

CAPÍTULO 6.

LA MINERÍA DE DATOS EN LA EDUCACIÓN SUPERIOR

Carmen Selene Pineda Morán¹

Ronald Madera Terán²

Ronald Rusbell Madera Arnao³

Yeni Madera Arnao⁴

Jesús Chuchón Vilca⁵

Sebastián Huangal Scheineder⁶

Jhon Dany Castañeda Requejo⁷

William Zacarías Ojeda Pereda⁸

¹: Ingeniero de sistemas, Magister en Gestión Pública, Doctora en Educación.

 <https://orcid.org/0000-0001-9198-6765>

²: Doctor en Educación, Doctor en Ciencias e Ingeniería.

 <https://orcid.org/0000-0002-7818-4511>

³: Master universitario en Diseño y Gestión Ambiental de Edificios.

 <https://orcid.org/0000-0003-2449-540X>

⁴: Bachiller en Ingeniería Civil.

 <https://orcid.org/0009-0002-8911-2916>

⁵:  <https://orcid.org/0000-0001-7319-2259>

⁶: Ingeniero Químico, Doctor en Ciencias Ambientales y Docente Universitario.

 <https://orcid.org/0000-0002-4864-1857>

⁷: Ingeniero Agrónomo, Magister en Ciencias con mención en Proyectos de Inversión. Docente Universitario.

 <https://orcid.org/0000-0001-7261-1634>

⁸: Contador público, Maestro en Ciencias Económicas, Doctor en Contabilidad. Docente Universitario.

 <https://orcid.org/0000-0001-5224-6115>

Ingreso: 11/03/24 • Aprobado: 30/04/24

RESUMEN

El propósito de este estudio está centrado en hacer converger desde la revisión de distintos casos, desarrollados por los investigadores más reputados vinculados a los estudios sobre el uso de la minería de datos en la educación superior. Para lograr la meta de esta investigación se hace uso de una profundización conceptual y de casuísticas, a través de distinto material bibliográfico, que se toma de los últimos 10 años en investigaciones sobre ambas categorías, la relevancia está en que se hace una revisión de múltiples posturas, en ciertos casos encontrando concordancias y en otras discrepancias. Todo con la finalidad de poder llegar a una reflexión basada en un debido estudio documentado.

Palabras clave: Ansiedad, covid-19, trabajadores informales, efectos, pandemia.

INTRODUCCIÓN

En los últimos años las actividades y el desarrollo de nuevas tecnologías han creado la necesidad de almacenar grandes volúmenes de información, donde todo ese flujo de datos recopilados ha permitido satisfacer las necesidades diarias de las organizaciones, pero ha presentado un problema inherente en las capacidades humanas para analizar y transformar la información en conocimiento útil y relevante que apoye a la toma de decisiones (Bienkowski et al, 2012). Por lo general, tales datos guardan información valiosa, pero por tratarse de una recopilación masiva de información se debe recurrir al empleo de técnicas y herramientas de Minería de Datos (Por sus siglas en inglés DM – Data Mining), que posibiliten la extracción de conocimiento en forma de reglas y patrones a partir de dichos datos, es decir, es un proceso de análisis de datos que consiste en buscar o encontrar tendencias o variaciones de comportamiento en los datos denominados patrones. La minería de datos en el contexto educativo desarrolla métodos y aplica estadísticas, aprendizaje automático (machine learning) y técnicas de minería de datos para analizar los datos recopilados durante la enseñanza y el aprendizaje (Ramaswami et al, 2019). Con esto pone a prueba las teorías del aprendizaje e informa la práctica educativa

