

Tweet Retrieval and Analysing the Trends

Utkarsh Malik, Harpreet Kaur, Aditi Chaudhary



Abstract: We can't disregard the importance of Social Media in Today's Technology Era. Internet is almost in every hand. People uses various Social Media platforms to express themselves and their thinking about various topics such as Politics, Entertainment, Sports, etc. In the Data Science industry, trend analysis can be used for several purposes like marketing or product analysis. Twitter data has been used to analyze political polarization and the spread of protest movements. Twitter is one of the most popular social media platform that allows the users to spread and share information. Twitter publishes the list of recent or latest topics named as "Trending Topics" which shows all the happenings in the world and what are the people's opinions about those topics. This Trend Analyzer will work on a given set of tweets and generates a graph based on the tweets and shows the comparative popularity of the used hashtags. This Analyzer will examine a set of tweets using Python and text-processing techniques.

Keywords: In The Data Science Industry, Trend Analysis Can Be Used For Several Purposes Like Marketing Or Product Analysis.

I. INTRODUCTION

Twitter has rapidly gained popularity among researchers as well as the students, policymakers, representatives, and the normal public. Twitter is the gold mine of data. Many users gets confused at some point to understand what exactly Twitter is, what it does and how they could use it. But still the tetchy nature of tweets shows that the Twitter is widely used by smartphone users who generally don't prefer reading long content stuffs.

Twitter allows its users to:

- Reachout a large number of audience quickly through tweets and retweets.
- Follow and observe the work of other experts who are in the same field.
- Build relationships with peoples around the world.
- Keep up-to-date with the latest news and happenings around the globe.
- Look for feedback and give feedback to others on their work.
- Put up with the discussions, events, conferences that are happening around the world that you can't attend in person.
- Express themselves and who they are as a person.
- Easily upgrades the research by providing links to your blog stories, journal articles, and news items.

Manuscript received on June 12, 2021.

Revised Manuscript received on June 22, 2021.

Manuscript published on July 30, 2021.

* Correspondence Author

Utkarsh Malik*, Department of IT & CSE, Meerut Institute of Engineering and Technology, Meerut (U.P), India.

Harpreet Kaur, Department of IT & CSE, Meerut Institute of Engineering and Technology, Meerut (U.P), India.

Aditi Chaudhary, Department of IT & CSE, Meerut Institute of Engineering and Technology, Meerut (U.P), India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

II. INFORMATION RETRIEVAL FROM TWITTER

How Tweets are Fetched?

For fetching the tweets from Twitter, Python has a very useful API "Tweepy" which helps in ifetching the information(tweets) from Twitter. Twitter allows the user to access its data using Twitter API or Tweepy. It also provides access to most of the functionalities of Twitter. It includes classes and methods. Tweepy is installed using **pip** which is a Python package manager. To authorize our app to access Twitter on our behalf, we need to use the OAuth Interface and create the required authenticon credentials i.e. consumer key, consumer secret, access key, and access secret. These keys will help the API for authentication.

Steps to obtain keys:

- 1) Login to the Twitter Developer Section.
- 2) Go to the "Create an App" option.
- 3) Fill in all the details of the application form such as App name, Application description, Use of the App and etc.
- 4) Click on "Create your Twitter Application" option.
- 5) Details of new app will be shown along with the Authentication credentials.

How Trend Determine?

Trends are determined by default based on your following, your interest, your geographical location. Trends and Hashtags are mostly grouped together as they are mostly related to a same topic. The topic that is popular now at the moment helps in discovering the latest emerging topic of discussion on Twitter and hence becomes a Trend.

III. METHODOLOGIES

A. DATASET

We will collect tweet data through Twitter streaming API's. We can use any keyword, hashtag or username as Input to download the tweets related to them.

B. FOR COLLECTING DATA

While collecting data, we are limited in two ways:
Cannot collect data from the past (a year ago, etc).

Twitter only provides a sample of its data for free (like 1% of its data). However, a 1% sample of data is in order of some million tweets a day. Many Social Media companies have APIs which are made available to third-party developers and researchers. Twitter has many API's available.



C. PREPROCESSING

Pre-processing of tweet data is done in several stages. Firstly, after downloading the tweets, we have to extract the text data and discard the unwanted media such as videos, audios, images, etc. Then, we store the English text which is retrieved from the tweet and remove the @, #, URL, and other punctuation from the stored data.

STEPS TO PREPROCESS DATA:

1. We have to create DataFrame using pandas.
2. We will be using .Cursor() to search through twitter for the required tweets.
3. We will iterate over each tweet in the list for extracting information about each tweet.

