DATA ANALYSIS ON LIVE SOCIAL MEDIA DATA

[1]Manohar M, [2]Sowmya M S , [3] Alok Kumar Pani
[1] Associate professor, [2] Assistant Professor, [3] Assitant professor
[1,3]Department of CSE, Faculty of Engineering, CHRIST (Deemed to be University) Bangalore,
[2] Department of ISE, School of Engineering, Jain University, Bangalore.

ABSTRACT

Natural Language Processing, Machine Learning and Artificial Intelligence are raisingadvancements. Their application opens incredible prospects and opportunities for the execution of computerized DMS (Decision Making systems).There are various Natural language processing techniques that are used to build applications and projects which can perform numerous tasks such as information retrieval, clustering, classification etc. Since the earlier decades people have been encountering exponential development in the utilization of online assets uniquely in web-based social networking and micro blogging sites such as Facebook, Twitter alongside some versatile applications like Whatsapp, hike etc. Many organizations consider their resources as an important source of marketing data. We consider theidea of micro blogs on which people post ongoing messages about their sentiments, feelings and assessments on an assortment of topics, converse current issues, grumble and express constructive notion for the items they use in everyday life. This data gives significant reaction by enabling them to additionally build up the upcoming age of their product. As of we know by ventures, the greater part of the study organizations is essentially taking choices in light of such sort of data got from web. Twitter is an online web application which encloses rich total of information that could be structured, semi-organized and unstructured. In this paper, a prediction model proposed which is made by performing sentiment analysis on twitter data by extracting tweets and categorizing them based upon their polarity.

Keyword: DMS (Decision Making systems), NLP (Natural Language Processing)

1. INTRODUCTION

According to [1] the objective of Natural Processing is to make the computer a fluent user of ordinary (human) language". The expansion in innovation utilization and the information that is gathered nowadays have constrained us to utilize more improved strategies to gather information. Natural Language Processing can resolve issues identified consistent employments in human-PC correspondence and upgrade the exactness and the procedures that is used to manipulate information in content arranging / text formatting. Here we introduce about the Sentiment analysis, Python and NLP are being explained. Sentiment analysis is one of the mostly used applications in natural language processing domain. Many people use the internet platform to express their feelings, emotions, sentiments, issues etc. The key role of sentiment analysis is to process this data and extract useful information which can be helpful in making smarter business decisions. It is the process of collecting and analyzing data based upon the individual's feelings. It is performed using various machine learning techniques, NLP strategies etc. Its emphases on recognizing whether a given piece of text is subjective or objective. If it is subjective, then whether it is negative, positive or neutral. Sentiment analysis is basically performed at document, phrase and sentence level. In document level, the synopsis of whole record is taken to know on what's more and to examine later. In phrase level investigation of expressions, a sentence is considered for checking the polarity. In sentence level, each sentence is arranged in specific class to give assessment. Sentiment analysis is domain centered, i.e., results of one domain cannot be applied to another domain. Sentiment analysis is connected in numerous genuine situations for instance to get audits of a film or any product, to get money related report of any company, for forecasts, promotion and so on. Twitter is a micro blogging stage where anybody can compose short type of message which is called tweets. The measure of information related with twitter is exceptionally huge. This information is unstructured and written in regular language. The venture is centered on Twitter assessment examination.

