

Predicción temprana de deserción mediante aprendizaje automático en cursos profesionales en línea

(Early dropout prediction via machine learning in
professional online courses)

Ignacio Urteaga

Laura Siri

Universidad Tecnológica Nacional / Universidad de Buenos Aires (Argentina)

Guillermo Garófalo

Universidad Tecnológica Nacional (Argentina)

DOI: <http://dx.doi.org/10.5944/ried.23.2.26356>

Cómo referenciar este artículo:

Urteaga, I., Siri, L., y Garófalo, G. (2020). Predicción temprana de deserción mediante aprendizaje automático en cursos profesionales en línea. *RIED. Revista Iberoamericana de Educación a Distancia*, 23(2), pp. 147-167. doi: <http://dx.doi.org/10.5944/ried.23.2.26356>

Resumen

A pesar de las ventajas del *e-learning*, esta modalidad de aprendizaje es proclive a la deserción. Estudios anteriores mostraron que se pueden aplicar técnicas de aprendizaje automático a los registros de interacciones entre estudiantes y la plataforma para predecir el abandono. En esa línea, este trabajo intenta encontrar modelos predictivos de deserción en cursos virtuales que duran entre seis y dieciséis semanas, utilizando registros de Moodle correspondientes a las dos primeras. Se evaluó la sensibilidad, especificidad y precisión de los modelos, pero se priorizó más en qué medida dichos modelos facilitaban evitar la deserción mediante acciones de retención efectivas en costo. Específicamente, se usaron datos de varias cohortes de cuatro cursos de temáticas y duraciones distintas, dictados por la Secretaría de Extensión de la Universidad Tecnológica Nacional de la República Argentina, Regional Buenos Aires, entre febrero de 2018 y octubre de 2019. Se usaron distintos algoritmos para generar modelos predictivos y optimizarlos hacia la mitigación de la pérdida económica causada por la deserción. Se analizó si alguno en particular generaba los mejores modelos para todos los cursos. Se estudió si convenía construir modelos separados por curso o bien uno para todo el conjunto de los datos de los cuatro cursos. Como conclusión, se encontró que sí es posible construir modelos predictivos exitosos y que el algoritmo que produjo los mejores modelos

fue una red neuronal en tres de los cuatro cursos. Asimismo, resultó mejor el modelo que ajustó cada uno por separado.

Palabras clave: enseñanza a distancia; abandono de estudios; aprendizaje automático; predicción; modelos matemáticos; algoritmos.

Abstract

Despite the advantages of e-learning, this way of learning is prone to dropping out. Previous studies show that machine-learning techniques can be applied to records of interactions between students and the platform to predict abandonment. In this line, this work tries to find predictive dropout models in virtual courses that last between six and sixteen weeks, using Moodle logs from the first two. Models' sensitivity, specificity and precision were evaluated, but priority was given to the extent to which these models made it easier to avoid attrition through cost-effective retention actions. Specifically, data from several cohorts of four courses with different themes and durations were used. All four dictated by the Secretariat of Extension of the National Technological University of the Argentine Republic, Regional Buenos Aires between February 2018 and October 2019. Different algorithms were used to generate predictive models and optimize them in order to mitigate the economic losses caused by attrition. It was analyzed if any one in particular generated the best models for all courses. It was studied whether it was convenient to build separate models per course or one for the entire data set of the four courses. It was found that it is possible to build successful predictive models and that the algorithm that produced the best models was a neural network in three of the four courses. The model that fit each one separately turned out better.

Keywords: distance study; dropout; machine learning; forecasting; mathematical models; algorithms.

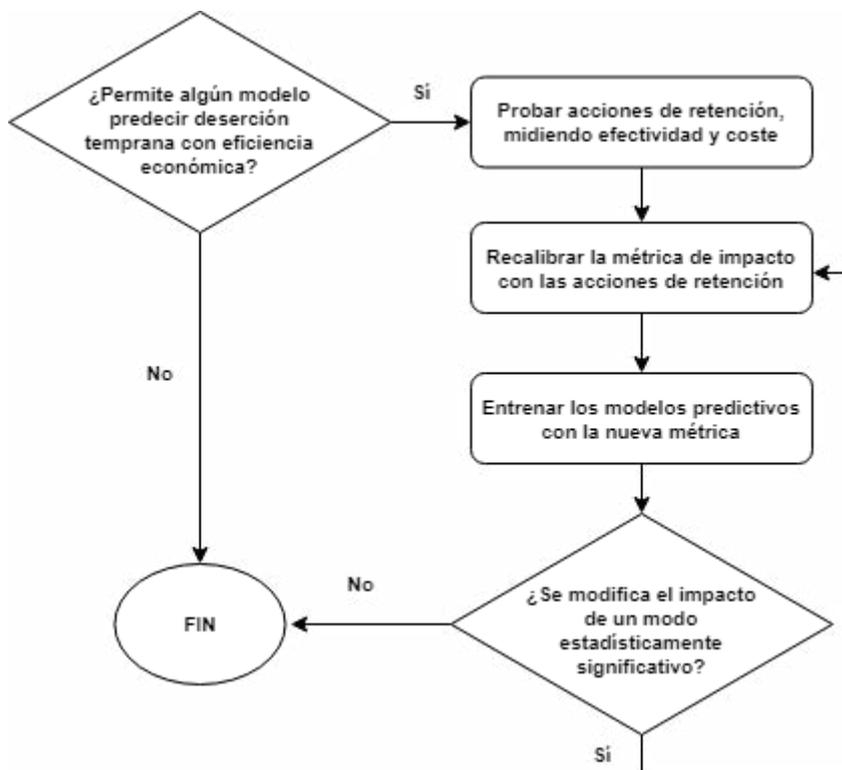
La educación a distancia basada en soportes digitales, o *e-learning*, ofrece múltiples ventajas, por ejemplo, ahorro en tiempo y costos de desplazamiento y flexibilidad horaria. Estos atractivos son especialmente relevantes para profesionales graduados que trabajan o tienen obligaciones familiares y aun así desean perfeccionarse. Sin embargo, muchos comienzan sus estudios bajo esta modalidad y abandonan en el camino. En el caso de posgrados o capacitaciones profesionales de pago, este abandono implica perder dinero invertido.

