

STUDY OF SOCIAL IMPACT AND INFLUENCE OF TWEETS IN CONNECTION WITH KERALA FLOOD

A Project Report Submitted By
ATHIRA P

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE (M.SC. COMPUTER SCIENCE)
(University of Calicut)
AT



DEPARTMENT OF COMPUTER SCIENCE
FAROOK COLLEGE (AUTONOMOUS)
FAROOK COLLEGE P.O., CALICUT, KERALA, INDIA
15-MARCH- 2019

ACKNOWLEDGEMENT

First of all, I am grateful to THE ALMIGHTY GOD for the blessings showered upon me over the years; it goes without saying that without them, this project would not complete. I wish to express my sincere thanks to The principal of the college Dr: K.M. Naseer, for providing me with all the necessary facilities for the successful completion of the same.

I would like to express my deep sense of gratitude to Dr.V Kabeer HOD, Department of Computer Science for his valuable guidance and consistent encouragement during the course of project work and for its successful completion. I wish to express my sincere thanks to Dr.P.P Abdul Haleem (Project Coordinator), Mr.Mohammed Shameer M.C, Mrs.Nusrath, Mrs.Rasiya Anwar, Mr.Sameer V.V for their valuable guidance, suggestions, encouragement and inspiration extended to me throughout my project. I am also thankful to Dr. V Kabeer project guide for his creative ideas and suggestions throughout the project. Last, but not the least, I am grateful to my parents and friends for the way they cooperated. Expected me to achieve success and have always stirred my ambition to do the best.

ATHIRA P

DECLARATION

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person or material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Dated: 15-March- 2019

Signature of the Author: _____
ATHIRA P

DEPARTMENT OF COMPUTER SCIENCE
FAROOK COLLEGE (AUTONOMOUS)

Date: **15-March- 2019**

This is to certify that the project report entitled **STUDY OF SOCIAL IMPACT AND INFLUENCE OF TWEETS IN CONNECTION WITH KERALA FLOOD** submitted by ATHIRA P to the University of Calicut for the award of the Master of Science (M.Sc.) in Computer Science is a bonaide record of the project work carried out by him/her under my supervision and guidance. The content of the report, in full or parts have not been submitted to any other Institute or University for the award of any other degree or diploma.

Research Supervisor: _____
Dr. V. Kabeer

Head of the Department: _____
Dr. V. Kabeer

This is to certify that the candidate was examined by us in the Project Viva Voce Examination held on and his/her Register Number is FKARMCS001.

Examiners:

Examiner 1: _____

Examiner 2: _____

Table of Contents

Table of Contents	v
List of Tables	vii
List of Figures	viii
Abstract	ix
1 Introduction	1
1.1 Project Category	2
1.2 Project Profile	2
1.2.1 Tools/Platform	2
2 Problem Analysis and Scope of the System	3
2.1 Problem Definition	3
2.2 Need Identification	4
2.3 Problem Domain	4
2.4 Objectives	4
2.5 Methodology	4
2.6 Scope of the Solution	5
2.7 Motivation	5
3 Analysis and Software Specification	6
3.1 Literature Review	6
3.2 Existing System	7
3.3 Proposed System	8
3.4 Requirement Specification	8
3.4.1 Functional Requirement	8
3.4.2 Non-Functional Requirement	9

3.4.3	Technical Requirements	9
3.4.4	Feasibility Study	11
4	Design	13
4.1	Users of the System	13
4.2	Modules of the System	14
4.2.1	Data Collection	14
4.2.2	Preprocessing	14
4.2.3	Classification	15
4.2.4	Result Graph	16
4.3	Data Flow Diagram	16
4.4	User Interface Layout	19
4.5	Database and Design	21
4.5.1	List of Entities and Attributes	22
4.5.2	ER(Entity Relationship) Diagram	23
4.5.3	Structure of Tables	24
5	Implementation	25
5.1	Tools/Script for Implimentation	25
5.2	Coding	27
5.2.1	Downloading Tweets	27
5.2.2	Preprocessing	28
5.2.3	Retrieved tweets are displaying using .NET	29
5.2.4	Dispalying preprocessed and classified tweets	30
5.2.5	Classification	34
5.3	Important Screenshots	36
6	Testing	39
6.1	Introduction	39
6.2	Testing Methodigies Adopted	39
6.2.1	Unit Testing	39
6.2.2	Integrated Testing	40
6.2.3	System Testing	41
7	Conclusion	42
	Bibliography	43
	Publications Out of Project Work	45

List of Tables

4.1	Twitter Table	24
4.2	Preprocess Table.	24
4.3	Result Table.	24

List of Figures

4.1	Level 0	17
4.2	Level 1	18
4.3	Download Tweets	19
4.4	Load Dataset	20
4.5	Preprocessing	20
4.6	Classification	21
4.7	List of Entities and Attributes	22
4.8	ER Diagram	23
5.1	Tweets downloading	36
5.2	Tweets Loading	37
5.3	Tweets Preprocessing	37
5.4	Tweets Classification	38
5.5	Result	38

Abstract

Micro blogging sites are rich in sources for a varied kind of information. This is a common place where people exchange their on various issue; it could be on ongoing trends. Based on their experiences they share a comment or complaint on any issues and express their thoughts in terms of positive and negative sentiment. In todays world, Govt and Non-Govt organization require feedback from active user, on a particular issue or on an event. So here is a challenge to analyze or detect and accomplish the global sentiment.

Twitter is a large platform where different opinion are presented and exchanged. Being a rich source of user generated data with their emotions and opinion, Twitter data can be used to analyze the social issues and its impact. In this proposed system Twitter data is used to analyze the social impact and influence of tweets in connection with Kerala flood. For data collection, Twitter provide a Twitter API which provide an authorized user to download tweets for data analysis. The data collected from twitter are preprocessed using stemming algorithm, to improve the classification accuracy. The preprocessed tweets are then classified using Naive Bayes algorithm, into positive, negative and neutral classes. By analyzing this classes, the overall opinion of society on Kerala flood can be identify.

Chapter 1

Introduction

Twitter is popular for its massive spreading of instant messages and the nature of freedom a user has. Twitter is a large platform where different opinion are presented and exchanged. No matter where the users come from, educated or not, or from different religion, they comment, discuss, argue over topics they are interested in and share their own feelings freely. Within 140 characters, users can post their feelings as short messages. This short messages contain emotional state of individual as well as emotions of a group of people that the user represent.

To analyze the public mood on a particular topic, normally people use conventional system like questionnaire or survey which is time consuming and not cost effective. In this study Twitter is used as the source for data. Download the Twitter data in connection with Kerala flood using Twitter APIs, and preprocessed using stemming method. Then preprocessed tweets are classified into positive, negative and neutral classes. By analyzing this classes further identify whether the opinion of the people are favoring or not in the matter of Kerala flood. This work is an extension of sentimental analysis using twitter as a tool.

1.1 Project Category

The project titled STUDY OF SOCIAL IMPACT AND INFLUENCE OF TWEETS IN CONNECTION WITH KERALA FLOOD comes under the category of Research and Development.

