

Stock Market Analysis Based on Artificial Neural Network with Big data

Miss.Minal P. Bharambe
Information Technology
PICT Pune.
Pune, India.
minal.bharambe@gmail.com

Prof. S.C.Dharmadhikari
Information Technology
PICT Pune.
Pune, India.
scdharmadhikari@pict.edu

Abstract— Nowadays financial market is becoming famous among the society due to lots of the online transactions. As such data contains information in many forms which is dynamic in nature, we can easily co- relate it with Big Data. One part of the financial market is that of Stock Market, as these both entities are dependent on each other. The point to focus is the unpredictable behaviour of the market. If we analyse stock market efficiently then we can make some effort for predicting respective market behaviour. Up till now multiple algorithms are applied but as the efficiency matters one has to go beyond them. One of the famous solution is with the help of neural network. Neural networks can easily work with complicated data. If we use neural network for Stock Market Analysis it will come up with the good outcomes for financial market.

Keywords- EMH, ANN, bag of words, noun phrases, Prediction System, Stock price classification.

I. INTRODUCTION (HEADING I)

Recently stock market has attracted the whole world towards itself, where forecasting leads to the successful market prediction. Predictability is directly proportional to investing and trading in stock market. The tool which monitors, predicts and regulators the market will help to take corrective measures[3]. Stock market deals with stocks having different industrial groups which cover ups the whole financial market. According to market status the actions are taken by the investors regarding sales and purchase. There are various factors available that affects the market status, like news releases on earnings and profits, and future estimated earnings, announcement of dividends, change of management etc. Initial research in financial and stock trading issues lead to the identification of some factors that are considered among experts to influence the price of a stock. At first, it is a reasonable thought that the behavior of an investor depends on the size of the owner company. Furthermore, we could identify the following influence factors:

Table 1 : Possible stock price influence factors

Factor	Description
Business Sector	Stock belonging Sector
Historical Behavior	Fluctuation of Stock over time
Stock Earnings	Percentile difference of the stock price value over period of time
Financial position of a company	The financial status of company
Volume	No of dealings taking place
Uncertainty	Availability of unpredictable factors
P/E factor	Price per annul earnings

In 1994, Haykin defines neural network as a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. The brain resembles in two respects: (1) Knowledge is acquired by the network through a learning process, and (2) Interneuron connection strengths known as synaptic weights are used to store the knowledge[4]. In last few years, application of neural network in stock market forecasting has been increased. With neural network market patterns can be extracted. A neural network is parallel, distributed processor having simple processing unit which is able to store experimental knowledge and can make use of it. Neural network can derive meaning from complicated and imprecise data remarkably. Extracting patterns and detecting trends are complex for both humans and computer techniques, but we can handle them effectively with neural network[4].

The paper is structured as follows: In Section II, we propose overview of existing system where the key challenges for existing systems are summarized. Section III focuses on proposed system. Section IV describes the techniques we have been using in proposed system and we conclude the paper in Section V.

II. LITERATURE REVIEW

The efficient market hypothesis (EMH) [Fama 1964] is the one that take into concern the Security. In EMH, it is assumed that the level of security matters with the information available and the people utilizing it. EMH has been broken into three forms by Fama's theory: weak, semi-strong, and strong. In weak EMH only historical information is embedded in the current price, in semi-strong EMH all historical and currently public information in the price is incorporated. In strong EMH the historical, public, and private information, such as insider information, is included. It is believed that the market reacts instantaneously to any given news and that it is impossible to consistently outperform the market. A different perspective on prediction is given by walk theory [Malkiel 1973]. In this theory, stock market prediction is believed to be impossible where prices are determined randomly and the market is outperforming infeasibly. In Random walk theory all public information is assumed to be available to everyone. However, even with such information, future prediction is ineffective with random walk theory. The price of a stock can be determined with inflation, joblessness, and return on equity (ROE), debt

levels, and individual price to earnings (PE) ratios. The technical analysis depends on historical and time-series data. These strategists believe that market timing is critical and opportunities can be found through the careful averaging of historical price and volume movements and comparing them against current prices. Technical analysis is considered to be more of an art form rather than a science and is subject to interpretation.

