Emotion Detection from Tweets and Emoticons Using Machine Learning

Submitted in partial fulfilment of the requirements
of the degree of

Bachelor of Computer Engineering

BY

Jain Rajat Ashok (Roll No. 22)

Kshatriya Sawan Ranjan (Roll No. 32)

Dubey Arun Kumar Vyas (Roll No. 14)

Shukla Pratyush Suresh (Roll No. 58)

Guide

Prof. K. Jayamalini

Department of Computer Engineering

Shree L. R. Tiwari College of Engineering

Kanakia Park, Mira Road (E), Thane - 401 107, Maharashtra.

Year 2019

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Jain Rajat Ashok

Roll No.:22

Kshatriya Sawan Ranjan

Roll No.:32

Dubey Arun Kumar Vyas

Roll No.:14

Shukla Pratyush Suresh

Roll No.:58

Date: 26th April, 2019

Shree Rahul Education Society's (Regd.)
SHREE L. R. TIWARI COLLEGE OF ENGINEERING
Kanakia Park, Near Commissioner's Bungalow, Mira Road (East), Thane 401107, Maharashtra
(Approved by AICTE, Govt. of Maharashtra & Affiliated to University of Mumbai)
Tel. No.: 022-65295732 / 022-65142376 | Email: slrtce@rahuleducation.com | Website: www.slrtce.in

CERTIFICATE

This is to certify that the project entitled **"Emotion Detection from Tweets and Emoticons Using Machine Learning"** is a bonafide work of

Jain Rajat Ashok (22)

Kshatriya Sawan Ranjan (32)

Dubey Arun Kumar Vyas (14)

Shukla Pratyush Suresh (58)

submitted to the University of Mumbai in partial fulfilment of the requirement for the award of the degree of **"Bachelor of Engineering"** in **"Computer Engineering"**.

Signature of Supervisor/Guide

Name: Prof. K Jayamalini

Date: 26th April, 2019

Signature of the HOD

Name: Dr. Vinayak D. Shinde

Date: 26th April, 2019

Signature of the Principal

Name: Dr. S. Ram Reddy

Date: 26th April, 2019

Shree Rahul Education Society's (Regd.)
SHREE L. R. TIWARI COLLEGE OF ENGINEERING
Kanakia Park, Near Commissioner's Bungalow, Mira Road (East), Thane 401107, Maharashtra
(Approved by AICTE, Govt. of Maharashtra & Affiliated to University of Mumbai)
Tel. No.: 022-65295732 / 022-65142376 | Email: slrtce@rahuleducation.com | Website: www.slrtce.in

Project Report Approval

This project report entitled "**Emotion Detection from Tweets and Emoticons Using Machine Learning"** by **Jain Rajat Ashok, Kshatriya Sawan Ranjan, Dubey Arun Kumar Vyas, Shukla Pratyush Suresh** is approved for the degree of Bachelor of Engineering in Computer Engineering.

Examiners

1. Name: _____

   Signature: _____

2. Name: _____

   Signature: _____

Date: 26th April, 2019

Place: Mira Road

Acknowledgement

A few sublime human experiences defy expressions of any kind, and a feeling of true gratitude is one of them. We, therefore, find words quite inadequate to express our indebtedness to our Guide **Prof. K. Jayamalini** for her virtuous guidance, encouragement and help throughout this work. Their deep insight into the problem and the ability to provide solutions has been immense value in improving the quality of project at all stages. This experience of working with them shall ever remain a source of inspiration and encouragement for us.

We express our thanks to **Prof. Vinayak D. Shinde**, HOD (CS), SLRTCE,Mira Road , for extending his support that he gave truly help the progression of the project work.

Our sincere thanks to **Dr. S. Ram Reddy**, Principal, SLRTCE, Mira Road for providing me the necessary administrative assistance in the completion of the work. We are extremely grateful to the celebrated authors whose precious works have been consulted and referred in our project work. We also wish to convey our appreciation to our friends who provided encouragement and timely support in the hour of need.

Special thanks to our Parents whose love and affectionate blessings have been a constant source of inspiration in making this a reality.

All the thanks are, however, only fraction of what is due to Almighty for granting us an opportunity and the divine grace to successfully accomplish this assignment.

Jain Rajat Ashok

Kshatriya Sawan Ranjan

Dubey Arun Kumar Vyas

Shukla Pratyush Suresh

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN	Artificial Neural Network
ER	Entity Relationship
ML	Machine Learning
EDFTML	Emotion Detection from tweets and emoticons using ML
NN	Neural Network
GUI	Graphical User Interface
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory

Abstract

With the rise of social networking epoch, there has been a surge of user generated content. Micro blogging sites have millions of people sharing their thoughts daily because of its characteristic short and simple manner of expression. We propose and investigate a paradigm to mine the sentiment from a popular real-time micro blogging service, Twitter, where users post real time reactions to and opinions about “everything”. Social networking sites like twitter, Facebook, etc. are the great source of communication for internet users. So these become an important source for understanding the opinions, views or emotions of people. We extract data, i.e. tweets from Twitter in real time and apply machine learning techniques to convert them into a useful form and then use it for building sentiment classifier. Given a piece of written text, the problem is to categorize the text into one specific sentiment polarity i.e. positive, negative or neutral

Chapter 1

Introduction

1. Introduction

Here we will elaborate the aspects like description, problem statement, scope, motivation and also the objectives of our project

Description

Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis, which is also known as opinion mining, studies people's sentiments towards certain entities. Internet is a resourceful place with respect to sentiment information. From a user's perspective, people are able to post their own content through various social media, such as forums, micro-blogs, or online social networking sites. From a researcher's perspective, many social media sites release their application programming interfaces (APIs), prompting data collection and analysis by researchers and developers. For instance, Twitter currently has three different versions of APIs available, namely the REST API, the Search API, and the Streaming API. With the REST API, developers are able to gather status data and user information; the Search API allows developers to query specific Twitter content, whereas the Streaming API is able to collect Twitter content in real time. Moreover, developers can mix those APIs to create their own applications. Hence, sentiment analysis seems having a strong fundament with the support of massive online data.

Microblogging websites have evolved to become a source of varied kind of information. This is due to nature of micro blogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, 2 complain, and express positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these microblogs to get a sense of general sentiment for their product. Many time these companies study user reactions and reply to users on microblogs. One challenge is to build technology to detect and summarize an overall sentiment. Our project analyzes tweets by the peoples on certain products of companies or brands or performed by political leaders. In order to do this, we analyzed tweets from Twitter. Tweets are a reliable source of information mainly because people tweet about anything and everything they do including buying new products and reviewing them.

Problem statement

Despite the availability of software to extract data regarding a person's sentiment on a specific product or service, organizations still face issues regarding the data extraction. With the rapid growth of the World Wide Web, people are using social media such as Twitter which

generates big volumes of opinion texts in the form of tweets which is available for the sentiment analysis. This allows a huge volume of information from a human viewpoint which makes it difficult to extract sentences, read them, analyses them tweet by tweet, summarize them and organize them into an understandable format in a timely manner.

Informal language refers to the use of colloquialisms and slang in communication, employing the conventions of spoken language such as ‘would not’ and ‘wouldn’t’. Not all systems are able to detect sentiment from use of informal language and this could create a problem for the analysis and decision-making process. Emoticons are a pictorial representation of human facial expressions, which in the absence of body language serve to draw receiver's attention to the tenor or temper of a sender's nominal verbal communication, improving and changing its interpretation.

For example, ☺ indicates a happy state of mind and ☹ indicates a sad state of mind. Systems currently in place do not have sufficient data to allow them to draw feelings out of the emoticons. As humans often started using emoticons to properly express what they cannot put into words. Short-form is widely used even with short message service (SMS). The usage of short-form will be used more frequently on Twitter so as to help to minimize the characters used. This is because Twitter has put a limit on its characters to 140.

Sentiment analysis has turned out as an exciting new trend in social media with a large amount of practical applications that range from applications in business to government use. Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis is also known as opinion mining. Sentiment Analysis is used to classify the reviews using the sentiment of the words into positive or negative. Using the sentiment expressed in the words, opinions on any entity can be categorized into positive or negative. For example, the sentence, ‘I am not excited by this product though it is quite cheap’ expresses a negative sentiment about the product.

Scope and Motivation

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover, the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). This could be done by analyzing overall public sentiment towards that firm with respect to time and using economics tools for finding the public sentiment. Firms can also estimate how well their product

is responding in the market, which areas of the market is it having a favorable response and in which a negative response.

Project Objectives

Sentiment classification is a way to analyze the subjective information in the text and then determine the opinion. Sentiment analysis is the procedure by which information is extracted from the opinions and emotions of people in regards to entities, events and their attributes. In decision making, the opinions of others have a significant effect on customer's ease, making choices with regards to online shopping, choosing events, products, entities.

The objective of this project is to show how sentimental analysis can be used to recognize user's mood or emotion using tweets which has been extracted in real time. The learning algorithm will then learn what our emotions from statistical data then determine the emotion.

Chapter 2

Literature Review

2. Literature Review

Here we will elaborate the aspects like the literature survey of the project and what all projects are existing and been actually used in the market which the makers of this project took the inspiration from and thus decided to go ahead with the project covering with the problem statement.

2.1 Literature Survey Papers

SR NO.	YEAR &AUTHOR	TITLE	ABSTRACT
1.	Antonio Lopardo, Marco Brambilla 2018	Analyzing and Predicting the US Midterm Elections on Twitter with Recurrent Neural Networks	We propose a method and a system that aim to gauge local support for the two major US political parties in the 68 most competitive House of Representative districts during the mid-term elections. We analyze tweets explicitly posted from locations within each district. To distinguish between Republican and Democratic tweets, we adopt a **RNN-LSTM** binary classifier which reached validation accuracy of 85% over individual tweets, despite the highly implicit and short content shared on the social network. The method was able to predict the correct winner on 60% of the highly districts.
2.	Rohith.V, D.Malathi 2018	Sentiment Analysis On Twitter: A Survey	Sentiment Analysis deals with the polarity of the given context. Real-time scenarios of Sentiment analysis include Product Reviews, Irony and Sarcasm detection, Auto-Correct feature, and so on. The fundamental Sentiment analysis algorithms that perform feature extraction include **Naive-Bayes algorithm, Support Vector Machines and Natural Language Processing**. These methods can

			label either positive or negative, find the overall polarity and so on. We have written this Survey paper which elucidates in detail the existing methods for Sentiment analysis.
3.	Fenna Miedema Prof. dr. Sandjai Bhulai 2018	Sentiment Analysis with Long Short-Term Memory networks	Sentiment classification is a task where text documents are classified according to their sentiment. **Recurrent Neural Networks and Long Short-Term networks** are both models that are often used for sentiment analysis. The goal of this research is to find out why these models work well for sentiment analysis and how these models work. A shortcoming of the Recurrent Neural network is that it is only capable of dealing with short-term dependencies. Long Short-Term networks address this problem by introducing a long-term memory into the network. The model correctly classified **86.74%** of the reviews in the validation set.
4.	Ph.D. Candidate: Abdalraouf Hassan, Advisor: Ausif Mahmood 2017	Sentiment Analysis With Recurrent Neural Network And Unsupervised Neural Language Model	This paper describes a simple and efficient Neural Language Model approach for text classification that relies only on unsupervised word representation inputs. Our model employs **Recurrent Neural Network Long Short-Term Memory (RNN-LSTM)**, on top of pre-trained word vectors for sentence-level classification tasks. In our hypothesis we argue that using word vectors obtained from an unsupervised neural language model as an extra feature with RNN-LSTM for Natural Language Processing (NLP) system can increase the performance of the system.

5.	Bhumika Gupta, Monika Negi, Kanika Vishwakarma, Goldi Rawat, Priyanka Badhani 2017	Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python	Twitter sentiment analysis is an application of sentiment analysis on data from Twitter (tweets), in order to extract sentiments conveyed by the user. In the past decades, the research in this field has consistently grown. The reason behind this is the challenging format of the tweets which makes the processing difficult. The tweet format is very small which generates a whole new dimension of problems like use of slang, abbreviations etc. In this paper, we aim to review some papers regarding research in sentiment analysis on Twitter, describing the methodologies adopted and models applied, along with describing a generalized Python based approach.

2.2 Related Work

Sentiment Analysis has been an attractive object of study for AI researchers, computational linguists, cognitive scientists and neurobiologist The first papers on SA were focused on reviews for movies (Pang, Lee, and Vaithyanathan 2002), (Turney 2001), (Pang et al. 2002), (Pang and Lee 2005) and (Popescu and Etzioni 2005) or products (Hu and Liu 2004b), (Popescu and Etzioni 2005). Following those studies, other focused on the analysis of sales of products such as books, movies and videogames based on customer's opinions (Chevalier and Mayzlin 2006), (Mishne and Glance 2006), (Liu et al. 2007), (Zhu and Zhang 2010). With the rapid growth of social media like Twitter, more attention was drawn towards social media content as in (Jansen et al. 2009), (Asur and Huberman 2010), (Arias, Arratia, and Xuriguera 2013). In addition to Sentiment Analysis and the impact of people's opinion on sales, extensive research has been done on separate fields of finance and economics. As demonstrated by Lemmon and Portniaguina (2006) and Han (2008) there is a correlation between the sentiment and confidence of the investors and the stock market. Moreover [8] Gilbert and Karahalios (2010) show that "estimating emotions from weblogs provides novel information about future stock market

prices.", while Bollen, Mao, and Zeng (2011) explored the fact that national events affect people's emotions and the relationship of their emotions to the value of Dow Jones Industrial Average (DJIA).

Due to these findings, more work has been done in the last years on the subject (Oh and Sheng 2011; Zhang, Fuehres, and Gloor 2011; Makrehchi, Shah, and Liao 2013; Si et al. 2013; Smailović et al. 2013; Sprenger et al. 2014; Sprenger et al. 2014). Quoting Mitchell et al. (2013)"Companies should pay more attention to the analysis of sentiment related to their brands and products in social media communication as well as in designing advertising content that triggers emotions." 14 The ability to measure public opinion on social and political affairs is critical for political parties. The usual methods such as polls are expensive, they may not be accurate and the results are not representative of the public sentiment.

Overall polls are unreliable. In addition, getting people's opinion by asking questions is not the best method of collecting useful data. Thus, Sentiment Analysis in social media like Twitter may provide an alternative measure of public opinion and extract useful data [12] (Ceron, Curini, and Stefano 2012; O'Connor et al. 2010; [11] Stieglitz and DangXuan 2012; Zhou et al. 2013). For example, Diakopoulos et al. (2010) present an analysis of ephemeral changes of sentiment in reaction to the first U.S. presidential debate video in 2008. Further work has been done in other areas of sentiment analysis in social media. Sakaki et al. (2010) propose a method of detecting major events by analysing the stream text in Twitter and at Culotta (2010) propose methods of identifying influenza-related messages. Data from Twitter can be used to analyze public emotion, demography, health characteristics and the "geography of happiness" (Mitchell et al. 2013), a term describing the correlation of sentiment to place. Studying virality in Twitter and the correlation of viral messages with sentiment, [10] Hansen et al. (2011) showed that "news with negative sentiment is more likely to become viral, while in the non-news segment this is not the case".

2.3 Existing System

2.3.1 Exploring Human Emotion using Twitter

Sentiment analysis or opinion mining on twitter data is an emerging topic in research. In this model, they have described a system for emotion analysis of tweets using only the core text. Tweets are usually short, more ambiguous and contain a huge amount of noisy data; sometimes it is difficult to understand the user's opinion. [6] The main challenge is to feature extraction for the purpose of classification and feature extraction depends on the perfection of pre-processing of a tweet. The pre-processing is the most difficult task, since it can be done in various ways and the methods or steps applied in pre-processing are not distinct. Most of the researches in this topic, have been focused on binary (positive and negative) and 3-way (positive, negative and neutral) classifications. In this paper, we have focused on emotion classification of tweets as multi-class classification. [6] We have chosen basic human emotions (happiness, sadness, surprise, disgust) and neutral as our emotion classes. According to the experimental results, our approach improved the performance of multi-class classification of twitter data.

