Scholarly Research Journal for Interdisciplinary Studies,
Online ISSN 2278-8808, SJIF 2016 = 6.17, www.srjis.com
UGC Approved Sr. No.49366, JULY-AUG 2017, VOL- 4/35

A REVIEW OF APPLICATION OF DATA MINING TECHNIQUES IN HIGHER EDUCATIONAL INSTITUTIONS

Dhanushya Mary Solomon[1], Anurag Tripathi[2], Ph. D. & Kalpana Salunkhe[3], Ph. D.

[1]*Lecturer (Research Scholar, Spicer Adventist University, Aundh Road, Pune-7)*

[2]*Assistant Professor, Alard Institute of Management Sciences, Pune - 7*

Abstract

Data mining is a process that helps to discover new patterns in large data sets. Different types of mining algorithms have been proposed by different researchers in recent years. This research paper has done some reviews on previous research works done on predicting students' performance by using various data mining methods and giving an early alert to their teachers. It also helps the teachers to modify their teaching methods if their current teaching methods give a poor performance from their students' side. Predicting students' performance is essential to be conducted to prevent failures of students. The main aim of this research work is to use suitable data mining algorithms on educational dataset. This research paper also discusses various algorithms of data mining in educational data mining which assists the teachers to help their students in improving their academic performance and experience. Every educational institutions depends on the students' academic performance and their satisfaction.

Introduction

Data mining is the efficient discovery of valuable, non-trivial information from a large collection of data. In other words it is the analysis step of the Knowledge Discovery in Database (KDD). It refers to extracting or mining knowledge from large amount of data. The extraction of data is typically used to help an organization cut costs in a particular area, increase revenue or both often facilitated by a data mining application, its primary objective is to identify and extract patterns contained in a given data set. [9]Most importantly, data mining techniques aim to provide insight that allows for a better understanding of data and its essential features. Companies and organizations can employ many different types of data mining techniques. While they may take a similar approach, all usually strive to meet different goals.

In educational institutions, data mining techniques have been applied in both, learning and administrative oriented issues. In learning the process can split into learner oriented and educator oriented. In learner oriented the focus is on supporting the student to learn more

effectively by suggesting new contents; in the later , the goal is to provide the educator a tool to empower him/her so he can guide the learner more effectively. [4]

Applying data mining (DM) in education is an emerging interdisciplinary research field also known as educational data mining (EDM). It is concerned with developing methods for exploring the unique types of data that come from educational environments. Its goal is to better understand how students learn and identify the settings in which they learn to improve educational outcomes and to gain insights into and explain educational phenomena. Educational information systems can store a huge amount of potential data from multiple sources coming in different formats and at different granularity levels.

Although, using data mining in higher education is a recent research field there are many works in this area. That is because of its potential to educational institutes. Data mining can be used in educational field to enhance our understanding of learning process to focus on identifying, variables related to the learning process of student. We could actually improve students' achievement and success more effectively in an efficient way using educational data mining techniques. It could bring the benefits and impacts to students, educators and academic institutions.

Research Statement Argument

Severe challenges are being faced by students in higher education. Institutions would like to know which students will need assistance in order to graduate. One way to effectively address these students is through the analysis and presentation of data and data mining. Data mining enables organizations to use their current reporting capabilities to uncover and understand hidden patterns in vast databases. These patterns are then built into data mining models and used to predict individual behavior with high accuracy. As a result of this insight, institutions are able to allocate resources and staff more effectively. Data mining may give an institution the necessary information so that they can take action before a student drop out.

Objectives

The main objective of this research work is to review various related research work done on Educational Data Mining (EDM) techniques to predict students' academic performance earlier. Studying various data mining techniques and applying them in educational institutions can act as an effective tool that provides information to change educators' practices and make an alert to help students to get back on track. Each particular educational problem has a specific objective with special characteristics that require a different treatment

of the mining problem. The issues mean that traditional DM techniques cannot be applied directly to these types of data and problems. As a consequence, the knowledge discovery process has to be adapted and some specific DM techniques are needed. Therefore a review is done on various related work which will give us an insight on various data mining techniques which will be suitable for different educational problems.

Related Previous Work

Garg and Sharma [2], focused on comparative analysis of data mining techniques and algorithms in order to predict students' placement in order to provide the overall placement of the college. Their experiments were carried out using data mining tools in WEKA.

Amjad Abu Saa [1], reviewed multiple decisions tree techniques and algorithms and their performance and their accuracies were tested and validated. It was noticed that Naïve Bayes algorithm worked better with the dataset than others. CHRT showed more accuracy of 40%. This study also presents high potential results in the data mining analysis of the Naïve Bayes model.