D. TREND DETECTION

We can determine the trend by analyzing the bar graph or the column graph or pie chart. And we can determine all the trending hashtags correlated to a particular topic.

E. BAR GRAPH

We will collect the latest tweets of specific topics and then filter them for multiple hashtags, after analyzing each tag we will generate the comparison graph for the tweets.

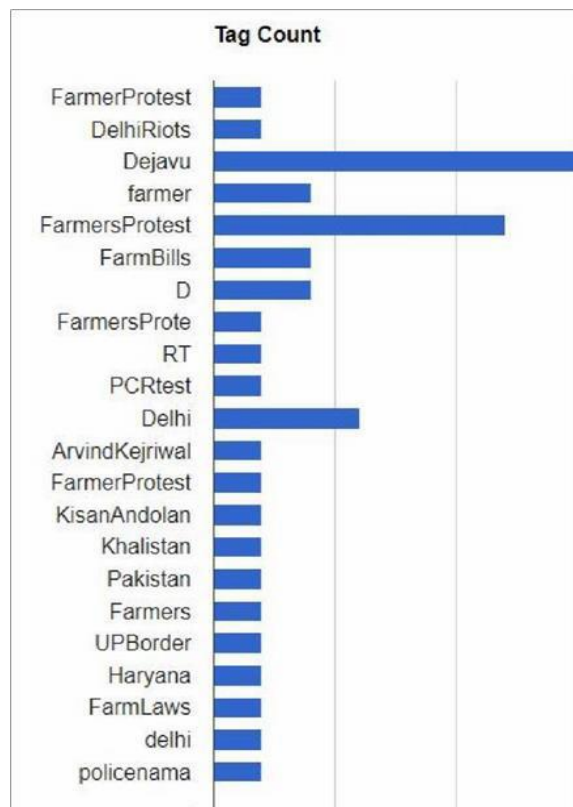


Figure 3.1

IV. RESULT

According to the aimed model, the final outcome of this paper leads to the development of a Twitter Trend Analyzer that retrieves the tweet and helps in determining the trend. Results will be generated in the graphical form. This Analyzer has three options for the user, so accordingly will have three different categories of result generation.

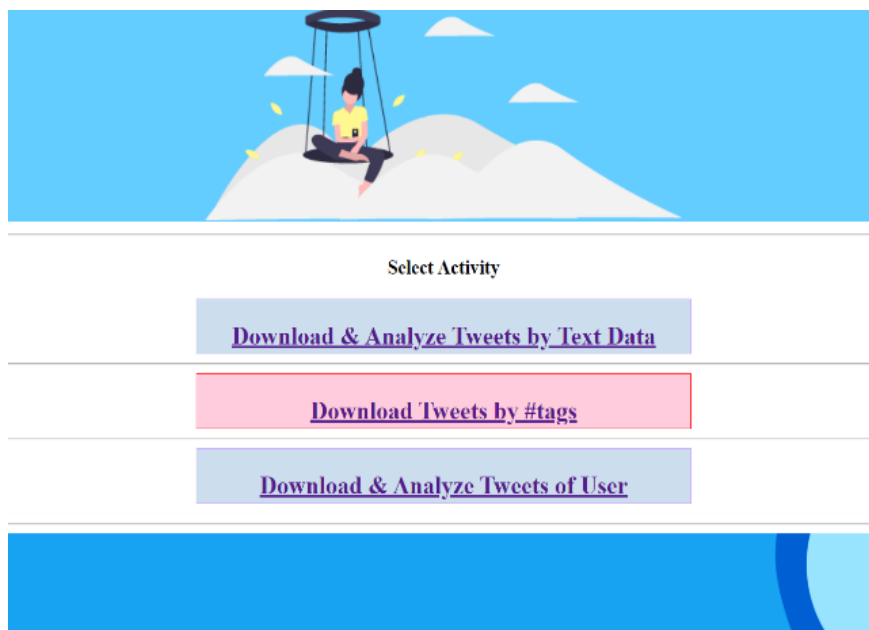


Figure 4.1



a) Download and Analyze Tweets by Text Data-

The user will input any topic according to their choice and the number of tweets they want to fetch. After clicking

on the search button, the results will be in the graphical form.