1.2 Project Profile

1.2.1 Tools/Platform

Hardware Requirements

- Intel core i3 and above.
- 4 GB RAM or above.
- 320 GB Hard Disk or above.
- 2 GB Graphics

Software Requirements

- Python 3.7.2
- Microsoft Visual Studio 2010
- Microsoft Excel 2013

Chapter 2

Problem Analysis and Scope of the System

2.1 Problem Definition

Twitter is a large platform where different opinion are presented and exchanged. Because of this rich content user generated data with their emotions and opinion [1] [3] this system use twitter data to analyze the social impact and influence of tweets in connection with Kerala flood. Twitter provide a Twitter API which provide an authenticated user account to extract tweets for data analyzes. This downloaded tweets need to be preprocessed before classification using stemming algorithm. For this study, Naive Bayes classification [2] technique is used to classify this preprocessed tweets into three category like positive, negative and neutral. After this classification, the probability of users who tweets positively, negatively and neutrally on this topic are to be plotted on a graph. According to this graph, the conclusion can be made on the overall opinion of users on Kerala flood. This conclusion can be used to analyze its social impact and influence son society.

2.2 Need Identification

There are different data mining research going on Twitter analysis. As per the Literature review, Several researchers attempted to analyze and classify emotional data on twitter to find what kind of emotions are likely to spread, comparison of classification accuracy and detecting the community on twitter. But there is few work related to analyze the social impact and influence Kerala flood related tweets. So this system analyze the tweets on Kerala flood and study the social imapct of these tweets using Naive Bayes classification.

2.3 Problem Domain

Tweets on Kerala flood are downloaded from Twitter using Twitter API and preprocessed. The preprocessed tweets are classified into positive and negative classes using Naive Bayes classifier. The result of the classification can be used to study the overall opinion of our society on Kerala flood.

2.4 Objectives

- To classify emotional tweets on twitter
- To analyze the influence of emotional tweets on Kerala flood and study its social impact

2.5 Methodology

Current methodologies for developing such data mining application are mostly based on Machine Learning technique offered by different programming languages

such as Python, R etc. In this proposed system the tweets regarding Kerala flood are downloaded using hash tag from Twitter. The downloaded tweets are preprocessed and then classified using Naive Bayes classifier into positive, negative and neutral classes. The result of the classification depicted on a graph.

2.6 Scope of the Solution

This project analyze the tweets on the topic Kerala flood and study its social impact and influence on society. Twitter is a large scale data platform on which number of studies cab be made on any intrested topic, but in this study, limiting its scope into Kerala flood related tweets. There are many attributes related to each tweets. In this work download Kerala flood tweets and their corresponding author id.

2.7 Motivation

Motivation behind the development of this system is that emerging use of twitter analysis in different fields. All kind of celebrities and common people are using twitter for various purpose and they mark their opinion and protest against some things in twitter. So in today's environment twitter analysis has importance. This leads to the development of Twitter analysis in connection with kerala flood and find how people react during calamity.

Chapter 3

Analysis and Software Specification

System analysis is a problem-solving technique that decomposes a system into its component pieces for the purpose of studying how well those component parts work and interact to accomplish their purpose.

3.1 Literature Review

There are many different types of research studies related to emotion extraction from natural language [1][2][3]. In many studies emotions are classified into two classes positive and negative or multidimensional. According to Maryam Hasan [4] Twitter messages as input data set, as they provide a very large, diverse and freely available ensemble of emotions. Using hash-tags as labels, our methodology trains supervised classifiers to detect multiple classes of emotion on potentially huge datasets with no manual effort. So to analyze the social opinion about a current issue like Kerala flood twitter is a better data rich platform.

2 Naive Bayes is one of the popular classification algorithm used to classify emotions. Ruan [4] proposed a method to find pessimist and optimists. Their result showed that Naive Bayes had highest accuracy among other classification algorithm. Because of

this I used to prefer Naive Bayes algorithm to classify emotional tweets with more accuracy.

In “Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach” [5], instead of presenting the sentiment polarity of each tweet relevant to the topic, the proposed system focus the study on hash tag-level sentiment classification. This task aims to automatically generate the overall sentiment polarity for a given hashtag in a certain time period, which markedly differs from the conventional sentence-level and document-level sentiment analysis.

Existing works related to user relationship in twitter mainly concentrate on community detection [6][7][8] , comparison of classification accuracy of different algorithm [4] or kinds of emotions that are likely to spread on different topic.

There are no studies analyze social impact and influence of tweets in connection with Kerala flood. This work analyzes users who post tweets on Kerala floods and classify users into three group positive, negative and neutral and then analyze its social impact.

3.2 Existing System

Large number of twitter analysis tools is available in market. Most of them are general purpose and don’t analyze the collected tweets from twitter regarding a particular hashtag. There are systems which list the most trending hashtag for the ease of searching process of users in their interest. Very few of the tools display the tweets for direct analysis on a particular hashtag.

3.3 Proposed System

Social impact and Influence analysis on Kerala flood propose a system that can analyze tweets from twitter on kerala flood and study its social impact. Proposed System perform classification of tweets into positive, negative and neutral classes using Naive Bayes classifier which improve the reliability of classification result. It also represent the analysis result in graphical format for the user readability.

3.4 Requirement Specification

3.4.1 Functional Requirement

A functional requirement denotes a function of a system or its component. A function is described as a set of inputs, the behavior, and outputs. Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that denote what a system is supposed to accomplish.

The functional requirements are:

- It perform various kinds of functions such as download tweets, preprocessing tweets, and classification of tweets.
- It provide complete facility to add data to data store.
- Fetch tweets according to the user input.

- Analyzed result shown in graphical format.

3.4.2 Non-Functional Requirement

The non-functional requirements of the system are:

- **Performance:** My system will be fast and scalable.
- **User interfaces** will be attractive.
- **Usability:** Easy to use
- **Reliability:** User generated data from Twitter

3.4.3 Technical Requirements

Already said that the proposed system is technically feasible. Technologies are used in the proposed system are python and .Net.

Hardware Requirements

- **Processor :** Intel core i3 and above.
- **Hard Disk :** 320 GB

- **RAM : 4 GB**
- **Graphics : 2 GB**

Software Requirements

- **OS : Windows 8 or above**

Windows 8 Start screen, showing default live tile arrangement. Windows 8 is a personal computer operating system that was produced by Microsoft as part of the Windows NT family of operating systems.

- **Language : Python 3.7.2**

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. It provide number for libraries for data mining process.

- **Front End : Microsoft Visual Studio 2010**

Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs, as well as websites, web apps, web services and mobile apps. Visual Studio uses Microsoft software development platforms such as Windows API, Windows Forms, Windows Presentation Foundation, Windows Store and Microsoft Silverlight. It can produce both native code and managed code.

- **Data Base : Microsoft Excel 2013**

Microsoft Excel is a spreadsheet program included in the Microsoft Office suite of applications. Spreadsheets present tables of values arranged in rows and columns. In this system tweets are stored in excel format.

