Analyzing and Predicting Students performance using Artificial Intelligence: A Case Study in Sultanate of Oman

Fatma Mohammed Haider AL-Zadjali, Dr. Syed Zakir Ali

Student of MSc-IT, Department of Computing, Faculty, Department of Computing
Middle East College, Muscat, Sultanate of Oman
Corresponding Author: Fatma Mohammed Haider

Abstract: *This research proposed to analyze and predict grade 10 students' performance in Mathematics for the Sultanate of Oman. Several factors were considered in this research that may influence students' performance such as factors related to the student textbook, school environment, factors associated with the teacher, and factor related to student behaviors and human and physical resources. The primary data was collected using questionnaires and secondary data such as students result in mathematics subject collected from the Ministry of Education (educational portal database). A sample of 1300 students with the low performance in Mathematics, was taken from 13 schools especially government schools in the Muscat region for the school year 2018-2019 to develop classification model which can efficiently classify and predict grade 10 students' performance in Mathematics based on some related attributes which can be helpful to guide the students with the low performance initially. For the classification model, two algorithms have been applied Neural Network and Decision tree and selected the best algorithm based on the accuracy. The results of the experiments shown that neural network classifier outperforms the decision tree, by achieving overall accuracy of 91%, both the classifiers perform better with attribute filter while removed of zero effect variables. The finding revealed that the frequent absence of the teacher and student's internal mark (class test performance) are the most effective factors that affect students' performance. This research will help the teachers to know the students' performance in advance and take suitable action at the right time, as well as provide appropriate counseling and advising to the student, hence failing ratio will reduce.*
Keywords: Machine learning process, Neural Network, Decision Tree

----

I.   Introduction

The past several years have observed a rapid development in the application in the field of Artificial Intelligence in the educational area, supported by the evidence that it helps educational institutions to learn useful and new knowledge about students. Educational data mining is an emerging discipline involved with developing approaches for examining the unique types of data that come from educational institutions. One of the important areas of data mining application is the development of student models that would classify and predict student performance to help educators to enhance the structure of their course and early identify those students who need special attention (Wakelamet al.2015).

Education is considered as an essential part of human resource progress as it is the first step for every human movement and ensures attainment of skills and knowledge that allow individuals to enhance their quality of life and increase their efficiency and effectiveness. Increase of social effectiveness will lead to producing of new sources, which improve the economic development of a country (Farooqet al.2011).

Students' mathematical achievement is very significant at the national level as shown by the intense interest of Sultanate of Oman on participating in Trends in International Mathematics and Science Study (TIMSS). TIMSS is an international assessment of the mathematics and Science achievement of the fourth and eighth-grade students around the world, conducted every four years in the United States (NCES 2019).

As per the rating system in the Ministry of Education the results measures on five benchmarks are Excellent (90-100), Very Good (80-89), Good (65-79), low performance (50-64) and fail (0-49). According to the data given by the Ministry of Education, Sultanate of Oman, there are a total of 8342 students in the Muscat region who are enrolled for grade-10 in the academic year 2018-19. After analyzing the data, it is found that the total number of students with low performance are 2233 and the total number of students who failed in first semester are 2186.

These statistics proved that students of grade 10 have a weakness in mathematics hence ; there should be some investigations to identify the factors that affect students' performance in mathematics as limiting the number of students who fail in the exams is considered as an important task therefore the capability to predict weak students before the final exam is very significant for the teacher because by identifying students with a

---

low performance the teacher will be able to inform the students during their study and provide them additional support to improve their performance.

According to Rashid and Aziz, 2016 there are several factors influence students' performance such as economic status, social background, geographical belongingness, age, parents' educational level, gender, etc. And this affects the achievement of the target of any educational system.

This research attempts to investigate the factors that affect grade10 students' performance in mathematics for the school year 2018-2019 especially government schools in the Muscat region. Classification algorithms are used to evaluate and predict students' performance at the end of the semester based on several attributes such as student textbook, school environment, teacher, student behaviors, class test performance, and human and physical resources.

II.   Related Work

During the last few years, many studies have been conducted to analyze and predict student's performance using artificial intelligence techniques. In this section, several research articles are being considered to analyze them and find out different techniques and algorithms that were applied, investigated factors that affect students' performance and results obtained by the researchers.

