

Introducción a Data Mining

Trabajo Práctico N° 1 - Introducción

1. Responda a la pregunta **¿Qué es *Data Mining*?** Empleando no más de 500 palabras. En su respuesta tome en cuenta lo siguiente:
 - o ¿Es una moda?
 - o ¿Es una simple transformación de tecnología desarrollada a partir de bases de datos, estadística y aprendizaje automático?
 - o Describir los pasos de *data mining* visto como un proceso de descubrimiento de conocimiento.

2. Fundamente si cada una de las siguientes actividades es o no una tarea de *data mining*:
 - o Dividir los clientes de una compañía de acuerdo con su género.
 - o Dividir los clientes de una compañía de acuerdo con su rentabilidad.
 - o Calcular el total de ventas de una compañía
 - o Clasificar una base de datos de estudiantes basado en los números de identificación.
 - o Predecir los resultados de arrojar un par de dados.
 - o Predecir el precio futuro de las acciones de una compañía usando registros históricos.
 - o Monitorear los latidos del corazón de un paciente para buscar anomalías.
 - o Monitorear ondas sísmicas para detectar actividades de terremotos.
 - o Extraer las frecuencias de una onda de sonido.

3. Presente un ejemplo donde *data mining* es crucial para el éxito de un negocio, o de una organización. Detalle ¿Qué tareas de *data mining* son fundamentales en ese caso?

4. Suponga que Usted tiene un empleo como consultor en data mining para una compañía de un buscador de Internet. Describa cómo puede *data mining* ayudar a la compañía por medio de ejemplos específicos de cómo se pueden aplicar técnicas como clustering, clasificación, mining de reglas de asociación y detección de anomalías.

5. Para cada uno de los siguientes conjuntos de datos, explicar si la privacidad de datos es un aspecto importante o no:



- a. Datos censales del período 1980-1998.
 - b. Direcciones de IP y tiempos de visita de usuarios Web que visitan su sitio web.
 - c. Imágenes satelitales de una región.
 - d. Nombres y direcciones de e-mail recolectadas de la web.
6. ¿Cuáles son las diferencias y similitudes primarias de SEMMA (© SAS) y CRISP-DM?