International Journal for Rapid Research in Engineering Technology & Applied Science
Vol VII Issue VII July-2021
ISSN (Online): 2455-4723

Prediction & Analysis Students Performance Using EDM over Decision Tree: Machine Learning Theory

Monica*, Prof. Mohit Jain**
BM College , RGTU Bhopal, Indore, 452001,India*
BM College , RGTU Bhopal, Indore, 452001,India**
*monikameravi@gmail.com *, bmctmohitcs@gmail.com***

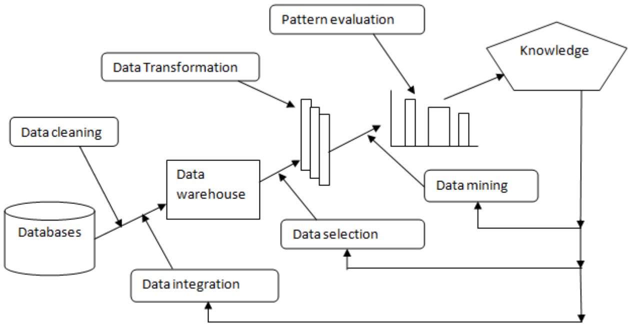
Abstract:-Machine Learning is very emerging technology that is used in each and every system. Education data mining is very important disciplines, because the result of student is so important for their future and the amount of data in education system is increasing day by day .In education it is relatively new but its importance increases because of increasing database. There are many approach for measuring students' performance .data mining is one of them .With the help of data mining the hidden information in the database is get out which help for improvement of students' performance education data mining is used to study the data available in education field to bring the hidden data i.e. important and useful information from it . There are many methods of machine learning which is used to analysis of students' performance clustering method like K-means 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 classification, regression, time series, and neural network can also be applied.

Key words: -Data mining, EDM, K-means, Decision Tree, Students data.

I. INTRODUCTION

Education is very important for student's life. Higher education institute are focus on analysis of every objects because of private participation. Data mining techniques is applied in many fields like marketing, medicine, fraud detection,web, and engineering and in the field if education it becomes more useful.The main aim of data mining is to know hidden information from large set of data. There are many method of machine learning which is used to predict students' performance. Clustering is a method which is more efficient as compare to other method, and K-means is one of them. Data mining provides various methods analysis; these include classification, association, k-means,decision tree, regression, time series, neural network,etc.

Application of data mining in the educational system is directly help to analysis of participants in the education system. The students also recommend many activities and task [2].Data mining is also used to show how students use material ofparticular course. In teaching environment trainer are able to obtain feedback on students [6]



Data mining Process

II. RELATED WORK

FatmaChiheb[1]-In these paper it discuss the case of Algerian university. They used the decision tree method to predict the performance of students. They used j48 algorithm for it.They applied j48 to weka to the data and obtain the decision tree.They follow the CRISP-DM method.They have the data set and for prediction they used the previous result for e.g. the result of student in five semestersand six semesters is based on the one, two three, four semester result.

Md.Hedayetul Islam Shovon[2]-Data clusteringtechniques i.e. K-means is applied. Data clustering is process of extracting unknown, hidden patterns from large data set. In this model they use internal and external assessment for prediction. This model helps to weak students to identify their score before the exam. Graph shows the relationship between GPA and attendance and also number of students and percentage of student regarding to GPA. And from it they show the percentage of students getting high, medium, and low gpa.

M.Durairaj[3]-Educational details and performance is based upon various factors like personal details,social etc. WEKA toolkit is used they collect the data set of college students real time data that describe the relationship between learning behavior of students and their academic performance, the data set contain students detail of different subject marks in semester which is subjected to the data mining process. In these K-means clustering is used and from the total number of 300 student record dataset, they choose 38 students record for our analysis .The confusion matrix is there to shows pass, fail, and absence for the exam. They compare the weighted average for decision tree and naviebayes techniques.

Mr.Shashikantpradipborgavakar[4]-Here the data clustering is used as k-means clustering to evaluate students' performance. Their performance is evaluated on the basic of class test,mid test,and final test.In their model they measured by internal and external assessment, in which they tale class test marks,lab performance, quiz etc. and final grade of students is predicted They generate the graph which shows the percentage of students getting high, medium, lowgpa.

EdinOsmanbegovic[5]-In these paper supervised data mining algorithm were applied.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 variable like Gender,GPA,Scholarships,High school, Entrance Exam,Grade,etc. are taken for the performance.Naive Bayesalgorithm, multilayer Perceptron, J48issued. Theresult indicates that the naïve Bayes classifier outperforms in predication decision tree and neural network method. These will help the student for future.

Qasem A.Al-Radelideh[6]-The title of the paper is "Mining student data using decision tree".They use data mining process for student performance in university courses to help the higher education management.Many factorsaffect the performance.They use classification technique for building the reliable classification model,the CRISP-DM (cross-industry standard process for data mining) is adopted .These method 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 table in proper format, the classification model were building 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 result.

