

**MINERÍA DE DATOS EDUCACIONALES: DESCUBRIR TESOROS OCULTOS
DURANTE EL APRENDIZAJE**

**EDUCATIONAL DATA MINING: DISCOVER HIDDEN TREASURES DURING
LEARNING**

Roberto García Sánchez, Mgtr.

 <https://orcid.org/0000-0001-5631-8892>

Universidad Católica de Santiago de Guayaquil (Guayaquil, Ecuador)

roberto.garcia02@cu.ucsg.edu.ec

Jorge Gil Mateos, Ph.D.

 <https://orcid.org/0000-0002-5384-6419>

Universidad de La Habana (La Habana, Cuba)

jorgegil@cepes.uh.cu

ARTÍCULO DE REVISIÓN BIBLIOGRÁFICA

Recibido: 12 de agosto de 2023

Aceptado: 11 de diciembre de 2023

RESUMEN

El presente trabajo aborda el tema de la minería de datos educacionales, una disciplina relativamente joven, con el objetivo de presentar definiciones, conceptos y relaciones sobre minería de datos, huella digital y minería de datos educacionales. La metodología empleada para este objetivo se sustenta en una amplia revisión bibliográfica sobre el estado de desarrollo de la minería de datos educacionales. Además, se presentan algunas de las herramientas, técnicas y utilidades más significativas en este campo, como la clasificación y predicción, clustering, detección de valores extremos, minería de relaciones, análisis de redes sociales, procesos de minería, minería de texto, procesamiento para análisis humano y descubrimiento mediante modelos. Como principal conclusión, se destaca la imperativa necesidad de incorporar la minería de datos en el ámbito educacional, con el propósito de esclarecer el camino a los docentes, encargados de administrar los procesos de formación, y autoridades a que se interesen en incursionar en este campo en favor de la educación virtual o en línea para revalorizar la educación y favorecer el proceso de enseñanza-aprendizaje en favor del estudiante.



Palabras clave: Minería de datos, minería de datos educacionales, big data, huella digital

ABSTRACT

The present work addresses the topic of educational data mining, a relatively young discipline, with the objective of presenting definitions, concepts, and relationships about data mining, digital footprint, and educational data mining. The methodology used for this objective is based on an extensive bibliographic review of the state of development of educational data mining. Additionally, some of the most significant tools, techniques, and utilities in this field are presented, such as classification and prediction, clustering, detection of extreme values, relationship mining, social network analysis, mining processes, text mining, processing for human analysis, and discovery through models. The main conclusion is the imperative need to incorporate data mining in the educational field, with the purpose of clarifying the path for teachers, those in charge of managing training processes, and authorities interested in entering this field in favor of virtual or online education to revalue education and favor the teaching-learning process in favor of the student.

Keywords: Data mining, educational data mining, big data, digital fingerprint

INTRODUCCIÓN

La digitalización de la mayoría de los ámbitos del quehacer humano ha supuesto la generación de una inmensa cantidad de datos que son colectados por sistemas y que posteriormente son analizados y pueden recrear el comportamiento exacto que ha desarrollado un proceso. Es tan grande el volumen de datos, que el entendimiento humano es sobrepasado, y se requiere aplicar técnicas y procesos que permitan depurar los datos, y generar información útil (Romero, Ventura, et al., 2010; Uddin & Lee, 2016). Esta información es cada día más valiosa, quien posea la información, poseerá el poder; una empresa que conozca el comportamiento de sus clientes, con la información suficiente será capaz de ofrecerle lo que necesita, posiblemente incluso con anterioridad a que el propio cliente sea consciente de ello. La información no permite predecir el futuro, ni tan siquiera poder predecir una catástrofe natural, o qué número saldrá premiado en la lotería, pero sí puede decir mucho sobre el comportamiento de las cosas, y en base al aprendizaje de similitudes de casos, poder identificar y predecir con un alto porcentaje situaciones que puedan ser utilizadas en beneficio (Huebner, 2013; Jiménez & Álvarez, 2010; Maheshwari, 2015; Morales, Trujillo, & Santana, 2019).

Trabajar con datos significa que es necesario coleccionarlos y transformarlos para poder convertirlos en información. Para conseguir esto, se deben afianzar procesos estadísticos, análisis y técnicas computacionales de cierta complejidad. Por ejemplo, en el ámbito de los negocios, debido al interés de incrementar la rentabilidad e ingresos económicos, se han desarrollado con creces diversas técnicas para explotar los datos, desde índices bursátiles, inversiones, proyecciones, etc.; algunas de las aplicaciones más importantes y destacadas son:

- Diseño de estrategias de negocio¹ basadas en información concreta.
- Conocer las preferencias de los usuarios.
- Facilitar la búsqueda de información relevante.
- Realizar predicciones sobre el comportamiento de los clientes.
- Detectar el riesgo de abandono de los clientes.

Sin embargo, hay muchos otros campos que han realizado significativos avances, como en medicina, que, con la ayuda de diagnósticos asistidos por inteligencia artificial, o la recolección de diferentes datos de los pacientes, alimentación, pasos caminados, signos vitales, permiten a los médicos realizar un mejor seguimiento de sus pacientes (Galetsi, Katsaliaki, & Kumar, 2020; George, George, & George, 2020).

En la educación también ha existido un progreso significativo en el empleo de la minería de datos y el Big Data², se han realizado diversas implementaciones con casos de éxito, algunos de los cuales se presentarán más adelante en este trabajo.

La detección de datos digitales de los usuarios es un proceso que permite coleccionar o registrar todo el recorrido e intercambio de los usuarios con sus pares, o con la plataforma digital en concreto. Esa marca que dejan los usuarios, y que permite rastrear una conducta, es lo que se ha dado en llamar huella digital.

La huella digital en la educación

Se considera que la huella digital es el conjunto de datos que se genera, considerando que todos los usuarios interactúan con la plataforma de teleformación o Espacio Virtual de Enseñanza y Aprendizaje (EVEA), y entre ellos, cada actividad queda registrada, caracterizando de este modo a cada usuario. Así los antecedentes y los datos de rendimiento del alumno, cada acción que realice (leer un archivo, participar en un foro, enviar mensajes o visitar enlaces

¹ Se entiende como estrategia de negocio al conjunto de actividades que involucra la formulación e implementación de los principales objetivos e iniciativas en cualquier área, no es exclusiva a los negocios.

