On the application of data mining algorithms for predicting student performance: a case study

Abdullah Alsheddy[†] **and** **Mohamed Habib** [††]

College of Computer and Information Sciences,
Al Imam Mohammad Ibn Saud Islamic
University, Riyadh, Saudi Arabia

College of Computing and Informatics,
Saudi Electronic University, Riyadh, Saudi Arabia
Faculty of Engineering, PortSaid University, Egypt

Summary
Data mining methods are often implemented at various sectors today for analyzing available data and extracting information and knowledge to support decision-making. Educational data mining is an emerging discipline, concerned with developing methods and algorithms that discover knowledge from data originating from educational environments. This paper study the application of data mining methods, namely classification and clustering, on undergraduate Information Technology student's data in a blended-learning university. A classifier that implements a decision tree algorithm is used to predict student performance at the end of the program based on students' socio-demographic variables and achieved results from the first (i.e. preparatory) year of the study program. The classifier is also used to derive critical courses that can serve as indicators for students' performance. A clustering algorithm, on the other hand, is applied for identifying the attributes of withdrawn students from this degree program. The results reveal the high potential of data mining applications in providing insights for students, academic advisors and university management, which can help in improving the quality of the educational processes and enhancing learning experience.

Key words:
Knowledge Discovery, Educational Data Mining, Decision trees, Clustering.

1. Introduction

Besides having the role of producing and disseminating information, universities are fundamentally institutions that educate students. Due to the advancement in technologies, universities record vast amount of data about students and their performance. These data contain valuable information about students, and usually are only used individually and for official uses (transcript, registration, etc.). However, these data could be used in academic consulting of students. In order to do so, they need to be analyzed, which requires a separate discovery investigation [1].

In order to analyze large data and obtain meaningful and operational information, a class of techniques referred to as knowledge discovery in databases (KDD) [2] have been widely used to acquire as much information out of the data. Data mining is a phase within the KDD process where a set of techniques and tools are applied to test and extract valuable hidden information from the data [3, 4]. Data mining is the automatic extraction of implicit and interesting patterns from large data collections [5]. Pattern here implies potentially valuable trends and relationships in data. It is usually applied in many applications, such as marketing, medical diagnosis, fraud detection and scientific discovery [6, 7].

The application of data mining techniques to educational data is known as Educational Data Mining (EDM) [8]. EDM is an emerging area that focuses on building up techniques for exploring data and discovering significant patterns from educational data. It attempts to empathize educational data from a new point of view rather than what data were originally intended to study [9]. EDM embraces a range of data mining techniques, to support relationship analysis, classification, clustering, elaborate educational hypotheses, and provide learning support [10].

The knowledge gained from EDM, is crucial to administration and decision makers in order to assist in inferring the overall performance of students in the system and help in improving the student's learning outcomes and the courses activities. Additionally, this gained knowledge may be valuable not only to the administration but also to students, as it can be adjusted towards different ends for different sharers in the process [11]. It could be oriented towards students in order to recommend scholars' activities, resources, advise path pruning or simply connections that would favor and improve their learning or to administration in order to get more objective feedback [12].

In this study, we aim to analyze the performance of students pursuing a 4-year Bachelor degree program in the discipline of Information Technology (IT). The first year of this program is a Preparatory Year (PY), and passing this year is a prerequisite to register for the remaining years of the program. A key feature of this program is the adoption of a blended learning model that is a hybrid of conventional face-to-face and online learning so that educational activity occurs both in the classroom and online. The rationale is to discover different insights about

Manuscript received October 5, 2017
Manuscript revised October 20, 2017

student performance in this blended learning environment. Such insights are very valuable to students, college and decision makers for improving learning experience and institutional effectiveness. In particular, this study examines three research questions:

- Question 1: Using classification techniques, a question will be whether we can predict the performance of students (i.e. students' Final GPA) at an early stage of the program using students' GPA by the end of the PY and some demographic/personal attributes. Such a classifier should enable the university administration to develop easy-to-implement educational policies.
- Question 2: Another question will be whether we can use classification techniques to identify the critical courses that affect the students' performance in the considered degree program. The performance of students in these courses can then serve as effective indicators of the students' final GPA.
- Question 3: The last question will be whether we can use clustering techniques to identify the attributes of withdrawn students from the considered program. The importance of this question is due to the unusually high student withdrawal rate from the considered BSc in IT program. The clustering approach should help in increasing student retention rate.

