

Pfizer Vaccine Tweet Sentiment Analysis

¹Gaurav Bhati, ²Ashok Kumar Meghvanshi, ³Jyothi Kiran Reddy Ragi, ⁴Parshant Kumar, ⁵Ashish Shrivastava

¹bhatisgaurav@gmail.com

^{1,2,3,4}UG Student, ⁵Assistant Professor,

^{1,2,3,4,5}School of Computer Science and Engineering ,Lovely Professional University, Phagwara, India.

Abstract : From the WHO pandemic declaration and government-initiated actions against coronavirus illness (COVID-19), sentiments close COVID-19 have evolved speedily.

This article provides associate degree analysis on however individuals react to an outbreak, what quantity they're tuned in to the illness and its symptoms, what preventive measures they're taking and whether or not individuals are following government's tips and guidelines etc. Understanding the posts on social media pages throughout an outbreak permits health agencies and volunteers to higher assess and perceive the public's insouciances, sentiments and desires so as to deliver applicable and effective data.

IndexTerms – Corona Virus , Pfizer, Sentiment Analysis.

I. Objective: - This study aimed to examine worldwide trends of four emotions—fear, anger, sadness, and joy—and the narratives underlying those emotions during the COVID-19 pandemic, and the effectiveness of the Pfizer vaccine.

II. Methods: - Almost a million social media tweet posts made during the early phases of the COVID-19 outbreak and were collected using various kinds of keywords like “Wuhan,” “corona,” “nCov”, ‘vaccine’, ‘Pfizer’ and “covid” as search keywords.

III. INTRODUCTION: -

The coronavirus malady (COVID-19) pandemic has infected people in further than two hundred countries and resulted in several deaths. After the WHO's pandemic declaration and government-initiated actions against the malady, sentiments concerning COVID-19 are stressful. About 90% of today's information has been provided throughout the last 2 years and obtaining insight into this huge scale information isn't trivial [1,2]. There are some works that are done on Facebook [3-7] sentimental analysis however during this paper we tend to approach targeting the Twitter sentimental analysis. within the past decade, social media analytic tools are used to watch public sentiments and communication patterns of public health emergencies just like the VHF and Zika epidemics. Some works are used on metaphysics to grasp the text [8]. Although several studies have investigated general sentiment valences and discourse topics, specific emotions are found to be a lot of closely joined to psychological processes and behaviors than the general positive and negative valences. Therefore, we tend to postulate that distinct emotions rising from social media and their underlying narratives are extremely relevant to this COVID-19 crisis and may offer unjust insights into the potency of public health electronic communication.

Since the user could be a social media platform and informal language used by the user which tends to be completely different from the traditional text and a few works that address this issue just like the work for brief informal texts [9] and there is the varied quantity of applications of sentiment analysis news [10].The use of short-form is going to be used a lot oftentimes on Twitter so as to attenuate the characters used. as a result of Twitter has placed a limit on its characters up to one hundred forty[11]. F o r e x a m p l e, ‘asap’ refers to be as soon as possible. notably, we tend to target four emotions: worry, anger, sadness, and joy. in keeping with Plutchik's Wheel of Emotions, fear-anger and sadness-joy are the fundamental feeling pairs of opposite experiences. worry is an unpleasant feeling usually arising from danger or uncertainties caused by circumstances, whereas anger results from uncertainties caused by others. unhappiness could be a negative feeling full-fledged usually when unpleasant circumstances that are totally out of one's management, and joy could be a positive feeling when pleasant events are appraised as bound and in check. work the evolution of those four basic emotions will demonstrate the ever-changing dynamics of the public's expertise to the crisis.

