

# An analysis on Stock Market Prediction using Data Mining Techniques

S.Prasanna

School of Information Technology

VIT University

Vellore,India

[sprasanna@vit.ac.in](mailto:sprasanna@vit.ac.in)

Dr.D.Ezhilmaran

School of Advanced Sciences

VIT University

Vellore,India

[ezhilmaran.d@vit.ac.in](mailto:ezhilmaran.d@vit.ac.in)

**Abstract**—Stock market data analysis needs the help of artificial intelligence and data mining techniques. The volatility of stock prices depends on gains or losses of certain companies. Many people consider stock market prediction as gambling. However it is possible to generate constructive patterns by the analysis of stock prices. Data mining techniques can be applied on past and present financial data to generate patterns and decision making algorithms. In this paper we have discussed several attempts made by researches for stock price prediction. These works show that data mining techniques can be applied for evaluation of past stock prices and acquire valuable information by estimating suitable financial indicators.

**Keywords** : Stock price, inflation, patterns, AI techniques.

## I. INTRODUCTION

Stock market is the bone of fast emerging economies such as India. Major of capital infusion for companies across the country was made possible only thru shares sold to people. So our country growth is tightly bounded with the performance of our stock market. Almost all the developing nations rely on their stock market for further strengthening of their economy. Any way in developing economies less than 10% of people are engaging themselves with stock market investment fearing the volatile nature of stock market. Many people felt that buying and selling of shares is an act of gambling which is a wrong notion. Majority of financial researchers agree that stock market is the only place where investor are getting consistent inflation beaten returns for so many years. Considering the fact of lack of knowledge and awareness across the people stock market prediction techniques plays a very crucial role in bringing more people into market as well as to retain the existing investors. Also the prediction techniques must be treated like astrology or gambling. The applied techniques must yield consistent accurate results with certain level of accuracy always in order change the mindset of passive investors.

By examining the literature stock market prediction techniques can be grouped into four types. 1) Technical analysis approach, 2) Fundamental analysis approach, 3) Time series prediction and 4) Machine learning algorithmic methods. Technical approach is otherwise referred as chartist approach. They are generating predictions based on the historical price values of selected stocks. Fundamental analysis approach is finding out the true value of a stock and compares it with the current trading levels and recommends buying of stock which is traded lesser than its true value. In case of Time series prediction linear flow prediction models are generated and historic patterns are traced. Those linear models can be classified into two types based on number of variables involved in their estimation. The one variable model is referred as univariate estimation model and multiple numbers of variables are considered in multivariate regression models. Finally Machine learning methods consider set of samples in order to generate linear and non linear patterns. The aim is to find out an expression that can generate the data.

### *Role of data mining in stock market*

Many researchers attempts to predict stock prices by applying statistical and charting approaches. But those methods lacks behind heavily due to human biased decisions on stock market based on day to day mind set of human behavior. By applying data mining in a suitable way hidden patterns can be uncovered which was not possible by traditional approaches. Also by applying business intelligence future price prediction with increased accuracy levels are possible with data mining techniques. The huge amount of data generated by stock markets forced the researchers to apply data mining to make investment decisions. The following challenges of stock market can be effectively addressed by mining techniques.

- 1) Future stock price prediction
- 2) To generate effective patterns of past data for further analysis.
- 3) To optimally utilize the capital of share holders.
- 4) For the growth of country economy.
- 5) To bring more investors to stock market who are lacking in analysis.
- 6) To stabilize the market.
- 7) To increase transparency in the market.
- 8) To check corruptive practices.
- 9) To bring more lazy and tech savvy investors into market.

## II. RELATED WORKS

Prediction of stock price variation is a very challenging task and the price movement behaves more like a random walk and time varying. In recent times, researchers have used various types of AI techniques to make trading decisions. Here, we present a brief review of some of the significant researches. A Sheta [1] has used Takagi-Sugeno (TS) technique to develop fuzzy models for two nonlinear processes. They were the software effort estimation for a NASA software projects and the prediction of the next week S&P 500 for stock market. The development of the TS fuzzy model can be achieved in two steps 1) the determination of the membership functions in the rule antecedents using the model input data; 2) the estimation of the consequence parameters. They used least-square estimation to estimate these parameters. The results were promising.

