Recognizing Contextual Polarity in Net neutrality using Sentimental Analysis

Dr C Manju

Assistant Professor,Department of Computer Science
Kanchi Mamunivar Center for PostGraduate Studies, Pondicherry,India
manjuc76@gmail.com

Abstract— **Net Neutrality is all about the neutrality in the web space. This means, no special favour to be shown to any data in the internet for whatsoever reasons.. Internet's most powerful tool, the Social media has become an essential component of our daily life. Social media helps people, who had never been heard, heard. Anyone can freely express his/her mind without feeling intimidated. This paper deals with data from one of the most popular social media, Twitter, for understanding public perceptions and misperceptions of net neutrality. A Sentimental analysis on tweets related to net neutrality is done country wise and overall during the period 2016-2017. The classification is done using machine learning technique. Processing of tweets data is done to find sentiments of people and polarity of their tweets which helps us to determine the attitude of the people around the world.**

Keywords- Sentimental Analysis; Net Neutrality; Naive Bayes Classifie;rLexicons;Polarity

I. INTRODUCTION

Sentimental Analysis can be defined as a contextual understanding of a sentence expressed in a social media or opinion mentioned by a person in form of a text. It is a part of Natural Language processing and detection of attitude of people or a society on a particular issue. It is a part of subjectivity analysis which includes review mining, opinion mining, appraisal extraction. The sentimental analysis can be in sentence level, document level and future level.[1]

Feedback is a mechanism of expression of our views or attitude of people whatever thing that is happening in and around us. It will be there in a launch of new product, new policy decisions of the Government, new events by tweets, Face books or Instagrams. We can say these as opinion retrieval system in social media.[2][3]

Net neutrality is a government mandated policy for all internet service providers to take all information in the internet to all users without any discrimination in terms of platform, website, content, equipment. No special favour to be shown to any data in the internet for whatsoever reasons. A predominant opinion about net neutrality is that an extremely popular device like internet, if remains neutral, can be the most useful device that human race has ever invented. The practice of certain internet service providers allowing certain data free of cost and slowing down some data or selling some data for a price are all considered to be against net neutrality. In the face of tough competition between ISPs, many have resorted to such practices to attract more customers. Though to some extend it helps certain section of the customers, to the larger extend, major part of the customers stand to lose.21st century's 17 years saw revolutionary changes in the way humans live. Internet played a big role in these developments . Unbridled availability of data across any kind of network was the key to the success of internet and it helped people bring the world to their finger tips. With the advent of competition among the internet service providers, to gain upper hand over or to beat the competition, each ISP tried to attract customers with luring offers. The most attractive among the offers was to offer social media portals free of cost. This in turn resulted in the ISPs deliberately throttling other data in the otherwise free world of internet.

In this paper analysis of the public opinion about net neutrality is done using twitter data is done to find its sentiments and polarity . Analysis is done using the tweets of people among various geographic regions , thereby getting the opinion of people all over the world in this context.

II. SENTIMENTAL ANALYSIS APPROACHES

Sentimental analysis can be done in different methods. There are various approaches to evaluate sentimental analysis in a text. These are Machine Learning Approach, Rule based approach and Lexical based approach.[4]

A. Machine Learning Approach

Machine learning approach can be of two types. They are supervised learning and unsupervised learning. The supervised learning algorithms can be used to determine sentiments in the text. The most commonly used supervised learning technique include

- Support Vector Machine

SVM algorithm is a 'simple' linear classification/regression algorithm[6]. It tries to find a hyper plane which separates the data in two classes as optimally as possible.

Here, as optimally as possible means that as much points as possible of label A should be separated to one side of the hyper plane and that of label B to the other side, while maximizing the distance of each point to this hyper plane.[10]
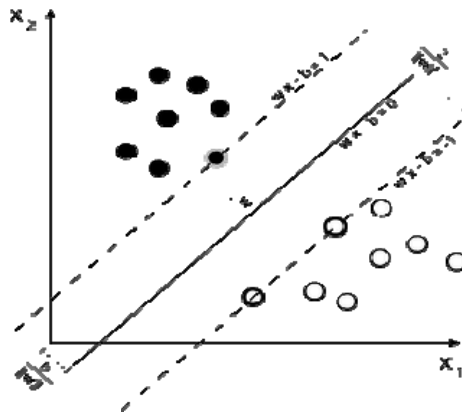

Figure 1 SVM Classifier

For the polarity classification there are three classes (positive, neutral, negative) and for sentiment Classification we can have even more than that The classical SVM system requires that the dataset is linearly separable, i.e. there is a single hyper plane which can separate the two classes. For non-linear datasets a Kernel function is used to map the data to a higher dimensional space in which it is linearly separable. The classical SVM system is a binary classifier, meaning that it can only separate the dataset into two classes. To deal with datasets with more than two classes, usually the dataset is reduced to a binary class dataset with which the SVM can work.

