

Predicción de riesgo académico usando notas, asistencia a clases y clics en el LMS

Academic risk prediction using grades, class attendance, and clicks in the LMS

Gabriel Zúñiga Varela¹, Carles Lindín²,
 Teresa Sancho Vinuesa¹

¹ Universitat Oberta de Catalunya, España

² Universitat de Barcelona, España

gzunigav@uoc.edu , carles.lindin@ub.edu , tsancho@uoc.edu

RESUMEN. Pocos estudios con Analítica del Aprendizaje han intentado predecir los resultados de conjunto de un año académico. Esta investigación desarrolló un modelo predictivo del riesgo de suspender el primer año en un Grado de Negocios (i.e., obtener menos créditos de los necesarios para aprobar), utilizando Regresión Logística con datos de dos cohortes de estudiantes (n=1046). El modelo utiliza la tasa de asistencia, calificaciones de evaluación continua de tres asignaturas y los clics en la LMS, del primer semestre del año. Se probó con 74 estudiantes de una cohorte diferente, obteniendo una precisión alta, pero una sensibilidad baja. Los resultados sugieren que la asistencia a clases, las habilidades de comunicación y la competencia numérica son transversales al éxito académico. El artículo revela una relación no lineal entre la actividad en la LMS y la media académica y propone un método para tratarlo.

ABSTRACT. Few studies with Learning Analytics have attempted to predict the overall outcomes of an academic year. This research developed a predictive model of the risk of failing the first year in a Business Degree (i.e.; earning fewer credits than needed to pass), using Logistic Regression with data from two cohorts of students (n=1046). The model uses attendance rate, continuous assessment grades for three subjects, and clicks on the LMS, from the first semester of the year. It was tested with 74 students from a different cohort, obtaining high precision but low sensitivity. The results suggest that class attendance, communication skills, and numerical competence are cross-cutting to academic success. The paper reveals a nonlinear relationship between LMS activity and academic mean, and proposes a method to address it.

PALABRAS CLAVE: Analítica del aprendizaje, Modelo predictivo, Educación superior, Sistema de gestión del aprendizaje, Riesgo académico.

KEYWORDS: Learning analytics, Predictive model, Higher education, Learning management system, Academic risk.

1. Introducción

Las tecnologías digitales han favorecido la emergencia de cantidades ingentes de datos de todo tipo en entornos educativos. Entre ellos destacan las huellas digitales que dejan los estudiantes al usar sistemas de gestión del aprendizaje (Learning Management Systems, LMS). Al mismo tiempo, las tecnologías digitales han facilitado el trabajo de estructurar, integrar, analizar y sacar provecho a dichos datos. La Analítica del Aprendizaje (Learning Analytics, LA) se ha desarrollado en la última década como una disciplina que gira en torno a la explotación de estos datos con el propósito de mejorar la comprensión del aprendizaje y facilitarlo, sobre todo en entornos educativos formales (Siemens, 2013).

Sin embargo, algunos autores critican el hecho de que se haya investigado poco sobre la efectividad de las intervenciones derivadas de modelos de LA (Ifenthaler & Yau, 2020; Larrabee Sønderlund, Hughes & Smith, 2019). Para superar esta deficiencia, recientemente se ha profundizado en el estudio de las razones por las cuales los estudiantes no tienen éxito, con el fin de darles apoyos específicos. Este enfoque ha sido denominado Educación de Precisión (Lu et al., 2018; Tempelaar, Rienties & Nguyen, 2021). Un abordaje prometedor es el de complementar la información aportada por las calificaciones o la actividad en la LMS con información sobre disposiciones de los estudiantes, particularmente de naturaleza socio-cognitiva, con el fin de afinar las intervenciones de apoyo (Tempelaar, Rienties & Nguyen, 2017).

La contribución que puede hacer la LA a las instituciones de educación superior no es menor. En España, el fracaso en primer año de estudios superiores es una de las principales causas de abandono de los estudios (Fernández Mellizo-Soto, 2022), fenómeno que ha sido detectado también en otros entornos, existiendo evidencias de que este abandono puede ser hasta 4,2 veces mayor entre estudiantes que fracasan (Ajjawi et al., 2020). El bajo desempeño académico tiene consecuencias individuales en la autoestima (Crocker et al., 2003), mientras que el abandono tiene además consecuencias económicas e incluso puede incidir en la interrupción de la movilidad de trayectoria social de los estudiantes (Byrom & Lightfoot, 2013).

La Analítica Predictiva del Aprendizaje (Predictive Learning Analytics, PLA), se centra en la generación de modelos para anticipar los resultados académicos de los estudiantes (Brooks & Thompson, 2017; Fiallos Quinteros, Jiménez Builes & Branch Bedoya, 2022). Abundantes trabajos se han centrado en predecir los resultados de estudiantes en asignaturas, por ejemplo, si suspenden o no, o si aprueban el examen final (Arnold & Pistilli, 2012; Choi et al., 2018; Kokoç, Akçapınar & Hasnine, 2021; Martínez-Carrascal et al., 2019; Waddington et al., 2016). Otras investigaciones han desarrollado modelos para predecir directamente la persistencia en los estudios (Aguiar et al., 2014; Foster & Siddle, 2020; Milliron, Malcolm & Kil, 2014) o el abandono (Aulck et al., 2019; García-Tinizaray, Ordoñez-Briceño & Torres-Díaz, 2015). Respecto a la predicción de asignaturas individuales, se ha visto que incluso dentro de una misma institución, los modelos predictivos no son necesariamente transferibles de una materia a otra (Conijn et al., 2017), sugiriendo que hay competencias subyacentes a la predicción que son específicas según la asignatura. Por otro lado, las investigaciones que utilizan el abandono o la persistencia como variable a predecir pueden perder de vista que hay distintos factores que revierten en dejar o continuar los estudios además del desempeño académico. En cambio, hay pocos estudios que intenten predecir los resultados académicos globales de un estudiante (Mansouri, ZareRavasan, & Ashrafi, 2021), es decir, la nota media o los créditos obtenidos al final de un año académico. Estudiar esta variable puede dar indicios sobre las competencias transversales necesarias para tener un buen rendimiento en materias de distinta naturaleza. Para cubrir esta brecha, el objetivo de esta investigación es dar respuesta a la siguiente pregunta:

¿Qué modelo construido con indicadores disponibles durante el primer semestre del año permite anticipar los resultados académicos en el conjunto de asignaturas a fin de año?

2. Revisión de la literatura

2.1. El éxito y el fracaso académico en educación superior



El conocimiento común indica que la inteligencia es la principal variable explicativa del éxito académico. En efecto, las habilidades de razonamiento pueden llegar a dar cuenta significativa de la varianza del rendimiento académico con un robusto $r=0,44$ encontrado entre medidas de este constructo (tests) y la nota media acumulada en estudios universitarios (Sackett et al., 2009). Sin embargo, en las últimas décadas, algunos académicos han explorado el fenómeno desde perspectivas más amplias. Dejando de lado variables exógenas al individuo como su origen sociocultural, factores institucionales, metodologías de enseñanza y características de los profesores, se han identificado diversos factores no intelectuales que influyen en los resultados académicos. En primer lugar, hay factores psicológicos como la autoeficacia académica, las metas de desempeño y la regulación del esfuerzo asociadas a las calificaciones universitarias (Richardson, Abraham & Bond, 2012). La motivación y el perfil vocacional también han sido identificados como críticos en el desempeño académico, específicamente, en el fracaso en los estudios (Marín Sánchez, Infante Rejano & Troyano Rodríguez, 2000). Otras investigaciones han apuntado a variables conductuales como los hábitos de estudio (Credé & Kuncel, 2008), considerados un tercer pilar explicativo del rendimiento universitario junto con los ya mencionados factores intelectuales y no intelectuales. La asistencia a clases (Karnik, Kishore & Meraj, 2020) es otro factor conductual estudiado, aunque en este caso las evidencias son menos sólidas, pues la relación puede deberse a un efecto mediado por otras variables, como el interés de los estudiantes en la asignatura (Andrietti & Velasco, 2015). Lu y Cutumisu (2022), usando precisamente técnicas de LA, han hallado recientemente que la actividad en las plataformas de aprendizaje y las calificaciones de evaluación continua en dichas plataformas median la relación entre asistencia y rendimiento. Finalmente, otras investigaciones apuntan a factores de personalidad como el optimismo (Ickson, Kaplan & Slobodin, 2020), ser concienzudo (Cárdenas Moren et al., 2020; Ickson, Kaplan & Slobodin, 2020) o la faceta de perseverancia asociada al temple (Credé, Tynan & Harms, 2017) como explicación de los resultados académicos.

