ANALYSIS OF TWITTER REVIEWS ON COVID-19

Mansi Saini
Student
Department of Computer Science
Indraprastha College For Women University of Delhi
sainimansi759@gmail.com

Khushboo Verma
Student
Department of Computer Science
Indraprastha College For Women University of Delhi
dukhushme@gmail.com

Surbhi
Student
Department of Computer Science
Indraprastha College For Women University of Delhi
hrsurbhi1512@gmail.com

Mr. Harendra Pratap Singh
Mentor
Department of Computer Science
Indraprastha College For Women University of Delhi
harendrapratap11@gmail.com

Dr. Archana Singhal
Mentor
Department of Computer Science
Indraprastha College For Women University of Delhi
harendrapratap11@gmail.com

Abstract- Currently, a newly born genome COVID declared as pandemic has spread globally and led to thousands of deaths, having an enormous impact on health, systems, and economies. COVID pathogenic viral infection is seeking attention worldwide, whether forcefully or being concerned. In such high social pacing, social media platforms have emerged as an alternate way for people to express their thoughts and reviews. TWITTER is one of the most popular platforms used worldwide by people to express their opinions. Using TWITTER API or Beautiful soul the latest reviews and comments related to COVID-19 will be extracted and analyzed. The motivation behind this review and analysis is to get feedback about how people think about the current situation using their positive or negative sentiment. Analysis has been done covering various aspects on the basis of economic conditions. Data mining tools can easily resolve those issues that traditionally were too time-consuming. Data mining of social networks can be done using graph mining methods such as classification, prediction, detection, pattern measurement, efficiency, and modeling, which are the basic part of data mining.
The global knowledge base should be strengthened for evidence-based decision-making in preventing and addressing the COVID-19 pandemic issues.

1 INTRODUCTION

We got introduced to a newly born genome COVID which is now declared

as Pandemic. This virus has spread globally and led to Millions of deaths and having an enormous impact on our Health Systems and Economics. This Worse situation in India is a wide part of the world pandemic (COVID-19) caused by Severe Acute Respiratory Syndrome Coronavirus. This Pathogenic viral infection seeking attention Worldwide whether forcefully or being concerned. It has given rise to millions of cases all over the country. So, for understanding the current thinking of people, with the help of Social media platforms (Twitter). So, Tweets have been extracted to analyse the views of people for COVID 19. The purpose behind extracting tweets is to identify the opinions/sentiments of Twitterati (Twitter users) as Good, Bad, or Neutral. The Classification of tweets can be done using the Graph Mining methods such as Classification, Prediction, Detection, Pattern Measurement, Efficiency, Data Processing, and Modelling which are the basic parts of Data Mining.

2 BACKGROUND

In such high social pacing, people got the alternate ways to express their thoughts and reviews on social platforms. Platforms like Twitter, Facebook, Instagram have the most appropriate and unbiased views about anything as these are the widely used applications where used applications where people feel free to discuss anything trendy. We have come across thousands of reviews based on the surviving pandemic, COVID 19 which expresses how people are thinking about this, then and now. We will be analysing reviews as positive and negative retrieving them from a renowned platform (Twitter) and will be showing the analysed result. So, for doing that we collected our data from Twitter, which is the best platform to know about feedbacks of people regarding any issue. It also classifies the trending hot topics in a linear and transparent way which is easy to reach and the source we used to do that is Data mining, Data mining implies analysing data patterns in large batches of data using one or more software. Data mining is used to discover patterns and relationships in the data in order to help make better business decisions. As the name justifies it mines the data for the purpose we are inspecting for. It has many task-based techniques

3 TECHNOLOGY

For sentiment analysis Twitter data has been used and sentiment analysis is done through Python to state whether a tweet is positive, negative or neutral. For accessing Twitter data (Data mining) one must have a Twitter developer Account which provides a self-serve tool to create and manage Twitter projects and apps. We can mine Twitter data using Twitter API or tweepy and store them in a csv file. CSV stands for comma separated values. CSV file allows us to save data in tabular form but values are separated by commas.

A. Why CSV?

CSV files are plain text files having tabular form of data therefore, creating them is more convenient for developers. It is easy to import them into a storage database. Here we used a csv file for storing tweets, in our csv file each tweet is stored in each row, so

extracting tweets for further processing becomes easy.

Tweepy is a python library that can be installed using the pip command. Authorized access to twitter can be done using the OAuth interface through consumer key and consumer secret and access tokens. Cursor interface of tweepy is used to iterate through different types of objects. We iterated through every tweet and saved them in a csv file.

Retrieved tweets have additional data like URL, username, etc. We have to filter extracted information in order to remove additional data. Re module of python provides us functions to check if a string matches a given regular expression, with help of these functions, additional data can be removed from tweets. For classification of tweets on the basis of their sentiment a sentiment analysis tool called VADER has been used. VADER stands for Valence Aware Dictionary and sEntiment Reasoner. VADER is a rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.

