LITERATURE REVIEW

Keogh and Kasetty (2003) They have execute beginning experiment on the data using the chi-square test and decision tree which are important to the stock market index of kula lumpur. The experiment of authors produces classification rules which lead to inter-relationships among the stocks in terms of their trading performance with regards to klci. They have used the trend analysis, segmentation and clustering in this paper.

Zhang and Zhou (2004) They have discussed KDD in the reference to financial application. They referred that developing data mining techniques have shown huge potentials in financial applications and will continue to flourish in the new knowledge-based economy. They list the data mining algorithm are Association rule mining (ARM) uncovers attractive connection patterns among a large set of data items by showing attribute-value conditions that occur together frequently. Classification and prediction is models that explain and differentiate data classes or concepts, Clustering analysis segments a huge set of data into subsets or clusters. Each cluster is a collection of data objects that are similar to one another within the same cluster but dissimilar to objects in other clusters. Sequential pattern and time-series mining looks for patterns where one event (or value) leads to another later event (or value).

Ting et. al (2006) They used the Sequential and non-sequential association Rule mining (ARM) were used to perform intra and inter-stock pattern mining. They deriving the probable Association among the stocks, authors observed the performance of Stocks with respect to a specific stock index.

Boginski et. al (2006) Authors describe the mining association rules in network approach which is Exploring the association and hidden relationships between item sets within a transaction. They have give and experiment of stock with example “if the stock prices of transflo and parsian go down, at 80% of probability, the price of iran khodro goes down”. That is proved by them in 40% of the Transactions. Here they take the probability posed as “confidence” and the percentage of the transactions cover this rule is called “support” with regards to the association rule.

Sudhanshu et. al (2009) In the presented research paper authors have applied genetic algorithm with minimum complexity over the rules fetched from Apriori association rule mining and used baye’s theorem on the generated rules. Their proposed system for generating association rule by genetic Algorithm by frequent item sets are generated using the Apriori association rule mining algorithm, and genetic algorithm item sets to generate the rules containing positive attributes, the negation of the attributes with the resultant part consists of single attribute and more than one attribute. The results are in this paper promising as the discovered rules are optimized rules. They describe that addition to the positive associations, negative associations can provide valuable information.

Kannan and Bhaskaran (2009) They have applied pruning or grouping rules by some means is necessary to get very important rules or knowledge. In ARM, large number of Association rules or patterns or knowledge is generated from the large volume of dataset. They describe two ways for selecting very appealing rules is using interestingness measures to rank and select a small set
of rules of different characteristics. Another way is forming groups or clusters of rules and selecting very important rules from each cluster. They have described method of grouping or clustering rules and selection of cover rules.

**Jingrong and Chunyu (2009)** In this article, they systematically, deeply and comprehensively analyzed and studied the association rule data mining technology. They used to correlation to measure the relations among item sets, and gave the computations of support level and confidence level of negative association rule based on traditional association rules, and analyzed. During the demonstration test of the algorithm, the results indicated that the algorithm was effective.

**Hajizadeh et. al (2010)** This provides an overview of application of data mining techniques such as decision tree, neural network, association rules, factor analysis and etc in stock markets. Also, this paper reveals progressive applications in addition to existing gap and less considered area. This paper has mainly concentrated on applications of the algorithms.

**Nikfarjam et. al (2010)** They states problems of data mining in stock market for data mining methods including in making interpretations, incorporating relations and probabilistic learning. The data mining techniques outlined in this paper advances pattern discovery methods that deals with complex numeric and non-numeric data, involving structured objects, text and data in a variety of discrete and continuous scales (nominal, order, absolute and so on). Also, this paper shows benefits of using such techniques for stock market forecast.

**Tiwari et. al (2010)** They proposed a hybrid system by using Classification methods and hhmm which describes the future market trends based on historical stock market data. This paper presents a hybrid system based on decision tree rough set, for predicting the trends in the BSE with the combination of hierarchical hidden markov model. They present future trends on the bases of price earnings and dividend.

**Hao (2010)** Author underscore that traditional sequential pattern mining only focuses disclose the order of item sets, but a sequential Pattern with the time intervals between successive item sets is more valuable than a traditional Sequential pattern. To solving Problems, a new algorithm for discovering target-oriented time-interval sequential patterns is presented in this paper. The main concept of this algorithm is to reverse the original sequence Dataset to enhance the efficiency for searching the target patterns. In addition, the clustering Analysis is used to automatically generate the suitable time partitions between successive Item sets.

**Yassir and Nayak (2010)** In this paper authors focused on data mining subject and issue data mining steps like association of events that can be correlated. Sequences as one event leads to another. Classification through the recognition of patterns. Forecasting which is a natural extrapolation from the other results and can facilitate more accurate projections.

