

A Project Report on  
Twitter Sentiment Analysis

Submitted in partial fulfilment for the award of the degree of

B.Tech (Branch)

By

Preetham Reddy P - 18BCE0482

Rohith Krishna - 18BCE0537

Ritesh Chowdary K - 18BCE0609

Under the guidance of  
Prof. Deepikaa S  
Assistant Professor (Junior)



**VIT<sup>®</sup>**  
**Vellore Institute of Technology**  
(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF COMPUTER SCIENCE & ENGINEERING  
April 2020

# Table of Contents

Abstract

Acknowledgement

1. Introduction

1.1. Motivation

1.2. Aim of the proposed Work

1.3. Objective(s) of the proposed work

1.4. Report Organization

2. Literature Survey

2.1. Survey of the Existing Models/Work

2.2. Summary/Gaps identified in the Survey

3. Proposed System Requirements Analysis and Design

3.1. Introduction

3.2. Requirement Analysis

3.2.1. Stakeholder Identification

3.2.2. Functional Requirements

3.2.3. Non Functional Requirements

3.2.4. System Requirements

3.2.4.1. H/W Requirements(details about Application-Specific Hardware)

3.2.4.2. S/W Requirements(details about Application-Specific Software)

3.2.5. Software Requirement Specification document

3.2.6. Work breakdown structure

3.2.7. Pert chart (with task table)

3.2.8. Gantt Chart

4. Design of the Proposed System

4.1. Introduction

4.2. High level Design (Framework, Architecture or Module for the Proposed System(with explanation))

4.2.1. Architecture design (choose the appropriate pattern with justification)

4.2.2. Architecture diagram (explanation)

4.2.3. UI design

4.3. Detailed Design (ER Diagram/UML Diagram/Mathematical Modeling)

4.3.1. ER Diagram

4.3.2. UML diagram (Use case, class, Statechart, Activity and interaction diagrams)

5. Implementation and Testing (Snap shots with description)

5.1. Implementation details (snapshots)

5.2. Testing

5.2.1. Types of Testing

5.2.2. Testcases (for all modules as per the template)

6. Conclusion, Limitations and Scope for future Work

7. References

## Abstract

Social media has received more attention nowadays. Public and private opinion about a wide variety of subjects are expressed and spread continually via numerous social media. Twitter is one of the social media that is gaining popularity.

Twitter offers organisations a fast and effective way to analyse customers' perspectives toward the critical to success in the market place. Developing a program for sentiment analysis is an approach to be used to computationally measure customers perceptions. Results classify customers perspective via tweets into positive and negative .

This has a very big use and impact on how large organisations advertise ,release their products . Once peoples opinion about a subject is gathered its analysed and the results are used by the organisations to know the opinion of their customers

Also opinions of a people of group of people can be used for politicians and law enforcement authorities .The objectives of the study are first, to study the sentiment analysis in microblogging which in view to analyse feedback from a customer of an organization's product; and second, is to develop a program for customers' review on a product which allows an organization or individual to sentiment and analyses a vast amount of tweets into a useful format.

We take single tweets and response to a tweet (thread) and analyse them to find out the positive, negative and neutrality percentage . This is done by using the twitter API called tweepy this gives us a JSON file with all the processed data which is used to train our model and when we give in a new a new tweet it can analyse the tweet nature .

## Acknowledgement

This is to certify that the project work entitled “Twitter sentimental Analysis” that is being submitted by “ Preetham Reddy P(18BCE0482) Rohith Krishna (18BCE0537), Ritesh Chowdary K (18BCE0609)” for CSE3001 Software engineering is a record of bonafide work done under supervision of Prof. Deepika.S , School of Computer Science Engineering , VIT UNIVERSITY

The contents of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted for any other CAL course

## 1. Introduction

### 1.1. Motivation

Being extremely interested in everything having a relation with the Machine Learning, the independent project was a great occasion to give us the time to learn and confirm our 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. That's why we decided to conduct my project around the Machine Learning.

### 1.2. Aim of the proposed Work

The main aim of this project is to provide the user with extensive twitter sentiment analysis just by entering the twitter handle name. This is important as all the companies and creators need to know how the audience is perceiving their product.

### 1.3. Objective(s) of the proposed work

- Make a neural network model for analysing tweets and replies of tweets of users.
- Make an interface as the model needs to be used by people who may or may not have prior knowledge of programming.

### 1.4. Report Organization

Daily users of twitter , Government analysts and Corporate analysts are the main users of our process model as they are daily users of our application for analysis purpose or the trends

## 2. Literature Survey

### 2.1. Survey of the Existing Models/Work

The literary review of this research paper was done with 10 relevant research papers. All of which have been cited.

[1] In this paper the author choses the topic for sentiment analysis to be done as the presidential elections , using tweet data obtained from social media: Twitter, by crawlers , they do the text processing from data obtained and use Naive Bayes method topredict the class. Afterward, compare with other methods such as SVM and KNN.In this study only two emotions were used for analysis, positive and negative emotion which are two classes used for classification of tweets to carry out the study , results depicted that Naïve Bayes method

has a better accuracy level 80.90% compared to using other methods, such as KNN which only has an accuracy rate of 75.58% and an accuracy rate using SVM which is 63.99%.

[2] The two different machine learning techniques of data analysis of twitter are discussed like Naïve Bayes, SVM and Maximum Entropy Method. The analysis of twitter data is being done in various aspects to mine the sentiments. This study defines the concept of opinion in sentiment analysis of Twitter. This study has tweets being classified into positive, negative and neutral reactions, the results of the study showed that the study shows that the machine learning method such as Naïve Bayes has the highest accuracy and can be considered as the baseline learning methods as well as in some cases Maximum Entropy methods are very effective.

