Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Sixth Information Systems International Conference (ISICO 2021)

The influence of fake accounts on sentiment analysis related to COVID-19 in Indonesia

Rivanda Putra Pratama, Aris Tjahyanto*

Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Abstract

Cases of the spread of COVID-19 that continue to increase in Indonesia have made the level of public satisfaction with the government in dealing with this virus fairly low. One way to measure the level of community satisfaction is by analyzing social media. Sentiment analysis can be used to analyze feedback from the public. Research related to sentiment analysis has been mostly carried out, but so far, it has focused more on opinions contained in sentences and comments and has not considered the subject of the account that posted it. On the other hand, the use of fake accounts or bots on social media is becoming more and more prevalent, so that the credibility of opinion makers is reduced. Based on these problems, this research conducted several experiments related to sentiment analysis using a machine learning approach and fake account categories to see the influence of fake accounts on sentiment analysis. The data used in this research were taken from social media Twitter. The results showed that there was an influence from fake accounts that can reduce the performance of sentiment classification. The experimental results of the two algorithms also prove that the Support Vector Machine algorithm has a better performance than the Naïve-Bayes algorithm for this case with the highest Accuracy value of 80.6%. In addition, the results of the sentiment visualization showed that there was an influence from fake accounts which actually leads to positive sentiment although it is not significant.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the Sixth Information Systems International Conference.

Keywords: Sentiment analysis; fake account; twitter; machine learning

* Corresponding author. Tel.: +62 81331550105.
E-mail address: aristj@its.ac.id
1. Introduction

Coronavirus Disease 2019 (COVID-19) is a heartbreaking disaster that affects all countries in the world, including Indonesia. Based on data from the official website of the COVID-19 Handling Task Force, there are around 67 million people exposed to COVID-19 and around 1.5 million people died. Meanwhile, in Indonesia, there are nearly 600,000 people affected and around 18,000 people died because of this virus [1]. At the beginning of the emergence of this pandemic, several countries implemented lockdown policies to prevent the spread of COVID-19. This policy has an impact on all segments of human life, starting from education, transportation, tourism, and the worst is the economy. Many schools, offices, and tourist attractions have been closed to massively reduce physical contact between people. People are encouraged to work from home if it is still possible. In the Indonesian economic sector, it also shows a continuous decline, until the end of 2020, Indonesia runs into a recession.

To accelerate the handling of COVID-19 and to recover and transform the national economy, the Indonesian government has formed a Task Force. The government uses online media as a medium of communication and information to the public to support the tasks of the COVID-19 Handling Task Force. Information technology that is increasingly developing makes any information related to COVID-19 easily accessible, anytime and anywhere. The results of research conducted by HootSuite and We Are Social from January 2019 to January 2020, show that internet users in Indonesia are currently increasing by 17% from the previous year. It has been noted that active internet users currently reach 175 million people or 64% of the total population, and active users of social media reach 160 million people or 59% of the total population [2].

The increasing number of internet users in Indonesia shows the era of information openness which is increasingly widespread, making it easier for people to obtain information from the internet. This encourages the government to disseminate information related to COVID-19 through online media. In addition, online media can also be used to get feedback from the community through social media. Based on this feedback, the government can measure the level of satisfaction of the community with the government's performance in dealing with COVID-19.

Feedback from the community can be interpreted as either satisfaction or dissatisfaction. Satisfaction from the community can be measured as a positive opinion. Meanwhile, dissatisfaction from the community can be measured as a negative opinion. Both forms of opinion from this feedback can be classified using sentiment analysis techniques. Sentiment analysis techniques in some studies are also known as opinion mining. Sentiment analysis is a technique for analyzing opinions, sentiments, appreciation, emotions about a product, service, organization, individual, and their attributes [3]. Sentiment analysis techniques can also be used in analyzing people's opinions of a piece of text to determine whether the sentiment is positive, negative, or neutral [4].

