Análisis de la Minería de Datos en el ámbito de la Educación

Fecha de recepción: 03/08/2019 Fecha de aceptación: 13/10/2019

Viviana Elizabeth Jiménez Chaves*
Miguel García Torres**
José Luis Vázquez Noguera***

Resumen

El trabajo presenta un análisis sobre la Minería de datos en el ámbito educativo. En la revisión bibliográfica se analizaron los artículos publicados en las bases de datos e índices: Springer Link, Scielo, Ebsco, Redalyc, Google Académico, Dialnet, Researchgate, Scopus, Latindex. Concentrados en textos que se refieren a América Latina como región o a un conjunto de países, en lugar de a países particulares de la región. Asimismo, en textos exclusivamente basados en investigaciones en educación. Como resultado de la investigación se cuenta con un material de consulta sobre el estado del arte del tema Minería de Datos en Educación o *Educational Data Mining* en sus siglas en inglés. En conclusión, la Minería de datos en el ámbito educativo constituye un nuevo recurso para llegar a la detección y previsión de los fenómenos en el ámbito educativo.

Palabras clave: Educación, Minería de datos, Educational Data Mining, Data Mining.

Abstract

This paper presents an analysis Data mining in the educational field. In the literature review the articles published in the databases and indexes were analyzed: Springer Link, Scielo, Ebsco, Redalyc, Google Académico, Dialnet, Researchgate, Scopus, Latindex. We focus on texts that refer to Latin America as a region or to a set of countries,

* Universidad Americana. Dirección de Investigación. Asunción, Paraguay. Email: viviana.jimenez@americana.edu.py
** Universidad Pablo de Olavide off Competer Science. Sevilla, España. Email: mgarciat@upo.es
*** Universidad Nacional de Asunción, Facultad Politécnica. San Lorenzo, Paraguay. Email: jlvazquez@pol.una.py

rather than to particular countries in the region. We have also focused on texts exclusively based on research in education. And as a result have a reference material on the state of the art of the topic Data Mining in Education or Educational Data Mining in its acronym in English. In conclusion, data mining in the educational field is a new resource to reach the detection and forecast of phenomena in the educational field.

Keywords: Education, Data mining, Educational Data Mining. Data Mining.

Introducción

El *Data Mining* o minería de datos tiene sus orígenes en los años 60 y 70 con varios autores como Rakesh Agrawal, GioWiederhold, Robert Blum y Gregory Piatetsky-Shapiro y otros, empezaron a consolidar los términos de Minería de Datos y *Knowledge Discovery in Databases*-KDD. (G.; Smith P.; Ramasasmy U. Fayyad, U.M.; Piatetskiy-Shapiro,2006), quienes fueron los pioneros en esta técnica, marcando esa época y ubicando a los datos en un sitio de preponderancia en el ámbito de la investigación en el área de la informática.

Una definición de *Data Mining* de acuerdo a Usama Fayyad, 1996; "Es un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos". El empleo de técnicas y herramientas de Minería de Datos facilitan la extracción de conocimiento en forma de reglas y patrones a partir de 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. (Ballesteros, 2014). La evolución de sus herramientas en el transcurso del tiempo puede dividirse en cuatro etapas principales presentadas en la siguiente Tabla N°1.

Tabla 1

Etapas

Etapas	Año
Colección de Datos	1960
Acceso de Datos	1980
Almacén de Datos y Apoyo a las Decisiones	1990
Minería de Datos Inteligentes	Finales de 1990

Fuente: Elaboración propia.

En el Paraguay la Minería de datos se inició aproximadamente en el año 2006 con el Taller Seminario *Current Trends* In Data *Science* dictado por el Dr. *Randy Paffenroth,* profesor de Ciencias Matemáticas en el *Worcester Polytechnic Institute* (WPI), siendo docente permanente del programa de *Data Science* (Ciencia de los Datos) de dicha Universidad. El curso se desarrolló en la Facultad Politécnica de la Universidad Nacional de

Asunción, entidad que se ha constituido en una de las propulsoras de la mencionada técnica y se ha definido con una línea de investigación en dicha casa de estudios.

