Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2016, 3(1): 26-33

Research Article ISSN: 2394 - 658X

Stock Market Analysis Using Artificial Neural Network on Big Data

Minal Punjo Bharambe and S C Dharmadhikari

Department of Information Technology, Pune Institute of Computer Technology, Pune, India
minal.bharambe@gmail.com

ABSTRACT

Big Data concern with large amount of data which is complex and continuously increasing. We can relate such Big Data with Stock Market Prediction System. It becomes the challenging task for time series prediction of financial market. It will predict the future for stock prices, whether increase or decrease. There are numbers of machine learning techniques applied to predict the stock market such as Genetic Algorithm (GA), Support Vector Machines (SVM) and Artificial Neural Network (ANN). We can get more accurate prediction value with the help of artificial neural network. We have used multilayer perceptron neural network. Artificial Neural Network can recollect information of any number of years and it can foresee the element in view of the past information. If we co-relate data mining with neural network then we can get lots of achievements. Utilizing these facilities for big data is not yet deployed. So we are just trying to utilize them and come up with the solution which is suitable for Stock Market Analysis.

Key words: Artificial Neural Network, MLP, Prediction System, Stock Price Classification

---

INTRODUCTION

Researcher's attentions are attracted greatly for many years in forecasting stock return or a stock index. It involves an assumption of fundamental information that is publicly available in the past that has some projecting relationships to the future stock returns or indices. It can be represented by interest rates and exchange rates, consumer price, growth rate of industrial production, income statements and dividend yields which can be categorized as economic variables, industry specific information and company specific information respectively. Huge Data begins with huge volume, (HACE portrayed information) heterogeneous as diverse sorts of representation for same people, self-ruling sources as every information source has the capacity produce and gather data without including any brought together control to give dispersed and decentralized control, and looks to investigate mind boggling and advancing connections among data[1]. One of the legitimate hotspots for the business investigation is budgetary online groups. Online networking substance can be examined with 3 noteworthy methods; we can say them as the wellsprings of examination and 3 levels. To start with source is group delivered unstructured stream of web activity as posts and twits regarding client, time and stock. A marker of the supposition communicated is second source. Market forecast models are in view of an opinion file that gives the day by day crude movement a positive/negative heading. Third one is behavioral/social data instead of content. All these three sources are amassed at client level, group level and multi-group level.

Artificial Neural Network (ANN) is a technique that is heavily researched and widely used in stock market prediction [3-4]. Due to dynamic nature of market, it becomes difficult to predict it. In technical analysis, it is believed that market timing is key. The Charts and modeling techniques are utilized to identify trends in price and volume which relies on historical data in order to predict future outcomes. Stock market research encapsulates two elemental trading philosophies; fundamental and technical approaches [Technical-Analysis 2005]. In fundamental analysis, stock market price movements are believed to derive from a security's relative data. Fundamentalists use numeric data, for example, profit, proportions, and administration viability to focus future gauges. In specialized investigation, it is accepted that market timing is key. Experts use outlines and displaying strategies to distinguish drifts in cost and volume. These later people depend on authentic information keeping in mind the end goal to foresee future results.

Data Mining

Data mining is a method of extracting unknown projecting information from large databases which is a widespread technology that helps organizations to focus on the most important information in data repositories [8]. Knowledge driven decisions helps for future trends prediction with data analysis. It helps in finding the hidden patterns, missed projecting information. On the other hand, neural network is another significant method for stock predictions due to their ability in dealing with uncertain, fuzzy and insufficient data which may fluctuate rapidly in very short period of time. Plentiful research and application of neural network has proved its effectiveness over traditional methods that exclude artificial intelligence.

Neural Network

A neural network is a massively parallel distributed processor made up of simple processing unit which has a capability for storing experiential knowledge and making it accessible for utilization. A neural system is an enormously parallel appropriated processor made up of straightforward preparing unit which has a characteristic inclination for putting away experiential learning and making it accessible for utilization. Neural network can be used in speech recognition, signal processing, financial forecasting, monitoring of process control and monitoring samples for analysis.