There has been a lot of research on sentiment analysis. However, research so far has focused more on the opinions contained in sentences and comments and does not consider the subject of the account that posted it. On the other hand, the use of fake accounts or bots on social media is becoming more and more prevalent, so that the credibility of opinion makers is reduced. To solve this problem, research was carried out related to the influence of fake accounts on sentiment analysis. Some tests will be conducted to get better sentiment classification performance. The results of sentiment visualization using fake accounts and without using fake accounts will show the influence of fake accounts on sentiment analysis.

2. Related Works

Most research on sentiment analysis uses a machine learning approach. The machine learning approach uses a number of features and algorithms to generate classification classes from the processed data [5]. Eliacik and Erdogan [6] conducted research related to microblogging obtained from Twitter. They use the Naïve-Bayes algorithm and Support Vector Machine as a classification algorithm. They also compared the initial classification results with the classification results combined with the PageRank, InterRank, and XiangRank methods to measure the influence of influential users. Kolog et al. [7] researched sentiment analysis and social phenomena about student life. They collected a dataset of student life through distributed questionnaires. They divide clusters of student life data using the K-Means algorithm. Furthermore, they compare the results of sentiment classification using several algorithms, including Sequential Minimum Optimization, Multinomial Naïve-Bayes, and J48 Decision Tree. Chouchani and Abed
[8] researched social influences that can improve sentiment analysis on social networks. They analyzed famous figures through their social media accounts. They use the SampleRank algorithm to measure the influence of famous figures.

In addition to the machine learning approach, several kinds of research use the lexicon-based approach to classify sentiments. The lexicon-based approach generally uses a dictionary as a basis for classifying sentiments. Bae and Lee [9] use a lexicon-based approach to conduct sentiment analysis on the Twitter dataset to see public sentiment towards microblogging from popular figures. They also conducted an analysis related to the influence of popular figures through user graphs of their social networks. Bravo-Marquez et al. [10] used a lexicon-based approach with the Annotate-Sample-Average algorithm to generate training data from the Twitter dataset.

Research related to the detection of fake accounts on social media is also quite a lot. Some of the approaches that are often used include filtering, rules, and machine learning [11]. Machine learning is the approach that has the best results in detecting fake accounts. In a research conducted by Akyon [12], a number of features of accounts deemed fake were used as features. Furthermore, a number of account datasets are classified and compared using the Naive Bayes algorithm, Support Vector Machine, Logistic Regression, and Neural Network. Velayutham and Tiwari [13] in their research used a friend-to-follow ratio called reputation, as well as a number of features of fake accounts as features. The dataset of the accounts is classified using the Naive Bayes algorithm. Ersahin et al. [14] in his research developed a discretization technique for a number of features of fake accounts. This technique is called Entropy Minimization Discretization and aims to improve the performance of the Naive Bayes algorithm in the classification of fake accounts. Jia et al. [15] developed a new method called SybilWalk to overcome some of the limitations of the Random Walk algorithm in classifying fake accounts.

In summary, this paper will use a machine learning approach for sentiment analysis. This paper will also compare the Naive Bayes algorithm and the Support Vector Machine in sentiment classification. The difference between the related work and this paper is that we will research the influence that fake accounts have on the results of a sentiment analysis regarding the performance of the Indonesian government against COVID-19.

3. Methodology

There are five steps that we will do in this paper. Each step in this methodology is carried out to test whether there is an influence of fake accounts on sentiment analysis, both on the performance of sentiment classification and overall sentiment results. The first step is data collection. The tweets data were collected as a dataset in this paper. The tweets data were taken through comments on the Twitter account of the Ministry of Health of the Republic of Indonesia (@KemenkesRI) regarding COVID-19. This data source was chosen because most of the information, news, feedback, and public complaints related to COVID-19 in Indonesia are on the Twitter account of the Ministry of Health of the Republic of Indonesia, so it can be assumed that tweets on this account can represent all Indonesian people. Tweet data is taken using a crawling technique with several parameters. Detailed parameters and descriptions of tweet data can be seen in Table 1.

