JHSS (Journal of Humanities and Social Studies)
https://journal.unpak.ac.id/index.php/jhss
Volume 08, Number 01, Page 298-302
e-ISSN: 2598-120X; p-ISSN: 2598-117X

ANALYZING TWITTER SOCIAL MEDIA USER RESPONSES TO FS CASES USING SENTIMENT ANALYSIS AND TEXT MINING

Muhammad Rafiansyah [a*)], Nurvita Trianasari [a)]

[a)] *Telkom University, Bandung, Indonesia*

[*)]*Corresponding Author: muhammadrafiansyah001@gmail.com*

Article history: *received 09 August 2023; revised 12 September 2023; accepted 04 October 2024* *DOI:https://doi.org/10.33751/jhss.v8i1.8640*

Abstract. There are several factors behind the research on analyzing the response of social media users, especially on Twitter, to the Ferdy Sambo case. Some of these factors include the increasing use of social media and the widespread adoption of platforms like Twitter in Indonesia to comment on or express opinions about various matters or cases, particularly the Ferdy Sambo case which is the focus of this research. The purpose of this study is to determine how Twitter users respond to the Ferdy Sambo case and to assess the sentiment analysis results from Twitter users based on three categories: positive, negative, and neutral sentiments. Through the utilization of sentiment analysis and text mining, this research aims to uncover commonly used words in comments and responses from Twitter users regarding this case. These frequently used words will then be visualized using word clouds to display both their prevalence and frequency. Based on the analysis conducted, the findings reveal that out of the 15,002 data points, the words or opinions expressed by Twitter users are predominantly negative. These outcomes are further substantiated by accuracy, precision, and recall rates calculated at 97%, 95%, and 90%, respectively. The implications of this study are anticipated to offer insights into the responses of social media users towards a particular case, wherein negative situations generally elicit pessimistic opinions and vice versa. Moreover, it is envisaged that this research could serve as a foundation for future studies on similar subjects or employing similar methodologies.

Keywords: sentiment analysis; twitter; word cloud; negative; neutral; Ferdi Sambo

I. INTRODUCTION

A case within the Indonesian police institution in mid-2022, precisely between July and August, garnered significant public attention. The incident involved a shootout among police officers, resulting in the death of Brigadier Nofriansyah Yosua Hutabarat. Subsequent investigations unveiled that this was a premeditated murder, orchestrated by FS, Brigadier Nofriansyah Yosua Hutabarat's superior. The facts that came to light prompted various public opinions on the matter, which were expressed through social media platforms, particularly Twitter. This mirrors the actions of Twitter users during the Christchurch shooting terrorism case in 2019 [1].Twitter stands as one of the most extensively used media platforms by internet users globally, including Indonesia. In Indonesia, Twitter holds the fifth position among the most prevalent social media platforms [2]**.** The number of Twitter users in Indonesia has consistently grown since 2019 with 6.425.000 user until 2022 with 18.450.000 [3]**.** With so many Twitter users, the FS case certainly became a topic of conversation. This case became a trending topic for a considerable time on Twitter. With this case trending, of course, there are many diverse responses about this case, ranging from positive, negative, to neutral responses.

In this study, there exists a research gap compared to previous studies that have tackled the same research theme, particularly concerning the sentiments of internet users, especially those on Twitter, toward a trending case. A previous research work authored by Samah Mansour titled " Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining" illustrates that people employ nearly identical words when tweeting about ISIS. This shared usage of both positive and negative terms reflects how people perceive this organization as a terrorist group, regardless of their country of origin. Furthermore, most users view ISIS as a source of threat and fear, irrespective of their geographical location [4]**.** Given the variations in the research subjects and objects in this study, it is hoped that this research will bridge the existing gap in understanding public sentiment towards a case. In the context of this study, the focus is on public sentiment towards the FS case on Twitter, involving the Indonesian population who use and express their opinions on this platform.

According to Henry Fayol, the concept of management involves the process of planning, organizing, coordinating, and controlling/directing available resources to

Volume 08, Number 01, Page 298-302
e-ISSN: 2598-120X; p-ISSN: 2598-117X

achieve goals effectively and efficiently [5]. On the other hand, as stated by Dr. Malayu S. P. Hasibuan, management is considered both an art and a science for overseeing and handling existing resources, including both human and non-human elements. These resources are processed and organized to attain specific objectives [6].

