Lockdown Countdown: Lockdown Sentiment Analysis on Twitter Using Artificial Neural Network

Lidia Sandra
Psychology Department
Krida Wacana Christian University
Jakarta, Indonesia, 11470
lidia.sandra@ukrida.ac.id

Sabar Aritonang
Information Techology Department
Universitas Islam Indonesia
Yogyakarta, Indonesia, 55584
sabarmail.art@gmail.com

2021 International Conference on Data Science and Its Applications (ICoDSA) | 978-1-6654-4303-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICODSA53588.2021.9617212

***Abstract*— Within a few months the COVID-19 virus (Corona Virus Disease 2019) spread to all corners of the world including Indonesia. This infectious disease appeared in the city of Wuhan, China at the end of December 2019. In order to prevent all countries from implementing a lockdown policy to fully respect freedom, dignity and human rights, Indonesia is also obliged to prevent the occurrence of a public health emergency that is troubling the world. To avoid public health risk factors, the Central Government and Regional Governments have an important role to carry out health quarantine. The government has the obligation and responsibility to protect public health from this COVID-19 virus. In this paper we are doing a sentiment analysis on the lockdown policy called PPKM (*Pemberlakuan Pembatasan Kegiatan Masyarakat*, in English: Enforcement of Restrictions on Community Activities) from twitter using *Artificial Neural Network* (ANN). From the analysis it is found that ANN have the highest AUC score with 95.76 percent. It is also found that neutral sentiment dominating with 48.9 percent while positive sentiment falls second with 28.9 percent and negative sentiments with 22.2 percent. We set the goal for this paper to be able to give an analysis that is beneficial for further research and policy making.**

Keywords—sentiment analysis; classification; twitter; lockdown; artificial neural network; ANN

I. INTRODUCTION

COVID-19 (Corona-virus disease 2019) is a disease that first found in People's Republic of China, in the city of Wuhan, capital city of Hubei province in December, 2019. Since then, the disease spread all around the world and cause a worldwide pandemic that is still ongoing until now. This disease can spread among human which means contagious and caused by SARS-CoV-2 (severe acute respiratory syndrome corona-virus 2) [1]. People who contracted COVID-19 may have varieties of symptoms according to the severity of the disease and it will change as the time goes on [2]. These symptoms are ranging from mild to severe, the most common mild symptoms are fever, the loss of appetite, fatigue, loss of smell, shortness of breath, cough, muscle aches and pain [3][4][5]. While the severe symptoms could be mentioned as follows, difficulty in waking, minor confusion, bluish face and lips, coughing blood, persistent chest pain, decreasing white blood cells, kidney failure, and a high fever [6][7].

The first COVID-19 case confirmed and reported in Indonesia was on March 2020. Two people are confirmed to contract COVID-19 from a Japanese citizen [8]. On April, the pandemic spread over 34 provinces in Indonesia while DKI Jakarta, West Java, and Central Java as the most number of people contracted with the virus [9]. Until this paper is being written (30 July 2021) there are 3.372.374 reported positive cases in Indonesia which is the highest in South East Asia. On the number of deaths, Indonesia is the third highest in Asia with 92.311 deaths [10]. It is also reported that 2.730.720 people already recovered, leaving 549.343 still need medical care from the hospital [11]. As a countermeasure of the pandemic happening in Indonesia, some of the region or provinces applying a large-scale social restriction called PSBB (*Pembatasan Sosial Berskala Besar,* in English: Large-scale Social Restriction) on 2020. This policy later on changed with a community activities restriction enforcement called PPKM (*Pemberlakuan Pembatasan Kegiatan Masyarakat,* in English: Enforcement of Restrictions on Community Activities) [12].

These so called "lock-downs" as a restriction causes a different response from the citizens on the policy. One of the platforms or medium for Indonesian citizen to amplify their opinion is Twitter. Twitter is a social media service and also an online micro-blog for its users. Their services allow the users to read and send text-based messages up to 140 characters. On 7 November 2017, these restrictions are being lifted, so users now can send their messages up to 280 character that they call a Tweet [13]. Twitter was born on March 2006 from the hands of Jack Dorsey. Since published, Twitter has become one of the ten top most visited website in the internet [14]. Twitter's influence and users also expanded to Indonesia. Since then, Twitter has been a part of a social activity in Indonesia - not only as a place to give and share thoughts, but also as a spreading medium of information and resources.

