

PAPER • OPEN ACCESS

Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm

To cite this article: Pristiyono *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1088** 012045

View the [article online](#) for updates and enhancements.

Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm

Pristiyono^{1*}, Mulkan Ritonga², Muhammad Ali Al Ihsan³, Agus Anjar⁴, Fauziah Hanum Rambe⁵

Universitas Labuhanbatu, Indonesia

*barb0112128004@gmail.com

Abstract. As of January 2021, with 2,066,175 deaths, 95,612,831 confirmed cases have been reported globally. Indonesia's COVID-19 Task Force report shows that there are currently 27,203 deaths, with reported cases exceeding 951,651, among the highest in Asia. The President of the Republic of Indonesia created a national team to speed up the production of vaccines for COVID-19. It stipulates that the government will arrange the provision, delivery, and vaccination of COVID-19 vaccines. The vaccination scheme would then become the pros and cons of Indonesian society. This research assesses the opinion of the Indonesian people through a social network analysis of the COVID-19 vaccine in January 2021. We used sentiment analysis using Naïve Bayes Algorithm by crawling Twitter data with 'Vaccine COVID-19' as keywords. We perform the data crawling process manually using the access token received from the Twitter API using the Rapid miner tools to extract the requested information and data. Data crawling continued with the Drone Emprit Academic Streaming Public Twitter Tool because of limited manual crawling resulting in more than 6000 tweets related to selected keywords on January 15-22, 2021. The result of sentiment measurement with over 3.4 thousand negative tweets (56%), over 2.4 thousand positive tweets (39%), and the remaining 301 tweets (1%) was neutral during the period of the week.

1. Introduction

It is suspected that the coronavirus originated from Bats last year in Wuhan in 2019. With the advancement of the virus by a human to human transmission, it reached humanity through the species boundary[1]. It has spread to 188 countries and 25 territories across the globe since its inception in November 2019, prompting elaborate attempts by WHO and governments to control the outbreak, largely due to the highly contagious nature of this virus. As of January 2021, with 2,066,175 deaths, 95,612,831 confirmed cases have been reported globally[2]. Indonesia's Covid-19 Task Force report shows that there are currently 27,203 deaths, with reported cases exceeding 951,651, among the highest in Asia[3].

Since COVID-19 is new to humanity and the essence of defensive immune responses is poorly known, it is unknown which vaccination techniques are going to be most effective. Therefore, designing diverse vaccine platforms and methods in tandem is crucial. Indeed, experts around the world have been racing to produce COVID-19 vaccines since the epidemic began, with at least 166 candidate vaccines now in preclinical and clinical production[4]. A new pandemic vaccine



development paradigm has been suggested to address the immediate need for a vaccine, compressing the development period from 10-15 years to 1-2 years. There remains, however, a lack of clarification as to what a secure and immunologically efficient COVID-19 vaccine strategy could constitute, how to identify good endpoints in vaccine efficacy testing, and what to expect over the next few years from the global vaccine initiative. While the characteristics of vaccines can differ, providing clear proof of direct and indirect safety may help plan how these vaccines can be used in an organized manner[5].

A national team was created by the President of the Republic of Indonesia to accelerate the production of vaccines for COVID-19. On 6 October 2020, the President signed and released a Presidential Declaration calling for a vaccine and introducing a vaccination campaign to tackle the COVID-19 pandemic. It stipulates that the government will arrange the provision, delivery and vaccination of COVID-19 vaccines. The Government of Indonesia plans to collect 30 million vaccine doses by the end of 2020 through a bilateral deal with numerous vaccine producers and an additional 50 million vaccine doses by the beginning of 2021[6].