Ching Long Su *et al.* [2] have developed a self-organized, five-layer neuro-fuzzy model to model the dynamics of stock market by using technical indicators. The model effectiveness in prediction and forecasting was validated by a set of data containing four indicators: the stochastic oscillator (%K and %D), volume adjusted moving average (VAMA) and ease of movement (EMV) from TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index). A modified moving average method can be proposed to predict the input set for the neuro-fuzzy model in forecasting stock price. Simulation results have shown that the model was effective in prediction and accurate in forecasting. The input error from the prediction of the modified moving average method attenuated significantly by the neuro-fuzzy model to yield better forecasting results.

M.H. FazelZarandiet *al.* [3] have developed a type-2 fuzzy rule based expert system for stock price analysis. Interval type-2 fuzzy logic system permitted to model rule uncertainties and every membership value of an element was interval itself. The proposed type-2 fuzzy model applied the technical and fundamental indexes as the input variables. The model can be tested on stock price prediction of an automotive manufactory in Asia. Through the intensive experimental tests, the model had successfully forecasted the price variation for stocks from different sectors. The results were very encouraging and implemented in a real-time trading system for stock price prediction during the trading period.

Robert K. Lai *et al.* [4] have established a financial time series-forecasting model by evolving and clustering fuzzy decision tree for stocks in Taiwan Stock Exchange Corporation (TSEC). The forecasting model integrated a data clustering technique, a fuzzy decision tree (FDT), and genetic algorithms (GA) to construct a decision-making system based on historical data and technical indexes. The set of historical data can be divided into k sub-clusters by adopting K-means algorithm. GA was then applied to evolve the number of fuzzy terms for each input index in FDT so the forecasting accuracy of the model can be further improved. A different forecasting model was generated for each sub-cluster. In other words, the number of fuzzy terms in each sub-cluster was different. Hit rate applied as a performance measure and the proposed GAFDT model had the best performance when compared with other approaches on various stocks in TSEC.

Shyi-Ming Chen and Yu-Chuan Chang [5] have presented a method for multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. First, the proposed method constructed training samples based on the variation rates of the training data set and then used the training samples to construct fuzzy rules by making use of the fuzzy C-means clustering algorithm, where each fuzzy rule corresponded to a given cluster. Then, they determined the weight of each fuzzy rule with respect to the input observations and used such weights to determine the predicted output, based on the multiple fuzzy rules interpolation scheme. They applied the proposed method to the temperature prediction problem and the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) data. The experimental results have shown that the proposed method can be produced better forecasting results than several existing methods.

WengLuen Ho *et al.* [6] have proposed an interday financial trading system with a predictive model empowered by a novel brain-inspired evolving Mamdani-Takagi-Sugeno Neural-Fuzzy Inference System (eMTSFIS). The eMTSFIS predictive model possessed synaptic mechanisms and information processing capabilities of the human hippocampus, resulted in a more robust and adaptive forecasting model as compared to existing econometric and neural-fuzzy techniques. The trading strategy of the proposed system was based on the moving-averages-convergence/divergence (MACD) principle to generate buy-sell trading signals. By introducing forecasting capabilities to the computation of the MACD trend signals, the lagging nature of the

MACD trading rule can be addressed. Experimental results based on the S&P500 Index confirmed that eMTSFIS was able to provide highly accurate predictions and the resultant system was able to identify timely trading opportunities while avoiding unnecessary trading transactions. These attributes enabled the eMTSFIS-based trading system to yield higher multiplicative returns for an investor.

EsmaeilHadavandiet *al.* [7] have presented an integrated approach based on genetic fuzzy systems (GFS) and artificial neural networks (ANN) for constructing a stock price forecasting expert system. At first, they used stepwise regression analysis (SRA) to determine factors which have most influence on stock prices. At the next stage, they divided their raw data into k clusters by means of self-organizing map (SOM) neural networks. Finally, all clusters were fed into independent GFS models with the ability of rule base extraction and data base tuning. They evaluated capability of the proposed approach by applying it on stock price data gathered from IT and Airlines sectors, and compared the outcomes with previous stock price forecasting methods using mean absolute percentage error (MAPE). Results have shown that the proposed approach outperformed all previous methods, so it can be considered as a suitable tool for stock price forecasting problems.

