Applied Computer Systems
ISSN 2255-8691 (online)
ISSN 2255-8683 (print)
December 2021, vol. 26, no. 2, pp. 122–131
https://doi.org/10.2478/acss-2021-0015
https://content.sciendo.com

Academic Performance Modelling with Machine Learning Based on Cognitive and Non-Cognitive Features

Bridgitte Owusu-Boadu[1], Isaac Kofi Nti[2*], Owusu Nyarko-Boateng[3], Justice Aning[4], Victoria Boafo[5]
[1] *Brivink Consult & Technology, Sunyani, Ghana*
[2,3] *Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana*
[4] *Department of Computer Science, Sunyani Technical University, Sunyani, Ghana*
[5] *Department of Computer Science & ICT, Mampong Technical College of Education, Mampong, Ghana*

Abstract – **The academic performance of students is essential for academic progression at all levels of education. However, the availability of several cognitive and non-cognitive factors that influence students' academic performance makes it challenging for academic authorities to use conventional analytical tools to extract hidden knowledge in educational data. Therefore, Educational Data Mining (EDM) requires computational techniques to simplify planning and determining students who might be at risk of failing or dropping from school due to academic performance, thus helping resolve student retention. The paper studies several cognitive and non-cognitive factors such as academic, demographic, social and behavioural and their effect on student academic performance using machine learning algorithms. Heterogenous lazy and eager machine learning classifiers, including Decision Tree (DT), *K*-Nearest-Neighbour (KNN), Artificial Neural Network (ANN), Logistic Regression (LR), Random Forest (RF), AdaBoost and Support Vector Machine (SVM) were adopted and training was performed based on k-fold (*k* = 10) and leave-one-out cross-validation. We evaluated their predictive performance using well-known evaluation metrics like Area under Curve (AUC), F-1 score, Precision, Accuracy, Kappa, Matthew's correlation coefficient (MCC) and Recall. The study outcome shows that Student Absence Days (SAD) are the most significant predictor of students' academic performance. In terms of prediction accuracy and AUC, the RF (Acc = 0.771, AUC = 0.903), LR (Acc = 0.779, AUC = 0.90) and ANN (Acc = 0.760, AUC = 0.895) outperformed all other algorithms (KNN (Acc = 0.638, AUC = 0.826), SVM (Acc = 0.727, AUC = 0.80), DT (Acc = 0.733, AUC = 0.876) and AdaBoost (Acc = 0.748, AUC = 0.808)), making them more suitable for predicting students' academic performance.**

Keywords – **Academic performance, AdaBoost, artificial neural network, decision tree, educational data mining, k-nearest neighbour, logistic regression, machine learning, naïve Bayes, random forest, support vector machine.**

I. Introduction

The academic performance of students is one of the integral mechanisms for evaluating an academic institution. It also helps design operative instruments that improve students' academic outcomes and avoid dropout, among other things. For example, West African second-cycle schools are usually ranked yearly based on the general performance of their students in the West African Examination Council (WAEC). Likewise, in Ghana, the rank of basic educational (primary and junior secondary schools) institutions is partially based on their students' performance in the Basic Education Certificate Examination (BECE). Therefore, a means to correctly articulate students' academic performance before their final examination at all education levels is critical to all academic institutions. Recently, there has been a decline in the general performance of students in science, mathematics and information technology. Some nationalists blame the use of foreign languages for teaching as responsible for poor academic performance and even underdevelopment among students [1].

However, other studies [2]–[5] have argued that students' academic performance goes beyond the language for means of instruction; it fundamentally includes both cognitive (such as high school grade point average) and non-cognitive characteristics. The non-cognitive characteristics include student engagement, behavioural observations at school, the general views of family and friends concerning schooling, gender, place of birth and involvement in extracurricular activities.

Recently, advancement in technology has resulted in a large amount of collected cognitive and non-cognitive educational data. Nevertheless, analysing big data to reach insightful information is challenging for humanity using traditional techniques [6]. However, Data Mining (DM) methods can be used effectively to learn treasured and essential hidden knowledge from these data. Thus, Educational Data Mining (EDM), i.e., analysing data of educational institutions with DM methods, is valuable for dramatically improving students' academic performance. Educational Data Mining implements data mining methods for analysing available data at educational institutions [6]. Although EDM leads to knowledge discovery,

* Corresponding author's e-mail: Isaac.nti@uenr.edu.gh

Machine Learning Algorithms (MLAs) provide the needed tools for this purpose. Hence, critical analysis and processing of educational data with MLAs can lead to valuable statistics concerning students' knowledge, association and academic performance.

This paper seeks to examine the degree to which cognitive and non-cognitive characteristics influence the academic performance of students at second-cycle schools using MLAs, as well as compares and evaluates the predictive performance of different classification methods for predicting students' academic performance. Specifically, we seek to:

i. Examine which factors, cognitive or non-cognitive or both, are significant predictors of students' academic performance.
ii. Perform a comparative analysis of different MLAs for predicting student academic performance to identify which algorithm improves prediction accuracy.
iii. Predict the average grade score of a student in mathematics, science and information technology and estimate whether a student is at risk of failing the final examination based on the best classification algorithm from the comparative analysis.
iv. Examine the effect of k-fold and leave-one-out cross-validation training techniques on the prediction accuracy of different MLAs for predicting students' academic performance.

The following research questions have been formulated:

RQ1: Which MLAs are appropriate for effective and efficient prediction of students' academic performance?

RQ2: Which cognitive and non-cognitive factors affect students' academic performance and predictive performance of MLAs?

We hope that the outcome of this study will help identify: (i) different cognitive and non-cognitive factors that significantly influence students' academic performance; (ii) a training technique (i.e., k-fold cross-validation or leave-one-out) that is more suitable for training MLAs to predict students' academic performance; (iii) the appropriate MLAs for predicting students' academic performance and determine the best approach for performing student performance prediction. This paper uses various machine learning classifiers, including Decision Tree (DT), K-Nearest-Neighbour (KNN), Artificial Neural Network (ANN), Logistic Regression (LR), Random Forest (RF), AdaBoost and Support Vector Machine (Radial Basis Function). These techniques have been chosen for this study based on their straightforward implementation and efficiency in classification tasks in different fields, such as engineering [7]–[10], finance [11]–[14] and education [15]–[18].

The rest of the paper is structured as follows: Section 2 presents review of literature. Section 3 presents the study methodology. Then, the study results are considered in Section 4, while conclusions and future research areas are discussed in Section 5.

II. Related Works

Some researchers attempted to predict students' academic performance based on machine learning algorithms using cognitive or non-cognitive factors. In this section, we present a few of them published between 2017 and 2020.

