

# I CONGRESO IBEROAMERICANO DE DOCENTES

CONGRESO VIRTUAL DEL 26 NOVIEMBRE AL 08 DICIEMBRE DE 2018

ALGECIRAS (CÁDIZ) DEL 06 AL 08 DICIEMBRE DE 2018

Actas del Congreso Iberoamericano de Docentes

Estado del Arte sobre el Uso de Minería de Datos  
para el Aprovechamiento de Recursos Digitales en  
Educación Superior

Miller Gonzalez

Edimer Santos

Carolina Mejía

ISBN: 978-84-948417-0-5

Edita **Asociación Formación IB.**

Coordinación editorial: **Joaquín Asenjo Pérez, Óscar Macías Álvarez, Patricia Ávalo Ortega y Yoel Yucra Beisaga**

Año de edición: **2018**

Presidente del Comité Científico: **César Bernal.**

El I Congreso Iberoamericano de Docentes se ha celebrado organizado conjuntamente por la Universidad de Cádiz y la Asociación Formación IB con el apoyo del Ayuntamiento de Algeciras y la Asociación Diverciencia entre otras instituciones.

<http://congreso.formacionib.org>



red  
iberoamericana  
de docentes



formaciónib))

# Estado del Arte sobre el Uso de Minería de Datos para el Aprovechamiento de Recursos Digitales en Educación Superior

Miller Gonzalez, Edimer Santos y Carolina Mejía, *Facultad de Estudios en Ambientes Virtuales, Universidad EAN, Bogotá, Colombia*

[mgonzal64666@universidadean.edu.co](mailto:mgonzal64666@universidadean.edu.co), [esantos18740@universidadean.edu.co](mailto:esantos18740@universidadean.edu.co)

[cmejiaco@universidadean.edu.co](mailto:cmejiaco@universidadean.edu.co)

## Resumen

En este estudio se realiza una revisión del uso de la minería de datos para el aprovechamiento de los recursos digitales en educación superior. Planteando inicialmente un recorrido sobre el concepto e inicio de la minería de datos y su uso inicial en educación e investigación mostrando allí que la minería de datos es una herramienta de alta actividad en educación permitiendo una investigación pre, durante y post obre un objetivo determinado y conlleva a unas pequeñas transformaciones de las prácticas en ambientes virtuales. Desde la investigación se muestra como la metodología de datos es usada en distintos campos y su influencia en otros, que en un principio no se tenía como finalidad pero que resulta ser útil en un estudio determinado al analizarse en tres etapas distintas.

*Palabras clave: Minería de datos, educación superior, educación virtual, recursos digitales, Colombia.*

## Introducción

En documento tiene tres partes, la primera muestra la construcción del problema, sus objetivos y justificación, además el campo de acción y alcance del estudio; la segunda muestra la construcción del estado del arte a nivel general sobre la minería de datos, su definición, marco histórico, aplicaciones y énfasis; en la última se muestra cómo la minería de datos es utilizada en educación superior para descubrir diferentes factores relacionados con la permanencia de los estudiantes durante su formación.

## El problema

La educación superior en Colombia ha existido desde el Siglo XVI, en el siglo XIX se fundan las primeras universidades estatales. Para el siglo XX se empieza a desarrollar la libre cátedra y aparecen las universidades privadas. A finales del siglo se crean los estatutos y la autonomía universitaria. En las siguientes décadas se inicia el proceso de evaluación de Alta Calidad basados en los estándares del Sistema Nacional de la Información de la Educación Superior (SNIES) para el 2015 solo el 16% contaba con esta calificación, (Melo, Ramos & Hernández, 2017).

De otro lado, la educación virtual es un fenómeno que se está dando a nivel mundial permitiendo que los estudiantes puedan acceder a cursos de nivel técnico hasta doctoral desde cualquier lugar del mundo; para ello es necesario una metodología que muestre los aspectos más sobresalientes para estudiar de forma virtual, presencial o mixta.

En ese sentido, el estudio quiere construir un estado del arte que permita conocer las estrategias que evidencian cómo se podría lograr un alto nivel de compromiso y dedicación por parte de los estudiantes sin descuidar su estilo de vida o relaciones sociales, ya que esta “ayuda a responder a los desafíos que han de enfrentar los universitarios, en tres espacios definidos: el de las posibilidades, el de las tecnologías y el de la pedagogía” (López C., 2006).

## **Enfoque del estudio**

La minería de datos, “también conocida como Descubrimiento de Conocimiento en Bases de datos (sus siglas en inglés son “KDD – Knowledge Discovery in Databases”), es el campo que nos permite descubrir información nueva y potencialmente útil de grandes cantidades de datos” (Jiménez, Hugo, & García, 2010) y por medio de estas técnicas, encontrar factores que ayuden a identificar el éxito de los estudiantes que ingresan a la universidad en la modalidad virtual.

