UADE

Sistemas de soporte a las decisiones

2do cuatrimestre 2009

Ing. Silvia Cobialca
Lic. Cecilia Ruz
Lic. Néstor Martínez

UADE

Preguntas y respuestas

Preguntas:

- ¿Cuál es el valor de los clientes?
- ¿Cuáles son los clientes que tienen mayor probabilidad de desertar?
- ¿Cuáles son los productos que se venden en forma conjunta? ….

Respuestas:

- Están en los datos del usuario
- Se necesitan herramientas especiales para encontrarlas

UADE

Principios básicos

- Los negocios no existen en forma aislada
- El éxito de un negocio no depende solamente de cómo se desarrolle sino también de cómo le va a la competencia
- Conocer información sobre el propio negocio permite actuar pro-activamente en lugar de re-activamente

UADE

Business Intelligence

- "Es un paraguas bajo el que se incluye un conjunto de conceptos y metodologías cuya misión consiste en mejorar el proceso de toma de decisiones en los negocios basándose en hechos y sistemas que trabajan con hechos"

Howard Dresner
(Gartner Group), 1989

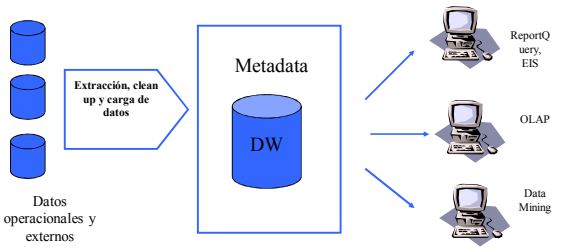
UADE

B.I.: recursos y herramientas

- Fuentes de datos : warehouses, data marts, etc
- Herramientas de administración de datos
- Herramientas de extracción y consulta
- Herramientas de modelización (Data Mining)

UADE

Las herramientas



UADE

Perspectiva histórica

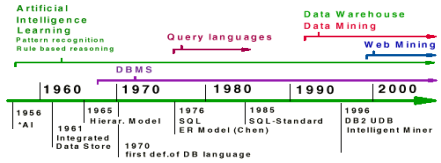
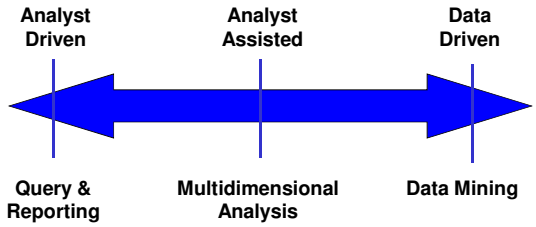


Figure 3-1 A historical view of data mining

Fuente: Mining your Own Business Data Using DB2 Intelligent Miner for Data

UADE

Knowledge discovery



UADE

Evolución desde los querys hasta la Minería de Datos

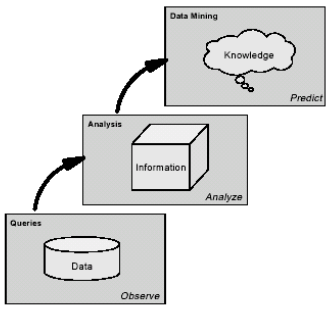


Figure 1. Evolution from Queries to Data Mining

UADE

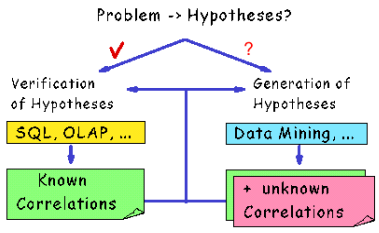
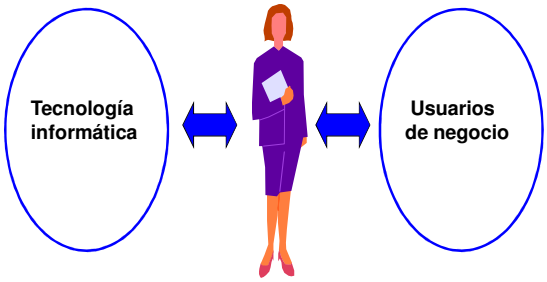


