International Journal of Advanced Computer Research, Vol 11(56)
ISSN (Print): 2249-7277 ISSN (Online): 2277-7970
http://dx.doi.org/10.19101/IJACR.2021.1152024

Research Article

Prediction of mathematics performance using educational data mining techniques

Paul K Mushi[1*] and Daniel Ngondya[2]
Computing Science Studies Department, Faculty of Science and Technology, Mzumbe University, Tanzania[1]
Telecommunication Department, College of Informatics Virtual and Education, University of Dodoma, Tanzania[2]

Received: 25-July-2021; Revised: 23-September-2021; Accepted: 26-September-2021

Abstract
Higher Learning Institutions (HLIs) nowadays store a large amount of students' data. However, these data are not widely used to solve the students' academic problems at the institutions such as poor performance in some courses. Educational Data Mining (EDM) is a technology that can be applied to predict the performance of students from the dataset at HLIs. This study intended to solve the problem of poor performance in mathematics by management degree students at HLIs using EDM techniques and Mzumbe University (MU) in Morogoro, Tanzania as a case study. A quantitative research approach was applied based on the design science steps. Secondary data were collected to create the dataset through a review of documents from the examination, admission, accommodation, and accounts offices, as well as the Department of Mathematics and Statistics from the Main and Mbeya campuses of MU. Different Machine Learning (ML) algorithms were applied on the training set (60%) such as K-Nearest Neighbor (K-NN), Random Forest (RF), Decision Tree (DT), Support Vector Classification (SVC), and Multilayer Perceptron (MLP). Machine Learning algorithms were validated using a 10-fold cross-validation and validation dataset (20%) and the best algorithms were established to be RF, DT, and K-NN. Further evaluation of these three ML algorithms using 20% of the dataset demonstrated that the RF algorithm was the best for model development for the prediction of mathematics performance with an accuracy of 99% and F1-scores of 99% and 100% for the fail and pass classes respectively. Moreover, DT could generate rules that can be applied to recommend the minimum grade of D in ordinary level mathematics for admission into the University for Management Degrees to reduce the failure rates at HLIs.

Keywords
Educational data mining, Machine learning, Random forest, Decision tree, Mathematics, Performance prediction.

1.Introduction
Higher Learning Institutions (HLIs) are critical for educating people to increase knowledge and innovations in the world. Cognizant of this, both private sectors and governments invest heavily in universities worldwide to sponsor students for the improvement of economic growth, societal development, and self-employment [1]. Parents often have high expectations for good performance and the successful completion of studies by their children. Hence, different stakeholders in HLIs such as parents or guardians, sponsors, government, students, lecturers, and decision-makers have obligations to realize good academic performance by scholars.

Although students in HLIs devote their time to acquire knowledge and skills, they may pass or fail in various programmes in both developed and developing countries. This can be attributed to demographic factors (age, gender, and marital status), economic factors of parents, living locations, and sponsorship availability [2].

Numerous studies have been conducted on mass failure in computer science and engineering disciplines in HLIs [3]. One study concerned the prediction of mathematics performance at the secondary school level [4]. Research was done using Data Mining (DM) classification algorithms to predict the performance of students in Introductory Programming that is mandatory for first-years in the Computer Science Discipline in affiliated colleges of the University of Madras to predict if the students were likely to pass or fail [5]. A study conducted at the

*Author for correspondence

Federal University of Technology Akure in Nigeria considered students pursuing Quantity Surveying, Estate Management, Architecture and Industrial Design in the Construction Discipline [6]. The researcher assessed factors affecting the performance of undergraduate students in the courses to improve their performance. In particular, the study was concerned with the assessment of academic (continuous assessment, Grade Point Average, and entry standards) and non-academic (economic issues and family matters) factors.

In Tanzania, a study at Mkwawa University College of Education concerning the academic performance of students in HLIs [7]. The study focused on the quality of academic advice given to students in the early stages when they join Universities. The study showed that during the first semester of the first year of the 2015/2016 academic year, almost half of the science-based education students (46% of 338) failed in one course. The researcher concluded that there was need to revisit and improve student advisory services to achieve better retention and graduation rates.

Considering Mzumbe University (MU) as a case study, Management programmes in HLIs with mass failures. These Management programmes differ in their admission/entry criteria and may include students with backgrounds in science or art subjects. Besides, the entry criteria for mathematics and other courses undertaken by students after admission also differ. The present study intended to focus on students in four Management programmes who were admitted without consideration of their mathematics background. The programmes were Bachelor of Local Government Management (BLGM), Bachelor of Health System Management (BHSM), Bachelor of Human Resource Management (BHRM), and Bachelor of Public Administration-Records and Archives Management (BPA-RAM). The programmes were offered at the School of Public Administration and Management (SOPAM) and MU Mbeya Campus College (MCC). The programmes had the same admission/entry criteria and it was mandatory to study Mathematics during the first semester of the first year [8]. Therefore, there would be a trend of students' mass failure in mathematics for Management programmes and there was a need to predict the performance at the earlier stages for first-year students to mitigate the failure rates.

Various Educational Data Mining (EDM) techniques such as Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbor (K-NN), Artificial Neural Network (ANN), Random Forest (RF), and Decision Tree (DT) have been used for the prediction of the students' performance. The present study was based on EDM techniques such as ANN, K-NN, and SVM, DT, and RF to come up with the best ML algorithm for model development of predicting mathematics performance of management students who are more likely to fail in mathematics at HLIs.

The present study was motivated by three reasons. First, the need to explore the research in ML by applying EDM techniques. The second reason was to solve the problem of mass failure in mathematics for a particular discipline in HLIs that has not been widely researched. The third reason was the consideration of mathematics performance at HLIs apart from the high, ordinary, or primary level schools. Worldwide HLIs are facing the problem of high failure rates among students in various disciplines such as engineering, computer science, management, and business. Consequently, students may have to undergo supplementary examinations, carryovers, or discontinuation during the semester or at the end of each academic year. Owing to this, many students may not complete their studies and gain meaningful employment. In addition, these failures may also lead to loss of funds by parents and government through Higher Education Students' Loans Board (HESLB), and unhealthier and unhappier lives in the surrounding societies. At Mzumbe University, mathematics is one of the courses with high failure rates among students pursuing Management degree programmes. The main objective of this study was to develop a mathematics performance predictive model for Management degree programmes using various EDM techniques. The specific objectives were; i) to identify the requirements for training ML algorithms in mathematics performance prediction, ii) to identify the best-trained ML algorithms in mathematics performance prediction, and iii) to evaluate the best selected trained ML algorithms in mathematics performance prediction.

2.Literature review

Different studies have been reviewed in this study regarding performance prediction in mathematics using EDM techniques. The research papers reviewed were those that helped to meet the objective of the study as described in the following paragraphs. The authors have reviewed the article on a systematic review on DM for mathematics and science education [9]. This article reviewed 64 articles based on the research topics and the DM techniques used. The review demonstrated that DM in mathematics and science education has commonly been applied in performance prediction. Moreover, student modeling was observed to be not only the popular research topic and classification but also the widely applied EDM technique in the study.

Related works have also been based on multi-models of EDM for predicting student performance in mathematics, for instance, in a case study on high schools in Cambodia [10]. The study applied nine algorithms to obtain the best models and the four leading algorithms were 1-NN (94.95%), and the three-ensemble tree-based algorithms; Boosted C5.0 (95.67%), Bagged CART (95.60%), and RF (96.69%). The study established that the RF algorithm was the best compared to others in classifying the performance levels of students in mathematics into four levels as poor, average, good, and excellent. The RF algorithm was still the best when Predictive Mean Square Error (PMSE) was applied and the result was 0.013 whereby the less the value, the better the goodness of fit of the model.

A survey on the prediction system for student academic performance using EDM is also available [11]. The study showed the research gaps in EDM to identify the weak students who may fail at universities, high schools, and colleges. It was based on educational data from various systems such as admission, registration, course management, and syllabus management systems. The study showed the efficiency of the classification techniques that use multi-class classification which gives the best accuracy while requiring less execution time in performance prediction.

Likewise, another study applied SVM and DT algorithms to predict the pass rates of students in mathematics and Portuguese [12]. The study involved experiments to find out the accuracies of the trained ML algorithms when all features were involved. The results of mathematics performance prediction were 91.5, 92.6, 72.4, and 88.3% for DT, SVM, RF, and ANN respectively. Nevertheless, when other predictor variables were not involved in the prediction of mathematics performance, the results were 83.9, 87.3, 52.7, and 81.3% for DT, SVM, RF, and ANN algorithms respectively. Cognizant of these variations, other predictor variables that depend on each other need to be considered in training the ML algorithms to have higher accuracies. Hence, in the present study, all the predictor variables for the prediction of mathematics performance were involved apart from finding the significance of the predictor variables to the outcome variable and found few of them had little influence.

Another related study compared the 5-level grading that had an accuracy of 71.14% for the RF algorithm and binary level grading in mathematics that had an accuracy of 91.39% [13]. The study showed that the DT algorithm had an accuracy of 73.42% when the 5-level grading was applied which increased to 89.11% when the binary level grading was used. Therefore, it was concluded that the accuracies of the trained algorithms increased as the level of the classes to the target variable decreased. Owing to this, the present study opted to base on binary classification by applying the ML algorithms in the prediction of mathematics performance.

Another researcher compared several selected studies in DM that applied classification algorithms such as DT, ANN, SVM, RF, K-NN, and NB [14]. The study showed that the mentioned algorithms were capable of processing a large amount of dataset and predicting categorical class labels using the training set to classify data, and consequently, could classify newly available data. Therefore, in the present study that intended to predict mathematics performance, classification algorithms were applied to categorize the remarks into the pass and fail class labels using the training data set after which the testing data set was used to predict the students' performance.

Another related study involved the RF ML algorithm to predict students' performance prediction and correctly classified 701 instances while only 297 instances were incorrectly classified and the accuracy level was 72.04 [15]. Therefore, in the present study, the RF algorithm is one of the trained ML algorithms that were used to predict students' performance in mathematics.

