Journal of Economic Sciences
Volume 4, Issue 2
https://www.jesciences.com

Predicting Students' Academic Outcome in Econometrics: A Comparative Analysis between Traditional and Machine Learning Methods

Hamzat Salami[1,*], Cletus Usman Idoko[1], Idris Ahmed Sani[1]

Affiliations

1. Prince Abubakar Audu University, Anyigba, Kogi State, Nigeria

*Corresponding Author Email: salami.h@ksu.edu.ng

Timeline

Received: Sep 19, 2025
Revised: Dec 22, 2025
Accepted: Dec 23, 2025
Published: Dec 24, 2025

DOI
https://doi.org/10.55603/jes.v4i2.a6

Abstract

The Sustainable Development Goals acknowledge education as an important producer of human capital and sustainable economic development. As the educational engagements becoming complex, present a challenge to the applicability of traditional methods of analysis. This paper is a comparison of traditional and AI-enhanced machine learning (ML) methods to analyze student performance in Econometrics at Prince Abubakar Audu University, Nigeria. The study utilized 897 students sample and 13 features of student's demography and academic performance attributes in Econometrics prerequisite undergraduate Economics courses as Statistics for Economists, Mathematics for Economists and Macroeconomics courses. The logistic regression model that explained traditional methods, found six predictors that significantly affected the academic performance of students in Econometrics but with a low accuracy. On the contrary, ML algorithms used in the study, such as K-Nearest Neighbor, Random Forest, and Support Vector Machine (SVM), worked much better. SVM had the highest accuracy of 85.48%. The study findings show the effectiveness of AI-enhanced methods in processing complex educational data compare to conventional approach. The research contributes to the desire need of more widely used of the innovative approach to produce credible information and inform decision-making based on the data to improve the learning outcomes and quality of education in Nigerian higher institutions.

Keywords: *Education Data Mining; Traditional Analytical Methods; Machining Learning Classifiers*

JEL Classification: *A22, C01, C53, C59*

1. Introduction

Education can be both formal and informal; specifically, it is the formalized system through which societies pass knowledge down through the generations (Johnson & Majewska, 2022). Moreover, the profound impact of education on human development and economic growth is indisputable. In fact, the neoclassical theory of economic growth has long recognized education as a pivotal determinant of human capital development, leading to sustainable economic progress (Jhingan, 2017; Mankiw, 2020). As a result, it has become one of the main prioritized pillars of the Sustainable Development Goals (SDGs); in particular, one of the primary objectives is SDG 4, which seeks inclusive, equitable, and quality education (Opesemowo & Adekomaya, 2024). Furthermore, the quality of education is typically assessed through funding, delivery of instruction, and academic performance (Olalekan et al., 2020; Wijaya & Saputri, 2019), which are veritable yardsticks to evaluate educational achievements across countries of the world.

However, the significance of proper educational funding in the human capital development of any economy cannot be overstated. Recognizing this, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) recommended in 2015 that member countries allocate 4 to 6% of their GDP or 15 to 20% of public expenditure to the education sector in their annual budget allocations (UNESCO, 2021). Despite these recommendations, many countries continue to lag behind in educational funding, perpetuating disparities in resources and educational opportunities. Consequently, this disparity is a stark indicator of why developed countries surge ahead in fields like science and technology. Indeed, the level of education offered in developed economies is renowned for its high standards, and the quality of education stands as a well-established marker that distinguishes advanced nations from their less developed counterparts (Suresh & Kumaravelu, 2017). Therefore, education funding serves as a vital indicator of the quality of education within a nation.

In addition to education funding as an indicator of qualitative education for human capital development, students' academic outcomes are assessed through course grades or cumulative grades for the award of degree certificates (Zhao et al., 2023). This assessment demonstrates the level of knowledge and skills comprehension among learners (Olalekan et al., 2020). Moreover, the measurement of performance, either on a course basis or on a general basis of graduation, assists in early interventions, which can decrease failure rates and therefore act as a dependable system quality indicator in educational research (Alwarthan et al., 2022). While traditional analytic options, which are proponents of the classics known for explanatory simplicity and ease, have long been useful in educational studies for analyzing educational data (Ahmed et al., 2021; Ali et al., 2013; Ampofo & Osei Owusu, 2015; Idoko & Emmanuel, 2015; Martin Sanz et al., 2017; Tadese et al., 2022), nevertheless, as the volume of data and educational dynamics grows, these approaches are becoming less effective at solving nonlinear relationships and problems with big data (Jiao et al., 2022). Thus, this has led to disruptive analytical methods using AI and machine learning made possible through modern educational analytical theories (Hasegawa et al., 2020).

Consequently, there is an increasing desire to utilize artificial intelligence (AI) and machine learning (ML) driven analytical tools to analyze the large amounts of education data that are created daily across the globe (Rajula et al., 2020). Specifically, Educational Data Mining (EDM), or the practice of analyzing the demographic and academic attributes of students to determine their performance, has been gaining more significance (Nabil et al., 2021; Yağcı, 2022). Studies have shown that AI enhanced EDM is very effective in personalized learning, curriculum design, teacher assistance, predictive modeling, and early intervention (Olalekan et al., 2020; Yakubu & Abubakar, 2022). In addition, it increases teaching and learning engagement and enhances learning outcomes (Opesemowo & Adekomaya, 2024).

