Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 82 (2016) 80 – 89

Procedia Computer Science

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

Predicting Students' Performance in University Courses: A Case Study and Tool in KSU Mathematics Department

Ghada Badr[a,b]*, Afnan Algobail[a], Hanadi Almutairi[a], Manal Almutery[a]

[a]*King Saud University, College of Computer and Information Sciences, Computer Science Department, Riyadh, Kingdom of Saudi Arabia*
[b]*IRI-The City of Scientific Research and Technological Applications, University and Research District, P.O. 21934 New Borg Alarab, Alex, Egypt*

Abstract

Educational data mining is a growing field that uses the data obtained from educational information systems to discover knowledge and find answers to questions and problems concerning the education system. High dropout rates and poor academic performance among students are examples of the most common issues that affect the reputation of an educational institution. Students' academic records can be analyzed to explore the factors behind these phenomena. This paper discusses the building of a model to predict the performance of students in a programming course based on their grades in courses in other subjects. A classification based on an association rules algorithm is used to build a classifier to help evaluate the student's performance in the programming course. This model aims to reduce dropout levels by helping student predict their likelihood of success in a course before they enroll in it. In addition, course instructors will be able to enhance student performance in the course by better estimating their abilities to learn the subject matter and adjusting their teaching strategies and methods.

Peer-review under responsibility of the Organizing Committee of SDMA2016

Keywords: Data mining; classification; prediction; higher education; association rules; CBA algorithm.

1. Introduction

Every day, we generate a huge amount of data from different sources such as social networks, business transactions, and clinical records. These data are stored in databases as row data, and we do not benefit from the potentially useful information that we could extract from them. However, various data mining (DM) techniques and

* Corresponding author. Tel.:011-80-51941; fax: +0-000-000-0000 .
E-mail address: ghbadr@ksu.edu.sa

Peer-review under responsibility of the Organizing Committee of SDMA2016
doi:10.1016/j.procs.2016.04.012

tools have been developed to turn this growing volume of data into valuable information. DM, or Knowledge Discovery in Database (KDD), is defined as "the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner [1]." In other words, DM can be useful and effective in extracting important, relevant information that has not previously been discovered. DM methods and techniques provide significant potential to organizations and researchers to discover implicit information in various areas, including bioinformatics, genetics, and education.

In the past few years, DM has been successfully deployed to enhance the quality of learning and teaching in higher educational institutions. In addition, DM techniques and tools have been found capable of explaining the causes of longstanding issues faced by the higher education sector, including rates of course failure and dropout and poor academic performance. To improve these rates, various data mining methodologies, including classification, clustering, regression, and association rules, can be used to predict students' future grades. In fact, predicting students' future performance based on their past academic performance and that of previous students is a common data mining task. The ultimate aim of such predictions is to help students find how well they will do in a particular course before they register in it so as to avoid having to drop out. In addition, it will help the course instructor identifying students at risk and take appropriate actions and adopt new strategies to improve student success.

This paper presents a data mining model for predicting student performance in a programming course based on their performance in English and mathematics courses. The Classification Based on Association rules (CBA) algorithm was used to build a classification model for predicting students' performance. As its subject, the study used the academic records of mathematics students who graduated from King Saud University (KSU) between 2008 and 2014.

The rest of this paper is organized as follows: Section 2 contains background information on DM in higher education and classification association rules for DM. Section 3 introduces some related research. Section 4 proposes overall work that can be used for educational DM research. Section 5 presents the main idea behind the project, a case study proposing an application to predict students' performance in a programming course based on their performance in mathematics and English courses. Section 6 summarizes the results. Section 7 presents discussion and conclusions.

2. Background

2.1. Data mining in higher education

Data mining can be applied in various fields to enhance the overall performance of a system. This can be done by extracting from a stored dataset important, relevant information that has not previously been discovered. The knowledge thus obtained can contribute to resolving many issues and improving the current system.

There is increasing interest in using DM in the educational field. In fact, applying traditional DM techniques to educational data such as student academic records is referred to as Educational Data Mining (EDM). EDM is defined as "the process used for transforming raw data compiled by education systems into useful information that could be used by lecturers to take corrective actions and answer research questions [2]." In other words, EDM can help institutions examine and improve the student learning process.

Understanding the student learning process plays an important role in developing an institution's educational process. Such understanding offers several advantages, including improving the outcomes of student learning and enabling planning to assist weaker students. Consequently, the number of students failing or dropping out of courses will decrease.

2.2. Classification Association Rules Mining (CARM)

There are many techniques for mining large amounts of data to explore and extract knowledge from it. The two most significant DM methods are classification rule mining and association rule mining. The former generates an accurate classifier by selecting a small group of rules from the dataset. The latter aims to discover and identify all rules within some minimum support and confidence constraints.