Other researchers conducted a study to predict the performance of engineering students and trained DT algorithms that had an accuracy of 67.8% [16]. From the Confusion Matrix, it was clear that the true positive rate of the model for the failure class was 0.786 accurate for Iterative Dichotomiser 3 (ID3) and C4.5 DT which means that the generated model successfully identified the students who were likely to fail. The study recommended that ML algorithms such as C4.5 DT can learn effective predictive models from students' data accumulated from the previous years. The factors considered by the study were student's programmes, admission types, gender, grades in high school and senior secondary, living locations, medium of teaching, family sizes, family annual income and family status, father's and mother's occupations, and qualifications. The present study, therefore, classified the dependent variable remarks as either pass or fail, considering the predictor variables from five previous academic years (2014/2015 to 2018/2019). The aim was to have more data for training and testing the ML algorithms and one of the ML algorithms trained in the study was DT which has previously been shown to have the highest accuracy to be considered for training.

Comparative analysis of selected studies for binary outcome prediction such as the pass or fail has been done [3]. Linear Regression and Support Vector Machine algorithms may be applied to predict numerical score outcomes such as predicting students' performance in the final examination (FE) [17]. However, the actual prediction of scores has historically been less accurate compared to those that predict whether the student has passed or failed a particular course. This is why in this study; classification of the target variable was involved as either pass or fail for the prediction of mathematics performance by students pursuing Management degrees.

One study used the DM classification algorithms to predict students' performance in Introductory Programming in the computer science discipline in affiliated colleges of the University of Madras [5]. First-year students had to take that course but the performance was not good, hence, the research was conducted to predict earlier the students who were likely to pass or fail. The classification algorithms applied were MLP, DT (J48), DT (REPTree), Sequential Minimal Optimization (SMO), and NB, and their predictive accuracies were 93.23, 92.3, 91.03, 90.03, and 84.46% respectively. The predictive variables considered in the study were gender, higher secondary education, the medium of instruction, private coaching, school location, grade in the course at college, and the target variable was grade in course at test. Through this review, it was necessary to apply classification algorithms to predict students' performance in mathematics under the Management discipline taken during the first year in the first semester.

Various predictor variables for the student to pass or fail in mathematics at the universities include the mathematics background at the ordinary level [18] and continuous assessment, or coursework [19]. Other factors that contribute to the prediction of students' academic performance regardless of the specific course that may be employed in the mathematics performance were determined from the reviewed studies. The predictor variables determined were personal attributes such as gender, age, interest in the study, admission type, and study behavior [20]. Family attributes such as parents' qualification and occupation, family income, status, and support for study (sponsorship) were found to influence the prediction of students' academic performance [21]. Furthermore, academic attributes such as high school grades, students' previous semester marks, class test grades, seminar performance, assignment performance, class and lab work attendance, and previous school marks were also found to contribute to the prediction of students' academic performance [21]. Therefore, some of the predictor variables identified from previous similar studies were used due to data availability, and some of the predictor variables were added from the field to form the dataset to train the ML algorithms for the prediction of students' performance in Mathematics.

From the review of related studies, the present study dwelt on classification algorithms such as KNN, DT, RF, ANN, and SVC to predict the students' performance in a particular subject. These algorithms have performed better in classifying the performance of students who are likely to fail or pass. Also, from the related works presented earlier, the present study dealt with the Management discipline since other disciplines such as science and engineering have been researched more in ML regarding performance prediction. Moreover, the present study observed from related works that Mathematics as a subject has been researched more in primary and high-level schools compared to the university level as one of the subjects that many students fail, hence the need to research it at the university level.

3.Materials and methods

3.1Dataset descriptions

Students' secondary data were collected from both MU campuses; Main and Mbeya. Data were collected to form a total of 3347 instances before cleaning and preprocessing. Students' data were gathered from the admissions, accommodations, examinations, and accounts office as well as the mathematics and statistics studies (MSS) department for five years; 2014/2015 to 2018/2019. The students had registration numbers that were unique for identification. Data collected from the MSS department was the number of instructors involved in teaching mathematics in different academic years. Students' loan allocation data were obtained from the accounts office to know which students had government loans and those who had support from their parents and guardians. Likewise, students' living locations were obtained from the accommodations office to whether the students were living on or off-campus. Students' coursework (CW), remarks, and final examination (FE) results for mathematics were collected from the examinations office. Additionally, data collected were age, ordinary level mathematics grades, gender, and entry categories from the admission office.

The block diagram of this study for mathematics performance prediction starting from the input data to the output information is shown in *Figure 1*. Data obtained from the field was taken for cleaning and preprocessing to remove outliers so that only the necessary information was taken forward and integrated to get a dataset. Thereafter, the useful features are identified and data transformed into the form that ML algorithms can be applied. Then results of the models were obtained and evaluated for knowledge discovery. In the end, useful information was obtained as output.

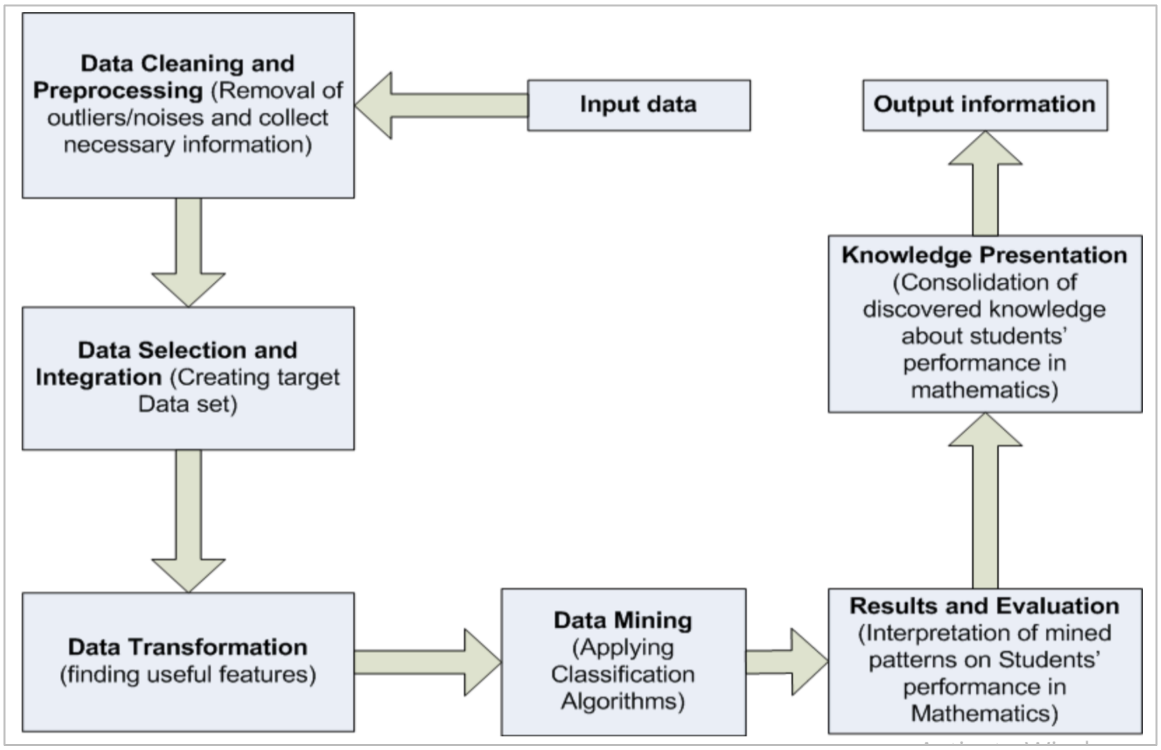

Figure 1 A block diagram for the prediction of mathematics performance. data collection (input data) was followed by data cleaning and preprocessing, data selection and integration, then data transformation, data mining, and the results and evaluation before consolidating the discovered knowledge about students' performance and obtaining the output information

3.2 Data cleaning and pre-processing

After data collection from the field, it was necessary to analyze them using Microsoft Excel (2016 version) and Jupyter Notebook (version 5.7.4) application programs. In excel, functions such as normsdist function were applied to obtain the p-values after getting the z-values in Z-test. Likewise, the find and replace functions in excel were applied in dataset creation as they were easy to access and made data easily visible during manipulation. Python version 3 was used as a powerful programming language that can manipulate data and perform feature engineering. Jupyter Notebook was used for code editing and displaying the analyzed data as it was easy to inspect data and features through graphs and tables [22]. In data analysis, the mentioned tools were able to process 3259 instances of the dataset after cleaning.

Before the application of ML algorithms on training datasets for performance prediction in mathematics, it was necessary to understand, clean and prepare the data. In variable identification, it was necessary to identify the predictor (input) variables and the target variables (output) for performance prediction in mathematics. Also, data types were identified on collected data such as numeric and character, and categories of variables as either categorical or continuous were identified.

In univariate analysis, the variables had to be explored individually using the methods and statistical measures that depended on the category of the variable as either categorical or continuous variables. In the case of continuous variables, it was necessary to understand the central tendency (mean, median, mode, max, and min) and the spread of variables (range, quartile, variance, and standard deviation) using statistical metrics visualization methods such as histograms and box plots which made it easy to identify missing values and outliers. For categorical variables, a frequency table was used to understand the distribution of each category with the count and count percentage metrics for each category, and a bar chart was used to visualize the categorical variables.

In the bi-variate analysis, it was essential to understand the relationship (association and disassociation) between two variables using Bi-variate analysis methods that depend on whether the variables were both continuous (Scatter plot and correlation), both categorical (chi-square test), or categorical and continuous variables (Z-test, T-test or ANOVA) were applied to assess the significance between variables. Moreover, the scatter plot and correlation were applied to find the relationship between numerical variables such as CW and FE. Similarly, the Z-test was applied to find the association between the categorical variables with two classes and the numerical variable such as CW and remarks (Pass or fail). Likewise, ANOVA was applied to find the relationship between the categorical variable with more than two classes such as mathematics ordinary level results in grades (A, B+, B, C, D, E, and F) and numerical variable CW or FE.

Regarding the treatment of missing values, being that a good predictive model should have no missing values in the dataset, the present study employed methods such as removing the entire rows with the missing values. The method was applied because some of the students had neither CW nor FE, as well as remarks had no values. It was necessary to handle them to avoid wrong prediction or classification and biased models. In this study, 88 instances were removed out of 3347 instances as some of the students had postponed studies or examinations. As a result, 3259 instances remained to form a dataset.

