

# Towards to a Predictive Model of Academic Performance Using Data Mining in the UTN - FRRe

DAVID L. LA RED MARTÍNEZ, MARCELO KARANIK, MIRTHA GIOVANNINI,  
REINALDO SCAPPINI

Educational Research Group / Department of Information Systems Engineering  
Resistencia Regional Faculty, National Technological University, Argentine

French 414, (3500) Resistencia, Chaco, Argentina

[laredmartinez@gigared.com](mailto:laredmartinez@gigared.com) [mkaranik@frre.utn.edu.ar](mailto:mkaranik@frre.utn.edu.ar) [meg\\_c51@yahoo.com.ar](mailto:meg_c51@yahoo.com.ar)

[rscappini@gmail.com](mailto:rscappini@gmail.com)

## Abstract

Students completing the courses required to become an Engineer in Information Systems in the Resistencia Regional Faculty, National Technological University, Argentine (UTN-FRRe), face the challenge of attending classes and fulfilling course regularization requirements, often for correlative courses. Such is the case of freshmen's course Algorithms and Data Structures: it must be regularized in order to be able to attend several second and third year courses. Based on the results of the project entitled "Profiling of students and academic performance through the use of data mining", 25/L059 - UTI1719, implemented in the aforementioned course (in 2013-2015), a new project has started, aimed to take the descriptive analysis (what happened) as a starting point, and use advanced analytics, trying to explain the why, the what will happen, and how we can address it. Different data mining tools will be used for the study: clustering, neural networks, Bayesian networks, decision trees, regression and time series, etc. These tools allow differ-

ent results to be obtained from different perspectives, for the given problem. In this way, potential problematic situations will be detected at the beginning of courses, and necessary measures can be taken to solve them. Thereby, the aim of this projects is to identify students who are at risk of abandoning the race to give special support and avoid that situation. Decision trees as predictive classification technique is mainly used.

**Keywords:** academic performance; data warehouses; data mining; predictive models.

## 1 Context

In search for reasons for the high degree of student's loss in early years, the project "Profiling of students and academic performance through the use of data mining", 25/L059 - UTI1719 was proposed and implemented in the course Algorithms and Data Structures of the Information Systems Engineering (UTN-FRRe) (2013-2015). As a result, the profiles of academic performance of students according to their social status, family,

academics background, etc., were established and linked with their academic performance in the aforementioned course.

The project “Way to a Predictive Model of Academic Performance Using Data Mining” UTI3808TC (2016-2018) was started as a follow-up for the referred project, within the scope of Educational Research Group (GIE) of the FRRe. The project is funded by the UTN, and then approved and inserted into the Education Ministry's Incentives Program.

## 2 Introduction

The University now faces the challenge of improving academic quality, focusing not only on the teaching-learning process, but also regarding other variables, such as permanent evaluation processes' systematization [1]. Among these variables, the profile study of academic performance of students stands out.

Academic performance is defined as the subject's productivity, nuanced with its activities, features and the more or less correct perception of the assigned goals [2].

When assessing the academic performance, generally the elements that influence performances are discussed in greater or lesser extent, such as, among others, socioeconomic factors, the breadth of curriculum, teaching methodologies, previous knowledge of students [3].

Several studies have shown that the factor most related to educational quality is the student as co-producer himself, measured by household socioeconomic status where they come from [4], and it has become evident that student productivity is higher for women, for younger students and for

those who come from homes with more educated parents [5].

Also, the contrast between those who work and study and only study has also become evident, showing no differences in academic performance for the two sets [6].

The problem of locating good predictors of future performance to reduce academic failure in graduate programs has received special attention in the US. [7], finding that classification techniques such as discriminant analysis or logistic regression are more suitable than multiple linear regression when predicting academic success/failure.

The diversity of studies on academic performance shows that no single way exists to evaluate it. Therefore, the determination of groups or classes of students is an element to consider in order to establish the causes of the problems related to their performance. Moreover, problems could vary depending on the regional context and social reality in which the student is embedded. That is, no tools can be applied to all areas, and results can not be extensible to account for all possible situations. This clearly shows the need to identify specific profiles in educational institutions, adapting tools to each situation in particular.

So, the need for implementing a mechanism to determine the student's characteristics arises, analyzing the existence of relationships and student behavior patterns that enables the clear definition of student profiles. In order to do so, data mining techniques for descriptive modeling is an alternative to use [8] [9] [10].

