Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11485
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Sentiment Analysis About COVID-19 Booster Vaccine on Twitter Using Deep Learning

Elly Indrayuni[1)*], Acmad Nurhadi[2)]
[1,2)]Universitas Bina Sarana Informatika, Indonesia
[1)]elly.eiy@bsi.ac.id, [2)] achmad.ahh@bsi.ac.id

Submitted : xxxxx | **Accepted** : July 13, 2022 | **Published** : July 15, 2022

Abstract: The rapid spread of COVID-19 cases to various countries has made the COVID-19 outbreak a global pandemic by the World Health Organization (WHO). The effect of the designation of COVID-19 as a pandemic has prompted the government to take preventive action against vaccination, as well as the WHO which has asked the public to immediately get a third or booster dose of vaccine. Various responses regarding the COVID-19 booster vaccine continue to emerge on social media such as Twitter. Twitter is often used by its users to express emotions about something either positive or negative. People tend to believe what they find on social networks, which makes them vulnerable to rumors and fake news. Sentiment analysis or opinion mining is one solution to overcome the problem of automatically classifying opinions or reviews into positive or negative opinions. In this study, the Deep Learning algorithm was used to analyze public opinion sentiment regarding the COVID-19 booster vaccine on Twitter. The data collection method used is crawling data using an access token obtained from the Twitter API. Meanwhile, to evaluate the model, the K-fold Cross-Validation method is used. The results of testing the model obtained the highest accuracy value at iterations = 10, which is 82.78% with AUC value = 0.836, precision = 83.33% and recall = 95.89%.

Keywords: covid-19, deep learning, sentiment analysis, twitter, vaccine

INTRODUCTION

The World Health Organization (WHO) declared the COVID-19 outbreak or corona virus to be a global pandemic on March 11, 2020. The determination of the pandemic status was due to the rapid and widespread spread of COVID-19 cases to various countries. COVID-19 is not only an infectious disease that is transmitted through contact and tiny droplets produced when people cough, sneeze or talk, now COVID-19 is a source of depression, stress and anxiety due to misleading information posted on social media (Chakraborty et al., 2020). The determination of COVID-19 as a pandemic has prompted the government to take preventive action against vaccination (Lestandy et al., 2021).

The World Health Organization (WHO) asks people around the world to immediately get the third dose of vaccine or booster vaccine. Various responses regarding the COVID-19 booster vaccine continue to emerge on various social media platforms. Social media such as Twitter, Facebook, and Instagram are the most popular communication media in society today. Twitter is often used by users to express emotions about something, either positive or negative (Ramdhani & Al-Fadillah, 2021). Based on research from Bond High Plus (Himalay, 2021) in 2021 on the internet every minute there is a fairly dense activity such as 1.4 million people scrolling on the Facebook application, 200,000 people tweeting on the Twitter application and 414,764 applications have been downloaded from the Google application provider. Play and App Store use smartphones every minute (Indrayuni et al., 2021). The majority of people consume news from social media for the first time, compared to other traditional sources such as television, newspapers and others. People tend to believe what they find on social networks, which makes them vulnerable to rumors and fake news (Lestandy et al., 2021).

Sentiment analysis or commonly known as opinion mining is a branch of research from text mining that aims to determine public (audience) perceptions or subjectivity to a topic, event, or problem (Rachman & Pramana, 2020). Sentiment analysis is a classification task that classifies a text into a positive or negative orientation (Haddi et al., 2013). Therefore, sentiment analysis or opinion mining is one solution to overcome the problem of automatically classifying opinions or reviews into positive or negative opinions (Indrayuni, 2018).

*Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v7i3.11485 p-ISSN : 2541-044X

Research on sentiment analysis has previously been carried out, including analysis of public opinion sentiment on the effect of PSBB on Twitter with the Decision Tree-KNN-Naive Bayes algorithm (Syarifuddin, 2020), sentiment analysis of COVID-19 tweets using the Deep Learning model (Chakraborty et al., 2020), sentiment analysis of online learning on twitter during the COVID-19 pandemic using the Naive Bayes method (Samsir et al., 2021). In research (Syarifuddin, 2020) using the Decision Tree-KNN-Naive Bayes algorithm, the results of this study show the accuracy values of Decision Tree, KNN, and Naïve Bayes of 83.3%, 80.80%, and 80.03%. The results for the precision of Decision Tree, K-NN, and Naïve Bayes are 81.06%, 82.72%, and 87.54%. Meanwhile, the results for Recall from Decision Tree, K-NN, and Naïve Bayes were 87.17%, 74.41%, and 62.71%. So in this study it can be concluded that the Decision Tree algorithm is the best classification in sentiment analysis. The second study, namely (Chakraborty et al., 2020) analysis of COVID-19 tweet sentiment using the Deep Learning method. This study resulted in a high accuracy value of 81%. Further research on sentiment analysis is (Samsir et al., 2021) for sentiment analysis of online learning during the COVID-19 pandemic on Twitter. The results of this study obtained a precision value of 97.15% using the Naive Bayes algorithm. Based on the three studies, it can be seen that the Naive Bayes algorithm produces the highest accuracy value. However, in research conducted by (Syarifuddin, 2020) the Decision Tree algorithm produces the highest accuracy value compared to Naive Bayes. Meanwhile research (Chakraborty et al., 2020) using Deep Learning produces high accuracy values as well.

Referring to previous research, this study has something in common, namely discussing sentiment analysis from tweets, while the difference lies in the topics and methods implemented, in this study the topic discussed was the COVID-19 booster vaccine sentiment analysis using the Deep Learning method for English texts.

LITERATURE REVIEW

Crawling data is a stage in research that aims to collect or download data from a database. Data collection from this study is data downloaded from the twitter server in the form of users and tweets and their attributes (Eka Sembodo et al., 2016).

The Twitter API or Twitter Application Programming Interface (API) is a program or application provided by Twitter to make it easier for other developers to access information on the Twitter website (Eka Sembodo et al., 2016).

Data pre-processing is the process of cleaning and preparing text for classification. Online text usually contains a lot of noise and uninformative parts such as HTML tags, scripts, and advertisements (Haddi et al., 2013). Pre-processing aims at extracting and cleaning tweets that will be used at the processing stage (sentiment analysis / tweet classification) (Arief & Imanuel, 2019).

Sentiment analysis is a strategy for checking judgments person or group; for example, some brand followers or individual customers in correspondence supporting customer representatives. Equal to content investigation, sentiment analysis finds customer opinions on various topics, including purchasing goods, provision of services, or promotional presentations (Alsaeedi & Khan, 2019). Sentiment analysis is the process of determining sentiment and classifying the polarity of the text in a document or sentence so that the category can be determined as positive, negative, or neutral (Samsir et al., 2021).

Deep learning is a sub-domain of machine learning that consists of algorithms called neural networks, which are proposed to represent a high-level generalization of data processing through multiple layers that are piled up among each other alternating linear and nonlinear transformations (Litjens et al., 2017). Deep learning algorithms are best suited when applied to massive image-based data sets to determine and test new imaging attributes (Kaur et al., 2021).

METHOD

This study applies the Deep Learning algorithm in analyzing public opinion sentiment on the COVID-19 booster vaccine through Twitter social media. Data was collected using data crawling techniques on Twitter. The stages of this research are shown in the research framework in Figure 1.

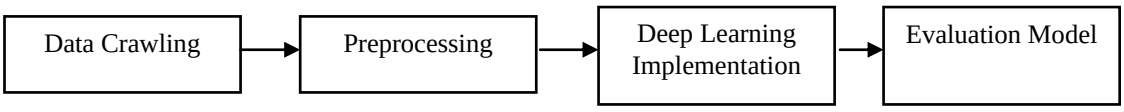

Fig. 1 Stages of Research Using Deep Learning Algorithm

*Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11485
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Data Crawling

Crawling data is done by providing keywords in a certain period related to the COVID-19 booster vaccine. The data crawling process is carried out using the Rapidminer tool using an access token obtained from the Twitter API. Tweet data crawled from Twitter is stored in the form of a CSV file.

Preprocessing

This data preprocessing stage is carried out to prepare the text before it is used in other processes. In this study, the preprocessing techniques used include filtering, cleaning, tokenization and stopword removal processes.