Big Data refers to a collection of data that surpasses the processing capacity of conventional database systems. This can involve data being too extensive, moving too rapidly, or not being compatible with a typical database architecture. To extract value from such data, alternative methods of processing must be chosen [7]. Additionally, according to Doug Laney, big data encompasses the increasingly varied data that companies and related organizations acquire. He states that this involves not only the volume of data but also the speed at which data is generated or updated [8].

Text mining is a term that refers to techniques capable of analyzing semi-structured and unstructured text data. Essentially, text mining is an interdisciplinary field related to Information Retrieval, Data Mining, Machine Learning, Statistics, and Computational Linguistics [9]. According to Marti A. Hearst, Text Mining can be defined as the discovery of new and previously unknown information by computers, automatically extracting information from various sources. The key to this process is successfully combining information taken from various sources [10]. This perspective aligns with the viewpoint of Harlian Milkha, who defines text mining as data mining in the form of text, with the data source usually derived from documents, and the goal being to identify words that can represent the document's contents, enabling connectivity analysis between documents [11].

Machine learning is one of the branches of artificial intelligence. According to [12], a computer scientist and one of the pioneers of machine learning, machine learning is the study of computer algorithms that can automatically enhance the performance of computer programs based on previous data. Machine learning operates by collecting, analysing, and contrasting both small and large data sets to identify patterns and explore distinctions. There are three types of machine learning: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is an algorithm that models the relationships and dependencies between the predicted target outputs and inputs, allowing us to predict output values for new data based on pre-established relationships and dependencies. Unsupervised learning, on the other hand, is a machine learning algorithm primarily used for pattern detection and descriptive modelling [12].

Social media data analysis is a complex process due to the unstructured nature of the data. There are two major approaches to processing social media data, namely based on content and structure. The content approach generally uses methodologies such as Sentiment Analysis, while the structural approach mostly employs Social Network Analysis (SNA) [13].There are four levels in sentiment analysis, namely document, sentence, entity, and aspect levels [14]. The approach in this research utilizes a sentiment analysis approach.

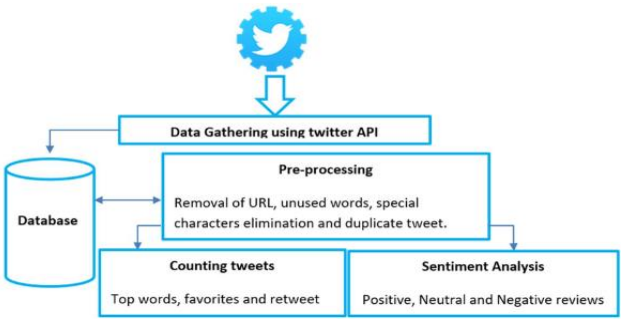

Figure 1. Conceptual Framework

Source : [15] [15]

This research aims to analyze the opinions of Twitter users, as Twitter is one of the social media platforms that can be used to express opinions through writing or text called "tweets." This analysis is conducted using the sentiment analysis method, which is a process that automates the extraction of opinions, attitudes, views, and emotions from text, such as tweets, through Natural Language Processing (NLP). The subject of this research is FS, who is an Inspector General of the Indonesian National Police. Regardless of any controversies surrounding the case he is currently involved in, which has become a topic of debate in ongoing trials, the researchers do not intend to discuss the truth or controversies of the case itself. Instead, the researchers aim to understand how Twitter users' opinions or responses towards Mr. FS as a whole are reflected through the Sentiment Analysis & Text Mining method. Sentiment analysis involves categorizing opinions in text into categories such as "positive," "negative," or "neutral" [4]. In this research, sentiment analysis is employed to determine whether opinions are positive or negative.

II. RESEARCH METHODS

Based on the methodological category, this type of research is included in qualitative research because in this study, the focus is the opinion conveyed through tweets on twitter social media, of course, the tweets on twitter contain a collection of words about the FS case. Qualitative research is research designed to tell researchers how (process) and why (meaning) something happens as it does. Qualitative research includes "a range of interpretive techniques that seek to describe, decode, translate, and otherwise understand the meaning, not the frequency, of certain naturally occurring phenomena in the social world" [16]. Then, according to the research objectives, this research can be categorized as descriptive research, because this research aims to describe the opinions of Twitter users on the FS case. Furthermore, based on the unit of analysis, this research is sourced from individual data of each Twitter user, especially those related to their opinions on the FS case through tweeting words on Twitter. The researcher's involvement in this research is minimal because the research intervention in this research only includes data collection and analysis. And the time of implementation of this research is only done once, so it can

Volume 08, Number 01, Page 298-302
e-ISSN: 2598-120X; p-ISSN: 2598-117X

be categorized as research with a cross-sectional implementation time.

