



La Minería de Datos y el Negocio:

Aplicaciones, Metodologías y
Técnicas

María Esther Ordóñez O. - 2008

 **ACIS** **XXVIII Salón de Informática**
LOS DATOS: Materia prima de la información y
real valor de las organizaciones.





Agenda

- Minería de Datos en el contexto de BI y DWH
- Aplicaciones de Minería de Datos
- Principales Tareas y Técnicas
- Metodología para procesos de Minería





Definición de BI

Conjunto de Tecnologías y Aplicaciones que permiten Recopilar, Almacenar, Analizar y Tener Acceso a datos, de tal manera que los usuarios de la Organización pueden tomar mejores decisiones .

Transformar los datos operacionales de una empresa en información “accionable”





Definición de BI

Información accionable

Información que realmente habilita y optimiza el proceso de toma de decisiones y la definición de estrategias y acciones encaminadas a mejorar el desempeño del negocio



Fundamento de BI



BOGEGAS DE DATOS

 **ACIS** **XXVIII** *Salón de Informática*
LOS DATOS: Materia prima de la información y
real valor de las organizaciones.



Sistemas Fuente (Source)

Area de Preparación de Datos - ETL (Data Staging Area)

Servidor de Presentación BODEGA DE DATOS

Usuario Final Acceso a Datos



Ralph Kimball

Definición de Minería de Datos

Exploración y Análisis de grandes cantidades de datos, utilizando métodos automáticos o semi-automáticos, con el objeto de descubrir patrones significativos y reglas de comportamiento



Aspectos a tener en cuenta....

- DM tiene a su alrededor elementos éticos y legales
- Pueden verse como métodos de discriminación
- La razón para un rechazo no puede ser “...la red neuronal dijo que NO!!!!”
- Condiciones bajo las cuales se recoge la información - las personas deben ser informadas de los objetivos del proceso



Por qué es Factible HOY

- Los datos necesarios se están produciendo y recopilando en DWH
- La potencia computacional está al alcance (\$)
- CRM es importante en el negocio
 - Todo negocio es un negocio de servicios
 - La información se percibe como un producto
- Disponibilidad de productos comerciales





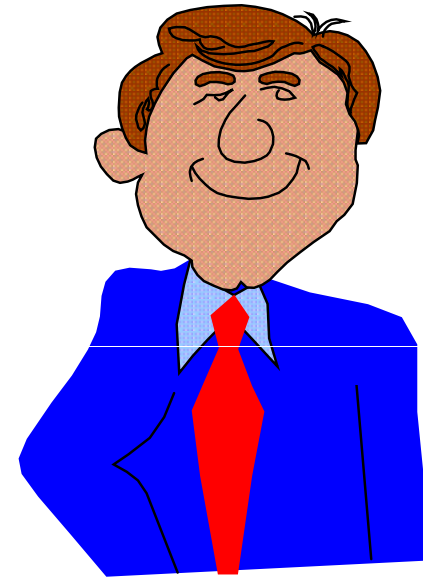
Agenda

- Minería de Datos en el contexto de BI y DWH
- Aplicaciones de Minería de Datos
- Principales Tareas y Técnicas
- Metodología para procesos de Minería



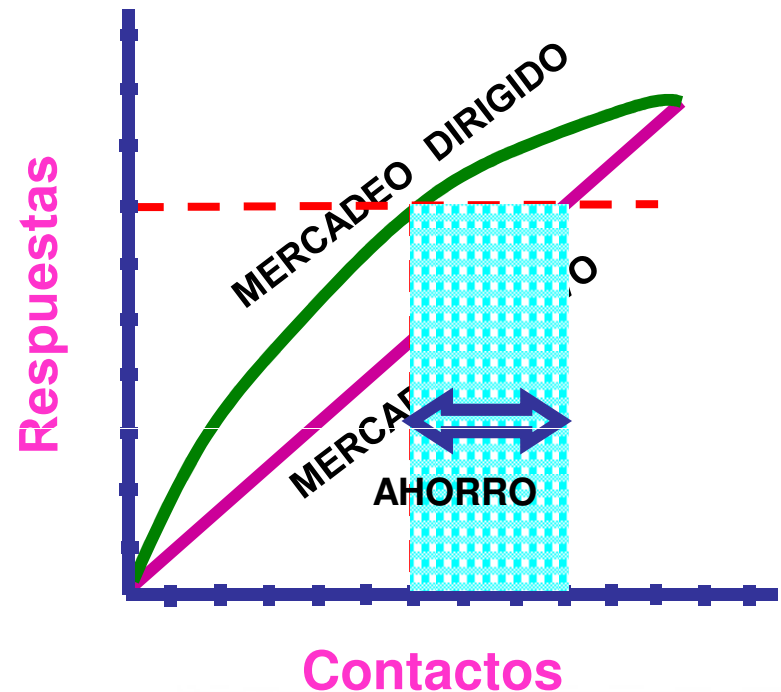
Aplicaciones Generales

- Conocimiento y Fidelización de clientes
 - Segmentación
 - Cross-sell y Up-sell
 - Manejo del ciclo de vida del cliente
 - Conocimiento Comunitario