3.4.4 Feasibility Study

The feasibility study, is an evaluation and analysis of the potential of a proposed project which is based on extensive investigation and research to support the process of decision making. This evaluation is very important because, the decision regarding the further procedure in the project is taken based on the same. It studies how feasible development of the software is going to be in terms of following aspects.

Technical feasibility

The technical feasibility assessment is focused on gaining and understanding of the present technical resources of the organization and the applicability to the expected needs of the proposed system. The proposed system is developed using .Net and Python which are available at free of cost. The system doesn't need any additional hardware requirements. The system collect data from twitter which provide data at free of cost.

Economical feasibility

This assessment typically involve cost and benefit analysis. In my system Twitter is main data source which is freely provide Twitter API to download tweets .The system use .NET and Python which is also freely available on internet. So the system doesn't need additional hardware and software requirements. So the system is economically

feasible.

Operational feasibility

Operational feasibility is measure how well a proposed system solves the problem within the given scope. In my system large dataset is retrieved from twitter and its analysis can be done using Naive Bayes classification using python libraries and the interface is designed in such a way that the user can be familiar with the system easily without any training. So the system is operationally feasible.

Chapter 4

Design

Software design is the process by which an agent creates a specification of a software artifact, intended to accomplish goals, using a set of primitive components and subject to constraints. Software design is the process of conceptualizing the software requirements into software implementation. This is the initial phase within the software development life cycle (SDLC)—shifting the concentration from the problem to the solution.

This section describes about the users of the system, modules, and a detailed diagrammatic description about what is being done.

4.1 Users of the System

The users of the system are psephologist, bussiness people, and admin. They are the main users of the system.

4.2 Modules of the System

4.2.1 Data Collection

In this module data collection is main process. In my system Twitter is the main data source where rich set of data is present on the topic ‘Kerala flood’. Data access from twitter have four steps. The first step is, input keyword ‘Kerala flood’ on which the tweets are collected from Twitter. In the next step the Twitter will provide an API key to access the twitter data. Using this keys user can make a JSON request for retrieving tweets. By receiving this request, Twitter collect the data on the given topic and provide to the user. The data which is in Java Script Object Notation format which stored in Excel file.

4.2.2 Preprocessing

In this module data preprocessing is the main process. Data which is collected from twitter contain noise and unnecessary data. So the preprocessing of data is needed to convert the data into a machine-readable format. Preprocessing means we need remove punctuations, accent marks and white spaces, expanding abbreviations etc. After preprocessing the preprocessed tweets are stored into a preprocessed data store.

In this system python libraries for stemming is used for preprocessing. Stemming and Lemmatization are Text Normalization (or sometimes called Word Normalization) techniques in the field of Natural Language Processing that are used to prepare text, words, and documents for further processing.

Natural Language Tool Kit (NLTK) is a Python library to make programs that work with natural language. It provides a user-friendly interface to datasets that are over

50 corpora and lexical resources such as WordNet Word repository. The library can perform different operations such as tokenizing, stemming, classification, parsing, tagging, and semantic reasoning.

For example, 're' is Python's Regular Expressions (RegEx) library, which takes care of parsing strings and modifying them in an efficient way without having to explicitly iterate through the characters comprising the particular string. ' nltk' is the Natural Processing Toolkit, which is one of the most commonly used Python libraries out there. It takes care of any processing that we need to perform on text to change its form or extract certain components from it. The class constructor removes stop words.

4.2.3 Classification

In this module the preprocessed data is retrieved and classified into two category using Naïve Bayes classification algorithm. The tweets are classified into positive, negative and neutral classes. Naive Bayes Classifier is a classification algorithm that relies on Bayes' Theorem. This theorem provides a way of calculating a type or probability called posterior probability, in which the probability of an event A occurring is reliant on probabilistic known background. By analyzing the probability of each class we study the overall opinion of society on Kerala flood.

Naive Bayes Classifier

Naive Bayes is a probabilistic technique for constructing classifiers. The characteristic assumption of the naive Bayes classifier is to consider that the value of a particular feature is independent of the value of any other feature, given the class variable. Naive Bayes classifiers have good results in complex real-world situations.

An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters necessary for classification and that the classifier can be trained incrementally.

4.2.4 Result Graph

Tweets are downloaded from twitter using Twitter APIs which have the key-value pair format. In this system, tweets and its corresponding user-id are downloaded and stored into an Excel file. After preprocessing the tweets are stored into preprocessed data store. The preprocessed data is classified into three classes positive, negative and neutral. In this study 500 tweets are downloaded using Twitter API and after preprocessing there are 62 tweets are present. The preprocessed tweets are classified into three classes as positive, negative and neutral. In the positive class there are 21 tweets, 25 tweets are in negative class and 16 tweets in neutral class. From this study we can conclude that overall opinion of society is negative.

4.3 Data Flow Diagram

Data flow diagrams are used to graphically represent the flow of data in a Software system. DFD describes the processes that are involved in a system to transfer data from the input to the file storage and reports generation. DFD graphically representing the functions, or processes, which capture, manipulate, store, and distribute data between a system and its environment and between components of a system. The visual representation makes it a good communication tool between User and System designer.

