International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 5 517 – 521

Twitter Sentiment Analysis of Current Affairs

Mahavar Anjali B
Dept. Computer Science & Engineering
Parul Institute of Technology, Parul University
Limda, Vadodara, India
me.anjalimahavar@gmail.com

Priya Pati
Dept. Computer Science & Engineering
Parul Institute of Technology, Parul University
Limda, Vadodara, India
Priya.pati@paruluniversity.ac.in

Abhishek Tripathi
Dept. Computer Science & Engineering
Parul Institute of Technology, Parul University
Limda, Vadodara, India
Abhishek.tripathi@paruluniversity.ac.in

Abstract— Sentiment Analysis is an important type of text analysis that aims to support judgment making by extracting &analyzing opinion oriented text. Identifying positive & negative opinions & measuring how positively & negatively an entity is regarded. sentiment analysis on social media data while the use of machine learning classifier for predicting the sentiment orientation provide a useful tool for users to monitor brand or product sentiment. File level sentiment analysis is used which consists of Term Frequency (TF) and Inverse Document Frequency (IDF) values as features along with Fuzzy Clustering which results in positive and negative sentiments. As more & more user articulate their views & opinion on twitter. So twitter becomes valuable sources of people's opinions. Tweets data can be used to infer people's outlook for marketing & social studies. Twitter sentiment analysis that can stain the general people's opinion in regard to social event which are going to be in current on twitter. In this research will take current scenarios which are going to be on twitter as an example for sentiment analysis. In these will use the proposed feature extraction model with emoticons and Synonym using SVM classifier. Using this can obtain greater accuracy as compared to previous research work. This research is the comparative analysis with different classifiers to identify public's opinion.

Keywords- *Data mining, Sentiment analysis, Emoticon, Naïve bayes, SVM.*

*****

I. INTRODUCTION

Sentiment Analysis is the task of identifying whether the opinion expressed in a text is positive, negative or neutral about specific given topic. **E.g. "I am so happy today, good morning to everyone"** , is general positive text & the text is : **" kabali is such a good movie highly recommended by 9/10 ",** express positive sentiment towards the movie , named kabala, which is considered as the topic of this text.Sometimes identifying the exact sentiment is not so clear even for humans. **E.g. " I am surprised so many people put kabali in their favourite film ever list, I felt it was a good watch but definitely not that good".** The sentiment expressed by the author towards the movie is probably positive but not as good as in the message.

This paper propose to combine many feature extraction techniques like emoticons, exclamation and question mark symbol, word gazetteer, unigrams to design more accurate sentiment classifications system. This paper presents empirical comparisons of six supervised algorithms that is Naïve Bayes, Bayes Net, Discriminative Multinomial Naïve Bayes, Sequential Minimal Optimization, Hyper pipes, Random Forest. This paper used the unigram and word gazetteer method with feature extraction [1].This paper proposes feature engineering and Dynamic Architecture for Artificial Neural Networks. This paper used different five methods in feature engineering 1) frequency analysis 2) affinity analysis 3) negation and valance shifter analysis 4) feature sentiment scoring 5) aspect categorization. This paper classified tweets into five categories that is strongly positive, mildly positive, neutral, mildly negative, strongly negative [2].This paper proposes a feed forward neural network (NN) for sentiment analysis. In this, tweets are collected from Twitter API. The average accuracy of this paper is 74.15% [3].In this, first analysis surveyed a group of participants for their perceived sentiment polarity of the most frequent emoticons. The second analysis examined clustering of words and emoticons to better understand the meaning conveyed by the emoticons. The third analysis compared the sentiment polarity of micro blog posts before and after emoticons was removed from the text [4].we first pre-processed the dataset, after that extracted the adjective from the dataset that have some meaning which is called feature vector, then selected the feature vector list and thereafter applied machine learning based classification algorithms namely: Naïve Bayes, Maximum entropy and SVM along with the Semantic Orientation based Word Net which extracts synonyms and similarity for the content feature [5].In this [6], paper propose a simple and completely automatic methodology for analyzing sentiment of users in Twitter. Firstly, builds a Twitter corpus by grouping tweets expressing positive and negative polarity through a completely automatic procedure by using only emoticons in tweets.In this paper [7], use the different machine learning techniques for classifying tweets. Sentiment analysis in twitter is difficult to due to its sort lengths, presence of slang words, emoticons and misspellings in tweets.

