

A Review on Prediction of Stock Market using Various Methods in the Field of Data Mining

T. Shobana and A. Umamakeswari*

School of Computing, SASTRA University, Tirumalaisamudram, Thanjavur – 613401, Tamilnadu, India; shobanathiyagarajan490@gmail.com, aum@cse.sastra.edu

Abstract

Objectives: In the current emerging competitive market, predicting the stock returns as well as the company's financial status in advance will provide more benefits for the investors in order to invest confidently. Stock prediction can be done by using the current and previous data available on the market. **Methods:** The performance metrics that need to be attained in case of stock prediction are accuracy, scalability and less time consumption. There are many researches done so far in order to predict the stock market to achieve the defined metrics. Many models have been available in the field of data mining for predicting the stock market such as if-then-else rules, Artificial Neural Network (ANN), Fuzzy systems, Bayesian algorithm and so on. **Findings:** In this paper, the various methods available and used for predicting the stock market are discussed. This survey helps to know which technique is the best to use in the field of predicting stock market in the area of mining. **Applications:** Forecasting and predicting the trends of market is the most important applications of stock market. It also uncovers the future market behavior which always helps the investors to understand when and what stocks can be purchased for the growth of their investment. For this reasons, many of the researches have been done so far in the area of analyzing the stock market using data mining.

Keywords: Data Mining, Methodologies, Prediction, Review, Stock Market

1. Introduction

Predicting the stock market is an important field of identifying the future value of the price for better exchange of finance between the companies having the shares. In the field of forecasting the financial status, stock market prediction has become a major one. Once the financial status of a company has been accurately predicted, the investors will be confidently investing their money. Even though many attempts have been made to predict the financial market's performance, none of the techniques have pro-

vided a constant result since it is a difficult task because of noisy and time varying data and also every technique used has its own disadvantages. The stock market prediction is divided into 2 groups¹. 1. Trend Prediction model: It is used for prediction by obtaining a relationship between various technical variables and the stock price movement. 2. Time series forecasting: It forecasts by analyzing the historical return of stock and it is used for predicting the future stock price return. Many popular models and methods are used for predicting the stock which includes Support Vector Machine (SVM), Decision Tree (DT),

*Author for correspondence

ANN, Genetic Algorithm (GA), and Bayesian Network (BN). Recently, it has been found that a successful selection of features related to predicting stock movements will help providing more accurate result. In that context, many feature selection algorithm has also been used in order to choose exact features which has more impact on the movement of stock returns. Sequential forward feature selection² is one of the techniques used. We can also apply any machine language algorithms like SVM to discover the features³. Many researchers have predicted the stock trend by hybridizing the feature selection algo-

rithm with machine learning algorithm³ which provided the prediction accuracy to certain extent. Out of all available models used for predicting stock market, some of the main models which provide advantages in all perspectives like time, accuracy are discussed in this paper.

2. Stock Prediction Methodologies

Various models used for the prediction of stock market and the methods are explained in the Table 1.

Table 1. Various methods used for predicting stock market

Sl.no.	References	Dataset type	Method Used	Functionalities
1	In ⁴	Time series data	4 stage ANN	The best Performance only with window size 3 as it increases linearly with No. of data
2	In ⁵	15 years of data-Time series data	3 stage Artificial Chemical Reaction neural Network and MLP	Good performance with high non linearity capturing capability using MLP trained data
3	In ⁷	Istanbul Stock Exchange data	MLP with Feed forward Network	More appropriate GFF Network
4	In ⁹	Time series data	Type-2 Fuzzy System	Outperforms the various traditional methods as well as type-1 fuzzy approach
5	In ¹⁰	Time series data	Particle Swarm Optimization	Lesser computational complexity

Table 1 Continued

6	In ¹¹	Time Series data	PSO with portfolio selection algorithm	Provides Minimum Risk with Maximum Profit
7	In ¹³	Trend Prediction data, Bombay Stock Exchange data	Hidden Markov Model with decision tree	The best prediction compared to MLP
8	In ¹⁴	Financial Time series data	Fuzzy DT with genetic algorithm	The best performance in predicting and with almost 80% the best system performance
9	In ¹⁵	India cements stock price index	GA with neuro- fuzzy systems	Promising for extracting trading rules

2.1 ANN and its Variants

In⁴, ANN model was used to predict the closing price using economic and financial theory which combines technical analysis, fundamental analysis and time series analysis addressing the percentage of accurate predictions. They adopted four prediction stages in their work such as:

- (1) Problem domain understanding
- (2) Sample collection
- (3) Input preprocessing and
- (4) Modeling and prediction.

