

# Techniques and Tools Used for Sentiment Analysis using Twitter Data

Sparsh Rastogi<sup>1</sup>, Riya Goyal<sup>2</sup>

<sup>1</sup>B. Tech Scholar, Galgotias University, Uttar Pradesh, India

<sup>2</sup>B. Tech Scholar, Galgotias University, Uttar Pradesh, India

\*\*\*

**Abstract:** -The era of web 4.0 has led to an exponential growth of the data being posted and shared on the internet. The semantic linkage of data shows its importance for early decisions for improving the services and demand supply chain. It requires extensive analysis of the opinions/ postings that have been posted on social media such as Twitter and Google+. Various social sites have emerged like Facebook, Twitter that allows the people to share their views over any topic freely with any hesitation. Twitter is one of the most used microblogging site used all over the world by a large amount of users to express their views on anything or everything in the form of tweets. Tweets are combination of text, emoticons, hashtags etc. of small length expressing high opinion or expressions. But, the major challenge in such data analysis starts with fetching a collective opinion of people using general methods such as polls and surveys etc. The study related to 'Twitter' analysis has already been done and prominently known as sentiment analysis. The analysis of twitter data would be helpful in understanding people's point of view on a particular product, services, personality, organisation, situation, regions, government actions, decisions etc. Our work mainly highlights the tools and techniques widely used for sentiment analysis using 'Twitter' data. We covered from the techniques used for tweets extraction to its classification

**Keywords-** Twitter, Microblogging

## 1. Introduction

Sentiment is a feeling of like or dislike towards an object, topic or situation . Sentiment analysis or opinion mining, is the study of sentiment people express towards something [4]. Sentimental analysis of twitter can be used to analyse the sentiments of the user tweets that have been posted by the users. It will categorise the reviews as positive or negative.

It can be implemented in python and R but here we will mention certain libraries and toolkits that will work along with python. Sentiment analysis can be said as a sub field of NLP whose application range from different fields be it technical to political [9]. Natural Language Toolkit (NLTK) is an integral part of this process as this process deals with textual data. Tweets can then be categorised into three classification namely positive, negative and neutral classification.

Decision making in sentiment analysis actually requires information about what do people think [3]. As we move to new era of technology, more and important information would be required by the people for doing their task. To utilize their time efficiently, organizations have made fast and reliable search engines which are capable of giving results on the basis of keywords what users provide for e.g. Google, YouTube, Facebook and they can move ahead in this field to satisfy their customers. But as there are certain restrictions on these social media platform as what to upload and write, people might sometime not be able to express their opinions on free will. So as to minimize these issues, government gave access of twitter to everyone through which they can share their views by write a tweet related to any topic. If the blog provides illegal or wrong information or gives useful data, then that microblog can be identified by the research team on the basis of sentimental analysis based on which they can be rewarded or punished by law.

With a tremendous pace, the blogs are uploaded on twitter also known as real time blogs. So to categorize them on their semantic concepts and sentiments is the greatest challenge. To make it possible, classification of our blogs as per their text is necessary which is done by sentimental analysis of twitter. Sentiment analysis is the process of studying emotions which have been presented in the form of text [5]. Till now the data has been fetched from review sites which give the reviews of movies, songs, applications only but as sentimental analysis of twitter came then the data is being fetched from real time blogs as well as reviews sites which increases the scope of getting data through which we can provide almost accurate information for our users. Many people have discussed various models performing sentiment analysis of twitter. Apporv Agarwal, Boyi Xie, Iliia Vovsha, Owen Rambo, Rebecca Passonneau [1] have discussed a tree kernel based model and a feature based model and compared it with the traditional unigram model and found out that both the models outperformed the traditional model.

Lisa Branz and Patricia Brockmann [2] collected two set of tweets, first were based on the software engineering and second were based on people working on software engineering that could be related to any topic. They used WEKA for sentiment analysis of these data sets and tried to calculate specific emotion for every tweet. They also calculated the gender specific scores for the tweets they collected.

