

# Evaluación económica de intervenciones educativas en la LOMLOE\*: propuestas de mejora con inteligencia artificial

## *An economic evaluation of educational interventions in the LOMLOE: Proposals for improvement with Artificial Intelligence*

**Dra. María Teresa BALLESTAR.** Profesora Asociada. Universidad Rey Juan Carlos ([teresa.ballestar@urjc.es](mailto:teresa.ballestar@urjc.es)).

**Dr. Jorge SAINZ.** Catedrático. Universidad Rey Juan Carlos ([jorge.sainz@urjc.es](mailto:jorge.sainz@urjc.es)).

**Dr. Ismael SANZ.** Profesor Titular. Universidad Rey Juan Carlos ([ismael.sanz@urjc.es](mailto:ismael.sanz@urjc.es)).

### Resumen:

El objetivo de esta investigación es demostrar la necesidad de evaluar económicamente la LOMLOE, especialmente tras la inversión de los fondos *EU Next Generation* que abren nuevas oportunidades de las que carecía la ley en su redacción inicial. Las Administraciones públicas tienen el reto de emplear esa inversión adicional de forma eficiente.

Nuestro análisis demuestra que los modelos de inteligencia artificial pueden predecir si los programas de apoyo educativo ayudarán a incrementar la probabilidad de que estudiantes rezagados superen 4.º de la ESO (Educación Secundaria Obligatoria). De esta forma, se puede calcular el retorno social de los pro-

gramas de apoyo educativo y contribuir a su diseño *ex-ante* para lograr que los alumnos tengan mayores tasas de éxito.

Para completar los modelos ya utilizados por Administraciones públicas, empleamos modelos de Machine Learning (ML) robustos como árboles de decisión CHAID y redes neuronales artificiales para analizar las características de los grupos de estudiantes y la intervención en la que han formado parte. Las conclusiones permiten mejorar los programas de refuerzo educativo de los próximos años para apoyar a los alumnos con menos posibilidades de éxito académico.

**Descriptor:** análisis de políticas públicas, *Machine Learning*, eficiencia educativa, LOMLOE.

\* Ley Orgánica 3/2020, de 29 de diciembre de 2020, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación. Fecha de recepción de la versión definitiva de este artículo: 13-12-2021.

Cómo citar este artículo: Ballestar, M. T., Sainz, J. y Sanz, I. (2022). Evaluación económica de intervenciones educativas en la LOMLOE: propuestas de mejora con inteligencia artificial | *An economic evaluation of educational interventions in the LOMLOE: Proposals for improvement with Artificial Intelligence*. *Revista Española de Pedagogía*, 80 (281), 133-154. <https://doi.org/10.22550/REP80-1-2022-09>  
<https://revistadepedagogia.org/>

## Abstract:

This research aims to demonstrate the need for an economic evaluation of the Organic Law 3/2020, of 29 December, which amends Organic Law 2/2006, of 3 May, on Education (LOMLOE), especially after the investment of EU Next Generation funds that open new opportunities that were lacking in the initial drafting of the law. The challenge for Public Administrations is to use this additional investment efficiently.

Our analysis shows that artificial intelligence models can predict whether educational support programmes will help increase the likelihood that students who lag behind will pass the 4th grade of ESO. In this way, we can calculate the social return of these

programmes and contribute to their *ex-ante* design to achieve higher success rates for students.

To complement the models already used by public administrations, we use robust Machine Learning (ML) models such as CHAID decision trees and artificial neural networks to analyse the characteristics of the groups of students and the intervention they have been part of. The conclusions allow us to improve educational reinforcement programmes in the coming years to support students with lower chances of academic success.

**Keywords:** Public policy analysis, Machine Learning, educational efficiency, LOMLOE.

## 1. Introducción

Uno de los principios en la gestión educativa por parte de organismos internacionales como la OCDE (Organización para la Cooperación y el Desarrollo Económicos) es la evaluación de políticas educativas tanto globales como de intervenciones concretas (Golden, 2020). El resultado del análisis lleva a una mayor eficacia e impacto, tanto a nivel social como individual, de las actuaciones educativas, lo que revierte en el alumnado y en una dinámica positiva para el sistema en general (OECD, 2018). Este principio informativo aparece en casi todas las legislaciones europeas y, en el caso español, se encuentra recogido en Ley Orgánica 2/2006, de 3 de mayo (LOE) que señala en su Artículo 2 bis que el «El funcionamiento del Sistema Educativo Espa-

ñol se rige por los principios de [...] eficiencia en la asignación de recursos públicos, transparencia y rendición de cuentas».

Nuestro objetivo es demostrar que la evaluación económica de las políticas educativas sirve para optimizar y priorizar los objetivos legislativos establecidos por la LOMLOE. Este hecho demostraría que la ley se enfrenta a una laguna en su construcción que afecta a su evaluación de la gestión, evaluación que si es un requerimiento para la justificación de los fondos *EU Next Generation* para educación que solo en 2021 representan en los presupuestos generales del estado 1852.5 millones de euros.

Va a ser esta inversión europea la que permitirá implantar y evaluar la ley, ya

que en su Memoria de Análisis e Impacto Normativo (MAIN), que recoge los aspectos técnicos de la norma, se señala que «[...] el proyecto no supone incremento o disminución del gasto público». Como contrapartida la Comisión exige que se evalúe si la población adquiere efectivamente las competencias para competir a nivel global (Crescenzi et al., 2021; Porte y Jensen, 2021).

La Comisión Europea y UNESCO reconocen la importancia de este tipo de actuaciones tanto en la elección de reformas como en su posterior análisis, y la necesidad de que haya datos disponibles sobre intervenciones educativas para su posterior análisis, pese a las reticencias de algunas administraciones (European Commission, 2010; Yusuf, 2007). Aunque el planteamiento preferido para este análisis es el enfoque experimental, en la realidad educativa existen problemas para llevarlo a cabo por diversos motivos: como el coste económico, la dificultad de establecer grupos de control aleatorios, los problemas de gestión, ansiedad, etc. (Golden, 2020; Slavin, 2016).

Para solventar esta propuesta se están utilizando un arsenal de nuevas técnicas que están en pleno desarrollo y que pueden servir como propuesta, entre las que se encuentra la inteligencia artificial (IA) (Ballestar et al., 2019; Chassignol et al., 2018; Chatterjee y Bhattacharjee, 2020). Siguiendo la estrategia de triangulación de Ballestar et al. (2020), evaluamos el uso de distintas estrategias de IA para la evaluación del retorno de intervenciones educativas. Para ello, profundizaremos en la intervención realizada por la Consejería de Educación de la Junta de Comunidades

de Castilla y León denominada *Programa para la Mejora del Éxito Educativo*, que analizaremos en su edición de 2020, tras el cierre de los colegios por la pandemia. La Junta realizó una convocatoria de ayudas económicas para que los centros educativos pudieran ofrecer aulas de apoyo durante el mes de julio a los alumnos de sexto de primaria, cuarto de ESO y segundo de Bachillerato en las áreas de lengua castellana y literatura, matemáticas o inglés con dificultades educativas. El programa ha contribuido a lo largo de los años a que un número importante de alumnos de la comunidad promocionen de curso. De forma general para todos los cursos que disfrutaban del programa, se observó en el curso académico 2019-20 una mejora de la promoción del 5 % en los estudiantes que participaron en el programa, en comparación con los que estando en condiciones de participar no lo hicieron.

## 2. La intervención

Para diseñar cualquier intervención educativa Slavin (2016) y Golden (2020) reconocen como fundamental que esté basada en evaluaciones científicas. Probablemente, las clases de apoyo que componen el *Programa para la Mejora del Éxito Educativo* en Castilla y León sea una de las más fundamentadas científicamente en los últimos años. Las investigaciones realizadas por la *Education Endowment Foundation*<sup>1</sup> demuestran que las tutorías en pequeños grupos se encuentran entre las medidas para las que existe evidencia empírica de mayor efectividad a nivel internacional. Como tal han demostrado ser un buen complemento para un sistema educativo que está diseña-

do para hacer pasar a un gran número de alumnos de curso a curso, pero que no funciona para todos los estudiantes.

Kraft (2015) y Burgess (2020) analizan los factores que contribuyen a su éxito: la participación por personal seleccionado y formado, coordinado con sus docentes habituales y en grupos reducidos. El impacto es relevante: Nickow et al., (2020) muestran, a partir del estudio de 96 artículos de experimentos aleatorios, que el efecto de las tutorías en pequeños grupos es importante y significativo (37 % de la desviación estándar).

