



Sentiment Analysis of Public Comments on Government Policies in Handling COVID-19: the Case of Indonesia

**Prayudi¹,
Muhammad Edy Susilo²,
Mangaras Yanu Florestiyanto^{3*}**

¹Communication Science, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia,
Email:- prayudi@upnyk.ac.id

²Communication Science, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia
Email:- muh_edy_susilo@yahoo.co.id

^{3*}Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia
Email:- mangaras.yanu@upnyk.ac.id

***Corresponding Author:-** Mangaras Yanu Florestiyanto

* Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia,
Email:- mangaras.yanu@upnyk.ac.id

Abstract.

Twitter is a social networking site that has a large user base. On Twitter, users may interact and talk about concepts and events. Social media usage is growing, which opens up new opportunities to study various communication aspects and trends. One of these is social media analysis, which has the specific goal of analyzing data to generate information. Twitter may be used to gain a thorough picture of current events and significant trends, such as how the Indonesian government handled the COVID-19 outbreak. Since the year 2020 began, COVID-19 has spread over the globe. The most recent information and viewpoints about the coronavirus are presented by various people, including media organizations and governmental bodies. In this study, tweet data was extracted from Twitter using the python programming language. After undergoing a number of pre-processing steps to clean up the data and get it ready for feature extraction, sentiment analysis, and classification, including positive sentiment, negative sentiment, and neutral sentiment using the Naïve Bayes algorithm, the data was then used in the study. The hashtag term "omicron" was used to collect data from Twitter. The three-week data search period runs from February 7 until February 27, 2022. A total of 106,834 pieces of data were successfully extracted from Twitter. The findings of the data analysis show that negative attitudes account for up to 46.44% of the total.

Keywords: covid-19, sentiment analysis, Twitter

1 Introduction

The Coronavirus pandemic, also known as COVID-19, started in Wuhan, China, early in 2020. In Indonesia and other parts of the world, Covid-19 has spread quickly. The Indonesian government has established many measures to combat the pandemic outbreak that should stop the virus from spreading [1]. Numerous studies on the first and second waves of Covid-19's spread revealed that the government lacked professionalism and failed to put the appropriate policies in place. Therefore, in the third wave, public attention would naturally be focused on the Indonesian government's management of Covid-19. The public's opinions on this issue, both for and against, are pretty diverse. These remarks are accessible through social media, which provides a platform for anybody to express their opinions openly. Twitter is one of the social media platforms that the public uses the most frequently to comment on and express ideas regarding COVID-19 [2]. Of course, if these comments are appropriately analyzed, they can offer valuable information [3]. These tweets' data can be used as a significant data source (big data).

Twitter is a well-liked social media platform on the web. Users can connect and discuss ideas and events on Twitter. The increased use of social media creates new chances to examine various features and communication patterns. For instance, social media data can be reviewed to learn more about topics, trends, influential people, and other information [4], [5]. One use of text mining that uses Twitter tweets is social media analysis, particularly on Twitter. One goal of text mining is to explicitly process old data to produce new information results [3]. As a result, this method is beneficial for deciphering information about a specific subject being discussed by Twitter users. Twitter can aid with insight into popular trends and significant events like government responses to the COVID-19 pandemic in Indonesia, even though it cannot be called an entire public conversation.

Sentiment analysis is one of the findings that may be obtained from text-mining tweet data [6]. Sentiment analysis will categorize the polarities of a collection of texts to determine if the opinions expressed are positive, negative, or neutral. On the basis of textual data, sentiment analysis is typically used to reveal public opinion on a topic, service satisfaction, and policy. From this, the propensity of public conversation messages on the subjects brought up is understood. Several earlier research on sentiment analysis used the terms "covid-19" or similar words on tweet data, such as Hendrawati and Yanti, attempted to utilize Twitter by employing sentiment analysis to determine the opinions on Covid-19 tweets that contained positive, negative, or neutral sentiments using Multi-layer Perceptron (MLP) and Backpropagation with Adam optimization. Bulk tweets concerning Covid-19 that were tweeted in Indonesian from November 2019 were gathered for this study, and the MLP Backpropagation model was trained from them. With f1-scores for the positive, negative, and neutral classes of 0.77, 0.75, and 0.5 from the maximum value of 1, this model was able to achieve an accuracy of up to 70%. This study demonstrates that the Covid-19-related tweets from Indonesia may be successfully classified using the current approach [7].