Shihab Elbagir and Jing Yang [10] have used twitter streaming API for getting the data and have used Machine Learning Techniques and SciKit for the classification. They have used the TF-IDF model for the feature extraction purpose combined with n gram technique ranging from 1 to 3. Multiple classification algorithms were used such as multinomial naïve bayes classifier, SVC, Bernoulli Naïve Bayes etc. Later the accuracies were calculated for each algorithm.

Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, Ming Zhang [11] have worked on sentiment analysis of twitter data using a graph based hashtag sentiment classification approach. They have illustrated that three types of information can be used for the classification of twitter data i.e. polarity of the sentiment of tweets containing hashtags, relationships and co-occurrences of hashtags in tweets and the literal meaning of hashtags. Later they have shown by their work that the performance can be improved drastically if the literal meaning of the hashtags is considered.

Brinda Hegde, Nagshree HS, Madhura Prakash [12] have performed sentiment analysis using machine learning techniques. They took the demonetization tweets as the data and the preprocessing was done using NLTK and SciKit learn and then classification algorithms such as Naïve bayes, Logistic Regression and Support Vector Machine were used and their accuracies were calculated.

## 2. LITERATURE SURVEY:

Sentimental analysis has its base working upon Natural Language Processing Toolkit combined with various scraping tools and classification algorithms. Processing can be performed on many small features to achieve classification (Positive, Negative and Neutral). Real time microblogs which contains negative as well as positive speeches, emoticons are being posted in every passing second and collecting those tweets will come up as a new and difficult challenge. Some of the blogs include emoticons, hashtags which help in displaying a strong opinion.

### 2.1 Related Work:

Using sentimental analysis on twitter is the latest trend among the researchers checking the potential applications and science based trials [6]. Challenges are not something in which we have to use our brains but it is a blogging challenge to find the users perspective, emotions, feelings, logical thoughts and so on and so forth the basis of their microblogs.

- There are different users using twitter from different locations.
- There are wide range of blogs written every second which makes our data more thorough.
- Microblogging is used by different people from actors, ministers, entrepreneurs to each layman for displaying their opinions on different task with different. With this we get large amount of data.

J. read [7] has worked on the use of emoticons to display opinions or sentiments in tweets as he states in one of his works "using emoticons to reduce dependency in machine learning techniques for sentimental classification". This means by detecting the emoticon we can easily and quickly find the sentiments of the user comparatively to what they have written because if he/she writes a blog then we need machine learning algorithms with a proper procedure and depicted models prepared by using Naive Bayes' classifier and support Vector Machine(SVM).

The same kind of work is also done by A.pak and A. Paraoubeck [8] who states that "twitter as a corpus for sentiments analysis and opinion mining". The only difference is that they can do their work as fast as J. read but classify the blogs into positive and negative classification. They extract the data from twitter which was written by famous newspapers like "Washington posts" and the "The New York Times". This classification was done by Naive Bayes' classifier as well as N-gram and POS-tags as features.

Our research paper is based on the combination of various techniques that can be used for the sentiment analysis of twitter.

## 3. DATA DESCRIPTION:

Twitter is a social networking site and microblogging site in which users can write or in the twitter language 'tweet' anything that comes to their mind. It is a real time networking service where users can write messages which is known as tweets. To write any tweet we have a word limit of 140 lengths of words so people make large use of short forms, acronyms, emoticons, hashtags and the slang words which becomes a challenge for a model to classify. Microblogging are short and fast messages. We can use two resources for pre-processing of twitter data. Firstly, an emoticon dictionary in which we provide multiple emoticons and add labels to classify the emoticon as positive or negative like 😊 can be given

the label of positive and ☹ this can be given the label of negative. Secondly we can create an acronym dictionary containing acronyms which can be fetched easily from online sources in which we can provide large amount of acronyms such as IDK as I don't know, LOL as laugh out loud, bff as best friends forever, RIP as rest in peace and many more.

