



Sentiment Analysis of the Impact of Covid-19 on Indonesia's Economy through Social Media Using the ANN Method

Astri Pertiwi¹, Agung Triayudi², Endah Tri Esti Handayani³

¹Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Jalan Sawo
Manila, Jakarta Selatan, 12520, Indonesia

Email: ¹astripertiwi51@gmail.com, ²agungtriayudi@civitas.unas.ac.id, ³endahtriesti@civitas.unas.ac.id

*Email Penulis Korespondensi : agungtriayudi@civitas.unas.ac.id

ARTICLE INFO

ABSTRACT

Article history:

Received: 04/04/2020

Revised: 20/04/2020

Accepted: 30/05/2020

Keywords:

Artificial Neural Network;

Covid-19;

Sentiment Analysis;

Text Mining;

Machine Learning

Social media platforms and micro blogging websites are rich sources of user-generated data. The internet has changed the way people express their thoughts and feelings through social media, like Twitter. Twitter as a place for information flow, Twitter is a rich source for learning about people's opinions and sentiment analysis. This study aims to determine how much the impact of the co-19 pandemic on economic conditions in Indonesia, measured from the resulting community sentiment. To solve the problem in this case study using a method that is for pre-processing data using cleansing, filtering, tokenization and stemming. The classification process uses the Artificial Neural Networks (ANN) method. The data used are Indonesian-language tweets with the keywords Covid-19, Ekonomi Indonesia, MSME, storylineeverline and phk, with a dataset of 1600 tweets. The results of this study are sentiment analysis of the impact of co-19 on the Indonesian economy. Results of training data SMEs get an accuracy of 99.37% and Ojekonline data get an accuracy of 99.25%.

Copyright © 2020 Jurnal Mantik.

All rights reserved.

1. Introduction

Beginning in March 2020, Indonesia began to be infected with the covid-19 virus. The origin of this virus was identified as originating from Wuhan, China. Then this virus spread to various other countries in Europe such as Italy, Spain, England, Germany and others. The virus spreads to the Americas, namely the United States, Brazil and others. In addition to Europe and America, the virus spread to Asia, such as South Korea, Japan, Singapore, Malaysia, including Indonesia. Therefore, the World Health Organization (WHO) officially announced the covid-19 outbreak as a global pandemic. It was announced on March 11, 2020. According to data on the official website (www.worldmeter.info/coronavirus) This pandemic has spread to 210 countries in the world (accessed April 21, 2020). Since the emergence of the most positive corona / covid-19 virus cases in DKI Jakarta, the central government has taken steps to stop the spread of the virus, namely by calling for Work From Home, Social Distancing to Large-Scale Restrictions (PSBB). In addition to DKI Jakarta, the provincial government others also began to take the same steps to stop the spread of the virus from continuing to spread.

The impact of this pandemic causes reduced mobility. The author tries to research the impact of this pandemic affecting economic sectors, such as Micro, Small and Medium Enterprises (MSMEs), Ojog Online, to the risk of termination of employment. To solve the problem in the case study of sentiment opinion analysis of the impact of co-19 on the Indonesian economy by applying the Artificial Neural Networks (ANN) method. The research is measured from opinions / sentiments that are raised in the community through social media namely Twitter, not only positive or neutral opinions but also negative ones. The development of the pandemic is so rapid, which is very easily accessed through social media.

Social media platforms and micro blogging websites are rich sources of user-generated data. [1] The internet has changed the way people express their thoughts and feelings through social media. Twitter as a place for information flow, Twitter is a rich source for learning about people's opinions and sentiment analysis. [3] Sentiment analysis has now become the dominant approach used to extract sentiments and judgments about certain topics from online sources, which are divided into three categories namely positive, negative and neutral. [2]



There are several studies on sentiment analysis using Machine Learning, [11] in research on the same topic, namely covid-19 using LSTM Neural Networks modeling to obtain an accuracy of 81.15%. Then in [10] research on the classification review of higher education institutions with a comparison of 4 methods, namely Logistic Regression get an accuracy of 79%, 80% for SVM, and 79% for RFC, and the results of the best sentiment analysis accuracy given ANN get an accuracy of more than 86% . While this research will apply the ANN method to classify sentiment data, the method was taken from previous research [12] 'Sentiment Analysis and Prediction using Neural Networks' using the ANN method in sentiment analysis for 'iPhoneX' with 3 sources of social media platforms namely Twitter, Facebook and Website which get 87% accuracy, the use of ANN method in sentiment classification and analysis can overcome this problem, because Neural network types are very efficient in predicting results with high accuracy. The author also concludes that the Machine Learning approach and ANN implementation will result in a better classification.