Este tipo de intervenciones es una de las pocas que vienen recogidas en nuestra legislación. La LOMLOE en su art. 4.4 señala que uno de sus objetivos es facilitar «[...] el acceso a los apoyos que el alumnado requiera» y el incremento de estas medidas a lo largo de todo el proceso formativo de forma individualizada (art. 20 bis) «[...] para evitar la repetición escolar, particularmente en entornos socialmente desfavorecidos». De hecho, la norma deja una redacción muy clara sobre sus objetivos a la repetición, por ejemplo, en la nueva redacción del art. 28: «La permanencia en el mismo curso [de secundaria] se considerará una medida de carácter excepcional y se tomará tras haber agotado las medidas ordinarias de refuerzo y apoyo para solventar las dificultades de aprendizaje».

Estas medidas de apoyo están presentes en la legislación educativa a lo largo de este siglo. Así la *non nata* Ley Orgánica 10/2002, de 23 de diciembre, de Calidad de la Educación (LOCE) recogía en su art. 2 el derecho de los estudiantes a «[...] recibir las ayudas

y los apoyos precisos para compensar las carencias y desventajas de tipo personal, familiar, económico, social y cultural, especialmente en el caso de presentar necesidades educativas especiales, que impidan o dificulten el acceso y la permanencia en el sistema educativo». Las distintas normas desde entonces han recogido preceptos similares de protección en línea con las recomendaciones de los organismos especializados (Gouédard et al., 2020; Pont y Montt, 2014; Schleicher, 2020) como la literatura sobre el retorno educativo (Brunello y Paola, 2014; de la Fuente y Jimeno, 2009; Doncel et al., 2014) que establecen la importancia de la adquisición de competencias, hecho que pasa a ser fundamental con la irrupción de las nuevas tecnologías (Ballestar et al., 2020, 2022; Goos et al., 2009; Gregory et al., 2019).

De acuerdo con el art. 9 de la LOMLOE corresponde a las comunidades autónomas (CCAA) llevar a cabo las actuaciones dirigidas a la reducción y prevención del fracaso escolar y del abandono escolar temprano a través de los programas de cooperación territorial. La Memoria de Análisis e Impacto Normativo de la Ley dotó en la redacción original de la ley, previa a la pandemia, con 45 millones de euros por año para este tipo de programas durante el periodo 2020-2023.

Las CCAA ya venían destinando una importante inversión a este tipo de proyectos. La Consejería de Educación de Castilla y León venía desarrollando desde el curso 2007-2008, entre otros, el *Programa para la Mejora del Éxito Educativo*. Sus objetivos son, en línea con la LOE, LOMCE (Ley Orgánica para la Mejora de la Calidad

Educativa, 2013) y LOMLOE «[...] contribuir a la mejora del modelo educativo en la Comunidad de Castilla y León y facilitar el éxito y la continuidad del alumnado en el sistema educativo» garantizando «la progresión didáctica del alumnado y su promoción en el sistema educativo, reduciendo el abandono escolar temprano, fomentando su eficaz inserción en el mercado laboral y optimizando al mismo tiempo el clima de convivencia en los centros docentes».

Son los centros los que solicitan estas medidas para favorecer el éxito educativo, la integración del alumnado y la involucreción de las familias para aumentar la tasa de promoción y titulación del alumnado, con especial atención al alumnado más vulnerable de 4.º de ESO, 6.º curso de Educación Primaria y 2.º de Bachillerato de centros sostenidos con fondos públicos. El programa se imparte en 93 centros en lengua, matemáticas e inglés de cara a reforzar el fin de ciclo y con la participación de estudiantes que asisten voluntariamente, autorizados y apoyados por su familia. Los requisitos de acceso son presentar una necesidad específica de apoyo educativo o estar en situación de vulnerabilidad socioeducativa y es impartida por docentes especializados.

La primera evaluación realizada por la junta es prometedora: para el curso 2019-2020 los alumnos que participan en el programa aumentan su probabilidad de aprobar lengua en un 24.6 %, inglés en un 13.8 % y matemáticas en un 8.4 % frente a los no inscritos, generando un alto nivel de satisfacción tanto en docentes (85.7 %) como familias (88.2 %).

Este primer análisis nos sirve como base para realizar nuestra investigación sobre la eficiencia económica de las intervenciones recogidas en la LOMLOE. Para ello, planteamos clasificar en grupos a los estudiantes que participan en el *Programa para la Mejora del Éxito Educativo* en Castilla y León en 4.º de la ESO y predecir la probabilidad de superar dicho curso escolar para cada uno de los grupos de estudiantes. Esto permitirá realizar pronósticos individuales para estudiantes en base a sus características, detalles del programa especial de apoyo realizado y su desempeño en el programa. Esto permite estimar la probabilidad de éxito del programa y calcular su retorno social (ROI) a través de un modelo de triangulación que emplea diferentes metodologías de ML como árboles de decisión CHAID y redes neuronales artificiales para evaluar políticas educativas.

### 3. Análisis empírico

La información empleada para el análisis contiene los datos disponibles de los alumnos de 4.º de la ESO participantes en el *Programa para la Mejora del Éxito Educativo* en el curso 2019-2020. Se dispone de 1739 registros que corresponden con los estudiantes que participaron en alguno de los tres programas implementados en la comunidad autónoma. De ellos, el 47.27 % son niñas y el 52.73 % son niños. De estos estudiantes, el 47.15 % (820 estudiantes) han repetido ya algún curso.

Los estudiantes pueden participar en tres tipos de programas. El programa C2 consiste en un acompañamiento del estudiante durante todo el año académico

y acumula el 76.37 % de los estudiantes (1328). El programa C3 consiste en la asistencia de clases extraordinarias durante verano en el mes de julio e incluye al 17.02 % de los estudiantes (296). Finalmente, el programa C2C3, una combinación de los dos anteriores, representa el 6.61 % de los estudiantes (115). Dichas actuaciones se imparten en las nueve provincias de Castilla y León, concentrando el 66.36 % de los estudiantes en capitales de provincia y el 33.64 % en otras localidades, con una ratio de éxito en la finalización de los estudios de 4.º de la ESO del 85.34 % para aquellos que terminaron el programa.

Una de las innovaciones de nuestra investigación es la aplicación de una metodología de triangulación que consiste en el desarrollo de más de un método cuantitativo, aplicando diferentes enfoques, con el objetivo de enriquecer los resultados, así como, confirmar por partida doble los resultados obtenidos (Ballestar et al., 2020). En la primera fase, se desarrolla un modelo de Machine Learning (ML) basado en árboles de decisión CHAID (Chi-square Automatic Interaction Detector) que determine cuáles son las variables relevantes a la hora de diseñar y predecir la tasa de éxito esperada que tendrán los participantes en el programa. El objetivo es identificar los factores a tener en cuenta a la hora de diseñar una intervención de estas características, así como calcular su tasa de éxito y, en consecuencia, el retorno social de la inversión realizada.

En la segunda fase, el primer modelo se evalúa mediante la realización de un

modelo completamente nuevo que utiliza una metodología ML basada en redes neuronales artificiales perceptrón multicapa con programación hacia atrás (ANN-MPL) que debe validar los resultados obtenidos anteriormente.

### 3.1. Definición de las variables de los modelos de ML

Los árboles de decisión CHAID y las redes neuronales artificiales perceptrón multicapa (ANN-MLP) son métodos de aprendizaje automático supervisados. Este tipo de modelos describen y explican las relaciones subyacentes existentes entre las diferentes variables de entrada con la finalidad de predecir el valor de la variable objetivo, a través de un proceso de entrenamiento usando una muestra de datos que contiene tanto los valores para las variables de entrada como de salida (Maimon y Rokach, 2005). Como métodos supervisados, la precisión de sus clasificaciones y predicciones estará condicionada a la calidad de la muestra disponible para el entrenamiento, validación y test de los modelos (Aad et al., 2012; Ballestar et al., 2018; Li y Eastman, 2006).

Se ha realizado un proceso de minería de datos sobre la base de origen con el objetivo de crear una única tabla para el desarrollo de los modelos de ML. Esta tabla contiene 1739 registros de estudiantes y 21 variables que capturan la información para la caracterización del estudiante, así como su desempeño y resultados tanto en el programa que ha participado como en el curso de 4.º de la ESO.

Mediante la realización de análisis descriptivos y causales iterativos, 5 de estas variables fueron identificadas como estadísticamente significativas y relevantes para el análisis empírico y desarrollo de los dos modelos de ML. De estas variables, 4 actúan como variables de entrada de información en los modelos y 1 como variable de salida u objetivo a determinar o predecir (Tabla 1). Por lo tanto, ambos modelos

utilizan las mismas 5 variables y son entrenados con la misma base de datos.

La variable objetivo o de salida, el porcentaje de estudiantes que no ha superado el curso de 4.º de la ESO (0 en la variable *finish\_studies*), es 14.66 %, mientras que los que sí superan el curso que representan el 85.34 % (valor 1 de la variable *finish\_studies*), dejando una muestra desbalanceada.