The sampling method employed in this study is the use of Non-probability Sampling (NPS) techniques. Non-probability Sampling is a sampling technique that does not provide equal opportunity for every element or member of the population to be selected as a sample [16]. The sample for this research consists of tweets from Twitter users. The keyword used in the data search process is "FS," and the data was collected in Indonesian over a period of 273 days, from July 07, 2022, to April 06, 2023. This resulted in a total of 15.002 tweets collected from Twitter social media. These tweets were then stored in CSV format for further processing and analysis.

Data analysis techniques are carried out in several stages. First, Text mining, an automated text mining technique, is used to identify, extract, manage, integrate, and efficiently exploit knowledge from text [17]. The tools used to support the text mining process in this method of collecting text data are Orange Data Mining version 3.34, using data from Twitter social media. Subsequently, the collected text data will be processed through the sentiment analysis method. In the second step, as stated by [18], one of the initial stages in facilitating data classification is data preprocessing. Generally, the stages or steps of data preprocessing in text classification involve four steps, namely Tokenization, Stopword Removal, Lowercase Conversion, Stemming. The third step, following data collection, involves performing sentiment analysis. As explained by Mansour (2018) [4], sentiment analysis utilizes natural language processing to classify text into three categories: positive, negative, and neutral.

III.RESULT AND DISCUSSION

A. Data Characteristics

In this study, data were obtained from Twitter social media using the keyword "Ferdy Sambo". The data were collected within a period of 273 days, starting from July 07, 2022, to April 06, 2023. The details of the collected data are as follows:

Table 1 Total Twitter Data

	Key Word
	Ferdy Sambo
Total Data	15.002

Source: processed by author, 2023

B. Results of Sentiment Analysis

Based on the 15.002 data that have been obtained using the keyword "Ferdy Sambo," 8.209 data are categorized as negative sentiment, 4.848 data are categorized as neutral, and 1.945 data are categorized as positive, as shown in the figure.

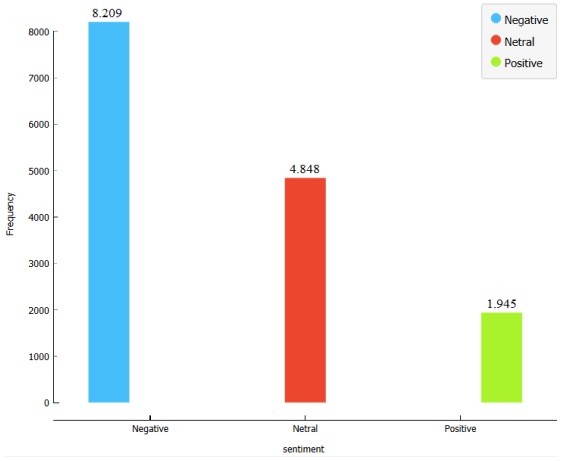

Figure 2 Distribution Bar Chart of Sentiment Analysis on the Keyword "Ferdy Sambo"

Source: processed by author, 2023

The results of the Sentiment Analysis evaluation for the keyword "Ferdy Sambo" can be observed in the confusion matrix table shown in Figure 3.

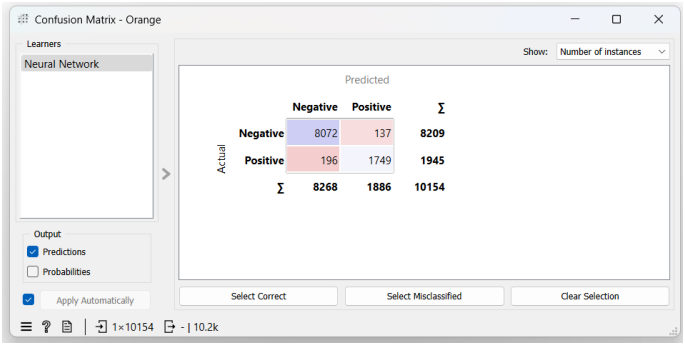

Figure 3 Confusion Matrix for the Keyword "Ferdy Sambo"

Source: processed by author, 2023

Based on the Confusion Matrix image above, to measure the performance of sentiment analysis classification, several calculations need to be performed, including accuracy rate, precision rate, and recall rate. The calculation formulas are as follows:

- Tingkat Akurasi/*Accuracy*:

$$\frac{TP+TN}{TP+FP+FN+TN} = \frac{1749+8072}{1749+137+196+8072} = \frac{9821}{10154} = 0{,}97 = 97\%$$

