



Texto livre

ISSN: 1983-3652

Universidade Federal de Minas Gerais - UFMG

Norambuena, Javier Mella; Badilla-Quintana, María Graciela; Angulo, Yaranay López

Modelos predictivos basados en uso de analíticas de
aprendizaje en educación superior: una revisión sistemática

Texto livre, vol. 15, e36310, 2022

Universidade Federal de Minas Gerais - UFMG

DOI: <https://doi.org/10.35699/1983-3652.2022.36310>

Disponible en: <https://www.redalyc.org/articulo.oa?id=577170677011>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en [redalyc.org](https://www.redalyc.org)



Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Modelos predictivos basados en uso de analíticas de aprendizaje en educación superior: una revisión sistemática

Modelos predictivos basados en uso de analítica da

aprendizagem no ensino superior: uma revisão sistemática

Predictive models based on the use of learning analytics in higher education: a systematic review

Javier Mella Norambuena  ^{*1,2}, María Graciela Badilla-Quintana 
^{†3} y Yaranay López Angulo  ^{‡4,5}

¹Universidad Católica de la Santísima Concepción, Programa de Doctorado en Educación, Concepción, Chile.

²Universidad Técnica Federico Santa María, Departamento de Ciencias, Concepción, Chile.

³Universidad Católica de la Santísima Concepción, Centro de Investigación en Educación y Desarrollo, Concepción, Chile.

⁴Universidad Santo Tomás, Facultad de Ciencias Sociales y Comunicaciones, Escuela de Psicología, Concepción, Chile.

⁵Universidad de Concepción, Departamento de Psicología, Facultad de Ciencias Sociales, Concepción, Chile.

Resumen

Los métodos tradicionales de predicción del riesgo académico en ocasiones presentan limitaciones para la identificación oportuna, por otro lado, las Analíticas de Aprendizaje presentan ciertas ventajas. El objetivo de este estudio fue analizar características de los modelos predictivos basados en analíticas de aprendizaje en Educación Superior. Se realizó una revisión sistemática de las bases Web of Science, Scopus y Eric usando las palabras clave "analítica de aprendizaje" y "predicción". Se seleccionaron 12 investigaciones que cumplieron con los criterios de inclusión. Los resultados indicaron que el 100% de los estudios buscaron predecir el rendimiento académico, se incluyen variables de analíticas, sociodemográficas y sociocognitivas como predictoras. El sistema de gestión de aprendizaje más usado fue Moodle de cursos *blended learning* y online. Los estudios se desarrollaron principalmente en Europa; las muestras fueron de hasta 500 participantes de Ingeniería y Tecnología. El tipo de análisis más frecuente fue regresión en software R y SPSS. La mayoría logró un modelo de predicción grande ($R^2 > .30$). Se concluye que la construcción actual de modelos de predicción de abandono universitario posee importantes limitaciones.

Palabras clave: Modelo predictivo. Analítica de aprendizaje. Educación superior. Revisión sistemática.

Resumo

Os métodos tradicionais de previsão de risco acadêmico às vezes apresentam limitações para identificação oportuna. Por outro lado, a Analítica da Aprendizagem (*Learning Analytics*) apresenta certas vantagens. O objetivo deste estudo é analisar características de modelos preditivos baseados na análise da aprendizagem no Ensino Superior. Uma revisão sistemática dos bancos de dados Web of Science, Scopus e Eric foi conduzida usando as palavras-chave "análise de aprendizagem" e "predição". Foram selecionados doze estudos de pesquisa que preenchem os critérios de inclusão. Os resultados indicam que 100% dos estudos buscaram prever o desempenho acadêmico, incluindo variáveis analíticas, sociodemográficas e sociocognitivas como preditores. O sistema de gerenciamento de aprendizagem mais comumente utilizado foi o Moodle para aprendizagem combinada e cursos *on-line*. Os estudos foram realizados principalmente na Europa, sendo as amostras de até 500 participantes de Engenharia e Tecnologia. O tipo de análise mais frequente foi a regressão nos softwares R e SPSS. A maioria conseguiu um grande modelo de previsão ($R^2 > .30$). Conclui-se que a atual construção de modelos de previsão de abandono escolar tem limitações importantes.

Palavras-chave: Modelo preditivo. Analítica da aprendizagem. Educação superior. Revisão sistemática.

*Email: javier.mella1988@gmail.com

†Email: mgbadilla@ucsc.cl

‡Email: yara13190@gmail.com

 **Textolivre**
Linguagem e Tecnologia

DOI: 10.35699/1983-3652.2022.36310

Sección:
Artículos

Autor correspondiente:
Javier Mella Norambuena

Editor de sección:
Hugo Heredia Ponce
Editor de maquetación:
Anna Izabella Pereira

Recibido el:
24 de septiembre de 2021
Aceptado el:
7 de noviembre de 2021
Publicado el:
2 de febrero de 2022

Esta obra está bajo una
licencia «CC BY 4.0».



Abstract

Traditional prediction methods are time-consuming and limited in identifying students at academic risk in a timely manner. On the other hand, Learning Analytics has certain advantages. The aim of this study was to analyze characteristics of predictive models based on learning analytics in higher education. A systematic review of Web of Science, Scopus and Eric databases was conducted using the keywords "learning analytics" and "prediction". Twelve research studies that met the inclusion criteria were selected. The results indicated that 100% of the studies sought to predict academic performance, including analytical, sociodemographic and sociocognitive variables as predictors. The most used learning management system was Moodle for blended learning and online courses. The studies were mainly developed in Europe; the samples consisted of up to 500 participants from Engineering and Technology. The most frequent type of analysis was regression in R and SPSS softwares. Most achieved a large prediction model ($R^2 > .30$). It was concluded that the current construction of dropout prediction models has important limitations.

Keywords: Predictive model. Learning analytics. Higher education. Systematic review.

1 Introducción

Actualmente para implementar procesos de enseñanza-aprendizaje efectivos, y a gran escala se necesitan las Tecnologías de la Información y la Comunicación (TIC), especialmente en la Educación Superior (ZHANG; GHANDOUR y SHESTAK, 2020). Con la expansión de la revolución digital y un rápido cambio en las tecnologías, los datos educativos aumentan a un ritmo vertiginoso (HOODA, 2020; NGUYEN y col., 2021). En esta línea, la incorporación de entornos virtuales de aprendizaje (EVA), o sistemas de gestión de aprendizaje (LMS por su nombre en inglés), se introdujeron de manera importante en las universidades desde comienzos del siglo XXI (ROJAS CASTRO, 2017). Un LMS es un paquete de software basado en la web que está diseñado para planificar, implementar y evaluar el aprendizaje, facilitar la interacción de los estudiantes, dar retroalimentación sobre el rendimiento y gestionar las actividades de los estudiantes. Las universidades usan los LMS, siendo los más comunes Moodle, Edmodo, Canvas, Schoology, Blackboard Learn (ZHANG; GHANDOUR y SHESTAK, 2020).

El uso de entornos virtuales de aprendizaje genera una huella digital asociada a registros de entrada y acciones, entre otras, que aporta una gran cantidad de información acerca de las actividades de los usuarios (KERIMBAYEV y col., 2020). Estos datos usualmente se les conoce como big data educacionales o big data a secas, que son entendidos como un conjunto de datos cuyo tamaño excede la capacidad del software tradicional para su captura, gestión y análisis (PICCIANO, 2012; TULASI, 2013). El surgimiento de la huella digital trae consigo una serie de nuevas áreas de estudio, tales como las analíticas o la minería de datos educacionales acerca del comportamiento de los estudiantes en estos ambientes o sobre las interacciones que allí ocurren (KLAŠNJA-MILIĆEVIĆ; IVANOVIĆ y BUDIMAC, 2017). En un contexto de alto interés por parte de las universidades para mejorar resultados de aprendizaje y éxito académico, la aparición de la analítica contribuye a la satisfacción de estas necesidades.

En los últimos años, las Analíticas de Aprendizaje (en adelante AA) han ido incrementando progresivamente y emergieron como un campo separado de investigación. Las AA son una nueva herramienta tecnológica considerada como una práctica de minería de conjunto de datos de las instituciones de Educación Superior para generar inteligencia procesable a través de técnicas de modelado estadístico y predictivo con el propósito de mejorar la toma de decisiones, el resultado y éxito de los estudiantes (HOODA, 2020). Los estudios sobre AA se pueden agrupar en tres períodos: 1) los estudios realizados hasta el 2010, 2) los estudios realizados entre 2011-2014 y 3) los estudios realizados desde el 2015 (ŞAHIN y YURDUGÜL, 2019). Actualmente, su crecimiento ha sido reconocido en el campo de la investigación y se ha documentado ampliamente, donde herramientas analíticas sofisticadas son usadas para mejorar el aprendizaje y la educación, especialmente en los sistemas de gestión del aprendizaje (CECHINEL y col., 2020).

Las AA se comprenden como el uso, la evaluación, la obtención y el análisis de información estática y dinámica sobre los estudiantes y los contextos de aprendizaje, para el modelado, la predicción y la optimización de procesos de aprendizaje, entornos de aprendizaje, así como para la toma de decisiones

educativas (HOODA, 2020). Es decir, se entiende como la medición, recopilación, análisis y reporte de datos sobre los estudiantes y sus contextos, con el propósito de comprender y optimizar el aprendizaje y los entornos en los que ocurre y por tanto su objetivo es procesar datos educativos para ofrecer información significativa relacionada con los perfiles de los estudiantes, materiales de aprendizaje, y el contexto de aprendizaje (NGUYEN y col., 2021).