Según una revisión de Kara, Erdoğan, Kokoç y Cagiltay (2019), en el caso de los adultos matriculados en cursos *online* los motivos de abandono son internos, externos y relacionados con el programa educativo. Además, dependen de características individuales como la edad, el género, los conocimientos, las habilidades y el contexto. Sin duda, es importante conocer qué factores de los anteriormente mencionados inciden en la deserción y en qué grado, pero no siempre la institución puede rastrearlos. En todo caso, si ya desertaron, es tarde para hacer algo al respecto. Además, en el caso de cursos de pocas semanas, son factores en los que difícilmente

haya tiempo de intervenir (Baker, Lindrum, Lindrum y Perkowski, 2015). Por lo tanto, es útil intentar predecir el abandono antes de que ocurra y aplicar acciones retentivas. Pero, si el ejercicio de dichas acciones retentivas no fuera efectivo en costo para la institución, aplicarlas podría ser antieconómico y, por lo tanto, la predicción algorítmica de posibles desertores no serviría mucho, por más sensible, específico o preciso que fuera un modelo.

Nuestro trabajo apunta, justamente, a encontrar mediante algoritmos de aprendizaje automático el mejor modelo en términos de eficacia en costos para predecir tempranamente deserción en ciertos cursos universitarios breves en línea y de pago, con el fin de guiar futuras acciones retentivas y minimizar así el impacto económico del abandono. La intención no fue trabajar sobre las causas del problema y asumimos que probablemente sean similares a las reseñadas por Kara et al. (2019), entre otros. El desafío fue, más bien, encontrar un modo de predecir lo antes posible quiénes serán desertores en un contexto en el cual no hay tiempo o medios de conocer sus motivos y donde la principal fuente de datos consiste en las interacciones de los estudiantes con la plataforma virtual. Lo ideal sería que los docentes no utilicen tiempo (que tiene un costo) para ejercer este tipo de acciones con quienes no van a desertar y que sí lo hagan con los que estén en riesgo. Por eso, para medir la efectividad de la predicción, priorizamos hacerlo en términos del beneficio económico de cada deserción evitada y del costo de cada acción de retención realizada. Denominamos impacto a la métrica así generada. El alcance de este artículo es responder la pregunta inicial del programa recursivo ilustrado en la figura 1, teniendo en mente que la ejecución iterada de las acciones de dicho programa, a su vez, conducirá a optimizar la respuesta.

Figura 1. Programa recursivo



En síntesis, nuestros objetivos fueron:

1. Verificar si es posible predecir deserción solamente con los datos de las interacciones entre estudiantes y la plataforma en las dos primeras semanas de un curso.
2. Ver qué algoritmo produce los mejores modelos predictivos en términos de la métrica de impacto.
3. Comparar las ventajas de ajustar un solo modelo predictivo conjunto para todos los cursos que formaron nuestro corpus de prueba o hacerlo con cada uno por separado.

El análisis se realizó con los datos de cuatro programas de formación profesional dictados por el Centro de E-Learning de la Secretaría de Extensión Universitaria de la Universidad Tecnológica Nacional de la República Argentina, Facultad Regional Buenos Aires (UTN.BA), entre febrero de 2018 y octubre de 2019 en su

campus virtual basado en Moodle. Cada uno de esos cursos fue repetido en varias ocasiones, con el mismo temario, y se distinguen por un número de versión. En total, se consideraron cuatro versiones de la Diplomatura en Business Intelligence (DBI), cinco de la Diplomatura en bases de datos (DBD), cinco del Posgrado en Gestión de Proyectos (PGP) y cuatro del curso Data Scientist con R (R). Ninguno de estos cursos dura menos de seis semanas ni más de dieciséis.

A continuación, se reseñan investigaciones recientes acerca de la factibilidad de predecir deserción en *e-learning* utilizando únicamente los registros de interacciones entre estudiantes y la plataforma. Luego se sintetizan las características de la enseñanza virtual en la UTN.BA y, específicamente, para sus programas de Extensión Universitaria. En la sección Materiales y Métodos se detallan las métricas utilizadas para evaluar la capacidad predictiva de los modelos generados con cada uno de los algoritmos elegidos. En particular, se presenta la métrica de impacto y se explica por qué se consideró oportuno añadirla a otras más usuales, como sensibilidad-especificidad y precisión. En la sección de Resultados se muestran los valores de impacto, sensibilidad-especificidad y precisión de los mejores modelos obtenidos para los distintos algoritmos, donde se destacan los generados por redes neuronales. Finalmente, en la sección de Discusión y Conclusiones, se interpretan dichos valores y se postulan líneas de trabajo futuro para ajustar y generalizar aún mejor los hallazgos obtenidos en la presente investigación. Como allí se detalla, el algoritmo que produjo los mejores modelos de los utilizados fue una red neuronal de tipo perceptrón multicapa en tres de los cuatro cursos y resultó mejor el modelo que ajustó cada uno por separado que el que intentó hacerlo sobre todos en conjunto.

Antecedentes

En general, puede decirse que es necesario prevenir la deserción para que la educación, tanto *online* como presencial, alcance su potencial. Es decir, frenarla antes de que se produzca y determinar cuanto antes qué tipo de ayuda necesita el estudiante para persistir (Tinto, 2017). Por otra parte, las tasas de abandono y procrastinación son mayores en los cursos *online* que en los presenciales (Baker et al., 2015). Por eso, predecir deserción estudiantil en enseñanza a distancia en plataformas virtuales es una de las mayores preocupaciones de la Educational Data Mining (EDM) o Minería de Datos Educativo (García Saiz, 2016). Esta área de estudios analiza datos generados en la gestión educativa para crear modelos predictivos basados en diversos algoritmos matemáticos, aunque no necesariamente optimizados con base en los aspectos económicos que nuestro trabajo ha elegido enfatizar.

En una plataforma Moodle, por ejemplo, queda registrada cada interacción de los estudiantes con el sistema, algo que no tiene un correlato en la educación presencial y que puede aprovecharse para predecir y prevenir deserción. Esos datos se generan a lo largo de cada cursada, desde la matriculación hasta la aprobación, y son accesibles tanto para los sectores administrativos como para el equipo docente.

Con un adecuado procesamiento, pueden brindar no solamente conocimiento accionable para mejorar un curso en marcha, sino para futuras cursadas (Casey y Azcona, 2017).

