Jurnal Mantik, 6 (4) (2023), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Analysis of twitter user sentiment on the monkeypox virus issue using the nrc lexicon

Muhammad Khahfi Zuhanda[1*], Anil Hakim Syofra[2], Devy Mathelinea[3], Prana Ugiana Gio[4], Yuan Anisa[5], Nanda Novita[6]

[1,5,6]Faculty of Engineering, Universitas Medan Area, Indonesia
[2]Faculty of Teaching and Education, Universitas Asahan, Indonesia
[3]Department of Mathematics, Universitas Sumatera Utara, Indonesia
[4]Departement of Vocational, Universitas Sumatera Utara, Indonesia

ARTICLE INFO

Article history:

Received Jan 30, 2023
Revised Feb 17, 2023
Accepted Feb 28, 2023

Keywords:

Monkeypox
NRC Lexicon
Sentiment analysis
Text mining
RStudio

ABSTRACT

Monkeypox virus is closely related to variola (smallpox virus) and causes a smallpox-like disease. The increase in monkeypox cases has caused the general public to be involved in providing responses to seek and share information related to monkeypox on the internet, especially on social media platforms. This study aims to analyze a collection of 5000 tweets on August 5, 2022, for sentiment analysis using the NRC lexicon. Of the 5,000 tweets that have been extracted, it is obtained that the words that Twitter users often use are "health", "emergency", "public", "covid", and "declares". By using the classification using the NRC lexicon comparison, we found that the emotion type of fear was the most widely used emotion, which had a presentation of 19.73%, followed by anticipation emotion at 16.78%, sadness 14.77%, trust at 13.90%, anger 9.99%, shock 9.14%, disgust 8.12%, and happy 1288 7.90 %. The negative sentiment that often appears on Twitter is equal to 51.92%, and positive sentiment has a percentage of 48.08%. The negative sentiment words that appear most often are "emergency", "virus", "disease", "shit", and "risk". The positive sentiment words that appear most often are "public", "vaccine", "sex", "contact", and "united". The analytical method with the lexicon method is very well used in analyzing various emotions and sentiments on social media.

Corresponding Author:

Muhammad Khahfi Zuhanda,
Faculty of Engineering,
Universitas Medan Area,
Kolam Street Number 1 Medan Estate / PBSI Building Street, 20223, Medan, Indonesia
Email: khahfi@staff.uma.ac.id

1. INTRODUCTION

Monkeypox disease will be horrendous throughout 2022. The monkeypox virus causes this disease. This disease is a zoonotic disease that once attacked humans but has now reappeared. In 1970, monkeypox was detected in the Democratic Republic of the Congo (DRC) in a nine-month-old boy (Thakur, 2022). Monkeypox virus is closely related to variola (smallpox virus), which belongs to the family Poxviridae, subfamily Chordopoxvirinae, and genus Orthopoxvirus and causes a smallpox-like disease. Symptoms of the disease usually

develop within two weeks of exposure. Initial symptoms typically include fever, muscle and back pain, headache, chills, fatigue, and swollen lymph nodes (Berthet et al., 2021).

The increase in cases of monkeypox has caused the general public to be involved in providing responses to seek and share information related to monkeypox on the internet, especially on social media platforms. Twitter is a social media platform widely used by various age groups worldwide. Currently, there are more than 400 million monthly active users on Twitter. Analyzing people's tweet data on Twitter has been of great interest to the scientific community in recent years, as can be seen from this current work in which relevant tweets were mined to analyze dataset tweets on COVID-19 (Mathur et al., 2020a), Supply Chain (Zuhanda et al., 2023), European Migration Crisis (Ferra & Nguyen, 2017), Vehicle Routing (Zuhanda, Suwilo, et al., 2022, 2023), Inflammatory Bowel Disease(Kesavarapu et al., 2021), and mysterious hepatitis (Zuhanda, Desniarti, et al., 2022).

The Twitter dataset serves as a data source for various applications. It uses case scenarios related to studying conversational paradigms and investigating the underlying patterns of information-seeking and sharing behavior. For example, the Twitter dataset regarding COVID-19 has been used for topic detection and sentiment analysis (Kausar et al., 2021), classifying language-agnostic discourse of tweets (Liu et al., 2018), monitoring people's emotions (Mathur et al., 2020b), studying the dynamics of users' emotions (Duncombe, 2019), conversation analysis (Pérez-Dasilva et al., 2020), studying the evolution of public sentiment over time, (Garcia & Berton, 2021) and several other applications.

