

Henry Fernando Vallejo Ballesteros ^a; Edelmira Guevara Iñiguez ^b; Segundo Rafael Medina Velasco ^c

Minería de Datos

*Revista Científica Mundo de la Investigación y el Conocimiento. Vol. 2 núm.,
especial, mayo, ISSN: 2588-073X, 2018, pp. 339-349*

DOI: [10.26820/recimundo/2.esp.2018.339-349](https://doi.org/10.26820/recimundo/2.esp.2018.339-349)

Editorial Saberes del Conocimiento

Recibido: 03/12/2017

Aceptado: 12/02/2018

- a. Universidad Estatal de Bolívar; hvallejo@ueb.edu.ec
- b. Universidad Estatal de Bolívar; eguevara@ueb.edu.ec
- c. Universidad Estatal de Bolívar; smedina@ueb.edu.ec

RESUMEN

La realización de base de datos se ha vuelto una acción fundamental para las empresas, ya que les permiten crear estrategias para conseguir nuevos clientes o fidelizar a los habituales. Pero a consecuencia de la generación masiva de datos, nos encontramos frente a un problema, la infoxicación, disponemos de tanta información, que a veces es imposible organizarla con efectividad. Por ello, la clave está en descubrir patrones o algoritmos para sacarle el máximo partido, y aquí es donde entra en juego el Data Mining o minería de datos. El Data Mining es un conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos que expliquen el comportamiento de estos datos. A pesar de que la idea del Data Mining puede parecer una innovación tecnológica muy reciente, en realidad este término apareció en los años sesenta conjuntamente con otros conceptos como por ejemplo, el data fishing o data archeology. No obstante, no fue hasta los años ochenta cuando empezó su consolidación. La minería de datos surgió con la intención o el objetivo de ayudar a comprender una enorme cantidad de datos, y que estos, pudieran ser utilizados para extraer conclusiones para contribuir en la mejora y crecimiento de las empresas, sobre todo, por lo que hace a las ventas o fidelización de clientes. Por tanto, los datos son el medio o la base para llegar a conclusiones y transformar estos datos en información relevante, para que las empresas puedan abarcar mejoras y soluciones que les ayuden a conseguir sus objetivos.

Palabras claves: Datos, Minería, Algoritmo, computación, sistema.

ABSTRACT

The realization of the database has become a fundamental action for companies, since they allow them to create strategies to get new clients or retain the usual ones. But as a result of the massive generation of data, we are facing a problem, infoxication, we have so much information, that sometimes it is impossible to organize it effectively. Therefore, the key is to discover patterns or algorithms to get the most out of it, and this is where Data Mining or data mining comes into play.

Data Mining is a set of techniques and technologies that allow to explore large databases, automatically or semiautomatically, in order to find repetitive patterns that explain the behavior of these data.

Although the idea of Data Mining may seem a very recent technological innovation, in fact this term appeared in the sixties along with other concepts such as, for example, data fishing or data archeology. However, it was not until the eighties when its consolidation began.

Data mining arose with the intention or aim of helping to understand a huge amount of data, and that these could be used to draw conclusions to contribute to the improvement and growth of companies, especially as regards Sales or customer loyalty.

Therefore, the data are the means or basis to reach conclusions and transform this data into relevant information, so that companies can include improvements and solutions that help them achieve their objectives.

Keywords: Data, Mining, Algorithm, computing, system.

Introducción.

La revolución digital ha hecho posible que la información digitalizada sea fácil de capturar, procesar, almacenar, distribuir, y transmitir. Con el importante progreso en informática y en las tecnologías relacionadas y la expansión de su uso en diferentes aspectos de la vida, se continúa recogiendo y almacenando en bases de datos gran cantidad de información.

Descubrir conocimiento de este enorme volumen de datos es un reto en sí mismo. El avance de la tecnología para la gestión de bases de datos hace posible integrar diferentes tipos de datos, tales como imagen, video, texto, y otros datos numéricos, en una base de datos sencilla, facilitando el procesamiento multimedia.

