

# APPLICATION OF DATA MINING IN EQUITY MARKET

Akshay Chelawat<sup>1</sup>, Rishi Agarwal<sup>2</sup>, Ashutosh Das<sup>3</sup>, Nalini N<sup>4</sup>

School of Computer Science and Engineering, VIT University, Vellore, TN-632014, INDIA

## Abstract:

The current paper is focused on investigating the approaches of forecasting of financial time series paying the main attention to the tools, which are taken from the methodology of Data Mining which are neural networks and clustering. The paper has the qualitative design and it is based on the studying of the secondary sources.

**Keywords**—Data mining, Securities, Wiener process, Heston's Model, Brownian Motion, Clustering , Neural Networks

## I. INTRODUCTION

THERE are many approaches to solving the problem of forecasting of financial time series, which are using methods from the most diverse sections of mathematics. All these methods can be divided into two classes: the first category includes those methods that use some prior information about the characteristics of the time series, while the second category includes those methods that do not take into account any additional information except input data.

The first type, for example, includes methods that are based on the solution of stochastic differential equations. In these methods, it is argued that the dynamics of the price change of the asset is subject to certain patterns, which are described by the stochastic differential equations. This assumption rather accurately simulates the reality. Therefore these methods have wide application in the field of prediction of financial time series.

The second type includes probabilistic and statistical analysis methods, neural networks and clustering methods. The advantage of these methods lies in the high sensitivity to changing the characteristics of the time series (retraining) and potentially higher prediction accuracy since these models allow us to identify hidden data dependencies, which cannot be established a priori.

In the literature, the first mention of stochastic differential equations appeared in the early twentieth century in works that were related to the description of Brownian motion. The emergence of stochastic differential equations required the creation of a proper calculus, which was called later the theory of stochastic integrals. The theory received the wide application in physics, biology, and chemistry, as well as in probability theory and financial mathematics. The most famous and often used example of stochastic differential equations is the equation with the so-called white noise, which is considered as a derivative of the Wiener process.

### A. Importance of Equity Market or Share Market

The Equity Market or Stock Market is one of the most efficient source to raise money for an individual and organization, if invested wisely, along with debt markets or bond markets. Issuing shares enables companies to be traded publicly, and raise extra money for growth of the company by selling shares of ownership of the company in a share market. Share market allows the investor or the shareholders to sell the shares in a very convenient way. It is an attractive feature of investing in stocks or bonds, compared to investing in other less liquid sources like property and other unmovable assets.

From the past results and experiences we can say that the value of stocks and other assets plays a vital

role in the dynamics of economic activity, and may influence or can be an indicator of mood of market and society. An economy where the stock market points are on the rise (in India SENSEX and NIFTY) are considered to be an up-and-coming economy. The behavior of the stock market usually indicates the economic strength of the country.

Rise and fall of share prices also affect the prices of everyday products and their amount of consumptions. Therefore, central banks keep an eye on the behavior of the stock market. One of the purpose of Central Banks is also to maintain the financial stability.

The clearinghouse for each and every transaction is facilitated by an Exchange, which acts as a collector and deliverable of the shares, and payment is guaranteed to the vendor of an equity. The counterparty may default on the transaction is the risk faced by any buyer or seller that is being eliminated by the above mentioned.

The economic process is facilitated by the smooth functioning of these activities wherein the production of goods, services and employment is being promoted by cheaper costs and industrial risks. In this way increased affluence is assumed to be contributed by the financial system, though some discrepancy exists as to whether the ideal financial system is market or bank-based.

An enhanced degree of inspection of the influence of the structure of equity markets (called market microstructure) has been prompted by recent events like the global financial Crisis, in specific to the solidity of the financial system and the transference of systemic risk.

### *B. Trends in Equity Market*

Equity Markets basically shows two trends, Bullish and Bearish.

#### a) Bullish Market Trend

Bullish investors are the ones who believe that the price will go up. The one who invests is bullish on the latent stock if they purchase call. That is, the stock value will rise is their belief and have bought the stocks at a particular value referred to as the strike price by indemnifying for the correct.

A capitalist is moreover found as bullish on the stock if he is the one selling puts. To buy the stock is a covenant exercised by the one selling puts, hence, the stock value will rise is his belief.

