

Performance using  
data mining  
techniques education  
essay



**ASSIGN  
BUSTER**

The Internet has become a common medium that improves the education. E-learning is a process in which education is in digital learning method. E-learning mainly focuses on learner-centric training rather than teacher-centric training, which has been in practice in traditional teaching. The amount of data stored in educational database increasing rapidly without any benefit to the management. These databases contain hidden information for improvement of students' performance. In our research, we will use educational data mining to discover useful knowledge from graduate students data, these data will collect from faculty of Science and Technology - university science Islam Malaysia (USIM). We will apply data mining techniques such as association, classification, clustering and outlier detection rules. In each of these four tasks, we will extract knowledge that describes and improve students' performance.

Keywords: E-Learning, Students' Performance, Data Mining.

## **Introduction**

The Internet has become a pervasive medium that has changed completely the education environment, and the way of knowledge that shared throughout the world. It provides an easy way to search and access to any information that you need. E- Learning has made knowledge accessible to a large number of people. Kumar, Bharadwaj and Pal (2012). The students' performance plays an important role in producing the best quality graduates and post-graduates who will become great leader and manpower for the country thus responsible for the country's economic and social development. The performance of students in universities should be a concern not only to the administrators and educators, but also to corporations in the labour

<https://assignbuster.com/performance-using-data-mining-techniques-education-essay/>

market. Academic achievement is one of the main factors considered by the employer in recruiting workers especially the fresh graduates. Thus, students have to place the greatest effort in their study to obtain a good grade in order to fulfil the employer's demand.

There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, Data mining concepts and techniques can be applied in E-Learning to discover knowledge that comes from educational environments. There are many Data Mining techniques such as Naïve Bayes, Neural Networks, Decision Trees, K-Nearest neighbor, and many others, to offer useful knowledge about the learning process for instructors. Heikki, Mannila (1996) Knowledge discovery in databases (KDD), often called data mining, aims at the discovery of useful information from large collections of data. U. Fayyad, Piatetsky, G. Shapiro, and P(1996). The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data.

Data mining techniques are used to extract or "mining" knowledge and discover hidden patterns from large volumes of data, this knowledge is helpful in decision making, data mining is actually part of the knowledge discovery process Baradwaj, B. and Pal, S. (2011). Mining online learning events is becoming promising area for research and development, particularly when the business in education is growing impressively Margo Hanna (2004). Higher education is become a big business, with growing of IT technology supporting online learning. The world is going fast towards online learning, so we can see a lot of open universities that provide courses online <https://assignbuster.com/performance-using-data-mining-techniques-education-essay/>

through the Internet. So one can study and take the exam and get certified whenever and wherever he/she. According to Porter and Cunningham (2005), data mining is the process of extracting useful information from any form of data, with emphasis to numeric data.

## **Literature review**

Romero, C. and Ventura, S. (2007), have a survey on educational data mining between 1995 and 2005. They concluded that there is growing interest in data mining and the evaluation of online educational systems, educational data mining a rising and promising area of research. Al-Radaideh, Q., Al-Shawakfa, E. and Al-Najjar, M. (2006). applied the data mining techniques to evaluate student data and study the main attributes that may affect the student performance in courses. The extracted classification rules are based on the decision tree as a classification method. It allows students to predict the final grade in a course under study. Baradwaj, B. and Pal, S. (2011), used the classification task to evaluate student' performance, they used decision tree method for classification. The goal of their study is to extract knowledge that describes students' performance in end semester examination. This study helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising.

Chandra and Nandhini Chandra, E. and Nandhini, K. (2010), applied the association rule on students' failed courses to identify the reason for students' failure patterns courses and suggest relevant causes of the failure to improve the low capacity students' performances. It also reveals some hidden patterns of students failed courses which could serve as base for <https://assignbuster.com/performance-using-data-mining-techniques-education-essay/>

academic planners in making academic decisions and an aid in the curriculum re-structuring and modification, it will help to improve students' performance and reducing failure rate. Mohammed M and Alaa M (2012), applied educational data mining to extract useful knowledge from graduate students data to improve graduate students' performance, and overcome the problem of low grads of graduate students. Han and Kamber (2000) describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process. Pandey and Pal (2011) conducted study on the student performance based by selecting 600 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayes Classification on category, language and background qualification, it was found that whether new comer students will performer or not. Galit (2007) gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams.

Hijazi and Naqvi (2006) conducted as study on the student performance by selecting a sample of 300 students (225 males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as " Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance.

Z. J. Kovacic (2010) presented a case study on educational data mining to identify up to what extent the enrolment data can be used to predict student's success. The algorithms CHAID and CART were applied on student enrolment data of information system students of open polytechnic of New Zealand to get two decision trees classifying successful and unsuccessful students. The accuracy obtained with CHAID and CART was 59.4 and 60.5 respectively. Ayesha, Mustafa, Sattar and Khan (2010) describe the use of k-means clustering algorithm to predict student's learning activities. The information generated after the implementation of data mining technique may be helpful for instructor as well as for students.

