Analysis of Twitter Users Sentiment against the Covid-19 Outbreak Using the Backpropagation Method with Adam Optimization

Theresia Hendrawati[1], and Christina Purnama Yanti[2]

[1,2] Informatics Department, STMIK STIKOM Indonesia
Denpasar - Bali, Indonesia
theresia.hendrawati@stiki-indonesia.ac.id

Abstract - This research tries to take advantage of Twitter by analyzing Indonesian-language tweets that discuss the Covid-19 virus outbreak to find out what Twitter users think about the Covid-19 virus outbreak. This study tries to analyze sentiment to see opinions on Covid-19 tweets that contains Posittive, Negative or Neutral sentiments using Multi-layer Perceptron (MLP) using Backprograration with Adam optimization. We collected 200 documents (tweets) in Indonesian about Covid-19 that were tweeted since November 2019 and then trained them to get our MLP Backpropagation model. Our model managed to get an accuracy of up to 70% with f1-scores for positive, negative, and neutral classes respectively 0.77, 0.75, and 0.5 from a maximum value of 1. This shows that our model is quite successful in carrying out the sentiment classification process for Indonesian tweets with the Covid-19 theme.

Index Terms—COVID-19, Backpropagation, Twitter, Sentiments.

I. INTRODUCTION

Social media is an entertainment medium that is of concern to its users. Various important information about politics, economy, society, and culture is presented by social media [1]. Twitter is one of the most popular social media in society today [2]. Based on a survey, the number of Twitter users in Indonesia in 2010 was at the top compared to other countries, with 20.8% of users accessing Twitter.com [3]. However, the large number of users on social media does not necessarily make the information disseminated reliable by its users. At the beginning of 2020, the world was shocked by the outbreak of a new virus, namely the new type of coronavirus (SARS-CoV-2) and a disease called Coronavirus disease 2019 (COVID-19). The origin of this virus comes from Wuhan, China. Until now, it is confirmed that there have been 65 countries infected with this one virus [4]. This research tries to take advantage of Twitter by analyzing Indonesian-language tweets that discuss the Covid-19 virus outbreak to find out what Twitter citizens think about the Covid-19 virus outbreak. To get accurate results, you need the right algorithm and method. Therefore an API from the Twitter platform is needed to retrieve data, after obtaining the data, it is followed by sentiment analysis which is useful for analyzing comments on Twitter to be translated into something more meaningful [1]. Several similar studies have been carried out, such as Brian Adrianto's research entitled "Analysis of Radical Content Sentiment Through Twitter Documents Using the Back Propagation Method" which discusses the content analysis on Twitter related to radical content that can harm society [5]. Another research by Maharadhika with the research title "Analysis of Sentiment against Joko Widodo's Government on Twitter Social Media Using the Naive Bayes Classifier Algorithm" about the opinions of Indonesian citizens expressed on social media in the form of praise, criticism, and insults. Based on previous research, this study tries to conduct sentiment analysis to see an opinion or opinion on a problem or object that contains negative, neutral, or positive sentiments using the Backprograration algorithm with Adam Optimization.

II. RELATED WORKS

Previous studies have conducted similar studies. Adrianto's research with the title "Analysis of Radical Content Sentiment through Twitter Documents Using the Back Propagation Method". This study discusses public opinion that is too excessive, causing radical actions on social media which can harm several parties. Therefore, tweets are classified on Twitter to share positive and negative content [5]. Another research is Syadid's research entitled "Analysis of Netizen Comment Sentiment on the 2019 Indonesian Presidential Candidates from Twitter Using the Term-Frequency-Inverse Document Frequency (TF-IDF) Algorithm and the Multi-Layer Perceptron Neural Network (MLP) Method". This study classifies positive and negative

sentiments in the tweets of the 2019 Indonesian presidential candidates by using 3 training data training scenarios, namely scenario 1 with 700 data, scenario 2 with 800 data, and scenario 3 with 900 data and the highest score is obtained in scenario 3 [6]. Another research is Mahardhika's research entitled "Analysis of Sentiment Against Joko Widodo's Government on Twitter Social Media Using the Naïve Bayes Classifier Algorithm" which classifies public opinion regarding Joko Widodo's government whether it is praising, criticizing, or insulting. The data used amounted to 400 tweets where 300 training data and 100 test data were obtained 96% for negative sentiment and 98% for the positive sentiment [7].

