ISSN: 0975-766X
CODEN: IJPTFI
Research Article

Available Online through
www.ijptonline.com

SENTIMENT CLASSIFICATION ON SOCIAL NETWORK DATA

I.Mohan*[1,] M.Moorthi[2]
Research Scholar, Anna University, Chennai.
Associate Professor, Prathyusha Engineering College Chennai.
Email: itmoha@gmail.com

Received on: 25-03-2017 *Accepted on: 28-04-2017*

Abstract:

“What do the people think?” is an important factor that is needed to be considered during the decision making process. Hence analysing the sentiments of the people are more important. Sentiment analysis is particular to a topic. I.e., classifiers can perform well only on a particular topic. If the topic differs classifiers may not be able to perform well. This is considered to be a major drawback in the case of sentiment analysis on social network data. Social network data are varied and this increases the complexity on classification of data. An universal labelling of data are more complex on the other hand. Sentiment classification of data aims at analysing and classifying the various diversified data to determine whether the data falls under positive, negative or neutral category. Sentiment analysis mainly deals with determining the polarity and the classification of emotions. Classification involves the process of splitting up the data into text and non-text features. Further the algorithms are used to classify the data. Classification involves in two process i.e., Polarity classification and emotion classification. Finally, a visualization graph is drawn to visualize the classification.

Keywords: Social media, sentiment classification, svm, naive Bayes.

1.Introduction: Sentiment analysis is the process of analysing the opinions , feelings and attitude of the speaker about a particular product, topic, task, organization etc. Hence, it is known as opinion mining. The growing social media has attracted the people to post their emotions, feelings and suggestions as comments. The opinions of the people not only depicts the emotions but also have business values. But, it is an complex to find out the overall opinion and suggestions of the people. To classify their opinion, we need sentiment classification system which would drastically reduce the work of the human and would classify huge number of social network data.

Social network data are varied. But sentiment classifiers always concerns on a particular subject or a topic. Classifier which perform well in one domain may not work well in the other. This drawback is mainly due to the different language constructs and their usage. For e.g.in a product review, play game comment is treated to be of negative value, while the same comment is treated as a positive comment in the game review. In social media, people use different opinions and suggestions on different topics and domains. Hence, a topic based classification is much needed to classify the data.

Social network data: social networking sites has become a part of everyone's life. People post their feelings and opinion in social media. Social media act as a platform which allows exhibiting the opinions of the people. Opinions of the people in social media are considered to be the social network data. Each of social network data are of economic , political and business value. These social network data may also reflect the opinion of people in general. Sentiment classification is performed on the social network data.