The vaccination scheme would then become the pros and cons of Indonesian society. Among 112,888 respondents from all provinces in Indonesia, in a survey conducted by the WHO in partnership with the Indonesian Ministry of Health, NITAG, and UNICEF, approximately 65 percent of respondents indicated their readiness to receive COVID-19 vaccine is given by the Indonesian Government, while about 8 percent said they would not take it. More than 27 percent of the remaining respondents raised concerns regarding the government's intent to administer the COVID-19 vaccine[7]. About 30 percent of respondents indicated that they or their close associates, such as family members, colleagues, or neighbors, had contracted the infection with COVID-19, and they were found to be more likely to consider the vaccine. Respondents raised substantial questions about the safety and efficacy of vaccines, expressed a lack of faith in vaccines, and raised doubts about the type of haram-halal vaccines. The most common explanations for non-vaccination were vaccine safety worries (30 per cent); vaccine efficacy doubt (22 per cent); loss of faith in vaccinations (13 per cent); apprehension of side effects such as fever and discomfort (12 per cent); and religious convictions (8 percent)[7].

This research assesses the opinion of the Indonesian people through a social network analysis of the COVID-19 vaccine in January 2021. We used sentiment analysis by crawling Twitter data with 'Vaccine Covid-19' as keywords. This method is achieved by writing up Twitter opinions about an event or issue so that positive or negative opinions can be concluded using the Naive Bayes Algorithm.

2. Methodology

2.1 Sentiment Analysis

Sentiment analysis is the process of determining the sentiment and categorizing the polarity of the text in a document or phrase so that divisions can be described as positive, negative, or neutral sentiments. Today, academics are commonly using emotion analysis as a branch of computer science research. Social networks, such as Twitter, are widely used in sentiment analysis to classify the public's perception[8]. This paper will use the SLR approach to review research on the Naïve Bayes algorithm with the problem of attribute independence assumptions. Systematic Literature Review (SLR) is a process for identifying, assessing, and interpreting all available research with a view to providing answers to specific RQs[9][10].

2.2 Text Mining

Text mining is intended to produce information from a set of documents. Text Mining can generate information by processing, grouping, and analyzing large volumes of unstructured data. Data mining is used to extract valuable information from a collection of records and data references in a text in an unstructured format [11]. The method of extracting information in text mining will lead to a feeling analysis that emotionally labels a statement whether it is positive or negative. Text mining objects are non-structured or semi-structured documents.

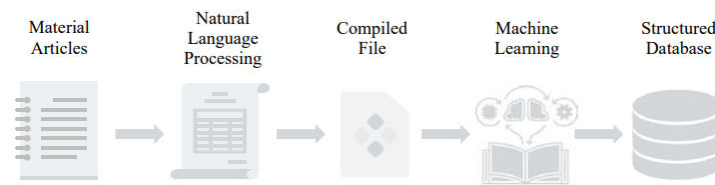


Figure 1. Text Mining Process

2.3 Naïve Bayes Algorithm

The Naïve Bayes method is a classification method for text mining used in sentiment analysis. This approach is theoretically good in terms of data consistency and calculation classification. Naïve Bayes is commonly used in classification techniques, particularly Twitter, using a variety of methods such as Unigram Naïve Bayes, Multinomial Naïve Bayes, and Maximum Entropy Classification. The main feature of the Naïve Bayes classification is to obtain a strong hypothesis of any condition or event. Calculation of the probability groups of Naïve Bayes uses the Bayes algorithm approach by using equations[12].

$$P(Y|X) = \frac{P(x|y)P(y)}{P(X)} \quad (1)$$

Equation (1) indicates that Y is a particular class, X is data of an undefined class, while P (Y | X) is the likelihood of a condition-dependent hypothesis, while P (Y) and P (X | Y) is the previous probabilities of a class based on a hypothetical condition, while P (X) is the probability of Y. Text mining efficiently removes the necessary information from a variety of records. Figure 1 displays unstructured text in organized content data in text mining and then stored in a structured database[13]. Figure 2 shows an opinion mining strategy using the Naïve Bayes process on Twitter. Data crawling is achieved by supplying keywords over a specific amount of time. The labeling process for sentiment assessment is finished after the data is obtained. The next step is a pre-processing and systematic transformation of data collection. At this point, a cleaning process is carried out to minimize noise and delete stop words to eliminate useless words such as 'I,' 'and' or '[14].