2. LITERATURE REVIEW

Opinion mining problem is a machine learning process that has been an enthusiasm to concentrate for the current years. Through the related works done, it is possible to take care of this problem. Despite the fact that few striking works have begun in this field, a completely customized and very proficient plan has not been displayed till now. This is because of the unstructured nature of natural language. The Dictionary of Natural language is extremely colossal that things end up being even hard. Various difficulties still happen in the field of machine learning. These issues must be embraced freely and those outcomes can be utilized to grow better techniques to do sentiment analysis. These days, there is an overall discourse with respect to ways & methodologies on the control of the information and the extraction of valuable data out of them. Machine learning procedures are associated with this exchange. Clever framework or intelligent systems that will gain from information and would have the capacity to anticipate or make proposals are required. Before we jump into more profundity about this quickly developing innovation, it is extremely imperative to show two verifiable logical assumes that have added to the machine learning and measurable learning systems. The first is Alan Turing. Turing was the immense donor in the field of counterfeit consciousness and machine learning of his chance. His impersonation diversion, broadly known as"Turning Test" has opened the uncountable possibilities that software engineering could offer. Uniquely, in the event that we consider that the entire world was endeavoring to recoup from Second World War, we can likewise consider him as an awesome identity that has put his own stone to the sciences and social development. Turing's amusement had three taking an interest subjects. Two subjects were people and one was the PC. One of the human had the part of the questioner. Moreover, the two was the qualification on who is the human and which is the PC after an ether a progression of inquiries utilizing print. The objective of the PC was to trick the investigative specialist and the objective of the second human was to convince the investigator that the other member is the machine. The amusement was one of the main trials on machine learning and counterfeit consciousness. The faultfinder on the consequences of Turing's amusement is out of the extent of this research work. The second individual and a logical identity we couldn't overlook, is Noam Chomsky. Chomsky's exploration field shifts from philosophical-political audits, to limited state machines and during that time he has affected numerous ages of individuals in all kind and zones he had been included. Concerning paper we unmistakable his work on the field of natural language processing and language modeling. He was the primary who considered limited state machines as a syntax arrangement and translated a finite state language as a language that have been created by the finite state grammar. A ton of advance has been done in the field of machine learning after the 1950s with numerous authentic logical figures in light of the fact that their work and they enthusiasm inspires until today. These days the assets and access to data is to a great degree simple and quick. Turing and Chomsky impacted in view of their enthusiasm to find and change the world. What they have through their exploration and eagerness is remarkable with respect to the learning of their opportunity and access to the assets.

Natural Language Processing (NLP) is imperative piece of data gathering conduct has dependably been to discover what other individuals think. With the expansion in number of individuals, the quantity of sentiment rich assets, for example, online survey locales and individual sites, chances to do different things has additionally be expanded. Along these lines the part of information is created. The information produced can be utilized and handled to get insights about publics intrigue and in this way the moves can be made. The sudden emission of action in the zone of conclusion mining and sentiment analysis, which manages the computational treatment of supposition, sentiment, and subjectivity in content, has subsequently happened in any event to some degree as an immediate reaction to the surge of enthusiasm for new frameworks that arrangement specifically with feelings as a top of the line question. This study covers strategies and methodologies that guarantee to straightforwardly empower opinion-oriented information-seeking systems. Our emphasis is on strategies that try to address the new difficulties raised by sentiment-mindful applications, when contrasted with those that are as of now show in more customary actuality based analysis. We incorporate material on summarization of evaluative content and on more extensive issues in regards to protection, control, and financial effect that the advancement of conclusion arranged data get to administrations offers ascend to.

[2] In this paper the authors have presented analysis for sentiment behavior of Twitter data. They has used naïve Bayes and fuzzy Classifier to classify the sentiment of a tweet i.e positive, negative or neutral. They present experimental evaluation of the dataset and classification results which proved that combined proposed method is more efficient in terms of Accuracy. [3] The author says that educational data mining is a prominent area to get information in educational fields using various data mining algorithms. In this paper they have used few learning algorithms to effectively rate the faculty belonging to an educational institute on the basis of feedback submitted by the students. The proposed model uses sentimental analysis and machine learning classifier algorithms for capturing the emotions from the student's feedback system. This model gives an accurate and efficient way to rate the faculty belonging to a particular educational institute. [4] The creators of this paper infer that the Social Networking locales are the assets which contain immense information. For instance, Twitter produces a huge number of bytes of the information. Comparatively Facebook, OLA, Uber they additionally create a large number of data. This information can be utilized for different purposes like business or social reason. Breaking down information from these interpersonal interaction Website is one of the new patterns for some, business related companied. Decision battles, audits about some new development can be successfully handle by utilizing sentimental analysis. Their proposed work in this paper essentially assesses sentimental analysis of Twitter information utilizing StandfordNLP (Natural Language Processing). [5] In this paper the author fundamentally discusses assessment mining. Author says that it is an art to trace the mood or feelings of people about a specific thing or a domain from a big set of opinions or reviews available online. In this work, a novel approach has been proposed which depends on SentiWordNet, which produces tally of score words into seven classifications, for example, solid positive, positive, frail positive, nonpartisan, feeble negative, negative and solid negative words for the supposition mining undertaking and assessed