IV. Literature Review: -

Within 20 years, there have emerged 3 extremely infective and deadly human coronaviruses, particularly SARS-CoV, MERS-CoV, and SARS-CoV-2. The economic burden and health threats caused by these coronaviruses are very dreadful and obtaining a lot of serious health issues because of the increasing range of worldwide infections and attributed deaths of SARS-CoV-2 and MERS-CoV. sadly, specific medical countermeasures for these hCoVs stay absent. Moreover, the quick unfold of information concerning the continuing SARS-CoV-2 pandemic unambiguously places the virus aboard associate annoying infodemic and causes supererogatory worldwide panic. Lessons learned from SARS-CoV and MERS-CoV, timely updated data of SARS-CoV-2 and MERS-CoV, and summarized specific information of those hCoVs are very priceless for effectively and expeditiously contain the happening of SARS-CoV-2 and MERS-CoV. By gaining a deeper understanding of hCoVs and also the sicknesses caused by them, we are able to bridge information gaps, offer cultural weapons for fighting and controlling the unfold of MERS-CoV and SARS-CoV-2, and prepare effective and sturdy defense lines against hCoVs that will emerge or reemerge within the future. to the current finish, the progressive information and examination of the biological options of those deadly hCoVs and also the clinical characteristics of sicknesses caused by them are consistently summarized in Traits of SARS-CoV, MERS-CoV, and SARS-CoV-2.

Animal hosts: -

The potential animal hosts of SARS-CoV, MERS-CoV, and SARS-CoV-2 were summarized. Notably, the outbreaks of these hCoVs square measure associated with interactions between humans and animals, especially, each SARS-CoV and SARS-CoV-2 emerge from wet markets in China. Considering early severe acute respiratory syndrome patients were related to wild animal markets in the province, SARS-CoV was thought to emerge from wild animals (included palm civets) that were sold in these markets. afterward, a strain of CoV shared extremely homologous similarity to SARS-CoV (99.8%) was isolated from palm civets from wild animal markets, therefore palm civets-derived CoVs were believed to be able to switch their hosts to human, inflicting the human-to-human transmission. to boot, the development that some severe acute respiratory syndrome patients (3/4) had had a transparent contact history with palm civets throughout the discontinuous occurred in the province was noticed, thereby epidemiologically supporting the previous assumption that palm civets act as vital animal hosts of SARS-CoV. Thus, it's cheap to believe that palm civets were vital intermediate hosts for SARS-CoV supported this robust proof. afterward, students found that palm civets on farms square measure for the most part free from SARS-CoV infection whereas about eightieth of the civets sold in AN animal market was serologically positive (high level of SARS-CoV antibody), indicating that no widespread infection in wild civets. Thus, palm civets became the intermediate hosts of SARS-CoV in all probability by obtaining infected throughout trade activities in or before reaching these wet markets [27, 28]. Afterward, a strain of SARS-like- CoV was isolated from Chinese horseshoe nuts, sharing 88–92% genomic identity to CoVs from humans or viverrine mammal cats, powerfully indicating that nuts square measure natural. hosts for SARS-CoV. MERS-CoV is additionally believed to originate from nuts.

The ribonucleic acid fragment obtained by PCR amplification of supermolecule isolated from bat stool shared 100% ester identity with MERS-CoV from AN infected case living within the same space, indicating nuts were in all probability the supply for MERS-CoV. Then, a bat-CoV was incontestable to carry a detailed biological process relationship with MERS-CoV.

