Performance Analysis of Machine Learning Algorithms in Prediction of Student Academic Performance

.

ABSTRACT

The advancement in technology has contributed largely to the application of data mining in education in recent times. However, selecting appropriate algorithm(s) to “mine” knowledge about educational data presents a difficult challenge to researchers and analyst. This paper contributes to the use of classification algorithms in academic performance prediction. The predictive ability of four popular algorithms; C4.5 Decision tree (CDT), Multilayer Perceptron (MLP), Naïve Bayes (NB) and Random Forest (RF) algorithms were compared. The models were built using student dataset from selected private senior high schools in Ghana. The comparative analysis of the algorithms was made based on their Accuracy, Recall, Specificity, F-Measure and Running time. On all the training and test ratios; 80:20, 70:30 and 10-fold cross validation, the results indicated that all the algorithms performed well in the classification. However, the Naïve Bayes algorithm performed significantly better than the MLP and CDT on some ratios. The running time of the NB, CDT and RF were the quickest while MLP took the longest time.

Keywords: Data Mining, Algorithms, Machine Learning, Classification, Prediction, Student Performance, Multilayer Perceptron Algorithm, Naïve Bayes Algorithm, C4.5 Decision Tree

1. INTRODUCTION

Data mining is currently the subject of extensive research. It can be defined as the systematic discovery of facts and patterns in large data collections. Data mining techniques can be used in a variety of industries, including business, medicine, and fraud detection. Recent years have seen the emergence of the multidisciplinary research field known as "Educational Data Mining (EDM)," which applies data mining to the field of education. Understanding student behavior and learning as well as enhancing educational results are the key objectives of EDM [10]. The academic performance of students to some extent is vital to their success in society. Students who are successful academically are more likely to have higher chances of getting good employment opportunities. These students are often well prepared to meet the demands of the constantly changing world than poor performing students.

As the number of students continue to increase, the data available has also increased exponentially. The use of these data by educational institutions has often been to make simple searches and prepare reports. However, more can be done with this data to help with decision making. Institutions can use knowledge extracted from data to determine or predict students' performance, attitude towards learning, possible school dropouts, etc. By using data mining techniques to process educational data, it is now possible to derive rules and predictions about students. Predictions can be made about a student's success or failure using a variety of data regarding the student's socioeconomic surroundings, learning environment, previous grades, or course notes.

There are several data mining approaches that can be used on a dataset depending on the task that requires knowledge. Choosing a suitable modeling approach for a given task is important when building models for prediction. One of such methods is Classification which is a supervised learning method. This approach of data mining maps training instances to pre-determined classes. A training set of pre-determined classes is studied by a learning algorithm in the first phase of classification to develop a model and a test set is utilized in the second phases to measure the model's accuracy [12].

Many data mining algorithms such as Decision trees (DT) and Naïve Bayes (NB) have been used to solve various classification problems with varying results which allows for improvements by combining multiple techniques or altering the ratios of training and testing. Much work is still needed in classifying students' data due to its high dimensionality. This study seeks to assess the ability of four commonly used algorithms in prediction of students' academic success on a binary classification task.

We put forward the following contributions:

1. Pre-processing techniques employed on data from selected private senior high schools in Ghana.
2. Models built for prediction using C4.5 decision tree (CDT), Multilayer Perceptron (MLP), Naïve Bayes and Random Forest (RF)
3. Evaluation of the performance of the algorithms based on Accuracy, Specificity, Recall, F-Measure and Running time.

The findings of this study provide researchers and school administrators' direction in discovering and selecting algorithm(s) to classify student datasets in order to identify good and poor students. It also provides insight into the application of classification algorithms on real-world sets and bridges the gap between theoretically projected outcomes against those that were observed.

2. RELATED WORKS

Educational Data Mining makes it possible to find answers to fundamental issues about education using data relating to specific educational contexts [19]. Studies have been done in this area because of the relevance of this field to stakeholders in discovering answers to questions about education using data related to educational settings and making future decisions. Fernandes et al., [9] predicted the academic performance of students in a district of Brazil, who were in public schools. The prediction was made on two distinct datasets, with the first set consisting of students' pre-enrolment variables, and the second set, two months post enrolment academic data of students. The classification models were built using Gradient Boosting Machine. From the findings, student grades and "absences" were identified as the most relevant attributes for predicting students end of year academic performance.

Nguyen et al. [15] assessed the performance of Bayesian networks and Decision trees in prediction of undergraduate and postgraduate student academic achievement using data from two educational institutions. The decision tree algorithm achieved the best accuracy in predicting the performance of students. It was found that decision trees were suitable for finding students who excelled academically. In a related study to categorize student performance into distinct classes and stages, Bharti [5] also revealed that Nave Bayes delivers the best Specificity when compared to the C.5 decision tree. According to Osmanbegovi and Sulji [16], when predicting True Negative values, the Nave Bayes algorithm outperforms MLP and C.45.

Amjad, Al-Emran and Shaalan [2] conducted a study to review and analyze 420 data mining research articles between 2009 and 2018. Results from the study showed that decision trees, Naïve Bayes classifiers, and artificial neural networks were the most used data mining algorithms in predicting students' factors. Students' previous grades and class performance, students' e-Learning activity, students' demographics, and students' social information were also found to be variables that had a high correlation with students' academic performance prediction.