utilizing machine learning calculations like Naive Bayes, SVM. [6] This paper portrays that the point is to dissect the sentimental impact of posts and afterward look at the outcome on various web-based social networking stages. Tremendous measures of posts are produced on social media platforms consistently. Public/People are especially inquisitive in finding the impact among them. Most analysts estimated the impact of a post through the quantity of answers it got. In any case, we don't know whether the impact is made emphatically or contrarily on different posts if their sentimental data isn't considered. In this paper, three research questions are raised and approaches are proposed for the measure of sentimental impact of posts. At last, a preparatory examination is outlined and done with some fascinating outcomes found. [7] This paper introduces a movie genre preference predictive model usable by small and medium-sized enterprises (SME's) who are in need of a data-based and analytical approach to stock proper movies for local audiences and retain more customers. In this paper they have used classification models to extract features from one thousand customers to predict their movie genre preference. In the implementation, a Gaussian kernel support vector machine (SVM) classification model and a logistic regression model were established to extract features from sample data. [8] This paper states that with the advent of social media and the exponential growth of information generated by online users, how can they find out some useful patterns from it which can help users to conclude something. Probabilistic matrix factorization is a method which can be used to handle massive amounts of data by learning low dimensional approximation matrices, but various works have ignored these relationships among users and resources. In this paper, a method is proposed to be based on tag weight of users and resources (TWPMF), which our use custom tags to more accurately identify the user interests and resource characteristics, so the influential neighbors are more accurately. The real-world data sets demonstrate that TWPMF algorithm can get more accurate rating predictions. [9] This paper expresses that the amount of information being made and prepared day by day has developed exponentially with the presentation of the web and web-based social networking. Once the information is accessible, it can be a colossal battle to discover something valuable from it. A standout amongst the most well-known uses for the substantial amounts of information is to make models to foresee the conduct or inclinations. One imperative use of forecast is anticipating budgetary results utilizing datasets. As a particular case, this investigation centers around the utilization of Twitter information gathered paving the way to a motion picture's opening end of the week to foresee its income through the span of every one of the opening ends of the week. Because of the absence of promptly accessible information, the information must be first assembled week after week utilizing Twitter's API and related outsider libraries. Development of the prescient model depends on a few machine learning calculations utilizing an arrangement of highlights got from client tweets. The outcomes demonstrate that their prescient model can be utilized to decide the achievement of motion pictures amid the opening end of the week by forecast the gross every day esteem. [10] This paper discusses the expanded enthusiasm for utilizing web-based social networking to anticipate social unrest. The endeavors have been made toward robotized unrest expectation, and afterward to recognize and channel them, which can be given to downstream clients to encourage analysis. In this paper they train a supervised classifier that can name Arabic language tweets as significant to unrest with high dependability. They look at the connection between preparing information size and execution and research approaches to upgrade the model building process while limiting expense. [11] This paper fundamentally discusses how identity separates a person with another. By knowing and understanding a people identity, numerous favorable circumstances can be acquired. Together with quick development of innovation, knowing a people identity should be possible consequently. Psychology research inquires about propose that specific identity qualities have connection with linguistic behavior. Upheld by the notoriety of web-based social networking, foreseeing people identity from their post wind up conceivable. Most existing investigates have done comparative approach in anticipating identity from online networking. [12] The paper briefly tells that the most powerful medium for communication among the individuals to share their valuable thoughts are Online Social Networks (OSNs). 'Twitter' is one of the most popular OSN rich with public data/tweets. In this paper author used Twitter Streaming API 'streamR' which is provided by 'R' statistical programming language, to extract the real-time tweets from Twitter. The tweet has many attributes which can be further analyzed to find most significant information about the Twitter user. We considered three attributes: screen-name, follower-count and friend-count. Twitter data is scaled up from gigabytes to petabytes and standalone system could not withstand or process this huge data due to hardware constraints. We used the prevailing parallel computing environment provided by 'Hadoop' with 'Python' programming language to analyze the Twitter users' whose follower-count and friend-count is less than 5000. We identified the user with maximum follower-count as the influential user who can contribute and maximize the information diffusion in Twitter. [13] This paper states that Sentiment analysis has been one of the most researched topics in Machine learning. The roots of sentiment analysis are in studies on public opinion analysis towards the beginning of twentieth century, however the flare-up of PC based sentiment analysis just happened with the accessibility of subjective content in Web. The task of producing compelling sentence model that catches both syntactic and semantic relations has been the essential objective to improve sentiment analyzers. This approach automates the whole process otherwise done using advance NLP techniques. It is a modular approach analyzing syntactic and context based relation from word level to phrase level to sentence level and then to document level. [14] This paper states that with the emergence of Web 2.0 and the development of social media platforms, more and more users are inclined to share their own opinions more comfortably on the Internet. Facing a large number of unstructured comments from social platforms, it is urgent to analyze and judge the tendency of emotion expressed in the text by Natural Language Processing. In this paper, the machine learning methods of sentiment analysis are described in detail. This paper introduces the popular sentiment analysis techniques from the perspective of machine learning technologies, including Support Vector Machine method, Naive Bayes method. Finally, the evaluation methods and challenges are given. [15] This paper talks about one of the most important services for the users using the internet is the Micro blogging which has evolved to be very important in today's life. A huge number of internet users share feelings on various parts of their life regularly in some of the most common sites, such as, Twitter, Facebook etc. Some of the politicians will be eagerly waiting to understand whether the individuals and the people are interested to support the respective politicians and the political parties.Thus analyzing the micro blogs and then getting the result out of it can be useful and the decisions can be made according to that. [16] This paper mostly talk about the share trading system exercises. As of late, stock market activities are getting to be needy greatly via web-based networking media communications to convey huge data for a broad number of clients. This obliges structures to scale speedily to suit the surge of new and in addition existing clients heading off to the proposals in view of data removed from web-based social networking. In this work, they proposed an approach for appointing scores to vocabulary things utilizing strategic relapse based relative scoring to address both the proposition quality and the structure adaptability.

