iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** I **Month of publication:** January 2018

DOI: http://doi.org/10.22214/ijraset.2018.1162

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 6 Issue I, January 2018- Available at www.ijraset.com

Sentiment Analysis of Social Media Text Data using Back Propagation in Artificial Neural Networks

Neha Sharma[1], Sarang Mandloi[2]
[1, 2] *M. Tech Scholar, Department of CSE, AITR Indore, Assistant Professor, Department of CSE, AITR Indore*

Abstract*: Artificial Neural Networks have gained tremendous importance off late due to its large data handling and analyzing capability. One of the major areas of research that has been impacted by it is the field of social media and related applications. This paper proposes a mechanism through which text data from social media applications can be classified. Here data from twitter has been used to form a database. The data mined is then pre-processed and subsequently a Neural Network is trained to classify data into three categories. The categories of data used to train the neural network are tweets corresponding to sad, neutral and happy moods of users tweeting social media data. It has been found that the proposed system achieves around 97% accuracy mainly due to the efficacy of back propagation mechanism of the designed artificial neural network model.
Keywords*: Artificial Neural Network (ANN), Text Mining, Back Propagation, Jacobian Matrix, Levenberg Marquardt Algorithm, Mean Square Error (MSE), Training States, Mean Square Error.

I. INTRODUCTION

Text data mining has gained a lot of importance off late due to its diverse applications related to social media, data analytics, business and political prediction mechanisms etc. Sentiment analysis is the computational study of text opinions, emotions, and attitude towards an entity. Sentiment analysis helps in the decision making also. With the help of sentiment analysis any company can analysis of product review in positive and negative. Another category can be included which can be thought of as a neutral category. The applications of such a technique can be diverse fields where human sentiment could substantially affect significant outcomes. One of the techniques that is used these days s the use of artificial intelligence to accomplish such tasks. Data mining is AN application of knowledge process within which skilled patterns and knowledge is extracted. This extracted info is consumed victimization applications and actual time programs for creating decisions. Tremendous quantity of consumer's entry web, in between they're in the slightest degree times still connected via their buddies creating use of those offerings. Typically tiny human grasping nature over web invitations results in hidden risks. Analysing data with all its parts (e.g. temporal and geographical) can cause data which can produce easier following and keeping check of your voters in elections or opinion polls etc. Moreover, the gathered data can cause you to plenty of palm in influencing them and knowing what data, influencing youth, but gaining data on totally different parts of population conjointly.

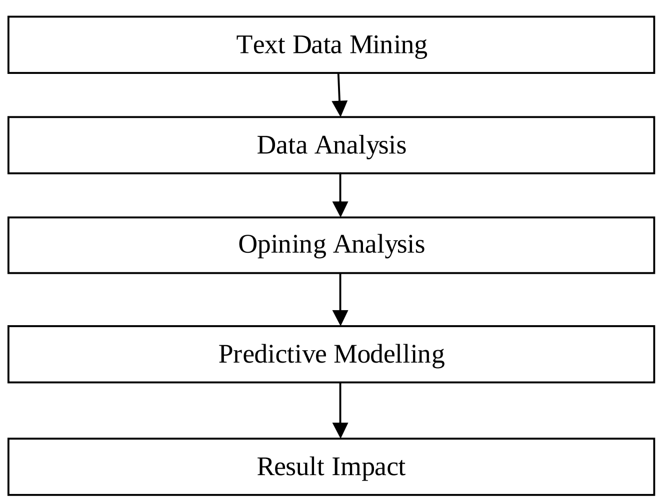

Fig.1 Basic Text Mining and Predictive Modelling Approach

Should the knowledge lack context, analysis becomes a tough draw back. Statements could become disrespectful, or lose their wit. Easy sentences might need their which implies inverted. It is, therefore, necessary to know the context of data creation, publication and consumption. Hence techniques are sought after which would accurately classify data classes and yield accurate results.