Kunwar Singh Vaisla and Dr. Ashutosh Kumar Bhatt [6] proved that neural network (NN) outperform statistical technique in forecasting stock market prices. They can work with sufficient data but reduces efficiency for complex data. Dase R.K. and Pawar D.D. [7] tried to sum up the application of Artificial Neural Network for predicting stock market. A neural network (NN) has the ability to extract useful information from large set of data. They have presented a review of literature about application of artificial neural network (ANN) for stock market predictions.

III. TECHNIQUES

A neural network is made up of a set of connected neurons, with the reception of input cells performing some transformations on it and generating outcome for further processing. ANN can provide us a powerful pattern classification and pattern recognition. Artificial neural network can work with non-linear, complex, imprecise and noisy data[5].

A. Basics of Neural Network :

A neuron is a real function of the input y_1, \dots, y_k , then the output will be generated with following function.

$$f(x_j) = f\left(\alpha_j + \sum_{i=1}^k w_{ij}y_i\right)$$

Where f is a function, and its graphical representation is as follows.

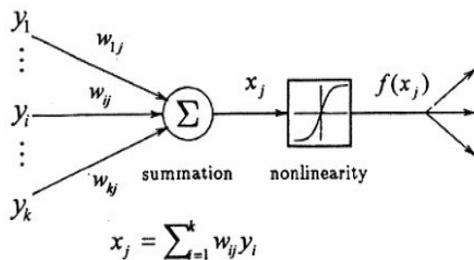


Fig 1 : Graphical Representation of single neuron

B. Feature of Neural Network :

1. Output patterns are generated by mapping input patterns.

2. Trained objects can be generated from untrained objects.
3. Featured with generalization.
4. Easily differentiate noise pattern in fault tolerant way.

C. Architecturs of Neural Network :

ANN is the data processing system based on neurons which works like the human brain.

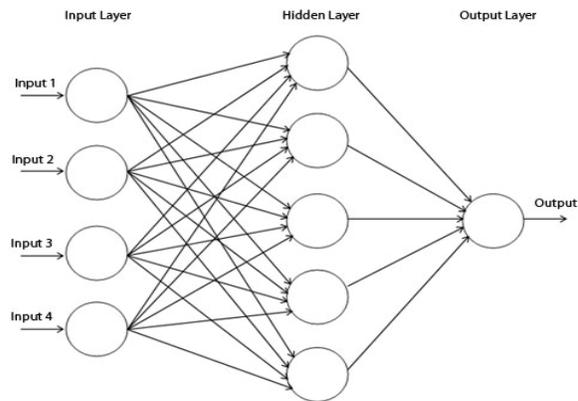


Fig 2: Neural Network

Feed Forard Network :

The Feed Forward networks are the networks having flow of information through hidden network in one direction from input layer to output layer. There is no concept of looping.

Recurrent Network:

The recurrent network are the network having atleast one feedback loop where output of each neuron is get fed to itself as an input.

D. Types of Neural Network :

There are 3 types of neural network :

1. Multilayer perceptron
2. Radial Basis Function Networks
3. kohnen Self Organizing Feature Maps

IV. PROPOSED SYSTEM

1.Data Collection

In order to train, validate and test the neural network, data is required and we collected five years historical data of various companies (IT and non-IT) from yahoo finance.

2.Data pre-processing

The data must be prepared such that it covers the range of inputs for which the network is going to be used. Since the performance and reliability of the output from

the neural network mainly depends on the quality of the data, therefore, the data must be pre-processed before it is fed to a neural network. First of all, we applied attribute relevance analysis on data so as to remove unwanted attributes from data and then the data was normalized in the range -1 to 1 using min-max normalization technique. Since the input is in the normalized form, the output we get is also in the normalized form and hence, it must be renormalised so as to have actual value. In order to train the network, we divided the data into three subsets [6] : Training Data Set : This data set was used to train the network. The gradient was computed and biases and the weights of the connections between the neurons were adjusted accordingly. Validation Data Set: This data was used to save the weights and biases at the minimum error and to avoid network over fitting data. Testing Data Set: This data set was used to test the performance of the network.

