doi.org/10.26398/IJAS.0031-011

ANALYSIS OF THE PREDICTION ABILITY OF A UNIVERSITY SELF-EVALUATION TEST: STATISTICAL LEARNING METHODS FOR PREDICTING STUDENT PERFORMANCE

Eni Hasa, Leonardo Grilli

Department of Statistics, Computer Science, Applications "G. Parenti"
University of Florence, Italy

Abstract*. The School of Economics and Management of the University of Florence uses a self-evaluation test as an instrument to assess the competencies of candidates who want to enrol in the three-year degree program. The aim of this study is to assess if the self-evaluation test scores give a gain in predicting the student performance when added to available student characteristics, such as the high school career. The student performance is measured by three binary indicators based on the number of credits gained after one year. For each binary outcome, the prediction is carried out using both logistic regression and random forest, using two alternative sets of predictors: (i) student characteristics; (ii) student characteristics and test scores. The predictive ability is assessed using 10-fold cross-validation. The main finding of the analysis, which refers to the academic year 2014/ 2015, is that the self-evaluation test scores do not help in predicting student performance once student characteristics are properly exploited.*

Keywords: *Cross-Validation, Logistic regression, Random Forest, Self-evaluation test, Student performance.*

1. INTRODUCTION

The School of Economics and Management of the University of Florence uses a self-evaluation test as an instrument to assess the competencies of candidates who wish to enrol in the three-year degree program. The test contributes to the orientation for the choice of the university course. Our study has the primary aim of evaluating whether the self-evaluation test provides valuable information for predicting student performance when added to the already available student characteristics. In order to be confident that our findings do not critically depend on the prediction method, we pursue the ancillary aim of implementing and comparing different prediction methods. The ability to predict student performance is pivotal for the university management in the processes of allocating resources and planning actions to support weak students. The literature on predicting university student

performance is large, see Grilli *et al*. (2016) for an overview. Here we consider the case of the School of Economics and Management of the University of Florence, which was analysed for an earlier academic year (2008/2009) using an approach based on quantile regression for counts (Grilli *et al*., 2016) and binomial mixture modelling (Grilli *et al*., 2015). The current paper differs from previous ones because of the focus on prediction rather than statistical modelling.

The self-evaluation test of the School of Economics and Management is compulsory, but it does not preclude the enrolment. It consists of 24 items with multiple responses, one of which is correct. There are three sections on logic, reading (comprehension of a text) and mathematics. The candidates have 20 minutes for each session. A correct answer yields 1 point, a wrong answer yields -0.25 points, whereas no points are assigned if the answer is not provided. Therefore, each candidate has separate scores on logic, reading and math. Candidates with a total score lower than 8 can still enrol, but they have to recover the gap by studying some material provided by the university.

The measures of the student performance are based on the number of credits gained after one year. The credits are defined in accordance with the European Credit Transfer and Accumulation System (ECTS), allowing the comparison between different courses of European universities through an assessment of the required workload to achieve the learning outcomes. A credit usually corresponds to 25 hours of work including lessons, exercises, and home study. Each course has a given number of credits, which are acquired when the students passes the exam. Typically, a year of a degree program corresponds to 60 credits.

In the analysis we consider three alternative outcomes based on the number of credits, namely obtaining at least 1 credit, at least 20 credits, and at least 40 credits. Such thresholds are chosen to match some well-defined targets for the university management. First of all, it is of paramount importance to predict whether a student will get some credits or will fail at all. Moreover, the Italian government adopts the percentage of students obtaining at least of 20 credits in a year as a criterion for allocating a quota of the government funding to the university[1]. For other purposes the threshold of obtaining at least 40 credits is relevant. Therefore, we focus the analysis on the indicators based on three relevant thresholds, namely at least one credit, at least 20 credits, and at least 40 credits.

For each binary outcome about student performance, we carry out the prediction using both logistic regression and random forest, using two alternative sets of predictors: (i) student characteristics; (ii) student characteristics and test

[1] http://attiministeriali.miur.it/anno-2016/dicembre/dm-29122016.aspx

scores. The predictive ability is assessed using 10-fold cross-validation.

