Research Article

EMT: Ensemble Meta-Based Tree Model for Predicting Student Performance

Ammar Almasri [ID],[1] Erbug Celebi,[1] and Rami S. Alkhawaldeh[2]

[1]*Department of Management Information Systems, Institute of Graduate Studies and Research, Cyprus International University, Lefkosa, North Cyprus 99258, Turkey*
[2]*Department of Computer Information Systems, The University of Jordan, Aqaba 77110, Jordan*

Correspondence should be addressed to Ammar Almasri; 20161093@student.ciu.edu.tr

In recent decades, predicting the performance of students in the academic field has revealed the attention by researchers for enhancing the weaknesses and provides support for future students. In order to facilitate the task, educational data mining (EDM) techniques are utilized for constructing prediction models built from student academic historical records. These models present the embedded knowledge that is more readable and interpretable by humans. Hence, in this paper, the contributions are presented in three folds that include the following: (i) providing a thorough analysis about the selected features and their effects on the performance value using statistical analysis techniques, (ii) building and studying the performance of several classifiers from different families of machine learning (ML) techniques, (iii) proposing an ensemble meta-based tree model (EMT) classifier technique for predicting the student performance. The experimental results show that the EMT as the ensemble technique gained a high accuracy performance reaching 98.5% (or 0.985). In addition, the proposed EMT technique obtains a high performance, which is a superior result compared to the other techniques.

1. Introduction

With the growing expansion in large data warehouse, the necessity for analysing the data and extracting the useful information becomes a common question and a rich scholarly topic of examination for many researchers [1]. Data mining techniques are used as analytical tools to extract the hidden knowledge in a form of models from such data warehouses [2, 3]. Deploying it, there are many application areas that adapt data mining techniques in their systems such as finance, marketing, economy, telecommunication, medicine, healthcare, and student performance applications [4]. However, due to the importance of predicting the performance of the students and digitizing the systems of the universities, academic institutes construct a large volume of data pertaining to students using computerized forms [5, 6]. It becomes essential for these academic institutes to transfer the huge amount of data and to manipulate it into useful

knowledge [7]. This aims to help instructors, employees, and authorities to facilitate their missions through analysing data. Therefore, the knowledge plays a vital role in pushing the wheel of education further by providing or extracting significant knowledge [5] by using data mining methods [7].

The educational data mining (EDM) is a mature field that concerns with the process of applying data mining techniques such as classification or clustering on educational data, to better understanding the learning process and students' achievements [8, 9]. This is conducted by finding the knowledge that is interpretable and understandable by humans [10]. In particular, the classification techniques are used to analyse educational data in order to predict students' academic performance. These results enhance the overall university performance and provide a more successful learning environment [10].

The problem of predicting student performance, therefore, is formulated as input features such as GradeID,

topic, and raised hands along with related labels. The features represent an analysis of the student historical data, whereas the label represents the actual performance. The observed records represent the training set that a classifier is derived. In particular, the formulation is as follows: assume that there is a set of students $s \in (\mathbf{S})$ and their historical academic performance records $r \in (\mathbf{R})$ that are gathered during specific semesters. The relevant features are extracted from such R for each student φ $(F_1, F_2, F_3, \ldots, F_n)$, and they are related with an academic performance $P(s)$ in order to build a training set $\mathbf{Ts}$. A classifier is built by mapping the set of features into their corresponding label value using the training set $\mathbf{Ts}$:

$$F\left(\varphi\left(F_1, F_2, F_3, \ldots, F_n\right)\right) \longrightarrow P(s). \qquad (1)$$

The academic institutes aim to enhance the performance of the students by monitoring their historical records in different semesters. The process consumes time and efforts to gather such information for making decision especially for many students. Many works have revealed different approaches that help in providing a decision for improving student performance. These approaches depend on building classifier models that also determine the performance of unknown students using fewer features. Although these techniques have an effect in improving the prediction performance, they do not provide a comprehensive and unified approach that accurately determines the performance. Therefore, the contributions comprise three folds which include three research questions as follows:

(1) Studying different features extracted from student historical academic records and analysing the correlation and relationship between features and their labels (or student performance). These correlations determine the importance of the features on predicting the student performance.

RQ1: what is the relationship between the extracted features and the student performance in order to build prediction systems?

(2) Several ML techniques from different theoretical families are examined on predicting the student performance that indicate how diversity is using these techniques and to what extent they help to improve the performance.

RQ2: how effective is employing various ML techniques in predicting the academic performance of the students and how to exploit the *techniques to improve the performance of new students?*

(3) A proposed approach combines two classification techniques. The approach follows the hypothesis that the performance results of ensemble classifiers built from different techniques are better than using a simple technique of classification models. A proposed model called EMT prediction model is built along with a thorough analysis of its performance on the academic student training set.

RQ3: what is the prediction performance of the EMT technique for estimating student performance?

The paper is organized as follows: Section 2 discusses related works and clarifies the weaknesses that derive the assumption in the proposed techniques. Section 3 presents the methodology for studying the extracted features, examining the conducted ML techniques and building a cluster-based prediction model. Section 4 investigates the experimental settings, results, and discussions, followed by conclusions and future works in Section 5.

2. Related Work

Several studies have emerged to enhance student performance [5, 11, 12]. These either focus on the effect of variant factors on student performance or build an appropriate classification model to predict the future unknown performance. Therefore, this section discusses the two aspects with a plethora of research.

2.1. The Student Performance Factors. In literature, there are many features that were studied which influence the final prediction of students' results. Table 1 summarizes the most features that were used to predict student performance. Such features were extracted from students' historical records and have been used as input for constructing prediction models.

The researchers investigated the factors that influence the academic performance of the students, especially those of low academic performance in educational institutions as shown in Table 1. The factors that are gathered during the academic life of the students include CGPA (grade point average), internal assessment, students' demographic information, external assessments, extracurricular activities, high school background, social interaction network, and psychometric. The most frequent factors from the reviewed studies are CGPA, internal assessment, and students' demographic information such as gender, age, and salary income. In respect to the other factors, the external assessments also used to predict student performance that includes the behavior of students outside the class. The other important factors that have not been used frequently in previous studies are the extracurricular activities, high school background, social interaction network, and psychometric.

2.2. Prediction Methods. The essential aspect having more attention in the area of student performance prediction is the prediction models that are used for classifying the student performance. There are variants of ML techniques categorized into a set of families where each family reflects a specific theoretical idea. As discussed in the Introduction, the EDM field has studied different ML techniques to determine these techniques obtaining a high accuracy to predict the future performance of students [7]. Table 2 summarizes the most used classification algorithms, which were used to predict student performance in educational dataset; several works have been investigated to find the best algorithms that can be used to predict the future student performance.

TABLE 1: Summary of the common attributes used in predicting student performance.

