

MINERÍA DE DATOS EN EDUCACIÓN: UNA REVISIÓN LITERARIA

DATA MINING IN EDUCATION: A LITERARY REVIEW

Jessica Johanna Morales Carrillo¹, Viviana Katherine Trujillo Utreras¹, Sulay Katherine Cevallos Molina¹, Santana Cedeño Hiraída Monserrate²

¹Grupo de Investigación SISCO, Carrera de Computación, ESPAM MFL, campus politécnico sitio El Limón, Calceta, Manabí, Ecuador

²Facultad de Ingeniería en Sistemas, Uleam, Av. Circunvalación - Vía a San Mateo Manta - Manabí - Ecuador

Email: jmorales@espam.edu.ec

Información del artículo Resumen

Tipo de artículo:
Artículo de revisión

Recibido:
11/01/2019

Aceptado:
23/05/2019

Licencia:
CC BY-NC-SA 4.0

Revista
ESPAMCIENCIA
10(1):14-20

La presente investigación fue desarrollada con el objetivo de determinar cuáles son las principales técnicas de minería de datos aplicadas en la educación, para mostrar el nivel de coincidencias e identificar qué necesidades satisfacen el uso de dichas técnicas. Mediante el uso de investigación documental por revisión de artículos científicos se filtró, clasificó y seleccionó las publicaciones que presentan mayor sustento para el análisis realizado, recopilando un total de 25 artículos, publicados entre el 2010 y 2018, escritos generalmente en inglés y español. Realizada la recolección de información se elaboró una matriz para la aplicación de minería de datos en los artículos seleccionados que contiene las características a evaluar (técnicas y herramientas), esto permitió realizar un análisis estadístico del nivel de coincidencia entre las características evaluadas y determinar cuáles fueron las más utilizadas. En el análisis de los artículos encontrados se observó que la mayoría están orientados al desarrollo de métodos de descubrimiento que analizan los datos de repositorios educacionales para comprender a los estudiantes, el entorno en el que aprenden y los factores que afectan la calidad del rendimiento académico. Con base en la investigación se determina que, una de las técnicas más experimentadas es la clasificación y la herramienta más usada es el software WEKA (Waikato Environment for Knowledge Analysis).

Palabras clave: Minería de datos, investigación documental, clasificación, árbol de decisión, weka.

Abstract

The present research was developed with the aim of determining what the main techniques of data mining applied to education are, in order to show the level of coincidence and to identify what needs meet the use of such technics. Through the use of desk research of reviewing scientific articles, publications that show greater support for the purposes of the analysis were filtered, classified, and selected, a total of 25 articles published generally in English and Spanish between 2010 and 2018 were compiled. After data compilation a matrix was developed for the application of data mining for the selected articles which contain the characteristics to be evaluated (technics and tools). This allowed to perform a statistical analysis regarding the level of coincidence of the evaluated characteristics to determine which were predominantly used. Analyzing the articles, it was found that most of them are focused on the development of discovery methods that analyze educational data repository in order to understand students, learning environment, and factors that influence academic performance, the research determined that one of the most experienced techniques is the classification and the most used tool WEKA (Waikato Environment for Knowledge Analysis) software.

Keywords: Data mining, desk research, classification, decision tree, weka.

INTRODUCCIÓN

Eckert y Suénaga (2013) expresan que en la actualidad la sociedad se encuentra en la época de la información,

donde los datos se incrementan de forma considerable; de tal manera, esta información ha sido y es un recurso que permite al ser humano tomar decisiones basados en la experiencia, tal como lo indican Morales y Parraga-Alava

(2018). Sin embargo, el exceso de datos no siempre aumenta el conocimiento porque al procesarlos resultan difíciles, por lo que es conveniente poseer información necesaria para así incrementar su efectividad. Troche (2014) resalta la necesidad de almacenar información y menciona que la experiencia nos ha enseñado que esta información con el tiempo llega a ser un insumo importante para mejorar y evitar los errores del pasado. A partir de la literatura citada, existe la necesidad de analizar información histórica a fin de contar con un repositorio de información preprocesada, de alto grado de consistencia, de manera que se ponga de manifiesto la minería de datos (MD) a través de las distintas técnicas de análisis de información, es entonces que se pone en manifiesto la minería de datos (MD).

