

# Minería de datos: Concepto, características, estructura y aplicaciones

MARÍA ISABEL ÁNGELES LARRIETA\* y ANGÉLICA MARÍA SANTILLÁN GÓMEZ\*

## Introducción

**E**l objetivo del presente artículo es despertar el interés de las personas que dentro de las empresas son responsables de tomar decisiones directivas que permitan descubrir en los datos históricos información útil para fundamentar mejor sus decisiones, con apoyo de la minería de datos. Por ello se expondrán las características de estos programas, su estructura general y algunas de sus aplicaciones.

A lo largo de la vida de una empresa se acumulan grandes cantidades de datos que son almacenados, algunos de ellos serán usados, otros se acumularán hasta perderse por falta de actualidad o por cambios en las políticas de manejo de datos.

Ahora con el desarrollo de los sistemas de cómputo, las empresas tienen la capacidad de almacenar y acceder, en archivos o bases de datos, grandes cantidades de datos históricos sobre las operaciones diarias de su negocio; información que en su momento fue usada para satisfacer las necesidades propias de la empresa y como soporte de las decisiones. Todos esos archivos contienen normalmente gran cantidad de datos que serían de utilidad si fuera posible aprovecharlos mediante procesos que arrojarían información útil.

La mayoría de las organizaciones no sufre por falta de datos, sino más bien por exceso, por lo que cada vez es más complicado buscar datos específicos y significativos que permitan tener una visión más completa y clara de la situación operacional de la empresa para mejorar la manera en que se toman las decisiones.

Las áreas de sistemas han venido trabajando para crear extractos de información de las bases de datos operacio-

nales y almacenar estos datos en archivos, tratando de responder a las peticiones de los usuarios que necesiten obtener información que les ayude a tomar mejores decisiones. Las necesidades de información han hecho que se diseñen sistemas de información ejecutiva y de apoyo a la toma de decisiones —“tienen como objetivo primordial proveer de toda la información necesaria a los ejecutivos de alto nivel para apoyarlos en la toma de decisiones, además de que les permite tener acceso rápido y efectivo a la información compartida y crítica del negocio”—<sup>1</sup>; sin embargo, las demandas de las empresas, con relación a la información, van más allá de simples consultas, tabulaciones cruzadas o reportes consolidados; lo que ha hecho que se creen nuevas formas de análisis de la información, con ventajas respecto de las que se conocían porque incorporan hechos sistemáticos que relacionan más de dos variables.

## Concepto

La minería de datos es el proceso que tiene como propósito descubrir, extraer y almacenar información

\*María Isabel Larrieta, egresada de la Licenciatura en Informática de la FCA-UNAM, actualmente es jefa del Departamento de Estadística de la Unidad de Sistemas de la Facultad de Medicina de la UNAM.

\*Angélica María Santillán Gómez, egresada de la Licenciatura en Informática de la FCA-UNAM, en este momento labora como técnico administrativo en la Subdirección de Sistemas de la Dirección General de Personal de la UNAM.

<sup>1</sup> GONZÁLEZ BONILLA, Marisol, *Introducción a los sistemas de información ejecutiva y la evaluación de herramientas para su desarrollo*, Soluciones Avanzadas, Núm. 35, 15 de julio 1996.

relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones globales, tendencias, desviaciones y otros indicadores aparentemente caóticos que tienen una explicación que pueden descubrirse mediante diversas técnicas de esta herramienta.

El objetivo fundamental es aprovechar el valor de la información localizada y usar los patrones preestablecidos para que los directivos tengan un mejor conocimiento de su negocio y puedan tomar decisiones más confiables.

### **Ventajas sobre otras herramientas de manejo de datos**

Las características que destacan la mayor parte de los fabricantes de estas herramientas son:

- La minería de datos auxilia a los usuarios empresariales en el procesamiento de reservas de datos para descubrir relaciones de las que, en algunos casos, anteriormente ni siquiera se sospechaba.
- La información obtenida a través de la minería de datos ayuda a los usuarios a elegir cursos de acción y a definir estrategias competitivas, porque conocen información que sólo ellos pueden emplear.
- Los seres humanos tienen la capacidad para percibir excepciones y anomalías rápidamente pero no tienen la habilidad para inferir relaciones que en grandes volúmenes de datos, por lo que la minería de datos, mediante modelos avanzados y reglas de inducción, puede examinar gran cantidad de datos y encontrar patrones difíciles de identificar a simple vista.
- Puede trabajar siguiendo los mismos criterios con grandes cantidades de información histórica.
- El proceso de búsqueda puede ser realizado por herramientas que automáticamente buscan patrones porque así están programadas y despliegan los tópicos más importantes.

