

A REVIEW: APPLICATION OF DATAMINING TOOLS FOR STOCK MARKET

Mrs. Keerti. S. Mahajan¹ & R. V. Kulkarni²

¹Research student, Chh. Shahu Institute of Business Education
And Research Center, Kolhapur, India

²Professor and HOD, Chh. Shahu Institute of Business Education and Research Center, Kolhapur, India

ABSTRACT

One of the most important problems in modern finance is finding efficient ways to summarize and visualize the stock market data to give individuals or institutions useful information about the market behavior for investment decisions. Therefore, Investment can be considered as a one of the fundamental pillars of national economy. The single and most important reason is to earn returns on investment. At the present time, many investors look to find a criterion for comparison stock together and selecting the best. The investors try in choosing the type of strategy, to choose strategies that maximize the earning value of the investment process.

The enormous amount of valuable data generated by the stock market has attracted researchers to explore this problem domain using different methodologies. Potential significant benefits of solving these problems motivated extensive research for years. Data mining is used for mining data from databases and finding out meaningful patterns from the database. Many organizations are now using these data mining techniques. So the research in data mining has gained a high attraction due to the importance of its applications and the increasing information.

Key words: Stock market, data mining, decision tree, neural network, clustering, association rules, factor analysis, time series.

INTRODUCTION:

Data mining, the science and technology of exploring data in order to discover previously unknown patterns, is a part of the overall process of knowledge discovery in databases (KDD). In today's computer-driven world, these databases contain massive quantities of information. The accessibility of this information makes data mining important and necessary. Financial institutions such as stock markets produce huge datasets that build a foundation for approaching these enormously complex and dynamic problems with data mining tools. Potential significant benefits of solving these problems motivated extensive research for years. The research in data mining has gained a high attraction due to the importance of its applications and the increasing information. Specifics of data mining in finance are coming from the need to accommodate specific efficiency criteria (e.g., the maximum of trading profit) to prediction accuracy, coordinated multiresolution forecast (minutes, days, weeks, months, and years), Be able to benefit from very subtle patterns with a short life time, and incorporate the impact of market players on market regularities (Boris and Evgenii, 2005).

STOCK MARKET:

A stock market or equity market, is a private or public market for the trading of company stock and derivatives of company stock at an agreed price; these are securities listed on a stock exchange as well as those only traded privately. The expression "stock market"

refers to the market that enables the trading of company stocks collective shares, other securities and derivatives. The stocks are listed and traded on stock exchanges which are entities a corporation or mutual organization specialized in the business of bringing buyers and sellers of stocks and securities together.

It is nowadays a common notion that vast amounts of capital are traded through the stock markets all around the world. recently the markets have become a more accessible investment tool, not only for strategic investors but for common people as well. Consequently they are not only related to macroeconomic parameters, but they influence everyday life in a more direct way. Therefore they constitute a mechanism which has important and direct social impacts. The characteristic that all stock markets have in common is the uncertainty, which is related to their short and long-term future state. This feature is undesirable for the investor but it is also unavoidable whenever the stock market is selected as the investment tool. The best that one can do is to try to reduce this uncertainty. Stock market prediction is one of the instruments in this process. Stock market forecasting includes uncovering market trends, planning investment strategies, identifying the best time to purchase the stocks and what stocks to purchase. Financial institutions produce huge data sets that build a foundation for approaching these enormously complex and dynamic problems with data mining tools. Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Application of decision tree in stock markets:

Decision trees are powerful and popular tools for classification and prediction. In data mining, a decision tree is a predictive model which can be used to represent both classifiers and regression models.

Decision trees are excellent tools for making financial or number based decisions where a lot of complex information needs to be taken into account. They provide an effective structure in which alternative decisions and the implications of taking those decisions can be laid down and evaluated. They also help you to form an accurate, balanced picture of the risks and rewards that can result from a particular choice. In a stock market, how to find right stocks and right timing to buy has been of great interest to investors. To achieve this objective, Muh-Cherng et al. (2006) present a stock trading method by combining the filter rule and the decision tree technique Listed companies' financial distress prediction is important to both listed companies and investors. Jie and Hui (2008) present a data mining method combining attribute-oriented induction, information gain, and decision tree, which is suitable for preprocessing financial data and constructing decision tree model for financial distress prediction. Accurately, forecasting stock prices has been extensively studied. Jar-Long and Shu-Hui (2006) provided a proposal to use a two-layer bias decision tree with technical indicators to create a decision rule that makes buy or not buy recommendations in the stock market.