Prabha and Shanavas [18] conducted a study to compare the performance of Naïve Bayes, Multilayer Perceptron, ZeroR, C4.5 decision tree and Random Forest using dataset collected from "Maths Tutor", an e-learning tool to find the best algorithm in predicting students' performance. The records of 120 students from the sixth standard were used to evaluate the performance of the algorithms using five (5) different criteria; number of correctly classified instances, mean absolute error and RMSE rates, the time taken to build the model and the ROC area. The 10-fold cross validation was used for classification. From the results, the Multilayer Perceptron and C4.5 decision tree achieved the highest accuracy and ROC curve weight average of 1, however, the error rates and time taken by the C4.5 decision tree were significantly lower than the MLP. Fan, Liu and Chen [8] designed a Physical Education (PE) management system and applied Decision tree ID3 algorithm to analyze PE students' performance in a way that would benefit physical education instruction. According to the findings, using the management system to analyze students' physical education performance can significantly lessen the effort that PE teachers must bear while also allowing them to focus more on the quality of physical education.

Kabakchieva [13] evaluated algorithms from four different families in predicting the performance of students at a Bulgarian university based on their personal and pre-university characteristics. The Naïve Bayes, Bayes net, OneR, JRip, C4.5 decision tree and K-Nearest Neighbor algorithms were applied in classifying students into variety of classes. To evaluate the models, the True Positive Rate and Precision were used as criteria. The results indicated that C4.5 decision tree was the best algorithm as it achieved the highest overall accuracy. Cortez and Silva [6] concluded that Decision Tree, Random Forest, and Neural Networks had a high predictive accuracy in classifying students' performance into two and five classes. They compared the predictive accuracy of these four algorithms in predicting students' failure in Mathematics and Portuguese.

3. METHODOLOGY

3.1 Study Area and Data Collection

The study was carried out in the Kwahu West Municipal District of Ghana. The municipal lies between latitudes 6°30° North, and 7° North and longitudes 0° 30° West and 1° West. It has an approximate total land area of 414km². There are 17 communities in the municipal with Nkawkaw as its capital. Five (5) private senior high schools were selected randomly from the municipal. Data was obtained from 456 students of the senior high schools. The dataset consisted of students' socio economic and prior enrollment variables. Table 1 shows the detailed description of the dataset.

Table 1. Description of Variables

Variables	Description	Possible Values	Data Type
SSG	Junior High School Grade in Social Studies	1, 2, 3, 4, 5, 6, 7, 8, 9	Numeric
SG	Junior High School Grade in Science	1, 2, 3, 4, 5, 6, 7, 8, 9	Numeric
EG	Junior High School Grade in English Language	1, 2, 3, 4, 5, 6, 7, 8, 9	Numeric
MG	Junior High School Grade in Mathematics	1, 2, 3, 4, 5, 6, 7, 8, 9	Numeric
fInc	Income level of family	High, Medium, Low	Nominal
mumEdu	Education level of mother	No formal education, Primary, Junior High School, Secondary, Tertiary	Nominal
fatherEdu	Education level of father	No formal education, Primary, Junior High School, Secondary, Tertiary	Nominal
mStatus	Parents' marital status	Single, Married	Nominal
fSize	Size of family	<=3, >3	Nominal
Sex	Sex of student	Male, Female	Nominal
eStatus	Employment status of parent	Employed, Unemployed	Nominal

3.2 Classification Framework

In order to "mine" information from data, several important processes must be completed. A data mining process may be based on a specific framework. Figure 1 presents the study's framework. This consist of handling data with missing values and outliers, feature selection, implementation of algorithms, comparison, and results. The framework was adopted from

[20] with some moderations. The Feature selection stage was introduced, as results from literature show increased accuracy of machine learning algorithms when optimal variables are used in creating models.

