FINAL PROJECT REPORT

SENTIMENT ANALYSIS OF COVID-19 TWEETS AND VISUALIZATION TWEETS

INDEX

12.1 Source Code

INTRODUCTION

1.1 Overview

Sentiment analysis is the technique by which computer can perceive the tone of the text written and some tasks can be automated according to the intention i.e., positive, negative or neutral without any human intervention. In this project, we have build a sentiment analysis model and a sentiment predictor model deployed using a Visualization Dashboard using Jupyter Notebook, Flask, HTML and CSS.

The following steps have been followed in the project making :-

- Scrap the tweets from Twitter to perform sentiment analysis
- Form sentiment analysis model using the nltk library
- Use the scrapped tweets and the analyzed sentiments as a training set to train the Random Forest classifier
- Use classifier to predict the sentiments of the tweets scrapped live in the Visualization Dashboard (constructed using HTML and CSS for the frontend and Flask for the backend)

1.2 Purpose

Here, we have built a tweets mining web application using HTML, CSS, Flask and Machine Learning models. The web application will function using the search query. The web app will surf the most recent and live tweets from Twitter and predict their sentiments using the Random Forest classifier model in the back end. The topic on which the tweets will be scrapped will be given by the user in the search box to analyze and predict the sentiments.

LITERATURE SURVEY

2.1 Existing Problem

The problem of internet trolling has been increasing day by day and with the exponential growth of the social media, it has been impossible for the people of prominent ranks in the administration to wrap their heads around the emotions of people.

Hence, there needs to be some technology which can help the concerned authorities to predict the sentiments of people and take decisions based on these predictions. Such a technology should be able to access the live sentiments of people that is it should be capable enough to update itself automatically with the live data.

2.2 Proposed Solution

By using Twitter API, Tweepy and csvWriter, we have created a csv file of 1,00,000 tweets related to COVID-19. These tweets are further cleaned and pre processed before the sentiment analysis.

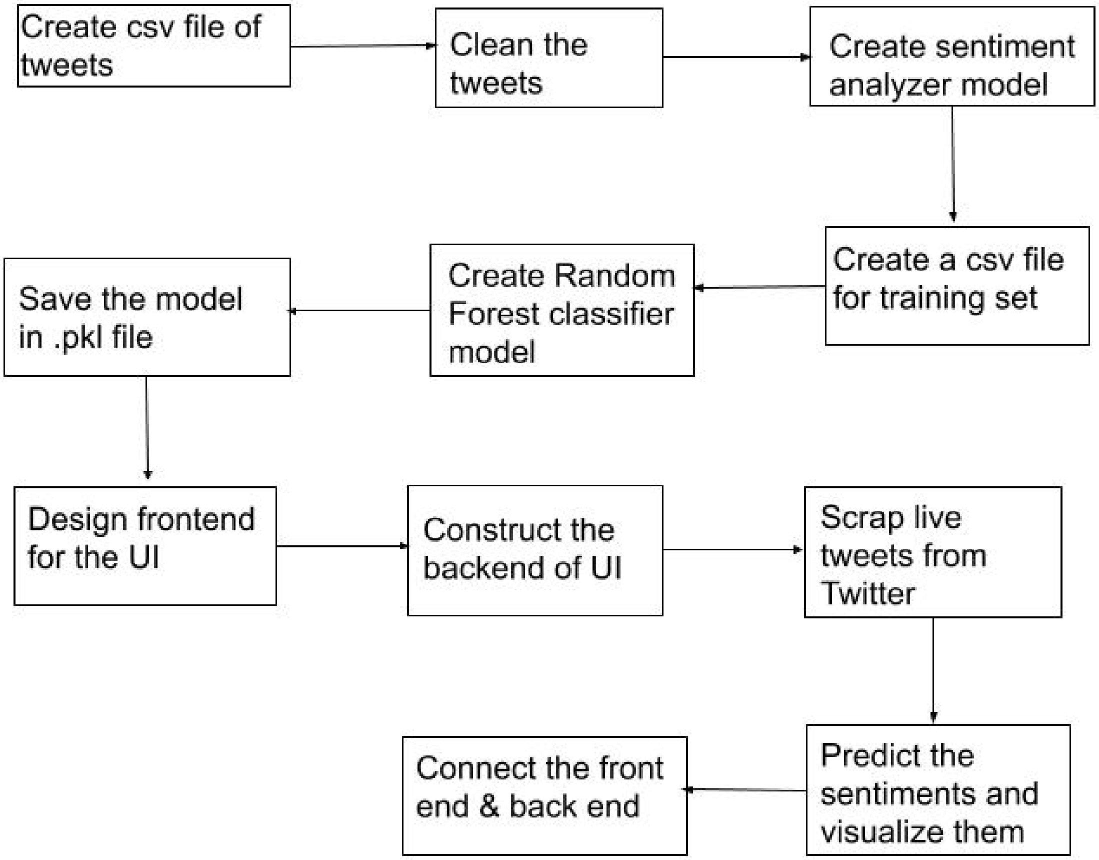
By using the nltk sentiment intensity analyzer, we analyze the sentiments of the tweets as positive, negative or neutral. This data along with the analyzed sentiments is converted into a csv file which will act as the training set for the classifier model.