Diagrammatic representation of Sentiment Analysis on twitter data

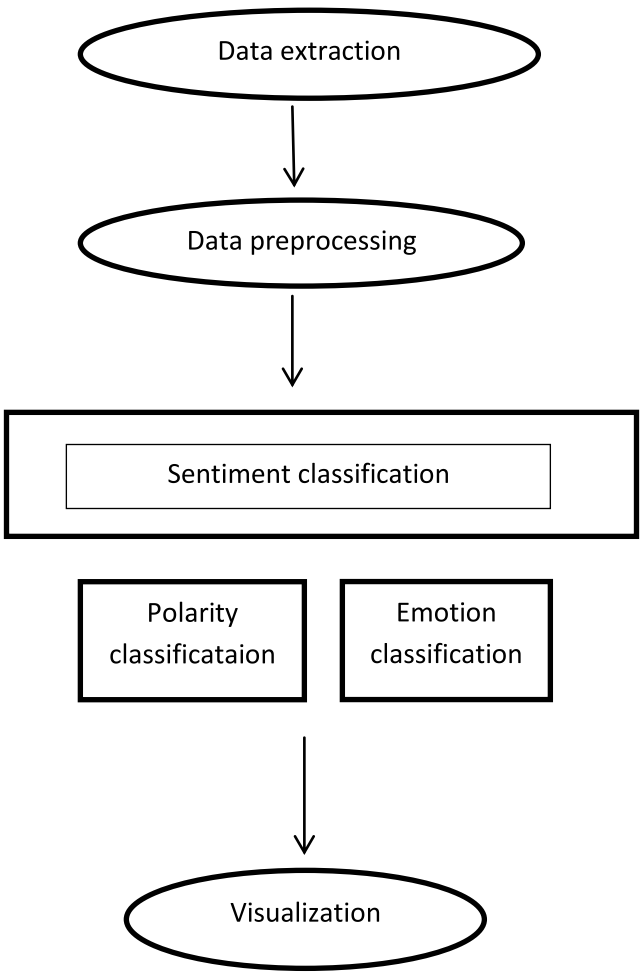

Fig.1.Sentiment Analysis on Twitter.

2. Three Different Classes of Sentiment Analysis

Sentiments can be classified into three different classes .i.e. positive, negative and neutral sentiments.

a. Positive Sentiments: These are the good (noble) words about the product in concern. If the positive sentiments are increased, it is denoted as good. In case of product reviews, if the positive reviews about the product are more when compared to the negative reviews, then we can conclude it is bought by many customers.

b. Negative Sentiments:

These are the bad (immoral) words about the product in consideration. If the negative sentiments are increased, it is rejected from the preference list. In case of product reviews, if the negative reviews more than the positive review on a project, then we can conclude no one is intending to buy it.

c. Neutral Sentiments: These are neither good nor bad words about the product. Hence it is neither preferred nor ignored.

3. Three Levels of Sentiment classification:

There are three different levels of sentiment classification. i.e. word level, phrase level and document level sentiment classification.

a. **Word Level Classification**:

This level of classification is carried out on the basis of the words that indicate the sentiment about the target. The word maybe noun, adjective or adverb. word level classification gives more accurate classified sentiments.

b. **Phrase Level Classification**:

This level of classification falls in good as well as bad category. The phrase signifying the attitude is found out from the sentence and the classification is done. But then it sometimes gives incorrect results if a negative word is added in front of the phrase. The phrase denotes the combination of two or more words that are not closely related to each other.

c. **Document Level Classification**:

In this level of classification, single document is considered about the prejudiced text. A single evaluation about the single subject from the document is considered. Then at times it is not useful in the case of blogs and forums as a customers might

compare one product with the another product which has similar features. Yet again the document may consist of the unrelated sentences which don't look like an opinion about the product.

4. Challenges in sentiment analysis

Sentiment related words and terms are of more significance. But the problem with the words and phrases are more complex to resolve i.e, sentiment lexicons are important but they do not provide the information that are needed for sentiment analysis and classification directly.

- In most of the times, the positive and the negative words have different orientations. For e.g, the word war is a negative word but at the same time, in the sentence “world war should be stopped” it denotes positive features.
- At some cases, the sentence containing the sentiment words fails to express the emotion of the sentiment. For.eg “why I am not happy?”. Here the sentiment word happy fails to express the own emotion.
- Sarcastic sentences always tend to be difficult for the sentiment classification.
- Some sentences may either contain the sentiment words or not. At such cases,it’s difficult to handle the sentiment of the sentence.

5. Literature survey

Twitter Sentiment Analysis: The Good the Bad and the OMG:

This paper deals with the investigations on the usage of the linguistics words that express the sentiments of the tweets in the twitter. This paper have estimated the already existing lexical resources as well as the features that are used to capture the information about the innovative and informal language in the, micro-blogs. To solve this problem, supervised learning method is introduced but influences the existing hash tags in twitter data.

Interpreting the Public Sentiment Variations on Twitter:

Twitter sentiment analysis is an important research area for academic as well as business fields for decision making like for the seller to decide if the product should be produced in a large quantity as per the buyers feedback and for the students to decide if the study material to be

referred or not. in this work, Shulong Tan et al. have proposed LDA based two models to interpret the sentiment variations on twitter i.e.-LDA to distill out the foreground topics and RCB-LDA to find out the reasons why public entiments have been changed for the target.