**Kumar et. al (2010)** In this paper authors used clustering techniques. Clustering techniques are used to group up the relevant data sets. Hierarchical and partitioned clustering techniques are used for the clustering process. The clustering process is the complex task with high process
time. The system is implemented as four major modules data set management, pattern extraction, clustering process and performance analysis. The data sets are preprocessed before the pattern extraction process. This concept is applied in the simultaneous pattern extraction and Clustering scheme. The Simultaneous pattern extraction and clustering system is implemented to cluster the breast cancer patient diagnosis Records. The current system is implemented as a stand-alone Application.

Nair et. al (2010) They proposed system which include genetic algorithm optimized decision tree-support vector machine (SVM) hybrid, which can predict one-day-ahead trends in stock markets it can adapt itself to the changing market conditions and most of the attempts at stock market trend prediction have approached it as a regression problem, they check the Performance of the proposed hybrid system is validated on the historical time series data from the BSE-Sensex. The system performance is then compared to that of an artificial neural network (ANN) based system and a naïve Bayes based system. It is found that the trend prediction accuracy is highest for the hybrid system and the genetic algorithm optimized decision tree- SVM hybrid system outperforms both the artificial neural network and the naïve bayes based trend prediction systems.

Setty et. al (2010) In this paper Companies were clustered under sector wise based on the financial ratio analysis & clustering analysis and for better investment. Data analysis like clustering, classification or anomaly analysis. Their main research work was set to develop the clusters of NIFTY companies for better investment. Price per earnings ratios were calculated for all the 50 NIFTY companies during years 2008-2009 & 2009-2010. and clustering of companies under sector wise were made based on the financial ratio analysis and clustering analysis.

Kerana and Rangaswamy (2011) Authors presents an efficient partition algorithm for mining frequent item sets(pafi) using Clustering. This algorithm finds the frequent item sets by partitioning the database transactions into clusters. Clusters are formed based on the similarity measures between the transactions. Then it finds the frequent Item sets with the transactions in the clusters directly using improved Apriori algorithm which further reduces the number of scans in the database and hence improve the efficiency. Clustering can be considered the most important unsupervised learning Problem.

Cenk and Ceren (2011) In this study, information on data mining concept and data mining methods has been rendered, and acknowledge that the purpose of data mining is to create decision making models for estimation of the behaviors in the future based on the analysis of the past activities. Data analysis techniques that have been traditionally used for such tasks include regression analysis, cluster analysis, numerical taxonomy, multidimensional analysis, other multivariate statistical methods, stochastic models, time series analysis, nonlinear estimation techniques, and others. Data mining can be used in the field of finance, particularly for early determination of financial failure, determination of financial information manipulation and presupposition of bear and bull periods in the stock exchanges.

Prasad and Prasad (2011) This paper presents a distributed association rule mining algorithm by utilizing the global count on frequent item sets. Authors observed that, Distributed Count
Association Rule Mining provides an efficient method for generating association rules from different datasets, distributed among various sites. Association rules provide information of this type in the form of "if-then" statements. These rules are computed from the data and association rules are probabilistic in nature.

Jadhav et. al (2011) In this paper, implementation of a system for pattern discovery using association rules is discussed as a method for web usage mining. Different transactions that are closely related to each other are grouped together by the use of clustering approaches on the preprocessed dataset. The analysis of such clusters will lead to discovery of strong association rules. They obtained all significant association rules between items in the large database of transactions. The relation between different page requests was found. The support and the confidence values of extracted rules are considered for obtaining the interest of the web visitors.

Kumar and Chandrashekharan (2011) They proposed attribute correction using data mining which follows two approaches, clustering techniques for context independent correlation and association rule for context-dependent correction. Context-dependent means that attribute values are connected regards to the reference data values, to which it is most similar and also take values of other attributes into consideration. This algorithm makes use of association rule to discover validation rule for data set. Apriori algorithm is used to generate frequent item sets. Context-independent correction means that all the records attributes are examined and cleaned in isolation without regards to values of the attribute of a given records.

Kumar and Kalia (2011) Authors have applied association rules algorithm to a real time dataset, and a group of association rules and frequent item sets are obtained. First item sets are mined for a given minimum support and based on these item sets obtained, the association rules are computed for a given minimum confidence. The pattern so generated helps investors to build their portfolio and using these patterns investment strategies may be planned. Further it is observed that a portfolio is safely built with major large cap companies as it is evident that mainly pattern recommends. Further multiple minimum supports specified based on the different characteristics of association rules in order to discover frequent item sets.

Dwivedi (2012) Author highlight that the data mining algorithm is the mechanism that creates a data mining model. To create a model, an Algorithm first analyzes a set of data and looks for specific patterns and trends. The algorithm uses the results of this analysis to define the parameters of the mining model. These parameters are then applied across the entire data set to extract actionable patterns and detailed statistics.