Aplicaciones Generales

- Optimización de Campañas de Mercadeo
 - Mercadeo Masivo vs. Mercadeo Dirigido



Aplicaciones Generales

- **Detección de Fraudes**
 - Identificación de patrones de comportamiento normal / fraudulento



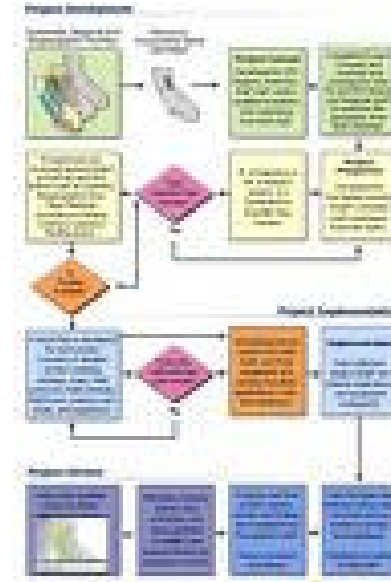
Aplicaciones Generales

- Manejo de deserción de clientes (churn – attrition)
 - Identificar potenciales desertores para aplicar estrategias de retención (si lo amerita)



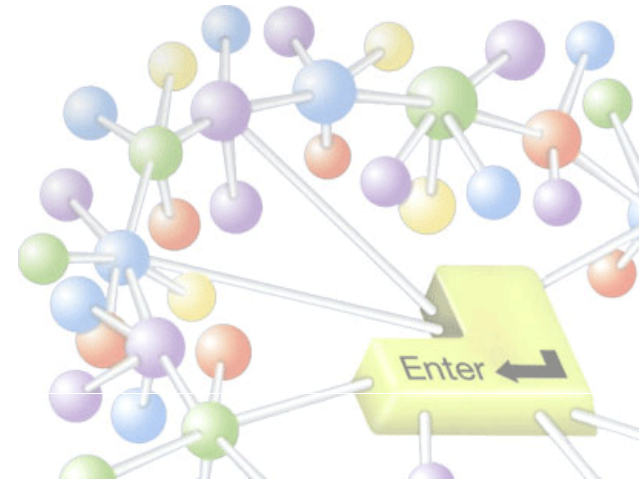
Aplicaciones Generales

- Eficiencia de Procesos
 - Se aplican técnicas de DM a casos ya resueltos para determinar reglas
 - Control estadístico de procesos de manufactura



Bioinformática

- Gestión y análisis de datos biológicos
- Predicción de estructura de proteínas, genoma humano
- Técnicas de predicción para determinar químicos con mayor probabilidad de producir drogas útiles



Telecomunicaciones

- Detección de Fraudes
- Perfilación de clientes
- Definición de productos (planes familia y amigos, identificación de clientes residenciales/oficina, llamadas internas)
- Aislamiento de fallas en red
- Localización de nuevos recursos (antenas)



Medicina

- Análisis características de grupos humanos sensibles a cierto tipo de enfermedades - Diagnóstico
- Mezcla con GIS para determinar condiciones geográficas asociadas a la presentación de enfermedades
- Planeación de recursos para atender necesidades de población



Almacenes de Retail

- Marcas propias
- Combos de productos
- Administración espacio en góndolas



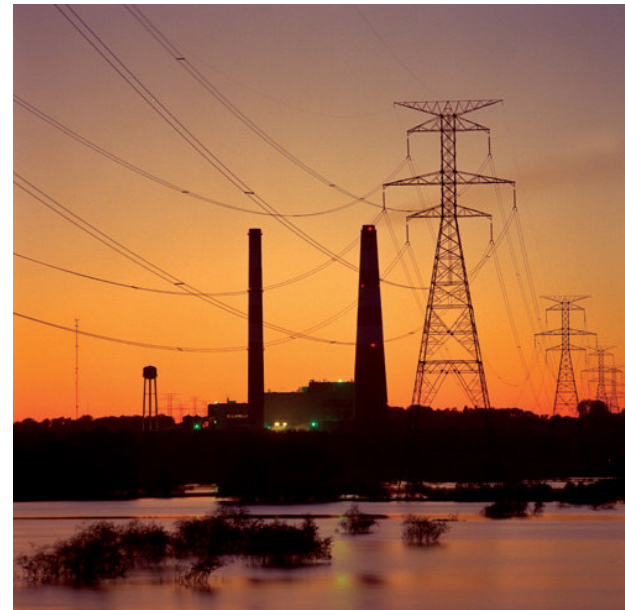
Entidades Financieras

- Control de Fraude
- Nivel de “involucramiento” de los clientes



Empresas de Servicios Públicos

- Venta de espacios de mercado dirigido
- Identificación de datos errados o fraudes por desviación en patrones de consumo