In turn, predictive modeling can be used to analyze a database and identify certain essential characteristics about the data set to predict the behavior of some variable [11].

### **3 Lines of Research, Development and Innovation**

Information Systems Engineering (ISI) is taught at the UTN-FRRe. This degree has several first year courses that are specific to the Systems Engineering scope, usually generating the highest student's losses. One of those courses is Algorithms and Data Structures.

Under the project "Profiling of students and academic performance through the use of data mining", we worked on identifying the variables that account for unequal academic performance of students in that course, having developed descriptive models of academic performance as a result; with this year's project "Way to a Predictive Model of Academic Performance Using Data Mining" we'll look for development of predictive models for academic achievement.

Given the results in assessments made during the course teaching, we focused on determining to what extent the unequal academic performance is influenced by socioeconomic and attitudinal variables such as: Middle School procedence, educational level of parents, socioeconomic status, age, gender, generally perspective to the study, use of support tools (virtual campus) [10].

The universe consisted of students able to attend the course in 2013, 2014 and 2015

and the unit of analysis was each of those students.

The analysis of the results was based on considering as a mining parameter the final status of the student, which reflects his/her course status at the end of the school year. Students who did not approve nor midterms or recuperatory exams were considered "Free" students; those who managed to pass 3 partial examinations (with or without a recuperatory instance) with a grade greater than or equal to 60% but did not reach at least 75% in all of them were considered "Regular" students; and those who approved all partial greater than or equal note 75% were considered "Promoted" students.

### **4 Results and Objectives**

The followings are the results of the previous project, and also the starting point for the beginning project.

We have obtained the following results: 81.42% of students in "Free" condition, 10.62% "Regular" students and only 7.96% "Promoted" students.

In the following comments it will be consider as "high" academic performance to that achieved by students with final status of "Promoted"; "medium" performance by students with situation of "Regular"; and those with "Free" situation have achieved "low" performance; in turn "academic success" is considered to achieve "high" and "medium" performance, and "academic failure" equals "low" performance.

Then, some of the aspects that is considered appropriate to note will be discussed:

- Given the type of Mid School the students come from, it was observed that for all categories of academic achievement, most students come from schools of provincial and municipal level, but with significant differences in percentage according to academic performance: high 78%, average 67% under 61%. This indicates that the type of secondary school the student comes from is related to his/her academic performance, noting that the highest percentage of participation of public schools (both city and county state) belongs to the category of higher academic performance.

- Considering the amount of hours per week that students dedicated to study, it was found that 56% of those with high academic performance have spent more than 20 hours a week studying. That percentage drops to 50% for the average academic performance and to 46% for underachievement. In addition, 22% of those with high academic performance have spent up to 10 hours per week to study; the percentage increases to 27% for low academic performance. This indicates a direct relationship between the dedication to study and academic success.

- Considering how important is the study for students, 89% of those with high academic achievement have given more importance to study than fun, that percentage drops to 50% for the average academic performance and 64 % for poor academic performance, being 64.6% for the total population. In addition, 11% of those who have high academic achievement have given more importance to study than work, that percentage increases to 33% for the average academic per-

formance and 18% for poor academic performance. This indicates a relationship between academic success and the importance given to the study in relation to fun and work.

- Regarding the mother's latest studies (the highest level achieved), it was observed that 22% of students with high academic performance have mothers with postgraduate studies; that percentage is reduced to 7% for the low academic performance, being 7.08% for the total population. In addition, 33% of those with high academic performance have mothers with university degree, and that percentage decreases to 25% for the average academic performance and 17% for low academic performance. This indicates a relationship between academic success and the level of education attained by the mother.

- Regarding the father's latest studies (highest level) it was observed that 11% of those with high academic performance have fathers with postgraduate studies, that percentage is reduced to 1% for low academic performance, being 1.77% for the total population. In addition, 44% of students with high academic performance are children of fathers with university degrees, the percentage goes down to 25% for the average performance and 21% for under academic performance. This indicates a relationship between academic success and the education degree the father has obtained.