Adebiyi et. al (2012) They proposed hybrid approach that combines the variables of technical and fundamental analysis for the creation of neural network predictive model for stock price prediction. The technical analysis variables are the core Stock market indices (current stock price, opening price, closing price, volume, highest price and lowest price etc.) While the fundamental analysis variables are company performance indices (price per annual earning, rumor/news, book value and financial status etc.).
Rana and Morena (2011) They have proposed a methodology based on cluster analysis to group stocks in the Indian stock market. Based on their end of day data, they demonstrate with example how an investor can maximize profit by trading a stock in a cluster with regards to the movement of the associated stock of the cluster. They have used K-means algorithm and detect the cluster based upon the percentage change of their end of day data listed on the NSE.

Sharma and Bhatia (2012) Authors defined time series as a set of data generated sequentially in time. The time series models assume that in the absence of major disruptions to critical factors of a recurring event, the data of this event in the future will be related to that of the past events and can be expressed via models developed from the past events.

Shaikh and Chhajed (2012) They review the different financial forecasting using data mining mainly focused on clustering. Clustering analysis segments a large set of data into subsets or clusters. Each cluster is a collection of data objects that are similar to one another within the same cluster but dissimilar to objects in other clusters. In other words, objects are clustered based on the principle of maximizing the intra-class similarity while minimizing the inter-class similarity. Sequential pattern and time-series mining example, clustering techniques can be used to identify stable dependencies for risk management and investment management.

Hilage and Kulkarni (2012) In this paper authors has reviewed the literature of data mining techniques such as Apriori Algorithm, Association Rules, Decision tree, and Rule Induction Technique and Neural network. This review of literature focuses on how data mining techniques are used for different application areas for finding out meaningful pattern from the database.

Dang and Mehta (2012) In the present research work authors focused on dataset for Bombay stock exchange (BSE), India. This dataset shows five attributes as date, open, high, low and close. Authors have used date and open attribute has been taken care off for determining the sequential pattern. There were 252 tuples ranging from January 2011 to December 2011. Where date states the temporal dataset, open shows the open price of the shares in stock market. A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. To carry on this research work trend analysis procedure has been opted. This procedure is used to ascertain the sequential pattern among the temporal datasets. It is the child procedure of time series analysis. In it statistical data is identified and characterized; to retrieve interesting sequential pattern in temporal data sets. The procedure of trend analysis has been applied in Minitab 15.0; which is well known data mining tool used for the purpose of forecasting and prediction.

Raheja and Kumar (2012) In this paper authors have focused on clustering technique of data mining in the distributed database. Without clustering computation time will always be high and so to reduce the computation time clustering technique has been applied. The computational time of traditional approach with proposed approach will vary as the number of Databases increases and thus showing major difference between traditional approach and proposed approach. The major problem with frequent set mining methods is the explosion of the number of results and so it is difficult to find the most interesting frequent item sets and so the concept of locally frequent item sets has been highlighted in this paper.
Kumar et. al (2012) The main idea of the authors of proposed new method is to develop a simple, but yet powerful, that captures the data stream content in a Window by using sliding-window technique to find regular items. The experimental results show the effectiveness of VDSRP-method in finding regular patterns in a data stream. In this paper authors presented a VDSRP method which is much better than the existing RSP-tree Algorithm because it uses sliding-window technique and the advantages of vertical database Format. This method is very simple to use with simple operations like arrays, unions, intersection, Deletion etc. To find out regular patterns over data streams. Authors experiment results perform well in both execution and memory consumption.

Kambey et. al (2012) Authors have analyzed the different method used in stock market and found that Application Of Association Rules in Stock Markets The Enormous amount of valuable data generated by stock Market has attracted researchers to explore this Problem domain using different methodologies. Data mining application have various challenges which are: Scientific research that relates to creation of Knowledge from stock market data set, better stock price prediction that concerns With the purchasing and sale of the items. To develop feasible and efficient methods For finding the useful patterns and future Trends, To utilize the capital resources of the Investors, To boost the economy, To create the interests in the favor of the Stock market, To protect investors and investment, To maintain market stability, To check out the all fraud and illegal trade Practices.

Argiddi and Apte (2012) In this paper Author use fragment based mining approach which focuses on minimizing the length of the transaction table of the stock market, based on some common features among the attributes which indirectly minimize the complexity involved in processing. Fragment based mining deals mainly with reducing the time and space complexity involved in processing the data. The result of fragment based mining promising one. Fragment based mining can potentially be used for predictions and recommendations stock trading platforms. In this research author have taken the original data sets of Bombay stock exchange (BSE) of different companies such as Infosys, tcs, and oracle etc from yahoo finance and try to find the association among the large scale it companies and small scale it companies. Fragment based mining algorithm gets accurate results with less time and space complexity as compared to fiti algorithm.

