

Minería de Datos

III. Experiencias Exitosas en la Empresa

Pedro Larrañaga, Iñaki Inza, Abdelmalik Moujahid

Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad del País Vasco

MMCC, 2006-2007

Índice

- 1 Casos exitosos en empresas
- 2 Software de minería de datos
- 3 Convirtiendo mis datos en conocimiento
- 4 Conclusiones

Índice

- 1 Casos exitosos en empresas**
- 2 Software de minería de datos
- 3 Convirtiendo mis datos en conocimiento
- 4 Conclusiones

Casos exitosos en empresas

Empresas

- Caja Madrid
- Iberia
- La Caixa
- Mapfre
- Axa Seguros e Inversiones
- Estrella Seguros
- ING Direct
- Repsol YPF
- Bankinter

Un tesoro en cada empresa

Inteligencia de negocio (Business Intelligence)

- Ampliar la visión estratégica
- Reducir el riesgo y la incertidumbre en la toma de decisiones
- Construir ventajas competitivas a largo plazo

Un tesoro en cada empresa

Parte esencial en la gestión empresarial

- ¿**Quienes** son nuestros **clientes**?
- ¿**Donde** están?
- ¿**Cómo** son?
- ¿Que pautas de **comportamiento** tienen?
- ¿**Que** es lo que **quieren**?

Un tesoro en cada empresa

Parte esencial en la gestión empresarial

- ¿Quienes son los **clientes** verdaderamente **rentables**?
- ¿Que puedo **ofrecerles**?
- ¿Cual es la forma mas **rentable** de hacerlo?
- ¿Que probabilidad hay de que **me abandonen**?

Un tesoro en cada empresa

Parte esencial en la gestión empresarial

- ¿Cuanto **me cuesta cada** uno de mis **productos** o servicios?
- ¿Cuales son **rentables**?
- ¿Cuanto me cuesta **conseguir un cliente**?
- ¿Que **departamentos funcionan**?
- ¿Como **dimensionar** cada uno de ellos?

Un tesoro en cada empresa

Parte esencial en la gestión empresarial

- Manejo de la información de negocio
- Factor clave de la gestión empresarial
- Soluciones tecnológicas para lograrlo

Caja Madrid



- Cuarta entidad financiera española
- 244.000 millones de euros
- 12.500 empleados
- 1920 oficinas
- 6,4 millones de clientes

Caja Madrid



Conocimiento de su cartera de clientes

- En 1992 preocupación por **conocer mas acerca de sus clientes**
- **Primera segmentación** usando variables como edad, nivel de renta y la actividad del cliente
- **Segunda segmentación** añadiendo compromiso con la compañía y rentabilidad del cliente
- **Mejora** un 50 por ciento los ratios de **eficiencia de la campaña de marketing**

Caja Madrid



El valor del cliente

- ¿Quiénes son los **clientes rentables** y quiénes no?
- Invertir **mas recursos de marketing en los primeros**
- **Fidelización de clientes rentables:**
 - Conseguir y mantener la **satisfacción** de los mejores clientes
 - Conocer la **probabilidad de abandono** de cada cliente por medio de **modelos predictivos** que proporcionan alarmas

Iberia



Iberia Plus

- Sistema acumulativo de **puntos**
- Permite optar a **vuelos gratuitos**
- **Categorías** según uso (plata, oro, platino)
- Utilizado como **fuentes de datos del comportamiento del cliente**: nivel de gasto, frecuencia de viajes, destinos, información que solicita, reclamaciones que efectúa, ...



Segmentar

- Localizar **grupos de clientes** con características comunes para efectuar **acciones comerciales** muy **definidas** y con la máxima **garantía de eficiencia**
- 18 **campañas de marketing** dirigidas a cada segmento

Acciones para la fidelización

- Los **100.000 mas rentables generan el 20 por ciento** de los ingresos totales
- **Invitaciones** a pruebas deportivas, actuaciones musicales, exposiciones,
- Pasar **de clase turista a "business class"** sin pedirlo expresamente, a partir de la información que aparece en la pantalla en el mostrador de facturación