Tabla 1. Descripción de las variables de los modelos de ML: árbol de decisión CHAID y red neuronal artificial perceptrón multicapa (ANN-MLP).

| Variables de entrada      | Descripción  |
|---------------------------|--|
| <i>years_repetition</i>   | Variable numérica discreta. Número de cursos que ha repetido el estudiante. En el caso de que el estudiante no haya repetido ningún curso el valor será 0. |
| <i>student_repetition</i> | Variable booleana. Valor 1 si el estudiante ha repetido algún curso; Valor 0 si no ha repetido ningún curso.   |
| <i>kind_program</i>       | Variable categórica que indica cuál de los tres programas académicos ha realizado el estudiante: C2; C3; C2C3.   |
| <i>finish_program</i>     | Variable booleana. Valor 1 si el estudiante ha finalizado el programa académico de apoyo; Valor 0 si no lo ha llegado a finalizar.                         |
| Variable de salida        | Descripción  |
| <i>finish_studies</i>     | Variable booleana. Valor 1 si el estudiante ha superado el curso 4º de la ESO; Valor 0 si no ha llegado a superar el curso.                                |

Fuente: Elaboración propia.

### 3.2. Primera fase del análisis empírico: modelo predictivo del éxito del programa especial de apoyo a la educación

En esta fase se desarrolla un modelo consistente en un árbol de decisión CHAID o *chi-squared automatic interaction detection* (Kass, 1980), cuyo objetivo es doble: clasificación y predicción. Este modelo clasificará a cada uno de los estudiantes en grupos homogéneos en base a las variables explicativas de entrada o predictores

que interactúan significativamente con la variable dependiente de salida u objetivo (Akin et al., 2017; Khosravi et al., 2019). El algoritmo identifica los estudiantes que probablemente pertenezcan a un grupo en particular y, al mismo tiempo, identifica las reglas que utilizará en predicciones de futuros casos de estudiantes que no hayan participado. En nuestro caso estas agrupaciones están constituidas en función de las características del estudiante, del programa

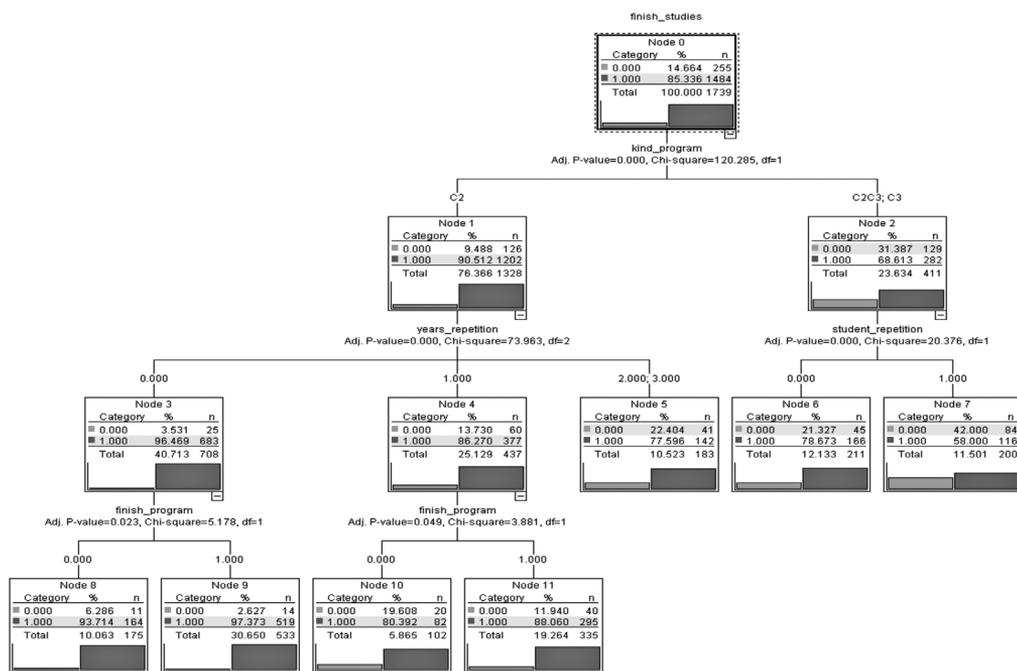
en el que participa y su desempeño, y posibilitan cumplir el segundo objetivo de determinar y predecir el porcentaje de aprobados en 4.º de la ESO tanto para cada uno de los grupos como individualmente (Ramaswami y Bhaskaran, 2010). Finalmente, la predicción de este porcentaje de éxito permite calcular como reforzar cada uno de los programas por tipología de estudiante.

La elección de árboles de decisión CHAID frente a otros métodos, como los propuestos para variables educativas por, por ejemplo, Cruz-Jesus et al., (2020) es que tienen la capacidad de generar árboles no binarios. Además, también admiten variables de entrada tipo numérico continuas o categóricas. Al ser no paramétrico puede manejar tanto relaciones lineales

como no lineales entre las variables explicativas y la variable de salida y también puede gestionar de forma muy eficiente grandes volúmenes de datos, incluso en tiempo real (Chassignol et al., 2018; Khosravi et al., 2019).

Para el desarrollo y entrenamiento del modelo se ha utilizado una metodología de entrenamiento, comprobación y validación (*training, testing, validación* [TTV]), que consiste en entrenar el modelo con el 70 % (1203 registros) de la muestra y realizar la comprobación y su validación con el restante 20 % (367 registros) y 10 % (169 registros) de la muestra respectivamente (Ballestar et al., 2019). El Gráfico 1 muestra el diagrama de árbol de decisión CHAID obtenido para el modelo de éxito del programa.

GRÁFICO 1. Diagrama de árbol de decisión CHAID para el modelo de éxito de los programas de apoyo a estudiantes para la superación del curso 4.º. ESO.



Fuente: Elaboración propia.

### 3.2.1. Primera fase del análisis empírico: Evaluación del modelo.

Los indicadores más relevantes de CHAID son la precisión en la clasificación, sensibilidad, especificidad, el área debajo de la curva ROC y el coeficiente de GINI. Estos indicadores de precisión y la matriz de confusión, que contiene el porcentaje de casos clasificados tanto correcta como incorrectamente para los dos posibles valores de la variable dependiente, están disponibles en la Tabla 2. Los indicadores han sido calculados tanto para el total de la muestra, como para cada una de las submuestras de entrenamiento, comprobación y validación del modelo. Por lo tanto, observamos que el modelo tiene una precisión en la clasificación de estudiantes y predicción de su porcentaje de estudiantes aprobados del 70.73 % (error del 29.27 %). Este porcentaje de estudiantes clasificados correctamente es muy similar entre todas las submuestras de entrenamiento, comprobación y validación, por lo que podemos afirmar que el modelo no ha sido sobreentrenado.

El porcentaje de positivos verdaderos, denominado *sensibilidad*, es del 71.43 %. Este valor explica el porcentaje de estudiantes que superan el curso y que han sido correctamente clasificados por el árbol de decisión CHAID en base a las características del estudiante, el programa de apoyo realizado y su resultado en dicho programa. El porcentaje de negativos verdaderos, también denominado *especificidad*, es del 66.67 %. Este valor es el porcentaje de estudiantes que no superan 4.º de la ESO y que han sido correctamente clasificado en base a las mismas variables de entrada. Los valores complementarios son el porcentaje de falsos positivos, 33.33 %, corresponde con el porcentaje de estudiantes que no habiendo su-

perado 4º de la ESO fueron clasificados por el modelo como que si lo habían conseguido. Finalmente, el porcentaje de falsos negativos es del 28.57 %. Este valor corresponde al porcentaje de estudiantes que habiendo superado el curso fueron clasificados por el modelo como que no lo habían superado.

