PREDICTION OF THE STUDENT SUCCESS RATE BY MEANS OF QUALITY TEACHING SURVEY VARIABLES APPLYING A MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) MODELS.

Ramiro Martís[2], Javier Alonso[2], Carlos Catalán[2], Ramón Fuentes[1] & Alberto A. Suárez[2]

Abstract.

The present research applies a multivariate adaptive regression splines (MARS) model to predict the success rate of students, employing a set of variables obtained from a quality teaching survey.

The survey measures student satisfaction as regards quality of teaching. The data correspond to the academic year 2008-2009 at the University of Oviedo. This survey has been conducted by the University of Oviedo Technical Quality Unit each year since 2001.

The aim of the present research paper was to calculate a forecast model able to predict the success rate of the students in each subject using some of the items of the quality teaching survey referred to above as predictive variables. The results show the existence of a clear relationship between student perception of lecturer performance and the student success rate.

Keywords:

Quality Higher Education, Quality Teaching Survey, Multivariate Adaptive Regression Splines Model, Success Rate of Students, Academic Achievement, Satisfaction.

[1] Faculty of Economics. Department of Applied Economic Research. University of Alicante.
Email: rfuentes@ua.es. Telephone and fax number: +34 965 90 97 11. Postal address: Facultad de Económicas. Universidad de Alicante. Ap. Correos 99. E-03080. Spain.

[2] University of Oviedo, Technical Unit of Quality, 3 Principado St., mezzanine floor, 33007 Oviedo (Spain). http://www.uniovi.net/calidad/ suarez@uniovi.es

1. Introduction.

Assessment of the activity of teaching is based primarily on the criteria and guidelines of the European Higher Education Area and on the legal requirements established in Spain by Fundamental Law 4/2007 and Royal Decree 1393/2007.

As the European Conference at Bergen (2005) established at the suggestion of the *European Association for Quality Assurance in Higher Education* (ENQA), Universities should set up internal systems of quality assurance capable of especially guaranteeing quality in the teaching activity, thus ensuring that teaching staff is both qualified and competent. In 2007, the Spanish Agency for Quality Assessment and Accreditation (its Spanish acronym, ANECA) implemented a *Support Programme for Assessing Teaching Activity* (DOCENTIA), one of its goals being to *guarantee the quality of university teaching, fostering its recognition, development and improvement.*

The University of Oviedo obtained a positive assessment report from ANECA in the Procedure for Assessing the Teaching Activity of University of Oviedo Staff on 22nd April 2008; teaching activity being understood as the set of actions, whether in-person or not, carried out by teaching staff at the University of Oviedo within the framework of the teaching/learning process. Within this context, the General Teaching Survey on students and teaching staff is a fundamental tool applied at the University of Oviedo to monitor the perception of Teaching Quality in its twofold aspects: imparted and received. To this end, 10-point Lickert scale questionnaires are employed, the survey unit being the subject-lecturer-group. This procedure is carried out twice a year coinciding with the two semesters. In the first period, the questionnaire appraises those subjects imparted during only the first semester, while the questionnaire employed in the second period surveys those subjects taught only in the second semester as well as those imparted on an annual basis.

To complete this information, achievement analyses[3] are carried out on each subject in each degree course. This includes data on new enrolments, teaching load, dropping out of the course and graduation, as well as the average time taken to obtain each one of the degrees conferred by the University of Oviedo.

The aim of this study is to analyse whether a relation exists between student satisfaction as regards quality of teaching and academic results with the aim of predicting subject success rates. Thus, depending on the satisfaction scores for the first semester, it should allow those responsible for academic matters to initiate corrective actions or improvements that would enable an increase in the final success rate.

To do so, a comparison is made per subject of the average student satisfaction score in each item on the satisfaction questionnaire versus the success rate for the academic year 2008-09.

[3] The primary aim of the Academic Achievement Study is to provide information on the actual situation and evolution of University of Oviedo degrees as regards progress and results in student learning.

The first step is to ensure the validity of the survey, which refers to the degree to which said instrument enables what it is said to measure to be measured. Normally, three complementary ways of demonstrating validity are referred to, as well as different procedures for verifying said demonstration [1], namely the validity of the contents, the criterion and the construct [1]. This survey has been used at the University of Oviedo since 2001. It has been systematically designed and revised by academic professionals and its application is approved each year by the University Governance Board.