Similarly, advancements in other areas such as medicine and materials science include extensive comparisons between traditional and AI procedures (Amin et al., 2021; Hasegawa et al., 2020; Rajula et al., 2020; Sheetal et al., 2023). However, empirical comparisons of traditional and AI supported analytics remain rather scarce in other areas of study. This divide indicates that there is a necessity to consider the predictive capabilities of both methods in light of the influx of educational information and data concerning Internet of Things (IoT)–facilitated information systems. To address this lack of empirical comparison in social science educational engagements, this paper fills such a requirement by examining a sample from the Economics Department at Prince Abubakar Audu University (PAAU) in Nigeria. Specifically, this dataset contains demographic and educational information of graduates who were enrolled from 2016 to 2020. Finally, this paper draws a comparison between conventional analytical methods and ML algorithms in student performance prediction in Econometrics using classifiers such as K Nearest Neighbor, Random Forest, and Support Vector Machine (SVM).

2. Literature Review

2.1 Conceptual Review

Education is a structured process of transmitting knowledge formally or informally, vital for individual and societal progress (Johnson & Majewska, 2022). A qualitative education system relies on evaluating students' academic performance, which involves analyzing the vast data generated daily in higher institutions. Such analyses provide insights that enhance learning and inform decision-making processes (Zhao et al., 2023). Students' academic performance measurement analyses through the traditional analytic methods and tools are dominant of education data mining (Ahmed et al., 2021; Ali et al., 2013). These techniques are robust, interpretable, and foundational for education research. However, the complexity and nonlinearity of modern educational data challenge their applicability, necessitating advanced approaches (Hasegawa et al., 2020).

Modern education data analysis leverages artificial intelligence (AI) and machine learning (ML) to process complex datasets, addressing the limitations of traditional methods (Kehinde et al., 2022). These techniques enable personalized learning, curriculum improvement, predictive analytics, and early interventions (Nabil et al., 2021). AI and ML systems efficiently analyze large datasets, learn patterns, and make precise predictions on unseen data (Olatunji et al., 2023). They also offer scalability and flexibility, making them ideal for the dynamic needs of tertiary institutions (Zualkernan, 2021). Therefore, while traditional methods remain foundational, modern AI and ML techniques provide advanced tools for addressing the complexities of contemporary educational data, driving qualitative improvements in learning outcomes.

2.2 Review of Related Works

In contrast to fields like medicine, material science, and disease diagnosis (Amin, et al., 2021; Hasegawa et al., 2020; Rajula, et al., 2020; Sheetal, et al., 2023) where comparative analyses between traditional and modern data analysis methods have been extensively delineated to facilitate the adoption of approaches, studies in education have largely been conducted independently created a wider gap in the literature of educational data mining. The following are the review studies in this area of endevours.

Tadese et al., (2022) study utilized conventional analytical tools to identify the determinants of academic performance among university students in Southern Ethiopia through an institution-based cross-sectional study spanning 29 days. With a total of 659 enrolled students, data collection used a self-administered questionnaire. Employing a multistage sampling technique, the study revealed associations between smoking, age, field of study, and academic performance. This study emphasized on the qualitative aspects of students' academic performance factors with traditional analytics approach of correlation and ANOVA. Alani and Hawas, (2021) conducted a similar study to ascertain the determinants affecting students' academic performance at Sohar University, Oman. Utilizing a quantitative survey methodology involving structured questionnaires administered across various faculties, the study identified environmental factors and teaching techniques as significant determinants of academic performance at the university. In the same vain, Ahmed et al. (2021) conducted a study at Kogi State University, Anyigba, to examine factors influencing students' academic performance using a traditional binary logistic regression approach. The research aimed to uncover and address strategic solutions for improving academic outcomes among students. The authors selected several key variables for analysis, including students' Joint Admission Matriculation Board (JAMB) examination scores (representing prior academic achievement), the number of friends on social networking sites, and the alignment of classroom knowledge with real-world applications (indicating students' motivation). Other factors considered were lecture attendance frequency, weekly study hours, discussion of lecture ideas (academic habits), and the time spent on additional jobs alongside academic responsibilities. The study findings revealed that these variables negatively impacted students' academic performance in their course of disciplines at higher institution of learning.

Idoko and Emmanuel (2015) conducted a survey research to explore factors influencing teacher effectiveness in teaching Economics in Nigerian secondary schools. Findings revealed teaching strategies, staff welfare, motivation, and cognitive experience as pivotal determinants of students' performance in Economics. The study advocated for the overhaul of teaching methods, improved staff welfare, and enhanced recruitment policies to optimize Economics education in secondary schools. Ampofo and Osei-Owusu (2015) investigated factors influencing academic performance in Senior High Schools in Ghana. Their findings identified parental education, academic ambition, and students' efforts as key determinants. In addition, Ali et al., (2013) explored factors affecting academic performance among graduate students at Islamia University of Bahawalpur Rahim Yar Khan Campus, Pakistan. Utilizing regression analysis, they revealed age, socioeconomic status, and study hours as significant contributors to academic performance, advocating for the incorporation of such variables in academic evaluation models. The reviewed studies approaches and findings elaborated on the simplicity of conventional analytical methods in deterministic and associativity modelling with poor predictive capacities which formed the strength of machine learning modelling analytics.