II. ARTIFICIAL NEURAL NETWORKS

An artificial neural network tries to emulate the human brain which is a natural neural network. Neural networks are gaining immense importance due to the fact that they can be used to implement artificial intelligence. The artificial neural network paradigm is similar to the human brain in the following ways:

A. It has a highly non-linear structure.
B. It can accept data in a parallel manner.
C. It can learn and adapt synonymous with the human brain's nature.

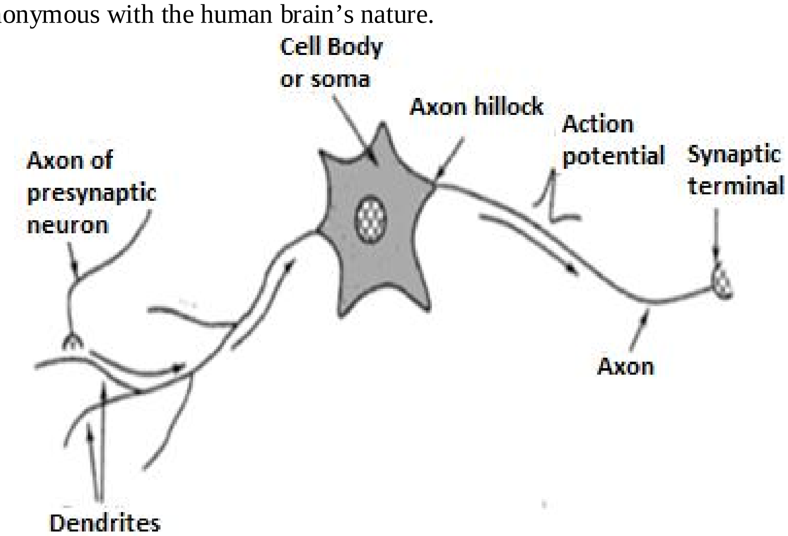

Fig.2 The biological model of neural network

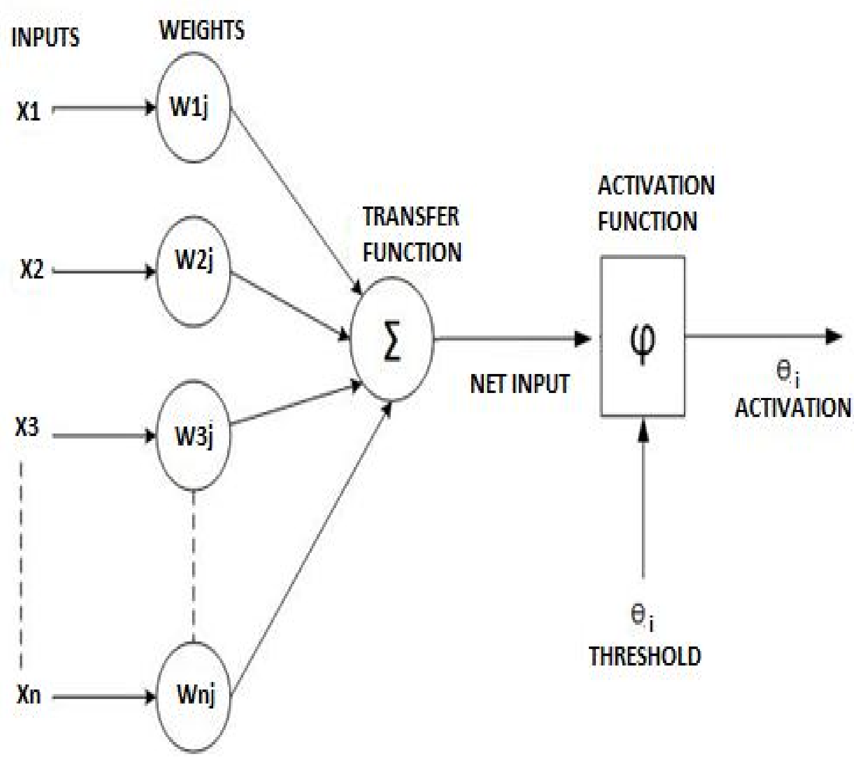

Fig.3 Mathematical model of a neural network.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 6 Issue I, January 2018- Available at www.ijraset.com

III.BACK PROPAGATION IN NEURAL NETWORKS

While various types of neural networks can be designed, yet one of the most effective forms of neural networks is the mechanism of back propagation in which the errors of prediction are used to update the weights of the existing neural networks with an aim of reducing the errors in subsequent predictions. The block diagram of a back propagation mechanism in neural networks is shown in the following diagram.