In the proposed research work we investigate Twitter as information data source for open vocabulary identity forecast in Indonesia. They break down and look at three changed factual model and discover relationship about which identity qualities are connected with linguistic behavior.

Summary of the proposed research work is given below:

- Is social listening mostly on user comments and conversation?
- Are the comments or conversation on social media accessible?
- Is the social media data possibly beneficial for sentiment analysis?
- How can the tweets be explicitly characterized?
- Is the integration of tweets helpful for real - world sentiment analysis applications?

3. PROPOSED SYSTEM

Sentiment analysis is one of the most significant area of concern for grouping sentiment analysis from any given information .It can be achieved on any data by using numerous machine learning methods. For instance at whatever point an item is to be propelled by a particular organization, the clients might want to think about the item ratings, reviews and nitty gritty depictions about it. So sentiment analysis can help in breaking down marketing, advertising and for making new techniques for advancing the item.

Sentiment analysis is substitute for Data mining for Recognition of human behavioral contain from online data. Reviews are the supportive content for any industry to identify the response of the reviewer. So the problem statement is to perform sentimental analysis on different domain and study the way it works. Absolutely its ordering the extremity of given content at either record level , sentence level or expression level to choose whether the communicated assessment in an archive , a sentence or an element highlight/aspect/feature is positive , negative or impartial.

The main objective of the work is to perform sentiment analysis on different domains.To achieve this objective, we used text classifiers to perform live sentiment analysis i.e., Sentiment analysis on the tweets that are live streamed.

• A thorough study of existing techniques and libraries for performing data analysis in python.
• Collection of useful information/relevant data from Twitter by the use of Twitter API
• Pre-processing of thedata collected is performed so that it can be used for mining.
• Storing the data in relevant data-frame.
• Application of text classification techniques and analysis of data to obtain information.
First, we have to register our client application with Twitter for which we need consumer token and consumer secret.
• We use Tweepy: a twitter library for authentication and setup Twitter's API withour access keys.
• Using the API object we stream the tweets in real time.
• Then we perform the pre-processing on the tweets, so that they could be fit for mining.
• Once the tweet cleaning and organizing is done, we pass these tweets to text classifier (Textblob), which is a python text classifier.
• We then obtain the polarity values for the information we've passed and could get the result of sentiment of domain analyzed.