### **Estructura**

- Algoritmos o programas de búsqueda mineros.

La minería de datos hace uso de programas de búsqueda para detectar desviaciones, tendencias y patrones ocultos en los datos históricos.

Los mineros son programas pensados y creados por el usuario, en los que se emplean técnicas diferentes para la explotación de los datos, tales como *cluster*, asociaciones, clasificación, visualización, redes neuronales, algoritmos genéticos, detección de desviaciones, entre otros. Todos ellos requieren bases de datos de tamaño considerable para que puedan ser eficientes.

La función de los programas mineros es correlacionar los criterios de selección y búsqueda con los datos históricos; si encuentran algo interesante lo presentan al usuario como un hallazgo.

Los programas mineros trabajan con procesos automáticos principalmente sobre bases de datos relacionales para buscar datos extraños,<sup>2</sup> patrones, tendencias o desviaciones; pueden ser ejecutados fuera de las horas pico, usando tiempos de máquina excedentes de noche o en horas de poco proceso, lo que los convierte en ayudantes importantes.

Una ventaja de los mineros es que no requieren *hardware* especial o dedicado. Trabajan en las redes de oficinas nacionales o regionales, utilizando por las noches el servidor de la base de datos relacional, y las PC's o estaciones de trabajo ya existentes. Es decir trabajan sobre datos ya recolectados, en máquinas ya existentes, realizando labores útiles mientras los usuarios no se encuentran trabajando.

---

<sup>2</sup> Son datos desconocidos o poco comunes en la información que se está manejando.

- Datos históricos (en dónde buscan).

Son datos estables y coherentes que se van acumulando a lo largo de la vida operativa de una empresa.

- Criterios de búsqueda (qué se busca).

Son las normas, tendencias y patrones desde los cuales los programas mineros realizarán el proceso de selección y búsqueda en los datos históricos. La prioridad de búsqueda, los criterios de interés y las explicaciones de situaciones extrañas son definidas por el usuario. Una vez establecidos los criterios de selección y búsqueda. Se analizan los datos históricos reportando los hallazgos inmediatamente en un archivo para su posterior revisión y decisión final.

- Almacenamiento de hallazgos (Cofre de tesoros).

Los hallazgos son los datos resultantes de correlacionar los criterios de selección y búsqueda con los datos históricos. El ser humano desempeña un papel fundamental, ya que sólo él puede decidir si este patrón, tendencia o criterio, tiene importancia, pertinencia y utilidad para la empresa.

## Ciclo de la minería de datos

### ¿Cómo funciona la minería de datos?

El proceso de la minería de datos es un ciclo, debido a que los resultados obtenidos pueden alimentar nuevamente dicho proceso; intervienen, principalmente, cuatro pasos que se describen a continuación:<sup>3</sup>

1. Los usuarios de la información deberán identificar los problemas del negocio y las áreas en donde los datos pueden dar valor agregado a la empresa, esto es: a raíz de un problema surge la necesidad de analizar a detalle los datos de la empresa para poder encontrar posibles soluciones al mismo, o bien, información que haga que las decisiones tomadas sean lo más ciertas posibles. Asimismo, es importante identificar las áreas en donde la información es muy cambiante, pero

primordial para la competitividad de la empresa. Para esto pueden manejarse diferentes criterios, no se puede decir específicamente cuáles son los correctos debido a que esto depende de las características de la empresa, pero el objetivo a perseguir es determinar los criterios, ideas, normas y cuestionamientos que fungirán como entrada para el proceso de minería de datos.

2. El usuario para analizar la información histórica seleccionará el algoritmo o algoritmos adecuados de minería. Posteriormente, estos algoritmos son traducidos a programas mineros que realizarán las búsquedas con los criterios previamente definidos.

