iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IV **Month of publication:** April 2017

DOI: http://doi.org/10.22214/ijraset.2017.4095

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Twitter Dual Sentiment Analysis (TDSA)

Neethu V M[1], Dr. S. Brilly Sangeetha [2]
[1]M.tech Student, [2]Associate Professor & Head, Department of Computer Science and Engineering
IES College of Engineering Thrissur, Kerala, India

Abstract: ***Nowadays, Most of the people's interested area is social medias. They use social medias to know the reviews about various topics and areas. Sentiment analysis is the one of the technology used in social medias to know the reviews about movies, company products, social topics etc. The most popular technology used for sentiment analysis is Bag of Words (BOW). It is a uni-gram model. Which keep track of each words to decide whether the sentiment belong to positive or negative. But it has limitations due to the polarity shifting. So an alternative technology used for sentiment analysis is Dual Sentiment Analysis (DSA), which means considering two sides of one review, that is original review and its corresponding reverse review and check the probability of whether it belongs to positive class or negative class or neutral. In this paper, the Dual Sentiment Analysis is performed on the data from Twitter. Twitter is one of the social media, where peoples can put their reviews about a specific topic or any areas. So here using technology for sentiment analysis is Twitter Dual Sentiment Analysis (TDSA) for the analysis of twitter datas.***

Keywords: Sentiment analysis, Bag of words, Dual Sentiment Analysis, Twitter Dual Sentiment Analysis.

I. INTRODUCTION

Sentiment analysis is one of the popular and emerging area, it systematically identify and extract the information. So sentiment analysis is also called opinion mining. It has many applications in real life. In social media peoples are communicating through various applications they prefer social medias to exchange information and communicating current trends, news's etc. After sentiment analysis of these data, we can differentiate what is significant and what is insignificant. Data available from social medias has more benefits for analysing the user opinion and their interests, for example measuring the lags and performance on a currently released product. Nowadays, the Internet plays a important role and it has widely influence on each aspect of human beings because of its wide range of resources. In general more people would like to spend their time on the Internet particularly in order to make different types of social groups and then try to talk with each other as regularly to enable the connection between them to become closer. Hence, Social Network Analysis has become an extensively applied method in research and business for searching into the web of relationships on the individual, organizational and societal level. With more and more computing power, the attractiveness of social networking websites such as Face book, Twitter, LinkedIn etc., and Big Data collection techniques, the demand for solid expertise in Social Network Analysis has recently exploded. Sentiment Analysis can be performed on mainly 3 levels these are, document level, sentence level and entity level. A single review about a topic is considered in the document level sentiment analysis. The polarity of each and every sentence is calculated in the sentence level sentiment analysis.

Twitter is a one of the popular social media in real time to express the opinions and interests of a person or group about a particular topic to appear going on a timeline. The message which is displayed on Twitter is called Tweet. There are many users which are made as friends and followings, tweets and their timeline are the main key components of Twitter. The sorted collection of multiple tweets is the timeline. A person can express his view in front of the world in various forms like images, text, videos etc. Because of popularity of Twitter as an information source, it led to development of applications and research in many areas. Twitter is mainly used by various companies to check the reviews of their products and to know their higher valued customers and to reconfigure their products if needed.

II. RELATED WORK

Twitter is popular online social networking media launched in March 2006.It enables users to send and read tweets with about 140 characters length. Currently twitter acts as opinionated Data Bank with large amount of data available used for sentiment analysis. Twitter is very convenient for research purposes because there are very large numbers of messages, many of which are publicly available, and obtaining them is technically simple compared to blogs from the web. Twitter data is collected for analysis using Twitter API. Two widely used approaches used for the same are Machine Learning & Dictionary Based approach. We are using Dictionary Based approach for analyzing the sentiments of data posted by different users. Then polarity classification of this data is done i.e. Tweets collected after analyses are classified into three categories as Positive, Negative and Neutral. Result of this is

©IJRASET: All Rights are Reserved

depicted by using PIE Chart. Sentiment analysis is done by using NLTK toolkit [7].

