

Social Network Analysis Using Twitter Data

Rishabh Mishra¹, Pradyumn Katre², Ritesh Srivastava³

^{1,2,3} Department of Computer Science & Engineering
Galgotias College of Engineering and Technology
Greater Noida, India

Abstract - Social media is a well-known platform for user to create, share and check the new information which is being updated on a daily basis. The world has become a global village because of the utilization of the internet and social media like Facebook, Twitter and various other social media sites. These platforms have become a wide source of growth and spreading of daily opinions related to feelings, thoughts, and moods, and also what is going on in today's world. Sentiment-Analysis is a combination of two words Sentiment and Analysis. Any opinion of an individual through which the feelings, attitude, and thoughts can be expressed is known as sentiment. The process of recognizing and classifying general public opinions, about any topic expressed in any piece of text, or in social media especially in order to determine whether the writer's attitude towards a particular topic, product, is negative, positive or neutral is known as analysis. Twitter, being the most popular micro-blogging social site is used to collect data to perform analysis. As Twitter's special and unique infrastructure and almost, total availability of its data, ensured its high popularity among researchers. Twitter is the most popular platform, when it comes to social media research both in academics and in industry; no other social media website has attracted as much attention from academics or industry. Twitter has an open infrastructure which allows any user to follow another user, and it provides almost 100% of its data through its API (Application Program Interface). This survey paper briefs about studies and approaches which are being used in on-going project of developing fact extraction technology on social media analysis of tweeter data. This survey paper brief about recent studies or approaches which have been used in social media analysis using the twitter data.

Key Words: — sentiment analysis, machine learning, twitter, tweets, analysis abstraction.

1.INTRODUCTION

Social media is a topic which is very much relevant in today's time, and they called social media has touched upon almost every aspect of life whether it is personal or professional or whether it is political, so for that matter you are connecting with your friends on Facebook, Twitter etc. or if you are going shopping on amazon or other shopping sites or if you are writing an article on a political issue you are all the time connected with social media, and we are getting to know views or information from all these mediums [7]. Now

before we get into the details of social media, we need to understand that what exactly is social media. Media are instrument of commission but when we say social media it becomes a social instrument of commission [10]. These media have evolved from a broadcast model to a network model. When we say broadcast model there is one person or a group of people who are creating a message, who is decelerating yokes to the rest of the people which means it is from one to many kinds of model, but today we have moved from many to many kinds of models which means from broadcast to network model. Network model is a model where many people are creating several messages all around the world and giving, sharing and lighting it on social media. We need to keep in mind that as media professionals or as communicators whether they are in the field of management by the themes of media or for that matter in the field of education. We need to effectively use the sports in a very good manner and whether we like it or not. This social media is going to grow as people are not willing to learn, as people are willing to share, as people are willing to comment. This social media is going to grow and it is going to grow multiple.

Twitter, one of the most popular social media websites went live on 21st March, 2006 [2]. It is a microblogging website and its micro blogs are commonly known as tweets. First tweet that was tweeted on twitter was "just setting up my twitter." Which was tweeted by its founder Jack Dorsey [3].

Originally tweets were restricted to 140 characters but it's word limit was increased to 280 except for Chinese, Japanese and Korean [4]. In September 2019, Twitter was ranked the 32nd most visited websites in the world by Alexa's web traffic analysis [5]. A user can send its tweet to his or her followers and his followers can in turn like or retweet the tweet. All the major celebrities and famous personalities around the world have their presence on Twitter and they interact with their fans on Twitter [11]. Twitter is a major trend setter in the world as what trends on Twitter trends in the world. People can follow up or give their opinion about the topic using the hashtags (#) followed by the topic [1]. During the 2016 U.S. presidential election, Twitter was the largest source of breaking news on the day, with 40 million election-related tweets sent by 10:00 p.m. (Eastern Time) that day [6].