afterward, the flexibility of replicating in nuts while not generating symptoms of MERS patients was incontestable, suggesting that nuts were ideal reservoirs for MERS-CoV. The intermediate reservoir role of Arabian camel camels for MERS-CoV was supported by lush proof. 2 medical specialty studies illustrated that MERS-CoV was current in Arabian camel camels and indicated potential cross-infection between them and humans. The high genomic identity of MERS-CoV isolated from Arabian camel camels and humans was afterward documented (99.2–99.5%). Afterward, the reservoir risk and natural host-to-human transmission role of Arabian camel camels were confirmed by many serologic studies. The origin of SARS-CoV-2 is additionally refined. the same as SARS-CoV, the emergence of SARS-CoV-2 Was thought of to be related to trade activities in a very wet market in the city. Researchers found that SARS-CoV-2 and BatCoV RaTG13 (a bat-CoV) were genetically similar and assumed that nuts can be the natural reservoirs for SARS-CoV-2. Another study incontestable that the similarity in ordination between SARS-CoV-2 and therefore the CoV isolated from scaly anteater (pangolin-CoV) was high however below that between SARS-CoV-2 and RaTG13 (91.02% vs. 96.2%) [38]. These findings were echoed by 2 different analyses, during which their genomic similarity square measure ninety.03 and 92.4%, severally. students conjointly analyzed the biological process relationships among these CoVs, noticing that RaTG13 and SARS-CoV-2 Were sorted along, and pangolin-CoV was their nearest common antecedent. Taken along, Zhang et al. prompt that pangolin-CoV is another closely connected kin of SARS-CoV-2, and pangolins instead of nuts can be the natural reservoirs for SARS-CoV-2 and RaTG13 [38]. though RaTG13 and SARS- CoV-2 share the best similarity concerning the general genomic sequence, SARS-CoV-2 exhibits the best sequence similarity (97.4%) to pangolin-CoV in terms of the receptor-binding domain (RBD), however, RBD sequence similarity between RaTG13 and SARS-CoV-2 is way less (89.2%). Additionally, notably, six key RBD residues of SARS-CoV-2 and pangolin-CoV square measure fully identical whereas just one organic compound is that the same between RaTG13 and SARS-CoV-2 concerning these six residues. These findings rendered the problem that that one is that the natural reservoir remains polemic, nevertheless, it's in agreement that there exist different animals acting as intermediate hosts. A study speculated snakes square measure in all probability the intermediate hosts as a result of the same similar sequence usage bias was found among SARS-CoV-2, a bat-derived SARS-like-CoV, and snakes.

V. EXPERIMENTAL ANALYSIS: -

Python is used for implementing various classification algorithms. The Twitter data is collected by creating an app in Twitter that interacts with the Twitter API and tweets are downloaded using words that match the search keywords. while particular tweets are searched on Coronavirus. From the Twitter API, confidential keys such as a consumer key, a consumer secret, an access token, and an access token secret are generated. These keys help to provide user authentication towards accessing Twitter API. The default permissions from a Twitter account are read-only. Using these keys, the respective keywords such as Corona Virus, Pfizer, a vaccine from Twitter are filtered in the particular location and languages. The results of the corona virus cure using the vaccine are generated using sentiment analysis and machine learning of the collected tweets.

VI. DESIGN AND IMPLEMENTATION: -

This research paper is based upon the twitter sentiment analysis by using the API's which is provided by Twitter.

Following is the approach which we used: -

1. Download testing data set as input in the program.
2. Cleaning the tweets by removing stop key words.
3. Tokenize each word in the dataset.

4.Methodology.

VII. IMPLEMENTATION: -

In this paper, we are using python for implementation of sentiment analysis. We had used several packages like tweepy, textblob etc. We can install this library by following: -

- . pip install tweepy.
- . pip install textblob.

The next step is downloading the dictionary by python -m textblob.download_corpora.of

It is used for text processing and uses NLTK and corpora is a large and structured set texts which we need for analyzing tweets.

VIII. METHODOLOGY: -

IX. Connection Twitter API: -

The dataset was a real-time stream of tweets collected using Twitter API .to get access to Twitter API we have applied for credentials through the Twitter developer account. Using these credentials, we have connected to Twitter API and collected tweets using the trigger word Pfizer. This credentials are shown in Fig.1.