Early applications of sentiment analysis are mainly focused on classifying movie reviews or product reviews as positive or negative for identifying positive and negative reviews, but many recent applications involve opinion mining in ways that require a more detailed analysis of the sentiment expressed in texts. Which is mainly used by the companies to determine areas of a product that need to be improved by summarizing product reviews to see what parts of the product are generally considered to be good or bad by users? Another application requiring a more detailed analysis of sentiment is to understand where political writers fall on the political spectrum, something that can only be done by looking at support or opposition to specific policies A couple of others applications, like allowing politicians who want a better understanding of how their constituents view different issues, or predicting stock prices based on opinions that people have about the companies and resources involved the marketplace, can similarly take advantage of structured representations of opinion. These applications can be tackled with a structured approach to opinion extraction. One specific application of sentiment in NLP (Natural Language Processing) that can be used for this purpose is sentiment analysis. It can be used to identify and extract subjective information from the information source collected. With all these processes and methods, it is possible to build a system which can extract application dependent information, process it and produce data which can be used for studying and deductions based on the information retrieved. Much of the current opinion mining research has focused on business and e-commerce applications, such as product reviews and movie ratings [5].

There is a methodological framework to collect, pre-process, analyze and map citizen opinion from Twitter in helping the Governments monitor their citizens' current moods is proposed based on the prior works. Naïve Bayes classifier is used to build a sentiment classifier, which employs a variety of features including a specific feature called emoji. This sentiment model out performs the top system in the task of Sentiment Analysis in Twitter in SemEval-2013 in terms of averaged scores. The novel feature emoji has proved to be useful for Sentiment Analysis in Twitter data. It can also apply to real-world tweets and present how Government agencies can track the fluctuation of citizens' moods using mapping techniques [4].

III. SENTIMENT ANALYSIS TECHNIQUES

A. Supervised Approach

Supervised approach is called Machine learning. Machine learning technique utilizes training data to build predictive model. Predictive models such as decision trees, logistic regressions or neural networks are serviced to make prediction on documents which are present outside the training set. This approach has advantage as it is based on learning patterns that are beneficial in making automated and efficient predictions. Also the algorithms are able-bodied of discovering complex and unimagined patterns that would be beyond what a human could wean. However it has drawbacks as large training data is necessary to build the model and consolidating it is time consuming and challenging. A rating is needed to be provided for every document, and if there are attributes of documents it should provide a rating for each of these as well. Another complication unstrings if two different reviewers assign two different sentiment ratings to the same document, then this can involve unexpected errors in building and measuring the performance of model.

B. Unsupervised Approach

Unsupervised approach is known as Natural language processing. Natural language processing (NLP) is an era of artificial intelligence that contributes with automatically extracting meaning from natural language text. It utilizes entities and syntactic patterns in the text to understand its meaning. It also services an amalgamation of language dictionaries, linguistic constructs like parts of speech, noun phrases along with a range of operators. The major role of rule-based methods is that it supplies freedom for the rule developers to use their domain knowledge to devise rules for analysis purpose. Rule-based methods are totally unsupervised and they do not require any training data. This is a main advantage in real-life applications where training data is scant. Additionally it provides the amenity to refine the rules over a time based on the feedback from analysts or subject-matter experts to adjust the models. The major problem with NLP approach is that they need a lot of human involvement in developing the rules and it completely rely on the domain knowledge of rule developers.

C. Basic Concepts

1) *Sentiments:* Sentiment is an emotion of people, that can be express through any social networks. The study of emotions in text can be conducted from two points of view. Firstly, one can investigate how emotions influence a writer or speaker of a text in choosing certain words and/or other linguistic elements or expressions.

©IJRASET: All Rights are Reserved

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

2) *Sentiment Analysis:* Sentiment Analysis is process of systematically identifying and extracting opinions from a piece of text, especially the main aim is to determine the writer's attitude towards a particular topic, product, specific area etc. is positive, negative, or neutral. "Sentiment Analysis is the task of identifying positive and negative opinions, emotions, and evaluations" of peoples.

3) *Dual Sentiment Analysis:* A new model of sentiment analysis called dual sentiment analysis (DSA) is to address problems for sentiment classification. First it proposes a data expansion technique by creating a sentiment reversed review for each training and test original review. Here use mainly two algorithms, dual training algorithm and dual prediction algorithm. First algorithm used to implement dual sentiment is dual training algorithm which is used to make the use of original and reversed training reviews for learning a sentiment training classifier and a dual prediction algorithm to classify the test reviews by considering two sides of one review.

4) *Tweets:* Tweets are short length messages and have a maximum length of 140 characters. This limits the amount of information that the user can share with every message. Due to this reason, users use a lot of acronyms, hash tags, emoticons, slang and special characters. Acronyms and slang such as 2moro for tomorrow and so on are used to keep sentences within the word limit. People also refer to other users using the @ operator. Users also post URLs of web pages to share information. Emoticons are a great way to express emotions without having to say much. Here assigning 1 for positive sentiments, -1 for negative sentiments and 0 for neutral sentiments.

