

HERRAMIENTAS DE ADMISIÓN Y RENDIMIENTO ACADÉMICO: EVIDENCIAS DE UN PRIMER CURSO EN UN GRADO EN ADMINISTRACIÓN DE EMPRESAS

JUAN MANUEL LÓPEZ ZAFRA

RICARDO A. QUERALT SÁNCHEZ DE LAS MATAS

CUNEF

SONIA DE PAZ COBO

Universidad Rey Juan Carlos

Las herramientas de admisión se han convertido en medios imperativos para que las escuelas privadas gestionen tanto el espacio limitado del que disponen como la búsqueda de la excelencia. La competencia por los mejores estudiantes de España se está haciendo cada día más presente por la presencia de diferentes elementos: las universidades públicas mejoran cada año, con algunos de ellas encabezando los distintos rankings de excelencia; la

crisis demográfica provoca a su vez una disminución en el volumen de estudiantes que acceden a la universidad; y el alza en los precios, que hace aún más difícil llegar a los mejores estudiantes. En el ámbito de la administración de empresas, la situación es paradigmática, pero no es exclusiva de ella. En el área de atención médica, el meta análisis realizado por Campbell y Dickson (1996) mostró los promedios de puntos de grado en enfermería y cursos relacionados como los mayores predictores cognitivos del éxito estudiantil; otros elementos como la educación y la edad de los padres se encontraron como los mejores predictores sociodemográficos. El debate sobre las pruebas de admisión como el SAT (el estándar utilizado entre las universidades estadounidenses) es presentado por Bollinger (2005). En el área de las escuelas de medicina, que se enfrentan a un gran número de solicitantes con plazas limitadas, Urlings-Strop *et al* (2013) trataron de «determinar la contribución relativa de las etapas académicas y no académicas a las diferencias encontradas en el

rendimiento de los estudiantes», comparando tanto a los estudiantes admitidos en la lotería como a tres grupos de participantes en el procedimiento de selección, y concluyendo que «la menor tasa de abandono escolar de los estudiantes seleccionados está relacionada tanto con la autoselección de participantes antes del inicio del procedimiento de selección como con la parte académica del procedimiento de selección».

En el presente artículo presentamos la investigación sobre el uso de un algoritmo supervisado, a saber, un árbol de regresión, para predecir la puntuación de los estudiantes admitidos en una escuela de negocios española de gestión privada. El objetivo principal es entender los efectos de las características definidas en el proceso de admisión para evaluar tanto la validez del proceso como la clasificación final del estudiante después de un año en la escuela. Después de mostrar el alcance de la investigación y los datos y la metodología en vigor, el análisis tanto de las estadísticas descriptivas como de los

resultados del árbol de regresión demostrará que el proceso de admisión en vigor funcionaría correctamente, incluso si se pudiera establecer algún ajuste fino para un rendimiento aún mejor.

ALCANCE DE LA INVESTIGACIÓN ▼

Con el fin de cumplir con los requisitos de la presente investigación, hemos accedido a los datos del proceso de admisión de los estudiantes que solicitaron, fueron seleccionados y finalmente admitidos en un centro universitario privado especializados en Administración de Empresas y Derecho de Madrid. Como es habitual en las escuelas y universidades privadas españolas, los estudiantes deben aprobar un proceso de admisión más allá de la *Prueba oficial de Acceso a la Universidad* (PAU), también conocido como *Selectividad*. Desde 2017, la prueba ha sido reemplazada por la Evaluación de Acceso a la Universidad (EBAU). Sea cual sea la prueba en vigor, tanto en los centros como en las universidades privadas existen herramientas de admisión adicionales en las que los estudiantes deben mostrar capacidades normalmente no capturadas en su expediente académico, y superar las competencias académicas habituales evaluadas por la prueba oficial de acceso a la universidad; de todos modos, ésta debe ser aprobada para que un estudiante tenga acceso en cualquier título oficial en España.