[3] In this paper, a systematic analysis of the tweets to give results was taken out from Twitter. Bag of words for positive and negative analysis were in form of txt files were taken to compare the tweets and classify them into positive, negative and neutral tweets. The application built was able to generate word cloud for most frequent words found in the tweets retrieved. Visualization was focused extensively to ensure better analysis and ease of understanding the exact opinion portrayed by user in tweets by analyzing emoticons.

[4] In this paper the authors from different points of view to mine the opinion. This survey paper discussed different techniques of sentiment analysis and methodology for sentiment analysis. If we are doing twitter sentiment analysis, it is necessary to know about the twitter, about extricating the tweets, its structure, and their significance.

[5] Neethu M. S. and Rajasree R used twitter posts on electronic products, compared the accuracy between different machine learning algorithms and further improved the accuracy by replacing repeated characters with two occurrences, including a slang dictionary and taking emoticons into consideration.

[6]. In addition, the area of neural networks has been investigated for performing sentiment analysis on benchmark datasets consisting of online product reviews. Beshpalov, Bai, Qi and Shokoufandeh carried out binary classification on Amazon and Trip Advisor datasets using a Perceptron classifier and obtained one of the lowest error rates among their experiments of 7.59 and 7.37 on the two datasets respectively.

[7] In this paper different configurations of deep learning methods based on CNN and LSTM networks are tested for sentiment analysis in Twitter data. In this paper different configurations of deep learning methods based on CNN and LSTM networks are tested for sentiment analysis in Twitter data. It was observed that when CNN and LSTM networks are combined together they perform better than when used alone. This is due to the effective dimensionality reduction process of CNN's and the preservation of word dependencies when using LSTM networks.

[8] Jotheeswaran and Koteeswaran performed binary classification on the IMDB dataset by employing a Multi-layer Perceptron Neural Network and using Decision Tree-based Feature Ranking for feature extraction and a hybrid algorithm (based on Differential Evolution and Genetic Algorithm for weight training, thereby obtaining a maximum classification accuracy of 83.25%.

[9] Santos and Gatti developed a deep convolutional neural network model for the sentimental analysis and obtained an accuracy of 85.7% and 86.4% on the aforementioned Stanford Sentiment Treebank and Stanford Twitter Sentiment Corpus (which is bounded by its classification based on emoticons) respectively

[10] Researchers have also been working upon prediction of accuracy of tested dataset using Machine Learning algorithms. Kanakaraj and Guddeti used Natural Language Processing Techniques for sentiment analysis and compared Machine Learning Methods and Ensemble Methods to improve on the accuracy of the classification

## 2.2. Summary

In the above literature survey we had the liberty to come across various models adopted for training the machine learning model to use it then on to perform an effective sentimental analysis of tweets on a person's or organizations twitter handle , different classifiers were adopted , the perceptron classifier for instance gave low error rates as much as mid 7 %and we also came to know that machine learning approach produced better results as supposed to natural language processing approach for sentimental analysis in machine learning approach the analysis was performed between supervised and unsupervised learning methods along with the hybrid , which stood as clear winner , producing better accuracy in classification of tweets than both the supervised and unsupervised models

In most of the papers, the authors were using different methods like support vector machines (SVM), Decision Tree, etc. Although Neural Networks are very power hungry and take a lot of time for training and testing there is no doubt that they give the best results in today's available models. Mainly in workloads like Natural Language Processing Deep learning models make a lot of sense. And on top of that we are using word to vector models for better usage of resources and better accuracy.

## 3. Proposed System Requirements Analysis and Design

### 3.1. Introduction

In this phase of the report we will extensively mention and discuss about the requirement analysis ,and design aspect of the project that includes user requirements , functional and non functional requirements and system requirements and stake holder identification

### 3.2. Requirement Analysis

#### 3.2.1. Stakeholder Identification

The stakeholders for our projects are the groups of highly keen people active on social media ,use twitter to be specific as per this application , to know about the positive , negative reactions to their recent posts , which can be used to analyse peoples reaction to his tweet , it can also be used by officials of organization releasing products into market to know the



pulse of the product released in the market , which help them to make changes accordingly , and by government analysts to analyse the reaction of people towards the decision passed by the government

### 3.2.2. **Functional Requirements**

Functional requirements are functions or features that must be included in the system to meet the business needs and be acceptable by the user.

In our application functional requirements are:

- (i)The system should be able to read the data regarding the tweets from the dataset
- (ii)The system should analyze the data and classify each tweet polarity

### 3.2.3. **Non Functional Requirements**

Non functional requirements is the description of features, attributes of a system as well as constraints that may limit the boundaries of the proposed system.

They are generally based on performance, control, security efficiency and services  
Non functional requirements for our application are :

- (i) User friendly and interactive interface
- (ii) System should provide better accuracy in classification of tweets
- (iii) To perform with efficient throughput and response time

### 3.2.4. **System Requirements**

#### 3.2.4.1. H/W Requirements(details about Application-Specific Hardware)

- 1.4GB RAM
- 2. Intel I5 or later processor
- 3. 128GB SSD or HDD
- 4. Internet Connectivity for Tweepy API

#### 3.2.4.2. S/W Requirements(details about Application-Specific Software)

- i)Any operating system with python installed
- libraries:
- ii)tensorflow

- iii)sklearn
- iv)tkinter
- v)tweepy
- vi)textblob
- vii)pandas

### 3.2.5. Software Requirement Specification document

#### User Requirements:-

the User of this application is provided with an feature for sentimental analysis of tweets that are collected from twitter.

