HINDUSTAN
INSTITUTE OF TECHNOLOGY & SCIENCE
(DEEMED TO BE UNIVERSITY)
CHENNAI

PROJECT REPORT

On

SENTIMENT ANALYSIS TOOL USING NATURAL LANGUAGE PROCESSING

BACHELOR OF TECHNOLOGY

(Information Technology)

Submitted by

Jeyavasan.T (20134034)

Name of the team members

Jeyavasan.T (20134034)

Venkatraman.M (20134018)

Submitted to

Dr. Saranya S

Assistant Professor

IT Department, HITS

IV SEMESTER

DESIGN PROJECT(ITB4243)

DEPARTMENT OF INFORMATION TECHNOLOGY

HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE

CHENNAI – 603 103

MAY 2022

HINDUSTAN
INSTITUTE OF TECHNOLOGY & SCIENCE
(DEEMED TO BE UNIVERSITY)
CHENNAI

BONAFIDE CERTIFICATE:

Certified that this Design project report "**SENTIMENT ANALYSIS TOOL USING NATURAL LANGUAGE PROCESSING**" is the bonafide work of Jeyavasan.T (20134034), Venkatraman.M (20134018) Team members who carried out the Design project work under my supervision during the academic year 2021-2022.

SIGNATURE

Supervisor
Dr. Saranya S
Assistant Professor
IT Department, HITS

INTERNAL EXAMINER
Name:
Designation:

EXTERNAL EXAMINER
Name:
Designation:

Project Viva-Voce conducted on _____

HINDUSTAN
INSTITUTE OF TECHNOLOGY & SCIENCE
(DEEMED TO BE UNIVERSITY)
CHENNAI

ACKNOWLEDGEMENT:

At first, we would like to thank Almighty God for the idea and opportunity to work on this project. We thank Dr. S. Saranya Assistant Professor, Department of Information Technology for their strong support and encouragement for the project "**SENTIMENT ANALYSIS TOOL USING NATURAL LANGUAGE PROCESSING**".

We thank all the faculty members and technical staff of the Department for their support and suggestions of the design project development.

Our Team Members:

Jeyavasan.T (20134034)

Venkatraman.M (20134018)

TABLE OF CONTENTS

1) ABSTRACT

An online reputation is one of a company's most valuable assets. Dealing with a negative review on social media can be costly if it's not handled properly. Sentiment analysis lets you monitor what's being said about your product or service, as well as track your progress on social media collecting real time reviews about products and services. The use of sentiment analysis on Twitter provides a large number of fascinating possibilities. The ability to analyze tweets in real-time today, and to determine the sentiment to follow behind each message, has added a completely new dimension to social media monitoring. Improvements of this project over the existing project is to collect data only from twitter to reduce unwanted confusions in the raw data which might lead to false results. By looking at social media interactions and what they reveal about consumers behind the screens, far beyond the surface level of the number of likes, comments, and shares, sentiment analysis will try to fully comprehend and comprehend the importance of social media interactions. A variety of groups will continue to make use of this technology, including brands, public figures, governments, NGOs, and educational institutions.

Existing Sentiment Analysis tools are meant for the usage of Data Scientists with a complex usage where one has to learn to use the program, but in this project, there is a user-friendly interface that can be used by anyone without any prior knowledge. This project will be showing user product reviews and respective sentiments with a good UI and UX for convenience of the user. Additional feature in this project is phrase level sentiment analysis in which it analyses the user input phrase and predicts the sentime

2) INTRODUCTION:

In this paper we discuss on how to implement Sentiment Analysis Tool to track user opinions and real time review of our products by scraping data from Twitter and analyzing it. The front-end is designed using HTML CSS and the back-end is implemented with Python Flask micro web-framework which is very light weight and easy to use. With the help of sentiment analysis modules called Text blob we identify the sentiments of the tweet. Text blob provides a sentiment score which will be helpful to segregate the tweets on the basis of Negative, Positive and Neutral. On the phrase level sentiment analysis section, we take the give phrase by user as an input and predict the sentiment for it and shows result sentiment. Sentiment analysis result is obtained in the form of a score called Sentiment score which lies between -1 and 1. Where -1 represents highest score in Negative, 0 represents Neutral and 1 represents the highest score in positive.

3) PROBLEM STATEMENT:

Besides reviewing system on e-commerce platforms such as Amazon, there is no clear way of analyzing feelings and emotions of consumers towards a product manufactured by companies. This may lead to customer dissatisfaction, bad reputation or angry customers for the company and the products. So, with help of Natural language processing one can determine close to accurate results of what people might actually feel, with help of real-time reviews from Twitter where people often register their opinions and reviews of products they use, this project is tool that segregates the tweets of query searched based on their opinions with a user-friendly interface.

For companies to enhance their products according to the need of their customers and consumers, for customers to obtain better customer satisfaction. Sentiment analysis provide better clarity in how a product is received compared to rating systems, hence this tool is created which will be efficient and also user friendly for the users. Segregation of reviews/tweets based on their sentiment drastically helps in identifying the negative reviews and overall analytics of emotions showed towards the product or company.

4) PROPOSED SYSTEM:

An idea which can overcome the disadvantages being faced by traditional survey method to get people opinions, to develop a Machine Learning Model by training the model to categorize the tweets based on sentiment of the tweet and make the model as accurate as possible, first the user will give input i.e. the keyword for extracting the tweets and then the extracted tweets will be categorized by the Machine Learning Model which will be either positive ,negative or neutral tweet and then the output will be displayed in graphical manner for better understanding of the results.

Advantages of Proposed System:

- There is no need to manually start a survey because in twitter there are already available tweets which are opinions of the people.
- There is no need to manually take tweets one by one.
- The user just has to download the application.
- There are no external hardware components required.
- No need to create a dataset for tweets, since live tweets can be extracted directly.

Limitations of the proposed system:

- Only English language is supported, other language support is yet to be developed.
- Only limited to one social media that is twitter, other social media support yet to be added.

5) LITERATURE SURVEY:

The following papers and journals have been referred for better understanding and idea creating process of this project.

