International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 03 Issue: 06 | June -2019 ISSN: 2590-1892

Real Time Sentiment Analysis on Twitter Data Using Natural Language Processing

Poornima G H[1], Vidyarani H J[2]

[1] Student, Dr.Ambedkar Institute of Technology, Bangalore, poornimagh07@gmail.com

[2] Assistant Professor, Dr.Ambedkar Institute of Technology, Bangalore, vidyagau@gmail.com

ABSTRACT

In today's world Social networking sites like Twitter, Facebook are the great source of communication for internet users, so it becomes an important source for understanding the opinions, views or emotions of the people on products, celebrities and also some places. In this paper we introduced a generalized form of sentiment analyzer for opinion mining on the twitter data, in which retrieve the tweets from the twitter the twitter based on the user requirement by using the tweepy and Twitter API keys and that tweets can be preprocessed by using the NLTK libraries and then it can be preprocessed by using the NLTK libraries and then it can be sentiment analysis and classified into positive, Negative and neutral tweets by using the textblob and result can be representation by using Pandas and Matplotlib and by using machine learning techniques find the accuracy of the sentiment analyzer.

Keyword: Opinion mining, Data analytics, Sentiment Analysis, TextBlob, Tweepy.

I.INTRODUCTION

With the large amount of increase in the web technologies, the No of people expressing their views and opinion via web are increasing. This information is useful for everyone like business, government and individual with 500+ million tweets per day, twitter is becoming a major source of information. Twitter [1] is a micro blogging site, which is popularly known for its short messages known as tweets. It has basically limit of 140 characters but from 2017 its doubled .Twitter has base of 240+ million active users and hence it is a useful source of information. The users often discuss their personal views on various subjects and also on current affairs via tweets. Out of all popular social Medias like Facebook, twitter, Google. Twitter‟s audience varies from regular users to celebrities, president. Therefore it is possible to collect tweets of users from twitters and internet groups.

1.1 Sentiment analysis:

Sentiment analysis uses data mining techniques to extract and capture data for analysis in order to get the Subjective opinion of a document or collection of documents,

Sentiment analysis is an old concept but technologies are changed day by day, the first sentiment analysis development using Php and then using java, R, etc. Now the sentiment analysis is one of the trending technologies in the data science and the natural language processing.

Sentiment analysis can be used for business, politics and also public sector. In marketing field companies use to understand the customers feelings towards products and brand. In political field it is used to keep of political view.

1.2 Proposed System

In the proposed system, we will streaming of tweets from twitter using Twitter Api based on the query. That can be developed using python language. The collected tweets will be subjected to preprocessing and it can be analyzed, categorized and also stored in a database. The sentiment will be represented in graphical manner and getting accuracy of the system by using machine learning technique such as SVM, Maximum Entropy, Naïve Bayes for getting the pression, recall and fmeature by using the nltk package.

The proposed system is more effective than existing one. This is because we will be able to know how the statics determined from the representation of the result.

Advantages of proposed system

1. In the proposed system, we will retrieve tweets from twitter using Twitter API based on the query.
2. The proposed system can be developed using python language
3. The collected tweets will be subjected to preprocessing and it can be analyzed, categorized and also stores in a database.
4. Generalized form of sentiment analysis, in which varieties of data can be analyzed.

II. RELATED WORK

Over the years many sentiment classification systems have been developed using many techniques. These systems have

Been implemented using many machine learning algorithm. In [2] The propose a twitter sentiment classification System for us airline service analytics which is designed by taking the data from the kaggle data set. The author have been used word embedding‟s and document vector technique used for data processing and machine learning classification algorithm can be used for classification of the tweets and find the accuracy. In [3] the Author proposes a study of twitter Sentiment Analysis using machine learning Algorithms on python which is explained on twitter sentiment analysis data extraction from the twitter by using WEKA mining tool and sentiment classifiers which are machine learning classifiers are used for sentiment classification. According to the survey the existing sentiment analysis systems are developed by using the traditional technology and old techniques, many systems are used existing data set for the experiment.