Deep Learning Implementation

Tweets that have gone through the preprocessing stage will be classified according to their class including positive, negative, or neutral opinions using the Deep Learning algorithm. The application of the Deep Learning algorithm uses the iterations value to find the highest accuracy value.

Evaluation Model

Sentiment classification testing is done by measuring accuracy, precision, and recall from the calculation results of the Deep Learning algorithm. Meanwhile, to evaluate the model, the K-fold Cross-Validation method is used.

RESULT

The data crawling process uses the keywords 'covid19 vaccine', 'vaccine', 'booster' and the hashtag #vaccinboostercovid19 which were selected in January 2022. In this study, the results of crawling data in the form of a CSV file are then labeled as including positive, negative or neutral tweets. The tweet data generated through this data crawling process is 947 tweets, with the categories of positive tweets being 584 tweets, negative tweets being 206 tweets, and neutral: 157 tweeting. In this study, tweet data used only two categories of tweets, namely positive tweets and negative tweets. The data crawling process using Rapidminer can be shown in Figure 2.

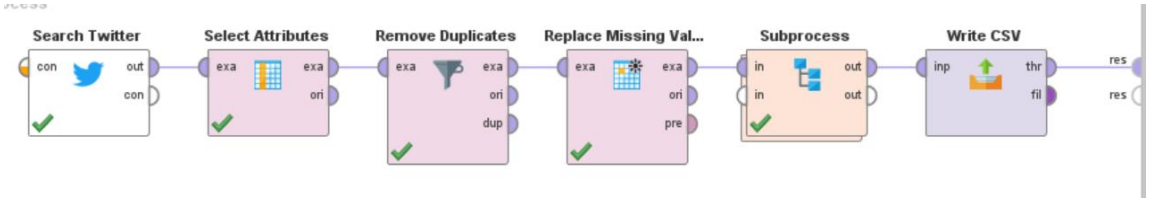

Fig. 2 Data Crawling Process Using Rapidminer

Figure 2 shows the data crawling process using Rapidminer. Several operators are used in the process. Search Twitter is used to connect to Twitter by entering the access token obtained from the Twitter API. For the filtering process, the select attributes operator is used which is useful for retrieving the required attributes such as username and text. Furthermore, the cleaning process is carried out using the remove duplicates operator to remove duplicate data during the crawling process. The replace missing value operator is applied to keep no missing values, then the subprocess operator is used to remove words that are not important. The last operator is write CSV which is used to store data in the form of a CSV file.

The next stage before the sentiment classification process using the Deep Learning algorithm is applied, namely the stage of the labeling process. This labeling is used to determine which tweets are categorized as positive tweets or negative tweets. The results of labeling can be seen in Table 1.

Table 1. Data Labeling

Labels	Tweet
Positive	COVID-19 death rates among the fully vaccinated population are rising. As vaccine effectiveness wears off, experts in the US encourage people to get a booster shot.
Positive	If you're aged 16+ and not had your #COVID19 booster yet, you're still eligible to get it.
Positive	The CDC recommends that people who received the Johnson & Johnson vaccine get the Pfizer or Moderna booster. The CDC advises you to mix and match (i.e., get a different COVID-19 booster than their initial vaccine). Find a provider by going to https://t.co/wqsjbTyPkE. https://t.co/aGu3IgbYhG
Negative	I feel very uncomfortable taking the third vaccine while these two confuse a “booster”

*Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11485
e-ISSN : 2541-2019
p-ISSN : 2541-044X

	with a “third dose”. They are not the same and nobody can confirm which one it is that they planned for me. https://t.co/l3Rrt7o7jZ
Negative	My sister wasn't so lucky got 2 vaccine booster shot flu shot still got covid-19 and she got really sick ended in the hospital

After the labeling stage, the next step is preprocessing. This stage is the stage where the data is prepared to become data that is ready to be analyzed (Syarifuddin, 2020). To select and separate a sentence into several words, the tokenization process is carried out during data preprocessing. In addition, the tokenization process can eliminate a punctuation mark, symbol or character that is not a letter so that it will make it easier to process data. At the preprocessing stage, stopword removal is also carried out to remove conjunctions or words with affixes so that the words stored are important words that have meaning. The results of preprocessing will produce the frequency of words that appear the most and are visualized into positive and negative word clouds as shown in Figure 3.

