

Forecasting Students' Academic Performance in Educational Data Using Machine Learning Techniques

Najah Al-Shanableh

<https://orcid.org/0000-0001-9877-8782>

Al al-Bayt University, Mafraq, Jordan

Mazen S. Alzyoud

<https://orcid.org/0000-0003-4729-2103>

Al al-Bayt University, Mafraq, Jordan

Ahmed Khalil

<https://orcid.org/0000-0002-9082-2280>

Higher Colleges of Technology, Sharjah, UAE

Mohamed Benlamine

<https://orcid.org/0000-0001-9858-5526>

Higher Colleges of Technology, Sharjah, UAE

Insaf Kraidia

Al-Ahliyya Amman University, Amman, Jordan

Sadeq Damrah

Australian University, West Mishref, Kuwait

Atena M. Tabakhi

<https://orcid.org/0000-0001-5317-8425>

Washington University in St. Louis, St. Louis, USA

Received: August 14th, 2025 | **Accepted:** January 9th, 2026

ABSTRACT

This study benchmarks multiple machine learning models to predict student academic performance. The research analyzes data from students in mathematics and Portuguese language courses, examining the relationship between various factors and academic performance. The benchmark implementation includes data preprocessing, exploratory data analysis, feature engineering, model training, and hyperparameter tuning for both regression (predicting final grades) and classification (predicting pass/fail outcomes) tasks. The findings demonstrate that ensemble methods, particularly gradient boosting models, outperform other algorithms with root mean square error of 3.34 for regression and F1 score of 0.88 for classification after hyperparameter tuning. Feature importance analysis reveals that past failures, alcohol consumption, study time, and parent education level are among the most influential predictors of academic performance. The results provide valuable insights for educational stakeholders to implement targeted interventions for at-risk students and improve overall academic outcomes.

KEYWORDS

Educational Data Mining (EDM), Knowledge Discovery, Classification, Feature Extraction, Data Mining, Prediction, Clustering, Student's Behavior, Student's Academic Performance, Personalized Learning

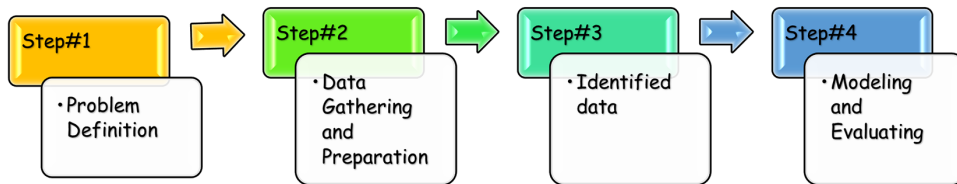
INTRODUCTION

Educational data mining (EDM) has emerged as a vital field applying data mining and machine learning techniques to educational datasets to enhance understanding of students' learning processes and factors affecting their academic performance. As educational institutions accumulate extensive data on student demographics, behaviors, and outcomes, there is a growing opportunity to leverage this information to identify patterns, predict performance, and develop targeted interventions. The data exploration and presentation process in EDM consisted of the steps shown in Fig. 1.

DOI: 10.4018/IJICTE.399756

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Figure 1. Educational Data Mining Process (EDM)



These steps, such as collecting data, preprocessing it, analyzing it, interpreting the results, and displaying them visually, are all needed to gain useful conclusions from educational data. With these realizations, strategies can be put in place to boost how effective teachers are and how much students participate, which eventually results in improved student outcomes (Inthanon & Wised, 2024). Using EDM, institutions help students learn better by making learning environments that suit their unique requirements and preferences. When these methods are applied, they can greatly improve how institutions operate by matching what learners need (Cheng, 2017).

Student academic performance prediction remains a challenging yet crucial task in educational research. Accurate prediction models can help educators identify at-risk students early, understand factors influencing learning outcomes, and develop personalized interventions to improve academic achievement. Traditional statistical methods often fail to capture complex, nonlinear relationships between various factors and student performance, making machine learning approaches particularly valuable in this domain.

Using machine learning algorithms such as regression and classification reveals important connections in educational data and helps improve how well student performance is forecast. Such advancements make it possible to help students early, greatly improving their learning (Al Breiki et al., 2019). On top of that, the integration of new algorithms, including ensemble methods and neural networks, can make predictions more accurate, which benefits educational planning.

Furthermore, using educational data mining provides a way to anticipate student outcomes and uncovers insights about the wider role of socioeconomic factors in learning. Studies have made clear that academic success is shaped by parental education and family dynamics, so any intervention should look at these factors to succeed (Li, 2024). If demographic data is included in predictive models, educators have a better chance to create approaches that fit the unique needs of every student population. This way of thinking highlights the need for a combined data approach that looks after academic outcomes and the cultural and social backgrounds of students, which helps every student achieve success. Using these advanced forecasting tools means students can learn more effectively and feel even more supported in the classroom.

This research aims to benchmark¹ multiple machine learning algorithms for predicting student academic performance using the University of California at Irvine (UCI) Student Performance dataset, which contains information on Portuguese secondary school students. We approach this as both a regression problem (predicting final numeric grades) and a classification task (predicting whether students will pass or fail) while also identifying the most influential factors affecting student outcomes. The insights gained from this study can inform educational policies, guide intervention strategies, and contribute to the broader field of EDM.

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/forecasting-students-academic-performance-in-educational-data-using-machine-learning-techniques/399756

Related Content

Online program Assessment: A Case Study of the University of Illinois at Urbana-Champaign Experience

Faye L. Lesht, Rae-Anne Montague, Vaughn J. Page, Najmuddin Shaikand Linda C. Smith (2008). *Online and Distance Learning: Concepts, Methodologies, Tools, and Applications* (pp. 2673-2689).

www.irma-international.org/chapter/online-program-assessment/27579

Intelligent and Interactive Chatbot Based on the Recommendation Mechanism to Reach Personalized Learning

Ching-Bang Yaoand Yu-Ling Wu (2022). *International Journal of Information and Communication Technology Education* (pp. 1-23).

www.irma-international.org/article/intelligent-and-interactive-chatbot-based-on-the-recommendation-mechanism-to-reach-personalized-learning/315596

The Ecosystem of Human Learning: The Neuroscience of Education

Adriana V. Braescu (2019). *Handbook of Research on Ecosystem-Based Theoretical Models of Learning and Communication* (pp. 86-104).

www.irma-international.org/chapter/the-ecosystem-of-human-learning/223572

Affective Tutoring Systems: Enhancing e-Learning with the Emotional Awareness of a Human Tutor

Nik Thompsonand Tanya Jane McGill (2012). *International Journal of Information and Communication Technology Education* (pp. 75-89).

www.irma-international.org/article/affective-tutoring-systems/70920

The Next Generation of E-Learning: Strategies for Media Rich Online Teaching and Engaged Learning

Daniel Tiong Hok Tan, Chye Seng Leeand Wee Sen Goh (2007). *Future Directions in Distance Learning and Communication Technologies* (pp. 222-242).

www.irma-international.org/chapter/next-generation-learning/18754