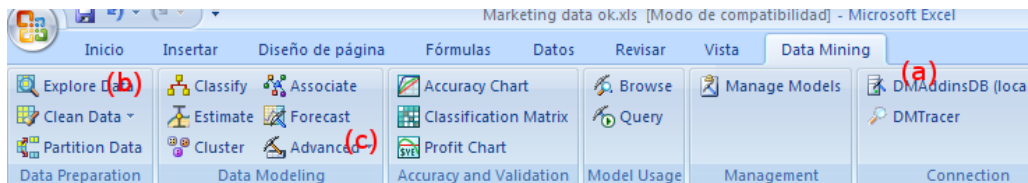


# Sistemas de soporte a las decisiones

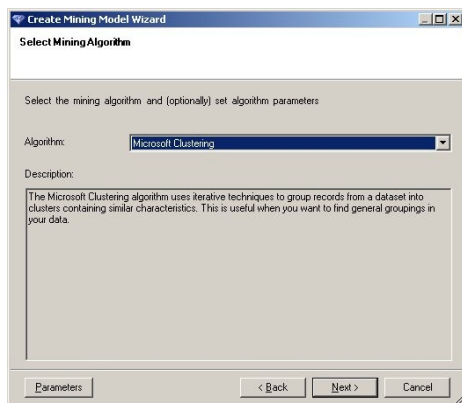
## Segundo Parcial

Instrucciones para general arboles de decisión y clusters, utilizando MS Excel 2007, SQL Server 2005 y el add in para data mining.

Asumimos que el add in para Excel esta bien configurado, el SQL Server tiene una base de datos con los permisos adecuados y hay una conexión correcta. Caso contrario ver la ayuda del add in en la sección Quick Start para configurarlo correctamente, y ejecutar el programa "Server configuration utility" que trae el add in, con permisos de administrador, que resolverá teóricamente el problema.



1. Cargar el archivo Excel con los datos. Seleccionar las herramientas Data mining.
2. Para cada variable (columna) debemos asegurarnos que los datos son correctos, no contiene basura. Para eso, analizamos cada columna.
  - a. Seleccionar una columna completa.
  - b. Ir a pestaña Data Preparation, icono Explore Data (imagen, icono B), y ver si hay datos incorrectos. Eliminar las celdas basura, o directamente descartar toda la columna si no sirven los datos para nuestro propósito (la columna la descartamos en el paso 3, poniendo Do Not Use en el wizard).
  - c. Repetir para todas las columnas.
3. Seleccionar todos los datos. Chequear que la conexión a SQL Server sea correcta (imagen, icono A)
4. Ir a la pestaña Data Modeling, Advanced (imagen, icono C), Create Mining Model.
5. Luego de la pagina de bienvenida, y de aceptar el rango de datos, podemos elegir el algoritmo a utilizar y sus parámetros. Esto nos permite crear un cluster o un árbol de decisión. Aquí las instrucciones se dividen en dos.



## Para crear un árbol de decisión.

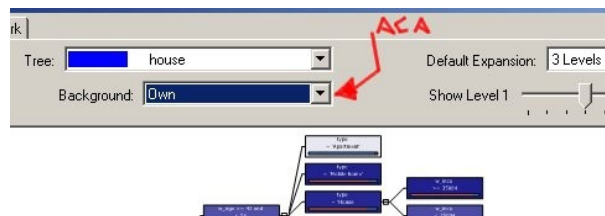
En algoritmo, seleccionar Microsoft Decision Trees.

Luego, en la siguiente pantalla, marcar solo el ID como Key, el resto como Input y prestar atención a las variables que son continuas que no estén marcadas como discretas.

Acordarse de descartar las columnas que analizamos como basura en el paso 2 poniendo "Do Not Use".

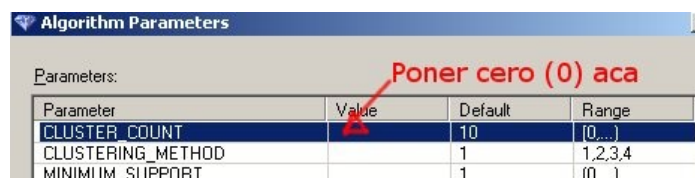
**Importante: debemos marcar la variable que queremos predecir como "Predict Only"**

**Importante: cuando se tiene abierto el árbol, RECORDAR seleccionar el caso que nos interesa chequear. (Background : ahí elegir lo que queremos ver, y buscar cual lo predice mas, es decir investigar en los cuadritos cual tiene el mayor porcentaje de lo que nos interesa, y en el recuadro inferior derecho veremos los parámetros que predicen esa población)**



## Para crear un cluster

En algoritmo, seleccionar Microsoft Clustering. **Atención, tocar el botón inferior izquierdo "Parameters", y colocar 0 en CLUSTER\_COUNT, Value = 0.**



Luego, en la siguiente pantalla, marcar solo el ID como Key, el resto como Input y prestar atención a las variables que son continuas que no estén marcadas como discretas.

Acordarse de descartar las columnas que analizamos como basura en el paso 2 poniendo "Do Not Use".

# **Minería de datos**

Bajo el nombre de minería de datos se engloban un conjunto de técnicas encaminadas a la extracción de conocimiento procesable, implícito en las bases de datos.

Las bases de la minería de datos se encuentran en la inteligencia artificial y en el análisis estadístico y mediante los modelos extraídos utilizando técnicas de minería de datos se aborda la solución a problemas de predicción, clasificación y segmentación.

## ***Árbol 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.

Un árbol de decisión tiene unas entradas las cuales pueden ser un objeto o una situación descrita por medio de un conjunto de atributos y a partir de esto devuelve una respuesta la cual en ultimas es una decisión que es tomada a partir de las entradas.

## ***Cluster***

Un algoritmo de agrupamiento (en inglés, 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.

Un algoritmo de clustering permite extraer representantes de un conjunto de datos, que pueden ser posteriormente usados para transmisión, para eliminación de ruido o con una fase posterior de calibración, para clasificación de vectores en diferentes conjuntos.