AUTHOR'S NAME	TITLE	JOURNAL NAME	PUBLICATION YEAR	OBJECTIVES	LIMITATIONS
Nabeela Altrabsheh, Mihaela Cocea,Sanaz Fallahkhair.	Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback.	IEEE	2014	To identify students' issues with lectures using Sentiment analysis	Takes data from only students feedback and use to generate an analysis.
Farha Nausheen ,Sayyada Hajera Begum	Sentiment analysis to predict election results using Python.	IEEE	2018	To Predict Election Results Using Python	Only takes election data for an input
Toufique Ahmed, Amiangshu Bosu, Anindya Iqbal, Shahram Rahimi	SentiCR: A customized sentiment analysis tool for code review interactions.	IEEE	2017	Sentiment analysis for Code review interactions in online forums	The input filtering can be very hard since coding forum's use code snippets which acts as stop words
Xiaobo Zhang ,Qingsong Yu	Hotel reviews sentiment analysis based on word vector clustering.	IEEE	2016	Hotel reviews sentiment analysis based on word vector clustering.	Only works based on word vector clusters

Meylan Wongkar; Apriandy Angdresey	Sentiment Analysis Using Naive Bayes Algorithm of The Data Crawler: Twitter	IEEE	2019	To obtain data with data crawler and analyze sentiment	Data crawler is not very resource efficient
V. Prakruthi; D. Sindhu; Dr. S. Anupama Kumar	Real Time Sentiment Analysis of Twitter Posts	IEEE	2018	To take real-time twitter posts and analyze the sentiment.	The latency is very high so most recent posts will not be added.
Rasika Wagh; Payal Punde	Survey on Sentiment Analysis using Twitter Dataset	IEEE	2018	To analyze sentiments of Surveys on dataset	Real time tweets not obtained, uses previously stored datasets.
Md. Rakibul Hasan; Maisha Maliha; M. Arifuzzaman	Sentiment Analysis with NLP on Twitter Data	IEEE	2019	To analyze sentiment of twitter data	Real time tweets not obtained, uses previously stored datasets.
Adyan Marendra Ramadhani; Hong Soon Goo	Twitter sentiment analysis using deep learning methods	IEEE	2018	To analyze sentiment using deep learning techniques	Not user-friendly interface complex usage
C.R. Nirmala; G.M. Roopa; K.R. Naveen Kumar	Data analysis for unemployment crisis	IEEE	2019	To analyze sentiments for unemployment crisis	Very specific data such as Unemployment crisis data only obtained.

6) OBJECTIVES AND SCOPE OF THE PROJECT:

6.1) OBJECTIVES:

- This Project will be useful for companies to identify issues on their product and rectify as soon as possible after analysing the product reviews from Twitter.

- This project is based on web so it will be accessible across all platforms and can be used with ease.

- Product producers can predict if their product will be well received when they release a promotion for the product and market further with the newly collected data.

- Information is wealth as they say, it is crucial to collect user data in all forms for improving your products and services with help of the power of python and machine learning we collect information.

- This project uses python, docker and cloud service for hosting, so it is robust libraries, flexible, fast and highly reliable.

- Collecting data about Positive, Negative and Neutral reviews of a product and plotting graphs using data points.

- Show the required data in a user-friendly UI.

6.2) SCOPE OF THE PROJECT:

As a result of a deeper and better understanding of the feelings, emotions, and sentiments of the key, high-value audiences of a brand or organization, members of these audiences will increasingly receive experiences and messages that are customized and directly connected to their wants and needs. In order to further segment markets, organizations can take into account audience members' actual feelings about the brand or their use of social media. Instead of segmenting based on age, gender, income, and other surface demographics, groups can segment by how they feel about the brand or how they use social media.

Despite the fact that some people shudder at the thought of companies learning more about them, more precise targeting means in the near future, we will no longer have to scratch our heads and wonder why we are seeing advertisements for products we would never think of buying. Therefore, with the sweep of advertising tactics, spraying products all over the place and exhausting our brains with irrelevant ads is all almost dead, and we will soon see a time when all marketing messages we see will be relevant and useful to us. This is an important goal to be achieved through sentiment analysis, and one of the major sections as part of the process.

7) REQUIREMENT SPECIFICATIONS

7.1) SOFTWARE REQUIREMENT SPECIFICATION:

- Python
- Docker (for deployment and containerization)
- HTML, CSS
- Text blob python module
- Regular expression (regex) python module (Re)
- Flask Micro-Web framework
- Amazon Web Services for cloud (for deploying the docker container)
- NLP package.
- Twitter API (Elevated Access)
- Twitter developer account
- Tweepy Module

7.2) HARDWARE REQUIREMENT SPECIFICATION:

- Intel core i7 processors with 3.2 GHz (4.6 GHz turbo).
- 6 physical / 12 logical cores
- 32 GB of memory
- Network Bandwidth 10 Gbps
- EBS Bandwidth 8,000 Mbps

8) SYSTEM DESIGN AND METHODOLOGY:

MODULES OF THE PROJECT:

This project consists of the following four modules:

Module 1.Collection of Data:

Collecting and storing data from Twitter as it is the most used social media for product reviews and complains. Users use twitter as a platform for rising their concerns.

Module 2.Filtering of Data:

Raw data from twitter might contain some unwanted substrings such as '@' mentions and Words that doesn't suggest any emotions such as "is" "was".

Module 3.Identifying the sentiment:

Identifying the sentiment based on the text provided by implementing NLP (Natural Language Processing) with use of Python modules.

Module 4.Showcasing the tweets with its respective sentiment:

Take the sentiment analysis data identified and use it for generating graphs other representation of data in the front-end.

FLOWCHARTS:

Following two diagrams shows the connections between different stages of a process or parts of a system and workflow.

Real-time tweets based sentiment analysis option works on the basis of the following process.