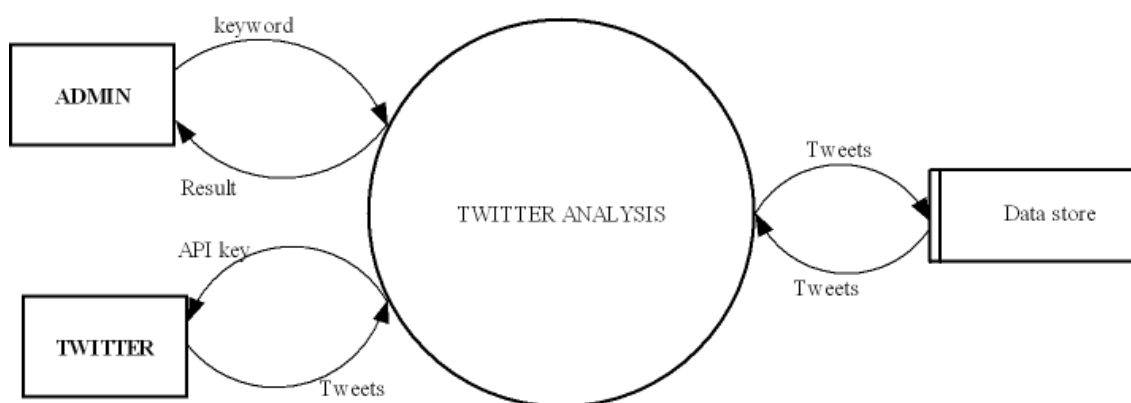


Figure 4.1: Level 0

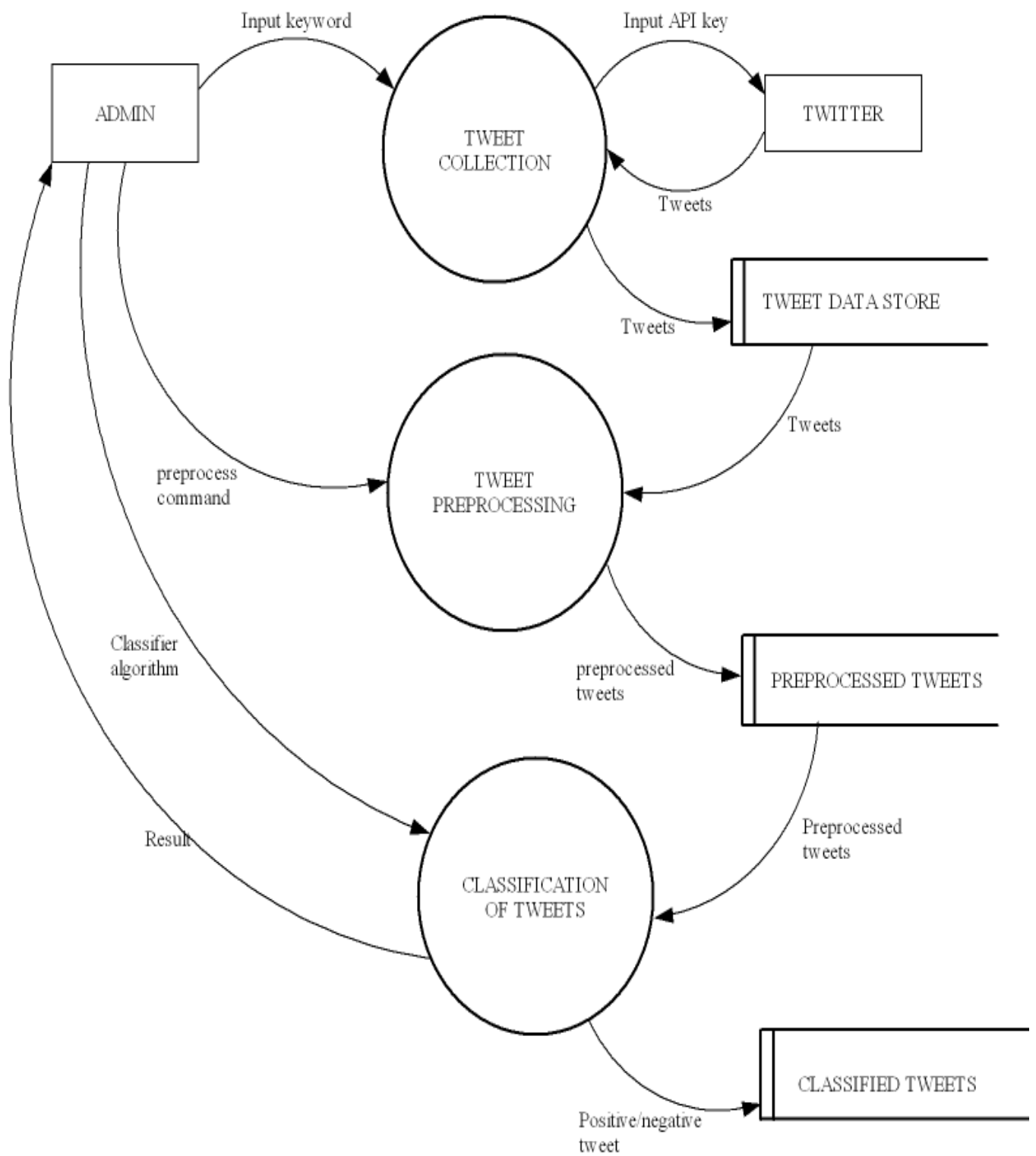


Figure 4.2: Level 1

4.4 User Interface Layout

These are the user interface design of the system "STUDY OF SOCIAL IMPACT AND INFLUENCE OF TWEETS IN CONNECTION WITH KERALA FLOOD", It is mainly focused on the user's experience and interactions. The goal of user interface deesign is to make the user's interaction more friendly, efficient and simple as possible. Here main activity layouts of system are shown below,

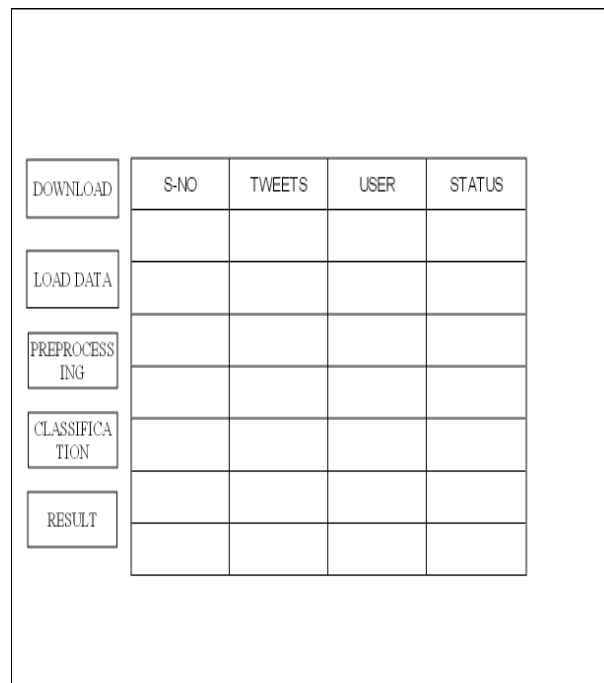


Figure 4.3: Download Tweets

hash tag

DOWNLOAD	S-NO	TWEETS	USER	STATUS

LOAD DATA

PREPROCESS
ING

CLASSIFICA
TION

RESULT

Figure 4.4: Load Dataset

DOWNLOAD	S-NO	TWEETS	PREPROCESSED

LOAD DATA

PREPROCESS
ING

CLASSIFICA
TION

RESULT

Figure 4.5: Preprocessing

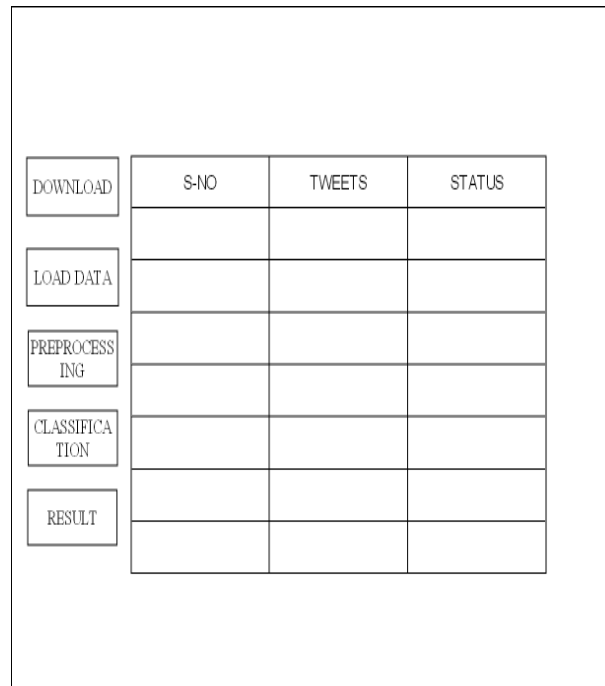


Figure 4.6: Classification

4.5 Database and Design

Database design run in parallel with the application design. As information is collected about what is to be done and what data is need to be entered, stored, messaged and printed on reports. One of the main tasks in building a new system is determining the contents and structure of the database. The type of retrieval and recording required by the user and availability of inputs determine what data has to be stored. The database is a collection of related data.