#### b) Bearish Market Trend

Investors are found as bearish if they believe that a stock value can plummet. To sell the stock to the purchaser at a particular value is a covenant faced by the one selling call and the stock value will plummet is his belief thereby making him defeatist. In order to sell the stocks at a finer price to the vendor of the put contract the one who buys a put wants the price to plummet. They are moreover contemplated to be bearish.

## II. RELATED WORK

We have conducted the massive review of the literature to find, which methods of Data Mining are appropriate for the forecasting of financial time series. However, we have detected that actually only two of them are effective. There are neural networks and clustering. That is why the describing of other methods is excessive and the review should be decreased to only these two approaches.

As we have mentioned previously, we have made the wide review of the secondary sources. However, we are planning to decrease the number of considered articles to only five sources. The articles, which we plan to use, are related to the topic of research and were published from 1995 to 2013. We plan to prepare the short describing the model of stochastic volatility and Heston's model, while the discussion of neural networks and clustering will be made in a detailed way. As we have stated previously, it is due to the fact that we are focused on methods of Data Mining.

## III. APPLICATIONS OF DATA MINING

Since Nineteen Eighties equity market is being manipulated by data mining. Prediction of

commodity index, prediction of stock value, management of portfolio, portfolio risk management, detection of tendency etc. are the numerous facets of equity market where data mining is being employed. Association rules, decision trees, Particle Swarm Optimization, fuzzy logic, neural networks etc. are the numerous ways that are used.

#### *A. Stock Price Prediction*

Stock value estimation refers to estimating the value of the stocks. This estimation is everlasting or interim. By employing basic and scientific scrutiny stock estimation is performed. Interest rate, rate of exchange, yield of dividend, etc. are numerous economic science factors on which basic scrutiny depends. History replicates itself is the assumption on which scientific scrutiny is based upon. Starting price, ending price, maximum price, minimum price, etc. are the factors on the basis of which current price is estimated. Numerous other scientific measures may be covered as well. Stock values are also effected by numerous societal and governmental facets besides the above mentioned. Movable mean line, purpose and strategies of line are also covered by the standard ways used for stock estimation. Yet the aforementioned strategies do not have high rate of success.

#### *B. Stock Index Prediction*

Stock index is calculated as a weighted average of a particular stock market. Indices are made out of the stocks that are either belongs to same exchange, same business or same corporations. Dow Jones Industrial Average (DJIA), FTSE 100 index, BSE Sensex, S&P 500 index, DAX, Nifty are some of the popular stock indices. To get an overall perspective about the economic health of a business or exchange we generally look for the stock index as they act as a benchmarks. On the basis of the geography of the companies which are involved in the index calculation, indices are classified as local and global. Stock value, size of company and variety of shares are the parameters on the basis of

which stock indices are calculated and referred to as price weighted, market price weighted and market share weighted indices respectively. E.g. NASDAQ and S&P 500 are market share weighted indices.

Many brokers, dealers and stock market investors have always been interested in prediction of stock indices. However, estimating the best time to buy and sell has always been very difficult due to numerous factors to be taken into account that can affect the stock market.

In depth knowledge of the lead-lag relations timing within completely discrete variables and also understanding which constraints are more necessary to witness as indications, for analyzing how the market moves are required while predicting.

With increasing monetary market volatility and capital flows, it becomes very difficult to make decisions and to make decisions, prediction should be very accurate.

#### *C. Portfolio Management*

The investments made by an investor, which can be either buying or selling of stocks, bonds, annuities, mutual funds, or other investments are recorded in compiled form, known as Portfolio and the person who manages the portfolio is known as Portfolio Manager.

Portfolio Managers tend to diversify their investments for two purposes, to maximize the return and to minimize the risk, this process is known as Diversification.

The profits or loss including financial gain and change in values which are earned on investments are referred to as return. When this term comes into account in case of portfolio, this is referred to as weighted return of stocks.

