Real Time Twitter Sentiment Analysis For Product Reviews

Priyanka Pateriya[1], Ratish Agrawal [2], Santosh Sahu [3], Sachin Goyal [4]

[1234] SOIT, RGPV Bhopal

priyanka.pateriya.26@gmail.com, ratish@rgtu.net, santoshsahu@rgtu.net, sachingoyal@rgtu.net

Abstract— Customer Opinions play a very crucial role in daily life. When we have to take a decision, opinions of other individuals are also considered. Now-a-days many of web users post their opinions for many products through blogs, review sites and social networking sites. Business organizations and corporate organizations are always eager to find consumer or individual views regarding their products, support and service. In ecommerce, online shopping and online tourism, it is very crucial to analyze the good amount of social data present on the Web automatically therefore, it is very important to create methods that automatically classify them. Opinion Mining sometimes called as Sentiment Classification is defined as mining and analyzing of reviews, views, emotions and opinions automatically from text, big data and speech by means of various methods. In this paper we are fetching real time reviews from the social site twitter and apply various text mining techniques to preprocess the data and than apply an machine learning approach through which we can use naïve bayes classification algorithm to classify the text into various emotions and polarities.

Keywords—opinion mining, sentiment analysis, e-commerce, text mining, twitter, naive bayes, SVM.

I.INTRODUCTION

Sentiment Analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall tonality of a document. In recent years, the exponential increase in the Internet usage and exchange of public opinion is the driving force behind Sentiment Analysis today. The Web is a huge repository of structured and unstructured data. The analysis of this data to extract latent public opinion and sentiment is a challenging task.

The analysis of sentiments may be document based where the sentiment in the entire document is summarized as positive, negative or objective. It can be sentence based where individual sentences, bearing sentiments, in the text are classified. SA can be phrase based where the phrases in a sentence are classified according to polarity. Sentiment Analysis identifies the phrases in a text that bears some sentiment. The author may speak about some objective facts or subjective opinions.

It is necessary to distinguish between the two. SA finds the subject towards whom the sentiment is directed. A text may contain many entities but it is necessary to find the entity towards which the sentiment is directed. It identifies the polarity and degree of the sentiment. Sentiments are classified as objective (facts), positive (denotes a state of happiness, bliss or satisfaction on part of the writer) or negative (denotes a state of sorrow, dejection or disappointment on part of the writer). The sentiments can further be given a score based on their degree of positivity, negativity or objectivity.

Opinion mining involves analyzing opinions, sentiments or mentality of the writer from the written text. Online opinions have indirect influence on the business of several ecommerce sites. Those sites market their products and the web users go through the reviews of the product before buying that product. Many organizations utilize opinion mining systems to track customer reviews of products sold online.

II. LITERATURE REVIEW

Sentiment analysis or opinion mining is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has gain much attention in recent years. In this paper, we aim to tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis. A general process for sentiment polarity categorization is proposed with detailed process descriptions. Data used in this study are online product reviews collected from Amazon.com. Experiments for both sentence-level categorization and review-level categorization are performed with promising outcomes. At last, we also give insight into our future work on sentiment analysis.

Text mining, additionally spoken as text data processing, is that the method of extracting attention-grabbing and non-trivial patterns or information from text documents. It uses algorithms to remodel free flow text (unstructured) into information that may be analyzed (structured) by applying applied mathematics, Machine Learning and tongue process (NLP) techniques. Text mining is Associate in Nursing evolving technology that permits enterprises to know their customers well, and facilitate them in redefining client wants. As e-commerce is changing into a lot of and knowledgeable, the quantity of client reviews and feedback that a product receives has adult chop-chop over a amount of your time. For a well-liked quality, the quantity of review comments will be in thousands or maybe a lot of. This makes it tough for the manufacturer to scan all of them to create Associate in Nursing aware call in rising product quality and support. once more it's tough for the manufacturer to stay track and to manage all client opinions. this text tries to derive some significant info from quality reviews which can be employed in enhancing quality options from engineering purpose of read and helps in rising the support quality and client expertise.

The customer review is important to improve service for company, which have both close opinion and open opinion. The open opinion means the comment as text which shows emotion and comment directly from customer. However, the company has many contents or group to evaluation themselves by rating and total rating for a type of services which there are many customer who needs to review. The problem is some customers given rating contrast with their comments. The other reviewers must read many comments and comprehensive the comments that are different from the rating. Therefore, this paper proposes the analysis and prediction rating from customer reviews who commented as open opinion using probability's classifier model. The classifier models are used case study of customer review's hotel in open comments for training data to classify comments as positive or negative called opinion mining.