Emoticons	Polarity
☺ :3 :c) :D	Positive
☹ :c :[	Negative

Certain aspects of tweets are as follows: -

- 1) Length of tweets: To write tweets, there is a word limit of 140 characters which differentiates it from previous sentimental analysis and also different from reviews based on films and product reviews.
- 2) Way of Writing: The availability of repeated characters, acronyms, emoticons and misspelled words are more in tweets comparatively to other messages due to short length of tweets.
- 3) Availability: The availability of data in public domain of twitter is more as compared to privacy settings of Facebook thus the data is found at all time at ease.
- 4) Regular Update: The blogs cannot be changed or updated for a long period of time because they are long in size while comparatively tweets are short in size and many people write their views every minute or a second which makes the tweets real time. That's why tweets are real time messages which are updated regularly and give the real time feeling.

There are few terminologies which are associated with tweets as follows: -

- Emoticons: Emoticons are used to define the user mood so precisely that it can't be expressed in words. Emoticons are nothing but facial expressions showed by pictorial representation using letters and punctuation. It helps the user to show their feelings.
- Tagging: It is a fashion for today's people to tag someone as to increase their views on its post or to spread their news around everyone. It is done by "@" symbol which targets the user you want.
- Hashtags: Hashtags are basically used to increase their views on their tweets or blogs. Its main function is to mark topics.
- Acronym: Just because of the limited size of tweets people use shortcuts in words to express more by writing less. This is in trend also like if people wants to say take care then they just write 'TC', as soon as possible as 'ASAP' etcetera and sometimes user uses special symbols.

The tweets were written in twitter by users from all over the world in many different languages. All the data is stored at a one place which is called as twitter storage through which we can access large amount of data which is available publically because tweets can be written in public domain field and in real time so we always have updated data. Due to the different languages of tweets, twitter use googles translate to convert the data into a common language. They can classify the data as positive, negative and neutral by human detector methods. The tweet which does not come under any category of classification would be labelled as junk which does not come under any sentimental analysis. [1]

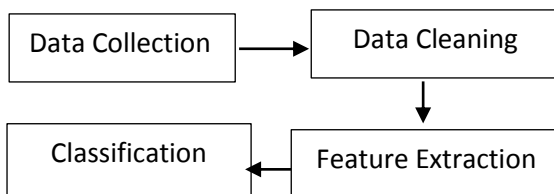
Table 1. Summary of researches selected for review

Author	Data Used	Amount of data used.	Feature extraction techniques used	Classification techniques used	Accuracy
Niketan Jivane 20 Nov, 2018.[13]	Movie Reviews	Training data-1800 Testing data-150	Bag of words	Naïve Bayes SVM	87.5% 90.71%
Shihab Elbagir, Jing Yang [10]	Publically available twitter	Training data-6000 Testing data-	TF-IDF	Naïve Bayes	71.02%

	datasets	4000			
Ali Hasan, Sana Moin et. Al.[14]	Election Tweets	Training data-1690 Testing data-400	Bag of words	Naïve Bayes SentiWordNet	79% 54.75%
Geetika Gautam [15]	Publically available twitter datasets		Bag of words	Naïve Bayes SVM Lexicon approach	82.2% 85.5% 89.9%
Apoorv Agarwal et.al[1]	Manually annotated tweets	11875 tweets	Word Net and POS tagging	Unigram Senti features	56.58% 56.31%

#### 4. SYSTEM ARCHITECTURE:

This section discusses the techniques of various steps involved in the process of sentiment analysis of twitter data.



4.1 DATA COLLECTION: - It can be defined as the process of collecting data from twitter for the purpose of development of the project. Now there are two techniques of collecting data or scraping of data. The first technique is useful for the people of non-development background.

There are various automated tools present that can help the users to collect real time data from twitter directly without writing codes.

Some of these tools are as follows: -

*Zapier*- This platform enables teams from different departments to use two apps that they require on daily basis simultaneously. You just need to choose a trigger app (from which data needs to be collected) and action app (where you store the data). So you can use this to store data from twitter to spreadsheets.

*Tweet Download*- It allows you to download tweets, replies and mentions from your own account.

As for the users with developing or technical background, the widely used software for accessing public data from twitter is Twitter API. The developers can write their own code to interact with the API or can use the libraries provided by different languages as per their comfort.

Twitter API provides two set of API: -

*Twitter Streaming API*- It enables the user to have a flow of tweets for a searched keyword or hashtag in real time. It has a free as well as a paid version.