Portfolio management makes decisions about diversifying investments and policies, it also deals with making decisions to fulfill the objectives of the investments, and asset allocation. The economical frontier of portfolios is that the set of portfolios that provides the best returns for each level risk (or equivalently, portfolios with the lowest risk for a given level of returns). Investment management consists of three phases: strategic asset allocation,

tactical asset allocation, and stock picking. Strategic asset allocation could be a long-run allocation strategy that implies selecting the market to invest in, choosing the type of assets to buy and therefore the distribution of these within the portfolio in accordance with the investor's objectives. Tactical asset allocation consists of frequently adjusting the portfolio, in a very systematic or discretionary approach, to take advantage of the short-term opportunities. Stock picking consists of choosing the most effective stocks to be incorporated within the portfolio. It's the foremost time intense section that impacts the return of the portfolio. Numerous 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.

#### *D. Discerning tendencies in equity market*

Scouring of key arrangements among the sections of securities market information can be done by using stock section info in fusion with subsequences of series of time. Logical conduct specific to their sector is demonstrated by a key arrangement which is an illustrative cluster of stocks. At a similar time several key arrangements might coexist in an exceedingly section. By employing mining to the data received from insider trading which is legal, the tendencies may be deduced. Firms' officers are needed to pass through the SEC sales data and buying of their firms' stock because of which the aforementioned information is obtainable.<sup>[10] [11]</sup> Following are the varied tendencies that have already been deduced:

- i) Guessing unusual equity returns: Due to the inside knowledge available, inside traders tend to make unusual returns. People on the outside can make increased gains if they get to know this inside knowledge. Unusual returns can be guessed better by the people on the outside using inside knowledge, if they concentrate on aspects like the financial condition and stability of the corporate. Also, analyzing

the coming months and predicting the fluctuations in the equity market for a particular stock. This kind of prediction system has been implemented using neural networks.<sup>[10] [11]</sup>

- ii) Oil prices and equity division returns: Since there have been huge ups and downs in the prices of oil recently, it can be clearly seen that oil plays a very important role in the international market. Thus, the impact of oil prices on the equity market should be analyzed further and deeper. In USA there is bidirectional overflow from oil market to equity market whereas European equity markets may experience an unidirectional overflow. This condition that causes the equity market to fluctuate due to oil prices has been analyzed using vector autoregressive-generalized conditional heteroscedasticity model.<sup>[10] [11]</sup>

#### *E. Recommender Mechanisms*

Actions like buying and selling of securities, as well as holding them and administration of portfolios by brokers as well as investors are carried out by taking assistance from Recommender Mechanisms. Finding products, guessing changes in the equity markets, short listing appropriate stocks to buy etc. are some of the main areas in which Recommender Mechanisms are applied.<sup>[10] [11]</sup> The various kinds of information and knowledge used in these mechanisms are:

- Information like global and local financial figures, interest rates, financial indexes, connections between companies that are public etc. These types of information are considered far sighted.
- Immediate or short sighted data like present prices of securities, gold, commodities, etc.
- Selection of data mining methods in deciding procedure.
- Company guidelines of the monetary corporation.

- Private expertise of the asset consultants.

Usually, keeping an eye on the future, an logical model is picked by the asset expert that uses the information available to derive some opening decision-making rules. These rules are then pooled together with his asset knowledge, signs of few picked immediate information and corporation rules of the investment organization to be included with last decisions.<sup>[10][11]</sup>

Security brokers manage majority of the buy as well as sell actions. Recommender mechanisms help investors to employ cheap brokers, who demand less money, when compared to complete service brokers, who demand more money for giving suggestion associated to trading. An important road-block to the usage of a recommender mechanism for equity market trading is that the inconsistency among the necessities of the asset analysts and their capability to tell their requirements to the electronic system. It is very problematic for a simple trader to implement data mining methods to the data himself as the difficulty of the overall data mining procedure. The recommender mechanisms stock up information in multiple styles according to the requirements, make documentation of the trades done, resources remaining, transactional information and give advice on which security to buy and which to sell. The numerous recommender mechanisms are modeled on association principle mining, commerce cleverness, sequential knowledge mining, multi-objective element group optimization, principle training etc.<sup>[10][11]</sup>

#### IV. RESULT

The model of stochastic volatility assumes that the dynamics of the price change of a non-dividend asset is subject to the law of geometric Brownian motion, but it assumes that the volatility of the price of an asset is not a constant but a random process.

