Using Data Mining Techniques with Open Source Software to Evaluate the Various Factors Affecting Academic Performance: 
A Case Study of Students in the Faculty of Information Technology

Feras Hanandeh, Hashemite University, Faculty of Information Technology, Zarqa, Jordan
Majdi Y. Al-Shannag, Yarmouk University, Faculty of Information Technology, Irbid, Jordan
Maha Mahdi Alkhaffaf, World Islamic Sciences University, Department of Management Information Systems, Amman, Jordan

ABSTRACT

This research paper studies the different factors that could affect the Faculty of Information Technology students’ accumulative averages at Jordanian Universities, by verifying the students’ information, background and academic records. It also has the objective to reveal how this information will affect the students to obtain high grades in their courses. The information of the students is extracted from the students’ records and its attributes are formulated as a huge database. Then, a free open source software (WEKA) which supports data mining tools and techniques are used to decide which attribute(s) will affect the students’ accumulative averages. It was found that the most important factor affects the students’ accumulative averages, is the student acceptance type. A decision tree model and rules are also built to determine how the students can get high grades in their courses. The overall accuracy of the model was 46.8% which is an accepted rate.

KEYWORDS

Academic Performance, Data Mining Techniques, Faculty of Information Technology, FOSS, Jordanian Universities, Jordan, Open Source Software

INTRODUCTION

Many factors may affect the students’ performance, like their average at high school, which may reflect their academic level. Also, the way students are admitted in the universities, who pays for tuition, …etc. are other factors which might affect the students’ performance.

The students’ environment, or the place where the student lives, may also affect his/her performance and study. Gender could also be one of the factors that may affect the students’ choices of study, where traditional families treat males in a different way than females.

The academic year, students’ specialization and the number of credit hours finished by the student are also potential factors affecting his/her performance, because new students usually get higher marks in the first few semesters and then their grades start to drop in the next semesters.

All of the above factors will be studied and analyzed by using data mining techniques and functions with open source software. There are numerous techniques which can be used in data mining, like classification, clustering, association, etc. But, we will focus mainly on the classification technique.
The most famous open source application that supports data mining techniques is the WEKA application, “WEKA is the product of the University of Waikato (New Zealand) and was first implemented in its modern form in 1997. It uses the GNU General Public License (GPL). The software is written in the Java language and contains a GUI for interacting with data files and producing visual results (think tables and curves)” (Abernethy, 2010). Therefore, this application will be used to build a classification and clustering models that proves our research.

The benefits of our research are: first, to give advices for students where (In which University) and what (specialization) they have study to get high grades. Second, to prove that the acceptance type for the students is very important factor that affecting on their grades.

Most of the researcher in this area collected the data from one university or one school, but we collected the data from eight universities from Jordan, and we had got a very huge data; so that we had got an accuracy which not reaches to 50%.

LITERATURE REVIEW

Ramaswami and Bhaskaran (2010) constructed a prediction model called CHIAD to predict the performance in higher secondary school education. They collected the input data for this model from regular students, schools and chief educational officers of different district schools, with a total of 772 student records. They found numerous factors affecting the students’ performance, like: “medium of instruction, marks obtained in secondary education, location of school, living area and type of secondary education”.

Al-Radaideh et al. (2011) provided a classification approach (data mining technique) to guide students in the basic education stage in selecting their academic tracks. They developed a decision tree classification, then they extracted a set of rules with an overall accuracy of 87.9%.

Win and Miller (2005) employed two methodologies to determine the factors that influence university students’ academic performance (analogous to an input-output approach and random coefficients’ model) by using data of first year students at the University of Western Australia in 2001. They found that high school is the most affecting factor of the academic performance of the students at the university. Further, immersion and reinforcement were affecting the performance. Their results showed that there is a strong positive relationship between the first-year mark and the Tertiary Entrance Rank (TER). It was also shown that the type of school (governmental schools or otherwise) has an effect on TER.

Chamorro-Premuzic and Furnham (2003) studied students’ neuroticism, psychotics and conscientiousness, using them as factors that affect the students’ academic performance in: exams, final-year projects, student absenteeism and essay writing. They found that neuroticism weakens academic performance, whereas conscientiousness strengthens it. Also, they found that psychotics limit academic performance. Then, they provided evidence to supporting good personality measures in academic selection procedures.

Vialardi et al. (2010) used data mining techniques to present a recommendation system in order to make students aware when choosing their courses to avoid difficulties and workload, so that this system will make students proceed in a right learning path and then improve their performance.

Vialardi et al. (2011) applied the CRISP-DM methodology to data from students of the Computer Science Department at the University of Lima to build a recommendation system by using the students’ academic performance record to support the enrollment process. They used two attributes (course difficulty and grades obtained in related courses). After using many methods, they found that bagging is the best one. They developed a system called “Student Performance Recommender System” and
Open Source Software Adoption: Anatomy of Success and Failure
www.igi-global.com/article/open-source-software-adoptions/2768?camid=4v1a

The System for Population Kinetics: Open Source Software for Population Analysis
www.igi-global.com/article/system-population-kinetics/38904?camid=4v1a