Regarding outliers, any observation in the dataset that appeared far away and diverged from the overall pattern was considered an outlier and needed to be treated using methods such as binning and mean imputation. An outlier may cause wrong representation on the dataset as it may affect the mean of the variable in the dataset. In the present study, there were outliers in the age predictor variable in that only four students were aged between 40 and 49. Hence, the binning method was applied to group them in the range of 40-49. Variable transformation was done as a step in feature engineering as the science of extracting more information from the existing data. In variable transformation, the binning method was applied to categorize the variables, for instance, categorical variables v numerical. Variable transformation was done for all categorical attributes using label encoder in SciKit-learn library in Python. Variable creation was done as a step-in feature engineering to obtain more variables from the existing variables by applying methods such as creating derived variables using such as age from the difference between current year and date of birth.

3.3 Feature importance, extraction and selection

Not all the collected must be used to form the dataset for training ML algorithms. Some might be dropped if found not to have importance in improving the performance of the trained ML algorithms and others might be generated from other features. In the present study, the chi-square metric was applied to find the relationship between the two categorical variables; the independent and the dependent variables such as between entry category and remarks, age group and remarks, mathematics ordinary level grades and remarks, living location and remarks and gender with remarks as one of the feature selection techniques [23]. The significance level used was 0.05 to compare with the p-values calculated from the chi-square, Z test, and ANOVA. Furthermore, feature importance determination was performed to calculate the scores of the predictor variables to the target variable remarks with two classes as pass or fail which were calculated from SelectKBest class in scikit-learn library.

3.4The training of machine learning algorithms

The dataset was divided into 60% for training the ML algorithms, 20% for validation of the trained ML algorithms, and 20% for testing the trained ML algorithms. The division of the dataset was considered after the review of the study for reducing dropout rates using ML approaches [24]. Various EDM techniques such as SVM, K-NN, ANN, and DT have been used for the prediction of students' performance have been used for the prediction of students' performance [25, 26]. The present study applied some of the ML algorithms using the data available at the University such as K-NN, SVC, RF, DT, and MLP in the prediction of mathematics performance as they are widely used in EDM with good performance predictions. To train ML algorithms better, it is recommended to train the ML algorithms with a dataset in the range of 60-80% and test with 4-20% [27]. The purpose is to avoid lowering the samples in the trained ML algorithms as it would result in algorithms with very few samples.

3.5 Evaluation of trained machine learning algorithms

The trained ML algorithms were evaluated based on available metrics such as accuracy and correctly and incorrectly predicted instances by Confusion Matrix [28]. Upon the training of the ML algorithms, a validation technique known as K-fold cross-validation (10-fold cross-validation in the present study) was applied as the normally preferred method. This allowed the ML algorithms to be trained on multiple train-test splits and with a small dataset compared to the Hold out validation technique [29]. Hence, it indicated how well the trained ML algorithms performed on unseen data.

Classification evaluation metrics were used to check the performance of the trained ML algorithms to discriminate among the results [30]. Since this study is solving the classification problem, a Confusion Matrix as a table was applied to outline different predictions and test results then contrasted them with real-world values. The evaluation metrics applied to the trained ML algorithms in the present study were recall, precision, accuracy, and F-measure (F1). The recall was calculated as the ratio of the positive class that was correctly detected which indicated how good the trained ML algorithm was to recognize a positive class as shown in Equation (1).

$$Recall = \frac{TP}{(TP+FN)} \quad (1)$$

Precision was calculated as the accuracy of positive class that computed how likely the positive class prediction was correct as shown in Equation (2).

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

Specificity was considered as the ratio of actual negatives from which the trained ML algorithm predicted as a negative class or true negative as shown in Equation (3).

$$Specificity = \frac{TN}{(TN+FP)} \quad (3)$$

In some cases, researchers try to obtain the best precision and recall simultaneously by applying the F1 Score which is the harmonic mean for the precision and recall values. The formula for calculating the F1 score is shown in Equation (4)

$$F - Measure = \frac{2xPrecision\ x\ Recall}{(PRecision+Recall)} \quad (4)$$

In the provided equations, TP and FN represent True Positives and False Negatives respectively. The higher the F1 score, the more the predictive power of the trained classification ML algorithm. Hence, a score was close to 1 meant a perfect trained ML algorithm for prediction. However, a score closes to 0 meant a decrease in the predictive capability of the trained ML algorithm [31]. Therefore, the evaluation metrics applied in the present study were accuracies, recall, precision, and F-measure because the study involved classifying the students as either pass or fail in mathematics. One of the studies employed the mentioned evaluation metrics when comparing classification algorithms such as ANN, NB, DT, LR, and SVM to find the best algorithm for predicting students' performance in board examinations [32].

The present study employed the philosophical Worldview of post-positivism that deals with the cause-and-effect relationship type of research which is a quantitative research approach [33, 34]. The quantitative research approach based on the design science research was used with process steps that included awareness of the problem, suggestions by setting objectives, development, evaluation, and the conclusion [35]. *Figure 2* shows the research method that was applied in this study from the data collection up to the knowledge obtained after applying ML algorithms on the training dataset for the prediction of students' performance in mathematics.

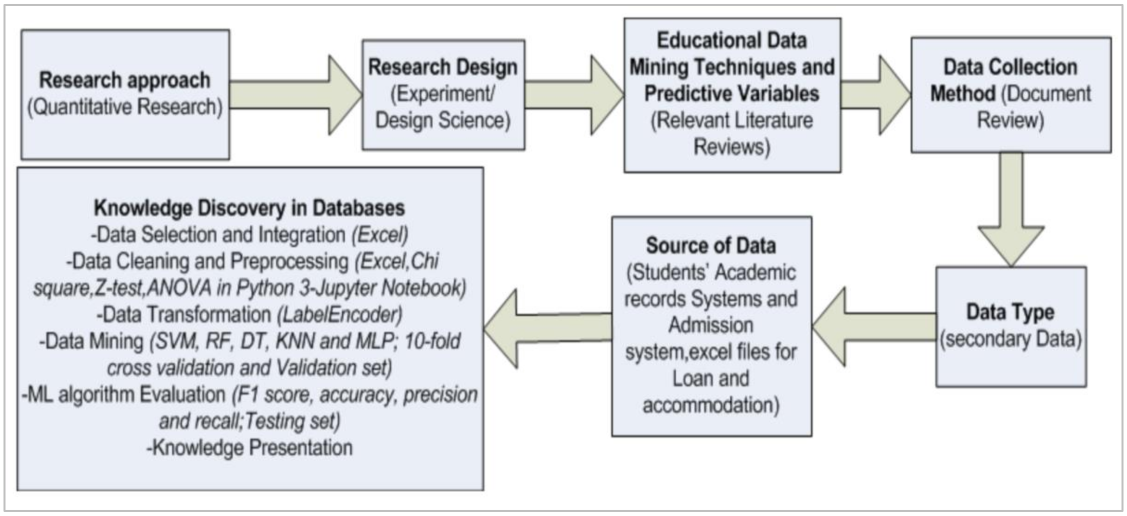

Figure 2 Research methodology. The study involved Quantitative Research approach and Experimental Research Design. A literature review of the available Data mining techniques was done, followed by data collection using Document Review (Secondary data) for knowledge discovery using various algorithms

4.Results

4.1Dataset descriptions

The present study showed that mathematics had high failure rates among the student datasets as shown in *Figure 3*. The year with the highest failure rates in mathematics was the 2016/2017 academic year while the failure rates for the rest of the years ranged from 20 -29%.

Figure 4 indicates the performance of students pursuing four management programmes in mathematics at MU from 2014/2015 up to the 2018/2019 academic year during the first year in the first semester. In the present study, the dataset was explored using frequency distribution of the remarks variable which is categorical. It has been observed that 2403 students (73.7%) had a pass class while the fail class had 856 students (26.3%) from 2014/2015 to 2018/2019 academic years during the first semester of the first year as shown in *Figure 5*.

Mathematics ordinary level grades were explored as shown in *Figure 6* whereby there were 1480 students (45.4%) with F grade, 87 students (2.7%) with E grade, 1309 students (40.2%) with D grade, 326 students (10%) with C grade, 38 students (1.2%) with B grade, 9 students (0.2%) and 10 students (0.3%) with A grade from the 2014/2015 academic year to 2018/2019.

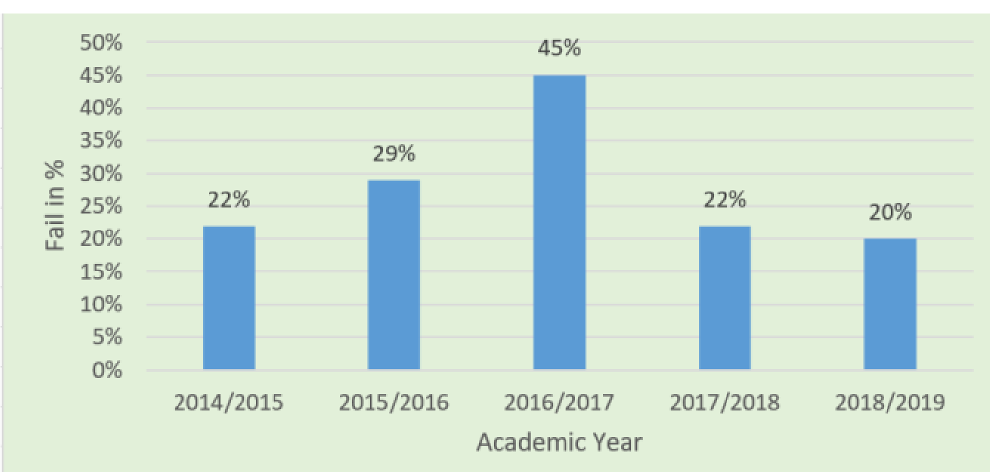

Figure 3 Failing of students in mathematics in management degree students. The 2016/2017 academic year had the greatest failure rates in Mathematics (45%) followed by 2015/2016 at 29%