The model of Heston's stochastic volatility better approximates the theoretical distribution of the asset

price to the real (empirical) distribution in comparison with the model of geometric Brownian motion with constant volatility. Heston's model takes into account the correlation between price and volatility, which allows us to better reflect potential market volatility, giving more accurate estimates of future values of asset quotes. Despite the above advantages, the Heston's model has many parameters that do not change with time, which indicates the limited accuracy of the calculations.

The mathematical concept of a neural network emerged from the mid-1940s after the publication of an article by W. McCulloch and W. Pitts on the logical calculation of ideas and nervous activity. Since the 1950s, the theory of neural networks was actively developed in the field of solving problems of artificial intelligence. Neural networks have become widely used in solving problems of prediction, pattern recognition, discriminant analysis, and others.

Neural networks are a system of artificial neurons, which are connected and interacting with each other. Each neuron deals only with signals that are periodically received and periodically sent to other neurons. Being connected to a sufficiently large network with controlled interaction, groups of neurons are able to solve rather complex tasks of cybernetics, adaptive control, forecasting, etc.

Neural networks are a system of artificial neurons, which are connected and interacting with each other. Each neuron deals only with signals that are periodically received and periodically sent to other neurons. Being connected to a sufficiently large network with controlled interaction, groups of neurons are able to solve rather complex tasks of cybernetics, adaptive control, forecasting, etc.

Despite the above advantages, practice shows that training of neural networks on raw data does not give the expected quality of forecasting. Before submitting data to the input of a neural network, it is almost always necessary to convert them to a form in which the neural network could better catch the hidden laws. Such data preparation is called

preprocessing. Usually, some algorithm is applied to the raw data, which is converting them into a more understandable form for the neural network. As practice shows, data preprocessing is the key to the efficient functioning of neural networks.

Clustering is the procedure for splitting of all the elements of a certain set into groups according to the principle of their similarity. In one group (class or cluster) there should be those elements of the set that are in a sense similar to each other. At the same time, any pair of elements from different classes must have significant differences from each other (within the same measure of similarity). Clustering is one of the most extensive areas of Data Mining. The methods of cluster analysis can solve the problems of generation and testing of hypotheses, the development of classification and so on.

Clustering methods can be used to achieve a variety of purposes. First, cluster analysis is applicable to a deeper understanding of the data by identifying the cluster structure. Splitting a set into groups of similar objects makes it possible to simplify and improve the quality of further data processing by applying to each cluster its analysis method. Secondly, clustering is applicable to the compression of the analyzed data. After splitting into clusters, further research is conducted on the most typical representative, while the remaining elements of the cluster are not used in the future. Thirdly, with the help of clustering, atypical objects can be detected. With its help, those elements of the set, which do not resemble any existing cluster, can be found.

Regardless of the subject of the study, the use of cluster analysis involves the following stages:

#### *A. Sampling for clustering*

The input data must satisfy the criteria of homogeneity and completeness. Homogeneity requires that all sampling objects should have the same nature and should be described by a similar set of characteristics. The criterion of completeness of data imposes a restriction on the representativeness of the sample.

#### *B. Pre-training of data*

In many cases, it is necessary to use some transformations of the original data, which would contribute to the improvement of the quality of clustering. Moreover, in almost all cases, normalization of data is required.

#### *C. The introduction of the feature of space*

To each object of the initial sample, a certain feature vector is put in correspondence, which is a set of characteristics of a particular object.

#### *D. The introduction of a measure of similarity (distance between vectors) in the space of features*

The smaller the distance between a pair of objects, the more they are similar. At the same time, the more distance, the fewer similarities they have.

In order to get closer to understanding how clustering is applicable to the forecasting of financial time series, it is necessary to plunge into the stock terminology. In the securities market, the concept of the trend is widely used. The concept of a trend is used to describe current market sentiment (market trends) to changes in stock quotes. Typically, three main types of the trend are singled out in the market: the ascending trend (when the price of the asset is prone to growth), the downward (when the asset price is downgraded), and the lateral one (when the price of the asset fluctuates around its current value). The main property of the trend is that it has a certain length in time. If somehow it is possible to reliably determine that at some point in time the element of the time series falls into one of three trends, then with some degree of certainty it can be judged that several points that are following the selected element will belong to the same trend. If at the initial moment of time we will occupy the appropriate position for the trend (for the uptrend the long position, and for the downtrend the short position), then after a certain period of time it will be possible to close the position with the profit.

