

APLICACIÓN DE LA MINERÍA DE DATOS EN LA EDUCACION EN LINEA

DATA MINING APPLICATION IN THE VIRTUAL EDUCATION

MSc. Alveiro Alonso Rosado Gómez, Esp. Alejandra Verjel Ibáñez

Universidad Francisco de Paula Santander Ocaña, Facultad de ingenierías, Programa de ingeniería de sistemas.

Resumen: Este artículo hace una revisión de la aplicación de la minería de datos en entornos educativos y de cómo esta permite extraer el conocimiento necesario para identificar el comportamiento de los estudiantes al interactuar con materiales y tutores, el conocimiento obtenido soporta la validez de nuevos modelos educativos que fomenten el aprendizaje. El material consultado se centra en la búsqueda de elementos que tiene la aplicación de la minería de datos, como son: el aprendizaje, la técnica, el algoritmo, la herramienta o software empleado y el conjunto de datos que fueron aplicados en la investigación. Entre los resultados que se encontraron esta la poca especificación que se hace sobre las características del conjunto de datos que se trabaja y el procesamiento que se le hace a este mismo. Por otra parte se encontró que se están aplicando combinación de aprendizajes con el fin de describir y clasificar los datos obtenidos.

Palabras clave: Análisis del aprendizaje, Extracción del conocimiento, Interacción en línea, Minería de datos educativa.

Abstract: This paper reviews the application of data mining in educational environments and how this can extract necessary to identify the behavior of students, tutors and resources in order to establish new models to promote learning knowledge. Research materials focusing on the search for elements of the application of data mining, such as: learning, art algorithm, the tool or software DM and the data set that were applied in the investigation. Among the results we found this little specification is made on the characteristics of the data set and processing work that is done at the same. Moreover it was found that combination of learning is applied in order to describe and classify data.

Keywords: Learning analytics, Knowledge discovery, online interaction, Educational data mining.

1. INTRODUCCIÓN

La aplicación de la minería de datos (DM, Data Mining) en la educación viene teniendo un lugar importante dentro de las investigaciones que se adelantan sobre la información que se almacena dentro de ámbito educativo (Peña-Ayala, 2014). La minería de datos educativa (EDM, Educational Data Mining) o minería de datos aplicada a la educación, es una disciplina emergente que busca desarrollar nuevos métodos para explorar la

información que se genera dentro de los ambientes educativos con el fin de entender la forma en que los estudiantes aprenden (Huapaya *et al.*, 2012), para poder tomar las decisiones adecuadas que garanticen el éxito en el proceso educativo (International Educational Data Mining Society, 2011). Esto incluye tanto a los sistemas educativos tradicionales como los sistemas educativos virtuales (Jormanainen & Sutinen, 2013), con sus entornos de aprendizaje y algunas iniciativas que incluyen la adaptación y mejoramiento de los

materiales didácticos de esas plataformas (Romero & Ventura, 2007).

Entre las diferentes formas que es aprovechada la DM se pueden mencionar las propuestas por (Shu-Hsien *et al.* 2012), los cuales clasificaron la aplicación de la minería de datos en redes neuronales, rendimiento de algoritmos, predicción, análisis de la arquitectura de software, agentes inteligentes, modelamiento, sistemas basados en conocimiento, optimización de sistemas y sistemas de información. Por otro lado (Romero & 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, 2013; Quinteros, *et al.*, 2016), detección de comportamientos indeseables, caracterización de estudiantes, análisis de redes sociales, desarrollo de mapas conceptuales, currículo y organización de las actividades escolares.

Se estudia sobre las recomendaciones de la DM puede dar sobre la interacción del estudiante con la plataforma y de cómo esta puede afectar el rendimiento dependiendo de elementos propios de la usabilidad (Joo & Alvarado, 2013), como puede ser el color, el tipo de fuente, la ubicación de los elementos dentro del entorno, la navegación del curso entre otros (Aponte *et al.*, 2012); esto se debe a que los materiales en línea son mejor recibidos por los estudiantes si están adaptados con su estilo de aprendizaje y si tienen en cuenta la necesidades especiales que puede llegar a tener un estudiante con discapacidad (Parack *et al.*, 2012; Lozano, 2012). Otro aspecto que se encontró que está ligado rendimiento, es la participación activa dentro de un ambiente de aprendizaje en línea (Coldwell *et al.*, 2008); esto se logra en parte monitoreando cual es el comportamiento dentro de un entorno virtual de aprendizaje; aspectos como el tiempo que tarda en resolver una evaluación (Jarusek & Pelánek, 2011) y el tiempo de interacción con la plataforma permite están ligados al resultado final dentro del curso (Romero, *et al.*, 2010; Dominguez, *et al.*, 2010).