Past studies [8], [9] have also shown that one of the approaches that can be used as a feature of texts and sentences in documents is tf-idf. Therefore, we decided to try to use this feature to conduct a sentiment analysis in this study.

III. PROPOSED APPROACH

In this study, we analyze the sentiment of the Covid-19 problem in the community using data taken from *Twitter*. We use Multi-layer Perceptron (MLP) architecture, Backpropagation with Adam optimization. The data we use is taken from *Twitter* using the *Twitter* API, the data collected is 200 data taken from several twitter tags: Covid, Covid-19, Coronavirus which were tweeted from November 1, 2019, to August 30, 2020. This data is then labeled manually into 3 classes of 'Positive', 'Negative', and 'Neutral'.

After getting text data from *Twitter*, the next step is to do pre-processing steps to get text data that we will look for features. The data pre-processing steps that are passed are as follows:

1. Performing the process of removing elements in the text, namely numbers, punctuation marks, and special characters.
2. Perform the process of eliminating stop words in sentences.
3. Perform the stemming process to get basic words.
4. The process of tokenizing to get the words in a sentence.

After getting the results of pre-processing data, we perform the feature extraction process. The feature extraction process carried out in this study is to look for the value of the Term Frequency and Inverse Document Frequency.

Term Frequency (*tf*) gives us the frequency of the word in each document in the corpus. It is the ratio of the number of times the word appears in a document compared to the total number of words in that document. It increases as the number of occurrences of that word within the document increases. Each document has its own *tf*. Equation (1) shows the equation for calculating the value of *tf*.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}} \tag{1}$$

Inverse Data Frequency (*idf*): used to calculate the weight of rare words across all documents in the corpus. The words that occur rarely in the corpus have a high *idf* score. It is given by the equation below. Equation (2) shows how we calculating the value of *idf*.

$$idf(w) = \log(\frac{N}{df_t}) \tag{2}$$

Combining these two (*tf* and *idf*) we come up with the TF-IDF score (*w*) for a word in a document in the corpus. It is the product of following Equation (3).

$$w_{i,j} = tf_{i,j} \times \log(\frac{N}{df_i}) \tag{3}$$

Where *tf*_{i,j} is number of occurrences of i in j, *df*_i is number of document containing i and N is total of number documents.

After the feature data (*tf-idf*) for each document (tweet) is obtained, then we can use this feature to carry out the classification model training process that we propose. We built a classification model using the Multi-layer Perceptron model using Backpropagation with 100 hidden layers, a logistical activation function, and Adam as the optimizer. We use the Backpropagation to classify documents to three classes (Back propagation output is 3). Of the 200 document dataset, 60 pieces of data (30%) were used as validation data, while 140 (60%) of other data were used for the training process.

IV. RESULT

After going through 470 epochs in the training process using the training data that had been prepared, the MLP Backpropagation that we built succeeded in achieving convergence. This model produces an accuracy of 70% with the f1-score for the 'Positive' class is 0.77, the 'Negative' class is 0.75 and the 'Neutral' class is 0.5. Table 1 shows the performance of the model we built.

TABLE I
MLP BACKPROPAGATION PERFORMANCE

Class	Precision	Recall	F1-Score	Documents
Positive	0.71	0.83	0.77	24
Negative	0.71	0.79	0.75	19
Neutral	0.64	0.41	0.50	17

TABLE II
MLP BACKPROPAGATION PERFORMANCE SUMMARY

Class	Precision	Recall	F1-Score	Documents
Accuracy			70	60
Macro Avg	0.69	0.68	0.67	60
Weighted Avg	0.69	0.70	0.69	60

From the information shown in Table 1 and Table 2 we know that of the 60 test data provided, 70% can be classified properly by the model we built. To find out how well the model is performing, we also look at the f1-score value from

the test data provided, for the 'Positive' class the f1-score obtained is 0.77 from a maximum of 1, for the 'Negative' class it gets a value of 0.75 while for the 'Neutral' class got not so good score at 0.5.

