Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/120330

Utilizing the Educational Data Mining Techniques "Orange Technology" for Detecting Patterns and Predicting Academic Performance of University Students

A. S. Abdelmagid[*] *and A. I. M. Qahmash*

Department of Educational Technology, College of Education, King Khalid University, Abha, Kingdom of Saudi Arabia

Received: 1 Oct. 2022, Revised: 28 Oct. 2022, Accepted: 20 Nov. 2022.
Published online:1 Mar. 2023.

Abstract: This study aimed at detecting the educational patterns and predicting the academic performance of university students through the “Orange” technology for data mining. To achieve this aim, a set of electronic courses taught to King Khalid University students through the Blackboard Learning Management System were selected. For knowledge detection, the "K-Means" clustering algorithm was used to detect patterns, while "Linear Regression, Random Forest, KNN, Tree, SVM" algorithms were used to predict students' academic performance. The results indicated that the "K-Means" aggregation algorithm collected students' scores in three main layers: the highest was in the first and second classes, while the lowest was in the third layer. As for predicting academic performance, the results indicated that students' academic performance can be predicted through activities and quarterly tests for all courses except for one Course in which academic performance can be predicted through the semester tests only, and the quarterly activities do not contribute to predicting the students' academic performance. The Linear Regression algorithm is the most contributing algorithm in predicting the academic performance of students, while the SVM algorithm is the least.

Keywords: Data mining, Orange technology, Educational patterns, Academic prediction, Artificial intelligence algorithms.

1. Introduction

Undoubtedly, the development of science, economics, and information and communication technology has helped greatly in increasing the amount of digital data in recent years, as traditional and statistical analysis techniques are no longer able to deal with the huge amounts of such data. Data Mining has emerged as one of the successful solutions for analyzing the content of digital data by converting it from mere accumulated and incomprehensible information (data) into information of cognitive value that decision-makers can exploit and benefit from [1].

Data Mining is the discovery of knowledge from data, or it is the process of analyzing data from different perspectives, extracting relationships between them and summarizing them into useful information such as: Information that can contribute to increasing profits, reducing costs, or both, or revealing information and finding useful information through the use of a set of complex statistical tools or artificial intelligence applications [2].

Data is divided into "Structured Data" and "Unstructured Data." Structured Data is stored in data fields that can be searched, analyzed, and managed. Unstructured Data cannot be easily categorized such as: Images, graphs, videos, web pages, wikis, tweets, Facebook posts, chats and more. Although these types of files have their own internal system, they are "unstructured" because their data is not completely consistent as a database. Moreover, there is also "Semi-structured Data." It is a mixture of the two aforesaid types, but it lacks a regular environment, such as: Word processing software [3] All of the above data categories are mined by organizations with the aim of using them in: Predictive analytics or machine learning or just for maintaining a large repository of data. Mined data, whatever its form or category, is known as “Big Data” [4].

Decisions are made in data science based on the information available from the data, aiming at not only producing a theory or law but also a decision or procedure that is consistent with the available data; for example: Amazon uses the big data that it collects through customers’ operations to make decisions related to the process of marketing its products, and to make suggestions based on data analysis. Therefore, Amazon is not only interested in reaching a theory about human behavior, but it is more preoccupied with researching the highest possible accuracy in predicting and directing consumer behavior [5]. Therefore, [6] study analyzed Google and Facebook. This study concluded that "Facebook" uses user data to develop and improve its services, by collecting data about users in terms of their uses and interaction with

[*]Corresponding author e-mail: abdelmagid@kku.edu.sa

the services provided by "Facebook" and the people and things they interact with, to personalize the content, that is, by providing services compatible with the user's preferences.

Data Mining is used to solve four issues: Discovering relationships, analyzing links, and making choices, where the data necessary is evaluated to make decisions, determine preference methods and propose the best one, and the fourth issue which is prediction such as: Predicting market needs and customer requests in the future, and finally determining the best procedures for carrying out work. There are four main issues for data mining, including: Prediction, Classification, Clustering Analysis, and Association Rules Discovery. The study of concluded that the prediction strategy is the most used strategy in academic research in data mining [7].

The experiences of Arab universities in data analysis include the experience of King Saud University by introducing "ITQAN" system for data analysis. This system aims to achieve the highest quality standards for the university and to reach all its applications to be standard institutional applications and to achieve added value for the university and help university officials to make the appropriate decision [8].

Analytics has been used in educational settings for nearly twenty years to develop predictive models for administrative functions such as: Registry Administration. When considering applications for admission, some educational institutions estimate the likelihood of applicants obtaining successful educational outcomes, such as: Continuance graduation based on models developed from estimates of the probability of their former students achieving successful results. These estimates were derived from pre-application demographic and academic preparation data related to results [9].

Data mining in educational systems is powerfully revealing useful information that can be used in formative assessment to assist teachers in providing a scientific basis for their decisions about designing or modifying learning environments. Moreover, the application of data mining in educational systems is a retrieval circular process consisting of: Formulating, testing, and revising hypotheses. Academic officials plan, design, develop, and maintain educational systems, and learners interact with these systems. The application of data mining methods starts from all available information about courses, students, usage, and interaction to discover useful information that helps teachers and administrators improve the e-learning process [10].

[11] indicates that e-learning management systems such as: Blackboard provides data, but does not explain the importance of such data; as the registers record a lot of data about the behavior of the learners in the learning management system. Nor does it explain why some of this data is important and some is not; therefore, more research and studies should be conducted to determine why some unknown data is important. The results of this study indicated that the time spent by students on Moodle system, the number of downloads, and the frequency of logging into the system did not have a significant effect on students' learning.

The stages of designing electronic courses begin with analyzing data related to these courses and the surrounding learning environment, such as: E-learning management systems through which courses are offered, educational social media, electronic forums, and others. All these environments result in a huge amount of data related to learning, whether it is structured data stored in databases or unstructured such as: Texts, graphics, sounds, etc., as a result of the learner’s interaction with its electronic courses [12].

In addition, the use of data mining techniques in the field of education facilitates exploring and focusing on the significant information in databases. Data mining techniques also focus on building future predictions and exploring behavior and trends which allows making decisions correctly and on time [13]. Predictive Analytics is also the significant goal of big data, which is a sub-discipline of data mining concerned with prediction potentialities. This technology uses measurable variables to predict the future behavior of an individual or organization where multiple predictions are combined into a single predictive model [14].

The process of managing higher education institutions is one of the difficulties facing those in charge of them due to their large size and the multiplicity of their data sources. Any educational institution includes many users who use Educational Data Mining (EDM) according to their needs and goals in the educational institution. Therefore, the primary goal of Educational Data Mining (EDM) is to use the huge amount of educational data to achieve a better understanding of its development, and to secure broader information therein [15] .In this aspect, the results of [16] study concluded that the use of data mining in higher education institutions is still in its infancy, despite its importance and the development of its various methods and benefits, which can provide us with the results of the correlations between the different patterns that show the strengths and weaknesses of the developmental decision-makers to provide what is useful in evaluating the performance within the educational institution for the quality of educational outcomes and outputs.

To achieve the goals of sustainable development and competitiveness in educational institutions in Arab universities, these institutions must embrace the big data revolution, and the ability to analyze and exploit it. Big data analysis of e-learning management systems in universities such as: Blackboard and other e-learning management systems help bring

value and meaning to empirical research in the humanities; as most studies lack accumulation, that is, the results of one another are built upon in order to help students who study through these systems. Therefore, researchers often find themselves in front of a large number of studies on e-learning that do not collect a general result (s), which reduces the educational benefit therein [17].

2. Research Problem:

Universities graduate Lots of students annually in different disciplines, resulting in the availability of large amounts of data that have not been developed and coordinated in a database. Many faculties exploit such data in order to be optimally used and invested, and to benefit from it in identifying the student's behavior during his academic period, and the factors and reasons that lead to his academic retardation. Consequently, the student's characteristics were not counted, and the reasons of his academic retardation or excellence in studying were analyzed, and the student's performance indicators for success in academic subjects were not studied to assist decision makers in developing policies capable of developing the educational process, and reducing the phenomenon of academic failure and retardation [18].