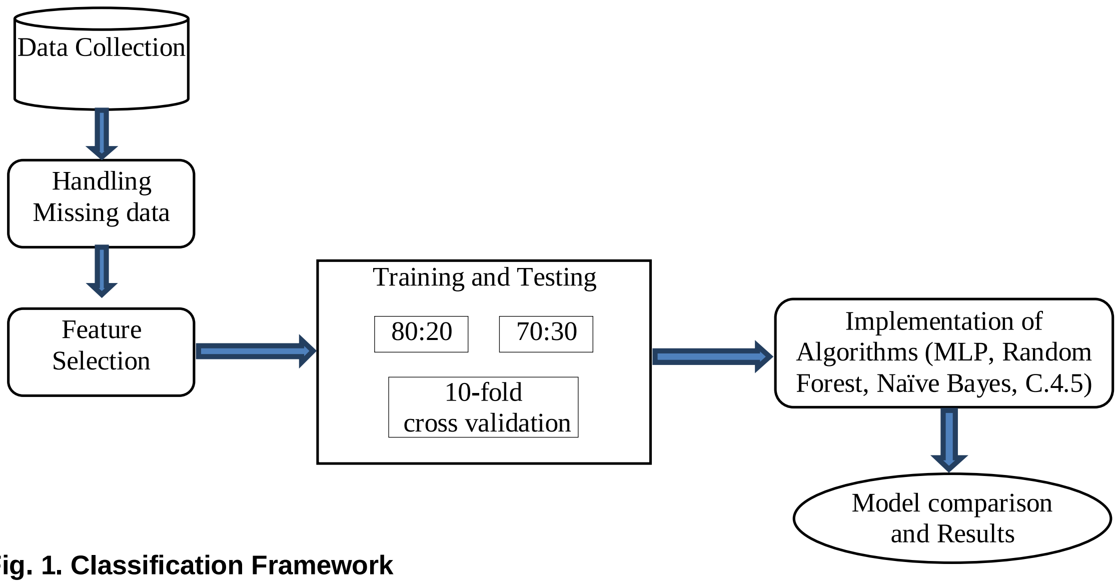

Fig. 1. Classification Framework

3.3 Data Pre-processing

In order to "mine" information from data, several important processes must be completed. A data mining process may be based Each figure should have a caption. The caption should be concise and typed separately, not on the figure area. Figures should be self-explanatory. Information presented in the figure should not be repeated in the table. All symbols and abbreviations used in the illustrations should be defined clearly. Figure legends should be given below the figures.

3.4 Feature Selection

Reducing total attributes in a dataset by removing unimportant variables can contribute to improved performance of data mining algorithms [14]. In addition to the full dataset (FSet), a reduced dataset (SetA) containing five highest ranked attributes were generated and used for the classification. The value of Information gain (IG) of attributes were used to rank the variables. Table 2 shows the two distinct datasets. To calculate Information gain, the posterior entropy was subtracted from the Class entropy. Entropy measures the degree of "impurity". When it's close to 0, it suggests the dataset contains less impurity. The entropy is reduced the most by an attribute that contains the most information.

Information gain (IG(Y)) and Entropy were calculated using the formula:

$$IG(Y) = Info(X) - Info_y(X)$$

Where $Info(X)$ is the Entropy of the class and $Info_y(X)$ is the Posterior Entropy.

$$Entropy, Info(X) = -\sum_{j=1}^{c} p(X_j) \log_2 p(X_j)$$

Where X is the data sample and $p(X_j)$ is the proportion of X in respect to class j.

Table 2. Datasets for Classification

Set	Number of Variables	Variables
FSet	11	EG, SSG, SG, MG, eStatus, mumEdu, fatherEdu, mStatus, fInc, fSize, Sex
SetA	5	EG, SSG, SG, MG, mumEdu

3.5 Training, Testing and Implementation of Algorithms

To carry out the experiments, models were built from two datasets; FSet and SetA, each containing different combinations of variables using Decision tree, Multilayer Perceptron, Naïve Bayes and Random Forest algorithms. The models were built by the algorithms using training and test ratios of 80:20, 70:30 and 10-fold cross validation. In building models using the 80:20 and 70:30 ratios, the dataset was split into two (training: testing) for training and testing purposes. The 10-fold cross validation divides dataset into 10 mutually exclusive subsets of approximately equal size. The training and testing of the algorithm are done 10 times, with one subset serving as the test set and the remaining nine serving as the training set each time. To calculate for the classifier's accuracy, the total number of accurate cases is divided by the total classification. Weka was used for the implementation of the algorithms. The details of each algorithm are presented in this session.

3.5.1 C4.5 Decision Tree

The C4.5 algorithm learns and generates trees from a training dataset that is used to classify new datasets. The C4.5 algorithm comes with added features and methods which make it easier to create simpler trees without compromising accuracy. It also handles both continuous and discrete data and uses pessimistic pruning to improve classification accuracy by removing non-essential branches from the decision tree. The root node of the tree, which is the Class's best splitting attribute was expanded and used to partition samples into many parts.

The information gain was computed for each variable and the one with the highest value was chosen as the root node. When selecting the next split attribute, the same procedure was utilized, without considering previously used split attributes. This was repeated until all the attributes had been used or the classification was completed. Overfitting was avoided by pruning or removing non-essential nodes. Pruning was done after generating the trees by substituting a child node for a parent node if the error rate validation does not reduce.

3.5.2 Naïve Bayes

The Bayes theorem provides the foundation for the Naive Bayes algorithm which assumes a high level of independency. It assumes that attributes are conditionally independent. In supervised settings, NB performs simple calculations and learns quickly. The NB algorithm predicts the class of unknown dataset using the formula:

$$p(c_i | s) = \frac{p(s| c_i)p(c_i)}{p(s)}$$

Where $p(c_i | s)$ = the posterior probability of c_i given s, $p(s| c_i)$ = the likelihood of s occurring given c_i,
$p(s)$= the likelihood of s occurring and
$p(c_i)$ = the likelihood of c_i occurring.

3.5.3 Multilayer Perceptron

Multilayer Perceptron algorithm consists of layers that cycle through weights of supplied data and maps them to correct outputs. It is from the family of Artificial Neural Networks. The MLP is a widely used algorithm for classification that locates weight that capture hidden Input / Output mapping in a set of instances. The number of weights to be updated by the algorithm was set at 0.3. This enables the training of the network while the weight amount is controlled. The momentum rate was set at 0.2 to prevent the network from reaching a local minimum too soon and conversely avoid overshooting the function's global minimum.