Figure 3-2 Standard and data mining approach on information detection

Fuente: Mining your Own Business Data Using DB2 Intelligent Miner for Data

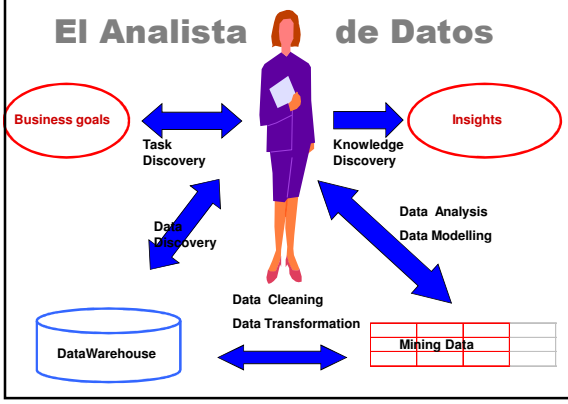
UADE

El Analista de Datos



UADE

El Analista de Datos



UADE

El Analista de Datos

- Es el vínculo entre las áreas de tecnología informática y las áreas de negocios
- Traduce los requerimientos de información en preguntas apropiadas para su análisis con las herramientas de minería.
- Realimenta el Data Warehouse de la compañía con nuevos criterios de data cleaning y data validation.

UADE

Habilidades requeridas

- Data manipulation (SQL)
- Conocimiento de las técnicas de minería y análisis exploratorio
- Habilidad de comunicación (interpretación) de los problemas de negocio
- Creatividad

UADE

Data Warehouse

- Data Warehouse es una colección de datos integrada, no volátil, que varía en el tiempo y organizada a partir de un tema para soportar decisiones de negocio

  Bill Inmon (1990)
- Es una copia de los datos transaccionales específicamente orientada para querys y análisis

  Ralph Kimball

UADE

Data Marts

- Técnicamente es un subconjunto del DW orientado a una finalidad específica de negocio : marketing, finanzas, producción, etc
- El término se utiliza también para identificar soluciones alternativas a un DW corporativo más reducidas y de menor costo y tiempo de implantación.

UADE

Herramientas de explotación del DW

- Herramientas de visualización
- Reporting
- OLAP
- Data Mining

UADE

OLAP

- On Line Analytical Processing
- Permiten la elaboración de vistas multidimensionales del DW para optimizar performance
- Están soportadas por motores de administración del DW que admiten la construcción de estos "cubos"

UADE

OLAP

- Herramientas útiles y poderosas para acceder a Bases de Datos y Data Warehouses y obtener "reportes" de información.
- La tecnología OLAP commplementa las actividades de Data Mining y supera las posibilidades del SQL

UADE

Data Mining (1997)

➔ **Data Mining** : es el proceso de exploración y análisis - de manera automática o semiautomática - de los datos para obtener patrones significativos y reglas de negocio.

➔ Michael Berry, Gordon Linoff
Data Mining for marketing, sales and customer support.
Wiley. USA, 1997

UADE

Data Mining (2000)

- .. nos gusta la noción de que los patrones deben ser significativos …
- .. Si hay algo que rechazamos es la frase "por medios automáticos o semiautomáticos", no porque no sea cierto - sin automatización es imposible minar grandes cantidades de datos - sino porque entendemos que se ha puesto demasiado énfasis en la automatización y no suficiente en las etapas de exploración y análisis
- .. Data Mining es un proceso

Michael Berry, Gordon Linoff
Mastering Data Mining. Wiley. USA 2000

UADE

Qué NO es Data Mining

- No es un producto que se compra enlatado sino una disciplina que debe ser dominada.
- No es una solución instantánea a los problemas de negocio.
- No es un fin en sí mismo sino un proceso que ayuda a encontrar soluciones a problemas de negocio.