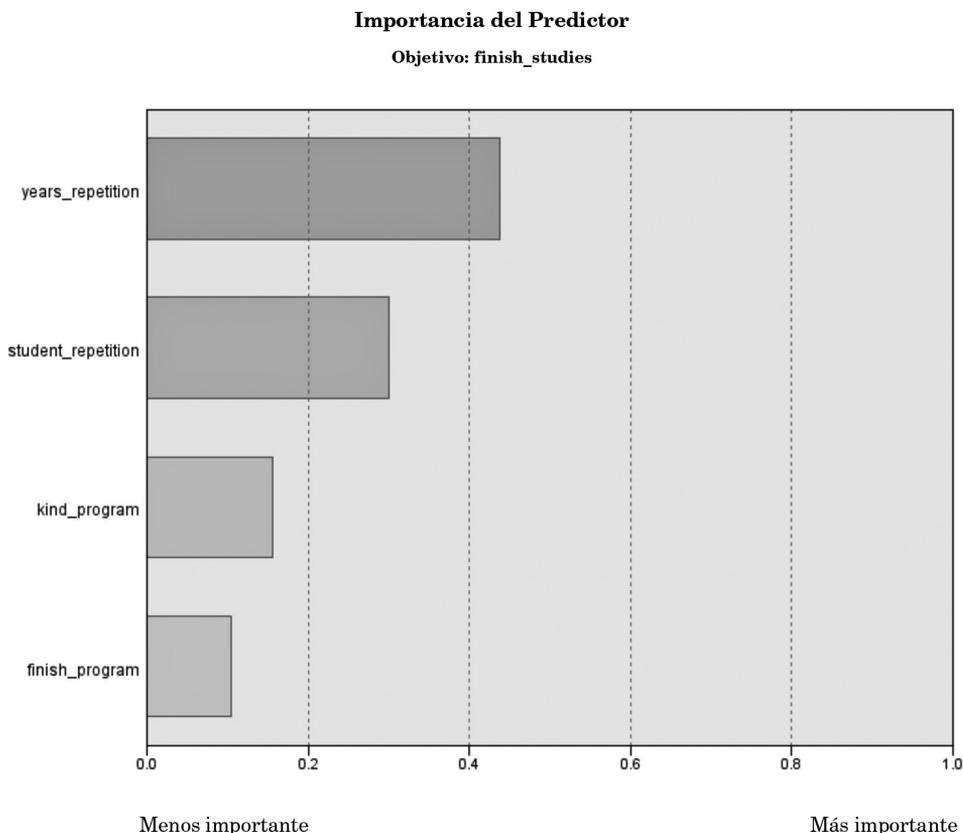
En esta investigación preferimos utilizar como principal medida de precisión del árbol de decisión CHAID el indicador de área debajo de la curva (AUC) ROC, ya que resulta más robusto que el indicador de precisión en la clasificación cuando se trabaja con muestras no balanceadas, como es nuestro caso (Tabla 2) (Dželihodžić y Jonko, 2016; Yin et al., 2013). El área debajo de la curva (AUC) ROC para el total de la muestra tiene un valor de 0.762, similar a las de las submuestras de entrenamiento, comprobación y validación del modelo (0.763, 0.760 y 0.755 respectivamente). Los valores de AUC pueden oscilar entre 0.5, que es el peor valor posible e implicaría que el modelo hace clasificaciones al azar y 1, que es el mejor valor y significaría que el modelo hace clasificaciones perfectas. Por lo tanto, se puede concluir que la calidad de este modelo de árbol de decisión CHAID es bueno (Hosmer Jr. et al., 2013). De forma complementaria, también se ha calculado el coeficiente de GINI, que se puede utilizar alternativamente al AUC, ya que están estrechamente relacionados. El coeficiente de GINI representa dos veces el área comprendida entre la curva ROC y la diagonal y oscila entre los valores 0 y 1. En nuestra investigación el coeficiente de GINI para el total de la muestra tiene un valor de 0.523, similar a la de las submuestras de entrenamiento, comprobación y validación del modelo (0.527, 0.519 y 0.510 respectivamente).

Tabla 2. Precisión del modelo y matriz de confusión del árbol de decisión CHAID.

| Precisión del Modelo |                       | Matriz de Confusión        |       |            |                      |                |                  |                   |
|----------------------|-----------------------|----------------------------|-------|------------|----------------------|----------------|------------------|-------------------|
|                      |                       | Predicción                 |       |            | Tamaño muestra       |                |                  |                   |
|                      |                       | Matriz de Confusión        |       |            | Predicción           |                |                  |                   |
|                      |                       | 0                          | 1     | 0          | 1                    | 0              | 1                |                   |
| <b>Muestra</b>       | <b>Tamaño Muestra</b> | <b>Porcentaje Correcto</b> |       | <b>AUC</b> | <b>GINI</b>          | <b>Muestra</b> | <b>Observado</b> | <b>Porcentaje</b> |
| <b>Total Muestra</b> | 1739                  | 70.73 %                    | 0.762 | 0.523      | <b>Total Muestra</b> | 0              | 170              | 85                |
|                      |                       |                            |       |            |                      | 1              | 424              | 1060              |
|                      |                       |                            |       |            |                      |                |                  | 28.57 %           |
|                      |                       |                            |       |            |                      |                |                  | 71.43 %           |
| <b>Submuestra</b>    |                       |                            |       |            |                      |                |                  |                   |
| <b>Entrenamiento</b> | 1203                  | 70.91 %                    | 0.763 | 0.527      | <b>Entrenamiento</b> | 0              | 117              | 62                |
|                      |                       |                            |       |            |                      | 1              | 288              | 736               |
|                      |                       |                            |       |            |                      |                |                  | 28.13 %           |
|                      |                       |                            |       |            |                      |                |                  | 71.88 %           |
| <b>Comprobación</b>  | 367                   | 69.21 %                    | 0.760 | 0.519      | <b>Comprobación</b>  | 0              | 34               | 16                |
|                      |                       |                            |       |            |                      | 1              | 97               | 220               |
|                      |                       |                            |       |            |                      |                |                  | 30.60 %           |
|                      |                       |                            |       |            |                      |                |                  | 69.40 %           |
| <b>Validación</b>    | 169                   | 72.78 %                    | 0.755 | 0.510      | <b>Validación</b>    | 0              | 19               | 7                 |
|                      |                       |                            |       |            |                      | 1              | 39               | 104               |
|                      |                       |                            |       |            |                      |                |                  | 27.27 %           |
|                      |                       |                            |       |            |                      |                |                  | 72.73 %           |

Fuente: Elaboración propia.

GRÁFICO 2. Importancia relativa de las variables de entrada en el árbol de decisión CHAID.



Fuente: Elaboración propia.

El Gráfico 2 muestra la importancia relativa de cada una de las variables predictoras significativas en el árbol de decisión CHAID. La importancia de estos predictores se determina calculando la reducción de la varianza de la variable objetivo (finalización o no de 4.º de la ESO por estudiante) que se puede atribuir a cada predictor mediante la realización de un análisis de sensibilidad (Saltelli et al., 2004). En esta investigación, la variable que acumula el 44 % de la importancia relativa es el número de años que ha repetido el estudiante (*years\_repetition*), seguida de la variable que indica si el estudiante ha re-

petido o no (*student\_repetition*) con una importancia del 30 %. A continuación, con un 16 %, se encuentra el tipo de programa en el que participó el estudiante (*kind\_program*) y, finalmente, con un 10 % si el estudiante finalizó el programa especial de apoyo (*finish\_program*).

### 3.2.2. Primera fase del análisis empírico: Análisis de los resultados del modelo árbol de decisión CHAID

El modelo ML basado en CHAID da lugar a un árbol en el que el primer grupo, denominado *nodo raíz*, representa el conjunto

total de la muestra, en este caso, de los 1739 estudiantes analizados. El algoritmo divide este conjunto en dos o más categorías denominadas *nodos padre o iniciales*. Debajo de los nodos padre se encuentran vinculados a ellos los nodos hijo. Las categorías que se encuentran en el último nivel del árbol de decisión se denominan *nodos terminales*. En lo que respecta a la jerarquía, los *nodos padre o iniciales* ejercen la mayor influencia sobre el *nodo raíz*, correspondiente a la variable dependiente que pretende explicar, mientras que ejercen una menor influencia sobre los *nodos terminales*.

Como se observa en el Gráfico 1, el número total de nodos es de 12, distribuidos de la siguiente forma en 3 niveles de profundidad: 1 nodo raíz (nivel 0), 4 nodos padre (distribuidos en los niveles 1 y 2) y 7 nodos terminales (distribuidos en los niveles 2 y 3). Además, en base al análisis de evaluación del modelo realizado en el apartado anterior se ha demostrado que este modelo tiene una buena capacidad de segmentación y predicción (ROC de 0.762).

#### 3.2.2.1. Nivel 0 del Árbol de decisión CHAID

En el Nivel 0 del árbol de decisión CHAID, el Node 0 es la variable a predecir, que determina si el estudiante ha finalizado con éxito 4.º de la ESO (*finish\_studies*). En este nivel se observa que, para el total de la muestra, el 85.34 % ha superado el curso, mientras que el 14.66 % no lo ha conseguido.

#### 3.2.2.2. Nivel 1 del Árbol de decisión CHAID

En este nivel, la variable más relevante para definir los dos primeros nodos padre (Node 1 y Node 2) es el tipo de programa en el que el estudiante ha participado

(*kind\_program*). Cada uno de estos nodos dará lugar a un subárbol que usarán diferentes variables predictoras para definir los siguientes niveles. En este nivel, la variable de mayor relevancia es el programa en el que participó el estudiante, lo llegue a finalizar o no. El 90.51 % de los estudiantes que ha participado en el programa C2 (acompañamiento del estudiante durante todo 4.º de la ESO) ha superado el curso 4º de la ESO, en comparación con el 68.61 % de aprobados para los que han participado en los programas C3 y C2C3. Por lo tanto, el programa con mayor éxito en sus objetivos es el C2.

