Students Performance Prediction Using Decision Tree Technique

Dr. Anjali B Raut

Department of Computer Science andEngineering, H.V.P.M.'s C.O.E.T., Amravati, Amravati, Maharashtra, India.

Ms. Ankita A Nichat

Department of Computer Science and Engineering, H.V.P.M.'s C.O.E.T., Amravati, Maharashtra, India.

Abstract

Analyzes data mining methods and techniques students' data to construct a predictive model for students' performance prediction. Data mining is also use for sorting the educational problem by using analysis techniques for measuring the student performance. In this paper, measuring student performance using classification technique such as decision tree. In this research, the paper has been focused classification techniques which are used to analyze performance by the scope of knowledge. Giving the details about the results, and the specific needs of studies to improvement, such as the accompaniment of students along their learning process, and the taking of timely decisions in order to prevent academic risk and desertion. Lastly, some recommendations and thoughts are laid out for the future development of performance.

Keywords : Data mining, Classification algorithms, decision trees.

INTRODUCTION

Classification techniques are using for education data modeling. There is an increase in their application within the last six years. Researchers prefer to apply a single technique in their studies on student evaluations like those mentioned above. Educational data mining applied many techniques are K- nearest neighbor, decision tree, Naïve Bayes, Neural network, Fuzzy, Genetic and other techniques are applied in the environment

In addition, these methods lack the ability to reveal useful hidden information. Exam

model is being constructed which is beneficial for both college and students. With this model, institutes can register and host exams. Students can give exams and view their results with guidance and instructor can evaluate the student performance. This model is an attempt to remove the existing flaws in the manual system of conducting exams. Examination System fulfills the requirements of the institutes to conduct the exams. Thus the purpose of the model is to provide a system that saves the efforts and time of both the institutes and the students. Institutes enter the questions they want in the exam. These questions are displayed as a test to the eligible students. The answers enter by the students are then evaluated and their score is calculated and saved. This score then can be accessed by the institutes to evaluate their performance. In this study measures the student performance by using data mining technique like classification, decision tree algorithm using to build the classifier model on base on dataset composed of responses of students to courses evaluation questions.

LITERATURE REVIEW & RELATED WORK

Mustafa Agaoglu [1] research in educational mining focuses on modeling student's performance instead of instructors' performance. One of the common tools to evaluate instructors' performance is the course evaluation questionnaire to evaluate based on students' perception. In this study, four different classification techniques, –decision tree algorithms, support vector machines, artificial neural networks, and discriminant analysis– are used to build classifier models. Their performances are compared over a dataset composed of responses of students to a real course evaluation questionnaire using accuracy, precision, recall, and specificity performance metrics. Although all the classifier models show comparably high classification performances, C5.0 classifier is the best with respect to accuracy, precision, and specificity. In addition, an analysis of the variable importance for each classifier model is done. Accordingly, it is shown that many of the questions in the course evaluation questionnaire appear to be irrelevant. Furthermore, the analysis shows that the instructors' success based on the students' perception.

Tripti Mishra,Dr. Dharminder Kumar,Dr. Sangeeta Gupta [2] use different classification techniques to build performance prediction model based on students' social integration, academic integration, and various emotional skills which have not been considered so far. Two algorithms J48 (Implementation of C4.5) and Random Tree have been applied to the records of MCA students of colleges affiliated to Guru Gobind Singh Indraprastha University to predict third semester performance. Random Tree is found to be more accurate in predicting performance than J48 algorithm.

Keno C. Piad,Menchita Dumlao,Melvin A. Ballera,Shaneth C. Ambat [3] predicts the employability of IT graduates using nine variables. First, different classification algorithms in data mining were tested making logistic regression with accuracy of

78.4 is implemented. Based on logistic regression analysis, three academic variables directly affect; IT_Core, IT_Professional and Gender identified as significant predictors for employability. The data were collected based on the five year profiles of 515 students randomly selected at the placement office tracer study.

Bipin Bihari Jayasingh [4] initiates a sample study that is taken for a particular institution, in the particular environment, for the particular batch and particular set of students. The sample data are collected from a classroom by distributing the questionnaire attempted by two different batches of student having questions pertaining to Inquiry based and deductive learning. The system is developed and tested twice after teaching the content using inductive method and implemented using attribute relevance, discriminant rules of class discrimination mining. The results are visualized through bar charts and shows that the two batches of learners of different years have different learning characteristics.

S. M. Merchán [5] presents and analyzes the experience of applying certain data mining methods and techniques on 932 Systems Engineering students' data, from El Bosque University in Bogotá, Colombia; effort which has been pursued in order to construct a predictive model for students' academic performance. As an iterative discovery and learning process, the experience is analyzed according to the results obtained in each of the process' iterations. Each obtained result is evaluated regarding the results that are expected, the data's input and output characterization, what theory dictates and the pertinence of the model obtained in terms of prediction accuracy. Said pertinence is evaluated taking into account particular details about the population studied, and the specific needs manifested by the institution.

