



## Web Sentiment Analysis: Comparison of Predicted Results with Neural Networks

**PAPPU RAJAN\***Department of Computer Science  
St.Xavier's College ,Palayamkottai  
Tamil Nadu , India**S.P.VICTOR**Department of Computer Science  
St.Xavier's College ,Palayamkottai  
Tamil Nadu , India

**Abstract**— As people are free to say their opinions on anything using various social networking sites like Twitter, Facebook, Discussion forums, and blogs. Particularly Microblogging and text messaging have emerged and become dominated tool over the web. Sentiment analysis, also called Opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics and their attributes. Opining Mining is a new emerging and developing major research area and also it is one of the subdivision of Web Mining system. The significant features of Sentiment is to analysing the overall sentiment for each object by computing the weighted average for all the sentiments in the textual data. In this paper objective deals how neural network systems can be used to compare sentiments with the stock prices for better accurate prediction stock movements.

**Keywords**— Sentiment analysis, Opinion mining, Neural Networks

### I. INTRODUCTION

Natural Language processing tasks is dominating system in early 1990s for all IT fields but later different systems are arriving, among these sentiment or opinion mining also one of the field. All this problems solved by single system that system is called opinion mining or sentiment mining. Here in this paper we used sentiment analysis for predicting stock prices. In this work, we have calculated sentiment score and then the score was compared with stock price of the selected companies from our research. Our previous work was collecting data from twitter. Using score analysis to calculate overall sentiment score of the company. After made sentiment score calculation, these values was used to compare stock price of the company which will help us to predict accuracy of the system. In this paper deals how neural network systems can be used to compare sentiments with the stock prices for better accurate prediction stock movements.

### II. REVIEW OF LITERATURE

William b. Claster and dinh quoc hung et al. [2010] , they proposed opinion model for multidimensional space measuring sentiment . They were following different methodology for self-organizing map with visual model containing comparison of 10 movies with respect to each particular sentiment using twitter tweets.

containing comparison of 10 movies with respect to each particular sentiment using twitter tweets.

Ronan Collobert et al. [2011]. They proposed neural network architecture which can be applied NLP tasks for part of speech, tagging, role labeling. They have developed various algorithm for the above tasks. This system was used to building freely available tagging system also produce good quality performance with minimal computational requirements feasibility.

Bichen Zheng et al.[2013], They analyzed Customer behavior prediction using neural network approaches. In this work deals with multilayer perception, finding relationship with input and output layer. Finally discussed prediction accuracy of ANN comparing with SVM.

### III. COMPARISON OF PREDICTED RESULTS WITH NEURAL NETWORKS

#### A.Motivation and Objective

From the review of literature, a number of approaches are used to improve the accuracy using sentiment score with other techniques. Here we are trying to find the relationship between the Twitter sentiment and stock prices to determine whether the sentiments are good indicators of future stock price movements using ANN. In this research, we are going to using ANN for compare the predicted values which was derived from ALM.

#### B.Data Set

Tweets were collected from five companies and also found sentiment score using positive and negative score analysis for our previous work. The Stock prices of below mentioned companies were collected from the official website of National Stock Exchange (NSE) for a period from 1<sup>st</sup> November, 2013 to 20<sup>th</sup> December 2013. The following are the companies according to verticals:

TABLE . I  
LISTS OF COMPANIES

Company Name	Type of Industry
Bharti Airtel Ltd.	Telecommunication - Services
Titan Industries Ltd.	Retail
Bosch Ltd.	Automobiles
Tata Consultancy Services Ltd.	Computers – Software
Colgate Palmolive (India) Ltd.	FMCG

**C. Overall opinion analysis**

In the previous work, we have found overall sentiment score using score analysis. The lexical sentiment score analysis was performed over five companies. We have analyzed about 2290 tweets per day. So for 50 days from 1<sup>st</sup> November, 2013 to 20<sup>th</sup> December, 2013, a total of 1,14,500 tweets were analyzed for the entire project. The following chart represents overall sentiment trends of the above mentioned company.