Mishra [7], used data mining model namely decision tree, neural network and Naïve Bayes Classifier. They were used to predict the overall absorption rate for every branch as well as the time it takes for all the students of a particular branch to get placed. This research work also puts forward the data mining algorithm namely C4.5* stat for numeric data sets which have been proved to have competent accuracy over standard bench marking data sets called UCI datasets. To predict students placement in order to provide the overall placement of the college.

Sumitha and Vinodhkumar [5], designed a student's data model using J 48 algorithm which proved to be an efficient algorithm in terms of accuracy identified by a comparative study of data mining classification algorithm. The researchers used a questionnaire to collect the real data from the students that described the relationship between learning behavior and academic performance. Next only the required data for data mining was selected. The variables for judging the learning and academic behavior of students used in the questionnaire are students' demographic details, school details, attendance, CGPA and final grade in last semester. The WEKA explorer application was used to build the classification model using different algorithm and each of them using different classification technique.

Kalpana and Venkatalakshmi [3], applied data mining techniques to discover knowledge based on the clustering methods such as centroid based, distribution based and density based

clustering. Cluster included groups with small distance among the cluster members. Also they clustered the students into groups using K-means clustering algorithm. Finally the distance-based approach and density-based approach were used. Each one of these tasks were used to improve the performance of graduate students.

Hamsa, Indradevi and Jbilant [6], used two selected classification methods; Decision Tree and Fuzzy genetic algorithm in order to develop students' academic performance prediction model, for the Bachelor and Master degree students in Computer Science and Electronics and Communication streams. Parameter like internal marks, sessional marks and admission score were selected to conduct this work. In decision tree, predefined strict decisions are followed, so students at the barrier of success will be identified as at risk. So in lecturer's point of view, average-level students will also be under the care of experts. So in the final exams, good result can be expected. In fuzzy genetic algorithm, a student having a low score in some attributes, have a chance to be at safe due to the high score attained from other attributes which make student comfortable. It also gives students a mental satisfaction. But the lecturer will have attention on them indirectly.

P. Kavipriya [8],made a review on different students' performance based on data mining techniques under different circumstances. This review helps in the speed of predicting result. To improve the accuracy and quality of predicting students' performance, Support Vector Machine algorithm is proposed which is showing the greatest accuracy among other techniques. The comparison is shown in the below table.

Table 1: Comparison of Classification Methods

Classification Methods	Accuracy Rate
Decision Tree	94%
Naïve Bayes	96%
Neural Network	91%
Support Vector Machine	97%
DT-J48	95%

Conclusion

A review of various data mining techniques is presented in this paper. This review of the paper work has come to a conclusion that data mining techniques is an efficient method for identifying the students' performance. It gives an alarm to the instructors by predicting the students' failure well in advance and make them to succeed in their examinations. This kind of analysis also will help the institution to identify the students who need special attention

and avoid failures and dropouts. For future work a model can be developed for monitoring the students' performance using these data mining techniques.

Bibliography

Amjad Abu Saa, (2016).''Educational Data Mining and Student's Performance Prediction''. International Journal of Advanced Computer Science and Application. Volume 7, No – 5.

Sumit Garg, Arvind K. Sharma, (2013). "Comparative analysis of data mining techniques on education dataset". International Journal of Computer Application. Volume 74, No – 5.

J. K. JothiKalpana& K. Venkatalakshmi, (2014).''Intellectual performance analysis of students by using data mining techniques". International Journal of Innovation Science, Engineering and Technology, Volume 3, Special Issue 3.

Camito Ernesto Lopez Guarin (2013). "Data Mining Model to Predict Performance at the Uniersidad Nacional de Colombia".

R. Sumitha& E. S. Vinothkumar (2016).''Prediction of students outcome using data mining techniques". International Journal of Scientific Engineering and Applied Science, Volume 2, Issue – 6.

Hashmia H., Simi Indiradevi, Jubilan K.K.,(2016). "Students' academic performance prediction model using decision tree and fuzzy genetic algorithm". Science Direct, Procedia Technology 25.

Arun Mishra, (2017), "Development of data mining model for the evaluation of human skill placement in the engineering sector". International Journal of Advance Research, Ideas and Innovations in Technology. Volume 3. Issue – 3.

P. Kavipriya, (2016), "A review on predicting students' academic performance earlier, using data mining techniques". International Journal of Advanced Research Computer Science and Software Engineering. Volume 6, Issue 12.

Nilesh Magar (2016), "Data mining & Data warehousing". Vision Publication.