

Admission tools and academic performance: evidence from a first course in a bachelor's degree in business administration

Juan Manuel López Zafra¹, Ricardo A. Queralt Sánchez de las Matas², Sonia de Paz Cobo³

¹Métodos Cuantitativos, CUNEF, Spain, ²Métodos Cuantitativos, CUNEF, Spain,

³Economía Aplicada I e Historia e Instituciones Económicas, Universidad Rey Juan Carlos, Spain.

Abstract

Admission tools have become imperative means for private schools to handle both limited space and the search of excellence. We use a supervised algorithm to predict the score of admitted students in a private-run Spanish business school. The main target is understanding the effects of the features defined in the admission process to assess both the validity of the process and the final ranking of the student after one year in the school, trying to ascertain what is the best mix of the variables in place to forecast the final score of the students when ending their first year in the BBA; along with the mix, we also want to define the decision rules allowing the best prediction. The results will prove that the present admission process in place is working properly even if some fine tuning could be set in place for an even better performance.

Keywords: *admission tools, competition, private education, business schools, regression trees*

1. Introduction

Admission tools have become imperative means for private schools to handle both limited space and the search of excellence. The competition for the best students in Spain is getting heavier with the presence of different factors: public universities getting better every year, with some of them topping the rankings; a lowering in the waves of students accessing the university due to the demographic crisis; and a hike in prices that makes even more difficult to reach the best students. In the area of business administration, the situation is paradigmatic, but not exclusive. In health care area, the meta-analysis conducted by Campbell and Dickson (1996) showed grade point averages in nursing and related courses as the greatest cognitive predictors of student success; other elements as parental education and age were found as the greatest demographic predictors. The debate on the admissions tests as the SAT one (the used-to-be standard among American universities) is brought by Bollinger (2005). In the area of medical schools, that face large number of applicants with limited student places, Urlings-Strop et al (2013) tried to “determine the relative contribution of the non-academic and academic steps to differences found in student performance”, comparing both lottery-admitted students and three groups of selection procedure participants, and concluding that “the lower dropout rate of selected students is related to both self-selection of participants before the start of the selection procedure and the academic part of the selection procedure.”

In the present paper we present the research on the use of a supervised algorithm, namely a regression tree, to predict the score of admitted students in a private-run Spanish business school. The main target is understanding the effects of the features defined in the admission process to assess both the validity of the process and the final ranking of the student after one year in the school. After showing the scope of the research and the data and methodology in place, the analysis of both the descriptive statistics and the results from the regression tree will prove that the present admission process in place is working properly even if some fine tuning could be set in place for an even better performance.

2. Scope of the research

In order to meet the requirements of the present research, we've accessed the data of the admission process of the students who applied, were selected and finally enrolled Colegio Universitario de Estudios Financieros (CUNEF), a private business school affiliated of Universidad Complutense de Madrid (UCM). As it's usual in Spanish private schools and universities, students must pass an admission process further than the official University Access Test (PAU for its initials in Spanish, aka *Selectividad*). Since 2017, the test has been replaced by the University Access Evaluation (EVAU). Whatever the test passed, in both private schools and universities there are additional admission tools in which students must

show capacities normally not captured in their academic record and exceeding the usual academic skills assessed by the official university access test; anyway, the latter must be passed for a student being granted access in any official degree in Spain.

The admission process in place was built on different elements: an interview, a mathematical proprietary test, a reading comprehension test, a psychological test and the results the students achieved during their 3rd and 4th years of the Compulsory Secondary Education (ESO for its initials in Spanish). Other elements in place were the grade in their first year of the two-year baccalaureate and the type of baccalaureate followed. Given the national nature of the ESO and given that these two courses close a complete period (something the first year of baccalaureate doesn't accomplish), we have considered the average of this two years of ESO will depict a better image of the capabilities of the student, giving up the last two.

