

A PROJECT REPORT
ON
“TWITTER TREND ANALYSIS”

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE
DEGREE OF
BACHELOR OF COMPUTER ENGINEERING

BY
SANKALP NILEKAR (B-17)
SUDEEP RAWAT (B-31)
RAHUL VERMA (B-52)

UNDER GUIDANCE OF
PROF. PRAVIN RAHATE



UNIVERSITY OF MUMBAI
DEPARTMENT OF COMPUTER ENGINEERING
DATTA MEGHE COLLEGE OF ENGINEERING
PLOT NO.98 SECTOR-3, AIROLI, NAVI MUMBAI
ACADEMIC YEAR 2019-20



DATTA MEGHE COLLEGE OF ENGINEERING

AIROLI, NAVI MUMBAI

CERTIFICATE

This is to certify that the project entitled “**TWITTER TREND ANALYSIS**” is bona fide work of **SANKALP NILEKAR, SUDEEP RAWAT and RAHUL VERMA** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Undergraduate in Computer Engineering**.

Prof. Pravin Rahate
Project Guide

Dr. A. P. Pande
Head of the Department

Dr. S.D. Sawarkar
Principal



DATTA MEGHE COLLEGE OF ENGINEERING

AIROLI, NAVI MUMBAI

PROJECT APPROVAL

This project report entitled “**TWITTER TREND ANALYSIS**” by
SANKALP NILEKAR, SUDEEP RAWAT and RAHUL VERMA is approved for the
degree of **Computer Engineering**.

Internal Examiner

Date:

Place:

External Examiner

Date:

Place:

DECLARATION

We declare that, this written submission represents our ideas in our 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 our 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.

Name of the Students

Signature

Sankalp Nilekar (B-17)

Sudeep Rawat (B-31)

Rahul Verma (B-52)

ACKNOWLEDGEMENT

Motivation and guidance are the keys towards success. I would like to extend my thanks to all the sources of motivation.

We would like to grab this opportunity to thank **Dr. S. D. Sawarkar, Principal** for encouragement and support he has given for our project.

We express our deep gratitude to **Dr. A. P. Pande, Head of the Department** who has been the constant driving force behind the completion of this project.

We wish to express our heartfelt appreciation and deep sense of gratitude to my project guide **Prof. Pravin Rahate** for his encouragement, invaluable support, timely help, lucid suggestions and excellent guidance which helped us to understand and achieve the project goal. His concrete directions and critical views have greatly helped us in successful completion of this work.

We extend our sincere appreciation to all Professors for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individually.

We are also thankful to all those who helped us directly or indirectly in completion of this work.

Place:

Date:

Name of the students

Sankalp Nilekar

Sudeep Rawat

Rahul Verma

ABSTRACT

The community of users participating in social media tends to share common interests at the same time, giving rise to what are known as social trends. A social trend reflects the voice of a large number of users which, for some reason, becomes popular in a specific moment. Through social trends, users, therefore, suggest that some occurrence of wide interest is taking place and subsequently triggering the trend. In this work, we explore the types of triggers that spark trends on the microblogging site Twitter and introduce a typology that includes the following four types: news, ongoing events, memes, and commemoratives. The user will be allowed to search for the latest trends by inputting a keyword into the search field. Based on user-provided keyword, the system will search for similar keywords in database and summarize the total count to provide the trending tweets on twitter. The trending tweets with the hashtag (#) will be displayed first and then the rest words will be displayed. By clicking on every trending tweet, the user commented tweets will be displayed. User can view all the tweets from the searched keyword. One of the main features on the homepage of Twitter shows a list of top terms so-called trending topics at all times. These terms reflect the topics that are being discussed most at the very moment on the site's fast-flowing stream of tweets. In order to avoid topics that are popular regularly (e.g., good morning or goodnight on certain times of the day), Twitter focuses on topics that are being discussed much more than usual, i.e., topics that recently suffered an increase of use, so that it trended for some reason.

Trending topics have attracted big interest not only among the users themselves but also among other information consumers such as journalists, real-time application developers, and social media researchers. Being able to know the top conversations being discussed at a given time helps keep updated about current affairs, and discover the main concerns of the community. Twitter defines trending topics as “topics that are immediately popular, rather than topics that have been popular for a while or on a daily basis”. However, no further evidence is known about the algorithm that extracts trending topics. It is assumed that the list is made up by terms that appear more frequently in the most recent stream of tweets than the usual expected.

INDEX

	Acknowledgement	i
	Abstract	ii
	List of Figures	iii
	List of Tables	iv
1.	Introduction	1
1.1	Aim and Objective	1
1.2	Problem Statement	2
1.3	Scope of the Project	3
2.	Literature Review	5
3.	Requirement Analysis	7
3.1	Use Case Diagram	8
3.2	Proposed Model	9
3.3	Process	10
3.4	Software and Hardware Requirements	12
4.	Project Design	13
4.1	Collecting data via Twitter API	13
4.2	Tweets Clusters	13
4.3	Twitter API	13
5.	Design Implementation	16
5.1	GUI	16
5.2	Chart	17
6.	Technologies Used	18
6.1	Python	18
6.2	API	18
7.	Testing	20
8.	Results	22
9.	Conclusions	26
9.1	Contributions	26
9.2	Future Work	26
10.	References	27

LIST OF FIGURES

Sr. No.	Name of the Figure	Pg. No.
1.	Use Case Diagram for Twitter Trend Analysis	8
2.	Model for Twitter Trend Analysis	9
3.	Diagram of Process of Twitter Trend Analysis	11
4.	Results	22

LIST OF TABLES

Sr. No.	Name of the Table	Pg. No.
1.	Summary of Literature Review	6
2.	Test Cases	21

Chapter 1

INTRODUCTION

1.1 Aim and Objective:

In the times of information age, the magnitude of online social media activity has reached an unprecedented level. Hundreds of millions of users spend hours online every day to stay connected and communicate with the rest of the world. Millions of users participate in these social networks of Social awareness streams. [19] People generate huge amount of data every day on various social media networks, which in aggregate indicate the interests and current attention of the local and global communities. There are many events and topics discussed on Twitter. Some topics may get a lot of attention and some may not. Some of these topics become very popular and focus of interests for a large number of people. The connections and the nature of social network let information disseminate to a large number of other people, a phenomenon known as going “viral”. These popular topics of discussions are also called “trends” in the social network. These trends are very dynamic and temporal in nature which exposes the aggregate interests and attention of global and local communities.

Trends in social networks are of high significance and a major point of interest in both the industry [19] and the research community [8, 15]. Many applications on web and business can be immensely benefitted from knowing what is currently “trending”, which represents an answer to the age-old query what are people talking about [9]. From stock exchange making real-time decision to search engines delivering more updated, relevant search results. Twitter is one of the most popular social networking and micro-blogging service, which had more than 200 Million registered users by 2013, producing 400 Million tweets every day [17]. As a microblogging website it allows its users to create a short text message of 140 characters as their posts called “tweets”. There are also many different ways for users to update their tweets, including the mobile phone, web and text messaging tools [14] and so on. Twitter is also very real-time in nature. In pasts, several events were reported on twitter as news hours

earlier than the mainstream media. [10]. Hence twitter is a very robust source for getting the real-time trends in the web.

The numbers of active users and tweets generated daily are enormous and hence, they collectively can give crucial clues to several interesting problems such as public opinion analysis and hot trend detection. Twitter employs a social model called following [14], in which the user is allowed to choose any other users that they want to follow without any permission or reciprocating by following them back. The one they follow is their friend, and they are the follower. Being a follower on Twitter means that they receive all the updates of their friends [4]. This makes **twitter a directed social network** where directed links could represent anything from intimate friendships to common interests, or even a passion for breaking news or celebrity gossip. Such directed links determine the flow of information and hence indicate a **user's influence** on others a concept that is crucial in sociology and viral marketing. The major drawback with using Twitter as a source of information is that not all of the tweets are informative. Contrary to it, the majority of tweets are “chatty” or “spammy” in nature [7]. So, it's crucial to filter out this noise from data to use the useful “informative” tweets. Hence, we need a system, which could separate useful tweets, which essentially means a classifier model to classify “informative” tweets from “chat” tweets. This system should work in a single pass and also be robust and fast enough to process up to 400 million tweets a day. Our attempt in this thesis is to provide such a model of a framework which given topic wise tweets clusters will be able to detect current trends and also predict some upcoming trends in their early stage using the social graph of twitter.

1.2 Problem Statement:

Our problem statement can be formally defined as follows.

Create an automated system, which takes in a continuous stream of raw tweets, processes these tweets to filter out “noise” and get relevant informative tweets. Further mine these “filtered” tweets to detect and predict evolving “trending” topics in their very early stage.

