

Educational Data Mining: The Case of the Department of Mathematics and Computing in the Period 2009-2018

Authors : Mário Ernesto Siteo, Orlando Zacarias

Abstract : University education is influenced by several factors that range from the adoption of strategies to strengthen the whole process to the academic performance improvement of the students themselves. This work uses data mining techniques to develop a predictive model to identify students with a tendency to evasion and retention. To this end, a database of real students' data from the Department of University Admission (DAU) and the Department of Mathematics and Informatics (DMI) was used. The data comprised 388 undergraduate students admitted in the years 2009 to 2014. The Weka tool was used for model building, using three different techniques, namely: K-nearest neighbor, random forest, and logistic regression. To allow for training on multiple train-test splits, a cross-validation approach was employed with a varying number of folds. To reduce bias variance and improve the performance of the models, ensemble methods of Bagging and Stacking were used. After comparing the results obtained by the three classifiers, Logistic Regression using Bagging with seven folds obtained the best performance, showing results above 90% in all evaluated metrics: accuracy, rate of true positives, and precision. Retention is the most common tendency.

Keywords : evasion and retention, cross-validation, bagging, stacking

Conference Title : ICKDDM 2023 : International Conference on Knowledge Discovery and Data Mining

Conference Location : Tbilisi, Georgia

Conference Dates : October 02-03, 2023