Decision Trees for Very Early Prediction of Student's Achievement

1st Eyman Alyahyan
College of Computer Science and Information Technology
Imam Abdulrahman Bin Faisal University
P.O. 1982, Dammam, Saudi Arabia
eaalyahyan@iau.edu.sa

2nd Dilek Düştegör
College of Computer Science and Information Technology
Imam Abdulrahman Bin Faisal University
P.O. 1982, Dammam, Saudi Arabia
ddustegor@iau.edu.sa

***Abstract*—The prediction of students' academic achievement is crucial to be conducted in a university for early detection of students at risk. This paper aims to present data mining models using classification methods based on Decision Trees (DT) algorithms to predict students' academic achievement after preparatory year, and to identify the algorithm that yields best performance. The students' academic achievement is defined as High, Average, or Below Average based on graduation CGPA. Three classifiers (J48, Random Tree and REPTree) are applied on a newly created dataset consisting of 339 students and 15 features, at the College of Computer Science and Information Technology (CCSIT). The outcome showed the J48 algorithm had an overall superior performance compared to other algorithms. Feature selection algorithms were used to reduce the feature vectors from 15 to 4 resulting in improvements in performance and computational efficiency. Finally, the results obtained help to pinpoint the preparatory year courses that impact graduation CGPA. Timely warnings, and preemptive counseling towards improving academic achievement is possible now.**

Keywords—Prediction, Academic Achievement, Decision Tree (DT), J48, Random Tree, REPTree.

I. INTRODUCTION

Employment of graduates students is a problem in Saudi Arabia; only 48% of them are employed [1] . This rate is even larger for technical fields like CS and IT [1]. While this is certainly a multi-dimensional problem, feedback from stakeholders (e.g. Aramco employees that are members of the College Board), and potential employers (workplaces where our students perform their COOP) acknowledged that unemployment is more spread among students with low CGPA.

This study uses Data Mining (DM) tools to analyze available data from past batches of students at the College of Computer Science & IT (authors' institution) and extract useful information to explain the phenomena of low CGPA. To be more precise, in this paper, we aim to use classification methods based on Decision Trees (DT) algorithms to develop a model to predict students' academic achievement. To enable an early intervention, the prediction needs to be done as early as possible. In this study, we chose to predict after the preparatory year, that is to say at least four years before graduation. Several DT methods are used to identify the best model, per obtained accuracy. Significant factors that influence student achievement are also identified. Therefore, this work investigates the possibility of adopting the results of decision tree algorithms in support of academic decisions on improving academic performance, thus, to provide an appropriate road map for students and lecturers.

Although many similar studies have been published, either on the prediction of academic achievement of students in higher education [2]–[5] or on the identification of the most important factors affecting student achievement [6]–[8], they are mainly limited to predicting achievement in a course, or a specific exam. This study aims to predict graduation level achievement right after preparatory year, thus four years ahead. To the best of our knowledge there is no study investigating academic success prediction possibilities with such a large prediction window. The results of this study are particularly very useful for the Saudi universities, to start addressing from an academic success perspective the unemployment issue among young Saudi graduates.

II. LITERATURE REVIEW

A. *Prediction of Students' Academic Achievement at Degree Level*

R. Asif et al [9]report that it is possible to predict the academic achievement, at graduation time, of undergraduate students in a four-year academic program. They used the student's pre-university factor as well as transcripts information from the first and second year of the degree program, with satisfactory accuracy. In their study, they used the graduation CPGPA target parameter, and grouped it into five different categories: A (90-100%), B (80-89%), C (70-79%), D (60-69%) and E (50-59%). They used Naive Bayes, K-Nearest Neighbor classifier, Neural Networks, Rule Induction, Decision Tree, Random Forest and X-means algorithms on a sample of 210 undergraduate students of the department of Information Technology at the University of Engineering in Pakistan. The results showed that the Naive Bayes rated the best performance by 83.65%.

