

## ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

USING DATA MINING TECHNIQUES TO EXPLORE PATTERNS OF ACADEMIC  
ACHIEVEMENT EFFECTS FOR HIGH SCHOOL STUDENTSKarrar Hussein ALI<sup>1</sup><sup>1</sup>Altınbas University, Institute of Graduate Studies, Department of Information Technology, Istanbul.  
kmlove196@gmail.com ORCID No: 0000-0002-6924-739XSefer KURNAZ<sup>2</sup><sup>2</sup>Altınbas University, Faculty of Engineering and Natural Sciences, Department of Computer  
Engineering, Istanbul.  
mailto:sefer.kurnaz@altinbas.edu.tr ORCID No: 0000-0002-7666-2639

GELİŞ TARİHİ/RECEIVED DATE: 05.02.2020 KABUL TARİHİ/ACCEPTED DATE: 17.05.2020

**Abstract**

This research presents an applied study of the field of knowledge discovery of educational data using data mining techniques, focusing on the development of teaching and learning, by discovering the main patterns of testing the data of the academic student of the intermediate level (baccalaureate) in Baghdad - Iraq, from 2010 to 2019 to get results on the Academic Performance Index. In this study, we discover some major patterns data, Some of this patterns association exists between student changed summation and the student gain level for some subjects, also the relation between summation degrees with degree gained from some subject. This research attempts to read this result and interpretation, supply and verification level and its type to supply to the ministry decision-maker. We choose data mining technique because it's better to use the benefit of quantity data; we use a different way from data mining technique to support discovery result clusters using (k-means) and classification use a decision tree, after first pre-processing data for database and restriction like logical data warehouse shape, we use k-means algorithm of clusters technique and (J48) algorithm of the classification technique of the decision tree, this different way and algorithms application through WEKA tool, which supports more algorithms and way of data mining Last deductive abstract and suggests some recommendation which interest for the decision-maker. Results from this research built a logical data warehouse & applying the algorithm of data mining's algorithm, besides the difficulties of some subjects which may form its tough words or other disqualification planning

**Keywords:** Data mining; Predicative models; Classification; Decision tree; Performance prediction.

**LİSE ÖĞRENCİLERİNİN AKADEMİK BAŞARIYA ULAŞMA DAVRANIŞLARININ VERİ  
MADENCİLİĞİ YÖNTEMLERİ İLE İNCELENMESİ****Özet**

Bu araştırma, veri madenciliği tekniklerini kullanarak bilginin keşfinde uygulamalı bir çalışma sunuyor, Bu çalışmanın temel amacı, 2010'den 2019'e kadar olan üçüncü Orta Sertifika - Bakalorya için öğrencilerin

akademik verilerinde mevcut bazı kalıpları keşfetmektir. Daha sonra Irak Eğitim Bakanlığındaki karar vericilerin eğitim politikalarını desteklemek için akademik performansla ilişkin genel göstergeler geldi. Özellikle veri hacminin yanı sıra, bu verilerin nispeten büyük zaman boyutu da arama sonuçlarından destek aldığından. Bu araştırmada, bu verilerde baskın olan bazı veri kalıplarının, eğitim açısından önemli göstergeler sağlayabilecek bir dizi örgütün varlığına göre özetlendiğini keşfettik. Bu kalıplardan, öğrencinin genel ortalaması ile bazı derslerin öğrenci başarısı arasında bir korelasyon vardır ve bazı derslerde elde edilen tahminde öğrenci başarısının değerlendirilmesi arasındaki ilişki vardır. Bu araştırma, bu sonucu okuma, yorumlama ve sunum seviyesini okumaya ve seviyesini ve kalitesini bakanlıktaki karar vericiye sunarak seviyesini doğrulamaya çalışır. Veri madenciliği teknikleri, bu verinin boyutundan yararlanmak için en uygun olarak seçilmiştir ve çünkü karar vermeyi desteklemek için sıklıkla kullanılan akıllı tımdengelim algoritmaları kullanılmaktadır. Bulguları desteklemek için farklı veri madenciliği teknikleri metotları kullanıyoruz, yani küme oluşturma ve sınıflandırmada k-means algoritmasını kullanarak küme oluşturma işleminin ardından (logical data warehouse) veritabanının ilk işlenmesi ve yeniden yapılandırılmasından sonra karar ağaçları kullanılarak kümeleme. Kümeli teknolojiye K-means algoritması, karar ağacı için sınıflandırma tekniğindeki algoritma (J48), bu yöntemler ve algoritmalar weka aracı kullanılarak uygulanmıştır. Veri madenciliğinde birçok algoritma ve yöntemi destekleyen. Araştırma sonuçlarına göre mantıksal bir veri ambarı ve bazı veri madenciliği algoritmalarının Milli Eğitim Bakanlığı veri tabanına uygulanması, Öğrenci kayıtlarıyla ilgili diğer önemli sonuçlara ek olarak, birçok dersin öğrencileri bırakması ve çalışma dışı bırakması gibi, Çalışma planları ve müfredatta ya zor bir kelime ya da kusur.

**Anahtar Kelimeler:** Veri Madenciliği, Sınıflandırma, Karar ağacı, Performans tahmini, Tahmini modeller

## 1. INTRODUCTION

In today's world, the age of knowledge explosion is surrounded by a huge amount of information in various aspects of life and different images, and information has become the most important components of contemporary life. This massive flow of information is due to the development of transport and information processing of communication devices and computers. Information, computers and electronic communication systems, automation systems or computerized computer-based systems applications that have proven to be better than traditional systems, Important. in several aspects such as the application of e-learning. Currently, the use of e-services is one of the most important in most government institutions. Sahay, S. K et.al, NGOs with these huge sizes of data, orthodox methods of data analysis, which are a mixing of statistical methods and somewhat computer systems designed. to manage databases, deal with this type of data. Among the most recent exercised sciences in this field are KDD in databases at the top of DM and discovery science in discovery of knowledge of data in databases (KDD) Fayyad, et.all. as an attached image. In academic and academic institutions, scientific knowledge can be used to explore knowledge and explore data to improve academic performance by devising these data types of institutions for students and faculty Márquez-Vera, et.all. This is related to key performance indicators such as student achievements. duration rate, and staff fineness. Educational institutions, for example, can portend students who teardrop out students who will be students who have low academic accomplishment will graduates and other strategically information, and then can survey and develop educational politicking to help these student upgrade their education or guide them to disciplines That fit with their abilities, preparations, preferences, abilities and other politics and measures that will mend academic showing at the educational institution Prajapati, M. M, et.all.

## 2. PROBLEM OF STUDY

The low level of educational attainment is a big problem that needs to be solved. It is multidimensional problem that is sometimes psychological and sometimes a social problem. The problem of the study is determined by finding the factors that lead to the low achievement in the grades of students in the stage of education in the baccalaureate exams. The third average in Iraq.

## 3. LITERATURE REVIEW

This section presents a series of previous studies that are directly and indirectly related to the topic. Many studies around the world have applied data mining algorithms to discover knowledge. Educational applications for data mining in education, focusing on inputs and outputs of the educational process. And how they affect each other:

Osmanbegovic, Edin, Suljic, Mirza, They collected student data during the summer semesters of the 2010-2011 academic year from the University of Tuzla, university of Economics, between first-year student and data taken through enrollments. Success was evaluated with the success grade of the exam. Student influence of social and demographics variable, result from high school and from the entrance exam, and attitude towards study that can have an impact on success. The purpose of the study was to find the best technique for portend student performance. Osmanbegovic, E., & Suljic, et.all .

