LITERATURE REVIEW

Raju P. S. et. al. (2014) Data mining and Customer Relationship Management is needed in Banking and Retail Industry. This paper includes various tasks and applications of data mining useful in these industries. The bank and retail industry realizes that data mining is useful process for decision making and gives advantages in competitive environment in the future.

Vidhate D. R. (2014) This paper concentrates on different observations of researchers for the development of super bazaar with consumer response. In this, knowledge mining is used to analyze customer buying behavior in super bazaar.

Lalithdevi B. et. al. (2013) This paper explains about Data Mining on Web log data. In this, activities of web usage mining and technologies used in each task explain in detail. Data preparation and pattern generation techniques are elaborated. That is useful in finding navigation patterns. At last concludes that many mining algorithms use the sequential pattern generation method and rest use ad hoc methods.

Bhaise R. B. et. al. (2013) Education Data Mining is the main theme of this paper, for better student’s education. For this, the author used K-Means or Clustering techniques on the sample data. This technique is used to analyze the data from different dimensions and categorize the data. They made clusters according to the student’s performance in the examination. The information generated after implementing mining technique is very much useful for teacher as well as student.

Borkar S. and Rjeswari K. (2013) Association rule mining is useful to evaluate student performance in the study. In this paper, for data analysis Weka tool is used. The main goal of this paper is predicting student performance in the university exam on basis of the criteria internal exam, assignment, attendance, etc. This paper concluded that the result of university result will be improved of the poor students by giving extra efforts in their unit test, attendance, assignment and graduation.

Chaurasia V. and Pal S. (2013) This paper gives the survey information of different data mining techniques for medical person for decision making. From this, the doctors can predict the presence of heart disease. This paper used Naive Bayes, J48 Decision Tree, Bagging techniques in the field of heart disease diagnosis. As a result, the bagging algorithm is better from others because it gives human readable classification rules.
Dr. Dhenakaran S. S. and Kumar M. S. (2013) In this paper, the authors concentrated on the cluster analysis technique. For that they collected the data of rural village of Sivagangai District like population, literacy rate, sex ratio, child population, etc. They used two methods for clustering that are Principal Component Analysis and K-means Clustering. In which, they found that the PCA method is better to reduce the dataset from high dimension data to low dimension.

Joshi A. and Kaur R. (2013) Clustering is an unsupervised learning technique that finds data from unlabelled collection. It is used in data preprocessing step of data mining which will be useful in other algorithms. This paper gives the comparative analysis for the various clustering techniques like K-Means, hierarchical, DBSCAN, OPTICS, STING.

Dr. Chitra K. and Subashini B. (2013) The author includes in this paper is various data mining techniques and its application in banking like fraud prevention and detection, acquiring new customers, customer retention, automatic credit card approval, provide segment based products, marketing and risk management. The paper gives the listing of algorithms of supervised and unsupervised learning. So through mining environment, decision making process becomes very fast.

Kaur J. and Pandya J. (2013) Data Archeology means automatic finding of unknown patterns of data. It also means that automatic prediction of the behavior. It gives the new way to marketers and analyst to find unknowing facts in the business. After discovering this, the paper also gives the outline for application of data mining and stages & techniques of knowledge mining. So it was concluded that for finding useful and unknown knowledge from large amount of data stored in some data source, data mining is the best technique.

Nadaf M. and Kadam V. (2013) Telecommunication data needs data mining to gain some important knowledge regarding customers. Main data mining processes in these fields are Customer Segmentation, Profiling, Data Preparation and Clustering. Data mining in telecom industry reduces the much more human efforts.

Ramageri B. M. and Dr. Desai B. L. (2013) The paper explains about how data mining can be useful in retail industry. This paper reviews the research about data mining and its applications form past to present and gives outline for how it will be useful for attracting profitable customers. For this, the authors follow the phases like business understanding, data understanding, data preparation, modeling, evaluation and employment.
Rathee A. and Mathur R. P. (2013) In this paper, the study of education database is done, which contain hidden knowledge for improving students performance. This paper gives the comparative study of decision tree algorithms like ID3, C4.5 and CART and as a result C4.5 is more accurate. The predictions obtained from the algorithms help the teacher to identify poor students and improve their performance.

Singh A. and Rana A. (2013) Cluster Analysis helps to categorize the customers in Automobile Industry according to their needs, which is known as customer segmentation. Customer retention is also done using data mining. For this, the authors used two-step clustering and K-means clustering techniques. Data is collected by taking different attributes and found correlation between different attributes. This all can be helpful to make passive customers to active customers.

