Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 82 (2016) 65 – 71

Procedia Computer Science

Symposium on Data Mining Applications, SDMA2016, 30 March 2016, Riyadh, Saudi Arabia

Predicting Critical Courses Affecting Students Performance: A Case Study

Yasmeen Altujjar, Wejdan Altamimi, Isra Al-Turaiki*, Muna Al-Razgan

Information Technology Department
College of Computer and Information Sciences
King Saud University
Riyadh 12372, Saudi Arabia

Abstract

Predicting student academic performance is one of the important applications of educational data mining. It allows academic institutions to provide appropriate support for students facing difficulties. Classification is a data mining technique that can be used to build prediction models. In this paper, we use the ID3 decision tree induction algorithm to build prediction models for academic performance. Our models are built based on records for female students in the Bachelors program at the Information Technology (IT) department, King Saud University, Riyadh, Saudi Arabia. The results indicate that reliable predictions can be achieved based on the performance of students in second year courses. We also identify key courses that can be used as performance predictors. We believe our findings are useful for decision makers at the IT department.

Peer-review under responsibility of the Organizing Committee of SDMA2016

Keywords: Academic performance; decision tree; ID3; data mining

1. Introduction

Data mining is defined as the process of extracting useful and novel information from large amounts of data[1]. It has been applied to provide useful solutions in many areas such as: business, finance, medicine, and healthcare. A relatively new field of data mining applications is *Educational Data Mining* (EDM). Emerging in 2005, EDM is concerned with developing data mining techniques to discover knowledge from data obtained from educational settings[2]. The main goal of this new field of research is to support decision making in academic institutions by analyzing educational data[3]. The information produced by EDM can be useful to several stakeholders in education. For example, it can help instructors evaluate course structure and teaching strategies. In addition, students can get course recommendations based on their progress. Student advisors can benefit from EDM to predict student performance. Predicting low performing students at early stages allows providing additional support for them. According to Baker

* Corresponding author. Tel.: +0-966-11805-2909.
E-mail address: flower.132@hotmail.com, aspirant-1-@hotmail.com, ialturaiki@ksu.edu.sa, malrazgan@ksu.edu.sa

Peer-review under responsibility of the Organizing Committee of SDMA2016
doi:10.1016/j.procs.2016.04.010

and Siemens[4], EDM methods can be classified into: prediction, structure discovery, relationship mining, distillation of data, and discovery with models.

The applications of data mining in the educational context have witnessed rapid growth. There are many factors that led to the growing interest in educational data mining. With the advances in technology, universities are able to accumulate huge amount of academic and non-academic data about their students. Educational institutions are utilizing many resources, such as Learning Management Systems and Student Information Systems, that are generating volumes of data[5]. Well-developed data mining techniques can play an important role in analysing this data.

In this paper, we use data mining techniques to build a model to predict academic performance. The goal of the study is to reveal the courses affecting low academic performance at the Information Technology department (IT) for female students at the College of Computer and Information Sciences, King Saud University, Riyadh. The model will allow the IT department to make the right decisions to monitor and support students and to enhance the quality of the program.

The rest of the paper is organised as follows. In section 2, we review related work in the prediction of academic performance using data mining techniques. Section 3 describes the dataset used, the data mining technique applied, and the experimental results. Finally, section 4 concludes the paper with our findings and recommendations.

2. Literature Review

Predicting student performance is an important application of educational data mining. There are many contributions in this field using different data mining tools and techniques.

Bhardwaj et al.[6] constructed a model to predict academic performance of students enrolled in the Bachelors of Computer Applications programs in five colleges of Dr. R. M. L. Awadh University in India. The data consisted of 300 records for students including: 226 males and 74 females and featured many factors such as: academic, social, demographic, and psychological attributes. The Naive Bayes classification algorithm was used to build the model. It was found that among the most affecting variable in student academic performance are: student grade in the senior secondary school, place of residence (e.g. town, village), and the language of instruction.

In New Zealand, Carnegie et al.[7] used students results at secondary school to predict their performance during the first year in college. The study focused on predicting success in engineering programs. Several models using J48 decision tree classification algorithm were built for this purpose. The models demonstrated that the *Guaranteed Entry Score* is an informative predictor of performance, but alone is not good enough for prediction. For students in the Engineering and Computer Systems specialisation, the study concluded the importance of calculus and physics at the merit and excellence level.

de Morais et al.[8] used data mining techniques to analyse the performance of students at an English e-learning course. The data was obtained for 120 students, 6592 steps, and 11394 transactions. *K-means* algorithm was applied to cluster students based on their answers. Grouping similar students help in selecting future learning activities for each group based on their performance. Regression analysis was then applied to predict the behaviour of students in each cluster. The study showed that the two variables: number of correct answers and the number of correct first attempts, are important for three out of the five obtained clusters.

Lopez et al.[9] developed two classification models to predict the loss of academic status. Decision tree and Naive Bayes algorithm were used to build the models. The study was based on student data from two undergraduate engineering programs at the Universidad Nacional de Colombia. The data of 1532 students consisted of admission and academic data. The former was obtained from the Admission unit and included: initial academic information, demographic and socio-economic, and academic potential. The academic data was obtained from the Academic Information System and included: student records, academic period, program, and GPA. The first model used admission data to predict the loss of academic status at a particular academic period. The second combined both admission and academic data. The effect of unbalanced classes in the data was considered and recovered by applying a cost-sensitive technique in the decision tree model. Models accuracy was evaluated using different data settings. The results showed that Naive Bayes performed better in terms of inbalanced accuracy (up to 85%) in the fourth period. However, decision trees results were more consistent among all periods, which means it is more reliable.

Hashim et al.[10] applied the C4.5 decision tree classification algorithm to predict the performance of student at Alneelain University in Sudan. Academic data of 124 graduate students was obtained from at the Mathematical