El proceso de admisión se basó en diferentes elementos: una entrevista, una prueba específica de matemáticas, una prueba de comprensión lectora, una prueba psicológica y los resultados que los alumnos lograron durante sus cursos 3º y 4º de Educación Secundaria Obligatoria (ESO). Otros elementos considerados fueron la calificación en su primer año de bachillerato y el tipo de bachillerato seguido. Dada la naturaleza nacional de la ESO, y dado que estos dos cursos cierran un periodo completo (algo que el primer año de bachillerato no logra), hemos considerado que la media de estos dos años de ESO representa una mejor imagen de las capacidades del alumno, renunciando a los dos últimos.

En el momento de desarrollar esta investigación, según el portal web Universia (<http://www.universia.es/estudios/grados/dg/269>), había 2.981 grados oficiales en España; en el área específica de Administración de Empresas, 137 facultades imparten el grado oficial, tanto como de forma individual (102) como junto con otro (35, las llamadas dobles titulaciones). Del total de centros que impartían la asignatura, 90 eran públicos y 47 privados: 11 en Madrid, 11 en Barcelona, 7 en Valencia, 4 en Alicante, 4 en Guipúzcoa, 3 en Valladolid, 2 en Ávila y 1 en La Rioja. Según el *Avance de la Estadística de Estudiantes* del curso académico 2017-2018 (http://www.educacionyfp.gob.es/servicios-al-ciudadano-mecd/estadisticas/educacion/universitaria/estadisticas/alumnado/2017-2018_Av/Grado-y-Ciclo).

html, Ministerio de Educación y Formación Profesional), durante ese año hubo 96.281 estudiantes matriculados en alguna universidad española en cursos de grado en administración de empresas (ADE); 80.323 de ellos (83,4%) a través de enseñanza presencial. El 16,9% (13.557) de la última cifra eran alumnos de universidades privadas. En centros adscritos a universidades oficiales, todos ellos con docencia presencial, había 6.746 alumnos, y 5.350 de ellos (79,3%) en centros adscritos a universidades públicas. Si sólo prestamos atención a aquellos en su primer año, y simplemente considerando los estudios puros de ADE (dejando de lado, por tanto, otros estudios similares), estamos hablando de un total de 34.144 estudiantes durante el año académico 2017-2018; de ellos, 27.320 escogieron formación presencial, y 5.381 siguieron sus estudios de primer curso en algún colegio de la comunidad autónoma de Madrid. Sólo el 19,8%, es decir, 1.066 de ellos, estaban matriculados en una escuela privada o universidad (no hay desglose para centros adscritos).

Las cifras explican la extrema competencia a la que se enfrentan las escuelas privadas en los estudios de administración de empresas. En ese mercado, la demanda de centros privados tiene que hacer frente a algunos competidores públicos de calidad con precios fuera del mercado, que o bien son capaces de alcanzar economías de escala, derivadas de su gran cantidad de servicios comunes, o no reflejan el coste real de los servicios ofrecidos. Los centros privados deben entonces romper esa barrera ofreciendo servicios (incluyendo, por supuesto, el servicio de la educación, el más importante desde lejos, pero no el único) que deben ser percibidos como mejores; sólo en ese caso, aquellos que exigen educación estarán dispuestos a pagar, en caso de que puedan hacerlo, un coste adicional, que claramente puede ser bastante importante. La lucha por la «medalla de oro» en la educación superior ya se ha convertido en una obsesión en Europa, como Bollinger (2005) señala para los Estados Unidos. Según el *Boletín Oficial de la Comunidad de Madrid* (BOCM, 2018), las tablas oficiales de precios públicos para los estudios de grado con un nivel 3 de experimentación (como los de ADE) muestran que el coste total de un grado de ADE en Madrid (el segundo más caro, justo después de la Comunidad Autónoma de Cataluña), 4 años y 240 ECTS (*European Credit Transfer and Accumulation System*, el sistema oficial de contabilidad en la UE para estudios de grado) alcanzaba los 5.133,6 euros, en caso de que el alumno no necesitase más que una sola convocatoria ordinaria y su, posible, extraordinaria. Un solo año en CUNEF tenía un precio de 10.500 euros, el equivalente en la universidad San Pablo - CEU (<https://notasdecorte.es/universidad-ceu-san-pablo>) era de 10.860 euros, 11.624,75 euros en la Universidad Pontificia Comillas – ICADE (<https://www.comillas.edu/grados/grado-en-ade-administracion-y-direccion-de-em-presas-e2>), y 21.000 euros en IE University (<https://>

www.ie.edu/es/universidad/admisiones/tasas-y-ayuda-financiera/metodos-de-pago/). Los tres primeros comparten el 28,2% del mercado total de estudiantes de primer curso en el grado de ADE en universidades privadas españolas.

