STM JOURNALS

International Journal of
Education Sciences

ISSN: 3048-9784
Volume 3, Issue 2, 2026
July–December
DOI (Journal): 10.37591/IJES

https://journals.stmjournals.com/ijes

Research IJES

Building a Fair and Transparent Student Performance Predictor Using Interpretable Machine Learning

Preyash Choudhary[1,*], Ankur Das[2], Sonu Gupta[3]

Abstract

This research incorporates a new, ethical framework that will help predict student performance based upon equitable access for all students, with a specific focus on the principles of ethics (fairness, transparency). The primary focus of this work is to use a scaled Random Forest Regressor (RFR) model to project Semester Three CGPA of MCA students, by including several different enrichment factors into a single comprehensive enriching dataset containing sem1 and sem2 GPAs, 10th and 12th percentiles, as well as additional dynamic behavioral indicators such as attendance rates and study hours, and demographic and engagement factors including gender and extracurricular activities among others. In integrating SHapley Additive exPlanations (SHAP) with an ensemble RFR model creates a combined effect of enabling the separation of the 'black-box' effect of the RFR model. SHAP provides mathematically sound, local, and global attribution to features to allow for the transformation of a passive predictor into a personalised, causal diagnostic tool. Additionally, the methodology utilises a specialised gender sensitivity audit that will enable organisations to identify and quantify and report potential 'disparities' that could occur at the group level. This satisfies the ethical requirement for equity in such organisations. The final outcome is an IML-enabled predictive system that is made available to you through a live web application. It provides not only a prediction method but also an ethical, actionable framework for higher education for the development of ethical and equitable decision-making and for providing educational interventions in the area of higher education.

Keywords: Interpretable machine learning (IML), student performance prediction, fairness in AI, SHAP, random forest, educational data mining (EDM), transparency

INTRODUCTION

The influx of Machine Learning (ML) into Educational Technology (EdTech) will create a transformation to the way we support students and allocate resources. The increasing use of algorithmic prediction of academic success and the identification of students at risk of failure has allowed for personalized learning paths. However, the implementation of powerful predictive models, such as complex ensemble models like Random Forest Regressors or neural networks, brings along a unique challenge: lack of transparency. When a student is flagged as "at-risk" or when a recommendation is made for a particular academic path, the absence of an easily identifiable and verifiable reason behind the prediction creates a "black box." This loss of transparency diminishes student, parent, and educator confidence and severely limits the ability of faculty to give targeted and effective guidance to students. Without a clear understanding of which input characteristics (e.g., previous GPA, attendance,

***Author for Correspondence**
Preyash Choudhary
E-mail: preyashchoudhary28@gmail.com

[1]Research Scholar, Department of CSE, Thakur Institute of Management Studies, Career Development & Research ,Mumbai, Maharashtra, India
[2]Research Scholar, Department of CSE, Thakur Institute of Management Studies, Career Development & Research ,Mumbai, Maharashtra, India
[3]Assistant Professor, Department of CSE, Thakur Institute of Management Studies, Career Development & Research ,Mumbai, Maharashtra, India

Received Date: April 15, 2026
Accepted Date: September 23, 2026
Published Date: September 24, 2026

Citation: Preyash Choudhary, Ankur Das, Sonu Gupta. Building a Fair and Transparent Student Performance Predictor Using Interpretable Machine Learning. Journal of Education Sciences. 2026; 3(2): 48–56p.

© STM Journals 2026. All Rights Reserved

study hours) had the highest weight in predicting a low score, any intervention will likely be generic and ineffective. The following is an updated and revised version of the original content. The updated content is more formal and contains grammatical errors and other issues with the original text [1].

A significant issue with the lack of transparency in predictive modeling is the potential to introduce algorithmic bias. For example, if a model learns from historical data that exhibits societal inequities (for example, inequities in access to education, demographics, etc.), the algorithm may develop a system that either inadvertently replicates or enhances these biases, resulting in negative impacts to various marginalized student populations [2].

