Scientific Journal of Computer Science
futuristech PT. Teknologi Futuristik Indonesia
Date of publication March 29, 2026, date of current version March 29, 2026.
Digital Object Identifier 10.64539/sjcs.v2i2.2026.451

Article

Predictive Analytics Model for Adaptive Teaching in Open and Distance Learning Institutions: Machine Learning Approach

Danladi Moses Adayilo[1,*], Ishaq Oyebisi Oyefolahan[1], Juliana Ngozi Ndunagu[2], Nwando Anekwe[3], Ebenezer Malcalm[4], Khanyisile Twabu[5]

[1] Africa Centre of Excellence on Technology Enhanced Learning, National Open University of Nigeria, Plot 91, Abuja 900108, Nigeria; e-mail: ace22160005@noun.edu.ng, danladimosesadayilo@gmail.com (D. M. Adayilo), ioyefolahan@noun.edu.ng (I. O. Oyefolahan).
[2] Department of Computer Science and Information Technology, Faculty of Computing, National Open University of Nigeria, Plot 91, Abuja 900108, Nigeria; e-mail: jndunagu@noun.edu.ng (J. N. Ndunagu).
[3] Directorate of Management Information System, National Open University of Nigeria, Plot 91, Abuja 900108, Nigeria; e-mail: nanekwe@noun.edu.ng (N. Anekwe).
[4] Management Studies, Ghana Communication Technology University, Accra, PO Box 100, Accra North, Ghana; e-mail: emalcalm@gctu.edu.gh (E. Malcalm).
[5] Department of Information Communication Technology, University of South Africa, Pretoria, 0002, South Africa; e-mail: mayekk@unisa.ac.za (K. Twabu).
* Correspondence

The authors received no financial support for the research, authorship, and/or publication of this article.

Abstract: The study investigates the application of predictive analytics model in adaptive teaching within Open and Distance Learning (ODL) institutions. The aim of the study lies in addressing the ongoing challenges of high dropout rates and low student engagement, particularly in developing countries. The research gap is the underutilisation of predictive analytics to personalise interventions and enhance learning outcomes in ODL environments. The study employs mixed-method research design including machine learning algorithms with Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and XGBoost, in predicting students at risk of academic failure and providing personalised interventions. A dataset of 5,000 students from the National Open University of Nigeria was used to trained and test the model. Model validation metrics used includes: accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC. More so, (n=1050) participants took part in the experimental and control group including semi-interview, enabling real world application of predictive model. Key findings indicated that Random Forest had the highest ROC-AUC (98.38%), followed by XGBoost (97.76%). Nevertheless, Logistic Regression and SVM outperformed the others in accuracy (97.43%), precision (97.65%), recall (95.95%), and F1-score (96.79%). These results show that adaptive teaching, supported by predictive analytics, is associated with improved student engagement and contributes to reducing dropout rates. The challenges such as data quality, privacy, trust and algorithms bias should be addressed. The study suggest that predictive analytics is capable of transforming teaching methods in ODL institutions, improve personalised and effective learning. Future study should focus on model optimisation and integration with other educational technologies.

Keywords: Predictive analytics; Adaptive teaching; Open and distance learning; Student engagement; Educational technology.

1. Introduction

The advancement of the digital learning environment rapidly has shifted the teaching and learning institutions with unmatched opportunities of data driven decision making in learning institutions. In the Open and Distance Learning (ODL) settings, the conventional one size fits all models of instruction have been demonstrated to be not sufficient to cater to the needs of diverse learners, which also leads to ongoing issues in terms of student engagement, academic achievements, and retention rates [1]-[7].

Predictive analytics used to make predictions based on historical and real time educational data and to forecast student performance, has also become a promising solution to these problems as it gives educators an opportunity to predict at risk students and adjust their instructional methods accordingly [3], [7]-[9].

Despite the predictive models and adaptive learning have become increasingly popular, the current literature in ODL has predominantly studied descriptive analytics or broadly speaking learning analytics models, with very few having studied model implementation, performance, and its application in pedagogical practise in a specific institutional context of an ODL settings [10]-[12]. Another research gap exists where predictive outcomes are directly connected to adaptive teaching strategies which dynamically respond to the individual learner profile especially in the specific context of ODL settings where face to face interactions remain restricted [4], [11], [13], [14]. Most of the empirical studies make use of quantitative research approach without the qualitative aspects. There is scanty empirical evidence on predictive analytics model in developing countries within ODL settings [2], [15]-[17].

The current research will fill these gaps by using a model of predictive analytics to enhance adaptive teaching in ODL institutions especially in terms of student engagement, retention, and performance. Through a synthesis of quantitative and qualitative research, the paper is also able to measure the effectiveness of predictive models in recognising at-risk students and discuss the perceptions and ethical issues of faculty and students regarding the adoption of predictive models. The findings will provide useful information to the ODL institutions particularly in developing nations where such interventions can greatly improve the learning outcomes.

The research main contributions lie in the fact that predictive analytics models are used to detect at-risk students of failure in their academic life early enough and provide them with tailored intervention depending on their learning profiles. The proposed work involves implementing machine learning models such as Logistic Regression, Support Vector Machines (SVM), Random Forest, and XGBoost to demonstrate how predictive analytics can be utilised to change the teaching approaches and make the students more successful in the online distance learning environment. Additionally, the study also highlights ethical considerations linked with the use of student information, and ensures that privacy and fairness are the key concerns in the predictive models. The study answered two research objectives including (1) to evaluate the effectiveness of predictive analytics-based models in predicting student engagement and academic performance in open and distance learning environments and (2) to investigate the faculty and students' perception in adoption of predictive analytics model in open and distance learning.

The theoretical framework of the research is based on a number of key theories which justify the introduction of adaptive teaching based on the use of predictive analytics in open and distance learning (ODL) settings [18]. Constructivism/Mastery Learning Theory focuses on learning as a constructive and active process in which learning takes place through the experiences of learners as they create knowledge. Predictive analytics, through customization of learning material and adapting to the needs of students, conforms to the concept of mastery learning, in which students will move forward when they have mastered every concept [1]. According to the Predictive Analytics Learning Theory, the use of historical and real-time data can be used to predict the intentions of students and adjust the teaching process to enhance learning results with the help of a machine learning model [2]. Technology Acceptance Model (TAM) and Unified Theory of Acceptance and Use of Technology (UTAUT) are interested in the factors that impact the adoption of new technologies, including predictive analytics in education, with particular attention to the predictive ease of use and usefulness in motivating educators to make adaptive learning systems [19]-[21]. Self-Regulated Learning explains the importance of students in enabling themselves to learn by setting goals, tracking, and regulating their learning plans through the help of predictive analytics which provide them with personalized recommendations [13], [17], [22]. According to the Socio-Technical System theory, both technology and human interaction should be combined in the most effective way to reach the best results, with the predictive systems and facilitators focusing on the collaboration to make educational experiences better [23]. Ethics and Governance theory talks about the moral component of the use of student data to predictive model and will ensure privacy, equity, and transparency of decision-making practises [9], [17], [24]-[28]. These frameworks collectively provide a comprehensive insight into how predictive analytics can be integrated into teaching practices and are associated with increased student engagement, performance, and retention in ODL settings, as illustrated in Figure 1.