Chui et al. [19] adopted a generative adversarial network-based deep SVM predictive model to predict students' performance. The paper reported that the proposed ICGAN-DSVM outperformed other state-of-the-art works by 8–29 % in AUC, specificity and sensitivity. Likewise, Hamoud, Hashim and Awadh [20] predicted students' academic performance using tree algorithms (J48, Random Tree and REPTree). The study outcome showed that the J48 algorithm outperformed Random Tree and RepTree algorithms. A predictive model based on a Convolutional Neural Network (CNN) was applied in [21] to predict students' academic performance. The proposed CNN model outperformed related works in terms of Recall, F1-score and Precision. Ahmad and Shahzadi [22] proposed an ANN-based student academic performance predictive framework. They reported prediction accuracy compared with other well-known MLAs. Burman and Som [23] employed a multi-classifier SVM to classify students into high, average and low groups based on their academic scores. The paper recorded higher prediction accuracy with the RBF kernel than using a linear function. Aggarwal et al. [16] experimented on student data containing academic and non-academic features using six machine learning classification algorithms. They found that multi-layer perceptron and Random Forest were the most hopeful classifiers for predicting academic performance.

Similarly, Hussain et al. [17] used PART, RF, J48 and Bayes classification algorithms to predict the academic performance of 300 students and reported RF to offer better classification accuracy and minor error. Likewise, Almarabeh [18] performed a comparative analysis of five classifiers, namely, Neural Network, Bayesian Network, J48, Naive Bayes and ID3, for predicting students' academic performance. The outcome of the study showed that the Bayesian Network classifier outperformed all other classifiers.

In reality, student academic performance prediction is essential for student academic development. Hence, several studies have been carried out in this field, but most of these studies used single factors like cognitive or non-cognitive but not both. However, several factors (cognitive and non-cognitive) affect students' academic performance and each of these factors possesses unique characteristics [6], [21]. Therefore, a blend of both is anticipated to give better results than single data [24]. On the other hand, studies that combined them randomly selected the MLA. Besides, as pointed out in the literature [25]–[27], the performance of MLAs is highly affected by the dataset characteristics. Hence, a single MLA might not perform well on all types of datasets. Again, in most previous studies, several researchers have proposed different classification models for students' academic performance based on different features; however, the optimisation effects of these significant features on students' performance classification was ignored.

III. Material and Methods

This section discusses the methods and techniques adopted in the study to accomplish the study goal.

A. Dataset and Collection Tool

The study population consisted of all third-year students (1133) in three (3) reputable private second-cycle institutions in Kumasi, Ghana. However, four hundred and eighty (480) students were sampled using the stratified random sampling technique. We concluded on this sample size based on Yamane (1967) as defined in (1) [28], using a 5 % margin of error.

$$n = N\left(1 + Ne^2\right)^{-1} \tag{1}$$

The researchers adopted a self-design questionnaire for data collection. It consisted of two sections; the first section was designed to obtain demographic information such as gender, place of birth, nationality, level, class group, relation to guardian, parent satisfaction with school and more. The second section was carefully designed to obtain quantitative information such as the number of student absence days, the frequency of students raised hands in class, how often students visit school library, student involvement in group discussions, number of times student visited school notice board. Participants were to indicate on a scale of 1 to 10 for a given term (3 months). The consent of school authorities and students was sought in this study. The terminal scores of students in mathematics, science and information technology were average as the final score (f_{score}) of a participant in this study, as defined in (2).

$$f_{\text{score}} = \frac{Math_{\text{score}} + Science_{\text{score}} + IT_{\text{score}}}{3} \tag{2}$$

f_{score} was then categorised into three groups:

if $f_{\text{score}} \leq 54$, *performance = low* (*L*)[0]
elseif $f_{\text{score}} \leq 69$, *performance = medium* (*M*)[1]
else performance = high (*H*)[2]

Thus, the study was carried out using sixteen features (see Table I). Fifteen served as independent (predictor) variables and one (performance) as a dependent (response) variable. Thus, the study dataset was a 480×16 matrix.

TABLE I
STUDY VARIABLES

Features	Symbol
1. Gender	G
2. Nationality	N
3. Place of birth	PB
4. Level	L
5. Class group	CG
6. Topic (math, science and IT)	TP
7. Term	TR
8. Relation to guardian	RG
9. Number of times student raised hands in class	NTSRHC
10. Visit school library	VSL
11. Notice board visit	NBV
12. Involvement in group discussions	IGD
13. Parent answering survey	PAS
14. Parent satisfaction with school	PASS
15. Student Absence Days	SAD
16. Performance (L = 0, M = 1, H = 2)	

B. Description of the Selected Algorithms

There are two types of learners in classification: Lazy Learners (LL) and Eager Learners (EL). The LL store the training data until classification using the testing data when it appears and has more significant predicting times, e.g., K-Nearest Neighbours and case-based reasoning. While the EL are trained before being data for prediction, they work the whole space based on a single hypothesis and fast make predictions, e.g., Naive Bayes, Artificial Neural Networks, Decision Trees, etc. In this study, we adopted a combination of LL and EL machine learning algorithms. A brief description of the algorithms used in this paper is presented as defined in [29]–[33]. Readers unfamiliar with machine learning and its associated algorithm are referred to [29]–[31], [33] for detailed tutorials on machine learning. We selected these techniques based on the performance in several related fields as reported in the literature [34]–[36] and their easy implementation and less computational time.

1. *K*-Nearest Neighbours (KNN) is easy to understand and implement supervised MLA for classification or regression tasks. It adopts the resemblance between new and available datasets and puts the new into the most similar group to the available groups. It is a non-parametric algorithm, i.e., it does not make any guess on primary data. It is sometimes referred to as an LL algorithm since it does not learn from the training dataset instantly instead keeps it. At the classification period, it completes an action on the dataset.

2. Decision Tree (DT) is a simple supervised MLA for regression and classification tasks; however, it is mainly used to classify problems. It is tree-like in construction, where inner nodes characterise the features of a dataset, branches signify the decision rules, and each leaf node signifies the result.

3. Random Forest (RF) is a supervised MLA commonly used for classification and regression tasks due to its straightforwardness and variability. It is an ensemble learning technique that combines the outcome of two or more decision trees via majority voting or averaging technique to enhance prediction accuracy depending on the problem at hand.

4. Multilayer Perceptron (MLP) is a deep ANN that learns a function f(.) : $R^m \rightarrow R^o$ by training on a dataset (DS), where (m) is the dimensions of DS and (o) is the output dimensions. It has at least three (3) layers of nodes, namely, (i) the input layer for receiving g the input signal, (ii) output layer for making a decision or forecast about the input and (iii) hidden layer(s) sandwich between (i) and (ii) for all computation of the MLP; this layer has an arbitrary number. They are often used for supervised ML tasks. Apart from the input layer, every other node is a neuron that uses a nonlinear activation function.

5. Support Vector Machine (SVM) is a supervised MLA for classification, regression and outliers' challenges. It aims at finding an optimal minimal hyperplane (h) *N*-dimensional space (*N* is the number of features in the dataset) that best divides a given dataset into classes. It is capable of handling both continuous and categorical datasets. The SVM creates an (*h*) in multi-dimensional space to separate different classes by generating optimal (*h*) iteratively, which is used to curtail an error. As a result, it gives better accuracy compared to other MLA techniques like LR and DT.

