

Twitter Sentiment Analysis

Tripti Tripathi¹, Aman Kumar Singh², Shalini Bhagat³, Mr.Basetty Mallikarjuna⁴

B.Tech CS, Galgotias University, Greater Noida, India^{1,2,3}

Guide, Galgotias University, Greater Noida, India⁴

Abstract: It is known that social media is one of the largest sources of unstructured data. Analyzing that data and harvesting meaning out of that is a tedious job. Recently opinion mining has become an emerging topic due to the vast amount of opinioned data available on the various social networking sites. Microblogging has appeared relatively recently, and twitter is the most popular microblogging sites used by the people. It is one of the biggest free, open data-source. In the world today, twitter often witnesses a lot of opinions. Opinion mining and sentiment analysis help researchers to gain insight into public emotions. In this paper, Twitter is used as a source of opinioned data. Twitter APIs are used for the collection of tweets. In this paper, R is used for the acquisition, preprocessing, analyzing the tweets, then sentiment analysis is performed based on the different approaches. In this paper, Tweets were collected from the period of Jan 2019 to March 2019. Using that tweets, sentiment analysis was performed to gain the opinion polarity of the folks concerning general elections held in India. Two candidates were considered for this study: Candidate-1 and Candidate-2. It was concluded that Candidate-1 is more liked and is famous as compared to Candidate-2. The result obtained in our paper was in full compliance with the actual election results obtained in May 2019.

Keywords: Twitter mining; Text Mining; Sentiment Analysis; Social media

I. INTRODUCTION

It is known that just a small percentage of the whole data which is produced daily all over the world is structured in nature, rest of about 79% of the data is unstructured in nature meaning thereby they don't have any predefined archetype. Social media is one the largest source of unstructured data. Now days, people have become so habitual to social media, that they cannot imagine a day without it. People express their happiness, anger, sadness and every other emotion of the various social networking sites. They not only use it as a communication medium, also they keep on posting their day to day experiences on it. Social media, in other words have become a pool of sentiments. The unstructured data is dirty in nature. Therefore, working with it is a tedious job. Recently text mining and sentiment analysis have gained huge attention because of the availability of masses of data from number of social networking sites.

In the older days, when one wanted to know about what people feel and think about for instance a particular product/item they used to do surveys. Surveys were expensive and time-consuming activity of gaining insight of public emotions. Now days opinion mining is being done on the large amount of data. One can obtain large amount of opinioned data from these social networking sites. Twitter is one among the most popular micro blogging site, it has huge amount of opinioned data [1].

Text mining or text analytics is the process of harvesting the meaning out of large amount of data. Sentiment analysis and opinion mining can be used interchangeably. It is the process of extracting sentiments, emotions, thinking, opinions out of data.

Sentiment Analysis is widely practiced to get insight about the feelings and attitude of the people on the internet; it is done to measure the performance of various brands, products etc. For instance: Macy's is a leading retail chain/store. It uses Sentiment analysis as one of the ways to improve customer experience and business growth.

II. THE WORLD OF SOCIAL MEDIA

In the world today, people are connected to each other via internet. Three decades ago, expressing opinions and making one's voice reachable on the fly to the world was a dream. Now with the advent of the social media, it has become a reality. In world today one cannot imagine their life without the social media; it is a growing phenomenon due to advances in information technology.

Virtual world has arisen not only as a communication media, but people also consider it as a necessary platform for expressing their views, opinions, sentiments on the various topics. People here not only connect by their blood relations or if acquainted with each other, but also people connect on social media based on the similar interests, hobbies, specialization etc. There are millions of active users of social media and users of social media are expected to increase up to 3 billion by 2021.

US, China, India has the largest social media users. It is known that only about 21% of the whole data produced is structured in nature, rest of the data is unstructured data. Most of the unstructured data is the textual data. It is mainly



produced by the various social networking sites. This data, if processed properly can be an information source and can be of great importance to us. Social media is used by both professional and non-professional users.

Tons of reviews or opinions produced and shared by users daily all over the world. If this data is refined and processed properly, then it can be used in various sectors for improving the various aspects of life.

A. Twitter Mining

Twitter mining is the hot topic these days, as it provides the important information which is used and applied in various fields. It is one of the major research areas. By using the various public APIs various tweets can be collected and analysed for research purpose. Through authenticated requests twitter APIs are established.

OAuth that is open standard authorization, which is required to get access of the protected data of the twitter. Users of the twitter come from every grounds of life. Its users vary from the celebrities to the normal users, company owners and representatives, politicians, also the country's presidents and prime ministers [14]. Therefore, it is possible to collect tweets of the persons from the different social and interest groups. Researchers find this information very useful to know about people's sentiments and opinions in defined number of words, political parties tend to get to know about whether people are in their support or not.

