

A REVIEW PAPER ON STUDENT PERFORMANCE USING DECISION TREE ALGORITHMS

¹Pratibha V.Jadhav ²Dr.Vaishali V. Patil,³Dr.S.D.Gore
¹Ph.D research scholar ²Assistant Professor, ³Professor
¹JJTU, Rajasthan. ²TC College, Baramati. ³SPPU Pune.

Abstract: The achievement of any educational institute is depending upon the performance of students. The prediction of student performance and its analysis are very essential factor for improvement of institutes. There are many attributes which affects student performance like attendance, internal marks, final marks etc. This type of prediction helps teachers and institutes to identify weak students to improve their marks and for taking variety of quick decisions on any situation. In data mining, there are various methods like classification, regression, association and clustering used for analysis of performance. These techniques are used in education field known as Knowledge discovery database (KDD). This is used to find latest hidden patterns from student data base for the analysis of performance. This paper presents a survey of the literature on student performance using Decision tree and different algorithms of decision tree in Educational field.

Index Terms: Performance of the students, decision tree, data mining, Educational data mining

1. Introduction:

In today's educational institutes, student performance is very important part of achievement for any institute. Many institutes are used final grades for evaluating the performance the students. This grade contains many attributes such as attendance, internal marks, final marks and extracurricular activities. These attributes are affects the performance of the students. It is very essential to predict and analysis the performance of students. This prediction helps teachers to identify weak students and to improve their marks and used for taking variety of quick decisions on any situation. Various data mining methods are used for prediction and analysis of student performance. In this paper, we briefly reviewed the concepts of data mining, Educational data mining, and uses of decision tree classification techniques.

2. Data mining:

Data mining is a method of extracting information from large amounts of data and it ignores rest. It is related with theories, methodologies and computer systems for mining from large amount of data. There are various techniques in Data mining to predict future developments and performance. These techniques are helps to take knowledge-driven decisions and for taking variety of quick decisions on any situation. Basically Data mining uses additional direct approach and it can work easily on large projects or in the form of unbalanced data which requires constant updating. Generally, it contains planning phase, risk analysis phase, development and testing phase and evaluation phase [1].

3. Educational data mining:

The modern educational institutes and organizations are apply data mining methods to take better decisions to grow and enhancements of educational organization. The powerful tools of data mining are uses for analyzing and extracting the unseen designs and interactions from the data. This technique is applied in educational environments for preprocessing of the data. The information is generated from educational institutes where research work is happen called as Educational Data Mining [2].

4. Decision tree:

Decision tree is one of most popular and powerful classification techniques in data mining. All science and technology are exploring large and compound type of data to find some relevant and useful information for modeling and making the process more efficient, cost-effective and accurate. Basically, Decision tree is executed in decision theory and statistics. The decision tree algorithm is very beneficial in data mining for handling a variety of nominal, numeric and textual type of response documents and various datasets where datasets having a large number of errors and missing values[3]. In this paper, a survey is related with the concepts of data mining in Educational field, different classification techniques used in Educational data mining for improvement of student performance. This paper focuses the role of decision tree, different algorithms of Decision tree and comparative study of other classification techniques with Decision tree.

5. Review of Educational data mining using Decision Tree:

Shahiri et al had suggested that in Educational data mining method, Prediction of student's performance was constructed by using different tools of data mining namely, Regression, classification and categorization. The most famous task is classification which is used to predict the student performance. Different algorithms are used in classification for predicting the student performances. Amongst all algorithms, Decision tree, Artificial Neural Networks, Naïve Bayes, K-Neighbor and Support Vector Machine are popular[4].

Ankita A Nichat et al had focused on the improvement of prediction and classification techniques such as Decision Tree, based on their academic performance. In this study they had constructed a classification model which classifies student performance and guide the students towards the achievement of good score and would help them to taking timely decisions. They have used C4.5, ID3 and adapted a greedy approach. In this algorithm there was no backtracking and this design was based upon cross-sectional area. This research would help to the teacher that on which topics student was weak to prevent academic risk and desertion [5].

