A STUDY ON PREDICTIVE DATA MINING ALGORITHMS

M.Shanmugapriya,
M.Phil Research Scholar,
P.G & Research Department of Computer Science,
Siri PSG Arts and Science College for Women,
Sankari,Tamilnadu,India.

K.Sudha,
Assistant Professor,
P.G & Research Department of Computer Science,
Siri PSG Arts and Science College for Women,
Sankari,Tamilnadu,India.

Abstract: Discovery of a predictive learning function that classifies data item into one of several predefined classes. Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. The algorithm then encodes these parameters into a model called a classifier, Here Discussed classification algorithms Bayesian Classification, Support Vector Machines (SVM), K-Nearest Neighbor Classifies, Genetic Algorithm, Artificial Neural Networks, Decision Tree, Naïve Bayes,ID3,C4.5.

Keywords: Datamining, Prediction , K-Nearest Neighbor Classifies, Genetic Algorithm, Artificial Neural Networks, Decision Tree, Naïve Bayes,ID3,C4.5.

I.INTRODUCTION

Data mining is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules. The overall purpose is to support decision making by turning collected data into actionable information. Using this perspective, data mining is the key activity in a larger process called knowledge discovery in databases (KDD). A generic description of KDD is given in Figure 1.

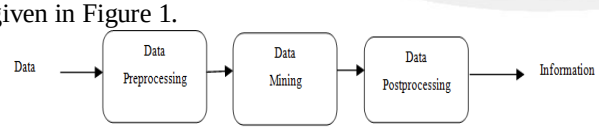

Figure 1: The KDD process.

The steps in the KDD process contain [1]:

- **Data cleaning:** It is also known as data cleansing; in this phase noise data andirrelevant data are removed from the collection.
- **Data integration:** In this stage, multiple data sources, often heterogeneous, arecombined in a common source.
- **Data selection:** The data relevant to the analysis is decided on and retrieved from thedata collection.
- **Data transformation:** It is also known as data consolidation; in this phase theselected data is transformed into forms appropriate for the mining procedure.
- **Data mining:** It is the crucial step in which clever techniques are applied to extractpotentially useful patterns.
- **Pattern evaluation:** In this step, interesting patterns representing knowledge areidentified based on given measures.
- **Knowledge representation:** It is the final phase in which the discovered knowledgeis visually presented to the user. This essential step uses visualization techniques tohelp users understand and interpret the data mining results.

Data can come from several different sources and in a variety of formats. The purpose of preprocessing is to transform the input into an appropriate form for data mining. Preprocessing typically involves steps like fusing data from multiple sources, selecting the data relevant for the mining task and cleaning data; e.g. handling missing values and outliers. The output of preprocessing is a standard data matrix; i.e. a vector of objects (tuples or instances) where each instance is a set of attributes values. If there are n instances and each instance has attributes, the standard data matrix thus has n rows and p columns. Data mining uses the preprocessed data to produce models, typically used for either description or prediction. The purpose of the post processing step is to make sure that only valid and useful data mining results are actually used. Post processing often includes activities like hypothesis testing and visualization.

The process of transforming data into information is, in practice, always context dependent; i.e. KDD is at all times performed in a specific situation and with a specific purpose. Although most research focuses on data mining techniques (the technical context), it is important to realize that for executives the ultimate goal isto add business value. This viewpoint is sometimes referred to as the business context of data mining. KDD, as described above, is actually part of a larger process called the virtuous cycle of data mining [1]; see Figure 2. In the virtuous cycle, KDD represents the activity transform data into actionable information using data mining

techniques. To exploit the full potential of the techniques, data mining must be part of company's strategy; i.e. data mining could typically be considered as part of customer relationship management.

