Sentiment Analysis
On Tweeter Data

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 write short status updates of maximum length 280 characters. It is a rapidly expanding service with over 200 million registered users - out of which 100 million are active users and half of them log on twitter on a daily basis - generating nearly 250 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.

Introduction

In the past years, the young generation people are moving towards the social media like Google Plus, WhatsApp, Facebook, Twitter, etc. The social media is also revolving with those people to get them involved by making current trending insights concepts that is trending within a second. In the recent years, the people are exposing their social related issues through several social media by comments, reviews, posts, hashtags, emoji's, etc. which was followed by many people and those tweets become popular soon. Moreover, the social media is also bringing tremendous opportunity platform for businesses to connect with the consumers so easily. People rest on mostly user produced content like, comments, over online for making the decision. Example: if anyone has to buy a product or make a decision, they initiallysearch its reviews online, converse about it on social media. The content that is displayed for that product is mainly taken into the point as well as the discussion in the social media is also noticed and these made the way to make our business a success. To automate our analysis based on the reviews or comments in the social media by the people, for a sentimental analysis. Sentimental Analysis (SA) is introduced to the world to tell us the information is correct or wrong in each scenario using the social media tags. Thus, we can know about how world or people are reacting to every aspect currently going in the world.

What is sentiment analysis?

Sentiment Analysis is the process of 'computationally' determining whether a piece of writing is positive, negative or neutral. It's also known as **opinion mining**, deriving the opinion or attitude of a speaker.

Why sentiment analysis?

Business: In marketing field companies use it to develop their strategies, to understand customers' feelings towards products or brand, how people respond to their campaigns or product launches and why consumers don't buy some
products.

Politics: In political field, it is used to keep track of political view, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results as well!

Public Actions: Sentiment analysis also is used to monitor and analyse social phenomena, for the spotting of potentially dangerous situations and determining the general mood of the blogosphere.

Prerequisites:

Basic programming knowledge

Although Python is highly involved in this mini-project, it is not required to have a deep knowledge in the language, as long as you have basic programming knowledge.

Installed tools

For this program, we will need Python to be installed on the computer. We will be using the libraries twitter, nltk, re, csv, time, and json. You are likely to have to install the first two libraries. The rest already come with the Python interpreter. It doesn't hurt to check that they're up-to-date though.

Data set splitting concept

This is critical to fully understand the process pipeline. You only need to know the difference between Training and Test data sets, and in what context each one is used.

Basic RESTful API knowledge

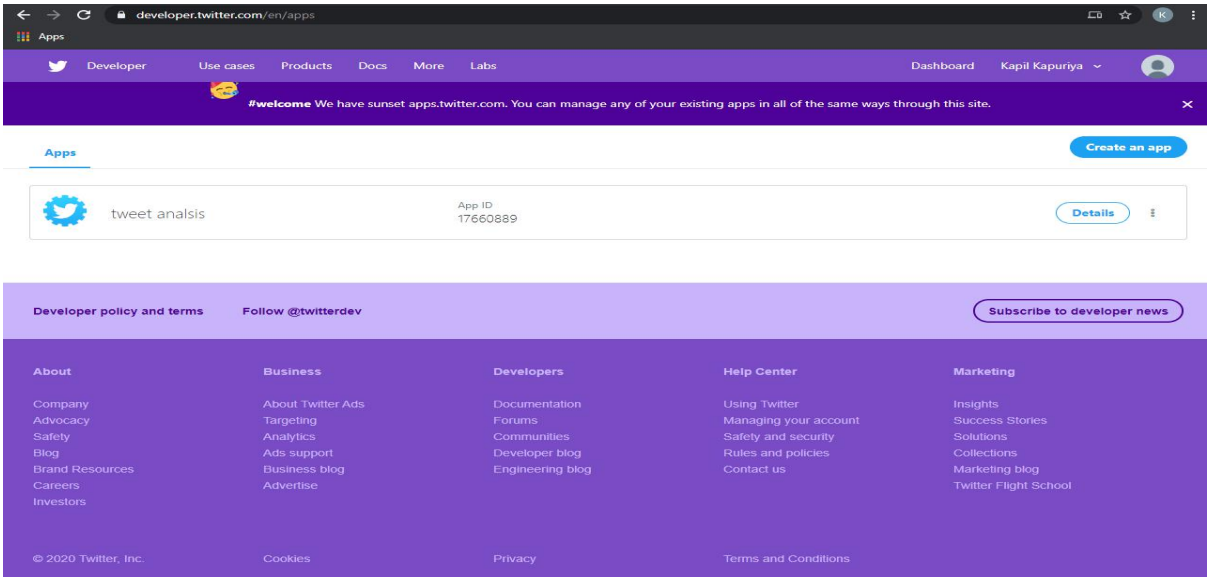
This is not crucial, but it could help. We will be using the Twitter API here and there in the code, making normal calls to the API and dealing with the JSON objects it returns.