Rashid and Aziz (2016) used artificial neural network technique to investigate the factors that are affecting student's performance on General English courses for first-year students at Salahaddin University. They used Five hundred students' data from six departments at the engineering college, and they also collected data regarding student's tutor's category, parental socioeconomic status, high school scores, and high school type. From the data analysis, they determined that departments and tutors are critical factors affecting academic performance and factors like GPA, gender, Age and parental degree do not have any effect on academic achievement.

Lakshmi, Martin and Venkatesan (2013) conducted a study to predict the most impacting factor that affects the performance of student using a genetic algorithm. They selected 120 students from the IT department and analysis quantitative factors such as mathematical, theoretical, departmental, elective subjects, attendance, lab, and project. After an investigation of factors, they found that mathematical subject and theoretical subject have a high impact on the students' performance.

Ruby and David (2015) conducted a study to compare the predicted results of the students with 12 attributes such as parent education, family income, attendance, CGPA, previous course studied, theory marks, stay, urban and the student performance predicted by recognized high influence factors using data mining multi-layer perceptron algorithm. 165 PG students of computer application course from Arts and Science college data collected. The authors proved that the chosen attributes that identified are high influencing factors in predicting student performance using MLP algorithm and they believed that their study would help the institution to know students' academic status in advance and focus on weak students to enhance their performance.

In 2017 Yassein, Helali and Mohamed studied factors that affect module success rate and students' performance. They collected a sample of 150 students' records from Najran University in Saudi Arabia; they did a correlation between students and modules attributes individually. C5.4 algorithm applied to predict the success rate, and they found that there is a relation between assignment, practical sessions and success rate of modules, this means that modules which consist of practical activities has a success rate much better than modules depends on a theoretical. Also, modules that include assignments have a better success rate. The research indicated that the most factors that affect student performance are attendance, final exam and midterm exam grades.

Mousa and Maghari (2017) used multiple data mining classification algorithms such as (Decision tree, Naïve Bayes and K-NN) to proposed student performance prediction model. One thousand one hundred student records were collected from gaze preparatory male school and observed that the decision tree gives the best results. The results of their study indicate that first term and previous year results have more effect on the students' performance and social circumstances have little influence on the students' performance.

Apolinar (2019) conducted a study to predict students' performance in data structures and algorithms subject of business information technology course. A sample of 108 students' records was selected; the J48 algorithm was used to create a decision tree model using WEKA application software and for model accuracy receiving operating curve and K-fold cross-validation was used. The research focused on variables of data structures and algorithms subjects such as quizzes, lab exercises, midterms, and the final exam. The results confirmed that the decision tree model was able to predict the accuracy of failed around 91.67% and 85.31% for a pass and 79.41% for conditional. The student should score in midterms more than 66.12% and 72.30% in the final exam to pass in the subject. That analysis indicated that the final exam is the most factor affect student performance.

Hamoud, Hashim and Awadh (2018) applied data mining techniques and developed a model based on decision tree algorithms using wake application software, and three classifiers were used such as Random tree, J48 and REP tree to predict factors affect student failure and success. The authors deepened on primary data they designed questionnaires on Google form to the students of computer science and Information Technology at the University of Basrah. The surveys consisted of 60 questions which considered as an input variable and covered different fields such as social activity, health, academic performance, and relationships. The number of respondents was 161 students and comparison between three classifiers was done to measure their performance. The results indicated that J48 algorithm is the best algorithm based on its performance and attributes like credits, GPA, fresh food and list important notes had the most significant effect on student success, On the other hand, work, gender, stage, status, and age has less impact.

Mythili and Shanavas (2014) conducted a study to evaluate and analyze the performance of school students using data mining classification techniques such as Random forest, J48, decision table, multilayer perceptron, and IBI. The authors focused on factors like parents' educational level, attendance, economic status, student result, and gender as input variables; they used a sample of 260 students' records. The results of the analysis conclude that attendance related to student performance and parents' educational, gender, economic status, and results are potential factors that affect students' performance. The research indicated that Random Forest performance is better than other selected algorithms.

