

Sentimental Analysis of Tweets about Covid-19 and Covid-19 Vaccines

Lavanya A P¹, Sarika M²

Assistant Professor¹, PG Scholar²
Department of Computer Science and Engineering^{1,2}
Sona College of Technology^{1,2}
lavanya.cse@sonatech.ac.in¹, sarikamurali97@gmail.com²

ABSTRACT:

The COVID-19 pandemic, is otherwise called the Covid pandemic, is a continuous genuine worldwide issue everywhere on the world. The flare-up first became known in December 2019 in Wuhan, China. This was announced pandemic by the World Health Organization on eleventh March 2020. Coronavirus infection contaminated on individuals and executed a huge number of individuals in the United States, Brazil, Russia, India and a few different nations. Since this pandemic keeps on influencing a great many lives, and various nations have depended on one or the other halfway or full lockdown. A COVID 19 immunization is an antibody proposed to give gained resistance against extreme intense respiratory condition Covid 2 (SARS CoV 2), the infection causing Covid illness 2019 (COVID 19). Individuals took web-based media stages to share their feelings, need and assessments about the Covid antibody. In this exploration work, slant examination on the tweets of individuals has been led. A dataset of in excess of 50,000 tweets with hashtags like #covid-19, #COVID19, #CORONAVIRUS, #CORONA, #StayHomeStaySafe, #Covid_19, #CovidPandemic, #covid19, #Covid_vaccine, #Vaccine #Lockdown, #Quarantine, #quarantine, and so on was considered in this examination. In view of the tweets posted in English an assessment examination was performed. This exploration was led to see how individuals from various tainted nations adapt to the circumstance. The tweets were gathered, preprocessed and afterward text mining calculations utilized lastly opinion investigation have been done and given the outcomes. The motivation behind this research paper to think about the sentiments of individuals' opinion on Covid19 and Covid Vaccine.

Keywords: covid, negative, neutral, positive, sentiment, subjectivity, vaccine.

1. INTRODUCTION

Covid infection (COVID-19) first was distinguished in Wuhan, China in December 2020. The infection has spread all through the world covering each district. In 3-4 months this pestilence has upset the entire world. The world has seen numerous pandemic periods, yet this pandemic today stimulates serious financial issues on a nation scale just as at small size. People may encounter maniacal side effects because of pandemic and countries may endure monetary downturn because of individuals with voyaging limitation, and all exercises relating to financial aspects have been shut and social removing has been forced. 21 million individuals all throughout the planet were accounted for positive for COVID-19 by mid-August 2020 and almost 773,072 were dead. Now COVID-19 is increasing at very fast rate, in the name of Covid second wave. As in the month of May 2021, total of 167,210,469 cases were reported all over the world. COVID-19 has affected more than 215 countries. The top 10 countries which have been severely affected by COVID-19 as on May 2021, includes USA (33,882,449 cases), Brazil (16,047,439 cases), India (26,608,138 cases), Russia (5,001,505 cases), South Africa (1,632,571 cases), France (5,593,962 cases). Different utilization of interpersonal interaction locales, similar to Twitter, speeds up the way toward sharing data and having sees on local area occasions and wellbeing emergencies. Coronavirus has been one of Twitter's moving zones all through January 2020 and it has kept on being

discussed up until now. Since more nations have received isolate measures, individuals have progressively depended on different online media destinations to hear news and communicating their point of view.

Twitter information is valuable in uncovering public discussions and sentiments about energizing issues and genuine information on arising pandemics. In the progressing COVID-19 pandemic, a few government organizations all throughout the planet use Twitter as one of the vital methods for contact to oftentimes trade strategy updates and news identified with COVID-19 with the overall population. Expanding quantities of studies have been gathered from Twitter information since the COVID-19 episode to comprehend the overall population's responses and discussions identified with COVID19. Expanding quantities of studies have been gathered from Twitter information since the COVID-19 episode to comprehend the overall population's responses and discussions identified with COVID19. Researchers use Tweets gathered between second February and fifteenth March 2020 to follow theme demonstrating and notion examination to comprehend key subjects and sentiments around COVID-19. Specialists and people intellectually more impacted by the plagues are the destined to talk about it on informal organizations like Twitter, which have gotten critical in our everyday lives.