Reliability is a basic characteristic of the measurement obtained when applying a measuring tool, as it refers to the accuracy and/or consistency of the values obtained. In the present case, reliability was determined via the internal consistency method calculating Cronbach's Alpha Coefficient[4] on the fifteen items on the questionnaire, obtaining a value of $\alpha = 0.97$, thus allowing us to state that the questionnaire has a higher degree of reliability[5] or generalisability.

Matriz de correlaciones inter-elementos

	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6	ITEM7	ITEM8	ITEM9	ITEM10	ITEM11	ITEM12	ITEM13	ITEM14	ITEM15
ITEM1	1.000	.793	.765	.694	.663	.613	.603	.609	.634	.651	.629	.616	.624	.677	.686
ITEM2	.793	1.000	.748	.657	.698	.590	.573	.582	.593	.612	.602	.596	.595	.631	.650
ITEM3	.765	.748	1.000	.775	.680	.677	.614	.641	.680	.677	.655	.637	.644	.700	.726
ITEM4	.694	.657	.775	1.000	.711	.729	.641	.625	.706	.715	.646	.642	.679	.742	.755
ITEM5	.663	.698	.680	.711	1.000	.639	.594	.561	.599	.632	.583	.592	.613	.667	.672
ITEM6	.613	.590	.677	.729	.639	1.000	.672	.624	.666	.668	.636	.610	.635	.712	.718
ITEM7	.603	.573	.614	.641	.594	.672	1.000	.588	.628	.675	.610	.653	.669	.646	.678
ITEM8	.609	.582	.641	.625	.561	.624	.588	1.000	.775	.708	.807	.696	.672	.666	.743
ITEM9	.634	.593	.680	.706	.599	.686	.628	.775	1.000	.794	.770	.739	.766	.750	.646
ITEM10	.651	.612	.677	.715	.632	.668	.675	.708	.794	1.000	.750	.735	.762	.744	.798
ITEM11	.629	.602	.655	.646	.583	.636	.610	.807	.770	.750	1.000	.810	.783	.705	.798
ITEM12	.616	.596	.637	.642	.592	.610	.653	.696	.739	.735	.810	1.000	.858	.695	.800
ITEM13	.624	.595	.644	.679	.613	.635	.669	.672	.766	.762	.783	.858	1.000	.738	.835
ITEM14	.677	.631	.700	.742	.667	.712	.646	.666	.750	.744	.705	.695	.738	1.000	.847
ITEM15	.686	.650	.726	.755	.672	.718	.678	.743	.846	.798	.798	.800	.835	.847	1.000

Table 1. Inter-element correlation matrix

Reliability was calculated on a response rate that reached 43.2%, thus allowing 83,723 questionnaires to be processed. Bearing in mind that the survey is voluntary, biases could be deduced in the intentionality of students to respond to the questionnaire. However, specific control variables, such as the class attendance rate, allow us to consider the goodness and validity of the survey, as no anomalous circumstances were observed in the profile of the students who responded to it. [2]

In contrast with academic achievement variables such as success, falling behind and dropping out the course [3], the investigation of academic factors related to teaching staff and the student body [4] acquire a certain degree of predominance in the monitoring of degrees.

The General Teaching Survey process has obtained the ISO 9001:2008 Standard Certificate of Approval since 2001.

[4] Cronbach's Alpha statistic reflects the degree of covariance in the items comprising the test and ranges between 0 and 1.
[5] Bisquerra (0.75) and Peterson (0.85).

The decision to choose this rate is due to the fact that, besides allowing us to ascertain the relation between satisfaction and success, it could allow us to predict success in the subject.

The alternative hypothesis is that the perceived satisfaction of students during the first semester allows us to predict success in the subject on completing the course.

2. Theory.

2.1. Multivariate adaptive regression splines (MARS).

Multivariate adaptive regression splines (MARS) is a multivariate nonparametric regression technique introduced by Friedman in 1991. Its main purpose is to predict the values of a continuous dependent variable, $\vec{y}(n\times 1)$, from a set of independent explanatory variables, $\vec{X}(n\times p)$. The MARS model can be represented as:

$$\vec{y} = f\left(\vec{X}\right) + \vec{\varepsilon} \quad 2)$$

where $\vec{\varepsilon}$ is an error vector of dimension $(n\times 1)$.