Las AA contribuyen a ayudar a Instituciones de Educación Superior (IES) en muchos aspectos del aprendizaje y la enseñanza utilizando los datos generados sobre los estudiantes y su experiencia de aprendizaje. Así, es posible convertir los datos sin analizar en información procesable que mejoran las decisiones relacionadas con la educación, informando y empoderando a las autoridades institucionales, académicos, estudiantes e investigadores de los problemas que se identifican mediante técnicas de minería de datos y aprovechar el juicio humano (GUERRA y col., 2020; HOODA, 2020). Tradicionalmente, estas decisiones eran tomadas por autoridades universitarias basadas en presunciones e hipótesis, procesos que además consumen mucho tiempo y limitan su calidad (HOODA, 2020).

Respecto a los fenómenos educativos clave a los cuales aportan las AA, y áreas más frecuentemente enfocadas por parte de los investigadores sobre su uso en IES, son los métodos para mejorar el desempeño, el comportamiento, predecir el rendimiento y la retención de los estudiantes, especialmente considerando que muchos estudiantes no completan sus cursos (GUERRA y col., 2020). En este contexto, las AA pueden proveer de información relevante como la identificación de estudiantes en riesgo, la posibilidad de construir un apoyo adaptativo en su viaje de aprendizaje o proporcionarles apoyo adicional para hacer frente a los problemas académicos, requisitos y expectativas (IFENTHALER y YAU, 2020b). Las AA ayudan a comprender el comportamiento de estudio, permiten brindar retroalimentación personalizada, facilitan la comprensión por parte de los docentes de los procesos de aprendizaje de sus estudiantes y orientan al diseño diferenciado y específico de intervenciones (IFENTHALER y YAU, 2020b; ŞAHİN y YURDUGÜL, 2019). En este sentido, Zhang, Ghandour y Shestak (2020), concluyó que las AA se pueden utilizar para predecir el éxito de los estudiantes y estimular mejores resultados durante el estudio.

Los modelos explicativos del éxito académico en la Educación Superior incluyen factores sociodemográficos de los estudiantes, por ejemplo género, origen étnico, antecedentes familiares; su capacidad cognitiva o rendimiento académico previo, como el promedio de notas; atributos individuales, como los rasgos personales e influencias contextuales motivacionales o psicosociales; así como factores relacionados con el curso, como el aprendizaje activo, la atención, o factores ambientales relacionados con la integración académica y social de apoyo. La posibilidad de recopilar y almacenar datos para los factores mencionados anteriormente y combinarlos constituye una enorme posibilidad que en la actualidad permiten las AA (IFENTHALER y YAU, 2020b).

Los sistemas de gestión del aprendizaje (SGA) suelen centrarse en la organización de cursos de los profesores e incluyen la gestión de los estudiantes en un entorno de aprendizaje en línea moderno. Las instituciones educativas utilizan sistemas tecnológicos como Moodle, Edmodo, Canvas, Schoology, Blackboard Learn, entre otros (ZHANG; GHANDOUR y SHESTAK, 2020). Al acceder con una cuenta personal, la actividad de cada estudiante queda en un archivo de registro. Las AA básicamente buscan responder a preguntas como: de qué, por qué, quién y cómo, para la optimización de estos entornos de aprendizaje utilizando los SGA como fuentes de datos que ofrecen mucha información y recursos a los estudiantes, tutores, autoridades, instituciones, investigadores y diseñadores.

Aunque se ha evidenciado la importancia de las AA para evaluar y predecir el desempeño de los estudiantes, mejorar las prácticas de aprendizaje y enseñanza, y tomar decisiones más rápidas basadas en datos (HOODA, 2020) es un área del conocimiento que aún tiene varias limitaciones. Por ejemplo, el desarrollo de investigación es alto en países anglosajones pero es incipiente en la región latinoamericana (CECHINEL y col., 2020) y autores como Ifenthaler y Yau (2020b) señalan que se requiere ampliar la evidencia sobre la efectividad de intervenciones basadas en las AA. Para esto es importante conocer los modelos predictivos disponibles en el área.

Respecto a la sistematización de la literatura para identificar la productividad de las AA en contexto de educación superior, se han realizado algunas aproximaciones teóricas, pero estas se han focalizado en cuestiones relacionadas con facilitar el éxito del estudio en la continuación y finaliza-

ción, del aprendizaje de los estudiantes (IFENTHALER y YAU, 2020a), en la eficacia de las AA en la intervenciones en términos de retención de estudiantes o éxito académico (LARRABEE; HUGHES y SMITH, 2019), y otras relacionadas con identificar los beneficios de utilizar las AA en la Educación Superior (BANIHASHHEM y col., 2018). Sin embargo, no se ha encontrado evidencia de sistematizaciones que identifiquen los modelos desarrollados y probados empíricamente basados en AA para explicar fenómenos en la Educación Superior. Por lo anterior, en este artículo se propone sistematizar la evidencia disponible sobre las AA que aporte información para responder a la siguiente pregunta de investigación: ¿qué características tienen los modelos predictivos basados en AA usados en Educación Superior? Los objetivos específicos son: (1) Determinar las características que tienen los modelos predictivos basados en AA en Educación Superior respecto de variables predictoras consideradas, fenómeno que busca explicar, tipo de LMS como fuente extracción de datos, tipo de modelo, software para el análisis y resultado del modelo; (2) Describir las características que tienen los participantes en los modelos predictivos que usan AA en Educación Superior.

2 Método

Para la revisión y selección de artículos, se utilizó la metodología de revisión sistemática basada en las directrices internacionales de PRISMA (MOHER y col., 2015), en dos procesos consecutivos. El primero tuvo por objetivo identificar la muestra de estudios y el segundo elaborar una matriz que da cuenta del protocolo de extracción de la información de los estudios para su posterior análisis.

2.1 Proceso 1. Selección de la muestra de estudios

El proceso de selección de los artículos implicó 5 fases específicas: identificación, duplicado, elegibilidad, selección y sesgo, las que se pueden apreciar en la Figura 1. La fase de identificación consistió en la búsqueda de artículos en las bases de datos Web of Science (WOS), Scopus y Eric usando como palabras clave “analíticas de aprendizaje” y “predicción” o “predictivo” en inglés, las cuales fueron validadas por 6 jueces expertos. El período de búsqueda no se restringió debido a lo acotado de los resultados en indagaciones preliminares. Se aplicaron los filtros proporcionados por cada base de datos. La exploración en las bases de datos y búsqueda de los estudios se realizó el 22 de mayo de 2021. La fase de duplicado implicó la eliminación de estudios repetidos dejando solo uno. La fase de cribado consistió en una revisión por el investigador principal y dos jueces independientes, a los cuales se les presentó como protocolo para llevar a cabo la eliminación de estudios en esta fase la descripción del objetivo de la revisión sistemática y las palabras clave que debían identificar en el título y resumen. Se eliminaron artículos que no contenían las palabras clave o no respondían al objetivo de la presente investigación. La etapa de selección implicó la descarga de los archivos en formato pdf de los estudios y la lectura completa para la aplicación de los criterios de exclusión previamente establecidos. Se incluyeron investigaciones empíricas cuantitativas sobre AA con diseños predictivos en estudiantes universitarios cuyos datos para el análisis fueron extraídos de LMS institucionales. El proceso completo de esta parte del método se puede observar en la Figura 1.

2.2 Proceso 2. Proceso de extracción de la información de la muestra de estudios seleccionados

En la Tabla 1 se presenta la descripción de la información extraída de los estudios que fueron incluidos en esta revisión sistemática y en la Tabla 2 la matriz de datos.

3 Resultados

Se han separado los resultados según variable y se presentan a continuación.

Variable predicha en los modelos que usan AA

En los modelos predictivos que usan AA, se identificó que en el 100% de las investigaciones fue el rendimiento académico de los estudiantes la variable predicha.

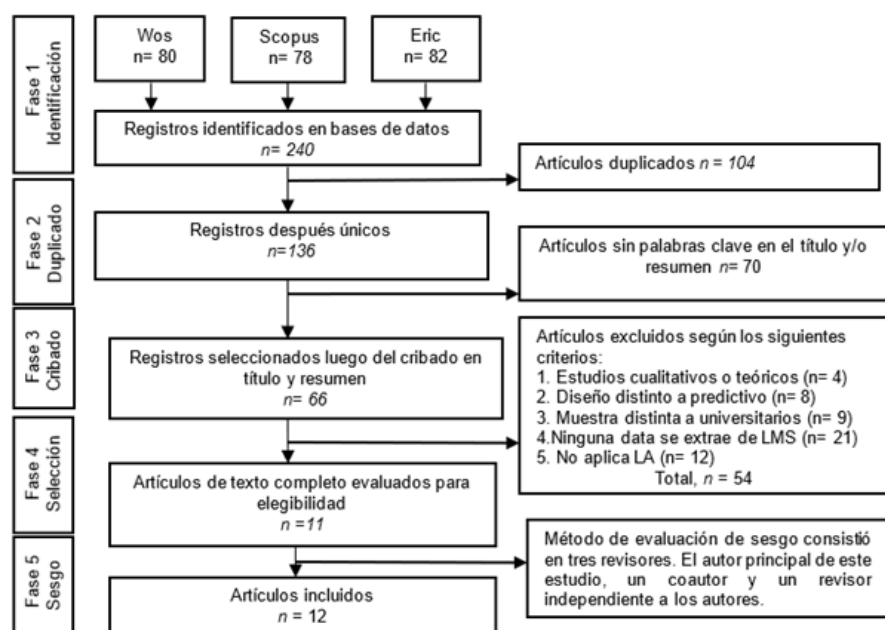


Figura 1. Flujograma de proceso completo identificación de estudios.

Fuente: Elaboración propia.

Tabla 1. Matriz explicativa de la información que fue extraída de la muestra de estudios seleccionados.

