# Analysis of Sentiment Techniques for Twitter Data

Mitali K. Shah A[1], Dr Satyen Parikh B[2]
[1]*PhD Scholar, Ganpat University*
[2]*Dean, Ganpat University*
*(E-mail: shahmitali07@gmail.com)*

***Abstract*—**Sentiment analysis is to rectify writer's attitude or to understand his/her opinion in the form of expressed words or text. Running era is totally depends on the social media where each one expresses their feelings, emotions, own thoughts and opinions. To generalize it and to have better understanding of this analysis, NLP would be the main tool. Reason behind the linking of NLP and sentiment analysis is, working procedure and algorithm of NLP is used to interpret the text and make it analyzable. It also deals with the interaction and interpretation of computer and human natural language. In the area of sentiment analysis, NLP has been primarily used also equally import for artificial intelligence. Opinion mining is also part of Sentiment analysis. It makes use of NLP in order to categorize the opinions of people with respect to products or the reviews. It deals with opinions and perspective of human related to emotions and attitude about some occurrence or the events. The sentiment analysis is done to check the positive, negative and neutral opinion of users about products to check its popularity or importance in the market. This research work is solely based on the type , method and steps.

***Keywords*—***Twitter, Sentiment analysis, Machine learning*

## I. INTRODUCTION

### 1.1 Sentiment analysis

Sentiment analysis is defined as the opinion mining that every individual has feelings, thoughts, and emotion for a particular event. To determine whether the given task is positive, negative or neutral, sentiment analysis (SA) has been used. It shows that sentiment analysis can be interpreted as a task of a classification in which each sentiment represents a new category. The computer science and artificial intelligence field Natural Language Processing (NLP) plays an important role in dealing with interaction between human and computer language. This field is particularly of use to merchants, stock traders, and in election works. Due to change in marketing level competition and the needs of the people a lot of research is needed in sentiment analysis [1].

Now a days Social Media such as twitter or other blogging sites are the base for the advertising market as well as sharing views about product [2], [3], [4], [5]or movies [6].It also becomes first choice for reviewing events or political aspiration [7], etc.

Main attraction of twitter is "express emotions with minimum number of words". That characteristic of twitter for not allowing more than 280 words makes it unique from other microblogging web media. One more advantage with respect to user is that there is no dependency over user to user. Means, one should not require to be followed by someone to see the his/her tweeter activity and likewise it is possible to have influential users, whose followers are numerous and whose tweets have usually huge impact on the public opinion [8],[9]. or even to ''create'' ones[10].

The speaker attitude and opinion are determined by sentiment analysis that's why it is also derived as opinion mining. Opinion mining involves building a system to collect and categorize opinions about a product. The further improvement has been performed on data collected from different users to analyze existing opinions. The social network sites act as a medium where user can post their opinions. It is mainly done through blog posts, Online forums, social media and product review websites etc. The use of Facebook, Twitter, Google, Instagram like social network sites has changed the people way to express their views or opinions.

A Platform has been provided to businesses through social media for sharing and advertisings their products with customers. Nowadays people depend on the content or data provided online for decision making. For the estimation of products and its usage many companies using sentiment analysis. The main cores of the sentiment analysis are the attitudes, appraisals, sentiments and opinion. It also helps policy makers and politicians to figure out the issues such as public grievances, public policies and issues. Sentiment analysis helps customer to buy a product and understand products and its services. The factual data present mainly focus on processing, searching or analyzing the retrieval textual information technique. It's very challenging to develop new applications due to huge growth of available information gathered from blogs and social networks. Twitter is mainly used to tweet ongoing messages using microblogging and social networking sites. New challenges have been created due to different unique characteristics present in tweets that also shape the different domains and methods conveyed in sentiment analysis

### 1.2 Steps for the Sentiment analysis

1 Data Collection

2 Data Preparation
3 Develop Model
4 Train and Update

### 1.3 Various Faces of Sentiment analysis

1 Data Extraction: - The input data set consists of real time tweets that have been extracted from Twitter using Tweepy.
2 Data Pre-processing: - In the pre-processing phase, the input data set is cleaned and transformed into a form suitable for feature extraction.
3 Feature Extraction: - The pre-processed data is fed as input to the feature extraction algorithm. In this step weights are assigned to the keywords thus preparing them for classification. In this paper N-Gram modelling technique has been used for designing the feature extraction algorithm.
4 Sentiment Classification: - In based on their polarity. In this work, k-nearest neighbour classifier has been used for classification into positive, negative and neutral classes.

