

# **Finding trending Hashtags and Time series analysis in Twitter**

## **Project Report**

**Team Members:**

KRISHNA GAPILESWAR V R – 2015103513

NAVEEN N S – 2015103561

## **Abstract:**

With advent of micro blogging sites, the data that is generated is humungous. Twitter is one of the major players of micro blogging sites. Every second, on average, around 6,000 tweets are tweeted on Twitter, which corresponds to over 350,000 tweets sent per minute, **500 million** tweets per day and around **200 billion** tweets per year. On average, around **6,000** tweets are tweeted on twitter every second. These tweets are treasure of data that can be analysed to find trending topics, mindset of the masses and even influences the most sought after political elections of the world. Our project is aimed at finding the hashtags in the tweets both in home timeline and in a specific user's timeline. The hashtags are obtained from tweets and analysed for their frequency and the time period in which the users tweet the most.

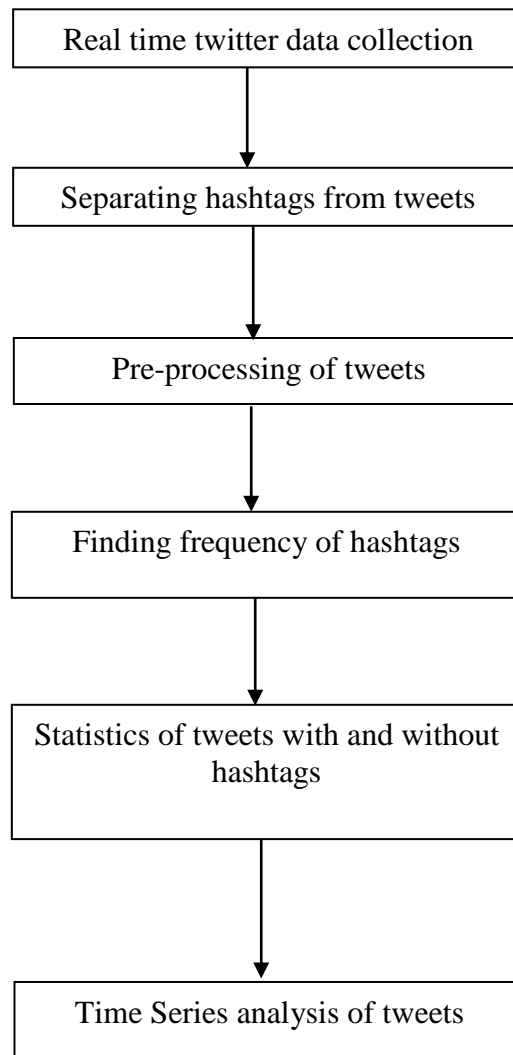
## **Existing methods:**

- Twitter Developer APIs.
- Python offers tweepy module for this purpose.
- In general, several websites like sprout social, hashtagify to find the trending hashtags.

## **Modules:**

- 1)twitterClient.py
- 2)twitterGetHomeTimeline.py
- 3)twitterGetUserTimeline.py
- 4)twitterStreaming.py
- 5)twitterHashtagFrequency.py
- 6)twitterHashtagsStats.py
- 7)twitterTimeSeries.py

## Block Diagram:



## Details:

Module Name	Responsibility
twitterClient.py	To set the twitter developer credentials
twitterGetHomeTimeline.py	To get the tweets from the home timeline
twitterGetUserTimeline.py	To get the tweets from a specific user timeline
twitterStreaming.py	To get the tweets associated with specific hashtag
twitterHashtagFrequency.py	To find the frequency of hashtags
twitterHashtagsStats.py	To find the total tweets, percentage of tweets with and without hashtags
twitterTimeSeries.py	To get a graphical view to analyse the time at which the tweets are posted

## Process flow:

With the help of twitter developer credentials access to the twitter api is obtained. Then a home timeline or particular user's timeline tweets are obtained in the json format or tweets embedded with specific hashtags can be obtained and stored in json format. Tweets are then pre-processed for case sensitivity and frequency of hashtags are obtained. Statistics about the tweets with and without hashtags is obtained. Finally the time series analysis is done and output is plotted in the form of graph.

Sample output:

