

Introducción al Análisis Inteligente de Datos (KD)

Tratamiento Inteligente de Datos
Master Universitario en Ingeniería Informática



**UNIVERSIDAD
DE GRANADA**

Gabriel Navarro (gnavarro@ugr.es, gnavarro@decsai.ugr.es)

Objetivos

- ❑ Entender el concepto de Minería de Datos y Extracción del Conocimiento/Tratamiento Inteligente de Datos
- ❑ Conocer áreas de aplicación de la Minería de Datos y la Extracción de Conocimiento/Tratamiento Inteligente de Datos
- ❑ Conocer la relación de la Minería de Datos con otras disciplinas
- ❑ Conocer las distintas fases que componen el proceso de Extracción de Conocimiento/Tratamiento Inteligente de Datos

Índice

- ❑ Concepto de DM y KD
- ❑ Fases de la Extracción de Conocimiento (KD)
 - Integración y recolección
 - Limpieza, transformación y selección
 - Minería de Datos
 - Evaluación
 - Difusión y uso
- ❑ Áreas de aplicación

Concepto de DM y KD

Varios conceptos/terminologías relativos al análisis de conjuntos de datos

- ❑ Inteligencia de Negocio/Business Intelligence
- ❑ Extracción de conocimiento/Knowledge Discovery
- ❑ Tratamiento Inteligente de Datos/Intelligence Data Analysis
- ❑ Minería de Datos/Data Mining
- ❑ Big Data

Diferentes autores los definen de diferentes formas

Concepto de KD y DM

Para intentar describir esos términos, tres conceptos parecidos pero no iguales:

- ☐ Datos
- ☐ Información
- ☐ Conocimiento

Dependiendo del contexto...

se usan indistintamente!

Concepto de KD y DM

Datos

- Un dato es una representación formal de un hecho, concepto o instrucción
- Son realidades concretas en su estado primario
- Los datos están constituidos por los registros de los hechos, acontecimientos, transacciones, etc.
- Son números, letras, símbolos, imágenes y sonidos, que describen objetos, condiciones o situaciones

Es algo objetivo

Concepto de KD y DM

Por ejemplo, análisis de la cesta de la compra

- ❑ Un supermercado obtiene de las cajas los **datos**:
 - 1 – True, False, False, True, False, True, True, False,...
 - 2 – False, True, False, False, True, False, False, True,...
 - 3 – False, False, True, False, True, False, False, False,...
 - 4 – False, True, True, False, True, False, False, False,...
 - 5 – True, True, False, False, False, True, False, True,...
 -

No tienen ninguna utilidad al estar desubicados

Concepto de KD y DM

Información:

- Información es un conjunto de datos organizados de tal modo que adquieren un valor adicional más allá del propio
- Ese valor es el significado que el ser humano da a los datos
- La información, de forma contraria a los datos, implica que los datos están procesados de tal manera que resulten útiles o significativos para el receptor de los mismos
- Los datos se pueden considerar la materia prima para obtener información

Es un dato... con cierta carga subjetiva

Concepto de KD y DM

Por ejemplo, análisis de la cesta de la compra

Nº	Huevos	Aceite	Pañales	Vino	Leche	Mantequilla	Salmón	Lechuga	...
1	x			x		x	x		
2		x			x			x	
3			x		x				
4		x	x		x				
5	x	x				x		x	

Al contextualizar los datos de venta se obtiene la siguiente **información** sobre la cesta de la compra

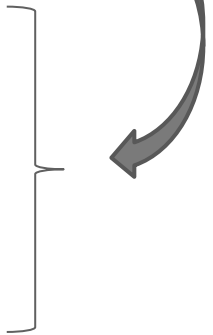
Concepto de KD y DM

Formas de obtener información de los datos:

- Contextualizando
- Categorizando, dando unidades de medida para interpretarlos
- Corrigiendo, eliminando errores e inconsistencias
- Condensando, eliminando redundancias, resumiendo

La subjetividad

Necesitamos un procesamiento para darle valor a los datos, darle un significado o situarlo en un contexto



Concepto de KD y DM

Conocimiento

- Patrones obtenidos a partir de la información
- El conocimiento es subjetivo:
 - depende de las intenciones (objetivo del aprendizaje)
 - debe ser *inteligible* para el que aprende o el que encarga el aprendizaje (usuario)
- Deben ayudar a tomar decisiones

Concepto de KD y DM

Por ejemplo, análisis de la cesta de la compra

Nº	Huevos	Aceite	Pañales	Vino	Leche	Mantequilla	Salmón	Lechuga	...
1	x			x		x	x		
2		x			x			x	
3			x		x				
4		x	x		x				
5	x	x				x		x	

Analizando los datos se obtiene la regla (**conocimiento**)
compra pañales, entonces compra leche

Concepto de KD y DM

Por ejemplo, análisis de la cesta de la compra

Nº	Huevos	Aceite	Pañales	Vino	Leche	Mantequilla	Salmón	Lechuga	...
1	x			x		x	x		
2		x			x			x	
3			x		x				
4		x	x		x				
5	x	x				x		x	

A partir de este conocimiento tomamos decisiones. Por ejemplo,
ubicar juntos leche y pañales

Posible ventaja competitiva respecto a otros

Concepto de KD y DM

Otro ejemplo, tenemos los **datos** desestructurados

- ❑ 15, 60.000, 2.200, Si, 2, ..., No
- ❑ 2, 30.000, 3.500, Si, 0, ..., Si
- ❑ 9, 9.000, 1.700, Si, 1, ..., No
- ❑ 15, 18.000, 1.900, No, 0, ..., Si
- ❑ 10, 24.000, 2.100, No, 0, ..., No
- ❑ ...

Concepto de KD y DM

Contextualizando, es **información** de clientes de un banco

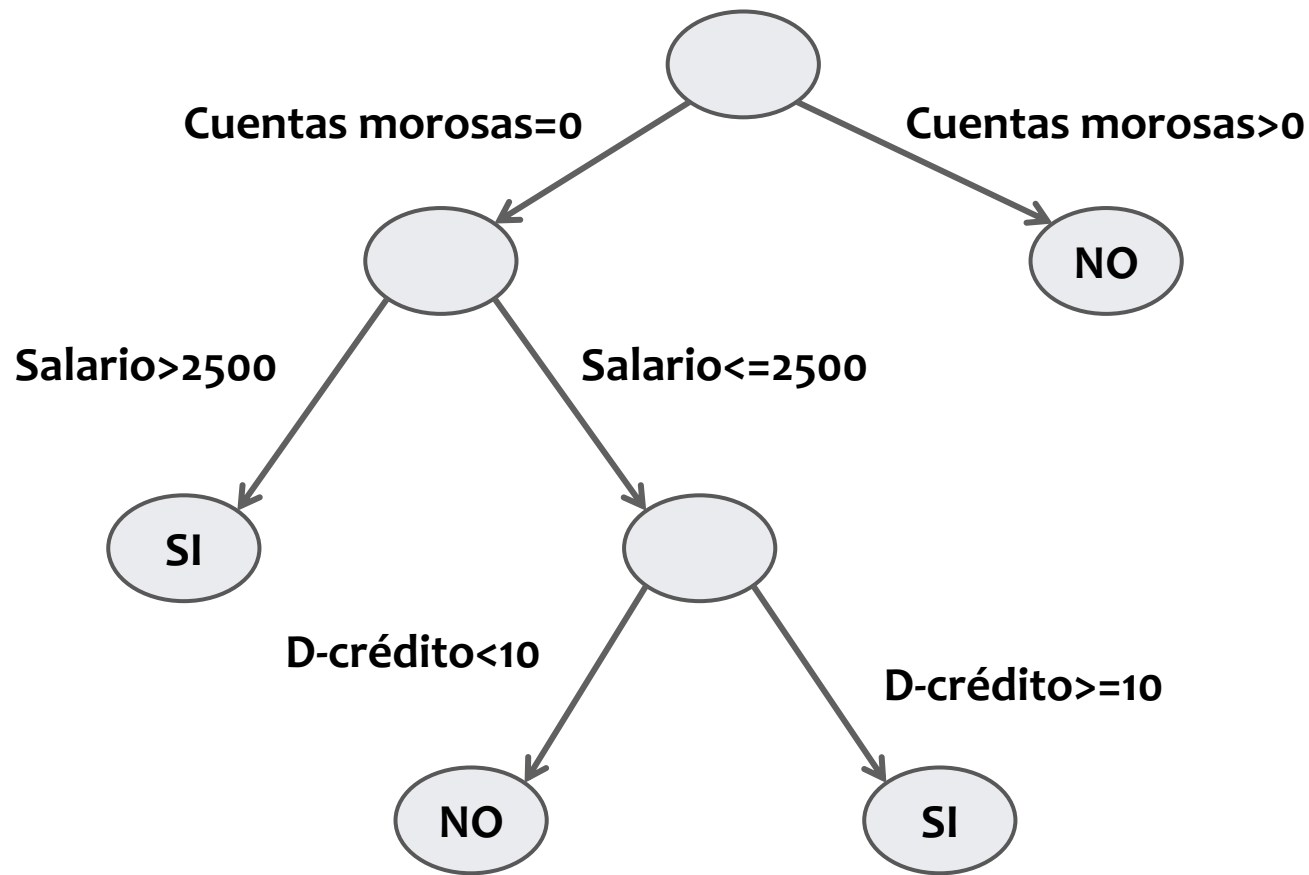
Id	D-crédito (años)	C-crédito (euros)	Salario (euros)	Casa propia	Cuentas morosas	...	Devuelve- crédito
101	15	60.000	2.200	Si	2	...	No
102	2	30.000	3.500	Si	0	...	Si
103	9	9.000	1.700	Si	1	...	No
104	15	18.000	1.900	No	0	...	Si
105	10	24.000	2.100	No	0	...	No
...

Al banco le interesa que se devuelvan los créditos

¿qué hacer con un nuevo cliente?