To make it easier to see the data from the validation results that we do, we display Figure 1 with a confusion matrix from the test data used to validate our model.
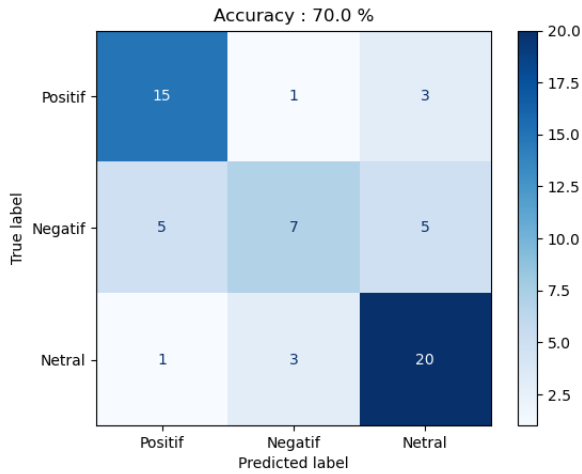

Fig. 1. Confusion Matrix of proposed Model

To see how the learning curve of the model we built can be seen in Figure 2. Figure 3 and Figure 4 also give us useful information about our proposed model training.
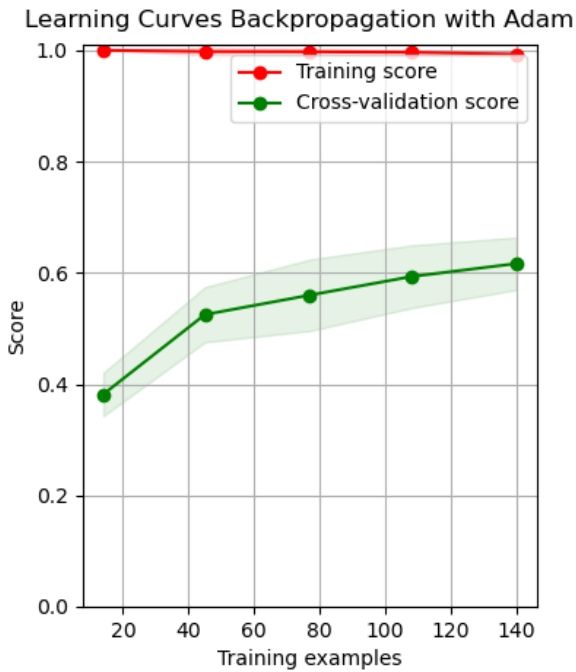
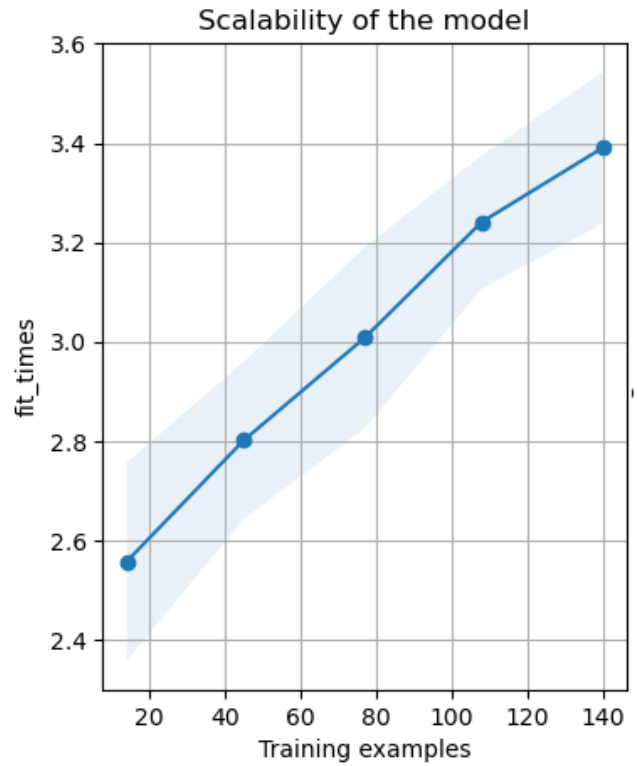Fig. 2. Learning curve of proposed model
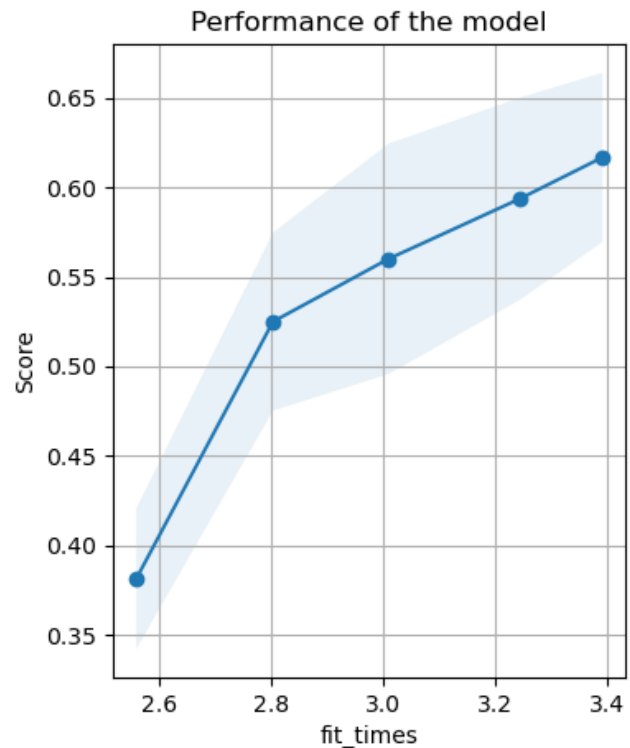