4.5.1 List of Entities and Attributes

Tables	Attributes
Twitter table	sno tweet userid
Preprocess Table	sno tweet preprocessedtweet
Result Tabel	sno tweet status

Figure 4.7: List of Entities and Attributes

4.5.2 ER(Entity Relationship) Diagram

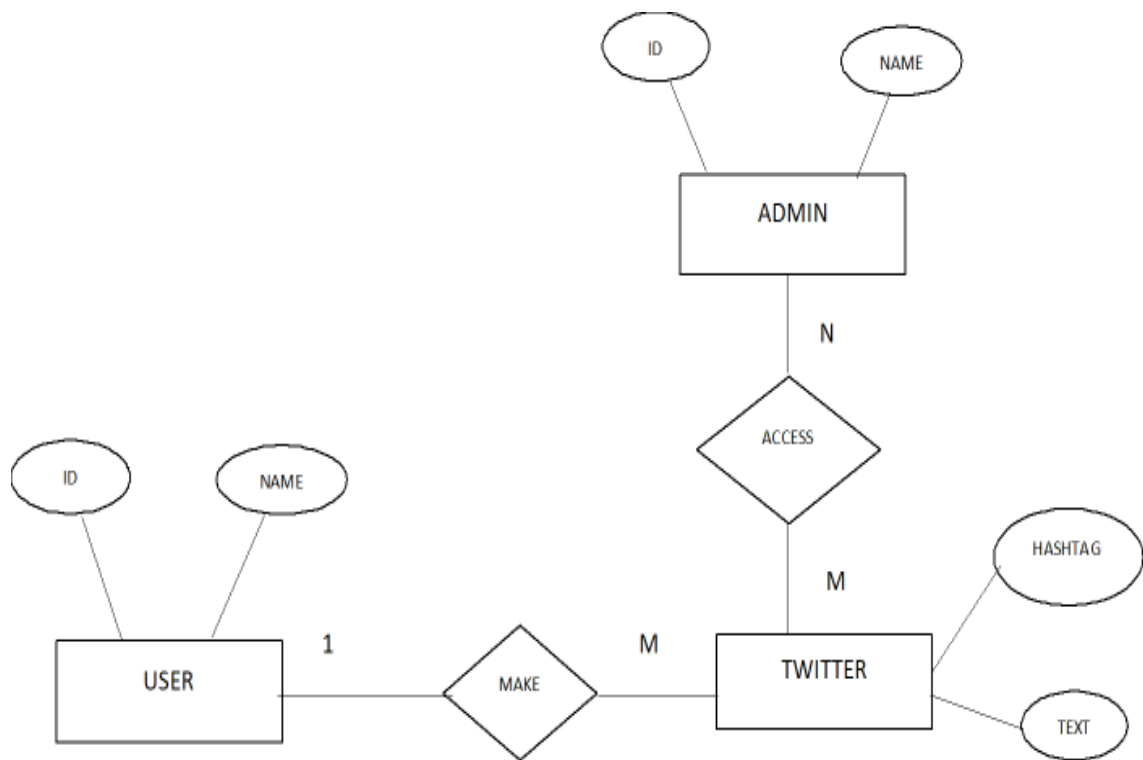


Figure 4.8: ER Diagram

4.5.3 Structure of Tables

Name	Type	Null	Default	Extra
sno	int(50)	no	no	Auto-Increment
tweet	varchar(200)	no	no	
userid	varchar(500)	no	no	

Table 4.1: Twitter Table

Name	Type	Null	Default	Extra
sno	int(50)	no	no	Auto-Increment
tweet	varchar(200)	no	no	
preprocessedtweet	varchar(200)	no	no	

Table 4.2: Preprocess Table.

Name	Type	Null	Default	Extra
sno	int(50)	no	no	Auto-Increment
tweet	varchar(200)	no	no	
status	varchar(200)	no	no	

Table 4.3: Result Table.

Chapter 5

Implementation

Implementation is the process that actually yields the lowest-level system elements in the system hierarchy. The purpose of the implementation process is to design and create (or fabricate) a system element conforming to that element's design properties and/or requirements. The element is constructed employing appropriate technologies and industry practices

5.1 Tools/Script for Implimentation

- **Python 3.7.2**

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are

available in source or binary form without charge for all major platforms, and can be freely distributed.

- **Microsoft Visual Studio 2010**

Visual Studio is a Integrated Development Environment(IDE) developed by Microsoft to develop GUI(Graphical User Interface), console, Web applications, web apps, mobile apps, cloud, and web services etc. With the help of this IDE, you can create managed code as well as native code. It uses the various platforms of Microsoft software development software like Windows store, Microsoft Silverlight, and Windows API etc. It is not a language specific IDE as you can use this to write code in C Sharp, C++, VB(Visual Basic), Python, JavaScript, and many more languages. It provides support for 36 different programming languages. It is available for Windows as well as for macOS.

- **Microsoft Excel 2013**

Excel is a commercial spreadsheet application produced and distributed by Microsoft for Microsoft Windows and Mac OS. It features the ability to perform basic calculations, use graphing tools, create pivot tables and create macros. Excel has the same basic features as all spreadsheet applications, which use a collection of cells arranged into rows and columns to organize and manipulate data. They can also display data as charts, histograms and line graphs.

5.2 Coding

Only the core sections of source code are discussed here.

5.2.1 Downloading Tweets

```
# coding: utf-8

# In [10]:

from twython import Twython

APP_KEY = 'V5Xudq1NhUsmhQqLiQTGFfGqV'
APP_SECRET = '5Tkp8phsGPGUgpMxOj4N5gaoZe5sIGMO3Iw2GPWPSf7vsQ3P5d'
OAUTH_TOKEN = '1070929351313485824-yOhTOUHVWOKh9d9hF3Q59w1XuN3rtD',
OAUTH_TOKEN_SECRET = '
    aIMAbmDhavPyFtFqo8ol1akO4zrw0Xuv1EiXVrzHAJkDP'
twitter = Twython(APP_KEY, APP_SECRET,
                  OAUTH_TOKEN, OAUTH_TOKEN_SECRET)

f="E:\TWITTER\HASHTAG.txt"
f1=open(f,"r")
h=f1.read()
h

res=[]
res=twitter.search(q=h, lang='en')

tweets = res['statuses']
tweets
```

```
texts = [tweet['text'] for tweet in tweets]
id1=[tweet['id_str'] for tweet in tweets]
```

```
import pandas as pd
```

```
df = pd.DataFrame.from_dict({ 'TWEET':texts , 'USER':id1 })
```

```
#df.to_excel('E:\\TWITTER\\c1.xls', header=True, index=False)
df.to_excel('E:\\TWITTER\\testdata.xls', header=True, index=False
)
```

```
f="E:\\TWITTER\\demo.txt"
f2=open(f,"w")
f2.write("successful")
f2.close()
```