The article is structured as follows. In Section 2 we describe the data used for the analysis. In Section 3 we show the results of predicting student performance using logistic regression, whereas in Section 4 we tackle the same task using random forest. In Section 5 we summarize the results and draw some conclusions.

2. DATA DESCRIPTION

The original data set is a merge of the administrative career archive and the test archive. We have 978 observations and 56 variables. We delete 19 observations because they are students with a diploma obtained abroad. Further, 2 observations were deleted because the high school degree was unknown. The variables include information about single exams, but we will only use summary measures over the academic year. The goal is to evaluate the addition of test information on predicting the student performance, besides student's background characteristics. We delete 88 students exempted from the test because they were changing their degree program and already passed a test or acquired at least 18 credits. Therefore, we limit the analysis to 869 students who took the test and enrolled at the School of Economics and Management in the 2014/2015 academic year[2].

The total test score in the dataset ranges from -2.25 to 24, with first quartile 10.75, median 13.25 and third quartile 16. The mean and standard deviation are 13.35 and 3.76, respectively. On the basis of previous findings (Grilli *et al*. 2016), we do not use the total score, but the partial scores in the three sections of the test (Logic, Reading and Mathematics). A few students (5.18%) obtained a total score less than 8, so they can still enrol, but they have to recover the gap by studying some material provided by the university. We could pick up those cases through an indicator, but this indicator turns out to be negligible in predicting the outcome, so it is not used in the analyses presented later on.

As in Grilli *et al.* (2016), we divide the predictors into two groups:

– Student characteristics (measured before the test): Gender (1 if is male), Residence (1 if the student is resident in Florence, Arezzo, Pisa, Pistoia, Prato), High School irregular career (1 if age at high school diploma > 19), High school type (Scientific, Classics, Technical, Other), High school grade (from 60 to 100).
– Test scores: partial scores on Logic, Reading and Mathematics.

[2] https://www.economia.unifi.it/upload/sub/test-autovalutazione/bando-test-autovalutazione-2014-15.pdf

We evaluate the prediction ability of student characteristics and test scores on the performance during the first year. As anticipated in the introduction, we choose to measure student performance by three binary outcomes corresponding to well defined targets for the university management: students obtaining at least one credit, at least 20 credits and at least 40 credits. Given that the first year of the considered degree programs include 6 exams of 9 credits each, we the binary outcomes are defined as in Table 1.

Tab. 1: Definition of the binary outcomes representing student performance

Symbol	*Exams*	*Credits*
Y_1	≥ 1	>0
Y_3	≥ 3	≥ 20
Y_5	≥ 5	≥ 40

The number of passed exams after one year (with frequencies in parenthesis) are 0 (283), 1 (158), 2 (112), 3 (114), 4 (83), 5 (69), 6 (50). The mean is 1.96 and the standard deviation is 1.91. It is worth to note the high number of students passing no exam, which is a common problem in the Italian university system.

Table 2 shows how student characteristics are related to test scores and outcomes. The high school grade, which is the only numerical variable, is divided into quartiles. The relationships have the expected signs, for example test scores and outcomes are better for students who attended a lyceum (classics or scientific) and who obtained a higher grade. In order to interpret the differences in test scores, note that they are expressed in points and the standard deviations are 1.99 for logic, 1.42 for reading and 1.91 for mathematics.

The percentages of the outcomes show substantial variation across the levels of student characteristics. Therefore, student characteristics can be exploited to predict student performance. However, most universities do not trust student characteristics. In particular, the high school grade is considered to be unreliable due to the heterogeneity of the criteria for assigning grades, and the mismatch between the competencies evaluated at high school and those required for a given degree program. Nonetheless, Grilli *et al*. (2015) and Grilli *et al*. (2016) show that the self-evaluation test of the School of Economics and Management of the University of Florence does not help much in improving the prediction of the performance. In order to investigate this issue, here we devise a systematic study of the gain in prediction yielded by the self-evalutaion test over the student characteristics, considering several binary outcomes and using two well known prediction methods, namely logistic regression and random forest.