3. Neural Network Creation and Training[8]

In this step neural network was created with two layers one hidden layer and one output layer .Of course, input layer is essential. Artificial Neural Networks depend on the following parameters:

- Number of layers
- Number of neurons in input layer
- Number of neurons in hidden layer
- Momentum
- Learning rate
- Number of training iterations that are required to obtain the best result
- Transfer function used for hidden and output layer
- Training algorithm used
- Learning function used.

The network was created with some initial values of above mentioned network parameters. Then, these parameters were varied and the results were observed. The network was trained using back propagation algorithm with the aim to improve the network performance i.e. to reduce mean square error (mse). In this algorithm, the network is trained by repeatedly processing the training data set and comparing the network output with the actual output and reducing the error to the minimum possible. If the error between network output and the actual falls below the threshold value, then the training stops otherwise weights of the connections between various neurons are modified so as to reduce mse. The modifications are done in the opposite direction i.e. from output layer through each hidden layer down to the first hidden layer. Since the modifications in the weights of the

connections are done in the backward direction so the name given is back propagation. Transfer functions calculate layer's output from its net input. Hyperbolic tangent sigmoid transfer function and Log-sigmoid transfer function can be used for hidden layer and output layer. We have used Log-sigmoid transfer function for hidden layer as well as output layer.

4. Network Validation

After training the network, it was validated using validation data so as to improve the network performance.

5. Using the Network

After validating the network, it was tested using the test data set. The testing was performed on ten different companies (IT and non IT) and 100 tests were performed for each company.

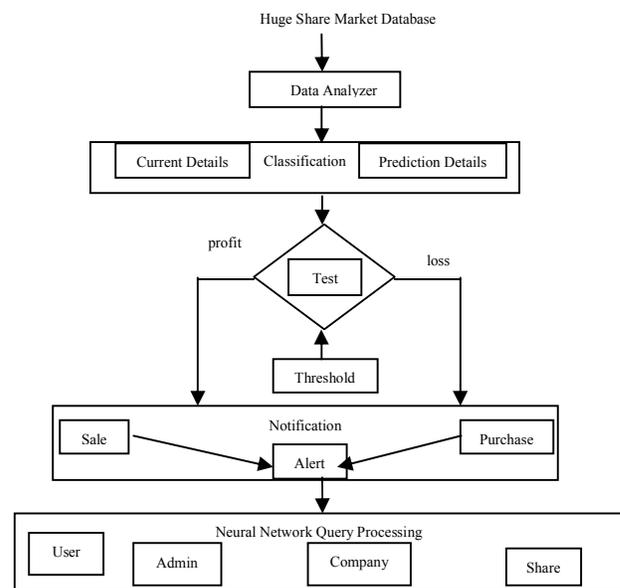


Fig 3. Block diagram of stock market prediction

A Stock Market Prediction System can be co-related with A Big Data Processing framework. At Tier I , Data Analyzer module does low level data accessing, processing and computing. At Tier II, Classification module does predication based on predefined threshold. At Tier III Neural network Back propagation algorithm is applied[1].

A huge share market database is provided as an input for to the whole prediction system. Analysis is done by the Data analyzer module. Two categories as current details and prediction details are generated based on the performed analysis. Prediction details and current details act as a backbone of decision making system. Input to that decision making system are stored prediction details, current details and threshold value. With the predefined threshold value, the processing about either profit or loss can be conducted. The alert is generated according to the

status whether in profit or loss. If the share holder is in profit, shares can be auctioned and if he is in loss, then it can be procured. Now as the market value grows shares can be sold. The pictorial representation of stock market prediction is shown in above diagram.

In financial analysis, accuracy matters a lot so accuracy in Stock Market's increasing or decreasing nature is beneficial to market analysis[9]. Multilayered classified set provides fine granularity[16]. To provide security we can find out outliers which identifies spammers[10].

There can be single source knowledge and multisource knowledge. By improving single source knowledge we can provide efficiency in handling big data, dynamic data[11]. We can use multisource knowledge[13],[14]. We can further perform analysis with whatever data we have[12]. For database system by scaling up large amount of data we can provide intelligent learning[15].

V. EXPERIMENTAL RESULTS

A. Registration Form

We have created two registration forms, one for user and another for company. In user registration, the user will do registration for selling and purchasing the shares. In company registration, company will register and will fill all details about the shares and will sell their shares.

