

Proceso de Modelado de Datos

1. Definición del objetivo:

Se selecciona la variable *target* del modelo, que en este caso será “**Survived**”, indicando si la persona sobrevivió o no.

2. Depuración de variables irrelevantes:

Se eliminan las columnas “**name**”, “**cabin**” y “**ticket**”, ya que no aportan información significativa para el análisis o podrían introducir ruido.

3. Transformación de variables categóricas:

Se modifica el campo “**sex**” para convertirlo en una variable numérica:

- Femenino → 1
- Masculino → 0

4. División del conjunto de datos:

Los datos se separan en dos subconjuntos:

- **80%** destinado al entrenamiento del modelo (*Training set*).
- **20%** reservado para su validación (*Test set*).

5. Evaluación inicial con Matriz de Confusión:

Los resultados obtenidos fueron:

- Verdaderos Positivos (VP): 41
- Verdaderos Negativos (VN): 94
- Falsos Positivos (FP): 22
- Falsos Negativos (FN): 21

6. Medición del desempeño general:

A través de la herramienta *Test and Score*, se obtuvo una **exactitud (Accuracy)** de **0.803**, lo que indica que el modelo alcanza un nivel de **confiabilidad satisfactorio** para su propósito.