3.5.4 Random Forest

Random Forest is an ensemble machine learning algorithm that uses several classification trees, thus earning it the name, Random Forest. RF algorithm is considered to be a highly rigorous learning algorithm in recent times, in spite of its biasness towards classification of high-level categorical attributes of datasets consisting of varied levels of categorical data [7]. The algorithm functions by first gathering from the given dataset, random samples. Next, each sample is used to build a decision tree. The predicted results are then voted on to select the classification with the most votes. In training and testing of the two datasets, the maximum depth of trees configuration of the algorithm was set to 16.

3.6 Model Performance Evaluation

The algorithms were evaluated using the Accuracy, Recall, Specificity, F-measure and Running time. All these metrics, except for Running time were calculated using the classification results in the Confusion Matrix (see Table 3), where TP is True Positives, TN is True Negatives, FP is False Positives and FN is False Negatives.

Table 3. Confusion Matrix

Classified as			
Positive	**Negative**		
TP	FN	Positive	**← Actual Class**
FP	TN	Negative	

3.6.1 Accuracy

Accuracy is the proportion of correctly classified instances of a model. It was calculated as:

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

3.6.2 Recall

Recall is the percentage of correctly classified positive class samples. It was calculated as:

$$Recall = \frac{(TP)}{(TP + FN)}$$

3.6.3 Specificity

Specificity is the percentage of correctly classified negative instances. Specificity was computed as:

$$Specificity = \frac{(TF)}{(TF + FP)}$$

3.6.4 F-Measure

F-Measure is the harmonic mean of the Precision and Recall. The Precision and Recall values were used in calculating F-Measure. The formulas are as follow:

$$F - Measure = \frac{(2 \times Precision \times Recall)}{((Precision + Recall)}$$

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

4. RESULTS AND DISCUSSION

The Weka Experimenter was used to run the full datasets as well as the reduced set using the four algorithms to see if a base algorithm was considerably better or worse than the others. The Weka Experimenter examines algorithms outputs based on a set criterion. The Naïve Bayes algorithm was selected as the base algorithm against which the C4.5 decision tree, Random Forest and MLP algorithms were compared. A two-tailed t-test [14] with significant level of 0.5 was used for the experiment. A classifier which was significantly better than the base classifier was tagged with (v) and if worse, tagged with asterisk (*). No tag simply means the classifier was neither better nor worse as compared to the base classifier.

4.1 Classification Algorithms on FSet

Fig. 2a, 2b and 2c visualize the performance of the algorithms on FSet in terms of Accuracy, Recall, Specificity and F-Measure using training and test ratios of 80:20, 70:30 and 10-fold cross validation respectively. The results show that all four algorithms performed well on the dataset. The highest accuracy of the NB and RF were 86.76% and 80.95% respectively at 10-fold cross validation. The MLP and CDT’s highest accuracy was 81.11% and 85.08%, which was achieved at 80:20 and 70:30 ratios respectively. The classification accuracy of the algorithms on all the three ratios were significantly the same except for MLP at 10-fold cross validation, which was significantly worse than the base algorithm, NB.

In terms of the True positives or Recall, neither algorithms performed worse nor better than the base algorithm, NB. The algorithms showed very high results on all the different training and testing ratios. From the results, the NB and RF achieved highest Recall of 89% and 87% respectively at 10-fold cross validation. The MLP and C4.5 performed best in Recall at 70:30 ratio with 83% and 91% respectively. The performance of the NB, MLP, CDT and RF on all the training and testing ratio were consistent, with the algorithms achieving a very high rate of 80% and above. However, the MLP performed significantly worse than NB at 10-fold cross validation. It achieved an F-Measure of 81% as compared to the 89% of the NB algorithm. The results also shows that Specificity for NB, CDT, MLP and RF were not significantly different, except at 80:20 ratio where the CDT was significantly lower than the NB. The CDT achieved a Specificity of 71% as compared to the 86% by the base algorithm.

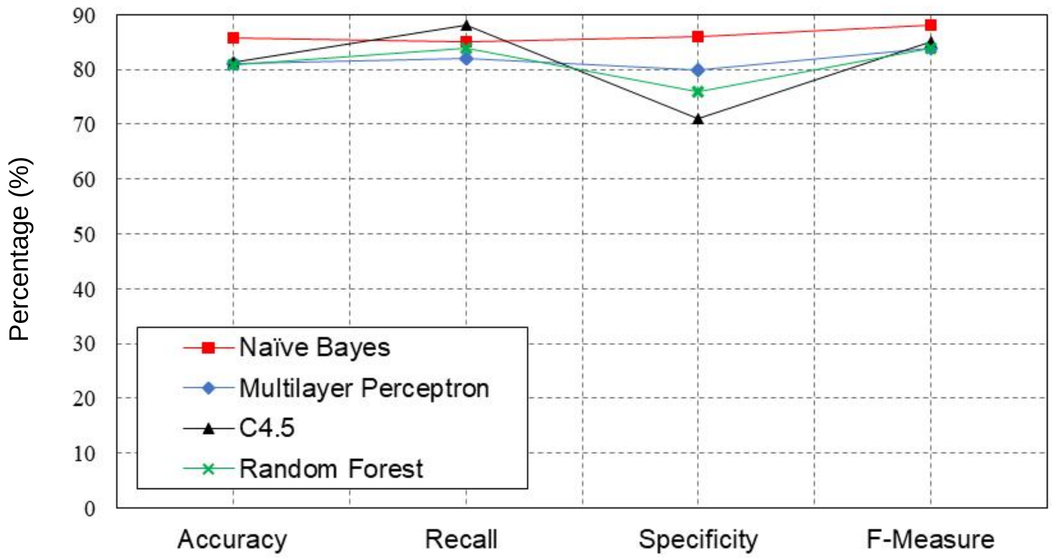

Fig 2a: Performance of Algorithms on 80:20 ratio (FSet)