Concepto de KD y DM

Analizando la tabla, llegamos al siguiente **conocimiento**



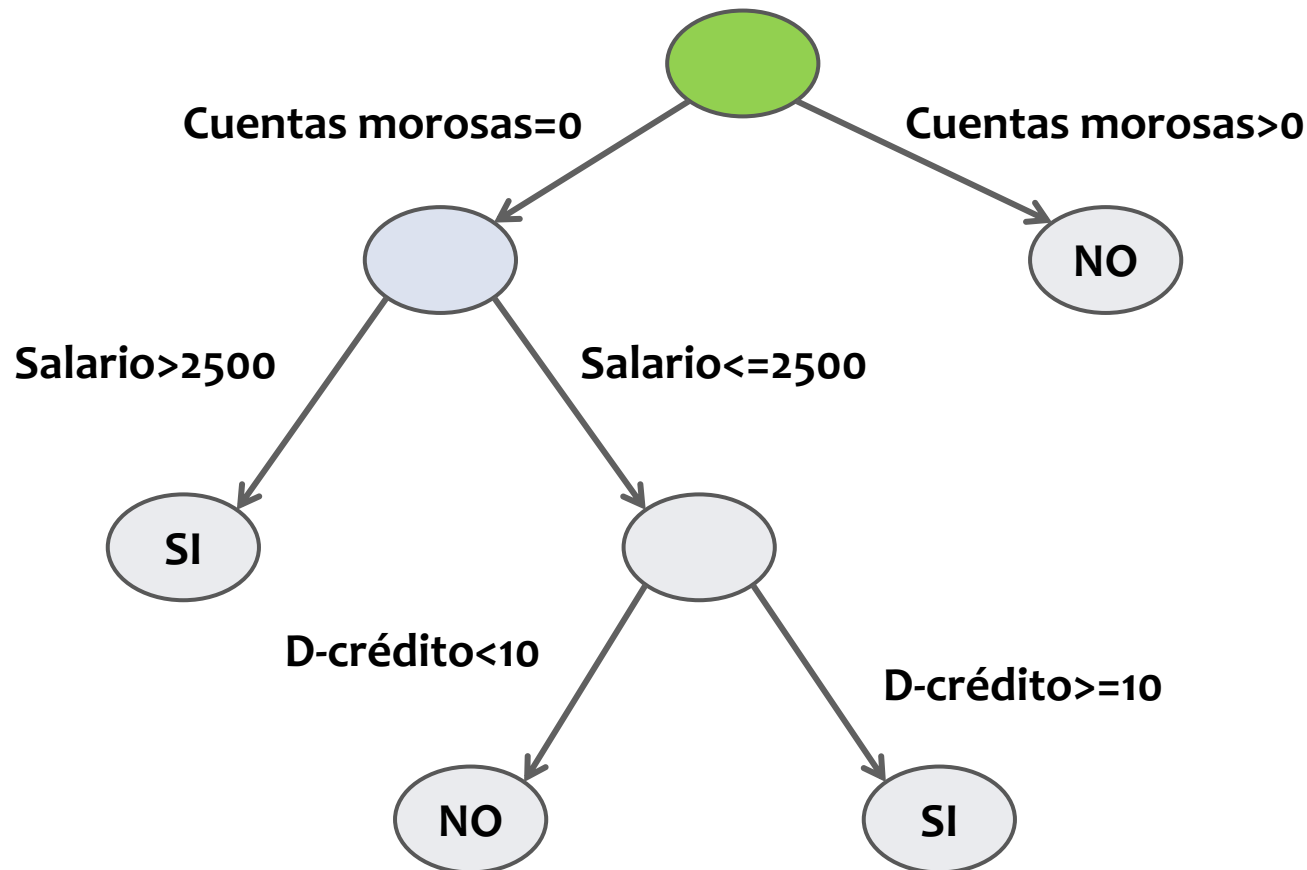
Concepto de KD y DM

Si llega un nuevo cliente, nos **ayuda a decidir** si se le concede el crédito o no

Id	D-crédito (años)	C-crédito (euros)	Salario (euros)	Casa propia	Cuentas morosas
121	20	150.000	1500	No	0

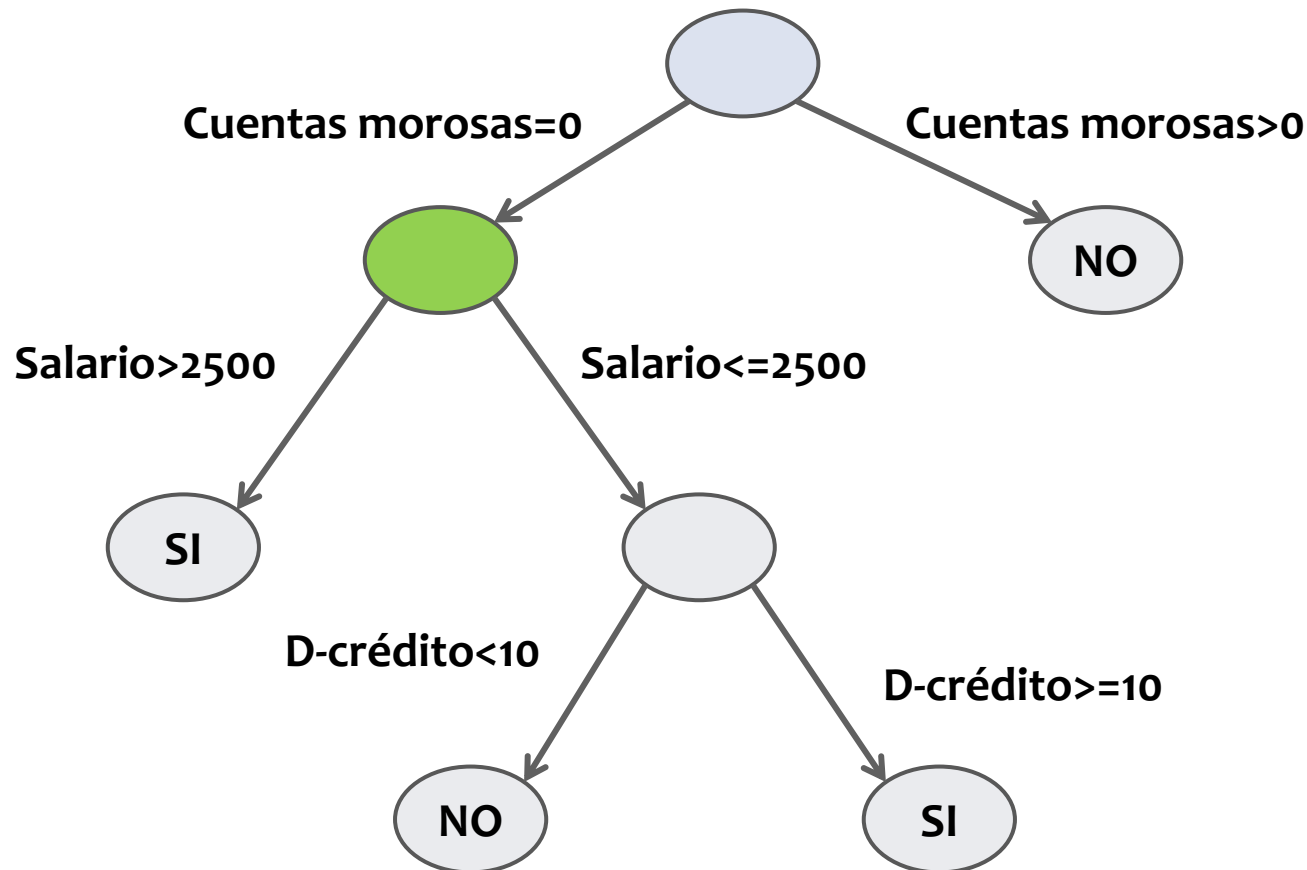
Concepto de KD y DM

Id	D-crédito (años)	C-crédito (euros)	Salario (euros)	Casa propia	Cuentas morosas
121	20	150.000	1500	No	o



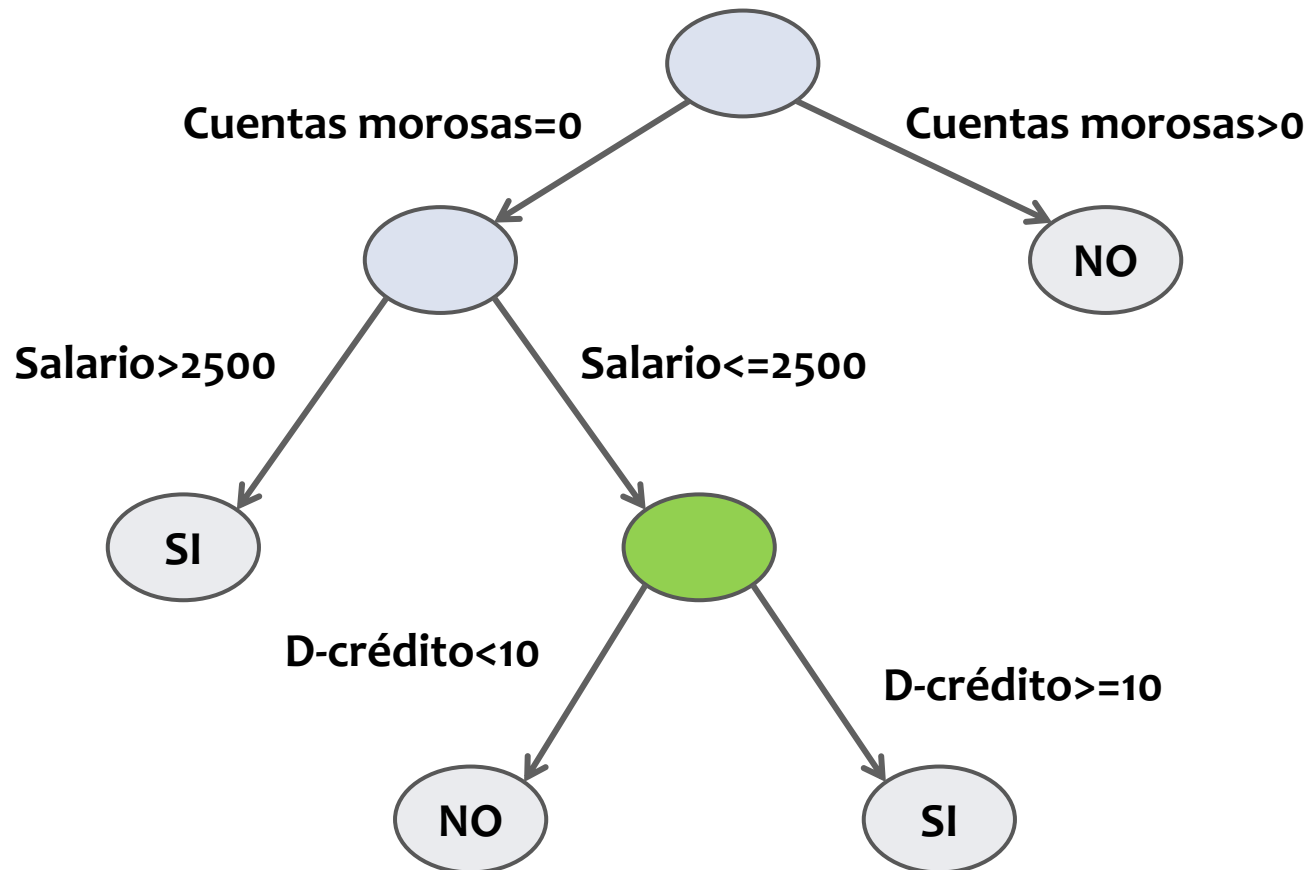
Concepto de KD y DM

Id	D-crédito (años)	C-crédito (euros)	Salario (euros)	Casa propia	Cuentas morosas
121	20	150.000	1500	No	0