La naturaleza compleja del éxito, o de su contrapartida, el fracaso académico, hace relevante el esfuerzo continuo por identificar sistemas de indicadores diversos que permitan anticiparlo y que ayuden a comprender por qué no se logran los resultados de aprendizaje esperados. En los últimos años, una nueva disciplina, la LA ha emergido con este propósito, tema que trataremos en el siguiente apartado.

2.2. La Analítica del Aprendizaje

La abundancia de datos y la mejora de modelos para explotarlos ha permitido abordar problemas conocidos a partir de nuevas evidencias. En su libro *Everybody lies*, Seth Stephens-Davidowitz (2017) explica cómo las huellas que dejan los usuarios de Internet revelan comportamientos no reconocidos de las personas, a veces, incluso inconscientes, que desafían los hallazgos de la investigación social tradicional a menudo apoyada en autopercepciones y valoraciones subjetivas. Llevado al ámbito educativo, este cambio quiere decir que la LA sirve para saber cosas que no se pueden ver de otras maneras (Pistilli & Heileman, 2017), con un potencial fascinante para mejorar la comprensión que tenemos del aprendizaje, de cómo incitarlo, facilitarlo o acelerarlo.

Al mismo tiempo, la facilidad para combinar y procesar los datos permite la elaboración de métodos más sofisticados para hacer predicciones (Pistilli & Heileman, 2017), lo cual constituye un segundo tipo de problemáticas abordadas por la LA. Se trata de usar datos socio demográficos, académicos (calificaciones) o de la actividad del estudiante en las LMS para determinar si se encuentran o no en situación de riesgo académico. Hay consenso en que es posible hacer este tipo de predicción con alto nivel de exactitud (Ifenthaler & Yau, 2020; Pistilli & Heileman, 2017).

En paralelo, con el desarrollo de nuevos modelos, una parte importante de la investigación gira en torno a cómo aprovechar esta información para la mejora de los resultados académicos. Una línea de estudios versa sobre los sistemas de alerta que entregan información predictiva a estudiantes y profesores y evalúan su impacto en los resultados académicos. En esta línea, las investigaciones arrojan indicios positivos (Bañeres et al., 2020; Foster & Siddle, 2020) en el sentido que los sistemas de alerta permiten comunicar predicciones precisas que parecen estar asociadas a, como mínimo, ligeras mejoras en los resultados.

La disciplina se encuentra en un punto en que todavía enfrenta varios desafíos. En primer lugar, la mayoría de los modelos predictivos desarrollados son específicos del contexto en que se desarrollan, es decir, de la asignatura para la cual han sido creados y por lo tanto, su portabilidad es baja (Conijn et al., 2017; Guzmán-Valenzuela et al., 2021). En segundo lugar, se ha encontrado que mucha de la investigación realizada adolece de una adecuada definición de los constructos con los que trabaja, como retención o éxito académico y que falta desarrollo de estudios a gran escala, longitudinales y con diseño experimental, que arrojen evidencias definitivas sobre el impacto de las Analíticas del Aprendizaje en el aprendizaje (Ifenthaler & Yau, 2020). Finalmente, una parte importante de los esfuerzos han apuntado a la mejora en la precisión predictiva de los modelos y se han preocupado menos de los aspectos educativos o del aprendizaje en sí (Guzmán-Valenzuela et al., 2021). En este sentido, algunas de las líneas de investigación actuales apuntan a la necesidad de entender por qué ciertas variables son predictivas y por qué lo son más que otras (e.g. Lu et al.; 2018). En relación con este déficit, algunos académicos hacen ver la necesidad de aplicar la LA con sentido común, no dejarse deslumbrar por el algoritmo y, sobre todo, no caer en el riesgo de perder de vista al estudiante (Scholes, 2016).

Una línea de investigación emergente y prometedora de los últimos años es la Educación de Precisión cuyo objetivo es “la mejora del diagnóstico, predicción, tratamiento y prevención de los resultados de aprendizaje” (Lu et al., 2018: pp. 221), intentando evitar proponer soluciones estándar a estudiantes distintos, aunque sus problemáticas sean similares. Investigaciones que se han hecho bajo este marco, han complementado la LA con información sobre disposiciones de los estudiantes para identificar maneras más adaptadas e individualizadas de intervenir y ayudarlos a tener éxito académico (Tempelaar, Rienties & Nguyen, 2017, 2021).

2.3. La predicción de los resultados académicos

“La analítica predictiva es un conjunto de técnicas usadas para hacer inferencias sobre eventos futuros” (Brooks & Thompson, 2017: p. 61) que intenta anticipar el valor de datos desconocidos basándose en observaciones de hechos pasados.

2.3.1. Cómo se predice

En un primer tiempo, las universidades estaban centradas en predecir la retención de los estudiantes en intervalos fijos de tiempo (Larrabee Sønderlund, Hughes & Smith, 2019). Estas predicciones utilizaban variables como el estrato socioeconómico, las calificaciones obtenidas en educación secundaria y resultados en exámenes estandarizados (tests) de aptitud académica como el SAT (Atkinson, 2001; Westrick et al., 2015), incluso al final de los estudios de grado (Schmitt et al., 2009). Sin embargo, investigaciones de gran escala sugieren que la predicción de la continuación de estudios y la graduación en un programa se pueden mejorar si se consideran las calificaciones obtenidas una vez iniciados los estudios (Aulck et al., 2019).

En cuanto a las variables utilizadas para hacer la predicción, se han utilizado calificaciones parciales (Bañeres et al., 2020) o combinaciones de datos sociodemográficos y datos académicos (Aulck et al., 2019). En los últimos años, la gran revolución ha sobrevenido con la consideración de la actividad en plataformas digitales, ya sea en clases presenciales (Choi et al., 2018) o en metodologías de aprendizaje combinado o blended learning (Conijn et al., 2017; Lu et al., 2018; Martínez-Carrascal, Márquez Cebrián, et al., 2020). En este último caso, las investigaciones han tenido en cuenta el tiempo que los estudiantes dedican a actividades (Tempelaar, Rienties & Nguyen, 2017), con qué frecuencia consultan los materiales, a qué aspectos del contenido se pone atención en cada momento (Casey & Azcona, 2017), o incluso patrones inferidos de la actividad de los estudiantes en las LMS a partir de distintos parámetros (Kokoç, Akçapınar & Hasnine, 2021; Martínez-Carrascal, Valderrama & Sancho-Vinuesa, 2020). El uso de esta actividad en las LMS ha permitido adelantar las predicciones sin perder precisión y son más informativos de cómo se preparan los estudiantes para enfrentar desafíos académicos, pero lo que representan todavía es considerado incierto (Conijn et al., 2017), lo que ha llevado a académicos como Knight y Buckingham-Shum a hacer un llamado a avanzar “de los clics a los constructos” (2017, p. 19).



Los estudios que han comparado clasificadores, las técnicas cuantitativas que soportan los modelos de predicción, arrojan evidencias no concluyentes. Por ejemplo, Casey y Azcona (2017) encontraron mejores resultados con el Árbol de Decisiones, al compararlo con otros como la regresión lineal, la regresión logística, diversos clasificadores de Naive Bayes, entre otros; mientras que Bertolini y colaboradores (2021) hallaron que métodos conjuntos de machine learning tienen un mejor desempeño comparados con técnicas tradicionales; y Auck et al. (2019) consideraron la regresión logística como más efectiva. No obstante, incluso la regresión lineal, un método clásico, ha sido encontrada más efectiva que otras técnicas como la regresión logística (Choi et al., 2018).

Las predicciones se intentan hacer lo antes posible. En el caso de asignaturas semestrales, ha habido predicciones interesantes ya a partir de la tercera semana de clases (Bertolini, Finch & Nehm, 2021). Cuando la variable dependiente es la graduación del programa, los datos del primer año permiten una predicción pertinente (Aulck et al., 2019).

2.3.2. Qué se hace con la predicción

Un uso frecuente de las predicciones obtenidas con PLA es la creación de sistemas de alerta que entregan retroalimentación a estudiantes, profesores o, la mayor parte del tiempo, ambos. Course Signals (Arnold & Pistilli, 2012) fue uno de los primeros sistemas que, utilizando información académica, sociodemográfica y de actividad en la LMS de los estudiantes, les daba retroalimentación sobre su probabilidad de aprobar una asignatura en un momento dado del semestre, mejorando la tasa de superación con respecto a aquellos cursos donde no se utilizaba. Bañeres y colaboradores (2020) han dado cuenta de un sistema similar puesto en marcha en una universidad de educación completamente en línea.

