A Survey of Stock Price Prediction & Estimation Using Data Mining Techniques

S. Prasanna
School of Information Technology VIT University Vellore, Tamil Nadu, India.

Dr. D. Ezhilmaran
School of Advanced Sciences VIT University Vellore, Tamil Nadu, India.

Abstract
The application of AI techniques for stock price prediction leads to voluminous growth of wealth of investors with the advent of technology. Several prediction and estimations are coming up for almost all sectors of the market. Particularly any kind of stock price prediction is not at all possible without excessive data manipulation which can be done effectively only thru data mining. The systematic statistical manipulation of data can be done effectively only by applying suitable business intelligence and AI techniques. Particularly Indian stock market is running in more complex scenario and needs excessive data mining. Several works regarding stock price prediction is done for international as well as Indian market with the advent of data mining techniques. In this review we attempt to narrate some important works done for stock price prediction using data mining techniques.

Keywords: Inflation, Pattern, AI techniques, BI, clustering

Introduction
In fast growing economies like India stock market movement greatly influences all citizens irrespective of the participation of people. The foundation knowledge of stock market is essential for all citizens for country’s development. The participation of more and more people add voluminous data to stock market. The participation of foreign investors/traders makes stock market subjective to several local, national, international issues. In fast growing economies such as India the growth is tightly tied up with the stock market. In India the stock prices are also influenced by WPI(whole sale price index) and CPI( consumer price index). Many times the market is highly volatile and the share holders are affected by sudden unexpected profit/losses. These profits and losses incurred in stock market is also affecting the people who are not participating in it.
This fact clearly indicates that the country’s growth is purely dependent on the stock market and its price movement. Many traders and investors involved in stock market is not doing proper analysis. Many people thought that this is an act of gambling or astrology or luck and bad luck. Several data mining techniques have broken the myth and generated predictive patterns which are promising growth of portfolios. All variants of prediction techniques can be classified in to four significant types

- Technical charting approach
- Variable model
- Fundamental analysis approach
- Machine learning algorithm based methods
- Time series prediction

Technical charting approach
The charting approach is basically categorized as technical approach. It deals with voluminous historical data of stock prices of the concerned stocks.

Variable Model
This approach is working on examining the selected parameters analysis to predict the future price of stocks.

Fundamental analysis approach
This approach is alternately referred as true or real price prediction which primarily focus on fundamentals of the company instead of price movement. It give weightage to true value prediction instead of current price movement.

Machine learning algorithms
This method attempts to predict the movement of stock prices based on training given with the past value movements.

Time Series analysis
This method considers the time as important parameter to generate series of stock price movement.

Significance of Data mining for Stock price predictions
Several works are done to predict stock price movements based on charting historical values. Many of them failed to deliver promising results since it cannot accommodate the actual movement. The data mining techniques can address such issues
The hidden patterns can be discovered by applying data mining pattern matching techniques. It serves as the preliminary stage for gathering business information to estimate future needs. Many of the conventional methods failed to deliver in this issue. Data mining techniques effectively addresses the challenges in which the traditional methods failed to deliver promising results. Some of them are

- Stock price prediction
- Pattern generation with historical data
- Effective utilization of business capital
The economic development
Investor friendly analysis

Related Works
Prediction of stock valuation difference is very baffling task. Various AI techniques are applied on trading decisions of most of the traders and investors. At this point we have made a review of some outstanding research done in this area. Haoming Huang [1] created a generic membership function name as Irregular shaped membership function(ISMF). It is applied with hierarchical coevolutionary genetic algorithm which is used to automatically derive each input feature in it. This system overcomes buy and hold with real world financial data. Trading signals are generated by price percentage oscillator as the main technical indicator.

Chang Liu and Hafiz Malik [2] proposed a work with respect to return and volatility. They are used to sort out low performing sector in the market. Apart from predicting investment decisions buy/sell his work also have a strategy for maximizing investment gains. The resultant trading decisions leads to higher profitability of investors.

Xiaoxiao Guo, [3] focuses on supply chain management under long cycle. Their approach combining inventory with information searched on web to conduct a demand prediction to optimize inventory. Back propogation NN is used to train the predictive model. A traditional inventory policy is configured by calculating a normal distribution of demand with history data. Then the inventory cost is estimated. The results promise the inventory policy lies on demand prediction and is superior in terms of overall cost of inventory.