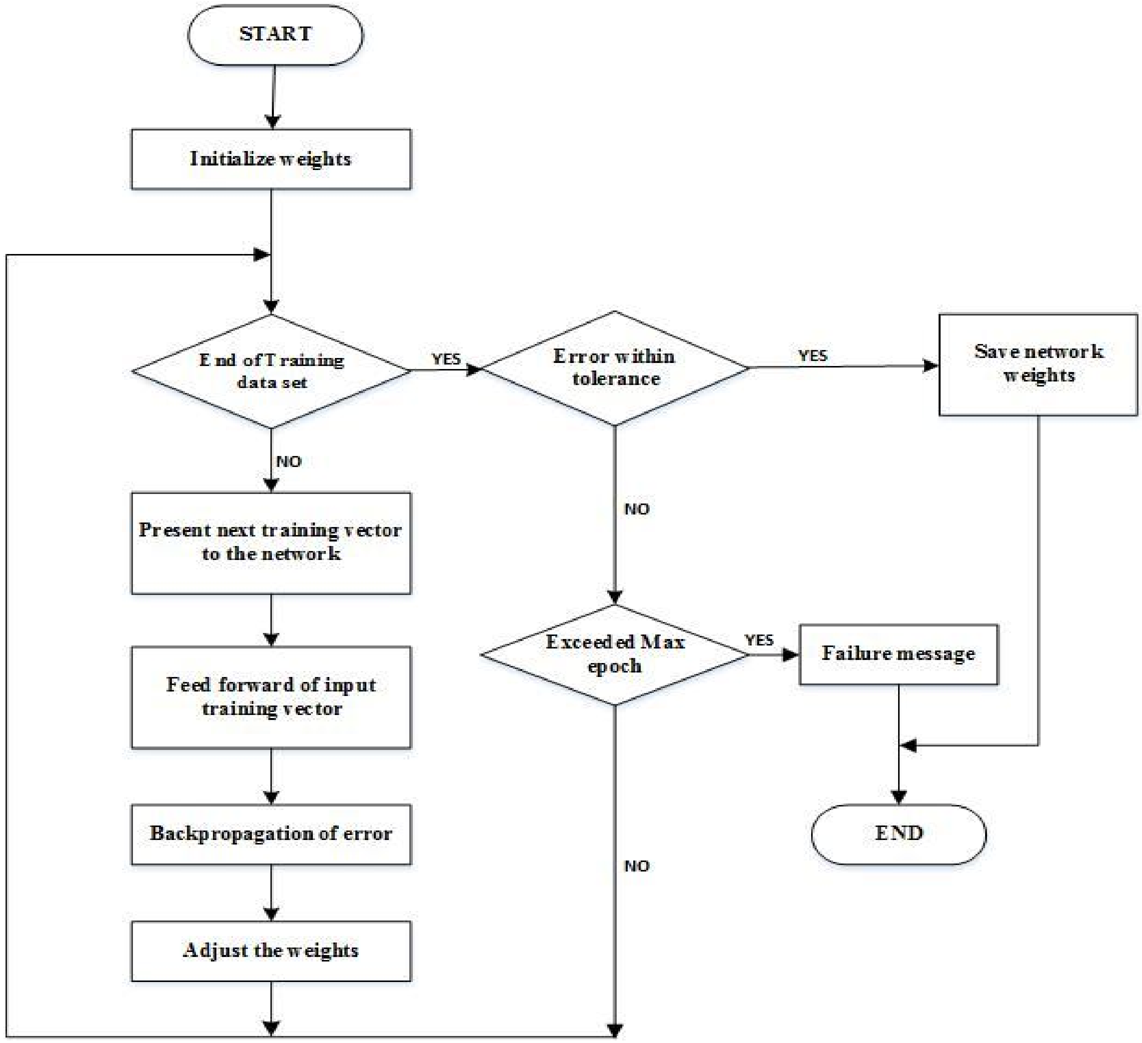

Fig.4 Back Propagation in Artificial Neural Networks

It can be seen from the above figure that the errors of iterations are fed back to the neural network which updates the weights of the neural network. The process continues till the errors attain tolerable limits or the epoch decided for the iterations get exhausted. The manners in which the errors are updated make a significant difference in the updating of weights. The Levenberg Marquardt (LM) algorithm is used in the present approach to employ back propagation in the neural network design. The LM algorithm can be considered to be a hybrid mixture of the steepest dissent and the Gauss-Newton method. It is quick and is stable thereby rendering an efficient algorithm. The mathematical modeling of the LM algorithm can be understood as follows:

Let the errors in iteration 'i' be designated by e_i

The weight of iteration 'i' be designated by w_i

Then the weight of the subsequent step w_{i+1} is given by:

$$W_{i+1}= Wi-(J_K J_K^T-\mu I)e_{Ij}J_k^T \qquad (1)$$

Here J_K stands for the Jacobian Matrix

J_K^T stands for the Transpose of the Jacobian Matrix.

µ stands for step size

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 6 Issue I, January 2018- Available at www.ijraset.com

I stands for an Identity Matrix
J_K is the second order derivative of error (e) with respect to weight (w)

