Journal of Artificial Intelligence and Metaheuristic

ISSN: 2833-5597 Vol: 11, No. 02, 2026 PP. 20–34
Article DOI: https://doi.org/10.54216/JAIM.110202

FULL LENGTH ARTICLE

A Hybrid Al-Biruni Earth Radius–Stochastic Fractal Search Optimized XGBoost Model for Accurate Student Performance Prediction

Abdelaziz A. Abdelhamid[1,*] **Amal H. Alharbi**[2]

[1] Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt
[2] Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

Emails: abdelaziz@cis.asu.edu.eg · ahalharbi@pnu.edu.sa

Received: December 09, 2025 Revised: February 10, 2026 Accepted: April 14, 2026 ⋆ Corresponding author

ABSTRACT

The ability to accurately predict student academic performance is now an established pillar of contemporary educational data mining, driven by the desire for data-driven interventions and personalized learning. This paper presents a mixed-methodology optimization and machine learning model that combines the Al-Biruni Earth Radius-Stochastic Fractal Search (BER-SFS) algorithm with eXtreme Gradient Boosting (XGBoost) to improve predictive accuracy and model stability in forecasting student performance. The proposed BER-SFS + XGBoost model is an optimized system that systematically optimises the hyperparameter based on dual exploration-exploitation mechanism to reduce the Mean Squared Error (MSE) of the baseline XGBoost model of 0.0226 to 0.00029 and the Root Mean Squared Error (RMSE) of the baseline XGBoost model to 0.1504 to 0.00194 and the coefficient of determination R2 of 0.84362 to 0.9019 Comparison to the other metaheuristics, including the Grey Wolf Optimizer (GWO), Particle Swarm Optimization (PSO) and Whale Optimization Algorithm (WOA) proved the superiority of the suggested hybrid model in all metrics of the evaluation. These findings support the potential to combine the dynamics of geometric exploration provided by BER with those of diffusion-based refinement offered by SFS to achieve high generalization and minimal bias. The implications extend beyond predictive analytics, demonstrating that the hybrid metaheuristic optimization approach can serve as a scalable, explainable method for adaptive educational systems, enabling early detection of at-risk learners and the development of data-driven academic support mechanisms.

Keywords: Educational Data Mining (EDM) ▪ Hybrid Metaheuristic Optimization Al-Biruni Earth Radius–Stochastic Fractal Search (BER–SFS) ▪ eXtreme Gradient Boosting (XGBoost) ▪ Student Performance Prediction.

1. INTRODUCTION

One of the most significant areas of contemporary educational research is students' academic performance. On other tiers, schools also face the challenge of ensuring that student performance is treated holistically rather than limited to grades or test scores [1]. Student performance measurement is not merely a combination of independent variables, including cognitive outcomes, student engagement and motivation, socio-economic status, and school support systems. Such variables are generally dynamic and create patterns that cannot be easily analysed using conventional methods. As described in the study by [2], student performance does not depend solely on academic ability but also on personal, environmental, and institutional factors that influence learning

© 2026 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

outcomes differently, and in most cases, nonlinearly. With such complexity, there is an increased demand for systematic, data-driven methods to analyze and predict student performance. The rise of digital learning programs, online learning systems, and academic management systems has led to a surge in educational content, commonly known as educational big data. This information has enormous potential to understand learner behavior, identify at-risk students, and improve institutional decision-making. Nevertheless, even in the wake of the opportunities it offers, educational data is characterized by high dimensionality, noisy missing values, and nonlinear relations between variables, making it very difficult to find a model and make predictions. The traditional statistical methods, despite their usefulness in inferential analysis, do not usually adequately describe these complex patterns of dependency . This restriction highlights the importance of more complex computational mechanisms that can acquire complex representations of student data and uncover hidden information. In this respect, Artificial Intelligence (AI) and Machine Learning (ML) paradigms have been agents of change in data analytics, enabling the modeling of complex systems and the generation of predictive knowledge [3]. Intelligent tutoring systems, automated evaluation, and adaptive learning have emerged in education due to advances in AI and ML. DL models, in particular, have demonstrated exceptional ability to process extensive, heterogeneous data from various education sources, such as student data systems, online education records, and evaluation systems [4]. Such models can handle multidimensional data to identify nonlinear relationships and make reliable predictions about student performance. On the other hand, regression-based machine learning models, such as Linear Regression, Support Vector Regression, and particularly tree-based ensemble regression models, are specifically designed to predict continuous target values, e.g., exam scores or grade point averages. The models enable researchers and educators to predict academic performance using a wide range of predictors, including behavioral, psychological, and socio-economic attributes [5]. These models prove invaluable for studying the fundamentals of student success, as they are interpretable and can be applied to related studies. Nonetheless, predictive models are prone to the quality and format of input data; thus, their quality and reliability vary dramatically. Educational data is heterogeneous, consisting of numerical, categorical, and text documents whose contributions to the performance forecast may vary. Outliers, missing values, irrelevant or redundant features, and so on can significantly corrupt the learning process unless properly addressed. Powerful preprocessing of data and feature engineering are thus very essential steps towards creating predictive models that are stable to use [6]. It involves cleaning, normalizing, and transforming the data to ensure consistency and relevance, as well as selecting the most informative set of features that are useful to the model's performance in a meaningful manner . This could cause them to overfit, make poor generalizations and exhibit poor predictive performance on unseen data. Dimensionality reduction and advanced feature selection techniques can be used to identify the most crucial variables with the most significant impact on student outcomes. Moreover, big data analytics enables researchers to uncover latent trends in large volumes of data, helping them better understand the connection between educational roots and performance indicators. Any efficient machine learning framework is based on optimization. Optimization algorithms are critical for improving the operational efficiency of predictive models by fine-tuning parameters and selecting the most pertinent features from vast amounts of data. Classical optimization methods tend to be unable to cope with the complexity and non-convexity of future ML tasks, and particularly high-dimensional search spaces. Such optimization algorithms have subsequently become more visible due to their ability to avoid local minima and efficiently search the global solution space using metaheuristic methods [7]. These algorithms are based on natural, physical, or evolutionary processes and have demonstrated significant efficiency in feature selection and hyperparameter optimization [8]. The objectives of feature selection are to determine the subset of variables that best predicts and, at the same time, minimizes model complexity and computational costs of the model inferred by that subset of variables. Hyperparameter optimization, in turn, aims to optimize the set of learning algorithm parameters, such as learning rate, tree depth, and regularization, to ensure the learning algorithm's functioning is robust and stable. More sophisticated ML models, such as the eXtreme Gradient Boosting (XGBoost) algorithm, have also been shown to be highly effective when combined with metaheuristic optimization algorithms. To illustrate, an Al-Biruni Earth Radius (BER) optimizer with the Stochastic Fractal Search (SFS) algorithm has been shown to add value to the process of determining optimal feature subsets and parameter settings. These hybrid optimization methods provide an effective and accurate way to maximize predictive performance compared with traditional grid or random search methods, which may be computationally expensive and often get stuck in local optima. It is due to the potent combination of a gradient-boosting system, XGBoost, and the exploratory capabilities of metaheuristic algorithms that generate models not only of superior accuracy but also of superior efficiency, and that can be used in educational data mining systems. The primary objectives of this paper will therefore be as follows:

1. To investigate the application of regression-based machine learning models, with a particular focus on XGBoost, for predicting student performance indices in an educational context.

2. To investigate the further improvement of the predictive power of XGBoost with the addition of an optimization algorithm called BER-SFS, to optimize the feature selection and hyperparameter optimization of the model.

3. To compare the performance of the optimized XGBoost model with other machine learning and optimization methods, it is essential to identify the advantages of the proposed model.

4. To determine the usefulness of metaheuristic optimization methods in enhancing the accuracy, consistency and interpretability of student performance prediction models.

5. To offer an in-depth analysis of the model performance in terms of various statistic measures and visualization to bring insights about the practical use of the model.

The remainder of this paper will be organized as follows: Section 2 will provide a critical survey of historical methodologies that have been applied thus far and their flaws in predicting student performance. Section 3 describes datasets used in this study, the machine learning models developed, and the optimization methods used. The analysis of the experimental results is presented in Section 4 to demonstrate the efficiency of the proposed approach and compare it with existing approaches. Section 5 explains the results in detail, extracting their implications, and indicating the spheres of practical application and potential restrictions of the suggested model. Finally, the conclusion and recommendations for future research are presented in Section 6, and it is stated that optimized ML models, enabled by efficient optimization algorithms, can shift the focus of educational analytics and decision-making. This theoretical framework thus provides a broad background on the same, along with the scope, significance, and value of the present study.