Peña-Ayala (2014), organiza las investigaciones antes mencionadas por criterios de aplicación, quedando estructuradas de la siguiente forma: los estudios centrados en el comportamiento del estudiante son los más comunes con un 21.62%, los siguen los de rendimiento estudiantil con 20.72% y los relaciones con la evaluación 20.27%, el porcentaje más bajos están los relacionados con

el acompañamiento y retroalimentación a los estudiantes con un 9.46% y por último los trabajos centrados en el currículo con un 8.56%.

Uno de los elementos que incide con el rendimiento académico es la elección de la carrera profesional; si el estudiante está satisfecho con su currículo obtendrá mejores resultados y demorara menos tiempo en la universidad (Campagni *et al.*, 2015). Pero qué pasa cuando el estudiante tiene inconvenientes con sus materias y se le dificulta su aprendizaje (Castro & Lizasoain, 2012), uno de esos casos es cuando los currículos incorporan cursos que desarrollan temas relacionados con los lenguajes de programación, para esos casos EDM, trata de establecer los comportamiento del estudiante (Jin, *et al.*, 2012; Stahovich & Lin, 2016). Todo esto se logra con la aplicación de aprendizaje de máquina para automatizar actividades como la evaluación y el análisis de elementos como las líneas de código, los errores que este puede tener y el tiempo que transcurre entre una compilación y otra, pueden ayudar a determinar el estilo de programación de cada estudiante y por lo tanto conocer sus fortalezas y debilidades en disciplina (Blikstein, 2011).

El aporte EDM, en el análisis de estos comportamientos permite al docente tener argumentos e información precisa y rápida de sus estudiantes que de la forma tradicional; en otros enfoques no es sencillo medir de forma rápida y aislada las limitaciones de los educandos; identificando cuales son las barreras que le impiden dominar las competencias propuestas para proponer un acompañamiento y asistencia particular que garantice una apropiación de la temática del curso (Fernandez-Medina *et al.*, 2013).

La EDM, va evolucionando conforme las formas de aplicarla cambian, los elementos que la integran y que se generan a partir de los cientos, miles y millones de registros (Chen *et al.*, 2017) que se generan por la demanda y utilización de los recursos educativos en línea (Ballesteros Román *et al.*, 2014), generando nuevas corrientes de investigación y por lo tanto nuevas comunidades de investigadores; es por esta razón que además de hablar de EDM, se habla de Aprendizaje y Conocimiento Analítico (LAK, Learning Analytics and Knowledge) y también se encuentra la Sociedad para la Investigación del Aprendizaje Analítico (SoLAR, Society for Learning Analytics Research); estas dos comunidades abarcan diferentes tópicos relacionados con el aprendizaje y

la aplicación de inteligencia artificial y DM (Piety *et al.*, 2014).

2. METODOLOGIA

EDM, tiene un *journal* que integra investigaciones que se vienen realizando sobre este tema, en él se puede encontrar las aplicaciones de DM en el contexto educativo (International Educational Data Mining Society, 2011), como complemento de estos referentes se buscaron temas relacionados en los sitios de publicaciones científicas de la ACM (Association for Computing Machinery) y dentro del sitio de ElSevier (ScienceDirect).

Los criterios de revisión de los artículos fueron elementos que están presentes en la aplicación de DM y que son relevantes al momento de emprender proyectos que se enmarcan en este tema. Estos elementos son: el tipo de aprendizaje, la técnica aplicada, el algoritmo, la herramienta o software de DM y el conjunto de datos que lo integran.

2.1 Aprendizaje

En la minería de datos existen dos tipos de aprendizaje, lo cuales son el aprendizaje supervisado o predictivo y el aprendizaje no supervisado o descriptivo, el primero se basa en entrenar un set de datos de un modelo para que por medio de diferentes datos se pueda predecir el comportamiento de una variable (Sanchez-Santillan *et al.*, 2016); el segundo método permite agrupar datos de forma rápida y también son llamados métodos simétricos o indirectos; las observaciones se clasifican en grupos que no son conocidos con anticipación, las variables pueden estar conectadas según vínculos desconocidos, así, todos los atributos disponibles son evaluados en el mismo nivel y no hay casualidades (Emtiyaz & Keyvanpour, 2011).