Pulakkazhy S. et. al. (2013) This paper reveals about data mining techniques used in banking sector. Using this techniques the decision making will become easy and fastest. Banks are using data mining in marketing, credit risk management, detection of money laundering, liquidity management, investment banking and fraud detection. The detected patterns help the bank to predict future events.

Agarwal B. B. and Tayal S. P. (2012) This book give some basic guidance about what is data mining, how it works, its evolution, etc. It includes the explanation about different data mining methods that are predictive which includes classification, regression, time series analysis, prediction and descriptive which includes clustering, association rules, sequential patter discovery, summarization. It also gives advantages and disadvantages about data mining.

Dharker S. and Rajavat A. (2012) By classification algorithms, the authors implement healthy diet recommendation system through web data mining. They compare two decision tree algorithms that are ID3 and c4.5 on the basis of accuracy and time performance analysis. Using these algorithms, they find the user access pattern. Then derived the result as ID3 algorithm is worked on each and every instance very accurately. C4.5 takes less time but will not work on each instance.

Hooda R. and Gill N. S. (2012) The paper gives the concept about data warehousing and data mining. Data warehouse is a central data storage that stores data from various resources.
Data mining means finding unknown facts from these data. This paper gives the outline about the history of these concepts.

**Dr. North M. (2012)** Data preparation is the most important phase in the data mining. It is the starting phase of it. In that, anomalies will be removed through data scrubbing. Rapid Miner is the tool that is used to handle missing data. Data reduction is done for missing values. This is known as filtering. Handling of inconsistent data and irrelevant data is also needed.

**Raval K. M. (2012)** This paper explains about different data mining techniques that includes Association, Classification, Clustering, Prediction and Sequential Patterns, etc. These techniques are useful for finding patterns to predict future trends in business.

**Suriya S. et. al. (2012)** Association rule mining means taking items without knowing the relations between them and find some important rule between them. The authors explained Apriori algorithm with support and confidence key terms. They reviewed new things regarding association rule mining like w-support, alarm association rules, association matrix mining algorithm, data mining for tax inspection and traffic management, etc. This brief analysis concluded that association rules used with any application give effective results.

**Mahendra T. V. et. al. (2012)** This paper concentrates on data mining algorithm for cloud data with sector/sphere framework and association rules. Cloud means a structure that incorporates various data centers and treats as single point of access for the people on the internet. Sector and sphere are for the application that computes the data as small as possible.

**Tiwari M. et. al. (2012)** This research is on using data mining techniques for retailers to predict customer behavior and provide better satisfaction level. Then find out accuracy of different data mining algorithms on different type of data. For this testing, the tool used is Weka.

**Dr. Kumar V. and Chadha A. (2012)** The fundamental thing in this paper is to use data mining methodologies to analyze student performance in their study. In this, association rules are used to compare student performance at graduate and post graduate level and predict their success or failure. For association rules finding, the authors use the Tanagra tool. After rules finding, it reveals various factors like student interest, curriculum design, teaching method which can affect student.
Vohra G. and Dr. Bhushan B. (2012) The focus is on analysis of life insurance data using data mining techniques. The factor analysis technique is used on one region data by asking questionnaire. Reliability of this data is tested by Cronbach’s alpha. And adequacy is tested by Kaiser-meyer-oklin and bartlett’s test of sphiricity. Then correlation between variables is found. The principal component analysis has been done and found that most of the factors affect the customer’s behavior.

Ramamohan Y. et. al. (2012) Data mining tools are used for prediction in the business and make fruitful decisions. This paper gives the brief overview for different data mining tools like Weka, Tanagra, Rapid Miner, DBMiner, Witness Miner, Orange.

Deshpande A. and Halarnkar P. N. (2011) I studied about the topics Concept Description and Association rule mining from this book. The authors explain about Data Generalization which abstracts a large set of task relevant data in a database from a low conceptual level to higher one. It has methods categorized in Data Cube approach (OLAP) and Attribute oriented approach (AOI). In Association Rule mining, they explain about market basket analysis which based on the theory that if you buy a certain group of items, you are more likely to buy another group of items. For this methods used are Apriori algorithm and Frequent Pattern Tree.