The recent monkeypox outbreak has also led to increased research and development in this area. These include – studies on outbreaks in Europe and North America (Mahase, 2022), studying increasing cases in the UK(Ismail et al., 2021), investigating methods of transmission of the virus through sexual contact (Heskin, 2022), analyzing public attitudes towards monkeypox (Dell'Amico et al., 2020), predicting the burden and duration of these outbreaks (Guagliardo et al., 2020), use of the Jynneos vaccine for vaccinating persons at risk of occupational exposure (Hooper et al., 2004), studying the incubation period (Harapan et al., 2020), and guidelines for pregnant individuals with exposure to monkeypox virus (Mbala et al., 2017).

However, none of the previous studies focused on mining Tweets about the Monkeypox outbreak for sentiment analysis using the NRC Lexicon method. This study proposes to analyze a collection of 5000 tweets on August 5, 2022, for sentiment analysis of the collected tweets. To achieve this goal, we arrange the next sequence in this research as follows. Section 2 presents the research methods used to achieve the research objective, namely analyzing the sentiments and emotions of Twitter users regarding the monkeypox outbreak. This analysis also presents word clouds, emotion-type diagrams, and sentiment diagrams of words that Twitter users frequently tweet. Section 3 of this study presents the results and discussion. Furthermore, the end of this paper presents conclusions regarding the research results obtained and is followed by references.

2. RESEARCH METHOD

2.1 Extract tweets

To be able to extract tweets from Twitter takes a few steps. First, create a Twitter application. The Twitter account will act as an interface to the Twitter API. Authenticating using RStudio requires the ROAuth package. Twitter development will provide an authentication ID code such as consumer key, consumer key secret, access token, and access token secret to extracting Twitter tweets. The redirect to the URL is done automatically by clicking on the authorized app and is verified using the OTP during the authentication process (Verma et al., 2019).

2.2 Data pre-processing

The extracted data is still in the form of raw tweets. So this tweet needs a cleaning process because the extracted tweet still contains data that is not required for analysis. It is essential to do this process to remove this unnecessary data from the extracted Twitter dataset. Things that need to be removed, such as HTML links, numbers, punctuation '@', stopwords, emoticons, RTs, changing tweets to lowercase before analyzing sentiment, so that the resulting data set only retains valuable information for analysis (Grover et al., 2020).

2.3 Lexical Analysis

Lexical analysis is an analysis that uses a lexicon-based approach. In 2006, Hui and Liu created a lexicon dictionary to parse emotional sentiments. This database contains 2,317 positive opinions and 3,338 negative opinions. Each word in the tweet is compared to the lexical database stored in the RStudio working directory, and the positive and negative sentiments are then analyzed by this method. (Zad et al., 2021).

2.4 Classification

In building the emotion classification model, tweets that have been cleaned and broken down into words will be categorized into eight emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The R package "Syuzhet" is used to classify this, and from this package, there is a lexicon NRC dictionary that we can use to organize emotions. The Hui and Lui, sentiment word database types positive and negative opinion words (Hassan et al., 2022).

2.5 Visualization

In visualizing, we use the "ggplot2" (Tyner et al., 2017), "plotly" (Plotly, 2019), and "wordcloud2" (Bashri & Kusumaningrum, 2017) packages in RStudio to present graphical bar charts and word clouds. The bar graph represents the results of the sentiment analysis that has been done, and the word cloud visualizes words that users often use to make them more attractive.

3. RESULTS AND DISCUSSIONS

In this study, we extracted 5,000 tweets from August 5, 2022. The top 5 data extracted using RStudio can be seen in Table 1. Table 1 presents the id of the tweet that made the tweet, the date and time of the tweet, the sentence of the tweet, and the device used in the tweet.

Table 1. Top 5 Tweets before data pre-processing

Screen_name	Created_at	Text	Source
Jxss_9	2022-08-05 08:13:50	It blows my mind when I bring up the monkeypox to my coworkers they don’t know wtf I’m talking about ???	Twitter for iPhone
RajalakshmiA4	2022-08-05 08:13:36	August 6th, CIDS Kerala, ID Saturday, 7pm \nJoin us for discussions on community onset infections and on 'monkeypox lessons' on	Twitter for Android

		various clinical presentation in Kerala, India and across the globe.\n@CidsIndia @jestjac @PriyaNori @SHEA_Epi https://t.co/TfWsVvuFCE	
Daviemoo	2022-08-05 08:13:34	No idea why but I'm very nervous about t'old monkeypox vaccine today. Don't know why. But it doesn't matter, I'm more arsed about making sure I don't accidentally make another viral outbreak worse- all, apparently without the help of the government. Shocker.	Twitter for iPhone
seasidebest	2022-08-05 08:13:30	@UVC_Far What are the effects of Far Uvc on MONKEYPOX???	Twitter for Android
seasidebest	2022-08-05 08:11:57	@UVRayLights What are the effects of Far Uvc on MONKEYPOX?\nThank you\nJohn	Twitter for Android

After cleaning, a word-for-word decapitation process is carried out for analysis. This process is obtained by calculating the frequency of each word that appears so that it can be presented in Table 3. The words that often appear are "health" 953 times, "emergency" 862 times, "public" 644 times," covid" 578 times, and "declares" 626 times. Table 3 presents ten words that are often used by Twitter users. Figure 1 is a word cloud of words that are often used by people who are tweeted via social media Twitter.

