Predict Students' Academic Performance based on their Assessment Grades and Online Activity Data

Amal Alhassan[1], Bassam Zafar[2]

Information Systems Department, FCIT
King Abdulaziz University, Jeddah, Saudi Arabia

Ahmed Mueen[3]

CIT Department, Faculty of Applied Studies
King Abdulaziz University, Jeddah, Saudi Arabia

Abstract—The ability to predict students' academic performance is critical for any educational institution that aims to improve their students' learning process and achievement. Although students' performance prediction problem is studied widely, it still represents a challenge and complex issue for educational institutions due to the different features that affect students learning process and achievement in courses. Moreover, the utilization of web-based learning systems in education provides opportunities to study how students learning and what learning behavior leading them to success. The main objective of this research was to investigate the impact of assessment grades and online activity data in the Learning Management System (LMS) on students' academic performance. Based on one of the commonly used data mining techniques for prediction, called classification. Five classification algorithms were applied that decision tree, random forest, sequential minimal optimization, multilayer perceptron, and logistic regression. Experimental results revealed that assessment grades are the most important features affecting students' academic performance. Moreover, prediction models that included assessment grades alone or in combination with activity data perform better than models based on activity data alone. Also, random forest algorithm performs well for predicting student a cademic performance, followed by decision tree.

Keywords—Predict student performance; learning management system; data mining; educational data mining; classification model

I. INTRODUCTION

Educational data mining (EDM) is an emerging field in data mining; aims to transform data accumulated in the educational system into information help educational institution to make informed decisions [1]. EDM uses data mining tools and techniques in education field to analyze student performance, predict their outcomes to help students at risk of academic failure, and provide feedback for the faculties and instructors to improve student outcomes and their learning process [2]. Most of the previous works have proved the effectiveness of data mining to address various educational issues. Student performance prediction is one of the most important issues studied by data mining techniques.

Moreover, the growing use of the internet in education produced a new context called web-based learning or learning management system (LMS). LMS is a web-based application for managing online learning. LMS allows an educational institution to manage students, monitor their participation and tracking their progress via the system [3]. LMS can provide accurate insight into student online activity and their learning

behavior because all data related to students' actions and events are monitored and recorded [4]. These data can be useful to analyze students learning behavior and create prediction models for their performance.

Predicting student performance is a crucial issue for each educational institution aims to improve students' performance and their learning process. Based on the prediction output, an educational institution can support those identified as low performing students. Although predicting students' performance is widely studied, it still a challenge and complex process because students' performance influenced by different features such as demographic, social, academic, economic, and other environmental features [5, 6]. Cognition of these features contributes to control their impact on student performance.

The main objective of this research is to investigate the impact of assessment and activity features from LMS on students' academic performance. Based on one of the most common data mining techniques for prediction, namely classification. Five classification algorithms are applied include decision tree (J48), random forest (RF), sequential minimal optimization (SMO), multilayer perceptron (MLP), and logistic regression (Logistic) for predicting students' performance.

The rest of this paper is structured in six sections. In Section 2, a review of the related work is presented. In Section 3, concepts and definitions related to this research are introduced. Section 4 explains the methodology followed to predict students' academic performance and identify the important features that affect their performance. Section 5 discusses the experimental results with previous works. Section 6 presents the conclusion and limitations of this study. The insights about future work are provided in Section 7.

II. LITERATURE REVIEW

In recent years, many researchers studied features extracted from Learning Management System (LMS) and whether can be used as predictors for students' academic performance. As in [7] researcher investigated the important behavioral indicators from LMS to predict student outcomes in online courses. The researcher considered indicators that reflect regular study as important features can be used to predict student outcomes. Other researchers investigated the impact of students' participation in an online discussion forum on their academic performance [8, 9]. In [8] the authors used qualitative, quantitative, and social network forum indicators to predict student performance. While in [9] students' performance was

predicted based on participation in the discussion forum and their academic records (e.g. assignments, quizzes, and exams).

Moreover, the impact of student online activity on their academic performance have been studied in different forms. In [10] researcher looked into four features of student activity on Moodle that are the number of viewed files, exchanged messages, completed quizzes, and content created by the student. While in [12] performance of 22 students was predicted based on their academic records and time spent by a student on Moodle. However, in [11] researchers considered online assessment data as indicators for student activity. They examined activity data from LMS in the form of assessments and exams to improve student engagement in a blended learning. Another study has predicted the performance of students based on enrollment data and activity on LMS [6]. They considered the heterogeneity of different students' sub-groups to predict their performance based on important enrollment data (e.g. gender and attendance type) and the level of their online activity.

Other studies looked into the different feature sets to predict student academic performance rather than all features in the dataset. As in [13] authors investigated the influence of different feature sets such as course features, student features, behavioral features, and past performance in the course on students' performance. Also, in [14,15] they examined the impact of different feature sets such as demographic, academic, behavioral, and extra features related to parents' participation in the learning process and student absence days. Furthermore, other researchers proposed the use of sub-groups (or sub-datasets) to construct effective prediction models, as in [6] they divided students' dataset into sub-datasets using enrollment and activity data to predict their academic performance. Moreover, sub-datasets is used in [16] to predict student dropout at academic institutions using enrollment data, first-term, and second-term data.