M. Mayilvaganan,D. Kalpanadevi [7] research, the paper have been focused the improvement of Prediction/ classification techniques which are used to analyze the skill expertise based on their academic performance by the scope of knowledge. Also the paper shows the comparative performance of C4.5 algorithm, AODE, Naïve Bayesian classifier algorithm, Multi Label K-Nearest Neighbor algorithm to find the well suited accuracy of classification algorithm and decision tree algorithm to analysis the performance of the students which can be experimented in Weka tool.

Crist´obal Romero [8] Educational data mining (EDM) is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. EDM uses computational approaches to analyze educational data in order to study educational questions. This paper surveys the most relevant studies carried out in this field to date. First, it introduces EDM and describes the different groups of user, types of educational environments, and the data they provide. It then goes on to list the most typical/common tasks in the educational environment that have been resolved through data-mining techniques, and finally, some of the most promising future lines of research are discussed.

PROPOSED METHODOLOGY

The major objective of the proposed methodology is to build the classification model that classifies a students' performance. The classifiers, has been built by combining the Standard for Data Mining that includes student performance and finally application of data mining techniques which is classification in present study.

In other words, using this Decision tree algorithm, we wanted to be able to guide student towards achievement of good score that we felt they would enjoy doing.Tree-based methods classify instances by sorting the instances down the tree from the root to some leaf node, which provides the classification of a particular instance. Each node in the tree specifies a test of some attribute of the instance and each branch descending from that node corresponds to one of the possible values for this attribute [11].

The decision trees generated by C4.5 can be used for classification. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller sub lists.

We also using Generalized Sequential Pattern mining algorithm for predicting the student's performance as pass or fail. Once the student is found at the risk of failure he/she can be provided guidance for performance improvement.

RESULT AND DISCUSSION

In result student getting the report of given test is show in figure. Figure 1 shows the test given by student marks obtains, total marks, and provide suggestion to study for improving performance.
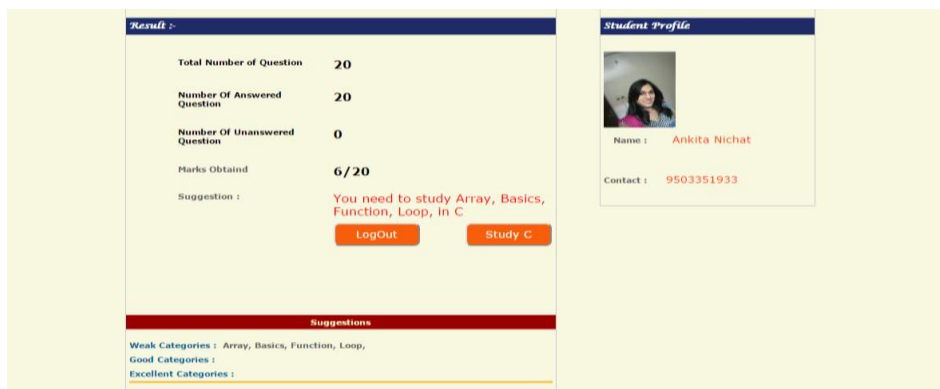
Fig 1: Result of test given

Figure 2 shows the student result to teacher. Which show how many tests are given by student and their performance according to category, weak concept, etc.
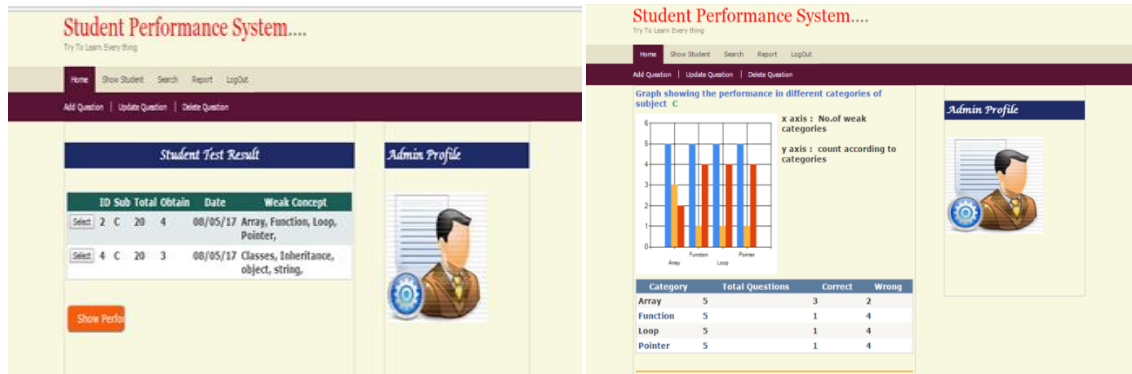