Factors or features	Description	References
(CGPA)	Cumulative grade point average	[11, 13–17]
Internal assessment	Students' behavior in the class such as exams marks, lab tests, class tests, attendance, and discussions	[5, 11, 12, 13, 18]
Students' demographic	Students' characteristics that include gender, age, salary income, and family background	[14, 19–22]
External assessments	Factors coming from student's environment: student's behavior out of the class such as extracurricular activities, high school background, social interaction network, and psychometric	[13, 17, 23–26]
Extracurricular activities	Activities which are not essential for normal class activities	[11, 13, 17, 27]
High school background	Preadmission courses for universities	[20–22]
Social interaction network	Interacting among students or instructors by using social networks	[12, 20, 24, 28]
Psychometric	These attributes can be collected after enrolment such as attitude, motivation, personality, and learning strategies	[27, 29, 30]

TABLE 2: Summary of the common classifiers which are used in predicting student performance.

Research study	Main results
[28]	(i) *SMO*, *NaiveBayesSimple*, and *BayesNet* obtained the highest accuracy and F-measure.
[17]	(i) Decision tree models (REPTree) had a high prediction accuracy. (ii) *REPTree* was less sensitive to missing values than *J48*.
[15]	(i) *Neural network* and *Naive Bayes* classification with SMOTE technique had a high accuracy with 75%. (ii) Naive Bayes and neural network models produced almost similar accuracy level when the discretization method was applied. Decision tree had less accuracy for both methods.
[31]	(i) Three decision tree algorithms *C4.5*, *CART*, and *ID3* were applied and the result indicated that C4.5 is the best classifier for prediction of student performance.
[29]	(i) Six classification algorithms considered were Naïve *Bayes (NB)*, *Unpruned Decision Tree5 (DT)*, *logistic regression*, *support vector machine using an ANOVA kernal function (SVM)*, *neural network (NN)*, and *k-nearest neighbor (k-NN)*. (ii) The results indicated that all algorithms had a good predictive accuracy for young students and KNN predicted very well for old students and the rest of the classifiers were poor.
[27]	(i) Two algorithms were used: *J48* and *random tree*. (ii) The result showed that random tree model was more accurate than J48.
[14]	(i) *NBTree* classification was performed with a pretty good accuracy.
[11]	(i) This study used *C4.5*, *AODE*, *Naïve Bayesian*, *multi label k-nearest neighbor* algorithms. (ii) The result concluded that multilabelled k-nearest neighbor had the best accuracy among the others (C4.5, AODE, and Naïve Bayesian).

TABLE 2: Continued.

Research study	Main results
[20]	(i) Three algorithms were compared: *C4.5*, *multilayer perceptron*, and *Naive Bayes*. (ii) Naive Bayes has a good prediction accuracy.
[4]	(i) Four classifiers were investigated: *J48 DT*, *NB*, *SMO*, and *MLP*. (ii) The results show that J48 DT algorithm achieves the best performance compared to the other algorithms with an accuracy of 84.8%.
[2]	(i) *BN*, *NB*, *SVM*, *C4.5*, and *CART* are used to build the learning model to predict student performance. (ii) *SVM* is the best classifier compared to the other BN, NB, C4.5, and CART.
[1]	(i) Two main algorithms, *decision stump* and *J48*, were applied. (ii) J48 provides more accuracy.

Romero et al. [28] found that sequential minimal optimization (SMO), NaiveBayesSimple and BayesNet techniques had the highest value to accuracy and F-measure. Jishan et al. [15] indicated that neural network and Naive Bayes classification models using SMOTE technique in an imbalance dataset had an accuracy of 75%, while the Naive Bayes and neural network models produced almost similar accuracy level when the discretization method is applied. By applying the decision tree, the result indicated that C4.5, CART, and ID3 had the best classifiers for the prediction performance of students [31].

A study was conducted in [20] to compare the C4.5 with multilayer perceptron and Naive Bayes. The results indicated that the Naive Bayes method has a good prediction accuracy among the other classifiers. Similarly, the decision tree models (REPTree) had a highest prediction accuracy [17].

Many researchers in their recent papers, as shown in Table 2, have made significant results in their understanding of appropriate factors that influence the students' achievement.

Also, they have shown that all classifiers can predict student performance with reasonable results. Moreover, selecting the most fitting technique is not the same in the previous studies since they have different datasets in different contexts as shown in Table 1.

3. Ensemble Meta-Based Tree Model

The performance of students in academic institutions provides an indication of how much effort such institutions must maintain to improve the low or even medium performance. The importance of employing ML techniques, exploiting the historical data of the students in order to predict unknown or future performance, has witnessed an intuitive attention motivating us to construct a model for predicting the unknown labels of future instances. The proposed method dubbed EMT is an ensemble technique which combines the best-selected techniques as the final prediction model. A proposed methodology is followed to show how the technique is constructed in an accurate manner. The methodology comprises three folds which include preprocessing phase, construction phase, and the evaluation phase as shown in Figure 1. These phases help to increase the EMT model performance to predict the student performance. Each phase is discussed as follows.

3.1. Preprocessing Phase. The constructed model of performance prediction depends on historical records of the students as a training set. Thus, gathering such records reveals utmost importance in order to increase the accuracy of the constructed model. The process of aggregating the records is conducted in a variety of techniques such as student interviewing, questionnaires, online student evaluation, etc. In most cases, the historical records are organized in unstructured forms, such as documents, which complicate the task of extracting the patterns from them. Therefore, converting the documents into an appropriate structural form simplifies the job of ML in building the predictive models. The training set maintains a set of features selected from the documents (or students' records) of 480 records along with student performance of low level (≤69%), middle level (70% to less than or equal to 89%), and high level (90% to less than or equal to 100%) [10]. The dataset includes sixteen attributes from a registration office that has information of preregistration students. These attributes are gathered during an academic semester. To clean the dataset, the outliers such as inconsistent and missing values are removed. The dataset is reduced to 13 attributes and 400 records, and the results are shown in Table 3 (for more descriptions, see Appendix A).

3.2. Construction Phase. The training set is ready to be used for constructing the predictive models of student performance. There is a set of ML techniques from different families that could be used to construct the predictive models. These families differentiate from each other in the theoretical process used for building the model. These

models can be deployed in systems for classifying future instances. Machine learning techniques construct a hypothesis from a space set of hypotheses using in the training set. Each hypothesis represented as a mathematical model (or pattern) to map the input instances into output labels. In machine learning, there are models classified into groups of the same theory but with slight differences in technical parts. In this paper, a set of learning techniques selected from different families are used. The classifiers in terms of family perspective are summarized as follows which are also used in the related works [32]:

(1) *Bayes*: This is a family that relies on the probability theory for building a classifier. This family builds a classifier on a form of rule-based or network models reflecting the probability of classes given specific features. In particular, the model uses the prior probability, the probabilities of observing various data given the hypothesis, and the observed data itself in the construction process such as NaiveBayes and BayesNet.