In this paper will be discussed the result of a sentiment analysis method using *Artificial Neural Network* (ANN) on the lockdown policy during a period. ANN is a computing system that generates similarity with biological neural networks that is found in animal brains. ANN is an adaptive system that can change its structure to solve a problem based on the external or internal information that flows within the network itself [15]. To put simply, ANN is a tool of a non-linear data statistic modelling that is use to model a complex relationship between input and output to find the patterns in the data [16].

II. RELATED WORK

The analysis of the sentiment from public responses on the implementation of lock-downs during the COVID-19 pandemic was conducted to measure public's opinion to this. The opinion could be categorized as a response that could be Neutral, Positive or Negative sentiments on Twitter. In the previous papers that contains different studies, many of them used multiple machine learning algorithm in doing the analysis. Some of these algorithms are *Support Vector Machine (SVM)*, *Logistic Regression*, *K-Nearest Neighbors,* and *Naive Bayes*.

Sentiment analysis on PPKM using different kinds of machine learning algorithms has been done in other literature. One of the cases using Support Vector Machine algorithm to do a sentiment analysis in the city of Ambon, Indonesia. The result of this literature is that there are 28% positive sentiment, 27% negative sentiments, and 45% neutral sentiments. It is found that the sentiments of the citizen with the plan of implementing PPKM in their city is fairly balanced from the positive and negative sentiments, but dominated by neutral sentiments [17].

In another literature was found a similar results from different country. In Saudi Arabia, the reaction of the citizen towards the pandemic from Twitter are found that most of the people having a positive comment in the beginning phase of the pandemic but gradually growing negative comment as the time went on [18]. The uses of different algorithm such as Random Forest and Naive Bayes Classifier are found on a sentiment analysis paper about PPKM policy. From this literature, it can be found that the result of classifying through Random Forest achieved the best accuracy by using Unigram with the number of 99.5% while using Naive Bayes Classifier achieved the number of 97.9%. From this comparison can be concluded that different uses of the algorithm in doing a sentiment analysis can affect and improve the result [19].

From the last sentiment analysis done, we can see that the support that is given to the policy from government is received well by the community, proven with the data taken from the sentiment analysis. Also understanding that the positive comments are the major components of the existed sentiments yet the negative one comprised of violations that still happened even after the policy is implemented.

This research uses data processing libraries such as NumPy, panda, and regex to pre-process and translate the text. To concoct a *Artificial Neural Network* model, we need those as well to be able to do a classification on the sentiments about this PPKM. The *Artificial Neural Network* model created was evaluated, and was found to have an accuracy of 86.2%.

III. Method

Sentiment analysis was carried out to examine the public's response to the implementation of PPKM or Pemberlakuan Pembatasan Kegiatan Masyarakat (*Enforcement of Restrictions on Community Activities*) during the pandemic. Different tools such as Pandas and Tweepy are used to mine data from Twitter. Function are defined for the scrapping and text pre-processing, determining the keywords, amount of Tweets, and range of the period. The stages in conducting sentiment analysis in research are Data Collection, Pre-processing, Feature Extraction, Modeling. Figure 1 shows a flow chart of the stages carried out in the study.

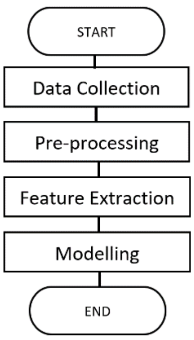

Fig. 1. Flowchart

A. Data Collection

Multiple tools are used to extract data from Twitter. These tools are Pandas, Tweepy, and others. Requesting Twitter API is the first thing to be done to get the permission on the database of Twitter accounts and the general data. Then, we define the function for scrapping and the preparation for text pre-processing later on. These steps including defining the keywords needed to mine certain Tweets, the number of Tweets needed, and the time period. Other than that, there are different parameters that are being set such as the Tweet ID, the content of it, the amount of re-tweet that the content has, the amount of favorite that the tweet has, and location coordinate of the author.