*Standard Search API*- It enables a user to search for historical data related to a particular keyword or hashtag. It can provide data for up to seven days ago.

Twitter API is a python wrapper for executing the requests. Tweepy library can be used to connect to the twitter streaming API and collect the tweets in real time using python.

**4.2 DATA CLEANING:** - Now the data collected contains various noise in the form of punctuations, emoticons, hashtags, tags, url's etc. We can remove the punctuations, replace emoticons with the labels from the emoticon dictionary, replace acronyms by the labels from the acronym dictionary and mark tags and url by specific symbols such as T or U respectively. Then we remove the stop words from the text which are conjunctions, prepositions etc. as they do not provide any meaning to the opinion. NLTK for python provides a data dictionary containing list of stop words in 16 different languages. Then we need to perform tokenization. It is the technique of splitting up of data into sentences or words known as tokens. NLTK provides tokenization package containing various methods for the process.

**4.3 FEATURE EXTRACTION:** - It is the process of dimensionality reduction of data to extract features useful for the classification. These features can be fed to the classification algorithms to get the required results. The machine learning classification algorithms cannot directly work on textual data therefore it needs to be converted into numbers for the working of algorithms thus feature extraction is necessary as it gives data that can be fed to the machine learning algorithms.

There are multiple techniques that can be used for feature extraction. Some of them are as follows: -

**POS tagging-** It can be done with the help of NLTK package and WordNet dictionary which is openly available for download. Part of speech tagging is necessary as the text is made up of multiple parts of speech such as nouns, pronouns, verbs, adverbs, adjectives etc. but not all of them are required for displaying sentiments or opinions like nouns and pronouns do not display any kind of feeling or opinion thus part of speech tagging helps us to select adjectives, verbs, adverbs from the text that are used for displaying opinions or sentiments.

**Bag of Words model-** It is one of the simplest model that can be used for feature extraction. It consists of two parts, a vocabulary of words and measure of words present. The basic idea is to count the number of times a word occurs in a document which can be used as weight or feature for classification. You can replace the count of words by calculating frequency for a word, or by simply representing the presence of word in a document in binary format. You can decide how complex you want your model to be.

**TF-IDF model-** The term frequency inverse document frequency is one of the most widely used model for feature extraction in terms of sentiment analysis. It assigns importance to word present in document. First, (t) the term frequency of a word is calculated as the total number of occurrences of a word in the document divided by the total number of words in the document. Inverse document frequency (IDF) calculates the importance of a word in the document by the formula  $IDF(t)=\log(N/DF)$  where N is the total number of documents and DF is the total number of documents in which the terms t is present.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}}$$

$$IDF = \log \left( 1 + \frac{\text{Total Documents}}{\text{Documents with Keyword}} \right)$$

**N-gram model-** This model can be combined with bag of words model for better results. A N gram model breaks a sequence of text into tokens of size N. N can hold any value 1, 2, 3..... For example, consider a sentence "I saw a beautiful dress in the store". If we use value of N as 1, it is called a unigram and it will produce tokens as 'I', 'saw', 'a', 'beautiful', 'dress', 'in', 'the', 'store'. If we use the value of N as 2 then it will produce the tokens as 'I saw', 'saw a', 'a beautiful', 'beautiful dress', 'dress in', 'in the', 'the store'. As it is clear from the example that bigrams or n grams are more useful than unigrams as 'beautiful dress' will make more impact than just 'dress'.

**4.4 SENTIMENT CLASSIFICATION:** - It can be done by three approaches. First is the use of machine learning classification algorithms such as Naïve Bayes, SVM, decision tree etc. second is the lexicon based approach which uses a pre compiled list of known sentiment terms and third is the deep learning approach.

**Machine learning approach: -**

**Naïve Bayes classifier-** This algorithm uses a probabilistic approach for the classification. Its base is the Bayes theorem. Naïve Bayes theorem works on the assumption of feature independency that is predictors or features are independent of

each other as presence of any feature does not have any kind of effect on any other feature therefore the algorithm got its name as naïve.

$$\hat{y} = \underset{y}{\operatorname{argmax}} P(y) \prod_{i=1}^n P(x_i|y)$$

There are various forms of Naïve Bayes algorithms present such as multinomial, Bernoulli, Gaussian but the algorithm used for classification of twitter is multinomial naïve bayes theorem.