By using Random Forest classifier available in the scikit-learn library, we train the classifier model to predict the sentiments of live tweets. As aforementioned, the tweets data along with their analyzed sentiments are used to train this classifier model. This classifier model is further saved as a .pkl file to be used in the back end of the dashboard.

By using the HTML and CSS, we design the front end of the Visualization Dashboard and Flask is used in the back end part. The classifier model saved earlier will be initialized and written again in the flask file which will then predict and classify the live tweets sentiments scrapped using Tweepy which is also included in the same code file. The user can visualize the sentiments using the pie chart which is made using the Google charts.

THEORITICAL ANALYSIS

3.1 Block Diagram


Create csv file of tweets
Clean the tweets
Create sentiment analyzer model
Create a csv file for training set
Create Random Forest classifier model
Save the model in .pkl file
Design frontend for the UI
Construct the backend of UI
Scrap live tweets from Twitter
Predict the sentiments and visualize them
Connect the front end & back end

3.2 Hardware / Software Designing

The software or programming languages used in the project are :-

1. Python (Jupyter Notebook) and its various libaries and modules like
   - Tweepy
   - Re
   - Nltk
   - String
   - Numpy
   - Pandas
   - Flask

- Pickle
- Csv
- Scikit-learn etc.

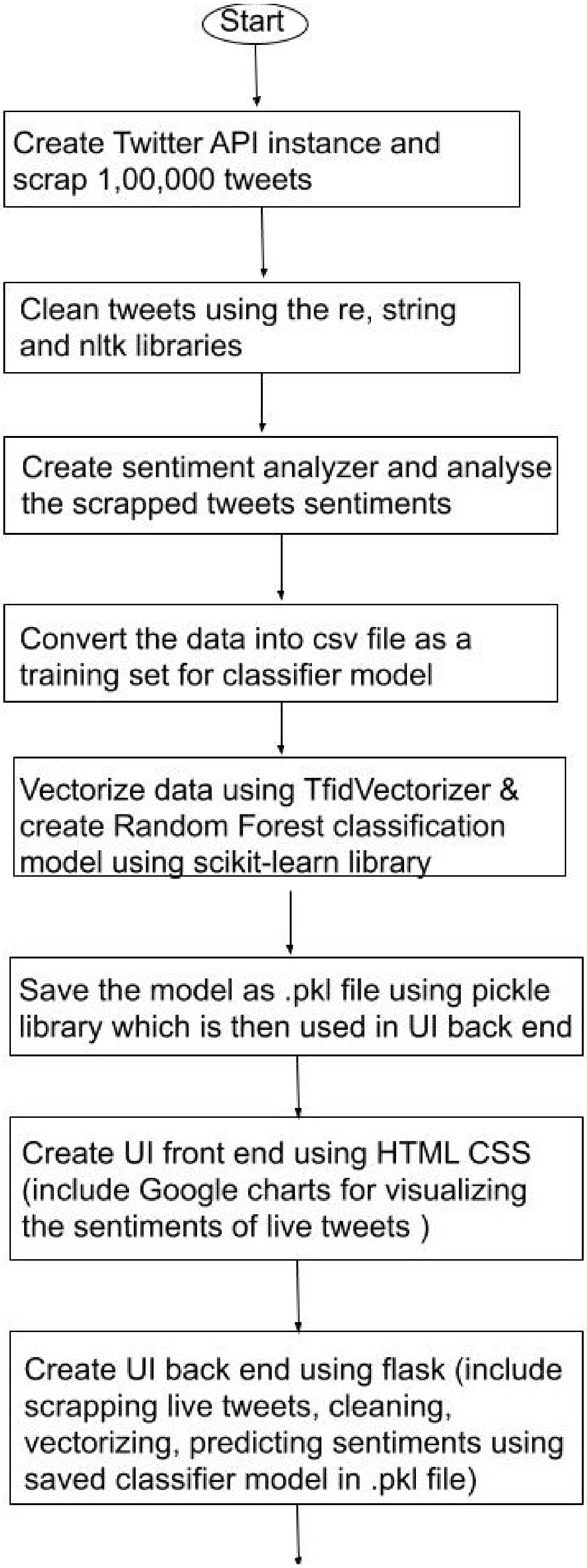
2. HTML
3. CSS
4. Google Charts

EXPERIMENTAL INVESTIGATIONS

In the process of developing this project, we have undergone many investigations related to the following topics.

1. Sentiment Analysis model
2. Vectorization
3. Classification through Random Forest classifier
4. Flask
5. Integrating flask with HTML
6. Google Charts

FLOWCHART



Start

Create Twitter API instance and scrap 1,00,000 tweets

Clean tweets using the re, string and nltk libraries

Create sentiment analyzer and analyse the scrapped tweets sentiments

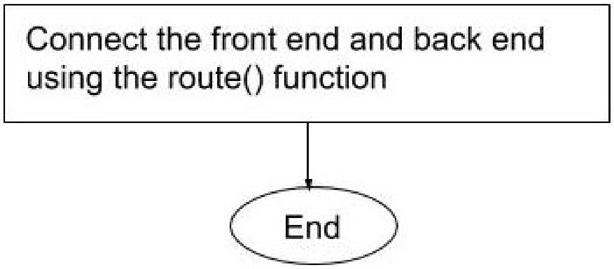
Convert the data into csv file as a training set for classifier model

Vectorize data using TfidVectorizer & create Random Forest classification model using scikit-learn library

Save the model as .pkl file using pickle library which is then used in UI back end

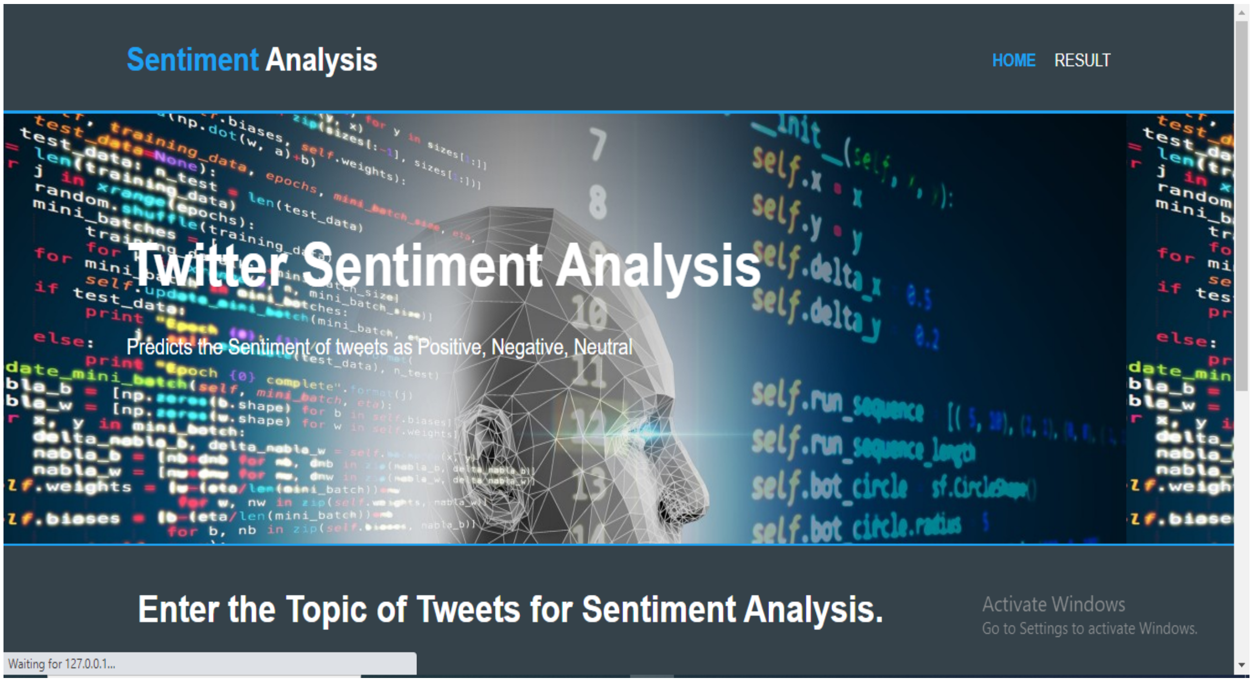
Create UI front end using HTML CSS (include Google charts for visualizing the sentiments of live tweets)

