INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023

# An Enhanced Student Engagement and Academic Performance Predictive System

**Ubani Kingsley Chukwuemeka[1], Obayi Adaora Angela[2], Oloruntobe Samson Abiodun[3], Agbo Jonathan Chukwunike[2], Cassia Anwar C.[4]**

[1]*Department of Computer Science, Federal Polytechnic Oko.*

[2]*Department of Computer Science, University of Nigeria, Nsukka.*

[3]*Department of computer science, caleb university, imota, lagos.*

[4]*Department of Computer Science, Federal Polytechnic Nekede Owerri.*

DOI: https://doi.org/10.51583/IJLTEMAS.2023.12506

Received: 27 April 2023; Revised: 08 May 2023; Accepted: 11 May 2023; Published: 07 June 2023

**Abstract:** Student engagement is one of the important constructs that is used to understand the behavior of the student towards the teaching-learning process, and it determines the students' academic performance. The aim of this study is for predictive analytics to work on the comprehension of student commitment and scholarly execution and anticipate students who are in danger of low execution or commitment right on time before the evaluation to work with conceivable mediation to further develop the learning results in advanced education. This research adopts the process of machine learning such as linear regression, decision tree, naïve Bayes, KNN, Kmeans in order to identify the most effective determinants for student academic performance prediction. The result of this study shows that after testing the five attributes, we discovered so far that the attributes that has impact on student evaluation are their Race/Ethnicity and Parental level of Education. Thus, the early prediction of student performance can trigger educators to track student dropouts in a particular course at an early stage. The model can also be used as an early warning system to identify failure students in the classroom by the course coordinators and educators, to take strategic decisions to improve student performance.

**Keywords:** Students engagement, academic performance, machine learning.

## I. Introduction

Student engagement alludes to a significant commitment all through the learning climate. It is best perceived as a connection between the understudy and the school, instructors, understudies, guidance and educational program (Martin and Torres, 2016). Understudy commitment is one of the significant builds that is utilized to comprehend the way of behaving of the understudy towards the educating growing experience. Understanding the way of behaving of understudies in the scholastic foundations will give a brief look at how the directions and scholarly practices are heading on in the college. Students' commitment has three aspects which are social, close to home, and mental. Social commitment alludes to understudy's cooperation in scholar and extracurricular exercises. Profound commitment alludes to Students positive and negative response to companions, educators and school. While mental commitment discusses Students mindfulness and eagerness to dominate troublesome abilities (Fredericks et al., 2004). Understudy commitment isn't simply advantageous to scholastic status of the school yet to its monetary life as well. As portrayed by Markwell (2007) when colleges and universities are progressively centered around the significance of effort to graduated class and other possible companions of the establishment with the end goal of extraordinarily expanding humanitarian help for advanced education, it is turning out to be all the more broadly perceived that how drawn in understudies are and feel themselves to be during their Student years will have an extraordinary bearing on how associated and strong towards the foundation they are probably going to be in later years. Student's execution prescient examination plays had an essential impact in schooling as of late. It considers the comprehension understudies' learning ways of behaving, the distinguishing proof of in danger Students, and the advancement of bits of knowledge into educating and learning improvement. Student execution expectation has been an extraordinary expansion to the intellectual/training industry. Having the option to anticipate the presentation of Students in light of specific highlights and attributes assists schools with arranging appropriately on center regions for various classes of understudy and in particular distinguish misbehavior. The forecast of scholarly execution, in its most fluctuated structures, has gotten extensive consideration in the space of training, particularly concerning advanced education, Zhang, (2018). Prescient examination in training includes construing dubious future occasions or results connected with learning or educating. A few errands foresee a part of educating, like course enlistment, understudy maintenance, or the effect of a given educational procedure on students. Different errands center on learning and students' viewpoints, for example, anticipating scholarly achievement, course grades or ability procurement. Student

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023

execution expectation is one of the huge regions since it tends to be utilized to improve student scholarly execution and diminish understudy whittling down, for instance, by utilizing early advance notice frameworks to help in danger students (Akçapınar, 2019, p.16). This data can be utilized to foster mediations and backing programs that can assist with further developing student commitment and scholastic execution, and at last to help student achievement.

## II. Student's Engagement in Academics

Students' engagement is a huge consider upgrading their advantage and mindfulness toward advancing by diminishing their disappointments, Handling and Makar (2008). In this specific situation, understudy commitment in numerical assignments and exercises might improve accomplishment by affecting understudies' close to home, mental, and social turn of events, (Christenson et al., 2012). Commitment is a significant develop in math learning, (Lehtinen et al., 2017). That associates hypothetical and reasonable issues in mental commitment and understudies' admittance to information and data to take care of numerical issues, (Handling and Makar, 2008). This scholastic commitment happens when Students plunge profound into learning exercises, when they are intellectually and sincerely consumed by the review materials, and frequently while collaborating with peers. It goes past "surface learning" (Hattie, 2003) like substance retention and satisfying prerequisites to accomplish a passing grade for a course. It brings Students into extraordinary reasoning exercises like examining and figuring out ideas, supporting methods, and finding meaning. Additionally, includes social collaboration with peers and the educator, through trading encounters, information, sentiments, and backing. The following are instances of Students commitment towards scholastics: Behavioral commitment, Emotional commitment, Cognitive commitment.