2.3.2 Tweet modelling with LSTM recurrent neural networks for hash tag recommendation

The hash symbol, called a hash tag, is used to mark the keyword or topic in a tweet. It was created organically by users as a way to categorize messages. [7] Hash tags also provide valuable information for many research applications such as sentiment classification and topic analysis. However, only a small number of tweets are manually annotated. Therefore, an automatic hash tag recommendation method is needed to help users tag their new tweets. Previous methods mostly use conventional machine learning classifiers such as SVM or utilize collaborative filtering technique. In this model a hash tag recommendation as a classification task but propose a novel recurrent neural network model to learn vector-based tweet representations to recommend hashtags. [7] Experiments on real world data from Twitter to recommend hashtags show that our proposed LSTM-RNN model outperforms state-ofthe-art methods and LSTM unit also obtains the best performance compared to standard RNN and gated recurrent unit (GRU).

2.4 Proposed System

The figure 2.3 depicts the flow chart of emotion detection from tweets and emojis using machine learning.

In following diagram, we first login with the valid login details i.e. username and password then we search the word that is to be searched. Afterword's we extract the data or tweets from the twitter using twitter API, we then preprocess the data from raw data to understandable data.

From there we use feature extracting process, Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power; also it may cause a classification algorithm to overfit to training samples and generalize poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Many machine learning practitioners believe that properly optimized feature extraction is the key to effective model construction.

After that we use the algorithm that we want to use in our project. Training data is the set of data used in an algorithm, training data means separating data into training and testing sets is an important part of evaluating data mining models. Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing. In machine learning, the study and construction of algorithms that can learn from and make predictions on data is a common task. Such algorithms work by making data-driven predictions or decisions through building a mathematical model from input data. Then the tweet classifier classifies the tweets according to its polarity and the process ends.

Algorithm Used:

Recurrent neural network (RNN) work as a powerful set of artificial neural network algorithm. A version of recurrent neural network works was used by DeepMind in their work playing video games with autonomous agents. Recurrent neural network work differs from feed forward neural network work because they include a feedback loop, whereby output from step x-1 is feedback to the neural network to affect the outcome of step x, and so forth for each subsequent step. For example, if a neural network is exposed to a word letter by letter, and it is asked to guess each

following letter, the first letter of a word will help determine what a recurrent neural network thinks the second letter will be, etc.

Long Short Term Memory neural network works – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. LSTM models are a variety of RNN. In RNN the prediction in sequence, where the hidden layer from one prediction is the hidden layer of the next prediction this will assign a memory to the neural network work, therefore, results 'from earlier estimation could lead to improve future predictions. LSTM gives RNN more features to an extreme control over memory; this aspect control how much the present input matters for forming the new memory, also how much the past memories matters in creating the new memory, and what parts of the memory are essential is producing the output.

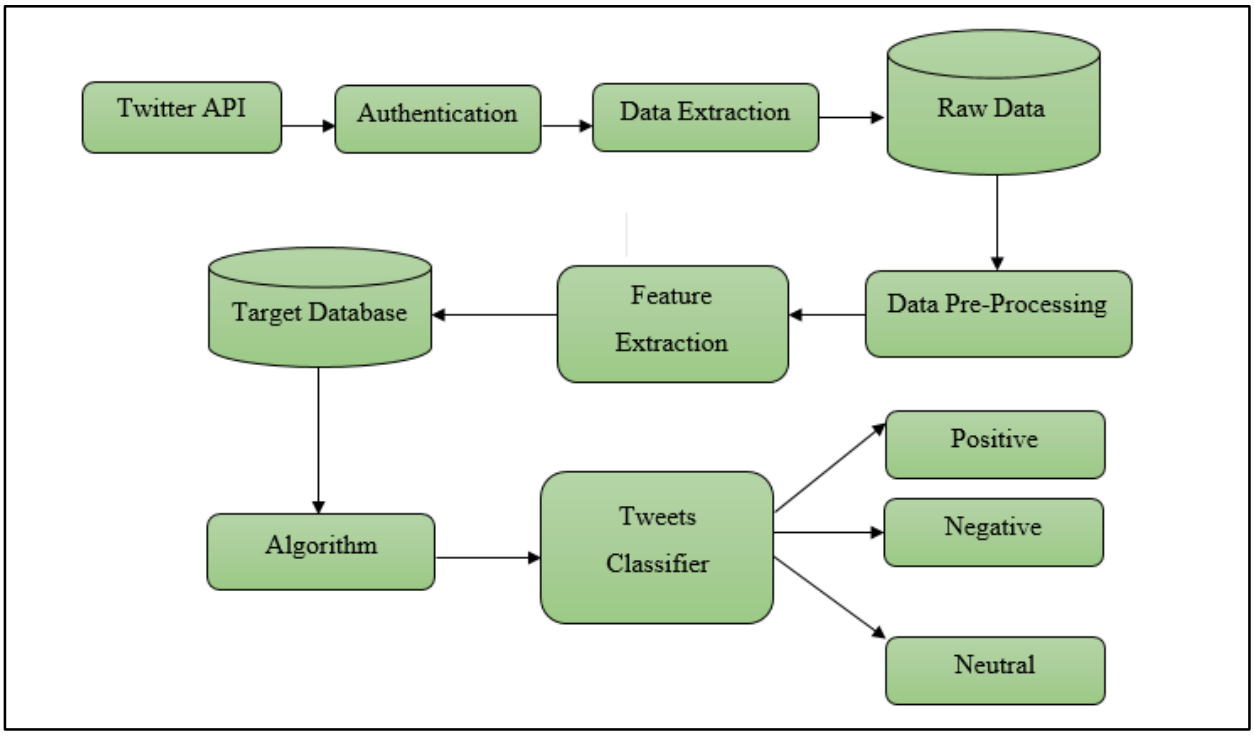

Figure 2.3 Block Diagram of EDFTML

Chapter 3

Requirement Analysis and Planning

3. Requirement Analysis and Planning

In requirements analysis encompasses those tasks that go into determining the needs or conditions to meet for a new or altered product or project, taking account of the possibly conflicting requirements of the various stakeholders, analyzing, documenting, validating and managing software or system requirements. Project planning is part which relates to the use of schedules such as Gantt charts to plan and subsequently report progress within the project environment. Initially, the project scope is defined and the appropriate methods for completing the project are determined

3.1 Functional Requirement

Functional requirement are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

- System should be able to process new tweets stored in database after retrieval
- System should be able to analyse data and classify each tweet polarity

3.2 Nonfunctional requirements

3.2.1 Performance

3.2.1.1 Real-Time: The software will provide up-to-date information, limited only by the rate of Twitter input.

3.2.1.2 System Resource Consumption: Resource consumption of this application should not reach an amount that renders the computer unusable. The system must be able to support multiple users. There's no initial target user base but the system should be able to support many users and allow for future expansion to different social networks such as Facebook & Google+. The response time of the system must allow for all actions to be performed in real time, the results must be available for review instantly.

3.2.2 Reliability

The reliability of our system is predicted to be very high and critical for proper functioning of the overall system. The software will meet all of the functional requirements without any

unexpected behavior. At no time should the output should display incorrect or outdated information without alerting the user to potential errors. The system must be available 24 hours a day 7 days a week as it may be used at any time, throughout many various locations to access all available options

3.2.3 Availability

The software will be available at all times for the users. The functionality of the software will depend on any external services such as internet access that are required.

3.3 System requirements

This section contains all of the functional and quality requirements of the system. It gives a detailed description of the system and all its features.

3.3.1 Hardware requirements

3.3.1.1: Since the application must run over the internet, all the hardware shall require to connect internet will be hardware interface for the system. As for e.g. Modem, WAN – LAN, Ethernet Cross-Cable.

3.3.1.2: Screen resolution of at least 800X600 is required for proper and complete viewing of screens. Higher resolution will be accepted.

3.3.1.3: Processor: Intel i3 and above
Ram: 2GB
Hard Disk: 500 GB

3.3.2 Software requirements

3.3.2.1 Inputs The software will receive input from two sources. First, the user interface and second, the Twitter API. The user interface will supply the keywords and the analysis session duration, while the Twitter API will supply the Tweet text.

3.3.2.2 Outputs The output will portray the current mood of the Twitter community. If available, historical data will be displayed in a graph.

3.3.2.3 Operating System The software will run on the Windows operating system.

3.4 Project Planning

3.4.1 Timeline Project

The following tables give the project plan for phase 1 and 2 of the project:

PHASE 1

Activity	Description	Effort in person weeks	Deliverable
Phase 1			
P-01	Information gathering	2 week	Requirement Gathering
P-02	Project analysis	1 week	Existing System Study & Literature Survey
P-03	Feasibility study	3 weeks	Feasibility Study
P-03	Documentation	1 week	Proposal Report
P-04	Designing	3 weeks	Module description
P-05	Project implementation	4 weeks	Implementation of important modules
	Total	14 weeks	

Table 0-1 Project Timeline Phase -1

PHASE 2

Table 0-2 Project Timeline Phase - 2

Activity	Description	Effort in person weeks	Deliverable
Phase 2			
P-01	Detailed Design	5 weeks	Documentation
P-02	Complete Implementation and Integration	5 weeks	Code and module of product release
P-03	Testing and Bug fixing	3 week	Test Report
P-04	Release/Deployment	Included in above	Product Release
P-05	Closure	One week	Documentation
	Total	14 weeks	

3.4.2 Task distribution

Table 0-3 Task Distribution

GROUP MEMBERS NAME	POSITION OF EVERY MEMBER	TASK PERFORMED BY MEMBER
Jain Rajat Ashok	Team Leader	Analysis, Designing, Testing and Coding.
Kshatriya Sawan Ranjan	Team Member	Analysis, Coding, Designing, Testing
Dubey Arun Kumar Vyas	Team Member	Concept, Documentation, Testing,
Shukla Pratyush Suresh	Team Member	Documentation, Designing, Testing.

Chapter 4

Analysis Modeling

4. Analysis Modeling

In this chapter, all the aspects of the proposed system will be covered in the diagrammatic manner and provides the detailed manner of the system.

4.1 Behavioural Modeling

UML behavioural diagrams visualize, specify, construct, and document the dynamic aspects of a system. The behavioural diagrams are categorized as follows: use case diagrams, interaction diagrams, state–chart diagrams, and activity diagrams.

4.1.1 Use Case Diagram

Fig 4.1 shows Use-Case Diagram

User in the Figure 4.1 needs to perform functionalities like:

1. Registration/Login: User need to first register using name and password and based on that user needs to login for using system.

2. Twitter Authentication: Authentication is done on the basis of 4 keys for using api to get live tweets.

3. Search word: user needs to enter a topic on which the live tweets will be downloaded and its polarity will be determined by model.

4. Data pre-processing/Feature extraction: In this data preprocessing like having uniform casing, remove urls, shorten slang words etc. is done and feature extraction like post tagging is done.

5. Opinion Identification: System will be trained using dataset and on live tweets it will predict its polarity.

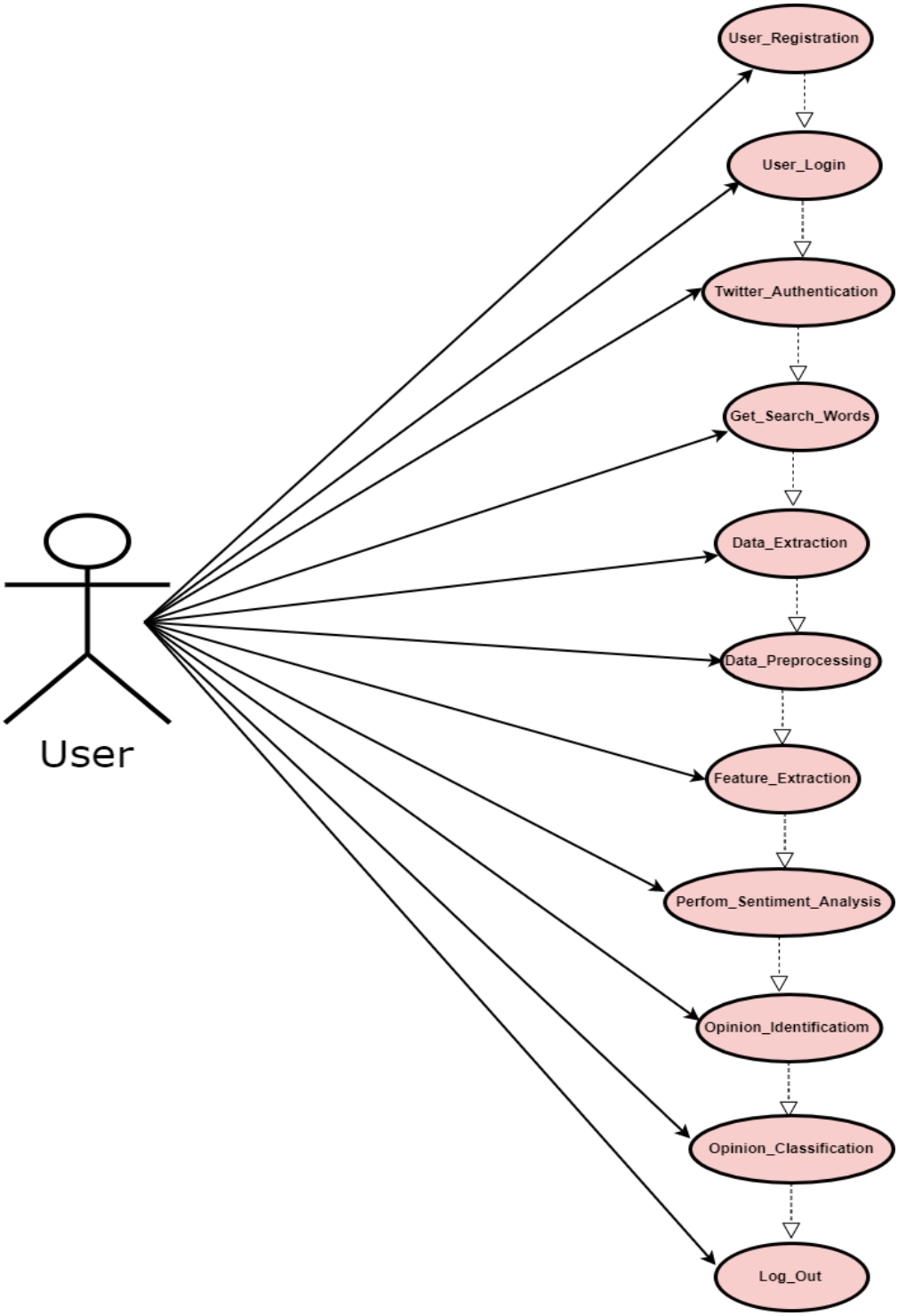

Figure 4.1 Use Case of EDFTML

4.1.2 Sequence Diagram

UML sequence diagrams are used to show how objects interact in a given situation. An important characteristic of a sequence diagram is that time passes from top to bottom: the interaction starts near the top of the diagram and ends at the bottom (i.e. Lower equals later). Here in Figure 4.2, there is one entity User.

User will first login in web application which will be checked from database. User will then search tweets of topic they want and then that related tweets will be stored in database. After that data pre-processing and feature extraction will be done and their results will be stored in database. Then sentiment analysis will be performed and users will get the opinion when it is completed. After that user can logout.