Tab. 2: Test scores and binary outcomes of performance by student characteristics. School of Economics and Management, University of Florence. Accademic year 2014/2015.

	Test scores					***Outcomes***		
	freq.	*total*	*logic*	*reading*	*math*	*%Y_1= 1*	*%Y_3= 1*	*%Y_5= 1*
All students	869	13.35	4.98	4.63	3.74	67.4	36.4	13.7
Gender								
Female	363	12.44	4.48	4.47	3.49	69.1	35.8	11.8
Male	506	14.00	5.34	4.75	3.91	66.2	36.8	15.0
Far-away resident								
Yes	145	12.57	4.66	4.67	3.27	60.4	26.4	9.0
No	724	13.51	5.05	4.62	3.83	68.8	38.3	14.6
High School type								
Scientific	297	14.78	5.27	4.88	4.64	75.4	45.1	21.2
Classics	67	13.60	5.26	4.85	3.49	80.6	44.8	16.4
Technical	327	12.64	4.89	4.53	3.23	63.6	32.1	9.7
Other	178	12.16	4.57	4.32	3.27	56.2	26.4	7.3
High School late graduation								
Yes	108	12.25	4.72	4.54	3.00	45.3	14.8	4.6
No	761	13.51	5.02	4.64	3.84	70.6	39.4	15.0
High School grade								
1st quartile	268	12.28	4.54	4.29	3.44	48.9	16.0	2.6
2nd quartile	180	12.93	4.73	4.66	3.53	64.4	26.1	7.8
3rd quartile	208	13.73	5.23	4.63	3.86	76.9	43.8	17.3
4th quartile	213	14.69	5.51	5.02	4.15	84.0	63.4	29.1

3. LOGISTIC REGRESSION

We use logistic regression to predict student performance measured by the three binary outcomes Y_1, Y_3, Y_5 defined in Table 1. First we use only student characteristics (measured before the test), then we add the self-evaluation test scores and assess the gain in prediction. For any binary outcome, we denote with p_i the success probability of student i and with x_{ij} the predictor j of student i. Logistic regression (Agresti 2015) is defined by

$$\pi_i = \frac{\exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)}{1+\exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)} \tag{1}$$

We fit logistic regression with the `glm` command of the software R using default settings (maximum likelihood through iteratively reweighted least squares).

3.1 MODEL FITTING

We work on the entire dataset with 869 students who did the self-evaluation test and enrolled in the academic year 2014/2015. We fit models separately on the three binary outcomes Y_1, Y_3, Y_5 defined in Table 1. We fit two models for each binary outcome: the first model includes only student characteristics, while the second model also include test scores. The estimated coefficients are displayed in Table 3, along with asterisks to denote statistically significant predictors. There are no interactions among the predictors as they are not statistically significant.

Tab. 3: Logistic regression: output for each of the three binary outcomes defined in Table 1. First model with student characteristics, second model with student characteristics and test scores. School of Economics and Management, University of Florence. Accademic year 2014/2015.

	Y_1		Y_3		Y_5	
Student characteristics						
Intercept	***-4.12	***-4.38	***-7.72	***-8.54	***-11.11	***-11.87
Male	0.06	0.01	*0.39	0.21	**0.67	0.50
Far-away residence	**-0.60	*-0.54	***-0.85	**-0.73	*-0.77	-0.58
High School (HS) type						
scientific (baseline)	-	-	-	-	-	-
classics	0.24	0.35	-0.10	0.08	-0.53	-0.24
technical	***-0.80	**-0.66	***-0.98	**-0.64	***-1.49	***-1.05
other	***-1.09	***-0.97	***-1.19	***-0.87	***-1.66	**-1.20
HS late graduation	**-0.65	**-0.62	**-1.01	**-0.95	-0.76	-0.63
HS grade	***0.07	***0.07	***0.10	***0.09	***0.11	***0.10
Test scores						
Logic		-0.00		*0.11		0.05
Reading		0.03		0.04		0.02
Mathematics		*0.10		***0.18		***0.27