### 2.1 Student Performance Prediction

Predicting students' performance is vital in issues connected with advanced education as well likewise as to profound learning and its relationship to instructive information. Expectation of understudies' exhibition offers help in choosing courses and planning suitable future review plans for understudies. With the quick development in the volume of instructive information, systemic methodologies have proceeded to develop and mature, alongside complex information examination strategies

(Dutt et al. 2017). The term 'understudy execution' is utilized diversely across various investigations. There are by and large two sorts: execution at the program level (Burgos et al. 2018) and at the course level (Marbouti et al, 2015). For the program level, expectation undertakings can be connected with recognize the likelihood of understudy dropouts or graduates from a degree program (e.g., a four-year certification). For the course level, understudy execution is characterized as understudies' learning results like tasks or evaluations in their courses after a review period (Lemay et al, 2020). Its expectation errands can zero in on anticipating understudies' last scores or grades, or their pass or bomb status toward the finish of the course (Yang et al, 2020). In danger understudies are the people who are bound to bomb the course (Akçapınar et al, 2019). The point of this kind of expectation task is to recognize in danger understudies and assist them with accomplishing their scholastic objectives. This examination centers around the course-level execution expectation and researches understudy learning ways of behaving on different learning undertakings. Prescient examination at the course level can give important bits of knowledge to improving instructing and learning. It can assist educators with grasping understudy conduct (e.g., the all out time in learning content survey) (Oliva-Cordova et al. 2021). It can likewise give understudies who experience gaining troubles the potential chances to get convenient mediations from their teachers (Mangaroska et al. 2018) and change their learning methodologies. Besides, the joining of prescient investigation has been displayed to upgrade instructive methodologies according to a foundation point of view. Past examination has utilized highlight significance examination in execution forecast assignments to decide the impact of understudy related highlights on scholarly execution (Waheed et al. 2020). Distinguishing these significant highlights can give important bits of knowledge to further developing educating and learning (Behr et al, 2020). Various methodologies and strategies are taken on in understudy execution forecast undertakings. From the information utilization point of view, relative examination of the prepared datasets (e.g., producing different capabilities through highlight extraction) to approve prescient models has less consideration. Besides, in grouping undertakings for anticipating understudy execution, a few information mining calculations have demonstrated especially famous, including Choice Tree, K-Closest Neighbor, Backing Vector Machines, Credulous Bayes, Irregular Woodland, Helped Trees, Versatile Supporting and Slope Supporting (Helal et al, 2019). As well as accomplishing elite execution in prescient demonstrating undertakings, the fame of specific calculations may likewise be connected with their capacity to produce reasonable result. This implies that these calculations can give clear and natural clarifications of their expectations, which can be helpful for deciphering the consequences of the model. Logical models are critical in taking care of issues in training, including foreseeing understudy execution (Alamri et al, 2021). Reasonable models, for example, tree-based calculations (53%) and rule-based calculations (33%) are more regularly utilized than profound learning calculations (6%) (Alamri et al, 2021). Profound learning models are examples of "black-box" models, implying that the manner in which models work can't be made sense of or figured out by human. In any case, late examinations have uncovered a change in center from logical models or "white-box" models to more complicated "black-box" models that are fit for tackling testing issues (Minar et al, 2018). Understudy's exhibition can be primarily arranged into: Graphic examination,

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023

Prescriptive Investigation, symptomatic Examination, and prescient Investigation. While expressive examination is the sort of examination of information that depicts, show or sum up data of interest in a valuable manner with the end goal that examples (Melody et al., 2013). Prescriptive Examination is an interaction that investigates information and gives moment suggestions on the most proficient method to upgrade strategic policies to suit different anticipated (Sanjay and Alamma, 2016).

**2.2 Review of Related Literature**

In this section, we examines some of the previous approaches used by researchers for academic engagement and student's performance. Below, we give a brief review of research studies that have been conducted using different approaches.

As per (Iqbal et al, 2017), applied machine learning techniques to predict students' grades in various courses for the dataset of the Electrical Designing Division at Data Innovation College (ITU) in Lahore, Pakistan. Their review demonstrated that the Limited Boltzmann Machine (RBM) method is reasonable for displaying plain information and showed improved results than different strategies utilized in foreseeing the understudies' presentation in a specific course. The examination in Zohair, (2019) showed that SVM is best performs for basic information in foreseeing an understudy's grade. The productivity of SVM in preparing the little dataset size in delivering higher characterization precision for anticipating understudies' exhibition likewise has been upheld in (Anderson and Anderson, 2017)..