6. Adaptive Boosting (AdaBoost) is the first hands-on boosting algorithm developed in 1996 by Freund and Schapire. It is a boosting method for ensemble ML, where weights are re-assigned to a particular example, with higher weights to wrongly categorised examples.

7. Logistic Regression (LR) is an MLA describing data and explaining the association between one dependent binary feature and two or more ordinal, nominal, interval, or ratio-level independent features.

TABLE II
HYPERPARAMETER SETTINGS

MLA	Hyperparameters
LR	penalty= l2, solver='lbfgs', max_iter =100, multi_class='auto', tol=0.0001, C=1.0
MLP	hidden_layer_sizes=100, alpha=0.15, learning_rate='0.001', activation='relu', solver='adam', batch_size='auto', hidden_layer_sizes=[50, 50]
SVM	C=100, kernel='rbf', degree=3, gamma='0.1', tol=0.001, learning_rate =' adaptive', max_iter =1000
RF	criterion = gini, min_impurity_decrease = 0.001, max_depth =6, min_sample_split = 6, n_estimators = 190, min_sample_leaf =6
AdaBoost	base_estimator=None, n_estimators=60, learning_rate=1.0, loss='linear', algorithm='SAMME'
KNN	n_neighbors=8, weights='uniform', algorithm='auto', leaf_size=30, p=2, metric = manhattan
DT	criterion='gini', splitter='best', max_depth=16, min_samples_split=2, max_features=log2

C. Study Framework

Fig. 1 shows the framework for the implementation of the study goal. Python, Scikit-learn library (https://scikit-learn.org) was used for all experiments in this paper. A total of 480 records with sixteen features were obtained from the participants. Firstly, we pre-processed our dataset by replacing missing values (where students did not respond to question) with average values. Then, we encoded all categorical data into numerical using the label encoding approach in the Scikit-learn library. Some of the attributes in our dataset were different; in size, we scaled the dataset in a range (0 to 1), using the min-max normalisation technique (3).

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \quad (3)$$

The clean dataset was then partitioned into 75 % for training and 25 % for testing. 10-fold cross-validation and leave-one-out techniques were used to train the models separately in this study.

D. Definition of the Classification Problem

Let N be a sequence; data point $x^{(i)} \in \Re^n$, $1 \le i \le N$; each having n characteristic features $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}\}$; each element $x^{(i)}$ is assigned a label $y^{(i)}$. For $n, c \in N$, a set of c labels I, and a sequence $x^{(i)} \in \Re^n$, $y^{(i)} \in I$, $1 \le i \le M$, find $f : \Re^n \rightarrow I$ such that $f(x^{(i)}) = y^{(i)}$ for all $1 \le i \le N$.

In this study, $c = 3$. Our goal is, given a pre-labelled training dataset DS = $(x^{(i)}, y^{(i)})$, $1 \le i \le M$, $M < N$, we attempt to make a machine learning algorithm find a function $f : \Re^n \rightarrow \{-1, 0, +1\}$.

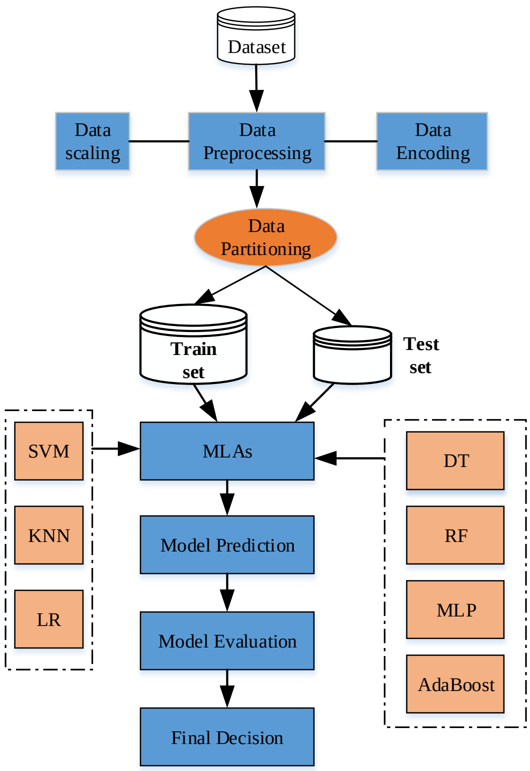

Fig. 1. Study framework.

E. Evaluation Metric

The prediction performance of the selected algorithms was compared using five well-known evaluation metrics for evaluating classification tasks in machine learning, namely, Area under Curve (AUC), F-1 score, Precision, Accuracy, Kappa, Matthews correlation coefficient (MCC) and Recall. A detailed definition of these metrics is available in [37]. Finally, the best classification algorithm for predicting the students' academic performance is selected based on the comparative analysis of the evaluation metrics.

F. Experimental Setup

All experiments in the current study were carried on a Lenovo (20EGS12E00) laptop, Intel® core™ i5-4340M CPU @ 2.90GHz (4 CPUs) 12GB memory. Table II shows the hyperparameter stings for the MLAs used in this study.

IV. Results and Discussions

This section presents the outcome of our experiments.

A. Evaluation of Cognitive and Non-Cognitive Factors Affecting Students' Academic Performance and Predictive Performance of MLAs

Concerning RQ1 (Which cognitive and non-cognitive factors affect students' academic performance and predictive performance of MLAs?), this study selected fifteen features reported in the literature to affect students' academic performance. We measured their degree of importance with student academic performance. The aim was to examine which among these features was a high predictor of students' academic performance. Table III shows the outcome of this study. The features were ranked using information gain ratio and Gini decrease. A feature with an importance measure closer to one is considered a feature with higher importance. The outcome shows that SAD is the most significant predictor of students' academic performance, followed by visit-school-library (VSL), the number of times students raised hands in class (NTSRHC), parent answering survey (PAS) and relation to guardian (RG).

TABLE III
Features Importance Ranking

Features	Gain ratio	Gini
SAD	**0.410**	**0.131**
VSL	**0.195**	**0.145**
NTSRHC	**0.181**	**0.139**
PAS	**0.152**	**0.055**
RG	**0.129**	**0.049**
NBV	0.127	0.098
PASS	0.111	0.040
G	0.055	0.019
N	0.052	0.045
PB	0.051	0.046
IGD	0.044	0.038
TP	0.023	0.030
L	0.019	0.019
TR	0.012	0.005
CG	0.006	0.003

The outcome suggested that student's absenteeism is more likely to affect their academic performance than any other factor. Likewise, students who regularly visit the school library for further learning are more likely to perform better than those who do not. Also, the outcome shows that students who contribute in class either by answering questions or asking a question are more likely to perform well academically than their counterparts who do not engage in class activities. Furthermore, the outcome suggests that parent involvement in their academic activities either by helping them with their homework and other school activities contributes effectively to their academic performance.

