CENTRAL ASIAN JOURNAL OF THEORETICAL AND APPLIED SCIENCES

Volume: 04 Issue: 12 | Dec 2023 ISSN: 2660-5317
https://cajotas.centralasianstudies.org

Machine Learning Implementation in Sentiment Analysis of Covid-19 Pandemic Handling Policies in Indonesia

David Hermansyah[1], Budi Santoso[2], Anggit Wikanningrum[3]
[1,2,3] Lecturer, Informatics Engineering, Universitas Dr.Soetomo, Surabaya, Indonesia
[1]david.hermansyah@unitomo.ac.id, [2]budi.santoso@unitomo.ac.id, [3]anggit.wikanningrum@unitomo.ac.id

Received 19th Sep 2023, Accepted 28th Nov 2023, Online 30th Dec 2023

Abstract: *The behavior of Indonesian society in expressing opinions has shifted towards the use of internet-based social media. The use of social media to voice aspirations also occurred during the emergence of the Covid-19 pandemic in Indonesia, especially in response to the government's policies in handling the outbreak. The purpose of this research is to determine the effectiveness and influence of government policies on the community, one of which is by conducting sentiment analysis on the Twitter social media platform. Previous studies have not specifically analyzed the sentiment of public reactions to government policies in the case of the Covid-19 pandemic or have only discussed the legal aspects of those policies. This research employs a sentiment classification method using machine learning into three categories: positive, neutral, and negative. The analysis is carried out on the occurrence of predefined phrases related to government policies in tweets circulating on social media. The results of this study indicate that public sentiment towards the government's handling of the Covid-19 pandemic tends to be neutral and positive, but the negative sentiment carries a stronger tendency despite its smaller quantity.*

Keywords: *machine learning, sentiment analysis, Covid-19, government policies*

Introduction

The social behavior of Indonesian society, particularly in choosing media to voice opinions and respond to current issues, has shifted towards online social media on the internet. Millions of users, both individuals and institutions, have utilized social media as a form of existence in the virtual world. They use social media to build social relationships, share information, and events (K. Sivarajah, 2019). This is supported by massive developments in internet network infrastructure, mobile technology, and the purchasing power of the population. If in the past channels for expressing opinions were dominated by limited print and electronic media, there are now various forms of online social media available and freely accessible. Thus, any prominent issue that arises in Indonesia is almost impossible to escape the attention and response of the public, in the form of comments, criticism, and suggestions through online social media. This includes every policy of the Government of the Republic of Indonesia, which will always be monitored through public responses, one of which is through social media. The current issue at the time of this research is the global outbreak of the Covid-19 virus, including in Indonesia. The COVID-19 pandemic or coronavirus, which spread globally, began at the end of 2019 in Wuhan, China. In Indonesia, the first official announcement of a Covid-19 case was made on March 1, 2020 (Covid-19 Task Force, 2020).

Although social media users, especially on Twitter, had previously discussed the likelihood of the virus entering Indonesia, this was denied by statements from several officials of the Indonesian government. As of September 11, 2020, the number of positive Covid-19 cases in Indonesia has reached 207,203 cases, with 8,456 reported deaths, and 147,510 individuals declared recovered (DetikHealth, 2020).

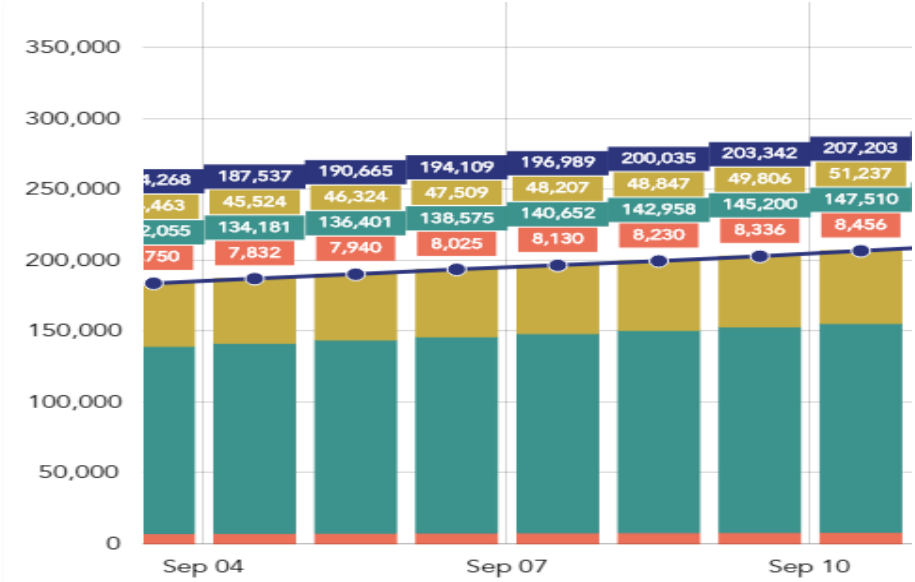

Figure 1. Daily graph of Covid-19 exposure in Indonesia in September 2020 (Covid-19 Task Force, 2020)