The Twitter messages made through Twitter are named as Tweets. These information are accessible in open space. It would thus be able to be taken as crude information essentially for the extraction of conclusions, for the investigation of client satisfaction and for various rating strategy plans and, at last, an investigation of slant has been led. Indeed, even the online buys these days occur based on individuals' assessments about various items. For its energy, publicists and purchasing groups need to invest additional time assessing the customer experience. This exploration study has been led to recognize the estimations of individuals of eleven diverse contaminated nations with COVID19 and distinguishes what feelings individuals have been sharing from various pieces of the world. The nations chose for the investigation are top 10 [1] tainted nations in addition to one more country from Gulf district, for example Oman. The nations chose for the examination are USA, Brazil, India, Russia, South Africa, Peru, Mexico, Chile, Spain, UK and Oman.

2. RELATED WORKS

Researchers investigated Twitter information for continuous projections of flu spread and other transferable flare-ups [13]. Analysts estimated the arising hazard in a flare-up of flu in 2009 by breaking down tweet watchwords and estimating the occurrence of infection progressively and the endeavors to forestall sickness [15]. All through the 2014 flare-up of the Ebola infection, Twitter clients shared significant wellbeing data from news sources with top Twitter exercises inside 24 hours of the news occasions [12]. [5] Investigated the sentiments concerning COVID-19, accordingly analyzing the sensations of different individuals about the pandemic. Therefore, the twitter API used to acquire valuable Covid tweets, and afterward examined dependent on certain, negative, and nonpartisan feelings with the assistance of AI strategies. Also, creators utilized NLTK library for pre-preparing of got tweets and the Textblob dataset has been utilized to assess the tweets, after that the energizing outcomes demonstrates good, negative, impartial sentiments all through different perceptions.

[2] Illustrate perceptions into the improvement of uneasiness feeling over the long run as COVID-19 hit the most significant levels in the United States, utilizing text based spellbinding investigation helped by suitable text representation. [7] Researchers proposed a strategy utilizing Latent Dirichlet Allocation for topic modeling to recognize subjects examined in the tweets to arrange the different themes talked about during Coronavirus pandemic. [14] Studied the emotional changes utilizing Twitter posts. [17] Sentimental analysis has a few applications, for instance in organizations, for surveys on items that permit

organizations to comprehend input from clients and web-based media audits to dissect client surveys. [16] Reveals the seven (7) long stretches of nostalgic audit of twitter. Since tweets on Twitter are a specific dislike a customary book, a few different works tackle this worry, like the work on short and brief writings. In [11], creator assessed the information with a huge amount of tweets that were taken as large information and accordingly recorded the words, sentences or entire records. Creators utilized the straight strategy to gauge tweet divisions.

3. PROPOSED WORK

The proposed work plans to accomplish a wistful examination about Coronavirus tweets and Coronavirus vaccine tweets.

It's been a difficult year of disorder, obliteration, pain, and misery, however the worldwide rollout of COVID-19 antibodies has started sensations of alleviation and recently discovered good faith for so many. The conversation of inoculation progress, openness, adequacy, and results is continuous, and it is pervading through reports and Twitter circles every single day. Accordingly, the inspiration for this undertaking is to enlarge my point of view on the condition of the worldwide pandemic by bridling the force of Twitter information. The sentimental analysis has been done to get a general information about the polarity of the tweets.

3.1 DATA COLLECTION

The proposed work utilizes a Kaggle dataset called "All COVID-19 Vaccines Tweets". The information was gathered utilizing a Python bundle called Tweepy. The proposed work was carried out with the dataset comprising of around 3000 tweets. The information acquired was investigated and the fundamental pre preparing works were done to make the information reasonable for the research work.

3.2 TEXTBLOB SENTIMENT ANALYSIS

Sentimental analysis is the way toward identifying positive or negative assumption in text. Pivoting to the supposition investigation segment of this work, we can take this instinct of a portion of the tweets being useful and a portion of the tweets being obstinate to segment the more noteworthy talk into isolated arrangements of tweets with comparative quantitative highlights. These highlights can be acquired utilizing a Python bundle called TextBlob, which gives an API to NLP assignments, for example, parts of speech tagging, noun phrase extraction, sentimental analysis, classification, prediction, summarization and more. TextBlob is one of the python libraries in Natural Language Processing.