How to perform twitter sentiment Analysis:

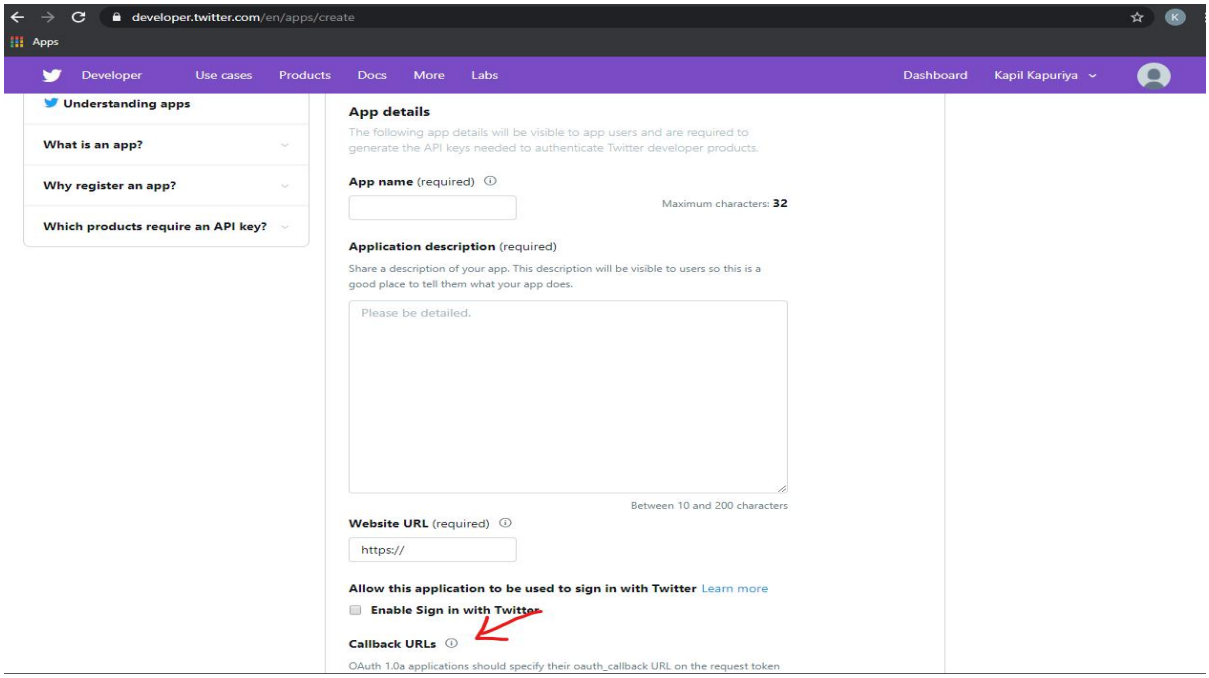
Create a ‘Consumer Key’, ‘Consumer Secret’, ‘Access token’ and ‘Access Token Secret’.