The burgeoning field of education data mining has seen various studies aimed at predicting student success and performance utilizing modern analytical tools of artificial intelligence and machine learning techniques in course-based education data mining. Ozkan et al., (2023) examined factors contributing to student success in College Algebra courses, achieving an 85% accuracy rate in predicting course outcomes. Similarly, Nabil et al., (2021) and Zualkernan, (2021) employed machine learning algorithms to predict students' performance in programming and engineering courses, respectively, showcasing high prediction accuracies.

Omolewa et al., (2019) and Tsiakmaki et al., (2018) delved into education data mining to predict course-based student performance, emphasizing the significance of demographic and academic features in predicting academic outcomes with above the bench mark accuracies and precisions among the used algorithms. O'Connell et al., (2018) investigated factors influencing student performance in mathematics courses, demonstrating the importance of past performance and experiences in predicting student success. Despite these advancements, there remains a dearth of comparative analyses between traditional and modern methods of education data analysis, particularly in Nigerian tertiary institutions. Hence, the study conducted at Prince Abubakar University, Anyigba, Kogi State, Nigeria, bridged this gap and contribute to the existing knowledge in education data mining, providing essential insights to guide decision-makers in the education sector.

2.3 Theoretical Framework

The adopted theoretical framework for this study revolves around the field of learning analytics. Learning analytics is a multidisciplinary approach that combines education, data science, and technology to analyze and interpret data related to the learning process. Siemens and Baker, (2012) revealed that learning analytics comprises education data mining and learning analytics knowledge. The formal is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in while the latter is defined as the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which learning occurs (Gasevic, et al., 2011; Romero & Ventura, 2010). However, the theoretical framework of this study relied on the EDM frame work aspect of learning analytics adopted in the studies (Nabil, et al., 2021; Ozkan & Stelmach 2023; Zualkernan 2021; Yauri, 2022). The adoption of EDM as a component of learning analytics theoretical framework of this study helps to provide a holistic link and explanations for every components of the study.

3. Methodology and Model

3.1 Description of Research Techniques

The study involved a comparison of traditional (logistic regression) and the innovative AI-based methods from simple algorithm to complex ones of K-Nearest Neighbor (K-NN), Random Forest and Support Vector Machine in forecasting students' academic outcomes in Econometrics of PAAU Economics graduates. The study sample size of 897 observations consisted of demographic and academic features were analyzed in traditional analytical environment of (E-view 12) and machine learning modelling platform of (WEKA).

3.2 Data Description and Measurement

The research is aimed at the demographic and academic performance predictors of the chosen Economics students of PAAU. The predictor variables are ID of students, their gender and the prerequisite courses, which are Principles and Theories of Economics, Mathematics for Economists and Introduction to Statistics. These courses cut across mathematics, statistics and macroeconomics theories. Econometrics I and II are target class variables. Such requirements are provided during the first and second years, which is followed by studying Econometrics in the third year. The research involved the use of 13 attributes and 897 observations to examine the performance of the students in Econometrics based on their performance in other related courses.

Table 1 describes the academic characteristics of the performance of students. The research classifies the demographic characteristics into male or female and academic performance in terms of traditional university grading scale (100-0 percent) as outlined in Table 1. The grading level goes between A (Excellence) 100-70 percent, B (Very good) 69-60 percent, and F (Fail) 39-0 percent. The Pass and Fail predictor variables will be binary-coded as 1 or 0 to be assembled by machine learning classification using WEKA. These modifications are needed in proper analysis.

Table 1: Academic Features Measurement Description

Sr. No	Grading (percentage %)	Remark	Grade classified to categorical variable
I	100-70 (A)	Excellency	PASS
II	69-60 (B)	Very good	
III	59-50 (C)	Good	
IV	49-40 (D or E)	Poor	
V	39-0 (F)	Fail	FAIL

3.3 Data Evaluation and Validation

The iterating process of machine learning model involves the performance measurement phase that includes data partitioning, data training, data testing, evaluation hyper parameter optimization discussed below:

Data Training and Testing

In machine learning model building, the research findings are becoming bifurcated into the training and test set data. Types of exist machine learning partitioning include percentage slip and N-fold cross validation. This analysis took percentage of slip partitioning of 80 and 20 percent model training, model testing and 5 fold cross validation was adopted in hyper parameters tuning to avoid the machine learning problem of course of dimensionality overfitting or under fitting Olatunji et al., (2022)

Evaluation Criteria

Post-training and post-testing of machine learning models are a significant part of the performance measurement of validation of results. The performance of the model is evaluated using the evaluation metrics of accuracy, precision, recall, F- measure, or mean squared error and confusion matrix depending on the problem. The primary performance measure is machine learning model testing accuracy that calculates the proportion of correctly classified instances to the total instances of testing set. Precision is the ratio of true positive to all predicted positive, recall is the ratio of true positive to all actual positive, and F-measure is the performance of each of the classes (Alassaf, 2018).