5.2.2 Preprocessing

```
f1="E:\\TWITTER\\row.txt"
rd=open(f1,"r")
s=rd.read()
rd.close()
s
```

```
import nltk
from nltk.tokenize import sent_tokenize, word_tokenize
words=word_tokenize(s)
#print(words)
t=[word for word in words if word.isalpha()]
print(t)
from nltk.corpus import stopwords
sw=set(stopwords.words('english'))
t1=[w for w in t if not w in sw]
print(t1)
```

```
from nltk.stem import PorterStemmer
```

```
ps=PorterStemmer()
f="E:\\TWITTER\\stem1.txt"
f1=open(f,"w")
for word in t1:
    print(ps.stem(word))
    f1.write(ps.stem(word)+" ")
f1.close()
```

5.2.3 Retrieved tweets are displaying using .NET

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Drawing;
using System.Data;
using System.Linq;
using System.Text;
using System.Windows.Forms;
using System.IO;
using System.Diagnostics;
using System.Data.OleDb;
using System.Threading;

namespace TWITTERANALYSIS
{
    public partial class TWT4 : UserControl
    {
        Class1 c = new Class1();
        public TWT4()
        {
            InitializeComponent();
        }

        private void label1_Click(object sender, EventArgs e)
        {
        }
    }
}
```

```

private void button2_Click(object sender, EventArgs e)
{
    StreamWriter s = new StreamWriter(@"E:\TWITTER\
        HASHTAG.txt");
    s.Write(textBox1.Text);
    s.Close();

    Process p = new Process();
    ProcessStartInfo ps = new ProcessStartInfo();
    ps.Arguments = @"C:\Users\Athira\Desktop\python\
        hasread2.py";
    ps.FileName = @"C:\Users\Athira\Anaconda3\python.exe
        ";
    Process.Start(ps);

    Thread.Sleep(5000);

    OleDbCommand cmd = new OleDbCommand();
    cmd.CommandText = "select * from [Sheet1$]";
    dataGridView1.DataSource = c.get(cmd);

}

private void TWT4_Load(object sender, EventArgs e)
{

}

}

```

5.2.4 Displaying preprocessed and classified tweets

```

private void button3_Click(object sender, EventArgs e)

```

```

{
    t2 = new TWT2();
    panell.Controls.Clear();
    panell.Controls.Add(t2);

    // preprocess.py
    if (File.Exists(@"E:\TWITTER\preprocessed.xls"))
    {
        File.Delete(@"E:\TWITTER\preprocessed.xls");
    }

    exekmean(@"E:\TWITTER\preprocess.py");
    while (true)
        if (File.Exists(@"E:\TWITTER\preprocessed.xls"))
            break;

    OleDbCommand cmd = new OleDbCommand();
    cmd.CommandText = "select * from [Sheet1$]";
    DataSet ds = new DataSet();

    DataTable dt = c.getpreprocessed(cmd);
    t2.dataGridView1.DataSource = dt;

}

private void button4_Click(object sender, EventArgs e)
{
    panell.Controls.Clear();
    t3 = new TWT3();
    panell.Controls.Add(t3);

    if (File.Exists(@"E:\TWITTER\review.xls"))
    {
        File.Delete(@"E:\TWITTER\review.xls");
    }

    exekmean(@"E:\TWITTER\classify.py");
    while (true)
        if (File.Exists(@"E:\TWITTER\review.xls"))

```



```

        break;

        OleDbCommand cmd = new OleDbCommand();
        cmd.CommandText = "select * from [Sheet1$]";
        DataSet ds = new DataSet();

        DataTable dt = c.getout(cmd);
        t3.dataGridView1.DataSource = dt;

    }

    private void exekmean(string nam)
    {
        Process p = new Process();
        ProcessStartInfo start = new ProcessStartInfo();
        start.Arguments = nam; //ok
        start.UseShellExecute = false;
        start.FileName = @"C:\Users\Athira\Anaconda3\python.exe";
        start.RedirectStandardOutput = true; // Any output,
        // generated by application will be redirected back
        start.RedirectStandardError = true;
        using (Process process = Process.Start(start))
        {
            using (StreamReader reader = process.
                StandardOutput)
            {
                string result = reader.ReadToEnd();
            }

            process.WaitForExit();
        }
    }

    private void button5_Click(object sender, EventArgs e)
    {
        panell1.Controls.Clear();
        panell1.Controls.Add(new TWT5());
    }

```

```

    }

    private void button6_Click(object sender, EventArgs e)
    {

        //showoutput.py

        panel1.Controls.Clear();
        TWI5 tw5 = new TWI5();
        this.panel1.Controls.Add(tw5);

        if (File.Exists(@"E:\\TWITTER\\output.png"))
        {
            File.Delete(@"E:\\TWITTER\\output.png");
        }

        exekmean(@"E:\\TWITTER\\showoutput.py");
        while (true)
            if (File.Exists(@"E:\\TWITTER\\output.png"))
                break;

        tw5.pictureBox1.ImageLocation = "E:\\TWITTER\\output.
            png";
    }

    private void button7_Click(object sender, EventArgs e)
    {
        this.Close();
    }

    private void label1_Click(object sender, EventArgs e)
    {

    }

}

}

```

5.2.5 Classification

```
# coding: utf-8
```

```
# In[46]:
```

```
import numpy as np
import pandas as pd
```

```
df = pd.read_excel('E:\\TWITTER\\preprocessed.xls')
df.head()
```

```
# In[50]:
```

```
import nltk
import nltk.sentiment
import nltk.sentiment.util
import sys
import io
from io import StringIO
```

```
# In[5]:
```

```
cnt=0
for index, row in df.iterrows():
    stdout_ = sys.stdout
    stream = StringIO()
    sys.stdout = stream
    nltk.sentiment.util.demo_liu_hu_lexicon(row['tweets'], plot=
        False)
    sys.stdout = stdout_
    sentiment = stream.getvalue()
    sentiment = sentiment[:-1]
    df.at[cnt, 'res'] = sentiment
    #re=nltk.sentiment.util.demo_liu_hu_lexicon(row['tweets'],
        plot=False)
    cnt+=1
```