Información extraída	Descripción
ID	Identificador de la muestra de estudios seleccionados para la presentación de los resultados de esta RSL.
Cita	Según formato de revista.
Variable predicha (dependiente)	Variable que busca predecir el modelo hipotetizado.
Variables predictoras (independiente)	Variables incluidas en los modelos predictivos correspondientes a las extraídas de los sistemas LMS institucionales.
Tipo de LMS	Tipo de sistema de gestión del aprendizaje donde son extraídos los datos de analíticas de aprendizajes.
Muestra	n: Corresponde al total de la muestra usada en el estudio. país: corresponde al país de la institución donde se realizó el estudio.
Modalidad de estudio	Formato de educación: Híbrida, <i>blended-learning</i> , presencial.
Tipo de modelo	Corresponde al análisis estadístico usado para probar el modelo hipotetizado (ejemplo, regresión, redes neuronales etc..).
Software	Software estadístico con el que se realizó el tratamiento de los datos y análisis respectivos para probar el modelo hipotetizado.
Resultado del modelo	Corresponde a los estadígrafos o coeficientes que dan cuenta del nivel de predicción del modelo (ejemplo: nivel de significancia dado por el valor p, porcentaje de predicción; coeficiente de determinación R ²).

Fuente: Elaboración propia.

Categorías de variables predictoras usadas en los modelos de AA

Los resultados de esta revisión muestran que las variables incluidas en los modelos predictivos que usan AA en Educación Superior integran 3 grandes categorías: (1) variables sociodemográficas; (2) datos extraídos de LMS; y (3) variables sociocognitivas. El 100% de los estudios incluyen diferentes datos de extracción de los LMS, mientras que dos investigaciones incluyeron además variables sociodemográficas (ID: 2, 10), y dos investigaciones incluyeron variables sociocognitivas (ID: 7, 11).

Resultados sobre el tipo de LMS

El tipo de LMS identificado en el 92% (n=11) de los estudios fue Moodle. Un estudio usó datos extraídos de la LMS Blackboard (ID: 11).

Resultados sobre la modalidad de estudio

Las modalidades de estudio donde se desarrollaron las investigaciones corresponden a: 5 investigaciones en modalidad *Blended Learning* (ID: 1, 2, 3, 4, 11); 5 corresponden a modalidad Online (ID: 12, 5, 8, 9, 10); y en dos estudios no se especificó la modalidad (ID: 7, 6).

Resultados sobre la muestra: país

Los estudios se realizaron en nueve países distintos (Ver tabla 2) de los cuales fueron agrupados en categorías según continente (Ver Figura 2).

Tabla 2. País de los participantes de los estudios.

País	ID	Frecuencia
Países bajos	1,8,11	3
Australia	2,5	2
Italia	4	1
España	7	1
Turquía	12	1
Taiwan	3	1
Israel	9	1
Ecuador	6	1
Estados Unidos	10	1

Fuente: Elaboración propia.

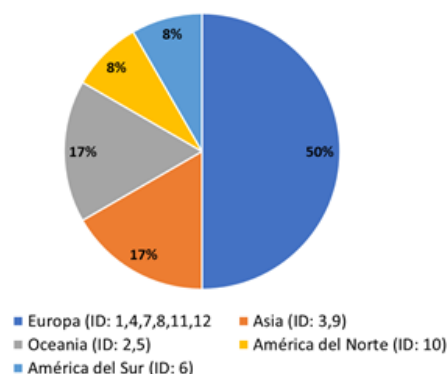


Figura 2. País de los participantes agrupados por continente.

Fuente: Elaboración propia.

Resultados sobre la muestra total

Para el tamaño muestral se establecieron 5 rangos (ver Figura 3), siendo el tamaño de muestra más usado en los estudios el rango entre 101 a 500 estudiantes (34%).

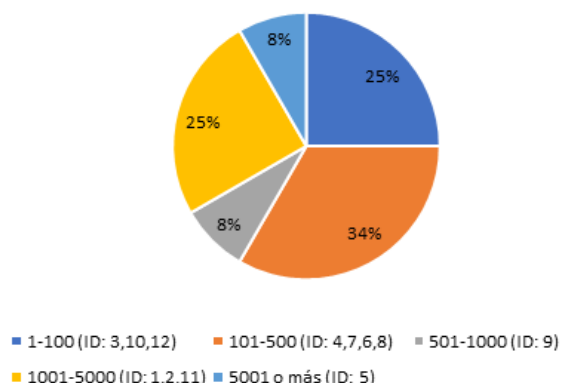


Figura 3. Rangos de tamaños de muestras usadas en los estudios.

Fuente: Elaboración propia.

Resultados sobre la carrera

En relación con las carreras que fueron parte de los estudios, dos investigaciones incluyeron carreras del área de la salud (ID: 7, 8). También se identificó un estudio que integró una carrera perteneciente al área de educación (ID: 4), un estudio que incluyó el área de economía (ID: 11), y otro estudio que incluyó una carrera del área de ingeniería. Además, en 7 investigaciones, no se especificaron qué carreras incluyeron (ID: 12, 1, 2, 3, 5, 9 y 10).

Resultados sobre la muestra: cursos

Los cursos a los que pertenecían los estudiantes fueron clasificados según su correspondencia con las áreas OCDE (Ver Figura 4). La categoría más frecuente incluyó asignaturas del área de Ingeniería y Tecnología.

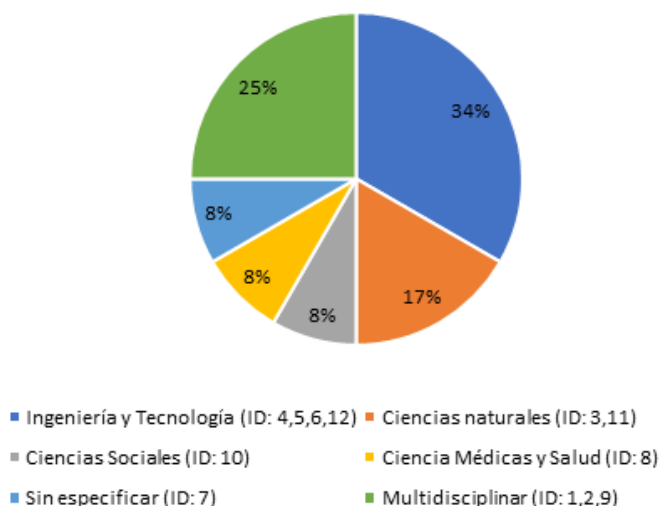


Figura 4. Cursos por área OCDE.

Fuente: Elaboración propia.

Resultados sobre el tipo de análisis

Respecto a los tipos de análisis de predicción empleados en los estudios, el más frecuente fue el análisis de regresión lineal, se identificaron siete estudios (58%) que los usaron en sus modelos predictivos (ID: 1, 2, 3, 7, 9, 10, 11), la red neuronal se usó en dos estudios (17%) (ID: 5 y 6) y otros análisis identificados fueron la Distancias Levenshtein entre las cadenas relativas a trayectorias (ID: 4); y el Algoritmo KNN (ID: 8), ambos tipos de análisis usados en un estudio. Además, un estudio usó más de un tipo de análisis (ID: 12).

Resultados sobre el tipo de software

Con relación al software usado para realizar los análisis de datos, los más frecuentemente usados fueron el software R (ID: 1, 6, 8) y el SPSS (ID: 2, 9 y 10), ambos usados por tres investigaciones respectivamente. Otros estudios usaron el software Python (ID: 5), Orange Data Mining software (ID: 12). Un estudio usó dos softwares (ID: 11), específicamente SPSS y Mplus. Tres investigaciones no especificaron el software que usaron para los análisis (ID: 3, 4).

Resultados de los modelos de predicción

En los modelos identificados para predecir el rendimiento académico, se observó, en primer lugar, que en cuatro estudios hubo variabilidad en los rangos de varianza explicada según los modelos que fueron probados. En dos de los estudios esta variabilidad fue grande donde los modelos explicaron entre el 8% y el 37% (ID: 1) y el 7% y 43% de la varianza (ID: 11); mientras que otros dos la variabilidad de los modelos fue pequeña entre ellos, fluctuando entre 30% y 38% varianza explicada (ID: 3) y 88,3% y un 90,9% de la varianza explicada (ID: 8). En segundo lugar, se encontró un estudio que logró una predicción media dado que el modelo explicó un 21% de la varianza. En tercer lugar, se observó que siete estudios lograron una predicción grande (ID: 4, 5, 6, 7, 9, 10, 12).

Los modelos que explican más de un 80% de la varianza son investigaciones realizadas en cursos de modalidad online. Es posible observar que en cursos formativos en modalidad *Blended Learning* los modelos predictivos explicaron entre el 7% y el 77% (ID: 1, 2, 3, 4, 11); mientras que en los cursos en modalidad Online explicaron entre un 32% y un 90,9% (ID: 5, 8, 8, 10 y 12). Este análisis muestra que los modelos predictivos basados en analítica de aprendizaje son más efectivos para explicar la variable de rendimiento académico en cursos universitarios de modalidad online.

Tabla 3. Matriz de extracción de información de la muestra de estudios seleccionados.