Machine Learning Process
Data Collection

A sample of 1300 students with low performance was taken from 13 schools in Muscat Region for the school year 2018-2019; where six schools are Male's, and seven are Female's schools. The descriptive analysis was done using Microsoft Excel as depicted in Figuer1 , the collected data shows that the 66% are male out of which 540(41%) failed, and 329(25%) with the low performance and 34% are female students; where 164(13%) failed and 267(21%) with the low performance.
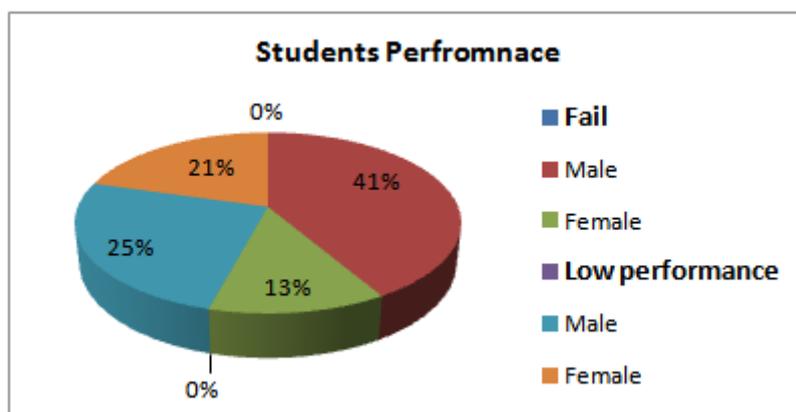

Figure 1: Total number of the students

Data Collection Procedure

This research used both primary and secondary data; where the primary data was directly obtained from the students' using questionnaire. The surveys involved of items that handled demographic data on gender, factors related to student textbook, school environment, factors associated with the teacher, factor related to student behaviors and factors related to human and physical resources. The items of the survey were Closed Ended questions on ordinal and nominal scale and consisted of 26 items. The copies of the questionnaire was distributed by the researcher after notification letter send to the schools by **Directorate General of Education in Muscat Governorate** of the classes that were be used in the study. The teachers administered the questionnaire to the selected students and got the responses back at the same time and give it back to the researcher. The questionnaire filled by 1300 students as mentioned before. Students were asked to include their name and school name to the answer sheet to link it with the secondary data. Secondary data such as student's exam results and class test performance in mathematics subject of the school year 2018-2019 was collected from the Ministry of Education (Educational portal database).

Analyzing and Predicting Students performance using Artificial Intelligence: A Case Study in Sultanate of Oman

Data Preprocessing

In this study, all answers of the questionnaire are converted from a nominal and ordinal to the numeric data type. Questions with the answers Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree are represented in form of 1,2,3,4,5 where 1= Strongly Disagree,2= Disagree,3= Neutral,4= Agree,5= Strongly Agree and Male=1 and Female=2 and students internal marks are in the range of 1 to 40. The output variable is student performance at the end of the semester and converted to zero and one; where Zero = Fail and one=low performance. As per rating scale followed by the Ministry of Education, marks between 0-49 = Fail and marks between 50-64 consider as low performance. The input variables with their normalized values presented in Table1.At this stage data has been checked for completeness, all questionnaires were answered with the response rate of 100% and from secondary data file which consists of students' marks; students with zero score has been removed from the data set. The data has been collected in the form of MS Excel files and upload in R studio software; Data normalized using Min-Max normalization known as feature scaling to make a linear transformation on the data to scale the data in the range of zero and one. As normalization is required for most of algorithm to improve the performance of the model .Min-Max normalization performed by the equation as follows; where min and max are the minimum and maximum values of the attribute X and X is the set of observed values of X (KUMAR JAIN and BHANDARE, 2014).

$$(xscaled) = \frac{x - xmin}{xmax - xmin}$$

After the normalization, the data has been split randomly in to training set 80% and testing set 20%.

Filtering

Filtering is the process of checking the unwanted attribute; as removing attributes with zero effect will enhance the performance of the model. Filtering is considered as one method of feature selection (Sharma, 2018).