## II. LITERATURE REVIEW

**Shahnawaz, et.al, (2017)** presented that sentiment analysis is the process to identify the opinion or feelings expressed in the opinioned data, in order to find the attitude of writer towards the particular topic whether it is positive, negative or neutral. It provides idea to the customer to identify the product or service is satisfactory or not before the customer buys it. Public opinions on different types of social media are the major concern of the scientific communities and business world to gather and extract public views. Inadequacy accuracy, inability to perform well in different domain and performance are the main issues in the current techniques. Author concluded, by using semi-supervised and unsupervised learning-based models, it will be easy minimize lack of labeled data if enough unlabeled data is available. [11]

**Chhaya Chauhan, et.al, (2017)** major researches in the field of current time are natural language processing, text analysis, text preprocessing, stemming etc. to analyze the unstructured data. In order to generate desired results, different techniques and tools are used. Basically, these techniques allow a computer to understand what is being said by humans. Sentiment of a text or sentence is determined using different techniques, as internet has a large repository of natural language. Companies these days are not fully aware of customer requirements and needs as people share their thoughts and reviews which are subjective in nature. To understand the sentiments of the people products reviews are necessary, therefore a summary of positive and negative reviews is needed to be generated [12]. Author's main focus was on the review of algorithms and techniques used to extract feature wise summary of the product and analyzed them to form an authentic review. In future, focus will be on higher level natural language processing tasks. Best techniques or tools should be used for more accurate result in which only those keywords are used which are in dataset and rest of the words are eliminated by the system.

**Pragya Juneja, et.al, (2017)** the objective of this research is to classify twitter data into sentiments of positive and negative by using different supervised machine learning classifiers to predict the Delhi Corporation Elections results. As today social networking sites like Twitter, Facebook, Instagram etc. plays a major role in sentiment analysis. Basically, sentiment analysis is used to identify the opinion and emotional states of the people in order to extract positive and negative views. Twitter analysis has to be done firstly, by extracting the twitter posts twitted by user and analyzes the present scenario, for example it was shown that BJP is more successful political party based on people opinion [13]. Therefore, author concluded that sentiment analysis can be utilized for any purpose based on the tweets they collected like marketing, finance, media, and entertainment and many more.

**Jamil Hussain, et.al, (2017**) Due to tremendous growth in communication technologies and population the use of social-networking sites has been increased. As with the advancement of technology, mini computers and smart phones are easily available to the human pockets with ease to share and express new ideas using social media platform like Facebook, Blogs, and Twitter etc. In this paper, author presented how to find the depression level of a person like its state of low mood and aversion to activity that can affect a person's thoughts, behavior, feelings and sense of well-being. It can be observed by extracting emotions from the text, using emotions theories, machine learning techniques and natural language processing technique on different social media. In this research, they have compared different existing classifiers such as SVM, ME and NB in order to measure depression by sentence level sentimental analysis. The author had adopted feature selection, voting model techniques and performed different experiments using twitter and 20newsgroups dataset. The results showed that SVM is superior to Naïve bayes in terms of accuracy [14].

**Zahra Rezaei et.al, (2017**) large amount of data has been created using social networking sites and users share their ideas and small messages using micro blogging website. Microblog sites are mainly used to share their opinions and to post their daily views, using microblogging service popularly known as Twitter. General overview of a specific topic has become a new challenge as number of tweets increases on daily basis. To predict the sentiment on twitter, algorithms will be used in real-time and under limited time as twitter messages are constantly increasing and arriving at very high speed. Hoeffding tree algorithm has been used as popular tool in mining data streams and appropriate approach in order to obtain splitting attribute as data in twitter follows data stream model. To enhance the performance filtering and wrappers techniques were used. In twitter language common preprocessing task is used before applying any algorithms

[15]. Hoeffding tree has the high process time, in order to eliminate this Mcdiamid tree algorithm is used. McDiarmid tree performance is better than Hoeffding tree but the accuracy of both are same. Due to large amount of twitter data, minimum processing time is required for the sentimental analysis therefore, McDiarmid Tree is better than Hoeffding tree.