- Tingkat Presisi/*Precision*:

$$\frac{TP}{TP+FP} = \frac{1749}{1749+137} = \frac{1749}{1886} = 0{,}95 = 95\%$$

- Tingkat *Recall*:

$$\frac{TP}{TP+FN} = \frac{1749}{1749+196} = \frac{1749}{1945} = 0{,}90 = 90\%$$

C. Word Cloud Results

Before entering this stage, the Word Cloud process is assisted by Pre-process Text Stopwords. The aim is to remove irrelevant keyword content that doesn't pertain to the topic of

Volume 08, Number 01, Page 298-302
e-ISSN: 2598-120X; p-ISSN: 2598-117X

discussion, in accordance with the chosen keyword theme. After Pre-process Text Stopwords is completed, the word cloud process follows. A word cloud is a visualization of a collection of various words. The size of the words in the word cloud indicates how frequently a word appears in the data. In other words, the larger the size of the word, the more often it appears. This can be observed in the word cloud results. Here are the outcomes of the word cloud visualization, showcasing the most frequently appearing keywords or words in the context of "Ferdy Sambo":

Figure 4 Word Cloud Visualization with the Keywords "Ferdy Sambo"
Source: processed by author, 2023.

It was found that the word "sambo" had the highest repetition count of 14,559 in the tweets, followed by the word "ferdi" with 13.822 repetitions.

D. Word Frequency

The frequency of negative and positive words for each word obtained through text mining techniques using the keyword "Ferdy Sambo" can be observed in the following table and diagram:

Table 2 Frequency of Negative Words

Word	Negative Word Total
Mati	6.120
Hukuman	4.186
Pembunuhan	1.562
Voni	4.486
Divoni	2.662

Source: processed by author, 2023.

Table 3 Frequency of Positive Words

Word	Positive Word Total
Agung	554
Hidup	418
Berharap	199
Teravrbukti	456
Kasih	299

Source: processed by author, 2023.

Table 4 Frequency of Neutral Words

Word	Neutral Word Total
Sambo	14.559
Ferdi	13.822
Brigadir	2.376
Hakim	3.271
Chandrawati	1.455

Source: processed by author, 2023.

E. Discussion of Research Findings

From the 15,002 data points, the results indicate that the majority of opinions or tweets on Twitter are negative regarding the shooting and murder case involving FS, accounting for a total percentage of 55% or 8.209 tweets. This is followed by neutral opinions, comprising 32% or 4.848 tweets, and positive opinions, accounting for 13% or 1.945 tweets. The sentiment analysis outcomes, primarily reflecting negative opinions, are consistent with the accuracy, precision, and recall rates calculated using the aforementioned formulas. These rates are as follows: an accuracy rate of 97%, precision of 95%, and recall rate of 90%.

This sentiment tendency is also evident from the frequency of word usage in the Twitter tweets. Some of the most frequently appearing negative words include “mati”, “pembunuhan”, “hukuman”, “vonis”, and “divonis” Moreover, several neutral words are recurrent in users' tweets, including “Sambo”, “Ferdi”, “Ferdy Sambo”, “Chandrawati”, “brigadir” and “hakim”. On the other hand, several positive words were used in tweets related to FS, such as “Agung”, “hidup”, “berharap”, “terbukti”, dan kasih.

The outcomes of this research can be employed in data management. By filtering out off-topic opinions of internet users or refining each existing opinion, it becomes possible to discern internet users' sentiments about a given subject—be it positive, negative, or neutral. Consequently, management can establish the subsequent course of action based on the data analysis findings using a sentiment analysis approach.

F. Comparison of Research Results

Regarding the analysis results, there are comparable sentiment analysis outcomes between this research and a previous study on ISIS titled "Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining." In both investigations, the prevalent terms in Twitter users' opinions often carry negative connotations. This similarity suggests that topics associated with adverse aspects like murder, terror, shooting, and the like tend to generate negative expressions.

From this comparison, it can be deduced that subjects inclined towards negativity tend to evoke corresponding negative opinions, and conversely. A study titled "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis" by Widodo Budiharto and Meiliana supports this notion [15]. It reveals that a majority of the received opinions are positive. This occurrence can be attributed to the study's topic being positively inclined, especially among the supporters of the presidential candidates expressing their favorable viewpoints.

