

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

## **Abstract**

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

The referred survey measures the students satisfaction with the quality teaching. These data correspond to the academic year 2008-2009 at Oviedo University. This questionnaire is yearly performed by the Technical Quality Unit of the University of Oviedo since 2001.

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

## **Keywords:**

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



## 1. Introduction.

La evaluación de la actividad docente se fundamenta prioritariamente en los criterios y directrices del Espacio Europeo de Educación Superior y en los requerimientos legales establecidos por la Ley Orgánica 4/2007 y el Real Decreto 1393/2007.

Tal como se estableció en la Conferencia Europea de Bergen (2005), a propuesta de la *European Association for Quality Assurance in Higher Education* (ENQA), las Universidades deben articular sistemas internos de garantía de calidad, capaces de garantizar en especial la calidad de la actividad docente, asegurando un profesorado cualificado y competente. En 2007 la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) puso en marcha en 2007 un *Programa de Apoyo a la Evaluación de la Actividad Docente del Profesorado* (DOCENTIA) siendo uno de sus objetivos *garantizar la calidad de la enseñanza universitaria, favoreciendo su reconocimiento, desarrollo y mejora*.

La Universidad de Oviedo obtiene el informe de evaluación positivo de la ANECA en el Procedimiento de Evaluación de la Actividad Docente del Profesorado de la Universidad de Oviedo, el 22 de abril de 2008, entendiéndose por actividad docente el conjunto de actuaciones, presenciales y no presenciales, realizadas por el profesorado de la Universidad de Oviedo en el marco del proceso de enseñanza-aprendizaje. En este contexto la Encuesta General de Enseñanza a estudiantes y profesorado es una de las herramientas fundamentales que se aplica, en la Universidad de Oviedo, para el seguimiento de la percepción, en sus dos vertientes, impartida y recibida, sobre la Calidad Docente. Para ello se emplean cuestionarios con escala Lickert de diez puntos, y la unidad de encuesta es asignatura-profesor/a-grupo. Este procedimiento se realiza dos veces al año coincidiendo con los dos cuatrimestres. Encuestando en el primer periodo sobre las asignaturas con docencia sólo durante el primer cuatrimestre, y en el segundo periodo, aquellas asignaturas con docencia, en el segundo cuatrimestre y anual.

Para completar esta información se realizan análisis de rendimiento<sup>1</sup> de cada asignatura en cada titulación, entre los que se incluyen datos de nuevo ingreso, dedicación lectiva, éxito y rendimiento de cada asignatura, abandono y graduación, así como la duración media de cada uno de los estudios en la Universidad de Oviedo y en la media de las universidades españolas.

---

<sup>1</sup> El Estudio de Rendimiento Académico tiene como objetivo primordial informar sobre la situación y evolución de las titulaciones de la Universidad de Oviedo en relación al progreso y los resultados del aprendizaje de los estudiantes.



El objetivo de este estudio es analizar si existe relación entre la satisfacción de los estudiantes y los resultados académicos con la finalidad de predecir el éxito en las asignaturas. De esta manera en función de los resultados de satisfacción del primer cuatrimestre, permitiría a los responsables académicos iniciar acciones correctoras o de mejora que permitiesen incrementar la tasa de éxito final.

Para ello se compara por asignatura, la valoración media de satisfacción de los estudiantes en cada ítem de la encuesta de satisfacción con la tasa de éxito del curso académico 2008-09.

El primer paso es garantizar la validez del instrumento, hace referencia al grado en que dicha herramienta permite medir lo que dice medir. Normalmente suele hacerse referencia a tres formas complementarias de manifestarse la validez, al tiempo que a distintos procedimientos para verificar su manifestación (Bisquerra, 1989), que son la validez de contenido, de criterio y de constructo (Bisquerra, 1989). Esta encuesta se aplica en la Universidad de Oviedo desde 2001, y ha sido diseñada y revisada sistemáticamente por profesionales académicos, aprobándose anualmente su aplicación en Consejo de Gobierno de la Universidad.