At the moment, according to the Universia web portal (<http://www.universia.es/estudios/grados/dg/269>), there are 2981 official bachelor's degrees in Spain; in the specific area of Business Administration, 137 certified schools teach the official bachelor's degree, both as a single degree (102) or along with a second one (35, the so-called double degrees). From the total number of schools teaching the subject, 90 are public and 47 are private-run: 11 in Madrid, 11 in Barcelona, 7 in Valencia, 4 in Alicante, 4 in Guipúzcoa, 3 in Valladolid, 2 in Ávila and 1 in La Rioja. According to the *Avance de la Estadística de Estudiantes* (Students' Statistics Preview) from the academic year of 2017-2018 (http://www.educacionyfp.gob.es/servicios-al-ciudadano-mecd/estadisticas/educacion/universitaria/estadisticas/alumnado/2017-2018_Av/Grado-y-Ciclo.html, Ministerio de Educación y Formación Profesional), during that year there were 96281 students enrolled in some Spanish university in courses conducting to a bachelor's degree in business administration (BBA); 80323 of them (83.4%) through in-person classes. 16.9% (13557) of the last figure were taking courses in private-run universities. In the affiliated centers, all of them with in-person classes, there were 6746 students and 5350 of them (79.3%) in some of the schools affiliated to public universities. If we just pay attention to those in their first year, and just considering the pure BBA studies (so neglecting other similar studies), we're talking about a total of 34144 students during the academic year of 2017-2018; of them, 27320 selected in-person courses and 5381 followed their first-year studies in some school in the autonomous region of Madrid. A tiny 19.8%, that is, 1066 of them, were enrolled in a private school or university (there's no breakdown for affiliated schools).

These figures account for the extreme competition the private schools are facing in the area of business administration. In that market, the demand meets private centers along with some quality public competitors with out-of-the-market prices that do not reflect the actual cost of the offered services. Private centers must then break the wall offering services (including, of course, the service of education, the most important from far but not the only

one) that must be perceived as better; just in that case, those demanding education will be ready to pay, in case they can afford it, an extra cost that can clearly be quite important. The fight for the “golden medal” in higher education has already become an obsession in Europe, as Bollinger (2005) reported for the United States. According to the *Boletín Oficial de la Comunidad de Madrid* (BOCM, 2018), the official tables of public prices for the degree studies with a level 3 in experimentalism (as those of BBA) show that the total cost for a 4-year, 240 ECTS (European Credit Transfer and Accumulation System, the official accounting system into the UE for degree's studies) credits BBA in Madrid (the second most expensive, just after the autonomous region of Catalonia) will add up to 5133.6 euros, in case the student needs no more than a single call and its retake. A single year in CUNEF is priced 10500 euros, the equivalent at Universidad San Pablo – CEU (<https://notasdecorte.es/universidad-ceu-san-pablo>) is 10860 euros, 11624,75 euros at Universidad Pontificia Comillas – ICADE (<https://www.comillas.edu/grados/grado-en-ade-administracion-y-direccion-de-emresas-e2>), and 21000 euros at IE University (<https://www.ie.edu/es/universidad/admisiones/tasas-y-ayuda-financiera/metodos-de-pago/>). The last three share the 28.2% of the total market of first-year students in a BBA in private-run Spanish universities.

3. Objectives and methodology

The targets of the present research are the following: starting from the available information from the admission tests (we leave the interview aside), we will try to ascertain what is the best mix of the variables in place to forecast the final score of the students when ending their first year in the BBA in CUNEF, as later explained; along with the mix, we also want to define the paths or decision rules allowing the best prediction. The main task is enhancing the tests in place and introducing improvements in the admission process, if any.

The research has followed the usual methodology of a data science project: definition of the question to be answered, data exploratory analysis, set-up and tuning of the model, achievement and analysis of the outcomes. The phases are depicted, in one way or another, in Peng and Matsui (2016), Hardoon and Shmueli (2013) or Provost and Fawcett (2013), among many others. Data come from the internal CUNEF data lake, and for the statistical processing we have employed R version 3.4.3 over RStudio version 1.1.414, the last available in the moment the research started; for the present review the versions were 3.5.2 and 1.1.463, respectively.