Arab universities possess big data, however, there is a continuous problem of processing this data remains in terms of size, coordination, and variety. Big Data focuses on the technical dimension, while Smart Data is more concerned with the analytical dimension, the value of data and its integration in decision-making processes, and academic prediction in various educational institutions [19]. Arab universities are interested in establishing online learning communities such as Blackboard system through which students can securely exchange resources, manage discussions, and exchange ideas with students, as well as students communication with each other and their teachers. Despite the interest of Arab universities in providing these systems, these universities are not interested in analyzing the big data of the students' electronic courses through these systems, and as a result, many students attribute their success or failure to external rather than internal factors [20].

Therefore, institutions of higher education including universities, colleges and higher institutes, are blamed by all segments of society for the failure of these institutions to qualify their graduates and provide them with the necessary skills to succeed in life and at work. This social reproach continues, and it has never stopped, rather, it is getting harsher nowadays for two reasons: The rapid pace of change, and its wide scope. Perhaps the traditional higher education's continuous slowness in responding to technical, social and economic shifts adds even greater pressure thereon. Modern technologies are developing rapidly, which makes it impossible to monitor them, as they affect various areas of life, work and production methods. Predictions that were described as imaginary have become a reality, not only in developed countries, but also in developing ones [21].

There are many studies concerned with studying the use of mining in the field of commerce, and studying its impact on the process of developing these systems, but its use in the field of education is still in its infancy. Educational research usually uses traditional methods such as: Interviews, learners' activities in the academic stage, and SPSS, Excel programs to collect data related to their educational experience. These methods are of limited use because they are time and effort consuming, and they always stop at a certain limit of short term and benefit, especially with regard to making the appropriate decision in the educational process [22]. The real problem with big data is the ability to analyze and understand this data in order to use the results of this analysis to improve the quality of human life, and to develop knowledge. Big data has come to evolve, and the only way to deal with it is to be aware of how to analyze and make good use of it. The interest in analyzing big data in universities is still weak in practice. There is a lot of theoretical writings about big data and its analysis, but the ability of educational institutions to analyze this data is still very weak [23].

The University of Khalid has a huge database on students, programs, courses, faculty members, staff, outputs, educational results, and others since its establishment until now. This huge amount of data is not exploited and has not been used effectively in knowing the indicators of success and failure. Also, the prevailing patterns affecting the performance of students and academic departments, which is one of the most important elements of evaluating the educational process, have not been known [24] - [25].

Analyzing big data for students' electronic courses taught to them via e-learning management systems (Blackboard) at King Khalid University provides a difficulty for the educational institution, especially unstructured data such as: Pictures, tweets, students' comments on Facebook, blogging, etc. Therefore, attention must be paid to analyzing the data of these courses in order to understand students' behavior, predict their achievement and skills needed by the labor market, and reach knowledge based on scientific foundations that help decision-makers build a high-quality e-learning environment.

Based on the above, the problem of this research lies in the lack of interest in analyzing the results of university students in their complete online courses to discover the educational patterns prevailing among university students in courses

(111 SLM, 112 SLM, 113 SLM, and 202 ARB), as well as the weak interest in predicting student achievement and their future skills. Accordingly, the current research attempts to address this weakness through the use of educational data mining techniques using the "Orange" technology, via algorithms for educational patterns and predicting the academic performance of university students.

3. Research Questions:

This research answers the following questions:

1- What are the prevailing educational patterns in the courses of (111 SLM, 112 SLM, 113 SLM, 202 ARB) that affect the achievement of King Khalid University students, which can be discovered using data mining algorithms through "Orange" technology?

2- What are the best algorithms that can be used to predict the academic achievement of King Khalid University students in the courses (111 SLM, 112 SLM, 113 SLM, and 202 ARB) through "Orange" technology?

4. Research Aims:

This research aims at:

1- Analyzing the educational data of King Khalid University students for their online courses through Blackboard system at the university.

2- Discovering the prevailing patterns in the data of students who study online courses through Blackboard system at the university.

3- Assisting decision-makers in making appropriate developmental decisions related to predicting the academic performance of students who study online courses through Blackboard system.

4- Presenting some Linear Regression, Random Forest, KNN, Tree, SVM, K-Means algorithms used in data mining via Orange technology to help teachers and decision-makers develop fully online courses on Blackboard system.

5- Presenting some Linear Regression, Random Forest, KNN, Tree, SVM, K-Means algorithms used in data mining via Orange technology to predict students' results and help in making the appropriate decision.

5. Research Importance:

1- Encouraging university education officials to pay attention to data mining in the university education system.

2- Encouraging those in charge of e-teaching to understand students' learning styles and provide them with appropriate educational content to improve educational content and develop the Blackboard system.

3- Encouraging university education officials to pay attention to employing data mining techniques in order to predict the future academic performance of students.

4- Presenting several algorithms that can help university education officials evaluate and assess the educational system, and take appropriate developmental decisions.

6. Research Methodology:

This research employed data mining and machine learning methods to discover knowledge from the results of male and female students at King Khalid University as well as analyzing student data, identifying prevailing patterns and predicting potential future changes, and develop electronic courses on the "Blackboard" system, through the use of several algorithms via Orange technology and open source software.

7. Research Limitations:

This research is limited to the following:

1- King Khalid University students.

2- 111 SLM, 112 SLM, 113 SLM, 202 ARB Courses. These courses are full online courses through Blackboard system.

3- The first semester of 2021/2022 AD.

4- Algorithms: Linear Regression, Random Forest, KNN, Tree, SVM, K-Means.

8. Research Terms:

1- Data Mining:

It is a process through which the raw data in courses (111 SLM, 112 SLM, 113 SLM, 202 ARB) are converted into useful information and knowledge through the use of "Orange" technology to identify the learning patterns of King Khalid University students, predict their academic performance, and develop their e-learning environments by using Linear Regression, Random Forest, KNN, Tree, SVM, K-Means algorithms.

2- Orange Program:

It is an open source smart application for educational data mining based on Python language, in which knowledge is analyzed and extracted from the data of King Khalid University students in courses (11 Salam, 112 Salam, 113 Salam, 202 ARB) in a visual way through a set of smart tools and various algorithms that help to understand data of King Khalid University students and predict their academic performance scientifically.

3- Learning Styles:

It is the general form of the scores of King Khalid University students in courses (11 SLM, 112 SLM, 113 SLM, and 202 ARB) as a result of their performance in activities, semester and final exams. It can be identified through the use of the K-Means algorithm through "Orange Technology".

4- Academic Prediction:

It is the estimate or prediction of the final scores of King Khalid University students based on the analysis of their scores in the semester activities and tests in the courses (111 SLM, 112 SLM, 113 SLM, and 202 ARB) through Linear Regression, Random Forest, KNN, Tree, SVM algorithms via "Orange" technology.

Research Methodology and Procedures

This research employed data mining methods to discover knowledge through the results of students' online general preparation courses studied via Blackboard system in order to identify learning patterns that affect their academic achievement, as well as predicting the academic performance of students in the future through the use of Linear Regression, Random Forest, KNN, Tree, SVM, K-Means algorisms available via Orange Technology.

9. Research Procedures:

In order to identify the effectiveness of using educational data mining techniques "Orange Technology" to discover patterns that affect students' academic achievement at King Khalid University, and to predict their academic performance, the following was conducted:

First: Research Population:

King Khalid University students who study online general preparation courses fully during the first semester of the academic year (2022/21 AD) via Blackboard system. Table (1) shows these materials and their number.

Table 1: General online preparation courses for university students

S	Course Name	Number of Courses	Numbers of Selected Courses	Numbers of Students
1	111 SLM	37	4	717
2	112 SLM	11	4	730
3	113 SLM	26	4	760
4	202 ARB	25	4	750
	Total	99	16	2957

Table (1) shows that (99) courses are provided entirely through Blackboard system at King Khalid University. (16) Courses were chosen randomly from these courses. The number of male and female students in these courses reached (2957). This data was obtained from the Learning Data Analytics Unit after the approval of the Deanship of E-Learning at the University.

Second: Initial Data Processing Phase:

The used database including its mined data represents the score of King Khalid University students for online general preparation courses taught via (Blackboard) system. Some fields that are not important in the search were excluded, namely student's name and gender, as well as the username, and students who did not submit activities, or are in progress to take semester or final exams. The focus in the data was on the scores of student activities, and their semester and final exams, as shown in Table (2).

Table 2: Scores of activities and semester and final exams for general preparation courses

S	Course Name	Activities scores	Midterm scores	Final exam scores
1	111 SLM	50	20	30

2	112 SLM	50	20	30
3	113 SLM	50	20	30
4	202 ARB	60	20	20

Third: Converting data into (Comma Separated Value .CSV) format:

Orange program allows dealing with CSV files. The format of the data available in Excel program has been converted to CSV files as in Fig (1).