II. METHODOLOGY

In our system, we have proposed a methodology that is divided into different stages as shown in Figure 3.1. The five stages are as follows:

1) Collection of tweets
2) Pre-processing
3) Feature Extraction
4) Classification
5) Predictive Analysis based on Application Related

1. Collection of tweets

For our system, we gathered our dataset by consulting the Twitter API and making use of word spotting based on occurrence of the word we are querying the recent tweets.

2. Pre-Processing

The procedure for pre-processing consists of the following steps:

i. Removing all non-English Tweets.
ii. Converting all the tweets collected to the lower case.
iii. Removing the URLs – erased all string that describes links or hyperlinks present in the tweets.
iv. Replacing any usernames present in the tweets to @username – removed the username and because these are not considers for sentiments.
v. Removing any unnecessary characters, extra spaces etc.
vi. Remove all the number from tweets and also remove all words which don't start with an alphabet, for example 9th, 9:15am.
vii. Removing punctuation like commas, single/double quotes question marks, etc. at the beginning and end of each word in a tweet. E.g. Happy!!!!!! Replaced with Happy.
viii. Converting the hash tags to normal words because hash tags can provide some helpful information, so it is useful to replace them with the literally same word without the hash. E.g. #Happy replaced with Happy.

3. Feature Extraction

i. **Use of Negation Method:** The appearance of negative words may change the opinion orientation like not happy is equivalent to sad.

ii. **Use of Unigram Model:** The feature extraction method, extracts the aspect (adjective) from the dataset. Later this adjective is used to show the positive and negative polarity in a sentence which is useful for determining the opinion of the individuals using unigram model. Unigram model extracts the adjective and segregates it. It discards the preceding and successive word occurring with the adjective in the sentences. For above example, i.e. "Driving Happy" through unigram model, only Happy is extracted from the sentence. Once the tweets are filtered, the output of the feature extractor is a list of the feature words present in the tweet.

4. Classification

i. Support Vector Machines Classifiers (SVM):

The main principle of SVM is to determine linear separators in the search space which can best separate the different classes. In figure 3.2 there are 2 classes x, o and there are 3 Hyper planes A, B and C. Hyper plane A provides the best separation between the classes, because the normal distance of any of the data points is the largest, so it represents the maximum margin of separation.

ii. Naive Bayes Classifier (NB)

The Naive Bayes classifier is the simplest and most commonly used classifier. Naive Bayes classification model computes the posterior probability of a class, based on the distribution of the words in the document. The model works with the BOWs feature extraction which ignores the position of the word in the document. It uses Bayes Theorem to predict the probability that a given feature set belongs to a particular label.

$$P(\text{label}|\text{features}) = \frac{P(\text{label}) * P(f1|\text{label}) * \ldots * P(fn|\text{label})}{P(\text{features})}$$

Where,

P(label) is the prior probability of a label or the likelihood that a random feature set the label.

P(features) is the prior probability that a given feature set is occurred.

P(features | label) is the prior probability that a given feature set is being classified as a label.