Few step for creat a key:

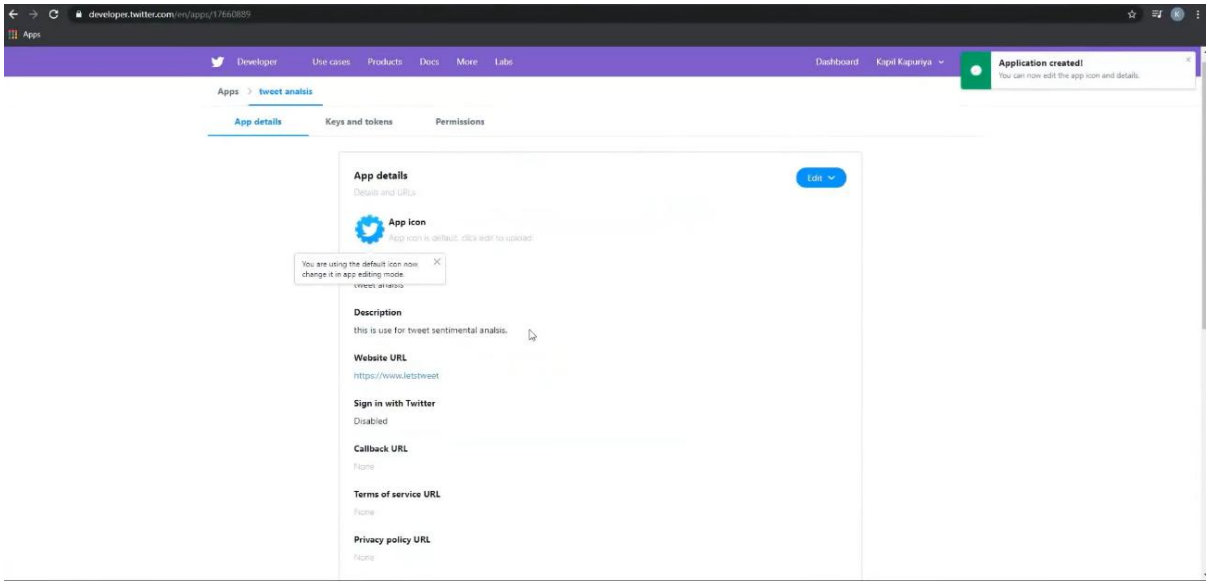
- Open this https://developer.twitter.com/en/apps and click the button: ‘Create New App’



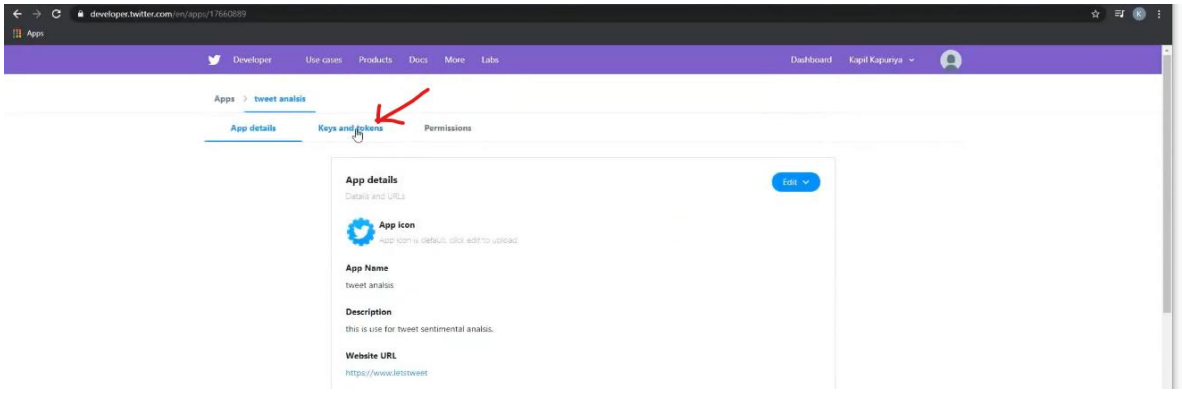
- Fill the application details. You can leave the callback url field empty.



