# Academic Performance Profiles In Algorithms and Data Structures of UTN - FRRe

DAVID L. LA RED MARTÍNEZ, MARCELO KARANIK, MIRTHA GIOVANNINI, NOELIA PINTO
Resistencia Regional Faculty, National Technological University, Argentine
French 414, (3500) Resistencia, Chaco, Argentine
laredmartinez@gigared.com

## Abstract

Academic performance is a critical factor considering that poor academic performance is often associated with a high attrition rate. This has been observed in subjects of Algorithms and Data Structures of Information Systems Engineering career (ISI) of the National Technological University, Resistencia Regional Faculty (UTN-FRRe), situated in Resistencia city, province of Chaco, Argentine, where the poor academic performance is observed at very high rates (between 60% and about 80% in recent years). In this paper, we propose the use of data mining techniques on performance information for students of the subject mentioned, in order to characterize the profiles of successful students (good academic performance) and those that are not (poor performance). This article describes the data models and data mining used and the main results are also commented.

**Keywords**: academic performance profiles; data warehouses; data mining, knowledge discovery in databases.

## 1 Introduction

Clearly, academic performance is a critical factor take into account that, frequently, underachievement is associated with a high dropout rate. This is precisely what has been repeatedly observed in subjects of the first level of the Engineering in Information Systems career (ISI) of the National Technological University Resistencia Regional Faculty (UTN-FRRe), located in the city of Resistencia, Chaco province, Argentine, including Algorithms and Data Structures where underachievement is observed at very high rates (between 60% and 80% in recent years).

Specifically, academic performance is defined as the productivity of the individual, qualified by their activities, features and more or less correct perception of the assigned tasks [1]. Academic performance is affected by a multitude of heterogeneous factors (internal and external) that influence student performance. Profiling is a widespread activity in many areas, and it is analogous to the process of identifying and classifying patterns. The information is organized in large data warehouses (DW) that, after pre-cleaning, it is analyzed by algorithms that perform data mining (DM).

In this paper, we propose the use of DM techniques on information about student performance of the Algorithms and Data Structures professorship, in Information Systems Engineering career that is dictated in the Resistencia Regional Faculty at the National Technological University (Chaco, Argentine). This article is structured as follows: in Section 2 are detailed concepts and works related to the measurement of academic performance. The concepts related to DW and DM are presented in Section 3. In Section 4 are described the scope of the proposal and the model used. In Section 5 the results are shown. Finally, in Section 6 some conclusions are presented in relation to the work done.

## 2 Academic Performance

There are several ways to assess student achievement. In general, it involves determining the actual production of a student regarding formal activities. Another way is to use indicators such as graduation rates, differentiated by types of institutions and analyzing student achievement from individual data [2] or through entry qualifications to university, performing the analysis of data using the statistical technique of ROC (Receiver Operating Characteristic) curve [3]. The cognitive aspects were the basis of the early research on the learning process; after researchers discovered the importance of affective components and their decisive influence on learning [4].

It has been shown in several studies that the most related factor to educational quality are the students themselves, measured by household socioeconomic status where they come from [5] and it has shown that students productivity is higher for women, for younger students and those from households with more educated parents [6], having great importance the relationship between hours worked and academic performance [7].

It has also been shown that variables such as study planning, intelligence, teacher support, study, time, environmental conditions of study, and involvement were part of the prediction equation of multiple regressions, which explain 25.70% of variance of academic performance in high school [8]. The diversity of studies on academic performance shows that there is no single way to evaluate it. Moreover, problems can vary depending on the regional context and the social reality in which the student is inserted. This clearly indicates the need to identify profiles in specific educational institutions by adapting the tools to each particular situation.

## 3 Data Handling

The correct data organization added to a suitable model of managing them, can provide a clear view of the drawbacks in the performance of students. In this sense, there are tools in the area of Artificial Intelligence, specifically to the Business Intelligence (BI), such as Data Warehouses (DW) and Data Mining (DM), used to discover hidden knowledge in large volumes of data that can be used to determine patterns and profiles properly.

A DW is a collection of data-oriented topics, integrated, nonvolatile, time variant, which is used to support the process of managerial decision making [9] [10] [11] [12].

The methodologies to be followed for the development of DW depend largely on the size of DW to create and the promptness with which the DW is required [13] [14].