Many works employed feature selection algorithms to create an effective classification model by excluding irrelevant and redundant features from the dataset [9, 3, 17, 18]. Feature selection algorithms can be divided into two basic groups are filter and wrapper methods. Different feature selection algorithms have been applied and compared in past works, as in [19] comparative study conducted using filter-based methods to evaluate the performance of the classification algorithm before and after feature selection application. Moreover, in [20] the performance of different filter-based methods was compared for predicting students' performance in the final exam. Also, in [21] researchers evaluated and compared the performance of different filter and wrapper methods on the dataset that has been gathered for predicting students' grades in the final examination.

Classification is one of the most common data mining techniques for prediction. Classification is a supervised learning process that predicts the class label of the target variable for a given dataset. In the classification model, the dataset is partitioned into two sets are training set for the learning process and test set to implement the classification process. Several classification algorithms have been used in previous works to predict students' academic performance

[22]. In this research, five classification algorithms are used include Decision Tree (J48) [23], Random Forest (RF) [24,2,12], Sequential Minimal Optimization (SMO) [13], Multilayer Perceptron (MLP) [25] and Logistic Regression (Logistic) [26] for predicting students' academic performance. These algorithms are used depending on their effectiveness in previous works for predicting students' performance.

Decision Tree (J48) is widely used for classification. J48 uses the C4.5 algorithm for constructing a decision tree. It similar to a tree structure and consists of three types of nodes are root, internal, and leaf nodes. This method partition the training set into several subsets recursively using the best features selected by merit criteria until it reaches termination, the termination occurs when all features values belong to a class label [27]. Random Forest (RF) constructs multiple decision trees instead of a single decision tree. Trees are constructed based on different samples and features selected randomly from the dataset to form the forest. It gets the result of the prediction from each created decision tree and selects the best prediction result based on the voting process [28, 29].

Sequential Minimal Optimization (SMO) uses an optimization technique for training support vector machine (SVM) [6]. SMO performs the classification process by finding the linear hyper-plane that can differentiate between classes very well. It can also deal with non-linear classification problems using "kernel" technique to convert low dimensional data space to a higher dimension that allows classifying the data [30]. Logistic regression (Logistic) is used in classification problems for prediction based on probability concept. This algorithm differs from linear regression by using "logistic function" instead of linear function for mapping the values of the prediction to probabilities. The probability of a dependent variable that has a binary value is predicted using a set of different independent values [30, 29]. Multilayer Perceptron (MLP) is a multilayer network of interconnected neurons. Neurons are represented in three layers include input, hidden, and output layer. MLP uses "sigmoid function" in hidden and output layers to predict probability [30]. This algorithm learns in the training process by adjusting the weights iteratively using a backpropagation function to attain sufficiently good output [31].

Previous works have studied the impact of different features on students' academic performance, but few works have focused on the impact of assessment and activity data together. Moreover, most of the previous works have used the whole dataset to construct the prediction models. These comprehensive models may unuseful to identify the effect of different features on student performance. However, this work contributes by investigating the impact of assessment and activity data jointly and separately. Sub-datasets is used to create prediction sub-models instead of the whole dataset; sub-datasets have been used in [6]. This work differs from [6] by studying other features related to students' assessment data and their online activity in the form of course access and mobile course access measurements. Additionally, feature selection using two different methods are applied to identify the most important features that affect students' academic performance. Finally, the performance of created prediction models is evaluated and compared.

III. BACKGROUND

A. Imbalanced Class Distribution

The imbalanced class distribution occurs in the classification model when the number of instances in one class is significantly lower than the number of instances in the other class. A class with a small number of instances called minority class, while the class with a large number of instances known as the majority class. The performance of machine learning algorithms is best when the classes are almost balanced in the dataset. Hence, the application of machine learning algorithms on an unbalanced dataset leads to bias the result into the majority class [30]. Several solutions have been proposed in previous works to handle an imbalance in the dataset [32]. This research looked into feature selection and sampling algorithms for solving the imbalance problem.

B. Feature Selection

Feature selection considered one of the most important data pre-processing steps, is used frequently in previous works to identify the relevant features as a subset of the original features in the dataset [3]. The subset produced by feature selection allows classifiers to reach the optimal performance and can be a good solution for imbalanced classes' distribution [32, 33]. This research looked into two methods of feature selection are filter and wrapper methods. Filter method uses a ranking technique to rank the features; the highly ranked features are applied to the classifier, while other features are excluded from the dataset [3]. While wrapper method selects a subset of features using an induction algorithm as a "black box" method to search for a good subset of features [20]. The accuracy of the inducted algorithm is estimated using the techniques of accuracy estimation.

C. Sampling

Sampling (or Resampling) is a technique used to resample the dataset artificially for balancing the number of instances in the classes [34]. It is considered a data pre-processing step and can be achieved by two ways are under-sampling the majority class and over-sampling the minority class.

D. Environment

All algorithms used in this research are implemented using the Waikato Environment for Knowledge Analysis (WEKA), which has been developed by Waikato University in New Zealand [31]. WEKA is a software tool based on java language, provides several algorithms for machine learning and data mining application.