La revisión realizada por Larrabee Sønderlund y otros investigadores (2019) han reportado que, aunque todavía hay escasas evidencias, estos sistemas aportan mejoras significativas de calificaciones e incluso entre un 11 y un 25% más de superación en los estudios. Es probable que el tipo de retroalimentación arrojado por estos sistemas, su frecuencia y las condiciones en que se entrega, sean clave para su efectividad. Por ejemplo, algunos estudios muestran que la retroalimentación entregada por sistemas de alerta puede tener una influencia negativa en la orientación a metas de maestría de los estudiantes (Lonn, Aguilar & Teasley, 2015) e incluso han tenido el efecto contrario al propósito inicial, como ocurrió en el caso de una versión modificada del ya mencionado Course Signals en que se encontró un 11,5% más de riesgo de abandono entre estudiantes que recibieron la retroalimentación en comparación con un grupo de control (Jayaprakash et al., 2014).

2.3.3. Qué se predice

Las investigaciones con PLA se reparten en dos grupos. En primer lugar están aquellas que se centran en asignaturas específicas, ya sea la predicción de los resultados en un examen final (Casey & Azcona, 2017; Choi et al., 2018; Huang et al., 2020; Tempelaar, Rienties & Nguyen, 2017); o en la probabilidad que tienen de aprobar o suspender la asignatura (Jayaprakash et al., 2014; Kokoç, Akçapınar & Hasnine, 2021; Martínez-Carrascal et al., 2019). Un segundo grupo, más reducido, de estudios usan como variable dependiente la persistencia (Aguilar et al., 2014) o abandono en un programa (Aulck et al., 2019; García-Tinizaray, Ordoñez-Briceño & Torres-Díaz, 2015), el progreso del estudiante (Foster & Siddle, 2020); o el tiempo que tardarán en graduarse (Herzog, 2006). Solo se ha podido identificar un estudio donde la variable a predecir sea una medida global de rendimiento, en este caso, la nota media en un Master en línea (Mansouri, ZareRavasan, & Ashrafi, 2021).

Cabe hacer notar que en ninguno de los artículos académicos revisados se detectó una discusión en profundidad sobre la validez de estas variables operacionales respecto al aprendizaje de los estudiantes, confirmando la aprensión expresada por Ifenthaler y Yau (2020) de que “existen debilidades obvias en la definición de variables clave como la retención o el éxito en los estudios” (p. 1981) en su propia revisión de investigaciones.

3. Metodología

Para responder la pregunta de investigación, se realizó un estudio de caso ex post facto de corte cuantitativo con estudiantes de primer año de un Grado de Negocios de una escuela de negocios europea con campus en Barcelona. En el primer año se encuentran las mayores tasas de fracaso académico y abandono de los estudios, tendencia consistente con otros sistemas de educación superior como el español (Lassibille & Navarro Gómez, 2008).

3.1. Población y contexto

Los datos utilizados para desarrollar el modelo (datos de entrenamiento) provienen de las cohortes de primer año de programa en 2019-2020 y 2020-2021 de dos campus (n=1046); mientras que los datos para ponerlo a prueba provienen de la cohorte 2021-2022 solo del campus español (n=74).

El programa está reconocido por las autoridades del país de origen de la escuela, y consta de tres años académicos. El currículo académico es común, por lo que Responsables Académicos coordinan a equipos de profesores con niveles similares de conocimiento experto y competencia pedagógica. La evaluación continua se realiza a través de una variedad de instrumentos como exámenes parciales, ejercicios, presentaciones, ensayos; siendo común la evaluación de trabajos o proyectos desarrollados con un equipo de trabajo de entre dos y cinco estudiantes. Los instrumentos de evaluación continua de algunas asignaturas pueden diferir entre campus, pero los resultados son supervisados por los Responsables Académicos con el fin de controlar sesgos de evaluador. En cambio, los exámenes finales son idénticos o comparables en cuanto a estructura, dificultad y duración, difiriendo en algunas preguntas específicas del contexto local, por ejemplo, de casos vistos en clase que difieren de un país a otro. El cálculo de la media de cada asignatura se hace con la misma fórmula de ponderación en ambos campus. En la mayoría de los casos, la nota de evaluación continua tiene un peso del 40% en la media final de la asignatura, mientras que el examen final cuenta por el 60% restante. Los estudiantes estaban organizados en las mismas secciones en todas las asignaturas del año. En la Tabla 1 se muestra la frecuencia de casos en cada categoría incluida en la muestra.

		Campus		Total
		BCN	Francia	
		Recuento	Recuento	
Género	Desconocido	24	43	67
	Femenino	61	407	468
	Masculino	78	433	511
Total		163	883	1046

Tabla 1. Estudiantes por campus, según género. Fuente: Elaboración propia.

Los estudiantes de la muestra tienen diversas nacionalidades y provienen de distintos sistemas de estudios secundarios.

Los datos de calificaciones y créditos fueron obtenidos de un sistema de información donde son rigurosamente registrados y mantenidos a perpetuidad. Las calificaciones de evaluación continua normalmente se tienen en el mes de noviembre, mientras que las calificaciones finales de cada asignatura están disponibles en enero del año siguiente. Al tratarse de un estudio ex post facto, no se pudo determinar con precisión el momento en que la predicción está disponible.

En esta escuela, la asistencia a clases es obligatoria y se controla por lista o registro con una tarjeta. Durante el período académico 2020-2021, el porcentaje de asistencia considera las clases que fueron presenciales y aquellas que se hicieron online a través de herramientas de videoconferencia. En este último caso, es posible que el indicador sea menos fiable dada la tendencia de los estudiantes a conectarse sin poner la cámara.

El género es declarado por el propio estudiante cuando se matricula en el programa u obtenido de su documento de identidad.

Por último, Moodle es la plataforma de gestión del aprendizaje utilizada por la escuela. La información sobre actividad es recuperada por el Departamento de Informática que facilita los datos agregados para uso de las secretarías académicas.

3.2. Variables

El riesgo académico, que es la variable a predecir, es una variable binaria, donde el caso “positivo” (1) es obtener menos de 54 créditos ECTS en asignaturas al final del año académico (excluyendo módulos no completados como las prácticas); mientras que el caso “negativo” (0) es obtener 54 créditos o más en el mismo periodo. Los créditos se obtienen al alcanzar una media ponderada igual o mayor al 50% de la calificación máxima. El programa permite que los estudiantes con una nota inferior al 50% realicen un examen de recuperación para conseguir los créditos faltantes cuya nota, en caso de ser mayor, reemplaza la de la media de la asignatura con un tope del 60% de la nota máxima. Algunas asignaturas están excluidas de esta posibilidad, por ejemplo, Ventas. Es importante explicar que, durante el periodo lectivo los estudiantes pueden obtener hasta 64 créditos ECTS, por lo que tienen un margen de créditos que pueden suspender sin arriesgarse a no ser admitidos en segundo año. También cabe señalar que los estudiantes que no alcanzan los criterios para aprobar deben repetir el año completo, no solo algunas asignaturas, por lo que hay mucho en juego al no aprobar.

Las calificaciones en asignaturas previas (e.g. Aguiar et al., 2014; Choi et al., 2018) o quizzes de la propia asignatura (Bañeres et al.; 2020; Choi et al.; 2018; Lu et al.; 2018; Tempelaar, Rienties & Nguyen; 2017, 2021) son uno de los indicadores más utilizados, por lo que las calificaciones han sido parte de las variables predictoras consideradas en esta investigación. En concreto, todas las calificaciones obtenidas en el primer semestre del año (entre septiembre y diciembre). Las calificaciones de evaluación continua son más interesantes para el modelo, puesto que están disponibles normalmente en noviembre, adelantando el momento en que se puede hacer la predicción.

La actividad de los estudiantes en la LMS también se ha usado extensamente como variable predictora de resultados académicos (Aguiar et al., 2014; Arnold & Pistilli, 2012; Bertolini, Finch & Nehm, 2021; Casey & Azcona, 2017; Conijn et al., 2017; James, 2018; Jayaprakash et al., 2014; O. H. T. Lu et al., 2018; Martínez-Carrascal, Márquez Cebrián, et al., 2020; Tempelaar, Rienties & Nguyen, 2017, por citar algunos trabajos). En el caso de esta investigación, los datos disponibles de actividad en la plataforma de aprendizaje virtual es el número de clics en el conjunto de asignaturas durante el periodo estudiado.