What is interesting is that not all social media users simply believe the official statements of government officials in the case of the Covid-19 pandemic. Unlike in the time before the development of information and communication technology, where almost all public media were controlled by the government. Currently, social media platforms are beyond the government's control, allowing the public to express their aspirations without difficulty. Meanwhile, the Indonesian government must be cautious and meticulous in conveying official information, as data and facts are easily accessible to the public. Therefore, any carelessness in delivering official information can lead to criticism and even public backlash through online social media.

The presence of online social media, especially Twitter with its hashtag feature, further expands the reach of disseminating information about the COVID-19 pandemic, especially if it becomes a trending topic. A trending topic will attract more social media users to find out the contents of that hashtag. The dissemination of information through Twitter, commonly referred to as tweets, produces a vast amount of digital data. This data can be essential and beneficial for the public or organizations when processed and analyzed properly. One form of processing textual digital data known in the world of information technology is sentiment analysis. Sentiment analysis is a method to explore whether the content collected is in a positive, negative, or neutral state (R. S. Luqyana, 2018). What is interesting is that not all social media users simply believe the official statements of government officials in the case of the Covid-19 pandemic. Unlike in the time before the development of information and communication technology, where almost all public media were controlled by the government. Currently, social media platforms are beyond the government's control, allowing the public to express their aspirations without difficulty. Meanwhile, the Indonesian government must be cautious and meticulous in conveying official information, as data and facts are easily accessible to the public. Therefore, any carelessness in delivering official information can lead to criticism and even public backlash through online social media.

The presence of online social media, particularly Twitter with its hashtag feature, is increasingly expanding the reach of disseminating all information about the COVID-19 pandemic, especially when it

becomes a trending topic. A trending topic attracts more social media users to explore the content of that hashtag. The delivery of information through Twitter, commonly known as tweets, produces a vast amount of digital data. This data can be crucial and beneficial for the public or organizations when processed and analyzed properly. One form of processing textual digital data known in the world of information technology is sentiment analysis. Sentiment analysis is a method to explore whether the collected content is in a positive, negative, or neutral state (R. S. Luqyana, 2018).

The Indonesian government officially declared the Covid-19 pandemic as a National Disaster Status through Presidential Decree Number 12 of 2020, dated April 13, 2020, concerning the Declaration of a Non-Natural Disaster, the Spread of Corona Virus Disease 2019 (COVID-19), as a National Disaster (Government of Indonesia, 2020). With this Presidential Decree, the government is granted legitimacy to design, allocate funds, and implement specific measures for handling the Covid-19 pandemic in Indonesia (Fitri, 2020). The initial step in implementing this national disaster status is the enforcement of Large-Scale Social Restrictions (PSBB) and social distancing policies.

As is customary with emergency policies, they will always receive both positive and negative opinions from the public as they directly impact people's lives. According to a survey conducted by the LIPI Population Research Center, "the survey results note that 39.4% of businesses have stopped, and 57.1% have experienced a production decline. Only 3.5% were unaffected" (Karunia, 2020).

In addition to manifesting in the virtual world through social media, public reactions to government policies on handling Covid-19 also extend to the offline world. Various demonstrations in different regions express dissatisfaction with the impacts of central or regional government policies. On the other hand, the significant number of medical personnel who have sacrificed their lives on the front lines of handling Covid-19 patients raises deep concerns. It is suspected that they did not have adequate medical equipment, making them highly susceptible to the virus, especially in the early days of the Covid-19 outbreak in Indonesia.

Previous research on sentiment analysis related to government policies has been conducted by (Akhmad Sa'rony, 2019), focusing on sentiment analysis of the relocation policy of the Indonesian capital. Meanwhile, (Fitri, 2020) emphasized the legal aspects of declaring the Covid-19 pandemic as a national disaster in her research.

In this study, sentiment analysis is used to analyze the opinions, feelings, and views of Indonesian society regarding several government policies related to handling the COVID-19 pandemic in Indonesia. The digital text data analyzed are obtained from capturing the Twitter user data stream discussing these government policies. Due to the global spread of the pandemic, including in Indonesia, and the implementation of social interaction restrictions affecting social conditions, various issues—both positive and negative—circulate, causing significant panic on social media (Martin, 2020).

Research Method

The research methodology for sentiment analysis of Twitter users regarding the Indonesian government's policies on handling the Covid-19 pandemic is divided into four stages: data collection, pre-processing, feature extraction, and classification. If needed, visualization of results in graphic form is added as an interaction with the readers. The complete diagram of the methodology can be seen in Figure 2.