- **Naïve Bayes:**

Naïve Bayes [8] classifiers are studying the classification task from a Statistical point of view. Bayes theorem can be written as

Posterior probability = conditional probability · prior probability /evidence   -(1)

Using Bayes' rule, this posterior probability can be rewritten as:

$P(\omega j \mid xi) = P(xi \mid \omega j) \cdot P(\omega j) P(xi)$ -(2)

Since the marginal probability is equal for all classes, it can be disregarded and the equation becomes: $P(xi \mid \omega j)$ be the probability of observing sample xi given that belongs to class ωj .

If $P(+ \mid x) \geq P(- \mid x)$ classify as positive , else classify as negative.   ---(3)

- **N Gram Approach**

N gram is a set of co-occurring words in a text. N gram is used for developing features for supervised machine learning model such as decision tree; naïve bayes. N gram tokenization is generally used before removing the stop words ,and is extracted to unigrams, bigrams and combination of unigrams and bigrams. As bigram contains pair of words. First word can be a stop word and stop word normally contains more information. Like unigram "Good" is positive word. But bigram "Not good" is caring negative meaning. In most of the research word, unigram and bigram are highly recommended for sentiment analysis.[5][6].

III.   METHODOLOGY

Sentimental analysis is the process of finding sentiments from the text. Net neutrality is the policy of non favoritism towards data among Internet Service Providers. It is the government policy to provide neutrality of usage of internet among various users. In this paper we are going to have a sentimental analysis of opinion of the people among on net neutrality The data to be analysed is taken from twitter API interface.

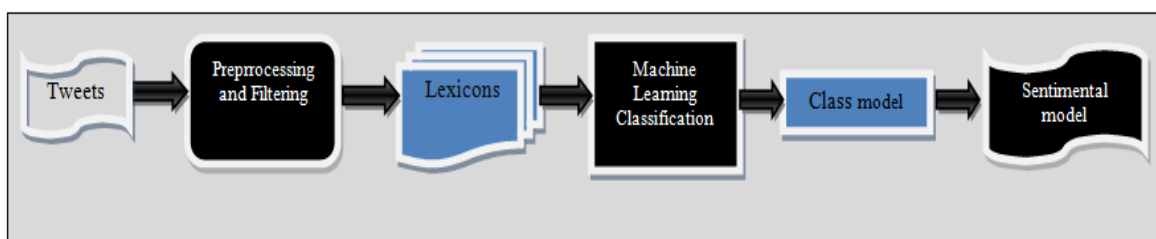The process can be diagrammatically represented as below.

Figure 2:Model Representation

A. Extracting Tweets

The tweets are extracted from Twitter application program interface. Twitter authenticated objects are createdand then tweets are extracted [9] using RTSUDIO.

B. Pre processing and Filteration

The tweets posted by the people will be highly unstructured and also non grammatical. They contain a lot of meaningful and meaningless vocabulary. Extensive usage of acronyms will also be there. The data received from tweets is converted into CSV file and necessary cleaning of text is done by removing unnecessary data such as tags, whitespace, YRL numbers , stop words etc. Filtration is done using lexicon based approach which uses positive and negative words. It is the process of extracting sentiment from the tweets. It is used for findingpositive and negative words and thereby producing sentiments.[7]