The rest of this paper is organized as follows. It begins with a review of related work in section 2. Then, an overview of the data mining techniques used in this study is described in section 3. The adopted solution framework is introduced in section 4. The results are presented and discussed in section 5. Finally, the paper concludes in section 6.

2. Related Work

There have been many studies in the past few decades aimed at applying data mining techniques to educational data obtained from various learning environments, including traditional classroom, online and blended learning. This section provides an overview of some recent research on EDM that focus on predicting student performance and retention. Recent surveys of EDM are presented in [13-15].

One of the major techniques in EDM that has been widely applied for prediction is classification. [16, 17] successfully used decision tree to classify students into three groups: the 'low-risk' students, who have a high probability of succeeding, the 'medium-risk' students, who may succeed according to the measures taken by the university, and the 'high-risk' students, who have a high probability of failing. In [12], they applied various classifiers, including decision tree, rule induction and fuzzy rule learning, toward figuring out what factors are predictive of student failure or non-retention in a course or degree program. While, in [18] authors used regression modeling to predict students' test scores by integrating timing information and the amount of assistance a student needs to solve problems. This research [19] predict students' success/failure and grade in a course by using social variables like age, sex, marital status, and nationality. This research examines the performance of several popular classification and regression algorithms. Decision trees and SVM obtained the best results in classification, while SVM, Random Forest, and AdaBoost.R2 obtained the best results in regression. In [20] authors predict students' performance in a course based on their performance in prerequisite courses and midterm examinations. They developed 24 predictive mathematical models and a SVM model. The SVM model yields the highest percentage of accurate prediction.

Another commonly applied method in EDU is clustering, which is used to group students with similar attributes. There have been studies that attempted to cluster students based on academic performance in examinations [21, 22]. In [23] they found typical behaviors in forums such as high-level workers, i.e. students that read all messages and post many messages in the forum, or lurkers, i.e. students who read all messages without posting any. They used a two-stage analysis strategy based on an agglomerative hierarchical clustering algorithm to identify different participation profiles adopted by learners in online discussion forums. The obtained final clusters actually group learners with similar activity patterns and allow to satisfactorily identifying different participation profiles in online discussion forums. In addition, [24] applied hierarchical cluster analysis (HCA) to identifying groups of students with similar performance from kindergarten levels until the end of high school. The results demonstrate that student grades are useful in identifying who may drop out. In [25], authors has clustered students' interaction data to build profiles of students, while applying Model-based clustering. The results confirmed the potential of clustering techniques to support evaluation of students' collaborative activities. In [26], X-means clustering with Euclidean distance is applied to investigating how students' performance progresses during their studies, and how this progress is related to the various courses. [27] Applies a clustering technique to group students based on their interaction with the learning management system represented by the time between various events such as mouse clicks and page loads; as well as the various assessment scores, including quizzes, session tests and course final grades.

3. Data Mining Techniques

There are a number of popular methods within EDM, which can be applied to uncover hidden knowledge from educational data. This section provides an overview of the two methods used in this work: classification and clustering.

3.1 Classification

Classification [32] is a particular case of prediction when the goal is to infer a target attribute (predicted variable) from some combination of other attributes (predictor variables), and the predicted variable is a categorical value. In EDM, prediction can be used for predicting student performance and detecting their behaviors. Some popular classification methods in educational domains include decision trees, random forest trees, neural networks, and Naïve Bayes. Decision trees are white-box classifiers that produce comprehensible and useful results, in contrast to black-box classifiers such as neural networks. This feature is very important for decision making in EDM, and therefore the decision trees technique is implemented in this work.

