



Dharmsinh Desai University  
Department Of Computer Engineering  
Faculty of Technology

B. Tech. CE Semester – VI

Subject : System Design Practice

**A Project Report**  
**On**  
**“ Twitter Sentiment**  
**Analysis”**

**BY:**

Thakkar Sachin (CE134) (17CEUOG044)  
Sanghani Khantil (CE116) (17CEUOG029)

**Guided By:**

Prof. Ankit P Vaishnav

**DHARMSINH DESAI UNIVERSITY**



**CERTIFICATE**

**This is to certify that the project carried out in the subject of System Design Practices entitled “Sentimental Analysis” and recorded in this report is a work of**

**1) Sanghani Khantil**

**ROLL NO: CE-116**

**ID: 17CEUOG029**

**2) Thakkar Sachin**

**ROLL NO: CE-134**

**ID: 17CEUOG044**

**Of Department of Computer Engineering, semester VI. They were involved in Project developing during academic year 2019 -2020.**

Prof. Ankit P. Vaishnav  
Asst. Prof,  
Department of Computer Engineering,  
Faculty of Technology,  
Dharmsinh Desai University, Nadiad

Dr .C.K. Bhensdadia  
Head,  
Department of Computer Engineering,  
Faculty of Technology,  
Dharmsinh Desai University, Nadiad

## Contents

<b>Chapter 1: Abstract .....</b>	<b>4</b>
<b>Chapter 2: Introduction .....</b>	<b>5</b>
<b>Chapter 3: Flow Control.....</b>	<b>6</b>
<b>Chapter 4: Designs.....</b>	<b>7</b>
4.1    Use Case Diagrams. ....	7
4.2    Activity Diagrams.....	8
<b>Chapter 5: Implementation Detail.....</b>	<b>9</b>
<b>Chapter 6: Testing .....</b>	<b>10</b>
<b>Chapter 7: Screenshots.....</b>	<b>11</b>
<b>Chapter 8: Conclusion .....</b>	<b>13</b>
<b>Chapter 9: Limitations and Future Extensions.....</b>	<b>14</b>
<b>Chapter 10: Bibliography.....</b>	<b>15</b>

### **1) Abstract**

Twitter Sentiment Analysis is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. It has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviours.

The project can be used to analyse polarity of a given statement or word. The current world scenario can be known in very precise manner by user because on performing Sentiment Analysis on recent Twitter data system can easily known the current world scenario. When used in this way, the domain is limited and the targeted characteristics are known beforehand. This can be fed to the project to greatly increase the accuracy of the results provided.

This report aims to provide reader with conceptual understanding of the system, and also act as a working guide.

## 2) Introduction

This project, on sentimental analysis, aims to determine the polarity of a given set of statements (positive, negative or neutral).

This is achieved by lexicographic analysis of words in the sentence. A list of words segregated into positive and negatives. The system then looks for these known words, and associates them with other enhancers, diminishers or inverters, thus extracting the sentiment of the set of statements.

Further, the system tries to associate the sentiment with an appropriate noun, thus providing a more comprehensive analysis of statements provided.