Asimismo, se ha incluido la asistencia a clases teniendo en cuenta la relación positiva que se ha encontrado entre dicha variable y el rendimiento académico (Bevitt, Baldwin & Calvert, 2010; Karnik, Kishore & Meraj, 2020), ya sea directa o indirectamente (Lu & Cutumisu, 2022).

Al no disponer de medias de educación secundaria o pruebas estandarizadas, la única variable sociodemográfica a la que ha habido acceso es el género, que en algunos estudios ha sido tenida en cuenta como predictiva (Tempelaar, Rienties & Nguyen, 2021). En la Tabla 2 se presentan las variables tenidas en cuenta en el estudio.

Variable	Descripción operacional	Mínimo	Máximo	Media	Desviación estándar
Riesgo Académico	Suspende=1, Aprueba=0	0	1		
Género	Femenino = 1, Masculino = 2	1	2		
Asistencia 1er semestre	Tasa de asistencia a clases durante el primer semestre (%)	1	100	90,2115	12,36646
ECTS	Número de ECTS obtenidos antes de Prácticas y Servicio Ciudadano	0	64	58,28	10,028
Media S1	Meda académica ponderada del primer semestre	5,00	90,35	67,8190	10,09212
Nota Des Prof	Nota de la asignatura Desarrollo Profesional (100% evaluación continua)	0	100	75,6725	13,64296
CC Conta Fi	Nota de Evaluación continua de la asignatura Contabilidad Financiera	0	100	66,4329	14,2586
CC Eco	Nota de Evaluación continua de la asignatura Economía	0	100	75,2313	21,12246
CC Dig Tools	Nota de Evaluación continua de la asignatura Herramientas Informáticas	0	100	75,216	16,11444
CC Fund Mkt	Nota de Evaluación continua de la asignatura Fundamentos del Marketing	3	92,25	67,193	10,9745
Media Ventas	Media de la asignatura de Ventas (100% evaluación continua)	0	100	68,7874	15,17098
Total clics	Total de clics en la LMS durante el semestre	21	368	134,18	37,444

Tabla 2. Estadística descriptiva de las variables consideradas. Fuente: Elaboración propia.

3.3. Estrategia y técnicas de análisis

El análisis preliminar de los datos se hizo utilizando correlaciones. El algoritmo de clasificación para desarrollar el modelo predictivo fue la regresión logística, con el procedimiento “Avanzar por pasos (Wald)”, siguiendo las recomendaciones de Berlanga-Silvente y Vilá-Baños (2014), que tiene como ventaja la posibilidad de reducir el riesgo de sobreajuste y provee una probabilidad de riesgo para cada uno de los casos (Choi et al., 2018).

Se realizaron pruebas con distintas combinaciones de variables para obtener los índices más altos de sensibilidad con un modelo parsimonioso, es decir, con el menor número de variables posible. La sensibilidad y la especificidad se calcularon de acuerdo a las siguientes fórmulas (Hosmer, Lemeshow & Sturdivant, 2013):

Sensibilidad: $TP/TP + FN$

Especificidad: $TN/TN + FP$

Precisión: $TP/TP + FP$

Donde TP (true positive) es el número de estudiantes para quienes el modelo predice correctamente que suspenderán; FN (false negative), los estudiantes identificados en riesgo que, en realidad, aprueban; TN (true negative), los estudiantes que el modelo predice correctamente que aprobarán y, finalmente FP (false positive), los estudiantes que suspenden, aunque el modelo predecía que aprobarían.

El modelo reducido se comparó con uno que incluye todas las calificaciones de evaluación continua como predictores utilizando el método Receiver Operating Characteristic, ROC (ver Casey & Azcona, 2017; Jayaprakash et al., 2014). Una precisión de 75% de acierto es considerada aceptable (Hosmer, Lemeshow & Sturdivant, 2013; Lu et al., 2018).

Todos estos análisis se realizaron con IBM® SPSS® Statistics versión 27.



3.4. Consideraciones éticas de la investigación

La investigación siguió el Código de Integridad en la Investigación de la Universitat de Barcelona.

La Dirección del Programa ha autorizado el uso de los datos de los estudiantes para la investigación. Se trabajó con información académica anonimizada (individualizada con códigos). Puesto que el modelo a construir representa una primera aproximación y no se puso a prueba hasta fin de año (julio), no se informó a los estudiantes sobre la predicción que el modelo hacía en su caso, ni hubo posibilidad de que la predicción afectara de alguna manera el resultado académico.

La investigación se realizó de la manera más rigurosa, honesta y respetuosa posible, velando por la calidad de los datos a tratar en cada una de las fases. No recibió financiación y no se detectan conflictos de interés.

4. Resultados

4.1. Relación entre rendimiento académico, calificaciones, asistencia y género

El análisis de correlaciones de Pearson confirma, sin sorpresas, relaciones significativas entre calificaciones y créditos obtenidos a fin de año.

La tasa de asistencia fue confirmada como un factor relevante, puesto que aparece relacionada directa y significativamente ($r=0,509$, $p<0,01$) a los resultados académico generales (Nº de ECTS) y, en general, a la media de cada asignatura.

Se ha sugerido que mujeres y hombres exhiben comportamientos diferenciales que influyen en su rendimiento (Balkis & Duru, 2017). En la muestra no fue posible identificar el género de 67 estudiantes (6,41% del total), posiblemente porque ya no son estudiantes del programa, lo cual puede haber introducido un sesgo en el análisis si este abandono se debió a fracaso académico. Al computar la variable como ordinal con los datos disponibles, asignando el valor de referencia más bajo (1) al género femenino, la correlación de esta variable resulta ser inversamente proporcional y significativa ($p<0,05$) con las medias de varias asignaturas. Es decir que, en esta muestra las mujeres tienen de media mejores resultados que los hombres. Sin embargo, el método de “Avanzar por pasos (Wald)” excluyó esta variable como predictor en los modelos predictivos.

4.2. Relación no lineal entre clics y rendimiento académico

También hay casos perdidos en el número de clics en la LMS (6,02%), posiblemente por las mismas razones que en el caso del género y con consecuencias similares respecto a la introducción de un sesgo muestral. Después del análisis inicial, se encontró una correlación significativa y positiva, pero débil, entre el número de clics y varias de las variables de desempeño académico, durante el semestre y a fin de año, entre ellas, el número de créditos obtenidos a fin del año académico (VD). Sin embargo, al graficar la relación entre la media académica del primer semestre y el número de clics, se detectó una relación curvilínea, como se puede apreciar en la Figura 1. Esto sugiere que los estudiantes cuya actividad es muy baja o muy alta tienden a tener un peor rendimiento.

En la Figura 1 se puede visualizar esta relación que queda confirmada por la mejora del R^2 a 0,192 ($p<0,01$) al comparar un modelo de regresión de la media del primer semestre (Media S1) que solo usa el total de clics (a la izquierda de la figura) y otro modelo que incluye además el total de clics al cuadrado (en el centro).

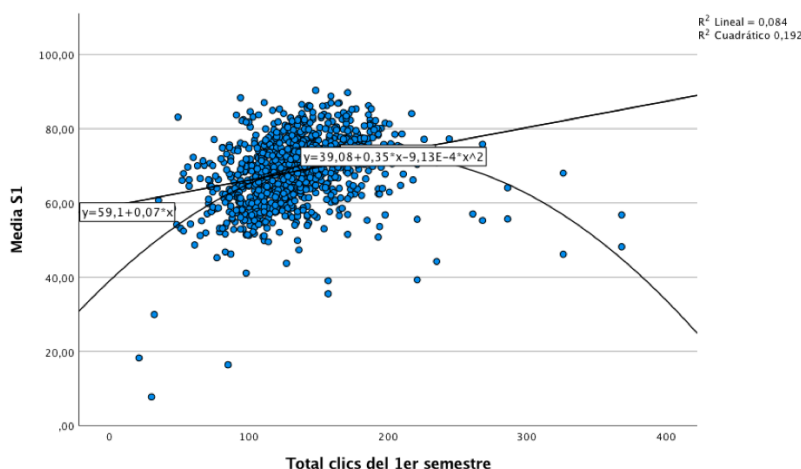


Figura 1. Diagrama de dispersión entre Clics y media del primer semestre. Fuente: Elaboración propia.

