

Análisis del aprendizaje: una revisión sistemática de literatura

Learning analytics: a systematic literature review

Victor Daniel Gil Vera*

Recibido: 22/03/2017

Aprobado: 12/07/2017

Resumen: la mayoría de algoritmos utilizados en el análisis de datos están diseñados de acuerdo con las capacidades de potencia y flexibilidad más que por su sencillez, y son demasiado complejos de utilizar en el contexto educativo. El objetivo de este trabajo es presentar una revisión de literatura sobre el análisis del aprendizaje en la educación superior: problemas, limitaciones, técnicas y herramientas empleadas. Se utilizó la metodología de la revisión sistemática de literatura para responder a tres preguntas de investigación tomando como base publicaciones científicas. Se concluye que se deben implementar, adaptar o desarrollar algoritmos predeterminados para el contexto educativo y, también, construir herramientas para el análisis de datos educacionales que cuenten con interfaces intuitivas y fáciles de utilizar.

Palabras clave: análisis del aprendizaje, análisis del aprendizaje multimodal, intercambio de datos, aprender a aprender, aprendizaje permanente.

Abstract: the majority of algorithms used in the data analysis are designed according to the capacities of power and flexibility rather than for its simplicity and are too complex to use in the educational context. The objective of this paper is to present a literature review on the learning analytics in higher education: problems, limitations, techniques and tools used. The systematic literature review methodology was used to answer three research questions on the basis of scientific publications. This paper concludes that it must implement, adapt or develop algorithms for the educational context and must be build tools for the analysis of educational data with intuitive interfaces and easy to use.

Keywords: learning analytics, learning to learn, lifelong learning, multimodal learning analytics, data exchange.

□Artículo de revisión. Este artículo es resultado del proyecto: construcción de un algoritmo predictivo para el análisis del aprendizaje.

*MSc, PhD(c). Docente Investigador Universidad Católica Luis Amigó. Estudiante de Doctorado Universidad Nacional de Colombia – Sede Medellín. Medellín, Colombia. victor.gilve@amigo.edu.co

INTRODUCCIÓN

Debido a la constante generación de información en escuelas, colegios y universidades relacionada con los estudiantes y sus contextos, es necesario adaptar algoritmos utilizados en el análisis de datos a las necesidades y realidades que se presentan en la educación superior, que permitan hacer análisis y pronósticos relacionados con el rendimiento académico, deserción, etc.

El análisis del aprendizaje se refiere a la recolección, análisis y presentación de datos sobre los alumnos y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que se producen (Peña, 2017) Surge como una disciplina emergente que busca mejorar la enseñanza y el aprendizaje por medio de una evaluación crítica de datos y la generación de patrones relacionados con los hábitos y respuestas de los estudiantes, que permitan proporcionar retroalimentación oportuna (Peña, 2017). Entre las principales problemáticas de la implementación de algoritmos encontradas en la revisión del estado del arte se destacan las siguientes: los usuarios tienen que proporcionar los valores apropiados a los parámetros para obtener buenos resultados y modelos, por lo tanto, deben poseer una cierta cantidad de conocimientos especializados para poder encontrar la configuración correcta (Drachsler y Greller, 2012). También, tienen que seleccionar el método o algoritmo específico que quieren utilizar de una amplia gama de métodos y algoritmos disponibles de análisis de datos. La mayoría de los algoritmos para el análisis de datos necesitan ser configurados antes de ser ejecutados, y no están diseñados para usuarios con pocos conocimientos en programación (Romero y Ventura, 2010).

Este trabajo busca responder a las siguientes preguntas de investigación: *P1. ¿Cuáles son los principales problemas de la implementación de algoritmos para el análisis de datos en la educación? P2. ¿Cuáles son las posibles soluciones? P3. ¿Qué casos reales de aplicación del análisis del aprendizaje se han reportado en la educación superior?* Se empleó la metodología de la revisión sistemática de literatura (RSL), la cual consiste

en una revisión sistemática, rigurosa, ordenada y auditable cuyo objetivo es responder preguntas de investigación, y toma como base investigaciones y publicaciones científicas realizadas con anterioridad (B. Kitchenham, 2004).

El artículo se compone de las siguientes secciones: definiciones y conceptos básicos, protocolo de la revisión, proceso de búsqueda de documentos (selección de documentos, evaluación de la calidad y extracción de datos), resultados del proceso de búsqueda, selección de documentos y evaluación de la calidad, análisis de los estudios seleccionados como un todo y discusión de las respuestas a las preguntas de investigación, conclusión.