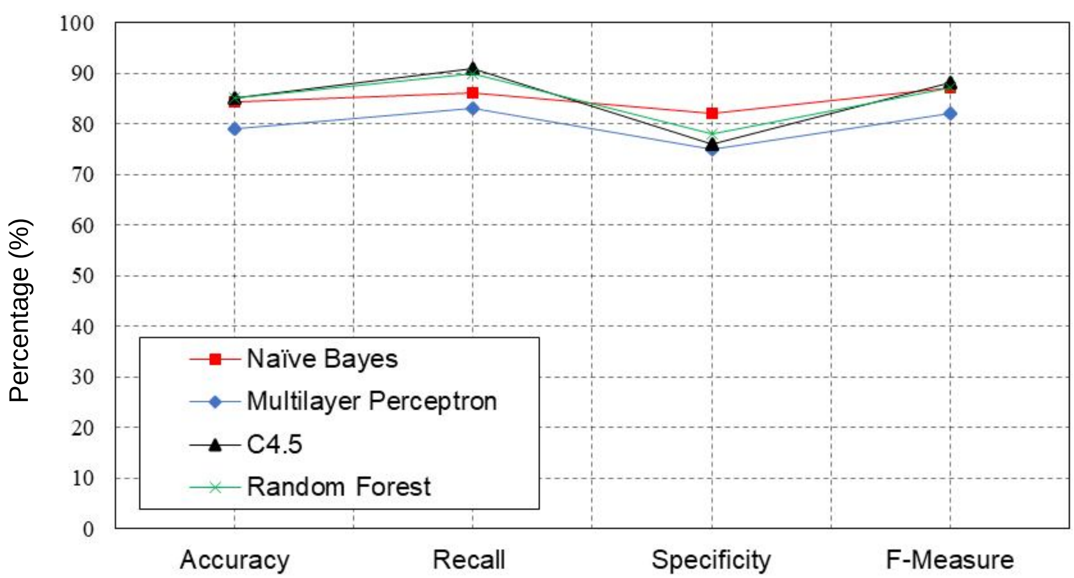

Fig 2b: Performance of Algorithms on 70:30 ratio (FSet)

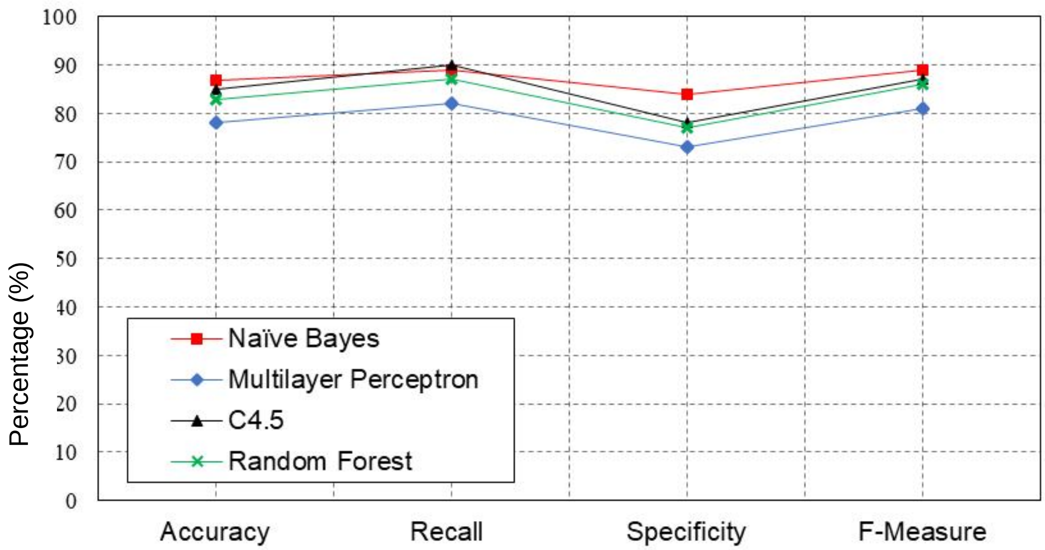

Fig 2c: Performance of Algorithms on 10-fold cross-validation (FSet)

The running time of the algorithms as presented in Table 4 were also significantly the same for NB, CDT and RF. However, the MLP took a longer running time than all the algorithms on all the ratios.