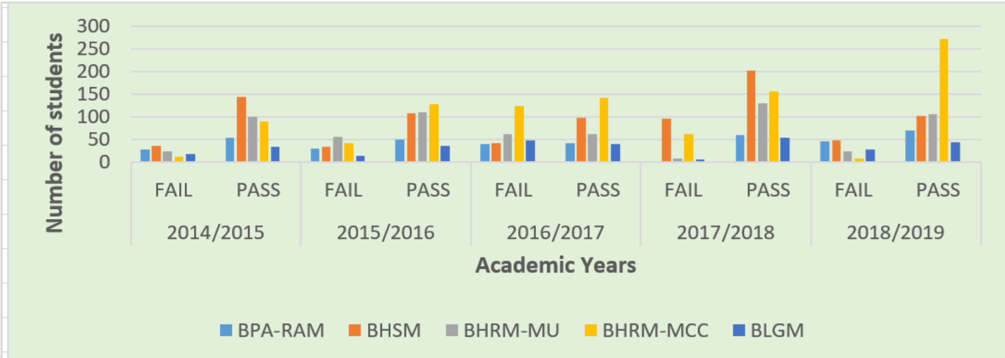

Figure 4 Performance of students pursuing management degrees in mathematics. BLGM = Bachelor of Local Government Management, BHSM = Bachelor of Health System Management, BHRM-MU = Bachelor of Human Resource Management at Mzumbe University, BHRM-MCC = Bachelor of Human Resource Management at Mzumbe University Mbeya Campus College, and BPA-RAM = Bachelor of Public Administration-Records and Archives Management. BHRM-MCC had the highest pass rates in 2018/2019 but tended to be lower than BHSM in 2017/2018 and 2014/2015

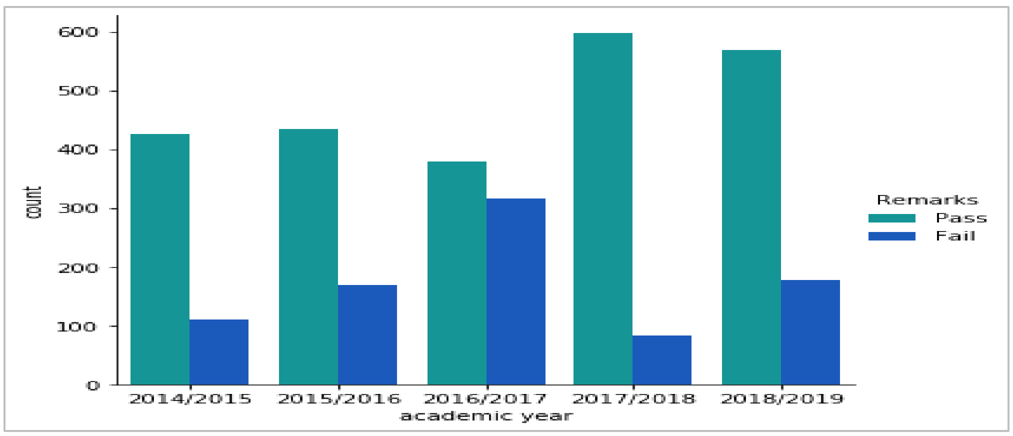

Figure 5 Remarks distribution in dataset. Although the pass tended to be higher than the fail marks in all years, the fails were also remarkably high, especially in the 2016/2017 academic year

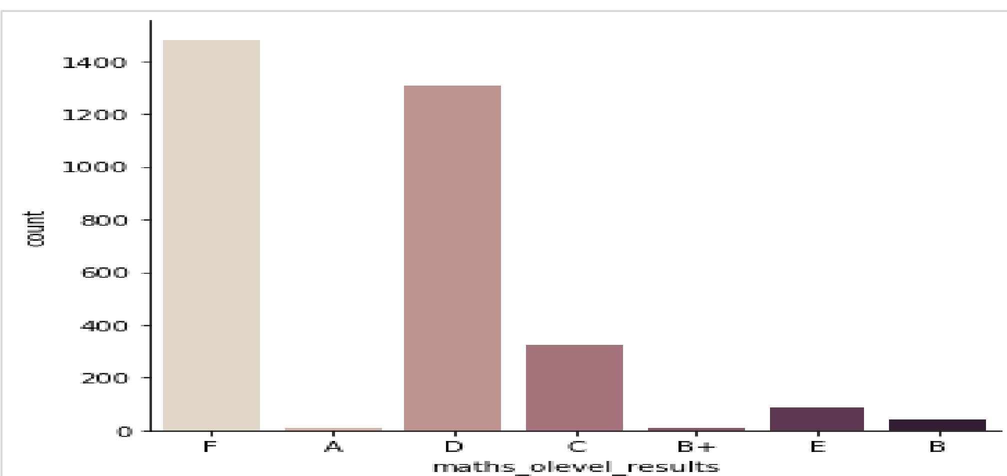

Figure 6 Mathematics ordinary level grades in the dataset. Students who passed ordinary level mathematics well (A, B, B+, C) were very few. Some students passed ordinary level mathematics marginally (D), and the rest failed the examination (F)

It was important to analyze the numerical variables which were FE, CW, and age of the students by using univariate analysis metrics. Mean, standard deviation, minimum, first quartile, second quartile, third quartile, and maximum values were applied and their histograms are shown in *Figure 7*.

Figure 7 shows the three histograms for CW, FE, and student age (age) against the number of students. Coursework included marks for test 1, test 2, individual assignment, and group assignment to make a total of 50% before doing the final examination that have 50% marks. The study showed that the maximum CW was 47 while the minimum was 0.5 with a mean of CW is 27.21 and a median of 27. The first, second, and third quartiles of CW were 23, 27, and 31.65 respectively. For the FE, the maximum and minimum marks were 50 and 0 respectively with a mean of 23.77 and a median of 23. The first, second, and third quartiles for FE were 18, 23, and 30 respectively. Likewise for age, the maximum was 49 and the minimum was 17 with a mean of 22 and a median of 21while the first, second, and third quartiles were 20, 21, and 23 respectively.

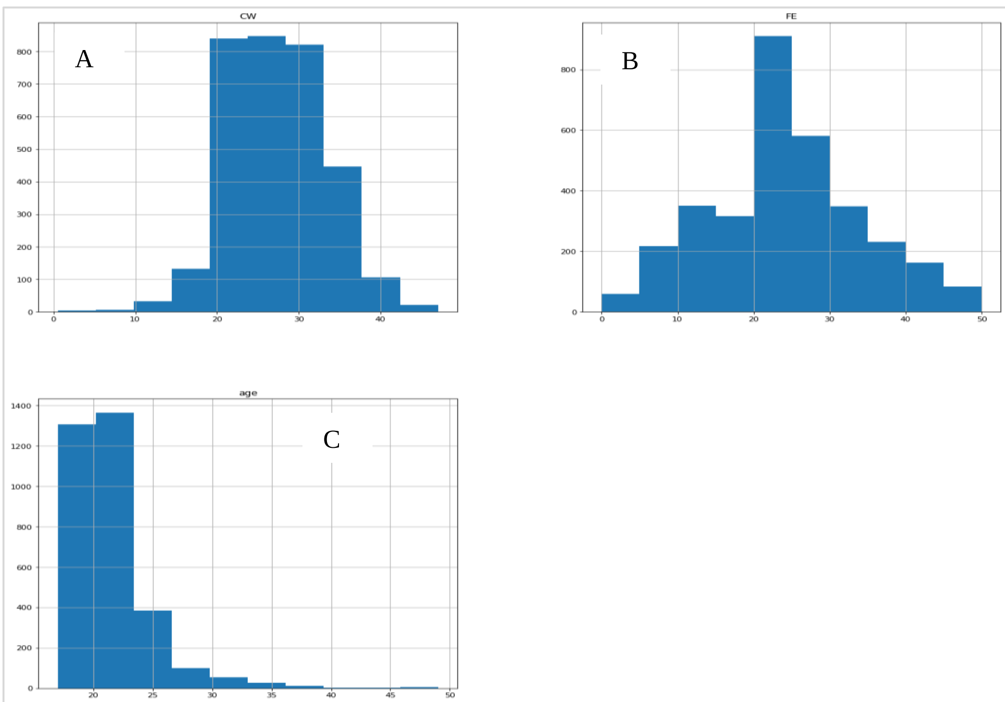

Figure 7 Univariate analysis on Course Work (A), Final Examination (B), and student age (C) in the dataset

Furthermore, the living location of the students was explored using frequency distribution where 679 students (20.8%) were accommodated in university hostels and 2580 students (79.2%) were accommodated in private hostels out of university as shown in *Figure 8*.

Likewise, the frequency distribution of the entry category was explored as either Recognition of Prior Learning (RPL), diploma, or form six as the entry qualification to join the first year in degree management programmes at MU from 2014/2015 to 2018/2019 academic years. It was found that 719 students (22.1%) were from diploma and 2534 students (77.7%) were from form six and 6 students (0.2%) were from RPL entry equivalent as shown in *Figure 9*.

Moreover, the dataset was explored to understand the distribution of the sponsorship variable to students pursuing management degree programmes at MU. The proportion of students who were government loan beneficiaries was 44.5% in five years from 2014/2015 to 2018/2019 academic years. While 55% of the students had private sponsorship during the first year in the first semester as shown in *Figure 10*.

The frequency distribution for gender attributes showed that the males and females in the dataset were 44.5 and 55.5% respectively during the first semester of the first year from 2014/2015 to 2018/2019 academic years (*Figure 11*).