Existing studies reviewed those educational predictive analytics as a subdiscipline of learning analytics that has quickly developed to utilise student information and enhance teaching, learning, and retention within online contexts. In its simplest form, predictive analytics employs statistical and machine learning techniques to the past and present educational data to predict the student performance, predict at risk students, and customise the interventions to promote the personalised learning process in open and distance learning (ODL) contexts [5], [29], [30]. The field of research falls under the category of more than descriptive analytics, which just summarises the behaviour of the student, instead of proactive forecasting,

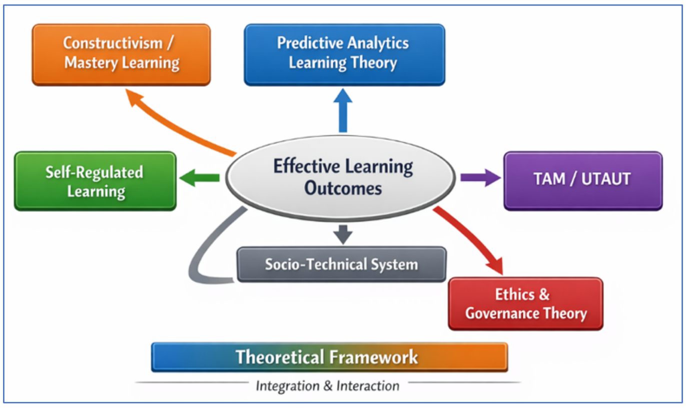

Figure 1. Theoretical Framework of the Study [2], [4], [17].

which allows the instructional strategies to be adjusted in real time [31], [32]. The existing empirical evidence shows that the predictive accuracy of identifying the at-risk learners and putting measures into change is over 85%, which is higher than its counterparts in other assessment procedures [7], [8], [33]. The AI conceptual researches inspired the adaptive systems to not just consider the predictive feature but also the pedagogical implementation required to ensure the transformation of teaching process thrives in any institutional setting [10], [11].

The structure of the paper is as follows: Section 2 describes the methodology that was applied in this study in detail, including data collection process, predictive model development and data validation methods. Section 3 reports on the results and discuss the findings with the focus on the effectiveness of the predictive models and their role in the student engagement and retention. Lastly, Section 4 wraps up the paper by summarising the main findings, contributions, limitations, and future research direction.

2. Methodology

2.1. Research Design

The research design is a mixed methods approach, which uses quantitative predictive modeling and qualitative measurement of adaptive teaching. The quantitative approach was utilised in order to understand real world problem while qualitative aspects was used to get responses from the participants regarding their perceptions and ethical concerns of the implementation of predictive analytics model [4], [5], [17]. The mixed-method approach is used to predict adaptive teaching interventions and while qualitative is used to evaluate them.

2.2. Experimental Design

The goal is to determine how predictive analytics is effective in enhancing student engagement, performance, and retention with the help of interventions in the form of individualized treatments. The interventions consist of:

- **Experimental Group:** The students in the experimental group were provided with individual, adaptive teaching intervention, powered by predictive analytics, including personalised content recommendations and facilitator intervention actions powered by at-risk predictions.
- **Control Group:** This is the baseline group as students were taught through the conventional mode of teaching and did not have any personalised interventions. The control group was adopted to compare the influence of the predictive model-driven adaptive teaching. They had been exposed to the usual modes of learning, and no special or anticipatory procedures.

The experimental and control groups were further categorised into courses and pre-tests and post-tests were taken to determine the effectiveness of interventions.

This methodology is integrated with seven components including data sources, data preprocessing, feature selection, predictive model development, model validation, adaptive teaching engine, ethics and governance. The study aims at designing and deploying a model of predictive model that will support adaptive teaching interventions in open distance learning institutions [8], [10], [34].

The components of the predictive model design is described below:

2.2.1. Data Sources

Here student information is uploaded in (csv format) or through the LMS, and Survey. The raw data is taken into the system as input to analysed the data. The institutional dataset of 5000 students' records were obtained from the National Open University of Nigeria, consisting of Learning Content Management System (LCMS) and the Directorate of Examination and Assessment (DEA). The dataset comprised of the student demographics and performance data, as well as of the behavioural characteristics that were retrieved in the Learning Content Management System (LCMS) and the Directorate of Examination and Assessment (DEA) of the National Open University of Nigeria. These variables includes students id, programme, gender, course code, Tutor marked Assignment, previous examination scores, grades, total scores, predicted outcomes, attendance , engagement (timestamps [comprising of the time and frequency by which students engage with online resources], clicks [monitoring the quantity of times students engaged with online resources], and forum activity [measuring how students participate in discussion forums]), played a key role in predicting student engagement, as well as identifying those students who are at-risk of their academic failure.

2.2.2. Data Preprocessing

This features cleanse and purifies the raw data. This stage entails that the processing of missing values, normalisation, and making the dataset analysis-suitable. Technologies and tools used in the study include: Data processing and modelling tools: Python 3 (Streamlit, Pandas, Scikit Learn, Numpy) [2], [3], [33], structured institutional student records of 5000 dataset, and exploratory data analysis Visualisation libraries (Statamodel, Matplotlib, Seaborn).

2.2.3. Feature Selection

Extraction of relevant features on the dataset. The step is a process that will identify the most significant features that can lead to the accuracy of the model and assist in the reduction of noise and the enhancement of the performance.

2.2.4. Model Development

This is the core of the system where machine learning models are developed and trained to give a response to a particular student that is at risk of academic failure. The machine learning models utilised in this study include (Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and XGBoost.

2.2.5. Model Training

The first step was to split the dataset into training and testing set with 80 / 20 split. The initial model training was done on this split with 80 percent of the data used to train the models and the other 20 percent serving to conduct a preliminary test of the model performance. In this way, the preliminary assessment of how well the model generalised to unseen data was possible.

The dataset trained and tested on the predictive models consisted of a 5,000-student sample, used to predict whether a student was at-risk depending on a variety of features (e.g., attendance, engagement, previous exam scores). The sample of 1,000 participants, however, was selected to undergo the pre-test and post-test analysis because of the logistics involved and the nature of the intervention. The sample was selected on the basis of certain criteria, i.e., willingness and availability to test and demographic representation [15], [34].

Thus, the 1,000 participants were selected from the bigger group of 5,000 students, and their involvement in the pre-test and post-test analysis was specifically intended to assess the effectiveness of the predictive model-based interventions. This helps to make sure that the connection between the two datasets is well understood and that there is no misunderstanding over the sampling approach.

2.2.6. 10-Fold Cross-Validation

Following the initial training and testing, 10-fold cross-validation was utilised in order to test the model performance in a stronger way. Cross-validation reduces the risk of overfitting or bias that may be caused by the random split and give a more valid estimate of the model. In 10-fold cross-validation the data is broken down into 10 subsets (folds). The model is trained on 9 of these folds and tested on the remaining fold and this is repeated 10 times with each fold serving as the test set once. The combination of both approaches helped the study to train the models with a large fraction of the data and also evaluate the models in a manner that reduced bias and overfitting. Primary training and testing were divided into two parts, the first 80/20 split, and the second cross-validation through 10 folds as a complete assessment of the overall performance of the model on various subsets of the data [4], [7].

2.2.7. Model Validation

This section presents the evaluation of the developed machine learning models for predicting student performance and at-risk behavior. Model performance was assessed using standard classification metrics, including accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC. These metrics provide complementary perspectives on model effectiveness, particularly in the presence of class imbalance [12], [35].

Accuracy:

Let y ∈ {0, 1} denote the true class label, where 1 represents an at-risk student and 0 represents a not-at-risk student, and let y^\hat{y}y^ denote the predicted class label. Accuracy is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Accuracy provides an overall measure of correct classification. However, in imbalanced datasets such as ODL contexts—where at-risk students form a minority class—accuracy alone may not be sufficient to evaluate model performance comprehensively.

Recall:

Recall (also known as sensitivity) measures the proportion of actual at-risk students correctly identified by the model:

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

High recall indicates that the model successfully identifies a large proportion of at-risk students, which is important in educational settings where missed interventions may negatively affect student outcomes.

Precision:

Precision measures the proportion of predicted at-risk students who are truly at risk:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

Figure 2: Proposed High Level System Architecture of Predictive Model.

High precision indicates that the model produces fewer false alarms, ensuring that interventions are directed toward students who genuinely require support.

F1-score:

To balance the trade-off between precision and recall, the F1-score is used as a harmonic mean of the two metrics:

$$F1 = 2 \text{ x} \frac{\text{Precision x Recall}}{\text{Precision + Recall}} \tag{4}$$

The F1-score is particularly useful in imbalanced classification problems, as it provides a single measure that captures both false positives and false negatives.

ROC-AUC:

The Receiver Operating Characteristic Area Under the Curve (ROC-AUC) evaluates the model's ability to discriminate between classes across different classification thresholds. It can be interpreted as:

$$\text{ROC} - \text{AUC} = \text{P}(\hat{y}_{\text{positive}} > \hat{y}_{\text{negative}}) \tag{5}$$

A higher ROC-AUC value indicates better separability between at-risk and not-at-risk students.

PR-AUC:

The Precision–Recall Area Under the Curve (PR-AUC) measures the area under the precision–recall curve and is particularly informative for imbalanced datasets:

$$\text{PR} - \text{AUC} = \int_0^1 \text{Precision}\ (R)\ dR \tag{6}$$

PR-AUC summarizes the trade-off between precision and recall across different thresholds and provides a more sensitive evaluation of performance on the minority (at-risk) class.

2.2.8. Adaptive Pedagogy Engine

The forecasting of individualised student takes place at this stage. This aspect renders the educational process more engaging, and identification of the at-risk students is conducted and interventions are comprised of tailored learning tracks, recommendations and feedbacks.

2.2.9. Ethics and Governance

The privacy and ethical concerns have been ensured by anonymising the identify of learners and adhering to the institutional data management policies regarding research data. Ethical approval was obtained from the University Research Ethics Committee and informed consent from the participants was also secured. Adherence to the ethical guidelines resolves the issues such as privacy of data, open access and unbiased use of predictive models and give ethical control over the pipelines. The study is in conformity with the Nigeria Data Protection Act (NDPA), 2023 and European Union General Protection Regulation (GDPR) 2018 [17], [24], [26], [27].

The Figure 2 shows the Proposed High Level System Architecture of Predictive Model, which is a combination of machine learning and adaptive pedagogy, that will

Table 1. Summary of demographic attributes of the participants.

Attribute	Classification	Frequency	Percentage (%)
Gender	Female	450	42.8
	Male	570	54.2
	Don't want to say	30	3
Age	18 – 29	650	61.9
	30 – 39	350	33.3
	40 – 49	30	2.9
	50 +	20	1.9
Role	Students	1000	95.4
	Faculty	15	1.4
	Administrators	15	1.4
	IT/Analyst Staff	10	0.9
	Instructional Designers	10	0.9
Qualification	B.Sc	10	20
	M.Sc	25	50
	PhD	15	30
Experience	3 years	10	20
	4 – 6 years	25	50
	7 years and above	15	30
Working Knowledge of Predictive Analytics	None / Basic	15	30
	Intermediate	25	50
	Advanced	10	20
Assigned Group	Experimental	500	50
	Control	500	50

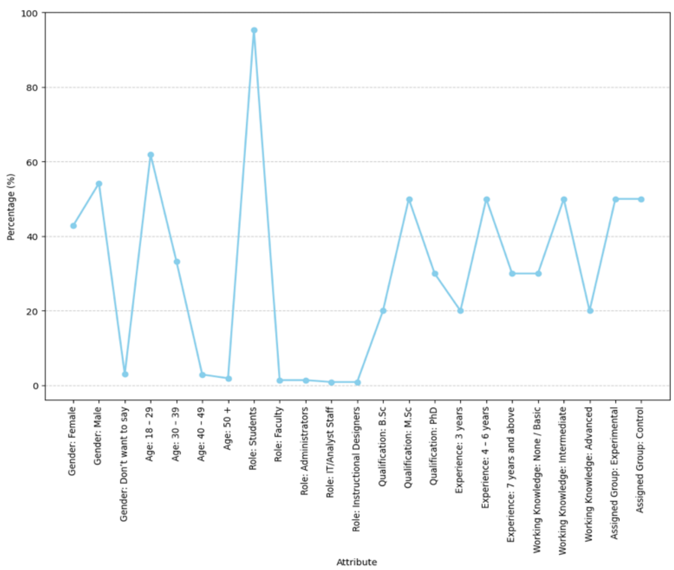

Figure 3. Summary of Demographic Attributes of the Participants.

ensure the efficient and ethical use of student data to bring about better educational outcomes.

Qualitative approach involving semi-structured interview of fifty (50) faculty members, and students on usefulness of adaptive recommendations. The mixed methods validation also guaranteed that the model does not only make predictions but is also valuable to the teaching practice in an actual ODL settings [5], [6].

3. Result and Discussion

3.1. Result

The predictive model was implemented using Core i5 processor system with a minimum of 16 GB RAM and 500 GB solid state drive/hard disk drive (SSD/HDD). Python 3 with the required libraries such as (Streamlit, Pandas, Numpy, Scikit-learn, XGBoost, LightGBM, Matplotlib, statsmodels among others. The result of findings is guided by the two research objectives including (1) to evaluate the effectiveness of predictive analytics-based models in predicting student engagement and academic performance in open and distance learning environments and (2) to investigate the faculty and students' perception in adoption of predictive analytics model in open and distance learning.

Table 2. Comparison of Performance Metrics of the Model.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	PR-AUC
Random Forest	96.85%	96.77%	95.38%	96.07%	98.38%	97.58%
XGBoost	96.78%	96.66%	95.66%	96.22%	97.76%	96.60%
Logistic Regression	97.43%	97.65%	95.95%	96.79%	97.50%	96.06%
Support Vector Machine (SVM)	97.43%	97.65%	95.95%	96.79%	97.12%	96.84%

Table 3. Impact of Predictive Analytics Model on Students Engagement and Retention (Experimental and Control Group).

Course Code	Group	Pre-Test Mean	Post-Test Mean	Gain	SD	t-value	df	p-value	Effect Size (Cohen's d)
BIO101	Experimental (72)	52	65	13	9.19	12.04	71	<0.001	1.41
	Control (72)	45	49	4	2.38	-	-	-	-
CIT104	Experimental (72)	52	68	16	11.31	12.04	71	<0.001	1.42
	Control (72)	46	48	2	1.41	-	-	-	-
CHM101	Experimental (72)	50	66	16	11.31	12.03	71	<0.001	1.42
	Control (72)	44	49	5	3.54	-	-	-	-
CHM192	Experimental (72)	52	70	18	12.73	12	71	<0.001	1.41
	Control (72)	45	50	5	3.54	-	-	-	-
GST101	Experimental (72)	54	68	14	9.9	11.97	71	<0.001	1.41
	Control (72)	46	49	3	2.12	-	-	-	-
MTH101	Experimental (70)	50	65	15	10.61	12.6		<0.001	1.51
	Control (70)	46	49	3	2.12	-	-	-	-
PHY101	Experimental (70)	52	68	16	11.32	11.85		<0.001	1.41
	Control (70)	45	49	4	2.83	-	-	-	-

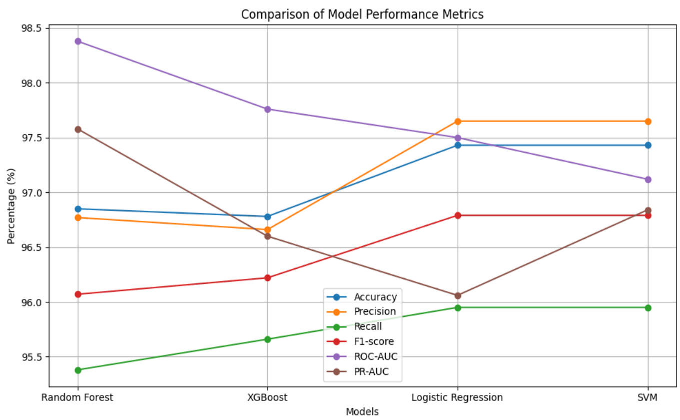

Figure 4. Line Chart Performance Metrics Comparison of the Model.