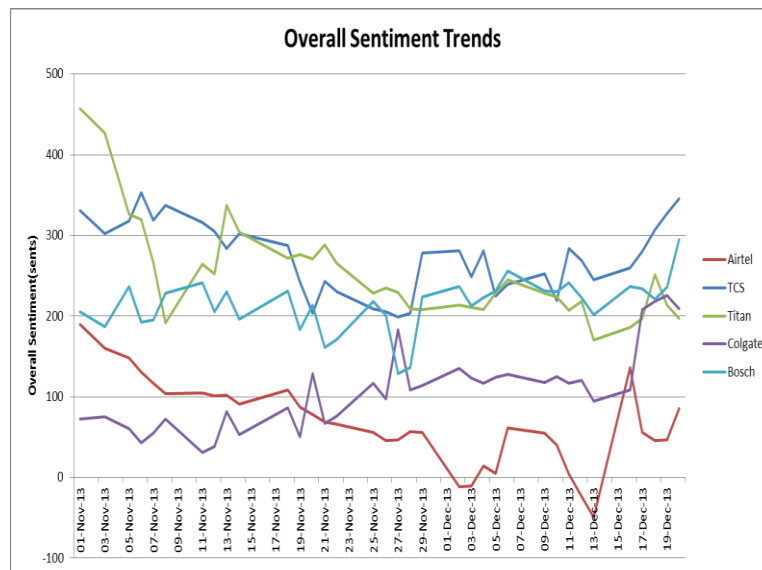


Fig.1 Overall sentiment of the Companies

**D..Neural Networks**

Neural Networks are well set of organized three layers system. These interconnected layers are input, hidden, output. The connections between nodes in each layer can have associated with weights. The weights are iteratively adjusting by algorithm to minimize error and getting better accurate prediction. The neural network is a set of programs and data structures that resemble operation of the human brain. In general neural networks are very effective for event prediction. The neural network will be using different fields and related with different applications. Here we pointed out only few. Process of estimating future events is called sales forecasting. ANN is having plenty facilities to predict sales forecasting which can be helpful to improve the business performance by the way of finding demand of the product, measuring inventory control, understanding consumer buying behaviours. In this research we are going to using ANN for compare the predicted values which was derived from ALM. MLP and RDF are neural networks predictive applications, there can be used to build models for scoring. These are functions of predictors that minimize the prediction error of target variable. Here the predictors are called inputs or independent variable and target variable are called outputs.

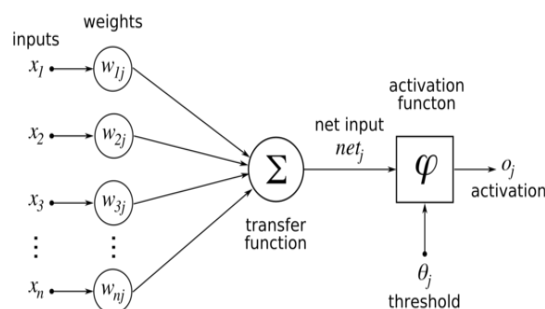


Fig.2 . The architecture of Artificial Neural Network

There are three layers in neural network such as input, hidden, output. The Following fig. shows the architecture of ANN. The learning function or the activation function that was used is sigmoid function. A log-sigmoid function, also known as a logistic function, is given by the relationship:

$$\sigma(t) = \frac{1}{1 + e^{-\beta t}}$$

Where  $\beta$  is a slope parameter

The input layer consist of N units of  $x_i$  ( $i=1,2,3...N$ ) and the hidden layer consist of processing entities of  $W_{ij}$  and one output activation layer  $O_j$ .  $w_{ij}$  is the connection weights between input units and hidden processing units,  $net_j$  is the connection weights between hidden processing units and the output unit. Here  $\phi$  is activation functions for hidden processing elements and output unit respectively. In this research we describe the comparison of prediction using Neural Networks (i.e.) Multilayer Perceptron (MLP) and Radial Basis Function (RBF). The closing prices of all the four companies were predicted by taking the opening price and overall sentiment of that company using both the types of Neural Networks techniques.