Según Pérez et al. (2014) el objetivo principal de la minería de datos consiste en extraer información oculta de un conjunto de datos. En esta misma línea Jaramillo y Paz (2015) señalan que la minería de datos es el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos, para así encontrar modelos a partir de los datos. La MD es la etapa de descubrimiento de conocimiento de la base de datos o también conocida como KDD por sus siglas en inglés Knowledge Discovery from databases (La Red et al., 2015). Para mejor comprensión del proceso de la MD Natek y Zwilling (2014) explican los pasos para el análisis: primero, crear los datos conjuntos; segundo, definir la herramienta de minería de datos a utilizar; tercero, evaluar las técnicas de minería de datos; y cuarto, analizar los datos por cada modelo y elegir el mejor.

Riquelme et al. (2016) complementan lo anterior y explican que la minería de datos es un campo interdisciplinario ya que existen numerosas áreas donde se puede aplicar, prácticamente en todas las actividades humanas que generen datos como, por ejemplo: comercio y banca, medicina y farmacia, seguridad y detección de fraude, recuperación de información no numérica, y astronomía.

Actualmente se ha incrementado la aplicación de la DM en el sector educativo. Fernández et al. (2017) define la minería de datos educacional (EDM, Educational Data Mining) como la aplicación tradicional de técnicas de minería de datos tradicional en el análisis de datos educativa dirigida a la resolución de problemas en el contexto educativo para entender más sobre la actuación académica, estilos de aprendizaje, y diversas cuestiones directamente vinculadas a los estudiantes.

Para Menacho (2017) menciona que uno de los retos que tienen que enfrentar las instituciones de educación superior para ofrecer una mayor calidad educativa, es mejorar el rendimiento académico de los estudiantes. Por

otro lado García et al. (2016) señalan que el abandono de los estudios superiores es un fenómeno cuyo impacto se refleja en el incremento de estudiantes en la universidad, lo que indica un importante desaprovechamiento de recursos.

La aplicación de la minería de datos educacional en diversas investigaciones plantea una solución para la mejora del rendimiento académico, pero el análisis de grandes cantidades de datos y la experimentación de diferentes métodos y técnicas para encontrar un modelo que sea fiable suponen de tiempo y esfuerzo. Existen técnicas de investigación documental que facilitan el análisis de toda la información existente acerca de un tema en común, aplicando técnicas que pueden determinar la concordancia de los resultados y comparar las observaciones de los diferentes autores para llegar a una conclusión acertada. De acuerdo con lo expuesto y a través del análisis bibliográfico se puede determinar cuáles son las principales técnicas de minería de datos aplicados en la educación, así como las herramientas empleadas con mayor frecuencia para este fin.

MATERIALES Y MÉTODOS

Para el desarrollo de este trabajo se aplicó la técnica de investigación documental por revisión de artículos científicos, de acuerdo con el siguiente flujo de procesos:

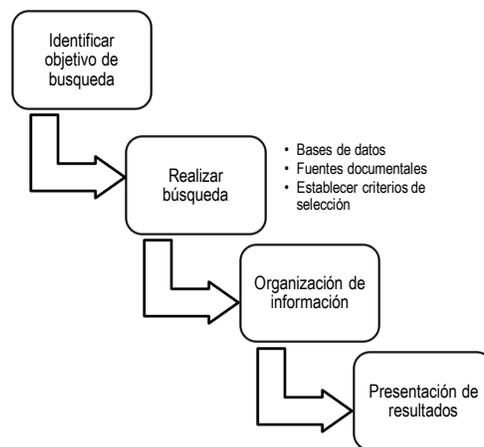


Figura 1. Diagrama de flujo de proceso metodológico

En la identificación del objetivo, se planteó la búsqueda de información de artículo que plantearan el uso de técnica de minería de datos en educación, así como la herramienta empleada para el análisis; En la búsqueda se utilizaron varias fuentes documentales, entre las principales tenemos Springer y Google académico, utilizando los siguientes descriptores:

- Minería de datos aplicada a la educación
- Técnicas de minería de datos educacionales

- Aplicación de técnicas de minería de datos en la educación.

En cuanto a la fecha, se estableció como criterio considerar las publicaciones a partir del 2010 (la búsqueda bibliográfica se la realizó en el mes de junio del 2018).

Una vez realizada la selección de los artículos, se continuó con la organización de información documental electrónica, donde se analizaron y escogieron los temas en relación con la investigación, para ser plasmados en una matriz organizada de forma cronológica considerando los criterios de título, técnica, autores, año y país de la investigación.