The user is provided with analysis of tweet he wants to analyze , he can see the positive threads , negative threads and the neutral threads of a post which helps him to know the reaction of general public towards the tweet.

#### Functional Requirements :

Functional requirements are functions or features that must be included in the system to meet the business needs and be acceptable by the user.

In our application functional requirements are:

- (i) The system should be able to read the data regarding the tweets from the dataset
- (ii) The system should analyze the data and classify each tweet polarity

#### Non Functional Requirement:

Non functional requirements is the description of features, attributes of a system as well as constraints that may limit the boundaries of the proposed system.

They are generally based on performance, control, security efficiency and services Non functional requirements for our application are :

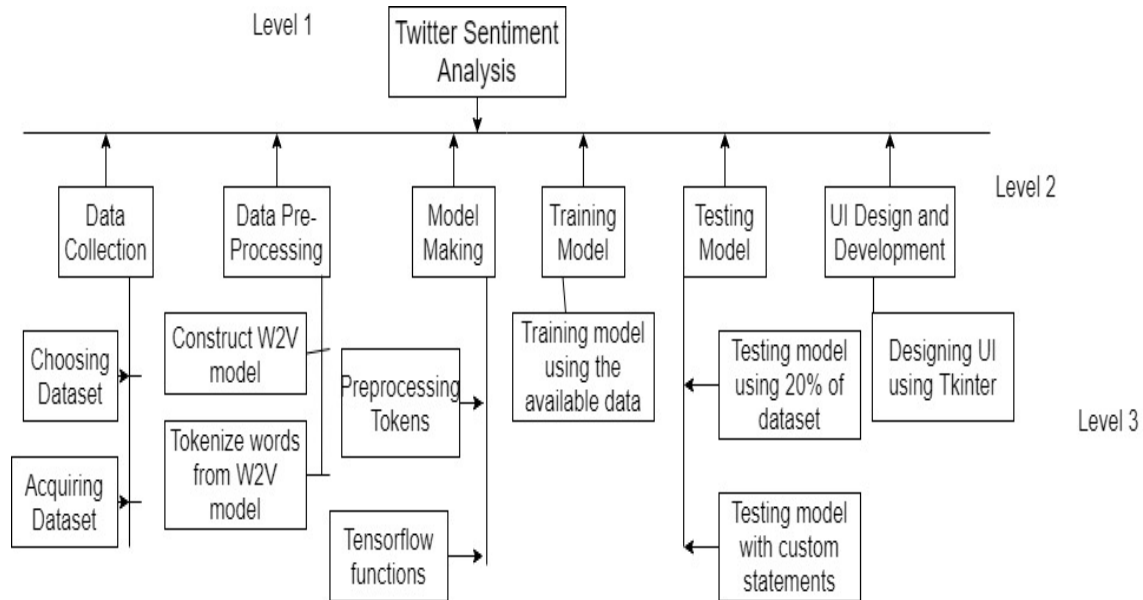
- (i) User friendly and interactive interface
- (ii) System should provide better accuracy in classification of tweets
- (iii) To perform with efficient throughput and response time

#### Other Requirements:

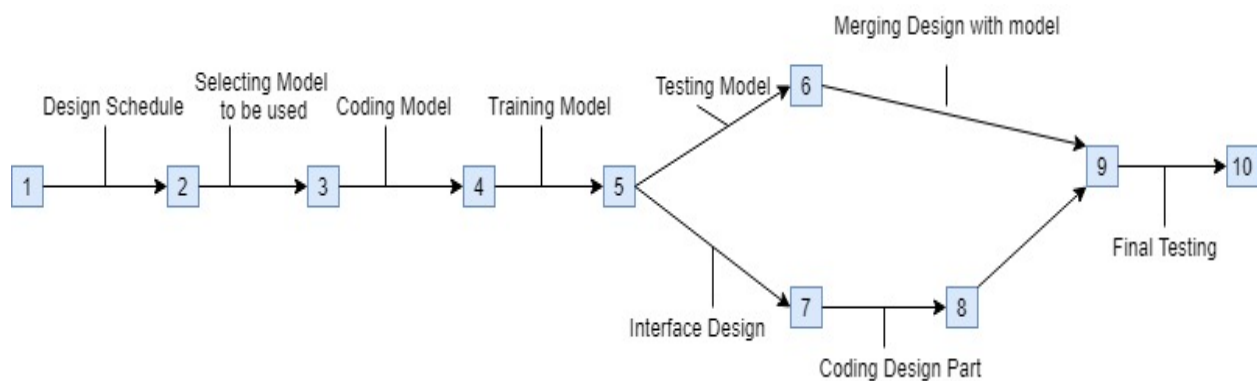
- i)Linux Operating System/Windows
- ii)Python Platform(Anaconda2,Spyder,Jupyter)

- iii)NLTK package,
- iv)Modern Web Browser
- v)Twitter API, Google API

### 3.2.6. Work breakdown structure



### 3.2.7. Pert chart (with task table)



Task	Prerequisite Task	Time Required
1	-	-

2	1	1
3	2	1
4	3	5
5	4	3
6	5	3
7	5	2
8	7	3
9	6,8	1
10	9	2

### 3.2.8. Gantt Chart

Task Name	July	Aug	Sept	Oct	Nov	Dec
Coding Model						
Training Model						
Testing Model						
Interface Design						

## 4. Design of the Proposed System

### 4.1. Introduction

First the dataset is run through various pre-processing techniques for the proper cleaning and formatting of data. From the data in the dataset we construct a word to vector model which is primarily used to map words which have similar/same meanings.

