International Journal of Engineering Education Vol. 34, No. 3, pp. 895–904, 2018
Printed in Great Britain
0949-149X/91 $3.00+0.00
© 2018 TEMPUS Publications.

Hybrid Intelligent System to Predict the Individual Academic Performance of Engineering Students*

ESTEBAN JOVE[1,2], JOSÉ ANTONIO LÓPEZ-VÁZQUEZ[1], MARÍA ISABEL FERNÁNDEZ-IBÁÑEZ[1], JOSÉ-LUIS CASTELEIRO-ROCA[1,2] and JOSÉ LUIS CALVO-ROLLE[1]
[1] University of A Coruna, Department of Industrial Engineering, 19 de febrero Avenue, 15495, Ferrol, A Coruna, Spain.
[2] University of La Laguna, Department of Computer Science and System Engineering, Astroísico Francisco Sánchez Avenue, 38200, La Laguna, Tenerife, Spain. E-mail: esteban.jove@udc.es (Esteban Jove), jose.lopez@udc.es (José Antonio López-Vázquez), isabel.fibanez@udc.es (María Isabel Fernández-Ibáñez), jose.luis.casteleiro@udc.es (José-Luis Casteleiro-Roca), jlcalvo@udc.es (José Luis Calvo-Rolle).

The present research deals with the problem of the low academic performance in engineering degrees. It has been done by analyzing the academic results of students that have attended to the bachelor's degree in industrial engineering. The dataset used was the official registered grades of their academic records in the University of A Coruña. In this work, several advanced regression techniques have been applied to create models of the academic itinerary during each student career. The created models can predict the number of attempts in every subject that the students would take in the first academic year. The attempts of the subjects from the rest of the years can be predicted by incorporating to the model the results of the already passed subjects. The developed model allows to achieve the proposed goal of predicting the academic performance of each student.

Keywords: academic performance; SVM; MLP

1. Introduction

The academic performance of university students is one of the main concerns on the high education system in Spain [1]. Different education parameters such as the exam absence and failure rates, the Grade Point Average (GPA) or the number of attempts needed to pass an exam show a low academic performance that is kept along the years [2].

Several researches performed on technical engineering degrees, like [3], show that only the 11% of the students obtain the graduate after the three years established by the educational curriculum. These studies also evidence that the average time spent to finish the degree is 5.41 years, the dropout rate is around 70% and the performance rate is 56%.

In addition to the studies mentioned above, more recent works, like [4, 5], accomplished on the Spanish University System (SUE, Sistema Universitario Español) show that engineering and architecture performance rates remain around the 60%, being the lowest of the SUE. Compared to studies with better performances, such as health sciences degrees, the engineering and architecture rates are 20% lower.

Researches on several university systems [6, 7] show that different factors can be used to predict the academic performance. The average academic record achieved at secondary school is the best predictor for university success/failure. Furthermore, according to [1] and [8], the recent academic results are more significant over the current performance. Hence, for the rest of the university courses, the best performance predictors are based on the previous courses results instead of the academic record of secondary school.

Most of the studies based on university academic performance prediction or learning model, were focused on the estimation of students' success or failure over a precise career at a specific university. The global performance during their studies is also taken into account [9, 10]. In the previously mentioned works, the used techniques are based on traditional statistic algorithms and advanced datamining [11]. Some studies were focused on predicting student success/failure possibilities [12–14]. Other works analyze the global performance over a specific career or subject [15–18]. The most used methods in Educational Data Mining (EDM) are the Decision Trees, Artificial Neural Networks (ANN) and Bayesian Networks, being the most common applied techniques: regression, classification and clustering [19, 20]. In [21] is shown a tool for achieving a student learning model during a specific work session. A method to improve the students' group assessment is described in [22]. [23] defines three system-independent classifications of interactions and evaluates the relation of their components with academic performance.

The aim of this research is the prediction of how many attempts will expend each student to pass a specific subject. It is necessary to emphasize the

* Accepted 12 January 2018.

difficult task of this prediction due to the fact that the studied problem has a very high non-linear component. Additionally, the created model could be a useful tool for helping the students to take decisions, for instance, to know in what subjects they are going to need to work more. Also, a knowledge base creation to advise the new cases could be implemented [24].

The above mentioned high non-linear component was the main reason to use intelligent regression techniques and to create a hybrid model, which gives more advantages over these kind of problems [25, 26]. The regression algorithms are commonly used to solve other non-linear problems like [27, 28], where their use leads to better performance. Usually, in order to create hybrid models, the dataset is partitioned by a clustering algorithm and, then, the regression techniques are trained and tested for each cluster. Several researches like [29, 30] show the achieved error reduction when the dataset is clustered before the regression techniques were applied, with a significant global improvement.

This paper is structured as follows: after this introduction, the case of study is described in the next section. Then, the regression techniques used, and the model approach are explained. In the results section, the outcomes and the best hybrid model are presented. Finally, the conclusions and future works are exposed.

2. Case of study

The aim of this work is to predict the students' academic performance using intelligent regression techniques. A general view of the case of study is presented in this section before explaining and implementing the regression techniques.

2.1 Degrees under study

This research analyzes the academic performance of two different degrees taught at the Polytechnic University College of the University of A Coruña (UDC): automation and industrial electronics engineering degree and electrical engineering degree. Both degrees were based on a 1997 educational curriculum of 236 credits, divided into three academic years. However, as Fig. 1 shows, only 13% of the graduates finished their studies in those three years.

A recent study performed on the 2015/2016 course confirms that this low performance trend is steady during the years; the performance rate of the electrical engineering degree and the automation and industrial electronics engineering degree were 53.96% and 50.34% respectively.