Los resultados del análisis de los datos mediante gráficos estadísticos permitió determinar el impacto de la

aplicación de técnicas minería de datos en la educación lo que contribuye en generar elementos de sustento para la utilización de dichas técnicas.

RESULTADOS Y DISCUSIÓN

Selección y organización de los documentos electrónicos: Se realizó la búsqueda de información en la plataforma Springer para obtener un buen porcentaje de publicaciones en el idioma inglés, también se buscó en la web específicamente en Google académico para obtener artículos de sustento teórico y científicamente respaldados.

Cuadro 1. Matriz resumen de documentos electrónicos seleccionados para análisis.

Título del proyecto	Técnicas	Herramientas	Autores	País
A Data Mining Approach to Identify the Factors Affecting the Academic Success of Tertiary Students in Sri Lanka	Clasificación	Software R	(Sanvitha et al., 2018)	India
Predicción del rendimiento académico aplicando técnicas de minería de datos	Clasificación	Weka	(Menacho, 2017)	Perú
Educational Data Mining: Discovery Standards of Academic Performance by Students in Public High Schools in the Federal District of Brazil	Clasificación	H2O	(Fernandes et al., 2017)	Brasil
Minería de datos aplicada para la identificación de factores de riesgo en alumnos	Asociación, Agrupamiento.	Weka	(Reyes et al., 2017)	México
Minería de datos en egresados de la Universidad de Caldas	Clasificación	SQL	(Bedoya et al., 2016)	Colombia
Prediction of Student Dropout Using Personal Profile and Data Mining Approach	Clasificación	Weka	(Meedech et al., 2016)	Tailandia
Aplicación de la minería de datos de la educación en línea	Clasificación	Cross-Platform Platform-Specific, Ad Hoc Tools, Learning Analytics Learning Analytic Frameworks and Tools	(Rosado y Veriel., 2016)	España
Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación superior particulares	Clasificación	Rapid Miner	(Estrada-Danell et al., 2016)	México
Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje	Clasificación	RapidMiner	(Jaramillo y Paz, 2015)	Ecuador
Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos	Clasificación	SPSS Clementine v.9.0	(Alcover et al., 2015)	España
Mining Educational Data to Predict Students' Academic Performance	Clasificación	NetBeans IDE, Weka	(Saleem et al., 2015)	Arabia
Caracterización de la deserción estudiantil en educación superior con minería de datos.	Clasificación, Agrupamiento	Weka	(Jiménez y Toledo, 2015)	Colombia
Using Data Mining Techniques to Detect the Personality of Players in an Educational Game	Clasificación	Weka	(Keshtkar et al., 2014)	USA
Data Mining and Social Network Analysis in the Educational Field: An Application for Non-Expert Users	Clasificación	E-Learning Web Miner (EIWM)	(García et al., 2014)	España
Aplicación de técnicas de Minería de Datos al análisis de situación y comportamiento académico de alumnos de la UGD	Clasificación, Agrupamiento, Asociación	Weka	(Eckert y Suénaga, 2013)	Argentina
Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil	Clasificación	Weka	(Pereira et al., 2013)	Colombia
Mining for Marks: A Comparison of Classification Algorithms when Predicting Academic Performance to Identify "Students at Risk"	Clasificación	Weka	(Mashiloane y Mchunu, 2013)	Sudáfrica
Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo	Clasificación	Weka	(Ballesteros et al., 2014)	México
Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades	Clasificación	SPSS AHA	(Kotsiantis, 2012)	Grecia
Aplicando minería de datos al marketing educativo	Agrupamiento	RapidMiner	(Pinzón, 2011)	México
A data mining approach to guide students through the enrollment process based on academic performance	Clasificación	CRISP-DM methodology	(Vialardi et al., 2011)	Perú
Accuracy Comparison of Predictive Algorithms of Data Mining: Application in Education Sector	Clasificación	Software de código abierto Krime Moodle - Logs del Sistema de Gestión de Aprendizaje de Código Abierto (LMS)	(Sharma y Mavani, 2011)	Berlín
Classification and Prediction of Academic Talent Using Data Mining Techniques	Clasificación	Weka y ROSETTA toolkit	(Jantan et al., 2009)	Malaysia
Minería de Datos en la Educación	Clasificación	Answer Tree, SPSS	(Jiménez y Álvarez, 2010)	España
Sistemas de gestión de contenidos de aprendizaje y técnicas de minería de datos para la enseñanza de ciencias computacionales. Un caso de estudio en el norte de Coahuila	Agrupamiento	Moodle y Weka	(Olague et al., 2010)	México

Revisión y análisis de los artículos: Se realizó un análisis exhaustivo y detallado de los documentos seleccionados lo que permitió observar y conocer que técnicas de MD son más acogidas en el sector de la educación, que herramientas utilizan para su aplicación y en qué países se están implementando.