(2) *Functions*: The classifiers in this family aim to construct a model of function. The function consists of input features and output labels. There are many techniques to map the inputs to the outputs, such as the neural network that uses the feed-forward and backpropagation methods for updating the network, in order to enhance the prediction by reducing the error of loss function. The other techniques also include logistics (regression), support vector machine (i.e., hyperplanes), etc.

(3) *Lazy*: The classifier models use the training set to find the output label from the most similar common labels in the training set. For instance, the future instance features are compared with the training set features, where the labels of the most similar instances are used to predict the value of that future instance. There are a set of techniques in this family such as IBk, KStar, and LWL.

(4) *Meta*: This family suggests that using a set of ensemble techniques increases the performance of the prediction models. Notably, the ensemble classifier is built by combining a set of weak learning classifiers, either by using voting or weighting techniques of their results. Each classifier is built by using a random sample of the training set. There are many techniques in this family such as AdaBoost, Bagging, and LogitBoost.

(5) *Trees*: This family builds classifiers in a form of trees where the nodes represent the attributes except the leave nodes represent the labels, while the arcs (or edges) represent the values of that attribute. The attribute is selected using different methods, which depend on entropy information that reflects the importance of attribute with low chaos. There are a set of techniques that differentiate with each other in selection methods such as decision tree (or J48), LMT, and random forest.
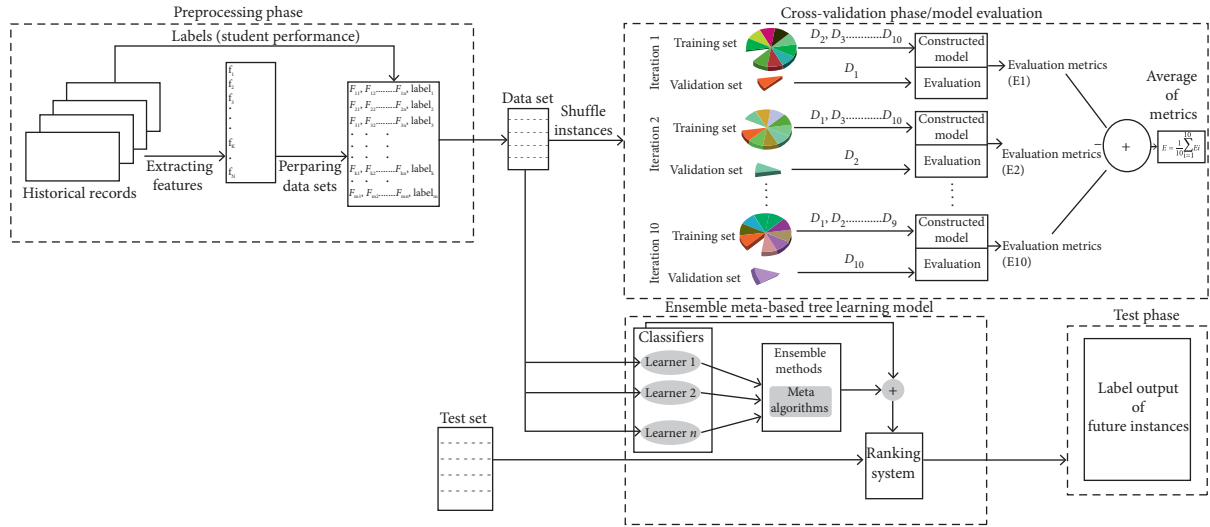
Figure 1: The architecture and evaluation of EMT technique.

Table 3: The main features used in predicting the performance of students.

Information type	Attribute name	Values
Part A Student information	(i) Gender	{Male, Female}
Part B Education information	(i) Stage_ID (ii) Semester (iii) Topic	{LowerSchool, MiddleSchool, HighSchool} {First, Second} {IT, Math, Arabic, Science, Quran, English, Spanish, French, Biology, Chemistry, Geology, History}
Part C Students' behaviors	(i) Student_Absence_Days (ii) Raised_Hands (iii) Visited_Resources (iv) Announcements (v) Discussion	{Under-7, Above-7} The values are equal or greater than 0 The values are equal or greater than 0 The values are equal or greater than 0 The values are equal or greater than 0
Part D Family behaviors	(i) Relation (ii) Parent_Answering_Survey (iii) Parent_School_Satisfaction	{Father, Mother} {Yes, No} {Good, Bad}
Part E Students' status	(i) Class	{L, M, H}

The EMT is an ensemble bagging technique that combines two methods from boosting and the best ML technique of whole families into a predictive model as shown in Figure 1. The boosting method is the most popular and powerful data mining algorithm that obtains a high prediction performance by combining several weak learners. It is an iterative process, in which a classifier is generated as a weak learner at each iteration. Then, an ensemble boosting classifier combines these weak learners with a coefficient weight for each learner into a strong prediction model with high accuracy performance. In this paper, the AdaboostM1 method is used to train more than one learner to solve the problem of student prediction performance. The best prediction model from these learners is selected as one of the EMT two methods. On the other hand, the bagging methods evaluate the output of the weak learners using two techniques. The techniques involve voting and averaging methods. In the voting method, the most popular class of weak learners is used as a final output result of the bagging

learner, while the averaging method estimates the average values of the weak learners as a final output.

In summary, the EMT technique combines the best classifier of the whole families on the same training set with the best learners of AdaboostM1 method as a bagging technique using the voting method between the two techniques.

3.3. Evaluation Phase. In order to evaluate the EMT technique to generalise it for future instances, the 10-fold cross-validation technique is used. The evaluation model part in Figure 1 demonstrates such technique. As shown, every 10 parts hold the same number of instances divided in a random way. At each iteration, one part is used as a validation set, while the other parts are combined as a training set. As a result, the training set is used for building the classifier, whereas the validation set is used to evaluate the classifier by predicting the class of its instances. Herein, a set

of evaluation metrics is also used to examine the performance of the constructed classifier model at each iteration. However, the evaluation values of all iterations are averaged into a final output evaluation metric value which is used as an estimator of the classifier performance. In the evaluation metric perspective, a set of evaluation metrics are used to evaluate the EMT technique, which are accuracy, F-measure, and ROC metrics. The accuracy metric represents the number of instances that are classified correctly over all instances using the three classes, which are low, medium, and high performance. The F-measure is an evaluation metric that combines and obtains the benefits of two evaluation metrics, which are Precision and Recall. The Precision (also called positive predictive value) is the percentage of relevant instances of the retrieved ones, while Recall (also known as sensitivity) is the percentage of retrieved and relevant instances over all the total number of relevant ones. The ROC area (or curve) is used to examine how the classifier can differentiate between positive and negative instances and identify a threshold for separating them:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})},$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})},$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}, \tag{2}$$

$$\text{F-measure} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}.$$

TP refers to true positive, which is the number of instances that are correctly labelled as belonging to the positive class. TN refers to true negative, which is the number of negative instances that are correctly classified as negative. FP refers to false positive, which is the number of positives that are incorrectly classified as negative. FN refers to false negative, which is the number of negative instances that are incorrectly classified as positive.