The sequence diagram is shown below**:**

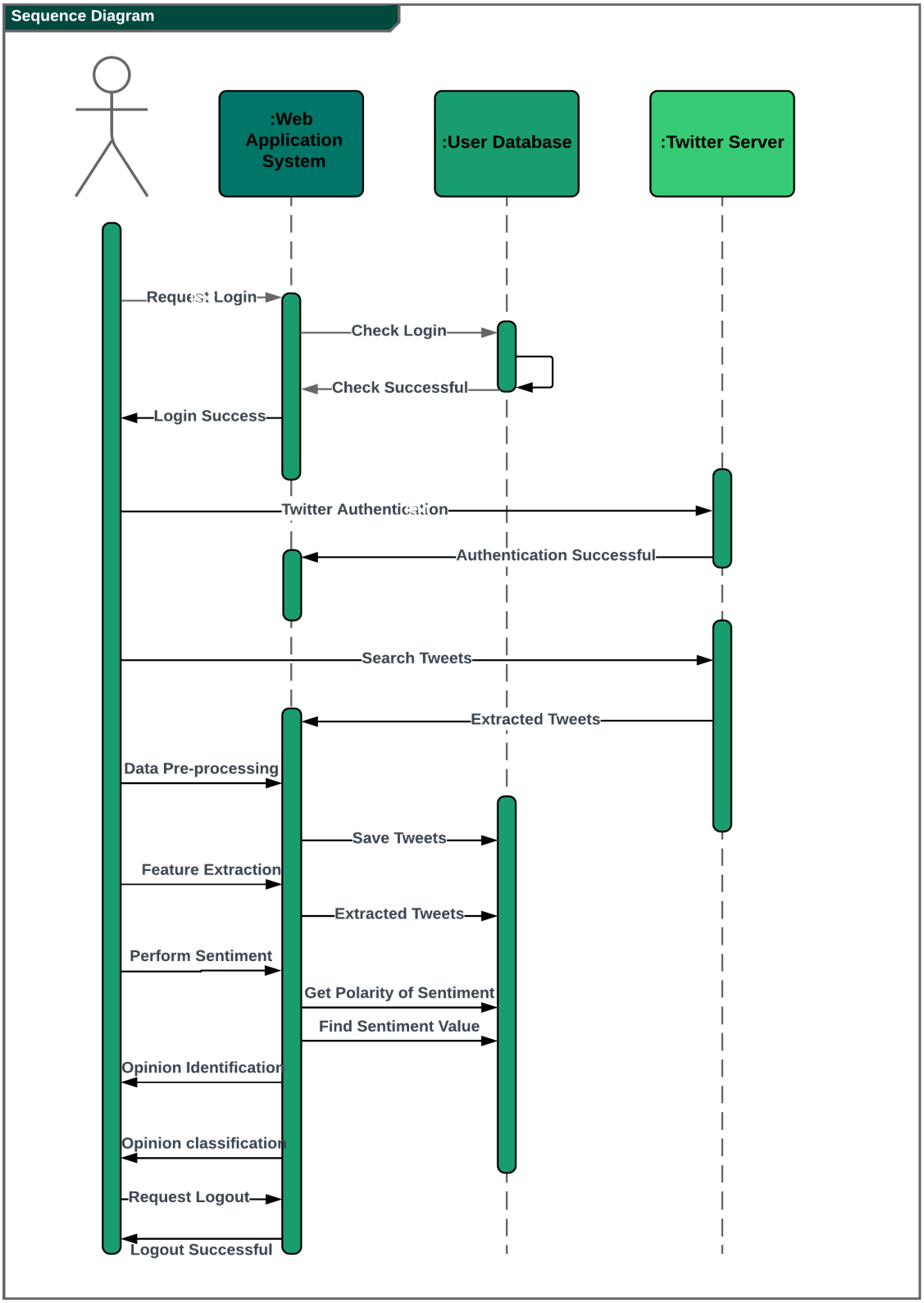

Figure 4.2 Sequence Diagram of EDFTML

4.1.3 Activity Diagram

It starts with starting point as shown in Figure 4.3, then the user will login which will be checked from database whether the entered username and password. Then the user will search word and after that data pre-processing and feature extraction will take place.

Based on that sentiment analysis will be done and once the user will get opinion identification then the user can logout.

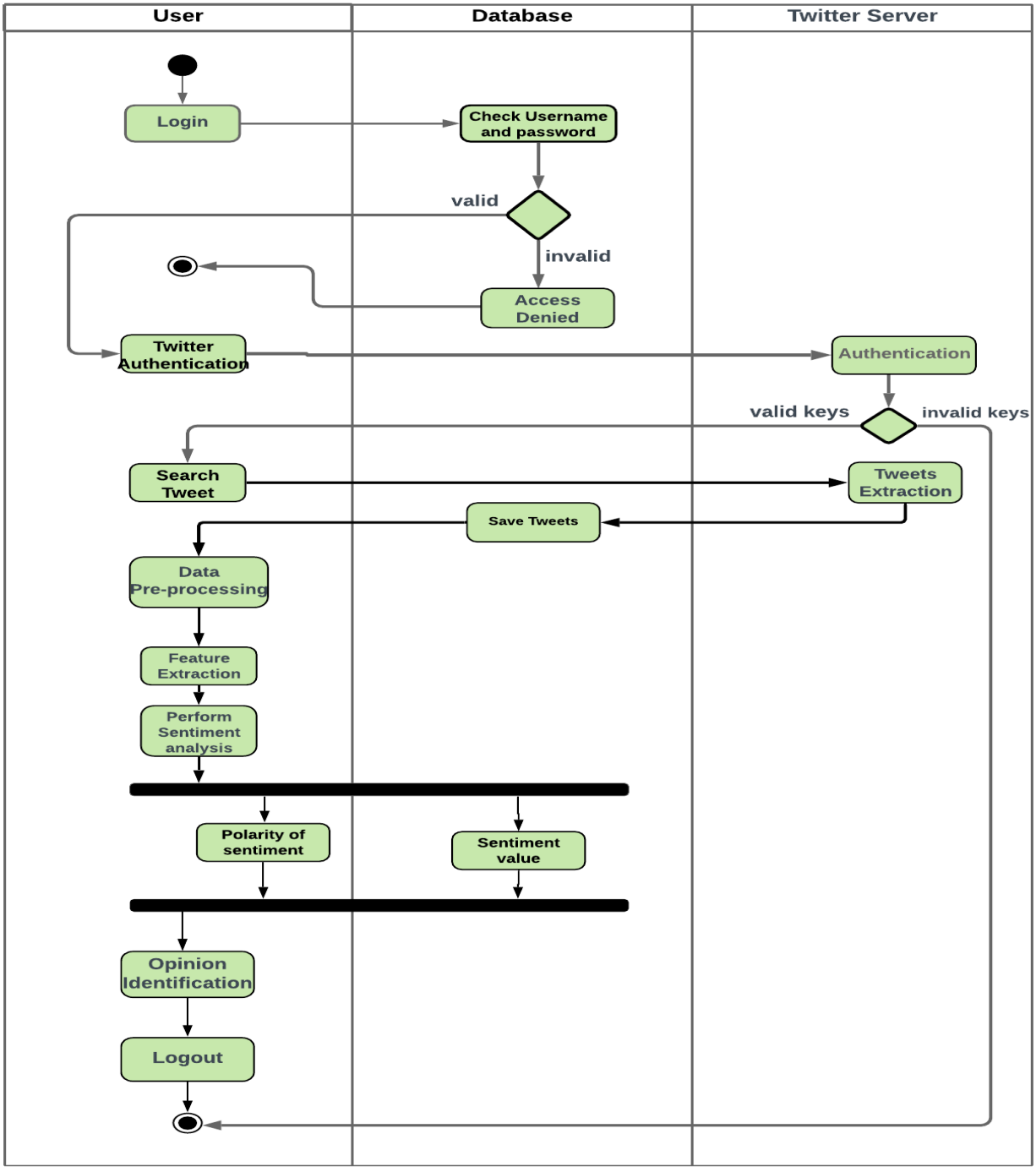

Figure 4.3 Activity Diagram of EDFTML

4.1.4 State Chart Diagram

Figure 4.4 represents various states that the system takes while processing. The foremost state is related to webcam which will start recording the video according to the timer when the timer is set. After the video is recorded the state is where the entire processing of the video takes place which will then lead to state where the system will calculate the drowsiness of the driver

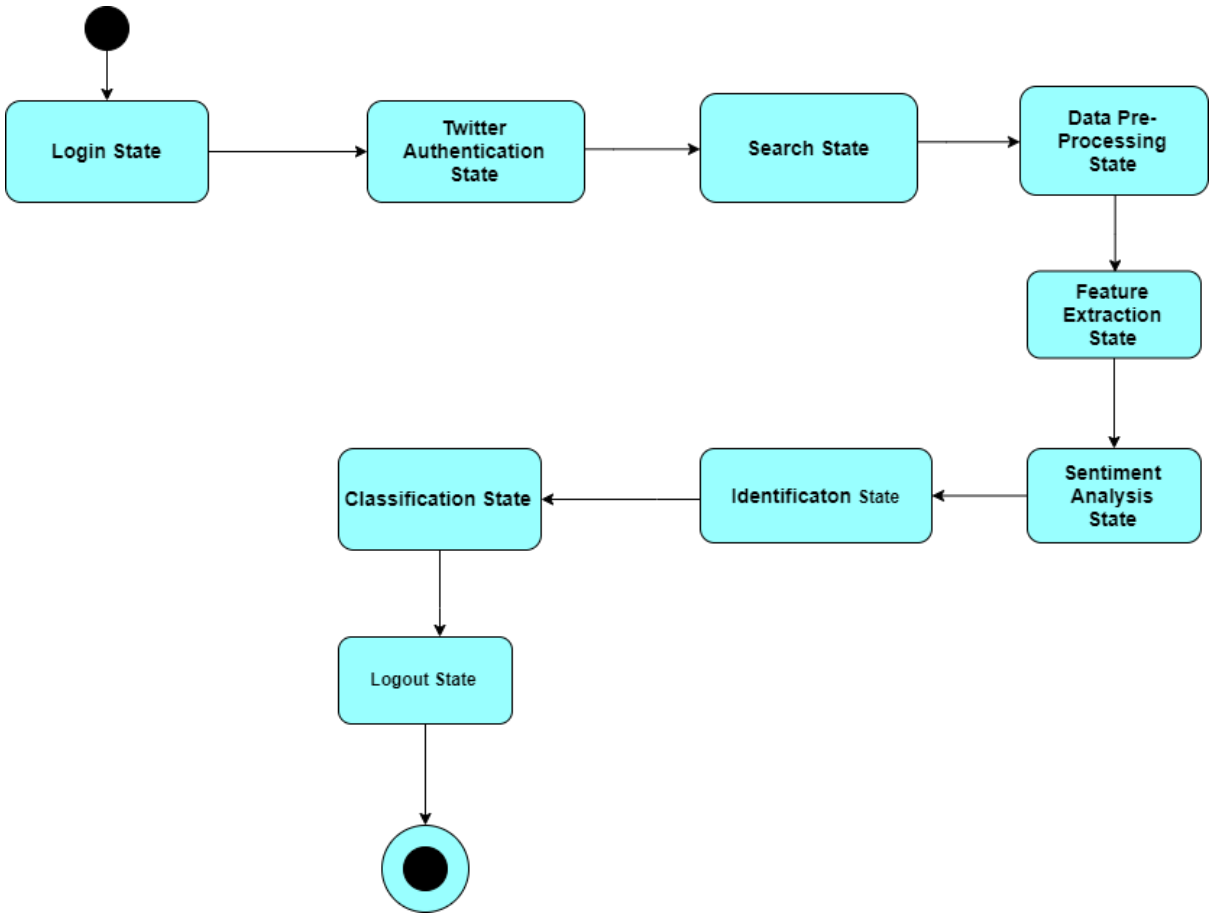

Figure 4.4 State Chart Diagram of EDFTML

4.2 Data Flow Diagram

4.2.1 DFD – Level 0

In Figure 4.5, there are various levels for emotion detection from tweets and emoticons. The level 0 indicates that the user logins into system and then search tweets topic which will be used for retrieving tweets and for data pre-processing and feature extraction. When sentiment analysis is applied the user will get back its result.

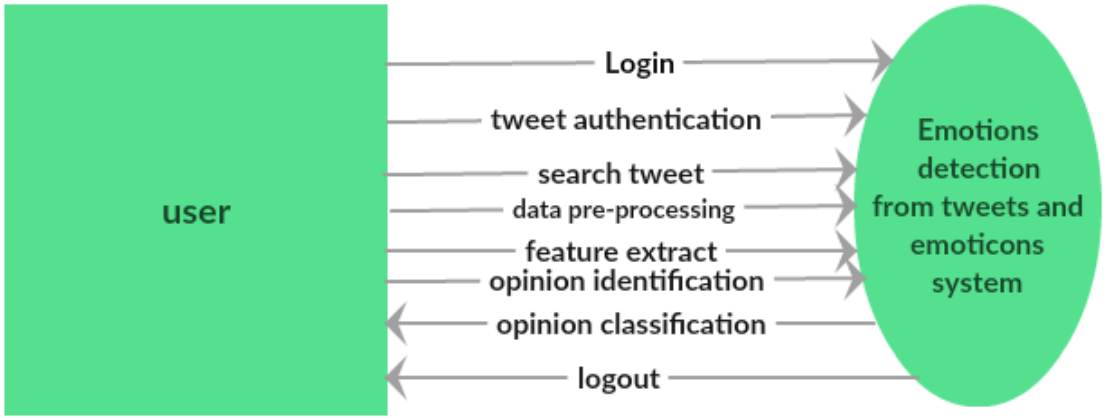

Figure 4.5 DFD level 0 of EDFTML

4.2.2 DFD - Level 1

Figure 4.6 contains next level i.e. level 1 of dfd contains 4 processes namely Login, data pre-processing, feature extraction, sentiment analysis. Login Process handles the Login Module (user). Data pre-processing handles search tweets which is stored in database once the tweets are retrieved. Based on the data pre-processing database feature extraction is done and stored in database through which sentiment analysis is done which users gets.

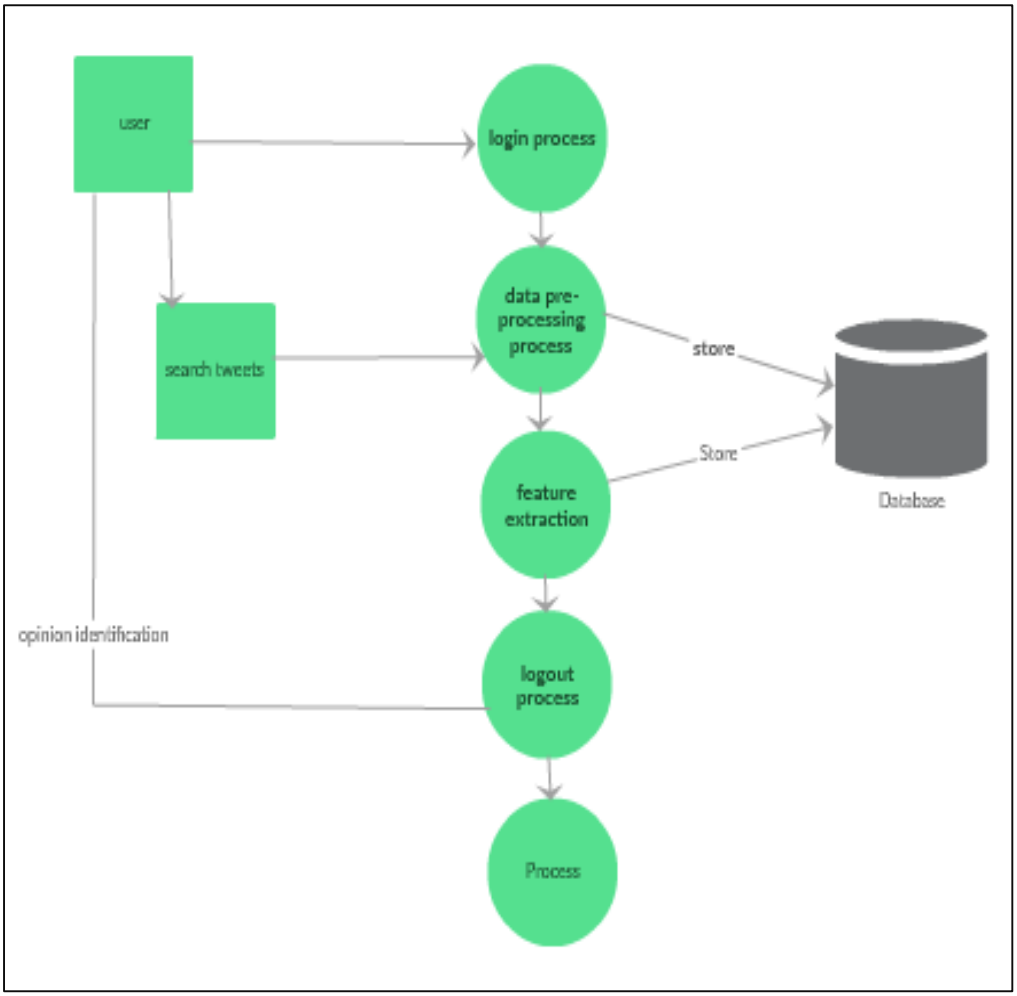

Figure 4.6 DFD level 1 of EDFTML

4.2.3 DFD – Level 2

In Figure 4.7 and figure 4.8, the level 2 of dfd explains the Data pre-processing and Sentiment analysis process in detail. The first process data pre-processing, in this user search tweets topic which will be retrieved from real time tweets data and based on that data pre-processing process is done which is cleaned and stored in database.