**Wei Zhao, Ziyu Guan et.al, (2017)** to ease new buyers in making good decisions products review are necessary. A new technique has been introduced for opinion mining which help us to determine the positive and negative of a post or review. For solving sentiment classification problems deep learning has effective means and without using human efforts a neutral network represents. Success of deep learning solely depends on the large-scale training data. In this paper, they had given a review on different sentiment classification using purposed deep learning framework that employs commonly existing rating. Adding a classification layer and learning a high-level representation are the two steps consists in purposed framework. In order to achieve supervised fine tunings a level sentences are used and on the top of embedding layer a classification layer is added on the other hand in the first step rating information is used to capture the general sentiment distribution. The long short memory and conventional feature extractors are used for low level network structure that helps in modeling review sentences. The Amazon data sets have been used that contained 1.1 weakly review sentences and 11,754 labeled review sentences. To check the proposed framework different experiments have been performed that show its superiority over baselines [5]

**Pulkit Garg, et.al, (2017)** as surveyed that social media has becoming a medium for online sharing by the increase of a greater number of people coming online. In this paper, we study post- terror attack tweets by extracting it from twitter. The flow data posted on twitter is used to study factors like last retweet, number of retweets and number of favorites. Maximum number of retweets indicates maximum reach. It creates widespread reaction on the social media. Governments are concentrating on digitalizing the whole nation. Due to increase in number of people, huge data is generated. Author discussed the Uri Terror Attacks that show more negative tweets tend to survive as compare to positive tweets, although their amount is low. It will lead to public unrest if people start targeting a community and provide negative information. Misleading information, the trends of retweets and number of favorites are the future scope to study its flow and survival [16].

**Ana Valdivia et.al, (2017)** Researchers working in the field of natural language processing and text mining received a lot of attention on sentiment analysis. To operate all domains there was lack of annotated data being used that hampered the accuracy of sentiment analysis. The issue is geared up after attempting many attempts. In this paper, authors' provided techniques and systematic literature review on cross-domain sentiment analysis. According to author there was no perfect solution hence to solve the problems of cross domain sentiment analysis different techniques, methods and approaches had been used to develop more accurate data in near future. The fuzzy majority based on aggregating polarity for several sentiment analyses the use of Induced Ordered weighted is proposed. Author's main focused on removing those neutral reviews labeled by accord of collections [17].

## III. TABLE OF COMPARISON

| Authors' Names | Year | Description | Outcomes |
|---|---|---|---|
| Shahnawaz, Parmanand Astya | 2017 | Inadequacy accuracy, inability to perform well in different domain and performance are the main issues in the current techniques. | Author concluded, by using semi-supervised and unsupervised learning-based models, it will be easy minimize lack of labeled data if enough unlabeled data is available. |
| Chhaya Chauhan, Smriti Sehgal | 2017 | To understand the sentiments of the people products reviews are necessary, therefore a summary of positive and negative reviews is needed to be generated. | Author's main focus was on the review of algorithms and techniques used to extract feature wise summary of the product and analyzed them to form an authentic review. |
| Pragya Juneja, Uma Ojha | 2017 | The objective of this research is to classify twitter data into sentiments of positive and negative by using different supervised machine learning classifiers to predict the Delhi Corporation Elections results. | Therefore, author concluded that sentiment analysis can be utilized for any purpose based on the tweets they collected like marketing, finance, media, and entertainment and many more. |
| Jamil Hussain, | 2017 | In this paper, author presented | The author had adopted feature |

