International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 3 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT20637

Twitter Trend Analysis

Sankalp Nilekar, Sudeep Rawat, Rahul Verma, Pravin Rahate
Computer Engineering, Datta Meghe College of Engineering Airoli, Navi Mumbai, Maharashtra, India

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

Keywords : Social Trends, Hashtag, Twitter, Term Frequency-Inverse Document Frequency

I. INTRODUCTION

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.

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 expose 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 and the

CSEIT20637 | Accepted : 01 May 2020 | Published : 06 May 2020 | May-June-2020 [6 (3) : 94-103]

research community. 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. 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. 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 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. 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, 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. 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. 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.

II. LITERATURE REVIEW

Trend analysis and based on that predicting public opinions. It plays important role, many researcher working on automatic technique of extraction and analysis of huge amount of twitter data.

In 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. 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 three different information retrieval metrics precision, recall, and f-measure.

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. Author propose model which predict public opinion on political event by Appling different classifier which predict that whether mood is positive or negative. 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, a test study on the classifier was conducted which showed the result with the accuracy of 85%.

The authors 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 preprocessing stage of classification for high accuracy. Trend analysis performs using SVM in that two pre classified datasets of tweets are used then do comparative analysis, they use measures Precision, Recall and F- Measure.

Some researchers had an approach where posted tweets from the Twitter micro-blogging site are subjected to preprocessing and classified based on their emotional content as positive, negative and neutral or irrelevant; and compares the performance of various classifying algorithms based on their precision and recall in such cases. Further, the paper also discusses the applications of this research and its limitations. A number of machine learning like Naïve Bayes and Random Forest models performed sentiment analysis on product review data. Some work in this field included experiments with mood classification on blog posts. One of the researches also deals with review of aspect-based opinion polling from unlabeled free-form textual customer reviews without requiring customers to answer any questions. The tweet retrieval process needs access tokens from the twitter developer site and a piece of code which perform the operation of retrieving those tweets.

B. Software requirements:
1) Windows 7 or higher.
2) Python 3.0 or higher.

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

III. REQUIREMENT ANALYSIS

Requirements analysis involves all the tasks that are conducted to identify the needs of different stakeholders. Therefore, requirements analysis means to analyze, document, validate and manage software or system requirements. High-quality requirements are documented, actionable, measurable, testable, traceable, helps to identify business opportunities, and are defined to a facilitate system design. After the extensive analysis of the problems in the system, we are familiar with the requirement that the current system needs. These requirements are listed below:

A. Hardware requirements:

1) Processor – i3.

2) Hard Disk – 5 GB.

3) Memory – 1GB RAM.

4) Internet Connection.

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.

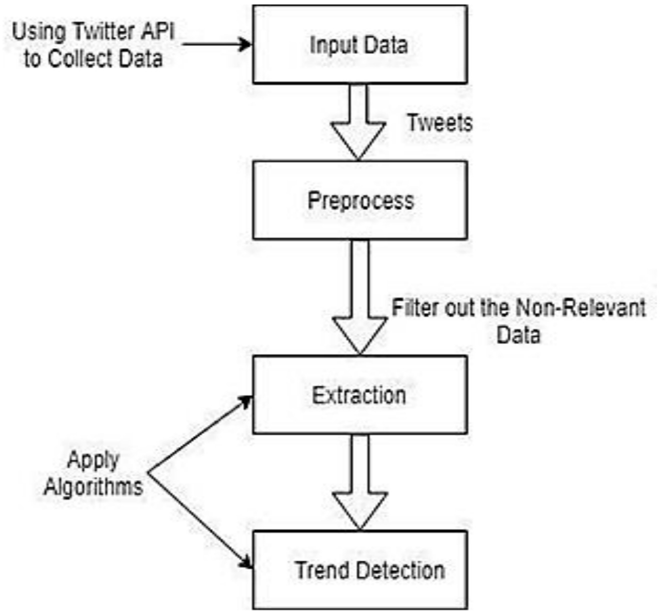
Video content performance:

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. information, economy, control and security efficiency and services.