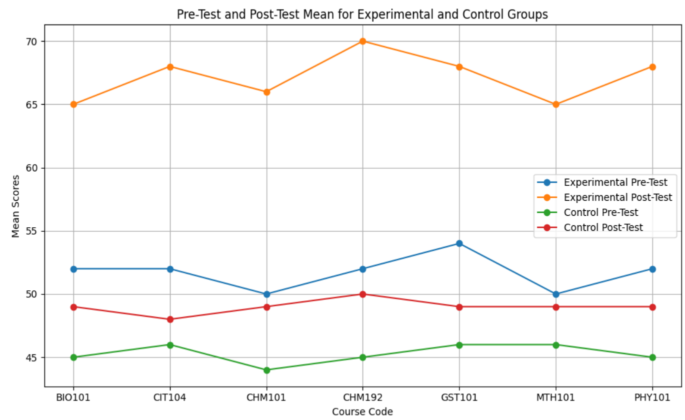

Figure 5. Line Chart of Pre-test and Post-test for Experimental and Control Groups.

The 5000 student dataset was obtained from the National Open University of Nigeria which was used to develop, trained and test the predictive model on 80/20 ratio data split. The data structure consists of student-id, gender, programme, course code, attendance, engagement TMA scores, previous examination scores, total score, grades, and predicted outcome.

The demographic attributes of (n=1050) participants in the study is presented in Table 1 and Figure 3. Table 1 and Figure 3, shown that the distribution of various attributes of the survey instrument that completed the study in the experimental and control group is mainly composed of majority males who constituted the majority of 54.2% with the other women constituting 42.8 and 3% not having disclosed their gender. Speaking of age, the biggest proportion of the participants is 18 to 29 years old, and they constitute 61.9% of the sample. The age group of 30-39 years spans the next biggest percentage (33.3%,), and the other age groups (40-49 years and 50+) constitute a figure of a much smaller percentage; that is, 2.9% and 1.9%.

In terms of roles, students constitute the huge majority of the sample with 95.4%, a very low percentage of faculty members (1.4%), administrators (1.4%), IT/Analyst staff (0.9%), and instructional designers (0.9%). Regarding educational degrees of the faculty members, administrators, IT/Analyst staff, and instructional designers, the majority of them have M.Sc. (50%), followed by Ph.D. (30%), and B.Sc. (20%).

Regarding terms of professional experience of the faculty members and others, half of the respondents (50) have a 4 to 6 years of experience, 30% above 7 years of experience, and 20% of 3 years of experience. In terms of their familiarity with predictive analytics, 50% of the faculty members and other respondents have intermediate level of knowledge, 30% of them have no or basic knowledge and 20% of them are advanced. Finally, the number of the participants (students) is equal between the experimental group and the control group, with each having 50 percent of the sample.

Table 3 and Figure 3 gives a clear overview of the demographic and professional demographic of the respondents with students as the majority and the faculty members, administrators, IT/Analyst staff, and instructional designers as minority having intermediate experience and knowledge in predictive analytics.

3.1.1. Quantitative Findings

Research Objective (RO) 1: The study forecast at-risk students and implement personalised interventions that helped to achieve better academic results using the combination of machine learning algorithms, including Logistic Regression, Support Vector machine (SVM), Random Forest, and XGBoost.

As presented in Table 2, Random Forest recorded the best performance in ROC-AUC (98.38%) and PR-AUC (97.58%) followed by XGBoost with 97.76% and 96.60%. However, Logistic Regression and SVM outperformed models in terms of accuracy (97.43%), precision (97.65%), recall (95.95%) and F1-score (96.79%) respectively.

Out of 5000 institutional datasets from the National Open University of Nigeria that was uploaded to the predictive model, 2005 students were flagged by the model as at-risk of academic failure representing 40.1% of the sample dataset, while 2,995 students were identified as pass students representing 59.9% of success rate. As presented in Table 2 and Figure 4, to determine the performance of the four predictive models; XGBoost, Logistic Regression, Random Forest, and Support Vector Machine (SVM), several measures were used to give the full picture of the effectiveness of the predictive model, which were accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC in predicting at-risk students and recommended intervention.

Logistic Regression and Support Vector Machine are at par with both having an accuracy of 97.43%, precision of 97.65%, recall of 95.95% and F1-score rate of 96.79%. While XGBoost is marginally higher than Random Forest in Precision (96.78% vs. 96.77) and Recall (95.66% vs. 95.38) but Random Forest is higher in ROC-AUC (98.38% vs. 97.76). The scores of Logistic Regression and Support Vector machine are similar on most metrics, which indicates that both approaches are equally useful in this study. The most balanced model appears to be Logistic Regression and Support Vector Machine which has high accuracy, precision, recall and F1-scores.

XGBoost and Random Forest are almost similar, except that the latter appears to be slightly better in terms of ROC-AUC and PR-AUC.

The predictive performance of all four models is very high, and the accuracy values are more than 96% which means that the models are very reliable to classify student learning outcomes. There is also similarity in the values of recall in the models (Recall of 96%) implies that the same model can be effective in identifying a student who is indeed at risk, which is especially important in the early warning system when it is expensive to miss the vulnerable learners. The Random Forest recorded the best performance with 98.38% of ROC-AUC and 97.57% of PR-AUC while Logistic Regression and SVM outperformed other model in terms of accuracy, precision, recall and F1-score.

The Table 3, revealed the seven (7) courses (BIO101, CIT104, CHM101, CHM192, GST101, MTH101, and PHY101) that were used to conduct the experiment. The actual participation in the experiment was 1000 participants. The two groups (Experimental and Control) were assigned 72 participants each for BIO101, CIT104, CHM101, CHM192, GST101 while, MTH101, and PHY101 had 70 participants each bring the total number to 1000 participants.

The students that participated in the study were 1000 in number and were assigned to experimental and control groups in equal number. Table 3 and Figure 5 shows the results of pre-test and post-test scores of the experimental and control groups in various courses, respective gains, standard deviation or SD, t-value, degrees of freedom (df), p-value, and the effect size (Cohen d).

In case of BIO101, the experimental group had a significant increase in scores, with 13 being the mean increase, a difference of 52 in the pretest to 65 in the post-test. The t-value was 9.19 and the p-value was below 0.001 which means a significant effect of a high extent. The magnitude of the effect (Cohen d) is 1.41 implying a big effect. The control group, on the contrary, gained less by 4, with a pre-test and post-test mean of 45 and 49, respectively. The control group has no significant difference because it is not the focus in the study.

In CIT104, score of the experimental group improved by 16 points, with a pre-test mean of 52 and a post-test mean of 68, t-value of 11.31 and p-value < 0.001. The d value of Cohen of 1.42 indicates that the effect size is large. The control group, however, recorded a smaller improvement of 2 where the pre-test and post-test were 46 and 48 respectively and the p-value was not found to be significant.

In the case of CHM101, the experimental group received 16 points, with mean 50ina range of 66 with a t-value of 11.31, p-value less than 0.0001, and a Cohen d of

Table 4. Qualitative Findings.