As per (Jens, 2013), the effect of students' engagement on their academic achievement through the course of their school professions was researched. This was accomplished by estimating the impacts of individual parts of commitment on the accomplishment of 1281 understudies from the Dutch locale Twente. Contrasts in these impacts for understudies of various stages in their profession were estimated utilizing a directed relapse approach. For each part of the tried commitment model, an impact on accomplishment could be found, yet the sort of impact contrasted relying upon the phase of the understudies' school profession and whether consequences for science or language accomplishments were estimated. The outcomes demonstrate that to comprehend the connection between understudy commitment and accomplishment, one needs to consider the various parts of commitment with regards to the understudies' stage in their school vocation.

In (Armando, 2019), determined the extent of student engagement at Partido State College and investigated the elements influencing understudy commitment. The review utilized engaging correlational technique. An instructor made survey was utilized to accumulate information. The general weighted normal for two semesters were utilized to decide the scholarly presentation of the respondents. Centered bunch conversation was utilized to approve the information got from the surveys. A sum of 300 and five understudies from the School of Training participated in the review. Mean and positioning, Pearson second relationship, and various relapse were utilized to treat the information. The review uncovered that the degree of understudy commitment along social, profound and mental commitment were high with a mean of 2.84. It was figured out that scholastic execution of the respondents was excellent (GWA=1.83). The correlational examination tracked down that educator ($r=.125$, $p=.029$), school ($r=.143$, $p=.013$), and family factors ($r=.106$, $p=.028$) were possitively connected with understudy commitment, while the Numerous Straight Relapse investigation uncovered that there was somewhat low level of change (1.8%) yet shows that the elements were critical indicators of understudy commitment $F(3, 301)=2.905$. Besides, it was figured out that conduct, close to home and mental commitment were decidedly corresponded to the scholarly presentation of the understudies. The instructor, the school, and the guardians ought to have solid coordinated effort to give more open doors to understudies to expand their university engagement.

As per (Courtney, 2020), used the multidimensional construct of student engagement to predict students' academic achievement. Student engagement was investigated by investigating factors connected with mental, conduct, and close to home commitment. It was anticipated that factors connected with close to home commitment (i.e., social help and test tension), conduct commitment (i.e., concentrate on ways of behaving and hesitation), and mental commitment (i.e., objective direction, coarseness, locus of control, and metacognition) would have a fundamentally connection to generally school GPA. A hieratical various relapse was utilized to investigate the connection between understudy commitment and in general school GPA. The general model that contained proportions of profound, social and mental commitment was huge and made sense of 57% of the difference in understudies' general school GPAs. As anticipated, earlier accomplishment as estimated by understudy perusing ACT scores were a critical, special indicator of generally school GPA. This connection stayed huge in each step of the model. Factors connected with profound commitment (i.e., social help), and conduct commitment (i.e., concentrate on ways of behaving) were found to have a huge connection to by and large school GPA. Not at all like what was conjectured, factors connected with mental commitment (i.e., objective direction, coarseness, locus of control, and metacognition) were not found to anticipate generally school GPA remarkably.

Brendan,et al.(2022), examine students reading behaviorutilizing a computerized course reading framework while taking an open-book test according to the point of view of commitment and execution to recognize the techniques that are utilized. We make models to foresee the presentation and commitment of students before the beginning of the evaluation and concentrate perusing conduct qualities utilized when the beginning of the appraisal in an advanced education setting. It was tracked down that techniques, for

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023

example, changing and reviewing are marks of how a student will act in an open digital book evaluation. Low performing understudies exploit the open digital book strategy of the evaluation and utilize a methodology of looking for data during the evaluation. Likewise contrasted with execution, the expectation of in general commitment has a higher precision, and hence could be more fitting for recognizing mediation up-and-comers as an early-advance notice intercession framework.

## III. Machine Learning Algorithm

Machine learning is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, much like humans do. With many companies collecting corporate and customer data, machine learning is the key to processing and using the massive quantity of data to identify profitable opportunities or risks. (Xin, 2018). Besides data mining, machine learning is used to compute complex tasks that humans could not feasibly perform, including facial recognition, network surveillance, automation, online search and healthcare. With the infinite amount to cases for machine learning, there are a multitude of algorithms that are used by programmers. Each algorithm differs in their approach and the type of problem that they are built to solve.

### 3.1 General Approach for Machine Learning

The general methodologies for machine learning are classified into four: supervised learning algorithms, semi-supervised learning algorithms, unsupervised learning algorithms, reinforcement learning (Guo, 2016).

1. Supervised Learning: algorithms that have input variables(x) and an output variable(y) and the algorithm learns the mapping function from the input to the output.
2. Unsupervised learning: algorithms must learn relationships between elements in a data-set and classify the raw data without "help".
3. Reinforcement Learning: concerned with how software agent sought to take actions in an environment to maximize some notion of cumulative reward.