IV. DESIGN PROCESS

The model has following steps:

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



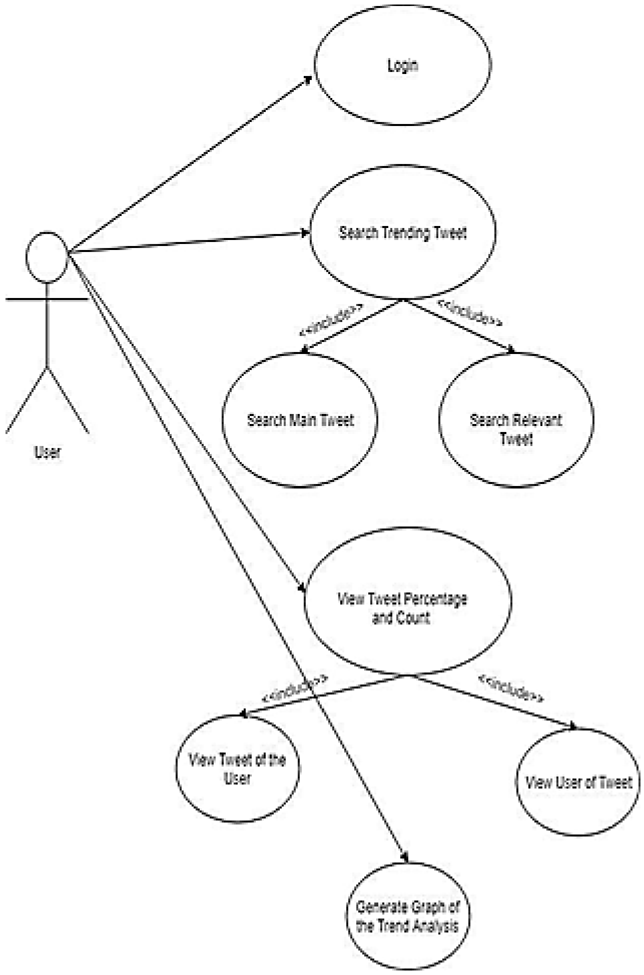
1) Dataset: Collect tweet data through twitter streaming API. Which download tweets in JSON format. We can apply keyword, hashtag, username to download tweets related to them.

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 ways 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

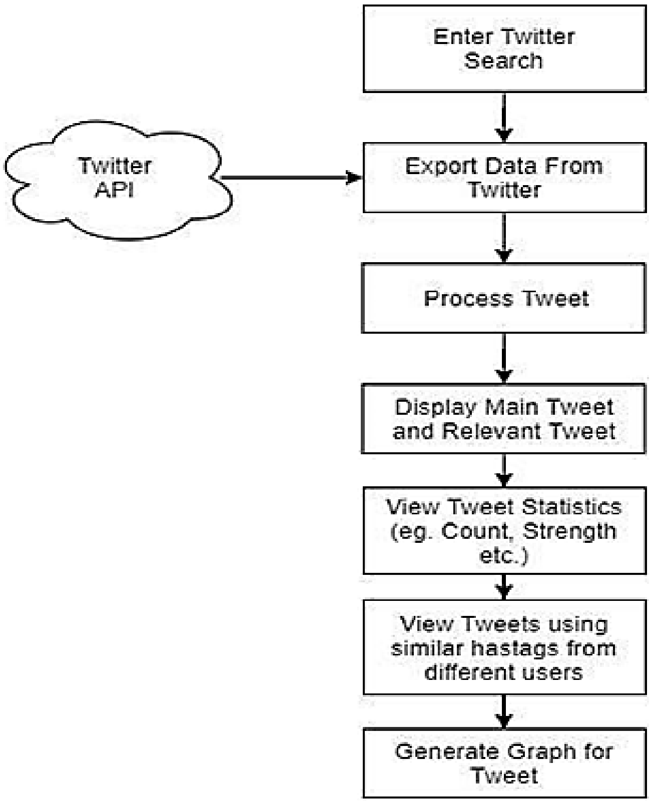
4) Trend Detection: We can determine trend by using TF- IDF calculation. And predict positive, negative, neutral mood tendency by applying machine learning algorithms. Show the graph for the given input by the user.