Actualmente, la data mining se ha extendido al sector educativo en todos sus niveles, varias áreas de las ciencias han aplicado DM con el propósito de estudiar el comportamiento de los alumnos bajo diferentes sistemas de enseñanza – aprendizaje. Entre los diversos datos que se pueden generar se tienen el tiempo en resolver un problema, oportunidades permitidas a los estudiantes, si la respuesta fue acertada o equivocada (Ballesteros et al, 2023). La investigación sobre minería de datos educativos y análisis de aprendizaje está comenzando a responder preguntas cada vez más complejas sobre lo que sabe un estudiante o si está comprometido. La literatura científica actual evidencia nuevas técnicas para la construcción de modelos y nuevos tipos de datos de sistemas de aprendizaje prometedores para predecir los resultados

de los estudiantes. La minería de datos educativos está surgiendo como un área de investigación con un conjunto de métodos y enfoques de investigación computacionales y psicológicos para comprender cómo aprenden los estudiantes. Los nuevos métodos y herramientas de aprendizaje interactivo apoyados por computadora (sistemas de tutoría inteligentes, simulaciones, juegos) han abierto oportunidades para recopilar y analizar datos de los estudiantes, descubrir patrones y tendencias en esos datos y hacer nuevos descubrimientos y probar hipótesis sobre cómo aprenden los estudiantes (Bienkowski et al, 2012). Los datos recopilados de los sistemas de aprendizaje en línea se pueden agregar sobre un gran número de estudiantes y pueden contener muchas variables que los algoritmos de minería de datos pueden explorar para la construcción de modelos.

En otras palabras, el propósito con DM es procesar automáticamente grandes cantidades de datos para encontrar conocimiento útil para un usuario y satisfacer sus metas respecto al conocimiento a detalle de un sistema de información sobre un contexto educativo. Por lo tanto, el objetivo de la presente investigación fue estudiar la aplicación de la ciencia de datos como herramienta para mejorar los procesos educativos universitarios.

MARCO CONCEPTUAL

Las bases conceptuales tienen su centro en dos categorías primordiales para profundizar en la problemática de la investigación. Por un lado, la minería de datos o data mining y, por el otro, los procesos educativos de enseñanza – aprendizaje, esto debido a que el análisis de grandes cantidades de datos para resolver problemas educativos ha propiciado el desarrollo de la Minería de Datos Educativos (MDE), siendo esta una disciplina dedicada a desarrollar métodos para explorar los datos provenientes de ambientes relacionados a

la educación, para tratar de entender mejor a los estudiantes, profesores y demás involucrados en sus entornos educativos con el fin de mejorar los procesos y el aprendizaje (Escobar, 2017, p. 204).

La minería de datos educativos permite el desarrollo de métodos para el descubrimiento de tipos particulares de datos que provienen de plataformas educativas, y el uso de esos métodos para comprender mejor a los estudiantes y sus entornos de aprendizaje (Baker & Yacef, 2009).

Descubrimiento de Conocimiento en Bases de Datos (KDD – Knowledge Data Discovery)

KDD es un proceso no-trivial de identificación de patrones válidos, novedosos y potencialmente útiles sobre un conjunto de datos, esto es, el objetivo es encontrar conocimiento útil, valido, relevante y nuevo sobre una determinada actividad. En este contexto, los datos hacen referencia a un conjunto de hechos o ejemplos en una base de datos y los patrones son resultados o expresiones en algún lenguaje que puedan describir de manera compacta los datos asimismo el termino no-trivial se comprende qué alguna búsqueda o inferencia es llevada a cabo, esto es, implica la búsqueda de modelos, estructuras, patrones o parámetros (Ballesteros et al, 2023, pp. 663, Durán y Costaguta, 2006)).

De acuerdo con Clancy y Gelinas (2016), KDD es un área interdisciplinaria que se centra en metodologías para extraer conocimientos útiles a partir de datos. La extracción de conocimientos a partir de datos se basa en investigaciones en estadística, bases de datos, reconocimiento de patrones, aprendizaje automático, visualización de datos, optimización e informática de alto rendimiento para ofrecer soluciones avanzadas de inteligencia empresarial y descubrimiento web.