De acuerdo con el análisis de las 25 revisiones bibliográficas, la técnica de minería de datos educacional que ha sido más utilizada, ha sido la de clasificación (Gráfico 1).

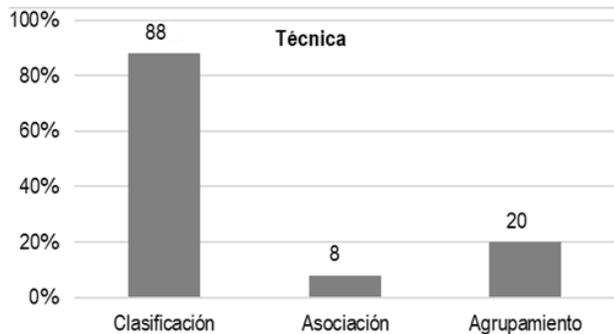


Gráfico 1. Nivel porcentual del uso de las técnicas MD en educación.

Existe un gran número de herramientas para hacer análisis de la información en MD y en este caso existe poca coincidencia ya que las herramientas usadas en las investigaciones son diferentes, incluso el software Weka tiene 37,50 % de similitud (Gráfico 2).

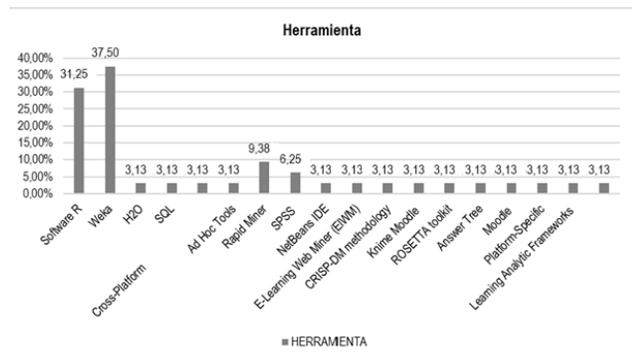


Gráfico 2. Nivel porcentual del uso de las herramientas de MD en educación.

Dentro de la investigación y de acuerdo con la documentación encontrada, los países que presentan mayor aplicación de MD en la educación presentan son México con 20%. España con 16% y Colombia con 12% (Gráfico 3). Es de mencionar que Ecuador a pesar de tener un porcentaje bajo en la aplicación de estas técnicas observamos que sí se están implementando y han obtenido buenos resultados específicamente en educación superior. México según el gráfico 4, ha tenido

una aplicación continua habiendo publicado varias investigaciones en diferentes años.

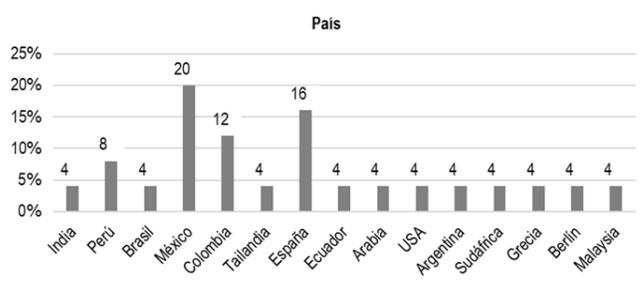


Gráfico 3. Nivel porcentual del uso de las técnicas de MD en países a nivel mundial.



Gráfico 4. Aplicación de técnicas de MD en Educación por año

De acuerdo con los resultados de esta investigación se muestra que el uso de estas técnicas es cada vez más influyente a nivel mundial para mejorar la educación con modelos aplicables, que permiten a los responsables de la educación, analizar el comportamiento del estudiante y determinar los factores que afectan en su rendimiento académico, lo que será determinante en la toma de mejores decisiones para aquellos estudiantes con dificultades en su proceso de formación.