A decision tree [28, 29] is a representative classifier that attempts to learn a classification tree that decides the value of a dependent attribute given the values of the independent attributes. A decision tree contains internal nodes and leaf nodes. Each internal node contains a logical test on an attribute (i.e. split). Connecting branches from an internal node to its children represent an outcome of the tests. Each leaf node represents a class label. Given a training set that consists of objects described by a set of attributes and a class label, the decision tree is built from this set in a recursive manner. The algorithm starts with an empty tree and the entire training set. In each recursion step, a test is created by choosing the attribute that can best split the training set at the current internal node into multiple subsets with maximal distinctness. This step is applied to every child nodes until all training examples at that node have the same value for the target class, in which a leaf node is crated with that class label .

3.2 Clustering

Clustering [32] is about collecting and presenting similar data items. Typically, some kind of distance measure is used to decide how similar instances are. Having defined the similarity measure, a cluster is therefore a group of items that are similar to each other within the group and dissimilar to the objects in other clusters. Clustering is particularly useful when the task is to find structure in the data without an a priori idea of what should be found. In EDM, clustering can be used to group students based on their learning and actions, or to group similar course materials.

A representative clustering algorithms is the expectation maximization (EM) algorithm [30,31]that is a mixture-based algorithm that finds maximum likelihood estimates of parameters in probabilistic models. Principally, the EM clustering algorithm alternates between two steps, namely Expectation step (E-step) and Maximization step (M-step). The E-step computes the membership probability of each data point in each cluster, while the M-step computes parameters maximizing the expected log-likelihood found on the E-step.

4. The Proposed Approach

The framework proposed in this work is shown in figure 1. It starts with a preprocessing phase, and then applying different mining algorithms over the extracted data. The last phase is the interpretation and visualization for the results.

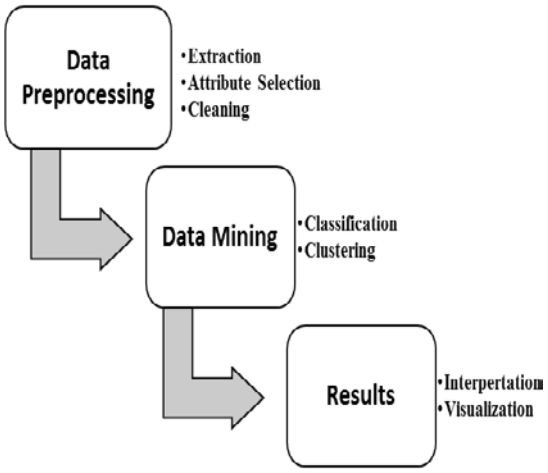


Fig. 1 Proposed Framework

4.1 Preprocessing phase

The data used in this study is extracted from the registration system (i.e. Banner) used by a public sector university in Saudi Arabia, which is employed to handle all students' data. The extracted dataset represents the data of all students (around 1980 student) enrolled in the BSc in Information Technology program since the academic year 2013/2014 until 2015/2016 (i.e. three academic years). The dataset comprises 42300 records, where each record

includes more than 20 attributes for each student and each course the student has registered in.

The dataset has to go through a preprocessing phase to clean the data from different errors and remove non-important and redundant attributes. The required data for evaluating students' performance are scattered over many tables; thus, this phase starts with attribute selection and merging multiple tables into a single table that contains the most important attributes for the required evaluation. This results in one table with 24 attributes. These attributes can be categorized into two main sets:

1. Students' personal information, including age, gender, and nationality.
2. Student academic performance, including course grade, semester's GPA and cumulative GPA.

All the irrelevant attributes are removed in the first step before merging the tables, as they do not offer any knowledge for the data processing. Although the dataset has some missing values, the total number of records with missing values for important attributes like age, gender, cumulative GPA, are only 300 records. Therefore, these records are removed from the dataset, since these values are very important for the applied mining processes.

The last step in this phase is the transformation of the dataset into a suitable format for the mining algorithms. The technique used in the transformation step depends on the applied mining algorithm on the dataset. Since we intend to apply various algorithms in this study, we leave the details of the transformation step for each mining algorithm to be discussed later.

