CHAPTER - 2

REVIEW OF LITERATURE
CHAPTER- 2

REVIEW OF LITERATURE

Processing and differentiating the massive data sets is carried out by automated tools to predict patterns of solutions in the area of finance, marketing, web based e-business and its allied regions and in stock market. But widespread review illustrates that customer behavior functions in stock market do not have the farthest expertise in computational grounds. The reviewed works were summarized in view of feasibility to opt the application on the proposed intelligent system design.


G. McQueen and S. Thorley [24] interpret an algorithmic knowledge engineering in current terms of technical analysis stock charting, which is traditionally practiced in a very subjective and non computational fashion.

A. Lo et al. [23] tests price charting heuristics using kernel regression for pattern identification and find marginally positive results.
with filter rules using ten patterns on price and evaluates performance using the distribution of returns, rather than excess profits, and work only with single-day returns.

T. Roberts [26] suggests the Efficient Markets Hypothesis (EMH). EMH may be interpreted to mean that market prices are best described as a random walk, and past price and volume information is worthless for predicting future market price behavior. Some researchers believe that the EMH scientific research program is somewhat in a “degenerate” stage, and progress in understanding financial markets requires the emergence of new, positive computational research programs. The EMH research program and paradigm, surveyed and exemplified, concentrates on developing and defending analytic information models of rational expectations which predict the evolution of the hybrid information system.

The research group R. Sullivan, A. Timmermann and H. White [27] has developed information processing models in asset markets which have established a strong, new research program replacing EMH. The models of the researcher S. Sunder et al [28] in asset market information processing are concerned with decision making where market reasoning mechanism actually works. Much of these works
results from laboratory simulations and computer-based market simulation models and waits to be empirically tested.

J. Campbell, S. Grossman and J. Wang [20] find that individual stock price changes accompanied by high stock market trading volume tend to be reversed, but that this is less true of individual stock price changes associated with low stock market trading volume. S. Brown, W. Goetzmann and A. Kumar [19] conclude, “trading volume is driven mainly by non-informational trades, while stock price movements are driven primarily by informational trades.” Both of these studies imply that volume is the result of factors which are not related to the intrinsic value or future price of the security. On the other hand, K. Chan, A. Hameed and W. Tong [21] find “higher profits for momentum portfolios implemented on markets with higher volume in the previous period, indicating that return continuation is stronger following an increase in trading volume. This result confirms the informational role of volume and its applicability in technical analysis.”

T. Bollerslev and D. Jubinski [18] and S. Neftci [25] are examples of studies which document positive contemporaneous correlation between trading volume and volatility, that is, absolute change in price. These studies continue the work on the mixtures-of-
distributions hypothesis originally formulated by P. Clark [22], by which volume and price are affected by the same “news” arrival and are variable so that “good news” causes a price increase and “bad news” causes a price decrease of stock volumes.

Predicting the future is a dominant problem category in finance and banking. Wilson and Sharda [17] demonstrate the superiority of ANNs over discriminate analysis method in predicting bankruptcy of a firm. Their ANN based system achieved a prediction accuracy of 97%.

T.K. Sung, N. Chang and G. Lee [51] show another application in this area, that a RI based model could achieve higher prediction accuracy than the discriminate analysis model in predicting bankruptcy under varied economic conditions.

Messier and Hansen [52] report the use of RI to predict defaulters of loans. Their study shows that a RI based model can achieve better prediction accuracy than discriminate analysis. Barr and Mani [53] report the use of artificial neural network and RI to forecast the price of the S&P 500 Index. The ANN was trained using time series data with 21 indicator variables as input and achieved a prediction accuracy of 0.92 for the change in the movement of the
index. Carter and Catlett [54] describe the use of RI to assess the reliability of credit card applicants. Using a data set containing both continuous and discrete valued attributes, their study showed that RI is superior to discriminate analysis and rule-based expert systems in this application. REDON [55] ranks stocks based on their risk and return performance and allows the user to create portfolios of stocks based on his/her risk tolerance level.

Some data mining applications employ more than one machine learning technique. Kim and Noh [56] report an integrated system that combines case based reasoning (CBR) with ANN to forecast interest rates for corporate bonds and treasury bills. Their integrated model outperformed a random walk model in predicting US interest rates, but was not so successful in predicting Korean interest rates. Another such hybrid system uses CBR with RI to rate corporate bonds [57]. It provides superior judgment about bond rating situations by complementing knowledge gathered from rules on how to rate bonds with similarity metrics obtained from past bond rating situations. This system matched the S&P recommended ratings 90.4% of the time for companies with complete data and 84.4% of the time for companies with incomplete data. A hybrid RI–ANN approach is presented in [58]
for predicting the direction of the daily price changes in the S&P 500 stock index futures. The RI subsystem selects key indicator variables that are then used to train the NN subsystem, which makes recommendations on the direction of the price change.