### The Supervised Learning Approach

Supervised Learning is a machine learning paradigm for acquiring the input-output relationship information of a system based on a given set of paired input-output training samples. As the output is regarded as the label of the input data or the supervision, an input-output training sample is also called labelled training data, or supervised data. Occasionally, it is also referred to as learning with a Teacher, Learning from Labelled Data, or Inductive Machine Learning (Kotsiantis, 2017). The goal of supervised learning is to build an artificial system that can learn the mapping between the input and the output, and can predict the output of the system given new inputs. If the output takes a finite set of discrete values that indicate the class labels of the input, the learned mapping leads to the classification of the input data. If the output takes continuous values, it leads to a regression of the input. The input-output relationship information is frequently represented with learning-model parameters. When these parameters are not directly available from training samples, a learning system needs to go through an estimation process to obtain these parameters. Different form Unsupervised Learning, the training data for Supervised Learning need supervised or labelled information, while the training data for unsupervised learning are unsupervised as they are not labelled (i.e., merely the inputs). If an algorithm uses both supervised and unsupervised training data, it is called a Semi-Supervised Learning algorithm. If an algorithm actively queries a user/teacher for labels in the training process, the iterative supervised learning is called Active Learning. According to Taiwo, O, (2010) the supervised machine learning algorithms which deals more with classification includes the following: Linear Classifiers, Logistic Regression, Naïve Bayes Classifier, Perceptron, Support Vector Machine; Quadratic Classifiers, K-Means Clustering, Boosting, Decision Tree, Random Forest (RF); Neural networks, Bayesian Networks and so on.

1. **Linear Classifiers:** Linear models for classification separate input vectors into classes using linear (hyperplane) decision boundaries Louridas, P (2016). The goal of classification in linear classifiers in machine learning, is to group items that have similar feature values, into groups. Timothy Jason (2008), stated that a linear classifier achieves this goal by making a classification decision based on the value of the linear combination of the features. A linear classifier is often used in situations where the speed of classification is an issue, since it is rated the fastest classifier Taiwo, O, (2010). Also, linear classifiers often work very well when the number of dimensions is large, as in document classification, where each element is typically the number of counts of a word in a document. The rate of convergence among data set variables however depends on the margin. Roughly speaking, the margin quantifies how linearly separable a dataset is, and hence how easy it is to solve a given classification problem Setiono R. (2000).
2. **Logistic regression:** This is a classification function that uses class for building and uses a single multinomial logistic regression model with a single estimator. Logistic regression usually states where the boundary between the classes exists, also states the class probabilities depend on distance from the boundary, in a specific approach. This moves towards the

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023

extremes (0 and 1) more rapidly when data set is larger. These statements about probabilities which make logistic regression more than just a classifier. It makes stronger, more detailed predictions, and can be fit in a different way; but those strong predictions could be wrong. Logistic regression is an approach to prediction, like Ordinary Least Squares (OLS) regression. However, with logistic regression, prediction results in a dichotomous outcome Newsom, I. (2015). Logistic regression is one of the most commonly used tools for applied statistics and discrete data analysis. Logistic regression is linear interpolation Shai Shalev (2014).

3. **Naive Bayesian (NB) Networks:** These are very simple Bayesian networks which are composed of directed acyclic graphs with only one parent (representing the unobserved node) and several children (corresponding to observed nodes) with a strong assumption of independence among child nodes in the context of their parent Nilsson, N.J. (2005).Thus, the independence model (Naive Bayes) is based on estimating. Bayes classifiers are usually less accurate that other more sophisticated learning algorithms (such as ANNs). However, performed a large-scale comparison of the naive Bayes classifier with state-of-the-art algorithms for decision tree induction, instance-based learning, and rule induction on standard benchmark datasets, and found it to be sometimes superior to the other learning schemes, even on datasets with substantial feature dependencies. Bayes classifier has attribute- independence problem which was addressed with Averaged One-Dependence Estimators Neocleous C. (2002).
4. **Multi-layer Perceptron:** This is a classifier in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non- convex, unconstrained minimization problem as in standard neural network training. Other well- known algorithms are based on the notion of perceptron Tapas Kanungo, D. M. (2002). Perceptron algorithm is used for learning from a batch of training instances by running the algorithm repeatedly through the training set until it finds a prediction vector which is correct on all of the training set. This prediction rule is then used for predicting the labels on the test set Neocleous C. (2002).
5. **Support Vector Machines (SVMs):** These are the most recent supervised machine learning technique. Support Vector Machine (SVM) models are closely related to classical multilayer perceptron neural networks. SVMs revolve around the notion of a margin‖—either side of a hyperplane that separates two data classes. Maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalization error.
6. **K-means:** According to Nilsson, N.J. (2005), K- means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. K-Means algorithm is employed when labeled data is not available. General method of converting rough rules of thumb into highly accurate prediction rule. Given ―weak learning algorithm that can consistently find classifiers (―rules of thumb‖) at least slightly better than random, say, accuracy _ 55%, with sufficient data, a boosting algorithm can provably construct single classifier with very high accuracy, say, 99%.
7. **Decision Trees:** Decision Trees (DT) are trees that classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. Decision tree learning, used in data mining and machine learning, uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. More descriptive names for such tree models are classification trees or regression trees. Decision tree classifiers usually employ post-pruning techniques that evaluate the performance of decision trees, as they are pruned by using a validation set. Any node can be removed and assigned the most common class of the training instances that are sorted to it.
8. **Neural Networks:** Neural Networks (NN) that can actually perform a number of regression and/or classification tasks at once, although commonly each network performs only one. In the vast majority of cases, therefore, the network will have a single output variable, although in the case of many-state classification problems, this may correspond to a number of output units (the post-processing stage takes care of the mapping from output units to output variables. Artificial Neural Network (ANN) depends upon three fundamental aspects, input and activation functions of the unit, network architecture and the weight of each input connection. Given that the first two aspects are fixed; the behavior of the ANN is defined by the current values of the weights. The weights of the net to be trained are initially set to random values, and then instances of the training set are repeatedly exposed to the net. The values for the input of an instance are placed on the input units and the output of the net is compared with the desired output for this instance. Then, all the weights in the net are adjusted slightly in the direction that would bring the output values of the net closer to the values for the desired output. There are several algorithms with which a network can be trained Lemnaru C. (2012).