There are two key aspects to this problem:

- Creating a model to process real-time tweets feed to filter out the “noise” and output only relevant tweets.
- Creating a system to store the social graph of twitter users such that it can be efficiently used as a data structure for answering various user queries and get relations for a given user.

1.3 Scope of the Project:

- Twitter Trend Analysis helps us to know what people are talking about:

Twitter Trend analysis searches for the trends happening around the world and thus we can know about the things people are talking about or the things people are currently interested in talking. What we do here in Twitter trend analysis is filter out the noise in the tweets and provide the user with the relevant information. Using this we come to know what the users require and such data can be used by various companies to provide specific services to customers. This data provided by Twitter trend analysis is very helpful for activity-based advertisements which are helpful for both the company providing it and the user as well.

- We can explore the trending topics using Trend Analysis:

Twitter helps us to provide information instantaneously and thus it can be used to spread important information around the world without any delays and maximum coverage. News channels prefer twitter for giving any information due to its reach. Popular topics in the region can be found using Twitter Trend Analysis. This helps with getting people aware of the things that are currently trending in the region.

- It helps us reach a wider audience:

Twitter has a large user base, which could include your potential customers. Using hashtags can help you reach an audience interested in a particular topic or in a particular location. The platform allows direct two-way communication with your customers. Because it's a business interaction, if you do it well it shows your business in a positive light.

- Twitter Trend Analysis will help us remove spam:

Twitter is filled with spam messages or noise just like any other social media platform which introduces lots of problems in the work of many people. Twitter Trend Analysis helps us solve all these issues by filtering out the noise and displaying just the relevant information that is needed by the user. This helps in increased productivity and reduces the work of the users. Reducing spam also reduces duplication and virus attacks. The noise removal taking place in Twitter Trend Analysis is really important and is one the most important aspect of it.

- Future Tweets Will Be Searched on Search Engines:

If you have searched on Google, you must have seen that tweets appear among the search results. In the future twitter, we expect that searching tweets will apply on Google and Bing. This means that you will be able to search for a specific tweet. This will be achieved by entering the tweet wording on the search bar. The relating tweets will appear as tweets results on your browser. Currently, the tweets from big brands appear on Google searches. Google featured tweets will be within everyone's grasp soon.

Chapter 2

LITERATURE REVIEW

Trend analysis and based on that predicting public opinions. It plays important role, many researchers working on the automatic technique of extraction and analysis of huge amount of twitter data. In [1] the author compares six trend detection method and find that standard natural language processing technique perform well for social streams on a particular topic. They conclude that n-gram gives the best performance other than state-of-art techniques. In [4], the authors have used three different machine learning algorithms Naïve Bayes, Decision Trees and Support Vector Machine for sentiment classification of Arabic dataset which was obtained from Twitter. This research has followed a framework for Arabic tweets classification in which two special sub-tasks were performed in preprocessing, Term Frequency-Inverse Document Frequency (TF-IDF) and Arabic stemming. They have used one dataset with three algorithms and performance has been evaluated on the basis of three different information retrieval metrics precision, recall, and f-measure. In [6] author proposed supervised learning techniques to classify twitter trending topic for that they use text-based and network-based classifier and conclude C5.0 gave the best performance.

In [19] author propose model which predict public opinion on the political event by Applying different classifier which predicts whether the mood is positive or negative. In [26], the authors proposed a way to get the pre-labelled data from twitter which can be used to train the SVM classifier. They used the twitter hashtags to judge the polarity of a tweet. To analyze the accuracy of the proposed technique, a test study on the classifier was conducted which showed the result with the accuracy of 85%. The authors in [27] introduced a new technique to classify the sentiment of tweets as positive or negative. They presented and discussed the results of machine learning algorithms for twitter sentiment analysis by using distant supervision. Training data, the authors used consisted of tweets with emotions which were used as noisy labels. According to authors, the machine learning algorithms such as Naive Bayes, Maximum Entropy and SVM when trained with emotion tweets can have accuracy more than 80%. The study also highlighted the steps used in the preprocessing stage of classification for high accuracy. In [28] sentiment analysis performs using SVM in that two pre-classified datasets of tweets are used then do a comparative analysis, they use measures Precision, Recall and F-Measure.

Title of Paper	Author	Published in	Results
Sensing Trending Topics in Twitter	Luca Maria Aiello, Georgios Petkos, Carlos Martin, David Corney, Symeon Papadopoulos, Ryan Skraba, Ayse Göker, Ioannis Kompatsiaris	IEEE Transactions on Multimedia, Vol. 15, No. 6, October 2013.	Author compare six trend detection method and find that standard natural language processing technique perform well for social streams on particular topic. They conclude that n-gram give best performance other than state of art techniques
Comparison of Machine Learning Approaches on Twitter Trend Analysis	Altawaier, M. M., & Tiun.	International Journal on Advanced Science, Engineering and Information Technology, 6(6)	The Authors have used one dataset with three algorithms and performance has been evaluated on the basis three different information retrieval metrics precision, recall, and f-measure.
Twitter Trending Topic Classification	Kathy Lee, Diana Palsetia, Ramanathan Narayanan, Md. Mostofa Ali Patwary, Ankit Agrawal,	2011 11th IEEE International Conference on Data Mining	Author proposed supervised learning techniques to classify twitter trending topic for that they use text based and network-based classifier and conclude C5.0 gave best performance
Study using Support Vector Machines to Classify the Sentiments of Tweets	Zgheib, W. A., & Barbar	CS224N Project Report, Stanford, 1(2009), 12	The authors proposed a way to get the pre labeled data from twitter which can be used to train SVM classifier. They used the twitter hash tags to judge the polarity of tweet. To analyze the accuracy of proposed technique

Chapter 3

REQUIREMENT ANALYSIS

- Tweet impressions:

Under the Tweets section, you can find a list of all your Tweets and the number of impressions. You can see individual Tweet performance, as well as recent months or a 28-day overview of cumulative impressions. Capitalize on this information by repurposing Tweets that gained the most impressions, or creating Tweets on a similar subject. You can also use the cumulative overview to compare monthly activity.

- Tweet engagements and engagement rate:

Similar to impressions, the Tweets section also shows your Tweets engagement, or the number of interactions your Tweet has received, as well as the engagement rate, which is engagements divided by impressions. If your Tweets are receiving little engagement, you may want to rethink your subject matter and format, for instance, you may want to add photo or video to your content mix, which tends to generate more engagement.

- Event and trending topic data:

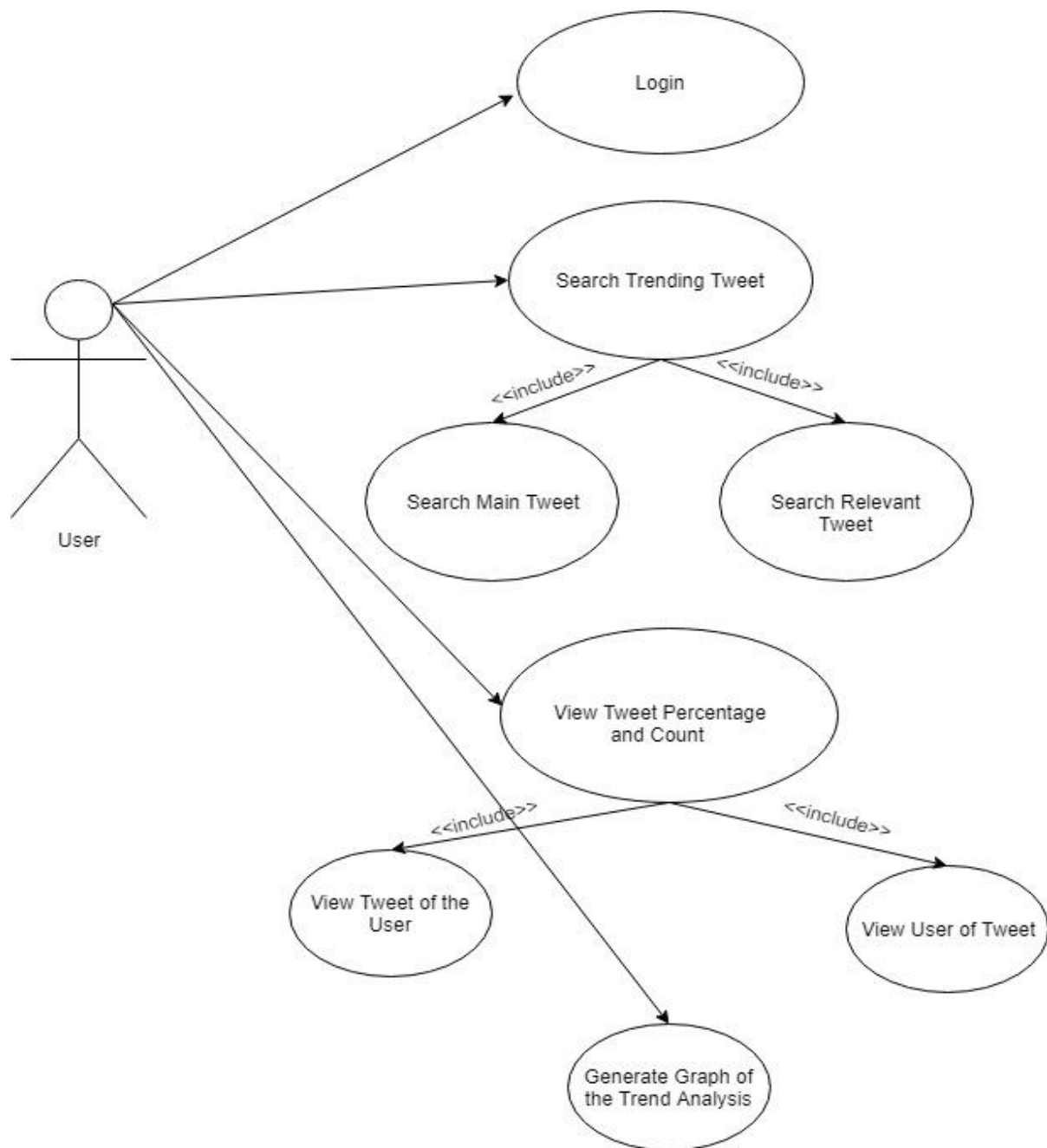
Discover upcoming holidays, events, and recurring trends, and find out who's Tweeting about them. This is great way to find potential new content ideas, and conversations to join in on.