III. METHODOLOGY

3.1 SYSTEM MODEL

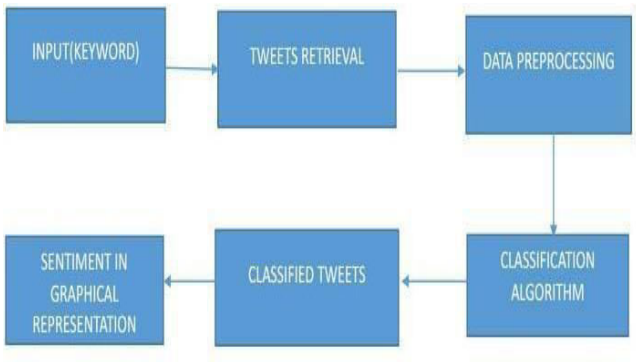

Figure 3.1 System Model

3.2 IMPLEMENTATION

a. GETTING TWITTERAPI KEY

1. Open twitter application web page and login, if necessary
2. Supply the necessary required fields, accept the terms Of Service, and solve the CAPTCHA
3. Submit the form
4. Go to the API Keys tab, there you will find your Consumer key and Consumer secret keys.
5. Collected data from the Twitter

b. COLLECTED DATA FROM THE TWITTER

By using tweepy can stream the data from twitter by using consumer key and consumer secrete keys? Extract the tweets about the particular subject given by the users. Tweepy is a python library module is used to stream live tweets directly from twitter in realtime.Tweepy is an open sourced hosted by GitHub and enables python to communicate with twitter platform and use its API

c. DATA PREPROCESSING

Raw tweets scraped from twitter generally result in a noisy dataset. This is due to the casual nature of people‟s usage of social media. Tweets have certain special characteristics such as retweets, emotions, user mentions, etc. which have to be suitably extracted. Therefore, and raw twitter data has to be normalized to create a dataset which can be easily learned by various classifiers. We have applied an extensive number of preprocessing steps to standardize the dataset and reduce its size. We first do some general preprocessing on tweets which as follows.

- Convert the tweets to lower case.
- Strip spaces and quotes (“and”) from the ends of tweets
- Replace 2 or more spaces with a single space.

d. SENTIMENT ANALYSIS

Sentiment Analysis is the process of computationally determine whether a piece of text is positive, negative or neutral. Sentiment Analysis can be done by using the textblob. Textblob is python library it is the module of an NLTK in which the textblob calculate the polarity of the twitter data. The polarity is in between -1.00 and +1.0.If the polarity of the tweets is +1.0 that tweet is labeled as a positive tweet, the polarity of the tweet is -1.0 that tweet labeled as a negative and the polarity of the tweet is 0.0 that can be labeled as neutral.

```
For Tweets in tweets:
    Print (Tweets. Text)
    Analysis=Textb1ob.Textb1ob (Tweet. text)
    po1arity +=analysis.Sentiment.Po1arity
    If analysis.Sentiment.po1arity ==0:
    Neutra1+=1
    Elif analysis.Sentiment.po1arity>0.00:
    Positive+=1
    Elif analysis.Sentiment.po1arity<0.00:
    Negative+=1
```

e. CLASSIFICATION

After analyzing the sentiment of each tweet the tweets can be classified into positive tweets, Negative tweets, neutral tweets and also calculate the percentage of the positive tweets negative and neutral tweets

```
Def calculate Percentage (a, b):
    Return 100*float (a)/float (b)
    positive-calculate Percentage(positive,numberOfTweets)
   negative=calculatePercentage(negative,numberOfTweets)
   neutral=calculatePercentage(neutral,numberOftweets)
```

f. GRAPHICAL REPRESENTATION

Graphical representation is the new concept introduced in the sentiment analysis, in which we using the pandas for the creating a data frame for the tweets and Matplotlib can be used for represents the sentiment analysis in the form of pie chart.

