

**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH
TECHNOLOGY****TWITTER ANALYTICS AND VISUALIZATION USING R****Neetu Anand *¹, Tapas Kumar ²**¹Maharaja Surajmal Institute, GGSIPU, New Delhi, India²Lingayas University, Faridabad, Haryana, India

DOI: 10.5281/zenodo.829749

ABSTRACT

It is an era of Internet technology various social networking sites are gain popularity worldwide. The data generated from these sites are growing day by day .Social media plays a very important role in most of the people's life. People communicate freely and share their comments, ideas on numerous events on any of the various social media like Twitter, Google+, Facebook, etc. Their views are useful for generating solutions and creating awareness about several problems .Twitter is the micro blogging site that has become very well-liked all over the world. These days, there is a continuing trend of posting every thought and feeling on any topic related to sports ,products, services, new announcements, politics, education etc. Due to this, analysis of these tweets has become a very important task so as to analyze the thoughts, opinions, feelings, sentiments, etc of various people. In this research paper, we have revealed all the steps required for the process of analysis of twitter data for opinion mining and also visualizing tweets on Google map.

KEYWORDS: Sentiment analysis, social media, opinion mining, Tweets, Data Analytics**I. INTRODUCTION**

Social networking sites provide an ease to communicate with any person living in any corner of the world and provides a real time environment for sharing thoughts, opinions, views and emotions with others which takes just a fraction of milliseconds or even less than that. The popular Social medias of nowadays are Facebook, Twitter, Google+ etc.

Twitter is the most widely used and it is a micro-blogging site which allows users to share, by using tweets which can be of maximum of 140 characters. Twitter not only allows sharing of textual data, but also enables us to share photos, videos or links to some websites or blogs. It is generating a huge amount of sentiment rich data in the form of tweets, status updates, blog posts, comments and reviews.[1] Data from social media sites are collected and analyzed by using mining techniques. Data science deals with big data, sparse data set, unstructured data set that require more diligence, data preparation and data conditioning. Big Data is mainly used in businesses or government organization to discover trends and patterns, which can support them in taking tactical decisions or spot a certain pattern or trend among the masses.

To inspect such a big volume of data, big data analytics is usually carry out using specific software tools for predictive analytics, data mining, text mining, forecasting, and data optimization. Some of the Big Data tools are like Apache Hadoop, Lumify, Apache Samoa, Rapid Miner, Mongo DB and R Programming Language for data analytics. In addition to data mining R provides statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others.

It is essential to analyze diverse types of data on Twitter. [2]Analysis helps to understand a person in a psychological manner means emotion of their minds, nature of the person etc. It gives a better view of positive and negative mindsets of people towards a particular thing, some rule, and any changes taking place or some product. Data analysis helps in analyzing various sentiments of a person, like if a person is happy, sad, angry, emotional, etc. A comment of a person could be judged as if it was said in a positive or negative manner, or if it was just neutral. Companies use twitter data Analysis for their products promotional activity, for services

provided by them to the customer by analyzing the feedback received. [3] They could also improvise the current one by considering the suggestions and negative feedbacks given by the customers. Customers will also be able to evaluate the product by analyzing the various responses of the other users towards it. That's why analyzing twitter data is seemly a very crucial task at the present time.

II. SENTIMENT ANALYSIS

Sentiment analysis is evaluating the people's decisions, attitudes, views, beliefs, emotions for individual, product, organization, topic, services, etc. Therefore, Sentiment Analysis is becoming a trendy these days. [4] Sentiment analysis relies heavily on pattern recognition and a basic understanding of key words. Sentiment analysis can be categorized as:

- i. Sentiment Opinion extraction
- ii. Opinion mining
- iii. Sentiment mining
- iv. Subjectivity analysis

