Mining Educational Data to Reduce Dropout Rates of Engineering Students

Saurabh Pal
Head, Department of Computer Applications,
VBS Purvanchal University, Jaunpur (U.P.), India
E-mail: drsaurabhpal@yahoo.co.in

Abstract— In the last two decades, number of Engineering Institutes and Universities grows rapidly in India. This causes a tight competition among these institutions and Universities while attracting the student to get admission in these Institutions/Universities. Most of the institutions and courses opened in Universities are in self finance mode, so all time they focused to fill all the seats of the courses not on the quality of students. Therefore a large number of students drop the course after first year. This paper presents a data mining application to generate predictive models for student's dropout management of Engineering. Given new records of incoming students, the predictive model can produce accurate prediction list identifying students who tend to need the support from the student dropout program most. The results show that the machine learning algorithm is able to establish effective predictive model from the existing student dropout data.

Index Terms— Educational Data Mining, Machine Learning Algorithms, Dropout Management, Predictive Models

I. INTRODUCTION

Data Mining is a technology used to describe knowledge discovery and to search for significant relationships such as patterns, association, and changes among variables in databases. The discovery of those relationships can be examined by using statistical, mathematical, artificial intelligence and machine learning techniques to enable users to extract and identify greater information and subsequent knowledge than simple query and analysis approaches.

A large number of Engineering students drop the course after first year. Based on review of the literature, various reasons were identified for the measuring 'relative importance' of student drop out. These reasons were then broadly grouped in to three basic categories:

- Personal reasons,

- Program/ course related reasons.

Student dropout is a challenging task in higher education [1] and it is reported that about one fourth of students dropped college after their first year [1-3]. Recent study results show that intervention programs can have significant effects on dropout, especially for the first year. To effectively utilize the limited support resources for the intervention programs, it is desirable to identify in advance students who tend to need the support most.

In this paper, we describe the experiments and the results from a data mining techniques for the students of Institute of Engineering and Technology of VBS Purvanchal University, Jaunpur to assist the student dropout program on campus. The development of machine learning algorithms in recent years has enabled a large number of successful data mining projects in various application domains in science, engineering, and business [4, 5]. In our study, we apply machine learning algorithm to analyze and extract information from existing student data to establish predictive model. The predictive model is then used to identify among new incoming first year students those who are most likely to benefit from the support of the student retention program.

The main objective of this paper is to use data mining methodologies to find students which are likely to drop out their first year of engineering. In this research, the classification task is used to evaluate previous year'sstudent dropout data and as there are many approaches that are used for data classification, the ID3, C4.5, CART and ADT decision tree methods is used here. Information like grade in High School, grade in Senior Secondary, student's family income, parents qualification etc. were collected from the student's management system, to predict list of students who need special attention.

The rest of this paper is organized as follows: Section 2 presents different type of decision trees for machine learning Section 3 describes background and history of educational data mining. Section 4 describes the methodology used in our experiments about applying data mining techniques on the educational data. Section 5 illustrates the results obtained. Finally we conclude this paper with a summary and an outlook for future work in Section 6.