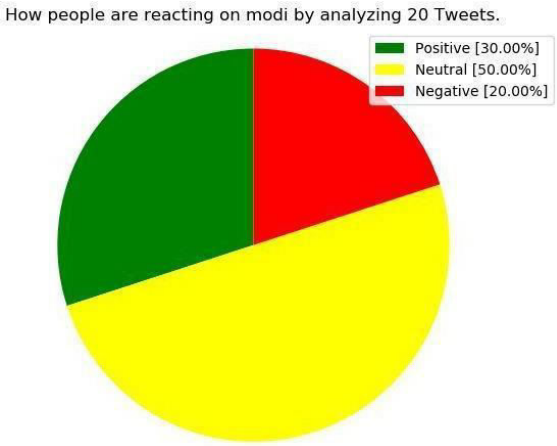

Figure: 3.2: Graphical Representation

g. CLASSIFIERS

SVM

SVM, also known as support vector machines, is a non-probabilistic binary linear classifier. For a training set of points (xi , yi) where x is the feature vector a y is the class, we want to find the maximum margin hyper plane that divides the points with yi =1 and yi =-1.

The equation of the hyper plane as follow

$$w.x - b = 0$$

We want to maximize the margin, denote by Y

$$\max \Upsilon \, , s.t \, . \forall i \, , \, \Upsilon \leq yi \, (\, w. \, xi + b)$$

NAÏVE BAYES

Naïve Bayes is a simple model which can be used for text Classification. In this model, the class $\hat{c}$ is assigned to a tweet t, where

$$\hat{c} = \arg\max P \, (c \mid t)$$

$$P \, (c \mid t \,) \alpha P \, (c) \prod_{}^{n} P \, (fi \mid c)$$

MAXIMUM ENTROPY

Maximum Entropy classifier model is based on the Maximum entropy classification principle. The main idea behind it is to choose the most uniform probabilistic model that maximum entropy, with given constraints. In binary classification problem like the ones we are addressing, it is the same as using 1ogistic regression to find a distribution over the classes.

The model represented by

$$P_{ME} (c \mid d, \lambda) = \frac{\exp[\sum i \lambda i fi \, (c, d \,)]}{\sum_{c} \exp[\sum i \lambda i fi \, (c, d \,)]}$$

Here, c is the class; d is the tweet and λ is the weight vector.

IV. RESULT AND DISCUSSION

a. Twitter API key

Getting the twitter API key by login through the twitter API keys and getting these twitter API key

Consumer Key='prfhQ8yGvV2HFZqez'
Consumer Secret='9tzd6nQLqTc48JfVK3LXg0D5M8w3T6ztZnV8sKGiXUIdZ'
Access Token=„89063637377GjevBxjFLbug6npdr3S4KdOnfxHclCZ"
Access Secret='TUKKLKyzoPcMXmdvLAIu8wyOxA0iAXqnDCy6uARSghss6'

b. Collecting the twitter data

Collecting the live twitter data by using these twitter API keys and also for accessing these keys using a tweepy package

.

	Tweets	User	ser_follower	User_location	rt_count	tweet_date
0	RT @ajaishukla: Air Chief Marshal Modi claims his	Pandit Ji	120		3864	2019-05-12 14:16:42
1	RT @theskindoctor13: "Sir remember u had put d	ANKUR	23	New Delhi, Inc	1500	2019-05-12 14:16:40
2	RT @dhruv_rathee: This meme is more scientifica	non linear	40		782	2019-05-12 14:16:40
3	RT @dhruv_rathee: This meme is more scientifica	kashif palawkar	116		782	2019-05-12 14:16:39
4	RT @dhruv_rathee: 3 Conclusions I draw from thi	Bibhu	1317	Bhubaneswar,	5922	2019-05-12 14:16:37
5	RT @dhruv_rathee: This meme is more scientifica	Rahul	94	Jaipur, India	782	2019-05-12 14:16:37
6	RT @dhruv_rathee: 3 Conclusions I draw from thi	Tabish khan تاج خان	4		5922	2019-05-12 14:16:37
7	RT @TajinderBagga: . @MamataOfficial Didi, Com	Chowkidar Joy Thompson	107	Kottayam-Mal	1708	2019-05-12 14:16:36
8	RT @ShekharGupta: In 2015, Time hailed Modi on	Parry Ravindranathan	1648	Hong Kong	32	2019-05-12 14:16:35
9	New post added at Mumbai Press News - Latest N	Mumbai Press	225	Mumbai, India	0	2019-05-12 14:16:35
10	RT @TilottamaChakr1: @MakrandParanspe @narc	Sushant Shrivastava	11		3	2019-05-12 14:16:33
11	RT @milindkhandekar: Modi's Balakot secret: I th	mafzal	25	New Delhi, Inc	690	2019-05-12 14:16:33
12	RT @RoshanKT_1: @narendramodi First Time Vot	Chowkidar अ कु र सि ह प	1635	Chandauli-Var	2	2019-05-12 14:16:33
13	RT @dhruv_rathee: When Time criticized Manmo	#11GloriousYrsOfVivianDsena	472	Mumbai, India	4309	2019-05-12 14:16:32
14	RT @dhruv_rathee: This meme is more scientifica	Meer Nadeem Khan	189	Bengaluru Sou	782	2019-05-12 14:16:32
15	RT @Vishj05: Modi: I suggested that the clouds co	Libid one out ®	228	India	516	2019-05-12 14:16:32
16	Why 'award wapsi' brigade silent on Alwar rape: I	CHOWKIDAR PR HARI KRISHN	18		0	2019-05-12 14:16:32
17	So that's the benchmark of congress. Excellent!!	A M Gomekar	84	Nagpur	0	2019-05-12 14:16:30

