(IJIASE) 2016, Vol. No. 2, Jan-Dec **e-ISSN: 2454-9258, p-ISSN: 2454-809X**

APPLICABILITY OF MACHINE LEARNING TECHNIQUE IN DEVELOPING A STUDENT ASSESMENT AND PERFORMANCE FRAMEWORK

Gautam Anand

Student, Sanskriti School, Chanakyapuri

ABSTRACT

Data mining is an emerging technology that is used in each and every system. Education data mining is a beneficial discipline because the amount of data in education system is increasing day by day. Its usage in higher education is relatively new but its importance increases because of expanding database There are many approaches for measuring students' performance ,data mining is one of them .With the help of data mining the hidden information in the database gets out which helps in improvement of students performance . Education data mining is used to study the data available in education field to bring the hidden data, i.e. essential and useful information from it . There are many methods of data mining used to analyse a student's performance classification, method like decision tree is most used to measure the students' performance. With the help of these, it is easy to improve the result and future of students. More methods like clustering, regression, time series, and neural network can also be applied.

I. INTRODUCTION

Higher education is essential for a student's life. Higher education institutes are focused on an analysis of every object because of private participation. Data mining techniques are applied in many fields like marketing, medicine, fraud detection, web, engineering etc. The main aim of data mining is to know hidden knowledge.DM provides various methods of analysis; these include classification,association,k-means,decision tree, regression, time series, neural network, etc. Application of data mining in the educational system directly helps in an analysis of participants in the education system. The students also recommend many activities and tasks[1].Data mining is also used to show how students use material of a particular course. In teaching environment, trainer can obtain feedback on students [5].

II. RELATED WORK

EdinOsmanbegovic [1]-In these paper supervised data mining algorithm were applied.A different method of data mining was compared. The data were collected from the survey conducted during the summer semester at the University of Tuzla. Many variables like Gender,GPA, Scholarships, High school performance ,Entrance Exam Grades,etc. are taken for the performance.Naive Bayesalgorithm, Multilayer Perceptron, J48issued. The result indicates that the naïve Bayes classifier outperforms in the predication decision tree and neural network method. These will help the student for the future.

Qasem A.Al-Radelideh [2]-The title of the paper is "Mining student data using a decision tree." They use the data mining process for students' performance in university courses to help the higher education management.Many factors affect the performance.They use a classification technique for building a reliable classification model; the CRISP-DM (cross-industry standard process for data mining) is adopted. These methods consist of five steps, i.e. collecting the relevant features of the problem, Preparing the data, Building the classification model, Evaluating the model and finally, future prediction. The data were collected in the table in the proper format; the classification model was built using the decision tree method. Many rules were applied. The WEKA toolkit is used. Different classification methods were used like ID3,C4.5 and naïve Bayes and accuracy were in the table as a result. J.K. JothiKalpana[3]-"Intellectual performance analysis of students by using data mining techniques "This paper focuses on the prediction of school in different levels such as primary,secondary,higher level. Clustering method such as centroid based distribution based and density-based clustering is used. The data were collected from Villupuram College. These methods were used for improving the performance as the students.

Cristobal Romero [4]-"Educational data mining;A Review of state of the art." EDM, i.e. educational data mining, is an emerging discipline.EDM process converts raw data coming from the educational system into useful information.DM techniques are used, i.e. association rule mining for selecting weak students.Several classification algorithms were applied in order to group students.EDM tools were designed for educators. Romero [5]-"Educational data mining survey from 1995 to 2005 "There is also web-based education in the computer-aided instructions in the specific location.Web-based training is so popular now a days that predication of its level also becomes useful. Data processing is done to transform the original data into a suitable shape. Web mining is there to extract knowledge from the web. Clustering,classification is used. In these, it says that the prediction of performance in e-learning is also so important.

S.Kotsiantis[6]-"Predicting students' performance in distance learning using machine learning techniques" Many universities are giving distance learning education, so predicting the performance of students in that becomes so important.A machine learning algorithm is so useful for many types of learning tasks — this paper Uses ML techniques to predict students' performance in distance learning system.Set of rules are planned.A decision tree is used; ANN is also inductive learning based on computational models.Set of the attribute are taken and divided into groups.There is an ANOVA test result.It showed that the best algorithm is naïve Bayes with 66.49% accuracy in the data is taken.

Pooja Thakar [7]-"Performance analysis and prediction in educational data mining: A Research Travelogues' "Lots of data is collected in scholarly databases.In order to get benefits from such big data tools are required.University produces lots of students, and its performance prediction is essential. Set of weak students are taken, and predication with data mining techniques is used.This paper says that many models are required for instruction.

V.Shanmugarajeshwari [8]- "Analysis of students' performance evaluation using classification techniques "The author used the classification techniques for predication of student's performance

in the education system.The data is collected, and preparation is done for pre-precessing the checking. It calculates the entropy, Info Gain, Ratio then the information gain for evaluating of these.Classification technique is used. A decision tree is built and finally gain ratio is estimated.