In [2], Analytics companies develop the ability to support their decisions through analytic reasoning using a sort of maths and mathematical techniques. Thomas Devonport in his book titled, "Competing on analytics: The new science of winning", claims that an enormous proportion of highperformance companies have high analytical skills among their personnel. On the other hand, a recent study has in addition conspicuous that over fifty 9 of the organizations do not have information required for decision-making. Learning "Data Analysis with R" not exclusively adds to existing analytics info and methodology, but in addition equips with exposure into latest analytics techniques in addition as prediction, social media analytics, text mining on. It provides an opportunity to work on real time info from Twitter, Facebook & amp; various social networking sites.

Twitter is one of most popular social networking site where people are expressing their views, opinion and emotions liberally. These tweets are recorded and analysed to mine emotions of people related to a terrorist attack (Uri attack). Present study retrieve tweets about Uri attack and find emotions and polarity of tweets. To mine emotions and polarity in tweets, text mining techniques are used. Approximately 5000 tweets are recoded and pre-processed to create a dataset of frequently appearing words. R is used for mining emotions and polarity.

III PROBLEM DEFINITION

The existence and the constructive stability of the trade and business is dependent on establishing a competitive dominance through effective and aggressive marketing strategies. With the prolific amount of information that is continuously being made available through the electronic media, the web users are unable to take advantage of these resources due to the lack of appropriate tools to utilize. Due to this it has become strenuous for web users to access relevant information productively. Furthermore, the significant increase in the number of websites puts forward a challenging task to organize the contents of the websites to cater to the needs of the users. An effective and broadly used remedy to this is the „Information filtering" method applied to manage abundant information flow.

Text mining [7] help an organization derive potentially valuable business insights from text-based content such as word documents, email and postings on social media streams like Facebook, Twitter and LinkedIn. Data mining or Text mining plays a important role in decision making because through these mining techniques we can analyse the data and on the basis of result we can take a decision.

Sentiment classification is a technique to focus on the sentiments or opinions expressed in an article or conveyed orally. The term sentiment includes emotions, conclusions, behavior and others. In this paper, the work concentrates on human readable text writing on the e-commerce sites.

IV PROPOSED WORK

Machine learning approaches are quite popular in the sentiment analysis domain. These approaches involve tokenization and preprocess the text and then classify the text into various emotions and polarities [09]. The aggregated sum of these scores determines the sentiment behind the text. It is generally classified as positive, negative or neutral. There are many machine learning classification algorithm are present like SVM, Random forest, Naïve Bayes and in this paper we are using Naïve Bayes classification because its gives 81% accuracy as compared to SVM who give 67% accuracy. So Naïve Bayes classification algorithm give better accuracy as compared to other classification algorithm.
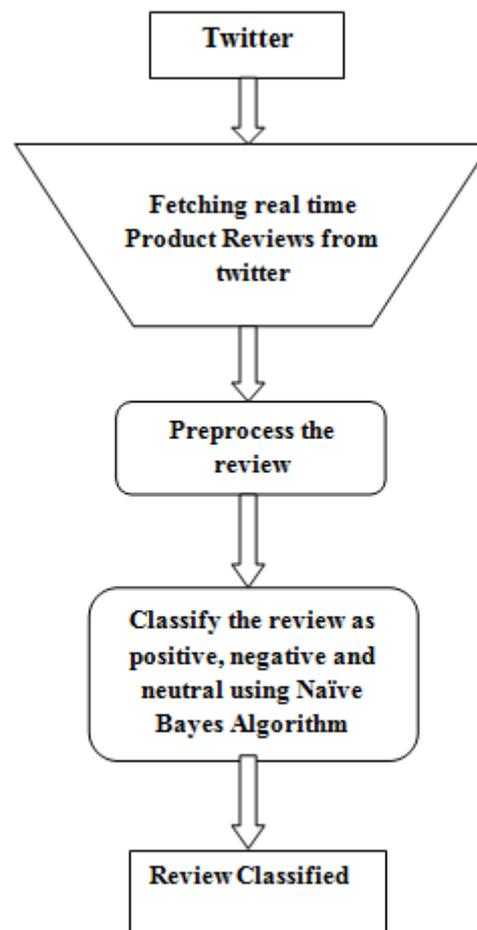
Figure 1. Flow Diagram

VII EXPERIMENTAL & RESULT ANALYSIS

All the experiments were performed using an i5-2410M CPU @ 2.30 GHz processor and 4 GB of RAM running Windows. After that we can install r base core on windows and Rstudio and than we are fetching real tweets and after that performing Sentiment analysis on that collected tweets. So, to achieve this we are going to follow the following methods:

➤ Collecting Tweets review.
➤ Preprocess the review text.
➤ Classify the review.