Due to the reasons explained before, the UDC understands that it is necessary to study the problem of the low academic performance at the engineering degrees, with the aim of improving these rates.

2.2 Dataset

The dataset was obtained in two different stages:

- First, the UDC provided confidential academic information of 2,736 students who joined the university between the academic years 1996/1997 and 2009/2010. With the aim of acquiring a general view of the dataset under study, a preliminary statistical analysis of different variables such as: gender, access method, admission grade, number of years spent to obtain the graduate, or the geographical origin was accomplished.
- During the second stage, more detailed information about the students, like their academic records was requested to the UDC.
- Table I shows an example of how these academic records are structured. Two groups of students are discarded before this study: the ones who could not finish the degree and the ones who enrolled the UDC before 2001, due to the lack of numeric grades on their academic records. The final dataset used to develop the model approach is composed by the academic records of 225 automation and industrial electronics graduates and 254 electrical engineering graduates. In order to make the model capable of predicting the number of attempts in the widest range of students, all the data is considered to obtain the model. For this reason, different outliers that consist on students with extremely good grades and students very poor academic records were not excluded from the original dataset.

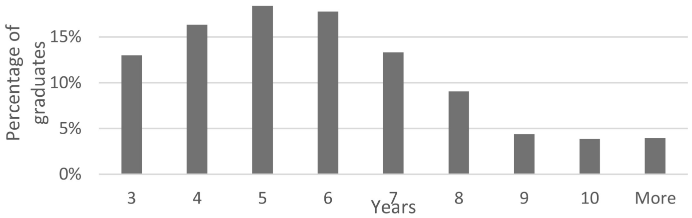

Fig. 1. Years needed to finish the degree.

Table 1. Example of information contained in the academic record

Academic Record							
			Subject 1		. . .	Subject *n*	
Student code	**Access method**	**Admission grade**	**Grade**	**Attempts**		**Grade**	**Attempts**
22002	H. school	7.5	8.5	1		6	2
32002	A. degree	7.0	5.0	3		8.0	1

The information considered to achieve the academic performance predictive model is summarized as follows:

- University admission grade: it is the average grade obtained by the students in their previous studies.
- Access method: it represents the way used by a student to enroll in the university. Two main cases are considered: high school or associate degree.
- Subject grades obtained by the student along the degree.
- Number of attempts spent by the student to pass each subject.

In case of missing academic information of some students, imputation techniques could be applied [31] to retrieve the information.

3. Model approach

The scheme defined for each model is shown in Fig. 2. Taking into account the system behavior and the test accomplished, the dataset can be divided in several operation ranges. Thus, several clusters are created and, for each one, a regression model is obtained with a single output that represent the number of attempts spent to pass a subject. The global model has two set of inputs (the grades obtained for the already passed subjects and the number of attempts to pass them). The cluster selector block connects the chosen models with the outputs. On each cluster block, only the best model is implemented. The cluster selection for a specific input is based on the Euclidean distance between the input and the centroids on each cluster.

The process followed to train and test the models is shown on Fig. 3. The initial dataset was conditioned to ensure that all samples have the same data.

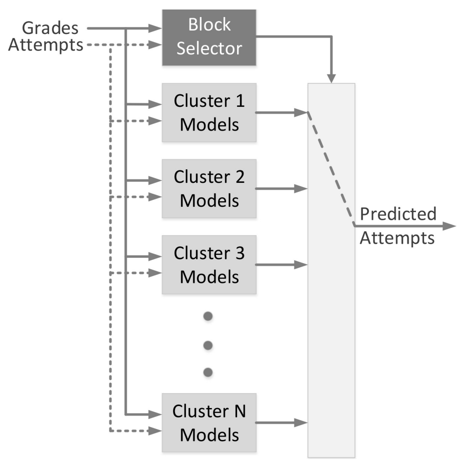

Fig. 2. Model approach.

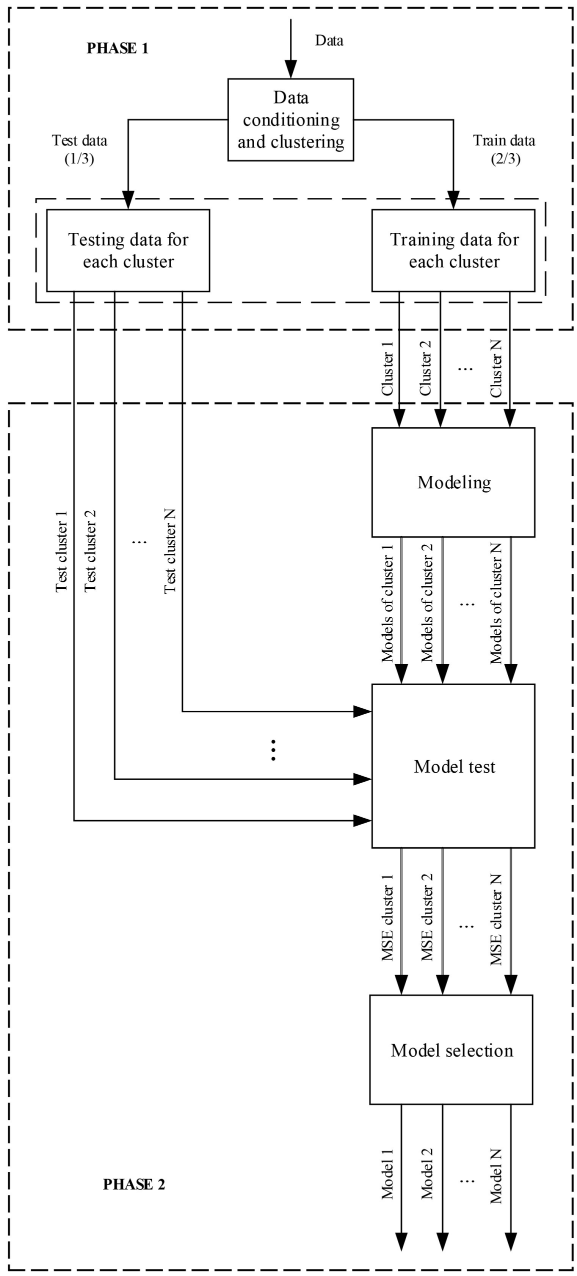

Fig. 3. Modeling process.

