Escola d' Estiu de Doctorat fac. Pedagogia

Taller Minería de datos aplicados a la educación

1ª parte

Introducción a la minería de datos

27 de junio de 2011

Mercedes Torrado

Departamento Métodos de Investigación y Diagnóstico en Educación (MIDE)

Para citar la obra:
Torrado, M. (2011) *Minería de datos aplicados a la educación.* Barcelona: Universidad de Barcelona. Deposito Digital http://hdl.handle.net/2445/19862

Introducción a la Minería de datos

- *INTRODUCCIÓN*
- *CONCEPTO*
- *PROCESO DE MINERÍA DE DATOS*
- *EJEMPLOS*
- *PROGRAMAS INFORMÁTICOS*
- *OTRAS APLICACIONES*

Introducción a la Minería de datos

Los avances tecnológicos en las últimas décadas nos han facilitado enormemente el acceso a grandes volúmenes de datos.

La cantidad de información que manejamos hoy en día nos obliga a abordar el estudio de los datos/información desde una **perspectiva global y no parcelada**

La preocupación por disponer de información suficiente para la toma de decisiones.

¿Cómo podemos analizar tal cantidad de información e identificar aquella que nos permita tomar decisiones y mejorar?

Escola d' Estiu de Doctorat Fac. Pedagogia

Torrado, M. (2011) ***La Minería de datos en educación (MDE)***

Introducción a la Minería de datos

En los años 90 apareció el concepto **DATA MINING.**

Esta técnica se vinculó estrechamente con la dirección de empresas y en concreto al marketing.

La minería de datos o Data Mining puede definirse como una extracción de información desconocida no trivial y potencialmente útil de una gran cantidad de información

Bajo este término se engloban un conjunto de técnicas de análisis cuyo objetivo es extraer conocimiento implícito de la base de datos.

Escola d' Estiu de Doctorat fac. Pedagogia

Torrado, M. (2011) ***La Minería de datos en educación (MDE)***

Introducción a la Minería de datos

DEFINICIÓN

El ***Data mining*** también es considerado como una tecnología emergente que parte, por un lado de las técnicas estadísticas y por otro de las técnicas de inteligencia artificial Aluja, 2001 (1)

Estadística se ha preocupado más por la posible generalización de los resultados

Inteligencia artificial – ofrece soluciones algorítmicas a los datos

La **Minería de datos** comprende un conjunto de técnicas para la descripción y predicción a partir de grandes masas de datos (Viera et al., 2009: 12)

[1] Tomàs Aluja en su artículo *La minería de datos, entre la estadística y la inteligencia artificial* publicado en el 2001 en la revista **QÜESTIIO** (vol 25, 3, p 479-498) hace todo un repaso de los orígenes de la Minería de datos en cuanto a los elementos correspondientes de la Estadística y de la inteligencia artificial

Equivalencias de nomeclaturas entre la Estadística y la Inteligencia Artificial

(Aluja, 2001: 482)

Inteligencia artificial	Estadística
Red (network)	Modelo
Ejemplos (patterns)	Observaciones, individuos
Inputs, outputs, features	Variables
Inputs	Variables explicativas
Outputs, targets	Variables de respuesta
Errores	Residuos
Training, learning	Estimación
Función de error, coste	Criterio de ajuste
Pesos, coef. sinápticos	Parámetros
Aprendizaje supervisado	Regresión, discriminación
Aprendizaje no supervisado	Clasificación

Introducción a la Minería de datos

Data mining se asocia a:

Minería predictiva

Usa principalmente técnicas estadísticas

Triangulación estadística

(Torrado, 2007)