USE CASE DIAGRAM:



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


Enter Twitter Search
Twitter API
Export Data From Twitter
Process Tweet
Display Main Tweet and Relevant Tweet
View Tweet Statistics (eg. Count, Strength etc.)
View Tweets using similar hastags from different users
Generate Graph for Tweet

V. DESIGN AND IMPLEMENTATION

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.

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.

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.

VI. TECHNOLOGIES USED

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.

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.

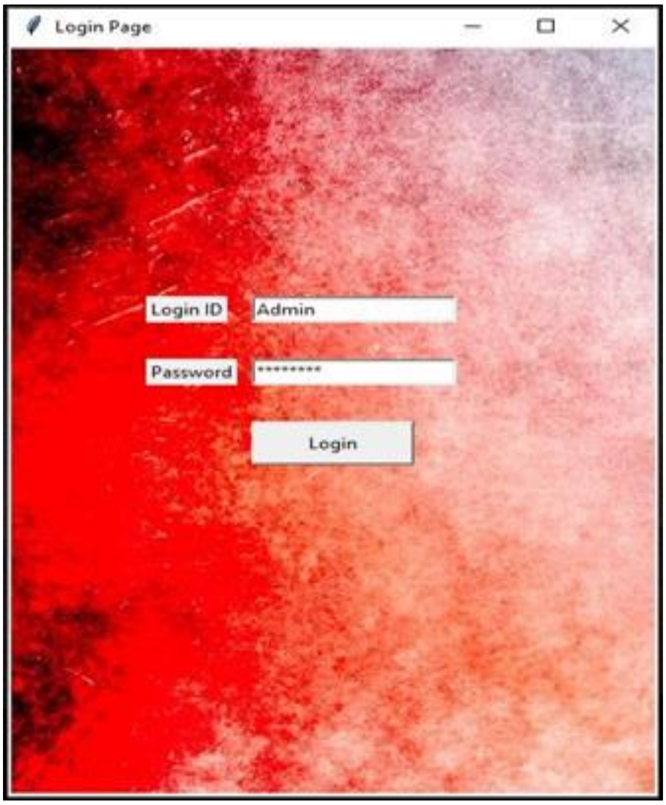
Today, with the rise of REST and web services over HTTP, the term is often assumed to refer to APIs of such services when given no other context (see the Web APIs section).

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.

Twitter API : When someone wants to access our 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 you before they can access your information. These permissions are not granted by default; you choose on a per-application basis whether to provide this access, and can control all the applications authorized on your account. The Twitter APIs include a wide range of endpoints.

VII. RESULTS

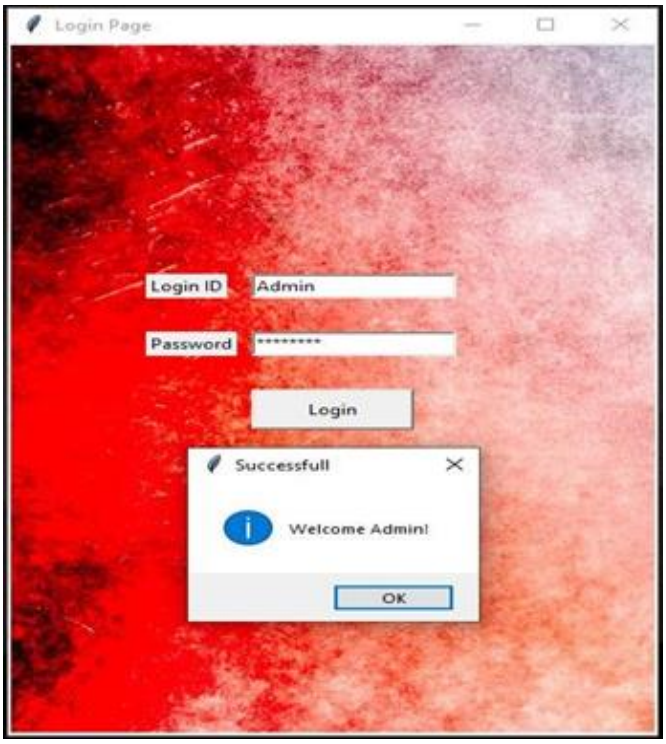
1. Login Page:



