**TWITTER SENTIMENT ANALYSIS USING VADER**

**D.Chakradhar**

(SRM Institute of Science and Technology, City Campus,

dchakri123@gmail.com)

**S.Thiagaraj**

(SRM Institute of Science and Technology, City Campus,

thiagaraj.sundaram@yahoo.in)

**Somen Kumar**

(SRM Institute of Science and Technology, City Campus,

[kingsomen333@gmail.com)](mailto:kingsomen333@gmail.com))

**Teja Rajashekar.B**

(SRM Institute of Science and Technology, City Campus,

[tejarajashekar@gmail.com)](mailto:kingsomen333@gmail.com))

**Dr. G. Pavai Anand**

(Assistant Professor,SRM Institute of Science and Technology, City Campus

[pavai\_gops@yahoo.co.in](mailto:pavai_gops@yahoo.co.in))

**ABSTRACT**:

Identifying the thought process of humans has been an important part of information gathering. With the increasing availability and popularity of opinion-rich resources such as micro blogging websites, online review sites and personal blogs, new opportunities and challenges have risen as people now can, and do, actively use information technologies to seek out and understand the opinions of others. Social media sites have emerged as one of the most important platforms to raise opinions and influence the way any business is commercialized. User views are utilized to analyze how the propagation of information impacts the lives in a large-scale network like Twitter. Twitter is an online news and social networking service on which users post and interact with messages known as "tweets". The process of extracting meaningful customer insight in the form of text from the users to rate the raw data obtained in terms of sentiment score is known as Sentiment analysis. Twitter sentiment analysis refers to the use of tweets to determine the polarity and inclination of vast population towards a specific topic, event, item or entity. The applications of such analysis can be observed on a widespread scale during elections, movie promotions, brand endorsements and many other fields .

The primary aim of this project is to successfully analyzing sentiment score in noisy twitter streams. This report describes and analyses in detail on the design of sentiment analysis by extracting a vast number of tweets using the twitter API. Results are then classified in accordance with user's perception via tweets into two categories namely positive and negative. Various methodologies ,models and techniques applied and adopted in this reportto carry out sentiment analysis on twitter data is also discussed in detail.

**1. INTRODUCTION:**

As the Internet is engulfing our world, its horizons are expanding ten folds on a daily basis. Social Media and Micro blogging platforms like Twitter, Tumblr, Facebook, Reddit have become dominant in spreading encapsulated news and trending topics across the globe at lightning pace

In recent years Twitter has emerged as a major micro-blogging website, amassing over 150 million users generating over 800 million tweets every day. Users share their opinions by composing tweets, that have a maximum limit of 140 characters. A topic is said to be trending if more and more users contribute their opinion and judgements towards the topic using a common hashtag, thereby making it a valuable source of online perception. Such trending topics generally garner awareness or are use with the intention of promote public figures, political campaigns during elections, product endorsements and entertainment like movies, award shows. Large organizations and firms take advantage of people's feedback to improve their products and services which further help in enhancing marketing strategies.

With innumerable tweets coming up every day, companies are able to model in customer with the help of graph and pie charts to graphically illustrate the sentiment reflected in their tweets. Despite being a goldmine of opinion, twitter language is unstructured. Users inevitably compose tweets that is completely devoid of sentence structure using slang, abbreviations, emoticons etc. Along with this, people also use sarcastic and polysemy language in their tweets . To overcome these obstacles

sentiment analysis is used to give accurate result in terms of percentage sentiment on a particular scale.

**2.PROBLEM STATEMENT:**

The aim of this project is to design a Twitter Sentiment Engine that will classify a tweet into one of the following three categories , namely : positive/negative/neutral. It also aims to vastly improve on the accuracy of the currently existing engines, analysis and algorithms by using a noise free training data, obtained by using an intensive approach for subjective/objective classification to eliminate out as many neutral and unwanted tweets as possible.

**3.LITERATURE REVIEW:**

Some of the existent research and scholarly works in the field of Machine Learning and data

mining to analyse sentiments using Twitter API and prepare prediction model for various applications are summarized in this section. As the number of interaction on available social platforms keep shooting up exponentially, information available on such micro-blogging sites like Twitter are becoming vast and can be extracted to turn into business objectives, social campaigns, marketing and other promotional strategies as explained in [4]. The benefit of social media to know public opinions and extract their emotions are considered by authors in [2] and explained how twitter gives advantage politically during elections. Further, the concept of the hashtag is used for text classification as it conveys emotion in few words. They indicated how previous research work suffered from lack of training set and missed some essential features of target data. The two stage approach opted for their framework consisted of first collecting and preparing training data from twitter using data mining. Then the conveying relevant features are then propounded into the