We have proceeded through classification and regression trees (CART) for discovering the main drivers of the process. Other approaches, such as the one by Hoefler and Gould (2000) were based both on traditional statistical techniques (linear and non-linear regression) and nontraditional – specifically, a neural network. Even if the latter was seeing as powerful by

the authors, we gave it up due to its difficulty in both implementing and understanding. In our case, and given metric variables, the regression trees are of interest. Initially introduced by Breiman, Friedman, Olshen and Stone (1984), this type of tools are likely the simplest to implement and interpret among classification algorithms. On their utility we can name the seminal work by Murthy (1998), the subsequent by Rokach and Maimon (2005) or the very new by Anandajayam and Sivakumar (2018).

Among the advantages of the technique, we must emphasize their ability to handle errors and missing value in the datasets, and the fact that they are nonparametric, meaning that there are no assumptions about the space distribution and the classifier structure. On the other hand, there are some disadvantages, being the most cited the possibility of overfitting due to the high flexibility of the technique (Pandya and Pandya, 2015; Ye and Hou, 2018; Fratello and Tagliaferri, 2019).

4. Analysis

Our dataset consists in the 313 students of the years 2015 and 2016, all of them finally studying a single bachelor's degree in business administration (BBA) in CUNEF, both in its bilingual English-Spanish version or the single Spanish one. For the present research, we considered their gender, their results in the mathematics test, the comprehension reading and the psychological ones (all together properly averaged into the variable NOTAPBAS, *score*), along with their final grades in the 3rd and 4th courses of the Compulsory Secondary Education (ESO for its initials in the Spanish official naming, *Educación Secundaria Obligatoria*, averaged into PBAESP2). The students have been classified according to their results in the 10 different subjects that define the official syllabus of the Universidad Complutense de Madrid (<https://economicasyempresariales.ucm.es/estudios/grado-ade-estudios-estructura>). The final grades of the student in each subject have been weighted according to the call the student has passed the exam: 1 if in the first one, .75 if in the second, .5 if in the third and .25 if in the fourth one. Given that the admission tests under scrutiny took place during the first semesters of years 2015 (for those entering the university September that year) and 2016 (for those entering September 2016), and that the research was originally conducted during the last three months of 2017, the maximum span of any student is two years in the same subject, giving a maximum of four calls for any subject. Those who entered in September 2016 accounted for a maximum of two calls for the same first-year subject.

Once we provided our 313 students with their final grade in the 100-points scale, we proceeded to sort them, defining three groups for the task of the present research: those in the top quartile (top 25), those in the first decile (bottom 10), and those in the middle of them, the so-called main group.

4.1. Descriptive statistics

For the total of the 313 students in the sample, 132 (42.2%) were women; the ratio holds for the students in the bottom 10 (13 of the total 31 were women) but completely reverse in the top 25, where 44 of the total 78 were women, a 56.4%. According to their academic provenance, the outcomes were, again, not surprising. In Spain, there are four different possible baccalaureates (humanities and social sciences, technology, natural and health sciences) that allow a student entering the university; the first one is the usual way to enter a BBA, and that is the case for the 78.3% of the sample versus 69.2% for those in the top 25 and 77.4% in the bottom 10 (non-significant difference from the main group); just 13.7% of the students in the main group came from a technical baccalaureate versus the 23.1% coming from the same one in the top 25 and 10% in the bottom 10; 6.7% from the one in health studies in the main group versus 7.7% in the top 25 and 6.5 in the bottom 10 (non-sig); and if a mere 1.3% from the one in arts live in the main group, there were none in the top 25 and 6.5% in the bottom 10.