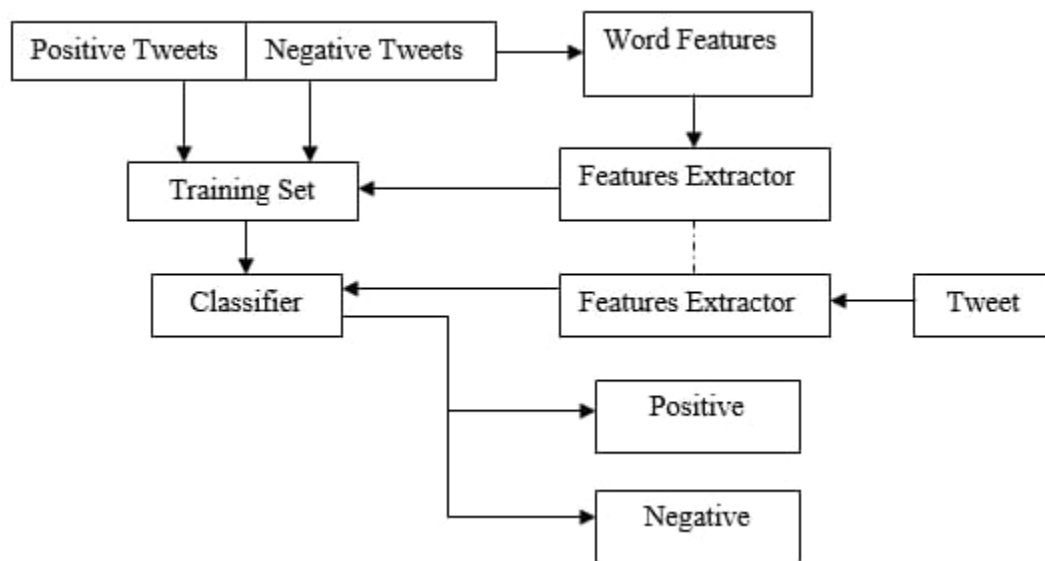
Then the model is constructed with keras by TensorFlow. We then train the model using the word to vector data obtained the previous steps. Then we attach the model to the tweepy API using the predict() function. The output from the predict function is then showed in the User Interface designed

### 4.2. High level Design (Framework, Architecture or Module for the Proposed System (with explanation))

#### 4.2.1. Architecture design (choose the appropriate pattern with justification)

For our project we initially take a set of positive and negative tweets . Using these tweets we create a training set . From the training set we extract the word features and feed it to the training set .Once our classifier is trained we give the tweet required to be analysed . The NLP feature extractor tokenises and vectorizes the tweet . Once we get the vectorised tweet the classifier compares checks if the tweets is positive or negative and gives the tweet analysis

#### 4.2.2. Architecture diagram (explanation)



#### 4.2.3. UI design

Twitter Sentimental Analysis

Enter the twitter handle you want to analyze:

Twitter Sentimental Analysis

Enter the twitter handle you want to analyze:

Positive tweets percentage: 17.7777777777778 %

Negative tweets percentage: 15.5555555555555 %

Neutral tweets percentage: 66.6666666666667 %

Positive tweets:

@Kwa\_Mamkhize Who would give out a whole million for likes ah yohh ur not bill gates

RT @F\_Sanchez\_Drago: Gane quien gane, ja, ja, ja... Dedico esta carcajada al New York Times, al Washington Post, a la CNN, a los periodistas...

Great H2 news @BillGates 🙄🙄🙄 <https://t.co/0i2EPy9xw>

RT @wiseconnector: Don't say you don't have enough time.

You have exactly the same number of hours per day as:

➡ Mark Zuckerberg

➡ Bil...

@Cr8zyB\_95 @Conserva\_Diva Nope, watch this and follow diagram to my best knowledge, remember "they" wanted a legal... <https://t.co/GQ0uY9sF4f>

Bill Gates books to read during pandemic for inspiration 2020

- The Splendid and the Vile: A Saga of Churchill, Fa... <https://t.co/fQyNSZRkbb>

RT @JonErichman: Increase in net worth today:

Jeff Bezos: +\$10.5 billion

Mark Zuckerberg: +\$8.1 billion

Larry Page: +\$4.1 billion

Sergey...

4 books Bill Gates recommends reading right now @CNBC <https://t.co/AUJNNGOzQD>

Negative tweets:

RT @WikiAnons: Anthony Fauci & CDC funded by Bill Gates knew flu vax plus flu caused enhanced flu which #COVID-1984 hospitals treated wrong...

RT @InnoEnergy\_S: Bill Gates hjälper EU bygga väteekonomi | @InnoEnergyEU startar ett center som ska främja ett europeiskt ekosystem för gr...

RT @goranalubuja: Sad dolazi NAJBITNIJA informacija. Ta organizacija je na 8. mestu donatora SZO svetske zdravstvene organizacije, na čelu B...