PLANEAMIENTO DE LA REVISIÓN

Esta sección presenta los parámetros utilizados en el diseño de la investigación, los cuales fueron establecidos tomando como referencia las publicaciones de (B. . Kitchenham, 2004) y (B. A. Kitchenham y Charters, 2007). Se realizó una búsqueda en las bases de datos científicas: Scopus, IEEE Xplore, ISI Web of knowledge. Las palabras clave de búsqueda fueron obtenidas a partir de los títulos de artículos y libros especializados sobre análisis del aprendizaje; posteriormente se obtuvieron sinónimos y palabras similares. Al final, se conformaron las cadenas de búsqueda con las combinaciones de las palabras: "Learning analytics", "21st century skills", "multimodal learning analytics", "data sharing", "learning to learn", "lifelong learning", "transferable skills", y "complex systems" en el título, resumen y palabras clave. Las cadenas de búsqueda empleadas fueron:

- (1) (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (21st century skills))
- (2) (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (multimodal learning analytics))
- (3) (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (data sharing))
- (4) (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (learning to learn))
- (5) (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (lifelong learning))
- (6) (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (transferable skills))

(7) (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (complex systems))

Se incluyeron artículos, artículos de conferencias, libros y capítulos de libros publicados en Scopus, IEEE Xplore, e ISI Web of knowledge. Se descartaron publicaciones de internet, resúmenes de conferencias y editoriales. Se recopilieron las publicaciones encontradas en la revisión y se jerarquizaron las temáticas de cada una con el fin de responder las preguntas de investigación.

DISCUSIÓN

Esta sección presenta el análisis de los resultados encontrados en la revisión, los cuales están enfocados en responder las tres preguntas de investigación planteadas:

P1. ¿Cuál es el principal problema de la implementación de algoritmos para el análisis de datos en la educación?

Según Drachsler y Greller (2012), algoritmos para el análisis del aprendizaje aún no han sido desarrollados y los existentes (árboles/reglas de decisión, regresión, agrupación, estadísticos descriptivos, visualización, series de tiempo/ análisis de secuencias, reglas de asociación, métodos de ensamble, minería de texto, vectores de soporte, etc.) pueden resultar útiles únicamente a desarrolladores (Romero y Ventura, 2010). Sumado a esto, no hay herramientas generales que puedan ser aplicadas a cualquier sistema educativo. Debido a lo anterior, se necesitan algoritmos/modelos/ herramientas prediseñadas para el contexto educativo, que permitan normalizar los datos de entrada y salida (preproceso - postproceso), emplear información semántica e integrar el dominio del conocimiento educativo. En la revisión del estado del arte se identificaron las diferentes técnicas utilizadas en el análisis del aprendizaje, sus principales aplicaciones y algunos ejemplos de investigaciones realizadas. Ver Tabla 1.

Tabla 1. Técnicas empleadas en el análisis del aprendizaje

Técnica	Uso / Aplicación	Ejemplos
Predicción	Predecir el rendimiento y detectar comportamientos de los estudiantes.	(Khousa, Atif, y Masud, 2015), (Cambuzzi, Rigo, y Barbosa, 2015) y (Harrison, Villano, Lynch, y Chen, 2015)
Agrupamiento	Agrupar materiales similares o estudiantes basados en sus patrones de aprendizaje e interacción.	(Aguar, Chawla, Brockman, Ambrose, y Goodrich, 2014), (Asif, Merceron, y Pathan, 2015) y (Scheffel et al., 2012)
Detección de valores atípicos	Detectar estudiantes con dificultades o procesos de aprendizaje irregulares.	(Grau-Valldosera y Minguillón, 2011), (Manso-Vázquez y Llamas-Nistal, 2015) y (Sinclair y Kalvala, 2015)
Minería de relación	Identificar relaciones en patrones de comportamiento de los estudiantes y diagnosticar las dificultades de estos.	(Kim, Jo, y Park, 2016), (Pardo et al., 2015) y (Piety, Hickey, y Bishop, 2014)
Análisis de redes sociales	Interpretar la estructura de las relaciones en las actividades de colaboración e interacción con herramientas de comunicación.	(Hecking, Ziebarth, y Hoppe, 2014), (Tervakari et al., 2013) y (Vozniuk, Holzer, y Gillet, 2014)