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023

### 3.2 Data Collection

In this study, real data of students (high school, college, associate degree, bachelors and master's degree) were collected from schools in Nigeria, January 2022 to March 2023. The dataset contains data of 1000 students (Male 503 and Female 503). The information on the attributes in the dataset is shown below.

Table 1. Attributes used in the dataset

| Attribute Name | Abbreviation | Description | Type | Possible Values |
|---|---|---|---|---|
| Class | Class | Class assign for the CSA course | Nominal | DDT1A, DDT1B, DDT1C, DDT1D |
| Year | Year | Year of student intake | Numeric | [2022 - 2023] |
| Gender | Gender | Students gender | Nominal | Female, Male |
| Continuous Assessment Mark | CAM | Marks obtained from quiz in class | Numeric | [47 – 96 |
| Test preparation Course | TPC | Marks obtained from problem based task, and test in class | Numeric | |
| Final Examination Mark | FEM | Marks obtained by students in the final examination | Numeric | [19 - 90] |
| Total Mark | TM | Marks obtained by the percentage of CAM and FE based on the course Curriculum | Numeric | [38 - 91] |
| Final Grade Points | FGP | Grade points obtained from student CGPA | Numeric | [0.67 – 4.00] |
| Group | Group | Group of grade to categorize the student performance | Nominal | 'Excellent', 'Good', 'PASS' 'Fail' (E, F). |
| Final Grade | FG | Student grade achievement based on TM students | Nominal | (A), (B1, B2), (C, D), (E, F). |

### 3.3 Data Pre-processing

We apply to pre-process the collected data to prepare it for the selection of machine learning algorithms such Linear Regression, decision tree, naïve Bayes, KNN, Kmeans. We tested the four attributes given and discovered so far that the attributes that has the impact on students evaluation are their tribe and parents education. We removed some irrelevant attributes based on feature selection in WEKA (Waikato Environment for Knowledge Analysis version 3.8.3) to develop machine learning models. We select feature selection using the WrapperSubsetEval technique that uses the BestFirst search method to pick the relevant attributes in the dataset.

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023

Based on the results, we removed all data related to class, group, gender, education, lunch and test preparation course except for the courses (Maths, Writing and Reading), Group and performance were selected in the final dataset. To monitor the students pass or fail in a course, we grouped the students' performance into five classes including 'Excellent' (A), 'Good' (B1, B2), 'PASS' (C, D), 'Fail' (E, F).

### 3.4 Design Model and Experimental Process

In this study, three classification and regression models were implemented which includes KNN, Naïve Bayes (NB), and decision tree (DT) to predict student FG performance. We used WEKA to conduct the experiment by using ten folds cross-validation whereby our dataset is partitioned into a training (90%) and testing (10%) set for evaluation. The three different predictive models were compared. The accuracy results are presented in detail in the following section.

### 3.5 Respondents' General Profiles

There are four (4) questions in this part which are; (i) gender, (ii) Group, (iii) education level (iv) test prepared course (v) Lunch. Information from this part aimed to know the respondents' background who were involved in the test. In table 2 below the frequency distribution and the percentage of respondents are involved. Based on Table 2, this research's respondents were made up of 1000 respondents of which 49.74% were female and 50.24% were male. In terms of group, 8.09% were from group A. 16.08% were group B, 16.98% were group C, 20.47% were group D, and 39.36% were group D. While in education Level, 180 respondent were high school, 272 respondent were college, 176 respondent were associate degree, 88 respondent were bachelor, and 285 respondent were master's degree. In the Test prepared course, 65.93% did not complete the course, and 34.06% completed the course. In terms of lunch, 42.25% were free/reduced and 57.74% were standard.