Final	Midterms	Total1	AC5	AC4	AC3	AC2	AC1	Total
30	20	47	10	10	10	10	7	97
30	20	50	10	10	10	10	10	100
30	19	50	10	10	10	10	10	99
30	20	50	10	10	10	10	10	100
30	20	40	9	10	5	7	9	90
29	20	49	10	9	10	10	10	98
30	20	50	10	10	10	10	10	100
30	20	50	10	10	10	10	10	100
30	20	48	10	10	10	10	8	98
30	20	50	10	10	10	10	10	100
28	20	49	9	10	10	10	10	97
30	20	50	10	10	10	10	10	100
30	20	47	10	10	8	10	9	97
30	20	50	10	10	10	10	10	100
30	20	50	10	10	10	10	10	100
30	20	50	10	10	10	10	10	100
30	20	49	9	10	10	10	10	99
30	20	50	10	10	10	10	10	100
30	20	50	10	10	10	10	10	100
30	20	50	10	10	10	10	10	100
30	20	45	8	9	10	10	8	95
30	20	50	10	10	10	10	10	100

Fig. 1. Student data after modification and conversion to CSV file

Fifth: Description of student data (Meta Data):

The database of King Khalid University students includes several fields, as shown in Table (3).

Table 3: General preparation courses database fields

S	Field Name	Data Type	S	Field Name	Data Type
1	Family name	Nominal	7	AC3	Numeric
2	First name	Nominal	8	AC4	Numeric
3	User	Numeric	9	AC5	Numeric
4	Total	Numeric	10	Total1	Numeric
5	AC1	Numeric	**11**	Midterms	Numeric
6	AC2	Numeric	**12**	Final	Numeric

Table (2) shows that the data type in the (Family name) field and the (First name) field is (Nominal), while the rest fields have (Numeric) data.

Sixth: Orange Program and Data Mining:

Orange software or technology is open source software for complex data mining and clustering methods, advanced data analysis, and graph generation. This program is an easy-to-use graphical interface, provided with a Python language library, and a website (https://orangedatamining.com), as in Figure (2) [10]. This Research uses Orange program.

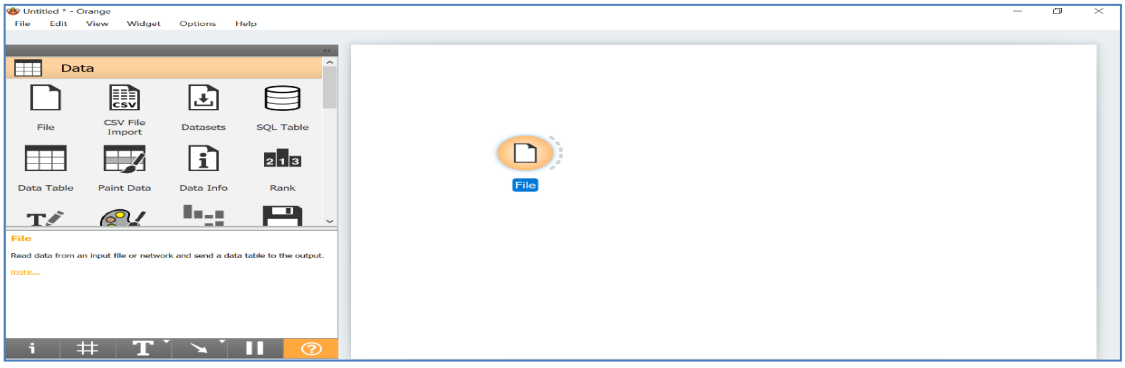


Fig. 2. Orange Data Mining Program

Seventh: Techniques and algorithms used in the research:

This study employed Classification, Clustering, and Predictions techniques, and Linear Regression, Random Forest, KNN, Tree, SVM, K-Means algorithms. These algorithms discover knowledge in the databases uploaded to the Orange program.

10. Research Findings and Discussion

After cleaning the data and uploading it to Orange, the research answer its questions as follows:

First Question Answer:

1- What are the prevailing educational patterns in the courses of (111 SLM, 112 SLM, 113 SLM, 202 ARB) that affect the achievement of King Khalid University students, which can be discovered using data mining algorithms through "Orange" technology?

To answer this question, the relevant literature, studies, and researches concerned the classification of educational data were reviewed, such as the study of [26]. The research employed K-Means algorithm through Orange program, as follows:

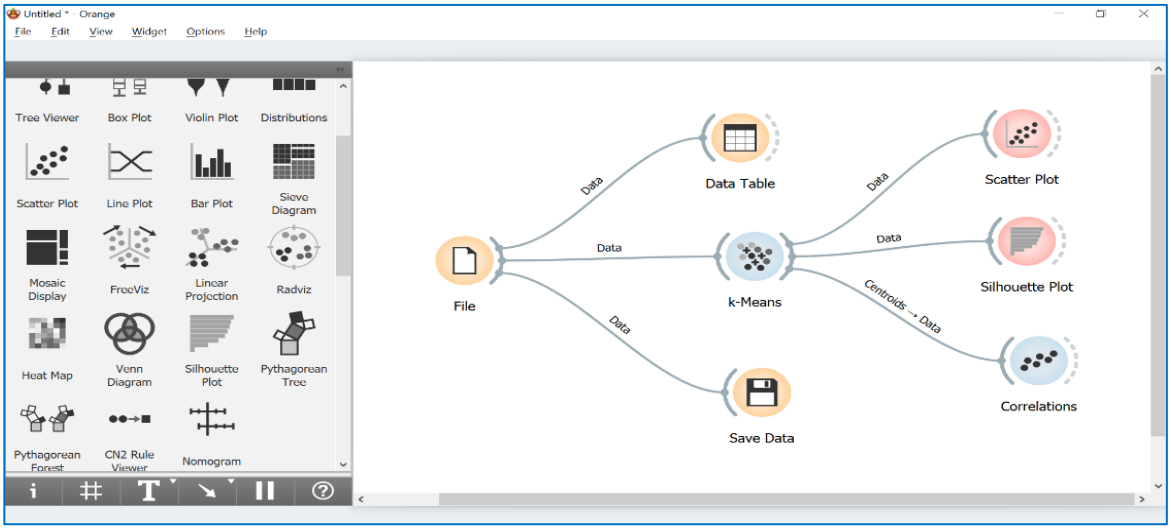


Fig. 3 In the previous figure, the (Orange) program and the (K-Means) algorithm were used in the courses (111SLAM, 112SLAM, 113SLAM, 202ARAB) in order to find correlations between big data as well as graphs in order to clarify the areas of data correlation.

After cleaning the data and uploading it to the Orange program, the research questions were answered as follows:

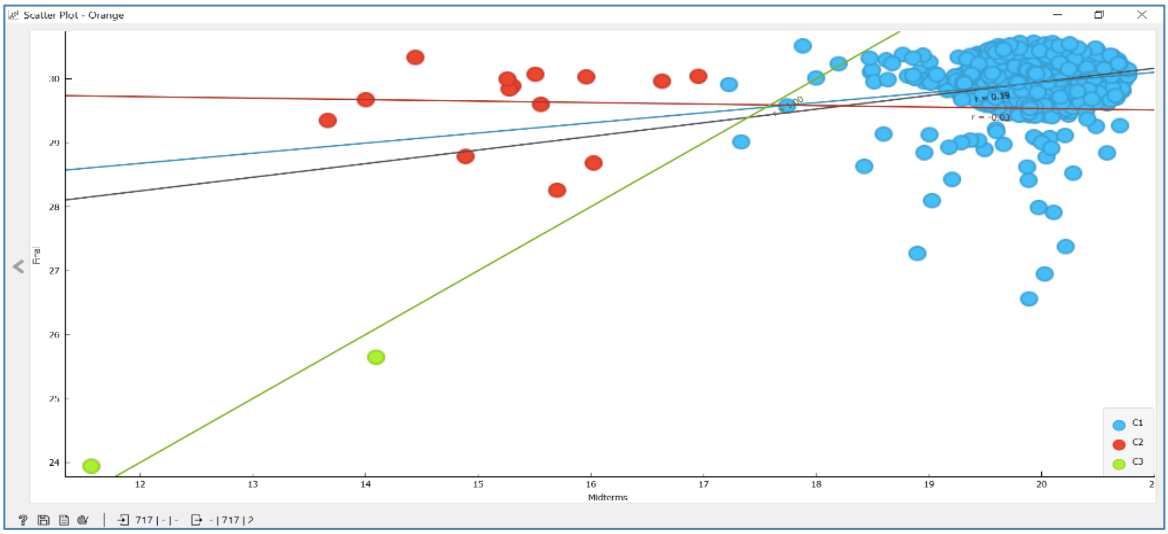


Fig. 4 The previous figure shows the division of students' scores in electronic courses into three layers using the (K-Means) algorithm for (111 Salam) course, the results were as the following figure:

The previous figure for (111 Salam) course shows that Orange program has divided the students' scores in this course into three main layers:

1- First layer: It is the majority layer. The students' midterm scores are (18-20), and (28-30) for the final exam.