People use casual approach in writing tweets; therefore, the tweets has the noisy data, spelling mistakes, grammatical and punctuation errors. Therefore, analysing tweets is a tedious job. Through twitter we get a vast amount of data, but the data obtained is unstructured in nature. If that data is analysed, then more value can be added to it.

Text mining is the process of analysing the day to day produced data which is written in our natural language or spoken language. To accomplish this text mining make use of the natural language processing [16][17]. Text mining/analytics uses natural language processing to accomplish this. NLP is a component of the text mining which basically helps the machine to read the text and to derive high quality information from the text.

NLP uses the principles of the computer science and artificial intelligence to deal with the human language, also study the data and derive useful information from it. In general terms the text analytics is a process and NLP are the method to carry it out.

For data collection we use tweepy library to collect tweets This library allows Python to access the Twitter platform/database using its API. To install tweepy you can write the following command:

```
pip install tweepy
```

B. Sentiment analysis of tweets

In this, an attempt to gain an insight of the textual tweet posted on the twitter. For instance if we talk about two political leaders then, about the candidates for the prime minister's position of the two most popular parties prior to the election is made that is Candidate-1 and Candidate-2. It is known that abundance of the information can be gathered from the tweets, if proper processing is done then the information obtained can be used for decision making, management and political campaigns. Opinion mining can be performed on the twitter data to get insights of the sentiments of the public. This information is very useful for various organizations as well as the different political parties. It is known that feelings, emotions, sentiments are of great importance in human life. Mining such opinions is called sentiment analysis [3].

It is an emerging field which identifies the polarity, relevance and objectivity of the text. With the help of this, text can be classified into categories like positive, negative and neural. In this; reviews, blogs, microblogs written by the humans are examined, to get insight about the feelings, attitudes of the people on the internet for monitoring the performance of the political parties, governments, products, movies etc. Assigning labels of sentiments to a text, which tells about the polarity of the text is called Sentiment Analysis (SA). It can be classified into three levels: document level, sentence level or aspect level. By document level it is meant that the opinions are classified for the entire document. For this to happen, document should consist of the single topic.

There is no provision for the document containing multiple topics. In sentence level, all the sentences are examined to find their polarity, it could be either positive or negative, mixed opinions may or may not be considered for a sentence. Aspect level can also be known as entity level, in this all the aspects of the sentences are considered, meaning thereby in what context sentence talks about. Entity level approach is a detailed approach but it quite complicated to implement.

III. RELATED WORK

For data collection we use tweepy library to collect tweets Tweepy: This library allows Python to access the Twitter platform/database using its API. to install tweepy you can write the following command

```
pip install tweepy.
```

For data processing the most important we use textblob library TextBlob: It is a Python library for processing textual data. It helps in diving into common natural language processing (NLP) tasks such as part-of- speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. tweepy and NLP tool TextBlob



library to analyse the polarity, as well as the subjectivity of a tweet on the specified subject or topic. Textblob helps us to get the polarity and subjectivity of the tweet

Polarity: Defines the positivity or negativity of the text; it returns a float value in the range of “-1.0 to 1.0”, where ‘0.0’ indicates neutral, ‘+1’ indicates a very positive sentiment and ‘-1’ represents a very negative sentiment. Subjectivity: Defines the text on the basis that how much of it is an opinion vs how factual it is. where ‘0.0’ is very objective and ‘1.0’ is very subjective.

For generating word cloud , word cloud, pandas, numpy, matplotlib and pillow library has been used:

The numpy library is one of the most popular and helpful libraries that is used for handling multi-dimensional arrays and matrices. It is also used in combination with pandas library to perform data analysis.

For visualization, matplotlib is a basic library that enables many other libraries to run and plot on its base including wordcloud that you will use in this tutorial. The pillow library is a package that enables image reading.. Pillow is a wrapper for PIL - Python Imaging Library. We will need this library to read in image as the mask for the wordcloud. And then for showing the output there will be a web app that has been created using streamlit Streamlit is an open-source app framework is the easiest way for data scientists and machine learning engineers to create beautiful, performant apps in only a few hours! All in pure Python. And this app will be the interface.

IV. APPROACH AND DESIGN

A. Lexicon Based Approach for Sentiment Analysis

- The words, phrases, idioms, adjectives that have a semantic meaning is known as lexicon, this approach relies on the lexicons for classifying the data.
- In other words, the lexicon is composed of set of positive, negative or neutral. It requires a scoring function to score each of the sentences according to negative, positive words, for determining the sentiment polarity.
- There are numerous ways of making the lexicons, one such method is based on dictionary approach. It begins with the basic sentimental words and is made large by adding similar words and antonyms to it. Once it is made, it can be used to perform matching in the words in text document and the lexicon words.
- Here words are treated as numbers. Positive number implies the positive words (PW) and Negative number implies the negative words (NW). Polarity is assigned as +1, -1, 0 for positive, negative and neutral text respectively.
- Hiu and Liu developed an opinion lexicon for general purpose, which will classify English words into positive and negative words. There are about one thousand nine hundred sixty-seven positive words and four thousand seven hundred eighty-three negative words. Final Sentiment score of the tweets is calculated by subtracting the positive tweets from the negative tweets.