A comprehensive predictive model must have mechanisms of both explanation and equity auditing as part of the development process.

The objectives of this research are to address both of these challenges by accomplishing three interrelated goals:

Robust Predictive Modeling

To create an accurate and reliable predictive model for predicting Semester 3 CGPA based on an extensive range of student-based information, such as academic, behavioral, and demographic information. The goal of this objective is to establish a performance benchmark for the model [3].

Conversion of an Inherently Opaque Model into a White Box Model using Interpretable Machine Learning (IML)

The goal of this objective is to convert an inherently opaque predictive model into a white-box model by using IML methods. Specifically, IML uses SHAP values to produce localized, quantitative, and individualized causal variables for each prediction to bridge the gap between predictive accuracy and predictive model transparency [4].

Fair and Ethical Auditing

Establishing an innovative, ethically compliant, and comprehensive methodology for the prediction, analysis, and management of academic success (academic achievement) with regards to one's personal characteristics (such as Gender) through the use of a fair and unbiased predictive model that provides both individualized analysis and systemic analysis of performance disparities between groups of individuals. The model provides a means to proactively monitoring how individuals are performing academically, thus providing an opportunity to develop interventions and effective support for the advancement of their academic success (achievement). The product of this research will provide an innovative, ethically compliant, and complete approach to predicting, analyzing, and supporting the academic success of all students regardless of their individual characteristics [5].

LITERATURE REVIEW

Student Performance Prediction using Machine Learning

The prediction of student performance has gained much prominence in EDM since it allows identifying at-risk students at an early stage and helps make timely educational interventions. In conventional prediction models, the most popular method was using statistical modeling techniques like linear and logistic regression that generated understandable results but were not able to capture the complexity of the relationship between the educational variables. The recent developments in the field of machine learning have led to the emergence of prediction algorithms like Decision Trees, SVM, Random Forest, Gradient Boosting, XGBoost, and Artificial Neural Networks [6].

Early research on forecasting student performance demonstrated that CGPA (Cumulative Grade Point Average) from previous semesters was the most significant predictor for students' success in their future academic pursuits. As such, the primary focus of this early research was predictive

© STM Journals 2026. All Rights Reserved

modeling based on CGPA alone; however, this approach did not take into consideration that CGPA is not the only aspect of student engagement and success. Moreover, the scope of subsequent research expanded beyond CGPA, discovering that student engagement metrics such as attendance frequency and course participation increased the one-dimensional predictive utility of CGPA as a mechanism for forecasting student's future academic performance. In addition to expanding the definition of factors utilised to predict a student's performance in college to include the student's high school GPA (grade point average), other forms of non-CGPA data that support predictive modelling are also being incorporated into predictive analysis models [7].

Machine Learning Models for Student Prediction

Out of different kinds of machine learning algorithms, Random Forest is one of the most popular algorithms used today owing to its reliability, non-susceptibility to overfitting, and capability to analyze heterogeneous educational data. Different from individual decision trees, Random Forest merges different decision trees through ensemble learning, resulting in low variability while retaining high precision. Many authors have found that Random Forest always performs better than traditional statistical methods in terms of predicting students' performance based on previous academic data, attendance, test scores, and behavioral characteristics [8].

The use of advanced predictive modelling techniques (e.g., Random Forest Regression) has enhanced the accuracy of predictions regarding future student performance due to the ability to use ensemble methods for modelling complex, multidimensional interactions of student data without major susceptibility to overfitting [9].

Explainable Artificial Intelligence (XAI)

Even though they perform very well in prediction tasks, complex machine learning models tend to be black box models and hence their decision processes cannot be easily understood. One approach that can be used to increase transparency in such models is called Explanable Artificial Intelligence (XAI). Of the various XAI approaches that exist today, one that has been extensively used is SHapley Additive exPlanations (SHAP) since it is mathematically consistent in its feature attribution using cooperative game theory. It can provide local as well as global explanations for any model [10].

