Sentiment Analysis of Events from Twitter Using Open Source Tool

Rajkumar S. Jagdale[1], Vishal S. Shirsat[2], Sachin N. Deshmukh[3]

[1, 2, 3] Department of CS and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad
[1] rajkumarjagdale@gmail.com; [2] vss.csit@gmail.com; [3] sndeshmukh@hotmail.com

Abstract— Now a days, Popularity of Internet has been rapidly increased. Sentiment analysis and opinion mining is the field of study that analyses people's opinions, sentiments, evaluations, attitudes, and emotions from written language. User generated contents are highly generated by users. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. It is difficult to analysis or summarize the user generated content. Most of the users writes their opinions, thoughts on blogs, social media sites, E-commerce site etc. So These Contents are very important for individuals, industry and research work to take decisions. For this Sentiment analysis and opinion mining research is hot research area which comes under the Natural Language processing. R is open source tool. In this Paper we are elaborating the different approaches of Sentiment Analysis and Opinion Mining for different dataset and find out the which approach is best for which dataset which will help to researchers to select approach and dataset. In proposed work we collected tweets using R tool of different events from twitter and did pre-processing and calculate sentiment score from that events. We plot Wordcloud of particular event which highlight the frequent term from tweets and also calculated numbers of positive, negative and neutral tweets from each events.

Keywords— Sentiment Analysis and Opinion mining, Natural language processing, SentiWordNet, General Inquirer

I. INTRODUCTION

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service. The aim of Sentiment analysis or Opinion mining is to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The basic task of Sentiment Analysis is to classify polarity of the given word, phrase, sentence, documents etc. Also we can find the different emotional states such as "Happy", "Sad", "Angry", "Fear", "Surprise" etc. If we think about business intelligence, sentiment analysis used in different ways. For example, in marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features.

There are diverse challenges in sentiment analysis. One is an opinion word that is considered to be positive in one situation may be considered negative in another situation. A second challenge is every time people don't express their opinion in same way. Most traditional text processing depends on the difference between two

pieces of the word. In Sentiment Analysis, however, "the picture was nice" is very different from "the picture was not nice". People can be contradictory in their opinions. People are giving both positive and negative comments, which is somewhat manageable by analysing sentences one at a time. Peoples express their opinion on different informal medium like twitter, blogs, Facebook, Amazon etc. which are human readable but difficult to understand for machine. Sometimes even other people have difficulty understanding what someone thought based on a short piece of text because it lacks context. For example, "That movie was as good as its last movie" is entirely dependent on what the person expressing the opinion thought of the previous model. Languages that have been studied mostly in sentiment analysis are English and in Chinese. Presently, there are very less researchers who did research in other languages like Arabic, Italian and Thai. This survey focuses from 2004 onwards, and which contains different methods on different dataset like Movie reviews, Product Reviews.

According to Bing Lui Sentiment Analysis has three main levels. Following are the three levels of Sentiment Analysis.

1.1 Levels of Sentiment Analysis
a.   Document level Sentiment Analysis

In this Sentiment Analysis level whole document has analysed and classify whether a whole opinion document expresses a positive or negative sentiment [1], [2]. In one document only reviews of one product has been reviewed. And task is to find out the opinion about that product. So this task is broadly known as document-level sentiment classification. In this level, expressed opinion is on single entity. This is not applicable when there is document which contains multiple product reviews.

b.   Sentence Level Sentiment Analysis

In this level, task goes to every sentence and determine whether the sentence expresses the positive, negative or neutral opinion. This level attentively related to Subjectivity Classification [3], which distinguishes objective sentences and subjective sentences. Objectives sentences express factual information about sentences where Subjective sentences express the subjective information about sentences. Many objective sentence can involve opinions. This task is known as Sentence Level Sentiment Analysis.

c.   Aspect level Sentiment Analysis

Aspect Level sentiment Analysis was earlier called Feature level (feature-based opinion mining and summarization) Sentiment Analysis [4]. Document and Sentence Level Sentiment Analysis do not find out what exactly people like or did not like. It achieves finer-grained analysis. In this level directly looks at the opinion itself instead of looking to documents, paragraphs, sentences, clauses or phrases. This level consider the entity, aspect of that entity, opinion of aspect, opinion holder and time. Because of these parameters this level can find what actually people like means which feature of product mostly likes by customers and also on which time. This task is more interesting and more difficult too.