B. Performance Measure of Seven Different MLAs for Predicting Students' Academic Performance

Regarding RQ2 (Which MLAs are appropriate for effective and efficient prediction of students' academic performance?), this study applied seven machine learning classification algorithms to classify students' academic performance. In addition, this study employed five well-known evaluation metrics to evaluate the predictive performance for classifying students' academic performance based on two training techniques.

Fig. 2 shows the comparative performance of the selected seven MLAs using a 10-fold cross-validation training technique. We observed that the RF obtained the highest prediction accuracy of 0.765, followed by the SVM (0.746) and the LR (0.7416). Thus, in terms of prediction accuracy, these three algorithms outperformed the KNN (0.7127), DT (0.696), AdaBoost (0.712) and MLP (0.716). Regarding the area under the curve metrics, RF measured 0.893, LR – 0.856. However, the KNN outperformed DT, AdaBoost, MLP and the SVM. Accordingly, the AUC, which is a key indicator for measuring the overall performance and the general measure of model accuracy as pointed out in [38], suggests that the RF is more appropriate for predicting students' academic performance than KNN, SVM, DT, AdaBoost, MLP and LR based on a 10-fold cross-validation training technique. In terms of precision and recall metrics, the RF recorded 0.767, 0.776; MLP – 0.725, 0.72 and LR – 0.756, 0.747. The difference in MLA performance based on different evaluation metrics suggests that using single metrics to compare algorithms is not a good standard since each metric gives a unique characteristic. In Aggarwal et al. [16], MLP outperformed RF. However, our outcome disagrees (see Fig. 2). In contrast, it agrees with Hussain et al. [17] who reported that RF outperformed PART, J48 and Bayes classifiers.

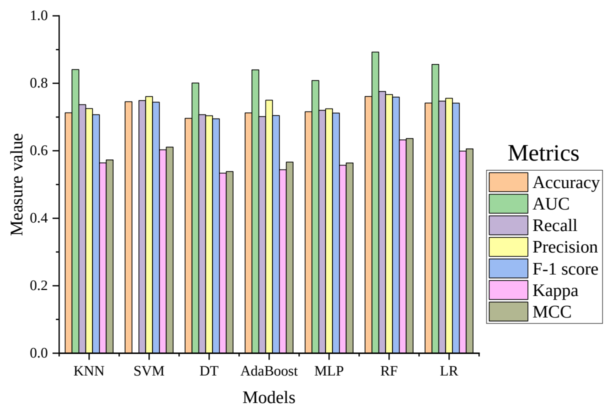

Fig. 2. MLAs performance with a 10-fold cross-validation training technique.

Fig. 3 shows the performance of the 7 MLAs in predicting students' academic performance based on the leave-one-out training technique. Similar to the outcome with the 10-fold cross-validation technique in this study (see Fig. 2), the LR, RF and MLP outperformed the KNN, SVM, DT and AdaBoost in accuracy and AUC. Nevertheless, we observed a slight increase in the prediction accuracy and AUC for all the MLAs. Thus, LR obtained an accuracy of 0.779 and AUC of 0.90, RF (Acc = 0.771, AUC = 0.903), MLP (Acc = 0.760, AUC = 0.895), KNN (Acc = 0.638, AUC = 0.826), SVM (Acc = 0.727, AUC = 0.80), DT (Acc = 0.733, AUC = 0.876) and AdaBoost (Acc = 0.748, AUC = 0.808). The higher the AUC, the better the performance of the machine learning model is at predicting false (0) as false (0) and true (1) as true (1). The AUC of the RF shows that it is more suitable for predicting students' performance based on cognitive and non-cognitive factors, likewise, in F1-score, precision and Recall (see Fig. 3). The slight increase in model performance with leave-one-out training techniques over the k-fold cross-validation shows that the leave-one-out training technique is more suitable for training models in predicting students' academic performance. The logistic regression obtained better accuracy between 74.2 % and 77.9 % in both training techniques.

The Receiver Operating Characteristics (ROC) curve visualises machine learning classifier performance. It is one of the critical essential evaluation metrics for assessing the classification performance of any model. Figs. 4–9 show the ROC of the RF, AdaBoost, DT, KNN, MLP and LR classifiers, respectively. For example, from Fig. 4, it can be seen that the RF has approximately 76 %–84% chance to distinguish students by low class (0), medium-class (1) and high class (2).

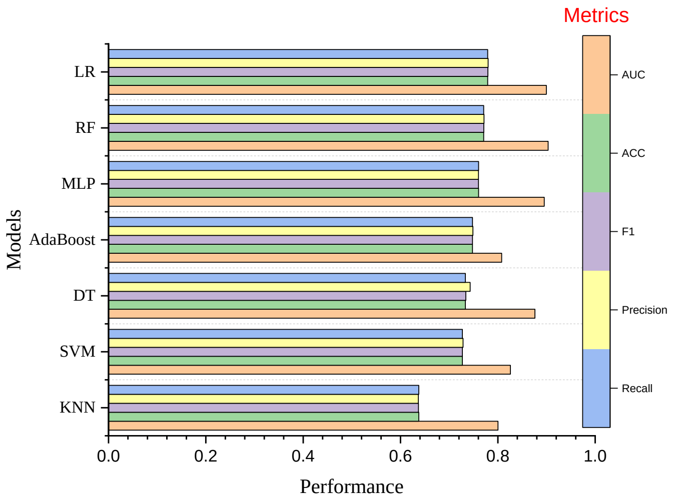

Fig. 3. MLA performance with the leave-one-out training technique.

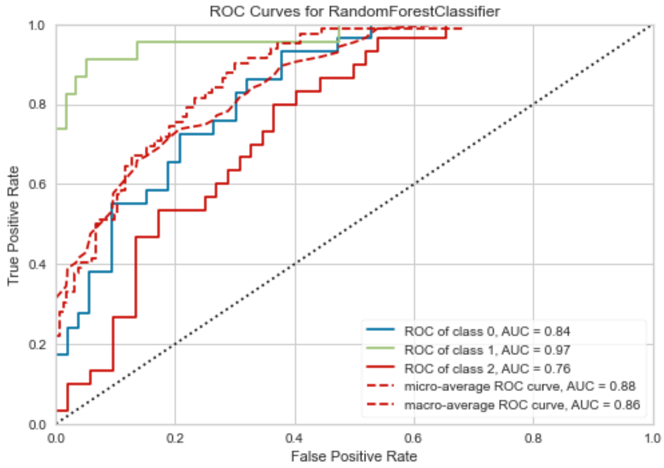

Fig. 4. ROC curve for RF classifier.