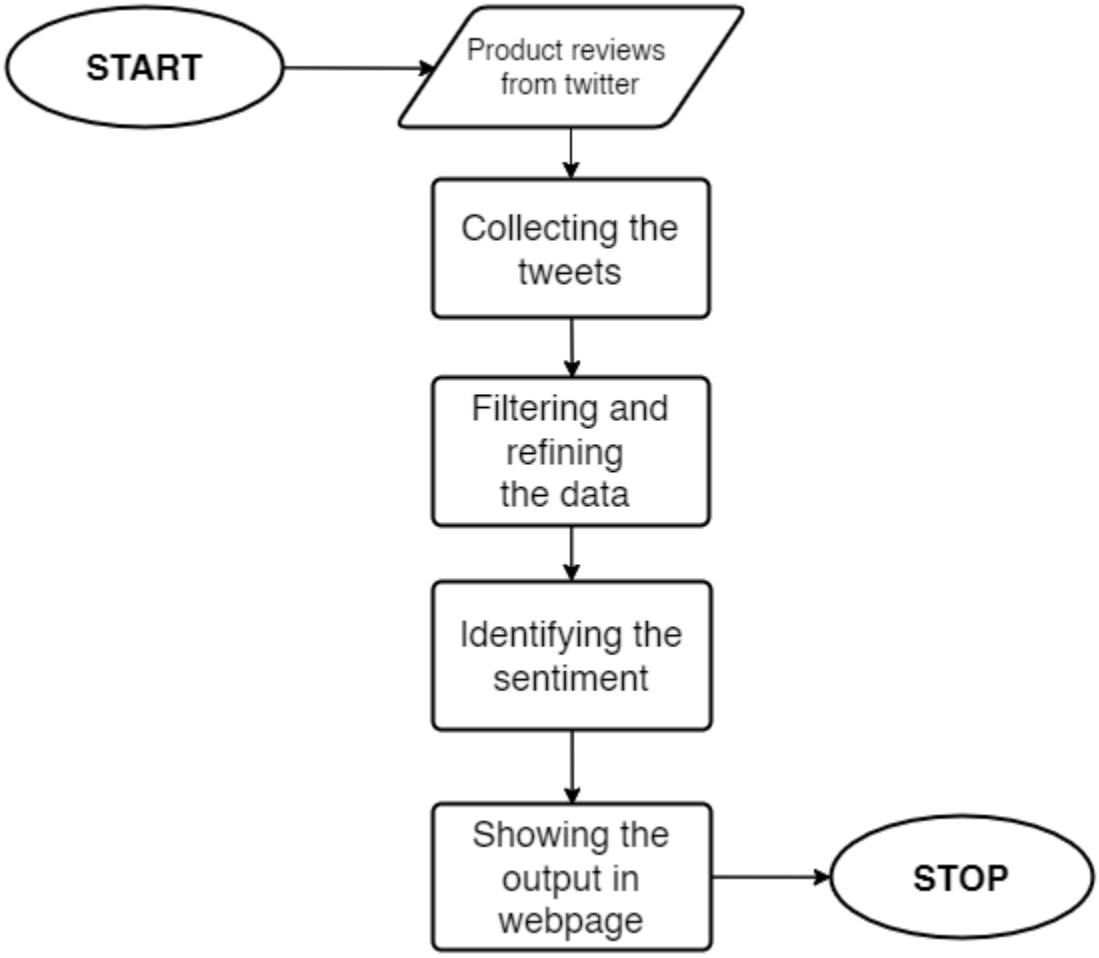

FIG 8.1: TWITTER PREDICTION SYSTEM PROCESS AFTER USER INPUTS THE QUERY AND COUNT.

Phrase based sentiment analysis option is available in the web application and it works on the basis of the following process.

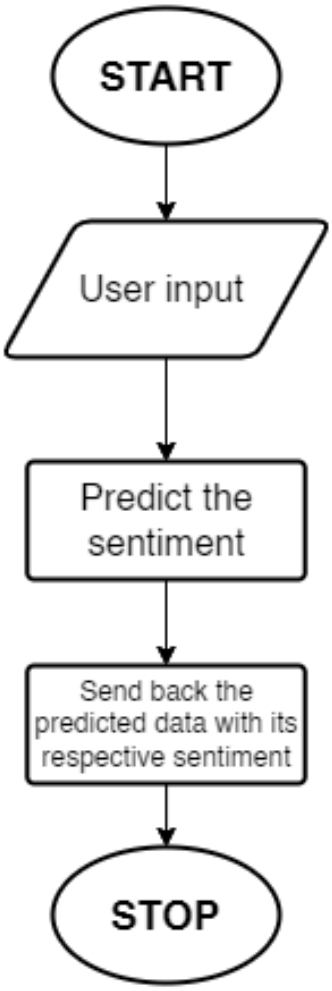

FIG 8.2: FLOW DIAGRAM FOR PHRASE BASED PREDICTION SYSTEM.

Use case diagram:

This use case diagram represents actors that interacts with the system , the system itself, the use cases, or services, that the system knows how to perform, and the lines represent relationships between these elements.