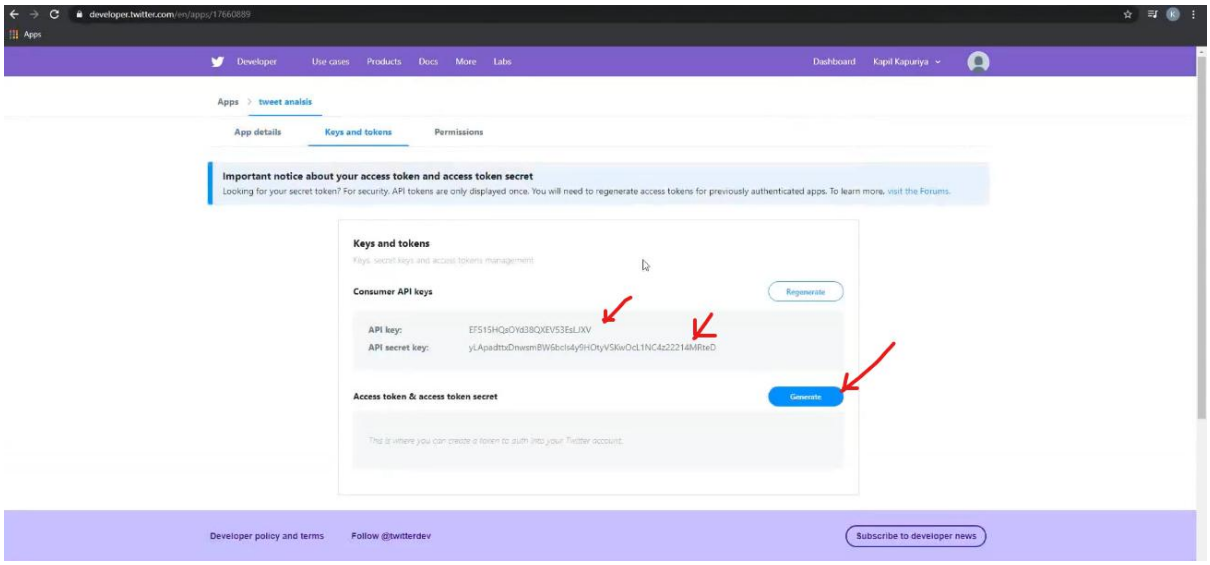
- Once the app is created, you will be redirected to the app page.