Un aporte reciente en este sentido es el de Evangelista (2019), el cual usa datos históricos de cursos en Moodle para desarrollar modelos que predicen deserción. Esta investigación considera atributos como tiempo de visionado de materiales, vista de recursos, exámenes realizados, respuestas en foros de discusión y vistas realizadas en fines de semana. Utiliza algoritmos como *J48*, *Random Forest*, *JRip* y *OneR* y encuentra que algunas interacciones, como actividades completadas, arrojan mayor valor predictivo que otras. Por otra parte, una revisión sistemática publicada por Félix, Ambrósio, Lima y Brancher (2018) cita numerosos artículos datados entre 2009 y 2016 que usan minería de datos para predecir no solamente deserción sino, en general, rendimiento, de estudiantes en plataformas Moodle, usando los registros del sistema. Previamente, Félix, Ambrósio, Neves, Siqueira y Brancher (2017), con una herramienta llamada Moodle Predicta, generaron reportes basados en el análisis predictivo de las interacciones con la plataforma y el algoritmo de clasificación “Bayes ingenuo”.

Otro estudio, de Buschetto Macarini, Cechinel, Batista Machado, Faria Culmant Ramos y Munoz (2019), comparó diferentes interacciones ocurridas en un *Learning Management System* (LMS), como Moodle, y las clasificó en cognitivas, sociales y docentes. Es decir, se usó el clásico esquema de Garrison, Anderson y Archer (1999). El objetivo fue analizar si el solo uso de registros de interacciones en una plataforma permitía predecir el fracaso estudiantil. Mediante diversos algoritmos, concluyeron que sí. Además, observaron que buena parte de los estudios existentes no detectaban estudiantes en riesgo sino hasta la mitad del curso y subrayaron que ellos pudieron hacerlo con base en las primeras ocho de un total de quince semanas. Logros análogos tuvieron Gray y Perkins (2019) en su estudio sobre predicción temprana de desempeño estudiantil en cursos virtuales usando aprendizaje automático y registros del LMS.

Por otra parte, Murray, Pérez, Geist y Hedrick (2013) mostraron que los estudiantes con mayores tasas de acceso a los materiales de un LMS obtenían mejores notas. Jokhan, Sharma y Singh (2019) utilizaron datos obtenidos de un curso semipresencial, mediante un *plugin* de captura en Moodle y un modelo de regresión. Concluyeron que el promedio de ingresos semanales al sistema y el de tareas completadas eran buenos predictores del desempeño general. Cohen (2017) también utilizó solamente registros de una plataforma y encontró que la actividad de los estudiantes podía ayudar a predecir a aquellos en riesgo. Usman, Salisu, Barroon y Yusuf (2019) analizaron interacciones con árboles de decisión, Bayes ingenuo y *k-nearest neighbors* (k-NN). Observaron que los primeros tenían en su caso mayor poder predictivo del desempeño. En cambio, Romero, Zafra, Luna y Ventura (2013) aplicaron 21 clasificadores a la actividad en Moodle de estudiantes de siete cursos para estudiar el rendimiento de distintas técnicas y concluyeron que no existe una mejor que el resto en todos los casos.

Estudios como los anteriormente mencionados muestran que se puede intentar predecir desempeño y, en particular, deserción aplicando algoritmos de aprendizaje automático al registro de las distintas interacciones con la plataforma virtual. Nuestro trabajo intenta, además, optimizar los modelos priorizando métricas de impacto económico del abandono.

En general, aunque parte de la bibliografía consultada menciona el efecto económico de deserción para los estudiantes, no encontramos en los análisis basados en algoritmos que pudimos rastrear una validación de su valor basada en métricas económicas específicas. Sin embargo, como en ciencia de datos ese enfoque forma parte habitual de los objetivos, nos pareció interesante aportar esa mirada, sin desconocer que los perjuicios de la deserción exceden este aspecto.

Contexto institucional

La Facultad Regional Buenos Aires (UTN.BA) es una de las 29 que conforman la Universidad Tecnológica Nacional (UTN) en la Argentina. Dicta nueve carreras de ingeniería, tres licenciaturas, siete tecnicaturas y veinte carreras de posgrado. Asimismo, su Secretaría de Cultura y Extensión Universitaria (SCEU) brinda una amplia oferta de actividades de formación profesional extracurriculares. Las tecnicaturas (formaciones técnicas breves), diplomaturas y posgrados de la SCEU otorgan certificados de aprobación emitidos por la Universidad y, en general, son propuestas de pago destinadas a profesionales graduados de diversas carreras.

Los cursos que la SCEU brinda mediante su Centro de *e-learning* se dictan mediante una modalidad colaborativa, constructivista y dialógica en una plataforma Moodle. Eso implica, como indica la página web institucional, la implementación de contenidos teórico-prácticos digitalizados en línea, materiales de lectura asociados a dichos contenidos, foros temáticos proactivos de socialización y discusión, así como evaluaciones periódicas de los temas y contenidos.

Los cursos virtuales de la ACEU, en general, están organizados en módulos de como mínimo un mes. La modalidad es básicamente asincrónica pero al menos una clase por módulo se dicta en forma sincrónica. Esta no es de asistencia obligatoria y queda grabada para quienes deseen verla en otro momento. Otros espacios de comunicación son los foros. Tanto en ellos como en las clases sincrónicas, el diálogo entre docentes y estudiantes es real, aunque mediado, en términos de la categorización de García Aretio (2011). Hay también instancias de diálogo simulado como, por ejemplo, en ciertos videos introductorios. Asimismo, no existen instancias presenciales.

MATERIALES Y MÉTODOS

Los conjuntos de datos usados en este trabajo provienen de cuatro cursos virtuales dictados por la SCEU UTN.BA durante 2018 y 2019. Se eligieron específicamente los

consignados en la tabla 1 por una cuestión de posibilidad institucional de acceso a la información. Sus duraciones son diferentes y atraen a personas de los más diversos orígenes profesionales. En conjunto, se registraron en la plataforma 591 individuos. Sin embargo, el total de matriculaciones en cursos por parte de estudiantes fue de 654, pues algunos estudiantes tomaron más de una de las capacitaciones consideradas en este trabajo. En esa cantidad de matriculaciones hubo 141 deserciones. En el caso de cursar más de uno, hay estudiantes que pudieron abandonar un curso y no otro.