*Support Vector Machine*- It is a supervised machine learning technique. SVM is known to perform quiet well in the case of sentiment analysis. SVM performs the classification by plotting the features in multiple dimensions and draws a hyperplane between them that separates both of the classes. The hyperplane is chosen such that the distance of nearest vectors of both the classes is maximum. This is known as optimal hyperplane.

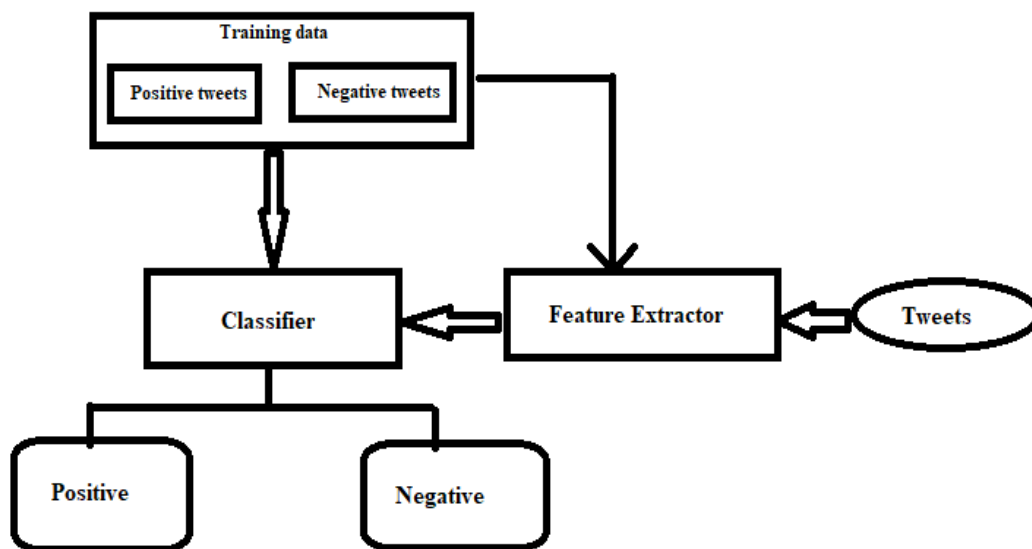


Fig.1 Machine learning based classification framework.

*Lexicon based approach*: -

*Dictionary based*- As the name suggests; this approach makes use of a dictionary of words. First of all, you would need to make or get a dictionary of positive as well as negative words. Then you can compare the words in your text to be analysed with the words present in the dictionary. If the word is found then the score is assigned, if the word is not found then you look for a synonym on the WordNet dictionary and match the synonyms with the word in your dictionary and assign the same score as that of the synonym. Then you can update your dictionary with the new word and assign a score. Later you can compute an average score for the entire text.

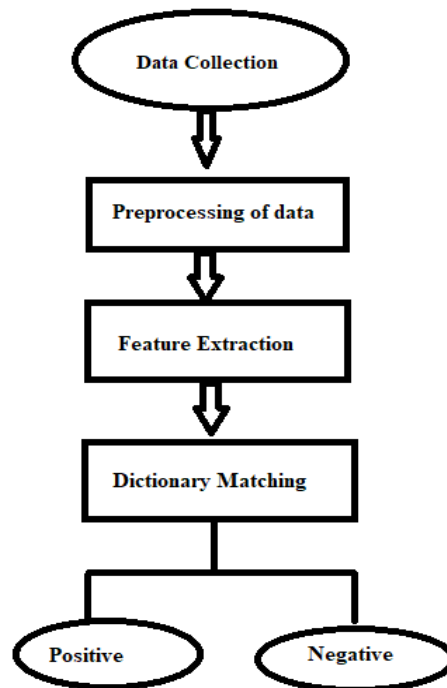


Fig.2 Dictionary based classification framework

Deep Learning based approach: -

Long Short Term Memory- These are a type of RNN (Recurrent neural network) This network is capable of remembering the sequence of the past data, which in the case of sentiment analysis of twitter is words and make a decision on the sentiment of the word. It is a sequence of copies of a cell and where output of each cell is forwarded as input to the next.