Dr. Dhanabhakyam M. and Dr. Punithavalli M. (2011) Market Basket Analysis is the best example of Association rule mining in which we find out association of customer buying patterns from the items that customers put in their basket. This paper does the work for existing data mining algorithm for market basket analysis. In which, Apriori algorithm with some modification, is very much useful to reduce time complexity and improve accuracy.

Prasad P. and Dr. Malik L. G. (2011) Clustering technique is explained in this paper, to segment customer profile for retail store. Retail data mining is used to identify the customer buying patterns and behaviors which will be helpful in improving customer service and maintaining customer retention. This paper describes clustering techniques of k-means and expectation maximization and how they applied in customer segmentation. In this, weka tool is used to generate clusters.

Olaniyi A. S. et. al. (2011) In this, the study of regression analysis is presented for the stock market data for share price prediction. Data is collected from Nigerian Stock Exchange. The
founded patterns will help the stock brokers and investors for making difficult decisions and it will reduce the risk in the stock market.

Chauhan R. et. al. (2010) This explains the use of data mining for promotional selling of luxury items through e-retailing. For this, apply clustering and data visualization technique for predictive analysis in marketing. For clustering, the authors use k-means clustering algorithm. In this paper, use the data in .arff file format which is supported by weka tool. The results show that online traffic, website stickiness and online sales turnover is the major factor that effecting e-retailing.

Mudimigh A. et. al. (2009) This paper is on extracting knowledge from customer’s data through data mining techniques and improving business performance. The authors studied about CRM with Data Mining. Methodology follows the steps: Customer Inquiry, Clusters of customers, Rule Induction Engine, Customer understanding, Action by Organization.

Hamilton A. et. al. (2008) This invention provides a method for gaining access to more complete information about the customer and targeted to derive direct benefit from the sale. This system gives highly correlated data. These centralized data gives benefits to the retailers, seller and customers.

Bramer M. (2007) Clustering means grouping together objects that are similar to each other or dissimilar to the objects belonging to other clusters. The object represented either in 2-D space or 3-D space according to attributes. This book also explains about finding centroid for cluster. The K-means clustering and hierarchical clustering methods are explained in detail for clustering.

Han J. and Kamber M. (2006) The book is presented from database perspective. Various knowledge discovery process like data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation have been discussed. Here, I have the answers for, on which kind of data, data mining can be applied? And what kinds of patterns can be mined? The classification of data mining that is database systems statistics, machine learning, visualization and information science is discussed.

Saida H. (2006) In this invention, the data is accessed from relational database and then factor analysis function is performed on the data to create factor loading matrix. Factor analysis is a technique that uncovers various factors regarding customer purchasing behavior. In that, customer transaction data are analyzed for affinity groups and customer destination
segments. Affinity groups indicate the frequency with which various products are purchased both together and separately. The customer destination segments are uncovered by clustering tool.

**Busche F. D. and Marcotte D. (2003)** The aim of this invention is to find the locations of products within a retail space using a position identifying system. Data mining algorithms are used to generate input data for forming a set of spatial relationships. Spatial relationships of products allow one to enhance product purchase by the relative location of products to another. Using data mining algorithms, association of products, classifications of behaviors and prediction was done. So, design a new store layout that will helpful in purchasing more products in short period of time.

**Bohr M. B. and Cunningham S. W. (2002)** In this invention, a computer implemented data mining system analyzes data using Gaussian Mixture Model. For that, methods like Ad hoc querying, Online Analytical Processing, Statistical Packages were used. This invention allows to gain better understanding of customer behavior by cluster analysis. Gaussian Mixture models are the form of Clustering. The data for clustering consists of customer transaction i.e. baskets. These clusters are understood by marketing or merchandising decision makers and used for promotional campaign and predicting future transaction.

**Shaw M. J. et. al. (2001)** Proper methodology for data mining and knowledge management is useful for managing marketing knowledge and supporting marketing decisions. Different data mining tasks are incorporated like dependency analysis, class identification, concept description, deviation detection, data visualization. The knowledge discovery is a process that extends the collections of data mining tasks into knowledge management.

**Chou P. B. et. al. (2000)** This invention concentrates on the selection of customer and market data using data mining techniques for targeted marketing. Here, market segmentation has been done using Tree Induction and CLIQUE clustering algorithm. From these two, Tree induction is preferred because of non overlapping partitions. In the next step, cumulative statistics is also calculated for the customer’s data and the market data.