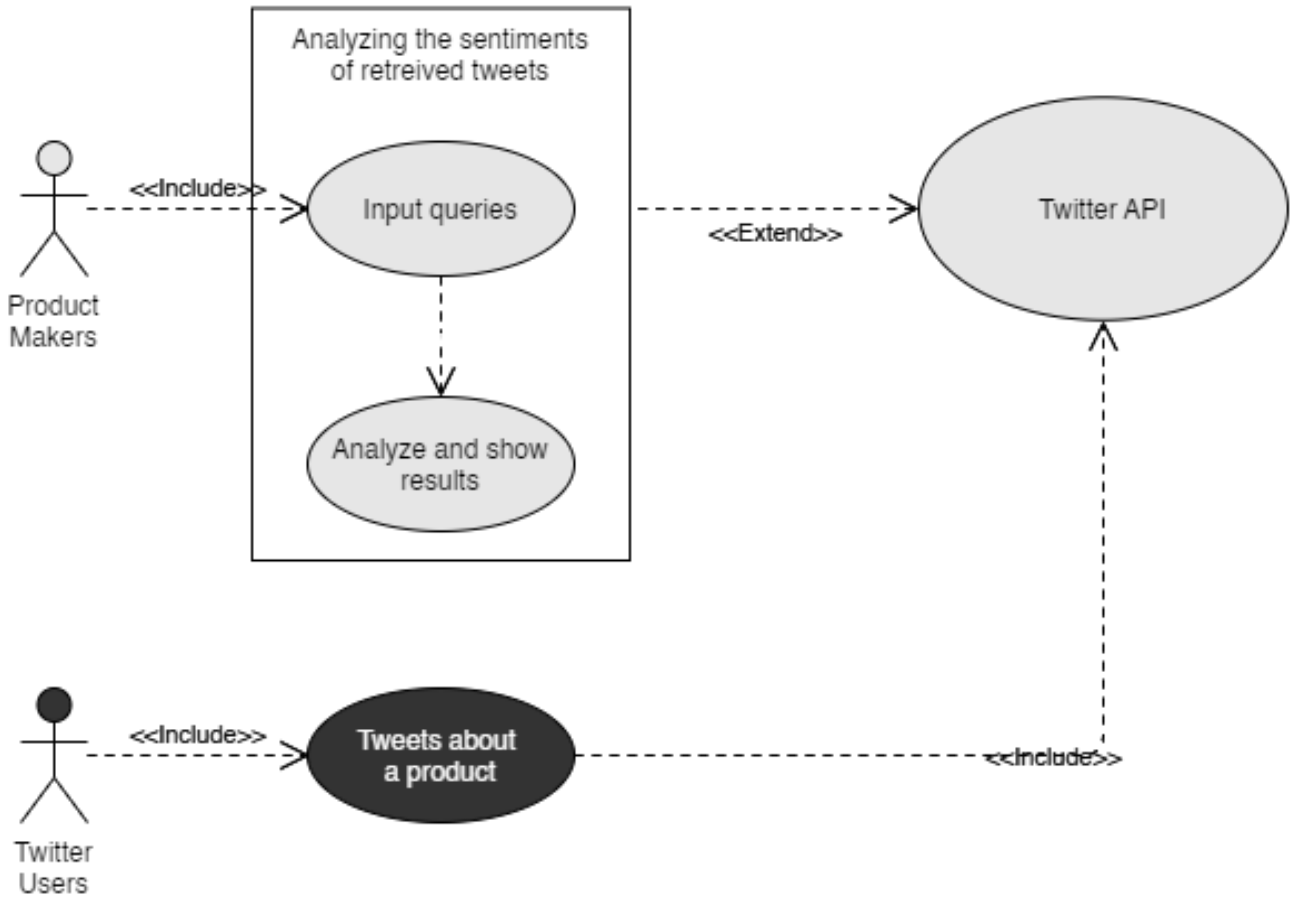

FIG 8.3 USE CASE DIAGRAM PROVIDES OVERVIEW OF BASIC WORKING PRINCIPLE AND THE ACTORS.

Input (Keyword):

User input the query and count of the tweets that need to be retrieved, Data in the form of raw tweets is acquired by using the Python library “tweepy” which provides a package for simple twitter streaming API. This API allows two modes of accessing.

• Specific keyword to track/search for in the tweets

• Specific Twitter user according to their name

Tweets Retrieval:

Since human labelling is an expensive process, we further filter out the tweets to be labelled so that we have the greatest amount of variation in tweets without the loss of generality. The filtering criteria applied are stated below:

• Remove Retweets (any tweet which contains the string “RT”).

• Remove tweets ‘@’ Mentions (tweets which other people might be tagged in).

• Remove non-English tweets (by comparing the words of the tweets with a list of 2,000 common English words, tweets with less than 15% of content matching threshold are discarded)

Labelling the tweets with Sentiment:

Labels the tweets according to their respective sentiment that was calculated with help of sentiment score obtained with help of Textblob’s output.

Store the tweets with their respective sentiment in a variable and use the variable to print out the results in the HTML page “/predict”

Phrase Level:

User input:

Here we take the user input from the HTML forms and stored it in a python variable as string to process later.

Predicting the sentiment:

The stored user input is then passed on to the Textblob method to find the sentiment polarity and store it in a variable and mask it with the respective sentiment.

Present the output:

Show the user input phrase with respective sentiment in “result1.html”.

Classification Algorithm:

The algorithm used in this project for classification is Naïve Bayes algorithm.

- Naive Bayes algorithm is based on the Bayesian theorem and used for solving classification problems.
- In general, it is used primarily in the context of text classification systems that contain a high-dimensional training dataset.
- One of the easy and most effective methods for determining the classification of data is the unsupervised learning algorithms known as Naive Bayes Classifiers, the algorithms are used to build fast machine learning models that can produce predictions quickly.
- It is a probabilistic classifier, which means that it is predicting an unobservable outcome based on the likelihood that it will happen.
- The Naive Bayes Algorithm has been popular for many purposes such as spam filtering, sentiment analysis, and classifying articles.

Positive:

If the entire tweet has a positive/happy/excited/joyful attitude or if something is mentioned with positive connotations. Also, if more than one sentiment is expressed in the tweet but the positive sentiment is more dominant.

Example: “4 more years of being in hell Iraq then I move to India :D”.

Negative:

If the entire tweet has a negative/sad/displeased attitude or if something is mentioned with negative connotations. Also, if more than one sentiment is expressed in the tweet but the negative sentiment is more dominant.

Example: “I want an android now this iPhone is boring”.

Neutral:

If the creator of tweet expresses no personal sentiment/opinion in the tweet and merely transmits information. Advertisements of different products would be labelled under this category.

Example: “I will be sharing tweets about Donald Trump”

Twitter API with Elevated access:

A Twitter API is a set of programmatic endpoints that is available to programmers to utilize to build an understanding or build upon a conversation taking place on Twitter.

You can use this API to discover and get, interact with, or create a range of resources, including the ones listed below.