Figure 4.1: Example of twitter data

c. Sentiment analysis

Each tweet can be classified as a positive tweet, negative tweet and also neutral. And each tweet can be labeled by this classification.

# Tweets	sentiment
@Resistadelphia1 @ProudResister They want a theocracy, not the democracy this nation was founded on. Their beliefsâ€¦ https	neutral
The World Igbo Congress has concluded a preliminary session for its upcoming convention slated for August 30 to Sepâ€¦ https://	neutral
RT @kylegriffin1: Schiff: "The attorney general of the U.S. misled the country about an investigation implicating the president. Th	positive
RT @BriannaWu: We give lip service to motherhood, but fail at childcare access, equal pay, higher ed affordability and parental leave.	
Thâ€¦	negative
RT @SafetyPinDaily: House Democrat subpoenas six years of Trump tax returns | Guardian https://t.co/B8A4S4JIpx	neutral
RT @Indopolity: Something big is cooking in Congress ecosystem. In an unprecedented development, CJI Gogoi decides to presid	positive
RT @Lawrence: Exactly 1 Republican member of Congress understands Trump tariffs are sales taxes paid in America by Americans	positive
RT @vanessacvenus: A Dangerous Trend: Muslims Running for Office https://t.co/96TKNoJhTq This was written last year. We now	negative
@realDonaldTrump @EdRollins @LouDobbs You have an AG who has committed contempt of Congress. Why are you trying toâ€¦	neutral
RT @politico: House Intelligence Committee Chairman Adam Schiff said Congress should consider legislation that would make it	negative
RT @StephenKing: Bill Cohen, on Trump: â€œHe is basically thumbing his nose at Congress and saying, â€˜I donâ€™t recognize yo	neutral
RT @MichaelArt123: Adam Schiff: We have to proceed in Court. The judge in the Oversight Cmte appeal has said we will expedite	neutral
RT @NEWS9TWEETS: #Congress President @RahulGandhi claims 'love will triumph over #BJP's hatred'. Which party according to y	positive
RT @Billbrowder: Good to see US Secretary of State Pompeo taking tough on Russia. To prove his sincerity, he should submit to C	positive
RT @krassenstein: Three things America needs to get rid of:	

Figure 4.2: Example of sentiment analysis

d. Graphical representation

After the sentiment analysis of the each tweet calculate the percentage of the positive, negative and neutral tweet and then this sentiment analysis can be represented in the form of pie chart

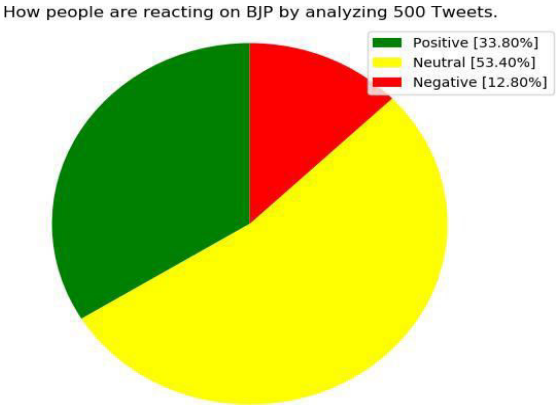

Figure 4.3: Example of Graphical representation

e. Accuracy

By using the machine learning classification algorithm find the accuracy of this sentiment analysis system and compare between the different algorithms. The below table shown the result.

