IBM Hack Challenege Project Report

On

Sentiment Analysis of COVID-19 Tweets – Visualization Dashboard

Submitted in the fulfillment of the IBM HackChallenge 2020

Team - Code Smashers

Ishan Kesarwani

Shivam Shivhare

School of Research and Technology IBM- ICE,

People's University, Bhopal (M.P)

(2020)

Table of Contents

ABSTRACT

This project addresses the problem of sentiment analysis in twitter; that is classifying tweets according to the sentiment expressed in them: positive, negative or neutral. Twitter is an online micro-blogging and social-networking platform which allows users to express their views in maximum length 280 characters. It is a rapidly expanding service with over 350 million registered users .Which 300 million are active users and half of them log on twitter on a daily basis - generating nearly 500 million tweets per day. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analysing the sentiments expressed in the tweets. Analysing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to develop a sentiment analysis and visualization model to help the government and individual to monitor and tackle the current Covid-19 situation.

CHAPTER 1: INTRODUCTION

Sentiment is an individual's personal view or opinion which he/she expresses. Analysing such kinds of human sentiments is known as sentiment analysis.

Sentiment analysis is the data mining which identifies and extracts subjective information which might be used for some specific purposes, and helps the society to understand the social sentiment of the people. In brief, it is the process of understanding the emotional sense behind a person's views. For analyzing the sentiments, the data is to be gathered. The data can be gathered from a number of sources such as Google, Twitter, Facebook, Pinterest et cetera. We are working on twitter sentiment analysis so we will gather data (sentiments) by calling the twitter API (tweepy).

Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics but as we all know, everyone is not using twitter so we would not get the exact data but what we will get is the sentiments of the small section of the society. We will perform the analysis on this extracted data (sentiments) and will visualize it graphically.

Sentiment analysis can be used in various ways. Some of them are as follows –

1. Businesses
2. Governments
3. Healthcare
4. Education

We will be performing sentiment analysis for the healthcare sector where we are analysing the people's mind on the novel corona outbreak and the issue caused due to this Covid-19 such as the imposing of the lockdown by the government to prevent the spread of the disease. With the help of our application the user will be able to track the current sentiments of the people and will also be able to search any #tag which the user wants. In the application the real time data analytics will be done.

CHAPTER 2: OBJECTIVE

In view with the current scenario of the outbreak of the Covid-19 it is extremely necessary to have an application which can monitor and analyze the people's sentiments. In this twitter sentiment analysis, our main objective is to extract the sentiments of the people on the outbreak of Covid-19 and the extension of the lockdown. We will analyze this information and also visualize it graphically. Such information may be of high value to different sectors. Talking about the Government, the government will get to know the sentiments of its people and take necessary actions to handle the scenario. This may also help the government to enact better policies for the people of the country.

The benefit of such kind of twitter sentiment analysis is not restricted to only one sector but it might affect a number of people in a positive sense. Now considering the scenario of the insurance sector, the companies may launch new policies seeing the current market conditions. The health sector will get to know about the medical needs of the people of the country. Similarly, there are other sectors also which will be benefited with this twitter sentiment analysis.

To achieve this goal we will be calling the twitter API (tweepy) and get the tweets of the people. As we are working on the sentimental analysis we will get the comments (sentiments) from the tweets. For authentication, we will import the OAuthHandler from the tweepy module and we will use TextBlob to tokenize the comments and NLTK for the analysis and extract the useful information from it.TextBlob aims to provide access to common text-processing operations through a familiar interface on which we will perform sentiment analysis. After getting the tweets tokenized, the polarity (positive, negative, neutral) will be calculated. Once all the useful data is extracted. For handling the huge amount of data pandas library will be used.

After all the analysis of data is done we use Python Django framework for visualizing the data in various ways. We represent the real-time data in the form of various kinds of graphs, charts, and statistics.

CHAPTER 3: LITERATURE SURVEY

3.1 Existing Problem

There are many websites which show twitter sentiment analysis but as one can find they do not represent the information in a clear manner as it is quite difficult to distinguish between the kind of sentiment. The biggest drawback of all these web applications is that they can analyse only the pre-searched keywords or #tags making the application useful to only the limited section of the society.

3.2 Proposed Solution

In our twitter sentiment visualization, we took care of all these drawbacks and have represented the information in a very clear manner that even the person with non technical background will be able to easily monitor the sentiments. The biggest feature which we have added is that we have a search bar to allow our users to search for any relevant #tags at real-time.

CHAPTER 4: REQUIREMENTS

4.1 Functional Requirements

- The application is hosted on the IBM cloud.
- A search bar will be implemented to search for the particular #tag/keywords and the application will show suggested #tag.
- The data related to the searched #tag will be subjected to data analytics and visualization.
- The sentiment analysis will be done on the twitter data and the polarity (positive, negative, or neutral) of the data will be calculated to know the sentiment of the people.
- The data will be visualized in different graphs.
- The graphs will be changed with the change in the searched hashed tag.
- The data will be shown in the real-time.
- Different #tag will represent different graphs.
- Different kinds of graphs will be represented such as –line chart, bar graph, and pie chart.

4.2: Non- Functional Requirements

4.2.1: Safety Requirements

Since our system is hosted on the IBM Cloud, it will take daily backup.

4.2.2: Security Requirements

IBM Cloud provides us security feature that can improve our system requirements.

4.2.3: Software Quality Attributes

- Availability: Since we are hosting our project on the server it will be available all the time.
- Correctness: The system should collect real time data from the twitter and perform appropriate analysis and visualization on the collected data.

- Maintainability: The system should keep all the data on the cloud.
- Usability: The system should satisfy the maximum number of users need.

4.3: System Requirements

The system requirements or software requirements is a listing of what software programs or hardware devices are required to operate the program properly.
The hardware requirements are the requirements of a hardware device. Most hardware only has operating system requirements or compatibility. For example, a printer may be compatible with Windows XP but not compatible with newer versions of Windows like Windows 10, Linux or the Apple macOS.

4.3.1: Hardware Requirements

- Processor (CPU) with 1 gigahertz (GHz) frequency or above.
- A minimum of 1GB of RAM.
- A minimum of 20GB of hard disk.
- Keyboard and a Mouse or some other compatible pointing device

4.3.2: Software Requirements

To develop the web application the following requirements should be fulfilled-
Operating System:

Any of the following operating system will be required.