III. PROPOSED METHOD

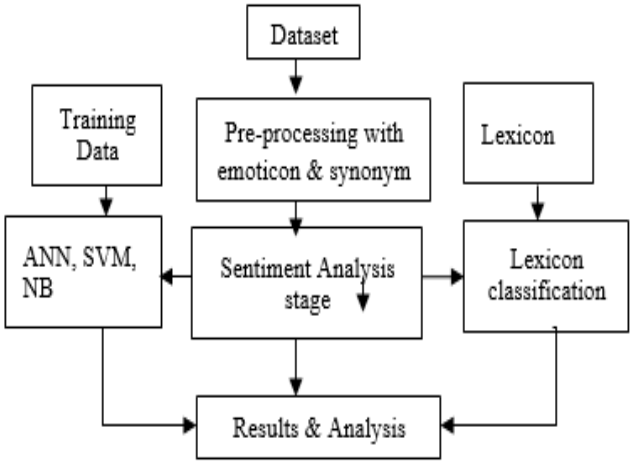

Fig 1: Proposed diagram

Emoticons

In this feature, entered the emoticons in the form text can also predict as pictorial emoticons.

 ISSN: 2321-8169
517 – 521

Example:

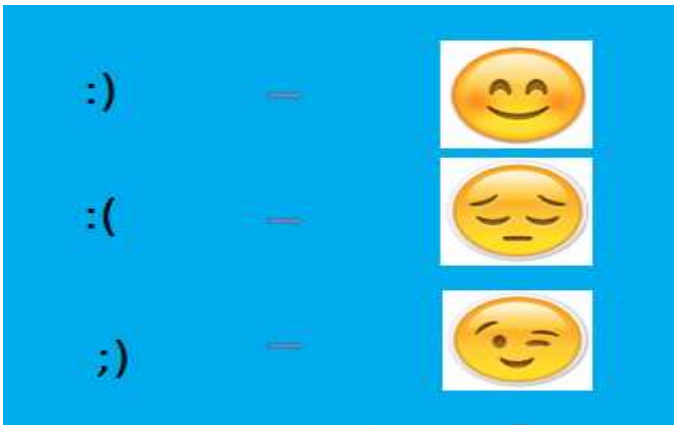

Fig 2: Example of emoticon

Emoticon & Its Meaning [7]

ICON	*MEANING*
:-) =) :) 8) :] =| => 8-) :-> :-] :") :')	smiley
:3 :> :') :-3 => :-> :-V =v :-1	happy face
^_^ ^L^ ^)^	happiness
:* :-*	kiss, couple kissing
;-) ;) ;-] ;] ;-> ;> %-}	wink, smirk
<3	heart
:-D :D =D :-P =3 xD	laughing
:P =P	tongue sticking out, playful
O.o o.O	surprised
:v	gape
B) B-) B| 8|	feel cool
:'-)	tears of happiness
:!	exclamation
:-X	Sealed lips, wearing braces
=* :-* :*	kiss

Fig 3: Emoticon's meaning

Replacing synonym

It takes all synonyms of word and search out with the dictionary and then give the results.
Example:
Good = skilful, descent, nice etc.

Experimental Parameters:

Accuracy (AC) is determined as:

$$AC = \frac{TP + TN}{TP+TN+FP+FN}$$

Where,
TP is true positive, TN is true negative, FP is False Positive, and FN is False Negative.

Recall (R) is determined as:

$$R = \frac{TP}{TP + FN}$$

Recall is also known as True Positive Rate (TPR).

Precision is determined as:

$$P = \frac{TP}{TP + FP}$$

IV. EXPERIMENTAL AND ANALYSIS

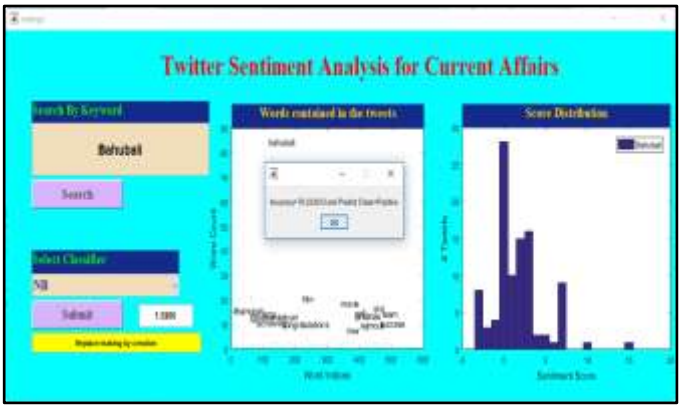

Fig 4: NB Classification

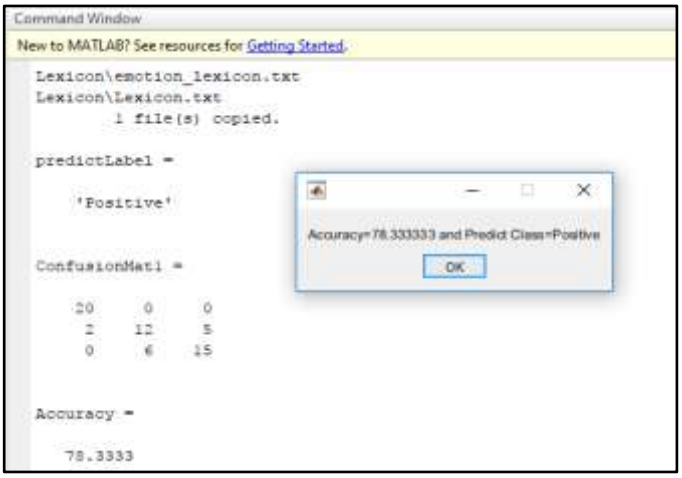

Fig 5: NB Confusion Matrix

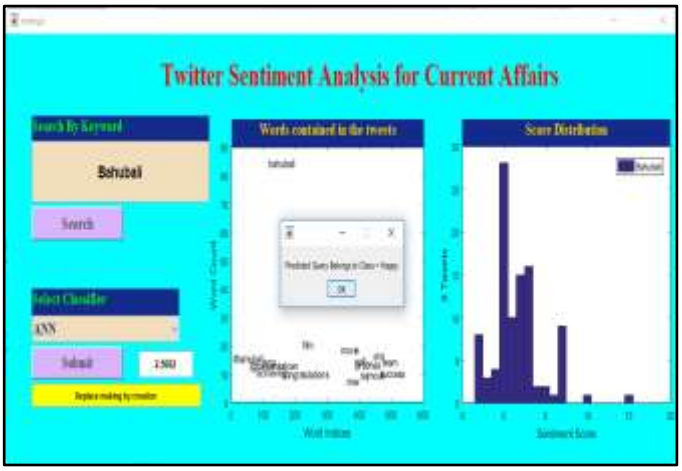

Fig 6: ANN Classification

ISSN: 2321-8169
517 – 521

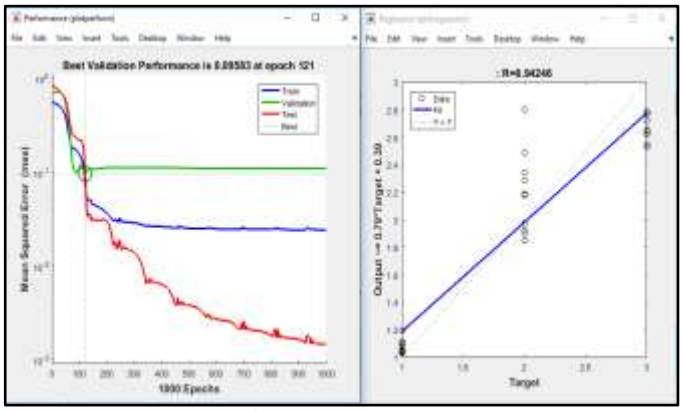

Fig 7: Performance & Regression Graph

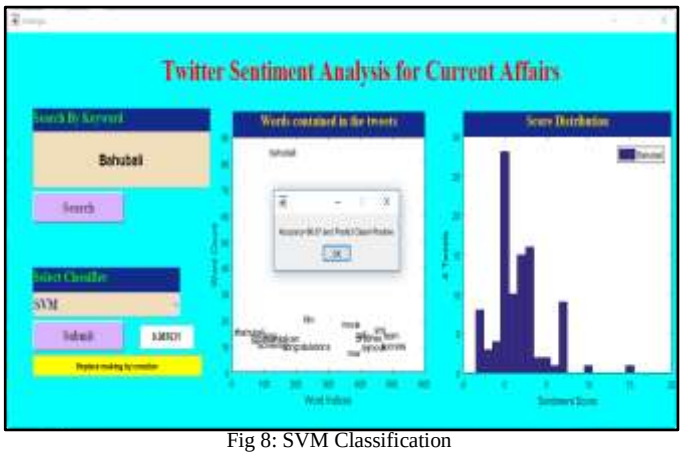

Fig 8: SVM Classification

```
Command Window
New to MATLAB? See resources for Getting Started.
  Predicted Query Belongs to Class = 1
  Positive
  SVM (1-against-1):
  accuracy = 86.67%
  Confusion Matrix in Persentage:
     100     0     0
       0   100     0
       0    40    60


  precision =

      1.0000
      0.7143
      1.0000


  recall =

      1.0000
      1.0000
      0.6000

  Predicted Query Belongs to Class = 1
  Positive
```

Fig 9: SVM Confusion Matrix

Table 1: Parameters

Classifier	Accuracy	Time
Naive Bayes	78.33%	1.92s
Support Vector Machine	86.67%	1.12s
ANN	94.24%	2.56s