Existen varias dificultades que pueden interferir con el resultado que se obtenga del análisis y esto es porque los datos se pueden encontrar en diferentes formas, formatos y en múltiples sistemas, aunado a que pueden provenir de fuentes internas o externas; para resolver este problema actualmente se ha hecho uso del *data warehouse*, que pretende reunir los datos más importantes de la empresa en una especie de base de datos corporativa, la cual requiere una gran cantidad de *gigabytes*, no siempre disponible en las organizaciones, sin embargo, es posible hacer minería de datos sin necesidad de tener el *data warehouse*, pero es muy importante tener claro que la información deberá estar lo más uniforme y congruente posible, ya que mucho depende de esto la certidumbre de los resultados que arroje.

3. Incorporar la información obtenida a través del proceso de minería de datos al proceso de toma de decisiones; así como presentar los hallazgos encontrados a los responsables de las operaciones de forma que la información obtenida pueda integrarse en los procesos de la empresa y pueda aplicarse en la solución de los problemas.

<sup>3</sup> J.A. Berry, Michael, Gordon, Linoff, *Data Mining Techniques*, p.22-35

4. Medir los resultados: Medir el valor de los hallazgos encontrados, que se proporcionan al tomador de decisiones con relación a la solución de los problemas identificados y a los criterios definidos en el primer punto.

## Aplicaciones

Las operaciones comerciales, de algunas de las grandes empresas, hoy en día se basan en informes periódicos producidos por consultas en bases de datos pregrabadas: ¿cuál es la cifra de ventas en una tienda?, ¿qué tendencias se derivan de las cifras de ventas?, ¿en dónde tienen una mejor aceptación los productos: en las tiendas del centro o en las tiendas de la periferia? Los informes con referencias cruzadas forman la base de la mayoría de las decisiones de los ejecutivos. Las consultas se elaboran de manera interactiva con el usuario para garantizar que la información presente una estructura eficiente. Estos informes se han elaborado para responder a preguntas recurrentes.

Por ejemplo, un analista tiene la siguiente suposición: "En el oeste se venden más camisas de manga corta que en el este". Lleva a cabo una consulta en la base de datos, obtiene un informe que confirma o no su suposición. Una forma de probar la hipótesis anterior es la presentada por la minería de datos, donde se sigue todo un proceso de búsqueda de patrones y criterios definidos por el usuario hasta llegar a la información relevante que apoye o niegue su hipótesis.<sup>4</sup>

Los hallazgos encontrados por los programas mineros ayudan a los directivos a analizar los hábitos de los clientes a fin de satisfacer mejor su demanda, mejorar la administración de inventarios y, en general, aumentar sus márgenes de utilidad.

Los principales distribuidores de refrescos y golosinas, como por ejemplo los de papas "Chips", utilizan soluciones de información estratégica para maximizar sus beneficios y mantener un alto nivel de satisfacción entre sus clientes mediante la entrega a tiempo de productos de alta calidad. Para ello administran los movimientos a través de la red de distribución de acuerdo con la información acerca de lo que se vende en las tiendas; estos datos permiten conocer las variaciones de temporada. No obstan-

te, en este negocio, si uno se basa solamente en esta situación y en las previsiones derivadas de ellas, pronto desaparece. Para competir se necesita información en tiempo real sobre lo que sucede; por ello, los conductores de camión llevan ordenadores conectados por radio para informar lo que ven cada vez que visitan a un minorista. Con esta información inmediata, las principales empresas pueden hacer ajustes sin demora y sacar el mejor provecho al inventario perecedero de la red de distribución. Además, la información procedente de las aplicaciones de información estratégica sobre lo que se vende y lo que no, permite a las empresas cambiar su producción en fábrica para adaptarse a la demanda.

Otro caso es el de la moda femenina, un análisis oportuno de la información constituye la clave del éxito en este negocio. La moda es estacional, pues su ciclo de vida es aproximadamente de tres meses en promedio, tiene una vida útil corta en almacén, un alto índice de rotación de existencias y da elevados beneficios. Además es de suma importancia detectar rápidamente las nuevas tendencias. Así, si usted puede predecir exactamente las prendas que se comprarán durante una semana determinada en una tienda, podrá reducir el inventario de la tienda y el desembolso del capital y aumentar sus beneficios. El análisis diario por producto de la información sobre ventas de un comercio puede incrementar su margen en determinado porcentaje.