4.2 Data Mining phase

In this phase, different data mining algorithm are applied on the dataset using the WEKA software [33]. WKEA is considered as a fast-growing, open source software that is fully implemented in Java. It provides a collection of machine learning and data mining algorithms for data pre-processing, classification, regression, clustering, association rules, and visualization. Among these algorithms, we apply two data mining techniques, namely classification and clustering for various purposes.

The first research question in this study aims to predict the performance of a student by the end of the program as early as possible, using a classification technique. This is achieved by using Decision Trees to generate a decision tree used for predicting students' cumulative GPA based on their obtained GPA in the first (i.e. preparatory) year. This classification technique is also employed to address the second research question in this study that aims to identify critical courses that influence students' performance in the considered academic program. The classifier algorithm is used to generate a decision tree that relates students' cumulative GPA to their grades obtained in different courses. To generate these decision trees, the J48 classifier for C4.5 algorithm [34] is applied.

On the other hand, the third research question of this study aims to identify the demographic/personal attributes of withdrawn students from the considered Information Technology program. In this regard, the clustering technique is used to identify the main attributes of withdrawn students, by applying the EM-clustering algorithm.

5. Results and Discussions

5.1 Early prediction of students' performance using classifiers

The aim of this trial is to find a relation between students' cumulative GPA at the end of the degree program and their GPA obtained at the end of the first year that serves as a preparatory year. Thus, the J48 classifier has been applied on the provided dataset to produce the required decision tree. The extracted dataset undergoes a second phase of preprocessing to convert the data into suitable format for the decision tree algorithm. Table 1 describes the attributes of the dataset used in this trial, which contains 1981 instances. Then a discretization process is applied to the dataset as given Table 2.

Table 1: Dataset Attributes

Attribute	**Description**
Age	Students age
Gender	Male / Female
Level	Student Level (1 – 8)
Nationality	Student Nationality
Preparatory Year's GPA (PYGPA)	Continuous Value for GPA
Cumulative GPA	Continuous Value for the current cumulative GPA

