Ensemble Approach for Predicting Student's Academic Performance

L. A. Muhammad, M. S. Argungu

***Abstract*—**Educational data mining (EDM) has recorded substantial considerations. Techniques of data mining in one way or the other have been proposed to dig out out-of-sight knowledge in educational data. The result of the study got assists academic institutions in further enhancing their process of learning and methods of passing knowledge to students. Consequently, the performance of students boasts and the educational products are by no doubt enhanced. This study adopted a student performance prediction model premised on techniques of data mining with Students' Essential Features (SEF). SEF are linked to the learner's interactivity with the e-learning management system. The performance of the student's predictive model is assessed by a set of classifiers, viz. Bayes Network, Logistic Regression, and Reduce Error Pruning Tree (REP). Consequently, ensemble methods of Bagging, Boosting, and Random Forest (RF) are applied to improve the performance of these single classifiers. The study reveals that the result shows a robust affinity between learners' behaviors and their academic attainment. Result from the study shows that the REP Tree and its ensemble record the highest accuracy of 83.33% using SEF. Hence, in terms of the Receiver Operating Curve (ROC), boosting method of REP Tree records 0.903, which is the best. This result further demonstrates the dependability of the proposed model.

***Keywords*—**Ensemble, bagging, Random Forest, boosting, data mining, classifiers, machine learning.

I. Introduction

EDUCATION has been recognized to be a major concrete cornerstone among other stones for national development. It is also recognized to be the most important component of human resources development and is accorded a pride of place in many countries' national development activities. There is no doubt that the importance of education cannot be belittled because there is no country that can succeed without educating its people. Reference [9] opined that education helps to improve security, health, prosperity, ecological balance in the world, and other national achievements. The progress of any country could be rated by the quality of the educational system citizens are exposed to. Reference [10] explained that most of the sectors of a nation have been experiencing dramatic changes in running their day-to-day activities and these innovation changes do not exclude the educational sector in its functioning. Most of the sectors in several nations have invested much in information and communication technologies to ease their duties, and this does not leave the education sector out. The rise or advent of information and communication technologies in higher education called for changes in the medium students receive knowledge and also affected the way tutors teach the students from traditional to modern methods [10].

Despite the advent of these technologies into the field of education, still, the field is faced with several challenges, and one of the challenges that face the field of education is the student's academic failure and dropout. Reference [10] also revealed that several higher institutions that have inculcated ICT into their academic system, collected massive volume of data from students through their Learning Management System, and these data are available in different digital formats: files, documents, sound, records, scientific, video data, and many other data formats. However, the reasonable way for universities to convert or transform this huge volume of data into meaningful information or knowledge for efficient decision-making over the student is havoc [10]. The use of these gigantic student data collected to predict academic performance is one of the hazardous issues faced by the educational sector. However, it was submitted in [8] that the educational authorities and institutes are putting in more effort, working hard to minimize the failure ratio of the student by predicting the performance through the obtained data. Predicting students' academic performance early in their stay in higher institutions is one of the tremendous solutions to curb students' failure or drop-out. It was revealed that students predicted with high academic performance will not have any problems or issues from completing their studies well [8]. Nevertheless, predicting academic performance is not an easy task to carry out, because it involves huge data which cannot be treated manually and this requires an automated system. To this end, this called for machine learning techniques to build a model to automatically predict students' academic performance.

However, there are many approaches that the advent of information technologies has brought into the education sector which proposed to predict students' academic performance and data mining is one of the most popular techniques embraced to evaluate and analyze students' performance. EDM is focused on developing, researching, and applying computerized procedures to detect patterns in extensive numbers of data in education which in the other way difficult to analyze due to the enormous volume of data within which they exist [10]. Moreover, EDM has emerged as a research area in recent years aimed at analyzing the unique kinds of data that arise in educational settings to resolve educational research issues.

L. A. Muhammad is with the Department of Information Technology, Nigeria Correctional Services, National Headquarters Abuja, Nigeria (phone: +234 8036105857; e-mail: waleelawal01@gmail.com).

M. S. Argungu is with the Department of Computer Science, Kebbi State University of Science and Technology Aliero, Kebbi State, Nigeria (corresponding author, phone: +2348100543186 e-mail: sm279arg@gmail.com).

Complex machine learning problems can be grouped into four core problem types: Classification, Regression, Clustering, and Rule extraction.