Another study by Machuca et al. used the hashtag #coronavirus. It successfully attracted 1,600,000 tweets from January 2020 to July 2020, showing that classification using the binary logistic regression algorithm has an accuracy of 78.5%. Opinions that have been successfully classified are generally dominated by positive sentiment. To summarize, 54% of users show positive feelings, and 46% show negative emotions [8].

Corona, COVID, and COVID-19 were used as keywords in a study by Vijay et al. in India. To acquire the polarity of each tweet that is utilized as a dataset for this study's training and testing, natural language processing (NLP) is used. The conclusion is that occasionally, over the research period, from January to May 2020, the positive polarity increased, and the negative polarity shrank to 19% [9]. Manguri et al.'s study utilizing the terms "coronavirus" and "covid-19." With the help of the Naive Bayes classification technique, Manguri et al. analyze tweet data. According to the study's findings, the neutral number reaches a very high polarity of more than 50%, with the majority being objectively documented at up to 64% [10].

This study aims to assess the trend of information connected to government policies in dealing with the third wave of Covid-19 attacks based on big data on Twitter using the data mining approach, per the description provided before. This study's examination of information trends focuses on sentiment analysis of government policies relating to the management of COVID-19.

2 Literature Review

Sentiment analysis on Twitter and other social media platforms is a hot topic among scholars. This section includes several important articles that have made significant advances to our understanding of user actions and circumstances in many global scenarios.

Text analysis has been applied in a number of different contexts over the years, including document clustering [11], email sorting [12], irony and sarcasm detection [13], hate speech detection [14], sentiment prediction, and opinion mining [15], biomedical text mining [16], [17], content mining [18], question answering system [19], and many others. Emotional analysis has also been heavily used in Twitter data [20]–[22]. Due to the vast amount of data, intriguing discoveries are sometimes concealed in plain sight.

Boldog et al. [23] describe how the strict communal quarantine brought on by the COVID-19 Pandemic has affected Filipinos, particularly in Luzon. Based on user tweets, the researcher also examines how a strict communal quarantine and other pandemic consequences may affect a person's personal lifestyle. From retrieved tweets, natural language processing techniques are used to gauge user attitudes. Opinions are used as analysis data. In order to assess the consequences of the strict community quarantine in the Luzon region, a qualitative method was also applied.

Twelve states' worth of tweets were gathered and examined by Dubey [24] for this study. All of the tweets were related to the unique COVID-19 illness, and they were all gathered between March 11 and March 31, 2020. The purpose of this investigation is to understand how people are responding to the illness epidemics in various nations. There is no question that he had to do certain essential actions, such as pre-processing and eliminating unnecessary information from the tweets, in order to complete the job. The results of these studies indicate that while the majority of individuals in these civilizations tend to think positively and believe that things will get better, it is also important to note that there are still indications of dread and melancholy. Due to epidemics and pandemics affecting significant populations, four governments, particularly those from the European continent, believe they cannot trust the situation.

In a research conducted by Vijay et al. in India, the terms Corona, COVID, and COVID-19 were utilized as keywords. Natural language processing (NLP) is used to obtain the polarity of each tweet that is used as a dataset for this study's training and testing. The study's findings indicate that the positive polarity occasionally grew and the negative polarity occasionally shrank to 19% over the research period, which spanned from January to May 2020 [9]. Using the names "coronavirus" and "covid-19," Manguri et al. [10] examine Twitter data using the Naive Bayes classification algorithm. The neutral percentage reaches a very high polarity of more than 50%, with the majority being objectively confirmed at up to 64%, according to the study's conclusions. The effects of COVID-19 were examined and depicted by Ra et al. [25] by doing sentiment analysis on Twitter datasets using algorithms and machine learning techniques. This research tries to comprehend both the extremely favorable and extremely negative attitudes held by the general people worldwide. This study shows that the Naive Bayes machine learning strategy improved categorization and is regarded as a fundamental idea in learning. Another ensemble approach that employs sentiment scores as the input function for machine learning classifiers, Max Entropy, SVM, Boosting, Decision Tree, and Random Forest is also raised by this research. As a consequence, the hybrid strategy LogitBoost outperformed the competition with 74% accuracy.

