www.ijecs.in
*International Journal Of Engineering And Computer Science ISSN: 2319-7242*
*Volume 6 Issue 2 Feb. 2017, Page No. 20273-20280*
*Index Copernicus Value (2015): 58.10, DOI: 10.18535/ijecs/v6i2.21*

# Classifying sentiments of twitter data using Bay's neural network

***Megha Saraswat[1], Mr. Rahul Patel[2]***
[1]Acropolis Institute Of Technology and Research
Indore Madhya Pradesh
meghasaraswat92@gmail.com
[2]Acropolis Institute Of Technology and *Research*
Indore Madhya Pradesh
rahulpatel@acropolis.in

***Abstract*— *Text mining is a classical domain of research and development, in this technique the text data is used for preparing the data models for classifying similar patterns of text documents. In order to perform this classification and categorization algorithms are implemented to perform the required task. But the text patterns in different social sites and blogs are not only categorized in their subjective similarities. These texts can also be classified on the basis of author's sentiments or the emotional aspects. Therefore the identification of emotional components and features from the social site text and classify the posts on the basis of their sentiments is respectively new domain of research and development. In this presented work the micro-blog text is classified according to the text sentiments and emotional features. Therefore the twitter micro text is used for training and testing of supervised data models. In this context a supervised hybrid classification technique is developed using the Bay's classifier and the Back propagation neural network. The key role of the Bay's classifier is to find the emotional components in terms of positive wordlist and negative word list. Additionally using these components the message is encoded in numerical strings. These numerical strings are further used with the BPN algorithm for performing the training and testing operations. In addition of that to improve the classification performance of text the abbreviations and similes are also recovered as the emotional components. The implementation of the proposed technique is performed using JAVA technology and for classifiers WEKA library is used. After the implementation the performance of the proposed technique is evaluated and compared with the similar classification technique. In comparative performance study the proposed model found efficient and accurate as compared to traditionally used technique*.

***Keywords*— Text mining, classification, sentiments, neural network, supervised learning**

## I. INTRODUCTION

Data mining is a classical domain of data analysis based on computational algorithms. These algorithms applied on raw set of data to find refined information from the data. Therefore according to the nature of data the techniques are also varied. In this presented work the social media based text analysis is primary aim of the study. By nature the text data is found in unstructured manner additionally there are no class labels are associated with the data. Therefore classification of text based on the sentiments becomes more complicated. In data mining, for unstructured data analysis the clustering approaches or unsupervised approaches are used. But the performances of the unsupervised approaches are not much effective as compared to the supervised learning algorithms. Thus in this work the supervised algorithm is used for computing the classes of the data.

In this context, the supervised learning techniques are need to be train on the specific kind of patterns and after training the learning algorithm able to classify the data or recognize the pattern. Here for training purposed the twitter dataset is used, the twitter data set contains the user ID, user twit and their positive or negative orientation. By which the classifier first perform training on the input data and then the trained data model is used to test the new and similar kinds of patterns. Additionally to enhance the performance of the traditional classification technique a hybrid classification algorithm is proposed for implementation. The key advantage to use the hybrid classifier is that, such techniques combine the goodness of more than one classification algorithms.

## II. PROPOSED WORK

This section provides the detailed discussion about the proposed working model for text classification based the text sentiments analysis. Thus the entire functional and conceptual concepts are reported in this chapter using the methodology and the proposed algorithm design steps.

### A. System overview

The social media becomes virtual playground for the new generation young persons. A rich number of youngsters are participating on the social media for communicating with their friends and posting their images and others. Therefore in their virtual social life they also communicate with their emotions and express their thinking using the social media text. In order to analyses the emotional expressions hidden on the posted text need to analyses the text using the NLP based techniques. In order to do this this presented work is aimed to classify the social media text according to their positive or negative orientation of user emotions.

