Using Sentimental Analysis in Prediction of Stock Market Investment

Sunil Kumar Khatri[1], Ayush Srivastava[2]

[1,2]*Amity Institute of Information Technology*
Amity University Uttar Pradesh, Noida, India
[1]*sunilkkhatri@gmail.com, skkhatri@amity.edu*
[2]*ayush.idea77@gmail.com*

Abstract: **Sentimental Analysis is one of the most popular technique which is widely been used in every industry. Extraction of sentiments from user's comments is used in detecting the user view for a particular company. Sentimental Analysis can help in predicting the mood of people which affects the stock prices and thus can help in prediction of actual prices. In this paper sentimental analysis is performed on the data extracted from Twitter and Stock Twits. The data is analyzed to compute the mood of user's comment. These comments are categorized into four category which are happy, up, down and rejected. The polarity index along with market data is supplied to an artificial neural network to predict the results.**

Keywords: **Sentimental Analysis, Stocks Markets, Media, Moods, Artificial Neural Network.**

I. INTRODUCTION

Sentimental Analysis also known as Opinion Mining is an area that uses Natural Language Processing and Text Analysis that helps in building a system that identifies and extract information in source material. An initial task in sentimental analysis is to determine the polarity of a specified text at the document level, sentence level or aspect level. In core, it is a process that helps in determining the emotional level behind a sequence of words, used to gain an insight of speaker's attitude, opinions and emotions expressed in a sentence. Sentimental Analysis is very useful in social media monitoring as it provides an insight of public opinions for certain topics. The uses of sentimental analysis are very extensive and powerful. Sentimental Analysis provides the ability to extract insight from social data which is broadly used by various organizations across the world[1]. Prediction of the market and stock prices for a company has always been a wide area for the researchers to work upon. A company can be successful in long run only if its consumers are happy with its performance and are giving positive feedback for its products. Expedia Canada used this technique to quickly understand consumer attitude which increased in a negative feedback towards one of their television advertisements[2]. Sentimental Analysis was applied to huge scale twitter data in order to find the collective mood states and an exactness of 86.7% was found of Dow Jones Industrial Regular daily directions[3]. Sentimental Analysis is a widely used field giving many benefits to every industry. Thus if sentiments are correctly categorized and their polarity are correctly determined they can be helpful in enhancing a company's performance and making its investors happy.

In our research work we have performed analysis on sentiments collected from yahoo and trained the artificial neural network with the results and stock prices of five top I.T. companies to predict the return of investment for the future day.

II. LITERATURE REVIEW

A growing amount of literature is devoted to developing new tools and models for sentimental analysis. Previous studies had concentrated on a large group of population using social information in prediction of consumer's attitude towards a company[4]. Numerous studies showed that using a mix approach can improve classification scheme [5]. The most common use of sentimental analysis is analyzing of twitter tweets and demonstrating the top trends in marketplace [6] Sentimental analysis is also used in sales forecast of a product by examining tweets and posts from face book[7]. Sentimental analysis had been conducted on stock linked tweets which were collected for a period of 6-month[8]. In order to reduce noise, selection of tweets containing tags of top 100 companies was considered. Each tweet was classified using a Naive Bayes method and a set of 2,500 tweets were trained. Results displayed that sentiment indicators are related with unusual returns and stock volume is linked with trading volume[9]

Sentimental Analysis was applied to tweets extracted from Twitter and news headlines to generate new predictors for investment[10]. From the collected data, they choose a random sample and defined each tweet as bullish or bearish if it contains those terms. They displayed that Twitter sentiment indicator and the occurrence of monetary terms on Twitter are statistically significant predictors of regular market returns. Sentimental analysis was also performed on a micro blogging service entirely devoted to stock market[11] They collected 62,100 blog posts from stocktwits.com, for a time span of three months. The sentiment of the posts was classified using a machine learning algorithm known as J48 classifier to generate a learning model. They proved that the mined sentiment have strong analytical value for coming market directions. Sentimental Analysis was used to forecast the closing index of Tata Services and an accuracy of 85.99% was found in the

process[13].Sentimental analysis is often used to build a social behavior graph on human's online behavior to find the correlation between trading and volume prices of stocks [14].Sentimental Analysis was also performed on the data extracted from SentiWordNet using a hybrid selection model to show how market trends affects a product popularity and rate[15].