In another study from A. Adekitan and O. Salau, conducted at Covenant University in Nigeria [10], the authors applied DM methods to predict the graduation CGPA for engineering students. Therefore, six different DM algorithms were applied. For the classification technique, the highest accuracy of 89.15% was obtained by logistic regression. The results indicated that the CGPA of students in the first three years of study could reasonably determine the final CGPA for the fifth year of engineering students.

In another similar study, A. R. Olusola et al. selected 101 undergraduate students from the Architecture department at Olabisi Onabanjo University, Nigeria [7]. They used DM methods to predict students' academic performance depending only on the pre-admission test and their high school background. The dependent variable CGPA was classed into binary class ("Pass"/"Fail"); such that students with CGPA more than 2.4 (on a scale of 5) were categorized as "Pass," otherwise they were categorized as "Fail." The outcomes of this study confirmed that the previous academic

978-1-7281-5467-1/20/$31.00 ©2020 IEEE

achievement was a good indicator of the academic performance of architecture students. Besides, the Support Vector Machine classifier made a better prediction than the Logistic Regression. Furthermore, they identified that, courses like Math, BIO, and PHY had the most significant effect on the students' academic performance.

All these past studies clearly show that a prediction window of four years is very rare, and bulk of studies in this field investigate shorter prediction window (e.g. one or two years maximum).

B. Prediction of Students' Academic Achievement in Saudi Universities

On the course level, a study proposed a model to predict the academic performance of the students based on their forum participation and academic record [3]. In this study, A. Mueen et al. applied three different classification techniques namely, Naive Bayes, C4.5 Decision tree and Multilayer Perception neural network MLP on the students' data from King Abdulaziz University. They collected learning activities of the students from the Learning Management System LMS of undergraduate students who studied the Advanced Operating System and Programming Fundamental courses which were conducted during two semesters. The Results proved that the Naive Bayes classifier beat the other two classifiers by obtaining the overall prediction accuracy of 86%.

In a similar study, H. Almarabeh study's aimed at assessing the performance of university students through the applied DM classification techniques [4]. He used five different classification techniques: ID3, Naive Bayes, Bayesian Network, Multilayer Perception neural network MLP and J48 Decision tree on 225 students from Health Sciences College of Science at king Saud Bin Abdulaziz University. The results showed that the Bayesian Network algorithm outperformed the other algorithms with 92% accuracy.

M. A. Al-Barrak and M. Al-Razgan presented a case study in the mining of educational data to predict the final CGPA of the students at an early stage, using prior academic achievement data (from mandatory university courses only), and the pre-university data [8]. They applied the J48 Decision Tree algorithm on 236 undergraduate female students who were chosen from the Faculty of Computer Science at King Saud University. The students' final GPA value was selected as a predictor parameter and categorized into five different classes: excellent, very good, good, average and poor. In this study, in order to predict students' final GPA, classification rules were discovered based on students' grades in mandatory courses. And the most important courses in the study plan, those that had a big impact on the students' final GPA, were evaluated.

Very few similar studies have been conducted for Saudi Universities, most of them looking at shorter prediction window, or using only few DM algorithms.

III. Decision Trees

DT is a modeling technique of DM, that is used to predict and classify specific data objects based on a pre-created model using a training dataset with the similarly attributes [11]. A top-down tree is created based on the features of a particular data set.

These trees include the middle nodes and leaf nodes that are considered decisions or actions, and the edges emerging from the middle nodes are considered as conditions. DT work by starting from the roots, then crossing into real conditions until the leaf comes, then the decision is made. DT represent decision rules; the condition that gives a procedure is reached by a series of conditions starting from the root of the decision tree and ending in the paper, which refer to the procedure [12]. The division of the attribute area into branches and sheets is the key point in all DT algorithms the satisfactory stopping condition organizes and classifies the data [13]. Figure 1 illustrates a decision tree that can be used to select a database.