Application of Neural network in stock market:

Neural networks have been successfully applied in a wide range of supervised and unsupervised learning applications. Neural network methods are commonly used for data mining tasks, because they often produce comprehensible models. A neural network is a computational technique that benefits from techniques similar to ones employed in the human brain.

The main advantage of neural networks is that they can approximate any nonlinear function to an arbitrary degree of accuracy with a suitable number of hidden units (Hornik et al., 1989). The development of powerful communication and trading facilities has enlarged the scope of selection for investors. David Enke and Suraphan Thawornwong introduced an

information gain technique used in machine learning for data mining to evaluate the predictive relationships of numerous financial and economic variables. Neural network models for level estimation and classification are then examined for their ability to provide an effective forecast of future values. A crossvalidation technique is also employed to improve the generalization ability of several models. The results show that the trading strategies guided by the classification models generate higher risk-adjusted profits than the buy-and-hold strategy, as well as those guided by the level-estimation based forecasts of the neural network and linear regression models.

Application of Clustering in stock market:

Clustering is a tool for data analysis, which solves classification problems. Its objective is to distribute cases (people, objects, events etc.) into groups, so that the degree of association can be strong between members of the same cluster and weak between members of different clusters. In clustering, there is no preclassified data and no distinction between independent and dependent variables. Instead, clustering algorithms search for groups of records (the clusters composed of records similar to each other). The algorithms discover these similarities. This way each cluster describes, in terms of data collected, the class to which its members belong. Clustering is a discovery tool. It may reveal associations and structure in data which, though not previously evident, nevertheless are sensible and useful once found.

As part of a stock market analysis and prediction system consisting of an expert system and clustering of stock prices, data is needed. Stock markets are recently triggering a growing interest in the physicists' community. Basaltoa et al. (2005) apply a pair wise clustering approach to the analysis of the Dow Jones index companies, in order to identify similar temporal behavior of the traded stock prices. The objective of this attention is to understand the underlying

dynamics which rules the companies' stock prices. In particular, it would be useful to find, inside a given stock market index, groups of companies sharing a similar temporal behavior. To this purpose, a clustering approach to the problem may represent a good strategy.

Application of Association rules in stock market:

In data mining, association rule is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro describes analyzing and presenting strong rules discovered in databases using different measures of Interestingness (Piatetsky-Shapiro, 1991). Based on the concept of strong rules, Agrawal et al. introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) systems in supermarkets (Agrawal et al., 1993). Association rule mining finds interesting associations and/or correlation relationships among large set of data items. Association rules shows attributed value conditions that occur frequently together in a given dataset. Mining association rules on large data sets has received considerable attention in recent years. Association rules are useful for determining correlations between attributes of a relation and have applications in marketing, financial, and retail sectors. Furthermore, optimized association rules are an effective way to focus on the most interesting characteristics involving certain attributes. Optimized association rules are permitted to contain uninstantiated attributes and the problem is to determine instantiations such that either the support or confidence of the rule is maximized.

Application of Factor analysis in stock market:

Factor analysis is particularly useful in situations where a large number of variables are believed to be determined by a relatively few common

causes of variation. Also, it should be particularly useful for analyzing financial markets because if financial markets are efficient, nominal returns will be affected by default and market risk and by expected inflation and inflation uncertainty. Michael Flad et al. provided an empirical analysis of the

common factor driving the intraday movements of the DAX and the DJIA during overlapping trading hours. Based on a minute-by-minute dataset spanning from

March to December, 2003, they estimate a bivariate common factor model for the two indices. By explicitly modeling the two stock indices, they implicitly assume that news on economic fundamentals is aggregated in both equity market and that, therefore, both stock indices are linked by a common trend of cumulated random information arrivals. They compute various measures of information leadership and find that the DJIA is the predominant source of price relevant information flowing into the transatlantic system of stock indices.

Recent research on dynamic factor models finds that the information in a large number of economic time series can be effectively summarized by a relatively small number of estimated factors, affording the opportunity to exploit a rich base of information more likely to span the information sets of financial market participants than in previous analyses. In doing so, their study contributes to the empirical literature by evaluating both the potential role of omitted information in the estimated risk–return relation as well as the robustness of previous results to conditioning on richer information sets .