II. DATA SOURCE

Improvement quality of service is depends upon the user's opinion .Using this opinions, individual or industry can find out the popular event or products in the world of Internet. User are giving their opinions on different social media websites and E-commerce sites. Like twitter, Facebook, Amazon, and Flipkart etc. Many of these blogs contains different reviews of different products, events, issues etc.

a.   Blogs

Blog is a web site on which someone writes about personal opinions, activities, and experiences. Blogging is one of the most valuable tools that businesses have to engage with customers and ultimately make their lives easier. With the increasing growth of User generated Content on the internet, blogging pages are increasing rapidly. For expressing personal opinion Blog pages are most popular. Bloggers record the daily events in their lives and express their opinions, feelings, and emotions in a blog [5]. In Sentiment analysis and opinion mining research blogs are used as source of user's opinion and used in many of the studies [6], [7], [8].

b.   Data Set

In the field of Movie review classification movie review dataset are mostly used which is available on website (http://www.cs.cornell.edu/People/pabo/movie-review-data). For multi-domain sentiment (MDS) dataset following website (http:// www.cs.jhu.edu/mdredze/datasets/sentiment) is used by researchers. In multi-domain sentiment analysis dataset contains four different types of product reviews extracted from Amazon.com including Books, DVDs, Electronics and Kitchen appliances, with 1000 positive and 1000 negative reviews for each domain.

Another review dataset available on (http://www.cs.uic.edu/liub/FBS/CustomerReviewData.zip.) This dataset consists of reviews of five electronics products downloaded from Amazon and Cnet [9],[10],[11], [12],[13],[14],[15] [16], [17], [18].

Micropinion Generation Dataset (CNET) contains 330 review texts. The reviews are on products from various categories like TV, cell phones, gps etc. This dataset was used for text summarization of opinions. This dataset is available on (http://kavita-ganesan.com/micropinion-generation) website. MovieLens Dataset contains 100,000 ratings (1-5) from 943 users on 1682 movies by GroupLens Research Project at the University of

Minnesota. Each user has rated at least 20 movies. This dataset is available on http://grouplens.org/datasets/movielens/ website. Restaurant Review Dataset contains a total 52077 reviews. The fields contain rating information, review counts, percent and cuisine type which is available on (http://www.cs.cmu.edu/~mehrbod/RR/) website. SNAP Review Dataset contains a 34,686,770 Amazon user reviews from 6,643,669 users. This dataset was initially used for recommendation systems which is available on (http://snap.stanford.edu/data/web-Amazon.html) website.

c. Review sites

Opinion can be an important factor in making a purchasing decision for any user. A large and unstructured data is a growing body of user-generated reviews is available on the Internet. Products reviews or Movie reviews are based on opinions expressed in much unstructured format. Reviewers data for research is mostly crawled from different social media website or e-commerce websites like www.amazon.com (product reviews), www.yelp.com (restaurant reviews), www.CNET download.com (product reviews) and www.reviewcentre.com, which hosts millions of product reviews by consumers. Other than these there are different professional review sites such as www.dpreview.com , www.zdnet.com and consumer opinion sites on broad topics and products such as www .consumerreview.com, www.epinions.com, www.bizrate.com [19],[9],[20],[21].

d. Micro-blogging

Twitter is a popular microblogging service where users create status messages called "tweets". These tweets contains user's opinion, thoughts on different topics. Tweets are sometime used in classifying the sentiment about particular topic or review of products. Some Mobile reviews found on twitter.

III.RELATED WORK

In [1] work is related to document level sentiment analysis on movie review dataset. They applied different machine learning techniques i.e. (Naive Bayes, maximum entropy classification, and support vector machines). Dataset for this research is from IMDB website. They selected only reviews where the author rating was expressed either with stars or some numerical value. Features they have taken are unigrams, unigrams+bigrams, bigrams, unigrams+POS, adjectives, top 2633 unigrams, unigrams+position. Results got quite good in comparison to the human generated baselines. SVM works best as compare to other classifiers.

In [2] present a simple unsupervised learning algorithm for classifying a review as recommended or not recommended. Average Semantic orientation of the phrases in the review that contain adjectives or adverbs is used for classifying the reviews. This paper works on document level Sentiment Analysis. Point wise Mutual Information (PMI) is used calculate semantic orientation of phrase and the word. They took reviews from Epinions for different domains (Automobiles, Banks, Movies, and Travel Destinations). They got accuracy ranges from 84% for automobile reviews to 66% for movie reviews.

In [22] they have developed an algorithm for predicting semantic orientation. Algorithm designed for isolated adjectives, rather than phrases containing adjectives or adverbs. They used four step supervised learning algorithm to infer the semantic orientation of adjectives from constraints on conjunctions. In that they got accuracy for classification of adjectives ranging from 78 % to 92 % depending on amount of training data.

In [23] they developed such system that generates sentiment timelines. It tracks online discussions on movies and generate plot which contains number of positive sentiment and negative sentiment messages over time. They used specific domain lexicons for movies. It is used instead of a hand-built lexicon. This work is used in automatic review rating, tracking advertising campaigns, tracking public opinion for politicians, tracking financial opinions by stock traders, tracking entertainment and technology trends by trend analysers.

In [24] it is concerned with subjectivity tagging. They evaluated objectively present factual information. This paper identifies strong clues of subjectivity using the results of a method for clustering words according to distributional similarity. In 10-fold cross validation results, features based on both similarity clusters and the lexical semantic features are shown to have higher precision than features based on each alone.