Tabla 1. Cursos, participantes y desertores

Curso	Sigla	Perfil	Semanas	V	Matriculaciones	Deserciones	% Deserción
Diplomatura en Bases de Datos	DBD	Profesionales que quieran conocer fundamentos de Bases de Datos	12	1	40	1	2,5
				2	26	10	38,5
				3	28	9	32,1
				4	18	7	38,9
				5	30	8	26,7
Diplomatura en Inteligencia de Negocios	DBI	Profesionales que deseen usar datos para facilitar decisiones organizacionales	12	20	43	11	25,6
				21	44	14	31,8
				22	31	2	6,5
				23	33	8	24,2
				24	37	8	21,6
Posgrado en Gestión de Proyectos	PGP	Profesionales con título de grado que deseen conocer los principios del PMI	16	6	56	11	19,6
				7	45	10	22,2
				8	39	6	15,4
				9	73	17	23,3
Científico de Datos con R	R	Profesionales que quieran profundizar en R para Business Intelligence	6	11	23	0	0
				12	33	7	21,2
				13	27	9	33,3
				14	28	3	10,7

Fuente: Elaboración propia

Para considerar o no a alguien desertor, se eligieron los siguientes criterios:

- Quienes aprobaron el examen final en línea no fueron considerados desertores, sin importar su nivel previo de participación. Además, se les dio certificado de aprobación.

- Tampoco se consideró desertores a los que siguieron participando del curso sin presentar o sin aprobar el final. A estos, se les otorgó un certificado de participación, aunque no de aprobación.
- Quienes no aprobaron el final ni interactuaron con la plataforma en las últimas dos semanas de cursada sí se consideraron desertores y no se les dio ningún certificado, ya que ese es el tiempo que tenían para rendir dicho examen.

Nos enfocamos específicamente en ver si era posible hallar modelos algorítmicos para predecir si alguien se convertiría en desertor según los criterios arriba descritos en función de sus interacciones con la plataforma Moodle y evaluar la calidad de dichos modelos en términos de beneficio económico. Además, limitamos el análisis a los primeros 14 días porque, antes de dicho plazo, los estudiantes pueden pedir devolución del pago o cambio de curso. La idea era testear distintos algoritmos, ya que no se podía saber a priori cuál ajustaría mejor los datos.

En total, para los cuatro cursos y todas sus versiones, había en los registros de Moodle 314.783 interacciones. De esas, 8.754 se descartaron porque no figuraba información sobre el causante de la interacción. Quedaron entonces 306.029, de las cuales 291.132 correspondían a estudiantes. El resto eran de docentes y administrativos, por lo que no eran relevantes para este estudio. Estas interacciones de estudiantes con la plataforma fueron clasificadas en los 12 tipos consignados en la tabla 2.

Tabla 2. Tipos y cantidades de interacciones entre estudiantes y la plataforma Moodle

Interacción	Casos	Definición
Administrativa	354	Tarea de carácter administrativo, por ejemplo, cambio de clave
Búsqueda	65	Buscar algo dentro de la plataforma del curso
Clase	10.199	Asistir a una clase sincrónica o reproducirla grabada
Comentario	1.415	Explicaciones o comentarios presentados al adjuntar trabajos prácticos
Descarga	605	Descargar materiales del curso
Entrega	2.811	Subir trabajos prácticos para su evaluación
Examen	72.081	Realizar exámenes en línea
Foro	556	Suscribirse o desuscribirse a un foro no obligatorio
Grupo	136	Interactuar dentro del foro para estudiantes que forman un equipo dentro de un curso
Informe	3.881	Descargar un reporte de Moodle, por ejemplo calificaciones
Publicar	1.720	Hacer comentarios en algún foro
Vista	212.206	Leer online materiales del curso

Fuente: Elaboración propia

En realidad, en el Moodle de la UTN-BA, a cada interacción se le asigna en los registros un tipo entre 60 posibles para el paquete de cursos aquí considerados. Sin embargo, para los fines de este estudio, no tenía sentido considerar distintos a algunos de los mencionados tipos. Por ejemplo, figuran tipos como “Vista de la Unidad 1 del PGP6”, “Vista de la Unidad 3 de R11” o alusiones a un foro en particular. Al consolidar bajo una sola denominación (por ejemplo, “Vista” y “Foro”) y omitir referencias específicas, solo quedaron 12 tipos.

Para poder operar sobre el flujo continuo de interacciones con la plataforma a lo largo del tiempo, hacía falta dividirlo en intervalos discretos. Luego de probar varios, se verificó que dividir los primeros 14 días de cada cursada en siete períodos de dos días producía una mejor predicción. Denominamos “atributo”, para cada combinación de alumno y curso, a la cantidad de veces que se produjo una interacción de cada uno de los 12 tipos para cada estudiante y versión de cada curso en un período de dos días. Se generó para cada atributo un nombre compuesto por un prefijo con una letra y un número referentes al período y un sufijo que da cuenta del tipo correspondiente. Por ejemplo, el atributo “A2dasvista” refleja la cantidad de vistas que hizo un estudiante durante los primeros dos días del inicio del curso.

Se creó una tabla de atributos que describe la participación de cada estudiante en las dos primeras semanas de cursada, con un registro por participante, curso y versión. Cada registro contenía un identificador del participante, el curso, la versión, los valores de los atributos y si desertó o no¹.

El valor de algunos atributos resultó nulo para todos los estudiantes de todos los cursos, por lo que no aportaban valor predictivo y fueron descartados. Corresponden a situaciones inexistentes, como entregar un examen el primer día. Otros presentan valores pequeños que podrían considerarse estadísticamente no significativos y conducir a un sobreajuste. Para evitarlo, se dividieron los datos aleatoriamente en una parte para entrenar (70%) y otra para medir la bondad del ajuste (30%). Se consideró esta opción mejor que desechar los tipos de interacciones poco poblados, porque aún pequeñas cantidades de ciertas interacciones podrían tener valor predictivo y no se quería omitirlas a priori. Por ejemplo, podría ser que alguien nunca hubiera descargado materiales y, justo antes de desertar, hubiera decidido bajar todos para conservarlos.