Asimismo, la tecnología en el siglo XXI es vista como la herramienta para cambiar y mejorar la perspectiva del mundo educativo que requiere de una transformación de las prácticas in situ y aún de las que se practican a distancia o virtual. Cardona (2002) replantea este esquema y brinda herramientas en relación a la instrumentalidad para el aprendizaje usando la internet.

Además de la tecnología, está la necesidad de interacción y la comunicación, por lo que “la interactividad y control están determinados por las capacidades y recursos tecnológicos de que dispone el emisor y, sobre todo, el receptor” (Cardona, 2002).

## **Objetivos**

- Desarrollar el estado del arte en el uso de minería de datos para el aprovechamiento de recursos digitales en educación superior con la posibilidad de implementación dentro de la Universidad EAN.
- Determinar cómo se usa la minería de datos en la educación superior, específicamente en la educación virtual

## **Justificación**

Según la revista Dinero (2017), alrededor del 37% de estudiantes que ingresan a la educación superior no finaliza sus estudios. Es por esta razón que actualmente los programas de carácter virtual en Colombia han crecido hasta en un 98% (Díaz, 2018), convirtiéndose en una oportunidad tangible de alcanzar un título de formación profesional.

Las técnicas de minería de datos están enfocadas hacia los profesores y hacia los estudiantes; cuando se basa en los primeros busca ayudar a mejorar el funcionamiento de las plataformas, los contenidos, las evaluaciones y la retroalimentación dentro de la misma. Cuando están enfocadas a los segundos se enfoca en cómo mejorar el aprendizaje e-learning.

## **Estado del Arte**

En este estudio, se consideró importante concentrar los esfuerzos en sistematizar; el contexto, alcance, aplicaciones y técnicas de la minería de datos, así como las aplicaciones de la minería de datos en entornos educativos.

## **Minería de datos**

En la actualidad se encuentran las bases de datos y se requiere encontrar formas de analizar y aprender a organizar todo tipo de información puesta allí y detectar algunos patrones de búsqueda y caracterizarlos, buscando dar sentido al extenso campo de información y ser una herramienta que ayude a interpretar todo tipo de datos (Gilbert, Riquelme, & Ruiz, 2006).

Sabiendo que el flujo y producción de información crece exponencialmente, se requieren métodos que ayuden a su interpretación de tal manera que se pueda “extraer información útil en la toma de decisiones” (Gilbert et al., 2006); Así al tratar de mostrar que “los datos en bruto raramente son beneficiosos directamente” como una frase coloquial pero ha tomado mucha fuerza por el hecho que hay información errónea, falsa o de poco peso a nivel científico o carente de fuentes en sí misma (Gilbert et al., 2006).

El procesamiento en sus inicios era manual, ayudado por la estadística el analista podría convertirse en “procesador de preguntas sofisticado” (Gilbert et al., 2006), la tecnología avanza y la información también así que la capacidad humana ha sido sobrepasada; se han creado entonces metodologías inteligentes de análisis de datos KDD, en la que se busca extraer conocimiento de estos (Gilbert et al., 2006).

### ***Definición***

La minería de datos KDD es la que permite extraer información importante, no trivial, implícita, es decir, se cree desconocida y de un uso potencialmente grande, es decir, la minería de datos descubre patrones o relaciones entre datos, además de clasificar nuevos datos basados en los existentes y así permitir tomar decisiones en distintos campos de la investigación, educación y otros campos de estudio.

### ***Alcance***

Con base en el documento de Han (2007) se explican algunos de los alcances de la minería de datos: descripción multidimensional de los conceptos; frecuencias, patrones, asociaciones y correlaciones existentes entre los datos en comparación con las casualidades inmersas dentro de la información encontrada; clasificación de información y predicciones; predicción de números desconocidos u olvidados; análisis de conglomerados (clúster), de casos atípicos o de tendencias.

### ***Campos de aplicación***

El texto de Gilbert et al. (2006) nos muestra claramente los campos donde se ha hecho uso del KDD: Comercio y banca, medicina y Farmacia, Seguridad y detección de fraude, Recuperación de información no numérica, astronomía, Geología, minería, agricultura y pesca, Ciencias Ambientales, Ciencias Sociales y Planificación de ciudades. (pp. 14), además identifica que el KDD está presente en campos que permiten recoger datos. En educación se evidencia la recolección de datos desde su ingreso a la educación superior y posteriormente se pueden identificar aspectos sobre hábitos de estudio y características de cada una de las carreras.