III. METHODOLOGY

An Artificial neural network is a computational model which is based on the structure and functions like a biological neural network. Information that is supplied within a network affects it as it changes or learns with that information.

An artificial neural network is a structure which is similar to a human brain. It contains highly interrelated structure of neurons. These neurons act as an input which activates the system [1]. Thus an artificial neural network is mostly preferred by the researchers to use in problems which involves computational tasks, analysis, finding similarities and much more. Data used in our research work was primarily collected from stock dedicated website i.e Stock Twits. All the tweets were mined from Stock Twits on which sentimental analysis were performed. For the market data, yahoo acted as our means of source. The data extracted from yahoo consists five parameter as indexes which involves opening, closing, high, low and volume, out of which only first four were our target parameters. The market data used for this research work was collected from a period of 01/01/2015 to 22/02/2016.

The implementation process was performed in MATLAB. MATLAB delivers a best validation graph to show the error and the least value of epochs which will give best results.

IV. PROPOSED MODEL

Monitoring social trends have been always proven to be helpful in detecting user's point of view towards a particular company. In our research work, we have collected data for top five companies i.e. Apple, Microsoft, Oracle, Google and Facebook in form of comments and tweets from Stock Twits.

This data is further classified on the basis of polarity indexes used in our work which are happy, up, down and rejected to compute the overall index of the sentiments and to determine whether the sentiments are positive, negative or neutral towards a specific company. The overall index values and the market data of that company became the inputs which are passed to an artificial neural network to train and predict. This input data for an artificial neural network is in form of comma separated delimiter (.csv) workbook where first four columns determine the market data indexes and last column determine the overall index value of sentiments for that company. The entire data is passed to a neural network to train and predict the closing value of stock price for a company. This step is performed for all five companies and their closing price are compared at last to tell in which company an investor should invest.

The sentimental analysis of these tweets are calculated by computing the index values of each category and overall sentence level score is calculated using equation 1.

$$\text{Score} = \sum (\text{index value of each category}) / N \quad (1)$$

Where N implies numbers of parameters.

In our work we have taken four parameters which are happy, up, down and rejected. The score of the data is calculated in the range of [0,1] . This score is added to csv sheet of each company and is given to an artificial neural network to train and predict the closing value. The output prediction parameter is calculated through the daily return of investment method. The predicted value for a particular company is calculated using equation 2.

$$\text{Daily return} = \text{closing value} - \text{opening value} \quad (2)$$

We have trained the network using 75% of data and 15 % of data is used for testing purpose while remaining 10% of data is used for validation.

V. EXPERIMENTAL RESULTS

An artificial neural network is an effectual tool which helps in prediction and performing sentimental analysis because of its structure which represents the structure of human brain [1]. The sentimental score with market values were provided to an artificial neural network to predict the future market value.

Table I represents the sentimental score performed on the tweets collected for a period of month for each company. The score is calculated using equation 1.

TABLE I: SENTIMENTAL SCORE OBTAINED

Company Name	Parameters				Score
	Happy	Up	Down	Rejected	
Apple	0.54	0.56	0.35	0.37	0.46
Google	0.20	0.62	0.68	0.06	0.39
Microsoft	0.46	0.50	0.78	0.62	0.59
Oracle	0.44	0.30	0.42	0.60	0.44
Facebook	0.92	0.82	0.88	0.78	0.85

The sentence score were added to the market data and were given to a neural network as input. In our network we taken one hidden layer and numbers of neurons are changed to get better accuracy. In order to train our network we have used Levenberg-Marquardt algorithm and used mean square error to calculate errors.

An activation function is a function that transform input values to output values while restricting the output value of a neuron. In our network we have used log-sigmoid function as an activation function for our network. The standard equation of a sigmoid function is

$$f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

The neural network used in our research work is feed forward network which enables a neuron to send its output to another neuron as input and output of these neurons are transferred to hidden layer as input and whose output finally become the input of output layer.