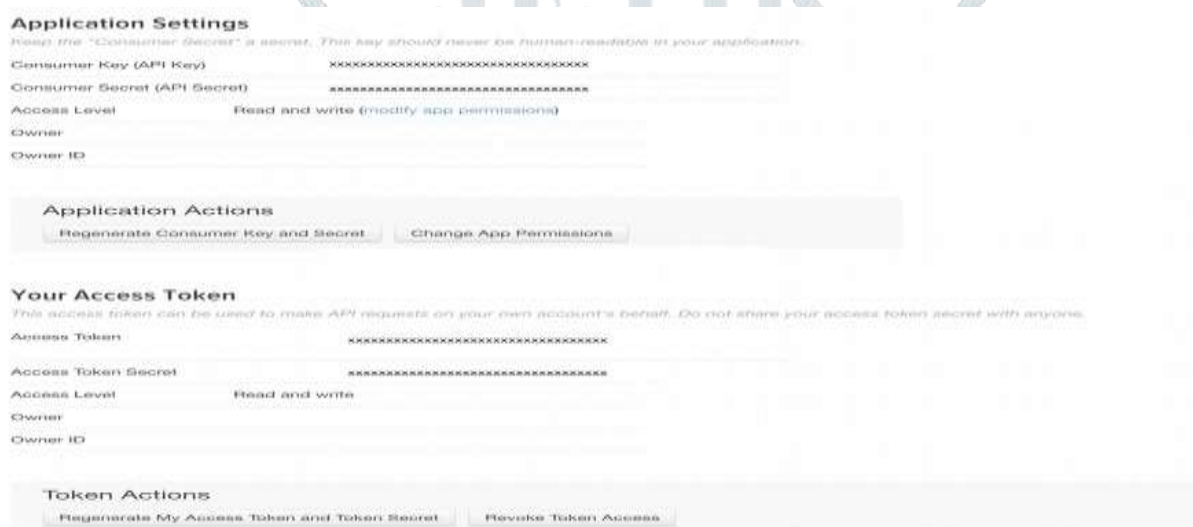


Fig [1] :- Login credentials for accessing data.

X. Data Cleaning: -

The collected tweets were then cleaned using python's regex by which we have removed contraction words, links, stop words and then used the emoji library to demojize the emojis in the tweets.

XI. Analyzing the Tweets: -

The cleaned tweets were then processed using the text blob library which produced the output of polarity of tweets, on the basis of this polarity the tweets were classified into three categories namely positive, neutral and negative.

XII. Results: -

Public emotions shifted powerfully from worry to anger over the course of the pandemic, whereas unhappiness and joy conjointly surfaced. Findings from word clouds counsel that fears around shortages of COVID-19 tests and medical provides became the progressively widespread purpose of dialogue.

Anger shifted from social phobia at the start of the pandemic to discourse around the stay-at-home notices. unhappiness was highlighted by the topics of losing friends and members of the family, whereas topics associated with joy enclosed words of feeling and healthiness. Below is the pie chart showing the result of the project(Fig.[2]) and in the next figure we had also add a word cloud (Fig.[3]).In addition we had also add graph for the frequency of words(Fig.[4]).