User Registration:-

In check function we are checking the value given by the user at the time of registration. The fields of the form while registration are:

1. User name
2. Password
3. Re-enter password
4. Address
5. Phone number
6. Email Address
7. Total amount that user have in bank account to purchase shares.

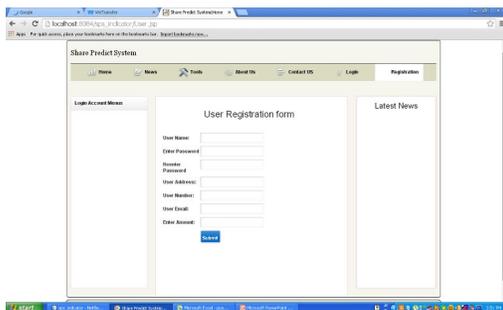


Fig 4: User Registration Form

Company Registration:-

In check function we are checking the value given by the user at the time of registration. The fields of the form while registration are:

1. Company name.
2. Company address.
3. Phone number.
4. Email Address.
5. Total amount that company have.
6. The value of share of that company.

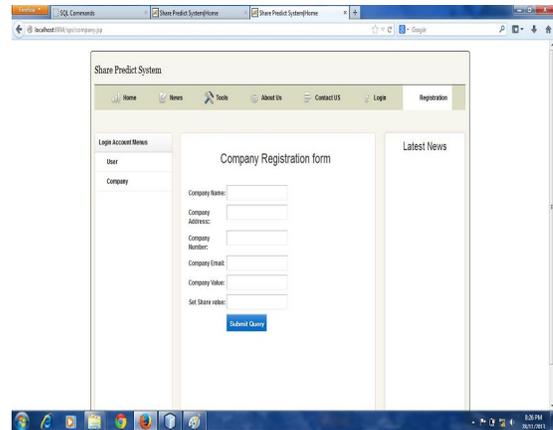


Fig 5: Company Registration Form

B. Login Form

We have created a login form to login for user and admin. In this check function will be used to check that whether user is logging in by correct name or not. The fields are Username and Password.

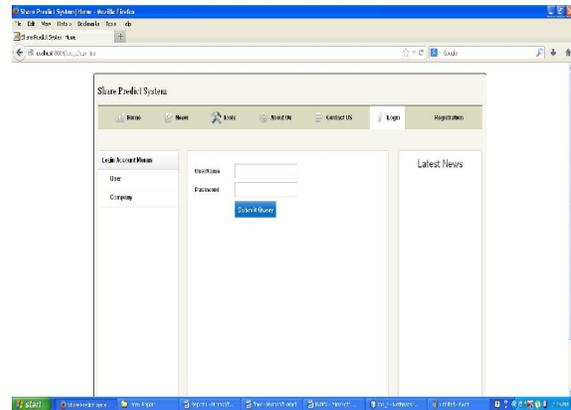


Fig 6: Login Form

C. Set Automation for Sale Form

In this form we have done the automation for selling the share. The shares will automatically get sold by the user when admin will do changes in the values of the shares. If the value of share updated by the admin match to the value set by the user, then only shares will get sell.

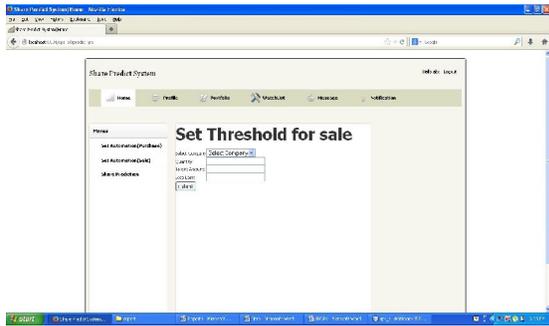


Fig 7: Set Automation for Sale Form

VI. RESULT ANALYSIS

The stock market prediction first build a decision-tree classifier using features set completely extracted from web traffic data of financial online communities. Our conducting tests showed how a classifier built bearing in mind both traffic and price-related features outperforms a market price-only classifier and the market benchmark, while a traffic only classifier outperforms all the other classifiers in predicting price increases, in the company of a gain of 4.2% on average and up to 25% compared to the market standard. Traffic-related features seem effective into predicting stock rises when assured levels of traffic are tied with stock size. The best predictive performances are achieved when information in relation to stock capitalization is tied with long-term and mid-term web traffic levels.

