MLKD 2024
The First International Conference on
Machine Learning and Knowledge Discovery
Amirkabir University of Technology, December 18-19, 2024

Find Effective Features and Optimal Algorithms for Predicting Student Performance in Mathematics and Science

Mahboubeh Molavi-Arabshahi* Narges Ghanbari†

Abstract

In this paper, using data from the 2019 TIMSS (Trends in International Mathematics and Science Study) test in Iran, we aim to find a method to predict students' scores in two subjects: mathematics and science. The major challenge with this database is the large number of features related to schools, teachers, and students, which makes finding and training a model for this classification difficult. Therefore, by applying feature selection and dimensionality reduction techniques, first on each type of feature and then on a dataset composed of all selected features, we attempt to find the best classification algorithm and the set of most impactful features on this performance. Finally, using the RFE (Recursive Feature Elimination) method for dimensionality reduction and three different classification algorithms, we identify 13 features that influence student scores. Additionally, we determine the algorithm that performs best on these selected features, which, in this research, is the gradient boosting algorithm.

Keywords: TIMSS 2019, Educational Data Mining, Feature Selection

*School of Mathematics and Computer Science, Iran University of Science and Technology, molavi@iust.ac.ir
†Department of Mathematical Sciences, Sharif University of Technology, narges.ghanbari81@sharif.edu