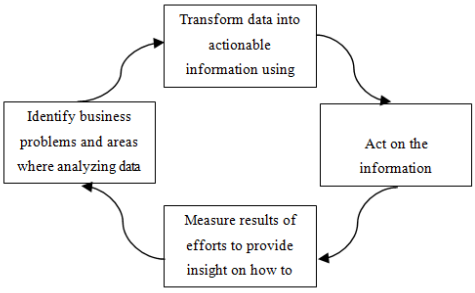

Figure 2: Virtuous cycle of data mining (adopted from [1]).
Whether the term data mining should be used for the entire virtuous cycle, the transform data into actionable information activity or just as one part of the KDD process depends on the abstraction level. In this thesis, data mining mainly refers to applying different data mining techniques, but the business context and its demands are also recognized. More specifically, the fact that results from data mining techniques ultimately should be used by human decision-makers places some demands on the data mining models. Since different techniques produce different kinds of models, these demands will in fact often determine which data mining technique to use.

II.DATA MINING TYPES

- **Predictive data mining**: The purpose of Predictive mining model is mainly to predict the future outcome than current behavior. The prediction output can be numeric value or in categorized form. The predictive models is the supervised learning functions which predicts the target value. Predictive methods to be utilized in the next phase of empirical study. Use some variables to predict unknown or future values of other variables in addition comparison between these supervised approaches was also conducted to get some insight about the strength and weaknesses of each approach since one of the aims of the study was to determine whether these methods were well suited for extracting the required knowledge. As a result, the predictive method will be able to predict in which cluster does the future student will fall into based on the enrollment information.
- **Descriptive Data Mining**: The second approach for mining data from large datasets is known as Descriptive data mining. It is normally used to generate correlation, frequency, cross tabulation, etc. This Descriptive method can be defined as to discover regularities in the data and to uncover patterns. This is also used to find interesting subgroups in the bulk of data. Some researchers used Descriptive to determine the demographic influence on some particular factors. Descriptive approaches employed in the study were identified, namely the Summarization and Clustering methods. Since the aim of the experiment was to study the patterns and getting some information within the enrollment data, no specific target has been identified. To this end, clusters were generated by Clustering method, and later used as target or output for Predictive approach.

III. LITERATURE REVIEW

Z. Huang, *et al.* [8] applied predictive analytics techniques to establish a decision support system for complex network operation management and help operators predict potential network failures and adapt the network in response to adverse situations. The resultant decision support system enables continuous monitoring of network performance and turns large amounts of data into actionable information.

R. M. Riensche, *et al.* [9] described a methodology and architecture to support the development of games in a predictive analytics context, designed to gather input knowledge, calculate results of complex predictive technical and social models, and explore those results in an engaging fashion.

Esther Ge, *et al.* [10]a number of possible data mining methods that can be applied to do the lifetime prediction of metallic components and how different sources of service life information could be integrated to form the basis of the lifetime prediction model. Researcher compare a number of data mining methods on the data sets provided by our industry partners and analyze what kind of methods are suitable for what kind of data. Firstly, traditional data mining methods like Naïve Bayes, Decision Tree, Neural Network, SVM and M5 etc were applied to build a number of independent predictors for each data set. The results indicate that the best method for predicting the service life depends on the data set used to train the model. Also analyzed the predicted service life from each data set using certain test cases. The testing shows that in some situations, inconsistent predicted results may be presented by the data mining systems due to using three different data sets (information) for a same test case.

Andrew Kusiak, *et al.* [11] have used data pre-processing, data transformations, and data mining approach to elicit knowledge about the interaction between many of measured parameters and patient survival. Two different data mining algorithms were employed for extracting knowledge in the form of decision rules. Those rules were used by a decision-making algorithm, which predicts survival of new unseen patients. Important parameters identified by data mining were interpreted for their medical significance. They have introduced a new concept in their research work, it has been applied and tested using collected data at four dialysis sites. The approach presented in their paper reduced the cost and effort of selecting patients for clinical studies. Patients can be

chosen based on the prediction results and the most significant parameters discovered.

IV.PREDICTIVE ANALYTICS TECHNIQUES

Predictive models analyze identify patterns in historical and transactional data to determine various risks and opportunities. Forecasting models capture relationships between many factors to allow assessment of the risks or potential associated with particular set of conditions, guiding decision making for candidate transactions. Three basic techniques for Predictive analytics are Data profiling and Transformations, Sequential Pattern Analysis and Time Series Tracking [5]. Data profiling and transformations are functions that change the row and column attributes and analyses dependencies, data formats, merge fields, aggregate records, and make rows and columns[4]. Sequential pattern analysis identifies relationships between the rows of data. Sequential pattern analysis involves identifying frequently observed sequential occurrence of items across ordered transactions over time. Time Series Tracking is an ordered sequence of values at variable time intervals at the same distance [4]. Time series analysis gives the fact that the data points taken over time.