Basically the text analysis and their classification is a task of text mining. In this technique the text is analysed and based on the similarity measures in text patterns the data is classified in multiple classes. In the similar ways in the sentiment based text analysis or the text orientation based text analysis the semantic similarity based classification or categorization is modelled. To perform such classification a hybrid model is proposed with the help of Bayesian classifier and back propagation neural network. Here the Bayesian classifier is implemented for finding the positive and negative features estimation. Additionally the neural network is used to learn on the computed features and after learning classify them using the

previous knowledge. The proposed technique contributes for the following main areas:

1. Enhances the text quality by recovering the complete text from the abbreviations of social media text
2. Enhance the text quality by recovering the hidden emotions using the placed similes

### B. Proposed methodology

The text mining for finding the sentiments and text orientation can be performed in two major phases. In first using the traditional user patterns a data model is prepared and in further using the prepared data model the classification of text according to their sentiments. That is the method of supervised learning, therefore to design the proposed technique there are two different modules are implemented.

**Training module**

The training model of the proposed concept is demonstrated using figure 2.1 this diagram contains the required components and subcomponents to report the functional aspects of the proposed system.

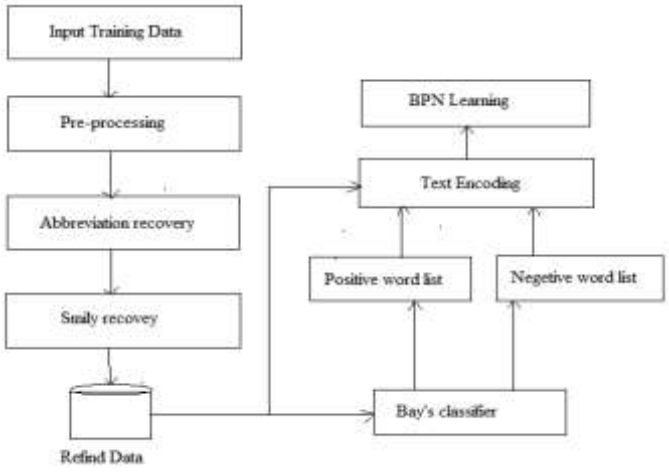


Figure 2.1 training module

**Input training data:** in order to find the text classification method according to the text sentiments the supervised learning technique is implemented with the system, thus for training the social media text and communication is required. Therefore in order to train the proposed data model the twitter dataset is used. The twitter dataset is a micro-blog data set which contains the user id, twits and the associated class labels. This dataset is sub-divided in a different number of instances for preparing the training and testing datasets.

**Pre-processing:** the pre-processing is a part of machine learning or data mining technique that enhances the quality of data to use with the classifier for effective learning. The pre-processing can include different methods and techniques to refine the noisy contents from the text data and refine for utilizing with the computational algorithm. In this purpose the text punctuations, special characters and unwanted symbols are reduced from the training test.

**Abbreviation recovery:** after pre-processing of the input training dataset the emotional features are tries to recover. Therefore in this phase the main concentration is placed for recovering the short forms of the text to their actual ful-form of the abbreviation. This text increases the amount of data in terms of size but the amount of features of text words are remain same.

**Smiley recovery:** after processing of data for the abbreviation, the text is again scanned for recovering the sentimental expressions by analyzing the input Smiley. Sometimes the social media user expresses their emotions only with the Smiley therefore the positive impact and negative impact of smiley is also need to obtain in the input text.

**Refined data:** finally the data becomes refined and complete as required for the learning and the classification. Thus this refined data is stored separately for further use with the learning of the pattern.

**Bay's classifier:** after pre-processing of the data now the refined text is used for learning with the proposed data model. Thus in first step the Bayesian classifier is implemented. The Bayesian classifier is worked on the basis of probability distributions of the similar word occurrence with both the kinds of text positive or negative. The bay's classifiers basic concept is demonstrated as:

The Naive Bayes classification algorithmic rule is a probabilistic classifier. It is based on probability models that incorporate robust independence assumptions. The independence assumptions usually don't have an effect on reality. So they're thought of as naive. You can derive probability models by using Bayes' theorem (proposed by Thomas Bayes). Based on the nature of the probability model, you'll train the Naive Bayes algorithm program in a much supervised learning setting. In straightforward terms, a naive Bayes classifier assumes that the value of a specific feature is unrelated to the presence or absence of the other feature, given the category variable. There are two types of probability as follows [29]:

- Posterior Probability [P (H/X)]
- Prior Probability [P (H)]

Where, X is data tuple and H is some hypothesis. According to Baye's Theorem

$$P\left(\frac{H}{X}\right) = \frac{P\left(\frac{X}{H}\right)P(H)}{P(X)}$$

Using the given hypothesis the individual text is processed and their probability distribution is computed between both the orientations. Therefore each individual words the word probability for positive and negative orientation is computed. Thus this phase is responsible to generate two different wordlist as:

- **Positive word list:** The positive wordlist contains the list of words and their probability for finding with the positive kinds of text.
- **Negative word list:** Similarly the negative word list maintains the list of words and probability which can most possibly occurred with the negative sentences.