#### IV. STEPS FOR COMPARISON OF PREDICTED RESULTS WITH NEURAL NETWORKS

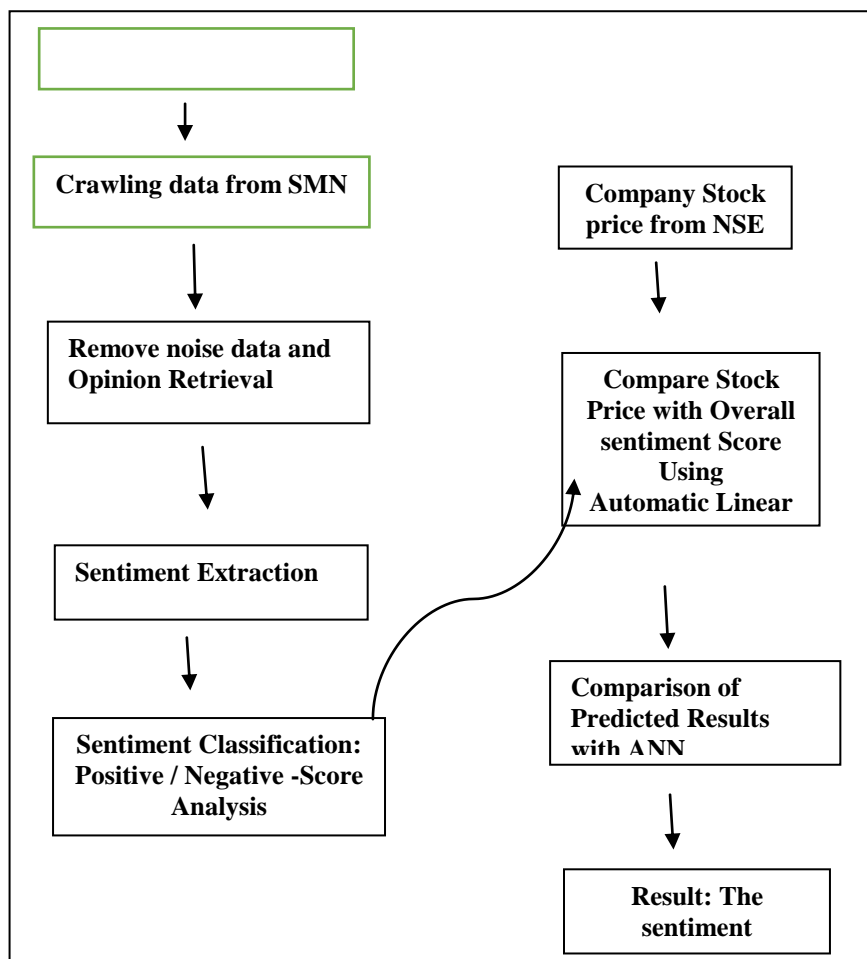


Fig.2 Steps for Comparison of Predicted Results with Neural Networks

In opinion mining are various types of sentiment analysis as: word level , feature-level, entity-level, sentence-level, document-level ., Data set were collected from different companies using Twitter by web crawling. Extracting data from SMN -Twitter : using Twitter API REST API s are having the following resources : Time lines , tweets search, streaming , direct message, friends and followers ,users ,suggested users ,favorites , lists, saved searchers, place and Geo , trends , spam reports, OAuth, help. These APIs use the pull strategy for data retrieval. To collect information a user must explicitly request it. Streaming APIs provides a continuous stream of public information from Twitter. These APIs use the push strategy for data retrieval. Once a request for information is made, the Streaming APIs provide a continuous stream of updates with no further input from the user. Opinion Retrieval involves retrieving desired information from bag-of-words or Twitter textual data to measure ad hoc information retrieval effectiveness in the standard way of methodology. A standard binary assessment of either relevant or non-relevant for each query-document pair. Sentiment Extraction: Finding or discovering of target entity. It uses various method to extract the sentiment from sentiment document using unsupervised learning, supervised learning and lexicon based approach. Sentiment Classification: Positive / Negative -Score Analysis: To find weather a piece of text is opinionated or not and to find the polarity of the text. This classification may be binary or multiclass classification. Up to this step we have already completed and calculated sentiment score. Now in this paper we going to do compare these score value with stock price of the company.