Técnica	Uso / Aplicación	Ejemplos
Proceso de minería	Analizar el comportamiento de los estudiantes.	(Menchaca, Guenaga, y Solabarrieta, 2015), (Vahdat, Oneto, Anguita, Funk, y Rauterberg, 2015) y (Wise, 2014)
Minería de texto	Analizar el contenido de los foros, chats, páginas web y documentos.	(Gasevic, Kovanovic, Joksimovic, y Siemens, 2014), (Lotsari, Verykios, Panagiotakopoulos, y Kalles, 2014) y (Prinsloo, Slade, y Galpin, 2012)
Destilación de datos para el juicio humano	Ayudar a los docentes a visualizar y analizar las actividades en curso de los estudiantes y el uso de la información.	(Aguiar et al., 2014), (Grann y Bushway, 2014) y (Swenson, 2014)
Descubrimiento con modelos	Identificar las relaciones entre los comportamientos del estudiante y las características o variables contextuales. Integración de modelos psicométricos en modelos de aprendizaje automático.	(Gibson, Kitto, y Willis, 2014), (Kovanović et al., 2015) y (Lockyer y Dawson, 2011)
Gamificación	Incluir posibilidades de aprendizaje lúdico para mantener la motivación; por ejemplo, la integración de los logros, puntos de experiencia o insignias como indicadores de éxito.	(Holman, Aguilar, y Fishman, 2013), (Øhrstrøm, Sandborg-Petersen, Thorvaldsen, y Ploug, 2013) y (Westera, Nadolski, y Hummel, 2013)
Aprendizaje de máquinas	Encontrar ideas ocultas en datos automáticamente (basado en modelos que se exponen a nuevos datos y se adaptan independientemente).	(Corrigan, Smeaton, Glynn, y Smyth, 2015), (McKay, Miller, y Tritz, 2012) y (Nespereira, Elhariri, El-Bendary, Vilas, y Redondo, 2016)
Estadística	Análisis e interpretación de datos cuantitativos para la toma de decisiones.	(Clow, 2014), (Khousa y Atif, 2014) y (Simsek et al., 2015)

Fuente: elaboración propia

Según KD Nuggets, (2017f), el top diez de los algoritmos más utilizados por los científicos de datos son:

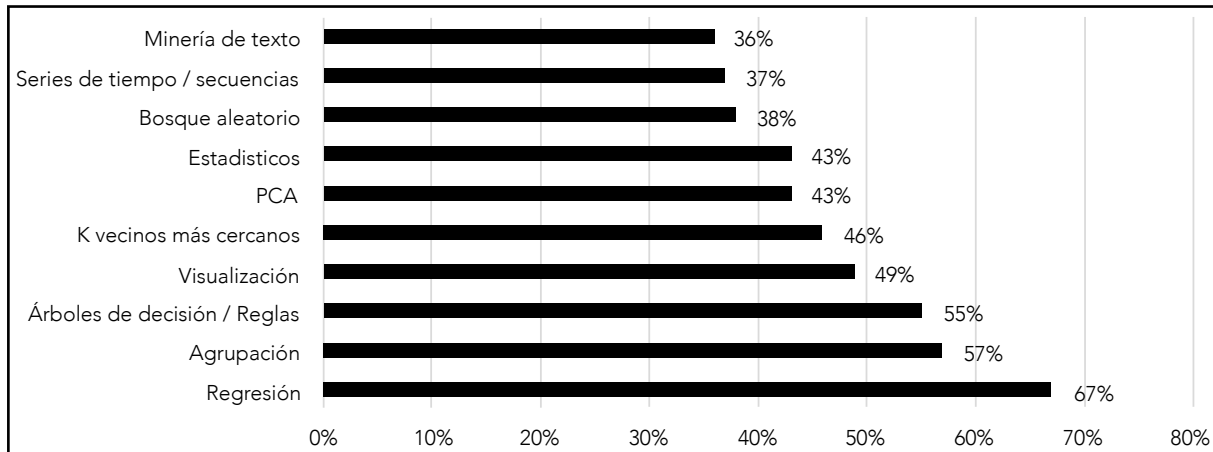


Figura 1. Algoritmos Análisis de Datos. Fuente: KD Nuggets Fuente: elaboración propia

A continuación, se presenta una breve descripción de cada uno de ellos:

Árboles/reglas de decisión: método supervisado (hay un conocimiento a priori) y no paramétrico, utilizado en el aprendizaje de clasificación y regresión. Permite crear modelos para predecir el valor de una variable de destino mediante el aprendizaje de reglas de decisión simples, inferidas a partir de las características de los datos (Scikit-Learn, 2017). Estos permiten hacer predicciones de pertenencia a un grupo utilizando criterios estadísticos (Mayor, 2015). Algunos algoritmos utilizados son: ID3, C4.5, C5.0 CART, Random forest y árboles de inferencia condicional.

- **ID3:** es uno de los algoritmos más simples para producir árboles de decisión con atributos y clases categóricas. Se basa en una medida denominada relación de ganancia (diferencia en la incertidumbre que se obtiene a través de una partición de los datos) para construir los árboles (Mayor, 2015). El objetivo es maximizar el poder predictivo del árbol mediante la reducción de la incertidumbre en los datos (Mayor, 2015). Está disponible para los usuarios de R a través del paquete `data.tree` (`install.packages("data.tree"); library(data.tree)`).

- **C4.5:** funciona de manera similar al ID3, pero utiliza la relación de ganancia como un criterio de partición (Mayor, 2015). Acepta la partición en atributos numéricos que dividen en

categorías. El valor de la división se selecciona con el fin de disminuir la entropía para el atributo. A diferencia del ID3, el C4.5 permite la poda, que es básicamente la parte inferior hasta la simplificación del árbol para evitar el sobreajuste de los datos del entrenamiento (Mayor, 2015). Está disponible para los usuarios de R a través del paquete `RWeka` (`install.packages("RWeka"); library(RWeka)`).

