LAUTECH Journal of Computing and Informatics (LAUJCI) – ISSN: 2714-4194
Volume 2 No.2, Sept 2021 – www.laujci.lautech.edu.ng

# Prediction of Students' Success in Secondary Education Using Selected Machine Learning Techniques

Ibrahim A. Adeyanju
Department of Computer Engineering, Federal University Oye- Ekiti, Nigeria
**ibrahim.adeyanju@fuoye.edu.ng**

Odunayo E. Oduntan
Department of Computer Science, The Federal Polytechnic, Ilaro, Nigeria
**odunayo.oduntan@federalpolyilaro.edu.ng**

Taofeeq A. Badmus
Department of Computer Engineering, Federal University Oye- Ekiti, Nigeria
**taofeeq.badmus@fuoye.edu.ng**

Mobolaji Adebayo
Department of Computer Engineering, Federal University Oye- Ekiti, Nigeria
**mobolaji.adebayo.1649@fuoye.edu.ng**

**ABSTRACT**

The success of secondary school students, in terms of passing final certificate exams, is an important factor that determines their progress to higher education and towards becoming skilled labour or entrepreneurs in future. The complexity of this problem is multi-faceted as student experiences and expectations are not same for all schools. However, failure to address this problem can result in a high percentage of unskilled workforce which has adverse effects on the development of any country. This paper evaluates the performance of three machine learning algorithms to predict success or failure of a final year secondary school student. The algorithms are Logistic Regression, Ada Boost with Decision stump and k-Nearest Neighbour (kNN). Experimental dataset was obtained from Kaggle, with 395 instances each originally having 31 attributes. A ten-fold cross validation evaluation methodology was employed in our experiments after feature selection with a best first attribute selection filter which reduced the attributes to five (5). We simulated the algorithms using WEKA 3.9.5. Ada Boost with Decision stump performed best among the three selected algorithms with an average accuracy of 71.65%, followed by Logistic Regression and kNN with 70.63% and 70.13%, respectively. We intend to experiment with data obtained locally from secondary schools within Nigeria to further validate the performance of the selected and other machine learning algorithms.

**Keywords:** Machine Learning, Logistic Regression, k-Nearest Neighbour, Ada Boost, Decision Stump, Secondary Education

## 1. INTRODUCTION

Student dropout is a major concern in the education and policy-making communities (Demetriou & Schmitz-Sciborski, 2011). About 40% of students seeking Bachelor degrees do not complete their degree within six years (Bean, 1990). Student attrition in educational institutions has a negative impact on all parties involved; students, institutions, and general public (Bowen, Chingos, & McPherson, 2009). Not considering the educational gain of a student before choosing to drop out, attrition results in direct financial losses and may create feelings of inadequacy that can lead to stigmatization within the society. Attrition is a cause of concern in many educational institutions all over the world. The factors that lead to attrition varies taking into the consideration geographical factors, ethnicity, education system of a country among others (Veenstra, 2009). Many studies have been carried out to find out the reasons which lead to attrition and also to overcoming the attrition problem through various approaches including intervention strategies (Stefanie, 2012; Ameri, Fard, & Chinnam, 2016; Aulck, et al., 2017; Dekker, Pchenenizkiy, & Vleeshouwers, 2009; Delen, 2011; Elbadrawy, et al., 2016; Xu, Moon, & Van der Schaar, 2017).

A pre-requisite to higher education is success in the final certificate exams in secondary (or high) school. This paper evaluates the performance of three selected machine learning algorithms, Ada Boost with Decision stump, K-Nearest Neighbour (KNN) and Logistic Regression for predicting the success of students in their final secondary school certificate exams. Section 2 discussed related work while selected machine learning algorithms used in this research are explained in Section 3. Experiment setup, results and conclusion are given in Sections 4, 5 and 6, respectively.