3 Research Methodology

Sentiment analysis of tweets in this study consists of several steps. First of all, collect tweet data. Furthermore, pre-processing of the data that has been collected, labeling, and followed by a classification process to classify tweets as having positive, negative, or neutral sentiments. The last step is to measure the performance of the classification model by using the confusion matrix.

3.1 Data collection

The data used is secondary data in the form of tweets from Twitter in Indonesian. The data is retrieved through a crawling process using Python. The population in this study were all tweets using the keyword (hashtag) "#omicron". The sample is tweet data from February 7 to February 27, 2022. The variables used are two attributes and one label. The attributes used include the date of the tweet and the tweet. The labels used for classification are labeling data from sentiment analysis. The label data types are categorical, namely positive, negative, and neutral.

3.2 Pre-processing

Pre-processing has 5 stages: cleansing, case folding, stopword removal, tokenizing, and stemming. Each step is described as follows:

a. Cleansing

Cleansing is a word-cleaning stage that does not affect the results of sentiment classification. The tweet document component has various attributes that do not affect sentiment. Examples are mentions starting with the '@' attribute, hashtags beginning with the '#' attribute, links starting with the 'HTTP', 'bit.ly' attribute, and the symbol character ~!@#\$\$%^&*()_+?<>.,?:{}[]|. Attributes that have no effect will be removed from the document and replaced with a space character.

b. Case folding

The case folding stage is the stage for converting capital letters in the review text into lowercase letters. This stage aims to make the review text document has a standard form.

c. Stopword removal

At the stopword removal stage, removing words that are not in accordance with the document topic will be carried out if there are words that do not affect the accuracy of the document sentiment classification process. The words to be omitted are collected in a stopword word database. If in the tweeted document, there is a word that matches the word in the stopword, then the word will be removed and replaced with a space character.

d. Tokenizing

This stage is the tokenization process, which is dividing the text contained in a sentence or paragraph by cutting words based on each word that composes them into a single piece. The words in the document in question are words separated by spaces. So that the result of this tokenization process is a single word that is entered into the database for weighting purposes.

e. Stemming

At this stage is stemming, which is a process carried out to change the words contained in a document into basic words using certain rules. The Indonesian stemming process is done by removing suffixes, prefixes, and confixes in the document.

3.3 Labeling

This stage is in the form of labeling Tweets with positive, negative, and neutral labels. Sentiment labeling on training data is done manually. This sentiment labeling is done so that the training data has a category so that the training process and data testing can be carried out. In doing this manual categorization, the comment data held is divided into 3 data categories, namely comments containing positive sentiments, comments containing negative sentiments, and neutral comments.

3.4 Sentiment classification

Based on prior knowledge, this Naive Bayes algorithm will forecast opportunities for the future. According to Taheri [26], Naive Bayes is a well-liked classification method with a high level of efficacy for data mining and machine learning. It is also simple to apply. The Naive Bayes classification technique takes conditional probability weights into account. The Naive Bayes classifier's structure is used to model and calculate the objective function, whose attribute is weight. The local optimization technique uses the quasecant approach to identify the ideal weight. The Naive Bayes classifier is used as the foundation for the suggested strategy. Results of numerical tests performed on several real-world data sets in binary classification, prove the effectiveness of the suggested approach.

4 Finding and Discussion

We gained a lot of knowledge about Indonesian people's emotional states and how those states change throughout the day throughout the analysis phase of the process. On February 9, 2022, when there was a total of about 750 talks, the distribution of conversations including the term "Omicron" during a period of seven days, from February 7 to February 13, 2022, was found to be the most active. As of now, there will be a minimum of 450 discussions on February 10, 2022. The distribution data is shown in Figure 1.



Fig. 1. Distribution of talks (07 February - 13 February 2022).

A total of 38.4K tweets in Indonesia with the term "Omicron" were produced as a consequence of the dialogue over the course of one week, and these tweets received up to 84.5K interactions. Positive interactions made up 7.4% of the tweets, negative conversations made up 48.8%, and neutral conversations made up 44.6%.