Minería de datos o Data mining

Este es un proceso que consiste en realizar una tarea de análisis de datos y una aplicación de algoritmos de análisis de datos, así como las limitaciones informáticas, el producto de un conjunto de documentos de datos. En líneas generales, consiste en explorar grandes cantidades de datos a partir de

procedimientos consistentes, como las reglas de asociación o las secuencias temporales, para la relación sistemática entre variables, detectando nuevos subconjuntos de datos (Alves y Lima, 2018).

Para Ballesteros et al (2023) esta etapa es la central del proceso KDD, con la finalidad de encontrar o descubrir los patrones de interés para el usuario final, estos patrones pueden ser grafos, reglas de asociación, clasificaciones, una red neuronal, clustering, entre otros (pp.664).

Métodos y técnicas de la minería de datos educativa

Para el desarrollo de softwares en la minería de datos educativa se necesita de clasificadores, agrupamiento y reglas de asociación, con lo cual se identificaría toda aquella información oculta para los diferentes actores dentro de las instituciones educativas.

Clasificadores para Minería de Datos Educativa

Con la clasificación se pretende definir un objeto en base de sus otras características, para la creación de modelos predictivos que describan clases de datos importantes o predicciones de tendencias futuras, tomando en cuenta que la DM busca obtener reglas para la división de los datos en una clase o categoría para la mejor comprensión de los datos (Ballesteros et al, 2023).

Clustering para la Minería de Datos Educativa

Los métodos de agrupamiento (Clustering) estudian los registros con etiquetas de clase los objetos de datos, de tal forma que los métodos de agrupamiento no están interesados en modelar un conjunto de relaciones y un conjunto de respuestas que pertenezcan a cada valor asociado. El clustering es utilizado para la generalización de etiquetas, esto es, la búsqueda y construcción de grupos formados de tal forma que los objetos (Patrones) dentro de un conjunto (un grupo) tienen una alta similitud en comparación entre sí denominada función de medida de asociación (similitud matemática en espacios métricos definida por medio de una norma de distancia), donde cada grupo formado puede ser interpretado como una clase de objetos de los cuales se pueden derivar reglas

(Ballesteros et al, 2023).

Proceso educativo: enseñanza –aprendizaje

El proceso educativo es definido por Chipana (2022) como un conjunto de actividades que configuran el acto educativo (enseñanza – aprendizaje), innovaciones, procedimientos, planificaciones y prácticas realizadas por quienes buscan el perfeccionamiento del hombre; esto da a lugar a considerar este proceso como el espacio y el conjunto de procedimientos que permiten la transmisión de conocimientos específicos o generales. Es importante resaltar que el proceso de enseñanza - aprendizaje los principales protagonistas son: el estudiante y el docente que cumple la función de ser el facilitador del proceso de aprendizaje.

Factores que inciden en el proceso de la enseñanza – aprendizaje

- El educando, quien lleva a cabo en sí el proceso de aprendizaje, ya sea individualmente o en grupo. El educando trae consigo un conjunto de factores que son apropiados mencionarlos: biológicos, psicológicos y sociales.
- El docente, quien es el actor que más incide en la comunidad educativa y tiene rasgos personales y profesionales. El docente debe considerar, durante el proceso de enseñanza, su actitud frente al estudiante y su postura frente al proceso como tal.
- Las condiciones ambientales, las cuales pueden influir de manera positiva o negativa en el educando, ya sea promoviendo, facilitando o impidiendo ciertas conductas, de manera que se va favorecido el aprendizaje. Por ejemplo, el ambiente físico de clases en la universidad o en la vivienda, los materiales y recursos didácticos, la organización, la relación con el educador, entre otros (Chipana, 2022).

METODOLOGÍA

La presente investigación fue de tipo documental, puesto que su centro de acción está relacionado al análisis bibliográfico de estudios relacionados con la minería de datos y los procesos de enseñanza – aprendizaje. Este capítulo tiene un enfoque teórico exploratorio, puesto que se parte desde una profundización inicial en lo conceptual que luego concretiza en las aplicaciones de minería de datos en la educación superior.

