Challenges and Advances in the knowledge of University Dropout and Academic Performance

(Desafíos y Avances en el conocimiento del Abandono y Rendimiento Universitario)

Adolfo Hernández

WEBINAR – ABANREDES

"PREDICTORS OF DROPOUT AND ACADEMIC PERFORMANCE THROUGH PATTERN RECOGNITION AND AUTONOMIC DIAGNOSIS"

FACULTAD DE COMERCIO Y TURISMO. 24 DE ABRIL DE 2024

Determinantes del rendimiento académico y abandono en la educación superior: análisis del impacto del uso de plataformas educativas y redes sociales

PID202020-116293RB-I00

Determinantes del rendimiento académico y abandono en la educación superior: análisis del impacto del uso de plataformas educativas y redes sociales

Determinants of academic performance and dropout in higher education: analysis of the impact of the use of educational platforms and social networks (ABANREDES)

R+D project, financed by the Ministry of Education and Science, 2020 call, 2021-2024

Presentation

A multidisciplinary research project, the result of the cooperation of specialists in various areas (Mathematics, Statistics, Economics and Business, Sociology, Education, etc.), whose main objective is to improve the quality and equity in higher education at through the implementation of educational interventions that reduce both university dropout and low academic performance.

Team

Principal Researchers (Universidad Complutense de Madrid - UCM):

Adolfo Hernández Estrada (Financial & Actuarial Economics & Statistics Department)

María Fernández Mellizo-Soto (Applied Sociology, Departmental Section at Faculty of Education)

Research Team: 7 from UCM, 2 from Comillas University, 1 from Valladolid University and 1 from Universidad Nacional de Concepción (Paraguay). 1 Research assistant.