Michael A [11]-These researches have applied the decision tree for predicting a student's final GPA.It used WEKA toolkit. It collects the data from CS College at king save university in the year 2012 was obtained from the institute.Each student record with different attributes.Student name,student id, final GPA,the semester of graduation etc.It is essential to improve the final GPA of the student.

Ben Daniel [13]-It applied big data analysis in higher education.KDD is an interdisciplinary area focusing on the method for identifying and extracting a pattern from large data sets.Big data help provide insight to support students learning needs.

TimmyDevasaia[18]-It used classification technique to predict student performance. A naive theorem is used various information like group action, class test,semester and assignment marks were collected from the student's previous knowledge to predict the performance of the student.

Ryan S.J.D.Baker[19]-"The state of educational data mining in 2009: A review and future vision" In these papers author reviews the trend in 2009 in the field of educational data mining. The year 2009 finds research communising of EDM and these moments in EDM bring unique opportunity.EDM categories in web mining, Statistics and Visualization,Clustering,Relationship mining, i.e. Association rule mining and causal data mining. There are many applications of EDM.These papers discuss the EDM.

Pooja M.Dhekankar[22]-"Analysis of student performance by using data mining concept "Data mining technique is used in many areas and in the educational field it becomes so crucial for future of the students. Students classification is done on the basis of students marks.Association rule,clustering outlier detection,classification is discussed in this paper.

Amjad Abu saa[28]-It applies c4.5, CART, ID5 algorithm for analysis of students performance. It takes various parameters for accuracy. The decision tree is built and based on it, student performance is predicted. Naive Bayes classification is also applied, which assumes that all given attributes in a dataset are independent. It creates different quantities predictive model by using different data mining tasks which are useful to predict students'grades.Various decision tree algorithms were implemented.Finally, we can say that it helps the university as well as students.

YoavBergner [24] et.al-It used collaborative filtering analysis of student data.There is logistic regression as collaborative filtering.There is parameter estimation. There is a simulated skill response. It applied the numerical method for analysing students response matrix with the goal of predicting response; it showed it naturally parameterizes series of models and multidimensional IRT.

(IJIASE) 2016, Vol. No. 2, Jan-Dec **e-ISSN: 2454-9258, p-ISSN: 2454-809X**

III. MACHINE LEARNING

It is the branch of science that works with the system in such a way that they automatically learn. It means that recognizing and understanding the input data and moving decision on the support data. The name machine learning was come in 1959 by Arthur Samuel. They evolved from the study of pattern, AI, computational theory. Machine learning constructs the algorithm that can learn and make predictions. Machine learning closely related to statistics which help in forecasting. It is tough to take the division for their problem and algorithm is developed. Their algorithm is based like statistics , logic etc.

Types of Machine Learning: - Supervised Learning: - In this there is desired input with the desired output. In addition to taking feedback about the accuracy of prediction. It can apply what is learned in the past to the new set of data using a suitable example to feedback about future events. There are known training data sets and starting from the analysis; the learning calculation creates a capacity to make a forecast about the yield. The framework can give focus on any new input. The learning calculation can likewise look at its return and discover blunder so as to alter the model.

Unsupervised Learning: - In these, we do not have any target to predict. It is used for clustering in different groups. It is utilized when the data used to prepare is neither ordered nor unravelled. It ponders how frameworks information would be able to capacity to depict concealed structure from unlabelled data. It models positive information and can draw data from informational index to portray covered structure.

Semi-Supervised M.L.: - It is between supervised and unsupervised learning. It means it uses both labelled and unlabelled data for training. It can be said that it uses a small amount of labelled data and a large amount of unlabelled data. The system there is for learning accuracy.

Types ofSupervised Learning Algorithms

Logistic Regression.

Decision Tree.

Support vector machine.

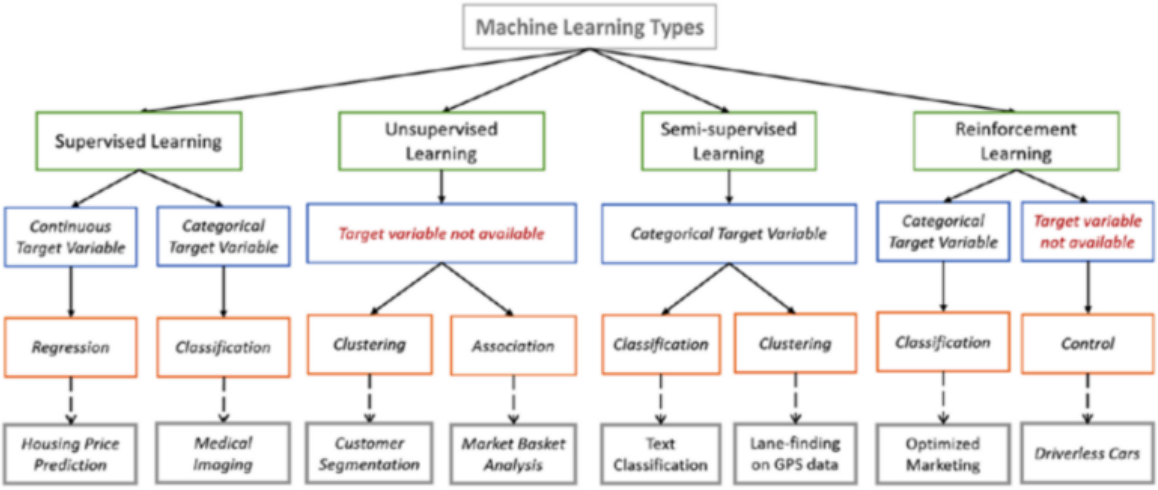
K-Nearest neighbours.