#### 3.2.2.3. Nivel 2 del Subárbol 1 (Izquierda) de decisión CHAID

La siguiente variable más relevante habiendo participado en el programa C2 es el número de veces que han repetido curso (*years\_repetition*). De forma general, ya sabemos que en el Node 1 aprueban el 90.51 % de los estudiantes, pero este porcentaje puede variar mucho dependiendo de si el estudiante ha repetido algún curso o no previamente: los estudiantes que no han repetido curso aumentan su porcentaje de aprobado hasta el 96.46 % (Node 3), mientras que aquellos que han repetido una vez lo ven reducido hasta el 86.27 % (Node 4) y los que han repetido dos o más veces hasta el 77.59 % (Node 5). Los no repetidores obtienen mejores resultados en el curso 4º de la ESO en comparación con los que han repetido algún curso en alguna ocasión.

En este Nivel 2 encontramos los nodos Node 3, 4 y 5. Por un lado, los nodos Node 3 y 4, en los que los estudiantes no han repetido curso o lo han hecho solo una vez,

dan lugar a los nuevos subárboles del Nivel 3. En este Nivel 3 la variable más discriminante es la que determina si el estudiante llegó a finalizar el programa C2 en el que participó (`finish_program`). Por otro lado, Node 5 del Nivel 2 es un nodo terminal (no tiene más subniveles), lo que implica que la probabilidad de superar 4.º de la ESO para los estudiantes de este grupo es independiente de su desempeño en el programa C2, al contrario que para los nodos Node 3 y 4.

#### 3.2.2.4. Nivel 3 del Subárbol 1 (Izquierda) de decisión CHAID

En el Nivel 3 se encuentran 4 nodos terminales: Node 8 y Node 9 corresponden a los estudiantes que habiendo participado en C2 no han repetido anteriormente y proceden del subárbol de Node 3. Los estudiantes del Node 9 han finalizado el programa C2 por lo que el 97.37 % han logrado superar el curso 4.º de la ESO, la cifra de mayor éxito de todos los segmentos del árbol de decisión CHAID. Mientras que en Node 8 no han llegado a finalizar el programa C2 y su porcentaje de éxito disminuye al 93.71 %.

Node 10 y Node 11 corresponden a los estudiantes en C2 que han repetido curso en una ocasión y proceden del subárbol de Node 4. Los estudiantes del Node 11 han finalizado el programa C2 por lo que el 88.06 % han logrado superar el curso 4.º de la ESO, mientras que los de Node 10 ven reducida esta cifra hasta el 80.39 % por no haber finalizado el programa C2. Por lo tanto, los estudiantes que finalizan el programa C2 obtienen mejores resultados en el curso 4.º de la ESO en comparación con los que no han llegado a finalizar el programa.

#### 3.2.2.5. Nivel 2 del Subárbol 2 (Derecha) de decisión CHAID

En el nivel 2 del subárbol 2 (derecha, haber formado parte de C2C3 o C3) es si el estudiante ha repetido alguna vez curso (`student_repetition`). Sabemos que en el Node 2 aprueban tan solo el 68.61 % de los estudiantes, pero este porcentaje puede variar mucho dependiendo de si el estudiante es repetidor o no. Se obtiene que los estudiantes que no han repetido ningún curso aumentan su porcentaje de aprobado de 4.º de la ESO hasta el 78.67 % (Node 6), mientras que aquellos que han repetido curso ven reducido este porcentaje hasta el 58 % (Node 7). Por lo tanto, los no repetidores obtienen mejores resultados en el curso 4.º de la ESO en comparación con los que han repetido en alguna ocasión. En este subárbol 2 (derecha) ya no existen más niveles siendo esta variable la última relevante para generar más grupos de estudiantes.

### 3.3. Segunda fase del análisis empírico: análisis de robustez y calidad predictiva del modelo

En esta investigación se aplica una metodología de triangulación para comprobar la robustez de los resultados obtenidos en la primera fase. Para ello, se desarrolla un modelo predictivo ML basado en una red neuronal artificial perceptrón multicapa (ANN-MLP). Ambos modelos, utilizan la misma base de datos con las mismas variables de entrada (caracterización del estudiante, desempeño y tipo de programa realizado) y salida (desempeño del estudiante en 4.º de la ESO) (Tabla 1), pero están contruidos utilizando métodos de ML completamente diferentes, con el objetivo de verificar si obtienen resultados confirmatorios (Wolszczak-Derlacz y Parteka, 2011).

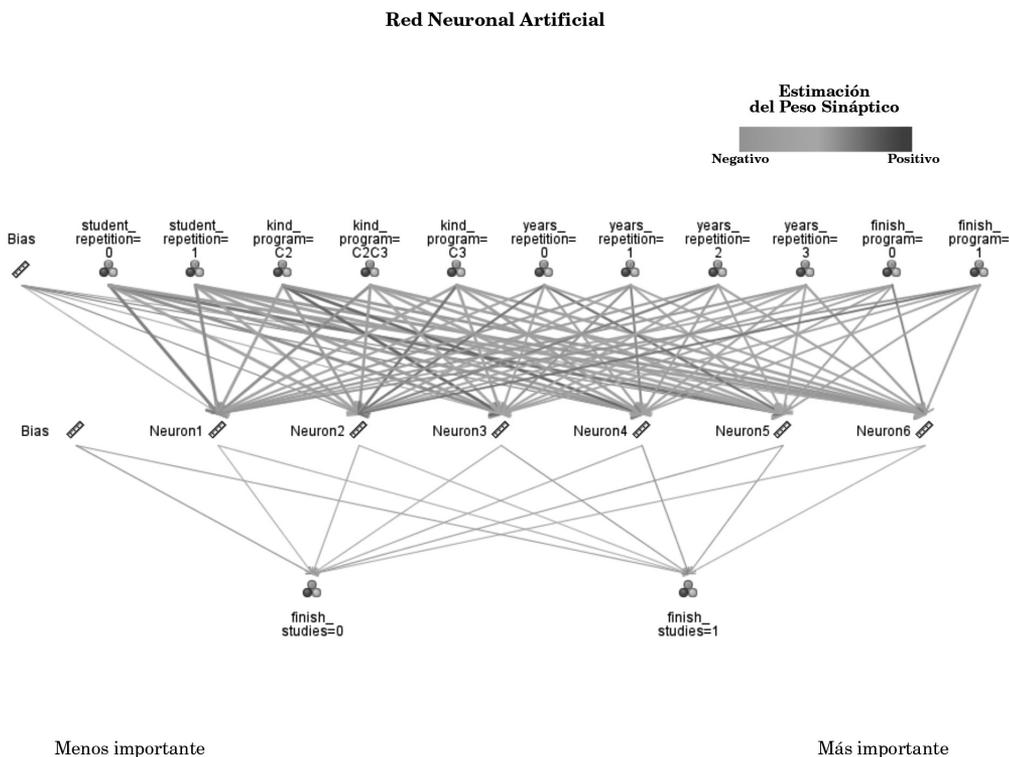
Se eligió ANN-MLP por la capacidad de las redes neuronales para manejar relaciones complejas entre variables, tanto lineales como no lineales, lo que las convierte en alternativa a modelos predictivos más tradicionales como regresiones logísticas donde la variable a predecir es dicotómica (Paliwal y Kumar, 2009). De esta forma, se asegura que no existen ningún tipo de relaciones que se hayan dejado de identificar.

A la hora de entrenar la ANN-MLP se debe tener en cuenta si la distribución de la variable dependiente es desbalanceada. Este sesgo en nuestra muestra podría impactar negativamente en el proceso de

entrenamiento de la ANN-MLP. Para evitarlo, se aplica un método de *oversampling* sobre el grupo menos representado para entrenar el modelo con el 70 % (2098 registros) de la muestra ya balanceada, garantizando que ambos grupos de estudiantes están igualmente representados a la hora de entrenar el modelo.

Por otro lado, la validación y comprobación del modelo ANN-MLP se realiza respectivamente con el 20 % (367 registros) y 10 % (169 registros) de la muestra, esta vez, sin balancear. Como en el primer modelo, se sigue también una metodología, entrenamiento, comprobación y validación (TTV).

GRÁFICO 3. Arquitectura de la ANN-MLP para predecir el éxito de los programas de apoyo a estudiantes para la superación del curso 4.º de la ESO.



Fuente: Elaboración propia.

El Gráfico 3 muestra la arquitectura utilizada para predecir el éxito de los programas de apoyo a estudiantes para la superación de 4.º de la ESO formada por tres capas: la primera capa de entrada consta de once unidades, correspondientes a los posibles valores de las cuatro variables de entrada relativos a las características de los estudiantes, su desempeño y el programa realizado (Tabla 1). La siguiente es la intermedia y se denomina capa oculta, consta de seis unidades o neuronas. Finalmente, la última consta de dos unidades, correspondientes a la predicción sobre el desempeño del estudiante en 4.º de la ESO (Tabla 1). El tipo de función de activación de la capa oculta y la capa de salida corresponden con una tangente hiperbólica y *softmax* respectivamente.

