

UNIVERSITY OF SCIENCE AND TECHNOLOGY
COLLEGE OF GRADUATE STUDIES AND ACADEMIC
ADVANCEMENT

Faculty of Computer Science and Information Technology

Model for Prediction final *Grade Point Average*
(Case Study: University of Science and Technology)

By

Mohammed Abd El_Hadi Mohammed_Khair

A Thesis

Submitted to the College of Graduate Studies and Academic Advancements
in Partial Fulfillment of the Requirement for the Degree of Master of
Science in Information Technology

Supervisor

Dr. Abdelhamid Salih

October 2016

Abstract

Data mining is a process of discovery important information and extract useful patterns from large collections of data. Educational data mining helping universities management ,teachers and students to improve students performance. Classification is a data mining (machine learning) technique used to predict group membership for data instances and its used when we predict some things by using the previous available information. In this research we investigate to build a model to predict the students final GBA based on their grades in specialization courses, the data sets of students were obtained from the University of Science and Technology (sudan). We applied three algorithms J48, JRIP and PART to know the best algorithm to use it on datasets. This model was built using Weka tool is the most famous tool used in data mining and extraction of important patterns. We found that j48 is the best algorithm because it has highest accuracy (Precision) among of the three algorithms. After investigating the experiments from the result we obtained the prediction model help university management and students to know important courses that affect on students final GBA that which helping to improve students performance and help the teachers and the university administration to improve the teaching process.

المستخلص

تنقيب البيانات هو عملية إكتشاف المعلومات المهمة واستخراج الأنماط المفيدة من مجموعات كبيرة من البيانات. تنقيب البيانات التعليمي يساعد إدارة الجامعات والأساتذة والطلاب لتحسين مستوى أداء الطلاب. التصنيف هو تقنية تنقيب البيانات (تعلم الآلة) تستخدم للتنبؤ بأعضاء المجموعة لحالات البيانات والتي تستخدم عندما نتوقع بعض الأشياء عن طريق استخدام المعلومات السابقة المتوفرة. في هذا البحث بحثنا لبناء نموذج للتنبؤ بال GBA النهائي للطلاب على أساس درجاتهم في دورات التخصص، وقد تم الحصول على مجموعات بيانات الطلاب من جامعة العلوم والتقانة (السودان). طبقنا ثلاثة خوارزميات (J48, JRIP, PART) للتعرف على أفضل خوارزمية لاستخدامها على مجموعات البيانات. تم بناء هذا النموذج باستخدام أداة WEKA وهي الأداة الأكثر شهرة المستخدمة في تنقيب البيانات واستخراج الأنماط الهامة. وجدنا أن J48 هي أفضل خوارزمية لأنها تحتوي على أعلى دقة (Precision) بين الخوارزميات الثلاثة. بعد البحث في نتائج التجارب حصلنا على نموذج للتنبؤ يساعد إدارة الجامعة والطلاب على معرفة الدورات الهامة التي تؤثر على ال GBA النهائي للطلاب مما يساعد على تحسين أداء الطلاب ومساعدة المعلمين وإدارة الجامعة لتحسين العملية التعليمية.



1.1 Introduction

Predicting student performance is the main objective of higher education institutions to provide quality education to its students, improve student performance by extract knowledge from the students database to discover and get useful patterns and important information can be used in decision support, forecasting, prediction and assessment.

The primary goal of data mining is to discover knowledge from the huge amount of data , There are different ways used for knowledge discovery from large data such as classification, clustering, prediction, association rule etc. In data mining, classification is a simple technique to extract data that is used to group of pre-sorted for example to build a model that can classification hug data set, using a decision tree and neural network classification is performed on a huge amount of data. Educational data mining (EDM) provides various methods for prediction of student`s performance, which improve the future result of student [1].

Data mining is a powerful technique for the analysis of important information from the data warehouse. It is the extraction of predictive information hidden to improve the decision-making process. The main objective of Educational data mining is to improve the academic performance of students. Educational data mining techniques has become a rapidly growing in the field of education. It helps to analyze student activities and improve their performance [2].

Classification is Consists of two steps: supervised learning of a training set of data to create a model, and then classifying the data depending to the model. Decision trees are a top-down approach to classification that divides the data into leaf and node divisions until the entire set has been analyzed [3].

The decision tree is a technique to help in decision-making which provide high classification accuracy with a simple representation of collected KDD. It help experts to validate and classify the results and outcomes of tests and analyze various new symptoms of diseases based on data [4].

1.2 Problem Statement

The problems that facing universities and students is a problem of prediction academic performance of students and predict final students final GBA. Universities don't have the prediction model, so it need prediction model to discover the important courses that affect on students results and improve the students performance and to identify students who need special attention and help the university to focus on courses that affect the students final GBA and increase educational dose of these courses.

1.3 Research Objective

This research aims at the following:

1. To collect the data set from the University of Science and Technology/ Faculty of Computer Science and Information Technology / Department of Information Technology/ Bachelor of Information Technology so that we could build a predictive model.
2. To build a predictive model using classification technique to help in predicting the performance of the students by discovering of the the most important course in the each academic year and discovering of the most important course which affect on the student final GBA in all the years of study.
3. To evaluate the model through the use of several different method.

1.4 Research Methodology

Build a model through the collect the data set and then processed and prepared the data, and use Weka data mining tool and prediction via classification data mining technique through the use of J48 algorithm, JRIP algorithm and PART algorithm, evaluation process, and get the Results and Discussion and finally obtain prediction model.

1.5 Research Structure

This research contains five chapters Chapter two "Literature Review" we discuss data mining and Classification background and related work. Chapter three "Methodology" we describe research methodology. Chapter four "Experimental and results" we explain the results of the experiment and discussed. Chapter five "Conclusions and recommendations" we show Conclusions and recommendations and references.