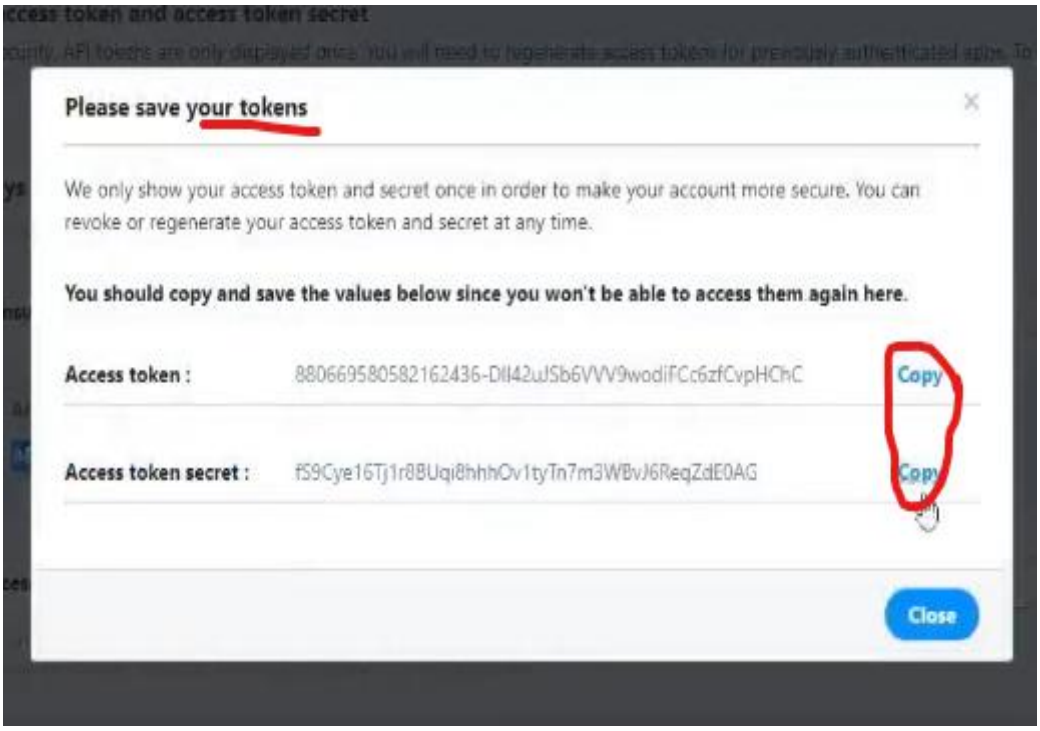


- Open the ‘Keys and Access Tokens’ tab.



- Copy ‘Consumer Key’, ‘Consumer Secret’, ‘Access token’ and ‘Access Token Secret’.





Twitter Sentiment Analysis Python:

Analysis of Twitter Sentiment using Python can be done through popular Python libraries like Tweepy and TextBlob.

Tweepy:

Tweepy, the Python client for the official Twitter API supports accessing Twitter via Basic Authentication and the newer method, OAuth. Twitter has stopped accepting Basic Authentication so OAuth is now the only way to use the Twitter API.

Tweety gives access to the well documented Twitter API. Tweepy makes it possible to get an object and use any method that the official Twitter API offers. The main Model classes in the Twitter API are Tweets, Users, Entities, and Places. Access to each returns a JSON-formatted response and traversing through information is very easy in Python.

TextBlob:

TextBlob, one of the popular Python libraries for processing textual data, stands on the NLTK. It works as a framework for almost all necessary tasks, we need in Basic NLP (Natural Language Processing). TextBlob has some advanced features like –

Sentiment Extraction

Spelling Correction

TextBlob is useful for Twitter Sentiment Analysis Python in the following ways:

Tokenization:

TextBlob can tokenize the text blocks into different sentences and words. This makes reading between the lines much easier.

Noun Phrases Extraction using TextBlob:

The noun is mostly used as an Entity in sentences. It is also one the most important NLP utility in Dependency Parsing. This is how different nouns are extracted from a sentence using TextBlob –

Part-of-Speech Tagging using TextBlob:

TextBlob is also used for tagging parts of speech with your sentences. For example –

```
In [9]: blob = TextBlob("Machine Learning Coding is best thing in the world")
        blob.tags

Out[9]: [('Machine', 'NNP'),
         ('Learning', 'NNP'),
         ('Coding', 'NNP'),
         ('is', 'VBZ'),
         ('best', 'JJS'),
         ('thing', 'NN'),
         ('in', 'IN'),
         ('the', 'DT'),
         ('world', 'NN')]
```

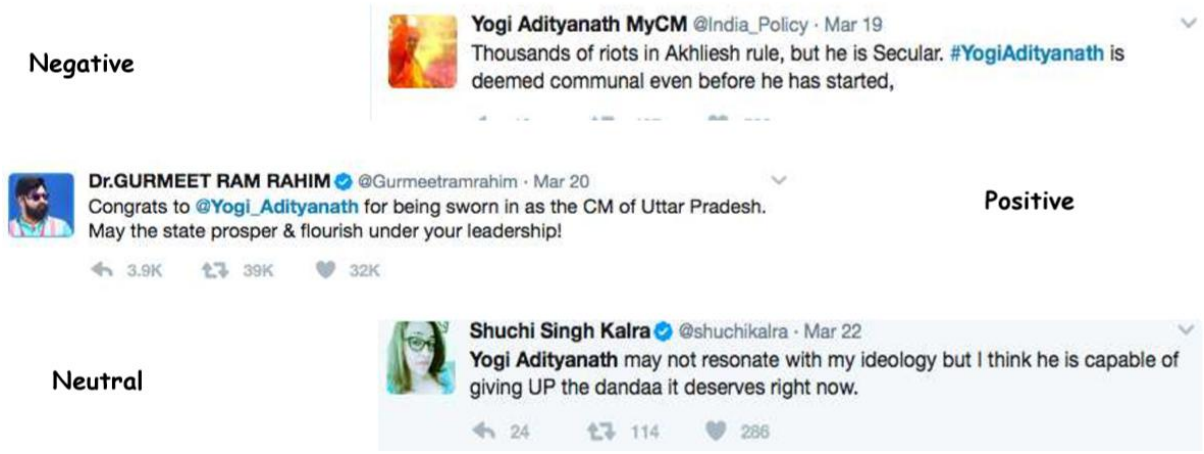
N-Grams with TextBlob:

Here N is basically a number. N-Gram is basically a chunk of words in the group. For a deep understanding of N-Gram, we may consider the following example-

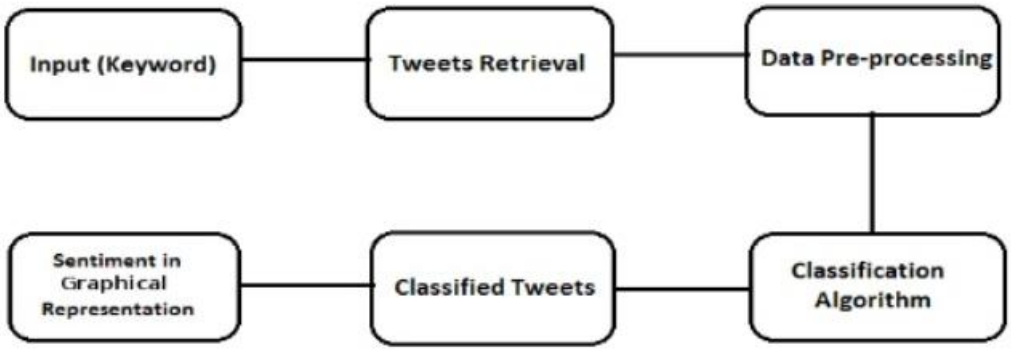
```
In [10]: blob = TextBlob("Machine Learning Coding is best thing in the world")
         blob.ngrams(n=3)

Out[10]: [WordList(['Machine', 'Learning', 'Coding']),
          WordList(['Learning', 'Coding', 'is']),
          WordList(['Coding', 'is', 'best']),
          WordList(['is', 'best', 'thing']),
          WordList(['best', 'thing', 'in']),
          WordList(['thing', 'in', 'the']),
          WordList(['in', 'the', 'world'])]
```

Classifying Sentiments :

