Y. Mo, and R. Hoefler, "Toward human-centric urban infrastructure: Text mining for social media data to identify the public perception of COVID-19 policy in transportation hubs," *Sustain. Cities Soc.*, vol. 76, no. October 2021, p. 103524, 2022, doi: 10.1016/j.scs.2021.103524.
- [21] D. Obembe, O. Kolade, F. Obembe, A. Owoseni, and O. Mafimisebi, "Covid-19 and the tourism industry: An early stage sentiment analysis of the impact of social media and stakeholder communication," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100040, 2021, doi: 10.1016/j.jjime.2021.100040.
- [22] N. Zhang, R. Liu, X.-Y. Zhang, and Z.-L. Pang, "The impact of consumer perceived value on repeat purchase intention based on online reviews: by the method of text mining," *Data Sci. Manag.*, vol. 3, no. September, pp. 22–32, 2021, doi: 10.1016/j.dsm.2021.09.001.
- [23] P. Aliandu, "Sentiment Analysis to Determine Accommodation, Shopping and Culinary Location on Foursquare in Kupang City," *Procedia Comput. Sci.*, vol. 72, pp. 300–305, 2015, doi: 10.1016/j.procs.2015.12.144.