MARS can be considered as a generalisation of ‘classification and regression trees’ (CART) [5], and is able to overcome some limitations of CART. MARS does not require any *a priori* assumptions about the underlying functional relationship between dependent and independent variables. Instead, this relation is uncovered from a set of coefficients and piecewise polynomials of degree q (basis functions) that are entirely “driven” from the regression data $\left(\vec{X},\vec{y}\right)$. The MARS regression model is constructed by fitting basis functions to distinct intervals of the independent variables. Generally, piecewise polynomials, also called splines, have pieces smoothly connected together. In MARS terminology, the joining points of the polynomials are called knots, nodes or breakdown points. These will be denoted by the small letter t. For a spline of degree q, each segment is a polynomial function. MARS uses two-sided truncated power functions as spline basis functions, described by the following equations [6]:

$$\left[-(x-t)\right]_+^q = \begin{cases} (t-x)^q & if\ x < t \\ 0 & otherwise \end{cases} \quad 3)$$

$$\left[+(x-t)\right]_+^q = \begin{cases} (t-x)^q & if\ x \geq t \\ 0 & otherwise \end{cases} \quad 4)$$

where $q(\geq 0)$ is the power to which the splines are raised and which determines the degree of smoothness of the resultant function estimate.

The MARS model of a dependent variable $\vec{y}$ with M basis functions (terms) can be written as [19-20]:

$$\hat{y} = \hat{f}_M(\vec{x}) = c_0 + \sum_{m=1}^{M} c_m B_m(\vec{x}) \qquad 5)$$

where $\hat{y}$ is the dependent variable predicted by the MARS model, c_0 is a constant, $B_m(\vec{x})$ is the *m*-th basis function, which may be a single spline basis functions, and c_m is the coefficient of the *m*-th basis functions.

Both the variables to be introduced into the model and the knot positions for each individual variable have to be optimized. For a data set $\vec{X}$ containing *n* objects and *p* explanatory variables, there are $N = n \times p$ pairs of spline basis functions, given by Eqs. (3) and (4), with knot locations x_{ij} ($i = 1, 2, ..., n$; $j = 1, 2, ..., p$).

A two-step procedure is followed to construct the final model. First, in order to select the consecutive pairs of basis functions of the model, a two-at-a-time forward stepwise procedure is implemented [19-20]. This forward stepwise selection of basis function leads to a very complex and overfitted model. Although it fits the data well, such a model has poor predictive abilities for new objects. To improve the prediction, the redundant basis functions are removed one at a time using a backward stepwise procedure. To determine which basis functions should be included in the model, MARS employs the generalized cross-validation criterion [19-20] (*GCV*), The GCV is the mean squared residual error divided by a penalty dependent on the model complexity. The GCV criterion is defined in the following way:

$$GCV(M) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}_M(\vec{x}_i)\right)^2}{\left(1 - C(M)/n\right)^2} \qquad 6)$$

where $C(M)$ is a complexity penalty that increases with the number of basis functions in the model and which is defined as:

$$C(M) = (M + 1) + d\,M \qquad 7)$$

where *M* is the number of basis functions in Eq. (5), and the parameter *d* is a penalty for each basis function included into the model. It can be also regarded as a smoothing parameter. Large values of *d* lead to fewer basis functions and therefore smoother function estimates. For more details on the selection of the *d* parameter, see Ref. [7]. In our study, the parameter *d* equals 2, and the maximum interaction level of the spline basis functions is restricted to 3.

The main steps of the MARS algorithm, as applied here, can be summarized as follows [6]:

1. Select the maximum allowed complexity of the model and define the *d* parameter.

Forward stepwise selection:

2. Start with the simplest model, i.e. with the constant coefficient only.
3. Explore the space of the basis functions for each explanatory variable.
4. Determine the pair of basis functions that minimizes the prediction error and include them in the model.
5. Go back to step 2 until a model with predetermined complexity is derived.

Backward stepwise selection:

6. Search the entire set of basis functions (excluding the constant) and delete from the model the one that contributes least to the overall goodness-of-fit using the GCV criterion.
7. Repeat 5 until GCV reaches its maximum.

The predetermined complexity of MARS model in step 3 should be considerably larger than the optimal (minimal GCV) model size M^*, so choosing the predetermined complexity of the model as greater than $2M^*$ generally suffices [7].