They configured a model of varying window sizes (number of points of monthly series that is the neural network's input) for obtaining various performances. They achieved the best performance with window size 3. When they increased the window size, they found it very difficult to get the desired result. So, they suggested combining this method with any other traditional techniques to get better accuracy and performance. In⁵ predicted the stock market indices using an Artificial Chemical

Reaction Neural Network (ACRNN) which inculcated Artificial Chemical Reaction Optimization (ACRO) to train the Multi-Layer Perceptrons (MLP) model for the purpose of predicting the time series data. They modeled and tested this method using 15 years of historical data with the following stages:

- (1) Reactant encoding
- (2) Enthalpy of a reactant
- (3) ACRNN training using ACRO algorithm.

This model showed a high nonlinearity capturing capability of MLP model since unbiased estimation cannot be possible with linear techniques. They suggested hybridization of this technique with robust evolutionary search algorithms which provides more accuracy on predicting the stock indices. In⁶ used a fused model of ANN with a Discrete Wavelet Transform to identify the market price value. The model provided a better accuracy over the models which did not incorporate wavelet transform since it provided an efficient way to predict the pattern for getting quick and accurate result.

2.2 MLP with Feed Forward Networks

In⁷ used a Generalized Feed Forward (GFF) networks with a MLP to predict the stock market of Turkish's values and Stock Exchange of Istanbul. Index value, exchange rate and interest rate of previous day along with five dummy variables were used as input for the system. 90% of data set was used as training data and the rest 10% were used for testing purpose. Training and testing were done by GFF as well as MLP. They educated that among all the ANN models, GFF model was found to be more relevant and appropriate for prediction of stock market. In⁸, they discussed a system for forecasting the long and short term price value of the company using MLP which provided the results with minimum number of epochs and took very less time for training the data.

2.3 Fuzzy Systems

In⁹ introduced a three phase stock market model for predicting the future market price movement. In the initial stage, an analysis was carried out with multiple regression technique to declare all the inputs which provide a strong affiliation with the result. Creation of prediction model was done by implementing the type-2 evolutionary based Fuzzy clustering model in the second stage. In the third stage, type-2 fuzzy neural network was modeled for future stock prediction. To avoid the problems associated with cluster optimization, differential evolution based algorithms were used. They concluded that this method outperforms the various traditional technique as well as fuzzy type-1 approach.

2.4 Particle Swarm Optimization Technique

In¹⁰ developed a model for predicting various stock indices using Particle Swarm Optimization (PSO) technique. Mean Square Error (MSE) was minimized by the use of PSO and obtained a long term prediction performance compared to the performance obtained by MLP model so far. It confirmed that it was computationally efficient and was giving wiser accuracy compared to the entire standard MLP based methods. They also analyzed that PSO is one of the promising forecasting models for stock market

prediction because of its lesser computational complexity. In¹¹ suggested many variations of PSO for the prediction of complex and intricate problems. Some of the modifications that was suggested in the basic PSO are:

- a) Dynamic multi-swarm PSO,
- b) Multi-swarm & multi-the best PSO
- c) Dynamic multiple swarm PSO with inter-regional search
- d) Multi-swarm PSO technique, e) Master-slave swarm evolutionary algorithm
- f) Velocity Clamping
- g) PSO with passive congregation method where first three were said to be the external modification technique and the other two were called as internal modification technique.

They discussed that PSO with portfolio selection algorithm can also be used to predict stocks to have minimum risk with maximum profit. In¹² used PSO to predict the stock volatility with less computational complexity and high accuracy.

2.5 Hidden Markov Model with DT Based Hybrid Technique

In¹³, they used a DT rough set based hybrid system to predict the future trends on stock market. They used DT for selecting and extracting relevant features. Once it was extracted, rough set classifier was used to predict the next day's trend using the features selected. They used Hidden Markov Model for the evaluation of future prediction. They validated the result using Bombay Stock Exchange (BSE-SENSEX) and acquired 92% of accuracy in predicting the future trends. They concluded that the DT rough set based hybrid model with Hidden Markov Model (HMM) is the best for predicting the stock market trend prediction.