² Es el término en inglés utilizado para denominar un gran volumen de datos estructurados o no estructurados, que se almacenan por para luego ser analizados.

recomendados, por ejemplo) deja una huella digital (Calvet & Juan, 2015). La información que es registrada y almacenada, y en general en la minería de datos ha sido caracterizada por un modelo de 4V's por IBM (IBM, 2017) y extendido al modelo presentado por Khan, Uddin y Gupta, de 7V's (Khan, Uddin y Gupta, 2014), en los que consta:

1. Volumen: tamaño de la información generada, dependerá de la cantidad de usuarios, actividades y tiempo.
2. Velocidad: la velocidad con la que se genera la información, cuanta más información se registre, más rápidamente crecerá la cantidad total.
3. Variedad: diferentes formas de datos, por cada tipo de acción, tarea, actividad.
4. Veracidad: autenticidad de los datos, que los datos registrados se correspondan a la realidad.
5. Validez: corrección y precisión con la que han sido registrado los datos.
6. Volatilidad: retención relevancia e importancia, qué tanto tiempo serán válidos los datos sin perder su valor.
7. Valor: resultado deseado, que permita obtener la información necesaria para la toma de decisiones.

En los EVEA, el objetivo principal de la minería de datos consiste en extraer la información oculta de un conjunto de datos, la huella digital de los actores, es decir, obtener conocimiento útil y comprensible (Jaramillo & Paz-Arias, 2015; Pérez-Palacios et al., 2014), que permita la retroalimentación los procesos de enseñanza-aprendizaje, en todas sus etapas, desde su diseño, implementación como desarrollo de este. Estas técnicas son relativamente jóvenes en la educación, y aún no son incluidas en las prácticas comunes de la inmensa mayoría de las comunidades docentes, por lo que la investigación científica y la innovación tienen un nicho amplio de desarrollo.

Por su relevancia en la esfera educativa se abordarán en los siguientes apartados, de una manera más extensa, las características de la minería de datos aplicada al área de la educación, exponiendo las herramientas, técnicas y utilidades más significativas al respecto; intentando pautar una suerte de ruta por donde pueden conducirse intervenciones educativas.

REVISIÓN DE LA LITERATURA

La Minería de Datos Educativas

La educación es uno de los procesos más importantes en la formación del ser humano, tanto a nivel individual, como colectivo. Individual, pues permitirá a cada persona formarse, adquirir

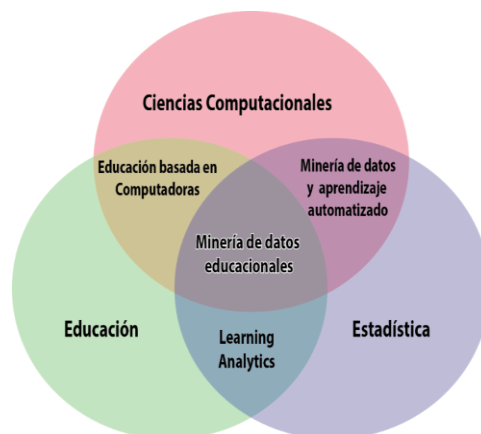
conocimientos, capacidades y habilidades que le permitan desarrollarse como persona y profesionalmente. Y colectivo, porque es la educación que reciba la población la que depare su futuro, los niños de hoy serán los gobernantes de la próxima generación. Por tanto, parece algo por lo cual merece la pena apostar, y en este sentido, aprovechar la minería de datos puede ayudar a conocer mejor el proceso educativo, identificar debilidades, incrementar las fortalezas, saber más de lo que ocurre en los procesos educativos con tecnología, entre otras aplicaciones (Baker, 2010; Baker & Inventado, 2014; Shaun et al., 2008; Siemens & Baker, 2012a).

Entre los retos para ofrecer una mayor calidad educativa que enfrentan actualmente las instituciones de educación superior está el de mejorar el rendimiento académico de sus estudiantes (Menacho Chiok, 2017), y frenar el abandono de los estudios superiores, un fenómeno que repercute directamente en el desaprovechamiento de recursos (García, et al., 2016). Entendiendo que el rendimiento es reflejo directo de los aprendizajes de los estudiantes, por ello es hacia el aprendizaje hacia donde se deben dirigir los esfuerzos con la minería de datos educacionales.

La base de la minería de datos educacionales (EDM, de sus siglas en inglés, Educational Data Mining) reside en la intersección de tres grandes áreas del conocimiento: educación, estadística y ciencias computacionales. En la Figura 1 se puede observar cómo interacciones entre estas ciencias generan subcampos especializados, como son la educación basada en computadoras, la minería de datos y aprendizaje automatizado, analítica de aprendizaje (LA, de sus siglas en inglés Learning Analytics) y su triple intersección, la minería de datos educacionales (Romero & Ventura, 2013; Siemens & Baker, 2012b).

Figura 1.

Principales áreas relacionadas con la minería de datos educacionales



Nota. Tomado de Romero y Ventura (Cristobal Romero & Ventura, 2013)

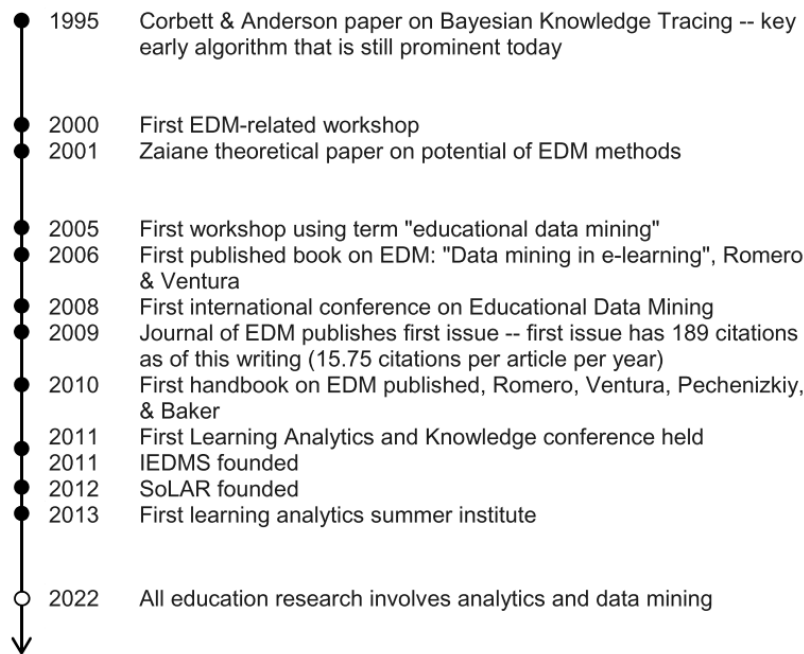
Evolución de la Minería de Datos Educativos

