

## MINERÍA DE DATOS (DATAMINING)

### CONCEPTOS:



La minería de datos aparece en nosotros como una de las tecnologías para el manejo, ordenamiento y análisis de información que aprovecha la capacidad existente hoy día de procesamiento, almacenamiento y transmisión de datos a gran velocidad y bajo costo.

La minería de datos trabaja con los datos como materia prima a la que hay que transformar de su estado bruto a formas que expresen comportamientos, patrones, tendencias, tipos, clases, dimensiones, parámetro, secuencias, asociaciones, que puedan generar algún modelo que nos permita comprender mejor un sistema, o sea extraer la esencia contenida en las bases de datos, extraer el metal precioso, el oro de la montañas de datos, encontrar la veta de conocimiento escondido en su interior y entonces saber; de que, con que, cuando, como, de qué manera opera, como falla, como trabaja, como se desarrolla, como crece, como aumenta o disminuye la eficiencia y productividad de un sistema y entonces poder tomar decisiones mejor fundamentadas y de mayor consistencia validacional. La minería de datos se entiende como “Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos” (Fallad y otros 1996).

### AUTOR:

- Gloria Yáñez Gayosso

### FUENTE:

(Página 1)

[http://www.ingenieria.buap.mx/DOCUMENTOS/REVISTA/REV\\_11/art\\_4.pdf](http://www.ingenieria.buap.mx/DOCUMENTOS/REVISTA/REV_11/art_4.pdf)

La Minería de datos es la parte central del proceso KDD, en la cual se buscan o encuentran patrones de interés para el usuario. Los patrones descubiertos pueden ser subgrafos, reglas de asociación, árboles de clasificación, una red neuronal entrenada, entre otros.

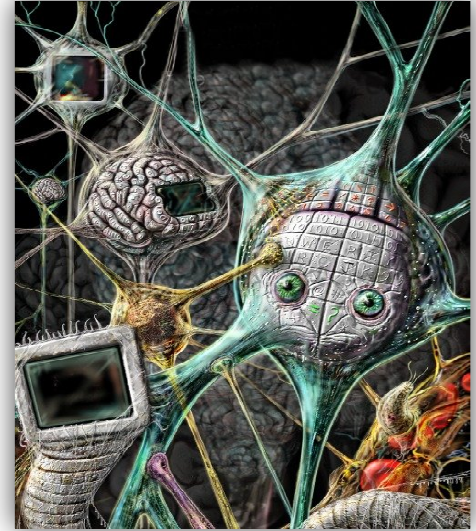
AUTORES:

- Ivan Olmos Pineda
- J. A. González Bernal.

FUENTE:

(Página 3)

<http://www.itpuebla.edu.mx/Eventos/MemoriasyResSemanaInformatica2007/02-%20IvanOlmosPineda.pdf>



La minería de datos o *data mining* (DM a partir de ahora). DM es un campo interdisciplinar con el objetivo general de predecir resultados y/o descubrir relaciones en los datos. DM puede ser descriptivo, descubrir patrones que describen los datos, o predictivo, para pronosticar el comportamiento del modelo basado en los datos disponibles.

Típicamente un algoritmo de DM tiene tres componentes: el modelo, el criterio de preferencia o elección y el algoritmo de búsqueda. El modelo con dos posibles tipologías según su función o su representación. En el primer caso puede ser de clasificación, regresión, clustering, de generación de reglas, reglas de asociación, modelos de dependencia o análisis de secuencias. Según su representación puede ser redes neuronales, árboles de decisión, discriminación lineal, etc. Cada modelo tiene unos parámetros que deben ser determinados mediante un algoritmo de búsqueda que optimiza los parámetros del modelo según el criterio de elección o preferencia que hace un mejor ajuste del modelo a los datos.

FUENTE:

<http://www.lsi.us.es/redmidas/>

### APLICACIÓN:

Las técnicas de minería de datos se están utilizando desde hace varios años para la obtención de patrones en los datos y para la extracción de información valiosa en el campo de la Ingeniería del Software...

Leer mas sobre el artículo leer la página 4 de este enlace --->

<http://www.sc.edu/jiwdocoj/remis/docs/minerw.pdf>

### CASOS:

En muchos cursos sobre minería de datos se cuenta una bonita historia sobre una gran cadena estadounidense de supermercados, Wal-Mart, que realizó a finales de los años 90 un análisis de los hábitos de compra de sus clientes.



Sorprendentemente, descubrieron una correlación estadísticamente significativa entre las compras de pañales y cerveza: los viernes por la tarde, los hombres entre 25 y 35 años que compraban cerveza también compraban pañales.

Después de un análisis detallado, este resultado se explica de forma bastante curiosa. Como los pañales son bastante voluminosos, las mujeres habitualmente mandaban a sus maridos a comprarlos. Los maridos y padres, jóvenes entre 25 y 35 años (rango medio de edad para tener niños tan pequeños), solían ir a la compra los viernes, algo reticentes, en el último momento posible. Estos pobres padres, con una vida social no demasiado boyante, a la vez que compraban pañales para sus bebés, aprovechaban para comprar cerveza, ya que no podrían salir a tomarlas al *pub*.