$$\frac{\partial^2 e}{\partial w^2}=J_K \qquad (2)$$

IV. PROPOSED SYSTEM

The final flowchart of the proposed algorithm can be proposed as follows

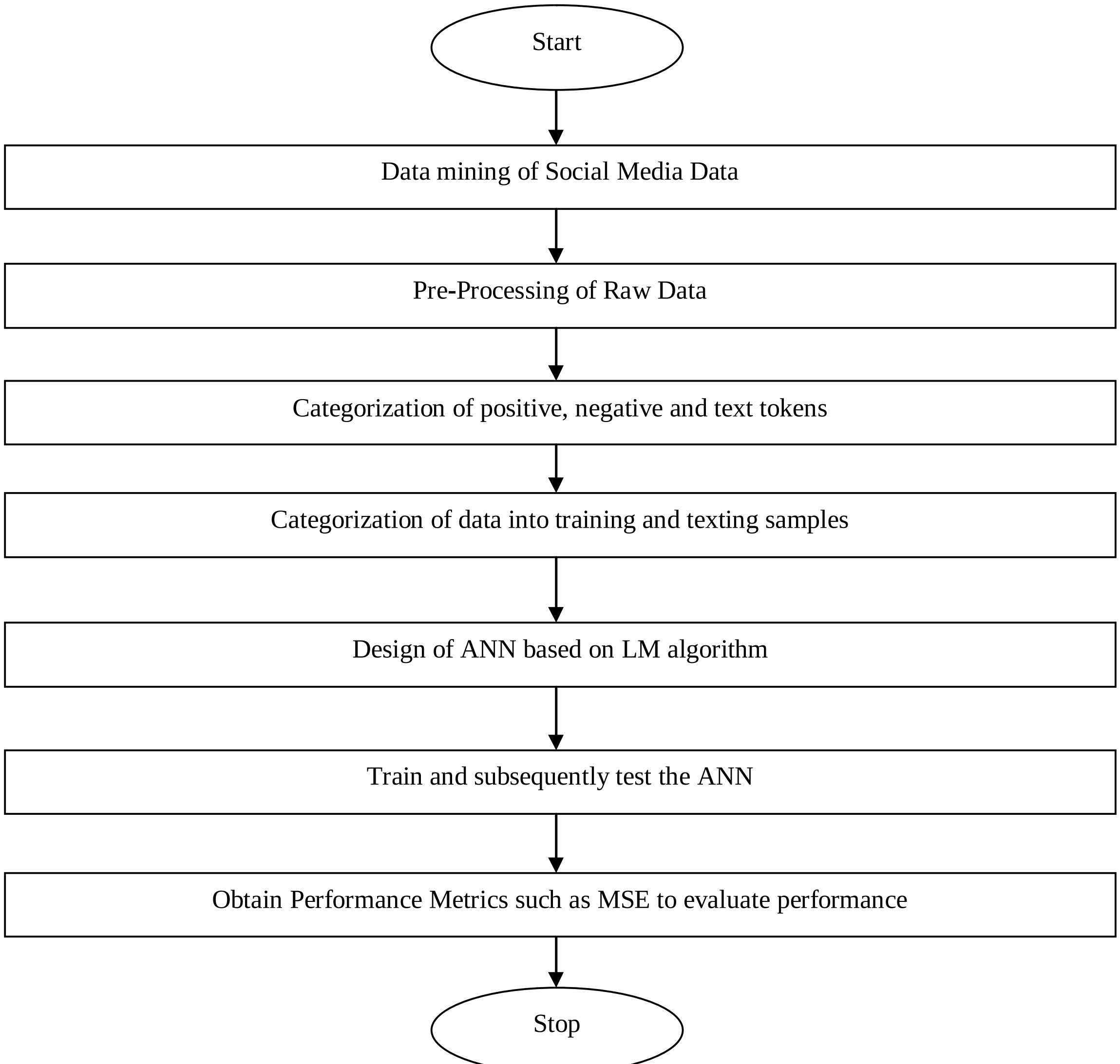

Fig.5 Flowchart of Proposed Algorithm

It should be noted that the performance metric for deciding the efficacy of the proposed algorithm is mean square error defined as:

$$\frac{\sum_{i=1}^{i=N} e_i^2}{N} = MSE \qquad (3)$$

And $e = y_p - y_a$

Here y_p represents predicted output and e represents the error in prediction
And y_a represents actual output
The mean square error evades the possibility of negative and positive errors getting cancelled out thereby rendering greater accuracy to the performance of the system.
Pre-processing involves eliminating single characters and special characters with nulls since they are assumed to have negligible effect on classification. A total of 997 tweets are used for the training and testing process. As a standard convention, 70% of the data has been used for training and 30% of the data has been used for testing.

V. RESULTS

The results obtained based on the proposed algorithm are:

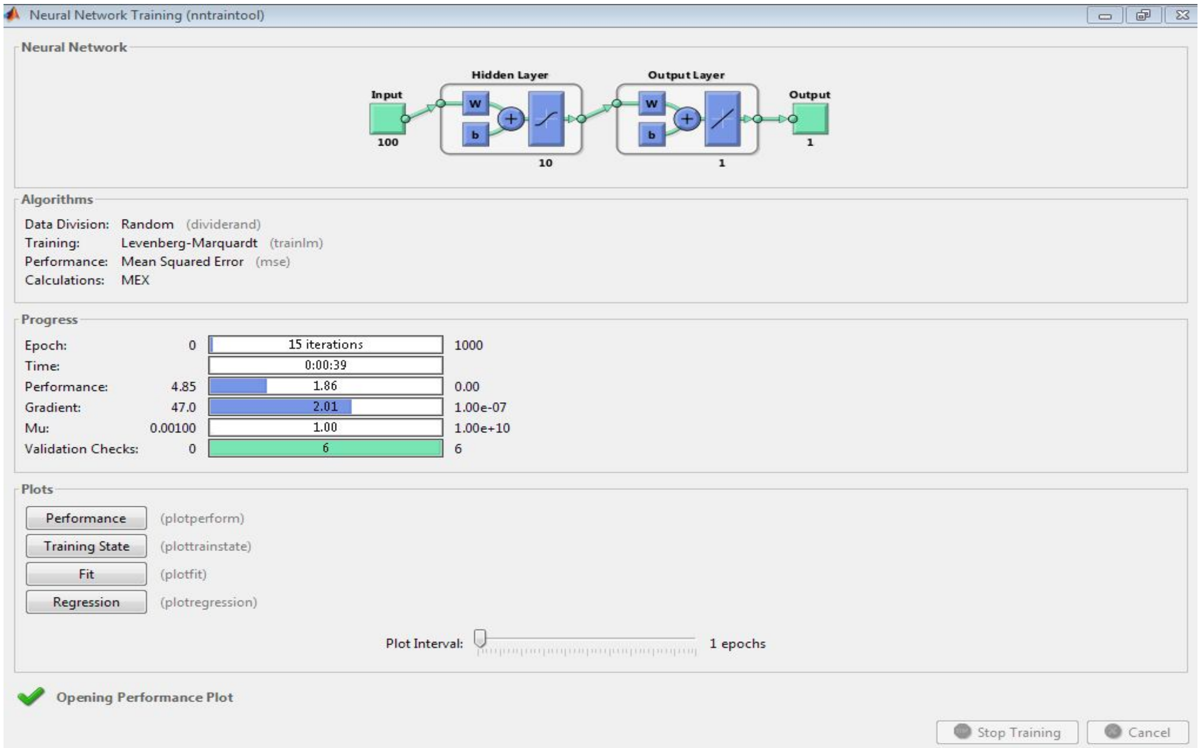

Fig.6 Training the designed Neural Network

It can be seen that the system is designed for 10 neurons in the hidden layer.

The performance function is Mean Square error. The mean square error is 4.85. This signifies that the accuracy of the system can be computed as follows:

$$Accuracy = \frac{Number\ of\ accurate\ Predictions}{Total\ number\ of\ predictions} \qquad (4)$$

Thus the accuracy of the proposed system is 95.15%.