2. LITERATURE REVIEW

Educational mining has developed into a valuable tool for identifying latent trends in educational data nd predicting academic student performance. According to the research by , a new machine learning-based model was proposed to estimate undergraduate students' final exam grades based on their midterm results. On a dataset of 1854 students who enrolled in the Turkish Language-I course at a Turkish university in the 20192020 fall semester, the paper assessed the performance of a variety of machine learning algorithms, including random forests, support vector machines, logistic regression, and k-nearest neighbors. The results showed that the model had an accuracy rate of 70-75% based on three primary parameters: midterm grades, department data and faculty data. Predictive models are essential in higher education learning analytics, as they enable early detection of students at high risk of failure and the selection of appropriate machine learning strategies. With the rise of technology, student performance prediction has become a significant research field, and data mining techniques are instrumental in analyzing educational data. A study by compared several resampling algorithms, including Borderline SMOTE, SMOTE-ENN, SVM-SMOTE, and SMOTE-Tomek, to achieve good predictive performance. The results, using classifiers such as Random Forest, K-Nearest Neighbors, and XGBoost with Random hold-out and 5-fold Shuffle cross-validation, revealed that balanced data improved classifier accuracy. The best method was SVM-SMOTE, and the highest predictive accuracy was achieved with Random Forest. Similarly, the article by [9] evaluated students' performance prediction during online learning following the COVID-19 pandemic. Based on the results of the digital electronics lab sessions, 86 statistical features were filtered and extracted from the data to identify significant predictors. Five classifiers were tested, including Random Forest, Support Vector Machine, Naive Bayes, Logistic regression, and Multilayer Perceptron. The Random Forest model scored 97.4%, while the other models scored lower across various validation cases. On the whole, the two works emphasize the importance of data set balancing and the use of strong classifiers, such as Random Forests, in educational data mining to improve prediction accuracy. The COVID-19 pandemic has dramatically changed the education sector, leading to widespread adoption of online learning platforms. As found in [10], online classes require student engagement to be as effective as traditional in-person learning. The study introduced deep learning algorithms that can identify and analyze students' emotions (anger, happiness, sadness, and surprise) in real time. Based on face landmark recognition, emotion recognition, and survey data gathered during a one-hour session, the researchers developed a process to estimate a mean engagement score. This online learning platform improves interactivity and the quality of online learning. Moreover, Massive Open Online Courses (MOOCs) are becoming increasingly successful, measured by student satisfaction rather than traditional measures such as completion rates. As noted by , increased satisfaction not only leads to greater reach and brand strength among institutions but also generates potential revenue. Based on Moore's theory of transactional distance, the paper used supervised machine learning, sentiment analysis, and hierarchical linear modeling to examine 249 MOOCs and 6,393 student feedback items. The results highlight that engagement and satisfaction among learners are essential to the success of online education environments. These findings indicated that instructor quality, course content, assessment strategies and schedule are essential contributors to satisfaction, but course structure, course duration, interaction and perceived difficulty are not. The research will add value to the existing literature by identifying the key variables at the learner and course levels that affect satisfaction, and by describing the implications for MOOC instructors and developers. Emergency remote learning, now popular in higher education because of COVID-19, has sparked debate over what makes students satisfied with this learning model. The article by [11] aimed to identify variables predicting satisfaction among 425 undergraduate students at a self-funded university in Hong Kong, where Moodle and Microsoft Teams are the leading learning platforms. The researcher compared the results of multiple regression models and machine learning methods and found that random forest recursive feature elimination and elastic net regression were more accurate, with an explained variance of 65.2 percent. The findings revealed that overall satisfaction with emergency remote learning was low (4.11 out of 7), despite students' technological abilities and access to devices and Wi-Fi. Preference was most likely predicted by face-to-face learning, followed by instructor efforts, appropriately adjusted tests, and the perception of good online learning delivery. These results show that assessment plans should be revised and enhanced, and that classes should be offered in an interactive, organized format, with the learning culture and program in mind. Massive open online courses (MOOCs) feature extensive open-access materials and facilitate self-directed learning, but it isn't easy to measure students' performance on these platforms. The report by [12] showed that applying big data and artificial intelligence to MOOCs enhances understanding of educational data and the learning processes of both students and teachers. This study has developed two deep neural network models: one that predicts students' performance outcomes based on video-watching behaviors, and another that predicts students' ability to answer test questions successfully after completing the exercises. Data from two courses on the National Tsing Hua University MOOCs platform were used to

validate these models. The video-based model was quite accurate in predicting performance, and the model used as an exercise was very effective at predicting outcomes with respect to conceptual knowledge. The use of artificial intelligence (AI)-based applications is crucial for identifying failing students early and enabling instructors to address the issue in time and help students attain academic success. With reference to [13], early prediction of academic performance helps institutions of higher education in applying targeted enrollment strategies and interventions to advance learning outcomes. The paper aimed to predict the cumulative grade point average (CGPA) of postgraduate students, using a sample of 635 master's students at a Malaysian private university. The six machine learning models were tested using the coefficient of determination, mean square error, and mean absolute error. The ANN had the best predictive performance, with an error of 0.08 points and an R^2 of 0.89. Conversely, the Gaussian Process Regression model explained 71% of the variance, with training and test errors of 0.095 points. The research implied that the involvement in the research and living conditions should be introduced as additional variables to improve the model's accuracy, and that a more extensive study with a broader scope of data should be considered in the future. In the same vein, the article by [14] highlighted the radical impact of artificial intelligence and machine learning in education. They conducted research on AI and ML awareness, best practices, and attitudes in higher education, using survey data from 103 Serbian students and statistical analysis. The results indicated that these technologies foster skill development and collaboration, as well as exposure to research settings, highlighting their potential to personalize learning and maximize educational outcomes. Performance Factors Analysis (PFA) is an established Knowledge Tracing technique that has been applied in adaptive learning systems and shown excellent predictive reliability relative to other techniques. As mentioned in the article by [15], previous improvements to PFA primarily focused on students' behavior to enhance pedagogical outcomes. New technical improvements have been created with the integration of machine learning, however. The researchers presented an ensemble-based solution with the Random Forest, AdaBoost, and XGBoost to improve the classical PFA model. Empirical assessments across three datasets found that the scalable XGBoost model was far superior to the original PFA, yielding significant gains in predictive power. Likewise, the study by also highlighted the importance of forecasting the most significant academic outcomes, such as grade point average (GPA), retention, and degree completion, which should enable the implementation of early interventions in higher education. Conventional predictive models usually assume linear relationships between variables and are therefore not flexible. To solve this, the researchers used a multilayer perceptron artificial neural network with the backpropagation algorithm to predict the academic performance of 655 higher education students. The model had high predictive power, with learning strategies as the best predictors of GPA, coping strategies as the best predictors of degree completion, and background factors as the best predictors of risk of dropout. These results indicate that modern machine learning methods have the potential to improve predictive performance and support the use of data in educational decision-making. Recent studies point to the growing nature of machine learning and artificial intelligence in improving education by predicting performance, analyzing engagement, and preventing dropouts. According to [16], transfer learning can enhance the predictive accuracy of student performance models by leveraging domain-specific knowledge, especially when data is limited. In their research, they used deep neural networks on higher education data and found that transfer learning helps identify students at high risk of failing. A simplified training vector-based support vector machine for effectively predicting marginal and at-risk students was introduced in the study by [17]. The model reduced training time by eliminating unnecessary training vectors without compromising predictive power, which exceeds 91 percent for both marginal and at-risk students, and therefore serves as a powerful early intervention tool. As stated in , Virtual Learning Environments (VLEs) play an essential role in learning monitoring and in enhancing academic performance. The study employed tree-based and neural network models to demonstrate that the frequency of online resource use is a strong predictor of student success, findings further bolstered by a case study of master's students at the University of Salamanca. To help young learners comprehend the concept of artificial intelligence at an earlier stage, [18] created an interactive K-12 course that introduced students to machine learning, neural networks, and ethics using Google's Teachable Machine. Using case studies of 108 students, it was found that participants learned the basic concepts of ML and created working classification models, independent of demographic factors. A longitudinal study by [19] of more than 400 publications on AI and deep learning in education from 2000-2019 found a changing trend in the field. The researchers observed a decline in traditional studies in instructional design. They proposed advancing research on learning analytics and student profiling, which focus on the active development of AI-based pedagogy. Online learning is also predominantly affected by academics' emotions. According to [20], using deep learning and aspect-based sentiment analysis on data collected from a Massive Open Online Course (MOOC) provided insight into how students felt about it. The findings showed that positive emotions were greatest at the onset of courses, then leveled off, and that teacher and course dimensions were more influential than student characteristics. Resolving the issue of student dropout, paper [21] suggested a data structure based on the information provided by 19 Brazilian schools to train dropout prediction models. With high performance scores, such as 95% accuracy and 93% recall, the study also revealed differences across levels of education and possible reasons behind dropout, which can be put into practical implementation. Equally, in a comparable study, the authors of the article by [22] tested dropout rates in the first year of engineering in Chile through multiple machine learning models. Decision trees trained with gradient boosting were found to be the most accurate, and results showed that higher mathematical test scores decreased the risk of dropping out. In contrast, higher language test scores inversely predicted the likelihood of dropping out. Lastly, in e-learning, where data are limited, the paper by discussed dropout prediction. The study, conducted over four years of academic testing, showed test accuracy of 77-93 percent and highlighted the need for rigorous performance assessment to eliminate bias from small data sets. All these studies collec-

tively show that research on educational data mining is being shaped by a range of machine learning methods, including conventional classifiers, deep learning, sentiment analysis, and transfer learning. These approaches, as described in Table 1, can advance predictive modeling in the fields of MOOCs, VLEs, online learning, and dropout prediction, and can offer significant data-driven educational enhancement tools.