OBJETIVOS Y METODOLOGÍA

Los objetivos de la presente investigación son los siguientes: partiendo de la información disponible de las pruebas de admisión (dejando de lado la entrevista), trataremos de determinar cuál es la mejor combinación de las variables empleadas para pronosticar la calificación final de los estudiantes al terminar su primer año en el grado de ADE; junto con esa combinación, también pretendemos definir las rutas o reglas de decisión que permitan la mejor predicción. La tarea principal es estudiar la posibilidad de mejorar las pruebas actuales e introducir mejoras en el proceso de admisión, si eso fuera asimismo posible.

La investigación ha seguido la metodología habitual de un proyecto de ciencia de datos: definición de la pregunta a responder, análisis exploratorio de datos, configuración y ajuste del modelo, análisis de los resultados y comunicación de los mismos. Las fases se representan, de una manera u otra, en Peng y Matsui (2016), Hardoon y Shmueli (2013) o Provost y Fawcett (2013), entre muchos otros. Los datos proceden del lago de datos interno del centro, y para el tratamiento estadístico empleamos la versión 3.4.3 R sobre Rstudio, versión 1.1.414, el último disponible en el momento en que comenzó la investigación.

Hemos procedido a través de la clasificación y regresión de árboles (CART) para descubrir los principales impulsores del proceso. Otros enfoques, como el de Hoefler y Gould (2000) se basaron tanto en técnicas estadísticas tradicionales (regresión lineal y no lineal) como en no tradicionales, como las redes neuronales. Incluso si este último se observaba como poderoso por los autores, renunciamos a él debido a su dificultad tanto en la implementación como en la comprensión y puesta en práctica en la línea de ejecución. En nuestro caso, y dadas las variables métricas, los árboles de regresión resultan de especial interés. Inicialmente introducidas por Breiman, Friedman, Olshen y Stone (1984), este tipo de herramientas son probablemente las más simples de implementar e interpretar entre los algoritmos de clasificación. En ese sentido, podemos nombrar la obra seminal de Murthy (1998), la posterior de Rokach y Maimon (2005) o la más novedosa de Anandajayam y Sivakumar (2018).

Entre las ventajas de la técnica, debemos enfatizar su capacidad para controlar los errores y los valores perdidos en los conjuntos de datos, y el hecho de no ser paramétricos, lo que significa que no hay suposiciones sobre la distribución del espacio y la estructura del clasificador. Por otro lado, presenta algunas desventajas, siendo la más citada la posibilidad de

sobreajuste debido a la alta flexibilidad de la técnica (Pandya y Pandya, 2015; Ye y Hou, 2018; Fratello y Tagliaferri, 2019).