There are some advanced Predictive analytic techniques like Classification-Regression, Association analysis, Time series forecasting to name a few. Classification uses attributes in data to assign an object to a predefined class or predict the value of numeric variable of interest [4]. Regression analysis is statistical tool for the study of relations between variables. Association analysis describes significant associations between data elements. Time series analysis is employed for forecasting the future value of a measure based on past values [5].

V. PREDICTION MODEL

Data mining methods are applied to all three data sets to build three predictors first. After that, these three predictors can make predictions for user's inputs. The final predicted life is either a multiple choice provided by three predictors or a value combined from the outputs of three predictors.

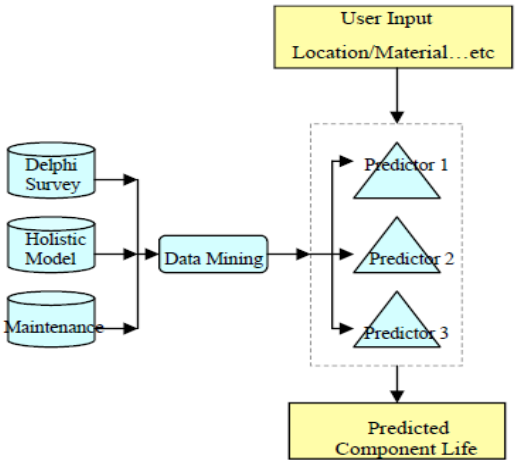

Figure 3: Overflow of Prediction Model

In order to get accurate predicted results, applied various data mining methods including Naïve Bayes, KNearest Neighbors (KNN), Decision Tree (DT), Neural Network (NN), Support Vector Machine (SVM) and M5Model Trees on these data sets. Naïve Bayes is statistical-based algorithm. It is useful in predicting the probability that a sample belongs to a particular class or grouping (Fayyad, Piatetsky-Shapiro, & Smyth, 1995a).KNN is based on the use of distance measures. Both DT and NN are very popular methods in data mining. DT is easy to understand and better in classification problems while NN cannot produce comprehensible models in general and is more efficient for predicting numerical target. Support vector machine is relatively newmethod. It can solve the problem of efficient learning froma limited training set.

VI.PREDICTIVE DATA MINING (PDM) ALGORITHMS TO COMPARE

A) DECISION TREE (DT) ALGORITHM

Decision tree is a predictive data mining techniques often used in clinical medicine to easily visualize, and understand resistant to noise in data. And is applicable in both regression and association data mining tasks [9] capable of handling continuous attributes, which are essential in case of medical data e.g. blood pressure, temperature, etc. It is a non-parametric supervised learning method used for classification to create models that predicts the value of a target variable by learning simple decision rules inferred from the data features [10]. Decision trees are significantly faster than neural networks with a shorter learning curve that is mainly used in the classification and prediction to represent knowledge. The instances are classified by sorting them down the tree from the root node to some leaf node. The nodes are branched based on if-then condition [11]. The variants of decision tree algorithm includes CART, ID3, C4.5, SLIQ, and SPRINT [12]. C4.5 algorithms is an extension of ID3 algorithms that uses concept of divide-and-conquer to construct a tree from a training set [4], and each path in the decision tree can be regarded as a decision rule [14]. The decision tree is built of nodes which specify conditional attributes – symptoms ,$S=\{S_1,S_2,\ldots S_i\}$ branches which show the values of $V_{i,k}$ i.e. the h-th range for i-th symptom and leaves which present decisions $D=\{d_1,d_2,\ldots d_i\}$ and their binary values,$W_{dk}=\{0,1\}$. A sample decision tree is presented in Fig 4, and as a set of association rules in equation.