La EDM puede ser vista desde dos perspectivas: una comunidad de investigación o un área de investigación científica. Desde la visión de comunidad de investigación se considera cercana al LA, de hecho, muchos de los miembros pertenecen a ambas comunidades, con colaboraciones o incluso constituyéndose como una competencia amistosa. EDM surge por primera ocasión en 2005 en un taller, y se convirtieron en una conferencia anual a partir de 2008, siendo en 2009 cuando se conforma como una revista científica y se crea la International Educational Data Mining Society en el año 2011 (Baker & Inventado, 2014) (Figura 2). Desde la perspectiva de área de investigación científica, EDM se relaciona con el análisis a gran escala de datos educativos, enfocado en métodos automatizados. En este sentido, comparten muchas similitudes con LA. Se han conformado dos grupos de investigación enfocados al análisis de la información registrada por los EVEA, por un lado, la minería de datos educativos y por el otro, analítica de aprendizaje, ambos han cobrado un gran interés por varios motivos:

- a. Existe interés en emplear los datos para realizar una mejor toma de decisiones, como sucede en la inteligencia de negocios o analítica (Daradoumis, et al., 2010).
- b. Existe estadística potente, machine-learning y métodos de minería de datos y técnicas para búsqueda de patrones y de construcción de modelos de predicción o reglas de decisión que pueden ser adaptadas de manera fácil a los datos educativos;
- c. generar información es relativamente fácil, y las capacidades de la tecnología en cuanto a almacenamiento y procesamiento permiten gestionarlo de manera solvente;
- d. debido a la crisis financiera y la fuerte competencia, las universidades están bajo presión para reducir costos e incrementar los ingresos provenientes de las demandas crecientes en educación de países en desarrollo, teniendo en cuenta una reducción de la deserción y mejorar la calidad de los cursos.

Figura 2.

Línea de tiempo de los hitos más significativos en EDM



Nota. Modificado a partir de Baker, & Inventado (2014)

La Sociedad Internacional de Minería de Datos Educativo,s considera la EDM como una disciplina emergente, preocupada por desarrollar métodos para explorar los tipos únicos de datos que provienen de entornos educativos, y usarlos para comprender mejor a los estudiantes y los entornos que en que aprenden. Por su parte, la Sociedad para la Investigación de Learning Analytics asume a las Learning Analytics como la medición, recopilación, análisis e informe de datos sobre los alumnos y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que se produce (Siemens & Baker, 2012b).

La principal meta de EDM y LA es extraer información de los datos educacionales para poder dar soporte a la toma de decisiones relacionadas con la educación (Daradoumis, Juan, Lera-López, Faulin, 2010) y la necesidad de tener una retroalimentación objetiva para que los instructores evalúen sus cursos y su eficacia en el proceso de aprendizaje, monitorear el proceso de aprendizaje, con el fin de ayudar a detectar dificultades rápidamente, mejorar los contenidos. Sin embargo, puede convertirse en un proceso complejo, y que consume bastantes recursos y tiempo si no se realiza con las herramientas adecuadas (Juan, Daradoumis, Faulin, & Xhafa, 2009).

Entre las diferencias está que los investigadores en EDM están más interesados en métodos automatizados para el descubrimiento de datos educativos, mientras que los investigadores de LA están más interesados en los métodos dirigidos por humanos para explorar los datos educativos. Los métodos automatizados para el descubrimiento pueden ayudar a lograr la mejor predicción posible; y los métodos de descubrimiento dirigidos por humanos pueden dar como resultado modelos de fenómenos más interpretables y comprensibles (Siemens & Baker, 2012a).

Otra diferencia es que los investigadores en EDM enfatizan el modelado de construcciones específicas y las relaciones entre ellas; y los investigadores en LA enfatizan una comprensión más holística de los sistemas de las construcciones. Los investigadores en EDM buscan aplicaciones en la adaptación automatizada, como apoyar a los alumnos mediante el uso de software educativo para identificar una necesidad y cambiar automáticamente para personalizar la experiencia del alumno (Arroyo et al., 2007; Baker et al., 2006; Corbett & Anderson, 1994); y los investigadores en LA buscan formas de informar y capacitar a los instructores y aprendices, como informar a los instructores sobre las formas en que los estudiantes específicos están luchando, para que el instructor pueda contactar al alumno (Arnold, 2010).

Ambas tendencias tienen un fuerte énfasis entre la teoría en las ciencias del aprendizaje y la filosofía de la educación; y en muchas ocasiones los investigadores que publican en las conferencias EDM y LA utilizan la teoría de las ciencias del aprendizaje y la educación para guiar su elección de análisis, y su objetivo es contribuir a la teoría con los resultados de sus análisis (Siemens & Baker, 2012a).

Hasta aquí queda más claro que el escenario en que se desarrollan estas dos tendencias (EDM y LA) es un escenario de investigadores y de docentes con un alto nivel de motivación por la inclusión de la minería a favor del aprendizaje en EVEA. Por lo que su introducción depende de desarrollos tecnológicos no generalizados hasta hoy.

Aplicaciones de la Minería de Datos Educativas

La minería de datos educativos es un campo de estudio que utiliza técnicas de análisis de datos para descubrir patrones, tendencias y relaciones en conjuntos de datos educativos. Estas aplicaciones son variadas y abarcan desde la mejora de la enseñanza hasta la toma de decisiones institucionales. A continuación, se describen algunas de las aplicaciones más relevantes de la minería de datos educativos:

- **Predicción del Rendimiento Estudiantil:** utilización de modelos predictivos para identificar a estudiantes en riesgo de bajo rendimiento (Romero & Ventura, 2010).
- **Personalización del Aprendizaje:** adaptación de métodos de enseñanza y recursos según las necesidades individuales de los estudiantes (Baker & Siemens, 2014).
- **Mejora de la Retención Estudiantil:** identificación de factores que contribuyen a la deserción estudiantil y desarrollo de estrategias para mejorar la retención (Arnold & Pistilli, 2012).
- **Evaluación de Efectividad Docente:** análisis de datos para evaluar el impacto de las estrategias de enseñanza y la eficacia de los docentes (Kotsiantis, Pierrakeas, & Pintelas, 2004).
- **Detección de Plagio:** identificación de posibles casos de plagio mediante el análisis de patrones de similitud en trabajos académicos (Sánchez-Rada, Iglesias, & García-Serrano, 2011).
- **Optimización de Programas Académicos:** utilización de datos para evaluar y mejorar la eficacia de los programas académicos (Romero, López, Luna, & Ventura, 2013).