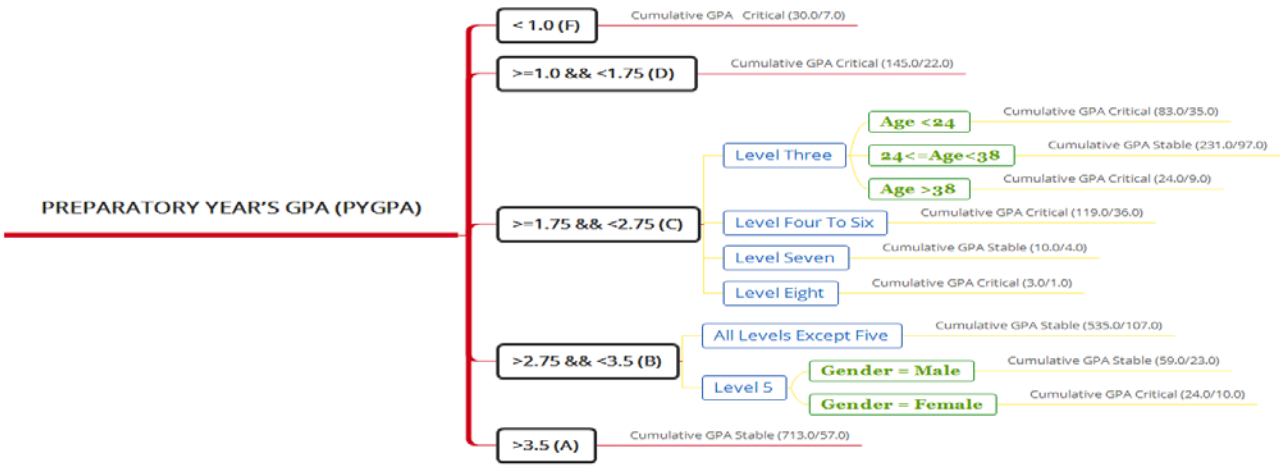


Fig. 2 A Decision Tree of Preparatory Year's GPA vs. Cumulative GPA

Table 2: Discretization Rules

Attribute	Discretization Criteria
Age	3 classes: Age < 24 years 24 years <= Age <= 38 years Age > 38 years
Gender	Male / Female
Level	Student Level: 1 to 8.
Nationality	Student Nationality
Preparatory Year's GPA (PYGPA)	5 classes: PYGPA < 1.0 (PYGPA = F) 1.0 <= PYGPA < 1.75 (PYGPA = D) 1.75 <= PYGPA < 2.75 (PYGPA = C) 2.75 <= PYGPA < 3.5 (PYGPA = B) 3.5<= PYGPA <4 (PYGPA = A)
Cumulative GPA (GPA)	2 Classes GPA < 2.0 (Critical status) GPA >= 2.0 (Stable status)

A decision tree is evaluated on test split of 60% for training and 40% for test. The percentage of Correctly Classified Instances is 78.8 %, while that of the Incorrectly Classified Instances is 21.2%. Figure 2 shows a decision tree, and Table 3 shows for each predicted class, the following performance measures:

- The True Positive (TP) rate: a measure of the proportion of positives that are correctly identified as such.
- The False Positive (FP) rate: a measure of the ratio between the number of negative events wrongly categorized as positive, and the total number of actual negative events.
- Precision: a measure of the fraction of relevant instances among the retrieved instances.
- Recall: a measure of the fraction of relevant instances that have been retrieved over the total amount of relevant instances).

The resulted decision tree reveals that 85.8% of students who obtained a GPA less than 1.75 at the end of the preparatory year (i.e. PYGPA is either F or D) are in critical status during the degree program as they find difficulty to raise their cumulative GPA to at least 2.0. On the other hand, 92.6% of students whose PYGPA is A, progress smoothly during the study of the degree program as they usually maintain an cumulative GPA that is greater than or equal 2.0.

Table 3: Detailed Accuracy by Class

Class	TP Rate	FP Rate	Precision	Recall
Cumulative GPA is Critical (<2.0)	0.447	0.071	0.721	0.447
Cumulative GPA is Stable (> 2.0)	0.929	0.553	0.803	0.929
Weighted Avg.	0.788	0.412	0.779	0.788

These results concluded from the decision tree are very helpful to decision makers and academic advisors to guide students in their tracks. It also helps in developing effective conditions for students who can continue to complete the BSc in IT degree program.

5.2 Identifying critical courses using classifiers

The second research question in this study is about deriving critical courses in the considered BSc in IT

program, where students' performance in these courses can be effective indicators for their final GPA in this academic program. In this experiment, a dataset of 2137 instances has been used, each of which comprises 44 attributes like student age, gender, cumulative GPA, and the grade obtained by every student in any course that student enrolled in. Discretization is applied on this dataset based on the rules shown in Table 2.

A decision tree algorithm is applied on the dataset that is divided into a training (66%) and testing (34%) subsets. Figure 3 shows part of the resulted decisions tree. The Correctly Classified Instances represent 74.4% of the total instances, while the Incorrectly Classified Instances are 25.6%. Table 4 shows the TP rate, the FP rate, Precision and Recall for the discovered decision tree.

Table 4: Detailed Accuracy by Class

Class	TP Rate	FP Rate	Precision	Recall
Cumulative GPA is Critical (<2.0)	0.301	0.068	0.65	0.301
Cumulative GPA is Stable (> 2.0)	0.932	0.699	0.759	0.932
Weighted Avg.	0.744	0.512	0.727	0.744

It is concluded from the decision tree that the Computer Programming I course, which is one of the level 3 courses, is the first critical course for students. Students' performance (i.e. their grade) in this course gives an indication to how they will perform in the next levels of this degree program. In addition, the decision tree suggests that the major courses that are critical to students' progress in this academic program are Computer Programming I, Computer Organization, Discrete Mathematics, Computer Programming II, Human Computer Interaction, Introduction to Database and Computer Networks. According to the study plan for the considered Information Technology degree program, these courses appear in level 3, 4 and 5.

These results are very helpful for the academic advisor to guide students during their study by supporting students who are at-risk, or to stimulate further the students showing promise.