The above given company stock prices above mentioned companies are collected from the official website of National Stock Exchange (NSE) for a period from 1<sup>st</sup> November, 2013 to 20<sup>th</sup> December 2013. The same period we already collected twitter data from the same companies and to using sentiment score analysis technique to calculate sentiment score values. These sentiment score value can be compared with Stock Price to finding the perfect predictors with their values using ALM . Finally compare the predicted result to find prediction accuracy using ANN.

*A.Steps for Comparison of Predicted Results with Neural Networks*

- Step 1 Read opinion data set
- Step 2 Clean up the opinion, to remove noise data
- Step 3 Split whole opinion based sentence into opinion word
- Step 4 Find the number of positive and negative opinion
- Step 5 Compare number of positive and negative opinion from multi set
- Step 6 Score opinion: Number of positive words – number of negative words
- Step 7 Collect Company Stock price from NSE
- Step 8 Compare Stock Price with Overall sentiment Score
- Step 9 Find the Predictors using ALM
- Step 10 Comparison of Predicted Results using ANN
- Step 11 Result: The sentiment prediction with accuracy

**V. RESULTS AND DISCUSSION**

*A..Comparison of Predicted Company Closing Prices with Artificial Neural Networks*

F-tests are performed by using company Actual value with ALM Predicted values, MLP Predicted Value and RBF Predicted Value. The following table represents the F-test for analyzing the variances between the samples:

**TABLE. 2 F-TEST TWO-SAMPLE FOR VARIANCES OF COMPANY (TCS)**

		N	Std. Deviation	Variance	df	F Critical two-tail	Result
Pair 1	Company Closing Price	34	41.37169	1711.616	33	0.32	Cannot Reject Null Hypothesis because $p > 0.05$ (Variances are the same)
	ALM Predicted Value	34	34.75203	1207.696	33		
Pair 2	Company Closing Price	34	41.37169	1711.616	33	0.57	Cannot Reject Null Hypothesis because $p > 0.05$ (Variances are the same)
	MLP Predicted Value	34	37.48353	1405.017	33		
Pair 3	Company Closing Price	34	41.37169	1711.616	33	0.69	Cannot Reject Null Hypothesis because $p > 0.05$ (Variances are the same)
	RBF Predicted Value	34	38.60305	1490.180	33		

From the table above we can see that the variances for Company Actual closing Price and ALM Predicted Value are 1711.62 and 1207.70. The variance for MLP Predicted Value is 1405.01 and RBF Predicted Value is 1490.18. The Company actual closing price has a standard deviation of 41.37, while value predicted using ALM has a standard deviation of 34.75, price predicted using Multi-layer Perceptron has a standard deviation of 37.48 and Radial basis function (RBF) achieved much less deviation of 38.60. The total sample size taken for this analysis is represented by ‘N’ which in this case 34. The F-Critical value (two-Tail) for the pair 1 (i.e.) Company Closing Price and ALM Predicted value is 0.32 which is greater than 0.05, therefore we accept the null hypothesis (H0) that the variances are the same. Similarly for the second pair (i.e.) Company closing Price and MLP Predicted Value is 0.57 which is greater than 0.05, therefore the variances are statistically same. The F-critical value for pair 3 (i.e.) Company Closing Price and RBF Predicted Value is 0.69 which is also greater than 0.05, so we accept the null hypothesis that means variances are the same.