3. MATERIAL AND METHODS

The overall methodological framework adopted in this study is illustrated in Figure 1. The workflow outlines the complete process from data acquisition to model evaluation, beginning with the *Student Performance Dataset*, which undergoes comprehensive preprocessing steps including handling of missing data, removal of null values, and normalization to ensure data quality and consistency. The processed dataset is then partitioned into training (80%) and testing (20%) subsets for model development. Baseline machine learning models such as *XGBoost*, *Support Vector Regression*, *Decision Tree*, and *Gradient Boosting Regressor* are initially trained, followed by optimization through the *Al-Biruni Earth Radius (BER)* algorithm integrated with the *Sine–Cosine Function Search (SFS)* framework. Additional metaheuristic optimization techniques—*BERSFS*, *GWO*, *PSO*, and *WOA*—are employed to enhance model performance. Finally, the optimized models are evaluated using multiple performance metrics to assess prediction accuracy and robustness.

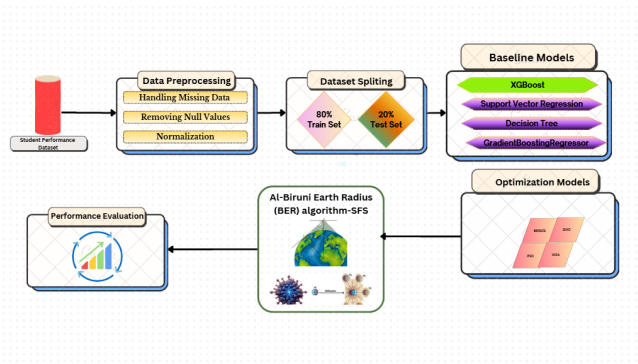

Figure 1. Workflow of the Proposed Methodology for Student Performance Prediction

3.1 Dataset Description

The Student Performance Dataset is an artificial dataset designed to examine the primary variables that determine students' academic performance. It has a total of 10,000 student records, where each record represents a single learner and contains information pertinent to a set of predictor variables and a single target variable, the Performance Index. This data will provide a methodological basis for research on the relationships among behavioral, academic, and lifestyle variables that affect general academic performance. The predictor variables in the dataset include the following and will be treated as independent inputs in the analysis. Each feature makes its own contribution to the determinants of academic performance. The dependent or target variable in this dataset is the performance index, which measures an individual student's overall academic performance. The information is presented in Table 3. The dataset would help investigate the relationships between the predictor variables and the target variable, enabling researchers and data analysts to determine the extent to which study habits, previous performance, extracurricular participation, and sleep patterns are positively related to overall academic success. Besides, it can serve as a powerful testing and benchmarking tool for machine learning models, particularly those aiming to predict student performance using regression-based methods. The data is well-structured and can be used for model validation, feature selection, and optimization experiments. It is necessary to add that the Student Performance Dataset is entirely artificial and was created solely for demonstration and experimentation. Thus, the relationships and correlations among the variables may not reflect the pattern of student performance in the real world. Nonetheless, the data set provides a controlled testing environment for algorithms, enabling the evaluation of predictive models and the analysis of feature significance in educational data mining. The performance of the aggregate students provides priceless statistics on the variability and central tendency of overall performance across the whole dataset. The Performance Index is roughly normally distributed, with the most frequent in the middle ranges of performance and less frequent at the higher and lower ends, as demonstrated by Figure 2. This tendency indicates that the data are evenly distributed and that the performance measure is appropriate for representing and capturing a range of student performance.

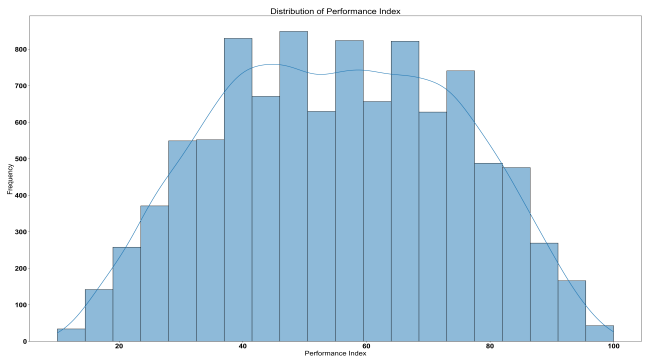

Figure 2. Distribution of Performance Index

To obtain a detailed picture of the underlying trends in the data, one will need to analyze the distributions of each predictor variable. The distribution plots of the main independent variables — Hours Studied, Previous Scores, Extracurricular Activities, Sleep Hours, and Sample Question Papers Practiced —are presented in Figure 3. These visualizations show how variable and dispersed each feature can be, making them valuable markers of students' behavior and study habits. The figure, for example, demonstrates that participation in extracurricular activities is highly binary. On the contrary, variables related to the study, i.e., hours studied and frequency of practice, are comparatively homogeneous. Such an exploratory analysis is an initial step toward establishing relationships among these factors and their combined impact on the Performance Index.

3.2 Data Preprocessing

Data preprocessing is an essential step in preparing raw data for machine learning. The accuracy, stability, and interpretability of predictive models depend directly on the quality of preprocessing. In this paper, several preprocessing

Table 1. Summary of key literature: focus area, methodology, and findings.

Reference	Focus Area	Methodology	Key Findings and Contributions
[23]	Imbalanced performance prediction	Resampling (Borderline SMOTE, SMOTE-ENN, SVM-SMOTE, SMOTE-Tomek); RF, KNN, XGBoost; 5-fold validation	Balanced datasets improve accuracy; SVM-SMOTE most effective; Random Forest achieved highest predictive accuracy.
[9]	Online engagement → performance	86 statistical features; RF, SVM, NB, LR, MLP	RF achieved 97.4% accuracy; feature engineering critical for predicting online lab performance.
[10]	Online engagement detection	Deep learning; facial landmarks and emotion recognition	Real-time emotional analytics enhance engagement and digital learning quality.
[24]	MOOC satisfaction drivers	Gradient boosting, sentiment analysis, hierarchical linear modeling (249 MOOCs, n=6393)	Instructor quality, content, and assessment most influence satisfaction; course duration and structure less significant.
[11]	Emergency remote learning satisfaction	Regression & ML with recursive feature elimination; elastic net ($R^2 = 0.652$)	Moderate satisfaction (4.11/7); strongest predictors: instructor effort, face-to-face preference, and assessment suitability.
[12]	MOOC performance prediction	Two deep neural network models (video behavior & exercise-based)	Accurately predicts student outcomes; early identification of underperforming learners.
[13]	Postgraduate CGPA prediction	Six ML models (ANN, GPR, etc.)	ANN achieved best performance ($R^2 = 0.89$); integrating contextual features improves accuracy.
[14]	AI/ML in higher education	Survey of 103 Serbian students; statistical analysis	AI/ML fosters collaboration, skill development, and research access in higher education.
[15]	Knowledge Tracing (PFA variants)	Ensemble models (RF, AdaBoost, XGBoost)	XGBoost significantly improves prediction accuracy over traditional PFA.
[25]	Academic outcome prediction	Multilayer perceptron with backpropagation	Learning strategies predict GPA; coping strategies predict completion; background predicts dropout risk.
[16]	Transfer learning for performance	Deep neural networks using cross-course transfer	Effective at-risk prediction using related datasets; validates transfer learning in education.
[17]	Early risk detection (RTV-SVM)	Reduced training-vector SVM	59.7% reduction in training vectors; maintains 92–93% accuracy for at-risk and marginal students.
[26]	VLE usage → success	Tree-based models and ANN on public dataset + case study	Frequency of VLE access is a key predictor of academic success; validated in master's cohort.
[18]	K–12 ML education	Interactive course (Teachable Machine project)	Students grasped ML concepts and built classifiers regardless of gender or modality; strong motivational outcomes.
[19]	20-year AI-in-education trends	Content analysis of 400+ research papers (2000–2019)	Trend shift from tech-based design to student profiling and analytics; mapped evolution of AI in education.
[20]	Academic emotions in MOOCs	BERT-based aspect sentiment analysis	Academic emotions improve early; identified teacher, course, and platform-level affective patterns.
[21]	Explainable dropout prediction	Classifiers + model explainability (19 Brazilian schools)	Achieved AUPRC 89.5%, Precision 95%, Recall 93%; identified causes of dropout for targeted intervention.
[22]	First-year dropout (engineering)	Eight ML models; site-specific vs. pooled	Site-specific models outperform pooled; GBDT best; math score lowers dropout risk.

Table 2. Description of Predictor Variables

Feature Name	Description
Hours Studied	Refers to the amount of time that a given student spent studying during a specific time. This variable is one of the primary predictors of effort and academic engagement in the study.
Previous Scores	Denotes the scores obtained by students in their previous assessments or tests. It acts as a proxy for prior academic performance and learning consistency.
Extracurricular Activities	Indicates whether a student participates in extracurricular activities, represented as a categorical variable with two possible values: *Yes* or *No*. This feature helps assess the balance between academic and non-academic commitments.
Sleep Hours	Denotes the mean number of hours of sleep the student gets each day. This is a potentially effective factor since adequate rest is a key element of cognitive performance and concentration.
Sample Question Papers Practiced	Measures the number of practice papers that the student has done before the assessment. This variable indicates how prepared the student is for exams and how well he/she understand test formats.

Table 3. Description of Target Variable

Variable Name	Description
Performance Index	The target variable is each student's academic performance. The index ranges from 10 to 100, with higher values indicating improved results. It is a composite quantitative indicator of student success, influenced by the predictor variables listed in Table 2. The values are simplified and rounded to the nearest number to make them simpler and more homogeneous.