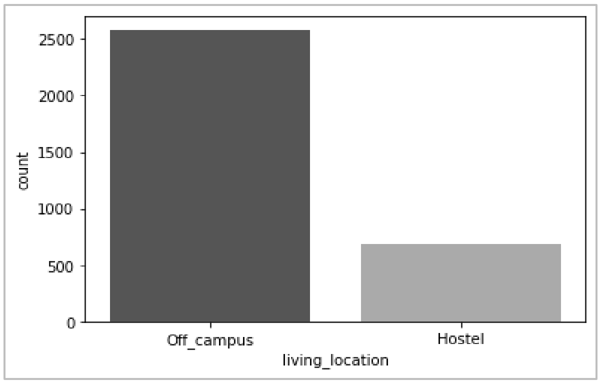

Figure 8 Living location of the students in the dataset. Most students lived off-campus and very few of them lived on campus (in hostels in the university)

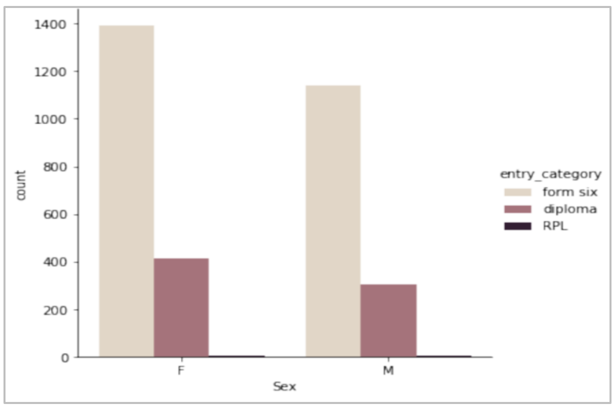

Figure 9 Entry category distribution in dataset. M = Male and F = female. The males who joined the university with form six and diploma grades were more than the females in each case

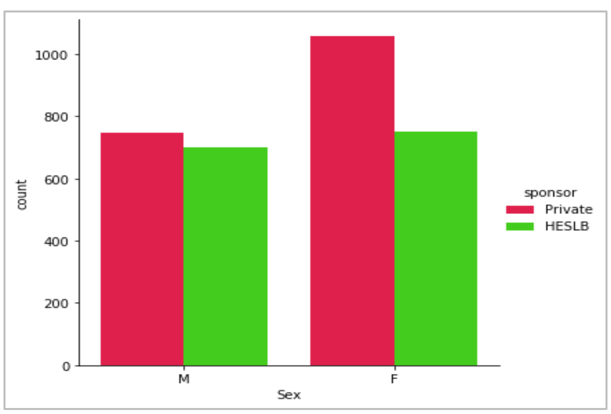

Figure 10 Sponsorship distribution in dataset. M = Male and F = Female. The females who had government and private sponsorships in the university were more than the males

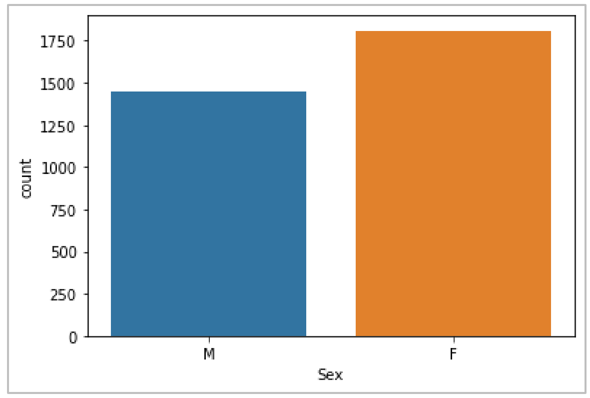

Figure 11 Gender distribution in dataset. M = Male and F = Female. The females were more than the males

4.2 Identifying the requirements for training machine

Learning algorithms. It was necessary to identify the requirements before training the ML algorithms for the prediction of performance in mathematics. The requirements included identification of the predictor variables, dataset creation after data collection from the field, and feature importance, extraction, and selection. Some predictor variables have a direct relation to the output variable remarks such as age group, entry category, mathematics grades at O level, CW, campus location, number of instructors, and FE (*Table 1*). Moreover, some of the predictor variables such as sex, sponsorship, and living location were shown not to influence the performance of the students in mathematics. However, the relationship between the variables was calculated and even combined to find the relationship to the output variable as shown in *Table 1*. If there was a relationship between variables, 'Yes' was applied, and vice versa.

From the feature importance determination shown in *Figure 12*, the score results were such that the FE at the University was the feature with the highest score of 139 that contributed highly to the prediction of students' performance in mathematics. The next feature was ordinary level grades in mathematics with 118.5 scores that also contributed to the prediction of mathematics performance at the University. The other feature was the entry category with 24.4 scores that contributed to the prediction of mathematics performance in the. Also, CW had scores of 20.6 to the target variable remarks in the prediction of mathematics performance. Likewise, the age of the students had scores of 3.5 contributions to the target variable in the prediction of mathematics performance while the campus location and number of instructors both had 1.6 scores. The other three predictor variables were gender, sponsorship, and living locations with scores of 1.1, 0.4, and 0.3 respectively regarding their contribution to the prediction of mathematics performance at the university.

Table 1 Relationship between variables in the dataset

Variables	Age group	Sex	Entry category	Maths ordinary level results	Sponsorship	Living location	CW	FE	Number of instructors	Campus location	Remarks
Age group	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Sex	No	Yes	No	Yes	Yes	No	No	No	Yes	Yes	No
Entry category	Yes	No	Yes	Yes	No	No	No	No	Yes	Yes	Yes
Maths ordinarylevel results	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Sponsorship	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes	No
Living location	Yes	No	No	Yes	Yes	Yes	No	No	No	No	No
CW	No	No	No	No	No	No	Yes	Yes	No	No	Yes
FE	No	No	No	No	No	No	Yes	Yes	No	No	Yes
Number of instructors	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Campus location	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Remarks	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes

CW = Coursework and FE = Final Examination

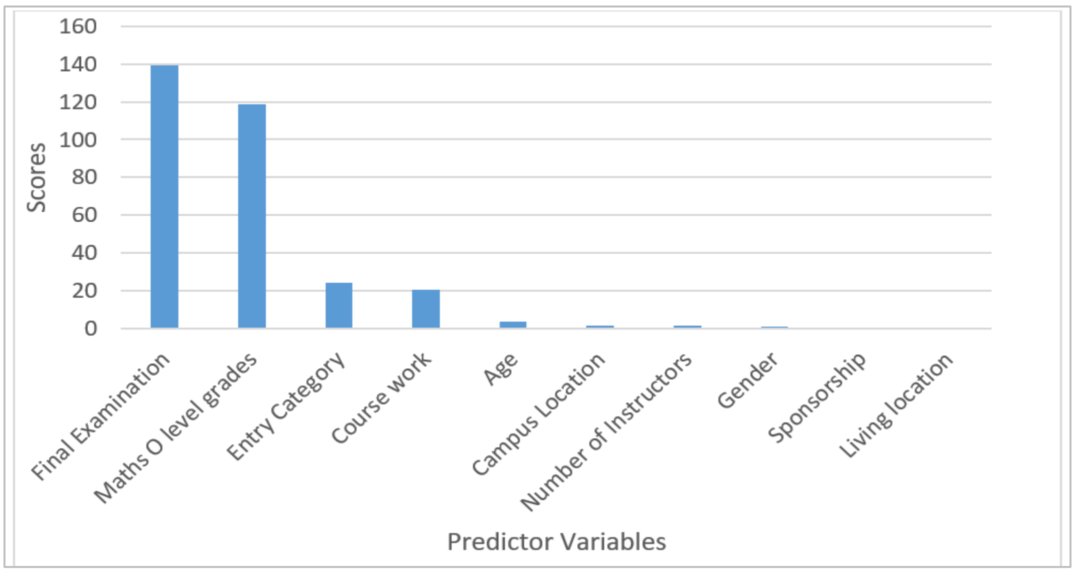

Figure 12 Feature importance to the target variable. The final examination had the greatest role in the prediction of performance in mathematics, followed by the ordinary level entry grades, Entry category, and Course work. The rest of the variables had very minimal influence on the prediction of mathematics performance

The instances involved in the present study were 3259 after removing 88 instances of missing values, irrelevant rows, and outliers after data collection. Moreover, 10 predictor variables were considered after checking their significance on the output variable. Attributes such as CW, FE, ordinary level mathematics grads, sponsorship, living location, age, gender, number of instructors, campus location, and entry categories were used as shown in *Table 2*. The number of instructors involved in teaching mathematics during a particular academic year was identified in the field. From the field, it was found that the number of instructors might differ because mathematics has statistics topics that other instructors may be included to support in teaching. The predictor variable was extracted from the workload allocation for each academic year from 2014/2015 to 2018/2019.

Figure 13 describes the DT classifier from the root node up to the branch nodes that classify the students' performance in mathematics as either Pass or Fail. Some of the rules or patterns have been generated from the decision tree that may be applied to draw conclusions in mathematics performance based on the predictor variables. From rule 1, if the score of the student in the final examination is above 19.45 and course work is below or equal to 16.5 while during admission mathematics grade in ordinary level was F then the student is likely to fail in mathematics at the University. Note that corresponding numerical values of the mathematics ordinary grades from A to F are from 0 to 6 respectively.