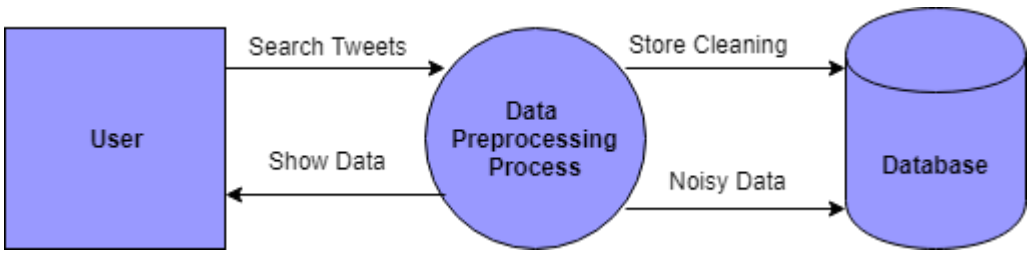

Figure 4.7 DFD level 2.1 of EDFTML

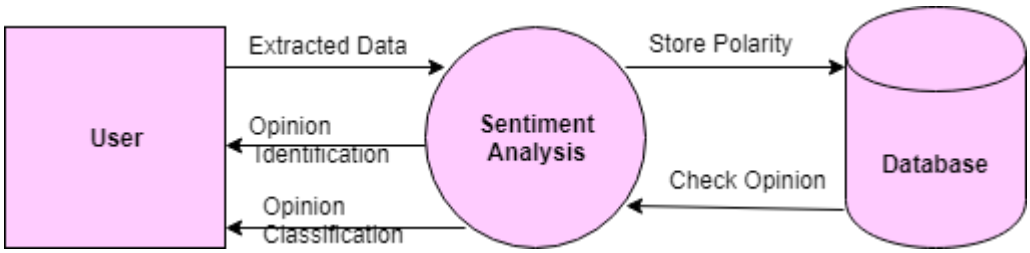

Figure 4.8 DFD level 2.2 of EDFTML

The second process Sentiment analysis process, in this extracted data is passed and sentiment analysis is done which will then store polarity of tweets in database and then user can know about opinion of tweets through opinion identification.

4.3 Database Design

4.3.1 ER Diagram

In Figure 4.9, ER diagram for emotion detection from tweets and emoticons is shown. ER stands for Entity Relationship. This diagram indicates various entities involved and how these entities are related to each other. The entities involved in our system are Registration, Login, Tweets, Tweets sentiment value and Tweets polarity. The communication between these entities is shown in the figure below.

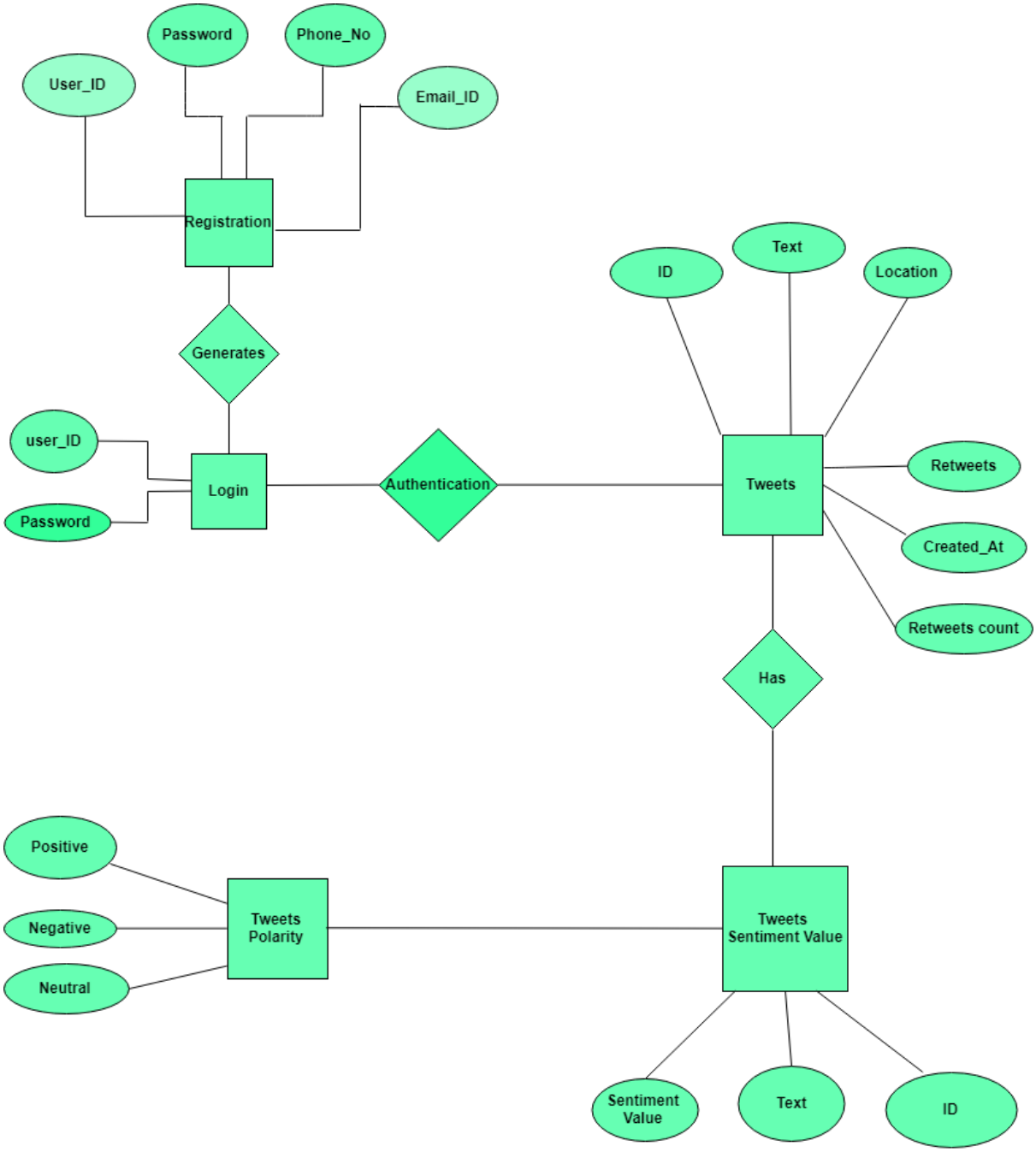

Figure 4.9 ER Diagram of EDFTML

Chapter 5

System Design

5. System Design

Design will elaborate the process of describing, organising and structuring the components of the system both at the architectural level and at the detailed level.

5.1 Architecture

The Figure 5.1 shows architecture of the system. The sentences that represent observations or attitude that is expressed as positive or negative are called as sentiments. The following figure 5.1 describes an overall system design of twitter sentiment analysis. First the user registers in the system by creating an account and then login using username and password to access system. The users post their tweets in twitter. These tweets are then extracted in real time using twitter API in the form of raw data which are then saved in database.

The raw dataset is converted into target dataset through Data Pre-Processing and Feature Extraction. The features of the words are selected and then machine learning techniques are applied on extracted features to classify them into its sentiment polarity that is namely positive, negative.

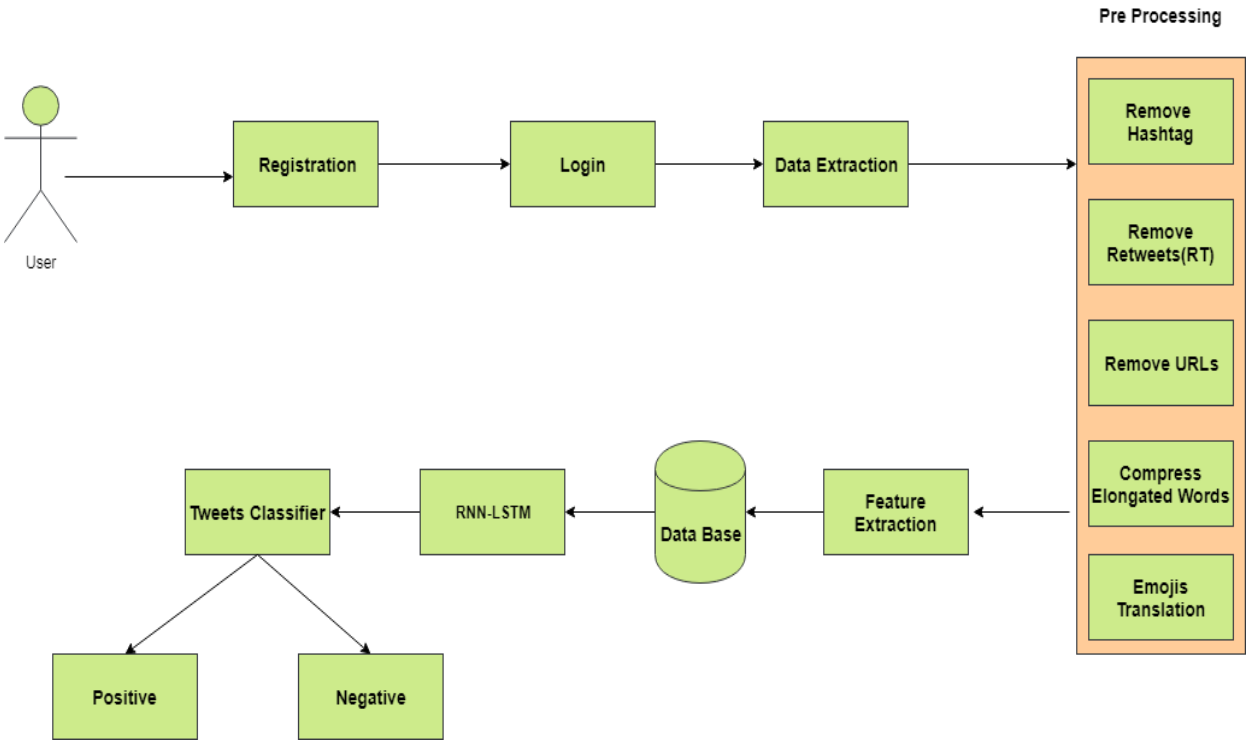

Figure 5.1 Architecture of EDFTML

5.2 Class Diagram

The class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing and documenting different aspects of a system but also for constructing executable code of the software application.

The class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object oriented systems because they are the only UML diagrams which can be mapped directly with object oriented languages.

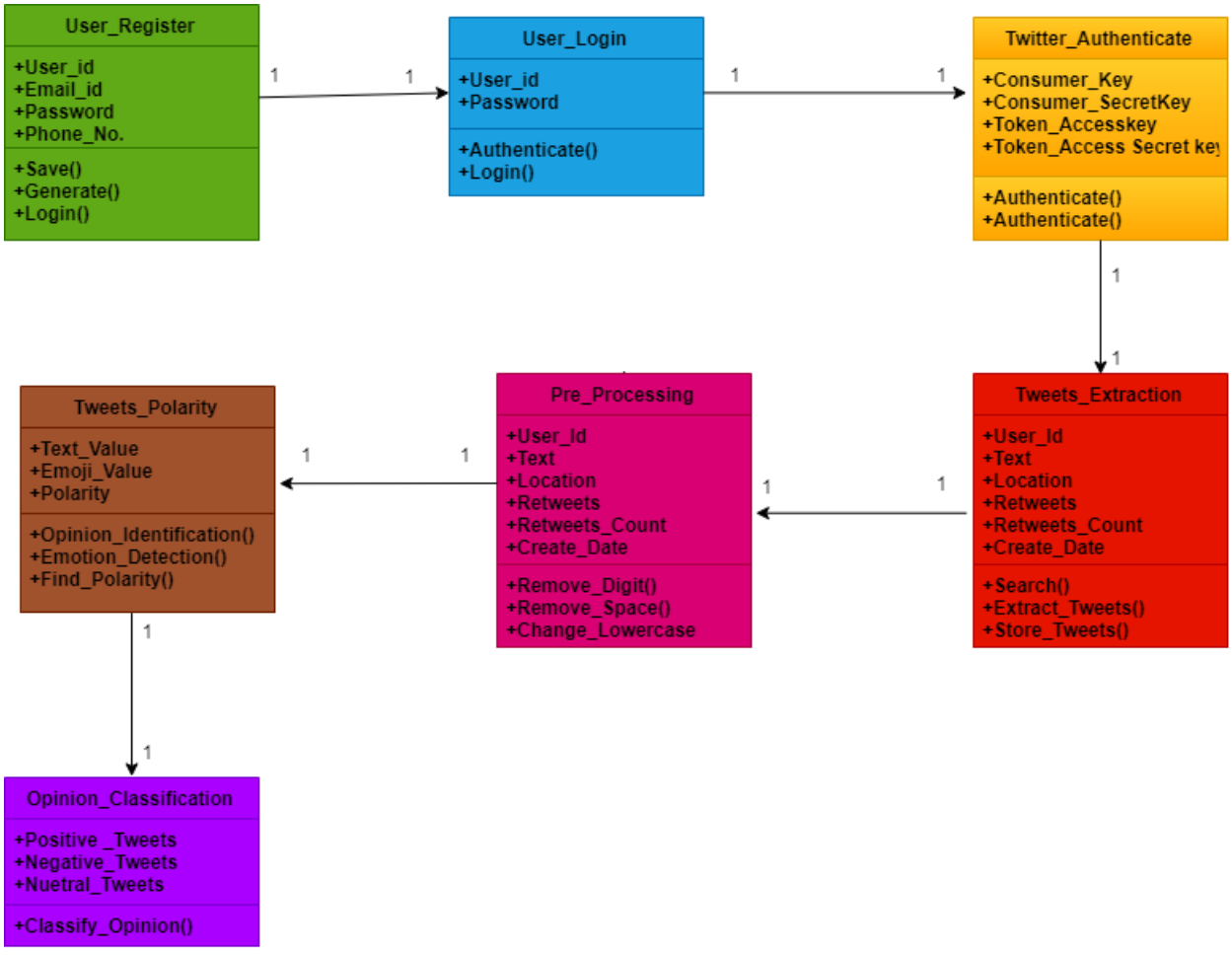

Fig.5.2 Class Diagram of EDFTML

5.3 Object Diagram

Figure 4.6 shows a static view of the structure of twitter sentiment analysis. Here as we can see, there are various objects UR1, UL1, TA1, TP1, PP1, TE1, OC1 are few examples to be named. There are various attributes associated with each object, for example, for the object UR1, there are attributes like User_id, Email_id, Password, Phone_no Similarly there are attributes for other objects as well. Here the object diagram is used to render a set of objects and their relationships as an instance.