- Windows 7, Windows 8 or Windows 10
- Mac OSX 10.8, 10.9, 10.10 or 10.11
- Linux

Software Required:

- Python Django framework (3.0.6)
- IBM Cloud (Used for hosting the application.)

- Any of the PyCharm IDE / IntelliJ IDEA IDE.

4.4 Software Model - Waterfall Model

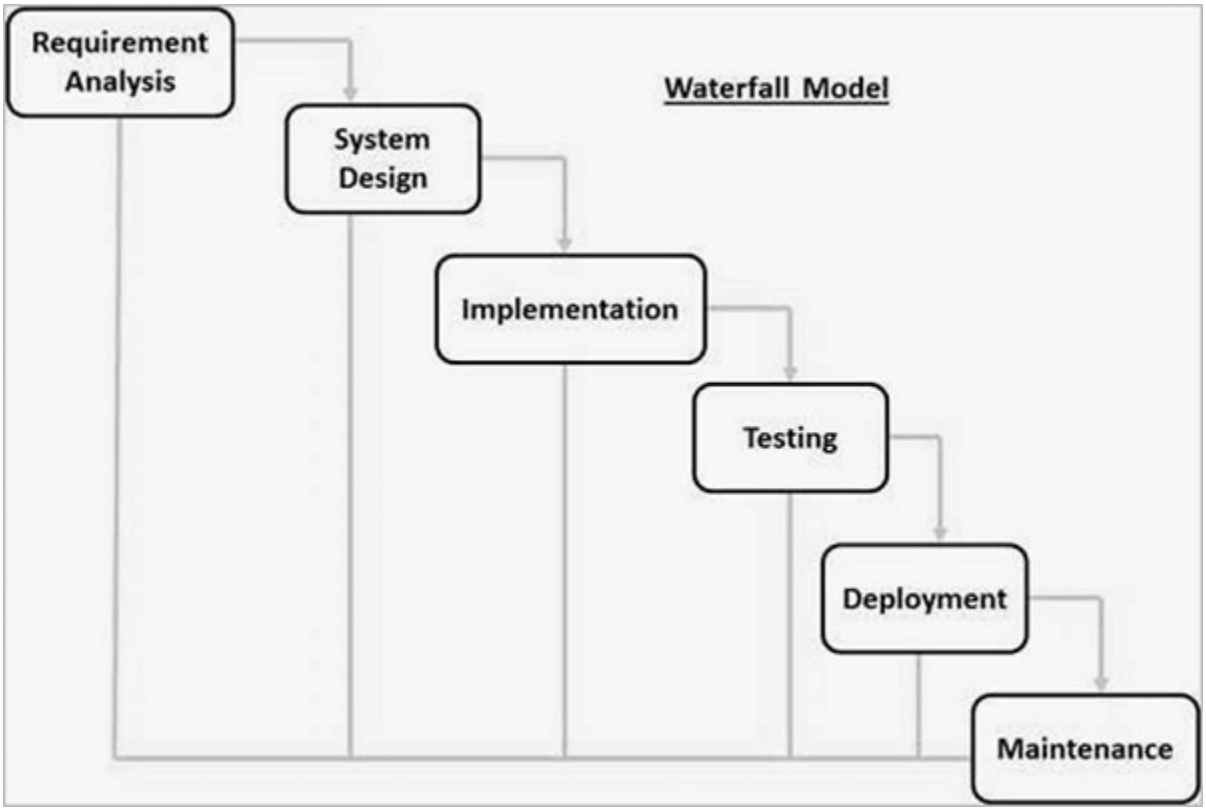


Fig. 4.4 Waterfall Model

The Waterfall Model was the first Process Model to be introduced. It is also referred to as a **linear-sequential life cycle model**. It is very simple to understand and use. In a waterfall model, each phase must be completed before the next phase can begin and there is no overlapping in the phases.

The waterfall Model illustrates the software development process in a linear sequential flow. This means that any phase in the development process begins only if the previous phase is complete. In this waterfall model, the phases do not overlap.

The sequential phases in Waterfall model are −

- **Requirement Gathering and analysis** − All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification

document.

- **System Design** − The requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.
- **Implementation** – With the inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.
- **Integration and Testing** − All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.
- **Deployment of system** − Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market.
- **Maintenance** − There are some issues which come up in the client environment. To fix those issues, patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

4.5 Feasibility Study

The prime focus of the feasibility is evaluating the practicality of the proposed system keeping in mind a number of factors such as cost efficiency, time management and memory management. The following factors are taken into account before deciding in favor of the new system.

4.5.1 Economic Feasibility-

As the application is developed to perform data analytics and visualization on the Covid-19 and related tweets the application is available to monitor the data free of cost.

4.5.2 Technical Feasibility-

To keep the track of the sentiment of the people due to the novel Covid-19 and the lockdown

enforced by the government to lessen the effects of Corona Virus, the tweets of the people with the related #tags will be gathered and the data analytics will be done to illustrate the accurate sentiment of the people. The same will be shown graphically.

4.5.3 Operational Feasibility- It's easy to use interface helps in efficiently monitoring the real time data. Various kinds of graphs are used to better illustrate the sentiment of the people in different parameters.

CHAPTER 5: PROJECT MODULES

1. Search #tag – The application allows its users to search for some specific #tag. Although analytics and visualization results on some of the #tags related to Covid-19 are already displayed on the main page. The data related to searched #tag will appear above the pre-displayed results.
2. Monitoring and filtering data – The data will be monitored on the real time basis. The data can also be filtered on the basis of the searched data.
3. Visualizing through graphs – The analysis will be performed on the basis of the tweets of the people fetched using the twitter.