Then, the data is divided in N clusters prior to train the models; this step is necessary to ensure that there are testing and training samples in all groups. After the clusters were created, the data from each one is divided using Hold Out validation by using two third of the data to train and one third to test the model. Different algorithms were tested, selecting the best one for each cluster depending on the Mean Square Error (MSE) achieved with each regression technique. Once all the best models for each cluster were selected, the performance of each hybrid model (with different number of clusters) is compared using a mean of the errors, taking into account the number of samples in each cluster.

It is remarkable that the model number of inputs for each course is different. This is a common case due to the fact that in the last course, for example, the students should have more passed subjects than in the first course. Moreover, the first course models have only two inputs, the access grade and the way used to enter in the university. The different models are classified in three types:

- The first model type has only two inputs: the access method and the admission grade. With this data, the models predict the number of attempts for the first course subjects.
- The second model type predicts the number of attempts for the second degree course. In this case, the inputs are the ones used on the first model type and the grades and attempts of the first course subjects.
- In the third model type, the inputs include the subjects of the second course and the outputs are the information for the last course subjects.

The three different model types are presented in Fig. 4.

3.1 Used techniques

The techniques tested in the study to achieve the best models are described below.

3.1.1 Data clustering. The K-Means algorithm

K-Means is an unsupervised technique of data grouping where similarity is measured [32], [33]. This clustering algorithm tries to organize unlabeled feature vectors into clusters or groups, in such way that samples within a cluster are similar to each other [34]. K-Means algorithm is a commonly used partitional clustering algorithm with square-error criteria, which minimizes the error function shown in Equation 1.

$$e = \Sigma_{k=1}^{C} \Sigma_{x \in Q_k} ||x - c_k||^2 \qquad (1)$$

The final clustering will depend on the initial cluster centroids and on the value of K (number of clusters). Choosing K value is the most critical election because it requires certain prior knowledge of the number of clusters present in the data, which is highly doubtful. The K-Means partitional clustering algorithm is computationally effective and works well if the data are close to its cluster and the cluster is hyperspherical in shape and well-

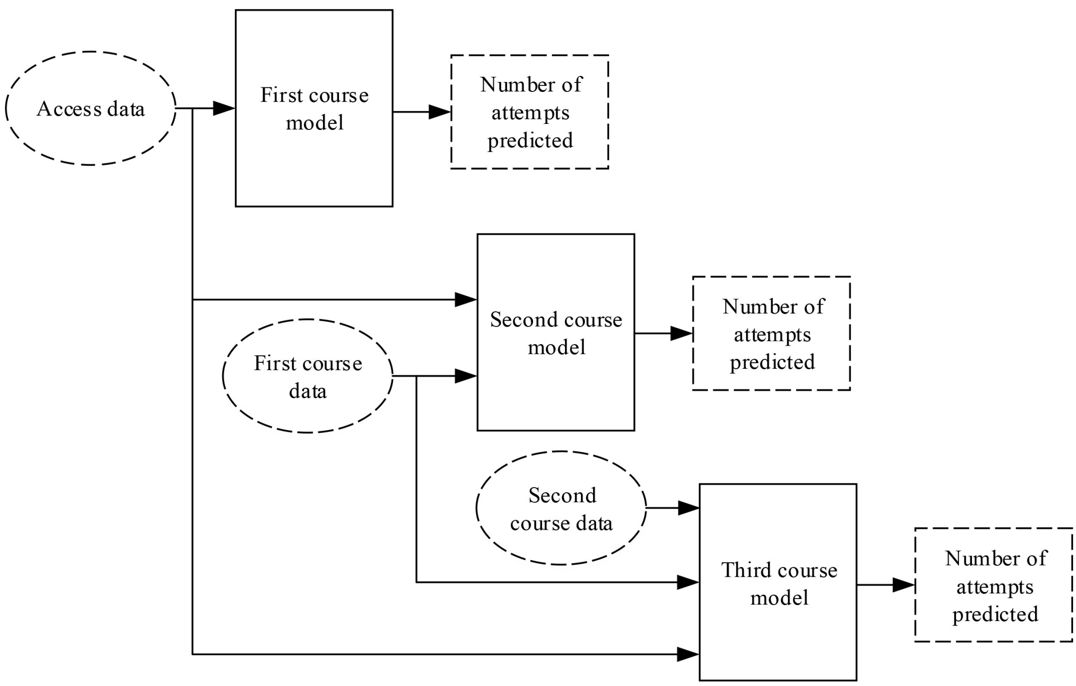

Fig. 4. Model types.

separated in the hyperspace. As the final division depends on the initial cluster centroids, in order to create the best division, K-means algorithm should be used several times and the best division (in terms of distance between centroids) is chosen. The final centroids must be stored to assign new data to correct cluster.

3.1.2 Regression techniques

Polynomial regression

Generally, a Polynomial regression model [35–39] may also be defined as a linear sum of basis functions. The number of basis functions depends on the number of the model inputs, and the degree of the polynomial used.

