Abstract

School dropout is a frequent problem in Brazil that has a professional and personal impact. The governing authorities seek to reduce this problem in education. Thus, the identification of the factors that cause the dropout rate and its prediction in higher education institutions are difficult tasks. Therefore, three combined models are proposed that use groupings and regression predict school dropouts in Higher Education Institutions (HEIs) in Brazil. The proposed models make the combination of algorithms, K-means with Linear Regression (LR), K-means with Robust Regression (RR), and Kmeans with Support Vector Regression (SVR). Four classic algorithms for evaluating our combined models (SVR, Bagging, LR, RR) are selected for comparison. The methodology utilized in this work was the Cross-Industry Standard Process for Data Mining (CRISP-DM). A comparative analysis performed with classic algorithms presents the efficiency and reliability of the proposed models for the school dropout problem.
9. Roberta Andrade de Araújo FAGUNDES and Francisco José de Azevêdo CYSNEIROS. Métodos de regressão robusta e kernel para dados intervalares. 2013.
19. OD Oyerinde and PA Chia. Predicting students’ academic performances—a learning
A Combined Model based on Clustering and Regression to Predicting School Dropout in Higher Education Institution


Index Terms

Computer Science          Data Mining

Keywords