Table II shows the output value calculated for each company with the mean square error.

TABLE 2: RESULTS OBTAINED

Company Name	No. of Neurons	MSE
Apple	10	0.14
Google	12	0.27
Microsoft	10	0.18
Oracle	9	0.22
Facebook	15	0.28

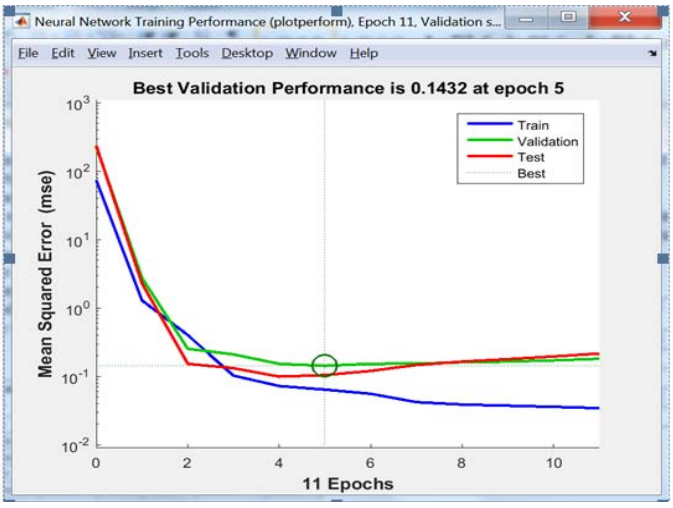

Fig. 1. Implementation of Apple's data

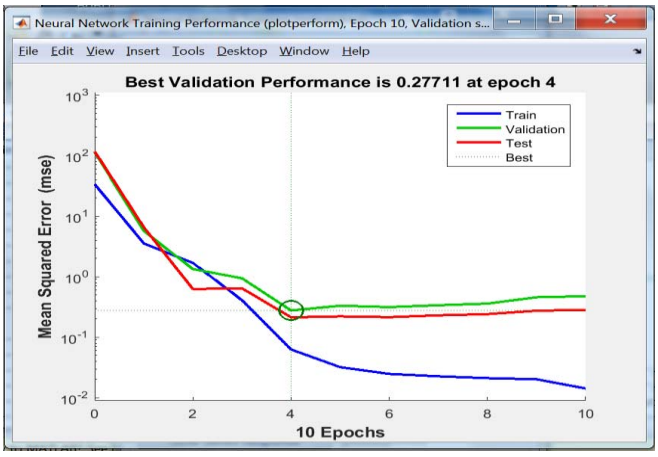

Fig. 2. Implementation of Google's data

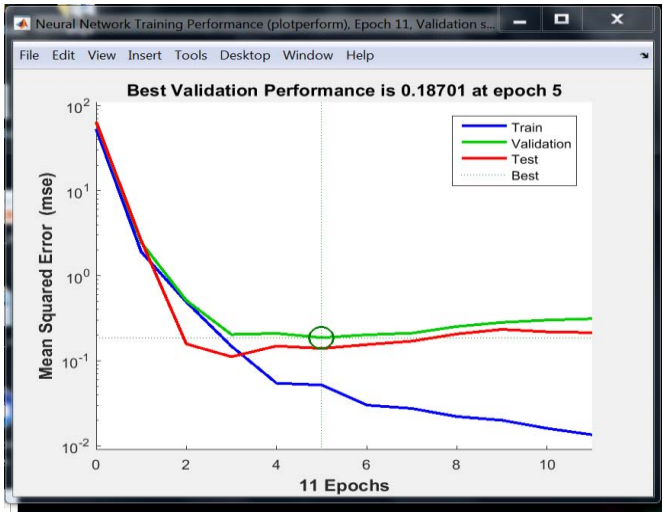

Fig. 3. Implementation of Microsoft's data

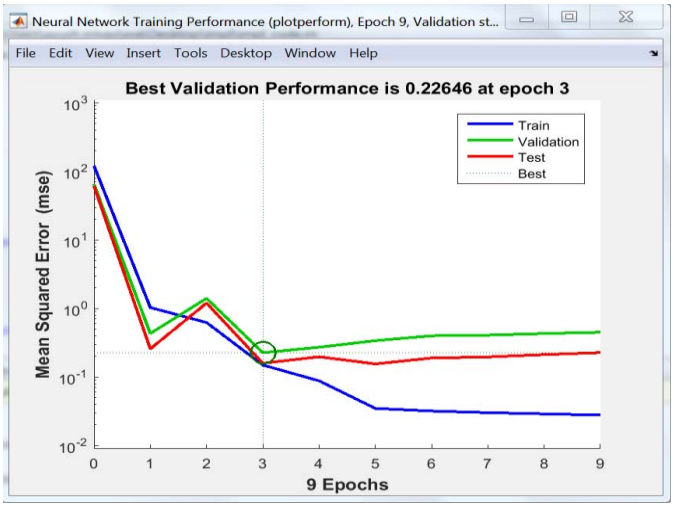

Fig. 4. Implementation of Oracle's data

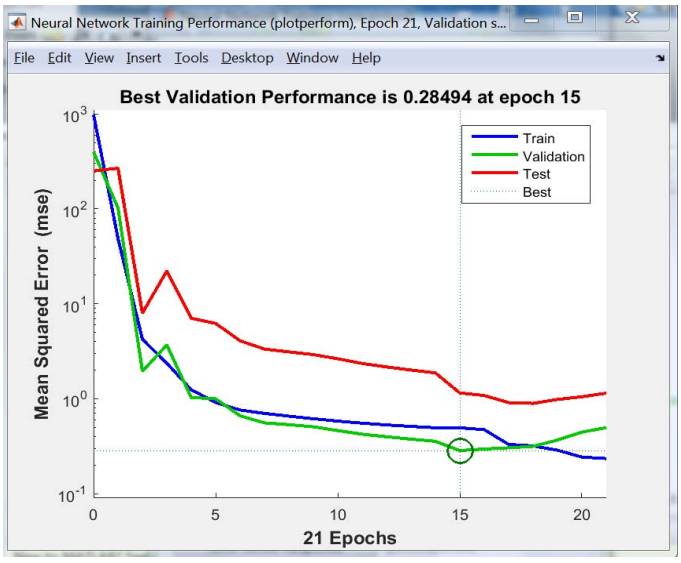

Fig. 5. Implementation of Facebook's data