Entidades Gubernamentales

- Análisis de conexiones (Link Analysis sobre reportes de sus agentes para encontrar pistas y asociaciones)
- Identificación de patrones sospechosos en traslados internacionales de fondos
- Lavado de activos



Brokers de Información

- Se “venden” servicios de información para mercadeo (Supermercados, Tarjetas de Crédito)
- Se recopila información para ofrecer servicios a terceros (Hábitos de TV, Recetas Farmacéuticas)



Agenda

- Minería de Datos en el contexto de BI y DWH
- Aplicaciones de Minería de Datos
- Principales Tareas y Técnicas
- Metodología para procesos de Minería



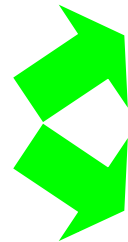
Estrategias de DM

Prueba de Hipótesis

Estrategia Top-Down: Sustenta o descarta ideas preconcebidas

Descubrimiento de Conocimiento

Estrategia Bottom-Up: Parte de los datos para descubrir



Dirigido

No Dirigido

Minería de Datos - Tareas

- Clasificación
 - Estimación
 - Predicción / Forecasting
 - Asociación / Agrupamiento por afinidad
 - Segmentación / Clustering
 - Descripción y Perfilación
 - Análisis Textual
- Dirigido
- No Dirigido

Clasificación

- Establecer una o más variables discretas de un objeto, con base en otros atributos del conjunto de datos - Analizar características de un nuevo objeto y asignarlo a una clase particular predefinida

- Clasificar solicitud de crédito en riesgo alto - medio – bajo
- Determinar qué teléfonos corresponde a máquinas de fax
- Identificar Reclamos de Seguro fraudulentos
- Clasificar persona como potencial “respondedor” a oferta
- Clasificar a un cliente dentro de un perfil particular

Estimación

- Establecer el valor de una variable continua, los resultados pueden ser ORDENADOS – Similar a Clasificación
- Estimar el valor del ingreso total de un grupo familiar
 - Determinar probabilidad de transacción sea fraudulenta
 - Estimar número de hijos en un grupo familiar
 - Estimar el valor del ciclo de vida de un cliente
 - Estimar probabilidad con que persona responde a campaña



Predicción

- Similar a clasificación o estimación, sólo que se refiere a identificar un comportamiento o valor estimado futuro

- Predecir qué clientes desertarán en los siguientes 6 meses
- Predecir el monto de saldo transferido si un prospecto de TC acepta la oferta de transferencia
- Predecir qué suscriptores de teléfonos ordenarán servicios de valor agregado

Asociación / Affinity Grouping

- Detectar eventos que ocurren de manera simultánea

- Un cliente que compra cerveza, compra pañales con prb P1
- Un cliente que compra Pizza, compra Vino con prb P1
- Un cliente que compra Vino, compra Pizza con prb P2

Segmentación / Clustering

- Dividir población heterogénea en grupos más homogéneos

- Identificar segmentos de clientes para un negocio particular

Descripción

- Describir un comportamiento en una base de datos compleja para aumentar el conocimiento y entendimiento sobre gente, productos, procesos etc. – Visualización – Diferenciación
- Establecer que las mujeres presentan menor siniestralidad en seguros de automóvil que los hombres
- Identificar las características de personas que apoya uno u otro partido político

Análisis Textual

- Convertir información des-estructurada en información estructurada (análisis de términos)

- Retroalimentación de datos en Call Center
- Clasificación de textos por temas

Técnicas

- Modelos de regresión lineal - logística
- Árboles de Decisión (regresión/modales)
- Redes Neuronales
- Detección de Clusters
- Market Basket Analysis
- Razonamiento Basado en memoria
- Link Analysis
- Algoritmos Genéticos





Agenda

- Minería de Datos en el contexto de BI y DWH
- Aplicaciones de Minería de Datos
- Principales Tareas y Técnicas
- Metodología para procesos de Minería



El ciclo de Minería de Datos



Consideraciones

- DM puede verse como un proceso de “Machine Learning”
 - Lenguaje de descripción de conceptos (language bias)
 - Orden en que se busca en el espacio de búsqueda (search bias)
 - Manera como se evita el overfitting al conjunto de entrenamiento (overfitting-avoidance bias)



Consideraciones

- La mayoría de métodos aprenden mediante ejemplos
- Los “ejemplos” se establecen mediante INSTANCIAS que tienen ATRIBUTOS (numéricos / categóricos o nominales) y están “pre-clasificados”



Consideraciones

◆ Preparar Datos

- Incluir información derivada o calculada
- Dividir en 3 conjuntos: Entrenamiento – Prueba – Evaluación
- Incluir casos “positivos” y “negativos”
- Oversampling



Ciclo de Vida – Minería de Datos