In [25] classify a document's polarity on a multi-way scale and expanded the task of classifying a movie review as either positive or negative to predicting star ratings on either 3 or 4 star scale. They checked human performance at the task. Applied algorithm is Meta algorithm, Based on a metric labelling. This Meta algorithm can give best performance over both multi-class and regression versions of SVMs when we employ a novel similarity measure appropriate to the problem. They used movie review Dataset.

In [4] they determined the opinions or sentiments expressed on different features or aspects of entities, e.g., of a cell phone, a digital camera, or a bank. They analyses the different aspect of different entities. This paper performed three tasks i.e. (1) mining product features that have been commented on by customers (2) identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative; (3) summarizing the results. In their proposed technique they did POS tagging, Frequent Features Identification, Opinion Words Extraction, and Orientation Identification for Opinion Words, Infrequent Feature Identification, Predicting the Orientations of Opinion Sentences and Summary Generation. They got the results of opinion sentence extraction and sentence orientation prediction i.e. For Opinion sentence extraction for five products average Precision and recall is 0.64 % and 0.69 % respectively. They got Sentence Orientation accuracy 0.84 %.

In [26] Sentiment model has proposed to predict sales performance by taking information from blogs and formulated as a two-class classification problem. They also collect for each movie one month's Box office data (daily gross revenue) from the IMDB web- site. In it they proposed PLSA for sentiment mining. They proposed ARSA: A SENTIMENT-AWARE MODEL to provide product sales predications based on the sentiment information captured from blogs.

3.1 Sentiment Analysis Process

Sentiment Analysis is classified into two main approaches i.e. Supervised Learning Approach and Un-supervised Approach. In Sentiment Analysis Process Following Steps are necessary.
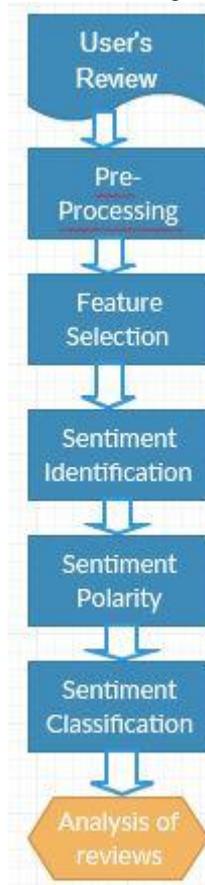

Figure 1. Sentiment Analysis Process

a.   Collection of User's Reviews

Reviews are necessary for doing the Sentiment Analysis Task. For the Collection of reviews there are different techniques which are used in this survey. Products reviews are collected from different E-commerce website like Amazon.com, Epinion.com and Cnet etc. and Movie reviews are collected from IMDB and other websites. The reviews can be a structured, semi-structured and unstructured type. Sentiment Analysis research, there are open source framework where researcher can get their data for the research purpose. R [27] is one of the programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. By installing required packages and authentication process of social website, to crawl the reviews from that site is easy task. Once we have our text data with us then we can use that data for Pre-processing purpose.

b.   Pre-Processing

Data pre-processing is done to remove the incomplete noisy and inconsistent data. Data must be pre-processed before using in feature selection task. In pre-processing following are some tasks:

- Removing URLs, Special characters, Numbers, Punctuations etc.
- Removing Stopwords
- Removal of Retweets (in case of twitter dataset)
- Stemming
- Tokenization

c.   Feature Selection

Feature selection from pre-processed text is the difficult task in sentiment analysis. The main goal of the feature selection is to decrease the dimensionality of the feature space and thus computational cost. Feature

selection will reduce the overfitting of the learning scheme to the training data. In [1] different machine learning algorithms were analysed on a movie review dataset with different feature selection techniques features are usually unigrams, biagrams and ngrams. POS tagging is used in feature selection techniques.

d. *Sentiment Word Identification*

Sentiment word identification is a fundamental work in numerous applications of sentiment analysis and opinion mining, such as review mining, opinion holder finding, and review classification. Sentiment words can be classified into positive, negative and neutral words.

e. *Sentiment Polarity Identification*

The basic task in SA is classifying the polarity of a given text at the document, sentence, or feature. The polarity is in three category i.e. Positive, Negative and Neutral. Polarity identification is done by using different lexicons e.g. Bing Lui sentiment lexicon, SentiWordNet etc. which help to calculate sentiment score, sentiment strength etc.

f. *Sentiment Classification*

Sentiment classification of movie review dataset and product review dataset is done using supervised machine learning approaches like naïve Bayes, SVM, Maximum Entropy etc. Accuracy is depends on which dataset is used for which classification methods. In the case of Supervised machine learning approaches Training dataset is used to train the classification model which then help to classify the test data.

g. *Analysis of Reviews*

Finally Analysis of result is important to make decision to individual and industry. In case of movie reviews if more result is positive then user can decide to go that movie .Analysis is used in business intelligence.