In this context it is noticed that some of the words can only found in positive text and some of the words can only be found in negative sentences. In addition of that, it is also possible for some words can be found in both the kinds of wordlist. A

simple example of positive and negative word list is reported using table 2.1.

| Words | Positive word list | Negative word list |
|---|---|---|
| Good | 0.75 | 0.25 |
| WOW | 1 | 0 |
| Painful | 0 | 1 |

Table 2.1 positive and negative word list

**Message encoding:** after preparing the list of positive and negative words the text instances in the dataset is encoded based on their class labels. For example a simple sentence for their positive and negative orientations is encoded in the following manner.

"Ram is a good boy" and "Ram is not a good boy"

In both the text sentences the almost the participating words are similar and only a single word is different from each other and by which the entire sentence is identified on the basis of orientation. Therefore the encoding of both the sentences is similar but a single word makes the sentence negative or positive.

**BPN training:** after encoding of the messages the BPN (back propagation neural network) is used for providing the training. The BPN algorithm training process is demonstrated as:

The implementation of neural network is defined in two phases' first training and second prediction: training method utilizes data and designs the data model. By this data model next phase prediction of values is performed [30].

**Training:**

1. Prepare two arrays, one is input and hidden unit and the second is output unit.

2. Here first is a two dimensional array $W_{ij}$ is used and output is a one dimensional array $Y_i$.

3. Original weights are random values put inside the arrays after that the output is given as**.**

$$x_j = \sum_{i=0} y_i W_{ij}$$

Where, $y_i$ is the activity level of the $j^{th}$ unit in the previous layer and $W_{ij}$ is the weightof the connection between the $i^{th}$ and the $j^{th}$ unit.

4. Next, action level of $y_i$ is estimated by sigmoidal function of the total weighted input.

$$y_i = \left[\frac{e^x - e^{-x}}{e^x + e^{-x}}\right]$$

When event of the all output units have been determined, the network calculates the error (E) given in equation**.**

$$E = \frac{1}{2}\sum_i (y_i - d_i)^2$$

Where, $y_i$ is the event level of the $j^{th}$ unit in the top layer and $d_i$ is the preferred output of the $j_i$ unit.

**Testing module**

After training of the proposed model for text classification the classification labels are need to recover for the input test dataset. Therefore a new module is implemented for recognizing the class labels of the unlabelled text. The proposed technique for testing operation is reported using figure 2.2 and their components are given as follows.

**Test dataset:** as discussed in the training module the twitter dataset is divided in different size of instances for training and testing purpose. Therefore the similar size of data is prepared for testing of the proposed data model. This phase is responsible for accepting and reading of the input test dataset.

**Pre-processing:** as similar in the training data pre-processing the similar functions are used for performing the pre-processing of the data. Therefore first the unwanted characters and punctuations are removed from the test dataset. In further as the abbreviations and smileys are recovered in training session in similar manner the text is recovered.

**Message encoding:** additionally in this phase the message is encoded similarly as performed in the training module. This encoding is taken place with the help of previously defined:

- Positive word list
- Negative wordlist

**Trained neural network:** finally the encoded message is passed to the trained neural network, trained neural network is accept the encoded text data and produces the following two outcomes:

- **Class labels:** for individual instances of the data a class label is returned by the classifier to define the orientation of the input test set
- **Performance:** additionally the performance of the classifier is also computed based on their accuracy and other parameters