RESULTADOS

La investigación en torno a la minería de datos en el sector educativo y análisis de aprendizaje ha dado respuestas a preguntas que eran cada vez más complejas los últimos años, sobre lo que sabe un estudiante. En la práctica son muchas las aplicaciones de esta herramienta, de acuerdo con lo investigado en la literatura, sobre todo en el aprendizaje en línea.

Estas áreas de aplicación son:

1. Modelado del conocimiento del usuario, del comportamiento del usuario y de la experiencia del usuario

Por lo general, los datos recopilados de los estudiantes tienen un comportamiento en particular, ante esto, los investigadores y desarrolladores crean y ajustan modelos que representan tales datos, especialmente habilidades y conocimientos. La idea es personalizar y adaptar los comportamientos del

sistema a las necesidades específicas de los usuarios, de modo que los sistemas "digan" lo "correcto" en el momento "correcto" y de la manera "correcta" (Gerhard, 2001). En otras palabras, inferir y predecir lo que sabe un usuario.

El modelo de conocimiento del usuario ha sido el más adoptado para diseñar sistemas de recomendación, sistemas expertos y sistemas de tutoría inteligente. Con este último, por ejemplo, se puede decidir qué problemas plantear a los alumnos durante la evaluación.

El método más popular para desarrollar este tipo de modelos es el de Cobertt y Anderson, el cual está basado en redes bayesianas (Corbet y Anderson, 1994). Otro método empleado, pero que ha surgido más recientemente, es el de Baker, este aplica el enfoque de aprendizaje automático o machine learning. Con ello se podría determinar, por ejemplo, si un alumno ha adivinado o si ha cometido un error, con esto se logra una precisión de hasta el 48% en las predicciones.

El modelado del conocimiento de los estudiantes es un componente común del software de aprendizaje comercial. Como ejemplo del uso de este modelo, el software de aprendizaje puede recopilar datos como, por ejemplo, cuántos minutos se dedican a una unidad, cuántas sugerencias se utilizaron y errores comunes. Estos modelos ayudan a los maestros a utilizar diferentes estrategias de instrucción para diversos grupos de estudiantes.

En relación con el modelado del comportamiento del usuario, se basa en los mismos tipos de datos de aprendizaje utilizados para predecir el conocimiento del usuario más otras medidas, como cuánto tiempo ha pasado un estudiante en línea (o en el sistema), si un estudiante ha completado un curso, cambios documentados en el aula o la escuela, contexto, asistencia, tardanzas y, a veces, el nivel de conocimiento de un estudiante.

2. Elaboración de perfiles de usuario

Un perfil de usuario es una colección de datos personales que describen las características esenciales de un usuario (Feijoo, et al, 2020). En el contexto educativo, esto consiste en la construcción y aplicación de perfiles de estudiantes utilizando minería de datos y algoritmos de aprendizaje automático. Teniendo

estos perfiles se pueden proporcionar entornos de aprendizaje adaptados y personalizados para individuos o grupos de estudiantes para maximizar la eficacia y eficiencia del aprendizaje, ajustados a las preferencias, intereses, antecedentes e incluso objetivos de aprendizaje de los estudiantes. Con los perfiles se puede predecir el comportamiento de los estudiantes.

Pero ¿qué información se utiliza para crear los perfiles? La respuesta es datos personales de aprendizaje, datos demográficos de los estudiantes, o en ambos (Bienkowski et al, 2012). Además de poder predecir el comportamiento, se pueden clasificar a los nuevos estudiantes, con el objetivo de brindar apoyo cuando se detecten comportamientos perjudiciales para el aprendizaje, o para aprender formas de apoyar el comportamiento comprometido.