Table 2 List of attributes that formed a dataset

S. No.	Variable name	Description	Possible values
1	Age	Student's age	Quantitative data (0-60)
2	Sex	Gender of the student	Qualitative data (Male or female)
3	Entry category	Whether the student is admitted through either form six, diploma or RPL	Qualitative data (form six, diploma or RPL)
4	Maths ordinary level results	Grades obtained at ordinary level	Qualitative data (A, B+, B, C, D, E or F)
5	Sponsorship	Financial supporter of the student at university	Qualitative data (private or loan-HESLB)
6	Living location	The place where students live while at university	Qualitative data (hostel or off campus)
7	Course work (CW)	Marks obtained before FE that include assignments and tests	Quantitative data (0-50)
8	Final examination (FE)	Marks obtained in university final examination	Quantitative data (0-50)
9	Number of instructors	Number of lecturers involved in mathematics course for particular academic years	Quantitative data (1 or 2)
10	Campus location	Location of the campus student admitted	Qualitative data (town or rural)
11	Remarks	Status of the course as either student passed or failed	Qualitative data (pass or fail)

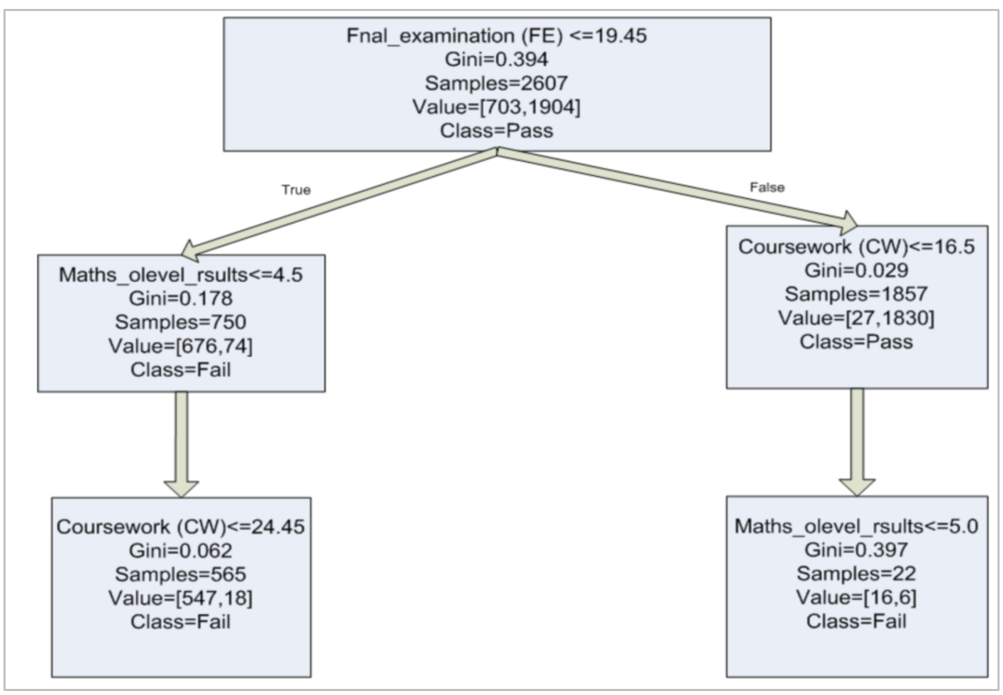

Figure 13 Decision tree (DT) classifier for pattern and knowledge discovery. The DT classifier was used determine pattern for knowledge discovery in mathematics performance. The DT resulted in rules such as Rule 1; if FE > 19.45 and CW<= 16.5 and maths_O-level results > 5.0 then Remarks = Fails and Rule 2; if FE <= 19.45 and maths_O-level results>4.5 and CW <= 24.45 then Remarks = Fail

Moreover, from rule 2 in *Figure 14*, it can be observed that, if the final examination is less than or equal to 19.45 and course work is less than or equal to 24.45 while the mathematics grade in the ordinary level during admission was E or F. Then, the student is likely to fail in mathematics at university from which 547 students from branch node majority belonged to the Fail class. Therefore, from the two samples of observations in patterns from decision tree classifiers, the admission office at University may consider the minimum grade at the ordinary level in mathematics to be admitted at management degree programmes from D and above to reduce the number of failing students in mathematics. This is because from the dataset number of the fail class was 856 students and the majority of students who had failed mathematics in University had also failed at the ordinary level with E or F grade.

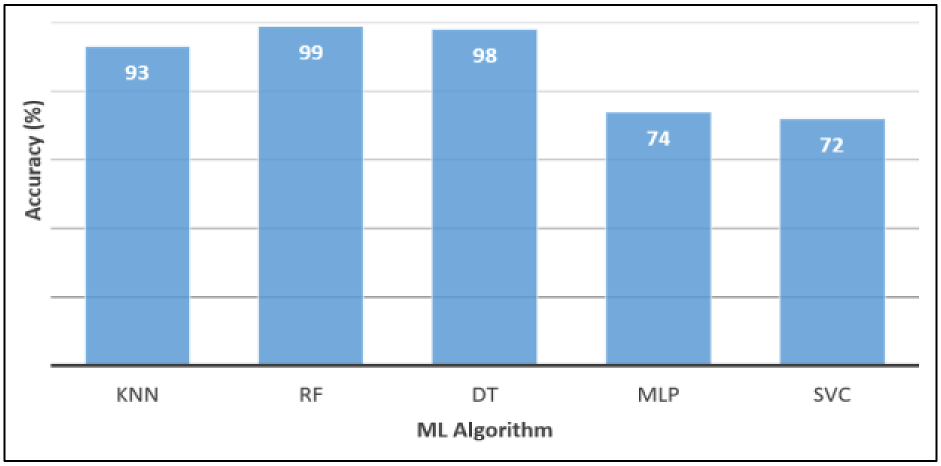

Figure 14 Validation results on accuracy for trained ML algorithms. KNN =K-Nearest Neighbors, RF = Random Forest, DT = Decision Tree, MLP = Multilayer Perceptron, and SVC = Support Vector Classification The accuracy of the algorithms were very high for KNN (93%), RF (99%), and DT (98%)

4.3Identifying the best-trained ML algorithms

This task was necessary to determine the best-trained ML algorithms and select one among them for the prediction of mathematics performance. *Figure 14* depicts the validation results of the trained ML algorithms (K-NN, RF, DT, MLP, and SVC) on accuracy. The accuracy was used to determine the best-trained ML algorithms for better generalization to the unseen dataset (testing set) for the prediction in mathematics performance. The KNN algorithm was one of the best-trained ML with a validation result of 93% accuracy. Similarly, the validation result of the RF ML algorithm was 99% accuracy which was the highest among other algorithms. Likewise, the DT ML algorithm had validation results of 98% accuracy as the second algorithm with the highest accuracy. The MLP and SVC had validation results of 74 and 72% accuracy respectively which did not perform better compared to the other three mentioned algorithms. *Figure 15* depicts the validation results of the trained ML algorithms on F1-score which is an important metric for classification problem that indicates how well the trained ML algorithm predicts the pass or fail classes. F1-score is also an important metric for the validation of trained ML algorithms as it takes precision and recall simultaneously.

Figure 15 shows the five trained ML algorithms against the other performance metric which is F1-score as widely used in classification problems. The higher the F1-score value, the better the trained ML algorithm. The K-NN, RF, and DT algorithms had the best F1-score validation results compared to MLP and SVC. For K-NN, 87% F1-score (Fail class) and 96% F1-score (Pass class) were obtained while for the RF algorithm, the F1-scores for the Fail and Pass classes were 98 and 99% respectively. In the DT algorithm, the F1-score results were 96% and 99% for the fail and pass classes respectively. For the MLP algorithm, 1% was the F1-score for prediction of fail class which shows that it could not classify well while 85% F1-score for pass class. Likewise, SVC had a 62% F1-score for the fail class and 78% F1-score for the pass class. From the validation results of accuracy and the F1-score of trained ML algorithms, the consideration of accuracy only for the selection of the best-trained ML algorithms is not enough in classification problems. Hence, the F1-score should also be considered to have the best ML algorithms for the prediction of unseen data during evaluation. The purpose of validation results for the trained ML algorithms was to avoid overfitting or underfitting when comparing them against the evaluation results of the testing set (unseen data) during prediction.

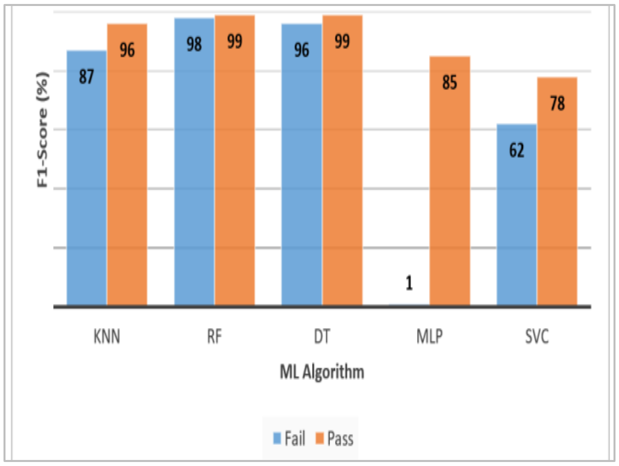

Figure 15 Validation results on F1-Score for the trained ML machine learning algorithms

4.4 Evaluation of the best selected trained ML algorithms

After training ML algorithms and selecting the best as shown in section 4.2, different evaluation metrics for ML algorithms were applied to determine the performance prediction in mathematics using the testing set (20% of the dataset). Evaluation metrics such as precision, recall, accuracy, and F1-score were calculated from the TP, TN, FP, and FN values of the Confusion Matrix for each classifier. The testing set was used in the evaluation of the trained ML algorithms to find the values of the evaluation metrics such as correctly classified classes, incorrectly classified classes, precision, recall, F1-Score, and accuracy as shown in *Table 3*.

Figure 16 shows the number of classified instances concerning the best ML algorithms. The RF algorithm was the best as it correctly classified 649 instances during the prediction of students' performance in mathematics and incorrectly classified only 4 instances out of 653 instances from the testing set. Likewise, the DT ML algorithm classified correctly 646 instances and 7 instances incorrectly and the K-NN algorithm classified 628 instances correctly and 25 instances incorrectly.