5) *Dictionary of Negative and Positive Words:* The dictionary of negative and positive words is a dataset containing around more negative and positive words. This dataset is used to determine the numeric features of number of negative and positive words in the tweets, based on which sentiment classification is done. The process of stemming, also performed on this dataset, so that it maps to the training and test dataset.

IV. DUAL SENTIMENT ANALYSIS

Propose a simple yet efficient model, called dual sentiment analysis (DSA), to address the polarity shift problem in sentiment classification Figure 1. Show a System of proposed architecture. A data expansion technique is used by creating sentiments-reversed reviews. The original and reversed reviews are established in a one-to-one correspondence & dual training (DT) algorithm and a dual prediction (DP) algorithm correspondingly, to make use of the original and reversed specimen in pairs for training a statistical classifier and make predictions. DSA structure is as polarity (positive-negative) classification to 3-class (positive, negative, neutral) sentiment classification. To curtail DSA's dependency on an external antonym dictionary, we finally develop a corpus-based method for construct pseudo-antonym dictionary. The pseudo antonym dictionary is language-independent and domain adaptive makes DSA model possible to be applied into a wide range of applications.

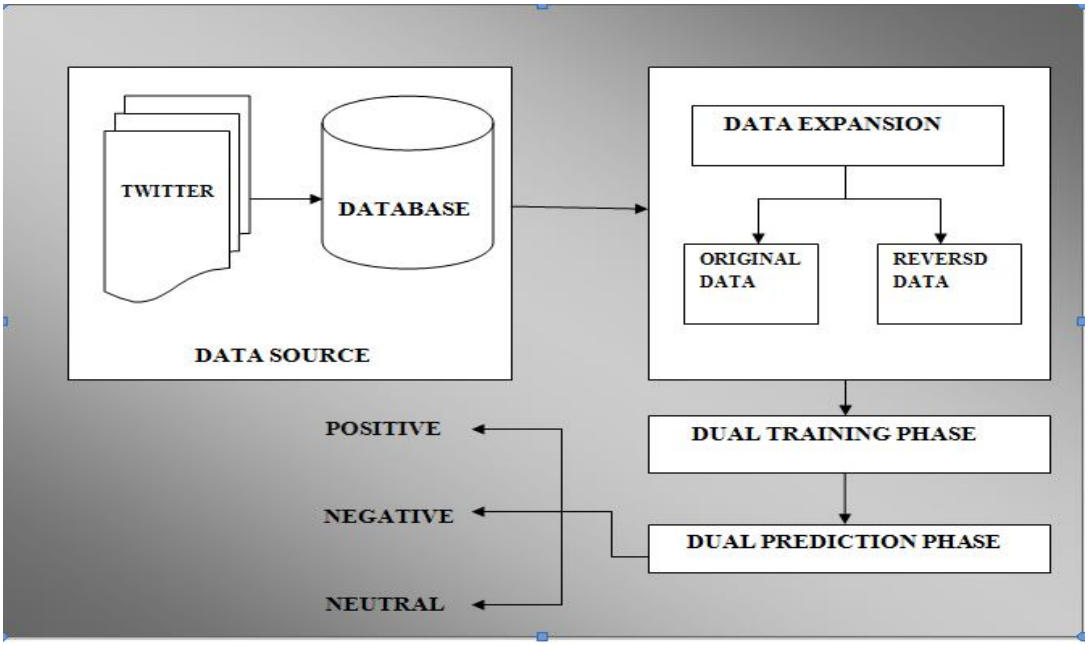


Fig 1: System Architecture

©IJRASET: All Rights are Reserved

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A. Data Expansion Technique

Data Expansion is the first method of dual sentiment analysis, it is based on an antonym dictionary, in this for each original review, and the reversed review is created according to the following rules:

1) *Text Reversion:* If there is a negation, we first detect the scope of negation. All sentiment words out of the scope of negation are reversed to their antonyms. In the scope of negation, negation words (e.g., ―no‖, ―not‖, ―don't‖, etc.) are removed, but the sentiment words are not reversed.
2) *Label Reversion:* For each of the training review, the class label is also reversed to its opposite (i.e., positive to negative, or vice versa), as the class label of the reversed review.

a) Example1:

Original Review: I don't like this picture. It is ugly.
(Negative).
Reversed Review: I like this picture. It is beautiful.
(Positive)

b) Example2

Original Review: The dress looks pretty. But it is costly. (Neutral)
Reversed Review: The dress looks less attractive. But it is cheap. (Neutral)

3) *Corpus based Dictionary:* In information theory, the mutual information (MI) of two random variables is a quantity that measures the mutual dependence of the two random variables. MI is widely used as a feature selection method in text categorization and sentiment classification. First, choose all adjectives, adverbs and verbs in the training corpus as candidate features, and use the MI (Mutual Information) metric to calculate the relevance of each candidate feature to the Positive (+1), Negative (-1) class, Neutral (0) class respectively. Then, rank two groups of features in a decreasing order. For original sentiment words create its corresponding antonym words from antonym dictionary to create reverse sentiment words. Corpus based dictionary is used for data expansion phase.

