



## APPLICATIONS OF DATA MINING IN STOCK MARKET

SAVINDERJIT KAUR<sup>1\*</sup> AND VEENU MANGAT<sup>2</sup>

<sup>1</sup>Department of Information Technology, UIET, PU, Chandigarh, India.

<sup>2</sup>Assistant Professor, Department of Information Technology, UIET, PU, Chandigarh, India.

\*Corresponding Author: Email - savinderjit.kaur@gmail.com

Received: December 12, 2011; Accepted: January 15, 2012

**Abstract-** Data mining is being actively applied to stock market since 1980s. The various aspects of stock market to which data mining has been applied include predicting stock indices, predicting stock prices, portfolio management, portfolio risk management, trend detection, designing recommender systems etc. The various algorithms and methods which have been used for the same include neural networks, association rules, decision trees, fuzzy logic, PSO (Particle Swarm Optimization), DE (Differential Evolution), GA (Genetic Algorithm) etc.

**Key words** - stock price, stock index, portfolio management, trends in stock market, recommender systems, abnormal stock returns, oil prices.

**Citation:** Savinderjit Kaur and Veenu Mangat (2012) Applications of Data Mining in Stock Market. Journal of Information and Operations Management ISSN: 0976-7754 & E-ISSN: 0976-7762, Volume 3, Issue 1, pp-66-68.

**Copyright:** Copyright©2012 Savinderjit Kaur and Veenu Mangat . This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Introduction

Stock market is a place where buying and selling of stocks/shares takes place. This trading can be done manually on trading floor or online. When an investor buys stocks of a certain company he becomes a part owner of that company according to the number of shares held by him. To make maximum profit, right investment should be made at the right time. According to the time for which stocks are held, stock trading can be classified as position trading, swing trading and day trading. In day trading, stocks are held for less than a day; they are even bought and sold within minutes. In swing trading, stocks are held for about one to five days, so decisions are based more on emotions than on fundamental values. In position trading stocks are held for about five days to six months, so decisions are made based on fundamental changes in values of stocks. Stock Market prediction is an attractive field for research due to its commercial applications and the attractive benefits it offers.

An important hypothesis related to stock market which has been debated and researched time and again is EMH (Efficient Market Hypothesis). According to EMH, the stock market immediately reflects all of the information available publicly. But in reality, the stock market is not that efficient, so the prediction of stock market

is possible. Generally there are three main schools of thoughts regarding such predictions. The first school believes that no investor can achieve above average trading advantage based on historical and present information. The second view is that of fundamental analysis in which macroeconomic variables are taken into consideration for prediction. Technical analysts have presented the third view on market price prediction. They believe that there are recurring patterns in the market behavior which can be identified and predicted [1]. Stock market follows stochastic, non parametric and non linear behavior. There are three kinds of common stock forecasting models, namely, Multi-layer Perception (MP), Radial Basis Function (RBF), and Support Vector Machine (SVM) [2]. The different facets of the stock market (like stock prices, stock indices, portfolio management etc.) which have been researched extensively using different algorithms are being discussed below.

### Aspects

#### 1. Stock Price Prediction

Stock price prediction deals with predicting the price of the stocks. This prediction can be long term or short term. Stock prediction has been done using fundamental and technical analysis. Fundamental analysis is based on various macroeconomic factors like interest

rate, exchange rate, dividends yield etc. Technical analysis is based on the belief that history repeats itself. It predicts current price based on opening price, closing price, highest price, lowest price etc. Various other technical indicators can also be included. In addition to these, various social, psychological and political factors also affect the stock prices. The traditional methods used for stock prediction included moving average line, point and line methods. But these methods have low success rate. The methods generally used for prediction can be classified as fundamental (ARIMA, GARCH) and those based on artificial intelligence (ANN, SOM, PSO, rough set). ARIMA and GARCH are linear based models so they can not accurately predict the stock market which is nonlinear. Artificial intelligence methods are non linear in nature, capable of learning and are more accurate. ARIMA (Autoregressive Integrated Moving Average) has been outperformed by ANN (Artificial Neural Networks) [3]. The various versions of ANN which have been implemented include amnesic neural networks, BPNN (Back Propagation Neural Network) and feed forward neural networks. Amnesic neural networks are based on human cognitive ability of forgetting with time, so the data from the past has less effect on the output value [4]. BPNN provides the capability of reducing errors in predicted price by providing feedback [5]. Feed forward neural networks have least success rate at price prediction as there is no option of providing feedbacks or making any other adjustments. Different algorithms can be used to train neural networks (like PSO, DE) to further improve the prediction success [6]. Fuzzy logic has been combined with decision trees and rough set theory for successful price prediction. SVM (Support Vector Machine) surpasses all other methods in prediction till now. Different hybrid methods have been developed by combining one with other. These hybrid methods are good at prediction as they combine the best features of all the involved algorithms. Sentiment analysis and text mining have also been used for price prediction.