Supervised Learning Model to predict the results of Elections held in USA in 2016. After collecting and

preprocessing the tweets, training data set was created first by manual labelling of hashtags and forming clusters. The data was then subjected to online Sentimental Analyzer VADER which predicts the polarity in percentage. This approach considerably reduced the number of tweets and training set involved. Support Vector Machine(SVM) and Naive Bayes was further applied as the classification algorithm to determine the polarity of tweets. Multistage Classification approach was used where an

entity classifier receives general class of tweets and categorize them with respect to individual candidates for comparison. The metric they used to determine the winner was the “PvT ratio” which refers to the proportion of Positive number of tweets to total count of tweets for the respective candidate

1. **Sentiment Analysis using Python**

Sentiment refers to a view or an opinion held by a person. Sentiment analysis refers to the accurate prediction of emotions in a word, sentence or corpus of documents

It is the process of calculating sentiment and finding out the user’s point of view in that particular statement or sentence. It is a classification technique which is used to derive opinion from the tweet and formulates a score which reflects the inclination of the user towards a particular topic as convey through the tweet. It is intended to serve as an application to understand the attitudes, opinions and emotions expressed within an online mention. The intention is to gain an overview of the wider public opinion behind certain topics. Sentiments are subjective to the topic of interest. As sentiment is subjective to say the least, it is necessary to determine which Features are important and need to be extracted from the text. For example, we can have two-class tweet sentiment classification (positive and negative) or three class tweet sentiment classification (positive, negative and neutral). Sentiment analysis precisely, is he aforementioned task of categorizing conversations into positive, negative or neutral labels. The basic steps for performing sentiment analysis includes data collection, pre-processing of data, feature extraction, selecting baseline features, sentiment detection and performing classification either using simple computation or else machine learning approaches. The dimension of the sentiment class is a crucial factor in deciding the efficiency of the model. The output to be calculated has to be programmed in terms of a class in the sentiment calculator. The greater the efficiency of the program the better and refined are the results. Sentiment analysis approaches can be broadly categorized into two classes – lexicon based and machine learning based. Lexicon based approach is an example of unsupervised learning as it proposes to perform analysis using lexicons and a scoring method to evaluate opinions. Machine learning approach on the other hand involves the use of feature extraction and training the model using feature set and some dataset.

**4.1 TWITTER SENTIMENT ANALYSIS**

Twitter sentiment analysis primarily aims to categorize the tweets into three common classes, namely, positive sentiment class, negative sentiment class, neutral sentiment class. Calculation can be done to add more features like most used words by a user, impressions of a certain tweet, frequently used words ,most used emoticon and average sentiment of the complete data can also be calculated.

Performing sentiment analysis is challenging on Twitter data, the reasons being:

1.Restricted Tweet size: Only a maximum of 140 characters can be used, tweets are generally compact statements, which results in a sparse set of features. Also the use of abbreviations, slang and emoticons in the tweets make it difficult to analyze the data.

2.Use of Slang: Slang consists of informal English words and phrases which are more common in speech than writing. Because of the evolutionary use of slangs, it is difficult for any system/algorithm to accurately detect it.

Twitter Features: The use of hashtags, user reference, emoticons, user names and URLs require different processing than other words.

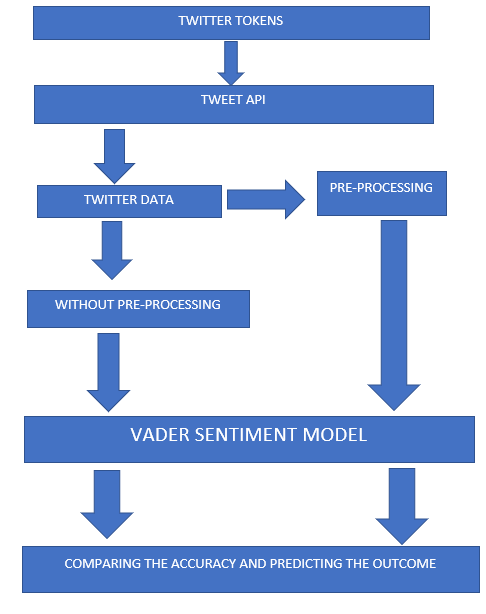
User Variety: Users express their opinions in a variety of ways, some using different language in between, while others using repeated words or symbols to convey an emotion. Some only use series of emoticons to represent a visual picture, while some on the other hand, use sarcastic statements which are hard to identify

Apart from these, we also face another set of problems in feature extraction which is the result of inadequate information available via tweets.