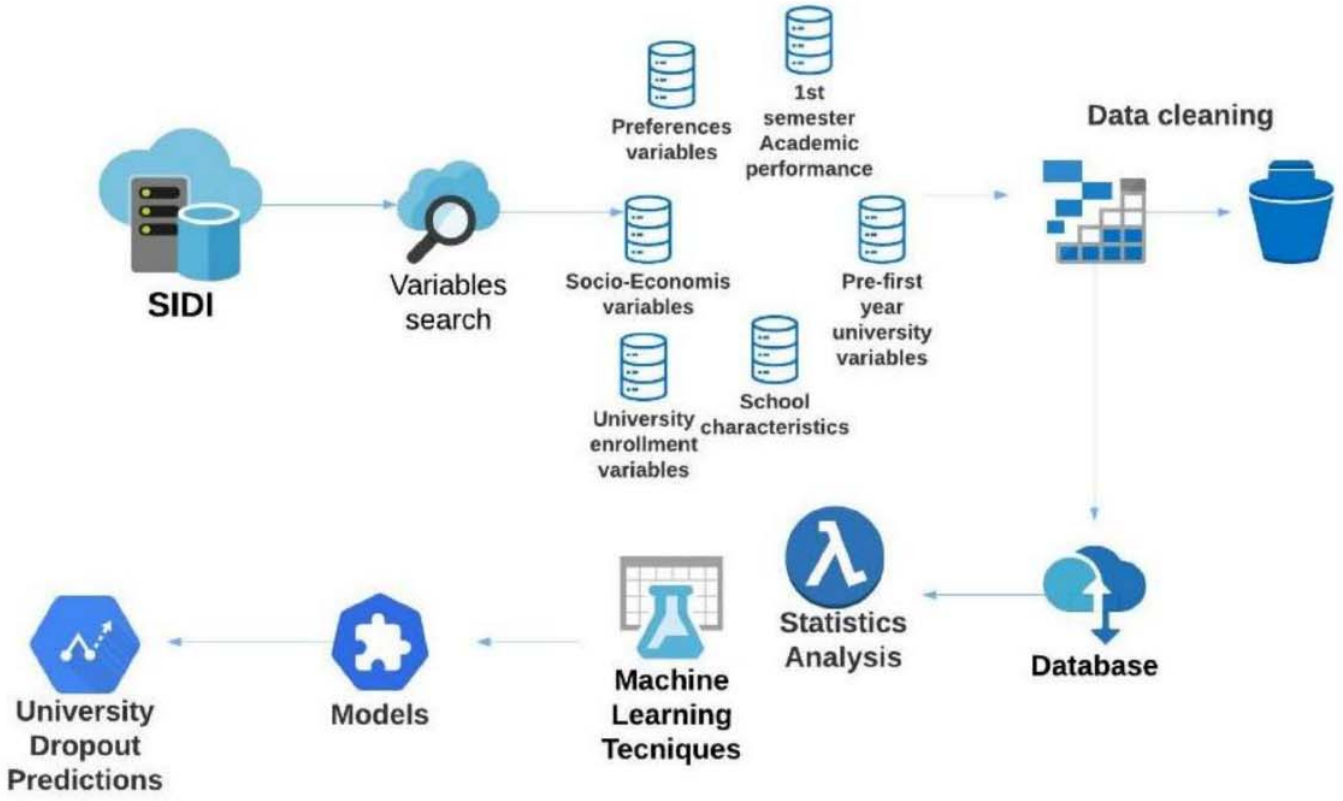
Team

Objectives and Methodology

Main goals: achieving the implementation of educational interventions to reduce both university dropout and low academic performance, with the purpose of improving the quality and equity in higher education

Methodology: models ("**Machine Learning**") for the prediction of academic performance and dropout are built, using ("**variables**") personal, social, academic and technological factors (digital interactions on educational platforms and social media).

Data

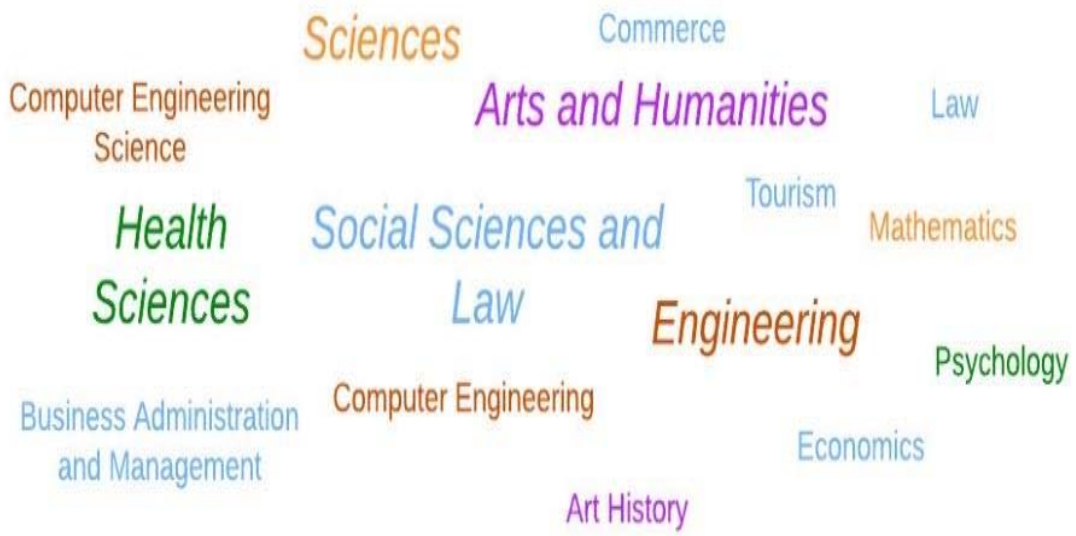


Institutional Intelligence Center of the UCM http://www.ucm.es/cii

Determinantes del rendimiento académico y abandono en la educación superior: análisis del impacto del uso de plataformas educativas y redes sociales
PID202020-116293RB-I00

CENTRO DE INTELIGENCIA INSTITUCIONAL

Data



Determinantes del rendimiento académico y abandono en la educación superior: análisis del impacto del uso de plataformas educativas y redes sociales

PID202020-116293RB-I00

CENTRO DE INTELIGENCIA INSTITUCIONAL

Some results ("work in progress") 1

Segura, M., Mello-Román, J.D., Hernández, A. (2022) Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role?. *Mathematics* 2022, 10(18), 3359, Special Issue Advances in Artificial Intelligence and Statistical Techniques with Applications to Health and Education. https://www.mdpi.com/2227-7390/10/18/3359.