Nandagopal et. al (2012) In this paper authors break the barrier of transactions and extend the scope of Mining association rules from traditional single-dimensional, intra transaction associations to n dimensional, Inter-transaction associations. Mining inter-transaction Associations pose more challenges on efficient processing than mining intra-transaction associations because the number of potential association rules becomes extremely large after the boundary of Transactions is broken. Authors conducted tests using the data set collected from different Stock exchange (se),various experimental results are reported by comparing with real life and Synthetic data sets .This study introduce The notion of n-dimensional inter-transaction association rule, define its measurements: support And confidence and develop an efficient algorithm called modified Apriori.
**Maheshwari (2012)** In this paper author apply association rule for mining the data which helps to remove the redundant rule and helps in compact representation of rules for user. In this paper authors, optimized algorithm has been proposed for online rule generation. The advantage of this algorithm is that the graph generated in our algorithm has less edge as compared to the lattice used in the existing algorithm. This algorithm generated all the essential rules and no rule is missing. The use of non redundant association rules help significantly in the reduction of irrelevant noise in the data mining process. The advantage of this algorithm is that the graph generated in our algorithm has less edge as compared to the lattice used in the existing algorithm.

**Kaur and Mangat (2012)** Authors in this paper describe the various application of data mining in stock market. They told that 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. 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).

**Karthik et. al (2012)** In this paper technique of integrated clustering and Weighted Rule Mining is described, to find the stock market trend with index dependency hybrid clustering and association algorithm is not appropriate. Statistic analysis techniques are not suitable for trend analysis with index relationship. The stock market transactions data is analyzed with clustering and weighted rule mining techniques. The K-means clustering algorithm is used to cluster the transaction with respect to the market flows. The market trade transactions are divided into three zones such as up trend, down trend and stable zone. The weighted rule mining technique is applied to fetch patterns from the indexes, sector indexes and company price values. Apriori algorithm is modified to carry out weighted rule mining process. The system produces the market trade trend flow with market indexes and sector index values.

**Kumar et. al (2012)** Main objective of authors in this paper is to compare various classification algorithms that have been frequently used in data mining for decision support systems. Three decision trees based algorithms, one artificial neural network, one statistical, one support vector machines with and without ad boost and one clustering algorithm are tested and compared on four datasets from different domains in terms of predictive accuracy, error rate, classification index, comprehensibility and training time. Experimental results demonstrate that Genetic Algorithm (GA) and support vector machines based algorithms are better in terms of predictive accuracy. Genetic algorithm is better in terms of comprehensibility as it is independent of data set size. K-Mean, neural network learning and logistic regression have no comprehensibility. Decision Tree (QUEST) is better in term of comprehensibility as like genetic algorithm, it is also independent of dataset size. QUEST is better in terms of classification index. K-means, Neural network, logistic regression.

**Nancy and Ramani (2012)** In this paper authors begin with the research uses only ten Universities and highlights the formation of association rules between the attributes or variables and explores the association rule between a course and gender, and discovers the influence of gender in studying a course. They attempt to cover the main algorithms used for clustering, with
a brief and simple description of each. The previous research with this dataset has applied only regression models and this is the first time to apply association rules. The Classical Algorithm for Mining Association Rule named Apriori is used in the dataset. The original Dataset undergoes various tasks like conversion to excel format, discretization of data to make it suitable for applying the data to Apriori Algorithm. Knowledge patterns regarding the association between the major (course) and gender were identified.

**Kumar and Valli (2013)** They introduce RFPID-algorithm to mine regular frequent Patterns using vertical data format which performs better with large number of transactions and long Item sets with one database scan. The other advantage of vertical data format is, it uses simple operations like Unions, intersections, deletions, simple arrays etc., and also it judges non frequent and non regular item sets before generating the candidate sets.

**Das and Uddin (2013)** Author state that Prediction in any field is a complicated, challenging and daunting process. Employing traditional methods may not ensure the reliability of the prediction. In this paper, authors are reviewing the possibility of applying two well-known techniques neural network and data mining in stock market prediction. As neural network is able to extract useful information from a huge data set and data mining is also able to predict future trends and behaviors. Therefore, a combination of both these techniques could make the prediction much reliable. That data mining and neural network is very effective techniques to deal with unpredictable data like stock data or any other data that needs prediction.

**Prasanna and Ezhilmaran (2013)** In this paper authors have discussed several attempts made by Researches for stock price prediction. These works show that data mining techniques can be applied for evaluation of past stock prices and acquire valuable information by estimating suitable financial indicators. There are two types of prediction methods are implemented by several researches to generate useful extracts. They are fundamental approaches and technical indicator based approaches. This stock data is highly volatile and unpredictable which makes data mining and AI techniques as suitable once. They suggest that Back propagation algorithm for training and suitable AI technique applied on some fundamental approaches may render promising results.