Table 2: Respondents' Profiles

| S/N | Respondents' background | Frequency | Percentage (%) |
|---|---|---|---|
| 1. | Gender | | |
| | Male | 503 | 50.24 |
| | Female | 497 | 49.74 |
| 2. | Group | | |
| | Group A | 81 | 8.09 |
| | Group B | 161 | 16.08 |
| | Group C | 170 | 16.98 |
| | Group D | 205 | 20.47 |
| | Group E | 394 | 39.36 |
| 3. | Education level | | |
| | High school | 180 | 17.98 |
| | College | 272 | 27.17 |
| | Associate degree | 176 | 17.58 |
| | Bachelor's degree | 88 | 8.79 |
| | Master's degree | 285 | 28.47 |
| 4. | Test prepared course | | |
| | None | 660 | 65.93 |
| | Completed | 341 | 34.06 |
| 5. | Lunch | | |
| | Free/reduced | 423 | 42.25 |
| | Standard | 578 | 57.74 |

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023

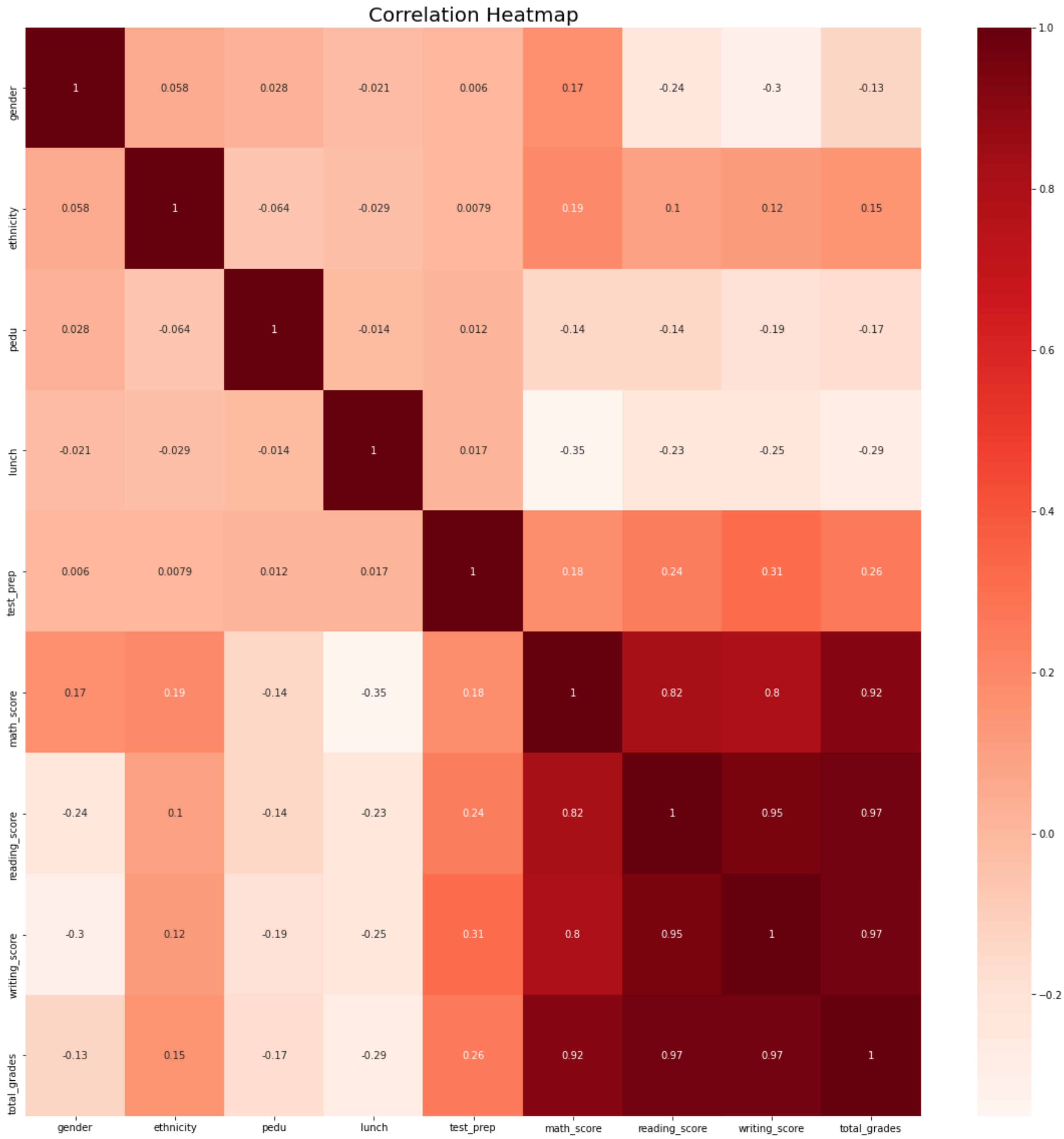


Fig 3: Student engagement data and academic performance data

Table 3: Accuracy of different evaluation metrics used

| ML Algorithms | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| SVM | 95 | 96 | 94 | 95 |
| Decision Tree | 92 | 92 | 92 | 92 |
| KNN | 93 | 82 | 72 | 93 |
| Logistic Regression | 78 | 93 | 93 | 77 |