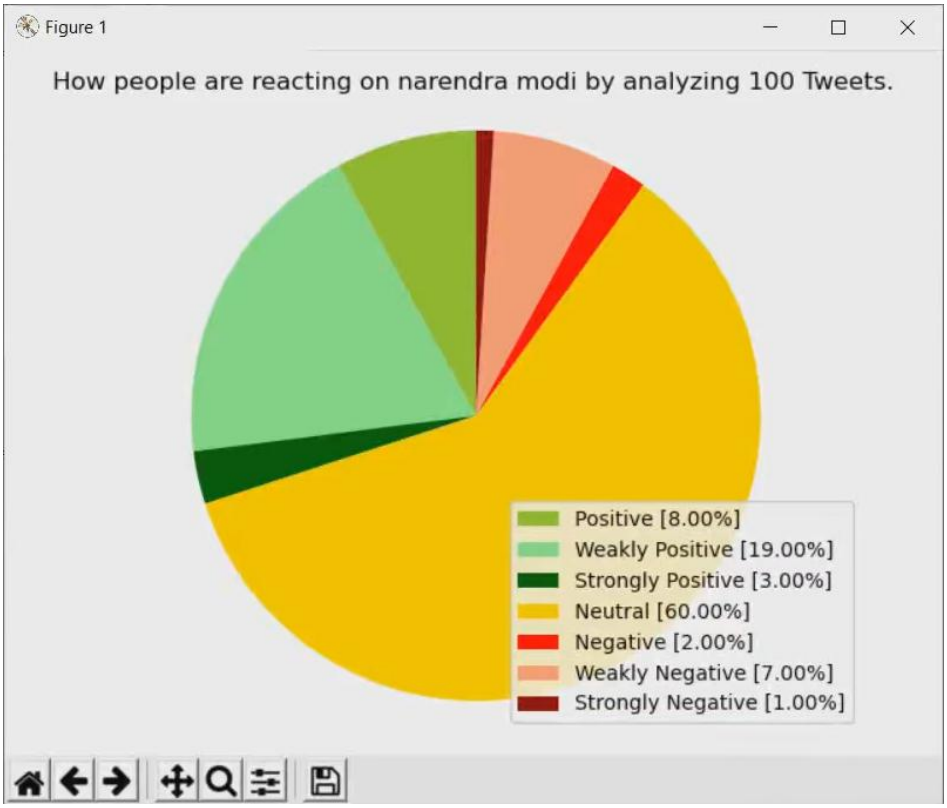


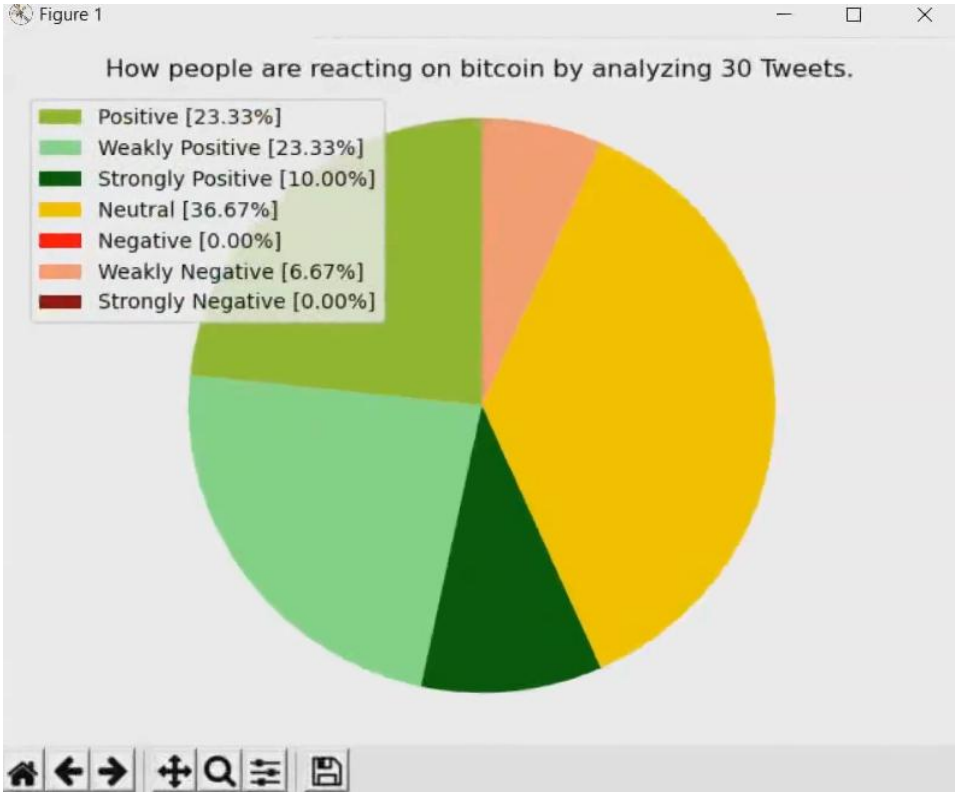
Working Process:



In this project, first I have collected the data and stored them in a file then read the data from the file and word tokenized them using python's Natural Language Toolkit Library(nltk). There is a list available of positive and negative words on internet that will have all the negative or positive words, stored the words in the dataframe and then compared words of my tweets with the respective words in positive and negative dataframes and generated net positive score and negative score for each(I have collected 200 tweets(as twitter maximum limit is 200 tweets at a time) about 'Narendra Modi' and 'Arvind Kejriwal' and then analyzing them and displaying the final result in a pie chart to determine whether people are speaking about whom Narendra Modi or Arvind Kejriwal).

Expected output:





References:

https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/

https://www.digitalvidya.com/blog/twitter-sentiment-analysis-introduction-and-techniques/

https://realpython.com/twitter-sentiment-python-docker-elasticsearch-kibana/