And Alom, et al., (2018) trajectory's the students of Australia inception from their elementary school year 1 to accomplishment of their high school successfully and furthermore it too trajectory's their consent into the higher educational universities or foundations. The criteria on which they are calculating the successive rates of the students is the gender of the students. In this paper data mining software Wilson calculator is used which is a practical meta-analysis effect size calculator and Orange so as to analyze data. Orange for given data sets provide the predictive modeling and the visualizations solutions. Alom, B. M. M., & Courtney, et.all.

In 2013 Marquez Vera and Sergey Rovira proposed a method, and they used real data for a group of 670 students from the Zacatecas School, located in Mexico and data from the University of Barcelona. They used the classification method, machine learning in the white box, induction rules and decision tree algorithms. Three hypothesis-based trials were conducted to find the failure rate and the dropout rate at school. They used the method of selecting features, out of seventy-seven traits, only fifteen were considered, which is considered the best in education systems. Dimensional modeling and statistical techniques are also implemented in this work. The Weka tool is used, and the results are not represented in graphical form for a better understanding. The problem of data imbalance is resolved efficiently. In order to predict student performance based on pre-university and personal characteristics. Márquez-Vera, et.all

N. Rachbure and W. Punlumjeak suggested a comparison of four methods for selecting IG, GAs, SVM, Min-Red and Max-Rel features with four supervised technologies: KNN, DT, NB and NN. They found that Max-Rel and Min-Red are the best method with a accuracy of 99.12% using KNN. N. Hafieza, AA Aziz and Ahmad of UniSZA proposed a framework to predict the academic performance of students for undergraduate

students in the first year of the CS course between July 2006/07 and July 2013/14 using the DT, NB and RB rating and found that RB is the best model with a 71.3% accuracy Aldikanji, E., & Ajami, et.all.

Asif, R., et al (2017) Studies the performance related to the education of the students. The data taken of the students is focused on two aspects. Firstly, achievement of the student is predicted at the completion of the four year study programme. Then predictions are combined with the progressions. The outcomes is the generation of two groups of students, low and high grade achieving students. By this teachers get to support the students at the low level by giving them activities and the task and to the high level students more opportunities are given Asif, R., Merceron, A., Ali, et.all.

Kabakchieva used a set of educational data to analyze by using data mining algorithms. He used two techniques: Decision Tree and Bayes. Using the Weka application tool, data for 10,330 students and 20 teachers were collected and considered. Weka classification filters are implemented on datasets by an algorithm, where JRip and J48 provide accurate and reliable results. Bayes workbook and kNN workbooks are inaccurate Kabakchieva, et.all.

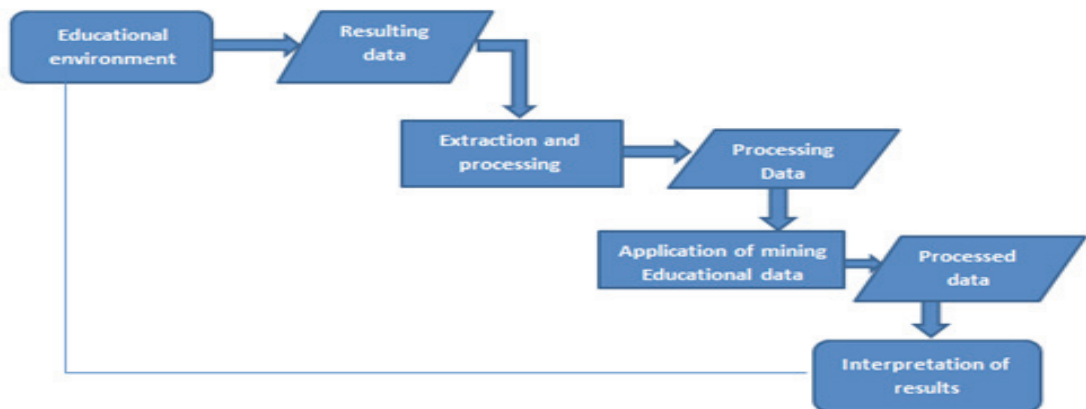
Another method proposed by Ahmed and his colleagues is used to predict student performance using the Weka tool by applying the included decision tree techniques, artificial intelligence, classification methods, neural networks, grouping, regression, and associate rules. ID3 methods have been used. On a set of data consisting of 1547 record used to predict performance. Used to implement decision tree. This method does not verify attributes such as mood, presence, and environmental factors. Another proposed model is based on longitudinal data derived from data from Gwinnett County public schools, and students who entered the 8th grade assessment in science and mathematics are implemented by Tamhan and his colleagues Aunsan, S., & Thammaboosadee, et.all. For missing values, the use of logistic regression, dependency, and Bayes classifier and decision trees means implementation techniques. They use the SPSS and weka designer taking into account demographics. From all these techniques logistic regression provides the best results. There was a problem filling the missing values. Produces noise forecasts and not a single workbook is enough for entire student data. Aggregation should be a practical option for optimizing risk prediction of student performance. Behavioral data and relevant recording factors should also be considered.

Tavares, R., et al,(2017) The prime focus of this paper is to enhance the digital educational resources in primary school for science education by implementing on the data mining techniques. There is a lot of impact on the students for the self regulated learning after adopting the learning approach. The analyses of the students behaviour is done after getting the particular help and the recommendations Tavares, R., Vieira, R., & Pedro, et.all.

#### **4. ABOUT EDM**

The term spread with the first workshop on the concept of EDM in 2005, which since 2008 has become an annual conference. A periodical has been created to publish the latest research on EDM Dellinger Dissertation,et.all.( EDM) One of the most important systems promoted, by the state and society is the educational system, so most of the scientific researches and studies, have spread to find ways to develop

the educational systems. EDM is represents a bridge between education, and computer science. Special sections are used for computer science, DM and machine learning. DM is used to detect hidden patterns in irregular data and turn them into regular and useful data. EDM is an emerging propriety that is concerned with develop methods to explore the unique type of data taken from educational environments and using these techniques to better understand studentes and the environments in which they are learning. Extraction of student performance records stored on the computer is on of the core areas in this discipline, and the extraction of registration data is another key area. The main uses for EDM implicate predicting student execution and study the learning process. to recommend improvement to existing educational practice. Baker, R. S., Martin, et.all, The extraction of educational data can be consider one of the learning science and an area of data extraction, and analyzes the learning process of the relevant areas EDM follows the same approach as in traditional method of DM from the need to understand the environments with which to learn and then collect the data and then clean and arrange the acquisition of techniques that can be applied and finally interpret the results and verify Dellinger Dissertation,et.all. the validity of the techniques second-hand. taking into account the distinct method and objective And techniques used resulting from the specificity of the educational environments and the purpose of exploration. EDM can be summarized as follows Baker, R. S. J. D, & Yacef, K, et. al



**Figure 1.** General framework in EDM

## 5. DATA COLLECTION METHODOLOGY

Data were obtained from various sources, such as sites for the dissemination of study results as well as the Directorate of Education Baghdad - Iraq, which is the result of a baccalaureate degree in Baghdad governorate. These data are the results of the third grade students and contain a number of data (108700) record of male and female students of (50) schools. From the year 2010 until 2019. These included morning schools, evening schools as well as private schools and finally foreign schools. Samples were collected from all schools. These schools are widespread in agricultural and low-income areas as well as in the city center. So that the results are accurate which is a degree of students in eight materials prescribed in Certificate third Average, namely: - Islamic Education, Arabic language, English, mathematics, social studies, biology, chemistry and finally physics.