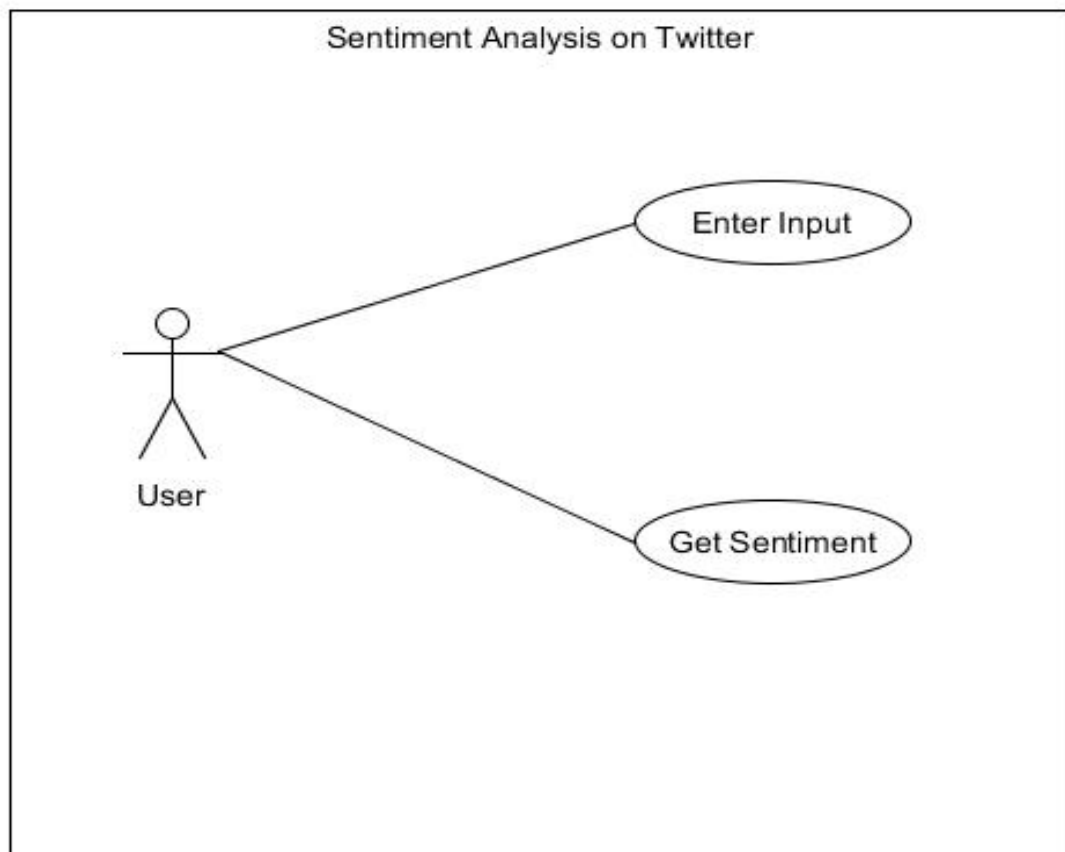
The project is provided CLI version. The former can be used to see how the program analyzes the set of statements provided, while the latter provides a better visualization in form of graphs. The core program is implemented in python. Matplotlib is used to create graphs. And those graphs are output to user in .png file.