Create UI back end using flask (include scrapping live tweets, cleaning, vectorizing, predicting sentiments using saved classifier model in .pkl file)


Connect the front end and back end using the route() function
End

RESULT

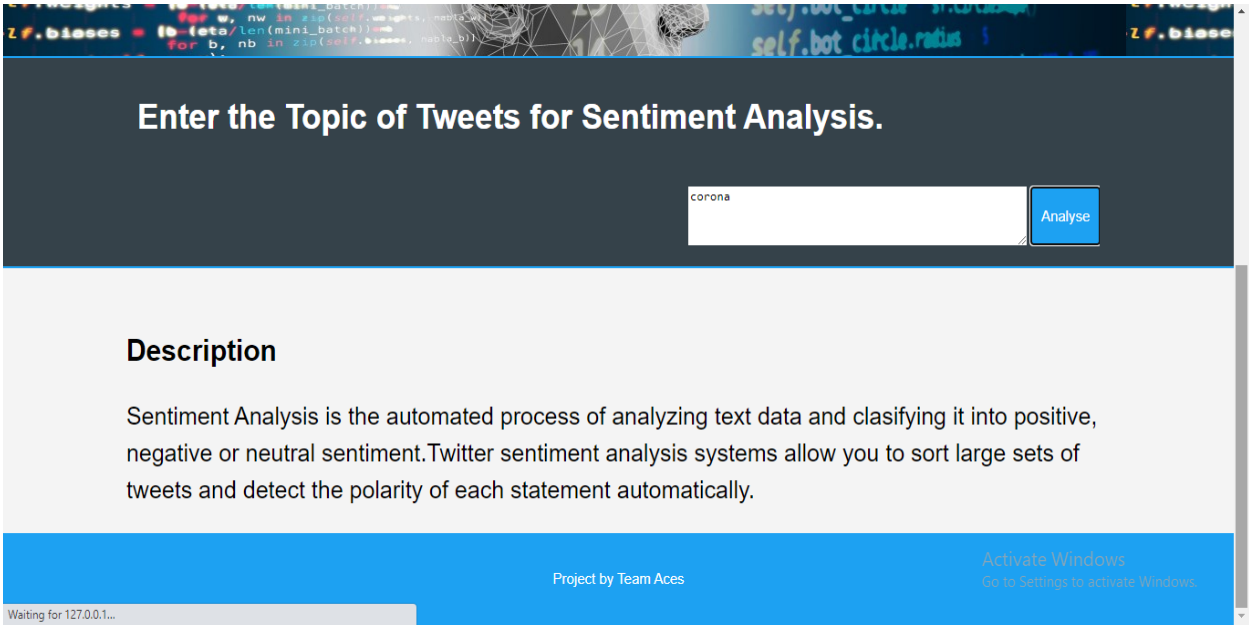
USER INTERFACE HOME PAGE :-



Twitter Sentiment Analysis

Predicts the Sentiment of tweets as Positive, Negative, Neutral

Enter the Topic of Tweets for Sentiment Analysis.