The most interesting and popular analysis based over the past years is the Sentiment analysis [8], [14], [16], [18]. The opinions that are held by any number of individuals are reviewed and analyzed using sentiment analysis [12], [19], [21], [38]. These reviews can be related to an event, brand, person or product. Earlier, magazines, newspapers and other

sources were used to express people's views. However, with the advancement in technology the people have begun to express their feelings on different social networking and microblogging sites. In a productive manner, the opinions of individuals have been extracted, studied and then evaluated by researchers. Twitter has gained the highest popularity in comparison to all other microblogging platforms in the past few years. It can be considered as a valid indicator for the sentiments of people. Different ways have been developed by several media organizations to mine the twitter information conduct training, testing and analysis, the tweets are collected using API [9]. On any imaginable topic the messages are posted by Twitter users. This is different from other microblogging sites in which only particular topic and purpose is discussed. Real time: Since the blogs are longer and a huge amount of time needs to be invested, these blogs are updated at longer time intervals.

2.LITERATURE SURVEY

Sentiment Analysis of twitter data is usually considered as much more complicated than the analysis of conventional texts such as books or review documents. This is due to the fact of short length of tweets, regular use of informal and irregular words, and the rapid evaluation of language that is used in twitter. A large amount of work has been conducted in Twitter sentiment using various techniques and methods which gave us some useful findings and discoveries.

Lexicon - based technique, N-gram Technique, Maximum entropy technique to discuss the importance. This survey discusses the importance and effects of sentiment 191 analysis challenges in sentiment evaluation based on two com- 192 prisons among forty- seven papers [13].

A research was conducted by using Twitter Classification Algorithm Called SentiDiff, which concluded that mining sentiment polarities expressed in Twitter messages is a meaningful while challenging task. Most of the existing solutions to twitter sentiment analysis only consider textual information of Twitter messages, and cannot achieve satisfactory performance due to unique characteristics of Twitter messages [15].

While detection and classification of social media-based extremist affiliations using sentiment analysis using Baseline method it was found that the proposed method outperformed the comparing methods in terms of better precision, recall, f-measure and accuracy [34].

A study conducted by machine learning approach for Transportation sentiment analysis using word embedding and ontology- based topic modeling yielded This offers a text classification system that identifies the most relevant transportation texts in social media and analyzes them to examine traffic control management and transportation services. Indeed, our proposed system efficiently classifies extremely ambiguous text and determines transportation polarity. This new approach not only improves the performance of LDA but also outperforms topic2vec document representation methods with transportation

datasets. It integrates lexicons into a pre-trained word embedding model that increases the accuracy of sentiment classification. It can also represent each word in the corpus with a low- dimensional vector and with semantic meaning. Furthermore, this approach can be linked to different information extraction, text mining, and sentiment analysis systems, since it can generate topics, features, and opinions from unclear text, and represents these extracted data accurately in order to improve the performance of sentiment classification. This approach can be helpful for medical application [17].

Sentiment analysis techniques were studied for the analysis of twitter data using N-gram algorithm, SVM algorithm and Sentiment Analysis Techniques and performed the sentiment analysis of twitter data in four steps namely collection of data, preprocessing the data, applying N- Gram algorithm and then SVM algorithm for classification [35].

Using Deep learning, Twitter sampling and lexicon Pam clustering, a research was conducted for a private firm with the name of 'HALAL', which the objective of knowing the people view about the company, the study showed and proved that people's sentiment towards and halal tourism and halal cosmetics is positive. A model with 93.78% of accuracy was used for analysis of data obtained from twitter [36].

Detection of signal in social media using tweets volume, hashtags with the help of a combination of threshold algorithm and gaussian algorithm to allow the firms and brands to monitor their position and reputation from detected signals in tweets' volume, hashtags and sentiments. While most of the researches used static threshold technique to detect spikes but, in this study, automatic moving threshold algorithm, Gaussian algorithm and hybrid algorithm (Combination of threshold algorithm and Gaussian algorithm) was used for detecting the spikes in data. This study showed that Gaussian and hybrid algorithm performed better than static threshold technique [20].

Twitter Sentiment in New York City Parks as measure of well beings in 2019 by using Data collection, Data Processing, Natural language processing methods and find It establishes the usefulness of Twitter based analytics in comparing user sentiment internal and external to urban parks, with application to New York City [33].

Clause-Newman-Moore algorithm was used to check if a person checks or looks up for health information online and mainly about HPV, it was found that 80% of the people are checking out health related information and tweets online [37].