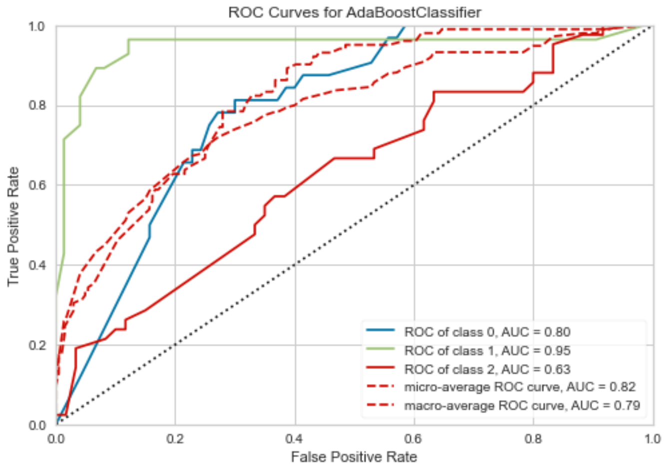

Fig. 5. ROC curve for AdaBoost classifier.

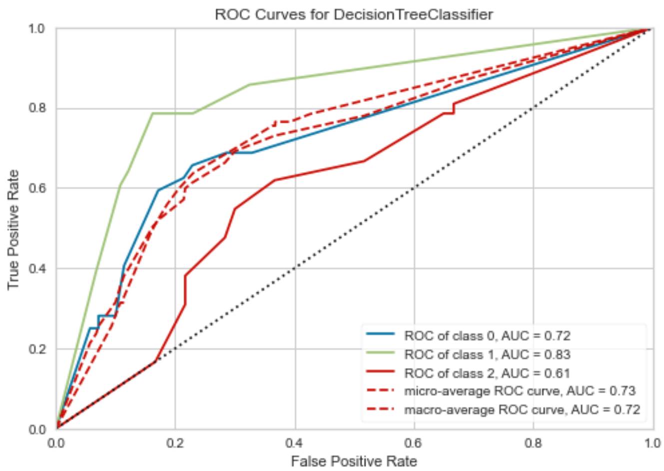

Fig. 6. ROC curve for DT classifier.

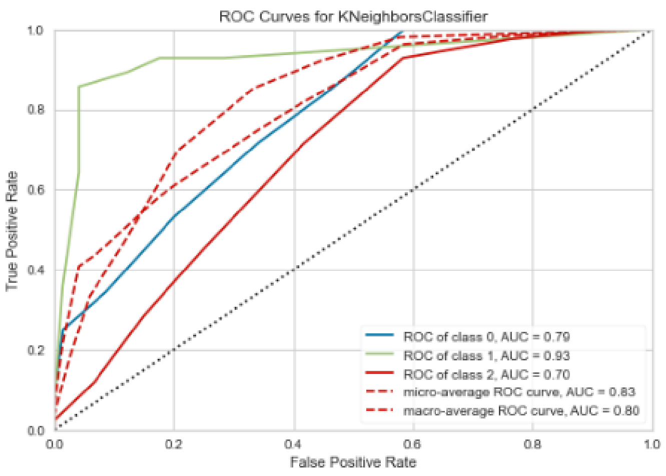

Fig. 7. ROC curve for KNN classifier.

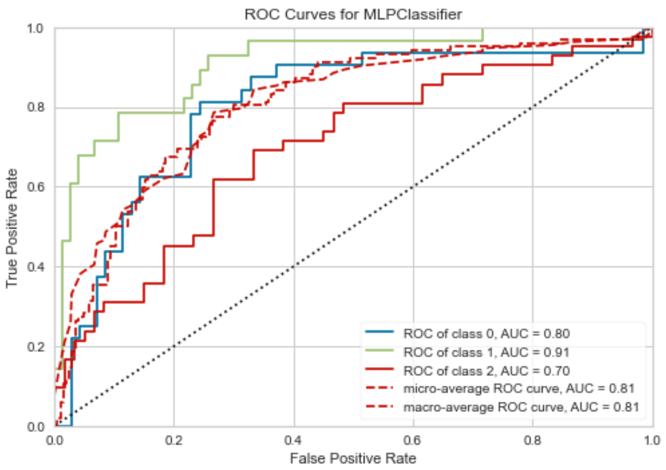

Fig. 8. ROC curve for MLP classifier.

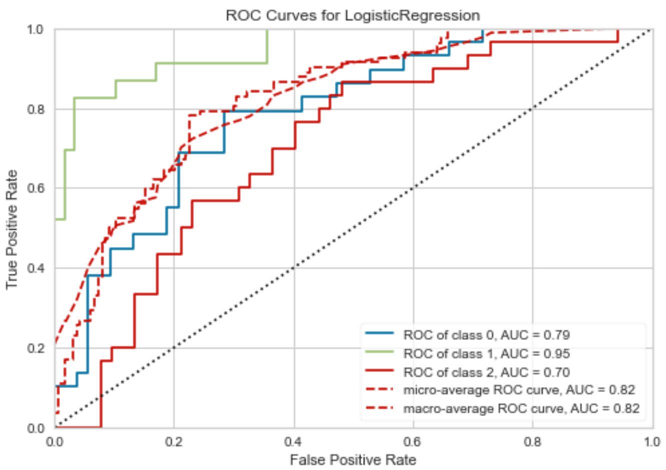

Fig. 9. ROC curve for LR.

Figs. 10–16 show the class report of the AdaBoost, DT, KNN, LR, MLP, RF and SVM, respectively. The case reports show the precision, Recall and Support of the models. The Support, which defines the number of examples of the true response in each class of target values, demonstrates that the RF outperformed all other classifiers.

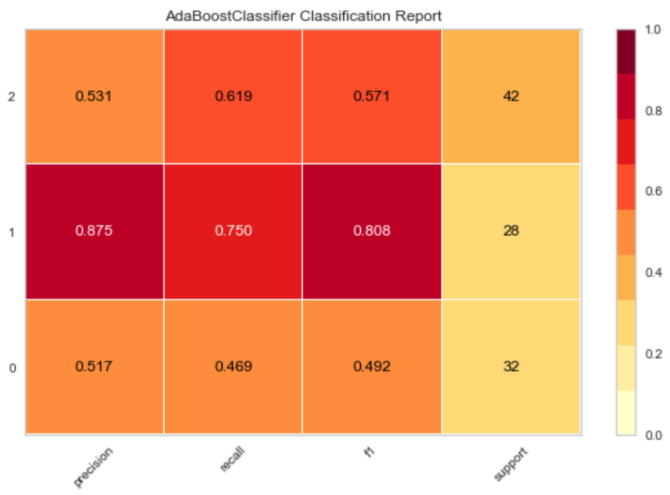

Fig. 10. Classification report for AdaBoost.

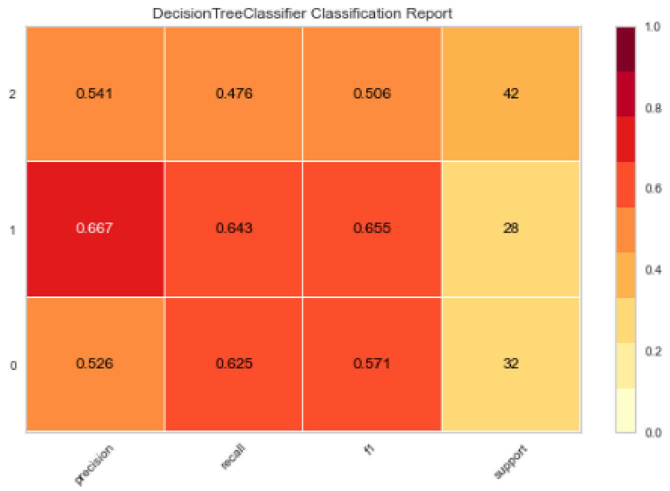

Fig. 11. Classification report for DT.