$$(S_1,v_{1,1})\cap(S_2,v_{2,1})\Rightarrow(d_1=1)$$

$$(S_1,v_{1,1})\cap(S_2,v_{2,2})\Rightarrow(d_1=0)$$

$$(S_1,v_{1,2})\cap(S_3,v_{3,1})\Rightarrow(d_2=1)$$

$$(S_1,v_{1,2})\cap(S_3,v_{3,2})\Rightarrow(d_2=0)$$

$$(S_1,v_{1,3})\Rightarrow(d_3=1)$$

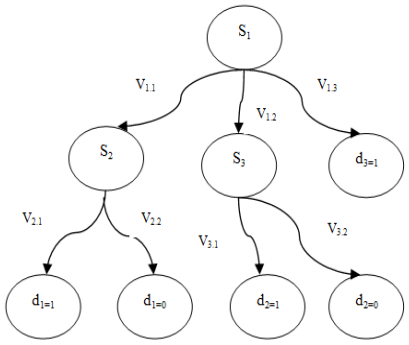

Figure 4:Sample decision tree applicable in clinical medicine

B) ARTIFICIAL NEURAL NETWORKS (ANN'S) ALGORITHM

The main function of Artificial Neural Networks is prediction. Despite its complexity and difficulty in understanding the predictions, it has been successfully applied in clinical medicine especially in prediction of coronary artery disease, EEG signals processing and the development of novel antidepressants [9]. ANN is biologically inspired, highly sophisticated analytical techniques, capable of modelling extremely complex non-linear functions [15], that are modelled based on the cognitive learning process and the neurological functions of the human brain, consisting of millions of neurons interconnected by synapses [16], and capable of predicting new observations after learning from existing data. Neural networks are also capable to predict changes and events in the system after the process of learning [11]. The interconnected sets of neurons are divided into three: input, hidden, and output ones. In clinical medicine, the patient's symptoms could serve as input set S, and disease couldserve as output set D to the neural network. The hidden neuron processes the outcome of preceding layers.

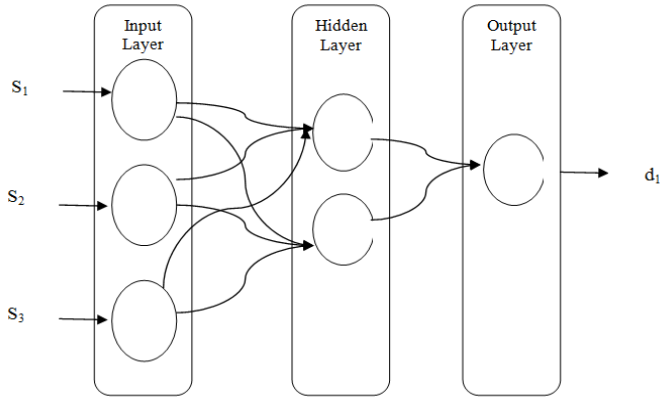

Figure 5: Graphical depiction of ANN for clinical diagnosis for symptoms
(S_1,S_2,S_3) like age, blood pressure, etc. and diagnosis (d_1)

The attributes that are passed as input to the network form the first layer, and the direction of the network symbolizes the dataflow during the process of prediction. The process of learning in ANN is to solve a task T , having a set of observations and a class of functions F, which is to f*€ F find as the optimal solution to the task [9]. The most popular of the ANN being Multilayer Perceptron algorithm. MLP is most suitable for approximating a classification function, and consists of a set of sensory elements that make up the input layer, one or more hidden layers of processing elements, and the output layer of the processing elements [13]. The Multi-Layer Perceptron (MLP) with back-propagation (a supervised learning algorithm) is arguably the most commonly used and well-studied ANN architecture capable of learning arbitrarily complex nonlinear functions to arbitrary accuracy levels [8]. It is essentially the collection of nonlinear neurons (perceptron's) organized and connected to each other in a feed forward multi-layer structure as shown in Fig [5], defines ANN based on: a) Interconnection pattern between different layers of neurons; b) Learning process for updating the weights of the interconnection; and c) Activation function that converts a neuron's weighted input to its output activation.