Si bien estas son algunas de las aplicaciones más relevantes, se requiere puntualizar que el aprovechamiento de la información dependerá en gran medida en el diseño de los recursos que conforman los cursos, tareas, actividades y otros elementos que forman parte del proceso de enseñanza aprendizaje mediante un EVEA. Por tanto, las utilidades que se presentan pueden no ser válidas en cualquier escenario, más bien requiere que se considere en ciertas condiciones en los diseños de los cursos para una explotación óptima.

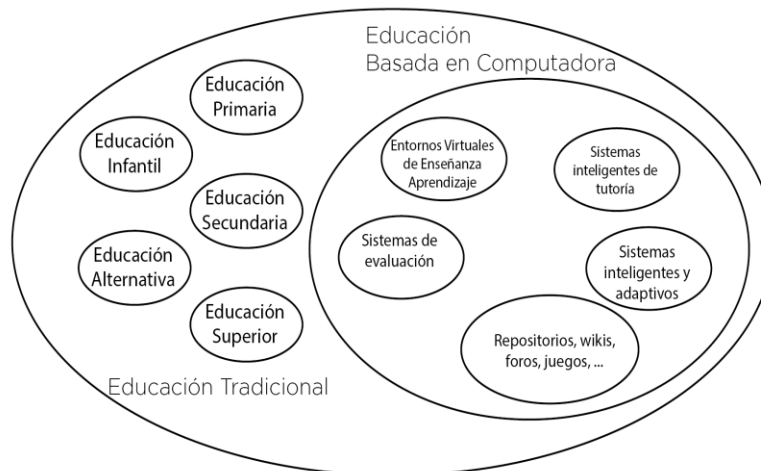
El proceso de la Minería De Datos Educativas

Es necesario comprender el proceso de enseñanza-aprendizaje que se desarrolla en un EVEA para poder realizar un correcto aprovechamiento al realizar el diseño de una estrategia de análisis de los datos educativos registrados en la plataforma. En el modelo de educación tradicional la mayoría del proceso de enseñanza-aprendizaje se realiza de manera presencial y síncrona, se debe cambiar hacia el paradigma de la educación basada en el uso de la computación, y los nuevos instrumentos que aparecen, tales como los EVEA, sistemas de evaluación, sistemas inteligentes de tutoría, repositorios, los Entornos Personales de Aprendizaje (PLE, del inglés Personal Learning Environment) wikis, etc. Es así como los diferentes niveles de educación incorporan elementos de la educación basada en computadora (Ver Figura 3), dando lugar a nuevos procesos, actividades, y en general, un mayor protagonismo del estudiante, y por ello se vuelve indispensable conocer comportamientos, patrones y cualquier información que

conlleve a desarrollar con éxito los cursos que implican el uso de TIC (Jiménez & Álvarez, 2010; Papamitsiou & Economides, 2014; Ray & Saeed, 2018; Romero, Ventura, & García, 2008; Romero & Ventura, 2013).

Figura 3.

Tipos de educación tradicional y basada en ambientes y sistemas computacionales



Nota. Tomado de Romero y Ventura, (2013)

Se pueden considerar diferentes perspectivas, por una parte, desde el punto de vista educacional y experimental, se puede ver como un ciclo iterativo de la hipótesis formativa, probar y refinar. En este proceso, la meta no es únicamente convertir los datos en conocimiento, sino filtrar el conocimiento para la toma de decisiones sobre como modificar el entorno educacional para fortalecer el aprendizaje de los estudiantes. Este tipo de evaluación formativa de un programa se lleva a cabo durante su desarrollo, con el propósito de mejorar continuamente el mismo. Otro aspecto a analizar puede ser el uso de los estudiantes del sistema, es una forma de evaluar el diseño instruccional, sea cual sea el modelo de diseño instruccional elegido, de una manera formativa, de tal manera que ayude a los diseñadores educacionales a mejorar el material instruccional y a descubrir modelos o patrones que pueden ser usados para establecer una base pedagógica para la toma de decisiones cuando se diseña o modifica un ambiente pedagógico (Baker & Inventado, 2014; Romero & Ventura, 2013).

Visto así, hasta aquí queda más claro otro grupo de destinatarios de la EDM: los diseñadores instruccionales; además de los investigadores y docentes altamente motivados por la inclusión de la minería de datos educacionales.

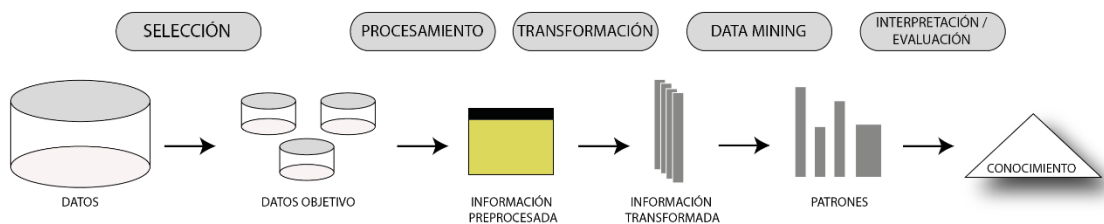
Desde el punto de vista de la minería de datos, la EDM puede ser similar a cualquier otro proceso de minería de datos, aunque más adelante se abordarán las diferencias y características específicas en cada paso.

El proceso de aplicación de la minería de datos (EPM, de sus siglas del inglés, Educational Process Mining) contempla la manipulación de la información registrada en los EVEA para descubrir, analizar y obtener una representación gráfica que ayude a comprender de manera más sencilla la información que se encierra en estos datos (Bogarín, Cerezo, & Romero, 2018).

En la Figura 4 se puede observar el proceso completo que conlleva el EPM, se parte de una abundante colección de datos, que deben ser seleccionados, de tal manera que se descarten aquellos que no sean útiles para el objetivo fijado. El siguiente paso consiste en realizar un procesamiento previo a la transformación que permita agrupar, y preparar los datos para que se puedan aplicar las técnicas de minería de datos, y que a partir de la detección de patrones se realice la interpretación y evaluación de la información que se ha transformado en conocimiento.

Figura 4.

Proceso de la minería de datos educacionales



Para un mejor entendimiento, se describen en la siguiente tabla algunas de los datos que se pueden considerar para la explotación de los datos que se registran en los EVEA, de tal manera que se pueda aplicar minería de datos educacionales y obtener información valiosa.

Tabla 1.