C. Classification

Classification is the process of identifying to which of a set of categories, a new observation belongs on the basis of a dataset containing observations (or instances). In this paper supervised machine learning approach-Naïve Bayes classification is used for classification purposes,. It calculates the posterior probability of a class which is based on words distributed in the tweets. It will predict the probability of whether the given feature set belongs to a particular label of content. Finally it will determine sentiments in the text. It is assumed that lexical resources can be evaluated using semiwordnet and wordnet which specifies positivity and negativity of text and semantic net which calculates MPQA in the range of positivity to negativity. The classification process eventually finds emotion, sensitivity, attention and pleasantness of data.
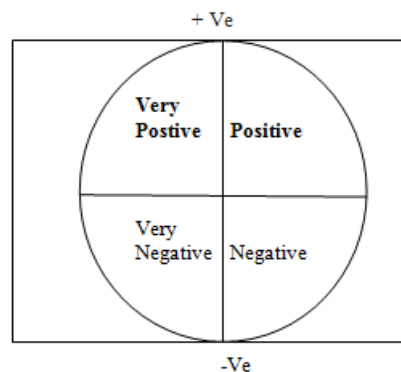

Figure 3:classification of lexicons

D. Sentimental and Polarity calculation

From the lexicons, various sentimental words in positivity and negativity will be categorised and sentimental score found.

Sentimental Score = $\sum$+ve words -$\sum$-ve words.

Polarity computation involve finding positive, negative and neutrality of the lexicons of the text tweeted by the people on specified subject.

IV. EXPERIMENTAL RESULT ANALYSIS

RSTUDIO tool software is used for the design and documentation of overall datamining process. We have implemented our model using the Naïve Bayes classifier for finding sentiments and polarity of the data.

A. Extracting Tweets

The process involve collection of data from twitter API by searching tweets from different geographic locations around the world.The study was done on data collected from India ,Brazil, European Union, United Kingdom ,China, France, Germany, Singapore, Spain and Solvenia. The tweets can be extracted from R tool using the following instructions.

library("bitops", lib.loc="~/R/win-library/3.4")

detach("package: twitter", unload=TRUE)

library("twitteR", lib.loc="~/R/win-library/3.4")

library("twitteR", lib.loc="~/R/win-library/3.4")

library("ROAuth", lib.loc="~/R/win-library/3.4")

consumer_key<-'edS0ckU41tL34DZach11Lt0a3'

consumer_secret<-'NYQMUR3hzyPVKkqRhlUXqyXslA8MSnHKrYwCZze7aIuLlEnag9'

access_token<-'946333528261668865-EKoyAloDbKd49SvUlcW0sLOVO32z4RP'

access_secret<-'8qcVhrEul38gF60jrLdrujv5cN7Ra5RaBJ6XsZrd5sHsP'

setup_twitter_oauth(consumer_key,consumer_secret,access_token,access_secret)

B.   Pre-processing and Filtration

 Pre-processing is an important in all data mining applications which transform raw data into formatted data. Here It involves tokenizing, pruning and stemming. Tokenization process splits the text of a document into sequence of tokens. The splitting points are defined using all non letter characters. This results in tokens consisting of one single word (unigrams). The Pruning here involves removing all URLs, hash tags, correction of spellings, punctuation, removal of stop words, removing non English texts etc. After tokenisation and pruning, we will have a set of tokens in terms of positivity and negativity. These tokens are filtered to get most frequently occurring tokens to find out sentiments of tweets generated by people. Stemming is a technique that is used to find the root or stem of a word. The filtered token set undergoes stemming to reduce the length of words until a minimum length is reached[10]. This Machine Learning and Applications resulted in reducing the different grammatical forms of a word to a single term. After this process the result generated can be represented diagrammatically as in figure 4,
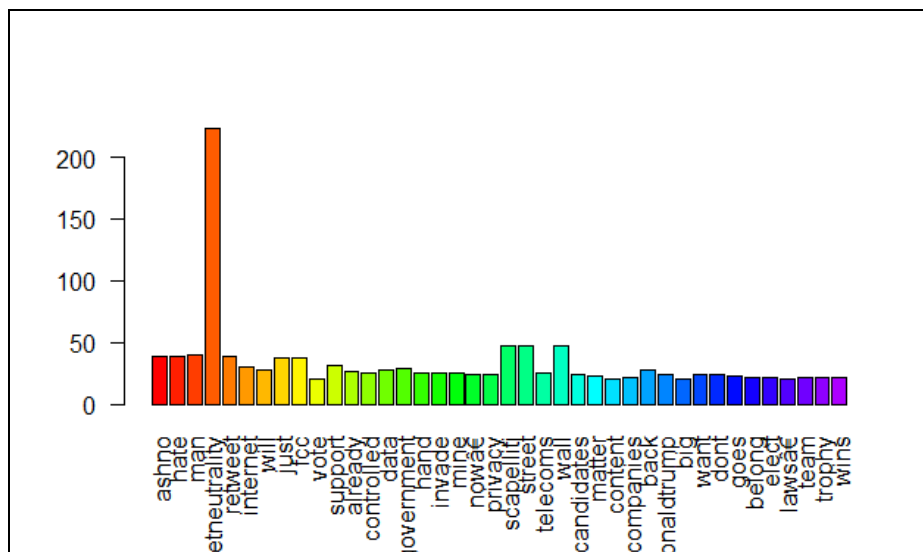