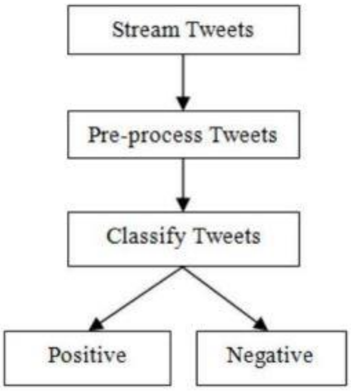


Fig. 1 System architecture

4. EXPERIMENTAL RESULT AND ANALYSIS

The results of sentiment analysis performed tells us the polarity of the tweets being extracted, i.e., the percentage of positive, negative and neutral tweets being present in the data obtained. Two types of text classifiers are being used. One is python's Textblob and another is Naive Bayes classifier. Textblob works on figuring out the polarity of sentences as discussed in earlier chapters. Naive Bayes classifier basically uses Bayes Theorem to calculate the probability that agiven article be an appropriate to a specific tag.

- Bayes Theorem:

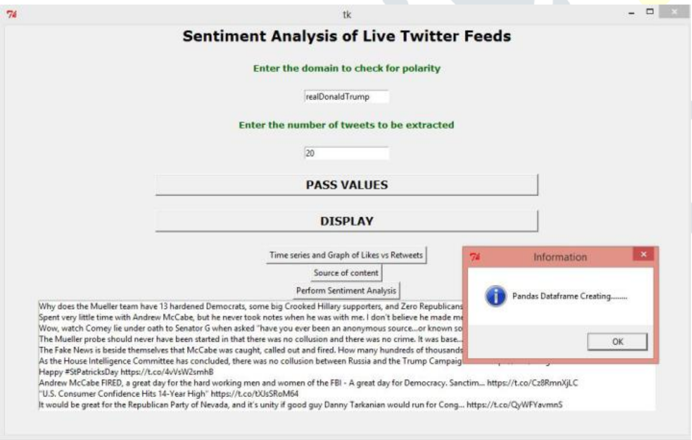
P(A/B) = (P(B/A) * P(A))/ P(B)

Where:

• P(A) is probability of given hypothesis H being true. This is also known as the prior probability.

• P(B) is the probability of the evidence (regardless of the hypothesis).

• P(B/A) is the probability of the evidence given that hypothesis is true.


tk
PASS VALUES
DISPLAY
Time series and Graph of Likes vs Retweets
Source of content
Perform Sentiment Analysis
Information
OK

Sentiment Analysis of Live Twitter Feeds

Enter the domain to check for polarity

realDonaldTrump

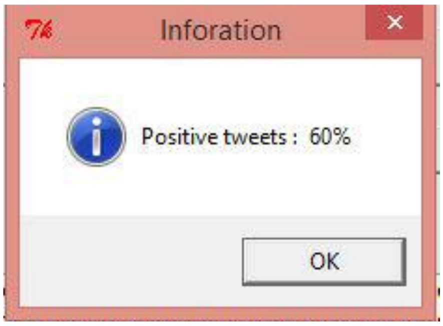
Enter the number of tweets to be extracted

20

Pandas Dataframe Creating........

```
Why does the Mueller team have 13 hardened Democrats, some big Crooked Hillary supporters, and Zero Republicans
Spent very little time with Andrew McCabe, but he never took notes when he was with me. I don't believe he made me
Wow, watch Comey lie under oath to Senator G when asked "have you ever been an anonymous source...or known so
The Mueller probe should never have been started in that there was no collusion and there was no crime. It was base...
The Fake News is beside themselves that McCabe was caught, called out and fired. How many hundreds of thousands
As the House Intelligence Committee has concluded, there was no collusion between Russia and the Trump Campaig
Happy #StPatricksDay https://t.co/4vVsW2smhB
Andrew McCabe FIRED, a great day for the hard working men and women of the FBI - A great day for Democracy. Sanctim... https://t.co/Cz8RmnXjLC
"U.S. Consumer Confidence Hits 14-Year High" https://t.co/tXJsSRoM64
It would be great for the Republican Party of Nevada, and it's unity if good guy Danny Tarkanian would run for Cong... https://t.co/QyWFYavmnS
```