- Taking into account the views of students on the use of ICT, it was observed that 56% of those with high academic performance felt that ICTs facilitate the teaching process, that percentage drops to

50% for average performance academic and 53% for low. In addition, 33% of students with high academic performance considered that mastering ICT is essential for professional practice, that percentage rises to 42% for medium and low academic performance. This would indicate that most students with higher academic achievement would be concentrated more in the teaching/learning process than the possible future work on his/her field of expertise.

All the above lets us conclude that academic success (and failure) are related to the type of high school the student has attended, dedication of the student measured in weekly hours of study, relative importance given to study versus fun and work, and educational level of parents [10].

*Overall goals of the new project*

Obtain the necessary knowledge to develop a predictive model to determine potentially problematic academic performance, from the implementation of data mining techniques.

*Specific objectives of the new project:*

- Use techniques of Data Mining on Data Warehouse built with data from students surveyed during 2013, 2014 and 2015 and extend the survey of student data for 2016 and 2017 to gain knowledge of variables with the more influence in academic performance.
- Analyze and observe the model's behavior under typical conditions of academic performance, obtaining the necessary knowledge to debug the predictive model.
- Validate the model by follow-up of selected student-type, obtaining a reliable predictive model.

- Using the developed predictive model of academic performance to predict the probability that a student leaves the course, given its socio-economic and academic characteristics.

## 5 Human Resources Development

The team is composed of a Director (Doctor, Category II IP, Category A UTN) a Co-Director (Doctor, Category IV IP, Category C UTN), a researcher (Specialist) doing her master's thesis, a researcher (Magister) and two fellows. We are currently working on the definition of master's thesis plans related to the theme project.

### Acknowledgements

This work has been done within the framework of the accredited research project coded UTI3808TC, of the Secretariat of Science, Technology and Graduate of the National Technological University, Argentina.

Our special thanks must go to the Ing. Noelia Pinto by the revision of the English version of this paper.

### References

- [1] Briand, L. C., Daly, J., y Wüst, J. (1999) A unified framework for coupling measurement in object oriented systems, **IEEE Transactions on Software Engineering**, 25, 1.
- [2] Maletic, J. I., Collard, M. L., y Marcus, A. (2002) Source Code Files as Structured Documents, in **Proceedings 10th IEEE International Workshop on Program Compre-**

- hension** (IWPC'02), Paris, France, pp. 289-292.
- [3] Marcus, A. (2003) **Semantic Driven Program Analysis**. Kent State University, Kent, OH, USA, Doctoral Thesis.
- [4] Maradona, G. y Calderón, M. I. (2007) Una aplicación del enfoque de la función de producción en educación. **Revista de Economía y Estadística**, Universidad Nacional de Córdoba, XLII. Argentina.
- [5] Porto, A. y Di Gresia, L. (2003) **Características y rendimiento de estudiantes universitarios**. El caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata. Documentos de Trabajo, Universidad Nacional de La Plata.
- [6] Reyes R, S. L. (2004) El Bajo Rendimiento Académico de los Estudiantes Universitarios. Una Aproximación a sus Causas. **Revista Theorethikos**. Año VI, N° 18, enero-junio, El Salvador.
- [7] Wilson, R. L.; Hardgrave, B. C. (1995) Predicting graduate student success in an MBA program: Regression versus classification". **Educational and Psychological Measurement**, 55, 186-195. USA.
- [8] La Red Martínez, D. L.; Podestá, C. E. (2014) Data Mining to Find Profiles of Students; Volume 10 – N° 30; **European Scientific Journal** (ESJ); pp. 23-43; ISSN N° 1857-7881; University of the Azores, Portugal.
- [9] La Red Martínez, D. L.; Podestá, C. E. (2014) Contributions from Data Mining to Study Academic Performance of Students of a Tertiary Institute; Volume 02 – N° 9; **American Journal of Educational Research**; pp. 713-726; ISSN N° 2327-6126; U.S.A.
- [10] La Red Martínez, D.; Karanik, M.; Giovannini, M. Y Pinto, N. (2015) Academic Performance Profiles: A Descriptive Model Based on Data Mining; Volume 11 – N° 9; **European Scientific Journal** (ESJ); pp. 17-38; ISSN N° 1857-7881; University of the Azores, Portugal.
- [11] Connoly, T. y Begg, C. (2005) **Sistemas de bases de datos. Un enfoque práctico para diseño, implementación y gestión**. 4ta Edición. Ed. Pearson Addison Wesley.