Por otra parte, no tiene sentido para todo algoritmo hablar de la relevancia predictiva de cada atributo en particular. Por ejemplo, en una red neuronal multicapa no resulta evidente en qué medida los diferentes pesos de las diversas capas filtran el efecto de un atributo determinado sobre la predicción de deserción. Sin embargo, para un árbol de decisión se puede extraer sin dificultad la importancia relativa de los atributos para la predicción. Por ejemplo, si con los datos consignados en nuestra tabla de atributos entrenamos un árbol de decisión mediante el paquete *R rpart*, se puede estimar la relevancia predictiva de cada uno, como se enumera en la tabla 3.

Tabla 3. 10 atributos con mayor relevancia predictiva

Predictor	Importancia
A2dasvista	4,1
C6dasvista	3,0
F12dasclase	2,0
E10dasvista	2,0
B4dasvista	1,7
F12dasvista	1,6
D8dasvista	1,3
G14dasclase	1,3
D8dasexamen	1,1
B4dasclase	1,0

Fuente: Elaboración propia

De esta manera, se observa que los atributos más relevantes se refieren a vista de materiales y a participación en las clases sincrónicas, pero la cantidad de participaciones en clase representa solo un 3% de los eventos y, sin embargo, resulta relevante para la predicción. Este ejemplo muestra el sentido de evitar excluir los atributos con bajo conteo de eventos.

La definición de impacto

En estadística, se considera que una predicción es positiva cuando afirma que se encontrará el resultado que se desea identificar (en nuestro caso, la deserción). Pero esa predicción positiva podría ser verdadera o falsa, según se verifique o no en la realidad lo que dice. A la inversa, una predicción es negativa si sostiene que se producirá un resultado diferente (en nuestro caso, la permanencia en el curso). Y esa afirmación, igualmente, podría ser verdadera o falsa, según se verifique o no en los hechos posteriores. Estas cuatro situaciones se grafican en la tabla 4. Claramente, predicción positiva o negativa no equivale a deseable o indeseable, ya que la deserción no es deseable.

Tabla 4. Definición de falsos y verdaderos positivos o negativos

	El modelo predice deserción	El modelo NO predice deserción
El alumno deserta	Verdadero positivo (TP)	Falso negativo (FN)
El alumno NO deserta	Falso positivo (FP)	Verdadero negativo (TN)

Fuente: Elaboración propia

A partir de los datos de entrenamiento se construyeron los distintos modelos para cada algoritmo. Una vez obtenidos los modelos, se los aplicó al conjunto de datos de prueba, se contó la cantidad de FP, TP, FN y TN y se calcularon estas métricas tradicionales:

$$\text{Sensibilidad} = \frac{TP}{TP + FN}$$

$$\text{Especificidad} = \frac{TN}{TN + FP}$$

$$\text{Precisión} = \text{Valor Predictivo Positivo} = \frac{TP}{TP + FP}$$

Sin embargo, los FP, TP, FN y TN no tienen el mismo impacto en la institución y en los participantes. Era más eficiente, entonces, aquel algoritmo que predijese mejor el más grave de los cuatro.

A continuación, había que pensar en términos de qué criterio resultaba operativamente más conveniente medir dicha gravedad. Se decidió considerar peor aquella alternativa que generara mayor impacto desde el punto de vista económico. Por supuesto, sin desconocer que la deserción afecta negativamente también al estudiante a nivel psicológico, a la reputación de la universidad y a la comunidad que se pierde de contar con un profesional más preparado, entre otros perjuicios. Sin embargo, aunque estas variables pueden intentar medirse, son más subjetivas e intangibles que la cantidad de dinero invertido y perdido o la cantidad de dinero recuperado al evitar la deserción.

Se decidió estimar el valor de un modelo como el dinero recuperado por su aplicación sobre el total de la pérdida por deserciones que igual tendrían lugar si el modelo fuera perfecto, definido de la forma siguiente:

$$\text{Impacto} = \frac{\text{Dinero recuperado}}{\text{Pérdida por deserciones}}$$

Donde:

Dinero recuperado = *Precio* × *Retenciones exitosas* - *Costo de acciones de retención*

“Precio” se refiere al precio de cada curso, mientras que las “retenciones exitosas” son todos los verdaderos positivos, multiplicados por la tasa de éxito de las acciones de retención:

$$\text{Retenciones exitosas} = TP * \text{Tasa de éxito de las acciones de retención}$$

El costo de las acciones de retención es la cantidad de positivos que produce el modelo, multiplicada por el costo de cada acción de retención:

$$\text{Costo de las acciones de retención} = (TP + FP) * \text{costo de una acción de retención}$$

En un modelo perfecto, todos los casos de deserción serían identificados y así se podría ejercer sobre ellos acciones de retención. En nuestro modelo, esa cantidad de casos se reparte entre TP y FN. Por lo tanto, en el modelo perfecto:

$$\text{Pérdida MP} = (TP + FN) * (\text{Precio} * \text{tasa de retención} - \text{costo de retención})$$

“Pérdida MP” representa las pérdidas que de igual manera habría, aún con un modelo perfecto.

Si se reúnen todas estas ecuaciones, queda la siguiente fórmula:

$$\text{Impacto} = \frac{TP \times vTP + FP \times vFP}{(TP + FN) \times vTP}$$

Donde:

$$vTP = \text{Precio} * \text{tasa de retención} - \text{costo de una acción de retención}$$

$$vFP = -\text{costo de una acción de retención}$$

Si bien los valores de los cursos han ido cambiando a lo largo del tiempo, sus precios relativos han permanecido más o menos constantes de acuerdo a lo indicado en la tabla 5. Las variaciones en la cantidad de inscriptos se deben a acciones diferentes del equipo comercial, a factores estacionales y a circunstancias singulares. Por ejemplo, el inicio del curso DBDv1 se postergó varios meses hasta reunir suficientes interesados.

Tabla 5. Precio de los distintos cursos

Curso	Precio
Diplomatura en Bases de Datos	AR\$12.000
Diplomatura en Inteligencia de Negocios	AR\$15.000
Posgrado en Gestión de Proyectos	AR\$30.000
Data Scientist con R	AR\$10.000

Fuente: Elaboración propia

Se podría objetar que no se cuenta en el presente con valores experimentales reales para la tasa de retención y el costo de la acción de retención. Por ello, se decidió partir de la efectividad y costo de acciones de retención ejecutadas en circunstancias similares del pasado, con la intención de recurrir a un proceso de aproximaciones sucesivas para ir ajustando modelos y acciones como el ilustrado en la figura 1.