CHAPTER 6: SYSTEM ARCHITECTURE

6.1 Client-Server Architecture

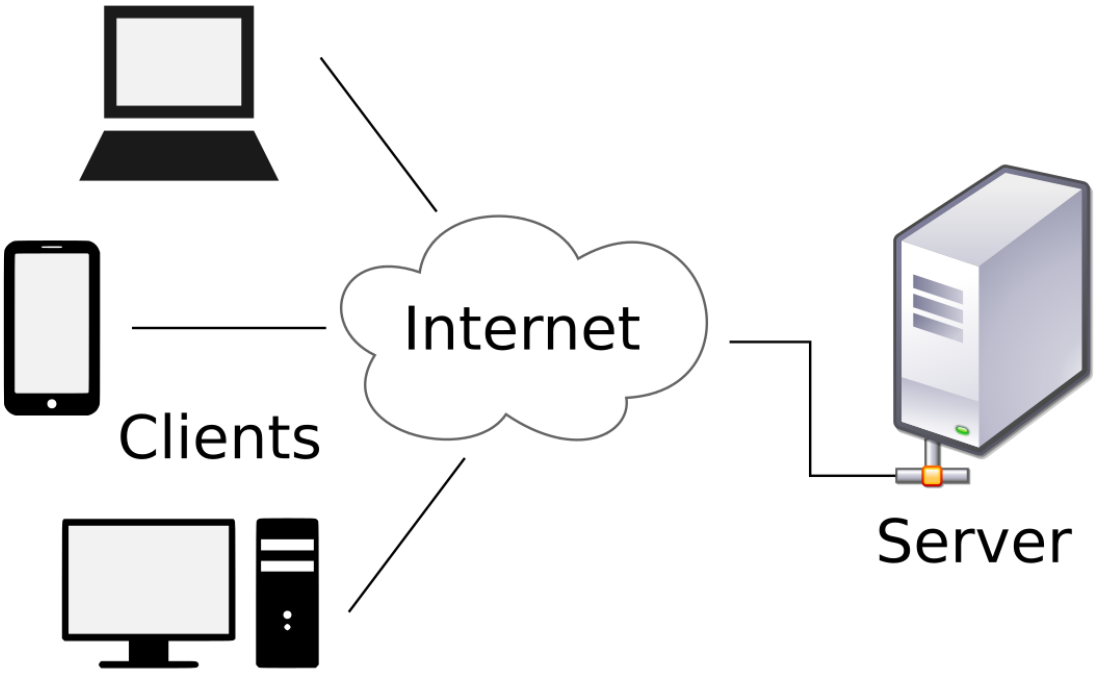


Fig. 6.1 Client - Server Architecture

The Internet has brought revolutionary change in the world of technologies, bringing the entire globe interconnected. But it follows some specific architecture and structure for communication. The popular is the client-server architecture which is a computing model, where the server hosts, distributes and controls the majority of the resources as well as services to be used by the client. Such structural designs are made up of one or more client systems connected to central or main servers through a network, which we usually know as an Internet connection. All such systems associated with it share computing resources.

The client-server architecture is also termed as a network-computing structure because every request and their associated services are distributed over a network. In the client-server architecture, when the client computer sends a request for data to the server through the internet, the server accepts the requested, process it and deliver the data packets requested back to the

client. One special feature is that the server computer has the potential to manage numerous clients at the same time. Also, a single client can connect to numerous servers at a single timestamp, where each server provides a different set of services to that specific client.

CHAPTER 7: DESIGN AND IMPLEMENTATION

7.1 User Classes and Characteristics

1. Users – The user will directly land to the home page. The users can search for the related #tag and the data analytics and visualization to perform sentiment analysis will be done on that data. Apart from the searched #tags sentiment analysis related to Covid-19 and Lockdown extension tweets will be shown.

7.2 Use Cases

Actor	Use Cases
Users	a. Search b. Fetch Twitter Data c. Sentiment analysis d. Sentiment Visualization

7.3 Use Case Diagram

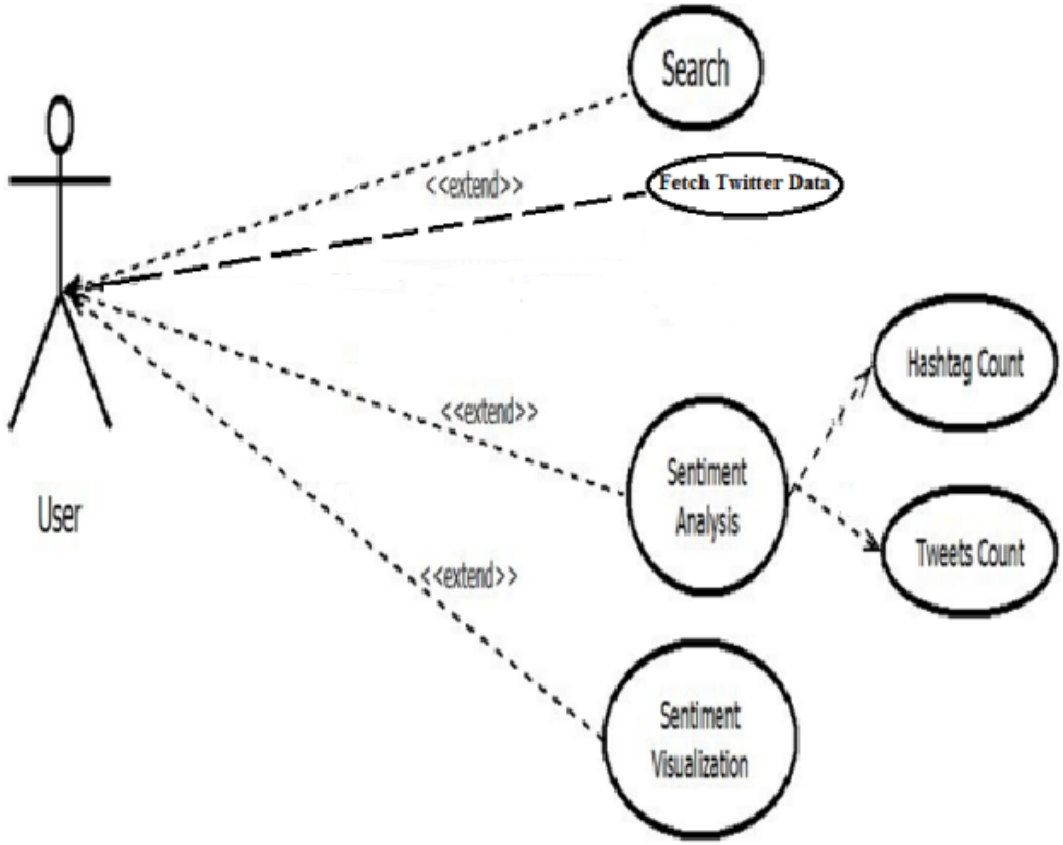


Fig 7.3.1 – User Use Case

7.4 Use Case Description

Use Case Number 1-

Title: Search the #tag.