### 3) FlowControl

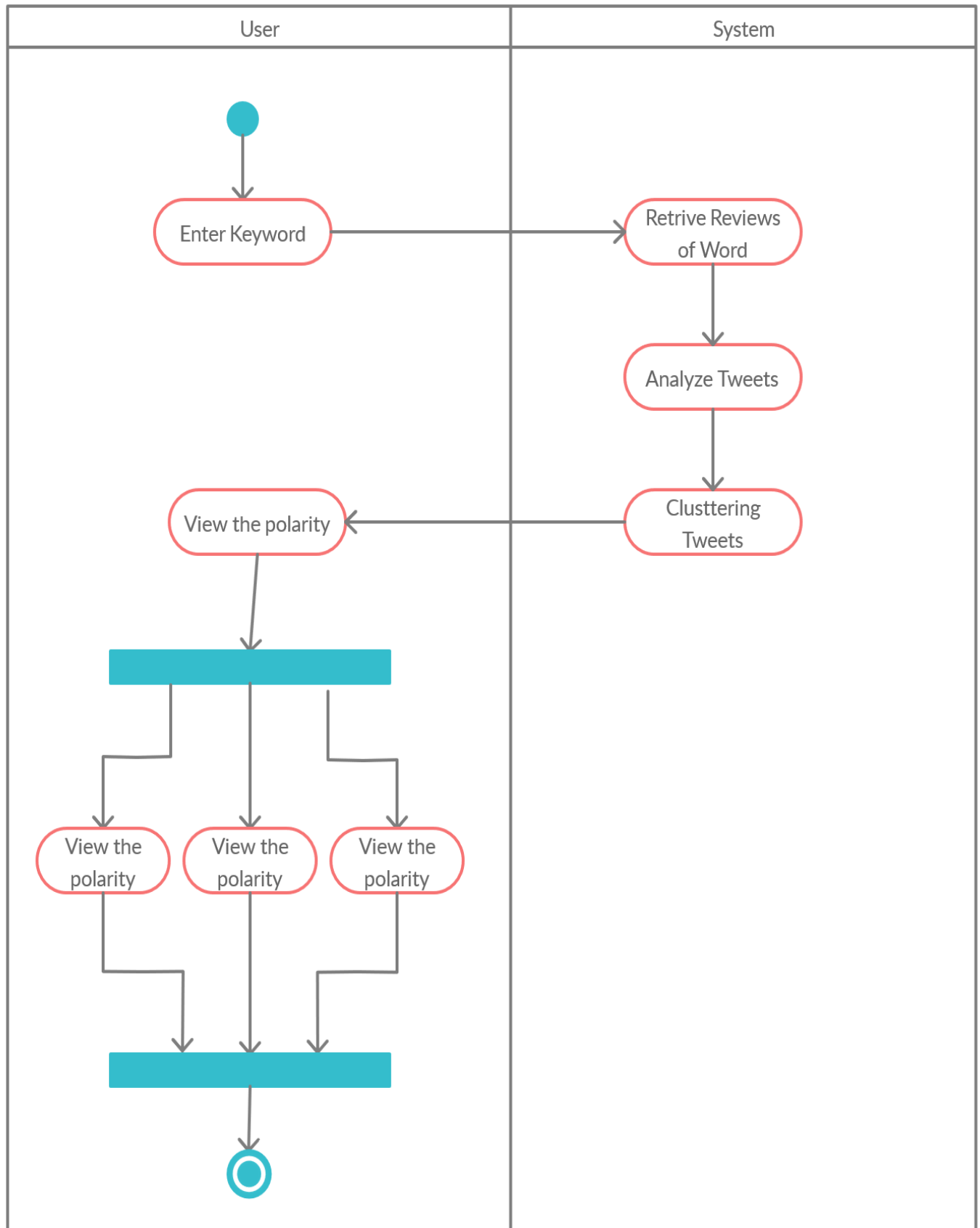
- Connect to the Twitter API
- Search Tweets (using tweepy.Cursor)
- Clean Tweets (tokenize or preprocess tweets)
- Building Vocabulary
- Matching tweets against our vocabulary
- Train and Test the Model
- Extra Features by analysis like,
  - Bar Charts
  - Pie Charts
  - Can facility of #hashtags or keyword search
  - Word Cloud for Positive Tweets

## 4) Diagrams

### 4.1) Usecase Diagrams



## 4.2) Activity Diagrams





## 5) Implementation Details

The program consists of CLI version. This version having two functionalities (User can Enter Query and Visualization as an output).

The file named “main.py” is done the main job of the system. It contains several important jobs described as follows.

By using `search(query,n,lang= 'en')` method of `tweepy.API` class, system get recent top “n” tweets of English language on the query which user has entered

For Clean Tweets : Alters case of the sentence, and performs optimization by removing unnecessary words like urls, single characters, special characters etc. After that this function do tokenizing the words, stemming the words and expanding the words to more readable format (i.e. n't to not).

For obtain trained data model which is already trained in file named “lstmmodel.sav” which is pickle file. By that system can get predicted value. After that System do analysis based on predicted value. And at the end user obtain output within 3 polarities (Positive, Neutral, Negative) and also Charts as an output for the given query.

Matplotlib is used to create graphs to show output in visualize form to user. User get Bar-chart, Pie-Chart and Word-Cloud for positive tweets as an .png file as output.

## 6) Testing

The program has been tested on several types of input datasets, to determine how well it is able to guess the sentiment.