E. Performance Evaluation Measures

This research used different evaluation measures that have been used in the literature to evaluate and compare the performance of classification models. These measures are (1), (2), (3), (4), (5), (6), and (7):

- Accuracy [35]: is the common measure used to evaluate the performance of classifiers, calculates the ratio of correctly classified instances to the total number of instances.

$$Accuracy = TP + TN/(TP + FN + FP + TN) \qquad (1)$$

- Precision [36]: is used to evaluate model exactness. It represents the ratio of true positive instances from all instances classified as positive by a classifier.

$$Precision = TP/(TP + FP) \qquad (2)$$

- Recall [36]: is used to evaluate model completeness. It represents the ratio of true positive instances classified correctly by the classifier.

$$Recall/TPR = TP/(TP + FN) \qquad (3)$$

- F-measure [36]: is used to get the average value of precision and recall. It used commonly by researchers to compare different classifiers performance.

$$F - measure = 2 \times \frac{precision \times recall}{precision + recall} \qquad (4)$$

- Area under ROC curve (AUC) [35]: is used to evaluate the capability of classification model to distinguish between classes. Its value figures out the tradeoff between true positive rate (TPR) and false positive rate (FPR) for a given classification model.

$$TPR = TP/(TP + FN), FPR = FP/(FP + TN) \qquad (5)$$

- Kappa value [6, 29]: is used to measure the accuracy of the classifier compared to the expected random classifier accuracy.

$$kappa = \frac{total\ accuracy - random\ accuracy}{1 - random\ accuracy} \qquad (6)$$

- Root Mean Squared Error (RMSE) [17]: is used to compare prediction errors by evaluating the difference between the actual value and prediction value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Predected_i - Actual_i)^2}{n}} \qquad (7)$$

IV. RESEARCH METHODOLOGY

To predict students' academic performance, the methodology suggested in this research follows five main phases include data collection, data pre-processing, sub-datasets generation, classification algorithms application, and evaluation (see Fig. 1).

A. Dataset

Student data used in this research was obtained from the Deanship of E-Learning and Distance Education at King Abdulaziz University. Data include 241 records for undergraduate students were gathered from six different courses delivered from 2017 to 2019 in the Department of Information Systems, Faculty of Computing and Information Technology. Students' data include assessment grades and activity data on the blackboard. All students' data were extracted from the Learning Management System (LMS) into several Excel files. One file for students' activities on the blackboard and 26 files dedicated to the assessment grades data.
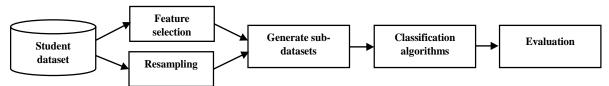
Fig. 1. Method Suggested for Predicting Students' Academic Performance.

During the data collection process, the file that contains students' activity and their IDs is merged with the files that include student IDs and corresponding assessment grades. Then, data cleaned by deleting fields that have few entries and zero values. After that, data transformation is performed. Data transformation is a critical step to convert the data from the format of the source file to the format of the destination file. In this case, the created Excel file converted into (CSV) format then to (ARFF) format to be compatible with the WEKA software tool for data mining application.

The features extracted from the LMS include students' assessment grades and measurements of their online activity on the blackboard. These features are categorized into three major groups that are assessment grades, course access measurements, and mobile course access measurements. The description of these features and their type are shown in Table I.

In King Abdulaziz University (KAU), student performance is assessed using the course grading system. In this system, each course is given a sum of 100 marks distributed for the midterm exams, final exam, and course-work (e.g. quizzes, assignments, projects, and labs work). The final mark earned by a student in the course is corresponding to a letter symbol for the grade [37]. Hence, in this classification problem, students are classified into low-performing students who earned grades D+, D, and F, and high-performing students who earned grades A+, A, B+, B, C+, and C in the course.

B. Data Pre-Processing

Data pre-processing is an essential phase before classification algorithms application. In this research, pre-processing phase includes two steps are feature selection and sampling. After that, the results of feature selection and sampling algorithms are compared to find better algorithm to deal with the imbalance in the dataset and enhance the accuracy of classification algorithms.

Feature selection is applied to select a subset of features that have a greater impact on student academic performance. Moreover, the subset produced by feature selection allows classifiers to reach optimal performance and can be a helpful solution for imbalanced class distribution in the dataset [32, 33]. Therefore, six different filter and wrapper methods are applied on student dataset. Three filter methods are applied include Correlation Attribute Evaluation, Information Gain Attribute Evaluation, and CFS Subset Evaluation [19, 20]. Besides three popular machine-learning algorithms include Decision Tree (J48), Naive Bayes (NB), and K-Nearest Neighbor (IBK in WEKA) are used to implement wrapper method [21]. The results of these six feature selection algorithms show that assessment grades are the most important features that affect student academic performance.

Correlation and Information Gain algorithms give the same high ranking for six features that are assignments mark, final exam, second midterm exam, lab mark, quizzes mark, and first midterm exam. While CFS subset selects four features that are assignments mark, quizzes mark, second midterm exam, and final exam as highly influential features.