As there are both advantages and disadvantages to each type of model, there has been an increased need for Model Agnostic Interpretability Techniques, as the inherent complexity of RFR models and other ensemble or deep learning models makes Interpreting these models very complex. The use of Interpretable Machine Learning (IML) techniques to explain the prediction of the model associated with "black-box" models provides a unique and powerful approach to provide an "explanation" for why a certain prediction was made. The two individuals credited with the first use of IML techniques were Lundberg and Lee in 2017 when they created the concept of SHapley Additive exPlanations, more commonly called SHAP [11].

The derivation of the SHAP value is based on an analysis of how cooperative game play works. The authors have shown that if certain conditions (local accuracy, missingness, consistency) are met, the Shapley Value will be the only unique value that meets these criteria. Additionally, while other techniques such as Permutation Importance only provide a global ranking of features, SHAP provides the "local" account of why the model predicted a specific value. This capability provides the "gold standard" of providing interpretable explanations in sensitive fields, such as education, where it allows advisors to not only understand what a model predicts, but also precisely why the model predicts the outcome it does for an individual student. This also means that an advisor will have access to a global picture of feature importances derived from aggregating SHAP values across all predictions made for an individual student [12].

Fairness in Educational AI

With the rise in the use of Artificial Intelligence in higher education, it is imperative to mention that

© STM Journals 2026. All Rights Reserved

the sole focus on prediction accuracy is not enough for the proper use of AI. There are various ethical requirements which should be met by any educational prediction system as well. Multiple works have shown that predictive models trained using educational data can inadvertently adopt demographic biases especially those related to protected characteristics like gender and socio-economic status.

From our research, we have determined that a model should also take into consideration the associated biases and impacts on different demographics. In this case, we are looking at educational data mining's potential to introduce a disparity impact on students due to an algorithmic bias related to gender. Based on judicial mandates, we will compare aggregated mean predicted performance values for students from protected gender classes in our model and measure the differences through systematic use of a verified process for auditing and for reporting systematic discrepancies in predicted outcomes based on protected classes. In addition to our fairness metric, we integrate interpretability within the model for full transparency to ensure that our model meets ethical standards of equity and anti-discrimination [13].

Research Gap

Even though prior work in this field has been able to build machine learning algorithms capable of accurately predicting the performance of students, there are several drawbacks that have not yet been addressed. Current literature mostly deals with optimizing prediction accuracy without providing any kind of explanation for the prediction itself. Researches that utilize SHAP mostly focus on explainability and do not incorporate the assessment of fairness in their prediction algorithm. Educational AI research related to fairness mostly assesses demographic gaps without giving any personalized explanation for teachers and students. Thus, there is a need for a unified approach that can provide accurate prediction, explainability, and fairness audit at the same time [14].

METHODOLOGY

Data Preparation and Feature Engineering

The dataset used for building the predictor was obtained from *Dataset.csv*. Several preprocessing steps were performed to ensure the data was clean, consistent, and suitable for model training.

To address missing values, all numerical attributes—including GPA scores, attendance records, and subject marks—were examined. Any missing entries were replaced using the mean value of the respective column. This approach is appropriate when the data follows a roughly normal distribution and helps preserve overall trends without introducing bias [15].

Since the original dataset did not contain a target variable for the third-semester GPA, a simulated target named Simulated_Sem3_GPA was created. This variable represents a realistic academic outcome and was generated as a weighted combination of the input features, with the addition of a small Gaussian noise term, N(0,0.1).). The purpose of this simulation was to create a controlled setting where the relationship between features and the target is meaningful yet not perfectly deterministic. This allows the Random Forest Regressor (RFR) to learn the underlying patterns effectively while also enabling reliable validation of model interpretability using SHAP values. Categorical variables such as Gender and Extra-Curricular Activities were converted into numerical form to make them compatible with the machine learning model. Gender was encoded as a binary variable (Male = 1, Female = 0), while participation in extra-curricular activities was similarly transformed into a numerical representation. This encoding ensures seamless integration of categorical information into the predictive framework [16].

