DROPOUT-PERMANENCE ANALYSIS OF UNIVERSITY STUDENTS USING DATA MINING

Amelec Viloria, David Martínez Sierra, Martha García Samper, Wilmer Orlando Cadavid Basto, Alberto Roncallo Pichón, Hugo Hernández-Palma, Victoria Diago Orozco, Sadhana J. Kamatkar

Abstract.

Dropout is a rejection method present in every educational system, related to the various selection processes, academic performance, and the efficiency of the system in general, that is, the result of the combination and effect of different variables. In this sense, the dropout of university students related to their academic performance is a matter of concern since several years ago. Academic information is analyzed in order to identify factors that influence students´ dropout at the University of Mumbai, India, by using a data mining technique. The data source contains information provided to the entrance (personal and educational background) and that is generated during the study period. The data selection and cleansing are made using different criteria of representation and implementation of classification algorithms such as decision trees, Bayesian networks, and rules. the following factors are identified as influential variables in the desertion: approved courses, quantity and results of attended courses, origin and age of entry of the student. Through this process, it was possible to identify the attributes that characterize the dropout cases and their relationship with the academic performance, especially in the first year of the career.

Keywords:

University dropout, Data mining, Classification algorithms, University studies