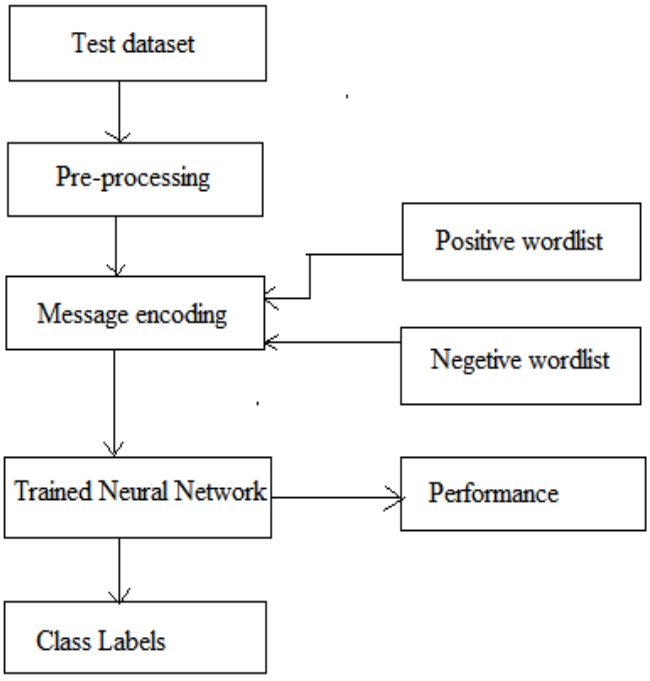


Figure 2.2 Testing module

### C. Proposed algorithm

The previous phase provides the understanding about the proposed data model, this section summarize the proposed model using the simple algorithm steps.

Input: training dataset $D_{train}$, testing dataset $D_{test}$

Output: class labels C

Process:

1. $R = readDataset(D_{train})$
2. $P = preprocessData(R)$
3. $R_c = recoverAbb(R)$
4. $R_f = recoverSmily(R_c)$
5. $[Plist\ Nlist] = Bays.Train(R_f)$
6. $for(i = 1; i \leq D_{train}.length; i++)$
7. $E_i = encodeMessage(D_{train}(i))$
8. $end\ for$
9. $T_{model} = BPN.train(E)$
10. $[C\ Pr] = T_{model}.classify(D_{test})$
11. $return\ [C\ Pr]$

Table 2.2 proposed algorithm

## III. Result Analysis

After successfully implementation of the proposed system and traditional SVM based technique the performance of the system is computed and compared using line graphs. The comparative performance of the system is demonstrated in terms of different performance parameters. These parameters and their observations are reported in this section.

### A. Accuracy

Accuracy of the classification techniques shows the amount of data which is accurately identified during the class label prediction. Thus the amount of accurately predicted class labels is termed as the accuracy of the classification. The accuracy of any classifier in terms of accuracy is computed using the following formula.

$$accuracy = \frac{total\ correctly\ classified\ data}{total\ data\ input\ for\ classification} X100$$

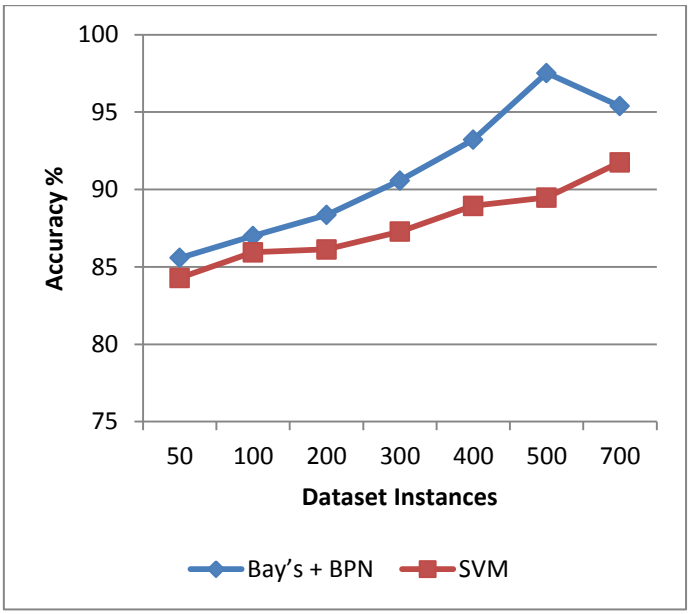


Figure 3.1 accuracy

| Dataset size | Bay's + BPN | SVM |
|---|---|---|
| 50 | 85.58 | 84.29 |
| 100 | 87 | 85.94 |
| 200 | 88.35 | 86.13 |
| 300 | 90.57 | 87.28 |
| 400 | 93.21 | 88.94 |
| 500 | 97.52 | 89.48 |
| 700 | 95.38 | 91.75 |