Predecir el rendimiento de los estudiantes en instituciones de educación superior ha sido un tema de interés ya que de esto depende el éxito académico; este fundamento se plantea en el trabajo realizado por Khalaf et al. (2018), el estudio indica que el proporcionar factores que aumentan la tasa de éxito y reduce el fracaso de los estudiantes es profundamente útil para las organizaciones educativas, y como solución plantea la minería de datos para encontrar patrones ocultos y dar sugerencias que mejoren el rendimiento de los estudiantes. Esta investigación presenta un modelo basado en algoritmos de árbol de decisión y sugiere el mejor algoritmo apoyado en el rendimiento, para ello se empleó tres clasificadores (J48, Random Tree y REPTree). Es así como esta publicación propone que los algoritmos de minería de datos y, especialmente, los algoritmos de árbol de decisión pueden ser la mejor solución para predecir el rendimiento de los estudiantes, indicando que los resultados del modelo muestran una

serie de factores (atributos) que afectan los resultados de precisión del árbol y el desempeño académico del estudiante en general. Con lo planteado y de acuerdo con los resultados obtenidos en esta revisión literaria, se pretende continuar con la investigación para aplicar minería de datos según la técnica más utilizada.

CONCLUSIONES

La aplicación de la minería de datos, según la revisión literaria, ha sido experimentada con mayor frecuencia en la educación desde el año 2013. Así mismo, se determinó que uno de los algoritmos más utilizados en este ámbito son los algoritmos dependientes de la técnica de clasificación debido que éstos permiten obtener varias soluciones. Todos los artículos revisados, además, concuerdan que la aplicación de la minería de datos aplicada a los datos históricos en la educación, permitirá corregir las estrategias de enseñanza – aprendizaje.

LITERATURA CITADA

Alcover, R; Benlloch, P; Blesa, M; Calduch, M; Celma, C; Ferri, J; Hernández, L; Iniesta, J; Ramírez, A; Robles, J. 2015. Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos. *Revista en Ciencias de Computacion*. XII: 163-170.

Ballesteros, A., Sánchez-Guzmán, D. and García, R. 2014. Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. *Lat. Am. J. Phys. Educ.* 7(4).

Bedoya, O; López, M; Marulanda, C. 2016. Minería de datos en egresados de la Universidad de Caldas. Medellín, CO. *Revista Virtual Universidad Católica del Norte*. 6:110-124.

Eckert, K y Suénaga, R. 2013. Aplicación de técnicas de Minería de Datos al análisis de situación y comportamiento académico de alumnos de la UGD. Argentina. XV Workshop de Investigadores en Ciencias de la Computación.

Estrada-Danell, R; Zamarripa-Franco, R.; Zúñiga-Garay, P.; Martínez-Trejo, I. 2016. Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación superior particulares. *Revista Electrónica Educare* (3) 20.

Fernández E., Carvalho R., Holanda M., Van Erven G. 2017. Educational Data Mining: Discovery Standards of Academic Performance by Students in Public High Schools in the Federal District of Brazil. *Advances in Intelligent Systems and Computing*, 569: 287-296

García, M., Bernardo, A., Rodríguez. L. 2016. Permanencia en la universidad: la importancia de un buen comienzo. Oviedo-Asturias, España *Open Access funded by Instituto de Ciencias de la Educación*, 44(1):1-6.

García, D; Palazuelos, C; Zorrilla, M. 2014. Data Mining and Social Network Analysis in the Educational Field: An Application for Non-Expert Users. *Educational Data Mining. Studies in Computational Intelligence*. 524:411-439.

Jaramillo, A. y Paz, H. 2015. Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje. Loja, Ecuador. *Revista Tecnológica ESPOL – RTE*, 28 (1):64-90.

Jantan, H; Hamdan, A; Othman, Z. 2009. Classification and Prediction of Academic Talent Using Data Mining Techniques. *Knowledge-Based and Intelligent Information and Engineering Systems*. Notas de la conferencia en Ciencias de la computación. 6276:491-500. <https://doi.org/10.1109/ISDA.2009.64>.

Jiménez, J. y Toledo, S. 2015. Caracterización de la deserción estudiantil en educación superior con minería de datos. Pasto-Nariño, CO. *Revista Tecnológica ESPOL – RTE*. 28: 447-463.

Jiménez, G. y Álvarez, H. 2010. Minería de Datos en la Educación. ES. *Revista de Inteligencia en Redes de Comunicación*. p 2-9.

Keshtkar, F; Burkett, C; Li, H; Graesser, A. 2014. Using Data Mining Techniques to Detect the Personality of Players in an Educational Game. *Studies in Computational Intelligence*. 524:125-150.

