

Early detection of university students in potential difficulty

A-S. Hoffait ^{*} M. Schyns [†]

This paper presents a novel approach, based on data mining methods, for the identification of freshmen who have a higher probability to face major difficulties to complete their first year. This is obviously a major concern, for the students who would just need some extra or specific help for being able to succeed, but also for the universities who have to maintain a high level of education with limited resources. We focus more specifically on early detection. The goal is to detect these students at registration time based on information easily available at this time; so as to be able to start academic achievement support before the start of the year or, in some cases, to help the student to select the most suitable orientation. We rely on some indicators of past performances and some environmental factors already identified in the literature. Our contribution is also to adapt three data mining methods, namely random forest, logistic regression and artificial neural network algorithms, to reach a high level of accuracy. We refine the conventional classification by creating subcategories for different levels of confidence: high risk of failure, risk of failure, expected success or high probability of success. Our methodology is illustrated on the real case of the University of Liège, a major University in Europe. With our approach, we can identify, with a confidence level of 90%, 10.5% students who will have strong difficulties if nothing is done to help them.

^{*}University of Liege-HEC Management School-QuantOM – Belgium, Email: ashoffait@ulg.ac.be

[†]University of Liege-HEC Management School-QuantOM – Belgium, Email: M.Schyns@ulg.ac.be.