LITERATURE REVIEW

There are several theories available for predicting future stock market. Analysis on large data sets is highly important in data mining. Boonkiatpong and Sinthupinyo [5] give a new approach which can work efficiently with the neural networks on large data sets. Here separate segments were formed to divide the data, and same network structure is used to learn them. Then all weights from the set of networks are integrated. The results from the experiments show that their proposed method back propagation Neural Network can preserve the accuracy while the training time is dramatically reduced. According to their experiment that weights of small dataset converged faster than the original dataset. However, the weights trained from sub dataset did not achieve better result when they were tested on the original dataset. Hence, they introduced weight integration approach. After the rule integration, the accuracy of the weight obtained from the proposed method is better than the weight set which is the best among the sub datasets. Wu and Wu [1] present a HACE theorem that characterizes the features of the Big Data revolution, and from the data mining perspective it proposes a Big Data processing model. This information driven model includes interest driven collection of data sources, mining and investigation, user interest modelling, and security and protection considerations. The analysis done contains the challenging issues in the data-driven model and also in the Big Data revolution. The essential challenge is that a Big Data mining framework needs to consider complex relationships between samples, models, and data sources, along with their evolving changes with time and other possible factors. Also it faces the problem of autonomous information sources and the variety of the data collection environments, often result in data with complicated conditions, such as missing/uncertain values.

Kannan et al [6] shows Data analysis is one way of predicting of future stocks prices using data mining techniques. Five methods such as Typical Price (TP), Bollinger Bands, Relative Strength Index (RSI), CMI and Moving Average (MA) of analyzing stocks were combined to predict the day's closing price better than level of significance. Various global events and their issues are also investigated. These indicators support both numerically and graphically. This algorithm was able to predict if the following day's closing price would increase or decrease better than chance (50%) with a high level of significance. Furthermore, this shows that there is some validity to technical analysis of stocks. This is not to say that this algorithm would make anyone rich, but it may be useful for trading analysis. The algorithm performed well on half of the stocks and not so well on the other half of the stocks. Babulo et al [2] describes various Neural Network models for stock prediction. The prediction was done by, modular neural network, ARIMA-based neural networks, genetic algorithm, recurrent network, back propagation network, radial basis function, branch network, functional link artificial neural network, feed forward neural network, fuzzy neural network etc. Analysis of all these Neural Network models is performed in this paper, as well as the future work. The problem of stock index prediction is one of the most popular targets for various prediction methods in the area of finance and economics. In the past many Computational Intelligence techniques have been applied to this task including neural network, fuzzy and hybrid models or genetically developed prediction rules. Despite enormous previous efforts and a wide range of methods applied to this problem, efficient stock market prediction remains a difficult task mainly due to complex and varying in time dependencies between factors affecting the price. Stock market prediction without sentiment analysis using a web-traffic based classifier and user-level analysis provides the predictive power of online community traffic with regard to stock prices [7]. Using the largest dataset to date, spanning 8 years and almost the complete set of SP500 stocks, they trained a classifier using a set of features entirely extracted from web-traffic data of financial online communities. The classifier is shown to outperform the predictive power based on price time-series, and the performances were similar as the classifier built considering price and traffic features together. The best predictive performances are achieved when long-term and midterm web traffic levels were formed. Traffic-related features seem effective in predicting stock rises when certain levels of traffic are coupled with stock size.

PROPOSED WORK

Architectural Model of Proposed System

Proposed system architecture of stock market prediction is shown in fig.1, Different modules used in it are as follows -

Data Analyzer

The system that has to be developed has been classified in two types one is on the static database and another is on real time system. We handled the static database with oracle 10g as the background and for real time system we will use the yahoo finance dataset.

Classification

Central input to the prediction system is share details of current market and history of share market. The prediction system will predict proposed rate for shares. Prediction details and current details will supply to decision making system. Input to that decision making system is stored prediction details. Decision making system takes the decision that whichever we have to auction the share or we have to procure it. When the worth of share increases, the shareholder can auction shares to get the turnover and when value of shares decreases, the shareholder can procure the shares having a smaller amount market value and he can auction that shares when the cost of that share will increase,. To perform this activity shareholder will set threshold value in system. The alert is given to the share holder at whether he is in profit or in loss. If he got profit, those shares can be auction by him and if he is in loss, then he can procure shares of fewer market prices and then he can sale those shares after growth in their market value.

Sell and Purchase based on threshold

Threshold is used to store value in system by this value shareholder take decision to be sale or purchase. If he is in profit, he can sell those shares and if he is in loss, then he can purchase shares. Threshold is set for the company for both sell and purchase. While setting the threshold for sell we have consider the factors for the company like number of shares, target amount and the loss. Parameters like company name, number of shares, and amount are considered for setting the threshold for purchase.

