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

Faculty of Computer Science and Information Technology

Educational Data Mining Using Classification Technique (J48 Algorithm)

by
Mayada Mohamed Abdalla

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. Atif Ail Mohmmed

September 2016

Abstract

Educational data mining (EDM) is a new growing research area and the essence of data mining concepts are used in the educational field for the purpose of extracting useful information on the behaviors of students in the learning process.

Educational data mining aims to analyze the students' performance, understand learning behaviors, and highlight factors that affect learning process in a particular educational system in the purpose of increasing passing ratio for the students.

There are huge data in academic institutions, this study use it to improve the performance of students and take academic decisions.

J48 Algorithm has been used, which is one of the classification methods for data mining in academic institutions algorithm, and we want to use it to improve student academic performance and take decisions.

The main goal here is to use classification algorithms in data mining techniques for build a data model to help academic institutions in support of their decisions, improve the academic performance of students. And Measure prediction accuracy compared to the actual results of students.

After conducting experiments outcome reached is a model to predict who can help academic institutions in decision-making and improve the performance of students academically.

المستخلص

تنقيب البيانات التعليمية هو مجال جديد في الأبحاث المتنامية وجوهر مفاهيم تنقيب البيانات يستخدم في المجال التعليمى لغرض استخراج معلومات مفيدة عن سلوكيات الطلاب في عملية التعلم.

ويهدف تنقيب البيانات التعليمي لتحليل أداء الطلاب، فهم سلوكيات التعلم، والتركيز على العوامل التي تؤثر على عملية التعلم في نظام تعليمي معين بغرض زيادة اجتياز نسبة الطلاب.

هناك بيانات ضخمة في المؤسسات الأكاديمية ونريد أن استخدامها لتحسين أداء الطلبة واتخاذ القرارات الأكاديمية.

تم استخدام خوارزمية(J48) التى هي واحدة من طرق التصنيف لتنقيب البيانات في المؤسسات الأكاديمية، ونحن نريد استخدام هذه البيانات الضخمة لتحسين أداء الطلبة واتخاذ القرارات الأكاديمية.

الهدف الرئيسي هنا هو استخدام تنقيب البيانات لبناء نموذج بيانات للمؤسسة الاكاديمية يمكنها الاستفادة منها في دعم قراراتهم أو لتحسين الأداء الأكاديمي للطلاب أكاديميا وباستخدام خوارزميات تصنيف وقياس دقة التنبؤ مقارنة بنتائج الطلاب الحقيقية.

بعد اجراء التجارب النتيجة التي تم التوصل اليها هي نموذج للتنبؤ الذي يمكن أن يساعد المؤسسات الاكاديمية فى اتخاذ القرار وتحسين اداء الطلاب اكاديميا.

1.1 Introduction

Data Mining is used to extract meaningful information and to developed significant relationship among variables stored in large data repositories. Educational data mining is an emerging discipline concern with developing methods for exploring the unique type of data that come from educational setting and using those methods to better understand students and the setting which they learn in as defined by the educational data mining community. Education is an essential element for the progress of country. Mining in educational environment is called educational data mining. It is concerned with developing new methods to discover knowledge from educational database. Educational data mining provides a set of techniques, which can help the educational system to overcome these issues. The objective of this research is to introduce educational data mining by describing step by step process using technique of C4.5 (Classification Methods).

This study will help the academic institutions to reduced drop-out ratio to a significant level and improve the performance of students.

1.2 Research Problem

In academic institutions data repositories, there are a large amount of academic data for students, which represents the academic results of previous years.

Analyzing this data using data mining technique can be useful to support Academy decisions or decisions that could lead to improved student performance, which difficult to be done using traditional tools.

The problem which this research aim to find out its solution is how to use a data mining tools in this huge data in academic institutions for the purpose of improving the performance of students and help academic decisions.

1.3 Research Objectives

The main objective of this research is to use data mining technique to build classification model for academic institution which can help to support decisions or to improve the academic performance of students using classification algorithms J48.

1.4 Research Methodology

The methodology used in this research based on the following topics:

1. First of all a sample of data was collected from the University of Science & Technology.
2. The data was preprocessed using cleaning data reduction techniques.
3. converted into CSV format and then to ARFF format .
4. Implement four experiments using WEKA machine learning tool to classify the processed data sample using a decision tree algorithms (J48).
5. The result of experiments are compared with actual students' academic results to decide whether the prediction is carried out correctly or not.

1.5 Thesis Structure

Chapter two explain data mining concepts, Process KDD, Data Mining Applications, Data Mining Life Cycle, Educational data mining, Phases of Educational Data Mining, and Related Work. Chapter three discuss Data Mining Methods, Components of Data Mining Algorithms, Classification, Classification Techniques, (C4.5), and pseudo code. Chapter four the Implementation. Chapter five conclusion and recommendations.