ANÁLISIS

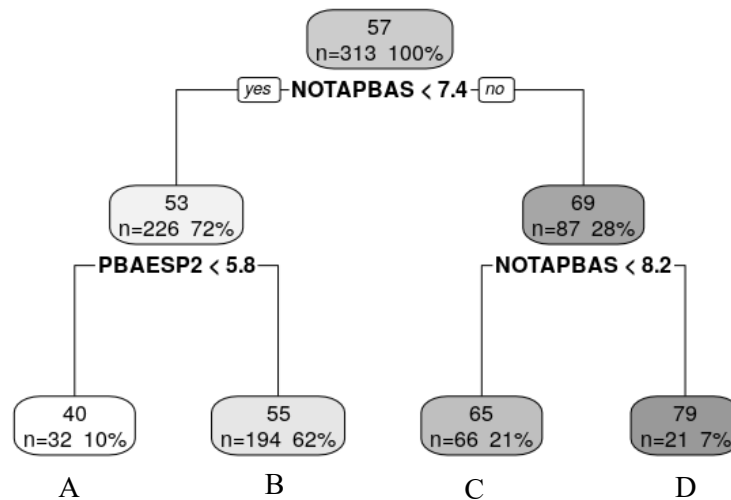
Nuestro conjunto de datos está formado por los 313 alumnos de los años 2015 y 2016, todos ellos estudiantes del grado de Administración de Empresas en CUNEF, tanto en su versión bilingüe inglés-español como en español solamente. Para la presente investigación, consideramos su género, sus resultados en la prueba de matemáticas, de comprensión escrita y las pruebas psicológicas (todas ellas, de forma conjunta, correctamente promediadas en la variable NOTAPBAS, puntuación), junto con sus calificaciones finales en los cursos 3º y 4º de Educación Secundaria Obligatoria (ESO, promediadas en PBAESP2). Los alumnos han sido clasificados según sus resultados en las 10 asignaturas diferentes que definen el plan de estudios oficial de la Universidad Complutense de Madrid (<https://economicasyempresariales.ucm.es/estudios/grado-ade-estudios-estructura>), que por entonces era el oficial de CUNEF, al ser centro adscrito. Las calificaciones finales del estudiante en cada asignatura se han ponderado de acuerdo con la convocatoria en la que el estudiante ha aprobado el examen: 1 si lo hizo en la primera, 0.75 si en la segunda, 0.5 si en la tercera y 0.25 si en la cuarta. Dado que las pruebas de admisión analizadas tuvieron lugar durante el primer semestre de los años 2015 (para los que entraron en la universidad septiembre de ese año) y 2016 (para los que entraron en septiembre de 2016), y que la investigación comenzó durante los últimos tres meses de 2017, el lapso máximo de cualquier estudiante es de dos años en la misma materia, generando un máximo de cuatro convocatorias para cualquier asignatura. Los que entraron en septiembre de 2016 tuvieron un máximo de dos convocatorias para la misma asignatura del primer año.

Una vez que asignamos a nuestros 313 estudiantes su calificación final en la escala de 100 puntos, procedimos a ordenarlos, definiendo tres grupos para la tarea de la investigación actual: los del cuartil superior (top 25), los del primer decil (bottom 10), y los que estaban entre ellos, el llamado grupo principal.

Estadísticas descriptivas

Para el total de los 313 estudiantes de la muestra, 132 (42,2%) eran mujeres; la proporción se mantiene para los estudiantes de la parte inferior 10 (13 del total 31 eran mujeres) pero cambia completamente en el top 25, donde 44 del total de 78 eran mujeres, un 56,4%. Según su procedencia académica, los resultados no fueron, de nuevo, sorprendentes. En España, en el momento de la investigación existían cuatro bachilleratos diferentes (humanidades y ciencias sociales, tecnología, ciencias naturales y de la salud) que permiten a un estudiante entrar en la

FIGURA 1
ÁRBOL DE REGRESIÓN FINAL



Fuente: Elaboración propia

universidad; el primero de ellos es la vía de acceso «natural» al grado en ADE, y esto fue el caso para el 78,3% de la muestra; en el top25, sólo el 69,2% accedieron a través de este bachillerato, mientras que los alumnos del bottom 10 mantuvieron la proporción del grupo general sin diferencias significativas (77,4%); sólo el 13,7% de los estudiantes del grupo principal procedían de un bachillerato técnico frente al 23,1% que venía del mismo en el top 25 y el 10% en el 10 inferior; 6,7% de la de estudios de salud en el grupo principal frente al 7,7% en el top 25 y 6,5% en el inferior 10 (diferencia no significativa); y si sólo un 1,3% del total procedían del bachillerato de artes en el grupo principal, no había ninguno en el top 25 mientras que un 6,5% en el bottom 10 procedían de este bachillerato.