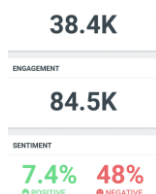


Fig. 2. Sentiment analysis for 07 February - 13 February 2022.



Fig. 3. The top theme for 07 February - 13 February 2022.

At the same time, several top themes were produced, dominated by "omicron", "covid", and "variant". In terms of location, Figure 4 shows the dominance of tweets about Omicron dominated by Jakarta at 79.5%, followed by West Java at 3.7%, East Java at 3.3%, Central Java at 2.4%, etc.

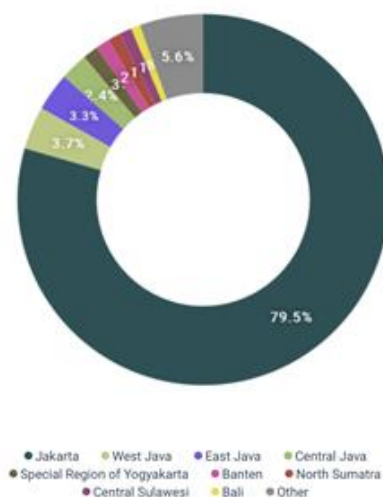


Fig. 4. Distribution of tweets by location for 07 February - 13 February 2022.

On February 19, with a total of more than 1100 talks, the distribution of conversations including the term "Omicron" during a period of seven days, namely between February 14 and February 20, 2022, was found to be the most active. While just 300 tweets were sent on February 14, which was the day with the lowest amount. This distribution can be seen in Figure 5.



Fig. 5. Distribution of talks (14 February - 20 February 2022).

Figure 6 displays the outcomes of the talks that took place between February 14 and February 20, 2022, resulting in a total of about 43.5K tweets and a total of 135.8K interactions. There were 9.5% of tweets with a positive attitude, followed by tweets with a neutral mood of up to 40.9% and tweets with negative sentiment of up to 49.6% based on the total number of tweets. And Figure 7 shows the top themes for the period 14 February to 20 February 2022.

RESULTS	43.5K
ENGAGEMENT	135.8K
SENTIMENT	9.5% POSITIVE, 49.6% NEGATIVE

Fig. 6. Sentiment analysis for 14 February - 20 February 2022.



Fig. 7. The top theme for 14 February - 20 February 2022.

The top 9 geographical areas where Twitter users spread messages about Omicron were identified by this Twitter analyst's findings. These are the figures:

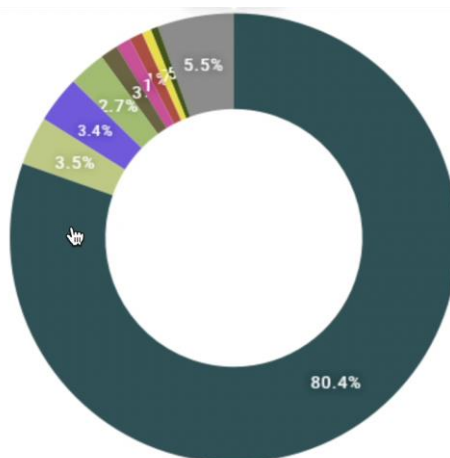


Fig. 8. Distribution of tweets by location for 14 February - 20 February 2022.

West Java had 3.5% of the 35K tweets, East Java had 3.4% of the 35K tweets, and Central Java had 2.7% of the 1.2K tweets. Jakarta had 80.4% of the 35K tweets. DIY had 562 tweets or 1.3% of all tweets; Banten had 464 tweets or 1.1% of all tweets; North Sumatra had 422 tweets or 1%; Bali had 296 tweets or 0.7%, and South Sulawesi had 231 tweets or 0.5%.

On February 25, 2022, with a total of over 700 talks, the distribution of conversations including the term "Omicron" throughout a period of seven days, namely between February 21 and February 27, 2022, was found to be the most active. The day with the fewest discussions, February 26, 2022, saw less than 100. The distribution of talks for the time span between 21 February - 27 February 2022 can be seen in the following figure.



Fig. 9. Distribution of talks (21 February - 28 February 2022).