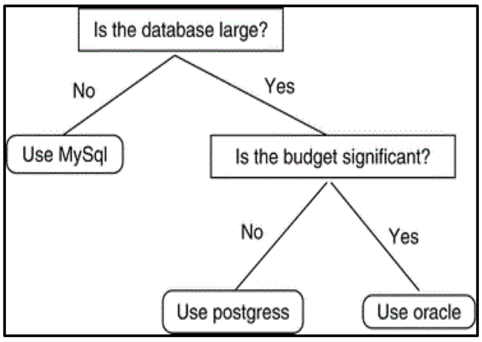

Fig. 1. Example of a decision tree for selecting a database [12]

The DT classification technique is performed in two phases [14]: construction trees and pruning. The tree construction phase follows the method from top to bottom. During this phase, the tree is frequently broken until the data items belong to the same class label. This phase is very daunting and consumes a lot of calculations where the training dataset is reprocessed frequently. The tree pruning phase is carried out in a bottom-up manner. Pruning a very large tree required to produce high-quality trees and improve the accuracy of the algorithm results, removes part of the bottom of the tree that has learned noise from the data to fix the circularization generalization of the classifier (overfitting) [15].

In a DM environment, we will find that the decision tree classifiers are particularly attractive because of several reasons: first of all, the resulting model is easy to understand and interpret by humans [15]. Secondly, they are especially proper for exploratory knowledge discovery since the decision tree classifiers are non-parametric. Thirdly, comparing with the other methods, the decision tree classifiers can be built faster. Finally, other classification models can't reach the accuracy of the decision tree classifiers [16].

In this research, three DT algorithms were used on CCSIT data, namely, J48, RepTree, and Random Tree. Each one are briefly explained in the following.

A. J48

The J48 algorithm is an application of the C4.5 algorithm [17] which is known in the Weka DM tool as "J48" which is an open-source Java implementation of the C4.5. It is the common useful decision tree technique for classification problems. This technique adopts a binary tree to model the classification process. After the tree is constructed, the algorithm is implemented in each group in the database and results in the classification of that group [18]. Given that this classifier is one of the commonly used tools in Weka which provides stability between accuracy, speed, and interpretability of results. Further, this algorithm classifies data into a decision tree in which we can easily identify students at risk [19].

B. *RepTree*

The Reduced Error Pruning (REPTree) is a decision tree learning algorithm. It considered a fast classifier based on the policy of calculating the acquisition of information with entropy and reducing the error caused by variance [20]. REPTree creates many trees and implements the regression tree logic to change duplicates. Next, the algorithm determines the best of all spawned trees. Depend on variance and result information, the classifier creates a regression decision tree. Moreover, this classifier pruning trees based on the use of the rear stabilization method and low error pruning. As in C4.5, this classifier can also work with missing or incomplete values by dividing identical states into parts [21].

C. *Random Tree*

The Random Tree (RT) algorithm defines a test based on a certain number of random features in each node without pruning. Usually, random trees refer to those created randomly and have nothing to do with machine learning. The advantage of building a random tree is efficient training and minimum memory requirements. To build a random decision tree, the RT algorithm uses only one pass on the data [20] [21].

IV. EMPIRICAL STUDIES

A. *Description of the Dataset*

For this study, data was extracted from the Students Information System (SIS) of Imam Abdulrahman bin Faisal University, where various students' information are collected, like student demographics, information related to their studies including their transcripts. The dataset was limited to students who graduated from the CCSIT, including both CS and CIS departments. Hence, students from the first batch of Fall 2009 up to the batch of Spring 2013 were included. Initially, 41,585 records were provided, which corresponded to 612 undergraduate students obtained from 6 batches, from 2009 batch up to the 2014 batch that graduated in Spring 2018. Concerning these raw data, an initial data preparation was performed consisting of selection and cleaning steps to remove redundant or irrelevant attributes/ instances. As well as removing noise and missing values. Accordingly, the remaining data was associated with 339 students with 15 features. The approach of data preparation is summarized in figure 2.

We used pre-university attributes i.e. (High School background and admission tests marks), and the students' grades in preparatory year, and demographics (see Table 1). The students' academic achievement was addressed as multi-class classification problem where the class (target) was classified based on the CGPA at graduation time. The first class called *High* was defined for a CGPA more than or equal to 4.50. The second class called *Average* included CGPA between 3.75 and 4.50 (not inclusive), and finally the third class called *Under Average* for CGPAs less than 3.75 on a 5-point scale.