## 2. RELATED WORK

An investigation to understand university student retention from a data mining perspective was conducted (Zhang, et al., 2011). They built their analysis on a dataset, combining several resources from university institutions such as the library usage, online resources and the student record system. They discovered that Naive Bayes approach outperforms support vector machines and decision trees in predicting student success. The use of survival analysis modelling to study student retention was developed (Ameri, Fard, & Chinnam, 2016; Ameri, 2015; Wang, Li, & Reddy, 2017; Tarmizi, Mutalib, AbdulHamid, & AbdulRahman, 2019). Their work was used to identify at-risk students using Cox proportional hazards model (Cox) and applied time-dependent Cox (TD-Cox). This approach captures time-varying factors and leverages that information to provide more accurate prediction of student dropout, using the dataset of students enrolled at Wayne State University (WSU) starting from 2002 until 2009. Certainly, subjects in survival analysis are usually followed over a specified period of time and the focus is on the time at which the event of interest occurs (Liang, Li, & Zheng, 2016). Thus, the benefit of using survival analysis over other methods is the ability to add the time component into the model and also effectively handle censored data.

A new data transformation model, which is built upon the summarized data matrix of link-based cluster ensembles (LCE) was previously proposed (Iam-On & Boongoen, 2017). The aim of the conducted study was to establish the clustering approach as a practical guideline for exploring student categories and characteristics. This was accomplished using educational dataset obtained from the operational database system at Mae Fah Luang University, Chiang Rai, Thailand. Like several existing dimensionality reduction techniques such as Principal Component Analysis (Oduntan & Adeyanju, 2017) and Kernel Principal Component Analysis, this method aims to achieve high classification accuracy by transforming the original data to a new form. However, the limitation of the new technique was the demanding time complexity, such that it may not scale up well to a very large dataset. Whilst worst-Case Traversal Time (WCT-T) is not quite for a highly time-critical application, it can be an attractive candidate for those quality-led works, such as the identification of those students at risk of under achievement.

Machine learning techniques (Adeyanju, Fenwa, & Omidiora, 2013; Mitchell, 1997) have been applied in various Massive Open On-line Course (MOOC) platforms such as Coursera and edX which are among popular used platforms for generating datasets to be used in student dropout prediction (Chen, Zhao, Boyer, Veeramachaneni, & Qu, 2017; Wang, Yu, & Miao, 2017; Yang, Piergallini, Howley, & Rose, 2014; Fei & Yeung, 2015; McAnulla, Ball, & Knapp, 2020; California State University, Chico State, 2021).

## 3. SELECTED MACHINE LEARNING ALGORITHMS

In this section, we discuss the details of the three selected machine learning algorithms; AdaBoost, Logistic regression and k-Nearest Neighbour.

### 3.1 AdaBoost

AdaBoost (Freund & Schapire, 1996) is one of the ensemble boosting classifiers. It combines multiple classifiers to increase the accuracy of classifiers. AdaBoost classifier builds a strong classifier by combining multiple poorly performing classifiers so that one can get high accuracy strong classifier. The basic concept behind AdaBoost is to set the weights of classifiers and training the data sample in each iteration, such that it ensures the accurate predictions of unusual observations. Any machine learning algorithm can be used as base classifier if it accepts weights on the training set. Adaptive Boosting algorithm can be used on extracted features as weights will be assigned to each instance, with higher weights to incorrectly classified instances. Figure 1 illustrates how the AdaBoost algorithm works.

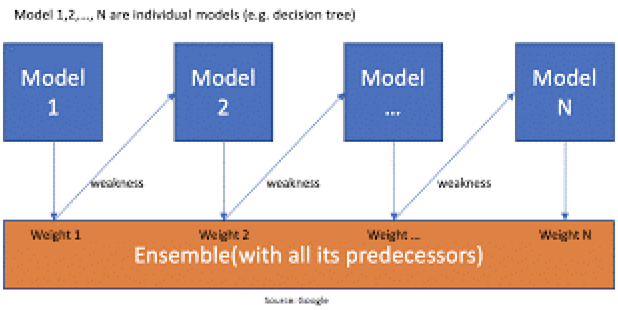


**Fig 1: AdaBoost Algorithm**

### 3.2 Logistic Regression

Logistic regression (le Cessie & van Houwelingen, 1992) is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. Mathematically, a logistic regression model predicts P(Y=1) as a function of X. it is a predictive analysis algorithm and based on the concept of probability (Olanloye, Oduntan, & Olasunkanmi, 2018). A number of parameters weighing the feature variables are calibrated in order to find an optimal fit to the dependent variable given a fixed functional form. In logistic regression this functional form is an S-curve in between the values of 0 and 1. How the S is shaped is determined by estimating fitting parameter values in this case using an iteratively re-weighted least squares method. Regression uses a more complex cost function, this cost function can be defined as the **'Sigmoid function'** or also known as the 'logistic function'; shown in Figure 2. The function maps any real value into another value between 0 and 1. In machine learning, sigmoid was used to map predictions to probabilities.