Theme	Faculty Members Responses (Mentioned in Frequency)	Students Responses (Mentioned in Frequency)	Theoretical Framework Alignment
Knowledge of Predictive Analytics (T1)	80% have a general knowledge; 20% do not know the predictive model and how it is used in instruction.	70% know of personalised interventions; only 30% understand the predictive analytics fully.	Constructivism/Mastery Learning: Predictive analytics makes the learning individual and assists students to learn at their speed [1].
Perceived Advantages of Adaptive Teaching (T2)	75% believe that personalised feedback and content suggestions help them to become more engaged and improve their student performance.	60% value personalised feedback and recommendations of the resources that can help them to better their student performance	Theory of Predictive Analytics Learning: Predictive analytics based adaptive teaching can improve learning through providing data-driven, personalised interventions [2], [21].
Obstacles in Implementation (T3)	60% reported that the time and effort required to apply adaptive teaching a practical activity might be too high, particularly in large classes.	25% believe that adaptive teaching is too general, and 20% would like to communicate more with the facilitators	TAM/UTAUT: Technology acceptance is a challenge because the user has problems in adoption and implementation when they find the system to be challenging to use or not beneficial [3], [21].
Facilitator Feedback (T4)	65% say that they have positive feedback regarding the adaptive teaching methods though state that they should balance the automated suggestions with human judgement.	30% indicate that the interventions were useful, but that they need to be more closely aligned with the needs of learners.	Self-Regulated Learning: When students have adaptive interventions which help them to control their learning process, then effective learning takes place [8].
Impact on Student Engagement (T5)	70% think adaptive teaching practises increase student engagement through the timely intervention and customised content.	55% indicate an increase in engagement through individual learning paths and specific interventions.	The socio-technical system entitles the positive understanding of technology and human judgment would enhance the performance of a system and the interest shown by a user in the learning systems [36], [37].
Impact on Academic Performance (T7)	65% report on the improvement of student performance as a result of the early intervention, which is founded on predictive outcomes.	45% of them affirm that the adaptive recommendations allowed them to improve grades by providing them with extra materials on challenging subjects.	Mastery Learning: Mastery through predictive interventions, teaches content progressively, enabling students to perform better [11].
Data Quality Concerns (T8)	40% have questions about the credibility of engagement data that is being used in predictive models as its effectiveness is not completely clear to them	30% are uncertain about the effectiveness of some recommendations because they do not fully understand the process of using the recommended resources.	Ethics and Governance Theory: When predictive models are endeavoured to succeed, securing data privacy and ethical application of student data is vital in ensuring the success and trustworthiness of the model [13], [28].
Support and Satisfaction (T9)	80% report feeling that predictive models are useful, but feel some scaling and data accuracy concerns	80% reported that they were satisfied with the adaptive teaching of predictive models especially in personalised feedback.	Socio-Technical System: Finding the balance between technology and the interaction with the user is a key element to the successful learning and the satisfaction of the user in adaptive teaching [4], [10].

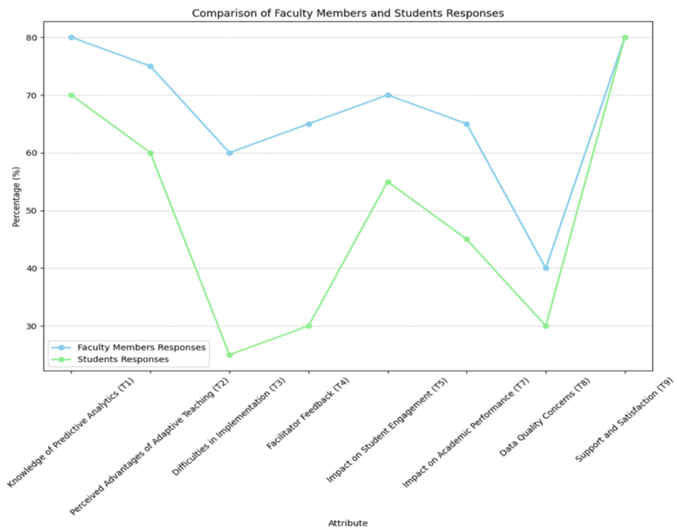

Figure 6. Comparison on Perceptions and Ethical Issues of Faculty Members and Students Responses.

1.42, which can be said to have a large effect size. The difference of 5 points between the control and the 44 and 49 was not significant.

In CHM192, the experimental group had a score improvement of 18, 52-70, with t-value 12.73 and p-value of under 0.001. Cohen d was also large with a value of 1.41. The control group increased by 5 points and the difference was not significant. In the case of GST101, its experimental group raised the duration of the test by 14 points with the post-test mean of 68, t=9.9, p= < 0.001 which indicates a substantial increase. The Cohen d value of 1.41 shows that the effect is large. There was a small increase of 3 points in the control group, but it is not significant. In MTH101, the experimental group had a 15-point improvement, and a mean of the change of 50 to 65. The t-test value of 10.61 and p-value of 0.001 are very significant improvement with a Cohen d of 1.51 which is a very large effect. The control group scored 3 points only.

Finally, in the case of PHY101, the experimental group improved by 16 points starting with a pre-test mean of 52, and a post-test mean of 68, t-value of 11.32, and p-value of below 0.001. The Cohen d of 1.41 indicates a huge effect size. The control group however exhibited lesser gain of 4 points. Overall, the post-test scores of the experimental groups were significantly higher and the effect sizes were high in all courses, which proves that interventions were very effective. The control groups on the other hand registered very little changes, which suggest the need for the experimental interventions.

3.1.2 Qualitative Findings

Research Objective (RO) 2: The responses from semi-structured interview of the study by faculty members and students is presented in Table 4.

Table 4 and Figure 6, illustrates responses of the faculty members and students regarding different features of adaptive teaching and predictive analytics. The faculty members tend to be more knowledgeable about predictive analytics (80%) than students (70%). Both groups note the usefulness of customised feedback and content suggestions, although faculty members have a higher appreciation of these (75% vs. 60%). There are also higher problems with adaptive teaching implementation, which are reported by faculty members in large-size classes (60% vs. 25%).

With respect to feedback, 65% of faculty members are attracted to the usefulness of the adaptive approaches but emphasise the need to balance the use of automation and human judgement whereas only 30% of students consider the interventions helpful. Faculty members who engage in adaptive teaching (70% vs only 55% of students) on engagement believe adaptive teaching to be a factor that enhances student engagement. The two groups agree that early interventions enhance academic performance with 65% of the faculty members and 45% students affirming the same. Faculty members and students raise data quality concerns by 40% and 30-percent, respectively, and each one of them demonstrates some degree of uncertainty regarding the effectiveness of the recommendations.

Regarding satisfaction, both faculty members and students (80% each) claim their satisfaction with predictive models though faculty members express their concerns regarding scaling and accuracy of data. On balance, the faculty members are more inclined to a more informed and positive perception of adaptive teaching, and students accentuate the importance of the more adequate correspondence to their learning needs.

3.2. Discussion

Predictive analytics in the current research was implemented through multiple machine learning models that are associated with improvements in student engagement, performance, and retention in the Open and Distance Learning (ODL) settings. The relationships between the machine learning analysis and the educational experiment (pre-test and post-test) can be viewed in light of the way in which the machine learning models could pinpoint at-risk students and enable personalised interventions which were followed with the help of pre-and post-test outcomes.

The 5000 datasets that was uploaded to the predictive model consist of various data structure including student-id, programme, gender, course code, attendance, engagement, TMA score, previous exams score, total score, grades, and predicted outcome. representing 40.1% of the sample dataset, while 2,995 students were identified as pass students representing 59.9% of success rate. The study has shown that predictive analytics based adaptive pedagogical models hold huge potential in enhancing student engagement, performance, and retention in open and distance learning institutions (ODL). The application of machine learning algorithms, including the Logistic Regression, Support Vector machine (SVM), random forest and XGBoost, have been identified to be effective in identifying at-risk students and providing them with personalised interventions. The rate of accuracy in the usage of these models was also astonishing because all the models got an accuracy level of more than 96, a fact which underlines the effectiveness of the models in terms of ability to predict the student learning outcomes besides being able to offer the right interventions in due time.

Random Forest was the most successful model in terms of ROC-AUC of (98.38%) and PR-AUC of (97.58%) showing that it is most likely to be able to identify students at risk and high performing students. However, Logistic Regression and SVM outperformed Random Forest and XGBoost in terms of accuracy (97.43%), precision (97.65%), recall (95.95%) and F1-score (96.79%). The four models are effective in forecasting at-risk students hence their balanced nature in regards to false positives and false negatives. XGBoost and Random Forest were a little bit distinct whereby XGBoost attained a higher level of accuracy and recall and a lower level of ROC-AUC score. However, the four models were highly predictive, which means that they can be applied in the educational systems intended to enhance student performance.

Although the Logistic Regression and the SVM had similar results in various indicators (e.g., accuracy, precision, and recall), the Random Forest was chosen as the best model with higher results in the ROC-AUC (98.38%) and PR-AUC (97.58%). The two metrics are especially relevant in the case of imbalanced datasets, like the one in this paper, where the percentage of students at risk is lower than the percentage of non-at-risk students.

