Analysis of twitter sentiment in COVID-19 era using fuzzy logic method

Devi Ajeng Efrilianda[1], Erika Noor Dianti[2], Oktaria Gina Khoirunnisa[3]
[1,2,3]Department of Computer Science, Universitas Negeri Semarang, Indonesia

ABSTRACT

The sentiment is an assessment of attitudes towards certain events or things. Collecting opinion is known as a sentiment from existing data. This technique can also help analyze the opinions given by people in assessing certain objects. The best available source for gathering sentiment is the internet. In the era of the Covid-19 pandemic, many people access social media, especially Twitter to give their opinion on certain objects. Twitter is known as the social media that is accessed by users to post their opinions online. By using soft computing, especially fuzzy logic, it is possible to design, create and build bots that can analyze user opinions on Twitter. This model is used for data sentiment analysis on Twitter. The results showed that the sentiment analysis during the Covid-19 period was still dominant with positive tweets. As many as 48% of tweets are positive, 30% are negative tweets and 22% are neutral tweets. The use of applications to identify tweet sentiment during Covid-19 uses a combination of fuzzy logic methods with artificial intelligence. With the help of the Twitter API, you can get tweet data during the Covid-19 pandemic so you can find out the frequently used tweet sentiments.

Corresponding Author:

Devi Ajeng Efrilianda,
Department of Computer Science,
Universitas Negeri Semarang,
Sekaran, Gunungpati, Semarang, Indonesia.
Email: deviajeng@mail.unnes.ac.id

1. INTRODUCTION

This year there has been a change in lifestyle almost all over the world. This is due to the emergence of a new type of virus that attacks humans, namely Coronavirus Disease 2019 (Covid-19). Coronavirus itself is a group of viruses from the Ortho Coronavirus subfamily that can cause disease not only in humans but also in birds and mammals [1]. This virus then spread from the Chinese city of Wuhan to cause a large and prolonged pandemic all over the world until now. At the same time, the current condition of the Covid-19 pandemic affects all areas of life from the economy, society, politics, education, and so on.

Because the Covid-19 pandemic did not end, the government then implemented Large-Scale Social Restrictions (PSBB) which impacted the community to implement work from home (WFH) and distance learning for students and office workers. Therefore, to support these conditions, people are required to use gadgets. Over time, people are more active in socializing indirectly in cyberspace via the internet which makes social media users increase. Now more than 100 million active users open applications such as social media [2].

Social media is a persistent, internet-based, personal mass communication channel that facilitates the perception of interaction among users, derives value primarily from user-generated content, and is untrained [3]. The most commonly used social media networks such as Facebook, Youtube. Instagram, and Twitter [4]. Twitter itself has become one of the most popular social media in the world with more than 200 million active users and 10.6 billion tweets worldwide. Each user can easily provide their arguments, stories, expressions,

breaking news, or hot topics via Twitter. Indonesia, as one of the largest countries in the world with a population of more than 200 million people and has a large number of active Twitter users as well [5].

Social media, especially Twitter, have become increasingly popular with tweets discussing the Covid-19 pandemic in cyberspace with the emergence of hashtags about events that occurred during the Covid-19 pandemic making it a trending topic. That way, many responses or perceptions emerge from the community. The data obtained from Twitter will later be processed and analyzed to be useful for society and an organization. Sentiment analysis is used in this study to analyze arguments, stories, expressions of an individual on events that occurred during the Covid-19 pandemic through text mining data posts on Twitter. Sentiment analysis is a study that analyzes people's arguments and opinions towards entities such as services in text or products [6], [7].

Based on the description above, this research will apply a sentiment analysis based on an opinion by sorting the topics from those that are often discussed to those that are rarely discussed. This research uses fuzzy logic to design, create and build bots that can analyze user opinions on Twitter. Fuzzy logic is useful for processing and evaluating information [8]. The purpose of this study is to analyze public sentiment towards events during Covid-19, to find out the trending order of topics on Twitter during the pandemic, and to design, build, and build bots that can analyze user opinions on Twitter.

