

A Novel Machine Learning Algorithm for Students Performance Analysis

Ayushi Mehra, Dr S. R. Patra

Department of Computer Science and Engineering
Bhopal Institute of Technology & Science, Bhopal (M.P.)

Abstract— Machine Learning is a field that is used in every system. Machine learning is used in the educational system, in pattern recognition, Games, Industries. In the education system, its importance becomes more because of the future of the students. Education data mining is very useful because the amount of data in the education system increases day by day. Higher education is relatively new, but its importance increases because of the increasing database. There are many approaches to measuring students' performance. K-means is one of the most efficient and used methods. With the help of data mining, the hidden information in the database is getting out, which helps improve students' performance. The decision tree is also a method used to predict the students' performance. Presently, the main problems that educational institutions face are with the growth of data and the need to use this data to enhance the quality. One of the basic techniques often used in analyzing data sets is clustering. This study makes use of cluster analysis to segment students into groups according to their characteristics. An unsupervised algorithm like K-means is discussed. Education data mining is used to study the data available in the education field to bring hidden data, i.e., important and useful information. With the help of these, it is easy to improve the result and future of students.

Introduction

Clustering methods in machine learning have been applied in many applications such as fraud detection, banking, academic performance and instruction detection. Developments in data have added great challenges to educational institutions and enable institutions to make better-informed decisions. Clustering is a common technique often applied to data analysis. Higher education institute is focused on the Analysis of every object because of private participation. Machine Learning provides various methods: classification, association, k-means, decision tree, regression, time series, neural network, etc. Application of data mining in the educational system directly helps to Analysis of participants in the education system. The students also recommend many activities and task. Many variables stand in the way of students receiving high grades. The faculty can target these variables in designing methods to enhance students' learning and academic performance by evaluating student records and graphs and assessing student performance. Data mining is also used to show how students use material of a particular course. In teaching, the environment trainer can obtain feedback on students.

Related Work

Fatma Chiheb[1]- Decision tree approach has been used in this research paper. The decision tree is developed using J48 Algorithm. The classification is performed using Weka and CRISP-DM. They collect data about graduate and postgraduate students of the companies. It is a Muslim university. This data has been collected from the department of computer science. They have tested the decision tree, which will advise on the best input and output. Degrees are known as attributes of students' success allocated.

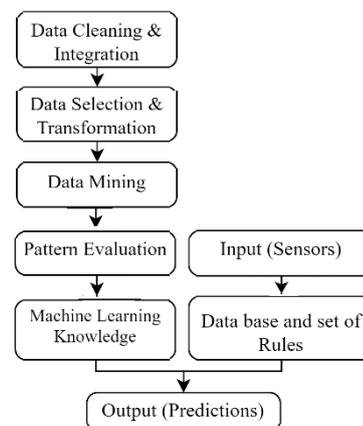


Fig.1: Different stages of the data mining process

Shanmugam Rajeshwari[2]- They score the students based on their results. The input data is collected from Ayya Nadar Janaki Ammal College, Sivakasi, Tamil Nadu. However, methods for feature selection, the number of methods is discussed. Training data is applied to a given data set to create a classifier model. Machine learning was used to assess students' ability to succeed.

M. Durairaj[3]- Educational information and success are determined by factors such as gender, diet, occupation, and others. Weka toolkit is used to collect the data set of college students' real-time data describing their learning behavior and academic success; the data set includes students' detail of different subject marks in the semester, subject to the data mining process. Using K-means clustering, the number of students at this university is divided into 38. The uncertainty matrix is there to indicate pass, fail, and absence for the test. They compare the usefulness of the decision tree and the naive Bayes models.

Mr Shashikant Pradip borgavakar[4]- The clustering is used to compare the performance of students. Their success will be measured based on the midterm exam, final exam, and graduation exam. In their model, they used internal and external assessment data, and they

created a graph which shows the percentage of students who get high, medium, and low scores for school grades.

Edin Osman Begovic[5]-In these paper supervised data mining algorithm was 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, Entrance Exam, Grade, etc., are taken for the performance. Naive Bayes algorithm, Multilayer Perceptron, J48 issued. 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.