TABLE I.   DATASET DESCRIPTION WITH FEATURES AND THEIR TYPE

category	Feature	Description	Type
Assessment data	Assign_Mark	Assignments mark	Numeric
	Quiz_Mark	Quizzes mark	Numeric
	MidTerm_1	First midterm exam mark	Numeric
	MidTerm_2	Second midterm exam mark	Numeric
	Lab_Mark	Lab work mark	Numeric
	Final_Exam	Final exam mark	Numeric
Course access data	Crse_Access	Number of times course access	Numeric
	Crse_Item_Access	Number of times a course item access	Numeric
	Crse_Interaction	Number of course interactions	Numeric
	Crse_Item_Interaction	Number of a course item interactions	Numeric
	Crse_Item_ Mins	Time spent on course item in minutes	Numeric
	Content_Access	Number of times content access	Numeric
	Assessment_Access	Number of times assessment access	Numeric
Mobile course access data	Mob_Crse_Access	Number of times a course access via mobile	Numeric
	Mob_Crse_Item_Access	Number of times a course item access via mobile	Numeric
	Mob_Crse_Interaction	Number of course interactions via mobile	Numeric
	Mob_Crse_Access_Mins	Time spent on a course access via mobile in minutes	Numeric
	Mob_Content_Access	Number of times a content access via mobile	Numeric
	Mob_Assessment_Access	Number of times an assessment access via mobile	Numeric

The subsets produced by wrapper methods show that Wrapper-J48 algorithm selects two features are assignments mark and final exam. While Wrapper-NB subset includes four features that are assignments mark, first midterm exam, final exam, and assessment access. The Wrapper-IBK algorithm determines only one feature is the assignment mark as the most important feature.

For sampling, three algorithms are applied on students dataset include random over-sampling of the minority class (Resample), random under-sampling of majority class (SpreadSubsample), and synthetic minority over-sampling technique (SMOTE), which have been used in [30, 34].

1) Comparison and evaluation results: To compare feature selection and sampling algorithms, these algorithms are applied along with five classification algorithms include J48, RF, SMO, MLP, and Logistic. The performance of these algorithms is evaluated and compared using 10-folds cross-validation and accuracy metric. Evaluation and comparison results of feature selection and sampling algorithms are presented in Table II and Fig. 2.

Results in Table II and Fig. 2 show that both feature selection and sampling algorithms improve the performance of classifiers. For feature selection algorithms, the subset produced by Wrapper-J48 attains the highest accuracy value of 98.42 when classified also using J48 algorithm. While the CFS subset obtains the second highest accuracy value of 97.38 when classified using MLP.

Also, Wrapper-IBK achieves the highest accuracy value of 97.10 for RF algorithm, CFS subset achieves the highest accuracy value of 95.27 for Logistic algorithm, Wrapper-NB achieves the highest accuracy value of 93.53 for SMO.

TABLE II.   ACCURACY RESULTS FOR FEATURE SELECTION AND SAMPLING ALGORITHMS

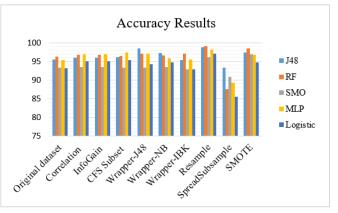| Algorithm	Produced dataset	J48	RF	SMO	MLP	Logistic
	Original dataset	95.51	96.35	93.32	95.31	93.16
Feature selection	Correlation	96.01	96.76	93.40	96.88	95.02
	InfoGain	96.01	96.76	93.40	96.88	95.02
	CFS Subset	96.05	96.46	93.28	**97.38**	**95.27**
	Wrapper-J48	**98.42**	97.05	93.28	97.09	94.19
	Wrapper-NB	97.25	96.67	**93.53**	95.84	94.65
	Wrapper-IBK	95.35	**97.10**	92.87	95.43	92.83
Sampling	Resample	**98.75**	**99.17**	96.17	**98.08**	**97.04**
	SpreadSubsample	93.25	87.50	90.75	89.25	85.50
	SMOTE	97.38	98.51	**96.90**	96.77	94.75

Fig. 2.   Accuracy Rsults for Feature Selection and Sampling Algorithms.

Thus, there is no one feature selection algorithm obtains better accuracy results for all classifiers. However, it observed that the subset produced by the Wrapper-J48 performs better than other subsets, by achieving accuracy results above 97.00 when classified using J48, RF, and MLP.

Results of sampling algorithm in Table II and Fig. 2 show that Resample algorithm obtains the highest accuracy values of 98.75, 99.17, 98.08, and 97.04 for J48, RF, MLP, and Logistic respectively. While SMO obtains its highest accuracy value of 96.17 with SMOTE. Furthermore, SpreadSubsample algorithm obtains the worst results of accuracy, even worse than the original dataset.

For both feature selection and sampling algorithms, Resample algorithm achieves the highest accuracy results with all classifiers, except SMO obtains the best accuracy with SMOTE. However, the use of SMO with Resample algorithm does not result in poor performance, but its performance considered better with SMOTE.

Therefore, Resample algorithm is used to balance the dataset and create more accurate prediction models for students' performance. Hence, SMO is excluded and used the remaining four classification algorithms that are J48, RF, MLP, and Logistic for creating prediction models of students' academic performance.

C. Generate Sub-Datasets

To investigate the impact of student assessment grades and activity data jointly and separately. Students' dataset is partitioned into six sub-datasets based on the three major groups of features (in Table I). The generated sub-datasets are described in Table III.

D. Predicting Students' Academic Performance

After resampling and generate sub-datasets, sub-models are constructed in each sub-dataset displayed in Table III. Additionally, the base model is constructed using "All features" to evaluate the performance of the sub-models compared to the base model. These prediction models are created using four classification algorithms include J48, RF, MLP, and Logistic.