Sentiment analysis is mainly used in marketing, political and sociological affairs. Sentiments classification can be a twin classification (positive or negative) or compound classification (extremely negative, negative, neutral, positive or extremely positive). Happiness, pleasure, cheerfulness, ecstasy, delight represents positive sentiments and irritate, anger, rage, depression, anxiety, grief represents negative sentiments. Various tools for detecting the feelings polarity are Emoticons, LIWC, SentiStrength, Senti WordNet, SenticNet, Happiness Index, AFINN, PANAS-t, Sentiment140, NRC, EWGA and FRN. The sentiment analysis can be performed at any level in the text viz. Word level, Sentence Level, Feature level, Document level SA and are used in various types of applications. [5] Some of the applications of sentiment analysis are listed below:

- i. Opinion finding: Businesses and Organizations which desire consumer opinions for their products they manufacture and service they provide.
- ii. Views: Individuals who make judgment to buy products or services based upon word of mouth or on-line reviews, or using public opinion example cases concerning politics or local issues
- iii. Online advertising: On-line promotional activities where in social media or an organization may place an advertisement in response to a positive review of a product.
- iv. Analyzing Feedback: To can get an idea of how happy customers are with the products, services from the ratio of positive to negative tweets about them and their experiences can be used for spotting dissatisfaction or the problems and it may also be used in positive manner to promote the upcoming deals .
- v. Prediction of Prices of commodities and shares evolution: To foretell market movements based on news bulletin, blogs and reaction of people on social media
- vi. Politics: what do people think about this candidate or issue? .It is also used for politic Voting
- vii. Public sentiments: It may be for any movie, product, legal matter, Policy or government-regulation proposals, Real-world events and many more.

The sentiment expressed may be of any states like:

- i. Attitudes: A tendency to respond positively or negatively towards a certain idea, object, person or situation.
E.g. Liking, loving, hating, valuing, desiring
- ii. Emotion: Conscious experience, expressions which can be seen by others.
E.g. Surprise, sad, anticipation, fear, proud, disgust
- iii. Mood: a temporary state of mind or feeling.
E.g. Excited, Giddy, irritable, restless, sad, Stressed, Weird, buoyant, Optimistic, Peaceful
- iv. Personality traits: characteristics that are the embodiment of an individual's or habitual patterns of behavior
E.g. Confident, anxious, Observant, Bossy, hostile, jealous

The sentiment classification approaches can be classified as [6]:

- i. Machine learning: It is used for predicting the polarity of sentiments based on trained as well as test data sets.
- ii. Lexicon based: It uses a predefined list of words, where each word is associated with a specific sentiment.



- iii. Hybrid: It has the potential to improve the sentiment classification performance and is the combination of both the machine learning and the lexicon based approaches.

III. LITERATURE SURVEY

A lot of researches have been done by various researchers in the field of sentiment analysis of twitter data. Barbosa et al. [7] designed a two phase automatic sentiment analysis method for classifying tweets. In the first phase they classified tweets as objective or subjective and then in second phase, the subjective tweets were further classified as positive or negative. Luo et al. [8] highlighted the challenges and efficient techniques to mine opinions from Twitter tweets. In [9] a survey on several techniques of sentiment analysis of product reviews for example machine learning, semantic orientation, opinion polling, holistic lexicon-based approach and many others is carried out. The survey highlighted that sentiment analysis/opinion mining play very important role to make decision about product /services. In another survey on approaches used for sentiment analysis [10] three approaches for performing sentiment extraction are described: subjective lexicon approach: is a list of words is assigned a score that indicates its nature in terms of positive, negative or objective; n-gram modeling approach: that can use uni-gram, bi-gram, tri-gram or combination of these for the sentiments classification; machine learning approach: performs the semi and/or supervised learning through the extraction of the features from the text and learn the model. [11] Bollen, Mao and Pepe described how text-mining can be used to extract mood and sentiment information from tweets over a period of time.