## 2. Stock Index Prediction

Stock index is the statistical average of a particular stock exchange. Indexes are composed of the stocks that are either part of the same exchange, same industry or same companies. NYSE Composite index, S&P 500 Composite Stock Price Index, DAX, Nifty are a few stock indexes to be named. Stock indexes give an overall perspective of the economic health of a certain industry or stock exchange, so they act as benchmarks. The indexes can be classified as local and global depending on the geography of the corporations included. The stock indexes can be calculated on bases of stock price, size of company and number of shares. These indexes are called price weighted, market value weighted and market share weighted indexes respectively. E.g. Dow Jones Industrial Average is price weighted average index. Forecasting stock indexes has always been a topic of interest for most stock investors, dealers and brokers [7]. Nevertheless, finding out the best time to buy or sell has always been very difficult because of the numerous factors to be taken into consideration that may affect the stock market. Forecasting requires a clear understanding of the timing of lead-lag relations among different variables, the statistical significance of these lead-lag relations and learning which variables are more important ones to watch as signals for predicting the market moves. Better forecasting is the key element for better financial decision making in the increasing financial market volatility and

internationalized capital flows. The various methods which have been used for stock index prediction include ANN, association rules, mobile data mining, adaptive bacterial foraging optimization (ABFO) and BFO (Bacterial foraging optimization), genetic neural networks, hybrid system of Markov chain and fuzzy stochastic system etc..

## 3. Portfolio Management

In broader terms, portfolio refers to the collection of investments held by an investor. A Portfolio is built by buying stock, bonds, annuities, mutual funds, or other investments belonging to different companies. Diversification is used to maximize the returns. Return refers to the profit or loss earned on investments, including income and change in value. The portfolio return is the weighted return of the stocks [8]. Portfolio management deals with making decisions about investment mix and policy, matching investments to objectives, asset allocation for individuals and institutions, and balancing risk against performance. The efficient frontier of portfolios is the set of portfolios that offers the greatest returns for each level risk (or equivalently, portfolios with the lowest risk for a given level of returns) [8]. Investment management consists of three phases: strategic asset allocation, tactical asset allocation, and stock picking [9]. Strategic asset allocation is a long-term allocation strategy that implies choosing the market to invest in, selecting the kind of assets to buy and the distribution of these in the portfolio in accordance with the investor's objectives. Tactical asset allocation consists of regularly adjusting the portfolio, in a systematic or discretionary way, to take advantage of the short-term opportunities [10]. Stock Picking consists of selecting the best stocks to be incorporated in the portfolio [10]. It is the most time consuming phase which impacts the return of the portfolio. Various market multiples such as price/earnings (P/E), book/market (B/M), sales/price (S/P), and cash flow/price (CF/P) influence the return of portfolio. Basu (1977) reported that low P/E stocks have abnormally high returns even after controlling for the market risk. Fama and French (1992) showed that B/M is more strongly positively related to stock returns than earnings/price (E/P) [11]. Barbee (1989) stated a positive relationship between the sales/price ratio and future stock returns. Barbee, Mukherji and Raines (1996) documented that S/P has a stronger positive relation with stock returns than B/M. Lakonishok, Shleifer and Vishny (1994) indicated larger return spreads for higher CF/P stocks than for stocks with high B/M or E/P. A plausible reason for the different explanatory powers that the various market multiples have for stock returns is the varying influences of accounting methods on the different accounting numbers used to compute the multiples [11]. The various methods used for portfolio management include clustering, factor models and equal correlation coefficient models, fuzzy logic, Cox-Ingersoll-Ross (CIR) framework, neural networks, hybrid system of ARX (Autoregressive exogenous), grey system and rough set theories etc.

## 4. Recommender Systems

Recommender systems help the brokers and investors alike to make decisions regarding buying, selling and holding of stocks and management of portfolios. Typical applications of recommender systems are product finders, stock market predictors, shopping guides, etc [12]. The different kinds of information and knowledge used in these systems include:

- long term data such as world and local economic figures, interest

rates, financial indexes, interrelationships of public corporations, etc.;

- Short term data such as current prices of stocks, gold, commodities, etc.;
- Choices of analytical techniques in decision-making process;
- Corporate policies of the financial company;
- Personal experience of the investment experts [13].

Traditionally, an investment expert would make his decision by choosing some analytical models to apply on the long term data to yield some preliminary decision guidelines. These guidelines are then combined with his investment experience, indications of some selected short term data and corporate policies of the investment company to generate final decisions [13].