TABLE III. STUDENTS' DATASET AND SUB-DATASETS DESCRIPTION

Dataset	Sub-dataset	Description
All features		Full students' dataset include assessment grades and blackboard activity data.
	Assessment only	Assessment grades without blackboard activity data.
	Assessment + Course access	Assessment grades with course access measurements.
	Assessment + Mobile course access	Assessment grades with mobile course access measurements.
	Course access + Mobile course access	All blackboard activity data without assessment grades.
	Course access only	Course access measurements only.
	Mobile course access only	Mobile course access measurements only.

The performance of base models and sub-models is evaluated and compared using different evaluation measures, which have been used in [6]. In this research, models are trained and tested using 10-folds cross-validation method [3]. In this method, the dataset is divided into ten equal subsets for training and testing. Each subset is run ten times, in each time 90% of instances are trained while 10% of instances are used for testing the model, tested instances in each iteration are different. Then the average of results is computed as the final result.

E. Results

To evaluate and compare the performance of prediction models, first, precision, recall, F-measure, Kappa value, and area under the ROC curve (AUC) are measured. Second, as a complement for previous measures, accuracy and the root mean squared error (RMSE) are measured [6]. All these evaluation measures are computed using 10-folds cross-validation method for all classifiers, where the better results are boldfaced.

1) Evaluate and compare the performance of sub-models to the base model based on Precision, Recall, F-measure, Kappa value, and Area under ROC curve (AUC): Table IV shows the results of Precision, Recall, F-measure, Kappa value, and Area under ROC curve (AUC) achieved by J48, RF, MLP and Logistic classifiers for the base model and sub-models. Results in Table IV show that created sub-models using "assessment only", "assessment + course access", "assessment + mobile course access" features and base model have the same high performance using random forest and J48 classifiers in terms of evaluation measures. For these sub-models and base model, random forest achieves the highest result of 0.99, 0.98, and 1 in terms of f-measure, kappa, and AUC respectively. While J48 achieves the second highest result of 0.99, 0.98, and 0.99 for f-measure, kappa, and AUC, respectively.

TABLE IV. RESULTS OF FIVE EVALUATION METRICS (PRECISION, RECALL, F-MEASURE, KAPPA, AND AUC) FOR ALL DATASETS WITH THE FOUR CLASSIFICATION ALGORITHMS

	Dataset	Precision	Recall	F-measure	Kappa	AUC
J48	**All features**	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**
	Assessment only	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**
	Assessment + course access	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**
	Assessment + mobile course access	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**
	Course access + mobile course access	0.94	0.93	0.93	0.86	0.94
	Course access only	0.93	0.91	0.91	0.83	0.93
	Mobile course access only	0.86	0.81	0.80	0.62	0.88
RF	**All features**	**0.99**	**0.99**	**0.99**	**0.98**	**1**
	Assessment only	**0.99**	**0.99**	**0.99**	**0.98**	**1**
	Assessment + course access	**0.99**	**0.99**	**0.99**	**0.98**	**1**
	Assessment + mobile course access	**0.99**	**0.99**	**0.99**	**0.98**	**1**
	Course access + mobile course access	0.97	0.96	0.96	0.93	**1**
	Course access only	0.97	0.97	0.97	0.94	0.99
	Mobile course access only	0.89	0.85	0.84	0.69	0.95
MLP	All features	0.98	**0.98**	**0.98**	0.96	0.99
	Assessment only	0.98	0.97	0.97	0.95	0.99
	Assessment + course access	0.98	**0.98**	**0.98**	0.95	0.98
	Assessment + mobile course access	**0.99**	**0.98**	**0.98**	**0.97**	**0.99**
	Course access + mobile course access	0.87	0.86	0.86	0.72	0.88
	Course access only	0.86	0.85	0.84	0.69	0.87
	Mobile course access only	0.75	0.71	0.70	0.42	0.81
Logistic	All features	0.97	0.97	0.97	0.94	0.97
	Assessment only	0.92	0.91	0.91	0.82	**0.98**
	Assessment + course access	0.97	0.97	0.97	0.94	**0.98**
	Assessment + mobile course access	**0.98**	**0.98**	**0.98**	**0.95**	0.97
	Course access + mobile course access	0.76	0.75	0.75	0.50	0.80
	Course access only	0.72	0.71	0.71	0.42	0.75
	Mobile course access only	0.66	0.63	0.61	0.26	0.69

Moreover, the sub-model that generated based on "assessment + mobile course access" features outperforms to its base model and other sub-models when using MLP and Logistic algorithms. This sub-model with MLP algorithm achieves results higher than other models in terms of precision and kappa value of 0.99 and 0.97 respectively. Also, this sub-model with logistic algorithm obtains results higher than other models in terms of precision, recall, f-measure and kappa value of 0.98, 0.98, 0.98 and 0.95, respectively.

However, the performance of created sub-models using activity data only (such as "course access only", "mobile course access only", and "course access + mobile course access") deteriorate comparing to performance of the base model. Hence, base model performs better than sub-models created based on activity data only by achieving values above 0.97, 0.94, and 0.97 for f-measure, kappa, AUC, respectively for all classifiers.