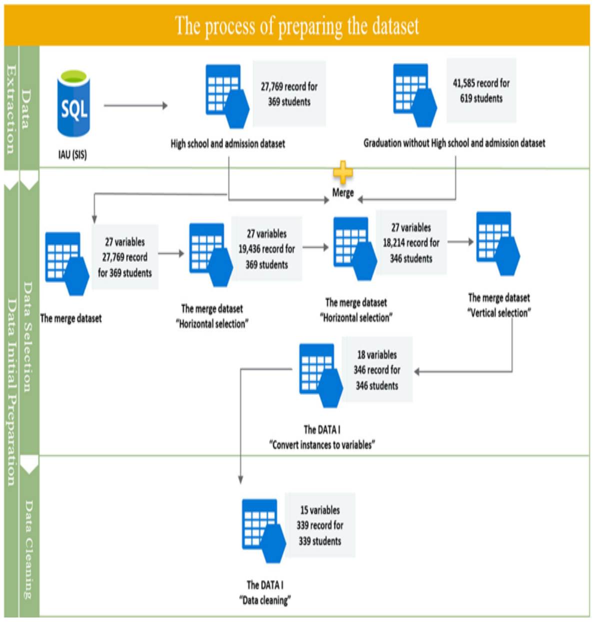

Fig. 2. Preparing dataset process

TABLE I. ATTRIBUTES DESCRIPTION

Variable Category	Name	Description	Initial value	Transformation
Demographic	Gender	Student sex	F,M	F = 0, M = 1
	Country	Student nationality	SAU, BHR, EGY, IND, IRQ, ….	SAU=0, NOT=1
Admission tests marks + High School background	SAT1	Standard Achievement Admission Test	98 - 60	-
	GAT	General Aptitude Test	96-59	-
	SGPA	Students grade point average for high school (100-point scale).	100-80	-
University background	MATH111	Mathematics 1	A+ / A / B+ / B / C+ / C / D+ / D / FP / F	-
	COMP131	Computer Skills		
	LRSK141	Learning and Searching skills		
	PHEDU152	Health and Physical Education		
	ENG101	General English Language		
	MATH112	Mathematics 1		
	CMSK142	Communication Skills		
	COMP122	Computer Applications		
	CGPA _1st year	CGPA for Prep year	5-2.6	-
Class (Output variable)	Student_achievement	CGPA for 5th year	High/Average/Under Average	

FP stands for FAILED then PASSED, F stands for FAILED

B. Experimental Setup

The experiment was performed using WEKA (Waikato Environment for Knowledge Analysis), a program that includes a set of machine learning algorithms that can be applied to datasets. WEKA also has options for pre-processing data, specifying features, and applying techniques to split the dataset for training and testing [22].

The dataset is first prepared and pre-processed for the experiment. So, the missing values of the nominal variables in the datasets were replaced with the mode of the attribute. This step was carried over WEKA using an unsupervised attribute filter (replace missing values). Then we discretized all numeric variables by binning the data [23]. It was a necessary step to increasing the accuracy of the models by overcoming the problem of outliers and easier to understand features. This step was performed via WEKA using an unsupervised attribute filter (Discretize) using 10 bins.

Next, the optimized parameters for J48, Random Tree, and REPTree were specified. This step was performed by adjusting the confidence factor J48, while the seed and minimum total weight for RT, as well as the seed changing parameter values, it was found that J48 performed better with an accuracy of 69.02%, while REPTree had an accuracy of 67.5%. RT also performs well with 61.9%.

Besides, the correlation coefficients and information gain between each feature and the target variable were calculated to rank the attributes for the feature selection. The results can be seen in Table IV and Table V.

Subsequently, the ability to improve the accuracy of classification performance was investigated by selecting features and identifying important features to predict students' academic achievement. Two methods of feature selection were investigated, feature selection based on correlation and the InfoGain feature selection method. The models were tested on the selected attributes using 10-fold cross validation and optimal parameters for each classifier. The results can be seen in Table VI and Table VII.