Mrinal Pandey et al has proposed a classification model namely decision tree algorithm for prediction the future grades of the engineering students in their final university examinations. Here, data mining techniques were used for absorbing the relevant information from large and compound data for analysis of data and predictions. Classification is most popular techniques in data mining. Classification is an unsupervised technique which uses to differentiate the predefined class labels. There are various popular classification techniques namely, Bayesian network, neural network, Genetic algorithm and Decision tree. There were many

factors to affect the performance of students. Decision tree was developed by considering some major factors and classification of students along with their grades. Four different decision tree algorithms were compared namely, J48, NB tree, Reptree and Simple cart. Out of these algorithms, J48 decision tree was found to be the best suitable algorithm for construction of model. This technique helps to predict future development based on previous records. WEKA software is used for construction and evaluation of model [6].

Kolo et al had given a decision tree approach for predicting academic performance of the students. This was necessary to predict student performance for improvement of the quality of education. This research has considered some factors that affect the student performance namely, personal, social and economic. The study was carried out to identify factors that will affect the academic performance of students and predict the academic performance of students as either pass or fail with the help of decision tree [7].

Acharya et al were applied different Machine Learning Techniques for prediction of student results and to predict the performance of student early by using previous semester results. The objectives of the study were to identify the attributes and examination pattern of students, perform and classify features selections which were used for classification. Three learning algorithms were used such as decision tree C4.5, Naïve Bayes and 1-Nearest Neighborhood. To analyze the various algorithms, Paired t-test were used .To check the efficiency of model, Kappa statistic and F-test values were used [8].

Daud et al had predicted the student performance by using different types of characteristics like academic record, family assets, income of a family and personal information. Study has been observed that expenditure of a family and their personal information might be affects the performance of the students and decrease the interest and concentration in their studies. In this proposed method, SVM, BN and NB classifiers were performed better than C4.5 and CART [9].

Madhankumar R et al were predicted the performance of students related to previous academic performance parameters, involvements in socio economic activities. To construction of the model the Decision tree algorithm has used and compared the impact of various attributes on students performance. This would help to identify the weak students and management to take timely decisions to improve the student performance [10].

Twinkle Chawla et al has found the factors that affect the performance of students and to find weak students that could give some helpful action to increase their academic performance. They have utilized greedy techniques to produce decision tree. J48 algorithm was used for splitting the datasets and also analyzed the normalized information gain [11].

Kabakchieva et al has been focused on development of data mining models for predicting the performance of student. Author has been used classification methods such as a Rule Learner, Decision tree, a neural network and K-Nearest Neighbour method for predicting a performance of students. The four classification models were compared and evaluated with Kappa Statistic, True positive, false positive rates, Precision, Recall, F-Measure and ROC Area [12].

Zeng et al had used survey data to develop a decision Tree model for forecasting the popularity of large number of Chinese colleges with average passing score. Author has been proposed efficient decision tree based on splitting criterion and decision tree pruning [13].

Khobragade et al had proposed a model on prediction of academic failure of students using Data mining Techniques. The objective of the research was to identify the failure of students to avoid them from dropping out and improve their academic performance. The Decision tree algorithm would help the Teacher and students to improve their performance by changing some regulation in their teaching methodologies. This study would help to teachers and Principal to make suitable provisions to increase the ability of students in their academics [14].

Pumpuang et al has proposed and compared the classifier algorithms for model of Course Registration Planning. The proposed model was used to predict the grade point average (GPA) of undergraduate students. NB Tree was showed the best predictive measures and selected as Course Registration Planning model for the students [15].

Taruna et al had compared five Classification Techniques namely, Decision tree, Naïve Bayes, Bayesian Network, NB Tree and K-Nearest Neighbor methods for engineering students for predicting performance of grades. Attributes were selected by Gini index, Information gain and Gain ratio for decision tree model. As a result, Decision tree J48 was best appropriate classifier for the datasets [16].