| Parameters    | Descriptions                                |
|--------------|---------------------------------------------|
| Conversation ID | Conversation ID of fetched tweet data.     |
| Username     | Username of Twitter account.                |
| Date         | Date when the tweet was made.               |
| Time         | Time when the tweet was made                |
| Retweet      | Status that shows whether the tweet made is a retweet or not. |
| Tweet        | The contents of the tweet.                  |

The tweets data were taken from January 2021 to June 2021. The total tweets data obtained were 4,170 tweets. The second step is to remove fake account tweets. Fake and real account data will be identified using the Tweetbotornot tool. Tweetbotornot is a tool that can detect whether a Twitter account is a social bot [16]. Tweetbotornot was
developed with the R programming language. The probability scale of this tool starts from 0 to 1 with a statement that Twitter accounts that have a probability value above 0.5 will be identified as social bots. This tool will work based on the username contained in the tweet data. This tool uses three parameters that are used as a reference in identifying the authenticity of the account. Details of account authenticity reference parameters and features of Tweetbotornot can be seen in Table 2.

| Parameters                  | Features                                      |
|-----------------------------|-----------------------------------------------|
| User-Level                  | Account Name, Account Creation Date, Number of Tweets, Number of Followers, and Number of Following |
| Top-Level Tweeting Patterns | Frequency of Tweets Sent, Number of Original Tweets (Non-Retweet), and Number of Likes |
| Text-Based Patterns in User Tweets | Number of Retweets, Number of Mentions, and Number of Hashtags |

The results of identifying the authenticity of the account using Tweetbotornot obtained 1,952 tweets data originating from fake accounts. Furthermore, the tweets data originating from fake accounts were deleted so that a dataset without fake accounts was obtained with a total of 2,218 tweets. There are three parameters obtained from the identification of fake accounts which can be seen in Table 3.

| Parameters             | Descriptions                        |
|------------------------|-------------------------------------|
| Username               | Username of Twitter account.        |
| Probability Score      | The probability score of the account.|
| Label                  | A label indicating whether the account is identified as a real or fake account. |

The third step is data pre-processing. The tweets data that have been collected are then cleaned by removing the stop words, which is a collection of words that have no meaning. The tweets data were also converted to a lowercase form to make it easier for the machine to classify sentiments. The tweets data that have been cleaned are then labeled according to the type of tweet, namely positive, neutral, and negative. From the data of all tweets that have been labeled, the data obtained are tweets with a positive label of 1,062 tweets, a neutral label of 1,889 tweets, and a negative label of 1,219 tweets. The fourth step is data processing which is processing the data by training the dataset and run the algorithm. The first training begins by using an entire tweet dataset consisting of real accounts and fake accounts, and the dataset is taken randomly. The second training was carried out without using fake accounts. The total number of datasets in the first training was also made equal to the number of datasets without using fake accounts in the second training to avoid unbalanced data. The fifth step is the result which is analyzing and discussing the score based on the result on the algorithm produced from data processing. In machine learning, there are several metrics to evaluate how accurate the algorithm is. In our research, we have used Accuracy, F1-Score, Precision, and Recall matrix as the performance metric. Classification accuracy is the number of correct predictions as a ratio of all predictions made. F1-Score is the harmonic mean of precision and recall. Precision is the ratio of items that are predicted as positive to that are positive. Recall is the ratio of items that are positive to that are predicted as positive by the system [14]. Next, the results were visualized between sentiment classifications using all tweets data and without using fake accounts. The steps are shown in Fig. 1.