2- Second layer: The students' midterm scores are (14-17), and (28-30) for the final exam

3- Third layer: It is the lowest layer. The students' midterm scores are (12-14), and (24-26) for the final exam. It represents a very small number of students. The following figure shows the three layers:

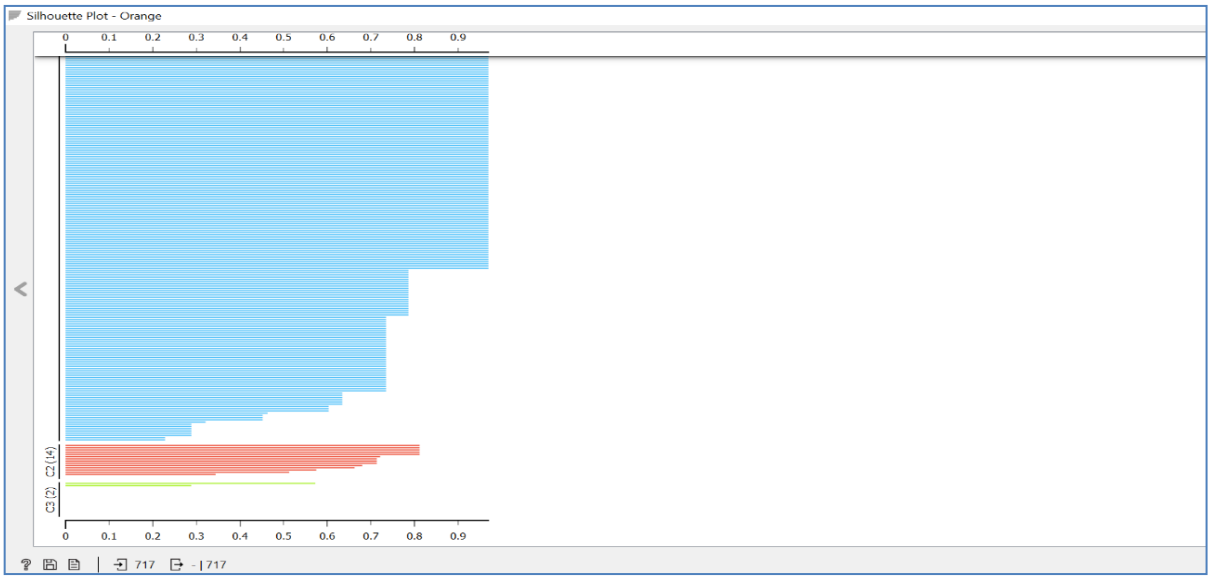


Fig. 5 This figure shows the distribution of students' grades in the semester and final exams in electronic courses. For (112 SLM) course, the results were as follows:

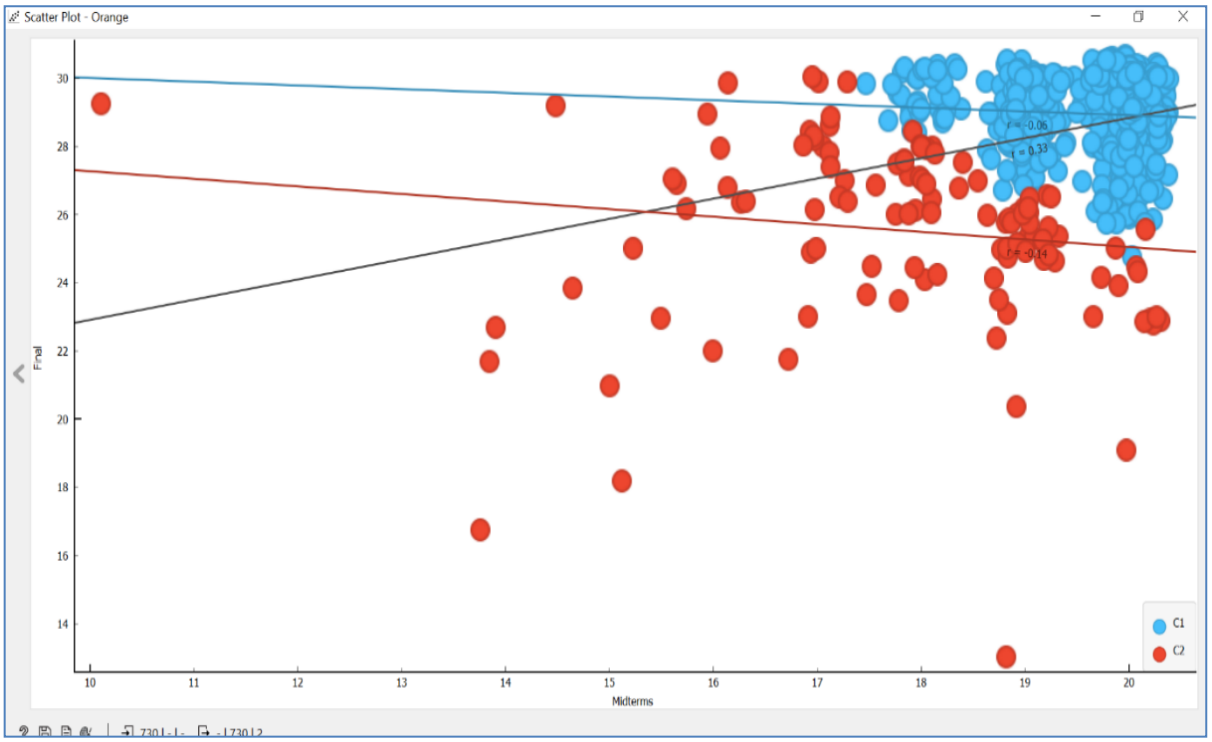


Fig. 6 The previous figure for (111 Salam) course shows that Orange program has divided the students' scores in this course into three main overlapping layers:

1- First layer: The students' midterm scores are (19-20), and (28-30) for the final exam

2- Second layer: The students' midterm scores are (14-17), and (24-28) for the final exam, which is an overlapping layer with the first one.

3- Third layer: Overlapping with the first and second layers, but to a small degree, as the midterm test scores were (14-18), and the final test scores were (17-30). The following figure shows these three layers:

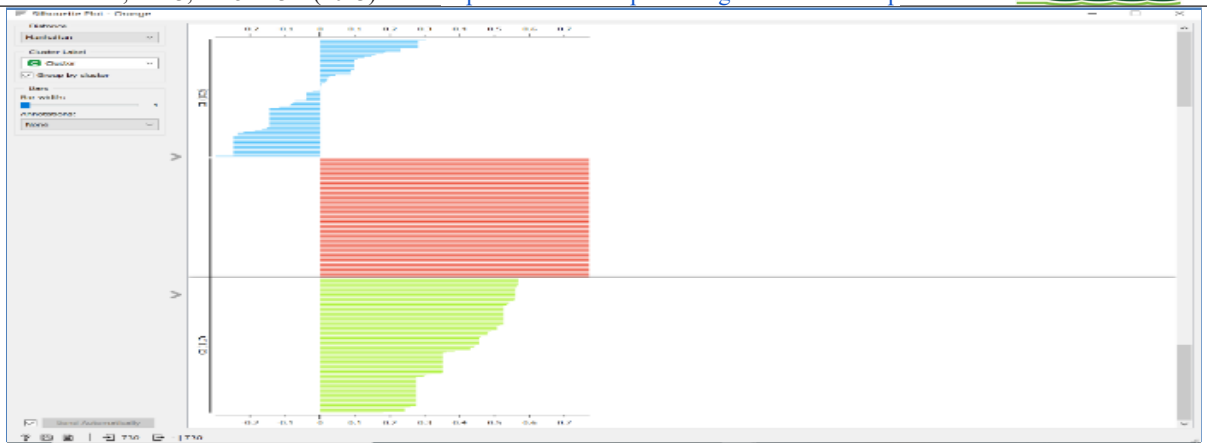


Fig. 7 This figure shows the overlap of students' scores (the middle region) in the semester and final exams in electronic courses.