- **C5.0:** es una versión mejorada del C4.5, incluye empuje y aventamiento. El objetivo del empuje es aumentar la fiabilidad de las predicciones realizando el análisis iterativamente y ajustando pesos de las observaciones después de cada iteración (Mayor, 2015). El aventamiento se refiere a la supresión de atributos inútiles para el análisis principal (Mayor, 2015). Está disponible para los usuarios de R a través del paquete `C50` (`install.packages("C50"); library(C50)`).

- **CART:** utiliza diferentes criterios estadísticos para decidir sobre las divisiones del árbol (Mayor, 2015). El atributo para hacer la partición es seleccionado con el índice de Gini como criterio de decisión. En árboles de clasificación, el índice de Gini se calcula como:

$$1 - \sum_{j=1}^c p_j^2 \quad (2)$$

Donde p es la probabilidad de cada posible partición sobre el atributo. Está disponible para los usuarios de R a través del paquete `rpart` (`install.packages("rpart"); library(rpart)`). CART permite predecir un resultado numérico. En los árboles de regresión, CART realiza una regresión y construye el árbol de una manera que minimiza el cuadrado de los residuos (Mayor, 2015).

- Random forest: utiliza el conjunto de aprendizaje (combinación de árboles CART) para mejorar la clasificación mediante un principio de voto (Mayor, 2015). En este se utiliza un algoritmo similar a CART para generar los árboles. Las diferencias incluyen el uso de empaquetamiento y selección de predictores al azar en cada partición del árbol (Mayor, 2015). El empaquetamiento busca reducir el impacto del error de medición (ruido) en los datos y por lo tanto evitar el desbordamiento (Mayor, 2015). Está disponible para los usuarios de R a través del paquete `randomForest` (`install.packages("randomForest"); library(randomForest)`).

- Árboles de inferencia condicional: a diferencia de los algoritmos anteriores, estos dependen de la significación estadística en la selección de atributos sobre los cuales se realiza la partición (Mayor, 2015). El algoritmo primero busca los atributos que predicen significativamente la clase, en una prueba de hipótesis que puede ser seleccionada en la llamada de la función (Mayor, 2015). Está disponible para los usuarios de R a través del paquete `partykit` (`install.packages(c("Formula", "partykit")); library(Formula); library(partykit)`).

Regresión: método utilizado en predicción, cuyo objetivo es intentar predecir una salida de valor real (valor numérico) de alguna variable (KDNuggets, 2017d). Los tipos de modelos de regresión pueden ser: regresión lineal simple, múltiple, o rectas de regresión. El primero es utilizado para examinar como un atributo afecta a otro, el segundo para evaluar como varios atributos afectan a otro (Mayor, 2015). Matemáticamente los modelos de regresión lineal se pueden expresar como:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_{p2} + \dot{E} \quad (1)$$

Dónde Y_t representa la variable dependiente, X_1, X_2, \dots, X_p , las variables explicativas, $\beta_1, \beta_2, \dots, \beta_p$, parámetros y \dot{E} error. Los modelos de regresión lineal simple analizan la relación entre un predictor (causa) y un criterio (consecuencia). Existen dos parámetros muy importantes que son el resultado de un análisis de regresión; el intercepto (valor promedio del criterio cuando el predictor es 0) y el valor de la pendiente (indica cuántas unidades, en promedio, el criterio cambia cuando el predictor se incrementa en una unidad) (KDNuggets, 2017d). Por su parte, los modelos de regresión lineal múltiple, miden el impacto de varios predictores sobre un único criterio, en lugar de solo uno en regresión simple (KDNuggets, 2017d). El valor de las observaciones puede ser calculado como la intersección más el coeficiente de la pendiente multiplicado por el factor predictivo más el residuo (KDNuggets, 2017d).

Agrupación: método de aprendizaje no supervisado utilizado para analizar datos que no incluyen clases pre-definidas. Las instancias de datos se agrupan maximizando la similitud entre clases iguales y minimizando la similitud entre clases diferentes (KDNuggets, 2017b). El algoritmo de agrupación permite identificar y agrupar las instancias que son muy similares, en contraposición a desagrupar las instancias que son menos similares unas de las otras (KDNuggets, 2017b). Un clúster es un grupo de casos, observaciones, individuos u otras unidades, que son similares entre sí en las características consideradas (Mayor, 2015). Los algoritmos de agrupamiento utilizan medidas de distancias entre los casos con el fin de crear grupos homogéneos de casos (Mayor, 2015). Algunos de los métodos utilizados son: jerárquico (aglomerativo), asignación de puntos, cohesión, K means, agrupamiento poblacional, algoritmo BFR (Bradley Fayyad Reina). Está disponible para los usuarios de R a través del paquete `clúster`.