La fiabilidad es una característica básica de la medición que se obtiene al aplicar una herramienta de medida, ya que hace referencia a la precisión y/o consistencia de los valores obtenidos. En este caso se determinó la fiabilidad a través del método de consistencia interna calculando el Coeficiente Alfa de Cronbach<sup>2</sup> sobre los quince ítems del cuestionario arrojando un valor de  $\alpha = 0,97$  lo que nos permite afirmar que el cuestionario tiene un elevado grado de fiabilidad<sup>3</sup> o generalizabilidad.

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	,686	,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	,846
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. Matriz de correlaciones inter-elementos

La fiabilidad se ha calculado sobre una tasa de respuesta que alcanza el 43,2%, permitiendo procesar 83.723 cuestionarios. Teniendo en cuenta que la encuesta es voluntaria, podrían deducirse sesgos sobre la

<sup>2</sup> El estadístico Alfa de Cronbach refleja el grado en el que covarían los ítems que constituyen el test y oscila entre 0 y 1.

<sup>3</sup> Bisquerra (0,75), y Peterson (0,85).



intencionalidad del estudiante al responder el cuestionario, sin embargo determinadas variables de control, como el índice de asistencia a clase, nos permiten considerar la bondad y validez de la consulta, al no detectar circunstancias anómalas en el perfil de los estudiantes que responden. (Harvey, 2003)

La investigación de factores académicos, relacionados con el profesorado y el alumnado (Tejedor y García-Valcarcel, 2007), en contraste con variables de rendimiento académico como, éxito, retardo y abandono (Cabrera Guillén, 2002) adquieren cierto predominio en el seguimiento de las titulaciones.

El proceso de la encuesta General de la Enseñanza está certificado según los requisitos de la Norma ISO 9001:2008, desde 2001.

La decisión de seleccionar esta tasa se debe porque además de permitarnos conocer la relación entre satisfacción y éxito, nos podría permitir predecir el éxito de la asignatura.

La hipótesis alternativa es que la satisfacción percibida por los estudiantes durante el primer cuatrimestre nos va permitir predecir el éxito en las asignaturas al finalizar el curso.

## 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,  $\mathbf{y}^p (n \times 1)$ , from a set of independent explanatory variables,  $\mathbf{X}^p (n \times p)$ . The MARS model can be represented as:

$$\mathbf{y}^p = f(\mathbf{X}^p) + \mathbf{e}^p \quad 2)$$

where  $\mathbf{e}^p$  is an error vector of dimension  $(n \times 1)$ .

MARS can be considered as a generalisation of ‘classification and regression trees’ (CART) (Hastie, T., Tibshirani, R., and Friedman, J.H. 2003), 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  $(\mathbf{X}^p, \mathbf{y}^p)$ . 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 (Sekulic, S., and Kowalski, B.R. 1992):

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

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

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  $\hat{y}^p$  with  $M$  basis functions (terms) can be written as [19-20]:

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

where  $\hat{y}^p$  is the dependent variable predicted by the MARS model,  $c_0$  is a constant,  $B_m(x^p)$  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  $X^p$  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  $x_{ij}$  ( $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, p$ ).

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. Such a model, although it fits the data well, 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 utilizes the generalized cross-validation [19-20] (*GVC*). The *GVC* is the mean squared residual error divided by a penalty dependent on the model complexity. The *GVC* criterion is defined in the following way:



$$GVC(M) = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_M(x_i))^2}{(1 - C(M)/n)^2} \quad 6)$$

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

$$C(M) = (M + 1) + dM \quad 7)$$

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 about the selection of the  $d$  parameter, see Ref. (Friedman, J.H. 1991). In our studies, 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 (Sekulic, S., and Kowalski, B.R. 1992):

1. Select the maximal 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 into the model.
5. Go 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 more than  $2M^*$  is enough in general (Friedman, J.H. 1991).