4. Results and Discussion

The experimental results are conducted to cover the research questions that ensure that the proposed technique for predicting the student performance gains high effective performance. The research questions clarify three important issues, which are (1) features analysis, (2) analysing the prediction models, and (3) proposing out a technique called ensemble meta-based tree model (EMT model).

4.1. Features Analysis. In order to evaluate the features and examine their effect on the student academic performance (i.e., class labels), the Pearson correlation method is used as an analytical technique [33]. Figure 2 shows the results of conducting the Pearson correlation method to predict the final grades of students. The attribute of visited resources occupies 0.436, the factor of student absence days gets 0.399,

raised hands obtains 0.376, announcement view achieves 0.33, the factor of relation gets 0.312, the attribute of parent answering survey obtains 0.272, the factor of parent-school satisfaction gets 0.196, the attribute of gender scores 0.164, the attribute of discussion occupies 0.160, stage attribute scores 0.077, semester factor occupies 0.075, and finally topic factor gets 0.06. The results indicate that the predictor variables have a statistically significant correlation or (p value < 0.01) with a class value.

The results, as shown in Figure 2, reflect strong positive correlations between all independent attributes and dependent attribute (or class). The correlation ranged between 0.06 as the lowest value for the topic and 0.436 as the highest value for visiting resources. It is worth mentioning that the factor of visiting resources highly affects the final results of students, which proves the fact that it highly influences their final results and it can raise them to get middle or high results instead of getting the low results. Therefore, academic institutions must focus on this factor by facilitating this service and encouraging their students to visit available resources. On the other hand, the topic indicates low correlation due to the various contents of courses that depend on the syllabus and the methods the instructors use during the class. These results might help instructors, administrators, and policymakers to make a better decision regarding the learning process. Finally, these results answer affirmatively *RQ1* that each factor behaves in a different way related the prediction class.

4.2. Prediction Model Analysis. Many researchers in their recent papers have made significant results in understanding the appropriate factors that influence the students' achievement. The question is how much effect these factors would be on the classification methods. In this section, 47 classification techniques are conducted using the dataset. The factors of this dataset are the extracted features along with the labels that represent the student performance. In order to evaluate the classification techniques, the evaluation phase is followed as discussed in Section 3. As shown in Table 4, the classifiers obtain consistent results for both F-measure and accuracy measures. These results show that the extracted features have a role to evolve the prediction models from students' historical records with reasonable accuracy.

Regarding classification outcomes, Table 4 shows the best algorithms of different families where each family behaves in different ways to extract patterns from the training set. Each family of ML techniques follows a theoretical framework for building the required model. As such, it is normal that the constructed model behaves in different ways for identifying the target classes (i.e., low, medium, and high) and provides diverse results. Each family has a model with best result which includes PART (rules), A2DE (Bayes), multilayer perceptron (functions), LocalKnn (lazy), and J48 (trees) algorithms with 91.8%, 89.5%, 91%, 92.8%, and 94.3% accuracy, respectively (see Appendix B). The most effective family in predicting the student performance is the tree family compared to the other families with high value in accuracy and F-measure. In particular, the J48 algorithm outperforms the whole 46 classifiers from different families
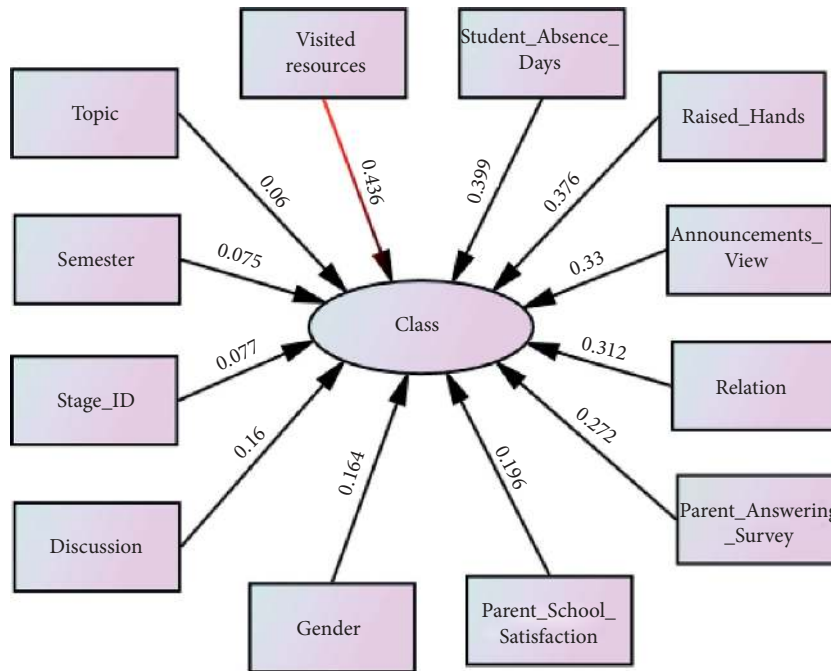
Figure 2: Pearson's correlation between each attribute and class.

Table 4: Comparison between five groups of classification algorithms (accuracy and F-measure values).

Algorithm's family	Algorithm name	Dataset	
		Accuracy	F-measure
Rules	ConjunctiveRule	0.633	0.533
	DecisionTable	0.850	0.849
	DTNB	0.855	0.854
	FURIA	0.888	0.887
	JRip	0.873	0.872
	LAC	0.780	0.781
	MODLEM	0.873	0.872
	Multiobjective EvolutionaryFuzzy	0.628	0.624
	NNge	0.830	0.831
	OLM	0.668	0.661
	OneR	0.695	0.692
	PART	**0.918**	**0.918**
	Ridor	0.875	0.874
	RoughSet	0.870	0.869
	ZeroR	0.433	0.261
Bayes	A1DE	0.880	0.880
	A2DE	**0.895**	**0.895**
	BayesNet	0.850	0.849
	NaiveBayes	0.773	0.769
Function	LibLINEAR	0.690	0.688
	LibSVM	0.540	0.470
	MLPClassifier	0.893	0.892
	Multilayer perceptron	**0.910**	**0.910**
	MultilayerPerceptronCS	0.910	0.910
	RBFClassifier	0.853	0.853
	SMO	0.870	0.870
Lazy	IB1	0.838	0.837
	IBK	0.838	0.837
	IBKLG	0.838	0.837
	KStar	0.863	0.862
	LocalKnn	**0.928**	**0.927**
	RseslibKnn	0.928	0.927

TABLE 4: Continued.