Scraping was done first to be able to analyze the sentiment according to the topic. The tweets that were scraped were ensured to contain the keywords of: *PPKM*, and *Pemberlakuan Pembatasan Kegiatan Masyarakat* (Enforcement of Restrictions on Community Activities).

B. Pre-processing

After the data-set are being gathered, pre-processing is being done to ensure the process of extracting the feature later on will be achieved effectively. First, we loaded up the list of the words that have a negative sentiment, positive sentiment, and the stop words in Bahasa Indonesia. Then it takes 10,000 data randomly to do a data training and modeling test, the list of words that has uploaded before are used to filter these datasets. This will filter out characters and words based on the stop words that we uploaded before.

The next step is to count the amount of positive and negative words based on the list that has been uploaded - then count the remainder of the total positive and negative sentiments. 0 will be used to represent neutral sentiments, 1 will represent positive sentiments, while -1 representing negative sentiments. It is found from the process that there is total 10,000 data that is taken randomly, 2,610 is a negative sentiment, 2,880 is positive, while the rest which is 4,510 are neutral sentiments.

C. Features Extraction

To extract all of the feature, first we will set all of the data will have 700 maximum feature each. So then we could split a word into one feature and convert it into the numbers. Once we split the data set, we test and train the data with different models.

With a *Gaussian Naive Bayes*, we define the model that is used to train the data. Then we fit the model and predict the output using the data test. The final part is to show the metric calculation and its accuracy. AUC score will be the next one to decide that the model fit or not. With this model the accuracy is on 61.1% with AUC score of 83.41%. The next model is *Logistic Regression* (LR), we define the model that is used to train the data. Then we fit the model and predict the output using the data test. The final part is to show the metric calculation and its accuracy. AUC score will be the next one to decide that the model fit or not. With this model the accuracy is on 86.55% with AUC score of 95.38%.

The next model after that is *Decision Tree* (DT). We define the model that is used to train the data. Then we fit the model and predict the output using the data test. The final part is to show the metric calculation and its accuracy. AUC score will be the next one to decide that the model fit or not. With this model the accuracy is on 86.4% with AUC score of

90.32%. For the *Support Vector Classifier* (SVC), we define the model that is used to train the data. Then we fit the model and predict the output using the data test. The final part is to show the metric calculation and its accuracy. AUC score will be the next one to decide that the model fit or not. With this model the accuracy is on 87.7% with AUC score of 95.66%.

Next model is *K-Nearest Neighbor* (KNN), we define the model that is used to train the data. Then we fit the model and predict the output using the data test. The final part is to show the metric calculation and its accuracy. AUC score will be the next one to decide that the model fit or not. With this model the accuracy is on 83.4% with AUC score of 95.16%. And for the last one would be *Artificial Neural Network* (ANN); we define the model that is used to train the data. Then we fit the model and predict the output using the data test. The final part is to show the metric calculation and its accuracy. AUC score will be the next one to decide that the model fit or not. With this model the accuracy is on 86.2% with AUC score of 95.76%.

D. Modelling

The best model is the one with the highest accuracy on making the prediction. As the precision from each of the model is almost similar one to another, we consider the highest AUC score from the model tested. Therefore, we choose Artificial Neural network to analyze the data-set with. These datasets were also split in such a way to ensure that no training data is used for testing, and vice versa. These datasets were then subsequently transformed using the created pipeline, to transform each record in both datasets into their corresponding numerical format. Afterward, the training data-set was fitted unto an Artificial Neural Network (ANN). The model was then evaluated with the testing data-set and was found to have an accuracy of 95.16%.