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  $f_i(\mathbf{x}_i^p)$  be the set of all single variable basis functions, i.e. basis functions that contain only  $\mathbf{x}_i^p$ . Similarly, let  $f_{ij}(\mathbf{x}_i^p, \mathbf{x}_j^p)$  be the set of all two-variable basis functions that contain the pairs of variables  $\mathbf{x}_i^p$  and  $\mathbf{x}_j^p$ , and  $f_{ijk}(\mathbf{x}_i^p, \mathbf{x}_j^p, \mathbf{x}_k^p)$  the set of all three-variable basis functions that contain the triplets of variables  $\mathbf{x}_i^p$ ,  $\mathbf{x}_j^p$  and  $\mathbf{x}_k^p$ . The MARS model can be rewritten in the following form:

$$\hat{f}(\mathbf{X}) = c_0 + \sum f_i(\mathbf{x}_i^p) + \sum f_{ij}(\mathbf{x}_i^p, \mathbf{x}_j^p) + \sum f_{ijk}(\mathbf{x}_i^p, \mathbf{x}_j^p, \mathbf{x}_k^p) \quad 8)$$

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 (Friedman, J.H. 1991). The two-variable interaction of a MARS model,  $If_{ij}(\mathbf{x}_i^p, \mathbf{x}_j^p)$ , is given by:

$$If_{ij}(\mathbf{x}_i^p, \mathbf{x}_j^p) = f_i(\mathbf{x}_i^p) + f_j(\mathbf{x}_j^p) + f_{ij}(\mathbf{x}_i^p, \mathbf{x}_j^p) \quad 9)$$

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  $f_i(\mathbf{x}_i^p)$  against  $\mathbf{x}_i^p$ . Two-variable interaction can be viewed by plotting  $If_{ij}(\mathbf{x}_i^p, \mathbf{x}_j^p)$  against  $\mathbf{x}_i^p$  and  $\mathbf{x}_j^p$  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. 1995):

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



where  $y_i$  is the value of dependent variable of the i-th object,  $\hat{y}_{-i}$  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:

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

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 is 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 into 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

Partiendo de objeto de la investigación, los datos proceden de dos tipos de variables, por un lado está la información procedente del cuestionario de estudiantes y en el otro los resultados de rendimiento académico.

La información del cuestionario de satisfacción se organiza en varios bloques de preguntas con temática semejante:

1. Cumplimiento Docente
2. Enseñanza
3. Actitud
4. Satisfacción General





El primer bloque, (Cumplimiento Docente) de 7 preguntas, recoge información de la percepción del alumnado acerca del cumplimiento docente, tanto en información sobre el programa, criterios de evaluación, adecuación de los contenidos de las clases, actividades y ejercicios que ilustran dichos contenidos, sistema de evaluación y por último la utilidad de la asistencia a clase y de las tutorías para preparar las asignaturas.

Las tres cuestiones siguientes (Enseñanza) analizan el desarrollo de las clases, conocimiento de la asignatura, claridad de exposición y materiales recomendados.

El tercer bloque (Actitud) se compone de tres preguntas referidas a la actitud mostrada por el profesorado hacia la asignatura y sus estudiantes.

El último bloque (Satisfacción General) valora la satisfacción general del alumnado con el trabajo realizado por el profesor o profesora y con lo aprendido en la asignatura.

La información de rendimiento académico responde a la variable:

Tasa de Éxito, (TÉX): definida como la relación porcentual entre el número total de créditos superados y el número total de créditos presentados por la totalidad de los estudiantes.

CÓDIGO	VARIABLE
PCD1	Información del Programa y Plan de Trabajo
PCD2	Criterios de Evaluación
PCD3	Contenidos
PCD4	Actividades y Ejercicios
PCD5	Sistema de Evaluación
PCD6	Asistencia a Clase
PCD7	Utilidad de la Tutoría
PENS1	Conocimiento de la Asignatura
PENS2	Explicación
PENS3	Materiales
PACTITUD1	Interés
PACTITUD2	Accesibilidad
PACTITUD3	Atención a Dificultades
PSATISF1	Aprendizaje
PSATISF2	Trabajo Profesor/a
TEX	Tasa de Éxito