2.   LITERATURE REVIEW

In a systematic literature review based on the reviewed papers, 7 papers are using the lexicon method, 10 papers using the machine learning method, and 7 papers mixing the two when applying sentiment analysis. The lexicon-based method is known as the unsupervised learning method. When conducting sentiment analysis most of the studies adopted the Sentiwordnet and TF-IDR methods. The Sentiwordnet method calculates based on positive or negative words. Meanwhile, the TF-IDR method converts numbers into words and is calculated using the frequency inversion document frequency method. Machine learning methods require training data to be processed. The methods that are often used in machine learning methods are the Naive Bayes and SVM, models. To improve results, combine the lexicon and machine learning methods. The data taken is mostly from the social media site Twitter [9].

The CorE Q9 bootstrap algorithm to find semantic lexicons that can be used to divide tweets into two categories: stressed and non-stressed. The Twitter data used in this article is collected via the Twitter API from January to April 2020 for the continental United States. One of the main innovations in our research is mapping the symptoms of stress that causes COVID-19 on a temporal scale. The algorithm takes a large, unnamed corpus from which it finds new related words and writes them into the wrong semantic category (for example, stress and non-stress in the case). Before the bootstrapping process, the pattern extraction was carried out on the unmarked corpus. It is used to extract all subject noun phrases, direct objects, or prepositional phrases. A universal sentence encoder is used to generate word embeddings. These text embeddings transform the tweet into a numeric vector, encoding the tweet text into the high-dimensional vector needed to find semantic similarities and perform classification tasks. The classifications used in the training process are SVM, logistic regression, naïve Bayes classification, and simple neural networks. By doing this, we were able to observe the spatiotemporal patterns of stress symptoms and answer questions about what are the main concerns regarding pandemics in different geographic areas on different time scales [10].

Sentiment analysis on Twitter users about the anti-LGBT campaign in Indonesia by seeking positive, negative, or neutral responses. Then use the Naive Bayes algorithm because it has a high degree of accuracy in analyzing sentiment. In this study, it can be concluded that on average giving neutral comments with an accuracy of 86.43% obtained through the Rapid Miner tool [11].

3.   METHOD

This research is a study by analyzing the responses or perceptions of the community in a situation that occurs. Therefore, this study uses two sources of information, namely, the collection of data on public sentiment on Twitter social media through hashtags related to events during the Covid-19 pandemic. The population obtained from Twitter residents was 212 tweets. This data collection is using a random sampling method. Furthermore, this study also uses literature reviews such as journals or books related to the research being carried out.

After determining the selected data source, data collection is then carried out. As previously mentioned, data was obtained from the Twitter social media application using queries or special hashtags regarding events that occurred during Covid-19. The use of queries here is to make it easier to find various public sentiments about events that occurred during Covid-19.

After the data is collected, it performs calculations using the fuzzy method. Fuzzy itself is seen as being able to decipher an input into an output without neglecting the existing aspects [12], [13].

Fuzzy logic itself is considered very easy to apply because of its flexible nature and can be based on human logic with everyday language that is easy to understand [4], [14]. This fuzzy calculation begins with converting linguistic variables into functions that describe variables in fuzzy form. Each of these functions represents a linguistic variable and describes a fuzzy set combined with certain criteria. After the set is described, the next step is to determine the if-then which explains how the relationship between the fuzzy set, namely max (representing and) and min (representing or). In the connection completion step, all estimates are combined on the final fuzzy set.

The simplest form of data collected is text on the Twitter application. Searching for topics on social media Twitter will also be easier with the hashtags of topics or related events. So, as mentioned earlier, this study collects data through hashtags used by Twitter users which will later be analyzed through data mining methods. Data mining itself is a method of finding information or patterns in selected data [15], [16].

3.1   Text analysis

Anyone can manually analyze a sentimental tweet. However, over time, there are more and more users of the Twitter social media application. This of course causes the analysis process to be less concise and time-consuming. However, this can be made easy with information technology. This process uses the fuzzy method. This method has the ability in the process of language reasoning (linguistic reasoning) [17]–[19]. In this process, we get 200 data which consists of which we sort into negative tweets and positive tweets.

3.2 Twitter API

Application Programming Interface or what we often API is a program used to retrieve or modify data [20]. The program provided by Twitter is intended to make it easier for someone to obtain information available on Twitter. The Twitter API itself will help in selecting data so that later the data used will be more concise. Some of the data obtained will be deleted later. This process is often called the preprocessing process [21], [22]. After that, proceed to the Case Folding Process where we will improve the tweet data that we get. The fix here is intended to homogenize the text into lowercase. Then it will be continued with the data cleaning process. This process aims to remove the RT component (retweet), URL, and username. Next, the last one is Stopword Removal. Stopword itself is a word that is in the data but is less helpful in the process of tweeting analysis [23].