Algoritmos y construcción de modelos

Existen múltiples algoritmos predictivos y es difícil saber a priori cuál presentará el mejor comportamiento para un problema concreto. En esta ocasión, se trabajó con los siguientes del paquete *CORELearn* (Robnik-Sikonja, Savicky y Robnik-Sikonja, 2018): *Random Forest*, *k-NN*, *Random Forest* con *k-NN* en las hojas con el modelo *rfNear* y *k-NN* con Bayes ingenuo. Se usó el paquete *rpart* con los métodos Anova, Class y Poisson para algoritmos de árboles de decisión (Therneau, Atkinson y Ripley, 2015). Se utilizó el paquete *NeuralNet* (Fritsch, Guenther y Guenther, 2019) con los modelos *rprop+* y *rprop-* para algoritmos de redes neuronales. En el caso de los árboles de decisión se varió sistemáticamente el parámetro de complejidad para buscar el máximo impacto.

Para las redes neuronales se utilizó un perceptrón multicapa con 10 neuronas ocultas en una única capa entre la de entrada y la de salida. Se buscó optimizar la respuesta variando el esquema de entrenamiento (*rprop+* y *rprop-*), así como la cantidad de neuronas en la capa oculta. Para explorar hasta qué punto las soluciones obtenidas eran estables y reproducibles, se ejecutaron cinco veces los correspondientes entrenamientos. Se calculó el impacto promedio sobre esas cinco corridas y su correspondiente desvío estándar. Como reflejo de la variación del desvío comparado con el promedio, se calculó la razón entre el desvío y el promedio y el desvío de los promedios calculados.

En vez de asumir las predicciones categoriales que algunos algoritmos producen, se tomó en todos los casos la respuesta continua, interpretada como una probabilidad, y se ajustó un umbral que optimizara el impacto. Para hacerlo, se tomó el intervalo entre la mínima y la máxima respuesta de cada modelo para todos los casos del conjunto de prueba y se partió en 100 pasos, en los cuales se colocó un umbral.

Todos los participantes para los cuales el modelo predijo una probabilidad de deserción mayor o igual al umbral fueron considerados posibles desertores. Todos aquellos cuya probabilidad de deserción quedó por debajo fueron considerados no necesitados de acciones de retención.

En principio, no se sabía cuál era el umbral que para un modelo dado brindara el mayor impacto. Pero si el umbral tiende al mínimo, la acción de retención debería ser aplicada sobre la totalidad de los participantes. Si tendiese al máximo, en cambio, no se la aplicaría a ninguno y se perderían los beneficios de retenerlos y de haber intentado predecir su deserción. Para buscar el umbral óptimo, entonces, se calculó el impacto de retención del modelo en pasos de un 1% desde la mínima respuesta continua hasta la máxima. Así se encontró en cada situación el umbral cuyo impacto fuera el mayor posible, se optimizó el resultado de cada modelo y se pudieron comparar los generados por los diferentes algoritmos.

RESULTADOS

Predicciones alcanzadas

Se presentan a continuación las mejores predicciones para cada curso en términos de sensibilidad-especificidad, precisión (también llamada valor predictivo positivo) e impacto.

Sensibilidad-Especificidad

Todos los mejores modelos de los cuatro cursos por separado presentaron una sensibilidad del 100%. Las especificidades fueron relativamente bajas: PGP, 26,7%; DBI, 41,9%; DBD, 27,6% y 64% para R. Esto se sigue de haber optimizado el impacto, dado que un falso positivo se consideró mucho menos gravoso que un falso negativo.

Podría parecer que una sensibilidad del 100% en todos los casos implica algún tipo de sesgo o proceso de selección de la información. Pero eso se debe a que el umbral que optimizaba el impacto sobre los conjuntos de entrenamiento fue lo suficientemente bajo como para garantizar la ausencia de falsos negativos sobre los conjuntos de prueba. Si no hay falsos negativos, la sensibilidad debe ser necesariamente del 100%. Esto se logró a costa de tener una muy baja precisión.

Precisión

La precisión medida para PGP fue de 27%. Para DBI se logró un 31%. Para DBD, fue de 28%. Para R, resultó de 25%.

Impacto

La métrica que llamamos impacto, considerada la más apropiada para decidir si el modelo predictivo agregaba valor, dio 97% para PGP; 94% para DBI; 92% para DBD y 88% para R.

Algoritmo óptimo por curso

Otra de las preguntas de investigación era si el mismo algoritmo servía para ajustar todos los cursos. Sin embargo, es interesante resaltar que, en tres de los cuatro, el mejor modelo se consiguió utilizando una red neuronal. Los mejores modelos hallados para cada curso se sintetizan en la tabla 6.

Tabla 6. Mejores modelos hallados para cada curso

Curso	Algoritmo con el que se entrenó el mejor modelo	Impacto (%)	Precisión (%)
DBI	k-NN	94	31
DBD	Redes Neuronales	92	28
PGP	Redes Neuronales	97	27
R	Redes Neuronales	88	25

Fuente: Elaboración propia

Sin embargo, la mejor red neuronal obtenida para DBI consiguió un impacto de 92% con una precisión de 25%, lo que quedó apenas debajo del mejor resultado encontrado, o sea, el correspondiente a k-NN.

En todos los casos, el desvío estándar dividido por el promedio del impacto en cada conjunto de cinco corridas idénticas fue menor al aumentar el número de neuronas ocultas, como se indica en la tabla 7.

Tabla 7. Variabilidad del Impacto vs. cantidad de neuronas en la capa oculta

Neuronas ocultas	Desvío / Promedio (%)	Promedio (%)
10	4	4
5	4	2
2	40	13
1	57	22

Fuente: Elaboración propia

Cabía preguntarse si incluir entre los predictores a las diferentes duraciones y precios hubiera podido mejorar el ajuste del modelo general y hacerlo superar a los modelos por curso. Para explorar esta posibilidad se recurrió al algoritmo *rpart* para conocer la importancia de los predictores después de incluir precio y duración. Se verificó que ni el precio ni la duración aparecían entre los 10 más importantes, y se deja para futuros trabajos explorar en mayor detalle esta posibilidad.