C) NAÏVE BAYES ALGORITHM

Naïve Bayes is a Bayesian Network based on Bayes rules of simple conditional probability that estimate the likelihood of a property given the set of input data. Naïve Bayes require small amount of training data to estimate parameters such as mean and variance necessary for classification . The structure of a Naïve Bayes model forms a Bayesian network of nodes with one node for each attribute. The nodes are interconnected with directed edges and form a directed acyclic graph [9] Bayesian networks uses directed acyclic graphs to model the dependencies among variables. Each node in the acyclic directed graph represents a stochastic variable and arcs represent a probabilistic dependency between a node and its parents [14]. The Bayesian Network is a way of representing probabilistic relationships between variables associated with an outcome of interest. The outcome of interest could be uncertainty essential during clinical diagnosis, prediction of patient's prognosis and treatment selection. The probabilities applied in the Naïve Bayes algorithm are calculated according to the Bayes' Rule. The probability of the likelihood of some conclusion S,given some evidence or observation T, where ∃ a dependence relationship between S and T, denoted as P(S\T), can be calculated based on Equation;

$$P(S \mid T) = \frac{P(T \mid S) * P(S)}{P(T)}$$

This conditional relationship allows an investigator to gain probability information about either Sor T with the known outcome of the other.

D) ID3 ALGORITHM

In decision tree learning, ID3 (Iterative Dichotomiser 3) is an algorithm invented by RossQuinlan used to generate a decision tree from the dataset. ID3 is typically used in the machinelearning and natural language processing domains. The decision tree technique involvesconstructing a tree to model the classification process. Once a tree is built, it is applied to eachtuple in the database and results in classification for that tuple. The following issues are faced bymost decision tree algorithms [2]:

- Choosing splitting attributes
- Ordering of splitting attributes
- Number of splits to take
- Balance of tree structure and pruning
- Stopping criteria

The ID3 algorithm is a classification algorithm based on Information Entropy, its basic idea isthat all examples are mapped to different categories according to different values of the conditionattribute set; its core is to determine the best classification attribute form condition attribute sets.The algorithm chooses information gain as attribute selection criteria; usually the attribute thathas the highest information gain is selected as the splitting attribute of current node, in order tomake information entropy that the divided subsets need smallest [4]. According to the differentvalues of the attribute, branches can be established, and the process above is recursively called on each branch to create other nodes and branches until all the samples in a branch belong to thesame category. To select the splitting attributes, the concepts of Entropy and Information Gain areused.

Entropy

Given probabilities p1, p2…,ps, where Σpi = 1, Entropy is defined as

$$H(p1, p2\text{…….},ps) = \Sigma - (pi \log pi)$$

Entropy finds the amount of order in a given database state. A value of H = 0 identifies aperfectly classified set. In other words, the higher the entropy, the higher the potential to improvethe classification process.

Information Gain

ID3 chooses the splitting attribute with the highest gain in information, where gain is defined asdifference between how much information is needed after the split. This is calculated bydetermining the differences between the entropies of the original dataset and the weighted sum ofthe entropies from each of the subdivided datasets. The formula used for this purpose is:

$$G(D, S) = H(D) - \Sigma P(Di) H(Di)$$

E) C4.5 ALGORITHM

C4.5 is a well-known algorithm used to generate a decision trees. It is an extension of the ID3algorithm used to overcome its disadvantages. The decision trees generated by the C4.5 algorithmcan be used for classification, and for this reason, C4.5 is also referred to as a statistical classifier.The C4.5 algorithm made a number of changes to improve ID3 algorithm [2]. Some of these are:

- Handling training data with missing values of attributes
- Handling differing cost attributes
- Pruning the decision tree after its creation
- Handling attributes with discrete and continuous values

Let the training data be a set S = s_1, s_2 ... of already classified samples. Each sample S_i = x_l, x_2... isa vector where x_l, x_2 ... represent attributes or features of the sample. The training data is a vectorC = c_1, c_2..., where c_1, c_2... represent the class to which each sample belongs to.At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits data setoff samples S into subsets that can be one class or the other [5]. It is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. Theattribute factor with the highest normalized information gain is considered to make the decision.The C4.5 algorithm then continues on the smaller sub-lists having next highest normalized information gain.