Dada esta relación cuadrática se decidió transformar la variable “Total clics” en “Abs. Z Clics” para determinar si se podía obtener una correlación más alta con las medidas de rendimiento académico. El proceso utilizado fue estandarizar dicha variable y computar sus valores absolutos, de modo que los valores cercanos o iguales al centro de la distribución se transforman en puntajes iguales o cercanos a 0, mientras que los niveles de actividad que se desvían de la media, por ser más altos o más bajos, corresponden a valores más altos en la nueva variable computada, “Abs. Z Clics”. La correlación con los ECTS a fin de año aumenta, sigue siendo significativa, y se transforma en negativa ($R = -0,419$, $p < 0,01$), implicando que mientras mayor es la desviación de la actividad de un estudiante en la LMS respecto a la media del grupo, ya sea por poca o demasiada actividad, peores los resultados académicos.

4.3. La pandemia de COVID-19

La pandemia del COVID-19 introdujo cambios a tener en cuenta. Durante el año académico 2020-2021 los exámenes se realizaron a distancia, sin vigilancia, y la tasa de asistencia a clases podía representar asistencia presencial o a través de Zoom. Para prever el impacto de semejante evento en el modelo y la relación entre variables, se compararon separadamente las correlaciones entre las medias de las asignaturas del primer semestre (incluida la media del semestre), la tasa de asistencia y la variable Abs. Z Clics para cada una de las cohortes, pre y post COVID-19. Aunque las magnitudes variaron, la mayoría de las relaciones estadísticamente significativas se mantuvieron al mismo nivel de significatividad, excepto en el caso de una asignatura (Desarrollo Profesional) que finalmente no fue incluida en el modelo y para la relación entre Abs. Z clics y la nota media de Economía que mantuvo el sentido de la relación, pero solo fue significativa al 95% en 2020-21 ($R = -0,219$ vs. $-0,106$). Se consideró seguro continuar utilizando el set de datos completo y no segmentado.

4.4. Modelo de predicción del riesgo académico

Para responder a la pregunta de investigación, se realizaron diversas iteraciones del procedimiento de regresión logística “Avanzar por pasos (Wald)” (ver Berlanga-Silvente & Baños, 2014). El modelo sintético con mayor sensibilidad quedó constituido por las calificaciones de evaluación continua de tres asignaturas, Marketing, Economía y Ventas. En todas las iteraciones, la variable computada “Abs. Z Clics” fue seleccionada. Cabe mencionar que, si bien la asistencia a clases quedó incluida en este modelo, era excluida en algunas combinaciones, posiblemente porque está fuertemente asociada a algunas calificaciones (en concepto de participación o por influir en algún tipo de sesgo de evaluación). Su incorporación se sostiene por consideraciones teóricas, como manifestación de esfuerzo y perseverancia del estudiante. En consecuencia, el modelo incluye tres tipos de variables: calificaciones de evaluación continua de algunas asignaturas, la asistencia a clases y actividad en la LMS en número de clics.

Para comprobar el supuesto de no colinealidad que requiere la regresión logística se hicieron regresiones lineales de cada una de las variables con las demás. Se tuvieron en cuenta el Índice de Tolerancia y el Factor de Inflación de la Varianza (Variance Inflation Factor, VIF). Los valores entre las variables incluídas cumplieron con los estándares comúnmente aceptados, esto es, Índice de Tolerancia inferior a 0,1 y un VIF menor que 10 (Regor, s. f.).

La prueba de Hosmer y Lemeshow de diez categorías muestra que no hay desviaciones significativas entre el número de casos pronosticados y el número de casos observados para cada uno de los valores de la variable dependiente (suspende o aprueba) en ninguna de las categorías ($\chi^2=4,729$, $p=0,786$), lo cual constituye una medida de validez del modelo (Hosmer, Lemeshow & Sturdivant, 2013). Adicionalmente, el R^2 de Nagelkerke es de 0,420. Considerando que este indicador, también llamado pseudo R^2 , tiene una naturaleza cercana al R^2 de la regresión lineal (esto es, el porcentaje de la varianza de la Variable Dependiente explicada por el modelo), este valor se puede tomar como un argumento adicional en apoyo al modelo (ver Tabla 3).

Variables en la ecuación	B	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
Asistencia 1er semestre	-0,024	0,012	4,157	1	0,041	0,977	0,955	0,999
Abs. Z Clics	0,671	0,170	15,614	1	0,000	1,956	1,402	2,728
CC Fund Mkt	-0,412	0,067	37,409	1	0,000	0,662	0,581	0,756
CC Eco	-0,184	0,028	42,154	1	0,000	0,832	0,786	0,879
Media Ventas	-0,125	0,049	6,442	1	0,011	0,882	0,801	0,972
Constante	9,045	1,228	54,290	1	0,000	8473,876		

B = Beta; gl=grados de libertad; Sig.=nivel de significación; EXP(B)=Razón de probabilidades; C.I.=Intervalo de Confianza.

Tabla 3. Indicadores del modelo reducido. Fuente: Elaboración propia.

En el caso estudiado, el riesgo real de fracaso es relativamente bajo, puesto que 88,6% de los estudiantes aprueba. Por esta razón se ha calculado también la tasa de Precisión que fue del 72,73%, indicando que potenciales intervenciones basadas en este modelo serán oportunas en la mayoría de los casos abordados.

4.5. Comparación entre un modelo reducido y otro con todas las calificaciones

Siguiendo el mismo procedimiento, se ensayó la predicción del riesgo académico utilizando la totalidad de las calificaciones de evaluación continua. La regresión logística utilizando el método “hacia delante: Wald”. En el séptimo paso, dicho método seleccionó la variable Abs. Z Clics, y todas las calificaciones de evaluación continua de las asignaturas del primer semestre excepto Desarrollo Profesional, pero excluyó la asistencia. Con un R^2 de Nagelkerke igual a 0,462, y una precisión de 72,13% de precisión, los indicadores de robustez del modelo con todas las calificaciones fueron positivos y similares al modelo reducido. En la Tabla 4 se pueden comparar también la sensibilidad, la especificidad y el porcentaje global entre ambos modelos. Es importante señalar que Hosmer et al. (2013) desaconsejan esta tabla de clasificación como criterio para juzgar el ajuste del modelo, aunque también indican que esta tabla “es más apropiada cuando la clasificación es la meta establecida del análisis” (p. 173), como es el caso presente.

			Modelo reducido			Modelo con todas las variables de calificaciones		
			Pronosticado			Pronosticado		
			Riesgo Académico			Riesgo Académico		
	Observado		Aprueba	Suspende	Porcentaje correcto	Aprueba	Suspende	Porcentaje correcto
Especificidad	Riesgo Académico	Aprueba	852	15	98,3	850	17	98,0
Sensibilidad		Suspende	72	40	35,7	68	44	39,3
Porcentaje global					91,1			91,3
a. El valor de corte es ,500								

Tabla 4. Comparación de la clasificación obtenida con el modelo reducido y el modelo con todas las calificaciones. Fuente: Elaboración propia.

Para una comparación de ambos modelos se llevó a cabo el análisis ROC (ver Figura 2), de uso común para valorar la efectividad de los clasificadores (e.g. Casey & Azcona, 2017). El análisis muestra que ambos modelos de predicción describen una discriminación que puede ser considerada excelente, puesto que el área bajo la curva (AUC) en ambos casos es mayor que el 80% (Hosmer, Lemeshow & Sturdivant, 2013), concretamente 87,7% para el modelo reducido y 89,6% para el modelo con todas las calificaciones.

Dado que los indicadores son similares en ambos modelos, la parsimonia y facilidad del modelo reducido lo hace más aconsejable para el uso práctico en el caso estudiado.

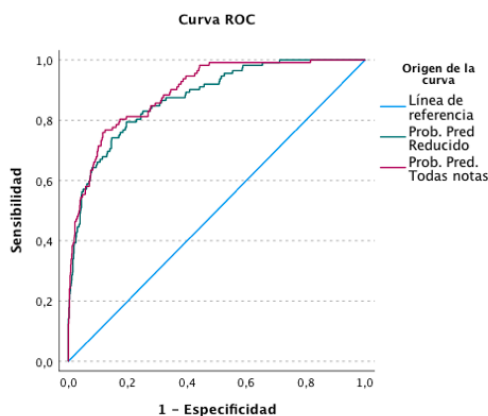


Figura 2. Curva ROC del modelo reducido y el modelo con todas las calificaciones. Fuente: Elaboración propia.