DM is the stage of knowledge discovery in databases (KDD). It is the consistent use of specific algorithms that generates a list of patterns from pre-processed data [15] [16] [17]. DM is closely linked to the DW since they provide historical information with which mining algorithms obtain the information needed for decision-making [18]. It also allows extracting patterns and trends to predict future behavior [19] [20]. Using DM, descriptive and predictive models can be generated [21] [22].

Currently there are several DM methodologies; the most widespread are the CRISP-DM and SEMMA [23] [24]. With these methodologies we try to explain the behavior of certain variables and to identify relevant issues within the academic performance. More detailed description can be seen in [25].

In this context, it is considered as academic performance, the results achieved in the assessments made during the course completed in 2013 (loading, filtering and information processing was performed in 2014). Low, medium, and high student profiles of achievement were searched using data mining on a data warehouse.

In similar work [26], it was proposed a model of data analysis that integrates academic and contextual information.

As it was mentioned in the previous section, the overall objective was to determine the variables that explain the unequal academic performance. To achieve this, the following activities are performed: a) gather information on the current situation regarding the academic performance of students, b) filter and debug information in the current databases, c) establish the relevant variables to describe the situation under study, d) determine how it affects each of the variables that were set to assess the situation of the student, e) determine how it affects each of the variables that were set to evaluate the academic context, f) establish actions aimed at improving indices academic performance of students.

Using the User-Driven technique we have pursued to determine performance profiles (low, medium and high) based on the results obtained by students in assessments, and then, relationships and correlations were look for among the variables mentioned in the previous section.

In the first instance, the information for the students was taken from the database of the academic system, from which the specific data of students and their grades were extracted, those that were considered as indicators of academic performance. Data under the socio-economic situation of the student and his family, as well as attitudinal aspects regarding the study and ICT were collected through a survey using a system of forms in an application online. This information was preprocessed, making a cleanup of inconsistent and missing data. The universe was made up of students able to study the subject in 2013 (we are working on the reporting burden of the course of 2014 and previous years, about 300 students per year) and the unit of analysis was each of those students.

DW structure, as it is shown in Figure 1, consists of a fact table and several dimension tables. The fact table includes specific student information and academic performance, while the dimension tables contain information that makes the description of socio-economic background of the student and family, their academic background in high school, and their attitude towards study and towards ICT.

The scheme of work was similar to described in [27], adapted to the particularities of the UTN - FRRe. The developed actions are indicated in the following section, where DM processes performed are commented, and also the main results obtained.

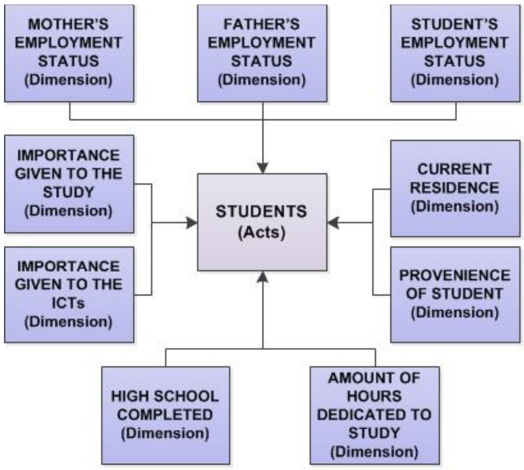


Figure 1. DW model used.

## 5 Detection of Academic Performance Profiles

The work done was divided into several stages such as selection, purification and data preparation, data mining, description of the results, which are explained below.

The selecting stage is characterized by the following. At this stage, different sources of information were selected which served as the basis for data mining stage. As a source of internal information, it was used the information from the corporate and professorship database, where the qualifications of the partial exams of students and their condition at the end of the completed study (Free, Regular, Promoted) are stored. To obtain external information was decisive direct participation of the student, because it was necessary to know information about personal issues that could not be achieved otherwise. Academic, socioeconomic and attitudinal data obtained in the above manner were used for the construction of DW which is then used for data mining processes.

The stage of purification and data preparation explained below.

The quality of the patterns obtained with data mining is directly proportional to the quality of the data used [28]. Based on this, the objective of this stage was the detection, correction, and removal of anomalous data.

Once refined information obtained by each student, we proceeded to the manual loading of qualifications corresponding to the three midterms, examination recovery and the final condition of the student at the end of the course.

As a final activity at this stage, and with full information, we proceeded to load DW. At the end of the process and before starting the next phase, 242 records were available.

The stage of data mining explained below. At this stage, it was selected DM techniques to use, creating corresponding mining flows, in which, the respective algorithms are parameterized. At first, we have started with the supervised classification technique with decision trees.