### 6. DATA PROCESSING AND CLEANING

After the data collection to be analyzed and concerning the results of intermediate education (Baccalaureate), where the total assessed materials for the third average of eight materials study excluding material languages for small numbers where and weak impact on the results of students section. This the exclusion of some non-important fields in the study, such as (school name and the name of the student and the pattern of education and number exam) for lack of importance in the search, the focus was only on the students' grades and gender during the subject of the study years, where the conversion of these grades to the nominal values for materials and total average student to estimate (aa, ba,bb,cb,cc,ff,absent) as well as the Gender types coded two digits (1,2). Figure2

Islamic	<=100	aa	<=89	ba	<=79	bb	<=69	cb	<=59	cc	<=49	ff	=0	absent
Arabic	<=100	aa	<=89	ba	<=79	bb	<=69	cb	<=59	cc	<=49	ff	=0	absent
English	<=100	aa	<=89	ba	<=79	bb	<=69	cb	<=59	cc	<=49	ff	=0	absent
Mathematics	<=100	aa	<=89	ba	<=79	bb	<=69	cb	<=59	cc	<=49	ff	=0	absent
Social	<=100	aa	<=89	ba	<=79	bb	<=69	cb	<=59	cc	<=49	ff	=0	absent
Biology	<=100	aa	<=89	ba	<=79	bb	<=69	cb	<=59	cc	<=49	ff	=0	absent
Chemistry	<=100	aa	<=89	ba	<=79	bb	<=69	cb	<=59	cc	<=49	ff	=0	absent
Physics	<=100	aa	<=89	ba	<=79	bb	<=69	cb	<=59	cc	<=49	ff	=0	absent
average	<=100	aa	<=89	ba	<=79	bb	<=69	cb	<=59	cc	<=49	ff	=0	Absent

Figure 2. Convert numeric to nominal

After data processing and cleaning is complete, the final count of data (108021) has become a record for all students. Figure 3



the year	Gender	I.D. estimate	Islamic	A.D. estimate	Arabic	E.D. estimate	English	M.D. estimate	Mathematics	S.D. estimate	Social	B.D. estimate	Biology	C.D. estimate	Chemistry	P.D. estimate	Physics	Total scores	average	h. lessons	result
2010	female	ba	84	ba	80	cb	67	ba	88	cc	59	aa	90	cb	64	cb	66	598	74.75	0	pass.
2010	female	bb	75	cc	51	cc	50	cc	50	cc	51	bb	70	cc	58	cc	50	455	56.875	0	pass.
2010	female	ba	80	bb	71	cc	53	cb	67	ff	36	bb	75	cc	50	cb	63	495	61.875	1	failed
2010	female	ba	80	cb	64	cc	50	ff	40	cc	50	cb	63	cb	65	cc	60	472	59	1	failed
2010	female	bb	71	cb	63	cc	50	ff	38	cc	50	cb	64	ff	31	cc	50	417	52.125	2	failed
2010	female	cb	66	cc	50	cc	50	ff	31	ff	22	cb	60	ff	27	ff	20	326	40.75	4	failed
2010	female	ba	85	ba	85	aa	91	aa	94	cc	59	aa	99	aa	91	aa	95	699	87.375	0	pass.
2010	female	absent	0	absent	0	absent	0	absent	0	absent	0	absent	0	absent	0	absent	0	0	0	8	failed
2010	female	ba	88	ba	85	bb	72	cb	63	cc	59	aa	100	ba	87	bb	79	633	79.125	0	pass.
2010	female	bb	79	cc	58	ff	36	bb	73	ff	39	cc	56	cc	50	cc	50	441	55.125	2	failed
2010	female	ba	82	bb	72	cc	59	bb	72	ff	33	ba	80	ff	36	cb	60	494	61.75	2	failed
2010	female	bb	73	cb	62	cc	57	cb	64	ff	37	ba	80	ff	26	cc	53	452	56.5	2	failed
2010	female	cb	64	cc	59	ff	38	cc	57	cc	50	cb	68	cc	50	cc	50	436	54.5	1	failed
2010	female	aa	91	ba	84	ba	85	cc	52	cc	50	ba	89	aa	91	cb	62	604	75.5	0	pass.
2010	female	ba	88	ba	87	cb	68	cb	66	cc	59	aa	92	cb	69	cb	61	590	73.75	0	pass.
2010	female	cc	57	cb	60	ff	42	cc	51	ff	26	cb	64	ff	30	cc	51	381	47.625	3	failed
2010	female	ba	89	cb	63	cb	62	ba	86	ff	27	cb	66	cc	50	cb	61	504	63	1	failed
2010	female	aa	91	cc	52	cc	53	ff	34	cb	68	ba	83	cc	54	cc	53	488	61	1	failed
2010	female	bb	76	cc	50	ff	27	ff	37	ff	38	ff	41	ff	33	ff	38	340	42.5	6	failed
2010	female	bb	71	cb	66	cc	53	cb	62	ff	35	cb	66	cb	62	bb	72	487	60.875	1	failed
2010	female	bb	79	cc	52	cc	50	ba	84	ff	23	cc	57	cb	69	ff	35	449	56.125	2	failed
2010	female	ba	85	cc	50	cc	56	cc	54	ff	36	cb	62	cc	59	cc	50	452	56.5	1	failed
2010	female	cb	68	bb	73	ff	41	cb	64	cb	68	cb	68	cc	50	cc	58	490	61.25	1	failed
2010	female	aa	94	cb	69	cb	69	ba	81	ba	83	bb	79	cb	69	cc	50	594	74.25	0	pass.
2010	female	ba	82	cb	60	cb	66	cc	50	cb	69	cb	68	cc	50	ff	42	487	60.875	1	failed
2010	female	ba	82	cc	50	cc	50	ff	21	bb	73	cc	50	cc	50	ff	25	401	50.125	2	failed
2010	female	bb	79	cc	57	cb	61	ff	33	cb	66	cc	50	cc	51	ff	38	435	54.375	2	failed
2010	female	ba	88	cc	50	cb	66	ff	40	bb	70	cb	65	ff	42	cc	50	471	58.875	2	failed
2010	female	aa	90	bb	72	cc	56	cb	67	bb	75	bb	72	ff	38	ff	42	512	64	2	failed
2010	female	ba	80	cb	61	cb	65	cc	50	bb	72	cb	60	ff	34	cc	51	479	59.125	1	failed
2010	female	aa	93	cc	54	ff	31	cc	59	cc	55	bb	70	ff	39	ff	33	434	54.25	3	failed
2010	female	aa	93	cb	67	cb	62	ba	88	cb	65	cb	67	bb	77	cc	59	578	72.25	0	pass.
2010	female	ba	85	cb	62	cc	58	cc	50	bb	71	cb	68	cc	50	ff	39	483	60.375	1	failed
2010	female	aa	95	cb	66	cc	52	cc	50	cb	68	cb	67	ff	42	ff	42	482	60.25	2	failed
2010	female	bb	78	cb	62	cc	50	cc	50	cb	68	bb	77	cb	62	cc	50	497	62.125	0	pass.

