



## Sentiment Analysis on Twitter Dataset

Abhigyan Tiwary\*, Dr.M Kumar\*\*

\* Ph.D Scholar, Mewar University Gangrar Chittorgarh, Rajasthan, INDIA, abhigyantiwary@gmail.com

\*\*Research Supervisor, Ph.D Scholar, Mewar University Gangrar Chittorgarh, Rajasthan, INDIA, prof.mkumar@gmail.com

### ABSTRACT

Now days the growth of social websites, blogging services and electronic media contributes huge amount of user give messages such as customer reviews, comments and opinions . Sentiment Analysis is important term of referred to collection information in a source by using NLP, computational linguistics and text analysis and to make decision by subjective information extracting and analyzing opinion, identifying positive and negative reviews measuring how positively and negatively an entity( public ,organization, product) is involved. Sentiment analysis is the area of study to analyze peoples reviews, emotion, attitudes and emotion from written languages. We examine sentiment analysis on Twitter data. This paper reports on the design of a sentiment analysis, extracting a vast amount of tweets. Prototyping is used in this development. Results classify customers’ perspective via tweets into positive and negative, which is represented in a pie chart and html page.

**Keywords-** Twitter, sentiment, opinion mining, social media, natural language processing

### INTRODUCTION

Nowadays, the age of Internet has changed the way people express their views, opinions. It is now mainly done through blog posts, online forums, product review websites, social media ,etc. Nowadays, millions of people are using social network sites like Facebook, Twitter, Google Plus, etc. to express their emotions, opinion and share views about their daily lives. Through the online communities, we get an interactive media where consumers inform and influence others through forums. Social media is generating a large volume of sentiment rich data in the form of tweets, status updates, blog posts, comments, reviews, etc. Moreover, social media provides an opportunity for businesses by giving a platform to connect with their customers for advertising. People mostly depend upon user generated content over online to a great extent for decision making. For e.g. if someone wants to buy a product or wants to use any service, then they firstly look up its reviews online, discuss about it on social media before taking a decision. The amount of content generated by users is too vast for a normal user to analyze. So there is a need to automate this, various sentiment analysis techniques are widely used.

Sentiment analysis (SA) tells user whether the information about the product is satisfactory or not before they buy it. Marketers and firms use this analysis data to understand about their products or services in such a way that it can be offered as per the user’s requirements. Textual Information retrieval techniques mainly focus on processing, searching or analyzing the factual data present. Facts have an objective component but, there are some other textual contents which express subjective characteristics. These contents are mainly opinions, sentiments, appraisals, attitudes, and emotions, which form the core of Sentiment Analysis (SA). It offers



many challenging opportunities to develop new applications, mainly due to the huge growth of available information on online sources like blogs and social networks. For example, recommendations of items proposed by a recommendation system can be predicted by taking into account considerations such as positive or negative opinions about those items by making use of SA.

### **SENTIMENT ANALYSIS**

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It's also referred as subjectivity analysis, opinion mining, and appraisal extraction. The words opinion, sentiment, view and belief are used interchangeably but there are differences between them.

- Opinion: A conclusion open to dispute (because different experts have different opinions )
- View: subjective opinion
- Belief: deliberate acceptance and intellectual assent
- Sentiment: opinion representing one's feelings

### **TWITTER SENTIMENTAL ANALYSIS**

Social networks is a rich platform to learn about people's opinion and sentiment regarding different topics as they can communicate and share their opinion actively on social medias including Facebook and Twitter. There are different opinion oriented information gathering systems which aim to extract people's opinion regarding different topics. The sentiment-aware systems these days have many applications from business to social sciences. Since social networks, especially Twitter, contains small texts and people may use different words and abbreviations which are difficult to extract their sentiment by current Natural Language processing systems easily, therefore some researchers have used deep learning and machine learning techniques to extract and mine the polarity of the text. Some of the top abbreviations are FB for Facebook, B4 for before, OMG for oh my god and so on. Therefore sentimental analysis for short texts like Twitter's posts is challenging

### **IMPLEMENTATION**

The approach to extract sentiment from tweets is as follows:

1. Start with downloading and caching the sentiment dictionary
2. Download twitter testing data sets, input it in to the program.
3. Clean the tweets by removing the stop words.
4. Tokenize each word in the dataset and feed in to the program.