Table 3 Evaluation of selected best trained ML algorithms

S. No.	ML algorithm	Support	Correctly classified instances	Incorrectly classified instances	Precision	Recall	F1-score	Accuracy
1	DT	Fail (157)	155	2	0.98	0.99	0.98	0.99
		Pass (496)	491	5	1.00	0.99	0.99	
2	RF	Fail (157)	156	1	0.98	0.99	0.99	0.99
		Pass (496)	493	3	1.00	0.99	1.00	
3	KNN	Fail (157)	146	11	0.91	0.93	0.92	0.96
		Pass (496)	482	14	0.98	0.97	0.97	

ML = Machine Learning, DT = Decision Tree, RF = Random Forest, and KNN = K-Nearest Neighbors

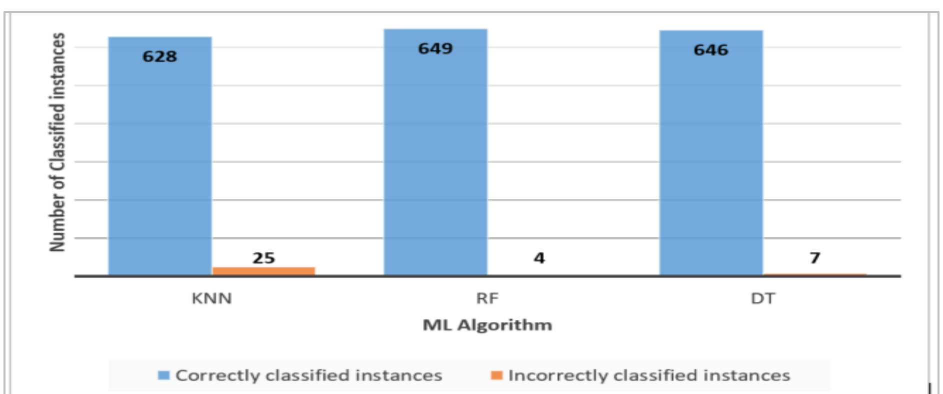

Figure 16 Number of classified instances concerning the best machine learning algorithms. The RF algorithm was the best as it correctly classified 649 instances on predicting students' performance in mathematics and incorrectly classified only 4 instances out of 653 instances from the testing set

Similarly, a comparison of the accuracies is shown in *Figure 17* which depicts the validation and testing results on accuracies of the selected best-trained ML algorithms. The selected trained ML algorithms had almost similar accuracies in validation and testing such as 96% accuracy in testing for the K-NN algorithm and 93% in the validation result which is an increase of 3%. Likewise, the accuracy during testing for RF was similar to that during validation at 99%. Furthermore, the accuracy in the testing set for DT has increased by 1% from 98% in the validation results to 99% in testing results. Therefore, from the results of *Figure 17*, it can be observed that there was no much difference in the accuracies of the best-trained ML algorithms when are tested with an unseen dataset.

Furthermore, the F1-scores in testing were compared with the validation results during the training of the ML algorithms as shown in *Figure 18*. For the RF algorithm in the fail class, F1-score increased by 1% from 98% in the validation results to 99% in testing results, and also by 1% from 99% in the validation results to 100% in testing results. Moreover, the DT algorithm in the fail class, F1-score increased by 11% from 87% in the validation results to 98% in testing results, and also by 3% from 96% in the validation results to 99% in testing results. For the K-NN algorithm, F1-score in the fail class decreased by 4% from 96% in the validation results to 92% in testing results and decreased by 2% from 99% in the validation results to 97% in testing results for the pass class.

From *Figure 18* it can be seen that the best-trained ML algorithms were validated and tested using F1-score performance metric for both fail and pass class. The results showed that the models were the best as there was no much difference in the F1-scores obtained during validation and testing of the models on an unseen dataset.

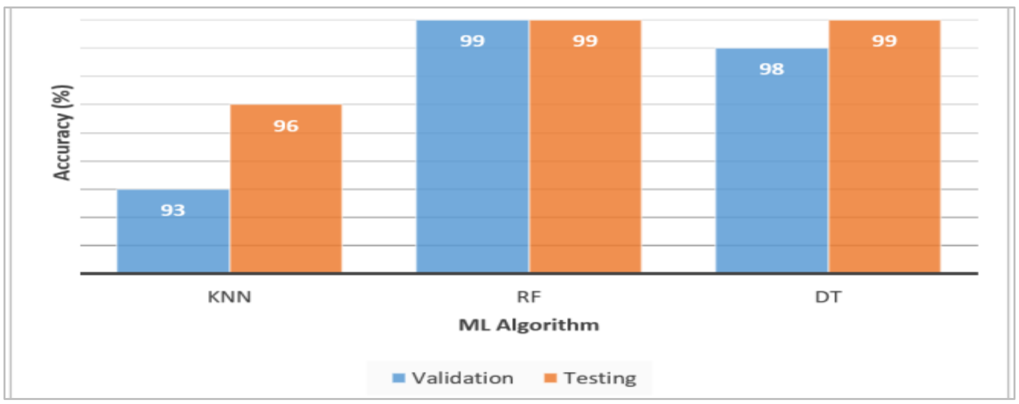

Figure 17 Comparison of validation and testing results on accuracy for selected best-trained ML Machine Learning algorithms. The selected trained ML algorithms had almost similar accuracies in validation and testing

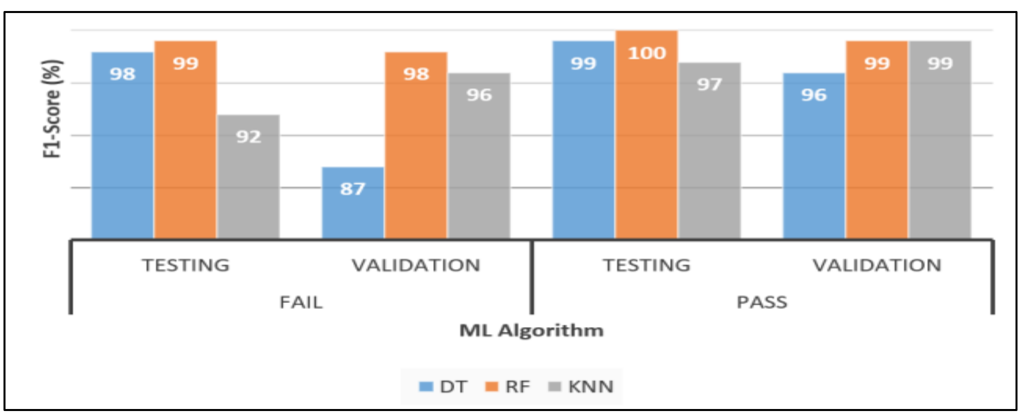

Figure 18 Comparison of validation and testing results on F1-score for selected best-trained ML Machine Learning algorithms

5.Discussion

5.1Identification of the requirements for prediction performance

In the present study, the instances collected were enough to form a dataset as it was compared with previous studies that researched the prediction of students' performance with 273 instances [33]. Also, another study researched with 270 instances in performance prediction using classification algorithms [28]. In one of the studies, the researcher predicted the students' performance in mathematics with 279 instances from 2007 to 2010 academic years basing on attributes such as oral, test, and final grades in the first and second semester [35]. Compared to this study, more instances (3259) were involved and 10 predictor variables were considered after checking their significance on the output variable.

From rule 1 (*Figure 12*), if the score of the student in FE is more than 19.45 and the CW is less than or equal to 16.5 while the ordinary level mathematics grade during admission was F, then the student is likely to fail in mathematics at the university. From rule 2 (*Figure 12*), if the FE is less than or equal to 19.45 and CW is less than or equal to 24.45 while the ordinary level mathematics grade during admission was E or F, then the student is likely to fail in mathematics at university from which 547 students from branch node majority belong to the Fail class. Therefore, from the two rules in patterns from DT classifiers, the university admissions office may consider the minimum grade in ordinary level in mathematics to be D and above for admission into Management degree programmes to reduce the number of failing students in mathematics. This is because from the dataset, the number of the fail class was 856 students, and a majority of students who had failed mathematics in university had also failed mathematics at the ordinary level with E or F grades.

5.2Training and validation of machine learning algorithms

The DT accuracy in the present study was 98%. The validation results were compared with a study that involved the training ML algorithms with all features due to the dependences of predictor variables to each other where the DT accuracy was 91.5% [12]. Moreover, the RF and MLP in Ma and Zhou [12] had accuracy levels of 72.4 and 88.3% respectively while in the present study, the values were 99 and 74% respectively. Likewise, the SVC algorithm in the present study was 72% while 92.6% in [12]. These results show that the trained algorithm can predict the performance in mathematics well when trained with all features due to their dependencies. Also, one of the studies used the F-measure as the evaluation metric when predicting the students' success in Introductory Mathematics [10]. Hence, the present study applied the F-measure metric to validate the trained ML algorithms well. The F1-score for the fail and pass classes indicates how correctly the trained ML algorithm can predict two classes respectively on the validation dataset. Since accuracy is the evaluation metric of the best-trained ML algorithms, also K-NN, RF, and DT had the best F1-score validation results compared to MLP and SVC. For K-NN, the F1-scores were 87 and 96% for the fail and pass classes respectively while for the RF algorithm, the F!-scores for the fail and pass classes were 98% and 99% F respectively. In the DT algorithm, the F1-score results were 96% and 99% for the fail and pass classes respectively. However, for the MLP algorithm, 1% was the F1-score for the prediction of the fail class and 85% for the pass class which shows its inadequacy at classifying well. Likewise, SVC had a 62% F1-score for the fail class and 78% F1-score for the pass class.

5.3Evaluation of trained machine learning algorithms

A comparison of the accuracies depicted the validation and testing results on accuracies of the selected best-trained ML algorithms (*Figure 17*). The selected trained ML algorithms had almost similar accuracies in validation and testing such as 96% accuracy in testing for the K-NN algorithm which had 93% in the validation result which was an increase of 3%. The accuracy during testing for RF was similar to that during validation. Furthermore, the accuracy in the testing set for DT increased by 1% from 98% in the validation results to 99% in the testing results. From the comparison of accuracies on validation and testing, there was no overfitting or underfitting on the best selected trained ML algorithms. Furthermore, the F1-scores in testing were compared with the validation results during the training of the ML algorithms (*Figure 18*). For the RF algorithm in the fail class, the F1-score increased by 1% from 98% in the validation results to 99% in testing results and by 1% from 99% in the validation results to 100% in testing results. Moreover, the DT algorithm in the fail class, the F1-score increased by 11% from 87% in the validation results to 98% in testing results, and by 3% from 96% in the validation results to 99% in testing results. For the K-NN algorithm, the F1-score in the fail class decreased by 4% from 96% in the validation results to 92% in the testing results and decreased by 2% from 99% in the validation results to 97% in testing results for the pass class.