Why Random Forest?

The Random Forest regression technique was chosen because it is highly accurate in predictions while also able to efficiently address the nonlinear relationship between the education variables. This method works through an ensemble learning process, which avoids overfitting since the predictions are made by averaging those from many decision trees.

© STM Journals 2026. All Rights Reserved

Additionally, Random Forest allows for the processing of both numeric and categorical data.

Building a Predictive Model

To build the predictive model we used Random Forest Regressor (RFR) with N=300 and a random state of 42. The RFR is an ensemble learning method that takes the results from many decision trees and finds the average of all tree outputs to provide a low variance and highly accurate output. The training data that was used is comprised of the complete engineered features, denoted as X, and was used to estimate the given simulated target of Simulated_Sem3_GPA:

X={Sem1_GPA,Sem2_GPA,…,Gender_enc; ExtraCurricularActivities_enc}

y=Simulated_Sem3_GPA

As a result of this exercise every student in the dataset has a corresponding Predicted_Sem3_GPA.

Why SHAP?

Even though the Random Forest algorithm is capable of delivering accurate predictions, the fact that it is an ensemble model restricts its interpretability. This is why SHAP was introduced to enhance the interpretability of the model. Each feature gets a contribution value from SHAP using the concept of cooperative game theory, which allows educators to know why a particular prediction was made.

Interpretable Machine Learning with SHAP

To Increase the transparency of the Random Forest Regressor (RFR) and provide explanations that support the prediction of the model SHAP (SHapley Additive exPlanations) was used. The SHAP values for the model were calculated using the shap.Explainer which was developed to support tree-based models. The SHAP value (ϕ i) represents the average contribution of an individual feature to the final prediction when examining all possible feature interactions combinations. The model prediction can be calculated as follows:

$$\text{Prediction}=\text{Base Value}+ \sum_{i=1}^{M} \phi_i$$

When interpreting model predictions using SHAP, there are two levels of interpretability: Local Interpretability: For each student, local explanations can be generated using waterfall or bar plots that visualize the effect of feature value on the predicted GPA. The plots clearly show the effect of each feature on the final prediction and whether their contribution increases or decreases the overall prediction from the average prediction (base value). The ability to provide feature-based feedback for each student offers the opportunity to provide more personalized support for their academic success. Global Interpretability: By collecting SHAP values for all students, the overall importance of each feature within the model can be determined. This allows for the identification of the most significant factors in determining GPA predictions and therefore highlights the key contributions to a student's performance [17].

Importance of Fairness Auditing

Accuracy of prediction does not necessarily ensure ethical deployment of AI. Since predictive models can accidentally detect the underlying demographic distribution in the educational data set, the model may create an imbalance in treating the various student populations. Thus, fairness auditing was included in assessing the balance of predictions made between the genders [18].

Fairness Auditing

As part of Fairness Auditing, a quantitative Fairness Audit was conducted by calculating and visualizing how well the Average Predicted Sem3 CGPA compares to those of other groups based on Gender attributes. These bar charts show that when comparing average predicted performance groups across all students using the Gender attribute as the Y-axis, it will illustrate where most disparity

© STM Journals 2026. All Rights Reserved

exists; this information will allow for educational authorities to take any necessary actions for educational equity and provide a means to better guide future educational research [19].

RESULTS AND DISCUSSION

The Random Forest model created proved to be able to learn the connection between students' past achievements, behavioral traits and their future academic results. The generated predictions show that the GPA of previous semesters still remains the most significant factor, while the attendance rate and time spent on studying have a considerable influence on the increase of prediction accuracy. The research results are in line with those of earlier Educational Data Mining researches [20].