Table 2. Name of the variables used in the model

## 4. Results and discussion

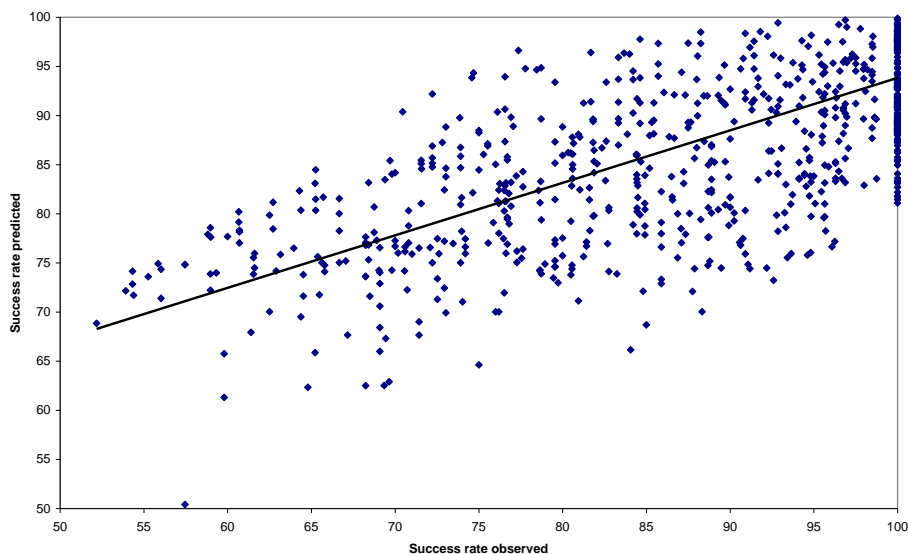
The MARS model was constructed for the output variable success rate. This regression model was build using the techniques in Friedman's papers "Multivariate Adaptive Regression Splines" (Friedman, J.H. 1991) and "Fast MARS". (Friedman, J.H. 1993)

The final model includes 23 basis functions, which are listed in Table 2 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 2) 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 the results of subjects (the other 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 success rate observed versus success rate predicted for the validation data is shown in Fig. 1. It can be observed how the highest the real success rate of the student the highest is the success rate predicted by the model. It can be observed in the same figure too that the model has difficulties predicting a success rate of 100%: (please see upper right part of the figure). The R-square obtained is 0.4662 and this would be higher if those subjects with a success rate of 100% were removed from the validation data set.

Los resultados de la figura 1, evidencian una clara agrupación de asignaturas cuya tasa de éxito alcanza el 100%. Analizando la tipología de estas asignaturas no se ha encontrado un perfil fijo que a primera vista nos haría intuir un carácter no obligatorio de las asignaturas, es decir asignaturas optativas y de libre configuración, sino todo lo contrario, incluye asignaturas obligatorias y optativas, no existiendo diferencias significativas en los resultados de los análisis, si omitimos las asignaturas no obligatorias.



**Figura 1. Scatterplot relating success rate observed versus success rate predicted (validation data).**

The importance of the explanatory variables in the MARS model is presented in Table 3. There are three main criteria to estimate the importance of the variables in a standard MARS model (Friedman, J.H. 1991) and they are presented in Table 3. 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 the same, but uses the 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 related with PCD (PCD5, PCD7, PCD1, PCD3, PCD4, PCD6), PENS (only PENS1), attitude (PACTITUD2, PACTITUD1, PACTITUD3) and finally satisfaction (PSATISF1).

La variable más influyente es el año académico, análisis predecesores nos impiden inferir que los estudiantes con menor capacidad abandonan en los primeros cursos, o que los mejores profesores dan clase en los cursos superiores. En un estudio realizado en 2001 en la Universidad de Oviedo, se concluyó que no se apreciaban diferencias significativas entre la baja tasa de éxito en los primeros cursos y la procedencia LOGSE o COU de los estudiantes (De Miguel 2001) en relación con el nivel. El origen, por tanto hay que buscarlo en el proceso de adaptación a los estudios universitarios (De Miguel 2001) o en otras variables.