Figure 3. After the cleaning process and data integration

## 7. TOOLS AND PROGRAMS USED IN THE DM

There are many products, tools or programs that the process of DM, which produces about 3 large and well-known companies (such as Microsoft or IBM or Spas) and other well-known companies and these programs, which is in the process of exploration that have been applied and the use of some techniques and algorithms their own in this research program (WEKA).

## 8. DATA CONVERSION (EXCEL) TO (.CSV).

Often data is displayed in the image (Excel) or a database, the original method of storage in WEKA is (ARFF). Where, this formula consists of a list of cases and the values of the properties of each of the cases separated by a comma. As well as that WEKA program allows export files of type (.CSV) and steps to convert the file as follows:

- from the button to open the download data on Excel.
- Press the Save As.
- Chose the file name from the dropdown menu and then change the file extension of the format (Excel) to (.CSV).

- Click Save. Figure4

result,average,Physics,P.D. estimate,Chemistry,C.D. estimate,Biology,B.D. estimate,Social ,S.D. estimate,Mathematics,M.D. estimate,English,E.D. estimate,Arabic,A.D. estimate,Islamic,I.D. estimate,Gender,the
pass,71.875,78,bb,81,ba,97,aa,67,cb,50,cc,50,cc,64,cb,88,ba,female,2010
pass,61.625,50,cc,50,cc,67,cb,66,cb,58,cc,50,cc,68,cb,84,ba,female,2010
pass,56.875,50,cc,60,cb,50,cc,52,cc,58,cc,59,cc,64,cb,62,cb,male,2019
failed,54.125,57,cc,50,cc,62,cb,50,cc,41,ff,50,cc,57,cc,66,cb,male,2019
pass,73.875,83,ba,61,cb,66,cb,74,bb,71,bb,73,bb,80,ba,83,ba,male,2019
failed,42,41,ff,50,cc,37,ff,30,ff,37,ff,31,ff,50,cc,60,cb,male,2018
failed,61.375,90,aa,54,cc,50,cc,64,cb,53,cc,40,ff,50,cc,90,aa,male,2018
failed,57.875,50,cc,52,cc,67,cb,50,cc,25,ff,79,bb,61,cb,79,bb,female,2015
pass,68.625,70,bb,69,cb,75,bb,61,cb,51,cc,66,cb,72,bb,85,ba,female,2015
failed,42.25,34,ff,31,ff,50,cc,51,cc,50,cc,26,ff,37,ff,59,cc,male,2014
failed,9.375,20,ff,0,absent,22,ff,0,absent,0,absent,0,absent,0,absent,33,ff,male,2014
failed,36.5,40,ff,39,ff,28,ff,40,ff,17,ff,20,ff,38,ff,70,bb,male,2014
failed,36,32,ff,30,ff,34,ff,42,ff,25,ff,25,ff,50,cc,50,cc,male,2014
failed,cb,61.75,60,cb,36,ff,80,ba,33,ff,72,bb,59,cc,72,bb,82,ba,female,2010
failed,cc,55.125,50,cc,50,cc,56,cc,39,ff,73,bb,36,ff,58,cc,79,bb,female,2010
pass,bb,73.75,61,cb,69,cb,92,aa,59,cc,66,cb,68,cb,87,ba,88,ba,female,2010

Figure 4. Conversion (Excel) to (.CSV)

### 9. CHOOSE THE APPROPRIATE ALGORITHMS FOR DM.

At this stage the use of Clustering technology and the application of the algorithm (k-means technique) using the classification decision tree algorithm (j48) using Wicca program.

### 10. IMPLEMENTATION OF THE ALGORITHM K-MEANS.

The k-means algorithm used to assemble multiple data (examples) depending on their characteristics to be in order and the assembly process by reducing the distances between the data and (clustering center) and is also one of the clustering algorithms that work on data mining After, entering data into your browser and make the filters operations on the data to fill in the missing values. We go to the option of cluster we go to the k-means algorithm selected. Then we go to the properties of the algorithm and we press to determine the number of clusters, where you select the five clusters, and then click on OK and then press start show the result as shown in Figure 5



```
Final cluster centroids:
Attribute      Full Data      Cluster#
                (108021.0)    0              1              2              3              4
-----
result         failed         failed         pass          failed         pass          failed
average        54.4155       17.0599       67.6714      52.7485       71.3722      49.5335
Physics        49.7124       11.9669       65.3004      45.5609       66.8055      45.6926
Chemistry      50.1409       11.7234       65.2995      46.7789       68.6911      45.0461
Biology        57.9692       14.8779       71.6373      57.327        76.42        52.7039
Social         56.0543       20.7154       67.9192      53.6196       70.336       53.3135
Mathematics    47.7834       10.624        65.6845      42.9732      67.3307      41.781
English        46.9466       14.1462       61.4946      45.1997      65.8201      39.4386
Arabic         57.2012       22.3972       66.2814      58.2096      72.4676      52.4155
Islamic        69.5238       30.0656       77.7625      72.3218      83.1172      65.8797
Gender         1.4507        1.1103        1              2              2              1
the year       2014.472     2013.5392     2013.9667    2014.9911    2014.3774    2014.5771

Time taken to build model (full training data) : 23.17 seconds

=== Model and evaluation on training set ===

Clustered Instances
0          8824 ( 8%)
1         14188 ( 13%)
2         26065 ( 24%)
3         21648 ( 20%)
4         37296 ( 35%)
```

Figure 5. Implementation of the algorithm k-means.

- The first cluster contains (8824) elements with a similarity percentage that reached 8%. We find that the highest achievement among the study materials is in the Islamic education with a percentage of 30%. In addition, the Arabic language and sociology are among the highest achievements. On the other hand, the mathematics material generally, has got the least achievement percentage that reached 10%. As well as chemistry and physics, we find similar in the collection, where the rate was 11%. And finally, English language and biology, we find similar in the collection, where the rate was 14%. As a result, we find that the Islamic education, Arabic language and sociology help in increasing the total achievement in the cumulative average. On the other hand, mathematics and chemistry help in decreasing the total achievement in the cumulative average.
- The second cluster contains (14188) elements with a similarity percentage that reached 13%. We find that the highest achievement among the study materials is in the Islamic education with a percentage of 77%. In addition, the biology and sociology are among the highest achievements. On the other hand, the English language material generally, has got the least achievement percentage that reached 61%. As well as chemistry, physics and mathematics, we find similar in the collection, where the rate was 65%. And finally, Arabic language reached 66%. As a result, we find that the Islamic education, biology, and sociology help in increasing the total achievement in the cumulative average. On the other hand, chemistry and English language help in decreasing the total achievement in the cumulative average.
- The third cluster contains (26065) elements with a similarity percentage that reached 24%. We find that the highest achievement among the study materials is in the Islamic education with a percentage of

72%. In addition, the Arabic language and biology are among the highest achievements. On the other hand, the mathematics material generally, has got the least achievement percentage that reached 42%. As well as English language and physics, we find similar in the collection, where the rate was 45%. As chemistry had 46%, and finally, sociology reached 53%. As a result, we find that the Islamic education, biology, and Arabic language help in increasing the total achievement in the cumulative average. On the other hand, Mathematics and English language help in decreasing the total achievement in the cumulative average.