También se cuenta que Wal-Mart utilizó este resultado para reubicar estos productos en lugares estratégicamente dispuestos: pusieron la cerveza cerca de los pañales. El resultado fue que los padres que habitualmente compraban cerveza después compraron

todavía más, al estar tan cómodamente situada. Además, los que antes no compraban cerveza, empezaron a hacerlo al estar tan a mano, justo al lado a los pañales. Así, las ventas de cerveza tuvieron un aumento espectacular.

Este es un buen ejemplo de los beneficios que puede aportar la Minería de Datos y, en particular, el análisis de la cesta de la compra (*market basket analysis*)

El problema es que esta historia existe en diferentes versiones, a veces se habla de 7 Eleven en vez de Wal-Mart, otras veces se adorna con datos específicos de porcentajes de incremento de ventas... Además no se conoce quiénes pudieron realizar esos estudios y, de hecho, no existe documentación específica de ningún proyecto en esta línea en ninguna de las dos organizaciones.

Por todo ello, no hay más remedio que pensar que se trata de una leyenda urbana, un bonito mito del mundo de la minería de datos.

AUTOR:

- Elizabeth Gutierrez Kafati

FUENTE:

[http://www.google.com.pe/imgres?imgurl=http://www.paisdigital.org/images/fotos%2520educacion/collage%2520edunova.jpg&imgrefurl=http://egkafati.bligoo.com/tag/datamining&usq=\\_\\_g9su\\_r4J-xjNVGzXzK0AO5-TN0g=&h=417&w=517&sz=96&hl=es&start=33&zoom=1&um=1&itbs=1&tbnid=oE8A-lipotRaiM:&tbnh=106&tbnw=131&prev=/images%3Fq%3Dminero%2Bde%2Bdatos%26start%3D20%26um%3D1%26hl%3Des%26sa%3DN%26ndsp%3D20%26tbs%3Disch:1](http://www.google.com.pe/imgres?imgurl=http://www.paisdigital.org/images/fotos%2520educacion/collage%2520edunova.jpg&imgrefurl=http://egkafati.bligoo.com/tag/datamining&usq=__g9su_r4J-xjNVGzXzK0AO5-TN0g=&h=417&w=517&sz=96&hl=es&start=33&zoom=1&um=1&itbs=1&tbnid=oE8A-lipotRaiM:&tbnh=106&tbnw=131&prev=/images%3Fq%3Dminero%2Bde%2Bdatos%26start%3D20%26um%3D1%26hl%3Des%26sa%3DN%26ndsp%3D20%26tbs%3Disch:1)

OTROS CASOS:

Autor:

- Sergio Hurtado Camarasa

<http://www.it.uc3m.es/~jvillena/irc/practicass/03-04/21.pres.pdf>

## **Ejemplos de uso de la minería de datos**

### **Negocios**

La minería de datos puede contribuir significativamente en las aplicaciones de administración empresarial basada en la relación con el cliente. En lugar de contactar con el cliente de forma indiscriminada a través de un centro de llamadas o enviando cartas, sólo se contactará con aquellos que se perciba que tienen una mayor probabilidad de responder positivamente a una determinada oferta o promoción.

En lugar de crear modelos para predecir qué clientes pueden cambiar, la empresa podría construir modelos separados para cada región y/o para cada tipo de cliente. También puede querer determinar qué clientes van a ser rentables durante una ventana de tiempo (una quincena, un mes) y sólo enviar las ofertas a las personas que es probable que sean rentables. Para mantener esta cantidad de modelos, es necesario gestionar las versiones de cada modelo y pasar a una minería de datos lo más automatizada posible.

### **Hábitos de compra en supermercados**

El ejemplo clásico de aplicación de la minería de datos tiene que ver con la detección de hábitos de compra en supermercados. Un estudio muy citado detectó que los viernes había una cantidad inusualmente elevada de clientes que adquirirían a la vez pañales y cerveza. Se detectó que se debía a que dicho día solían acudir al supermercado padres jóvenes cuya perspectiva para el fin de semana consistía en quedarse en casa cuidando de su hijo y viendo la televisión con una cerveza en la mano. El supermercado pudo incrementar sus ventas de cerveza colocándolas próximas a los pañales para fomentar las ventas compulsivas.

### **Patrones de fuga**

Un ejemplo más habitual es el de la detección de patrones de fuga. En muchas industrias como la banca, las telecomunicaciones, etc. existe un comprensible interés en detectar cuanto antes aquellos clientes que puedan estar pensando en rescindir sus contratos para, posiblemente, pasarse a la competencia. A estos clientes y en función de su valor se les podrían hacer ofertas personalizadas, ofrecer promociones especiales, etc., con el

objetivo último de retenerlos. La minería de datos ayuda a determinar qué clientes son los más proclives a darse de baja estudiando sus patrones de comportamiento y comparándolos con muestras de clientes que, efectivamente, se dieron de baja en el pasado.

