Analyzing the Popularity of A City (Hyd) Through Twitter Data

Vaddadi Vasudha Rani#[1] , Dr.K.Sandhya Rani[#2]

[1]*Vaddadi Vasudha Rani, Sr.Asst Professor, GMRIT, Rajam.*
[2]*Professor, CS Department ,SPMVV, Tirupati, INDIA*

Abstract

Anyone can have an opinion about anything. Nowadays we have many social networking sites to express our views on current topics or even about our personal life. For this micro-blogging is one of the option. Now Sentiment analysis can be easily defined as analysing and classifying a text as "a personal positive or negative feeling". When Twitter is considered any status, message created by a user can be called as a "tweet". These can be shared by other users among them which can be called as re-tweeting. This project aims to consider a dataset of tweets on a City [#Hyderabad] over a period of time and perform sentiment analysis on it. Then after analysing conclusions about which kind of tweets are being re-tweeted mostly can be made and this helps in predicting the popularity of the city.

Keywords **—** *City Popularity, tweets, sentiment analysis, re-tweet count.*

I. INTRODUCTION

Twitter is a social networking and micro-blogging service, enabling registered users to read and post short messages, so-called tweets. Twitter messages are limited to 140 characters and users are also able to upload photos or short videos. Tweets are posted to a publicly available profile or can be sent as direct messages to other users.

Twitter is one of the most popular social networks worldwide. Part of the appeal is the ability of users to follow any other user with a public profile, enabling users to interact with celebrities who regularly post on the social media site.

In the first Quarter of 2010 the number of monthly active users totaled to 30 million users. Now over 7 years this count multiplied 11 times. As of the fourth quarter of 2017, the micro-blogging service averaged at 330 million monthly active users.

There are many companies and media organizations which are increasingly seeking ways to mine Twitter for information about what people think and feel about their products and services. Most of the time they use sentiment analysis for this to find the people's perspective towards anything.

While there has been a fair amount of research on how sentiments are expressed in genres such as online reviews and news articles, how sentiments are expressed given the informal language and message-length constraints of micro-blogging has also been much studied.

It can be said that people tend to tweet about anything and sometimes everything that is part of their lives. Even though twitter contains data in the form of text, nowadays it also contains short video clips, audio clips, images, emojis etc. when we talk about performing sentiment analysis there is a large variety of data which can be analysed and selecting this domain is also very important. In our paper we use dataset that includes the name of the city, for which we would like to determine the popularity.

The existing sentiment analysis techniques are useful in various applications such as disaster relief and humanitarian assistance, marketing and trade predictions, checking political polls, advertising market, scientific surveys, checking customer loyalty, finding job opportunities, population health care and understanding students learning experiences.

In this paper we would like to go with the keyword which we have used in our project. That is, we have downloaded all the tweets mentioning the word "Hyderabad" and we get all the text from twitter that has this word as part of the tweet. This can be either directly or indirectly where the people use #Hyderabad.

II. METHODOLOGY

As we aim to perform sentiment analysis and visualize our results we need to follow some basic steps here. These are

- Creating a twitter application and authorizing through R.

- Downloading Tweets.

- Pre-Processing the downloaded tweets.

- Using a classifier on the Data to do Sentiment Analysis.

- Mapping the results and plotting graphs.

These five steps can be performed easily by anyone. Now in this section in detail discussion about the methodology used is done.

The first step requires having a twitter account as a prerequisite. Now if we need to get access to some tweets then we need to create an application or in simple terms a twitter app. Now we would be given an api_key and an api_secret.Then access_token and access_token_secret need to be generated. In RStudio store all these authentication values and after loading "twitteR" package, authorize using setup_twitter_oauth() and give your RStudio the chance to access twitter data.

After successful authorization the second step comes where we need to download the tweets that we require. This can be achieved by using the method searchtwitter() which takes our keyword as our input, a value n defining how many tweets needed to be downloaded and may other parameters. These can be further used to refine and filter the data which is important to us. This method returns a large list which we can convert to a data frame so that our work becomes easier. It is better to save this as a csv file so that even if the data is lost there is no need to perform the above steps every time.

Now the above two steps can be substituted by a simpler approach to download tweets which is by using an add-on called Twitter Archiver in Google Spread sheet. This approach is much easier but has its own limitations, so we won't be discussing here.