Table 2. Details of account authenticity reference parameters and features of Tweetbotornot

Table 3. Detailed parameters and descriptions of the identification of fake accounts.
4. Result and Discussion

At the experimental stage of sentiment analysis, the classification process is carried out using the Naïve-Bayes (NB) and Support Vector Machine (SVM) algorithms. These two methods are known to have good classification performance according to several studies related to machine learning. NB is the simplest and most widely used algorithm. The NB algorithm works by calculating the probability of a class appearing based on the distribution of words in a document using the Bag of Word (BOW) feature extraction [3]. Meanwhile, SVM is an algorithm that works by transforming training data into higher dimensions, where this algorithm will look for a hyperplane with the largest margin that separates data by class using training tuples [5]. However, basically, the SVM algorithm only supports binary classification. So, the use of this algorithm is slightly modified because this research uses three sentiment classes. The SVM algorithm in this research uses multiclass classification with a One-to-Rest approach, where one class will be separated from the other two classes by a hyperplane. The classification process begins with a random dataset of tweets with fake accounts and continues with a dataset of tweets without fake accounts. The dataset of tweets with fake accounts is equal to the dataset of tweets without fake accounts of 2,218 tweets. Each process is tested four times with a comparison ratio of training data and testing data for each experiment of 60:40, 70:30, 80:20, and 90:10. The experiments were carried out to obtain the best sentiment classification results. The first experiment was carried out using the NB algorithm. The results of each experiment for the NB algorithm using fake accounts can be seen in Table 4. Also, the results of each experiment for the NB algorithm without using fake accounts can be seen in Table 5.

| Data Ratio | Precision Positive | Precision Neutral | Precision Negative | Recall Positive | Recall Neutral | Recall Negative | F1 Score Positive | F1 Score Neutral | F1 Score Negative | Accuracy |
|------------|--------------------|------------------|-------------------|----------------|---------------|----------------|------------------|----------------|------------------|----------|
| 60:40      | 0.51               | 0.60             | 0.47              | 0.70           | 0.51          | 0.43           | 0.59             | 0.55           | 0.45             | 0.532    |
| 70:30      | 0.53               | 0.58             | 0.48              | 0.69           | 0.49          | 0.46           | 0.60             | 0.53           | 0.47             | 0.538    |
| 80:20      | 0.52               | 0.58             | 0.46              | 0.72           | 0.44          | 0.47           | 0.60             | 0.50           | 0.47             | 0.523    |
| 90:10      | 0.51               | 0.61             | 0.41              | 0.66           | 0.44          | 0.48           | 0.57             | 0.51           | 0.44             | 0.510    |
Table 5. NB algorithm experiment results without using fake accounts.

| Data Ratio | Precision | Recall | F1 Score | Accuracy |
|------------|-----------|--------|----------|----------|
|            | Positive  | Neutral| Negative | Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| |
| 60 : 40    | 0.45      | 0.64   | 0.45     | 0.59     | 0.53   | 0.47   | 0.51     | 0.58   | 0.46   | 0.528     |
| 70 : 30    | 0.51      | 0.65   | 0.39     | 0.73     | 0.48   | 0.43   | 0.60     | 0.55   | 0.41   | 0.527     |
| 80 : 20    | 0.53      | 0.65   | 0.44     | 0.60     | 0.51   | 0.54   | 0.56     | 0.57   | 0.49   | 0.545     |
| 90 : 10    | 0.48      | 0.69   | 0.55     | 0.64     | 0.64   | 0.48   | 0.55     | 0.66   | 0.51   | 0.590     |

From these results, it can be seen that the highest Accuracy score for the NB algorithm experiment using all tweets data is obtained at a data ratio of 70:30 with a percentage of 53.8%. Also, the lowest Accuracy score is obtained at a data ratio of 90:10 with a percentage of 51%. Meanwhile, the NB algorithm experiment without using fake accounts obtained the highest Accuracy score at 90:10 data ratio of 59%. Then, the lowest Accuracy score is obtained at a data ratio of 70:30 with a percentage of 52.7%. These results indicate that the sentiment classification process without using fake accounts is better than using all tweets data in the NB algorithm experiment.

Furthermore, the sentiment classification experiment is continued by using the SVM algorithm. The dataset used in this second experiment is the same as the dataset in the first experiment. The results of each experiment for the SVM algorithm using fake accounts can be seen in Table 6. Also, the results of each experiment for the SVM algorithm without using fake accounts can be seen in Table 7.

