

# LAS APLICACIONES EN EL SECTOR PRIVADO DE LA MINERÍA DE DATOS

Rafael Alvarado Cuevas<sup>1</sup>, Saul Eduardo Vences Amador<sup>2</sup>

*1, 2, Facultad de Informática Mazatlán, Universidad Autónoma de Sinaloa, México.*

## Abstract

El procesamiento y análisis de las grandes cantidades de datos que se producen en la actualidad, posibilitan el hallazgo de patrones y tendencias ocultos en los mismos, que impacta directamente en la toma de decisiones en diversas áreas de estudios. Se generan datos a gran velocidad y en grandes cantidades que requieren ser procesados para poder actuar de manera rápida. Un ejemplo es el caso de la observación de turnos que se generan en entidades bancarias, donde hay momentos del día en que se requiere modificar los esquemas de atención, según la abundancia de determinadas categorías de clientes o el incremento de demandas de determinados servicios.

Existen numerosas técnicas de minería de datos aplicables a distintos casos de análisis de datos, que permiten obtener ventajas de esas grandes cantidades de datos almacenados. En este artículo se presentan los tópicos de interés del proyecto Minería de Datos: su aplicación a repositorios de datos masivos.

Palabras clave: Información, Minería de Datos, Pymes.

## 1 INTRODUCCIÓN

El proceso de buscar en los datos para descubrir relaciones y predecir tendencias futuras tiene una larga historia. Conocido también como descubrimiento de conocimientos en bases de datos comprende tres disciplinas científicas entrelazadas: estadística, inteligencia artificial y machine learning.

Lo que se consideraba antiguo regresa otra vez como innovación debido a que la minería de datos continúa evolucionando para igualar el ritmo potencial sin límites del big data y poder de cómputo asequible.

En la actualidad, gracias a los avances tecnológicos se ha permitido llegar más allá de las prácticas manuales, tediosas y que toman mucho tiempo al análisis de datos rápido, fácil y automatizado. Mientras más compleja sea la información recopilada, mayor es la eficiencia que hay para descubrir ideas relevantes.

Una realidad actual de la minería de datos es su papel como una tecnología de apoyo que ha venido a resolver dos grandes retos:

Trabajar con volúmenes inmensos de datos para extraer y descubrir información de interés.

Utilizar técnicas para explorar, analizar, comprender y aplicar el conocimiento adquirido de grandes conjuntos de datos, así como para identificar tendencias y comportamientos que faciliten una mejor comprensión de los fenómenos y sirvan de ayuda en la toma de decisiones.

Debido al importante crecimiento que ha tenido, instituciones como el Instituto Tecnológico de Massachusetts la consideran como una de las diez tecnologías emergentes más importantes del siglo 21 que cambiara el sentido de investigación en el mundo.

El impulso al empleo de minería de datos se asocia a una serie de factores como [1]:

El exponencial crecimiento de la recolección de datos y la evolución del poder de cómputo.

El almacenamiento de los datos con la finalidad de tener acceso a una base de datos actualizada y confiable.

La creciente disponibilidad de información en Internet.

La presión competitiva del mercado en una economía globalizada.

El desarrollo de herramientas comerciales para llevar a cabo minería de datos.

Estos son algunos de los factores más importantes que han llevado a emplear la minería de datos en un campo amplio con aplicaciones en: finanzas, análisis de mercado, procesos industriales, medicina, biología, química, telecomunicaciones, biometría, medio ambiente, hidrología, entre otros.

Como una subárea específica de la Minería de Datos se puede citar al Data Stream Mining, que es el proceso de extraer conocimiento en estructuras de archivos continuas y con rápidas transiciones [2]. Los data streams se generan de forma continua y a altas velocidades. El origen de la información puede provenir de diversas fuentes, como registros generados por clientes que utilizan aplicaciones móviles, transacciones electrónicas, logs de navegación de una red, información de redes sociales, datos provenientes de dispositivos wearables, entre muchos otros ejemplos.

Los diferentes resultados aportan a las empresas visibilidad de numerosos aspectos del negocio y de las actividades de sus clientes, como la tasa de uso de un servicio, la actividad de un servidor, la ubicación geográfica de un móvil, personas, la afluencia de determinado tipo de usuarios, entre otros aspectos, y les permite responder con rapidez ante cualquier situación que surja. Por ejemplo, un banco podría analizar el incremento de determinada categoría de clientes en un momento dado y responder rápidamente habilitando más puestos de atención para los mismos.

Existen dos tipos de agrupamientos a los cuales se les considera principales: el agrupamiento jerárquico que se caracteriza por el desarrollo recursivo de una estructura en forma de árbol, y el agrupamiento particional que organiza los registros dentro de un número indefinido de grupos. Los métodos particionales tienen ventajas en aplicaciones que involucran gran cantidad de datos y los jerárquicos por lo general se utilizan para identificar el número deseado de grupos. Otro tipo de agrupamiento son las redes de Kohonen (Mapa auto organizativo) que son utilizados para agrupar un conjunto de datos en distintas regiones, especialmente cuando no se tiene información de los mismos.