Table 4. FSet Running Time Results (in seconds)

Model	80:20	70:30	10-fold cross- validation
Naïve Bayes	0.01	0.01	0.01
Multilayer Perceptron	0.15*	0.14*	0.17*
C4.5	0.01	0.1	0.01
Random Forest	0.02	0.01	0.02

4.2 Classification Algorithms on SetA

The results for SetA are presented in Fig 3a, 3b and 3c. The results show that the classification accuracy of the algorithms were significantly the same. Each of the algorithms performed very well in the binary classification. At 70:30 ratio, the results of the NB and RF were 88.76% and 83.38% respectively which were the classifiers' best accuracy. The MLP performed best in Accuracy (82.63%) at 80:20 whereas CDT's was 83.94% at 10-fold cross validation.

The Recall and F-Measure achieved by the four algorithms were neither significantly worse nor better than the other. The NB, MLP, CDT and RF performed very well in classifying the dataset. In terms of F-Measure, all the algorithms performed better at 10-fold cross validation. The results of the NB, MLP, CDT and RF were 90%, 85%, 87% and 85% respectively. The results also show that the highest Recall by all the algorithms were achieved at 10-fold cross validation. The Recall for NB, MLP, CDT and RF were 89%, 86%, 89% and 86% in that order.

The Specificity of the algorithms on all the ratios were neither significantly worse nor better, with the exception of MLP which performed slightly lower (71%) than the base algorithm NB (93%) at 80:20. All the other algorithms did very well classifying True negatives.

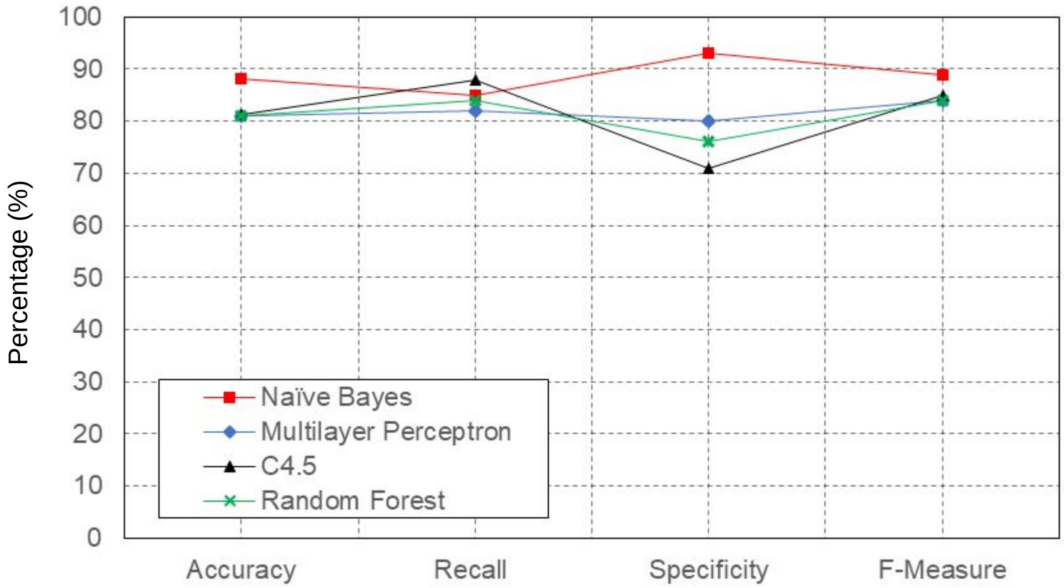

Fig 3a: Performance of Algorithms on 80:20 ratio (SetA)

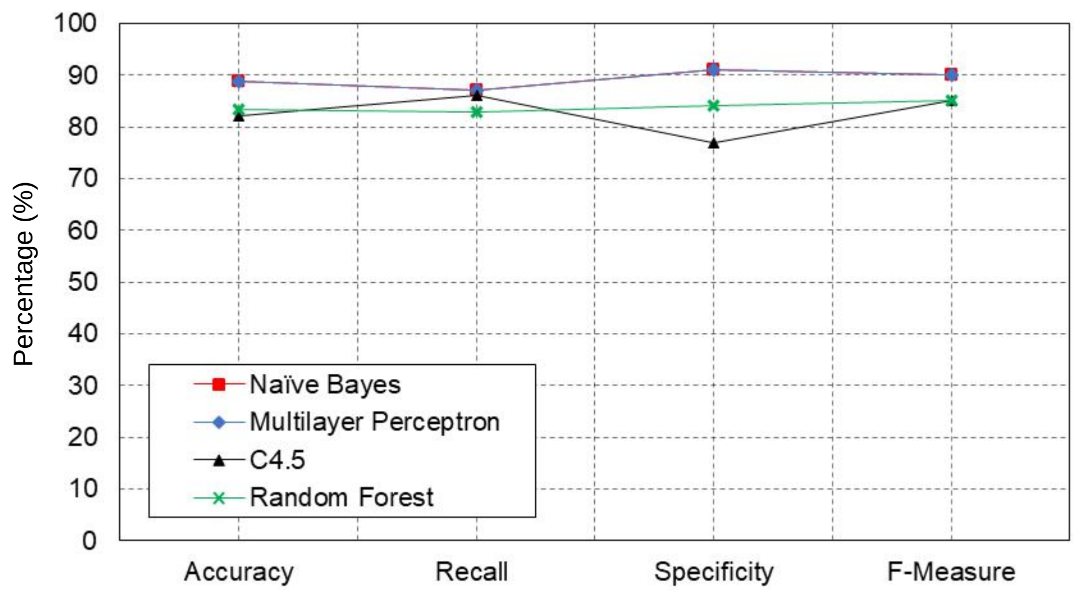

Fig 3b: Performance of Algorithms on 70:30 ratio (SetA)