**4.2 Methodology**

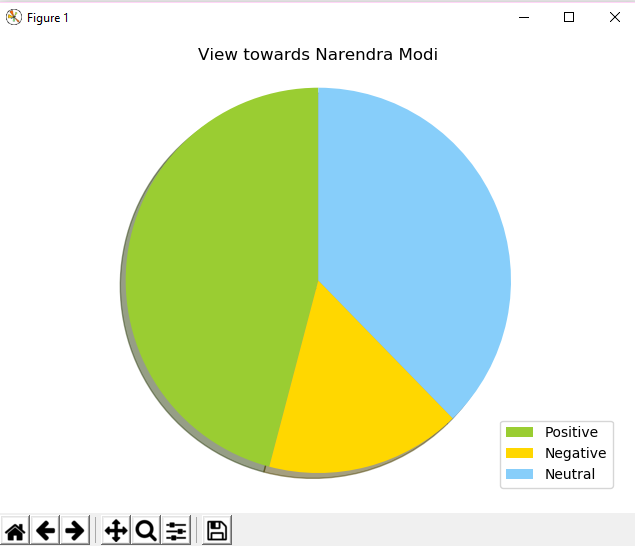
In order to perform sentiment analysis, we are required to collect data from the desired source (Twitter in this case). This data then undergoes various stages of pre-processing,

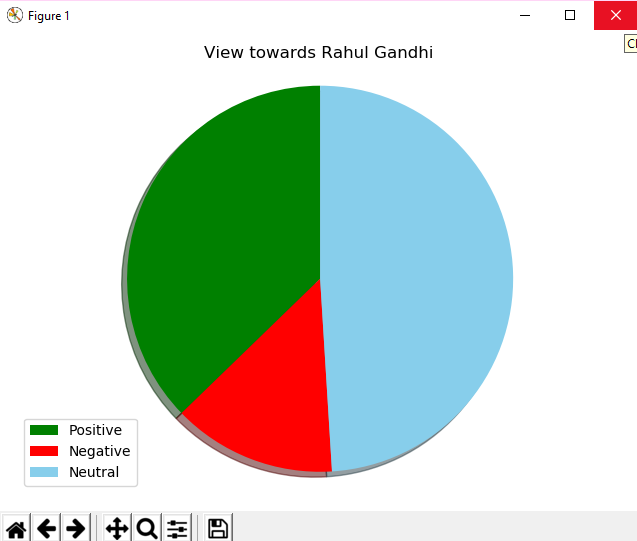
like removal of stopwords and removing hashtags, semicolons , emotes and other unwanted characters in the tweet data to make it more machine sensible than its previous form. Given below is flow- process representation of the pre-processing of the tweets. We use two different functions , one for tweets without pre-processing and one with pre-processing to compare the accuracy of the model and finally predict the outcome of the forthcoming election using the no. of positive sentiment tweets percentage for the two specified subjects, in this case Shri. Narendra Modi and Rahul Gandhi.



**DATA-FLOW DIAGRAM FOR THE PROJECT**

**5.RESULTS**





**6. CONCLUSION AND FUTURE SCOPE**

**6.1 CONCLUSION**

Twitter sentiment analysis using vader was implemented successfully. The analysis is used for predicting the outcome of the forthcoming elections. Two prominent leaders of our country, Shri. Narendra Modi and Shri. Rahul Gandhi were used as the subjects for our analysis Largely neutral sentiment was displayed towards both leaders, this indicates info about the leader’s day to day political activities being broadcasted by the twitter users that’s devoid of any sentiment. Thorough scrutinizing of tweets collected on a daily basis showcased more positive sentiment being displayed towards Shri. Narendra Modi than towards Shri. Rahul Gandhi.

Our analysis has helped us to conclude that Narendra Modi is the prospective candidate favored by the twitter users for the forthcoming elections

**6.2. FUTURE SCOPE**

The data available online is inevitable which can be further useful in analyzing sentiments of people online. Sentiment analysis or opinion mining finds its application in almost every field. Most brands are using sentiment analysis tools to view and analyze the opinions of the actual users who are using their brand products. Similar to this the brands can also have an idea of what kind of services are appreciated by the customers along with their products. With the help of opinion mining news and media houses are trying to analyze the image of political leaders either after or before elections. But still a lot of challenges come in the way of sentiment analysis as there are some cases when the reviews available are influenced or paid reviews. The future of sentiment analysis lies in not only attaining accurate results but also detecting fake reviews.

1. **REFERENCES**

Datasets ( Helpful for making the model ): https://stackoverflow.com/questions/24605702/good-dataset-for-sentiment-analysis http://www.sananalytics.com/lab/twitter-sentiment/ https://inclass.kaggle.com/c/si650winter11/data http://www.ai.stanford.edu/~amaas/data/sentiment/ https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/ Few lecture slides

https://lct-master.org/files/MullenSentimentCourseSlides.pdf

https://web.stanford.edu/class/cs124/lec/sentiment.

Few blogs:

http://www.datasciencecentral.com/profiles/blogs/text-classification-sentiment-analysis-

tutorial-blog

https://www.quantinsti.com/blog/sentiment-analysis-news-python/

Few papers:

1.Mouthami, K.; Devi, K . N.; Bhaskaran V. M. Sentiment Analysis and Classification Based on

Textual Reviews. In Proceeding 2013 of IEEE

International Conference on Information Communication & Embedded System, 2013

2.Liu, B.; Indurkhya, N.; Damerau. Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing. 2010.

SEARCH ENGINE

www.google.com

www.python.org.com