La variabilidad de los impactos fue pequeña para 5 y 10 neuronas ocultas y se incrementó con menos neuronas ocultas. Por otra parte, el impacto del modelo resultó dependiente de la cantidad de neuronas ocultas, como muestra la tabla 8.

Tabla 8. Impacto vs. cantidad de neuronas en la capa oculta

Neuronas ocultas	Impacto promedio (%)	Desvío estándar (%)
10	88	4
5	87	8
2	65	7
1	47	12

Fuente: Elaboración propia

Modelo general vs. modelos individuales

Para evaluar si el modelo general era mejor o peor que los individuales, se compararon con las tres métricas planteadas: sensibilidad-especificidad, precisión e impacto.

Sensibilidad-Especificidad

Todos los mejores modelos consiguieron para todos los cursos una sensibilidad del 100%. Consecuentemente, la comparación se apoyó en la especificidad. Por separado dio: PGP, 26,7; DBI, 41,9%; DBD, 27,6% y 64% para R. Al calcular especificidad con el modelo ajustado con todos los cursos, dio solo 18%. Por lo tanto, era el menos específico.

Precisión

La precisión o valor predictivo positivo, tampoco consiguió una ventaja decisiva para el ajuste en conjunto comparado con el de modelos separados para los distintos cursos. En DBI fue de 100% y en R, de 33%. Ambos obtuvieron valores mejores que la precisión correspondiente al modelo del conjunto de los cursos, que fue de 28%. DBD logró una precisión de 28% y PGP de 27%.

Impacto

En el caso de cada curso considerado en particular, se obtuvo un impacto de 97% para PGP; 94% para DBI; 92% para DBD y 88% para R. Al considerar todos, el impacto fue de 91%. Este 91% era un promedio entre todos los cursos y convenía comparar los impactos de los modelos individuales con el que tendría el modelo general al aplicarlo curso a curso. Así, se obtuvo 96% para PGP; 97% para DBI; 86% para DBD y 55% para R. Como se puede observar, para DBI se consiguió un mejor ajuste con el modelo general que con el particular. Para los demás cursos sucedió lo contrario. El que más impacto perdió al aplicar el modelo general fue el de R, lo cual es compatible con el hecho de que cuesta y dura menos que los demás.

En el caso de los modelos individuales, no tenía sentido incluir como predictores precios, duraciones y existencia de actividades grupales, ya que son los mismos para todas las versiones. Sin embargo, sí tenía sentido incluirlos al construir un modelo conjunto.

DISCUSIÓN Y CONCLUSIONES

Este trabajo confirma que es posible generar modelos para predecir la deserción en cursos de *e-learning* utilizando solamente el registro de interacciones entre los estudiantes y la plataforma. En particular, lo logra con datos de las primeras dos semanas de cursos que duran entre 6 y 16 semanas, lo cual es extremadamente temprano y permite la implementación de acciones de retención.

Por otra parte, en lugar de medir el aporte de cada modelo solo a través de métricas usuales como sensibilidad-especificidad y precisión, se construyó otra – el impacto – orientada a optimizar el resultado deseado en términos de evitación de pérdidas económicas, tanto para el estudiante como para la institución. De este modo, la relevancia del análisis no fue solo teórica, sino que intentó contribuir al mejor interés de la comunidad de aprendizaje. Se optimizó cada modelo mediante la aplicación de un umbral variable para decidir si predice una deserción o no. Al ajustar el umbral, se halló el punto de impacto máximo.

Los impactos de los modelos predictivos creados estuvieron en la zona de 90%. El algoritmo que produjo modelos más eficientes fue una red neuronal entrenada con retropropagación. Solo en un caso, con k-NN, el impacto fue marginalmente mejor que con una red neuronal (94% vs. 92%). Las redes neuronales generaron los mejores modelos con diez neuronas ocultas y se verificó la estabilidad de esa solución.

Se construyó un modelo único para los cuatro cursos analizados pero, en general, fue inferior a los modelos hechos por separado. Al entrenar modelos independientes para cada curso se obtuvo un mejor impacto que para el ajuste colectivo.

Quedaría por ver si los modelos construidos seguirán siendo aplicables en futuras versiones. Se espera realizar este análisis cuando esa información esté disponible.

También sería conveniente generar nuevos modelos predictivos para otros cursos. La intención es refinarlos a medida que se vayan incorporando datos concretos sobre el costo y la probabilidad de éxito de las acciones de retención llevadas a cabo. Se pretende ensayar distintos tipos de acciones y medir sistemáticamente su costo y probabilidad de éxito, con el fin de optimizar la métrica de impacto propuesta y ejecutar el programa recursivo ilustrado en la figura 1.

Además, se podría estudiar si otros algoritmos permiten obtener modelos de aún mejor impacto, en particular los de regresión logística, máquinas de soporte vectorial y reglas de asociación. También sería interesante aplicar minería de texto al contenido de los foros de los cursos, como en Sorour, Mine, Goda y Hirokawa (2015), entre otros. Asimismo, se podrían utilizar técnicas de análisis de redes sociales, como proponen Muñoz, Delgado, Rubio, Grilo y Basto-Fernandes (2017), para explorar si los estudiantes integrados en grupos –como en el caso del PGP– son menos proclives a la deserción.