Improvement in Performance using Hyper parameter Optimization

The hyper parameters of models can also be tuned by techniques such as grid search or random search to identify better hyper parameters in each model and modify the hyper parameters depending on the effect on model performance. The classification process of the study was used to tune all the hyper parameter processes that were specific to each classifiers of this study. So, it is worth mentioning that machine learning is an evolutional process, and models can be refined continuously either more data is obtained or more effective methods are invented

3.4 Data Analytical Procedure

Experimental test of this study was done using E-views 12 version and WEKA. The dichotomy dependent and independent variables of the performance evaluation of students were performed through a logistic regression model on the traditional platform of E-views with all the analyses corresponding to machine learning modelling performed on WEKA, a free machine learning analysis tool. It was utilized to preprocess and process data to obtain useful statistics of both the predictor and the predicted variable in the research. Moreover, the feature ranking was also done using the platform by determining the correlation of each feature with the class label. In this research, where the machine learning data analysis involved classification analysis, parameter tuning, feature selection, training and testing of the dataset were done in WEKA machine learning platform in the processes of data analysis. There were four classifiers (K-NN, RF and SVM) involved in the process. The dataset was partitioned using direct partitioning schemes in terms of model training and testing. In particular, 80:20 percentage cut was utilized to do parameter and feature selection. The model training and testing classification was also carried out using the same direct partitioning approach. Parameters The classification algorithms were all trained using optimized parameter values obtained using the optimization solution.

4. Results and Discussions

4.1 Data Presentation and Analysis

Table 2: Descriptive Statistics

Attribute	**Variable class**	**Missing value (%)**	**No. of male (%)**	**No. of female (%)**	**Sample size**
Gender	Categorical	Null	545	352	897
Attribute	**Variable class**	**Missing value (%)**	**No. of passed (%)**	**No. of failed (%)**	**Sample size**
ECO 102	Categorical	26 (3%)	718(80%)	179 (20%)	897
ECO 103	Categorical	23 (3%)	775(86%)	122 (14%)	897
ECO 104	Categorical	31 (3%)	781(87%)	116 (13%)	897
ECO 105	Categorical	32 (4%)	793(88%)	104 (12%)	897
ECO 106	Categorical	49 (7%)	749(84%)	148 (16%)	897
ECO 203	Categorical	30 (3%)	842(94%)	55 (6%)	897
ECO 204	Categorical	37 (4%)	824(92%)	73(8%)	897
ECO 205	Categorical	37 (4%)	710(79%)	187 (21%)	897
ECO 206	Categorical	48 (5%)	683(76%)	214 (24%)	897
ECO 207	Categorical	19 (2%)	762(85%)	135(15%)	897
ECO 208	Categorical	27 (3%)	755(84%)	142 (16%)	897
ECOMETRICS	Nominal	84 (9%)	688(77%)	209 (23%)	897

Table 2 provides descriptive statistics of students demographic and academic performance features, including gender distribution (61% male, 39% female) out of 897 sampled students. Academic features are presented by variable class, with missing values less than 50% filled with minority instances during data preprocessing. Each course's performance is detailed in terms of passed and failed percentages. For example, Macroeconomic Analysis (ECO 203) had the highest pass rate (94%) and lowest failure rate (6%). Econometrics (ECONOMETRICS) had the lowest pass rate (77%) and highest failure rate (23%). Gender distribution and academic performance metrics are vital for both the traditional and machine learning predictive model analyses. The Table 3 below presents the traditional analytical technique of logistic regression model.

Table 3 presents a model intercept parameter estimate of -4.369678 represents the likelihood ratio of null model of the logistic regression with 5% significant level ($p < 0.05$). It is an indicator that the model is unlikely to observe such a value if the true coefficient were zero. The logistic regression model results in Table 3 shown that the following predictor variables (*ECO 207, ECO 208, ECO 104, ECO 106, ECO 204, ECO 205)* have coefficients that are statistically significant ($p < 0.05$). The likelihood of a student passing Econometrics in the intermediate level of studies increases as he or she passes the above prerequisite courses with 5% level of significant. The results in Table 2 also revealed that the following independent variables such as *ECO 102, ECO 103, ECO 105, ECO 206, GENDER* have coefficients that are not statistically significant ($p > 0.05$) suggests that there is insufficient evidence to conclude that these variables have a significant effect on the likelihood chances of a student passing Econometrics in 300 Level if he or she passes the above listed courses putting into consideration the students' gender.