Algorithm's family	Algorithm name	Dataset	
		Accuracy	F-measure
	BFTree	0.905	0.905
	CDT	0.903	0.902
	DecisionStump	0.558	0.406
	ForestPA	0.910	0.910
	FT	0.883	0.883
	HoeffdingTree	0.773	0.769
	J48	**0.943**	**0.943**
Trees	J48Consolidated	0.940	0.940
	LADTree	0.900	0.900
	LMT	0.933	0.932
	NBTree	0.925	0.925
	RandomForest	0.935	0.935
	RandomTree	0.828	0.827
	REPTree	0.893	0.893
	SimpleCart	0.893	0.892

TABLE 5: Boosting and Bagging a nominal class classifier Results.

Algorithm name	Dataset			Boosting method			Bagging method		
	Accuracy	F-measure	ROC area	Accuracy	F-measure	ROC area	Accuracy	F-measure	ROC area
PART	0.918	0.918	0.973	0.955	0.955	0.994	0.925	0.925	0.989
A2DE	0.895	0.895	0.973	0.877	0.877	0.970	0.893	0.893	0.977
Multilayer perceptron	0.910	0.910	0.968	0.920	0.920	0.960	0.905	0.905	0.981
LocalKnn	0.928	0.927	0.943	0.925	0.925	0.982	0.918	0.917	0.982
J48	0.943	0.943	0.989	**0.983**	**0.982**	**0.996**	**0.945**	**0.945**	**0.994**

TABLE 6: Voting between boosted J48 and tree family algorithms.

Techniques	Algorithm name	Dataset		
		Accuracy	F-measure	ROC area
	BFTree + Adaboost_J48	0.950	0.950	0.996
	CDT + Adaboost_J48	0.973	0.972	0.996
	ForestPA + Adaboost_J48	0.980	0.980	0.997
Combination of classifiers	LADTree + Adaboost_J48	0.983	0.982	0.997
	LMT + Adaboost_J48	**0.983**	**0.982**	**0.998**
	NBTree + Adaboost_J48	**0.985**	**0.985**	**0.998**
	RandomForest + Adaboost_J48	0.980	0.980	0.997

+ means using voting method of the bagging technique.

with consistent results after applying the 10 cross-validation method of 377 correctly classified instances of 400 instances, which is 94.3%. The consistent results of the tree family relate to the feature selection criteria that are used to select the relevant features for building the tree model. The most correlated feature is selected to be at the root of the tree such as visiting resources, whilst the other features are selected subsequently based on their effectiveness till the leaf nodes that represent the decision classes. The analysis of ML techniques and examining their effects to predict the student performance leads us to answer the research question *RQ2*. The question studies the ability of ML techniques in EDM deployed in the academic institution systems. Hence, it is better to find the most promising and superior technique in a specific school.

4.3. Ensemble Meta-Based Tree Model. The proposed EMT model essentially combines two consistent ML techniques

into a voting bagging technique. The idea comes from aggregating the most effective algorithms into one technique that would enhance the prediction models. The EMT technique performs as an ensemble method to get more accurate prediction results than those from traditional learning methods that focus on one algorithm learner [34, 35].

Table 5 shows the results of using the best classifiers of different families alone and as functions in boosting and bagging methods. As shown, the classifiers in the boosting method reveal a high accuracy in comparison with the bagging method using accuracy, F-measure, and ROC area metrics. In particular, the J48 classifier obtains consistent results when it is used as a function in boosting method dubbed Adaboost_J48, which is increased by 4.1% from 0.943 to 0.983. Consequently, using the J48 classifier as a function in boosting method is the first function of the EMT model to be combined with the best classifier overall the other classification models.

CCI = LL (104) + MM (158) + HH (108)
ICC = LM (6) + LH (0) + ML (7) + MH (8) + HL (0) + HM (9)
Accuracy = CCI/(CCI + ICC) = 370/(370 + 30) = 0.925

Confusion matrix to analyze NB-Tree model				
Confusion matrix	Target (actual results)			NB-Tree
	L	M	H	
Model (predictive results) L	104	6	0	
M	7	158	8	
H	0	9	108	
CCI	ICI	Accuracy	Precision	Recall
370	30	0.925	0.925	0.925

(a)

CCI = LL (109) + MM (154) + HH (114)
ICC = LM (1) + LH (0) + ML (2) + MH (17) + HL (0) + HM (3)
Accuracy = CCI/(CCI + ICC) = 377/(377 + 23) = 0.942

Confusion matrix to analyze J48 model				
Confusion matrix	Target (actual results)			J48
	L	M	H	
Model (predictive results) L	109	1	0	
M	2	154	17	
H	0	3	114	
CCI	ICI	Accuracy	Precision	Recall
377	23	0.943	0.946	0.943

(b)

CCI = LL (110) + MM (169) + HH (115)
ICC = LM (0) + LH (0) + ML (2) + MH (2) + HL (0) + HM (2)
Accuracy = CCI/(CCI + ICC) = 394/(394 + 6) = 0.985

