**PROJECT REPORT**

**on**

**Twitter Sentiment Analysis**

**(CSE VI Semester Mini project PCS-604)**

**2021-2022**

****

**Submitted to: Submitted by:**

Mr. Saumitro Saurabh Pandey

(CC-CSE-J-VI-Sem) Roll. No.: 1918675

CSE-J-VI-Sem

Session: 2021-2022

**DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

**GRAPHIC ERA HILL UNVERSITY, DEHRADUN**

**CERTIFICATEH**

### Certified that Mr. Saurabh Pandey (Roll No.- 1918675) has developed mini project on “Twitter sentiment Analysis” for the CS VI Semester Mini Project Lab (PCS-604) in Graphic Era Hill University, Dehradun. The project carried out by Students is their own work as best of my knowledge.

Date:29/6/22

(Mr. Saumitro)

**Class Co-ordinator**

**CSE-J-VI-Sem**

(CSE Department)

GEHU Dehradun

(Mr. Saumitro)

**Project Guide**

Resource Person

(CSE Department)

GEHU Dehradun

**ACKNOWLEDGMENT**

We would like to express our gratitude to The Almighty God, the most Beneficent and the most Merciful, for completion of project.

We wish to thank our parents for their continuing support and encouragement. We also wish to thank them for providing us with the opportunity to reach this far in our studies.

We would like to thank particularly our project Co-ordinator and Project Guide Mr. Saumitro for his patience, support and encouragement throughout the completion of this project and having faith in us.

At last but not the least We greatly indebted to all other persons who directly or indirectly helped us during this work.

**Mr. Saurabh Pandey**

**Roll No.- 1918675**

**CSE-J-VI-Sem**

**Session: 2021-2022**

**GEHU, Dehradun**

**TABLE OF CONTENTS**

**1. INTRODUCTION**

1.1 About Project

1.2 Problem Statement

1.3 Motivations

1.4 Idea

**2. PROJECT**

2.1 Requirement Analysis

2.2 Methodology

2.3 Visual Representation

**3. SNAPSHOT OF PROJECT**

3.1 Web Page

3.2 Result Page

**4. CONCLUSION**

**REFERENCE**

**INTRODUCTION**

* 1. **ABOUT PROJECT**

Sentiment analysis is the task of finding the opinions and affinity of people towards specific topics of interest. Be it a product or a movie, opinions of people matter, and it affects the decision-making process of people. The first thing a person does when he or she wants to buy a product on-line, is to see the kind of reviews and opinions that people have written. Social media such as Facebook, blogs, twitter have become a place where people post their opinions on certain topics. The sentiment of the tweets of a particular subject has multiple usage, including stock market analysis of a company, movie reviews, in psychology to analyze the mood of people that has a variety of applications, and so on.

Sentiments of tweets can be categorized into many categories like positive, negative and neutral. The data, being labeled by humans, has a lot of noise, and its hard to achieve good accuracy.

* 1. **PROBLEM STATEMENT**

Generating statistical information regarding emotions, sentiments out of analysis of user’s opinions from tweets, which can be used as an inference to understand how users feel thereby improving users experiences regarding.

**1.3 MOTIVATIONS**

Being extremely interested in everything having a relation with the Machine Learning, the independant project was a great occasion to give me the time to learn and confirm my interest for this field. The fact that we can make estimations, predictions and give the ability for machines to learn by themselves is both powerful and limitless in term of application possibilities. We can use Machine Learning in Finance, Medicine, almost everywhere. That’s why I decided to conduct my project around the Machine Learning.

**1.4 IDEA**

This project was motivated by my desire to investigate the sentiment analysis field of machine learning since it allows to approach natural language processing which is a very hot topic actually. Following my previous experience where it was about face recognition and Machine learning model . I applied the same idea with tweets and try to figure out which is positive or negative.

**PROJECT**

**2.1 REQUIREMENT ANALYSIS**

1. SOFTWARE REQUIREMENT -
2. Django
3. Pycharm
4. Visual-studio code
5. Latest web browser
6. HARDWARE REQUIREMENT - Windows and RAM(8-GB) minimum

**2.2 METHODOLOGY**

The sentiment analysis of Twitter data is an emerging field that needs much more attention.We use Tweepy an API to stream live tweets from Twitter.User based on his interest chooses a keyword and tweets containing that keyword are collected and stored into a list .Then we make it a labeled dataset using pandas dataframe. .Next we perform pre-processing to clean,remove unwanted text,characters out of the tweets.Using textblob and setting the sentiment fields accordingly.Thus our train data set without preprocessing is readyThen we train our classifier by fitting the train data to the classifier ,there after prediction of results over unseen test data set is made which there after provides us with the accuracy with which the classifier had predicted the outcomes.There after we present our results in a pictorial manner which is the best way to showcase results because of its easiness to understand information out of it.