Since the variances between the variables are the same for all the cases, we can apply T-test assuming equal variances. The following table represents the T-Test for two sample assuming equal variances along with mean, variance, standard error mean, t-value, degree of freedom and the significance value (2-sided):

**TABLE . 3 T-TEST TWO-SAMPLE ASSUMING EQUAL VARIANCES OF COMPANY(TCS)**

		N	Mean	Variance	df	T Critical two-tail	Result
Pair 1	Company Close Price	34	2037.94412	1711.616	66	0.98	Cannot Reject Null Hypothesis because $p > 0.05$ (Means are the same)
	ALM Predicted Value	34	2038.10029	1207.696			
Pair 2	Company Close Price	34	2037.94412	1711.616	66	0.55	Cannot Reject Null Hypothesis because $p > 0.05$ (Means are the same)
	MLP Predicted Value	34	2032.25029	1405.017			
Pair 3	Company Close Price	34	2037.94412	1711.616	66	0.98	Cannot Reject Null Hypothesis because $p > 0.05$ (Means are the same)
	RBF Predicted Value	34	2037.75059	1490.180			

The mean for Company actual price and ALM predicted value are 2037.94 and 2038.10, whereas the mean for MLP Predicted value and RBF Predicted value are 2032.25 and 2037.75 respectively. The total sample size taken into account for the entire analysis is 34 for ease case. The degree of freedom is computed with the formula  $(N1+N2-2)$  which is 66. Here the main objective is to verify whether there is a significant difference in their means of two variables or the means are the same. So, in this analysis there are four variables taken into account (i.e.) the actual closing price of Company , ALM predicted value , MLP predicted value and RBF predicted value. The t-value for Company’s actual closing price and ALM predicted value is 0.98; which is greater than 0.05 ( $p>0.05$ ), therefore we accept the null hypothesis ( $H_0$ ), that there is no difference in the mean. The t-value for Company’s actual price and MLP predicted value is 0.55, which is very much higher than 0.05, so we cannot reject the null hypothesis that the means of the two variables are same. The t-value for Company’s actual price and RBF predicted value is 0.98 which is greater than the standard value of 0.05; therefore the null hypothesis is accepted that the means between the two variables are equal.

## VI. CONCLUSIONS

Sentiment analysis is also used to take importance decision which can be improve the business performance and their potential role as enabling technologies for other sub systems. In this study shows the company sentiment score and their stock price having strong relationship among them. The predicted results of Automatic Linear Modeling and Artificial Neural Networks are statistically similar but ALM is easier to perform in terms of analysis than complicated ANN. If the overall sentiment for a firm is consecutively declining for three days, then it is bound to decrease for another two days. The predicted values of closing prices by Automatic Linear Modelling and Neural Network MLP and RBF are compared and the variances are found to be equal. For further research we are going use ranking system which will help us to compare and lead for better accuracy for prediction.

## REFERENCES

- [1] Y.F. Sun, Y.C. Liang, W.L. Zhang, H.P. Lee, W.Z. Lin, and L.J. Cao. (2005), Optimal Partition Algorithm of the RBF neural Network and Its Application to Financial Time Series Forecasting, *Neural Computation & Application*, pp. 36–44.
- [2] S. Charles, and L. Arockiam, (2012), Feature Selection: A New Perspective, *International Journal of Data Mining and Knowledge Engineering*, Print: ISSN 0974-9683 & Online: ISSN 0974-9578.
- [3] Paudel, J. and Laux, J. (2010), A Behavioral Approach To Stock Pricing. *The Journal of Applied Business Research*, Vol. No. 26, No.3, pp. 99-106.
- [4] B. Pang and L. Lee. (2008) ,Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval*, Vol. 2, No. 1–2, pp. 1–135.
- [5] N.Jakob, S.H. Weber, M.C. Muller, I. Gurevych. (2009) ,Beyond the Stars: Exploiting Free-Text User Reviews to Improve the Accuracy of Movie Recommendations, *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*.
- [6] B. Jansen, M. Zhang, K. Sobel, and A. Chowdury. (2009), Twitter power: Tweets as electronic word of mouth, *Journal of the American Society for Information Science and Technology*, *Journal of the American Society for Information Science And Technology*, 60(11):2169–2188.
- [7] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. (2011) ,Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pp.227, New York, NY, USA. ACM.