In this research, **Information. Gain** function from **F-Selector** package was used to evaluate the attributes; the function tells how much information each independent variables gives about the dependent variable. Table 1 represented the weight of each attributes with the respect of the class. Variables name's represented in Table 2

Variables	Weight
V9	0.37365125
Internal Mark	0.34535484
V26	0.30257844
V14	0.19445182
V21	0.15877488
V22	0.13698284
Gender	0.10405403
V10	0.09542863
V15	0.07295616
V5	0.07204624
V19	0.07183412
V11	0.04912860
V25	0.02193181
V4	0.02002375
V3	0.00000000
V6	0.00000000
V7	0.00000000
V8	0.00000000
V12	0.00000000
V13	0.00000000
V16	0.00000000
V17	0.00000000
V18	0.00000000
V20	0.00000000
V23	0.00000000
V24	0.00000000

Table 1: Variables weights

Factor	Variable	Domain	Normalized Value
Demographic Data	Gender	Male-Female	Male=1,Female=2
	V3	The textbook questions and activities don't cover	1,2,3,4,5

		ideas related to mathematical procedures and patterns.	
School Text Book	V4	Topics in the textbook are not related to each other.	1,2,3,4,5
	V5	The design of the textbook in terms of colors, images, charts, font size, and arrangement of the ideas and its sequence doesn't help me to study.	1,2,3,4,5
Teacher	V6	The teacher doesn't diversify activities to develop my mathematical thinking.	1,2,3,4,5
	V7	The practical activities provided by the teacher are for low capabilities.	1,2,3,4,5
	V8	Lack of teacher motivation in teaching mathematics.	1,2,3,4,5
	V9	The Frequent absence of Math teacher.	1,2,3,4,5
	V10	The teacher focuses only on students with high performance.	1,2,3,4,5
	V11	Ensure to complete the curriculum according to the plan without considering achieving the learning outcomes.	1,2,3,4,5
	V12	Lack of mathematics learning tools available with the teacher.	1,2,3,4,5
	V13	The teacher doesn't encourage us to build our new knowledge in mathematics.	1,2,3,4,5
	V14	The teacher focuses only on activities mentioned in the textbook without using external resources.	1,2,3,4,5
Student	Internal Mark	Class test performance	1,2,3,4,5,6……40
	V15	Lack of training us with questions related to the high levels of critical thinking.	1,2,3,4,5
	V16	I don't trust in my abilities to write in the exam.	1,2,3,4,5
	V17	Mathematics subject is in the last classes of the daily school schedule.	1,2,3,4,5
	V18	I have health problems.	1,2,3,4,5
	V19	I spent most of my time watching TV, playing games, and using social media sites.	1,2,3,4,5
	V20	My weakness in reading impedes my understanding of the questions.	1,2,3,4,5
	V21	I don't have the basic mathematics skills	1,2,3,4,5
	V22	I have difficulty in understanding what is required when I answer questions	1,2,3,4,5
Human and Physical Resources	V23	Lack of facilities in schools to do our activities.	1,2,3,4,5
	V24	There is a lack of suitable environments in school for teaching and learning Mathematics.	1,2,3,4,5
	V25	Lack of teaching aids and sources in the school.	1,2,3,4,5
	V26	Lack of communication between schools and parents to follow students' performance.	1,2,3,4,5

Table2: Classification Input variables

Cross Validation

In this research, before building classification models 10 fold cross-validation was performed using two classification models are neural network (nnet) and classification tree (cart). The accuracy of algorithms as shown in table 3; from table 3 it is observed that the accuracy of two classifiers approximately is the same.

Model	Accuracy
Neural Network (nnet)	0.8876900
Classification tree (cart)	0.8815405

Table 3: The accuracy of classifier using 10 fold cross validation

Two classifiers have been selected for this study Neural Network Decision tree and by building these two classification models, we can get an idea of the accuracy of the algorithms on our test set.

Neural networks Classification Model

In this study **Resilient Back Propagation algorithm** used to train the neural network, the data have been trained with 2, 3, 4 and 5 hidden layers, the Growing method is followed to set the number of hidden layers; the 5 hidden layers gave the best result. Two cases have been studied in this research as follows.