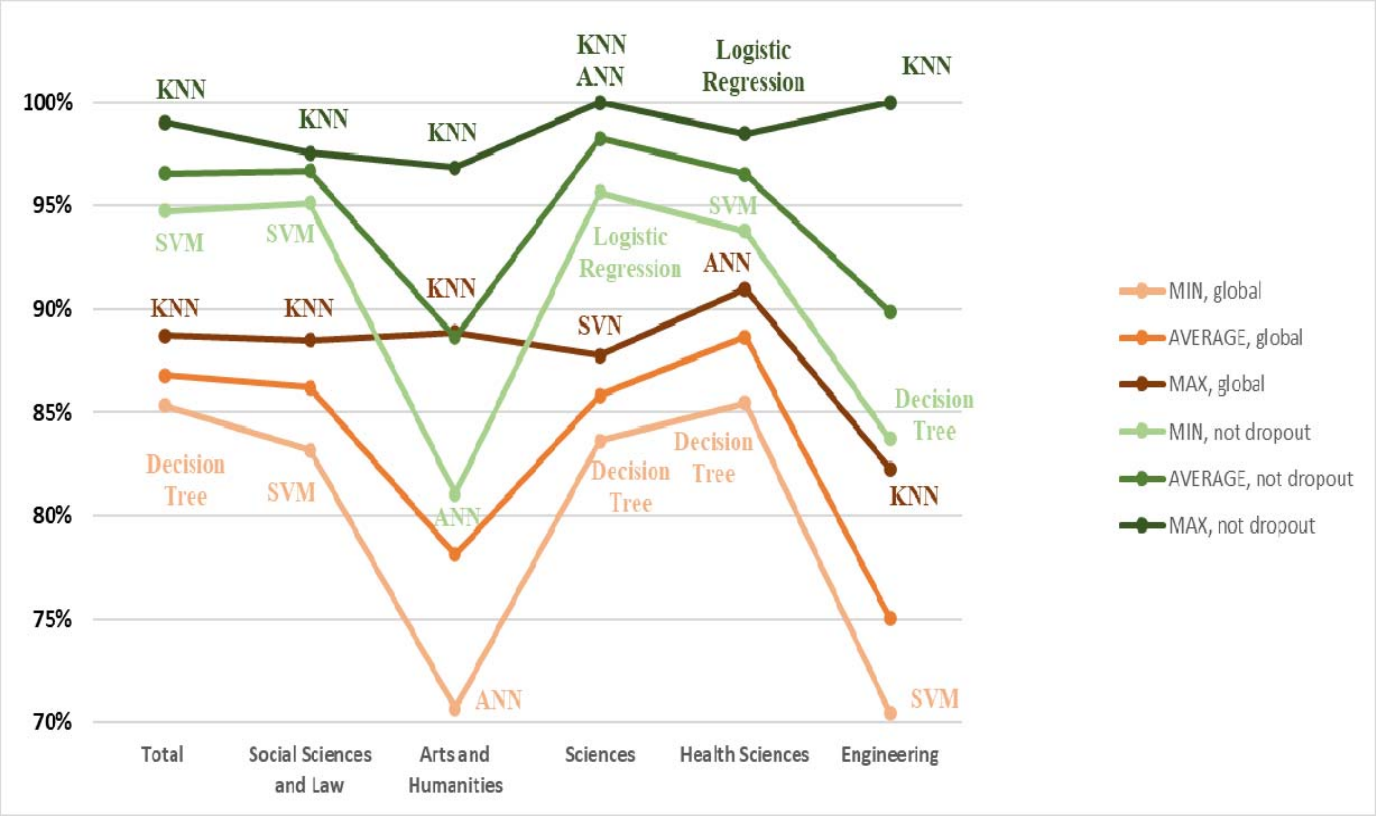
Feature Selection Process in order to identify the variables more correlated with dropout; then, some Machine Learning Models (Support Vector Machines, Decision Trees and Artificial Neural Networks) as well as a Logistic Regression

