

## TOMA DE DECISIONES CON DATA MINING

Ing. Jesús Salinas Flores

### RESUMEN

La información se ha convertido en algo cada vez más importante para nosotros. El presente estudio muestra una introducción al concepto de Data Mining como una herramienta a la que muchas empresas y organizaciones recurren para mejorar sus procesos de toma de decisiones.

### ABSTRACT

The information has been converted into one of the most important for us. The present study shows the concept an introduction to Data Mining as a tool which many companies and organizations to make use to improve the process of make decisions.

*"El problema ahora es que en una empresa por lo general se tiene gran cantidad de información almacenada en diferentes lugares ..."*

### INTRODUCCIÓN

Bill Gates en su libro "Camino al Futuro" da una visión sobre las implicaciones de la información en el mundo de los negocios. " La información se ha convertido en algo cada vez más importante para nosotros. La revolución de la información está en sus comienzos. El costo de las comunicaciones disminuirá tan rápidamente como lo ha hecho el costo de la informática. .... Las empresas que continúen confiando exclusivamente en estos métodos de toma de decisiones se arriesgan a perder ante los competidores que toman decisiones más rápidamente, al tiempo que dedican al proceso menos recursos y probablemente menos estratos de gestores." .

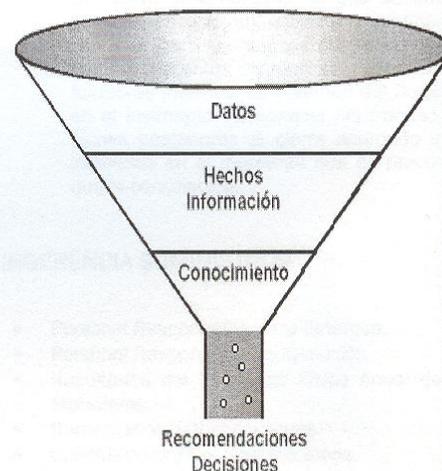
En estos tiempos, en que los negocios se hicieron más competitivos y complejos, los datos cada vez cobraron más importancia y se convirtieron en información vital para la toma de decisiones de los gerentes. El problema ahora es que en una empresa por lo general se tiene gran cantidad de información almacenada en diferentes lugares: un data warehouse, un data mart, una base de datos o información adquiridos a terceros. El Data mining es la forma más avanzada de extraer información y patrones de comportamiento, ocultos entre grandes cantidades de información; esto es; permite transformar la información disponible en conocimiento útil para la toma de decisiones.

### CONCEPTO

*Data Mining*, (minería de datos o manejo de datos) es el proceso de extracción de información significativa de grandes bases de datos, información que revela inteligencia del negocio a través de patrones ocultos, tendencias y correlaciones para permitir al usuario realizar predicciones que resuelven problemas del negocio proporcionando una ventaja competitiva.

Las herramientas de Data Mining predicen las nuevas perspectivas y pronostican la situación futura de las empresas, esto ayuda a las mismas a tomar decisiones de negocio proactivamente. Las técnicas de Data Mining están listas para su aplicación en los negocios ya que está soportada por tres tecnologías que ahora tienen el suficiente desarrollo:

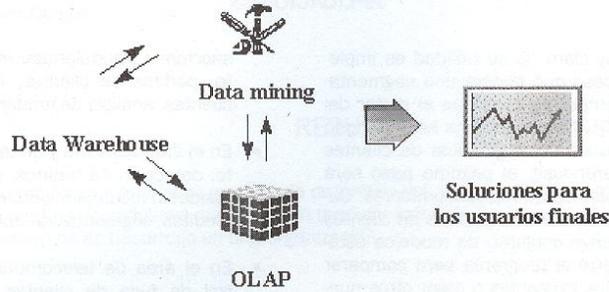
- Recolección masiva de datos
- El poder de las computadoras multiprocesadoras.
- Los algoritmos de Data Mining.



## ARQUITECTURA

El Data Mining debe estar integrada tanto con el Data Warehouse como con herramientas de análisis flexibles; muchas de las herramientas de Data Mining que funcionan fuera del Data

Warehouse requieren pasos adicionales para la extracción, importación y análisis de los datos. La estructura ideal sería la siguiente:



**Data Warehouse** (Bodega de Datos) es una arquitectura que involucra un proceso continuo que mezcla los datos de varias fuentes heterogéneas, incluye datos históricos y adquiridos para soportar la constante necesidad de consultas estructuradas y/o ad hoc, reportes para el soporte de decisiones. El Data Warehouse utiliza la tecnología asignada para el manejo corporativo de las bases de datos (por ejemplo: Oracle, Sybase, MS SQL Server).

