

# Analyzing COVID-19 Vaccination

POULAMI GHOSH<sup>1</sup>, SAYANI GHOSH<sup>2</sup>

<sup>1</sup>Dept. of CA, University of Engineering Management, Kolkata, INDIA

<sup>2</sup>Dept. of BCA, University of Engineering Management, Kolkata, INDIA

## Abstract

The outbreak of the coronavirus has resulted in unprecedented action, which has led authorities to decide to begin the blockade of the areas most hit by the infectious disease. Social media has been an important support for people during this difficult time. On November 9, 2020, when the first vaccine with an infection rate of 90% or higher was announced, social media responded with , and people around the world began to express the feelings of vaccination. It was no longer a hypothesis, but closer to ,every day to become a reality Therefore, it becomes imperative to verify some of the information posted on social media during the pandemic situation, specially related to Covid vaccines. To this end, it is necessary to correctly identify fact-checkable posts, so that their information content can be verified.In this work, we have addressed the problem to identify 3 types of classification on the Twitter microblogging site. We organized a shared task in the FIRE 2021 conference to study the problem of identifyefficient classifier for prediction tweets posted during a particular pandemic scenario (the Covid 19). This paper describes the dataset used in the shared task, and compares the performance of different classification that are provax, antivax and last neutral for identifying effective tweets related to Covid vaccines.We experimented with a classification-based approach. Our experiment shows that SVM classification performs well in order to effective posts.Using this support vector machine in order to solve the antivax, provax,neutral classification of twets .We're going to do this because vaccination is an important step for Covid19 so people can easily fix the news about the vaccine and grab their own slot.

## Keywords

Corona, Twitter, Microblogs, vaccine, classification, classifier

## 1. Introduction

The use of social media is escalating worldwide as lockdown is in place in some parts of the world and social distancing in other parts of the world . To exchange ideas and information regarding a series of aspects that occurred during this period. People also seem to be relying on information posted on social media. As a result, the social media platform is gaining increasing attention as it is a moderator channel between each individual and another of the world, and has become one of the fastest growing information systems for social applications. In this channel, individuals express different views, opinions, and feelings during various events triggered by the coronavirus pandemic.The coronavirus outbreak caused by the novel coronavirus SARSCoV2 has resulted in a series of changes in many aspects of the economic and social lives of many people. Since its onset, the coronavirus pandemic has continued to monopolize different regions of the world, reaching 220 countries and territories on December 9, 2020[1]. Among several

---

Forum for Information Retrieval Evaluation, December 13-17, 2021, India

✉ poulami.ghosh@uem.edu.in (P. GHOSH); sayanighosh37@gmail.com (S. GHOSH)



© 2021 Forum for Information Retrieval Evaluation, December 13-17, 2021, India

CEUR Workshop Proceedings (CEUR-WS.org)

well-known social media platforms, Twitter has gained particular attention because users can easily disseminate information about their opinions on a certain topic via a public message, called a tweet[2]. In addition to the information provided voluntarily by the user, a tweet may also contain information relating to the location of user and may contain links, emojis, and hashtags that may help users to better express their feelings. , which makes it an excellent source of valuable information [2]. Even more, Twitter has been used by government officials and politicians to inform the public about their activities or, in case , major events occurring . We introduced the problem to identify three classification on tweets as a shared task titled 'Information Retrieval from Microblogs During Disaster' (IRMiDis). Collects and annotates the COVID19 vaccination data set, determines the COVID19 vaccination posture detection classifier, and the number of reported events and tweets and posture (e.g., in favor, against or neutral) linked to There is. years of analysis by the media.Many teams participated in the IRMiDis shared task, and proposed several classifier to idenfy.The methodology was based on classification of tweets into three classes, viz. provax) and antivax and neutral using Support Vector Machine (SVM) model that is known to perform well for text classification. We have used SVM here because it is a very efficient simple classifier algorithm which is widely used for pattern recognition which can also have a very good classification performance than any other classifier.