Estadísticos descriptivos: permiten describir cuantitativamente una recopilación de información, y resumir a través de un solo número (media, mediana, moda, distribuciones de probabilidad, covarianza, correlación, etc.) a un conjunto de datos (KDNuggets, 2017).

Visualización: técnica utilizada para comunicar datos o información mediante la codificación de objetos visuales (puntos, líneas o barras) que figuran en los gráficos. El objetivo es comunicar información de forma clara y eficiente para los

usuários a través de medios gráficos (Friedman, 2008). Algunos de los métodos utilizados son Sampling, Shuffling, Sorting, Maze Generation, Using Vision to Think.

Series de tiempo/análisis de secuencias: permiten analizar el comportamiento de una variable en el tiempo, facilitando la toma de decisiones. Toman como base datos históricos para obtener una caracterización estadística de los enlaces existentes en el pasado y en el presente (Diebold, 2001), permiten considerar información temporal, estimar parámetros y proporcionar datos puntuales precisos sobre valores futuros (Pravilovic, Bilancia, Appice, y Malerba, 2017). Algunas de las técnicas utilizadas son; ARIMA, RNA, SVM, algoritmos genéticos para la predicción con series de tiempo.

Reglas de asociación: técnica para descubrir cómo los elementos se asocian entre sí (KDNuggets, 2017a). Hay tres formas comunes para medir la asociación; apoyo, confianza, elevación. Existen tres algoritmos para realizar búsquedas de reglas de asociación en bases de datos: Apriori, Partition, y Eclat.

Métodos de ensamble: técnica para combinar varios algoritmos de aprendizaje débiles en un

intento por construir un algoritmo de aprendizaje más poderoso, combinan múltiples hipótesis para formar una hipótesis mejor (KDNuggets, 2017c). Algunas de las técnicas utilizadas son; Boosting, Bagging, Combination Methods, Diversity, Ensemble Pruning.

Minería de texto: el análisis de texto o minería de texto es la extensión natural y esencial del análisis predictivo y ciencia de los datos (KDNuggets, 2017e). Algunos de los métodos utilizados son: clasificación, descubrimiento de asociaciones, agrupación.

Vectores de soporte: algoritmo de aprendizaje automático supervisado que puede ser utilizado tanto para los retos de clasificación o regresión (Ray, 2015). Sin embargo, se utiliza sobre todo en problemas de clasificación. En este algoritmo, representamos gráficamente cada elemento de datos como un punto en el espacio n-dimensional (donde n es el número de características que tiene). Los vectores de soporte son simplemente las coordenadas de una observación individual.

La Tabla 2, resume las principales ventajas y desventajas de las técnicas mencionadas anteriormente:

Tabla 2. Ventajas y desventajas

Técnica	Ventajas	Desventajas
Árboles/ reglas de decisión	Permiten analizar todas las opciones posibles de un conjunto de datos. Ayudan a identificar la probabilidad de que un evento determinado suceda.	Requieren un conocimiento previo. Presenta problemas cuando la cantidad de alternativas es grande.
Regresión	Permite establecer una relación lineal entre las variables dependientes e independientes.	Altamente susceptible a valores atípicos.
Agrupación	Permite resumir y entender los datos, solo requiere instancias pero no etiquetas.	Altamente sensible al ruido y a valores atípicos.

Técnica	Ventajas	Desventajas
Estadísticos descriptivos	Fáciles de calcular y de interpretar. Presentan rigor matemático.	Altamente sensibles a valores extremos, no son confiables en distribuciones asimétricas
Visualización	Permite la codificación de objetos visuales.	Los datos deben tener una distribución uniforme para que no haya vacíos.
Series de tiempo/ Análisis de secuencias	Permiten hacer inferencias sobre el futuro (predicciones).	Requieren información histórica. Altamente sensibles al ruido y a valores atípicos.
Reglas de asociación	Permiten establecer diversidad de reglas, cada una de las cuales puede tener una conclusión diferente.	Al tratar de encontrar patrones en un espacio de búsqueda amplio necesitan más tiempo de ejecución que un algoritmo de árbol de decisión.
Métodos de ensamble	Permiten reducir el sesgo y la varianza (aprendizaje supervisado).	La combinación de varios algoritmos a veces puede resultar poco eficiente e inadecuada para el tratamiento de los datos.
Minería de texto	Permite explotar fuentes de información, clasificar, categorizar, resumir textos, editar, mapear y visualizar.	No considera la perspectiva y el sentimiento al momento de realizar análisis cuantitativos, lo que puede conducir a interpretaciones equivocadas o al mal uso de la información.
Vectores de soporte	Pueden ser empleados tanto para hacer clasificaciones o regresiones. También en la solución de problemas lineales y no lineales	Malas clasificaciones debido al sobreentrenamiento ("Overtraining") o al aprendizaje inadecuado de las características de los datos.

Fuente: elaboración propia

P2. ¿Cuáles son las posibles soluciones al problema ?