Table 3: Logistic Regression Model Estimates

Variable	Coefficient	Std-Error	Z-Statistics	Probability
C	-4.369678	0.537356	-8.131806	0.0000
ECO 207	0.31114	0.246695	1.261232	0.2072
ECO 208	1.341571	0.222867	6.019612	0.0000
ECO 102	0.348156	0.226864	1.534645	0.1249
ECO 103	0.307088	0.258415	1.188351	0.2347
ECO 104	1.059775	0.248938	4.257189	0.0000
ECO 105	0.414784	0.275415	1.506031	0.1321
ECO_106	0.708188	0.233854	3.02834	0.0025
ECO 203	0.325616	0.356937	0.91225	0.3616
ECO 204	0.909954	0.318648	2.85567	0.0043
ECO 205	0.827031	0.215253	3.842142	0.0001
ECO 206	0.354121	0.213889	1.655629	0.0978
GENDER	-0.073295	0.195225	-0.37544	0.7073

Regression Statistics:

Pseudo R^2 = 0.229515,   *Count* R^2 = 0.1884 (19%)

Note: Chi-Square (*χ^2 cal.*) = 231.91, *Chi-Square* (*χ^2 statistics at 12 df with 0.05 significant level*) = 21.0261

The logistic regression model statistics of *Pseudo* R^2 of 0.229515 (23%) and *Count* R^2 of 0.1884 (19%) indicate poor fit of the model, meaning the predictor variables of the model under the traditional approach were not sufficient to explained the predicted variable despite the general model significant attainment through Chi-square. Chi-Square represents the overall significance of the model. The calculated chi-square value (231.91) obtained is compared to the critical chi-square value (21.03) at the specified degrees of freedom and significance level (12 df, 0.05). In this case, the calculated chi-square value exceeds the critical value, indicating that the model as a whole is statistically significant, an indication that the study worthwhile exploring with superior analytical approach.

Table 4: Result of Difference Feature Subsets

Number of Attributes	Accuracy of the Classifiers			Average Accuracy
	K-NN	RF	SVM	
12 (All)	84.02	78.43	82.89	82
10	86.59	81.56	84.92	84
9	84.36	82	84.36	83
5	84.92	83.24	85.48	85
1	79.32	79.33	80.45	80

Table 4 presents the features combinations classification process to determine the best performing accuracy among the selected algorithms. Among the attributes combination of various values, the combination of the first to five features presented in Table 4 (*ECO208, ECO104, ECO205, ECO102 and ECO204*) gave an optimal performing features combination with average accuracy of 85%. The classification process that resulted to the optimal features combination accuracies among the selected classifiers was carried out through percentage slip partitioning of 80% for model training and 20% for model testing evaluation processes with thorough parameters tuning of various selected machine learning classifiers.

Figure 1: Result of Best Performing Feature Subsets Diagram

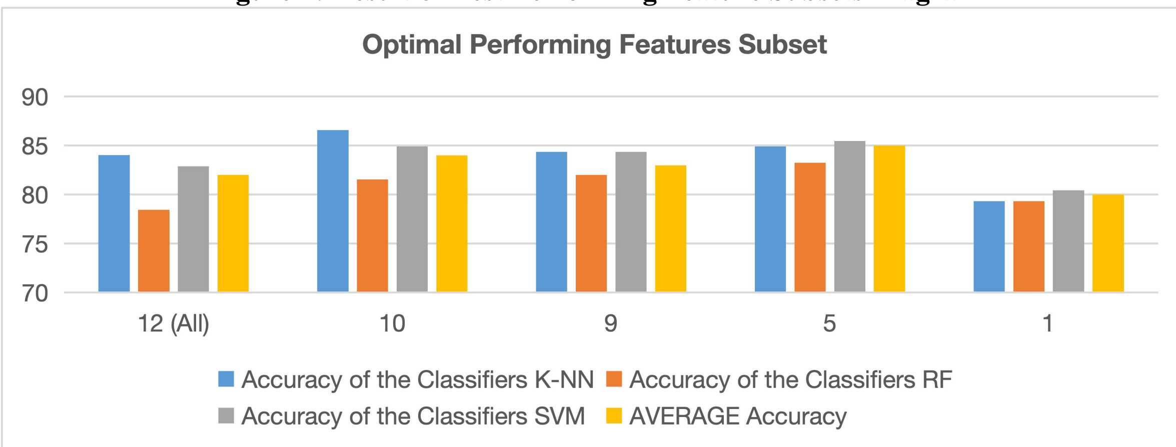

Figure 1 reveals how each selected classifier performed under difference values of attributes combinations. K-Nearest Neighbor performed best at tenth (10th) features combination of 86.59 %, Random Forest attained optimal at 5th features combination of 83.24%, Support Vector Machine achieved the highest performing attributes combination at fifth (5th) features conjoined of 85.48% average accuracy, the 5th features combination was voted as the best performing features subsets of the study and therefore adopted as the optimal features combination for the study classification analysis.

Hyper parameter optimization of each classifiers

Hyper parameter tuning is crucial in machine learning to optimize accuracy. The study hyper parameter tuning for each classifier including K-Nearest Neighbor used K=15 and Euclidean Distance, while Random Forest's best performance came with NumExecutive=1 and Seed=2. Support Vector Machine achieved optimal tuning at C=20 and PolyKernel. The below is classification performance on the optimal attributes of the study.

Table 5: Classification Performance on Optima Features

Classifier	Accuracy	Precision	Recall	f-measure
K-NN	84.92	0.84	0.85	0.82
RF	83.24	0.81	0.83	0.82
SVM	85.48	0.84	0.86	0.85

Table 5 presents the performance measures of each classifiers with Support Vector Machine excelling at 85.48% accuracy, 0.84 precision, 0.86 recall, and 0.85 f-measure. Consequently, Support Vector Machine was chosen as the study's classification model, given its superior performance compared to other classifiers as shown below.

