BARTOLOMEO, A. R. (2025). Reseña de tesis. *Modelos predictivos y explicativos del rendimiento académico: un estudio en la Facultad de Ciencias Económicas de la Universidad Nacional de Cuyo*. Doctorado en Ciencias Económicas, Facultad de Ciencias Económicas, Universidad Nacional de Cuyo. *Cuyonomics. Investigaciones en Economía Regional*, *9*(16), 156-162.

Reseña de tesis de doctorado

Modelos predictivos y explicativos del rendimiento académico. Un estudio en la Facultad de Ciencias Económicas de la Universidad Nacional de Cuyo

Tesis de doctorado en Ciencias Económicas, mención Administración
Facultad de Ciencias Económicas, Universidad Nacional de Cuyo
Mendoza, julio de 2025
197 páginas

Alejandro Ramón Bartolomeo
Facultad de Ciencias Económicas, Universidad Nacional de Cuyo
alejandro.bartolomeo@fce.uncu.edu.ar

URL de la revista: revistas.uncu.edu.ar/ojs3/index.php/cuyonomics
ISSN 2591-555X
Esta obra es distribuida bajo una Licencia Creative Commons
Atribución No Comercial – Compartir Igual 4.0 Internacional

Introducción

El propósito general del estudio es analizar la relación existente entre los indicadores más relevantes del rendimiento académico y las variables observadas en los estudiantes de la Facultad de Ciencias Económicas (FCE) de la Universidad Nacional de Cuyo (UNCUYO). El ámbito de la investigación se centra en los estudiantes de las carreras de Contador Público, Licenciatura en Administración y Licenciatura en Economía. La meta principal es utilizar los datos existentes en la institución para generar un modelo predictivo a través de técnicas de aprendizaje automático (*machine learning*) que permita anticipar el rendimiento académico de los alumnos universitarios en estas tres carreras.

La investigación se justifica por la magnitud del abandono de los estudios superiores en el sistema universitario argentino. Los indicadores nacionales señalan cierta volatilidad, ya que cerca del 37 % de los nuevos ingresantes no continúa estudiando al año siguiente, y aproximadamente el 23 % opta por una oferta académica distinta tras uno o dos años desde su ingreso. Además, solamente el 27,7 % de los egresados lo hace en el tiempo esperado de la carrera, lo que implica un retraso importante en el egreso. Conocer con anticipación las variables que podrían generar un problema de desempeño tiene un impacto directo en una adecuada administración de los recursos y en la gestión académica. La generación de un modelo predictivo, por lo tanto, es una piedra angular para que la gestión pueda anticipar acciones de soporte y contención a los estudiantes en riesgo.

El estudio está dividido en dos partes: la primera parte aborda el planteamiento general y el marco teórico (capítulos 1 a 6), mientras que la segunda parte desarrolla el estudio empírico (capítulos 7 a 14). La metodología empleada utiliza técnicas de *machine learning* que forman parte del mundo de la inteligencia artificial.

Hipótesis del estudio

La investigación se plantea con la siguiente hipótesis general: es posible desarrollar un modelo predictivo basado en variables personales, socioeconómicas y académicas, con especial énfasis en la edad y el promedio de notas del primer año, que permita predecir el desempeño académico de los alumnos de la Facultad de Ciencias

Económicas de la UNCUYO con una exactitud (*accuracy*) superior al 80 % y que, además, sea aplicable, con ajustes menores, a otras instituciones educativas y niveles académicos, como facultades y colegios secundarios.

Las principales hipótesis de trabajo elaboradas a partir de lo que se espera de la ejecución del modelo son: a) la edad es la variable demográfica de mayor peso predictivo en el desempeño académico, con un mejor rendimiento en estudiantes más jóvenes; b) el promedio de notas del primer año es un predictor más fuerte del éxito académico a lo largo de la carrera, y su inclusión en el modelo mejora significativamente la precisión del pronóstico; c) los estudiantes que trabajan a tiempo completo tienden a mostrar un desempeño académico inferior y mayores dificultades para egresar, y d) la edad y el promedio de notas del primer año son los principales predictores del rendimiento académico en todas las carreras.

Aspectos metodológicos

Enfoque y alcance

El estudio se inscribe en un enfoque cuantitativo con un diseño explicativo y transversal. El objetivo es identificar las variables que permiten predecir el desempeño académico de los estudiantes de la FCE-UNCUYO en 2022. Dada la accesibilidad a los registros de la institución, la investigación se llevó a cabo como un estudio censal que abarcó a toda la población de interés de primero a quinto año que realizó la reinscripción anual para 2022.

Fuente y operacionalización de datos

Los datos utilizados provienen del sistema informático de la FCE-UNCUYO, extraídos de la encuesta anual de reinscripción (datos personales, económicos y académicos) y del sistema Guaraní (trayectoria académica y notas).