Minería de datos para el descubrimiento del conocimiento

Usa principalmente técnicas de inteligencia artificial

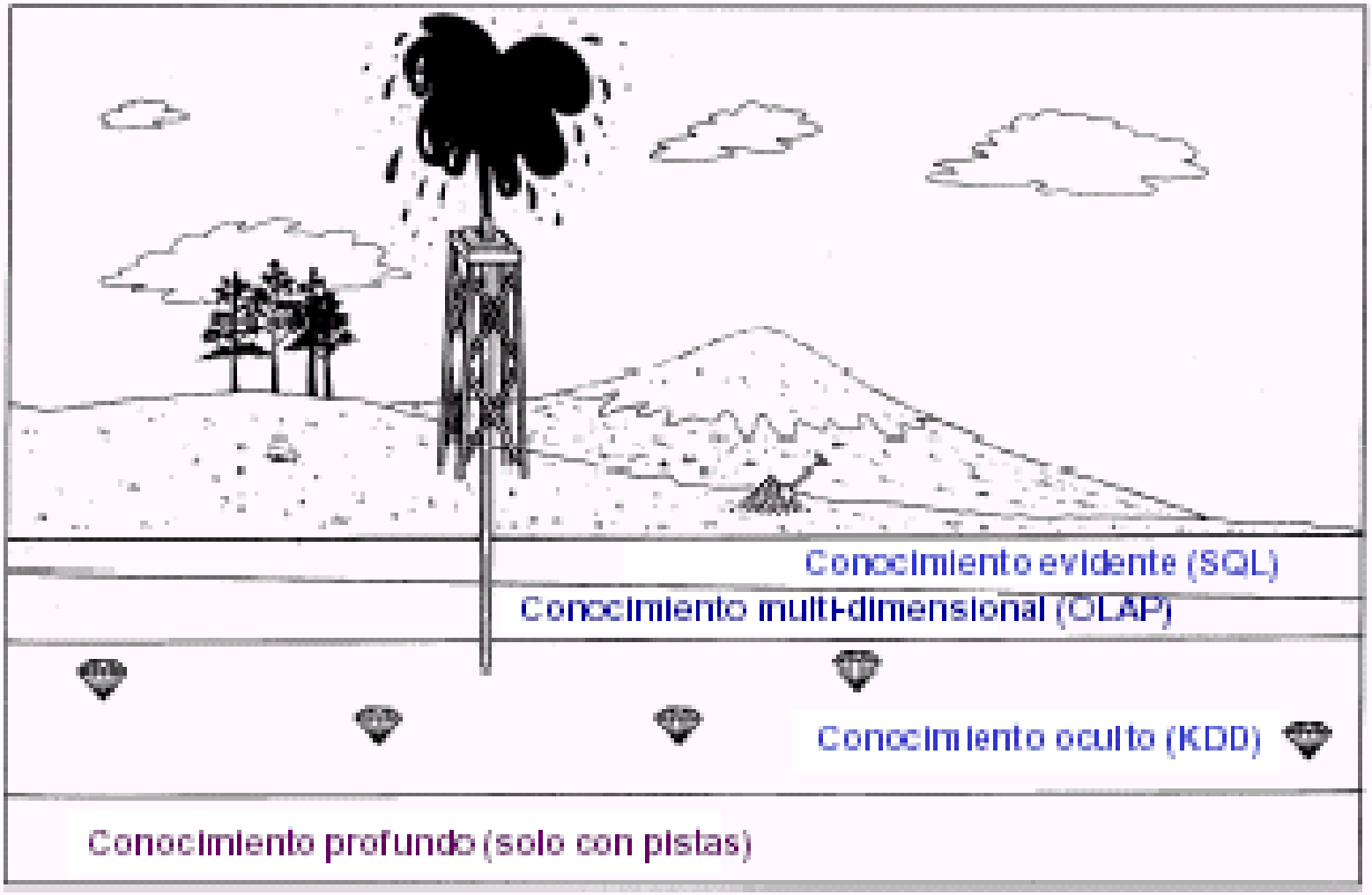
Knowedge Discovery in Databases

(KDD)

Escola d' Estiu de Doctorat Fac. Pedagogia

Torrado, M. (2011) ***La Minería de datos en educación (MDE)***

Distintos tipos de conocimiento



SQL: Structured Query Language
OLAP: Online Analytical Processing
KDD: Knowledge Discovery on Databases

Introducción a la Minería de datos

Proceso de Descubrimiento de Conocimiento de Bases de Datos(KDD)

Las **siglas KDD** fue creada en 1995 para designar el conjunto de procesos, técnicas que propician el contexto en el cual la minería de datos tendrá lugar”

(Viera et al., 2009)

Una posible definición:

La integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten información hacia la toma de decisiones

(Riquelme et al., 2006)

Introducción a la Minería de datos

La finalidad del KDD es:

- Procesar automáticamente grandes canti **DATO** de datos “brutos”
- Identificar los patrones más significativos y relevantes **INFORMACIÓN**
- Presentarlos como conocimiento apropiado para satisfacer las metas del usuario **CONOCIMIENTO**

Fuente de Datos | Pre-procesamiento | Exploración y transformación | Reconocimiento de patrones | Evaluación e interpretación

Fuente de datos

Bases de datos

Muestreo

Selección

Limpieza de datos

- Datos que no existen
- Datos no clasificados
- Identificación de extremos

Transformación de datos

- Reducir variables / dimensionalidad
- Creación de características
- Generación de variables nuevas

Modelado

- Clasificación
- Regresión
- Agrupamiento
- Asociación

Informes

PREPARACIÓN DE LOS DATOS

EXPLOTACIÓN

EVALUACIÓN

PROCESO KDD y de MINERÍA DE DATOS

Escola d' Estiu de Doctorat Fac. Pedagogia

Torrado, M. (2011) ***La Minería de datos en educación (MDE)***

Fuente de Datos | Pre-procesamiento | Exploración y transformación | Reconocimiento de patrones | Evaluación e interpretación

CLASIFICACIÓN

La finalidad es crear un modelo para poder predecir la pertenencia a un determinado grupo

EJEMPLO:
Diagnosticar alumnos de éxito académico

PRUEBAS:
Árboles de decisiones, análisis discriminantes, etc..