Application of time series in stock markets:

It is obvious that forecasting activities play an important role in our daily life. The traditional statistical approaches for time series can predict problems arising from new trends, but fail to forecast the data with linguistic facts. Furthermore, the traditional time series

requires more historical data along with some assumptions like normality postulates.

In recent years, many researchers have used fuzzy time series to handle forecasting problems. A number of researchers presented fuzzy time series forecasting models in the last 15 years (Tahseen and Syed, 2008).

Time series data is characterized as large in data size, high dimensionality and update continuously. Moreover, the time series data is always considered as a whole instead of individual numerical fields. Indeed, a large set of time series data is from stock market. Moreover, dimensionality reduction is an essential step before many time series analysis and mining tasks. For these reasons, research is prompted to augment existing technologies and build new representation to manage financial time series data. Tak-chung et al. (2008) have represented financial time series according to the importance of the data points.

REVIEW OF LITERATURE:

Chavoshi (2003),”the Tehran Stock Exchange: Artificial neural network models and multi-factor model”.

The research results stock returns using macroeconomic variables However, artificial neural networks are more successful at this and can significantly reduce the forecast error.

Avrma (2004), “Stock return predictability and asset pricing models”

Using asset pricing models.

In this research paper Avrma presented result on Asset allocations based on Conditional models outperform their unconditional counterparts that exclude return predictability.

Ravichandran et al., (2005),”Estimation of return On Investment In Share Market through ANN using artificial neural network.
Muh-Cherng, “A STOCK TRADING METHOD” by combining the filter rule and decision tree technique, 2006.

In a stock market, how to find right stocks and right timing to buy has been of great interest to investors. To achieve this objective muh-cherng presents a “STOCK TRADING METHOD”. The filter rule, have been widely used by investors, is used to generate candidate trading points. These points are subsequently clustered and screened by the application of a decision tree algorithm. Compared to previous literature that applied such a combination technique, this research is distinct in incorporating the future information into the criteria for clustering the trading points. TAIWAN and NASDAQ stock markets are used to justify the proposed method. Considering our Indian stock market, implementing decision tree method will be another challenge.

Namazi and Kyanmehr (2007),”Predict daily stock returns of companies listed in Tehran Stock Exchange using neural networks”.

Artificial neural networks have the ability to predict daily returns with the error rate are relatively good Multi-factor model is able to predict.

Rahmani and Saeedi (2008),”Performance Evaluation of logit models in forecasting stock returns”.

Out of 12 Portfolio only 4 portfolios have been successful Obtaining additional returns to the market and there was no significant difference Performance of two models logit and regression.

Jie and Hui, “Listed companies’ financial distress prediction”, 2008.

This case study highlights the importance of prediction to both listed companies and investors. Implementing a data mining method by combining attribute-oriented induction, information gain, and decision tree, this is suitable for preprocessing financial data and constructing decision tree model for financial distress prediction. The empirical experiment with 35 financial ratios and 135 pairs of listed companies as initial samples got satisfying results, which testifies to the feasibility and validity of the proposed data mining method for listed companies’ financial distress prediction.

Tae, “PORTFOLIO MANAGEMENT OF STOCK PRICES”, 2007

Accurate volatility forecasting is the core task in the risk management in which various portfolio pricings, hedging, and option strategies are exercised. So, tae proposes hybrid models with neural network and time series models for forecasting the volatility of stock price index in two view points: deviation and direction. It demonstrates the utility of the neural network forecasting combined with time series analysis for the financial goods.

Jar-Long and Shu-Hui’ “FORECASTING STOCK PRICES”, 2006

This research provides a proposal to use a two layer bias decision tree with technical indicators to create a decision rule that makes buy or not to buy recommendations in the stock market. A novel method designed for using two-layer bias decision tree to improve purchasing accuracy. Comparison with random purchases, the results indicate the system presented here not only has excellent out of sample forecasting performance, but also delivers a significant improvement in

investment returns for all listed companies. Additionally, the proposed system has few parameter requirements, stable learning, and fast learning speed. Increasingly, the system presented here has high accuracy given large amounts of varied test data, with testing periods that experienced structural change including both bull and bear markets. Based on all of the above, they believe the proposed bias decision model is very flexible, modular and easily understandable.

