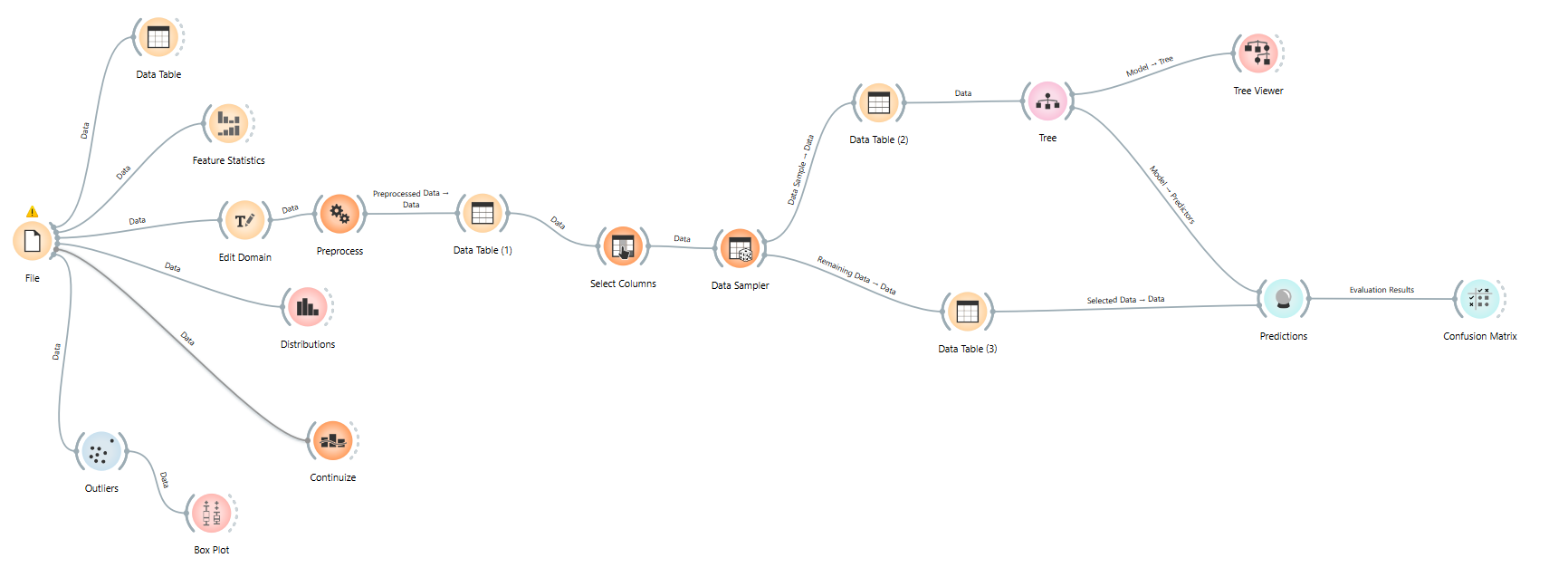
***Titanic, Orange***

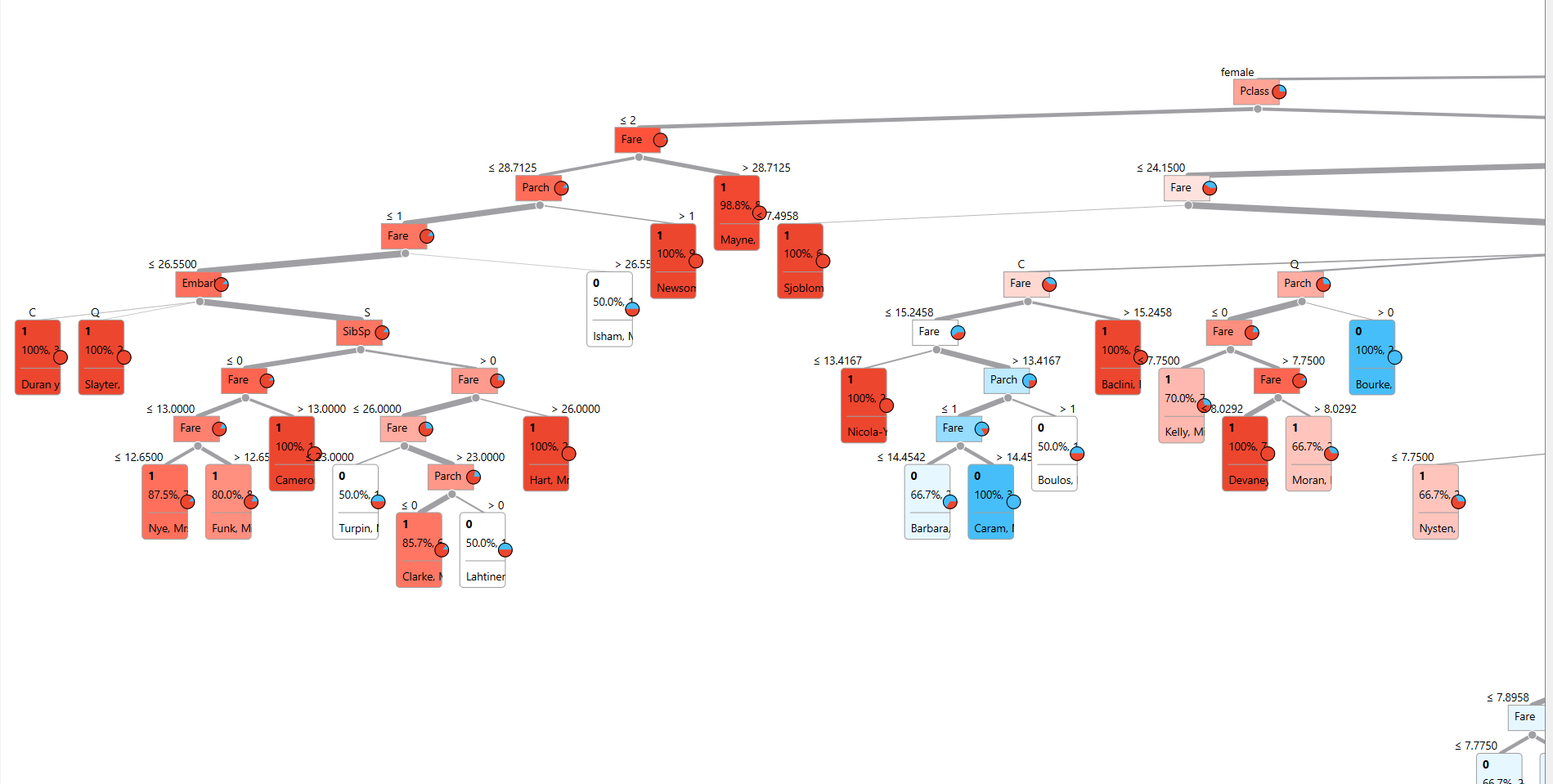


Para analizar el archivo Titanic empleamos Orange y armamos el siguiente flujo. Primero cargamos la tabla con File y exploramos el contenido con Feature Statistics, Distributions y Box Plot para identificar distribuciones y posibles outliers. Con Edit Domain revisamos tipos de variables y, con Preprocess, imputamos valores faltantes (por ejemplo Age) y transformamos variables categóricas (Sex, Embarked) a formato utilizable por los modelos; además normalizamos las variables numéricas cuando fue necesario.

A continuación usamos Select Columns para eliminar variables irrelevantes o problemáticas para la predicción (por ejemplo PassengerId, Name, Ticket, Cabin) y dejamos las características que creemos predictoras. Definimos Survived como la variable objetivo (target).

Dividimos la muestra con Data Sampler en un 80% para entrenamiento y 20% para prueba. Con los datos de entrenamiento ajustamos un modelo de Tree (árbol de decisión) y visualizamos su estructura con Tree Viewer para entender qué variables y cortes toma el modelo. Después aplicamos Predictions sobre el conjunto de prueba y evaluamos el desempeño con Confusion Matrix, obteniendo métricas clave como accuracy, precision y recall que nos permiten juzgar la capacidad del modelo para clasificar sobrevivientes/no sobrevivientes.