Modelado

- **Clasificación / asociación**
- Regresión
- Agrupamiento

Informes

EXPLOTACIÓN

PROCESO KDD y de MINERÍA DE DATOS

Escola d' Estiu de Doctorat Fac. Pedagogia

Torrado, M. (2011) ***La Minería de datos en educación (MDE)***

Fuente de Datos | Pre-procesamiento | Exploración y transformación | Reconocimiento de patrones | Evaluación e interpretación

REGRESIÓN

La finalidad es crear un modelo para poder predecir el valor de una variable dependiente a partir de otras independientes

EJEMPLO:
Estimar el rendimiento académico del primer año de carrera

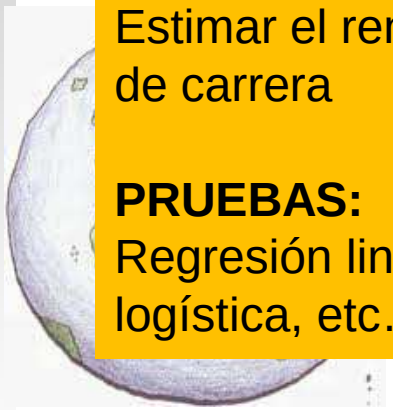
PRUEBAS:
Regresión lineal, redes neuronales, regresión logística, etc…

Modelado

- Clasificación /asociación
- **Regresión**
- Agrupamiento

Informes

XPLOTACIÓN



PROCESO KDD y de MINERÍA DE DATOS

Escola d' Estiu de Doctorat Fac. Pedagogia

Torrado, M. (2011) ***La Minería de datos en educación (MDE)***

Fuente de Datos | Pre-procesamiento | Exploración y transformación | Reconocimiento de patrones | Evaluación e interpretación

AGRUPAMIENTO / SEGMENTACIÓN

La finalidad es crear un modelo para poder agrupar con características similares

EJEMPLO:
Identificar perfiles de alumnos

PRUEBAS:
K-medias, Bietápico, etc..

Modelado

- Clasificación / asociación
- Regresión
- **Agrupamiento**

Informes

EXPLOTACIÓN

PROCESO KDD y de MINERÍA DE DATOS

Escola d' Estiu de Doctorat Fac. Pedagogia

Torrado, M. (2011) ***La Minería de datos en educación (MDE)***

LOS DATOS DEBEN SER (Viera et al., 2009)

Precisión – sin errores de medición
Consistencia – datos coherentes
Completos – sin falta de atributos
Relevancia – Concernientes al problema
No redundancia – Sin duplicar la misma información

PROCESO KDD y de MINERÍA DE DATOS

Torrado, M. (2011) ***La Minería de datos en educación (MDE)***

Introducción a la Minería de datos

Algunas aplicaciones (Riquelme, 2006)

Comercio y banca

Segmentación de clientes, previsión de ventas, análisis de riesgos

Medicina y farmacia

Diagnóstico de enfermedades y la efectividad de los tratamientos

Seguridad y detección de fraude

Reconocimiento facial, acceso a redes no permitidas,…

Astronomía

Identificación de nuevas estrellas y galaxias

Geología, minería, agricultura y pesca

Identificación de áreas de uso para distintos cultivos o pesca, explotación minera en base de datos de imágenes de satélites

Ciencias ambientales

Identificación de modelos de funcionamiento de ecosistemas naturales o artificiales

Ciencias sociales

Estudio de los flujos de opinión, identificar barrios con conflicto en función de valores socio-demográgicos

Introducción a la Minería de datos

Algunas aplicaciones en Educación MDE

En el ámbito educativo la aplicación de la minería de datos como técnica de análisis **se ubica en el entorno del sistema educativo y en concreto en Educación superior.** Las base de datos que se utilizan en los sistemas educativos permiten disponer de una gran cantidad de información, tanto de los estudiantes, trabajadores, departamentos, universidades, etc…., por ejemplo la base de datos UNEIX

- La MDE tiene como objetivo obtener una mejor comprensión del proceso de aprendizaje de los estudiantes y de su participación global en el proceso, orientado a la mejora de la calidad y rentabilidad del sistema educativo

(Winters, T, 2006)