Table 3. Top 10 words that have the most frequency

Num.	Word	Frequency	Num.	Word	Frequency
1.	health	953	6.	people	415
2.	emergency	862	7.	gay	369
3.	public	644	8.	vaccine	354
4.	covid	578	9.	outbreak	353
5.	declares	426	10,	amp	235

To analyze the tweet emotions of community Twitter users, we use the NRC lexicon in the Syuzet package in RStudio to classify the types of emotions people use. Table 4 presents a presentation of the number of emotions that have been classified. From the classification results, it was found that the emotion type of fear was the most widely used emotion, which had a score of 3217 (19.73%), followed by anticipation 2685 (16.78%), sadness (2409) 14.77%, trust 2267 (13.90%), anger 1629 (9.99%), surprised (1490) 9.14 %, disgusted (1324) 8.12%, and happy 1288 (7.90 %). Figure 2 presents a bar chart visualization for visualizing the types of emotions classified using plotly packages.

Figure 1. Wordcloud

Table 4. Total emotional score

Emotional type	Emotional score	Percentage (%)
anger	1629	9.99
anticipation	2685	16.46
disgust	1324	8.12
fear	3217	19.73
joy	1288	7.90
sadness	2409	14.77
surprise	1490	9.14
trust	2267	13.90

Table 5 presents the NRC lexicon classification of positive and negative sentiment scores. Table 5 shows that the negative sentiment that often appears on Twitter is 4656, with a percentage of 51.92% and the positive view has a score of 4312, with a rate of 48.08%. Figure 3 is a frequency diagram of positive and negative sentiments. The negative sentiment words that appear most often are “emergency” 862 times, “virus” 211 times, “disease” 158 times, “shit” 148 times, and “risk” 134 times. The word positive sentiment that appears most often is “public” 644 times, “vaccine” 354 times, “sex” 201 times, “contact” 133 times, and “united” 81 times.

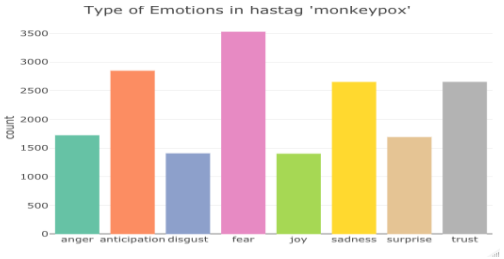

Figure 2. Bar chart of emotion scores

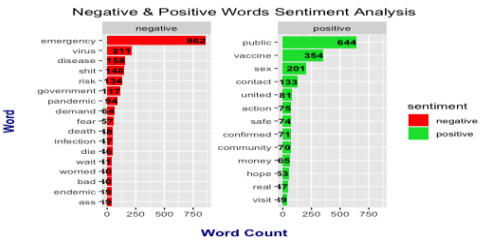

Figure 3. Word frequency diagram

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily (Grieshaber, 2020). The discussion can be made in several sub-chapters.

4. CONCLUSION

This research succeeded in finding words that are often used, emotional reactions, and public sentiment regarding the issue of the monkeypox virus. Of the 5,000 tweets that have been extracted, The result of the research found that the words frequently used by Twitter users were "health" 953 times, "emergency" 862 times, "public" 644 times, "covid" 578

times, and "declares" 626 times. By using the classification using the NRC lexicon comparison, The analysis found that the emotion type of fear was the most widely used emotion, which had a score of 3217 (19.73%), followed by anticipation 2685 (16.78%), sadness 2409 (14.77%), trust 2267 (13.90%), angry 1629 (9.99%), shocked 1490 (9.14%), disgusted 1324 (8.12%), and happy 1288 (7.90%). The negative sentiment that often appears on Twitter is 4656, with a percentage of 51.92%, and the positive view has a score of 4312, with a rate of 48.08%. The negative sentiment words that appear most often are "emergency" 862 times, "virus" 211 times, "disease" 158 times, "shit" 148 times, and "risk" 134 times. The word positive sentiment that appears most often is "public" 644 times, "vaccine" 354 times, "sex" 201 times, "contact" 133 times, and "united" 81 times. The lexicon NRC method effectively classifies the emotions and sentiments of frequently used words. However, the method still has limitations in analyzing if the user uses figurative words, such as satire, or uses the opposite word. Future research will address a bibliometric analysis of the monkeypox.