Naive Bayes.

Linear Regression.

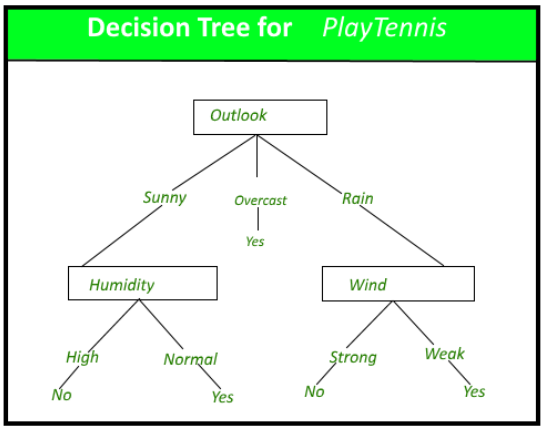
Types of Unsupervised Learning: - K - Mean clustering. Hierarchical clustering. I have a hidden Markov model. Data Mining-Data mining is the process of discovering patterns in extensive data set involving the methods at the intersection of machine learning, and the system. With the help of data mining tools, the prediction can be made quickly, there are significant data sets, and the pattern is identifying and establishing h a relationship to solve the problem through data analysis. It means

analyzing hidden trends of data according to different perspectives for categorization into useful information.

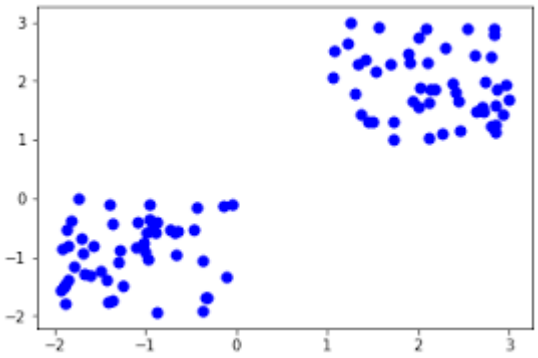


KNN: -K-nearest neighbours is the classification algorithm. On pattern recognition, it is the non-parametric method used for both classification and regression. In both cases, the input consists of the K-closest training example in the feature space. The output depends on whether k-nearest. It is to simple even with its simplicity; it gives highly competitive results. KNN has different names is K-nearest neighbour.

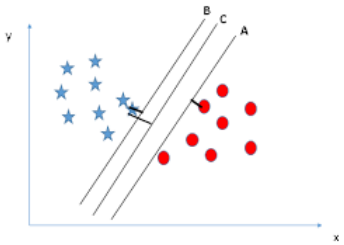
Memory – Based Neighbours Decision TreeAlgorithm-It is flow chart like structure. The tree consists of the root node branches and the leaf nodes. The internal node in the tree denotes test on an attribute. Each chapter indicates the outcome of the trial and the class label. Tree pruning is performed in order to remove anomalies in the training data due to noise. Classification technique-It is a method in which data is classified into different classes. The classification analysis is used to receive useful information from data or Metadata.In classification techniques, we will apply the algorithm to organize the data. Association Rule-It refers to the method that can help to identify the pattern in the database .These techniques are used to determine the hidden pattern in the data that can be used to identify variables within the data that frequently appear in the dataset .This rule is mostly used in analyzing customer behaviour. Clustering Analysis-cluster is a collection of similar data within a group.It means that the data of the same type are kept in within a group and others are in another group.Clustering analysis is the analysis of discovering groups and cluster in the data in such a way that the degree of association is the highest.



K-Means Clustering:-It is a definite method that is mostly used for cluster analysis. K – Mean clustering aim to partition n observation into k-clusters in which each view belongs to the cluster with the nearest mean.



SVM:-Support vector machine is a supervised learning algorithm; it is a classification concept. When we have a dataset with features and class; labels both then we can use support vector machine, but if a dataset does not have the output, it is considered as an unsupervised algorithm. There are two types of SVM- Linear Nonlinear in the linear model the training example are plotted in space. In the nonlinear model, our dataset is generally dispersed up to some expert.



(IJIASE) 2016, Vol. No. 2, Jan-Dec **e-ISSN: 2454-9258, p-ISSN: 2454-809X**

IV CONCLUSION

Machine learning is an emerging technology that is used everywhere nowadays in the bank, labs, telecom, industries. Data mining is part of it which helps in prediction; the future forecast is significant in many places . Many algorithms are built, and more and more research is going on. Every technology uses the concept of it. We survey many papers for the prediction of students' performance. Students performance is so essential for their future; it not only helps the students but also helps teachers, institutions and parents. Many prominent institutes use the concept of AI for prediction.