- Video content performance:

If you're using video as part of your content strategy, you can track your video views, as well see a bigger picture of how people are responding to your videos. For instance, are they watching it to completion? If you want to fine-tune your Twitter strategy, spending some time understanding your Twitter analytics is a great place to start. Get started by viewing your Twitter analytics dashboard today.

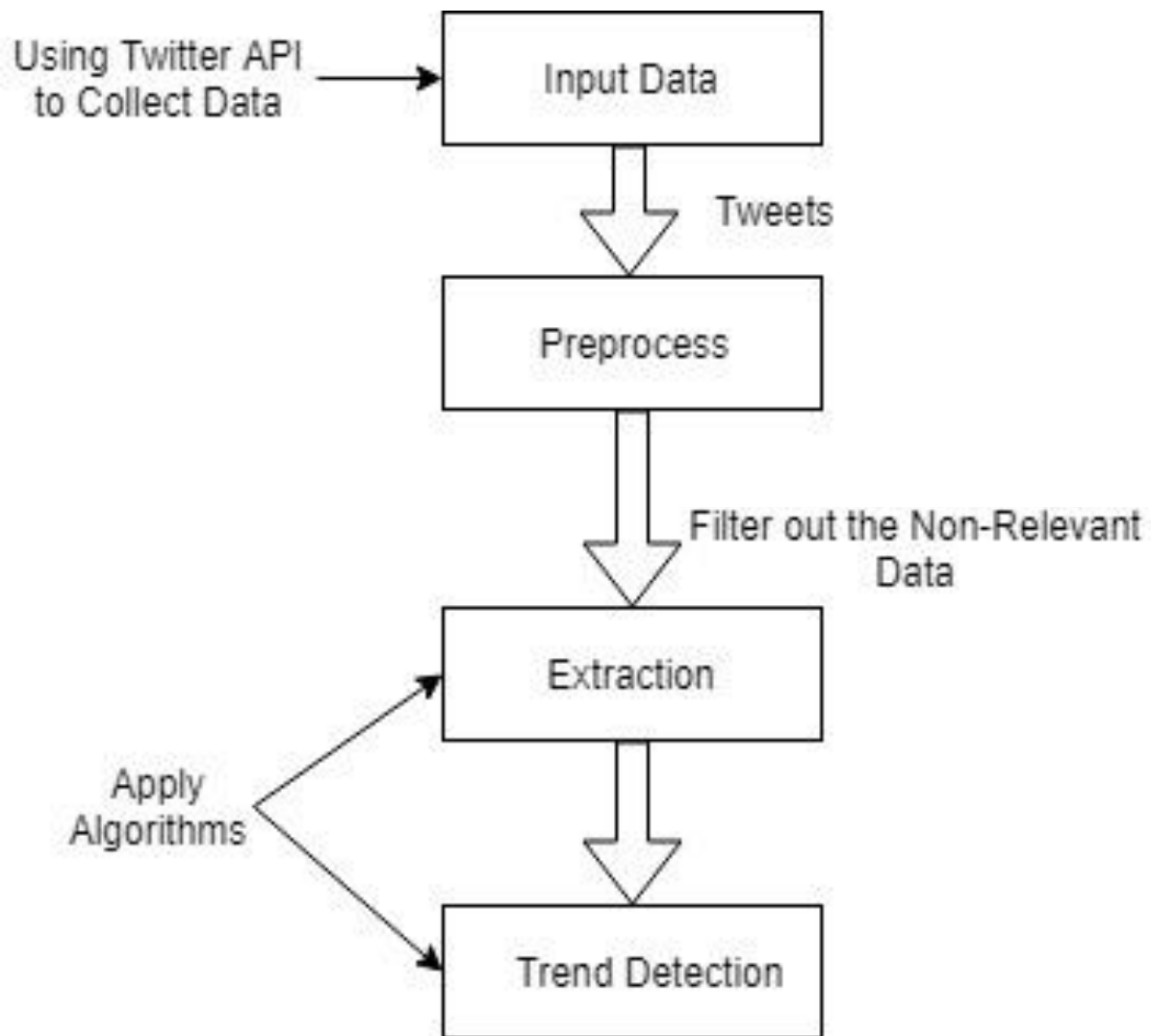
3.1 Use Case Diagram:

A Use Case Diagram is a graphic depiction of the interactions among the elements of a system. A Use Case Diagram contains four components- Boundary, Actors, use cases and Relationships among actors and use cases.



Use Case Diagram for Twitter Trend Analysis (fig. 3.1.1)

3.2 Proposed Model:



Model for Twitter Trend Analysis (fig. 3.2.1)

The model has following steps:

- Data collection of tweets
- Pre-process tweet
- Extraction
- Trend Detection

1) Dataset: Collect tweet data through twitter streaming API. Which download tweets in JSON format. We can apply keyword, hashtag, username to download tweets related to them.

2) Pre-processing: Tweet pre-processing module having several stages. After downloading tweets, we have to extract text data form that and discard video, audio, image etc. store English text which is retrieve form tweet. Then remove @, #, URL and other punctuation form tweets and apply stop word remove, word tokenize.

3) Extraction: After pre-processing stage next module is Extraction which is done in two way through Term frequency calculation and pos tagging. We can view the Count, Strength etc. after the extraction of the data, which can therefore, be used to view tweets using similar hashtags from different users

4) Trend Detection: We can determine trend by using TF-IDF calculation. And predict positive, negative, neutral mood tendency by applying machine learning algorithms. Show the graph for the given input by the user.

3.3 Process:

The system comprises of 4 major modules with its description as follows:

1) Login:

User need to login first using valid credentials to access the system.

2) Search for Latest Trends:

After successful login, user can search for latest trending tweets by entering the keyword in the search column.

3) View Latest Trending Tweet:

Based on user-inputted keyword, the search results will be displayed in form of trending tweets.