Figure 4.2

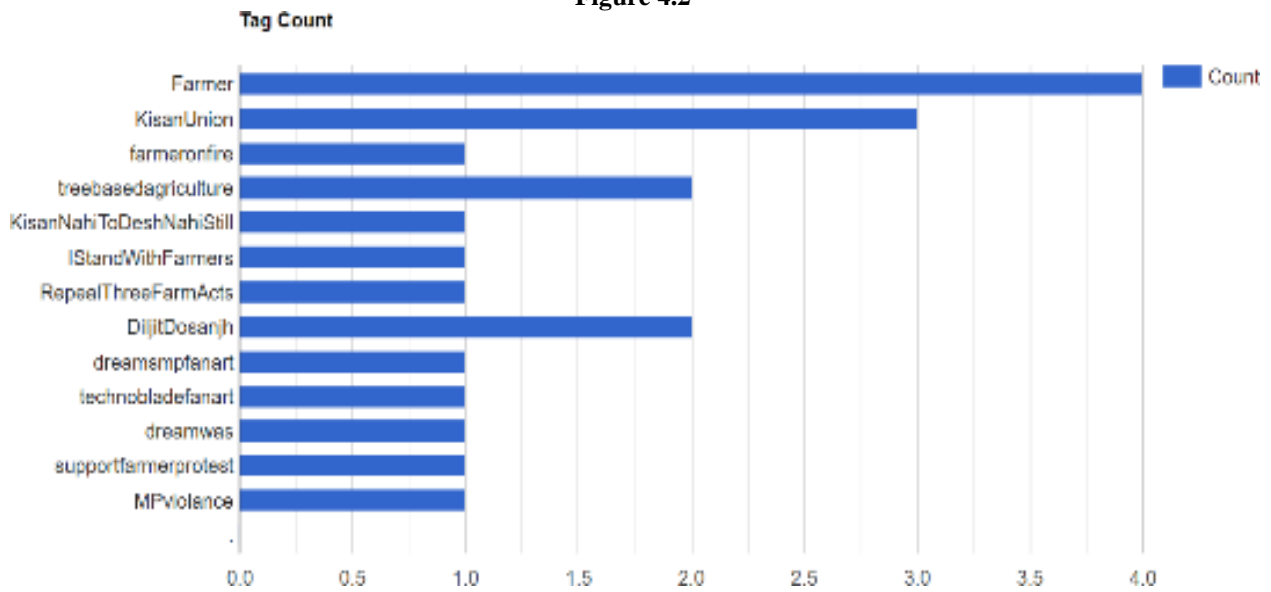


Figure 4.3

b) Download Tweets by Hashtags-

In this option, the user can search for more than one topic by separating them using commas between them. And then

the user will select the particular date from which he wants to fetch the data and again the count of tweets user wants. This option gives result in the form of text data.

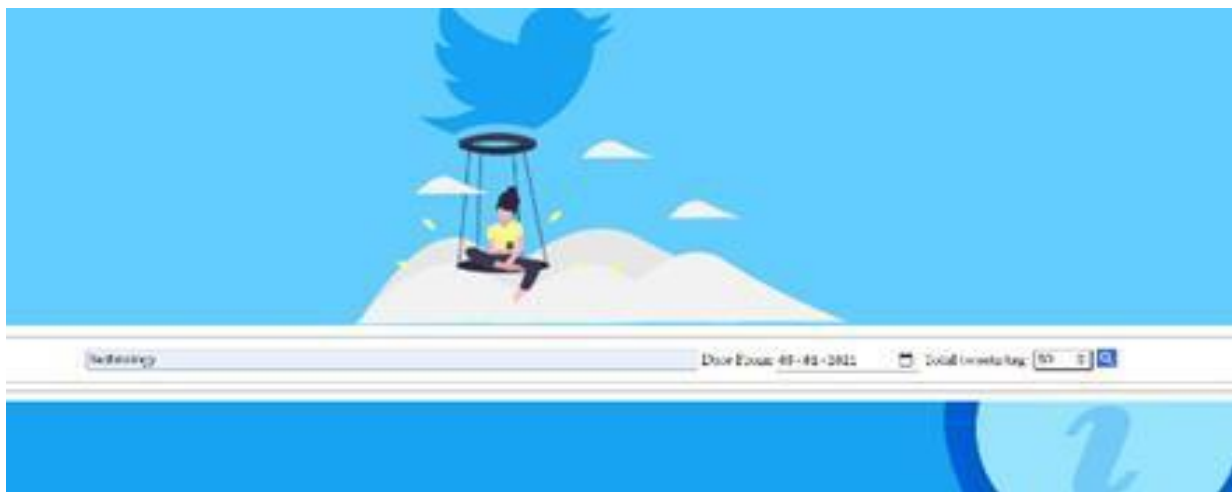


Figure 4.4

V. CHALLENGES AND BENEFITS

There are various challenges that need to be taken care of, while designing this Analyzer:

- **Ambiguity-** As Tweets consists of various phrases, keywords or hashtags that may create a confusion while retrieving and preprocessing.
- **Handle Real Time Data-** As Twitter cannot collect data from the past and only provides 1% of its data. Hence, retrieving the data and dealing with real time data is a challenge.

