

Unidad Temática 2 - Práctico Domiciliario 1

Ejercicio 1

1. "Handling Missing Values"

En la primer parte se utilizó el operador **"Select Attributes"** para quitar las columnas "cabin" y "lifeboat" del dataset de Titanic. El resultado de este proceso es un dataset con 1309 registros y 10 columnas.

Inicialmente el dataset de Titanic contiene 1309 registros, de los cuales 263 tienen valores faltantes en la columna "age". Para el tratamiento de estos valores faltantes se utilizó el operador **"Replace Missing Values"** con el método "Average" para reemplazar los valores faltantes por el promedio de los valores de la columna. El resultado de este proceso es un dataset con 1309 registros y sin valores faltantes en la columna "age".

No obstante aún quedan valores faltantes en las columnas "fare" y "embarked". Para el tratamiento de estos valores faltantes se utilizó el operador **"Filter Examples"** para filtrar los registros que contengan valores faltantes en las columnas "fare" y "embarked". El resultado de este proceso es un dataset con 1306 registros y sin valores faltantes en las columnas "age", "fare" y "embarked".

2. "Normalization and Outlier Detection"

En la segunda parte se utilizó el operador **"Select Attributes"** para quitar las columnas "name", "ticket number", "cabin" y "lifeboat" del dataset de Titanic. El resultado de este proceso es un dataset con 1309 registros y 8 columnas.

Lo siguiente fue normalizar los valores utilizando el operador **"Normalize"**, señalando todas las columnas y utilizando el método "Z-transformation". El resultado de este proceso es un dataset con 1309 registros y 8 columnas con valores normalizados.

Finalmente se utilizó el operador **"Detect Outliers (Distances)"** y con la ayuda de **"Filter Examples"** se filtro el atributo "outlier" en 'false' para detectar los outliers en el dataset. El resultado de este proceso es un dataset con 1299 registros y 8 columnas sin outliers.

Con la ayuda de un breakpoint en el operador **"Detect Outliers (Distances)"** se puede observar que los outliers se encuentran en los registros 50, 51, 114 al 117, 184, 253, 303 y 1172.