

La réussite universitaire : prédictions par génération de règles

Nadine Meskens*, Jean-Philippe Vandamme*, Abdelhakim Artiba**

*Facultés Universitaires Catholiques de Mons (FUCaM)

151 Chaussée de Binche, 7000 Mons – Belgique

{ meskens, vandamme }@fucam.ac.be

<http://www.fucam.ac.be>

** Ecole de technologie Supérieure (ETS)

1100, rue Notre-Dame Ouest, Montréal (Québec) H3C 1K3 – Canada

artiba@i4e2.org

<http://www.i4e2.org>

Depuis longtemps, l'échec scolaire en première année universitaire alimente bon nombre de débats. De nombreux psychopédagogues ont tenté de le comprendre puis de l'expliquer. De nombreux statisticiens ont quant à eux essayé de le prévoir. Nos recherches visent à établir un modèle permettant de déterminer le plus tôt possible dans l'année le groupe des étudiants de première année sur qui il faut cibler en priorité les ressources pédagogiques dont on dispose afin d'améliorer le taux de réussite. Pour cela, nous avons transposé sous forme de questionnaire les hypothèses posées dans de nombreux modèles théoriques.

Notre objectif final est de pouvoir classer les étudiants en trois groupes : le groupe des étudiants ayant une forte probabilité de réussir l'année ("low risk"), le groupe des étudiants qui peuvent éventuellement réussir moyennant des actions à mener ("medium risk") et le groupe des étudiants ayant une forte probabilité d'échouer (ou d'abandonner) ("high risk"). Cette classification a de moins en moins d'intérêt au fur et à mesure que l'année scolaire avance. Il n'est pas intéressant de devoir attendre février voire avril pour arriver à cibler correctement les étudiants qui ont vraiment besoin de mesures d'accompagnement. Notre objectif est donc bien de proposer une méthode permettant d'identifier les étudiants « à risque » le plus tôt possible dans l'année.

Afin d'atteindre notre objectif, nous avons d'abord dû déterminer quels sont les critères qui vont réellement nous permettre de réaliser des prédictions sur la réussite des étudiants. Une fois arrêté la liste des facteurs susceptibles de conduire au classement de ces étudiants en plusieurs groupes, un questionnaire permettant de relever cet ensemble de facteurs pour chacun des étudiants a été établi. Les questionnaires complétés ont conduit alors à la construction de la base de données qui nous sert à réaliser effectivement nos prédictions. En effet, nous avons besoin d'un tableau de données où chaque étudiant sera décrit selon un certain nombre de critères ou d'attributs tels que son âge, le niveau d'éducation de ses parents, ses perceptions par rapport au monde universitaire qui l'entoure, etc. Ensuite, après avoir récolté via ce questionnaire des données suffisamment nombreuses et diverses, l'objectif a été d'extraire de l'information via des méthodes statistiques ou de fouille de données pur et ainsi permettre la classification des étudiants en trois classes les plus homogènes possibles.

Cette recherche décrit la méthodologie adoptée et les variables qui ont été analysées. Ensuite, sur base de l'ensemble des questions qui ont été posées à 227 étudiants en première année universitaire en Belgique lors de l'année académique 2003-2004, nous montrerons les différents résultats obtenus par les méthodes des ensembles approximatifs et des arbres de décision et nous comparerons leurs performances avec celle de l'analyse discriminante linéaire.