IV. Result and Discussion

the results of the second sentiment analysis using the keyword "PPKM (Enforcement of Restrictions on Community Activities)". Sentiment analysis data analysis is carried out carefully as follows:

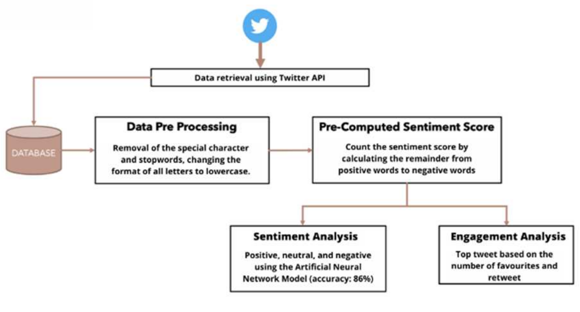

Fig. 2. Scheme of Analysis

	tanggal	id	text	rt	fav	place
0	2021-07-11 23:59:59	1414373940571480064	RT @Pencerah__: Saya heran sama sebagian masya...	91	0	NaN
1	2021-07-11 23:59:57	1414373928596762880	RT @geloraco: Dibentak Polisi saat Razia PPKM,...	1093	0	NaN
2	2021-07-11 23:59:56	1414373927070014976	https://t.co/TUSTFo80j0\n\nBeneran nih pak @jo...	0	0	NaN
3	2021-07-11 23:59:53	1414373914684236032	@shabrisabi @AREAJULID Itu kenapa skrg namanya...	5	24	NaN
4	2021-07-11 23:59:47	1414373889262572032	RT @TarizSolis: ~ Rekomendasi film pendek di Y...	4180	0	NaN

Fig. 3. Tweets Sample Scrapped in the Sentiment Analysis

The results of sentiment analysis using the PPKM keyword resulted in the largest percentage of sentiment being neutral (48.9%) followed by positive sentiment (28.9%) and then negative (22.2%). There seems to be no significant difference between positive and negative sentiments as shown in Figure 4 below.

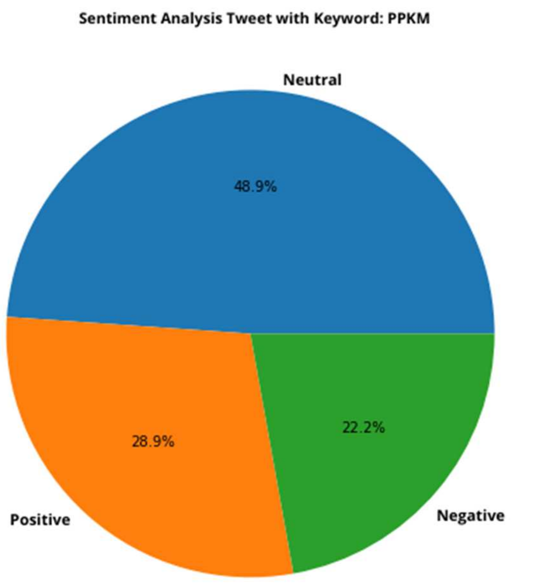

Fig. 4. Tweets Sample Scrapped in the Sentiment Analysis

From the analysis of the frequency of words that appear the most in each interesting sentiment, it is found that in all categories, the key words that appear still contain negative sentiments such as: *stress, bosan* (bored)*, mental, depresi* (depression)*, darurat* (emergency). The five words that most often appear in neutral sentiments by ignoring the word "ppkm" for example are the words: *stress, darurat* (emergency)*, butuh* (need)*, biar* (so on)*, Youtube*. The five words that most often appear in positive sentiments by ignoring the word "ppkm" for example are the words: *makan* (eat'food)*, gajian* (paycheck)*, lambat* (slow)*, darurat* (emergency)*, kota* (city). While the five words that most often appear in negative sentiments by ignoring the word "ppkm" for example are the words: *darurat* (emergency)*, tutup* (close)*, rakyat* (citizens)*, sehat* (health)*, jalan* (road).