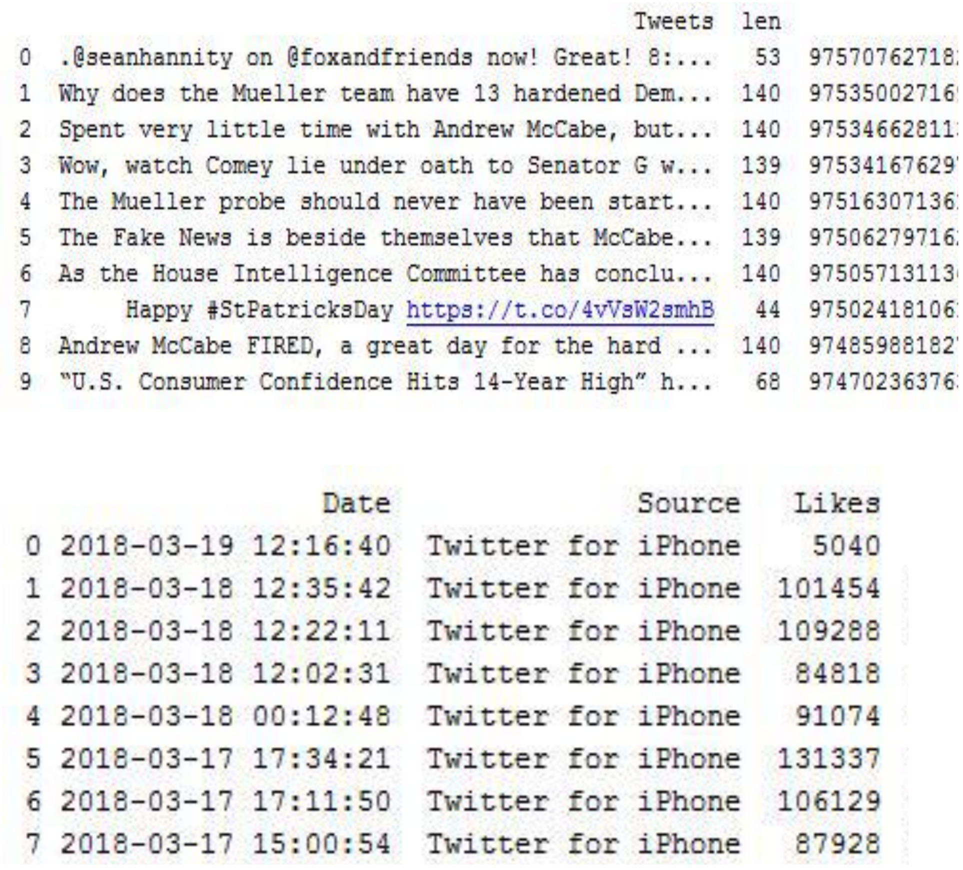
Fig. 2 Interface for the Python script



Positive tweets : 60%

Fig. 3 Sentiment result for Positive tweets

Both the classifiers gives us the results when performed sentiment analysis. Here two types of results are yielded. One is the analysis on User profile and being on any interested domain. Few snap shots of results being captured are displayed above in this page. Pandas data frame helped a lot in this project for data organizing and storing. As we are doing real streaming rather than storing the data a place after extraction, it helps in speeding up the process. Obtaining from a remote file is a time taking process compared to directly getting them from an advanced data structure.



```
                                            Tweets  len
0  .@seanhannity on @foxandfriends now! Great! 8:...   53  97570762718
1  Why does the Mueller team have 13 hardened Dem...  140  97535002716
2  Spent very little time with Andrew McCabe, but...  140  97534662811
3  Wow, watch Comey lie under oath to Senator G w...  139  97534167629
4  The Mueller probe should never have been start...  140  97516307136
5  The Fake News is beside themselves that McCabe...  139  97506279716
6  As the House Intelligence Committee has conclu...  140  97505713113
7        Happy #StPatricksDay https://t.co/4vVsW2smhB   44  97502418106
8  Andrew McCabe FIRED, a great day for the hard ...  140  97485988182
9  "U.S. Consumer Confidence Hits 14-Year High" h...   68  97470236376
```

```
                 Date              Source   Likes
0 2018-03-19 12:16:40  Twitter for iPhone    5040
1 2018-03-18 12:35:42  Twitter for iPhone  101454
2 2018-03-18 12:22:11  Twitter for iPhone  109288
3 2018-03-18 12:02:31  Twitter for iPhone   84818
4 2018-03-18 00:12:48  Twitter for iPhone   91074
5 2018-03-17 17:34:21  Twitter for iPhone  131337
6 2018-03-17 17:11:50  Twitter for iPhone  106129
7 2018-03-17 15:00:54  Twitter for iPhone   87928
```