$^{*}p$-value<0.05; $^{**}p$-value<0.01; $^{***}p$-value<0.001

As for student characteristics, in all six models the high school grade has a positive and significant effect, and students from technical and other high schools perform significantly worse than those from scientific high schools. The other predictors are not always significant, but their estimated coefficients have the same sign in all six models: the performance is higher for males and lower for students with far-away residence and for students with an irregular high school career. Among the scores of self-evaluation test, the Mathematics score always has a positive and statistically significant effect, whereas the Logic score is significant only for Y_3 and the Reading score is never significant. It is worth to note that the high school grade has an effect much larger than the test scores: for example, on

the logit scale for Y_1, an increase of one standard deviation of the high school grade amounts to adding 0.07×11.25=0.788, whereas an increase of one standard deviation of the mathematics score (the one with the largest coefficient) amounts to adding 0.10×1.91=0.191.

3.2 PREDICTION ERRORS

In order to evaluate the prediction ability, we cross-classify the binary response y with the binary prediction $\hat{y}$. There are four possible outcomes:

– True Positive (TP) if y=1 is classified as $\hat{y}$=1
– False Negative (FN) if y=1 is classified as $\hat{y}$=0
– True Negative (TN) if y=0 is classified as $\hat{y}$=0
– False Positive (FP) if y=0 is classified as $\hat{y}$=1

The true positive rate (or sensitivity) of a classifier is the proportion of correctly classified positives over total positives P($\hat{y}$=1|y=1); on the other hand, the false positive rate (or 1-specificity) is the proportion of incorrectly classified negatives over the total negatives P($\hat{y}$=1|y=0). The overall prediction error is:

$$\text{PE} = \frac{\text{FN} + \text{FP}}{\text{TP+FN+TN+FP}} \tag{2}$$

The prediction of the outcome y_i is requires a cut-off π_0 so that $\hat{y}_i = 1$ if $\hat{\pi}_i > \pi_0$ and $\hat{y}_i = 0$ otherwise. We use the standard cut-off equal to 0.5. The cut-off can be changed to control FN and FP rates if one of the two rates is too large.

We assess the predictive ability of the classification method by estimating the overall prediction error (2) through 10-fold cross-validation (Hastie *et al*., 2009). The folds are determined separately for each outcome (Y_1, Y_3, Y_5) as follows: (*i*) observations are divided in two strata according to the outcome (0 or 1); (*ii*) within each stratum, observations are randomly assigned to the folds. This procedure ensures that all folds approximately have the same proportions of 0's and 1's. The obtained folds are used for all the methods.

Table 4 displays the average prediction error from 10-fold cross-validation for logistic regression, separately for each binary outcome and type of predictors (student characteristics with or without test scores).

For both types of predictors, the highest prediction error concerns Y_1 (passing at least 1 exam). Therefore, the most difficult task is to predict which students will fail. The self-evaluation test scores do not help to reduce the prediction error, which even increase by more than 3 points in the critical task of predicting Y_1. It is also

Tab. 4: Average prediction errors from 10-fold cross-validation for logistic regression exploiting student characteristics with or without test scores. Overall error (false positive and false negative errors in parenthesis, respectively). School of Economics and Management, University of Florence, academic year 2014/2015.

Type of predictors	Y_1	Y_3	Y_5
Student characteristics	0.284 (0.077, 0.207)	0.251 (0.170, 0.081)	0.138 (0.114, 0.024)
Student characteristics & test scores	0.321 (0.174, 0.147)	0.261 (0.216, 0.044)	0.125 (0.102, 0.023)

worrying that the test scores dramatically push up the false positive error (i.e., predicting that a student passes at least 1 exam while she actually does not).

Logistic regression usually has a prediction performance which is satisfactory, but not as good as methods specifically designed for prediction. It is therefore worthwhile to run the prediction task on our data using a cutting edge method such as random forest.

4. RANDOM FOREST

Random forest (Breiman, 2001) is a prediction method based on trees (classification trees if the outcome is categorical or regression trees if the outcome is quantitative). We give a short outline of the method, referring to Hastie *et al*. (2009) and James *et al*. (2013) for further details.