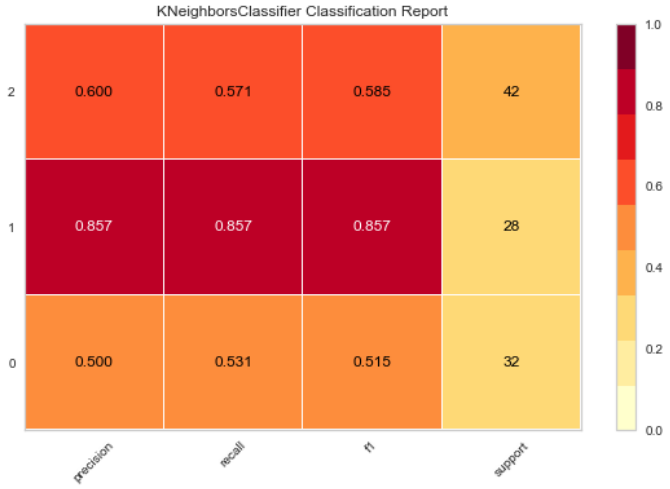

Fig. 12. Classification report for KNN.

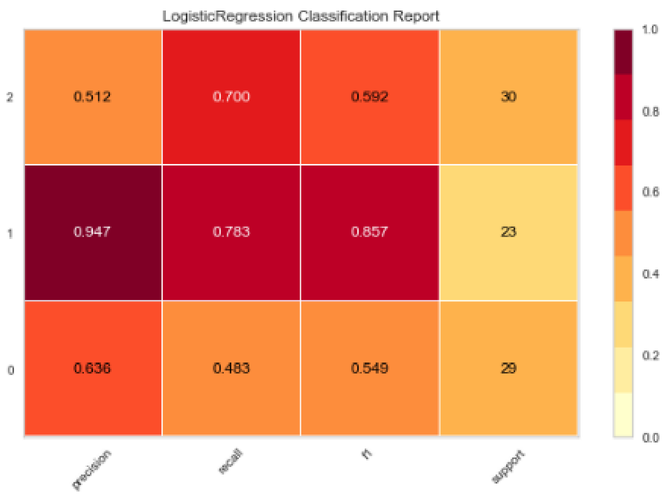

Fig. 13. Classification report for LR.

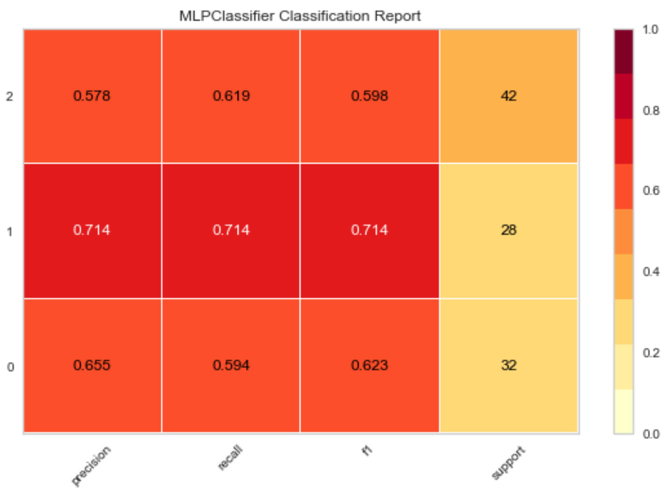

Fig. 14. Classification report for MLP.

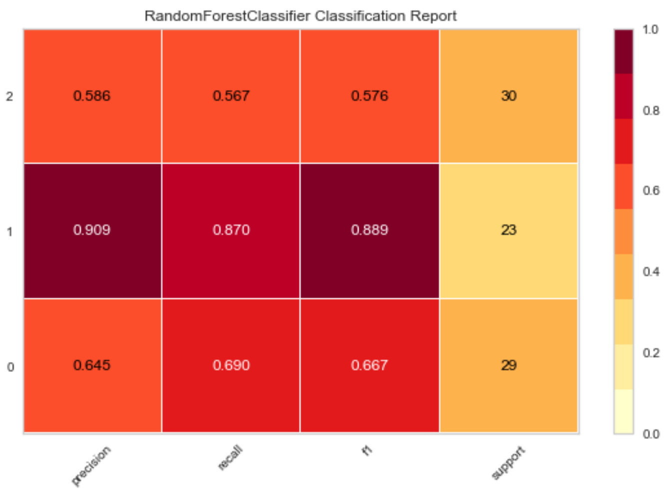

Fig. 15. Classification report for RF.

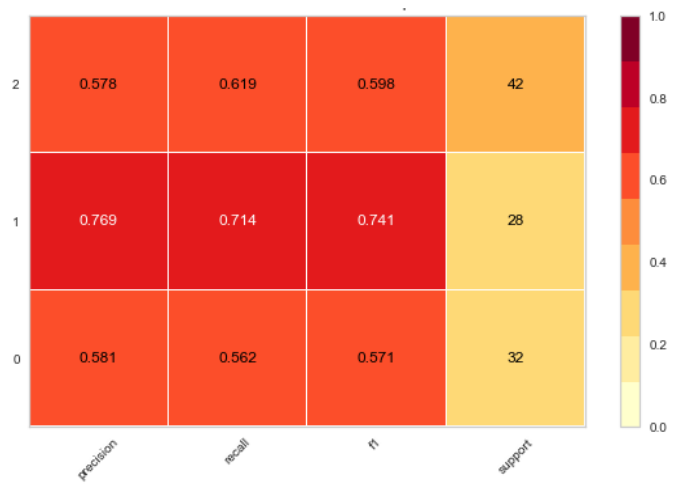

Fig. 16. Classification report for SVM.

V. Conclusion

Students' academic performance is a critical factor in students' development and measuring academic intuition performance. This study has evaluated the effect of fifteen (15) cognitive and non-cognitive factors believed to affect students' performance. It has also predicted the academic performance of students based on the most significant factors among the fifteen. Seven machine learning classification algorithms, namely DT, K- KNN, ANN, LR, RF, AdaBoost and SVM, have been used in our experiments. In addition, two training techniques, namely k-fold cross-validation and leave-one-out, have been adopted to train these algorithms. Five (5) evaluators have been employed, namely, AUC, F-1 score, Accuracy, Recall, Precision, Kappa and MCC, to measure the predictive performance of the seven classification methods

The outcome shows that SAD is the most significate predictor of students' academic performance, followed by the frequency of a student visit-school-library, the number of times students raised hands in class, parent answering survey and relation to guardian. Also, from the study outcome, it can be confirmed that MLAs can effectively and accurately predict students' academic performance. Furthermore, we have observed that training the MLAs with the leave-one-out technique yielded higher performance metrics than the *k*-fold cross-validation technique ($k = 10$). Thus, the study results show that the RF, LR, and MLP are perfect classifiers for classifying students' academic performance based on their achieved AUC and accuracy values.