Following the results, by using the correlation-based feature selection method and reducing the features from 15 to 4, we implemented the classifiers using different partition ratios. It was found that the most powerful accuracy of REPTree and RT was achieved with 50:50 ratio (50% for training data and 50% for testing data) while J48 achieved their best performance when 70:30 ratio (50% for training data and 50% for testing data). The results can be seen in Table VIII.

Finally, From the above experimental results, we choose the optimal options of the best subset features that achieve the best result of cross validation or partition ratio to develop the final model.

V. OPTIMIZATION STRATEGY

In the quest of optimizing classification results, CVParameterSelection search methodologies for parameters in Weka were applied [24], in order to enhance the accuracy and to obtain the best performance criteria based on the accuracy. All classifiers parameter tuning used 10-fold cross-validation with the partitioning of the training datasets. Table II shows the default parameters and the chosen optimal parameters for each classifier.

TABLE II. DEFAULT AND OPTIMAL PARAMETERS FOR EACH CLASSIFIER

Model	Parameters Values		
	Parameters	***Default value***	***Optimal value***
J48	Confidence Factor	0.25	0.326
	Minimum number of object	2	14
REPTree	Seed	1	25
RT	minNum	1	5
	Seed	1	4

As shown in Table III, comparing the new result of the optimal parameters with the result of the default parameters shows a significant increase in accuracy.

TABLE III. DEFAULT AND OPTIMAL PARAMETERS FOR EACH CLASSIFIER

Model	Performance (accuracy)	
	Default value	***Optimal value***
J48	67.2%	69.02%
REPTree	66.9	67.5%
RT	57.2	61.94

IV. RESULTS AND DISCUSSION

A. Effect of Feature Selection on the Dataset

Feature selection based on correlation and the InfoGain feature selection method were implemented with recursive feature elimination to find the subset achieving the highest performance and the most significant features to predict students' academic achievement.
The correlation coefficient was used to rank features based on Pearson's values from top to lowest relationship with the class variable (Student_achievement) as shown in Table IV. Likewise, the information gain was used to rank features based on measuring the information gain with respect to the class as shown in Table V.
Then, the recursive feature elimination procedure was applied to produce subsets by recursively eliminating the bottom half of the ranked features until a single feature remained.

TABLE IV. THE CORRELATION OF EACH FEATURE AND CLASS VARIABLE

Order	Attributes name	Correlation with class
1	CGPA _1st year	0.271
2	COMP131	0.2478
3	CMSK142	0.2294
4	MATH112	0.2287
5	COMP122	0.2257
6	SGPA	0.2009
7	ENG101	0.1973
8	SEX	0.1971
9	LRSK141	0.1914
10	MATH111	0.1835
11	COUNTRY	0.1407
12	PHEDU152	0.1358
13	SAT1	0.1074
14	GAT	0.0807

TABLE V. THE INFORMATION GAIN OF EACH FEATURE AND CLASS VARIABLE

Order	Attributes name	Information gain with class
1	CGPA _1st year	0.4993
2	MATH111	0.2839
3	ENG101	0.2693
4	MATH112	0.2379
5	COMP131	0.1948
6	GAT	0.1818
7	SAT1	0.1784
8	COMP122	0.1681
9	SGPA	0.1537
10	PHEDU152	0.137
11	LRSK141	0.1267
12	CMSK142	0.1176
13	SEX	0.0734
14	COUNTRY	0.0398

TABLE VI. CORRELATION-BASED FEATURE SELECTION RESULTS

Number of features	Features	*J48*	*REPTree*	*RT*	*AVG*
14 features	All	69.02%	67.5%	61.9%	66.1%
7 features	CGPA _1st year COMP131 CMSK142 MATH112 COMP122 SGPA ENG101	68.4%	67.8%	60.1%	65.4%
4 features	CGPA _1st year COMP131 CMSK142 MATH112	69.3%	68.1%	65.4%	67.6%
2 features	CGPA _1st year COMP131	64.01%	62.8%	65.4%	64.0%