2.2 Técnica

Las técnicas de DM son el resultado de un largo proceso de investigaciones; esta evolución se inicia cuando se empieza a almacenar la información de organizaciones en las computadoras, continúa con mejoras en el acceso de datos y más recientemente se han generado tecnologías que los usuarios naveguen por la información en tiempo real y la habilidad de almacenamiento de grandes bases de datos es crítica para la minería de datos (Chiarini *et al.*, 2010, Araque *et al.*, 2013). DM toma este proceso evolutivo y va más allá del acceso

retrospectivo de los datos y navegación analizando la información para luego mostrar resultados (Ngai *et al.*, 2009).

Las técnicas se aplican según el aprendizaje que se está aplicado es por esta razón que se agrupan en Predictivas y Descriptivas (Kurgan, 2006); las primeras están compuestas por tareas de calificación y regresión mientras que la segunda está conformada por agrupación, asociación y correlación (Rodríguez, 2010).

2.3 Algoritmo

La aplicación de algoritmos de minería han permitido detectar patrones en los datos y, por ende, crear modelos que sustenten la toma de decisiones, y así contribuir al mejoramiento de un problema en particular (Dueñas, 2009). Estos algoritmos son operaciones codificadas como un conglomerado finito de reglas que toman los datos como entradas y sus salidas como modelos o patrones (Gahegan, 2001).

2.4 Herramienta

Las herramientas en DM dan respuesta a preguntas que en el pasado eran difíciles de responder, existen muchas técnicas de DM, con sus ventajas y desafíos en diferentes tipos de aplicaciones, entre estas se encuentran las redes neuronales y la detección de interacción automatizado de chi-cuadrado (Rygielski *et al.*, 2002).

2.5 Datos

DM, está relacionada con el almacenamiento de datos, estos son obtenidos de operaciones diarias de las organizaciones y deben ser tratados para que permitan el análisis y toma de decisiones; estos beneficios se logran estableciendo aspectos como la limpieza de datos y el acceso a datos; en la limpieza de datos (Choi *et al.*, 2017), las instituciones se ven obligadas a estandarizar la variedad de datos que tienen, de manera que se establezcan categorías, esto unifica el manejo de los datos y los errores que estos puedan tener; en el acceso a los datos, se deben establecer métodos bien definidos, de manera que se pueda tener una interacción fácil con los datos y no tener inconvenientes para revisarlos posteriormente (Wu, 2013).

Siempre que las instituciones han establecido los datos para ser analizados, se utiliza comúnmente el

procesamiento analítico en línea online analytical processing OLAP, que se centra en brindar análisis de datos multidimensional, dirigidos a simplificar y apoyar el análisis de datos, para automatizar el proceso tanto como sea posible, el proceso de extracción del conocimiento knowledge discovery in databases KDD es un factor importante de los que está actualmente con el apoyo de la mayoría de los sistemas de bases de datos estándar (Fayyad *et al.*, 1996).

3. RESULTADOS

3.1 Aprendizaje

En cuanto a este elemento de DM, se puede mencionar que existen investigaciones aplicando aprendizaje no supervisado en donde integran las respuestas que los estudiantes realizan en una plataforma de aprendizaje en línea (Burgosa *et al.*, 2017) y la integran con el comportamiento de los mismos en una transmisión de video y como por medio de minería de datos y de texto muestran el comportamiento de los estudiantes en las respuestas que hacen a las preguntas del docente (Cigliuti *et al.*, 2014). Los resultados muestran que el programa académico que matriculan los estudiantes identifica la cantidad de interacción que estos van a tener en estas plataformas (He, 2013); esto se enmarca en la constante búsqueda de obtener mejores resultados al momento de interactuar con plataformas dedicadas a dar respuesta a las preguntas hechas por los usuarios (Atkinson *et al.*, 2013).

3.2 Técnica

Dentro de las técnicas se encontró que existe la tendencia de combinar descripción y clasificación; la primera permite agrupar los datos por su comportamiento y la segunda establecer a que grupo puede pertenecer un estudiante dependiendo de su comportamiento (Costa *et al.*, 2017).

Se busca agrupar los datos de acuerdo con su característica y dimensionalidad con el fin de obtener mejores resultados al momento de realizar DM (Azarnoush *et al.*, 2013).