5.3 Identifying the attributes of withdrawn students using clustering

The data obtained from the university in this study shows a high student withdrawal rate from the considered degree program. Thus, the third trial aims to identify the main demographic/personal attribute of withdrawn students. A dataset of 245 instances, which includes all withdrawn students from the considered degree program, has been used in this experiment. Each instance contains 6 attributes that are given in Table 5. Only the age attribute is discretized into three categories as shown in Table 2. Then, a clustering algorithm using the EM-clustering algorithm with 32 iterations performed on the dataset.

Table 5: Withdrawn Students Dataset Attributes

Attribute	Description
AGE	Students age
Gender	Male / Female
Level	Student Level (1 – 8)
Nationality	Student Nationality
Course Title	The title of course with F (failed) for this student
Cumulative GPA	Continuous Value for the final level GPA for student

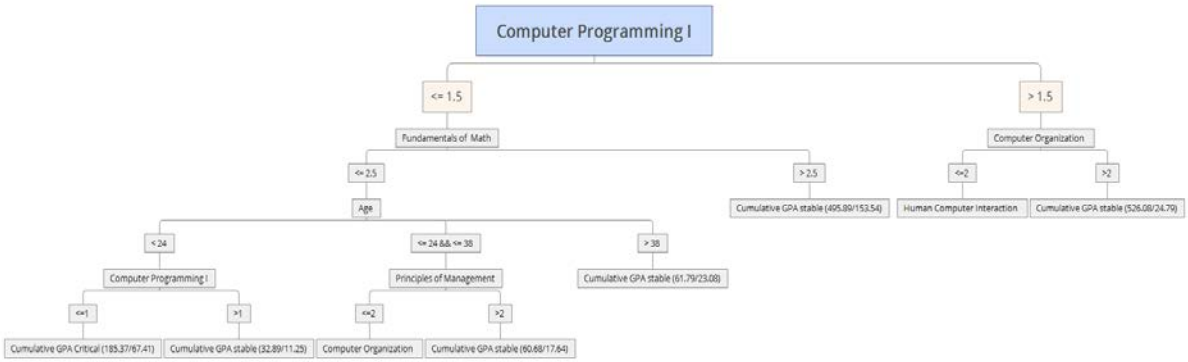


Fig. 3 Decision Tree for Critical Courses

Table 6: Discovered Clusters Description

Attribute		0	1	2	3
		19%	22%	9%	49%
AGE	< 24	13.6224	22.2987	19.4951	28.5838
	>= 24 && <=38	25.4312	18.4973	5.4467	93.6248
	> 38	11.6816	15.6013	1.285	1.432
	[total]	50.7351	56.3974	26.2268	123.6407
Gender	Female	45.3166	4.5807	21.3747	18.728
	Male	4.4186	50.8166	3.852	103.9128
	[total]	49.7351	55.3974	25.2268	122.6407
COURSE_TITLE	Technical Writing	7.7291	10.2569	3.425	17.5889
	Intro to Islamic culture	5.7209	6.5446	1.6087	15.1257
	Computer Programming I	8.5594	10.0576	4.7943	20.5887
	Computer Organization	8.9226	10.3379	3.6029	18.1366
	Discrete Mathematics	10.253	9.7883	4.758	22.2006
	Introduction to IT & IS	8.2024	8.5472	4.4507	20.7997
	Prof. Conduct &Ethics in Islam	1.2497	1.1024	1.4965	4.1515
	Operating Systems	1.8148	1.6478	2.9683	2.5691
	Principles of Management	1.7489	2.2644	3.0132	3.9735
	Software Engineering	1.0324	1.0159	1.0033	1.9484
	Islamic Economic System	1.0444	1.0216	1.0044	2.9296
	Statistics	1.7322	1.7248	1.2303	1.3126
	Computer Programming II	1.8386	1.0513	1.9383	2.1719
	IT Systems	1.8867	2.0366	1.9328	1.1439
	[total]	61.7351	67.3974	37.2268	134.6407
Cumulative GPA	< 2.0	6.087	51.8542	21.2831	20.7757
	>= 2.0	43.6482	3.5431	3.9437	101.865
	[total]	49.7351	55.3974	25.2268	122.6407