With a first degree polynomial, the linear sum could be defined as the one shown in equation 2 (for two inputs). The model becomes more complex as the degree increases, equation 3 shows a second degree polynomial.

$$F(x) = a_0 + a_1x_1 + a_1x_2 \tag{2}$$

$$F(x) = a_0 + a_1x_1 + a_1x_2 + a_3x_1x_2 + a_4{x_1}^2 + a_5{x_2}^2 \tag{3}$$

Although the polynomial regression algorithm is not an intelligent technique, it shows good results when the data is very disperse; in these cases, the other techniques use to over train the model and the results are not as good as they are expected.

Artificial Neural Networks (ANN): MultiLayer Perceptron (MLP)

A Multilayer Perceptron is a feedforward Artificial Neural Network [40, 41]. It is one of the most typical ANNs due to its robustness and relatively simple structure. However, the ANN architecture must be well selected to obtain good results. The MLP is composed by one input layer, one or more hidden layers and one output layer.

Each layer is made of neurons, with a specific activation function. In a usual configuration, all neurons from a layer have the same activation function. This function could be: Step, Linear, Log-sigmoid or Tan-sigmoid. Equation 4 shows the Tan-sigmoid function. All layers in MLP have weighted connections between neurons of each layer.

$$F(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}} \tag{4}$$

It is possible to define the output of a MLP as shown in equation 5 [42].

$$f_\theta(x) = \beta + \Sigma_{i=1}^{k} a_i \phi({w_i}^T x + b_i) \tag{5}$$

where:

$x = (x(1), \ldots, x(d))^T \in \mathfrak{R}^d$ is the vector of inputs
k is the number of hidden layers
ϕ is a bounded transfer function
$\theta = (\beta, a_1, \ldots, a_k, b_1, \ldots, b_k, w_{11}, \ldots, w_{kd})$ is the parameter vector of the model
$w_i = (w_{i1}, \ldots, w_{id})^T \in \mathfrak{R}^d$ is the parameter vector for the hidden unit i

The use of this type of ANN has offered successful results in a wide variety of application [43–45].

Support Vector Regression (SVR), Least Square Support Vector Regression (LS-SVR)

Support Vector Regression is a modification of the Support Vector Machines (SVM) algorithm used for classification. In SVR the basic idea is to map the data into a high-dimensional feature space F via a non linear mapping and apply linear regression in this space [46–48].

Least Square formulation of SVM, is called LS-SVM. The approximation of the solution is achieved by solving a system of linear equations, and it is comparable to SVM in terms of generalization performance [49], [50]. The application of LS-SVM to regression is known as LS-SVR [51], [52]. In LS-SVR, the insensitive loss function is replaced by a classical squared loss function, which constructs the Lagragian by solving a linear KarushKuhn-Tucker (equation 6).

$$\begin{bmatrix} 0 & {I_n}^T \\ I_n & K + \gamma_{-1} I \end{bmatrix} \begin{bmatrix} b_0 \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{6}$$

where:

I_n is a vector of n ones
T means transpose of a matrix or vector
γ a weight vector
b regression vector
b_0 is the model offset

In LS-SVR, only two parameters (γ, σ) are needed. Where, σ is the width of the used kernel [52].

Although the SVM is used commonly in classification, this modification of the algorithm shows good performance in terms of error when the number of samples is sufficient. If the number of samples is low, its use is not recommended because it would cause over training and the model could not be used with new data.

4. Results

A wide combination of tests was performed to select the best configuration for each model. All the used algorithms have the same configuration before the test. To ensure the best clustering partition, the K-Means algorithm was used with 20 replicates, with random initialization of the centroids; the final configuration is selected as the best of the 20 results obtained. The number of clusters in which the data is divided varies from 1 to 9. The criteria to check the performance of the number of clusters selected consists on the error obtained at the regression stage. The results show that, over 9 clusters, an increase in the number of clusters implies an increase in the error obtained. Besides, given the size of the dataset, the use of higher values of K could lead to clusters without a significant number of samples to accomplish the regression.

The MLP (ANN) regression algorithm was trained for different configurations; always with one hidden layer, but the number of neurons in the hidden layer varies from 1 (ANN 1) to 15 (ANN 15). The activation function of this internal neurons was tan-sigmoid for all tests and the output layer neuron had a linear activation function. The used training algorithm was Levenberg-Marquardt; gradient descent was used as learning algorithm and the set performance function was the Mean Squared Error.

The LS-SVR was trained with the self-tuning script implemented using the Matlab toolbox developed by KULeuven-ESAT-SCD. The kernel model was set to Radial Basis Function and the type was 'Function Estimation' to perform regression. The optimization function is 'simplex' and the cost-criterion is 'leaveoneoutlssvm' with 'mse' as a performance function.

For Polynomial regression, the order of the trained polynomial varies from 1st (Poly 1) to 3rd (Poly 3) order.

Table 2 shows the different algorithm configurations obtained for each hybrid model to predict the number of attempts needed to pass the Physics subject in the electrical engineering degree. In this specific subject, the best model performance was achieved by dividing the data into 5 clusters. A different cluster configuration was obtained for each subject.

Table 3 shows the Mean Absolute Error (MAE) obtained by the hybrid model in number of attempts prediction. In this table, it is possible to see the performance of the model. In the cases shown, the worst error obtained for the hybrid configuration is nearly 0.7, which means an error of 1 attempt.

Table 4 summarizes the results for best model achieved to predict attempts in the Physics subject in electrical engineering degree, including the regression technique parameters applied to each cluster.