Classification and prediction are the same except that prediction return continuous numerical value or score while classification returns a categorical value. Classification is a supervised learning technique (i.e. the class labels are known before the task) [4]. There are several classification learning schemes to implement classification or prediction: decision trees, backtracking, probability, logistic regression, and many others that are highly embraced to mine educational data [10]. The concept is that when any of these mentioned approaches are deployed to carry out classification as a lone classifier, it is referred to as a single classifier [4]. The idea is that all these single classifiers are attached to one weakness which mostly affects their model performances. The limitations attached to the performance of the use of single classifiers, hence, pave way for the introduction of the ensemble method of classification. The ensemble approach performs its operations by combining the results of several individual models to improve the reliability and performance of the model. Ensemble techniques have been adopted and popular among academicians, developers, and researchers for predictive modeling. There are various ensemble methods such as Bagging, Boosting, RF, Stacking, Voting, and so on [14].

Bagging as an ensemble method works by selecting tuples randomly (this principle or process is called Bootstrap Aggregating) while developing the model. The models are built simultaneously and the average is taken to form the overall decision (result) of the ensemble model. It thus decreases variance but it is prone to bias. Reference [14] studied the efficiency of the bagging method and it was submitted that bagging efficiency is relatively high. Boosting, on the other hand, improves bagging. The model is developed sequentially, the wrongly classified tuples by the previous classifier would now be assigned more weight to receive more attention in the next classifier in the series. In the end, the weighted average is taken to form the final decision. Boosting is susceptible to overfitting. Reference [16] reported in their study that boosting method is effective and efficient in prediction or classification. RF is an improvement on bagging. This is because the principle of Bootstrap Aggregating is extended to features. This method of the RF was used by Han [5] and the result has it that the method is indeed very efficient to predict.

Computing focuses much on timing, accuracy, and performance and among all. Different studies such as [2], [1], [14], and so on have modeled systems using both single classifiers and ensemble methods to predict students' academic performance. However, the model worked well to some level, there is more chance to help their models since higher institutions are dealing with many student data. Still, it was observed that there are still areas that can be further explored. Thus, this study attempts to model an automated student academic performance system on different learning schemes of the decision tree, logistic regression, backtracking, and probability upon various ensemble methods to predict students' performance for comparison and accuracy purposes.

Predicting students' academic performance in higher institutions is the main objective of this study to detect students that are at risk in their studies. However, to compare the performance of various learning schemes using an ensemble approach in predicting the performance of students in both training datasets and test datasets, this study is useful.

The above purpose could be achieved based on the following specific objectives: Design the three base classifiers (single classifiers) on the student's datasets (train and test), Model the base classifiers unto an ensemble based on Bagging, Boosting, and RF (train and test), Implement the proposed ensemble models, Evaluate the performances of the implemented models with the use of confusion matrix table i.e. accuracy, Error rate, ROC, etc.

II. Methodology

A. Proposed Model

This section presents the methodology adopted in carrying out this study. This is briefly represented in Fig. 1 [13].

B. Acquisition of Data Set

The dataset provided to experiment with this study was retrieved from UCI online machine learning repository and it is readily available for data mining. This dataset was systematically gathered from students through a Learning Management System (LMS) called Kalboard 360. The dataset is available at www.uci.com. The dataset contains records of 300 students with 22 attributes. It was released in 2018, for classification-associated tasks. A brief overview of the dataset is thus presented in Table I [6].

C. Data Preprocessing (Feature Selection)

As displayed in Fig. 1, the experiment started with the acquisition of the dataset, followed by preprocessing phase, at this phase feature selection is required. This eliminated all redundant and irrelevant features that the dataset carried. This study adopted Correlation Based Features Selection (CFS) to carry out the feature selection technique. On the WEKA platform, CFS is implemented as Correlation Attribute Evaluator on WEKA [6].

The same WEKA (Waikato Environment for Knowledge Analysis) was used for the development of ensemble models. There are several machine learning platforms such as WEKA, R, MATLAB, SPSS, and so on. WEKA was picked for this study because it has an updated algorithm coupled with its GUI friendly compared to "R".