The high accuracy can be attributed to the back propagation mechanism of ANN.

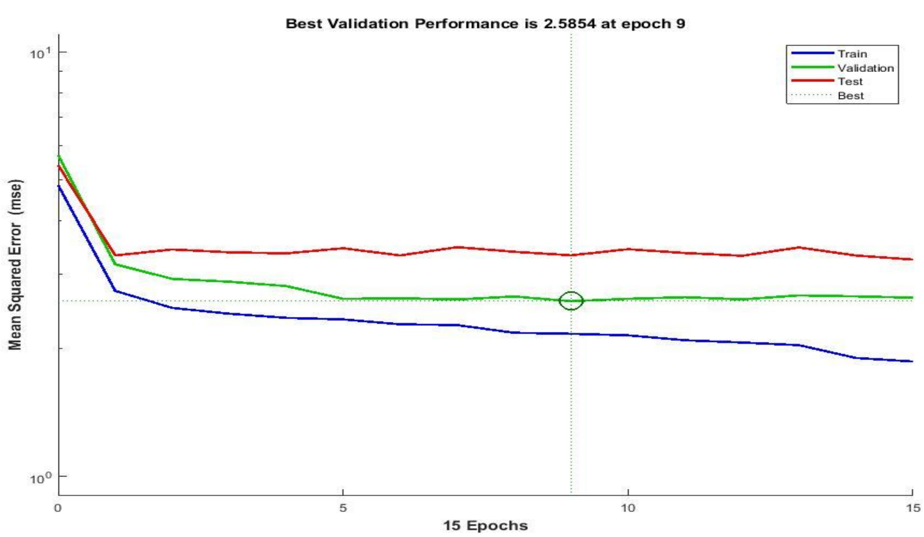

Fig.7 Mean Square Error as a function of number of epochs.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 6 Issue I, January 2018- Available at www.ijraset.com

It can be seen that the mean square error reduces down to 4.85 and the runs of iterations or epochs is 15 in all. After 6 epochs the MSE in the validation almost becomes constant. This indicates towards the fact that the algorithm converges quite quickly.