El bloque de variables PCD (cumplimiento docente) se revela como el más importante de cara a la determinación de la tasa de éxito, debido al número de variables que correlacionan con el éxito. Por orden de importancia son: *Sistema de Evaluación, Utilidad de la Tutoría, Información sobre el Programa y Plan de Trabajo, Contenidos, Actividades y ejercicios, Asistencia a Clase.*

Del bloque de Enseñanza la única que influye es la percepción subjetiva que tiene el estudiante sobre el conocimiento de la asignatura, como se puede comprobar en la figura 2 y en la tabla 3.

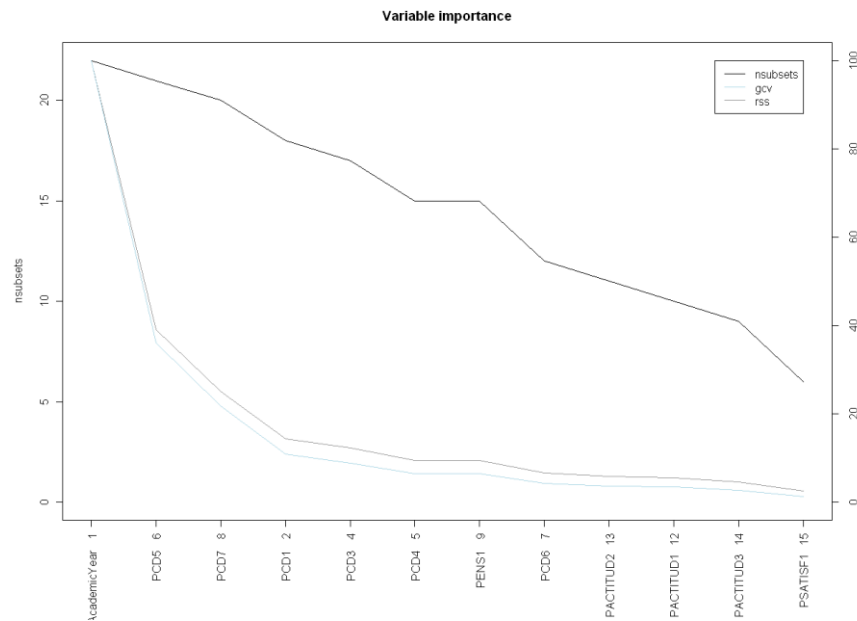


Figura 2. Importance of the explanatory variables in the MARS model

	nsubsets	GCV	RSS
<b>AcademicYear</b>	<b>22</b>	<b>100.00</b>	<b>100.00</b>
<b>PCD5: Sistema de Evaluación</b>	<b>21</b>	<b>36.03</b>	<b>38.99</b>
<b>PCD7: Utilidad de la Tutoría</b>	<b>20</b>	<b>21.61</b>	<b>25.11</b>
<b>PCD1: Información del Programa y Plan de Trabajo</b>	<b>18</b>	<b>18.88</b>	<b>14.38</b>
<b>PCD3: Contenidos</b>	<b>17</b>	<b>8.85</b>	<b>12.24</b>
<b>PCD4: Actividades y Ejercicios</b>	<b>15</b>	<b>6.43</b>	<b>9.49</b>
PENS1: Conocimiento de la Asignatura	15	6.43	9.49
PCD6: Asistencia a Clase	12	4.21	6.87
PACTITUD2: Accesibilidad	11	3.60	5.89
PACTITUD1: Interés	10	3.42	5.49
PACTITUD3: Atención a Dificultades	9	2.75	4.62
PSATISF1: Aprendizaje	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-PCD3)	3.696258
5	h(PCD4-6.28571)	-1.412395
6	h(PCD5-5.45455)	7.187522
7	h(PCD6-8.61538)	10.175454
8	h(PCD7-4.71429)	-3.363943
9	h(AcademicYear-4) * h(8.61538-PCD6)	1.302150
10	h(3-AcademicYear) * h(PACTITUD1-9.33333)	20.728190
11	h(PCD1-9.21429) * h(7.73913-PCD3)	11.601691
12	h(9.21429-PCD1) * h(7.73913-PCD3)	-0.420384
13	h(PCD1-7.8) * h(PCD5-5.45455)	-2.086423