Many mining tasks are implemented and they are readily available for use in WEKA. The implementation of these algorithms covers data preprocessing, classification, regression, clustering, and association rules; and a visualization facility is also embedded. But in this work, WEKA is employed to design and build the single classifiers of Bayes Network, RepTree, and Logit Boost. Consequently, the ensemble models of Bagging, Boosting, and RF were built with the aforementioned algorithms as base classifiers.

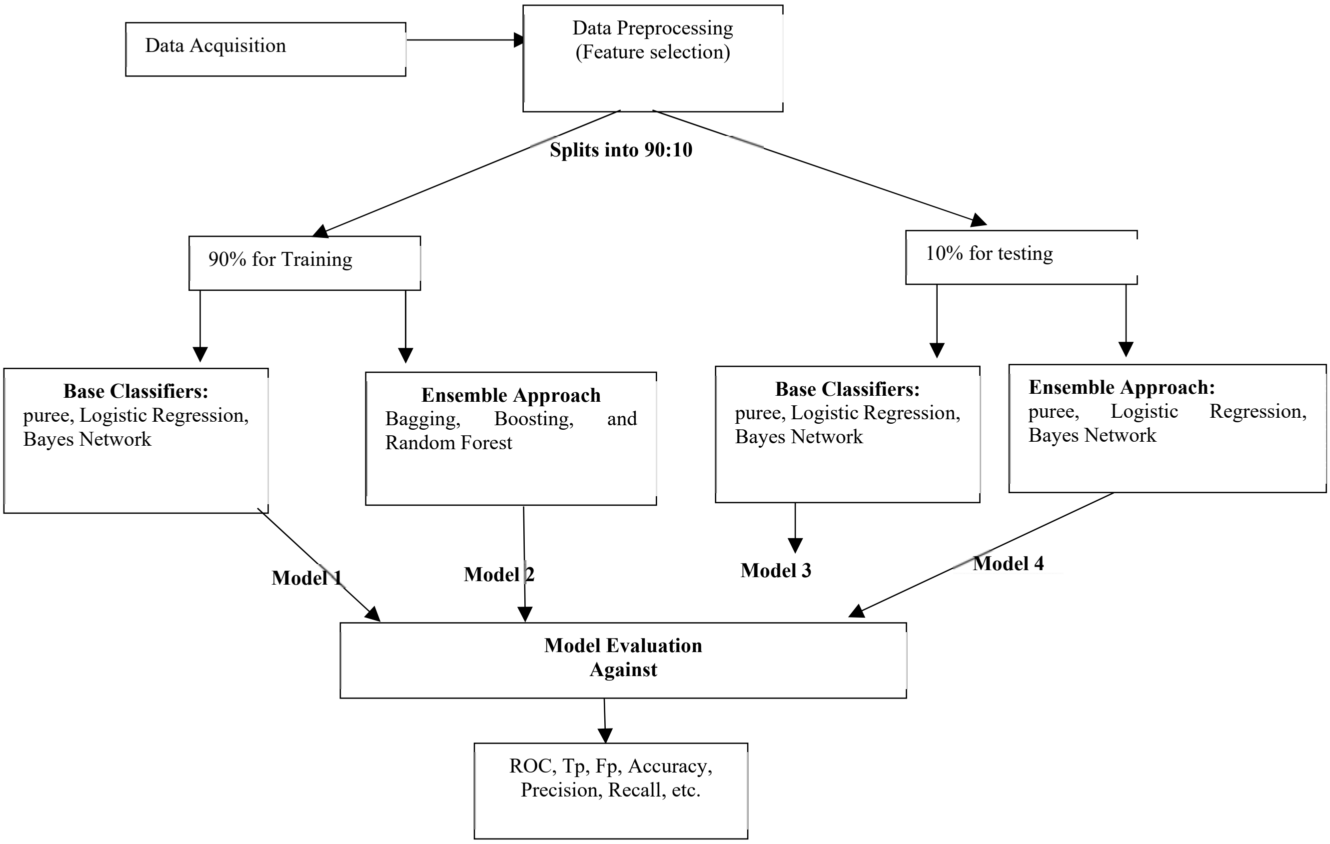

Fig. 1 Proposed model

TABLE I
ACQUISITION OF DATA SET

S/N	Attributes	Details
1	ge	M = 72, F=59
2	CST	G = 49, ST = 20, SC = 4, OBC = 57, MOBC = 6
3	tnp	Best = 9, Vg = 38, Good = 59, Pass = 25, Fail = 0
4	twp	Best = 5, Vg = 44, Good = 65, Pass = 17
5	iap	Best = 8, Vg = 63, Good = 53, Pass = 7, Fail = 0
6	esp	Best = 8, Vg = 42, Good = 54, Pass = 27, Fail = 0
7	arr	Y = 53, N = 78
8	ms	Married = 0, Unmarried = 131
9	ls	T = 39, V = 92
10	as	Free = 55, Paid = 76
11	fmi	Vh = 6, High = 15, Am = 27, Medium = 63, Low = 20
12	fs	Large = 2, Average = 40, Small = 89
13	fq	Il = 20, Um = 40, 10 = 23, 12 = 22, Degree = 27, Pg = 29
14	mq	Il = 27, Um = 52, 10 = 25, 12 = 17, Degree = 7, Pg = 3
15	fo	Service = 38, Business = 34, Retired = 3, Farmer = 27, Others = 29
16	Mo	Service = 12, Business = 1, Retired = 1, Housewife = 115, Others = 2
17	nf	Large = 58, Average = 43, Small = 36
18	sh	Good = 27, Average = 59, Poor = 45
19	ss	Govt = 91, Private = 40
20	Me	Eng = 62, Asm = 60, Hin = 7, Ben = 2
21	tt	Large = 10, Average = 43, Small = 78
22	atd	Good = 56, Average = 47, Poor = 28

D.Facts and Proposed Configuration Parameters in Building the Model

WEKA is a machine learning tool, in which all the aforementioned algorithms are by default set with some parameters. Thus, these parameters could be changed to conform to the given scenario. The changes of these parameters are done towards an improvement on the accuracy of algorithms, therefore if changing of parameters does not enhance the accuracy of techniques, the purpose is defeated. The study followed the array of the configuration setting as depicted below.

- Firstly, for feature selection:
  – The search method is the best first algorithm, the attribute selection mode is cross-validation, the number of threads is 1, and the pool size is 1.
- Secondly, for Base classifiers:
  – Bayes Network: Bayes Network learning uses various search algorithms and quality measures in which simple estimator is used for estimating the conditional probability tables of a Bayes network once the structure has been learned. This Bayes Network learning algorithm used a hill climbing algorithm restricted by an order of the variables and the score type is ENTROPY. A batch size of 100 is retained.
  – Simple Logistic: It is a classifier for building linear logistic regression models. Logistic Boost with simple regression functions as a base and learners are used for fitting the logistic models. The optimal number of Logit Boost

iterations to perform is cross-validated, which leads to automatic attribute selection. The heuristic stop is 50 while the batch size is retained at 100.

– REP Tree: Fast decision tree learner builds a decision/ regression tree using information gain/variance and prunes it using reduced-error pruning (with back fitting), only sorts values for numeric attributes once. Missing values are dealt by splitting the corresponding instances into pieces (i.e. as in C4.5). Also, the maximum depth is -1, the seed is 1, the minimum number is 2, the minimum variance probability is 0.001, the number of folds is 3 and the batch size is 100.

- Thirdly, for Ensemble Method [13]:

– Boosting (AdaBoost): This is the class for boosting a nominal class classifier using the Adaboost M1 method. Only nominal class problems can be tackled. AdaBoost often dramatically improves performance, but sometimes overfits. Also, the weight threshold is 100, the batch size is 100, the seed is 1, and the number of iterations is 10. The base classifiers are Simple Logistic, REP Tree, and Bayes Network. The base classifiers would be used separately.

– Bagging: This is the class for bagging a classifier to reduce variance, do classification and regression depending on the base learner. Also, the number of execution slots is 1, the batch size is 100, the seed is 1, and the number of iterations is 10. The base classifiers are Simple Logistic, REP Tree, and Bayes Network. The base classifiers would be used separately.

– Random Tree: It is a class for building a tree that considers Kalboard arbitrarily selected attributes at each node. In short, it is a class for constructing a forest of random trees. Random Tree also can do classification and regression depending on the base learner. Also, the maximum depth is 0, the number of execution slots is 1, the batch size is 100, the seed is 1, the bag size percent is 100, and the number of iterations is 100.