TABLE VII. INFOGAIN METHOD FEATURE SELECTION RESULTS

Number of features	Features	*J48*	*REPTree*	*RT*	*AVG*
14 features	All	69.02%	67.5%	61.9%	66.1%
7 features	CGPA _1st year MATH111 ENG101 MATH112 COMP131 GAT SAT1	69.6%	67.8%	59.5%	65.6%
4 features	CGPA _1st year MATH111 ENG101 MATH112	69.3%	67.5%	59.5%	65.4%
2 features	CGPA _1st year MATH111	66.9%	66.3%	64.3%	65.8%

After investigating the two feature selection methods as shown in Table VI and Table VII, it is evident that the best results were obtained by applying the correlation-based feature selection using the top 4 features: *CGPA _1st year, COMP131, CMSK142, MATH112,* that is to say, the CGPA student obtained in preparatory year, the letter grade obtained from COMP131 which is an introductory course to computer science, as well as the letter grade obtained from MATH 112 which strengthens basic math skills, which are broadly used in calculus. Also, CMSK14 provides the students an overview of the current concepts and theories in the area of communication.

B. Trying different partition ratio

After getting the best results by using the correlation-based feature selection method and reducing the features from 15 to 4, the performance of each classifier was examined by conducting several tests on the dataset using different partition ratios. The amount of training data varied from 50 to 80. Table VIII shows the results of different direct partition ratios applied to each classifier.

TABLE VIII. DEFAULT AND OPTIMAL PARAMETERS FOR EACH CLASSIFIER

Partition ratio	Performance		
	J48	*REPTree*	*RT*
50:50	63.9%	65.08%	60.3%
60:40	65.4%	62.5%	56.6%
70:30	66.6%	63.7%	58.8%
80:20	61.7%	61.7%	52.9%

C. The results of 10-fold validation comparison with Direct Partition techniques

On analyzing and comparing the two testing methods (10-Fold cross-validation and direct partition ratio), all classifiers - as can be noted in Table IX- obtained a better value in using 10-Fold cross-validation techniques.

TABLE IX. COMPARISON BETWEEN 10-FOLD CROSS VALIDATION AND DIRECT PARTITION THECHNIQUES

Techniques	Proposed model		
	J48	*REPTree*	*RT*
10-fold validation	69.3%	68.1%	65.4%
Partition ratio	66.6%	65.08%	60.3%

IV. FURTHER DISCUSSION

The final model for predicting students' achievement was developed using the best subset feature with the optimal parameters obtained as shown in Table X. The ideal results of each J48, REPTree and RT were obtained when using a 10-fold cross-validation technique. Among all classifiers, the J48 performed better than REPTree and RT for predicting students' achievement with reasonable accuracy of 69.3%.

As shown in Figure 3, Reducing features had a positive impact on accuracy. The classification performance was improved by selecting the 4 most predictive features, out of 15. From this process, we were able to determine what are the important features that had a significant impact on predicting academic achievement of CCSIT students; CGPA for Prep year, Computer Skills course, Communication Skills course, Mathematics course. By focusing on these three courses that all CCSIT students take during their preparatory year, one can identify early enough students potentially at risk of graduating with a low CGPA. Early intervention can eventually change the outcome.

The receiver operating characteristic (ROC) curve is an accuracy of a test to discriminate between classes. The nearer the ROC curve is to the upper left side, the greater the overall accuracy of the test and the separation capacity between classes. The area found under the curve for each classifier as shown in Figure 4, Figure 5, Figure 6, indicates that the J48 is the most appropriate classifier than the other classifiers.

TABLE X. RESULTS OF USING THE BEST OPTIONS BASED ON THE OPTIMAL PARAMETERS OBTAINED

Techniques	**Proposed model**		
	J48	***REPTree***	***RT***
10-fold validation	69.3%	68.1%	65.4%