4.6. Predicción para la cohorte 2021-2022

Una vez establecido el modelo, se probó con un grupo de estudiantes del campus español de la escuela de negocios, pertenecientes al mismo programa. De las 78 observaciones disponibles, no fue posible recuperar los registros de clics en el entorno virtual de aprendizaje de 4 estudiantes, con lo cual el grupo definitivo fue de 74 personas. Al clasificar como “en riesgo” a los estudiantes con una probabilidad predicha de no aprobar mayor que 0,5, el modelo predice que 6 estudiantes (8,3% de la muestra de test) se encontrarán en riesgo de no cumplir con el criterio de pase a segundo año del programa. Los resultados del modelo arrojan una alta precisión con cinco casos reales de riesgo de los 6 identificados, aunque con un alto número de falsos negativos (ver Tabla 5).

Predicción Observado	Aprueba	Suspende	Total general	Sensibilidad
Aprueba	49	1	50	
Suspende	19	5	24	20,83%
Total general	68	6	74	
Precisión		83,33%		

Tabla 5. Resultados de la predicción con datos de 21-22. Fuente: Elaboración propia.

Con estos argumentos, podemos contestar a la pregunta de investigación diciendo que, en esta escuela de negocios, un modelo que incluya las calificaciones de evaluación continua de Marketing, Economía y Ventas, la tasa de asistencia durante el primer semestre y la actividad en la LMS (total de clics en el periodo), permite anticipar quiénes estarán en riesgo de no obtener resultados satisfactorios a fin de año.

5. Conclusiones

Los resultados de este caso de estudio muestran que es posible predecir los resultados académicos globales al final del primer año utilizando indicadores disponibles en el primer semestre. La R^2 de Nagelkerke (0,420) y el AUC (87,7%) son indicadores que soportan el modelo. Globalmente, estos hallazgos confirman investigaciones previas que predecían la persistencia (Aguiar et al., 2014) o el abandono de un programa (Aulck et al., 2019) y contribuyen al acervo de este tipo de estudios, porque todavía hay pocas investigaciones (Mansouri, ZareRavasan, & Ashrafi, 2021) que utilicen LA para predecir resultados académicos globales, no solo en una asignatura específica.

La prueba del modelo con datos de la cohorte 2021-2022 señala que un 83% de los estudiantes identificados, efectivamente se encontraron en situación de riesgo. Como debilidad, en esta muestra de prueba, la sensibilidad es más bien baja (20,83%), que en términos prácticos indica que, si el modelo se hubiera usado para dar apoyos específicos, un grupo de estudiantes que necesitaba este apoyo habría pasado desapercibido.

El modelo reducido incluye las notas de evaluación continua de Ventas, Marketing y Economía. La nota de Ventas está calculada principalmente a partir de simulaciones y participación en clase. La nota de evaluación continua de Marketing refleja una combinación de apropiación de conceptos con algunos elementos cuantitativos, predominando los primeros. Por último, Economía es más bien lo contrario, requiriendo una mayor competencia numérica. Marcal et al. (2005) encontraron que las habilidades de comunicación de los estudiantes están asociadas a su rendimiento en cursos iniciales de Marketing, mientras que Liesz y Reyes (1989) detectaron fuertes asociaciones entre habilidades de razonamiento y desempeño en cursos de finanzas, una materia de estudios cercana a la Economía. Lo anterior sugiere que las notas de las asignaturas incluidas en el modelo se asocian a habilidades subyacentes de distinto tipo, que son transversales al desempeño académico general en este programa de estudios.

Otra variable incluida en el modelo es la actividad en la LMS (clics), cuya relación con el rendimiento académico se explica mejor a través de una relación cuadrática ($R^2=0,192$ vs. $R^2=0,084$ para una relación lineal). Esta relación no lineal se corresponde con lo encontrado por el estudio llevado a cabo por Cavanaugh, Hargis y Mayberry (2016) en una institución de educación superior solo para mujeres en el medio oriente. Ellos analizaron la relación entre el número de clics, el tiempo dentro de la LMS y las calificaciones de una asignatura de Introducción a la Investigación y encontraron una relación cuadrática muy similar. Los autores conjeturaron que un mayor número de clics puede deberse a distracción del estudiante o a menor preparación previa, pero no proponen una explicación definitiva. Por su parte, Conijn et al. (2017), después de investigar en 17 cursos combinados, señalaron que la actividad en la LMS no hacía una contribución significativa a la predicción cuando se contaba con las calificaciones. Estos autores abogan por una mayor comprensión de lo que representa esta actividad. Los resultados de nuestro estudio sugieren que los clics pueden asociarse con distintas disposiciones de los estudiantes. En efecto, no sorprende que estudiantes con bajo nivel de actividad

tiendan a presentar peor rendimiento, pues se asocia a un abordaje pobre de los deberes académicos. La interpretación es necesariamente distinta para los estudiantes que obtienen peores resultados, pero que son muy activos en la LMS durante el semestre, lo cual además de las conjeturas de Cavanaugh y colaboradores (2016), podría ser una manifestación de perseverancia (Credé, Tynan & Harms, 2017). Futuras investigaciones deberían determinar el mecanismo subyacente a esta relación peculiar y revisar estudios anteriores para determinar si la falta de importancia de esta actividad en los modelos, no es resultado de considerar que la relación es de naturaleza lineal.

Finalmente, no sorprende que la asistencia también aparezca como predictor dada la correlación alta y significativa con los créditos a fin de año (0,506 $p < 0,01$). Este resultado está en línea con la investigación existente (Andrietti & Velasco, 2015; Karnik, Kishore & Meraj, 2020), confirmando que asistir a clases regularmente revierte en mejores resultados académicos.

5.1. Limitaciones e investigación futura

Este trabajo cuenta con limitaciones de distinta índole. Primero, al tratarse del estudio de un solo caso, no es posible valorar la transferibilidad del modelo, lo cual ya ha sido señalado en otras investigaciones (Conijn et al., 2017). La mejora en este sentido es que el caso estudiado es un programa transnacional, lo cual aporta validez al modelo en entornos con distintas características culturales. Para superar esta limitación, futuros estudios tendrían que incluir programas similares de distintas instituciones, intentando desarrollar un modelo predictivo que pueda ser generalizable, por ejemplo, a distintos Grados en negocios.

Segundo, al ser un estudio *ex post facto*, hay un sesgo en la muestra utilizada para crear el modelo. Como se indicó, hay 71 casos para los cuales no se cuenta con los valores de género, y 67 para los que tampoco se cuenta con el número de clics. Es posible que estos datos no se encuentren en las bases de datos a que se tuvo acceso, porque se trata de estudiantes que han abandonado el programa. Investigaciones futuras podrían replicar el diseño presentado, asegurando la integridad de los datos, lo que sería posible si la institución o instituciones donde se investiga emprenden la pesquisa de manera prospectiva, velando porque los ficheros de registros conserven toda la información de la muestra desde el momento en que se inicia el trabajo.

También por la naturaleza retrospectiva del estudio no se pudo determinar con precisión la anticipación con la que se puede hacer la predicción. Los indicadores de asistencia y clics pueden obtenerse para el periodo que se desee, pero no puede saberse con precisión en qué fecha estaban disponibles las calificaciones de evaluación continua, por lo que no fue posible probar con datos a periodos regulares para identificar el momento más temprano en que se puede hacer una predicción fiable. Un futuro diseño podría explorar la posibilidad de hacer la predicción con medidas de evaluación continua a medida que se vayan obteniendo, intentando anticiparla al máximo.

La relación cuadrática entre los clics en la plataforma de aprendizaje y el rendimiento académico, de naturaleza más bien contra intuitiva, debería ser el objeto de un estudio. El propósito sería determinar qué variables explican sobre todo el extremo derecho de la distribución (estudiantes con mucha actividad, cuyos resultados académicos no son buenos), si se produce en otros contextos, si es un efecto estable o no a través del tiempo, si se replica en todas las asignaturas o solo en las de un tipo, entre otras incógnitas.

Para acabar, el modelo permite adelantar las intervenciones para ayudar a los estudiantes que más lo necesitan a evitar el fracaso a fin de año. Acciones basadas en este modelo deberían orientarse a indagar en las dificultades específicas de los estudiantes en riesgo de no superar el primer curso, contribuyendo a las todavía incipientes evidencias de cómo la LA puede mejorar los resultados académicos (Ifenthaler & Yau, 2020; Larrabee Sønderlund, Hughes & Smith, 2019). A pesar de las pistas que ofrecen las variables predictivas, el modelo continúa siendo una caja negra. Descifrar esta caja negra, continuar el esfuerzo por aportar teoría a las LA y avanzar “de los clics a los constructos” (Knight & Shum, 2017, p. 18) son los principales desafíos futuros.



