

IED09

EVALUATING KEY DETERMINANTS OF STUDENT PERFORMANCE: A DATA-DRIVEN APPROACH USING CORRELATION, REGRESSION, KNN, AND CART MODELS

Mehryar Parsi¹, Nibriti Poudel¹, Wen Cheng²

¹University of the West, Rosemead, CA, USA. ²Cal Poly Pomona, Pomona, CA, USA

Abstract

This paper investigates the key factors influencing student performance by applying multiple analytical methods, including correlation analysis, linear regression, K-Nearest Neighbors (KNN), and Classification and Regression Trees (CART). Using a dataset obtained from Kaggle, essential variables such as Hours studied, Attendance, Previous scores, Tutoring sessions, Physical activity, and Sleep hours were analyzed to understand their impact on exam scores. Results indicate that Hours studied and Attendance are the most significant predictors of student performance, while factors like Sleep hours and Physical activity show minimal influence. The linear regression model explained 59.8% of the variance in exam scores, with both KNN and CART models achieving high accuracy rates, confirming the effectiveness of machine learning techniques in predicting academic outcomes. These findings underscore the importance of academic engagement, particularly attendance and study habits, in determining student success.

Conference Track

Innovative Education