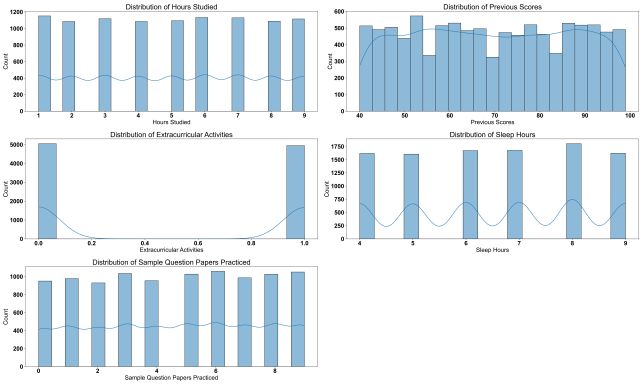

Figure 3. Distribution of Predictor Variables

functions were used to ensure that the Student Performance Dataset was free of inconsistencies, bias, and noise. These processes are quality control, imputation and encoding, scaling, correlation screening and assisted feature design. Preprocessing phase one entailed quality control to ascertain the integrity and reliability of the data. These involved confirming completeness, consistency and logical validity of any records. Redundant and abnormal records, e.g., implausible study hours or negative performance indices, were identified and eliminated. Outliers were identified using statistical techniques, such as summary and boxplots, and missing values and data distribution were evaluated to ensure the dataset was ready. This was done to ensure that subsequent analyses used correct and representative data. After quality control, data values were either missing or incomplete, and imputation techniques were used to fill the gaps. For numerical values (i.e., Hours Studied, Sleep Hours, and Previous Scores), missing values were imputed using the feature's mean or median, depending on its skewness. Mode imputations have been used to retain the most frequently occurring category for categorical variables such as Extracurricular Activities. Categorical variables were encoded as numbers to prepare the dataset for machine learning models. Extracurricular Activities were binary coded with 1 indicating Yes and 0 indicating No. The change made it easy to incorporate into regression-based algorithms, such as XGBoost and Support Vector Regression, that require numerical input features. Function scaling was applied to normalize the numerical variables and ensure that all features contributed equally to model training. Standardization per feature was adopted, and each variable was transformed to have a mean of zero and a standard deviation of one. The approach ensures that models that are sensitive to feature magnitudes are not biased toward larger-valued attributes, as typified by Support Vector Regression (SVR) and Gradient Boosting. The scaling parameters were trained only on the training data to avoid data leakage, and the same transformation was then applied to the test data. Mathematically, the standardized value z_i for each feature x_i was computed as:

$$z_i = \frac{x_i - \mu_x}{\sigma_x} \tag{1}$$

where μ_x and σ_x represent the mean and standard deviation of the feature, respectively. Correlation analysis was performed to eliminate redundancy and reduce multicollinearity among features. The Pearson correlation coefficient was estimated for all combinations of numerical variables to assess the linear relationship between them. Strongly correlated variables ($r > 0.85$) were identified, and redundant variables were removed to improve model generalization and minimize overfitting. Also, they visualized the predictor dependency structure using correlation heatmaps, which allowed them to retain only the most significant variables—Hours Studied, Previous Scores, and Sample Question Papers Practiced—for further analysis in downstream modeling. In this work, both domain knowledge and statistical evidence were used to make feature engineering and preprocessing choices. From the exploratory analysis, it was found that behavioral characteristics (e.g., hours of study, sleep duration) had nonlinear relationships with performance outcomes, which explains why nonlinear learners such as XGBoost and SVR should be included. To make the model interpretable, features were retained only if they made a meaningful contribution to academic prediction. The preprocessing pipeline was implemented to ensure reproducibility and future scalability, in case new data sources or features are introduced. All in all, the preprocessing phase produced a clean, standardized, and semantically consistent dataset that served as the basis for robust, comprehensible machine learning models.

3.3 Al-Biruni Earth Radius (BER) Algorithm

Unlike the theory in the 19th century, it relies on a methodology of the Persian polymath Al-Biruni in the 11th century to estimate the radius of the Earth. Al-Biruni cleverly used the calculations of the radius of the Earth as a result of geometrical observations made on a mountain peak. The first method he used to measure the mountain's height was to measure the angle from two points on level ground to the mountain's summit. The height of the mountain, h, is obtained by using the following equation:

$$h = \frac{d\left(\tan\theta_1 \times \tan\theta_2\right)}{\tan\theta_2 - \tan\theta_1} \tag{2}$$

Having known the height of the mountain, Al-Biruni estimated the angle of depression between the height of the mountain and the horizon. He then calculated the radius of the Earth, R, using trigonometric relationships as:

$$R = \frac{h\cos\alpha}{1 - \cos\alpha} \tag{3}$$

The geometric principle of this method is depicted in Fig. 4.

The specified optimization algorithm is rooted in the concepts Al-Biruni developed for exploring and exploiting in the context of cooperative optimization. This mechanism is based on swarm intelligence, in which agents (individuals) collaborate and communicate to accomplish a common task. Swarm organisms, such as ants and bees, are groups that form naturally and can explore the environment in parallel and communicate with one another. The BER algorithm mimics this joint move by dividing its population into two sets, balancing exploration and exploitation to discover new regions and optimize existing solutions. The design aims to prevent local optima from forming prematurely, since every subgroup is complementary to the others in the optimization process. Exploration is dynamically strengthened if the algorithm fails to achieve improved fitness after several rounds. Optimization algorithms aim to determine the best solution to a problem subject to

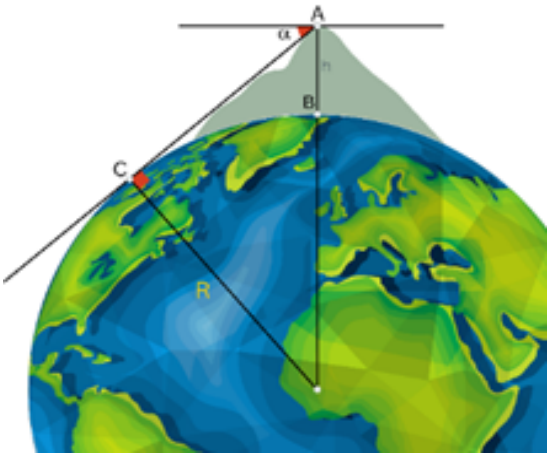

Figure 4. Calculation of Earth's radius based on the Al-Biruni principle.

given constraints. In BER, all the objects of the population are described as vectors of a d-dimensional nature:

$$\vec{Q} = \{Q_1, Q_2, \ldots, Q_d\} \in \mathbb{R}^d$$

where Q_i denotes a feature or parameter, and d represents the problem's dimensionality. The fitness function P evaluates the quality of each candidate solution. The goal of the optimization process is to maximize this fitness by iteratively updating the population vectors.

The algorithm requires initialization of parameters, including the population size, the upper and lower bounds of the variables, and the maximum number of iterations (Max_{iter}). A random population is generated at the beginning, and fitness values are computed for all individuals.

3.4 Exploration Operation

Exploration focuses on identifying promising regions in the search space and avoiding premature convergence. This stage allows the algorithm to investigate potential solutions near interesting regions using the following update equations:

$$Q(t+1) = Q(t) + D(2r_2 - 1) \tag{4}$$
$$D = r_1(S(t) - 1) \tag{5}$$

Here, $S(t)$ is the current solution vector, and D defines the search radius around each candidate. The coefficients r_1 and r_2 are random parameters used to enhance stochastic movement across the search space. The distance factor r is calculated as:

$$r = h\frac{\cos(x)}{1-\cos(x)} \tag{6}$$

where x is an angle between 0° and 180°, and h represents the metaphorical mountain height controlling exploration range.

3.5 Exploitation Operation

Exploitation aims to refine existing solutions and enhance convergence around optimal regions. It updates solutions using the following equations:

$$Q(t+1) = r^2(Q(t) + D) \tag{7}$$
$$D = r_3(L(t) - Q(t)) \tag{8}$$

where $L(t)$ represents the best-known solution, and r_3 is a random coefficient vector. A local search is then performed around the best solution using:

$$Q'(t+1) = r(Q^*(t) + k) \tag{9}$$
$$k = 1 + \frac{2t^2}{Max_{iter}^2} \tag{10}$$

If no improvement in fitness occurs over two consecutive iterations, mutation is applied as:

$$Q(t+1) = kz^2 - h\frac{\cos(x)}{1-\cos(x)} \tag{11}$$

Where z is a random number between 0 and 1, the elitism mechanism ensures that the best solutions are preserved across iterations, maintaining stability while enhancing diversity. The dual exploration–exploitation strategy of BER enables it to achieve efficient global optimization while avoiding local stagnation.