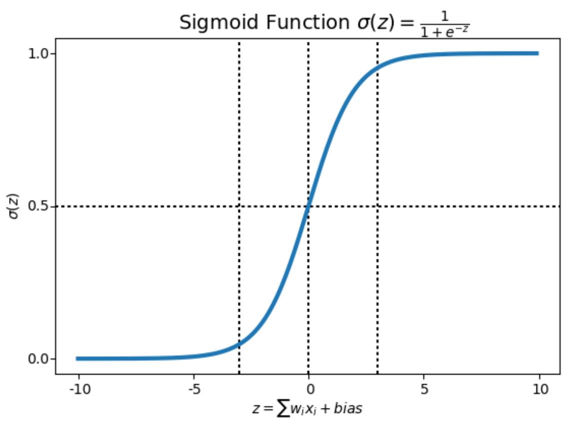


**Fig 2: The Sigmoid Function**

### 3.3 K-Nearest Neighbour

K-Nearest Neighbour (kNN) (Aha & Kibler, 1991) is widely used in the area of pattern recognition. Nearest-neighbour classifiers are based on learning by analogy, that is, by comparing a given test tuple with training tuples that are similar to it. The training tuples are described by n attributes. Each tuple represents a point in an n-dimensional space. In this way, all of the training tuples are stored in an n-dimensional pattern space. When given an unknown tuple, a k-nearest-neighbour classifier searches the pattern space for the k training tuples that are closest to the unknown tuple.

For K-Nearest Neighbour, the training tuples are described by n attributes. Each tuple represents a point in an n-dimensional space. In this way, all of the training tuples are stored in an n-dimensional pattern space. When given an unknown tuple, a k-nearest-neighbour classifier searches the pattern space for the k training tuples that are closest to the unknown tuple. These k training tuples are the k "nearest neighbours" of the unknown tuple. "Closeness" is defined in terms of a distance metric, such as Euclidean distance (Natarajan et al., 2012). For k-nearest-neighbour classification, the unknown tuple is assigned the most common class among its k nearest neighbours. When k = 2 as used in this work, the unknown tuple is assigned the class of the training tuple that is second closest to it in pattern space. Nearest neighbour classifiers can also be used for prediction, that is, to return a real-valued prediction for a given unknown tuple. In this case, the classifier returns the average value of the real-valued labels associated with the k nearest neighbours of the unknown tuple. Figure 3 illustrates the KNN algorithm. In its simplest form, the closest neighbour of a new test example is selected from the neighbourhood and its class is predicted as that of the test. In the figure, the test example will be predicted as yes based on the closest neighbour.

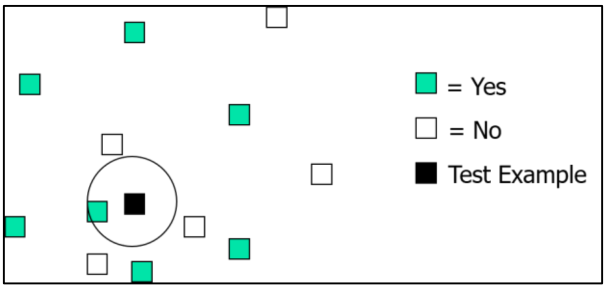


**Fig 3: kNN Algorithm illustrated**

## 4. EXPERIMENTAL DESIGN AND SIMULATION

This study elaborates the methods of developing the predictive model using selected classification algorithms which involves the following process; data collection and cleaning, feature selection and classification. Three (3) machine learning algorithms were used for experiments: AdaBoost with Decision stump, Logistic Regression and K-Nearest Neighbour with k=3. The performance of these algorithms on student attrition was evaluated using average accuracy, true positive rate (recall), precision and f-measure metrics. A ten-fold cross validation evaluation methodology was used in our experiments. Figure 4 shows the flow-diagram of the training process of the experiment.