Figure :4 Lexicons on tweets collected

C. Classification

Classification is done using Naive Bayes Classifier. In this classifier, a classification of words based on Bayes theorem i. After pre-processing and filtration we will have a bag of words. These words represent positive, negative and neutral opinion of the tweets generated by the people. The classifier resultant is based on the probability of occurrence of these words based on Bayes theorem.

People in certain areas are in favour of net neutrality. India has no laws in governing net neutrality. In 2017 TRAI released its recommendation on Net neutrality.So we can say India is one step closer to ensure the Net Neutrality. In India, the debate over net neutrality gained momentum when the telecom major Airtel announced charges for calls through Skype and Whatsapp using its network. Facebook's free basics campaign was to support free flow of data in network. net users were prompted to send messages in support of net neutrality through facebook. In February 2016, Telecom Regulatory Authority of India barred differential data pricing on content services and finally prohibited telecom service providers from discriminatory charges on data usage and favoured the net neutrality in India. TRAI's ruling also ended the free basics campaign in the country.Meanwhile, in US, the Federal Communications Commission repealed the net neutrality rules paving the way for service providers to introduce paid services or prioritise certain services

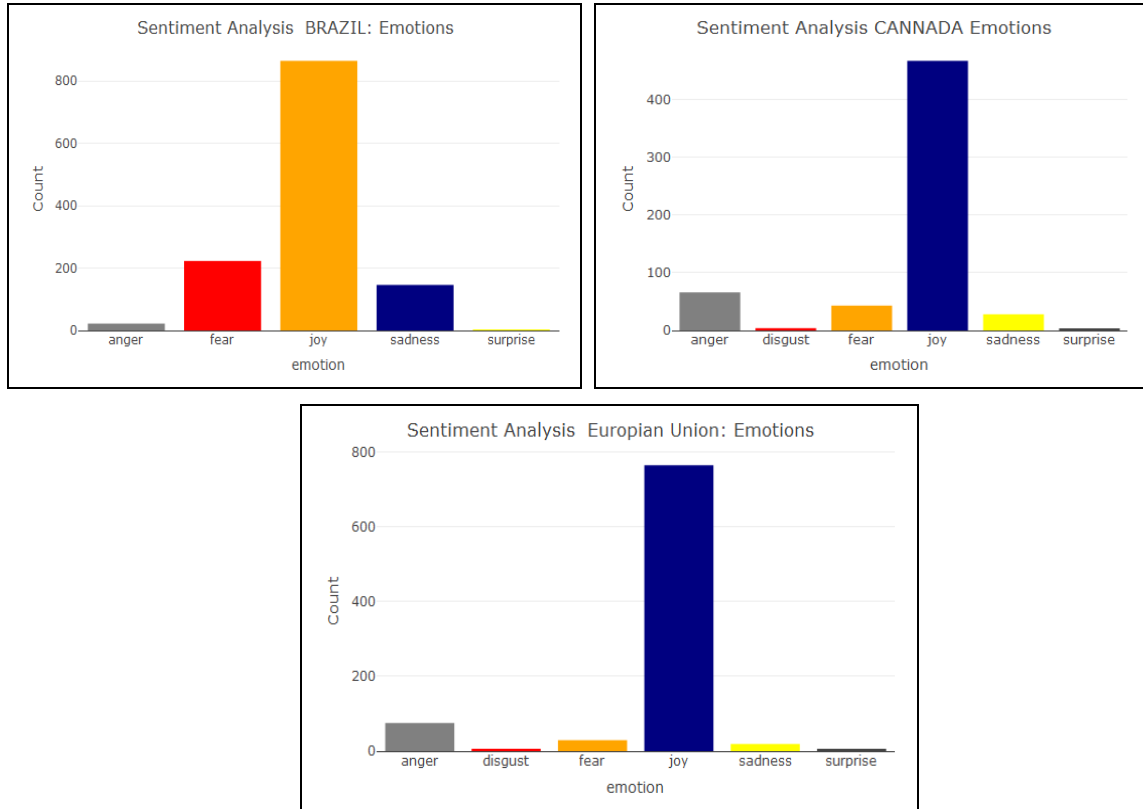The emotions of various countries is as in the figure 5