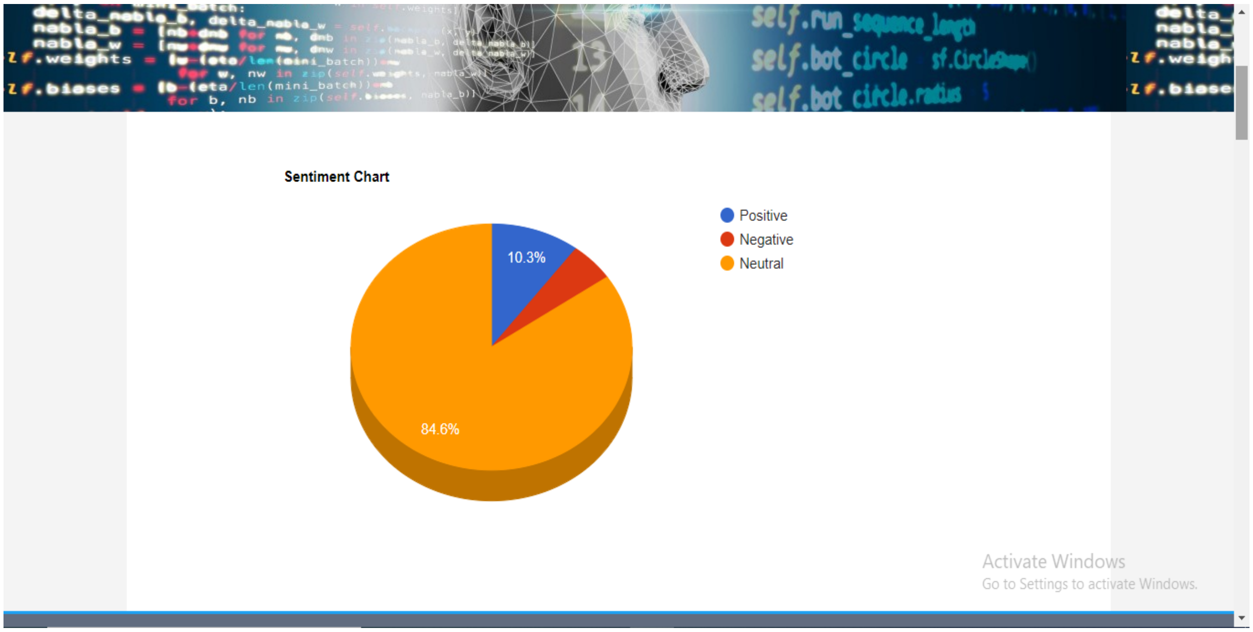
Enter the Topic of Tweets for Sentiment Analysis.

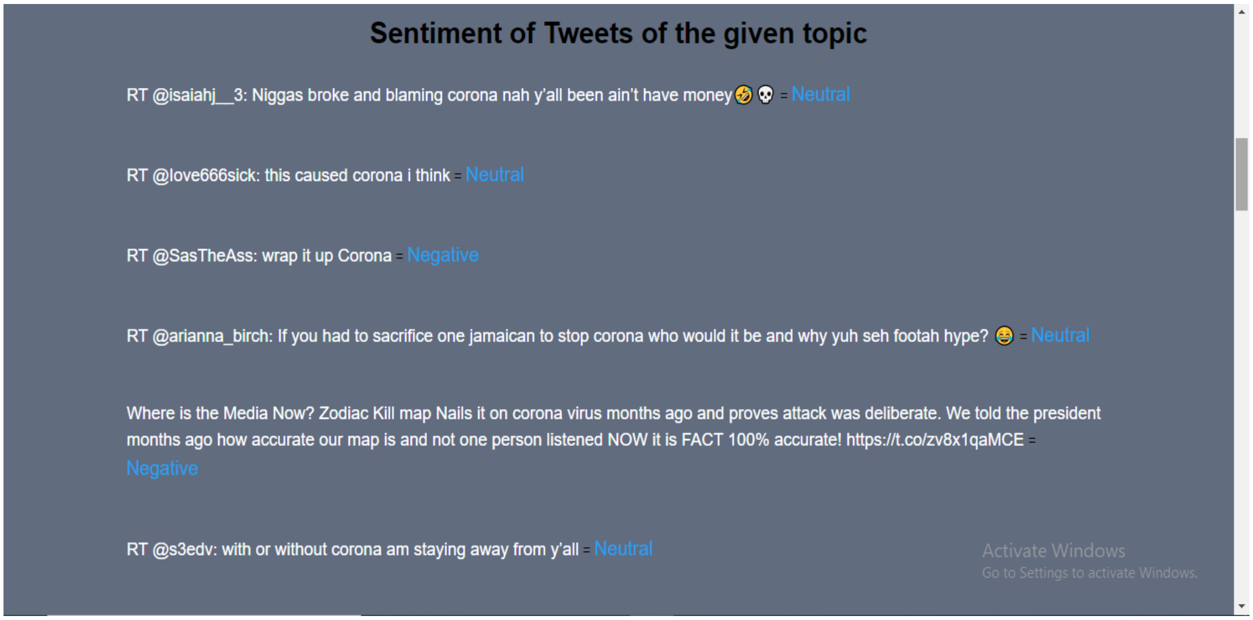
Description

Sentiment Analysis is the automated process of analyzing text data and clasifying it into positive, negative or neutral sentiment.Twitter sentiment analysis systems allow you to sort large sets of tweets and detect the polarity of each statement automatically.