La minería de datos nació con la idea de aprovechar dos cosas: la ingente cantidad de datos que se almacenaban en áreas como el comercio, la banca o la sanidad, y la potencia de los nuevos ordenadores para realizar operaciones de análisis sobre esos datos. El data mining permite encontrar información escondida en los datos que no siempre resulta aparente, ya que, dado el gigantesco volumen de datos existentes, gran parte de ese volumen nunca será analizado. La minería de datos es un proceso de identificación de información relevante extraída de grandes volúmenes de datos, con el objetivo de descubrir patrones y tendencias estructurando la información obtenida de un modo comprensible para su posterior utilización. (Mata, 2017)

Materiales y métodos

La metodología empleada en la presente investigación es una revisión bibliográfica documental de tipo no experimental. En vista que se utilizó como método de investigación

diferentes artículos y trabajos de investigación alojados en páginas web, referente a la temática del empleo de elementos informáticos dentro de los procesos administrativos.

La minería de datos surge a principios de los años ochenta cuando la Administración de Hacienda de Estados Unidos desarrolló un programa de investigación para detectar fraudes en la declaración y evasión de impuestos, mediante lógica difusa, redes neuronales y técnicas de reconocimiento de patrones. Sin embargo, su expansión se produce hasta la década de los noventa, principalmente debido a:

- a) El incremento en la potencia de procesamiento de las computadoras, así como en la capacidad de almacenamiento.
- b) El crecimiento de la cantidad de datos almacenados se ve favorecido no solo por el abaratamiento de los discos y sistemas de almacenamiento masivo, sino también por la automatización de trabajos y técnicas de acopio de datos (observación con nuevas tecnologías, entrevistas más prácticas, encuestas por internet, etc.).
- c) La aparición de nuevos métodos y técnicas de aprendizaje y almacenamiento de datos, como las redes neuronales, la Inteligencia Artificial y el surgimiento del almacén de datos (Basagoiti Astigarraga, 2007)

La minería de datos tiene como objetivo analizar los datos desde todas las perspectivas estratégicas para la organización, con el fin de transformarla en información útil y conocimiento, siendo de utilidad general para aumentar la facturación, ampliar el margen operativo, etc. En general y, concretamente en las empresas, la minería de datos se soporta mediante utilidades informáticas, onpremise o en cloud, que sirven de instrumentos para el análisis de datos. (kyocera Document Solutions, 2006)

Minería de Datos

Vol. 2, núm. Esp., (2018)

Henry Fernando Vallejo Ballesteros; Edelmira Guevara Iñiguez; Segundo Rafael Medina Velasco

Este tipo de proyectos es de amplia utilidad para analizar los datos desde diferentes dimensiones o puntos de vista, ordenando, clasificando, filtrando y resumiendo todas las relaciones que un dato puede tener dentro de la información de la empresa. Está centrado no solo en extraer conocimiento sino en encontrar las relaciones o correlaciones que la información, vista desde diferentes ámbitos, tiene con otros datos aparentemente no conectados y, generalmente, recogidos en enormes bases de datos relacionales. (Kyocera Document Solutions, 2006)

Este tipo de proyectos, en la actualidad, está penetrando fuertemente en las empresas debido a diferentes aspectos:

- a) Las continuas innovaciones en la capacidad de procesamiento, almacenamiento en disco y software permiten un nivel de análisis, eficiencia y facilidad en uso anteriormente no disponibles.
- b) Las empresas usan información “desconectada” de otros sistemas terceros, generalmente en la nube, que necesitan poder correlacionar con los datos que si se generan de manera interna.
- c) Se tiene la posibilidad de cruzar y relacionar información operativa y no operativa en los sistemas de información. Por ejemplo, cruzar resultados propios de venta con pronósticos del sector, para la toma de decisiones en los planes estratégicos.
- d) El manejo además de la información histórica con las relaciones que estos sistemas son capaces de desarrollar, nos permiten en muchas ocasiones plantear escenarios de comportamiento futuro de las tendencias del mercado, de nuestras ventas o de la evolución de cierta línea de negocio.