Table 6. SVM algorithm experiment results using fake accounts.

| Data Ratio | Precision | Recall | F1 Score | Accuracy |
|------------|-----------|--------|----------|----------|
|            | Positive  | Neutral| Negative | Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| |
| 60 : 40    | 0.87      | 0.76   | 0.72     | 0.72     | 0.88   | 0.64   | 0.79     | 0.82   | 0.68   | 0.774     |
| 70 : 30    | 0.91      | 0.76   | 0.78     | 0.69     | 0.90   | 0.72   | 0.79     | 0.83   | 0.75   | 0.795     |
| 80 : 20    | 0.84      | 0.72   | 0.76     | 0.67     | 0.88   | 0.66   | 0.75     | 0.79   | 0.71   | 0.756     |
| 90 : 10    | 0.86      | 0.76   | 0.81     | 0.76     | 0.87   | 0.70   | 0.81     | 0.81   | 0.75   | 0.792     |

Table 7. SVM algorithm experiment results without using fake accounts.

| Data Ratio | Precision | Recall | F1 Score | Accuracy |
|------------|-----------|--------|----------|----------|
|            | Positive  | Neutral| Negative | Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| Positive  | Neutral| Negative| |
| 60 : 40    | 0.82      | 0.72   | 0.75     | 0.65     | 0.88   | 0.62   | 0.72     | 0.79   | 0.68   | 0.744     |
| 70 : 30    | 0.83      | 0.72   | 0.79     | 0.64     | 0.90   | 0.63   | 0.72     | 0.80   | 0.70   | 0.756     |
| 80 : 20    | 0.91      | 0.77   | 0.80     | 0.67     | 0.92   | 0.74   | 0.77     | 0.84   | 0.77   | 0.806     |
| 90 : 10    | 0.93      | 0.72   | 0.77     | 0.58     | 0.91   | 0.68   | 0.72     | 0.80   | 0.72   | 0.761     |

From these results, it can be seen that the highest Accuracy score for the SVM algorithm experiment using all tweets data is obtained at a data ratio of 70:30 with a percentage of 79.5%. Also, the lowest Accuracy score is obtained at a data ratio of 80:20 with a percentage of 75.6%. Meanwhile, the SVM algorithm experiment without using fake accounts obtained the highest Accuracy score at 80:20 data ratio of 80.6%. Then, the lowest Accuracy score is obtained at a data ratio of 60:40 with a percentage of 74.4%. These results indicate that the sentiment classification process without using fake accounts is better than using all tweets data in the SVM algorithm experiment.

The results of the sentiment classification experiment on both algorithms showed that the sentiment classification process without using fake accounts is better than using all tweets data. These results prove that there is an influence from fake accounts that can reduce the performance of sentiment classification. From the experimental results it is also known that the SVM algorithm has the highest Accuracy value of 80.6%, while the NB algorithm only has the highest Accuracy value of 59%. Precision, Recall, and F1-Score in the SVM algorithm also show higher numbers
when compared to the NB algorithm. These results prove that sentiment classification using the SVM algorithm has better performance than the NB algorithm for this case.

Apart from the results of the sentiment classification process, there are also the results of the sentiment visualization of tweets. The sentiment visualization of tweets is displayed in a pie chart. The results of the tweets sentiment visualization using all tweets and without fake accounts can be seen in Fig. 2.

From the results of the visualization, it can be seen that if we use all tweets data, the percentage of positive sentiment is 25.47%, neutral sentiment is 45.3%, and negative sentiment is 29.23%. However, after the tweet data from fake accounts were removed, there was a change in the percentage to a positive sentiment of 24.39%, a neutral sentiment of 46.35%, and a negative sentiment of 29.26%. These results indicate that there was an influence from fake accounts which actually leads to positive sentiment although it is not significant. This result is contrary to the initial hypothesis of the research which assumes that fake accounts will lead opinions towards negative sentiment.