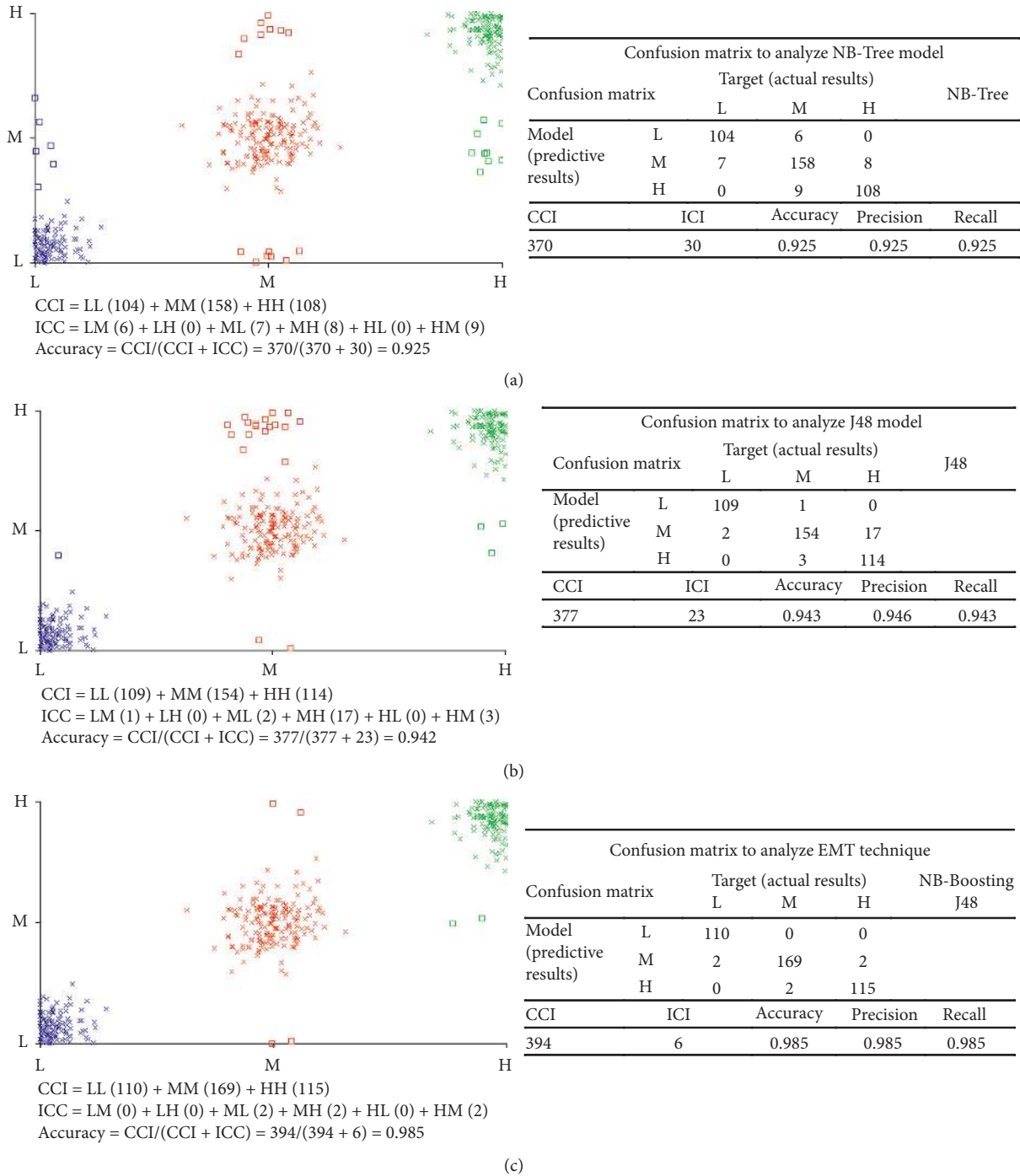| Confusion matrix to analyze EMT technique				
Confusion matrix	Target (actual results)			NB-Boosting J48
	L	M	H	
Model (predictive results) L	110	0	0	
M	2	169	2	
H	0	2	115	
CCI	ICI	Accuracy	Precision	Recall
394	6	0.985	0.985	0.985

(c)

FIGURE 3: Two classifiers and their combination in the EMT technique: (a) NB-Tree model; (b) J48 model (EMT model); (c) the EMT technique.

Combining multiple effective classifiers instead of using a single classifier into an accurate prediction model would obtain a high prediction performance of the students [35]. Thereupon, the EMT model ensembles a set of more effective classifier models that are derived from aggregating the first function of EMT model with the best tree-based classifier using voting method of bagging technique as shown in Table 6. The results are ranged between 0.95 for BFTree voting with Adaboost_J48 and 0.985 when ensemble NBTree voting with Adaboost_J48. The NBTree algorithm shows the robustness of integrating the Naïve Bayes technique as a decision tree on increasing the performance metric.

Figure 3 describes the results regarding the performance of students at the academic institution. In Figure 3(a) and based on the results, the accuracy score of NBTree classifier is 0.925. The results in Figure 3(b) represent the output of the J48 algorithm. Because of the increase in the numbers of accurate results of this figure, the total accuracy for the second figure reached to 0.943. Figure 3(c) represents the confusion matrix and the performance of the EMT model.
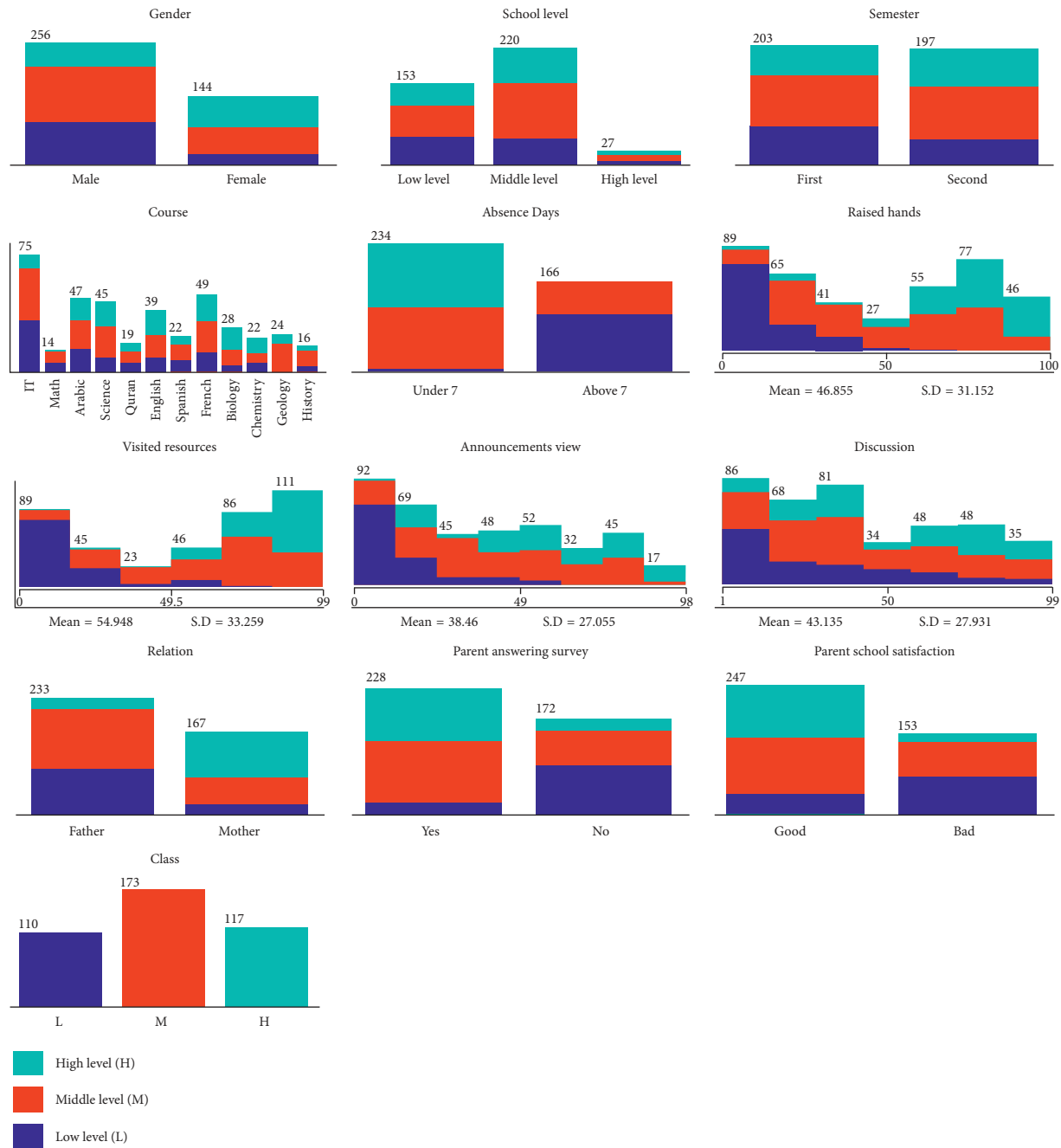
FIGURE 4