3.3 Algoritmo

La aplicación de DM para predecir el rendimiento académico de los estudiantes, aplicando técnicas como son J48, M5P y RepTree, este estudio encontró que el modelo generado con REPTree indicó 97,0588% en 66 casos eran correctamente

clasificado y 2,9412% (2 casos) incorrectamente clasificados. Por el lado de J48 encontraron 98,5294% (67 casos) que fueron correctamente clasificados y el 1,4706% (1 caso) incorrectamente clasificado. Para el caso de M5P con un coeficiente de correlación indicado de 0.9358 y una relación error absoluto del 32,2884%, lo cual indica que el poder de clasificación de esta técnica es menor que las otras dos (Natek & Zwilling, 2014).

Este artículo, busca encontrar el algoritmo que brinde el mejor resultado de predicción de rendimiento de los estudiantes en una institución de educación superior, ella se encontró que la mejor predicción fue hecho por el algoritmo *Multilayer Perception* (Kaur *et al.*, 2015).

3.4 Herramienta

Hernández y Conde (2014), enuncian cuatro (4), tipos de herramienta para hacer análisis de la información dentro de un entorno virtual de aprendizaje: la primera de ellas es *cross-platform and platform-specific*, las cuales generan un cuadro de mando que se encarga de almacenar y mostrar gráficamente los comportamientos de los distintos actores que intervienen con la plataforma. El segundo tipo de aplicación es la Ad hoc tools, la cual es construida con un fin específico, para monitorear un conjunto limitado de variables, lo cual produce que tenga problemas de escalabilidad. El tercer tipo de herramienta, son las *Learning Analytics tools*, son Software que se adapta o se instala como complemento de la plataforma de aprendizaje en línea y permiten monitorear el comportamiento del desarrollo del estudiante dentro del curso y su interacción con la plataforma (Márquez Vera *et al.*, 2012). Una de las características más relevantes de esta clase de herramienta es la forma de mostrar la información dado que permite visualizar los datos desde varios puntos de vista y con diferentes tipos de gráficos (Ranellucci *et al.*, 2016). La cuarta herramienta *Learning Analytic frameworks and tolos*, integra las mismas funciones que la *Learning Analytics tools*, pero integra funcionalidades que permiten analizar e interpretar el comportamiento de los estudiantes.

3.5 Datos

Existen repositorios de datos como el ofrecido por la del Centro de Estudio del Aprendizaje Pittsburgh (PSLC, Pittsburgh Science of Learning Center), el cual es alojado por la Universidad Carnegie Mellon, y tiene como finalidad es almacenar los

datos sobre las interacciones de los estudiantes con herramientas educativas en línea (Pavlik, 2013).

4. CONCLUSIONES

La aplicación de la minería de datos, abarca una gran cantidad de escenarios, entre ellos el de la educación. En este artículo se mostró como EDM, permite a los investigadores en educación, soportar y describir el comportamiento del estudiante al interactuar con su entorno de aprendizaje.

Uno de los aspectos que también se pudo intuir con las investigaciones citadas, fue el que la descripción de los atributos y el conjunto de datos en general, no se comenta de forma ampliada, lo cual no permite a otros investigadores tener criterios soportados sobre la forma de tratar los atributos y el proceso de adecuación de los datos que se debe tener cuando se desea realizar una investigación semejante.

Dentro de las tendencias que se encontraron están la del aprendizaje centrado en el diseño y la de realizar investigaciones que permitan conocer el comportamiento del estudiante para ofrecerle los recursos y temas más acordes con su forma de aprendizaje.