Agradecimientos

Agradecimientos a Diana Pérez Rojas por su contribución al ordenamiento de los datos y al Dr. Antonio Bartolomé Pina por su valiosa retroalimentación.

Financiación

Esta investigación no recibió financiación externa.

Cómo citar este artículo / How to cite this paper

Zúñiga Varela, G.; Lindín, C.; Sancho Vinuesa, T. (2025). Predicción de riesgo académico usando notas, asistencia a clases y clics en el LMS. *Campus Virtuales*, 14(1), 83-99.
<https://doi.org/10.54988/cv.2025.1.1434>

Referencias

- Aguiar, E.; Ambrose, G. A.; Chawla, N. V.; Goodrich, V.; Brockman, J. (2014). Engagement vs Performance: Using Electronic Portfolios to Predict First Semester Engineering Student Persistence. *Journal of Learning Analytics*, 1(3), 7-33. <http://dx.doi.org/10.18608/jla.2014.13.3>.
- Ajjawi, R.; Dracup, M.; Zacharias, N.; Bennett, S.; Boud, D. (2020). Persisting students' explanations of and emotional responses to academic failure. *Higher Education Research & Development*, 39(2), 185-199. <https://doi.org/10.1080/07294360.2019.1664999>.
- Andrietti, V.; Velasco, C. (2015). Lecture Attendance, Study Time, and Academic Performance: A Panel Data Study. *Journal of Economic Education*, 46(3), 239-259. <https://doi.org/10.1080/00220485.2015.1040182>.
- Arnold, K. E.; Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 267. <https://doi.org/10.1145/2330601.2330666>.
- Atkinson, R. (2001). Standardized tests and access to American universities. The 2001 Robert H. Atwell Distinguished Lecture. The annual meeting of the American Council on Education. Washington D.C., Estados Unidos. <http://dx.doi.org/10.1525/9780520933941-017>.
- Aulck, L.; Nambi, D.; Velagapudi, N.; Blumenstock, J.; West, J. (2019, julio). Mining University Registrar Records to Predict First-Year Undergraduate Attrition. In *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*. The 12th International Conference on Educational Data Mining (EDM 2019). Montreal, Canada.
- Balkis, M.; Duru, E. (2017). Gender Differences in the Relationship between Academic Procrastination, Satisfaction with Academic Life and Academic Performance. *Electronic Journal of Research in Educational Psychology*, 15(1), 105-125. <https://doi.org/10.14204/ejrep.41.16042>.
- Bañeres, D.; Rodríguez, M. E.; Guerrero-Roldán, A. E.; Karadeniz, A. (2020). An Early Warning System to Detect At-Risk Students in Online Higher Education. *Applied Sciences*, 10(13), 4427. <https://doi.org/10.3390/app10134427>.
- Berlanga-Silvente, V.; Baños, R. V. (2014). Cómo obtener un Modelo de Regresión Logística Binaria con SPSS. *REIRE Revista d'Innovació i Recerca en Educació*, 7(2), 105-118. <https://doi.org/10.1344/reire2014.7.2727>.
- Bertolini, R.; Finch, S. J.; Nehm, R. H. (2021). Testing the Impact of Novel Assessment Sources and Machine Learning Methods on Predictive Outcome Modeling in Undergraduate Biology. *Journal of Science Education and Technology*, 30(2), 193-209. <http://dx.doi.org/10.1007/s10956-020-09888-8>.
- Bevitt, D.; Baldwin, C.; Calvert, J. (2010). Intervening Early: Attendance and Performance Monitoring as a Trigger for First Year Support in the Biosciences. *Bioscience Education*, 15(1), 1-14. <https://doi.org/10.3108/beej.15.4>.
- Brooks, C.; Thompson, C. (2017). Predictive modelling in teaching and learning. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *The Handbook of Learning Analytics* (1st Edition, pp. 61-68). Alberta (Canada): Society for Learning Analytics Research (SoLAR). <https://doi.org/10.18608/hla17>.
- Byrom, T.; Lightfoot, N. (2013). Interrupted trajectories: The impact of academic failure on the social mobility of working-class students. *British Journal of Sociology of Education*, 34(5-6), 812-828. <https://doi.org/10.1080/01425692.2013.816042>.
- Cárdenas Moren, C.; Crawford Augant, K.; Crawford Labrin, B.; Soto de Giorgis, R.; de la Fuente-Mella, H.; Peña Fritz, Á.; Valenzuela Saavedra, M.; Hermosilla Monckton, P.; Álvarez Castellí, L. (2020). A quantitative analysis of the identification of personality traits in engineering students and their relation to academic performance. *Studies in Higher Education*, 45(7), 1323-1334. <http://dx.doi.org/10.1080/03075079.2019.1572089>.
- Casey, K.; Azcona, D. (2017). Utilizing student activity patterns to predict performance. *International Journal of Educational Technology in Higher Education*, 14(1), 1-15. <https://doi.org/10.1186/s41239-017-0044-3>.
- Cavanaugh, C.; Hargis, J.; Mayberry, J. (2016). Participation in the virtual environment of blended college courses: An activity study of student performance. *The International Review of Research in Open and Distributed Learning*, 17(3), 423-432.

