

University of Science and Technology
Faculty of Post Graduate studies
& academic advancement

Using data mining tools in academic record

Thesis in partial fulfillment of the requirement of the
degree of *M.Sc.* in Information System

By:

ahmed siraj eldin ahmed

Supervisor:

Dr. Atif Ali

June 2014

Abstract

The Growth of educational university can be measured in terms of successful students of the university. Data mining techniques is to process a dataset and identify the relevance of classification test data. Mining tools to solve large amounts of problems such as classification, clustering, association rule, neural networks. Machine learning data mining tool used for different analysis Waikato Environment for Knowledge Analysis is introduced by university of New Zealand Extraction of Evolutionary Learning not only analysis the data mining classifications but also the genetic evolutionary algorithms is the best efficient tool in learning. The aim of this research shares that of a previous research programmed that involved an academic advisor helping students to predict their final passing results based on their performance in several subjects in early semesters of their study period. In this research a more practical method is introduced and applied, based the classification of functions of data mining. The classification was performed using software based upon the Classification and Regression Trees (CART) algorithm decision tree represents a decision support tool very often used because it is simple to understand and interpret. Classification and Regression Trees CART is a technique formed by a collection of rules based on values of certain variables in the modeling data set. Presents of CART to build the decision tree we used free data mining tools presented the decision tree the results and the statistical information about the data used to generate the decision model. Data mining WEKA tools is peter way to predict knowledge from large data base.

1.1 Introduction:

Classification models are supervised learning methods for predicting the value of categorical target attribute unlike regression models which deal with Numerical attributes. The development of algorithms is a step in emulating the inductive opportunities afforded by classification extends into several different application domains in classification trees Bayesian methods, neural networks, logistic regression and support vector machines. Universities today are operating in a very complex and highly competitive environment. The main challenge for modern universities is to deeply analyze their performance to build a strategy for development and future actions. University management should focus more on the profile of admitted students, getting aware of the different types and specific student' s characteristics based on the received data. The main goal of the research is to reveal the high potential of data mining applications for university management to find out any patterns in the available data. The university management would like the strongest predictors of university performance by collected data.

Research Problem: 1.2

the amount of data stored in educational database increasing rapidly. These databases contain hidden information for improvement of student' s performance. This makes problem to analyze and prediction from this large and complex database. Discover hidden patterns that in the educational record.

Research Main Objectives: 1.3

Use data mining classification algorithm to analysis large data that stored in large and complex databases

Sub objective:

I would like to analyze data registered in the University of Technology & science to try forecast the students performance in our college using the data between 2008 and 2010 the data used to create our database include: GPA, recommendation and subject.

- Generation of a data source of predictive variables.
- Identification of different factors, which affects a student' s learning behavior and performance during academic career.
- Construction of a prediction model using classification data mining techniques on the basis of identified predictive variables.
- Validation of the developed model for students.

1.4 Methodology:

The data mining developing software tool are: MYSQL DATABASE ,EXCEL ,MS

ACCESS,SPSS,METLAB,WEKA,TANGAR,WEBMINER,V.B 6.0.

Classification data with WEKA is the best data mining tools. It can be loaded the data from various sources, including files URLs and databases. Supported file formats include WEKA own ARFF format, CSV, Lib SVM format. It is also possible to generate data using an artificial data source and edit data manually using a dataset editor. Doing the classification and regression algorithms to the data bring from the data base.

Research organization: 1.5

- Chapter one is introduce the problem of data in excel and how to solve this problem data mining classification and regression tree CART.
- Chapter Tow introduce to data mining and data mining tools like classification, Association rules and clustering.
- Chapter three is describe the CRISP model its important model in data mining process.
- Chapter four describing how the classification and regression tree CART work.
- Chapter five for related works paper of predicts student performance by CART algorithm.
- Chapter six for implementing CART algorithm in WEKA.
- Chapter seven for conclusion and recommendation.