Se definieron las variables a analizar en tres grandes grupos: personales, socioeconómicas y académicas. Un paso crucial fue la cuantificación del rendimiento académico, debido a la falta de estandarización que suele presentar la doctrina al respecto. El trabajo adoptó un enfoque mixto que combina el rendimiento como resultado (nivel de notas) y como proceso (regularidad).

El criterio de buen/mal desempeño se configuró como una combinación de estos dos aspectos, de acuerdo con los lineamientos de otros estudios. Se consideró «buen desempeño» a aquellos casos en que el promedio de notas (incluyendo aplazos) multiplicado por la cantidad de materias aprobadas por año (A/P, indicador de productividad media) resulta en un valor igual o superior a 28 (por ejemplo, nota promedio de 7 por 4 materias aprobadas al año).

Se generaron dos variables respuesta (*targets*) para los modelos: a) desempeño general (modelo 1), basado en el promedio de notas de toda la carrera; b) desempeño sin primer año (modelo 2 o ajustado), basado en el promedio de notas de segundo a quinto año, excluyendo las notas de primer año, que se usó como variable predictora independiente para evitar la endogeneidad.

Herramientas de modelado

Para el análisis y la modelización se empleó el lenguaje de programación Python y sus bibliotecas, con la herramienta PyCaret como núcleo del autoaprendizaje automático. El uso de la biblioteca PyCaret fue crucial porque permitió conocer el mejor modelo estadístico para los datos utilizados, ordenando los algoritmos de clasificación disponibles (como regresión logística, *gradient boosting*, *random forest*) según métricas de ejecución, como la exactitud (*accuracy*).

Conclusiones detalladas

La etapa de análisis gráfico preliminar (EDA) sugirió que la edad tenía una relación negativa con el desempeño (a mayor edad, peor desempeño) y que existía una correlación positiva relativamente fuerte (cercana a 0,53) entre las notas de primer año y las notas posteriores de la carrera. Estos hallazgos preliminares orientaron el modelado posterior.

Evaluación de los modelos predictivos (confirmación de la hipótesis)

Los resultados del modelado confirmaron la hipótesis principal: se lograron modelos con una exactitud (*accuracy*) mayor o igual al 80 %. El nivel de los indicadores de ejecución es muy satisfactorio, lo que implica que los modelos tienen un alto poder predictivo.

Modelo 1: desempeño general (algoritmo: *gradient boosting classifier*, GBM)

El algoritmo que demostró la mejor respuesta para el modelo de desempeño general fue el *gradient boosting classifier* (GBM). Exactitud (*accuracy*): el modelo logró una exactitud de 0,8182 (81,82 %) en la validación cruzada y 80,17) 0,8017 %) en los datos de prueba. Esto confirma que el modelo acierta en la predicción de buen/mal desempeño en aproximadamente el 80 % de los casos. Área bajo la curva (AUC): el valor fue de 0,89. Según la interpretación de Swets (1988), este valor se acerca a la «exactitud alta» (> 0,9), lo que denota la alta capacidad discriminante del modelo. Sensibilidad (*recall*): el modelo obtuvo una sensibilidad de 0,8162, lo que indica una

buena capacidad para clasificar correctamente a los estudiantes que realmente tienen mal desempeño. Coeficiente *kappa*: con un valor de 0,619, alcanzó un nivel de «coincidencia sustancial» entre la predicción y la realidad, una vez descontados los acuerdos por azar. Variables más importantes: el análisis de importancia de variables (RFECV) reveló que las variables esenciales en la predicción son la edad y las notas obtenidas en primer año, seguidas por la carrera (CPN o LA) y la cantidad de hijos.

Modelo 2: desempeño sin primer año (algoritmo: *logistic regression*, LR)

El mejor algoritmo en este caso fue la regresión logística. Aunque su desempeño fue ligeramente inferior al modelo 1, sigue siendo un modelo útil. Exactitud (*accuracy*): el valor de 0,8002 es similar al del modelo 1, lo que ratifica la alta predictividad. Sensibilidad (*recall*): es ligeramente menor (0,8002) que en el modelo 1, lo que señala una disminución pequeña en la capacidad de clasificar a los estudiantes con mal desempeño. Coeficiente MCC (*Matthews correlation coefficient*): el valor de 0,5848, aunque bueno, es más bajo que el 0,62 del modelo 1. Esta disminución es significativa y puede sugerir un leve desbalanceo en los datos, pero, sobre todo, subraya la importancia de las notas de primer año en la predicción.

La conclusión comparativa es que ambos modelos resultan altamente predictivos. La comparación entre el modelo 1 (GBM) y el modelo 2 (LR) demuestra que la inclusión de las notas de primer año en la variable respuesta del modelo 1 no invalida sus resultados y que este predictor es fundamental para la explicación del desempeño académico total.