- [8] J. Bollen, H. Mao, and X. Zeng.(2011) , On the Predictability of Stock Market Behavior using StockTwits Sentiment and Posting Volume ,Journal of Computational Science,Vol.2(1),pp.1-8.
- [9] R. Collobert, K. Kavukcuoglu, and C. Farabet. Torch. (2011) ,A matlab-like environment for machine learning. In BigLearn, NIPS Workshop Science , Vol.2(1),pp.1-8.
- [10] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa.(2011) , Natural language processing from scratch, Journal of Machine Learning Research, 12:2493–2537.
- [11] N. A. Diakopoulos and D. A. Shamma.(2010), Characterizing debate performance via aggregated twitter sentiment, In Proceedings of the 28th international conference on Human factors in computing systems, CHI '10, pp. 1195–1198, New York, NY, USA, ACM.
- [12] J. Weston, F. Ratle, and R. Collobert. (2008), Deep learning via semi-supervised embedding, In International Conference on Machine learning (ICML), pp. 1168–1175.
- [13] Hastie, T., Tibshirani, R & Friedman, J. (2011), The elements of statistical learning: Data mining, inference, and prediction, Second Edition Series, Springer Series in Statistics,XXII ,745,p.282 illus.
- [14] Kantardzic, M. (2011), Data mining: Concepts, models, methods, and algorithms ,(2nd ed, John Wiley & Sons, Inc. Hoboken, New Jerrcy, ISBN 9780 -9 -470 -89045-5.
- [15] D. C. Ciresan, U. Meier, J. Masci, J. Schmidhuber (2012), Multi-Column Deep Neural Network for Traffic Sign Classification Neural Networks, The International Joint Conference on Neural Network, IDSIA - USI - SUPSI | Galleria 2, Manno - Lugano 6928, Switzerland.
- [16] H. Mao, S. Counts, and J. Bollen.(2011), Predicting financial markets: Comparing survey, news, twitter and search engine data, arXiv preprint arXiv:1112.1051.
- [17] S. Kogan, D. Levin, B.R. Routledge, J.S. Sagi, and N.A. Smith.( 2009), Predicting risk from financial reports with regression, In NAACL '09, pp. 272–280.
- [18] Oh, C. and Liu Sheng, O. (2011) , Investigating Predictive Power of Stock Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movements, Knowledge Management and Business Intelligent ,Thirty Second International Conference on Information Systems, Shanghai,. pp. 5--17.
- [19] Helen Susannah Moat, Chester Curme, Adam Avakian, Dror Y. Kenett, H. Eugene Stanley and Tobias Preis (2013), Quantifying Wikipedia Usage Patterns Before Stock Market Moves.Scientific Reports .
- [20] Tobias Preis, Helen Susannah Moat and H. Eugene Stanley (2013), Quantifying Trading Behavior in Financial Markets Using Google Trends, Scientific Reports .
- [21] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welpe (2010), Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment, Association for the Advancement of Artificial Intelligence.
- [22] Davidov, D., Tsur, O., and Rappoport(2010), A Semi-supervised recognition of sarcastic sentences in twitter and amazon,CoNLL '10 Proceedings of the Fourteenth Conference on Computational Natural Language Learning,pp. pp. 107-116.
- [23] Bissattini C., Christodoulou K.(2013),Web sentiment analysis for revealing public opinions, trends and making good financial decisions, Journal of Advanced Research in Computer Science and Software Engineering.
- [24] Tulankar S.,Athale R.,Bhujbal S (2013), Sentiment analysis of equities using data mining techniques and visualizing the trends, International Journal of Computer Science Issues. , Vol. 10, Issue 4, No 2, ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784.
- [25] William B. Claster and DINH Quoc Hung, Subana Shanmuganathan.(2010), Unsupervised Artificial Neural Nets for Modeling Movie Sentiment , Computational Intelligence, Communication Systems and Networks (CICSyN),Second International Conference,IEEE, pp.349-354, print ISBN 978-4244-7837-8