Assessment of the effectiveness of the ensemble frameworks is done as follows. Each one of the models would be trained using a stratified ten-fold validation test mode in a Weka environment to test and evaluate the algorithms, which would use 10-fold cross-validation. In this process, the dataset would be divided into 10 subsets. In each iteration, one of the 10 subsets is applied as the sampling tuple and the other n-1 subsets form the later-to-be-used set for training still. Statistics of performance are computed across all 10 rounds. This presents a good indication of how well the classifier will achieve on concealed data. 90% of the data set is used in training or building the models while the remaining 10% is used to test the models.

E. Performance Evaluation and Measurement Terms

The evaluation of the models is done using the confusion matrix values as basic measurement parameters. Its derivatives are also used. This section presents the measurement parameters.

F. Confusion Matrix and Derivatives

- True Positive Rate: This returns the proportion of actual positives which are accurately classified in Immunotherapy/Cryotherapy data sets. It is calculated by dividing correctly classified instances by the total instances of the YES class label.
- False Positive Rate: This computes the proportion of actual negatives which are incorrectly identified. It is calculated by dividing negatives classified correctly by total negatives.
- Precision: These are the positive predictive values. The calculation is true positives divided by all positive results.
- Recall: The Recall is computed as true positives divided by true positive and false negative results.
- F-Measure: This term works by blending recall and precision scores into a unit measure of performance.

Apart from true positive rate to receiver operating characteristic area, others are based on correctly, incorrectly classified, and unclassified instances. The unclassified instances show the percentage of instances that were not classified. The percentage of correctly classified instances is often called the accuracy of sample accuracy.

The steps below show how to obtain value for accuracy:

$$\text{Sensitivity} = \frac{(\text{tpos})}{(\text{pos})} \tag{1}$$

$$\text{Specificity} = \frac{(\text{tneg})}{(\text{neg})} \tag{2}$$

where tpos is the number of true positives i.e., the correctly classified instances, pos is the number of positive instances, tneg is the number of true negatives i.e. the correctly classified instances, neg is the number of negative tuples, and pos is the number of false positives i.e. the wrongly classified instances. Therefore, it can be shown that accuracy is a function of sensitivity and specificity:

$$\text{Accuracy} = \left[\text{Sensitivity} \times \left(\frac{\text{pos}}{(\text{pos+neg})}\right)\right] + \left[\text{Specificity} \times \left(\frac{\text{neg}}{(\text{pos+neg})}\right)\right] \tag{3}$$

III. Related Work

Reference [12] investigated the high rate of failure among students of higher institutions. Three algorithms of data mining techniques were proposed and implemented to predict academic performance. At the end of the study, it was revealed that based on the algorithm proposed and implemented, C4.5 (known as J48 in the WEKA tool) has the highest prediction accuracy. However, the number of instances used for the study is considered minimal, the study could have considered larger instances to improve prediction accuracy

The study by [17] was to compare different data mining algorithms. The study used Cross Industry Standard Process (CRISP) data mining method, and the results showed that Naïve Bayes returns better prediction accuracy. However, it would have been better if the study juxtaposes other algorithms such as algorithms of other learning schemes or ensemble methods.

Reference [3] examined the hidden cause of students' failure to get employment after leaving school. With the deployment of data mining techniques, the study was able to monitor the academic performance of students. It was revealed that Naïve Bayes has a higher prediction accuracy of 65%. The study could have explored other algorithms such as algorithms of other learning schemes or ensemble methods.

The study conducted by [17] was carried out to unveil the causes of the poor learning attitude of students in the school. The study focused on the comparison of the performances of two data mining algorithms to predict student learning based on student records (data set). After the experiment, the result showed that the average percentage of both classifiers was above 60%, whereas Naïve Bayes has a higher precision average. However, ensemble methods could have been deployed to further solidify the study.

Reference [7] conducted a study to identify slow, average, and fast learners among students. The study employed the techniques of data mining to predict the academic performance of the student. Naive Bayes, J48, Zero, and Random Tree models were trained and tested on the dataset. In the end, the result showed that RF has higher accuracy than the other three algorithms. However, ensemble methods could have been deployed further to boost the prediction accuracy.

The study by [1] was conducted to predict the academic performance of the newly admitted student. Three classification techniques of machine learning were proposed and implemented to model the new data attributes features obtained from the learner's interactivity e-learning management system. Sequence to this, the result showed that learners' behavior has an impact on students' academic performance. However, further studies are required to validate this study.