La minería de datos, que permite la gestión en tiempo real de manera eficaz, es una herramienta aplicable a cualquier tipo de empresa. Una amplia gama de compañías pueden tener aplicaciones exitosas con ella. A continuación se describen otros ejemplos:

Particularmente en el área de mercadotecnia las empresas prefieren concentrar sus recursos en aquellas personas con mayores probabilidades de responder a su oferta, en lugar de dirigirse a miles de personas que no pueden convertirse en sus clientes. Si saben quiénes son estas personas, pueden efectuar un correo directo o una campaña de venta por teléfono destinada específicamente a

---

<sup>4</sup> Update IBM Software, *Gestión competitiva de la información*.



ello. El índice de respuesta a un correo directo masivo es generalmente bajo, entre 0.05 y 2%, si se definen los destinatarios con mayor eficacia, este índice aumentará considerablemente, reduciendo costos y aumentando los beneficios.

El uso de esta herramienta en el diseño de programas para clientes asiduos, es muy importante porque en los años 90 el servicio al cliente es la clave del éxito; conocer al cliente es el primer paso para satisfacer sus necesidades, dicen los especialistas. Para ello, las empresas pueden hacer un seguimiento de las compras del cliente en el nivel de canasta de compra o de lo que han comprado en una serie de visitas previas. La comprensión de tales hábitos de compra permite desarrollar promociones destinadas a un tipo concreto de clientes que, de tener éxito, aumentarán su fidelidad y frecuencia de compra.

En el uso del análisis de siniestros la necesidad de reducir costos y la creciente competencia han hecho que las aseguradoras revisen reclamaciones que reciben para obtener una mayor eficiencia organizativa y para medir el grado de satisfacción de sus clientes. La comprensión de los factores que determinan el periodo necesario para liquidar un siniestro puede reducir el tiempo medio para tramitarlo.

En general los ejemplos muestran las ventajas que tiene el uso en minería de datos para comprender mejor al cliente y que las empresas puedan desarrollar campañas de ventas más eficaces, así como una mercadotecnia mejor dirigida y estrategias innovadoras de desarrollo de productos que resulten en mayores ingresos y rentabilidad.

## Conclusión

Actualmente el valor de la información se ha acrecentado hasta convertirse en un activo estratégico para la competitividad de una empresa. Su unidad y consistencia son importantes, pues de estas características depende una buena parte de la confiabilidad de la información seleccionada para tomar decisiones.

La minería de datos ayuda a los directivos a obtener una visión más completa y detallada de su negocio ya que les permite buscar datos de sus operaciones cotidianas que se salen de los rangos que están considerados como normales de lo que, en parte, depende la confiabilidad de la información para la toma de decisiones.

En la medida en la que una empresa capte datos de sus operaciones cotidianas tendrá la oportunidad de correlacionarlos y hacer descubrimientos que le ayuden a identificar posibles clientes, puntos de venta, fraudes, entre otros.

La minería de datos tiene futuro dentro de las empresas, debido a que existen grandes bases de datos que contienen valores desaprovechados; los mercados están más saturados y se requieren de análisis intensos para captar la atención de los clientes.

En todo el proceso de la minería de datos, el ser humano es el factor más importante, ya que sólo él tiene la capacidad de analizar y decidir si los patrones, normas o funciones encontrados tienen importancia, pertinencia y utilidad para su empresa.

## Bibliografía

B. MOXON'S, *Defining Data Mining DBMS Online*, August 1996.

VIZCAÍNO, Carlos, *Tecnologías de información y estrategias de negocios, Soluciones Avanzadas*, Núm. 34, 15 de Junio 1996.

GUZMAN ARENAS, Adolfo, *Tecnologías de información y estrategias de negocios, Soluciones Avanzadas*, Núm. 34, 15 de junio de 1996.

GONZÁLEZ LOZANO, Marisol, *Introducción a los sistemas de información ejecutiva y la evaluación de herramientas para su desarrollo, Soluciones Avanzadas*, Núm. 35, 15 de julio de 1996.

HARJINDER S., Gill y Rao Prakash C., *Data Warehousin. La integración para la mejor toma de decisiones*, Prentice Hall Hispanoamericana S. A., 1996, pp. 182, 240.

J. A. BERRY, Michael y Gordon Linoff, *Data Mining Techniques*, United States of America, John Wiley & Sons, Inc., 1997, pp. 17-35.