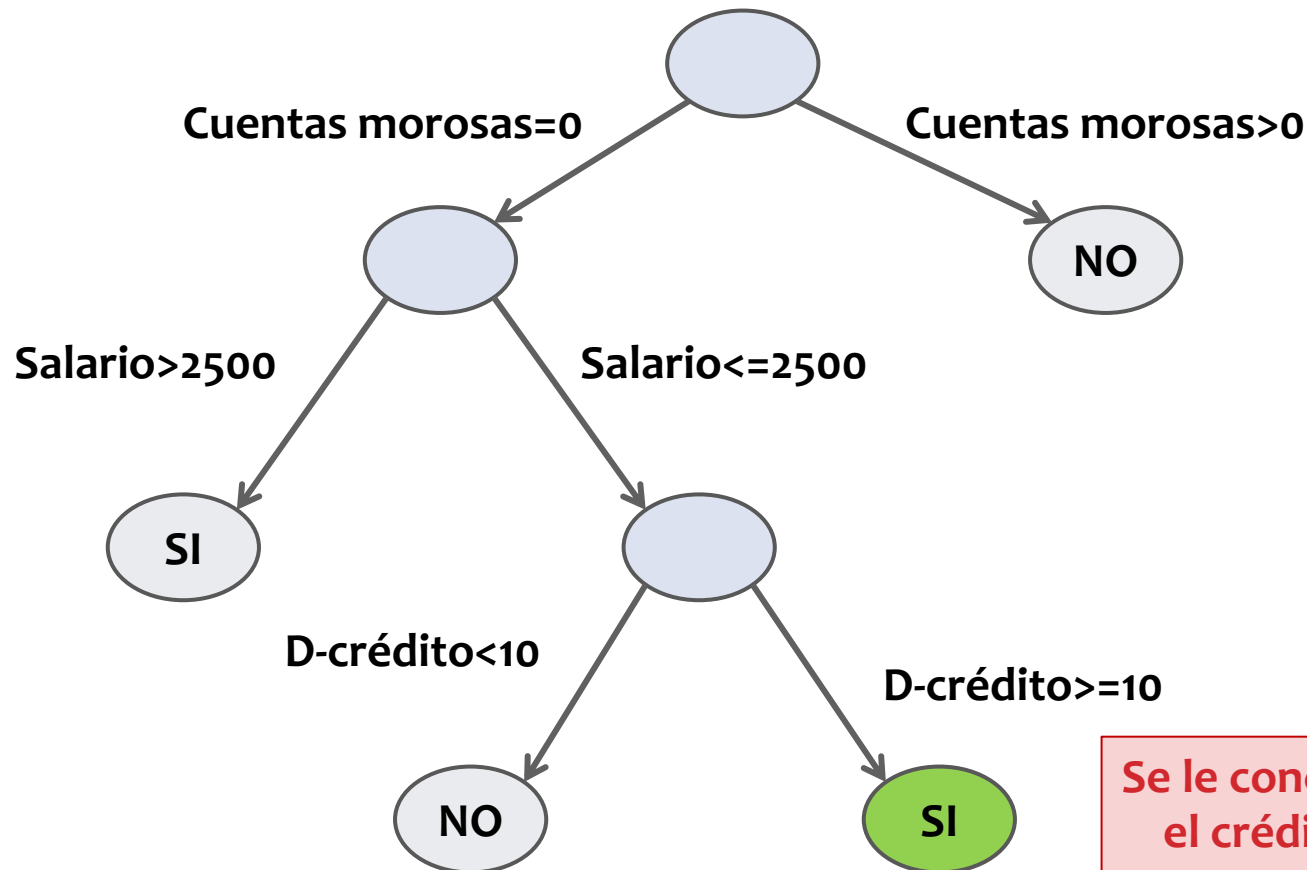
Concepto de KD y DM

Id	D-crédito (años)	C-crédito (euros)	Salario (euros)	Casa propia	Cuentas morosas
121	20	150.000	1500	No	0



Concepto de KD y DM

Id	D-crédito (años)	C-crédito (euros)	Salario (euros)	Casa propia	Cuentas morosas
121	20	150.000	1500	No	0



Concepto de KD y DM

Otro ejemplo, datos booleanos

❑ 0,1,0,1,1,0,0,0,0,1,0,0,0,1,0,0,1,0,0,0,0,...

❑ 0,0,1,1,1,0,1,0,0,1,0,0,0,1,1,0,1,0,1,0,0,...

❑ 0,1,0,0,1,0,1,0,0,1,0,0,0,1,0,0,1,0,1,1,1,...

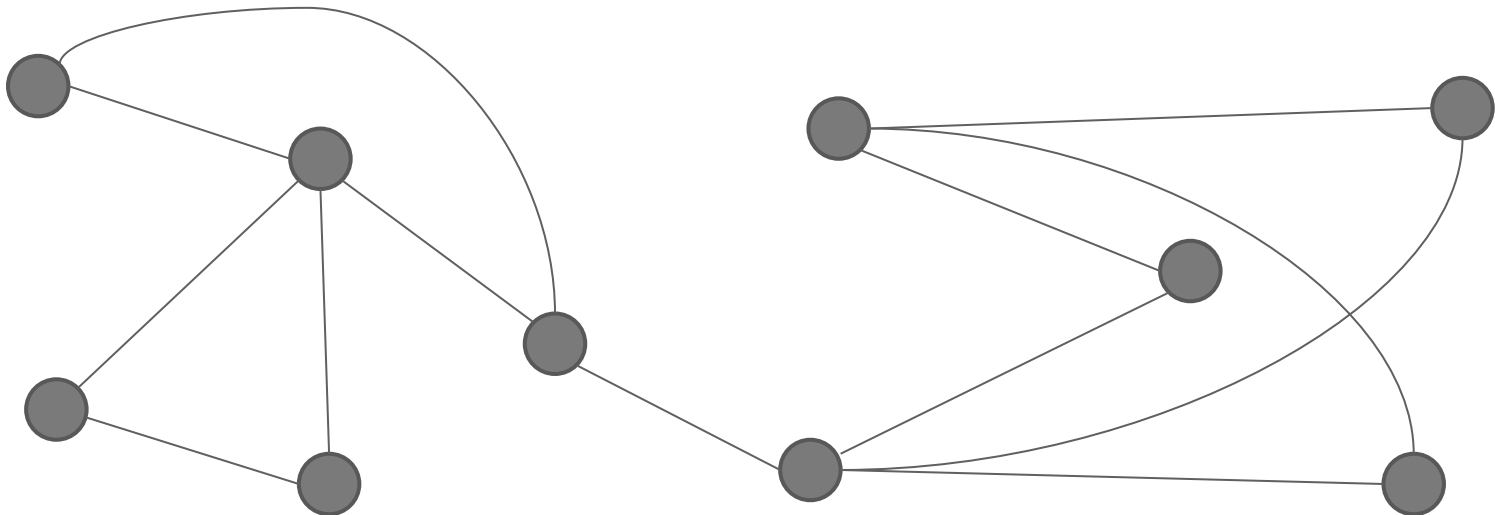
❑ 0,1,0,0,0,0,0,0,0,1,0,0,0,1,0,0,1,0,0,0,0,...

❑ 0,1,0,1,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,...

❑ ...

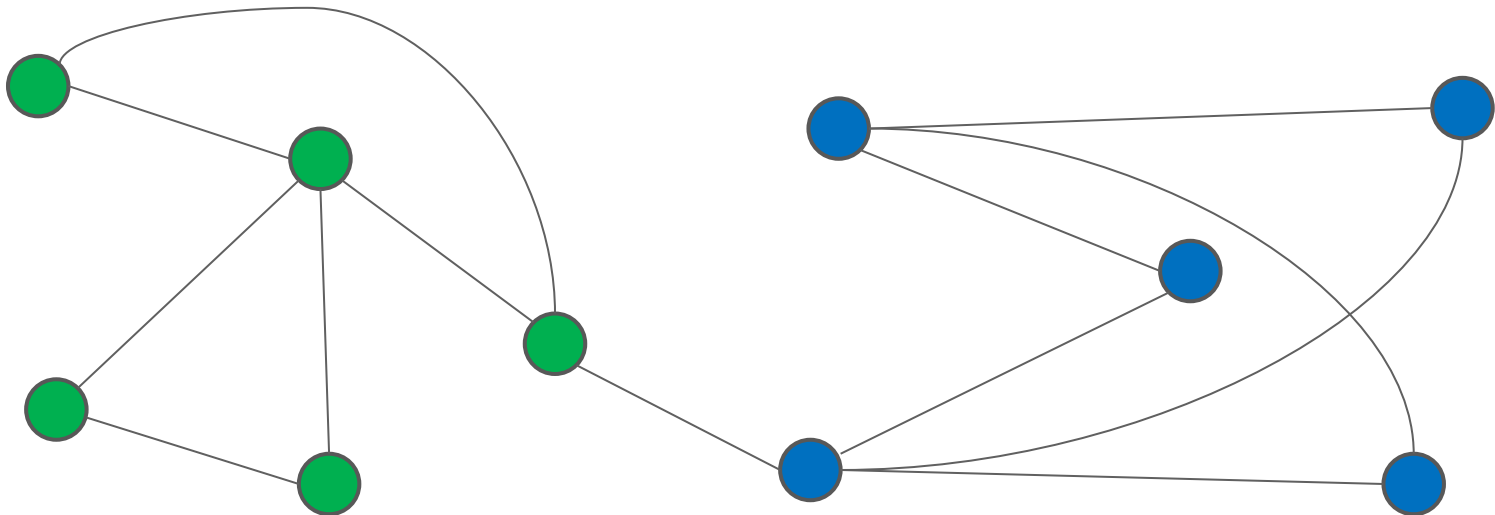
Concepto de KD y DM

Contextualizamos, matriz de adyacencia de una red social



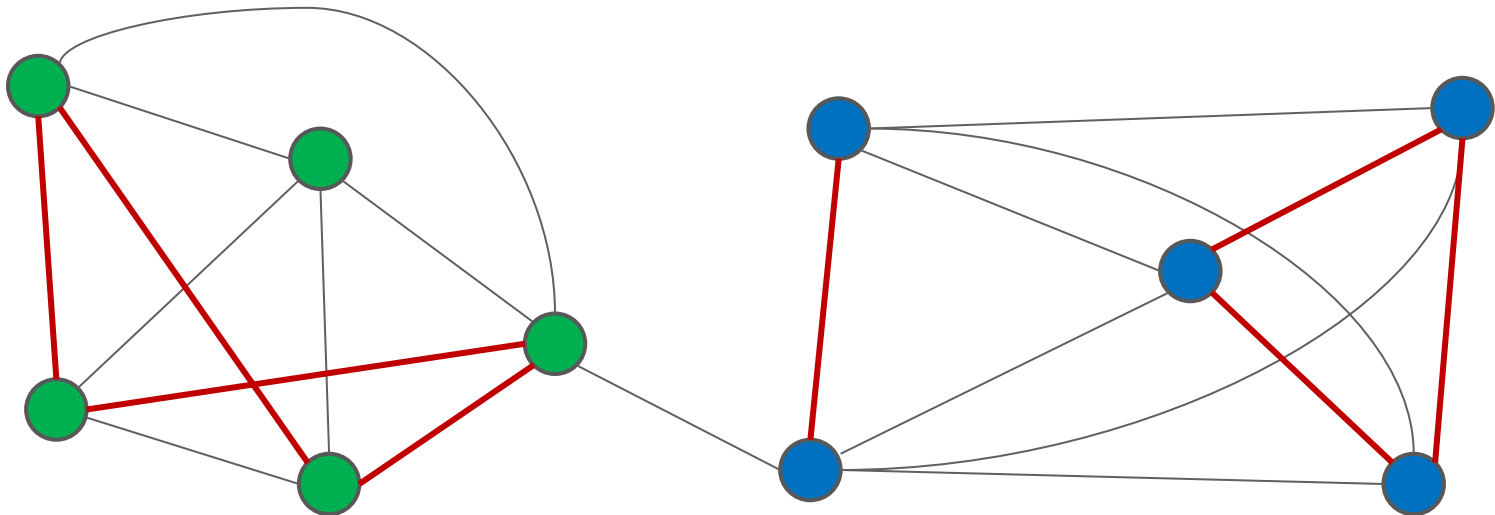
Concepto de KD y DM

Conocimiento, agrupamiento entre usuarios con amigos
“parecidos”



Concepto de KD y DM

Decisión, sugerir amigos entre usuarios del mismo grupo



Concepto de KD y DM

¿Cómo obtener ese conocimiento?