Introducción a la Minería de datos

Algunas aplicaciones en Educación MDE

- R. Alcover, J. Benlloch, P. Blesa, M. A. Calduch1, M. Celma, C. Ferri, J. Hernández-Orallo, L. Iniesta, J. Más, M. J. Ramírez-Quintana, A. Robles, J. M. Valiente, M. J. Vicent, L. R. Zúnica. (2007) **Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos.** XIII Jornadas de enseñanza universitaria de la informática. Teruel. España (disponible internet)

- Quiroga, E. (2008) **Minería de datos en educación superior aplicada a un modelo de alerta acdémica.** Chile

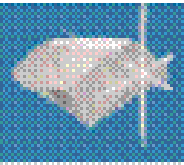
- Valero, S. (2009) **Aplicación de la minería de datos para predecir la deserción**. Universidad tecnológicva de Izúcar de Matamoros

Algunos ejemplos visuales

EJEMPLOS

Cerón, M.A. y Gómez, H. (2010) Minería de datos

(http://www.slideshare.net/04071977/mineria-de-datos)



Análisis de Canasta

Ejemplo
Un ejemplo tradicional de minería de datos es el relacionado con una búsqueda en una bodega de datos, de un negocio de cadena, de hechos comunes y relevantes:
Luego del proceso se dio como resultado la siguiente:

Si edad < 35;
y sexo = masculino;
y dia = jueves
entonces compras incluyen
pañales;
y cerveza

Esto sirvió para que empresa tomara medidas relacionada con la ubicación de ciertos productos en sitios comunes.

EJEMPLOS

Cerón, M.A. y Gómez, H. (2010) Minería de datos

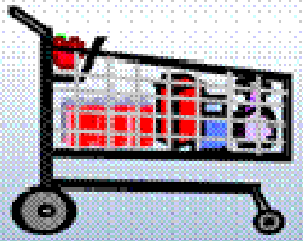
(http://www.slideshare.net/04071977/mineria-de-datos)

Análisis de Canasta (Market Basket Analysis)

Los hábitos de compra de los clientes pueden ser representados a través de asociaciones o correlaciones entre los diferentes productos que compran en sus "canastas".

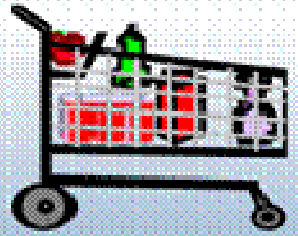
Cliente 1:

Arroz, puré, bebida



Cliente 2:

Arroz, helado, pan



Cliente 1:

Arroz, bebida, cerveza

2

Escola d' Estiu de Doctorat Fac. Pedagogia

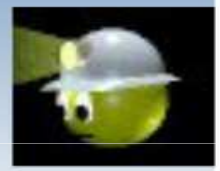
EJEMPLOS

Cerón, M.A. y Gómez, H. (2010) Minería de datos

(http://www.slideshare.net/04071977/mineria-de-datos)

Ejemplo

Gestión de personal de una empresa: ¿Qué clases de empleados hay contratados?



Datos:

Id	Salary	Married	Car	Children	Rent/ Owner	Union	Off sick/year	Work years	Gender
1	10000	yes	no	0	Rent	no	7	15	M
2	20000	no	yes	1	Rent	yes	3	3	F
3	15000	yes	yes	2	Owner	yes	5	10	M
4	30000	yes	yes	1	Rent	no	15	7	F
5	10000	yes	yes	0	Owner	yes	1	6	M
6	40000	no	yes	0	Rent	yes	3	16	F
7	25000	no	no	0	Rent	yes	0	8	M
8	20000	no	yes	0	Owner	yes	2	6	F
15	8000	no	yes	0	Rent	no	3	2	M
...	...	...	...	...	...	...	...	...	...

Modelo generado:

Minería de datos

Grupo 1: Sin niños y en una casa alquilada. Bajo número de uniones. Muchos días enfermos