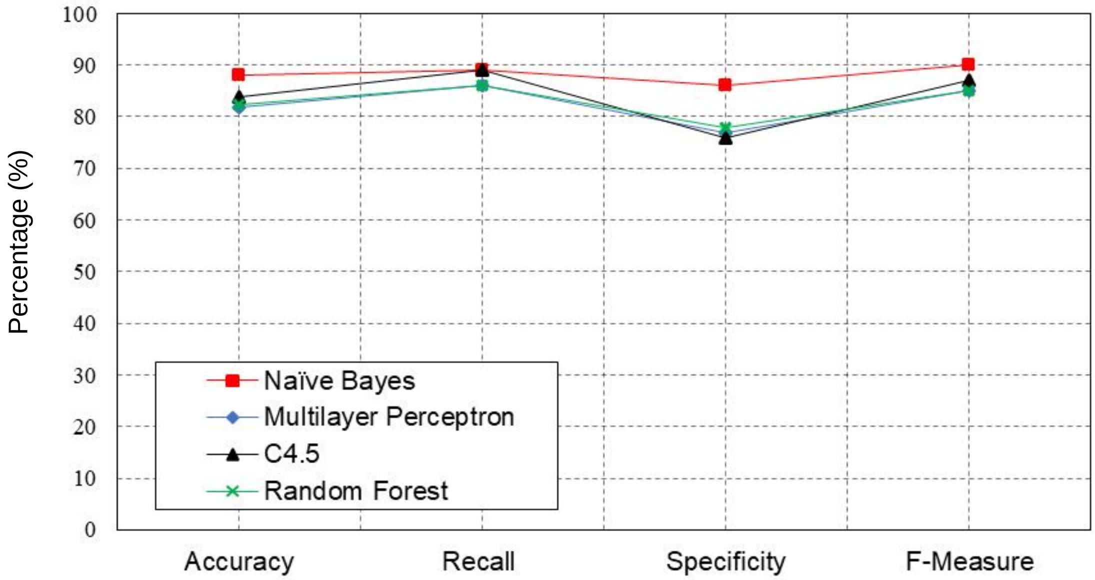

Fig 3c: Performance of Algorithms on 10-fold cross validation (SetA)

Also, the MLP, in terms of running time performed significantly worse than the three algorithms as shown in Table 5.

Table 5. SetA Running Time Results (in seconds)

Model	80:20	70:30	10-fold cross- validation
Naïve Bayes	0.01	0.01	0.01
Multilayer Perceptron	0.05*	0.04*	0.05*
C4.5	0.01	0.1	0.01
Random Forest	0.01	0.01	0.01

4.3 Critical Analysis of Classification Algorithms

The machine learning classifiers utilized in this work performed well on the two distinct datasets (FSet and SetA) using the different training and testing ratios. The NB, particularly, performed better on the reduced dataset (SetA) in classifying the Accuracy, Recall, Specificity and F-Measure on all the ratios. This is consistent with the findings of Anuradha and Velmurugan [3] which reported a high F-Measure for the Naïve Bayes algorithm in predicting the GPA of students using their grades. This could imply that Naive Bayes, as opposed to C4.5 and MLP, is more effective in predicting positive values. According to Ashari, Paryudi and Tjoa [4], Naïve Bayes is the best algorithm for classifying students' performance based on Precision, Recall, F- measure, Accuracy, and AUC as it could outperform Decision Tree and k-Nearest Neighbor. Also, Osmanbegovi and Suljic [16] reported that the Naïve Bayes algorithm is more efficient when compared with MLP and C.45 in predicting True Negative values.

Several reasons may have contributed to the performance of the NB algorithm. Even though most probability estimates are poor, it is possible for the right class to have the highest estimate, resulting in the accurate classification. The NB also has the capability to handle categorical data well, especially in small datasets than most algorithms [1]. The Maximum Likelihood approach that the Naive Bayes algorithm utilizes to categorize instances could be responsible for its good performance. This method chooses the set of values that gives the highest probable results for a given model and maps instances to the correct class so far as the class has the highest probability estimate.

From the result, there was no significant difference in the running time of the NB, RF and CDT. This is consistent with the results of [4] where decision tree was found to be very fast in terms of running time in building trees. The speed of NB could be because it has a linear learning time [17]. The results from this study suggest that NB, RF and CDT are computationally and resource-wise very efficient. Across all training sets, the MLP algorithm took much longer than the others. This could be because setting up MLP's network and learning from a training set requires more memory and runtime.

5. SUMMARY AND CONCLUSION

The study was aimed at evaluating the predictive ability of four popular algorithms; C4.5 Decision Tree, Random Forest, Naïve Bayes, and Multilayer Perceptron algorithms on a binary classification task. Assessment metrics which include Accuracy, Recall, F-Measure, Specificity and Running time were used in comparing the performance of the algorithms in predicting students' academic performance. Data was obtained from 456 students from senior high school in the eastern region of Ghana.