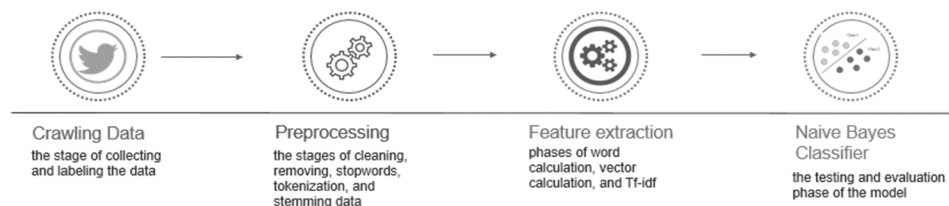


Figure 2. Research Stages Using Naïve Bayes Classification

The tokenization method is used to define words and split sentences into space-based phrases and punctuation marks. The last step in pre-processing is to modify the affixes to root words. The third stage of opinion mining is an extract intended to ease the classification of Naïve Bayes. This stage produces a model and is used to show the precision of the effects of classification[15].

3. Result and Discussion

We use the keyword 'COVID-19,' which is filtered by the keywords 'vaccine' and 'sinovac' in Indonesian tweets. The data crawling process is performed manually using the access token received from the Twitter API using the Rapid miner tools to extract the requested information and data. Data crawling continued with the Drone Emprit Academic Streaming Public Twitter Tool due to limited manual crawling resulting in 6000 tweets related to selected keywords on January 15-22, 2021[16].

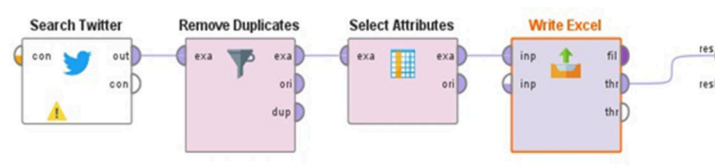


Figure 3. Data Crawling Process Using Rapid Miner

Figure 3 shows the Rapid Miner data crawling process. Some operators, such as Tweet, are used to connect to Twitter by submitting the access code token issued from the Twitter API. They eliminate duplicate operator is used to removing or erasing duplicate data during the data crawl process. The next operator is selecting attributes as seen in Figure 3, which is used to take the necessary attributes such as username and text. The last operator in excel, which is used to store data in Excel format. The effects of data extraction using Rapid miner are seen in Table 1 with a username and text attributes.

Table 1. Data Crawling (sample twit)

Username	Twit
@sofiesyarief	Vaksinasi ya bagus, tapi kan butuh waktu lama. Yang di depan mata adalah angka penularan makin tinggi dan ketersediaan fasilitas kesehatan makin langka. Masa iya gini terus sampai mayoritas penduduk disuntik vaksin?
@jokowi	Saya telah menerima suntikan vaksin Sinovac. Evaluasi BPOM, efikasinya 65,3 %, di atas standar WHO yang 50 %. MUI juga menyatakan vaksin Sinovac halal. Saya telah memerintahkan agar vaksinasi Covid-19 segera dilaksanakan di seluruh Tanah Air.
@geloraco	7 Hari Setelah Disuntik Vaksin, Bupati Sleman Umumkan Positif Covid-19 https://t.co/cpUNKEkZjf
@henrysubiakto	Ada hoax ttg chip di dlm vaksin covid-19. Vaksin itu cairan sdg Chip itu benda elektronik yg mahal, sekecil apapun ia bkn benda cair yg bisa mengalir bersama vaksin lewat jarum suntik. Hoax tdk masuk akal begini kok dipercaya?
@marlina_idha	WHO larang bukti Vaksinasi Covid-19 utk syarat bepergian. https://t.co/oxFAUB8Lft

The labeling process is carried out to decide if a tweet belongs to a positive class that involves praise, suggestions, input, and reflections on positive emotions such as joy, pleasure, and happiness. Meanwhile, the negative class comprises complaints, satirical phrases, critique, and reflections on negative emotions such as frustration, dissatisfaction, and disappointment. The results of the labeling are shown in Table 2.