4) View Tweets:

User can click on respective trending tweet to view the message twitted by other users

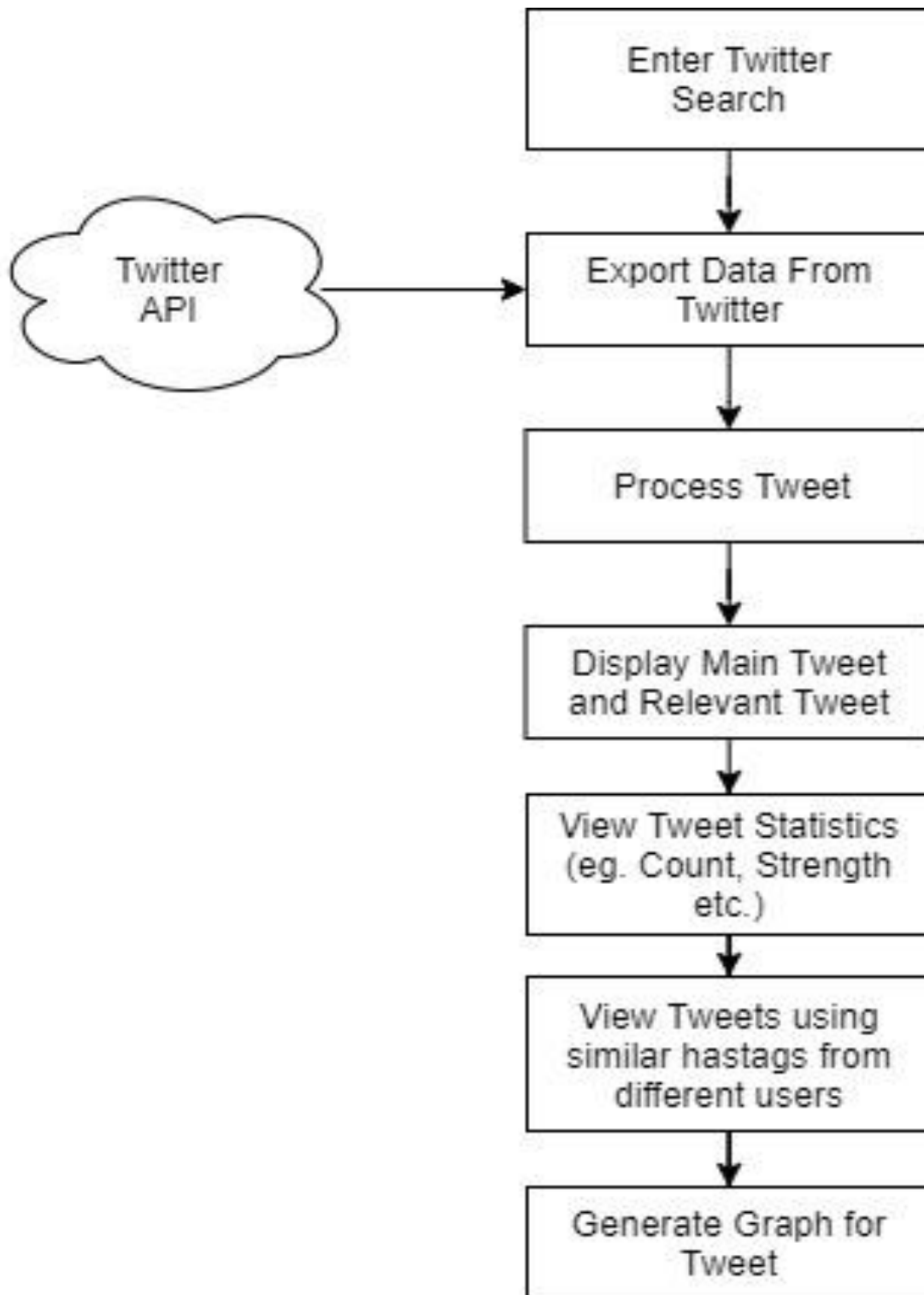


Diagram of Process of Twitter Trend Analysis (fig. 3.3.1)

3.4 Software and Hardware Requirements:

Software Requirements:

- Windows 7 or higher.
- Python 3.0 or higher.

Hardware Components:

- Processor – i3
- Hard Disk – 5 GB
- Memory – 1GB RAM
- Internet Connection

Chapter 4

PROJECT DESIGN

4.1 Collecting data via Twitter API:

As the mentioned earlier core of our approach is the social graph, which consists of twitter users and relations among them. The current size of twitter has more than 200 million nodes or users in it. [17] Twitter exposes these relations to the public via their API. [4] where relations like “followers” and “friends” (users a person follows) can be collected via API calls. The major blocking factor here is the rate limit to a number of calls, which can be made to API. We created more than 10 different API keys for created several instances of nodes making calls to Twitter API. We wrote a python library to communicate with the API. Also, the Twitter API was upgraded several times by Twitter resulting in a change of API methods and responses. Hence required changes in our code accordingly. Also due to heavy traffic or technical glitches, there were frequent outages in the API service causing further delay and blocking.

4.2 Tweets clusters:

We need all the tweets related to a particular topic together which we call a cluster. This tweets cluster tells us who all users are involved in a discussion of that particular topic.

4.3 Twitter API:

When someone wants to access Twitters APIs, they are required to register an application. By default, applications can only access public information on Twitter. Certain endpoints, such as those responsible for sending or receiving Direct Messages, require additional permissions from us before they can access your information. These permissions are not granted by default, we choose on a per-application basis whether to provide this access, and can control all the applications authorized on our account.

The Twitter APIs include a wide range of endpoints, which fall into five primary groups:

- Accounts and users:

Twitter allows developers to programmatically manage an account's profile and settings, mute or block users, manage users and followers, request information about an authorized account's activity, and more. These endpoints can help citizen services like the Commonwealth of Virginia's Department of Emergency Management which provides information to residents about emergency responses and emergency alerts.

- Tweets and replies:

Twitter makes public Tweets and replies available to developers, and allow developers to post Tweets via our API. Developers can access Tweets by searching for specific keywords, or requesting a sample of Tweets from specific accounts. These endpoints are used by NGOs like the UN to identify, understand and counter misinformation around public health initiatives. For example, in Indonesia, there were persistent rumors that vaccinations either contained pork product or caused infertility. Understanding how the rumors started and were spread allowed the UN to get a team on the ground to help dispel these myths, which caused particular concern in this Muslim-majority nation. Similarly, Twitter helps researchers listen for early symptomatic indications of disease outbreaks and monitor their spread. A team at Northeastern University recently developed a new flu-tracking technique using Twitter data that forecasts outbreaks up to six weeks in advance which is much earlier than many other models, without sacrificing accuracy.

- Direct Messages:

Twitter's Direct Message endpoints provide access to the DM conversations of users who have explicitly granted permission to a specific application. Twitter's DM APIs provide limited access to developers to create personalized experiences on Twitter, like Wendy's March Madness bracket builder. For accounts they own or manage, businesses can create these human- or chatbot-powered conversational experiences to communicate directly with customers for customer service, marketing, and brand engagement experiences.

- Ads:

Twitter provides a suite of APIs to let developers, like Sprinklr, help businesses automatically create and manage ad campaigns on Twitter. Developers can use public

Tweets to identify topics and interests, and provide businesses with tools for running advertising campaigns to reach the diverse audiences on Twitter.

- **Publisher tools and SDKs:**

Twitter provides tools for software developers and publishers to embed Twitter timelines, share buttons, and other Twitter content on webpages. These tools allow brands to bring live, public conversations from Twitter into their web experience and make it easy for their customers to share information and articles from their sites.

Chapter 5

DESIGN IMPLEMENTATION

5.1 GUI:

A GUI (graphical user interface) is a system of interactive visual components for computer software. A GUI displays objects that convey information, and represent actions that can be taken by the user. The objects change color, size, or visibility when the user interacts with them. GUI objects include icons, cursors, and buttons. These graphical elements are sometimes enhanced with sounds, or visual effects like transparency and drop shadows. A GUI is considered to be more user-friendly than a text-based command-line interface, such as MS-DOS, or the shell of Unix like operating systems. The GUI was first developed at Xerox PARC by Alan Kay, Douglas Engelbart, and a group of other researchers in 1981. Later, Apple introduced the Lisa computer with a GUI on January 19, 1983. A GUI uses windows, icons, and menus to carry out commands, such as opening, deleting, and moving files. Although a GUI operating system is primarily navigated using a mouse, a keyboard can also be used via keyboard shortcuts or the arrow keys. As an example, if you wanted to open a program on a GUI system, you would move the mouse pointer to the program's icon and double-click it. Unlike a command-line operating system or CUI, like Unix or MS-DOS, GUI operating systems are much easier to learn and use because commands do not need to be memorized. Additionally, users do not need to know any programming languages. Because of their ease of use and more modern appearance, GUI operating systems have come to dominate today's market. A pointing device, such as the mouse, is used to interact with nearly all aspects of the GUI. More modern (and mobile) devices also utilize a touch screen. However, as stated in previous sections, it is also possible to navigate a GUI using a keyboard.