Manualmente:

- Método tradicional
- Un experto analiza e interpreta los datos
- Realiza un informe que refleje tendencias o pautas

Concepto de KD y DM

¿Cómo obtener ese conocimiento?

~~Manualmente:~~

- Metodo tradicional
- Un experto analiza e interpreta los datos
- Realiza un informe que refleje tendencias o pautas

- Caro
- Lento
- Subjetivo
- Impracticable para mucho volumen de datos

Concepto de KD y DM

¿Cómo obtener ese conocimiento?

~~Manualmente:~~

- Metodo tradicional
- Un experto analiza e interpreta los datos
- Realiza un informe que refleje tendencias o pautas

- Caro
- Lento
- Subjetivo
- Impracticable para mucho volumen de datos

Minería de Datos
Data Mining

Proceso que intenta descubrir patrones en/ a partir de grandes volúmenes de datos

Concepto de KD y DM

¿Para qué queremos esto?

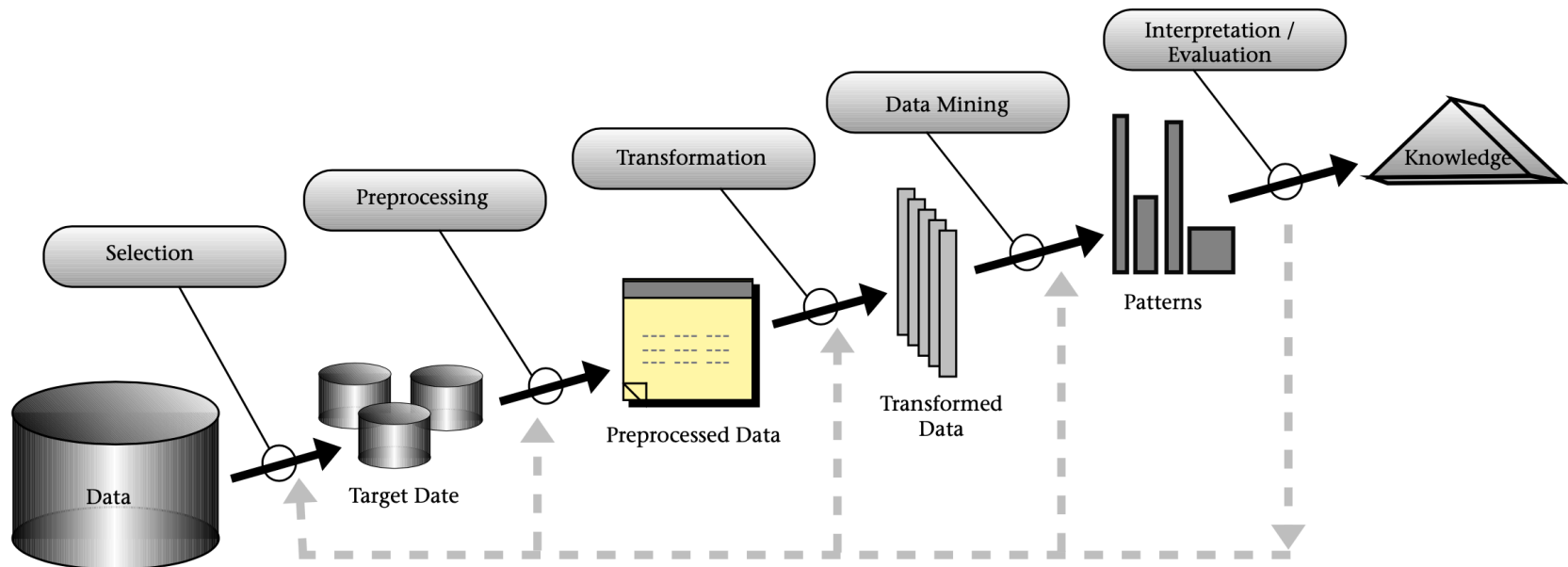
- ❑ Para realizar un **análisis exploratorio** de los datos. Sin saber nada de ellos, conocer algo interesante
 - Visualización
 - Análisis Estadístico
- ❑ Para obtener características comunes entre objetos de una base de datos/repositorio (**Técnicas descriptivas**)
 - Clustering
 - Reglas de asociación
- ❑ Para predecir valores de variables (**Técnicas predictivas**)
 - Clasificación
 - Regresión

Concepto de KD y DM

Extracción de Conocimiento (Knowledge Discovery, KD)

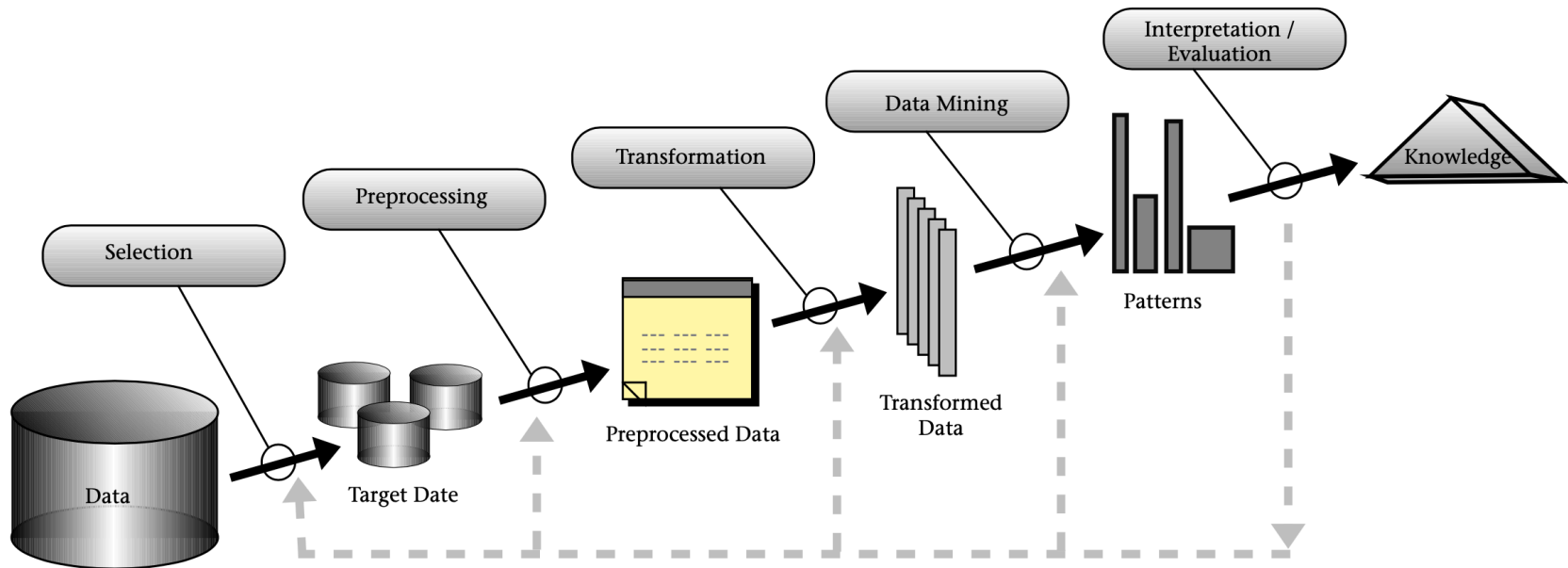
Proceso no trivial de identificar patrones **válidos** (para datos nuevos), **novedosos** (desconocidos para el usuario, **potencialmente útiles** y, en última instancia, **comprensibles** (si no, no será útil) **a partir de los datos** (no es subjetivo).

[Fayyad, U.M.; Piatetskiy-Shapiro, G.; Smith, P.; Ramasasmy, U. “Advances in Knowledge Discovery and Data Mining”, AAAI Press / MIT Press, 1996]



Concepto de KD y DM

DM es un paso del proceso de KD



Muchas veces, el preprocesamiento también se considera parte de la DM, porque se utilizan técnicas de DM

Análisis Inteligente de Datos (IDA) y KD se consideran sinónimos. KD tiene una connotación más empresarial e IDA, más científica

Big Data

Es una colección de datos grande, complejos, **muy difícil de procesar a través de herramientas de gestión y procesamiento de datos tradicionales.** Son datos cuyo volumen, diversidad y complejidad **requieren nueva arquitectura, técnicas, algoritmos y análisis** para gestionar y extraer valor y conocimiento oculto en ellos



Business Intelligence

No parece muy claro:

- a segment of information technology that comprises software systems that enable *finding, storing, organising and supplying data*; when incorporated into an information system, it enables company to utilise real-time analysis of information (*Information Technology Toolbox*)
- software that enables business users to *see and use large amounts of complex data* (e.g. multidimensional analysis, query tools, data mining tools) (*SDG Computing*)
- a category of applications and technologies for *gathering, storing, analysing, reporting on and providing access to data* to help enterprise users make *better business decisions* (Cognos)
- BI helps corporations transform their *operational data into actionable information*; helps meet query reporting and advanced analytical needs (*MycroStrategy*)
- a set of concepts and methods to *improve business decision making by using fact-based support systems* (Gartner Group)

Business Intelligence

“BI can be understood as a philosophy, which includes the strategies, processes, applications, data, products, technologies and technical architectures used to support the collection, analysis, presentation and dissemination of business information”

N. Dedić and C. Stanier, Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery

Es un conjunto de sistemas de información que trabajan de forma coordinada que ayudan a la empresa a tomar decisiones basadas en datos

- Sistemas de almacenamiento de datos (data warehouse)
- Sistemas de minería de datos (data mining)
- Herramientas de procesamiento analítico de datos (OLAP)
- Sistemas de administración de conocimiento (KBS)
- Herramientas de consulta y reporte de datos
- Tableros de información (Dashboards)
- Sistemas de ayuda a la decisión

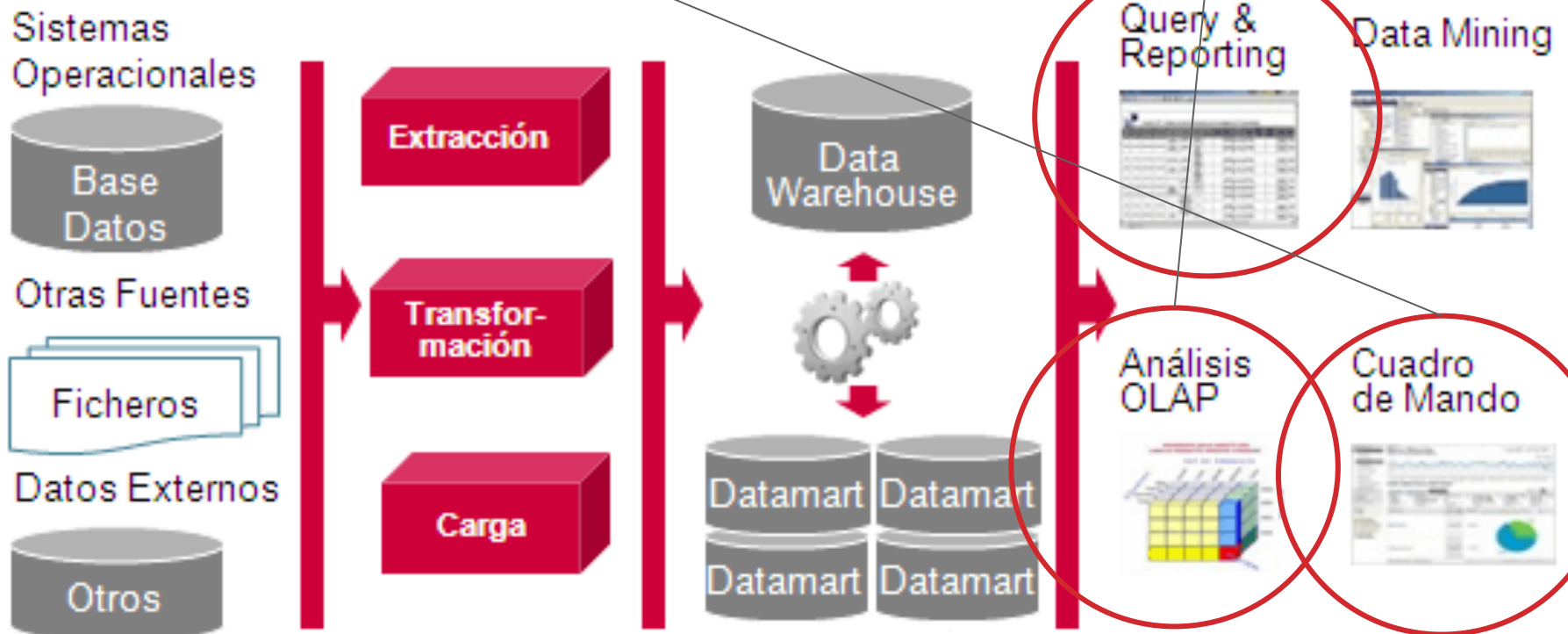
Arquitectura de un sistema de BI

Visualización de los indicadores empresariales:

- Cumplimiento
- Evaluación
- Eficiencia
- Alertas

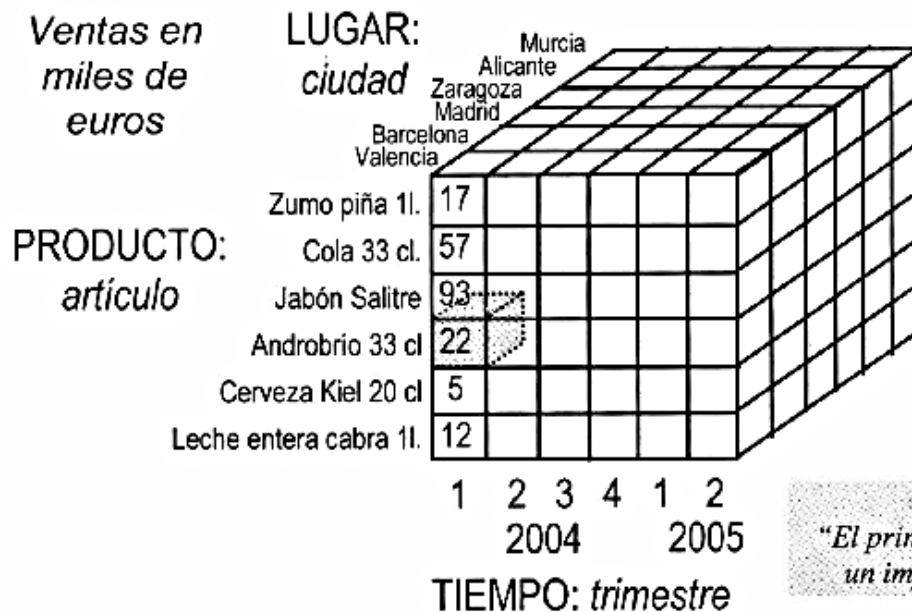
OnLine Analytical Processing:

- Análisis Multidimensional

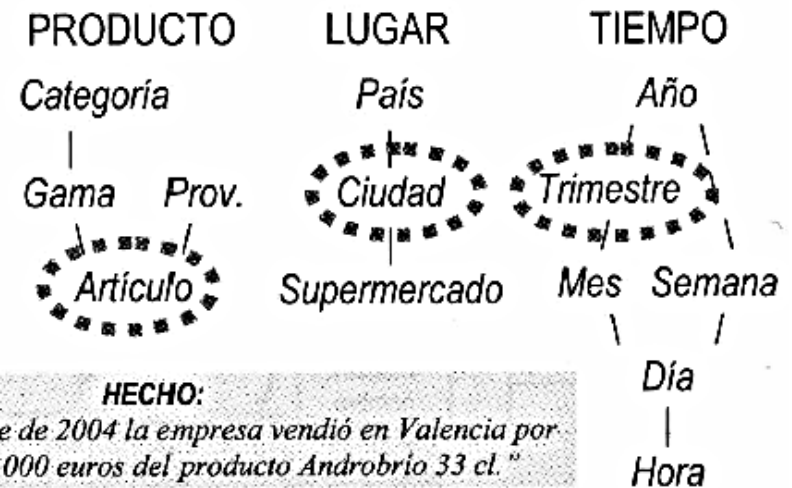


Datamart

Representación tridimensional de un datamart

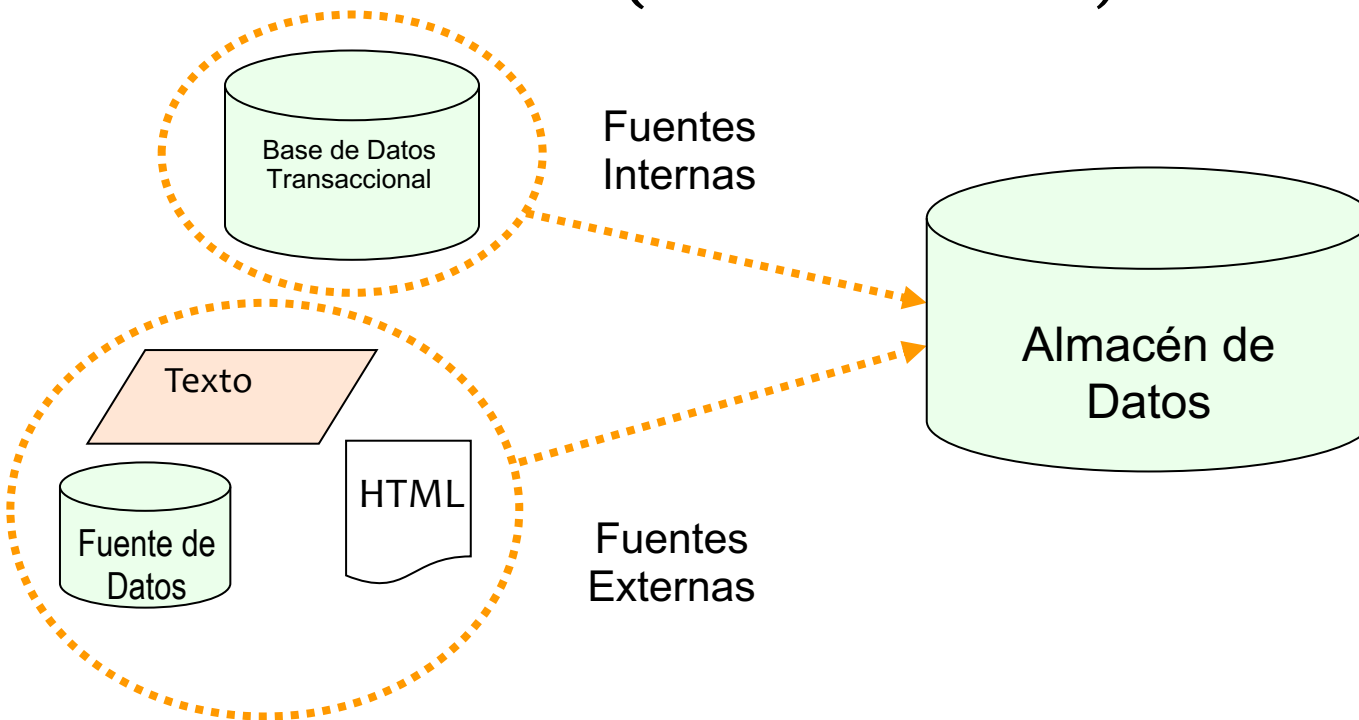


Jerarquía de dimensiones:



Concepto de DM y KD

La diversidad de los datos utilizados en BI hacen conveniente el utilizar almacenes de datos (Data Warehouse)



En esta asignatura utilizaremos solo bases de datos relacionales
KDD (Knowledge Discovery in Databases)

Concepto de DM y KD

Análisis OLAP (OnLine Analytical Processing, en DW)

- Análisis en tiempo real
- Para cruzar gran cantidad de información
- Exclusivamente **de consulta**
- Para realizar informes y resúmenes
- Requieren tiempo y recursos en BBDD

No es lo mismo OLAP que Data Mining

Análisis OLAP	Data Mining
¿Cuál es la proporción media de accidentes entre fumadores y no fumadores?	¿Cuál es la mejor predicción para accidentes?
¿Cuál es la factura telefónica media de mis clientes y de los que han dejado la compañía?	¿Dejara X la compañía? ¿Qué factores afectan a los clientes que se van?
¿Cuánto es la compra media diaria de tarjetas robadas y legítimas?	¿Cuáles son los patrones de compra asociados con el fraude de tarjetas?

Concepto de DM y KD

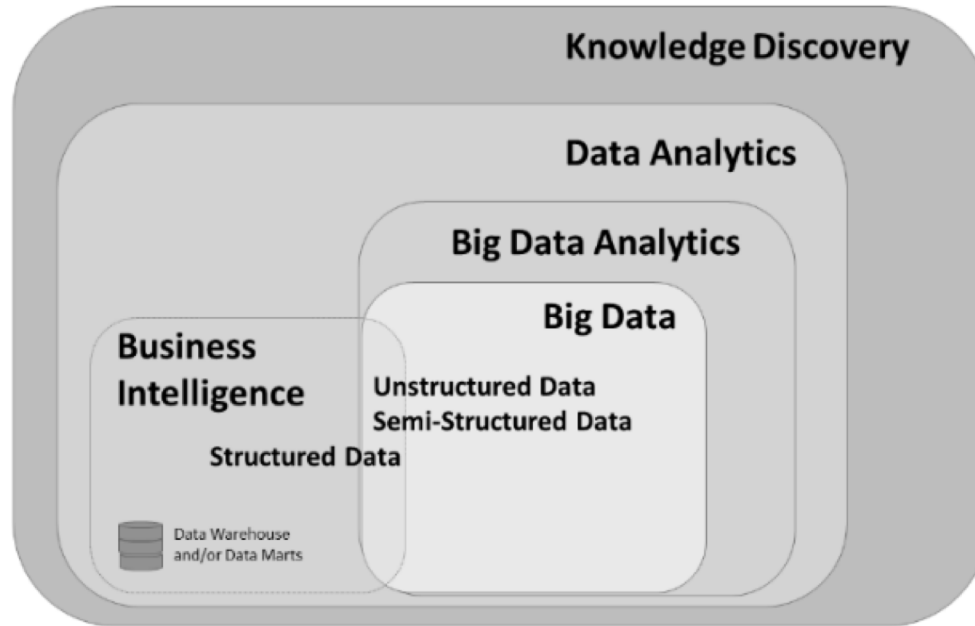
Análisis OLAP (OnLine Analytical Processing, en DW)

- Análisis en tiempo real
- Para cruzar gran cantidad de información
- Exclusivamente **de consulta**
- Para realizar informes y resúmenes
- Requieren tiempo y recursos en BBDD