Il a déjà un projet politique à base d'excréments humains. On va dire qu'il continue sur sa lancée.... <https://t.co/mW6i85SGI>

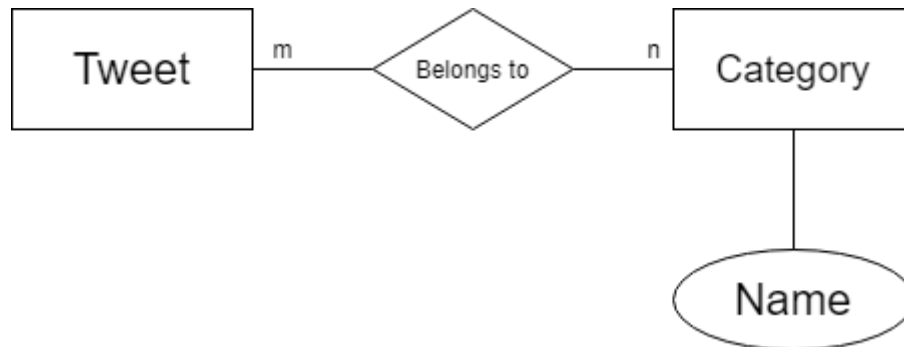
@durucysi @Karl\_Lauterbach Alles manipuliert. Von Bill Gates. Und Black Lives Matter. Und den Kommunisten.

RT @WorldTruthTV: George Soros and Bill Gates Exposed As The Force Behind Facebook's New 'Fake News' Detector <https://t.co/GeVvwBynaR> via @...

RT @AfricaRepublic: Massive calls for @realDonaldTrump to arrest Bill Gates for crimes against Humanity intensifies in America. White Ho...

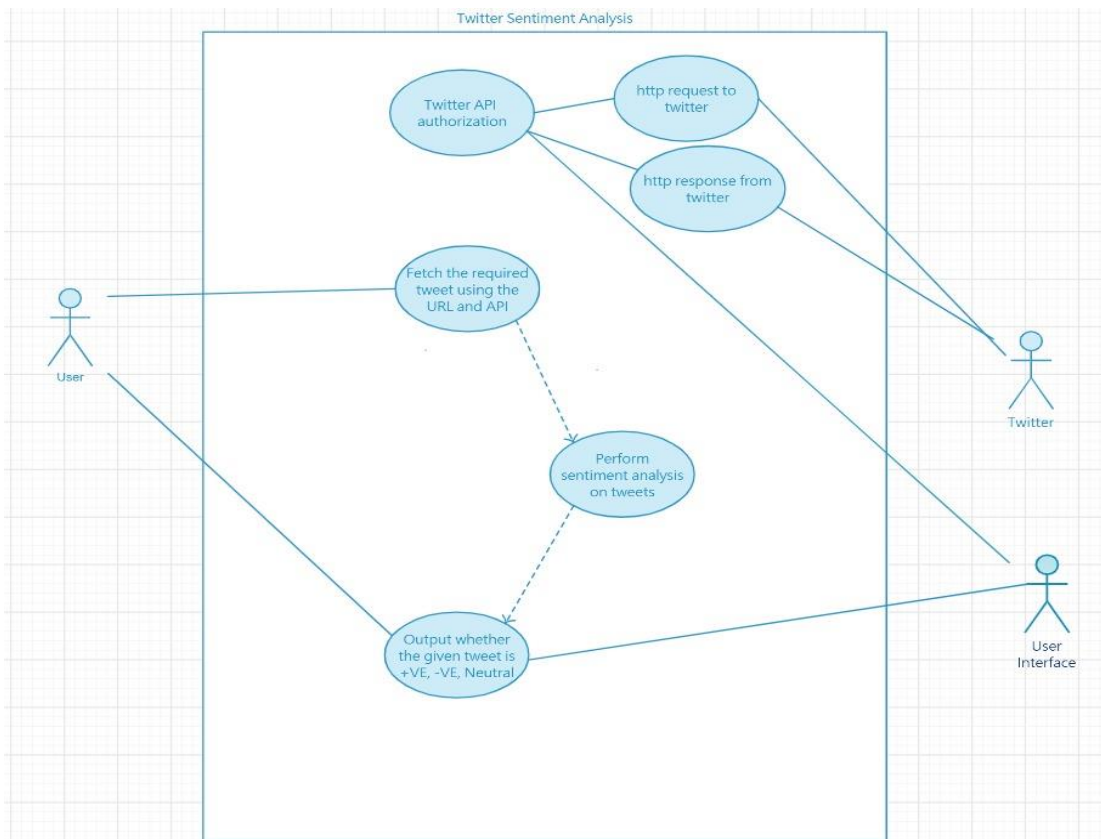
### 4.3. Detailed Design (ER Diagram/UML Diagram/Mathematical Modeling)