- The fourth cluster contains (21648) elements with a similarity percentage that reached 20%. We find that the highest achievement among the study materials is in the Islamic education with a percentage of 83%. In addition, the Arabic language and biology are among the highest achievements. On the other hand, the English language material generally, has got the least achievement percentage that reached 65%. Moreover, physics is nearly the same with the percentage of 66%. Also, and Mathematics had 67%. As chemistry had 68%, and finally, sociology reached 70%. As a result, we find that the Islamic education, Arabic language and biology help in increasing the total achievement in the cumulative average. On the other hand, English language and physics help in decreasing the total achievement in the cumulative average.
- Cluster fifth contains (37296) elements with a similarity percentage that reached 35%. We find that the highest achievement among the study materials is in the Islamic education with a percentage of 65%. In addition, the sociology and biology are among the highest achievements. On the other hand, the English language material generally, has got the least achievement percentage that reached 39%. As well as chemistry and physics, we find similar in the collection, where the rate was 45%. Moreover, Arabic language is nearly the same with the percentage of 52%, finally, where the mathematics reached 41%. As a result, we find that the Islamic education, biology, and sociology help in increasing the total achievement in the cumulative average. On the other hand, English language and mathematics help in decreasing the total achievement in the cumulative average.

#### • In other respects

1. Note the first cluster and the third cluster are similar in, the highest achievement rate in Islamic education and Arabic language, and the lowest achievement in mathematics.
2. The third cluster and the fifth cluster are similar in, the highest achievement rate in Islamic education. The lowest achievement in English and mathematics. Sociology and physics are similar in the collection.
3. The second cluster and the fourth cluster are similar, and the highest achievement rate in Islamic education, the lowest percentage in English.
4. Cluster second and fifth is different in achievement where the male is more than the female.
5. The third cluster is different in achievement where the female is more than the male.

**11. CAUSES THAT LEAD TO THE RATE IN ISLAMIC EDUCATION HIGH**

- A. Islamic education is the main religion in the country is easy for many students.
- B. The most successful and achievable subjects because the conservation side takes a large part of the exam scores, which helps the student to achieve higher.
- C. The formula became almost known to students, which gained them experience in solving the exam.
- D. The uses of creative thinking skills in the teaching of Islamic education and Arabic have an impact on their academic achievement.
- E. The language of the Qur'an is the Arabic language where students know it.

**12. LESS MATERIAL TAKEN FOR THE SUBJECTS OF ENGLISH AND MATHEMATICS FOR THE FOLLOWING REASONS**

- A. the difficult political and security events that Iraq is going through is one of the most important reasons that lead to low academic achievement of the student where it causes fear, anxiety, tension and psychological instability.
- B. Pass students in lower grades without the ability to English and mathematics.
- C. The spread of violence and physical and verbal punishment within the school and the family and the environment in which the student lives.
- D. The use of the method of conservation used in teaching lead to low achievement of students.
- E. Failure to take into account the individual differences when raising materials within the curriculum leads to low achievement of students.
- F. Teaching English and mathematics from non-specialist teachers.
- G. Failure to register teachers of English and mathematics in the appropriate training courses lead to low achievement of students.
- H. Lack of availability and use of modern devices and means lead to low achievement of students.
- I. Busy classes lead to low student achievement.

**13. ALSO FIND THAT THE HIGHEST PERCENTAGE OF FEMALES MORE THAN MALES AT THE COLLECTION TO THE FOLLOWING**

- A. The presence of desire when the females more than males.
- B. When the females more success because the females are more males interested.
- C. Places students in solving duties more than female students. As shown in Figure 6

```

Final cluster centroids:
Attribute      Full Data      Cluster#
              (108021.0)    0          1          2          3          4
-----
result        failed        failed        pass        pass        failed        failed
Gender        male         male         male        female      female        Male

Time taken to build model (full training data) : 1.51 seconds

=== Model and evaluation on training set ===

Clustered Instances
0          44936 ( 42%)
1          14188 ( 13%)
2          21647 ( 20%)
3          27039 ( 25%)
4           211 ( 0%)
    
```

**Figure 6.** Highest percentage of females more than males at the collection

**14. THE USE OF A CLUSTER OF ISLAMIC EDUCATION AND COMPARE IT WITH THE AVERAGE FINAL RESULT**

- A. The first cluster number (30836) the average collection (49.86) and the average score (35.62) and finally estimate was (ff).
- B. The second cluster number (17325) the average collection (85.49) and the average score (66.90) and finally estimate was (cb).
- C. The third cluster number (15317) the average collection (72.25) and the average score (64.66) and finally estimate was (cb).
- D. The fourth cluster number (9422) the average collection (91.96) and the average score (81.48) and finally estimate was (ba).
- E. The Fifth cluster number (35121) the average collection (71.69) and the average score (53.01) and finally estimate was (cc). Figure 7

```

Final cluster centroids:
Attribute      Full Data      Cluster#
(108021.0)    (30836.0)    0 (17325.0)    1 (15317.0)    2 (9422.0)    3 (35121.0)    4
-----
result         failed         failed         pass          pass          pass          failed
AVG.D.estimate cc             ff             cb             cb             ba             cc
average        54.4155       35.6289       66.9059       64.6643       81.4845       53.017
Islamic        69.5238       49.86         85.4965       72.2534       91.9659       71.6983
I.D.estimate   bb             cc             ba             bb             aa             bb

Time taken to build model (full training data) : 8.46 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      30836 ( 29%)
1      17325 ( 16%)
2      15317 ( 14%)
3      9422  ( 9%)
4      35121 ( 33%)
    
```

Figure 7. Use of a cluster of Islamic education

### 15. THE USE OF A CLUSTER OF ARABIC LANGUAGE AND COMPARE IT WITH THE AVERAGE FINAL RESULT

- A. The first cluster number (23067) the average collection (66.10) and the average score (63.53) and finally estimate was (cb).
- B. The second cluster number (10959) the average collection (73.84) and the average score (73.35) and finally estimate was (bb).
- C. The third cluster number (33974) the average collection (56.49) and the average score (54.56) and finally estimate was (cc).
- D. The fourth cluster number (6443) the average collection (83.73) and the average score (85.56) and finally estimate was (ba).
- E. The Fifth cluster number (33578) the average collection (41.27) and the average score (35.83) and finally estimate was (ff). Figure 8

```

Final cluster centroids:
Attribute      Full Data      Cluster#
(108021.0)    (23067.0)    0 (10959.0)    1 (33974.0)    2 (6443.0)    3 (33578.0)    4
-----
result         failed         pass          pass          failed         pass          failed
AVG.D.estimate cc             cb             bb             cc             ba             ff
average        54.4155       63.5342       73.3568       54.5694       85.5604       35.8375
Arabic         57.2012       66.1042       73.8414       56.4949       83.7385       41.2768

Time taken to build model (full training data) : 5.33 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      23067 ( 21%)
1      10959 ( 10%)
2      33974 ( 31%)
3       6443 (  6%)
4      33578 ( 31%)
    
```

Figure 8. Use of a cluster of Arabic language.