Among sub-models created based on activity data only, sub-model that represent the "course access only" features with random forest classifier achieves result better than other activity sub-models. This sub-model obtains values of 0.97, 0.94, 0.99 in terms of f-measure, kappa, AUC, respectively. Followed by the sub-model created using all activity data ("course access + mobile course access") with the random forest, it obtains results better than other sub-models of activity that have been created using J48, MLP, and Logistic algorithms. This sub-model obtains values of 0.96, 0.93, and 1.00 for f-measure, kappa, AUC respectively.

2) Evaluate and compare the performance of sub-models to the base model based on accuracy and root mean squared error (RMSE): Table V shows the results of accuracy and root mean squared error (RMSE) achieved by J48, RF, MLP and Logistic classifiers for the base model and sub-models. Results in Table V show that base model that represents "all features" with random forest superior to all other classifiers and models by achieving the highest accuracy value of 99.17. In addition, random forest ensures sub-models produced using "assessment only", "assessment + course access", and "assessment + mobile course access" features obtain high accuracy value close to 99.00 and low RMSE value of 0.06.

However, the two sub-models based on "assessment only" and "assessment + mobile course access" features with the J48 classifier outperform their base model and other sub-models by achieving the lowest root mean squared error (RMSE) value of 0.04 and accuracy value of 98.92. Moreover, the sub-model that generated using "assessment + mobile course access" features with MLP classifier outperform to its base model and other sub-models by achieving higher accuracy of 98.38 and lower root mean squared error (RMSE) value of 0.07. Also, the sub-model based on "assessment + mobile course access" features using the Logistic classifier outperforms its base model by achieving a higher accuracy value of 97.54.

TABLE V.   RESULTS OF ACCURACY AND ROOT MEAN SQUARED ERROR (RMSE) FOR ALL DATASETS WITH THE FOUR CLASSIFICATION ALGORITHMS

	Dataset	Accuracy	RMSE
J48	All features	98.75	0.05
	Assessment only	**98.92**	**0.04**
	Assessment + course access	98.75	0.05
	Assessment + mobile course access	**98.92**	**0.04**
	Course access + mobile course access	93.13	0.23
	Course access only	91.42	0.27
	Mobile course access only	81.13	0.35
RF	**All features**	**99.17**	0.07
	Assessment only	99.00	**0.06**
	Assessment + course access	99.04	**0.06**
	Assessment + mobile course access	99.00	**0.06**
	Course access + mobile course access	96.33	0.17
	Course access only	96.75	0.18
	Mobile course access only	84.58	0.30
MLP	All features	98.08	0.08
	Assessment only	97.25	0.12
	Assessment + course access	97.71	0.11
	Assessment + mobile course access	**98.38**	**0.07**
	Course access + mobile course access	86.21	0.33
	Course access only	84.63	0.34
	Mobile course access only	71.00	0.42
Logistic	All features	97.04	**0.12**
	Assessment only	90.92	0.20
	Assessment + course access	97.08	0.16
	Assessment + mobile course access	**97.54**	0.13
	Course access + mobile course access	75.00	0.43
	Course access only	71.21	0.45
	Mobile course access only	62.79	0.47

V.   DISCUSSION

This research was to investigate the impact of assessment and activity data on students' academic performance. Therefore, students' dataset was analyzed using different feature selection algorithms to identify important features that affect their academic performance. Moreover, the base model and sub-models based on assessment and activity features jointly and separately were constructed. Also, the performance of used classification algorithms was compared to find the best algorithm for classifying student performance.

Feature selection results revealed the important features that affect student performance are assessment grades, especially assignments mark and final exam. Hence, this research corroborates with the finding reached by [11, 13] they concluded that student performance significantly influenced by assessment data. In [13] they compared four feature sets include student and course characteristics, LMS features, and past performance including assessment grades. Their results demonstrated that student characteristics and the assessment grades had a larger impact on student performance than other features sets. While authors in [11] found a strong correlation between grades of assessments and examinations with students' final grades.

For the base and sub models, the experimental results showed that base model generated from "All features" dataset and classified using random forest algorithm outperforms other prediction models, by obtaining the best results for all performance evaluation measures, especially best accuracy value of 99.17. Moreover, sub-models that include the assessment grades separately or jointly with activity data obtain results better than sub-models rely on the activity data alone for prediction. Additionally, the findings reveal that sub-model generated based on "assessment + mobile course access" features perform well to predict student academic performance.

Researchers in [6] reported that the performance of created sub-models superior to the performance of the base model. In addition to the effectiveness of use students' sub-datasets to predict their academic performance. The results supports this fact to some extent, in terms of the usefulness of investigating the sub-datasets to predict students' academic performance and assess the impact of different features on their success in courses. However, the results revealed that base model based on all features and sub-models that included assessment data separately or jointly with activity features achieved high performance results. That indicates both base model and sub-model perform well to predict students' academic performance.

Regarding the impact of assessment and activity data, results showed prediction models that include assessment grades separately or jointly with activity data have superior prediction results compared to models based on activity data alone. This finding indicates assessment grades affect students' performance significantly, while activity data alone has less impact. Hence, this research corroborates with the finding have been reached by researchers in [11], they revealed a strong relationship between students' online activities in the form of assessments and exams with their final grades in the course. Their finding indicates the importance of assessment data in predicting students' achievement in the course. In addition to the usefulness of investigating students' online activity to assess its impact on academic achievement.