5. For each word, compare it with positive sentiments and negative sentiments word in the dictionary. Then increment positive count or negative count.
6. Finally, based on the positive count and negative count, we can get result percentage about sentiment to decide the polarity.

```
#connect all libraries
library(twitterR)
library(ROAuth)
library(plyr)
library(dplyr)
library(stringr)
library(ggplot2)
#connect to API
download.file(url='http://curl.haxx.se/ca/cacert.pem', destfile='cacert
reqURL <- 'https://api.twitter.com/oauth/request_token'
accessURL <- 'https://api.twitter.com/oauth/access_token'
authURL <- 'https://api.twitter.com/oauth/authorize'
consumerKey <- '_____' #put the Consumer Key from Twitter Applic
consumerSecret <- '_____' #put the Consumer Secret from Twitt
Cred <- OAuthFactory$new(consumerKey=consumerKey,
                          consumerSecret=co
                          requestURL=reqURL
                          accessURL=accessU
                          authURL=authURL)

Cred$handshake(cainfo = system.file('CurlSSL', 'cacert.pem', package =
save(Cred, file='twitter authentication.Rdata')
load('twitter authentication.Rdata') #Once you launch the code first ti
registerTwitterOAuth(Cred)
#the function of tweets accessing and analyzing
search <- function(searchterm)
{
#access tweets and create cumulative file
list <- searchTwitter(searchterm, cainfo='cacert.pem', n=1500)
df <- twListToDF(list)
df <- df[, order(names(df))]
df$created <- strftime(df$created, '%Y-%m-%d')
```

Figure 1-Extract Sentiment from Twitter in ‘R’

To connect to Twitter and query latest tweets, we need to create an account on twitter and define an application. Users need to go to the apps.twitter.com/app/new and generate the api keys.



Application Management

## Create an application

**Application Details**

**Name \***  
  
Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

**Description \***  
  
Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

**Website \***  
  
Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

**Callback URL**  
  
Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth\_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Figure 2- Twitter Application Management

## RESULT

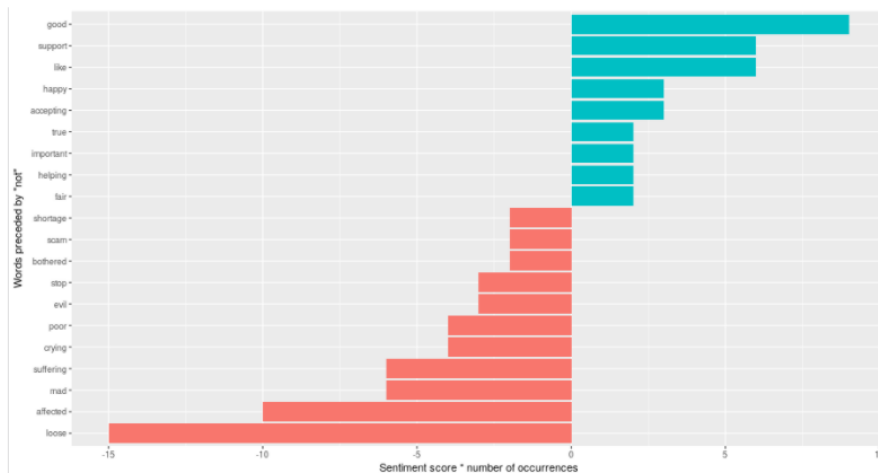


Figure-3 Sentiments score of twitter datasets



Figure-4 Word cloud of Demonetization



For the implementation of our dissertation use data collecting mechanism, Python’s API for Twitter – ‘Tweepy’ was used. It was imported in the python script. The python script written by the researcher for extracting twitter’s posts using Twitter API related to ‘Demonetization 2016’ is as follows.

The accuracy of the proposed “SVM, Naive Bayes’ classifier is found to be the highest.