| | | | |
|---|---|---|---|
| Anees Ul Hassan, Musarrat Hussain, Muhammad Sadiq, Sungyoung Lee | | how to find the depression level of a person like its state of low mood and aversion to activity that can affect a person's thoughts, behavior, feelings and sense of well-being. | selection, voting model techniques and performed different experiments using twitter and 20newsgroups dataset. The results showed that SVM is superior to Naïve bayes in terms of accuracy |
| Zahra Rezaei, Mehrdad Jalali | 2017 | Hoeffding tree algorithm has been used as popular tool in mining data streams and appropriate approach in order to obtain splitting attribute as data in twitter follows data stream model. To enhance the performance filtering and wrappers techniques were used. | Due to large amount of twitter data, minimum processing time is required for the sentimental analysis therefore, McDiarmid Tree is better than Hoeffding tree. |
| Wei Zhao, Ziyu Guan, Ziyu Guan,Long Chen, Xiaofei He,Deng Cai, Beidou Wang and Quan Wang | 2017 | A new technique has been introduced for opinion mining which help us to determine the positive and negative of a post or review. For solving sentiment classification problems deep learning has effective means and without using human efforts a neutral network represents. | The long short memory and conventional feature extractors are used for low level network structure that helps in modeling review sentences. The Amazon data sets have been used that contained 1.1 weakly review sentences and 11,754 labeled review sentences. |
| Pulkit Garg, Himanshu Garg, Virender Ranga | | In this paper, we study post-terror attack tweets by extracting it from twitter. The flow data posted on twitter is used to study factors like last retweet, number of retweets and number of favorites. | It will lead to public unrest if people start targeting a community and provide negative information. |
| Ana Valdivia, M. Victoria Luz´ and Francisco Herrera | 2017 | In this paper, authors' provided techniques and systematic literature review on cross-domain sentiment analysis. | The fuzzy majority based on aggregating polarity for several sentiment analyses the use of Induced Ordered weighted is proposed. Author's main focused on removing those neutral reviews labeled by accord of collections. |

## IV. CONCLUSION

This paper examines the different methodologies and steps for sentiment analysis. Numerous approaches have been designed in the recent years for sentiment analysis. However complete efficiency has not been achieved so far. Yet at the same time the nearness of unpredictable and objective sentences ruins the execution. The main challenges that come up are named entity recognition, anaphora resolutions, negation expressions, sarcasms, abbreviations, misspellings, etc.

## REFERENCES

[1] Devika, M. D., Cª Sunitha, and Amal Ganesh. "Sentiment analysis: A comparative study on different approaches." Procedia Computer Science 87 (2016): 44-49.

[2] O'Connor, Brendan, et al. "From tweets to polls: Linking text sentiment to public opinion time series." Icwsm 11.122-129 (2010): 1-2.

[3] Cabanlit, Mark Anthony, and Kurt Junshean Espinosa. "Optimizing N-gram based text feature selection in sentiment analysis for commercial products in Twitter through polarity lexicons." Information, Intelligence,

Systems and Applications, IISA 2014, The 5th International Conference on. IEEE, 2014

[4] Chauhan, Chhaya, and Smriti Sehgal. "Sentiment analysis on product reviews." Computing, Communication and Automation (ICCCA), 2017 International Conference on. IEEE, 2017.

[5] Zhao, Wei, et al. "Weakly-supervised deep embedding for product review sentiment analysis." IEEE Transactions on Knowledge and Data Engineering 30.1 (2018): 185-197.

[6] Hodeghatta, Umesh Rao. "Sentiment analysis of Hollywood movies on Twitter." Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM, 2013

[7] Soler, Juan M., Fernando Cuartero, and Manuel Roblizo. "Twitter as a tool for predicting elections results." Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on. IEEE, 2012.

[8] Cha, Meeyoung, et al. "Measuring user influence in twitter: The million follower fallacy." Icwsm 10.10-17 (2010): 30.

[9] Trusov, Michael, Anand V. Bodapati, and Randolph E. Bucklin. "Determining influential users in internet social networks." Journal of Marketing Research 47.4 (2010): 643-658.

[10] Messias, Johnnatan, et al. "You followed my bot! Transforming robots into influential users in Twitter." (2013)

[11] astya, parmanand. "Sentiment analysis: approaches and open issues." Computing, communication and automation (ICCCA), 2017 international conference on. Ieee, 2017.

[12] chauhan, chhaya, and smriti sehgal. "Sentiment analysis on product reviews." Computing, communication and automation (ICCCA), 2017 international conference on. Ieee, 2017.

[13] juneja, pragya, and uma ojha. "Casting online votes: to predict offline results using sentiment analysis by machine learning classifiers." 2017 8th international conference on computing, communication and networking technologies (ICCCNT). Ieee, 2017.

[14] hassan, anees ul, et al. "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression." Information and communication technology convergence (ICTC), 2017 international conference on. Ieee, 2017.