It is possible to analyse a MARS model using surface plots that visualise the interactions and effects between the basis functions. To illustrate this, some definitions will be introduced. Let $f_i(\vec{x}_i)$ be the set of all single variable basis functions, i.e. basis functions that contain only $\vec{x}_i$. Similarly, let $f_{ij}(\vec{x}_i, \vec{x}_j)$ be the set of all two-variable basis functions that contain the pairs of variables $\vec{x}_i$ and $\vec{x}_j$, and $f_{ijk}(\vec{x}_i, \vec{x}_j, \vec{x}_k)$ the set of all three-variable basis functions that contain the triplets of variables $\vec{x}_i$, $\vec{x}_j$ and $\vec{x}_k$. The MARS model can be rewritten in the following form:

$$\hat{f}(\vec{x}) = c_0 + \sum f_i(\vec{x}_i) + \sum f_{ij}(\vec{x}_i, \vec{x}_j) + \sum f_{ijk}(\vec{x}_i, \vec{x}_j, \vec{x}_k) \qquad 8)$$

where the first sum is over all single-variable basis functions, the second sum is over all strictly two-variable basis functions, and the third sum represents all three-variable basis functions. Eq. (8) is called ANOVA decomposition due to its similarity to the decomposition by ANOVA of experimental design [7]. The two-variable interaction of a MARS model, $If_{ij}(\vec{x}_i, \vec{x}_j)$, is given by:

$$If_{ij}(\vec{x}_i, \vec{x}_j) = f_i(\vec{x}_i) + f_j(\vec{x}_j) + f_{ij}(\vec{x}_i, \vec{x}_j) \qquad 9)$$

Higher level interactions can be defined in a similar way. The graphical presentation of the ANOVA decomposition facilitates the interpretation of the MARS model. The effect of a one-variable basis function can be viewed by plotting $f_i(\vec{x}_i)$ against $\vec{x}_i$. Two-variable interaction can be viewed by plotting $If_{ij}(\vec{x}_i, \vec{x}_j)$ against $\vec{x}_i$ and $\vec{x}_i$ in a surface plot.

2.2. Prediction ability of the MARS model.

The prediction ability of the MARS model can be evaluated in terms of the 'root mean squared error of cross-validation' (RMSECV) and the squared leave-one-out correlation coefficient (q^2). To compute RMSECV, one object is left out from the data set and the model is constructed for the remaining $n-1$ objects. Then the model is used to predict the value for the object left out. When all objects have been left out once, RMSECV is given by Friedman, J.H., and Roosen, C.B. [8]:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{-i}\right)^2}{n}} \quad 10)$$

where y_i is the value of dependent variable of the i-th object and $\hat{y}_{-i}$ is the predicted value of the dependent variable of the i-th object with the model built without the i-th object.

The value of q^2 is given as:

$$q^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{-i}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \quad 11)$$

where $\bar{y}$ is the mean value of the dependent variable for all n objects.

2.3. The importance of the variables in the MARS model.

Once the MARS model has been constructed, it is possible to evaluate the importance of the explanatory variables used to construct the basis functions. Since each explanatory variable can be incorporated in different basis functions, the importance of the variable is expressed as its contribution to the goodness-of-fit of the model. The scoring of the importance of variables in the MARS model is similar to the leave-one-out cross-validation concept. To calculate variable importance scores, MARS refits the model after deleting all terms involving the variable at issue and calculating the reduction in goodness-of-fit. The importance of the variables is a relative measure and scaled between 0 and 1. The most important variable is the one that, when dropped, decreases the model fit the most and it receives the highest score, i.e. 1. The less important variables receive the lower scores, which is the ratio of the reduction in goodness-of-fit of these variables to that of the most important variable.

3. Data.

In line with the research goal, the data proceed from two types of variables: on the one hand, the information obtained via the student questionnaire and, on the other, academic achievement results.

The information from the satisfaction questionnaire is organised in several blocks of questions covering similar themes:

1. Teaching Accomplishment
2. Teaching
3. Attitude
4. General Satisfaction

The first block (Teaching Accomplishment), made up of 7 questions, gathers information on student perception of teaching accomplishment, as regards both information and syllabus, assessment criteria, appropriateness of the contents of lectures, activities and assignments that illustrate said contents, system of assessment and, finally, the usefulness of attending lectures and tutorials in preparing subjects.

The following three questions (Teaching) analyse the way lectures are given, knowledge of the subject, clarity of exposition and recommended reading matter.

The third block (Attitude) is made up of three questions that refer to the attitude shown by teaching staff both towards the subject and towards their students.

The last block (General Satisfaction) scores the general satisfaction of students with the work done by the lecturer and with what they learned in the subject.

The information on academic achievement responds to the variable:

Success Rate, (SR): defined as the percent relation between the total number of credits passed and the total number of credits taken by the total number of students.