3.3 Fuzzy Application

The first step before we start applying fuzzy in this research is connecting the data with the existing database. Next, we enter the nature and hashtags text which will later be processed into text, and hashtags which will later be processed using fuzzy logic. At this stage, it will produce positive and negative scores for each tweet. After getting both of them, we do the calculation of each score which will then be subtracted from each other and produce the final score. After that, the data normalization stage is carried out. This data normalization aims to assist in data measurement and reduce redundancies [20]. After the weighting is done, we will begin to count and classify the data into several categories, namely very positive tweets, positive tweets, neutral tweets, negative tweets, and very negative tweets. The information about tweet analysis model can be seen at figure 1.
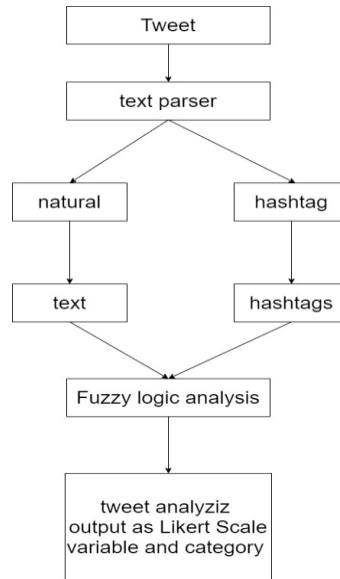
Figure 1**.** Tweet analysis model.

4. RESULT AND DISCUSSION

First, the tweet will be extracted or deciphered, using the fuzzy logic method, the sentiment value of the tweet will be calculated. If you have a tweet value, you can calculate the arithmetic average value. The results of this will know the percentage of positive, neutral, and negative values. Many of the conversations that occurred during Covid-19 discussed economic topics, case reports, and health. In [24] the authors analyzed Twitter data in the period 2 February-15 March 2020 from the identification of topics, 10 of which had positive sentiments and 2 were negative. Case [25] analyzed data from 01 January - 23 March found a greater number of positive and neutral tweets. Meanwhile, negative tweets were few. This is because people think the Covid-19 virus will end soon. However, as time went on negative tweets grew more and more.

Overall sentiment about Covid-19 is growing rapidly and the problem of the pandemic has become more complex as time goes on. In this study, the results showed that the sentiment analysis during the Covid-19 period was still dominant with positive tweets. As many as 48% of tweets are positive, 30% are negative tweets and 22% are neutral tweets. The use of applications to identify tweet sentiment during Covid-19 uses a combination of fuzzy logic methods with artificial intelligence. With the help of the Twitter API, you can get tweet data during the Covid-19 pandemic so you can find out the frequently used tweet sentiments.

5. CONCLUSION

Based on calculations using the fuzzy method, the results show that during the Covid-19 pandemic this study found 48% positive tweets, 30% negative tweets, and 22% neutral tweets. It can be concluded that there are more positive tweets than negative and neutral tweets. This is because people think that the pandemic will not last long, even though there are more and more negative tweets every day.

REFERENCES

[1] M. Mitra and R. K. Samanta, "A study on UCI hepatitis disease dataset using soft computing," *Model. Meas. Control C*, vol. 78, no. 4, pp. 467–477, 2017, doi: 10.18280/mmc_c.780405.

[2] Y. Hou, D. Xiong, T. Jiang, L. Song, and Q. Wang, "Social media addiction: Its impact, mediation, and intervention," *Cyberpsychology*, vol. 13, no. 1, 2019, doi: 10.5817/CP2019-1-4.

[3] R. A., E. S. -, A. A., and B. Ibrahim, "Modifications of particle swarm optimization techniques and its application on stock market: a survey," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 3, 2015, doi: 10.14569/ijacsa.2015.060315.

[4] K. Howells and A. Ertugan, "Applying fuzzy logic for sentiment analysis of social media network data in marketing," *Procedia Comput. Sci.*, vol. 120, pp. 664–670, 2017, doi: 10.1016/j.procs.2017.11.293.

[5] M. Ibrahim, O. Abdillah, A. F. Wicaksono, and M. Adriani, "Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation," *Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW 2015*, pp. 1348–1353, 2016, doi: 10.1109/ICDMW.2015.113.

[6] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information," *Comput. Intell. Neurosci.*, vol. 2015, 2015, doi: 10.1155/2015/715730.