The main objective of this research, however, is to show an application of sentiment analysis for NLP (based on the Artificial Neural Networks model) to find out how big the impact of the covid-19 pandemic on economic conditions in Indonesia. The author believes that the results of this study will help in understanding community concerns and needs with respect to issues related to covid-19. In addition, the authors' findings can help in improving the market strategy at umkm, as well as ojol affected by the covid-19 pandemic.

2. Research methods

2.1 Flow chart

In the sentiment analysis analysis of the impact of co-19 on the Indonesian economy, opinion is needed on Twitter social media in the form of a dataset, then it will go through a processing (pre-processing). Figure 1 shows the framework for sentiment analysis.

2.2 Data Crawling

In this process the data needed is public opinion on Twitter social media in the form of a dataset containing various Tweets. This process uses the Crawling technique, with the Tweepy library in the python programming language that is run with the Jupyter Notebook tools. Tweets used in Indonesian, which contain positive and negative sentiments.

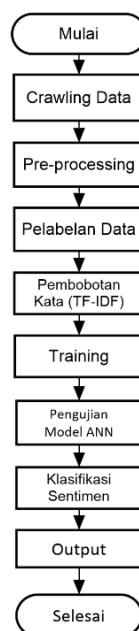


Fig 1. Flow chart System

The function of the Flowchart system above is to describe the steps of a sentiment analysis analysis of the impact of co-19 on the Indonesian economy.

2.3 Pre-processing

Pre-processing is the initial stage in processing the dataset. Before the dataset can proceed to the model,

the data must proceed to the pre-processing step, to remove irrelevant words that do not represent any sentiments such as punctuation, symbols, and links. [4]

Following are the stages in pre-processing, namely:



Fig 2. Pre-processing

a. Comment Selection

At this stage, selection of comments containing the word or hashtag (#) covid-19, EkonomiIndonesia, MSME, storylineeverline and phk. [5]

b. Cleansing

The sentences obtained are usually still noise, namely random errors or variants in the measured variable, for that, we must eliminate the noise. The omitted words are HTML characters, keywords, emotion icons, hashtag (#), username (@username), url (http://website.com), and email (name@website.com). [5]

c. Case Folding

Case folding is a stage in Pre-processing that aims to change each word form to be the same. This is done by changing the word to lower case or lowercase letters. [6]

d. Tokenization

The process of decomposing the entire text into words. For example, the phrase "information and communication technology" after tokenized will be: "technology", "information", "and", "communication". [8]

e. Stopword Removal

This process is used to eliminate words that are common but have no meaning (stopwords). Stopword removal is done using the Literature library. [6] For example: "who", "in", "to", "or", "there", "joined", "then", "and", etc. [4] [8]

f. Stemming

Stemming is a process to find the stem (basic word) from the word stopword removal (filtering). There are 2 rules for doing stemming, namely the dictionary approach and the rule approach. [8]

2.4. Data Labeling

Documents contained on the internet are uncontrolled data labeled (unsupervised data). So to be processed using supervised learning, a method is needed to label these data. The obstacle that arises is the amount of data that must be labeled, namely by calculating the amount of text contained in each document. [9] The labeling process is carried out as follows:

- a. Determine words that have positive as well as negative meanings.
- b. If the number of positive words > the number of negative words, the sentiment label is positive (label 1).
- c. If the number of positive words < the number of negative words, the sentiment label is positive (label 0).

2.5. Word Weighting (TD-IDF)

Data that has been processed certainly still contains words that are difficult to be processed by most machine learning classification algorithms. This is because most classification algorithms require data that is represented as a numeric value. Therefore, we will change the textual data that has gone through the pre-processing process into numerical form using the term-frequency-inverse-document-frequency (TF-IDF) technique. [10] The stages of weighting with TF-IDF are:

- a. Calculate the term frequency (tf t, d)
- b. Calculate weighting term frequency (Wtf)

$$Wtft, = \begin{cases} 1 + \log_{10} tft, d, & \text{if } tft, d > 0 \\ 0, & \text{otherwise} \end{cases} \quad [6]$$

- c. Calculate document frequency (df)
- d. Calculate the inverse document frequency (idf) weight

$$idf = \frac{\log_{10} N}{dft}$$

5. Calculate the weight value of TF-IDF [6]

$$W_{t,d} = W_{tft,d} \times id_{ft}$$

Information:

$t_{ft,d}$ = term frequency, $W_{tft,d}$ = term frequency weight

id_{ft} = number of documents containing the term

N = total number of documents

$W_{t,d}$ = TF-IDF weight

2.6. Artificial Neural Networks (ANN)

Artificial Neural Network (ANN) is a set of algorithms that are inspired by the workings of the biological nervous system, especially on human brain cells in processing information. ANN is one of the algorithms of Machine learning, ANN is also part of the discipline of Artificial Intelligence (Artificial Intelligence)

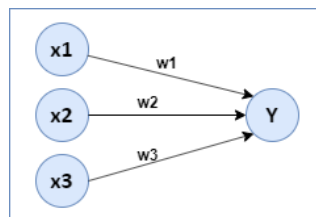


Figure 3 Simple Artificial Neural Networks

Y accepts input from neurons x1, x2, x3 with relationship weights w1, w2, and w3 in sequence. The three neuron impulses that are added after each input are multiplied by their respective weights, where:

$$Y_{in} = \sum x_i \cdot w_i / \text{net} = x_1 w_1 + x_2 w_2 + x_3 w_3$$

Y_{in} = The output, x_i = Input, w = Weight

So the magnitude of the impulse received by Y follows the activation function $y = f(\text{net})$. If the value of the activation function is strong enough, then the signal will be forwarded. The activation function value as the output of this network model can also be used as a basis for changing weights.

2.6. Single Layer Perceptron

Single Layer Perceptron is one of the simplest artificial neural network techniques, which consists of an input layer and an output layer. The basic elements of ANN consist of 3 main parts: weight, threshold, and activation function.

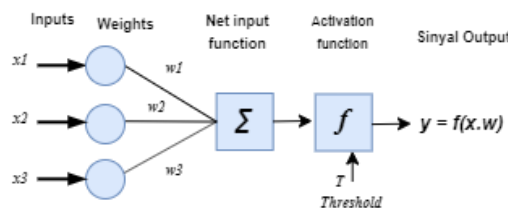


Figure 4 Single Layer Perceptron

The network above has 3 input values (x1, x2, x3) that will be processed, with each input going through a weighted path, then all the values are combined \sum the combined value is then processed by the activation function $f[]$ to produce the Y signal as the output. The activation function $f[]$ uses a certain threshold value to limit the output value to always be within the specified value limit.

2.7 ANN Training Process

Following are the steps in processing single perceptron learning, the training process is carried out using the perceptron learning rules as follows:

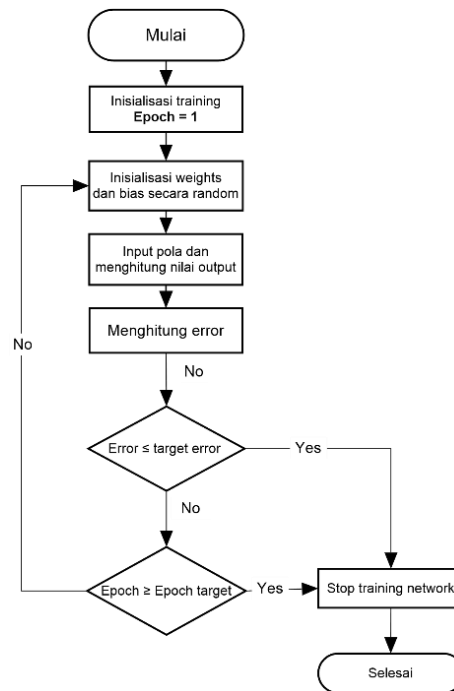


Fig 5. Flow chart ANN Training

- Initial initialization of training using epoch values starts from 1 epoch.
- Initialization sets the weights value, biased randomly.
- Pass the input to the neuron, then it will get an output value or called feedforward.
- Counting errors, if the smaller error is the same as the target error, then stop the training process. If not, then change the epoch value until it gets the smallest error.
- Furthermore, if the epoch value is greater than the target, the training process is stopped, but if you have not gotten the results according to the target, change the epoch value.
- Repeat these steps until there is no change in the error value, the error value \leq a threshold (usually close to 0), or have repeated the exercise process for T times (threshold).

3. Results and Discussion

In this study, the writer will describe the stages in the classification process to analyze the sentiment opinion of the impact of co-19 on the Indonesian economy. The dataset used was taken from Twitter social media using the Crawling technique. Tweets used in Indonesian, which contain positive and negative sentiments. Tweet with keyword or hashtag (#) covid-19, coronavirus, EkonomiIndonesia, UMKM, ceritadriversonline and phk. The classification process is using the ANN method. And processed by python programming on notebook jupyter tools.