Sentiment analysis of twitter data:

This paper was published in 2012. It introduced the machine learning technique to implement the sentiment analysis on data. Sentiment classification of data classified the data as positive, negative and neutral. They used two kinds of models:

tree kernel and feature based kernel model. Both the models leave behind the unigram baselines. They performed the feature analysis for the feature based approach that reveals the important significance which combine the polarity and parts-of-speech tags.

6. Sentiment classification on social data

In our proposed system, we perform sentiment classification on twitter data to classify the data into three categories.

- Positive
- Negative and
- Neutral

6.1 About twitter

Twitter is a popular social networking site and micro blogging service which allows the user to express their opinions and feeling through their posts, which are commonly known as Tweets. Tweets are very small messages which have a limit bound of 140 characters.

Due to this limitation, people use acronyms, emoticons, short words to express their feeling.

Following are the some of the terminologies that used in tweets

Target:

Twitter users uses the symbol "@" to refer the target user or micro blogger that will automatically alert the target user.

Emoticons: Emoticons are the pictorial representations of the feeling that are used to convey the feeling of the user quickly.

Hash tags: Hash tags are usually used to mark up the important topics. Hah tags increase the visibility of their tweets.

6.2 sentiment classification:

Word level and document level classification may also produce inaccurate results. Sometimes it is insufficient in many applications. Hence, we need to understand the sentiments of the tweets based on analysing the appropriate sentiment of the opinion.

For e.g., "yes…. I love # dark chocolates" .In this tweet, # dark chocolate is the entity, love is the sentiment. The opinion on this general aspect is positive.

Overall design of the sentiment analysis

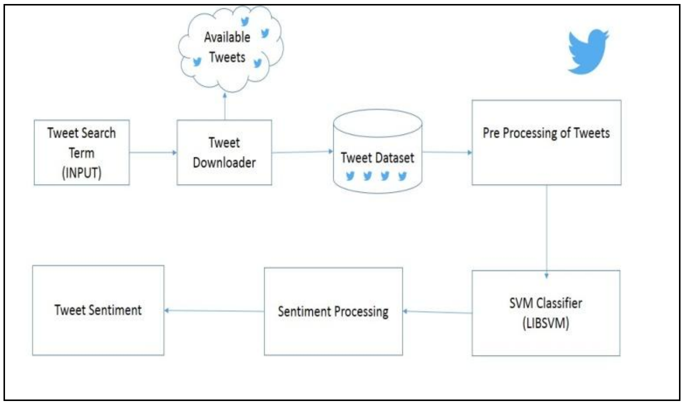

Fig .2. overall architecture of the process.

6.3 Sentiment analysis process:

Data extraction: twitter contains huge amount of data. Therefore, we need to extract the tweets on a particular topic from the twitter API.

Data pre-processing: This technique involves the cleaning of data by removing the punctuations, stem words, spell correction etc.

Applying classification algorithms: classifications algorithms are applied to categorize the tweets based on the polarity and emotion of the tweets.

Visualization: the result of the sentiment classification is represented in the graphs.

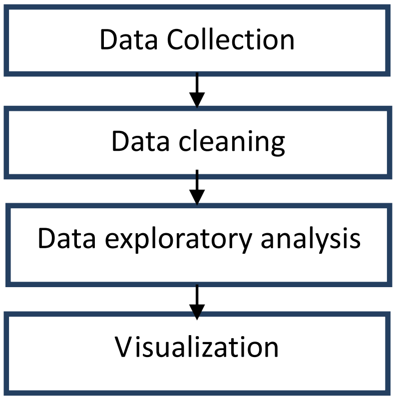

Fig.3. modules of sentiment analysis.

7. Implementation of the sentiment analysis:

7.1 **Data extraction**: Social network data are extracted from the social networking sites such as twitter. To extract a data from twitter, one must have an account in twitter. To access twitter data, we need to create an application on the developer site. Keys which are generated during the application creation are then used to extract tweets in R.