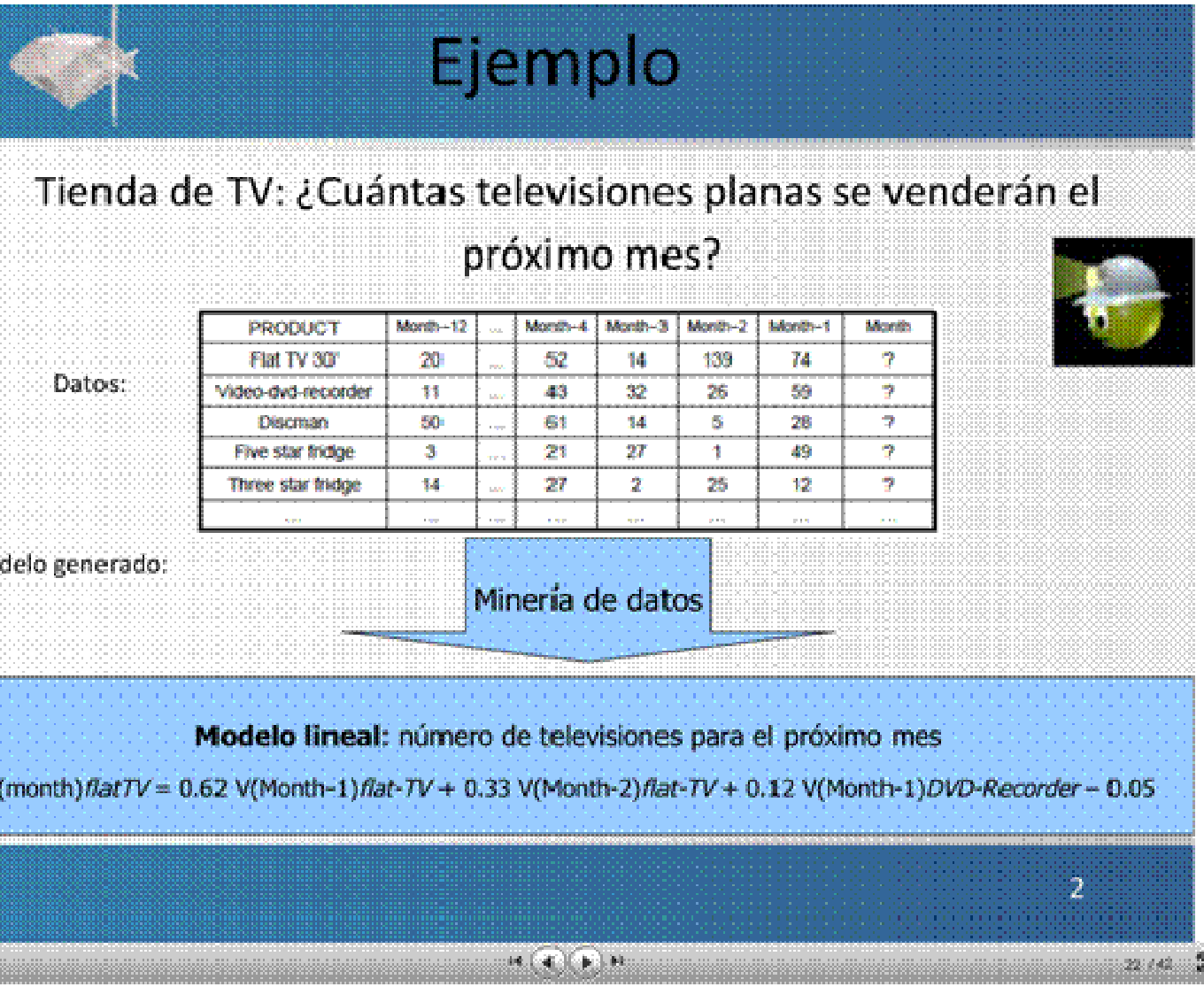
Grupo 2: Sin niños y con coche. Alto número de uniones. Pocos días enfermos. Más mujeres y en una casa alquilada

Grupo 3: Con niños, casados y con coche. Más hombres y normalmente propietarios de casa. Bajo número de uniones

EJEMPLOS

Cerón, M.A. y Gómez, H. (2010) Minería de datos

(http://www.slideshare.net/04071977/mineria-de-datos)



