Towards Prediction of Students' Academic Performance in Secondary School Using Decision Trees

Obadiah Matolo Musau[1], Kelvin Omieno[2], Raphael Angulu[3]

[1]*School of Computing and Informatics, Technical University of Mombasa, Kenya*

[2]*Department of Information Technology and Informatics, Kaimosi Friends University College, Kenya*

[3]*Department of Computer Science, Masinde Muliro University of Science and Technology, Kenya*

Abstract **- Prediction of students' academic performance with high accuracy is useful in many contexts. Institutions would like to know which students are likely to have low academic achievements or need assistance in order to finish their studies. Various machine learning techniques have been applied to create models to predict student's academic performance at various levels of study. This paper aimed to develop a machine learning model for prediction of secondary school students' academic performance. We collected records of 1720 former secondary school graduates from five public institutions in Kenya. Prediction was done by applying J48 Decision Tree, Naïve Bayes and Neural Networks Multilayer Perceptron classification techniques using WEKA machine learning environment. The study found out that J48 Decision Tree classifier predicted students' academic performance with higher accuracy than Naïve Bayes and Neural Networks classifiers. This knowledge will help educational institutions to accurately predict academic performance of the students.**

Keywords **-Prediction, J48 Decision Trees, Naïve Bayes, Neural Network, WEKA tool**

I. INTRODUCTION

In the recent past, application of machine learning techniques in the educational sector has grown exponentially. This has been spurred by the motivation that educators can now uncover new, interesting and useful insights about students (Livieris, Drakopoulou, Tampakas, Mikropoulos, & Pintelas, 2018). Machine learning has enabled the development of more accurate and efficient performance prediction models(Agrawal & Mavani, 2015; Usman & Adenubi, 2013) in the educational sector that have the ability to classify and identify weak students with low achievements than was previously possible(Ma, Yang, & Zhou, 2018; Kabakchieva, 2012;Moseley & Mead, 2008).

Machine learning represents promising areas of research in educational fields. With the abundance of educational datasets available, the demand for machine learning and other data mining techniques is in rise. Machine learning is often associated with educational data mining (EDM) which is an emerging interdisciplinary field that uses data mining and machine learning techniques to explore data from educational settings. The objective of EDM is to find out predictions and patterns that best characterize student's behaviour and performance. EDM is inspired by pattern recognitions and works by applying machine learning techniques to historical data to improve future decisions (Danso, 2006).

Students' success in learning is linked to several factors(UNICEF, 2009) that include student demographics, educational background, psychological, student academic progress and other environmental variables (Guo, et al., 2015). However, predicting performance of students that have diverse factors require more customized approach to address the diversity (Xu, Han, Marcu, & Schaar, 2017). Critical move towards attaining students' success in learning is to build a model that can continuously track and accurately predict students' future academic performance, such as what are they likely to get in final examination, given current and previous performance (UNICEF, 2009; Asif, Merceron, & Pathan, 2014).

In this paper, we propose the application of J48 Decision Tree for predicting student's academic performance. We evaluated the classification accuracy of J48 Decision Trees against other well-known classifiers; Naïve Bayes and Neural Network. Our objective was to identify which machine learning algorithm gave the best prediction accuracy. Decision Tree algorithm performed better than the other two algorithms.

II. LITERATURE REVIEW

Several studies have addressed the topic on student performance prediction. Livieris et al. (2018) conducted a study to predict secondary school students' academic performance in final examinations in the course of Mathematics. The study compared the effectiveness of two wrapper-based semi-supervised learning approach: self-training and Yet Another Two Stage Idea (YATSI) methods with neural network classifier in prediction of performance. The input data consisted of records of 3,716 students collected by the Microsoft showcase school Avgoulea-Linardatou between 2007 and 2016. The findings revealed that use of semi-supervised algorithms which utilize fewer labeled and many unlabelled data helps improve prediction accuracy and develop reliable prediction models.

Paulo & Silva (2008) applied Business Intelligence and Data Mining techniques to predict performance in mathematics and Portuguese language courses for secondary school student's. The input variables included mark reports, students' demographic, social and school related attributes such as student's age, alcohol consumption, and mother's education. Decision Trees, Random Forest, Neural Networks and Support Vector Machine techniques were applied to construct the performance prediction model. Oladokun et al. (2008) applied Artificial Neural Network to build a model for to build a model for prediction of secondary school students academic performance before being considered for university admission based on the Multilayer Perceptron Topology. The input variables included parental background, gender, ordinary level subjects' scores, subject's combination, matriculation examination, scores, type of school, location of school and age on admission from University of Ibadan-engineering department. The results showed that the model predicted more than 70% of prospective students' academic performance correctly.