The difference arises then between the top quarter of the students and the main group more than from the bottom decile and the latter: in the first case, we have more girls (1.3 vs 0.73 per boy in the main group) and a higher ratio of students coming from more technical baccalaureates: the general ratio of 3.8 humanities and social sciences per science student fall down to 2.25 in the top 25%.

The average score for the students at the end of the first year is 57.07 points (95% confidence interval -95% ci- 55.13 - 59.07) compared to 77.89 in the top quarter (95% ci 76.62 – 79.17) and 23.34 in the bottom decile (95% ci 18.90 – 27.78); their results in the ESO grade are 6.93 (in a 10 points scale, 95% ci 6.83 – 7.03), 7.6 for the top quarter (95% ci 7.38 – 7.83) and 6.28 (95% ci 6.00 – 6.56) for the lower decile. So there clearly exists a difference in terms of ESO grades between the three groups. And the presence of this difference along with the gender, in terms of academic performance, pushed us to analyze the issue through classification and regression trees.

4.2. The regression tree

Starting from our 313 complete observations, we build a regression tree where we try to predict the value of the end-of-year-one score from the values of gender (SEXO), average score in tests (NOTAPBAS) and average of third and yearth courses of ESO (PBAESP2). We used libraries `rpart` (recursive partitioning for classification, regression and survival trees, “an implementation of most of the functionality” of Breiman, Friedman, Olshen and Stone, 1984) and `rpart.plot` (for plotting 'rpart' models libraries). After pruning the tree, for avoiding overfitting, where the complexity parameter (`cp`) meets the minimum cross-validated error (`xerror`), starting at an initial level of `cp_i = 0.174943` we are able to diminish it to a more suitable level of `cp_f = 0.018036`, after just 3 splits. As it's known, the

best cp value is the one that minimizes the prediction error RMSE (root mean squared error). The variables in the final model are just NOTAPBAS and PBAESP2, with an importance of 57% and 43% respectively, showing that the internal tests conducted at CUNEF overperformed the average score of the secondary years. As Figure 1 depicts, for the 7% having a score over 8.2 in the internal tests (bottom right node, D), the predicted average score at the end of year 1 will be 78.52 points (rounded to 79 in the figure). The decision rules are easy then to interpret: apart from the latter, for those in the range of 7.4 to 8.2, the average score will be 64.48 points (node C). The 87 students with proprietary admission test score above 7.4 (nodes C and D) represent 27.8% of the total sample; among them, the minimum PBAESP2 value is 6.36. The lower 10% of this one, accounting 32 students, is built according to one main condition: whatever the score in the admission tests, if the average grade in the 3rd and 4th years of ESO is below 5.8, the average score ending 1st year of the BBA in CUNEF will be 40.12 points out of 100. As we can observe, there's a light difference between the scores in the sample and those (node A) predicted by the regression tree: the lower decile shown an average PBAESP2 of 6.28 points vs the predicted 5.8 by the regression tree.

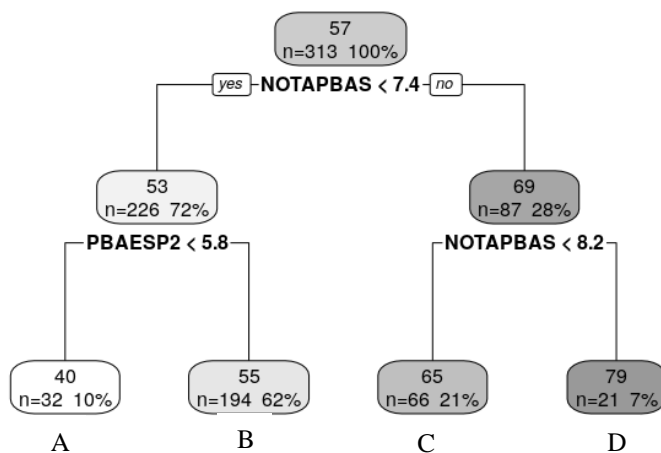


Figure 1. Final regression tree.