IV. DATA ANALYSIS TOOL R FOR TWITTER MINING

The R software provides the ability to download Twitter related data, and to be able to plot and analyze it. R is not the only analytics tool that can be used to analyze Twitter usage, patterns and networks but also it is free, open source, and has a well-supported user community. [12] R analysis packages are flexible and integrate easily with one another. This makes it a good choice for doing analytics. Some of the important packages of R include

- *twitterR*: Provides an interface to Twitter API.
- *ROAuth*: Allows users to authenticate to the server
- *plyr*: Manage a big problem by dividing it into manageable pieces, work on it and to put them together.
- *stringr*: *stringr* is responsible for all string functions in R environment. It is a simpler and easier to use package.
- *data.table* - An alternative way to organize data sets for very, very fast operations. Useful for big data.
- *maps* - Easy to use map polygons for plots.
- *ggplot2*: Used to implement graphics in R. It supports multiple data sources.
- *ggmap* - Download street maps straight from Google maps and use them as a background in your *ggplots*.
- *RColorBrewer*: Used for drawing maps shared according to a variable through palettes.
- *Devtools*: An essential suite of tools for turning any code into an R package.
- *caret* - Tools for training regression and classification models

V. EXPERIMENTS AND RESULTS

The method followed to predict sentiments and visualize tweets on Google Maps using R consisted of following steps:

- *Creating a Twitter Account and create a new app*: Go to Twitter Developer link and sign in with your credentials, after that go to twitter Apps and click on “Create New App” button for creating new application.
- *Extracting tweets*: Extract tweets of Budget 2017 using *twitterR* Package. *twitterR* package has search Twitter function that can search tweets based on a supplied search string.

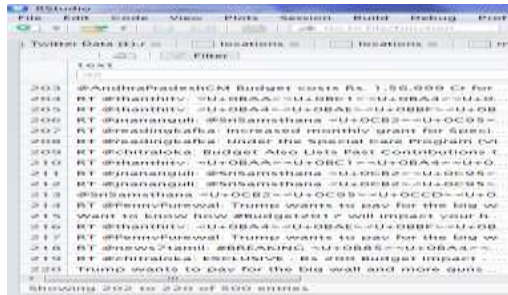


Fig1: Extracting 500 tweets having tag “Budget 2017”.

- *Sentiment analysis of twitter data:* It involves removing all the stop words and plots the results by considering various types of sentiments.



Fig2: Result of sentiment Analysis

- *Extracting Maps:* ggmap package has a function getmap that can download maps from Google Maps API.



Fig3: India Map extracted from Google API

- *Filtering Data:* Excluding rows from data frame where longitude/latitude=NA & taking only last two columns (unique data). Showing the final output of 251 Observation & 2 Variable.



id	lat	lon
1	85.832714	8.004300
2	75.832714	8.004300
3	86.832714	8.004300
4	76.832714	8.004300
5	74.832714	8.004300
6	73.832714	8.004300
7	72.832714	8.004300
8	71.832714	8.004300
9	70.832714	8.004300
10	69.832714	8.004300
11	68.832714	8.004300
12	67.832714	8.004300
13	66.832714	8.004300
14	65.832714	8.004300
15	64.832714	8.004300
16	63.832714	8.004300
17	62.832714	8.004300
18	61.832714	8.004300
19	60.832714	8.004300
20	59.832714	8.004300
21	58.832714	8.004300
22	57.832714	8.004300
23	56.832714	8.004300
24	55.832714	8.004300
25	54.832714	8.004300
26	53.832714	8.004300
27	52.832714	8.004300
28	51.832714	8.004300
29	50.832714	8.004300
30	49.832714	8.004300
31	48.832714	8.004300
32	47.832714	8.004300
33	46.832714	8.004300
34	45.832714	8.004300
35	44.832714	8.004300
36	43.832714	8.004300
37	42.832714	8.004300
38	41.832714	8.004300
39	40.832714	8.004300
40	39.832714	8.004300
41	38.832714	8.004300
42	37.832714	8.004300
43	36.832714	8.004300
44	35.832714	8.004300
45	34.832714	8.004300
46	33.832714	8.004300
47	32.832714	8.004300
48	31.832714	8.004300
49	30.832714	8.004300
50	29.832714	8.004300
51	28.832714	8.004300
52	27.832714	8.004300
53	26.832714	8.004300
54	25.832714	8.004300
55	24.832714	8.004300
56	23.832714	8.004300
57	22.832714	8.004300
58	21.832714	8.004300
59	20.832714	8.004300
60	19.832714	8.004300
61	18.832714	8.004300
62	17.832714	8.004300
63	16.832714	8.004300
64	15.832714	8.004300
65	14.832714	8.004300
66	13.832714	8.004300
67	12.832714	8.004300
68	11.832714	8.004300
69	10.832714	8.004300
70	9.832714	8.004300
71	8.832714	8.004300
72	7.832714	8.004300
73	6.832714	8.004300
74	5.832714	8.004300
75	4.832714	8.004300
76	3.832714	8.004300
77	2.832714	8.004300
78	1.832714	8.004300
79	0.832714	8.004300
80	0.832714	8.004300