A tree is a data-driven recursive partitioning of the space of the predictors. Trees are very flexible since no functional form is assumed, however their prediction performance is generally not good since they are quite sensitive to small changes in the training set. In other words, trees have low bias and high variance. Moreover, trees do not fully exploit the predictors as the fit is largely determined by a small set of strong predictors. A random forest is made of a set of trees grown on different versions of the data (thanks to bootstrap) using different predictors at each split of any tree (since a random subset of predictors is selected at each split). Averaging over the trees reduces the variance and generally yields accurate predictions.

The random forest algorithm can be summarised as follows.

1. Consider a training data set where $\boldsymbol{X}$ is the $n \times p$ matrix of the predictors and y is the $n \times 1$ vector of the outcome. Fix the number of bootstrap samples (equal to the number of trees) B and the number of predictors to be used at each split $m \leq p$.
2. For $b=1,2,\ldots,B$, do the following.

a. Create a bootstrap version of the training data by randomly sampling the n rows with replacement n times.
b. Grow a maximal-depth tree using the bootstrap version of the training data, sampling m of the p predictors at random prior to making each split.
c. Save the tree, as well as the bootstrap sampling frequencies for each of the training observations.

3. At any point $\boldsymbol{x}_0$ compute the prediction $\hat{y}_b(\boldsymbol{x}_0)$ using the b-th tree for $b=1,2,\ldots,B$, then compute the random forest prediction as the average:

$$\hat{y}(\boldsymbol{x}_0) = \frac{1}{B}\sum_{b=1}^{B} \hat{y}_b(\boldsymbol{x}_0)$$

In a classification setting, the b-th tree assigns the point of interest $\boldsymbol{x}_0$ to one of the outcome categories, say the c-th category: it is said that the b-th tree *votes* for the c-th category. Averaging across the trees yields the fraction of votes for any category. The standard way to summarise the votes of the trees is the *simple majority rule*, so that the random forest assigns $\boldsymbol{x}_0$ to the category which received most votes.

Any observation is expected to be out of a fraction $e^{-1}\approx 0.37$ of the bootstrap samples, thus it is not used in about 37% of the trees. This property is exploited to obtain the *Out-Of-Bag* (OOB) error: (1) for each observation $i=1,\ldots,n$ compute the OOB prediction, namely the average prediction across the bootstrap samples where i was not present, then compute the OOB error by contrasting the outcome y_i with the OOB prediction; (2) compute the overall OOB error by averaging across observations. The OOB error is a peculiar by-product of a random forest which is equivalent to leave-one-out cross-validation as long as the number of bootstrap samples B is large. The OOB error is used for tuning the parameters B (number of trees) and m (number of predictors to be used at each split): B should be large enough so that the OOB error stabilises, whereas m should be chosen to minimise the OOB error.

4.1 IMPLEMENTATION

We apply random forest to predict the performance of the 869 students who did the self-evaluation test and enrolled in the 2014/2015 academic year to the School of Economics and Management of the University of Florence. We replicate the analysis carried out in the previous section for logistic regression.

We implement random forest using the `randomForest` package of R. We adopt the default value for the number of trees (`ntree`), namely $B=500$, which turns out to be largely sufficient on the basis of the plot of the OOB error. Moreover, we adopt the default criterion for choosing the number of predictors randomly selected at

each split of the tree (`mtry`), namely $m=\sqrt{p}$ where p is the number of predictors. In our application, the predictors are p=5 (when using only student characteristics) or p=8 (when adding test scores): in both cases, the criterion suggests to randomly select m=2 predictors at each split. Trying with different values of m reveals that m=2 is optimal, or nearly optimal, in terms of OOB error in all the configurations of our application.

4.2 IMPORTANCE OF THE PREDICTORS

A useful tool accompanying a random forest is the variable importance plot, where the predictors are ranked according to their predictive power measured as the mean decrease in accuracy computed from permuting OOB data (e.g. Hastie *et al.*, 2009). In this way, categorical and numerical predictors are on the same scale and thus comparable.