Figure 10 displays the volume of talks in Indonesia including the term "Omicron," which amounts to about 24.8K tweets with an overall engagement of 84.5K. In contrast, good interactions made up 21.4% of these tweets, negative conversations made up 38.5%, and neutral conversations made up 40.1%. While "#PatuhHindariKerumunan", "#Omicron", and "#COVID19" are among the top conversational hashtags in Figure 11 for the period of February 21–27, 2022.

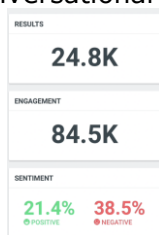


Fig. 10. Sentiment analysis for 14 February - 20 February 2022.



Fig. 11. The top theme for 14 February - 20 February 2022.

By location, the dominance of tweets about Omicron is dominated by Jakarta with a total of 79.8%. This can be seen in Figure 12.

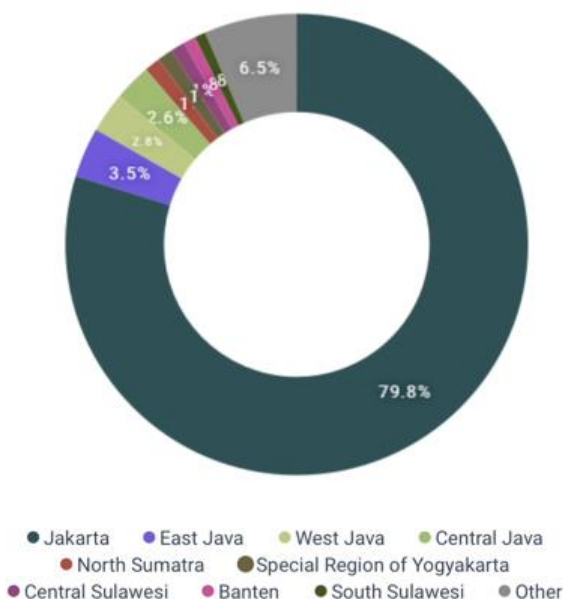


Fig. 12. Distribution of tweets by location for 14 February - 20 February 2022.

Overall, the analysis's findings indicate that Twitter users' attitude tends to be negative in relation to the 49,617 total tweets that were sent during the observation period (46.44%). The negative sentiment did not surpass the neutral sentiment until the third week. The table below displays these findings.

Table 1. The analysis's findings.

Sentiment	First week		Second Week		Third Week		Total	
Positive	2.843	7.4 %	4.138	9.5 %	5.318	21.4 %	12.299	11.5 %
Negative	18.442	48 %	21.607	49.6 %	9.568	38.5 %	49.617	46.44 %
Neutral	17.136	44.6 %	17.817	40.9 %	9.965	40.1 %	44.918	42.06 %
Total	38.421	100 %	43.562	100 %	24.851	100 %	106.834	100 %

Negative tweets often point the finger at the government since, up until the third attack, it had not been concentrating on resolving the Covid issue. Several concerns that are viewed as the government's attempts to hide its failure to deal with the epidemic emerged in some of the top tweets in this category.

With 44,918 tweets (42.06%), neutral sentiment is less prevalent than negative sentiment. Users that fit this description on Twitter frequently post about the value of adhering to health guidelines as well as the milder symptoms of the Delta Variant Covid-19 virus. Additionally, if taken in a more focused manner, neutral sentiment is mostly produced through the tweets of governmental organizations and online media accounts that disseminate information. Typically, tweets from these accounts take the form of website news leads. Thus, it frequently does not fall into either the positive or negative category.

Neutral tweets frequently urge the implementation of health protocols in response to the rising number of omicron infections in Indonesia. Several tweets from governmental organizations demonstrate this. However, the Polda Twitter account's tweets, not the Health Service's, were the ones that garnered more attention. For instance, the tweet from the @PIDM Bojonegoro account has the most interaction, with 48 likes and 45 retweets. This can mean that continuing with health precautions, such as using masks indoors and outside, is still important. Although tweets with neutral sentiment are high, the engagement of neutral sentiment tweets is not as high as negative sentiment tweets.