Case 1:

In this case, all 26 input variables have been used in the training; student result has been used as an output. Initially, neuralnet, NeuralNetTools, caret, packages were installed, and then the data divided into two sets training 80% and testing 20%. Initially The network of 2 hidden layers has been trained with err.fct = "sse", linear.output = FALSE, threshold =0.1, algorithm = "rprop+" and act.fact=logistic; where is **threshold** is a value specify the partial derivatives of the error function as stopping standards, **err.fct** is function used to calculate the error **'sse'** stand for the Sum of Squared Errors, linear output set to False because we are dealing with classification problem and algorithm used is Resilient Back Propagation as mentioned earlier because backprpopagation was tried in the beginning and was very slow.

The accuracy of the model with the all variables with two hidden layers was **0.8576923%,** then the model was trained with three hidden layers and the accuracy of the model was **0.85%**and with four hidden layers the accuracy was **0.8846154%;** with the five hidden layer the accuracy of the model was **0.8923077%** and misclassification error was **10%**, the model classified 15 students as Fail but in real they passed and classified **13** students as Failed where they were passed. The model with five hidden layers gives the best result; the results of five hidden layers are show in Table 4.

No	Factor	Quantity	Percentage
1	Number of Variables	26	_
2	Accuracy of the model	_	0.8923077
3	Misclassification	28	0.1076923
4	Number of classified passes as pass	111	_
5	Number of classified passes as Fail	15	_
6	Number of classified Fail as Failed	121	_
7	Number of classified Fail as pass	13	_

Table4: Five Hidden layers Results with all variables

Case 2:

Modified model, in this case all variables with Zero weight as shown in Table 3 have been removed, thence there are no zero affect variables .As it is seen from the results of table 3, the variables V3, V6, V7, V8, V12, V13, V16, V17, V18, V20, V23, V24 have zero effect on students performance. The data is retrained with 14 variables and 5 hidden layers and with the same training and testing set of case one training set of 80% and testing 20% .The variables of the new model represented in Table 5; the modified model is shown in Figure 3. However there are some variables such as V10, V15, V5, V19, V11, V25, and V4 with the small weights but the accuracy of the model **without** these variables led to 0.8730769%. While the accuracy of modified model is 0.9115385%, the results of predicated model shown in Table 6.

Variables	Weight
V9	0.37365125
Internal Mark	0.34535484
V26	0.30257844
V14	0.19445182
V21	0.15877488
V22	0.13698284
Gender	0.10405403
V10	0.09542863
V15	0.07295616
V5	0.07204624
V19	0.07183412
V11	0.04912860
V25	0.02193181
V4	0.02002375

Table5: input Variables of modified model

No	Factor	Quantity	Percentage
1	Number of Variables	14	_
2	Accuracy of the model	_	0.9115385
3	Misclassification	23	0.08846154
4	Number of classified passes as pass	113	_
5	Number of classified passes as Fail	12	_
6	Number of classified Fail as Failed	124	_
7	Number of classified Fail as pass	11	_

Table6: Five Hidden layers Results with 14 variables

Decision Tree Classification Model
Case 1:
In this case, all 26 input variables have been used in the training and same procedure follow as classification with neural network the data divided into two sets training 80% and testing 20% the model created using R-part packages. The model classified **14**students as Fail but in real they passed and classified **18** students as Failed where they were passed. The accuracy of the model with all variables was **87%** with misclassification of **12%** as shown in table 7.

No	Factor	Quantity	Percentage
1	Number of Variables	26	_
2	Accuracy of the model	_	0.8714859
3	Misclassification	32	0.1285141
4	Number of classified passes as pass	108	_
5	Number of classified passes as Fail	14	_
6	Number of classified Fail as Failed	109	_
7	Number of classified Fail as pass	18	_

Table7: Decision Tree classification Results with all variables

Case2:

In the second experiment all variables with Zero weight have been removed and the model classified **19**students as Fail but in real they passed and classified **10** students as Failed where they were passed. The accuracy of the model was **88%** with misclassification of **11%** as shown in table 8

No	Factor	Quantity	Percentage
1	Number of Variables	14	_
2	Accuracy of the model	_	0.8884615
3	Misclassification	29	0.1115385
4	Number of classified passes as pass	105	_
5	Number of classified passes as Fail	19	_
6	Number of classified Fail as Failed	126	_
7	Number of classified Fail as pass	10	_

