2nd Global Conference on Innovations in Education

12 - 13 December 2024 Dubai - UAE

Comparative Analysis of Machine Learning Models on Student Performance Data: Insights from Test Scores and Survey Data

Sanjana Sundararaman, and Dr. Maheen Hasib

Heriot Watt University, Scotland

Abstract

With the increasing use of digital learning platforms, large volumes of student data have become available for analysis. This paper investigates how machine learning, learning analytics, and educational data mining can be utilized to gain insights into student performance. Various predictive modeling techniques, including Random Forest (RF), K- Nearest Neighbour(KNN), and Decision Trees (DT), are evaluated for their ability to forecast student test scores. Clustering algorithms like K-means are employed to identify patterns within the data. The study integrates these predictive models with survey data collected from undergraduate students at Heriot-Watt University Dubai, aiming to identify factors that influence academic outcomes. The research uses comparative analysis across different machine learning models which is applied to both the survey data and Kaggle test score dataset. The analysis reveals that linear regression is the most effective model for the Kaggle test score dataset, while K-means clustering provides the best insights from the survey data. The survey model is determined to be more comprehensive due to its inclusion of more predictors. The findings contribute to understanding how data-driven approaches can support educational decisions and interventions while addressing ethical considerations and inclusivity in educational settings.

Keywords: AI in Education, Educational Data Mining, Machine Learning in Education, Predictive Modeling, Test Scores