Thus, each point of the financial time series will

be classified into one of the three clusters for an upward, downward or lateral trend. The main difficulty in solving this problem is the choice of the vector of features and the specification of the distance function between them. Currently, the clustering of financial time series is a very urgent topic, therefore, there are many articles offering different ways of introducing the feature vector and the distance function.

One way to dynamically cluster time supposes that clustering is not performed within the same time series, but clustering of a set of time series is performed. In other words, time series are divided into groups instead of their elements. The first element of the feature vector is its mathematical expectation (the expectation of many of its elements). As a second element, its standard deviation is calculated. The third element is the indicator trend. The fourth element is calculated based on the use of the coefficients of the expansion of the series by the method of main components. The fifth is the indicator of Hirst. As a function of similarity, the Euclidean distance is chosen, and the k-average algorithm is used for clustering.

In addition to partitioning the time series into disjoint clusters, there are algorithms of the Fuzzy clustering, when each element of the time series is mapped not to the class to which it belongs but to the vector of the probability that the element belongs to each of the existing clusters. Fuzzy clustering is very effective in forecasting financial time series since it provides the most honest answer to the question to which class the next point should be assigned. Moreover, such a probabilistic approach provides information that allows us to conduct the in-depth risk analysis that occurs during the building of trading strategies.

## V. CONCLUSION & FUTURE WORK

Such methods of Data Mining as neural networks and clustering are successfully used for forecasting

of financial time series. Among other methods, we can emphasize the models of stochastic volatility.

The using of cluster analysis should be conducted following the next compulsive stages of the process: sampling for clustering, pre-training of data, the introduction of the feature of space, the introduction of a measure of similarity (distance between vectors) in the space of features, application of one of the clustering algorithms and direct partitioning of the original set into groups.

Neural networks have several advantages: neural network analysis does not imply any restrictions on the nature of input information, the neural network is self-learning, neural networks are able to find the optimal indicators for a given time series and build on them an optimal prediction strategy, they do not use any prior statements about the nature of the data. In spite of these advantages, the quality of their work does not always correspond to expectations.

In the future, we plan to conduct the research, which will have the quantitative character and will be focused on the applying of Data Mining for the making predictions in the trading.

## REFERENCES

- [1] Hajizadeh, E., Ardakani, H. D., & Shahrabi, J. (2010). Application of data mining techniques in stock markets: A survey. *Journal of Economics and International Finance*, 2(7), 109.
- [2] Al-Radaideh, Q. A., Assaf, A., & Alnagi, E. (2013). Predicting stock prices using data mining techniques. In *The International Arab Conference on Information Technology (ACIT'2013)*.
- [3] Irigoien, I., Sierra, B., & Arenas, C. (2012). ICGE: an R package for detecting relevant clusters and atypical units in gene expression. *BMC bioinformatics*, 13(1), 30.
- [4] Lin, J. H., & Haug, P. J. (2006). Data preparation framework for preprocessing clinical data in data mining. In *AMIA Annual Symposium Proceedings (Vol. 2006, p. 489)*. American Medical Informatics Association.
- [5] Lee, Y. K., & Hoon, K. (1995). Brownian motion the research goes on.... This paper is available online.
- [6] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [7] Paranjape-Voditel, P., & Deshpande, U. (2013). A stock market portfolio recommender system based on

- association rule mining. *Applied Soft Computing*, 13(2), 1055-1063.
- [8] Khuat, T. T., & Le, M. H. (2017). An Application of Artificial Neural Networks and Fuzzy Logic on the Stock Price Prediction Problem. *International Journal on Informatics Visualization*, 1(2).
- [9] Sharmila, B., & Khanchana, R. (2017). A Survey on Analysis of Stock Market by using Data Mining Techniques. *Software Engineering and Technology*, 9(2), 27-30.
- [10] 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.
- [11] "Bioinfo Publication". *Bioinfopublication.org*. N.p., 2017. Web. 2 May 2017.