Figure 2 is the classification performance accuracy of the optimal features selected. The diagram revealed how Support Vector Machine outperformed other machine learning algorithms on the optimal features combination classification process of 85.48% while K-Nearest Neighbor, Random Forest have 84.92% and

83.24% respectively. This is a justification of adoption SVM model as best performing model for the study among other selected classifiers as confirmed by confusion matrices below.

Figure 2: Classification Performance Accuracy of Optimal Features Chart

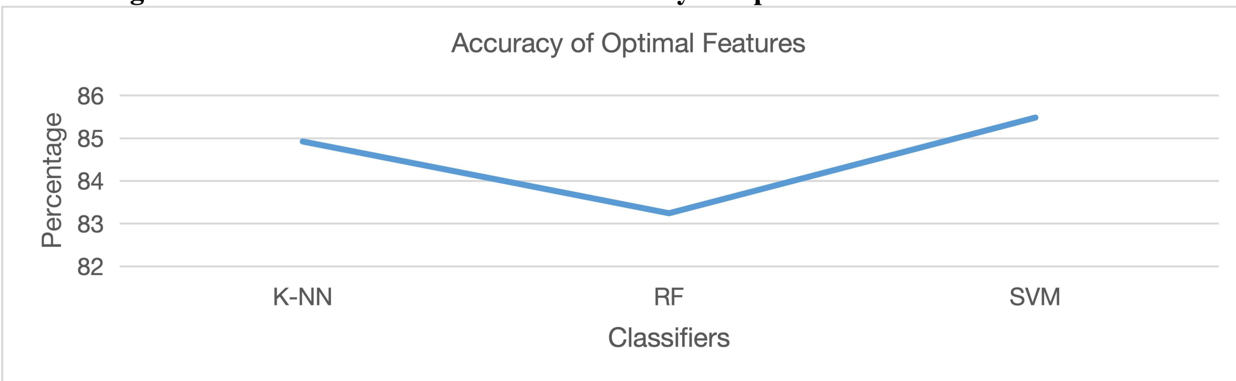

Table 6: Classification Confusion Matrices

	PREDICTED					
					K-NN	
					142 (TP)	2 (FN)
					25 (FP)	10 (TN)
		Random Forest				
A			136 (TP)	8 (FN)		
C			22 (FP)	13 (TN)		
T U	SVM					
A	136 (TP)	8 (FN)				
L	18 (FP)	12 (TN)				

According to Table 5, SVM shows the highest performance in terms of all measures (accuracy, precision, recall, and f-measure). The confusion matrices above exhibit further look into the actual and predicted classes by K-NN, RF and SVM in Tables 6. The most significant metric to examine in the confusion matrices above is the False Positive (FP) rate. False Positive rate is the ratio of the failed students in Econometrics but wrongly classified as passed students by a model. Such a student among other would exhibit indifference attitude to studies believing he or she is heading to passing grades in the course under study. The aim of machine learning analysis is to find and adopt the machine learning classifier that minimizes False Positive (FP) in the predicted outcomes to actual outcomes depicted by above matrices. The lowest FP rates were identified in SVM followed by RF and K-NN classifiers.

4.2 Discussion of Findings

Descriptive statistics was utilized to summarize the distribution of variables across gender and students' academic performance in the selected courses for the study. To address missing values, the related class's minority instance was used to fill in the gaps, following Olatunji's (2023) recommendations for handling missing data in machine learning model preprocessing. Analysis of the descriptive statistics revealed significant findings: 94% of students passed Macroeconomics Analysis 1 (*ECO 203*), with a 6% failure rate. In contrast, *Econometrics*, the target variable, exhibited the highest failure rate at 23%, with a 77% passed rate, consistent with the notion supported by O'Connell et al. (2018) that students tend to perform better in theoretical courses compared to quantitative ones in any discipline of study.

The logistic regression model estimates of the study identified several academic performance features significant in determining students' performance in predicted variable, as similar to the deterministic studies of Ahmed et al. (2021), Ali et al., (2013), Ampofo & Osei-Owusu, (2015), Idoko & Emmanuel, (2015), Martin Sanz et al., (2017) and Tadese et al., (2022). This finding is aligned with the ML algorithms' identification of an optimal set of five predictor variables for the ML model through feature selection, feature engineering, and parameter tuning across various selected classifiers of the study.

The study findings also revealed that despite the poor model performance of the conventional logistic regression, as indicated by *Pseudo* R^2 and *Count* R^2 (model accuracy), its overall significance through square-chi (χ^2) was overshadowed by the superior performance metrics of ML evaluation. The ML approach exhibited higher accuracy, precision, recall, and f-measure values, further validated by the minimal False Positive (FP) values within the model confusion matrices of the selected machine learning algorithms for the study. This conclusion is in line with the findings of Amin et al. (2021), Hasegawa et al. (2020), Rajula et al. (2020) and Sheetal et al. (2023), who confirmed in their various studies the superiority of artificial intelligence (AI)-enhanced machine learning and deep learning techniques over traditional approaches in data analytics. Although their studies were in medicine, disease diagnosis, and material sciences, this study has reaffirmed these findings in the field of education data analytics with data limitation to cover the wider range of student's population in Nigeria.

