

Using some machine learning methods to predict student performance

Nguyen Bich Quynh

Abstract: This study explores the application of several machine learning methods to predict student academic performance. Using a dataset comprising demographic, behavioral, and academic history features, we implemented and compared models including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and Gradient Boosting. The primary objective was to identify the most effective algorithm for early and accurate prediction of student outcomes, thereby enabling timely educational interventions. Model performance was evaluated based on accuracy, precision, recall, F1-score, and AUC-ROC. Experimental results indicate that ensemble methods, particularly Random Forest and Gradient Boosting, significantly outperformed simpler models, achieving high predictive accuracy. Feature importance analysis further revealed that prior academic achievement and engagement metrics were the most influential predictors. The findings demonstrate that machine learning can serve as a robust tool for identifying at-risk students, providing a data-driven foundation for personalized support strategies in educational settings.

Keywords: machine learning, prediction, student performance, educational data analysis