Data visualization has been used in some data mining applications also in the field of finance. MineSet [59] uses supervised and unsupervised learning for loan approval, churn management in banking, and detection of credit card frauds. It allows creation of visual decision trees for obtaining business rules. Gershon and Eick [60] discuss visualization based data mining applications in fixed income and derivative risk management, what-if analysis for decision support, and credit application evaluation. Shaw and Gentry [61] use inductive learning for risk classification in loan evaluation and bond rating. The goal of this system is to make generalizations about the outcome in each task using a large set of variables. It demonstrates the superior performance of RI over discriminate analysis.

GUHA, KEX, and KnowledgeSeeker [62] identify classes of accounts with interesting behavior patterns, such as periodically changing small credit and debit balances, from financial transaction databases. G.H. John, Y. Zhao [63] made intelligent mining which
combines the predictive features of ANN with visualization, to detect delinquent bank loans. It uses attributes such as loan-to-value ratio, origination amount, period and type of loan, etc., to predict whether a mortgage is likely to become delinquent. FALCON [64] uses an ANN based approach to identify suspicious credit card transactions whereas Kokkinaki [65] proposes the use of similarity trees for a similar application. FAIS [66] detects large cash transactions that indicate potential for money laundering operations.

Computer network also offers an attractive domain for data mining because of the intensive nature of applications in this domain. One class of applications involves finding trends and patterns in operational characteristics of networks to diagnose chronic faults. Sasisekharan and Seshadri [67] combine statistical methods and machine learning techniques to identify patterns of chronic problems in network operations, and use these patterns to predict potential faults in telecom networks.

R.T. Scarfe, and R. J. Shortland [68] identify the frequently occurring alarm episodes and encode their predictive characteristics as rules to predict likelihood of occurrence of such episodes. Fraud
detection is another important application area in telecommunications. R. Shortland and R. Scarfe [69]

B. Allen [70], S.S. Anand et al [71], and A. Berson et al [74] focused data mining applications in marketing to include retail sales analysis, market basket analysis, product performance analysis, and market segmentation analysis. British Telecom uses RI to distinguish between users and nonusers of telecommunication products [75]. This information facilitates tailoring new services to closely meet customers’ needs. Cross-sales, i.e. selling a product with enhanced features to current customers, can be improved by associating customer characteristics with product features. Anand et al. [72and 73] describe a RI based methodology for identifying target customers for a new insurance product offered by a bank from among the account holders of that bank.

IDEA [79] analyzes the effect of new promotions on market behavior in the telecommunications industry. It provides interactive querying and drill-down capabilities on customer characteristics. Bhattacharya [76] discusses the use of GA to classify customers that are most likely to respond to a marketing campaign. This study identifies the mailing depth (i.e. percentage of customers to whom the
advertisement needs to be sent) required to maximize return on advertisement under a budget constraint.

Data visualization has been used in conjunction with other machine learning techniques in marketing applications. WinViz [78] has been used to study transportation, food, shopping, and travel patterns for numerous households in Singapore. Visualization has also been used for analyzing the performance of consumer products and for identifying factors that affect this performance. This system enables the user to interactively compare performances of different stores with the benchmark. Systems specializing in market basket analysis include Rules Visualizer of MineSet and Nicheworks [77]. Both these tools use visualization to identify products that are frequently purchased together. Rules Visualizer found, for example, that grocery items like bread and dry-food condiments are usually sold together. Nicheworks discovered a strong relationship between purchase of discounted hardware items and plumbing items in hardware stores.

Based on the reviews mentioned above, it is discovered that the work attempted for making predictions on Safe Investor Trading in stock market business using either linear or non-linear methods is almost nil. If so by what technique is that business process mining still
going on? The answer to the question is that only by simple statistical
and probabilistic wear and tear predicate ethics the business pulls on.
This technology gap is sought to be bridged by a machine learning
process. The attempt uses the methodology of an intelligent hybrid
model of Rule Induction (RI) based Adaptive Fuzzy Apriori (AFA) -
Artificial Neural Network (ANN) data mining algorithm which aims at
a management information system in knowledge process outsourcing
by predicting the trading behavior of the investor that leads to a
profitable business in the stock market.