La Caixa



- Primera caja de ahorros de España
- 4800 sucursales
- 9 millones de operaciones anuales de venta de nuevos productos
- Mas de 200 productos financieros (ahorro, financiación, seguros, servicios)
- Multicanal (oficinas, mailing, telemarketing, internet, cajeros automáticos, ...)
- Datos de 36 meses, con variables sociodemográficas, de contratación, consumo, productos contratados: 500 variables estáticas + 200 variables de comportamiento
- 6 terabytes (6 millones de megabytes) de información

La Caixa



Segmentación

- **Optimizar los recursos del marketing:** cada cliente es tratado de forma individual y personalizada
- Información personalizada a cada cliente (**800.000 modelos diferentes de información**) con contenidos de interés general y personalizados

La Caixa



Oficina inteligente

- Cada vez que un cliente entra a una oficina, el empleado tiene **en pantalla toda la información relevante** acerca del mismo:
 - **Acciones comerciales ajustadas** al perfil del cliente
 - **Modelos de predicción** que asignan **probabilidades de contratación** para cada producto financiero

La Caixa



Detectando abandonos

- **Modelos predictivos** asignando a cada cliente una **probabilidad de abandonar la compañía**
- **Síntomas previos** al abandono:
 - Disminución del pasivo
 - Supresión de movimientos en cuenta
 - Eliminación de domicializaciones
 - Anulación de tarjetas

Mapfre



- 3,3 millones de clientes
- 3,9 millones de pólizas (hogar, fallecimientos, accidentes, comercio)
- 1,3 millones de clientes en pólizas de hogar
- 2,2 millones de pólizas de hogar
- Entre 1993 y 1997 incremento del número de pólizas anuladas del 137 por ciento

Mapfre



Conocer el valor del cliente de pólizas de hogar

- **Variables:** sociodemográficas (edad, zona residencia, consumo), seguros contratados en los últimos años, características de la vivienda, coste siniestralidad de cada cliente, ...
- **Modelos** (árboles de clasificación, redes neuronales, regresiones logísticas) para **predecir el valor del cliente** (previsión de gastos - estimación de gastos):
 - El 15 por ciento de los clientes proporcionaba el 80 por ciento del valor (grupo A)
 - El siguiente 25 por ciento de los clientes mas valiosos (grupo B)
 - El 50 por ciento siguiente (grupo C)
 - El 10 por ciento restante (grupo D) con valor negativo o nulo para la compañía

Mapfre



Acciones para incrementar el rendimiento de la cartera y evitar su abandono

- Segmentar clientes según su valor actual en alto y bajo. A la vez subdividirlos según su valor futuro en alto y bajo
- Crear distintos modelos predictivos de abandono de la póliza para cada uno de los cuatro grupos
- Acciones concretas para cada uno de los cuatro segmentos
- Se consigue romper la tendencia de 1997
- En 2003 tasa de anulaciones de pólizas en Hogar se reduce en un 40 por ciento (del 14,3 en 1997 al 8,5 en 2003)
 - En número de cancelaciones mas de 68.000 pólizas (45.000 clientes)
 - En volumen de negocio 3,7 millones de euros

Axa Seguros e Inversiones



- 1,5 millones de pólizas en el ramo del automóvil
- Finales de los años 90 ratios de abandono inquietantes
- ¿Quiénes son sus mejores clientes para fidelizarlos?
 - Conocer los clientes valiosos
 - Estimar para cada uno de ellos su probabilidad de abandono
 - Focalizar en dichos clientes acciones comerciales para evitar su fuga

Axa Seguros e Inversiones



Modelo de predicción de anulaciones de pólizas

- 26 variables, 176.000 pólizas, **12 por ciento de anulación**
- Paradigma de **regresión logística**
- **Conocimiento adquirido:**
 - Cliente **cuanto mas antiguo** en la compañía **menos probabilidad** de anulación
 - Cliente que **anula pólizas de otros ramos aumenta** tambien la **probabilidad** de anular pólizas de autos
 - Clientes que **mantienen póliza** en otros ramos tienen alta probabilidad de mantenerse **fieles** en autos
 - Tasa de **abandono crece** según la **antigüedad del auto**
 - **A partir de 30 años de edad del cliente**, la tasa de **anulación bajaba** considerablemente
 - Pólizas contratadas a través de **corredores libres más probabilidad** de anulación
 - Cuanto **más se fraccionaba el pago**, mas **alta tasa** de anulación
 - **Rapidez con la que la compañía respondía** ante un siniestro **bajaba** la probabilidad de anulación

