

# Early detection of university students in potential difficulty

A-S. Hoffait \*      M. Schyns †

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, 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 these students but also for the collectivity and the Universities. 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 University of Liège, as other Universities, has already taken many initiatives. But, if we were able to early identify with a high probability those students, they might develop adapted methods to attack the problem with more emphasis where it is more needed and when it is still possible.

Several papers have already addressed the same problem (e.g. [1]). Here, our contribution comes from the use of tools that are usually exploited in operation research. As case study the cohort of starting students at the University of Liège from year 2008 is analysed. Our contribution is multiple.

A decision tool is developed in order to identify on an early stage, the students who have a high probability to face difficulties if nothing is done to help them. Three standard datamining methods are considered : logistic regression, artificial neural networks and decision trees. Following preliminary results, the classification framework is modified in order to increase the probability of correct identification of the students. The classification is no longer restricted to two extreme classes, e.g. failure or success, but subcategories are constructed for different levels of confidence : high risk of failure, risk of failure, expected success or high probability of success. The algorithms are modified accordingly and are designed in order 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... The three classification techniques are compared on the data base and the results are discussed in

---

\*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.

the light of the literature review [2]. Sample selection bias is also examined since some students had to be removed from the analysis because they did not fill in questionnaire about the independent factors [3]. Finally, a “what-if sensitivity analysis” is performed. The goal is to measure in more depth the impact of some factors and the potential impact of some possible solutions, e.g., a complementary training or a reorientation.

**Keywords.** Datamining, logistic regression, artificial neural networks, decision trees, factors of success, case of study

## Références

- [1] ARIAS ORTIZ, E., AND DEHON, C. What are the factors of success at university ? A case study in Belgium. *CESifo Economic Studies* 54, 2 (2008), 121–148.
- [2] OLINSKY, A., QUINN, J., SCHUMACHER, P., AND SMITH, R. A comparison of logistic regression, neural networks and classification trees predicting success of actuarial students. *Journal of education for Business* 85 (2010), 258–263.
- [3] POIRIER, D. Partial observability in bivariate probit models. *Journal of Econometrics* 12, 2 (1980), 209–217.