- e) Los sistemas de minería de datos, en general, están también alineados con los paradigmas de BigData e Inteligencia Artificial que están llegando cada día más a las empresas. (kyocera Document Solutions, 2006)

La minería de datos hace referencia a un conjunto de algoritmos que permiten la identificación de patrones útiles y novedosos que se encuentran “ocultos” en grandes bases de datos. Previo al desarrollo de los modelos es necesario surtir 4 pasos contemplados por el KDD, los que en adelante se describen.

- a) Entendimiento del problema: se define la pregunta que se quiere resolver y se establecen los objetivos específicos a trabajar.
- b) Selección de datos: se determina qué bases de datos pueden ser utilizadas para analizar el problema y se consolidan los conjuntos de información.
- c) Pre-procesamiento y limpieza: se busca manejar los datos faltantes junto con los valores atípicos que se identifiquen en las bases de datos.
- d) Transformación de datos: se busca generar nuevos indicadores que puedan ser más representativos del fenómeno analizado o contener mejor información que la de las variables disponibles por sí mismas, tomando la información acumulada hasta este momento. (Basagoiti Astigarraga, 2007)

Las aplicaciones de minería de datos

La minería de datos se puede aplicar a una variedad de aplicaciones en prácticamente todas las industrias.

- a) **Los minoristas** pueden implementar minería de datos para identificar mejor qué productos es probable que compren las personas en función de sus hábitos de compra anteriores, o qué productos es probable que vendan en ciertos momentos del año. Esto puede ayudar a los comerciantes a planificar inventarios y diseños de tiendas.
- b) **Los bancos y otros proveedores de servicios financieros** pueden extraer datos relacionados con las cuentas, transacciones y preferencias de canales de sus clientes para satisfacer mejor sus necesidades. También pueden recopilar los datos analizados de sus sitios web y las interacciones en las redes sociales para ayudar a aumentar la lealtad de los clientes existentes y atraer otros nuevos.
- c) **Las empresas manufactureras** pueden usar la minería de datos para buscar patrones en el proceso de producción, de modo que puedan identificar con precisión los cuellos de botella y los métodos defectuosos y encontrar formas de aumentar la eficiencia. También pueden aplicar el conocimiento de la minería de datos al diseño de productos y realizar ajustes basados en los comentarios de las experiencias de los clientes.
- d) **Las instituciones educativas** pueden beneficiarse de la minería de datos, como el análisis de conjuntos de datos para predecir los comportamientos de aprendizaje futuros y el rendimiento de los estudiantes, y luego utilizar estos conocimientos para mejorar los métodos de enseñanza o los planes de estudios.

- e) **Los proveedores de atención médica** pueden extraer y analizar datos para determinar mejores formas de brindar atención a los pacientes y reducir los costos. Con la ayuda de la minería de datos, pueden predecir cuántos pacientes necesitarán cuidar y qué tipo de servicios necesitarán esos pacientes. En las ciencias de la vida, la minería se puede utilizar para obtener conocimientos de datos biológicos masivos, para ayudar a desarrollar nuevos medicamentos y otros tratamientos.
- f) **En múltiples industrias**, incluida la atención médica y el comercio minorista, puede utilizar la minería de datos para detectar fraudes y otros abusos, mucho más rápido que con los métodos tradicionales para identificar tales actividades. (Violino, 2017)

Los riesgos y desafíos de la minería de datos

La minería de datos viene con su cuota de riesgos y desafíos. Al igual que con cualquier tecnología que implique el uso de información potencialmente sensible o de identificación personal, la seguridad y la privacidad se encuentran entre las mayores preocupaciones.