Table8: Decision Tree classification Results with 14 variables

III. Discussion of the Results

Two classification techniques have been applied on the dataset to build the classification model; techniques are Neural Network algorithm and Decision tree algorithm. Algorithms applied in two cases first with all variables and second case with filtered variables. Comparison of algorithms represented in Table 9; as it is seen from the results neural network classifier has more accuracy then decision tree and both classifiers perform better with attribute filter while removing zero effect variables. The first row exposes the performance of the algorithms without attribute filter and the second row presents the performance of the algorithms after applying the attribute filter. The effectiveness of the attribute filter on both of algorithms is evident the accuracy in both algorithms are increased. The experiments have shown that Neural Network models can be successfully used to predict the expected performance of grade 10 student's in Mathematics at the Sultanate of Oman. In the first attempt, the network predicated 89 %(232 out of 260 tested students) and 91% for the modified model (237 out of 260 tested students).**Moreover Rashid and Aziz (2016)** applied neural network algorithm to investigate the factors that are affecting student's performance on General English courses for first-year students at Salahaddin University. Their results show that neural network model performed better when they remove zero affect variables which are consistent with our results.

Table9: Performance of the classifiers

Evaluation Criteria		Classifiers	
		Neural Network	Decision Tree
Without Attribute Filter	Accuracy	89%	87%
With Attribute Filter		91%	88%

This study aims to test and explore the process of applying neural network and decision tree algorithms with a questionnaire of students to find out the factors that affect student performance. Based on the results of the models it can be said that a number of variables (factors) affect students performance and the accuracy of algorithms. Factors like **student textbook** have no effect on student performance; **teacher-related factors** such as the frequent absence of the teacher and focusing only on activities mentioned in the textbook without using external resources have affect one student performance. **Student related factors** such as class test performance, lack of their basic mathematics skills and their difficulty in understanding what is required when they answer questions affect students' performance. **Human and physical related factors such as Lack** of communication between schools and parents to follow up students' performance affects student performance and Lack of teaching aids and sources in the school has less effect on students' performance. The questionnaire contains many unimportant questions and has zero effect on students' performance as shown in Table 1.From data analysis it can be concluded that the **frequent absence of the teacher and student's internal mark (class test performance)** are the most effective factors effect on students' performance. **Besides,** this result consistent

with the study done by Yassein, Helali and Mohamed in 2017 that midterm exam grades is on of most effective factor that effect student performance in Najran University.

IV.   Conclusion

This research has proposed to analyze the factors that affect grade 10 students' performance in Mathematics for Sultanate of Oman. This has been done in order to help the teachers to know the students' performance in advance and take suitable action on students such as individual counseling, appropriate advising etc., at the right time which can minimize the failure rate. This was achieved by implementing machine learning algorithms and developed classification models. For the classification model, supervised learning algorithms implemented using Neural Network and Decision tree. From the data analysis, it is concluded that the frequent absence of the teacher and student's internal mark (class test performance) are the most effective factors that effect on students' performance.