- Tweets
- Users
- Spaces
- Direct Messages
- Lists
- Trends
- Media
- Places

Text blob module:

TextBlob is a Python package with a straightforward API for interacting with its functions and doing basic NLP (Natural language processing) tasks. TextBlob is useful since it behaves similarly to Python strings. So, just like in Python, you can transform and play with it. I've listed some fundamental chores for you below. Don't worry about the syntax; it's merely to give you an idea of how closely TextBlob and Python strings are connected. TextBlob library also comes with a NaiveBayesAnalyzer. Which is used for text-classification.

Flask Framework:

Flask is a web framework and a Python module that makes it simple to create web applications. It has a simple and extensible core: it's a microframework without an ORM (Object Relational Manager) or other things like that. Advantage of using Flask over any other python web frameworks such as Django is Flask is very pythonic and can be learnt very easily and quickly

9) SYSTEM IMPLEMENTATION:

Flask app:

This code is the backend code for the web application done with flask micro web-framework in python. This serves as a webserver once executed and can be used with the URL that it provides.

```
import re
import tweepy
from tweepy import OAuthHandler
from textblob import TextBlob
from textblob.sentiments import NaiveBayesAnalyzer

from flask import Flask, render_template , redirect, url_for, request



def clean_tweet( tweet):

        return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])
|(\w+:\/\/\S+)", " ", tweet).split())

def get_tweet_sentiment( tweet):

        analysis = TextBlob(clean_tweet(tweet))
        if analysis.sentiment.polarity > 0:
            return "positive"
        elif analysis.sentiment.polarity == 0:
            return "neutral"
        else:
            return "negative"


def get_tweets(api, query, count=5):

        count = int(count)
        tweets = []
        try:
```

```
            fetched_tweets = tweepy.Cursor(api.search_tweets, q=query,
lang ='en', tweet_mode='extended').items(count)

            for tweet in fetched_tweets:

                parsed_tweet = {}

                if 'retweeted_status' in dir(tweet):
                    parsed_tweet['text'] =
tweet.retweeted_status.full_text
                else:
                    parsed_tweet['text'] = tweet.full_text

                parsed_tweet['sentiment'] =
get_tweet_sentiment(parsed_tweet['text'])

                if tweet.retweet_count > 0:
                    if parsed_tweet not in tweets:
                        tweets.append(parsed_tweet)
                else:
                    tweets.append(parsed_tweet)
            return tweets
        except tweepy.TweepyException as e:
            print("Error : " + str(e))

app = Flask(__name__)
app.static_folder = 'static'

@app.route('/')
def home():
  return render_template("index.html")
@app.route("/predict", methods=['POST','GET'])
def pred():
    if request.method=='POST':
            query=request.form['query']
            count=request.form['num']
            fetched_tweets = get_tweets(api,query, count)
            return render_template('result.html',
result=fetched_tweets)
```

```
@app.route("/predict1", methods=['POST','GET'])
def pred1():
    if request.method=='POST':
            text = request.form['txt']
            blob = TextBlob(text)
            if blob.sentiment.polarity > 0:
                text_sentiment = "positive"
            elif blob.sentiment.polarity == 0:
                text_sentiment = "neutral"
            else:
                text_sentiment = "negative"
            return render_template('result1.html',msg=text,
result=text_sentiment)


if __name__ == '__main__':

    consumer_key = 'EwG6T8KZTCuSfv6Wy2rfUu1gO'
    consumer_secret =
'R8dREd4HyxY6flL3OOHktuEfkTAXj66HZ5QGWGUsmoKfcaNhND'
    access_token = '405461195-
HdMbZqc7YmMP5yTMG5rix5nrahxGP72WG9VjF6w1'
    access_token_secret =
'9Zl6g93TtRvH3voFlOd6pbDwFGZ5A7YLDJnogrkm1O0NT'

    try:
        auth = OAuthHandler(consumer_key, consumer_secret)
        auth.set_access_token(access_token, access_token_secret)
        api = tweepy.API(auth)
    except:
        print("Error: Authentication Failed")

    app.debug=True
    app.run(host='localhost')
```

10) RESULTS:

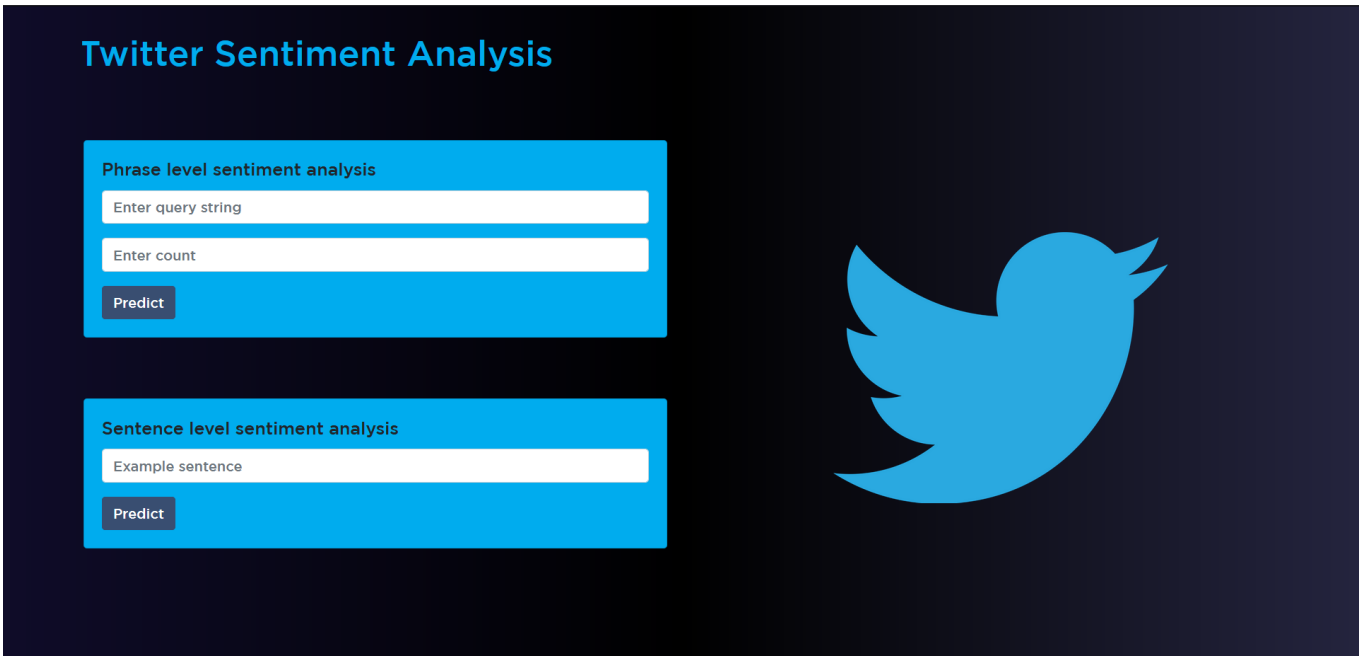

FIG 10.1: HOME PAGE OF THE WEB APPLICATION.

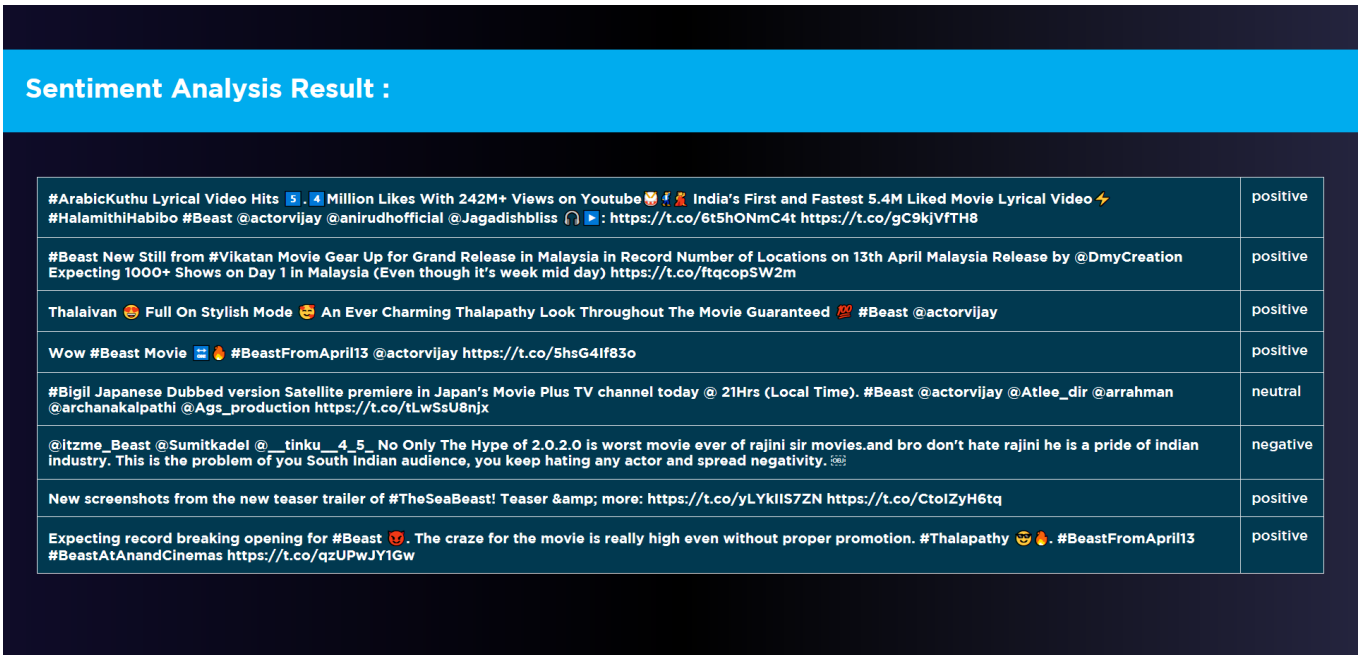

FIG 10.2: OUTPUT FOR THE QUERY "BEAST" & COUNT "8" IN TWITTER PREDICTION SECTION.