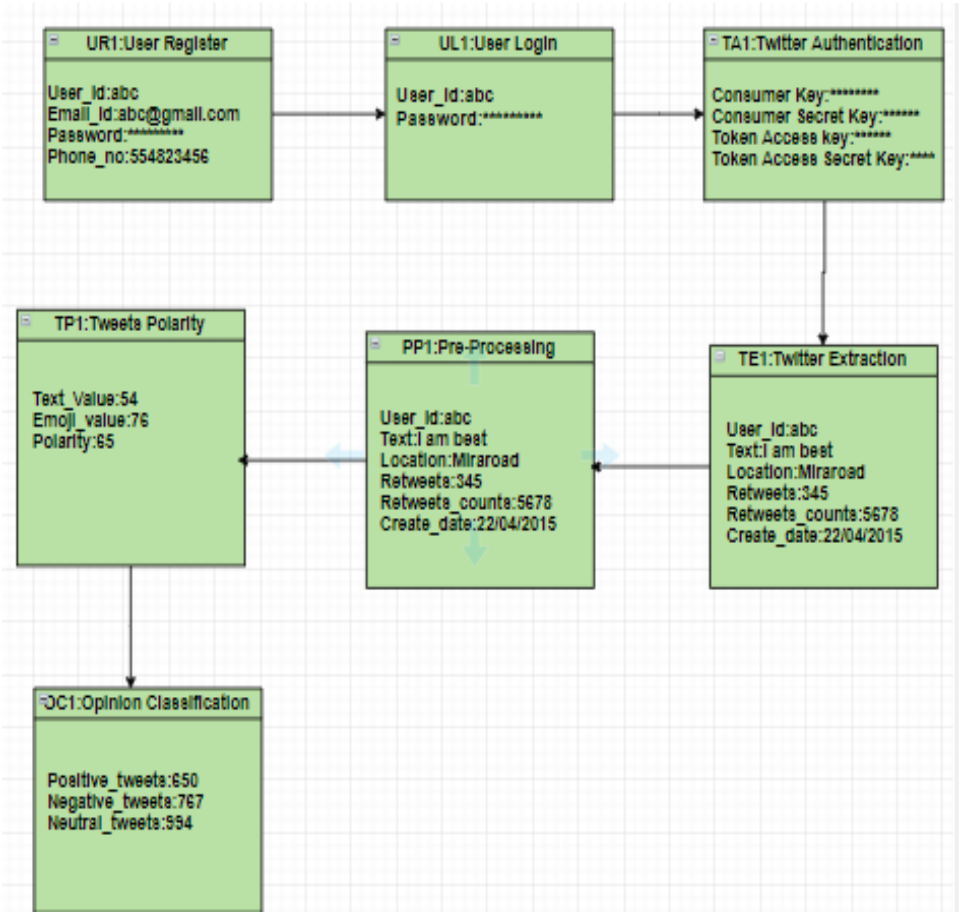

Figure 5.3 Object diagram of EDFTML

5.4 Collaboration Diagram

From the name Interaction it is clear that the diagram is used to describe some type of interactions among the different elements in the model. So this interaction is a part of dynamic behaviour of the system. This interactive behaviour is represented in UML by two diagrams known as Sequence diagram and Collaboration diagram. The basic purposes of both the diagrams are similar. The purpose of interaction diagram is**:**

To capture the dynamic behaviour of system. To describe the message flow in the system. To describe the structural organization of the objects. To describe the interaction among objects.

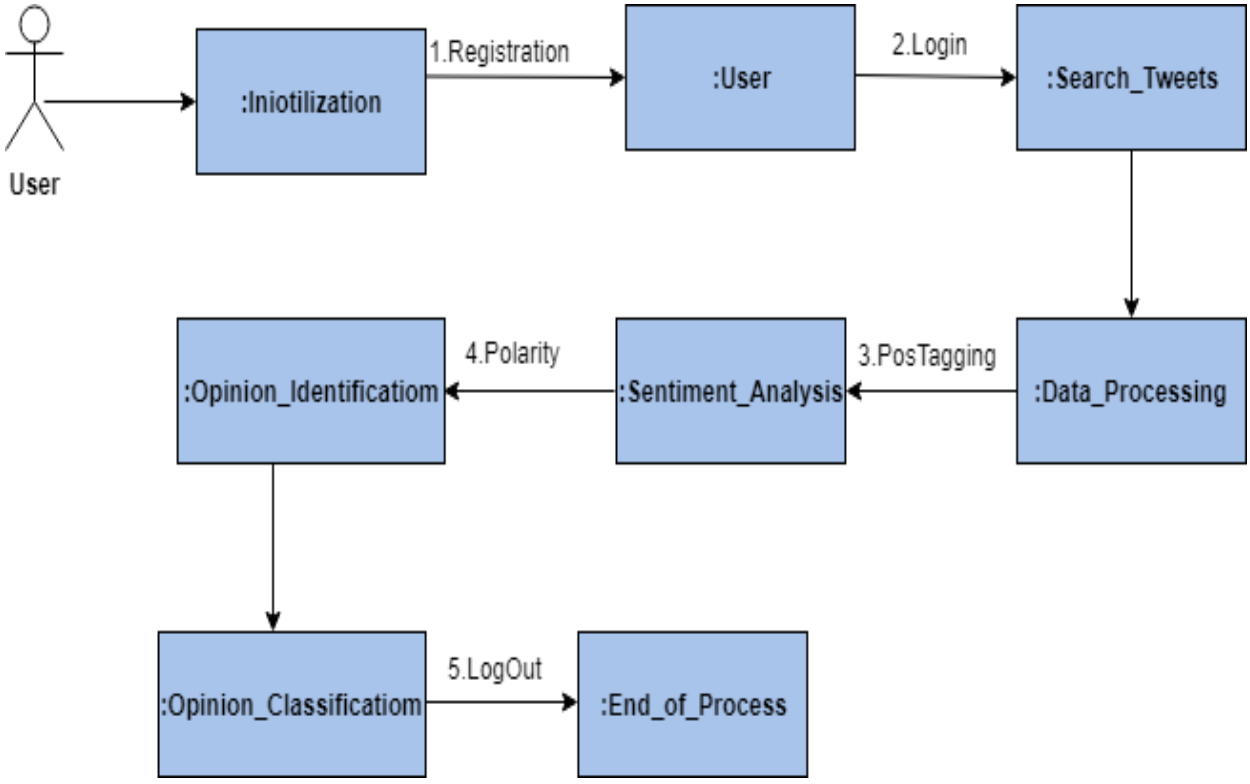

Fig 5.4 Collaboration diagram of EDFTML

5.5 Component Diagram

Component diagram is a special kind of diagram in UML. It does not describe the functionality of the system but it describes the components used to make those functionalities. Component diagrams are used during the implementation phase of an application. However, it is prepared well in advance to visualize the implementation details.

Here fig 5.5 component diagram shown below

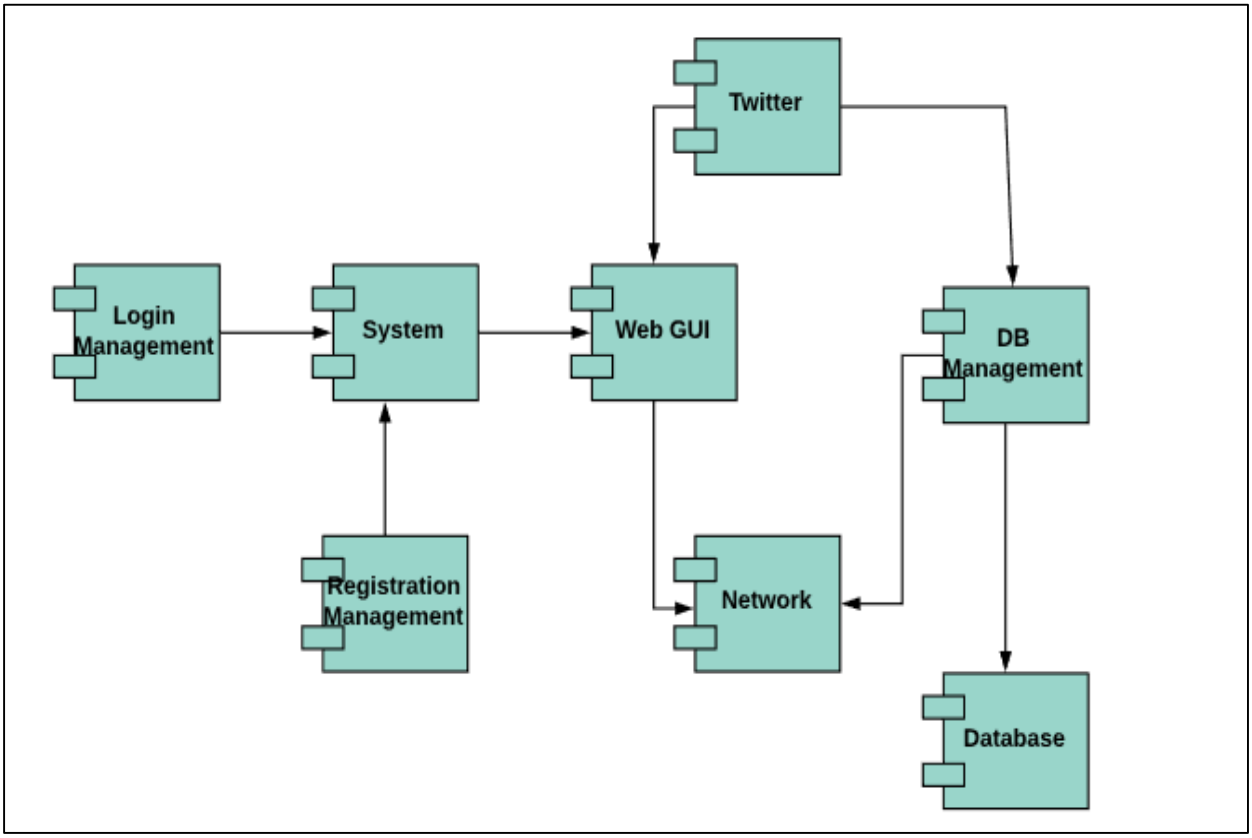

Fig. 5.5 Component Diagram of EDFTML

5.6 Deployment Diagram

The name Deployment itself describes the purpose of the diagram. Deployment diagrams are used for describing the hardware components where software components are deployed. Component diagrams and deployment diagrams are closely related.

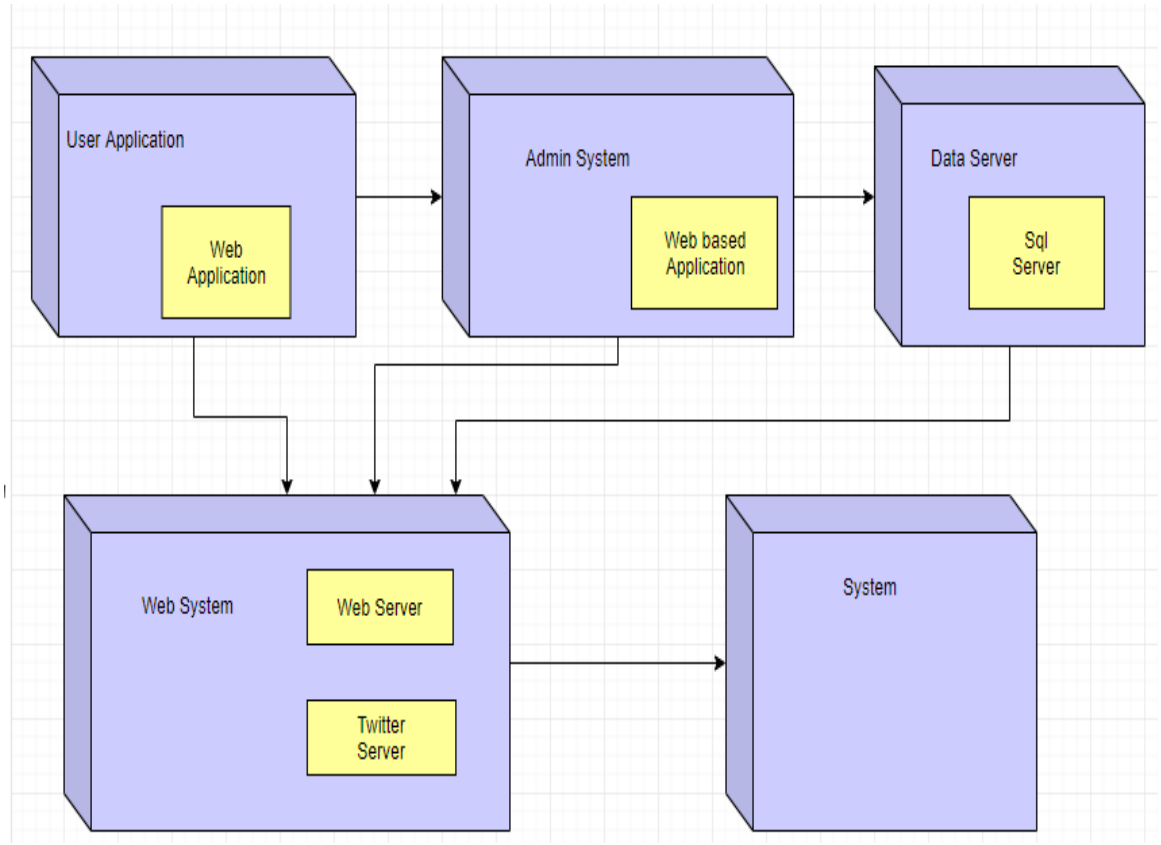

Fig. 5.6 Deployment Diagram of EDFTML

5.7 Graphical User Interface

In fig 5.7 shows the registration window which the users need to fill before using our model Emotion Detection from tweets.

All the information of the users will be then saved in database.

Fig 5.7 Registration Window

Once the user's information is stored in database, Users need to login with ID given to them.

Fig 5.8 Login Window

In figure 5.9 it shows an interface in which we manually enter a tweet for which our model shows polarity whether it is positive or negative.

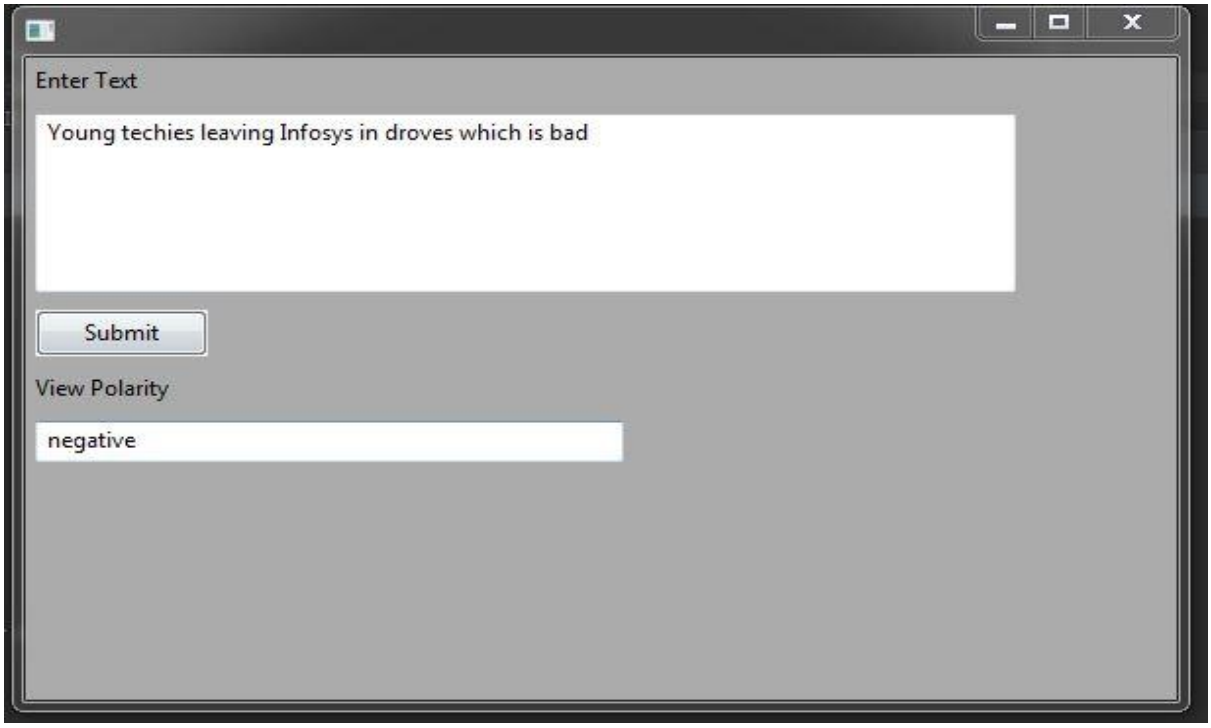

Figure 5.9 Result

Chapter 6

Implementation

6. Implementation

This chapter will give all the implementation details and methods used to implement the application in brief.

6.1 Using Tweepy

Tweepy is an open source python library, which provides a convenient way for accessing the Twitter API.