## **2. RELATED WORK**

There has been a lot of research in recent years on utilizing online social media (OSM) during disasters, which involves several challenges such as, parsing short and informal messages, handling information overload, and prioritizing different types of information. The reader is referred to [1],[3] for comprehensive surveys on using OSM for disaster informatics. Several studies have attempted to identify particular types of information from microblogs posted during disaster events. For instance, some studies focused on extraction of situational information [4],[1]or actionable posts[3], while some works attempted to identify more specific information such as need of resources or availability of resources . Many of these prior works use either classification-based approaches (e.g., classifying tweets into situational and non-situational classes [4]. Machine learning methods include classic machine learning and deep learning algorithms. Classical machine learning algorithms frequently used for posture detection are support vector machines (SVMs) and many types[5] In the present work, we focus on identifying effective classifier for classification on tweets data.

## **3. MICROBLOG DATASET**

The present work is a task (IRMiDis) that we organized at the Annual Conference of the Forum for Information Retrieval Evaluation (FIRE) 2021. The task was to identify antivax,provax,neutral tweets that is related to covid vaccine from among a large set of tweets posted during Covid 19. This section describes the dataset (that was used in the shared task).

### **3.1. Collecting tweets during Covid 19**

We used Twitter search API2 to crawl 100k English tweets related to Covid-19 Vaccine which have been occurring worldwide since November-December 2019 with the keywords "covid", "pandemic", and 'vaccine'. We then removed the dataset redundancy and nearly duplicate tweets to get a chronologically sorted set of unique English tweets (based on Twitter-assigned timestamps).

## **4. PROPOSED METHOD**

The design of our proposed system for classification of tweets into provax,antivax,neutral is shown in the Fig. 2. The steps of designed model are provided as follows.

## **5. METHODOLOGY**

The main objective of this work was to classify the tweets into provax, antivax and neutral checkable accurately from a set of tweets that were gathered. In order to fulfill our purpose, our proposed framework first collects the set of tweets and then Support Vector Machine classification was applied in order to recognize the provax,antivax, neutral tweets properly. Our proposed methodologies are 'semi-automatic' in nature, where some manual effort is employed to generate a training set for the classifier, and then the classification and ranking are automatic.

### **5.1. Pre-processing the tweets**

In the current study 2792 tweets have been used to reveal the antivax/provax/neutral sentiment from a tweet. The raw tweets are preprocess ,and in this steps we clean data, remove stop words, white spaces then the tweets are case-folded to lowercase.

### **5.2. Constructing training set for classification**

After stemming and lemmatization are applied,the pre processed level tweets are then transform to numeric feature vector using term frequency inverse document frequency. After transforming the unstructured tweet data into numeric structured data Then classifier SVM (described later) is trained over this training set containing 991 provax tweets and 791 antivax tweets and 1010 tweets are neutral and get training accuracy 0.67.

### **5.3. Support Vector machine**

Support vector machines (SVMs) are a family of supervised learning algorithms used for classification, regression , and other tasks such as outlier detection.Other classification algorithms suffer from over-compliance, but one of the advantages of SVM is that this situation will be difficult . Another advantage lies in the fact that, in addition to binary classification, multi-class classification can also be performed by by combining several binary classification functions. Thus, each class in finds a classifier that separates it from the other classes for each class considered individually at a time.The SVM algorithm builds an N-dimensional hyperplane model

	precision	recall	fi-score	support
0	0.71	0.57	0.63	209
1	0.66	0.77	0.71	246
2	0.65	0.65	0.65	243
accuracy			0.67	699
macro avg	0.67	0.66	0.67	698
micro avg	0.67	0.67	0.67	698

**Table 1**  
classification result

that assigns future instances to one of two possible output classes. In our task we are applied SVM to solve the various type of classification of tweets.

#### 5.4. Identify fact-checkable tweets

The training dataset was compared with the trained dataset to identify three type of classification provax, antivax, and neutral tweets.

## 6. RESULTS AND DISCUSSION

In this task we firstly Preprocess the tweets after that converts the tweets into a feature vector then transforming the unstructured tweet data into numeric structured data the same has been used to train a SVM in order to solve the antivax, provax, neutral classification of tweets and at last, after a set of test and set of sorters, we come to the conclusion that the accuracy of the final result is 0.380 and the macro F1 score is 0.370. The result is shown in table 1.