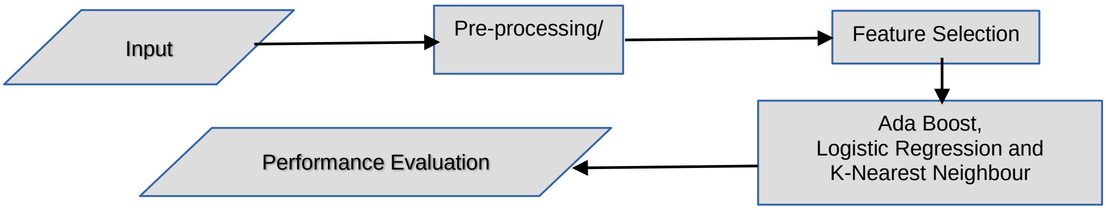


**Fig 4: Flow Diagram of the Training Process**

### 4.1 Experimental Dataset

The dataset comprises of 395 instances with 31 attributes having varied data types including numeric, nominal and Boolean. The freely available dataset was downloaded from Kaggle (www.kaggle.com/dipam7/student-grade-prediction/version/1). The dataset was originally from the UCL machine learning repository and contains student achievement in secondary education of two Portuguese schools collected using school reports and questionnaires (Cortez & Silva, 2008). Table 1 shows the dataset attributes and a short description of each attribute. In this study, the class attribute is "passed", which indicates if a student's success after secondary education.

**Table 1. Attributes of the Students' success dataset**

| | Attributes | Description | | Attributes | Description |
|---|---|---|---|---|---|
| 1 | school | student's school (binary: "GP" or "MS") | 16 | schoolsup | extra educational support (binary: yes or no) |
| 2 | sex | student's sex (binary: "F" - female or "M" - male) | 17 | famsup | family educational support (binary: yes or no) |
| 3 | age | student's age (numeric: from 15 to 22) | 18 | paid | extra paid classes within the course subject (Math or Portuguese) (binary: yes or no) |
| 4 | address | student's home address type (binary: "U" - urban or "R" - rural) | 19 | activities | extra-curricular activities (binary: yes or no) |
| 5 | famsize | family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3) | 20 | nursery | attended nursery school (binary: yes or no) |
| 6 | Pstatus | parent's cohabitation status (binary: "T" - living together or "A" - apart) | 21 | higher | wants to take higher education (binary: yes or no) |
| 7 | Medu | mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education) | 22 | internet | Internet access at home (binary: yes or no) |
| 8 | Fedu | father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education) | 23 | romantic | with a romantic relationship (binary: yes or no) |
| 9 | Mjob | mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other") | 24 | famrel | quality of family relationships (numeric: from 1 - very bad to 5 - excellent) |
| 10 | Fjob | father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other") | 25 | freetime | free time after school (numeric: from 1 - very low to 5 - very high) |
| 11 | reason | reason to choose this school (nominal: close to "home", | 26 | goout | going out with friends (numeric: from 1 - very low |

| | | | | | |
|---|---|---|---|---|---|
| | | school "reputation", "course" preference or "other") | | | to 5 - very high) |
| 12 | guardian | student's guardian (nominal: "mother", "father" or "other") | 27 | dalc | workday alcohol consumption (numeric: from 1 - very low to 5 - very high) |
| 13 | traveltime | home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour) | 28 | walc | weekend alcohol consumption (numeric: from 1 - very low to 5 - very high) |
| 14 | studytime | weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours) | 29 | health | current health status (numeric: from 1 - very bad to 5 - very good) |
| 15 | failures | number of past class failures (numeric: n if 1<=n<3, else 4) | 30 | absences | number of school absences (numeric: from 0 to 93) |
| | | | 31 | passed | did the student pass the final exam (binary) |

### 4.2 Pre-processing and Feature Selection

From the dataset, there were several non-numeric attributes that needed to be cleaned. Eight of such attributes had values- yes/ no and were reasonably converted into 1/0 (binary) values. Other attributes, like mother's job (Mjob) and father's job (Fjob), had more than two values and were used as nominal variables.  Each nominal value for a particular attribute was converted to a distinct dummy variable for ease of programming.