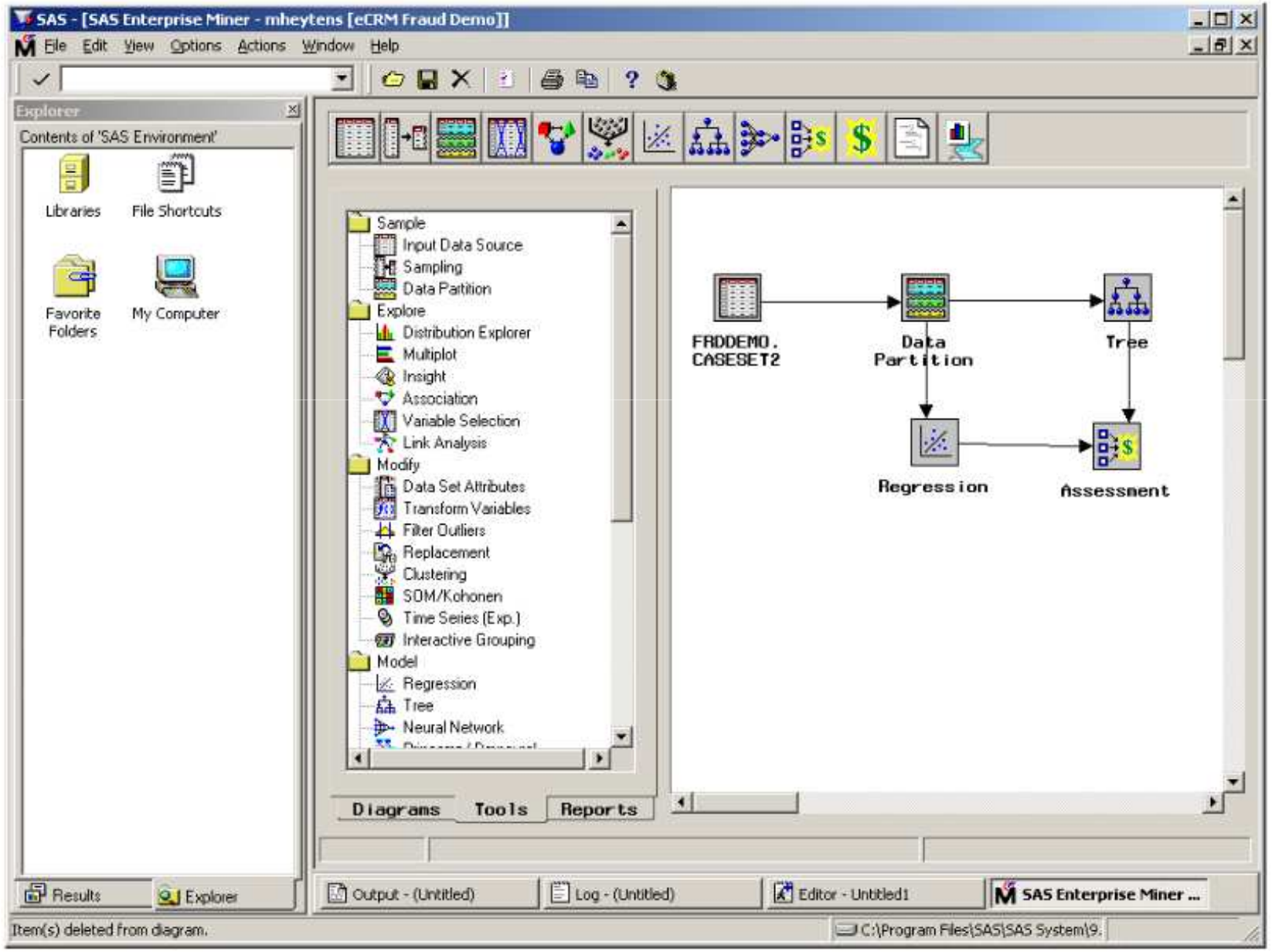
Introducción a la Minería de datos

PROGRAMAS INFORMÁTICOS

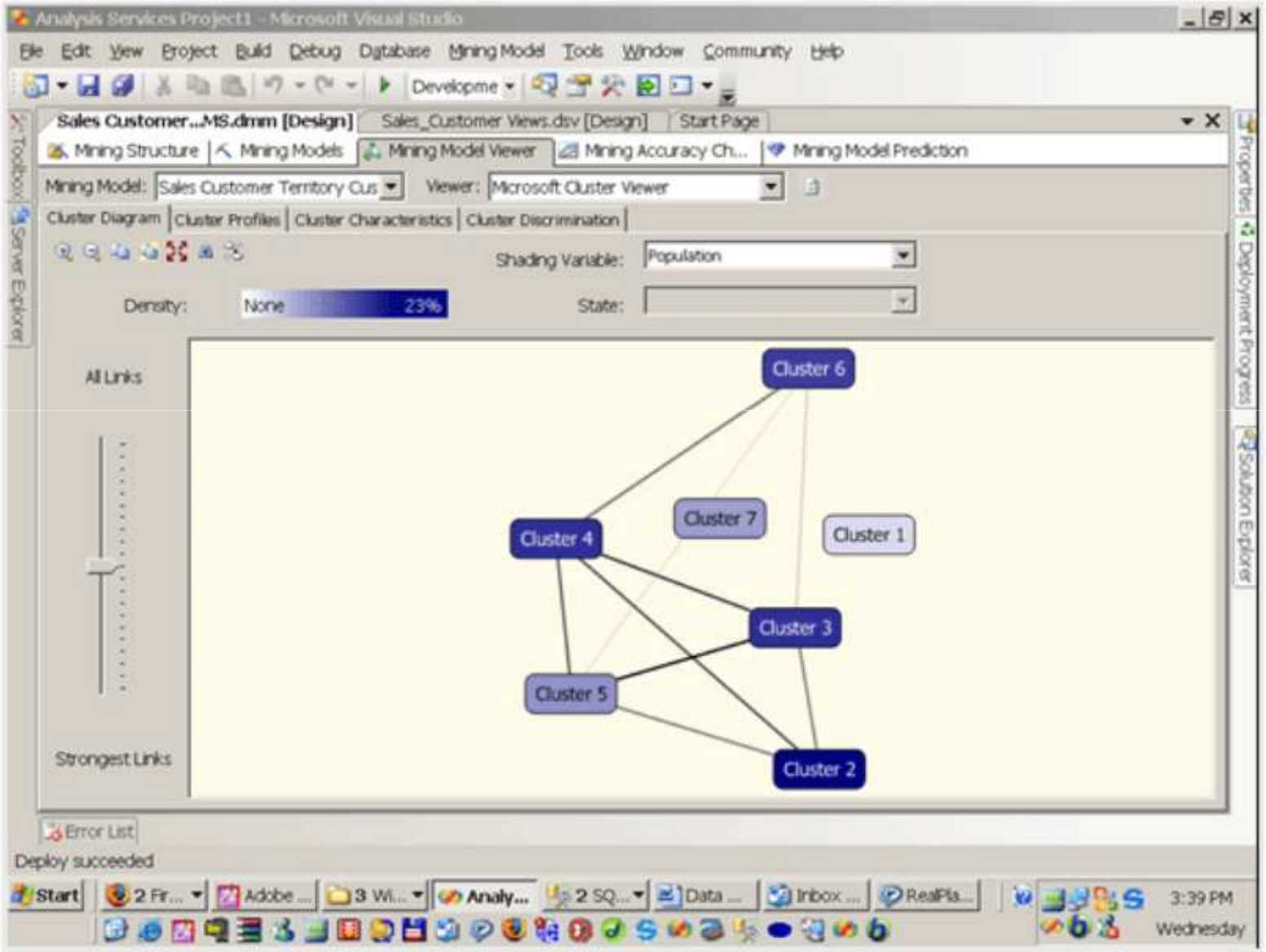
En esta última década han aparecido una serie de programas informáticos que nos han permitido analizar un gran volumen de datos