14	$h(7.8-PCD1) * h(PCD5-5.45455)$	-1.227524
15	$h(PCD3-8.5) * h(8.61538-PCD6)$	-14.394898
16	$h(6.28571-PCD4) * h(PENS1-7.3)$	-3.838091
17	$h(6.28571-PCD4) * h(7.3-PENS1)$	-0.540679
18	$h(8.61538-PCD6) * h(PSATISF1-7)$	2.064487
19	$h(8.61538-PCD6) * h(7-PSATISF1)$	-0.353424
20	$h(PCD7-4.71429) * h(PACTITUD2-9.16667)$	-9.390150
21	$h(PCD7-4.71429) * h(PACTITUD3-5.33333)$	0.601493
22	$h(4-AcademicYear) * h(PCD3-5) * h(8.61538-PCD6)$	0.168861
23	$h(4-AcademicYear) * h(5-PCD3) * h(8.61538-PCD6)$	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 2. Y las relaciones más interesantes se presentan en los gráficos siguientes.

En la figura 3 se puede comprobar como la tasa de éxito es más alta en los cursos superiores (4-6) y cuando la asistencia es mayor (5-10).

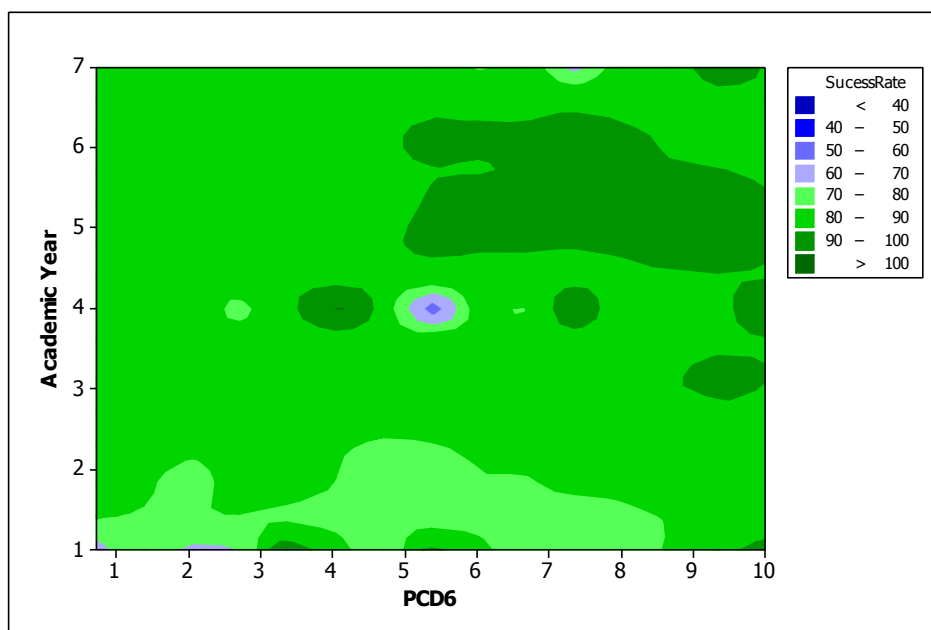
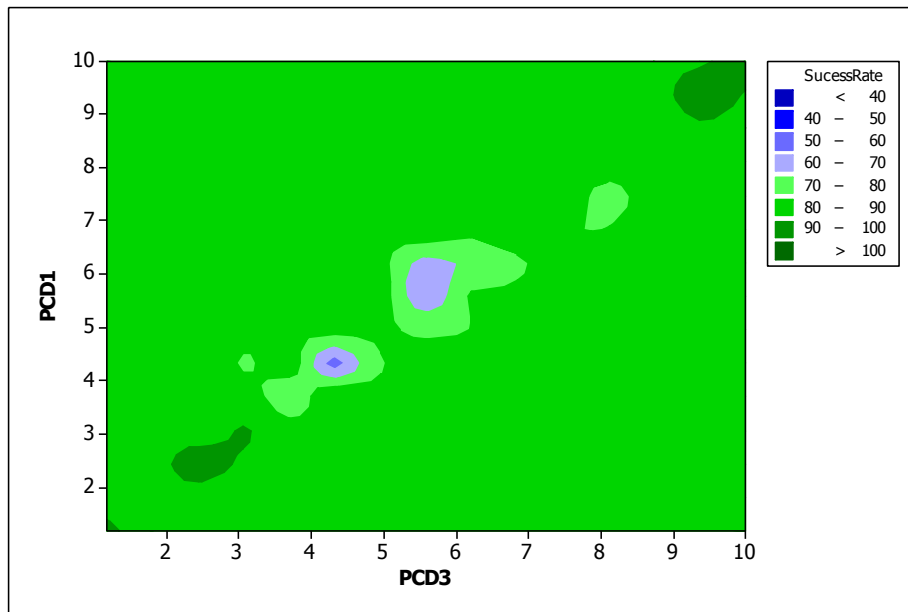