Homegrown Model Explanation and Personalization Predictions

As a major part of the implementation of the Interpretable Machine Learning (IML) method, the personal prediction page demonstrates how well IML can be applied. When a user selects a student, the Predictive Semester 3 Grade Point Average (GPA) and the SHAP explanatory bar chart will appear immediately. The use of visualized prediction explanations provides an opportunity to see both "local fidelity," confirming that the explanation of a given prediction is indeed tailored specifically to that instance of data [21]. Two cases taken from the model predictions give some examples of the range of prediction results: Case 1: High Achievement Student: The addition of Semester 1 GPA and Attendance Rate, where Semester GPA predicts far over the Semester 2 Student Average (i.e., Semester 2 GPA score [+0.80 points] based on Semester 3 prediction, Semester 1 GPA score [0.60 points]), would present both features as the most positively impacting on the prediction for this student. Case 2: At Risk Student: Conversely, a lower than average number of Study Hours Per Day (e.g., 1 hour) and lower Marks in Maths could easily present as the two largest negatively influencing features on their student prediction for this student (i.e., lower Study Hours Per Day –0.50 points; lower Marks in Maths –0.30 points).

The diagnostic feedback generated by the system is one of its most important contributions. Instead of saying "you are predicted to earn $7.50," it reveals, "Your prediction is primarily based on your excellent academic history as indicated by your previous GPAs, whereas study hours are having a negative impact; by addressing this area, you can improve your score further." The high level of actionable interpretability of this information makes it extremely valuable when used as part of academic advising; educators can focus their efforts on dynamic behavioral attributes related to student engagement rather than static behavioral attributes, such as historical GPA data. When combined with the use of SHAP visualizations, educators are able to switch from the role of passive predictors of academic performance, to that of active diagnostic tools used for providing academic counseling and supporting student academic success [22].

Although the focus of this analysis was to examine local measures of feature influence, the aggregation of SHAP values for all students made it possible to derive a crucial global picture of feature influence. Through the analysis of our Random Forest classification model's SHAP values, it became apparent that the features of prior academic record (Sem1_GPA), and Sem2_GPA, had the highest average magnitude of SHAP values and as such, are the most predictive features of student academic achievement. In addition, as dynamic behavioral factors (AttendanceRate, and

StudyHoursPerDay) were ranked as third and fourth in terms of average magnitude of SHAP values, they should be considered as the most readily manipulable levers for influencing future academic achievement, and should be prioritized by educational institutions as a part of their programs for supporting students academically[23].

Group-Level Fairness Analysis and Disparities

The Integrated Gender-wise Predicted CGPA Analysis is an important source of information about the overall behaviour of the predictive population for all of the predictions made using this model as well as being an information source fulfilling the research's mandate for transparency.

© STM Journals 2026. All Rights Reserved

The Gender Distribution Pie Chart is extremely important contextually, considering that imbalances in gender distribution (such as 70% Male and 30% Female gender distributions) will disproportionately affect the way that the predictive model performs on a specific gender (in this case, Female).

The subsequent Bar Chart showing the anticipated CGPA by Gender Distribution is intended to audit the group justice status of a given Population/Cohort being assessed for Predictive Model Fairness.

If there is a statistically significant gap in the anticipated CGPA for all groups (Males and Females) in the Bar Chart (i.e. a deviation from Statistical Parity) then the Model is flagging a possible Group Inequity issue.

For example, if Males have an average anticipated CGPA of
8.45 and Females have an average anticipated CGPA of
8.20, then the Model is flagging an Unequal Impact warning.

While the underlying Random Forest Model is trained on Simulated Data designed to learn specific weights associated with an Individual's/Group's features, this finding is critical.

In a real-world setting, the finding triggers two questions:

1. Is the difference a statistical artefact (e.g, random sampling)?
2. Is the difference a legitimate result of true underlying performance (i.e, differences in attendance rates or study hours between Males and Females) captured by the Model's features?
3. Is there a difference between the Genders due to bias? (In other words - Is the Genders itself carrying too much weight in the model, or is the model just reflecting bias, and therefore historical inequities that are inherent in the data it was trained on?).