Some results (“work in progress”) 1

Dropout detection does not work only with enrollment variables, but it improves after the first semester results. Academic performance is always a relevant variable, but there are others, such as the level of preference that the student had over the course that he or she was finally able to study.

The success of the techniques depends on the program areas. Machine Learning obtains the best results, but a simple Logistic Regression model can be used as a reasonable baseline.

Some results (“work in progress”) 1



Some results (“work in progress”) 2

Ana María Sánchez-Sánchez, Jorge Daniel Mello-Román, Marina Segura and Adolfo Hernández (2024). Identifying Determinants of Academic Success: A Machine Learning Approach in Higher Education. *Under revision*

Machine learning techniques to identify determinants of academic performance in first-year university students. 8,700 records from UCM corresponding to all incoming students in the academic year 2022-23, information was available on 28 variables related to university access, academic performance corresponding to the first year and socio-economic characteristics.

Some results ("work in progress") 2

Feature selection using Random Forest and Extreme Gradient Boosting (XGBoost) to identify the main predictors of academic performance and avoid overfitting in the models, followed by analysis with four different machine learning techniques: Linear Regression, Support Vector Regression, Random Forest, and XGBoost.

Influence of variables that had not appeared in the literature before: admission option and number of enrolled credits.

Some results ("work in progress") 2

Table S6. Importance of variables for RF and XGBoost

Importance Order	Random Forest[1]		XGBoost[1]	
	Variables	**%IncMSE**	**Variables**	**Gain**
1	Access grade	114.36	Access grade	0.4856
2	Academic amount	50.32	Academic amount	0.0816
3	No. of ECTS enrolled 1st semester	48.89	No. of ECTS enrolled 1st semester.	0.0747
4	No. of ECTS enrolled 1st year	39.19	No. of ECTS enrolled 1st year	0.0605
5	Scholarship holder	32.98	Age	0.0511
6	Gender	32.15	Admission option	0.0502
7	Admission option	31.13	Administrative fee	0.0410
8	Family township	28.27	Gender	0.0397
9	Location of the school	27.72	Father's or guardian's level of studies	0.0220
10	Admission study	24.75	Family township	0.0217

[1] Dependent Variable = First-semester grade.

Some results ("work in progress") 3

Arroyo-Barrigüete, J. L., Carabias-López, S., Hernández, A. Segura, M. (2023). Effect of advanced high school major on mathematical performance at university: a comparative study in Business Administration degrees. *Revista de Educación* 402, 109-132.
https://recyt.fecyt.es/index.php/Redu/article/view/96806

The major taken at high school is a very relevant variable in predicting average marks during the first year of business administration (BA) degrees. Comparison of the results at two different universities: 873 students at UCM and 822 at the Universidad Pontificia Comillas.

Some results ("work in progress") 3

Use of regression models combined with interpretable neural networks to ensure the robustness of the results