5.2 Chart:

Charts are often used to ease understanding of large quantities of data and the relationships between parts of the data. Charts can usually be read more quickly than the raw data. They are used in a wide variety of fields, and can be created by hand (often on graph paper) or by computer using a charting application. Certain types of charts are more useful for presenting a given data set than others. For example, data that presents percentages in different groups (such as "satisfied, not satisfied, unsure") are often displayed in a pie chart, but may be more easily understood when presented in a horizontal bar chart. [2] On the other hand, data that represents numbers that change over a period of time (such as "annual revenue from 1990 to 2000") might be best shown as a line chart:

- **Histogram:** A histogram is an approximate representation of the distribution of numerical or categorical data. It was first introduced by Karl Pearson. To construct a histogram, the first step is to "bin" (or "bucket") the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent, and are often (but not required to be) of equal size. If the bins are of equal size, a rectangle is erected over the bin with height proportional to the frequency—the number of cases in each bin. A histogram may also be normalized to display "relative" frequencies. It then shows the proportion of cases that fall into each of several categories, with the sum of the heights equaling 1.
- **Scatter Plot:** A scatter plot can be used either when one continuous variable that is under the control of the experimenter and the other depends on it or when both continuous variables are independent. If a parameter exists that is systematically incremented and/or decremented by the other, it is called the control parameter or independent variable and is customarily plotted along the horizontal axis. The measured or dependent variable is customarily plotted along the vertical axis. A scatter plot can be used either when one continuous variable that is under the control of the experimenter and the other depends on it or when both continuous variables are independent.

Chapter 6

TECHNOLOGIES USED

6.1 Python:

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library. Python was conceived in the late 1980s as a successor to the ABC language. Python 2.0, released in 2000, introduced features like list comprehensions and a garbage collection system capable of collecting reference cycles. Python 3.0, released in 2008, was a major revision of the language that is not completely backward-compatible, and much Python 2 code does not run unmodified on Python 3. The Python 2 language, i.e. Python 2.7.x, was officially discontinued on 1 January 2020 (first planned for 2015) after which security patches and other improvements will not be released for it. With Python 2's end-of-life, only Python 3.5.x and later are supported. Python interpreters are available for many operating systems. A global community of programmers develops and maintains CPython, an open source reference implementation. A non-profit organization, the Python Software Foundation, manages and directs resources for Python and CPython development.

6.2 API:

An application programming interface (API) is a computing interface to a software component or a system, that defines how other components or systems can use it. It defines the kinds of calls or requests that can be made, how to make them, the data formats that should be used, the conventions to follow, etc. It can also provide extension mechanisms so that users can extend existing functionality in various ways and to varying degrees. An API can be entirely custom, specific to a component, or it can be designed based on an industry standard to ensure

interoperability. Some APIs have to be documented, others are designed so that they can be "interrogated" to determine supported functionality. Since other components/systems rely only on the API, the system that provides the API can (ideally) change its internal details "behind" that API without affecting its users. Today, with the rise of REST and web services over HTTP, the term is often assumed to refer to APIs of such services when given no other context (see the Web APIs section). Sometimes the term API is, by extension, used to refer to the subset of software entities (code, subcomponents, modules, etc.) that serve to actually implement the API of some encompassing component or system. In building applications, an API (application programming interface) simplifies programming by abstracting the underlying implementation and only exposing objects or actions the developer needs. While a graphical interface for an email client might provide a user with a button that performs all the steps for fetching and highlighting new emails, an API for file input/output might give the developer a function that copies a file from one location to another without requiring that the developer understand the file system operations occurring behind the scenes.

Chapter 7

TESTING

The basic goal of software development process is to produce the software that has very few or no errors. In an effort to detect errors soon after they are introduced, each phase ends with verification activity such as review. However, most of these verification activities in the early phase of software development are based on human evaluation. So, it is impossible to detect all errors in that phase itself. Testing plays an important role in quality assurance of the software. It is dynamic method for verification and validation, where system to be tested, is executed under various constraints and its behavior or response is observed.

- **Test Plans:**

A test plan is a general document for the entire project that defines entire scope approach to be taken and schedule of testing as well as identifies the test item for testing. The test planning can be done well before the actual testing commences and also can be done in parallel in design and implementation phase.

- **Unit Testing:**

Unit tests are tests that exercise a class or module in isolation from rest of the application, and form a significant element of Test Driven Development where failing tests are written to allow the application to be ‘debugged into existence’. Here, various modules have been tested individually. This has been done manually to test cases with the help of which the application has been tested.

- **Performance Testing:**

Performance Testing has been done to measure the responsiveness of application to workload such as increased user request and the type of internet connection. The parameters were chosen randomly for anticipation of errors or bug which resolved until the application performed consistently.

- **Compatibility Testing:**

Compatibility Testing is a type of software testing used to ensure compatibility of the system application with various other hardware specifications. It is a type of nonfunctional testing. This type of testing helps us to find out how well the system application performs in a particular environment that includes hardware, network, operating system and other software. It is basically the testing of application or the product built with computing environment. It tests whether the software product that is built is compatible with other hardware, operating tests whether the software product that is built is compatible with other hardware, operating system, database or other system software or not. Twitter Trend Analysis is compatible with Windows 7 or later.

- **System Testing:**

System Testing is actually a series of different tests whose sole purpose is to exercise the full computer based system. System testing falls under the black box testing category of software testing. White box testing is the testing of internal workings or coding implementations of software application. In contrast, black box or system testing is the opposite. System Testing involves the external working of the application from user's perspective. Here, after testing each module individually, all the classes and/or modules are integrated and tested as a whole application.

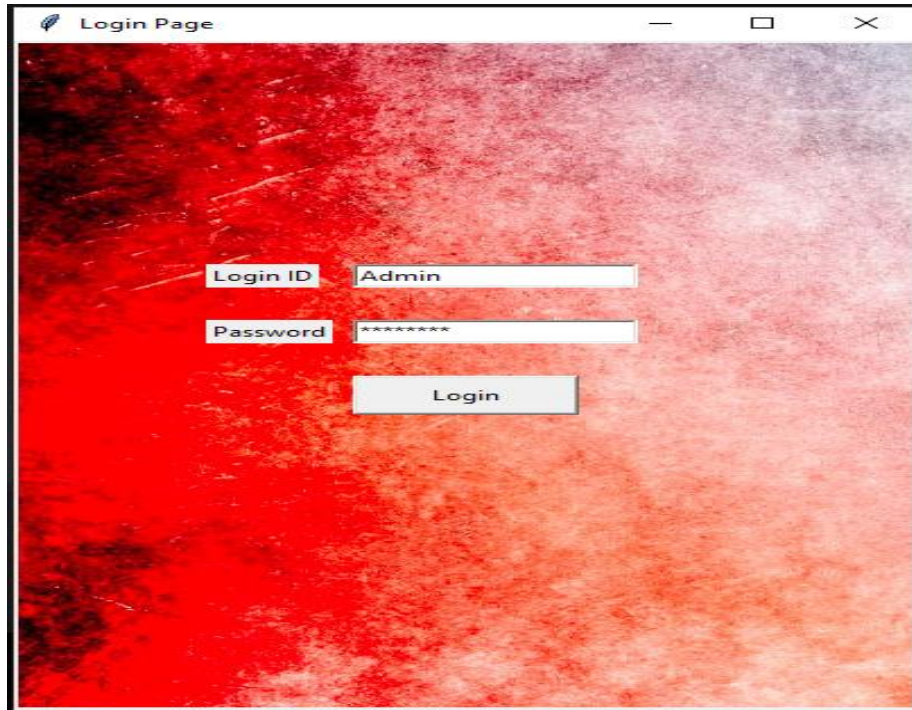
- **Test Cases:**

Sr. No.	Test Case Description	Expected Result	Actual Result
1.	User Login with correct user id and password	Successful login	Pass
2.	For valid input	Output Screen (Username, Location, Tweets, Date) should be displayed	Pass
3.	For invalid input	Display "Please enter a valid Hashtag"	Pass