Sus diferencias radican en la presentación e implementación. Pasan por las mismas etapas: colecta de datos, depuración y análisis

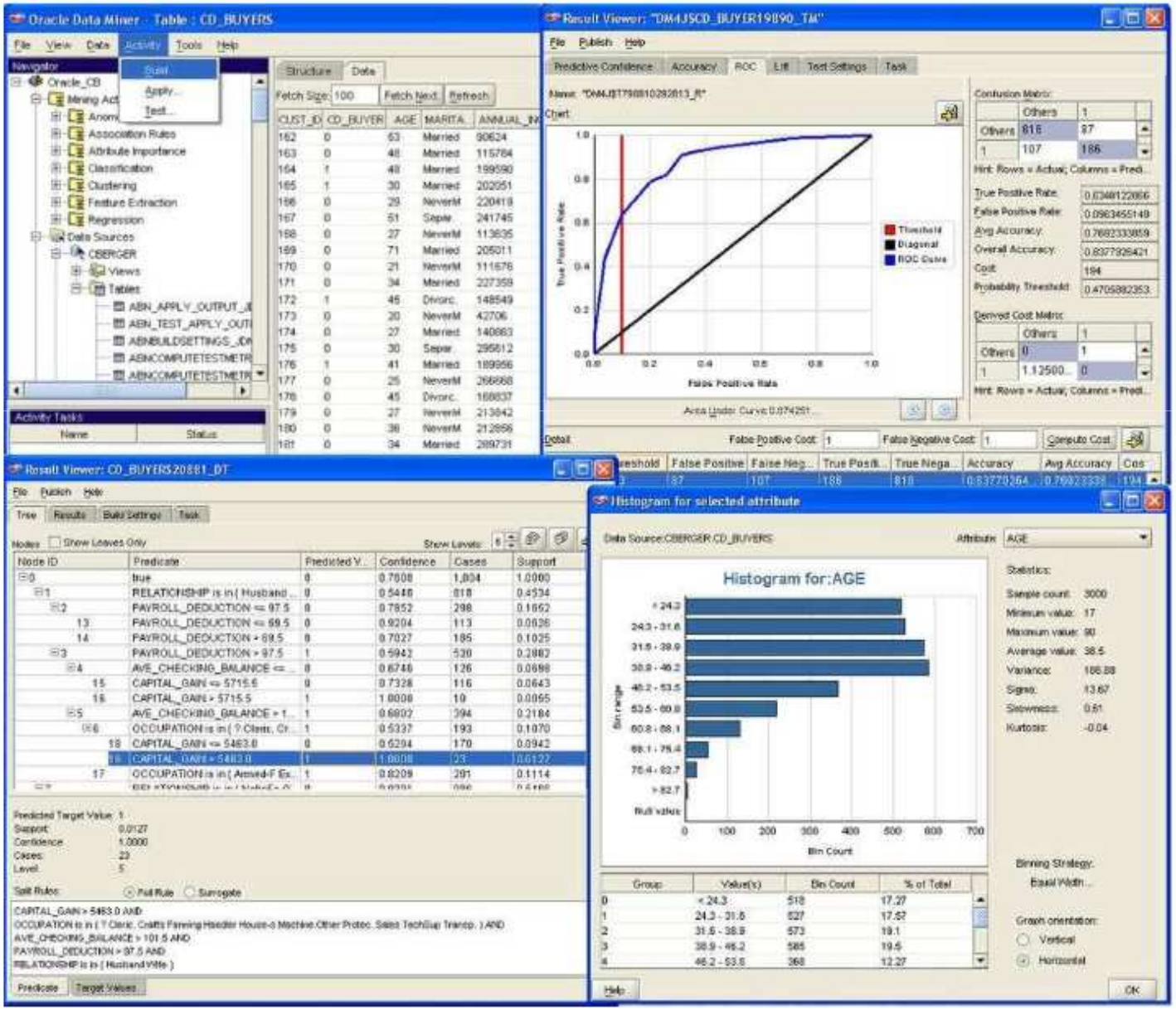
SAS System


SAS - [SAS Enterprise Miner - mheytens [eCRM Fraud Demo]]
File Edit View Options Actions Window Help
Explorer
Contents of 'SAS Environment'
Libraries
File Shortcuts
Favorite Folders
My Computer
Sample
Input Data Source
Sampling
Data Partition
Explore
Distribution Explorer
Multiplot
Insight
Association
Variable Selection
Link Analysis
Modify
Data Set Attributes
Transform Variables
Filter Outliers
Replacement
Clustering
SOM/Kohonen
Time Series (Exp.)
Interactive Grouping
Model
Regression
Tree
Neural Network
Diagrams
Tools
Reports
FRDDEMO.
CASESET2
Data
Partition
Tree
Regression
Assessment
Results
Explorer
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1
SAS Enterprise Miner ...
Item(s) deleted from diagram.
C:\Program Files\SAS\SAS System\9.

SQL Server Data Mining



ORACLE DATA MINING



Clementine


Clementine Data Mining System Version 5.0 (EVALUATION) - (c) ISL 1994-1998 - (None)
File Edit Tools Options SuperNode Displays Help
Clementine
Generated Models
EXECUTE
0.0M USED, 24.7GM FREE
Editing diagram
Sources
Record Ops
Field Ops
Graphs
Modelling
Output
Var. File
Select
Merge
Filter
Derive
Plot
Histogram
Build C5.0
Regression
Table
Analysis
ODBC
Sample
Balance
Type
Filler
Distribution
Web
Build Rule
GRI
Matrix
Statistics

Introducción a la Minería de datos

Text Mining

Si bien es cierto que existe una gran cantidad de información almacenada en bases de datos, **la existencia de un gran volumen de documentos hace necesario aplicar algún tipo de sistema de análisis.**

El análisis presenta un mayor nivel de complejidad y de dimensiones en cuanto a la categorización de texto y procesamiento de lenguaje natural. La minería de texto o text mining permite la extracción y recuperación de la información

PASW Text Mining (2010)

Internet Web Mining

Otra de las aplicaciones de la minería de datos consiste en aplicar sus técnicas a documentos y servicios Web (minería de Web) en concreto, **el análisis de datos por Internet y on line.**

¿cuáles son las páginas web más visitadas?

Bibliografía

Aluja, T (2001) *La minería de datos, entre la estadística y la inteligencia artificial*. **QÜESTIIÓ**, vol 25,3, p 479-498

Han, J. y Kamber, M. (2006) **Data mining, concepts and techniques**. USA

Hernandez Orallo J.L. (2004) **Introducción a la minería de datos**. New York: Pearson Prentice Hall

Pérez, C. Santín, D. (2007) **Minería de datos: técnicas y herramientas**. Madrid: Paraninfo

Riquelme, J.C.; Ruiz, R y Gilbert, K. (2006) *Minería de datos: conceptos y tendencias*. **Revista Iberoamericana de Inteligencia artificial**, 29, pp 11-18

Vieira Braga, L.P.; Ortiz Valencia, L.I.; Ramírez Carvajal, S.S. (2009) **Introducción a La Minería de Datos**. Rio de Janeiro: E-papers servicios editoriales

Winters, T (2006) Educational Data Mining: Collection and Analysis of Score Matrices for Outcomes- Based Assessment .USA, University of California: Riverside

¿PASAMOS A LA SEGUNDA PARTE ?

O

¿HACEMOS DESCANSO?