5. Conclusion

As we have shown along the present research, regression trees can be of help in the admission process. They are useful in defining the important variables (gender, even if not suitable for use according to a non-discriminating protocol, was wrongly presumed as a keystone) and they present clear, easy-to-follow decision rules for the management. In our case, they have also proven the importance of the admission tests as a feeding element above the official grades the students bring in their portfolio. Some problems have arisen, nevertheless, the most important one being the difference in the average grade that defines the lower decile of students in terms of their compulsory secondary education scores in 3rd and 4th courses. This is certainly due to an unavoidable fact: the small size of the sample, that prevents its correct split into a training set and a test one, essential for building a machine-learning model. Future work, along with a higher size, will observe additional possibilities according to new variables as the sending school; this will help to improve the definition of the decision rules in the pursue of excellence.

References

- Anandajayam, P; Sivakumar, N (2018) The Study on Predictive Analysis Algorithm: Survey. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Volume 3, Issue 1.
- BOCM (2018). Boletín Oficial de la Comunidad de Madrid, 31 de agosto de 2018. Decreto 126/2018, de 31 de julio, por el que se establecen los precios públicos por estudios universitarios conducentes a títulos oficiales y servicios de naturaleza académica en las Universidades Públicas de la Comunidad de Madrid, pp. 9 a 15.
- Bollinger, L. C. (2005). Competition in higher education and admissions testing. In Camara, W, and Kimmel, E. (Eds.) *Choosing Students: Higher Education Admissions Tools for the 21st Century*. Taylor and Francis.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984) *Classification and Regression Trees*. Wadsworth International Group.
- Campbell, A. R.; Dickson, C. J. (1996). Predicting student success: A 10-year review using integrative review and meta-analysis. *Journal of Professional Nursing*, 12(1), 47–59. [http://doi.org/10.1016/S8755-7223\(96\)80074-3](http://doi.org/10.1016/S8755-7223(96)80074-3)
- Fratello, Michele; Taglaferri, Roberto (2019) Decision trees and random forests. In Cannataro, Mario; Gaeta, Bruno, and Khan, Mohammad Asif (Eds) *Encyclopedia of bioinformatics and computational biology*, Vol. 1. Elsevier.
- Hardoon, D.; Shmueli, G. (2013) *Getting started with business analytics*. CRC Press.
- Hoefer, P.; Gould, J. (2000) Assessment of Admission Criteria for Predicting Students' Academic Performance in Graduate Business Programs, *Journal of Education for Business*, 75:4, 225-229, DOI: 10.1080/08832320009599019

- Murthy, Sreerama K. (1998) Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey. *Data Mining and Knowledge Discovery*, Volume 2, Issue 4, pp 345–389
- Pandya, Rutvija; Pandya, Jayati (2015) C5.0 Algorithm to Improved Decision Tree with Feature
- Peng, R.; Matsui, E. (2016) *The Art of Data Science*. Skybrude Consulting, LLC
- Provost, F.; Fawcett, T. (2013) *Data Science for Business*. O'Reilly Media Inc., EE. UU.
- Rokach, Lior; Maimon, Oded (2005) Top-Down Induction of Decision Trees Classifiers—A Survey. *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*. December. DOI: 10.1109/TSMCC.2004.843247
- Selection and Reduced Error Pruning. *International Journal of Computer Applications*, Volume 117, No. 16, Mayo
- Urlings-Strop, L. C., Stegers-Jager, K. M., Stijnen, T.; Themmen, A. P. N. (2013). Academic and non-academic selection criteria in predicting medical school performance. *Medical Teacher*, 35(6), 497–502. <http://doi.org/10.3109/0142159X.2013.774333>
- Ye, Jie; Hou, Li-duo (2018) Improvement and Application of Decision Tree C4. 5 Algorithm. *DEStech Transactions on Computer Science and Engineering*. DOI 10.12783/dtcese/CCNT2018/24686