```
# In[51]:
```

```
df
```

```
# In[52]:
```

```
df.to_excel("E:\\TWITTER\\review.xls")
```

5.3 Important Screenshots

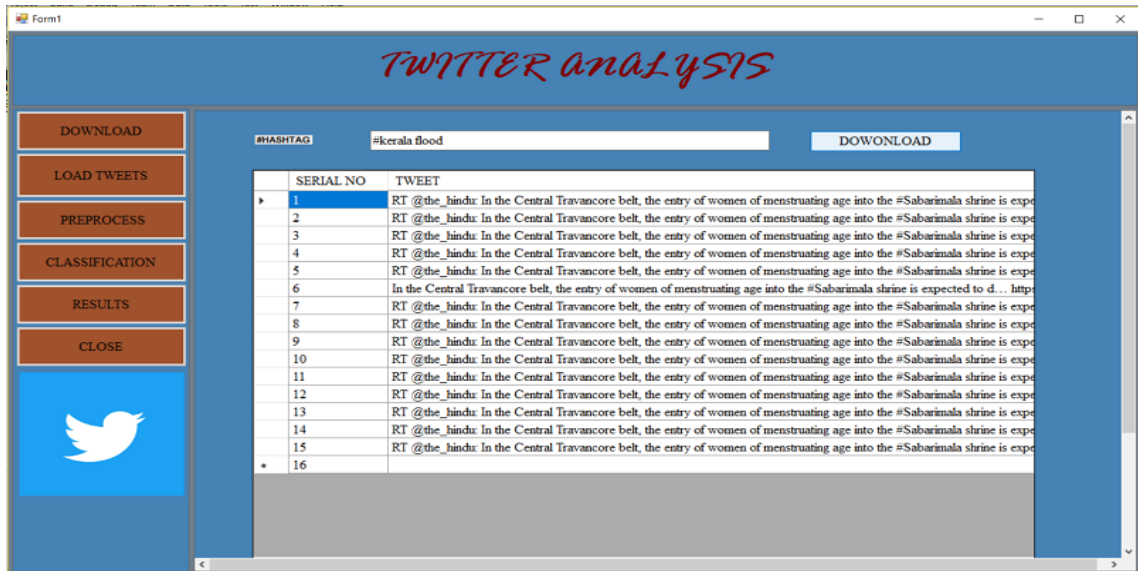


Figure 5.1: Tweets downloading

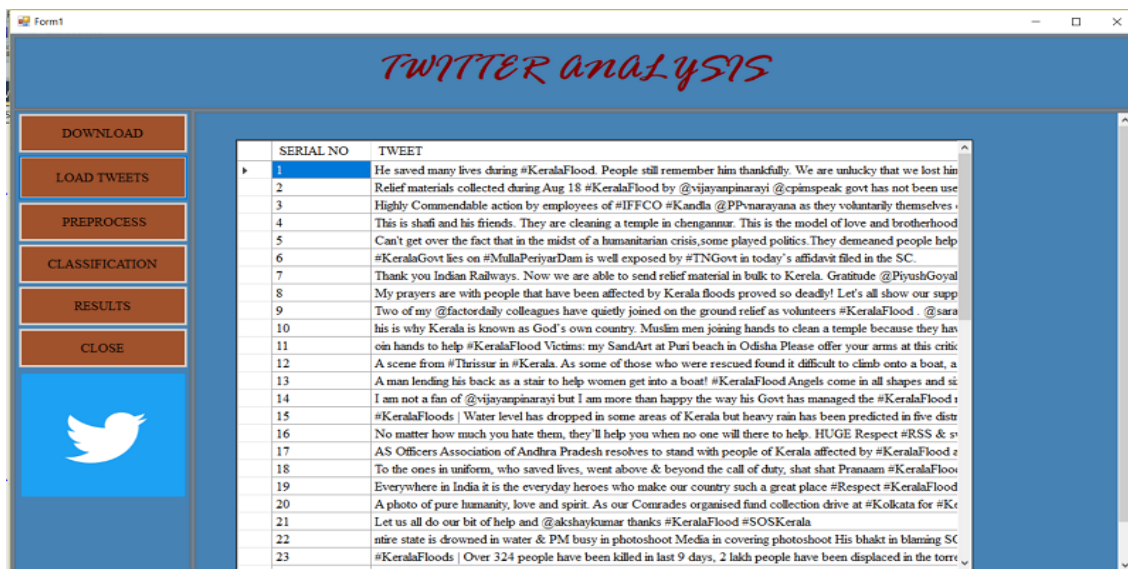


Figure 5.2: Tweets Loading

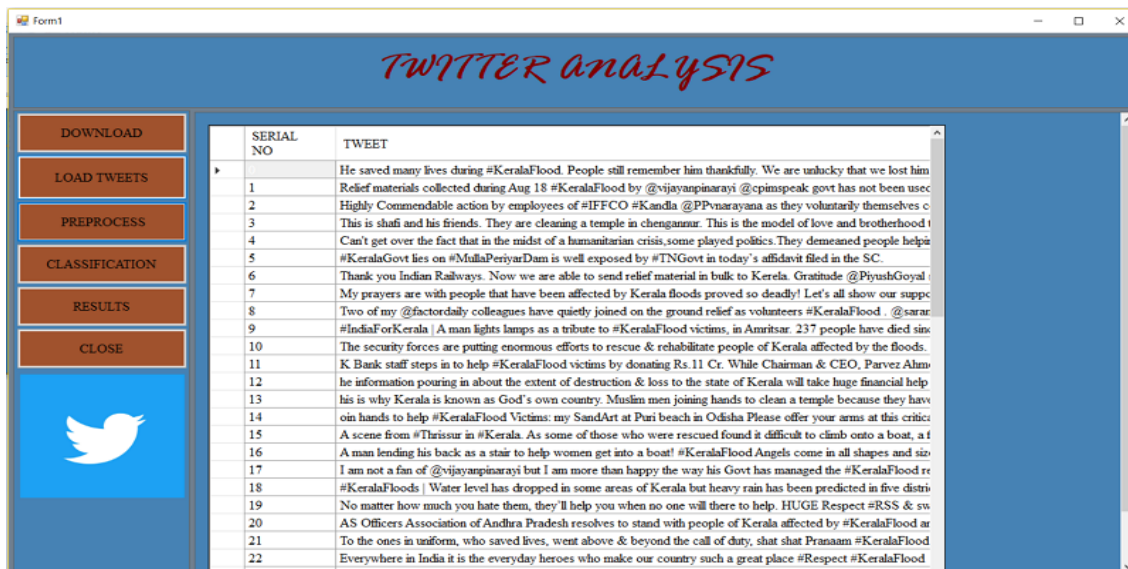


Figure 5.3: Tweets Preprocessing

Form1

TWITTER analysis

DOWNLOAD


LOAD TWEETS

PREPROCESS

CLASSIFICATION

RESULTS

CLOSE



	STATUS
	Negative
	Negative
	Positive
	Neutral
	Neutral
	Neutral
	Positive
	Positive
	Positive
	Negative
	Negative
	Negative
	Negative
	Neutral
	Negative
	Negative
	Neutral
	Positive
	Neutral
	Neutral
	Positive
	Positive
	Neutral
	Positive
	Positive
	Positive
	Positive

Figure 5.4: Tweets Classification

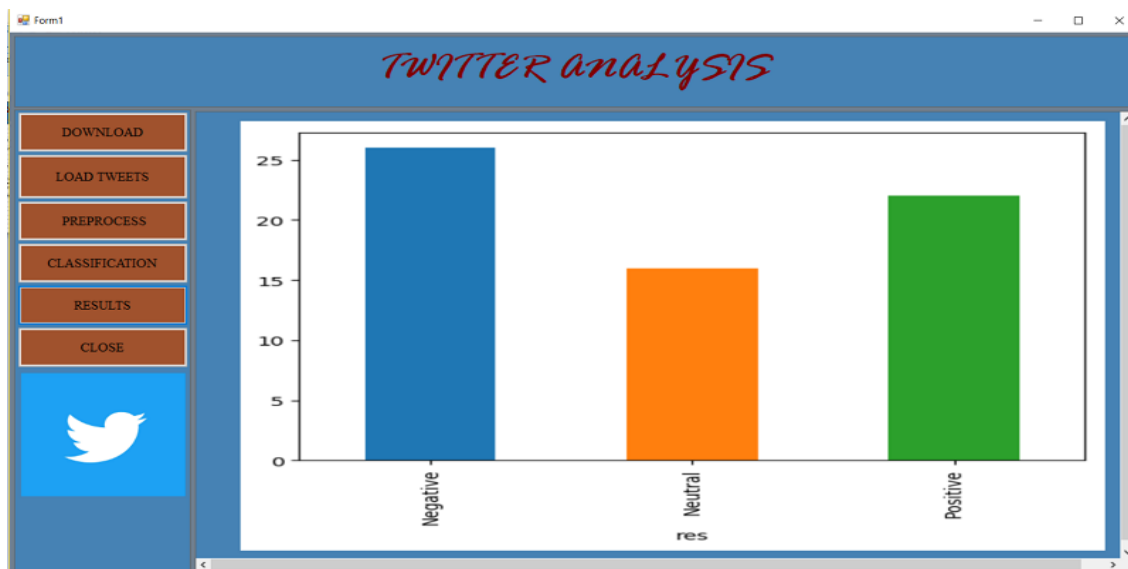


Figure 5.5: Result

Chapter 6

Testing

6.1 Introduction

Software testing is the critical element of software quality assurance and represents the ultimate review of specification, design, and coding. Testing is often used in association with the terms verification and validation. Verification is the checking or testing of an item for confirmation and consistency with an associated specification. Testing is just one Kind of verification, which also uses techniques such as reviews, analysis, inspections and walks through. Validation is the process of checking that what has been specified is what the user actually wanted.

6.2 Testing Methodigies Adopted

6.2.1 Unit Testing

Unit testing focuses on verifying the effort on the smallest unit of the software module. The the local data structure is examined to ensure that the data stored temporarily maintains its integrity during all steps in the algorithm execution. Boundary conditions are tested to ensure that the module operates properly at boundaries established to limit or restrict processing. In ANALYZING THE SOCIAL IMPACT OF

TWEETS IN KERALA FLOOD complete modules are tested one by one.

Test Case: 01

- Test Case ID: UT01
- Test Case Description: Without entering HASH TAG press download button
- Expected Result: it should prompt validation message

6.2.2 Integrated Testing

Integration testing is a systematic technique for constructing a program structure while conducting tests to uncover errors associated with integration. Data can be tested across an interface. One module can have an inadvertent, adverse effect on others. Integration testing is a systematic technique for constructing a program structure while conducting tests to uncover errors associated with integration. The major concerns of integration testing are developing an incremental strategy that will limit the complexity of entire actions among components as they are added to the system.

Test Case: 01

- Test Case ID: IT01
- Test Case Description: Check the details can be saved.
- Test Strategy: Different types of input data given to the program.
 1. Incomplete data
 2. Complete data

- Expected Result:
 1. If the input is incomplete, the program should give an appropriate message and set the focus in the corresponding component on the interface.
 2. If the data is complete then the program should execute correctly. The data should be saved and required output displayed.

6.2.3 System Testing

System Testing is a level of software testing where a complete and integrated software is tested. The purpose of this test is to evaluate the system's compliance with the specified requirements.

Test Case: 01

- Test Case ID: ST01