En la Tabla 3 se describen los indicadores de precisión y la matriz de confusión de la ANN-MLP que demuestran que la calidad de este modelo es buena: clasificación correcta del 71.19 % de los estudiantes y un AUC de 0.763 (Hosmer Jr. et al., 2013). Además, también se verifica que no existe sobre-entrenamiento ya que los indicadores obtenidos tanto para el total de la muestra como para las submuestras de entrenamiento, comprobación y validación son similares.

#### 4. Discusión

Los indicadores que hemos utilizado para evaluar la precisión de la ANN-MLP como predictor del éxito de los estudiantes son los mismos que los utilizados para el CHAID, facilitando así la comparación entre resultados, calidad en la predicción y robustez de ambos modelos.

En la Tabla 4 se realiza la comparativa entre los indicadores de la precisión en la clasificación y matriz de confusión de ambos modelos (Tabla 3 vs. Tabla 2). Se observa que las predicciones y precisión de la ANN-MLP son consistentes con las realizadas por el árbol de decisión CHAID. La ANN-MLP predice correctamente el 71.19 % (error del 8.81 %) de los estudiantes que superarán el curso de 4.º de la ESO. Esta cifra representa un 0.46 % (8 estudiantes) de mayor capacidad predictiva frente al 70.73 % de aciertos obtenidos por el árbol de decisión CHAID. Para el resto de los indicadores, como la sensibilidad, especificidad, falsos positivos y falsos negativos, las diferencias también son estadísticamente no significativas, oscilando entre -1.57 % (diferencia en la clasificación de 4 estudiantes) y 0.81 % (diferencia en la clasificación de 12 estudiantes). Lo mismo sucede con las diferencias existentes entre las áreas debajo de la curva (AUC) ROC y coeficiente de GINI. El AUC para la ANN-MLP y el árbol de decisión CHAID, 0.763 y 0.762 respectivamente y el coeficiente de GINI 0.525 y 0.523 respectivamente. En el Gráfico 4 se muestra la importancia relativa de las variables o predictores de entrada de la ANN-MLP siguiendo el método de la varianza y en total suman 1.

Si bien es cierto que tanto el árbol de decisión CHAID como la ANN-MLP utilizan la misma base de datos y variables en el modelo y consiguen alcanzar resultados, capacidad de predicción y robustez muy similares, la forma de hacerlo es muy diferente, tanto en los métodos utilizados como en el manejo e importancia de las variables que contribuyen en el modelo.

Tabla 3. Precisión del modelo y matriz de confusión de la red neuronal artificial perceptrón multicapa (ANN-MLP).

| Precisión del Modelo |                | Matriz de Confusión |       |       |                      |            |                |      |         |         |
|----------------------|----------------|---------------------|-------|-------|----------------------|------------|----------------|------|---------|---------|
| Muestra              | Tamaño Muestra | Porcentaje Correcto | AUC   | GINI  | Muestra              | Predicción |                |      |         |         |
|                      |                |                     |       |       |                      | Observado  | Tamaño muestra |      |         |         |
|                      |                |                     |       |       |                      | 0          | 1              | 0    | 1       |         |
| <b>Total Muestra</b> | 1739           | 71.19 %             | 0.763 | 0.525 | <b>Total Muestra</b> | 0          | 166            | 89   | 65.10 % | 34.90 % |
|                      |                |                     |       |       |                      | 1          | 412            | 1072 | 27.76 % | 72.24 % |
| <b>Submuestra</b>    |                |                     |       |       |                      |            |                |      |         |         |
| <b>Entrenamiento</b> | 2098*          | 69.78 %             | 0.767 | 0.535 | <b>Entrenamiento</b> | 0          | 720            | 354  | 67.04 % | 32.96 % |
|                      |                |                     |       |       |                      | 1          | 280            | 744  | 27.34 % | 72.66 % |
| <b>Comprobación</b>  | 367            | 70.03 %             | 0.755 | 0.510 | <b>Comprobación</b>  | 0          | 31             | 19   | 62.00 % | 38.00 % |
|                      |                |                     |       |       |                      | 1          | 91             | 226  | 28.71 % | 71.29 % |
| <b>Validación</b>    | 169            | 69.23 %             | 0.744 | 0,488 | <b>Validación</b>    | 0          | 15             | 11   | 57.69 % | 42.31 % |
|                      |                |                     |       |       |                      | 1          | 41             | 102  | 28.67 % | 71.33 % |

\**Overampled* para balancear la muestra.

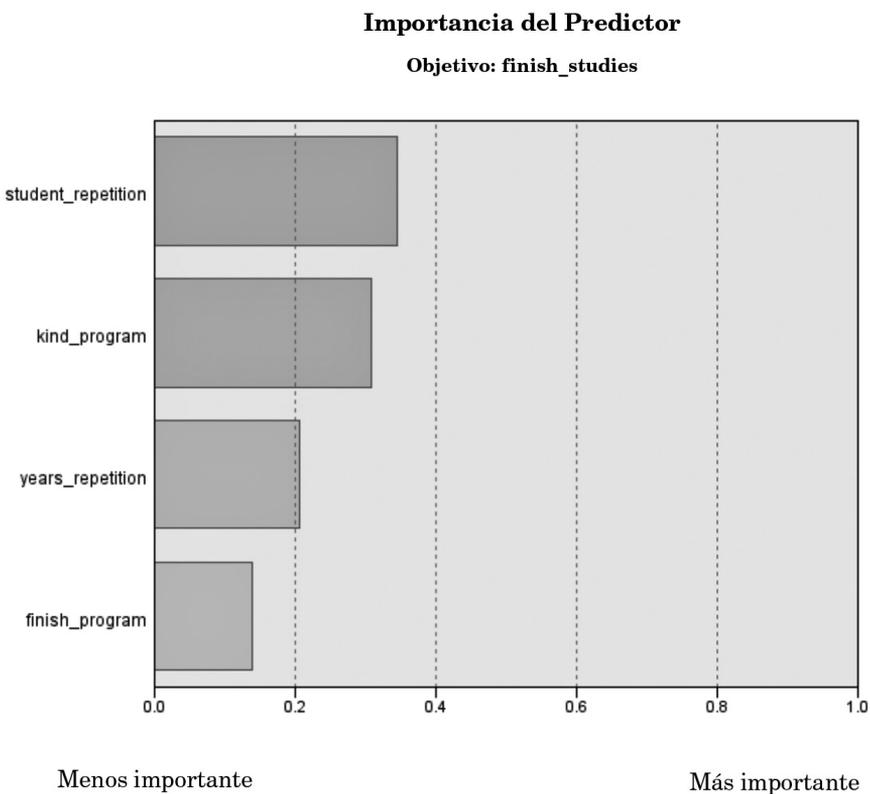
Fuente: Elaboración propia.

TABLA 4. Comparación indicadores: Precisión de la clasificación y matriz de confusión de la ANN-MLP en comparación con el árbol de decisión CHAID.

| Indicadores                | ANN-MLP comparado con Arbol CHAID |                        |
|----------------------------|-----------------------------------|------------------------|
|                            | Variación %                       | Variación clasificados |
| <b>Porcentaje Correcto</b> | 0.46 %                            | 8                      |
| <b>Sensibilidad</b>        | 0.81 %                            | 12                     |
| <b>Especificidad</b>       | -1.57 %                           | -4                     |
| <b>Falsos positivos</b>    | 1.57 %                            | 4                      |
| <b>Falsos negativos</b>    | -0.81 %                           | -12                    |
| <b>AUC</b>                 | 0.001                             | -                      |
| <b>GINI</b>                | 0.002                             | -                      |

Fuente: Elaboración propia.

GRÁFICO 4. Importancia relativa de las variables de entrada de la red neuronal artificial perceptrón multicapa (ANN-MLP).



Fuente: Elaboración propia.

En el caso de la ANN-MLP la variable que acumula mayor importancia relativa acumula hasta el 35 % y es la que indica si el estudiante ha repetido algún curso (*student\_repetition*), seguida del tipo de programa en el que participó el estudiante con una importancia del 31 % (*kind\_program*). A continuación, con una importancia del 21 % se encuentra la variable que recoge el número de cursos que ha repetido el estudiante (*years\_repetition*) y, finalmente, con una importancia del 13 %, la variable que expresa si el estudiante finalizó el programa especial de apoyo (*finish\_program*).