In general terms, all the models have a very good performance. Taking into consideration all the

Table 2. Algorithm configurations used to predict Physics subject attempts in electrical engineering degree

No. of clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
1	ANN 2								
2	ANN 3	Poly 1							
3	ANN 1	Poly 1	Poly 1						
4	LS-SVR	Poly 1	Poly 1	Poly 1					
5	**Poly 1**	**LS-SVR**	**Poly 1**	**ANN 1**	**ANN 2**				
6	ANN 2	Poly 1	ANN 1	ANN 1	ANN 1	LS-SVR			
7	Poly 1	ANN 1	LS-SVR	Poly 1	Poly 1	ANN 1	LS-SVR		
8	ANN 1	LS-SVR	Poly 1	ANN 1	Poly 1	Poly 1	ANN 2	Poly 1	
9	ANN 1	Poly 1	Poly 1	ANN 1	LS-SVR	ANN 1	Poly 1	ANN 3	LS-SVR

Table 3. Mean Absolute Error for Physics subject in electrical engineering degree

No. of clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
1	0.7848								
2	0.7111	0.7476							
3	0.5958	0.7591	0.7034						
4	0.6776	0.6215	0.6647	0.6872					
5	**0.6828**	**0.6154**	**0.6944**	**0.5900**	**0.6281**				
6	0.7793	0.5722	0.7836	0.6965	0.5990	0.6817			
7	0.7991	0.6580	0.5852	0.6823	0.6722	0.7382	0.6778		
8	0.7127	0.6524	0.7423	0.7837	0.6548	0.7208	0.6428	0.7157	
9	0.7574	0.7463	0.7097	0.6601	0.6957	0.6756	0.7788	0.7413	0.7097

Table 4. Model parameters in Physics subject in electrical engineering degree

Cluster	Regression technique	Model parameters	MAE
Cluster 1	Poly 1	Independent term = 3.2224 Access method coef. = 0.14716 Access grade coef. = 1.5371	0.6828
Cluster 2	LS-SVR	$\gamma = 1.7118^{\circ}$ $\sigma^2 = 3.4805$ $b = 1.5371$	0.6154
Cluster 3	Poly 1	Independent term = 2.6395 Access method coef. = 0.49412 Access grade coef. = −2.25239	0.6944
Cluster 4	ANN 1	$\omega_{hidden} = \begin{bmatrix} 358.9922 \\ -802.9025 \end{bmatrix}$ $\beta_{hidden} = [-95.0370]$ $\omega_{output} = [0.1169]$ $\beta_{output} = [-0.7655]$	0.5900
Cluster 5	ANN 2	$\omega_{hidden} = \begin{bmatrix} -10.2393 & -5.1983 \\ 8.3290 & -1.1225 \end{bmatrix}$ $\beta_{hidden} = \begin{bmatrix} 7.9124 \\ -6.1045 \end{bmatrix}$ $\omega_{output} = \begin{bmatrix} -0.2364 \\ 0.3611 \end{bmatrix}$ $\beta_{output} = [-0.4404]$	0.6281

Table 5. Intermediate result for Graphic Expression

Graphic Expression—Electrical Engineering Degree					
Number of clusters	**1**	**2**	**3**	**4**	**5**
Best algorithms	ANN-3	LS-SVR	Poly-1	Poly-1	**Poly-4**
MAE	0.7825	0.7993	0.8077	0.8473	**0.7822**
MSE	1.2015	0.9483	1.1468	0.9619	**1.0124**

subjects studies during the degree, the worst result is 1.25 absolute error. It means that the model proposed predicts the number of attempts needed to pass each subject with a maximum error of one attempt. The best absolute error is 0.31 and is the same for 3 different subjects.

Table 5 shows an example of the intermediate results obtained for the Graphic Expression subject. It is shown that for each subject, different configurations are studied, and the best hybrid model is chosen by the lower achieved error. For a right understanding, two error types were used, the MAE and the Mean Squared Error (MSE). The MSE is chosen to detect the best regression technique and the MAE is used due to the real meaning representation.

Figs. 5 and Fig. 6 show the MAE distribution for all the subjects of the degrees.

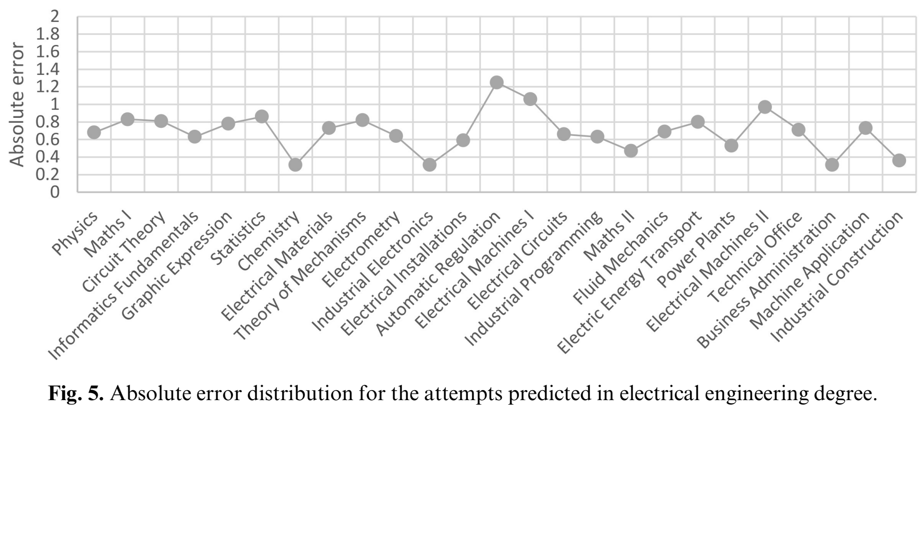

Fig. 5. Absolute error distribution for the attempts predicted in electrical engineering degree.