CODE	VARIABLE
PTA1	Information on the Syllabus and Work Schedule
PTA2	Assessment Criteria
PTA3	Contents
PTA4	Activities and Assignments
PTA5	System of Assessment
PTA6	Class Attendance
PTA7	Usefulness of the Tutorial
PTCH1	Knowledge of the Subject
PTCH2	Explanation
PTCH3	Materials
PATT1	Interest
PATT2	Accessibility
PATT3	Attention to Difficulties
PSAT1	Learning
PSAT2	Work of Lecturer
SR	Success Rate

Table 2. Names of the variables used in the model

4. Results and discussion.

The MARS model was constructed for the output variable Success Rate. This regression model was built using the techniques in Friedman's papers "Multivariate Adaptive Regression Splines" [7] and "Fast MARS". [9]

The final model includes 23 basis functions, which are listed in Table 4 together with their corresponding coefficients. Apart from the constant term and seven basis functions of level 1, there are 13 interaction terms of level 2 and 2 interaction terms of level 3 (see Table 4) This model was built using the results of 3,544 subjects as training data (80% of the total database) The validation was performed with 887 results of subjects (the remaining 20% of the total database). The inclusion of the results of a subject either in the training or validation subset was performed at random. A scatterplot relating observed success rate versus predicted success rate for the validation data is shown in Fig. 1. It can be observed that the higher the real success rates of the student, the higher the success rate predicted by the model. It can likewise be observed in the same figure that the model has difficulties predicting a success rate of 100%: (see upper right part of the figure). The R-square obtained is 0.4662, which would be higher if those subjects with a success rate of 100% were removed from the validation data set.

The results in Figure 1 evidence a clear grouping of subjects whose success rate reaches 100%. Analysing the typology of these subjects, at first sight a fixed profile has not been found that would lead us to infer that these subjects are non compulsory in character, i.e. optional or freely chosen subjects. To the contrary; the profile comprises both compulsory and optional subjects, with no significant differences existing in the results of the analyses if the non compulsory subjects are left out.

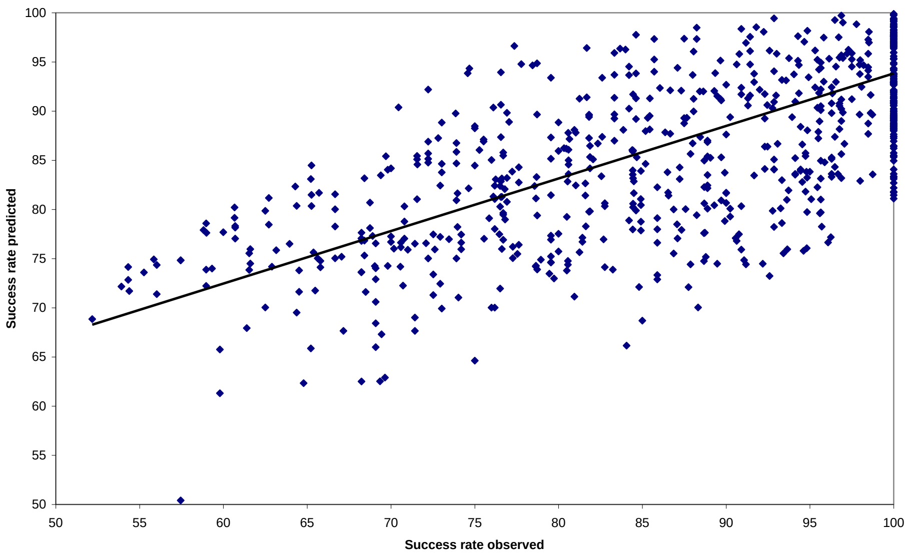

Figure 1. Scatterplot of observed success rate versus predicted success rate (validation data).

The importance of the explanatory variables in the MARS model is presented in Table 3, which includes three main criteria used to estimate the importance of the variables in a standard MARS model [7]. These criteria are as follows:

• The nsubsets criterion counts the number of model subsets that include the variable. Variables that are included in more subsets are considered more important. By subsets we mean the subsets of terms generated by the pruning pass. There is one subset for each model size, and the subset is the best set of terms for that model size. Only subsets that are smaller than or equal in size to the final model are used for estimating variable importance.

• The RR criterion first calculates the decrease in the RSS for each subset relative to the previous subset. (For multiple response models, RSS's are calculated over all responses.) Then, for each variable, it sums these decreases over all subsets that include the variable. Finally, it scales the summed decreases so the maximum summed decrease is 100. Variables which cause larger net decreases in the RSS are considered more important.