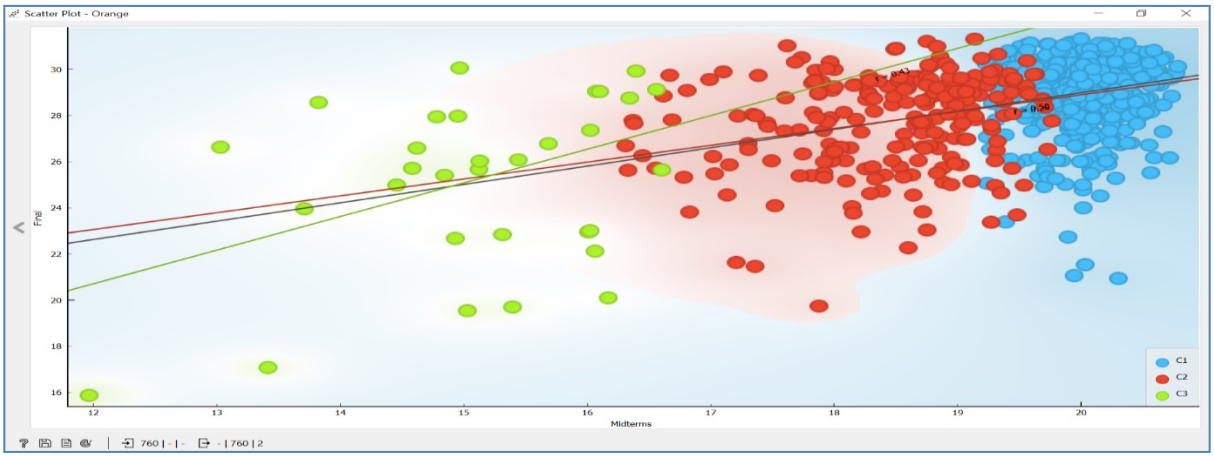


Fig. 8 In the previous figure, the (Orange) program and the (K-Means) algorithm were used to divide students' scores in electronic courses into three non-overlapping layers. for (113 Salam) course, the results were as follows:

The previous figure for the (113 SLM) course shows that (Orange) technology has divided students' scores in this course into three overlapping layers:

1- First layer: The students' midterm scores are (19-20), and (27-30) for the final exam.

2- Second layer: The students' midterm scores are (17-20), and (27-30) for the final exam, which is an overlapping layer with the first one.

3- Third layer: Overlapping with the first and second layers, but to a lesser extent, as students' midterm scores are (12-16), and (16-29) for the final exam the following figure shows these three layers:

Fig. 9 The previous figure shows the division of students' scores in electronic courses into three layers using the (K-Means) algorithm, the least of which is the third layer.

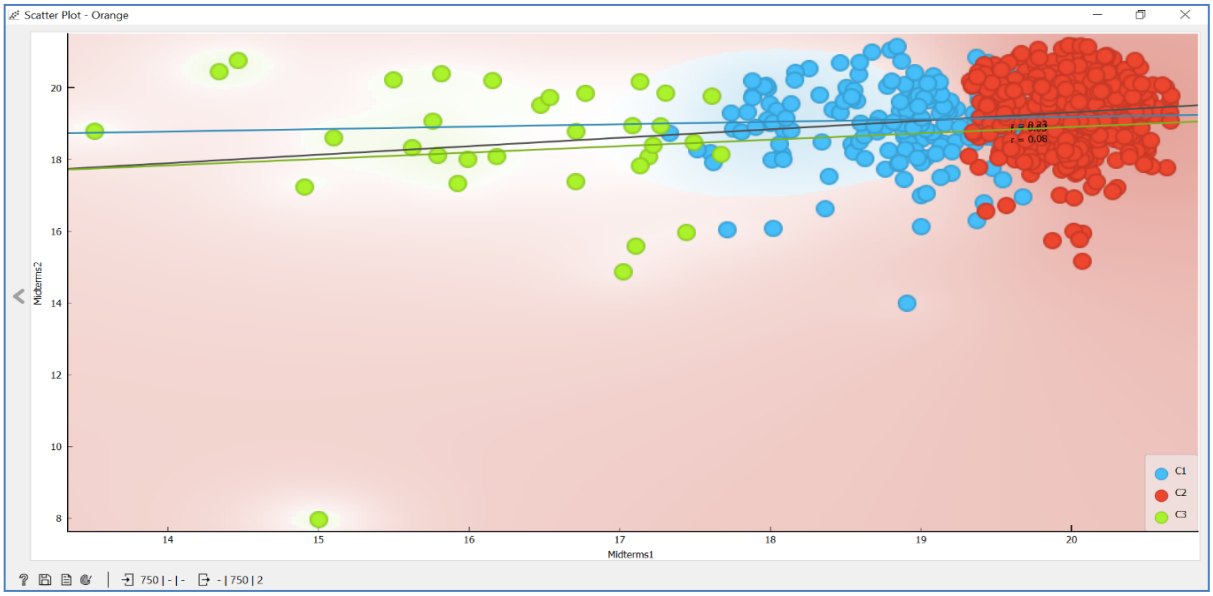


Fig. 10 the previous figure for the course (202 Arabs) shows that the (Orange) program has divided the students' grades in this subject into three overlapping layers: for (202 ARB) course, the results were as the following figure:

The previous figure for (202 ARB) course shows that Orange program has divided the students' scores in this course into three main overlapping layers:

1- First layer: The students' midterm scores are (19-20), and (17-20) for the final exam), whose final score was (20).

2- Second layer: This layer includes most of the students' scores. The students' midterm scores are (18-19), and (16-20) for the final exam, which is a layer overlapping with a small degree with the first layer.

3- Third layer: Overlapping with the second layer, but to a lesser degree. The students' midterm scores are (13-17), and (19-20) for the final exam. The following figure shows these three layers:

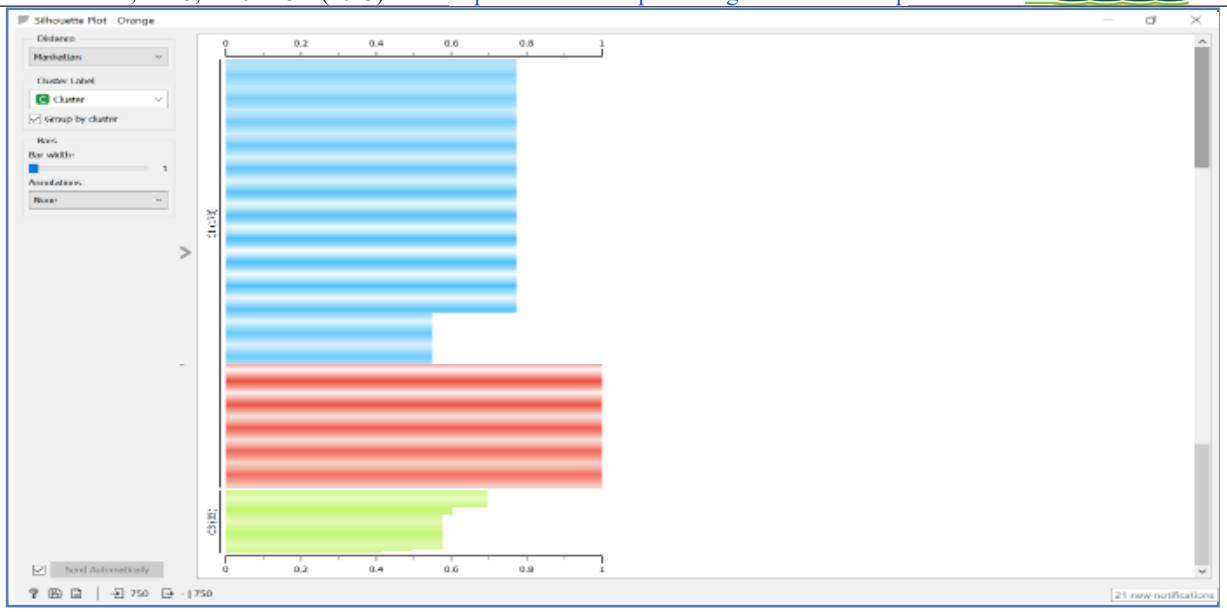


Fig. 11 the previous figure shows the division of huge scores for students in electronic courses into three layers using the (K-Means) algorithm, the least of which is the third layer.

