Assessment of Graduate Students' Performance using Data Mining: Proposed Research

Mrs. Priyanka Rahul Patil[1] Dr. R. S. Kamath[2]
[1]M. Phil. Student [2]Associate Professor
[2]Department of Computer Engineering
[1,2]Chhatrapati Shahu Institute of Business Education & Research, Kolhapur, Maharashtra, India

Abstract— Data mining provides a set of techniques which can help educational system to improve learning experience of students as well as increase their profits. Proposed research is designed to justify that various data mining techniques intent in extraction of the hidden knowledge from the student database and prediction of students' performance based on dependency parameters. This aids to identify the students who need special attention and allow the teacher to provide appropriate advising. The researcher has justified that various data mining techniques intent in extraction of the hidden knowledge from the student database and prediction of students' performance based on dependency parameters.

Key words: Data Mining Approach, R-Environment, Pre-established criteria

I. INTRODUCTION

The biggest challenges now-a-days, the educational institutions face, are the explosive growth of educational data and to use this data to improve the quality of managerial decisions to deliver quality education. Manual data analysis has been around for sometimes now, but it creates bottleneck for large data analysis. Data mining allows user to analyze data from different dimensions categorized it and summarize the relationship, identified during mining process. The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data. Data mining tools predict patterns, future trends and behaviors, allowing businesses to effect proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analysis of past events provided by retrospective tools typical of decision support systems.

Data mining is considered as the most suited technology appropriate in giving additional insight into the lecturer, student, alumni, manager, and other educational staff behavior and acting as an active automated assistant in helping them for making better decisions on their educational activities. The data mining techniques can help the institutes in extracting patterns like students having similar characteristics, association of students' attitude with performance, what factors will attract meritorious students and so on.

Proposed research presents data mining process for assessing students' academic performance using the R environment and selected R packages for computing. One way to achieve highest level of quality in education system is by discovering knowledge for the assessment of students' performance.

II. LITERATURE REVIEW

Several studies used data mining for extracting rules and predicting certain behaviors in educational domain. In this literature review, references of the similar work have taken and explained the same with respect to the proposed research. As a part of this research more than 20 papers has been explored and thoroughly studied. Some selected references for broad overview are highlighted here.

Romero and Ventura have a surveyed on educational data mining between 1995 and 2005 [9]. They concluded that educational data mining is a promising area of research and it has a specific requirement not presented in other domains. Thus, work should be oriented towards educational domain of data mining. El-Halees, explained a case study that used educational data mining to analyze students' learning behavior [5]. The goal of his study is to show how useful data mining can be used in higher education to improve students' performance. He used students' data from database course and collected all available data including personal records and academic records of students, course records and data came from e-learning system. Then, he applied data mining techniques to discover many kinds of knowledge such as association rules and classification rules using decision tree. Also he clustered the student into groups using EMclustering, and detected all outliers in the data using outlier analysis. Finally, he presented how can we benefited from the discovered knowledge to improve the performance of student.

Al-Radaideh et al. applied the data mining techniques, particularly classification to help in improving the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses [1]. The extracted classification rules are based on the decision tree as a classification method; the extracted classification rules are studied and evaluated. It allows students to predict the final grade in a course under study. Baradwaj and Pal, applied the classification as data mining technique to evaluate student' performance, they used decision tree method for classification [3]. The goal of their study is to extract knowledge that describes students' performance in end semester examination. They have used students' data retrieved from the student' database including Attendance, Class test, Seminar and Assignment marks. This study helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising.

Shannaq et al., applied the classification as data mining technique to predict the numbers of enrolled students by evaluating academic data from enrolled students to study the main attributes that may affect the students' loyalty (number of enrolled students) [11]. The extracted classification rules are based on the decision tree as a classification method, the extracted classification rules are studied and evaluated using different evaluation methods. It allows the University management to prepare necessary

resources for the new enrolled students and indicates at an early stage which type of students will potentially be enrolled and what areas to concentrate upon in higher education systems for support.

Chandra and Nandhini, applied the association rule mining analysis based on students' failed courses to identifies students' failure patterns [4]. The goal of their study is to identify hidden relationship between the failed courses and suggests relevant causes of the failure to improve the low capacity students' performances. The extracted association rules reveal some hidden patterns of students' failed courses which could serve as a foundation stone for academic planners in making academic decisions and an aid in the curriculum re-structuring and modification with a view to improving students' performance and reducing failure rate. Ayesha et al., used k-means clustering algorithm as a data mining technique to predict students' learning activities in a students' database including class quizzes, mid and final exam and assignments [2]. This correlated information will be conveyed to the class teacher before the conduction of final exam. This study helps the teachers to reduce the failing ratio by taking appropriate steps at right time and improve the performance of students.
Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes. Data Mining can be used in educational field to enhance our understanding of learning process to focus on identifying, extracting and evaluating variables related to the learning process of students as described by Alaa el-Halees [6].

Literature review reveals that data mining methods are often implemented at educational institutions for analyzing available data and extracting information and knowledge to support decision-making. Analysis will help the teachers to take proper attention towards the progress of the student during the course. It will help to build reputation of institute in the field of education. The analysis will reveal dependency factors of student's academic performance.

III. PROPOSED RESEARCH FRAMEWORK

The methodology will be followed for the present study is explained in this section. Logical flow of methodology is presented diagrammatically as follows in Figure 1. Data mining a computational method of processing data will be applied in to obtain useful knowledge from EDM dataset.