Axa Seguros e Inversiones



Remedio frente al abandono

- Buscando **actuaciones comerciales más efectivas**
 - **Dos grupos** de clientes (A con alta probabilidad de anulación, B con baja)
 - **Cada grupo se subdividió en cinco subgrupos** de igual tamaño. En cada subgrupo una **estrategia diferente**:
 - A.1 y B.1: ninguna actuación (12 % abandono)
 - A.2 y B.2: llamada (6 % abandono)
 - A.3 y B.3: llamada + servicio adicional (5 % abandono)
 - A.4 y B.4: contacto personal (3 % abandono)
 - A.5 y B.5: comunicación via correo + oferta servicios adicionales (10 % abandono)

Estrella Seguros



- Compañía de seguros **de las 10 mas grandes** de España
- Base de datos con **variables** específicas conteniendo:
 - Operativa del cliente con la entidad
 - Historial de adquisición de productos
 - Historial de siniestrabilidad

Estrella Seguros



Saneamiento. Adecuación. Abandono. Calidad del servicio

- **Saneamiento de la cartera de clientes:** modelos de segmentación de clientes y modelos de predicción que permitieron discriminar los clientes que merecía la pena conservar
- **Adecuar los precios a la realidad del mercado:** modelo para calcular las primas de los seguros de manera automática según los ratios de cada cliente, el tipo de cobertura y el vehículo. Criterios por los que se acepta o deniega una determinada contratación
- **Modelos de abandono de clientes:** probabilidad de que cada uno de los mejores clientes abandone la compañía en función de la edad, zona residencia, tipo vehículo, cobertura de riesgos
- **Control de calidad del servicio:** velocidad de los procesos de peritaje, liquidación de las indemnizaciones, quejas de los clientes
- **Modelos de predicción de compra:** para cada cliente y cada producto de la compañía

ING Direct



Banco singular

- **En 7 años:** 1.175.000 clientes, 15.000 millones de euros de recursos, **primer banco directo en España**, sexto por volumen de depósitos de ahorro
- Sólo **tres oficinas**
- **Teléfono e internet**
- Manera distinta de tratar al cliente: al identificarlo un **dispositivo inteligente conduce su llamada** (en tiempo real) al operador mas capacitado

Toda decisión respaldada por modelos matemáticos

- En **2002** tenía **400.000 clientes** con cuenta naranja (marketing tradicional)
- En **2005**, **mas de un millón de clientes** en 4 líneas de negocio (cuenta naranja, hipoteca naranja, productos de inversión y cuenta nómina)
- **Conocer a los clientes**: 20 variables (demográficas, comportamiento cliente en canales que mas utiliza, interés por diferentes productos, ...)
- **Segmentación** (perfiles homogéneos)
- **Modelos predictivos (regresión logística)** para probabilidad de adquirir los distintos productos financieros

Repsol YPF



Visa Repsol

- Carburantes cuota de **mercado del 50 %** en España
- Tarjeta Visa Repsol como **herramienta de fidelización**:
 - **Gratuita**
 - **2 % descuento en estaciones de servicio** del grupo Repsol YPF
 - Cantidad de información (**clientes no anónimos**)
 - Datos sobre el **comportamiento de compra**: nivel de consumo, frecuencia de compra, distribución geográfica, tipo producto, servicios de tienda, días de compra, horario, estacionalidad del gasto, ...
 - Perfil de usuarios: **7 grupos** (políticas de marketing diferenciadas)
 - **A cada cliente su campaña de marketing**: reducción de costes y aumento de respuestas positivas
 - Tarjeta Solred para **profesionales del transporte** (900.000 tarjetas)

Basilea II cambio de reglas de juego

- Mas **énfasis en los procedimientos de gestión**:
 - Tener activadas las operativas para **detectar y evaluar** de forma continuada los **riesgos de gestión**
 - **Pérdida esperada** a partir de la probabilidad de incumplimiento y la severidad de la pérdida
 - **Pérdida inesperada** derivada de la volatilidad de los mercados y de los cambios cíclicos en la economía
- **Modelos predictivos** estableciendo un **nivel de riesgo** (de 1 a 9) para cada operación (personas físicas y empresas (con ayuda de experto)) para determinar las coberturas de capital
- **Modelos de simulación de riesgo**: ver el comportamiento modificando las distintas variables que influyen en el negocio financiero

