

1. Carga y visualización de datos

Usamos el widget **File** para importar el dataset al entorno de Orange. Luego, mediante **Data Table**, se visualizaron las variables disponibles y se identificó la variable target del análisis: *Supervivencia*. Observamos que contábamos con 892 instancias y el rol de las distintas variables.

2. Exploración de las variables

Con el widget **Distributions** se analizaron las distribuciones para evaluar su relación con la supervivencia. En particular, la variable **Embarked** resultó relevante, ya que mostró diferencias claras en la proporción de sobrevivientes según el puerto de embarque: en S, aproximadamente el 66% no sobrevivió y el 33% sí lo hizo.

3. Preparación de los datos

La variable **Cabin** fue eliminada debido a la gran cantidad de valores faltantes, lo que impedía utilizarla como atributo confiable en el entrenamiento. Esta decisión permitió trabajar con un dataset más limpio y enfocado en las variables realmente relevantes para la predicción. A través de **Edit Domain**, vimos que se definió la variable target (*Survived*) con un 1 y 0 para su utilización en el entrenamiento.

4. División de los datos

Con el widget **Data Sampler**, se separó el dataset en dos subconjuntos: un **70% destinado al entrenamiento** y un **30% reservado para la evaluación**. Esta división permitió entrenar el modelo en una muestra representativa y luego validar su desempeño en datos no utilizados en la fase de entrenamiento.

5. Entrenamiento y evaluación del modelo

Se empleó el widget **Training** para generar el modelo de **árbol de decisión (Tree)**. Posteriormente, el rendimiento del clasificador se evaluó mediante **Predictions**, **Confusion Matrix** y **Test & Score**, lo que permitió obtener métricas de desempeño y analizar los aciertos y errores. Finalmente, a través de **Tree Viewer**, se interpretó la estructura del árbol, identificando las variables más relevantes en la predicción de la supervivencia.

6. Resultados

El modelo alcanzó una **puntuación de 0,745** indicando un rendimiento **bueno**, mostrando que el modelo logra distinguir de manera adecuada entre pasajeros que sobrevivieron y los que no.

La **Confusion Matrix** reflejó:

- De un total de 267 casos, el modelo clasificó correctamente 198 (141 verdaderos negativos y 57 verdaderos positivos).
- Se produjeron 69 errores (31 falsos positivos y 38 falsos negativos).

Esto significa que el modelo es más **preciso al identificar a los pasajeros que no sobrevivieron** (82% de acierto en esa clase), aunque presenta un desempeño algo menor en la identificación de los sobrevivientes.

La visualización en **Tree Viewer** evidenció que la primera variable de decisión fue el **sexo del pasajero**. En niveles siguientes aparecieron como relevantes la **clase del pasajero (Pclass)** y la **tarifa pagada (Fare)**, ambas coherentes con la evidencia de que los pasajeros de primera clase y con mayores recursos tenían más posibilidades de sobrevivir.

		Predicted		Σ
		0	1	
Actual	0	140	32	172
	1	43	52	95
Σ		183	84	267