Algorithm 1 Al-Biruni Earth Radius (BER) Optimization Algorithm

```
Initialize population S_i (i = 1, 2, ..., d), maximum iterations Max_iter, and fitness function F_n
Initialize BER parameters
Set iteration counter t = 1
Compute initial fitness F_n for all S_i
Identify best solution S*
while t ≤ Max_iter do
    for each solution in the exploration group do
        Compute r = h cos(x)/(1−cos(x))
        D = r_1(S(t) − 1)
        S(t+1) = S(t) + D(2r_2 − 1)
    end for
    for each solution in the exploitation group do
        Compute D = r_2(L(t) − S(t))
        S_1(t+1) = r^2(S(t) + D)
        k = 1 + 2t^2/Max_iter^2
        S_2(t+1) = r(S*(t) + k)
        Compare S_1(t+1) and S_2(t+1), select best S*
        if no fitness improvement for two iterations then
            Mutate solution: S(t+1) = kz^2 − h cos(x)/(1−cos(x))
        end if
    end for
    Update fitness F_n for all S
end while
return best solution S*
```

The BER optimization process consists of exploration and exploitation, as summarized in Algorithm 1, to make the search process diverse and ensure convergence to the global optimum. The Stochastic Fractal Search (SFS) algorithm is based on the mathematical concept of fractals and their self-

similarity. It builds upon the original Fractal Search (FS) algorithm. It adds stochastic diffusion-based operations based on the concept of Diffusion-Limited Aggregation (DLA), which form fractal-shaped structures in the search space. SFS dynamic propagation of the best candidate solution is through a diffusion process that maintains a balance between exploration and exploitation. The diffusion process about the best solution (BP) is depicted in Figure 5 where clusters are created around the best solution, including BP_1, BP_2, BP_3, BP_4, and BP_5 that help in steering the search to the global optimum.

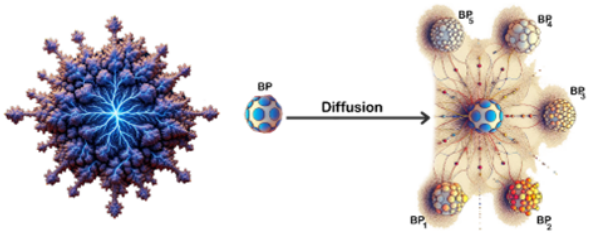

Figure 5. Illustration of the diffusion process in the Stochastic Fractal Search (SFS) algorithm.

This diffusion-based search and clustering will enhance search variety and avoid early convergence. Inclusion of random diffusion, together with local search, gives SFS a better ability to optimize the entire process of metaheuristic algorithms and also provides a stable convergence path for high-dimensional optimization problems.

3.6 Machine Learning Models

The study used numerous regression-based machine learning algorithms to predict students' performance, and each was chosen for its ability to learn complex relationships and handle diverse data types. Examples of models include eXtreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), Decision Tree (DT), and Gradient Boosting Regressor (GBR). The models are also suitable for continuous prediction problems, applied to predict academic performance indices, and, overall, provide a trade-off between interpretability, robustness, and predictive accuracy. Below are subsections that summarize the models, their theoretical foundations, and their applications to educational data mining.

eXtreme Gradient Boosting (XGBoost) XGBoost is an extreme gradient boosting algorithm that is scalable. It is a collective learning method that constructs a sequence of weak prediction models, typically decision trees. The successive models are trained to minimize the residual error of the previous model, thereby improving overall predictive performance. XGBoost offers several improvements over traditional boosting approaches, including regularization, tree pruning, and handling sparse data. Its objective role works in the following manner:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \Omega(f_k) \tag{12}$$

Where $l(y_i, \hat{y}_i^{(t)})$ represents the loss function measuring prediction error, and $\Omega(f_k)$ is a regularization term that controls model complexity, helping to prevent overfitting. XGBoost is flexible and can effectively handle missing data, nonlinear relationships, and large-scale data through parallel computation. Moreover, it combines first- and second-order gradients during optimization, which allows faster convergence and better generalization. XGBoost is also particularly effective for predicting student performance, as it can capture subtle interactions among variables such as study hours, past scores, and lifestyle factors, and predict high student performance.

Support Vector Regression (SVR) Support Vector Regression (SVR) is a variant of the Support Vector Machine (SVM) used for continuous prediction. The working principle of SVR is to map the input data into a high-dimensional feature space using kernel functions, and then fit a linear model in the transformed space. The algorithm is intended to reduce the prediction error to the required tolerance, represented by the variable ε, thereby introducing a trade-off between model complexity and accuracy. The optimization problem of SVR is:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{13}$$

Subject to:

$$\begin{cases} y_i - (w \cdot x_i + b) \leq \varepsilon + \xi_i \\ (w \cdot x_i + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Here, w represents the weight vector, b the bias term, C the regularization parameter, and ξ_i, ξ_i^* the slack variables representing deviations beyond the margin ε. The SVR will be able to learn a nonlinear relationship between the predictors and the target variable by using the kernel functions (radial basis function (RBF) or polynomial kernels). This makes it particularly appropriate for educational datasets in which relationships between qualities, including research time and performance, are not necessarily linear. SVR is computationally intensive but robust to outliers and offers high precision.

Decision Tree (DT) The Decision tree (DT) algorithm is a non-parametric supervised learning algorithm in which data sets are classified into subsets based on feature values, forming a tree-like structure with decision nodes and terminal leaves. Internal nodes are associated with decision rules specified by a feature threshold, and leaf nodes are associated with the predicted value of the target variable. The algorithm is a recursive breakdown of the data using impurity reduction measures (Mean Squared Error (MSE) or Mean Absolute Error (MAE) in a regression task. The MSE of a split of an experiment is determined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{14}$$

Decision Trees are easy to understand and interpret, and they are instrumental in educational analytics, where they can be equally crucial for gaining insights into the impact of individual characteristics, e.g., the time spent studying or participating in extracurricular activities, as well as for their high predictive accuracy. Nevertheless, the trees are likely to overfit without pruning, leading to poor generalization. To cope with this, various parameters, including the maximum depth, minimum sample split and pruning methods, are used to strike a balance between bias and variance.

Gradient Boosting Regressor (GBR) Gradient Boosting Regressor (GBR) is a powerful ensemble model that builds

predictive models sequentially. It sequentially integrates weak learners (mostly decision trees) so that each learner improves the errors of the prior ensemble. The iterative algorithm reduces a differentiable loss through the principles of gradient descent. The equation of the model update is obtained as follows:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x) \tag{15}$$

Where $F_m(x)$ is the ensemble model after m iterations, η is the learning rate controlling the contribution of each weak learner, and $h_m(x)$ denotes the newly added regression tree. The negative gradient of the loss function serves as the residual, guiding subsequent model adjustments. GBR is highly flexible and precise, and can handle nonlinear relationships and variable interactions. The performance, however, relies heavily on the tuning of its hyperparameters, including the learning rate, number of estimators, and tree depth. GBR can be effectively applied to educational data to identify subtle dependencies between behavioral and academic characteristics, yielding very accurate predictions of student performance. All models offer unique benefits for predicting academic performance. XGBoost and GBR are boosting-based approaches that are particularly good at reducing bias and variance through ensemble learning and regularization. SVR provides accurate regression properties for nonlinear mappings; as such, it is suitable for datasets with smooth, continuous relationships. Decision trees are easier to interpret and more transparent, but they provide educators with an overview of the factors with the most significant impact on performance. The summary of the models used in this study is provided in Table 4. Overall, the models offer a wide range of methodological approaches for predicting student performance, including linear and nonlinear methods. XGBoost and GBR are better at making predictions, SVR is more theoretically rigorous at continuous regression, and Decision Trees are easier to interpret. By comparing and contrasting these models, the proposed study will define the most efficient algorithmic approach to educational prediction, enabling data-driven interventions and academic decision-making.

3.7 Evaluation Metrics

To assess the prediction and reliability of the regression models, nine standard statistical measures of assessment were taken; Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Coversation Coefficient (r), Coefficient of Determination (R2), Relative root mean squared error (RRMSE), Nash -Sutcliffe Efficiency (NSE), and Willmott index of Agreement (WI). These metrics are used to determine the consistency, bias, and accuracy of observed and predicted values. The measures and mathematical forms of the measures are also summarized in Table 5.

The metrics provide a variety of ways of modelling performance. When MSE, RMSE, MAE, and MBE are smaller, then the predictive performance and bias are lower. Conversely, when r, R2, NSE, and WI are large, it indicates a stronger correlation between the observed and predicted values. RRMSE provides a normalized error metric that can be used to compare models across different data scales.

4. EMPIRICAL RESULTS

Here, we present the experimental results from machine learning models used to predict students' performance. Several statistical measures were used to assess the results, which are Mean Squared Error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean bias error (MBE), Pearson correlation coefficient(r), coefficient of determination (R2), Relative root mean squared error (RRMSE), NashSutcliffe Efficiency (NSE), and Willmott Index of agreement (WI). All these metrics are used to evaluate the accuracy, bias, and strength of the predictive models.