These results show that the prediction in mathematics performance using unseen data was successfully done using the RF and DT algorithms both of which had the highest accuracy of 99% followed by K-NN at 96%. Also, the prediction of mathematics performance on unseen data was successfully done for the fail class where the RF algorithm performed best with an F1-score of 99%, followed by DT with 98%, then K-NN with 92%. Similarly, in the pass class, the RF algorithm had the highest F1-score of 100%, followed by DT with 99%, and then K-NN with 97%. The results were compared with a previous study that compared the 5-level grading system which had an accuracy of 71.14% for the RF algorithm and 91.39% for the binary level grading in mathematics [13]. In the same study, the DT algorithm had an accuracy of 73.42% when the 5-level grading was applied and then the accuracy increased to 89.11% when the binary level grading was used. The results obtained in the present study had high accuracies that involved binary classification as to whether the students would pass or fail in mathematics since accuracy increases as the classification levels decrease. From this discussion, it can be concluded that the best ML algorithm for the prediction of mathematics performance in the present study was the RF with an accuracy of 99% and F1-scores of 99 and 100% for the fail and pass classes respectively. As a result, the RF predictive model for mathematics performance by Management degree students in the present study was the best. The HLIs that may apply the trained ML algorithm are those that offer art-based Management programmes such as the Institute of social work, Jordan University College, Local Government Training Institute, Moshi Co-operative University, and The Mwalimu Nyerere Memorial Academy, and others in Tanzania.

5.4 Limitation of the study

This study faced limitations in the data collection due to time and financial constraints. Therefore, this led to the elimination of some of the features such as the marital status of students, parents' educational status, and parents' economic status whose collection needed more time and financial support.

6. Conclusion and future work

The present study aimed at predicting the performance of Management degree students in mathematics MU in Tanzania, as a case study. The requirements for the performance prediction were to be determined such as the predictor variables to create a dataset to train ML algorithms. The dataset was created after the data collection based on the predictor variables. Then, 60% of the dataset built the training dataset, 20% of the dataset was used to validate the trained ML algorithms, and 20% of the dataset to evaluate trained ML algorithms. The study involved 5 ML algorithms training were selected based on their wide application and good performance in EDM for performance prediction as discerned from literature review. The ML algorithms that were trained in the present study were MLP, SVC, RF, DT, and K-NN, and during validation, the best algorithms were RF, DT, and K-NN. Further evaluation of the three best ML algorithms showed that the RF ML algorithm was the best in predicting mathematics performance with an accuracy of 99% and F1-scores of 99% and 100% for the fail and pass classes respectively. As a result, the RF predictive model for mathematics performance for Management degree students was established to be the best in the present study. Moreover, the DT algorithm managed to generate rules that were applied to recommend the minimum grade of D in ordinary level mathematics for admission into Management degree programmes to reduce the failure rates. Therefore, the study was successful in achieving all three specific objectives to reach the main objective of predicting the mathematics performance of Management degree students in HLIs using EDM techniques.

The present study did not address some issues that may be implemented by other researchers in the future as follows. The study was based only on the secondary data that were available which resulted in the use of a few features as predictor variables for the prediction of students' performance in mathematics. Therefore, other researchers may consider primary data to have more features by preparing questionnaires to capture data such as parents' occupations and education, students' relationship status, time taken to study mathematics privately, interests in mathematics, and family sizes. Moreover, the present study was based on HLIs for degree students, therefore, other researchers may apply the ML algorithms in primary and secondary schools, and colleges for certificate and diploma students and consider other programmes such as engineering, computer science, and business programmes at HLIs. Similarly, the study did not consider the online courses in electronic learning such as Moodle to get features like time spent working on courses in Moodle, uploaded assignments in Moodle, and participation in online group discussions. Therefore, other researchers may involve the study of online courses.

Likewise, the present study applied some of the performance metrics which are Confusion Matrix, accuracy, precision, recall, and F-measure in the validation and evaluation of the prediction model. Other researchers may include more performance metrics for model validation and evaluation such as PMSE and Root Mean Square Error (RMSE) for more comparison of models in selecting the best. Also, the present study trained 5 ML algorithms; K-NN, RF, SVC, DT, and MLP as the widely used algorithms in EDM for performance prediction of the students. Other algorithms such as NB and LR may be applied by other researchers to solve the problem of the students' performance at primary and secondary schools, colleges, and HLIs. Likewise, other researchers may consider working on the performance prediction of final-year students while considering all courses as to either the student will graduate or not. This will assist students to be advised early on measures to take to graduate, or avoid discontinuation.

Acknowledgment

The authors would like to express their special thanks to Mzumbe University, Tanzania for supporting this study.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] https://www.vistacollege.edu/blog/resources/higher-education-in-the-21st-century. Accessed 13 January 2020.
[2] Osmanbegovic E, Suljic M. Data mining approach for predicting student performance. Economic Review: Journal of Economics and Business. 2012; 10(1):3-12.
[3] Ashenafi MM. A comparative analysis of selected studies in student performance prediction. International Journal of Data Mining & Knowledge Management Process. 2017; 7(4):17-32.
[4] http://hdl.handle.net/1822/8024. Accessed 13 January 2020.
[5] Sivasakthi M. Classification and prediction based data mining algorithms to predict students' introductory programming performance. In international conference on inventive computing and informatics 2017 (pp. 346-50). IEEE.
[6] Olatunji SO, Aghimien DO, Oke AE, Olushola E. Factors affecting performance of undergraduate students in construction related disciplines. Journal of Education and Practice. 2016; 7(13):55-62.
[7] Fussy D. The status of academic advising in Tanzanian universities. KJEP.2018; 15(1):81-98.
[8] https://www.tcu.go.tz/sites/default/files/Admission%20Guidebook%20Direct%20Entry%2006.10.2020.pdf. Accessed 13 January 2020.
[9] Shin D, Shim J. A systematic review on data mining for mathematics and science education. International Journal of Science & Mathematics Education. 2021; 19(4):639-59.
[10] Sokkhey P, Navy S, Tong L, Okazaki T. Multi-models of educational data mining for predicting student performance in mathematics: a case study on high schools in Cambodia. IEIE Transactions on Smart Processing and Computing. 2020; 9(3):217-29.
[11] Ingale NV. Survey on prediction system for student academic performance using educational data mining. Turkish Journal of Computer and Mathematics Education. 2021; 12(13):363-9.
[12] Ma X, Zhou Z. Student pass rates prediction using optimized support vector machine and decision tree. In 8th annual computing and communication workshop and conference 2018 (pp. 209-15). IEEE.
[13] Ünal F. Data mining for student performance prediction in education. Data Mining-Methods, Applications and Systems. 2020.
[14] Nikam SS. A comparative study of classification techniques in data mining algorithms. Oriental Journal of Computer Science & Technology. 2015; 8(1):13-9.
[15] Pandey M, Taruna S. A comparative study of ensemble methods for students' performance modeling. International Journal of Computer Applications. 2014; 103(8):26-32.
[16] Yadav SK, Pal S. Data mining: a prediction for performance improvement of engineering students using classification. arXiv preprint arXiv:1203.3832. 2012.
[17] Widyahastuti F, Tjhin VU. Predicting students performance in final examination using linear regression and multilayer perceptron. In international conference on human system interactions 2017 (pp. 188-92). IEEE.
[18] Kumah MS, Region V, Akpandja TK, Region V, Djondo BI, Region V. Factors contributing to the poor performance in mathematics: a case study among students' in colleges of Education-Ghana. ResearchJournali's Journal of Mathematics. 2016; 3(2):1-2.
[19] Okwute AO, Musa CD. Investigating achievement of continuous assessment techniques on students' discovery in mathematics performance at SSSI level at GDSS bwari, bwari area council, Abuja, FCT. International Journal of Innovative Research in Information Security. 2018; 5(06):2014-9.
[20] Saa AA, Al-Emran M, Shaalan K. Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. Technology, Knowledge and Learning. 2019; 24(4):567-98.
[21] Kumar M, Singh AJ, Handa D. Literature survey on student's performance prediction in education using data mining techniques. International Journal of Education and Management Engineering. 2017; 7(6):40-9.
[22] Slater S, Joksimović S, Kovanovic V, Baker RS, Gasevic D. Tools for educational data mining: a review. Journal of Educational and Behavioral Statistics. 2017; 42(1):85-106.
[23] Zaffar M, Savita KS, Hashmani MA, Rizvi SS. A study of feature selection algorithms for predicting students academic performance. International Journal of

Advanced Computer Science and Applications (IJACSA). 2018; 9(5):541-9.
[24] Mduma N, Kalegele K, Machuve D. Machine learning approach for reducing students dropout rates. International Journal of Advanced Computer Research.2019; 9(42):156-69.
[25] Roy S, Garg A. Analyzing performance of students by using data mining techniques a literature survey. In uttar pradesh section international conference on electrical, computer and electronics 2017 (pp. 130-3). IEEE.
[26] Shahiri AM, Husain W. A review on predicting student's performance using data mining techniques. Procedia Computer Science. 2015; 72:414-22.
[27] https://www.researchgate.net/post/What_is_the_best_way_to_divide_a_dataset_into_training_and_test_sets. Accessed 08 February 2020.
[28] Saa AA. Educational data mining & students' performance prediction. International Journal of Advanced Computer Science and Applications. 2016; 7(5):212-20.
[29] https://medium.com/@eijaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f. Accessed 08 February 2020.
[30] Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process. 2015; 5(2):1-11.
[31] https://medium.com/analytics-vidhya/performance-metrics-for-classification-problem-74935775cdca. Accessed 20 February 2020.
[32] Rustia RA, Cruz MM, Burac MA, Palaoag TD. Predicting student's board examination performance using classification algorithms. In proceedings of the 2018 7th international conference on software and computer applications 2018 (pp. 233-7).
[33] Zohair LM. Prediction of Student's performance by modelling small dataset size. International Journal of Educational Technology in Higher Education. 2019; 16(1):1-8.
[34] Livieris IE, Drakopoulou K, Pintelas P. Predicting students' performance using artificial neural networks. In panhellenic conference with international participation information and communication technologies in education 2012 (pp. 321-8).
[35] Vihavainen A, Luukkainen M, Kurhila J. Using students' programming behavior to predict success in an introductory mathematics course. In Educational Data Mining 2013.

Paul K Mushi, is currently working as Assistant Lecturer in Faculty of Science and Technology, Mzumbe University, Tanzania. His areas of interests are Data Science, Machine Learning and Data Mining in Education.

Email: pkmushi@mzumbe.ac.tz

Daniel Ngondya, Ph.D, is currently working as Lecturer at The University of Dodoma's College of Informatics and Virtual Education. His research interests lies in Optimization of Utility Systems and Modeling of Intelligent Systems.

Email: dngondya@gmail.com