Table 3.1 accuracy

The comparative performance of traditional and proposed text classification technique according to the text orientation is given using table 3.1 and table 3.1. In this diagram the X axis shows the number of instances produced for training and testing and the Y axis shows the amount of accurately classified instances. Similarly the performance of the proposed hybrid classifier (Bay's and BPN) is demonstrated using blue line and the red line is used to demonstrate the traditional technique's (SVM) performance. According to the obtained performance the proposed technique enhances the classification rate of the proposed technique as compared to the traditional technique, thus the proposed method is more effective and accurate as compared to the traditional classification technique.

### B. Error rate

In data mining and machine learning the error rate shows the amount of data which is incorrectly recognized by the classifier is termed as the error rate of the system. The error rate of the classifier in terms of percentage can be computed using the following formula.

$$error\ rate = 100 - accuracy$$

Or

$$error\ rate = \frac{incorrectly\ classified\ data}{total\ input\ data\ for\ classification} X100$$

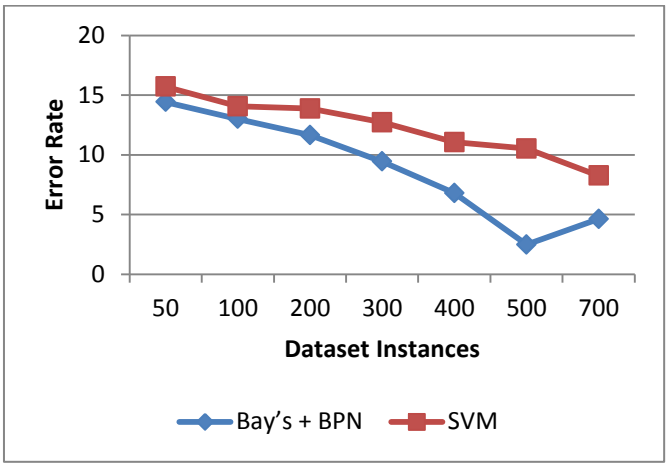


Figure 3.2 error rate

| Dataset size | Bay's + BPN | SVM |
|---|---|---|
| 50 | 14.42 | 15.71 |
| 100 | 13 | 14.06 |
| 200 | 11.65 | 13.87 |
| 300 | 9.43 | 12.72 |
| 400 | 6.79 | 11.06 |
| 500 | 2.48 | 10.52 |
| 700 | 4.62 | 8.25 |

Table 3.2 error rate

The error rate of the proposed hybrid classifier and traditional SVM classifier for finding the text sentiments are provided using figure 3.2 and table 3.2. In the given figure the red line shows the performance of the traditional SVM classifier and blue line is used to denote the performance of the proposed BPN and bay's based classification technique. Additionally for representing the performance of the system using line graphs the X axis shows the number of dataset instances provided for classification and Y axis shows the amount of incorrectly identified patterns among them in terms of percentage. According to the performance evaluation the error rate of both the classifier is enhances with the increasing amount of datasets but the proposed technique produces more accurate data as compared to the traditional approach. Therefore the proposed technique is more adoptable than the traditional classification technique.

**C. Memory usages**

Every computational algorithm consumes a considerable amount of main memory during their computation. This consumed memory unit is termed as the memory usages or space complexity of the algorithm. The space complexity of the processes in JAVA implementation is computed using the following formula.

$$memory\ usage = total\ memory - free\ memory$$

The comparative memory consumption of both the techniques proposed and traditional is reported using the figure 3.3 and table 3.3. To represent the performance of the proposed technique the blue line is used and for demonstration of traditional technique the red line is used. For performance demonstration the X axis of the line graph contains the amount of data produced for training and testing and the Y axis shows the amount of memory required for execution of the algorithm. According to the demonstrated results both the algorithms consumes similar patterns of memory therefore the memory consumption is depends on the amount of data for processing, additionally in the keen observations the proposed technique is resource preserving technique as compared to the traditional approach. Thus the proposed technique is more adoptable than the traditional technique.