The answer to the second question:

2- What are the best algorithms that can be used to predict the academic achievement of King Khalid University students in the courses (111 SLM, 112 SLM, 113 SLM, and 202 ARB) through "Orange" technology?

To answer this question, the relevant literature, studies, and research concerned with educational data mining and prediction were reviewed, such as: The study of [27] and the study of [28]. Linear Regression, Random Forest, KNN, Tree, SVM algorithms were used via Orange technology as follows:

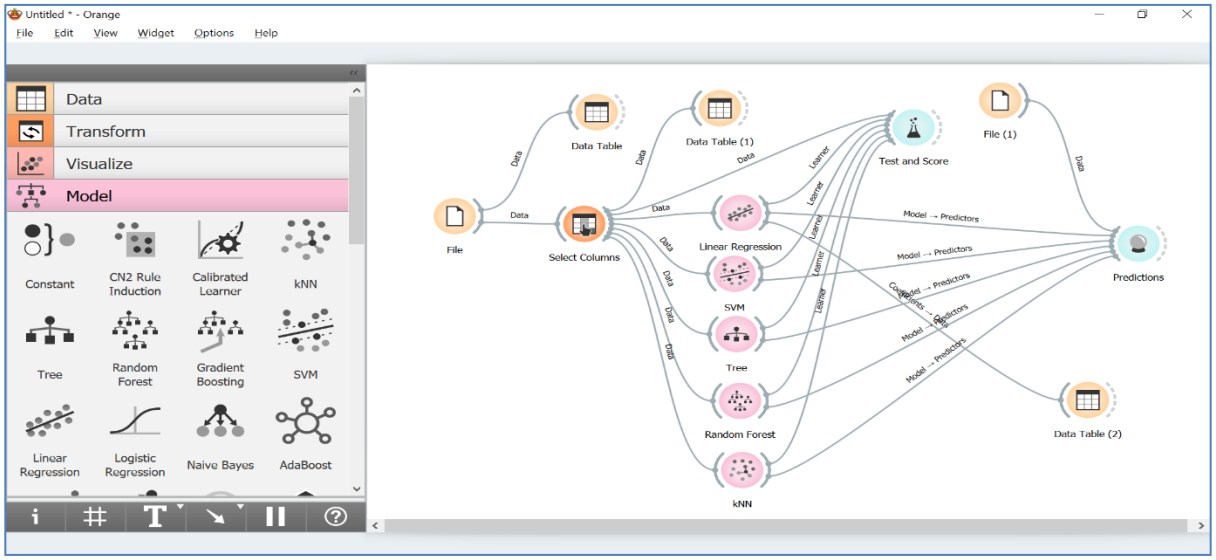


Fig. 12 the previous figure shows the use of Linear Regression, Random Forest, KNN, Tree, SVM algorithms via (Orange) technology.

After analyzing the data and using Linear Regression, Random Forest, KNN, Tree, SVM via Orange technology, the research questions were answered as follows:

1- For (111 Salam) course, the results were as follows:

Linear Regression, SVM, Tree, Random Forest, KNN algorithms were used to predict the students' data, and the results were according to Table (4).

Table 4: The results of the students' algorithms for (111 SLM) course

N	Model	MSE	RMSE	MAE	R^2
1	Random Forest	0.141	0.375	0.145	0.160
2	Tree	0.142	0.377	0.145	0.152
3	Linear Regression	0.146	0.383	o.146	0.125
4	KNN	0.163	0.403	.160	0.028
5	SVM	0.166	0.407	0.172	0.009

Table (4) shows that Random Forest algorithm is the best for classifying students' scores, followed by Tree, Linear Regression, KNN algorithm, then SVM algorithm. The following figure shows the academic prediction scores according to these algorithms:

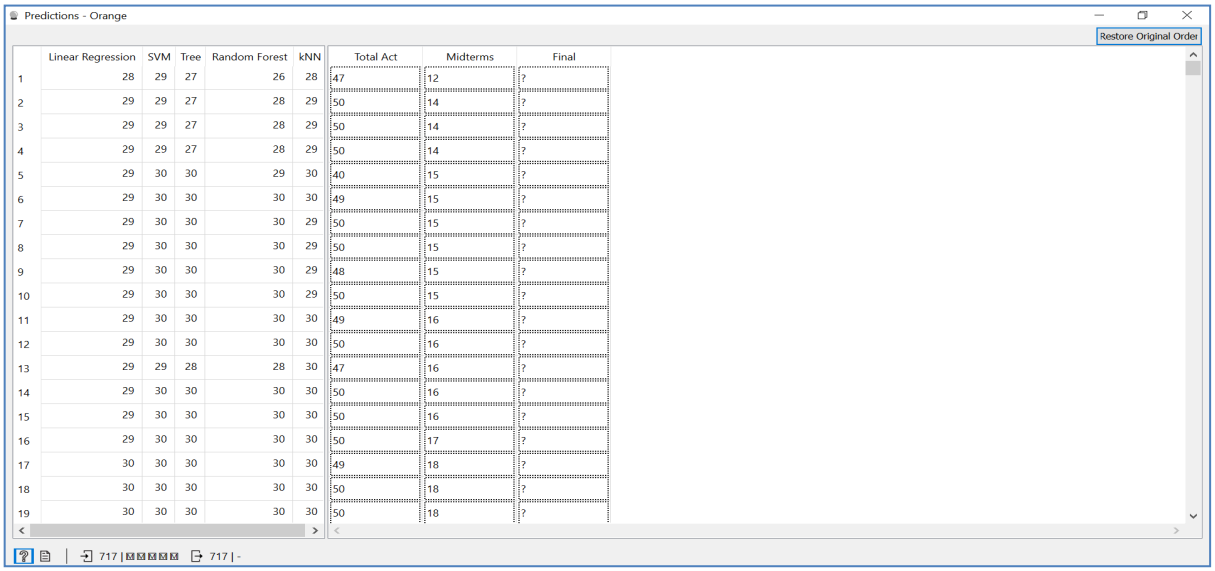


Fig. 13 the previous figure shows the students' scores in the semester and final exams, as well as the total score

Academic Performance Prediction Equation:

Student's final score=	The student's midterm score x 0.213	+	25,378

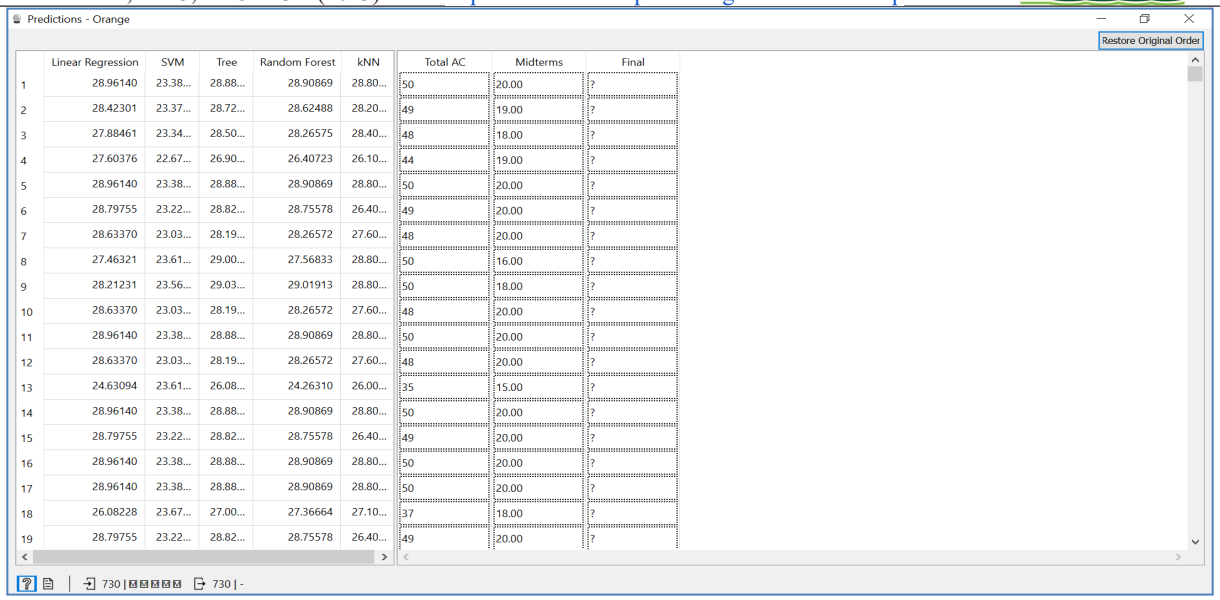
Based on the foregoing, it is clear that the students' scores in the midterm activities are not taken into account when predicting their final scores in the (111 SLM) course. Rather, only the students' midterm scores are taken into account in predicting their final scores.