To ensure effective prediction of the four algorithms, the dataset was thoroughly analyzed and pre-processed before the classification stage. A new dataset (SetA) which consists of the 5 highest attributes with Information Gain was generated from the full dataset (FSet) and together used for the classification. The training and testing ratios used for building models by the algorithms were 80:20, 70:30 and 10-fold cross validation.

The findings from the experiments showed that all the algorithms exhibited a high level of performance in the classification. On the full dataset, all the algorithms did well on the 3 training and testing ratios performed, with the Naive Bayes algorithm achieving the best prediction performance. However, the Multilayer Perceptron achieved lowest performance. On the reduced dataset (SetA), the performance of the algorithms improved further which is suggestive that using optimal attributes of a dataset can help increase the performance of the algorithm. Generally, the Naïve Bayes algorithm was significantly better than the Multilayer Perceptron, C4.5 and Random Forest algorithms which suggests that it is a good candidate for predicting students' academic performance. More investigation needs to be done with a hybrid model at various training and testing ratios and other classification algorithms for predicting students' academic performance.

6. RECOMMENDATIONS

From the conclusions drawn in the study, the following are recommended for further study:

1. This study considered only selected students' pre-enrollment and socio-economic variables. The performance of the classification algorithms using factors such as students affect, motivation and other non-cognitive variables may be considered.

2. Different feature selection methods other than the one tested may be employed in ranking and selecting attributes to improve prediction. For example, wrapper methods of selecting attributes may be studied.

3. The use of multiple algorithms in an ensemble method may be used for increased performance in terms of prediction.

REFERENCES

[1] Al-Barrak M., Al-Razgan M. (2016). Predicting students' performance through classification: A case study. Journal of theoretical and Applied Information Technology, 75(2), 167-175.

[2] Amjad A. S., Al-Emran M., & Shaalan K. (2019). Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. Technology, Knowledge and Learning, 24(4), 567-598.

[3] Anuradha, C., Velmurugan T. (2015). A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance. Indian Journal of Science and Technology, 8(15),1-12.

[4] Ashari A., Paryudi I., Tjoa A. M. (2013). Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool. International Journal of Advanced Computer Science and Applications, 4(11), 33-39.

[5] Bharti T. (2014). Data Mining with Big Data Using C4.5 and Bayesian Classifier. International Journal of Advanced Research in Computer Science and Software Engineering, 4(8), 97-107.

[6] Cortez P., & Silva A. M. G. (2008). Using Data Mining to Predict Secondary School Student Performance. In A. Brito, & J. Teixeira (Eds.), Proceedings of 5th Annual Future Business Technology Conference, Porto, 5-12.

[7] Cutler A., Cutler D. R., & Stevens J. R. (2012). Random forests, in Ensemble Machine Learning. New York, NY, USA: Springer, pp. 157 175.

[8] Fan Y., Liu Y., Chen H., & Ma J. (2019). Data mining-based design and implementation of college physical education performance management and analysis system. International Journal of Emerging Technologies in Learning, 14(06), 87-97.

[9] Fernandes E., Holanda M., Victorino M., Borges V., Carvalho R., & Van Erven G. (2019). Educational data mining: Predictive analysis of academic performance of public-school students in the capital of Brazil. Journal of Business Research, 94:335-343.

[10] Ghorpade S. J., Patil S. S., & Chaudhari R. S. (2020). Educational Data Mining: Tools and Techniques Study. International Journal of Research and Analytical Reviews, 7(4): 520 -525.

[11] Guyon I., & Elisseeff A. (2003). An introduction to variable and feature selection. The Journal of Machine Learning Research, 3, 1157–1182.

[12] Han J., & Kamber M. (2006). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco

[13] Kabakchieva D. (2013). Predicting student performance by using data mining methods for classification, Cybernetics, and Information Technologies, 13(1), 61-72.

[14] Nadeau C., & Bengio Y. (2000). Inference for the generalization error. Advances in Neural Information Processing Systems 12, MIT Press.

[15] Nguyen T. N., Janecek P., & Haddawy P. (2007). A comparative analysis of techniques for predicting academic performance. Frontiers in Education Conference - Global Engineering: Knowledge without Borders, Opportunities without Passports, 2007. FIE '07. 37th Annual, pp.T2G-7, T2G-12.

[16] Osmanbegovi E., & Suljic M. (2012). Data mining approach for predicting student performance, Journal of Economic and Business, X (1): 3-12.

[17] Pazzani M., & Bilsus D. (1997). “Learning and Revising User Profiles: The Identification of Interesting Web Sites,” Machine Learning, Vol. 27, 313 – 331.

[18] Prabha S. L., & Shanavas A. R. M. (2015). Performance of Classification Algorithms on Students’ Data – A Comparative Study, International Journal of Computer Science and Mobile Applications, 3 (9), 1-8.

[19] Romero C., & Ventura S. (2010). Educational data mining: a review of the state of the art. IEEE Trans. On Syst. Man and Cybern. Part C Appl. Rev. 40(6), 601–618.

[20] Huang, Shaobo, Predictive Modeling and Analysis of Student Academic Performance in an Engineering Dynamics Course. (2011). All Graduate Theses and Dissertations. 1086.