Simply put, here's how we can interact with the Twitter API; we can:

- Post a tweet
- Get timeline of a user, with a list of latest tweets
- Send and receive direct messages
- Search for tweets and much more

This library ensures that we can easily do these operations, and it also ensures the security and privacy of a user – for which we naturally need to have OAuth credentials configured.

```
18    import tweepy
19    from tweepy import OAuthHandler
```

```
33    consumer_key = 'UEpQIRozu0bMxhz6AjxpsfV9e'
34    consumer_secret = 'LqqjLzJMxqnTvVPEi8GX6Gdf8ifUQEgFaueKHGoCWQZmneOCqd'
35    access_token = '2839725024-4nFPskpjeY5mpRDkD2NqSMwITKxAzD9ftA9IPGZ'
36    access_token_secret = 'VZInAR71y7w3mZLF5Sr0YkSdAouxC7gYe1Ba2HUE7YMzn'
37
38    auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
39    auth.set_access_token(access_token, access_token_secret)
40    api = tweepy.API(auth)
41
```

Figure 6.1 Snapshot of Tweepy

6.2 Register/Login

In this the user first registers their name and password and then login to get access to our future modules.

```
42    with sqlite3.connect('quit.db') as db:
43        c = db.cursor()
44
45    c.execute('CREATE TABLE IF NOT EXISTS user (username TEXT NOT NULL ,password TEX NOT NULL);')
46    db.commit()
47    db.close()
48
```

```
    # Login Function
    def login(self):
        # Establish Connection
        with sqlite3.connect('quit.db') as db:
            c = db.cursor()

        # Find user If there is any take proper action
        find_user = 'SELECT * FROM user WHERE username = ? and password = ?'
        c.execute(find_user, [(self.username.get()), (self.password.get())])
        result = c.fetchall()
        if result:
            self.logf.pack_forget()
            self.head['text'] = self.username.get() + '\n Loged In'
            self.head['pady'] = 150
            main()
        else:
            ms.showerror('Oops!', 'Username Not Found.')

    def new_user(self):
        # Establish Connection
        with sqlite3.connect('quit.db') as db:
            c = db.cursor()

        # Find Existing username if any take proper action
        find_user = 'SELECT * FROM user WHERE username = ?'
        c.execute(find_user, [(self.username.get())])
        if c.fetchall():
            ms.showerror('Error!', 'Username Taken Try a Diffrent One.')
        else:
            ms.showinfo('Success!', 'Account Created!')
            self.log()
```

Figure 6.2 Snapshot of Login

6.3 Select file for training

In this the user must select file which contains tweets which is around 15,000 tweets used for training neural network.

10	Negative	Deer in the headlights RT @lizzwinstead: Ben Carson, may be the only brain surgeon who has performed a lobotomy on himself. #GOPDebate
11	Negative	RT @NancyOsborne180: Last night's debate proved it! #GOPDebate #BATsAsk @BadassTeachersA #TBATs https://t.co/G2gGjY1bJD
12	Negative	@JGreenDC @realDonaldTrump In all fairness #BillClinton owns that phrase.#GOPDebate
13	Positive	RT
14	Negative	Me reading my family's comments about how great the #GOPDebate was http://t.co/glaGjPygXZ
15	Neutral	RT @ArcticFox2016: RT @AllenWestRepub "Dear @JebBush #GOPDebate #NotAMistake http://t.co/TtFG7KYcd9"
16	Positive	RT @pattonoswalt: I loved Scott Walker as Mark Harmon's romantic rival in SUMMER SCHOOL. Look it up. #GOPDebate
17	Negative	Hey @ChrisChristie exploiting the tragedy of 9/11 for your own political gain is @rudygiulianiGOP's thing #GOPDebate
18	Negative	RT @CarolCNN: #DonaldTrump under fire for comments about women @PeterBeinart @SL_Schaeffer @IWF @@MyRkiger weigh in on #GOPdebate h
19	Negative	RT @johncardillo: Guess who had most speaking time at the #GOPDebate. @FoxNews moderators with 31.7% of time. http://t.co/2WSUT0c0Lx
20	Negative	reason comment is funny 'in case you're ignorant' is the #gop #tcot are the reason the government isn't working for the people #gopdebate
21	Negative	RT @PamelaGeller: Huckabee: Paying for transgender surgery for soldiers, sailors and airmen does not make our country safer #Ha #GOPDebate
22	Positive	RT @ChuckNellis: Cruz has class & truth, that gets my vote! #GOPDebate
23	Negative	RT @mchamric: RT â€œ@TeaTraitors: #GOPDebate was still Clown Show! I'm glad Head Clown Trump helping destroy GOP. http://t.co/nwGx8G8JWrâ
24	Negative	RT @erinmallorylong: No *I* hate Planned Parenthood and women more! NO I HATE PLANNED PARENTHOOD AND WOMEN MORE!!!!! #GOPDebate

Figure 6.3 Snapshot of dataset

6.4 Extracting Tweets on Specific Topic

The Sample screen shot of this module is shown in Figure 6.4.

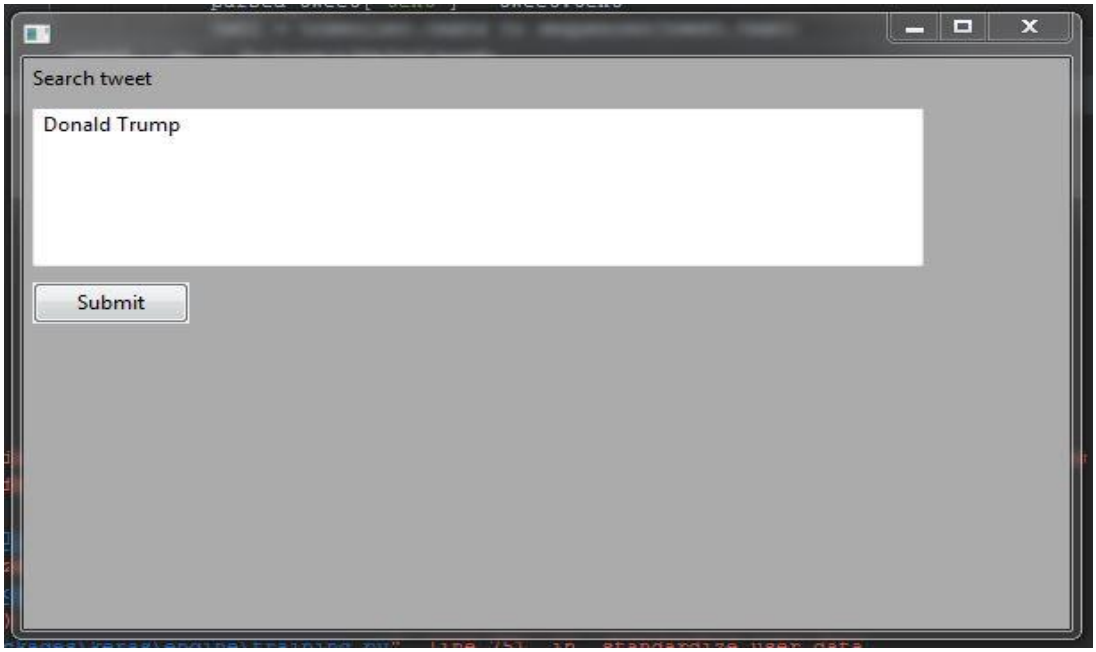

Figure 6.4 Search Window

```
try:
    fetched_tweets = api.search(q=txt_search, count=200)
    for tweet in fetched_tweets:
        parsed_tweet = {}
        senti_value = ''
        parsed_tweet['text'] = tweet.text
        tweets_live = tweet.text
        twt1 = tokenizer.texts_to_sequences(tweet.text)
        twt1 = pad_sequences(twt1, maxlen=30, dtype='int32', value=0)
        sentiment = model.predict(twt1, batch_size=1, verbose=2)[0]

        if (np.argmax(sentiment) <= 0.3):
            print(tweets_live + '--->> NEGATIVE')
            n_cnt += 1
            senti_value = 'NEGATIVE'

        elif (np.argmax(sentiment) == 1):
            print(tweets_live + '--->> POSITIVE')
            p_cnt += 1
            senti_value = 'POSITIVE'
        data_to_save.append([tweets_live, senti_value])
        df = pd.DataFrame(data_to_save)
        df.to_csv('result.csv')

except tweepy.TweepError as e:
    print("Error : " + str(e))

print("Final_Result_Positive--->" + str(p_cnt))
print("Final_Result_negative--->" + str(n_cnt))
```

Figure 6.5 Snapshot for Fetching tweets and predicting polarity

6.5 RNN-LSTM

The Sample screen shot of this module is shown in Figure 6.6

```
max_fatures = 4000
tokenizer = Tokenizer(num_words=max_fatures, split=' ')
tokenizer.fit_on_texts(data['text'].values)
X = tokenizer.texts_to_sequences(data['text'].values)
X = pad_sequences(X)

embed_dim = 128
lstm_out = 196

model = Sequential()
model.add(Embedding(max_fatures, embed_dim, input_length=X.shape[1]))
model.add(SpatialDropout1D(0.4))
model.add(LSTM(lstm_out, dropout=0.2, recurrent_dropout=0.2))
model.add(Dense(2, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
print(model.summary())

Y = pd.get_dummies(data['sentiment']).values
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.33, random_state=42)
print(X_train.shape, Y_train.shape)
print(X_test.shape, Y_test.shape)

batch_size = 64
model.fit(X_train, Y_train, epochs=7, batch_size=batch_size, verbose=2)

validation_size = 1500

X_validate = X_test[-validation_size:]
Y_validate = Y_test[-validation_size:]
X_test = X_test[:-validation_size]
Y_test = Y_test[:-validation_size]
```

Figure 6.6 Snapshot for LSTM

Chapter 7

Testing

7. Testing

In this section we mention the various tests conducted to test our application. It also explains various test Cases used to test the application.

7.1 Test Cases

Test Case Id: 01

Test Objective: To test the Registration Module

Item No	Test Condition	Operator Action	Input Specification	Output Specification (Expected Results)	Pass or Fail
1	Successful. Registration	1. Insert User name and Password. 2. Press Create Account.	User name and Password	1. System saves the User name and Password In database	Pass
2	Unsuccessful Registration	1. Insert User name and Password. 2. Press Login button.	User name and Password	1. System checks the User name and Password and Pops up message “Account Present”.	Pass

Test Case Id: 02

Test Objective: To test the Login Module

Item No	Test Condition	Operator Action	Input Specification	Output Specification (Expected Results)	Pass or Fail
1	Successful Login.	1. Insert User name and Password. 2. Press Login button.	User name and Password	1. System validates the User name and Password & provides access to the system & redirects to the Home Page	Pass
2	Unsuccessful Login due to Incorrect password.	1. Insert User name and Password. 2. Press Login button.	User name and Password	1. System validates the User name and Password and Pops up message “Invalid User ID or password”.	Pass

Test Case Id: 03

Test Objective: Input Search topic

Item No	Test Condition	Operator Action	Input Specification	Output Specification(Expected Results)	Pass or Fail
1	Input Search Topic	To find tweets Based on that topic	Name	Prediction on live tweets	Pass
2	Incorrect prediction Due to missing Value	Skip one field Of input	Null value	Error due to no Live tweets fetched	fail

7.2 Types of Testing Used

7.2.1 Black Box Testing

BLACK BOX TESTING, also known as Behavioural Testing is a software testing method in which the internal structure/design/implementation of the item being tested is not known to the tester. These tests can be functional or non-functional, though usually functional.

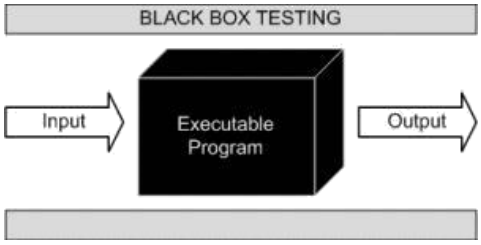

Fig 7.1 Black box Testing

This method is named so because the software program, in the eyes of the tester, is like a black box; inside which one cannot see. This method attempts to find errors in the following categories:

Incorrect or missing functions

Interface errors

Errors in data structures or external database access

Behaviour or performance errors

Initialization and termination errors

E.g.: A tester, without knowledge of the internal structures of a website, tests the web pages by using a browser; providing inputs (clicks, keystrokes) and verifying the outputs against the expected outcome.

7.2.2 White Box Testing

WHITE BOX TESTING (also known as Clear Box Testing, Open Box Testing, Glass Box Testing, Transparent Box Testing, Code-Based Testing or Structural Testing) is a software testing method in which the internal structure/design/implementation of the item being tested is known to the tester. The tester chooses inputs to exercise paths through the code and determines the appropriate outputs. Programming know-how and the implementation knowledge is essential. White box testing is testing beyond the user interface and into the nitty-gritty of a system.

This method is named so because the software program, in the eyes of the tester, is like a white/transparent box; inside which one clearly sees.

E.g.: A tester, usually a developer as well, studies the implementation code of a certain field on a webpage, determines all legal (valid and invalid) AND illegal inputs and verifies the outputs against the expected outcomes, which is also determined by studying the implementation code.

Chapter 8

Results

8. Results

8.1 Loading Dataset for Training

This chapter contains results of final product module wise. It contains snapshots of different modules of the system.

Figure 8.1 shows the interface for selecting file in which we have tweets i.e. dataset for training our model

.

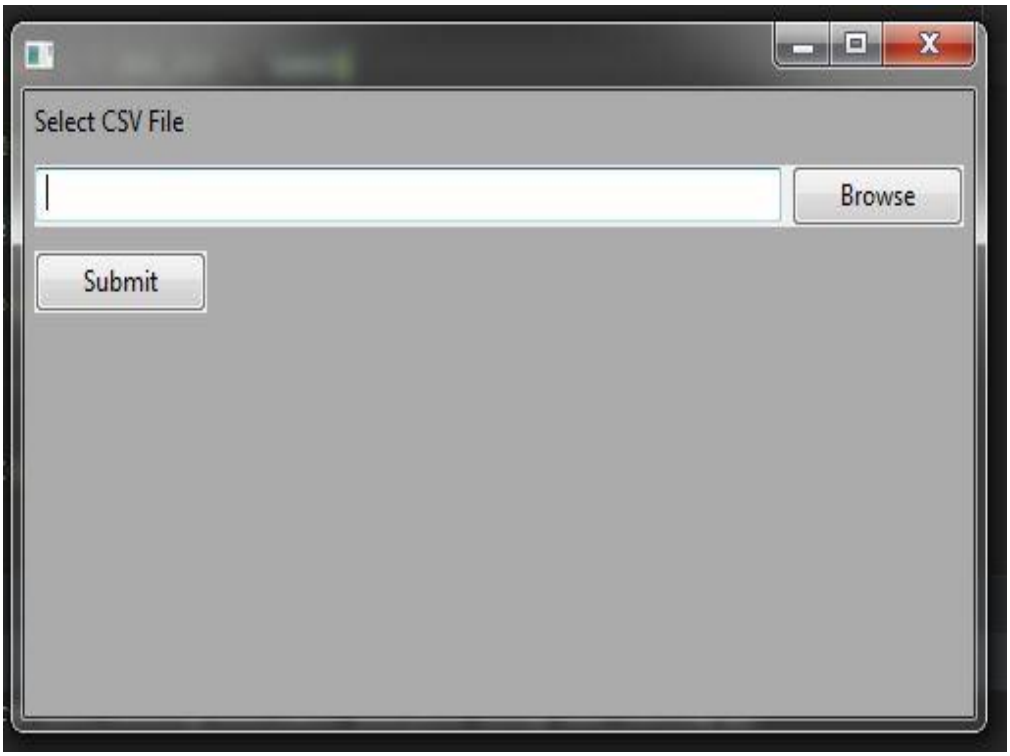

Fig 8.1 Selecting Dataset GUI

In dataset we have around 15,000 tweets classified as negative and positive used for training our model. When we train our model it shows following accuracy along with graph shown in figure 8.3 which shows how many tweets are positive or negative.

```
Epoch 1/7
2019-04-23 21:35:36.812857: I tensorflow/core/platform
 - 22s - loss: 0.4869 - acc: 0.7755
Epoch 2/7
 - 21s - loss: 0.3428 - acc: 0.8551
Epoch 3/7
 - 21s - loss: 0.2661 - acc: 0.8931
Epoch 4/7
 - 20s - loss: 0.2174 - acc: 0.9138
Epoch 5/7
 - 20s - loss: 0.1746 - acc: 0.9326
Epoch 6/7
 - 20s - loss: 0.1484 - acc: 0.9435
Epoch 7/7
 - 20s - loss: 0.1240 - acc: 0.9524
pos_acc 79.54545454545455 %
neg_acc 84.2964824120603 %
```