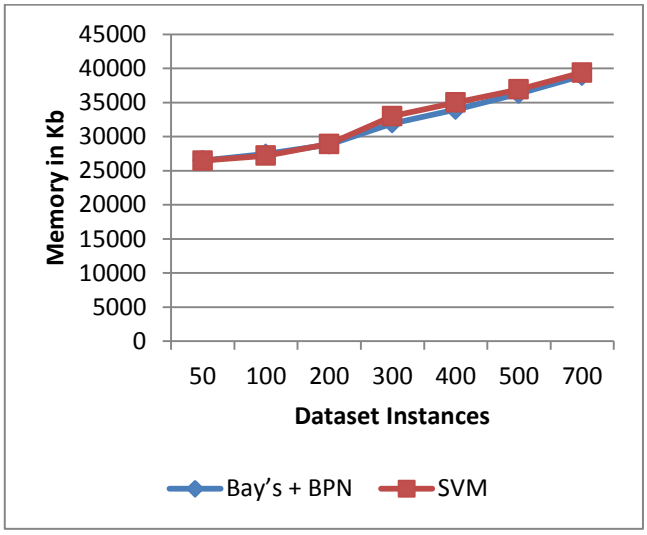


Figure 3.3 memory usages

| Dataset size | Bay's + BPN | SVM |
|---|---|---|
| 50 | 26493 | 26481 |
| 100 | 27463 | 27194 |
| 200 | 28844 | 28918 |
| 300 | 31947 | 32991 |
| 400 | 33928 | 34982 |
| 500 | 36281 | 36918 |
| 700 | 38918 | 39381 |

Table 3.3 memory consumption

### D. Time consumption

The supervised algorithms work on two different phases namely training and testing. During the training predefined data patterns are used for providing the training during this an amount of time is required for learning of algorithm to use with the classification or pattern recognition. This time consumption is also termed as the training time or time complexity of algorithm. The amount of required time is given using the following formula.

$$training\ time = training\ end\ time - training\ start\ time$$

The training time of both the implemented classification techniques are given using table 3.4 and table 3.4. The X axis of the diagram shows the number of instances produced for performing the classifier's training and the Y axis shows the amount of time consumed for training. Additionally the blue line denotes the amount of time consumed by the proposed algorithm and the red line shows the time consumed by the traditional WEKA based SVM classifier. According to the observed results the proposed technique consumes higher time as compared to the traditional classifier. Thus the proposed technique is profitable only when the accuracy is key area of concern not the time. Additionally the traditional technique is more adoptable for the time complexity.

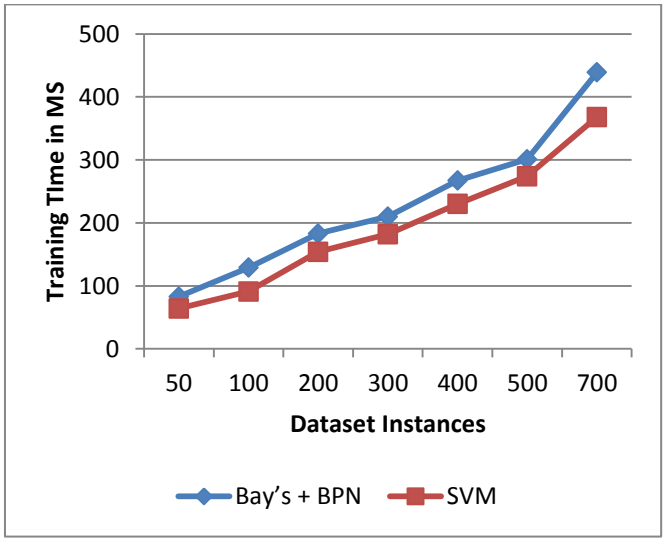


Figure 3.4 training time

| Dataset size | Bay's + BPN | SVM |
|---|---|---|
| 50 | 83 | 64 |
| 100 | 129 | 91 |
| 200 | 183 | 154 |
| 300 | 210 | 182 |
| 400 | 267 | 230 |
| 500 | 301 | 274 |
| 700 | 439 | 368 |

Table 3.4 time consumption

## IV. Conclusion

After implementation and analysis of the obtained performance of the proposed technique this chapter provides the conclusion of the conduced study work for classification of text according to the social media based user's text sentiments. Additionally the future extension of the proposed work is also suggested.