Of all the sentiments, the word "*darurat*" which means emergency, is the one that comes up the most. In the category of positive sentiment, the most talked about interesting words are *makanan* (food) and *gaji* (paycheck). Positive sentiments may arise from people who are not directly affected by the implementation of PPKM. *Makanan* (food) and *bayar* (pay) are the most basic of PPKM's influence now. In the category of negative sentiment, the most talked about interesting words are *tutup* (close) and *jalan* (road). When examined further from these words, they probably refer to the number of road closures in cities. It can be assumed that the people are harmed by the road closures or that the people are aware of the ineffectiveness of road closures during the PPKM process. See Figure 5-7 below.

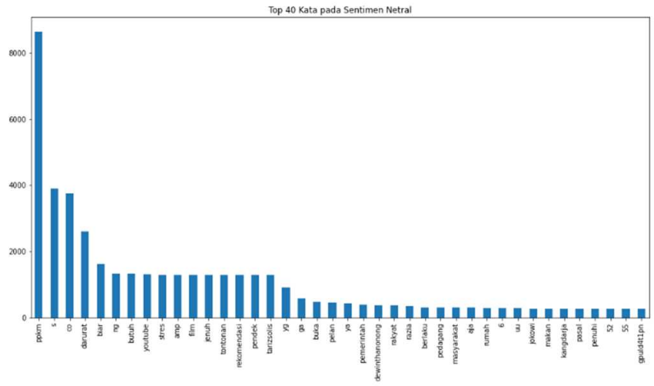

Fig. 5. Top 40 Words with Neutral Sentiments

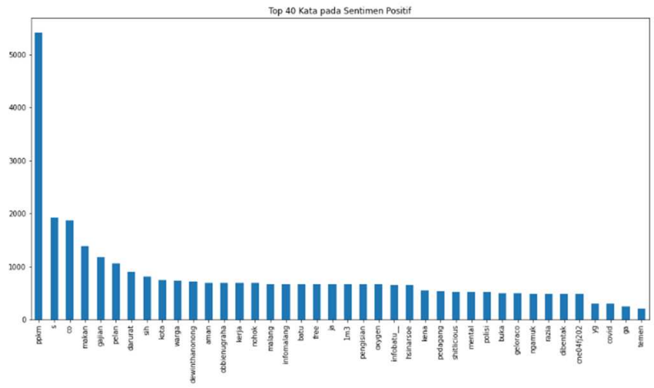

Fig. 6. Top 40 Words with Positive Sentiments

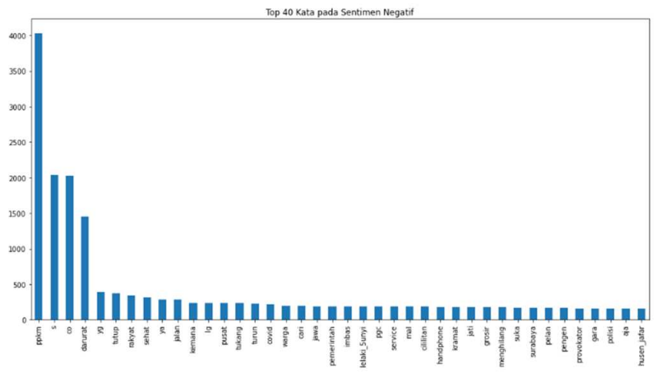

Fig. 7. Top 40 Words with Negative Sentiments

The highest number of re-tweets and favorites is held by neutral sentiment. This is also related to the result that neutral sentiment tweets are the most numerous. See Figure 8 and 9.