Variables clave que se pueden considerar al realizar EDM en datos de un EVEA

| Variable | Descripción |
|--------------------|--|
| Datos Demográficos | <ul style="list-style-type: none"> • Identificación del estudiante (ID del estudiante) • Género • Edad • Antecedentes académicos |

| | |
|---|--|
| Participación y Actividad | <ul style="list-style-type: none"> • Número de sesiones iniciadas. • Duración de cada sesión. • Frecuencia de participación en actividades en línea • Número de interacciones en foros y discusiones. • Participación en grupos de estudio virtuales. • Acceso a recursos educativos (lecturas, videos, etc.). |
| Rendimiento Académico | <ul style="list-style-type: none"> • Calificaciones en asignaturas específicas. • Resultados de evaluaciones y exámenes. • Tasas de finalización de actividades y cursos. • Tiempo dedicado a tareas y actividades específicas. |
| Interacciones con Contenido | <ul style="list-style-type: none"> • Navegación a través de módulos y lecciones. • Visualización de contenido multimedia. • Descargas de materiales de estudio. • Interacciones con simulaciones y actividades interactivas. |
| Comportamiento en Foros y Comunidades | <ul style="list-style-type: none"> • Participación en discusiones. • Publicación de preguntas y respuestas. • Retroalimentación dada y recibida. • Colaboración con otros estudiantes. |
| Conexiones y Comunicación | <ul style="list-style-type: none"> • Frecuencia y duración de la comunicación con profesores y compañeros. • Uso de herramientas de mensajería interna del EVEA. • Participación en sesiones de chat en tiempo real. |
| Historial de Navegación y Patrones de Uso | <ul style="list-style-type: none"> • Páginas web visitadas dentro del EVEA. • Secuencia de acciones realizadas durante una sesión. • Patrones de búsqueda dentro de la plataforma. |
| Variables Temporales | <ul style="list-style-type: none"> • Horario de acceso al EVEA. • Días y horas de mayor actividad. • Frecuencia de acceso en días específicos de la semana. |
| Indicadores de Deserción | <ul style="list-style-type: none"> • Periodos de inactividad prolongados. • Reducción en la frecuencia de participación. • Indicadores de falta de interés o compromiso. |
| Datos de Retroalimentación | <ul style="list-style-type: none"> • Encuestas de satisfacción del estudiante. • Retroalimentación sobre la usabilidad del EVEA. • Comentarios y sugerencias proporcionados por los estudiantes. |

Nota. Adaptado de varios autores

Por otra parte, y debido a la importancia para su comprensión por parte de desarrolladores e innovadores, se van a detallar cada uno de los pasos que conlleva el EPM (Bogarín, Cerezo, & Romero, 2018):

Selección de datos

La selección de los datos a recolectar depende del tipo de entorno virtual de enseñanza aprendizaje que se utiliza, y puede ser obtenida de diferentes fuentes, desde aspectos administrativos, observaciones, cuestionarios, medidas de uso de la propia plataforma, incluso hasta las calificaciones parciales y finales. Este paso puede parecer sencillo, pero es la base de todo el trabajo a desarrollar, por lo que determinar los datos que serán objeto del análisis es esencial.

Procesamiento

En el contexto de la educación, el proceso de la información es una tarea compleja y muy importante, y en muchas ocasiones ocupa gran parte del tiempo dedicado a la minería de datos educacionales. En primer lugar, la automatización del registro de la información generada es muy importante, ya que se manejará gran cantidad de información que será necesaria preprocesar. Los datos que se registran no siempre están en el formato apropiado, o la heterogeneidad de los mismos debido a la propia naturaleza de los datos educacionales conlleva la definición de estructuras de datos que representen de manera estandarizada, y respondan al tipo de problema que se desea resolver. Si es necesario se convierten los datos a un formato apropiado. El procesamiento conlleva la recolección de información que sea objeto de estudio, enfocándose en las cuestiones a ser resueltas, y asegurándose que está alineada a las preguntas planteadas.

Los EVEA pueden almacenar una abrumadora cantidad de datos de múltiples fuentes con diferentes formatos y niveles de granularidad³, y de significado que proporciona más o menos datos según sea el caso. Existen diversas variables sobre cada estudiante que pueden ser recolectadas para su análisis (Romero & Ventura, 2013), en la tabla se puede ver algunas de las que se podría tener en cuenta. Pero no solo en cuanto a los estudiantes, sino que es necesario tener en cuenta que cada dato registrado está asociado a algún actor, y requiere que sean manejados de una manera correcta, teniendo siempre en cuenta la confidencialidad de la información recolectada.

³ Se refiere a la especificidad a la que se define un nivel de detalle en una tabla, es decir, si hablamos de una jerarquía la granularidad empieza por la parte más alta de la jerarquía, siendo la granularidad mínima, el nivel más bajo.

Transformación

En primer lugar, la información educacional registrada que encontramos no está en el formato adecuado para poder realizar los análisis mediante los algoritmos, por lo que será necesario adecuar, mediante diversos procesos, los datos, a fin de que se adapten a las necesidades del caso.

Minería de datos

La mayoría de las técnicas de minería de datos incluyendo las técnicas de clasificación, clustering, y asociación se han utilizado con éxito en el ámbito de la educación. Sin embargo, las características que presentan los EVEA requieren que se apliquen diferentes tratamientos de información. Se debe diferenciar información sobre el aprendizaje, sobre los estudiantes, y otros elementos que puedan ser de interés. En el siguiente epígrafe se profundiza en los métodos de minería de datos (Romero & Ventura, 2013).

Interpretación y evaluación

El último peldaño del proceso consiste en obtener el conocimiento apropiado para la toma de decisiones que permitan mejorar y fortalecer el entorno educacional o el proceso de enseñanza y aprendizaje, tanto desde el punto de vista del docente, como del estudiante.

Los modelos o patrones resultados de aplicar las técnicas de minería de datos deben ser suficientemente comprensibles como para que permitan ser interpretados por los actores encargados de ejecutar la toma de decisiones. En este sentido, se pueden aplicar técnicas de representación visual, que sean más fáciles de interpretar.

Otra buena forma para mostrar los resultados, información, explicaciones, recomendaciones o comentarios a usuarios que no son expertos en minería de datos son los sistemas recomendadores⁴, que presenten una lista de sugerencias o conclusiones sobre los resultados y como aplicarlos (Romero & Ventura, 2013).

El EPM es laborioso y algo complicado, pero una vez desplegado estaremos en condiciones de disponer de información oculta y útil para los propósitos de intervenciones de mejora de la educación virtual.

⁴ Los sistemas recomendadores forman parte de un sistema de filtrado de información, los cuales presentan distintos tipos de temas o ítems de información que son del interés de un usuario en particular.

Tabla 2.