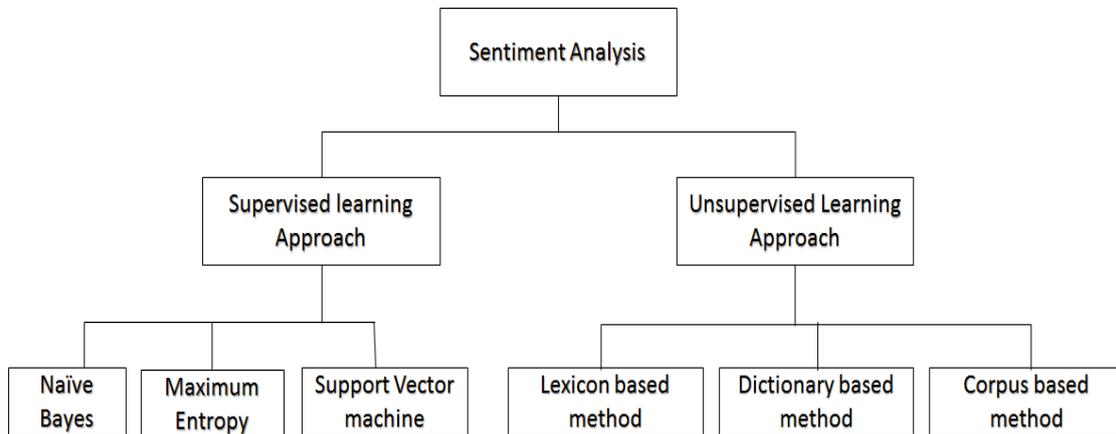### 3.1 *Sentiment Classification Approaches*

Figure 2: Sentiment Classification Approaches
There are two main approaches in sentiment analysis i.e. Supervised learning and Unsupervised Learning Approach

3.2.1 **Unsupervised Learning Approach**

This approach is used to classify the sentiment when we have training data and it can solve the problem of domain dependency and need to reduce the training data. Turney [2] used two seed words (poor and excellent) to calculate the semantic orientation of the phrase. Point wise mutual information has been used to find out association of seed words with their phrase. Sentiment of document is calculated as the average semantic orientation of all such phrases. This approach got 66% accuracy for the movie review dataset at document level.

A) *Lexicon based Methods*

Lexicon-based approaches for sentiment classification are based on the insight that the polarity of a piece of text can be obtained on the ground of the polarity of the words which compose it. [28] In this methods lexical resources which are concerned with mapping words to a categorical (Positive, Negative, Neutral) or numeric score calculated by the algorithm to get the overall sentiment of that text. Following are some lexical resources:

a. *SentiWordNet*

SentiWordNet [29] is a lexical resource used in the sentiment analysis applications. It gives an annotation based on three numerical sentiment scores (positivity, negativity, neutrality) for each WordNet synset [30]. It provides a synset-based sentiment representation, different senses of the same Term may have different sentiment scores. Word Sense Disambiguation (WSD) algorithm to identify the most promising meaning.

b. *WordNet-Affect*

479

WordNet-Affect [31] is a linguistic resource for a lexical representation of affective knowledge. It is an extension of WordNet which labels affective-related synsets with affective concepts defined as A-Labels (e.g. the term euphoria is labelled with the concept positive-emotion, the noun illness is labelled with physical state, and so on). The mapping is performed on the ground of a domain-independent hierarchy.

c.   MPQA

It [32] is related to subjectivity lexicons and provides lexicons of 8,222 terms labelled as subjective expressions which are gathered from different sources. This contains list of words along with their POS-tagging, labelled with polarity (positive, negative, neutral) and intensity (strong, weak).

d.   SenticNet

This [33] lexical resource is used for concept-level sentiment analysis. It depends on Sentic Computing [34] which is multi-disciplinary paradigm for Sentiment Analysis. SenticNet is able to associate polarity and affective information also to complex concepts like accomplishing goal, celebrate special occasion and so on. It gives sentiment score in range between -1 and 1 for 14,000 common sense concepts.

e.   Hu and Liu's lexicon

This opinion lexicon contains a list of positive and negative words or sentiment words for English. This list was compiled for [4] paper which contains 2006 positive and 4783 negative sentiment words.

B)   Dictionary Based Methods
f.   WordNet

WordNet [35] is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. In [36] polarity of a word is determined by measuring its shortest distance to 'good' and 'bad'. We extract the words that are contained by WordNet from our dictionary for comparison in our experiment.

g.   General Inquirer

The General Inquirer (GI) is an application in text analysis with one of the oldest manually constructed lexicons. The GI has been in development and refinement since 1966, and is designed as a tool for content analysis, a technique used by social scientists, political scientists, and psychologists for objectively identifying specified characteristics of messages [37]. The lexicon contains more than 11K words classified into one or more of 183 categories. In GI there are 1,915 words labelled Positive and the 2,291 words labelled as Negative. It is ideally used in several research to automatically determine sentiment properties of textual data.

h.   LIWC

LIWC is [38] text analysis software designed for studying the various emotional, cognitive, structural, and process components present in text samples. LIWC uses a proprietary dictionary of almost 4,500 words organized into one (or more) of 76 categories, including 905 words in two categories especially related to sentiment analysis. First category is Positive Emotions (e.g. Love, nice, good, great) which are 406 in numbers and Negative Emotions (Hurt, ugly, sad, bad, worse) which are 499 in number.

i.   AFINN

AFINN [49] is a list of English words rated for valence with an integer between minus five (negative) and plus five (positive). The words have been manually labelled by Finn Årup Nielsen in 2009-2011. The file is tab-separated. In first version AFINN-96 1468 unique words and phrases on 1480 lines and in second updated version AFINN-111 2477 words and phrases are present.