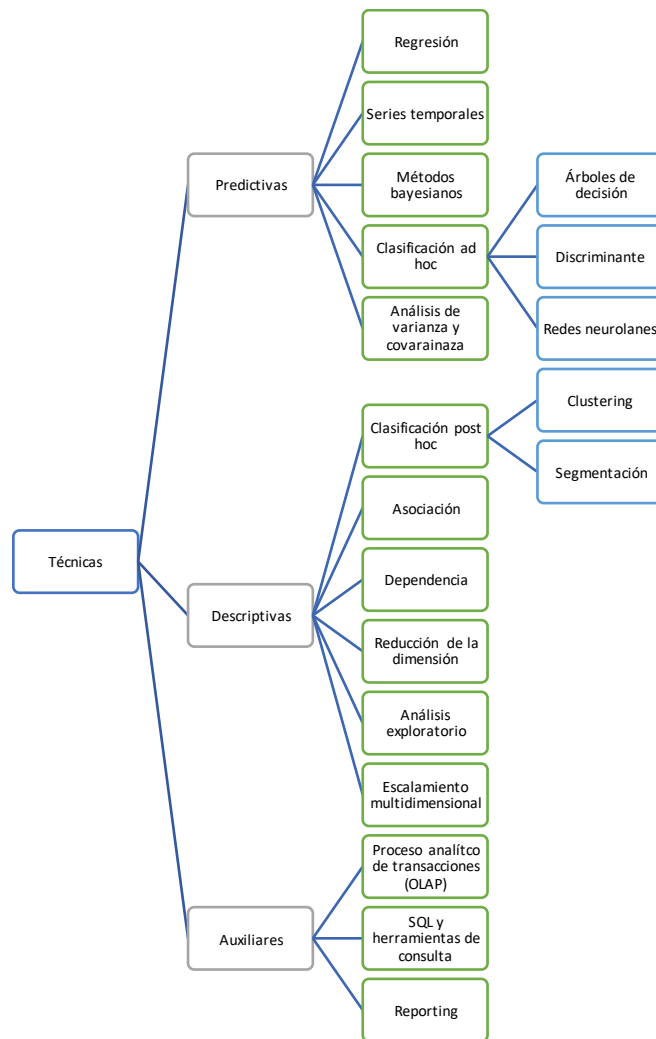
### ***Técnicas usadas***

Con base en el texto de Pérez & Santín (2007) se muestran las técnicas usadas: descriptivo o predictivo, luego se realiza un ajuste del modelo acorde con el objetivo de la investigación.

Las técnicas predictivas están basadas en conocimiento previo, diferenciando entre dos tipos de variables: independientes y dependientes; por lo tanto, se debe validar después de realizar el proceso, identificando 4 fases: identificación objetiva, estimación, diagnóstico y predicción.

Las técnicas descriptivas no adopta un valor a la variable (independiente o dependiente), en consecuencia, son construidos al reconocer patrones.

En la Figura 1 se muestran las técnicas usadas adaptadas de Pérez & Santín (2007).



**Figura 1.** Técnicas en Minería de Datos

**Fuente:** (Pérez & Santín, 2007)

De la figura 1 observamos que el KDD usa dos aspectos claves, la descripción y predicción, en educación se hacen las dos cosas al mismo tiempo, describiendo aspectos generales de un comportamiento o tarea y se predice el nivel de éxito sobre la misma, lo que lleva entonces a que se identifique en la descripción y que se requiere para que haya una predicción con un intervalo de confianza superior al 95%.

### ***Minería de datos en entornos educativos***

La aplicabilidad de la minería de datos con fines educacionales se sitúa en el año 2009 (Baker) con el nacimiento del grupo de investigación de Minería de Datos Educacional (Educational Data Mining, EDM, por sus siglas en inglés). A partir de allí se ha constituido como “una disciplina emergente preocupada por el desarrollo de métodos que exploren los datos únicos y a gran escala para entender mejor a los estudiantes y la manera en cómo estos aprenden” (Barnes, Heffernan, & Heiner, 2007); a partir del 2008 se han celebrado anualmente conferencias internacionales en este tema, siendo la más reciente en China en el año 2017.

De acuerdo con (Baker, n.d.), el procesamiento de grandes cantidades de datos ha permitido que la información sea procesada de manera ágil y dinámica con resultados altamente

confiables. La EDM ha permitido ser aplicado en aspectos como: Modelos en predecir comportamientos de los estudiantes, Descubrimiento y mejora en modelaje de datos para dominios de internet basados en algoritmos de búsqueda, Estudios para crear softwares de aprendizaje. El texto de Castro, Mugica, Nebot, & Vellido (n.d.) se muestra una definición similar.

Por otra parte, Romero & Ventura (2010) con base en una encuesta se resalta el punto de vista de la minería de datos en ambientes virtuales, se destaca la importancia del internet en todo el proceso renovado de formación académica. Así, la información se organiza en las siguientes categorías: instructores, estudiante, perfil, correlación entre el estudiante y conceptos, y creación de cursos.