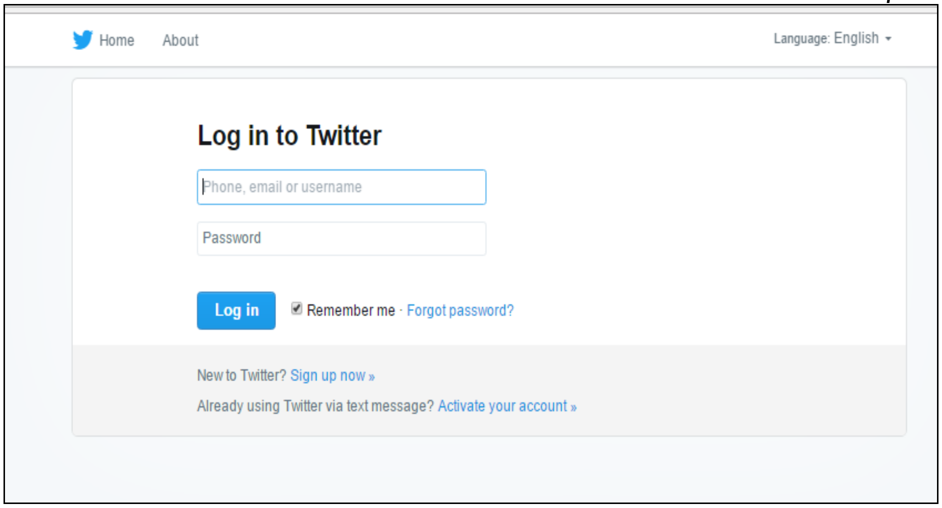

Fig.4 Twitter login page.

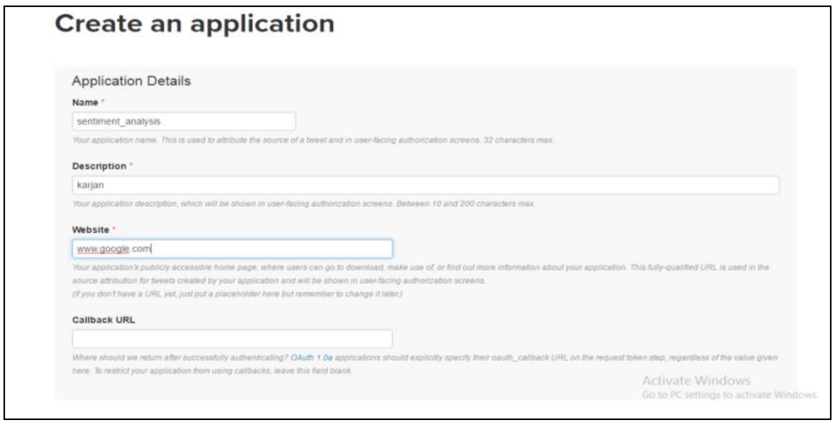

Fig .5 application creation.

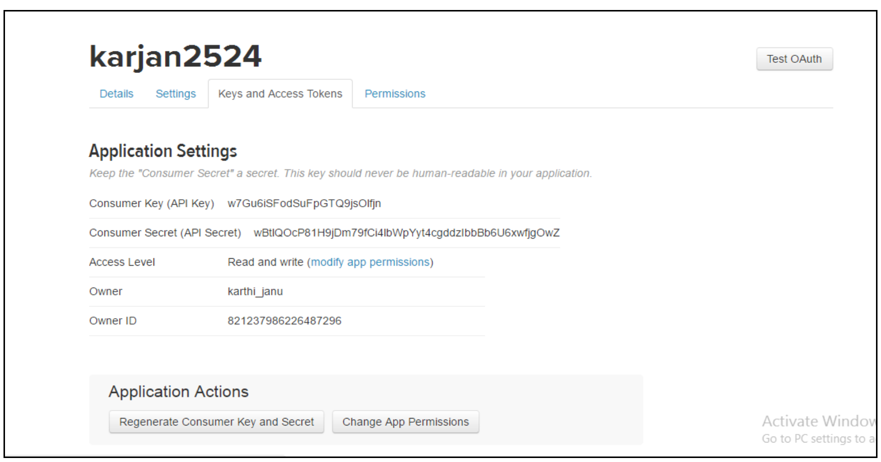

Fig.6 application settings which has the credentials.