Existen numerosas y variadas áreas de aplicación para la minería de datos, tales como negocios, *e-commerce y marketing* (Berry y Linoff, 2000; Koren et al., 2009), bioinformática (Bergeron, 2002), genética (Friedman et al., 2000), medicina (McCormick et al., 2012), astronomía (Burl et al., 1998), ingeniería (Grossman et al., 2001).

En el año 2008, emerge una nueva aplicación denominada en sus siglas en inglés como *Data Mining* (EDM) o Minería de Datos Educacionales, una disciplina dedicada a desarrollar métodos para explorar los datos provenientes de ambientes relacionados a la educación. Estas se han constituido en grandes herramientas para tratar de entender mejor a los estudiantes, profesores y actores en sus entornos educacionales. Al tiempo que los procesos educativos, mediante la identificación de soluciones de problemas en el aprendizaje en varios ámbitos. Por ejemplo: en el rendimiento académico en un área determinada, en el rendimiento en la educación virtual, en las dificultades en el aprendizaje de los alumnos en cualquier nivel educativo y últimamente también incursionando en la gestión educativa de la calidad.

De esta manera se define a la EDM partiendo de las principales teorías de la Minería de Datos y complementada con las definiciones sobre el impacto en los sistemas aprendizaje-enseñanza, se puede proponer lo siguiente: "La Minería de Datos Educativa-EDM es el empleo de las herramientas tecnológicas, algoritmos y las estrategias de análisis de información utilizadas por la Minería de Datos-DM, pero dentro de un contexto educativo para la búsqueda, análisis y la extracción de patrones de conocimiento, dónde se resuelvan problemas que mejoren el proceso enseñanza–aprendizaje a partir de modelos predictivos de forma cualitativa y cuantitativa".(Ballesteros, 2012).

Los autores (Romero y Ventura, 2010, Caballero A. et al., 2013), definieron unas categorías para la aplicación de EDM, como son Análisis y visualización de datos, soporte a la instrucción, comportamiento de los estudiantes, predicción del rendimiento académico. (Rosado Gómez, 2016).

En la educación, a través de la toma de decisiones, permitirá a las instituciones educativas incrementar el nivel de calidad que se brinda a través de nuevos conocimientos que pueden ser descubiertos en los datos que se encuentran en las bases de datos y en los resultados de actividades académicas como evaluaciones que podrán ser extraídas utilizando técnicas de minería de datos. (Chalaris et al., 2014).

Según Liñán y Pérez (2015), la popularidad por parte de los investigadores en los últimos años a la aplicación de minería de datos, en el sector de la educación, se debe a los siguientes factores:

Existe un interés en mejorar el proceso de toma de decisiones utilizando un enfoque basado en datos generados, como ocurre en inteligencia de negocios y analítica de datos.

En la actualidad, se conocen técnicas y métodos basados en estadística, sistemas expertos y minería de datos para buscar patrones en los datos y construir modelos de predicción o reglas que pueden ser adaptados fácilmente a los datos generados en este sector.

La generación de los datos por utilizar es relativamente fácil o ya existen, y su almacenamiento y procesamiento es posible con la tecnología de computación actual.

Dada la fuerte competencia y la crisis financiera, las universidades están bajo presión de reducir costos e incrementar las ganancias mediante la explotación de la creciente demanda en educación de calidad, reducir el abandono académico y mejorar la oferta educativa.

De las diferentes aplicaciones o tareas que existen en el ámbito de la educación que se pueden resolver mediante Minería de datos (Tabla 2), uno de los más antiguos y el más popular es la predicción de rendimiento académico de los estudiantes (Romero et al, 2013).