Implementar / adaptar/ desarrollar algoritmos predeterminados para el contexto educativo con el fin de simplificar la configuración y ejecución a los usuarios. Además, construir herramientas para el análisis de datos educacionales que cuenten con interfaces intuitivas y fáciles de utilizar, y con facilidades de visualización para educadores y diseñadores de e-learning (Romero y Ventura, 2010), también es importante desarrollar herramientas de pre-procesamiento específicas con el fin de automatizar y facilitar las tareas o funciones de pre-procesamiento que actualmente los usuarios de análisis de datos educacionales deben hacer manualmente.

P3. ¿Qué casos reales de aplicación del análisis del aprendizaje se han reportado en instituciones de educación superior?

- Universidad de Purdue: implementó un sistema de señalización e intervenciones para mejorar el éxito académico de los estudiantes, lo que aumentó las tasas de retención y graduación. El sistema de señales tenía como objetivo ayudar a los estudiantes a comprender su progreso lo suficientemente temprano para que fueran capaces de buscar ayuda.
- Universidad de Maryland: en esta universidad había un creciente uso de ambientes virtuales de aprendizaje (AVA), pero no había pruebas de que estos mejoraran el aprendizaje. Se realizó un proyecto de investigación institucional para conocer las correlaciones entre los datos de los AVA y las calificaciones de los estudiantes, y para conocer cómo utilizar mejor las predicciones para apoyarlos. La investigación también permitió analizar las posibles intervenciones y cómo identificar prácticas pedagógicas eficaces a partir de los datos.
- Instituto de Tecnología de Nueva York: desarrollaron su propio modelo predictivo con colaboración de personal de asesoramiento especializado, para identificar con un alto grado de precisión a estudiantes en riesgo. El objetivo era aumentar la retención de los estudiantes en el primer año de sus estudios mediante la creación de un modelo para identificar a los más necesitados de apoyo, y para proporcionar información acerca de la situación de cada alumno, que pudiera ayudar a los docentes en su labor.

- Marist College: desarrollaron un modelo predictivo de código fuente abierto para generar alertas tempranas, el cual ha sido transferido a diferentes instituciones, para definir estrategias de intervención que ayuden a estudiantes en riesgo.

- Universidad de Cowan: implementaron la tecnología C4S, que identifica automáticamente a los estudiantes que necesitan apoyo. Esta tecnología, permite al personal docente tener contacto con un gran número de estudiantes y gestionar una serie de intervenciones para cada uno.

- Universidad de Nueva Inglaterra: desarrollaron un proceso dinámico, sistemático y automatizado que captura el estado de bienestar de los estudiantes, el cual toma como base tres componentes principales: e –Motion (emoticones para expresar el estado de ánimo; feliz, neutro, triste), el Vibe (cajas de texto que reúnen los comentarios de los estudiantes) y el AWE (motor automatizado que analiza los datos procedentes de los sistemas corporativos y sugiere comportamientos de riesgo).

- Open University: implementó un programa de análisis estratégico de aprendizaje para mejorar el éxito de los estudiantes mediante la incrustación de la toma de decisiones basada en la evidencia en todos los niveles. Esta universidad ha venido desarrollando sus capacidades institucionales para fortalecer el análisis del aprendizaje, las cuales se organizan en torno a tres ejes principales: disponibilidad de datos, análisis y creación de conocimiento y procesos que influyen en el éxito de los estudiantes.

- Nottingham Trent University: implementó procesos para mejorar la retención, aumentar el sentido de pertenencia, y mejorar los resultados institucionales.

- Universidad de Wollongong: implementaron procesos que vinculan el uso de redes sociales a las prácticas pedagógicas (SNAPP: Social Networks Adapting Pedagogical Practice). Estos procesos permiten analizar discusiones en línea, descubrir patrones en foros, los cuales son moderados por un instructor, las intervenciones realizadas por el instructor pueden tener un impacto significativo en la experiencia de aprendizaje de los estudiantes.

- Universidades Abiertas de Australia: al ser una institución que ofrece educación a distancia, adaptaron un sistema para realizar la planificación personalizada para cada estudiante, el cual propone módulos alternativos y apoyo virtual según las necesidades de los mismos.

CONCLUSIONES

El manejo estratégico y sistematizado de grandes y complejas bases de datos está permitiendo solucionar problemáticas que se presentan en la educación. El uso de diferentes técnicas computacionales para el análisis de grandes volúmenes de información; estadística computacional, minería de datos, investigación de operaciones, matemática aplicada, optimización, aprendizaje estadístico y de máquinas, modelado estratégico, entre otras, permiten agregar valor a los datos y generar ventajas en la educación.