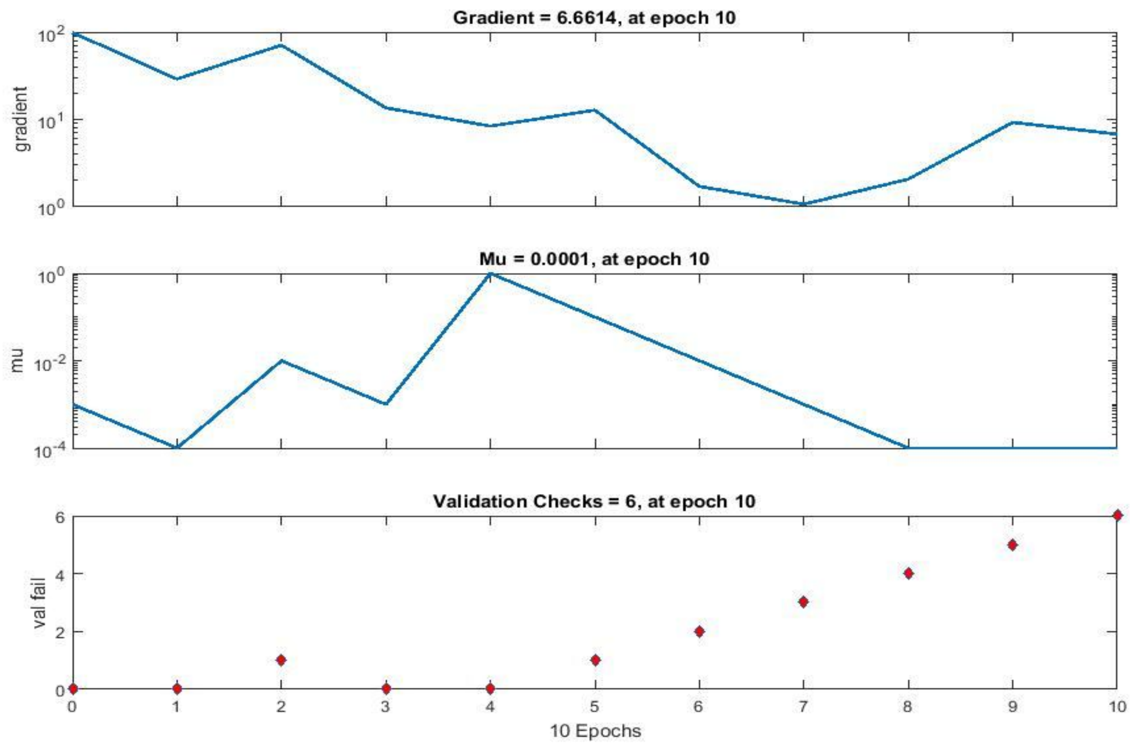

Fig.8 Training States of the ANN

The training states depict the variation of step size (μ) and gradient (g) with the number of epochs. It can be clearly seen that as the error falls by a maximum step size (μ), the gradient shows lesser rate of change thereby confirming the theoretical aspect of back propagation mentioned in the previous sections.

VI. CONCLUSION

It can be concluded from the mathematical background and subsequent results that the proposed system yields high accuracy for text mining data. The data used here are in the form of tweets. The training algorithm used is Levenberg-Marquardt (LM) which yields stability in error prediction. The proposed system uses 997 tweets out of which 70% has been used for training and the rest of the 30% have been used for testing. The data division in to training and testing data sets has been kept random. The number of neurons in the hidden layer has been kept as 10. The accuracy achieved by the propose system in terms of Mean Square Error (MSE) is 95.15%

REFERENCES

[1] Hanjun Lee Business School, Korea “The Influence Of Negative Emotions In An Online Brand Community On Customer Innovation Activities ” 2014 47th Hawaii International Conference on System Science -978-1-4799-2504-9/14 $31.00 © 2014 IEE

[2] Varsha Sahayak et. Al. “Sentiment Analysis on Twitter Data”, International Journal of Innovative Research in Advanced Engineering (IJIRAE), ISSN: 2349-2163, Volume 2, January 2015.