Osmanbegović and Suljić (2012) compared the performance of Bayesian classifier, neural networks and decision trees in predicting student performance. The classifiers were applied data collected from students of the Faculty of Economics in Bosnia and Herzegovina in a survey conducted between 2010- 2011. The results showed that Naïve Bayes classifier outperformed decision tree and neural network methods in prediction. Khasanah and Harwati (2017) conducted a comparative study to predict students' academic performance using Bayesian network and decision tree classification algorithms. The data consisted of 178 student data collected from student data base from Universitas Islam Indonesia's information system. The best prediction was obtained from Bayesian network classification algorithm.

Khan, Hayat, & Daud(2015) applied J48 decision tree algorithm on student data containing previous performance to build a model to predict the student final grade based on Secondary School Certificate (SSC) – part one marks from Islamabad Capital Territory in Pakistan. The required data was extracted from Federal Board of Intermediate and Secondary Education student database for the years 2005, 2006, 2009, 2010, and 2012. The predictive model obtained was able to correctly classify 1268 student out of 1500 with a prediction accuracy of 84.53%. Sharma and Santosh (2017) developed a model based on previous student performances to predict final student performance by applying the ID3 decision tree algorithm on student data from Gyan Ganga Institute of Technology and Sciences in India. Kabakchieva (2012) used Decision Tree, Neural Network and the k-Nearest Neighbor algorithms in a comparative study to develop a student performance prediction model for Bulgarian universities. The decision tree algorithm outperformed the other algorithms.

III. MACHINE LEARNING ALGORITHMS

According to Baradwaj & Pal(2011), the four main types of machine learning are supervised learning, unsupervised learning, semi-supervised learning and reinforced learning. In supervised learning, the objective is to build a prediction model for predicting the true labels of unseen future data(Livieris, Drakopoulou, Tampakas, Mikropoulos, & Pintelas,2018). The input dataset consists of labeled data (Dey, 2016). There are two types of problems in supervised learning; classification and regression problem (Baradwaj & Pal, 2011;Karthikeyan & Kavipriya, 2017). In this work, we applied the most commonly used supervised machine learning classifiers, decision trees, naïve bayes and neural networks, in predicting student performance (Saa, Al-Emran & Shaalan, 2019) to build the prediction model.

A. Decision Trees

Decision tree is a classification algorithm which uses a tree structure to build classification models(Khan, Hayat, & Daud, 2015). The tree is built through a recursive process which breaks down the set of training data into discrete groups in order to maximize the distance between groups. The tree consists of nodes and branches. The nodes represent attributes while the branches represent the values each node can take (Baradwaj & Pal, 2011; Dey, 2016). C4.5 is a common decision tree learning algorithm. Decision-tree based models require lots of data for the algorithms to get properly trained. Lack of data therefore may be a plausible explanation why some studies produce mixed outcomes in regard to prediction model performance(Xing, Rui, Eva, & Sean, 2015). The method is thus not suitable for prediction in smaller datasets. C4.5 decision tree algorithm is an implementation by J48 algorithm in WEKA software tool.

B. Naïve Bayes

Naive Bayes is a classification algorithm which consists of a collection of simple probabilistic classifiers which are based on Bayes' theorem. The algorithm makes two assumptions: that the predictive attributes are conditionally independent with familiar classification and; that there are no hidden attributes that could interfere with the process of prediction.Naive Bayes is a very robust model which has quite often outperformed sophisticated models. It provides a very efficient algorithm for data classification(Osmanbegović & Suljić , 2012).

C. Neural Networks

The Neural Network algorithm mimics the structure of the human brain (Agrawal & Mavani, 2015).It consists of a set of highly interconnected entities that mimic the human neurons referred to as processing unit (or artificial neuron). The processing units are interconnected (through synapses) to transmit signal from one neuron to another. The processing units have the ability to receive or accept a set of inputs (signals), process it and respond with an output to the neurons connected to it (Usman & Adenubi, 2013). A neuron has two modes of operation: the first mode is called the training mode whose objective is to determine the input-output mapping. This is achieved through training the network using a set of paired data to allow the neuron learn when to fire and when not to fire. The second mode is the "using mode" where the weights of the connections between neurons are then fixed and the network is used to determine the classifications of a new set of data(Livieris, Drakopoulou, & Panagiotis, 2012). At this point, the neuron will detect and fire the output associated to any input pattern. However, if the input pattern is not among the list of the taught input patterns, then the firing rule is applied to decide whether the neuron will fire or not fire. The signal is represented in form of a real number at any connection between the neurons.