Label	Twit
Negative	Vaksinasi ya bagus, tapi kan butuh waktu lama. Yang di depan mata adalah angka penularan makin tinggi dan ketersediaan fasilitas kesehatan makin langka. Masa iya gini terus sampai mayoritas penduduk disuntik vaksin?
Positive	Saya telah menerima suntikan vaksin Sinovac. Evaluasi BPOM, efikasinya 65,3 %, di atas standar WHO yang 50 %. MUI juga menyatakan vaksin Sinovac halal. Saya telah memerintahkan agar vaksinasi Covid-19 segera dilaksanakan di seluruh Tanah Air.
Negative	7 Hari Setelah Disuntik Vaksin, Bupati Sleman Umumkan Positif Covid-19 https://t.co/cpUNKEkZjf
Negative	Ada hoax ttg chip di dlm vaksin covid-19. Vaksin itu cairan sdg Chip itu benda elektronik yg mahal, sekecil apapun ia bkn benda cair yg bisa mengalir bersama vaksin lewat jarum suntik. Hoax tdk masuk akal begini kok dipercaya?
Negative	WHO larang bukti Vaksinasi Covid-19 utk syarat bepergian. https://t.co/oxFAUB8Lft

The next stage measures the frequency of terms that frequently appear in both groups to generate positive or negative and represent them using the word cloud as seen in Figure 4 and 5.



Figure 4. Word cloud Positive



Figure 5. Word cloud Negative

The sentiment analysis model is based on prepared training results that are evaluated on data testing to assess sentiments towards the COVID-19 vaccine. The text examined in this analysis is the text used in Indonesian tweets. The findings of the study data were tested to make 'Vaccine Covid-19' appear on Twitter in January 2021, with just 39% positive feelings. The negative opinion is dominated by 56 percent and the remaining 1% is neutral.

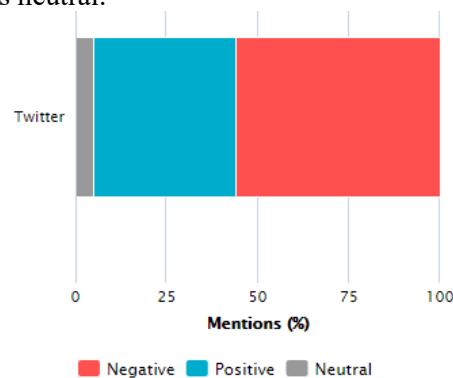


Figure 6. Sentiment Analysis

Figure 6 is the result of sentiment measurement with more than 3.4 thousand negative tweets, more than 2.4 thousand positive tweets, and the remaining 301 tweets were considered neutral during the period of the week. Based on these results, the government's plan to provide the COVID-19 vaccine is still getting negative sentiment because it has not been fully accepted in the community.

4. Conclusion

This research using the Naïve Bayes sentiment classification algorithm from Twitter data with the keyword 'COVID-19' filtered by the keyword 'vaccine' in Indonesian tweets in the second and third weeks of January 2021. Analysis in that period showed 39% positive sentiment, 56% negative sentiment, and 1% favorable sentiment. Negative opinion in that period was produced because the public did not believe that the vaccine was safe. Further study is expected to use various algorithms to achieve more precise results in the analysis of opinions.

References

- [1] X. Ou *et al.*, "Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV," *Nat. Commun.*, vol. 11, no. 1, p. 1620, Dec. 2020, doi: 10.1038/s41467-020-15562-9.
- [2] WHO, "WHO Coronavirus Disease (COVID-19) Dashboard," *covid19.who.int*, 2021. <https://covid19.who.int/> (accessed Jan. 22, 2021).
- [3] Xinhua, "Indonesia reports 11,703 new COVID-19 cases, 346 new deaths," *xinhuanet.com*, 2021. http://www.xinhuanet.com/english/2021-01/21/c_139687149.htm (accessed Jan. 22, 2021).