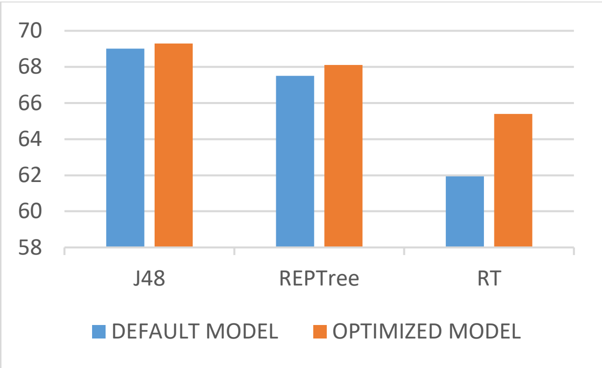

Fig. 3. Comparison between optimized models and default model

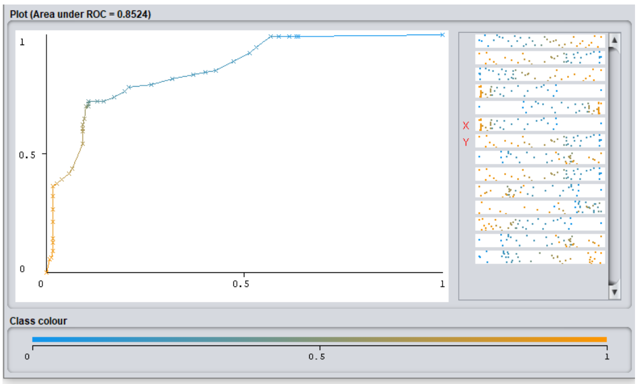

Fig. 4. J48 ROC curve for Under Average class

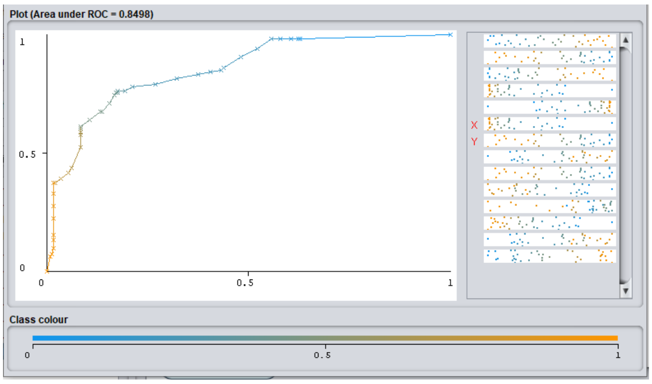

Fig. 5. REPTree ROC curve for Under Average class

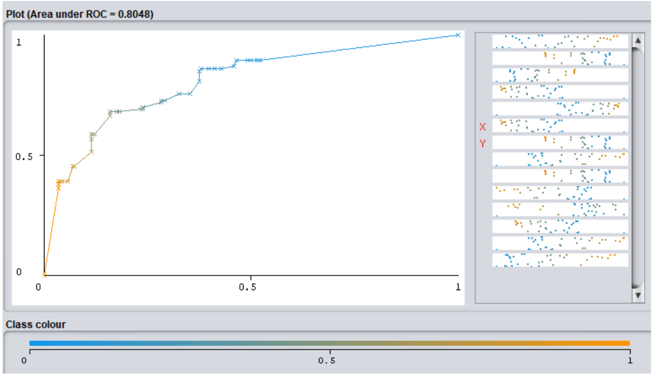

Fig. 6. RT ROC curve for Under Average class

VI. CONCLUSION

This research focused on the predictive ability of DM methods to predict students' achievement after preparatory year at the degree level in higher education. The students' achievement is based on the Grade Point Average (CGPA) defined as (high, average, or under average).

Throughout the experiment, we have implemented three decision tree classifiers; J48, REPTree and RT on the student dataset to predict the achievement of the student at graduation year. The results showed that the J48 classifier outperforms REPTree and RT for predicting students' achievement with reasonable accuracy of 69.3%. Moreover, the important features that had a significant impact on predicting academic achievement of CCSIT students were; CGPA for Prep year, Computer Skills course, Communication Skills course, Mathematics course. The results obtained will help to predict students' final achievement early enough to take effective countermeasures by providing timely warnings to students. Thus, the percentage of students who have low achievement can be reduced providing the right counseling.

REFERENCES

[1] "University disciplines .. and the structure of the labor market _ Economic newspaper," 2019. [Online]. Available: http://www.aleqt.com/2018/07/16/article_1420651.html. [Accessed: 30-Oct-2019].