REFERENCES

Bashri, M. F. A., & Kusumaningrum, R. (2017). Sentiment analysis using Latent Dirichlet Allocation and topic polarity wordcloud visualization. *2017 5th International Conference on Information and Communication Technology, ICoIC7 2017*. https://doi.org/10.1109/ICoICT.2017.8074651

Berthet, N., Descorps-Declère, S., Besombes, C., Curaudeau, M., Nkili Meyong, A. A., Selekon, B., Labouba, I., Gonofio, E. C., Ouilibona, R. S., Simo Tchetgna, H. D., Feher, M., Fontanet, A., Kazanji, M., Manuguerra, J. C., Hassanin, A., Gessain, A., & Nakoune, E. (2021). Genomic history of human monkey pox infections in the Central African Republic between 2001 and 2018. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-92315-8

Dell'Amico, M., Montemanni, R., & Novellani, S. (2020). Matheuristic algorithms for the parallel drone scheduling traveling salesman problem. *Annals of Operations Research*, *289*(2). https://doi.org/10.1007/s10479-020-03562-3

Duncombe, C. (2019). The Politics of Twitter: Emotions and the Power of Social Media. *International Political Sociology*, *13*(4). https://doi.org/10.1093/ips/olz013

Ferra, I., & Nguyen, D. (2017). #Migrantcrisis: "tagging" the European migration crisis on Twitter. *Journal of Communication Management*, *21*(4). https://doi.org/10.1108/JCOM-02-2017-0026

Garcia, K., & Berton, L. (2021). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing*, *101*. https://doi.org/10.1016/j.asoc.2020.107057

Grover, A., Kumar, P., & Garg, K. (2020). Improvisation of Cleaning Process on Tweets for Opinion Mining. *International Journal of Big Data and Analytics in Healthcare*, *5*(1). https://doi.org/10.4018/ijbdah.2020010104

Guagliardo, S. A. J., Monroe, B., Moundjoa, C., Athanase, A., Okpu, G., Burgado, J., Townsend, M. B., Satheshkumar, P. S., Epperson, S., Doty, J. B., Reynolds, M. G., Dibongue, E., Etoundi, G. A., Mathieu, E., & McCollum, A. M. (2020). Asymptomatic orthopoxvirus circulation in humans in the wake of a monkeypox outbreak among chimpanzees in Cameroon. *American Journal of Tropical Medicine and Hygiene*, *102*(1). https://doi.org/10.4269/ajtmh.19-0467

Harapan, H., Wagner, A. L., Yufika, A., Setiawan, A. M., Anwar, S., Wahyuni, S., Asrizal, F. W., Sufri, M. R., Putra, R. P., Wijayanti, N. P., Salwiyadi, S., Maulana, R., Khusna, A., Nusrina, I., Shidiq, M., Fitriani, D., Muharrir, M., Husna, C. A., Yusri, F., … Mudatsir, M. (2020). Acceptance and willingness to pay for a hypothetical vaccine against monkeypox viral infection among frontline physicians: A cross-sectional study in Indonesia. *Vaccine*, *38*(43). https://doi.org/10.1016/j.vaccine.2020.08.034

Hassan, M. K., Hudaefi, F. A., & Caraka, R. E. (2022). Mining netizen's opinion on cryptocurrency: sentiment analysis of Twitter data. *Studies in Economics and Finance*, *39*(3), 365–385. https://doi.org/10.1108/SEF-06-2021-0237

Heskin, J. et al. (2022). Transmission of monkeypox virus through sexual contact - A novel route of infection. *Journal Infection*, *85*(3), 334–363.