Furthermore, researchers reached a similar conclusion in [13]; they found past performance in the course (including assessment grades) and student characteristics have a greater impact on student performance, while LMS features had a lower impact. The experimental result support this fact, activity data alone have a lower impact on student performance compared to the assessment grades. However, assessment and activity data together enhance the accuracy of the prediction

model. Hence, this finding demonstrates the importance of including the assessment grades with activity data for the prediction model of students' academic performance.

However, researchers in [11] concentrated on online assessments alone as indicators for student activity. Others in [12] investigated only one feature of online activity data which is time spent by a student on Moodle, while Moodle (or LMS) provides more features that can be investigated. Moreover, the dataset used in [12] included 22 instances only; which can be considered a very small number of instances compared to datasets used in previous works. However, this work studied more features of student online activity than those examined in [11, 12], using dataset includes 241 instances. Also, this experiment studied the impact of students' online activity in other forms like course access measurements and mobile course access measurements as well as the assessment grades.

For classification algorithms, the experimental results revealed the random forest algorithm perform better compared to other classification algorithms. This finding is in accordance with findings reported by [12, 24, 2]; they also found random forest algorithm outperform other classification algorithms for student performance prediction using different features such as personal, academic, and activity data. Moreover, in this experiment, random forest algorithm ensures the highest performance results for base and sub models. Followed by decision tree algorithm by obtaining the second highest performance results. As random forest does not provide interpretable results, decision tree can be considered more useful.

VI. CONCLUSION

This research was to investigate the impact of assessment and activity data on students' academic performance. For this purpose, different feature selection algorithms were used to identify the important features that affect students' academic performance. Also, prediction models were constructed based on assessment and activity data jointly and separately using four classification algorithms that are decision tree, random forest, multilayer perceptron, and logistic regression.

Results of feature selection revealed that the most important features that affect student academic performance are assessment data, especially assignments mark and final exam. For prediction models, results demonstrated that both base model and sub-model perform well for predicting students' academic performance. Random forest outperformed other classifiers to predict students' performance by achieving the highest accuracy degrees for both base model and sub model, followed by decision tree. As the random forest does not provide understandable output, the decision tree can be considered more useful.

Furthermore, prediction models that included assessment data separately or jointly with activity data performed better than models based on activity data alone. This indicates that assessment data affect student performance significantly, while activity data have a lower impact. However, assessment and activity data together work better to enhance the accuracy of the prediction model. It is important to include assessment data

with activity data for the prediction model of students' academic performance.

However, certain limitations are observed in this research. The experiment was conducted using data of students for a specific department at faculty. Dataset had only 241 records and 19 features. These results might be different for another dataset with more records and other different features. Also, there might be a possibility of achieving more accurate results by other data mining algorithms.

VII. FUTURE WORK

In future work, this work can be further extended to predict students' academic performance using data from other faculties and different departments to generalize the results. Also, further work may visualize and interpret decision tree result to obtain understandable results help to support low-performing students. Moreover, the same features can be used with other data mining techniques such as regression to predict student final grade in the course, association rule to detect the relationships between students' final grade with their assessment and activity data.

REFERENCES

[1] A. AL-Malaise, A. Malibari and M. Alkhozae, "Students performance prediction system using multi agent data mining technique", International Journal of Data Mining & Knowledge Management Process, vol. 4, no. 5, pp. 01-20, 2014.

[2] S. Hussain, N. Abdulaziz Dahan, F. Ba-Alwi and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA", Indonesian Journal of Electrical Engineering and Computer Science, vol. 9, no. 2, p. 447, 2018.

[3] E. Amrieh, T. Hamtini and I. Aljarah, "Mining educational data to predict student's academic performance using ensemble methods", International Journal of Database Theory and Application, vol. 9, no. 8, pp. 119-136, 2016.

[4] P. Shayan and M. Zaanen, "Predicting student performance from their behavior in learning management systems", International Journal of Information and Education Technology, vol. 9, no. 5, pp. 337-341, 2019.

[5] N. A. Yassein, R. G. M. Helali, and S. B. Mohomad, "Predicting student academic performance in KSA using data mining techniques," Journal of Information Technology & Software Engineering, vol. 07, no. 05, 2017.

[6] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, D. Murray and Q. Long, "Predicting academic performance by considering student heterogeneity", Knowledge-Based Systems, vol. 161, pp. 134-146, 2018.

[7] J. You, "Identifying significant indicators using LMS data to predict course achievement in online learning", The Internet and Higher Education, vol. 29, pp. 23-30, 2016.

[8] C. Romero, M. López, J. Luna and S. Ventura, "Predicting students' final performance from participation in on-line discussion forums", Computers & Education, vol. 68, pp. 458-472, 2013.

[9] A. Mueen, B. Zafar and U. Manzoor, "Modeling and predicting students' academic performance using data mining techniques", International Journal of Modern Education and Computer Science, vol. 8, no. 11, pp. 36-42, 2016.

[10] N. Z. Zacharis, "Classification and regression trees (CART) for predictive modeling in blended learning" , International Journal of Intelligent Systems and Applications, vol. 10, no. 3, pp. 1–9, 2018.

[11] M. Ayub, H. Toba, M. Wijanto and S. Yong, "Modelling online assessment in management subjects through educational data mining", in 2017 International Conference on Data and Software Engineering (ICoDSE), 2017.