Thirdly we need to clean our downloaded data. This includes getting rid of URLs, Retweet Headers, references to other Screen Names, removing non-English characters and unnecessary spaces.
An example of this can be seen here. The text before pre-processing may be **"RT @CenInfoSecPPP: Addressing In Opening Ceremony Program of Bureau Office #DhartiTv at #Hyderabad. https://t.co/3eLeHq27IGðŸ' • Cc: @IamAmmâ€¦ ".** Now after pre-processing this may look like "Addressing In Opening Ceremony Program of Bureau Office DhartiTv at Hyderabad. "

Sample Data Download :

After Pre-Processing to have only Tweets:

The fourth step is a very critical one, and this is where we perform the actual Sentiment Analysis. We have used the sentiment() available from the sentimntr package.

- **Input:**
  Text M= {m$_1$, m$_2$, m$_3$......m$_n$}
- Dictionary of Polarized words:
  Jockers (2017) dictionary found in the lexicon package.
  - Total → 10,747 words
  - '+' Polarity → 3,586 words
  - '-' Polarity → 7,161 words
- **Output:**

  Positive messages p= {p$_1$, p$_2$...},
  Negative messages n= {n$_1$, n$_2$, n$_3$....},
  Neutral messages nu= {nu$_1$, nu$_2$, nu$_3$...}

  The working of the function would be

Step 1: Divide a Sentence into words $m_i = \{w_1, w_2, w_3.........\}$, i=1,2,....,n.

Step 2: if $w_i \in$ Dictionary of Polarized words

Return + ve polarity and - ve polarity.

Step 3: Calculate overall polarity of words=log(+ve polarity)-log(-ve polarity) by knowing orientation.

Step 4: Repeat step 2 until end of words.

Step 5: add the polarities of all words of the text, i.e. total polarity score of the sentence.

Step 6: Based on that polarity score, sentence can be positive, negative or neutral.

Step 7: repeat step 1 until M $\in$ NULL (end of text).

At last visualization is done where all our results will be shown using graphs and wordclouds. Here we aim to show a graph where the daily retweet count of each kind of sentiment for a period is plotted. A wordcloud can also be depicted which contains what words are mostly linked to the positive and negative sentiments.

A. *Result*

 1) *Module: 1*

The below plotted graph represents the relation between number of tweets in a particular day and tweets retweet count. The green color dots denotes the number of positive retweet count of a particular date, similarly the red color dots represents the number of negative retweet count and blue color dots as neutral count of retweets of a particular date in a month. By joining all the dots we got these lines which has plotted in the graph.
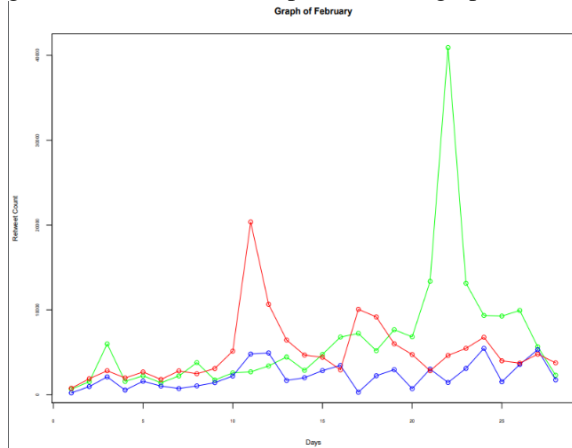
Fig 1: Graph Depicting How Different Types of Tweets are Being Retweeted Over the Month of February .

 2) *Module: 2*

The below word cloud has been generated by considering 100 positive tweets and 100 negative tweets. According to the polarity of the word, the size of the word has been displayed.
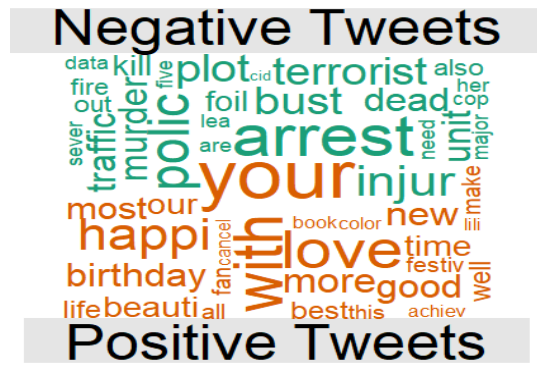
Fig 2. Word Cloud Sowing Most Frequent Positive and Negative Terms in our Data.

 3) *Module:3 – 3.1*

In this module we have considered some landmark places of the particular city which are mostly visited by the visitors and after performing sentiment analysis the place which has been repeated the most in the tweets ,the word cloud has generated. For example the word Mall has repeated most of the time in the tweets, followed by other places.

Fig3. Word Cloud of Top Tourist Landmarks in Hyderabad.

