



Tópicos Avanzados de Base de Datos.

UNIVERSIDAD DIEGO PORTALES.

Ayudantía Minería de Datos

- ▶ Profesor: Alejandro Romero Frantzen.
- ▶ Curso: Tópicos Avanzados de Base de Datos.
- ▶ Ayudante: Álvaro Cáceres Aravena.
- ▶ Correo: ayudantiaudp.bi@gmail.com

Minería de Datos.

► ¿Qué es minería de datos?

El **Data Mining** (Minería de datos), es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Objetivo

El objetivo general del proceso de Minería de Datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior.



Proceso de descubrimiento del conocimiento en base de datos.
Knowledge Discovery in Databases KDD.

Características

El proceso común de Minería de Datos se compone de cuatro etapas:

- ▶ **Determinación de los objetivos:** El cliente determina qué objetivos quiere conseguir gracias al uso del *Data Mining*.
- ▶ **Procesamiento de los datos:** Selección, limpieza, enriquecimiento, reducción y transformación de la base de datos.
- ▶ **Determinación del modelo:** Primero se debe hacer un análisis estadístico de los datos y después visualización gráfica de los mismos.
- ▶ **Análisis de los resultados:** En este paso se verifican si los resultados obtenidos son coherentes.

Aplicaciones

- ▶ Utiliza los métodos de la Inteligencia Artificial, aprendizaje automático, estadística y sistemas de bases de datos.
- ▶ Las herramientas de Minería de Datos predicen tendencias futuras y comportamientos.
- ▶ Además puede responder a preguntas que consumirían demasiado tiempo para resolverlas.
- ▶ Minería de datos es utilizado en **sistemas expertos**, que proveen herramientas típicas de soporte de decisión.

Técnicas Data Mining

Las técnicas para la Minería de Datos son:

- ▶ **Redes neuronales.**
- ▶ **Regresión lineal.**
- ▶ **Árboles de decisión.**
- ▶ **Modelos estadísticos.**
- ▶ **Reglas de asociación.**
- ▶ **Agrupamiento o *Clustering*.**

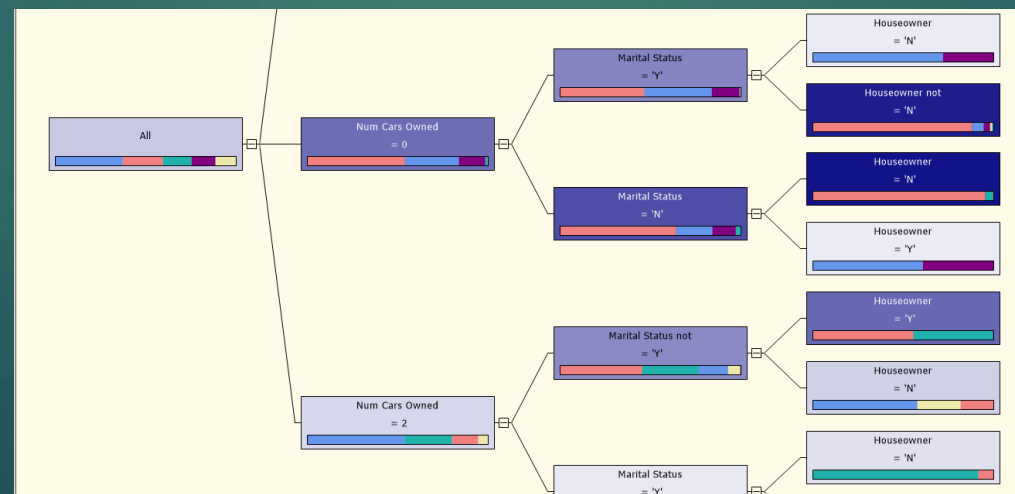


Técnicas Data Mining

- ▶ **Redes neuronales:** Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida.
- ▶ **Regresión lineal:** Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.

Técnicas Data Mining

- ▶ **Árboles de decisión:** Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.



Técnicas Data Mining

- ▶ **Modelos estadísticos:** Es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales y en la regresión para indicar los diferentes factores que modifican la variable de respuesta.
- ▶ **Reglas de asociación.** Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.

Técnicas Data Mining

- ▶ **Agrupamiento o Clustering:** Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes.