The neurons and connections normally have a weight that keeps on adjusting itself as learning proceeds by either increasing or decreasing the strength of the signal at a connection. Typically, the neurons may be assigned some threshold, in such a case, the signal is fired only if the aggregate signal crosses the threshold. Neurons in a neural network are usually organized in layers. The input signal traverses through all the layers from the first layer (also called input layer) through the network layers to the last layer (also called output layer). Each layer is designed to perform certain kinds of transformations on the input signal or data. Where necessary the traversal may traverse iteratively. Neural Networks have the capability of self-learning and self-adapting which makes it to be more efficient and accurate than other classification techniques (Baradwaj & Pal, 2011; Mitchell, 1997; Xing, Rui, Eva, & Sean, 2015).Multilayer Perceptron (MLP) algorithm is a popular neural networks and the widely used.

This study used 10-fold cross-validation to evaluate the model. Cross validation has over the years been used as a standard way of evaluating the performance of machine learning algorithms due to its ability to reduce variance.

IV. DATA DESCRIPTION

The source data for this study consisted of secondary school student's data which was collected through a questionnaire between January 2019 and April 2019 from five public institutions in Kenya. The initial data collected consisted of 1720 instances and each instance consisting of 60attributes. Using feature selection techniques, the number of attributes found to be more significant in predicting the class (KCSE) attribute were 15 out of 60 attributes. In Kenya, secondary school education consists of 4 years of schooling preceding 8 years of primary school education. The grading system in Kenya is an expanded letter grade ranging from A to E as follows: A is expanded to A, A-; B is expanded to B+, B, B-; C is expanded to C+, C, C-; D is expanded to D+, D, D- and E which is not expanded. This grades are based on a numeric 12-point scale where A is equivalent to 12 points representing excellent and, E is equivalent to 1 point representing poorest.

The mode of evaluation is an end of cycle (exit) examination called Kenya Certificate of Secondary Education (KCSE) administered at the end of the four years of secondary schooling. The aim of this study was to predict the KCSE grade a student is likely to score given previous performances, student demographic features and learning environmental factors. Table I shows the names, code and domain of the attributes for the model.

TABLE I ATTRIBUTES INFORMATION

Code	Name	Domain
MG	Mock Grade	{a,b,c,d,e}
F3G	Form 3 Grade	{a,b,c,d,e }
F2G	Form 2 Grade	{a,b,c,d,e }
F1G	Form 1 Grade	{a,b,c,d,e }
ME	Mothers Education	{none(1),primary education(2),secondary education(3),postsecondary (4),degree and above(5)}
FE	Father's Education	{none(1),primary education(2),secondary education(3),postsecondary (4),degree and above(5)}
NSF1	Subjects in Form 1	{Numeric}
NSF2	Subjects in Form 2	{Numeric}
AS	Assessment Style	{ formal(1),informal(2),all(3)}
Religion	Religion	{muslim,christian,others}
DF	Difficulties Paying Fees	{yes,no}
Internet	Access to Internet	{yes,no}
EC	Examination Challenges	{yes,no}
CL	Computer Laboratory	{yes,no}
KCSE	KCSE Grade (Class Attribute)	{a,b,c,d,e}

The distribution of the grades in the four years of secondary school is shown in Table II.

TABLE II DISTRIBUTION OF GRADES BY LEVEL

Grade	Level				
	Form 1	**Form 2**	**Form 3**	**Mock**	**KCSE**
A	725	160	57	47	61
B	800	1077	658	590	473
C	184	456	949	988	1029
D	9	25	56	94	153
E	2	2	0	1	4
Total	**1720**	**1720**	**1720**	**1720**	**1720**

A comparison of performance in each year of study is shown in Figure 1.

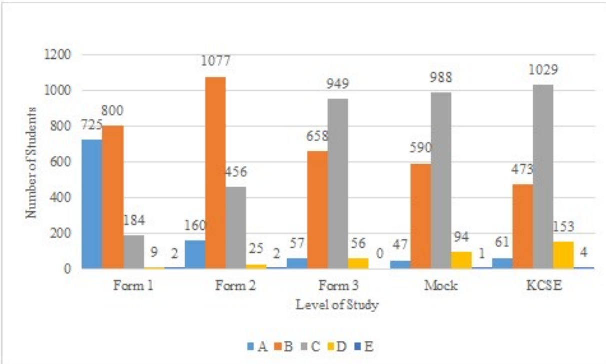

Fig.1 Distribution of Grades

As shown in Figure 1, it can be seen that the performance at Form 3 and Mock examination is almost similar to that of KCSE.