The study by [11] focused on a comparative analysis of EDM techniques. The results from the study showed that MLP, a neural network-based classification shows the best result of 74.8% compared to C4.5, Naïve Bayes, ID3, and CART. However, the study could have explored ensemble for comparison purposes.

The aim of [15] was to implement a classification technique in web usage mining to a financial institution that may assist the industry to identify web performance issues. The study proposed and implemented a k-nearest neighbor algorithm with standardized Euclidean distance to classify frequent access patterns. The result showed that implementation of algorithms for k-nearest neighbor is possible using web usage mining and it may probably assist the company to find interesting knowledge. The study could have explored other algorithms for comparison purposes.

Reference [16] Investigated the ensemble methods of Adaboost, Bagging, Dagging, and Grading in the prediction of students' academic performance. The results from the study revealed that the Boosting method with the Adaboost algorithm turned out the highest accuracy. However, other base classifiers of different learning schemes could be explored to further expand the comparison scope.

Reference [2] proposed and implemented a neural network called Deep Neural Network (DNN). The aim was to show the categorical class each student belongs to. Comparing the model with the existing studies that used the same dataset, the accuracy and outperforms were achieved. This study affords the institution of higher learning to offer a remedy to the potentially failing students. However, this study could have juxtaposed the results with other outstanding algorithms and methods.

This study proposed an ensemble of RF, boosting, and base classifiers to ameliorate the prediction of students' academic performance. The model considered Bayes Network and logistic regression as base classifiers. Consequently, WEKA has the capacity of ensemble the new method of Boosting, together with the RF act as open source tool used to apply for improving the higher performance of this single classifier. The finding reveals that there is significant robust affinity among the two, leaners' behaviors and their academic attainment. Having considered the significant studies done so far related to the prediction of student performance, it is noted that the benefit of using approaches might have not been explored yet. Specifically, other possible algorithms of ensemble methods might have not been employed and other attribute selection techniques are not been fully explored. Thus, this study investigates the effectiveness of Logistic Regression, REP Tree, and Bayes Network as base learners in the ensemble methods of Bagging and Boosting for predicting students' performance after performing feature selection.

IV. Result and Discussion

The effective features that are selected after applying CFS are presented. The results from the various classifiers' prediction and their corresponding ensemble methods are shown and discussed.

A. Effective Features

It is agreed that it is not all the attributes contained in the dataset will be used for the prediction, that is the main reason CFS is applied to decide the most important attributes in predicting student performance. Table II revealed the outcome of the attribute selected based on the application of CFS on the dataset. The search method of best first was used. Forward Search direction was employed in the search. A stale search after 5 node expansions was recorded. The total number of subsets evaluated stood at 196. The merit of the best subset found set at 0.182

TABLE II
Effective Features

S/N	Selected Attributes
1	Ge
2	Tnp
3	Twp
4	Esp
5	As
6	Fm
7	Mq
8	Nf

As shown in Table II, attributes are selected as the most important out of a total of 22 attributes. Hence, these attributes returned by this feature selection method are referred to as SEF in this study.

TABLE III
PREDICTION RESULTS USING BAYES NETWORK AS A SINGLE CLASSIFIER AND AS AN ENSEMBLE METHOD

Parameters	Bayes Network	Bagging (Bayes Network)	Boosting (Bayes Network)
Correctly Classified Instances (%)	74.7686	75.2316	74.7686
Incorrectly Classified (%)	25.2315	24.7685	25.2315
Mean Absolute Error	0.2209	0.2206	0.2681
Root Mean Squared Error	0.3522	0.3467	0.3539
Relative Absolute Error (%)	51.0745	50.9957	61.9703
Root Relative Squared Error (%)	75.7433	74.5448	76.0934
Tp Rate	0.748	0.752	0.748
Fp Rate	0.145	0.138	0.145
Precision	0.749	0.753	0.749
Recall	0.748	0.752	0.748
F-Measure	0.747	0.750	0.747
ROC AREA	0.868	0.874	0.845

Table III presents the result obtained when Bayes Network is deployed in predicting students' performance. It is the first model as a single classifier before being an ensemble. This is not conclusive since it is a training stage, what matters most is the success records during the testing stage. As observed in Table III, the bagging method of Bayes Network did well during the training stage with an accuracy of 75.23%.