TextBlob effectively utilized Natural Language ToolKit (NLTK) to accomplish its assignments. NLTK is a library which gives a simple admittance to a ton of lexical assets and permits clients to work with classification, clustering and numerous different assignments. TextBlob is a straightforward library which upholds complex examination and procedure on text based data. TextBlob's sentiment analysis utilizes two key measurements: polarity and subjectivity.

TextBlob returns polarity and subjectivity of a sentence. Polarity lies between [-1,1], -1 characterizes a negative feeling and 1 characterizes a good opinion. Subjectivity lies between [0,1]. Subjectivity evaluates the measure of closely-held conviction and authentic data contained in the content. The higher subjectivity implies that the content contains genuine belief as opposed to real data. While TextBlob is prepared on an enormous corpora and utilizes a genuinely dependable language model, its administration is most appropriate to perceive grammatical and semantic highlights from literary information instead of to comprehend the content. With this execution, the supposition, polarity and subjectivity about the Coronavirus tweets were found.

3.3 WORD CLOUD

Word cloud is a significant method of addressing the significant data at a glance. Word cloud in one such information perception methods. A word cloud is an assortment, or group, of words portrayed in various sizes. The greater and bolder the word shows up, the more frequently it's referenced inside a given content and the more significant it is. The word cloud is produced for the generally dataset to address the most rehashed words from the dataset. The word cloud is likewise produced dependent on the positive words, negative words and neutral words to give the general overall information.

4 RESULTS AND DISCUSSION

The sentimental analysis result obtained from the covid data set is represented in Figure 1. Positive, Negative and Neutral are the sentiments considered in this research work.

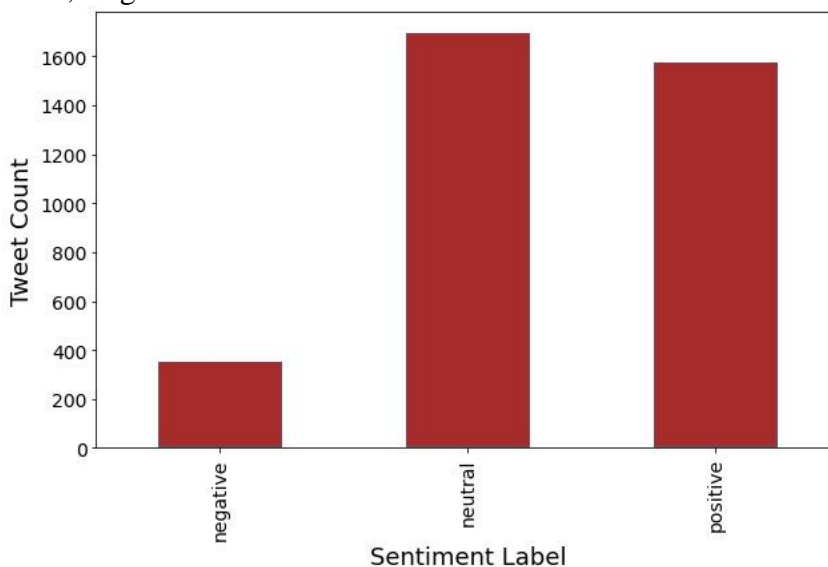


Figure 1 Sentimental Analysis

4.1 POLARITY

Polarity lies between $[-1,1]$, -1 defines a negative sentiment and 1 defines a positive sentiment.