### A. Conclusion

NLP (natural language processing) for finding the author's sentiments and emotions and their expressions are a new domain of study as compared to the traditional text classification. In this presented work the key area of interest is to investigate the contribution of the data mining techniques over analysing the text sentiments or orientation text expressiveness.

The main aim of the proposed technique is to enhance the classification accuracy of the traditional classification techniques. Thus among a number of available classifiers the Bayesian classifier and the back propagation neural network is used for implementing new technique. The hybrid technique contains the goodness of both the classification techniques thus promises to provide more accurate classification outcomes as compared to individual classifiers.

After text quality enhancement the bay's classifier is implemented for finding a word with the positive text and also with negative expression of text. Using this prepared probability distribution of individual words the back propagation neural network is trained and further utilized for the classification of text according to the text orientation discovery.

The implementation of the proposed technique is performed using the JAVA technology. Additionally for implementation of the traditional technique is provided using the WEKA based classifier. After implementation the experimental performance of both the techniques are computed and compared with the different performance factors. The computed parameters and the observations are reported using the table 4.1 as the performance summary.

| S. No. | Parameters | Proposed hybrid technique | Traditional technique |
|---|---|---|---|
| 1 | Accuracy | High | Low |
| 2 | Error rate | Low | High |
| 3 | Time consumption | High | Low |
| 4 | Memory consumption | Low | High |

Table 4.1 performance summary

### B. Future work

The main aim of preparing the binary classifier for finding the text orientation in terms of positive and negative is prepared successfully. Additionally the enhanced performance over the traditional technique is also obtained. In near future the

following areas of applications are suggested for future extension.

1. The method currently provides the binary classes thus need to improve the technique for the multiple class identification
2. Extend the technique for implementing with stock market forecasting and NEWS analysis
3. Extend the technique for monitoring the social media based communication among terrorist and other harmful activities

## REFERENCES

[1] Malhar Anjaria, Ram Mahana Reddy Guddeti, "Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning", 978-1-4799-3635-9/14/$31.00 ©2014 IEEE

[2] Kainaz Bomi Sheriwala, "Data Mining Techniques in Stock Market", INDIAN JOURNAL OF APPLIED RESEARCH, Volume 4, August 2014.

[3] S. L. Pandhripande and Aasheesh Dixit, "Prediction of 2 Scrip Listed in NSE using Artificial Neural Network", International Journal of Computer Applications (IJCA), Volume 134, No.2, January 2016.

[4] Zabir Haider Khan, Tasnim Sharmin Alin, Md. Akter Hussain, "Price Prediction of Share Market using Artificial Neural Network (ANN)", International Journal of Computer Applications (0975 – 8887) Volume 22– No.2, May 2011

[5] Kainaz Bomi Sheriwala, "Data Mining Techniques in Stock Market", INDIAN JOURNAL OF APPLIED RESEARCH, Volume 4, August 2014.

[6] Umajancy. S, Dr. Antony Selvadoss Thanamani, "An Analysis on Text Mining –Text Retrieval and Text Extraction", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 8, August 2013

[7] Soujanya Poria, Erik Cambria, Newton Howard, Guang-Bin Huang, Amir Hussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content", [8] Quanzeng You and Jiebo Luo, Hailin Jin, Jianchao Yang, "Cross-modality Consistent Regression for Joint Visual-Textual Sentiment Analysis of Social Multimedia", and/or a fee. Request permissions from permissions@acm.org WSDM'16, February 22–25, 2016, San Francisco, CA, USA c 2016 ACM

[9] Basant Agarwal, Namita Mittal, Pooja Bansal, and Sonal Garg, "Sentiment Analysis Using Common-Sense and Context Information", Hindawi Publishing Corporation Computational Intelligence and Neuroscience Volume 2015, Article ID 715730, 9 pages

[10] Sérgio Canuto, Marcos André Gonçalves, Fabrício Benevenuto, "Exploiting New Sentiment-Based Meta-level Features for Effective Sentiment Analysis", WSDM'16, February 22–25, 2016, San Francisco, CA, USA c 2016 ACM