6. Conclusion:

Decision tree algorithms such as ID3,J48,C4.5 ,NB tree, Reptree and Simple cart used for predicting student performance and helps to Teacher and students to improve their performance by changing some regulation in their teaching methodologies.

7. References:

- [1] Prof. Lumbini P. Khobragade, Prof. Pravin Mahadik "Predicting Students' Academic Failure Using Data Mining Techniques". *International Journal of Advance Research in Computer Science and Management Studies* Volume 3, Issue 5, May 2015 pg. 191-199
- [2] Priyanka Saini, Ajit Kumar Jain "Prediction using Classification Technique for the Students' Enrollment Process in Higher Educational Institutions" *International Journal of Computer Applications (0975 – 8887) Volume 84 – No 14, December 2013.*
- [3] Li, Linna, and Xuemin Zhang. "Study of data mining algorithm based on decision tree." In *Computer Design and Applications (ICCDA), 2010 International Conference on*, vol. 1, pp. V1-155.IEEE, 2010
- [4] Shahiri, Amirah Mohamed, and Wahidah Husain. "A review on predicting student's performance using data mining techniques." *Procedia Computer Science* 72 (2015): 414-422.
- [5] Ankita A Nichat, Dr. Anjali B Raut "Predicting and Analysis of Student Performance Using Decision Tree Technique" *International Journal of Innovative Research in Computer and Communication Engineering* ,Vol. 5, Issue 4, April 2017 ISSN(Online): 2320-9801 ISSN (Print): 2320-9798 page no. 7319-7328
- [6] Mrinal Pandey, Vivek Kumar Sharma "A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction" *International Journal of Computer Applications (0975 – 8887) Volume 61– No.13, January 2013*
- [7] Kolo, Kolo David, Solomon A. Adepoju, and John Kolo Alhassan. "A decision tree approach for predicting student's academic performance." *IJ Education and Management Engineering* 5 (2015): 12-19.
- [8] Acharya, Anal, and Devadatta Sinha. "Early prediction of student's performance using machine learning techniques." *International Journal of Computer Applications* 107, no. 1 (2014).
- [9] Daud, Ali, NaifRadiAljohani, RabeehAyazAbbasi, Miltiadis D. Lytras, Farhat Abbas, and Jalal S. Alowibdi. "Predicting student performance using advanced learning analytics." In *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 415-421. International World Wide Web Conferences Steering Committee, 2017
- [10] Madhan kumar R and Rajesh N "Predicting upcoming students performance using mining technique" *International Journal of Modern Trends in Engineering and Research (IJMTER) Volume 04, Issue 7, [July–2017] ISSN (Online):2349–9745; ISSN (Print):2393-8161*
- [11] Twinkle Chawla, Gurpreet Singh" *International Journal of Advanced Research in Computer Science*", vol.8 No. (7), July-August 2017, 1245-1248 ISSN No. 0976-5697

[12] Kabakchieva, Dorina. "Student performance prediction by using data mining classification algorithms." *International Journal of Computer Science and Management Research* 1, no. 4 (2012): 686-690..(2012)

[13] Zeng, Xiangxiang, Sisi Yuan, You Li, and QuanZou. "Decision tree classification model for popularity forecast of Chinese colleges." *Journal of Applied Mathematics* 2014.

[14] Khobragade, Lumbini P., and PravinMahadik. "Predicting Students' Academic Failure Using Data Mining Techniques." *International Journal* 3, no. 5 (2015).

[15] Pumpuang, Pathom, AnongnartSrivihok, and PrasongPraneetpolgrang. "Comparisons of classifier algorithms: bayesian network, C4. 5, decision forest and NBTree for course registration planning model of undergraduate students." In *Systems, Man and Cybernetics, 2008.SMC 2008. IEEE International Conference on*, pp. 3647-3651. IEEE, 2008.

[16] Taruna, S., and Mrinal Pandey. "An empirical analysis of classification techniques for predicting academic performance." In *Advance Computing Conference (IACC), 2014 IEEE International*, pp. 523-528. IEEE, 2014