TABLE IV
PERFORMANCE EVALUATION/MEASUREMENT OF THE REP TREE AND ITS ENSEMBLES DURING THE TRAINING STAGE

Parameters	REP Tree	Bagging (REP Tree)	Boosting (REP Tree)
Correctly Classified Instances (%)	81.25	81.25	81.25
Incorrectly Classified (%)	18.75	18.75	18.75
Mean Absolute Error	0.1835	0.1844	0.2393
Root Mean Squared Error	0.3033	0.2912	0.3237
Tp Rate	0.813	0.813	0.813
Fp Rate	0.143	0.122	0.143
Precision	0.742	0.807	0.742
Recall	0.813	0.813	0.813
F-Measure	0.775	0.804	0.775
ROC AREA	0.921	0.946	0.897

The results obtained during the testing stage showed that there is no difference between the single classifier of Bayes Network and its experimented ensembles. Table IV showed that Bayes Network and its ensembles recorded an accuracy of 81.25%. However, in terms of ROC Bagging (Bayes Network) achieved 0.946 which is the highest. Hence, Bagging (Bayes Network) could be safely picked as being the best in this case.

Table V presents the results of the prediction with REP Tree and its ensembles during the testing stage. The same value is recorded across the board here also. A single classifier of the REP Tree and its ensembles returned a prediction accuracy of 83.33%. However, selecting the best classifier will now involve the consideration of other factors such as the ROC. Boosting method records the highest value for ROC with 0.903 which is the closest to 1.

TABLE V
PERFORMANCE EVALUATION/MEASUREMENT OF THE REP TREE AND ITS ENSEMBLES DURING THE TRAINING STAGE

Parameters	REP Tree	Bagging (REP Tree)	Boosting (REP Tree)
Correctly Classified Instances (%)	83.3333	83.3333	83.3333
Incorrectly Classified (%)	16.6667	16.6667	16.6667
Mean Absolute Error	0.2227	0.254	0.2389
Root Mean Squared Error	0.3251	0.3419	0.3237
Tp Rate	0.833	0.833	0.833
Fp Rate	0.119	0.128	0.119
Precision	0.827	0.812	0.827
Recall	0.833	0.833	0.833
F-Measure	0.823	0.815	0.823
Roc Area	0.851	0.886	0.903

TABLE VI
PERFORMANCE EVALUATION/MEASUREMENT OF RF

Parameters	RF (Training)	RF (Testing)
Correctly Classified Instances (%)	74.537	79.1667
Incorrectly Classified (%)	25.463	20.8333
Mean Absolute Error	0.2232	0.2351
Root Mean Squared Error	0.3438	0.3446
Tp Rate	0.745	0.792
Fp Rate	0.150	0.175
Precision	0.746	0.806
Recall	0.745	0.792
F-Measure	0.745	0.784
ROC AREA	0.877	0.875

Table VI presents the prediction results of RF. It is observed that RF could match not up with the previous classifier of Bayes Network, Logistic Regression, and REP Tree, both as single classifiers and their ensembles.

With the results of this study, it can be deduced that the REP Tree performed better compared to others in predicting students' academic performance with a prediction accuracy of 83.33%. This means that 30 of 38 students (testing set) are correctly classified to the right class labels (High, Medium, and Low) and 8 students are incorrectly classified. The results of this study further prove the reliability of the proposed model. Compared to the study by [1], REP Tree performed better than all the single classifiers (C4.5, Neural Network, and Naïve Bayes) deployed together with their ensemble of bagging, boosting, and RF in the study. The highest value recorded by [1] for prediction during testing is 79.1% while 82.2% is recorded during validation.

V. CONCLUSION

It has been established that academic institutions across the world today are faced with concerns about abysmal academic performance. E-Learning and LMS have been a good source of a large amount of data about teaching and learning interaction. From this platform, interesting knowledge can be obviated from the data kept in LMS to enhance the achievement of students academically. In this study, student performance prediction models based on machine learning are proposed with subset features called SEF. SEF are related to learner interactivity with LMS. The performance of students' predictive models is evaluated by a set of classifiers, namely; Bayes Network, Logistic Regression, and REP Tree. Consequently, ensemble methods are applied to improve the performance of these single classifiers. Bagging, Boosting, and RF, which form the array of most frequently used ensemble methods as reported in different literature, are deployed. The obtained results reveal that there is a strong relationship between SEF and their academic achievement. The accuracy of the student's predictive model using SEF in the case of REP Tree as a single classifier and ensemble methods achieved 83.33% prediction accuracy. In terms of ROC, boosting method of the REP Tree achieved the best with 0.903.