Volume 08, Number 01, Page 298-302
e-ISSN: 2598-120X; p-ISSN: 2598-117X

IV. CONCLUSIONS

Drawing from the research question or problem statement explored by the author, and considering the research results and discussions conducted, a conclusion can be drawn. The majority of Indonesian people, particularly Twitter users, exhibit a negative sentiment towards the Ferdy Sambo case that unfolded within the country. This negative sentiment constitutes 55% of the total, while neutral sentiment accounts for 32%, and positive sentiment amounts to 13%. Furthermore, the accuracy rate of data analysis is determined to be 97%, with a precision rate of 95% and a recall rate of 90%.

The author has several suggestions for future research endeavours. Firstly, it is recommended that the data collection period be extended for a more comprehensive analysis. A larger dataset often leads to higher levels of accuracy, precision, and recall, thus enhancing the quality of data analysis. These three factors significantly contribute to evaluating the analysis's overall quality. Additionally, the author proposes exploring diverse data analysis methodologies in subsequent studies. While sentiment analysis has been employed in this research, expanding the analytical techniques could offer a broader understanding and deeper insights, ultimately yielding more significant benefits.

The author also hopes that government agencies in Indonesia, particularly the Indonesian Legal and Police agencies, will take more assertive actions against cases involving officials and public figures. Such actions are crucial to maintain the integrity of these institutions in the face of law violations committed by individuals. Even though these cases involve only a few individuals, they have the potential to tarnish the image of the respective government agencies. In today's era, where everyone can freely express their opinions, especially on social media, such cases can lead to varied public responses. This situation emphasizes the significance of addressing such matters promptly and effectively. Failing to do so could lead to the distortion of public opinions and perceptions. It could either strengthen public trust in these agencies or erode it, depending on the direction these agencies take in response.

REFERENCES

[1] H. Fakhrurroja, M. N. Atmaja and J. N. C. Panjaitan, "Crisis Communication on Twitter: A Social Network Analysis of Christchurch Terrorist Attack in 2019," *Conference: 2019 International Conference on ICT for Smart Society (ICISS),* November 2019.

[2] D. H. Jayani, "10 Media Sosial yang Paling Sering Digunakan di Indonesia," 26 Februari 2020. [Online].

[3] M. A. Rizaty, "Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022," 10 August 2022. [Online].

[4] S. Mansour, "Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Minin," *Procedia Computer Science,* vol. 140, p. 95–103, 2018.

[5] D. L. Fajri, "5 Fungsi Manajemen Menurut Henry Fayol - Nasional Katadata.co.id," 8 March 2022. [Online].

[6] M. S. P. Hasibuan, Organisasi dan Motivasi, Sinar Grafika Offset, 1996.

[7] I. O'Reilly Media, Big Data Now: 2012 Edition, O'Reilly Media, 2012.

[8] D. Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety.," *META Group Research Note 6.,* 2001.

[9] H. Jiawei, M. Kamber and J. Pei, Data Mining: Concepts and Techniques Third Edition, Waltham, MA: Morgan Kaufmann, 2012.

[10] M. A. Hearst, "Untangling Text Data Mining," *Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics,,* June 1999.

[11] H. Milkha, Machine Learning Text Categorization: Text Mining, University of Texas, 2006.

[12] T. M. Mitchell, Machine Learning, McGraw-Hill., 1997.

[13] I. Indrawati and A. Alamsyah, "Social Network Data Analytics for Market," *Conference: 2017 5th International Conference on Information and Communication Technology (ICoIC7),* pp. 215-219, 2017.

[14] P. K. Sari, A. Alamsyah and S. Wibowo , "Measuring e-Commerce service quality from online customer," *Journal of Physics,* pp. 1-6, 2018.

[15] W. Budiharto and M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," *Journal of Big Data,* vol. 5, no. 51, 2018.

[16] S. Metode Penelitian kuantitatif, kualitatif dan R & D, Penerbit Alfabeta, Bandung, 2018.

[17] S. Ananiadou, B. Rea, N. Okazaki and R. N. Procter, "Supporting Systematic Reviews Using Text Mining," *Social Science Computer Review,* vol. 27, no. 4, pp. 509-523, October 2009.

[18] D. R. Cooper and P. S. Schindler, Business Research Method, mc-Graw Hill Irwin, 2014.

[19] A. K. Uysal and S. Gunal, "The impact of preprocessing on text classification," *Information Processing & Management,* vol. 50, no. 1, pp. 104-112, 2014.

[20] W. Budiharto and M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," *Journal of Big Data,* vol. 5, no. 51, 2018.