UADE

Pilares del proceso de Data Mining

- Datos
- Algoritmos y técnicas
- Prácticas de modelización.

UADE

Disciplinas que se integran

- Inteligencia Artificial
- Estadística
- Tecnologías de soporte de decisiones (OLTP)
- Tecnologías de hardware y software

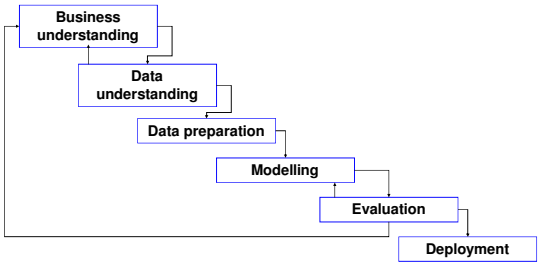
UADE

Etapas en el proceso de Data Mining

- Identificar el problema de negocio
- Transformar los datos en información
- Actuar a partir de los resultados
- Medir los resultados de las acciones

UADE

Un proyecto genérico



UADE

Data Mining y OLAP

- Las herramientas de reporting, OLAP y consulta responden efectivamente para la construcción de modelos descriptivos y retrospectivos para confirmar o rechazar hipótesis previas del usuario

UADE

Data Mining y OLAP

- Las herramientas de Data Mining permiten encontrar patrones no evidentes en los grandes volúmenes de información del DW y proponer modelos predictivos

UADE

Qué es la Estadística

- Es la disciplina que extrae información general a partir de datos específicos.
- Es el estudio de la estabilidad en la variación
- Es el arte de examinar, sumarizar y extraer conclusiones a partir de los datos.

UADE

Data Mining y Estadística

- Los métodos estadísticos son el corazón de muchas de las técnicas de minería de datos.
- Originalmente muchas de estas técnicas fueron diseñadas con propósitos confirmatorios.
- La estadística exploratoria aparece en los 70 con los aportes de J.Tuckey

UADE

Data Mining y Estadística

- En la Minería de Datos no se hacen supuestos a priori sobre la naturaleza de las variables y de las relaciones entre ellas (normalidad, linealidad, etc).
- Los algoritmos estadísticos se adaptan - para Minería de Datos - al procesamiento de grandes volúmenes de datos

UADE

Data Mining e IA

- La Inteligencia Artificial se integra a la Minería de Datos a partir de las redes neuronales artificiales
- Se utilizan para construir modelos predictivos no lineales que aprenden a través de entrenamiento y que se asimilan a los modelos de redes de neuronas biológicas.

UADE

Redes neuronales

- Las redes neuronales son adecuadas para problemas de tipo predictivo.
- Un problema apropiado para una red neuronal tiene tres características:
  - Se comprenden claramente los INPUTS
  - Se comprende claramente el OUTPUT
  - Existen ejemplos (experiencia) suficientes para entrenar a la red

UADE

Los modelos neuronales

- La red neuronal no produce reglas explícitas que describan el modelo
- Un modelo neuronal es tan bueno como lo es el set de datos usado para entrenar la red
- El modelo es estático y debe ser explícitamente actualizado agregando ejemplos recientes y re-entrenando la red para asegurar su vigencia y utilidad

UADE

Los modelos neuronales

- Con modelos neuronales se puede atacar una gran variedad de problemas y producir buenos resultados aún en dominios complejos con variables continuas y categóricas
- Son apropiados para tareas de clasificación y predicción cuando los resultados del modelo son más importantes que comprender cómo funciona el modelo.