La aplicación de la metodología de triangulación concluye que los modelos de árbol de decisión CHAID y la ANN-MLP obtienen resultados similares en cuanto a precisión de clasificación y predicción del éxito de los programas de apoyo a estudiantes para la superación de 4.º de la ESO y, por lo tanto, son modelos robustos que confirman sus resultados. De los tres programas analizados el que mayor porcentaje de éxito ha conseguido es el C2 que acompaña al estudiante durante todo el curso de 4.º de la ESO, alcanzando un 90.51 % de aprobados del curso, comparado con el 68.61 % de las otras dos opciones C2 (clases extraordinarias en el mes de Julio) y C2C3 (combinación de las opciones anteriores).

En términos de evaluación educativa los resultados son impactantes: se puede evaluar el retorno económico y social de una intervención en función de las distintas características de los estudiantes, como el aprovechamiento de los cursos, el tipo de curso o, por supuesto, si son repetidores o no.

## 5. Conclusiones

A través de nuestra investigación hemos contrastado la necesidad de evaluar económicamente la LOMLOE. La ley partía con la misma dotación presupuestaria que la LOMCE para intervenciones como la analizada (45 millones de euros). La llegada de fondos europeos abre nuevas posibilidades no contempladas en su diseño inicial, por lo que se abre el reto de emplear esa inversión adicional de forma eficiente.

Nuestro análisis demuestra que se puede predecir la probabilidad que tienen los estudiantes en riesgo de superar 4.º de la ESO en función del programa de apoyo en el que participan. De esta forma se puede contribuir al diseño de programas más eficientes y con mayores tasas de éxito.

Para completar los modelos ya utilizados por Administraciones públicas, empleamos modelos de ML robustos como árboles de decisión CHAID y redes neuronales artificiales para analizar las características de los estudiantes, la intervención en la que han formado parte y los resultados finales. Esto permite maximizar el retorno social de cada programa y apoyar a aquellos estudiantes con menos posibilidades de éxito.

Por ejemplo, nuestros resultados muestran que, en el *Programa para la Mejora del Éxito Educativo* en Castilla y León, independientemente del programa en el que el estudiante ha participado, aquellos que nunca han repetido tienen mayor probabilidad de aprobar 4.º de la ESO que los que lo han hecho. La probabilidad de superar el curso aumenta cuando los estudiantes, además de participar en los programas especia-

les, lo finalizan. Por este motivo, el grupo de estudiantes que alcanza mejores resultados es el de no repetidores que participan y finalizan el programa C2 con un 97.37 % de aprobados de 4.º de la ESO. Mientras que el que peores resultados obtiene es el de repetidores que han participado en los programas C2C3 y C3 independientemente de que lo hayan finalizado o no con un 58 % de aprobados en 4.º de la ESO. Nuestro análisis también muestra que realizar los apoyos educativos en julio no constituyen un gran valor añadido por cuanto que el programa C1 con acompañamiento durante el curso obtiene mejores resultados que el C3 que aúna este refuerzo durante el año escolar y uno añadido en el mes de julio. En la siguiente fase de investigación comprobaremos si estos diferentes resultados se deben a los propios alumnos o es por las características del programa.

Como corolario, también demostramos que los modelos de ML como CHAID y redes neuronales artificiales son buenos candidatos a la hora de construir modelos que den respuesta a las preguntas educativas planteadas, como la importancia de tutelar al estudiante para que finalice dicho programa y no lo abandone, especialmente en el caso de los jóvenes que han repetido para los que el programa supone una herramienta muy eficiente para evitar el fracaso escolar y una forma eficiente de inversión del dinero público.

## Nota

<sup>1</sup> <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/the-impact-of-covid-19-on-school-starters/>

## Referencias bibliográficas

- Aad, G., Abbott, B., Abdallah, J., Abdelalim, A. A., Abdesselam, A., Abdinov, O., Abi, B., Abolins, M., Abramowicz, H., Abreu, H., Acerbi, E., Acharya, B. S., Adams, D. L., Addy, T. N., Adelman, J., Aderholz, M., Adomeit, S., Adragna, P., Abye, T., ... Zwailinski, L. (2012). Performance of missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s} = 7$  TeV with atlas [Rendimiento de la reconstrucción del momento transversal perdido en colisiones protón-protón en  $\sqrt{s} = 7$  TeV con atlas]. *European Physical Journal C*, 72 (1), 1-35. <https://doi.org/10.1140/epjc/s10052-011-1844-6>
- Akin, M., Eyduran, E. y Reed, B. M. (2017). Use of RSM and CHAID data mining algorithm for predicting mineral nutrition of hazelnut [Uso del algoritmo de minería de datos RSM y CHAID para predecir la nutrición mineral de la avellana]. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 128 (2), 303-316.
- Ballestar, M. T., Doncel, L. M., Sainz, J. y Ortigosa-Blanch, A. (2019). A novel machine learning approach for evaluation of public policies: An application in relation to the performance of university researchers [Un novedoso enfoque de aprendizaje automático para la evaluación de políticas públicas: una aplicación en relación con el rendimiento de los investigadores universitarios]. *Technological Forecasting and Social Change*, 149. <https://doi.org/10.1016/j.techfore.2019.119756>
- Ballestar, M. T., García-Lázaro, A. y Sainz, J. (2020). Todos los caminos llevan a la educación: un primer análisis de la robotización, la educación y el empleo. *Papeles de Economía Española*, 166, 33-49.
- Ballestar, M. T., Garcia-Lazaro, A., Sainz, J. y Sanz, I. (2022). Why is your company not robotic? The technology and human capital needed by firms to become robotic [¿Por qué su empresa no está robotizada? La tecnología y el capital humano que necesitan las empresas para robotizarse]. *Journal of Business Research*, 142, 328-343. <https://doi.org/10.1016/j.jbusres.2021.12.061>
- Ballestar, M. T., Grau-Carles, P. y Sainz, J. (2018). Customer segmentation in e-commerce: Applications to the cashback business model [Segmentación de clientes en el comercio electrónico: Aplicaciones al modelo de negocio

- del cashback]. *Journal of Business Research*, 88, 407-414. <https://doi.org/10.1016/j.jbusres.2017.11.047>
- Brunello, G. y de Paola, M. (2014). The costs of early school leaving in Europe [Los costes del abandono escolar en Europa]. *IZA Journal of Labor Policy*, 3 (1), 1-31. <https://doi.org/10.1186/2193-9004-3-22>
- Burgess, S. (16 de junio de 2020). *How we should deal with the lockdown learning loss in England's schools [Cómo debemos afrontar la pérdida de aprendizaje en las escuelas de Inglaterra]*. VOX Cepr Policy Portal. <https://bit.ly/3Eom12M>
- Chassinol, M., Khoroshavin, A., Klimova, A. y Bilyatdinova, A. (2018). Artificial intelligence trends in education: A narrative overview [Tendencias de la inteligencia artificial en la educación: una visión general]. *Procedia Computer Science*, 136, 16-24. <https://doi.org/10.1016/j.procs.2018.08.233>
- Chatterjee, S. y Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling [Adopción de la inteligencia artificial en la enseñanza superior: un análisis cuantitativo mediante un modelo de ecuaciones estructurales]. *Education and Information Technologies*, 25 (5), 3443-3463. <https://doi.org/10.1007/s10639-020-10159-7>
- Crescenzi, R., Giua, M. y Sonzogno, G. V. (2021). Mind the Covid-19 crisis: An evidence-based implementation of Next Generation EU. *Journal of Policy Modeling*, 43 (2), 278-297. <https://doi.org/10.1016/J.JPOLMOD.2021.03.002>
- Cruz-Jesús, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M. y Rosa-Louro, A. (2020). Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. *Heliyon*, 6 (6), E04081. <https://doi.org/10.1016/j.heliyon.2020.e04081>
- De la Fuente, A., & Jimeno, J. F. (2009). The private and fiscal returns to schooling in the European Union [Los rendimientos privados y fiscales de la escolarización en la Unión Europea]. *Journal of the European Economic Association*, 7 (6), 1319-1360. <https://doi.org/10.1162/JEEA.2009.7.6.1319>
- Doncel, L. M., Durá, P., Grau, P. y Sainz, J. (2014). Fiscal sustainability and immigration in the Madrid region [Sostenibilidad fiscal e inmigración en la Comunidad de Madrid]. *International Migration*, 52 (6), 180-196. <https://doi.org/10.1111/imig.12004>
- Dželihodžić, A. y Jonko, D. (2016). Comparison of ensemble classification techniques and single classifiers performance for customer credit assessment [Comparación de las técnicas de clasificación por conjuntos y el rendimiento de los clasificadores individuales para la evaluación del crédito de los clientes]. *Modeling of Artificial Intelligence*, 3, 140-150.
- European Commission (2010). *EUROPE 2020: A strategy for smart, sustainable and inclusive growth [EUROPA 2020: Una estrategia para un crecimiento inteligente, sostenible e integrador]*. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:2020:FIN:EN:PDF>
- Golden, G. (2020). Education policy evaluation: Surveying the OECD landscape [Evaluación de la política educativa: un estudio del panorama de la OCDE]. *OECD Education Working Papers*, 236. OECD Publishing. [https://www.oecd-ilibrary.org/education/education-policy-evaluation\\_9f127490-en](https://www.oecd-ilibrary.org/education/education-policy-evaluation_9f127490-en)
- Goos, M., Manning, A. y Salomons, A. (2009). Job polarization in Europe [Polarización del empleo en Europa]. *American Economic Review*, 99 (2), 58-63. <https://doi.org/10.1257/aer.99.2.58>
- Gouédard, P., Pont, B. y Viennet, R. (2020). Education responses to COVID-19: Implementing a way forward. *OECD Education Working Papers*, 224. OECD Publishing. <https://doi.org/10.1787/8e95f977-en>
- Gregory, T., Salomons, A. y Zierahn, U. (2019). Racing With or Against the Machine? Evidence from Europe. *IZA Institute of Labor Economics*, 2019 (12063).
- Hosmer Jr., D. W., Lemeshow, S. y Sturdivant, R. X. (2013). *Applied logistic regression [Regresión logística aplicada]*. John Wiley & Sons.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data [Una técnica exploratoria para investigar grandes cantidades de datos categóricos]. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29 (2), 119-127. <https://doi.org/10.2307/2986296>
- Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, J. E., Toschi, M., Zisimopoulos, P., Sigaras, A., La-