Chi-Lin Lu and Ta-Cheng Chen,
“CLASSIFICATION RULES OF
INFORMATION TRANSFERENCY
LEVELS OF THE LISTED FIRMS”,
in Taiwan’s stock market.

Decision tree based mining techniques to explore the hidden knowledge of information disclosure status among the listed companies in Taiwan’s stock market. Moreover, the multilearner model constructed with decision tree algorithm has been applied. By using

Decision tree –based rule mining approach, the significant factors with the corresponding equality/ inequality and threshold values were decided simultaneously, so as to generate the decision rules , and thus it has provided effective decision supports for judging the information disclosure problems in Taiwan’s stock market.

Abhijit Dutta (2001)

Author has examined the investor’s reaction to information using primary data collected from 600 individual investor’s band observers that the individual investors are less reactive to bad news as they invest for longer period.

Prabina Das, S.Srinivasan and A. K. Dutta (2000)

Have studied the reaction of GDR (gross domestic ratio) prices and the underlying share prices to the announcement of dividends and found that the CAR for the GDR is mostly negative irrespective of the rate of dividend whereas the domestic share prices react in a more synchronous manner.

Jijo Lukose and Narayan Rao (2002)

Examined the security price behavior around the announcement of stock splits and around ex-split date.

Horvath and Zuckerman (1993)

Suggested that one’s biological, demographic and socioeconomic characteristics; together with his/her psychological makeup affects one’s risk tolerance level.

Mitra (1995)

Discussed factors that were related to individuals risk tolerance, which included years until retirement, knowledge sophistication, income and net worth.

Cohn, Lewellen et.al

Found risky asset fraction of the portfolio to be positively correlated with income and age and negatively correlated with marital status.

Llewellyn et.al

While identifying the systematic patterns of investment behavior exhibited by individuals found age and expressed risk taking propensities to be inversely related with major shifts taking place at age 55 and beyond.

Rajrajan V (1997, 1998, 2000 and 2003)

Author has classified investors on the basis of their demographics. An author has also brought out the investors characteristics on the basis of their investment size. He found that the percentage of risky assets to total financial investments had declined as the investor moves up through various stages in life cycle. Also investors' lifestyles based characteristics has been identified.

Al – Tamimi (2007)

Identified company fundamental factors (performance of the company, change in board of directors, appointment of new management, and the creation of new assets, dividends, earnings), and external factors (government rules and regulations, inflation, and other economic conditions, investor behaviour, market conditions, money supply, competitions uncontrolled natural or environmental circumstances) as influencers of asset prices. He developed a simple regression model to measure the coefficients of correlation between the independent and dependent variables.

Rigobon and Sack (2004)

Discovered that increases in war risk caused declines in Treasury yields and equity prices, a widening of lower-grade corporate spreads, a fall in the dollar, and arise in oil prices. A positive correlation exists between the price of oil and war. They argue that oil has significant impact on market.

Basaltoa,” Portfolio Optimization”, 2005

The identification of clusters of companies of a given stock market index can be exploited in the portfolio optimization strategies (Basaltoa et al., 2005). So the graph representation of the stock market data and interpretation of the properties of this graph gives a new insight in to the internal structure of the stock market. They have used “self organized” system, the edge density of the market graph, as well as the maximum clique size, steadily increases during the last several years, which supports the well-known idea about the globalization of economy which has been widely discussed recently and also they used tick-by tick transaction data, to study price clustering on the TOKYO STOCK EXCHANGE, which is computerized limit order market. As for the intraday pattern, the degree of price clustering is greatest at the market opening. Then, it decreases during the first half hour and reaches a stable level. It does not increase again near the market closing. There is no clear difference in clustering between call auctions and continuous auctions. Again in 2008 Shu-Hsien investigated stock market investment issues on Taiwan Stock Market using a two stage Data mining approach.

The first stage apriori algorithm is a methodology of association rules, which is implemented to mine knowledge patterns and rules in order to propose stock category association and possible stock category investment collections.

Then the K-means algorithm is a methodology of cluster analysis implemented to explore the stock cluster in order to mine stock category clusters for investment information. By doing so, they propose several possible Taiwan Stock Market portfolio alternatives under different circumstances.