3.1. Pre-processing Data

The data that has been obtained from the crawling process with several hashtags, is then processed by the data, which aims to facilitate the next process. Pre-processing used is cleansing, filtering, tokenization and stemming.

Table 1
Pre-processing

Before proceeding	After processing
b'RT @BerliandoHD: Ttwitter do your magic. \n Inviting these SME actors affected by COVID-19, we creative workers are ready to help, \xe2 \x80 \xa6'	berliandohd twitter your magic invites fellow colleagues umkm impact covid creative work ready to help
b'Cerita Ojol When Pandemic Corona, Income Drops Drastically, Fear Cannot Install Motorcycle Credit \xe2 \x80 \xa6 https://t.co/vjNYH0cpJa'	the story of ojol pandemic corona can plummet dramatically in fear of not repaying motorcycle loans

After pre-processing is done, then labeling process is done with label 1 for words that contain positive sentiment and label 0 for words that contain negative sentiment. The next process is weighting the word with the TF-IDF method, to calculate the terms in the document.

3.3 Training Process

At this stage, ANN will go through a training process with a ready dataset. Training and testing data used were 1600 tweets, with the hashtag category of 600 tweets and ojekonline 00 tweets, each of which contained a balanced amount of positive and negative sentiment data. In this process, previously the training process was carried out on the dataset to get accuracy, following the process

a. 1st attempt

In the first experiment, data training was carried out 5 times, using different epoch values to find out how much influence the epoch had. The dataset in the first try was in the MSME category.

TABLE 2
1ST TRAINING TRIAL

Training to	Batch Size	Epoch	Accuracy	loss
1	20	10	99.12%	0.0148
2		20	99.12%	0.2387
3		30	99.37%	-1.0294
4		40	99.25%	-2.7179
5		50	99.37%	-5,677
6		100	98.75%	-41,976

From the above experiments it can be seen that the 40th epoch has shown optimum accuracy results. Epoch value influences in obtaining the highest accuracy value. The author tries to add the epoch value to 100 but the results are not far off if doing the 50th epoch.

b. 2nd attempt

In the second experiment to continue the previous process, using the epoch value of 50. At this stage the author tries to use a different batch size for each training. This was done to find out how much influence the value of the batch size in the training process. Dataset in the 2nd trial with the Ojek category online.

TABLE 3
2ND TRAINING TRIAL

Training to	Batch Size	Epoch	Accuracy	loss
1	10	50	99%	0.0088
2	8		99%	0.087
3	4		98.87%	0.087
4	2		99.25%	0.0091
5	1		99%	0.087

From the results of the 2nd experiment, it can be seen that the value of the batch size affects the increase in accuracy results. If the smaller the batch size value, the better the accuracy obtained. However, if the value of the batch size gets smaller, the training process takes a long time. From the 1st and 2nd experiments of the training process above, the model that gets the highest accuracy results will be used in the next process, namely the sentiment classification testing process. Figure 6

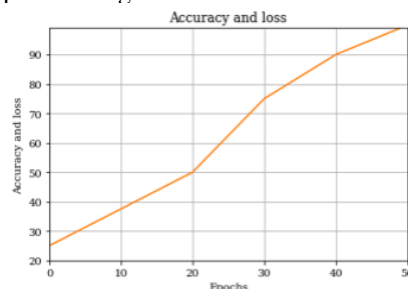


Fig 6. Graph of UMKM Training Results

The training model with an accuracy of 99.37% and a loss of 5.6% is the model of the dataset with the MSME category with the number of epoch 50 with batch size 20. Based on the first training with 50 epochs, training with the same epoch is carried out in the training process the second uses the Ojekonline category dataset.



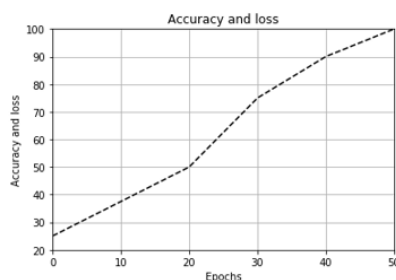


Fig 7. Ojekonline Training Results Graph

In this process, training with an accuracy of 99.25% and a loss of 0.91% is the model of the dataset with the Ojek online category with the number of epoch 50 with batch size 20. The next process the results of this training model will be used in the testing process.

c. Testing Process

In this testing process, the testing of the model data is carried out by carrying out the implementation of the model to determine the classification of data about the sentiment analysis of the impact of co-19 on Indonesia's economy. The data will be tested in the amount of 100 tweets taken randomly. The results of sentiment analysis testing on the 'umkm' and 'ojekonline' models will be shown in table 4.