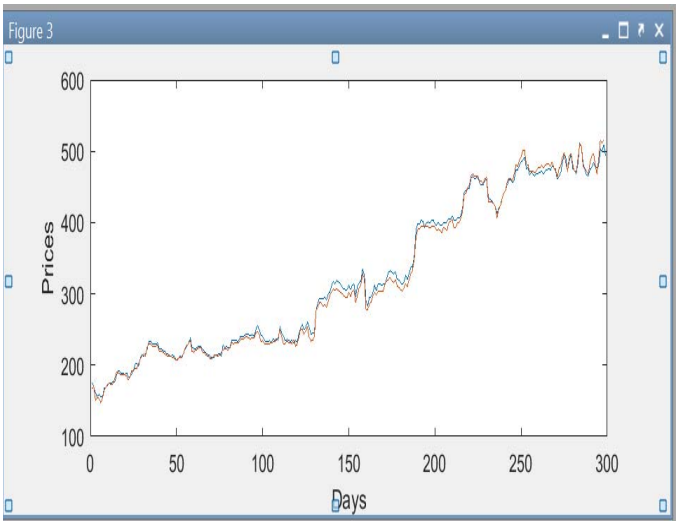

Fig. 6. Prediction graph for each day vs actual prices graph.

Table II and Fig. 1,2,3,4,5 indicates the implementation of Apple, Google, Microsoft, Oracle and Facebook data respectively with minimum mean square error.

Fig. 6 demonstrates the graph of predicted value for each day vs. the actual market closing prices of each day. The line in blue signifies the predicted price whereas line in red signifies actual price. Fig. 6 indicates the closing price of a company at the end of 300th days.This company's closing price is high but due to its low sentimental score it will not be recommended to user for investing.