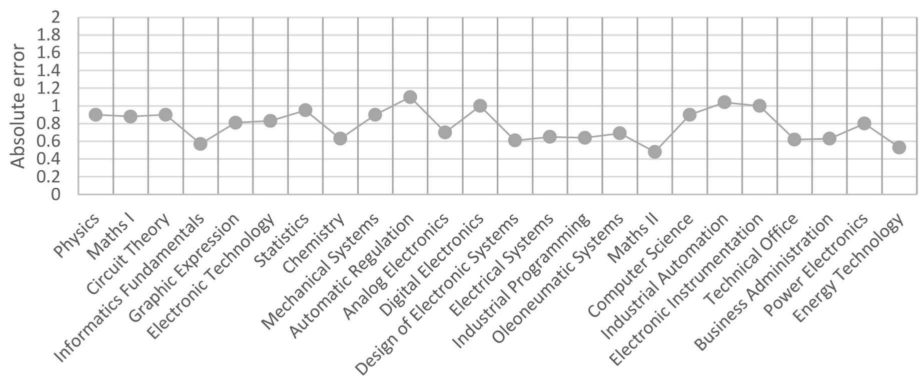

Fig. 6. Absolute error distribution for the attempts predicted in automation and industrial electronics engineering degree.

5. Conclusions

With the proposed model, the academic performance of the students was possible to predict. The good accuracy achieved can help the students to detect which subjects may represent more difficulties for themselves during the degree. This could be a useful tool to improve different parameters such as the academic performance, the GPA and the dropout rates in all academic years.

The best results were obtained with a hybrid model for each subject. As all the models are trained independently, each one is the best for a specific subject. The worst results were achieved for the prediction of Automatic Regulation subject in both degrees. In this subject, the absolute errors obtained in the attempts prediction were 1.25 attempts in the electrical engineering degree and 1.10 attempts in the automation and industrial electronics engineering degree. On the other hand, the lowest absolute errors were 0.31 in three subjects in the electrical engineering degree (Chemistry, Industrial Electronics and Business Administration). In the automation and industrial electronics engineering degree, the lowest error was 0.48 attempts in the subject Mathematics II. This means that the worst prediction gives an error of at most one attempt. Taking into consideration that only two attempts are allowed in one academic year, the proposed model achieves less than one year deviation.

It is remarkable that the predictions made on students with extremely high and low grades present higher errors than the predictions made the rest of students. These extreme cases produce an increase in the absolute error obtained in the prediction.

6. Future works

As future works, there are some different research lines to follow. The first one could be verifying the applicability of the achieved model over another degrees and universities. Due to the fact that the students, staff, and so on are different, the model may not be valid to predict this specific students' performance. Before its application, it is necessary to check its performance over a significant sample. However, it is possible to create a new model with the dataset of the specific case following the same methodology. The difference of this new model and the one proposed on this work would consist on a new configuration with its own parameter values.

Regarding the model implementation, it is important to check alternative methods for clustering and regression, especially, emergent techniques. Also, it is necessary to verify another indicator or/and criteria for the regression and clustering quality measurements to increase reliability and accuracy.

The proposed model is expected to be implanted soon with the aim of helping the students' decision-making in the framework of Tutorial Action Plans.

When the model is used for the students' advice, one of the main problems could be the students' performance change. Then, it is required to add adaptability to the model due to this variation. With this aim, it could be necessary to obtain new models frequently, adding the new students' performance. In case that this fact occurs, the use of the simplest modeling would be mandatory; maybe it could be necessary to obtain ad-hoc models. It implies, for instance, the use of global models instead of local models for different behavioral zones.

Creating students' profiles considering failure parameters such as abandonment and delays during studies could be very helpful. Then, it would be possible to mitigate common weaknesses that allow to improve the students' academic performance.

References

1. M. Miguel Díaz, *Evaluación del rendimiento en la enseñanza superior: resultados entre alumnos procedentes de la LOGSE y del COU*. Oviedo: Ministerio de Educación, 2002.
2. F. Javier and T. Tejedor, Causas del bajo rendimiento del