- <http://dx.doi.org/10.19173/irrodl.v17i3.1811>.
- Choi, S. P. M.; Lam, S. S.; Li, K. C.; Wong, B. T. (2018). Learning Analytics at Low Cost: At-risk Student Prediction with Clicker Data and Systematic Proactive Interventions. *Journal of Educational Technology & Society*, 21(2), 273-290.
- Conijn, R.; Snijders, C.; Kleingeld, A.; Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1), 17-29. <https://doi.org/10.1109/TLT.2016.2616312>.
- Credé, M.; Kuncel, N. R. (2008). Study Habits, Skills, and Attitudes: The Third Pillar Supporting Collegiate Academic Performance. *Perspectives on Psychological Science*, 3(6), 425-453. <https://doi.org/10.1111/j.1745-6924.2008.00089.x>.
- Credé, M.; Tynan, M. C.; Harms, P. D. (2017). Much Ado About Grit: A Meta-Analytic Synthesis of the Grit Literature. *Journal of Personality and Social Psychology*, 113(3), 492-511. <http://dx.doi.org/10.1037/pspp0000102>.
- Crocker, J.; Karpinski, A.; Quinn, D. M.; Chase, S. K. (2003). When Grades Determine Self-Worth: Consequences of Contingent Self-Worth for Male and Female Engineering and Psychology Majors. *Journal of Personality and Social Psychology*, 85(3), 507-516. <https://doi.org/10.1037/0022-3514.85.3.507>.
- Fernández Mellizo-Soto, M. (2022). Estudio sobre el Abandono de los Estudios de Grado en el Sistema Universitario Español. Ministerio de Universidades, Gobierno de España. (https://www.universidades.gob.es/stf/s/universidades/ministerio/ficheros/Informe_Ejecutivo_abandono_fin2_comenmtadoMinistro.pdf).
- Fiallos Quinteros, J. C.; Jimenez Builes, J. A.; Branch Bedoya, J. W. (2022). Teaching and learning analytics applied to programming courses. *Campus Virtuales*, 11(1), 35-49. <https://doi.org/10.54988/cv.2022.1.880>.
- Foster, E.; Siddle, R. (2020). The Effectiveness of Learning Analytics for Identifying At-Risk Students in Higher Education. *Assessment & Evaluation in Higher Education*, 45(6), 842-854. <http://dx.doi.org/10.1080/02602938.2019.1682118>.
- García-Tinizaray, D.; Ordoñez-Briceño, K.; Torres-Díaz, J. C. (2015). Learning analytics para predecir la deserción de estudiantes a distancia. *Campus virtuales*, 3(1), 120-126.
- Guzmán-Valenzuela, C.; Gómez-González, C.; Rojas-Murphy Tagle, A.; Lorca-Vyhmeister, A. (2021). Learning analytics in higher education: A preponderance of analytics but very little learning?. *International Journal of Educational Technology in Higher Education*, 18(1), 1-19. <https://doi.org/10.1186/s41239-021-00258-x>.
- Herzog, S. (2006). Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression. *New directions for institutional research*, 2006(131), 17-33. <http://dx.doi.org/10.1002/ir.185>.
- Hosmer, D. W.; Lemeshow, S.; Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Hoboken, N.J. (USA): Wiley.
- Huang, A. Y. Q.; Lu, O. H. T.; Huang, J. C. H.; Yin, C. J.; Yang, S. J. H. (2020). Predicting Students' Academic Performance by Using Educational Big Data and Learning Analytics: Evaluation of Classification Methods and Learning Logs. *Interactive Learning Environments*, 28(2), 206-230.
- Ickson, T.; Kaplan, O.; Slobodin, O. (2020). Does Optimism Predict Academic Performance? Exploring the Moderating Roles of Conscientiousness and Gender. *Studies in Higher Education*, 45(3), 635-647. <http://dx.doi.org/10.1080/03075079.2018.1564257>.
- Ifenthaler, D.; Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68(4), 1961-1990. <https://doi.org/10.1007/s11423-020-09788-z>.
- James, T. (2018). Empirical Results Using Learning Analytics in the Classroom. *College Quarterly*, 21(2), n.d.
- Jayaprakash, S. M.; Moody, E. W.; Lauría, E. J. M.; Regan, J. R.; Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, 1(1), 6-47. <http://dx.doi.org/10.18608/jla.2014.11.3>.
- Karnik, A.; Kishore, P.; Meraj, M. (2020). Examining the linkage between class attendance at university and academic performance in an International Branch Campus setting. *Research in Comparative and International Education*, 15(4), 371-390. <https://doi.org/10.1177/1745499920958855>.
- Knight, S., & Shum, S. B. (2017). Theory and learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *The Handbook of Learning Analytics* (1st Edition, pp. 17-22). Alberta (Canada): Society for Learning Analytics Research (SoLAR). <https://doi.org/10.18608/hla17>.
- Koç, M.; Akçapınar, G.; Hasnane, M. N. (2021). Unfolding Students' Online Assignment Submission Behavioral Patterns using Temporal Learning Analytics. *Educational Technology & Society*, 24(1), 223-235.
- Larrabee Sønderlund, A.; Hughes, E.; Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 50(5), 2594-2618. <https://doi.org/10.1111/bjet.12720>.
- Lassibille, G.; Navarro Gómez, L. (2008). Why do higher education students drop out? Evidence from Spain. *Education Economics*, 16(1), 89-105. <https://doi.org/10.1080/09645290701523267>.
- Liesz, T. J.; Reyes, M. G. C. (1989). The Use of Piagetian Concepts to Enhance Student Performance in the Introductory Finance Course. *Journal of Financial Education*, 18, 8-14.
- Lonn, S.; Aguilar, S. J.; Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*, 47, 90-97. <https://doi.org/10.1016/j.chb.2014.07.013>.
- Lu, C.; Cutumisu, M. (2022). Online engagement and performance on formative assessments mediate the relationship between attendance and course performance. *International Journal of Educational Technology in Higher Education*, 19(1), 1-23. <https://doi.org/10.1186/s41239-021-00307-5>.
- Lu, O. H. T.; Huang, A. Y. Q.; Huang, J. C. H.; Lin, A. J. Q.; Ogata, H.; Yang, S. J. H. (2018). Applying Learning Analytics for the Early Prediction of Students' Academic Performance in Blended Learning. *Educational Technology & Society*, 21(2), 220-232.
- Mansouri, T.; ZareRavasan, A.; Ashrafi, A. (2021). A Learning Fuzzy Cognitive Map (LFCM) Approach to Predict Student Performance. *Journal of Information Technology Education: Research*, 20, 221-243. <https://doi.org/10.28945/4760>.
- Marcal, L. E.; Hennessey, J. E.; Curren, M. T.; Roberts, W. W. (2005). Do Business Communication Courses Improve Student Performance in Introductory Marketing?. *Journal of Education for Business*, 80(5), 289-294. <https://doi.org/10.3200/JOEB.80.5.289-294>.



- Marín Sánchez, M.; Infante Rejano, E.; Troyano Rodríguez, Y. (2000). El fracaso académico en la universidad: Aspectos motivacionales e intereses profesionales. *Revista Latinoamericana de Psicología*, 32(3), 505-517.
- Martínez-Carrascal, J. A.; Campuzano, J.; Sancho-Vinuesa, T.; Valderrama, E. (2019). Predicting student performance over time. A case study for a blended-learning engineering course. *Proceedings of the Learning Analytics Summer Institute Spain 2019: Learning Analytics in Higher Education*, 2415, 43-55. (<http://ceur-ws.org/Vol-2415/paper05.pdf>).
- Martínez-Carrascal, J. A.; Márquez Cebrián, D.; Sancho-Vinuesa, T.; Valderrama, E. (2020). Impact of early activity on flipped classroom performance prediction: A case study for a first-year Engineering course. *Computer Applications in Engineering Education*, 28(3), 590-605. <https://doi.org/10.1002/cae.22229>.
- Martínez-Carrascal, J. A.; Valderrama, E.; Sancho-Vinuesa, T. (2020). Combining clustering and sequential pattern mining to detect behavioral differences in log data: Conceptualization and case study. *Proceedings of the Learning Analytics Summer Institute Spain 2020: Learning Analytics. Time for Adoption?*, 2671, 25-38. (<http://ceur-ws.org/Vol-2671/paper04.pdf>).
- Milliron, M. D.; Malcolm, L.; Kil, D. (2014). Insight and Action Analytics: Three Case Studies to Consider. *Research & Practice in Assessment*, 9, 70-89.
- Pistilli, M. D.; Heileman, G. L. (2017). Guiding Early and Often: Using Curricular and Learning Analytics to Shape Teaching, Learning, and Student Success in Gateway Courses: Guiding Early and Often. *New Directions for Higher Education*, 2017(180), 21-30. <https://doi.org/10.1002/he.20258>.
- Regorz, A. (s. f.). How to interpret a Collinearity Diagnostics table in SPSS. REGORZ STATISTIK / REGORZ STATISTICS. (http://www.regorz-statistik.de/en/collinearity_diagnostics_table_SPSS.html#references).
- Richardson, M.; Abraham, C.; Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological bulletin*, 138(2), 353-387. <http://dx.doi.org/10.1037/a0026838>.
- Sackett, P. R.; Kuncel, N. R.; Arneson, J. J.; Cooper, S. R.; Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance?. *Psychological bulletin*, 135(1), 1-22. <http://dx.doi.org/10.1037/a0013978>.
- Schmitt, N.; Keeney, J.; Oswald, F. L.; Pleskac, T. J.; Billington, A. Q.; Sinha, R.; Zorzie, M. (2009). Prediction of 4-Year College Student Performance Using Cognitive and Noncognitive Predictors and the Impact on Demographic Status of Admitted Students. *Journal of Applied Psychology*, 94(6), 1479-1497. <http://dx.doi.org/10.1037/a0016810>.
- Scholes, V. (2016). The ethics of using learning analytics to categorize students on risk. *Educational Technology, Research and Development*, 64(5), 939-955. <http://dx.doi.org/10.1007/s11423-016-9458-1>.
- Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57(10), 1380-1400. <https://doi.org/10.1177/0002764213498851>.
- Stephens-Davidowitz, S. (2017). *Everybody lies: Big data, new data, and what the internet can tell us about who we really are*. New York (USA): HarperCollins New York.
- Tempelaar, D. T.; Rienties, B.; Nguyen, Q. (2017). Towards Actionable Learning Analytics Using Dispositions. *IEEE Transactions on Learning Technologies*, 10(1), 6-16. <https://doi.org/10.1109/TLT.2017.2662679>.
- Tempelaar, D. T.; Rienties, B.; Nguyen, Q. (2021). The contribution of dispositional learning analytics to precision education. *Educational Technology & Society*, 24(1), 109-122.
- Waddington, R. J.; Nam, S.; Lonn, S.; Teasley, S. D. (2016). Improving Early Warning Systems with Categorized Course Resource Usage. *Journal of Learning Analytics*, 3(3), 263-290. <https://doi.org/10.18608/jla.2016.33.13>.
- Westrick, P. A.; Le, H.; Robbins, S. B.; Radunzel, J. M. R.; Schmidt, F. L. (2015). College Performance and Retention: A Meta-Analysis of the Predictive Validities of ACT® Scores, High School Grades, and SES. *Educational Assessment*, 20(1), 23-45. <https://doi.org/10.1080/10627197.2015.997614>.