Project Report On

Twitter Sentiment Analysis

Submitted by

Subham Mohapatra (181EE147)
Sachin Pandey (181EE237)
Ankit Rai (181EE109)

IV SEM B.Tech (IT - Minor)

Under the guidance of

Dr. Anand Kumar M
Assistant Professor
Dept of IT, NITK Surathkal

in partial fulfillment for the award of the degree

of

Bachelor of Technology

in

Information Technology

at

Department of Information Technology

National Institute of Technology Karnataka, Surathkal.

April 2020

Department of Information Technology

National Institute of Technology Karnataka, Surathkal

Mid Semester Evaluation (February 2020)

Course code : IT 252

Course Title: Minor Project

Title of the Project: Twitter Sentiment Analysis

Details of Project Group

Name of the Student	*Register No.*	*Signature with Date*
1. Subham Mohapatra	181EE147	
2. Sachin Pandey	181EE237	
3. Ankit Rai	181EE109	

Name of Project Guide:

Signature of the Project Guide: Dr. Anand Kumar M

Place:

Date:

ABSTRACT

Social Media in the current world scenario is a way to define the opinion and thought process of us, as individuals of a society. The major idea and opinion sharing takes place across many of these social media websites and it defines how we perceive our notion towards certain ideas and things. One such social media platform is twitter and through this project we built a website which monitors the tweets on an individual basis and gives us various information about a certain ‘hashtag’ which has been there on twitter.

Through this project we can view the positive, negative, factual and opinion related tweets related to a certain hashtag. This acts as a filter to go through the idea which is there on twitter and simply form our opinion by going through the top positive, negative, factual and opinion related tweets to that subject matter.

TABLE OF CONTENT

Chapter 1

INTRODUCTION

1.1 Objective

The objective is to have a website which gives various stats regarding a particular tweet hashtag. The website can have just the tweets shown as the stats or we can have a graphical representation of the stats. It can run locally after downloading the repository and running the commands.

The website will be made using flask and it will be using tweepy and textblob for the NLP part of the website. The frontend will be basic CSS, javascript and bootstrap. We will be later on adding visual graphs and circular charts for better representation of the data. We will use tweepy to interact with the twitter api and textblob will help us with the NLP as it is based on the famous NLP library ,i.e., NLTK.

It should be able to classify positive, negative, factual and opinion based tweets based on the provided hashtag. This will help the user to follow up with the idea of other users on the twitter platform.

1.2 Motivation

The motivation behind this project was to minimize the hassle while forming an opinion based on other people's thoughts present on the platform(Twitter). This will help the users to easily skim over the current trends on the platform without being bothered with the various non-sensible tweets on the platform.

With the addition of graphical representation the platform should become even more user-friendly and attractive.

Chapter 2

Requirements Analysis

2.1 OS Requirements

The website will run on the Operating Systems(Windows, MacOS, Linux) with the basic requirement of them having a web browser. To run the website locally the repository folder has a requirements.txt which has all the python installations that are required.

2.2 Application Requirements

The application requires the following packages:

- python 2
- Flask
- pip
- tweepy
- textblob
- pigar

We also need our very own twitter API credentials to work with the website. We also need to run the flask server to view our website.

Chapter 3

Implementation

3.1 Methodology

The basic idea is to collect tweets from twitter with the required hashtag and then we will use our model to classify the tweets into various categories based on the model. The website will take input from the user and will let it decide what category of tweets related to that topic does he/she want. Now, based on the category selected we will classify all the tweets related to that hashtag in an orderly manner.

For positive and negative tweet classification we will be using polarity subjects with range -1.0 to 1.0 in float format to classify all the tweets into the two categories. It will give us a set of approximately 13 tweets in real time and with each refresh we will get a new set of tweets related to that hashtag.

For opinion and factual classification we will be using the subjectivity of the tweet with range 0.0 to 1.0 in float format to classify them. The more subjective the tweet is, it will have higher subjectivity rating and hence will be classified opinions while the one with lower subjectivity rating will be classified as fact**s.**

3.2 Tools

The website is using general HTML, CSS, JS and Bootstrap for the frontend and then we use tweepy to interact with the twitter api and it helps us perform various tasks such as twitter Authentication(OAuth) with a breeze. Using the tweepy package we can easily collect data and train our model which uses the textblob package. Using this package we have a parameter known as sentiment which helps us to classify our data as it has parameters such as Polarity and Subjectivity which help us in our data classification.