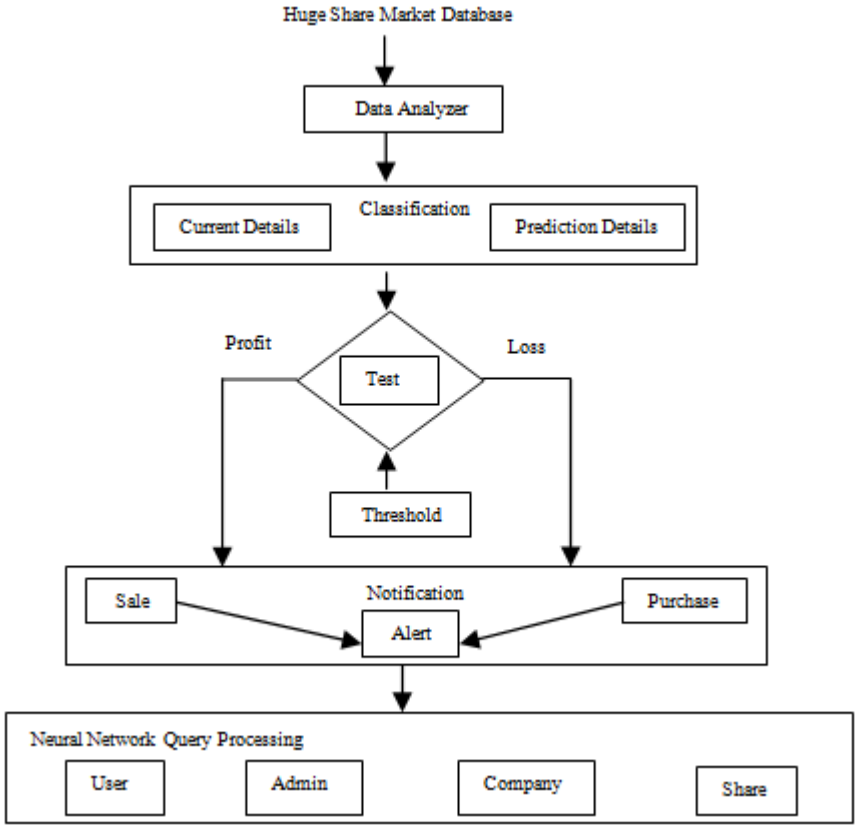

Fig. 1 Proposed System

Neural Network Query Processing

In a neural network, each neuron except the neurons on input layer receives and processes stimuli (inputs) from other neurons. The processed information is available at the output end of the neuron. The fig.2 illustrates the way in which each neuron in an MLP processes the information.

A neuron of the l th layer receives stimuli from the neurons of l-1 th layer, that is, $z_1^{l-1}, z_2^{l-1}, \ldots, z_{N_{t-1}}^{l-1}$. Each input is first multiplied by the corresponding weight parameter, and the resulting products are added to produce a weighted sum γ. This weighted sum is passed through a neuron activation function σ to produce the final output of the neuron. This output z_i^l can, in turn, become the stimulus for neurons in the next layer.

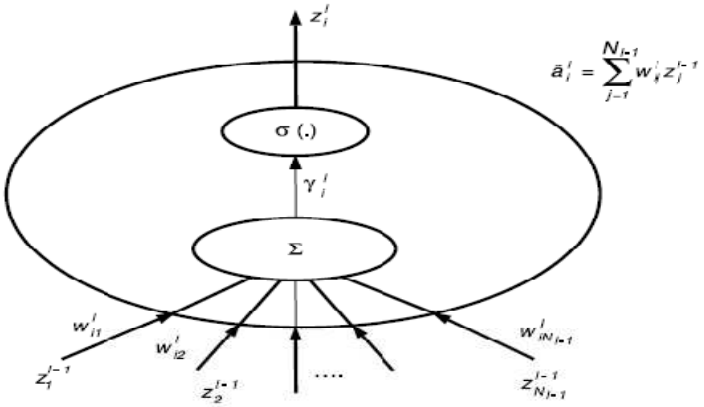

Fig.2 Information processing by i th neuron of l th layer [10]