Collecting Tweets on product :-

For collecting tweets we need a r package called twitteR and ROAuth package to authenticate user consumer and token keys.

```
> library(RColorBrewer)
>
> library(twitteR)
> library(ROAuth)
>
> consumer_key<-'VE12EJOT2niBwUONTt7Q5sA4u'
> consumer_secret<- 'J5fOrVLmldAafwS4Pxm3eRyB537MMrrsXPdayANnpLyalfpCsY'
> access_token<- '772457331795726337-6IHz8xax4nKXgYfGy7qnZvYmayoOv41'
> access_secret<- 'xRcOOuVc7nRp2Wmq55Z2lzTCzYiDu2daRJNbX1ScOCcHn'
>
>
> setup_twitter_oauth(consumer_key, consumer_secret, access_token,access_secret)
[1] "Using direct authentication"
>
> some_tweets = searchTwitter("starbucks", n=1500, lang="en")
|
```

For consumer key and access tokens we need to create a twitter app through which we can generate our twitter keys and put into twitter_oauth and than we are storing to collects tweets on starbucks with frame size of 1500 and stored into some_tweets variable.

Preprocess the tweets text:-

For performing text mining on tweets we need a package called tm package which internally consist natural language processing NLP package. Figure 2 shows the tweets text which is fetch from twitter and these text consist an number, urls, so we need to preprocess these text by using text mining techniques.

Figure 2. Text review fetch from twitter

We are preprocessing twitter text and tm package consist many function through which we can remove numbers, punctual, and special characters coming from text. And first we are converting all the text into lower case to remove noise and we can also eliminate url's coming from text. Figure 3 show the preprocess text by using tm package.
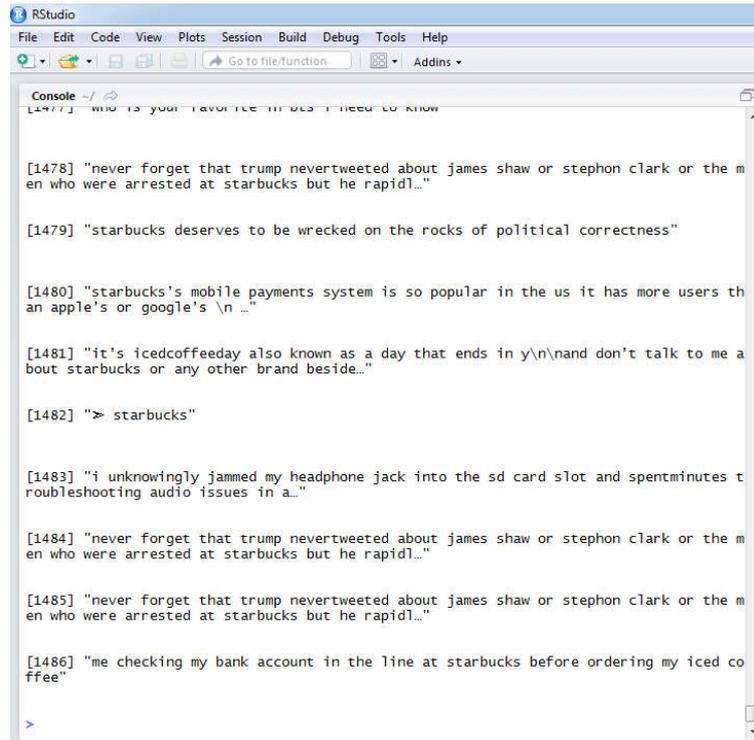
Figure 3. Pre-process text

After preprocessing the data we are applying naive bayes classfier to classify the text into various sentiment emotions and sentiment polarities. The naive bayes algorithm classifies the text into various emotions shown in figure 4.
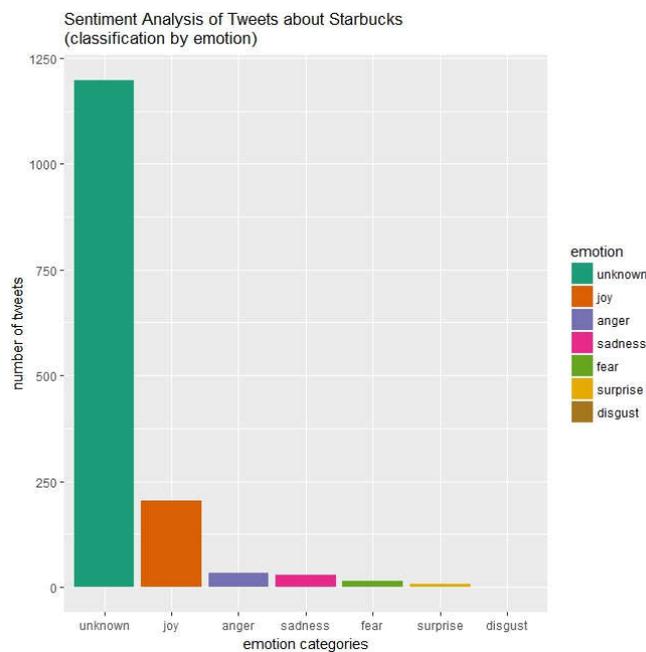