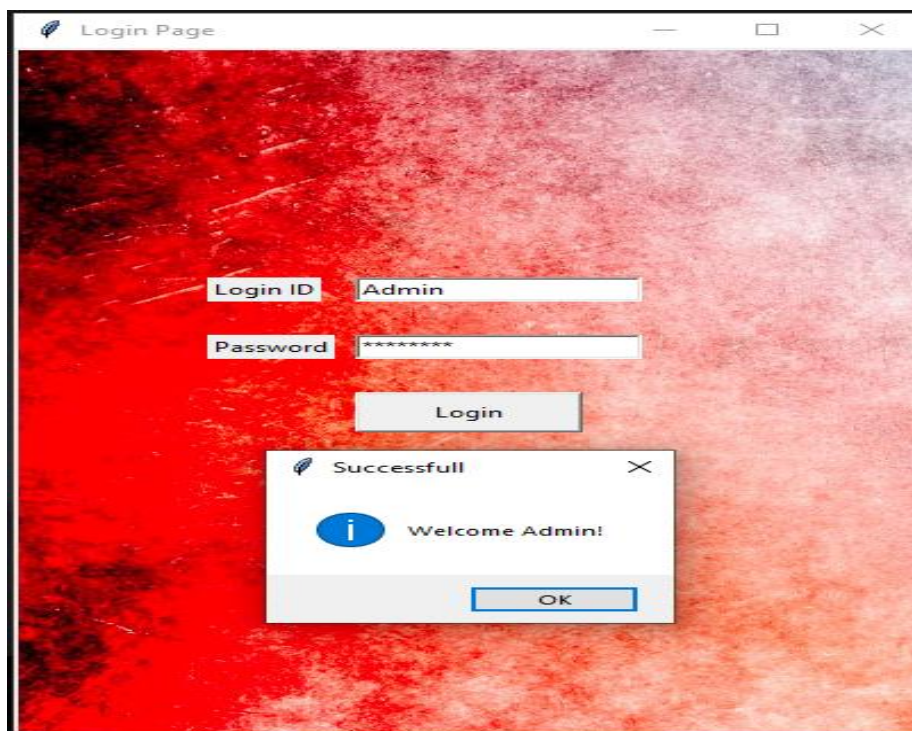
Chapter 8

RESULTS

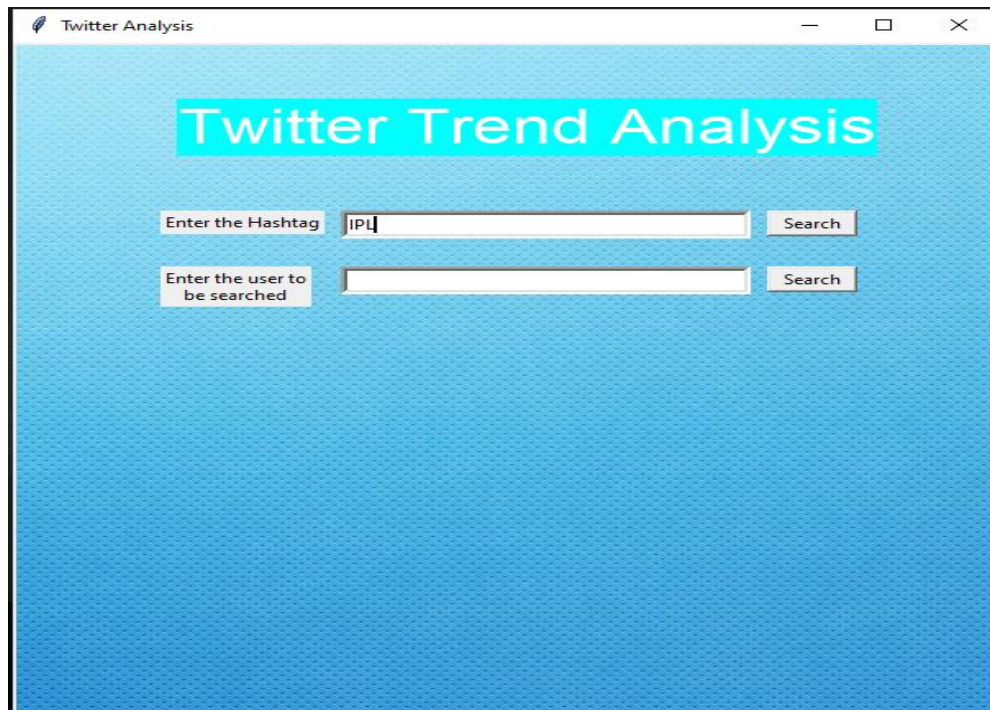
1) Login Page:



2) Login Attempt Successful:



3) Enter Hashtag Screen:

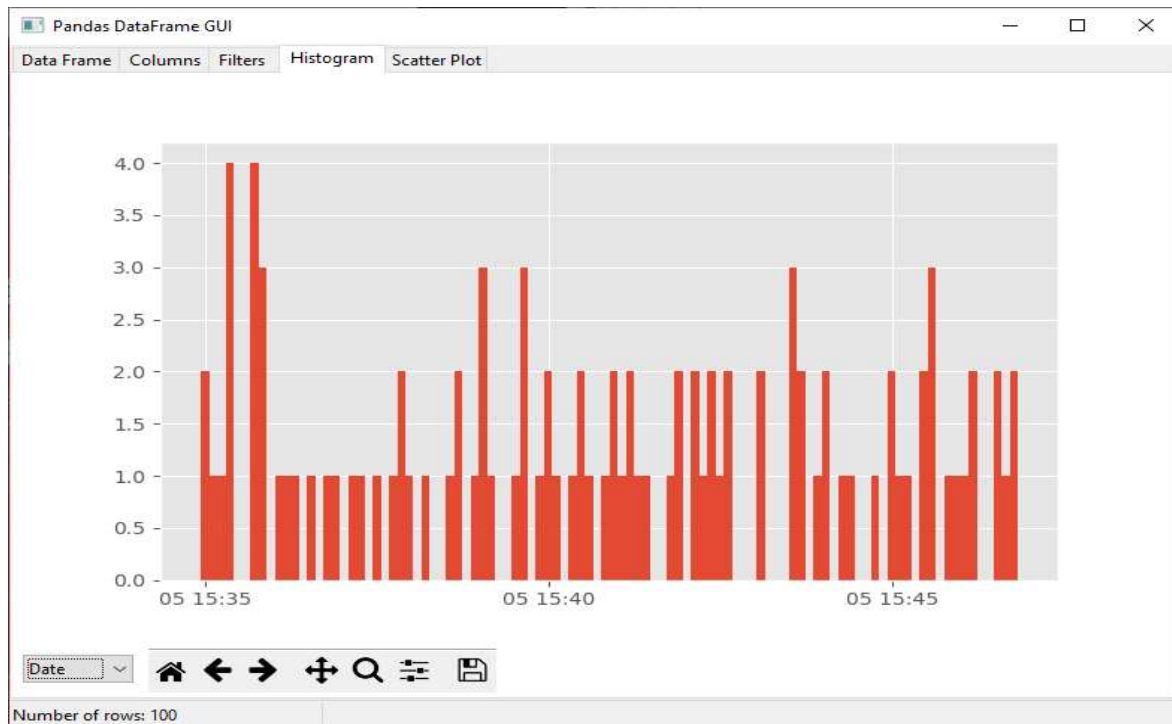


4) User, Tweet, Location, Date Output:

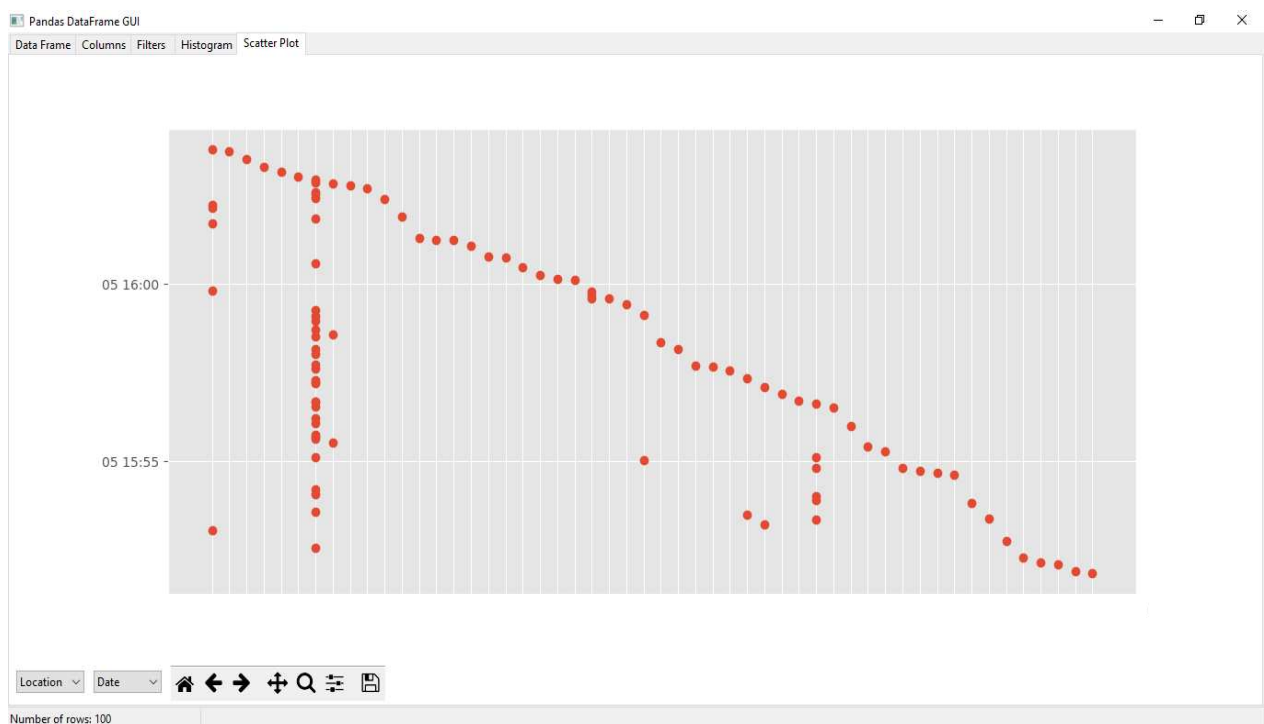
User	Tweet	Location	Date
king_18__	RT @SELFISH_DHONI: 1. Kohli's Winning % Greater than Dhoni 2. IPL Cup :#DHONI 3 Cup out of 11 IPL season. #ROHIT 4 Cup out of 7 Season...	Kurnool, India	2020-03-05 15:46:50
ManuDr9143516	@ABdeVilliers71Sir 2020 ipl trophy rcb win or loss...Do what you say (win are loss i am big fan of rcb) virat &abd	Karnataka, India	2020-03-05 15:46:47
sanji_3301	RT @comedibanda: RCB to other teams in IPL https://t.co/ser9aqX5l		2020-03-05 15:46:36
bharath9520	RT @Shebas_10dulkar: IPL Teams & Their H2H Records in IPL1(1/2) https://t.co/9fwy4i3X7T	sendamangalam,namakkal	2020-03-05 15:46:34
Saravana_Dfan	RT @Saravana_Dfan: Any @IonsdenKup Fans ூ This Year KXIP than IPL Winner Take \$S @கனத்திங்க சந்திப்புகடாம் @#IPL2020 https://t.co/zWAC...	தனுஷ் இதயத்தில்	2020-03-05 15:46:31
74Nitinip	RT @msdfansofficial: As quoted by T0I: CAC asked the candidates, 'Whether they would pick Dhoni if he had a good IPL, even though he hasn...'.	India	2020-03-05 15:46:14
Viper1veets	RT @KKRiders: Skipper DK &amp; Coach Baz @ Opponents ூ to Team mates ூ to Pillars of strength for KKR! How many #KKR captain-coach combos ...	Vaipulla Raja	2020-03-05 15:46:08
GOATian263	RT @HTSportsNews: #IPL2020 Franchises to protect to BCCI over additional staging fees Writes @rkrmasesh@BCCI @IPL https://t.co/ygWpdaRnA		2020-03-05 15:46:04
RajeshA68591831	SC to hear Dream 11 case ahead of IPL https://t.co/aFyBilvK	चमौली, उत्तराखण्ड, भारत	2020-03-05 15:45:54
sachisingh	RT @Real_Anju: @Dream11's alleged 2000cr sports betting case in the court of CJI himself tomorrow! Big momenthttps://t.co/CJUPnGvG5b	India	2020-03-05 15:45:47
SMOHAMMEDNOVIFA3	RT @SiDhoni: Name the player @s#IPL #Dhoni https://t.co/duF2wId9F	India	2020-03-05 15:45:36
dna	'I got plenty of time to plan: KKR spinner Kuldeep Yadav eyeing IPL 2020 to cement spot in India's T20 World Cup s... https://t.co/hQhpkLLKQc	India	2020-03-05 15:45:34
SELFISH_DHONI_	RT @M_off: Only Cricketer To Win the IPL Both As player & Captain...The name is... #RohitSharma @ImRo45 @mipaltan#Mumbaindians#O...		2020-03-05 15:45:33
ATHadishetty	RT @IPL: Can't wait for an even-tful start to this #VIVOIPL season. @ImRo45i @Kya is baar bhi @mipaltan ka #KheBolega? Find out LIVE, M...		2020-03-05 15:45:29
INDIANGAMERS8	RT @KhabriKaPaPa: The Biggest Event Of The Year Is Coming. How Many @IPL Fans Here?? If You Are RCB Fan Follow - @RCBOfficialC ONLY FO...	Uttar Pradesh, India	2020-03-05 15:45:26
lostt somewhere	RT @Chitrakshiv: DELHI CAPITALS will win IPL this year	भारत	2020-03-05 15:45:14
OmKrishnaPande5	RT @Manish31j: More than Coronavirus, we all afraid of IPL getting cancelled.	Gorakhpur, India	2020-03-05 15:45:05
nepalbasak4	RT @GOPALBASAK5: IPL Prize Money Reduced To Half As BCCI Seeks https://t.co/yoHru4BQMj https://t.co/gT7toWwL		2020-03-05 15:45:01
Saravana_Dfan	Any @IonsdenKup Fans ூ This Year KXIP than IPL Winner Take \$S @கனத்திங்க சந்திப்புகடாம் @#IPL2020 https://t.co/zWAC5h310s	தனுஷ் இதயத்தில்	2020-03-05 15:44:59
BackBench_Hut	RT @dhaikloketwast Sainanter playing for Chennai Super King this #IPL https://t.co/bsuMlsW0UM	#Lonely Wolf of Snow for...	2020-03-05 15:44:45
httwitees	RT @HTSportsNews: #IPL2020 Franchises to protect to BCCI over additional staging fees Writes @rkrmasesh@BCCI @IPL https://t.co/ygWpdaRnA	India	2020-03-05 15:44:22
HTSportsNews	#IPL2020 Franchises to protect to BCCI over additional staging fees Writes @rkrmasesh@BCCI @IPL https://t.co/ygWpdaRnA	India	2020-03-05 15:44:16
Mahit7HR	RT @AskRishabh: Ye IPL CSK ka hai https://t.co/uGis9QeLcG3	Hrithik World	2020-03-05 15:44:05
HindustanTimes	IPL 2020: Franchises to protest to BCCI over additional staging feeshttps://t.co/cOSqZ4w7f	New Delhi, India	2020-03-05 15:44:02
JayMaha59999893	RT @Real_Anju: @Dream11's alleged 2000cr sports betting case in the court of CJI himself tomorrow! Big momenthttps://t.co/CJUPnGvG5b		2020-03-05 15:43:58
rjpn2112	RT @Lakshmi: Once a King, will always be a King!#IPL Season, No Question, where do I stand? My allegiance will always be with Kings in...		2020-03-05 15:43:44
dk_munro	Here's what we're reading, thinking and talking about in the @munkschool's Innovation Policy Lab this week. Includi... https://t.co/2xtpYqOm9m	Ottawa/Toronto	2020-03-05 15:43:40
McTahir0724625	RT @IPL: Can't wait for an even-tful start to this #VIVOIPL season. @ImRo45i @Kya is baar bhi @mipaltan ka #KheBolega? Find out LIVE, M...	West Bengal, India	2020-03-05 15:43:37
SurpalinRath5	RT @Real_Anju: @Dream11's alleged 2000cr sports betting case in the court of CJI himself tomorrow! Big momenthttps://t.co/CJUPnGvG5b	Malpur, India	2020-03-05 15:43:34
pravyasofficial	RT @IPL: Can't wait for an even-tful start to this #VIVOIPL season. @ImRo45i @Kya is baar bhi @mipaltan ka #KheBolega? Find out LIVE, M...	India	2020-03-05 15:43:33
sanu_shadab	@KanglaKrait Something will change in history of cricket PORKistani will never become sincere in the matter of cr... https://t.co/hwj9j26d2	Royadh	2020-03-05 15:43:08
adityasethi002	RT @Segy5B13: The Biggest Event Of The YEAR Is Coming. How Many @IPL Fans Here?? If You Are RCB Fan Follow - @RCBOfficialC ONLY FOR RC...		2020-03-05 15:43:04
TakKaal	RT @Real_Anju: @Dream11's alleged 2000cr sports betting case in the court of CJI himself tomorrow! Big momenthttps://t.co/CJUPnGvG5b		2020-03-05 15:42:35