Table 6 shows the four clusters found by the EM algorithm. The results reveal that all withdrawn student are in level 3 which is the first semester after passing the PY. This suggests that this level is very critical for students. The first cluster shows that 19% of the withdrawn students are mainly female and mainly in age period between 24 and 38. Also, the fourth cluster indicates that male student in age between 24 and 38, who fail mainly in level 3 courses represents 49% of the withdrawn students. These results, indicates that mainly male and female students less than 38 years old have difficulties in level three, which lead them to withdraw. In addition, about half of the withdrawn students are male in age period between 24 and 38. Table 7, shows the results of evaluating the discovered model, with Log likelihood: -4.29866.

Table 7: Evaluated dataset over Clusters

Class	Clustered Instances
0	62 (25%)
1	59 (24%)
2	23 (9%)
3	101 (41%)

6. Conclusions

The present study investigates the effectiveness of applying data mining techniques, namely classification and clustering, to a real educational dataset obtained from a university that adopts blended learning. The results show the possibility of predicting students' performance based on some demographic/personal attributes and their marks and GPA at early stage of the program. In addition, the applied data mining techniques can help in identifying critical courses that affect the performance of students in the degree program, and identifying the attributes of withdrawn students.

The findings discovered by the applied techniques are very valuable to students, college and decision makers for improving learning experience and institutional effectiveness. For example, these insights can help in refining the program curriculum, as well as identifying high- and low-performing students in early years of the program, and then effectively guiding them during their study.

Since the dataset we considered here is for a blended-learning-based degree program, an interesting extension to

the current study is to extend the dataset to include attributes related to students' online participation and interaction. Then, data mining techniques can be applied to examine the relationship between students' online interactions and their course performance, as well as to evaluate the quality of course contents and instructor engagement.

References

[1] Guvenc, E., Student performance assessment in higher education using data mining. 2001, Institute of Science, Bogazici University.

[2] Lara, J., Lizcano, D., Martínez, M., Pazos, J., and T. Riera, A system for knowledge discovery in e-learning environments within the European Higher Education Area – Application to student data from Open University of Madrid, UDIMA. Computers & Education, 2014. 72(1): p. 23–36.

[3] Fayyad, U.M., Piatetsky-Shapiro, G., and P. Smyth, From data mining to knowledge discovery: an overview. Advances in knowledge discovery and data mining, 1996: p. 1–34.

[4] Romero, C. and S. Ventura, Data mining in e-learning. 2006, Southampton, UK: Wit Press.

[5] Klosgen, W. and J. Zytkow, Handbook of data mining and knowledge discovery. 2002, New York: Oxford University Press.

[6] Guruler, H., Istanbullu, A., and M. Karahasan, A new student performance analysing system using knowledge discovery in higher educational databases. Computers & Education, 2010. 55(1): p. 247–254.

[7] Han, J. and M. Kamber, Data Mining: Concepts and Techniques, 2nd edition. 2006: The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor.

[8] Howard, S., Ma, J., and J. Yang, Student rules: Exploring patterns of students' computer-efficacy and engagement with digital technologies in learning. Computers & Education, 2016. 101(1): p. 29–42.

[9] Baker, R.S.J.D., Data mining for education. International Encyclopedia of Education. Elsevier, 2010: p. 112–118.

[10] Baker, R.S.J.D. and K. Yacef, The state of educational data mining in 2009: A review d future visions. Journal of Educational Data Mining, 2009. 1(1): p. 3–17.

[11] Zorrilla, M.E., Menasalvas, E., Marin, D., Mora, E., and J. Segovia, Web usage mining project for improving web-based learning sites. Web mining workshop, Cataluna, 2005: p. 1–22.

[12] Romero, C., Ventura, S., and E. García, Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 2008. 51(1): p. 368–384.

[13] 13. Dutt, A., M.A. Ismail, and T. Herawan, A Systematic Review on Educational Data Mining. 2017. 5: p. 15991–16005.