[7]   I.E. Tiffani, "Optimization of naïve bayes classifier by implemented unigram, bigram, trigram for sentiment analysis of hotel review," *J. Soft Comput. Explor.*, vol. 1. no. 1, pp. 1–7, 2020.

[8]   R. Santhosh and M. Mohanapriya, "Generalized fuzzy logic based performance prediction in data mining," *Mater. Today Proc.*, 2020, doi: 10.1016/j.matpr.2020.08.626.

[9]   Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," *Procedia Comput. Sci.*, vol. 161, pp. 707–714, 2019, doi: 10.1016/j.procs.2019.11.174.

[10]   D. Li, H. Chaudhary, and Z. Zhang, "Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining," *Int. J. Environ. Res. Public Health*, vol. 17, no. 14, pp. 1–23, 2020, doi: 10.3390/ijerph17144988.

[11]   S. Kaur, P. Kaul, and P. M. Zadeh, "Monitoring the dynamics of emotions during Covid-19 using twitter data," *Procedia Comput. Sci.*, vol. 177, pp. 423–430, 2020, doi: 10.1016/j.procs.2020.10.056.

[12]   Sunjana, "Prediction of production using the fuzzy mamdani inference method," *Int. J. Adv. Sci. Technol.*, vol. 28, no. 6, pp. 136–139, 2019.

[13]   T. Yulianto, S. Komariyah, and N. Ulfaniyah, "Application of fuzzy inference system by Sugeno method on estimating of salt production," *AIP Conf. Proc.*, vol. 1867, 2017, doi: 10.1063/1.4994442.

[14]   A. Chakraborty, A. Kolya, and S. Dutta, "A fuzzy logic inspired approach for social media sentiment analysis via deep neural network," *Adv. Intell. Syst. Comput.*, vol. 706, pp. 175–185, 2018, doi: 10.1007/978-981-10-8237-5_17.

[15]   B. Çiğşar and D. Ünal, "Comparison of data mining classification algorithms determining the default risk," *Sci. Program.*, vol. 2019, 2019, doi: 10.1155/2019/8706505.

[16]   S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," *Procedia Comput. Sci.*, vol. 82, pp. 115–121, 2016, doi: 10.1016/j.procs.2016.04.016.

[17]   J. M. Sánchez, J. P. Rodríguez, and H. E. Espitia, "Review of artificial intelligence applied in decision-making processes in agricultural public policy," *Processes*, vol. 8, no. 11, pp. 1–23, 2020, doi: 10.3390/pr8111374.

[18]   T. Tashtoush, A. Alazzam, and A. Rodan, "Utilizing fuzzy logic controller in manufacturing facilities design: Machine and operator allocation," *Cogent Eng.*, vol. 7, no. 1, 2020, doi: 10.1080/23311916.2020.1771820.

[19]   Sulistiana and M. A. Muslim, "Support Vector Machine (SVM) optimization using grid search and unigram to improve e-commerce review accuracy," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 8–15, 2020.

[20]   A. Haque and T. Rahman, "Sentiment analysis by using fuzzy logic," *Int. J. Comput. Sci. Eng. Inf. Technol.*, vol. 4, no. 1, pp. 33–48, 2014, doi: 10.5121/ijcseit.2014.4104.

[21]   I. M. Yulietha, S. A. Faraby, Adiwijaya, and W. C. Widyaningtyas, "An implementation of support vector machine on sentiment classification of movie reviews," *J. Phys. Conf. Ser.*, vol. 971, no. 1, 2018, doi: 10.1088/1742-6596/971/1/012056.

[22]   A. Lestari and Alamsyah, "Increasing Accuracy of C4.5 algorithm using information gain ratio and adaboost for classification of chronic kidney disease," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 32–38, 2020.

[23]   D. Ramachandran and R. Parvathi, "Analysis of twitter specific preprocessing technique for tweets," *Procedia Comput. Sci.*, vol. 165, pp. 245–251, 2019, doi: 10.1016/j.procs.2020.01.083.

[24]   A.-A. A., A. D., H. M., H. M., and S. Z., "Top Concerns of tweeters during the covid-19 pandemic: Infoveillance Study," *J. Med. Internet Res.*, vol. 22, no. 4, p. e19016, 2020.

[25]   K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media," *Appl. Soft Comput. J.*, vol. 97, 2020, doi: 10.1016/j.asoc.2020.106754.