References

[1].   Wakelam ,E.Jefferies ,A, Davey ,N and Sun ,Y(2015)'The Potential for Using Artificial Intelligence Techniques to Improve e-Learning Systems ' [Online] Available from https:// pdfs.semanticscholar.org/ 641e/ 9fe2856abaa7dbf59c3712aa2ad470d49242.pdf[14 Apr 2019]
[2].   Farooq,M.Chaudhry ,A.Shafiq ,M .and Berhanu ,G.(2011)'FACTORS AFFECTING STUDENTS' QUALITY OF ACADEMIC PERFORMANCE: A CASE OF SECONDARY SCHOOL LEVEL ' Journal of Quality and Technology Management' [Online]II(II) ,01 - 14Available from https:// www.researchgate.net/profile/Muhammad_Farooq108/publication/284150574_Factors_affecting_students'_quality_of_academic_p erformance_A_case_of_secondary_school_level/links/578c9fbb08ae254b1de84371/Factors-affectingstudentsqualityofacademicperformanceAcaseofsecondaryschoollevel.pdf?_sg%5B0%5D=0PbV8xN20jraXU3DXpbVHA_ YiYxayipg4gbr0_wcc4t_bvuLxc14wrmWoK4uLwDFG06lMBrlYQOW1Cp9S2b0g.f8obC93LH8Mv8dnDmHUsOoPnfALTqc8f7p 0xlHkWkoi2KJgOp9jKmSP724wEcbTXD6IBkEjTVQXuVV2U48NpQ&_sg%5B1%5D=NcD3ZoNFnFCjOXioGkOCugQkuBbfE DglANlFcuohblFayKtkOQRMrhgnrafPfb3KZKNpdEeIlG2Bae5eJwc6ogYrm6It_rLEosSw5iL8wpa.f8obC93LH8Mv8dnDmHUsO o-PnfALTqc8f7p0xlHkWkoi2KJgOp9jKmSP724wEcbTXD6IBkEjTVQXuVV2U48NpQ&_iepl
[3].   National Center For Education Statistics(2019)Trends in International Mathematics and Science Study (TIMSS) [Online]Available from https://nces.ed.gov/timss/ [17 Apr 2019]
[4].   Rashid,T. and Aziz ,N. (2016)' Student Academic Performance Using Artificial Intelligence' ZANCO Journal of Pure and Applied Sciences [Online] 28 (2), 56-69. Available from https:// www.researchgate.net/ publication/291262353_Student_Academic_Performance_Using_Artificial_Intelligence [1 Mar 2019].
[5].   Lakshmi, T. Martin, A.andVenkatesan, V. (2013) 'An Analysis of Students Performance Using Genetic Algorithm' Journal of Computer Sciences and Applications [Online] 1(4), 75-79.Available from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.825.1810&rep=rep1&type=pdf [1 Mar 2019].
[6].   Ruby, J. and   David, K. (2015) 'Analysis of Influencing Factors in Predicting Students Performance Using MLP - A Comparative Study' International Journal of Innovative Research in Computer and Communication Engineering [Online] 3(2), 2320-9801.Available from https:// pdfs.semanticscholar.org/ bb03/ fff265554fb512380156552e87cf22f45342.pdf [3 Mar 2019].
[7].   Yassein ,N. Helali,R and Mohomad,S.(2017)'Predicting Student Academic Performance in KSA using Data Mining Techniques' Journal of Information Technology & Software Engineering [Online] 7(5), 2165-7866 .Available from https://www.omicsonline.org/open-access/predicting-student-academic-performance-in-ksa-using-data-mining-techniques-2165-7866-1000213.pdf [14 Mar 2019].
[8].   Mousa, H and Maghari, A. (2017) 'School Students' Performance Predication Using Data Mining Classification' International Journal of Advanced Research in Computer and Communication Engineering [Online] 6(8), 2278-1021. Available from https:// www.researchgate.net/ publication/ 321920848_School_Students% 27_Performance_Predication_Using_Data_Mining_Classification [21 Mar 2019].
[9].   Apolinar, M. (2019) 'Using Decision Tree Algorithm to Predict Student Performance' Indian Journal of Science and Technology [Online] 12(5), 0974-5645. Available fromhttp:// www.indjst.org/ index.php/indjst/article/ view/140987/100004[3 Apr 2019]
[10].   Hamoud ,A. Hashim ,A and Awadh ,W.(2018)'Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis' International Journal of Interactive Multimedia and Artificial Intelligence [Online] 5(2), .Available fromhttps:// www.ijimai.org/journal/sites/ default/files/files/ 2018/02/ ijimai_5_2_3_pdf_30196.pdf[3 Apr 2019]
[11].   Mythili, M.Shanavas,M (2014) 'An Analysis of students' performance using classification algorithms' IOSR Journal of Computer Engineering[Online]   16(1),   2278-0661.Available   fromhttps://   www.researchgate.net/   publication/ 314445897_An_Analysis_of_students%27_performance_using_classification_algorithms[4 Apr 2019]
[12].   KUMAR JAIN ,Y and BHANDARE ,S.(2014)'Min Max Normalization Based Data Perturbation Method for Privacy Protection'International   Journal   of   Computer   &   Communication   Technology[Online]3(4),0975   –7449Availablefromhttps://pdfs.semanticscholar.org/7c1c/97d1c37c94eceaa4b3c4fe7310e30f55e4eb.pdf?_ga=2.13971519.9671758 37.1562239985-1164588325.1551376892[26June 2019].