The accuracy of this figure increases to 0.985. This indicates an important fact regarding the improvement of the actual results and the tangible improvement in the performance of students. The proposed figure of this research points out to the development in the accuracy of student performance and the prediction, which is adopted after the boosting of the algorithms J48 and NBTree together. This, in turn, leads to the highest accuracy results to 0.985, unlike the observed accuracy results which are recorded after the use of the J48 algorithm or boosting method alone.

The question here is around how valuable is it to use the EMT model as a promising method for predicting the performance of students. The answer is proven, as shown in Table 6 and Figure 3, which particularly leads to confirm *RQ3*. The main purpose of the work here is to propose more effective and accurate prediction models to enhance the quality of educational institutions by deploying the models. This would be used to predict potential students who achieve low performance and understand why they achieved such results. Thus, the EMT prediction model and its results would help instructors, administrators, and policymakers to make a better decision especially for those who are expected to achieve low-level performance [5, 36]. The decision aims to reduce the number of expected potential students who

A matrix 3 × 3 to compare actual results and model results				
	Algorithm's family			
Confusion matrix	Target (actual results)			Algorithm name
	L	M	H	
Model (predictive results) L	LL	LM	LH	
M	ML	MM	MH	
H	HL	HM	HH	
Correctly classified instances		Incorrectly classified instances		Accuracy
CCI = LL + MM + HH		ICC = LM + LH + ML + MH + HL + HM		CCI/(ICC+CCI)

Confusion matrix to analyze PART algorithm
Algorithm's family: rules

Confusion matrix		Target (actual results)			Algorithm name PART
		L	M	H	
Model (predictive results)	L	108	2	0	
	M	3	151	19	
	H	0	9	108	
CCI		ICI		Accuracy	
367		33		0.918	

Confusion matrix to analyze A2DE algorithm
Algorithm's family: bayes

Confusion matrix		Target (actual results)			Algorithm name A2DE
		L	M	H	
Model (predictive results)	L	101	9	0	
	M	10	151	12	
	H	0	11	106	
CCI		ICI		Accuracy	
358		42		0.895	

Confusion matrix to analyze multilayer perceptron algorithm
Algorithm's family: function

Confusion matrix		Target (actual results)			Algorithm name multilayer perceptron
		L	M	H	
Model (predictive results)	L	102	8	0	
	M	7	153	13	
	H	0	8	109	
CCI		ICI		Accuracy	
364		36		0.910	

Confusion matrix to analyze LocalKnn algorithm
Algorithm's family: lazy

Confusion matrix		Target (actual results)			Algorithm name localKnn
		L	M	H	
Model (predictive results)	L	106	4	0	
	M	8	158	7	
	H	0	10	107	
CCI		ICI		Accuracy	
371		29		0.928	

Confusion matrix to analyze J48 algorithm
Algorithm's family: trees

Confusion matrix		Target (actual results)			Algorithm name J48
		L	M	H	
Model (predictive results)	L	109	1	0	
	M	2	154	17	
	H	0	3	114	
CCI		ICI		Accuracy	
377		23		0.943	

Figure 5

5. Conclusion and Future Work

The educational data mining (EDM) uses analytical tools from the data mining field that are exploited to explore the unique types of the dataset in the academic field. The tools convert educational system data that are not readable by a human into interpretable and valuable information that may possibly have a big impact on the educational research. In this paper, the EDM field is used to predict students' academic performance. The contributions are maintained into three folds which include (i) examining a set of features that reflect the performance of students using Pearson correlation, (ii) evaluating a set of learning models using ML techniques where these techniques use specifically a dataset of 400 instances collected in two semesters to be used in the construction phase (the dataset includes thirteen attributes, from a registration office, having preregistration students' information along with student performance as low, middle, and high), and (iii) proposing a prediction model for evaluating student performance on the same dataset, called EMT model. The EMT model is constructed by combining the most effective techniques studied from 47 learning techniques.

might achieve low-level performance and develop the learning system by adding new tools that improve the students' academic performance [5].

However, the results show strong correlations between the independent features and the final achievement of students ranging between 0.436 and 0.06, which leads to answer *RQ1*. Based on the 10-fold cross-validation method, a comparison between different ML techniques is conducted. Herein, a set of evaluation metrics are used to evaluate these techniques, which are accuracy, F-measure, and ROC metrics. The results found that PART, A2DE, multilayer perceptron, LocalKnn, and J48 algorithms have accuracy values of 91.8%, 89.5%, 91%, 92.8%, and 94.3%, respectively, which were selected from different five families of classifiers which ensure the validity of *RQ2*. Finally, based on the previous ML results, the J48 was selected to achieve the main goal which is *RQ3*. Moreover, experimental results were conducted to enhance the results on the best-selected classifier J48 by applying the ensemble method and voting the results with the tree family algorithms. The results have shown a significant improvement using the proposed EMT model algorithm. It has been proved that the proposed algorithm can achieve an accuracy of up to 98.5%; as a result, the *RQ3* is consequently achieved.

The main purpose of the work is to improve the quality of education institutions. The improvements are achieved by deploying the proposed predictive model that would be used to predict the performance of the students for especially those having low performance and understanding the reasons behind such results. As a result, the model provides the academic institute, including its management and teachers, with an accurate evaluation of their students at an early stage of the learning process to prevent them from getting a low result. Also, it helps to better allocate both resources and staff that result in increasing the effectiveness in education development.

In future work, the research suggests using more features such as examining of how the use of social media or babysitting would affect the performance of the students. In addition, extra experiments could be conducted using other data mining techniques such as clustering.

Appendix

A. The Description of the Dataset

In this study, 400 students' records with 13 features (or attributes) were used as a dataset. The dataset as shown in Figure 4 illustrates the definitions of features and their distributions with range values and types.