### ***Aplicabilidad de la minería de datos en educación***

De la aplicabilidad de la minería de datos en entornos educativos se resaltan: los hechos por Barnes, Becket, & Joazeiro de Baker (2008), Yacef, Zaïane, Hershkovitz, Yudelsohn, & Stamper (2012), Boticario et al. (2015), Castro et al. (n.d.)

De las plataformas actuales más importantes en el campo de los ambientes virtuales está el de García (2016).

De los objetivos de la minería de datos en ambientes educativos esta Romero, Ventura, Pechenizkiy, & SJD Baker (2011).

Estos análisis han podido generar conclusiones en los siguientes aspectos: Prediciendo el éxito académico, Prediciendo resultados del curso, Prediciendo nuevas tareas, Prediciendo habilidades, hábitos y motivación, Meta cognitiva. (Peña-Ayala, 2014).

De manera complementaria, el estudio de Espigares y García (2010) concluye que “la necesidad de que en el futuro se integren las herramientas de análisis estadísticos de la minería de datos educativos dentro de las propias plataformas de tele formación”

### **Conclusiones**

La minería de datos se ha convertido en un aspecto importante dentro de la evaluación de los ambientes educativos desde programas escolares para infantes o niños e inclusive más avanzados como en la educación profesional formal. Los avances estadísticos, el procesamiento ágil de datos a través de hardware de alta tecnología y nuevas técnicas de carácter predictivo han sido claves en la consecución de los resultados mostrados a través de este documento.

Sin embargo, el reto es alto en el cambio de paradigmas y uso de nuevos métodos tanto de aprendizaje como de enseñanza ya que al igual que en todo proceso comunicativo se debe garantizar que el mensaje sea transmitido desde su fuente (docente) hasta el receptor (estudiante) y el canal (plataforma educativa) lo haga de manera adecuada dando el espacio para que el contexto y retroalimentación sea el adecuado como se puede observar en (Hervás, Morales, & Soto, 2005)

Siguiendo a Romero et al. (2011) se encuentra que, dependiendo del tipo de estudiante, actividad del curso, fuente de información e inclusive aspectos demográficos, se puede predecir si un estudiante aprueba o no un curso o una tarea que podría determinar su permanencia en el curso al que se ha inscrito.

García (2016) y Boticario et al. (2015) muestran como los recursos digitales son importantes en los procesos de educación superior a nivel virtual y presencial usando técnicas de minería de

datos que arrojan predicciones sobre el comportamiento en los foros y las descripciones que se pueden hacer allí de acuerdo a las participaciones de cada uno de los integrantes.

## Referencias

- Baker, R. S. J. D. (n.d.). Data Mining for Education. *International Encyclopedia of Education*
- Barnes, T., Heffernan, N., & Heiner, C. (2007). Educational Data Mining. (Webpage)
- Barnes, T., Becket, J., & Joazeiro de Baker, R. (2008). Educational Data Mining 2008
- Boticario, C., Desmarais, M., Hershkovitz, A., Luna, J., Merceron, A., Mitros, P., ..., Vutura, S. (2015). Proceedings of the 8th International Conference on Educational Data Mining Madrid – Spain.
- Cardona, G. (2002). Tendencias educativas para el siglo XXI. *Edutec. Revista Electrónica de Tecnología Educativa*, 15 (Mayo)
- Castro F., Vellido A., Nebot À., Mugica F. (2007) Applying Data Mining Techniques to e-Learning Problems. In: Jain L.C., Tedman R.A., Tedman D.K. (eds) *Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence*, vol 62. Springer, Berlin, Heidelberg
- DINERO. (2017). Deserción y abandono de la educación universitaria en Colombia. Retrieved March 15, 2018
- García, D. (2016). Minería de datos aplicada a la enseñanza virtual: nuevas propuestas para la construcción de modelos y su integración en un entorno amigable para el usuario no experto. Universidad de Cantabria
- Gilbert, K., Riquelme, J. C., & Ruiz, R. (2006). Conceptos y Tendencias. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*. No, 29, 11–18
- Hervás, C., Morales, C. R., & Soto, V. (2005). Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web. In *Actas del III Taller Nacional de Minería de Datos y Aprendizaje* (pp. 49–56)
- López C., R. (2006). Hacia un sistema virtual para la educación en México. *Apertura*, 6(3), 6–23
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems With Applications*, 41, 1432–1462
- Pérez, C., & Santín, D. (2007). *Minería de datos: técnicas y herramientas*. Thomson.
- Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *APPLICATIONS AND REVIEWS*, 40(6)
- Romero, C., Ventura, S., Pechenizkiy, M., & SJD Baker, R. (2011). *Handbook of Educational Data Mining*
- Yacef, K., Zaïane, O., Hershkovitz, A., Yudelson, M., & Stamper, J. (2012). EDM 2012 5th International Conference on Educational Data Mining