Fig 8.2 Training Model

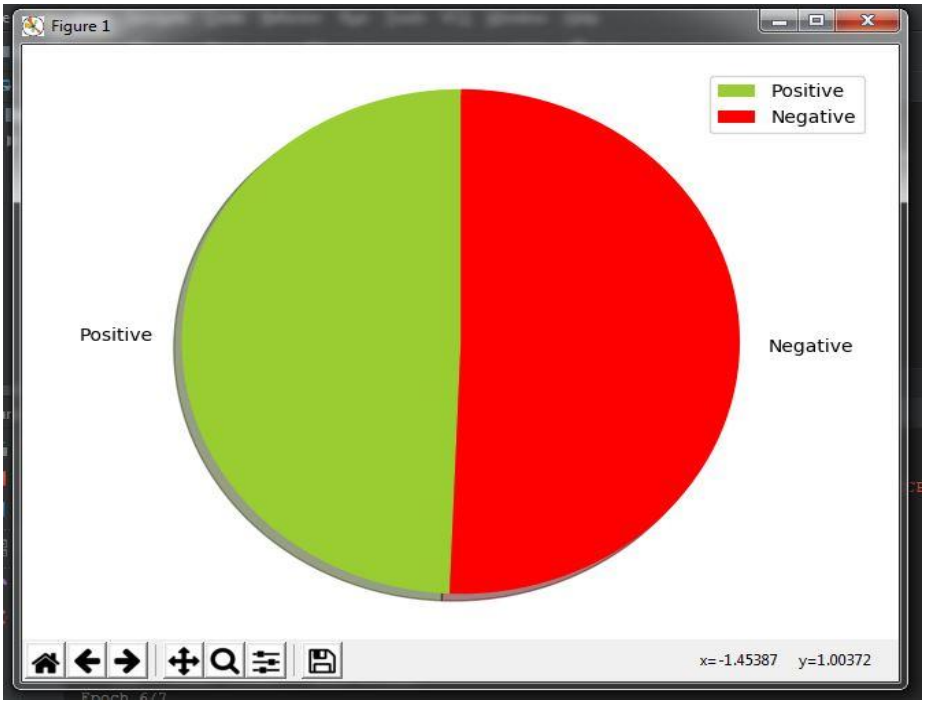

Fig 8.3 Graph based on Dataset

8.2 Extracting Real Time Tweets

In Figure 8.4 is the snapshot of the result generated by our program

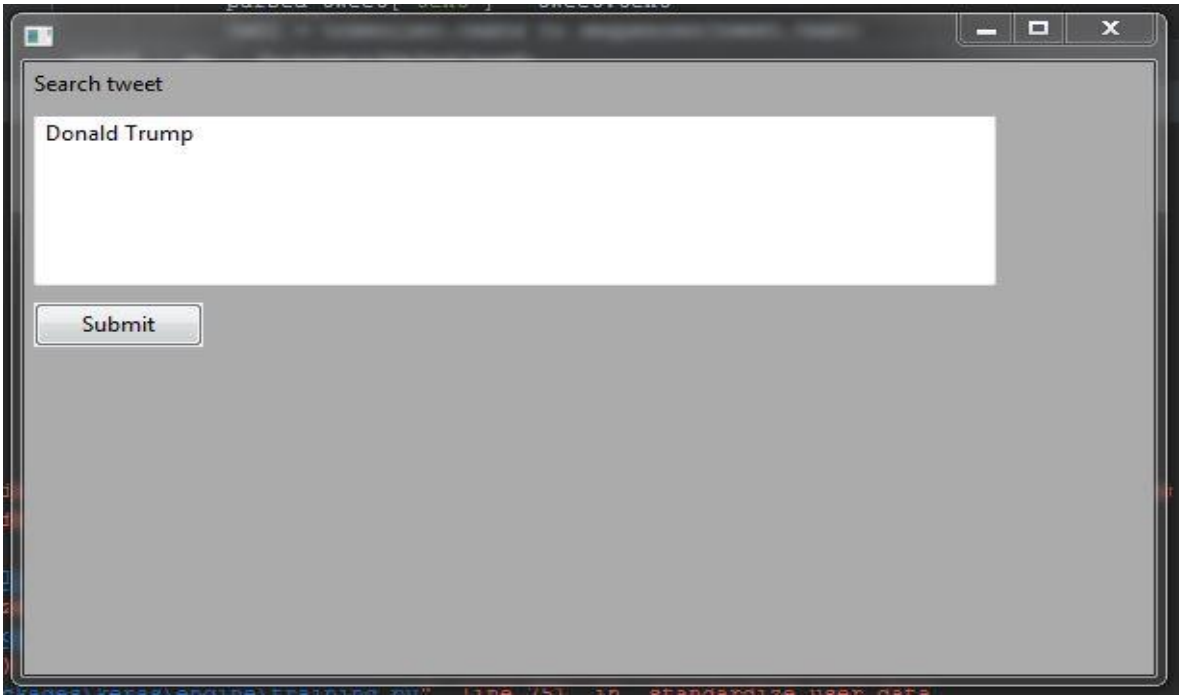

Figure 8.4 Search Window Result

When user enters search topic, based on that topic live tweets are extracted and system then predicts its polarity as shown in fig 8.5 and live tweets is saved in file along with its polarity as shown in fig 8.6

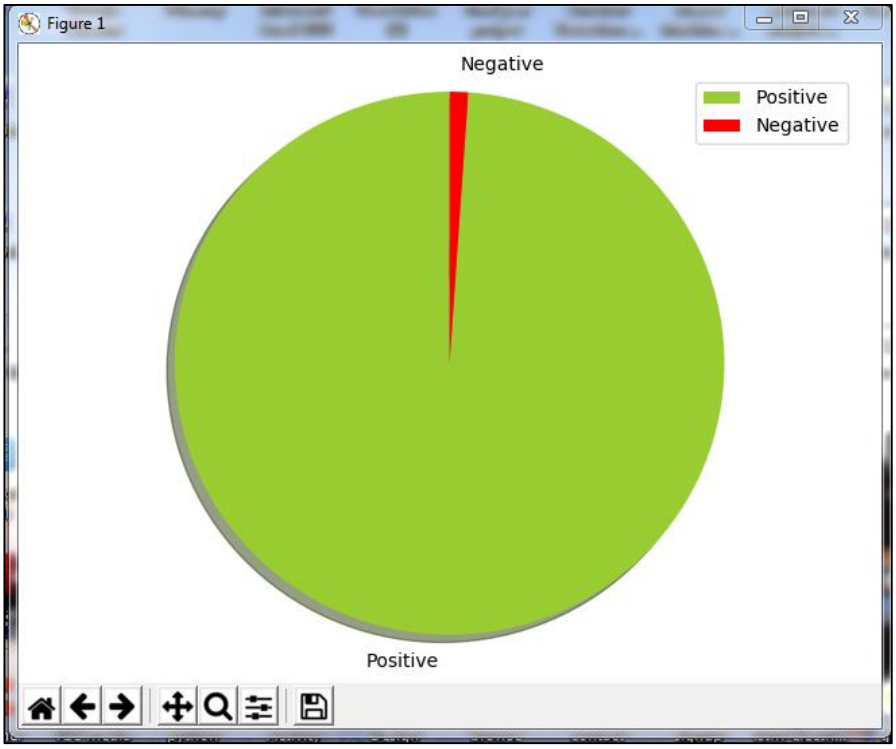

Figure 8.5 Graph of live tweets

65	RT @grantstern: Impeachment is a practical remedy for Congı	POSITIVE
66	RT @mrjamesob: Queen sends Prince Philip to pick up Donal	POSITIVE
67	RT @realDonaldTrump: â€œThe best thing ever to happen to	POSITIVE
68	Mueller report: Donald Trump failed us as commander in chie	NEGATIVE
69	RT @MrJonCryer: I wanna remind everybody that Donald Trui	POSITIVE
70	RT @realDonaldTrump: â€œThe best thing ever to happen to	POSITIVE
71	RT @tribelaw: A central focus of the impeachment hearings t	POSITIVE
72	RT @Back_dafucup: The Trump administration faces another	POSITIVE
73	@Stop_Trump20 Itâ€™s what Donald trump looks like with nc	POSITIVE
74	I cant think of any Americans who hate Theresa May like the I	POSITIVE
75	RT @acnewsitics: BREAKING: PresidentÂ Donald TrumpÂ and	POSITIVE
76	RT @TranslateRealDT: Donald Trump has now tweeted 53 tim	POSITIVE

Figure 8.6 Live tweets along with its polarity

8.3 Result Predicted by System

In Figure 8.7 is the snapshot of the result generated by our program

In this the user can enter a tweet to see what is its polarity predicted by system.

Figure 8.7 Result

9. Conclusion

9.1 Conclusion

A Prototype Emotion Detection from tweets is developed using RNN-LSTM. The models will be trained and validated against a test dataset. In this prototype, an improved RNN language model is put forward using LSTM. It is applied to achieve multi-classification for text emotional attributes, and identifies text emotional attributes more accurately. Even Today Sentiment Analysis is still a difficult and Complex problem in computer Science. Sentiments are express by Humans in Different Ways. The objective of this work is to create a system for finding the emotion from tweets and emoticons for users in a user friendly way.

9.2 Future Scope

This system has lot of future scope as the need for this system is felt very greatly. In future work, we could try these improvement programs or use different models combination to improve the performance of text sentiment analysis. The models will be trained and validated against a test dataset. We apply machine learning techniques to solve twitter sentiment analysis problem.

Also, Twitter sentiment analysis comes under the category of text and opinion mining. This research topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application issues with the slang used and the short forms of words. Many analyzers don’t perform well when the number of classes is increased. Also it’s still not tested that how accurate the model will be for other topics. Hence sentiment analysis has a very bright scope of development in future. Existing sentiment analysis models can be improved further with more semantic and insightful knowledge.

10. References

[1] Antonio Lopardo, Marco Brambilla "Analyzing and Predicting the US Midterm Elections on Twitter with Recurrent Neural Networks" 2018 IEEE International Conference on Big Data (Big Data)

[2] Rohith.V, D. Malathi "Sentiment Analysis on Twitter: A Survey" International Journal of Pure and Applied Mathematics Volume 118 No. 22 2018, 365-375

[3] Fenna Miedema "Sentiment Analysis with Long Short-Term Memory networks" Research Paper Business Analytics Vrije Universiteit Amsterdam August 1, 2018

[4] Abdalraouf Hassan "Sentiment Analysis with Recurrent Neural Network and Unsupervised Neural Language Model" March 2017

[5] Mitali Desai, Mayuri A. Mehta, "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey" IEEE System, International Conference on Computing, Communication and Automation (ICCCA2016)

[6] Abu Zonayed Riyadh, Nasif Alvi, Kamrul Hasan Talukder, "Exploring Human Emotion Via Twitter" IEEE System, 2017 20th International Conference on Computer and Information Technology (ICCIT), 22-24 December, 2017

[7] Jia Li, Hua Xu, Xingwei He, Junhui Deng and Xiaomin Sun, "Tweet Modeling with LSTM Recurrent Neural Networks for Hashtag Recommendation" IEEE System, 2016 International Joint Conference on Neural Networks (IJCNN)

[8] Gilbert, Eric and Karrie Karahalios. 2010. "Widespread Worry and the Stock Market." Proceedings of the 4th International AAAI Conference on Weblogs and Social Media 58–65.

[9] Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter Mood Predicts the Stock Market." Journal of Computational Science 2(1):1–8.

[10] Hansen, Lars Kai, Adam Arvidsson, Finn Aarup Nielsen, Elanor Colleoni, and Michael Etter. 2011. "Good Friends, Bad News - Affect and Virality in Twitter." Communications in Computer and Information Science 185 CCIS (PART 2):34–43.

[11] Stieglitz, Stefan and Linh Dang-Xuan. 2012. "Political Communication and Influence through Microblogging - An Empirical Analysis of Sentiment in Twitter Messages and Retweet Behavior." Proceedings of the Annual Hawaii International Conference on System Sciences 3500–3509.

[12] Ceron, Andrea, Luigi Curini, and M. Stefano. 2012. "Tweet Your Vote: How Content Analysis of Social Networks Can Improve Our Knowledge of Citizens' Policy Preferences. An Application to Italy and France." New Media & Society 16:1–24.

Appendix A

Publication: Rajat Jain, Sawan Kshatriya, Arun Dubey, Pratyush Shukla, "Emotion Detection from Tweets and Emoticons Using Rnn-Lstm", Journal of Emerging Technologies and Innovative Research (An International Open Access Journal & UGC and ISSN Approved) ISSN: 2349-5162

Journal of Emerging Technologies and Innovative Research

An International Open Access Journal

www.jetir.org | editor@jetir.org

Certificate of Publication

The Board of

Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162)

Is hereby awarding this certificate to

Rajat Jain

In recognition of the publication of the paper entitled

Emotion Detection from Tweets and Emoticons Using Rnn-Lstm

Published In JETIR (www.JETIR.org) ISSN UGC Approved & 5.87 Impact Factor

Published in Volume 6 Issue 4 , April-2019

EDITOR

EDITOR IN CHIEF

JETIR1904F44 Research Paper Weblink http://www.jetir.org/view?paper=JETIR1904F44 Registration ID : 206479

Journal of Emerging Technologies and Innovative Research

An International Open Access Journal

www.jetir.org | editor@jetir.org

Certificate of Publication

The Board of

Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162)

Is hereby awarding this certificate to

Sawan Kshatriya

In recognition of the publication of the paper entitled

Emotion Detection from Tweets and Emoticons Using Rnn-Lstm

Published In JETIR (www.JETIR.org) ISSN UGC Approved & 5.87 Impact Factor

Published in Volume 6 Issue 4 , April-2019

EDITOR

EDITOR IN CHIEF

JETIR1904F44 Research Paper Weblink http://www.jetir.org/view?paper=JETIR1904F44 Registration ID : 206479

Journal of Emerging Technologies and Innovative Research

An International Open Access Journal

www.jetir.org | editor@jetir.org

Certificate of Publication

The Board of

Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162)

Is hereby awarding this certificate to

Arun Dubey

In recognition of the publication of the paper entitled

Emotion Detection from Tweets and Emoticons Using Rnn-Lstm

Published In JETIR (www.JETIR.org) ISSN UGC Approved & 5.87 Impact Factor

Published in Volume 6 Issue 4 , April-2019

EDITOR

EDITOR IN CHIEF

JETIR1904F44 Research Paper Weblink http://www.jetir.org/view?paper=JETIR1904F44 Registration ID : 206479

Journal of Emerging Technologies and Innovative Research

An International Open Access Journal

www.jetir.org | editor@jetir.org

Certificate of Publication

The Board of

Journal of Emerging Technologies and Innovative Research (ISSN : 2349-5162)

Is hereby awarding this certificate to

Pratyush Shukla

In recognition of the publication of the paper entitled

Emotion Detection from Tweets and Emoticons Using Rnn-Lstm

Published In JETIR (www.JETIR.org) ISSN UGC Approved & 5.87 Impact Factor

Published in Volume 6 Issue 4 , April-2019

EDITOR

EDITOR IN CHIEF

JETIR1904F44 Research Paper Weblink http://www.jetir.org/view?paper=JETIR1904F44 Registration ID : 206479

Published Paper:

© 2019 JETIR April 2019, Volume 6, Issue 4 www.jetir.org (ISSN-2349-5162)

Emotion Detection from Tweets and Emoticons Using Rnn-Lstm

[1]Rajat Jain, [2]Sawan Kshatriya, [3] Arun Dubey, [4] Pratyush Shukla, Prof. K Jayamalini

[1]Author, [2]Author, [3]Author, [4]Author
Department of Computer Engineering
Mumbai University
Shree LR Tiwari College of Engineering, Thane, India

Abstract: Many micro blogging sites have millions of people sharing their thoughts daily. We propose and investigate the sentiment from a popular real-time micro blogging service, Twitter, where real time reactions are posted by the user and we find their opinions for almost about "everything". Social networking sites like twitter, Facebook, Instagram, Orkut etc. are the great source of communication for internet users. So this becomes an important source for understanding the opinions, views or emotions of people. We extract data, i.e. tweets from Twitter in real time and apply machine learning techniques to convert them into a useful form and then use it for building sentiment classifier. Given a piece of written text, the problem is to categorize the text into one specific sentiment polarity i.e. positive, negative. With the increase use of Internet and big explosion of text data, it has been a very significant research subject to extract valuable information from Text Ocean. To realize multi-classification for text sentiment and emoticons sentiments, this paper promotes a RNN language model based on Long Short Term Memory (LSTM). LSTM is far better than the traditional RNN. And as a language model, LSTM is applied to achieve multi-classification for text and emoticon emotional attributes.