7.2 **Data pre-processing**: Extracted tweets may also contain the noisy data. Those noisy data are needed to be pre-processed. pre-processing techniques involves the removal of

- blank spaces,

- @ people ,
- punctuations, ,
- integers,
- numeric characters and
- Duplicate tweets.

Other pre-processing activities include:

- **Converting to lower/upper case**: in order to simply the process, we need to convert the whole text into upper/lower case for the easy processing of the data.
- **Removing URL**: hyperlinks in tweets do not play much role in classification so they need to be removed.
- **Removing newline character**: these character indicate the newline represented by "\n", hence they are to be removed.

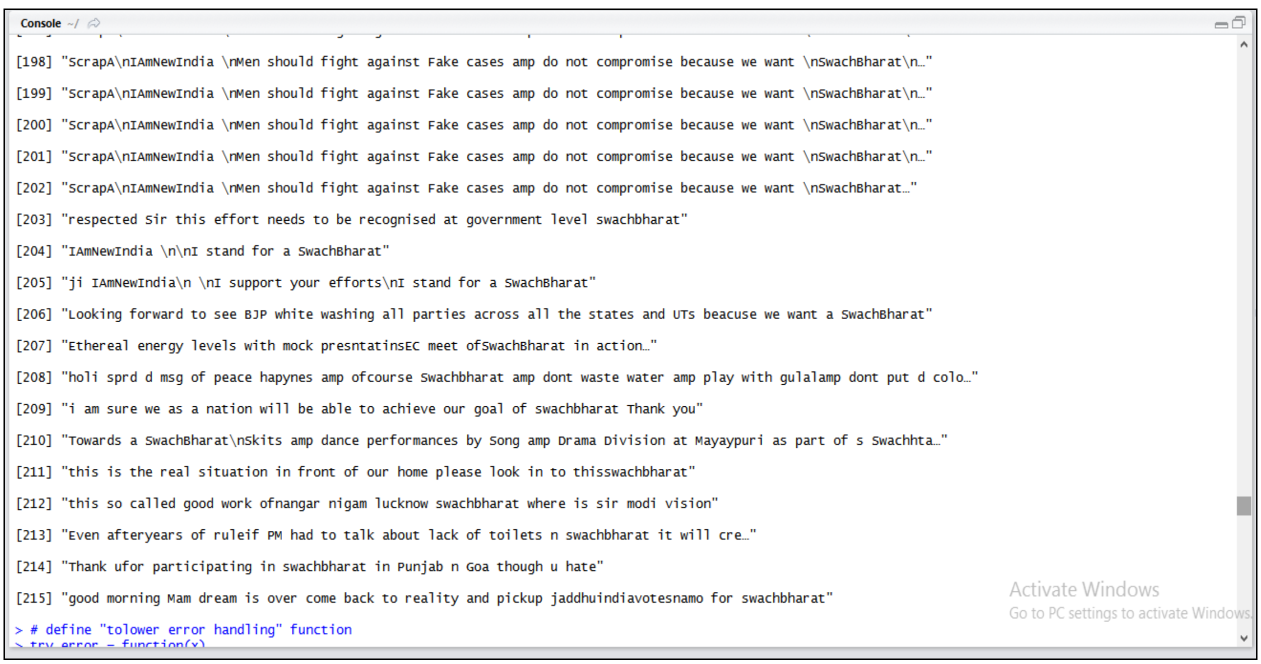

Fig.7 pre-processed dataset about swatch bharat.

7.3 Sentiment classification: Sentiment classification involves the classification of polarity and emotions of the data. Sentiment package is downloaded from the repository for the classification of tweets.