Number of rows: 100

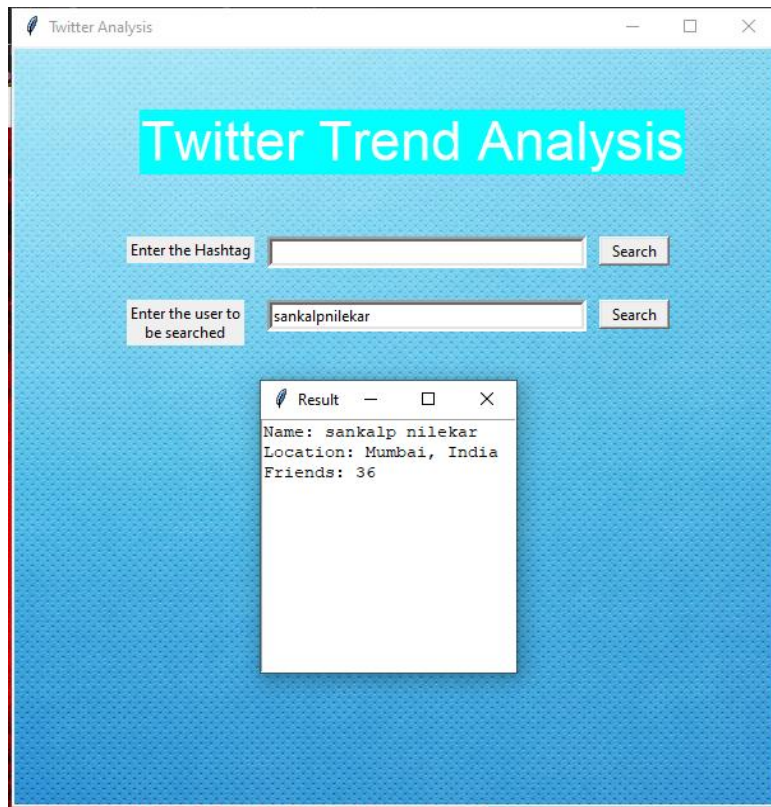
5) Histogram Output:



6) Scatter Plot Output:



7) User Search Screen:



The screenshot displays a window titled "Twitter Analysis" with a blue background. At the top, the title "Twitter Trend Analysis" is shown in a yellow box. Below this, there are two input sections. The first section is labeled "Enter the Hashtag" and contains an empty text box and a "Search" button. The second section is labeled "Enter the user to be searched" and contains a text box with the value "sankalpnilekar" and a "Search" button. A small window titled "Result" is open, displaying the following information:

```
Name: sankalp nilekar
Location: Mumbai, India
Friends: 36
```

Chapter 9

CONCLUSIONS

9.1 Contributions

The main contribution of this project report is to suggest a new unique way to analyze the trends in online social networks. We have identified some of the features, which can help develop a model, which can be used to classify “trends” and “non-trends” in very early stages. We have also developed a highly scalable and efficient model to filter noise from tweets. Also, this prediction model is generic enough to be applied to any social media network, which has a connection among users. Thus, we can conclude that by constructing evolving graphs for different topics and observing several topological properties of the graph we can distinguish “trend” from “non-trends” topics.

9.2 Future Work

We have certainly established the benefits of constructing a social graph of twitter. So resolving more relations for the users in the graph can be useful and can improve the performance of the model. The model’s success depends on the topic of wise clustering of tweets. Currently, we have used simple clustering which not very “strict”, other better clustering algorithms can be used. Storing and processing graphs have been the real challenge and bottleneck of the whole pipeline. It’s necessary to improve this step by exploring better ways to do the same. Other data structures and algorithms can be explored to process the graph faster.

Chapter 10

REFERENCES

- [1] Luca Maria Aiello, Georgios Petkos, Carlos Martin, David Corney, Symeon Papadopoulos, Ryan Skraba, Ayse Göker, Ioannis Kompatsiaris, Senior Member “Sensing Trending Topics in Twitter” IEEE, and Alejandro Jaimes IEEE Transactions on Multimedia, Vol. 15, No. 6, October 2013.
- [2] Soyeon Caren Han, Hyunsuk Chung, Do Hyeong Kim, Sungyoung Lee, and Byeong Ho Kang “Twitter Trending Topics Meaning Disambiguation” Springer International Publishing Switzerland 2014.
- [3] Arkaitz Zubiaga, Damiano Spina, Raquel Mart´inez, V´ictor Fresno “Real-Time Classification of Twitter Trends” Journal of the American Society for Information Science and Technology copyright @ 2013.
- [4] Altawaier, M. M., & Tiun, S. (2016) “Comparison of Machine Learning Approaches on Arabic Twitter Sentiment Analysis” International Journal on Advanced Science, Engineering and Information Technology, 6(6), 1067-1073.
- [5] Rong Lu and Qing Yang, “Trend Analysis of News Topics on Twitter”, International Journal of Machine Learning and Computing Vol. 2, No. 3, June 2012
- [6] Kathy Lee, Diana Palsetia, Ramanathan Narayanan, Md. Mostofa Ali Patwary, Ankit Agrawal, Alok Choudhary, “Twitter Trending Topic Classification” 2011 11th IEEE International Conference on Data Mining
- [7] Erwin B. Setiawan, Dwi H. Widyantoro, Kridanto Surendro, “Feature Expansion using Word Embedding for Tweet Topic Classification” IEEE, 2016.
- [8] M. J. Zaki, A. Siebes, J. X. Yu, B. Goethals, G. I. Webb, and X. Wu, editors. 12th IEEE International Conference on Data Mining, ICDM 2012, Brussels, Belgium, December 10-13, 2012. IEEE Computer Society, 2012.
- [9] J. Yang and J. Leskovec. Patterns of temporal variation in online media. In Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11, pages 177{186, New York, NY, USA, 2011. ACM.

- [10] A. Vakali, M. Giatoglou, and S. Antaris. Social networking trends and dynamics detection via a cloud-based framework design. In Proceedings of the 21st international conference companion on World Wide Web, WWW '12 Companion, pages 1213{1220, New York, NY, USA, 2012. ACM.
- [11] Yubao Zhang, Student Member, IEEE, Xin Ruan, Student Member, IEEE, Haining Wang, Senior Member, IEEE, Hui Wang, and Su He “Twitter Trends Manipulation: A First Look Inside the Security of Twitter Trending” IEEE transactions on information forensics and security, vol. 12, no. 1, january 2017.
- [12] Amina Madani, Omar Boussaid, Djamel Eddine Zegour “Real-time trending topics detection and description from Twitter content” Springer-2015.
- [13] Arkaitz Zubiaga, Damiano Spina, Raquel Martinez, Victor Fresno, “Real-Time Classification of Twitter Trends” American Society for Information Science and Technology 2013.
- [14] Arkaitz Zubiaga, Damiano Spina, Víctor Fresno, Raquel Martínez “Classifying Trending Topics: A Typology of Conversation Triggers on Twitter” ACM 2011.
- [15] María del Pilar Salas-Zárate, José Medina-Moreira, Paul Javier Álvarez-Sagubay “Sentiment Analysis and Trend Detection in Twitter” Springer 2011.
- [16] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '09, pages 42{51, New York, NY, USA, 2009. ACM.
- [17] G. H. S. Milstein, A. Chowdhury. Twitter and the Micro-Messaging Revolution: Communication, Connections, and Immediacy {140 Characters at a Time. O'Reilly Radar Report, November 2008.
- [18] A. Ritter, Mausam, O. Etzioni, and S. Clark. Open domain event extraction from twitter. In KDD, 2012.
- [19] A. Hernandez-Suarez, G. Sanchez-Perez, V. MartinezHernandez, H. Perez-Meana,K. Toscano-Medina, M. Nakano and V. Sanchez “Predicting Political Mood Tendencies based on Twitter Data”
- [20] A. Ritter, S. Clark, Mausam, and O. Etzioni. Named entity recognition in tweets: An experimental study. In EMNLP, 2011.
- [21] Amina Madani,Omar Boussaid, Djamel Eddine Zegour “Real-time trending topics detection and description from Twitter Content” Springer 2015.

- [22] M. Mathioudakis and N. Koudas. Twittermonitor: trend detection over the twitter stream. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, SIGMOD '10, pages 1155{1158, New York, NY, USA, 2010. ACM.
- [23] L. M. L. Delcambre and G. Giuliano, editors. Proceedings of the 2005 National Conference on Digital Government Research, DG. O 2005, Atlanta, Georgia, USA, May 15-18, 2005, volume 89 of ACM International Conference Proceeding Series. Digital Government Research Center, 2005.
- [24] L. Barbosa and J. Feng. Robust sentiment detection on twitter from biased and noisy data. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10, pages 36{44, Stroudsburg, PA, USA, 2010. Association for CompY. Altshuler, W. Pan, and A. S. Pentland. Trends prediction using social delusion models. In Proceedings of the 5th international conference on Social Computing, Behavioral-Cultural Modeling and Prediction, SBP'12, pages 97{104, Berlin, Heidelberg, 2012. Springer-Verlag. Computational Linguistics.
- [25] Zgheib, W. A., & Barbar, A. M. A Study using Support Vector Machines to Classify the Sentiments of Tweets.
- [26] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(2009), 12.
- [27] Munir Ahmad, Shabib Aftab, Iftikhar Ali “Sentiment Analysis of Tweets using SVM” International Journal of Computer Applications November 2017.