Fig 2: Student Performance

Analysis of time required for query processing and execution. Figure shows the Graphical Representation of Response Time requires for query processing and execution.
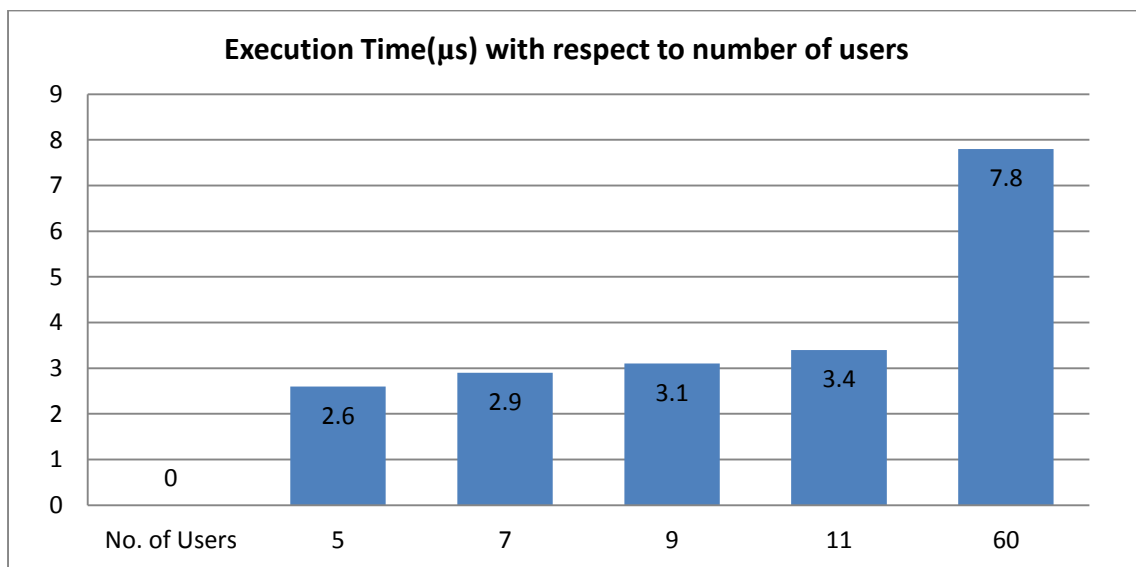

Fig 3: Graphical Representation of Response Time

CONCLUSION AND FUTURE DIRECTION

Academic success of students of any professional Institution has become the major issue for the management. An early analysis of students at risk of poor performance helps the management take timely action to improve their performance through extra

coaching and counseling. This paper focuses on analysis student academic performance by using advantage of data mining techniques model. Predicting student performance in a specific subject based on their performance of test result components during the performance by applying the C4.5 decision tree. Future work include applying data mining techniques on an expanded data set which consider extracurricular activities and other vocational courses completed by students, which we believe may have a significant impact on the overall performance of the students.

REFERENCES

[1]   Mustafa Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in   HigherEducation," *IEEE Access ,* Volume*: 4 ,*2016.

[2]   Tripti Mishra,Dr. Dharminder Kumar,Dr. Sangeeta Gupta,"Mining Students' Data for Performance Prediction," *in fourth International Conference on Advanced Computing & Communication Technologies,*2014.

[3]   Keno C. Piad, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat," Predicting IT Employability Using Data Mining Techniques," *in third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC),* 2016.

[4]   Bipin Bihari Jayasingh,"A Data Mining Approach to Inquiry Based Inductive Learning Practice In Engineering Education," *in IEEE 6th International Conference on Advanced Computing,*2016.

[5]   S. M. Merchán,"Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic,"*IEEE Latin America Transactions,* vol. 14, no. 6, June 2016.

[6]   Konstantina Chrysafiadi and Maria Virvou," Fuzzy Logic for adaptive instruction in an e-learning environment for computer programming," *IEEE Transactions on Fuzzy Systems ,*Volume*:* 23, Issue: 1, Feb. 2015.

[7]   M. Mayilvaganan,D. Kalpanadevi ," Comparison of Classification Techniques for predicting the performance of Students Academic Environment," *in International Conference on Communication and Network Technologies (ICCNT),* 2014.

[8]   Crist´obal Romero," Educational Data Mining: A Review of the State of the Art," *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews,* Vol. 40, No. 6, November 2010.

[9]   Priyanka Anandrao Patil, R. V. Mane," Prediction of Students Performance Using Frequent Pattern Tree," *Sixth International Conference on Computational Intelligence and Communication Networks,*2014.

[10] Behrouz Minaei-Bidgoli, Deborah A. Kashy , Gerd Kortemeyer, William F. Punch, "Predicting Student Performance: An Application Of Data Mining Methods With An Educational Web-Based System," *in 33'd ASEE/IEEE Frontiers in Education Conference, T2A*-13,November 5-4,2003.

[11] R.S.J.D Baker and K.Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions" , *Journal of Educational Data Mining*, 1, Vol 1, No 1, 2009.