7.3.1 Polarity classification:

- Polarity classification is classifying and categorizing a data under either positive, negative or neutral.
- SVM algorithm is used to classify the polarity of the tweets.

7.3.2 **Emotion classification:**

- Emotion classification involves the classification of the type of the emotion exhibited in the tweet. Emotions includes joy, anger, frustration, anticipation, surprise, happy, fear and sad.
- Naïve Bayes algorithm is used to classify the emotion of the tweets.

7.4 Visualization of the data:

- In R studio, ggplot2 package was installed for plotting the graph.
  - ➢ Polarity graph was generated.
  - ➢ Emotions graph was generated.
- word cloud package was installed for the visualization of word cloud.
  - ➢ word cloud was then generated.

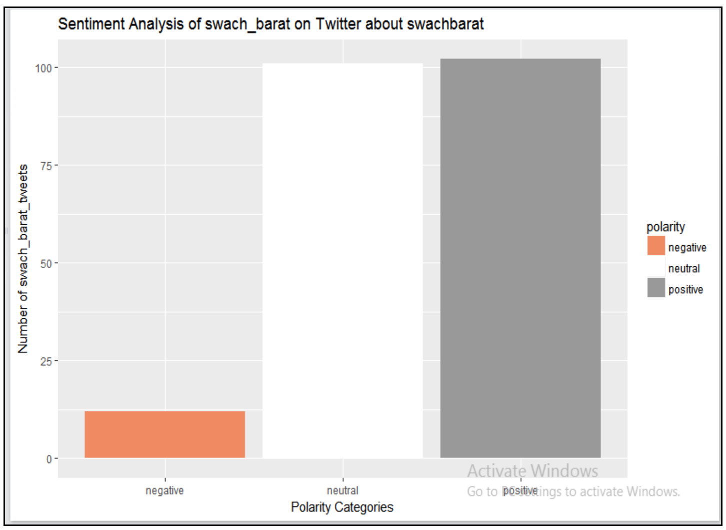

Fig.8 polarity classification on swatch bharath

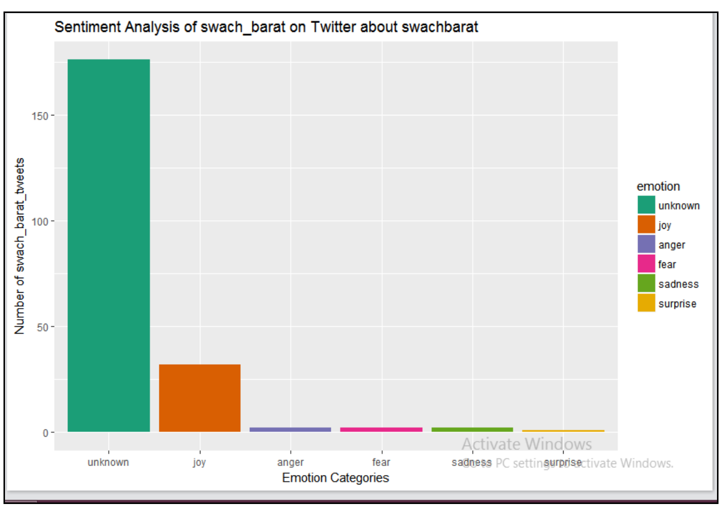

Fig.9.Emotion classification on swatch bharath data.