2.6 Genetic Algorithm

In¹⁴ developed a model for financial time series forecasting. Instead of having DT alone for every series of data, they integrated a data clustering technique along with

fuzzy DT and GA in which the entire historical data set was divided into k- clusters using K-means algorithm. Each cluster was checked by different forecasting model each clusters had different fuzzy terms for reducing the time taken for prediction. They concluded that performance of the system experienced was 82% and provided the best performance compared to other approaches. In¹⁵ proposed a data mining approach using GA to remove the problems with knowledge acquisition which are intrinsic for the construction and the maintenance of the stock market's rule-based applications. They modeled the system with basic GA four stages such as initialization, selection, Crossover, and mutation. The main aim of their GA model is to have the refined hypothesis and evaluation technique based on the fitness function. They used India Cements Stock Price Index (ICSPI) to train the data and to test the data. With the help of GA, they found six alternative rules which provided high return over a certain period of time. They concluded that GA is a promising method for extracting trading rules. They also suggested to use GA to train the weights when they used with neuro-fuzzy system for getting more accuracy¹⁶.

3. Conclusion

By the use of this various available techniques, it is possible to create a new technique to predict the future trends in stock market. It is possible to utilize any of the discussed techniques and develop a hybrid system for the prediction of financial status of a company accurately. But it is important to design the system accordingly by which the accuracy and performance can be increased with less computational complexity. This paper is a complete survey of methods and models used by many researchers in the application of data mining techniques for stock market prediction. This survey will also help in choosing the best algorithm for time series and trend prediction method based on their prediction metrics.

4. References

- Zhang X, Hu Y, Xie K, Wang S, Ngai EWT, Liu M. A causal feature selection algorithm for stock prediction modeling. *Neurocomputing*. 2014 Oct; 142:48–59.
- Altinbas H, Biskin OT. Selecting macroeconomic influencers on stock markets by using feature selection algorithms. *Procedia Economics and Finance*. 2015 Nov; 30:22–9.
- Ni LP, Ni ZW, Gao YZ. Stock trend prediction based on fractal feature selection and support vector machine. *Expert Systems with Applications*. 2011 May; 38(5):5569–76.
- de Oliveira FA, Nobre CN, Zárate LE. Applying ANNs to prediction of stock price and improvement of the directional prediction index – case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*. 2013 Dec; 40(18):7596–606.
- Nayak SC, Misra BB, Behera HS. ACRO of neural networks for efficient prediction of stock market indices. *Ain Shams Engineering Journals*. 2015 Oct 3.
- Chandar SK, Sumathi M, Sivanandam SN. Prediction of stock market price using hybrid of wavelet transform and ANN. *Indian Journal of Science and Technology*. 2016 Feb; 9(8):1–5.
- Egeli B, Ozturan M, Badur B. Stock market prediction using ANNs. *Proceedings of the 3rd Hawaii International Conference on Business*. Hawaii; 2003.
- Kamley, Jaloree S, Thakur RS. Long term and short term investment strategy for predicting the performance of BSE using MLP model. *Indian Journal of Science and Technology*. 2015 Sep; 8(22):1–9.
- Enke D, Grauer M, Mehdiyev N. Stock market prediction with multiple regression, fuzzy type-2 clustering and neural networks. *Procedia Computer Science*. 2011 Dec; 6:201–6.
- Prediction of S&P 500 and DJIA Stock indices using PSO Technique [Internet]. 2008 [cited 2008 Jun 1]. Available from: <http://ieeexplore.ieee.org/document/4630960/>.
- Jamous RA. Modifications of PSO techniques and its application on stock market: a survey. *International Journal of Advanced Computer Science and Applications*. 2015 Mar; 6(3):1–10.
- Mohapatra P, Das S, Patra TK, Anirudh M. Stock volatility forecasting using swarm optimized hybrid network. *International Journal of Emerging Trends and Technology in Computer Science (IJETTCS)*. 2013 May–Jun; 2(3):78–85.
- Tiwari S, Pandit R, Richhariya V. Predicting future trends in stock market by DT rough-set based hybrid system with HHMM. *International Journal of Electronics and Computer Science Engineering*. 2012; 1(3):1578–87.

14. Lai RK, Fan CY, Huang WH, Chang PC. Evolving and clustering fuzzy DT for financial time series data forecasting. *Expert Systems with Applications*. 2009 Mar; 36(2):3761–73.
15. Naik RL, Ramesh D, Manjula B, Govardhan A. Prediction of stock market index using genetic algorithm. *computer Engineering and Intelligent Systems*. 2012; 3(7):162–71.
16. Agrawal S, Jindal M, Pillai GN. Momentum analysis based stock market prediction using Adaptive Neuro-Fuzzy Inference System (ANFIS). *Proceedings of International Multi conference of Engineers and Computer Scientists*. 2010 Mar; 1:1–6.