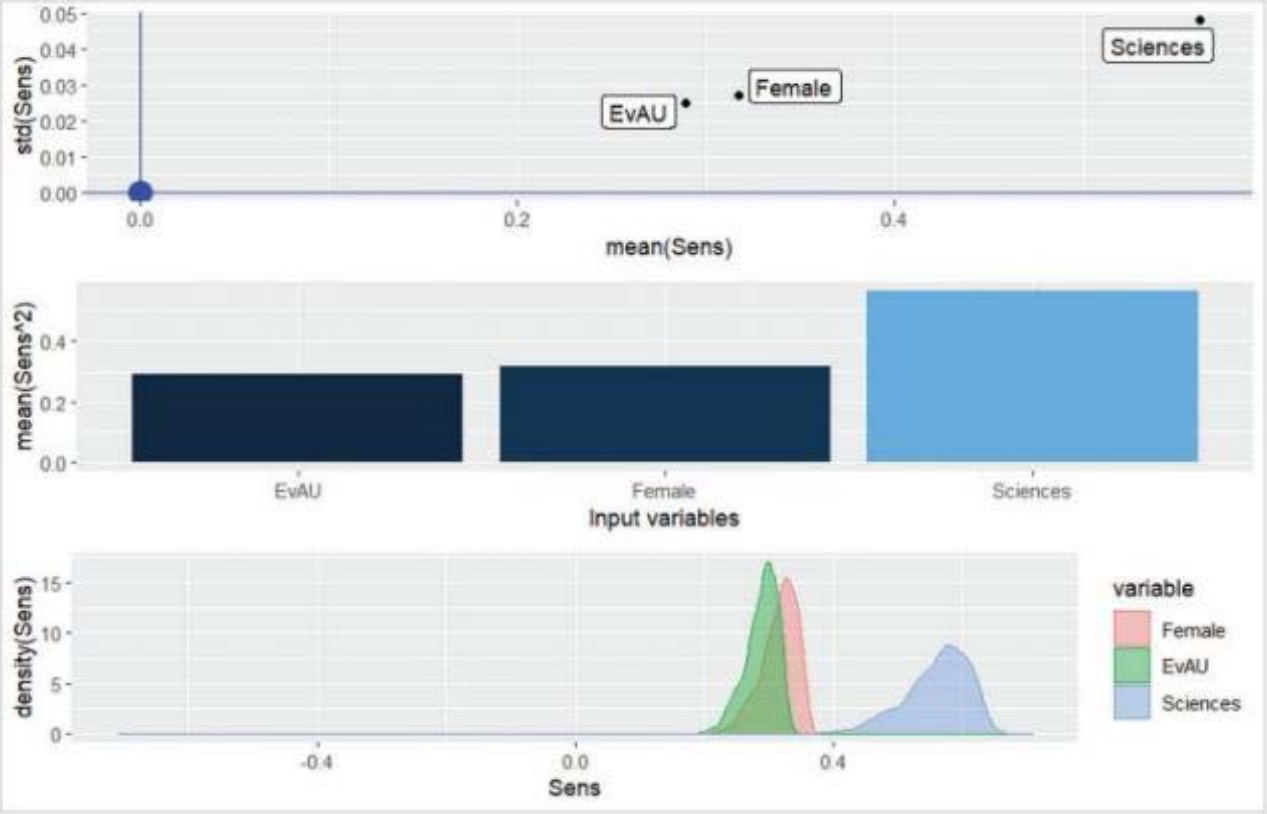
Results are virtually identical: students from the science major outperform their peers from the social sciences major in business mathematics. Two implications:

-) need to reflect on the focus and content of mathematics in the social sciences major, seeking greater alignment with the requirements of the degrees linked to this major;

-) regarding teaching practice, need to rethink teaching strategies in mathematics, bearing in mind the characteristics and learning styles of students from the social sciences major.

Some results ("work in progress") 3

FIGURE I. Neural network results, including the metrics proposed by Pizarroso *et al.* (2022), for Universidad Complutense de Madrid (average mark in Business Mathematics I and II)



Future

STEM subjects in non-STEM university degrees: gender gap, academic performance, and impact on dropout rates (RENDISTEM)

Proposal submitted: R+D project, Ministry of Education and Science, 2024 call

Determinantes del rendimiento académico y abandono en la educación superior: análisis del impacto del uso de plataformas educativas y redes sociales

PID202020-116293RB-I00

Information and contact

Web page:
https://www.ucm.es/abanredes-en/

YouTube:
https://www.youtube.com/@ABANREDES2020/videos

Blog: https://proyectoabanredes.blogspot.com/

Linkedin:
https://www.linkedin.com/company/proyecto-abanredes/

Email: abanredes@ucm.es