



Universidad Nacional Experimental Politécnica
"Antonio José De Sucre"
Vice-rectorado Puerto Ordaz
Departamento de Ingeniería Industrial
Sistemas de Información

DATA MINING Y EL VOLUMEN DE DATOS



Profesor:
MSc. Ing. Iván Turmero

Integrantes:
Castillo, Xiorle
García, Vanessa
Gil, Karlenys
Marcantonio, Patricia
Muñoz, Anick
Torrealba, Pedro

CIUDAD GUAYANA, MARZO DE 2010

INTRODUCCIÓN

El Data Mining es una de las principales herramientas que se utilizan dentro de los programas de gestión del conocimiento como soporte a la toma de decisiones. El fin es la extracción de información oculta o análisis de datos mediante técnicas estadísticas de grandes bases de datos. Las herramientas de data Mining o minería de datos pueden responder a preguntas de negocios empresariales a priori no planteadas o que pueden consumir demasiado tiempo para ser resueltas.

Las técnicas de data Mining se centran en analizar el gran volumen de datos.

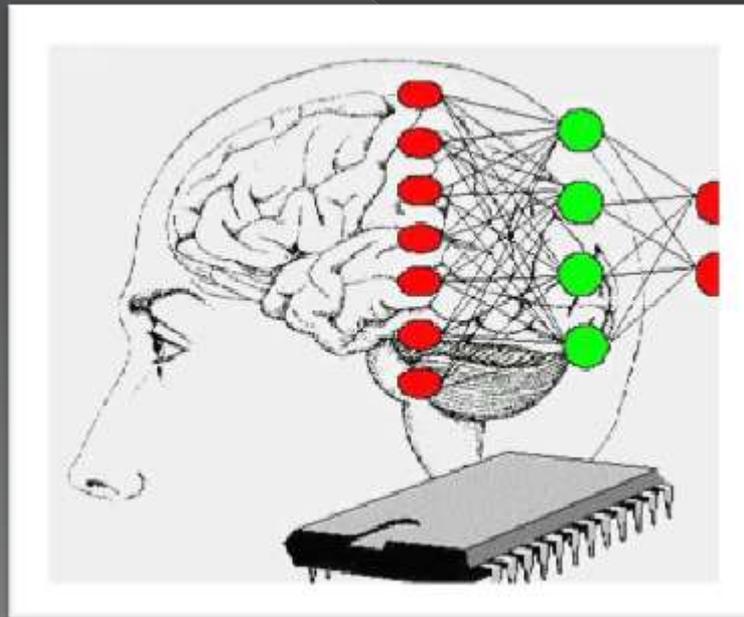
En definitiva, la minería de datos es una tecnología usada para descubrir información oculta y desconocida, pero potencialmente útil, a partir de las fuentes de información de la propia empresa. Obtiene un conocimiento de un negocio, utilizando técnicas de clustering, redes neuronales, árboles de decisión y reglas de asociación etc.

Minería de datos



Minería de datos

La minería de datos hace uso de todas las técnicas que puedan aportar información útil, desde un sencillo análisis gráfico, pasando por métodos estadísticos más o menos complejos, complementados con métodos y algoritmos del campo de la inteligencia artificial y el aprendizaje automático que resuelven problemas típicos de agrupamiento automático, clasificación, predicción de valores, detección de patrones, asociación de atributos, etc. Es, por tanto, un campo multidisciplinar que cubre numerosas áreas y se aborda desde múltiples puntos de vista, como la estadística, la informática (cálculo automático) o la ingeniería.



PROCESO

Un proceso típico de minería de datos consta de los siguientes pasos generales:

- Selección del conjunto de datos, tanto en lo que se refiere a las variables dependientes, como a las variables independientes, como objetivo, como posiblemente al muestreo de los registros disponibles.

1



- Análisis de las propiedades de los datos, en especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos (valores nulos).

2



- Transformación del conjunto de datos de entrada, se realizará de diversas formas en función del análisis previo, con el objetivo de prepararlo para aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema.

3



PROCESO

• Seleccionar y aplicar la técnica de minería de datos, se construye el modelo predictivo, de clasificación o segmentación.

4



• Extracción de conocimiento, mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables.