Tabla 2

Aplicación de Minería de Datos

Tarea	Objetivo/Descripción	Aplicaciones claves
Predicción	Inferir la variable objetivo desde la combinación de otras variables. Se aplican técnicas de clasificación, regresión y estimación de densidad.	Predicción de rendimiento académico y detectar comportamiento de estudiantes.
Agrupamiento	Identificar grupos con características similares.	Crear grupos de estudiantes con características similares en su patrón de interacción y aprendizaje
Minería de relaciones	Estudio de relaciones entre variables y descubrimiento de reglas.	Identificar relaciones en patrones de aprendizaje y diagnosticar dificultades en aprendizaje.
Destilación de datos	Representación de datos de manera más inteligible mediante resumen, visualización e interfaces interactivos.	Brindar herramientas de apoyo a los instructores para visualizar y analizar actividades relacionadas con sus estudiantes.
Descubrimiento con modelos.	Emplear modelos validados previamente a nuevos datos.	Identificación de relación entre el comportamiento y las características de los estudiantes o de variables contextuales
Detección de anomalías.	Detección de individuos con diferencias significativas.	Detección de estudiantes con dificultades o con aprendizaje irregular
Minería de textos	Extracción de información de calidad desde los datos.	Análisis de contenido de foros, chats, páginas web y documentos.
Seguimiento de conocimiento	Estimar el dominio de habilidades de estudiantes, utilizando modelos cognitivos y los registros de la evaluación como evidencia	Monitorear el nivel alcanzado por los estudiantes en el tiempo.
Factorización de matriz no negativa	Definición de matrices en base a resultados de evaluación que se descomponen en otras matrices que representan habilidades obtenidas.	Evaluación de habilidades.

Fuente: Romero y Ventura (2013), Shahiri y Husain (2015).

Antecedentes de algunas investigaciones en Latinoamérica que utilizaron la Minería de Datos

A nivel Latinoamericano, algunas instituciones educativas de nivel superior han desarrollado ciertos proyectos de investigación con respecto a la deserción estudiantil, orientados a la aplicación de minería de datos para el descubrimiento de patrones y

causas de la deserción, temas que fueron abordados teniendo en cuenta problemas que son comunes en el ámbito educativo de los países de Latinoamérica.

En la Universidad Nacional de Misiones (Argentina) se realizó una investigación sobre deserción estudiantil utilizando las técnicas de minería de datos. Su objetivo principal fue maximizar la calidad que los modelos tienen para clasificar y agrupar a los estudiantes, de acuerdo con sus características académicas, factores sociales y demográficos, que han desertado de la Carrera Analista en Sistemas de Computación de la Facultad de Ciencias Exactas, Químicas y Naturales analizando los datos de las cohortes entre los años 2000 al 2006 Pautsch (2009) y Pautsch et al., (2010).

En la Universidad Nacional del Nordeste (Argentina) se realizó un estudio cuyo objetivo principal fue aplicar técnicas de bodegas de datos y minería de datos basadas en *clustering* para la búsqueda de perfiles de los alumnos de la asignatura Sistemas Operativos de la Licenciatura en Sistemas de Información según su rendimiento académico, situación demográfica y socioeconómica, que permita conocer a priori situaciones potenciales de éxito o de fracaso académico (La Red et al., 2010).

La Universidad Tecnológica de Izúcar de Matamoros (México), se propuso una investigación cuyo objetivo fue identificar las causas que motivan la deserción de sus estudiantes desde que ingresan. Mediante la técnica de minería de datos clasificación y la herramienta *Weka,* encontraron relaciones entre atributos académicos que identifican y predicen la probabilidad de deserción y propusieron una herramienta para el tutor que le permite predecir la probabilidad de deserción de cualquier alumno en cualquier momento de su estancia escolar Valero (2009) y Valero et al., (2009).

Colombia, en la Universidad de La Sabana se realizó un proyecto de investigación donde el objetivo era seleccionar, de una base de datos de estudiantes, los atributos que tuvieran mayor incidencia en la deserción de la Universidad en los últimos cuatro años, con la técnica de minería de datos clasificación por Rough Sets utilizando el paquete ROSE2 (Restrepo et al., 2008).