To summarise the abovementioned results, the most significant factors that influenced students' academic performance classification are SAD, VSL, NTSRHC, PAS and RG. Thus, the higher the values of these influencing factors, the higher the predictive ability of classification. Therefore, the identified significant factors can help teachers design a learning activity that can promote the performance of students. The current study has used a dataset from three institutions in Kumasi, Ghana. In future, more datasets will be added from several private and public institutions to give a generalised report across the country. Also, different sampling techniques will be adopted to examine their effect on various machine learning algorithms.

Funding: Not applicable.
Conflicts of interest/Competing interests: The authors declare that there are no conflicts of interest or competing interests.
Availability of data and material: The data used in the study would be made available upon reasonable request.
Code availability: The data used in the study would be made available upon reasonable request.

REFERENCES

[1] E. I. Ani, “Debating the roots of poor academic performance in the West African subregion: The perspective of a philosopher,” *SAGE Open*, vol. 7, no. 2, Art no. 2158244017707795, May 2017. https://doi.org/10.1177/2158244017707795
[2] B. G. Adams, N. Wiium, and A. Abubakar, “Developmental assets and academic performance of adolescents in Ghana, Kenya, and South Africa,” *Child & Youth Care Forum*, vol. 48, no. 2, pp. 207–222, Nov. 2019. https://doi.org/10.1007/s10566-018-9480-z
[3] S. Venkatesh, Y. K. Rao, H. Nagaraja, T. Woolley, F. O. Alele, and B. S. Malau-Aduli, “Factors influencing medical students’ experiences and satisfaction with blended integrated e-learning,” *Medical Principles and Practice*, vol. 29, no. 4, pp. 396–402, Jul. 2020. https://doi.org/10.1159/000505210
[4] L. F. Casinillo, M. A. E. Palen, E. L. Casinillo, and P. G. Batidor, “Assessing senior high student’s learning experiences in mathematics,” *Indonesian Journal of Educational Studies*, vol. 23, no. 1, pp. 44–60, 2020. https://doi.org/10.26858/ijes.v23i1.13437
[5] C. Semeraro, D. Giofrè, G. Coppola, D. Lucangeli, and R. Cassibba, “The role of cognitive and non-cognitive factors in mathematics achievement: The importance of the quality of the student-teacher relationship in middle school,” *PLoS ONE*, vol. 15, no. 4, Art no. e0231381, 2020. https://doi.org/10.1371/journal.pone.0231381
[6] R. Ghorbani and R. Ghousi, “Comparing different resampling methods in predicting students’ performance using machine learning techniques,” *IEEE Access*, vol. 8, pp. 67899–67911, Apr. 2020. https://doi.org/10.1109/ACCESS.2020.2986809
[7] I. K. Nti, A. Y. Appiah, and O. Nyarko-Boateng, “Assessment and prediction of earthing resistance in domestic installation,” *Engineering Reports*, vol. 2, no. 1, Art no. e12090, Jan. 2020. https://doi.org/10.1002/eng2.12090
[8] I. K. Nti, M. Teimeh, A. F. Adekoya, and O. Nyarko-Boateng, “Forecasting electricity consumption of residential users based on lifestyle data using artificial neural networks,” *ICTACT Journal on Soft Computing*, vol. 10, no. 3, pp. 2107–2116, 2020.
[9] O. Nyarko-Boateng, A. F. Adekoya, and B. A. Weyori, “Predicting the actual location of faults in underground optical networks using linear regression,” *Engineering Reports*, vol. 3, no. 3, Art no. e212304, Mar. 2021. https://doi.org/10.1002/eng2.12304
[10] O. Nyarko-Boateng, A. F. Adekoya, and B. A. Weyori, “Tracing the exact location of failures in underground optical networks using LSTM deep learning model,” *Indian Journal of Science and Technology*, vol. 14, no. 4, pp. 297–309, 2021. https://doi.org/10.17485/IJST/v14i4.2008
[11] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “Efficient stock-market prediction using ensemble support vector machine,” *Open Computer Science*, vol. 10, no. 1, pp. 153–163, Jul. 2020. https://doi.org/10.1515/comp-2020-0199
[12] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “Random forest based feature selection of macroeconomic variables for stock market prediction,” *American Journal of Applied Sciences*, vol. 16, no. 7, pp. 200–212, Jul. 2019. https://doi.org/10.3844/ajassp.2019.200.212
[13] F. Ecer, S. Ardabili, S. S. Band, and A. Mosavi, “Training multilayer perceptron with genetic algorithms and particle swarm optimization for modeling stock price index prediction,” *Entropy*, vol. 22, no. 11, Art no. 1239, 2020. https://doi.org/10.3390/e22111239
[14] Isha, S. Dixit, M. K. Ahirwar, D. Sakethnath, and M. Rakha, “Stock prediction by analyzing the past market trend,” in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, Sep. 2021, pp. 1–4. https://doi.org/10.1109/ICRITO51393.2021.9596263
[15] I. K. Nti and J. A. Quarcoo, “Self-motivation and academic performance in computer programming language using a hybridised machine learning technique,” *International Journal of Artificial Intelligence and Expert Systems*, vol. 8, no. 2, pp. 12–30, 2019.
[16] D. Aggarwal, S. Mittal, and V. Bali, “Prediction model for classifying students based on performance using machine learning techniques,” *International Journal of Recent Technology and Engineering*, vol. 8, no. 2S7, pp. 496–503, Jul. 2019. https://doi.org/10.35940/ijrte.B1093.0782S719
[17] S. Hussain, N. A. Dahan, F. M. 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, pp. 447–459, 2018. https://doi.org/10.11591/ijeecs.v9.i2.pp447-459
[18] H. Almarabeh, “Analysis of students’ performance by using different data mining classifiers,” *International Journal of Modern Education and Computer Science*, vol. 9, no. 8, pp. 9–15, Aug. 2017. https://doi.org/10.5815/ijmecs.2017.08.02
[19] K. T. Chui, R. W. Liu, M. Zhao, and P. O. De Pablos, “Predicting students’ performance with school and family tutoring using generative adversarial network-based deep support vector machine,” *IEEE Access*, vol. 8, pp. 86745–86752, May 2020. https://doi.org/10.1109/ACCESS.2020.2992869
[20] A. K. Hamoud, A. S. Hashim, and W. A. Awadh, “Predicting student performance in higher education institutions using decision tree analysis,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 2, pp. 26–31, 2018. https://doi.org/10.9781/ijimai.2018.02.004
[21] M. Akour, H. Al Sghaier, and O. Al Qasem, “The effectiveness of using deep learning algorithms in predicting students achievements,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 388–394, 2020. https://doi.org/10.11591/ijeecs.v19.i1.pp388-394
[22] Z. Ahmad and E. Shahzadi, “Prediction of students’ academic performance using artificial neural network,” *Bulletin of Education and Research*, vol. 40, no. 3, pp. 157–164, 2018.
[23] I. Burman and S. Som, “Predicting students academic performance using support vector machine,” in *2019 Amity International Conference on Artificial Intelligence (AICAI)*, Dubai, United Arab Emirates, Feb. 2019, pp. 756–759. https://doi.org/10.1109/AICAI.2019.8701260
[24] V. Cernat and L. Moldovan, “Emotional problems and academic performance of students in manufacturing,” in *11th International Conference on Interdisciplinarity in Engineering*, Tirgu Mures, Romania, Oct. 2017. *Procedia Manufacturing*, vol. 22, pp. 833–839, 2018. https://doi.org/10.1016/j.promfg.2018.03.118
[25] D. Oreski, S. Oreski, and B. Klicek, “Effects of dataset characteristics on the performance of feature selection techniques,” *Applied Soft Computing*, vol. 52, pp. 109–119, 2017. https://doi.org/10.1016/j.asoc.2016.12.023
[26] I. H. Sarker, “Machine learning: Algorithms, real-world applications and research directions,” *SN Computer Science*, vol. 2, no. 3, Art no. 160, Mar. 2021. https://doi.org/10.1007/s42979-021-00592-x
[27] D. C. T. Hernández, “An experimental study of K* algorithm,” *International Journal of Information Engineering and Electronic Business*, vol. 7, no. 2, pp. 14–19, Mar. 2015. https://doi.org/10.5815/ijieeb.2015.02.03
[28] A. M. Adam, “Sample size determination in survey research,” *Journal of Scientific Research and Reports*, vol. 26, no. 5, pp. 90–97, June 2020. https://doi.org/10.9734/JSRR/2020/v26i530263
[29] J. G. Carbonell, R. S. Michalski, and T. M. Mitchell, “An overview of machine learning,” in *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, Eds. Morgan Kaufmann, vol. 1, 1983, pp. 3–23. https://doi.org/10.1016/B978-0-08-051054-5.50005-4
[30] G. Bonaccorso, *Machine Learning Algorithms*. Packt Publishing Ltd., 2017.
[31] A. Dey, “Machine learning algorithms: A review,” *International Journal of Computer Science and Information Technologies*, vol. 7, no. 3, pp. 1174–1179, 2016.
[32] S. Abirami and P. Chitra, “Energy-efficient edge based real-time healthcare support system,” in *The Digital Twin Paradigm for Smarter Systems and Environments: The Industry Use Cases*. *Advances in Computers*, P. Raj and E. Preetha, Eds. Elsevier,vol. 117, no. 1, pp. 339–368, Oct. 2020. https://doi.org/10.1016/bs.adcom.2019.09.007