Zero score states that there are no opinion words in the text, whereas positive score implies positive polarity of the text and negative score implies negative polarity of the text.

This approach produces good results for opinion classification.

The aim of the paper is grounded in analyzing the sentiments of the public and make a web app to display it using various methods. We plan to have two tabs in the web app first being the “tweet analyser tab” and other one “generate twitter data”. Show recent tweets tab: it will show the first five tweets fetched from the account, generate world cloud tab: it collects all the keywords from the fetched twitter data and makes a cool collage of words and shows it, visualization of sentiment analysis tab: it shows bar graph of negative, positive and neutral tweets.

Tweet analyser then has following three option tabs:

Show recent tweets, Generate word cloud, visualization of sentiment analysis. Generate twitter data element will fetch first hundred tweet from the account and tell its polarity, subjectivity and analysis(negative, positive or neutral).



	Tweets	Subjectivity	Polarity	Analysis
0	Best wishes to the people o...	0.6000	0.6438	Positive
1	Government of India has dec...	0.2292	0	Neutral
2	हमारे यहां कहते हैं कि बेटी जब बोल...	0	0	Neutral
3	Greetings to the people of ...	0.5273	0.4182	Positive
4	Sharing a compilation of th...	0.0667	0	Neutral
5	पिछले 7 सालों में हमने सरकार और ...	0	0	Neutral
6	कोरोना के खिलाफ लड़ाई में बहुत ब...	0	0	Neutral
7	At a time when everybody wa...	0.0833	0.1000	Positive
8	Our Nari Shakti is at the f...	0.4732	0.1161	Positive
9	Group Captain Patnaik, like...	0.3069	0.1292	Positive
10	Dinesh Upadhyay Ji belongs ...	0.6667	0.3333	Positive

IV. CONCLUSION AND BENEFITS

The beauty of this works brings together text mining and sentiment analysis which otherwise are discussed as separate topics, this work has amalgamated the two into one. The approach followed in this paper was sound and systematic. It included analysis of tweet text before performing sentiment analysis on them. The data set used in paper is primary and adequate in size. The tools selected in this paper was based on intensive literature review. It is known that the tweet has no particular format, the only limitation imposed is with the length of the text. As the tweets were collected from the online source, hence contained no sentiment labels. The absence of the sentiment labels and the noisy nature of the tweets added complexity to the paper.

You can develop a more insightful, data-based marketing strategy. Understand your customers and audience, Measure your marketing campaign, Give extra boost to your customer service, Improve the quality of your products you can combine the feedback of customers and hence work on the product quality as to what is more liked, Better crisis management, Seeing what people are saying about you can help you identify potential crises and PR disasters and address them before they blow up, Boost sales revenue And of course, the be all and end all for any business, the benefits of sentiment analysis translate directly into increased sales revenue. By listening to and analysing comments on Facebook and Twitter, local government departments can gauge public sentiment towards their department and the services they provide, and use the results to improve services such as parking and leisure facilities, local policing, and the condition of roads. Universities can use sentiment analysis to analyze student feedback and comments garnered either from their own surveys, or from online sources such as social media. They can then use the results to identify and address any areas of student dissatisfaction, as well as identify and build on those areas where students are expressing positive sentiments.

REFERENCES

- [1] EFTHYMIOS KOULOUMPIS AND JOHANNA MOORE, IJCSI INTERNATIONAL JOURNAL OF COMPUTER SCIENCE ISSUES, VOL. 9, ISSUE 4, NO 3, JULY 2012
- [2] S. Batra and D. Rao, "Entity Based Sentiment Analysis on Twitter", Stanford University, 2010
- [3] Saif M. Mohammad and Xiaodan Zhu, "Sentiment Analysis of social media texts", 2014
- [4] Ekaterina Kochmar, University of Cambridge, at the Cambridge coding Academy Data Science, 2016
- [5] Manju Venugopalan and Deepa Gupta, "Exploring Sentiment Analysis on Twitter Data", IEEE 2015
- [6] Brett Duncan and Yanqing Zhang, "Neural Networks for Sentiment Analysis on Twitter", 2017
- [7] Afroze Ibrahim Baqapuri, "Twitter Sentiment Analysis: The Good the Bad and the OMG!", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011