En México, en el Instituto de Estudios Superiores de Tamaulipas se realizó una investigación denominada: Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación superior particulares, consistió en analizar la forma en que la minería de datos (MD) permite optimizar el proceso de captación de matrícula, la intención de diseñar un modelo predictivo de gestión de matrícula para las IES particulares de México. Con un método correlacional, se utilizó un conjunto de datos de prospectos ficticios para crear un árbol de decisión con enfoque de criterio de disminución de entropía con el *software Rapid Miner.* Los resultados muestran que es posible construir y probar un modelo predictivo de gestión de matrícula como el

ZAM & EST propuesto por quienes escriben, para que las IES particulares puedan mejorar sus procesos de captación (Estrada et. al., 2016).

En Brasil, en la Universidad de Brasilia se realizó la investigación denominada Minería de datos educativos: un análisis de perfil de estudiantes brasileños, cuyo objetivo fue el análisis relacionado con la situación actual de la educación brasileña, así como el perfil de estudiantes e instituciones. Se analizó la modalidad de cursos ofrecidos por las instituciones educativas, investigando el perfil y la cantidad de estudiantes que tienen becas, y cómo ingresan a los cursos. Los resultados demostraron que el 30% de los estudiantes eligen no informar su color de piel; de los que informaron, el 37.4% son blancos y el 6% son negros. Aproximadamente el 82% de los estudiantes ingresaron a la educación superior mediante el método tradicional (examen de ingreso), y el 82.3% de los hombres y el 77.2% de las mujeres optaron por cursos presenciales. (Fernandes, 2019)

En Colombia se realizó el artículo científico de investigación cuyo objetivo era analizar la forma en que la minería de datos (MD) permite optimizar el proceso de captación de matrícula. Esto, con la intención de diseñar un modelo predictivo de gestión de matrícula para las IES particulares de México. Se analiza la situación actual de las instituciones de educación superior (IES) particulares en relación con su proceso de captación de matrícula y la aplicación de la MD en este. Con un método correlacional, se utilizó un conjunto de datos de prospectos ficticios para crear un árbol de decisión con enfoque de criterio de disminución de entropía con el *software Rapid Miner*. Los resultados muestran que es posible construir y probar un modelo predictivo de gestión de matrícula como el *ZAM & EST* propuesto por quienes escriben, para que las IES particulares puedan mejorar sus procesos de captación. (Estrada y otros, 2016)

La deserción estudiantil es uno de los fenómenos que más preocupa a los directivos de instituciones educativas en los diferentes niveles académicos. En el caso particular de la educación superior, dicho fenómeno es relevante para identificar características o factores que influyen en el problema. A continuación, el artículo que muestra los resultados del análisis sobre la deserción estudiantil corresponde a la Universidad Nacional de Colombia - sede de Manizales, donde se aplicó técnicas de minería de datos educativos. Para ejecutar la muestra, se utilizó un conjunto de datos con 655 registros en la prueba, reunidos desde el primer semestre académico (2009) hasta el segundo semestre académico (2014), y un *software* de código abierto llamado *Weka* cuya versión es 3.8.2. Además, las técnicas de clasificación, como los árboles de decisión. Y se usaron reglas de inducción; además, las técnicas de selección de atributos fueron aplicados para reducir la dimensionalidad del conjunto de datos iniciales. Se comprobó que en los procesos de ingreso y admisión a la institución se suelen tener muchos vacíos de información que hicieron que se tuviese que retirar una buena parte de la recolección de dicha información para posteriores periodos académicos. (Hernández y otros, 2018)

En Ecuador se realizó una investigación cuyo objetivo fue la implementación de minería de datos en la gestión académica de las instituciones de educación superior. La preocupación de los administradores educativos por mejorar la eficiencia académica y cumplir con el encargo social de potenciar el talento humano superando las limitaciones propias del estudiante y del entorno educativo, exige la implementación de procesos basados en el uso de tecnología e información. Este trabajo en particular hace una revisión sobre las técnicas de minería de datos aplicables para el análisis de los problemas concernientes a los actores académicos y el centro educativo. Versa la discusión y análisis sobre el papel de la minería de datos en la educación. Se describe su utilidad para analizar las distintas aristas de las actividades educativas y como su implementación podría ayudar a cada sector involucrado para mejorar la calidad en el desempeño de las instituciones educativas de una manera inclusiva. Se revisan las limitaciones de la minería de datos educativa. Se realiza una exploración de la literatura previa, para lo cual se realizaron búsquedas en diferentes revistas, publicaciones, artículos y otros. (Terán y otros, 2017)