Fig 4. Sample data set

5. CONCLUSION:

Sentiment analysis is used to identify people's opinion, attitude and emotional states. The views of people could be positive or negative. Commonly parts of speech are used as feature to extract the sentiment of the text. An adjective plays a crucial role in identifying the sentiment of the text from respective parts of speech. Sometimes when words having adverbs and adjectives together appear in the statement and sometimes there exists sarcasm in the statement, such cases becomes a difficulty for the algorithm to predict the opinion. Twitter is a massive source of data, which makes it more attractive for performing data analysis. We can perform analysis on as much data as we want but huge the data, more precise be the results. We have seen how analysis could be performed by real time streaming on users' profile, we can also form a word cloud which shows the words a person uses and could judge the behavior etc. The analysis that's performed for comparative purpose on different domains helps a lot in field of data analysis.For example,the waypeople reacted to different political parties, different people/celebrities,different electronicgadgets and what not.In this way sentiment analysis is really a great option forprediction purposes.It could also be used for financing, reviewing ,marketing, sales, stock prediction, election and many more.

REFERENCES

[1] Kristin M. Tolle,Hsinchun Chen," Comparing noun phrasing techniques for use with medical digital library tools", Journal of the American Society for Information Science banner, 2000.

[2] Ruchi Mehra, Mandeep Kaur Bedi,Gagandeep Singh,, Raman Arora, Tannu Bala, Sunny Saxena," Sentimental Analysis Using Fuzzy and Naive Bayes", International Conference on Computing Methodologies and Communication (ICCMC),IEEE, page 945-950, 2017.

[3] K. S. Krishnaveni ; Rohit R Pai ; Vignesh Iyer, "Faculty rating system based on student feedbacks using sentimental analysis", International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2017.

[4] Hase Sudeep Kisan; Hase Anand Kisan; Aher Priyanka Suresh, "Collective intelligence and sentimental analysis of twitter data by using StandfordNLP libraries with software as a service (SaaS)", IEEE International Conference on Computational Intelligence and Computing Research (ICCIC),2016.

[5] Shoiab Ahmed; AjitDanti, "A novel approach for Sentimental Analysis and Opinion Mining based on SentiWordNet using web data" Second International Conference on Inventive Communication and Computational Technologies (ICICCT) ,2018.

[6] Humera,Shaziya , G.Kavitha Raniah Zaheer, " Text Categorization of Movie Reviews for Sentiment Analysis " International journal of innovative research in science, engineering and Technology , vol. 4, issue 11, 2015.

[7] Haifeng Wang ; Haili Zhang "Movie genre preference prediction using machine learning for customer-based information" World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering Vol:11, No:12, 2017.

[8] Feng Xiong , YongJian Liu ; Qing Xie, " Recommendations Based on Collaborative Filtering By Tag Weights", 13th International Conference on Semantics, Knowledge and Grids (SKG), 2017.

[9] Steve Shim , Mohammad Pourhomayoun, " Predicting Movie Market Revenue Using Social Media Data" , IEEE International Conference on Information Reuse and Integration, 2017.

[10] Alan Mishler, Kevin Wonus, Wendy Chambers, Michael Bloodgood, "Filtering Tweets for Social Unrest" , IEEE 11th International Conference on Semantic Computing (ICSC), pages 17-23, 2017.

[11] Louis Christy Lukito, Alva Erwin, James Purnama, WulanDanoekoesoemo, "Social media user personality classification using computational linguistic" , 8th International Conference on Information Technology and Electrical Engineering (ICITEE), 2016.

[12] K Sailaja Kumar, D Evangelin Geetha, N. Nagesh, T V Sai Manoj, "Identify the influential user in online social networks using R, Hadoop and Python" , International Conference on Circuits, Controls, Communications and Computing (I4C), 2016.

[13] Anmol Chachra, Pulkit Mehndiratta, Mohit Gupta, "Sentiment analysis of text using deep convolution neural networks", enth International Conference on Contemporary Computing, IC3 2017.

[14] Peng Yang, Yunfang Chen, "A survey on sentiment analysis by using machine learning methods",IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2017.

[15] Siddu P. Algur, Rashmi H. Patil , "Sentiment analysis by identifying the speaker's polarity in Twitter data" , International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017.

[16] Kunal Bhargava , Rahul Katarya, "An improved lexicon using logistic regression for sentiment analysis", International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), 2017.