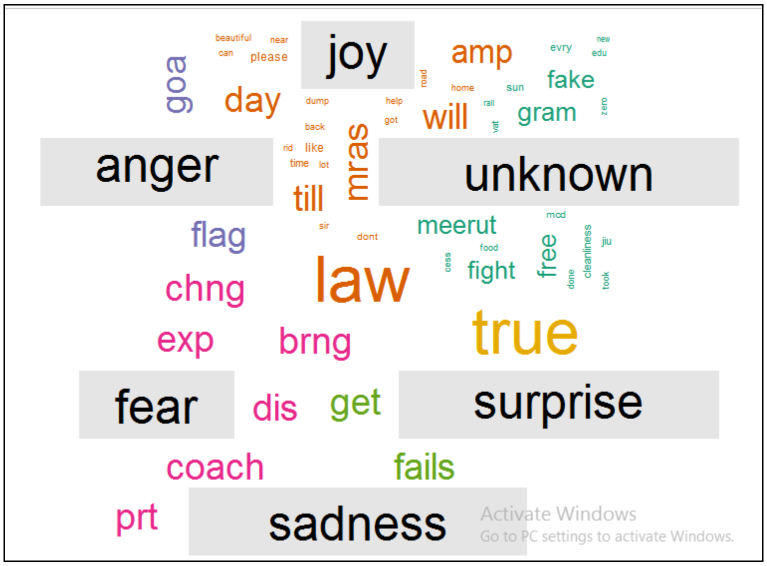

Fig 10. Word cloud.

8. Future enhancements: Further this paper can be extended with classifying the data along with the emoticons. Emoticon data can also reflect the opinion of the people. Classification of social network data along with the emoticon is a huge and complex task which is in research.

9.Conclusion:This paper provides a clear understanding of the sentiment analysis on social network data. Sentiment analysis has been a topic of research for years. Survey which are done in the field shows the evolution of research in the topic. Sentiment classification is a complex task which needs lots of research and analysis to predict the exact the output. Our paper has successfully analysed and classified the social network data. Limitations and enhancements in the paper will be done in the future work.

References:

1. Mohan I, "*Knowledge Discovery Using Big Data*" in Journal of Current Computer Science and Technology Volume 5 Issue 5 May 2015
2. Mohan. I, Sangavi. D, Priyanka. K "*Survey- Algorithms Used For Sentiment Analysis*," in International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 2 Issue 3 PP:961-973 March 2016 .
3. Mohan. I, Pradeep Kumar.S, Gokula Krishnan.S, Aravindhan. S "*Survey On Social Network Mining*," in International Journal of Computer Science and Mobile Computing, Volume 5 Issue 3 PP:467-481 March 2016 .
4. L. T. Nguyen, P. Wu, W. Chan, W. Peng, and Y. Zhang, "Predicting collective sentiment dynamics from time-series socialmedia," in Proc. 1st Int. Workshop Issues Sentiment Discovery Opinion Mining, 2012, p. 6.

5. M. Thelwall, K. Buckley, and G. Paltoglou, “Sentiment in twitterevents,” J. Am. Soc. Inform. Sci. Technol., vol. 62, no. 2, pp. 406–418,2011.
6. A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, “Sentiment analysis of twitter data,” in Proc. Workshop Lang. Soc. Media, 2011, pp. 30–38.
7. B. Liu, “Sentiment analysis and opinion mining,” Synthesis Lect. Human Lang. Technol., vol. 5, no. 1, pp. 1–167, 2012.
8. C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, “User-levelsentiment analysis incorporating social networks,” in Proc. 17thACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011,pp. 1397–1405.
9. J. Blitzer, M. Dredze, and F. Pereira, “Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification, ”in Proc. 45th Annu. Meeting Assoc. Comput. Linguistics, 2007, vol. 7, pp. 440–447.
10. F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, “Cross-domain co extraction of sentiment and topic lexicons,” in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics: Long Papers, 2012, pp. 410–419.
11. S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, “Cross-domainsentiment classification via spectral feature alignment,” in Proc.19th Int. Conf. World Wide Web, 2010, pp. 751–760.
12. I. Ounis, C. Macdonald, J. Lin, and I. Soboroff, “Overview of thetrec-2011 microblog track,” in Proc. 20th Text Retrieval Conf., ,http://trec.nist.gov/pubs/trec20/t20.proceedings.html
13. I. Soboroff, I. Ounis, J. Lin, and I. Soboroff, “Overview of the trec-2012 microblog track,” in Proc. 21st Text Retrieval Conf., 2012.
14. A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” CS224N Project Report, Computer Science Department, Stanford, USA, pp. 1–12, 2009.