Esquema proceso de aplicación de minería de datos educacionales

| Fase | Tareas |
|---------------------------------------|---|
| 1. Selección de Datos | <ul style="list-style-type: none"> • Seleccionar datos relevantes del EVEA, como frecuencia de acceso, participación en foros, tiempo dedicado a actividades y calificaciones finales de los estudiantes. |
| 2. Procesamiento | <ul style="list-style-type: none"> • Realizar limpieza de datos para manejar valores faltantes, erróneos, duplicados. • Convertir datos brutos en un formato adecuado para el análisis. |
| 3. Transformación | <ul style="list-style-type: none"> • Transformar los datos en secuencias temporales de actividades para cada estudiante. • Normalizar variables según sea necesario. |
| 4. Minería de Datos | <ul style="list-style-type: none"> • Aplicar algoritmos de minería de patrones de secuencias para descubrir patrones en las actividades de los estudiantes. • Utilizar técnicas de minería de datos • Identificar patrones frecuentes y secuencias significativas. |
| 5. Interpretación y Evaluación | <ul style="list-style-type: none"> • Relacionar los patrones descubiertos con el rendimiento académico de los estudiantes. • Interpretar los resultados para entender cómo ciertos comportamientos se correlacionan con el éxito académico. • Evaluar la eficacia de los patrones descubiertos en predecir el rendimiento académico. |

Técnicas de Minería De Datos Educacionales

En la sección anterior se ha descrito las fases para la aplicación de minería de datos educacionales sobre los datos registrados en un EVEA, y en la fase de “Minería de Datos” se debe aplicar técnicas particulares para el análisis de gran volumen de datos, a continuación, se han recopilado algunas de las técnicas más utilizadas en la minería de datos educacionales. Cabe señalar que es posible realizar combinaciones de estas técnicas, y se da por hecho que no están descritas todas las existentes (Mohamad & Tasir, 2013).

Clasificación y predicción

El objetivo de esta técnica es determinar el valor de una variable independiente correspondiente a un aspecto o atributo de los datos a partir de una combinación de otros aspectos, variables dependientes, mediante técnicas como la clasificación, regresión, estimación

de densidad. A partir de valores pasados esta técnica establece una serie de clases, que permitan crear un modelo que facilite la predicción de valores. Se utiliza especialmente para detectar el comportamiento y desempeño de los estudiantes basado en el comportamiento de estos (Baker & Gowda, s. f.; Kop, 2010; Ray & Saeed, 2018; Romero, Espejo, et al., 2010).

Clustering

Se identifican y agrupan aquellos datos que comparten alguna característica, dando lugar a la segmentación del conjunto de datos en varias categorías de menor tamaño. Es habitual utilizar alguna medida para definir dos instancias similares. En educación se puede usar para agrupar Recursos Educativos Abiertos (REA) similares, o estudiantes con comportamientos similares, o grupos de estudiantes para realizar acciones de recomendación de contenido, entre otros usos (Amershi & Conati, 2009; Romero, Ventura, Pechenizkiy, & Baker, 2010). Esta técnica de agrupamiento se puede utilizar en cualquier ámbito que involucre clasificación, incluso para determinar que tanto han colaborado los usuarios en foros de discusión (Anaya & Boticario, 2009).

Detección de valores más extremos

Consiste en descubrir puntos de la información que son significativamente diferentes del resto de los datos. Un valor extremo es una observación (o medida) que es mayor o menor que el resto de los valores de la información. Es habitualmente usado para detectar estudiantes con dificultades, desviaciones en las acciones o conducta del estudiante o del docente, o procesos irregulares de aprendizaje (Ueno, 2004).

Minería de relaciones

Se identifican las relaciones entre variables y normalmente se codifica como reglas que serán usadas para detectar alguna situación particular. Hay diferentes tipos de relaciones como puede ser la asociación (entre dos o más variables), patrones secuenciales (asociación temporal entre variables), correlación (correlación lineal entre variables), o casual. Es utilizada para determinar relaciones de patrones de comportamiento de los estudiantes y realizar diagnósticos de dificultades de aprendizaje, o errores que ocurren frecuentemente acompañados de otras situaciones (Merceron & Yacef, 2010).

Análisis de redes sociales

Tiene como objetivo entender y medir la relación entre dos entidades de información relacionadas. La técnica de relaciones sociales en términos de teoría de redes consiste en nodos que representan actores individuales y conexiones o enlaces que identifican el tipo de relación entre dos nodos. Su aplicación puede emplearse para interpretar y analizar la estructura y relaciones en tareas colaborativas y la interacción con las herramientas de comunicación (Rabbany, Takaffoli, & Zaiane, 2011).

Proceso de minería

Extraer conocimiento a partir del análisis de los eventos registrado por un sistema de información para tener una representación visual clara del proceso. Consta de tres subcampos: chequeo de composición de la información, modelo de descubrimiento y modelo de extensión. Su utilidad puede reflejar el comportamiento de los estudiantes luego de analizar las trazas registradas de su actividad secuencial de sus actividades, mediante la representación gráfica de su rendimiento (Romero, Ventura, Pechenizkiy, & Baker, 2010).

Minería de texto

Pretende obtener información de calidad a partir del análisis de texto no estructurado, mediante métodos como la categorización, text clustering, extracción de entidades o conceptos, producción de taxonomía granular, análisis sentimental (opiniones, sentimientos, ...), resumen documental, y los modelos de entidad relación. Se usa para el análisis del contenido de foros de discusión, chats, páginas web, documentos (Tane, Schmitz, & Stumme, 2004).

Procesamiento para análisis humano

Representar información de manera inteligible mediante el uso de resumen, visualización e interfaces interactivas para destacar información útil para la toma de decisiones. Por una parte, es relativamente fácil obtener estadísticas descriptivas de la información educacional para obtener características globales de la información, resumir y realizar reportes del comportamiento del estudiante. Por otra parte, para ver, explorar, entender grandes cantidades de información. Ha sido utilizado para ayudar a los educadores a visualizar y analizar la información de actividades de los estudiantes y el uso que realizan de la plataforma (Mazza & Milani, 2004).

Descubrimiento mediante modelos

El uso de un modelo previamente validado sobre un fenómeno como componente de otro análisis como la predicción o minería de relaciones. Es un área en desarrollo para la minería de datos educacionales, y se sustenta en la identificación de relaciones entre el comportamiento de los estudiantes y sus características o variables contextuales, el análisis de preguntas de investigación de diferentes contextos, y la integración de modelos psicométricos (Baker & Yacef, 2009; Bienkowski, Feng, & Means., 2012).

Es apropiado para estimar la adquisición de habilidades por parte de los estudiantes, utiliza un modelo cognitivo que mapea una resolución de problemas para las habilidades requeridas, y registra las respuestas correctas e incorrectas de los estudiantes como evidencia de su conocimiento en una habilidad en concreto; registra el conocimiento del estudiante en el transcurso del tiempo y lo parametriza en variables.