ID	Cita	Variable predicha	Variables predictoras	Tipo LMS/ Modalidad	Muestra	Tipo de modelo / Software / Resultado del modelo
1	(CONIJN y col., 2017)	Rendimiento académico	Datos extraídos de LMS: 1 Número total de clics 2 Número de sesiones en línea 3 Tiempo total en línea 4 Número de visitas a la página del curso 5 Irregularidad del tiempo de estudio 6 Irregularidad del intervalo de estudio 7 Mayor periodo de inactividad (min) 8 Tiempo hasta la primera actividad (min) 9 Tiempo medio por sesión (min) 10 Número de recursos consultados 11 Número de enlaces vistos 12 Número de páginas de contenido vistas 13 Número de visitas a los debates 14 Número total de mensajes de discusión 15 Número de cuestionarios iniciados 16 Número de intentos por cuestionario 17 Número de cuestionarios aprobados 18 Número de visitas a los cuestionarios 19 Número de tareas enviadas 20 Número de asignaciones (envío) 21 Número de ediciones del wiki 22 Número de visualizaciones del wiki	Tipo de LMS: Moodle Modalidad: <i>Blended Learning</i>	País: Países Bajos N total: 4.989 Carrera: No específica Curso: matemáticas, física y psicología	Tipo de modelo: Regresión lineal múltiple Software: R Resultado: Los modelos explicaron entre el 8% y 37% de la varianza.
2	(GAŠEVIĆ y col., 2016)	Éxito académico	Sociodemográficas 1 Edad 2 Sexo (hombre/mujer); 3 Estudiante internacional (sí/no); 4 Idioma hablado en casa (inglés/idioma distinto del inglés); 5 Lejanía del hogar (urbano/no urbano); 6 Acceso a la universidad (estudiante a tiempo completo/parcial); 7 Matriculación previa en el mismo curso (sí/no). 8 acceso al inicio del curso (acceso temprano/no acceso al curso/acceso tardío) Datos extraídos de LMS: 1 Foros 2 Inicios de sesión en el curso recursos 3 Envío de archivos Turnitin, tareas, libro, cuestionarios, comentarios, mapa, aula virtual, lecciones y chat.	Tipo de LMS: Moodle Modalidad: <i>Blended Learning</i>	País: Australia N total: 4134 Carrera: No se específica Curso: contabilidad comunicación, informática, economía, diseño gráfico, marketing, matemáticas biología.	Tipo de modelo: Regresión lineal múltiple y modelos de regresión logística Software: SPSS 1. Los modelos en promedio explicaron el 21% de la varianza.

3	(LU y col., 2018)	Rendimiento académico final	Datos extraídos de LMS: - Número de días que muestra actividad - Número de actividades que realiza - Número de días que ve vídeos - Número de vídeos que ve - Número de vídeos que ve completamente - Número de veces que hace clic en Buscar - Número de vídeos en los que hace clic en "Pausa" - Número de vídeos en los que hace clic en "Detener" - Número de veces que hace clic en "Reproducir" - Número de veces que un alumno hace clic en "Buscar hacia atrás" - Número de veces que realiza prácticas en línea - Número de unidades de Cálculo que un practica - Número de días que practica en línea - Suma de días de unidades de Cálculo practicadas - Puntuación semanal de la práctica en línea - Puntuación semanal de los deberes - Puntuación semanal en los exámenes - Número de veces que el alumno participa en la tutoría extraescolar * por semana MOOCs en línea	Tipo de LMS: Open Edx) MOOCs, utiliza datos del lms y de aplicación OAS through Maple T.A Modalidad: <i>Blended Learning</i>	País: Taiwán N total: 59 Carrera: No se es-pecifica Curso: Cálculo	Tipo de modelo: regresión de componentes principales Software: No declara Los modelos explicaron entre el 30% y 38% de la varianza.
4	(MIRANDA y VEGLIANTE, 2019)	Resultados de aprendizaje	Datos extraídos de LMS: Ruta de navegación de los alumnos entre los contenidos.	Tipo de LMS: Moodle (moocs) Modalidad: <i>Blended Learning</i>	País: Italia N total: 407 Carrera: Ciencias de la Educación. Curso: Ciencias básicas computacionales	Tipo de modelo: Distancias de Levenshtein entre las trayectorias Software: No declara Resultado: 1. El porcentaje de estimación del modelo fue de 77%.
5	(MONLLAÓ OLIVÉ y col., 2020)	Rendimiento académico	Datos extraídos de LMS: - Valoraciones positivas en las entradas de la base de datos - Valoraciones positivas en los mensajes del foro - El estudiante intentó al menos un cuestionario - Porcentaje de pruebas intentadas en el curso - El alumno ha fallado un intento de cuestionario	Tipo de LMS: Moodle (moocs) Modalidad: Online	País: Australia N total: 46.895 Carrera: No aplica Curso: Learn Moodle	Tipo de modelo: red neuronal Software: Python El porcentaje de precisión promedio del modelo fue de 88.81%.

6	(SALGADO y col., 2019)	Rendimiento académico	Datos extraídos de LMS: 1 Irregularidad del tiempo de estudio 2 Irregularidad del intervalo de estudio 3 Mayor periodo de inactividad 4 Tiempo hasta la primera actividad (min) 5 Tiempo promedio por sesión (min) 6 Número de intentos por prueba 7 Número de cuestionarios vistos 8 Número de cuestionarios pasados 9 Número de wiki vistas 10 Número de consultas al profesor 11 Participación en los foros 12 Participación en chat 13 Participación en video-colaboración 14 Número de recursos vistos 15 Número de enlaces vistos 16 Número de páginas de contenido vistas 17 Número de publicaciones de discusión vistas 18 Número total de publicaciones de discusión 19 Numero de cuestionarios iniciados 20 Número de asignaciones enviadas 21 Número de presentación vistas 22 Número de ediciones en la wiki	Tipo de LMS: Moodle (moocs) Modalidad: No declara	País: Ecuador N total: 300 Carrera: Sistemas Cursos: Programación I, Programación II, Estructura de Datos, Programación Distribuida, Aplicaciones de Minería de Datos.	Tipo de modelo: red neuronal Software: R El porcentaje de precisión promedio del modelo fue de 75.28%.
7	(SAIZ MANZANARES y col., 2018)	Resultados finales de aprendizaje	Datos extraídos de LMS: 1 Accesos a la información complementaria 2 Accesos a orientaciones para realizar actividad 3 Accesos a información sobre conceptos 4 Accesos al feedback Variables sociocognitivas: Estrategias Metacognitivas	Tipo de LMS: Moodle Modalidad: No declara	País: España N total: 122 Carrera: Terapia Ocupacional y Enfermería Curso: no especifica	Tipo de modelo: Regresión Lineal Software: No declara Los modelos en promedio explicaron en 64.1% de la varianza.
8	(SAQR; FORS y NOURI, 2019)	Rendimiento académico	Datos extraídos de LMS: 1 Temporalidad (participación diaria, participación semanal participaciones tempranas que corresponde al número de mensajes en el primer día de la semana; participaciones tardías; tendencias de participación a lo largo del día, la semana, el curso y todo el año). 2 Duración de la conexión 3 Número total de interacciones de un estudiante 4 Tamaño de los mensajes.	Tipo de LMS: Moodle Modalidad: Online	País: Holanda N total: 183 Carrera: Odontología Curso: Sistemas Corporales, Cirugía Dental, Neurociencia, y Principios de las Ciencias Odontológicas.	Tipo de modelo: KNN algorithm Software: R Los modelos explicaron entre el 88,3% y un 90,9% de la varianza.

9	(SOFFER y COHEN, 2019)	Nota final del curso	Datos extraídos de LMS: 1 Actividad de vídeo en días 2 Vídeos vistos (%) 3 Minutos de vídeo vistos (%) 4 Escribir en los foros 5 Lectura de los mensajes del foro 6 Entradas en la página de inicio del curso 7 Entradas de la unidad de aprendizaje 8 Entradas en la página media de la unidad 9 Entradas de material adicional 10 Tareas entregadas (%) 11 Entradas totales	Tipo de LMS: Moodle (moocs) Modalidad: Online	País: Israel N total: 646 Carrera: Programa de estudios complementarios Cursos: curso en el ámbito de los estudios de Asia Oriental; curso en el ámbito de los estudios africanos; curso en Historia de la Mirada; curso en las Bases Celulares.	Tipo de modelo: análisis de regresión Software: SPSS El modelo final explicó el 33% de la varianza.
10	(STRANG, 2016)	Rendimiento académico	Datos extraídos de LMS: 1 Entradas al curso, 2 Lecciones vistas, 3 Actividad de las tareas, 4 Publicaciones en el foro Variables demográficas para la 1 Edad 2 Género 3 Cultura de los estudiantes	Tipo de LMS: Moodle (moocs) Modalidad: Online	País: Estados Unidos N total: 45 Carrera: No específica. Curso: profesionalidad de división superior y gestión de recursos humanos	Tipo de modelo: análisis de regresión pasos hacia atrás (stepwise backward) Software: SPSS El modelo final explicó el 32% de la varianza.
11	(TEMPELAAR, D.; RIENTIES y NGUYEN, 2020)	Rendimiento académico	Datos extraídos de LMS: 1 BClicks como el número total de clics en BlackBoard. Variables sociocognitivas: 1 Emociones de logro, 2 Emociones epistémicas 3 Objetivos de logro 4 Motivación y compromiso 5 Actitudes hacia el aprendizaje 6 Enfoques del aprendizaje 7 Motivaciones académicas	Tipo de LMS: blackboard Modalidad: Blended Learning	País: Países Bajos. N total: 1080 estudiantes Carrera: programa de negocios y economía Curso: introducción a las matemáticas y la estadística	Tipo de modelo: regresión múltiple Software: SPSS y Mplus Los modelos explicaron entre el 7% y 43% varianza.