Index Terms **- RNN, LSTM, Twitter, Sentiment analysis, Opinion mining.**

I. INTRODUCTION

People are posting more and more text information on Internet, and it has been a great challenge to distinguish whether the information is useful or not. As a result, it is necessary to create models to dig out valuable information, which can be used for product reviews, movie reviews, politics, sentiment analysis. However, the traditional method is very limited, as it is unable to deal with a large amount of data timely. Consequently, people are giving more importance to the efficiency of the sentimental model, and RNN is a good model. Feed-forward networks are able to take into account only a fixed context length to predict the next word, recurrent neural networks (RNN) can take advantage of all previous words. Traditional RNN language model is going further in model generalization instead of considering only the several previous words (parameter n) the recursive weights are assumed to represent short term memory. More in general we could say that RNN sees text as signal having words. Long Short-Term Memory (LSTM) neural network is different type of RNN structure. It allows to discover both long and short patterns in data and helps to eliminate the problem of vanishing gradient by training RNN.

II. RELATED WORK

[1] In Antonio Lopardo, Marco Brambilla paper is based on how they adopted a RNN-LSTM binary classifier to reach a validation accuracy of 85% over individual tweets, despite the highly implicit and short content shared on the social network. The method was able to predict the correct winner on 60% of the highly competitive (and thus extremely hard to predict also with traditional methods) districts.

[2] In Rohith.V, D. Malathipaper is based on the comparison made by them on different artificial neural network algorithms and how they fair well.

[3] Fenna Miedema paper is based on rnn-lstm algorithm which shows that the efficiency of the algorithm increases when used with lstm.

[4] Abdalraouf Hassan, paper describes a simple and efficient Neural Language Model approach for text classification that relies only on unsupervised word representation inputs. Their model employs Recurrent Neural Network Long Short-Term Memory (RNN-LSTM), on top of pre-trained word vectors for sentence-level classification tasks.

III. PROPOSED SYSTEM

The objective of this research is to create an system which will predict the emotion of user whether is it positive or negative using RNN-LSTM. LSTM are a type of RNN. Prediction is sequentially in RNN, and the hidden layer from one prediction is the hidden layer of the next prediction which will assign a memory to the network. Results from past predictions can improve future predictions. LSTM gives RNN an extra aspect that gives it fine-grained control over memory. This aspect control how much the current input matters in creating the new memory, and how much the past memories matters in creating the new memory, and what are important in generating the output. Twitter is a great source for opinions of various kinds of events and products. Detecting the sentiment of these micro blogs is a challenging task which has attracted increased research interest in recent years.

The Figure1 shows block diagram of the system. The users post their tweets in twitter. These tweets are then extracted in real time using twitter API in the form of raw data which are then saved in database. The raw dataset is converted into target dataset through Data Pre-Processing and Feature Extraction. The features of the words are selected and then machine learning techniques are applied on extracted features to classify them into its sentiment polarity that is namely positive or negative.

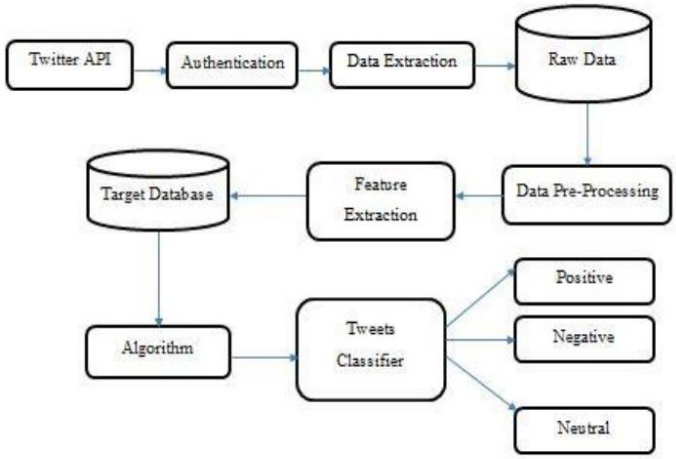

FIG 1: Proposed Block Diagram

3.1 Data Preprocessing

Mining of Twitter data is a challenging task. The collected data is raw data. In order to apply classifier, it is essential to pre-process or clean the raw data. The pre-processing task involves removal of hashtags, uniform casing and other Twitter notations (@, RT), emoticons, URLs, stop words, decompression of slang words and compression of elongated word . [5]The following steps show the pre-processing procedure.

- To Remove Twitter notations such as hashtags (#), retweets (RT), and account Id (@).
- Remove the URLs, hyperlinks.
- Remove the stop words such as is, am, are etc. The stop words do not affect any emotions; it is just to compress the dataset.
- Compress the elongated words such as happyyy into happy.
- Decompress the slag words such as g8, f9. Mostly slang words are adjectives or nouns and they contain the highest degree level of sentiments. So it is necessary to decompress them.
- Emoji Translation: In the social media, people always use emoji to express their moods. Therefore, emoji could be useful information for sentiment analysis. We use regular expression to find out emoji pattern and translate them into positive and negative words.

3.2 Feature Extraction

The pre-processed dataset has various discrete properties. In feature extraction methods, we extract different aspects such as adjectives, verbs and nouns and later these aspects are identified as positive or negative to detect the polarity of the whole sentence. [5]Followings are the widely used Feature Extraction methods.

- Negative Phrases: Negative words changes the meaning or orientation of the opinion. So it is evident to take negative word into account.
- Parts Of Speech (POS): Finding nouns, verbs, adjectives etc. as they are significant gauges of opinions.

3.3 Dataset for Training Neural Network

Dataset contains positive and negative tweets used for training our system for prediction of user emotion.

3.4 Work-Flow

1. USER REGISTRATION: The user registers itself with the system by entering the name and password.
2. USER LOGIN: The user then logins itself in the system by entering registered name and password.
3. SELECT DATASET: In this the user selects a file in which we have our dataset of positive and negative tweets for training our model.

10	Negative	Deer in the headlights RT @lizzwinstead: Ben Carson, may be the only brain surgeon who has performed a lobotomy on himself. #GOPDebate
11	Negative	RT @NancyOsborne180: Last night's debate proved it! #GOPDebate #BATsAsk @BadassTeachersA #TBATs https://t.co/G2gGjY1bJD
12	Negative	@JGreenDC @realDonaldTrump In all fairness #BillClinton owns that phrase.#GOPDebate
13	Positive	RT
14	Negative	Me reading my family's comments about how great the #GOPDebate was http://t.co/gIaGjPygXZ
15	Neutral	RT @ArcticFox2016: RT @AllenWestRepub "Dear @JebBush #GOPDebate #NotAMistake http://t.co/TtFG7KYcd9"
16	Positive	RT @pattonoswalt: I loved Scott Walker as Mark Harmon's romantic rival in SUMMER SCHOOL. Look it up. #GOPDebate
17	Negative	Hey @ChrisChristie exploiting the tragedy of 9/11 for your own political gain is @rudygiulianiGOP's thing #GOPDebate
18	Negative	RT @CarolCNN: #DonaldTrump under fire for comments about women @PeterBeinart @SL_Schaeffer @IWF @@MyRkiger weigh in on #GOPdebate h
19	Negative	RT @johncardillo: Guess who had most speaking time at the #GOPDebate. @FoxNews moderators with 31.7% of time. http://t.co/2WSUT0c0Lx
20	Negative	reason comment is funny 'in case you're ignorant' is the #gop #tcot are the reason the government isn't working for the people #gopdebate
21	Negative	RT @PamelaGeller: Huckabee: Paying for transgender surgery for soldiers, sailors and airmen does not make our country safer #Ha #GOPDebate
22	Positive	RT @ChuckNellis: Cruz has class & truth, that gets my vote! #GOPDebate
23	Negative	RT @mchamric: RT â€œ@TeaTraitors: #GOPDebate was still Clown Show! I'm glad Head Clown Trump helping destroy GOP. http://t.co/nwGx8G8JWrá
24	Negative	RT @erinmallorylong: No *I* hate Planned Parenthood and women more! NO I HATE PLANNED PARENTHOOD AND WOMEN MORE!!!!! #GOPDebate

FIG 2: Screenshot of Dataset

4. ENTER TWEET: User need to enter a tweet so that our system can predict whether it is positive or negative.
5. ENTER SEARCH TOPIC: User need to enter a topic on which live tweets will be downloaded and our system will predict whether it is positive or negative and a pie chart will also be plotted.

IV. ALGORITHM USED

Recurrent neural network (RNN) work as a powerful set of artificial neural network algorithm. A version of recurrent neural network works was used by DeepMind in their work playing video games with autonomous agents. Recurrent neural network work differ from feed forward neural network work because they include a feedback loop, whereby output from step x-1 is fed back to the neural network to affect the outcome of step x, and so forth for each subsequent step. For example, if a neural network is exposed to a word letter by letter, and it is asked to guess each following letter, the first letter of a word will help determine what a recurrent neural network thinks the second letter will be, etc.

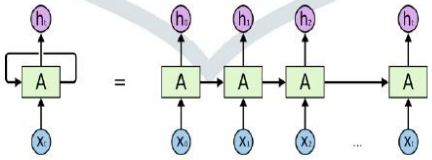

FIG 3: RNN

Long Short Term Memory neural network works – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. LSTM models are a variety of RNN. In RNN the prediction in sequence, where the hidden layer from one prediction is the hidden layer of the next prediction this will assign a memory to the neural network work, therefore, results 'from earlier estimation could lead to improve future predictions. LSTM gives RNN more features to a extreme control over memory; this aspects control how much the present input matters for forming the new memory, also how much the past memories matters in creating the new memory, and what parts of the memory are essential is producing the output.

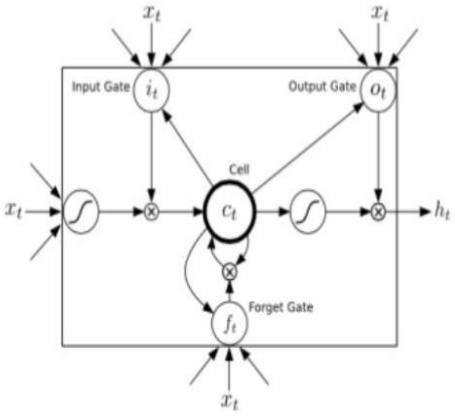

FIG 4: LSTM

Figure 4: shows the setup for LSTM using only last hidden layer for sentiment prediction and use softmax/cross-entropy loss associated with sentiment analysis.

V. RESULTS

This section deals with the results of the system implemented and are discussed below:

- The given figure below shows the Enter text window in which user need to enter tweet to find its emotion i.e. whether it is positive or negative

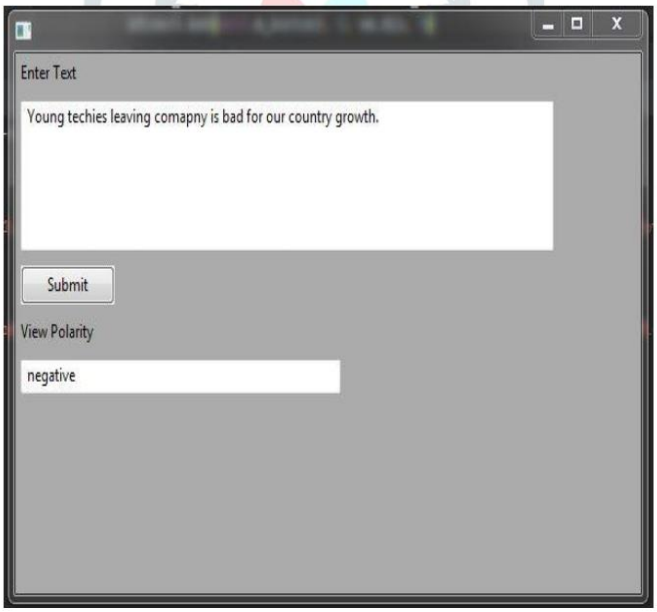

Fig 5: Enter Text Gui

- The given figure below shows the Search text window in which user can enter a topic on which live tweets will be downloaded and our system will predict its polarity.

Fig 6: Search Text Gui

- Also when the result is predicted , results are saved in file with its sentiment and a graph is shown as in figure 7

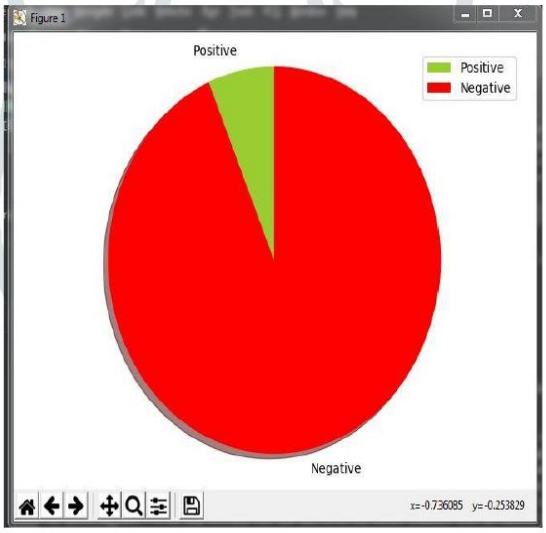

Fig 7: Search Text Gui

VI. CONCLUSION & FUTURE SCOPE

In this paper, an improved RNN language model is put forward using LSTM, which successfully covers all history sequence information and performs better than conventional RNN. It is applied to achieve multi-classification for text emotional attributes, and identifies text emotional attributes more accurately. Even Today Sentiment Analysis is still a difficult and Complex problem in computer Science. Sentiments are express by Humans in Different Ways. The objective of this work is to create a system for finding the emotion from tweets and emoticons for users in a user friendly way.

This system has lot of future scope as the need for this system is felt very greatly. The few numbers of future implementations are:

In future work, we could try these improvement programs or use different models combination to improve the performance of text sentiment analysis. The models will be trained and validated against a test dataset. We apply machine learning techniques to solve twitter sentiment analysis problem.

Also, Twitter sentiment analysis comes under the category of text and opinion mining. This research topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application issues with the slang used and the short forms of words. Many analyzers don't perform well when the number of classes are increased. Also it's still not tested that how accurate the model will be for other topics. Hence sentiment analysis has a very bright scope of development in future.

Existing sentiment analysis models can be improved further with more semantic and insightful knowledge.

VII. REFRENCES

[1] Antonio Lopardo, Marco Brambilla "Analyzing and Predicting the US Midterm Elections on Twitter with Recurrent Neural Networks" 2018 IEEE International Conference on Big Data (Big Data)

[2] Rohith.V, D. Malathi "Sentiment Analysis On Twitter: A Survey" International Journal of Pure and Applied Mathematics Volume 118 No. 22 2018, 365-375

[3] Fenna Miedema "Sentiment Analysis with Long Short-Term Memory networks" Research Paper Business Analytics Vrije Universiteit Amsterdam August 1, 2018

[4] Abdalraouf Hassan "Sentiment Analysis With Recurrent Neural Network And Unsupervised Neural Language Model " March 2017

[5] Mitali Desai, Mayuri A. Mehta, "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey" IEEE System, International Conference on Computing, Communication and Automation (ICCCA2016)