V. Results and Discussions

For this study, WEKA (Waikato Environment for Knowledge Analysis) software package was used. WEKA is an open-source software that was developed by Waikato University in New Zealand (Osmanbegovic & Suljic, 2012). It provides a collection of machine learning algorithms and stores data in a flat file format called ARFF (Attribute Relation File Format). WEKA is used under the GNU license for knowledge analysis (Osmanbegovic & Suljic, 2012). In order to get a better understanding of the significance of each variables to the output variable, we analysed the impact of each variables in relation to students' prediction success. To achieve this, we conducted several test using feature selection techniques; Gain Ratio, Info Gain and One R-test. The top ranked attributes were selected as shown in Table II.

We conducted several experiments in order to find out the predictive performance of each classifier for predicting students' academic performance. The predictive performance of each classifier were evaluated using several performance metrics that included prediction accuracy, correctly classified instances, incorrectly classified instances, precision and recall. Table III shows a summary the results obtained in each experiment.

TABLE III PREDICTIVE PERFORMANCE OF EACH CLASSIFIER

Evaluation Metric	Classifier		
	Naïve Bayes	J48	Multilayer Perceptron
Accuracy	70.1%	73.0%	66.7%
Correctly Classified Instances	1206	1256	1148
Incorrectly Classified Instances	514	464	572
Precision	0.695	0.719	0.657
Recall	0.701	0.730	0.667
F-Measure	695	0.714	0.661
ROC Area	0.797	0.758	0.746

From the results of the three classifiers as shown in Table III, it can be seen that in terms of prediction accuracy, J48 decision Tree classifiers achieved 73.0 % while Naïve Bayes classifier and Multi Perceptron classifiers achieved 70.1% and 66.7% prediction accuracy respectively.

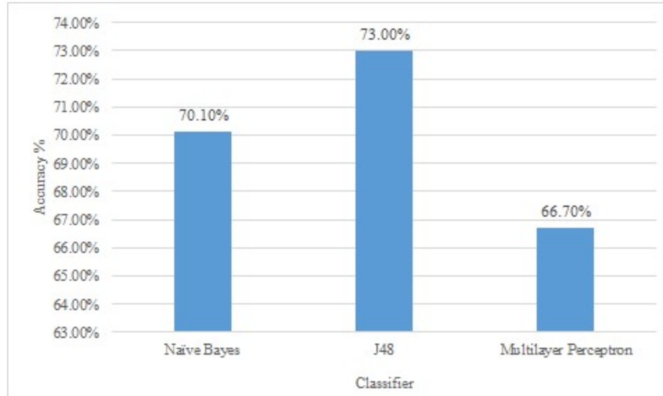

Fig.2 Prediction Accuracy of Classifiers

Figure 2 shows that J48 Decision Tree classifiers predicts better followed Naïve Bayes classifier while Multi Perceptron classifiers achieved the lowest prediction accuracy.

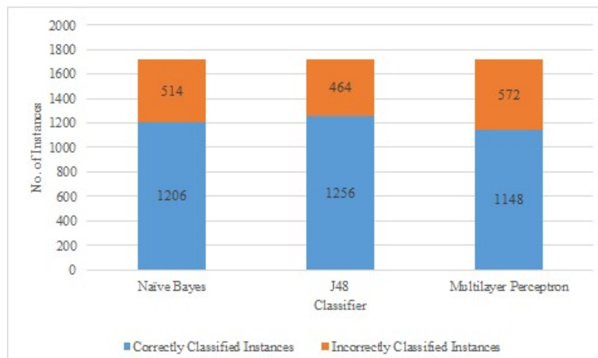

Fig. 3 Classification Error

Figure 3 shows the classification error of the three models. J48 decision Tree correctly classified 1256 instances out of 1720 instances, and 464 instances were incorrectly classified. Naïve Bayes classified 1206 instances correctly and 514 instances were incorrectly classified, Multilayer Perceptron correctly classified 1148 instances and incorrectly classified 572 instances out of 1720 instances.