B. Dual Training Phase

In the training stage, all of the original training samples are reversed to their opposites. We refer to them as "original training" set and "reversed training set" respectively. The original training specimens are reversed to their opposites. Indicate to them as "original training set" and "reversed training set. In our data expansion technique, there is a one-to one correspondence among the original and reversed reviews. The classifier is trained by maximizing a combination of the likelihoods of the original and reversed training samples. This process is called dual training. Note that our method can be easily adapted to the other classifiers such as Naïve Bayes and SVMs.

C. Dual Prediction Phase

Dual prediction works in addressing the polarity shift problem. This time we think "I don't like this book. It is boring" is an original test review, and "I like this book. It is interesting" is the reversed test review. Accordingly, it is very likely that the original test review will be misclassified as Positive. While in DP, due to the removal of negation in the reversed review, "like" this time the plays a positive role. Therefore, the probability that the reversed review being classified into Positive must be high. In DP, a weighted combination of two component predictions is used as the dual prediction output.

Notation	Description
x	The original sample
$\tilde{x}$	The reversed sample
$y \in \{0,1\}$	The class label of the original sample
$\tilde{y} = 1 - y$	The class label of the reversed sample
$D = \{(x_i, y_i)\}_{i=1}^{N}$	The original training set
$\tilde{D} = \{(\tilde{x}_i, \tilde{y}_i)\}_{i=1}^{N}$	The reversed training set
w	Weights of features in a linear model
$J(w)$	Log-likelihood function
$P(.\|x)$	Prediction for the original sample
$P(.\|\tilde{x})$	Prediction for the reversed sample
$P(.\|x, \tilde{x})$	Dual prediction based on a pair of samples

I. Table I: Notations In Dual Training And Dual Prediction

©IJRASET: All Rights are Reserved

V. CONCLUSIONS

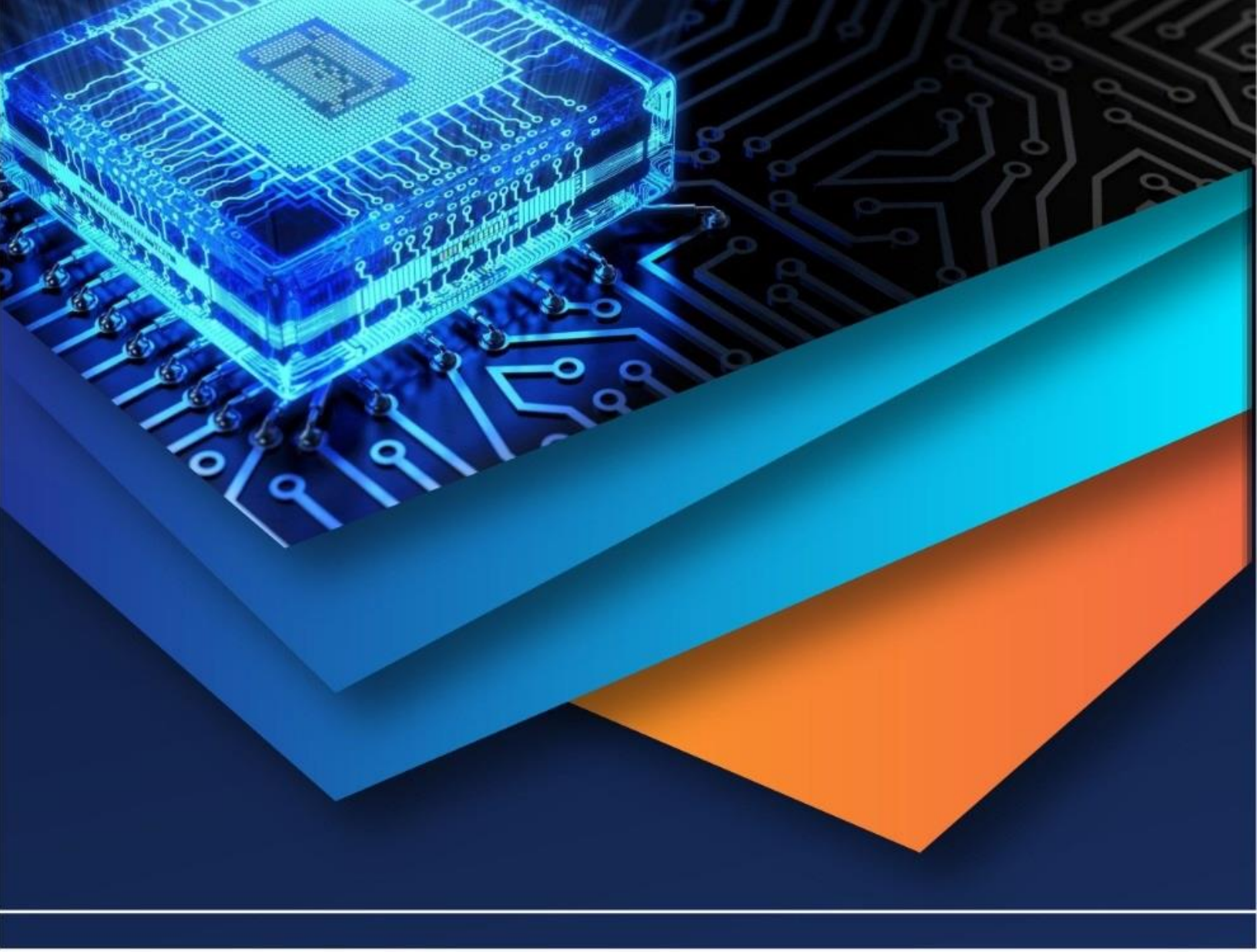
This work is mainly propose a novel data expansion approach, called dual sentiment analysis (DSA), to address the polarity shift problem in sentiment classification. The basic idea of DSA is to create reversed reviews that are sentiment-opposite to the original reviews, and make use of the original and reversed reviews in pairs to train a sentiment classifier and make predictions. DSA is highlighted by the technique of one-to-one correspondence data expansion and the manner of using a pair of samples in training (dual training) and prediction (dual prediction). A wide range of experiments demonstrate that the DSA model is very effective for polarity classification and it significantly outperforms several alternative methods of considering polarity shift. In addition, we strengthen the DSA algorithm by developing a selective data expansion technique that chooses training reviews with higher sentiment degree for data expansion.