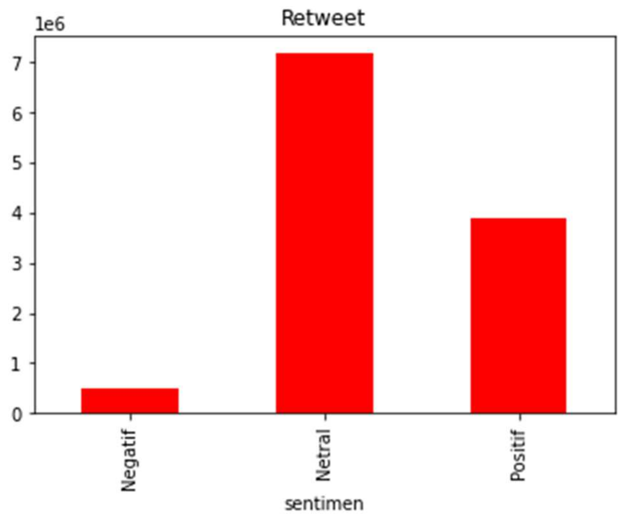

Fig. 8. Amount of Re-tweet on Each Sentiments

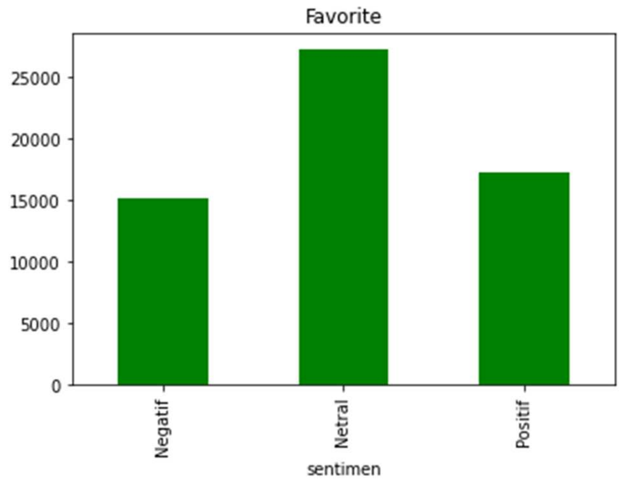

Fig. 9. Amount of Favourite on Each Sentiments

It can be concluded from the results of a netnographic [20] analysis using 25,000 netted tweets data, the biggest sentiment related to government intervention in the Covid-19 pandemic, namely PPKM has a neutral sentiment, and if you don't look at the neutral one, the largest sentiment is negative compared to positive. This shows public anxiety about the restrictions imposed. Negative sentiments that are still conveyed in a satirical tone so that they are classified more into neutral sentiments show the apathy of young people who ultimately tend to give in to situations that they feel cannot be controlled and hopes for a bigger role for the government to be able to cope with this pandemic better and faster.

V. CONCLUSION

After doing a sentiment analysis using Artificial Neural Network on the data-set, it can be concluded that. With the ANN model it can be achieved the highest AUC score than any other model tested, even though the accuracy is similar to one another. It is also found from the data-set that the biggest sentiment related to PPKM has a neutral sentiment. If we are ignoring the neutral one, the amount of the negative one is on par with the positive one, yet the positive sentiments still dominating more percentage than the negative one

An honest mirror from the youth and early adulthood community through social media also shows the frequent keywords related to psychological well-being disorders that arise such as boredom, stress, depression, confusion, homesickness, loneliness; Concerns and anxiety related to the general downturn of the economy through words and tweets that are often displayed, which shows that teenagers and early adults in general are experiencing anxiety related to the uncertain situation in the Covid-19 pandemic. The call is then often voiced by the community through the name that is most often mentioned in sentiment analysis, namely Joko Widodo, which contains hope for immediate treatment, and the addition of facilities to improve psychological, social, economic and health welfare.

Another interesting thing to study is that if we look at the tweets with the most number of engagement, it is the tweets from social media star called influencers. Their tweets have a huge opportunity to form other public sentiments and spread them with many re-tweets and likes. Contagious re-tweets from this so called influencers, able to give indication that tweets and opinions are contagious. The tendency for conformity to friends in tweeters, and the occurrence of polarization in opinion according to the circle is very high. This is interesting to study further to see the circle of influence and how big is the chance of spreading a tweet or opinion in this unlimited virtual space.

REFERENCES

[1] Zimmer C (26 February 2021). "The Secret Life of a Coronavirus – An oily, 100-nanometer-wide bubble of genes has killed more than two million people and reshaped the world. Scientists don't quite know what to make of it".

[2] "Clinical characteristics of COVID-19". European Centre for Disease Prevention and Control. Retrieved 2021-07-29.

[3] "Symptoms of Coronavirus". U.S. Centers for Disease Control and Prevention (CDC). 2021-02-22. Archived from the original on 2021-03-04. Retrieved 2021-07-29.