En definitiva, existen diversas técnicas para aplicar minería de datos en la educación, sin embargo, será necesario el análisis de los diferentes escenarios, que pudiera derivar en nuevas técnicas, y usualmente, la combinación de varias de ellas conlleva una mejora de los resultados obtenidos.

Aplicaciones de Minería De Datos Educacionales

La minería de datos educacionales se refiere al proceso de descubrimiento de patrones, tendencias y conocimientos útiles a partir de grandes conjuntos de datos educativos. Estos datos pueden provenir de diversas fuentes, como sistemas de gestión de aprendizaje, registros académicos, encuestas y evaluaciones. A continuación, se describen algunas aplicaciones clave de la minería de datos educacionales:

En la Universidad de Purdue (EE.UU.) se desarrolló el sistema "Course Signals" para mejorar el rendimiento estudiantil mediante el uso de analítica de aprendizaje. Utiliza datos como calificaciones previas, participación en clases y resultados de pruebas para predecir el rendimiento futuro de los estudiantes. Cuando identifica a estudiantes en riesgo es capaz de enviar alertas a profesores y estudiantes con recomendaciones específicas para mejorar. Este enfoque proactivo ha demostrado ser eficaz en la mejora del éxito académico (Arnold, K. E., & Pistilli, M. D., 2012).

La UNED de España implementó un sistema de predicción utilizando minería de datos para estudiantes de cursos en línea. Se analizan los datos de participación, interacción con el contenido del curso y desempeño en actividades, y mediante un modelo predictivo se identifica

a los estudiantes en riesgo de no completar con éxito el curso. Esto permitió a la institución tomar medidas preventivas, como proporcionar apoyo adicional y recursos personalizados, para mejorar la retención y el rendimiento académico (Kotsiantis, S. B., Pierrakeas, C., & Pintelas, P., 2004).

La Universidad Carlos III de Madrid (España) empleó modelos lógicos en un sistema de detección automática de dificultades para analizar algunos datos como: la participación en clases, el tiempo dedicado a las tareas y los resultados de evaluaciones. De este modo se logró identificar patrones que indicaban posibles dificultades de los estudiantes. Este sistema permitió intervenciones tempranas y personalizadas para abordar las necesidades individuales de los estudiantes y mejorar su experiencia académica (Sánchez-Rada, J. F., Iglesias, C. A., & García-Serrano, A., 2011).

La Universidad de Córdoba (España), aplicó minería de datos para predecir el rendimiento final de los estudiantes basándose en su participación en foros en línea. Se analizaron datos de interacción en los foros, incluyendo la calidad de las contribuciones y la frecuencia de participación con el objetivo de identificar patrones que sugiriesen un mayor compromiso académico y, por lo tanto, un mejor rendimiento, y la relación entre ambas variables. Esta aplicación permitió a la institución desarrollar estrategias para fomentar la participación y mejorar la calidad del aprendizaje en línea (Romero, C., López, M. I., Luna, J. M., & Ventura, S., 2013).

CONCLUSIONES

La incorporación de la minería de datos a las prácticas de formación virtual en un futuro muy cercano es un imperativo. Es una disciplina relativamente joven, sus desarrollos en la educación han sido poco considerados y esto justifica que no hayan sido incluidos en las prácticas comunes de la inmensa mayoría de las comunidades docentes, por lo que la investigación científica y la innovación tienen un nicho amplio de desarrollo.

En el presente trabajo se han expuesto el proceso para la aplicación de minería de datos educacionales, técnicas y utilidades más significativas, con el propósito de esclarecer el camino a los docentes, pero en especial a quienes se encargan de administrar los procesos de formación, y autoridades a que se interesen en incursionar en este campo en favor de la educación virtual o en línea.

Entre las técnicas presentadas destacan la clasificación y predicción, clustering, detección de valores extremos, minería de relaciones, análisis de redes sociales, procesos de minería, minería de texto y descubrimiento mediante modelos.

Queda patente que existen diversas técnicas para aplicar minería de datos en la educación y que, a menudo, la combinación de varias de ellas conlleva una mejora de los resultados obtenidos en pro de intervenciones para el fortalecimiento de la educación virtual.

Además, el estudio de casos de implementaciones exitosas son sin duda el aliento a continuar profundizando en un campo de actualidad que puede revalorizar la educación y favorecer el proceso de enseñanza-aprendizaje en favor del estudiante, el principal beneficiario, ya que debe revocar en un mejor desarrollo de sus conocimientos.

REFERENCIAS BIBLIOGRÁFICAS

- Amershi, S., & Conati, C. (2009). Combining Unsupervised and Supervised Classification to Build User Models for Exploratory Learning Environments. En *Journal of Educational Data Mining*, Article (Vol. 2, Número 1).
- Anaya, A. R., & Boticario, J. G. (2009). A Data Mining Approach to Reveal Representative Collaboration Indicators in Open Collaboration Frameworks. *EDM 2009*.
- Arnold, K. (2010). Signals: Applying Academic Analytics | *EDUCAUSE*. *Educause Quarterly*, 33, 1-10.
- Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267-270).
- Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., & Woolf, B. (2007). Repairing Disengagement With Non-Invasive Interventions. *Artificial Intelligence in Education*, 158(January).
- Baker, R. S. J. d. (2010). Data mining for education. *International Encyclopedia of Education*, 7, 112-118. <https://doi.org/10.4018/978-1-59140-557-3>
- Baker, R. S. J. D., Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., & Beck, J. E. (2006). Adapting to when students game an intelligent tutoring system. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4053 LNCS, 392-401. https://doi.org/10.1007/11774303_39
- Baker, R. S. J. D., & Gowda, S. M. (s. f.). *Automatically Detecting a Student's Preparation for Future Learning: Help Use is Key*.