## 16. THE USE OF A CLUSTER OF ENGLISH LANGUAGE AND COMPARE IT WITH THE AVERAGE FINAL RESULT

- A. The first cluster number (33580) the average collection (29.60) and the average score (35.83) and finally estimate was (ff).
- B. The second cluster number (10731) the average collection (66.62) and the average score (73.39) and finally estimate was (bb).
- C. The third cluster number (36262) the average collection (44.13) and the average score (55.24) and finally estimate was (cc).
- D. The fourth cluster number (6435) the average collection (82.34) and the average score (85.56) and finally estimate was (ba).
- E. The Fifth cluster number (21013) the average collection (58.62) and the average score (63.44) and finally estimate was (cb). Figure 9

```
Final cluster centroids:
Attribute      Full Data      Cluster#
                (108021.0)    0          1          2          3          4
                (33580.0)    (10731.0)  (36262.0)  (6435.0)  (21013.0)
=====
result         failed        failed      pass       failed     pass       pass
AVG.D.estimate cc            ff          bb         cc         ba         cb
average        54.4155      35.8394    73.3916   55.2443   85.5606   63.4422
English        46.9466      29.6034    66.6293   44.1323   82.345    58.6266

Time taken to build model (full training data) : 4.18 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      33580 ( 31%)
1      10731 ( 10%)
2      36262 ( 34%)
3       6435 ( 6%)
4      21013 ( 19%)
```

Figure 9. Use of a cluster of English language.

## 17. THE USE OF A CLUSTER OF MATHEMATICS AND COMPARE IT WITH THE AVERAGE FINAL RESULT

- A. The first cluster number (21366) the average collection (61.42) and the average score (63.46) and finally estimate was (cb).
- B. The second cluster number (10819) the average collection (69.77) and the average score (73.37) and finally estimate was (bb).



- C. The third cluster number (35791) the average collection (45.84) and the average score (55.10) and finally estimate was (cc).
- D. The fourth cluster number (6434) the average collection (84.84) and the average score (85.55) and finally estimate was (ba).
- E. The Fifth cluster number (33611) the average collection (27.00) and the average score (35.86) and finally estimate was (ff). Figure 10.

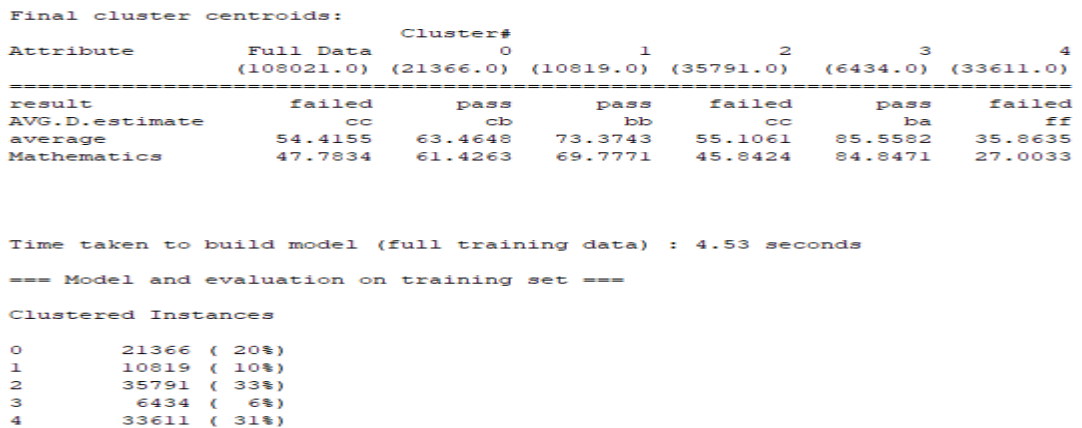


Figure 10. Use of a cluster of mathematics.

**18. THE USE OF A CLUSTER OF SOCIOLOGY AND COMPARE IT WITH THE AVERAGE FINAL RESULT**

- A. The first cluster number (34251) the average collection (55.31) and the average score (54.66) and finally estimate was (cc).
- B. The second cluster number (10937) the average collection (73.40) and the average score (73.36) and finally estimate was (bb).
- C. The third cluster number (22780) the average collection (66.28) and the average score (63.51) and finally estimate was (cb).
- D. The fourth cluster number (6437) the average collection (83.27) and the average score (85.55) and finally estimate was (ba).
- E. The Fifth cluster number (33616) the average collection (39.02) and the average score (35.87) and finally estimate was (ff). Figure 11

```

Final cluster centroids:
Attribute          Full Data          Cluster#
                  (108021.0)         0
                  (34251.0)         1
                  (10937.0)         2
                  (22780.0)         3
                  (6437.0)         4
                  (33616.0)
=====
result            failed            failed            pass            pass            pass            failed
AVG.D.estimate    cc              cc              bb              cb              ba              ff
average           54.4155        54.6635        73.3605        63.5115        85.5551        35.8716
Social            56.0543        55.3118        73.4071        66.282         83.2745        39.0218

Time taken to build model (full training data) : 4.21 seconds
=== Model and evaluation on training set ===

Clustered Instances
0          34251 ( 32%)
1          10937 ( 10%)
2          22780 ( 21%)
3          6437 ( 6%)
4          33616 ( 31%)
    
```

Figure 11. Use of a cluster of sociology.

**19. THE USE OF A CLUSTER OF BIOLOGY AND COMPARE IT WITH THE AVERAGE FINAL RESULT**

- A. The first cluster number (33944) the average collection (57.74) and the average score (54.57) and finally estimate was (cc).
- B. The second cluster number (10963) the average collection (79.73) and the average score (73.35) and finally estimate was (bb).
- C. The third cluster number (23038) the average collection (70.42) and the average score (63.51) and finally estimate was (cb).
- D. The fourth cluster number (6441) the average collection (89.87) and the average score (85.55) and finally estimate was (ba).
- E. The Fifth cluster number (33635) the average collection (36.46) and the average score (35.88) and finally estimate was (ff). Figure 12

```

Final cluster centroids:
Attribute          Full Data          Cluster#
                  (108021.0)         0
                  (33944.0)         1
                  (10963.0)         2
                  (23038.0)         3
                  (6441.0)         4
                  (33635.0)
=====
result            failed            failed            pass            pass            pass            failed
AVG.D.estimate    cc              cc              bb              cb              ba              ff
average           54.4155        54.5754        73.3565        63.5169        85.5571        35.8831
Biology            57.9692        57.7409        79.7309        70.4249        89.871         36.4659

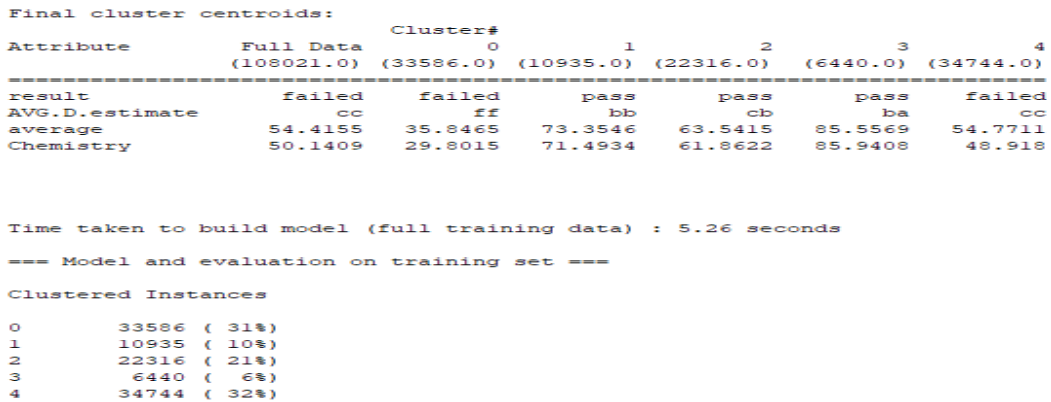
Time taken to build model (full training data) : 5.23 seconds
=== Model and evaluation on training set ===

Clustered Instances
0          33944 ( 31%)
1          10963 ( 10%)
2          23038 ( 21%)
3          6441 ( 6%)
4          33635 ( 31%)
    
```

Figure 12. Use of a cluster of biology.