[12] R. Hasan, S. Palaniappan, A. Abdul Raziff, S. Mahmood and K. Sarker, "Student academic performance prediction by using decision tree algorithm", in 2018 4th International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 2018.

[13] R. Conijn, A. Kleingeld, U. Matzat, C. Snijders and M. Zaanen, "Influence of course characteristics, student characteristics, and behavior in learning management systems on student performance", in 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 2016.

[14] M. H. Rahman and M. R. Islam, "Predict Students Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques", 2017 2nd International Conference on Electrical & Electronic Engineering (ICEEE), 2017.

[15] B. Francis and S. Babu, "Predicting academic performance of students using a hybrid data mining approach", Journal of Medical Systems, vol. 43, no. 6, 2019.

[16] G. Bilquise, S. Abdallah and T. Kobbaey, "Predicting student retention among a homogeneous population using data mining", in Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2019. AISI 2019, 2020, pp. 35-46.

[17] S. Hussain, N. Abdulaziz Dahan, F. Ba-Alwi and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA", Indonesian Journal of Electrical Engineering and Computer Science, vol. 9, no. 2, p. 447, 2018.

[18] P. Kumari, P. Jain and R. Pamula, "An efficient use of ensemble methods to predict students academic performanc", in 2018 4th International Conference on Recent Advances in Information Technology (RAIT), 2018, pp. 1-6.

[19] S. Gnanambal, M. Thangaraj, V. Meenatchi and V. Gayathri, "Classification algorithms with attribute selection: an evaluation study using WEKA", Int. J. Advanced Networking and Applications, vol. 09, no. 06, pp. 3640-3644, 2018.

[20] C. Anuradha and T. Velmurugan, "Performance evaluation of feature selection algorithms in educational data mining", International Journal of Data Mining Techniques and Applications, vol. 5, no. 2, pp. 131-139, 2016.

[21] A. Acharya and D. Sinha, "Application of feature selection methods in educational data mining", International Journal of Computer Applications, vol. 103, no. 2, pp. 34-38, 2014.

[22] A. Kumar, R. Selvam and K. Kumar, "Review on prediction algorithms in educational data mining", International Journal of Pure and Applied Mathematics, vol. 118, no. 8, pp. 531-537, 2018.

[23] M. Al-Saleem, N. Al-Kathiry, S. Al-Osimi, and G. Badr, "Mining educational data to predict students' academic performance," Machine Learning and Data Mining in Pattern Recognition Lecture Notes in Computer Science, pp. 403–414, 2015.

[24] K. T. S. Kasthuriarachchi, S. R. Liyanage, and C. M. Bhatt, "A data mining approach to identify the factors affecting the academic success of tertiary students in Sri Lanka", Lecture Notes on Data Engineering and Communications Technologies Software Data Engineering for Network eLearning Environments, pp. 179–197, 2018.

[25] F. Widyahastuti and V. Tjhin, "Predicting students performance in final examination using linear regression and multilayer perceptron", in 2017 10th International Conference on Human System Interactions (HSI), 2017, pp. 188--192.

[26] M. Bucos and B. Drăgulescu, "Predicting student success using data generated in traditional educational environments", TEM Journal, vol. 7, no. 3, pp. 617-625, 2018.

[27] N. Bhargav, G. Sharma, R. Bhargava and M. Mathuria, "Decision tree analysis on J48 algorithm for data mining", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 6, 2013.

[28] M. Khaldy and C. Kambhampati, "Resampling imbalanced class and the effectiveness of feature selection methods for heart failure dataset", International Robotics & Automation Journal, vol. 4, no. 1, pp. 1-10, 2018.

[29] C. Zabriskie, J. Yang, S. DeVore and J. Stewart, "Using machine learning to predict physics course outcomes", Physical Review Physics Education Research, vol. 15, no. 2, 2019.

[30] A. Verma, "Evaluation of classification algorithms with solutions to class imbalance problem on bank marketing dataset using WEKA",

International Research Journal of Engineering and Technology (IRJET), vol. 06, no. 03, pp. 54-60, 2019.

[31] R. Arora and S. Suman, "Comparative analysis of classification algorithms on different datasets using WEKA", International Journal of Computer Applications, vol. 54, no. 13, pp. 21-25, 2012.

[32] R. Longadg, S. S. Dongre and L. Malik, "Class imbalance problem in data mining: review", International Journal of Computer Science and Network (IJCSN), vol. 2, no. 1, pp. 1305.1707, 2013.

[33] M. Koutina and K. Kermanidis, "Predicting postgraduate students' performance using machine learning techniques", Artificial intelligence applications and innovations, pp. 159-168, 2011.

[34] P. Yıldırım, "Pattern classification with imbalanced and multiclass data for the prediction of albendazole adverse event outcomes", Procedia Computer Science, vol. 83, pp. 1013-1018, 2016.

[35] A. Tharwat, "Classification assessment methods", Applied Computing and Informatics, 2018.

[36] H. M and S. M.N, "A review on evaluation metrics for data classification evaluations", International Journal of Data Mining & Knowledge Management Process, vol. 5, no. 2, pp. 01-11, 2015.

[37] Undergraduate catalog. 2018, pp.11-12. [Online]. Accessed: Mar 26, 2020. Available :https://fcit.kau.edu.sa/aims/templates/catalog18-19-online1.pdf.