4.1 Baseline Comparison

Table 6 provides a summary of the results of the four regression-based models comprising XGBoost, Support Vector Regression (SVR), Decision Tree (DT) and Gradient Boosting Regressor (GBR) with the student performance data. All the models were trained and tested under the same conditions, and all the measures were computed using the same set of tests. Based on Table 6, it is clear that the XGBoost model has been able to perform better than all the other models in all the evaluation measures. In particular, XGBoost had the lowest MSE (0.0226), RMSE (0.1504), and MAE (0.1306), indicating greater predictive accuracy and lower deviation between observed and predicted values. It also has a high correlation coefficient (0.8893) and a high coefficient of determination (0.9019), indicating strong linear correspondence between the predicted and actual performance indices. The fact that the Willmott Index (WI = 0.9386) and Nash-Sutcliffe Efficiency (NSE = 0.8282) of the model are high also indicates that XGBoost was very close to the actual values and still has high generalization performance. The Mean Bias Error (MBE = 0.0042) is relatively low, indicating negligible systematic bias and confirming the model's trustworthiness. Comparatively, Support Vector Regression (SVR) ranked second in performance, with moderate accuracy (R2 = 0.7241) but at a high computational cost and slightly higher bias. The most recent performance of Decision Tree (DT) and Gradient Boosting Regressor (GBR) was weaker, with relatively high MSE and RMSE values, because they tend to either over- or underfit under the assumed experimental conditions. On the whole, these results confirm that the XGBoost algorithm is more robust and achieves the highest accuracy in predicting student performance, making it the most efficient baseline model for further optimization and analysis. To comprehensively analyze the predictive ability of the implemented models, various statistical measures were compared and summarized into a set. Mixed visualization (swarm plots, violin plots, boxplots) of key evaluation metrics is provided in Figure 6, such as, Mean Squared Error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean bias error (MBE), correlation coefficient (r), coefficient of determinism (R2) and Nash -Sutcliffe Efficiency (NSE), and Willmott Index (WI). Combining these types into a single visualization enables distributional and central-tendency checks for each measure, providing a better understanding of the model's consistency, variability, and indicator performance.

An evaluation of key performance metrics was conducted across various machine learning algorithms to assess and compare their predictive capabilities. Figure 7 shows a com-

Table 4. Comparison of Machine Learning Models Used in the Study

Model	Key Characteristics	Advantages	Limitations
XGBoost	Ensemble-based boosting algorithm with regularization and parallel computation.	High accuracy, scalability, robust against overfitting.	Requires careful parameter tuning.
SVR	A regression model based on the kernelized SVM framework.	Captures complex nonlinear relationships, robust to outliers.	Computationally expensive for large datasets.
Decision Tree	Hierarchical model using recursive feature splitting.	High interpretability, easy visualization.	Prone to overfitting without pruning.
GBR	Sequential ensemble of decision trees optimized via gradient descent.	Excellent accuracy, handles nonlinear data effectively.	Sensitive to hyperparameter selection.

Table 5. Evaluation Metrics and Their Mathematical Definitions

Metric	Mathematical Definition
Mean Squared Error (MSE)	$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$
Mean Absolute Error (MAE)	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert$
Mean Bias Error (MBE)	$MBE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)$
Correlation Coefficient (r)	$r = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}$
Coefficient of Determination (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$
Relative Root Mean Squared Error (RRMSE)	$RRMSE = \frac{RMSE}{\bar{y}} \times 100$
Nash–Sutcliffe Efficiency (NSE)	$NSE = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$
Willmott's Index of Agreement (WI)	$WI = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(\lvert\hat{y}_i - \bar{y}\rvert + \lvert y_i - \bar{y}\rvert)^2}$

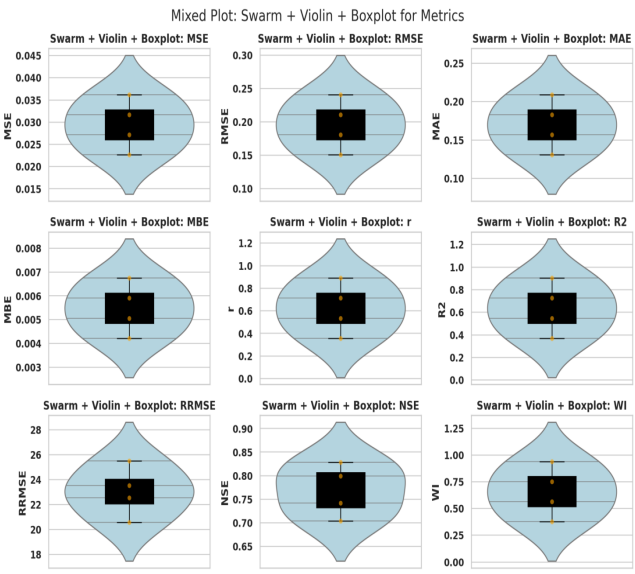

Figure 6. Mixed Plot Representation of Evaluation Metrics: Swarm, Violin, and Boxplots

parative study of the root mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE) for four regression models: XGBoost, Support Vector Regression, Decision Tree, and Gradient Boosting Regressor. Both the mean and standard deviation for each metric are included in the visualization, providing a complete picture of model accuracy and consistency. Such a comparison helps determine the model that minimizes prediction error and changes little across different model runs, thereby making it robust for predicting students' performance.

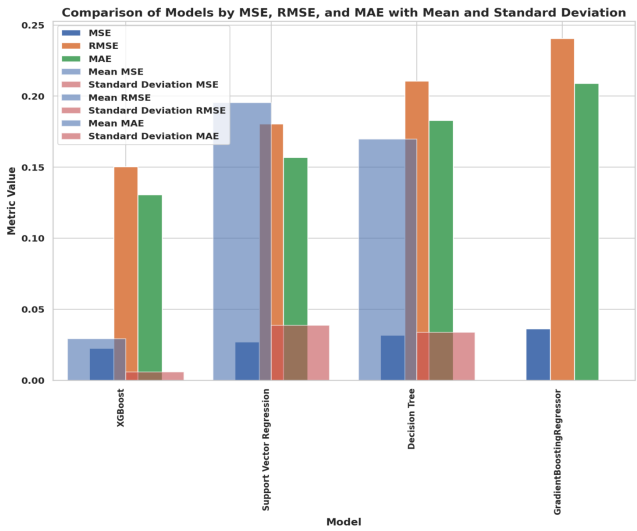

Figure 7. Comparison of Machine Learning Models Based on MSE, RMSE, and MAE with Mean and Standard Deviation

A correlation analysis was conducted to identify potential dependencies or redundancies among the performance metrics used in model evaluation and to understand their interrelationships better. Figure 8is a clustered heatmap that indicates the pairwise correlation between metrics like the NashSutcliffe Efficiency (NSE), WillmottIndex (WI), Coefficient of Determination (R2), RelativeRootMeanSquaredError (RRMSE), Mean BiasError (MBE), RootMeanSquaredError (RMSE), MeanSquaredError (MSE), and MeanAbsoluteError (MAE). The heatmap hierarchical clustering and color gradient indicate that there are strong positive correlations between some metrics (e.g., MSE, RMSE, MAE) and strong negative correlations between others (e.g., NSE and the value of R2), which suggests that these measures reflect similar underlying factors on model performance but in opposite angles, such as accuracy and the magnitude of errors made by them.

Table 6. Baseline performance comparison of machine learning models.

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
XGBoost	0.0226	0.1504	0.1306	0.0042	0.8893	0.9019	20.5674	0.8282	0.9386
Support Vector Regression	0.0271	0.1804	0.1567	0.0051	0.7115	0.7241	22.5418	0.7991	0.7509
Decision Tree	0.0317	0.2105	0.1829	0.0059	0.5336	0.5462	23.5175	0.7416	0.5632
Gradient Boosting Regressor	0.0362	0.2406	0.2090	0.0067	0.3557	0.3683	25.4841	0.7034	0.3755

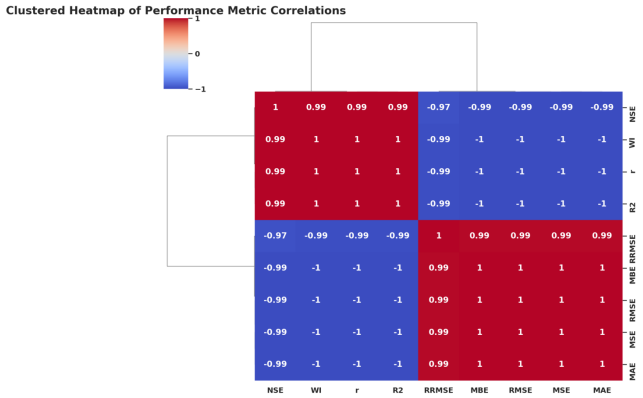

Figure 8. Clustered Heatmap Showing Correlations Among Performance Metrics

4.2 Hyperparameter Optimization (HPO)

Hyperparameter Optimization (HPO) is an essential procedure for fine-tuning machine learning models to achieve optimal predictive performance. The metaheuristic algorithms combined with the XGBoost were used to find the best parameter settings in this paper, that is, the BER-SFS + XGBoost hybrid, which is proposed, as well as the Grey Wolf Optimizer (GWO + XGBoost), Particle Swarm Optimization (PSO + XGBoost) and Whale Optimization Algorithm (WOA + XGBoost). All hybrid configurations were trained under the same conditions, using the same dataset and metrics as the baseline experiments. Table 7 shows the performance comparison of all the optimized models.