Fig. 3. Scalability of proposed Model

Fig. 4. Performance plot of proposed Model

V. CONCLUSION

This study shows that the MLP Backpropagation model generated from this study is able to classify the sentiments of Indonesian-language tweets that have the Covid-19 theme.

This model was built using a tweet dataset that has 3 different sentiments, namely positive, negative, and neutral. We found the MLP architecture, using logistic activation functions and the Adam optimization model to be good enough to build a classifier by which to perform the sentiment analysis process. Our models have an accuracy of up to 70%.

Although this approach and model has had considerable success in recognizing sentiment in documents/tweets, it needs to be seen again, this model still has weaknesses that can be seen in the resulting f1-score in the 'Neutral' class. This opens up opportunities for further research considering that our dataset may be too small to train sentiment in 'Neutral' class output.

REFERENCES

[1]　M. A. Maulana, A. Setyanto, and M. P. Kurniawan, "Analisis Sentimen Media Sosial Universitas Amikom Yogyakarta Sebagai Sarana Penyebaran Informasi Menggunakan Algoritma Klasifikasi SVM," *SEMNASTEKNOMEDIA ONLINE*, vol. 6, no. 1, pp. 1–2, 2018.

[2]　A. F. Hidayatullah and A. S. Azhari, "Analisis Sentimen dan Klasifikasi Kategori terhadap tokoh publik pada twitter," in *Seminar Nasional Informatika (SEMNASIF)*, 2015.

[3]　T. Pramiyati, A. Purwarianti, and I. Supriana, "KECENDERUNGAN PENILAIAN PENGGUNA INFORMASI TERHADAP TWEET (KICAUAN) PADA MEDIA SOSIAL TWITTER," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 7, no. 1, pp. 209–216, 2016.

[4]　Y. Yuliana, "Corona virus diseases (Covid-19): Sebuah tinjauan literatur," *Wellness Heal. Mag.*, vol. 2, no. 1, pp. 187–192, 2020.

[5]　B. Andrianto and S. A. Indriati, "Analisis Sentimen Konten Radikal Melalui Dokumen Twitter Menggunakan Metode Backpropagation," *J. Pengemb. Teknol. Informasi dan Ilmu Komput.*, vol. 2, no. 12, pp. 7380–7385, 2018.

[6]　F. Syadid, "Analisis sentimen komentar netizen terhadap calon presiden Indonesia 2019 dari twitter menggunakan algoritma term frequency-invers document frequency (tf-idf) dan metode multi layer perceptron (mlp) neural network," 2019.

[7]　Y. S. Mahardhika and E. Zuliarso, "ANALISIS SENTIMEN TERHADAP PEMERINTAHAN JOKO WIDODO PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVES BAYES CLASSIFIER," *SINTAK*, vol. 2, no. Nov, 2018.

[8]　I. W. A. Setyadi, D. C. Khrisne, and I. M. A. Suyadnya, "Automatic Text Summarization Menggunakan Metode Graph dan Ant Colony Optimization," *Maj. Ilm. Teknol. Elektro*, vol. 17, no. 1, pp. 124–130, 2018.

[9]　D. C. Khrisne and I. M. A. Suyadnya, "RUNCING: an Indonesian Text Summarization System Using Cat Swarm Optimization."