**20. THE USE OF A CLUSTER OF CHEMISTRY AND COMPARE IT WITH THE AVERAGE FINAL RESULT**

- A. The first cluster number (33586) the average collection (29.80) and the average score (35.84) and finally estimate was (ff).
- B. The second cluster number (10935) the average collection (71.49) and the average score (73.35) and finally estimate was (bb).
- C. The third cluster number (22316) the average collection (61.86) and the average score (63.54) and finally estimate was (cb).
- D. The fourth cluster number (6440) the average collection (85.94) and the average score (85.55) and finally estimate was (ba).
- E. The Fifth cluster number (34744) the average collection (48.91) and the average score (54.77) and finally estimate was (cc). Figure 13



**Figure 13.** Use of a cluster of chemistry.

**21. THE USE OF A CLUSTER OF PHYSICS AND COMPARE IT WITH THE AVERAGE FINAL RESULT**

- A. The first cluster number (9886) the average collection (58.22) and the average score (62.58) and finally estimate was (cb).
- B. The second cluster number (21738) the average collection (60.57) and the average score (66.11) and finally estimate was (cb).
- C. The third cluster number (31771) the average collection (49.28) and the average score (54.14) and finally estimate was (cc).
- D. The fourth cluster number (11072) the average collection (81.31) and the average score (81.05) and finally estimate was (ba).
- E. The Fifth cluster number (33554) the average collection (30.14) and the average score (35.82) and finally estimate was (ff). Figure 14.

```

Final cluster centroids:
Attribute          Full Data          Cluster#
(108021.0)        (9886.0)          (21738.0)         (31771.0)         (11072.0)         (33554.0)
-----
result            failed            failed            pass              failed            pass              failed
AVG.D.estimate    cc               cb               cb               cc               ba               ff
average           54.4155         62.8504         66.1146         54.1408         81.0534         35.8213
Physics           49.7124         58.2213         60.5722         49.2874         81.311          30.1455

Time taken to build model (full training data) : 5.37 seconds

=== Model and evaluation on training set ===

Clustered Instances
0          9886 ( 9%)
1         21738 ( 20%)
2         31771 ( 29%)
3         11072 ( 10%)
4         33554 ( 31%)
    
```

Figure 14. Use of a cluster of physics.

## 22. FROM FINDINGS THE IMPLEMENTATION OF THE ALGORITHM K-MEANS

- A. There is a significant similarity between some materials (Islamic Education, Arabic language and biology) and materials (English and math), where working as a factor for success and failure in the final average.
- B. through the above results proved that more materials which are very weak (English and math).
- C. The results showed in the case of student success and failure rate in Islamic education, it is considered in the final deposit rate.
- D. Proved that the success rate, is higher among females than males.

## 23. IMPLEMENTATION OF THE DECISION TREE

At this stage, the implementation of the decision tree on the data, where the results were as follows: Figure 15, Figure 16, table 1,2.

```

=== Run information ===
Scheme:          weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:        weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.Reorder-R1,3,4
Instances:       108021
Attributes:      10
result
F.D.estimate
C.D.estimate
B.D.estimate
S.D.estimate
M.D.estimate
E.D.estimate
A.D.estimate
I.D.estimate
av.D.estimate
Test mode:      evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree
-----
result = pass
| A.D.estimate = ba
| | C.D.estimate = cb
| | | M.D.estimate = ba
| | | | B.D.estimate = aa
| | | | | E.D.estimate = cb
| | | | | F.D.estimate = cb: bb (4.0)
| | | | | F.D.estimate = cc: bb (0.0)
    
```

Figure 15. Result decision tree

==== Evaluation on training set ====

Time taken to test model on training data: 1.45 seconds

==== Summary ====

Correctly Classified Instances	97963	90.6888 %
Incorrectly Classified Instances	10058	9.3112 %
Kappa statistic	0.8772	
Mean absolute error	0.0449	
Root mean squared error	0.1498	
Relative absolute error	20.7015 %	
Root relative squared error	45.4991 %	
Total Number of Instances	108021	

==== Detailed Accuracy By Class ====

**Table 1.** Detailed Accuracy By Class

TP Rate	Rate Fp	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.851	0.016	0.860	0.851	0.855	0.839	0.988	0.906	bb
0.859	0.026	0.934	0.859	0.895	0.854	0.959	0.928	cc
0.880	0.040	0.865	0.880	0.872	0.834	0.975	0.916	cb
1.000	0.032	0.928	1.000	0.962	0.947	0.984	0.928	ff
0.858	0.005	0.884	0.858	0.871	0.865	0.996	0.920	ba
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	absent
0.877	0.001	0.900	0.877	0.889	0.887	0.999	0.943	aa
0.907	0.028	0.907	0.907	0.906	0.879	0.976	0.924	Avg Weightd

==== Confusion Matrix ====

**Table 2** Confusion Matrix

a	b	c	D	E	f	g	<-- classified as
9348	1	1274	0	364	0	0	a = bb
10	27953	2096	2475	0	0	0	b = cc
967	1983	21556	0	0	0	0	c = cb
0	0	0	31716	0	0	0	d = ff
546	0	1	0	4199	0	149	e = ba
0	0	0	0	0	1832	0	f = absent
1	0	0	0	189	0	1359	g = aa

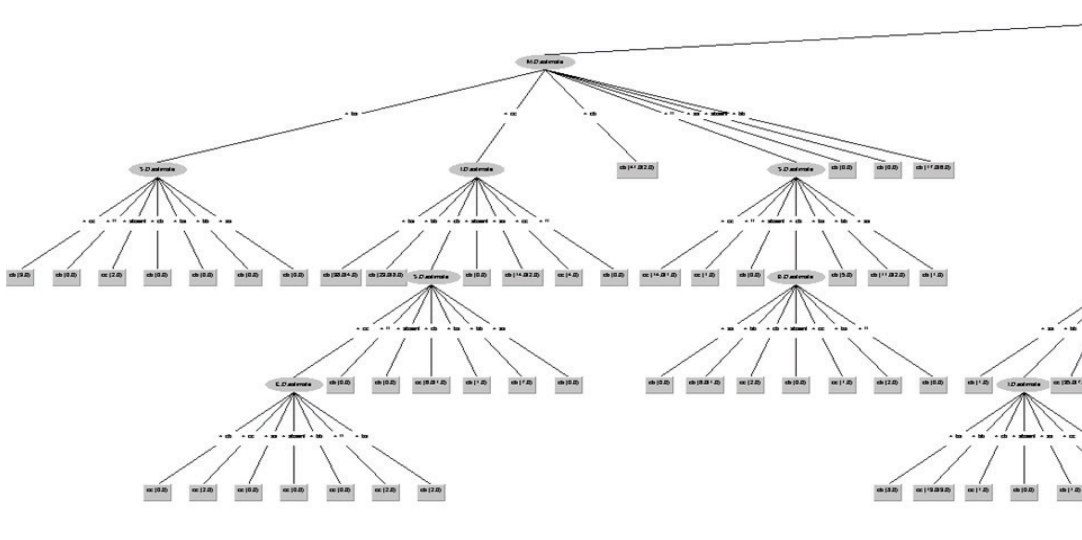


Figure 16. Visualize tree

## 24. DECISION TREE RESULTS

- A. The number of students failed (31716) students and the total absence of all subjects (1832) students out of (108021) students, then the number of successful students in the final rate (74743) students.
- B. Show the graduation of materials from the highest failures among the successful students (mathematics and English) down to the lowest failures, namely, Islamic education and biology.
- C. Where according to the grades of students in the final rate between the acceptable of less than 49 degrees and higher than 59 degrees, to excellent, which is higher than 89 degrees. The results were as shown in the tables 1,2.