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023

## IV. Result and Discussion.

In this study, we conducted four experiments using SVM, decision tree, KNN, Logistic Regression to predict the level of student engagement and academic performance. In table 6, it shows the different evaluation metrics used for the performance of all algorithm. The result of our prediction shows, 95% accuracy level for SVM, 96% precision level, 94% recall and 95% F1-score level. For decision tree, it shows 92% accuracy, 92 precision level, 92% recall and 92% F1-score level. For KNN, 93% accuracy, 82% precision, 72% recall and 93% F1-score level. For Logistic Regression, it shows that 78% accuracy, 93% precision level, 93% recall and 77% F1-score. Fig 9, shows that SVM has a high-performance level in predicting student engagement and academic performance. The graph above (fig 3), shows the overall course score/grade of students. In math, 26 students score from 0-39, 79 students score 40-49, 274 students score 50-59, 73 students score 60-64, 54 students score 65-69 and 445 students score 7- -100. In reading, 35 students score 0-39, 62 students score 40 – 49, 154 students score 50-59, 96 students score 60-64, 160 students score 65-69 and 520 students score 70-100. In writing, 26 students score 30-39, 71 students score 40-49, 59 students score 50-59, 89 students score 60-64, 231 students score 65-69 and 490 students score 70-100.

## V. Conclusion

This study focused on developing a predictive models to identify students at risk of disengaging or underperforming and also provide higher education institutions with insights and recommendations to improve student engagement and academic performance. It uses machine learning techniques such as KNN, decision tree, Logistic Regression and SVM in predicting the effect of student's engagement and academic performance. The model can also be used as an early warning system to identify failure students in the classroom by the course coordinators and educators, to take strategic decisions to improve student performance. Thus, the early prediction of student performance can trigger educators to track student dropouts in a particular course at an early stage. After testing the four attributes given, we discovered so far that the attributes that has impact on student evaluation are their ethnicity and parents education level.

## Reference


1. Anderson, T. and Anderson, R. (2017). 'Applications of Machine Learning To Student Grade Prediction in Quantitative Business Courses', Global Journal of Business Pedagogy, 1(3), pp. 13–22.
2. Alamri, R.; Alharbi, B. (**2021)** 'Explainable student performance prediction models: A systematic review. IEEE Access, 9, 33132–33143. [CrossRef]
3. Akçapınar, G.; Altun, A.; A¸skar, P. (**2019**), 'Using learning analytics to develop early-warning system for at-risk students. Int. J. Educ. Technol. High. Educ. 16, 40. [CrossRef].
4. Akram, A.; Fu, C.; Li, Y.; Javed, M.Y.; Lin, R.; Jiang, Y.; Tang, Y. (**2019). '**Predicting students' academic procrastination in blended learning course using homework submission data. IEEE Access **2019**, 7, 102487–102498. [CrossRef]
5. Behr, A.; Giese, M.; Theune, K. (**2020). '**Early prediction of university dropouts - A random forest approach. Jahrbücher Für Natl. Und Stat. **2020**, 240, 743–789. [CrossRef]
6. Brendan Flanagan* , Rwitajit Majumdar and Hiroaki Ogata (2022). Early-warning prediction of student performance and engagement in open book assessment by reading behavior analysis. Int J Educ Technol High Educ.
7. Burgos, C.; Campanario, M.L.; de la Peña, D.; Lara, J.A.; Lizcano, D.; Martínez, M.A. (2018). 'Data mining for modeling students'performance: A tutoring action plan to prevent academic dropout. Comput. Electr. Eng. **2018**, 66, 541–556. [CrossRef]
8. Christenson S. L., Wylie C., & Reschly A. L. (2012). Handbook of research on student engagement. Springer Science and Business Media. https://doi.org/10.1007/978-1-4614-2018-7
9. Courtney Brallier (2020). The effects of student engagement on academic achievement among college students. Middle Tennessee State University.
10. Dessart L., Veloutsou C., & Morgan-Thomas A. (2015). Consumer engagement in online brand communities: A social media perspective. Journal of Product and Brand Management, 24(1), 28–42. https://doi.org/10.1108/JPBM-06-2014-0635/FULL/XML
11. Fredricks JA, Blumenfeld PC, Paris AH.(2004). 'School engagement: Potential of the concept, state of the evidence. Review of Educational Research. 2004;74:59–109. [Google Scholar]
12. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. Neurocomputing, 187, 27–48. http://dx.doi.org/ 10.1016/j.neucom.2015.09.116.
13. Henrie C. R., Halverson L. R., & Graham C. R. (2015). 'Measuring student engagement in technology-mediated learning: A review. Computers & Education, 90, 36–https://doi.org/10.1016/J.COMPEDU.2015.09.005
14. Helal, S.; Li, J.; Liu, L.; Ebrahimie, E.; Dawson, S.; Murray, D.J. (2019). 'Identifying key factors of student academic performance by subgroup discovery. Int. J. Data Sci. Anal. **2019**, 7, 227–245. [CrossRef]