Fig. 3 Wordcloud positive opinion

Fig. 4 Wordcloud negative opinion

The sentiment classification testing process is carried out using Deep Learning. In this test, the iterations value is applied to produce the highest accuracy value. In Deep Learning, of course, you are familiar with the terms epoch, Batch Size, and Iterations because these three things are a solution for handling large amounts of data where our computers do not allow us to train so much data in one training. The iterations values applied in this algorithm are 3, 5, 8, 10, 12 and 15. The results of testing the sentiment classification of the Covid-19 booster vaccine on Twitter can be seen in Table 2.

Table 2. Sentiment Classification Test Results Using Deep Learning

Iteration	Acuracy	AUC	Precision	Recall
3	75.57%	0.701	76.17%	97.43%
5	77.34%	0.785	77.18%	98.46%
8	76.71%	0.732	77.78%	95.89%

*Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11485
e-ISSN : 2541-2019
p-ISSN : 2541-044X

10	82.78%	0.836	83.33%	95.89%
12	80.89%	0.832	82.17%	94.69%
15	79.75%	0.811	80.72%	95.38%

In table 2 it can be seen that the highest accuracy of sentiment classification test results using Deep Learning is obtained at iteration value = 10, where the accuracy value is 82.78% with AUC value = 0.836, precision = 83.33% and recall = 95.89%. The results of the classification test using Rapidminer can be seen in Figure 5. Meanwhile, the ROC curve which shows the AUC value = 0.836 can be seen in Figure 6.

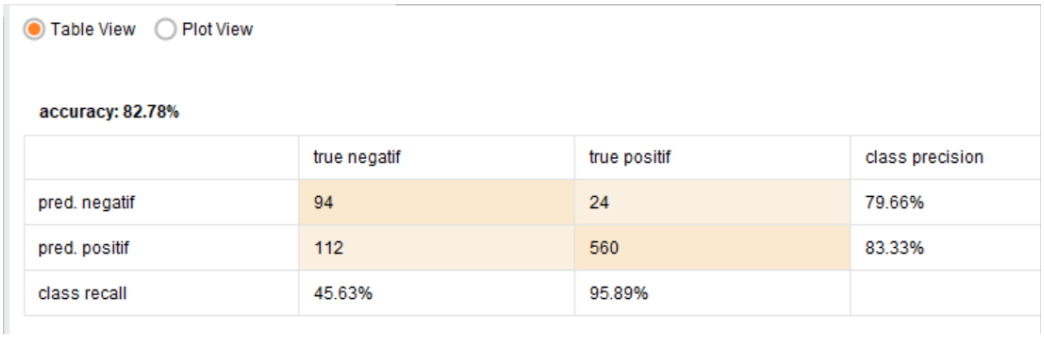
Table View Plot View

accuracy: 82.78%

	true negatif	true positif	class precision
pred. negatif	94	24	79.66%
pred. positif	112	560	83.33%
class recall	45.63%	95.89%	

Fig. 5 Test Results Using Deep Learning With Iteration Value = 10

Fig. 6 ROC Curve Test Results Using Deep Learning

DISCUSSIONS

This research has succeeded in building a sentiment analysis model by applying the Deep Learning algorithm. Based on the data collected using the data crawling technique on Twitter, 947 tweets were generated, with the category of positive tweets being 584 tweets, negative tweets being 206 tweets, and neutral: 157 tweeting. Based on this, it can be seen that tweets with positive sentiments have a greater frequency and number than tweets with negative or neutral sentiments. In this study, two categories of tweets were taken, namely positive and negative tweets. In the process of testing the sentiment classification of Twitter users towards the Covid-19 booster vaccine using the Deep Learning algorithm, 6 different iterations values were applied, namely 3, 5, 8, 10, 12 and 15. After applying the iteration values in the Deep Learning algorithm, it can be seen that the application iterations value affects the resulting accuracy value.