Activation Function

The most commonly used hidden neuron activation function is the sigmoid function given by

$$\sigma(\gamma) = \frac{1}{1+e^{-\gamma}} \tag{1}$$

The processing of activation function is shown as follows

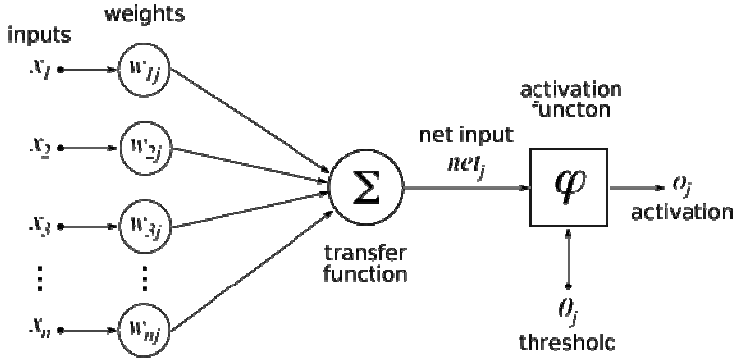

Fig. 3 Activation Function [10]

ALGORITHM USED

The algorithm for Prediction using Neural Network explains the multilayer perceptron (MLP) neural network. It is the feed forword neural network. Generally an input pattern x from the training set is given to MLP for getting the output y. In the MLP an input pattern, x, from the training set is sent. We have to compare y with target t, to get the error quantity. Use the error quantity to modify the weights, so next time y will be closer to t. Repeat with another x from the training set.

Procedure: Prediction using Neural Network [11]

1. Initialize the network, with weights between -1 and +1.
2. Present the first training pattern, and obtain the output.
3. Compare the network output with the target output.
4. Propagate the error backwards.

(a) Correct the output layer of weights using the following formula.

$$w_{ho} = w_{ho} + (\eta \delta_o o_h)$$

where w_{ho} is the weight connecting hidden unit h with output unit o, η is the learning rate, $o_h$is the output at hidden unit h. δ_o is given by the following.

$$\delta_o = o_o(1 - o_o)(t_o - o_o)$$

where o_o is the output at node o of the output layer, and $t - o$ is the target output for that node.

(b) Correct the input weights using the following formula.

$$w_{ih} = w_{ih} + (\eta \delta_h o_i)$$

where w_{ih} is the weight connecting node i of the input layer with node h of the hidden layer, o_i is the input at node i of the input layer, η is the learning rate.

δ_h is calculated as follows.

$$\delta_h = o_h(1 - o_h) \sum_o \delta_o w_{ho}$$

5. Calculate the error, by taking the average difference between the target and the output vector.

$$E = \frac{\sqrt{\sum_{n=1}^{p}(t_o - o_o)^2}}{p}$$

Where p is the number of units in the output layer.

6. Repeat from 2 for each pattern in the training set to complete one epoch.

7. Shuffle the training set randomly. This is important so as to prevent the network being influenced by the order of the data.

8. repeat from step 2 for a set number of epochs, or until the error ceases to change.

EXPERIMENTAL RESULTS

Experiment 1: Stock Market Analysis with DAX Dataset

We trained the network with the DAX (German stock index) data – for month march 2015 from 1st to 30th - to predict the value at 31.03.2015. As a strategy we take the sequences from 4 days to predict each 5th day. In the training set 5th day is the supervised value. For testing we'll use prepared data set in which the DAX data are given from the 27, 28, 29 and 30.03.2015 to predict the value at 31.03.2015.

Since the network is initialised with random weight values, the test results will differ from a calculation to calculation. After five tests it came out with the following prediction - results for 03.31.2015: 4084.61; 4081.28; 4073.08; 4075.22; 4087.42. The duration time was 3 sec.