• The GCV criterion is similar, but uses GCV instead of the RSS. Adding a variable can increase the GCV; i.e. adding the variable has a deleterious effect on the model. When this happens, the variable could even have a negative total importance and thus appear less important than unused variables.

The application of this criteria to the present model showed that the most important variables are Academic Year, some variables related with PTA (PTA5, PTA7, PTA1, PTA3, PTA4, PTA6), PTCH (only PTCH1), Attitude (PATT2, PATT1, PATT3) and finally Satisfaction (PSAT1).

The variable with the most influence is the Academic Year. Prior analyses do not allow us to infer whether students with a lower capacity drop out in their first years or whether the best lecturers teach the higher courses. In a study carried out in 2001 at the University of Oviedo, it was concluded that significant differences could not be found between the low success rate in the lower courses and whether the students had followed the LOGSE or COU curriculum at secondary school [10] as regards their level. The origin must therefore be sought in the process of adaptation to university studies [10] or in other variables.

The block of variables PTA (Teaching Accomplishment) is shown to be the most important in determining the success rate, due to the number of variables that correlate with success. In order of importance, these are: *System of Assessment, Usefulness of the Tutorial, Information on the Syllabus and Work Schedule, Contents, Activities and Assignments and Class Attendance.*

The only variable from the Teaching block that has an influence is the student's subjective perception regarding Knowledge of the Subject, as can be seen from Figure 2 and Table 3.

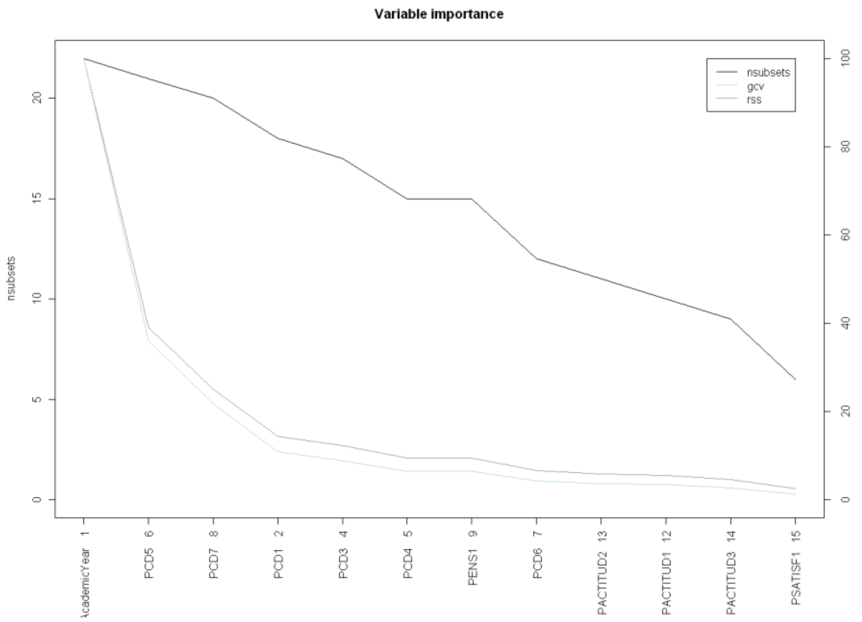

Figure 2. Importance of the explanatory variables in the MARS model

	nsubsets	GCV	RSS
AcademicYear	**22**	**100.00**	**100.00**
PTA5: System of Assessment	**21**	**36.03**	**38.99**
PTA7: Usefulness of the Tutorial	**20**	**21.61**	**25.11**
PTA1: Information on the Syllabus and Work Schedule	**18**	**18.88**	**14.38**
PTA3: Contents	**17**	**8.85**	**12.24**
PTA4: Activities and Assignments	**15**	**6.43**	**9.49**
PTCH1: Knowledge of the Subject	15	6.43	9.49
PTA6: Class Attendance	12	4.21	6.87
PATT2: Accessibility	11	3.60	5.89
PATT1: Interest	10	3.42	5.49
PATT3: Attention to Difficulties	9	2.75	4.62
PSAT: Learning	6	1.31	2.59

Table 3. Importance of the explanatory variables in the MARS model.