ROC-AUC is used to measure how the model can discriminate between at-risk and non-at-risk students and the discriminative power of Random Forest was the highest. While PR-AUC is the performance measurement of the model in relation to the positive class (i.e. at-risk students) which is essential in early intervention in educational institutions. The Logistic Regression and SVM were almost equal or even more accurate and precise; however, the value of ROC-AUC and PR-AUC in Random Forest makes it the most suitable due to its ability to detect at-risk students which is the primary purpose of the current research. The choice of the best model which is the Random Forest was not determined using one feature but rather its capability to be used everywhere to predict accurately the students under-at-risk and offer useful interventions especially when it comes to the goals of this study.

Among the 5,000 students, 1,000 were chosen to take part in the educational experiment, according to certain inclusion criteria, including willingness to take part in education experiment, accessibility to testing, and other aspects. These 1,000 participants are therefore described in the demographic table, which was evenly divided into experimental and control groups to implement the educational intervention, and the larger dataset was trained and evaluated the predictive models. This makes sure that the connection between the entire data and experimental sample is well comprehended and there is no misunderstanding about the scope or purpose of data. Although 50 faculty members were used in semi-interview, making the total number of both quantitative and qualitative to (n=1050) participants.

In the experimental and control group, the participants are mostly male (54.2%), with females at 42.8% while 3% prefer not to say their gender. The majority of participants are aged 18-29 (61.9%), followed by 30-39 years (33.3%). Students represent 95.4% of the sample, while faculty, administrators, IT/Analyst staff, and instructional designers are in the minority. Most respondents (faculty members) with advanced degrees hold an M.Sc. (50%), followed by Ph.D. (30%) and B.Sc. (20%). For experience 50% have 4-6 years and 50% have intermediate pred knowledge of predictive analytics. The experimental and the control groups are equally represented by the sample.

It was also confirmed that predictive analytics had an influence on student engagement and retention based on the pre-test and post-test scores of the experimental and control groups in various courses. The experimental group showed considerable improvements in scores in all situations, while the values of Cohen d values represent a large effect size. As an example, the mean change in the experimental group in BIO101 was 13 points, t-value of 9.19, and a p-value of less than 0.001, which represents a significant effect of the intervention. The control group demonstrated increase in 4 points, which was statistically insignificant when compared to the experimental group. The control group was meant to act as baseline over the experimental group, which had an intervention. The predictive model-based intervention was applied only to the experimental group thus the absence of the predictive model-based intervention in the control group. The fact that the results in the control group are not significant cannot be regarded as a drawback but as an anticipated consequence, as no intervention was used.

Other similar courses like CIT104, CHM101, CHM192, GST101, MTH101, and PHY101 also display the same trends in which the experimental group displayed a trend where they gained significantly in post-test scores as compared to the control group who had slight changes. The similarity in courses also establishes the efficacy of predictive analytics in improving student outcomes in learning, showing that predictive model-based personal interventions could have a strong influence on student engagement and student performance.

The qualitative findings revealed that the faculty members and students reported positive outcomes of the implementation of predictive analytics-based adaptive teaching. Most faculty members (80%) had a general concept of predictive analytics and 75% of the faculty members thought that personalised feedback and content recommendations helped to increase student engagement and better performance. Equally, students (70%) reported the value of personalised interventions in improving their learning experience, 60% of them were appreciating the feedback and the recommendations given by the predictive model.

Nonetheless, both the faculty members and the students were concerned with the problem of adopting the adaptive teaching techniques. Faculty members noted that it is time and work-consuming to make adaptive teaching possible, particularly in large classes; students expressed their concerns regarding the efficiency of some suggestions because they do not understand how the process works including ethics issues. These hurdles notwithstanding, the two groups realised the positive effects of the interventions on the student engagement and academic performance.

3.3. Benchmarking with Existing Studies

On the findings of current study, a comparison of the results with the similar studies in predictive analytics used in adaptive teaching shows that there are a few insights. The results of present study shows that the performance of predictive models, using four machine learning models including Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost. The key metrics of the performance revealed the accuracy of the Logistic Regression and SVM was 97.43%, and the accuracy of Random Forest and XGBoost was slightly lower (96.85% and 96.78% respectively). The highest ROC-AUC was obtained with random forest 98.38% and then XGBoost 97.76% compared to the work of [7] with 87% accuracy.

Also, an ensemble model of Random Forest and XGBoost have been found to be more effective than individual models in the same study, indicating that combining multiple models is more effective in predictive tasks in education. Other studies such as those of [3] and [4] also applied Random Forest and SVM in predicting student performance and the model attained a predictive accuracy of over 85% and 87% respectively. This can be compared to the findings of present study as every model performed at a level above 96% meaning that machine learning is very robust when used to predict the performance of students and is better than the previous studies.

More so, the present experiment indicated that there were large shifts in the study engagement in the experimental group, with large effect sizes (Cohens d 1.4), especially in courses such as BIO101 and CIT104. The gain in post-test scores in the experimental group was quite remarkable as opposed to the control group. The present study is the first to examine predictive analytics in relation to student engagement and retention, similar works like the one by [4] have demonstrated that predictive analytics lead to better post-test results after individualised interventions with 86% accuracy. Their research also used machine learning model and concluded that student success in online and distance learning relied on personalised interventions. Their study effect sizes were also significant which supports the present findings of the effectiveness of adaptive teaching but the present study is more significant than the previous studies because its uses seven courses in the experimental group as against one course used in the previous studies [4], [38].

The existing research also brought up crucial ethical issues, such as privacy of the data, trust, and quality of the data on engagement to make predictions. According to this, a recent study by [24] highlighted the importance of protecting the privacy of student data and ethical use of predictive models. This plays a significant role in keeping trust which can promote the use of predictive analytics in learning institutions. The present study has suggested ethical framework in compliance with the NDPA 2023 and GDPR 2018 [26], [27].

In general, the findings of the current study can be compared and, in certain ways, are better than the existing research in the domain looking at the accuracy and ROC-AUC of the models. The predictive models, especially Random Forest outperformed the other models in terms of ROC-AUC while Logistic Regression and SVM, model accuracy is higher and especially on predicting student performance, and the results on student engagement and retention are just a confirmation of the favourable effect of personalised intervention in adaptive teaching [4], [7].

3.4. Implications of the Study

There are a number of significant implications of this study to the implementation of predictive analytics in ODL settings as follows:

1) Personalisation of Learning: Predictive analytics has the potential to contribute to personalization of learning in ODL setting in a significant way. Educators can make the learning environment more responsive by forecasting the students at risk and offering special interventions that would satisfy the needs of individual students.
2) Better Retention and Performance: The results of the study show that adaptive teaching methodologies based on predictive analytics can contribute to improving student retention and student academic performance. This is especially true to the ODL institutions where in most cases the rate of students dropping out is greater than in the traditional face-to-face environment.

3) Scalability and Implementation Issues: The success of the predictive models in identifying at-risk students and tailoring the learning interventions to students, however, comes at a cost of resource-consuming nature of the latter, as well as the challenge of scalability. More studies are required to streamline such systems to be used on a large scale without overburdening the facilitators.
4) Training of the Facilitator: Faculty members need to have sufficient training on how to analyse the results of predictive models and apply the adaptive teaching methods in practise. The success of interventions based on the use of predictive analytics depends on the ability of the instructors to use the tools, which, in turn, is essential to ensure the success of the implementation.
5) Ethical Issues: Privacy, trust including bias and fairness is also noted as concerns in the implementation of predictive analytics model in ODL.