Figure 5 Sentiments generated by various 15 Sentiments generated by various Countries

The polarity is the measure of positive, negative and neutral of the opinion expressed in the tweets. The figure below represents the polarity of different countries.
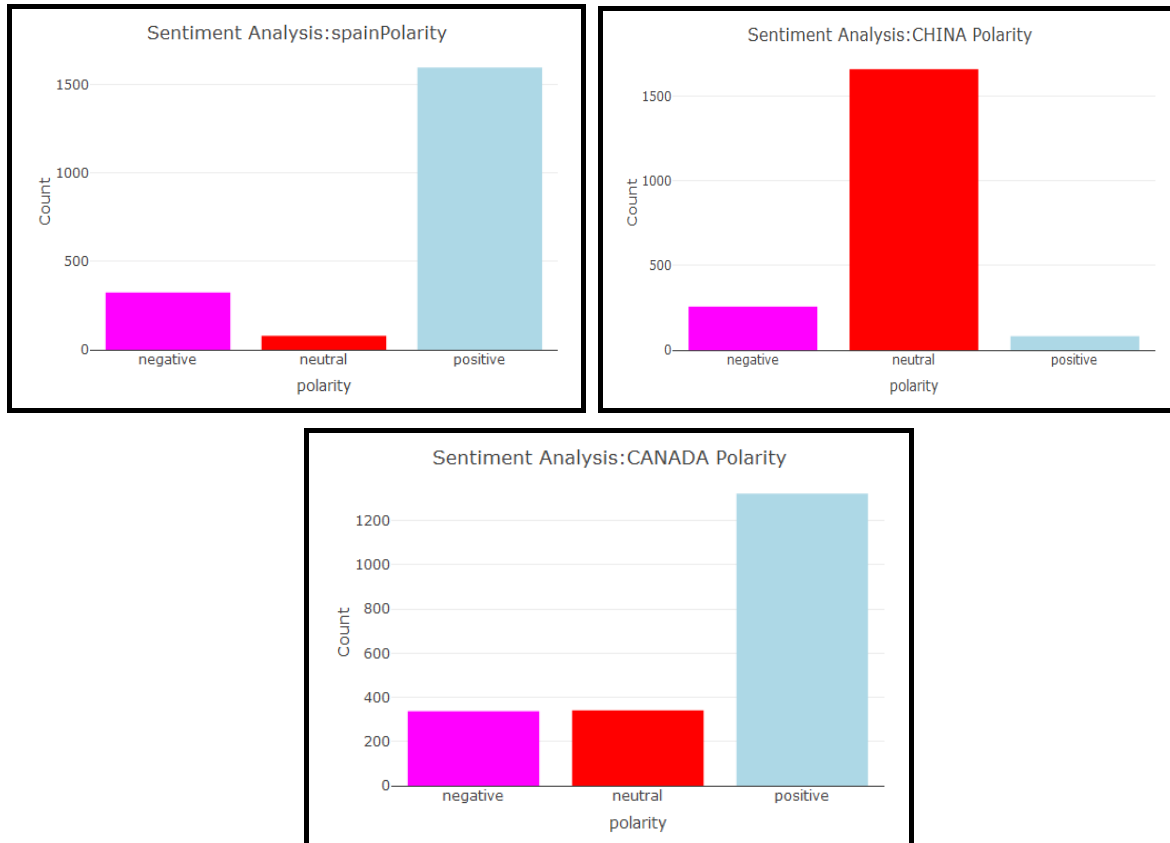

Figure 6. Polarity generated by various Countries

The result shows the polarity of the opinion of people differ from place to place .The analysis of polarity is evaluated and resultant is represented in a table is as follows