estudiante universitario (en opinión de los profesores y alumnos). Propuestas de mejora en el marco del EEES,'' pp. 443–473.
3. A. Espina, La Formación Técnica Postsecundaria y la Competitividad de la Economía Española, *Rev. Española Investig. Sociológicas*, **77**, 1997, pp. 69–115.
4. F. Michavila, La Universidad espa{ñ}ola en cifras, Madrid: CRUE, 2012.
5. C. Pérez, *El Rendimiento Académico en las Universidades Españolas*. Madrid: Conferencia de Rectores de las Universidades Españolas, 2012.
6. O. Frischenschlager, G. Haidinger and L. Mitterauer, Factors associated with academic success at Vienna Medical School: prospective survey, *Croat. Med. J.*, **46**(1), 2005, pp. 58–65.
7. M. Salanova, W. Schaufeli, I. Martinez and E. Breso, How obstacles and facilitators predict academic performance: the mediating role of study burnout and engagement, *Anxiety, Stress, Coping*, **23**, 2010, pp. 53–70.
8. F. Tejedor, Poder explicativo de algunos determinantes del rendimiento en los estudios universitarios, *Rev. Española Pedagog.*, **LXI**, 2003, pp. 5–32.
9. L. Cohen, L. Manion and K. Morrison, *Research methods in education*, Oxon: Routledge, 2007.
10. J. Heredia, A. Rodríguez and J. Vilalta, Predicción del rendimiento en una asignatura empleando la regresión logística ordinal, *Estud. Pedagógicos*, **XL**(1), 2014, pp. 145–162.
11. S. Palmer, Modelling engineering student academic performance using academic analytics, *Int. J. Eng. Educ.*, **29**(1), 2013, pp. 132–138.
12. N. T. N. Hien and P. Haddawy, A decision support system for evaluating international student applications, *Proc.—Front. Educ. Conf. FIE*, 2007, pp. 1–6.
13. S. Romero Morales, Cristóbal; Márquez Vera, Carlos; Ventura Soto, Predicción del Fracaso Escolar Mediante Técnicas de Minería de Datos, *IEEE-Rita*, **7**(3), 2012, pp. 109–117.
14. H. Wilhelmiina, Comparison of machine learning methods for intelligent tutoring systems, *ITS'06 Proc. 8th Int. Conf. Intell. Tutoring Syst.*, 2006, pp. 525–534.
15. B. Minaei-Bidgoli, D. a. Kashy, G. Kortemeyer and W. F. Punch, Predicting student performance: an application of data mining methods with an educational web-based system, *33rd Annu. Front, Educ, 2003, FIE 2003*, **1**, 2003, p. T2A_13–T2A_18.
16. D. Ibrahim, Zaidah, Rusli, Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision tree And Linear Regression, *Proc. 21st Annu. SAS Malaysia Forum*, no. September, 2007, pp. 1–6.
17. R. Shaun, J. De Baker and J. E. B. Eds, Educational Data Mining 2008 The 1st International Conference on Educational Data Mining, *Network*, 2008, p. 187.
18. K. Gaspar Fàbregas, M. A. Núñez Andrés, J. Rodriguez Jordana, J. Riba and F. de Paula, Influence of admission marks on the academic performance of technical architecture students, *Int. J. Eng. Educ.*, **32**(1(A)), 2016, pp. 209–218.
19. C. Romero and S. Ventura, Educational Data Mining: A Review of the State of the Art, *IEEE Trans. Syst. Man, Cybern. Part C Applications Rev.*, **40**(6), 2010, pp. 601–618.
20. V. Cakir and A. Gheorghe, Longitudinal Academic Performance Analysis Using a Two-Step Clustering Methodology, *Int. J. Eng. Educ.*, **33**(1), 2017, pp. 203–215.
21. M. Minović, M. Milovanović, U. Šošević and M. Á. C. González, Visualisation of student learning model in serious games, *Comput. Human Behav.*, **47**, 2015, pp. 98–107.
22. Á. Fidalgo-Blanco, M. L. Sein-Echaluce, F. J. García-Peñalvo and M. Á. Conde, Using Learning Analytics to improve teamwork assessment, *Comput. Human Behav.*, **47**, 2015, pp. 149–156.
23. Á. F. Agudo-Peregrina, S. Iglesias-Pradas, M. Á. Conde-González and Á. Hernández-García, Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported {F2F} and online learning, *Comput. Human Behav.*, **31**, 2014, pp. 542–550.
24. H. Alaiz Moretón, J. L. Calvo Rolle, I. García and A. Alonso Alvarez, Formalization and practical implementation of a conceptual model for PID controller tuning, *Asian J. Control*, **13**(6), 2011, pp. 773–784.
25. J. L. Casteleiro-Roca, J. L. Calvo-Rolle, M. C. Meizoso-López, A. J. Piñón-Pazos and B. A. Rodríguez-Gómez, Bio-inspired model of ground temperature behavior on the horizontal geothermal exchanger of an installation based on a heat pump, *Neurocomputing*, **150**, 2015, pp. 90–98.
26. J. L. Calvo-Rolle, Ó. Fontenla-Romero, B. Pérez-Sánchez and B. Guijarro-Berdinas, Adaptive Inverse Control Using an Online Learning Algorithm for Neural Networks, *Informatica*, **25**(3), 2014, pp. 401–414.
27. J. L. Calvo-Rolle, J. L. Casteleiro-Roca, H. Quintián and M. C. Meizoso-López, A hybrid intelligent system for PID controller using in a steel rolling process, *Expert Syst. Appl.*, **40**(13), 2013, pp. 5188–5196.
28. I. Machon-Gonzalez, H. Lopez-Garcia and J. L. Calvo-Rolle, A hybrid batch SOM-NG algorithm, in *Neural Networks (IJCNN), The 2010 International Joint Conference on*, 2010, pp. 1–5.
29. J. L. Casteleiro-Roca, J. A. M. Pérez, J. A. Reboso-Morales, F. J. de Cos Juez, F. J. Pérez-Castelo and J. L. Calvo-Rolle, An Intelligent Model for Bispectral Index (BIS) in Patients Undergoing General Anesthesia, in *International Conference on EUropean Transnational Education*, 2016, pp. 290–300.
30. J.-L. Casteleiro-Roca, J. L. Calvo-Rolle, J. A. Méndez Pérez, N. Roqueñí Gutiérrez and F. J. de Cos Juez, Hybrid Intelligent System to Perform Fault Detection on BIS Sensor During Surgeries, *Sensors*, **17**(1), 2017.
31. C. C. Turrado, M. del C. M. López, F. S. Lasheras, B. A. R. Gómez, J. L. C. Rollé and F. J. de C. Juez, Missing data imputation of solar radiation data under different atmospheric conditions, *Sensors*, **14**(11), 2014, pp. 20382–20399.
32. A. K. Qin and P. N. Suganthan, Enhanced neural gas network for prototype-based clustering, *Pattern Recogn.*, **38**(8), 2005, pp. 1275–1288.
33. J. Ye and T. Xiong, SVM versus least squares SVM, *J. Mach. Learn. Res.—Proc. Track*, **2**, 2007, pp. 644–651.
34. S. Kaski, J. Sinkkonen, and A. Klami, Discriminative clustering, *Neurocomputing*, **69**(1–3), 2005, pp. 18–41.
35. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
36. R. M. Heiberger and E. Neuwirth, Polynomial regression, in *R Through Excel*, Springer New York, 2009, pp. 269–284.
37. J. L. Casteleiro-Roca, J. L. Calvo-Rolle, M. C. Meizoso-López, A. P. Piñón-Pazos and B. A. Rodríguez-Gómez, New approach for the QCM sensors characterization, *Sensors Actuators A Phys.*, **207**, 2014, pp. 1–9.
38. X. Wu, *Optimal designs for segmented polynomial regression models and web-based implementation of optimal design software*. Stony Brook, NY, USA: State University of New York at Stony Brook, 2007.
39. Z. G. Zhang and S. C. Chan, On kernel selection of multivariate local polynomial modelling and its application to image smoothing and reconstruction, *J. Signal Process. Syst.*, **64**(3), 2011, pp. 361–374.
40. P. D. Wasserman, *Advanced methods in neural computing*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1993.
41. Z. Zeng and J. Wang, *Advances in neural network research and applications*, 1st ed. Springer Publishing Company, Incorporated, 2010.
42. J. Rynkiewicz, General bound of overfitting for MLP regression models, *Neurocomputing*, **90**, 2012, pp. 106–110.
43. J.-L. Casteleiro-Roca, E. Jove, F. Sánchez-Lasheras, J.-A. Méndez-Pérez, J.-L. Calvo-Rolle and F. J. De Cos Juez, Power Cell SOC Modelling for Intelligent Virtual Sensor Implementation, *J. Sensors*, vol. 2017, 2017.
44. J. L. Casteleiro-Roca, H. Quintián, J. L. Calvo-Rolle, E. Corchado, M. del Carmen Meizoso-López, and A. Piñón-Pazos, An intelligent fault detection system for a heat pump installation based on a geothermal heat exchanger, *J. Appl. Log.*, **17**, 2016, pp. 36–47.
45. J. L. Calvo-Rolle, I. Machón-González and H. López-