Índice

- 1 Casos exitosos en empresas
- 2 Software de minería de datos**
- 3 Convirtiendo mis datos en conocimiento
- 4 Conclusiones

Sistemas y herramientas de minería de datos

- Entornos clasificados en tres grupos:
 - **Librerías**: implementan las funcionalidades y utilidades básicas como acceso a datos, inducción de modelos, exportación de resultados. Facilitan el desarrollo de diseño de experimentos, contraste de modelos, creación de modelos combinados. Necesarios conocimientos de programación
 - **Suites**: capacidades para el preprocesado de datos, diferentes paradigmas, facilidades para el diseño de experimentos, interfaz que facilita la interacción entre el usuario y la herramienta
 - **Herramientas específicas**: se centran en un determinado paradigma, y/o en una determinada tarea de minería de datos. No necesarios conocimientos de programación

Librería: MLC++

<http://www.sgi.com/tech/mlc>

- MLC++ (**Machine Learning library in C++**)
- Hasta la versión 1.3x de dominio público (**Universidad de Stanford**)
- Versiones posteriores distribuidas por **Silicon Graphics**
- MLC++ integrado con **MineSet** herramienta de visualización
- Características:
 - **Acceso a datos**: archivos con formato plano siguiendo el estilo de los del repositorio UCI
 - **Transformaciones de datos**: reducido (eliminación de variables, discretizaciones, construcción de nuevas variables)
 - **Paradigmas**: tablas de decisión, árboles de clasificación, métodos Bayesianos, perceptrón multicapa

Librería: MLC++

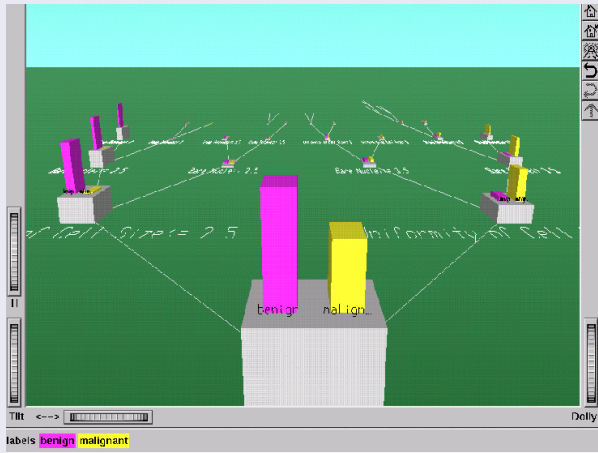


Figura: Visualización de un árbol de clasificación en MLC++ mediante la herramienta MineSet

Suite: SPSS Clementine

<http://www.spss.com/spssbi/clementine>

- **Herramienta visual** inicialmente desarrollada por ISL (Integral Solutions Limited)
- En la actualidad comercializada por **SPSS** con una arquitectura distribuida (cliente/servidor)
- Sistema **multiplataforma** (Windows, Sun Solaris, HP-UX, OS/400)
- Características:
 - **Acceso a datos**: tablas Excel, archivos planos, archivos SPSS
 - **Preprocesado de datos**: muestreo, particiones, reordenación de campos, fusión de tablas, recodificaciones
 - **Paradigmas**: árboles de clasificación, redes neuronales (Kohonen, perceptrón multicapa, RBF), agrupamiento (K medias), reglas de asociación (Apriori), regresión lineal, logística, combinación de modelos (boosting con C5.0)
 - Técnicas de **evaluación de modelos**
 - **Visualización** de resultados: visión global de todo el proceso
 - **Exportación**: generación automática de informes

Suite: SPSS Clementine

<http://www.spss.com/spssbi/clementine>



Figura: Entorno gráfico de Clementine

Suite: Weka

<http://www.cs.waikato.ac.nz/~ml/WEKA>

- WEKA (**Waikato Environment for Knowledge Analysis**) herramienta visual de libre distribución desarrollada en la Universidad de Waikato (Nueva Zelanda)
- Características:
 - **Acceso a datos**: datos cargados en formato ARFF (archivo plano organizado en filas y columnas), también se permite importar de diversos formatos
 - **Preprocesado de datos**: selección de variables, discretización, tratamiento de valores desconocidos, transformación de variables
 - **Paradigmas**: árboles de clasificación, tablas de decisión, vecinos más próximos, máquinas de soporte vectorial, reglas de asociación, agrupamiento, clasificadores Bayesianos, combinación de clasificadores
 - **Visualización**: filtrado, selección de variables, selección del modelo, diseño de experimentos, proceso global
 - Herramienta desarrollada en **Java**. Licencia GNU posibilita actualizar su código (gran número de proyectos asociados)