4. Conclusion and Future Work

The results of this research point to the high potential of the application of predictive analytics-based models in adaptive teaching within open and distance learning (ODL) institutions. The study has revealed that predictive analytics are associated with improvements in student engagement, student performance and retention through its ability to predict student academic performance and the detection of students at risk. The application of machine learning models, including Logistic Regression, Support Vector machine (SVM), random forest and XGBoost, showed that a random forest is best with an ROC-AUC score of 98.38%, and XGBoost came second with 97.76%. However, Logistic Regression and Support Vector Machine outperformed the other models with accuracy of (97.43%), precision of (97.65%), recall of 95.95%), and F1-score of (96.79%). The conclusion underline that these predictive models are strong instrument in changing the teaching strategies providing educators with the opportunity to offer individual interventions to fulfil the individual learning needs of individual students.

The research contributions in this work are relevant in the development of the use of predictive analytics in adaptive teaching in Open and Distance Learning (ODL) institutions. The study cannot overlook the important issues of student engagement and academic performance as well as student retention by creating and applying a machine learning-based predictive model. The study is the first of its kind since it integrates several algorithms, such as the Logistic Regression, Support Vector Machine (SVM), a random forest, and the XGBoost to identify the at-risk students and provide them with the personalised intervention. This method is especially helpful in ODL, where face-to-face communication is restricted, and it is possible to better adapt the teaching to the needs of single students. This research is also important to the field as it gives empirical evidence based on a large-scale sample of 5,000 students, where performance outcomes verify the efficiency of the predictive models in enhancing the outcomes of students. Besides, the work also revealed the significance of resolving ethical issues like privacy of data and bias in algorithms and suggests future research directions in optimising and scaled usage of the models in larger institutions of learning

In a nutshell, the field of predictive analytics realisation in the context of ODL is an attractive development of the educational technology domain. By leveraging the power of data to predict student performance and customise learning interventions, predictive analytics has become extremely valuable to the learning process and help to achieve better student outcomes by using the capability of data to anticipate the performance of students and tailor learning interventions [29], [31], [39], [40]. The research proves that the role of the human judgement and evidence-based stand can be considered as a key to the development of a more accepting and individualised learning process.

The research has some limitations as follows:

1) The accuracy of the predictions may have been affected by the issues with the data quality including the unfinished or inconsistent records of the students [14].
2) The predictive models may also have a scaling issue as the system would have to deal with a bigger amount of data in a variety of courses and cohorts.
3) Ethical concerns, especially in the privacy of student information and the control of predictive models, ought to be handled as well to ensure that such technologies are used in a manner that is responsible and transparent [26], [27].
4) The faculty members should also be sufficiently trained to be in position to comprehend and apply the insights of predictive nature in a manner that the interventions would make a difference and help the students [31].
5) Geographical limitation, as the study cannot be generalised since it was implemented in one ODL settings, (National Open University of Nigeria).

The future research should be centred on the following areas:

1) Long-term Impact: The long-term effect of predictive analytics-based interventions on the student success and retention after several terms would be an informative study of the sustainability of the interventions.
2) Scalability and Efficiency: Studies are necessary to fine-tune the predictive analytics system to use in large-scale implementation of the ODL

institutions to ensure that the system can process hefty amounts of data without affecting its performance or congesting the facilitators.

3) Combination with other E-learning Technologies: The combination of predictive analytics and other educational technologies, including gamification or AI-based tutoring systems, may be an engaging and holistic learning experience of students.

5. Declarations

5.1. Author Contributions

Danladi Moses Adayilo: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources; Visualisation, Writing - Original Draft, Project administration; **Ishaq Oyebisi Oyefolahan:** Writing—review and editing, Validation, Supervision; **Juliana Ngozi Ndunagu;** Writing—review and editing, Validation, Supervision; **Ebenezer Malcalm:** Writing—review and editing, Validation, Supervision; **Khanyisile Twabu:** Writing—review and editing, Validation, Supervision; **Nwando Anekwe:** Data curation.

5.2. Institutional Review Board Statement

Approval from the University Research Ethics Committee (UREC) was secured before embarking on this research.

5.3. Informed Consent Statement

All authors have read and agreed to the published version of the manuscript.

5.4. Data Availability Statement

The data presented in this study will be provided on request from the corresponding author.

5.5. Acknowledgment

There was no external funding. Grammarly AI was used in spell checking errors of the original draft manuscript.

5.6. Conflicts of Interest

The author declares no conflicts of interest.

6. References

[1] H. Shuaibu *et al.*, "AI-Enhanced Feedback Systems and Student's Learning Outcomes in Open, Distance and e-Learning," *Nigerian Open, Distance and e-Learning Journal (NODeLJ)*, vol. 2, pp. 90–107, May 2025. https://doi.org/10.60787/NODEL.V2.29.

[2] D. M. Adayilo, I. O. Oyefolahan, J. N. Ndunagu, C. Otuya, E. Malcalm, and K. Twabu, "AI-Powered Tutoring Systems for Personalised Learning Feedback in Developing Secondary Education Contexts", *J. Fut. Artif. Intell. Tech.*, vol. 2, no. 4, pp. 549–564, Dec. 2025. https://doi.org/10.62411/faith.3048-3719-290.

[3] R. R. Alonso, K. A. Carvajal, and N. R. Acevedo, "Role of Artificial Intelligence in the personalization of distance education: a systematic review," *RIED-Revista Iberoamericana de Educacion a Distancia*, vol. 28, no. 1, pp. 9–36, Jan. 2025. https://doi.org/10.5944/RIED.28.1.41538.

[4] M. F. Contrino, M. Reyes-Millán, P. Vázquez-Villegas, and J. Membrillo-Hernández, "Using an adaptive learning tool to improve student performance and satisfaction in online and face-to-face education for a more personalized approach," *Smart Learning Environments*, vol. 11, no. 1, Dec. 2024. https://doi.org/10.1186/s40561-024-00292-y.

[5] N. S. Alotaibi, "The Impact of AI and LMS Integration on the Future of Higher Education: Opportunities, Challenges, and Strategies for Transformation," *Sustainability (Switzerland)*, vol. 16, no. 23, Dec. 2024. https://doi.org/10.3390/su162310357.

[6] M. Mukred, U. A. Mokhtar, B. Hawash, H. AlSalman, and M. Zohaib, "The adoption and use of learning analytics tools to improve decision making in higher learning institutions: An

extension of technology acceptance model," *Heliyon*, vol. 10, no. 4, p. e26315, Feb. 2024. https://doi.org/10.1016/j.heliyon.2024.e26315.

[7] D. Green, E. Thompson, and I. Ochieng, "Predictive Analytics to Enhance Learning Outcomes: Cases from UK Schools," *J. Emerging Technologies in Education*, vol. 3, no. 1, pp. 34–43, Apr. 2025. https://doi.org/10.70177/jete.v3i1.2127.

[8] S. Ahmed, Md. S. Rahman, M. S. Kaiser, and A. S. M. S. Hosen, "Advancing Personalized and Inclusive Education for Students with Disability Through Artificial Intelligence: Perspectives, Challenges, and Opportunities," *Digital*, vol. 5, no. 2, p. 11, Mar. 2025. https://doi.org/10.3390/digital5020011.

[9] E.-Y. Seo, J. Yang, J.-E. Lee, and G. So, "Predictive modelling of student dropout risk: Practical insights from a South Korean distance university," *Heliyon*, vol. 10, no. 11, p. e30960, Jun. 2024. https://doi.org/10.1016/j.heliyon.2024.e30960.

[10] H. M. Qadir, M. T. Suleman, R. A. Khan, M. Sohaib, M. J. Hasan, and S. A. Hussain, "Optimizing learning outcomes: a deep dive into hybrid AI models for adaptive educational feedback," *J Big Data*, vol. 12, no. 1, p. 144, Jun. 2025. https://doi.org/10.1186/s40537-025-01187-6.

[11] A. Abukader, A. Alzubi, and O. R. Adegboye, "Intelligent System for Student Performance Prediction: An Educational Data Mining Approach Using Metaheuristic-Optimized LightGBM with SHAP-Based Learning Analytics," *Applied Sciences*, vol. 15, no. 20, p. 10875, Oct. 2025. https://doi.org/10.3390/app152010875.