E.venkatasanet.al[6]-In this article, the clustering and classification algorithm were compared using matrix laboratory software, for the initial data WEKA software is utilized. Data set of students were picked up from private arts and science colleges from Chennai city. Near about 573 students are there in the database. In the details, they take the internal exam and end semester exam details. An algorithm such as J48 was used allows the input attribute to get a classification model. Matrix Laboratory is used for measuring the operational of several data mining algorithm. There is a table for error measure.

A. Seetha ram Nagesh[7]-Prediction of students' performance is so important. Still, if it is predicted at an early stage, it becomes so useful for the students. Here, they applied k means clustering algorithm for analyzing the students result from data and predicting the students' performance. Unsupervised techniques are also called clustering techniques. The k means it is a partition-based clustering algorithm. The distance measure in k means clustering is Euclidean distance. Here the data set used was obtained from the information department of the engineering college. The attribute is aggregate and attendance for an experiment. They create the final output after clustering; they show by red, green, blue to differentiate the poor, average, good students.

Qasem A. Al-Radelideh [8]-The paper title is "Mining student data using decision tree". They use the data mining process for student performance in university courses to help the higher education management. Many factors affect performance. These methods consist of five steps: collecting relevant features, preparing data, constructing the classification model, testing the model, and then predictions for the future. The data were collected in a 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.

MashaelA[9]-These researches have applied a decision tree for predicting a student's final GPA. It used WEKA toolkit. It collects the data from C.s. College at king save university in 2012 were collected from the institute.

Each student record with different attributes. Student name, student id, final GPA, a semester of graduation etc. It is important to improve the final GPA of the student.

Ryan S. J. D. Baker [10]-"The state of educational data mining in 2009:A review and future vision "In these paper author reviews the trend in 2009 in the field of educational data mining. The year 2009 finds research communizing 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 data mining. There are many applications of EDM. These papers discuss the EDM.

Pooja M. Dhekankar[11]-"Analysis of student performance using data mining concept "Data mining technique is used in many areas and the educational field. It becomes so important for the students' future. Students classification is done based on students mark. Association rule, clustering outlier detection, classification is discussed in this paper.

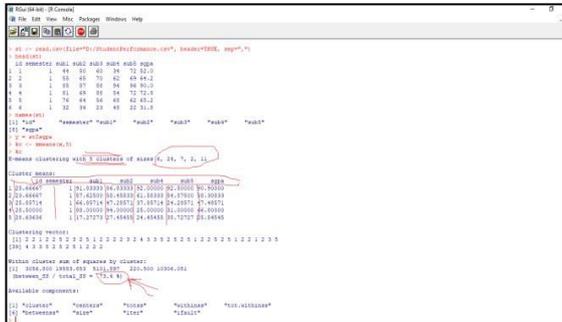
Amjad Abu saa[12]-It applies c4.5, CART, ID5 Algorithm for Analysis of students' performance. It takes various parameters for accuracy. The decision tree is built and based on its student performance is predicted. Naive Bayes classification is also applied, which assumes that all given attribute in a dataset is independent. It creates different predictive models by using different data mining tasks that effectively predict student grades .various decision tree algorithm was implemented. Finally, we can say that it helps the university as well as students.

Yoav Bergner[13] 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 numerical method for analyzing student response matrix to predict response; it naturally parameterizes a series of models and multidimensional IRT.

Existing System

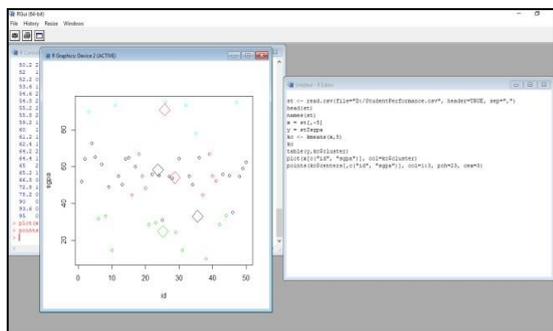
A decision tree is supervised techniques, and there are many methods to build the decision tree and predict the performance. There is a huge amount of data produced in the educational system. These can be exploited to extract useful knowledge. In today's system, lots of technique is used to predict students' performance. In the existing system, a decision tree is build using J48 Algorithm. There is a case of Algerian university in which a student's performance predicts using a decision tree. Decision tree method is unstable because the decision tree gives many possible answers. On changing the root node, it changes the tree and has a different prediction. There is a huge amount of data in the educational system in the existing system. They predict the performance based on the previous semester result. A decision tree is build using the J48 Algorithm, which is very hard to build because of its splitting. Tree algorithm uses many tests to determine the particular split. But even before that