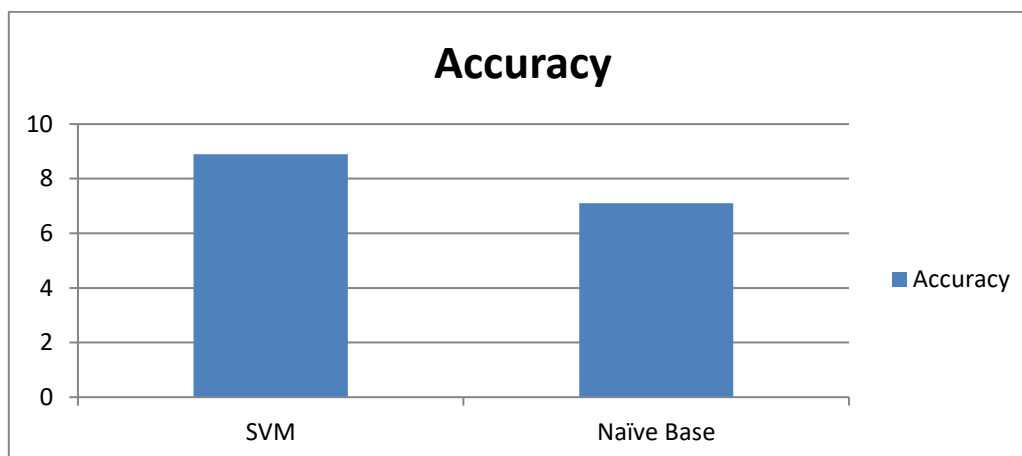


Table: 1 Performance Evaluations of SVM and naïve Base

The accuracy of classification of the sentiment for proposed approach is highest. It means that using the ‘SVM’ approach, 86.03% of the tweets in test data are correctly classified as showing the actual sentiment of the users towards the event ‘Demonetization in India 2016’.

## CONCLUSION

Sentiment Analysis provides huge number of opportunities by uncovering the opinions and views from the unstructured Twitter data set. To conclude, this research has explained that an efficient sentiment analysis can be performed on an event, Demonetization in India 2016. Throughout the continuation of this research various data analysis tools were applied to gather, clean, mine and determine review from the dataset. This analysis can help them to spot a positive turn in viewer’s opinion of their brand image. Uncovering positive trends early on can permit them to make educated decisions. It is shown in this research the approach of supervised machine learning classifier ‘SVM’ & Naïve Base so SVM has a major effect on the overall accuracy of the analysis. This approach has an accuracy of around 86.03% for classification. The simulation of the given technique was presented in ‘R’ language. With machine learning approach constantly are being developed and bettered, amazing amount of computational power becoming willingly available, and huge amount of data being uploaded to social media sites daily which lead sentiment analysis a standard practice for marketing and feedback.

## REFERENCES



1. Andrea Esuli and Fabrizio Sebastiani “SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining” 2006. In Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, IT. Forthcoming.
2. Haseena Rahmath P(2014) ”Opinion mining and sentiment analysis -challenges and applications”, International Journal of Application or Innovation in Engineering & Management, Volume 3, Issue 5, pp 401-403, May 2014
3. SMEUREANU, Cristian BUCUR “Applying Supervised Opinion Mining Techniques on Online User Reviews” Informatica Economică vol. 16, no. 2/2012
4. Chang, CH, Kayed, M, Girgis, MR & Shaalan, KF 2006, „A Survey of Web Information Extraction Systems“, IEEE T. Knowl. Data En., vol. 18, no. 10, pp. 1411-1428.
5. Hong, JL 2010, „Deep Web Data Extraction“, Proceeding of the IEEE International Conference on Systems Man and Cybernetics (SMC). Istanbul, pp. 3420-3427.
6. Jain, A, Jain, S, Shukla P & Bandiya, H 2012, „Towards Automatic Detection of Sentiments in Customer Reviews“, Int. J. Inform. Sci. Tech., vol. 2, no. 4, pp. 103-111.
7. Kamal, A & Abulaish, M 2013, „Statistical Features Identification for Sentiment Analysis using Machine Learning Techniques“, Proceeding of the International Symposium on Computational and Business Intelligence (ISCBI). New Delhi, pp. 178-181
8. Liu, B 2012, „Sentiment analysis and opinion mining“, Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167.
9. Miao, Q, Li, Q & Zeng, D 2010, „Mining fine grained opinions by using probabilistic models and domain knowledge“, Proceeding of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). Toronto, ON, vol. 1, pp. 358-365.
10. Ion Smeureanu & Cristian Bucur 2012, „Applying Supervised Opinion Mining Techniques On Online User Reviews“, Informatica Economică, vol. 16, no. 2, pp. 81-91.
11. Vu, TT, Pham, HT, Luu, CT & Ha, QT 2011, „A Feature-based Opinion Mining Model on Product Reviews in Vietnamese“, Studies in Computational Intelligence, Springer-Verlag, Berlin, Heidelberg, pp. 23-33.
12. Padmaja, S & Fatima, SS 2013, „Opinion Mining and Sentiment Analysis–An Assessment of Peoples“, Belief: A Survey. International Journal.
13. Vinodhini, G & Chandrasekaran, RM 2012, „Sentiment analysis and opinion mining: a survey“, International Journal, vol. 2, no. 6.
14. Buche, A, Chandak, D & Zadgaonkar, A 2013, „Opinion Mining and Analysis: A survey“, arXiv preprint arXiv:1307.3336.
15. Liu, B 2012, „Sentiment analysis and opinion mining“, Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167.