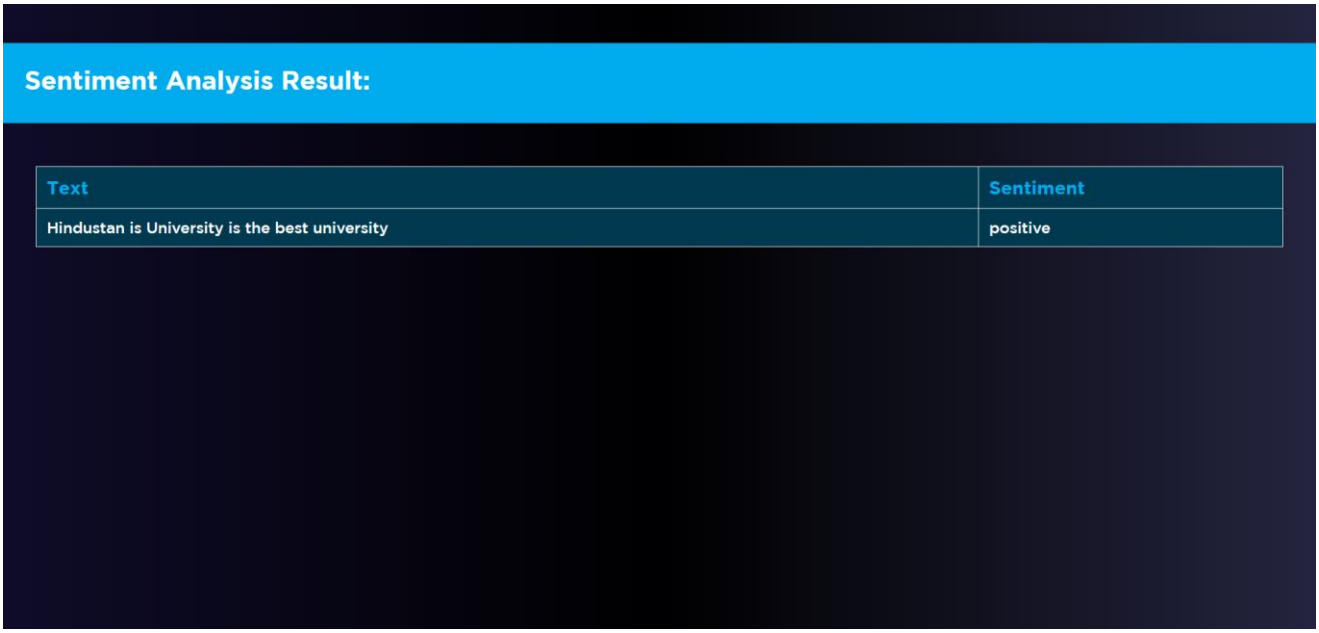

FIG 10.3: OUTPUT OF PHRASE BASED PREDICTION FOR THE INPUT “HINDUSTAN UNIVERSITY IS THE BEST UNIVERSITY” AND SENTIMENT IS POSITIVE.

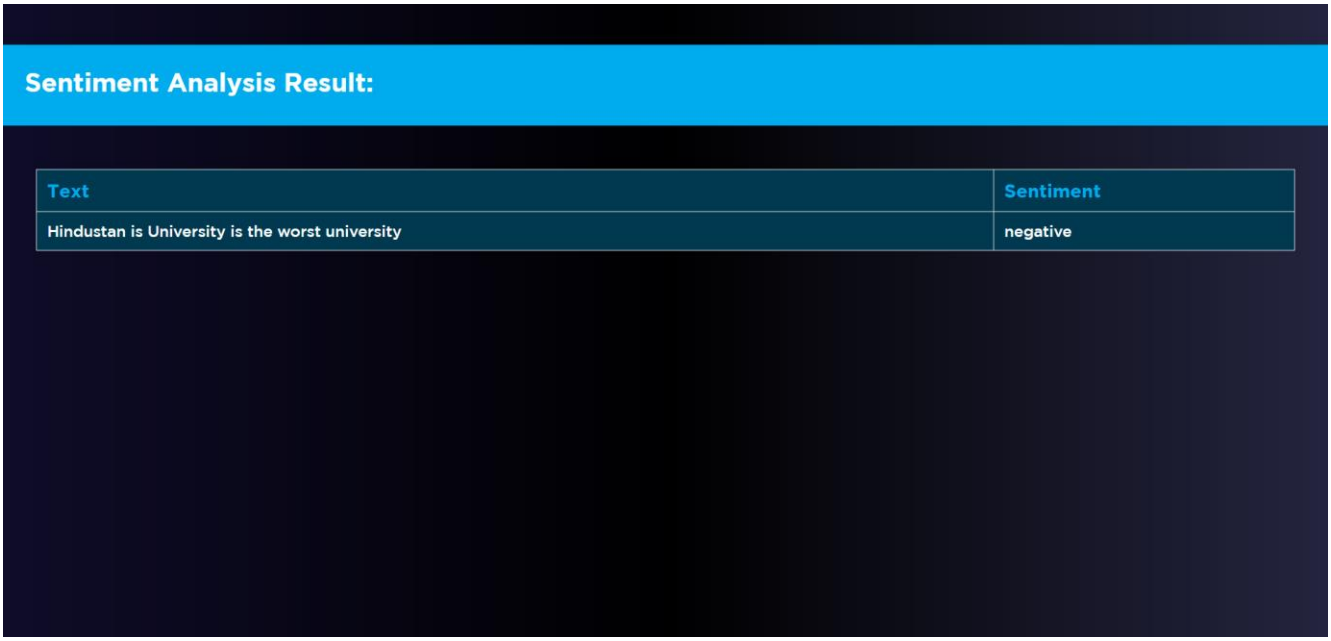

FIG 10.4: OUTPUT OF PHRASE BASED PREDICTION FOR THE INPUT “HINDUSTAN UNIVERSITY IS THE WORST UNIVERSITY” AND SENTIMENT IS NEGATIVE.