The analysis of the results was based on consideration as mining parameter the variable related to the final situation of the student, which reflects their status in the matter at the close of school term. It was considered as *Free* students, those that not approved neither midterms nor tests to recover; *Regular*, who managed to approve 3 exams (by retrieving them or not) with greater than or equal to 60% note, but did not reach at least 75% in all cases. Finally, the students in the *Promoted* state are those who approved all partial greater than or equal to 75% note.

Taking into account the above, we have obtained the following results: 81.42% of students in Free condition, 10.62% as Regular student, and only 7.96% as Promoted student. Thus, and always by basing the analysis according to Status parameter, it was considered different criteria for grouping data for the description of classes: dependence of secondary school, number of hours dedicated to the study, importance given to study, academic level of their mother, academic level of their father and use of ICT.

Finally, it is important to refer to the overall quality of the model used to classify the Final Status Student, which turned out to be 0.944, meaning that when estimating the situation based on the variables considered in the model, the estimate is correct in 94.4% of cases.

## 6 Conclusions

The processes of educational data mining made have produced a considerable volume of information, whose detailed study will consume a considerable amount of time, not only to the members of the research project but other areas since, as it is supposed, academic performance is influenced by socioeconomic and cultural background of the students and attitudinal aspects of them regarding the study and use of ICT.

In the following comments are considered *high* academic performance to that achieved by students with final status of *promoted*, *medium* performance to students with situation of *regular*, and performance *low*, the situation of students with *free*; at the same time it will be considered *academic success* to *high* and *medium* performance; and *academic failure* to *low* performance.

Considering the type of secondary school from which students come, it was observed that for all categories of academic achievement most students come from School of Provincial and Municipal level, but with significant differences in the percentages as, high academic performance: 78%, middle 67%, and low 61%. The highest percentage of participation of schools Provincial and Municipal level (State) falls under the category of higher academic performance.

In the face of the amount of hours per week that students devoted to the study was observed that 56% of those who have high academic performance have spent more than 20 hours per week to study, this percentage drops to 50% for the medium academic performance and 46% for poor academic performance. This indicates a direct relationship between the dedication to study and academic success.

Considering the importance that students give the study was observed that 89% of those who have high academic achievement have given more importance to study than fun. This percentage drops to 50% for the medium academic performance and 64% for poor academic performance. This indicates a relationship between academic success and the importance given to the study before the fun.

In consideration of recent studies of the mother (the highest level), it was observed that 22% of those who have high academic performance have mothers with postgraduate studies, this percentage is reduced to 7% for poor academic performance, being 7.08% for the total population. In addition, 33% of those who had a high academic performance are children of mothers with completed university studies, this percentage decreases to 25% for the medium academic performance and 17% for poor academic performance. This indicates a relationship between academic success and the level of education achieved by the mother.

Considering recent studies of the father (the highest level), it was observed that 11% of those who have high academic performance have parents with graduate studies, this percentage is reduced to 1% for poor academic performance, being 1.77% for the total population. In addition, 44% of those who had a high academic performance are children of parents with completed university studies, this percentage decreases to 25% for the medium academic performance and 21% for poor academic performance. This indicates a relationship between academic achievement and educational level achieved by the father.

Taking into account the views of students on the use of ICT it was observed that 56% of those who had a high academic achievement felt that it facilitates the learning process, this percentage is reduced to 50% for the medium academic performance, being 53% for poor. In addition, 33% of those who have high academic performance considered that the domain of ICT for professional practice will be essential; this percentage rises to 42% for medium and low academic performance. This would indicate that most students with higher academic performance would be concentrated more on the teaching – learning than in the possible future exercise of the profession.

Clearly, the model presented in this paper is suitable for the determination of profiles and constitutes a valid tool for academic management.

## Acknowledgements:

This work was performed within the framework of the accredited research project: “Determination of profiles of students and academic performance by using data mining”, code 25 / L059 - UTI 1719.

## References:

[1] Forteza, J. (1975) Modelo instrumental de las relaciones entre variables motivacionales y rendimiento. *Revista de Psicología General y Aplicada*, 132, 75-91. España.

[2] García, M. M.; San Segundo, M. J. (2001) *El Rendimiento Académico en el Primer Curso Universitario*. X Jornadas de la Asociación de Economía de la Educación. Libro de Actas, págs. 435-445. España.

[3] Vivo Molina, J. M.; Franco Nicolás, M.; Sánchez de la Vega, M. del M. (2004) Estudio del rendimiento académico universitario basado en curvas ROC. *Revista de Investigación Educativa*, RIE, Vol. 22, Nº 2, 2004, págs. 327-340. España.