Machine Learning based sentiment analysis for twitter accounts using Naïve Bayes and support vector machines (SVM) concluded that Text Blob and Wordnet use word sense disambiguation with greater accuracies and can be used further in predicting elections [22].

Hadoop Technology was used for sentiment analysis of twitter data using big data tools and found that that Twitter being the largest social networking site generates data of very

large amount because of millions of tweets and followers which are increasing per day so according to this whole processing and tweets are analyzed [23].

A text mining approach to the Syrian Refugee crises via sentiment analysis on twitter was done to conclude that Turkish tweets carried slightly more positive sentiments towards Syrians and refugees than neutral and negative sentiments, nevertheless the sentiments of tweets were almost evenly distributed among the three major categories [24].

Emotions Analysis of twitter data using hashtag emotions was carried out by using emotion analysis techniques by extracting emotion lexicons and discovered that afraid is the least expressed and sadness is the most expressed emotion on twitter [25].

Advance Machine Learning Technologies and opinion mining techniques was used for analysis of twitter with feature selection and classifier ensemble. This system discovered that the proposed system can be used for measuring users' opinion from their tweets which is very useful in many applications such as marketing, political polarity detection and reviewing products [26].

A model was made for sentiment analysis of unstructured social media text using machine learning algorithm and found that preparing a dataset by automatically collecting tweets using hashtags has an advantage as compared to the data set which is formed by manual annotation [27].

Twitter was used to predict the election result for candidate selection using Naïve Byes algorithm and concluded that the discrepancy between the real candidates who is predicted to win the election based on their popularity is more with other parameters and those who have only references with money power [28].

Twitter data was used for transit performance assessment, which is a framework for evaluating transit riders' opinions about quality service with the help of Machine learning techniques (LDA model) and suggested that in this study negative sentiments are mostly related to the performance of train and bus systems. It is clear that there are several negative tweets about Frontrunner and TRAX. Frontrunner is a commuter rail operated by UTA serving more than 16,000 riders each day [29].

World Environment day user opinions in twitter was studied using Machine learning; NVivo; textual analysis, python and SVM algorithm and concluded that the positive and negative tweets can establish on the key factors that most concern users tweets depending upon the hashtag on environment and public health such as climate change, global warming, extreme weather, water pollution, deforestation, climate risks, acid rain or massive industrialization expr # world environment day [30].

Air service quality was studies using twitter data as a resource using various ASQ methods conducted. Its finding indicates that the frequency of passenger references to the attributes of the scale differs significantly in some cases and that the discernment of these differences can provide

actionable insights for airport management when improving airport [31].

Interdisciplinary Optimism was studied using Machine learning algorithm, SVM, Naive Bayes method and after studying the Positive, negative and neutral tweets we found that the positive tweets are concerned to be the most positive experiences and optimistic thinking mostly among the youngsters for their bright future [32].

3.Application of Sentiment analysis of twitter data

The proposed work finds its application in many finds ranging from public health to disaster management. The techniques of sentiment analysis help in getting people opinion on various issues

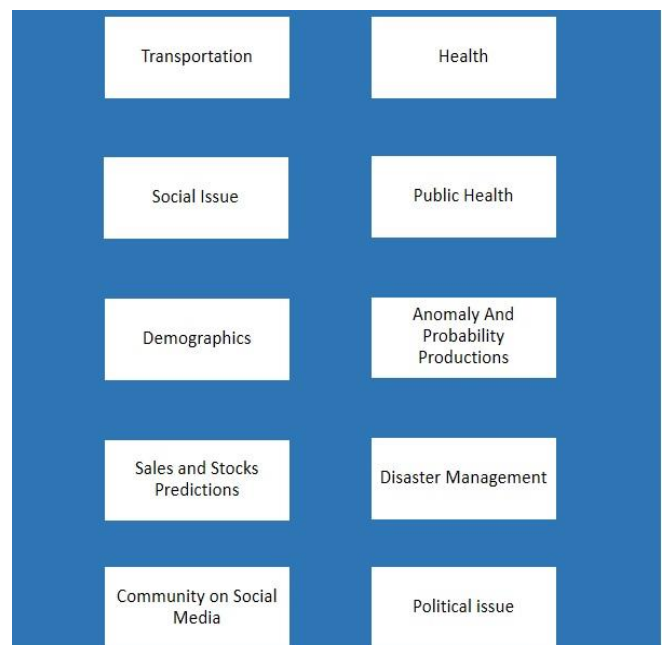


Fig. 1. Application of Proposed Approach

(i) Health- The techniques of sentiment analysis is used in analyzing patient feedback about your practice. It can help doctors and healthcare personnel gain a competitive edge over the competition and improve their service based on feedback provided.