En un nivel fundamental, los datos que se extraen deben ser completos, precisos y confiables; después de todo, lo está usando para tomar decisiones comerciales importantes y, a menudo, para interactuar con el público, reguladores, inversores y socios comerciales. Las formas modernas de datos también requieren nuevos tipos de tecnologías, como reunir conjuntos de datos de una variedad de entornos informáticos distribuidos (también conocido como integración de big data) y para datos más complejos, como imágenes y video, datos temporales y datos espaciales.

Conclusiones.

Nuestra capacidad para almacenar datos ha crecido en los últimos años a velocidades exponenciales. En contrapartida, nuestra capacidad para procesarlos y utilizarlos no ha ido a la par. Por este motivo, el data mining se presenta como una tecnología de apoyo para explorar, analizar, comprender y aplicar el conocimiento obtenido usando grandes volúmenes de datos. Descubrir nuevos caminos que nos ayuden en la identificación de interesantes estructuras en los datos es una de las tareas fundamentales en el data mining. (Estrada-Danell & Zamarripa-Franco, 2016)

El desarrollo de la tecnología de Minería de Datos está en un momento crítico. Existe una serie de elementos que la hacen operable, sin embargo, existen algunos factores que pueden crear un descrédito a la Minería de Datos, como ser; que los productos a comercializar son, en la actualidad, significativamente costosos, y los consumidores pueden hallar una relación costo/beneficio improductiva, que se requiera de mucha experiencia para utilizar herramientas de la tecnología, o que sea muy fácil hallar patrones equívocos, triviales o no interesantes, que no sea posible resolver los aspectos técnicos de hallar patrones en tiempo o en espacio, además, hoy en día, las corporaciones comercializan con millones de perfiles personales, sin que aquellos a que se refieren los datos intercambiados, estén en posibilidad de intervenir, entonces, se llega a pensar que presenta un peligro o riesgo para la privacidad de los clientes. (Ribas, 2014)

Recomendaciones.

Independiente de la técnica y el enfoque que se use, se debe seguir un sistema organizado de análisis. Por lo general, el data mining sigue cuatro etapas básicas:

-
- a) Seleccionar la muestra: Se debe elegir el conjunto inicial de datos que se analizará.
 - b) Registrar las características de los datos iniciales: El periodo de tiempo al que corresponden los datos, el segmento, época y cualquier otra característica, deben ser registrados antes de comenzar el estudio.
 - c) Aplicar la técnica o criterio de procesamiento: Al aplicar una de las técnicas descritas, se crea un modelo de conocimiento que luego puede ser aplicado a otros datos o a un subconjunto de la misma muestra.
 - d) Interpretar los datos: Las conclusiones obtenidas se deben validar revisando su coherencia y pertinencia para el estudio. Si no pasan la prueba, se puede volver a aplicar el proceso, usando otra técnica. (ida.cl, 2006)

Bibliografía.

- Basagoiti Astigarraga, R. (2007). *Técnicas de minería de datos aplicadas a series temporales bursátiles*. Madrid: Mondragón Unibertsitatea.
- Estrada-Danell, R., & Zamarripa-Franco, R. (2016). Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación superior particulares. *Revista Electrónica Educare*, 23(3), 1-21.
- ida.cl. (2006). Minería de Datos: Conceptos y Tendencias. *nteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 10(29), 11-18.
- kyocera Document Solutions. (2006). MINERÍA DE DATOS: HERRAMIENTA DE APOYO EN LA SELECCIÓN DE EQUIPOS DE PROYECTOS INFORMÁTICOS. *Ingeniería Industrial*, 2(3), 7-10.
- Mata, R. (13 de Enero de 2017). *icemd.com*. Obtenido de icemd.com: <https://www.icemd.com/digital-knowledge/articulos/mineria-datos-proceso-areas-se-puede-aplica/>
- Ribas, E. (Julio de 2014). *iebschool.com*. Obtenido de iebschool.com: <https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/>
- Violino, B. (25 de Agosto de 2017). *infoworld.com*. Obtenido de infoworld.com: <https://www.infoworld.com/article/3218151/data-mining/what-is-data-mining-how-analytics-uncovers-insights.html>