

Sudan University of Science & Technology

جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا

**Use of Data Mining Techniques to Find Relation between  
Attendance and Results in Learning**

إستخدام تقنية تنقيب البيانات للمقارنة بين الحضور ودرجات التحصيل في  
التعليم

بجث تكميلي لنيل درجة الماجستير في تقنية المعلومات

**By:** Rafea Abdelkafi Magzoub

**Supervisor:** Dr. Nisreen Beshir Osman

2015

i

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

{أَقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ} 1 {خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ} 2 {أَقْرَأْ وَرَبُّكَ الْأَكْبَرُ} 3 {الَّذِي عَلَّمَ بِالْقَلَمِ} 4 {الْإِنْسَانَ أَنَّهُ الْكَلِيمُ الْعَظِيمُ} 5

صدق الله العظيم

سورة العلق الآية ( 1- 5 )

وقال تعالى :

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

1 {يُنْفِقْ كَرِيماً} 2 {وَسِعَ كَرَمُكَ الْمَشْرِقَ وَمَغْرِبَ} 3 {لَيْسَ لَكَ تُسُوفٌ} 4 {فَلَيْسَ يُنْفِقُ فِيهَا أَنفَاهُ} 5 {الَّذِي كَلَّمَكَ اللَّهُ} 6 {ذُنُوباً} 7 {إِلَّا هُوَ لِيَأْتِلْهَا} 8 {اللَّهُ} 9 {وَيُعَدُّ} 10 {سُورٍ} 11 {سَوَاءً} 12 {}

صدق الله العظيم

سورة الطلاق (الآية 7)

## خلاصة البحث

يعتبر التنقيب عن البيانات أو ما يعرف بإستخلاص البيانات أو إكتشاف المعرفة من الطرق المثلى لتحليل البيانات من وجهات نظر مختلفة. ويرجع ذلك إلى المعلومات التي يتم الحصول عليها.

نظام التنقيب عن البيانات لا يستخدم فقط كأداة لتحليل البيانات، بل قد يتيح للمستخدم التحليل من عدد من الزوايا المختلفة وتلخيص العلاقات وإيجاد الارتباطات أو الأنماط بين عشرات الحقول.

هذه الأطروحة البحثية تقوم بالمقارنة بين التعليم الإلكتروني والتعليم التقليدي من خلال بيانات الطلاب.

تم إختيار الإداة "ويكا" لإجراء عملية التجميع والمقارنة بين الفئتين بإستخدام خوارزمية الـ K-means حيث أنها تستخدم تقنية التنقيب، كما تم إستخدام مايكروسوفت إكسل لحفظ البيانات. وخلصت النتائج الى ان التعليم الالكتروني افضل من التعليم التقليدي.

## **Abstract**

Data mining or data extraction is one of the best ways to analyze data from different points of view based on the obtained information. It is also known as discovery of knowledge

Data mining system is used as an analytical tool to analyze the data, as well as allow the user to analyze a number of different angles. Moreover, it summarizes the relations and finds the correlations or patterns among dozens of fields. In this research, data mining technique is used to compare between the traditional learning and electronic learning using student's dataset, from x university.

The tool that has been selected is WEKA. Clustering methods was used to compare between the two groups using K-means algorithms, also Microsoft Excel spread sheet was used to save the data. The results showed that the e-learning is better than the traditional learning.

## **Acknowledgement**

First without comparing, all thanks to the God, who help us to complete this research.

And second I would like to express my special appreciation and thanks to my supervisor *Dr. Nisreen Beshir Osman*, who have been a tremendous mentor for me. I would like to thank her for her encouragement through the research process till completion and her advice that have been priceless.

A special thanks to my family the words cannot express how grateful I am to my father, my mother for all of the sacrifices that offered to me, also thanks to my (brothers, sister).

And special thanks for each one who supported me and for those who helped me writing this research especially my friends.

Last but not least to the reader these lines, hoping to increase through the project Square tender.

# Contents

<b>1. Introduction</b>	<b>1</b>
1.1 Introduction .....	1
1.2 Statement of Problem.....	1
1.3 Research Objectives .....	2
1.4 Research Structure .....	2
<b>2. Literature Review</b>	<b>3</b>
2.1 What is a Data Mining? .....	3
2.1.1 Data Mining Model.....	3
2.1.2 Data Mining Tasks.....	4
2.1.3 The Component of Data Mining .....	5
2.1.4 Data Mining Process.....	6
2.1.5 Data Mining Techniques....	7
2.1.6 Challenges in Data Mining.....	9
2.2 Clustering .....	9
2.2.1 The Goal of Clustering.....	10
2.2.2 Possible Application of Clustering.....	10
2.2.3 Requirement of Clustering.....	11
2.2.4 Major Clustering Methods.....	12
2.2.5 Problems in the Clustering.....	15
2.3 Related Works.....	16
<b>3. Methodology</b>	<b>18</b>
3.1 Introduction .....	18
3.2 Developing an Understanding of the Application Domain.....	18
3.3 Preprocessing and Creating Operation Database.....	18
3.4 Data Mining.....	23
3.4 Interpreting Mined Pattern.....	23
<b>4. Result and Discussion</b>	<b>24</b>
4.1 Introduction.....	24
4.2 First Experiment.....	24

4.3 Second Experiment .....	26
4.4 Third Experiment.....	27
4.5 Results .....	28
4.6 Discussion.....	28
<b>5. Conclusion and Recommendation</b>	<b>29</b>
5.1 Conclusion.....	29
5.2 Recommendation.....	29
<b>References</b>	<b>30</b>

## **List of Tables**

2.1 Classification Algorithms.....	8
2.2 Clustering Algorithms.....	9
3.1 Normalize .....	22
3.2 Dataset of case study description .....	23
4.1 Result of the first experiment .....	26
4.2 Result of the second experiment .....	27
4.3 Result of the third experiment .....	28

## **List of Figures**

2.1 Data Mining Model .....	4
2.2 Steps of KDD .....	5
2.3 Clustering Analysis .....	10
2.4 K-means Lifecycle .....	13
3.1 Knowledge discovery from data in data mining.....	18
3.2 Samples from Dataset .....	20
3.3 Remove irrelevant Attributes.....	21
3.4 Handling Missing Values.....	22
3.5 Dataset CSV file .....	24
4.1 Result of the first experiment .....	26
4.2 Result of the second experiment .....	27
4.3 Result of the third experiment .....	28