Figura 3. Contour Plot of SuccessRate vs Academic Year; PCD6: Asistencia a Clase.

En la figura 4 se detecta que la tasa de éxito disminuye cuando la valoración de los estudiantes sobre la Información sobre el Programa y Plan de Trabajo; y los Contenidos están entre el intervalo 5 y 6.



**Figura 4. Contour Plot of SuccessRate vs PCD1: Información sobre el Programa y Plan de Trabajo; PCD3: Contenidos**



## 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 it is shown in Figure 2 and Table 3.

Los resultados obtenidos avalan la hipótesis alternativa, *la satisfacción percibida por los estudiantes durante el primer cuatrimestre nos va permitir predecir el éxito en la asignatura al finalizar el curso*, sin embargo la baja correlación entre las variables no nos permite garantizar la precisión del modelo.

Los resultados nos permiten predecir el éxito en una asignatura principalmente en función del año académico y de la satisfacción de los estudiantes en las preguntas del bloque de Cumplimiento Docente: *Sistema de Evaluación, Utilidad de la Tutoría, Información, Contenidos, Actividades, Asistencia a Clase*. Permitiendo a los responsables académicos, por ejemplo, iniciar acciones de mejora continua e innovación de la actividad docente, adecuar el plan docente (planificación, desarrollo y resultados) a las necesidades/expectativas de los estudiantes, realizar una valoración del esfuerzo docente en el cumplimiento de su actividad, entre otras.

La aplicación de este modelo predictivo nos permite extraer alguna conclusión más como la distribución de la importancia de los diferentes ítems del cuestionario para los estudiantes.

Como futura línea de investigación, en la que de hecho estamos trabajando consiste en analizar los factores de riesgo de este modelo.

## 6. Referencias Bibliográficas

- Bisquerra Alzina, R. (1989): *Métodos de Investigación Educativa*. Ceac. Barcelona.
- Cabrera Guillén, P. y Galán Delgado, E. (2002). Factores contextuales y rendimiento académico. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 5 (3).
- De Miguel Díaz, M (Director) (2001): *Evaluación del Rendimiento en la Enseñanza Superior*. Ministerio de Educación Cultura y Deporte.
- Friedman, J.H. (1991) 'Multivariate adaptive regression splines', *Ann. Stat.*, Vol. 19, pp. 1-141.
- Friedman, J. H (1993). "Fast MARS." Dept. of Statistics, Stanford University Technical Report.



- 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.
- Hastie, T., Tibshirani, R., and Friedman, J.H. (2003) *The Elements of Statistical Learning*, Springer-Verlag, New York.
- Harvey, L. (2003). Student feedback. *Quality in Higher Education*, 9 (1), pp3-20.
- Sekulic, S., and Kowalski, B.R. (1992) 'MARS: a tutorial', *J. Chemom.*, Vol. 6, pp. 199-216.
- 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.