**2.2.1 Extraction Of Data**

Tweets based on a keyword of user’s choice of interest have been collected using a twitter API known as Tweepy and stored into a list.This data set collected for sentiment analysis have tweets based on a keyword e.g.,tatasafari . Tweets mimicking various emotions as a dataset downloaded from twitter is used for sentiment analysis. In order to extract the opinion ,first of all data is selected and extracted from twitter in the form of tweets. After selecting the data set of the tweets, these tweets were cleaned from emoticons, unnecessary punctuation marks and a pandas dataframe was created to store this data in a specific transformed structure. In this structure, all the transformed tweets are in lowercase alphabets and are divided into different columns . The details about the steps adopted for the transformation of information are described in next subsections.

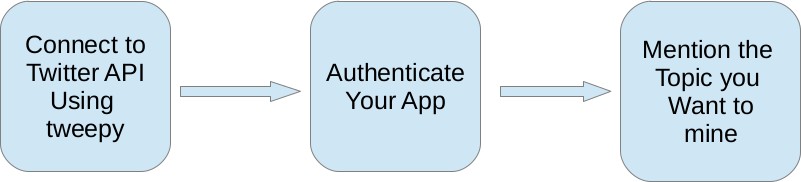


Fig. Extraction of data

**2.2.2 Pre-processing Of Data:**

Following are the Preprocessing steps that have been carried out: Removing Html tags and urls: Html tags and urls often have minimum sentiments thus they are removed from tweets. So, we create a function called cleantext() which cleans the tweet based on regular expressions.

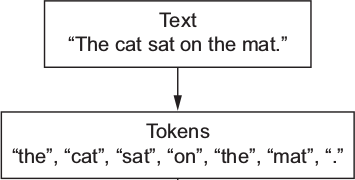
**Conversion to lowercase:**

To maintain uniformity all the tweets are converted to lowercase .This will benefit to avert inconsistency in data.Python provides a function called lower() to convert sentences to lower case.

After cleaning the tweets and converting them to lower case .We further perform some cleaning operations on the Tweets by passing the Tweets to TextBlob which creates a TextBlob object and perform the following operations on the tweets :-

**2.2.3 Tokenization:**

Tokenization is the process of converting text into tokens before transforming it into vectors. It is also easier to filter out unnecessary tokens. For example, a document into paragraphs or sentences into words. In this case we are tokenising the reviews into words



**2.2.4 Removing punctuations and special symbols:**

Apart from the considered set of emoticons punctuations and symbols like &,\,; are removed.

**2.2.5 Stop words removal:**

Stop words are the most commonly occuring words which are not relevant in the context of the data and do not contribute any deeper meaning to the phrase. In this case contain no sentiment.

"This is a sample sentence, showing off the stop words filtration."

['This', 'is', 'a', 'sample', 'sentence', ',', 'showing', 'off', 'the', 'stop', 'words', 'filtration', '.']

**After stop words removal:**

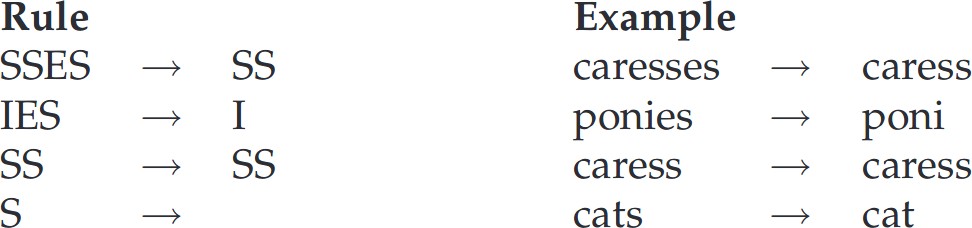
['This', 'sample', 'sentence', ',', 'showing', 'stop', 'words', 'filtration', '.']

**2.2.6 Stemming and Lemmatization:**

Sentences are always narrated in tenses,singular and plural forms making most words accompany with -ing,-ed,es and ies. Therefore,extracting the root word will suffice to identify sentiment behind the text.

Base forms are the skeleton for grammar stemming and lemmatization reduces inflectional forms and derivational forms to common base forms .