UADE

Customer Relationship Management

- Es el proceso que administra la relación entre la compañía y sus clientes
- Para que resulte exitoso resulta necesario identificar los patrones de consumo y comportamiento de los clientes

UADE

Data Mining - CRM

- Data Mining se utiliza para sistematizar los procesos de búsqueda de los predictores de comportamiento de los clientes en las etapas de diseño de campañas
- También se aplica para la medición de los resultados de la campaña y la realimentación del CRM

UADE

Problemas típicos de Data Mining

- Clasificación
- Estimación
- Predicción
- Agrupamiento a partir de reglas de asociación
- Clustering
- Descripción y visualización, etc

UADE

Problemas usuales

- Caracterización de perfiles de clientes para definir acciones de Up selling y Cross selling
- Tracking de campañas y predicción de respuesta / no respuesta
- Canasta de consumo de tarjetas de crédito y prevención de fraudes
- Modelos de predicción de abandono

UADE

Problemas usuales

- Programas de millaje y fidelización
- Consolidación de Bases de Datos propias con fuentes externas
- Web mining y análisis de tráfico y uso de recursos de e-business
- Definición de marcos muestrales para investigaciones de mercado y encuestas de customer satisfaction.

UADE

La elección del modelo para Data Mining

➔Principales objetivos del proceso de Data Mining
- predicción
- descripción

➔El método a utilizar depende de los objetivos perseguidos por el análisis pero también de la calidad y cantidad de los datos disponibles

UADE

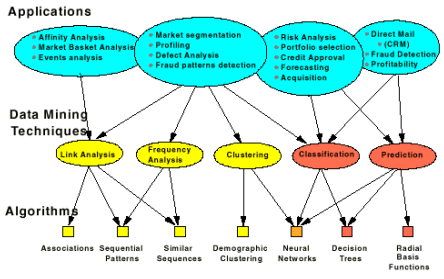


Figure 3-3 From applications to algorithms

Fuente: Mining your Own Business Data Using DB2 Intelligent Miner for Data

UADE

La evaluación de los modelos

- Cuán ajustado es el modelo?
- Es correcta su descripción de los datos observados?
- Cuanta confianza se puede tener en sus predicciones?
- Cuán comprensible es el modelo?

UADE

Las medidas

- La concordancia de un modelo predictivo con la realidad se mide con relación a la tasa de error, es decir, el porcentaje de casos clasificados o cuya predicción fue incorrecta.
- Para ello se dispone de datos de validación y testing sobre los que debe aplicarse periódicamente el modelo a modo de control.

UADE

Las medidas

- En el caso de los modelos descriptivos, una buena regla es la que proporciona la información más comprensible con la menor "longitud" de expresión de la regla.
- En definitiva, la medida más importante de efectividad es el retorno de la inversión

UADE

Un proyecto exitoso

- Un único project leader
- Un equipo multidisciplinario integrado por personas de las áreas de IT y de negocio
- Las unidades de negocio están involucradas desde el comienzo
- El área de IT está involucrada desde el comienzo
- Un pequeño proyecto piloto que muestre las ventajas de Data Mining

UADE

Para qué Minería de Datos

- La Minería de Datos es una herramienta eficaz para dar respuesta preguntas complejas de Inteligencia de Negocios
- Las herramientas disponibles permiten automatizar parte de la tarea de encontrar los patrones de comportamiento ocultos en los datos
- Pero ….

UADE

Qué no puede automatizarse (todavía)

- La elección de los problemas de negocio candidatos para tareas de Data Mining
- La identificación y recolección de los datos que contienen la información buscada
- El masajeo y tratamiento de los datos que posibilita la búsqueda de patrones
- El diseño y cálculo de variables derivadas

UADE

Qué no puede automatizarse (todavía)

- El plan de acciones que apoyándose en los resultados del modelo produzca el ROI
- La medición del éxito de las acciones realizadas a partir de los resultados proporcionados por Data Mining

UADE

Contenidos de la materia

- Introducción al Análisis de Datos
- Elementos de diseño de repositorios de datos para BI (Datawarehouses, Data Marts)
- Introducción a las herramientas OLAP
- Introducción a las herramientas de Minería de datos