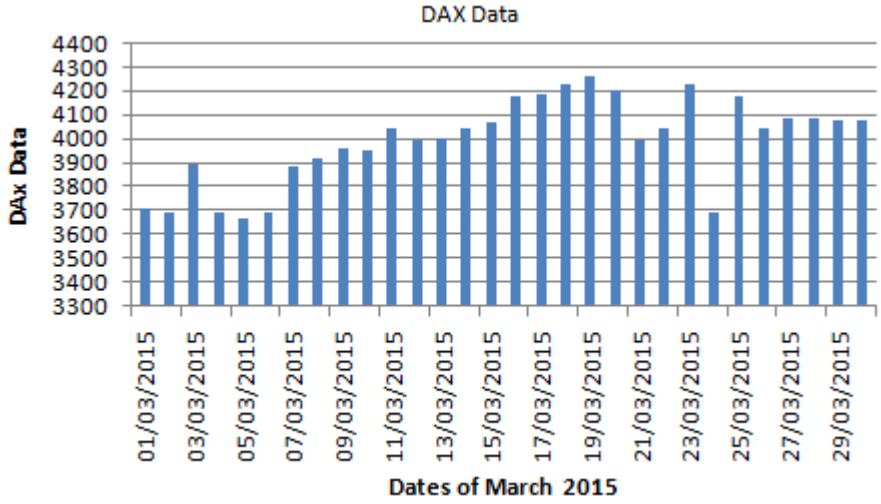

Fig. 4 DAX data for 1st -30th march 2015

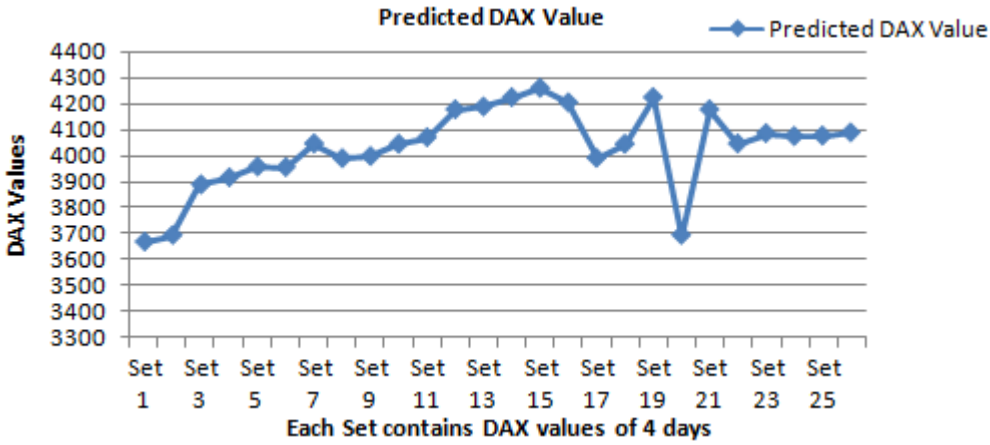

Fig. 5 Predicted DAX value for 31st March 2015

Experiment 2: Stock Market Analysis with Case Study

Case Study 1: Processing of neural network on 10 days

For the Case Study 1, we have used the 10 days testing dataset with 3 layers of neural network to predict the estimating closing price. Three iterations are used where the third iteration gives nearly the probable output. The detailed description is as given in following table -1. As shown in the table 1 the outcome from the first iteration is far away with the actual value. In second iteration we are closer to the output. And in the third iteration we are almost have the actual value.

Table-1 Comparison of Iterations on 10 days

No of Days	No of Layers	Iteration 1	Iteration 2	Iteration 3	Output
10	3	542.25	499.875	494.8	494

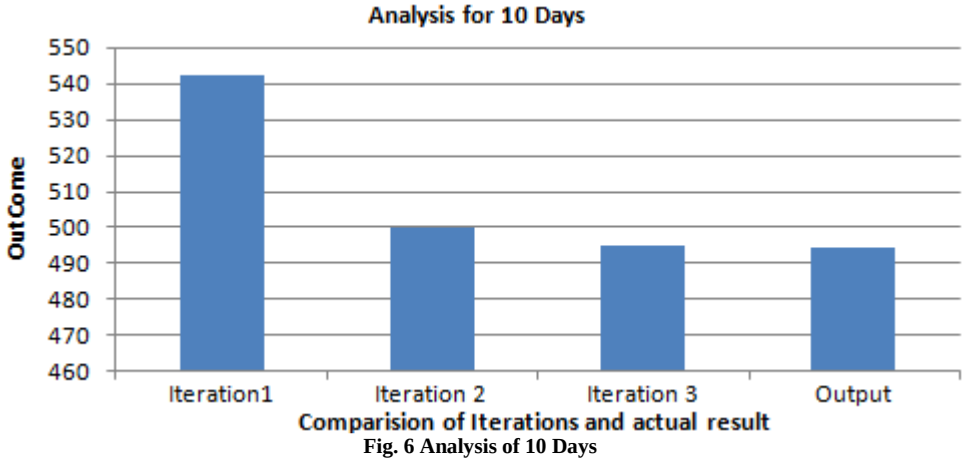

Fig. 6 Analysis of 10 Days

Case Study 2: Processing of neural network on 50 days

For the Case Study 2, we have used the 50 days testing dataset with same as it is predefined 3 layers of neural network to predict the estimating closing price. Three iterations are used where the third iteration gives nearly the probable output. The detailed description is as given in following table 2. As shown in the table -2 the outcome from the first iteration is 504.8 which is far away with the actual value 463. In second iteration we are closer to the output as we are getting 477.75 and in the third iteration we are getting 463.52 where actual value is 463.