Login ID: Admin

Password: *********

2. Login Attempt Successful:



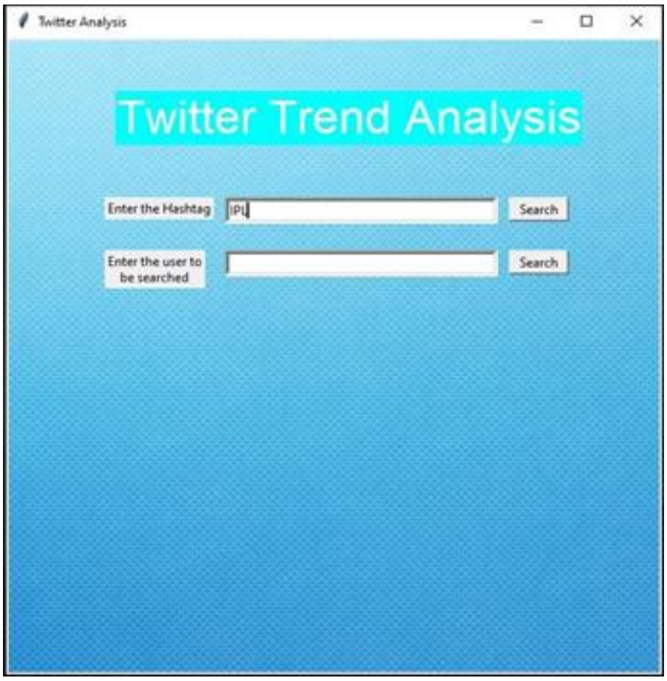
Login ID: Admin

Password: *********

Successfull

Welcome Admin!