REFERENCES

[1] Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's Academic Performance Using Ensemble Methods. *International Journal of Database Theory and Application*, *9*(8), 209–213. https://doi.org/10.14257/ijdta.2016.9.8.13

[2] Bendangnuksung, R. K & Prabu, E. P. (2018). Early Dropout Prediction Using Data Mining: A Case Study with High School Students. *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 13, Number 2 (2018) pp. 1171-1176

[3] Devasia, M. T., P, M. V. T., & Hegde, V. (2016). Prediction of Students Performance Using Educational Data Mining. In *Department of Computer Science Amrita Vishwa Vidyapeetham University, Mysuru Campus Mysuru, Karnataka, India* (p. 190).

[4] Gupta, A., Gupta, S., & Singh, D. (2015). A Systematic Review of Classification Techniques and Implementation of ID3 Decision Tree Algorithm. In *International Conference on System Modeling & Advancement in Research Trends (SMART)* (pp. 144–152).

[5] Han, H. U. I. X., & Jie, J. S. U. N. (2009). Financial Distress Prediction Based on Ensemble Classifiers of Multiple Reductions. *International Conference on Management Science & Engineering* (pp. 1247–1252).

[6] Hodžić, A., Kevrić, J., & Karadag, A. (2016). Comparison of Machine Learning Techniques in Phishing Websites. In *International Conference on Economic and Social Studies (ICESoS)) 2016* (pp. 249–256).

[7] Mhetre, V. (2017). Classification-Based Data Mining Algorithms to Predict Slow, Average, and Fast Learners in Educational System Using Weka. In *Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication* (pp. 475–479).

[8] Mousa, H., & Maghari, A. (2017). School Students â€™ Performance Prediction Using Data Mining Classification. *International Journal of Advanced Research in Computer and Communication Engineering ISO*, *6*(8), 136–141. https://doi.org/10.17148/IJARCCE.2017.6824

[9] Osokoya, I. J (2009). Mining Educational Data to Predict Student's Academic Performance Using Ensemble Methods. And Theory *International Journal Application*, Database of 9(8), 209–213.https://doi.org/10.14257/ijdta.2016.9.8.13

[10] Olukoya, B. M (2018). Mining Educational Data to Predict Student's Academic Performance. *International Journal on Recent and Innovation Trends in Computing and Communication,* 4(1), 1–5.

[11] Phil, M., & Shoba, S. A. (2017). Educational data Mining & Students Performance Prediction Using SVM Techniques. *International Research Journal of Engineering and Technology (IRJET)*, *4*(8), 1248–1254.

[12] Roy, S. (2017). Predicting Academic Performance of Students Using Classification Techniques. In *4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer, and Electronics* (pp. 568–572).

[13] Roy, S., & Garg, A. (2017). Predicting Academic Performance of Students Using Classification Techniques. In *4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer, and Electronics* (pp. 568–572). Mathura.

[14] Shet, S. (2014). Approach for Predicting Student Performance Using Ensemble Model Method. *International Journal of Innovative Research in Computer and Communication Engineering*, *2*(5), 161–169.

[15] Suharjito, Diana, & Herianto. (2017). Implementation of Classification Technique in Web Usage Mining of Banking Company. In *Proceeding - 2016 International Seminar on Intelligent Technology and Its Application, ISITIA 2016: Recent Trends in Intelligent Computational Technologies for Sustainable Energy* (pp. 211–218). https://doi.org/10.1109/ISITIA.2016.7828660

[16] Vinodhini, G., & Chandrasekaran, R. M. (2018). Predicting Students' Academic Performance in the University Using Meta Decision Tree Classifiers. *Journal of Computer Science*, *15*(12), 1–8. https://doi.org/10.3844/jcssp.2018.654.662

[17] Wati, M., Indrawan, W., Widians, J. A., & Puspitasari, N. (2017). Data Mining for Predicting Students ' Learning Results. In *Dept. of Computer Science and Information Technology Universitas Mulawarman Samarinda, Indonesia* (p. 28).