Tampoco OLAP es una parte de Data Mining

- Aunque los problemas de generalización y resumen se pueden considerar dentro de DM, realmente corresponden a la primera fase del proceso de KD
- La tecnología de DW es un enfoque comercial y se centra en implementación, desempeño y visualización interactiva
- La DM es más académica y desarrolla procesos más sofisticados
- Realmente son dos etapas de un proceso que pueden realimentarse

Concepto de DM y KD



N. Dedić and C. Stanier, Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery, Lecture Notes in Business Information Processing, vol 285. Springer, Cham. https://doi.org/10.1007/978-3-319-58801-8_10

Fases KD - Preparación de datos

La preparación de datos es importante porque

- ❑ los **datos reales pueden ser impuros**, pueden conducir a la extracción de patrones/reglas poco útiles
 - Datos incompletos, datos con ruido, datos inconsistentes
- ❑ puede generar un **conjunto de datos más pequeño** que el original, lo cual puede mejorar la eficiencia del proceso de Minería de Datos
 - Selección relevante, reducción de datos
- ❑ **genera “datos de calidad”**, los cuales pueden conducir a patrones/reglas de calidad

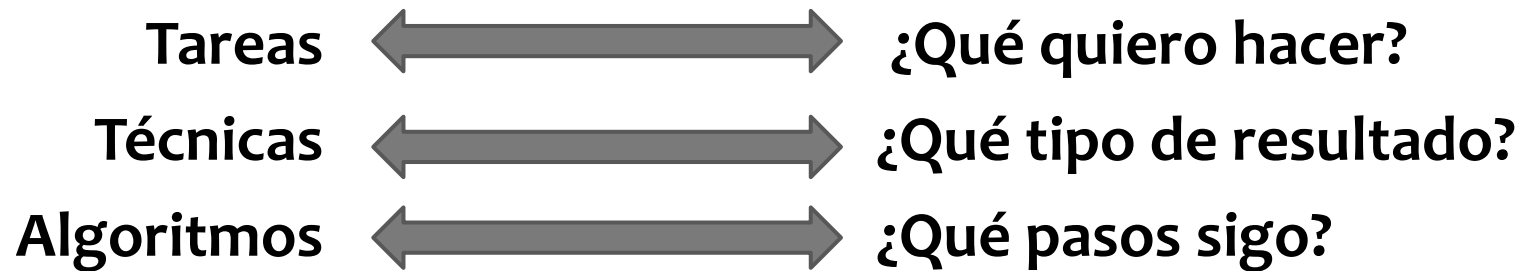
Fases KD - Preparación de datos

Las BBDD reales en la actualidad suelen contener datos ruidosos, perdidos y/o inconsistentes, fundamentalmente por su gran tamaño

En qué consiste la preparación de datos:

- ❑ **Integración** desde distintos orígenes/formato
- ❑ **Limpieza**. Eliminación de ruido e inconsistencias
- ❑ **Transformación**
 - Discretización, normalización,...
- ❑ **Reducción/Selección**
 - Tamaño, dimensión, eliminación redundancias,...

Fases KD - Minería de Datos



Fases KD - Minería de Datos

☐ Tarea predictiva

- Clasificación
- Regresión
- Series temporales
- Redes Bayesianas
- ...

¿Cuáles serán las ventas el año próximo?

¿Es esta transacción fraudulenta?

¿Qué tipo de seguro es más probable que contrate el cliente X?

Minería de Datos

Por ejemplo, en vista de los días anteriores,

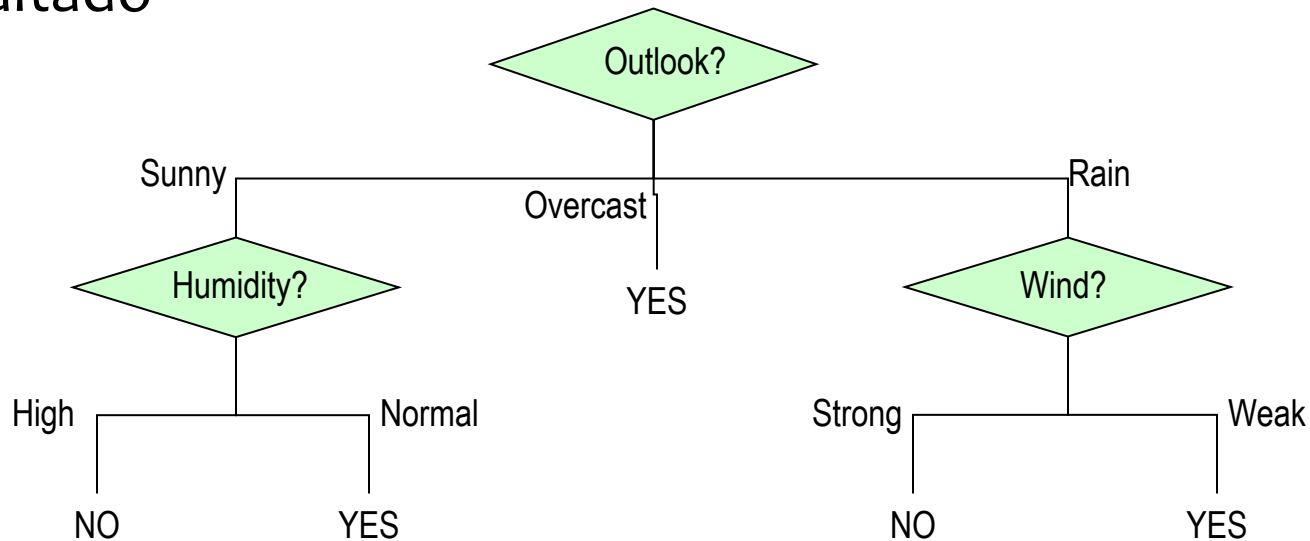
¿jugaré hoy al tenis?

(clasificar hoy entre categoría SI o categoría NO)

Day	Sky	Temperature	Humidity	Wind	PlayTennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Minería de Datos

- ❑ Seleccionamos técnica –Árbol de decisión
- ❑ Seleccionamos algoritmo –Algoritmo ID3
- ❑ Resultado

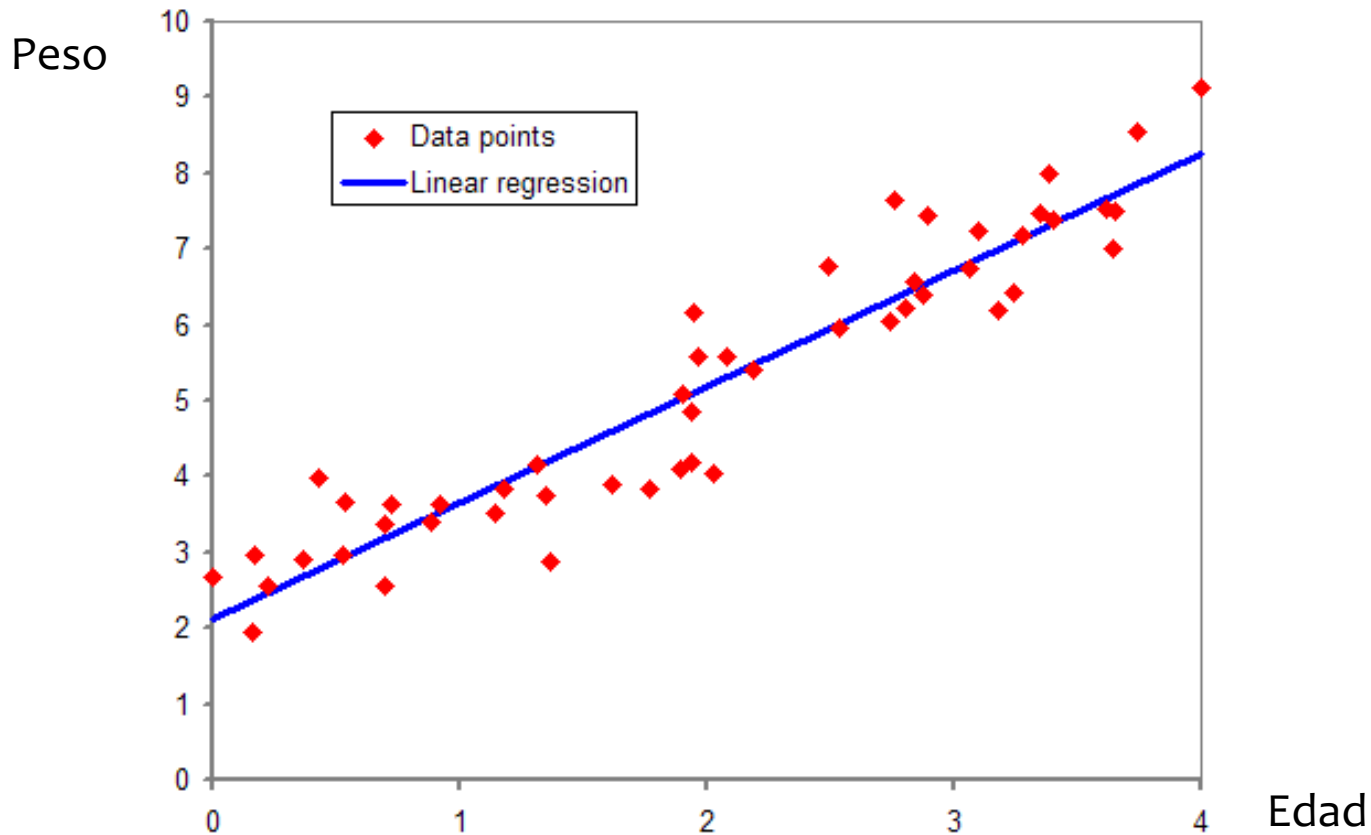


- ❑ Si (Outlook = sunny, Temperature = hot, Humidity = high, Wind = strong) entonces NO