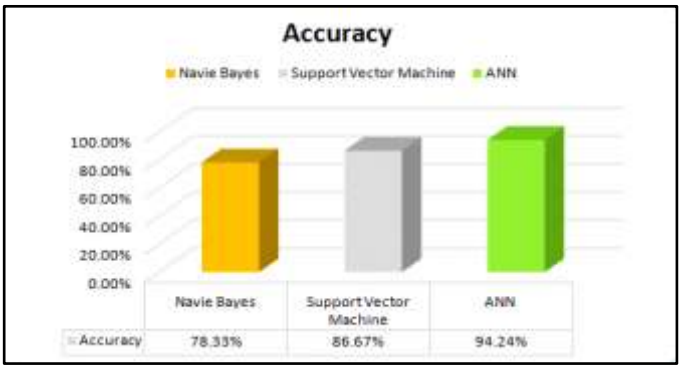

Fig 10: Accuracy Analysis

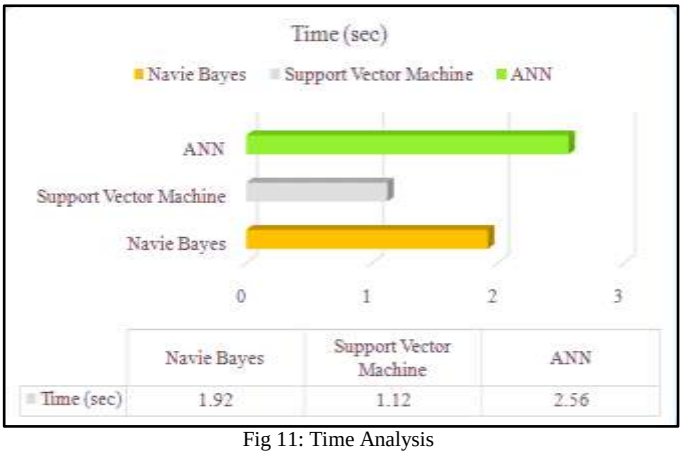

Fig 11: Time Analysis

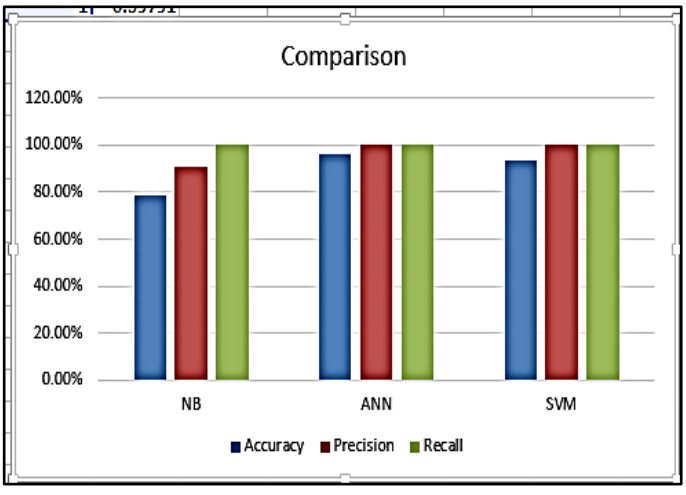

Fig 12: Comparison

CONCLUSION

As sentiments of individuals are extremely useful for people and company owner for making several decisions, introduced proposed Hybrid Polarity Detection System for Sentiment Analysis and summarization that uses new set of features, tries to improve the accuracy compare to state-of-the-art techniques to get the clear idea about the marketing research auditing, public opinion tracking, product reviewing, business research, political review, enhancing of web shopping bases, and so on. As per our experiment, we believe that as the part of Sentiment Analysis, moving towards word level Sentiment Features rather than manual text processing. Form the Experiments it has been observe that naïve bayes classifier have the Overfiitting problem and give 78.33% accuracy. To get better accuracy than NB, ANN is used which give the 86.67% accuracy. Further SVM is used with feature extraction model with emoticon and synonyms and it gives 94.24% accuracy.

ACKNOWLEDGMENT

I believe I have become a lot more adaptable than before. Countless people have helped me in many different ways. Let me try to remember a few. To those I am sure to miss out, my sincerest apologies. I am highly indebted to **Assistant Prof. Priya Pati, PIT** and **Assistant Prof. Abhishek Tripathi, PIT** for his guidance and constant supervision as well as for providing necessary information regarding this dissertation.

Thanks to my Family members my father, mother, grandfather, grandmother and many more for their support to carry out this dissertation. Thanks, to my brother for always supporting during my work.

REFERENCE

[1] Ajay Deswal and SudhirkumarSharma**"Twitter Sentiment Analysis Using Various Classification Algorithms"**, IEEE Conf(ICRITO)2016, DOI: 10.11.9/ICRITO.2016.7784960, Pages 251-257.

[2] David Zimbra, M.Ghiassi, Sean **lee "Brand Related Twitter Sentiment Analysis Using Feature Engineering And Dynamic Architecture for ANN "** IEEE Conf(HICSS)2016, DOI: 10.1109/HICSS.2016.244,Pages 1930-1938.

[3] Brett Duncan, Yanquing Zhang **"Neural Network for Sentiment Analysis on Twitter"** IEEE Conf(ICCI*CC) 2015 , DOI: 10.1109/ICCI*CC.2015.7259397, Pages 275-278.

[4] Fao Wang, jorge A Castanon, **"Sentiment Expression via Emotions on Social media ",** IEEE Conf(BD) 2015 , DOI:10.1109/Big Data.2015.7364034, pages 2404-2408.

[5] GeetikaGautam, DivakarYadav, **"Sentiment Analysis of Twitter Data Using Machine Learning Approaches and semantic Analysis "**, IEEE Conf(IC3)2014, DOI: 10.1109/IC3.2014.6879213, Pages 437-442.

[6] Diego Terrana, AgneseAugello, **"Automatic Unsupervised polarity Detection on a Twitter data stream",** IEEE Conf(ICSC) 2014, DOI:10.1109/ICSC.2014.17, Pages 128-134.

[7] Xujuan Zhou, Xiaohui Tao, Jianming Yong and ZhenyuYang**"Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis"**, IEEE 2014