2- For (112 SLM) course, the results were as follows:

Linear Regression, SVM, Tree, Random Forest, KNN algorithms were used to predict the academic performance of King Khalid University students. The results were according to Table (5).

Table 5: The results of the students' algorithms for (112 SLM) course

N	Model	MSE	RMSE	MAE	R^2
1	Linear Regression	3.012	1.735	1.224	0.149
2	Random Forest	3.204	1.790	1.224	0.095
3	KNN	3.362	1.834	1.288	0.050
4	Tree	3.440	1.855	1.264	0.028
5	SVM	24.565	4.956	4.585	-5.943



	Linear Regression	SVM	Tree	Random Forest	kNN	Total AC	Midterms	Final
1	28.96140	23.38...	28.88...	28.90869	28.80...	50	20.00	?
2	28.42301	23.37...	28.72...	28.62488	28.20...	49	19.00	?
3	27.88461	23.34...	28.50...	28.26575	28.40...	48	18.00	?
4	27.60376	22.67...	26.90...	26.40723	26.10...	44	19.00	?
5	28.96140	23.38...	28.88...	28.90869	28.80...	50	20.00	?
6	28.79755	23.22...	28.82...	28.75578	26.40...	49	20.00	?
7	28.63370	23.03...	28.19...	28.26572	27.60...	48	20.00	?
8	27.46321	23.61...	29.00...	27.56833	28.80...	50	16.00	?
9	28.21231	23.56...	29.03...	29.01913	28.80...	50	18.00	?
10	28.63370	23.03...	28.19...	28.26572	27.60...	48	20.00	?
11	28.96140	23.38...	28.88...	28.90869	28.80...	50	20.00	?
12	28.63370	23.03...	28.19...	28.26572	27.60...	48	20.00	?
13	24.63094	23.61...	26.08...	24.26310	26.00...	35	15.00	?
14	28.96140	23.38...	28.88...	28.90869	28.80...	50	20.00	?
15	28.79755	23.22...	28.82...	28.75578	26.40...	49	20.00	?
16	28.96140	23.38...	28.88...	28.90869	28.80...	50	20.00	?
17	28.96140	23.38...	28.88...	28.90869	28.80...	50	20.00	?
18	26.08228	23.67...	27.00...	27.36664	27.10...	37	18.00	?
19	28.79755	23.22...	28.82...	28.75578	26.40...	49	20.00	?

Fig. 14 the previous figure shows the students' scores in the semester and final exams, as well as the total score

Table (5) shows that the Linear Regression algorithm is the best for classifying students' scores, followed by Random Forest algorithm, then KNN, Tree, and SVM algorithm. The following figure shows the prediction degrees according to these algorithms:

Academic Performance Prediction Equation:

Student's final score = (ACTI) x 0.164 + Midterm Scores x 0.375 + 13,278

Based on the foregoing, it is clear that the students' scores in ACTI and the midterm exam are taken into consideration when predicting the students' final scores, although the midterm exam scores are greater than ACTI scores in predicting their final scores.

3- For (113 SLM) Course, the results were as follows:

Linear Regression, SVM, Tree, Random Forest, KNN algorithms were employed to predict the students' data, and the results were according to Table (6).

Table 6: The results of the students' algorithms for (113 SLM) course

N	Model	MSE	RMSE	MAE	R^2
1	Linear Regression	2.526	1.589	1.151	0.254
2	Random Forest	2.705	1.645	1.187	0.201
3	Tree	2.738	1.655	1.196	0.192
4	KNN	2.951	1.718	1.242	0.129
5	SVM	16.037	4.005	3.702	-3.735

Table (6) shows that the Linear Regression algorithm is the best for classifying students' scores, followed by Random Forest algorithm, then Tree, KNN, and SVM algorithm. The scores for predicting the academic performance of King Khalid University students were according to these algorithms as follows:

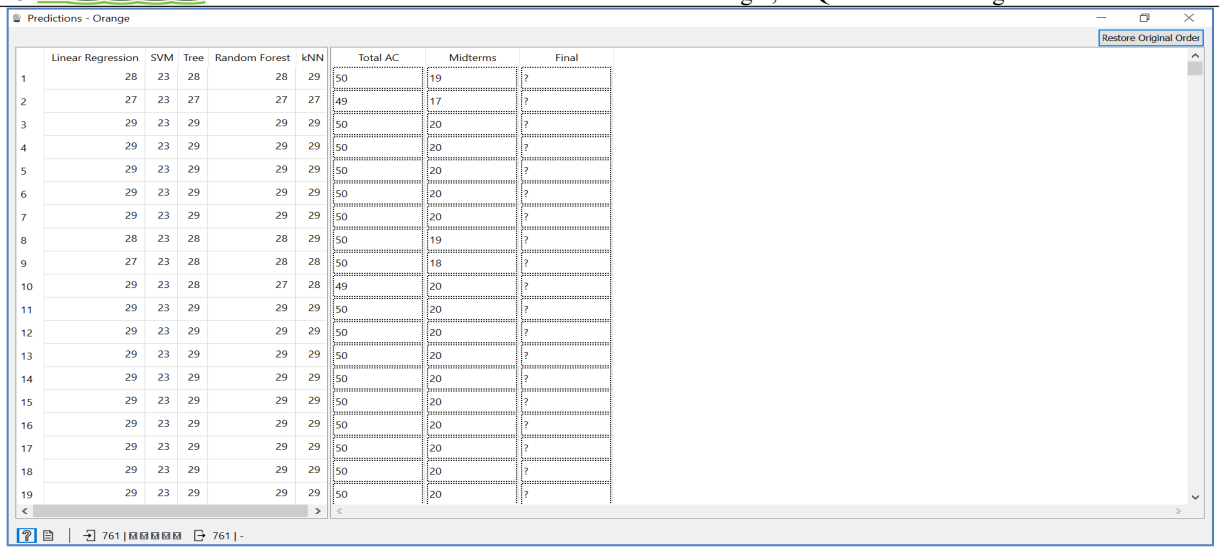


Fig. 15 the previous figure shows the students' scores in the semester and final exams, as well as the total score

Academic Performance Prediction Equation:

Student's final score = (ACTI) x 0.113 + Midterm Scores x 0.761 + 8,137

Based on the foregoing, it is clear that students' scores in ACTI and the Midterm exam are taken into consideration when predicting their final scores, although the midterm exam scores is greater than ACTI scores in predicting the students' final scores.

4- For (202 ARB) Course, the results were as follows:

Linear Regression, SVM, Tree, Random Forest, KNN algorithms were employed to predict the students' data, and the results were according to Table (7).