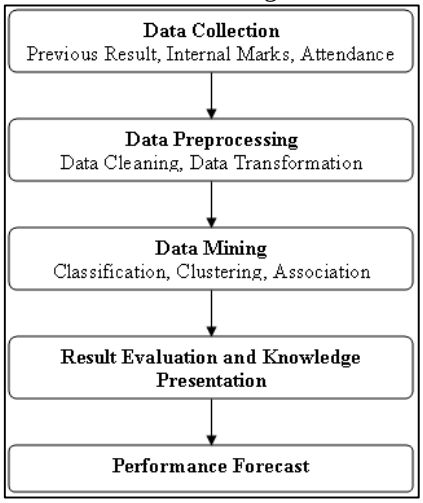

Fig. 1: Flow diagram of proposed research

A. Data Collection

The first step of the proposed research is to design sample and gather students' data. This includes constructing the dataset needed for data mining process. The proposed research planned to design data set consists of 300 records by collecting graduate students information from the department of Information Technology for a period of 1 year.

B. Data Preprocessing

Preprocessing includes finding incorrect or missing data. Erroneous data may be corrected or removed, whereas missing data must be supplied. Transformation is converting the data into a common format for processing. Some data may be encoded or transformed into more usable format. Data reduction, dimensionality reduction and data transformation method may be used to reduce the number of possible data values being considered.

C. Educational Data Mining

EDM is the exploration and analysis of large data sets, in order to discover meaningful pattern and rules. The key idea is to find effective way to combine the computer's power to process the data with the human eye's ability to detect patterns. The objective is to design and work efficiently with large data sets. After preprocessing the data, research applies data mining techniques - association, classification, clustering, and prediction etc for the analysis of educational data.

D. Knowledge Presentation

The discovered knowledge will used for prediction students' performance and constructive recommendation to overcome the problem of low grades of students. The model development, building classifiers and their evaluation, Educational data analysis using clustering will be carried out in the proposed research. The framework can be used as a basis for the prediction of students' performance in campus placement.

IV. CONCLUSION

Proposed research presents data mining process for assessing students' academic performance using R environment and selected R packages for computing. The main focus is to provide a set of techniques which can help educational system to improve learning experience of students as well as increase their profits. Authors will apply the data mining techniques such as association, classification, clustering and prediction for the assessment of educational data and present the extracted knowledge and describes its importance in educational domain by considering the data set of graduate students'. The discovered knowledge will be used for accessing students' performance and constructive recommendation to overcome the problem of low grades of students'. Thus the proposed research will assists an academic institute to improve the quality of education by analyzing the data and discover the factors that affect the academic results.

REFERENCES

[1] Al-Radaideh, Q., Al-Shawakfa, E. and Al-Najjar, M. (2006) 'Mining Student Data Using Decision Trees',

The 2006 International Arab Conference on Information Technology (ACIT'2006) – Conference Proceedings.
[2] Ayesha, S. , Mustafa, T. , Sattar, A. and Khan, I. (2010) 'Data Mining Model for Higher Education System', European Journal of Scientific Research, vol. 43, no. 1, pp. 24-29.
[3] Baradwaj, B. and Pal, S. (2011) 'Mining Educational Data to Analyze Student s' Performance', International Journal of Advanced Computer Science and Applications, vol. 2, no. 6, pp. 63-69.
[4] Chandra, E. and Nandhini, K. (2010) 'Knowledge Mining from Student Data', European Journal of Scientific Research, vol. 47, no. 1, pp. 156-163.
[5] El-Halees, A. (2008) 'Mining Students Data to Analyze Learning Behavior: A Case Study', The 2008 international Arab Conference of Information Technology (ACIT2008) – Conference Proceedings, University of Sfax, Tunisia, Dec 15- 18.
[6] Alaa el-Halees, "Mining students data to analyze e-Learning behavior: A Case Study", 2009.
[7] Brijesh Kumar Baradwaj, "Mining Educational Data to Analyze Students" Performance", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.
[8] Rajni Jindal and Malaya Dutta Borah, "A survey on educational data mining and research trends", International Journal of Database Management Systems (IJDMS) Vol.5, No.3, June 2013.
[9] Romero, C. and Ventura, S. (2007) 'Educational data Mining: A Survey from 1995 to 2005', Expert Systems with Applications (33), pp. 135-146.
[10] Ryan S.J.d. Baker. "Data Mining for Education", Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.
[11] Shannaq, B., Rafael, Y. and Alexandro, V. (2010) 'Student Relationship in Higher Education Using Data Mining Techniques', Global Journal of Computer Science and Technology, vol. 10, no. 11, pp. 54-59.
[12] Han, J. & M. Kamber, "Data mining: concepts and techniques", San Francisco: Morgan Kaufman (2001).
[13] Agathe MERCERON and Kalina YACEF, "Educational Data Mining: a Case Study", School of Information Technologies - University of Sydney, Australia
[14] Anand V. Saurkar et.al , "A Review Paper on Various Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4, April 2014
[15] Mohammed M. Abu Tair. "140 Mining Educational Data to Improve Students' Performance: A Case Study", International Journal of Information and Communication Technology Research, Volume 2 No. 2, February 2012
[16] http://www.rdatamining.com/
[17] Surjeet Kumar Yadavet. Al., "Data Mining Applications: A comparative Study for Predicting Student's performance", International Journal Of Innovative Technology & Creative Engineering (Issn:2045-711) Vol.1 No.12 December.