[2] N. A. Yassein, R. G. M Helali, and S. B. Mohomad, "Predicting Student Academic Performance in KSA using Data Mining Techniques," *J. Inf. Technol. Softw. Eng.*, vol. 07, no. 05, 2017.

[3] A. Mueen, B. Zafar, and U. Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques," *Int. J. Mod. Educ. Comput. Sci.*, vol. 8, no. 11, pp. 36–42, 2016.

[4] H. Almarabeh, "Analysis of Students' Performance by Using Different Data Mining Classifiers," *Int. J. Mod. Educ. Comput. Sci.*, vol. 9, no. 8, pp. 9–15, 2017.

[5] M. Al-Saleem, N. Al-Kathiry, S. Al-Osimi, and G. Badr, "Mining Educational Data to Predict Students' Academic Performance," Springer, Cham, 2015, pp. 403–414.

[6] R. Asif, A. Merceron, and M. K. Pathan, "Predicting Student Academic Performance at Degree Level: A Case Study," *Int. J. Intell. Syst. Appl.*, vol. 7, no. 1, pp. 49–61, 2015.

[7] R. O. Aluko, E. I. Daniel, O. Shamsideen Oshodi, C. O. Aigbavboa, and A. O. Abisuga, "Towards reliable prediction of academic performance of architecture students using data mining techniques," *J. Eng. Des. Technol.*, vol. 16, no. 3, pp. 385–397, Jun. 2018.

[8] M. A. Al-barrak and M. Al-razgan, "Predicting Students Final GPA Using Decision Trees : A Case Study," *Int. J. Inf. Educ. Technol.*, vol. 6, no. 7, 2016.

[9] R. Asif, A. Merceron, S. Abbas, and N. Ghani, "Analyzing undergraduate students ' performance using educational data

mining," *Comput. Educ.*, vol. 113, pp. 177–194, 2017.

[10] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," *Heliyon*, vol. 5, no. 2, p. e01250, 2019.

[11] J. Mesarić and D. Šebalj, "Decision trees for predicting the academic success of students," *Croat. Oper. Res. Rev.*, vol. 7, no. 2, pp. 367–388, Dec. 2016.

[12] A. Dobra, "Decision Trees," in *Encyclopedia of Database Systems*, New York, NY: Springer New York, 2018, pp. 1022–1023.

[13] M. Brijain, M. Brijain, R. Patel, M. Kushik, and K. Rana, "A Survey on Decision Tree Algorithm For Classification."

[14] Matthew N. Anyanwu and Sajjan G. Shiva, "Comparative Analysis of Serial Decision Tree Classification Algorithms," *Int. J. Comput. Sci. Secur.* , no. 1985–1553, 2009.

[15] A. Dobra, "Decision Tree Classification," in *Encyclopedia of Database Systems*, New York, NY: Springer New York, 2016, pp. 1–6.

[16] T. Lim, T. Lim, W. Loh, and Y.-S. Shih, "An Empirical Comparison of Decision Trees and Other Classification Methods," 1998.

[17] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[18] G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright, "A practical differentially private random decision tree classifier," *Trans. Data Priv.*, vol. 5, no. 1, pp. 273–295, Apr. 2012.

[19] L. Devasena, "Comparative Analysis of Random Forest, REP Tree and J48 Classifiers for Credit Risk Prediction," *Int. J. Comput. Appl.*, no. 0975 – 8887, pp. 30–36, 2014.

[20] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Elsevier Inc, 2016.

[21] S. K. Jayanthi and S. Sasikala, "REPTREE CLASSIFIER FOR IDENTIFYING LINK SPAM IN WEB SEARCH ENGINES," *IJSC*, vol. 3, no. 2, pp. 498–505, 2013.

[22] "Weka 3 - Data Mining with Open Source Machine Learning Software in Java." [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/index.html. [Accessed: 09-Sep-2019].

[23] S. Maheshwari, R. C. Jain, and R. S. Jadon, "A Review on Class Imbalance Problem: Analysis and Potential Solutions."

[24] R. Kohavi, "Wrappers for performance enhancement and obvious decision graphs," 1995.