Project by Team Aces

USER INTERFACE RESULT PAGE :-





Sentiment of Tweets of the given topic

RT @isaiahj__3: Niggas broke and blaming corona nah y'all been ain't have money🤣💀 = Neutral

RT @love666sick: this caused corona i think = Neutral

RT @SasTheAss: wrap it up Corona = Negative

RT @arianna_birch: If you had to sacrifice one jamaican to stop corona who would it be and why yuh seh footah hype? 😂 = Neutral

Where is the Media Now? Zodiac Kill map Nails it on corona virus months ago and proves attack was deliberate. We told the president months ago how accurate our map is and not one person listened NOW it is FACT 100% accurate! https://t.co/zv8x1qaMCE = Negative

RT @s3edv: with or without corona am staying away from y'all = Neutral

ADVANTAGES & DISADVANTAGES

The advantages of this project are as follows :-

- User can get to know about the live tweets & their sentiments
- Deployed classification model gives high accuracy
- User has the privilege to view the tweets & their predicted sentiments on any topic & on any number of topics using the search box available in the User Interface
- Software and programming languages used makes it easy to create and understand

The disadvantages of the project are listed below.

- The tweets of languages other than English are not included
- A certain sentiment may overpower the other two sometimes due to the words used in the tweets frequently

APPLICATIONS

The Twitter Sentiment Predictor finds its use in assessing the people's point of view on a certain topic before hand and then, the concerned authorities, organizations and companies can take decisions based on this assessment. With the privilege of scrapping of live tweets related to any topic and to any number of topics will make it easy to predict the emotions of people more accurately.

CONCLUSION

The project gives the knowledge of scrapping of live tweets from Twitter and how to analyze the sentiments of people using the sentiment intensity analyzer. It also makes one learn about the classification models to classify the sentiments of people based on the learning from the training set. A person can also gain the knowledge of integrating Flask in back end with HTML and CSS in the front end and how to use Google Charts to visualize the predicted results.

FUTURE SCOPE

The project can be further modified to fetch the tweets of languages other than English which can then be converted into English language using the language translation and then they can be further cleaned, analyzed and used to predict the sentiments using the classification model as used in the project. One only need to inculcate the step of language translation to convert the tweets of any language to English. This will make the project more complete in itself.

BIBLIOGRAPHY

Name of the Participant :- Ishika Singhal

College name :- Ajay Kumar Garg Engg. College

Problem Statement :- Sentiment Analysis of Covid-19 tweets and Visualization Dashboard

References taken :-

1. Twitter API : https://developer.twitter.com/en/docs

2. Tweets scrapping : https://towardsdatascience.com/how-to-scrape-tweets-from-twitter-59287e20f0f1

3. Sentiment Analysis model : https://www.kaggle.com/satanizer/covid-19-tweets-analysis

4. Random Forest classifier : https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

5. Flask : https://flask.palletsprojects.com/en/1.1.x/

6. Integration of flask with HTML : https://www.freecodecamp.org/news/how-to-build-a-web-application-using-flask-and-deploy-it-to-the-cloud-3551c985e492/

7. Google Charts : https://developers.google.com/chart

APPENDIX

Source Code for Flask and Sentiment Prediction (Back End) :
https://drive.google.com/file/d/1PIFy7ov9I_aBbkFHCNtmriIbdOnkjmsn/view?usp=sharing

Source code for HTML & CSS (Front End) :
https://drive.google.com/drive/folders/1qeBsRVoh1j2T0cJYm3Q1gpwMLB7S16Gh?usp=sharing
https://drive.google.com/drive/folders/1Xlg6mgvJi21A0inaVgtk84w9Et2NeW-W?usp=sharing

Source Code for Sentiment Analysis and Classification model :
https://drive.google.com/file/d/1fZQCoC164i5O8IJkw50r7PfOjuBc63kt/view?usp=sharing