Kelvin Simet *al.* [8] have proposed a method, 3D subspace clustering to generate rules to pick potential undervalued stocks. 3D subspace clustering is effective in handling high dimensional financial data and is adaptive to new data. The obtained results are not influenced by human's biases and emotions, and are easily interpretable. The conducted extensive experimentation in the stock market over a period of 28 years (from 1980 to 2007), and found that using rules generated by 3D subspace clustering algorithms, CATSeeker and MIC, results in 60% more profits than using Graham's rules alone.

Anil Rajput [9] have proposed a rule based classification method which generates rules with preprocessed data. It focus on each class separately and increases the probability of exact classification. PRISM algorithm is applied to generate rules by examining the training data and identifying suitable rules that covers majority of class. Also it measures the success of each rules. It serves as a suitable tool for generating rules for buy and sell.

Yin Song[10] proposed a method of analyzing coupled behavior of stocks in the market. CHMM(Coupled hidden markov model has been considered for coupled relations in hidden state space. He proposed a new graph based framework to find out abnormalities that might occur in coupled behaviors. He also apply the results to real world stock data and proved that their method beats the basic CHMM markov model in both technical and business measures.

### III. CONCLUSION

As per the discussed works above basically two types of prediction methods are implemented by several researches to generate useful extracts. They are fundamental approaches and technical indicator based approaches. Many researchers adopted technical indicator approaches only. Limited work is done with fundamental approaches which give plenty of opportunity for further research. Since the stock data is highly volatile and unpredictable it needs the intelligence of human for effective prediction. Also it needs rigorous training of old data for analysis. This temperament of stock data makes data mining and AI techniques as suitable once. Back propagation algorithm for training and suitable AI technique applied on some fundamental approaches may render promising results. survey

### References

- [1] A Sheta, "Software Effort Estimation and Stock Market Prediction Using Takagi-Sugeno Fuzzy Models", In Proceedings of the IEEE International Conference on Fuzzy Systems, pp.171-178, Vancouver, BC, 2006.
- [2] Ching Long Su, ChuenJyh Chen and Shih Ming Yang, "A self-organized neuro-fuzzy system for stock market dynamics modeling and forecasting", WSEAS Transactions on Information Science and Applications, Vol7.8, No.9, September 2010 .
- [3] M.H. FazelZarandi, B. Rezaee, I.B. Turksen and E. Neshat, "A type-2 fuzzy rule-based expert system model for stock price analysis", Expert Systems with Applications, Vol.36, No.1, pp. 139-154, January 2009.
- [4] Robert K. Lai, Chin-Yuan Fan, Wei-Hsiu Huang and Pei-Chann Chang, "Evolving and clustering fuzzy decision tree for financial time series data forecasting", An International Journal of Expert Systems with Applications, Vol.36, No.2, pp. 3761-3773, March 2009.
- [5] Shyi-Ming Chen and Yu-Chuan Chang, "Multi-Variable Fuzzy Forecasting Based On Fuzzy Clustering and Fuzzy Rule Interpolation Techniques", Information Sciences, Vol.180, No.24, pp. 4772-4783, 2010.
- [6] WengLuen Ho, Whye Loon Tung and Chai Quek, "Brain-Inspired Evolving Neuro-Fuzzy System for Financial Forecasting and Trading of the S&P500 Index", Lecture Notes in Computer Science, Vol.6230, pp.601-607, 2010
- [7] EsmaeilHadavandi, Hassan Shavandi and ArashGhanbari, "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting", Knowledge-Based Systems, Vol.23, No.8, pp. 800-808, December 2010.
- [8] Kelvin Sim, Vivekanand Gopalkrishnan, Clifton Phua and Gao Cong, "3D Subspace Clustering for Value Investing", IEEE Intelligent Systems, Vol. PP, No.99, pp. 1, 2012.
- [9] Anil Rajput , S.P. Saxena , Ramesh Prasad Aharwal and Ritu Soni, "Rule based Classification of BSE Stock Data with Data Mining", International Journal of Information Sciences and Application, Volume 4, Number 1 (2012), pp. 1-9.
- [10] Xianggao Cai, Su Hu, Xiaola Lin, "Feature Extraction Using Restricted Boltzmann Machine for Stock Price Prediction", 978-1-4673-0089-6/12/\$26.00 ©2012 IEEE.
- [11] Yin Song and Longbing Cao, "Graph-based Coupled Behavior Analysis: A Case Study on Detecting Collaborative Manipulations in Stock Markets", WCCI 2012 IEEE World Congress on Computational Intelligence June, 10-15, 2012 - Brisbane, Australia.