By presenting results with this clarity - you're moving from pure prediction towards promotion of equity. The clarity encourages analysis of counterfactual fairness, which is a method of analyzing what a predicted value would have been if the Gender of the student was changed to a different Gender while all other Performance features are held constant. Analysis of that nature is just one way that you can deploy mitigation strategies for biased algorithms (i.e., discussed later in the conclusion) before the model is used to make a decision about a student's academic experience [24].

Through providing SHAP (i.e., one explanation of local predictions) combined with the Fairness Bar Chart (i.e., a method to evaluate group disparity) - it creates a really strong ethical foundation for applying the model [25].

The application of SHAP greatly enhances the practicality of the suggested approach. Instead of providing just the forecasted GPA score, the model identifies the contribution of each individual feature used to make that particular prediction. The explanation allows the professors to develop interventions aimed at addressing modifiable behavioural factors, such as attendance and studying, and not past academic records [26-27].

CONCLUSION

In this study, a framework of interpreting and ensuring fairness of predictions regarding student academic performance using Random Forest Regression and SHapley Additive exPlanations (SHAP) is proposed. The developed framework shows that good predictive performance is possible and that interpretability and accountability in education can be ensured simultaneously. Based on the combination of academic, behavioral, and demographic variables, the framework makes predictions about Semester 3 CGPA and explains the factors affecting the prediction.

© STM Journals 2026. All Rights Reserved

Integration of SHAP adds practical significance to the predictive model because it changes its nature from the traditional black box model to the one that can generate interpretations. While the typical output of any model is the predicted GPA, the developed framework makes predictions and also reveals what factors contribute to each prediction through their contributions. Such factors are previous semester GPA, attendance, study hours, and subject performance.

Another contribution from this study is the addition of the fairness auditing technique to measure prediction results from various demographic groups. The gendered fairness approach encourages the responsible use of AI through enabling education institutions to detect any disparities before implementing prediction systems in their environment. Such a framework of prediction, transparency, and fairness constitutes an ethical way of employing machine learning techniques in the education sector.

This framework is useful in a number of ways. For instance, educational institutions can employ it to identify students who are at risk of poor performance, academic counseling, resource allocation, and decision making. In addition, the explanation capability of the model increases confidence levels of the teachers and students in the system.

However, some limitations of this work need to be highlighted. First, the use of a synthetic Semester 3 GPA as a target variable and the analysis of a dataset collected in one educational environment might constrain the applicability of the results. In future studies, validation of the suggested framework should be done on real-world multi-institutional data. It would also be valuable to test the suggested system for other protected features besides gender, apply more advanced criteria of fairness and methods of bias reduction, and compare the performance of the approach against deep neural networks and other state-of-the-art explainable AI techniques.

REFERENCES

1. Breiman L. Random forests. Machine learning. 2001 Oct;45(1):5–32.
2. Slack D, Hilgard A, Singh S, Lakkaraju H. Reliable post hoc explanations: Modeling uncertainty in explainability. Advances in neural information processing systems. 2021 Dec 6;34:9391–404.
3. Ribeiro MT, Singh S, Guestrin C. " Why should i trust you?" Explaining the predictions of any classifier. InProceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining 2016 Aug 13 (pp. 1135–1144).
4. Albanna B, Handl J, Heeks R. Publication outperformance among global South researchers: An analysis of individual-level and publication-level predictors of positive deviance. Scientometrics. 2021 Oct;126(10):8375–431.
5. Romero C, Ventura S. Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews). 2010 Jul 26;40(6):601–18.
6. Romero C, Ventura S. Educational data mining and learning analytics: An updated survey. Wiley interdisciplinary reviews: Data mining and knowledge discovery. 2020 May;10(3):e1355.
7. Asif R, Merceron A, Ali SA, Haider NG. Analyzing undergraduate students' performance using educational data mining. Computers & education. 2017 Oct 1;113:177–94.
8. Aldowah H, Al-Samarraie H, Fauzy WM. Educational data mining and learning analytics for 21st century higher education: A review and synthesis. Telematics and Informatics. 2019 Apr 1;37:13–49.
9. Saxena M, Singh NP. Role of Hybrid Feature Selection Algorithms in Foreseeing Student Performance. Sciences (IJTNS). 2025;5(1):43–75.
10. Nuankaew P, Nuankaew WS. Student Performance Prediction Model for Predicting Academic Achievement of High School Students. European Journal of Educational Research. 2022;11(2):949–63.