3. Modelado del dominio

Un dominio puede ser una materia, curso o área temática. Un modelado del dominio consiste en representar los conceptos clave inherentes al mismo e identificar las relaciones entre todos los conceptos de estudio clave. Con un modelado se puede conocer las necesidades de aprendizaje de un grupo de estudiantes, también se puede obtener más detalle como, por ejemplo, las áreas a cubrir de esas necesidades ya identificadas.

En esta fase se recomienda utilizar una taxonomía del dominio, asociaciones entre habilidades (como requisitos previos), respuestas de los usuarios (incluida la corrección) y acciones a lo largo del tiempo sobre recursos de aprendizaje individuales (como un concepto de unidad) (Bienkowski et al, 2012).

4. Análisis de tendencias

El análisis de tendencias consiste en recopilar información e intentar detectar un patrón secuencial o tendencia de la información a lo largo del tiempo. En educación, esto ayuda a responder preguntas como qué cambios se han producido en el aprendizaje de los estudiantes a lo largo del tiempo y cómo ha cambiado el aprendizaje.

En la minería de datos educativos, el análisis de tendencias a menudo se refiere a técnicas para extraer un patrón subyacente, que podría estar parcial o casi completamente oculto por datos que no contribuyen al patrón. Aunque los datos reales necesarios para el análisis de tendencias varían según la información de interés, normalmente se requieren datos longitudinales de al menos tres puntos en el tiempo.

Desafíos y consideraciones de implementación

Las tecnologías emergentes aplicadas a la educación superior hoy en días ofrecen a investigadores y desarrolladores oportunidades para crear entornos de aprendizaje personalizados, pues para ello se basan en grandes conjuntos de datos que pueden analizarse para respaldar la mejora continua. Sin embargo, estos beneficios dependen de la gestión de todos los datos que se pueden capturar en tiempo real entre muchos estudiantes. Un desafío para la implementación exitosa de técnicas de análisis de aprendizaje y minería de datos educativos es tener suficientes recursos técnicos para usar big data e incurrir en los gastos asociados con los servicios de software y el almacenamiento. Otra debilidad son los requisitos de memoria de la computadora necesarios para procesar algoritmos avanzados.

Tal vez, el reto más importante de la minería y el análisis de datos es saber cómo discriminar entre estos, es decir, centrar sobre las preguntas que deben responderse, asegurarse de que los datos coincidan con las preguntas y aplicar las técnicas analíticas correctas, esto permitirá tener perfiles y modelos más ajustados a la realidad y con menor rango de error.

Lo que se busca es poder concluir o tener indicadores que permitan tomar decisiones acertadas sobre la educación, en función de los datos recopilados que a su vez derivan de la interacción del estudiante con un sistema de aprendizaje. Ir más allá de un conjunto de datos y combinar múltiples fuentes de datos para proporcionar una visión integrada del progreso de un estudiante no es una labor fácil. Podrían tomarse datos erróneos o descartar datos valiosos con facilidad, lo cual dificulta la evaluación.

DISCUSIÓN Y CONCLUSIONES

Trabajar con datos mediante análisis y minería de datos es el auge en el sector comercial actualmente ya que, se han estado aplicando las diversas herramientas y técnicas que antes solo se utilizaban en investigación de ciertas disciplinas, y se les está enfocando a prestar servicios a los usuarios finales. Las instituciones de educación superior, hoy en día, están aplicando minería de datos para mejorar los servicios que brindan, así como tener objetivos visibles y mensurables, detectar áreas de mejora, establecer políticas y medir resultados.