[33] H.-B. Ly, T.-A. Nguyen, and B. T. Pham, “Estimation of soil cohesion using machine learning method: A random forest approach,” *Advances in Civil Engineering*, vol. 2021, Art no. 8873993, 2021. https://doi.org/10.1155/2021/8873993
[34] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “A comprehensive evaluation of ensemble learning for stock-market prediction,” *Journal of Big Data*, vol. 7, no. 1, Art no. 20, Mar. 2020. https://doi.org/10.1186/s40537-020-00299-5
[35] 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, Feb. 2013. https://doi.org/10.5120/10489-5242
[36] P. Kaur, M. Singh, and G. S. Josan, “Classification and prediction based data mining algorithms to predict slow learners in education sector,” in *3rd International Conference on Recent Trends in Computing*, Delhi, India, Mar. 2015. *Procedia Computer Science*, vol. 57, pp. 500–508, 2015. https://doi.org/10.1016/j.procs.2015.07.372
[37] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “A systematic review of fundamental and technical analysis of stock market predictions,” *Artificial Intelligence Review*, vol. 53, no. 4, pp. 3007–3057, Aug. 2019. https://doi.org/10.1007/s10462-019-09754-z
[38] A. A. Taha and S. J. Malebary, “An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine,” *IEEE Access*, vol. 8, pp. 25579–25587, Feb. 2020. https://doi.org/10.1109/ACCESS.2020.2971354

Bridgitte Owusu-Boadu hold Higher National Diploma (HND) Marketing and B. Sc. Human resources from Kwame Nkrumah University of science and Technology (KNUST). Current Bridgitte is the CEO of Brivink Consult and Technology. Her research interest is in human computer interaction, machine learning application in education, agriculture.
Email: bink.hope@gmail.com

Isaac Kofi Nti holds HND Electrical & Electronic Engineering, B. Sc. Computer Science, M. Sc. Information Technology and Ph.D. Computer Science. Dr. Nti is a Lecturer at the Department of Computer Science and Informatics, University of energy and natural resources (UENR), Sunyani Ghana. His research interests include artificial intelligence, energy system modelling, intelligent information systems and social and sustainable computing, business analytics and data privacy and security.
Email: Isaac.nti@uenr.edu.gh
ORCID iD: https://orcid.org/0000-0001-9257-4295

Owusu Nyarko-Boateng holds HND Electrical & Electronics Engineering, B. Sc. Computer Science, PGDE, and MSc Information Technology and PhD from University of Energy and Natural Resources, Ghana. Owusu Nyarko-Boateng has previously worked with MTN Ghana and Huawei Technologies (SA) for over ten (10) years. He has in-depth experience in telecommunications transmission systems, including fiber optics cable deployment for long-haul and short distance (FTTx), 2G BTS, WCDMA (3G), and 4G (LTE) plants installations and configurations. He also managed Huawei DWDM Optix8800 OSN, Huawei OptiX OSN1800 and optical distributor frames (ODF). His research areas: machine learning, artificial intelligence, computer networks and data communications, network security, fiber optics technologies, modelling transmission systems, 5G & 6G Technologies, Expert Systems, computational intelligence for data communications.
Email: owusu.nyarko-boateng@uenr.edu.gh
ORCID iD: https://orcid.org/0000-0003-0300-2469

Justice Aning received the M. Sc. degree in Information Technology from The Kwame Nkrumah University of Science and Technology (KNUST), Ghana, in 2017. He is currently a lecturer with the Department of Computer Science, Sunyani Technical University, Sunyani, Ghana. J. Aning has authored more than five papers in journals. His current research interests include Web design, Machine learning, intelligent systems for modelling and optimization.
Email: aning421@gmail.com

Victoria Boafo holds M. Sc. Information Technology from The Kwame Nkrumah University of Science and Technology (KNUST). Currently she teaches Information Communication and Technology (ICT) Mampong Technical College of Education, Mampong, Ghana. Her research interest is in human computer interaction, machine learning application in education and Information and communication technology.