5. Conclusion and Policy Implications

This paper contrasts the traditional and modern approaches to analysis, namely machine learning classification method with the analysis of student performance in one of the Economics based course, Econometrics, within Prince Abubakar Audu University in Nigeria. The research utilized 13 features based on Econometrics using a sample size of 897 students and their demographic and academic data. The traditional method was the logistic regression model which found the six significant predictor variables with limited accuracy. On the other hand, machine learning algorithms such as k-nearest neighbor, random forest, and support machine, were more successful than the logistic regression model. The accuracy and precision of SVM were the best. Due to the changing nature of the education processes, the research recommends that AI-enhanced machine learning methods should be highly utilized in education data analyses to deliver quality insights and make informed decisions about qualitative learning systems of higher institutions in Nigeria. Additional support of students in quantitative courses and recommendation of further studies in the longitudinal studies should be used to examine how predictors of academic performance change with age and their long-term influence on academic performance of the students with the machine learning analysis method.

Acknowledgement:
The author acknowledges the useful comments from the Editor and anonymous reviewers. Moreover, the researchers of this study are responsible for all errors.

Data Availability Statement:
Data is self-collected from readily available secondary sources of the institution of study and self-analyzed and provided at request..

Funding, if any:
No funding of any kind was provided for this research.

Conflict of Interest Disclosure Statement:
There is no conflict of interest among the authors of the study.

Ethical Approval:
Ethical approval has been obtained from the relevant forum of importance

References

Ahmed, J., Sule, B. O., Saka, S. A., Rasheedat, O. O., & Itopa, I. I. (2021). A Study on Factors that Influence Students' Academic Performance in Kogi State University Anyigba: A Binary Logistic Regression Approach. *International Journal of Mathematics and Statistics Studies*, 9(4), 14-16.

Alani, F. S., & Hawas, A. T. (2021). Factors Affecting Students' Academic Performance: A Case Study of Sohar University. *Psychology and Education*, 58(5), 4624-4635.

Alassaf, R. A., Alsulaim, K. A., Alroomi, N. Y., Alsharif, N. S., Aljubeir, M. F., Olatunji, S. O.,... & Alturayeif, N. S. (2018, November). Preemptive Diagnosis of Chronic Kidney Disease Using Machine Learning Techniques. In 2018 *International Conference on Innovations in Information Technology (IIT) (*pp. 99-104). IEEE.

Ali, S., Haider, Z., Munir, F., Khan, H., & Ahmed, A. (2013). Factors Contributing to the Students Academic Performance: A Case Study of Islamia University Sub-Campus. *American Journal of Educational Research*, *1*(8), 283-289.

Alwarthan, S. A., Aslam, N., & Khan, I. U. (2022). Predicting Student Academic Performance at Higher Education Using Data Mining: A Systematic Review. *Applied Computational Intelligence and Soft Computing*, *2022*.

Amin, M. N., Iqtidar, A., Khan, K., Javed, M. F., Shalabi, F. I., & Qadir, M. G. (2021). Comparison of Machine Learning Approaches with Traditional Methods for Predicting the Compressive Strength of Rice Husk Ash Concrete. *Crystals*, *11*(7), 779.

Ampofo, E. T., & Osei-Owusu, B. (2015). Students' Academic Performance as Mediated by Students' Academic Ambition and Effort in the Public Senior High Schools in Ashanti Mampong Municipality of Ghana. *International Journal of Academic Research and Reflection*, *3*(5), 19-35.

Gašević, D., Kovanović, V., & Joksimović, S. (2017). Piecing the Learning Analytics Puzzle: A Consolidated Model of a Field of Research and Practice. Learning: Research and Practice, 3(1), 63-78.

Hasegawa, D., Yamakawa, K., Nishida, K., Okada, N., Murao, S., & Nishida, O. (2020). Comparative Analysis of Three Machine-Learning Techniques and Conventional Techniques for Predicting Sepsis-Induced Coagulopathy Progression. *Journal of Clinical Medicine*, *9*(7), 2113.

Idoko, C. U., & Emmanuel, A. (2015). Teachers Effectiveness in Teaching Economics: Implication for Secondary Education. *International Journal of Innovative Research & Development*, *4*(2), 69-72.

Jiao, P., Ouyang, F., Zhang, Q., & Alavi, A. H. (2022). Artificial Intelligence-Enabled Prediction Model of Student Academic Performance in Online Engineering Education. *Artificial Intelligence Review*, *55*(8), 6321-6344.

Jhingan, M. L. (2017). *Macro-Economic Theory* (13th Ed). Flipcart: Virinda Publication Ltd.

Johnson, M., & Majewska, D. (2022). Formal, Non-Formal, and Informal Learning: What are They, and How Can We Research Them? *Journal of Education, Society & Multiculturalism*, *1*(2).

Kehinde, A. J., Adeniyi, A. E., Ogundokun, R. O., Gupta, H., & Misra, S. (2022). Prediction of students' performance with artificial neural network using demographic traits. In *Recent Innovations in Computing: Proceedings of ICRIC 2021, Volume 2* (pp. 613-624). Singapore: Springer Singapore.

Mankiw, N. G. (2020). *Principles of Economics* (9th Ed). Boston: Cengage Learning Publication Ltd.