Figure 1 shows the *variable importance plots* for random forest using both student characteristics and test scores, separately for each binary outcome. The high school grade is by far the most important predictor. The other high school features

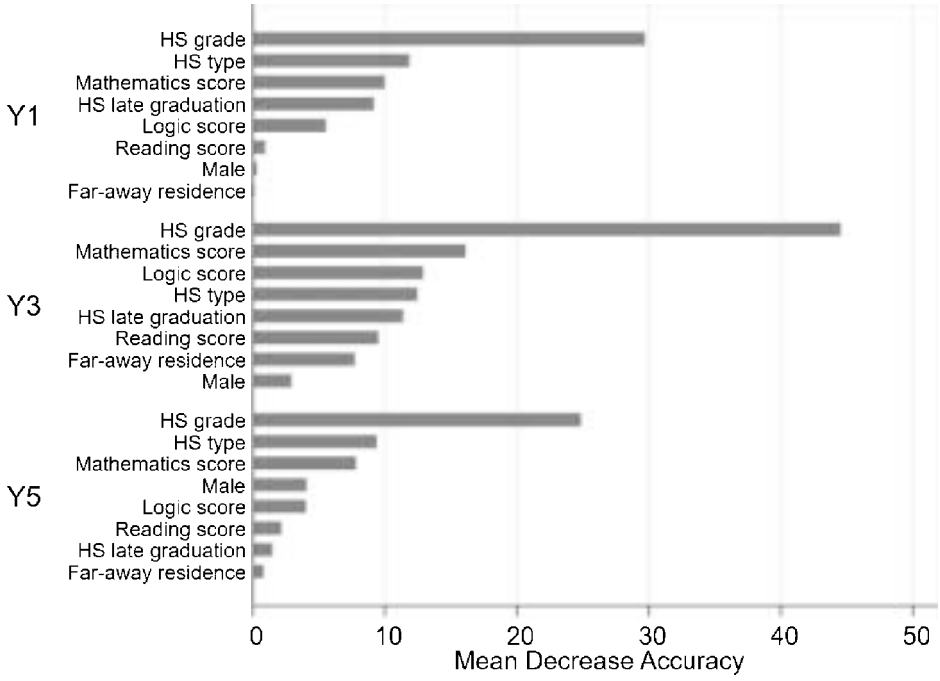

Fig. 1: Variable importance plots for random forests using student characteristics and test scores. School of Economics and Management, University of Florence, academic year 2014/2015.

are relevant, though in a different order: for example, late graduation is important to predict Y_1 (taking at least 1 exam), but not to predict Y_5 (taking at least 5 exams). For all binary outcomes, the test scores have the same order of importance: mathematics, logic, reading. The math score is the most important, likely because those competencies are key for the exam of mathematics, but also for other exams such as microeconomics and statistics. Test scores are more helpful to predict Y_3 than the other outcomes. However, in order to assess whether test scores actually improve prediction, we have to remove them from the set of predictors and compare the prediction errors.

4.3 PREDICTION ABILITY

To ensure comparability, we assess the prediction ability of random forest through a cross-validation on the same 10 folds used for logistic regression. Table 5 displays the average prediction error, separately for each binary outcome and type of predictors (student characteristics with or without test scores). The cross-validation errors are close to the OOB errors available after running the random forest procedure (for example, for the bottom left part of Table 5 the OOB errors are 0.308, 0.261, 0.124).

Tab. 5: Average prediction errors from 10-fold cross-validation for random forest exploiting student characteristics with or without test scores. Overall error (false positive and false negative errors in parenthesis, respectively). School of Economics and Management, University of Florence, academic year 2014/2015

Type of predictors	Y_1	Y_3	Y_5
Student characteristics	0.296 (0.070, 0.226)	0.251 (0.183, 0.068)	0.138 (0.123, 0.015)
Student characteristics & test scores	0.306 (0.088, 0.219)	0.247 (0.170, 0.077)	0.125 (0.114, 0.011)

The results of random forest are similar to those of logistic regression. In some configurations random forest is more accurate, in other configurations the reverse is true. Like for logistic regression, it is more difficult to predict Y_1 (passing at least 1 exam) than Y_5 (passing at least 5 exams). It is worth to note that the false positive error is smaller than the false negative error in predicting Y_1, while the reverse is true for the other two outcomes. It is confirmed that self-evaluation test scores give little additional predictive power: they slightly reduce the overall error in predicting Y_3 and Y_5 and, unfortunately, they increase the overall error in the critical task of predicting Y_1 (even if, contrary to logistic regression, the false positive error does not increase much).