Phase 2: Insert K cluster 1 (Choose random K element)
 Phase 3- Measure the arithmetic mean of each cluster.
 For each of the data records, determine which cluster to hold (the nearest cluster using a distance measure). It re-assigns each record to the most similar cluster and recalculates the mean of the entire cluster.



RESULT ANALYSIS

K-Means algorithm is used to predict the students' performance. It is stable and efficient as compared to the decision tree. In the dataset, we take the attribute-Student_id,-Unique id corresponds to every student. Semester (sem1-sem2)-Semester id correspond to semester i.e.(sem 1 or sem 2). Subject-marks (sub1-sub5)-Each subject mark corresponds to every student in both the semester. Sem Result (SGPA)-The percentage of those students in that particular semester.



When k=5 and the graph between id and SGPA

Table 1 Compression Study Existing Work versus Proposed Work

Parameters	Existing Work	Proposed Work
	Decision Trees	K-Mean
Average (SGPA)	Above 58%	Above 71 %
Execution Time	13.2 Seconds	10.32 Seconds

Conclusion

Machine learning is a very emerging technology that every place it used. Nowadays in the bank, labs, telecom, industrial every place machine learning is used. Data mining is part of it, which predicts prediction; future prediction is very important in many places that help so much. Many algorithms are built, and more and more research are going on every technology used the concept of it. We survey many papers for the prediction of students' performance. Decision tree method is used in many places, but on comparing to clustering techniques, i.e., k means it is less efficient, K means more

efficient and stable. Students' performance is important for their future; it helps students and helps teachers, institute parents. Many big institutes used the concept of A.I. for prediction.

References

- [1]. Fatma chiheb, Fatima Boumahdi, predicting students' performance using Decision trees: Case of an Algerian University.2017 International Conference on Mathematics and information technology, Adrar, Algeria -Dec 04-5,2017.
- [2]. V. Shanmuga Rajeshwari, Analysis of students' performance evaluation using classification techniques,978-1-4673-8437-7 IEEE.
- [3]. M. Durairaj-Educational data mining for predicting students' performance using a clustering algorithm, M. durairaj et al. (IJCSIT) International journal of computer science and information technologies vol.5(4),2014.
- [4]. Mr Shashikant Pradip Borgavakar, evaluating students' performance using K means clustering, International journal of engineering research and Technology(IJERT) vol.6 issue May 05 2017.
- [5]. Edin Osman Begovic, Mirza Suljic-Data mining approach for predicting student performance Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012.
- [6]. E. Venkatesan et al.-Prediction of students' academic performance using classification and clustering algorithm, International journal of pure and applied mathematics volume 116 no. 16 2017.
- [7]. A. Seetha ram Nagesh –Application of clustering algorithm for Analysis of student academic performance, International journal of computer sciences and engineering volume-6,issue-1.
- [8]. Qasem A. Al-Radaideh, Emad Al-Shawakfa - Mining Student Data Using Decision Trees, Research Gate Article 2006.
- [9]. Michael 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.
- [10]. Ryan S.J.D. Baker Kalina Yacef "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.
- [11]. Miss. Pooja M. Dhekankar Dinesh S. Datar "Analysis of Student Performance using Data Mining Concept" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 3 Issue: 5 2942 – 2944 IJRITCC | May 2015.
- [12]. Amjad Abu Saa, "Educational Data Mining & Students' Performance Prediction", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016.
- [13]. Yoav Bergner, Stefan Droschlery, Gerd Kortemeyerz, Saif Rayyan, Daniel Seat on and David E. Pritchard "Model-Based Collaborative Filtering Analysis of Student Response Data: Machine Learning Item Response Theory" Proceedings of the 5th International Conference on Educational Data Mining.