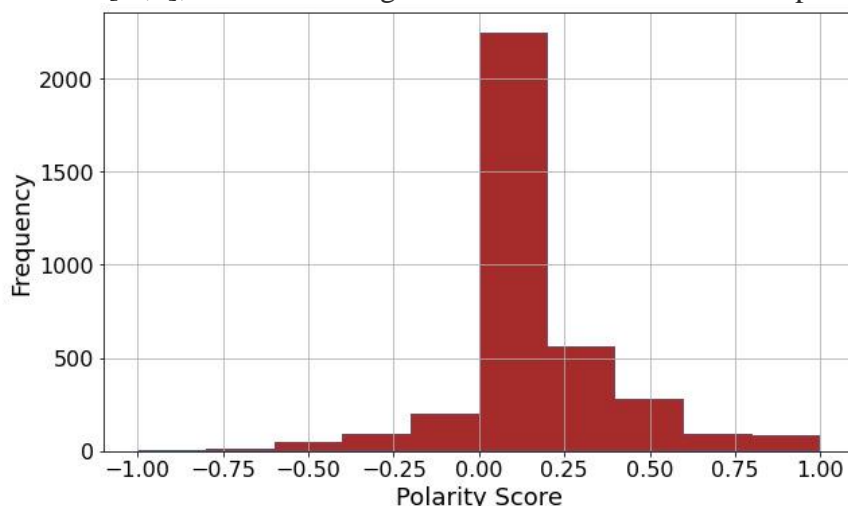


Figure 2 Polarity graph

From Figure 2, it is obvious that the greatest polarity falls between 0.00 to 0.25.



Figure 7 Word cloud for Negative words

5 CONCLUSION

The exploration work is created to contemplate the tweets posted by different clients from various pieces of the world. The outcomes acquired from this work show the opinion about the tweets. Most of the tweets fall under the 'positive' and 'neutral' assumptions. In light of the assumptions, word clouds are produced to show the most more than once utilized words in every classification separately. The outcomes give a by and large wistful examination about the Covid tweets and Covid Vaccine tweets.

REFERENCES

- [1] Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *Journal of medical Internet research*, 22(4), e19016.
- [2] Samuel, J., Ali, G. G., Rahman, M. M., Esawi, E., & Samuel, Y. (2020). COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. *Information*, 11(6), 314. doi:10.3390/info11060314.
- [3] Rosenberg, H., Syed, S., & Rezaie, S. (2020). The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Canadian Journal of Emergency Medicine*, 1-4.
- [4] Rufai, S. R., & Bunce, C. (2020). World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. *Journal of Public Health*.
- [5] C.Kaur and A. Sharma, "Twitter Sentiment Analysis on Coronavirus using Textblob," *EasyChair* 2516 -2314, 2020.
- [6] Budhwani H, Sun R. Creating COVID-19 Stigma by Referencing the Novel Coronavirus as the "Chinese virus" on Twitter: Quantitative Analysis of Social Media Data. *Journal of Medical Internet Research* 2020;22(5):e19301.
- [7] R. J. Medford, S. N. Saleh, A. Sumarsono, T. M. Perl, and C. U. J. m. Lehmann, "An" Infodemic": Leveraging High Volume Twitter Data to Understand Public Sentiment for the COVID 19 Outbreak," 2020.
- [8] Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health and Surveillance* 2020;6(2):e19273. doi:10.2196/19273.
- [9] Jagdale, Rajkumar S., Vishal S. Shirsat, and Sachin N. Deshmukh. "Sentiment analysis on product reviews using machine learning techniques." In *Cognitive Informatics and Soft Computing*, pp. 639-647. Springer, Singapore, 2019.

- [10] Uma Ramya V, ThirupathiRao K. Sentiment Analysis of Movie Review using Machine Learning Techniques. *International Journal of Engineering & Technology*. 2018;7(2.7):676.
- [11] Sulthana, A. R., Jaithunbi, A. K., & Ramesh, L. S. (2018). Sentiment analysis in twitter data using data analytic techniques for predictive modelling. *Journal of Physics: Conference Series*, 1000, 012130. doi:10.1088/1742-6596/1000/1/012130.
- [12] Padma T., Balasubramanie P., Knowledge based decision support system to assist work-related risk analysis in musculoskeletal disorder, *Knowledge-Based Systems*, Vol.22(1) PP: 72-78 DOI: 10.1016/j.knosys.2008.07.001,2009
- [13] Househ M. Communicating Ebola through social media and electronic news media outlets: A cross-sectional study. *Health Informatics J* 2016; 22:470–478.
- [14] Chorianoopoulos K, Talvis K. Flutrack.org: Open-source and linked data for epidemiology. *Health Informatics J* 2016; 22:962–974.
- [15] De Choudhury, M.; Counts, S.; Horvitz, E. Predicting Postpartum Changes in Emotion and Behavior via Social Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paris, France, 27 April–2 May 2013; pp. 3267–3276.
- [16] Signorini A, Segre AM, Polgreen PM. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. *PLoS ONE* 2011; 6:e19467.
- [17] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media*. Association for Computational Linguistics, 2011, June.
- [18] Wilson, T., Wiebe, J., & Hoffmann, P. Recognizing contextual polarity in phraselevel sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005, October.