[14] Silva, C., Silva, C.; Fonseca, J.(2016). Educational Data Mining: A literature review. Advances in Intelligent Systems and Computing. Rocha, A.; Serrhini, M.; Felgueiras, C. (Eds.) Europe and MENA Cooperation Advances in Information and Communication Technologies. Springer Verlag Proceedings. 2016.

[15] Slater, S., et al., Tools for Educational Data Mining: A Review. Journal of Educational and Behavioral Statistics, 2017. 42(1): p. 85–106.

[16] Dekker, G., M. Pechenizkiy, and J. Vleeshouwers, Predicting Students Drop Out: A Case Study. 2009. p. 41-50.

[17] F Superby, J., J.-P. Vandamme, and N. Meskens, Determination of factors influencing the achievement of the first-year university students using data mining methods. 2006.

[18] Feng, M., N.T. Heffernan, and K.R. Koedinger, Predicting State Test Scores Better with Intelligent Tutoring Systems: Developing Metrics to Measure Assistance Required, in Intelligent Tutoring Systems: 8th International Conference, ITS 2006, Jhongli, Taiwan, June 26-30, 2006. Proceedings, M. Ikeda, K.D. Ashley, and T.-W. Chan, Editors. 2006, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 31–40.

[19] Strecht, P., et al., A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. 2015.

[20] Huang, S. and N. Fang, Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers & Education, 2013. 61(Supplement C): p. 133–145.

[21] Bresfelean, V.P., et al., Determining students #x2019; academic failure profile founded on data mining methods, in ITI 2008 - 30th International Conference on Information Technology Interfaces. 2008. p. 317–322.

[22] P. Vandamme, J., N. Meskens, and J. F. Superby, Predicting Academic Performance by Data Mining Methods. 2007. 15: p. 405–419.

[23] Cobo, G., et al., Using agglomerative hierarchical clustering to model learner participation profiles in online discussion forums. 2012.

[24] Bowers, A.J., Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. 2010. 15.

[25] Talavera, L. and E. Gaudioso, Mining Student Data To Characterize Similar Behavior Groups In Unstructured Collaboration Spaces. 2004.

[26] Asif, R., et al., Analyzing undergraduate students' performance using educational data mining. 2017. 113.

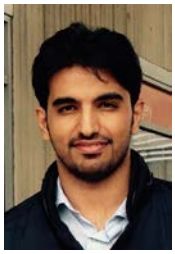
[27] Charitopoulos, A., M. Rangoussi, and D. Koulouriotis, Educational data mining and data analysis for optimal learning content management: Applied in moodle for undergraduate engineering studies. 2017. p. 990-998.

[28] Safavian, S.R. and D. Landgrebe, A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 1991. 21(3): p. 660–674.

[29] Quinlan, R., Induction of Decision Trees. 1986. 1: p. 81–106.

[30] Dempster, A.P., N.M. Laird, and D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 1977. 39(1): p. 1–38.

[31] Jin, X. and J. Han, Expectation Maximization Clustering, in Encyclopedia of Machine Learning, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 382–383.
[32] Sammut, C. and G.I. Webb, eds. Encyclopedia of Machine Learning. 2010, Springer US: Boston, MA.
[33] Weka, Weka. 2005, Machine Learning Group at the University of Waikato.
[34] Quinlan, R., C4.5: Programs for machine learning. 1993: Morgan Kaufman Publishers.



Abdullah Alsheddy received the B.S. in Computer Science from King Saud University in 2002, and MSc and PhD in Computer Science from University of Essex in 2007 and 2011, respectively. Since 2011, he is an assistant professor at Computer Science department, Al Imam Mohammad Ibn Saud Islamic University. His research interest includes optimization and machine learning.

Mohamed Habib received a first class-honors Bachelor of Engineering from Suez Canal University in 2000 and an MSEE and a Ph.D. from the Electrical Engineering Department at the Suez Canal University in 2005 and 2010, respectively. From 2010-Present, he joined Portsaid university as Assistant Professor. From 2014-Present he is a visiting Assistant Professor in Saudi Electronic University. He has many publications in Data Mining, Machine Learning and AI.