Feature selection was carried out using a supervised attribute filter. The filter uses subset evaluator which evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them (Hall, 1998). Subsets of features that are highly correlated with the class while having low intercorrelation are preferred. The filter also uses the best first search algorithm which searches the space of attribute subsets by greedy hill-climbing augmented with a backtracking facility. Best first search may start with the empty set of attributes and search forward, or start with the full set of attributes and search backward, or start at any point and search in both directions, by considering all possible single attribute additions and deletions at a given point (Mitchell, 1997; Witten, Frank, & Hall, 2011; The University of Waikato, 2020).

The feature selection process resulted in five most important attributes out of the original thirty-one attributes in the dataset. In other words, four attributes (reason, failures, schoolsup and gout; see Table 1 for details) and the class label (passed) as fifth attribute were left for our simulation experiments after feature selection.

### 4.3 Simulation

The machine learning algorithms and experiments were simulated using WEKA data mining tool (Witten, Frank, & Hall, 2011; The University of Waikato, 2020). The parameters for the three algorithms were left at the default values as set in WEKA, except for those stated. Figure 4 shows a screenshot of the WEKA explorer where the experiments were performed.
AdaBoost (Adaptive Boosting) algorithm used decision stump as the base learner. Default values were used for Logistic Regression while the number of nearest neighbours was set to 3 for the k nearest neighbour algorithm.

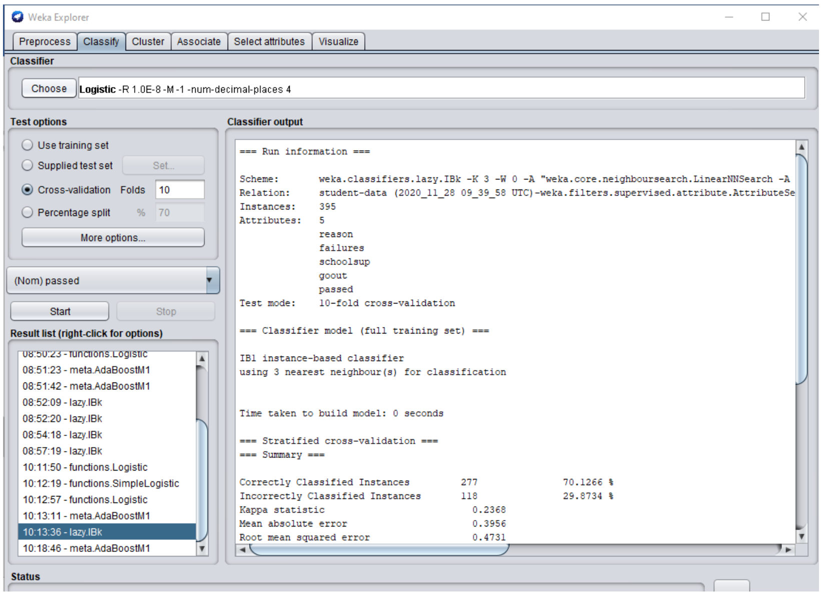


**Fig 4: Screenshot of the simulation experiments in WEKA**

## 5. RESULTS AND DISCUSSION

The results of the prediction using the following classification algorithms are shown in Table 2. Among the three classifiers used, Adaboost with Decision stump performed best with an accuracy of 71.65% and F-measure of 0.692. Adaboost also outperformed the other two classifiers with the precision and recall evaluation metrics. The Logistic Regression and kNN classifiers gave an accuracy of 70.63% and 70.13% respectively. Nevertheless, all three algorithms performed well on the dataset and can be used to predict secondary students' success in their final exams.

**Table 2: Evaluation results for the three selected machine learning algorithms**

| | **Adaboost with Decision Stump** | **Logistic Regression** | **K- Nearest Neighbour (k=3)** |
|---|---|---|---|
| **Accuracy (%)** | *71.6456* | 70.6329 | 70.1266 |
| **True Positive Rate (Recall)** | 0.716 | 0.706 | 0.701 |
| **Precision** | 0.700 | 0.688 | 0.680 |
| **Fmeasure** | 0.692 | 0.669 | 0.673 |

## 6. CONCLUSION

This work has employed machine learning classification techniques for predicting the success of secondary school students in their final exams. Adaboost with decision stump was discovered to perform best out of the three algorithms used in empirical experiments. We intend to experiment with data obtained locally from secondary schools within Nigeria to further validate the performance of the selected three and other machine learning algorithms. Future work will also investigate the use of the same algorithms for attrition prediction in tertiary institutions.