Khalaf Hamoud, A., Salah Hashim, A., and Aqeel Awadh, W. 2018. Predicting student performance in higher education institutions using decision tree analysis. *Revista Internacional de Multimedia Interactiva e Inteligencia Artificial*, 5(2).

Kotsiantis, S. 2012. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artif Intell Rev.* 37:331-344.

La Red, D., Karanik, M., Giovannini, M., Pinto, N. 2015. Perfiles de Rendimiento Académico: Un Modelo basado en Minería de datos. *Campus Virtuales*. Argentina, 4(1): 12-30.

Mashiloane, L; Mchunu, M. 2013. Mining for Marks: A Comparison of Classification Algorithms when

- Predicting Academic Performance to Identify “Students at Risk”. *Inteligencia minera y exploración del conocimiento. Conference notes in Computer Science*. 8284:403-414.
- Menacho, C. 2017. Predicción del rendimiento académico aplicando técnicas de minería de datos. La Molina-Lima, Perú. *Revista de Facultad de Economía y Planificación*, 78(1):26-33.
- Meedech, P; Iam-On, N; Boongoen, T. 2016. Prediction of Student Dropout Using Personal Profile and Data Mining Approach. *Intelligent and Evolutionary Systems, Proceedings in Adaptation, Learning and Optimization*. 5:143-155.
- Morales, J. y Parraga-Alava, J. 2018. How Predicting The Academic Success of Students of the ESPAM MFL?: A Preliminary Decision Trees Based Study. *IEEE*. 978-1-5386-6657-9/18
- Natek, S. y Zwilling, M. 2014. Student data mining solution–knowledge management system related to higher education institutions. *Celje, SI. Expert Systems with Applications*, 41(14):6400–6407.
- Olague, J; Torres, S; Morales, F; Valdez, A; Silva, A. 2010. Sistemas de gestión de contenidos de aprendizaje y técnicas de minería de datos para la enseñanza de ciencias computacionales. Un caso de estudio en el norte de Coahuila. Distrito Federal, MEX. *Revista Mexicana de Investigación Educativa*. 15:391-421.
- Pérez, T., Caballero, D., Caro, A., Rodríguez, P., Antequera, T. 2014. Applying data mining and Computer Vision Techniques to MRI to estimate quality traits in Iberian hams. *Caceres, España. Journal of Food Engineering*, 131:82–88.
- Pereira, R; Calderón, A; Jiménez, A. 2013. Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil. Manizales, CO. *Revista de Facultad de Ciencias e Ingeniería*. 28:31-47.
- Pinzón, L. 2011. Aplicando minería de datos al marketing educativo. CO. *Notas D Marketing*. 1:45-61.
- Reyes, A; Flores, A; Alejo, R; Rendón, E. 2017. Minería de datos aplicada para la identificación de factores de riesgo en alumnos. MX. *Revistas de Ciencias de la Computación*. 139:177–189.
- Riquelme, J., Ruiz, R., Gilbert, K. 2016. *Minería de Datos: Conceptos y Tendencias*. Valencia, España. *Revista Iberoamericana de Inteligencia Artificial*, 10 (29):11-18.
- Rosado, A. y Verjel, A. 2016. Aplicación de la minería de datos de la educación en línea. CO. *Revista Colombiana de Tecnologías de Avanzada*. 1: 92-96.
- Saleem, M; Kathiry, N; Osimi, S; Badr, G. 2015. Mining Educational Data to Predict Students’ Academic Performance. *Machine Learning and Data Mining in Pattern Recognition. MLDM 2015. Lecture Notes in Computer Science*. 9166:403-414.
- Sanvitha, K; Liyanage, S; Bhatt, C. 2018. A Data Mining Approach to Identify the Factors Affecting the Academic Success of Tertiary Students in Sri Lanka. *Software Data Engineering for Network Learning Environments Notes on Data Engineering and Communications Technologies*. 11:179-197.
- Sharma, M. y Mavani, M. 2011. Accuracy Comparison of Predictive Algorithms of Data Mining: Application in Education Sector. DE. *Journal of Computer Science*. 11:189–194.
- Troche, A. 2014. Aplicación de la minería de datos sobre bases de datos transaccionales. Salle, BO. *Fides et Ratio - Revista de Difusión cultural y científica de la Universidad La Salle en Bolivia*, 7(7):58-66.
- Vialardi, C; Chue, J; Peche, J. 2011. A data mining approach to guide students through the enrollment process based on academic performance. *User Modeling and User-Adapted Interaction*. 21: 217–248