(ii) Public Health- Various methods can be used for sentiment analysis of tweets in health care. There are many unstructured free-text tweets which are related to public health which being shared on twitter, which is becoming a popular ground for healthcare research.

(iii) Political Issues - The political tweets which are shared on Twitter can be analyzed which can be helpful in getting public opinion about a political issue.

(iv) Social Issues- A social issue is an issue which relates to people's personal lives and interactions, hence upon analysing the people's opinions about a social issue we can tell the general public opinion about an issue.

(v) Transportation- People express their opinion regarding the use of city public transportation via twitter. These opinions from the general public can be processed and analyzed to know the positive and negative opinions.

(vi) Disaster Management- Analysis of Disaster related tweets. In twitter is one of the methods that could gear up detecting tweets for situational awareness. Using it authorities can make out where they should look for particular information such as the most affected area, types of help to be deployed.

(vii) Sales and Stock Predication- People’s tweets about a particular brand and stocks can be analysis and its future standing thus, can be predicted.

4. PROPOSED APPORACH

The data is collected from twitter API using Tweepy, it is a Python library for accessing the Twitter API. It is great for simple automation and creating twitter bots. The data sets obtained from twitter is then pre-processed or cleaned from all the unwanted data like links, emoticons, and other non-relevant information which is not needed in the sentiment analysis. The features or keywords are then extracted from the processed data, which help us to classify the output. The processed data will be checked for keywords that will help to determine the nature of the tweet ie whether it is a positive tweet, negative tweet or a neutral tweet. After the given tweet is sentimentally analysed. It could be categorised effectively.

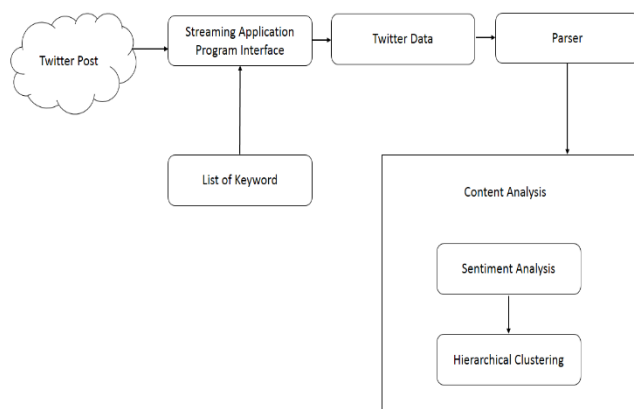


Fig. 2. Proposed Approach

Fig. 2 depicts the proposed approach. The data is taken from twitter from streaming Application programming interface (API), and the data taken is checked with a list of keywords provided. The twitter data which is in the form of TXT, JSON Format is passed to a parser, which converts the data in Javascript Object Notation(JSON) to Comma Separated Values (CSV). Content Analysis is then applied to the converted data, which is done in two steps. Sentiment Analysis is done using Support Vector Machine(SVM) and then Hierarchical Clustering executed using multiscale bootstrap resampling.

5. CONCLUSION

After studying the recent projects and studies related to social media analysis using twitter data we found that the, Twitter, being the most popular micro-blogging social site is used to collect data to perform analysis. As Twitter’s unique infrastructure and the near total availability of its data, ensured its high popularity among researchers. The popularity of using Twitter for social media research both in academia and in industry remains high; no other platform has attracted as much attention from academics. Twitter has many advantages over other social media sites because using tweepy API we can extract almost every data on twitter for processing.

Sentiment Analysis has various applications such as brand promotion and building, popular brands take the help of twitter to know the general opinion of the public regarding their products. Various algorithm such as Gaussian Algorithm, SVM and Naïve Byes Methods are used to tell the public opinion about the topics that are trending in world.