## 7. REFERENCES

[1] Adeyanju, I. A., Fenwa, D. O., & Omidiora, E. O. (2013). Comparison of Machine Learning Classifiers for Recognition of Online and Offline Handwritten Digits. *Computer Engineering and Intelligent Systems*.
[2] Aha, D., & Kibler, D. (1991). Instance-based learning algorithms. *Machine Learning, 6*, 37-66.
[3] Ameri, S. (2015). *Survival Analysis Approach For Early Prediction Of Student Dropout.* Wayne State University. .
[4] Ameri, S., Fard, M., & Chinnam, R. (2016). Survival Analysis Based Framework for Early Prediction of Student Dropouts. *25th ACM International on Conference on Information and Knowledge Management, CIKM,*, (pp. 903-912). New york.
[5] Aulck, L., Aras, R., Li, L., Heureux, L., Lu, P., & West, J. (2017). STEM-ming the Tide: Predicting STEM attrition using student transcript data. *ACM Knowledge Discovery and Data Mining Conference.* Nova Scotia.
[6] Bean, J. (1990). *Using Retention Management in Enrollment Management.* San Francisco.
[7] Bowen, W., Chingos, M., & McPherson, M. (2009). *Crossing the finish Line: Completing College at america's Public Universities.* New Haven: Princeton University Press.
[8] California State University, Chico State. (2021, January). January 2021- Continuation, Retention and the Timing of Student Attrition, Pt.1. *By the Numbers*. Retrieved from www.csuchico.edu/ir/by-the-numbers/continuation-retention-attrition.shtml
[9] Chen, Y. C., Zhao, M., Boyer, S., Veeramachaneni, K., & Qu, H. (2017). Visualizing learning patterns in Massive Open Online Courses for dropout reasoning and prediction. *IEEE Conference on Visual Analytics Science and Technology* (pp. 111-120). IEEE.
[10] Cortez, P., & Silva, A. (2008). Using Data Mining to predict secondary school student performance. In A. Brito, & J. Teixeira (Ed.), *Proceedings of the 5th Future Business Technology Conference (FUBUTEC 2008)* (pp. 5-12). Porto, Portugal: EUROSIS.
[11] Dekker, G., Pchenenizkiy, M., & Vleeshouwers, J. (2009). Predicting Students Drop out: A Case Study. *Educational Data Mining*.
[12] Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention:Research, Theory & Practice,*, 17–35.
[13] Demetriou, C., & Schmitz-Sciborski, A. (2011). Integration, motivation, strengths and optimism: Retention theories past, present and future. *National Symposium on Student Retention*, (pp. 300-312). Charleston.
[14] Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. (2016). Predicting Student Performance Using Personalized Analytics. 61-69.
[15] Fei, M., & Yeung, D. (2015). Temporal Models for Predicting Student Dropout in Massive Open Online Courses. *International Conference on Data Mining Workshop* (pp. 256-263). IEEE.
[16] Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm.