En la literatura, autores definen la minería de datos de diferentes maneras, algunas de ellas son:

Es el proceso de descubrir patrones en los datos, los cuales se presentan en grandes cantidades. Los patrones descubiertos deben ser significativos de manera que se permitan ventajas, por lo general, de tipo económicas [3].

Es el proceso eficiente, no trivial, de extraer información valiosa (patrones y tendencias) de una gran colección de datos [4].

Es el análisis de conjuntos de datos observados, a menudo extensos, para encontrar relaciones insospechadas y resumir los datos en forma comprensible y útil para el usuario de la información [5].

Es la exploración y análisis de grandes cantidades de datos para descubrir reglas y patrones significativos [6].

Las anteriores cuatro definiciones tienen en común la extracción de información potencialmente útil (patrones, asociaciones o relaciones entre los datos) para los usuarios finales. Algunas otras

definiciones consideran además del proceso de descubrir patrones válidos, la exploración y análisis de series extensas de datos. Debido a las diversas formas de definir la minería de datos, se resume el objetivo como el proceso de convertir datos extraídos de grandes volúmenes de información en conocimiento útil.

## 2 METODOLOGÍA

La metodología de esta investigación es exploratoria, se utilizaron como técnicas la revisión de artículos académicos, documentos sobre la aplicación vinculada a la temática, dando a conocer las principales medidas y herramientas usadas en cada caso referente a la minería de datos.

## 3 RESULTADOS

Según la metodología expuesta, se llevará a cabo un análisis en los casos sobre la aplicación de minería de datos para sectores empresariales e industriales; de los cuales se encuentran los siguientes casos en la literatura:

### 3.1 La operación cotidiana de la empresa

Todas las organizaciones y empresas coleccionan y administran datos de suma importancia relacionados con personas, procesos u otro tipo de actividades para las cuales fueron creadas. Los mas comunes son los relacionados con ventas de productos o servicios, empleados, pacientes o con clientes.

Dichas colecciones se pueden almacenar en discos de gran capacidad, que ya es posible comprar y tener en el hogar. Para darse una idea de su capacidad, pueden almacenar el numero del Registro Federal de Causantes (RFC) y la edad de cada uno de los habitantes de la Republica Mexicana, para lo cual basta un disco con capacidad de almacenamiento de un *terabyte* (mil *gigabytes*, equivalentes a un millón de *megabytes*) [7].

#### 3.1.1 Bodegas de datos para la toma estratégica de decisiones.

Con el paso del tiempo, los datos de las aplicaciones OLTP (Procesamiento de transacciones en línea) se transfieren con una serie de procesos conocidos como extracción, transformación y limpieza las colecciones ya conocidas llamadas bodegas de datos, donde uno de sus usos es el análisis; ya sea con el procesamiento analítico en línea (OLAP), o la minería de datos.

Ambos análisis se caracterizan por utilizar un gran numero de datos de interés que se generaron a través de varios días, meses o años, de acuerdo con el interés de la organización.

#### 3.1.2 La minería de datos en los negocios y el marketing.

La necesidad de información surge como consecuencia de dos factores: la incertidumbre y el coste de asumir en caso de equivocarse en la toma de decisiones. En la actualidad ambas características están presentes en niveles elevados para los negocios, haciendo necesaria la información para el éxito en el mercado, hasta el punto de que algunos autores hablan de una economía basada en la información y el conocimiento.

Para obtener la información necesaria se tiene que recurrir de una materia prima: los datos. Estos se encuentran hoy en día con gran facilidad gracias a la complejidad de captación, transmisión y gestión de los mismos que ofrecen las tecnologías de la información.

#### 3.1.3 La rentabilidad económica de las empresas a través de los procesos de minería de datos, como herramienta inteligente competitiva.

Es innegable que organizaciones comerciales, analistas de inversiones, entidades de crédito y demás usuarios especializados, cada vez solicitan más información sobre la rentabilidad de todas las actividades de negocio de la organización. ¿Están capacitados los solicitantes para procesar correctamente todo ese recurso de datos, cada vez más voluminoso y complejo? Nos inclinamos a pensar que no, por la siguiente razón: El volumen creciente de datos a que se enfrentan los usuarios,

especialmente aquellos especializados en economía, finanzas y marketing, está haciendo que la sobrecarga informativa convierta los datos en una masa en la cual no es posible ver con claridad los ámbitos y complejos procesos de la situación y perspectivas de la empresa. Ante este panorama nos preguntamos ¿Qué medidas pueden adoptarse? Cabe indicar que los avances tecnológicos han conducido a un mundo complejo y cambiante, pero contradictoriamente es una fuente de soluciones.

Siguiendo en el mismo orden de ideas, se resalta que la Minería de Datos o explotación de datos ha revolucionado y potenciado las bases de datos tradicionales, pues permite identificar mediante la aplicación automática de algoritmos recursivos, las variables más relevantes para una decisión concreta o los subconjuntos de entidades (empresas, clientes, entre otros) más homogéneos con arreglo a unas características determinadas.