Nº	Basis function	Coefficient
1	1	80.160733
2	h(AcademicYear-3)	3.417153
3	h(3-AcademicYear)	-8.253868
4	h(7.73913-PTA3)	3.696258
5	h(PTA4-6.28571)	-1.412395
6	h(PTA5-5.45455)	7.187522
7	h(PTA6-8.61538)	10.175454
8	h(PTA7-4.71429)	-3.363943
9	h(AcademicYear-4) * h(8.61538-PTA6)	1.302150
10	h(3-AcademicYear) * h(PATT1-9.33333)	20.728190
11	h(PTA1-9.21429) * h(7.73913-PTA3)	11.601691
12	h(9.21429-PTA1) * h(7.73913-PTA3)	-0.420384
13	h(PTA1-7.8) * h(PTA5-5.45455)	-2.086423
14	h(7.8-PTA1) * h(PTA5-5.45455)	-1.227524
15	h(PTA3-8.5) * h(8.61538-PTA6)	-14.394898

16	h(6.28571-PTA4) * h(PTCH1-7.3)	-3.838091
17	h(6.28571-PTA4) * h(7.3-PTCH1)	-0.540679
18	h(8.61538-PTA6) * h(PSAT1-7)	2.064487
19	h(8.61538-PTA6) * h(7-PSAT1)	-0.353424
20	h(PTA7-4.71429) * h(PATT2-9.16667)	-9.390150
21	h(PTA7-4.71429) * h(PATT3-5.33333)	0.601493
22	h(4-AcademicYear) * h(PTA3-5) * h(8.61538-PTA6)	0.168861
23	h(4-AcademicYear) * h(5-PTA3) * h(8.61538-PTA6)	0.513164

Table 4. The basis functions and the corresponding coefficients in the MARS model.

Due to the local properties of the MARS model, it is possible to gain some additional knowledge about the interaction between explanatory variables and the response variable by looking closely at the model's decomposition. This decomposition is presented in Table 4, while the most interesting relations are shown in the following graphs.

It can be seen in Figure 3 that the later the academic year (4th-6th) and the higher the level of class attendance (5-10), the higher the success rate.

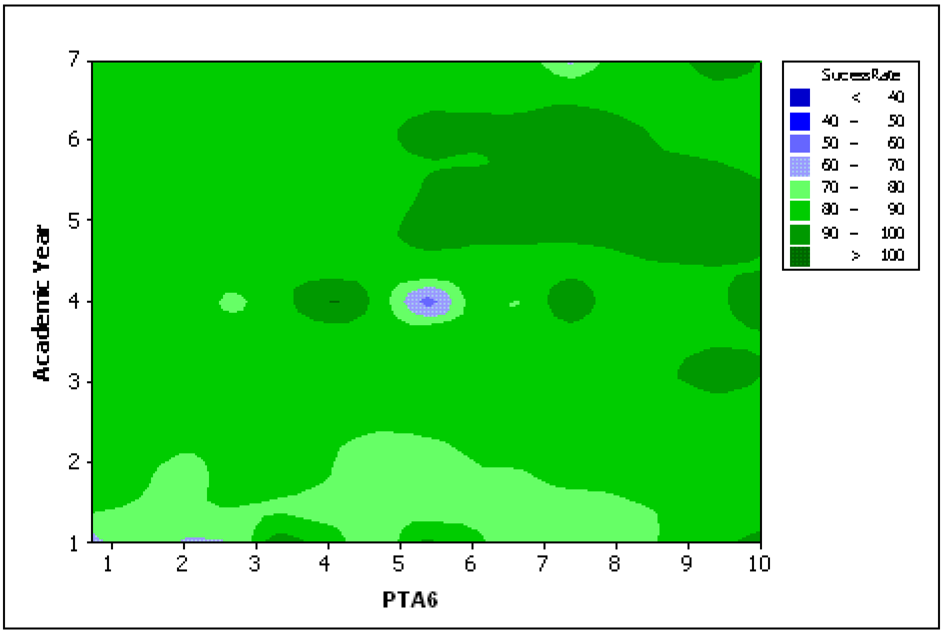

Figure 3. Contour Plot of Success Rate vs Academic Year; PTA6: Class Attendance.

It can be observed in Figure 4 that the success rate decreases when student assessments of Information on the Syllabus and Work Schedule and Contents fall within the interval 5-6.