Sr No.	Test Case	Output	Expected	Result
1	People fall, but we get up because the ground is no place for a champion.	Positive	Positive	Pass
2	The phone has great screen, but I did not like its speakers.	Neutral	Neutral	Pass
3	I did not like his acting.	Negative	Negative	Pass
4	Call quality is good, but the quality of features it provides as per price point is very bad.	Negative	Negative	Pass
5	Modi has build his Gujarat model so powerful that he had to hide it while trump visited Gujarat so that trump can not copy that.	Positive	Negative	Fail

## 7) Screenshots

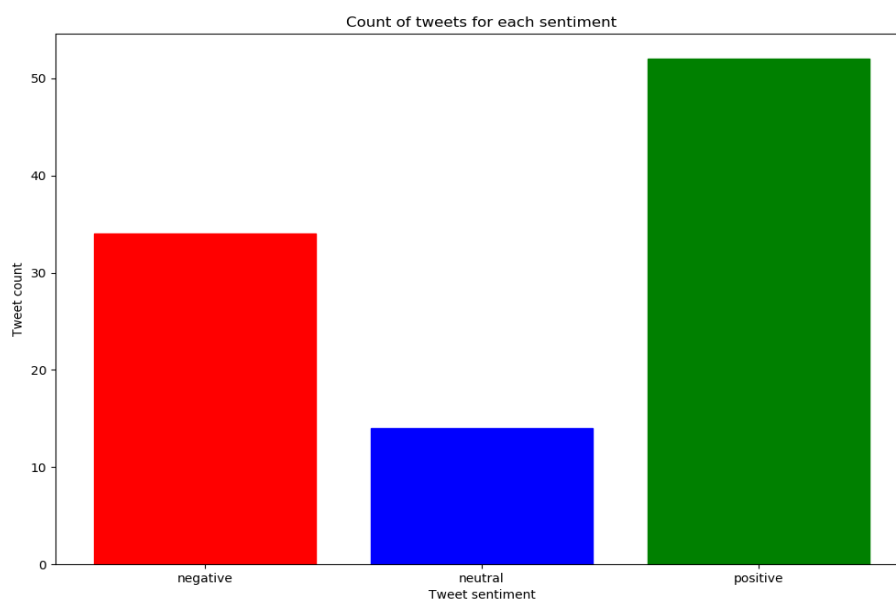
```
C:\Windows\System32\cmd.exe - python main.py
Microsoft Windows [Version 10.0.18363.778]
(c) 2019 Microsoft Corporation. All rights reserved.

S:\sentiment_final>python main.py
Using TensorFlow backend.
enter keyword
corona
```

*Asking for user input.*

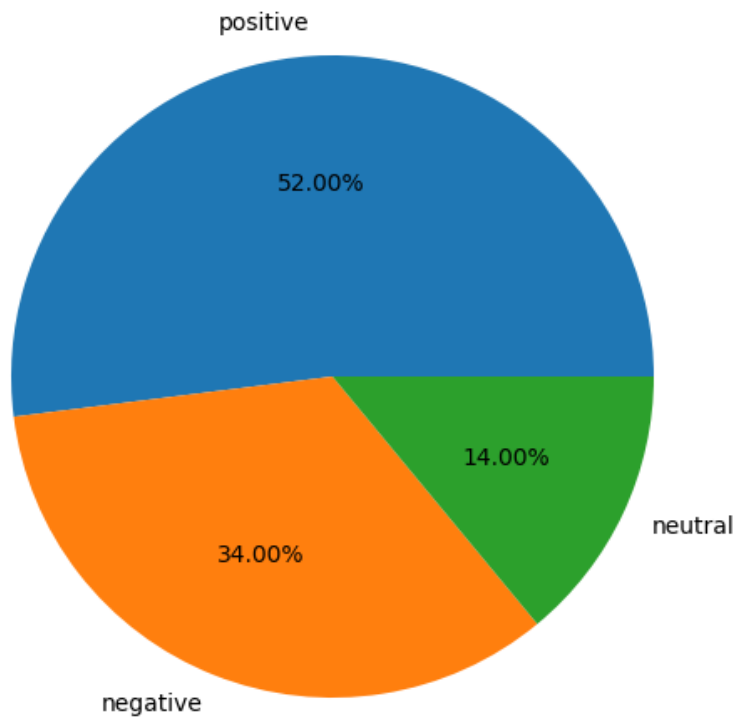
```
1024000/1024000 [=====] - 2539s 2ms/step - loss: 0.4801 - acc: 0.7680 - val_loss: 0.4639 - val_acc: 0.7770
Epoch 2/5
1024000/1024000 [=====] - 2574s 3ms/step - loss: 0.4612 - acc: 0.7792 - val_loss: 0.4573 - val_acc: 0.7809
Epoch 3/5
1024000/1024000 [=====] - 2126s 2ms/step - loss: 0.4547 - acc: 0.7829 - val_loss: 0.4540 - val_acc: 0.7832
Epoch 4/5
1024000/1024000 [=====] - 1900s 2ms/step - loss: 0.4497 - acc: 0.7860 - val_loss: 0.4514 - val_acc: 0.7851
Epoch 5/5
1024000/1024000 [=====] - 2055s 2ms/step - loss: 0.4457 - acc: 0.7884 - val_loss: 0.4494 - val_acc: 0.7859
[32m<keras.callbacks.History object at 0x000001A3C719CE48>[0m
[32mTesting the LSTM model[0m
320000/320000 [=====] - 149s 466us/step
Test accuracy: 0.786490625
```