VII. CONCLUSION AND FUTURE WORK

We can relate big data with Stock Market Prediction System with the huge share market database, and by applying neural network back-propagation algorithm. We can easily handle complex data with neural network. It can be used as automatic trader agents for effective market prediction. The future market direction can be predicted by the system.

The decision has been taken based on the historical analysis about buying or selling stocks with the back propagation. Up till now actual implementation is not conducted for big data they has simulated to the specific dataset. Our aim is to find most effective tool for predicting the market. It is applicable to parallel organization. It permits the solutions to problems where multiple constraints must be satisfied simultaneously. Finally, reconsidering the factors high accuracy and ease of maintenance we can achieve and improve it with neural network which will impact of our big data based stock market prediction system.

As for the future work, there is still big room for testing and improving the proposed model by evaluating

the model over the whole companies listed in the stock market. Also, the evaluation of a larger collection of learning techniques such as neural networks, genetic algorithms, and association rules can represent a rich area for future investigation. Finally, reconsidering the factors affecting the behavior of the stock markets, such as trading volume, news and financial reports which might impact stock price can be another rich field for future studying.

REFERENCES

- [1] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding. "Data Mining with Big Data," IEEE Transactions on Knowledge and Data Engineering, vol. 26, 2014.
- [2] S. Arun Joe Babulo et al, "Stock Market Indices Prediction with Various Neural Network Models" International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 3, pg. 42-46, March-2014.
- [3] Manna Majumder, MD Anwar Hussian, "Forecasting Of Indian Stock Market Index Using Artificial Neural Network", National Stock Exchange of India Limited, 2009.
- [4] Phichhang Ou, " Prediction of Stock Market Index Movement by Ten Data Mining Techniques ", Modern Applied Science, Vol. 3, No.12., December, 2009.
- [5] Amrender Kumar, "Artificial Neural Network for Data Mining", I.A.S.R.I., Library Avenue, Pusa, New Delhi -110012
- [6] Kunwar Singh Vaisla & Dr. Ashutosh Kumar Bhatt, "An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting", (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 06, pp. 2104-2109, 2010.
- [7] Dase R.K. & Pawar D.D, "Application of Artificial Neural Network for stock market predictions: A review of literature", International Journal of Machine Intelligence, ISSN: 0975-2927, Volume 2, Issue 2, pp. 14-17, 2010.
- [8] http://en.wikipedia.org/wiki/Artificial_neural_network
- [9] J. Bollen, H. Mao, and X. Zeng, "Twitter Mood Predicts the Stock Market," J. Computational Science, vol. 2, no. 1, pp. 1-8, 2011.
- [10] S. Borgatti, A. Mehra, D. Brass, and G. Labianca, "Network Analysis in the Social Sciences," Science, vol. 323, pp. 892-895, 2009
- [11] E.Y. Chang, H. Bai, and K. Zhu, "Parallel Algorithms for Mining Large-Scale Rich-Media Data," Proc. 17th ACM Int'l Conf. Multimedia, (MM '09), pp. 917-918, 2009.
- [12] R. Chen, K. Sivakumar, and H. Kargupta, "Collective Mining of Bayesian Networks from Distributed Heterogeneous Data," Knowledge and Information Systems, vol. 6, no. 2, pp. 164-187, 2004.
- [13] C.T. Chu, S.K. Kim, Y.A. Lin, Y. Yu, G.R. Bradski, A.Y. Ng, and K. Olukotun, "Map-Reduce for Machine Learning on Multicore," Proc. 20th Ann. Conf. Neural Information Processing Systems (NIPS '06), pp. 281-288, 2006.
- [14] X. Wu, C. Zhang, and S. Zhang, "Database Classification for Multi-Database Mining," Information Systems, vol. 30, no. 1, pp. 71-88, 2005.
- [15] X. Wu, "Building Intelligent Learning Database Systems," AI Magazine, vol. 21, no. 3, pp. 61-67, 2000.
- [16] M. Ye, X. Wu, X. Hu, and D. Hu, "Anonymizing Classification Data Using Rough Set Theory," Knowledge-Based Systems, vol. 43, pp. 82-94, 2013.