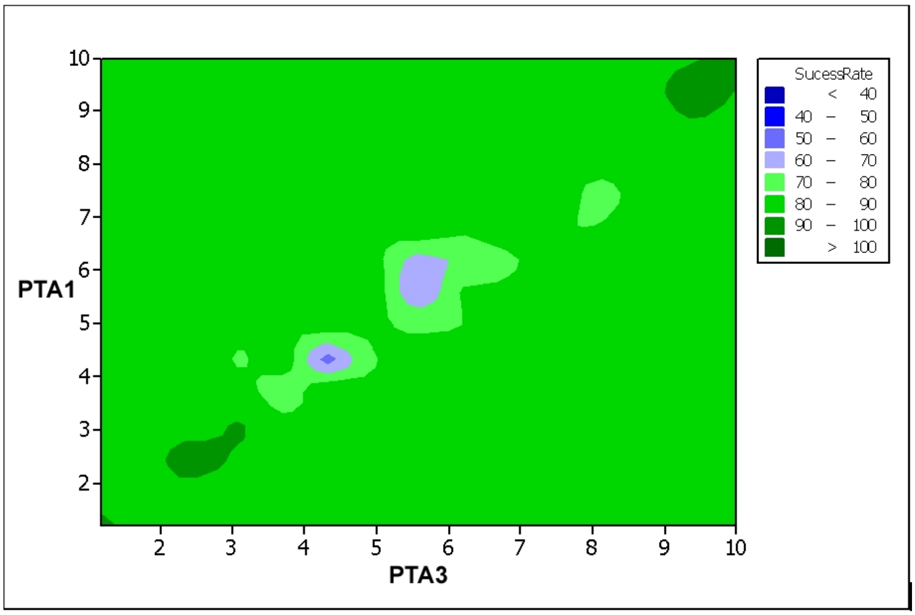

Figure 4. Contour Plot of Success Rate vs PTA1: Information on the Syllabus and Work Schedule; PTA3: Contents

5. Conclusions.

Due to the nonlinear and local character of the MARS model, very complex relationships in the data can be modelled, as is shown for the present data set. Although the interpretation of the basis functions is difficult, it is possible to evaluate the importance of certain variables for the model, as can be appreciated in Figure 2 and Table 3.

The results obtained support the alternative hypothesis: *the perceived satisfaction of students during the first semester allows us to predict success in the subject on completing the course*. However, the low correlation between the variables does not allow us to guarantee the accuracy of the model.

The results allow us to predict success in a subject mainly as a function of the academic year and of student satisfaction as measured in the block of questions on Teaching Accomplishment: *System of Assessment, Usefulness of the Tutorial, Information, Contents, Activities and Class Attendance.* This will enable those responsible for academic matters to initiate, for example, actions aimed at continuous improvement and innovation in the teaching activity, as well as to adjust the syllabus (planning, development and results) to student needs/expectations and assess the effort made by lecturers as regards teaching performance. The application of this predictive model allows us to extract further conclusions regarding the distribution of importance of the different items on the student questionnaire.

A future line of research, and one on which we are currently working, consists in analysing the risk factors of this model.

6. Bibliographical References.

[1] Bisquerra Alzina, R. (1989): *Métodos de Investigación Educativa*. Ceac. Barcelona.

[2] Harvey, L. (2003). Student feedback. *Quality in Higher Education*, 9 (1), pp3-20.

[3] Cabrera Guillén, P. y Galán Delgado, E. (2002). Factores contextuales y rendimiento académico. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 5 (3).

[4] Tejedor, F. J. y García Valcarcel, A. (2007). Causas del bajo rendimiento del estudiante universitario (en opinión de los profesores y alumnos). Propuestas de mejora en el Marco del EEES. *Revista de Educación*, 342. pp. 443-473.

[5] Hastie, T., Tibshirani, R., and Friedman, J.H. (2003) *The Elements of Statistical Learning*, Springer-Verlag, New York.

[6] Sekulic, S., and Kowalski, B.R. (1992) 'MARS: a tutorial', *J. Chemom.*, Vol. 6, pp. 199-216.

[7] Friedman, J.H. (1991) 'Multivariate adaptive regression splines', *Ann. Stat.*, Vol. 19, pp. 1-141.

[8] Friedman, J.H., and Roosen, C.B. (1995) 'An introduction to multivariate adaptive regression splines', *Statistical Methods in Medical Research*, Vol. 4, pp. 197-217.

[9] Friedman, J. H (1993). "Fast MARS." Dept. of Statistics, Stanford University Technical Report.

[10] De Miguel Díaz, M (Director) (2001): *Evaluación del Rendimiento en la Enseñanza Superior*. Ministerio de Educación Cultura y Deporte.