This research is a simple research because it only performs testing using only one algorithm, namely Deep Learning. The accuracy value generated in this study is 82.78%. In this study, any feature selection has not been applied. Feature selection can be used to improve performance in terms of speed intervals, predictive power and reduce dimensions. Feature selection is intended to eliminate irrelevant and redundant features that can affect the resulting accuracy value. In further research, to improve the quality of research on sentiment analysis or text

*Corresponding author

mining, feature selection can be applied to the Deep Learning algorithm so that it can produce better accuracy values than this research. Further research can also be carried out by comparing the test results using two algorithms and applying optimization techniques to the algorithm used to produce the highest accuracy value in sentiment analysis.

CONCLUSION

This research has succeeded in conducting a sentiment analysis on public opinion regarding the COVID-19 booster vaccine with English subtitles. The results show that the Deep Learning algorithm produces high accuracy values for sentiment analysis of twitter users towards the COVID-19 booster vaccine. Testing the Deep Learning algorithm model by applying the iterations value 6 times proves that the iterations value affects the resulting accuracy value. Based on the model test, the highest accuracy value is obtained at iterations = 10, which is 82.78% with AUC value = 0.836, precision = 83.33% and recall = 95.89%. From the description above, it can be concluded that the Deep Learning algorithm has performed well in analyzing the sentiments of twitter users towards the COVID-19 booster vaccine with English text.

REFERENCES

Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*, *10*(2), 361–374. https://doi.org/10.14569/ijacsa.2019.0100248

Arief, R., & Imanuel, K. (2019). Analisis Sentimen Topik Viral Desa Penari Pada Media Sosial Twitter Dengan Metode Lexicon Based. *Jurnal Ilmiah Matrik*, *21*(3), 242–250. https://doi.org/10.33557/jurnalmatrik.v21i3.727

Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing Journal*, *97*, 106754. https://doi.org/10.1016/j.asoc.2020.106754

Eka Sembodo, J., Budi Setiawan, E., & Abdurahman Baizal, Z. (2016). *Data Crawling Otomatis pada Twitter*. *October 2018*, 11–16. https://doi.org/10.21108/indosc.2016.111

Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *First International Conference on Information Technology and Quantitative Management*, *17*, 26–32. https://doi.org/10.1016/j.procs.2013.05.005

Himalay, P. (2021). *What Happen in an Internet Minute - Bond High Plus*. Bondhighplus.Com.

Indrayuni, E. (2018). Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film. *Jurnal Pilar Nusa Mandiri*, *14*(2), 175. https://doi.org/10.33480/pilar.v14i2.918

Indrayuni, E., Nurhadi, A., & Kristiyanti, D. A. (2021). Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc. *Faktor Exacta*, *14*(2), 64. https://doi.org/10.30998/faktorexacta.v14i2.9697

Kaur, H., Ahsaan, S. U., Alankar, B., & Chang, V. (2021). A Proposed Sentiment Analysis Deep Learning Algorithm for Analyzing COVID-19 Tweets. *Information Systems Frontiers*, *23*(6), 1417–1429. https://doi.org/10.1007/s10796-021-10135-7

Lestandy, M., Abdurrahim, A., & Syafa'ah, L. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(4), 802–808. https://doi.org/10.29207/resti.v5i4.3308

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., & Ghafoorian, M. Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88.

Rachman, F. F., & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Health Information Management Journal*, *8*(2), 100–109. https://inohim.esaunggul.ac.id/index.php/INO/article/view/223/175

Ramdhani, N., & Al-Fadillah, R. H. (2021). *Analisis Sentimen Pengguna Twitter Terhadap Belajar Daring Selama Pandemi Covid-19 Dengan Deep Learning*. *7*(2), 2021.

Samsir, Ambiyar, Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes. *Jurnal Media Informatika Budidarma*, *5*(1), 157. https://doi.org/10.30865/mib.v5i1.2604

Syarifuddin, M. (2020). Analisis Sentimen Opini Publik Terhadap Efek PSBB Pada Twitter Dengan Algoritma Decision Tree-KNN-Naive Bayes. *Inti Nusa Mandiri*, *15*(1), 87–94. https://doi.org/10.33480/inti.v15i1.1433

*Corresponding author