[4] Herrera Clavero, F. et al. (2004) ¿Cómo Interactúan el Autoconcepto y el Rendimiento Académico en un Contexto Educativo Pluricultural? *Revista Iberoamericana de Educación*. España.

[5] Maradona, G. & Calderón, M. I. (2004) Una aplicación del enfoque de la función de producción en educación. *Revista de Economía y Estadística*, Universidad Nacional de Córdoba, XLII. Argentina.

[6] Porto, A. & Di Gresia, L. (2000) *Características y rendimiento de estudiantes universitarios*. El caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata. Documentos de Trabajo UNLP, 24.

[7] Fazio, M. V. (2004) *Incidencia de las horas trabajadas en el rendimiento académico de estudiantes universitarios argentinos*.

Documentos de Trabajo UNLP, 52. Argentina.

[8] Marcelo García, C.; Villarín Martínez, M.; Bermejo Campos, B. (1987) Contextualización del rendimiento en bachillerato. *Revista de Educación*, 282, 267-283. España.

[9] Inmon, W. H. (1992). *Data Warehouse Performance*. John Wiley & Sons. USA.

[10] Inmon, W. H. (1996). *Building the Data Warehouse*. John Wiley & Sons. USA.

[11] Simon, A. (1997). *Data Warehouse, Data Mining and OLAP*. John Wiley & Sons. USA.

[12] Gutting, R. (1994) An Introduction to spatial database systems. *VLDB Journal*, 3, 357-399.

[13] Widom J. (1995) *Research Problems in data warehousing*. Conf. Information and Knowledge Management, Baltimore. U.S.A.

[14] Harinarayan V., Rajaraman, A., Ullman, J. (1996) Implementation data cubes efficiently. *ACM SIGMOD Record*, 25 (2), 205 - 216.

[15] Fayyad, U.M.; Grinstein, G. & Wierse, A. (2001) *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann. Harcourt Intl.

[16] Hand, D.J.; Mannila, H. & Smyth, P. (2000) *Principles of Data Mining*. The MIT Press. USA.

[17] Frawley, W., Piatetsky-Shapiro, G. & Matheus, C. (1992) Knowledge Discovery in Databases: An Overview. *AI magazine*, 13(3), 57.

[18] IBM Software Group. (2003) *Enterprise Data Warehousing whit DB2: The 10 Terabyte TPC-H Benchmark*. IBM Press. USA.

[19] Berson, A. & Smith, S. J. (1997) *Data Warehouse, Data Mining & OLAP*. McGraw Hill. USA.

[20] White, C. J. (2001) *IBM Enterprise Analytics for the Intelligent e-Business*. IBM Press. USA.

[21] Agrawal, R.; Shafer, J. C. (1996) Parallel Mining of Association Rules. *IEEE Transactions on Knowledge and Data Engineering*. December 1996. USA.

[22] Ballard, Ch.; Rollins, J.; Ramos, J.; Perkins, A.; Hale, R.; Dorneich, A.; Cas Milner, E. &Chodagam, J. (2007) *Dynamic Warehousing: Data Mining Made Easy*. IBM International Technical Support Organization. IBM Press. USA.

[23] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Renartz, T., Shearer, C., Wirth, R. (1999) *CRISP-DM 1.0.Step-by-step data mining guide*.

[24] Matignon, R. (2009) *Data Mining Using SAS Enterprise Miner*. U.S.A.: Wiley.

[25] La Red Martínez, D. L.; Karanik, M.; Giovannini, M.; Pinto, N. (2014). *Estudio del perfil de rendimiento académico: un abordaje desde Data Warehouring*. 2° Congreso Nacional de Ingeniería Informática / Sistemas de Información - CoNaIISI - 2014; ISSN N° 2346-9927; pág. 604-612; Universidad Nacional de San Luis, San Luis, Argentina.

[26] La Red Martínez, D. L.; Acosta, J. C.; Uribe, V. E.; Rambo, A. R. (2012) Academic Performance: An Approach From Data Mining. *Journal of Systemics, Cybernetics and Informatics*; V. 10 N° 1 2012, págs.66-72; USA.

[27] 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.

[28] Spositto, O., Etcheverry, M., Ryckeboer, H., & Bossero, J. (2010). *Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil*. En Novena Conferencia Iberoamericana en Sistemas, Cibernética e Informática, CISCI (Vol. 29, pp. 06-2).