Martin Sanz, N., Rodrigo, I., Izquierdo GarcÃa, C., & Ajenjo Pastrana, P. (2017). Exploring Academic Performance: Looking beyond Numerical Grades. *Universal Journal of Educational Research*, *5*(7), 1105-1112.

Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. *IEEE Access*, *9*, 140731-140746.

O'Connell, K. A., Wostl, E., Crosslin, M., Berry, T. L., & Grover, J. P. (2018). Student Ability Best Predicts Final Grade in a College Algebra Course. *Journal of Learning Analytics*, *5*(3), 167-181.

Olalekan, A. M., Egwuche, O. S., & Olatunji, S. O. (2020, March). Performance Evaluation of Machine Learning Techniques for Prediction of Graduating Students in Tertiary Institution. In *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)* (Pp. 1-7). IEEE.

Olatunji, (2022). Summer Machine Learning Classes: Machine Learning for All. YouTube Channel: Https://www.youtube.com/@asomai6923/videos for more videos.

Olatunji, (2023). ASOM Summer Machine Learning Training (Track 1 & 2): Machine Learning for All. YouTube Channel: Https://www.youtube.Ccm/@asomai6923/videos for more Videos.

Olatunji, S. O., Alsheikh, N., Alnajrani, L., Alanazy, A., Almusairii, M., Alshammasi, S.,& Alhiyafi, J. (2023). Comprehensible Machine-Learning-Based Models for the Pre-Emptive Diagnosis of Multiple

Sclerosis Using Clinical Data: A Retrospective Study in the Eastern Province of Saudi Arabia. *International Journal of Environmental Research and Public Health*, 20(5), 4261.

Omolewa, O. T., Oladele, A. T., Adeyinka, A. A., & Oluwaseun, O. R. (2019). Prediction of Student's Academic Performance Using K-Means Clustering and Multiple Linear Regressions. *Journal of Engineering and Applied Sciences*, *14*(22), 8254-8260.

Opesemowo, O. A. G., & Adekomaya, V. (2024). Harnessing Artificial Intelligence for Advancing Sustainable Development Goals in South Africa's Higher Education System: A Qualitative Study. *International Journal of Learning, Teaching and Educational Research*, *23*(3).

Ozkan, Z. A., Yu, Y., & Stelmach, E. (2023, June). Predicting Student Success in College Algebra Classes Using Machine Learning. In *2023 ASEE Annual Conference & Exposition*.

Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., & Fanos, V. (2020). Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. *Medicina*, *56*(9), 455.

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews), 40(6), 601-618.

Sheetal, A., Jiang, Z., & Di Milia, L. (2023). Using Machine Learning to Analyze Longitudinal Data: A Tutorial Guide and Best-Practice Recommendations for Social Science Researchers. *Applied Psychology*, *72*(3), 1339-1364.

Siemens, G., & Baker, R. S. D. (2012, April). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (Pp. 252-254).

Suresh, E. S. M., & Kumaravelu, A. (2017, June). The quality of education and its challenges in developing countries. In *2017 ASEE International Forum*.

Tadese, M., Yeshaneh, A., & Mulu, G. B. (2022). Determinants of Good Academic Performance Among University Students in Ethiopia: A Cross-Sectional Study. *BMC Medical Education*, *22*(1), 395.

Tsiakmaki, M., Kostopoulos, G., Koutsonikos, G., Pierrakeas, C., Kotsiantis, S., & Ragos, O. (2018, July). Predicting University Students' Grades Based on Previous Academic Achievements. In *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)* (Pp. 1-6). IEEE.

UNESCO, (2021). UNESCO Member States unite to increase investment in education. https://www.unesco.org/en/articles/unesco-member-states-unite-increase-investment-education

Wijaya, B. A., Kumar, V., Wau, B. F. J., Tanjung, J. P., & Dharshinni, N. P. (2022). Application of Data Mining Using Naive Bayes for Student Success Rates in Learning. *Jurnal Media Informatika Budidarma*, *6*(4), 1980-1987.

Wijaya, S. A., & Saputri, S. D. (2019). The Influence of Study Habits on Student Academic Achievement. Ekuitas*: Journal of Economic Education*, 7(2), 117-121.

Yağcı, M. (2022). Educational Data Mining: Prediction of Students' Academic Performance Using Machine Learning Algorithms. *Smart Learning Environments*, *9*(1), 11.

Yakubu, M. N., & Abubakar, A. M. (2022). Applying Machine Learning Approach to Predict Students' Performance in Higher Educational Institutions. *Kybernetes*, *51*(2), 916-934.

Yauri, R. A., Suru, H. U., Afrifa, J., & Moses, H. G. (2022). A Machine Leaning Approach in Predicting Student's Academic Performance Using Artificial Neural Network. *Journal of Computational and Cognitive Engineering*.

Zhao, L., Ren, J., Zhang, L., & Zhao, H. (2023). Quantitative Analysis and Prediction of Academic Performance of Students Using Machine Learning. *Sustainability*, *15*(16), 12531.

Zualkernan, I. (2021). Exploring Predicting Performance of Engineering Students Using Deep Learning. *International Association for Development of the Information Society*.