## 5. DISCUSSION

In this section we will discuss the internal working of the system i.e. how a new tweet is classified as positive or negative with the help of example. We will use the bag of words model along with naïve bayes classifier for our understanding.

Our aim is to classify a document (tweet) as positive or negative. Now, as we know bag of words model calculates the frequencies of each occurring word. To solve the problem using naïve bayes classifier, we need to divide the data set into two halves, a training dataset and a testing dataset. The training dataset needs to be labelled as positive or negative to help train the classifier to label the new incoming tweets. Some terminologies that will be used in this section are:-

P(+) which means the probability of word being positive

P(-) which means the probability of word being negative

Pdoc(+) which means probability of tweet being positive

Pdoc(-) which means probability of tweet being negative

We need to have a large training dataset and to train the algorithm. Here, I am taking a small example for better understanding of the process. Say for example we have a set of labelled movie reviews like: -

Doc	Movie Review	Label
1	I loved the story	+
2	I disliked the story	-
3	An awesome story, exciting story	+
4	Bad acting	-
5	Awesome acting, an exciting story	+

So, now we apply bag of words model on this data set. So the first task is to make a list of all the unique words that appear in the data set. Unique words list [I, Loved, the, story, disliked, An, Awesome, Exciting, Bad, acting] Now the next step is to create feature set. It can be represented in the form of an array with attributes as all the possible words and their values are the number of times these words occur as shown in Table 2.

Docs	I	Loved	The	Story	Disliked	An	Awesome	Exciting	Bad	Acting	Label
1	1	1	1	1							+
2	1		1	1	1						-
3				2		1	1	1			+
4									1	1	-
5						1	1	1		1	+

Table 2. Feature set for the given example

Now, applying naïve bayes to train the model, the first task is to calculate the probability of the positive tweets and the negative tweets so for this example  $P(+)=3/5=0.6$  and  $P(-)=2/5=0.4$ .

Now we need to calculate the probability of each word occurring as positive and then each word occurring as negative. The formula is

$$P(w_k|+) = \frac{nk+1}{n+|vocabulary|}$$

Where nk is the number of times the word k occurs in the particular case (+ or -). n is the total number of words in the particular case and vocabulary includes the total number of unique words. A 1 is added to nk so that is in any case the nk is 0 then the probability does not become zero. This is known as Laplace Smoothing. So the next step is to calculate the probability of each word as positive and negative where total number of words occurring in positive are 14 and total number of words occurring in negative are 6.

So the positive and negative probabilities of each word are calculated as Table 3.

Table 3. Probability of each word

	I	Loved	The	Story	Disliked	An	Awesome	Exciting	Bad	Acting
+ve	0.0833	0.0833	0.0833	0.2083	0.417	0.125	0.125	0.125	0.417	0.0833
-ve	0.125	0.0625	0.125	0.125	0.125	0.0625	0.0625	0.0625	0.125	0.125

Now, as our classifier is trained, we are capable of classifying a new tweet.

For example, we have a new tweet as "I disliked the bad acting".

To classify it as positive or negative, we have the formula

$$V_{nb} = \text{argmax } P(v_j) \prod P(w|v_j)$$

According to the formula, we need to calculate the probability of one case (positive or negative) into the probability of all the words belonging to the case. We need to do this for both the positive and negative cases. The argmax function will chose the maximum of both values and that label will be assigned to the new tweet.

If  $v_j$  is + then  $P(+)*P(I|+)*P(disliked|+)*P(the|+)*P(bad|+)*P(acting|+)= 6.03 * 10^{-7}$

If  $v_j$  is - then

$P(-)*P(I|-)*P(disliked|-)*P(the|-)*P(bad|-)*P(acting|-)= 1.22 * 10^{-5}$



Now we can see that the probability of  $v_j$  as negative is greater than the probability of  $v_j$  as positive therefore our new tweet is now classified or labelled as negative.