Chapter 4

Results and Analysis

4.1 Screenshots

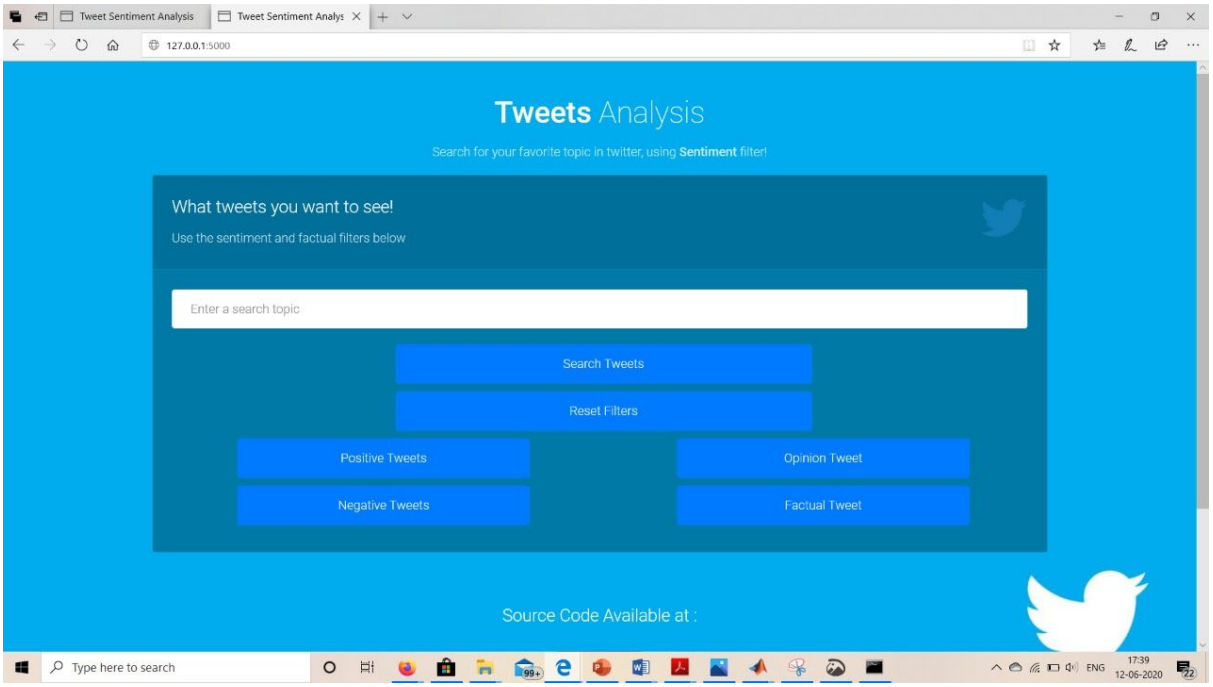

Fig 1. The basic look of the website

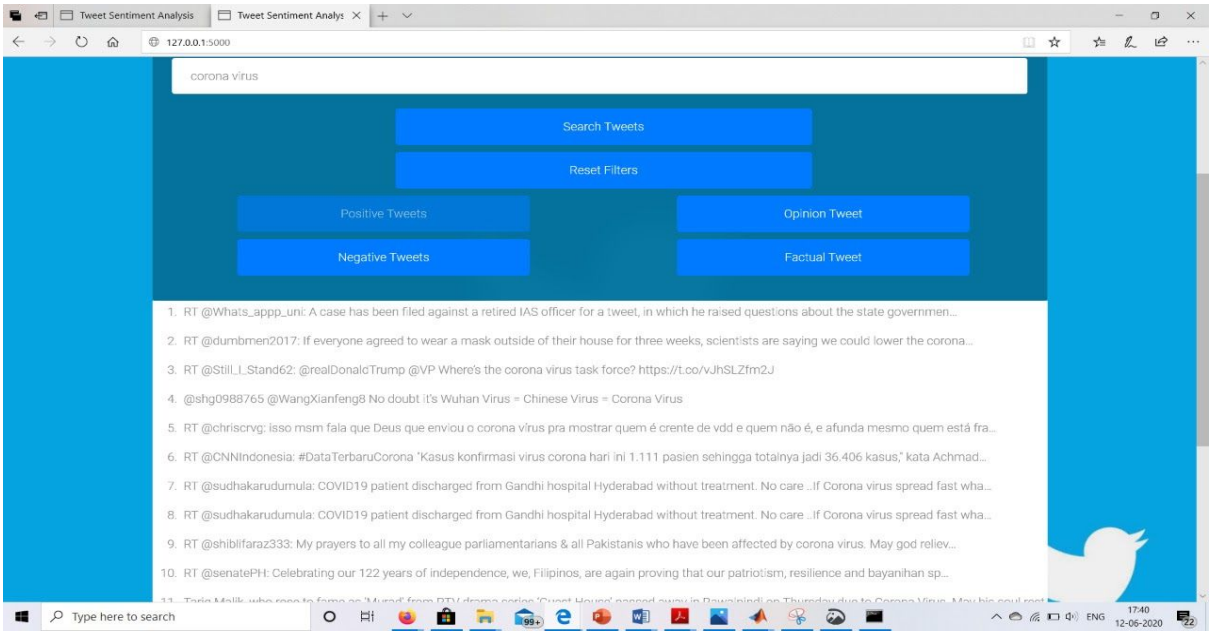

Fig 2. Positive search results for coronavirus

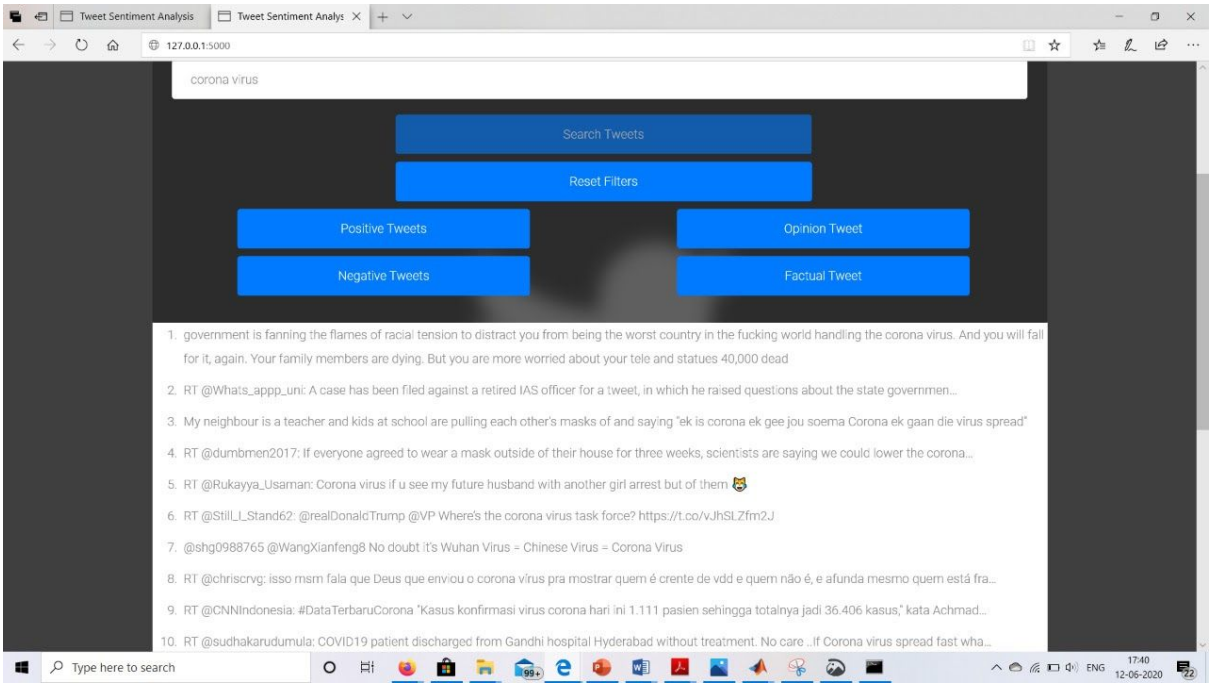

Fig 3. Negative search results for coronavirus

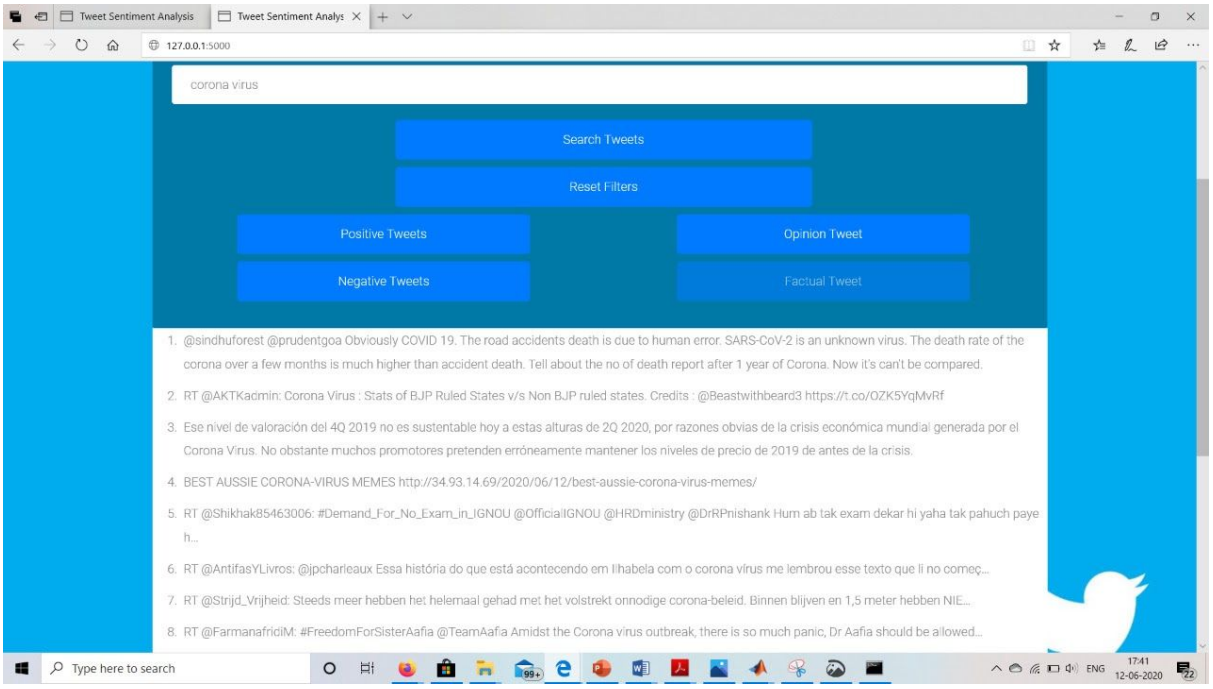

Fig 4. Factual search results for coronavirus

4.2 Source code

The project is open sourced on github and the complete source code and installation guide can be found on https://github.com/subham9499/Twitter-Sentiment-Analysis.

Chapter 5

Conclusion

5.1 Conclusion

The main objective was completed and the webapp will be able to classify tweets for the given hashtag into 4 different categories. The webapp takes input from the user and then accesses the twitter API to gather tweets with that hashtag and then use our model to classify based on the model selected by the user. The user has 4 options to select from for classifying the tweets related to a certain hashtag.

Background research took place for the project where we went through the various pre-existing models which helped us to decide the model which we will be using and also helped us take design of the website's frontend. We also came across various bugs and crashes which were solved using online information and by going through different projects.

5.2 Future Work

We are planning to add the following things to the project in future time:

1. Graphical Representation of data using circular and bar graphs.
2. Beautification of the site.
3. Optimizing the requirements file to overcome any bugs which we face.