	Accuracy	Precision	Recall	fmeasure
SVM	0.804	0.804	0.912	0.883
NaïveBayes	0.712	0.808	0.712	0.877
MaxEnt	0.696	0.801	0.696	0.666

Table 4.1 example of compare with algorithm

V. APPLICATIONS

a. Election Result Prediction

This sentiment analyzer can be used for election result prediction, now a days most of the people are expressing their opinion in social media, based on that data predicting the election result and also getting the people opinion based on the candidate, party and state. The example shown in below. In which compare with the two party.

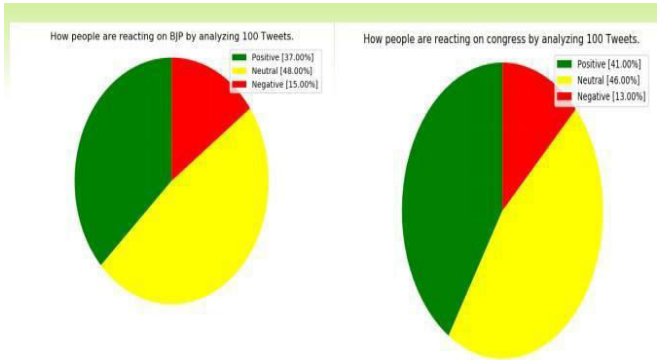

Figure 5.1: Election Prediction

b. Product Decision making

In sthis application the customer getting the information about the product by using this sentiment analyzer. Now a days online shopping is the very common and also the customer more believe about the reviews of the product. And compare with the different company product. One example for this application shown below.

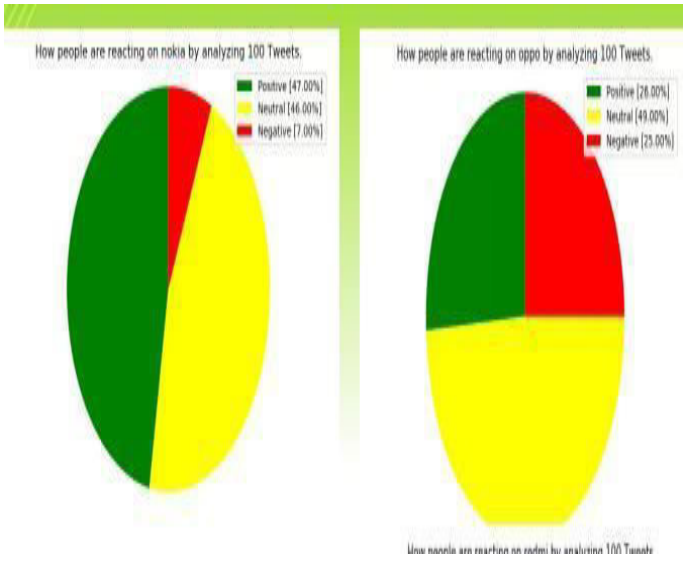

Figure 5.2: product decision

VI. CONCLUSION AND FUTURE ENHANCEMENT

Now days the people are more interact with internet, most of the peoples are getting the information about some topic through the social media. This sentiment analyzer give chance to user can getting the information about specific topic and sentiment analysis and give the accurate result of the product, people or company.

In this paper we introduced only for twitter data in the future developing a real time sentiment analysis system for other social media.

REFERENCES

[1] https://en.wikipedia.org/wiki/Twitter

[2] AnitaRane and Dr.AnandKumar "Sentiment Classification System of Twitter Data for US Airline service" 42nd IEEE International Conference on Computer Software &

Bhumika Gupta and Monika Negi, KanikaVishwakama,

Goldi "Study of twitter Sentiment analysis using Machine Learning Algorithms on Python" International Journal of Computer Applications(0975-8887)Volume 165 –No.9,May 2017

https://dev.twitter.com/apps/new
https://ww.nltk.org/modules/nltk/classify/maxent.html