NOTAS

- ¹ Aquí pueden consultarse las tablas de atributos y de resultados de los modelos, así como el código R ejecutado: https://drive.google.com/drive/folders/1REnt_GF8yOn1Lcs6AXTL82KqFUJZ8q-w?usp=sharing

REFERENCIAS

- Baker, R. S., Lindrum, D., Lindrum, M. J., y Perkowski, D. (2015). Analyzing Early At-Risk Factors in Higher Education e-Learning Courses. *International Educational Data Mining Society*.
- Buschetto Macarini, L. A., Cechinel, C., Batista Machado, M. F., Faria Culmant Ramos, V., y Munoz, R. (2019). Predicting Students Success in Blended Learning-Evaluating Different Interactions Inside Learning Management Systems. *Applied Sciences*, 9(24), 5523. <https://doi.org/10.3390/app9245523>
- Casey, K., y Azcona, D. (2017). Utilizing student activity patterns to predict performance. *International Journal of Educational Technology in Higher Education*, 14(1), 23. <https://doi.org/10.1186/s41239-017-0044-3>
- Centro de e-Learning de la SCEU-FRBA-UTN (2019). *Pautas generales para el cursado de cursos, diplomaturas, expertos y carreras online*. Recuperado de <http://www.sceu.frba.utn.edu.ar/e-learning/quienes-somos/309.html>
- Cohen, A. (2017). Analysis of student activity in web-supported courses as a tool for predicting dropout. *Educational Technology Research and Development*, 65(5), 1285-1304.
- Evangelista, E. D. (2019). Development of Machine Learning Models using Study Behavior Predictors of Students' Academic Performance through Moodle. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(6S3).
- Félix, I. M., Ambrósio, A. P., Neves, P. S., Siqueira, J., y Brancher, J. D. (2017). Moodle Predicta: A Data Mining Tool for Student Follow Up. *Proceedings of the 9th*

- International Conference on Computer Supported Education (CSEDU)*, 339-346.
- Félix, I. M., Ambrósio, A., Lima, P., y Brancher, J. D. (2018). Data Mining for Student Outcome Prediction on Moodle: a systematic mapping. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 29(1), 1393. <http://dx.doi.org/10.5753/cbie.sbie.2018.1393>
- Fritsch, S., Guenther, F., y Guenther, M. F. (2019). Package 'neuralnet'. *Training of Neural Networks*. Recuperado de <https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>
- García Aretio, L. (2011). Perspectivas teóricas de la educación a distancia y virtual. *Revista Española de Pedagogía*, 69(249), 255-271.
- García Saiz, D. (2016). *Minería de datos aplicada a la enseñanza virtual: nuevas propuestas para la construcción de modelos y su integración en un entorno amigable para el usuario no experto*. Tesis Doctoral, Universidad de Cantabria, Departamento de Ingeniería Informática y Electrónica, Cantabria.
- Garrison, D. R., Anderson, T., y Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The internet and higher education*, 2(2-3), 87-105.
- Gray, C. C., y Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education*, 131, 22-32.
- Jokhan, A., Sharma, B., y Singh, S. (2019). Early warning system as a predictor for student performance in higher education blended courses. *Studies in Higher Education*, 44(11), 1900-1911.
- Muñoz, A., Delgado, R., Rubio, E., Grilo, C., y Basto-Fernandes, V. (2017). Forum participation plugin for Moodle: Development and Discussion. *Procedia Computer Science*, 121, 982-989. <https://doi.org/10.1016/j.procs.2017.11.127>
- Murray, M., Pérez, J., Geist, D., y Hedrick, A. (2013, Julio). Student interaction with content in online and hybrid courses: Leading horses to the proverbial water. *Proceedings of the Informing Science and Information Technology Education Conference* (pp. 99-115). Informing Science Institute.
- Robnik-Sikonja, M., Savicky, P., y Robnik-Sikonja, M. M. (2018). Package 'CORElearn'. Recuperado de <https://cran.r-project.org/web/packages/CORElearn/CORElearn.pdf>
- Romero, C., Zafra, A., Luna, J. M., y Ventura, S. (2013). Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data. *Expert Systems*, 30(2), 162-172.
- Therneau, T., Atkinson, B., y Ripley, B. (2015). Package 'rpart'. Recuperado de <https://cran.r-project.org/web/packages/rpart/rpart.pdf>
- Kara, M., Erdoğan, F., Kokoç, M., y Cagiltay, K. (2019). Challenges Faced by Adult Learners in Online Distance Education: A Literature Review. *Open Praxis*, 11(1), 5. <https://doi.org/10.5944/openpraxis.11.1.929>
- Sorour, S. E., Mine, T., Goda, K., y Hirokawa, S. (2015). A Predictive Model to Evaluate Student Performance. *JIP*, 23, 192-201.
- Tinto, V. (2017 [Publicado online inicialmente el 11 de diciembre de 2015]). Through the Eyes of Students. *Journal of College Student Retention: Research, Theory & Practice*, 19(3), 254-269. <https://doi.org/10.1177/1521025115621917>
- Usman, U. I., Salisu, A., Barroon, A. I. E., y Yusuf, A. (2019). A Comparative Study of Base Classifiers in Predicting Students' Performance based on Interaction with LMS Platform. *FUDMA Journal of Sciences*. 3(1), 231-239.

PERFIL ACADÉMICO Y PROFESIONAL DE LOS AUTORES

Ignacio Urteaga. Profesor titular en ciencia de datos, análisis de negocios y gestión de proyectos en la Secretaría de Cultura y Extensión Universitaria de la Universidad Tecnológica Nacional (Argentina). Catedrático e investigador jefe en la Universidad del Salvador (Argentina), Maestría en Dirección de Sistemas de Información. Exgerente de investigación de datos en diversas empresas. MBA, PMP, ITILp y físico. Sus líneas de investigación son en ciencia de datos aplicada a problemáticas empresariales, salud y educación.

E-mail: jesus.urteaga@usal.edu.ar

Laura Siri. Comunicóloga graduada de la Universidad de Buenos Aires. Docente de Ciencias de la Comunicación en dicha universidad, en el área de políticas tecnológicas, y del área de Cultura y Extensión Universitaria de la Universidad Tecnológica Nacional de la República Argentina. Miembro desde 1994 de diversos equipos de investigación. Periodista y editora especializada en informática. Sus líneas de investigación principales son en el campo de estudios en Ciencia, Tecnología y Sociedad (CTS).

E-mail: lsiri@sociales.uba.ar

Guillermo Garófalo. Profesor de Historia e integrante del equipo docente del área de Cultura y Extensión Universitaria de la Universidad Tecnológica Nacional de la República Argentina. Su principal área de interés es gestión de proyectos educativos.

E-mail: guillermogarofalo@outlook.com

Dirección:

Secretaría de Cultura y Extensión Universitaria

Universidad Tecnológica Nacional

Medrano 951, 2º piso, Of. 206

Ciudad Autónoma de Buenos Aires (Argentina)

Fecha de recepción del artículo: 12/01/2020

Fecha de aceptación del artículo: 07/02/2020

Fecha de aprobación para maquetación: 27/03/2020