C)   Corpus Based Methods
j.   Darmstadt Service Review Corpus

It consists of consumer reviews annotated with opinion related information at the sentence and expression levels. In [39] they gave Niek Sanders: He has constructed a Twitter Sentiment Corpus that "consists of 5513 hand-classified tweets.

Author & Year	Dataset	Features & Techniques	Results			
			Accuracy	Precision	Recall	F1 Score
Kaiquan Xu(2011)[40]	Amazon Reviews	linguistic features, Multiclass,SVM,CRF	61.38	61.96	93.49	74.26
Long Sheng(2011) [41]	Movie Reviews	PMI,SO-A,SO-LSA, Back-Propagation neural network	64	60	98	75
Hanhoon Kang(2012) [42]	Restaurant Reviews	Unigram, Bigram, Improved Naïve Bayes I and NB	81.4	--	--	--

Lin Y, Zhang J (2012) [43]	Product Reviews	PMI, Lexicons, SVM,Unsupervised approach	--	82.62	85.26	83.92
Federico Neri (2012) [44]	Facebook posts Rai – the Italian public broadcasting service	words, phrases, sentences, Bayesian method, K-Means	--	93	87	--
Prashant Raina(2013) [45]	News Articles.	MPQA corpus, semantic parser, Sentic Computing, ConceptNet and SenticNet	71	Pos-46.3 Neg-61.6 Neu-90.9	Pos-79.3 Neg-70.5 Neu-69.8	Pos-58.5 Neg-65.8 Neu-79
Seyed-Ali Bahrainian (2013) [46]	Twitter dataset	Unigram feature, SVM, NB, MaxEnt Hybrid Approach	89.78	--	--	--
Emitza Guzman (2014) [47]	US App Store and Google Play.	Topics, N-top features, LDA,Topic Model, lexical sentiment analysis	--	91	73	--
Geetika Gautam (2014) [48]	Twitter dataset	Unigram, Naïve Bayes, Maximum entropy ,SVM, WordNet	NB-88.2 ME-83.8 SVM-85.5 SA-89.9	--	--	--

Table 1. Summary of Some Research Article

IV. PROPOSED METHODOLOGY

The methodology used for mining twitter dataset is shown in fig 3 and following steps are important in this methodology.

a. Data Collection

R is a programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing [27]. R has different packages which helps to get the social media data like Twitter, Facebook and also different packages for pre-processing on text / numeric data and visualization of results i.e. tm, stringr, ggplot etc. TwitteR package is used to get tweets from twitter using twitter API.

b. Pre-Processing

Pre-processing is very important in data mining and it effects on the accuracy of result. In Pre-processing tm package is used to get text from tweets and other pre-processing steps like Stopwords removal, removing spaces, punctuation, URLs and performing stemming (get the root of the words). After this step unstructured data represents in Term-Document Matrix.

c. Data Analysis

From the above step we got TDM and using TDM, it is easy to find association rules, finding more frequent terms and performing sentiment analysis using the lexicon-based approach, which uses a set of positive and negative words. Using Scoring Function score of every tweet has been calculated using Bing Lui lexicons [4].

d. Visualization of Result

Wordcloud package in R helps to represent the word cloud which display most frequents terms from the text data. In this methodology it shows the frequency of the words in the customer's tweets.

V. EXPERIMENTS AND RESULT

The experiments having collection of different hashtag (#) tweets. TwitteR package used to access the live tweets from twitter. ROAuth, TwitteR, Rcurl etc. packages enables authentication and access to twitter messages by using keyword search queries [27].

Topic name	No.of Tweets	Period
#Budget2016	10000	29 Feb 2016
#RailBudget2016	10000	25 Feb 2016
#Freedom251	10000	18 Feb 2016
#MakeInIndia	10000	23 Feb – 29 Feb 2016
#Oscars2016	10000	29 Feb 2016

#startup	10000	28 Feb – 29 Feb 2016	
#InternationalWomensDay	25000	08 Mar 2016	
#AsiaCupT20Final	20000	07 Mar 2016	
#IndvsPak	30000	20 Mar – 21 Mar 2016	
#ProKabaddi	5756	28 Feb – 06 Mar 2016	

Table 2.Twitter Dataset Used for Experimental Work

It used Unsupervised Learning approaches which contains lexicons in this methodology. Tweet has been accessed over time with different topics. Related to sentiment analysis and opinion mining for in this methodology lexicon based methods has been used. Fig 1 shows the Wordcloud of #MakeInIndia topic. For calculating score of each tweet we need the opinion lexicons and scoring function. Working of scoring function is as follows:

$$\text{Sentiment Score} = \Sigma \text{ positive words} - \Sigma \text{ Negative words} \qquad \text{(Equation 1)}$$

Score will be positive if the number of positive words are greater than number of negative words which called as positive polarity. Score will be negative if the number of negative words are greater than number of positive words which called as negative polarity. Score will be neutral if the number of positive and negative words are same or is no existence of any opinion words in the text.