En Paraguay, en la Universidad Americana se realizó la investigación denominada Análisis de la formación del profesorado en matemáticas en el sistema de educación primaria de Paraguay utilizando técnicas de aprendizaje automático. Los datos, recogidos de los profesores de matemáticas de la educación primaria de todo el país, como resultado arroja que existen dificultades en la capacidad de comprender las diferentes estructuras de los tipos de evaluación en las matemáticas. En este trabajo se utilizaron los siguientes clasificadores: red bayesiana (BN), clasificador Bayesiano ingenuo (NB) y máquinas de vectores soporte (SVM) (Jiménez, y Otros 2019).

Los principales enfoques de MDE (Minería de Datos en Educación) son:1) la predicción, cuyo objetivo es el desarrollo de un modelo que infiera de una variable a partir de combinación de datos disponibles; 2) *clustering con* el fin de agrupar datos naturalmente, separando el conjunto en una serie de categorías, descubriendo nuevos patrones de comportamiento; 3) búsqueda de relaciones entre variables, obteniendo asociaciones curriculares en secuencias de cursos y la identificación de estrategias que guíen los procesos de enseñanza (Jiménez y Álvarez, 2010)

Conclusiones

La minería de datos en educación y su implementación es una realidad en el ámbito de las ciencias, a través de la misma se solucionarán varios problemas en el ámbito educativo, tanto en los docentes como en los alumnos. En varios casos se puede predecir ciertos fenómenos que pueden ocurrir y prevenirlos a través de acciones puntuales, lo que hace que para el país sea de suma importancia, teniendo en cuenta la escasa evolución tanto en la calidad como en la inversión en el Sistema Educativo, lo que hace que esta nueva forma

de solución a través de una técnica informática unida a una cantidad importante de datos suministrados, conciba que la confiabilidad y validez en los resultados sean consideradas como viables.

Los resultados que surgen de la minería de datos en el ámbito educativo generarán la toma de decisiones en las instituciones educativas, las cuales incrementarán el nivel de calidad que se brinda a través de nuevos conocimientos que pueden ser descubiertos en los datos que se encuentran en las bases de datos y en los resultados de actividades académicas.

La generación y almacenamiento de datos, si bien aún no es una práctica en las instituciones educativas, sería una de las formas de llegar a posibles planteamientos de problemas, teniendo en cuenta la cantidad de datos que recogidos posibilitarían que se pueda detectar los mismos en el ámbito y llegar a las posibles soluciones a corto o largo plazo.

En cuanto al enfoque investigativo la minería de datos educacionales ofrece numerosas ventajas comparándola con los paradigmas más tradicionales de realizar investigación relativa a la educación, como experimentos de laboratorio, estudios sociológicos o investigación de diseño. En un futuro la creación de repositorios públicos de datos educacionales analizados y con argumentación científica será de suma importancia para incluso proponer políticas educativas efectivas en el país.

Referencias

Ballesteros Román, A. (2012). Minería de Datos Educativa Aplicada a la Investigación de Patrones de Aprendizaje en Estudiante en Ciencias. *Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada, Instituto Politécnico Nacional*, México.

Chalaris, M., Gritzalis, S., Maragoudakis, M., Sgouropoulou, C., & Tsolakidis, A. (2014). *Improving quality of educational processes providing new knowledge using data mining techniques*. Procedia-Social and Behavioral Sciences, 147, 390-397.

Estrada-DanelL, R., Zamarripa-Franco, R., Zúñiga-Garay, P., & Martínez-Trejo, I. (2016*). Contributions to the Enrollment Process with Data Mining in Private Higher Education Institutions*. Revista Electrónica Educare, 20(3), 1-21. https://doi.org/10.15359/ree.20-3.11

Fernandes, Eduardo y Holanda, Maristela y Victorino, Marcio & Borges, Vinicius & Carvalho, Rommel & Erven, Gustavo Van, 2019. *"Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil,"* Journal of Business Research, Elsevier, vol. 94(C), pages 335-343.