García, Neuro-Robust controller for non-linear systems, *DYNA*, **86**(3), 2011, pp. 308–317.
46. N. Cristianini and J. Shawe-Taylor, *An introduction to Support Vector Machines and other kernel-based learning methods*, New York, NY, USA: Cambridge University Press, 2000.
47. I. Steinwart and A. Christmann, *Support Vector Machines*, 1st ed. Springer Publishing Company, Incorporated, 2008.
48. V. Vapnik, *The nature of statistical learning theory*, Springer, 1995.
49. J. A. K. Suykens and J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.*, **9**(3), 1999, pp. 293–300.
50. R. Wang, A. Wang and Q. Song, Research on the alkalinity of sintering process based on LS-SVM Algorithms, in *Advances in Computer Science and Information Engineering*, **168**, D. Jin and S. Lin, Eds. Springer Berlin Heidelberg, 2012, pp. 449–454.
51. Y. Guo, X. Li, G. Bai and J. Ma, Time Series Prediction Method Based on LS-SVR with Modified Gaussian RBF, in *Neural Information Processing*, **7664**, T. Huang, Z. Zeng, C. Li and C. Leung, Eds. Springer Berlin Heidelberg, 2012, pp. 9–17.
52. L. Wang and J. Wu, Neural network ensemble model using PPR and LS-SVR for stock market eorecasting, in *Advanced Intelligent Computing*, **6838**, D.-S. Huang, Y. Gan, V. Bevilacqua and J. Figueroa, Eds. Springer Berlin Heidelberg, 2012, pp. 1–8.

Esteban Jove: Esteban Jove received the Master‘s Degree in Industrial Engineering from the University of Leon, Leon, Spain, in 2014. After two years working in the automotive industry, he joined the University of A Coruna, Spain, where he is Professor of Power Electronics in the Faculty of Engineering since 2016. He is he is a PhD student in the University of La Laguna and his research has been focused on using intelligent techniques to model nonlinear systems.

José Antonio López-Vázquez received the BS from University of Oviedo in 1989, the MS in Industrial Design and Product Development from the University of A Coruna in 2000, the MS in Integrated Project Management in 2010 and he received PhD degree from the University of A Coruna in 2015. Since 1990 to 1996 he was employed in several companies of the industrial sector, working in Industrial Engineering Project Management. In 1996 he joined the University of A Coruna, Spain, where he is Professor of Industrial Engineering Department and he is secretary of the Polytechnic University School. His main research interests have been centered in applying advanced regression techniques to study and predict the individual academic performance of engineering students.

María Isabel Fernández-Ibáñez received the MS in Architecture from the University of Santiago de Compostela in 1988 and she obtains the PhD degree in Architecture from the UDC in 2010. After several years working as arquitect, in 1993 she joined the University of A Coruna, Spain, where she is Professor in the Engineering Graphics area. Since 2010, she combines educational and research tasks with management, like deputy director of the Polytechnic University College. She belongs to the investigation group “Science and Technical Cybernetics”.

José-Luis Casteleiro-Roca received the BS from University of A Coruna in 2003, the MS in Industrial Engineering from the University of Leon in 2012, and now he is a PhD student in the University of La Laguna. He has been Technical Engineer in the Spanish Navy since 2004 and, since 2014, he is also part of the teaching and research staff of the UDC as a part-time associate professor. His main research interests are centered in applying expert system technologies to the diagnosis and control systems and in intelligent systems for control engineering and optimization.

José Luis Calvo-Rolle received the MS and PhD degrees in Industrial Engineering from the University of Leon, Leon, Spain, in 2004, and 2007, respectively. He is an Associate Professor of Automatic Control and the head of Industrial Engineering Department, Faculty of Engineering, University of A Coruna, Spain. His main research interests have been centered in applying expert system technology to the diagnosis and control systems and in intelligent training systems for control engineering, optimization and education.