Result of sentiment analysis are shown in Table 3 showing the distribution of positive, negative and neutral tweets of every topic. Wordcloud has been plotted of only one topic which is in fig 1. And Fig 2 shows the chart of table 3 which indicate the how polarity varies with different events on twitter dataset.

Topic Name	Positive tweets	Negative Tweets	Neutral Tweets
#Budget2016	3507	1555	4938
#RailBudget2016	3703	969	5328
#Freedom251	7564	581	1855
#MakeInIndia	3052	1012	5936
#Oscars2016	5539	694	3767
#startup	3907	799	5294
#InternationalWomensDay	15600	1724	7676
#AsiaCupT20Final	9742	2143	8115
#IndvsPak	16590	2606	10804
#ProKabaddi	3328	224	2204

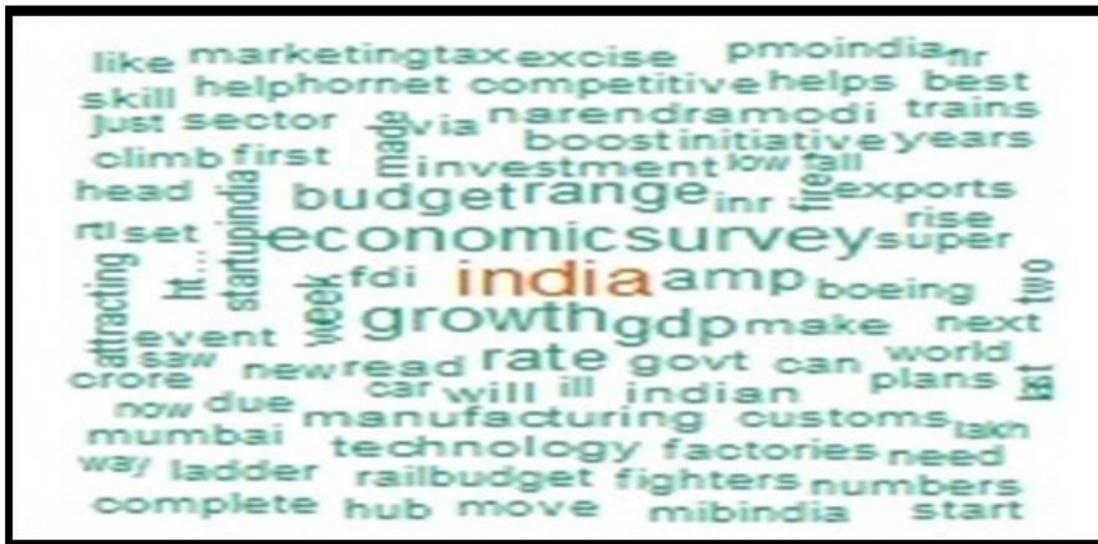Table 3. Result of Polarity of Twitter Dataset
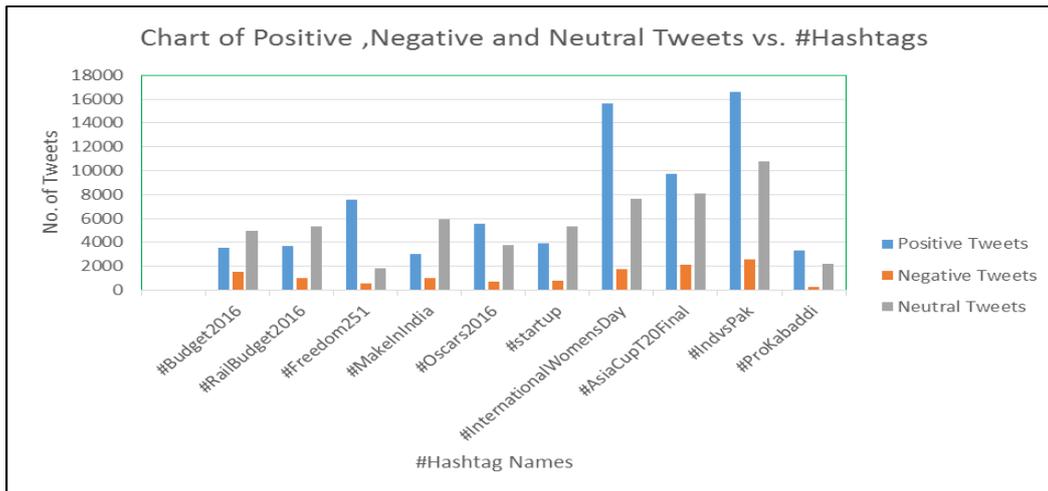

