

## Improve Student Performance Prediction Using Majority Vote Ensemble Model for Higher Education

**Authors :** Wade Ghribi, Abdelmoty M. Ahmed, Ahmed Said Badawy, Belgacem Bouallegue

**Abstract :** In higher education institutions, the most pressing priority is to improve student performance and retention. Large volumes of student data are used in Educational Data Mining techniques to find new hidden information from students' learning behavior, particularly to uncover the early symptom of at-risk pupils. On the other hand, data with noise, outliers, and irrelevant information may provide incorrect conclusions. By identifying features of students' data that have the potential to improve performance prediction results, comparing and identifying the most appropriate ensemble learning technique after preprocessing the data, and optimizing the hyperparameters, this paper aims to develop a reliable students' performance prediction model for Higher Education Institutions. Data was gathered from two different systems: a student information system and an e-learning system for undergraduate students in the College of Computer Science of a Saudi Arabian State University. The cases of 4413 students were used in this article. The process includes data collection, data integration, data preprocessing (such as cleaning, normalization, and transformation), feature selection, pattern extraction, and, finally, model optimization and assessment. Random Forest, Bagging, Stacking, Majority Vote, and two types of Boosting techniques, AdaBoost and XGBoost, are ensemble learning approaches, whereas Decision Tree, Support Vector Machine, and Artificial Neural Network are supervised learning techniques. Hyperparameters for ensemble learning systems will be fine-tuned to provide enhanced performance and optimal output. The findings imply that combining features of students' behavior from e-learning and students' information systems using Majority Vote produced better outcomes than the other ensemble techniques.

**Keywords :** educational data mining, student performance prediction, e-learning, classification, ensemble learning, higher education

**Conference Title :** ICCSGP 2022 : International Conference on Computers, Software, Graphics and Programming

**Conference Location :** New York, United States

**Conference Dates :** August 08-09, 2022