Suite: Weka

<http://www.cs.waikato.ac.nz/~ml/WEKA>

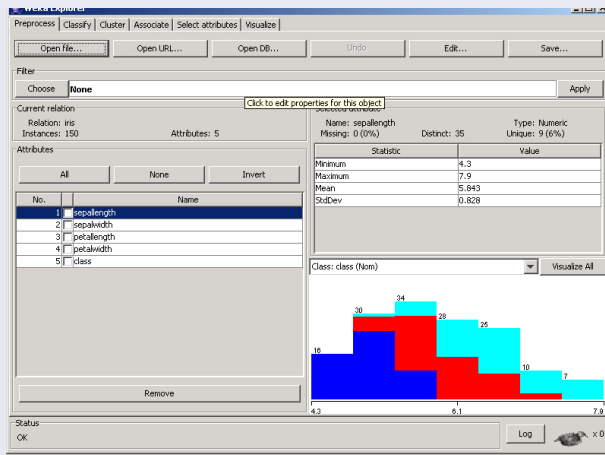


Figura: Weka en forma explorer

Suite: DB2 Intelligent Miner

<http://www-306.ibm.com/software/data/iminer>

- DB2 Intelligent Miner herramienta comercial distribuida y desarrollada por **IBM** para explotar masivos sistemas de información
- **DB2 Intelligent Miner for Data**: agrupamiento, asociaciones, patrones, clasificación, predicción, análisis de series temporales
- **DB2 Intelligent Miner Modelling**: asociaciones, árboles de clasificación, agrupamiento demográfico
- **DB2 Intelligent Miner Visualization**: visualizar y evaluar gráficamente los resultados
- **Plataformas**: Windows, Solaris, AIX, OS

Suite: DB2 Intelligent Miner

<http://www-306.ibm.com/software/data/iminer>

IBM United States [change] Terms of use

Home Products Services & industry solutions Support & downloads My IBM

Software > Information Management >

DB2 Intelligent Miner

DB2 Intelligent Miner Library

Warranty info

IBM's data mining capabilities help you detect fraud, segment your customers, and simplify market basket analysis. IBM's in-database mining capabilities integrate with your existing systems to provide scalable, high performing predictive analysis without moving your data into proprietary data mining platforms. Use SQL, Web Services, or Java to access DB2's data mining capabilities from your own applications or business intelligence tools from IBM's business partners.

Support and replacements

- **Announcement Letter 906-047**
March 14, 2006, IBM announced withdrawal from marketing and end of support for the Intelligent Miner tools. Effective June 14, 2006, they were withdrawn from marketing. Effective September 30, 2007, V8.1 will be withdrawn from support and effective April 30, 2008, V8.2 will be withdrawn from support.
- **DB2 Intelligent Miner for Scoring support**
- **DB2 Intelligent Miner Visualization support**
- **DB2 Intelligent Miner for Data support**
- **DB2 Data Warehouse Editions** is the replacement product.
- **DB2 Intelligent Miner Modelling support**

Information Management software

DB2 Data Warehouse Edition
The preferred platform for data warehousing

IBM Information
On Demand
Oct 14 - 19, 2007

About IBM Privacy Contact

Figura: DB2 Intelligent Miner

Suite: SAS Enterprise Miner

<http://www.sas.com>

- Una de las dos suites que proporciona **SAS System** (la otra es SAS Text Miner)
- **Arquitectura distribuida**. Potente interfaz gráfica
- **Metodología SEMMA** (Sample, Explore, Modify, Model, Assess)
- **Acceso datos**: archivo propio, Oracle, DB2, soporte para almacenes de datos
- **Preprocesado de datos**: transformaciones, missing, filtros valores extremos, muestreo
- **Paradigmas**: Árboles de clasificación, regresión lineal, logística, multclasificadores (ensemble, boosting, bagging)
- **Evaluación**
- **Visualización** resultados
- **Multiplataforma** (Windows, Linux, Solaris)