Description: The application will have a search bar to search for the relevant #tag.

Actor: Users

Pre condition: The users need to open the web application.

Post condition: Two types-

1. Success: The search bar provides suggestions of pre-saved #tags and searches for any new #tag if it is relevant.
2. Failure: Unable to find the searched #tag.

Path: Steps in the use case –

Primary:

1. Users need to click onto the search bar.
2. Some suggestions will be shown to the users.
3. The user can search for the relevant #tag.
4. If the #tag exists in the twitter, its related data will be displayed else the message will be displayed that no #tag found.

Exception Path:

1. Database error.
2. Internet connectivity error.
3. API calling error.

Use Case Number -2

Title: Fetch Twitter data

Description: The sentiments is done through twitter data

Actor: User

Pre condition: Twitter credentials should be correct.

Post condition: Two types-

1. Success: Data receive successfully.
2. Failure: Unknown API error.

Path: Steps in the use case –

Primary:

1. Tweepy module is use for accessing the Twitter API.
2. Using textblob library for processing textual data.
3. NLTK is use for natural language processing to tokenize the words,sentences,stemming and tagging.
4. NLP prediction will be used for Data prepartion,training,prediction & feature.
5. Polarity will be calculated to find out the sentiment of the tweet such as negative,Positive and Neutral.
6. Once the polarity is calculated sentiment analysis is done.

Exception Path:

1. Unknown API Error.
2. Server Connectivity Error.

Use Case Number – 3

Title: Sentiment Analysis

Description: The analytics of the sentiment of the people can be viewed in the form of the well structured tables.

Actor: User

Pre condition: The application should fetch the data.

Post condition: Two types-

1. Success: The data is displayed.
2. Failure: Fatal error.

Path: Steps in the use case –

Primary:

1. The users can search the #tags to monitor the real time analysis.
2. The users can also monitor the default data analysis displayed on the Covid-19, Lockdown Extension tweets.
3. The user finds analysis of sentiments of the people.
4. User can easily filter the sentiment day wise, week wise and month wise.
5. All the data displayed will be in real time.

Exception Path:

1. Database error.
2. Server Connectivity Error
3. API call error

Use Case Number -4

Title: Sentiment Visualization

Description: The visualization of the sentiments is done through different graphs.

Actor: User

Pre condition: Sentiment analysis data should be obtained.

Post condition: Two types-

1. Success: The users can see visualization of the sentiments of the people.
2. Failure: Fatal error.

Path: Steps in the use case –

Primary:

1. The user gets to see sentiment analysis.
2. The user can see visualization.
3. The user can filter the graphs to visualize the sentiments of the people with different kinds of the graphs.
4. The user can also visualize any other relevant #tag.

Exception Path:

1. Database error
2. Server Connectivity Error
3. API Call Error

7.5 System Flow Diagram:-

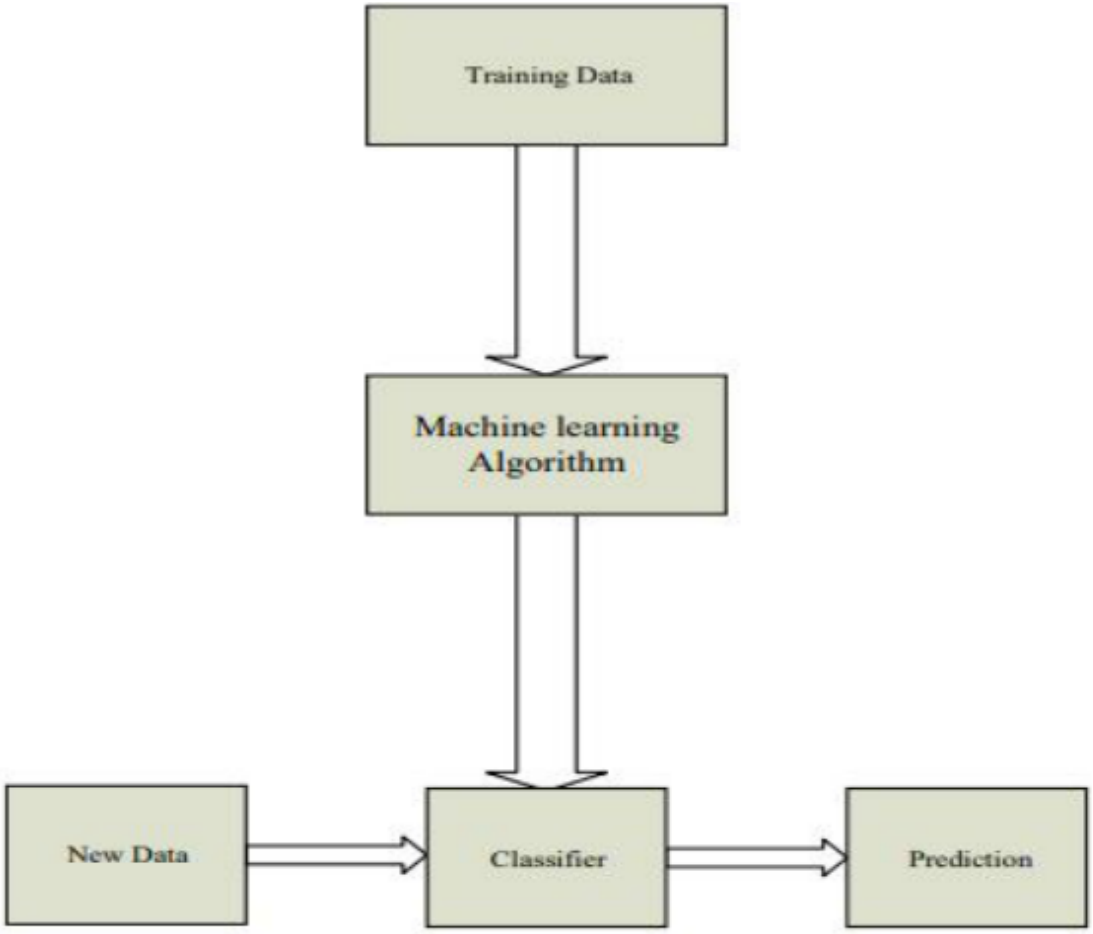


Fig 7.5.1 – System Flow Diagram for Sentiment Analysis using Twitter API

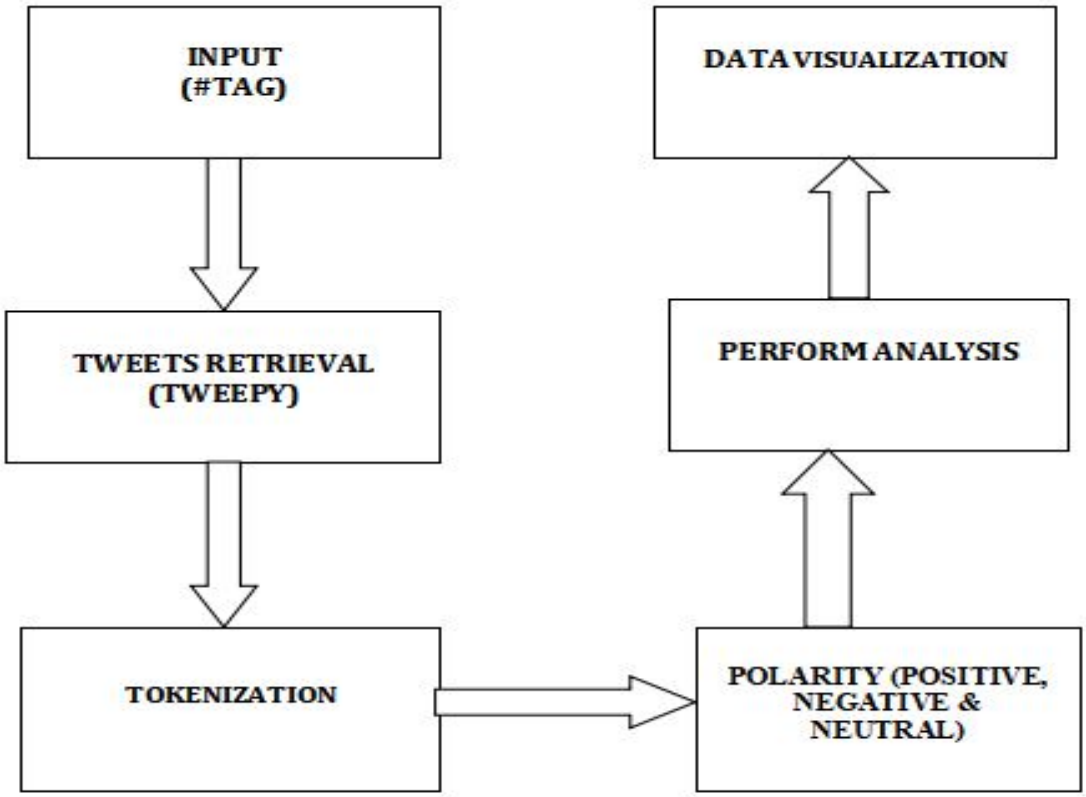
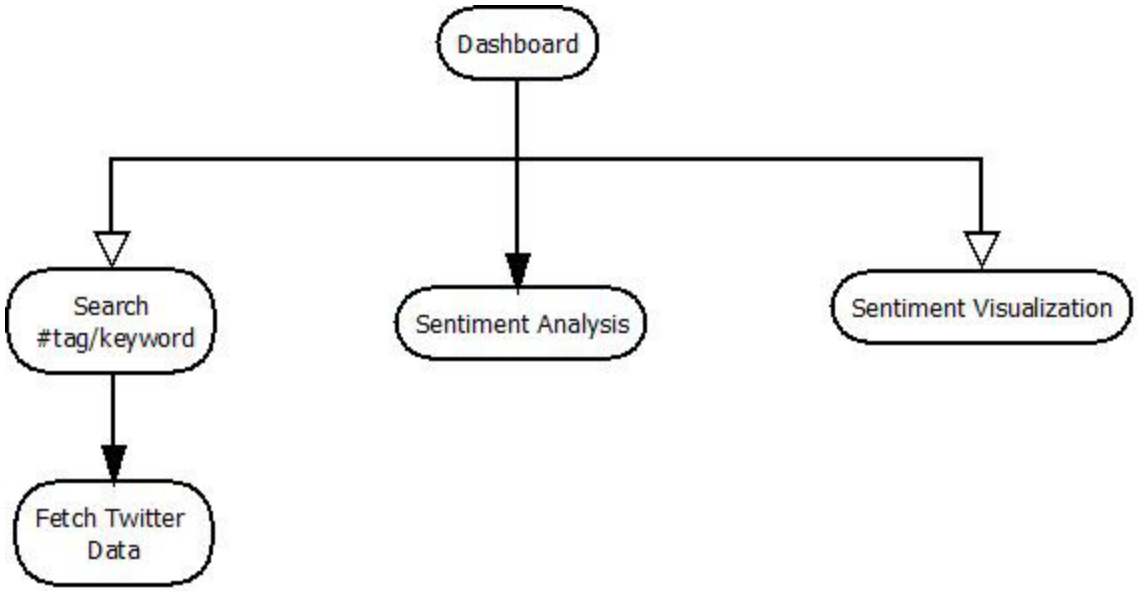