Debido al incremento vertiginoso de información en el sector educativo, y al ritmo acelerado de crecimiento de la educación virtual y a distancia, se hace necesario incrementar el número de expertos calificados en el análisis de grandes volúmenes de datos, especialmente en países de América Latina. En la educación, el análisis del aprendizaje se constituye en una herramienta de gran utilidad para mejorar la calidad académica y los procesos de enseñanza aprendizaje.

REFERENCIAS

- Aguiar, E., Chawla, N. V., Brockman, J., Ambrose, G. A., y Goodrich, V. (2014). Engagement vs performance: using electronic portfolios to predict first semester engineering student retention. In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (pp. 103–112). ACM.
- Asif, R., Merceron, A., y Pathan, M. K. (2015). Investigating performance of students: a longitudinal study. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (pp. 108–112). ACM.
- Cambuzzi, W. L., Rigo, S. J., y Barbosa, J. L. V. (2015). Dropout Prediction and Reduction in Distance Education Courses with the Learning Analytics Multitrail Approach. *J. UCS*, 21(1), 23–47.
- Clow, D. (2014). Data wranglers: human interpreters to help close the feedback loop. In Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (pp. 49–53). ACM.
- Corrigan, O., Smeaton, A. F., Glynn, M., y Smyth, S. (2015). Using Educational Analytics to Improve Test Performance. In *Design for Teaching and Learning in a Networked World* (pp. 42–55). Springer.
- Diebold, F. (2001). *Elements of forecasting*. (South y W. C. Publishing., Eds.) (2nd ed.). Australia.
- Drachler, H., y Greller, W. (2012). The pulse of learning analytics understandings and expectations from the stakeholders. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 120–129). ACM.
- Friedman, V. (2008). Data visualization and infographics. *Graphics, Monday Inspiration*, 14, 2008.
- Gasevic, D., Kovanovic, V., Joksimovic, S., y Siemens, G. (2014). Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. *The International Review of Research in Open and Distributed Learning*, 15(5).
- Gibson, A., Kitto, K., y Willis, J. (2014). A cognitive processing framework for learning analytics. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 212–216). ACM.
- Grann, J., y Bushway, D. (2014). Competency map: Visualizing student learning to promote student success. In *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 168–172). ACM.
- Grau-Valldosera, J., y Minguillón, J. (2011). Redefining dropping out in online higher education: a case study from the UOC. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 75–80). ACM.
- Harrison, S., Villano, R., Lynch, G., y Chen, G. (2015). Likelihood analysis of student enrollment outcomes using learning environment variables: A case study approach. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 141–145). ACM.
- Hecking, T., Ziebarth, S., y Hoppe, H. U. (2014). Analysis of dynamic resource access patterns in a blended learning course. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 173–182). ACM.
- Holman, C., Aguilar, S., y Fishman, B. (2013). GradeCraft: What can we learn from a game-inspired learning management system? In *Proceedings of the Third International Conference on Learning Analytics and*