En fin, estos métodos y procedimientos se han convertido en retos tecnológicos para procesar los datos y convertirlos en conocimiento útil para la toma de decisiones. Este camino se presenta como una opción para las organizaciones que quieran ser competitivas, valiéndose de la experiencia acumulada, la cual sin duda alguna constituye el principal activo del que se dispone para la creación de valor. De esta manera, una organización que reflexiona, documenta y aprende, está en condiciones de innovar y obtener ventajas competitivas [8].

### *3.1.4 Influencia de la tecnología para la consolidación de las actividades de negocios establecidas en minería de datos.*

Autores como Han y Kamber (2001) consideran que el auge que ha alcanzado actualmente el *Data Mining* es debido a que en el presente nos encontramos ante enormes cantidades de datos y con la urgente necesidad de transformarlos en información útil y conocimiento. Se dice que sin *Data Mining* somos “ricos en datos” pero “pobres en información”. El *Data Mining* es la evolución natural de la tecnología de información. Por lo antes expuesto, se puede inferir que las necesidades de información de las organizaciones han cambiado a lo largo del tiempo.

Hoy día, la creciente dinámica de mercado y competitividad llevan a la necesidad de contar con la información adecuada en el momento indicado y para ello los gerentes necesitan estar bien informados para poder tomar las decisiones de negocio apropiadas. Por otra parte, los datos con los que cuentan dichas organizaciones generalmente se encuentran dispersos a través de diversos sistemas, propiciando de esta manera que datos valiosos se pierdan. De esto último, se puede decir que lo que poseen las empresas son datos, no información [8].

En este orden de ideas, cabe resaltar la importancia de otras tecnologías de apoyo en los procesos de Minería de Datos, que han sido utilizadas desde hace tiempo, logrando la integración de éstas con la administración de datos, contribuyendo a mejorar la técnica. Thuraisingham (1999) agrega que los métodos estadísticos han propiciado la aparición de varios paquetes de software para computar sumas, promedios y distribuciones, que han ido integrándose con las bases de datos a explorar.

Considerando el criterio de Molina y García (2004), se puede decir que el aprendizaje automático consiste en la obtención de reglas de aprendizaje y modelos de los datos, para lo cual a menudo se necesita la ayuda de la estadística. Por esta razón, los métodos estadísticos y el aprendizaje automático son los dos componentes más importantes de la Minería de Datos. Existen otras tecnologías, entre las que se incluyen la visualización y el procesamiento paralelo, para dar apoyo a la toma de decisiones. Las técnicas de visualización ayudan a presentar los datos, para facilitar la minería de éstos. Por su parte, las técnicas de procesamiento paralelo, a su vez, colaboran para mejorar el rendimiento de la Minería de Datos. Los sistemas de apoyo a la toma de decisiones permiten discriminar los resultados, de manera que puedan ser utilizados en las funciones de dirección en las organizaciones [8].

## **4 CONCLUSIONES**

A lo largo de esta investigación se ha analizado el estado actual de una de las técnicas que más está influyendo a las actividades de negocio de las empresas y en la cual están involucrados un gran número de investigadores a nivel mundial, por las implementaciones, estrategias y beneficios que esta da. Como últimas palabras podemos aclarar que esta tecnología aún no alcanza su meta final, en la actualidad nos encontramos con un gran avance de esta misma, pero siendo conscientes que puede y llegará más lejos, no solo en el ámbito empresarial sino en un sinfín de entornos más.

## REFERENCIAS

- [1] DANIEL T. LAROSE. (2005). DISCOVERING KNOWLEDGE IN DATA. NEW JERSEY: John Wiley & Sons.
- [2] Raúl Mata. (2017). Minería de datos: qué es, cómo es el proceso y a qué áreas se puede aplicar. 25/09/2019, de ICEMD Sitio web: <https://www.icemd.com/digital-knowledge/articulos/mineria-datos-proceso-areas-se-puede-aplica/>
- [3] Witten IH, Frank E. (2005). Data Mining: Practical Machine Learning Tools and Techniques. San Francisco: Morgan Kaufmann.
- [4] S. Sumathi S.N. Sivanandam. (2006). Introduction to Data Mining and its Applications. Berlin: Springer-Verlag Berlin Heidelberg.
- [5] David Hand y Heikki Mannila. (2001). Principles of Data Mining. London: Bradford Books.
- [6] Michael J. A. Berry, Gordon S. Linoff. (2004). Data Mining Techniques. Indiana: John Wiley & Sons.
- [7] Martínez Luna, G. L. (2011). Minería de datos: cómo hallar una aguja en un pajar. Ingenierías, 14(53), 55-63.
- [8] Marcano Aular, Y. J., & Talavera Pereira, R. (2007). Minería de datos como soporte a la toma de decisiones empresariales. *Opción*, 23(52), 104-118.