5. CONCLUDING REMARKS

Our study considered the self-evaluation test of the School of Economics and Management of the University of Florence, academic year 2014/2015. The aim was to assess the contribution given by the test scores, in addition to student characteristics, to the prediction of student performance at the end of the first year. We measured student performance using the binary outcomes defined in Table 1: obtaining at least 1 credit (i.e., passing at least 1 exam), obtaining at least 20 credits (i.e., passing at least 3 exams), obtaining at least 40 credits (i.e., passing at least 5 exams). We exploited two prediction methods, namely logistic regression and random forest, and assessed their predictive ability through 10-fold cross-validation.

The implementation of logistic regression was straightforward as it is a consolidated and relatively simple method. Random forest is more complex, requiring critical choices such as the number of predictors to be used at each split of the tree.

The prediction ability of random forest turned out to be similar to logistic regression. Therefore, we argue that logistic regression is preferable as it simpler as for implementation and interpretation. A benefit of random forest is the associated *variable importance plot*, which allowed to compare both categorical and numerical predictors in terms of predictive power, revealing that the high school grade is by far the most important predictor. Both logistic regression and random forest could be refined, for example we could explore non-linearities and interactions in logistic regression. However, we do not pursue technical developments since they are unlikely to contribute to the issue of assessing the role of the self-evaluation scores in predicting student outcomes.

We analyse binary outcomes defined by applying thresholds to the number of gained credits, corresponding to the number of passed exams: this choice is motivated by their role as targets for the university management. Alternative approaches are possible, for example: (*i*) analysing the number of exams as a numerical outcome, which is not recommended since zero exams is a value of special interest which deserves a focussed prediction; (*ii*) analysing the number of exams as a categorical outcome with 7 levels (from 0 to 6 exams). The latter approach is methodologically adequate, but it complicates the implementation and the interpretation of the findings, without substantial insights for the university management.

From a substantive point of view, there are two main findings. First, the self-evaluation test scores did not help in predicting student performance once student characteristics were properly exploited. This pattern was detected for all binary outcomes in most scenarios, so we can argue that the information provided by the

pre-enrolment test is largely redundant. Clearly, the generalizability of this finding should be investigated with studies on other cohorts of students, degree programs, universities, and different kinds the test. Anyway, this finding does not mean that the pre-enrolment test is useless since there are other motivations beyond the predictive ability, such as giving the potential students an instrument for self-evaluation to orient their choices, and providing the university management with an 'objective' screening tool that avoids decisions explicitly based on student characteristics (which may be controversial).

The second main finding is that the three binary outcomes are not equally challenging in terms of prediction: in fact, it is rather more difficult to predict if a student will obtain at least one credit than to predict if a student will obtain more than 40 credits. This is bad news for the university management, which is comprehensibly worried about reducing the drop-out rate. A possible explanation is that a complete failure to get credits may be due to factors unrelated to the high school career or the self-evaluation test scores, such as the motivation of the student and accidental events like health problems or straits.

In conclusion, we recommend a revision of the self-evaluation test in order to effectively increase the information provided by the high school career. In particular, we argue that the test should include a section to evaluate some psychological traits of the candidate, such as the willingness to work hard and the motivation to undertake the degree program under consideration.

REFERENCES

Agresti, A. (2015). *Foundations of linear and generalized linear models.* John Wiley & Sons.

Breiman, L. (2001). Random forests. *Machine learning* 45: 5-32.

Grilli, L., Rampichini, C. and Varriale, R. (2015). Binomial Mixture Modeling of University Credits. *Communications in Statistics - Theory and Methods*. 44: 4866-4879.

Grilli, L., Rampichini, C. and Varriale, R. (2016). Statistical modelling of gained university. *Statistical Modelling*, 16: 47-66.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The elements of statistical learning: Data Mining, Inference, and Prediction.* Second Edition. Springer.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.