Fig 7.5.2 – System Flow Diagram for Sentiment Analysis using Twitter API

7.6 Control Flow Diagram:-



CHAPTER 8: ADVANTAGES AND DISADVANTAGES

The benefits of sentiment analysis

- Sentiment analysis is a useful tool for any organization, government or group for which public sentiment or attitude towards them is important for their success - whichever way that success is defined.
- On social media, blogs, and online forums millions of people are busily discussing and reviewing businesses, companies, and organizations. And those opinions are being ‘listened to’ and analysed.
- Those being discussed are making use of this enormous amount of data by using computer programs that don’t just locate all mentions of their products, services, or business, but also determine the emotions and attitudes behind the words being used.
- The results from sentiment analysis help businesses understand the conversations and discussions taking place about them, and helps them react and take action accordingly.
- They can create better products and services, and they can formulate the marketing messages they send out according to the sentiments being expressed by their target audience or customers.
- All of which adds up to increased sales and revenue.
- By listening to and analysing comments on Facebook and Twitter, local government departments can gauge public sentiment towards their department and the services they provide, and use the results to improve services such as parking and leisure facilities, local policing, and the condition of roads.
- Universities can use sentiment analysis to analyze student feedback and comments garnered either from their own surveys, or from online sources such as social media. They can then use the results to identify and address any areas of student dissatisfaction, as well as identify and build on those areas where students are expressing positive sentiments.
- Businesses can compare their results with those of their competitors to better understand people’s attitude to their business. They can identify where they may be excelling, or identify where there’s room for improvement compared to the competition.
- They can also conduct market research into general sentiment around key issues, topics, products, and services, before developing and launching their own new services, products or features.

Disadvantages of automated sentiment analysis

- Sentiment analysis tools can identify and analyse many pieces of text automatically and quickly.
- But computer programs have problems recognizing things like sarcasm and irony, negations, jokes, and exaggerations - the sorts of things a person would have little trouble identifying. And failing to recognize these can skew the results.
- 'Disappointed' may be classified as a negative word for the purposes of sentiment analysis, but within the phrase “I wasn't disappointed", it should be classified as positive.
- We would find it easy to recognize as sarcasm the statement "I'm really loving the enormous pool at my hotel!", if this statement is accompanied by a photo of a tiny swimming pool; whereas an automated sentiment analysis tool probably would not, and would most likely classify it as an example of positive sentiment.
- With short sentences and pieces of text, for example like those you find on Twitter especially, and sometimes on Facebook, there might not be enough context for a reliable sentiment analysis. However, in general, Twitter has a reputation for being a good source of information for sentiment analysis, and with the new increased word count for tweets it's likely it will become even more useful.

CHAPTER 9: Result

The result of this application is to visualize the analysis of sentiment via Twitter API.

Snapshots of the application are attached below:

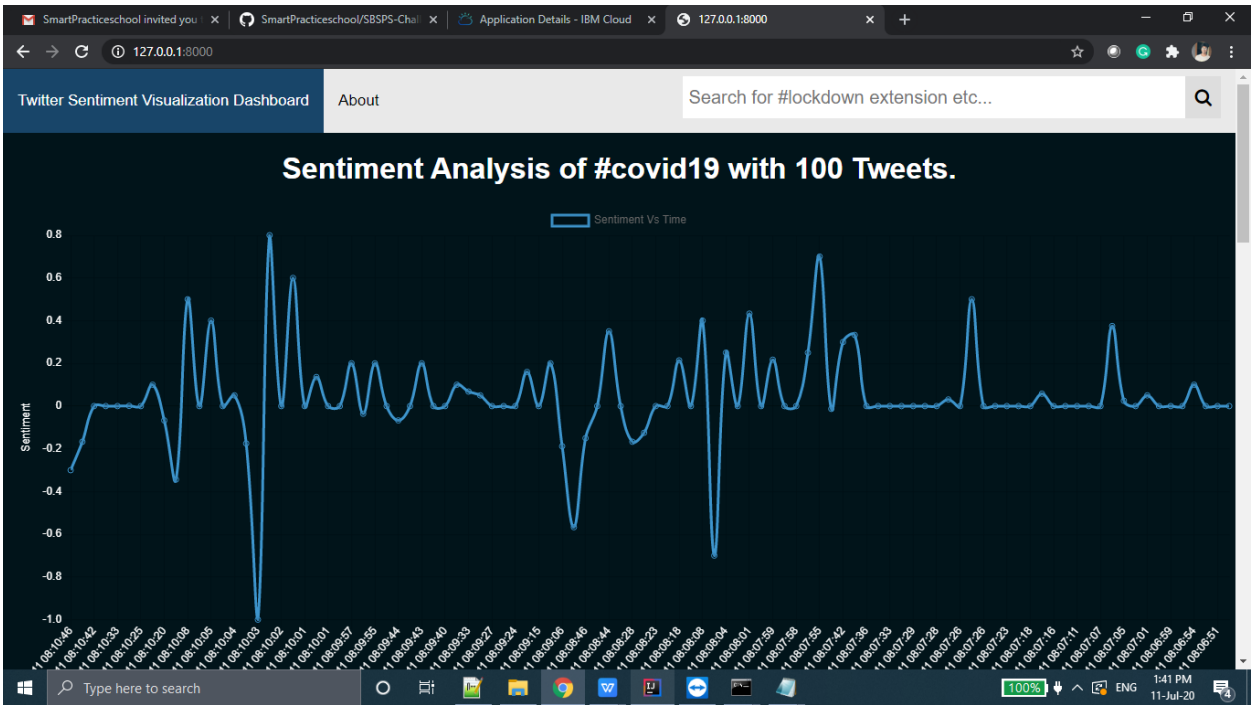


Fig-9.1 Line Chart Sentiment VS Time

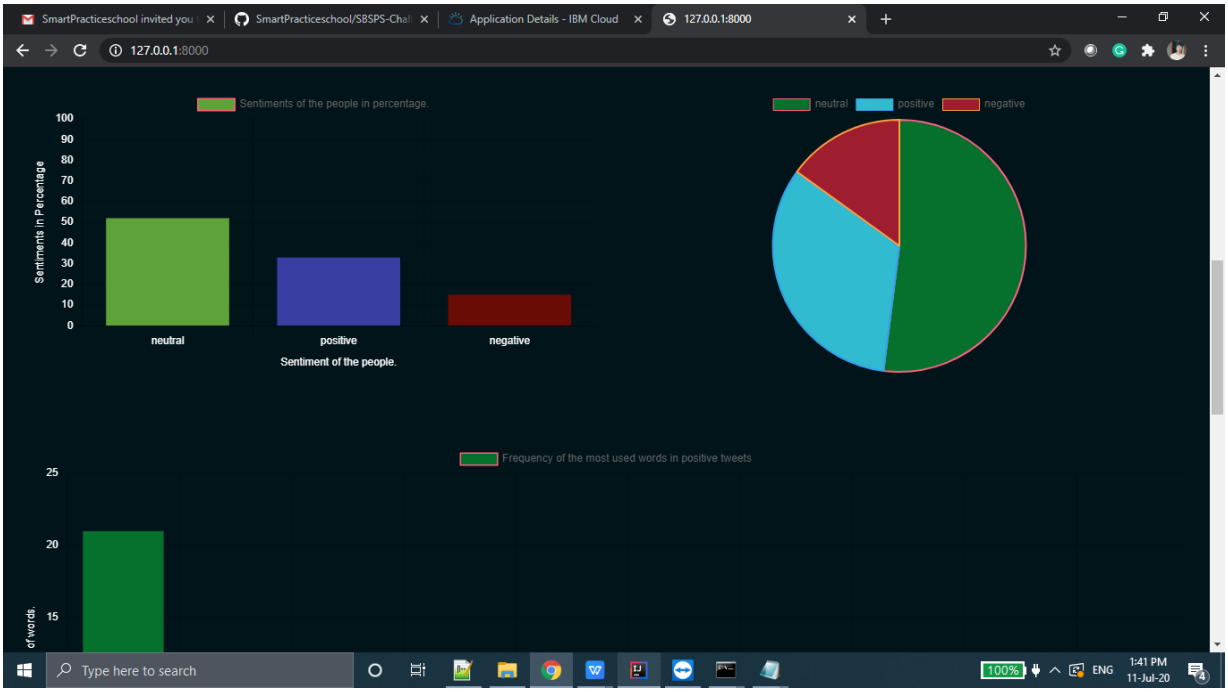


Fig-9.2 Bar & Pie Chart Sentiment of the people in percentage.

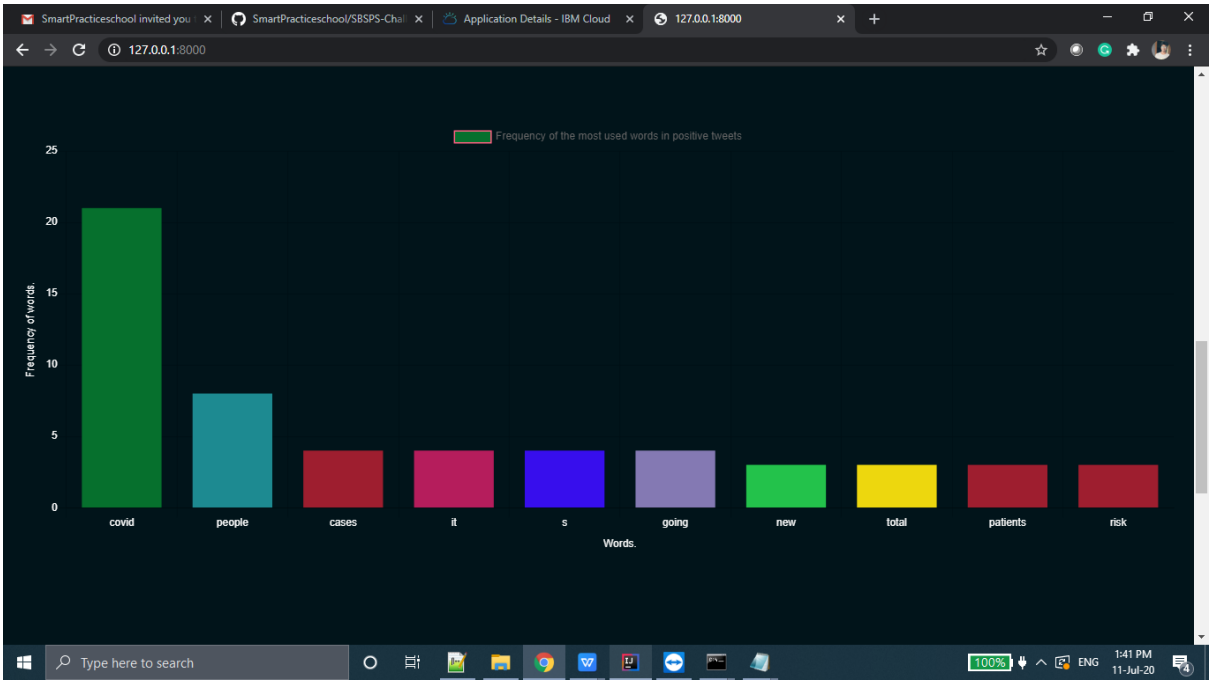


Fig-9.3 Word frequency of Positive Tweets.

