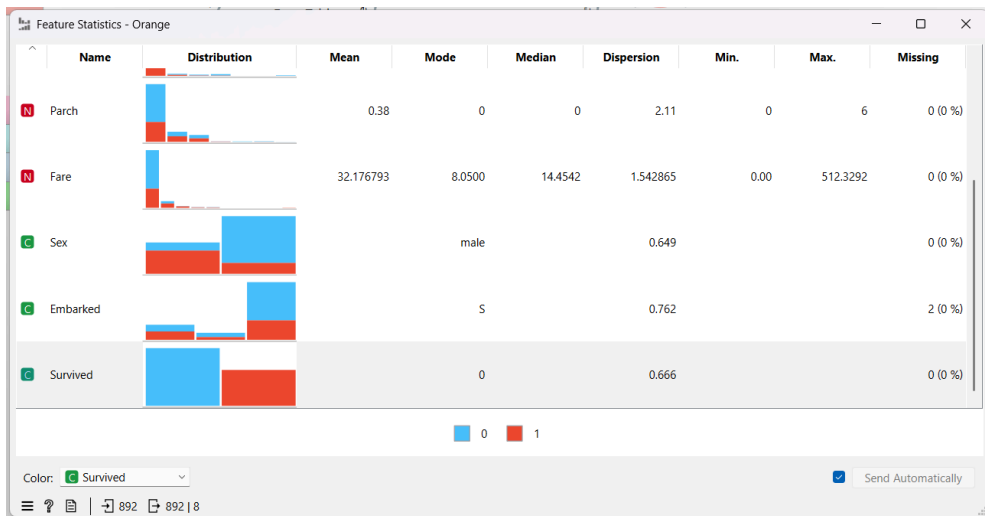


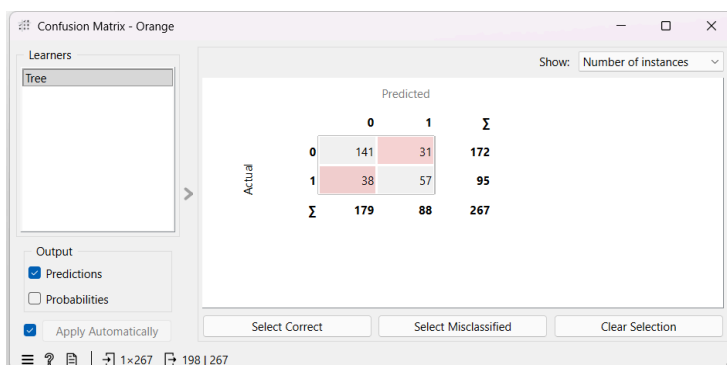
Trabajo 25/9



Primero con este Feature Statistics pudimos tener un primer vistazo de las variables del dataset. Nos dejó ver la media, mediana, mínimos y máximos de cada atributo y cómo están distribuidos. Ahí notamos, por ejemplo, que la mayoría de los pasajeros pagó tarifas bajas, que la mayoría eran hombres y que casi dos tercios no sobrevivieron. Esto nos ayudó a entender los datos de manera general antes de avanzar con el resto del análisis.

Después usamos el Data Sampler para dividir el dataset en entrenamiento (70%) y prueba (30%). El Training se usó para que el modelo aprenda y el Test para ver cómo respondía con datos nuevos

Después construimos el árbol de decisión que quedó bastante grande. Esto nos sirvió para visualizar cómo las variables, como el sexo, la tarifa y los acompañantes, se combinan para predecir la supervivencia. El árbol mostró que el sexo fue una de las variables más importantes, y también aparecieron divisiones por tarifa y otros factores. Con esto pudimos ver de manera clara qué condiciones aumentaban o reducían las chances de sobrevivir.



Después pasamos a las predicciones y analizamos los resultados en la matriz de confusión. Ahí vimos que el modelo acertó en la mayoría de los casos: 141 personas fueron clasificadas correctamente como no sobrevivientes y 57 como sobrevivientes. Sin embargo,

también hubo errores, ya que 31 sobrevivientes reales fueron predichos como no sobrevivientes y 38 no sobrevivientes como sobrevivientes. Esto nos mostró que el árbol tiene un desempeño aceptable, pero no perfecto, sobre todo a la hora de diferenciar a los que sí sobrevivieron.