Fig1 Wordcloud of #MakeInIndia

Fig 2 No. of Positive, Negative and Negative Tweets vs. #Hashtag

VI. CONCLUSION AND FUTURE WORK

In this proposed work, we used Twitter API using R tool which is open source. Tweets from twitter has been collected and gives to pre-processing task in that tool. R open source tool is used in text mining and also to crawl streaming data from social media like twitter and Facebook etc. Movie reviews data also pre-processed in R tool for sentiment analysis and opinion mining. As explained in section 3, there are different supervised and unsupervised approaches and different lexicons, dictionaries and corpus based methods which are very helpful in Sentiment Analysis. Different dataset are available for movie review, product review, Epinions dataset etc. In this method sentiment score has been calculated and counted number of positive, negative and neutral tweets for given #Hashtag and can predict the public opinion of particular event. As per above analysis of different #Hashtags tweets for sentiment analysis, individual and industry can find the public opinion behind that event. Table of summary shows the used methods and dataset for particular research group. Bing Lui lexicons are used to calculate sentiment score. Fig 1 shows the Wordcloud which display frequent word of event on plot and highlight that terms which helps easily on which users has given tress.

Future work about product review sentiment analysis is find out aspects and their polarity of the product which helps for consumer to take decision to buy online products on e-commerce site. Aspect level sentiment analysis gives detail information about that product and about movie review, director of that movie should know what exactly user know from particular movie which is possible in aspect based sentiment analysis. In Hotel reviews same like movie, hotel owner should know what items people like from their hotel and what other items need for customers.

REFERENCES

[1] Pang, Bo; Lee, Lillian; Vaithyanathan, Shivakumar (2002). "Thumbs up? Sentiment Classification using Machine Learning Techniques". Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 79–86.

[2] Turney, Peter (2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". Proceedings of the Association for Computational Linguistics. pp. 417–424.

[3] Wiebe, Janyce, Rebecca F. Bruce, and Thomas P. O'Hara. *Development and use of a gold-standard data set for subjectivity classifications*. In *Proceedings of the Association for Computational Linguistics (ACL-1999)*. 1999.

[4] Hu, Minqing and Bing Liu. *Mining and summarizing customer reviews*. In *Proceedings of ACM SIGKDD* International Conference on Knowledge Discovery and Data Mining (KDD-2004). 2004.

[5] Chau, M., & Xu, J. (2007). Mining communities and their relationships in blogs: A study of online hate groups. International Journal of Human – Computer Studies, 65(1), 57–70.

[6] Martin, J. (2005). Blogging for dollars. Fortune Small Business, 15(10), 88–92.

[7]   Godbole, N.; Srinivasaiah, M. & Skiena, S. (2007), Large-Scale Sentiment Analysis for News and Blogs, in 'Proceedings of the International Conference on Weblogs and Social media (ICWSM)'.

[8]   H. Tang, S. Tan, and X. Cheng. A Survey on Sentiment Detection of Reviews. Expert Systems with Applications. Elsevier. 2009, doi:10.1016/j.eswa.2009.02.063.

[9]   Hu, and Liu, "Opinion extraction and summarization on the web", AAAI. (2006), pp. 1621-1624.

[10] Long-Sheng Chen, Cheng-Hsiang Liu, Hui-Ju Chiu, "A neural network based approach for sentiment classification in the blogosphere", Journal of Informetrics 5 (2011) 313–322.

[11] ZHU Jian , XU Chen, WANG Han-shi, " Sentiment classification using the theory of ANNs", The Journal of China Universities of Posts and Telecommunications, July 2010, 17(Suppl.): 58–62 .

[12] Bo Pang, Lillian Lee.A Sentimental Education: Sentiment Analysis Using Subjectivity. Proceedings of ACL, pp. 271--278, 2004

[13] Bai, R. Padman, "Markov blankets and meta-heuristic search: Sentiment extraction from unstructured text," Lecture Notes in Computer Science, vol. 3932, pp. 167–187, 2006.

[14] Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," Computational Intelligence, vol. 22, pp. 110–125, 2006.

[15] Zhou, L. and P. Chaovalit (2008), Ontology-supported Polarity Mining, Journal of the American Society for Information Science and Technology, 59(1), 1-13.

[16] Yulan He,"Learning sentiment classification model from labeled features. "Proceedings of the 19th {ACM} Conference on Information and Knowledge Management, {CIKM} 2010, Toronto, Ontario, Canada, October 26-30, 2010.

[17] Rudy Prabowo, Mike Thelwall, "Sentiment analysis: A combined approach." Journal of Informetrics 3 (2009) 143–157.

[18] Rui Xia, Chengqing Zong, Shoushan Li, "Ensemble of feature sets and classification algorithms for sentiment classification", Information Sciences 181 (2011) 1138–1152.

[19] Popescu, A. M., Etzioni, O.: Extracting Product Features and Opinions from Reviews, In Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, 2005, 339–346.