12 (AKÇAPINAR; ALTUN y AŞKAR, 2019)	Rendimiento académico al final del trimestre	Datos extraídos de LMS:		Tipo de LMS: Moodle Modalidad: On-line	País: Turquía N total: 76 Carrera: No se es-pecifica Curso: Hardware de Computadora	Tipo de modelo: Multimétodo (Naive Bayes, Árbol de clasificación, Bosque aleatorio, Red neuronal de máquinas de vec-tores de soporte, Reglas CN2) Software: Orange data mining soft-ware (software de minería de datos de Orange)
		1	El número total de sesiones			
		2	La cantidad total de tiempo (en minu-tos) dedicado al aprendizaje			
		3	El número de días diferentes en los que se registra el registro de aprendizaje			
		4	El número total de visitas realizadas a los materiales de aprendizaje.			
		5	El número total de visitas realizadas a las páginas de contenido.			
		6	El número total de visitas realizadas a la página de notificación.			
		7	El número total de visitas realizadas a la página de anuncios.			
		8	El número total de visitas realizadas a la página de la publicación.			
		9	El número total de visitas realizadas a la página de discusión.			
		10	El número total de publicaciones crea-das			
		11	La cantidad promedio de tiempo (en segundos) dedicado a escribir publica-ciones			
		12	El número de días diferentes en los que se escriben las publicaciones.			
		13	El número de etiquetas nuevas creadas al escribir una publicación.			
		14	El número de etiquetas utilizadas en las publicaciones escritas.			
		15	La cantidad de veces que se usa copiar y pegar al escribir publicaciones			
		16	El número medio de pulsaciones de te-clas que realiza el alumno mientras es-cribe una publicación			
		17	El número medio de teclas de retro-ceso utilizadas por el alumno mientras escribiendo una publicación			
		18	La cantidad promedio de veces que la página perdió el foco al escribir una publicación.			
		19	El número de puestos abiertos			
		20	El número de comentarios creados			
		21	El número de publicaciones calificadas			
		22	El número de comentarios calificados			
		23	El número total de respuestas escritas en el foro de discusión			
		24	El número total de veces para abrir res-puestas en el foro de discusión.			
		25	El número de respuestas calificadas en el foro de discusión			
		26	El número de preguntas calificadas en el foro de discusión			

Fuente: Elaboración propia.

4 Discusión

Este estudio se propuso realizar una revisión exhaustiva de la literatura sobre los modelos predictivos en la Educación Superior basado en el uso de AA. Se presenta la discusión de los resultados de cada información extraída de las investigaciones para responder al propósito de este estudio, además se presentan limitaciones y futuras líneas de investigación.

Discusión sobre las categorías de variables predictoras usadas en los modelos de AA

Los resultados de esta revisión muestran que las variables incluidas en los modelos predictivos que usan AA en todos sus casos integran datos extraídos de las LMS institucionales, sin embargo, solo dos estudios incluyeron variables sociodemográficas y otros dos, variables sociocognitivas. Existe evidencia que los estudios en el campo de la AA que incluyen datos demográficos de los estudiantes y datos proporcionados por las LMS son identificadores eficaces del rendimiento "en riesgo". Sin embargo, la construcción de modelo con solo AA o aquellos que integran variables sociodemográficas no permiten una intervención personalizada para estos estudiantes (ZHENG y col., 2020). La información generada por estos modelos predictivos puede no ser adecuada para intervenciones pedagógicas debido a la incapacidad de explicar por qué los estudiantes muestran estos patrones de comportamiento. Para poder explicar el comportamiento de los estudiantes en las LMS se requiere incorporar variables sociocognitivas propias de la disposición al aprendizaje (por ejemplo, la autorregulación y las emociones) en los modelos convencionales de AA (TEMPELAAR, D. T.; RIENTIES y NGUYEN, 2017). Los modelos predictivos utilizados para los datos educativos siguen siendo inadecuados para ofrecer una imagen real del rendimiento académico de los estudiantes. Esto se debe a la falta de investigación sobre los modelos predictivos y los factores que afectan al rendimiento académico de los estudiantes (ZULKIFLI; MOHAMED y AZMEE, 2019). En definitiva, en el futuro, los modelos predictivos utilizados para el rendimiento académico de los estudiantes deberían tener en cuenta el último método de valoración de las evaluaciones basado en un sistema educativo moderno que haga hincapié en las habilidades blandas, las habilidades interpersonales y las capacidades de pensamiento de alto nivel.

Discusión sobre el tipo de LMS

En relación con los tipos de LMS en la mayoría de los estudios se usó Moodle y solo uno usó Blackboard. Uno de los elementos importantes de un LMS es la capacidad de mejorar de forma óptima la calidad del aprendizaje y aplicar un sistema de aprendizaje interactivo. Una revisión sistemática se propuso analizar las características de flexibilidad, facilidad de uso y accesibilidad de diferentes LMS que pueden utilizarse para los procesos de enseñanza y aprendizaje en el contexto de las instituciones de educación superior (KASIM y KHALID, 2016). Los resultados identificaron 6 LMS diferentes, donde tres de ellos eran de código abierto (Moodle, Sakai y ATutor), mientras que otras tres eran comerciales (Blackboard, SuccessFactors y Sum Total). Esto podría explicar por qué en los estudios identificados en la presente revisión, el LMS más frecuente fue Moodle y solo un estudio usó un LMS comercial, además, este último tiene menos disponibilidad de características para los usuarios. Específicamente el LMS Moodle se caracteriza por: (1) basado en la nube, (2) flexible, (3) fácil de usar, (4) capaz de integrarse con otros sistemas, (5) accesible, (6) uso amigable, (7) interacción sincrónica y asincrónica, (8) posibilidad de ver quién está en línea, (9) espacio personal para el borrador y diarios, así como para gestionar la información personal y privada, y (10) permite enviar y recibir mensajes personales con otros usuarios; sin embargo, el LMS Blackboard, solo dispone de las características 1, 2, 5, 6 y 7. Otra revisión teórica, con un propósito similar presentó una visión general de las opciones de LMS disponibles para las universidades (DOBRE, 2015). Según los antecedentes descritos, aunque un número importante de universidades usa el LMS Blackboard, este ha comenzado a enfrentarse desde 2002 a una competencia muy fuerte, especialmente de otros LMS comerciales (Canvas, Desire2Learn), así como de LMS de código abierto (Moodle y Sakai). Los LMS comerciales pueden ser fáciles de desarrollar y utilizar, pero tienen un costo elevado; por otra parte, los LMS de código abierto son gratuitos, pero el costo de mantenimiento y mejora son elevados. En definitiva, el factor crucial que influye en la satisfacción de los estudiantes es que las características disponibles en un LMS satisfagan sus necesidades y faciliten su uso. En el caso de los investigadores

es importante que los LMS permitan integrar acciones que faciliten el monitoreo del comportamiento de los estudiantes en las plataformas, lo que permitirá extraer información valiosa para predecir su rendimiento y consecuentemente disponer de esta información oportunamente para desplegar acciones de prevención de un rendimiento insuficiente. Finalmente, la efectividad y eficiencia de un LMS puede identificar y ayudar a resolver problemas potenciales (KADIR y AZIZ, 2016).

Discusión sobre la modalidad de estudio

Los resultados de esta revisión evidenciaron que las modalidades de estudio donde se desarrollaron las investigaciones corresponden a *Blended Learning* y Online. Llama la atención que ninguno de los estudios se realizó en estudiantes de carreras en modalidad presencial, es decir, en carreras de clases convencionales. Estas, si bien se realizan en la interacción misma del aula presencial, en general, cuentan con una plataforma basada en LMS de los cursos que están desarrollando los estudiantes en un determinado semestre. La investigación previa describe la enseñanza tradicional en el aula como aquella basada en técnicas de enseñanza monótonas y que carece de interacción, por lo que los estudiantes pierden la curiosidad por comprender el propósito de estudiar (PRIYAADHARSHINI y col., 2020). Los estudiantes esperan nuevas técnicas de enseñanza, asignaciones digitales y modelos de evaluación desafiantes. En la educación superior, se han introducido varios procesos nuevos de enseñanza-aprendizaje para motivar a los estudiantes y fomentar la práctica del autoaprendizaje, lo que facilita el desarrollo de mejores habilidades y conocimientos. Quizás es por esto, que este estudio no identificó investigaciones que usen AA en este ambiente presencial de aprendizaje. Sin embargo, con las diversas innovaciones en las TIC para la educación superior, el aprendizaje basado en diferentes métodos que incluyen el uso de las LMS se reconoce como una de las estrategias innovadoras de enseñanza-aprendizaje que han ido ganando interés en las universidades tradicionales. Por lo anterior, es esperable que las AA sean parte de las carreras que se desarrollan en la presencialidad, dado que como se ha avanzado, cuentan de igual forma con apoyo de diferentes LMS desde donde es posible extraer información de sus trayectorias académicas. Con el avance de la tecnología educativa y la necesidad de involucrar a los estudiantes del siglo XXI en las metas de aprendizaje, se requiere modificar el paradigma del aula de educación superior integrando los nuevos estilos de enseñanza innovadores que implican el uso de las TIC, contribuyendo a que los procesos de enseñanza-aprendizaje se agilicen con aulas inteligentes equipadas con actividades colaborativas donde los docentes aprovechan las herramientas TIC disponibles en los LMS y diferentes servicios en línea de los cursos disponibles de forma gratuita para las carreras en la educación superior (SRIMADHAVEN y col., 2020).

Discusión sobre la muestra: país

La presente revisión mostró que los estudios de AA se han desarrollado principalmente en Europa mientras que en otras regiones del mundo como Oceanía, América y Asia es incipiente y en África no se identificaron estudios, esto coincide con una investigación previa que ha advertido como las estrategias actuales de diseño de aprendizaje se han situado dentro de instituciones específicas en Europa (MITTELMEIER y col., 2018). Aunque pudiera señalarse de forma clara que se recomienda el incremento de las AA en las diferentes regiones del mundo, se debe tener presente que los diferentes recursos tecnológicos adoptados por las universidades permiten a los docentes diseñar experiencias de aprendizaje con propósitos específicos para los estudiantes; estos tienen un sentido según los diversos contextos institucionales y culturales y de qué manera, lo que implica la necesidad de evaluar críticamente la relevancia y adecuación de la implementación de AA considerando las diferentes realidades (ZABOLOTNIAIA y col., 2020). Por otra parte, autores han señalado el problema de las brechas en el desempeño de las instituciones de educación superior de África, identifica inmensas barreras estructurales y tecnológicas para la transformación efectiva y profunda de la enseñanza y el aprendizaje y que a su vez logren satisfacer las necesidades emergentes de equidad y acceso (LUSIGI, 2019). Con estos dos antecedentes discutidos, el desarrollo de estudios basados en AA en las diferentes regiones del mundo tiene un curso natural inherente a sus disposiciones y accesos tecnológicos.