## 25. GET OTHER RESULTS FROM WEKA.

- A. Islamic Education The number of successful (100789) where the students who rated AA (10651), who rated BA (22903), who rated BB (28009), who rated CB (22803), and who rated CC (16423).The number of repeaters (7232),who rated FF (4859), and who rated absentees (2373).
- B. Arabic language The number of successful (91410) where the students who rated AA (1459), who rated BA (6823), who rated BB (15596), who rated CB (25418), and who rated CC (42114).The number of repeaters (16611),who rated FF (13956), and who rated absentees (2655).
- C. English language The number of successful (63442) where the students who rated AA (2343), who rated BA (3898), who rated BB (6992), who rated CB (12164), and who rated CC (38045).The number of repeaters (44579),who rated FF (41220), and who rated absentees (3359).



- D. Mathematics The number of successful (65332) where the students who rated AA (3025), who rated BA (4922), who rated BB (8978), who rated CB (14142), and who rated CC (34265).The number of repeaters (42689),who rated FF (38750), and who rated absentees (3939).
- E. Sociology The number of successful (86365) where the students who rated AA (2606), who rated BA (7508), who rated BB (14773), who rated CB (22463), and who rated CC (39015).The number of repeaters (21656),who rated FF (18050), and who rated absentees (3606).
- F. Biology The number of successful (85590) where the students who rated AA (5553), who rated BA (10924), who rated BB (17028), who rated CB (20643), and who rated CC (31442).The number of repeaters (22431),who rated FF (18603), and who rated absentees (3828).
- G. Chemistry The number of successful (72223) where the students who rated AA (3058), who rated BA (5108), who rated BB (9391), who rated CB (15707), and who rated CC (38959).The number of repeaters (35798),who rated FF (31618), and who rated absentees (4180).
- H. Physics The number of successful (71923) where the students who rated AA (2136), who rated BA (4770), who rated BB (9296), who rated CB (15909), and who rated CC (39812).The number of repeaters (36098),who rated FF (31613), and who rated absentees (4485).
- I. Patterns resulting from the material in the case of failure in Islamic education and neighborhoods where the total number of students (6142), the number of successful students (21) students Duds (6121) student.
- J. Patterns resulting from the material in the case of success in mathematics, where the total number of students (65332), the number of successful students (59 412) students Duds (5920) student.
- K. Patterns resulting from the material in the case of success in the English language, where the total number of students (63442), the number of successful students (56 850) students Duds (6592) student.
- L. Patterns resulting from the material in the case of success in mathematics and English language, where the total number of students (48042), the number of successful students (46 630) students Duds (1412) student.

From these patterns, we conclude that the Islamic education and biology has a significant impact on the final rate increase and the success of students in the material. As well as the material of English and mathematics has an effect on the final cut and the rate of failure students.

## 26. CONCLUSIONS AND FUTURE WORK

The prediction of the performance of the student academic question of the process very important. Since the rise in the amount of learning data is very large that should be exploited and understood to develop the educational process. Where you in the thesis build a logical repository of data and the application of some mining algorithms in (50) schools, which included the results of the third grade students, which is a result of the baccalaureate degree in Baghdad, Iraq province . Where have reached some important conclusions, including the need to build an integrated, coherent and free of errors in the Directorate of

Education data warehouse. In addition, I noticed a relationship between some students and materials seep dropping out of the study because of the difficulty of vocabulary, the teacher teaching method or a defect in the school plans and curricula. And also it appeared to me and other important results related to student records such as final rates of relations and the period of interruption trends in academic students and the level of their performance by studying the baccalaureate. As well as after seeing the research appeared a correlation of specific materials low levels for students in which the results of a (English language and mathematics). There are several aspects have not been touched upon, such as the use of other ways cluster with the group of data for more information, and other aspects are not touched upon the lack of resources needed such as updating student data and update missing data for students in the database and the burdens of Iraqi education in order to be effective conclusion in the more patterns and relationships, such as health status, marital status, place of birth, the study of the teaching staff and the extent of data affected by the level of students and recent use of new techniques in data mining, such as genetic algorithm technology. All of these can be developed for future as works.

## REFERENCES

**Aldikanji, E., and Ajami, K.** 2016. Studying Academic Indicators within Virtual Learning Environment Using Educational Data Mining. *International Journal of Data Mining & Knowledge Management Process* 6(6), 29–42.

**Alom, B. M. M., and Courtney, M.** 2018. Educational Data Mining: A Case Study Perspectives from Primary to University Education in Australia. *International Journal of Information Technology and Computer Science* 10(2), 1–9.

**Asif, R., Merceron, A., Ali, S. A., Haider, N. G.** 2017. Analyzing undergraduate students' performance using educational data mining. *Computers and Education* (113), 177–194.

**Aunsan, S., and Thammaboosadee, S.** 2016. *Constructing A Risk Behavior Guideline for Adolescent Students Using Decision Tree* 01, 9–13.

**Baker, R. S., Martin, T., and Rossi, L. M.** 2016. Educational Data Mining and Learning Analytics. *The Handbook of Cognition and Assessment*, 379–396.

**Baker, R. S. J. D., and Yacef, K.** 2009. The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining* (1)1, 3–16.

**Dellinger Dissertation, J. T., Zhang, Y. L., Davis, B. W., Siemens, G.** 2019. *Pathway To Adopting Learning Analytics: Reconceptualizing the Decision-Making Process of K-12 Leaders in North Texas*.

**Fayyad, U.** 1996. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM* 39(11), 27–34

**Kabakchieva, D.** 2013. Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies* 13(1), 61–72.

**Márquez-Vera, C., Cano, A., Romero, C., Ventura, S.** 2013. Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied Intelligence* 38(3), 315–330.

**Osmanbegovic, E., and Suljic, M.** 2012. Data mining approach for predicting student performance. *Economic Review: Journal of Economics and Business* 10(1), 3-12.

**Prajapati, M. M., Patel, V. D., Patel, and D. M.** 2018. Experimental investigation of Wire EDM process parameters on surface roughness of AISI 304L during main cut and trim cuts. *International Research Journal of Engineering and Technology*, 2562–2565.

**Sahay S.K., and Sharma A.** 2019 A Survey on the Detection of Android Malicious Apps. *Advances in Computer Communication and Computational Sciences. Advances in Intelligent Systems and Computing. Vol. 924. Springer, Singapore*

**Tavares, R., Vieira, R., and Pedro, L.** 2017. A preliminary proposal of a conceptual educational data mining framework for science education: Scientific competences development and self-regulated learning. *2017 International Symposium on Computers in Education. SIIE 2017. 2018-January, 1–6.*