[20] Qingliang Miao, Qiudan Li, Ruwei Dai, "AMAZING: A sentiment mining and retrieval system", Expert Systems with Applications 36 (2009) 7192–7198.

[21] Xing Fang and Justin Zhan."Sentiment analysis using product review data". Journal of Big Data 2015. DOI: 10.1186/s40537-015-0015-2

[22] Hatzivassiloglou, V., & McKeown, K.R. 1997. Predicting the semantic orientation of adjectives. Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL (pp. 174-181). New Brunswick, NJ: ACL.

[23] Tong, R.M. 2001. An operational system for detecting and tracking opinions in on-line discussions. Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification (pp. 1-6). New York, NY: ACM.

[24] Wiebe, J.M. 2000. Learning subjective adjectives from corpora. Proceedings of the 17th National Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press.

[25] Pang, Bo; Lee, Lillian (2005). "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales". Proceedings of the Association for Computational Linguistics (ACL). pp. 115–124.

[26] Yang Liu, Xiangji Huang, Aijun An, Xiaohui Yu (2007). "ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs" SIGIR'07, July 23–27, 2007, Amsterdam, The Netherlands.

[27] R Core Team (2015). R: A language and environment for statistical computing. R   Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[28] Cataldo Musto, Giovanni Semeraro, Marco Polignano."A comparison of Lexicon-based approaches for Sentiment Analysis of microblog posts".8th International Workshop on Information Filtering and Retrieval Pisa (Italy) December 10, 2014

[29] Andrea Esuli Baccianella, Stefano and Fabrizio Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of LREC, volume 10, pages 2200-2204, 2010.

[30] George A Miller. WordNet: a lexical database for English. Communications of the ACM, 38(11):39{41, 1995.

[31] Carlo Strapparava and Alessandro Valitutti. Wordnet affect: an affective extension of wordnet. In LREC, volume 4, pages 1083{1086, 2004.

[32] Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. Language resources and evaluation, 39(2-3):165-210, 2005.

[33] Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. AAAI, Quebec City, pages 1515{1521, 2014.

[34] Erik Cambria and Amir Hussain. Sentic computing. Springer, 2012.

[35] http://wordnet.princeton.edu/

[36] Kamps, Maarten Marx, Robert J. Mokken and Maarten De Rijke, "Using wordnet to measure semantic orientation of adjectives", Proceedings of 4th International Conference on Language Resources and Evaluation, pp. 1115-1118, Lisbon, Portugal, 2004.

[37] Stone, P. J., Dunphy, D. C., Smith, M. S., and Ogilvie, D. M. (1966). The General Inquirer: A Computer Approach to Content Analysis. MIT Press.

[38] Clayton J. Hutto, Eric Gilbert."VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.".Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014.

[39] http://www.sananalytics.com/lab/twitter-sentiment/

[40] Kaiquan Xu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song, "Mining comparative opinions from customer reviews for Competitive Intelligence", Decision Support Systems 50 (2011) 743–754.

[41] Long-Sheng Chen, Cheng-Hsiang Liu, Hui-Ju Chiu, "A neural network based approach for sentiment classification in the blogosphere", Journal of Informetrics 5 (2011) 313–322.

[42] Hanhoon Kang, Seong Joon Yoo , Dongil Han," Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews". Expert Systems with Applications 39 (2012) 6000–6010.

[43] Lin Y, Zhang J, Wang X, Zhou A (2012) An information theoretic approach to sentiment polarity classification. In: Proceedings of the 2Nd Joint WICOW/AIRWeb Workshop on Web Quality, WebQuality '12. ACM, New York, NY, USA.pp 35–40

[44] Federico Neri , Carlo Aliprandi , Federico Capeci , Montserrat Cuadros , Tomas By, Sentiment Analysis on Social Media, Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), p.919-926, August 26-29, 2012

[45] Prashant Raina, "Sentiment Analysis in News Articles Using Sentic Computing", ICDMW, 2013, 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW) 2013, pp. 959-962, doi:10.1109/ICDMW.2013.27

[46] Seyed-Ali Bahrainian, Andreas Dengel, "Sentiment Analysis and Summarization of Twitter Data", CSE, 2013, 2013 IEEE 16th International Conference on Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on Computational Science and Engineering (CSE) 2013, pp. 227-234, doi:10.1109/CSE.2013.44

[47] Emitza Guzman, Walid Maalej, "How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews", RE, 2014, 2014 IEEE 22nd International Requirements Engineering Conference (RE), 2014 IEEE 22nd International Requirements Engineering Conference (RE) 2014, pp. 153-162, doi:10.1109/RE.2014.6912257

[48] Geetika Gautam and Divakar yadav, "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis", 978-1-4799-5172-7/, pp.437-442 2014 IEEE.

[49] Lars Kai Hansen, Adam Arvidsson, Finn Årup Nielsen, Elanor Colleoni,Michael Etter, "Good Friends, Bad News - Affect and Virality inTwitter", The 2011 International Workshop on Social Computing,Network, and Services (SocialComNet 2011).