G.; Smith P.; Ramasasmy U. Fayyad, U.M.; Piatetskiy-Shapiro.(2006) *Advances in Knowledge Discovery and Data Mining. AAA*I Press / MIT Press.

Hernández-Leal, E. J., Quintero-Lorza, D. P., Escobar-Naranjo, J. C., Ramírez-Gómez, J. S., & Duqueméndez, N. D. (2018). *Educational data mining for the analysis of student desertion.* Learning Analytics for Latin América 2018, 2231, 51-60.

Jiménez V, García M, Vázquez Noguera J, Cabrera C, Riego Esteche A, Divina F ,Marrufo Vázquez, M (2019) *Analysis of Teacher Training in Mathematics in Paraguay's Elementary Education System Using Machine Learning Techniques*. Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019). CISIS 2019, ICEUTE 2019. Advances in Intelligent Systems and Computing, vol 951. Springer, Cham.

Rosado Gómez, y A Verjel Ibáñez, A (2017) *Aplicación de la minería de datos en la educación en línea*. Revista Colombiana de Tecnologías de Avanzada. ISSN: 1692-7257 - Volumen 1 – Número 29 –92-98.

La Red, D., Acosta, J.C., Cutro, L., Uribe, V.E. & Rambo, A.R. (2010). "*Data Warehouse y Data Mining Aplicados al Estudio del Rendimiento Académico"*. Novena Conferencia Iberoamericana en Sistemas, Cibernética e Informática, CISCI 2010. Orlando: International Institute of Informatics and Systemics. Memorias CISCI 2010, Volumen I, pág. 289-294.

Liñán, L. C., & Pérez, Á. A. J. (2015). *Educational Data Mining and Learning Analytics: differences, similarities, and time evolution*. International Journal of Educational Technology in Higher Education, 12(3), 98-112.

Luna-Ramírez, Enrique, Hernández-Chessani, David', Soria-Cruz, Jorge y Cruz Valenzuela, Roberto. *Extracción de reglas a partir de conjuntos de datos del EXANI-II mediante el clasificador J48 de WEKA*. Revista de Docencia e Investigación Educativa 2016, 2-3: 35-40

Pautsch, J. (2009). *Minería de datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación.* Tesis de grado (Licenciado en Sistemas de Información), Posadas, Misiones (Argentina): Universidad Nacional de Misiones, pág.193.

Pautsch, J. La Res, D. y Cutro, L. (2010). *Minería de datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación*. [en línea]. Posadas: Universidad Nacional de Misiones. http://www.dataprix. com/files/Analisis%20de%20Desercion%20Univ_0.pdf. [fecha de consulta: 18/06/2019].

Terán, Harold Elbert Escobar; Saltos, Maritza Alcívar; De La Plata, Carlos Márquez; Terán, Charles Edisson Escob *Implementación de minería de datos en la gestión académica de las instituciones de educación superior*. Revista Didasc @ lia: Didáctica y Educación. Jul- Sep2017, vol. 8 Edición 3, p203-212. 10p.

Restrepo, M. y López, A. (2008). *Uso de la metodología Rough Sets en un modelo de deserción académica*. XIV Congreso Ibero Latinoamericano de Investigación de Operaciones, CLAIO, 2008. Cartagena: Universidad del Norte, Libro de Memorias CLAIO 2008, pág.108-109.

Romero, C., & Ventura, S. (2013). *Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.

Valero, S. (2009*). Aplicación de técnicas de minería de datos para predecir la deserción*. [en línea]. Izúcar de Matamoros: Universidad Tecnológica de Izúcar de Matamoros, 2009. http://www.utim.edu.mx/~svalero/docs/MineriaDesercion.pdf. [fecha de consulta: 10/06/2012].

Valero, S., Salvador, A. y García, M. (2009). *Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos*. [en línea]. Izúcar de Matamoros: Universidad Tecnológica de Izúcar de Matamoros. http://www.utim.edu.mx/~svalero/ docs/e1.pdf. [fecha de consulta: 10/06/2019].