Suite: SAS Enterprise Miner

<http://www.sas.com>



Figura: SAS Enterprise Miner

Herramienta específica: CART

<http://www.salford-systems.com/index.html>

- CART (**Classification and Regression Trees**): Árboles de clasificación y regresión (Universidad de Stanford)
- Comercializado por **Salford Systems**
- Acceso a **70 formatos** de archivos diferentes
- Capacidad de **visualización interactiva**
- **Información estadística**: Errores de clasificación, importancia de cada variable en la clasificación
- **Multiplataforma**: Windows, Linux, Unix

Herramienta específica: CART

<http://www.salford-systems.com/index.html>

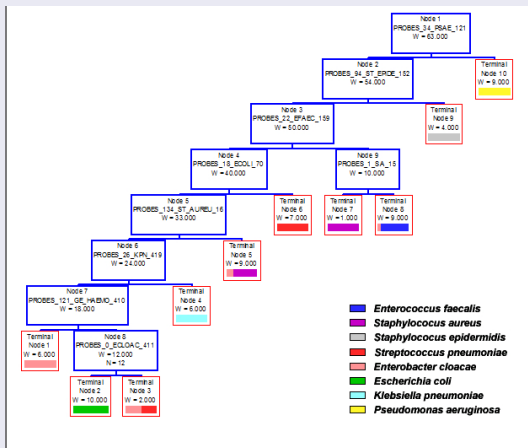


Figura: Árbol de clasificación obtenido con CART

Herramienta específica: AutoClass

<http://ic.arc.nasa.gov/ic/projects/bayes-group/autoclass>

- AutoClass **agrupamiento** mediante **técnicas Bayesianas**
- Desarrollado por la **NASA**
- **AutoClass C** es de **libre distribución**, pero no dispone de utilidades gráficas para la representación de la información
- AutoClass C **multiplataforma**: Windows, Unix (Solaris, SunOs), Linux, MS-DOS

Herramienta específica: AutoClass

<http://ic.arc.nasa.gov/ic/projects/bayes-group/autoclass>

The AutoClass Project

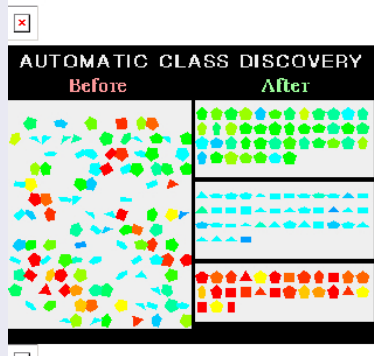


Figura: AutoClass clustering Bayesiano

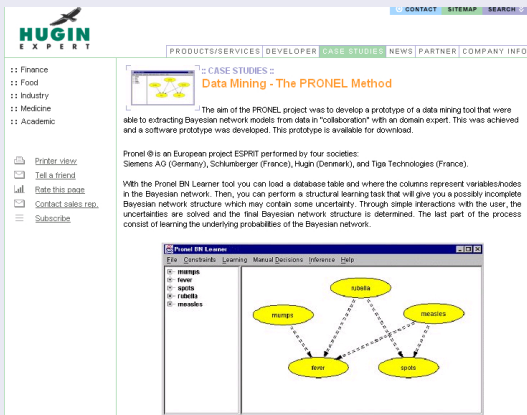
Herramienta específica: HUGIN

<http://www.hugin.com>

- Software comercial desarrollado inicialmente en la **Universidad de Aalborg** (Dinamarca)
- **Redes Bayesianas** y diagramas de influencia
- **Inferencia** (razonamiento) exacta y aproximada
- **Aprendizaje a partir de datos** (detectar independencias y score + búsqueda)

Herramienta específica: HUGIN

<http://www.hugin.com>



The screenshot shows the HUGIN EXPERT website. The header includes the HUGIN logo and navigation links: CONTACT, SITEMAP, and SEARCH. Below the header is a menu bar with links: PRODUCTS/SERVICES, DEVELOPER, CASE STUDIES (highlighted), NEWS, PARTNER, and COMPANY INFO. The main content area is titled 'CASE STUDIES' and features a link to 'Data Mining - The PRONEL Method'. A small diagram of a Bayesian network is shown next to the link. The text describes the PRONEL project's aim to develop a data mining tool by extracting Bayesian network models from data in collaboration with an expert. It mentions that the tool is available for download. Below the text, there is a screenshot of the 'ProneL BN Learner' software interface, which displays a Bayesian network diagram with nodes labeled 'mumps', 'fever', 'spots', 'rubella', 'measles', 'Rovar', and 'spots'. The interface includes a menu bar with options like File, Constraints, Learning, Manual Decisions, Inference, and Help. A list of nodes is visible on the left side of the interface.