J.K. JothiKalpana[7]-"Intellectual performance analysis of students by using data mining techniques "This paper focus on the prediction of school in different level such as primary,secondary,higher level. Clustering method such as centroid based distribution based and density based clustering are used. The data were collected from Villupuram College. There method used for improving the performance as the students.

Cristobal Romero[8]-"Educational data mining;A Review of the state of the art".EDM i.e. educational data mining is emerging discipline.EDM process converts raw data coming from 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[9]-"Educational data mining survey from 1995 to 2005"There is also web-based education in the computer aided instruction in the specific location.Web based education is so popular now a days that predication its level is also become useful.Data processing is done for transform the original data into suitable shape. Web mining is there for extract knowledge from the web. Clustering, classification is used.In these it says that the predication of performance in e-learning is also so important.

S.Kotsiantis[10]-"Predicating students' performance in distance learning using machine learning techniques" Many university are giving distance learning education so predicating performance of students in that become so important.Machine learning algorithm is so effective for many types of learning tasks.This paper Use ML techniques to predict students' performance in distance learning system.Set of rules are planned.Decision tree are used,ANN is also inductive learning based on computational models.Set of attribute are taken and divided into groups.There is ANOVA test result.It showed that best algorithm is naïve Bayes with 66.49% accuracy in the data it taken.

PoojaThakar[11]-"Performance analysis and prediction in education data mining: A Research Travelogues'"Lots of data is collected in educational databases.In order to get benefits from such big data tools are required.University produces lots of students and its performance predication is important.Set of weak students are taken and predication with data mining techniques is used.This paper says that many models are required for an instruction.

V.Shanmugarajeshwari[12]-"Analysis of students' performance evaluation using classificationtechniques "The author used the classification techniques for predication of student's performance in education system.The data is collected and preparation is done the preprocessing for checking. It calculates the entropy, Info Gain, Ratio then the information gain for evaluating of these.Classification technique is used.Decision tree is build and finally gain ratio is evaluated.

MashaelA[13]-These researches has applied decision tree for predicting students final GPA.It used WEKA toolkit .It collect the data from C.s. College at king save university in the year 2012 were collected from the institute.Each student record with different attributes, Student name,student id, final GPA,semester of graduation etc.It is important to improve the final GPA of the student.

Ben Daniel[14]-It applied big data analysis in higher education.KDD is an interdisciplinary area focusing on method for identifying and extracting pattern from large data sets.Big data help provide insight to support students learning needs.

TismyDevasaia[15]-It used classification technique to predict the student performance .Naive theorem is used various information like group action,class text,semester and assignment marks were collected from the students previous information to predict performance of the student.

Ryan S.J.D.Baker[16]-"The state of educational data mining in 2009:A review and future vision"In these paper author review the trend in 2009 in field of educational data mining. The year 2009 finds research communizing of EDM and these moment 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 application of edm.These papers discuss about the EDM.

IIIMACHINELEARNING

It is the branch of science that works with the system in such 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 prediction. It is very difficult to take the division for their problem and algorithm is developed. There algorithm are based like statics logic etc.

Application of machine learning:-

- Vision processing.
- Language processing.
- Forecasting.
- Pattern recognition.
- Games.
- Data mining.
- Robotics.
- Expert system.

Types of Machine Learning:-

Supervised Learning:-

In there is desired input with desired output. In addition to take feedback about the accuracy of predication. It can be apply what are learned in the past to the new set of data using the suitable example to feedback future events.

There are known training data set and starting from the analysis; the learning algorithm produces as function to make prediction about the output. The system is able to provide targets from any new input. The learning algorithm can also compare its output and find error in order to modify the model.

Unsupervised Learning:-

In these we do not have any target to predict. It is used for clustering in different groups. It is used when the information used to train isneither classified nor solved.

It studies how systems can information function to describe hidden structure from unlabeled data. It examples positive data and can draw information from data set to describe hidden structure.

Semi-Supervised M.L.:-

It is between supervised and unsupervised learning. It means it used both labeled and unlabeled data for training. It can be said that it used small amount of labeled data and the large amount of unlabeled data. The system in there is for learning accuracy.

IV EXISTING SYSTEM

There are huge amount of data produced in educational system. These can be exploited in order to extract the useful knowledge. In today's system lots of technique is used to predict the students' performance.There is a case of Algerian university in which student's performance is predicting using decision tree. Decision tree is build using the J48 algorithm.There is the huge amount of data in the educational system in the existing system they predict the performance on the basic of previous semester result.The J48 algorithm is used which is very hard to build because of its splitting. Weka toolkit is used and crisp-dm model is applied.

V PROBLEM STATEMENT

There is problem that, there is huge amount of data in the educational system, For predicting the students' performance there should be method which is more efficient and produced useful result. Decision tree is a classification technique which is less efficient as compare to clustering techniques J48 is a decision tree algorithm which is used for predicting student performance but it is less efficient as compare to k-means clustering techniques.Decision trees examine only a single field at a time, leading to rectangular classification boxes. This may not correspond well with the actual distribution of records in the decision space.