Stock brokers handle most of the buying and selling. Recommender systems can help investors to hire discount brokers, who charge less, instead of full service brokers, who charge more for providing advice regarding trading. A critical obstacle to the use of a recommender system for stock investment is the mismatch between the needs of the investment analysts and their ability to communicate their needs to the computer. It is very difficult for a normal trader to apply data mining techniques to the data on his own due to the complexity involved in the whole data mining process. The recommender systems store data in different formats according to the requirements, keep record of the investments made, money left, transactional details and make suggestions regarding buying and selling of stocks. The different recommender systems are based on association rule mining, business intelligence (BI), temporal data mining, multi-objective particle swarm optimization (MOPSO), rule induction etc.

### 5. Detecting trends in stock market

The stock sector information in conjunction with time series subsequences can be used for mining core patterns within the sectors of stock market data. A core pattern is a representative group of stocks that shows coherent behavior specific to their sector. [14]. Multiple core patterns may exist in a sector at the same time [14]. The trends can be detected by applying mining to the information obtained from legal insider trading. This data is available because officers of companies are required by law to submit to the Securities Exchange Commission a record of the sales and purchases of their companies' stock [15]. Temporal patterns have been studied using hybrid system of NN and GA [16]. The various trends which have been detected include:

*i) Predicting abnormal stock returns:* Insider traders usually make abnormal returns because of the insider information available. Outsiders who can get access to the insider information can also make increased profits. The ability of outsiders, using insider trading information, to predict abnormal returns can be increased by focusing on data such as the size of the company and the number of months in the future that are predictive for stock prices [15]. NN has been used for this type of prediction [15].

*ii) Oil prices and stock sector returns:* In view of the crucial role of oil in the global economy and its spectacular price fluctuations in the recent years, it is naturally opportune to ask questions about the impact of the price of oil on stock prices [17]. Research has shown that spillover is usually unidirectional from oil markets to stock markets in Europe, but bidirectional in the United States. This interrelationship between oil prices and stock market returns has been studied using vector autoregressive-generalized autoregres-

sive conditional heteroscedasticity (VAR-GARCH) model [17].

### Conclusion

Data mining has been used in stock market to make predictions regarding trends and prices to gain maximum profits. A lot of research has been carried out on its various aspects. There is still room for improving accuracy of these prediction methods by developing new hybrid methods and by improving the existing algorithms. More research can be done in detecting trends in the stock market like studying the abnormal stock returns, trends preceding and following executive stock options awards, book-to-market effect, bubble diagnosis, inter industry patterns etc.

### References

- [1] Majhi R., Panda G., Majhi B., Sahoo G. (2009) "Expert Systems with Applications, 10097-10104.
- [2] Zang X., Yu J. (2011) *IEEE*.
- [3] Wijaya Y.B., Kom S., Napitupulu T.A. (2010) *Second International Conference on Advances in Computing, Control, and Telecommunication Technologies* 176-179.
- [4] Ye Q., Liang B., Li Y. (2005) *IEEE*.
- [5] Jiawei Han and Micheline Kamber (2006) *Data Mining Concepts and Techniques*, 2<sup>nd</sup> ed.
- [6] Abdul-Salam M.E., Abdul-Kader H.M., Abdel-Wahed W.F. (2010) *IEEE*.
- [7] Wang Y., Cheng S., Hsu M.H. (2010) *Applied Soft Computing*, 613-617.
- [8] Nanda S.R., Mahanty B., Tiwari M.K. (2010) *Expert Systems with Applications*, 8793-8798.
- [9] Amenc N. and Sourd L.V. (2003) *Wiley Finance*.
- [10] Casanova I.J. (2010) *IEEE*.
- [11] Barbee Jr. W.C., Jeong J., Mukherji S. (2008) *Global Finance Journal*, 1-10.
- [12] Paranjape Voditel P., Deshpande U. (2011) *Second International Conference on Emerging Applications of Information Technology, published by IEEE Computer Society*.
- [13] Moon Y.S., Yau C. and Yip W.M. (1989) *IEEE*.
- [14] Wu J., Denton A., Elariss O., Xu D. (2009) *IEEE International Conference on Data Mining Workshops* 558-563.
- [15] Safer A.M. (2001) "Predicting Abnormal Stock Returns with a Nonparametric Nonlinear Method", *IEEE*.
- [16] Kim H., Shin K. (2007) *Applied Soft Computing*.
- [17] Arouri M.H., Jouini J., Nguyen D.K. (2011) *Journal of International Money and Finance*, 1386-1405.