5

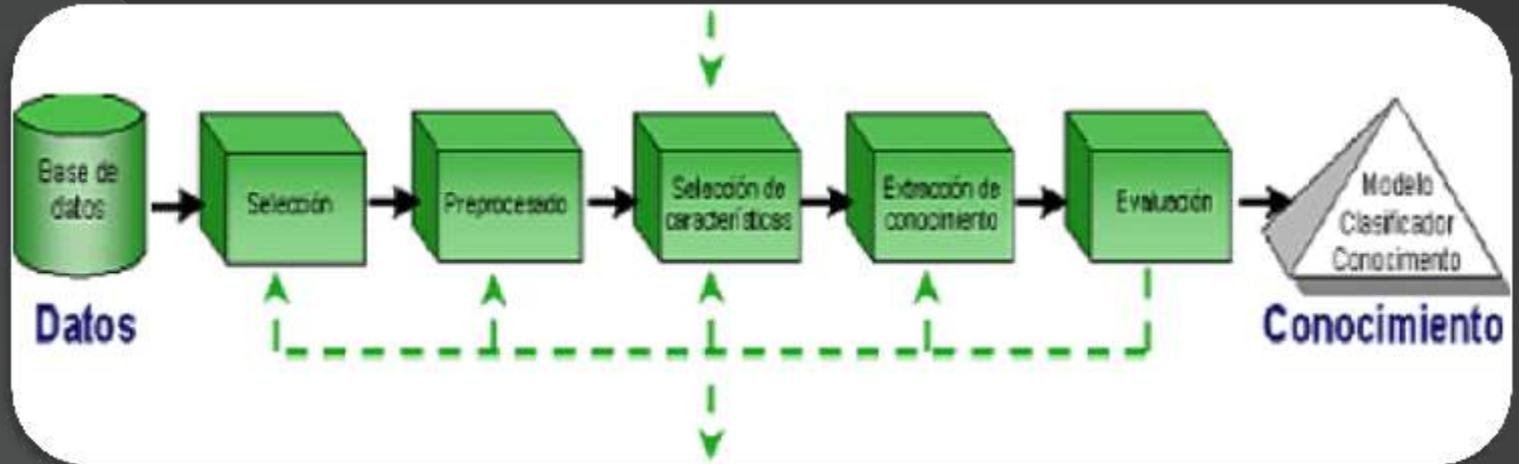


• Interpretación y evaluación de datos, una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias.

6



PROCESO



Si el modelo final no superara la evaluación, el proceso se podría repetir desde el principio o, si el experto lo considera oportuno, a partir de cualquiera de los pasos anteriores. Esta retroalimentación se podrá repetir cuantas veces se considere necesario hasta obtener un modelo válido.

Una vez validado el modelo, si resulta ser aceptable (proporciona salidas adecuadas y/o con márgenes de error admisibles) éste ya está listo para su explotación.

PROTOCOLO DE UN PROYECTO DE MINERÍA DE DATOS

Un proyecto de minería de datos tiene varias fases necesarias que son, esencialmente:

Comprensión del negocio y del problema que se quiere resolver.

Determinación, obtención y limpieza de los datos necesarios.

Creación de modelos matemáticos.

Validación, comunicación, etc. de los resultados obtenidos.

Integración, si procede, de los resultados en un sistema transaccional o similar.