Example: Cats is reduced to cat ,ponies is reduced to poni.

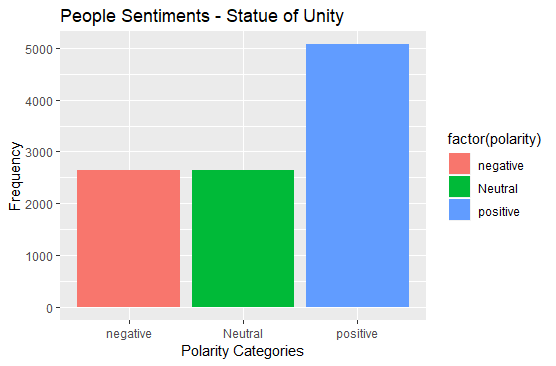
Stemming is a crude way of reducing terms to their root, by just defining rules of chopping off some characters at the end of the word, and hopefully, gets good results most of the time. The goal of both stemming and lemmatization is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.With that being said, stemming/lemmatizing helps us reduce the number of overall terms to certain “root” terms.

**2.2.7 Classification of Tweets:**

Now to classify tweets into positive , negative and neutral we used TextBlob.sentiment.polarity and TextBlob.sentiment.subjectivity function which gives us the sentiment score .On the pandas dataframe we applied getPolarity() and getSubjectivity() which used polarity and subjectivity operation of TextBlob library . The polarity function is used for generating the sentiment scores for each tweet and subjectivity is used to measure the personal opinion of the tweet users . The Sentiment Score of each tweet can be positive or negative or neutral on the basis of opinions of public .

**Sentiment score :**

The Senti-Strength Scale is used for measuring the polarity(Sentiment) score for positive ,negative and neutral tweets in which +1 means Positive , -1 for Negative and 0 for Neutral . The sentiment score is a more specific numerical illustration of the Opinions of people.



We create new columns named subjectivity and polarity in our pandas dataframe and stores the subjectivity scores and polarity scores in respective columns.

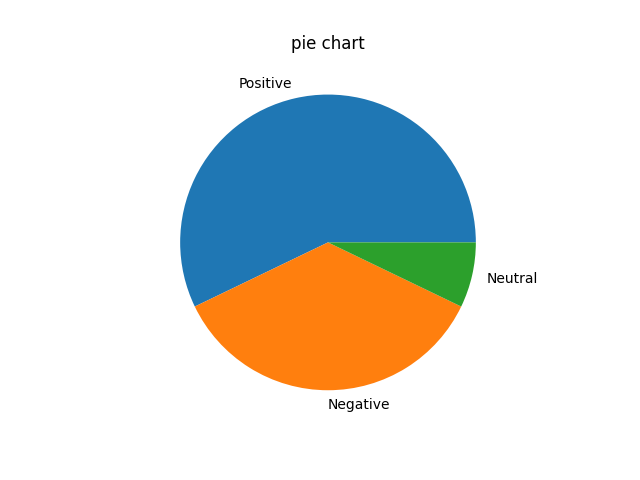
**2.2.8 Marking the tweets:**

We create a function getAnalysis() which marks the tweets as positive , negative and neutral based polarity scores and stores it into a new column named Analysis.

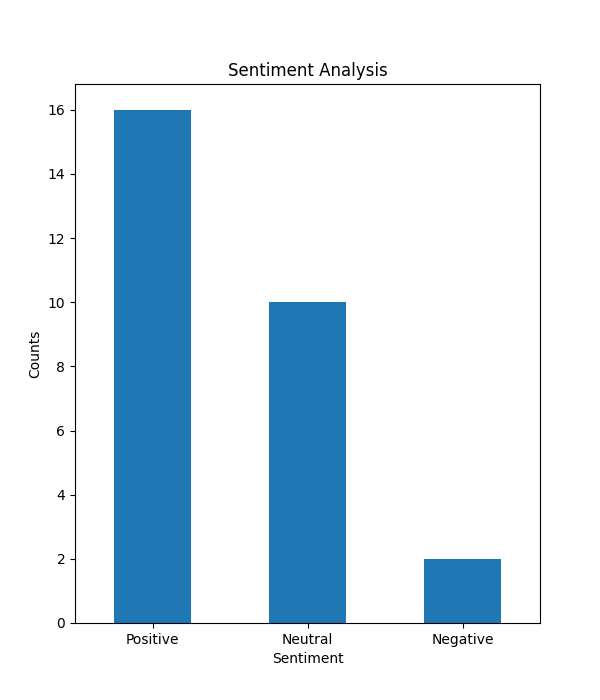
With the help of the Analysis column we create a 3 list ptweets , ntweets and neutweets whic contains the positive , negative and neutral tweets respectively.