- 2021).
- [4] M. Jeyanathan, S. Afkhami, F. Smaill, M. S. Miller, B. D. Lichty, and Z. Xing, "Immunological considerations for COVID-19 vaccine strategies," *Nat. Rev. Immunol.*, vol. 20, no. 10, pp. 615–632, Oct. 2020, doi: 10.1038/s41577-020-00434-6.
 - [5] M. Lipsitch and N. E. Dean, "Understanding COVID-19 vaccine efficacy," *Science (80-.)*, vol. 370, no. 6518, pp. 763–765, Nov. 2020, doi: 10.1126/science.abe5938.
 - [6] B. Hermawan, "Jokowi Bentuk Timnas Percepatan Pengembangan Vaksin Covid-19," *republika.co.id*, 2020. <https://republika.co.id/berita/qgcqby354/jokowi-bentuk-timnas-percepatan-pengembangan-vaksin-covid19> (accessed Jan. 22, 2021).
 - [7] WHO, "COVID-19 Vaccine Acceptance Survey in Indonesia," 2020.
 - [8] T. Mustaqim, K. Umam, and M. A. Muslim, "Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm," *J. Phys. Conf. Ser.*, vol. 1567, p. 032024, Jun. 2020, doi: 10.1088/1742-6596/1567/3/032024.
 - [9] R. Watrianthos, S. Suryadi, D. Irmayani, M. Nasution, and E. F. S. Simanjorang, "Sentiment analysis of traveloka app using naïve bayes classifier method," *Int. J. Sci. Technol. Res.*, vol. 8, no. 7, pp. 786–788, 2019, doi: 10.31227/osf.io/2dbe4.
 - [10] R. Watrianthos, M. Bobbi Kurniawan, Kusmanto, S. Budiman, and B. Ulya, "Mapping of Traffic Accidents in Labuhanbatu Regency using GIS Support," *J. Phys. Conf. Ser.*, vol. 1566, p. 012104, Jun. 2020, doi: 10.1088/1742-6596/1566/1/012104.
 - [11] S. Panjaitan *et al.*, "Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns," *J. Phys. Conf. Ser.*, vol. 1255, p. 012057, Aug. 2019, doi: 10.1088/1742-6596/1255/1/012057.
 - [12] Y. Findawati, I. R. I. Astutik, A. S. Fitroni, I. Indrawati, and N. Yuniasih, "Comparative analysis of Naïve Bayes, K Nearest Neighbor and C.45 method in weather forecast," *J. Phys. Conf. Ser.*, vol. 1402, p. 066046, Dec. 2019, doi: 10.1088/1742-6596/1402/6/066046.
 - [13] S. Kurniawan, W. Gata, D. A. Puspitawati, I. K. S. Parthama, H. Setiawan, and S. Hartini, "Text Mining Pre-Processing Using Gata Framework and RapidMiner for Indonesian Sentiment Analysis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 835, p. 012057, May 2020, doi: 10.1088/1757-899X/835/1/012057.
 - [14] Y. Pratama, A. Roberto Tampubolon, L. Diantri Sianturi, R. Diana Manalu, and D. Friez Pangaribuan, "Implementation of Sentiment Analysis on Twitter Using Naïve Bayes Algorithm to Know the People Responses to Debate of DKI Jakarta Governor Election," in *Journal of Physics: Conference Series*, Mar. 2019, vol. 1175, p. 012102, doi: 10.1088/1742-6596/1175/1/012102.
 - [15] R. Aulianita, L. Utami, N. Musyaffa, G. Wijaya, A. Mukhayaroh, and A. Yoraeni, "Sentiment Analysis Review Of Smartphones With Artificial Intelligent Camera Technology Using Naive Bayes and n-gram Character Selection," *J. Phys. Conf. Ser.*, vol. 1641, p. 012076, Nov. 2020, doi: 10.1088/1742-6596/1641/1/012076.
 - [16] I. Fahmi, "Drone Emprit Academic: Software for social media monitoring and analytics," *dea*, 2020. academic.droneemprit.id (accessed Oct. 31, 2020).