Yunus Yetisi [4] use ANNs with the set of input parameters of the market to predict NASDAQ’s stock value. In this work they use feed forward networks are used. Regression technique applied to confirm the network performance. The generated plots render the outputs for validation, training and test cases.

Anthony J. T. Lee, Ming-Chih Lin, Rung-Tai Kao [5] proposed HRK(Hierarchical agglomerative and recursive K-Means clustering) which predicts short term stock movements with reference to financial reports. This method contains three phases. First the financial reports are converted to feature vector and HRK method is applied to divide them into clusters. In second step K-Means clustering method partitions each cluster into sub clusters so that each sub cluster belongs to same class. Third for each sub cluster the centroid is chosen as representative feature vector. These vectors are used for stock price movement prediction.

Haoming Huang [6] applies the hierarchical coevolutionary fuzzy system named HiCEFS as predictive model. It employs a prudent trading strategy with respect to price percentage oscillator (PPO). To construct a precise predictive model Irregular shaped membership function(ISMF) is employed and hierarchical coevolutionary genetic algorithm (HCGA) adopted to automize ISMF for each input in HiCEFS.

Dongsong Zhang and Lina Zhou [7] address the need for atomized approaches for effective utilization of financial data of corporates and individuals in planning and decision making. It uncovers the hidden patterns and future trend prediction. The benefits are profit margin increase, cheaper cost and sound marketplace response. It also analyses several data mining techniques proposed for financial data analysis. Lay-Ki Soon [8] compares the numeric, symbolic data of stock market with respect to similarity. For normalized dataset the empirical study concludes that numeric stock data is more consistent when compared with symbolic stock data. It explores the possibility of combining numeric and symbolic data with stock market data on trend modelling. By incorporating temporal semantics of the dataset the growth of casual relationships between stocks with respect to time the results are interesting.

Depei Bao [9] utilizes high level representation of time series which is insensitive to noise then intuitive to humans. Professional investors gathers knowledge from technical indicators which generally depicts the aggregation of market on particular time period. By joining high level representation and probabilistic model the uncertainty and randomness is reduced further levels. Thus the prediction precision is improved.

Kelvin Sim [10] proposed a system based on graham’s rules. They propose 3D subspace clustering for rule generation to choose potential undervalued stocks. This is an effective method in dealing with multi dimensional financial data. It is also adaptive to new data. The results are not influenced by human emotions and biases. The results promises 60% more profits than simply applying graham rules alone.

Prasanna [11] proposed a method to estimate true value of stock price using hybrid Mcniven approach. The predictions are generated for three categories. Fair valued, Over valued and under valued stocks. This helps the investors to select good stocks which are fair valued thus increasing the profit for investors. This method yields better results than 3D subspace clustering method in selected cases.

Conclusion
The above survey is shortly explaining the works done for stock price prediction. As per the observations the data mining techniques are not utilized heavily due to varying nature of different country markets. Also extensive works are done for technical charting approach only which will be only useful for traders. Many other techniques such as fundamentals analysis the works are very scarce. The challenge with stock price prediction is it cannot be determined only with respect to stock historical data. It is also influenced by several other factors such as market sentiments, government policy decisions news and etc. So it needs data from different sources to be integrated for data mining. In that case the data pre-processing will become more complex since data originates from different sources and due to heterogeneous nature it is difficult to arrange it for further steps of data mining. So there are so many issues and challenges remaining in applying data mining techniques for stock price prediction. However year by year some of the key issues are addressed by researchers from different countries. There is not doubt data mining techniques if applied effectively by combining data from heterogeneous sources for stock market will be one and only method to increase the profits of investors/traders.
References


[2] Chang Liu and Hafiz Malik[2], ” A New Investment Strategy Based on Data Mining and Neural Networks”, 2014 International Joint Conference on Neural Networks (IJCNN) July 6-11, 2014, Beijing, China

[3] Xiaoxiao Guo, Chang Liu, Wei Xu, ” A prediction-based inventory optimization using data mining models”, 2014 Seventh International Joint Conference on Computational Sciences and Optimization


[5] Anthony J. T. Lee, Ming-Chih Lin, Rung-Tai Kao, ” An Effective Clustering Approach To Stock Market Prediction” Department of Information Management, National Taiwan University, Taipei, Taiwan