3. Enter Hashtag Screen:

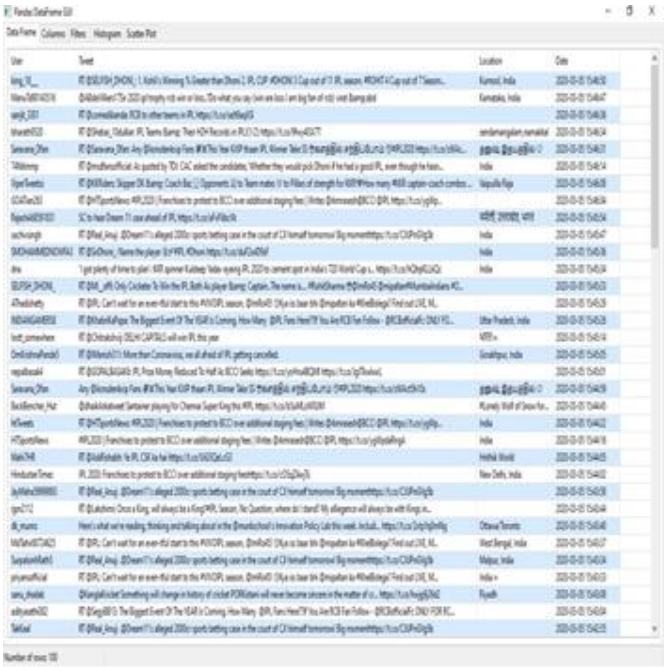


Twitter Trend Analysis

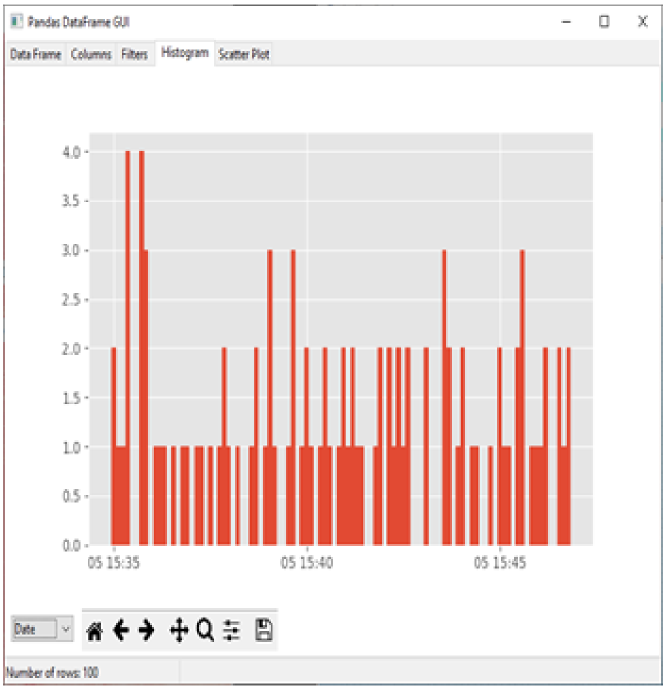
Enter the Hashtag: IPL

Enter the user to be searched:

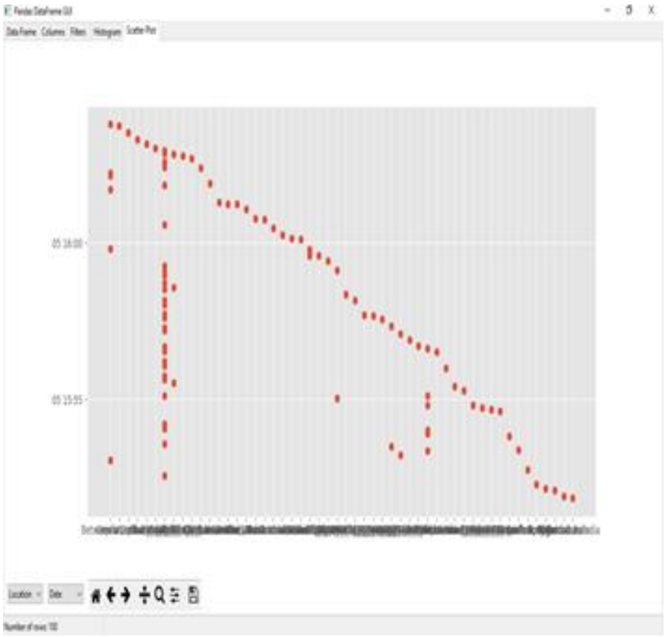
4. User, Tweet, Location, Date Output:



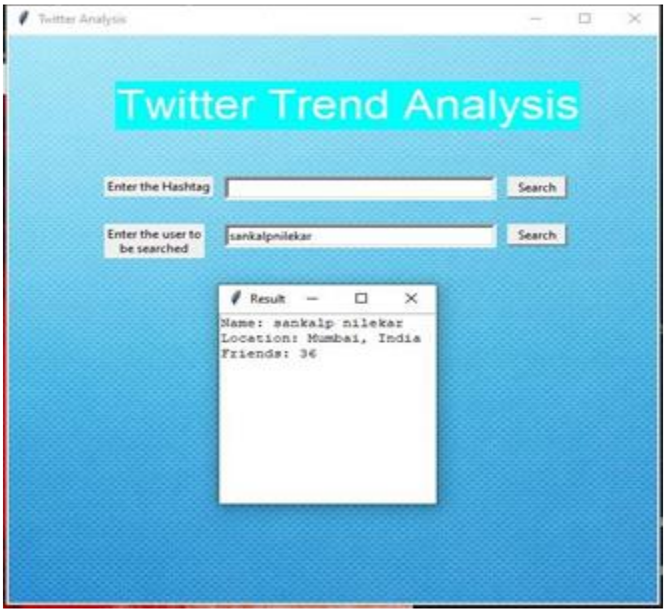
5. Histogram Output:



6. Scatterplot Output:



7. User Search Screen:



sankalpnilekar

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

VIII. CONCLUSION

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. 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. We have shown that how by constructing evolving graphs for different topics and observing several topological properties of the graph we can distinguish “trend” from “non-trends” topics.

IX. 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´ınez, V´ıctor 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]. 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

Cite this article as :

Sankalp Nilekar, Sudeep Rawat, Rahul Verma, Pravin Rahate , "Twitter Trend Analysis", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 3, pp. 94-103, May-June 2020. Available at doi : https://doi.org/10.32628/CSEIT20637
Journal URL : http://ijsrcseit.com/CSEIT20637