El procesamiento analítico en línea (On-Line Processing-OLAP) se refiere a la tecnología que permite a los usuarios de bases de datos generar resúmenes comparativos o descriptivos en línea de datos y otras interrogantes analíticas; es decir OLAP describe la tecnología asociada al acceso y análisis de datos en línea. Las técnicas OLAP pueden ser muy simples (tablas de frecuencia, estadística descriptiva, tabulaciones

cruzadas simples, etc.) o muy complejas (como ajustes estacionales). El Data Mining puede operar sobre cualquier clase de información sin procesar o incluso, no estructurado, también se pueden aplicar a datos y resúmenes generados por OLAP para proporcionar un conocimiento más profundo.

Los Data Warehouse almacenan datos pero carecen de las herramientas para analizarlos. Los productos OLAP responden a cuestiones del tipo ¿Qué?, como por ejemplo, qué región vende más, qué clientes son los más rentables, etc. Mientras que las técnicas de Data Mining responden a cuestiones estratégicas como ¿por qué las ventas están bajando?, ¿Por qué los clientes se van a otras compañías?, es decir, nos proporcionan información, conocimiento valioso con el que puede diseñar una política de ventas, marketing, etc.

*“El Data mining es la forma más avanzada de extraer información y patrones de comportamiento, ocultos entre grandes cantidades de información ..”*

## METODOLOGÍA DEL TRABAJO

La técnica con que trabaja el Data Mining se denomina “modelamiento”; el modelamiento es el acto de desarrollar un modelo de una situación donde ya se conocen las respuestas y aplicarlo luego en otra situación desconocida; es decir utilizando experiencias anteriores se obtienen patrones de comportamiento que pueden ser utilizados para luego realizar un modelo que permita pronosticar y predecir.

El proceso de Data Mining presenta tres etapas:  
 La exploración de la data o la información  
 La construcción del modelo  
 La validación/verificación del modelo

Entre las técnicas más utilizadas en el proceso de Data Mining podemos destacar las técnicas exploratorias de análisis gráfico y redes neuronales.

Las *técnicas exploratorias* incluyen técnicas como la revisión de la distribución de variable, la revisión de matrices de correlación, examinar tablas de frecuencias de múltiples entrada y técnicas para identificar modelos de datos multivariados incluyendo entre otros: análisis cluster, análisis de factores, análisis de series de tiempo, correlación canónica, análisis de correspondencia y árboles de decisión.

Las *redes neurales* son técnicas analíticas modeladas después del proceso hipotético de aprender el sistema cognoscitivo y las funciones neurológicas del cerebro y capaz de predecir nuevas observaciones a partir de otras observaciones después de ejecutar un proceso de

aprendizaje de los datos existentes. Una de sus principales ventajas es que, teóricamente, pueden aproximar cualquier función continua y por lo tanto, el investigador no requiere plantear ninguna hipótesis sobre el modelo actual.

## APLICACIONES

Un ejemplo muy claro de su utilidad es implementar un proceso que genere una segmentación de los clientes. Una vez que el motor del Data Mining (SQL, Oracle, etc.) ha seleccionado un grupo adecuado de segmentos de clientes de su Data Warehouse, el próximo paso será extrapolar los perfiles de los consumidores. Cada vez que llega una nueva cosecha de clientes se aplica un nuevo conjunto de modelos estadísticos y se corre el programa para comparar contra segmentos existentes o crear otros nuevos. Otras aplicaciones son:

- En el área de Marketing: segmentación de mercados, descubrimientos de tran-

sacciones fraudulentas, mercadeo directo, perfiles de clientes, focalización de clientes, análisis de tendencias.

- En el área bancaria y de tarjetas de crédito: detección de fraudes, análisis de morosidad o incumplimiento de pagos en los créditos, segmentación del mercado.
- En el área de telecomunicaciones: control de fuga de clientes, detección de fraudes, ventas cruzadas.
- En el área de seguros: análisis de riesgos, detección de fraudes, mercadeo.

*“El Data Mining permite transformar la información disponible en conocimiento útil para la toma de decisiones.”*

## BIBLIOGRAFÍA

- Data Mining: A Hands-On Approach for Business Professionals (Data Warehousing Institute Series) Robert Groth. Paperback (1997)
- Data Mining: Technologies, Techniques, Tools, and Trends. Bhavani Thuraisingham. (1998).
- Data Mining Solutions: Methods and Tools for Solving Real-World Problems Christopher Westphal, Teresa Blaxton. (1998).
- Data Mining: Building Competitive Advantage Robert Groth. (1999).
- El Data Warehouse. El Data Mining. Franco, Jean Michel. (1997).

Internet:

- El Data Mining con SPSS. <http://www.spss.es/ofertas/data-minig2.asp>
- Data Mining con STATISTICA. <http://www.statsoftinc.com/espana/datamine.html>
- El Data Mining vs. Estadística. [http://www.at-systems.es/datamining/datamining\\_estadistica.html](http://www.at-systems.es/datamining/datamining_estadistica.html)
- Data Mining con OBECAN. <http://www.obecan.ceyas.rcanaria.es/garoe/metodologia.html>
- Del Query al Data Mining. <http://www.swgreenhouse.com/hsdatmrt.html>