Table-2 Comparison of Iterations on 50 days

No of Days	No of Layers	Iteration 1	Iteration 2	Iteration 3	Output
50	3	504.8	477.75	463.52	463

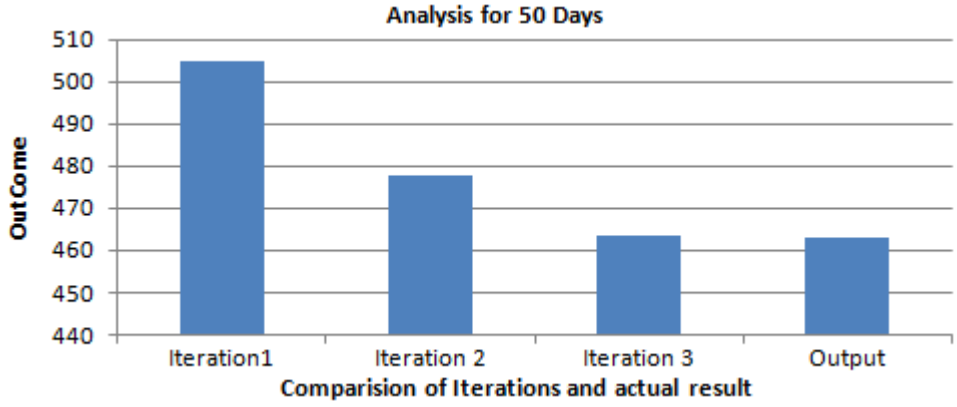

Fig. 7 Analysis of 50 Days

Case Study 3: Processing of neural network on 100 days

For the Case Study 3, we have used the 100 days testing dataset with same as it is predefined 3 layers of neural network to predict the estimating closing price. Three iterations are used where the third iteration gives nearly the probable output. The detailed description is as given in following table 3. As shown in the table-3 the outcome from the first iteration is 457.4 which is far away with the actual value 470. In second iteration we are closer to the output as we are getting 472.63 and in the third iteration we are getting 470.26 where actual value is 470.

Table-3 Comparison of Iterations on 100 days

No of Days	No of Layers	Iteration 1	Iteration 2	Iteration 3	Output
100	3	457.4	472.63	470.26	470