## **Research problem**

Students are the main assets of educational institutions. The students' performance is an important factor to produce the best quality graduates. Students have to place the greatest effort in their study to obtain a good grade to achieve their academic carrier. Many factors could act as barrier and catalyst to students, these factors will reflect on student to achieve a high grade or low grade.

USIM does not use any knowledge discovery process approach to get knowledge about student' performance, decision making in educational system need to use these knowledge:

To identify the weak students and help them to get better marks.

To reduce the problem of dropouts and low grade of graduate students.

## **Scope Of Research**

The data set will use in this research contains master students information will collect from the Faculty of Science and Technology (FST), USIM for a period of five years in period from 2007 to 2012. We will use data mining techniques on these data to extract helpful knowledge can be used to improve students' performance and graduate quality.

## **Research Objective**

To study students' performance in e-learning using data mining techniques

Discover hidden information of students' performance.

To identify the weak students and help to score better mark and reduce fail ratio.

To improve the performance of the students.

## **Research Outcome**

Provide to the decision maker helpful constructive recommendation to overcome the problem of low grade of graduate students. Improve students' academic performance.

## **Research Question**

What is the current level of graduate students at USIM?

Do USIM' decision maker have a good knowledge about students' performance?

Do students need special attention to improve their performance?

## **Research Methodology**

The methodology used in this research starts from the problem definition, then data collection and preprocessing, then we come to the data mining methods which are association, classification, clustering, and outlier detection, followed by the evaluation of results and patterns, finally the knowledge representation process.

Understanding the domain and problem definition.

Data collection and preprocessing: the data will be collected from Graduate Students data base, and the data will clean and transformed into a mineable format.

Apply data mining methods: the data mining methods will apply to discover and summarize knowledge about student performance as follow:

### **Association Rules**

Mining association rules searches for interesting relationships among items in a given data set.

### **Classification**

Classification is a data mining task that predicts group membership for data instances. The classification approaches are used to predict the Grade of the graduate student and how other attributes affect them.

We will use the follow method to predict the grade:



Decision Tree based Methods

Rule-based Methods

Neural Networks

Naïve Bayes and Bayesian Belief Networks

And we will compare between the result that we get from these techniques and we will try to identify which one is better to use because many researches using decision tree and they advice to use other techniques.

## **Clustering**

Clustering is a data mining task that finds groups of objects so that objects that belong to one cluster are more similar to each other than to objects belonging to different cluster

Followed by the evaluation of results and patterns.

Finally the knowledge representation.

## **Conclusion**

In this research we will discover hidden information from graduate student data collected from the Faculty of Science and Technology in USIM. The data include five years from 2007-2012. This data existing and available in the USIM database, particularly we will use association rules to discovering, Rule Induction, Neural network, Naïve bayes, and Decision Tree to predict grade of students. The result will contain helpful knowledge, this knowledge can provide to decision maker to improve the graduate quality.

## Research work plan

The work plan for this research will be as following table and figure:

**Table 1: The Gantt Chart Description**

**No**

**Task**

**Start**

**Duration**

**Date Start**

**Date End**

1

Project Title

0

6

9/15/2012

9/20/2012

2

Proposal Writing , Submission & Presentation

7

85

9/21/2012

12/14/2012

3

Literature Review (students' performance - e learning - data mining)

92

32

12/15/2012

1/15/2013

4

Data collection & preprocessing (Selction - cleaning - transformation)

124

76

1/16/2013

4/1/2013

5

Apply data mining methods

200

75

4/2/2013

6/15/2013

6

Paper 1 Writing & Submission

275

22

6/16/2013

7/7/2013

7

Paper 1 Presentation (conference article)

297

1

7/8/2013

7/8/2013

8

First Three Chapters Writings

298

55

7/9/2013

9/1/2013

9

Proposal Defense Presentation

353

29

9/2/2013

9/30/2013

10

Paper 2 Writing & Submission

382

27

10/1/2013

10/27/2013

11

Paper 2 Presentation (conference article)

409

1

10/28/2013

10/28/2013

12

Evaluation of results and patterns

410

74

10/29/2013

1/10/2014

13

Paper 3 Writing ( Journal article)

484

31

1/11/2014

2/10/2014

14

Paper 3 Presentation

515

1

2/11/2014

2/11/2014

15

Thesis writing

516

195

2/12/2014

8/25/2014

16

Submission

711

1

8/26/2014

8/26/2014

17

Viva

712

97

8/27/2014

12/1/2014

**\*Red color date is Milestones targeted**

**Figure 1: Gantt Chart**