© STM Journals 2026. All Rights Reserved

11. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From local explanations to global understanding with explainable AI for trees. Nature machine intelligence. 2020 Jan;2(1):56–67.
12. Freiesleben T, König G, Molnar C, Tejero-Cantero A. Scientific Inference with Interpretable Machine Learning: Analyzing Models to Learn About Real-World Phenomena: T. Freiesleben et al. Minds and Machines. 2024 Jul 15;34(3):32.
13. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM computing surveys (CSUR). 2018 Aug 22;51(5):1–42.
14. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable ai: A review of machine learning interpretability methods. Entropy. 2020 Dec 25;23(1):18.
15. Barocas S, Hardt M, Narayanan A. Fairness and machine learning. Recommender systems handbook. 2020;1:453–9.
16. Porras JM, Lara JA, Romero C, Ventura S. A case-study comparison of machine learning approaches for predicting student's dropout from multiple online educational entities. Algorithms. 2023 Dec 3;16(12):554.
17. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM computing surveys (CSUR). 2021 Jul 13;54(6):1–35.
18. Raftopoulos G, Davrazos G, Kotsiantis S. Evaluating fairness strategies in educational data mining: A comparative study of bias mitigation techniques. Electronics. 2025 May 1;14(9):1856.
19. Chernozhukov V, Karl Härdle W, Huang C, Wang W. Lasso-driven inference in time and space. The Annals of Statistics. 2021 Jun;49(3):1702–35.
20. James G, Witten D, Hastie T, Tibshirani R. Linear model selection and regularization. InAn introduction to statistical learning: with applications in R 2021 Jul 28 (pp. 225–288). New York, NY: Springer US.
21. Van Norren DE. The ethics of artificial intelligence, UNESCO and the African Ubuntu perspective. Journal of Information, Communication and Ethics in Society. 2023 Jan 31;21(1):112–28.
22. McCarthy J. Generality in artificial intelligence. Communications of the ACM. 1987 Dec 1;30(12):1030–5.
23. Choi WC, Lam CT, Mendes AJ. Analyzing the interpretability of machine learning prediction on student performance using shapley additive explanations. In2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE) 2024 Dec 9 (pp. 1–8). IEEE.
24. Alamri R, Alharbi B. Explainable student performance prediction models: A systematic review. Ieee Access. 2021 Feb 23;9:33132–43.
25. Hoq M, Brusilovsky P, Akram B. Analysis of an explainable student performance prediction model in an introductory programming course. InProceedings of the 16th International Conference on Educational Data Mining 2023 (pp. 79–90).
26. Le Quy T, Nguyen TH, Friege G, Ntoutsi E. Evaluation of group fairness measures in student performance prediction problems. InJoint European Conference on Machine Learning and Knowledge Discovery in Databases 2022 Sep 19 (pp. 119–136). Cham: Springer Nature Switzerland.
27. Sahlaoui H, Nayyar A, Agoujil S, Jaber MM. Predicting and interpreting student performance using ensemble models and shapley additive explanations. IEEe Access. 2021 Oct 29;9:152688–703.

© STM Journals 2026. All Rights Reserved