La investigación concluye que el éxito en el desarrollo de estos modelos predictivos, utilizando solo datos existentes en la institución, abre la puerta a futuras líneas de investigación:

- Exploración de nuevas variables: incorporar datos externos no disponibles actualmente, como el desempeño en el nivel medio (notas) o los resultados de exámenes de ingreso, aunque se requeriría estandarizar los criterios de evaluación.
- Variables comportamentales: estudiar el impacto de variables poco convencionales, como el comportamiento en el uso de Internet (uso académico *versus* uso lúdico), lo que podría predecir el éxito académico.

La metodología de aprendizaje automático permitiría a la FCE-UNCUYO generar modelos más precisos, cambiando el paradigma de modelado al elegir el mejor algoritmo en función de su ejecución, lo cual es una herramienta valiosa y eficiente para la gestión universitaria.

Para comprender la utilidad de estos modelos predictivos podemos imaginarlos como un radar sofisticado en un aeropuerto. El radar (el modelo predictivo) no solo detecta a los aviones (estudiantes), sino que, basándose en su velocidad inicial (no-

tas de primer año) y su patrón de vuelo (edad y regularidad), es capaz de anticipar con un 80 % de certeza qué aviones llegarán a tiempo a su destino (buen desempeño) y cuáles están en riesgo de desviarse o tener problemas (mal desempeño), lo que permitiría a la torre de control (la gestión académica) intervenir antes de que el problema se manifieste completamente.

Bibliografía

BARTOLOMEO, A. y MACHIN URBAY, G. (4-6 de octubre de 2018). *Análisis cuantitativo de los factores relativos al fracaso académico de los estudiantes de Matemática y Cálculo Financiero de la Facultad de Ciencias Económicas de la Universidad Nacional de Cuyo.* Trabajo presentado en las XXXIX Jornadas Nacionales de Profesores Universitarios de Matemática Financiera. Villa Mercedes, San Luis.

BARTOLOMEO, A. y MACHIN URBAY, G. (4-6 de noviembre de 2021). *Modelo de regresión logística para determinar el desempeño académico de los alumnos de la Facultad de Ciencias Económicas de la Universidad Nacional de Cuyo.* Trabajo presentado en las XLII Jornadas Nacionales de Profesores Universitarios de Matemática Financiera. Mendoza.

BARTOLOMEO, A. y MACHIN URBAY, G. (20-22 de octubre de 2022). *Algoritmos de clasificación para medir el desempeño académico de alumnos universitarios: autoaprendizaje automático con PyCaret.* Trabajo presentado en las XLIII Jornadas Nacionales de Profesores Universitarios de Matemática Financiera. La Plata, Buenos Aires.

BARTOLOMEO, A. y MACHIN URBAY, G. (19-21 de octubre de 2023). Análisis del desempeño temprano: un enfoque de aprendizaje automático en los alumnos de la Facultad de Ciencias Económicas de la Universidad Nacional de Cuyo. Trabajo presentado en las XLIV Jornadas Nacionales de Profesores Universitarios de Matemática Financiera. Paraná, Entre Ríos.

COSCHIZA, C.; FERNÁNDEZ, J.M.; GAPEL REDCOZUB, G.; NIEVAS, M. y RUIZ, H. (2016). Características socioeconómicas y rendimiento académico. El caso de una universidad argentina. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 14*(3), 51-76. DOI:10.15366/reice2016.14.3.003.

MASCI, C.; JOHNES, G. y AGASISTI, T. (2018). Student and school performance across countries: A machine learning approach. *European Journal of Operational Research, 269*(3), 1072-1085.

MIGUÉIS, V. L.; FREITAS, A.; GARCIA, P. J. y SILVA, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems, 115*, 36-51.

MUSSO, M. F.; HERNÁNDEZ, C. F. R. y CASCALLAR, E. C. (2020). Predicting key educational outcomes in academic trajectories: a machine-learning approach. *Higher Education, 80*, 875-894.

PORTO, A. y DI GRESIA, L. (2004). Rendimiento de estudiantes universitarios y sus determinantes. *Revista De Economía Y Estadística*, *42*(1), 93-113. https://doi.org/10.55444/2451.7321.2004.v42.n1.3800.

SWETS, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285-1293.

TAFANI, R.; BOSCH, E.; CAMINATI, R.; CHIESA, G.; BRANQUER, G.; ESTRADA, S.; GASPIO, N. y ROGGERI, M. (2014). Educación y salud como *input* del capital humano. Rendimiento académico de estudiantes de la Facultad de Ciencias Económicas. UNRC. *Revista de Salud Pública*, *15*(1), 65-67. Recuperado el 24/11/2025 de https://revistas.unc.edu.ar/index.php/RSD/article/view/7012.

TEJEDOR, F. J. y GARCÍA-VALCÁRCEL, A. (2007). Causas del bajo rendimiento del estudiante universitario (en opinión de los profesores y alumnos). Propuestas de mejora en el marco del EEES. *Revista de Educación*, *342*(1), 443-473.