CONCLUSION:

With the increase of economic globalization and evolution of information technology, financial data are being generated and accumulated at an unprecedented pace. As a result, there has been a critical need for automated approaches to effective and efficient utilization of massive amount of financial data to support companies and individuals in strategic planning and

investment decision making. Data mining techniques have been used to uncover hidden patterns and predict future trends and behaviors in financial markets. The competitive advantages achieved by data mining include increased revenue, reduced cost, and much improved marketplace responsiveness and awareness.

It is observed from the above literature that most of researches has been done on comparing simulated trading and forecasting results with results of other methods and real gain/loss like stock return predictability and asset pricing model, estimation of return on investment, forecasting stock pricing, portfolio optimization etc by TOKYO stock exchange, TEHRAN, TAIWAN and NASDAQ stock market.

But according to Indian scenario the situation is different from other countries because, The mentality of Indian investors are different, financial status is different, attitude of people is different. Therefore by considering the Indian stock market situation and the attitude of Indian people towards investment in stock market authors would like to research impact of different scripts of stock market by using applications of datamining tools to benefit Indian people.

REFERENCES

- Calderon T.G., and Cheh J.J., (2002), „A roadmap for future neural networks research in auditing and risk assessment”, International Journal of Accounting Information Systems, Vol. 3, No. 4, pp. 203-236.
- Das, S. and Chen, M. 2001. Yahoo for Amazon: Extracting Market Sentiment from Stock Message Boards, Proceedings of the 8th Asia Pacific Finance Association Annual Conference (APFA 2001), Bangkok, Thailand, July 22-25.
- Agrawal R, Imilienski T, Swami A (1993). Mining association rules between sets of items in large databases, In Proceedings of the ACM SIGMOD international conference on management of data.
- Basaltoa N, Bellottib R, De Carlob F, Facchib P, Pascasio S (2005). Clustering stock market companies via chaotic map synchronization,
- Chi-Lin L, Ta-Cheng C (2009). A study of applying data mining Approach to the information disclosure for Taiwan's stock market Investors, Expert Systems with Applications.
- David E, Suraphan T (2005). The use of data mining and neural Networks for forecasting stock market returns, Expert Systems with Applications.
- Leonid Churilov, Adyl Bagirov, Daniel Schwartz, Kate Smith, Michael Dally, Journal of management Information system: 2005, Data mining with combined use of optimization techniques and self organizing maps for improving risk grouping rules: application to prostate cancer patients
- Anthony Danna, Oscar H. Gandy, and Journal of business ethics: 2002, All that glitters is not gold: Digging Beneath the surface of data mining.
- AC Yeo, KA Smith, RJ Willis and M Brooks, Journal of the operation research society : 2002 , A mathematical programming approach to optimize insurance premium pricing within a data mining framework.
- Shakil Ahmed, Frans Coenen, Paul Leng, Knowledge Information System : 2006, Tree based partitioning of data for association rule mining
- Data Mining: Concepts and Techniques Jiawei Han and Micheline Kamber, Morgan Kaufmann
- Data Mining for Association Rules and Sequential Patterns Jean-Marc Adamo, Springer.
- Cowan A (2002). Book review: Data Mining in Finance, Int. J. forecasting.
- David E, Suraphan T (2005). The use of data mining and neural networks for forecasting stock market returns, Expert Systems with Applications.
- Defu Z, Qingshan J, Xin L (2004). Application of Neural Networks in

Financial Data Mining, Proceedings of world academy of science, Eng. Techno.

16.Hornik K (1989). Stinchcombe M. and White H.,
“Multilayer feed forward
Networks are universal approximators”, Neural
Networks.

17.Hsiao-Fan W, Ching-Yi K (2004). Factor Analysis in
Data Mining, Computers and Mathematics with
Applications. http://en.wikipedia.org/wiki/Stock_market
<http://www.anderson.ucla.edu/faculty/jason.frans/teacher/technologies/palace/datamining.htm>
http://www.resample.com/xlminer/help/Assocrules/associationrules_intro.htm

18.Huang K, Yu HK (2005). A type 2 fuzzy time series
model for stock Index forecasting, Physical.

19.Jar-Long W, Shu-Hui C (2006). Stock market trading
rule discovery using two-layer bias decision tree, Expert
Systems with Applications.

20.Jie S, Hui L (2008). Data mining method for listed
companies' financial distress prediction, Knowledge-
Based Systems.