VI. CONCLUSION

Sentimental Analysis provides the ability to analyze the opinions of people for a particular product or for a company. Prediction of stock market is really a hard nut to crack and requires lot of efforts. The market data if analyzed in a proper way can be very effectual in predicting a company's future. We have mined data and trained a neural network to predict the closing price. Though the closing prices are high of a company but due to sentence score, investment in that company will not be a good decision.

Instead of investing in a company whose closing prices are high, we will recommend you to invest in a company whose sentimental score is high and positive, there are high chances for its stock prices to go up in future. We have ensure that the error rate while performing all implementation is reduce to the least. This work can be extended for a better output if the data samples are taken for a much longer period.

ACKNOWLEDGEMENT

Authors express their deep sense of gratitude to the Founder President of Amity University, Dr. Ashok K Chauhan for his keen interest in promoting research in the Amity University and has always been an inspiration for achieving greater heights.

REFERENCES

[1] Sunil Kumar Khatri, Himanshu Singhal and Prashant Johri. Sentimental analysis to Predict Bombay Stock Exchange Using Artificial Neural Network, Proc. Of ICRITO,2014, pp. 380-384.

[2] Bhat, A.A. and Kamatath,S.S. Automated Stock Price Prediction and Trading Framework for Nifty Intraday Training, 4th International Conference on Computing Communication and Network Technologies (ICCNT),2013, pp. 1-6.

[3] Neethu M.S and Rajasree R. Sentiment Analysis in Twitter Using Machine Learning Technique, 4 ICCC NT,2013,pp. 1-5.

[4] Vincent Martin.Predicting the French Stock Market using Social Media Analysis, 8th International Workshop on semantic and social media adaption and Personalization, IEEE,2013, pp. 3-7.

[5] M.P.RajaKumar and Dr. V. Santhi. A Comparision of stock Trend Prediction using Accuracy riven neural network variance.Proc. Of INT, conf. on Control, Commmunication and power engineering,2012, pp. 73-79.

[6] Hui Song,Yingxiang Fan,Xiaoqiang Liu and Dao Tao. Extracting product features from online reviews for sentimental analysis, Computer Science and Converge Information Technology,2014, pp. 741-750.

[7] Xinzhi Wang and Xiangfeng Luo.Sentimental Space Based Analysis of User Personalized Sentiments.Ninth International Conference on Semantics, Knowledge and Grids,2013, pp. 151-156.

[8] Lu Yonghe and Chen Jianhua. Public Opinion Analysis of Microblog Content, Proc.Of International Con.On Information Science and application, 2014, pp. 1-5.

[9] Sprenger and Webye.Sentiment Analysis of Stock Market News with Semi-supervised Learning, International Conf. on Computer and Information Science,2012, pp. 325-328.

[10] Bin Wen, Wenhua Dai and Junzhe Zhao.Sentence. Sentimental Classification Based on Semantic Comprehension, International Symposium on ISICS,2012, pp. 458-461.

[11] M.P. Rajakumar, Dr.V.Shanthi. A Comparison of Stock Trend Prediction Using Accuracy Driven Neural network variants,Proc. of Int. Conf. on Control, Communication and Power Engineering, 2013,pp.73-79.

[12] Zheng Chen.Study of Stock Prediction Based on Social network, International conference of social Computing, 2013, pp. 913-916.

[13] A. Victor Devadoss; T Antony Alphonnse Ligori. Stock Prediction using Artificial Neural Network, International Journal of Data Mining Techniques and Applications Vol:02, December 2013, pp. 283-291.

[14] Phayung Meesad, Jiajia Li. Stock trend prediction relying on text mining and sentiment analysis with tweets, Fourth World congress on Information and Technology,2014,pp.257-262.

[15] Thamer M.jamel, Ban M.Khammas. Implementation of a Sigmoid Activation Function for Neural Network Using FPGA, 13th Scientific Conference of Al-Ma'moon University College, 2013, pp.1-10.