- Baker, R. S. J. D., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1(1).
- Baker, R. S. J., & Inventado, P. S. (2014). Chapter 4: Educational Data Mining and Learning Analytics. Springer, Chapter 4, 61-75.
- Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. *Learning Analytics: Foundations, Applications, and Future Directions*, 61-75.
- Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. Washington, DC: SRI International, 1-57.
- Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. *WIREs Data Mining and Knowledge Discovery*, 8. <https://doi.org/doi: 10.1002/widm.1230>
- Calvet Liñán, L., & Juan Pérez, Á. A. (2015). Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *RUSC. Universities and Knowledge Society Journal*, 12(3), 98. <https://doi.org/10.7238/rusc.v12i3.2515>
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253-278. <https://doi.org/10.1007/BF01099821>
- Daradoumis, T., Ardura, I. R., Faulin, J., Juan, A. A., Xhafa, F., & Lopez, F. J. M. (2010). Customer Relationship Management applied to higher education: developing an e-monitoring system to improve relationships in electronic learning environments. *International Journal of Services Technology and Management*, 14(1), 103. <https://doi.org/10.1504/IJSTM.2010.032887>
- Daradoumis, T., Juan, A. A., Lera-López, F., & Faulin, J. (2010). Using Collaboration Strategies to Support the Monitoring of Online Collaborative Learning Activity (pp. 271-277). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13166-0_39
- García, E., Bernardo, M., Gutiérrez, A. B., & Rodríguez-Muñiz, L. J. (2016). Persistence in university studies: The importance of a good start. En *Aula Abierta* (Vol. 44, Número 1, pp. 1-6). Elsevier Doyma. <https://doi.org/10.1016/j.aula.2015.04.001>
- Galetsis, P., Katsaliaki, K., & Kumar, S. (2020). Big data analytics in health sector: Theoretical framework, techniques and prospects. *International Journal of Information Management*, 50(March 2019), 206-216. <https://doi.org/10.1016/j.ijinfomgt.2019.05.003>

- George, D. M., George, A. E., & George, K. L. (2020). Big Data Just Got Bigger: Implications of Real-World Evidence and Patient-Entered Data on Health Care and Health Care Policy. *American Journal of Lifestyle Medicine*, 14(1), 40-42. <https://doi.org/10.1177/1559827619882605>
- Huebner, R. (2013). A survey of educational data-mining research. *Research in Higher Education Journal*, 19.
- IBM. (2017). Infographic: The Four V's of Big Data. IBM Big Data & Analytics Hub. <https://www.ibmbigdatahub.com/infographic/four-vs-big-data>
- Jaramillo, A., & Paz-Arias, H. (2015). Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje. *Revista Tecnológica ESPOL – RTE*, 28(1), 64-90.
- Jiménez, Á., & Álvarez, H. (2010). Minería de Datos en la Educación. *Revista de Inteligencia en Redes de Comunicación*, 2-29.
- Juan, A. A., Daradoumis, T., Faulin, J., & Xhafa, F. (2009). A data analysis model based on control charts to monitor online learning processes. *International Journal of Business Intelligence and Data Mining*, 4(2), 159. <https://doi.org/10.1504/ijbidm.2009.026906>
- Khan, M. A. U. D., Uddin, M. F., & Gupta, N. (2014). Seven V's of Big Data understanding Big Data to extract value. *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education - «Engineering Education: Industry Involvement and Interdisciplinary Trends»*, ASEE Zone 1 2014. <https://doi.org/10.1109/ASEEZone1.2014.6820689>
- Kop, R. (2010). The design and development of a Personal Learning Environment. *Archives des publications du CNRC*.
- Kotsiantis, S. B., Pierrakeas, C., & Pintelas, P. (2004). Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5), 411-426.
- Maheshwari, A. K. (2015). *Business Intelligence and Data Mining*. Business Expert Press, LLC.
- Mazza, R., & Milani, C. (2004). GISMO: A Graphical Interactive Student Monitoring Tool for Course Management Systems. *Computer Science*.
- Menacho Chiok, C. H. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26. <https://doi.org/10.21704/ac.v78i1.811>
-

- Merceron, A., & Yacef, K. (2010). Measuring Correlation of Strong Symmetric Association Rules in Educational Data (pp. 245-256). <https://doi.org/10.1201/b10274-20>
- Mohamad, S. K., & Tasir, Z. (2013). Educational Data Mining: A Review. *Procedia - Social and Behavioral Sciences*, 97, 320-324. <https://doi.org/10.1016/j.sbspro.2013.10.240>
- Morales, J., Trujillo, V., & Santana, H. (2019). Minería de datos en educación: una revisión literaria. *Espamciencia*, 10(1), 14-20.
- Papamitsiou, Z., & Economides, A. A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society*, 17(4), 49-64.
- Pérez-Palacios, T., Caballero, D., Caro, A., Rodríguez, P. G., & Antequera, T. (2014). Applying data mining and Computer Vision Techniques to MRI to estimate quality traits in Iberian hams. *Journal of Food Engineering*, 131, 82-88. <https://doi.org/10.1016/j.jfoodeng.2014.01.015>
- Rabbany, R., Takaffoli, M., & Zaïane, O. R. (2011). Analyzing Participation of Students in Online Courses Using Social Network Analysis Techniques. *International Conference on Educational Data Mining*.
- Ray, S., & Saeed, M. (2018). Applications of educational data mining and learning analytics tools in handling big data in higher education. *Applications of Big Data Analytics: Trends, Issues, and Challenges*, 135-160. https://doi.org/10.1007/978-3-319-76472-6_7
- Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2010). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, 21(1), 135-146. <https://doi.org/10.1002/cae.20456>
- Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, 68, 458-472.
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27. <https://doi.org/10.1002/widm.1075>
- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, 51(1), 368-384. <https://doi.org/10.1016/j.compedu.2007.05.016>
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. J. d. (2010). *Handbook of Educational Data Mining*. CRC Press.
-

- Shaun, R., Baker, J. De, & Eds, J. E. B. (2008). Educational Data Mining 2008 The 1 st International Conference on Educational Data Mining. Network, January, 187.
- Sánchez-Rada, J. F., Iglesias, C. A., & García-Serrano, A. (2011). Towards an automatic early detection of student's difficulty: Logical models rule. *Computers & Education*, 56(3), 573-582.
- Siemens, G., & Baker, R. S. J. D. (2012a). Learning analytics and educational data mining. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 252. <https://doi.org/10.1145/2330601.2330661>
- Siemens, G., & Baker, R. S. J. D. (2012b). Learning analytics and educational data mining: Towards communication and collaboration. *ACM International Conference Proceeding Series*, December, 252-254. <https://doi.org/10.1145/2330601.2330661>
- Tane, J., Schmitz, C., & Stumme, G. (2004). Semantic resource management for the web: An elearning application. *Proceedings of the 13th International World Wide Web Conference on Alternate Track, Papers and Posters, WWW Alt. 2004*, 1-10. <https://doi.org/10.1145/1013367.1013369>
- Uddin, M. F., & Lee, J. (2016). We Are What We Generate - Understanding Ourselves Through Our Data. *Procedia Computer Science*, 95, 335-344. <https://doi.org/10.1016/j.procs.2016.09.343>
- Ueno, M. (2004). Online Outlier Detection System for Learning Time Data in E-Learning and Its Evaluation. *Triangulation of Bayesian networks View project Learning Bayesian networks View project online outlier detection system for learning time data in e-learning and it's evaluation.*