Discusión sobre el tamaño de las muestras

En cuanto a los tamaños muestrales usados en los estudios, se identificó que el rango de mayor frecuencia fue entre 101 a 500 estudiantes. Considerando que las AA son parte de las Ciencias de Datos y de lo conocido como big data, era esperable que las muestras más frecuentes fueran de tamaño grande, sin embargo, fue de un tamaño típico y propio de análisis de datos tradicionales (WIBAWA y col., 2021). Una posible explicación de esto es que los estudios implementaron un tipo de análisis predictivo, que tiene por propósito determinar qué variables explican en este caso el rendimiento académico de los estudiantes, y para ello requieren considerar varios aspectos específicos para que el modelo propuesto logre un nivel de predicción alto, y en este sentido, muestras grandes pueden implicar una variabilidad en las características de estas y por tanto un modelo resultante de baja predicción.

Discusión sobre la carrera y asignatura

Este estudio reveló que más de la mitad de las investigaciones no especifica la carrera de los participantes. Los estudios que sí informaron la carrera señalan a las pertenecientes al área de educación, salud, economía e ingeniería. Por otra parte, se observó respecto de la asignatura, que prácticamente todos los estudios la describen, mostrando que pertenecen principalmente a cursos sobre Ingeniería y Tecnología. Esto es esperable, dado que como se ha advertido por diferentes investigaciones, las tasas de abandono y reprobación en cursos de esta área son alto (LÁZARO ALVAREZ; CALLEJAS y GRIOL, 2020). Dado que la presente revisión seleccionó investigaciones cuyo propósito es predecir el éxito de los estudios, es coherente que la mayor frecuencia de cursos sea parte de esta área disciplinar, puesto que históricamente sus estudiantes muestran dificultades para alcanzar las exigencias académicas.

Discusión sobre el tipo de análisis

Respecto de los tipos de análisis de predicción empleados en los estudios, el usado con más frecuencia fue el análisis de regresión lineal. Otros identificados que se usaron en menor medida fueron la red neuronal, la Distancias Levenshtein entre las cadenas relativas a trayectorias y el Algoritmo KNN. Una revisión sistemática tuvo por objetivo identificar los métodos de predicción del rendimiento académico de los estudiantes en la educación superior (ZULKIFLI; MOHAMED y AZMEE, 2019). Los autores refieren que los modelos predictivos pueden clasificarse en tres (clasificación, clúster y regresión), aunque cada grupo tiene varios métodos que se desarrollaron de acuerdo con sus respectivos objetivos. La clasificación es un método utilizado para identificar las categorías de grupos de nuevas observaciones conocidas en conjuntos de grupos existentes, los métodos de clasificación utilizados más frecuentes son: Clasificación de Bayes (BC), K-Vecino más cercano (KNN), máquina de vectores de soporte (SVM) y árboles de clasificación (CT). Los clústeres o conglomerados (agrupación) son un método que implica el proceso de dividir datos o poblaciones en grupos que tienen caracteres casi idénticos o patrones, pero difieren entre grupos, entre los métodos utilizados para la agrupación se encuentran: Análisis de componentes Principales (PCA), Clustering de K-medias (K-M C), y agrupación jerárquica (HC). El análisis de regresión es un proceso estadístico para estimar la relación entre las variables dependientes y las independientes. A través del modelo de regresión que se ha construido, la variable dependiente se puede predecir fácilmente utilizando la variable independiente deseada. Los resultados encontrados por estos investigadores evidenciaron que el método más frecuente usado es el de clasificación, seguido por el de regresión. Este resultado sería contradictorio al del presente estudio, que mostró a la regresión como el más usado. Esto se podría explicar, dado que la revisión que hicieron los investigadores corresponde a modelos predictivos convencionales. Es decir, no incluyen AA, donde ellos mismos discuten que hay métodos recientes que han incorporado diferentes elementos en sus métodos, como la minería de datos, que combina métodos entre aprendizaje automático, estadísticas y sistemas de bases de datos para identificar patrones en grandes conjuntos de datos. En el método de minería de datos predictivo, los modelos que utilizan el método de clasificación se han categorizado como aprendizaje supervisado definido como aprender a enseñar o entrenar máquinas para reconocer datos a través de buen etiquetado. Mientras tanto, la predicción mediante el uso de clúster es más

conocido como aprendizaje no supervisado, en minería de datos el que se lleva a cabo mediante máquinas para reconocer datos no etiquetados, pero permitiéndoles utilizar algoritmos contra los datos sin orientación. Respecto de los métodos de regresión, resultan desafiante puesto que los datos deben ser cuantitativos, normalizados, sin datos extremos y sin multicolinealidad. Datos con valores y atributos extremos utilizados en relación con los demás conduce a un uso inadecuado de estimaciones. Además, la cantidad de muestreo de datos utilizado en el método de minería de datos debe ser grande para que se pueda generalizar a la población.

Discusión sobre el tipo de software

Respecto del software usado para realizar los análisis de datos, los más frecuentemente usados fueron el software R y el SPSS. Existen muchas herramientas estadísticas que se utilizan en el análisis de datos. Cada herramienta tiene ventajas y desventajas y se ha hecho popular según la característica en la que ha destacado como costo, visualización, paquetes, estadísticas, capacidad de manejo de datos, gráficos y también el big data entre otros factores. Casi todas las áreas, incluyendo educación, han puesto énfasis en torno a la capacidad de hacer predicciones y descubrir patrones en datos. La ciencia de datos es el centro de esta revolución. Eso incluye minería de datos, aprendizaje automático y estadísticas metodologías para extraer conocimiento y apalancamiento de predicciones a partir de datos. El valor de las estadísticas radica en organizar, transformar y simplificación de datos. Para lograr esto, se requieren herramientas de análisis de datos estadísticos (BANSAL y SRIVASTAVA, 2018). En los entornos de LMS, el estudio de los datos a través de la analítica de big data es muy poderoso, especialmente para la toma de decisiones y el uso de datos estadísticos en este entorno rico en datos. Tanto Python como R pueden utilizarse para tomar decisiones que implican grandes datos. Por un lado, Python es perfecto para la enseñanza de la estadística en un entorno rico en datos. Por otro lado, aunque R es un poco más complicado, hay muchos programas personalizables (COLLIAU y col., 2021). En definitiva, los resultados del presente estudio, si bien identifican al Software R en las investigaciones analizadas como uno de los softwares más empleados, en la misma frecuencia de uso se identificó el software SPSS, y de acuerdo a lo señalado en la literatura, es el software como menos propiedad para ser usado en big data (BANSAL y SRIVASTAVA, 2018).

Discusión sobre los modelos de predicción

Respecto de la predicción de los modelos identificados para predecir el rendimiento académico, se observaron diferentes resultados en las investigaciones. Algunos no lograron un nivel mínimo de predicción para fenómenos de las ciencias sociales ($R^2 < .01$), mientras que otros lograron una predicción grande ($R^2 < .30$). Esta variedad en los resultados puede tener diferentes explicaciones. En primer lugar, podría deberse a que la AA para predecir fenómenos en educación, es una línea de investigación reciente y por tanto considera una disciplina emergente (Siemens, 2013). Lo anterior ha implicado la reciente exploración de modelos predictivos de fenómenos como el rendimiento académico o la deserción de los estudios. Los investigadores, por una parte, usan solo variables de AA en los modelos, mientras que otros, han incluido variables sociodemográficas, y otros variables sociocognitivas. Esto implica una integración de diferentes categorías de variables en los modelos que pudieran estar influyendo en el poder predictivo de estos. En segundo lugar, se dispone de diferentes modelos para realizar los análisis estadísticos, los cuales varían en sus procesos y consecuentemente en sus resultados de predicción. En tercer lugar, las muestras de los estudios fueron diversas en cuanto al tamaño y las carreras participantes.

Los métodos tradicionales de predicción consumen mucho tiempo y son incapaces de identificar un número similar de estudiantes en riesgo académico a tiempo. Los investigadores empiezan a explorar el uso de la AA para construir modelos de predicción del abandono que permitan identificar tempranamente aquellos estudiantes de riesgo y, a continuación, realizar intervenciones antes de que el estudiante en riesgo abandone (ZHENG y col., 2020). La AA busca mejorar los procesos de aprendizaje a través de mediciones sistemáticas del aprendizaje datos relacionados y para proporcionar retroalimentación informativa a estudiantes y profesores. Sin embargo, la construcción actual de modelos de predicción de abandono tiene aún importantes limitaciones (ZHENG y col., 2020).

Este estudio también reveló que los modelos predictivos basados en AA que son más efectivos para explicar la variable rendimiento académico en cursos universitarios son aquellos de modalidad online. Esto podría explicarse, dado que el comportamiento de estudio de quienes cursan este tipo de programas está principalmente concentrado en los LMS (MUBARAK; CAO y AHMED, 2021). En cambio, en Modalidad *Blended Learning*, el comportamiento de estudio depende solo en una parte de los LMS, pero otra se da en modalidad presencial, donde no es posible por medio de las AA hacer seguimiento de las diferentes interacciones y dinámicas de estudio.

Por lo anterior, es importante que en cursos *blended learning* o presencial, donde los LMS son solo una parte o un apoyo de la formación del estudiantado, se incluyan otras formas de medición de sus procesos autorregulatorios del estudio (patrones de aprendizaje), que permitan mejorar la predicción del rendimiento. Como, por ejemplo, las clásicas escalas de autorreporte, donde es el mismo estudiante quien autoinforma las características de sus procesos de disposición al estudio, monitoreo y autoevaluación del mismo.