TABLE 4
RESULTS OF SENTIMENT ANALYSIS WITH ANN

Sentiment	MSME	Ojekonline
Positive	83	77
Negative	17	23

In the results of the classification analysis of the co-19 impact opinion sentiment analysis on the 2 models it is known that the umkm and ojol sectors are indeed affected by the covid-19 pandemic but based on tweets in April within 7 days (provisions of the Twitter API) contain positive sentiments about Microbusiness Small and Medium Enterprises (MSMEs) namely solutions and marketing survive in the middle of the covid-19 pandemic, indicating that the umkm category outperformed the ojekonline category.

4. Conclusion

Based on research that has been done, it can be concluded that:

- a. Machine Learning, especially sentiment analysis, can be used to classify opinion sentiments of the 19th impact of Indonesia's economic conditions, which are focused on the umkm and ojekonline sectors.
- b. This study obtained the highest accuracy, namely 99.37% in umkm and 99.25% in the ojekonline category, by applying the Artificial Neural Networks (ANN) method. The results of this study are expected to be used as information on market solutions / strategies on the umkm and ojekonline affected by the co-19 pandemic so they can survive the economic crisis. And this research is expected to be able to provide information that not many people know so that this research will help in understanding the concerns and needs of the community in connection with issues related to covid-19
- c. For further research, developing research requires large amounts of data, as well as the time period of the data obtained in order to produce more detailed research, in accordance with the period of this pandemic impacting on economic conditions in Indonesia. Development of the method used can also use other Neural Networks methods such as Backpropagation Neural Networks.

5. Reference

[1] M. Ahmad, S. Aftab, SS Muhammad, and S. Ahmad, 'Machine Learning Techniques for Sentiment Analysis: A Review', *International Journal of Multidisciplinary Sciences and Engineering*, Vol.8, No.3, April (2017), [ISSN: 2045-7057].



- [2] S. Bharathi and A. Geetha, '*Sentiment Analysis for Effective Stock Market Prediction* ', *International Journal of Intelligent Engineering & Systems, ijies2017*', Vol. 10, No.3, (2017)
- [3] C. Kaur and A. Sharma, '*Twitter Sentiment Analysis on Coronavirus using Textblob* ', *EasyChair Preprint*, No.2974, March 16 (2020)
- [4] MO Pratama, W.Satyawan, R.Jannati, B.Pamungkas, Raspiani, ME Syahputra and I. Neforawati, '*The sentiment analysis of Indonesia commuter line using machine learning based on twitter data*', *2018 International Conference of Computer and Informatics Engineering (IC2IE), IOP Conf. Series: Journal of Physics: Conf. Series* 1193 (2019) 012029
- [5] GA Buntoro, '*DKI Jakarta Governor Candidate Sentiment Analysis 2017 on Twitter*', *Integer Journal*, Vol. 2, No. 1, March (2017): 32-41
- [6] WA Luqyana, I. Cholissodin, and Prime Hospital, '*Analysis of Cyberbullying Sentiments on Instagram Comments by the Support Vector Machine Classification Method*', *Journal of Information Technology Development and Computer Science*, Vol. 2, No. 11, p. 4707-4713 November (2018)
- [7] A. Triayudi, '*Convolutional neural networks for text classifications on Twitter*', *Journal of Software Engineering & Intelligent Systems*, Vol. 4, 31 December (2019), 2016-2019, ISSN: 2518-8739
- [8] WC Frans Mariel, S. Mariyah, and A. Permana, '*Sentiment analysis: a comparison of deep learning neural network algorithms with SVM and nave Bayes for Indonesian text* ', *International Conference on Data and Information Science*', *IOP Conf. Series: Journal of Physics: Conf.Series* 971(2018)
- [9] A. Futuhul Hadi, D. Bagus C. W, and M. Hasan, '*Text Mining on Twitter Social Media Case Study: Quiet Period 2017 Round 2 of DKI Jakarta Election, National Seminar on Mathematics and Its Applications*', October 21 (2017)
- [10] H. Akbar, '*Sentiment Analysis Review of Higher Education Institutions From Facebook Using Artificial Neural Networks*', *JIK: Journal of Computer Science*, Vol.4, No.1, June (2019).
- [11] H. Jelodar, Y. Wang, and R. Orji, '*Deep Sentiment Classification and Topic Discovery on Coronavirus Novels or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Networks Approach*', *bioRxiv preprint*, April 24 (2020) .
- [12] S. Paliwal, S. Kumar Khatri, and M. Sharma '*Sentiment Analysis and Prediction using Neural Networks, Proceedings*