**2.3 Visual Representation**

**Pie chart:**

****

**Bar graph:**



**Word-Cloud:**

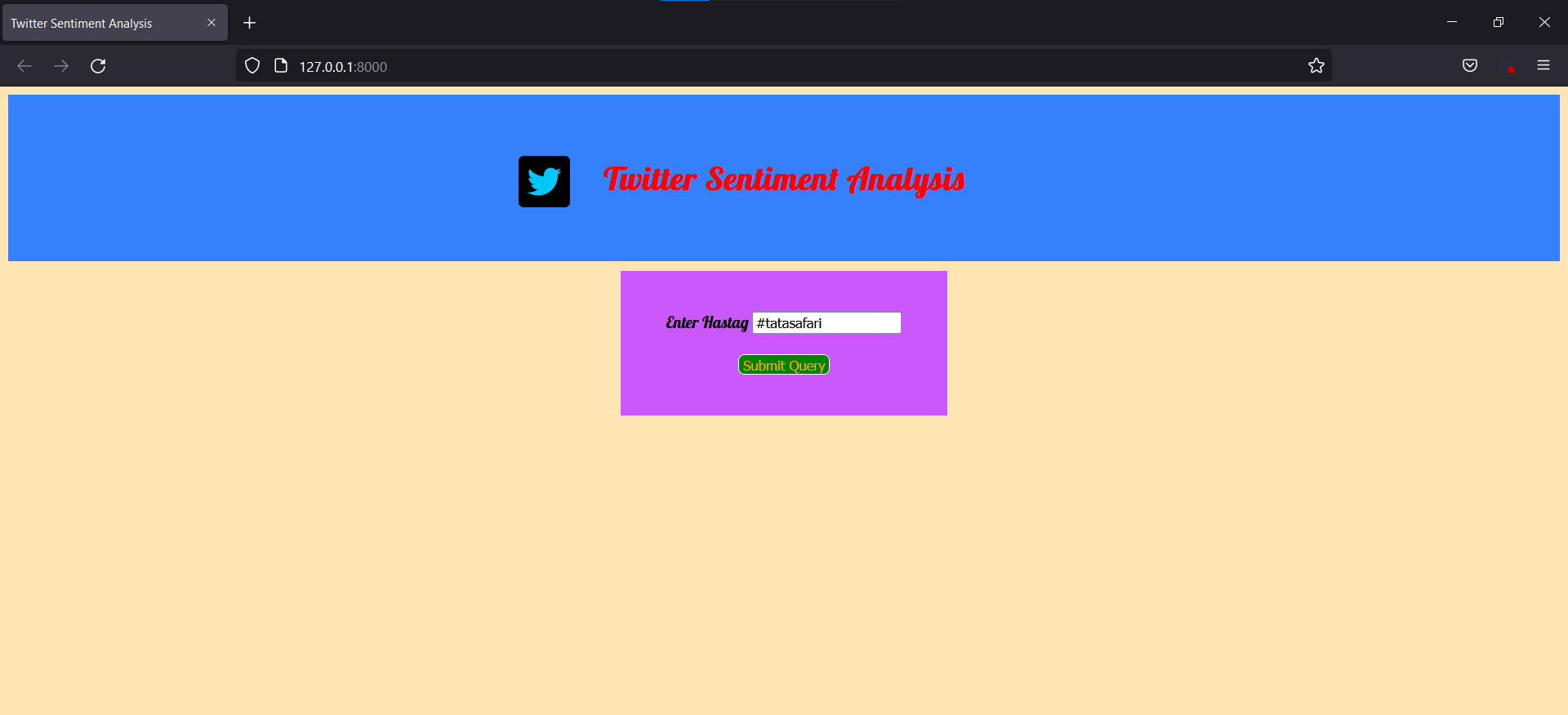


**Negative Word-Cloud:**

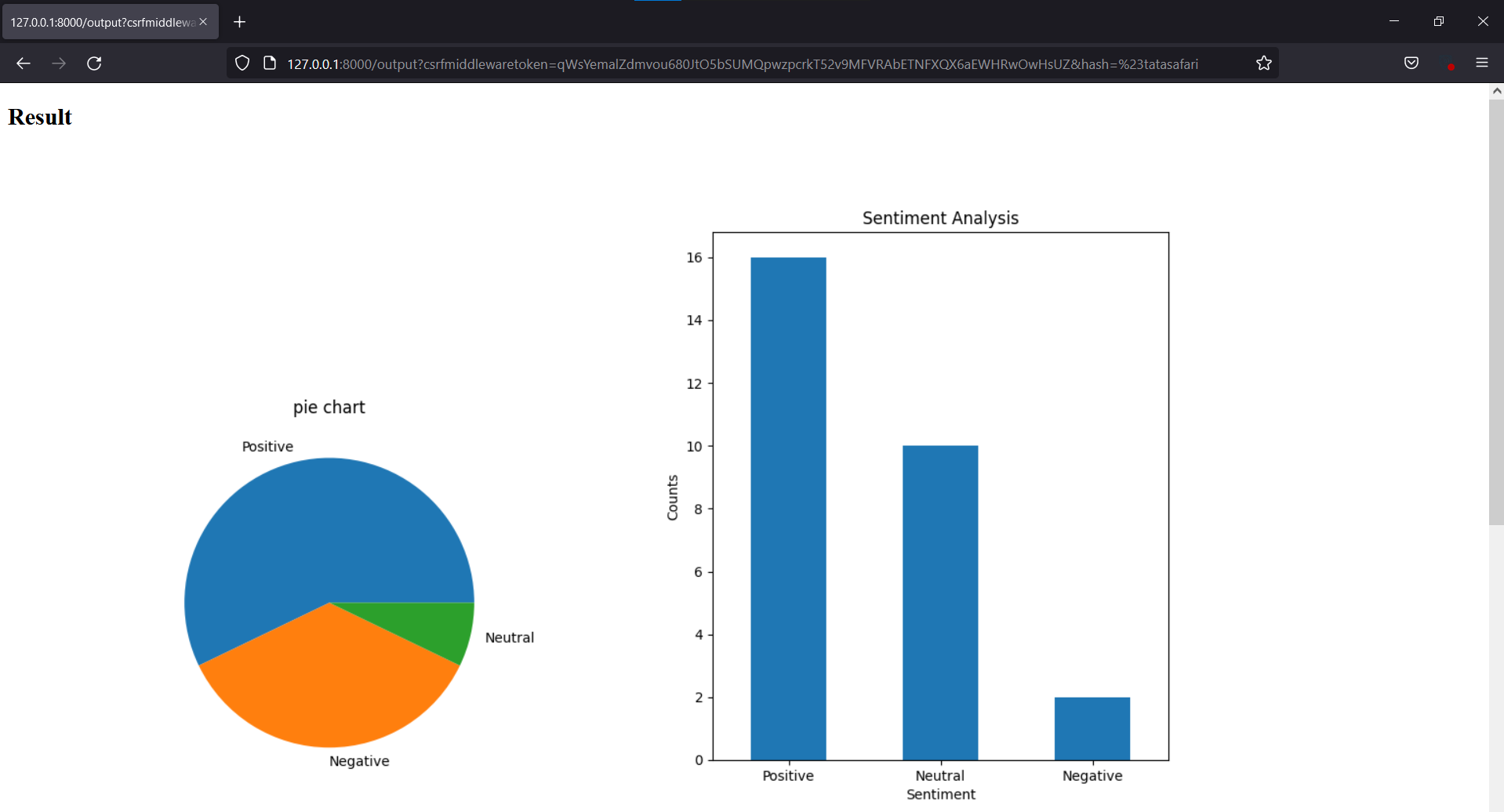


**SNAPSHOT OF PROJECT**

**3.1 WEB PAGE**



**3.2 Result**

**4. CONCLUSION**

We furnished results for Sentiment and Emotional Analysis on twitter data . With the help of TextBlob library which is based on Naive-bayes Classifier.We firmly conclude that implementing sentiment analysis using the TextBlob library will help in deeper understanding of textual data which can essentially serve a potential platform for businesses

**REFERENCE**

1. Shiv Dhar, Suyog Pednekar, Kishan Borad- Methods for Sentiment Analysis Computer Engineering, VIVA Institute of Technology, University of Mumbai, India.
2. <https://youtu.be/ujId4ipkBio>
3. <https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>
4. Ravinder Ahujaa, Aakarsha Chuga, Shruti Kohlia,Shaurya Guptaa,Pratyush Ahujaa- The Impact of Features Extraction on the Sentiment Analysis Noida, India.
5. Machine Learning Tom M. Mitchell