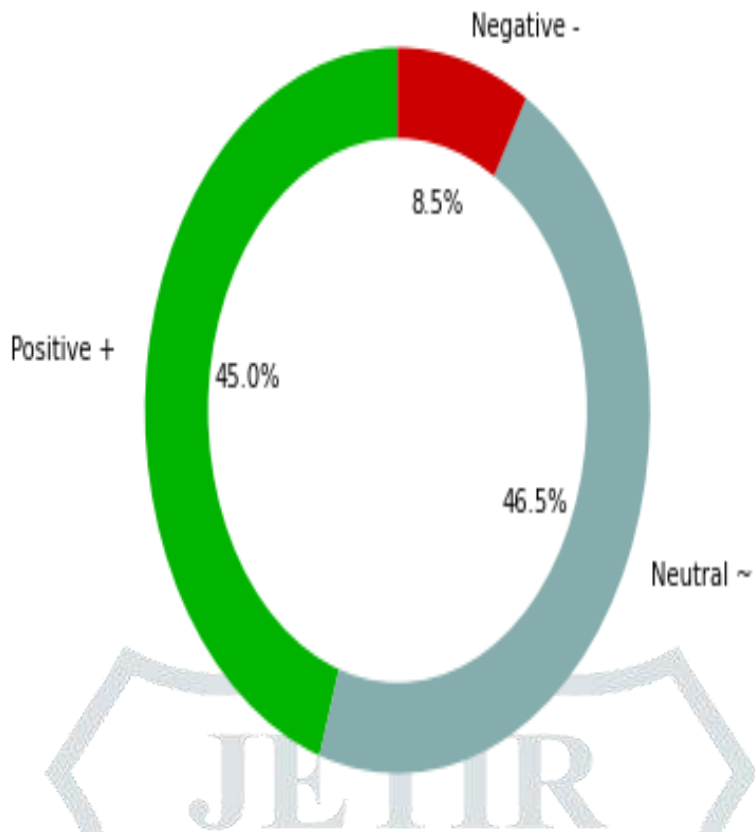


Fig [2]. Graphical Representation of Sentiment Analysis.



Fig [3]. Pfizer Word Cloud.