 4) *Module : 3.2*

By considering the above word cloud here we have plotted a bar graph shows that,which landmark place is mostly visited by the visitors. Mall is the place which is mostly visited by the visitors followed by park, masjid and charminar.
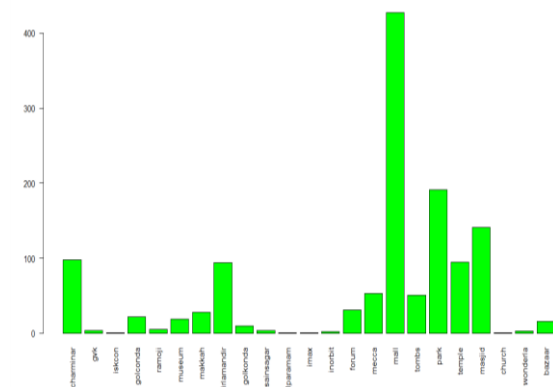
Fig 4.Graph of the landmarks and their mentioning in tweets.

III. RESULTS

A. Re-tweet Count:

	Positive	Neutral	Negative
count	179,666	63,151	141,064
percentage	46.80	16.45	36.74

Table 1: Re-Tweets Count from all three categories of Tweets Positive, Negative and Neutral

B. Top 5 Tourist Attractions and Contribution:

LOCATION	PERCENTAGE
MALLS	32.89 %
PARKS	14.71 %
TEMPLE	14.56 %
MASJID	10.86 %
CHARMINAR	7.55 %

Table 2: Top 5 Locations identified with contributions

IV. FUTURE SCOPE

Sentiment analysis has evolved very much from 2000's onwards. Currently there are many interesting domains in which this can be applied i.e. Opinion mining for reviews, ratings etc. This is mostly applied on health related issues and movies and celebrities tweets. Other Domains can be applicable with the same application are

1. An Item
2. A Social issue
3. A Trending Topic
4. A Celebrity

V. CONCLUSIONS

From this project we can conclude that for the city of Hyderabad mostly positive tweets were re-tweeted. So, most of the people responded towards positive messages. The type of people of this can be perceived as Positive. Now coming to the second part of our project, out of the 23 tourist locations that we have considered the most popular places were Malls and Parks followed by Temples and Masjids and then comes Charminar. This data can possibly help tourism package planners to dynamically create or modify packages for tourists and travellers to maximize their profits.

REFERENCES

[1] Salas-Zárate, M. D. P., Medina-Moreira, J., Lagos Ortiz, K., Luna-Aveiga, H., Rodríguez-García, M. Á., & Valencia-García, R. (2017). Sentiment analysis on tweets about diabetes: an aspect-level approach. Computational and mathematical methods in medicine, 2017.

[2] Vega-Oliveros, D. A., Berton, L., Vazquez, F., & Rodrigues, F. A. (2017). The Impact of Social Curiosity on Information Spreading on Networks. arXiv preprint arXiv:1706.07972.

[3] Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." Contemporary computing (IC3), 2014 seventh international conference on. IEEE, 2014.

[4] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., & Badhani, P. (2017). Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python. International Journal of Computer Applications, 165(9).

[5] Angiani, G., Ferrari, L., Fontanini, T., Fornacciari, P., Iotti, E., Magliani, F., & Manicardi, S. (2016). A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter. In KDWeb.

[6] Desai, Mitali, and Mayuri A. Mehta. "Techniques for sentiment analysis of Twitter data: A comprehensive survey." Computing, Communication and Automation (ICCCA), 2016 International Conference on. IEEE, 2016.

[7] Kouloumpis, Efthymios, Theresa Wilson, and Johanna D. Moore. "Twitter sentiment analysis: The good the bad and the omg!." Icwsm 11.538-541 (2011): 164.

[8] Zhang, S., Xu, K., Chen, X., & Liu, X. (2014). Characterizing Information Spreading in Online Social Networks. arXiv preprint arXiv:1404.5562.

[9] Batayneh, M. K., Marie, I., and Asi, I. (2008),"Promoting the use of crumb rubber concrete indeveloping countries." Journal of Waste Management,ELSEVIER, 28, 2171-2176.

[10] Egyptian Code Committee 203, (2003), "Experimentalguide for testing of concrete materials." Part 3 of theEgyptian code of practice for the design andconstruction of reinforced concrete structures.

[11] Eldin, N. N., and Senouci, A. B. (1993), "Rubber-Tyreparticles as concrete aggregate."Journal of Material inCivil Engineering, ASCE, 5(4), 478-496.

[12] https://www.rdocumentation.org/packages/twitteR/versions/1.1.9/topics/searchTwitter