*Thirteenth International Conference on Machine Learning*, (pp. 148-156). San Francisco.
[17] Hall, M. A. (1998). *Correlation-based Feature Subset Selection for Machine Learning.* Hamilton, New Zealand.
[18] Iam-On, N., & Boongoen, T. (2017). Generating descriptive model for student dropout: A review of clustering approach., . *Human-centric Computing and Information Sciences*, 7.
[19] le Cessie, S., & van Houwelingen, J. (1992). Ridge Estimators in Logistic Regression. *Applied Statistics, 41*(1), 191-201.
[20] Liang, J., Li, C., & Zheng, L. (2016). Machine learning application in MOOCs: Dropout prediction. *International Conference on Computer Science and Education* (pp. 52-57). ICCSE.
[21] McAnulla, S., Ball, S., & Knapp, K. (2020). Understanding student radiographer attrition: RIsk factors and strategies. *Radiography, 26*(3), 198-204. doi:10.1016/j.radi.2018.12.001
[22] Mitchell, T. (1997). *Machine Learning.* McGraw Hill.
[23] Oduntan, O., & Adeyanju, I. (2017). A Comparative Analysis of Modified Principal Component Analysis and Generalized Latent Semantic Analysis Approach to Automated Marking of Theory-Based Exams. *IOSR Journal of Mobile Computing and Application, 4*(2), 31-41.
[24] Olanloye, O., Oduntan, O. E., & Olasunkanmi, O. (2018). Using Predictive Machine Learning Regression Model Predict the Population of Nigeria. *Annals of Computer Science Series, 16*(2), 137 – 142.
[25] Stefanie, S. T. (2012). The Best laid Plans: Examining the Conditions Under which a planning Intervention Improves learning and reduces attrition. *Journal Of Applied Psychology*, 8-10.
[26] Tarmizi, S., Mutalib, S., AbdulHamid, N., & AbdulRahman, S. (2019). A Review on Student Attrition in Higher Education Using Big Data Analytics and Data Mining Techniques. *International Journal of Modern Education and Computer Science, 11*(8), 1-14. doi:10.5815/ijmecs.2019.08.01
[27] The University of Waikato. (2020). Waikato Environment for Knowledge Analysis (WEKA). Hamilton, New Zealand.
[28] Veenstra. (2009). A strategy for improving freshman college retention. Journal for Quality &. *Journal for Quality & participation, 31*(4), 19-23.
[29] Wang, P., Li, Y., & Reddy, C. (2017). Machine Learning for Survival Analysis: A Survey. *ACM Computing Survey, 1*(1).
[30] Wang, W., Yu, H., & Miao, C. (2017). Deep Model for Dropout Prediction in MOOCs. *International Conference on Crowd Science and Engineering* , (pp. 26-32).
[31] Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques.* Burlington, MA: Morgan Kaufmann.
[32] Xu, J., Moon, K., & Van der Schaar, M. (2017). A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs. . *IEEE Journal of Selected Topics in Signal Processing*, 742–753.
[33] Yang, D., Piergallini, M., Howley, I., & Rose, C. (2014). Forum Thread Recommendation for Massive Open Online Courses. *International Conference on Educational Data Mining (EDM)*, (pp. 257-260).
[34] Zhang, J., Zhan, Z.-h., Lin, Y., Chen, N., Gong, Y.-j., Zhong, J.-h., . . . Shi, Y.-h. (2011). Evolutionary Computation Meets Machine Learning: A Survey. *Computational Intelligence Magazine*, pp. 68-75.

**Author's Brief Profile**

Ibrahim A. Adeyanju is currently an Associate Professor (Reader) at the Department of Computer Engineering, Federal University Oye-Ekiti, Nigeria where he teaches and does research. His specialization is in Intelligent Systems with research interests covering different research areas of Computer Science and Engineering including Intelligent hardware and software systems, Pattern Recognition, Artificial Intelligence, Machine Learning, Embedded Systems, Information Retrieval and Networks. He has several peer-reviewed articles in reputable national and international journals as well as referred conference proceedings.

Odunayo Esther Oduntan, is a researcher and an academia in Computer Science. She has Ph.D, M.Sc, B.Sc(Computer Science) and a Graduate Student of Masters Degree in Machine Intelligence at the African Institute for Mathematical Sciences, Kigali, Rwanda. She is a Principal Lecturer in the Department of Computer Science, The Federal Polytechnic, Ilaro Ogun State. Her research interest is in machine intelligence; comprising machine learning, natural language processing and computer vision. She is a member of Black in AI, Data Science Nigeria (DSN), Women in Data Science, Google Scholar Forum, Computer Professionals of Nigeria (CPN) and the Nigeria Computer Society (NCS). She is happily married and blessed with children.

Taofeeq A. Badmus is currently a Lecturer at the Department of Computer Engineering, Federal University Oye-Ekiti, Nigeria where he teaches and does research. His research covers different areas of Computer Science and Engineering. He is currently studying for his PhD programme at the University of Nottingham, UK. He has several peer-reviewed articles in reputable national and international journals as well as referred conference proceedings.

Mobolaji Adebayo is a recent graduate of the Department of Computer Engineering, Federal University Oye-Ekiti, Nigeria. He did his undergraduate final year project under the supervision of the first author (Dr Adeyanju) which was extended for this journal article.