Table:1 Polarity Classification

Country	Negative	Neutral	Positive
India	256	1615	93
Brazil	315	306	1379
EU	248	425	1327
UK	168	35	1797
China	257	1661	82
France	267	1474	259
Germany	305	72	1623
Singapore	253	1644	103
Spain	324	79	1597
Cannada	537	341	1322
Slovenia	309	75	1616

The result can be graphically represented as follows
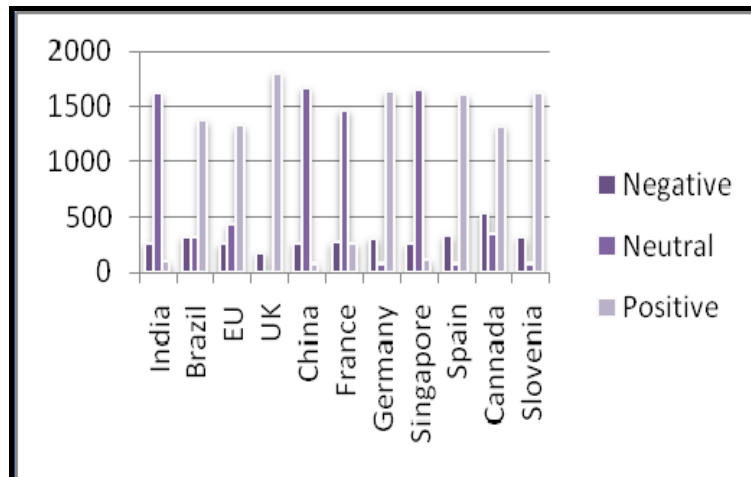

Figure 7 Polarity Graph

V.   CONCLUSION

Net neutrality is neutrality in web space. In this paper an analysis of twitter data is done from various countries all over the world, The data collected is analyzed to find the sentiments and polarity of the opinions expressed. The people had a versatile opinion regarding the topic . Among the countries evaluated , Solvenia, Spain, Germany, UK,EU have positive opinion towards net neutrality. The countries China, France, India, and Singapore have neutral opinion over the topic. It can be concluded that sentimental analysis can be used to evaluate the sentiments and polarity of people's opinion.

REFERENCES

[1]   T K ShivaPrasad, Jyothi Shetty,,Sentimental Analysis on Product Review, International Conference on Innovative Communication and Computations Technologies.2017 MarchISSN 1704223
[2]   Haseena Rahamath Opnion Mining and Sentimental Analysis-Challenges snd Applications, IJAEIM, ISSN 2329-4847
[3]   Devika M D, Sunitha Cª, Amal Ganesh ,Sentiment Analysis: A Comparative Study On Different Approaches, Fourth International Conference on Recent Trends in Computer Science & Engineering. Chennai, Tamil Nadu, India, Science Direct Procedia Computer Science 87 (2016) 44 – 49
[4]   A. Tamilselvi, M. ParveenTaj ,Sentiment Analysis of Micro blogs using Opinion Mining Classification Algorithm,International Journal of Science and Research (IJSR) ISSN (Online): 2319-7 IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727 PP 01-05 .
[5]   "Incorporating sentiment prior knowledge for weakly supervised sentiment analysis ",He Y., ACM Transactions on Asian Language Information Processing, Vol. 11(2).
[6]   Sentiment analysis in twitter using machine learning techniques, M s Neethu, R Rajasree ,2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT) ACC No14063416

[7] "Analyzing Influence of Emotional Tweets on User Relationships by Naive Bayes Classification and Statistical Tests", Kiichi Tago, Qun Jin, Service-Oriented Computing and Applications (SOCA) 2017 IEEE 10th International Conference on, pp. 217-222, 2017.

[8] Sentiment Analysis of Twitter Data: A Survey of Techniques Vishal A. Kharde, International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016

[9] Adaptive cotraining SVM for sentiment classification on tweets. Liu, S., Li, F., Li, F., Cheng, X., &Shen, In Proceedings of the 22nd ACMinternational conference on Conference on : information & knowledge management (pp. 2079-2088). ACM,2013.

[10] Feature Selection and Classification Approaches for Sentimental Analysis, Gautami Tripathi1 and Naganna S Machine Learning and Applications: An International Journal (MLAIJ) Vol.2,No.2,June2015 ISSN 2394-0840.