

# **A Survey on Analyzing COVID-19 Vaccines on Twitter Dataset Using Tweepy and Text Blob**

**Abhishek Akshay Chaudhri<sup>1</sup>, S.S. Saranya<sup>2</sup>, Sparsh Dubey<sup>3</sup>**

<sup>1</sup>Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur Campus, Tamil Nadu, India. E-mail: aa8937@srmist.edu.in

<sup>2</sup>Associate Professor, Department of Computer Science and Engineering, School of Computing, SRM Institute of Science and Technology, Kattankulathur Campus, Tamil Nadu, India. E-mail: saranyas6@srmist.edu.in

<sup>3</sup>Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur Campus, Tamil Nadu, India. E-mail: ss3742@srmist.edu.in

## **ABSTRACT**

This paper depends on utilizing twitter for sentiment analysis of the views of people on COVID-19 vaccines and eagerness to have the jab. An ever increasing number of individuals have begun posting on the web about whether they are in favor of these vaccines or not. Many mediums are there which give these analysis among which one such medium is Twitter which has gotten very mainstream as of late. Twitter is a famous microblogging website where clients are permitted a constraint of 280 characters; this sort of limitation causes the clients to be compact just as expressive simultaneously. Thus, it turns into a rich hotspot for opinion examination and belief mining. In this examination, we will investigate the eagerness of having different vaccine by people and their recommendation to others after having the same. In the wake of applying different models and machine learning algorithms to tweets information, we have discovered that it is for sure conceivable to anticipate the vaccine is being favored or not on a large scale.

## **KEYWORDS**

Sentimental Analysis, Tweepy, Textblob, Python.

## **Introduction**

Sentiment analysis can be defined as detailed study of NLP which includes the computational studying of opinions, sentiments and feelings expressed in textual format. The idea with which we initialized this research was to be able to tell the average polarity of the views of individuals regarding getting the vaccine shots getting released by various medical organisations during this difficult time of COVID-19, so that they can have a better view and mindset of it's pros and cons and related side effects. By this research, people will be able to make a clear mindset and decision on, if they want to have vaccine or not. Sentimental Analysis or Opinion Mining focuses at learning people's views, attitudes and feelings towards an entity. The entity can represent individual, events or issues. So the best place for people to approach and publish about anything and how they feel of that product is one and only Twitter platform. It can provide us the attitude, emotions and assessments. An enormous amount of researches have been performed in the field of sentimental analysis. But majority of them are aligned on classification of formal and bigger pieces of textual data like reviews. With the mass popularity of social network platforms and microblogging websites and a large amount of data accessible from these resources, research projects on sentimental analysis have witnessed a gradual domain shift. Twitter is the best source for analysing the data of all types because:

- The Application Program Interface is neat and comes with rich developing tools.
- The data it provides is full of informations and has a data-structure that is perfect for analysis.
- Twitter data is easily and freely available to all those with proper usage rights.

## **Literature Survey**

Sentiment Analysis is often done to analyse the nature of any comment or statements made by people acting or commenting to it. Today Sentimental Analysis has achieved the place where it can be used to find the nature of statement either positive or negative along with handling various topics also. There have been an intensive amount of works recently within the field of "Sentiment analysis" by numerous researchers. In the initial phases, it was capable of only performing classifications that were binary in nature i.e. classifying the comments wither positive or negative. There have been researches on various aspects of Covid-19 recently. There has been an enormous amount of efforts being put in this study by programmers with the help of various programming techniques. For sentimentally

analysing the factors, usually technique called fuzzy logic is used. There are algorithmic procedures that help constructing the fuzzy-domain sentimental ontology tree thus, supporting the reviews that have the extractions of sentiment word or sentences, unique feature of the items and relations amongst features and thus accurately predicts the polarities of the various reviews. After designing the functions for the method formulation and standardisation of the procedure of evaluation of the strength of the person's views in the presence of a modifier called as adverbial on the social platforms. In this the data has been taken from Twitter using the Tweepy API for analysing the corresponding data by the utilization of sentimental analysis. The data has to be cleaned before any preprocessing technique can be used on the same. The topics are often used in several contexts like a tweet or a retweet or a hashtag for instance. In other similar researches tweets have been carried out to examine the preference of people over a certain topics.

## Proposed System

First of all, the Twitter is connected through Tweepy API, a python internal library. It library crawls the necessary data from the twitter and stores locally. Before the data is stored, it is divided into positive and negative polarity ones, using python library called Textblob.

## Related Work

Shulong Tan[1] proposed the idea "to move one step further on different aspects of sentiment analysis to interpret sentiment variations. He observed that, emerging topics within the sentiment variation periods are highly related to the genuine reasons behind the variations. He proposed a Latent Dirichlet Allocation (LDA) model, Foreground and Background LDA model, which filters longstanding background topics". "To enhance the readability of the mined reasons, they selected the most representative tweets for foreground topics and develop another generative model called Reason Candidate and Background LDA which ranks them with respect to their popularity."