Election predictions can be made be made by analyzing peoples tweets, by doing sentiment analysis on the twitter data we can tell which candidate is more popular, which one has a more positive impact on the public.

References

- [1]. “The Infrastructure Behind Twitter: Scaling Networking, Storage and Provisioning” <https://www.infoq.com/news/2017/01/scaling-twitter-infrastructure/>.
- [2]. “Twitter” <https://www.britannica.com/topic/Twitter>.
- [3]. “TWITTER THEN AND NOW: HOW THE SITE HAS CHANGED SINCE ITS LAUNCH IN 2006” <https://www.independent.co.uk/life-style/twitter-launch-first-tweets-jack-dorsey-ceo-a8846261.html>.
- [4]. “Welcome to a world with 280-character tweets” <https://money.cnn.com/2017/11/07/technology/twitter-280-character-limit/index.html>.
- [5]. “twitter.com” <https://www.alexa.com/siteinfo/twitter.com>
- [6]. “For Election Day Influence, Twitter Ruled Social Media” <https://www.nytimes.com/2016/11/09/technology/for-election-day-chatter-twitter-ruled-social-media.html>
- [7]. “Tweeting Made Easier” https://blog.twitter.com/official/en_us/topics/product/2017/tweetingmadeeasier.html
- [8]. “Sentiment Analysis: Concept, Analysis and Applications” <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>
- [9]. “An Introduction to Text Mining using Twitter Streaming API and Python”