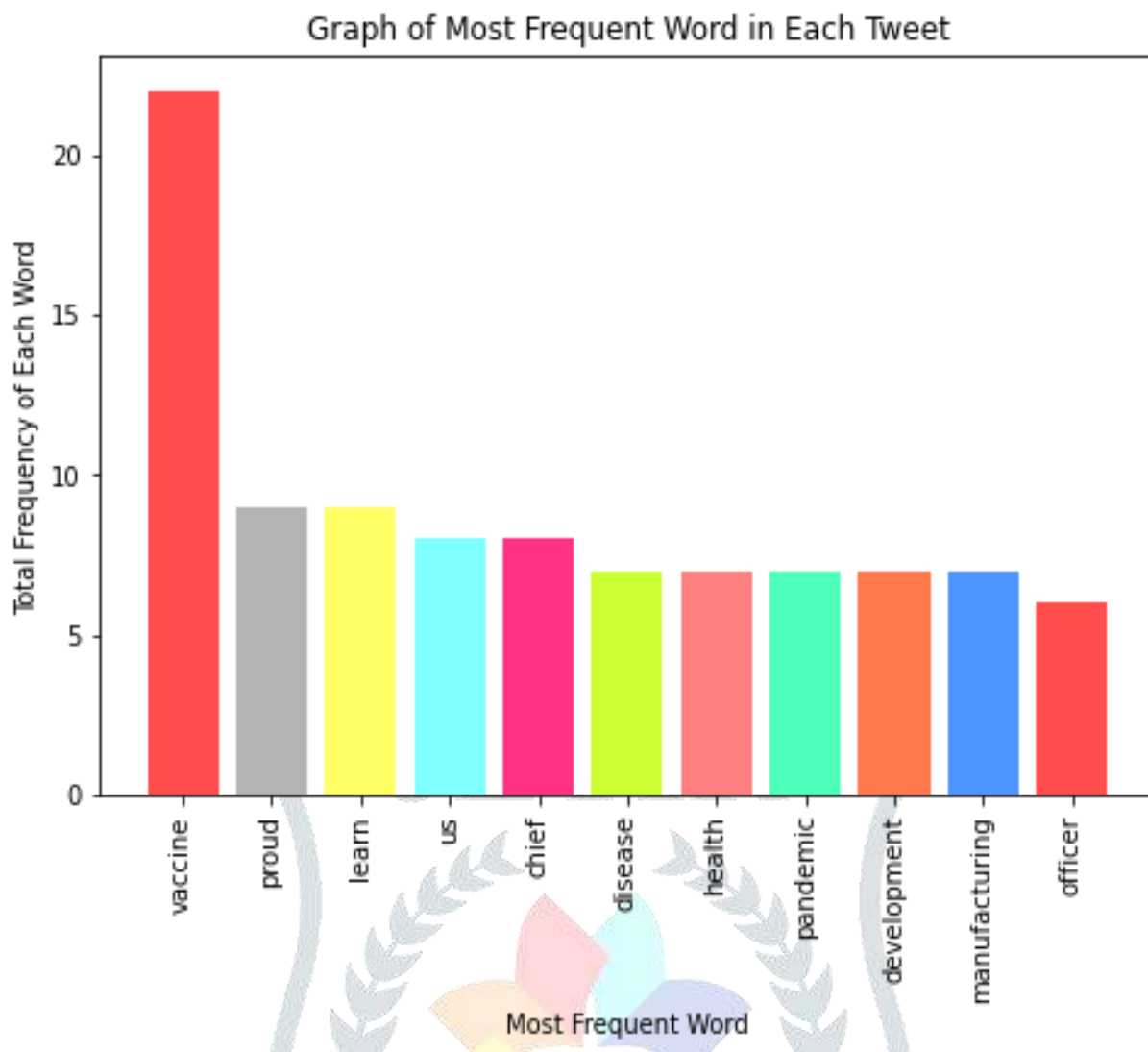


Fig [4]. Graph of Most Frequent Word in Each Tweet.

XIII. *Conclusion:* -

Overall, COVID-19 sentiments have shown fast evolutions inside the span of some weeks. Findings recommend that emotion-driven collective problems around shared public distress experiences of the COVID-19 pandemic area unit developing and embrace large-scale social isolation and also the loss of human lives. The steady rise of social group issues indicated by negative emotions must be monitored and controlled by complementing regular crisis communication with strategic public health communication that aims to balance public psychological eudaemonia. This works can continue to help to understands the perception of individuals concerning the coronavirus vaccine's impact on the public. In this analysis paper, we've got older the whole analysis of tweets concerning Covid-19 and that we had older the number of tweets and target the audience and got the wonderful results. In this analysis paper, we tend to display the results on commonplace and our model is ready to predict numerous sentiments of individuals. we've got used numerous sorts of models for coaching purposes. The results have shown that a good amount of sentiment is positive. Such solutions can play important role in the effectiveness of the vaccines.

XIV. *Future Work:* -

While the current research paper is able to analyze the tweets but, in the future, we can add some more features to it like Emoji, video, images, gifs, etc. As the data is increasing day to day so that this research paper acts as a foundation to analyze larger volumes of new data which is expected to help to build different kinds of models for the betterment of the recovery process.

XV. *REFERENCES:* -

[1] Bagheri, Hamid, and Abdusalam Abdullah Shaltoolki. "Big Data: challenges, opportunities and Cloud based solutions." *International Journal of Electrical and Computer Engineering* 5.2 (2015): 340.

- [2] Bagheri, Hamid, Mohammad Ali Torkamani, and Zhaleh Ghaffari. "Multi-Agent Approach for facing challenges in Ultra-Large-Scale systems." *International Journal of Electrical and Computer Engineering* 4.2 (2014): 151.
- [3] Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, 31, 527-541.
- [4] Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82-89.
- [5] Dasgupta, S. S., Natarajan, S., Kaipa, K. K., Bhattacharjee, S. K., & Viswanathan, A. (2015, October). Sentiment analysis of Facebook data using Hadoop based open-source technologies. In *Data Science and Advanced Analytics (DSAA), 2015*. 36678 2015. IEEE International Conference on (pp. 1-3). IEEE.
- [6] Trinh, S., Nguyen, L., Vo, M., & Do, P. (2016). Lexicon-based sentiment analysis of Facebook comments in Vietnamese language. In *Recent developments in intelligent information and database systems* (pp. 263-276). Springer International Publishing.
- [7] Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, 17, 26-32.
- [8] Boguslavsky, I. (2017). Semantic Descriptions for a Text Understanding System. In *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue"* (2017) (pp. 14-28).
- [9] Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50, 723-762.
- [10] Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., ... & Belyaeva, J. (2013). Sentiment analysis in the news. arXiv preprint arXiv:1309.6202.
- [11] D. Boyd, S. Golder, & G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter," *System Sciences (HICSS)*, 2010. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=542831. Introduction.