## 6. CONCLUSIONS

Twitter data analysis has been done to find or to categorise the opinions. Here in this paper, we have discussed various techniques that have been employed for this purpose. Multiple authors have discussed the working and criticalities of the different method to analyse the twitter data depicting advantages and drawbacks of multiple techniques. We found at that though bag of words is the majorly used technique for feature extraction, TF-IDF presents an enhanced and advanced option now for the same but the bag of words technique is comparatively more flexible as it can be adjusted according to the developers needs and complexity requirements.

Though SVM proves to give more accurate results as in *table 1* studies done by multiple authors prove that naïve bayes is the fastest algorithm for sentiment classification. In terms of sentiment classification, naïve bayes and SVM prove more effective on large data only as then they can be trained more accurately and give better results but on smaller data, they might prove inaccurate while lexicon based approach though easier is suitable for smaller data as it needs a dictionary to be created. Large data would require a larger dictionary to be created with high vocabulary and still many words could remain unaccounted or else it would result in a degraded performance.

## 7. FUTURE WORK

As many people have already been successful in categorising tweets as positive, negative or neutral, research can be done as to specifically categorise tweets with a particular emotion. It should consider the potential effects of the presence of hashtags, tags in the tweets and how can they contribute more in the classification of tweets more accurately. Hybrid models that combine various techniques can be used and tried to increase the accuracy and decrease the time. A detailed research can be done to dig past the classifications and understand the relevance of interaction between users and then try to depict the nature of the users.

## 8. REFERENCES

1. Sentiment Analysis of Twitter Data Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau Department of Computer Science Columbia University New York, NY 10027 USA.
2. Poster: Sentiment Analysis of Twitter Data: Towards Filtering, Analyzing and Interpreting Social Network Data Lisa Branz Department of Computer, Patricia Brockmann Department of Computer Science
3. Opinion Mining and sentiment analysis Bo Pang Yahoo! Research Sunnyvale, CA 94089 USA and Lillian Lee Computer Science Department Cornell University Ithaca, NY 14853 USA
4. Sentiment Analysis using product review Xing Fang and Justin Zhan Journal of Big Data.
5. W. Medhat et al. Sentiment analysis algorithms and applications: A survey, Ain Shams Eng J (2014), <http://dx.doi.org/10.1016/j.asej.2014.04.011>
6. Sentiment Analysis on Twitter Akshi Kumar and Teeja Mary Sebastian Department of Computer Engineering, Delhi Technological University Delhi, India. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012.
7. Jonathon Read. ACL student '05: Proceedings of the ACL Student Research Workshop June 2005
8. Twitter as a Corpus for Sentiment Analysis and Opinion Mining Alexander Pak, Patrick Paroubek University's de Paris-Sud, Laboratoire LIMSI-CNRS, Bâtiment 508, F-91405 Orsay Cedex, France
9. Sentiment analysis of twitter. Retrieved from monkeylearn.com

10. Sentiment Analysis of Twitter Data Using Machine Learning Techniques and Scikit-learn Shihab Elbagir† College of Computer Science and Technology Harbin Engineering University Harbin China, Jing Yang College of Computer Science and Technology Harbin Engineering University Harbin China
11. Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, Ming Zhang Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach
12. Sentiment analysis of Twitter data: A machine learning approach to analyse demonetization tweets. Brinda Hegde, Nagshree HS, Madhura Prakash. International Research Journal of Engineering and Technology (IRJET). June 2018
13. Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques. Akshay Amolik, Niketan Jivane, Mahavir Bhandari, Dr.M.Venkatesan International Journal of Engineering and Technology (IJET)
14. Machine Learning-Based Sentiment Analysis for Twitter Accounts Ali Hasan 1 , Sana Moin 1 , Ahmad Karim 2 and Shahaboddin Shamshirband 3
15. Geetika Gautam,, Yadav, D.: Sentiment analysis of twitter data using machine learning approaches and semantic analysis. 2014 Seventh International Conference on Contemporary Computing (IC3). (2014).