Figura: Redes Bayesianas en minería de datos

Índice

- 1 Casos exitosos en empresas
- 2 Software de minería de datos
- 3 Convirtiendo mis datos en conocimiento**
- 4 Conclusiones

Implementando un proyecto de minería de datos

Pilares de la minería de datos

- A. **Datos** como materia prima
- B. **Técnicas** de minería de datos
- C. La **modelización** propiamente dicha
- D. Gestión del **cliente** (en aplicaciones empresariales)

Implementando un proyecto de minería de datos

A. Datos como materia prima

- **Nos encontramos con:**
 - Distintos formatos, tipos, sistemas
 - Mucho ruido, datos incompletos
- **Debemos conseguir:**
 - Única tabla: registro (caso), columna (variable)
 - En modelización predictiva identificar una (o varias) variable(s) de interés

Implementando un proyecto de minería de datos

B. Técnicas de minería de datos

- **Preproceso**: imputar, discretizar, seleccionar variables
- **Clustering** (segmentación): a la búsqueda de perfiles (K -means)
- **Clasificación supervisada**: predecir el valor de una variable en base a otras (árbol de clasificación, clasificadores K -NN, regresión logística, redes Bayesianas, sistemas clasificadores, redes neuronales, inducción de reglas, máquinas de soporte vectorial, análisis discriminante, ...)
- **Evalución honesta** de la bondad de los modelos

Implementando un proyecto de minería de datos

C. La modelización propiamente dicha

- Modelar requiere de **conocimientos técnicos**, pero tambien de **experiencia, saber hacer, intuición**, ...
- Modelo computacional: **simplificación** de la parte de la realidad a la que hemos accedido (los datos) en base a un ente **abstracto** (el modelo)
- Analogía con la **escultura**, la **pintura**
- **No sólo ciencia tambien arte**

Implementando un proyecto de minería de datos

D. Gestión del cliente

- Cada empresa tiene varios **tesoros**
- Uno de ellos es la **cartera de clientes**
- Identificar los **rentables**, los **fieles**
- **Retenerlos**: elaborar acciones con tal objetivo
- Anticiparse a posibles **abandonos**
- **Minería de datos ayuda** para la identificación y retención (fidelización) de clientes

Implementando un proyecto de minería de datos

¿En quien me apoyo?

- **Software disponible** (no sólo se trata de usar el ratón) versus **programación "ad hoc"**
- Tan sólo organizaciones empresariales de determinado tamaño llegan a tener un **departamento de minería de datos** donde resuelven todo el proceso (de datos a conocimiento)
- Lo habitual **apoyarse** en **empresas asesoras** y/o en la **Universidad**
- Complementaridad ya que la **Universidad**:
 - Puede **aplicar metodología ya existente** con garantías
 - Puede **desarrollar nueva metodología** si el problema así lo requiere
- La **confidenciabilidad** de los datos está garantizada

Índice

- 1 Casos exitosos en empresas
- 2 Software de minería de datos
- 3 Convirtiendo mis datos en conocimiento
- 4 Conclusiones**

A modo de resumen

- **Minería de datos permite**, a partir del análisis de archivos y registros, descubrir perfiles, relaciones, reglas, asociaciones, excepciones que sean útiles para la toma de decisiones
- **Clustering** (segmentar)
- **Modelos predictivos**: diagnosticar, pronosticar
- **De datos a conocimiento** (sociedad del conocimiento)
- **Mi empresa se puede beneficiar** de esta nueva tecnología de la información
- Inteligencia computacional aplicada al negocio: *business intelligence*

Minería de Datos

III. Experiencias Exitosas en la Empresa

Pedro Larrañaga, Iñaki Inza, Abdelmalik Moujahid

Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad del País Vasco

MMCC, 2006-2007