Resulta clave comprender la actividad de estudio y aprendizaje del estudiantado para implementar acciones que permitan oportunamente retroalimentar su desempeño y promover estrategias efectivas para mejorar en una siguiente actuación, tal y como lo ha evidenciado la investigación de la autorregulación del aprendizaje y engagement académico. Para esto, tal como recomienda la investigación, se requiere nuevas propuestas de gestión de datos en las instituciones de educación superior, que permitan superar estos desafíos generando un proceso estandarizado de recolección de datos por medio de LMS y de escalas de medición de otras variables como las sociocognitivas, que no son informada por las AA y su trayectoria en investigación han demostrado su efectividad para predecir el desempeño académico, que sumado a las AA permitirían obtener modelos predictivos consistentes.

La presente revisión sistemática de la literatura tiene algunas limitaciones a considerar. Se revisaron tres bases de datos, pero no se incluyó la base de datos Scielo; y además carece de un análisis estadístico de la explicación de los modelos. Por tanto, un futuro estudio podría por una parte ampliar la revisión de estudios sobre AA incluyendo otras bases de datos y además realizar un meta análisis para evidenciar por medio de un mecanismo de análisis estadístico la predicción efectiva de los modelos propuestos en los estudios.

5 Conclusiones

A partir de los resultados se pueden establecer las siguientes conclusiones: 1) las variables que se incluyen en los modelos predictivos del rendimiento académico en la Educación Superior requieren incluir no solo AA, sino también otras variables psicoeducativas para comprender mejor el fenómeno; 2) los datos de analíticas son extraídos de los LMS institucionales y en todos los estudios analizados se trató de programas *blended learning* y online, por tanto, se requieren investigaciones de modelos predictivos que usen las AA en programas tradicionales de enseñanza universitaria; 3) las investigaciones sobre modelos predictivos en Latinoamérica son incipientes, se requieren investigaciones considerando las particularidades regionales; (4) finalmente, el poder predictivo del rendimiento académico en modelos que incorporan las AA es grande y por tanto prometedores para avanzar en las problemáticas actuales de la Educación Superior como el rendimiento académico y la permanencia.

Agencia de Financiamiento

Beca Doctorado Nacional Folio 21212274 de la Agencia Nacional de Investigación y Desarrollo de Chile (ANID).

Referencias

AKÇAPINAR, Gökhan; ALTUN, Arif y AŞKAR, Petek. Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, v. 16, n. 1, pág. 40, dic. 2019. DOI: 10.1186/s41239-019-0172-z. Disponible en: <<https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-019-0172-z>>. Acceso en: 27 ene. 2022.

- BANIHASHEM, Seyyed Kazem y col. Learning Analytics: A Systematic Literature Review. *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, v. 9, n. 2, jun. 2018. ISSN 2476-7263, 2476-7271. DOI: 10.5812/ijvlms.63024. Disponible en: <<http://ijvlms.com/en/articles/63024.html>>. Acceso en: 27 ene. 2022.
- BANSAL, Anmol y SRIVASTAVA, Satyajee. Tools Used in Data Analysis: A Comparative Study. *International Journal of Recent Research*, v. 5, n. 1, p. 15-18, 2018. Disponible en: <<https://www.ijrra.net/Vol5issue1/IJRR-05-01-04.pdf>>.
- CECHINEL, Cristian y col. Mapping Learning Analytics initiatives in Latin America. *British Journal of Educational Technology*, v. 51, n. 4, p. 892-914, jul. 2020. DOI: 10.1111/bjet.12941. Disponible en: <<https://onlinelibrary.wiley.com/doi/10.1111/bjet.12941>>. Acceso en: 27 ene. 2022.
- COLLIAU, Taylor y col. MatLab vs. Python vs. R. *Journal of Data Science*, v. 15, n. 3, p. 355-372, mar. 2021. DOI: 10.6339/JDS.201707_15(3).0001. Disponible en: <<https://jds-online.org/journal/JDS/article/402>>. Acceso en: 27 ene. 2022.
- CONIJN, Rianne y col. Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. *IEEE Transactions on Learning Technologies*, v. 10, n. 1, p. 17-29, ene. 2017. DOI: 10.1109/TLT.2016.2616312. Disponible en: <<http://ieeexplore.ieee.org/document/7589022/>>. Acceso en: 27 ene. 2022.
- DOBRE, Iuliana. Learning Management Systems for Higher Education - An Overview of Available Options for Higher Education Organizations. *Procedia - Social and Behavioral Sciences*, v. 180, n. 1, p. 313-320, mayo 2015. DOI: 10.1016/j.sbspro.2015.02.122. Disponible en: <<https://linkinghub.elsevier.com/retrieve/pii/S1877042815014536>>. Acceso en: 27 ene. 2022.
- GAŠEVIĆ, Dragan y col. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, v. 28, p. 68-84, ene. 2016. DOI: 10.1016/j.iheduc.2015.10.002. Disponible en: <<https://linkinghub.elsevier.com/retrieve/pii/S1096751615300038>>. Acceso en: 27 ene. 2022.
- GUERRA, Julio y col. Adaptation and evaluation of a learning analytics dashboard to improve academic support at three Latin American universities. *British Journal of Educational Technology*, v. 51, n. 4, p. 973-1001, jul. 2020. DOI: 10.1111/bjet.12950. Disponible en: <<https://onlinelibrary.wiley.com/doi/10.1111/bjet.12950>>. Acceso en: 27 ene. 2022.
- HOODA, Monika. Learning Analytics Lens: Improving Quality of Higher Education. *International Journal of Emerging Trends in Engineering Research*, v. 8, n. 5, p. 1626-1646, mayo 2020. DOI: 10.30534/ijeter/2020/24852020. Disponible en: <<http://www.warse.org/IJETER/static/pdf/file/ijeter24852020.pdf>>. Acceso en: 27 ene. 2022.
- IFENTHALER, Dirk y YAU, Jane Yin-Kim. Reflections on Different Learning Analytics Indicators for Supporting Study Success. *International Journal of Learning Analytics and Artificial Intelligence for Education (IJAI)*, v. 2, n. 2, pág. 4, jul. 2020. DOI: 10.3991/ijai.v2i2.15639. Disponible en: <<https://online-journals.org/index.php/i-jai/article/view/15639>>. Acceso en: 27 ene. 2022.
- IFENTHALER, Dirk y YAU, Jane Yin-Kim. Utilising learning analytics to support study success in higher education: a systematic review. *Educational Technology Research and Development*, v. 68, n. 4, p. 1961-1990, ago. 2020. DOI: 10.1007/s11423-020-09788-z. Disponible en: <<https://link.springer.com/10.1007/s11423-020-09788-z>>. Acceso en: 27 ene. 2022.
- KADIR, Aini Zuriyati Abdul y AZIZ, Nur Sukinah. Learning Management System of Higher Education Institution. *Indian Journal of Science and Technology*, v. 9, n. 9, mar. 2016. ISSN 0974-5645, 0974-6846. DOI: 10.17485/ijst/2016/v9i9/88717. Disponible en: <<https://indjst.org/articles/learning-management-system-of-higher-education-institution>>. Acceso en: 27 ene. 2022.
- KASIM, Nurul Nadirah y KHALID, Fariza. Choosing the Right Learning Management System (LMS) for the Higher Education Institution Context: A Systematic Review. *International Journal of Emerging Technologies in Learning (IJET)*, v. 11, n. 06, pág. 55, jun. 2016. DOI: 10.3991/ijet.v11i06.5644. Disponible en: <<http://online-journals.org/index.php/i-jet/article/view/5644>>. Acceso en: 27 ene. 2022.
- KERIMBAYEV, Nurassyl y col. Virtual educational environment: interactive communication using LMS Moodle. *Education and Information Technologies*, v. 25, n. 3, p. 1965-1982, mayo 2020. DOI: 10.1007/s10639-019-10067-5. Disponible en: <<http://link.springer.com/10.1007/s10639-019-10067-5>>. Acceso en: 27 ene. 2022.