Hooper, J. W., Thompson, E., Wilhelmsen, C., Zimmerman, M., Ichou, M. A., Steffen, S. E., Schmaljohn, C. S., Schmaljohn, A. L., & Jahrling, P. B. (2004). Smallpox DNA Vaccine Protects Nonhuman Primates against Lethal Monkeypox. *Journal of Virology*, *78*(9). https://doi.org/10.1128/jvi.78.9.4433-4443.2004

Ismail, S. A., Saliba, V., Lopez Bernal, J., Ramsay, M. E., & Ladhani, S. N. (2021). SARS-CoV-2 infection and transmission in educational settings: a prospective, cross-sectional analysis of infection clusters and outbreaks in England. *The Lancet Infectious Diseases*, *21*(3). https://doi.org/10.1016/S1473-3099(20)30882-3

Kausar, M. A., Soosaimanickam, A., & Nasar, M. (2021). Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak. *International Journal of Advanced Computer Science and Applications*, *12*(2). https://doi.org/10.14569/IJACSA.2021.0120252

Kesavarapu, K., Zelenkauskaite, A., & Nandi, N. (2021). Reach and Power of Physician-Initiated Tweets in a Twitter Inflammatory Bowel Disease Community. *Crohn's and Colitis 360*, *3*(3). https://doi.org/10.1093/crocol/otab052

Liu, Y., Wert, G., Greenawald, B., Boni, M. al, & Brown, D. E. (2018). Predicting violent behavior using language agnostic models. *IC3K 2018 - Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, *1*. https://doi.org/10.5220/0006933701020109

Mahase, E. (2022). Monkeypox: What do we know about the outbreaks in Europe and North America? *The BMJ*.

Mathur, A., Kubde, P., & Vaidya, S. (2020a). Emotional analysis using twitter data during pandemic situation: Covid-19. *Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020*. https://doi.org/10.1109/ICCES48766.2020.09138079

Mathur, A., Kubde, P., & Vaidya, S. (2020b). *Emotional Analysis using Twitter Data during Pandemic Situation: COVID-19*. https://doi.org/10.1109/icces48766.2020.9138079

Mbala, P. K., Huggins, J. W., Riu-Rovira, T., Ahuka, S. M., Mulembakani, P., Rimoin, A. W., Martin, J. W., & Muyembe, J. J. T. (2017). Maternal and Fetal Outcomes among Pregnant Women with Human Monkeypox Infection in the Democratic Republic of Congo. *Journal of Infectious Diseases*, *216*(7). https://doi.org/10.1093/infdis/jix260

Pérez-Dasilva, J. Á., Meso-Ayerdi, K., & Mendiguren-Galdospín, T. (2020). Fake news and coronavirus: Detecting key players and trends through analysis of twitter conversations. *Profesional de La Informacion*, *29*(3). https://doi.org/10.3145/epi.2020.may.08

Plotly. (2019). *Plotly: Modern Analytic Apps for the Enterprise - Plotly*. Plotly.

Thakur, N. (2022). MonkeyPox2022Tweets: A Large-Scale Twitter Dataset on the 2022 Monkeypox Outbreak, Findings from Analysis of Tweets, and Open Research Questions. *Infectious Disease Reports*, *14*(6), 855–883. https://doi.org/10.3390/idr14060087

Tyner, S., Briatte, F., & Hofmann, H. (2017). Network visualization with ggplot2. *R Journal*, *9*(1). https://doi.org/10.32614/rj-2017-023

Verma, P., Khanday, A. M. U. D., Rabani, S. T., Mir, M. H., & Jamwal, S. (2019). Twitter sentiment analysis on Indian government project using R. *International Journal of Recent Technology and Engineering*, *8*(3), 8338–8341. https://doi.org/10.35940/ijrte.C6612.098319

Zad, S., Jimenez, J., & Finlayson, M. (2021). *Hell Hath No Fury? Correcting Bias in the NRC Emotion Lexicon*. https://doi.org/10.18653/v1/2021.woah-1.11

Zuhanda, M. K., Anisa, Y., Desniarti, D., Hafiz, M., Syofra, A. H., Caraka, R. E., & Noh, M. (2023). Supply chain strategy during the COVID-19 terms: sentiment analysis and knowledge discovery through text mining. *Indonesian Journal of Electrical Engineering and Computer Science*, *30*(2), 1120. https://doi.org/10.11591/ijeecs.v30.i2.pp1120-1127

Zuhanda, M. K., Desniarti, & Hafiz, M. (2022). Twitter User Sentiment Analysis on the Mysterious Hepatitis Issue. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, *6*(1), 258–266. https://doi.org/10.31289/jite.v6i1.7326

Zuhanda, M. K., Suwilo, S., Sitompul, O. S., Mardiningsih, Caraka, R. E., Kim, Y., & Noh, M. (2023). *Optimization of Vehicle Routing Problem in the Context of E-commerce Logistics Distribution*.

Zuhanda, M. K., Suwilo, S., Sitompul, O. S., & Mardiningsih, M. (2022). A combination k-means clustering and 2-opt algorithm for solving the two echelon e-commerce logistic distribution. *Logforum*, *18*(2), 213–225. https://doi.org/10.17270/J.LOG.2022.734