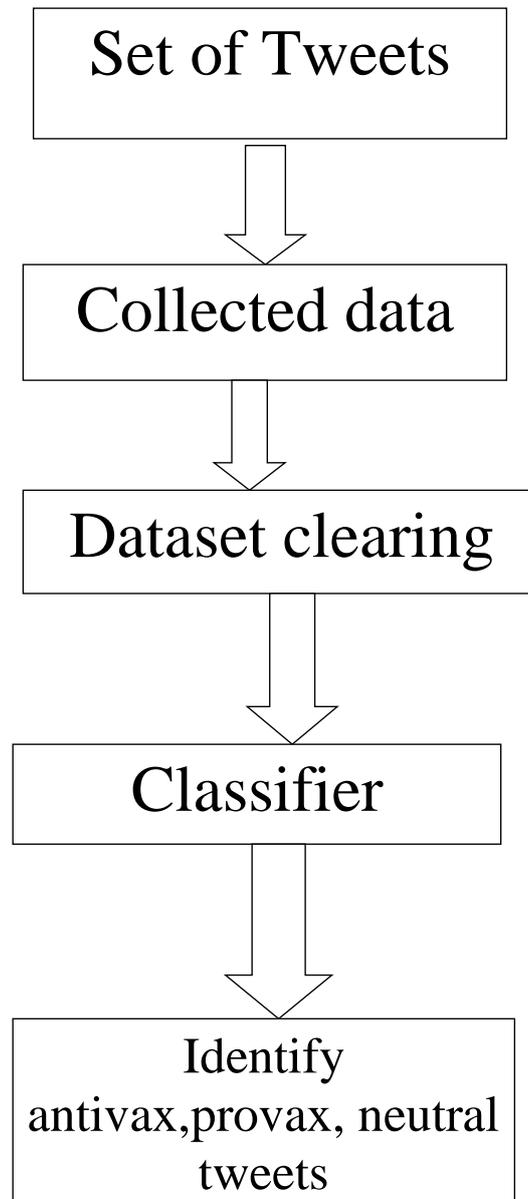
## 7. CONCLUSION

The purpose of this treatise was to monitor the number of Covid-19 vaccine twitter messages during the analysis period, matching major events reported by tweets for Covid-19 vaccination. The proposed approach categorized tweets. Covid-19 is divided into three major classes, Provac, Antivax and Neutral, in relation to vaccination. We educated the system on intermediate datasets to improve the accuracy of the system and evaluated the system on challenging datasets to prove robustness and durability. System reliability. It can be said that the system can be implemented with a dataset of tweets incidental to the recommendation and related large tweets in the future. And again, this operation may be further extended using other sorters to see the changes in the results.

## References

- [1] Worldometer, Coronavirus update (live): 63,777,845 cases and 1,477,777 deaths from covid-19 virus pandemic., 2020. URL: <https://www.worldometers.info/coronavirus/>.

- [2] E. D'Andrea, P. Ducange, A. Bechini, A. Renda, F. Marcelloni, Monitoring the public opinion about the vaccination topic from tweets analysis, *Expert Systems with Applications* 116 (2019) 209–226. URL: <https://www.sciencedirect.com/science/article/pii/S0957417418305803>. doi:<https://doi.org/10.1016/j.eswa.2018.09.009>.
- [3] K. Rudra, N. Ganguly, P. Goyal, S. Ghosh, Extracting and summarizing situational information from the twitter social media during disasters, *ACM Trans. Web* 12 (2018). URL: <https://doi.org/10.1145/3178541>. doi:10.1145/3178541.
- [4] K. Rudra, S. Ghosh, N. Ganguly, P. Goyal, S. Ghosh, Extracting situational information from microblogs during disaster events: A classification-summarization approach, in: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15*, Association for Computing Machinery, New York, NY, USA, 2015, p. 583–592. URL: <https://doi.org/10.1145/2806416.2806485>. doi:10.1145/2806416.2806485.
- [5] M. Basu, S. Ghosh, K. Ghosh, Overview of the fire 2018 track: Information retrieval from microblogs during disasters (irmidis), in: *Proceedings of the 10th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE'18*, Association for Computing Machinery, New York, NY, USA, 2018, p. 1–5. URL: <https://doi.org/10.1145/3293339.3293340>. doi:10.1145/3293339.3293340.



**Figure 1:** Proposed system for Classification of antivax, provax, neutral tweets