Sentiment analysis is essential for anyone who is going to make a decision. Sentiment analysis is helpful in different field for calculating, identifying and expressing sentiment. It is helpful for everyone when they want to buy a product and they can decide which product is best. Sentiment analysis is very important for Enterprises and helps them to know what customers think about their products. Therefore companies can take decisions about their products based on customer's feedback Thus companies can modify their products features and introduce new products according to customers' opinions in a better and faster way. This dual sentiment analysis can be performed on twitter datas. So the companies can easily check their reviews about their products

REFERENCES

[1] Rui xia, Feng Xu, Chengqing zong, Qianmu Li, Yong Q, "Dual sentiment analysis considering two sides of the one review", IEEE transactions on 2015.
[2] R. pormima, R. roopadevi., "Sentiment Analysis of two sides on review using Dual prediction", ISSN 2321 3361 2016 IJESC
[3] Kishori k pawar, Pukhraj P Shrishrimal, R. R. Deshmukh, "Twitter sentiment analysis a review," International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015 957 ISSN 2229-5518.
[4] Mengdi Li1, Eugene Ch'ng, Alain Chong, Simon, "The new eye of smart city : Novel citizen sentiment analysis in Twitter" International Doctoral Innovation centre, University of Nottigham china, IEEE 2016.
[5] Ravendra Ratan singh jandail, "A proposed Novel approach for sentiment analysis and opinion mining", international Journal of Ubi Comp vol 5,2014
[6] S. Fujita and A. Fujino, "Word sense disambiguation by combining labeled data expansion and semi-supervised learning method," Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), pp. 676-685, 2011.
[7] Prerna mishra, Dr ranjana Rajnish,Dr.pankaj kumar, "Sentiment Analysis of Twitter data: Case study on digital India", international conference on information technology,IEEE 2016.
[8] Alaxander park, Patrick paroubek, "Twitter as a corpus for sentiment analysis and opinion mining", Universite de paris-sud laborotaire LIMSI-CNRS,Batiment 508,201
[9] Akshikumar ,teeja mary Sebastian," Sentiment Analysis on Twitter", International journal of computer science issues, vol 9,issue 4, July 2012.
[10] Ahamed Abbasi, Stephen France, zhu zhang,, Hsinchun Chen, "Selecting Attributes for sentiment classification using Feature selection network" IEEE transactions 2011.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)