La diferencia más significativa se da entre el primer cuartil de los estudiantes (top25) y el grupo principal más que entre el decil inferior y el principal: en el primer caso, tenemos más chicas (1,3 vs 0,73 por varón en el grupo principal) y una mayor proporción de estudiantes procedentes de bachilleratos más técnicos: la proporción general de 3,8 estudiantes de humanidades y ciencias sociales por estudiante de ciencias cae a 2,25 en el top 25%.

La puntuación media de los estudiantes al finalizar el primer año es de 57,07 puntos (intervalo de confianza del 95% -95% ci- 55,13 - 59,07) frente a 77,89 en el grupo del cuartil superior top25 (95% ci 76,62 - 79,17) y 23,34 en el decil inferior bottom10 (95% ci 18,90 - 27,78); Sus resultados en la ESO son 6,93 puntos en el grupo principal (en una escala de 10 puntos, 95% ci 6,83 - 7,03), 7,6 para el cuartil superior top25 (95% ci 7,38 - 7,83) y 6,28 (95% ci 6,00 - 6,56) para el decil inferior. Así que, claramente, existe una diferencia en términos de calificaciones de la ESO entre los tres grupos. Y la presencia de esta

diferencia en términos de rendimiento académico, junto con el género, nos empujó a analizar el tema a través de árboles de clasificación y regresión.

El árbol de regresión

A partir de nuestra muestra completa de 313 observaciones, construimos un árbol de regresión donde tratamos de predecir el valor de la calificación al finalizar el primer año a partir de los valores de género (SEXO), puntuación media en las pruebas (NOTAPBAS) y promedio de tercer y cuarto curso de ESO (PBAESP2). Utilizamos bibliotecas 'rpart' (partición recursiva para árboles de clasificación, regresión y supervivencia, «una implementación de la mayor parte de la funcionalidad» de Breiman, Friedman, Olshen y Stone, 1984) y 'rpart.plot' (para la representación gráfica de las bibliotecas de 'rpart'). Después de podar el árbol, para evitar el sobreajuste, cuando el parámetro de complejidad (cp) alcance el error mínimo con validación cruzada ($xerror$), a partir de un nivel inicial de $cp_i = 0.174943$, fuimos capaces de reducirlo a un nivel más adecuado de $cp_f = 0.018036$, después de sólo 3 divisiones. Como se sabe, el mejor valor del parámetro de complejidad cp es el que minimiza el RMSE (*root mean square error*, raíz cuadrada del error cuadrático medio). Las variables en el modelo final son sólo NOTAPBAS y PBAESP2, con una importancia de 57% y 43% respectivamente, mostrando que las pruebas internas realizadas en CUNEF mejoraron la capacidad predictiva del valor medio de los cursos de ESO. Como se muestra en la Figura 1, para el 7% que tiene una puntuación superior a 8,2 en las pruebas internas (nodo inferior derecho, D), la puntuación media prevista al final del año 1 será de 78,52 puntos (redondeada a 79 en la cifra). Las reglas de decisión son fáciles de interpretar: aparte de estas últimas, para aquellos en el ran-

go de 7,4 a 8,2, la puntuación media será de 64,48 puntos (nodo C). Los 87 estudiantes con puntuación en la prueba de admisión interna por encima del 7,4 (nodos C y D) representan el 27,8% de la muestra total; entre ellos, el valor mínimo de PBAESP2 en 6,36. El 10% inferior, bottom 10, formado por 32 alumnos, se construye según una condición principal: cualquiera que sea la puntuación en las pruebas de admisión, si la nota media en los 3º y 4º de ESO está por debajo de 5,8, la puntuación media con la que acabará el primer año del grado en ADE en CUNEF será de 40,12 puntos sobre 100. Como podemos observar, hay una ligera diferencia entre las puntuaciones de la muestra y las previstas por el árbol de regresión (nodo A): el decil inferior mostró un PBAESP2 promedio de 6,28 puntos frente a los 5,8 previstos por el árbol de regresión.