- Knowledge (pp. 260–264). ACM.
- KDNuggets. (2017). Data Science Statistics 101. Retrieved June 3, 2017, from <http://www.kdnuggets.com/2016/07/data-science-statistics-101.html>
- KDNuggets. (2017a). Association Rules and the Apriori Algorithm: A Tutorial. Retrieved March 7, 2016, from <http://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html>
- KDNuggets. (2017b). Comparing Clustering Techniques: A Concise Technical Overview. Retrieved June 3, 2017, from <http://www.kdnuggets.com/2016/09/comparing-clustering-techniques-concise-technical-overview.html>
- KDNuggets. (2017c). Ensemble Methods: Elegant Techniques to Produce Improved Machine Learning Results. Retrieved March 6, 2017, from <http://www.kdnuggets.com/2016/02/ensemble-methods-techniques-produce-improved-machine-learning.html>
- KDNuggets. (2017d). Regression Analysis: A Primer. Retrieved March 6, 2017, from <http://www.kdnuggets.com/2017/02/regression-analysis-primer.html>
- KDNuggets. (2017e). Text Analytics, Text Mining. Retrieved March 6, 2017, from <http://www.kdnuggets.com/2015/04/statisticscom-text-analytics-text-mining-courses.html>
- KDNuggets. (2017f). Top Algorithms for Analytics. Retrieved March 7, 2017, from <http://www.kdnuggets.com/2011/11/algorithms-for-analytics-data-mining.html>
- Khousa, E. A., y Atif, Y. (2014). A Learning Analytics Approach to Career Readiness Development in Higher Education. In *International Conference on Web-Based Learning* (pp. 133–141). Springer.
- Khousa, E. A., Atif, Y., y Masud, M. M. (2015). A social learning analytics approach to cognitive apprenticeship. *Smart Learning Environments*, 2(1), 14.
- Kim, J., Jo, I.-H., y Park, Y. (2016). Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review*, 17(1), 13–24.
- Kitchenham, B. A., y Charters, S. (2007). *Procedures for Performing Systematic Literature Reviews in Software Engineering*. Keele University y Durham University, UK.
- Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., y Hatala, M. (2015). Penetrating the black box of time-on-task estimation. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 184–193). ACM.
- Lockyer, L., y Dawson, S. (2011). Learning designs and learning analytics. In *Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 153–156). ACM.
- Lotsari, E., Verykios, V. S., Panagiotakopoulos, C., y Kalles, D. (2014). A learning analytics methodology for student profiling. In *Hellenic Conference on Artificial Intelligence* (pp. 300–312). Springer.
- Manoso-Vázquez, M., y Llamas-Nistal, M. (2015). A Monitoring System to Ease Self-Regulated Learning Processes. *IEEE Revista Iberoamericana de Tecnologías Del Aprendizaje*, 10(2), 52–59.
- Mayor, E. (2015). Learning Predictive Analytics with R.
- McKay, T., Miller, K., y Tritz, J. (2012). What to do with actionable intelligence: E 2 Coach as an intervention engine. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 88–91). ACM.
- Menchaca, I., Guenaga, M., y Solabarrieta, J. (2015). Project-Based Learning: Methodology and Assessment Learning Technologies and Assessment Criteria. In *Design for Teaching and Learning in a Networked World* (pp. 601–604). Springer.
- Nespereira, C. G., Elhariri, E., El-Bendary, N., Vilas, A. F., y Redondo, R. P. D. (2016). Machine Learning Based Classification Approach for Predicting Students Performance in Blended Learning. In *The 1st International Conference on Advanced Intelligent System and Informatics (AISII2015)*, November 28-30, 2015, Beni Suef, Egypt (pp. 47–56). Springer.
- Øhrstrøm, P., Sandborg-Petersen, U., Thorvaldsen, S., y Ploug, T. (2013). Teaching logic through web-based and gamified quizzing of formal arguments. In *European Conference on Technology Enhanced Learning* (pp. 410–423). Springer.
- Pardo, A., Mirriahi, N., Dawson, S., Zhao, Y., Zhao, A., y Gašević, D. (2015). Identifying learning strategies associated with active use of video annotation software. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 255–259). ACM.
- Peña, A. (2017). *Learning Analytics: Fundamentals, Applications, and A View of the Current State of the Art* (94th ed.). Mexico DF: Springer Berlin Heidelberg.
- Piety, P. J., Hickey, D. T., y Bishop, M. J. (2014). Educational data sciences: framing emergent practices for analytics of learning, organizations,

- and systems. In Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (pp. 193–202). ACM.
- Pravilovic, S., Bilancia, M., Appice, A., y Malerba, D. (2017). Using multiple time series analysis for geosensor data forecasting. *Information Sciences*, 380, 31–52. <https://doi.org/http://dx.doi.org/10.1016/j.ins.2016.11.001>
- Prinsloo, P., Slade, S., y Galpin, F. (2012). Learning analytics: challenges, paradoxes and opportunities for mega open distance learning institutions. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 130–133). ACM.
- Ray, S. (2015). Understanding Support Vector Machine algorithm from examples. Retrieved March 6, 2017, from <https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code/>
- Romero, C., y Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 40(6), 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- Scheffel, M., Niemann, K., Leony, D., Pardo, A., Schmitz, H.-C., Wolpers, M., y Kloos, C. D. (2012). Key action extraction for learning analytics. In *European Conference on Technology Enhanced Learning* (pp. 320–333). Springer.
- Scikit-Learn. (2017). Decision Trees. Retrieved March 6, 2017, from <http://scikit-learn.org/stable/modules/tree.html#multi-output-problems>
- Simsek, D., Sándor, Á., Shum, S. B., Ferguson, R., De Liddo, A., y Whitelock, D. (2015). Correlations between automated rhetorical analysis and tutors' grades on student essays. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (pp. 355–359). ACM.
- Sinclair, J., y Kalvala, S. (2015). Engagement measures in massive open online courses. In *International Workshop on Learning Technology for Education in Cloud* (pp. 3–15). Springer.
- Swenson, J. (2014). Establishing an ethical literacy for learning analytics. In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (pp. 246–250). ACM.
- Tervakari, A.-M., Marttila, J., Kailanto, M., Huhtamäki, J., Koro, J., y Silius, K. (2013). Developing learning analytics for TUT circle. In *Open and Social Technologies for Networked Learning* (pp. 101–110). Springer.
- Vahdat, M., Oneto, L., Anguita, D., Funk, M., y Rauterberg, M. (2015). A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator. In *Design for Teaching and Learning in a Networked World* (pp. 352–366). Springer.
- Vozniuk, A., Holzer, A., y Gillet, D. (2014). Peer assessment based on ratings in a social media course. In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (pp. 133–137). ACM.
- Westera, W., Nadolski, R., y Hummel, H. (2013). Learning analytics in serious gaming: uncovering the hidden treasury of game log files. In *International Conference on Games and Learning Alliance* (pp. 41–52). Springer.
- Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (pp. 203–211). ACM.