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XII, Issue V, May 2023


15. Iqbal, Z., Qadir, J., Mian, A.N. and Kamiran, F. (2017) 'Machine Learning Based Student Grade Prediction: A Case Study', pp. 1–22.
16. Jenz Abbing (2013), the effect of student's engagement on academic achievement in different stages of their academic career from a drop out perspective. University of twente.
17. Kotsiantis, S. B. (2017). Supervised Machine Learning: A Review of Classification Techniques. Informatica 31 (2017). Pp. 249 – 268. Retrieved from IJSwebsite:2.4.3/index.php/informatica/article/download/148/140.
18. Lemay, D.J.; Doleck, T. (2020). 'Predicting completion of massive open online course (MOOC) assignments from video viewing behavior. Interact. Learn. Environ. **2020**, 1782–1793. [CrossRef]
19. Lemnaru C. (2012). Strategies for dealing with Real World Classification Problems, (Unpublished PhD thesis) Faculty of Computer Science and Automation, Universitatea Technica, Din Cluj-Napoca. Available at website: http://users.utcluj.ro/~cameliav/documents/TezaFinalLemna ru.pdf
20. Louridas, P., & Ebert, C. (2016). Machine Learning. IEEE Software, 33(5), 110-115. doi:10.1109/ms.2016.11.
21. Mangaroska, K.; Giannakos, M. (2018). 'Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. IEEE Trans. Learn. Technol. **2018**, 12, 516–534. [CrossRef]
22. Marbouti, F.; Diefes-Dux, H.A.; Strobel, J. (2015). 'Building course-specific regression-based models to identify at-risk students. In Proceedings of the 2015 ASEE Annual Conference & Exposition, Seattle, WA, USA, 14–17 June 2015; pp. 26–304.
23. Martin, J., & Torres, A. (n.d.). (2016). 'User's Guide and Toolkit for the Surveys of Student Engagement: The High School Survey of Student Engagement and the Middle Grade School Survey of Student Engagement. National Association of Independent Schools. Retrieved from https://www.nais.org/Articles/Documents/Member/2016 HSSSE-report-full-FINAL.pdf.
24. Neocleous C. & Schizas C. (2002). Artificial Neural Network Learning: A Comparative Review. In: Vlahavas I.P., Spyropoulos C.D. (eds)Methods and Applications ofArtificial Intelligence.
25. Nilsson, N.J. (2005). Learning machines. New York: McGraw-Hill.Published in: Journal of IEEE Transactions on Information Theory Volume 12 Issue 3, 1966. doi: 10.1109/TIT.1966.1053912 pp. 407 – 407. Available at ACM digital library website: http://dl.acm.org/citation.cfm?id=2267404
26. Oliva-Cordova, L.M.; Garcia-Cabot, A.; Amado-Salvatierra, H.R. (**2021).** 'Learning analytics to support teaching skills: A systematic literature review. IEEE Access **2021**, 9, 58351–58363. [CrossRef]
27. Setiono R. and Loew, W. K. (2000), FERNN: An algorithm for fast extraction of rules from neural networks, Applied Intelligence.
28. Shai Shalev-Shwartz and Shai Ben-David (2014). Understanding Machine Learning From Theory to Algorithms.
29. Taiwo, O. A. (2010). Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307-034-6, InTech, University of Portsmouth United Kingdom. Pp 3 – 31. Available at InTech open http://www.intechopen.com/books/new-advances-in- machine-learning/types-of-machine-learning-algorithms.
30. Timothy Jason Shepard, P. J. (2008). Decision Fusion Using a Multi-Linear Classifier. In Proceedings of the International Conference on Multisource-Multisensor Information Fusion.
31. Waheed, H.; Hassan, S.U.; Aljohani, N.R.; Hardman, J.; Alelyani, S.; Nawaz, (2020). 'R. Predicting academic performance of students from VLE big data using deep learning models. Comput. Hum. Behav. **2020**, 104, 106189. [CrossRef].
32. Yang, Y.; Hooshyar, D.; Pedaste, M.;Wang, M.; Huang, Y.M.; Lim, H. (2020). 'Prediction of students' procrastination behaviour through their submission behavioural pattern in online learning. J. Ambient. Intell. Humaniz. Comput. **2020**, 1–18. [CrossRef]
33. Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, et al., (2018). "Machine learning and deep learning methods for cybersecurity", IEEE Access, vol. 6, pp. 35365-35381.
34. Zhang, L., Li, K.F., (2018). Education analytics: Challenges and approaches, in: 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), p. 193–198. doi:10.1109/WAINA.2018.00086. www.kaggle.com/code/spscientist/student-performamce-in-exams/notebook