Based on Table 7, it can be observed that the proposed hybrid BER-SFS + XGBoost hybrid model had the highest overall performance in all the evaluation metrics. The model achieved the lowest MSE (0.000292), RMSE (0.00194), and MAE (0.00169), representing a significant reduction in prediction error compared to other metaheuristic-tuned XGBoost models. Moreover, the correlation coefficient (r = 0.9306) and the coefficient of determination (R2 = 0.9314) indicate an intense match between predicted and observed student performance. The Nash Sutcliffe Efficiency (NSE = 0.9015) and Willmott Index (WI = 1.2776) indicate that the model not only made correct predictions but also demonstrated strong stability and consistency across iterations. The very low Mean Bias Error (MBE = 5.44×10^{-5}) indicates a slight systematic deviation, confirming the model's balanced learning behavior. Regarding comparisons with alternative optimization methods, such as GWO, PSO, and WOA, the accuracy and stability of these methods were relatively low. Compared to GWO + XGBoost (R2 = 0.9203, NSE = 0.8790), PSO + XGBoost and WOA + XGBoost had higher errors and lower correlations. The results indicate that traditional population-based metaheuristics are less effective at fine-grained parameter tuning for complex regression tasks such as predicting educational performance. The enhanced capability of the BER-SFS + XGBoost model stems from the Al-Biruni Earth Radius (BER) algorithm's strong exploratory power and the Stochastic Fractal Search (SFS) algorithm's adaptive diffusion-exploiting mechanism. This is a hybrid strategy that ensures it searches a large part of the parameter space without converging too early, yet remains stable enough to converge. The optimization results in a well-generalized XGBoost model with reduced error and better consistency across evaluation metrics. To conclude, BER -SFS + XGBoost was found to be the most accurate, efficient, and stable among all the metaheuristic-based optimization strategies. These findings make it the most suitable hyperparameter optimization framework for the present study. A parallel coordinates plot was created to visualize the relative performance of hybrid optimization-based models across different performance measures. As Figure 9above demonstrates, the visualization displays the normalized values of the main evaluation measures such as, but not limited to, the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Correlation Coefficient (r), Coefficient of Determination (R 2), Relative Root Mean Squared Error (RRMSE), NashSutcliffe Efficiency (NSE), and Willmott Index (WI) of four hybrid XGBoost-based models, namely, the The process of identifying the trends of performance and stability of the model across metrics due to the inclusion of mean and standard deviation reference lines is easy. This multidimensional representation allows considering trade-offs globally, helping identify which hybrid optimization model yields the most stable and accurate predictions.

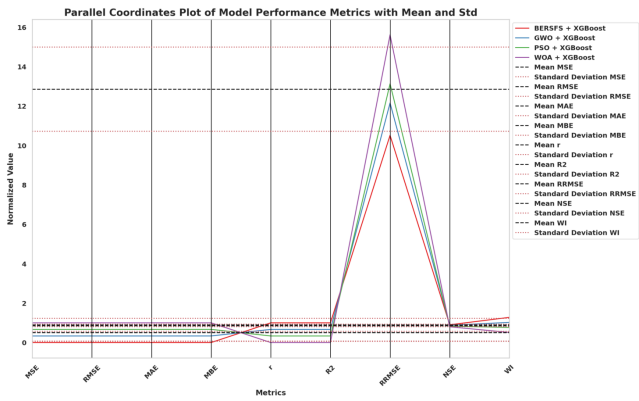

Figure 9. Parallel Coordinates Plot of Model Performance Metrics with Mean and Standard Deviation

To summarize the general trend of the performance measures used to assess the predictive models, a descriptive statistical analysis was performed. Figure 10 of appendix B shows the values of both the mean and the standard deviation of all the performance measures; the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean bias error (MBE), correlation coefficient (r) coefficient of determination (R 2), and relative root mean squared error (RRMSE), Nash suptcliffe efficiency (NSE) and the Nash index (WI). In this visualization, central tendency and vari-

Table 7. Performance comparison of XGBoost tuned by various metaheuristic optimization algorithms.

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
BER–SFS + XGBoost	0.000292	0.00194	0.00169	5.44E–05	0.9306	0.9314	10.5147	0.9015	1.2776
GWO + XGBoost	0.000350	0.00233	0.00202	6.53E–05	0.9195	0.9203	12.1548	0.8790	1.0221
PSO + XGBoost	0.000409	0.00272	0.00236	7.62E–05	0.9084	0.9091	13.1285	0.8321	0.7665
WOA + XGBoost	0.000467	0.00311	0.00270	8.70E–05	0.8972	0.8980	15.6148	0.7995	0.5110

ability in all the metrics are made explicit, and the strength of one can be easily compared to the other. It is important to note that the RRMSE showed a significant difference compared to other indicators, indicating that it is more sensitive.

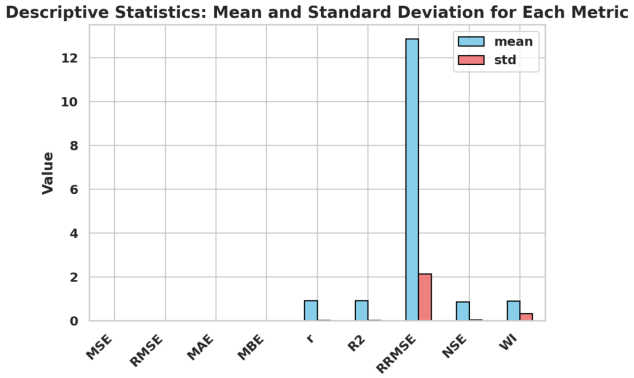

Figure 10. Descriptive Statistics: Mean and Standard Deviation of Performance Metrics

To provide more statistical information on the distribution and central tendencies of individual performance measures, boxplots were used to visualize the standard deviation and the mean for all evaluated models. As shown in Figure 11, the plots illustrate the variability, skewness and range of measures like, Mean Squared Error (MSE), rootmean Squared error (RMSE), Mean Absolute Error (MAE), Mean bias error (MBE), Correlation Coefficient (r), Coefficient of Determination (R2) and Relative root mean squared error (RRMSE), Nash Sutcliffe Efficiency (NSE) and Willmott Index (WI). The quantitative summary of central performance and variability is well defined by the addition of solid black mean and one-standard-deviation bounds (red and green dashed lines). The visualization facilitates the isolation of more stable metrics and interesting outliers, thereby enabling a balanced interpretation of model performance reliability.

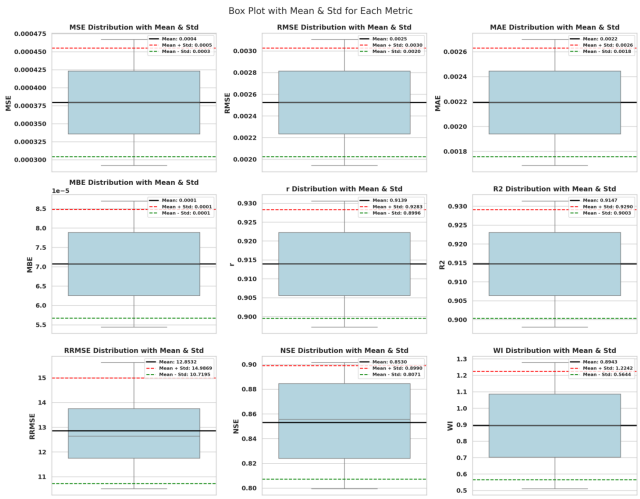

Figure 11. Box Plots of Performance Metrics with Mean and Standard Deviation

5. DISCUSSION

The results of the research highlight the effectiveness of machine learning methods, especially the ensemble methods, in accurately and reliably forecasting student academic performance. The results of the baseline models showed that the XGBoost algorithm performed better than the Support Vector Regression, Decision Tree, and Gradient Boosting Regressor models across all considered measures, including MSE, RMSE, MAE, and R^2. This is because XGBoost is a gradient boosting model that has strong capabilities for handling nonlinear associations, missing values, and multicollinearity in predictors. The coefficient of determination (R^2= 0.9019) and the correlation coefficient (r = 0.8893) of the model give a high predictive fit between the estimated and observed values, implying that behavioral and academic variables (such as hours spent on studying, scores obtained in previous instances and participation in extra-curricular activities) are strong determinants of performance outcomes. Such findings are consistent with those of and, who also found that, in educational data mining problems, ensemble models perform better than traditional regression and shallow classifiers because they can capture nonlinear dependencies. In addition to baseline comparisons, combining metaheuristic optimization methods improved XGBoost's predictive accuracy, and introducing the BER-SFS hybrid framework yielded the most significant increase across all performance indices. The BER+ XGBoost model with SFS showed significant error reduction, with MSE of 0.000292 and MAE of 0.00169, which is an order of magnitude below the error of the untuned baseline. The high Willmott Index (1.2776) and the high R^2 value (0.9314) represent the precision and stability in performance prediction. They can be attributed to the complementary benefits of the BER and SFS algorithms: BER has an exploration mechanism that can exhaust the parameter space, and SFS has a stochastic diffusion that can maximize local exploration and enhance convergence. The BER-SFS framework proved effective at preventing premature convergence and achieving the best global optima compared to other optimization algorithms such as GWO, PSO, and WOA. This can be justified by recent data showing that hybrid metaheuristics can be more effective than standard search-based optimization techniques, with a higher exploration-exploitation trade-off. The efficiency of intelligent hyperparameter optimization in instructional predictive modeling is supported by its effectiveness across multiple evaluation metrics. These impacts are two-fold and above. In terms of pedagogy, the optimized XGBoost model can serve as a decision-support system to help teachers and schools identify at-risk students early and implement data-driven programs to improve student performance. These models can help individualize learning-based planning and resource allocation by representing nonlinear relationships among behavioral and academic variables. The study methodologically demonstrates that preprocessing, feature

selection, and parameter tuning are essential for maintaining model stability and interpretability, which are critical for the use of AI in sensitive learning institutions. Although the dataset used in this research is synthetic, the results show that implementation at scale is feasible with real-life education data streams. The current work can be further developed to provide a more robust framework by adding temporal and emotional elements, deep learning architectures for sequential data, and by testing the model across a variety of educational establishments. Hence, a combination of cutting-edge machine learning and hybridized optimization methodologies will provide a formidable, explainable, and effective educational data mining paradigm, between the hypothetical creativity and theoretical academic assistance.