In its implementation, this study applies several concepts of text data mining :

Text Mining

Text mining, also referred to as data mining, is called such because the same algorithms can be used in both mining concepts. The significant difference between the two is that data mining involves more structured data, while text mining deals with specific features that are relatively unstructured and require

initial processing (preprocessing). According to (Yogapreethi, 2016), the main process of text mining consists of:

1. Information Extraction: an initial analysis of unstructured text by extracting information, including searching for relationships between phrases.
2. Clustering: the process of separating text into groups and assigning weights to each word.
3. Classification: the process of identifying the main topics of documents by determining metadata in the form of class labels.
4. Information Visualization: the process of displaying the results of pattern analysis, used for user interaction with documents.

Sentiment Analysis

The steps of sentiment analysis can be applied as follows (Kiruthika, 2016):

- Sentence Level: searching for positive, neutral, and negative sentiments in each line of text.
- Document Level: analyzing the sentiment of the entire text as positive, neutral, or negative.
- Aspect Level: grouping all attributes with the same sentiment results.
- User Level: utilizing visualization of sentiment analysis results as a means of interacting with users.

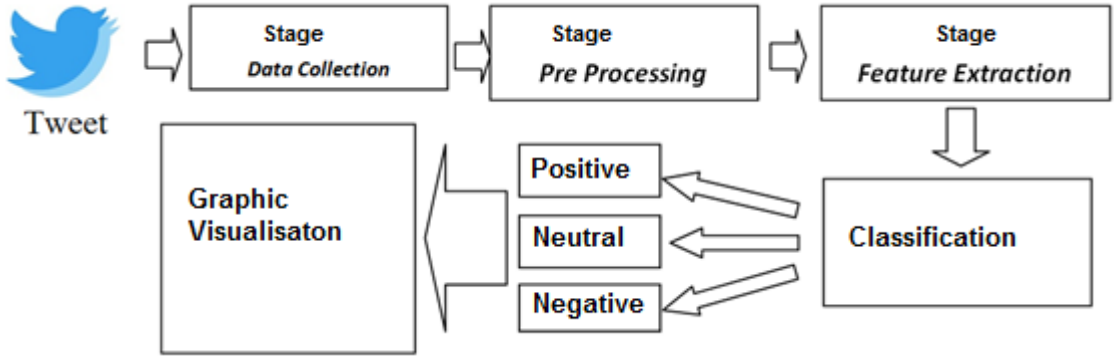

Figure 2. Bock Diagram of research stages

Explanation of the research stages is as follows:

a. Data Collection Stage

In the data collection stage, the process of collecting textual data from tweet streams sent by the public as a reaction to the policies adopted by the Indonesian government in handling Covid-19 in Indonesia is conducted. To obtain a set of textual data that meets the requirements of this research, several keywords representing the filtering process of Twitter text data streams are determined. In this research, the filtering keywords used are "vaccine," "PSBB" (Large-Scale Social Restrictions), "government," "Indonesia," and the main keyword "Covid-19." These keywords will be combined as phrases and used to limit the topic of tweets collected in this research. The data collection process is carried out through a tool in the form of a computer program connected to the internet and using the Python programming language.

b. Pre-Processing Stage

The pre-processing stage is required to ensure that the textual data generated from the previous process meets the requirements for further processing. Generally, tweets sent by Twitter users contain

abbreviations or other characters that complicate the analysis process. Therefore, these abbreviations and characters are removed without affecting the sentiment weighting of the entire sentence.

c. Feature Extraction Stage

The feature extraction stage is the process of searching for a set of words from tweet data that has passed through pre-processing. In other words, the text data entering this stage is already free from abbreviations or unnecessary characters, enabling word weighting analysis.

d. Classification Stage

This stage involves the determination of where the weighting results will be calculated and grouped into positive, neutral, or negative categories.

Results and Discussion

This section will discuss the testing and analysis results conducted on four keywords in tweets within the dataset.

The dataset used in this research was obtained from a previous study on lexicon-based sentiment analysis. There are two groups of tweet data successfully obtained on May 26, 2020, and May 28, 2020, with a total of 23,142 tweets. After filtering, only tweets containing the keyword "Covid-19" were obtained, totaling 19,980, with the following details:
Total data in tweets for "Covid-19": 19,980
Total tweets for May 26: 8,145
Total tweets for May 28: 11,678
Total Twitter accounts: 10,514

Testing was conducted by filtering specific keywords in the obtained tweet data, then analyzing the sentiment in each tweet using the TextBlob tool. In each tweet analysis process, the positive sentiment score will be calculated in the range of 0 to 1, while negative sentiment is in the range of -1 to 0. Neutral sentiment has no weight or is equal to 0.