The last category, positive sentiment, had a total of 12,299 (11.5%) tweets. This indicates that Twitter users still have an optimistic outlook despite the fact that more and more individuals are contracting the Omicron variety of the Covid-19 epidemic. This analysis showed nearly no tweets in favor of how the government is handling the epidemic, despite the fact that thousands of tweets were recognized as having favorable sentiment. Conversely, tweets with positive feelings frequently feature personal stories from Twitter users who are thankful to be well in the face of the growing third wave of Omicron assaults.

This study paints a picture of how Twitter social media users in Indonesia felt about the Omicron Covid-19 variant attack in Indonesia based on the sentiment analysis displayed in Table 1 and the textual analysis shown through the top tweet samples included in each sentiment category. Twitter users appear to be considering how the Indonesian government handled the first and second waves of the covid pandemic that happened in 2020 and 2021 in relation to the government's strategy for handling the Omicron variant onslaught. Government policies tended to change aimlessly and be implemented at this time (Jaffrey & Jaffrey, 2020; Lesmana & Aranditio, 2021; Lindsey & Mann, 2020; Prayudi et al., 2021; Prayudi & Sagita, 2021).

This study also suggests that the Ministry of Health and the Task Force for the Acceleration of Handling Covid-19, the authorized government entities, have not yet fully optimized their use of social media for crucial Covid-19 crisis communication. Due to the Covid-19 epidemic, people now perform activities that were formerly performed in real settings online, making them more susceptible to social media fraud. The sentiment analysis of this study, however, seldom ever came across tweets from legitimate government organizations.

5 Conclusion

This study's application of textual sentiment analysis was successful in categorizing tweets textually and increasing Indonesian Twitter users' engagement. It also clarifies the backdrop for why some tweets garner more interaction than others. By defining the text's context, textual sentiment analysis extends beyond count-based metrics to offer insight into the content.

This study demonstrates how vocal Indonesian Twitter users are about the third Covid-19 attack and how the Indonesian government is handling it. Users continue to believe that the Indonesian

government has not taken the third wave of the Covid-19 epidemic seriously, as seen by the dominance and high engagement of Twitter users' unfavorable sentiment data on the government's handling of the Omicron variety.

The government should carry out its responsibility to effectively engage the people through a variety of channels, including social media. The advantages of social media for the government include promoting government effectiveness, regaining the public trust that has been lost over concerns of government policy, addressing technological obstacles, and being utilized to communicate during crises and natural catastrophes.