Table 7: The results of the students' algorithms for (202 ARB) course

N	Model	MSE	RMSE	MAE	R^2
1	Linear Regression	0.771	0.878	0.668	0.146
4	KNN	0.900	0.949	0.703	0.003
3	Tree	0.908	0.953	0.701	-0.0006
2	Random Forest	0.909	0.953	0.706	-0.007
5	SVM	2.806	1.675	1.475	-2.108

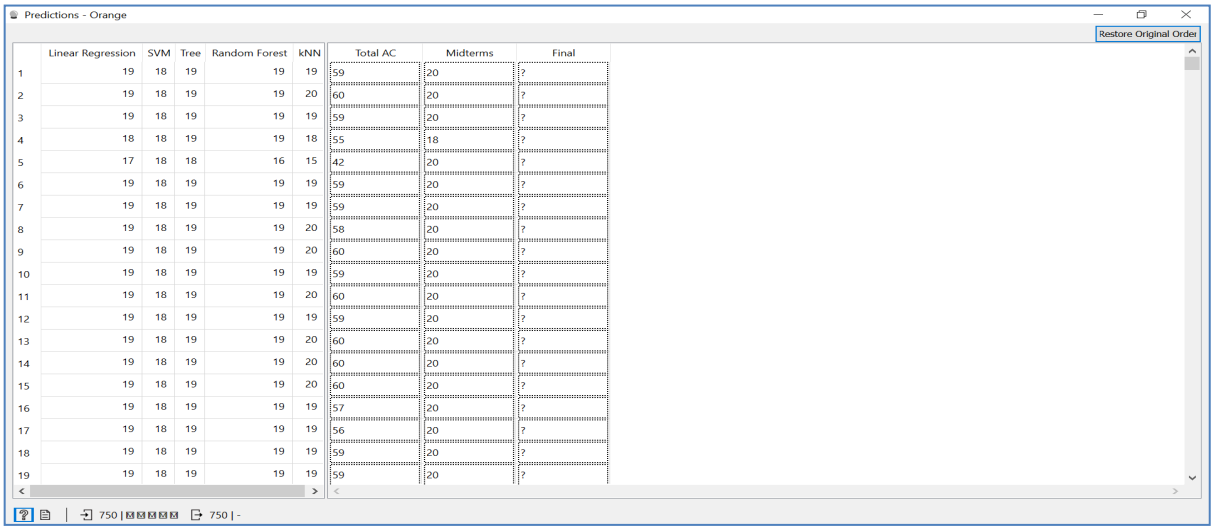


Fig. 16 the previous figure shows the students' scores in the semester and final exams, as well as the total score

Table (7) shows that the Linear Regression algorithm is the best for classifying students' scores, followed by KNN, then

Tree, Random Forest, and SVM. The following figure shows the scores for predicting academic performance according to these algorithms:

Academic Performance Prediction Equation:

Student's final score = (ACTI) x 0.147 + Midterm Scores x 0.186 + 6,897

Based on the foregoing, it is clear that the students' scores in the ACTI activities and the Midterms are taken into account when predicting the academic performance of King Khalid University students, and that the Midterm scores are greater than the ACTI scores in predicting students' final scores.

Suggested Researches:

In light of the findings, the research suggests making the following researches:

1- A comparison between Orange technology and Weka technology to discover educational patterns and predict the academic performance of university students.

2- Utilizing Orange technology to predict the academic performance of university students through the Logistic regression algorithm.

3- Utilizing Orange technology for semantic analysis of images and feelings, and predicting the academic performance of university students.

4- Utilizing Orange technology to deal with ROC Analysis and ML algorithms

Acknowledgements

The author extends their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through General Research Project under grant number (233/1443 AH).

Conflict of interest

The authors declare that there is no conflict regarding the publication of this paper.

References:

[1] Ahmed, Ahmed Abdel-Muttalib, Abdullah, Mohamed Hassan. Educational Data Mining using Apriori Algorithm to Evaluate Student Performance and Improve Production, *White Nile Journal of Studies and Research*, 1 (11), 9-36 (2018).

[2] Sayed, Ahmed Fayez, (2016). Open Source Data Mining Tools - Analytical Evaluation Study, *Taibah University Journal*, *10*(1), 791-865.

[3] Hilbert, M. Big data for development: A review of promises and challenges. *Development Policy Review*, *34*(1), 135-174 (2016).

[4] Abdullah, Salwa. *Artificial Intelligence*. Medad for publishing and distribution, (2021).

[5] Al-Mutairi, Abdullah. Big data and the Big Theory Dream, *Al-Faisal Scientific Journal*, Year 54, *55* (1), 33-38 (2017).

[6] Karati, Hana. *Using Big Data in Technology Companies and User Privacy* [Unpublished Master's Thesis], May 8 University, Algeria, (2017).

[7] Rostami, K. H., Samadi, S., Omrani, H., Margavi, A. K., Asadzadeh, H., & Nazari, H. (2011). Data Mining and Application in Accounting and Auditing. *Journal of Education and Vocational Research*, *2*(6), 211-215.

[8] Koksal, G., Batmaz, I., & Testik, M. C. A review of data mining applications for quality improvement in the manufacturing industry. *Expert Systems with Applications*, 38(10),

13448-13467 (2011).

[9] Al-Aklubi, Ali Dib. Big Data and Decision Making at King Saud University: Evaluation Study of Itqan system. *QSCINCE*, 1(2), 1-12(2018).

[10] Khamis, Mohamed Attia. *Recent Trends in Educational Technology and Areas of Research*, Academic Center for Publishing and Distribution, (2020).

[11] Mwalumbwe, I. & Mtebe, J., S. Using Learning Analytics to Predict Students' Performance in Moodle Learning Management System: A Case of Mbeya University of Science and Technology. *The Electronic Journal of Information Systems in Developing Countries*, 79(1), 1-13 (2017).

[12] Muhammad, Malik Khalid. A Framework based on Learning Analytics for Big Data in learning Management Systems to Develop the Design and Production of E-learning Courses. *Educational and Social Studies*, *24*(4), 343-426(2018).

[13] Ahmed, Ahmed Abdel-Muttalib, Abdullah, Mohamed Hassan. Educational Data Mining using Apriori Algorithm to Evaluate Student Performance and Improve Production, *White Nile Journal of Studies and Research*, 1 (11), 9-36 (2018).

[14] Simon, Phil. (2020). Too Big to Ignore, The Business Case for Big Data, (Translated by: Atef Mohammed Al-Omari, Mabrouk Saleh Al-Rakhmi), King Saud University Press.

[15] Nasser, Reem Jamal. Using Data Mining Techniques in Learning [Unpublished master's thesis], Al-Baath University, Syria, (2019).

[16] Horst, R. M. Higher Education Executives and Data-Driven Decision Making- A Phenomenological Study [Unpublished PhD thesis], Concordia University, Canada, (2020).

[17] Maqnani, Sabrina, Shabila, Muqaddam. The Role of Big Data in Supporting Sustainable Development in the Arab Countries, *Journal of Information and Technology Studies*, *1* (4), 1-14(2019).

[18] Youssef, Nawras Kazem, (2016). Does Demographic Data affect Student Graduation?: Analytical study using data mining, Journal of Humanities, Scientific and Social Sciences, 1(1), 238-265

[19] Ibrahim, Mahmoud Ahmed. The Feedback Pattern based on Educational Analyzes in an E-learning Environment to develop the Skills of Producing Websites and Self-Regulation among Primary School Students, *Journal of Education Technology - Studies and Research*, 33 (1), 1-75 (2017).

[20] Pitler, Howard, Hubbell, Elizabeth Ross, Kuhn, Matt, (2018). Using Technology With Classroom Instruction That Works (Translated by: Sous Mesto), Obeikan Library.

[21] Al-Mubarak, Saad Al-Badawi Al-Mubarak. *Using Data Mining Techniques to Discover Patterns of Academic Achievement Effects for Secondary School Students* [Unpublished Master's Thesis]. University of Neelain, The Sudan, (2017).

[22] Al-Nuwaisir, Khalid bin Rashid. *Twenty-First Century Skills: Education and the challenges of the digital age*. Takween Press, (2021).

[23] Floridi, Luciano. *The Fourth Revolution - How the Infosphere is Reshaping Human Reality*. Mohammed Rashid Al Maktoum Knowledge Foundation, UAE, (2017).

[24] Nayef, Qutayba Nabil, and Ayoub, Mohieddin Khalaf. Using K-Means Algorithm for Clustering in Data Mining with an Applied Reality, *Journal of Economic and*

Administrative Sciences, *22*(91), 389-406 (2016).

[25] Hussein, Alwani. *Data Management in the Era of National Transformation*, Elm company, (2016).

[26] Othman, Abdul Rahman. *Learn Hadoop, Umm Al Qura University*: College of Computer in Qunfudhah, KSA, (2017).

[27] Abu Qurun, Mazahir Abd al-Rahman, Hijazi, Muhammad Othman. *Comparison of Educational Data Mining Algorithms for Decision Support in Sudanese Higher Education Institutions A case study - Student Performance in the Faculty of Computer Science and Information Technology* - Al-Zaeem Al-Azhari University [unpublished PhD thesis], Al-Neelain University, The Sudan, (2018).

[28] Abdullah, Abu Bakr Muhammad Mustafa. *Data Mining of Primary Level Students to predict their Scores in Mathematics* [unpublished master's thesis], Africa International University, The Sudan, (2020).