[11] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li, "Dual Sentiment Analysis: Considering Two Sides of One Review", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 8, AUGUST 2015

[12] Aliaksei Severyn, Alessandro Moschitti, "Twitter Sentiment Analysis with Deep Convolutional Neural Networks", SIGIR'15, August 09 - 13, 2015, Santiago, Chile c 2015 ACM

[13] Rui Gaspar, Claudia Pedro, Panos Panagiotopoulos, Beate Seibt, "Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events", Computers in Human Behavior 56 (2016) 179-191

[14] Alexander Hogenboom, Flavius Frasincar, Franciska de Jong, Uzay Kaymak, "Using Rhetorical Structure in Sentiment Analysis"

[15] Pete Burnap, Omer F. Rana, Nick Avis, Matthew Williams, William Housley, Adam Edwards, Jeffrey Morgan, Luke Sloan, "Detecting tension in online communities with computational Twitter analysis", Technological Forecasting & Social Change xxx (2013) xxx–xxx

[16] Quan Zou, Jinjin Li, Qingqi Hong, Ziyu Lin, Yun Wu, Hua Shi, and Ying Ju, "Prediction of MicroRNA-Disease Associations Based on Social Network Analysis Methods", Hindawi Publishing Corporation BioMed Research International Volume 2015, Article ID 810514, 9 pages

[17] S. L. Pandhripande and Aasheesh Dixit, "Prediction of 2 Scrip Listed in NSE using Artificial Neural Network", International Journal of Computer Applications (IJCA), Volume 134, No.2, January 2016.

[18] Dr. B. Srinivasan and K. Pavya, "A STUDY ON DATA MINING PREDICTION TECHNIQUES IN HEALTHCARE SECTOR", International Research Journal of Engineering and Technology (IRJET), PP. 552-556, Volume 3, Mar-2016

[19] Vipin Kumar, Joydeep Ghosh and • David J. Hand, "Top 10 algorithms in data mining", Knowledge and Information System, PP. 1–37, (2008).

[20] Vapnik V (1995), the nature of statistical learning theory. Springer, New York.

[21] Kavitha G, Udhayakumar A and Nagarajan D, "Stock Market Trend Analysis Using Hidden Markov Models", available online: https://arxiv.org/ftp/arxiv/papers/1311/1311.4771.pdf.

[22] Haiqin Yang, Laiwan Chan, and Irwin King, "Support Vector Machine Regression for Volatile Stock Market Prediction", Intelligent Data Engineering and Automated Learning IDEAL, PP. 391- 396, Springer-Verlag Berlin Heidelberg 2002

[23] Data Mining - Classification & Prediction, available online: http://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm, [accessed 27 April 2016].

[24] Samuel Odei Danso, "An Exploration of Classification Prediction Techniques in Data Mining: The insurance domain", Master's Degree in Advanced Software Engineering, School of Design, Engineering, and Computing. Bournemouth University, September, 2006.

[25] Miloš Radovanović, Mirjana Ivanović, "Text Mining: Approaches And Applications", Abstract Methods and Applications in Computer Science (no. 144017A), Novi Sad, Serbia, Vol. 38, No. 3, 2008, 227-234

[26] Vishal Gupta, Gurpreet S. Lehal, "A Survey of Text Mining Techniques and Applications", Journal of Emerging Technologies in Web Intelligence, Vol. 1, No. 1, August 2009

[27] P. Bhargavi, B. Jyothi, S. Jyothi, K. Sekar, "Knowledge Extraction Using Rule Based Decision Tree Approach", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.7, July 2008

[28] Umajancy. S, Dr. Antony Selvadoss Thanamani, "An Analysis on Text Mining –Text Retrieval and Text Extraction",

International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 8, August 2013
[29] Roshani Choudhary, Jagdish Raikwal, "An Ensemble Approach to Enhance Performance of Webpage Classification", International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5614-5619
[30] Pratik Gite, Sanjay Thakur, "An Effective Intrusion Detection System for Routing Attacks in MANET using Machine Learning Technique", International Journal of Computer Applications (0975 – 8887), Volume 113 – No. 9, March 2015