*Successfully run and shows accuracy of 78.64% of the system.*

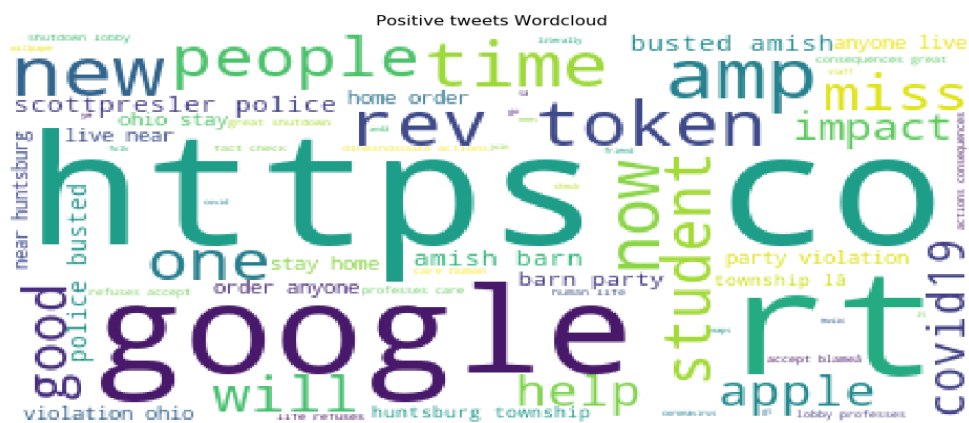


*Output Bar-Chart as .png file format*

## Twitter Sentiment Analysis



*Output Pie-Chart as .png file format*



*Word-Cloud for Positive Tweets as an output in .png file format*

## **8) Conclusion**

The system successfully classifies sentiments based on the user input query.

It also provides proper graphs as output. The graphs are Pie-charts, Bar-charts and Word-Cloud for positive tweets as .png file format which are visualized output for user.

## 9) Limitations and Future Extension

### Limitation:

The System cannot identify regional (other than English language) or context-specific words.

The System cannot identify polarity for diplomatic sentences contains sarcastic words.

Accuracy of the System is 78.64% (Approx. 79%).

### Future Extension:

The system identify more than these 3 polarities i.e. Positive, Negative and Neutral. For Example System perform identification based on sentiments like happy, sad, angry etc.

In addition of graphs output System shows “Positive Tweets Word Cloud” to user as an output.

## 10) Bibliography

### Websites:

<https://www.kaggle.com/>

<https://matplotlib.org/>

<https://www.tensorflow.org/>

<https://www.kaggle.com/kazanova/sentiment140>