### **Fraudes**

Un caso análogo es el de la detección de transacciones de blanqueo de dinero o de fraude en el uso de tarjetas de crédito o de servicios de telefonía móvil e, incluso, en la relación de los contribuyentes con el fisco. Generalmente, estas operaciones fraudulentas o ilegales suelen seguir patrones característicos que permiten, con cierto grado de probabilidad, distinguirlos de las legítimas y desarrollar así mecanismos para tomar medidas rápidas frente a ellas.

### **Recursos humanos**

La minería de datos también puede ser útil para los departamentos de recursos humanos en la identificación de las características de sus empleados de mayor éxito. La información obtenida puede ayudar a la contratación de personal, centrándose en los esfuerzos de sus empleados y los resultados obtenidos por éstos. Además, la ayuda ofrecida por las aplicaciones para Dirección estratégica en una empresa se traducen en la obtención de ventajas a nivel corporativo, tales como mejorar el margen de beneficios o compartir objetivos; y en la mejora de las decisiones operativas, tales como desarrollo de planes de producción o gestión de mano de obra.

### **Comportamiento en Internet**

También es un área en boga el del análisis del comportamiento de los visitantes sobre todo, cuando son clientes potenciales en una página de Internet. O la utilización de la información obtenida por medios más o menos legítimos sobre ellos para ofrecerles propaganda adaptada específicamente a su perfil. O para, una vez que adquieren un determinado producto, saber inmediatamente qué otro ofrecerle teniendo en cuenta la información histórica disponible acerca de los clientes que han comprado el primero.

### **Terrorismo**

La minería de datos ha sido citada como el método por el cual la unidad Able Danger del Ejército de los EE.UU. había identificado al líder de los atentados del 11 de septiembre de 2001, Mohammed Atta, y a otros tres secuestradores del "11-S" como posibles miembros de una célula de Al Qaeda que operan en los EE.UU. más de un año antes del ataque. Se ha sugerido que tanto la Agencia Central de Inteligencia y sus homóloga canadiense, Servicio de Inteligencia y Seguridad Canadiense, también han empleado este método.

### **Juegos**

Desde comienzos de la década de 1960, con la disponibilidad de oráculos para determinados juegos combinatoriales, también llamados finales de juego de tablero (por ejemplo, para las tres en raya o en finales de ajedrez) con cualquier configuración de inicio, se ha abierto una nueva área en la minería de datos que consiste en la extracción de estrategias utilizadas por personas para estos oráculos. Los planteamientos actuales sobre reconocimiento de patrones, no parecen poder aplicarse con éxito al funcionamiento de estos oráculos. En su lugar, la producción de patrones *perspicaces* se basa en una amplia experimentación con bases de datos sobre esos finales de juego, combinado con un estudio intensivo de los propios **finales de juego** en problemas bien diseñados y con conocimiento de la técnica, Ejemplos notables de investigadores que trabajan en este campo son Berlekamp en el juego de puntos-y-cajas (o Timbiriche) y John Nunn en finales de ajedrez.

### **Genética**

En el estudio de la genética humana, el objetivo principal es entender la relación cartográfica entre las partes y la variación individual en las secuencias del ADN humano y la variabilidad en la susceptibilidad a las enfermedades. En términos más llanos, se trata de saber cómo los cambios en la secuencia de ADN de un individuo afectan al riesgo de desarrollar enfermedades comunes (como por ejemplo el cáncer). Esto es muy importante para ayudar a mejorar el diagnóstico, prevención y tratamiento de las enfermedades. La técnica de minería de datos que se utiliza para realizar esta tarea se conoce como "*reducción de dimensionalidad multifactorial*".

### **Ingeniería eléctrica**

En el ámbito de la ingeniería eléctrica, las técnicas minería de datos han sido ampliamente utilizadas para monitorizar las condiciones de las instalaciones de alta tensión. La finalidad de esta monitorización es obtener información valiosa sobre el estado del aislamiento de los equipos. Para la vigilancia de las vibraciones o el análisis de los cambios de carga en transformadores se utilizan ciertas técnicas para agrupación de datos (*clustering*) tales como los Mapas Auto-Organizativos (**SOM**, *Self-organizing map*). Estos mapas sirven para detectar condiciones anormales y para estimar la naturaleza de dichas anomalías.

### **Análisis de gases**

También se han aplicado técnicas de minería de datos para el análisis de gases disueltos (**DGA**, *Dissolved gas analysis*) en transformadores eléctricos. El análisis de gases disueltos se conoce desde hace mucho tiempo como herramienta para diagnosticar transformadores. Los Mapas Auto-Organizativos (**SOM**) se utilizan para analizar datos y determinar tendencias que podrían pasarse por alto utilizando las técnicas clásicas **DGA**.

### **COMENTARIO PERSONAL:**

Este tema sobre el DATA MINING o Minería de Datos nos ayuda a facilitar el ordenamiento de nuestra información en una manera más rápida y segura.