Figure 4. Classification by emotions

We can also classify the sentiment polarity of tweets text and which is shown in figure 5.
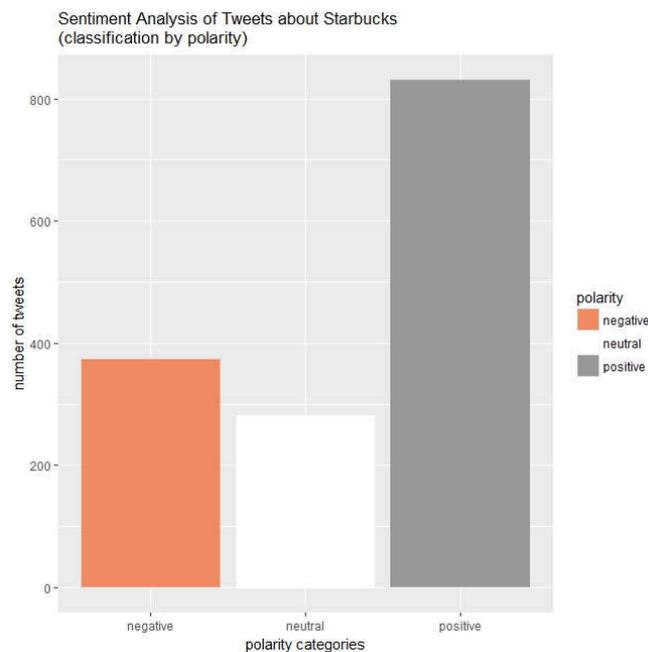

Figure 5. Classification by Polarity

VI   CONCLUSION

Opinion mining has become a fascinating research area due to the availability of a huge volume of user-generated content in review sites, forums and blogs. Opinion mining has applications in a variety of fields ranging from market research to decision making to advertising. With the help of opinion mining, companies can estimate the extent of product acceptance and can devise strategies to improve their product. Individuals can also use opinion mining tools to make decisions on their buying by comparing competitive products not just based on specifications but also based on user experience and public opinions. In this paper we are using naïve bayes classification algorithm to classify the review and we also calculate the accuracy of these classification algorithm and its give 81% accuracy . By using these we are classify the text into various emotions and polarity.

REFERENCES

[01] Xing Fang* and Justin Zhan, "Sentiment analysis using product review data" in Fang and Zhan Journal of Big Data (2015) 2:5 DOI 10.1186/s40537-015-0015-2.

[02] Chandrasekhar Rangu, Shuvojit Chatterjee, Srinivasa Rao Valluru, "Text Mining Approach for Product Quality Enhancement" in IEEE 2017.

[03] Shruti Kohli, Himani Singal, "Data Analysis with R" in 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing.

[04] Wararat Songpan, "The Analysis and Prediction of Customer Review Rating Using Opinion Mining" in 978-1-5090-5756-6/17/$31.00 ©2017 IEEE SERA 2017, June 7-9,2017, London, UK

[05] Dipak R. Kawade, Dr.Kavita S. Oza, "Sentiment Analysis: Machine Learning Approach" in International Journal of Engineering and Technology (IJET), Vol 9 No 3 Jun-Jul 2017.

[06] Porter M.F, Snowball: A language for stemming algorithms. 2001.

[07] Ning Zhong, Yuefeng Li, and Sheng-Tang Wu, "Effective Pattern Discovery for Text Mining", IEEE Transactions on Knowledge And Data Engineering, Vol. 24, No.1, January 2012.

[08] Timothy P. Jurka, Loren Collingwood, Amber E. Boydstun, Emiliano Grossman, and Wouter van Atteveldt- RTextTools: A Supervised Learning Package for Text Classification.

[09] Menaka S, Radha N- Text Classification using Keyword Extraction Technique, International Journal of Advanced Research in Computer Science and Software Engineering-Dec 2013

[10] K Nirmala,M Pushpa, Feature based text Classification using Application Term Set,International Journal of Computer Applications (0975 – 8887) Volume 52– No.10, August 2012

[11] Lokmanyathilak Govindan Sankar Selvan," A Framework for Fast-Feedback Opinion Mining on Twitter Data Streams", IEEE 2015.

[12] T. K. Das , D.P. Acharjya & M. R. Patra, " Opinion Mining about a product by Analyzing Public Tweets in Twitter ", ICCCI- 2014, Jan 03-05, 2014.