#### 4.3.1. ER Diagram

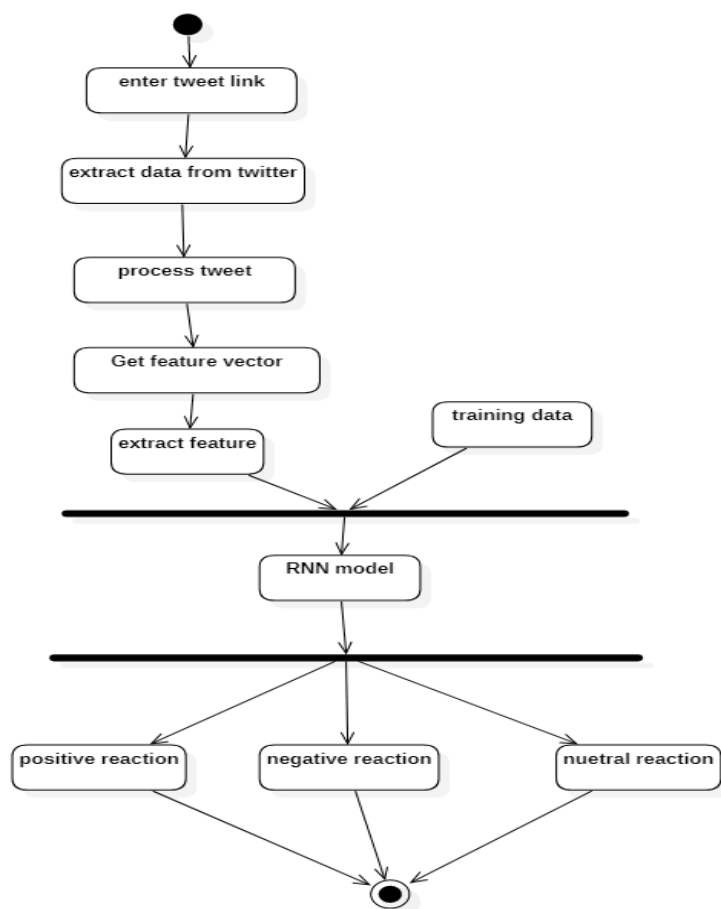


#### 4.3.2. UML diagram (Use case, class, Statechart, Activity and interaction diagrams)

##### USE CASE DIAGRAM

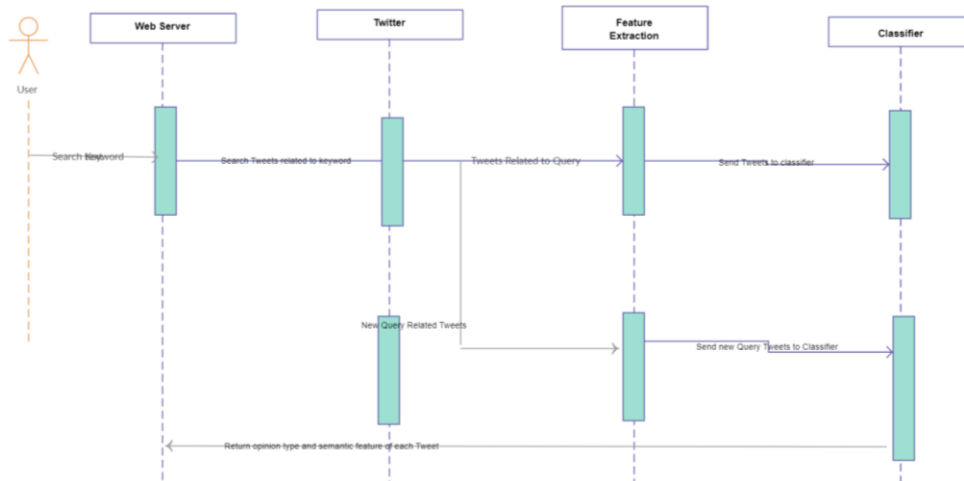


##### Activity Diagram

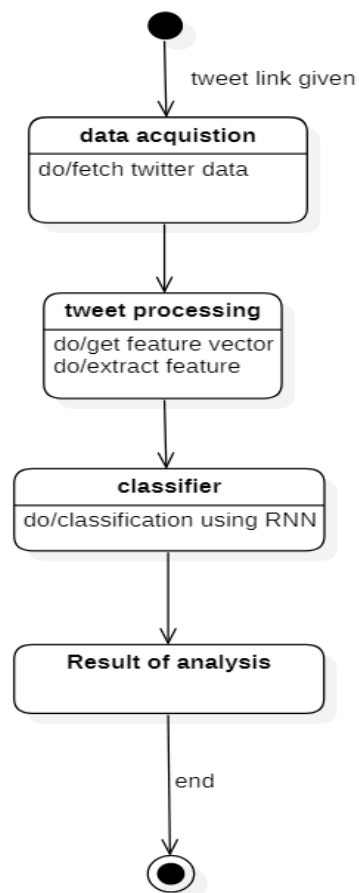


Sequence Diagram:

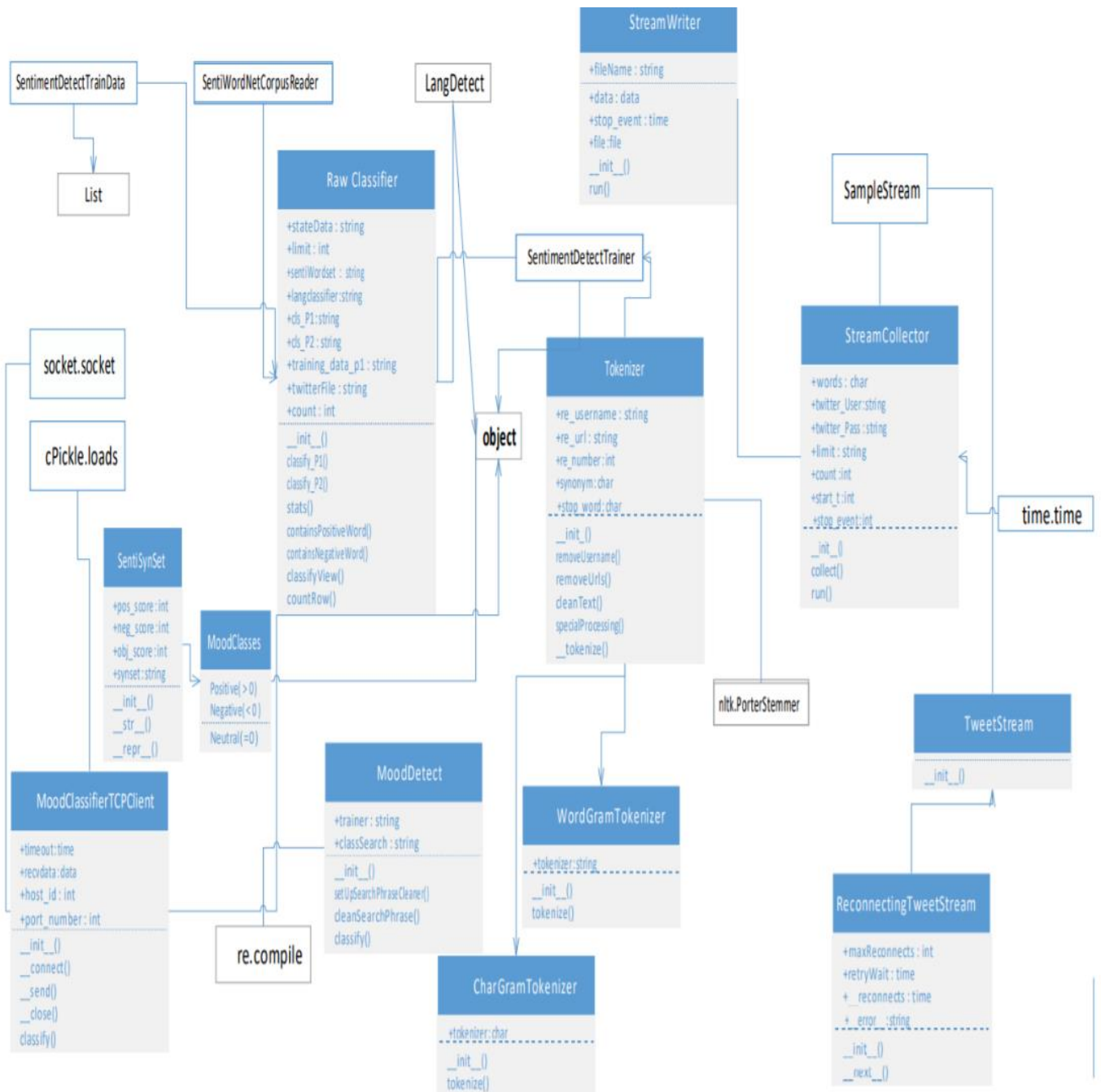




### STATE CHART DIAGRAM:



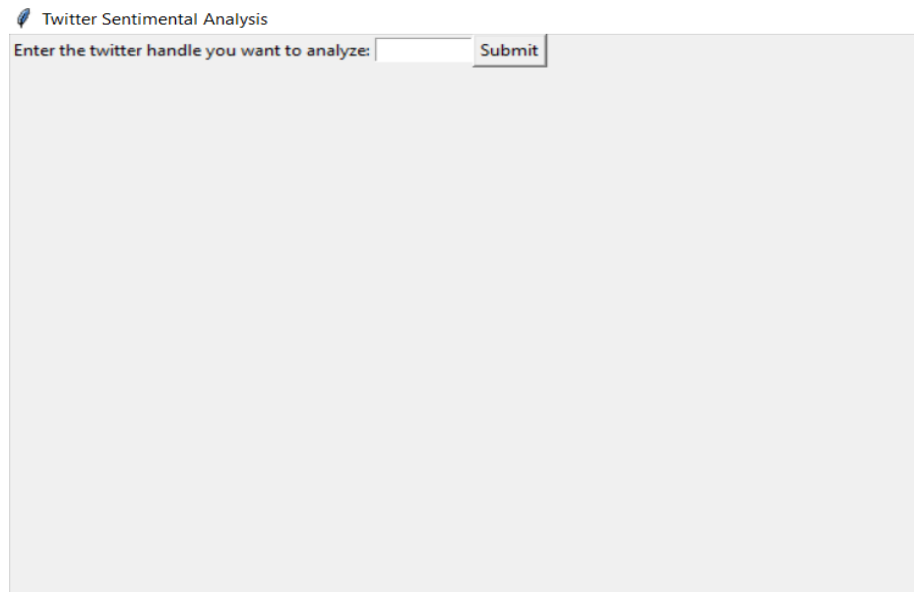
## CLASS DIAGRAM



## 5. Implementation and Testing (Snap shots with description)

### 5.1. Implementation details (snapshots)

**Initial screen after the application starts running , text box to enter twitter handle name**



**After a valid twitter handle is entered in text box provided and user clicks on submit**

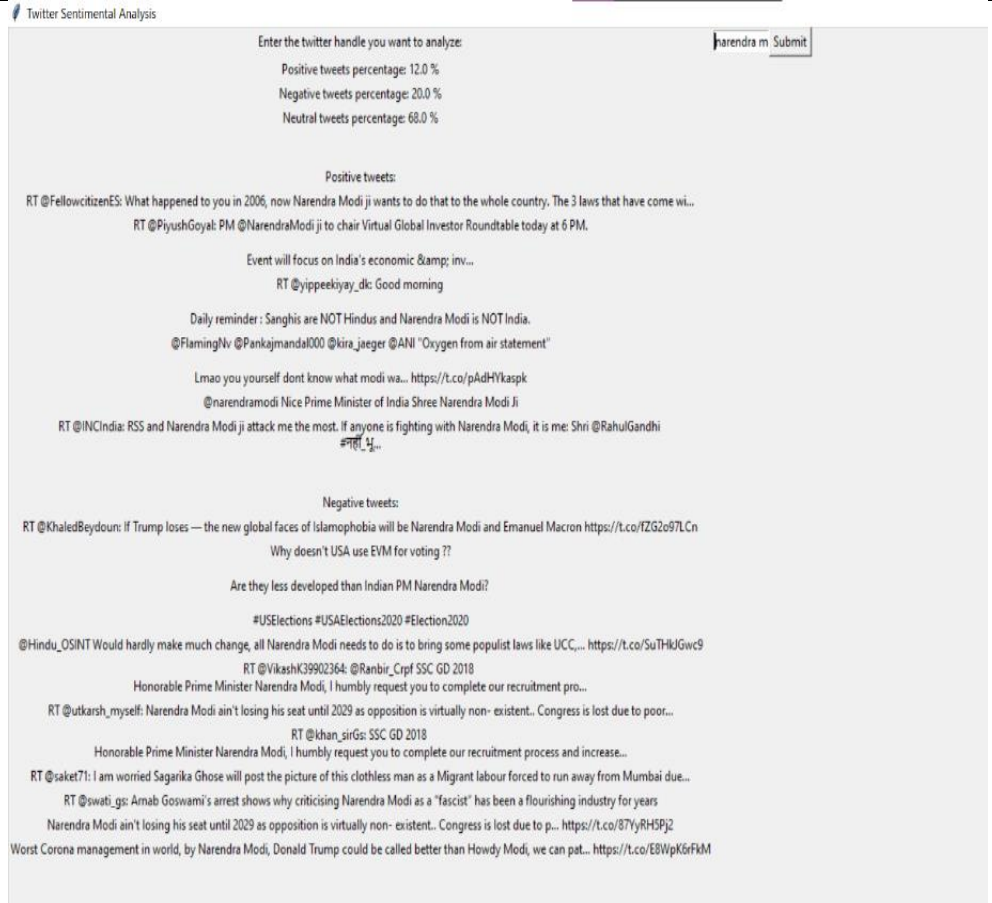


## 5.2. Testing

### 5.2.1. Types of Testing

We have adopted the approach of black box testing. In regards with the nature of our project we chose to do Equivalent class partitioning, where input data of software is divided into partitions for deriving test cases

### 5.2.2. Testcases (for all modules as per the template)

TEST CASE	INPUT	OUTPUT
C1	Narendra Modi	 The screenshot shows a web application titled 'Twitter Sentimental Analysis'. It has a text input field containing 'Narendra Modi' and a 'Submit' button. Below the input, the results are displayed: 'Positive tweets percentage: 12.0 %', 'Negative tweets percentage: 20.0 %', and 'Neutral tweets percentage: 68.0 %'. There are three sections of tweets: 'Positive tweets' with 10 tweets, 'Negative tweets' with 10 tweets, and 'Neutral tweets' with 10 tweets. The tweets are displayed in a list format with their respective text and user handles.
C2	asjdfshbf	Error

C1: a valid twitter account name

C2: an invalid twitter account name

## **6. Conclusions, Limitations and Future Scope of work:**

### **Conclusions:**

In this project we made a twitter sentiment analysis software which analyses tweets of the given twitter handle. We used Deep Learning techniques for the machine learning model in this software. Here, we used TensorFlow, pandas, sklearn, matplotlib, nltk among other libraries for making the machine learning model for the analysis. The method used here is called Natural Language Processing (NLP) where the language is processed through a trained NLP model which in-turn analyses the statement to positive, neutral or negative.

### **Limitations:**

Due to the lack of processing power we did not have the chance to train the model for desired number of epochs. As the model clearly did not touch its saturation limit the model would've performed better if there were more time and processing power to train the model. And as we had little knowledge to make UI in python, we could not do that great of an interface design. We used Tkinter, a basic interface making library in Python.