CONCLUSIÓN

Como hemos mostrado a lo largo de la presente investigación, los árboles de regresión pueden ser de ayuda en el proceso de admisión. Son útiles para definir las variables importantes (el género, incluso si no es adecuado para su uso de acuerdo con un protocolo no discriminatorio, se presumió erróneamente como una piedra angular) y presentan reglas de decisión claras y fáciles de seguir para la gestión. En nuestro caso, también han demostrado la importancia de las pruebas de admisión como elemento de alimentación por encima de las calificaciones oficiales que los estudiantes traen en su cartera. Algunos problemas han surgido, sin embargo, siendo el más importante la diferencia en la nota media que define el decil inferior de los estudiantes en términos de sus calificaciones de 3º y 4º cursos en Educación Secundaria Obligatoria. Esto se debe sin duda a un hecho inevitable: el pequeño tamaño de la muestra, que evita su correcta división en un conjunto de entrenamiento y uno de prueba, esencial para construir un modelo de aprendizaje automático. El trabajo futuro, junto con un tamaño más alto, observará posibilidades adicionales de acuerdo con nuevas variables como el colegio o instituto de origen del estudiante; esto ayudará a mejorar la definición de las normas de decisión en la persecución de la excelencia.

REFERENCIAS

- Anandajayam, P; Sivakumar, N (2018) The Study on Predictive Analysis Algorithm: Survey. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Volume 3, Issue 1.
- BOCM (2018). Boletín Oficial de la Comunidad de Madrid, 31 de agosto de 2018. Decreto 126/2018, de 31 de julio, por el que se establecen los precios públicos por estudios universitarios conducentes a títulos oficiales y servicios de naturaleza académica en las Universidades Públicas de la Comunidad de Madrid, pp. 9 a 15.
- Bollinger, L. C. (2005). Competition in higher education and admissions testing. In Camara, W, and Kimmel, E. (Eds.) *Choosing Students: Higher Education Admissions Tools for the 21st Century*. Taylor and Francis.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984) *Classification and Regression Trees*. Wadsworth International Group.
- Campbell, A. R.; Dickson, C. J. (1996). Predicting student success: A 10-year review using integrative review and meta-analysis. *Journal of Professional Nursing*, 12(1), 47–59. [http://doi.org/10.1016/S8755-7223\(96\)80074-3](http://doi.org/10.1016/S8755-7223(96)80074-3)
- Fratello, Michele; Tagliaferri, Roberto (2019) Decision trees and random forests. In Cannataro, Mario; Gaeta, Bruno, and Khan, Mohammad Asif (Eds) *Encyclopedia of bioinformatics and computational biology*, Vol. 1. Elsevier.
- Hardoon, D.; Shmueli, G. (2103) *Getting started with business analytics*. CRC Press.
- Hoefer, P.; Gould, J. (2000) Assessment of Admission Criteria for Predicting Students' Academic Performance in Graduate Business Programs, *Journal of Education for Business*, 75:4, 225-229, DOI: 10.1080/08832320009599019
- Murthy, Sreerama K. (1998) Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey. *Data Mining and Knowledge Discovery*, Volume 2, Issue 4, pp 345–389
- Pandya, Rutvija; Pandya, Jayati (2015) C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning. *International Journal of Computer Applications*, Volume 117, No. 16, Mayo
- Peng, R.; Matsui, E. (2016) *The Art of Data Science*. Skybrude Consulting, LLC
- Provost, F.; Fawcett, T. (2013) *Data Science for Business*. O'Reilly Media Inc., EE. UU.
- Rokach, Lior; Maimon, Oded (2005) Top-Down Induction of Decision Trees Classifiers—A Survey. *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*. December. DOI: 10.1109/TSMCC.2004.843247
- Ye, Jie; Hou, Li-duo (2018) Improvement and Application of Decision Tree C4. 5 Algorithm. *DEStech Transactions on Computer Science and Engineering*. DOI 10.12783/dtcse/CCNT2018/24686
- Urlings-Strop, L. C., Stegers-Jager, K. M., Stijnen, T.; Themmen, A. P. N. (2013). Academic and non-academic selection criteria in predicting medical school performance. *Medical Teacher*, 35(6), 497–502. <http://doi.org/10.3109/0142159X.2013.774333>