- <http://adilmoujahid.com/posts/2014/07/twitter-analytics/>
- [10]. "Social media platforms to follow 'code of ethics' in all future elections: EC"
<https://economictimes.indiatimes.com/news/politics-and-nation/social-media-platforms-to-follow-code-of-ethics-in-all-future-elections-ec/articleshow/71314195.cms>
- [11]. "19 Best Examples of How Top Brands Use Twitter"
<https://www.postplanner.com/blog/19-best-examples-of-how-top-brands-use-twitter/>
- [12]. "Understanding Sentiment Analysis: What It Is & Why It's Used"
<https://www.brandwatch.com/blog/understanding-sentiment-analysis/>
- [13]. Hussein, Doaa Mohey El-Din Mohamed. "A survey on sentiment analysis challenges." *Journal of King Saud University-Engineering Sciences* 30.4 (2018): 330-338.
- [14]. Srivastava, Ritesh, and M. P. S. Bhatia. "Ensemble methods for sentiment analysis of on-line micro-texts." *2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*. IEEE, 2016.
- [15]. Wang, Lei, Jianwei Niu, and Shui Yu. "SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis." *IEEE Transactions on Knowledge and Data Engineering* (2019).
- [16]. Srivastava, Ritesh, M. P. S. Bhatia, Veena Tayal, and J. K. Verma. "Framework for Real-World Event Detection Through Online Social Networking Sites." In *Data and Communication Networks*, pp. 195-203. Springer, Singapore, 2019.
- [17]. Ali, Farman, Daehan Kwak, Pervez Khan, Shaker El-Sappagh, Amjad Ali, Sana Ullah, Kye Hyun Kim, and Kyung-Sup Kwak. "Transportation sentiment analysis using word embedding and ontology-based topic modeling." *Knowledge-Based Systems* 174 (2019): 27-42.
- [18]. Tomar, Nandini, Ritesh Srivastava, and J. K. Verma. "Analysing Public Sentiment on GST Implementation in India." *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*. IEEE, 2018.
- [19]. Tomar, Nandini, Ritesh Srivastava, and J. K. Verma. "Mining Public Opinion on Aadhaar Card Linking in India." *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*. IEEE, 2018.
- [20]. Nazir, Faria, Mustansar Ali Ghazanfar, Muazzam Maqsood, Farhan Aadil, Seungmin Rho, and Irfan Mehmood. "Social media signal detection using tweets volume, hashtag, and sentiment analysis." *Multimedia Tools and Applications* 78, no. 3 (2019): 3553-3586.
- [21]. Srivastava, Ritesh, and Veena Mittal. "Most preferable combination of explicit drift detection approaches with different classifiers for mining concept drifting data streams." *International Journal of Data Science* 4.3 (2019): 196-214.
- [22]. Hasan, Ali, Sana Moin, Ahmad Karim, and Shahabuddin Shamsheerband. "Machine learning-based sentiment analysis for twitter accounts." *Mathematical and Computational Applications* 23, no. 1 (2018): 11.
- [23]. Malik, Monica, Sameena Naaz, and Iffat Rehman Ansari. "Sentiment Analysis of Twitter Data Using Big Data Tools and Hadoop Ecosystem." *International Conference on ISMAC in Computational Vision and Bio-Engineering*. Springer, Cham, 2018.
- [24]. Öztürk, Nazan, and Serkan Ayvaz. "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis." *Telematics and Informatics* 35.1 (2018): 136-147.
- [25]. Goel, Prerna, and Reema Thareja. "Emotion Analysis of Twitter Data Using Hashtag Emotions." *International Conference on Application of Computing and Communication Technologies*. Springer, Singapore, 2018.
- [26]. Fouad, Mohammed M., Tarek F. Gharib, and Abdulfattah S. Mashat. "Efficient Twitter sentiment analysis system with feature selection and classifier ensemble." *International Conference on Advanced Machine Learning Technologies and Applications*. Springer, Cham, 2018.
- [27]. Rout, Jitendra Kumar, Kim-Kwang Raymond Choo, Amiya Kumar Dash, Sambit Bakshi, Sanjay Kumar Jena, and Karen L. Williams. "A model for sentiment and emotion analysis of unstructured social media text." *Electronic Commerce Research* 18, no. 1 (2018): 181-199.
- [28]. Prabhu, BP Aniruddha, B. P. Ashwini, Tarique Anwar Khan, and Arup Das. "Predicting Election Result with Sentimental Analysis Using Twitter Data for Candidate Selection." In *Innovations in Computer Science and Engineering*, pp. 49-55. Springer, Singapore, 2019.
- [29]. Haghghi, N. Nima, Xiaoyue Cathy Liu, Ran Wei, Wenwen Li, and Hu Shao. "Using Twitter data for transit performance assessment: a framework for evaluating transit riders' opinions about quality of service." *Public Transport* 10, no. 2 (2018): 363-377.
- [30]. Reyes-Menendez, Ana, José Saura, and Cesar Alvarez-Alonso. "Understanding# WorldEnvironmentDay user opinions in Twitter: A topic-based sentiment analysis approach." *International journal of environmental research and public health* 15.11 (2018): 2537.
- [31]. Martin-Domingo, Luis, Juan Carlos Martín, and Glen Mandsberg. "Social media as a resource for sentiment analysis of Airport Service Quality (ASQ)." *Journal of Air Transport Management* 78 (2019): 106-115.
- [32]. Weber, Charlotte Teresa, and Shaheen Syed. "Interdisciplinary optimism? Sentiment analysis of Twitter data." *Royal Society open science* 6.7 (2019): 190473.
- [33]. Plunz, Richard A., Yijia Zhou, Maria Isabel Carrasco Vintimilla, Kathleen Mckeown, Tao Yu, Laura Ugucconi, and Maria Paola Sutto. "Twitter sentiment in New York City parks as measure of well-being." *Landscape and Urban Planning* 189 (2019): 235-246.
- [34]. Ahmad, Shakeel, Muhammad Zubair Asghar, Fahad M. Alotaibi, and Irfanullah Awan. "Detection and classification of social media-based extremist affiliations using sentiment analysis techniques." *Human-centric Computing and Information Sciences* 9, no. 1 (2019): 24.
- [35]. Tyagi, Priyanka, and R. C. Tripathi. "A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data." Available at SSRN 3368718 (2019).
- [36]. Mostafa, Mohamed M. "Clustering halal food consumers: A Twitter sentiment analysis." *International Journal of Market Research* 61.3 (2019): 320-337.
- [37]. Himelboim, Itai, et al. "A Social Networks Approach to Understanding Vaccine Conversations on Twitter:

Network Clusters, Sentiment, and Certainty in HPV Social Networks." *Health communication* (2019): 1-9.

- [38]. Tayal, Veena, and Ritesh Srivastava. "Challenges in Mining Big Data Streams." *Data and Communication Networks*. Springer, Singapore, 2019. 173-183.