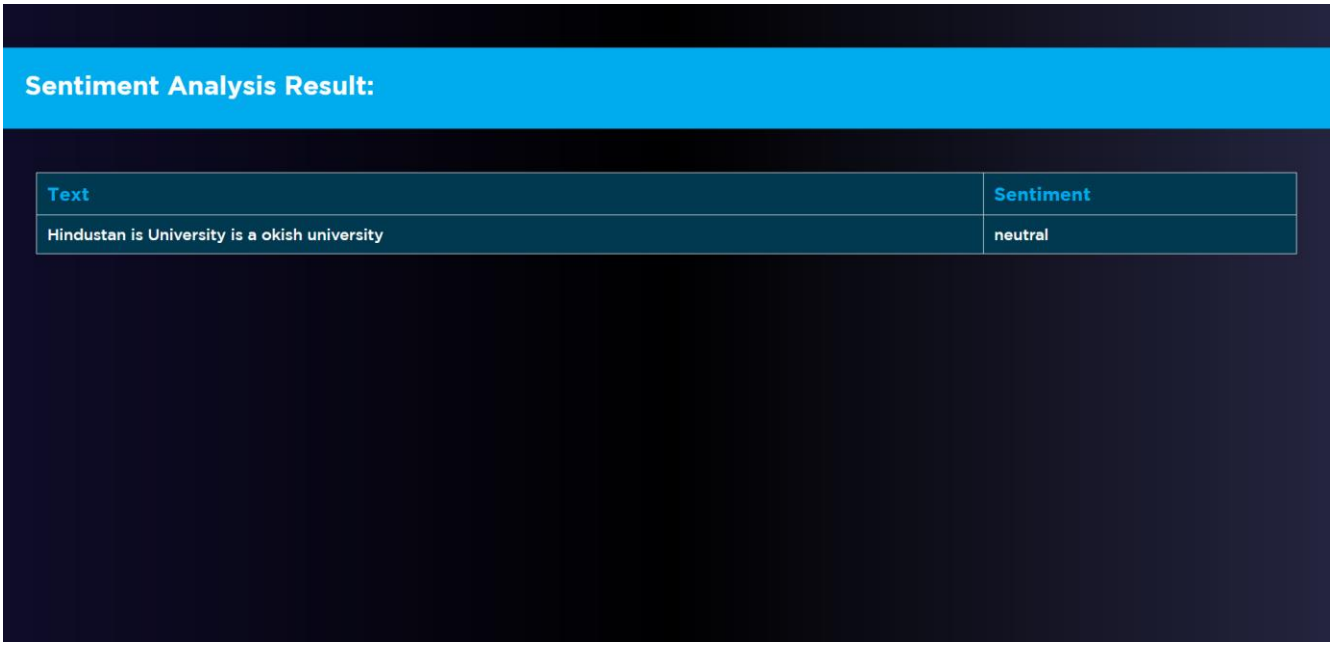

FIG 10.5: OUTPUT OF PHRASE BASED PREDICTION FOR THE INPUT “HINDUSTAN UNIVERSITY IS A OKISH UNIVERSITY” AND SENTIMENT IS NEUTRAL.

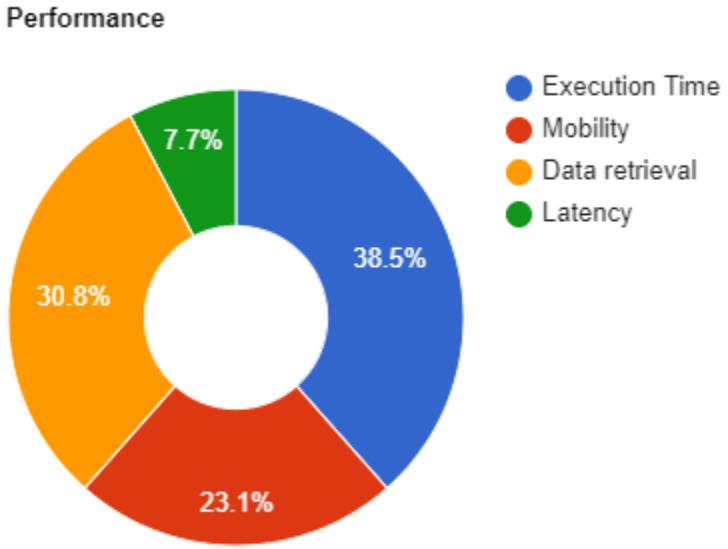

FIG 10.6: IMPROVEMENT OF PERFORMANCE BY PERCENTAGE IN PIE CHART.

11) CONCLUSION AND FUTURE WORKS:

In this paper we have discussed about Twitter sentiment analysis tool based on Natural language processing which is one of the major part of the domain of Machine learning. Main technology used was Python and TextBlob module and Tweepy module for accessing Twitter API. The sentiment classification is done based on sentiment score. These signify the positive, negative or neutral attitude of users towards a particular product. These predictions are helpful in segregating the reviews based on their sentiment and can be used to prevent loss of user trust, angry customers and user dissatisfaction. Phrase based predictions can be used for any other manual Sentiment predictions.

Future works planned is including other languages support and expanding the dashboard functionalities with more statistical visualization. The limitations of the project should overcome in the future. Mobile version for the same project is planned.

12) REFERENCES:

[1] V. K. Singh, R. Piryani, A. Uddin and P. Waila, “Sentiment analysis of movie reviews: a new feature-based heuristic for aspect-level sentiment classification,” Automation, Computing, Communication,Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on. IEEE, 2013, pp. 712-717.

[2] X. D. Chen,. “Research on sentiment dictionary based emotional tendency analysis of Chinese microblog,” Huazhong University of Science & Technology, 2012.

[3] K. Y. Chen and Z. S. He, “Sentiment classification of hotel reviews based on sentiment dictionary,” Modern Computer (Professional Edition), vol. 6, 2017, pp. 3-6.

[4] S. Y. Li, J. B. Gao and L. L. Xu, “Sentiment analysis solution based on hotel product reviews,” Computer Systems & Applications, vol. 26, no. 1, 2017, pp. 227-231.

[5] J. Jiao and Y. Zhou, “Sentiment polarity analysis based multidictionary,” Physics Procedia, vol. 22, 2011, pp. 590-596.

[6] H. L. Sang, J. Cui and J. W. Kim, “Sentiment analysis on movie review through building modified sentiment dictionary by movie genre,” vol. 22, no. 2, 2016, pp. 97-113.

[7] B. Pang, L. Lee and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol. 10, 2002, pp. 79-86.

[8] H. He, Z. Li, C. Yao and W. Zhang, “Sentiment classification technology based on markov logic networks,” New Review of Hypermedia and Multimedia, vol. 22, no. 3, 2016, pp. 243-256.

[9] Godbole, Namrata, Manja Srinivasaiah, and Steven Skiena. "Large-Scale Sentiment Analysis for News and Blogs." ICWSM 7.21 (2007): 219-222.

[10] Mondher Bouazizi, Tomoaki Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter", Access IEEE, vol. 5, pp. 20617-20639, 2017, ISSN 2169-3536.

[11] Boia, Marina, et al. "A:) is worth a thousand words: How people attach sentiment to emoticons and words in tweets." Social computing (socialcom),2013 international conference on. IEEE, 2013.

[12] Manuel, K., Kishore Varma Indukuri, and P. Radha Krishna. "Analyzing internet slang for sentiment mining." 2010 Second Vaagdevi International Conference on Information Technology for Real World Problems. 2010.

[13] Akcora, Cuneyt Gurcan, et al. "Identifying breakpoints in public opinion." Proceedings of the first workshop on social media analytics. ACM,2010.

[14] S. Y. Li, J. B. Gao and L. L. Xu, “Sentiment analysis solution based on hotel product reviews,” Computer Systems & Applications, vol. 26, no. 1, 2017, pp. 227-231.

[15] H. L. Sang, J. Cui and J. W. Kim, “Sentiment analysis on movie review through building modified sentiment dictionary by movie genre,” vol. 22, no. 2, 2016, pp. 97-113.