REFERENCIAS

- Andonie, R. (2010). Extreme Data Mining: Inference from Small Datasets. *International Journal of Computers, Communications & Control*, págs. 280-291.
- Aponte, F., Hoyos, J., & Monsalve, J. (2012). Minería de usabilidad aplicada a plataformas virtuales. *Revista Virtual Universidad Católica del Norte*, 27-43.
- Araque G J., Díaz R J.L., Gualdrón G O.E. (2013). Optimización del THD en un convertidor multinivel monofásico usando algoritmos genéticos. *Revista colombiana de tecnologías de Avanzada*. 1 (21). Pág. 60 - 66.
- Atkinson, J., Figueroa, A., & Andrade, C. (2013). Evolutionary optimization for ranking how-to questions based on user-generated contents. *Expert Systems with Applications*, 7060–7068.
- Ballesteros Román, A., Sánchez Guzmán, D., & García Salcedo, R. (2014). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. *Latin-American Journal of Physics Education*, 662-668.
- Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. *International Conference on Learning Analytics and Knowledge 2011* (págs. 110-116). Banff: ACM.
- Burgosa, C., Campanario, M., de la Peña, D., Lara, J., Lizcano, D., & Martínez, M. (2017). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers & Electrical Engineering*.
- Caballero Amaury, Velasco Gabriel, Pardo García A. (2013). Differentiations of objects in diffuse databases. *Revista colombiana de tecnologías de Avanzada*. 2 (22). Pág. 131 – 137.
- Campagni, R., Merlini, D., Sprugnoli, R., & Verri, M. (2015). Data mining models for student careers. *Expert Systems with Applications*, 5508–5521.
- Castro, M., & Lizasoain, L. (2012). Las técnicas de modelización estadística en la investigación educativa: minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales. *Revista Española de Pedagogía*, 131-148.
- Chen, Y., Crespi, N., Ortiz, A., & Shu, L. (2017). Reality mining: A prediction algorithm for disease dynamics based on mobile big data. *Information Sciences*, 82-93.
- Chiarini, M., Dutta, K., & Vandermeer, D. (2010). Using Data Mining Techniques to Discover Bias Patterns in Missing Data. *Journal of Data and Information Quality (JDIQ)*, 2.
- Choi, J., Kim, B., Hahn, H., Park, H., Jeong, Y., Yoo, J., & Kee, M. (2017). Data mining-based variable assessment methodology for evaluating the contribution of knowledge services of a public research institute to business performance of firms. *Expert Systems with Applications*, 37–48.
- Cigliuti, P., Martins, S., Rodríguez, D., & García, R. (2014). Identificación de Contenidos Débilmente Apropriados por Estudiantes Basada en Explotación de Información Caso de Estudio. *XII Workshop de Tecnología Informática Aplicada en Educación* (págs. 1-10). Buenos Aires: Universidad Nacional de La Matanza.
- Coldwell, J., Craig, A., Paterson, T., & Mustard, J. (2008). Online Students: Relationships between participation, demographics and academic performance. *The Electronic Journal of e-Learning*, 19–30.

- Costa, E., Fonseca, B., Marcelo, A., Ferreira, F., & Regod, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 247–256.
- Dominguez, A., Yacef, K., & Curran, J. (2010). Data mining to generate individualised feedback. *Intelligent Tutoring Systems. Lecture Notes in Computer Science*, 303–305.
- Dueñas, M. X. (2009). Minería de datos espaciales en búsqueda de la. *SciELO*, 137-156.
- Emtiyaz, S., & Keyvanpour, M. (Octubre de 2011). Customers Behavior Modeling by Semi-Supervised Learning in. *Advances in information Sciences and Service Sciences(AISS)*, 3(9).
- Fayyad, U., Piatetsky, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in. *AI Magazine*, 17(3), 37-54.
- Fernandez-Medina, C., Pérez-Pérez, J., Álvarez-García, V., & Paule-Ruiz, M. d. (2013). Assistance in computer programming learning using educational data mining and learning analytics. *Innovation and Technology in Computer Science Education* (págs. 237-242). Canterbury: ACM.
- Gahegan, M. (Septiembre de 2001). Data mining and knowledge discovery in the. *National Academies Computer Science and Telecommunications Board*.
- He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, 90–102.
- Hernández-García, Á., & Conde, M. (2014). Dealing with complexity: educational data and tools for learning analytics. *Technological Ecosystems for Enhancing Multiculturality* (págs. 263-268). Porto: ACM.
- Hernández, C., Contreras, B., & Torres, C. (2016). Desarrollo de Libros Electrónicos: “Taller Pedagógico”. *Revista Colombiana de Tecnologías de Avanzada*, 130-136.
- Huapaya, C., Lizarralde, Francisco, Arona, G., & Massa, S. (2012). Minería de datos educacional en ambientes virtuales de aprendizaje. *XIV Workshop de Investigadores en Ciencias de la Computación* (págs. 996-1000). Entre Ríos: Red de Universidades con Carreras en Informática (RedUNCI).
- International Educational Data Mining Society. (2011). *Educational Data Mining*. Recuperado el 1 de Septiembre de 2015, de <http://www.educationaldatamining.org/>
- Jarusek, P., & Pelánek, R. (2011). Problem Response Theory and its Application for Tutoring. *Fourth International Conference on Educational Data Mining*, (págs. 1-2). Eindhoven.
- Jin, W., Barnes, T., Stamper, J., John, M., Johnson, M., & Lehmann, L. (2012). Program representation for automatic hint generation for a data-driven novice programming tutor. *International conference on Intelligent Tutoring Systems* (págs. 304-309). Chania: Springer-Verlag.
- Jormanainen, I., & Sutinen, E. (2013). An open approach for learning educational data mining. *Koli Calling '13 Proceedings of the 13th Koli Calling International Conference on Computing Education Research* (págs. 203-204). Koli: ACM New York.
- Joo, J., & Alvarado, V. (2013). Evaluación multi criterio/multi objetivo aplicada a datos sobre educación: una primera aproximación. *Revista Educación y Tecnología*, 112-123.
- Kaur, P., Singh, M., & Josan, G. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. *Procedia Computer Science*, 500–508.
- Kurgan, L. (2006). A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review*, 21(01), 1-24.
- Lozano, C. (2012). El futuro de las tecnologías digitales aplicadas al aprendizaje de personas con necesidades educativas especiales. *Revista de Educación a Distancia*, 1-43.
- Márquez Vera, C., Romero Morales, C., & Ventura Soto, S. (2012). Predicción del Fracaso Escolar Mediante Técnicas de Minería de Datos. *IEEE-RITA*, 109-117.
- Natek, S., & Zwilling, M. (2014). Student data mining solution—knowledge management system related to higher education institutions. *Expert Systems with Applications*, 6400–6407.
- Ngai, E., Li, X., & Dorothy, C. (2009). Application of data mining techniques in customer relationship management: A literature review