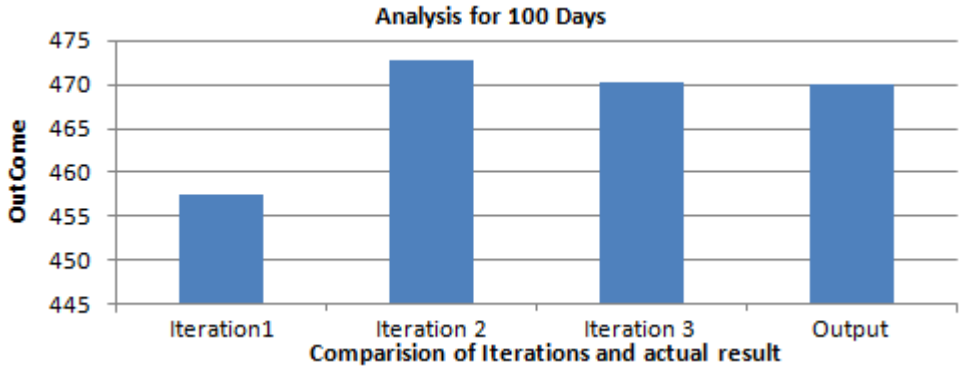

Fig. 8 Analysis of 100 Days

CONCLUSION

Selection of data and methods for data mining and neural network is an essential job in stock market prediction. With the combination of both data mining and neural network one can easily deal with unpredictable data like stock data. We can get more accurate prediction value with the help of neural network. Traffic-related features seem effective in predicting stock rises when certain levels of traffic are coupled with stock size. The best predictive performances are achieved when information about stock capitalization is coupled with long term and mid-term web traffic levels. The NN has ability to extract useful information from the dataset so it is widely play very important role in stock market prediction. These approaches are used to control and monitor the entire the stock market price behavior and fluctuation. As the Big Data is characterized by HACE we can relate it with Stock Market Prediction System. For that purpose we are using huge share market database, and by applying neural network back-propagation algorithm we can work with our system. Neural Network can be useful for handling complex data. Effective market prediction can help investors with trade advices or can be used as a component inside automatic trader agents. Sometimes prediction systems indirectly help traders by providing supportive information such as the future market direction.

Based on the knowledge extracted from the historical prices of such stocks the decision has been taken about buying or selling stocks. It will be conducted based on one of the data mining techniques; the back propagation. Up to this stage no one has applied strategy for big data they has simulated to the specific dataset. Our aim is to find most effective tool which will predict the market circumstances based on previous historical data and threshold that has been defined. As the neural network is more like real nervous system it is most likely applicable to parallel organization. It permits the solutions to problems where multiple constraints must be satisfied simultaneously. Neural network provides various benefits such as High Accuracy, Noise tolerance, independent from prior assumption, ease of maintenance, and problem tolerance. Finally, reconsidering the factors high accuracy and ease of maintenance we can improve the overall impact of our big data based stock market prediction system. In future, it can be further investigated with real time stock data from any market using neural network.

ANNs are used with reservation for their time-consuming training to get high precision. In the near future, we plan to keep on investigating the factors that affect system performance, and consider improving the speed and accuracy of the neural network. Additionally, future research should focus on the examinations of other types of networks that were rarely applied, such as Hopfiled's, Kohonen‟s, etc. Finally, almost all researchers emphasize the integration of NNs with other methods of artificial intelligence as one of the best solutions for improving the limitations. The Future scope of the work will concentrate towards trying different ANN architecture and identifying which suits well for prediction along with different training and learning functions.

REFERENCES

[1] Xindong Wu, Xingquan Zhu, Gong-Qing Wu and Wei Ding, Data Mining with Big Data, *IEEE Transactions on Knowledge and Data Engineering*, **2014**, 26 (1), 97-107.
[2] S Arun Joe Babulo, B Janaki and C Jeeva, Stock Market Indices Prediction with Various Neural Network Models, *International Journal of Computer Science and Mobile Applications*, **2014**, 2 (3), 42-46.
[3] Mizuno, Kosaka, Yajima and Komoda, Application of Neural Network to Technical Analysis of Stock Market Prediction, *Studies in Informatics and Control*, **1998**, 7,111-120.

[4] Jacek Mandziuk and Marcin Jaruszewicz, Neuro-Evolutionary Approach to Stock Market Prediction, *IEEE Conference on Neural Networks International Joint*, **2007**, 2515-2520.
[5] Kritsanatt Boonkiatpong and Sukree Sinthupinyo, Applying Multiple Neural Networks on Large Scale Data, *International Proceedings of Computer Science and Information Technology*, **2011**, 6, 189-193.
[6] K Senthamarai Kannan, P Sailapathi Sekar, M Mohamed Sathik and P Arumugam, Financial Stock Market Forecast using Data Mining Techniques, *International MultiConference of Engneers and Computer Scientists*, Hong Kong Place, **2010,** 1, 555-559.
[7] Pierpaolo Dondio, Stock Market Prediction without Sentiment Analysis: Using a Web-traffic based Classifier and User-level Analysis, *IEEE 46th Hawaii InternationalConference on System Sciences (HICSS)*, Wailea, HI, USA, **2013**, 3137-3146.
[8] Debashish Das and Mohammad Shorif Uddin, Data Mining and Neural Network Techniques in Stock Market Prediction: A Methodological Review, *International Journal of Artificial Intelligence & Applications*, **2013,** 4 (1), 117-127.
[9] Yahoo Finance, *Yahoo Finance Dataset*, Web. *http://finance.yahoo.com*, **21/2/2015**.
[10] Q Zhang and KC Gupta, Neural Networks for RF and Microwave Design, **2000**, Artech House, Inc. Norwood, MA, USA, 392.
[11] Leonardo Noriega, *Multilayer Perception Tutorial*, School of Computing, Staffordshire University, UK, **2005**.