[3] Xia Hu, Lei Tang, Jiliang Tang, Huan Liu, “Exploiting Social Relations for Sentiment Analysisin Microblogging”, permission and/or a fee.WSDM ’13, February 4–8, 2013, Rome, Italy.Copyright 2013 ACM 978-1-4503-1869-3/13/02

[4] Fei Jiang, Anqi Cui, Yiqun Liu, Min Zhang, and Shaoping Ma, “Every Term Has Sentiment:Learning from Emoticon Evidencesfor Chinese Microblog Sentiment Analysis”,c Springer-Verlag Berlin Heidelberg 2013

[5] Eric Baucom,AzadeSanjari, Xiaozhong Liu,Miao Chen, “Mirroring the Real World in Social Media: Twitter,Geolocation, and Sentiment Analysis”,Copyright 2013 ACM 978-1-4503-2415-1/13/10

[6] Min Wang,Donglin Cao, Lingxiao Li,Shaozi Li, RongrongJi, “Microblog Sentiment Analysis Based on Cross-mediaBag-of-words Model”,ICIMCS’14, July 10–12, 2014, Xiamen, Fujian, China.Copyright 2014 ACM 978-1-4503-2810-4/14/07

[7] Felipe Bravo-Marquez, Marcelo Mendoza,Barbara Poblete, “Combining Strengths, Emotions and Polarities forBoosting Twitter Sentiment Analysis”,WISDOM’13, August 11 2013, Chicago, IL, USACopyright 2013 ACM 978-1-4503-2332-1/13/0

[8] Pedro Calais Guerra, Wagner Meira Jr.,Claire Cardie, “Sentiment Analysis on Evolving Social Streams:How Self-Report Imbalances Can Help”,WSDM’14, February 24–28, 2014, New York, New York, USA.Copyright 2014 ACM 978-1-4503-2351-2/14/0

[9] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, “New Avenues in Opinion Mining and Sentiment Analysis,” Intelligent Systems, IEEE , vol.28, no.2, pp. 15-21, 2013.

[10] S. Baccianella, A. Esuli, and F. Sebastiani, ” SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining,” In Proc. of 7th Int’l Conf. on Language Resources and Evaluation, pp 2200-2204, 2010.

[11] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "SentiFul: A Lexicon for Sentiment Analysis," In IEEE Trans. On Affective Computing, vol. 02, issue no. 01, pp. 22-36, 2011

[12] Bollegala, D. Weir, and J. Carroll, "CrossDomainSentiment Classification using a SentimentSensitive Thesaurus," Knowledge and Data Engineering, IEEE Transactions, vol.25, issue no.08, pp. 1,

[13] R. Xia and C. Zong, “A POS-based Ensemble Model for Cross-domain Sentiment Classification,” In Proc. of 5th Int’l Joint Conf. on Natural Language Processing, pp. 614–622, 2011

[14] K. Ghag and K. Shah, "Comparative analysis of the techniques for Sentiment Analysis," In Proc. of Int’l Conf. on Advances in Technology and Engineering, pp. 1-7, 2013

[15] G. Paltoglou and M. Thelwall, “A study of Information Retrieval weighting schemes for sentiment analysis,” In Proc. of 48th Annual Meeting of the Association for Computational Linguistics, pp. 1386-1395, 2010

[16] B. Pang, L. Lee, and S. Vaithyanathan, ”Thumbs up? sentiment classification using machine learning techniques,” In Proc. of Conf. on Empirical Methods in Natural Language Processing, pp 79-86, 2002

[17] T. Mullen and N. Collier, ”Sentiment analysis using support vector machines with diverse information sources,’ In Proc. of Conf. on Empirical Methods in Natural Language Processing,” pp 412–418, 2004.