B. The Confusion Matrix Results for the Best Five Classifiers

In this part, confusion matrix is adopted as a tool to analyse the best five classifiers of each family by comparing the actual and the target outputs. The results shown in Figure 5 indicate that the best classifiers are PART, A2DE, multilayer perceptron, LocalKnn, and J48 algorithms with correctly classified instances as 367, 358, 364, 371, and 377, respectively.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Kalaivani, B. Priyadharshini, and B. S. Nalini, "Analyzing student's academic performance based on data mining approach," *International Journal of Innovative Research in Computer Science and Technology*, vol. 5, no. 1, pp. 194–197, 2017.

[2] A. Daud, N. R. Aljohani, R. A. Abbasi, M. D. Lytras, F. Abbas, and J. S. Alowibdi, "Predicting student performance using advanced learning analytics," in *Proceedings of the 26th International Conference on World Wide Web, Companion*, Perth, Australia, April 2017.

[3] S. Roy and S. N. Singh, "Emerging trends in applications of big data in educational data mining and learning analytics," in *Proceedings of 2017 7th International Conference on Cloud Computing, Data Science and Engineering-Confluence*, Noida, India, January 2017.

[4] A. M. Ahmed, A. Rizaner, and A. H. Ulusoy, "Using data mining to predict instructor performance," *Procedia Computer Science*, vol. 102, pp. 137–142, 2016.

[5] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Computers and Education*, vol. 113, pp. 177–194, 2017.

[6] J. A. Larusson and B. White, *Learning Analytics: From Research to Practice*, Vol. 13, Springer, Berlin, Germany, 2014.

[7] M. Wook, Z. M. Yusof, and M. Z. A. Nazri, "Educational data mining acceptance among undergraduate students," *Education and Information Technologies*, vol. 22, no. 3, pp. 1195–1216, 2016.

[8] R. S. J. D. Baker and K. Yacef, "The state of educational data mining in 2009: a review and future visions," *JEDM Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.

[9] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, 2010.

[10] E. A. 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.

[11] M. Mayilvaganan and D. Kalpanadevi, "Comparison of classification techniques for predicting the performance of students academic environment," in *Proceedings of 2014 International Conference on Communication and Network Technologies (ICCNT)*, Sivakasi, India, December 2014.

[12] C. Tucker, B. K. Pursel, and A. Divinsky, "Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes," *Computers in Education Journal*, vol. 5, no. 4, pp. 84–95, 2014.

[13] D. M. D. Angeline, "Association rule generation for student performance analysis using apriori algorithm," *The SIJ Transactions on Computer Science Engineering and its Applications (CSEA)*, vol. 1, no. 1, pp. 12–16, 2013.

[14] T. M. Christian and M. Ayub, "Exploration of classification using NBTree for predicting students' performance," in

Proceedings of 2014 International Conference on Data and Software Engineering (ICODSE), Bandung, Indonesia, November 2014.

[15] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, "Improving accuracy of studentsâ€™ final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique," *Decision Analytics*, vol. 2, no. 1, p. 1, 2015.

[16] K. F. Li, D. Rusk, and F. Song, "Predicting student academic performance," in *Proceedings of 2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, Taichung, Taiwan, July 2013.

[17] S. Natek and M. Zwilling, "Student data mining solution-knowledge management system related to higher education institutions," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6400–6407, 2014.

[18] Z. K. Papamitsiou, V. Terzis, and A. A. Economides, "Temporal learning analytics for computer based testing," in *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, Indianapolis, IN, USA, March 2014.

[19] U. Bin Mat, N. Buniyamin, P. M. Arsad, and R. Kassim, "An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention," in *Proceedings of 2013 IEEE 5th Conference on the Engineering Education (ICEED)*, Kuala Lumpur, Malaysia, December 2013.

[20] E. Osmanbegović and M. Suljić, "Data mining approach for predicting student performance," *Economic Review*, vol. 10, no. 1, pp. 3–12, 2012.

[21] V. Ramesh, P. Parkavi, and K. Ramar, "Predicting student performance: a statistical and data mining approach," *International Journal of Computer Applications*, vol. 63, no. 8, pp. 35–39, 2013.

[22] M. M. A. Tair and A. M. El-Halees, "Mining educational data to improve students' performance: a case study," *International Journal of Information*, vol. 2, no. 2, pp. 140–146, 2012.

[23] P. M. Arsad and N. Buniyamin, "A neural network students' performance prediction model (NNSPPM)," in *Proceedings of 2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, Kuala Lumpur, Malaysia, November 2013.

[24] A. Bogarín, C. Romero, R. Cerezo, and M. Sánchez-Santillán, "Clustering for improving educational process mining," in *Proceedings of Fourth International Conference on Learning Analytics and Knowledge*, Indianapolis, IN, USA, March 2014.

[25] C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy, "Visualizing patterns of student engagement and performance in MOOCs," in *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, Indianapolis, IN, USA, March 2014.

[26] S. Parack, Z. Zahid, and F. Merchant, "Application of data mining in educational databases for predicting academic trends and patterns," in *Proceedings of 2012 IEEE International Conference on Technology Enhanced Education (ICTEE)*, Amritapuri, India, January 2012.

[27] T. Mishra, D. Kumar, and S. Gupta, "Mining students' data for prediction performance," in *Proceedings of 2014 Fourth International Conference on Advanced Computing and Communication Technologies (ACCT)*, Rohtak, India, February 2014.

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

[29] G. Gray, C. McGuinness, and P. Owende, "An application of classification models to predict learner progression in tertiary education," in *Proceedings of 2014 IEEE International Advance Computing Conference (IACC)*, Gurgaon, India, February 2014.

[30] I. Hidayah, A. E. Permanasari, and N. Ratwastuti, "Student classification for academic performance prediction using neuro fuzzy in a conventional classroom," in *Proceedings of 2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, Yogyakarta, Indonesia, October 2013.

[31] K. Bunkar, U. K. Singh, B. Pandya, and R. Bunkar, "Data mining: prediction for performance improvement of graduate students using classification," in *Proceedings of 2012 Ninth International Conference on Wireless and Optical Communications Networks (WOCN)*, Indore, India, September 2012.

[32] H. W. Ian, F. Eibe, A. H. Mark, and J. P. Christopher, *Data Mining, Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers Inc, Burlington, MA, USA, 4th edition, 2016.

[33] M. Atherton, M. Shah, J. Vazquez, Z. Griffiths, B. Jackson, and C. Burgess, "Using learning analytics to assess student engagement and academic outcomes in open access enabling programmes," *Open Learning: The Journal of Open, Distance and e-Learning*, vol. 32, no. 2, pp. 119–136, 2017.

[34] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Burlington, MA, USA, 2016.

[35] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, CRC Press, Boca Raton, FL, USA, 2012.

[36] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," *Procedia Computer Science*, vol. 72, pp. 414–422, 2015.