A Comparative Analysis of Classification Models with Wrapper-Based Feature Selection for Predicting Student Academic Performance

Authors : Abdullah Al Farwan, Ya Zhang

Abstract : In today's educational arena, it is critical to understand educational data and be able to evaluate important aspects, particularly data on student achievement. Educational Data Mining (EDM) is a research area that focusing on uncovering patterns and information in data from educational institutions. Teachers, if they are able to predict their students' class performance, can use this information to improve their teaching abilities. It has evolved into valuable knowledge that can be used for a wide range of objectives; for example, a strategic plan can be used to generate high-quality education. Based on previous data, this paper recommends employing data mining techniques to forecast students' final grades. In this study, five data mining methods, Decision Tree, JRip, Naive Bayes, Multi-layer Perceptron, and Random Forest with wrapper feature selection, were used on two datasets relating to Portuguese language and mathematics classes lessons. The results showed the effectiveness of using data mining learning methodologies in predicting student academic success. The classification accuracy achieved with selected algorithms lies in the range of 80-94%. Among all the selected classification algorithms, the lowest accuracy is achieved by the Multi-layer Perceptron algorithm, which is close to 70.45%, and the highest accuracy is achieved by the Random Forest algorithm, which is close to 94.10%. This proposed work can assist educational administrators to identify poor performing students at an early stage and perhaps implement motivational interventions to improve their academic success and prevent educational dropout.