[12] Z. N. Khlaif *et al.*, "University Teachers' Views on the Adoption and Integration of Generative AI Tools for Student Assessment in Higher Education," *Education Sciences*, vol. 14, no. 10, Oct. 2024. https://doi.org/10.3390/educsci14101090.

[13] J. Chun *et al.*, "A Comparative Analysis of On-Device AI-Driven, Self-Regulated Learning and Traditional Pedagogy in University Health Sciences Education," *Applied Sciences*, vol. 15, no. 4, p. 1815, Feb. 2025. https://doi.org/10.3390/app15041815.

[14] M. Almutairi, A. Alamry, G. Rahman, and B. A. Mudhsh, "The role of AI-powered learning analytics in enhancing EFL curriculum design and learning outcomes in higher education," *International Journal of Innovative Research and Scientific Studies*, vol. 8, no. 7, pp. 477–487, Oct. 2025. https://doi.org/10.53894/ijirss.v8i7.10480.

[15] K. Almeman, F. E. Ayeb, M. Berrima, B. Issaoui, and H. Morsy, "The Integration of AI and Metaverse in Education: A Systematic Literature Review," *Applied Sciences (Switzerland)*, vol. 15, no. 2, Jan. 2025. https://doi.org/10.3390/app15020863.

[16] S. Weydner-Volkmann and D. Bär, "Student autonomy and Learning Analytics: Philosophical Considerations for Designing Feedback Tools," *Journal of Learning Analytics*, vol. 11, no. 3, pp. 1–14, Dec. 2024. https://doi.org/10.18608/jla.2024.8313.

[17] E. Salami and I. Nwankwo, "Regulating the privacy aspects of artificial intelligence systems in Nigeria: A primer," *African Journal on Privacy and Data Protection*, vol. 1, no. 1, Jul. 2024. https://doi.org/10.29053/ajpdp.v1i1.0011.

[18] A. Christodoulou and C. Angeli, "Adaptive Learning Techniques for a Personalized Educational Software in Developing Teachers' Technological Pedagogical Content Knowledge," *Front. Educ.*, vol. 7, art. 789397. https://doi.org/10.3389/feduc.2022.789397.

[19] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Q.*, vol. 13, no. 3, pp. 319–340, Sep. 1989. https://doi.org/10.2307/249008.

[20] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view", *MIS Quarterly*, vol. 27 no. 3, 425–478, (2003). https://doi.org/10.2307/30036540.

[21] M. N. Yakubu, N. David, and N. H. Abubakar, "Students' behavioural intention to use content generative AI for learning and research: A UTAUT theoretical perspective," *Education and Information Technologies*, vol. 30, no. 13, pp. 17969–17994, Aug. 2025. https://doi.org/10.1007/s10639-025-13441-8.

[22] Y. M. Hemmler and D. Ifenthaler, "Self-regulated learning strategies in continuing education: A systematic review and meta-analysis," *Educational Research Review*, vol. 45. Elsevier Ltd, Nov. 2024. https://doi.org/10.1016/j.edurev.2024.100629.

[23] P. E. Smaldino *et al.*, "Information architectures: a framework for understanding socio-technical systems," *npj Complexity*, vol. 2, no. 1, Apr. 2025. https://doi.org/10.1038/s44260-025-00037-z.

[24] Q. Liu and M. Khalil, "Understanding privacy and data protection issues in learning analytics using a systematic review," *British Journal of Educational Technology*, vol. 54, no. 6, pp. 1715–1747, Nov. 2023. https://doi.org/10.1111/bjet.13388.

[25] J. Zhu, "Innovative Exploration of the Integration Path of Artificial Intelligence Ethics and College Students' Ideological and Political Education," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024. https://doi.org/10.2478/amns-2024-2280.

[26] Federal Republic of Nigeria, "Nigeria Data Protection Act, 2023." 2023. [Online]. Available: https://cert.gov.ng/ngcert/resources/Nigeria_Data_Protection_Act_2023.pdf.

[27] European Union, "EU General Data Protection Regulation (GDPR), 2018." 2018.

[28] C. Mutimukwe, O. Viberg, L. M. Oberg, and T. Cerratto-Pargman, "Students' privacy concerns in learning analytics: Model development," *British Journal of Educational Technology*, vol. 53, no. 4, pp. 932–951, Jul. 2022. https://doi.org/10.1111/bjet.13234.

[29] S. Suharti and I. Hutabarat, D. C. Llamas, "Adaptive Learning Analytics for Tracking Student Performance and Predicting Academic Success in Digital Classrooms," *International Journal of Educational Technology and Society*, vol. 1, no. 3, pp. 34–43, Sep. 2024. https://doi.org/10.61132/ijets.v1i3.411.

[30] E. Ahmed, "Student Performance Prediction Using Machine Learning Algorithms," *Applied Computational Intelligence and Soft Computing*, vol. 2024, 2024. https://doi.org/10.1155/2024/4067721.

[31] A. Kaveri, J. Korpi, A. Silvola, and H. Muukkonen, "Staff perspectives on key dimensions of institutional learning analytics implementation," *Education and Information Technologies*, Oct. 2025. https://doi.org/10.1007/s10639-025-13790-4.

[32] A. Angeioplastis, J. Aliprantis, M. Konstantakis, and A. Tsimpiris, "Predicting Student Performance and Enhancing Learning Outcomes: A Data-Driven Approach Using Educational Data Mining Techniques," *Computers*, vol. 14, no. 3, p. 83, Feb. 2025. https://doi.org/10.3390/computers14030083.

[33] R. Manprisio, M. A. Salam, S. T. Mohmad, and S. V. Medasani, "Redefining Learning Paradigms: Integrating Artificial Intelligence into Modern Classrooms," *Ubiquitous Learning*, vol. 17, no. 2, pp. 157–177, 2024. https://doi.org/10.18848/1835-9795/CGP/v17i02/157-177.

[34] M. A. Aslam, F. Murtaza, M. E. U. Haq, A. Yasin, and N. Ali, "SAPEx-D: A Comprehensive Dataset for Predictive Analytics in Personalized Education Using Machine Learning," *Data*, vol. 10, no. 3, Mar. 2025. https://doi.org/10.3390/data10030027.

[35] A. Gaitantzi and I. Kazanidis, "The Role of Artificial Intelligence in Computer Science Education: A Systematic Review with a Focus on Database Instruction," *Applied Sciences (Switzerland)*, vol. 15, no. 7, Apr. 2025. https://doi.org/10.3390/app15073960.

[36] J. Garzón, E. Patiño, and C. Marulanda, "Systematic Review of Artificial Intelligence in Education: Trends, Benefits, and Challenges," *Multimodal technologies and Interaction*, vol. 9, no. 8, p. 84, Aug. 2025. https://doi.org/10.3390/mti9080084.

[37] A. Martín-Rodríguez and R. Madrigal-Cerezo, "Technology-Enhanced Pedagogy in Physical Education: Bridging Engagement, Learning, and Lifelong Activity," *Education Sciences*, vol. 15, no. 4, Apr. 2025. https://doi.org/10.3390/educsci15040409.

[38] C. Merino-Campos, "The Impact of Artificial Intelligence on Personalized Learning in Higher Education: A Systematic Review," *Trends in Higher Education*, vol. 4, no. 2, Jun. 2025. https://doi.org/10.3390/higheredu4020017.

[39] M. N. Gul, W. Abbasi, M. Z. Babar, A. Aljohani, and M. Arif, "Data driven decisions in education using a comprehensive machine learning framework for student performance prediction," *Discov Computing*, vol. 28, no. 1, p. 153, Jul. 2025. https://doi.org/10.1007/s10791-025-09585-3.

[40] T. N. Nguyen and H. T. Truong, "Trends and emerging themes in the effects of generative artificial intelligence in education: A systematic review," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 21, no. 4, pp. 1–11, 2025. https://doi.org/10.29333/ejmste/16124.