Fig4: Showing the filtered data

- *Plotting tweets on Google map:* Converting both columns of location as numeric. Mapping Map with Locations.



Fig5: Plotted tweets on Google Maps

VI. CONCLUSION

The processed data is analyzed and the results have been produced for various types of emotions of the collected tweets related to "Budget2017". The location of the collected tweets are analyzed and plotted on Google maps. The location information is mainly used to gain insight into the prominent locations discussing an event. Maps are an obvious choice to visualize location information and aid in analysis of tweets. Each tweet is identified by a dot on the map. Through this we can easily identify the regions of interest and regions of high density of twitter user. This information can be used for targeting the user for future communication depending upon the interest of people.

VII. ACKNOWLEDGEMENTS

I am thankful to my PhD thesis supervisor, Professor Tapas Kumar for guiding me in preparing this paper.

VII. REFERENCES

- [1] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldrige, "Twitter polarity classification with label propagation over lexical links and the follower graph", In Proceedings of the First Workshop on Unsupervised Learning in NLP, 2011.
- [2] Alec Go, R. Bhayani, L. Huang, "Twitter Sentiment Classification Using Distant Supervision", Stanford University, Technical Paper, 2009.
- [3] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis", Computational Linguistics, Volume 37, Number 2, pp. 267-307, 2011.
- [4] Bifet and E. Frank, "Sentiment Knowledge Discovery in Twitter Streaming Data", In Proceedings of the 13th International Conference on Discovery Science, Berlin, Germany: Springer, pp. 1-15, 2010,.



- [5] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: Sentiment classification using machine learning techniques", In Proceedings of the Conference on Empirical Methods in Natural Language Processing - Volume 10, EMNLP '02, pp. 79–86, Philadelphia, Pennsylvania, USA, 2002.
- [6] Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social Media Data", IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3 - 6, pp 91-96, 2013.
- [7] L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data", COLING 2010: Poster Volume, pp. 36-44, Beijing, August 2010.
- [8] Zhunchen Luo, Miles Osborne, Ting Wang, "An effective approach to tweets opinion retrieval", Springer Journal on World Wide Web, pp. 545-566, Volume 18, Issue 3, May 2015.
- [9] A. N. Jebaseeli, E. Kirubakaran, "A survey on sentiment analysis of (product) reviews", International Journal of Computer Applications", Volume 47– No.11, pp. 36-39, 2012
- [10] A. Kaur, V. Gupta, "A survey on sentiment analysis and opinion mining techniques", Journal of Emerging Technologies in Web Intelligence, 5(4), pp. 367- 371, 2013.
- [11] J. Bollen, H. Mao, A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 450-453, 2009.
- [12] Albert Bifet and Eibe Frank, "Sentiment knowledge discovery in twitter streaming data", In Discovery Science, pp.1–15, 2010.

CITE AN ARTICLE

Anand, Neetu, and Tapas Kumar. "TWITTER ANALYTICS AND VISUALIZATION USING R." *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY* 6.7 (2017): 496-501. Web. 15 July 2017.