Minería de Datos

Por ejemplo, teniendo en cuenta los datos de varios niños

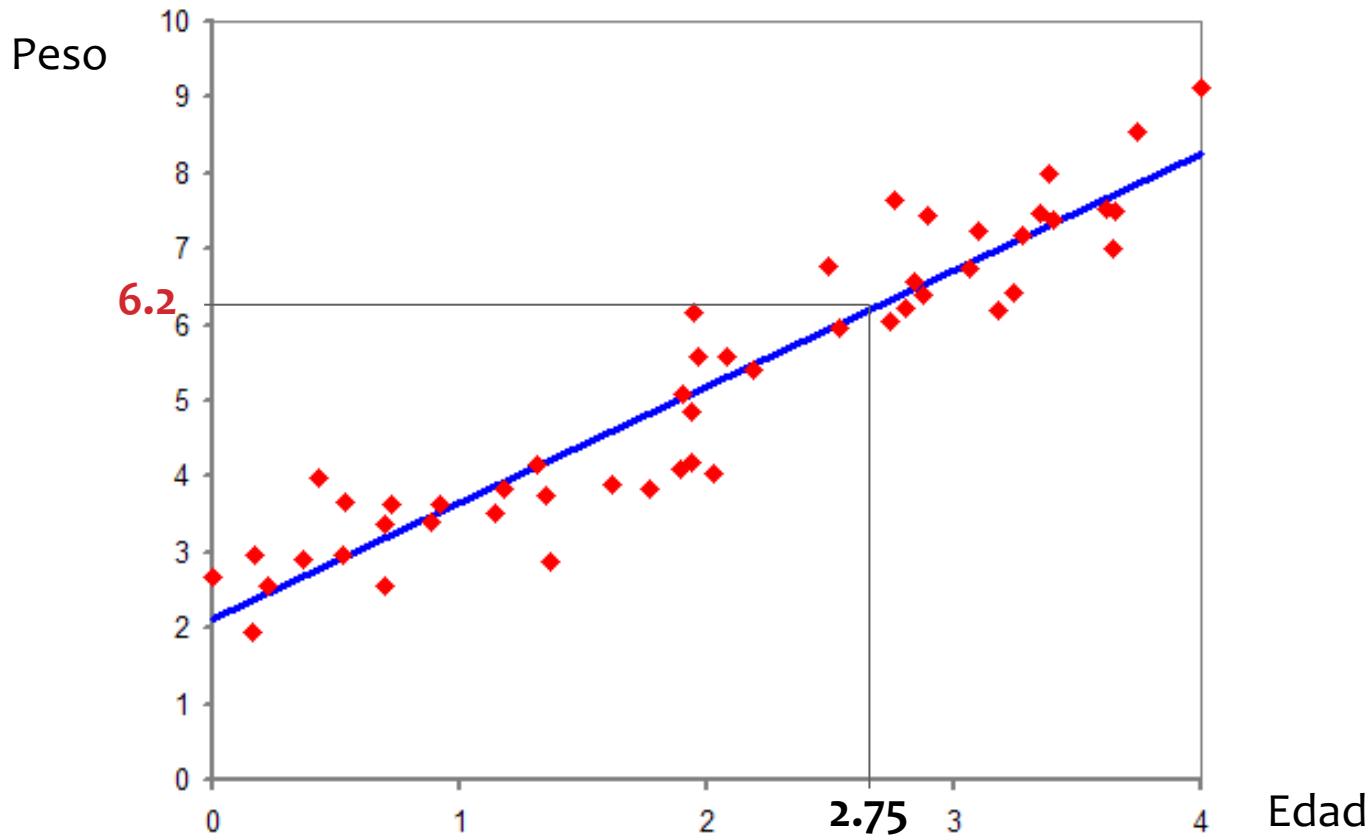
¿Cuánto estimo que pesará mi bebé?



Minería de Datos

Por ejemplo, teniendo en cuenta los datos de varios niños

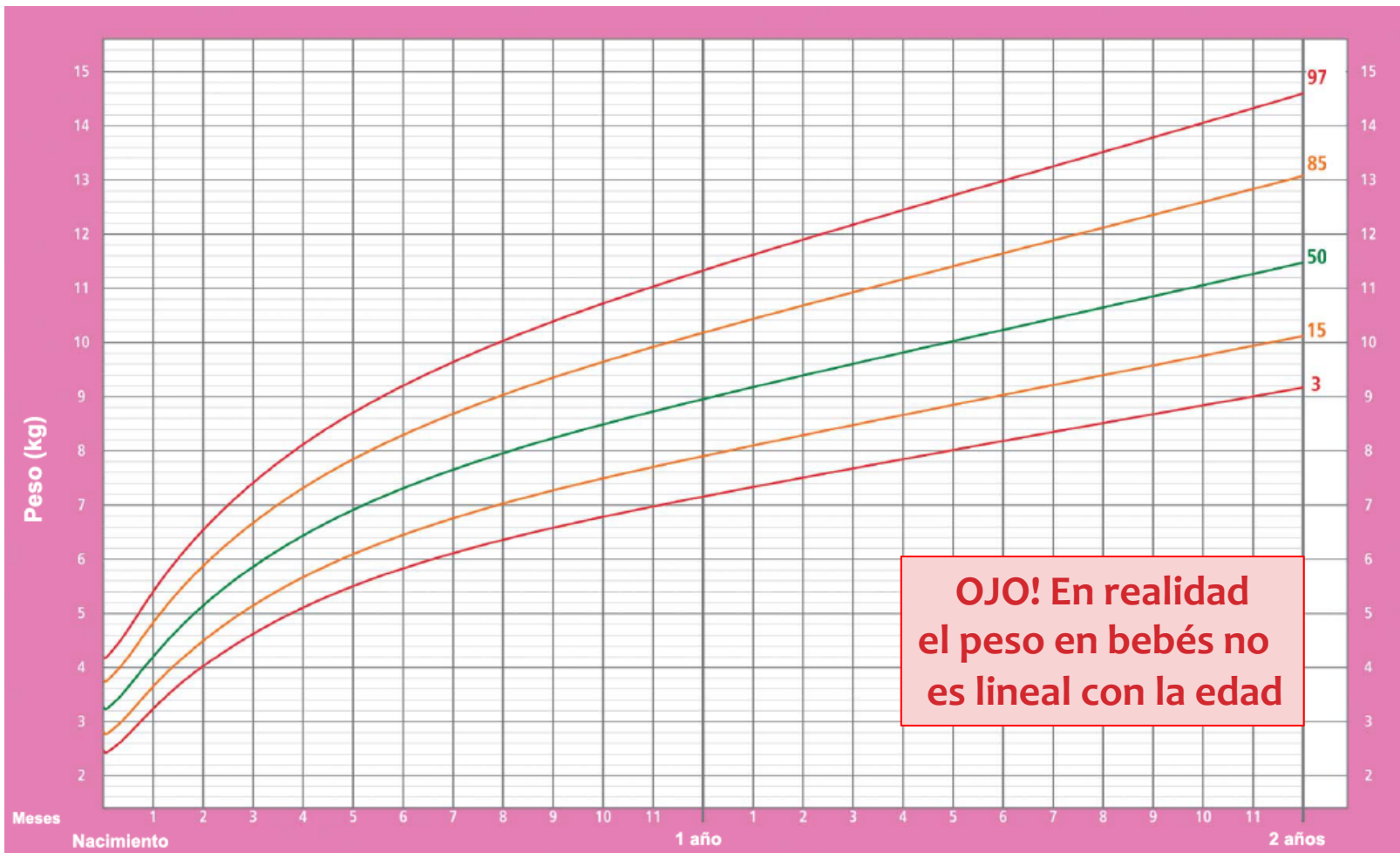
¿Cuánto estimo que pesará mi bebé?



Minería de Datos

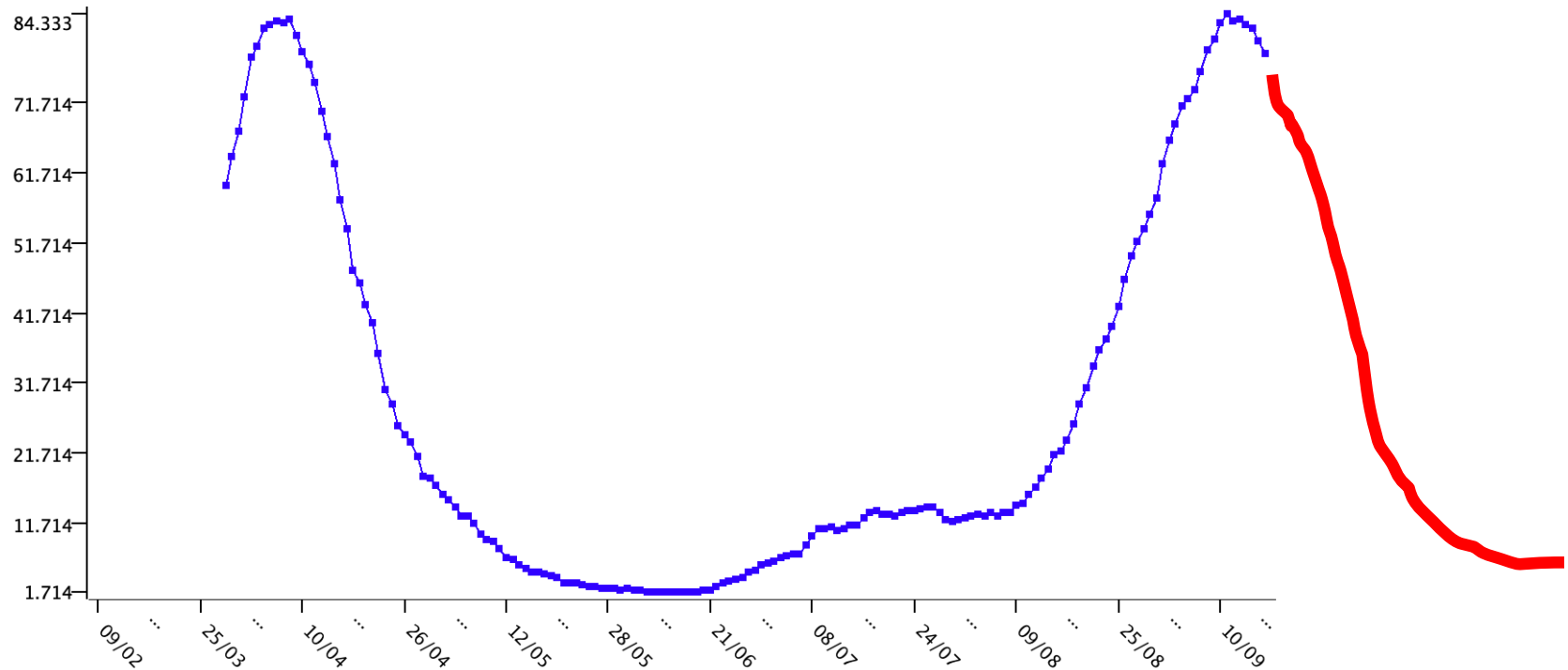
Por ejemplo, teniendo en cuenta los datos de varios niños

¿Cuánto estimo que pesará mi bebé?



Minería de Datos

A la vista de los datos de contagio en Granada, ¿qué se espera en las próximas semanas?



Minería de Datos

❑ Tarea descriptiva

- Clustering
- Correlaciones
- Reglas de asociación
- ...

Los clientes que compran pañales suelen comprar cerveza

El tabaco y el alcohol son los factores más importantes en la enfermedad XXXX

Los clientes sin televisión y con bicicleta tienen características muy diferenciadas del resto

Minería de Datos

¿Cómo categorizar a nuestros empleados?

#Ej	Sueldo	Casado	Coche	Hijos	Alq/Prop	Sindic.	Bajas/Año	Antigüedad	Sexo
1	10000	Sí	No	0	Alquiler	No	7	15	H
2	20000	No	Sí	1	Alquiler	Sí	3	3	M
3	15000	Sí	Sí	2	Prop	Sí	5	10	H
4	30000	Sí	Sí	1	Alquiler	No	15	7	M
5	10000	Sí	Sí	0	Prop	Sí	1	6	H
6	40000	No	Sí	0	Alquiler	Sí	3	16	M
7	25000	No	No	0	Alquiler	Sí	0	8	H
8	20000	No	Sí	0	Prop	Sí	2	6	M
9	20000	Sí	Sí	3	Prop	No	7	5	H
10	30000	Sí	Sí	2	Prop	No	1	20	H
11	50000	No	No	0	Alquiler	No	2	12	M
12	8000	Sí	Sí	2	Prop	No	3	1	H
13	20000	No	No	0	Alquiler	No	27	5	M
14	10000	No	Sí	0	Alquiler	Sí	0	7	H
15	8000	No	Sí	0	Alquiler	No	3	2	H

Minería de Datos

Tarea – Clustering

Algoritmo – K-medias

cluster 1: 5 examples

Sueldo : 22600

Casado : No -> 0.8

Sí -> 0.2

Coche : No -> 0.8

Sí -> 0.2

Hijos : 0

Alq/Prop : Alquiler -> 1.0

Sindic. : No -> 0.8

Sí -> 0.2

Bajas/Año : 8

Antigüedad : 8

Sexo : H -> 0.6

M -> 0.4

cluster 2: 4 examples

Sueldo : 22500

Casado : No -> 1.0

Coche : Sí -> 1.0

Hijos : 0

Alq/Prop : Alquiler -> 0.75

Prop -> 0.25

Sindic. : Sí -> 1.0

Bajas/Año : 2

Antigüedad : 8

Sexo : H -> 0.25

M -> 0.75

cluster 3: 6 examples

Sueldo : 18833

Casado : Sí -> 1.0

Coche : Sí -> 1.0

Hijos : 2

Alq/Prop : Alquiler -> 0.17

Prop -> 0.83

Sindic. : No -> 0.67

Sí -> 0.33

Bajas/Año : 5

Antigüedad : 8

Sexo : H -> 0.83

M -> 0.17

- GRUPO 1: Sin hijos y de alquiler. Poco sindicados. Muchas bajas.
- GRUPO 2: Sin hijos y con coche. Muy sindicados. Pocas bajas. Normalmente de alquiler y mujeres.
- GRUPO 3: Con hijos, casados y con coche. Propietarios. Poco sindicados. Hombres.