F) GENETIC ALGORITHM

In the computer science field of artificial intelligence, a genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover. GAs is inspired by Darwin"s Theory about Evolution "Survival of Fittest". GAs is adaptive heuristic search based on the evolutionary ideas of natural selection and genetics. Genetic algorithms find application in bioinformatics, phylogenetics, computational science, engineering, economics, chemistry, manufacturing, mathematics, physics and other fields [3]. GAs simulate the survival of the fittest among individuals over consecutive generation for solving a problem. Each generation consists of a population of character strings that are analogous to the chromosome that see in DNA. Each individual represents a point in a search space and a possible solution.

The individuals in the population are then made to go through a process of evolution. The GA maintains a population of n chromosomes (solutions) with associated fitness values. Parents are selected to mate, on the basis of their fitness, producing offspring via a reproductive plan. Consequently highly fit solutions are given more opportunities to reproduce, so that offspring inherit characteristics from each parent. As parents mate and produce offspring, room must be made for the new arrivals since the population is kept at a static size. Individuals in the population die and are replaced by the new solutions, eventually creating a new generation once all mating opportunities in the old population have been exhausted. In this way it is hoped that over successive generations better solutions will thrive while the least fit solutions die out. New generations of solutions are produced containing, on average, more good genes than a typical solution in a previous generation. Eventually, once the

population has converged and is not producing offspring noticeably different from those in previous generations, the algorithm itself is said to have converged to a set of solutions to the problem at hand.

Advantages of GA

- It can solve every optimisation problem which can be described with the chromosome encoding.
- It solves problems with multiple solutions.
- Since the genetic algorithm execution technique is not dependent on the error surface, can solve multi-dimensional, non-differential, non-continuous, and even non-parametrical problems.
- Structural genetic algorithm gives us the possibility to solve the solution structure and solution parameter problems at the same time by means of genetic algorithm.
- Genetic algorithm is a method which is very easy to understand and it practically does not demand the knowledge of mathematics.
- Genetic algorithms are easily transferred to existing simulations and models.

Disadvantages of GA

- Certain optimisation problems (they are called variant problems) cannot be solved by means of genetic algorithms. This occurs due to poorly known fitness functions which generate bad chromosome blocks in spite of the fact that only good chromosome blocks cross-over.
- There is no absolute assurance that a genetic algorithm will find a global optimum. It happens very often when the populations have a lot of subjects.
- Like other artificial intelligence techniques, the genetic algorithm cannot assure constant optimisation response times.
- It is unreasonable to use genetic algorithms for on-line controls in real systems without testing them first on a simulation model.

G) SUPPORT VECTOR MACHINE (SVM) ALGORITHM

In contemporary applications of machine learning, support vector machine (SVM) is known as one of the most precise and most powerful methods amongst well-known algorithms. Support vector machine is one of the methods of learning with supervisor that is used for classification and regression. This method is among relatively new methods, which in recent years has shown a good efficiency compared to older methods for classification including Multi-Layer Perception neural networks. The basis of SVM is to find a linear boundary amongst classes of data objects. This effort is made to select that line or hyper plane which has a wider safety margin. The solution of equation is finding the optimum line for data through optimization methods that are well-known methods in solving problems subject to constraint. If classes are not linearly separable, data should be taken to a very higher dimensional space so that the machine can classify highly complicated data. In order to solve the problem of very high dimensions using this method, Lagrange theorem is used to transform the intended minimization problem to its binary form that in which instead of the complicated functions, a simpler function named kernel function is used. This algorithm has powerful and flawless theoretical principles but needs dozen samples to be responsive to problems with number of dimensions.

In addition, efficient methods for SVM learning are growing rapidly. In a learning process having two classes, the aim of SVM is finding the best function for classification so that in data collection, the members of the two classes can be recognized. For data collections that are linearly resolvable, the scale of the best classification is determined geometrically. From sensory aspect, that margin which is defined as part of the space or the same parting between two classes is defined by hyperplane. Geometrically, margin corresponds the least distance between nearest data and a point on hyperplane. This geometric definition authorizes us to find how to maximize margins, although have incalculable hyperplanes and just a few deserve the solution for SVM. The reason that SVM emphasizes on the biggest marginfor hyperplane is that this matter better provides thegeneralization capacity of the algorithm. This not onlyhelps the efficiency of the classification and its precisionon training data but also prepares the ground for betterclassification of the future data.One of the problems of SVM is its calculative complexity.Nevertheless, this problem has been solved admissibly.One solution is that a big optimization problem is dividedto some smaller problems that each problem includes acouple of precisely selected variables that the problem canutilize them effectively. This process will be continueduntil all these segmented parts are solved.