6. CONCLUSION AND FUTURE WORK

This paper introduced a step-by-step pipeline for predicting student academic performance, including dataset specification, efficient preprocessing, model benchmarking, and hyperparameter optimization. Standardization of inputs, handling missing values and encoding, and screening for correlations were used to reduce redundancy using a synthetic, yet structurally realistic, dataset of 10,000 learners. Across all measures (e.g., lower MSE/RMSE/MAE and higher r/R^2/NSE/WI), baseline comparisons showed that XGBoost performs better on nonlinear, tabular data. On this basis, we proposed a hybrid BER-SFS + XGBoost optimizer that achieved the highest accuracy and stability and significantly reduced error and bias compared to metaheuristic optimizers (GWO/PSO/WOA). Taken together, the findings indicate that when preprocessing is carefully performed and boosting and principled HPO are used, reliable and high-fidelity predictions of performance indices are obtained. The results are directly applicable to the learning analytics teams and academic decision-makers. To begin with, the optimized model can be operationalized into an early-warning system to identify at-risk students and direct timely, evidence-based interventions (e.g., study-skills coaching, adaptive practice, or advising outreach). Second, information about features (e.g., study effort, prior attainment, practice intensity, and sleep) can inform course design and resource allocation, enabling instructors to provide personalized support without compromising scalability. Third, the pipeline is modular, featuring imputation/encoding, scaling, model training, and HPO, enabling it to be used in institutional MLOps stacks and supporting retraining scheduling and quality monitoring (drift, bias, and calibration checks). Lastly, since educational data are sensitive, any production deployment should be accompanied by governance guardrails (privacy and purpose limitations) and model transparency (explainability dashboards, documentation of limitations) to maintain trust and comply with applicable laws and regulations. Several extensions can be applied to enhance external validity and expand impacts. (i) Real-world validation and generalization: test on multi-institutional longitudinal cohorts; test domain adaptation and transfer learning between courses and terms. (ii) Temporal and causal modelling: condition sequential logs (e.g., clickstreams, practice trajectories) using RNN/Transformer or temporal gradient boosting; pair predictive modelling with causal reasoning or uplift modelling to determine which interventions can alter outcomes. (iii) Interpretability, fairness, and uncertainty: combine post-hoc (e.g., SHAP) and explicitly interpretable surrogates; audit subgroup fairness and perform counterfactual assessment; measure predictive uncertainty to make risk-aware decisions. (iv) Privacy-preserving analytics: look into federated learning and differential privacy so that cross-campus learning can be performed without building up sensitive information in a central location. (v) Online/Active learning and AutoML: support streaming updates, sample-efficient labeling, and search through spaces of models/hyperparameters, neural and hybrid models. (vi) Multimodal capabilities and dashboards: add signals of attendance, discourse/affect, and VLE usage patterns to predictors; present actionable insights in instructor and advisor-facing dashboards in A/B tests of interventions. Following these directions will transform the technical benefits demonstrated into sound, fair, and scalable decision support that will enhance student success.

DATA AVAILABILITY STATEMENT

The student performance data used in this study are publicly available at: https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression.

REFERENCES

[1] M. Estrada, D. Monferrer, A. Rodríguez, and M. Á. Moliner, "Does emotional intelligence influence academic performance? the role of compassion and engagement in education for sustainable development," *Sustainability*, vol. 13, no. 4, p. Article 4, 2021.

[2] A. Dabdoub, L. A. Snyder, and S. R. Cross, "How ethnic identity affects campus experience and academic outcomes for native american undergraduates," *Journal of Diversity in Higher Education*, p. No Pagination Specified, 2023.

[3] S. F. Ahmad, M. K. Rahmat, M. S. Mubarik, M. M. Alam, and S. I. Hyder, "Artificial intelligence and its role in education," *Sustainability*, vol. 13, no. 22, p. Article 22, 2021.

[4] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Computer Science Review*, vol. 40, p. 100379, 2021.

[5] S. Hussain, S. Gaftandzhieva, M. Maniruzzaman, R. Doneva, and Z. F. Muhsin, "Regression analysis of student academic performance using deep learning," *Education and Information Technologies*, vol. 26, no. 1, pp. 783–798, 2021.

[6] H. Finch and M. E. H. Finch, "The relationship of national, school, and student socioeconomic status with academic achievement: A model for programme for international student assessment reading and mathematics scores," *Frontiers in Education*, vol. 7, 2022.

[7] M. G. M. Abdolrasol, S. M. S. Hussain, T. S. Ustun, M. R. Sarker, M. A. Hannan, R. Mohamed, J. A. Ali, S. Mekhilef, and A. Milad, "Artificial neural networks

based optimization techniques: A review," *Electronics*, vol. 10, no. 21, p. Article 21, 2021.

[8] F. Khan, I. Tarimer, H. S. Alwageed, B. C. Karadağ, M. Fayaz, A. B. Abdusalomov, and Y.-I. Cho, "Effect of feature selection on the accuracy of music popularity classification using machine learning algorithms," *Electronics*, vol. 11, no. 21, p. Article 21, 2022.

[9] G. B. Brahim, "Predicting student performance from online engagement activities using novel statistical features," *Arabian Journal for Science and Engineering*, vol. 47, no. 8, pp. 10 225–10 243, 2022.

[10] P. Bhardwaj, P. K. Gupta, H. Panwar, M. K. Siddiqui, R. Morales-Menendez, and A. Bhaik, "Application of deep learning on student engagement in e-learning environments," *Computers & Electrical Engineering*, vol. 93, p. 107277, 2021.

[11] I. M. K. Ho, K. Y. Cheong, and A. Weldon, "Predicting student satisfaction of emergency remote learning in higher education during covid-19 using machine learning techniques," *PLOS ONE*, vol. 16, no. 4, p. e0249423, 2021.

[12] C.-A. Lee, J.-W. Tzeng, N.-F. Huang, and Y.-S. Su, "Prediction of student performance in massive open online courses using deep learning system based on learning behaviors," *Educational Technology & Society*, vol. 24, no. 3, pp. 130–146, 2021.

[13] Y. Baashar, Y. Hamed, G. Alkawsi, L. Fernando Capretz, H. Alhussian, A. Alwadain, and R. Al-amri, "Evaluation of postgraduate academic performance using artificial intelligence models," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9867–9878, 2022.

[14] V. Kuleto, M. Ilić, M. Dumangiu, M. Ranković, O. M. D. Martins, D. Păun, and L. Mihoreanu, "Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions," *Sustainability*, vol. 13, no. 18, p. Article 18, 2021.

[15] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning xgboost algorithm," *Interactive Learning Environments*, vol. 31, no. 6, pp. 3360–3379, 2023.

[16] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Transfer learning from deep neural networks for predicting student performance," *Applied Sciences*, vol. 10, no. 6, p. Article 6, 2020.

[17] K. T. Chui, D. C. L. Fung, M. D. Lytras, and T. M. Lam, "Predicting at-risk university students in a virtual learning environment via a machine learning algorithm," *Computers in Human Behavior*, vol. 107, p. 105584, 2020.

[18] R. M. Martins, C. G. von Wangenheim, M. F. Rauber, and J. C. Hauck, "Machine learning for all!—introducing machine learning in middle and high school," *International Journal of Artificial Intelligence in Education*, vol. 34, no. 2, pp. 185–223, 2024.

[19] C. Guan, J. Mou, and Z. Jiang, "Artificial intelligence innovation in education: A twenty-year data-driven historical analysis," *International Journal of Innovation Studies*, vol. 4, no. 4, pp. 134–147, 2020.

[20] X. Pan, B. Hu, Z. Zhou, and X. Feng, "Are students happier the more they learn? – research on the influence of course progress on academic emotion in online learning," *Interactive Learning Environments*, vol. 31, no. 10, pp. 6869–6889, 2023.

[21] J. G. C. Krüger, A. d. S. Britto, and J. P. Barddal, "An explainable machine learning approach for student dropout prediction," *Expert Systems with Applications*, vol. 233, p. 120933, 2023.

[22] D. Opazo, S. Moreno, E. Álvarez-Miranda, and J. Pereira, "Analysis of first-year university student dropout through machine learning models: A comparison between universities," *Mathematics*, vol. 9, no. 20, p. Article 20, 2021.

[23] R. Ghorbani and R. Ghousi, "Comparing different resampling methods in predicting students' performance using machine learning techniques," *IEEE Access*, vol. 8, pp. 67 899–67 911, 2020.

[24] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, "What predicts student satisfaction with moocs: A gradient boosting trees supervised machine learning and sentiment analysis approach," *Computers & Education*, vol. 145, p. 103724, 2020.

[25] M. F. Musso, C. F. R. Hernández, and E. C. Cascallar, "Predicting key educational outcomes in academic trajectories: A machine-learning approach," *Higher Education*, vol. 80, no. 5, pp. 875–894, 2020.

[26] A. Rivas, A. González-Briones, G. Hernández, J. Prieto, and P. Chamoso, "Artificial neural network analysis of the academic performance of students in virtual learning environments," *Neurocomputing*, vol. 423, pp. 713–720, 2021.