Minería de Datos

Cesta de la compra ¿Qué deducimos de ella?

	VINO "EL CABEZÓN"	GASEOSA "CHISPA"	VINO "TÍO PACO"	HORCHATA "XUFER"	BIZCOCHOS "GOLOSO"	GALLETAS "TRIGO"	CHOCOLATE "LA VACA"
T1	1	1	0	0	0	1	0
T2	0	1	1	0	0	0	0
T3	0	0	0	1	1	1	0
T4	1	1	0	1	1	1	1
T5	0	0	0	0	0	1	0
T6	1	0	0	0	0	1	1
T7	0	1	1	1	1	0	0
T8	0	0	0	1	1	1	1
T9	1	1	0	0	1	0	1
T10	0	1	0	0	1	0	0

Minería de Datos

Modelo – Reglas de asociación

Algoritmo -- Apriori

	VINO "EL CABEZÓN"	GASEOSA "CHISPA"	VINO "TÍO PACO"	HORCHATA "XUFER"	BIZCOCHOS "GOLOSO"	GALLETAS "TRIGO"	CHOCOLATE "LA VACA"
T1	1	1	0	0	0	1	0
T2	0	1	1	0	0	0	0
T3	0	0	0	1	1	1	0
T4	1	1	0	1	1	1	1
T5	0	0	0	0	0	1	0
T6	1	0	0	0	0	1	1
T7	0	1	1	1	1	0	0
T8	0	0	0	1	1	1	1
T9	1	1	0	0	1	0	1
T10	0	1	0	0	1	0	0

SI bizcochos "Goloso" Y horchata "Xufer" ENTONCES galletas "Trigo"

SI bizcochos "Goloso" Y galletas "Trigo" ENTONCES horchata "Xufer"

SI galletas "Trigo" Y horchata "Xufer" ENTONCES bizcochos "Goloso"

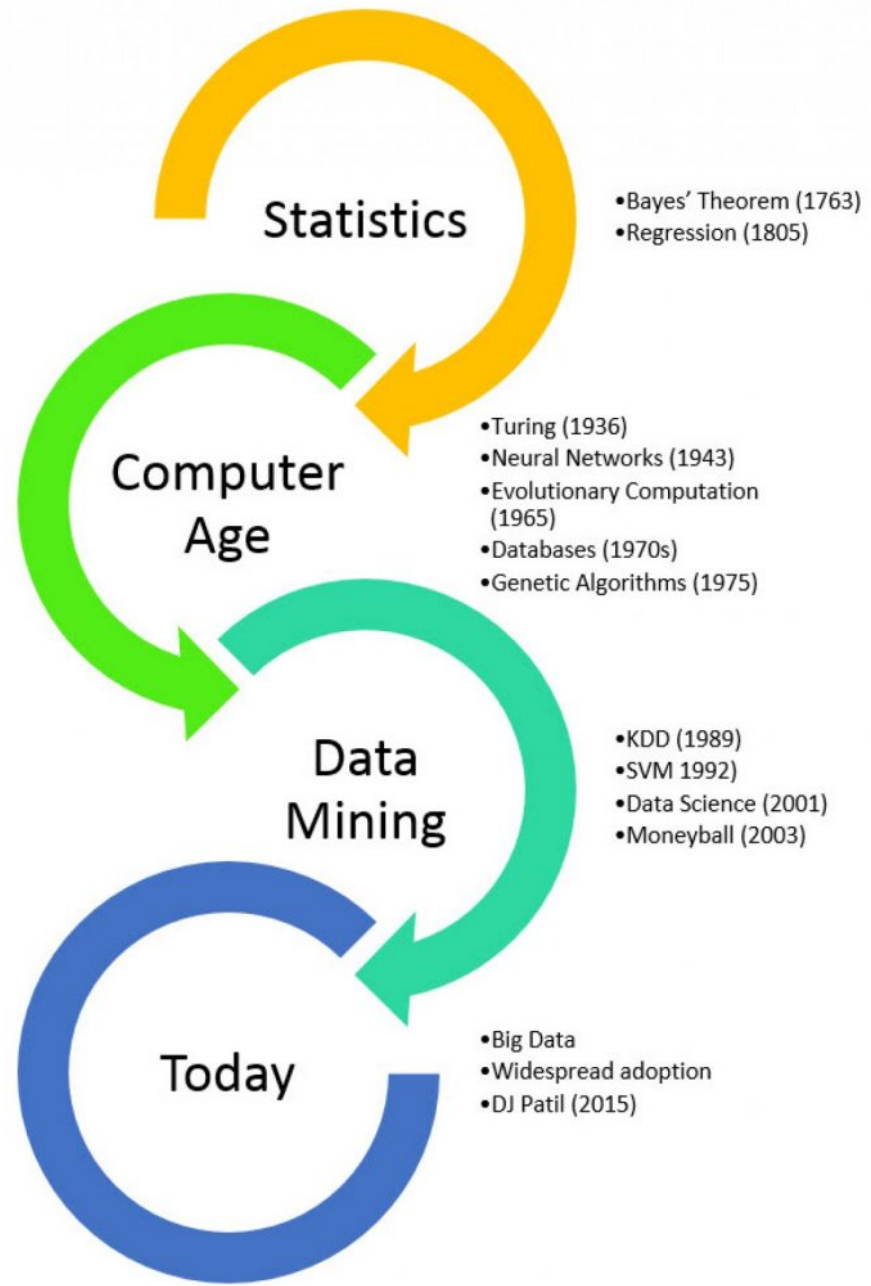
Minería de Datos

Compras a través de internet (asociaciones)

- ❑ Una persona compra un libro (producto) en Amazon
- ❑ Tarea: Recomendar otros libros (productos) que esa persona pueda comprar
- ❑ Amazon calcula reglas de asociación basándose en las compras de libros: clientes que compran “Advances in Knowledge Discovery and Data Mining”, también compran “Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations”
- ❑ El programa de recomendación es bastante exitoso

Minería de Datos

Data Mining



Fases de KD - Evaluación

- ❑ La fase de Data Mining puede producir varias hipótesis de modelos
- ❑ Es necesario establecer **qué modelos son los más válidos**
- ❑ Criterios: los patrones descubiertos deben ser
 - precisos,
 - comprensibles, e
 - interesantes (útiles, novedosos)

Se necesita establecer un **metodología** de evaluación y unas **medidas de evaluación**

Fases de KD - Evaluación

Las medidas de la evaluación dependen mucho del modelo:

- Clasificación: precisión predictiva (%acierto)
- Regresión: Error cuadrático medio
- Agrupamiento: Medidas de cohesión y separación entre grupos
- Reglas de asociación: cobertura, confianza...

La interpretación de los mejores modelos (visualización, simplicidad, posibilidad de integración, ventajas colaterales,...) ayuda a la selección del modelo(s) final(es)

Fases de KD - Difusión y uso

Una vez construido y validado el modelo puede utilizarse:

- para recomendar acciones
- para aplicar el modelo a diferentes conjuntos de datos

En cualquier caso, es necesario:

- Difusión: Elaboración de informes para su distribución
- Utilización del nuevo conocimiento de forma independiente
- Incorporación a sistemas ya existentes
- comprobar con el conocimiento ya utilizado para evitar inconsistencias y posibles conflictos

La monitorización del sistema en acción dará lugar a nuevos casos que realimentarán el ciclo del KDD. Las conclusiones iniciales pueden variar, invalidando el modelo adquirido

Áreas de aplicación

❑ *Aplicaciones empresariales / industriales*

- Toma de decisiones en banca, seguros, finanzas, marketing, control de calidad, retención de clientes, predicción, políticas de acción (sanidad, etc.), ...

❑ *Aplicaciones en investigación científica*

- Medicina, astronomía, geografía, genética, bioquímica, meteorología, etc.

Áreas de aplicación

☐ Análisis de riesgo en banca

- Detectar patrones de uso fraudulento en tarjetas
- Estudio de concesión de créditos y/o tarjetas
- Determinación del gasto en tarjeta por grupos
- Identificar reglas de comportamiento del mercado de valores a partir de históricos

☐ Análisis de riesgo en seguros

- Predicción de clientes propensos a suscribir nuevas pólizas
- Identificar grupos/patrones de riesgo
- Identificar tendencias de comportamiento fraudulento

☐ Ambos

- Identificación de clientes leales, identificación de fuga de clientes

Áreas de aplicación

❑ Medicina / diagnóstico

- Identificación de terapias para diferentes enfermedades
- Estudio de factores de riesgo en distintas patologías
- Segmentación de pacientes en grupos afines
- Gestión hospitalaria y planificación temporal de salas, urgencias,...
- Recomendación priorizada de fármacos para una misma patología
- Estudios en genética (ADN,...)
- Selección de embriones en reproducción artificial

Áreas de aplicación

- ❑ **Web Mining – Text Mining** / Minería de datos web – Minería de Textos
 - Realización de resúmenes
 - Minería de datos web en un sitio de e-comercio, generaría análisis del comportamiento y perfiles del visitante
 - En función de esto y de su actividad, generar perfiles de cliente y estudiar posibilidades de venta cruzada (*cross-selling*)
 - Recuperación de información (*Information Retrieval*)
 - Agrupamiento de documentos-páginas web

Importancia científica

❑ Revistas de investigacion

- ACM Transactions on Knowledge Discovery from Data
- IEEE Transactions on Knowledge and Data Engineering
- Data Mining and Knowledge Discovery
- Explorations Data & Knowledge Engineering
- Knowledge and Information Systems
- Machine Learning
- ...

Importancia científica

□ Congresos

- KDD: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
- ICDM: IEEE International Conference on Data Mining
- PKDD: European Conference on Principles of Data Mining and KD
- SDM: SIAM International Conference on Data Mining
- ICDE: International Conference on Data Engineering
- CIKM: ACM International Conference on Information and Knowledge Management
- ADMA: International Conference on Advanced Data Mining and Applications
- PAKDD: Pacific-Asia Conference on Knowledge Discovery and Data Mining
- MLDM: International Conference on Machine Learning and Data Mining
- DS: International Conference on Discovery Science
- ICDM: Industrial Conference on Data Mining

Bibliografía

- Introducción a la Minería de Datos. José Hernández Orallo, M.José Ramírez Quintana, Cèsar Ferri Ramírez. Pearson, 2004.
Capítulos 1,2,3,4,5 y 6