Disadvantages of decision trees:

- They are unstable, meaning that a small change in the data can lead to a large change in the structure of the optimal decision tree.
- They are often relatively inaccurate. Many other predictors perform better with similar data. This can be remedied by replacing a single decision tree with a random forest of decision trees, but a random forest is not as easy to interpret as a single decision tree.
- For data including categorical variables with different number of levels, information gain in decision trees is biased in favor of those attributes with more levels.
- Calculations can get very complex, particularly if many values are uncertain and/or if many outcomes are linked.

VI PROPOSED WORK

Prediction of students' performance can be done using Machine Learning algorithm. Clustering is a technology in which there is cluster with group of similar data. K means algorithm is used to predict the performance of students.K means is a unsupervised machine learning algorithm. K means clustering set the partition of n observations into k clusters in which each observations belongs to cluster with nearest mean.Cluster is measured with the mean value of the objects in a cluster, which can be viewed as the cluster centroid. The idea is to define K centers and one for each cluster. These centers should be placed I proper way because different location give different result.So better choice is to place them far away from each other.The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more.

K means Algorithm-

Step1-Accept the number of cluster to group data and the dataset to cluster as input values.

Step2-Inilize the first K cluster(Choose random K element)

Step 3-Calculate the arithmetic mean of each cluster formed in the dataset.

Step 4-K mean assign each record in the dataset to only one of the initial cluster (the nearest cluster using a distance measure).

Step 5-K means re-assigns each record in the dataset to the most similar cluster and recalculate the mean of the entire cluster in the dataset.

Advantages of K-means algorithm-

- ➢ Easy to implement.
- ➢ With a large number of variables, K-Means may be computationally faster than hierarchical clustering (if K is small).
- ➢ K-Means may produce higher clusters than hierarchical clustering.
- ➢ An instance can change cluster (move to another cluster) when the centroids are recomputed.

VIICONCLUSION

Machine learning is very emerging technology that every placed it used. Now days in bank, labs, telecom, industrial each and every place machine learning is used. Data mining is part of it which helps in prediction, future prediction is very important in many place which help so much. Many algorithm is build and more and more research is going on every technology used the concept of it. We survey many papers for prediction of students' performance .On comparing decision tree and k means it is seen that

k means is more efficient as compare to decision tree.Students performance is so important for their future it not only help student but also help teachers institute parents. Many big institutes used the concept of AI for prediction.

REFERENCES-

1. Fatmachiheb,FatimaBoumahdi- Predicting students' performance using Decision trees: Case of an Algerian University.2017 International conference on Mathematics and information technology, Adrar, Algeria –Dec 4-5,2017.

2. Md.Hedayetul Islam shovon, HanfuzaHaque-Prediction of students' academic performance by an application of K-means clustering algorithm. International journal of advanced research in computer science and software engineering, Volume 2,Issue 7 July 2012.

3. M.Durairaj-Educational data mining for prediction of students' performance using clustering algorithm,M.durairaj et.al (IJCSIT) International journal of computer science and information technologies vol.5(4),2014.

4. Mr.ShashikantpradipBorgavakar-Evaluating students' performance using K means clustering, International journal of engineering Research and technology(IJERT) vol.6 issue 05 May 2017.

5. EdinOsmanbegovic, MirzaSuljic -Data mining approach for predicting student performance Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012.

6. Qasem A. Al-Radaideh, Emad Al-Shawakfa - Mining Student Data Using Decision Trees, Research Gate Article 2006.

7. J.K. JothiKalpana, K. Venkatalakshmi- "Intellectual performance analysis of students by using data mining techniques", International Journal of innovative Research in science, Engineering and technology. Volume 3, Special issue 3, March 2014.

8. Cristobal Romero, Member, IEEE, Sebastián Ventura, Senior Member, IEEE "Educational data mining: A review of the state of the art" Page 21 Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews.

9. C. Romero *, S. Ventura "Educational data mining: A survey from 1995 to 2005" Science Direct 2006.

10. S.Kotsiantis, C.Pierrakers, P.pintelas-"predicting students' performance in distance learning using machine learning techniques" Taylor & Machine Group. Applied A.I. 18:411-426, 2004.

11. PoojaThakar, Anil Mehta, Manisha -"Performance Analysis and prediction in educational Data mining: A research travelogue" International Journal of Computer Applications (0975 – 8887) Volume 110 – No. 15, January 2015.

12. V. Shanmugarajeshwari, R. Lawrence "Analysis of student's performance Evaluation using Classification Techniques" 978-1-4673-8437-7/16 IEEE 2016.

13. Mashael A. Al-Barrak and Muna Al-Razgan "Predicating students final GPA using decision tree: A case study "International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016.

14. Ben Daniel "Big Data and analytics in higher education: Opportunities and challenges "British Journal of Educational Technology (2014).

15. Ms.TismyDevasia, Ms.Vinushree T P, Mr.VinayakHegde "Prediction of student's performance using educational data mining".

16. Ryan S.J.D. BakerKalinaYacef "The state of educational data mining in 2009 A review and future vision" Journal of Educational Data Mining, Article 1, Vol 1, No 1, Fall 2009.