**Determination of factors influencing the academic achievement of first-year university students using data mining methods**

* The research aims to classify students into three groups: the ‘low-risk’ students the ‘medium-risk’ students, and ‘high-risk’ students
* We need to create a database in which every student is described according to characteristics such as their age, their parent’s level of education, their perception of the university environment, etc.
* Idea is to determine if it is possible to predict a decision variable using the explanatory variables which we retained in the model
* Finally, it presents the results of the application of discriminant analysis, neural networks, random forests, and decision trees aimed at predicting those students' academic success.
* One variable in five proved to be correlated (of which more than one-third were very strongly correlated) with university performance. The most correlated is concerned with attendance at courses, previous academic experience mainly concerned with mathematics and study skills, and the estimated chance of success.
* We noted that 20% of our variables showed a significant correlation with academic success. This 20% of variables are found in each of the factor categories proposed in the model. The same is true for the variables used by each of the three methods of prediction that we compared in this research.
* With respect to the results obtained by the methods of prediction, we conclude that the rates of prediction obtained in validation are not remarkable. However, discriminant analysis, and to a lesser extent neural networks and random forests, seem to be able to lead to interesting results, on the condition
* However, in the future we increase the size of our samples from each university, incorporating data from an additional academic year, for example