- very, S., Cooper, L. A. D. y Hickman, C. (2019). Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization [El aprendizaje profundo permite evaluar y seleccionar de forma robusta los blastocistos humanos tras la fecundación in vitro]. *NPJ Digital Medicine*, 2 (1), 1-9.
- Kraft, M. (20 de agosto de 2015). What If Every Struggling Student Had a Tutor? [¿Y si todos los estudiantes con dificultades tuvieran un tutor?] *Education Week*. <https://www.edweek.org/education/opinion-what-if-every-struggling-student-had-a-tutor/2015/08>
- Ley Orgánica 10/2002, de 23 de diciembre, de Calidad de la Educación (LOCE). *Boletín Oficial del Estado*, 307, de 24 de diciembre de 2002, páginas 45188 a 45220. <https://www.boe.es/eli/es/lo/2002/12/23/10>
- Ley Orgánica 2/2006, de 3 de mayo de Educación (LOE). *Boletín Oficial del Estado*, 106. <https://www.boe.es/eli/es/lo/2006/05/03/2/con>
- Ley Orgánica 8/2013, de 9 de diciembre, para la Mejora de la Calidad Educativa (LOMCE). *Boletín Oficial del Estado*, 295, páginas 97858 a 97921. <https://www.boe.es/eli/es/lo/2013/12/09/8>
- Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación (LOMLOE). *Boletín Oficial del Estado*, 340, de 30 de diciembre de 2020, páginas 122868 a 122953. <https://www.boe.es/eli/es/lo/2020/12/29/3>
- Li, Z., & Eastman, J. R. (2006). Commitment and typicality measurements for fuzzy ARTMAP neural network [Medidas de compromiso y tipicidad para la red neuronal difusa ARTMAP]. *Geoinformatics 2006: Geospatial Information Science*, 6420. <https://doi.org/10.1117/12.712998>
- Maimon, O. y Rokach, L. (2005). *Data mining and knowledge discovery handbook [Manual de minería de datos y descubrimiento de conocimientos]*. Springer.
- Nickow, A., Oreopoulos, P. y Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence [Los impresionantes efectos de la tutoría en el aprendizaje de preescolar a 12 años: una revisión sistemática y un meta-análisis de las pruebas experimentales]. *National Bureau of Economic Research Working Papers Series*, 27476. <https://doi.org/10.3386/w27476>
- OECD. (2018). *Education policy outlook 2018: Putting student learning at the centre [Perspectivas de la política educativa 2018: Poner el aprendizaje de los estudiantes en el centro]*. OECD.
- Paliwal, M. y Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications [Redes neuronales y técnicas estadísticas: una revisión de las aplicaciones]. *Expert Systems with Applications*, 36 (1), 2-17.
- Pont, B. y Montt, G. (2014). *Education policy outlook. Spain [Perspectivas de la política educativa. España]*. OECD.
- Porte, C. de la y Jensen, M. D. (2021). The next generation EU: An analysis of the dimensions of conflict behind the deal [La UE de nueva generación: un análisis de las dimensiones del conflicto tras el acuerdo]. *Social Policy & Administration*, 55 (2), 388-402. <https://doi.org/10.1111/SPOL.12709>
- Ramaswami, M. y Bhaskaran, R. (2010). A CHAID based performance prediction model in educational data mining. *International Journal of Computer Science*, 7 (1), 10-18.
- Saltelli, A., Tarantola, S., Campolongo, F. y Ratto, M. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models (Vol. 1)*. Wiley Online Library.
- Schleicher, A. (2020). The impact of COVID-19 on education: Insights from education at a glance 2020 [El impacto de COVID-19 en la educación: visión de la educación en un vistazo 2020]. *OECD Journal: Economic Studies*. <https://www.oecd.org/education/the-impact-of-covid-19-on-education-insights-education-at-a-glance-2020.pdf>
- Slavin, R. E. (2016). Evidence-based education policies: Transforming educational practice and research [Políticas educativas basadas en la evidencia: transformar la práctica y la investigación educativas]. *Educational Researcher*, 31 (7), 15-21. <https://doi.org/10.3102/0013189X031007015>
- Wolszczak-Derlacz, J. y Parteka, A. (2011). Efficiency of European public higher education institutions: A two-stage multicountry approach [Eficiencia de las instituciones públicas europeas de enseñanza superior: un enfoque multinacio-

- nal en dos fases]. *Scientometrics*, 89 (3), 887-917. <https://doi.org/10.1007/s11192-011-0484-9>
- Yin, L., Ge, Y., Xiao, K., Wang, X. y Quan, X. (2013). Feature selection for high-dimensional imbalanced data [Selección de características para datos desequilibrados de alta dimensión]. *Neurocomputing*, 105, 3-11.
- Yusuf, A. (2007). *Standard-setting at UNESCO: Conventions, recommendations, declarations and charters adopted by UNESCO, 1948-2006 (Vol. 2) [La elaboración de normas en la UNESCO: Convenciones, recomendaciones, declaraciones y cartas adoptadas por la UNESCO, 1948-2006 (Vol. 2)]*. Martinus Nijhoff Publishers.

## Biografía de los autores

**María Teresa Ballestar** es Graduada en Estadística por la Universidad de Zaragoza, Licenciada en Investigación y Técnicas de Mercado y Master en Sociedad de la Información y el Conocimiento en la Universitat Oberta de Catalunya. Doctorada en metodologías y tecnologías Big Data aplicadas a la Economía en la Universidad Rey Juan Carlos. Durante los últimos años ha ocupado puestos de dirección y liderado proyectos de transformación digital, innovación, data analytics y data science. Además, es Profesora Asociada de la Universidad Rey Juan Carlos y ha publicado más de una decena de artículos científicos y de divulgación y colaborado con distintos medios de comunicación.

 <https://orcid.org/0000-0001-8526-7561>

**Jorge Sainz Orcid** es Licenciado en Ciencias Económicas por la Universidad Complutense de Madrid y en Derecho por la UNED. Doctor en Economía (URJC) y

MBA (especialización en finanzas y políticas públicas) Simon School, University of Rochester. Es Catedrático de Economía Aplicada en la Universidad Rey Juan Carlos de Madrid y Visiting Fellow en el Institute for Policy Research, University of Bath.

Ha sido Asesor del Gabinete de la consejera de Educación y Subdirector General de Investigación de la Consejería de Educación de la Comunidad de Madrid. En el Ministerio de Educación, Cultura y Deporte ha sido director general de Política Universitaria y Secretario General de Universidades.

 <https://orcid.org/0000-0001-8491-3154>

**Ismael Sanz** es Profesor Titular del Departamento de Economía Aplicada I de la Universidad Rey Juan Carlos. Doctor en Economía Aplicada en la Facultad de Ciencias Económicas y Empresariales de la Universidad Complutense de Madrid (UCM). Ha sido Director General de Innovación, Becas y Ayudas de la Consejería de Educación de la Comunidad de Madrid, Director del Instituto Nacional de Evaluación Educativa (INEE) del Ministerio de Educación, Cultura y Deporte (2012-2015) y Chair del Strategic Development Group de PISA de la OCDE (noviembre 2014-julio de 2015). Actualmente es Vicerrector de Calidad de la URJC. Autor de artículos en revistas de investigación de alto impacto.

 <https://orcid.org/0000-0003-1286-4124>