There are several benefits to the society:

- In Stock Market.
- In identifying the Breaking News.
- During the Elections.
- For Entertainment purposes.

VI. CONCLUSION

In this paper, we started with the very basics of Tweet Retrieval and then Determining the Trend. We demonstrated the steps to collect data, to preprocess the data and finally obtain the desired output. We outlined the results in the form of a graph and by analyzing the graph, we find out the most trending hashtags regarding that particular topic. There are several tools and techniques available for trend detection But there is no standard technique that has been established yet.

Hence, Comparative Analysis is required and needed to understand the efficiency. This paper is introductory in nature and hence deals with the basics.

FUTURE SCOPE

1. We can include images, audio, and videos as well for analyzing the hashtags present with them.
2. We can add sentimental analysis with the trend analyzer to enhance the features. So that the user can do both the tasks in the same project.
3. We can filter the downloaded tweet data according to our need.
4. We can apply translation on the downloaded data to retrieve the actual information of tweet i.e. in some other language.

REFERENCES

1. Sensing Trending Topics in Twitter, Luca Maria Aiello, Georgios Petkos, Carlos Martin, David Corney, Symeon Papadopoulos, Ryan Skraba, Ayse Göker, Ioannis Kompatsiaris, Senior Member, IEEE, and Alejandro Jaimes IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 6, OCTOBER 2013.
2. Arkaitz Zubiaga, Damiano Spina, Raquel Martinez, Victor Fresno, "Real- Time Classification of Twitter Trends" American Society for Information Science and Technology 2013.
3. Amina Madani, Omar Boussaid, Djamel Eddine Zegour "Real-time trending topics detection and description from Twitter content" Springer-2015.
4. Soyeon Caren Han, Hyunsuk Chung, Do Hyeong Kim, Publishing Switzerland 2014.
5. Ihler, Hutchins, Smyth, Adaptive Event Detection with Time Varying Poisson Processes. http://www.datalab.uci.edu/papers/event_detection_kdd06.pdf 2006.
6. S. Nikolov, Trend or No Trend: A Novel Nonparametric Method for Classifying Time Series. <http://dspace.mit.edu/bitstream/handle/1721.1/85399/870304955.pdf> 2011.
7. G. Salton and M. J. McGill, Introduction to Modern Information Retrieval. New York, NY, USA: McGraw-Hill.
8. Trend Analysis of News Topics on Twitter Rong Lu and Qing Yang,

AUTHORS PROFILE



Hi, This is **Aditi Chaudhary**, Born and brought up in Meerut (Uttar Pradesh). I am currently pursuing my B.tech (Information Technology) from MIET, Meerut. I did my Schooling from Meerut Only. It is my Final Year College Project that I've worked on with my 2 friends. I don't have any publications yet as this is our first work. I've did several Certification Courses during college that Includes Adobe Workshop, Introduction to ML and etc. I am a Animal Lover. I love pets and I also love Cooking. This is a very big achievement for us that our first research paper is going to be published.



Harpreet Kaur, I am pursuing my BTech final year in the stream of Information Technology from Meerut Institute of Engineering and Technology, Meerut. I have completed my Senior secondary from KK. Public School, Sardhana, 2017. I have completed my High school at St. Francis Convent school, Meerut, 2015. During my graduation, I have completed various

online courses-

1. Google Analytics, Google, April 2020
2. Artificial Intelligence, Accenture, May 2020
3. Artificial Intelligence, Accenture, May 2020
4. SQL Fundamentals, Solo Learn, May 2020
5. Introduction to Data Science in Python, Coursera, June 2020
6. HTML Basics. Solo Learn, July 2020

I have been working on a project named "Analyzing Twitter Trends" for my final year project, with my two team members.



Utkarsh Malik, I am currently pursuing B.Tech Final year in Computer Science & Engineering from MIET, Meerut. I have done various online courses from reputed universities in my graduation- Python Data Structures, Coursera, May 2020. Getting Started with Python, Coursera, May 2020. SQL Fundamentals, Solo Learn, April 2020. Google Analytics for Beginners, Google, April 2020.

I also developed my final year project called "Analyzing the twitter trend" along with my friends- Aditi & Harpreet

The main aim of our project is to build a twitter trend analyzer that will analyze a set of tweets using text-processing techniques. The trend analyzer will work on a given set of tweets and generates graphs based on the tweets and show the relative popularity of the used hash tags.

Along with my graduation I have participated in many cultural activities like -

Getting first Position in Rangotsav, College Fest Organizing blood donation camp in our college I also love to play outdoor games.