A. Why VADER?

VADER not only conveys that words like "happy"," good"," enjoy" etc. are positive and words like "bad", "fight"," difficult", etc. are negative but also understands that phrases like "not good", "did not love", etc. are negative and phrases like "not bad", "don't fight", etc. are positive. Moreover, it also understands the emphasis of capitalization of words.

B. Working of VADER.

Working of VADER depends on a dictionary that maps lexical features to emotion intensities. Sum of the intensity of each word in a text is called sentiment score. VADER expresses a text as positive, negative or neutral.

SentimentIntensityAnalyzer class of vaderSentiment.vaderSentiment module is used. polarity_scores() function will return a dictionary for each tweet having positive, negative, neutral and compound score as key and their respective score as values. We will check compound score on the basis of following criteria:

positive sentiment: (compound score ≥ 0.05)

$neutralsentiment : (compoundscore > -0.05) and (compoundscore < 0.05)$

$negativesentiment : (compoundscore \leq -0.05)$

4 RESULT REPRESENTATION

For representing the result, the pie graph and line chart has been plotted. These graphs are plotted using the pyplot interface of matplotlib library. The matplotlib is a python library, it provides users many interfaces and functions for 2D graphics similar to MATLAB. Matplot provides many named collections of methods, pyplot is one of them which allows users to construct 2D plots.

A. Why graphs?

We are working on thousands of tweets and we have checked all tweet's sentiment using VADER. After this we have a polarity score of all the tweets saved in a dictionary but showing overall sentiment is manually impossible, so we need some human readable content like graphs.

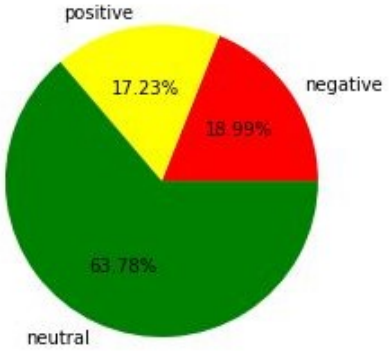
5 RESULT

Here our research comes to an end

taking up the result with it. The whole analysis shows that the positive content of the recent views analysis is 17.23% negative content is 18.99% and the tweets that are neutral results or unidentified by the provided library to be analysed as 63.78%.

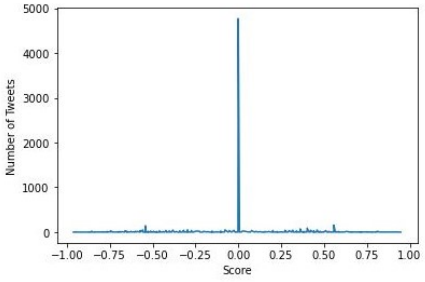
This shows that the public reviews have come to a changed view about COVID-19. Initially this pandemic and related circumstances were seen as a destructive timeline with every aspect. But now people's perspective has been changed. Now the recovered patients and concerned ones are keeping up motivating others to not to be scared of this. If anyone find themselves symptomatic just make yourself examined and fight with this just as the normal fever not letting it broke down on themselves. Also, other motivators are doing a great job posting daily reports, recovery rates, headline heroic stories to unite people with them.

The given pie graph shows a pictorial representation of the generated result:



The pie graph clearly shows that the yellow region that represents positive reviews of the total reviews is 17.23% , the red region that represents negative reviews of the total reviews is 18.99% and the rest reviews or the neutral or the unidentified reviews represented by green region is 63.78%. this is the appropriate data that we found using data mining tools.

Again, we plotted a line chart showing the points graph that was provided to each tweet and the frequency of the certain point. l0.25



The generated line chart show that the maximum number of times allotted score is 0. It may be unidentified or it may have equal number of positive and negative words/emoticons from the dictionary making the score neutral.

6 CONCLUSION

After obtaining the processed data, it can be concluded that the used data mining techniques such as Classification, Prediction, Detection, Pattern Measurement, Efficiency and Modelling, have resulted in more accurate and desired results. We successfully generated the favourable data through these techniques from the recent tweets and classified them in the required categories additionally analysed as graphical representations for viewer's comfort.

Data Pre-processing is the most fundamental step to handle large unstructured datasets to achieve the ac-

curacy of classified results.

Although this virus has threatened almost the whole world, still some people are there to maintain a sense of security and motivation for the sake of everyone.

7 ACKNOWLEDGMENT

We would like to thank you for our principal Dr. Babli Moitra Saraf who gave us a chance of doing this Research paper on the topic " Analysis of Twitter reviews on Covid 19" as well as our Mentor who help us a lot in writing this Research paper Mr. Harendra Pratap Singh and Dr. Archana Singhal.

8 REFERENCES

[1]https://www.infogix.com/top-5-data-mining-techniques/

[2]https://www.earthdatascience.org/courses/use-data-open-source-python/intro-to-apis/twitter-data-in-python/

[3]http://www.compciv.org/homework/assignments/twitter$_c$sv$_a$nalysis/

[4]https://developer.twitter.com/en/docs/developer-portal/overview

[5]https://www.geeksforgeeks.org/extraction-of-tweets-using-tweepy/#: :text=Twitter%20allows%20us%20to%20mine,available%20easily%20for%20each%20user.

[6]https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664