First test: filtering the keywords "Covid-19" and "president" or "government" or "Indonesia."

In the first test, the keywords "Covid-19" and "president" or "government" or "Indonesia" were chosen. This means only tweets containing these words will be analyzed, and their sentiment scores will be calculated. This test aims to observe public sentiment on Twitter regarding the Covid-19 pandemic, which is a topic of discussion for the central government. The data obtained from this first test is as follows:
Number of tweets meeting the keyword criteria: 3,775
Number of positive sentiments: 960
Number of neutral sentiments: 2,231
Number of negative sentiments: 584
Total sentiment score: -30.34

The test results indicate that out of 3,775 tweets containing the keywords "Covid-19" and "president" or "government" or "Indonesia," there is a tendency for neutral public sentiment and a higher positive public sentiment compared to negative sentiment. However, another fact was found that the total score (positive sentiment score minus negative sentiment score) reached -30.34. Despite the smaller number of negative sentiments, their score is higher. This may suggest that, on average, tweets with negative sentiment use words with a stronger tendency.

Second test: using the keywords "Covid-19" and "governor" or "mayor" or "regent" or "regional government" or "city government."

In the second test, the keywords "Covid-19" and "governor" or "mayor" or "regent" or "regional

government" or "city government" were selected. This test aims to observe public sentiment on Twitter regarding the Covid-19 pandemic involving both level 1 and level 2 governments as topics. The data obtained from this second test is as follows:
Number of tweets meeting the keyword criteria: 343
Number of positive sentiments: 137
Positive sentiment score: 19.67
Number of neutral sentiments: 170
Number of negative sentiments: 36
Negative sentiment score: -6.72
Total sentiment score: +12.95

The test results indicate that out of 343 tweets containing the keywords "Covid-19" and "governor" or "mayor" or "regent" or "regional government" or "city government," there is a tendency for neutral and positive public sentiment to be more dominant than negative sentiment. Another fact found is that the total score (positive sentiment score minus negative sentiment score) reached +12.95. This indicates that the public is relatively satisfied with the handling of Covid-19 in the region and expresses it through the social media platform Twitter.

Conclusions

Based on the testing and analysis conducted on public sentiment data through the Twitter social media platform on several topics related to the Covid-19 pandemic, it can be concluded that the public tends to express neutral and positive sentiments more than negative sentiments in response to the Covid-19 pandemic and its handling. However, some individuals express negative sentiments through more critical criticisms than positive sentiments. This is evidenced by the higher average score of negative sentiment, even though the number is smaller than positive sentiment.

References

1. Akhmad Sa'rony, A. P. (2019). "Analisis Sentimen Kebijakan Pemindahan Ibukota Republik Indonesia dengan Menggunakan Algoritme Term-Based Random Sampling dan Metode Klasifikasi Naïve Bayes". Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer .
2. Covid-19, S. (2020). "Peta Sebaran Covid-19. Retrieved from Satgas Penanganan COVID-19": https://covid19.go.id/
3. DetikHealth, F. A. (2020). "Update Corona Indonesia 10 September: Rekor 3.861 Kasus Baru, Total 207.203". Retrieved from DetikHealth: https://health.detik.com/berita-detikhealth/d-5167585/update-corona-indonesia-10-september-rekor-3861-kasus-baru-total-207203
4. Fitri, W. (2020). "Implikasi Yuridis Penetapan Status Bencana Nasional Pandemi Corona Virus Disease 2019 (COVID-19) Terhadap Perbuatan Hukum Keperdataan". SUPREMASI HUKUM.
5. K. Sivarajah, U. I. (2019). "Role of big data and social media analytics for business to business sustainability: A participatory web context". Ind. Mark. Manag.
6. Karunia, A. M. (2020). "LIPI: 39,4 Persen Usaha di Indonesia Terhenti Akibat Pandemi Covid-19". Retrieved from Kompas.com: https://money.kompas.com/read/2020/05/19/214500926/lipi--39-4-persen-usaha-di-indonesia-terhenti-akibat-pandemi-covid-19
7. Kiruthika, S. (2016). "Sentiment Analysis of Twitter Data". IJIET .
8. Martin, S. K.-S. (2020). "The pandemic of social media panic travels faster than the COVID-19 outbreak Dengue Tools View project GBD 2013-2015 Study View project". Journal of Travel Medicine .

9. Pemerintah RI. (2020). “Keputusan Presiden Nomor 12 Tahun 2020 tentang Penetapan Bencana Nonalam Penyebaran Corona Virus Disease 2019 (COVID-19) sebagai Bencana Nasional”.
10. R. S. Luqyana, W. A. (2018). ”Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine”. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer .
11. Yogapreethi, N. (2016). “A Review on Text Mining in Data Mining”. Int. Journal Soft Computing .