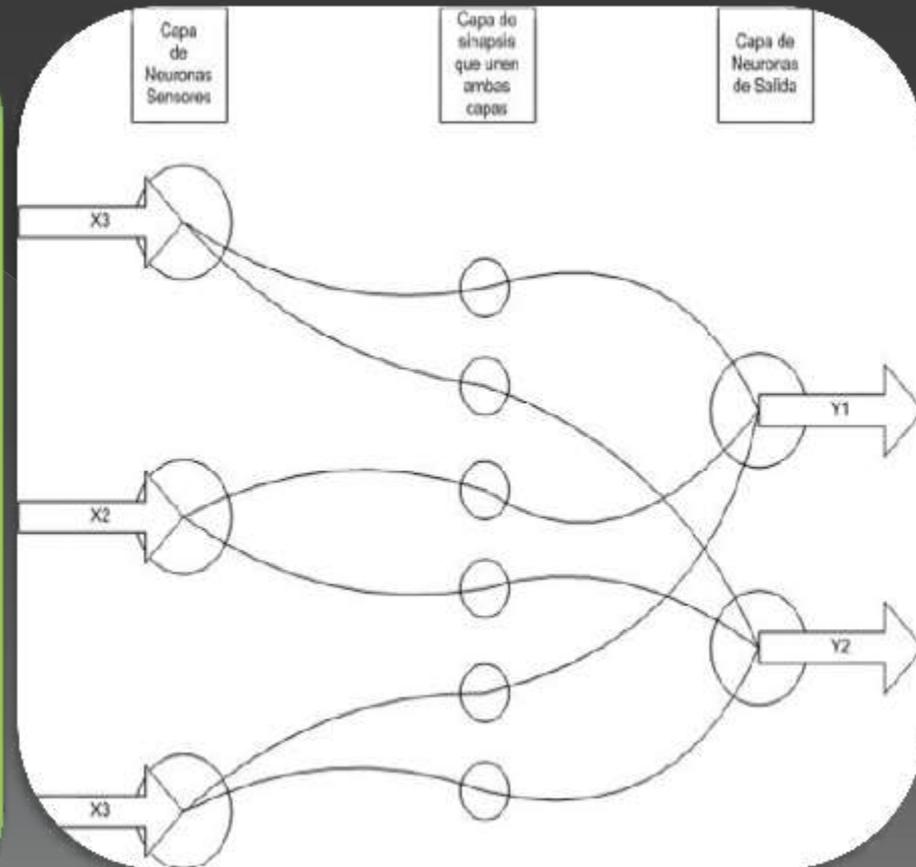
TÉCNICAS DE MINERÍA DE DATOS

Las técnicas más representativas son

1

Redes neuronales

Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida.



El Perceptrón

TÉCNICAS DE MINERÍA DE DATOS

2

Regresión lineal

Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.

3

Árboles de decisión

Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.

4

Modelos estadísticos

Es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales y en la regresión para indicar los diferentes factores que modifican la variable de respuesta

5

Agrupamiento o Clustering

Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes.



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.

HÁBITOS DE COMPRA EN SUPERMERCADOS



Un estudio muy citado detectó que los viernes había una cantidad inusualmente elevada de clientes que adquirían 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.

EJEMPLOS DE USO DE LA MINERÍA DE DATOS

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.

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, distinguirlas de las legítimas y desarrollar así mecanismos para tomar medidas rápidas frente a ellas.

EJEMPLOS DE USO DE LA MINERÍA DE DATOS

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.

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.

EJEMPLOS DE USO DE LA MINERÍA DE DATOS

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.

JUEGOS



Esta nueva área en la minería de datos consiste en la extracción de estrategias utilizadas por personas para los oráculos para determinados juegos combinatorios. Los planteamientos actuales sobre reconocimiento de patrones, no parecen poder aplicarse con éxito al funcionamiento de estos oráculos.

EJEMPLOS DE USO DE LA MINERÍA DE DATOS

GENÉTICA



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".

ANÁLISIS DE GASES



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.

EJEMPLOS DE USO DE LA MINERÍA DE DATOS

INGENIERIA 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).

Minería de datos y otras disciplinas análogas

DE LA ESTADÍSTICA

Análisis de varianza

Regresión

Prueba Chi-cuadrado

Análisis de agrupamiento o clustering

Análisis discriminante

Series de tiempo

EJEMPLOS DE USO DE LA MINERÍA DE DATOS

DE LA INFORMÁTICA

Algoritmos genéticos

Inteligencia Artificial

Sistemas Expertos

Sistemas Inteligentes

Redes neuronales

TENDENCIAS

La Minería de Datos ha sufrido transformaciones en los últimos años de acuerdo con cambios tecnológicos, de estrategias de marketing, la extensión de los modelos de compra en línea, etc. Los más importantes de ellos son:

1

La importancia que han cobrado los datos no estructurados (texto, páginas de Internet, etc.)

2

La necesidad de integrar los algoritmos y resultados obtenidos en sistemas operacionales, portales de Internet, etc.

3

La exigencia de que los procesos funcionen prácticamente en línea (por ejemplo, que frente a un fraude con una tarjeta de crédito).

4

Los tiempos de respuesta. El gran volumen de datos que hay que procesar en muchos casos para obtener un modelo válido es un inconveniente; esto implica grandes cantidades de tiempo de proceso y hay problemas que requieren una respuesta en tiempo real.

CONCLUSIONES

El Data Mining (minería de datos), es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

En resumen, el Data Mining se presenta como una tecnología emergente, con varias ventajas: por un lado, resulta un buen punto de encuentro entre los investigadores y las personas de negocios; por otro, ahorra grandes cantidades de dinero a una empresa y abre nuevas oportunidades de negocios. Además, no hay duda de que trabajar con esta tecnología implica cuidar un sinnúmero de detalles debido a que el producto final involucra "toma de decisiones".