[4] Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, Wade RG (23 June 2020). "The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries". PLOS ONE. 15 (6): e0234765. Bibcode:2020PLoSO..1534765G. doi:10.1371/journal.pone.0234765. PMC 7310678. PMID 32574165. S2CID 220046286.

[5] Islam MA (November 2020). "Prevalence of Headache in Patients With Coronavirus Disease 2019 (COVID-19): A Systematic Review and Meta-Analysis of 14,275 Patients". Frontiers in Neurology. 11: 562634. doi:10.3389/fneur.2020.562634. PMC 7728918. PMID 33329305.

[6] Saniasiaya J, Islam MA (April 2021). "Prevalence of Olfactory Dysfunction in Coronavirus Disease 2019 (COVID-19): A Meta-analysis of 27,492 Patients". The Laryngoscope. 131 (4): 865 878. doi:10.1002/lary.29286. PMC 7753439. PMID 33219539.

[7] Islam MA (April 2021). "Prevalence and characteristics of fever in adult and paediatric patients with coronavirus disease 2019 (COVID-19): A systematic review and meta-analysis of 17515 patients". PLOS ONE. 16 (4): e0249788. Bibcode:2021PLoSO..1649788I. doi:10.1371/journal.pone.0249788. PMC 8023501. PMID 33822812.

[8] Ratcliffe, Rebecca (2 Maret 2020). "First coronavirus cases confirmed in Indonesia amid fears nation is ill-prepared for an outbreak". The Guardian.

[9] "Indonesia confirms first cases of coronavirus". Bangkok Post. Reuters. 2 Maret 2020.

[10] "Coronavirus Update Worldwide". Worldometer. Retrieved 30-07-2021.

[11] "Peta Sebaran". Komite Penanganan COVID-19 dan Pemulihan Ekonomi Nasional. Accessed on 30 July 2021.

[12] "Soal PSBB Jawa-Bali, Pemerintah Kenalkan Istilah PPKM". CNN Indonesia (in Indonesian). 7 January 2021. Retrieved 31 May 2021

[13] "Top Sites". Alexa Internet. Accessed on 2021-07-30.

[14] D'Monte, Leslie (April 29, 2009). "Swine Flu's Tweet Tweet Causes Online Flutter". Business Standard. Accessed on July 30, 2021. Also known as the 'SMS of the internet', Twitter is a free social networking and micro-blogging service.

[15] Nasution, Darmeli; Harumy, T. Henny F.; Haryanto, Eko; Fachrizal, Ferry; Julham; Turnip, Arjon (2015-10). "A classification method for prediction of qualitative properties of multivariate EEG-P300 signals". 2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT). IEEE. doi:10.1109/icacomit.2015.7440180. ISBN 978-1-4673-7408-8.

[16] Kurt Hornik, Maxwell Stinchcombe, Halbert White, Multilayer feedforward networks are universal approximators, Neural Networks, Volume 2, Issue 5, 1989, Pages 359-366, ISSN 0893-6080, http://dx.doi.org/10.1016/0893-6080(89)90020-8.

[17] Tuhuteru, Hennie. (2 July 2020). "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine". Information System Development. Volume 5.

[18] Addawood, Aseel; Alsuwailem, Alhanouf; Alohali, Ali; Alajaji, Dalal; Alturki, Mashail; Alsuhaibani, Jaida; Aljabli, Fawziah. 2020. "Tracking And Understanding Public Reaction During COVID-19: Saudi Arabia As A Use Case".

[19] Indarbensyah, Putri P.P.E.; Rochmawati, Naim. 2021. "Penerapan N-Gram menggunakan Algoritma Random Forest dan Naïve Bayes Classifier pada Analisis Sentimen Kebijakan PPKM 2021". JINACS. Volume 02. p235-244. ISSN: 2686-2220.

[20] del Fresno, Miguel; López-Pelaez, Antonio (2014). "Social work and Netnography: The case of Spain and generic drugs". Qualitative Social Work. 13 (1): 85–107. doi:10.1177/1473325013507736. S2CID 144911827.