5. Conclusion

In this paper, several experiments were conducted to determine the influence of fake accounts on sentiment analysis. The tweets data obtained are divided into two datasets, including a dataset of all tweets using fake accounts and without using fake accounts. Each dataset was tested several times by classifying tweet sentiments. This research also compares sentiment classification using the Naïve-Bayes (NB) and Support Vector Machine (SVM) algorithms. The results of the sentiment classification experiment on both algorithms showed that the sentiment classification process without using fake accounts is better than using all tweets data. These results prove that there was an influence from fake accounts that can reduce the performance of sentiment classification. The experimental results of the two algorithms also prove that the SVM algorithm has a better performance than the NB algorithm for this case with the highest Accuracy value of 80.6%. In addition, the results of the sentiment visualization showed that there was an influence from fake accounts which actually leads to positive sentiment although it is not significant. This result is contrary to the initial hypothesis of the research which assumes that fake accounts will lead opinions towards negative sentiment. We will try several ways to get better results for future work, including using more extensive datasets and using other classification processes.

References

[1] Satgas Penanganan COVID-19. (2020) “Komite Penanganan COVID-19 dan Pemulihan Ekonomi Nasional [Title in English: Committee for Handling COVID-19 and National Economic Recovery].” [Online]. Available: https://www.covid19.go.id/.
[2] Hootsuite, and W. A. Social. (2020) “Digital 2020: Indonesia.” [Online]. Available: https://datareportal.com/reports/digital-2020-indonesia/.
[3] W. Medhat, A. Hassan, and H. Korashy. (2014) “Sentiment analysis algorithms and applications: A survey.” Ain Shams Engineering Journal
5 (4): 1093-1113.

[4] M. A. Fauzi. (2018) “Random forest approach for sentiment analysis in Indonesian language.” Indonesian Journal of Electrical Engineering and Computer Science 12 (1): 46-50.

[5] J. Han, M. Kamber, and J. Pei. (2012) “Data Mining: Concepts and Techniques.” Elsevier.

[6] A. B. Eliacik, and N. Erdogan. (2018) “Influential user weighted sentiment analysis on topic based microblogging community.” Expert Systems with Applications 92: 403-418.

[7] E. A. Kolog, C. S. Montero, and T. Toivonen. (2018) “Using machine learning for sentiment and social influence analysis in text.” in International Conference on Information Technology & Systems, Springer, Cham

[8] N. Chouchani, and M. Abed. (2020) “Enhance sentiment analysis on social networks with social influence analytics.” Journal of Ambient Intelligence and Humanized Computing 11 (1): 139-149.

[9] Y. Bae, and H. Lee. (2012) “Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers.” Journal of the American Society for Information Science and technology 63 (12): 2521-2535.

[10] F. Bravo-Marquez, E. Frank, and B. Pfahringer. (2016) “Annotate-Sample-Average (ASA): A new distant supervision approach for Twitter sentiment analysis.” in 22nd European Conference on Artificial Intelligence (ECAI), IOS Press

[11] E. Van Der Walt, and J. Eloff. (2018) “Using Machine Learning to Detect Fake Identities: Bots vs Humans.” IEEE Access 6: 6540-6549.

[12] F. C. Akyon, and M. Esat Kalfaoglu. (2019) “Instagram Fake and Automated Account Detection.” in 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE

[13] T. Velayutham, and P. K. Tiwari. (2018) “Bot identification: Helping analysts for right data in twitter.” in 2017 3rd international conference on advances in computing, communication & automation (ICACCA)(fall), IEEE

[14] B. Erşahin, Ö. Aktaş, D. Kilmç, and C. Akyol. (2017) “Twitter fake account detection.” in 2017 International Conference on Computer Science and Engineering (UBMK), IEEE

[15] J. Jia, B. Wang, and N. Z. Gong. (2017) “Random Walk Based Fake Account Detection in Online Social Networks.” in 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), IEEE

[16] Jon, M. Dustin, Fauzan, and M. Hamza. (2015) “Tweetbotornot.” [Online]. Available: https://github.com/mkearney/tweetbotornot.