### **Future Scope or work:**

As new machine learning models come, we can expect that this same work can be done with much more accuracy and more efficiency. As deep learning models are considered the best right now for Natural Language Processing, we used Neural Networks. With new machine learning concepts such as reinforcement learning we can expect better accuracy.

## **7. References:**

[1] Meylan Wongkar and Apriandy Angdresey, Sentiment Analysis Using Naive Bayes Algorithm of The Data Crawler : Twitter, 2019 Fourth International Conference on Informatics and Computing (ICIC)

[2] Lokesh Mandloi and Ruchi Patel, Twitter Sentiments Analysis Using Machine Learning Methods, 2020 International Anupama Kumar, Real Time Sentiment Analysis Of Twitter Posts, 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018 Conference for Emerging Technology (INCET)

[3] V. Prakruthi; D. Sindhu; Dr. S. ; Dr. S. Anupama Kumar, Real Time Sentiment Analysis Of Twitter Posts, 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018

[4] Rasika Wagh and Payal Punde , Survey on Sentiment Analysis using Twitter Dataset, 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)

[5] Neethu M. S. and Rajasree R., Sentiment Analysis in Twitter using Machine Learning Techniques, 4th IEEE International Conference on Computing, Communications and Networking Technologies, pp. 1-5, Tiruchengode, India, 2013

[6]Bespalov D., Bai B., Qi Y., and Shokoufandeh A., Sentiment classification based on supervised latent n-gram analysis, 20th ACM international conference on Information and knowledge management, pp. 375-382, New York, USA, 2011

[7]Dionysis Goularas and Sani Kamis, Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data, 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)

[8]Jotheeswaran J. and Koteeswaran S., Decision Tree Based Feature Selection and Multilayer Perceptron for Sentiment Analysis, Journal of Engineering and Applied Sciences, vol. 10, issue 14, pp. 5883-5894, January 2015

[9]dos Santos C. N. and Gatti M., Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts, 25th International Conference on Computational Linguistics, pp. 6978, Dublin, Ireland, August 2014.

[10]Bahrainian S.-A., Dengel A., Sentiment Analysis and Summarization of Twitter Data”, 16th IEEE International Conference on Computational Science and Engineering, pp. 227-234, Sydney, Australia, December 2013