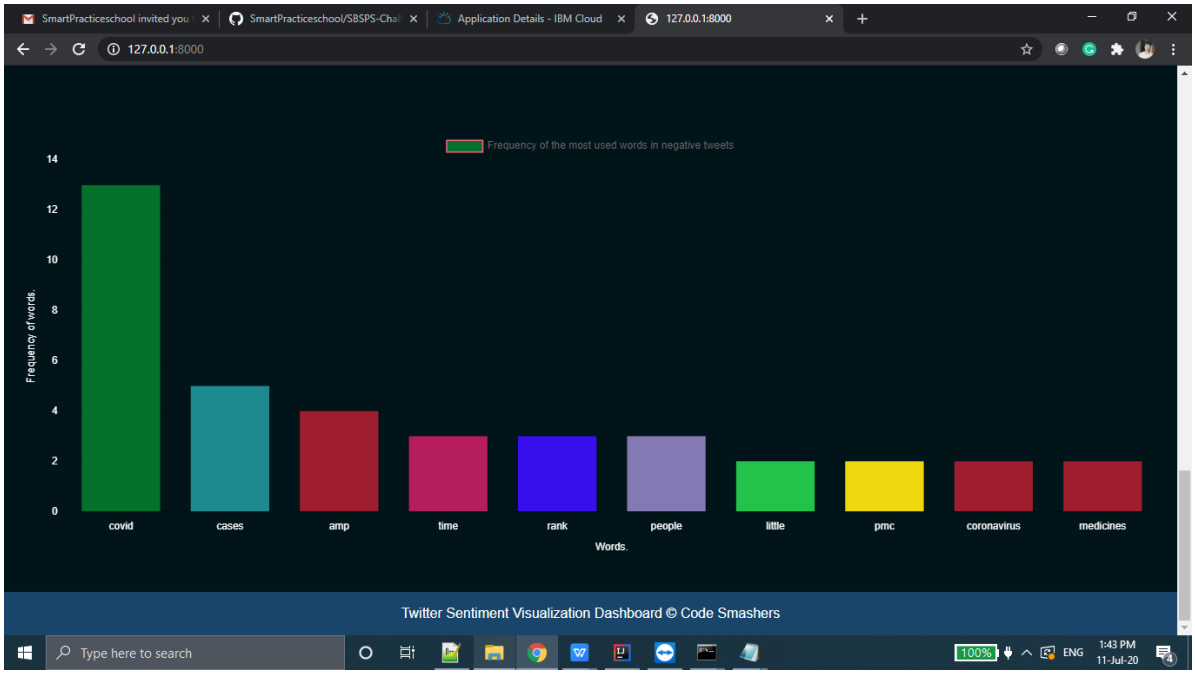


Fig-9.4 Word frequency of Negative Tweets.

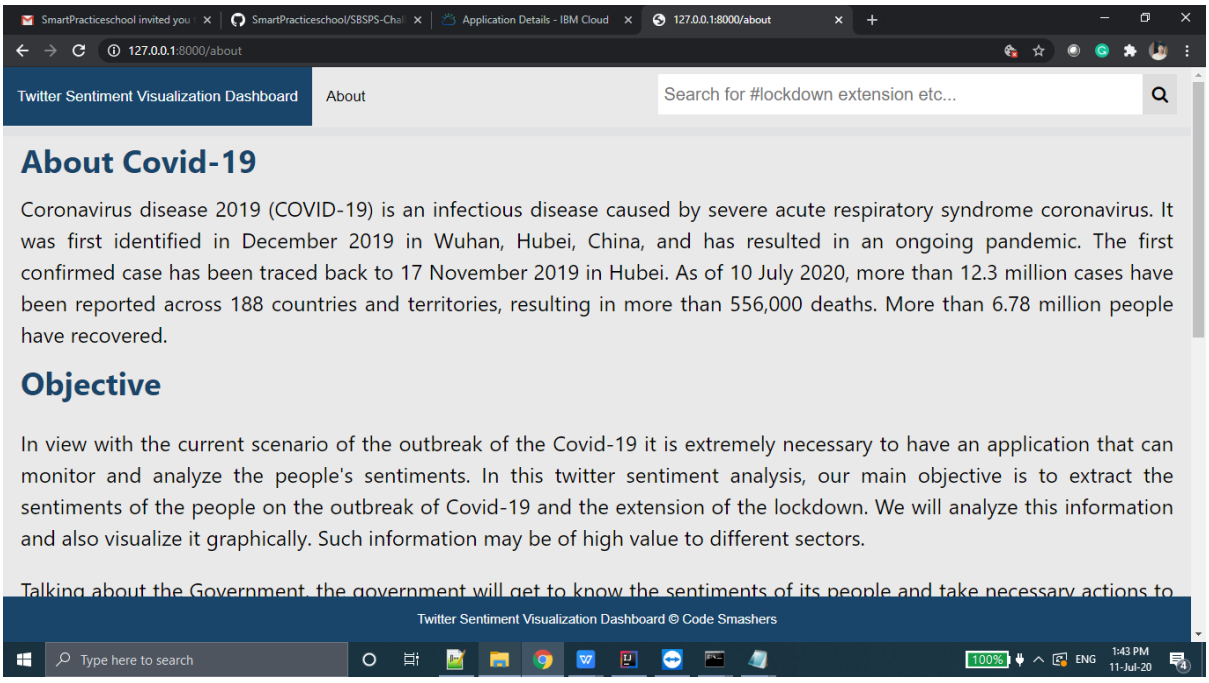


Fig-9.5 Twitter Sentiment Visualization Dashboard About Page.

CHAPTER 10: CONCLUSION

Nowadays, sentiment analysis is a trending topic in machine learning. Here we worked on a visualization dashboard to visualize the twitter sentiment analysis of people, which they tweet in twitter to know about the emotions, and sentiments. We have done sentiment analysis by gathering the actual data from tweets and performed data analytics to analyze the people's sentiment about Covid-19 or lockdown extension. We are detecting the sentiments of a huge amount of texts very accurately because the complexity in the English language is low and even more if we consider other languages such as Hindi it is difficult to sentiment. In this project, we tried to show the basic way of classifying tweets into a positive, negative, or neutral category. We have also made a graph between polarity and time which helps to better visualize the sentiment with time.

Chapter 11: FUTURE SCOPE

In twitter sentiment analysis and visualization we will visualize the tweets fetched through twitter streaming API and thus we will be able to move the graph at real-time once the connection to the twitter is made. We will also be filtering the tweets country-wise and state wise. After the state-wise filtration of the tweets is done it will also be visualized on the map of the respective country.

We will call the WHO API and show the covid-19 data on our dashboard.

Moreover, we can also add a dropdown of suggested #tags/keywords in the search bar.

CHAPTER 12: BIBLIOGRAPGHY

We have referred the following sites while developing this web application-

1. Towardsdatascience - https://towardsdatascience.com/. This site has a lot to offer for the newbies in the field of Data Science
2. AnalyticsVidya – https://www.analyticsvidhya.com/ .We have taken some help from this website also.
3. Geeks for Geeks – https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/ . This website has also helped in starting the project.
4. Medium.com - https://medium.com/better-programming/twitter-sentiment-analysis-15d8892c0082
5. IEEE website.
6. Chart.js -https://www.chartjs.org/docs/. For visualizing the graphs on Django.
7. ApexCharts.js- https://apexcharts.com/docs/installation/ For visualizing the graphs on Django.
8. YouTube – A lot of free contents are available for everyone on the youtube.
9. Twitter (tweepy) Documentation