- Test Case Description: Checks that the system work properly.
- Test Strategy : Compare the requirement specification with working of the system.
 1. Partially meets the requirement specification.
 2. Meet all the requirements.
 1. If the system meets only partial requirements in there- requirement specification, it will not get the desired result as needed
 2. When all the requirements are satisfied, the system works as needed.

Chapter 7

Conclusion

I would like to conclude that this system is so powerful because it performs different tasks. It fetch tweets on Kerala flood from twitter and store in the datastore until it analysed and result produced. The proposed system is more flexible and changes can be easily made. The system is developed with an insight into the necessary modification that may be required in the future. Hence the system can be maintained successfully without much effort. The project report entitled "STUDY OF SOCIAL IMPACT AND INFLUENCE OF TWEETS IN CONNECTION WITH KERALA FLOOD " has come to its final stage. The system has been developed with much care that it is free of errors and at the time it is efficient and more less time consuming one.

In this study 500 tweets are downloaded using Twitter API and after preprocessing there are 62 tweets are present. The preprocessed tweets are classified into three classes as positive, negative and neutral. In the positive class there are 21 tweets, 25 tweets are in negative class and 16 tweets in neutral class. From this study we can conclude that overall opinion of society is negative. The system can be enhanced with different classifier algorithms to improve the classification accuracy.

Bibliography

- [1] K. Tago , Q. Jin , *Influence Analysis of Emotional Behaviors to User Relationships Based on Twitter Data*, In Proceedings of the Fifth International Conference on Advanced Cloud and Big Data, 2017
- [2] C. dos Santos, M. Gatti, *Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts*, In Proceedings of COLING 2014, pp. 69–78, 2014.
- [3] M. Hasan, E. Rundensteiner, *EMOTEX: Detecting Emotions in Twitter Messages*, In Proceedings of the Sixth ASE International Conference on Social Computing (SocialCom2014). Academy of Science and Engineering (ASE), 2014
- [4] X. Ruan, S. Wilson, R. Mihalcea, *Finding Optimists and Pessimists on Twitter*, In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp. 320–325, 2016.
- [5]] Xiaolong Wangy , Furu Weiz, Xiaohua Liuz, Ming Zhouz, Ming Zhangy, *Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach*, CIKM’11, October 24–28, 2011, Glasgow, Scotland, UK. Copyright 2011 ACM 978-1-4503-0717-8/11/10 .

- [6] M. Ozer, N. Kim, and H. Davulcu, *Community detection in political Twitter networks using Nonnegative Matrix Factorization methods*, In Proceedings of Advances in Social Networks Analysis and Mining(ASONAM). pp. 81–88, 2016.
- [7] C. Hongyun, et al., *Sociallens: Searching and browsing communities by content and interaction*, In Proceedings of the Data Engineering(ICDE), 2017.
- [8] M. B. Hutair, I. Kamel and Z. A. Agbari, *Social Community Detection Based on Node Distance and Interest*, Big Data Computing Applications and Technologies (BDCAT), 2016.

Publications Out of Project Work

1. Athira P and V KAbeer, "ANALYZING THE SOCIAL IMPACT OF TWEETS IN KERALA FLOOD", *First National Conference On Computational Intelligence and Data Analysis (CIDA) 2019*, Calicut University, February, 2019.