VI.CONCLUSION

In this paper discussed about data mining methods for prediction techniques. In order to get accurate predicted, discussed various data mining methods including Naïve Bayes, KNearest Neighbors (KNN), Decision Tree (DT), NeuralNetwork (NN), Support Vector Machine (SVM) and M5Model Trees on these data sets.

REFERENCE

[1]. M. J. A. Berry and G. Linoff, Data Mining Techniques: For Marketing, Sales and Customer Support, Wiley, 1997.
[2]. M. J. A. Berry and G. Linoff, Mastering Data Mining – The Art and Science of Customer Relationship Management, Wiley, 2000.
[3]. M. Dunham, Data Mining – Introductory and Advanced Topics, Prentice Hall, 2003.
[4]. http://analyticsweb.com/predictive-analytics.
[5]. http://www.articlesbase.com/strategic-planning-articles/predictiveanalytics-1704860.html
[6]. M Zaman, Predictive analytics; the future of business intelligencewww.mahmoudyoussef.com

[7]. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.
[8]. Quinlan, J. R. (1992). *Learning with Continuous Classes.* Paper presented at the 5th Australian Joint Conference on Artificial Intelligence.
[9]. K. Aftarczuk, "Evaluation of selected data mining algorithms implemented in Medical Decision Support Systems," Blekinge, 2007.
[10]. B.Venkatalakshmi and M. Shivsankar, "Heart Disease Diagnosis Using Predictive Data mining," in 2014 IEEE International Conference on Innovations in Engineering and Technology (ICIET'14), Tamil Nadu, India, 2014.
[11]. B. Milovic and M. Milovic, "Prediction and Decision Making in Health Care using Data Mining," International Journal of Public Health Science (IJPHS), vol. 1, no. 2, pp. 69-78, December 2012.
[12]. M. Hemalatha and S. Megala, "Mining Techniques in Heath Care: A Survey of Immunization," Journal of Theoritical and Applied Informattion Technology, vol. 25, no. 2, pp. 65-70, 2011.
[13]. E. Osmanbegović and M. Suljić, "Data Mining Approach For Predicting Student Performance," Economic Review – Journal of Economics and Business, vol. X, no. 1 , pp. 3-12, May 2012.
[14]. R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: Current issues and guidelines," International Journal of Medical Informatics, vol. 77, no. 2, pp. 81-97, February 2008.
[15]. A. S. Fathima, D. Manimegalai and N. Hundewale, "A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the Arbovirus-Dengue," International Journal of Computer Science (IJCSI), vol. 8, no. 6, pp. 322-328, November 2011.
[16]. S. Gupta, D. Kumar and A. Sharma, "Data mining classification techniques applied for breast cancer diagnosis and prognosis," Indian Journal of Computer Science and Engineering, vol. 2, no. 2, pp. 188-193, 2011.
[17]. J. Nalavade, M. Gavali, N. Gohil and S. Jamale, "Impelling Heart Attack Prediction System using Data Mining and Artificial Neural Network," International Journal of Current Engineering and Technology, vol. 4, no. 3, pp. 1-5, June 2014.
[18]. B. K. Bhardwaj and S. Pal, "Data Mining: A prediction for performance mprovement using classification," International Journal of Computer Science and Information Security (IJCSIS), vol. 9, no. 4, pp. 1-5, 2011.
[19]. S. Floyd, "Data Mining Techniques for Prognosis in Pancreatic Cancer," 2007.
[20]. K. Srinivas, B. K. Rani and A. Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of HeartAttacks," International Journal on Computer Science and Engineering (IJCSE), vol. 2, no. 2, pp. 250-255, 2010.