Ahora, con los avances en los sistemas de aprendizaje adaptativo, existen posibilidades de aprovechar el poder de los circuitos de retroalimentación a nivel de profesores y estudiantes individuales. Medir y hacer visibles las actividades de aprendizaje y evaluación de los estudiantes abre la posibilidad de que los estudiantes desarrollen habilidades para monitorear su propio aprendizaje y ver directamente cómo su esfuerzo mejora su éxito. Además, los sistemas de aprendizaje adaptativo permiten a los educadores ver rápidamente la efectividad de sus adaptaciones e intervenciones, generando retroalimentación para la mejora continua. Resultados similares obtuvieron Rosado y Verjel (2016), quienes en su estudio exponen que es posible predecir la cantidad de interacción que tienen los estudiantes con las plataformas y que esto se logra con buena óptima agrupación de datos considerando las características y dimensionalidad de los mismos. Los autores también destacan que unas de las herramientas utilizadas son cross-platform, platform-specific, ad hoc tolos, learning analytic tools, learning analytic frameworks.

REFERENCIAS

- Alves, F. & Lima, D. (2018). Uso de la clasificación para el análisis y la minería de datos en la herramienta de enseñanza-aprendizaje Google Classroom. *Nuevas Ideas en Informática Educativa*, (14), pp. 589-594. <https://www.tise.cl/Volumen14/TISE2018/589.pdf>
- Baker, R., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1 (1), pp. 3–17. <https://doi.org/10.5281/zenodo.3554657>
- Bienkowski, M, Feng, M. & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. U.S. Department of Education. <https://files.eric.ed.gov/fulltext/ED611199.pdf>
- Chipana, F. (2022). Dinámica del proceso enseñanza – aprendizaje en educación superior. *Ciencia Latina*, 6(1). https://doi.org/10.37811/cl_rcm.v6i1.1827
- Clancy, T. & Gelinas, L. (2016). Knowledge Discovery and Data Mining: Implications for Nurse Leaders. *Journal of Nursing Administration*, 46(9), pp. 422-424. https://www.nursingcenter.com/journalarticle?Article_ID=3638930&Journal_ID=54024&Issue_ID=3638884
- Corbett, A., & Anderson, J. (1994). Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction*, 4(4), pp. 253–278. 10.1007/BFO1099821
- Durán, E., & Costaguta, R. (2007). Minería de datos para descubrir estilos de aprendizaje. *Revista Iberoamericana De Educación*, 42(2), 1–10. <https://doi.org/10.35362/rie4222430>
-

- Escobar, H., Alcívar, M., Márquez, C. & Escobar, C. (2017). Implementación de minería de datos en la gestión académica de las instituciones de educación superior. *Didasc@lia: Didáctica y Educación*, 8(3), pp. 203-212. <https://dialnet.unirioja.es/servlet/articulo?codigo=6662630>
- Feijoo, B. & Sádaba, C. (2020). ¿Nivel experto o ingenuo? Detección y confianza de los niños en la publicidad que reciben a través de sus dispositivos móviles. Perfiles de usuario. *ZER, Revista de Estudios de Comunicación*, 25(48), 231-248. <https://reunir.unir.net/handle/123456789/10119>
- Gerhard, F. (2001). User Modeling in Human-Computer Interaction. *User Modeling and User Adapted Interaction*, (11), pp. 65–86. <https://link.springer.com/article/10.1023/A:1011145532042>
- Morales, J., Trujillo, V., Cevallos, S., Cedeño, S. & Monserrate, H. (2019). Minería de datos en educación: Una revisión literaria. *Revista ESPAMCIENCIA*, 10(1), pp. 14-20. https://revistasespam.espam.edu.ec/index.php/Revista_ESPAMCIENCIA/article/view/180/174
- Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J. and Garcia, P. (2019). Using educational data mining techniques to increase the prediction accuracy of student academic performance. *Information and Learning Sciences*, 120(7/8), pp. 451-467. <https://doi.org/10.1108/ILS-03-2019-0017>
- Rosado Gómez, A., & Verjel Ibáñez, A. (2017). Aplicación de la minería de datos en la educación en línea. *Revista colombiana de tecnologías de avanzada (RCTA)*, 1(29), 92–98. <https://doi.org/10.24054/rcta.v1i29.194>