- KLÁŠNJA-MILIĆEVIĆ, Aleksandra; IVANOVIĆ, Mirjana y BUDIMAC, Zoran. Data science in education: Big data and learning analytics. *Computer Applications in Engineering Education*, v. 25, n. 6, p. 1066-1078, nov. 2017. DOI: 10.1002/cae.21844. Disponible en: <<https://onlinelibrary.wiley.com/doi/10.1002/cae.21844>>. Acceso en: 27 ene. 2022.
- LARRABEE, Anders; HUGHES, Emily y SMITH, Joanne. The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, v. 50, n. 5, p. 2594-2618, sep. 2019. DOI: 10.1111/bjet.12720. Disponible en: <<https://onlinelibrary.wiley.com/doi/10.1111/bjet.12720>>. Acceso en: 27 ene. 2022.
- LÁZARO ALVAREZ, Niurys; CALLEJAS, Zoraida y GRIOL, David. Predicting Computer Engineering students' dropout in Cuban Higher Education with pre-enrollment and early performance data. *Journal of Technology and Science Education*, v. 10, n. 2, pág. 241, sep. 2020. DOI: 10.3926/jotse.922. Disponible en: <<https://www.jotse.org/index.php/jotse/article/view/922>>. Acceso en: 27 ene. 2022.
- LU, Owen H. T. y col. Applying Learning Analytics for the Early Prediction of Students' Academic Performance in Blended Learning. *Educational Technology & Society*, v. 21, n. 2, p. 220-232, 2018. Disponible en: <<https://eric.ed.gov/?id=EJ1175301>>. Acceso en: 27 ene. 2022.
- LUSIGI, Angela. Higher Education, Technology, and Equity in Africa. *New Review of Information Networking*, v. 24, n. 1, p. 1-16, ene. 2019. DOI: 10.1080/13614576.2019.1608576. Disponible en: <<https://www.tandfonline.com/doi/full/10.1080/13614576.2019.1608576>>. Acceso en: 27 ene. 2022.
- MIRANDA, Sergio y VEGLIANTE, Rosa. Learning Analytics to support learners and teachers: the navigation among contents as a model to adopt. *Journal of e-Learning and Knowledge Society*, p. 101-116, oct. 2019. DOI: 10.20368/1971-8829/1135065. Disponible en: <https://www.je-lks.org/ojs/index.php/Je-LKS_EN/article/view/1135065>. Acceso en: 27 ene. 2022.
- MITTELMEIER, Jenna y col. Learning design in diverse institutional and cultural contexts: suggestions from a participatory workshop with higher education professionals in Africa. *Open Learning: The Journal of Open, Distance and e-Learning*, v. 33, n. 3, p. 250-266, sep. 2018. DOI: 10.1080/02680513.2018.1486185. Disponible en: <<https://www.tandfonline.com/doi/full/10.1080/02680513.2018.1486185>>. Acceso en: 27 ene. 2022.
- MOHER, David y col. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, v. 4, n. 1, pág. 1, dic. 2015. DOI: 10.1186/2046-4053-4-1. Disponible en: <<https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/2046-4053-4-1>>. Acceso en: 27 ene. 2022.
- MONLLAÓ OLIVÉ, David y col. A supervised learning framework: using assessment to identify students at risk of dropping out of a MOOC. *Journal of Computing in Higher Education*, v. 32, n. 1, p. 9-26, abr. 2020. DOI: 10.1007/s12528-019-09230-1. Disponible en: <<http://link.springer.com/10.1007/s12528-019-09230-1>>. Acceso en: 27 ene. 2022.
- MUBARAK, Ahmed Ali; CAO, Han y AHMED, Salah A.M. Predictive learning analytics using deep learning model in MOOCs' courses videos. *Education and Information Technologies*, v. 26, n. 1, p. 371-392, ene. 2021. DOI: 10.1007/s10639-020-10273-6. Disponible en: <<https://link.springer.com/10.1007/s10639-020-10273-6>>. Acceso en: 27 ene. 2022.
- NGUYEN, Andy y col. Design principles for learning analytics information systems in higher education. *European Journal of Information Systems*, v. 30, n. 5, p. 541-568, sep. 2021. DOI: 10.1080/0960085X.2020.1816144. Disponible en: <<https://www.tandfonline.com/doi/full/10.1080/0960085X.2020.1816144>>. Acceso en: 27 ene. 2022.
- PICCIANO, Anthony G. The Evolution of Big Data and Learning Analytics in American Higher Education. *Online Learning*, v. 16, n. 3, jun. 2012. DOI: 10.24059/olj.v16i3.267. Disponible en: <<https://olj.onlinelearningconsortium.org/index.php/olj/article/view/267>>. Acceso en: 27 ene. 2022.
- PRIYAADHARSHINI, M y col. Learning Analytics: Game-based Learning for Programming Course in Higher Education. *Procedia Computer Science*, v. 172, p. 468-472, 2020. DOI: 10.1016/j.procs.2020.05.143. Disponible en: <<https://linkinghub.elsevier.com/retrieve/pii/S1877050920314733>>. Acceso en: 27 ene. 2022.

- ROJAS CASTRO, Pablo. Learning Analytics. Una Revisión de la Literatura. *Educación y Educadores*, v. 20, n. 1, p. 106-128, feb. 2017. DOI: 10.5294/edu.2017.20.1.6. Disponible en: <<http://educacionyeducadores.unisabana.edu.co/index.php/eye/article/view/6412/4454>>. Acceso en: 27 ene. 2022.
- ŞAHİN, Muhittin y YURDUGÜL, Halil. Educational Data Mining and Learning Analytics: Past, Present and Future. *Bartın University Journal of Faculty of Education*, v. 9, n. 1, p. 121-131, 2019.
- SAIZ MANZANARES, María Consuelo y col. Detección del alumno en riesgo en titulaciones de Ciencias de la Salud: aplicación de técnicas de Learning Analytics. *European Journal of Investigation in Health, Psychology and Education*, v. 8, n. 3, pág. 129, dic. 2018. DOI: 10.30552/ejihpe.v8i3.273. Disponible en: <<https://formacionasunivep.com/ejihpe/index.php/journal/article/view/273>>. Acceso en: 27 ene. 2022.
- SALGADO, Nelson y col. Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. *Sistemas y Tecnologías de Información*, v. 1, n. 17, p. 258-266, 2019.
- SAQR, Mohammed; FORS, Uno y NOURI, Jalal. Time to focus on the temporal dimension of learning: a learning analytics study of the temporal patterns of students' interactions and self-regulation. *International Journal of Technology Enhanced Learning*, v. 11, n. 4, pág. 398, 2019. DOI: 10.1504/IJTEL.2019.10020597. Disponible en: <<http://www.inderscience.com/link.php?id=10020597>>. Acceso en: 27 ene. 2022.
- SOFFER, Tal y COHEN, Anat. Students' engagement characteristics predict success and completion of online courses. *Journal of Computer Assisted Learning*, v. 35, n. 3, p. 378-389, jun. 2019. DOI: 10.1111/jcal.12340. Disponible en: <<https://onlinelibrary.wiley.com/doi/10.1111/jcal.12340>>. Acceso en: 27 ene. 2022.
- SRIMADHAVEN, T. y col. Learning Analytics: Virtual Reality for Programming Course in Higher Education. *Procedia Computer Science*, v. 172, p. 433-437, 2020. DOI: 10.1016/j.procs.2020.05.095. Disponible en: <<https://linkinghub.elsevier.com/retrieve/pii/S1877050920314241>>. Acceso en: 27 ene. 2022.
- STRANG, Kenneth David. Can online student performance be forecasted by learning analytics? *International Journal of Technology Enhanced Learning*, v. 8, n. 1, pág. 26, 2016. DOI: 10.1504/IJTEL.2016.075950. Disponible en: <<http://www.inderscience.com/link.php?id=75950>>. Acceso en: 27 ene. 2022.
- TEMPELAAR, Dirk; RIENTIES, Bart y NGUYEN, Quan. Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application. Edición: Vitomir Kovanovic. *PLOS ONE*, v. 15, n. 6, e0233977, jun. 2020. DOI: 10.1371/journal.pone.0233977. Disponible en: <<https://dx.plos.org/10.1371/journal.pone.0233977>>. Acceso en: 27 ene. 2022.
- TEMPELAAR, Dirk T.; RIENTIES, Bart y NGUYEN, Quan. Towards Actionable Learning Analytics Using Dispositions. *IEEE Transactions on Learning Technologies*, v. 10, n. 1, p. 6-16, ene. 2017. DOI: 10.1109/TLT.2017.2662679. Disponible en: <<http://ieeexplore.ieee.org/document/7839177/>>. Acceso en: 27 ene. 2022.
- TULASI, Ben. Significance of Big Data and Analytics in Higher Education. In: BIG Data and Learning Analytics in Higher Education: Current Theory and Practice. Cham: Springer International Publishing, 2013. Disponible en: <https://doi.org/10.1007/978-3-319-06520-5_1>. Acceso en: 27 ene. 2022.
- WIBAWA, Basuki y col. Learning analytic and educational data mining for learning science and technology. In: TRANSFORMING Research and Education of Science and Mathematics in the Digital Age. Jakarta, Indonesia: [s.n.], 2021. v. 20, pág. 060001. DOI: 10.1063/5.0041844. Disponible en: <<http://aip.scitation.org/doi/abs/10.1063/5.0041844>>. Acceso en: 27 ene. 2022.
- ZABOLOTNIAIA, Mariia y col. Use of the LMS Moodle for an Effective Implementation of an Innovative Policy in Higher Educational Institutions. *International Journal of Emerging Technologies in Learning (iJET)*, v. 15, n. 13, pág. 172, jul. 2020. DOI: 10.3991/ijet.v15i13.14945. Disponible en: <<https://online-journals.org/index.php/i-jet/article/view/14945>>. Acceso en: 27 ene. 2022.
- ZHANG, Yaqun; GHANDOUR, Ahmad y SHESTAK, Viktor. Using Learning Analytics to Predict Students Performance in Moodle LMS. *International Journal of Emerging Technologies in Learning (iJET)*, v. 15, n. 20, pág. 102, oct. 2020. DOI: 10.3991/ijet.v15i20.15915. Disponible en: <<https://online-journals.org/index.php/i-jet/article/view/15915>>. Acceso en: 27 ene. 2022.
- ZHENG, Juan y col. Profiling self-regulation behaviors in STEM learning of engineering design. *Computers & Education*, v. 143, pág. 103669, ene. 2020. DOI: 10.1016/j.compedu.2019.103669. Disponible en: <<https://linkinghub.elsevier.com/retrieve/pii/S0360131519302222>>. Acceso en: 27 ene. 2022.

ZULKIFLI, Faiz; MOHAMED, Zulkifley y AZMEE, Nor Afzalina. Systematic Research on Predictive Models on Students' Academic Performance in Higher Education. *International Journal of Recent Technology and Engineering*, v. 8, 2S3, p. 357-363, ago. 2019. DOI: 10.35940/ijrte.B1061.0782S319. Disponible en: <<https://www.ijrte.org/wp-content/uploads/papers/v8i2S3/B10610782S319.pdf>>. Acceso en: 27 ene. 2022.

Contribuciones de los autores

Javier Mella Norambuena: Conceptualización, Curación de datos, Metodología, Redacción – borrador original; **María Graciela Badilla:** Conceptualización, Investigación, Supervisión; **Yaranay López Angulo:** Conceptualización, Visualización, Redacción – Revisión y edición .