- and classification. *Expert systems with applications*, 36(2), 2592-2602.
- Quinteros, O., Funes, A., & Ahumada, H. (2016). Extracción de conocimiento en el cursado del ciclo común de articulación de carreras de Ingeniería. XVIII Workshop de Investigadores en Ciencias de la Computación (págs. 223-226). Entre Ríos: Red de Universidades con Carreras en Informática (RedUNCI).
- Parack, S., Zahid, Z., & Merchant, F. (2012). Application of data mining in educational databases for predicting academic trends and patterns. *Proceedings of 2012 IEEE international conference on technology enhanced education (ICTEE)* (págs. 1-4). Kerala: IEEE.
- Pavlik, P. (2013). Mining the Dynamics of Student Utility and Strategy Use during Vocabulary Learning. *JEDM - Journal of Educational Data Mining*, págs. 39-71.
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 1432–1462.
- Piety, P., Hickey, D., & Bishop, M. (2014). Educational data sciences: framing emergent practices for analytics of learning, organizations, and systems. *International Conference on Learning Analytics And Knowledge* (págs. 193-202). Indianapolis: ACM.
- Ranellucci, J., Poitras, E., Bouchet, F., Lajoie, S., & Halle, N. (2016). Understanding Emotional Expressions in Social Media Through Data Mining. En *Emotions, Technology, and Social Media* (págs. 85-103). Academic Press.
- Rincón, I., Suarez, S., & Sánchez, E. (2015). Pensamiento Crítico de los Estudiantes que Utilizan Facebook Como Nueva Tecnología de Aprendizaje. *Revista Colombiana de Tecnologías de Avanzada*, 133-140
- Rodríguez, J. E. (2010). Fundamentos de minería de datos.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 135–146.
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on systems, man, and cybernetics, part C: applications and reviews*, 40(6), 601-618.
- Romero, C., Espejo, P., Zafra, A., Romero, J., & Ventura, S. (2010). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, 135–146.
- Rosado Gómez, A. (2013). Modelo de detección de estudiantes excluidos en carreras de ingeniería utilizando Minería de Datos. *Revista Ingenio UFPSO*, 47-53.
- Rygielski, C., Cheng, J., & Yen, D. (Noviembre de 2002). Data mining techniques for customer relationship management. *Technology in Society*, 24(4).
- Sanchez-Santillan, M., Paule-Ruiz, M., & Nuñez, J. (2016). Predicting Students' Performance: Incremental InteractionClassifiers. *L@S '16 Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (págs. 217-220). Edinburgh: ACM New York.
- Shu-Hsien, L., Pei-Hui, C., & Pei-Yuan, H. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications*, 11303-11311.
- Stahovich, T., & Lin, H. (2016). Enabling data mining of handwritten coursework. *Computers & Graphics*, 31-45.
- Wu, X. (26 de Junio de 2013). Data mining with big data. *IEEE Xplore*, 97 - 107.