VI. Conclusion

This study compared the performance of J48Decision Tree, Naïve Bayes and Multi Perceptron Neural Network algorithms inpredicting students' KCSE grade in secondary school. The performance metrics used to evaluate the models were accuracy, error rate, precision and recall. The results showed that J48 algorithm performed better in prediction of student's academic performance in secondary schools.

REFERENCES

[1] Livieris, I. E., Drakopoulou, K., Tampakas, V. T., Mikropoulos, T. A., &Pintelas, P. (2018). Predicting secondary school students' performance utilizing a semi-supervised learning approach. Journal of Educational Computing Research, 0735633117752614.

[2] Agrawal, H., &Mavani, H. (March 2015). Student Performance Prediction using Machine Learning. International Journal of Engineering Research & Technology (IJERT) , 4(03), 111-113.

[3] Usman, O., &Adenubi, A. (October 2013). Artificial Neural Network (ANN) model for predicting students' academic performance. Journal of Science and Information Technology, 1, 23-37.

[4] Ma, X., Yang, Y., & Zhou, Z. (April 2018). Using Machine Learning Algorithm to Predict Student Pass Rates In Online Education. Proceedings of the 3rd International Conference on Multimedia Systems and Signal Processing (pp. 156-161). ACM.

[5] Kabakchieva, D. (2012). Student performance prediction by using data mining classification algorithms. International Journal of Computer Science and Management Research, 1(4), 686-690.

[6] Moseley, L. G., & Mead, D. M. (2008). Predicting who will drop out of nursing courses: a machine learning exercise. Nurse education today, 28(4), 469-475.

[7] Danso, S. O. (2006). An exploration of classification prediction techniques in data mining: the insurance domain. Master Degree thesis.

[8] UNICEF, W. (2009). Abolishing School Fees in Africa: Lessons from Ethiopia, Ghana, Kenya, Malawi and Mozambique. Development Practice in Education.

[9] Guo, Bo, Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting students performance in educational data mining. Educational Technology (ISET), 2015 International Symposium, (pp. 125-128).

[10] Xu, J., Han, Y., Marcu, D., &Schaar, M. a. (2017). Progressive Prediction of Student Performance in College Programs. AAAI, 1604-1610.

[11] Asif, R., Merceron, A., &Pathan, M. K. (December 2014). Predicting Student Academic Performance at Degree Level: A Case Study. International Journal of Intelligent Systems Technologies and Applications, 01, 49-61.

[12] Paulo, C., & Silva, A. (2008). Using data mining to predict secondary school student performance. 5-12.

[13] Oladokun, V. O., Adebanjo, A. T., & Charles-Owaba, O. E. (2008). Predicting students academic performance using artificial neural network: A case study of an engineering course. The Pacific Journal of Science and Technology, 8(1), 72-79.

[14] Osmanbegović, E., &Suljić , M. (2012). Data mining approach for predicting student performance. Economic Review: Journal of Economics and Business, 10(1), 3-12.

[15] Khasanah, A. U., &Harwati. (2017). A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. IOP Conference Series: Materials Science and Engineering, 215(1).

[16] Khan, B., Hayat, M. S., &Daud, M. (2015). Final Grade Prediction of Secondary School Student using Decision Tree. International Journal of Computer Applications, 115(21), 32-36.

[17] Sharma, G., & Santosh, V. K. (2017). Analysis and Prediction of Student's Academic Performance in University Courses. International Journal of Computer Applications, 160(4).

[18] Baradwaj, B. K., & Pal, S. (2011). Mining Educational Data to Analyze Students‟ Performance. International Journal of Advanced Computer Science and Applications, 2(6), 63-69.

[19] Dey, A. (2016). Machine Learning Algorithms: A Review. International Journal of Computer Science and Information Technologies, 7(3), 1174-1179.

[20] Karthikeyan, K., &Kavipriya, P. (May 2017). On Improving Student Performance Prediction in Education Systems using Enhanced Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering, 7(5), 935-941.

[21] W Xing, W., Rui, G., Eva, P., & Sean, G. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. Computers in Human Behavior, 47, 168-181.

[22] Livieris, L. E., Drakopoulou, K., & Panagiotis, P. (2012). Predicting students' performance using artificial neural networks. 8th PanHellenic Conference with International Participation Information and Communication Technologies in Education, (pp. 321-328).

[23] Mitchell, T. M. (1997). Machine learning (Vol. 45(37)). Burr Ridge: IL:McGraw Hill, (pp. 870-877).

[24] Saa, A. A., 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, 1-32