Acknowledgments

References

- R. Djalante et al., "Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020," *Progress in Disaster Science*, vol. 6, p. 100091, Apr. 2020, doi: 10.1016/j.pdisas.2020.100091.
- F. C. Permana, Z. M. Wicaksono, C. Kurniawan, A. S. Abdullah, and B. N. Ruchjana, "Perception analysis of the Indonesian society on twitter social media on the increase in BPJS kesehatan contribution in the Covid 19 pandemic era," in *Journal of Physics: Conference Series*, Jan. 2021, vol. 1722, no. 1. doi: 10.1088/1742-6596/1722/1/012022.
- M. Syarifuddin, "ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN KNN," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, Aug. 2020, doi: 10.33480/inti.v15i1.1347.
- L. Zhang, J. Zhao, and K. Xu, "Who creates Trends in Online Social Media: The Crowd or Opinion Leaders?," *Journal of Computer-Mediated Communication*, vol. 21, no. 1, pp. 1–16, Jan. 2016, doi: 10.1111/jcc4.12145.
- D. Yeung, "Social Media as a Catalyst for Policy Action and Social Change for Health and Well-Being: Viewpoint," *J Med Internet Res*, vol. 20, no. 3, p. e94, Mar. 2018, doi: 10.2196/jmir.8508.
- R. Risnantoyo, A. Nugroho, and K. Mandara, "Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 4, no. 1, pp. 86–96, Jul. 2020, doi: 10.31289/jite.v4i1.3798.
- T. Hendrawati and C. P. Yanti, "Analysis of Twitter Users Sentiment against the Covid-19 Outbreak Using the Backpropagation Method with Adam Optimization," *Journal of Electrical, Electronics and Informatics*, vol. Vol. 5, no. 1, 2021.
- C. R. Machuca, C. Gallardo, and R. M. Toasa, "Twitter sentiment analysis on coronavirus: Machine learning approach," in *Journal of Physics: Conference Series*, Mar. 2021, vol. 1828, no. 1. doi: 10.1088/1742-6596/1828/1/012104.
- T. Vijay, A. Chawla, B. Dhanka, and P. Karmakar, "Sentiment Analysis on COVID-19 Twitter Data," Dec. 2020. doi: 10.1109/ICRAIE51050.2020.9358301.
- K. H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin, "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks," *Kurdistan Journal of Applied Research*, pp. 54–65, May 2020, doi: 10.24017/covid.8.
- K. M. Hammouda and M. S. Kamel, "Efficient phrase-based document indexing for Web document clustering," *IEEE Trans Knowl Data Eng*, vol. 16, no. 10, pp. 1279–1296, Oct. 2004, doi: 10.1109/TKDE.2004.58.
- X. Carreras and L. Marquez, "Boosting Trees for Anti-Spam Email Filtering," in *Proceedings of RANLP-2001*, Sep. 2001, pp. 58–64.
- U. Naseem, I. Razzak, P. Eklund, and K. Musial, "Towards Improved Deep Contextual Embedding for the identification of Irony and Sarcasm," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, pp. 1–7. doi: 10.1109/IJCNN48605.2020.9207237.
- U. Naseem, I. Razzak, and I. A. Hameed, "Deep Context-Aware Embedding for Abusive and Hate Speech detection on Twitter," 2019.
- U. Naseem, I. Razzak, K. Musial, and M. Imran, "Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis," *Future Generation Computer Systems*, vol. 113, pp. 58–69, Dec. 2020, doi: 10.1016/j.future.2020.06.050.
- U. Naseem, M. Khushi, V. Reddy, S. Rajendran, I. Razzak, and J. Kim, "BioALBERT: A Simple and Effective Pre-trained Language Model for Biomedical Named Entity Recognition," Sep. 2020.
- [17] U. Naseem, K. Musial, P. Eklund, and M. Prasad, "Biomedical Named-Entity Recognition by Hierarchically Fusing BioBERT Representations and Deep Contextual-Level Word-Embedding," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9206808.
- C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*. 2013.
- V. Gupta and G. S. Lehal, "A Survey of Text Mining Techniques and Applications," *Journal of Emerging Technologies in Web Intelligence*, vol. 1, no. 1, Aug. 2009, doi: 10.4304/jetwi.1.1.60-76.
- G. Rocha and H. Lopes Cardoso, "Recognizing Textual Entailment: Challenges in the Portuguese Language," *Information*, vol. 9, no. 4, p. 76, Mar. 2018, doi: 10.3390/info9040076.
- G. Carducci, G. Rizzo, D. Monti, E. Palumbo, and M. Morisio, "TwitPersonality: Computing Personality Traits from Tweets Using Word Embeddings and Supervised Learning," *Information*, vol. 9, no. 5, p. 127, May 2018, doi: 10.3390/info9050127.

- T. Ahmad, A. Ramsay, and H. Ahmed, "Detecting Emotions in English and Arabic Tweets," *Information*, vol. 10, no. 3, p. 98, Mar. 2019, doi: 10.3390/info10030098.
- P. Boldog, T. Tekeli, Z. Vizi, A. Dénes, F. A. Bartha, and G. Röst, "Risk Assessment of Novel Coronavirus COVID-19 Outbreaks Outside China," *J Clin Med*, vol. 9, no. 2, p. 571, Feb. 2020, doi: 10.3390/jcm9020571.
- A. D. Dubey, "Twitter Sentiment Analysis during COVID19 Outbreak," *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3572023.
- M. R. B. A, and S. K, "COVID-19 Outbreak: Tweet based Analysis and Visualization towards the Influence of Coronavirus in the World," *GEDRAG & ORGANISATIE REVIEW*, vol. 33, no. 02, Apr. 2020, doi: 10.37896/GOR33.02/062.
- S. Taheri and M. Mammadov, "Learning the naive Bayes classifier with optimization models," *International Journal of Applied Mathematics and Computer Science*, vol. 23, no. 4, pp. 787–795, Dec. 2013, doi: 10.2478/amcs-2013-0059.