

# Early detection of university students in potential difficulty

A-S. Hoffait <sup>\*</sup>      M. Schyns <sup>†</sup>

Access to the Belgian higher education system is easier and cheaper than in most foreign countries. Moreover, the quality of our Universities is acknowledged and the degrees they deliver could be considered as a must. As a consequence, lots of candidates apply. However, while access is relatively open, it is not a free lunch and a large proportion of candidates do not satisfy the requirements at the end of the first year. This has a cost for the students (fees, one year lost, frustration...) but also for the collectivity (the Universities are funded by the Government) and for the Universities (lots of ressources are needed). It is therefore important to try to identify as early as possible the students that could potentially be in difficulty during the first year. The Universities might then take appropriate measures to alleviate the problem. There are indeed various reasons that could explain a failure and some that could be dealt with : a wrong orientation (e.g. due to a lack of correct information), a misinterpretation of the requirements and expectations of a university degree, a huge difficulty in making the transition from the high education system to the higher education system, an insufficient mastery of some prerequisite concepts... The University of Liège, as other Universities, has already taken different initiatives. But, if we were able to early identify with a high probability those students, the Universities might develop adapted methods to attack the problem with more emphasis where it is more needed and when it is still possible.

Our contribution is multiple. First, the aim is to develop a decision tool able to identify the students who have a high probability to face difficulties if nothing is done to help them. For that, we consider three standard data-mining methods : logistic regression, artificial neural networks and decision trees and focus on early detection, i.e. before starting at the University. Secondly, we suggest to adapt these three methods as well as the classification framework in order to increase the probability of correct identification of the students. In our approach, we do not restrict the classification to two extreme classes, e.g. failure or succes, but we create subcategories for different

---

<sup>\*</sup>University of Liege-HEC Management School-QuantOM – Belgium, Email: ashoffait@ulg.ac.be

<sup>†</sup>University of Liege-HEC Management School-QuantOM – Belgium, Email: M.Schyns@ulg.ac.be.

levels of confidence : high risk of failure, risk of failure, expected success or high probability of success. The algorithms are modified accordingly and to give more weight to the class that really matters. Note that this approach remains valid for any other classification problems for which the focus is on some extreme classes ; e.g. fraud detection, credit default... Finally, simulations are conducted to measure the performances of the three methods, with and without the suggested adaptation. We check if the factors of success/failure we can identify are similar to those reported in the literature. We also make a “what-if sensitivity analysis”. The goal is to measure in more depth the impact of some factors and the impact of some solutions, e.g., a complementary training or a reorientation.