Xiaohui Yu[2] worked on predicting sales performance based on opinion mining. As posting reviews online is one of the easiest things to do as a consumer, they do a "case study in the movie domain, to understand and solve the problem of mining reviews for predicting product sales performance. For the sentiment factor, they proposed, Sentiment PLSA in which a review is converted into a document generated by number of hidden sentiment factors. Then, based on S-PLSFA, they proposed ARSA, an Autoregressive Sentiment-Aware model for sales prediction."

Xiaolong Wang[3] worked on "topic sentiment analysis in Twitter, which was a graph based hashtag sentiment classification approach". They said that hashtags are important factors, and also helps in better understanding of opinions. They said that hashtags should be considered as complete words in finding out polarity of the sentiment. They used SVM classifier to generate the hashtag sentiment; it was a language dependent model for twitter dataset.

Rincy Jose[4] talked about the prediction of election results using opinion mining on the internet. They selected various classifiers which were "SentiWordNet classifier, Naive Bayes classifier and Hidden Markov model classifier which proved to be more accurate". So after sentiment classification on various tweets they came to the conclusion that their ensemble approach was able to produce an accuracy of around 71.48% which was more than all the three classifiers put alone.

Prakruthi V[5] did "the real time sentiment analysis of posts. They evaluated the people's ideas and thinking and feelings regarding a person, trend, product or brand. The outcome was visualized using histogram and pie chart", and outcome represented in the form of neutral, negative and positive remarks.

Efthymios Kouloumpis[6] worked on the "investigation of the utility of linguistic features for detecting the sentiment of Twitter messages", thus helping in creating more accurate analysis of sentiments. They said that hashtags and emoticons contribute largely in the accuracy of the datasets. They focused on n-gram feature, lexicon feature to be able to work on datasets.

Fuji Ren, Ye Wu[7] proposed to predict the unknown user topic opinions which is a lexicon based learning algorithm and is also a language dependent. They tried to mathematically include topical and social context into the existing prediction model. They understood the correlation between social and topical context and also used TCS to measure the same. The results demonstrated that the proposed ScTcMF framework was indeed an improved approach than the

existing one.

Dhanush M[8] worked on sentiment analysis of a topic using tweepy, their main reason of this work is to “assess a point and furthermore compute what number of individuals have a positive or negative view, in the context of utilizing estimation examination taking surveys from twitter.” They used methodologies like, NLP, Doc2Vec which is a more mature version of Word2Vec, Naïve Bayes, and Logistics regression.

Ali Hasan[9] worked on “Machine Learning based Sentiment Analysis for twitter accounts”, they state that, “they have used hybrid approach that involves sentiment analyser which includes ML techniques, also providing comparison of various techniques of sentiment analysis such as Naïve Bayes, and SVM.”

## References

- [1] Shulong, T., Yang, L., Huan, S., Ziyu, G., Xifeng, Y., Jiajun, B., & Xiaofei, H. (2014). Interpreting the Public Sentiment Variations on Twitter. *IEEE Transactions on Knowledge and Data Engineering*, 26(5), 1158–1170. <https://doi.org/10.1109/tkde.2013.116>
- [2] Yu, X., Liu, Y., Huang, X., & An, A. (2012). Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain. *IEEE Transactions on Knowledge and Data Engineering*, 24(4), 720–734. <https://doi.org/10.1109/tkde.2010.269>
- [3] Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. *In Proceedings of the 20th ACM international conference on Information and knowledge management*, 1031-1040. <https://doi.org/10.1145/2063576.2063726>
- [4] Jose, R., & Chooralil, V. S. (2016). Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach. *In international conference on data mining and advanced computing (SAPIENCE)*, 64-67. <https://doi.org/10.1109/sapience.2016.7684133>
- [5] Prakruthi, V., Sindhu, D., & Kumar, S.A. (2018). Real time sentiment analysis of Twitter posts. *In 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, 29-34. <https://doi.org/10.1109/csitss.2018.8768774>
- [6] Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg!. *In Proceedings of the International AAAI Conference on Web and Social Media*, 5(1).
- [7] Ren, F., & Wu, Y. (2013). Predicting user-topic opinions in twitter with social and topical context. *IEEE Transactions on affective computing*, 4(4), 412-424.
- [8] Dhanush, M., Ijaz Nizami, S., Patra, A., Biswas, P., & Immadi, G. (2018). Sentiment analysis of a topic on twitter using tweepy. *International Research Journal of Engineering and Technology*, 5(5), 2881.
- [9] Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. *Mathematical and Computational Applications*, 23(1), 11.