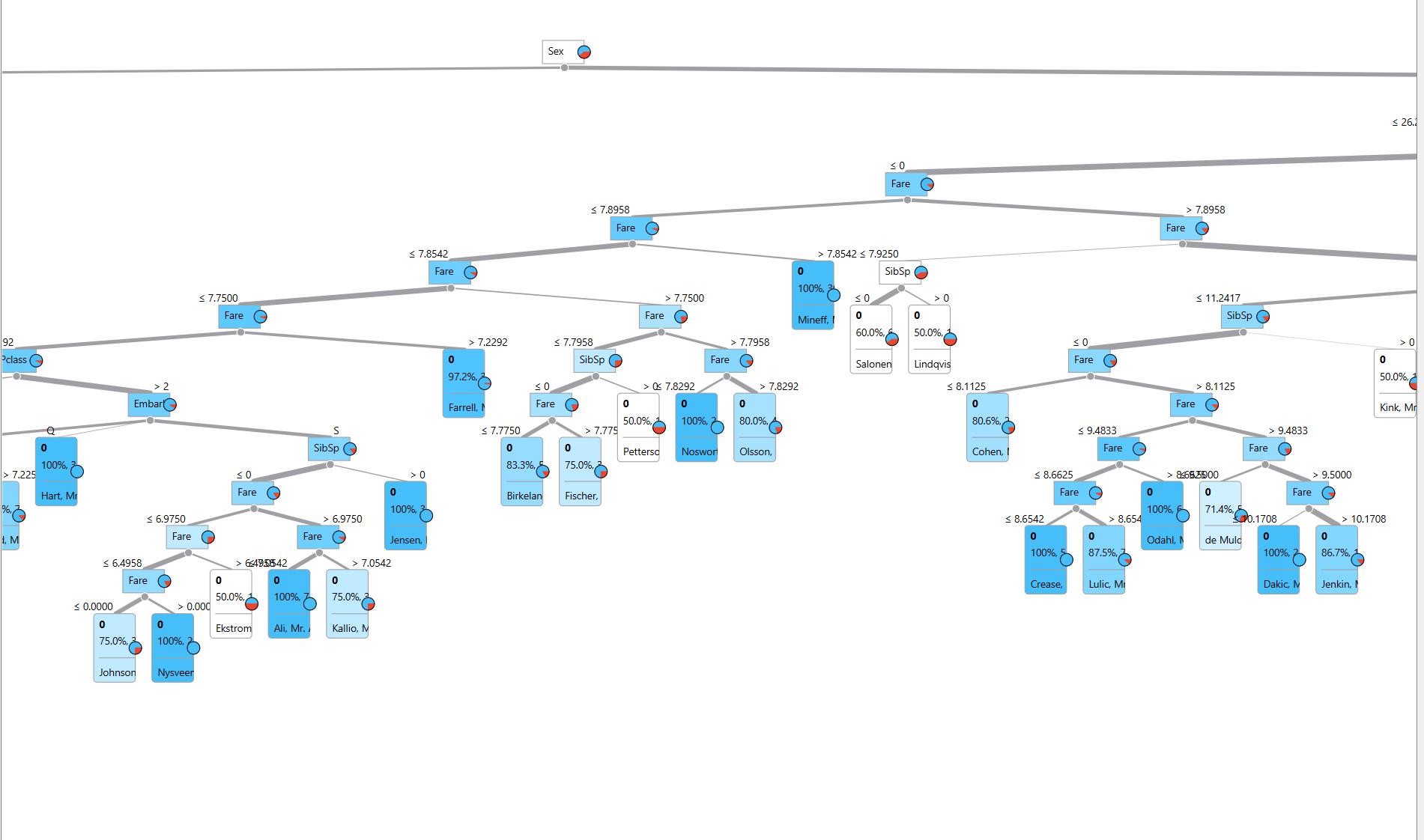
Finalmente interpretamos los resultados: la matriz de confusión muestra la cantidad de verdaderos positivos/negativos y falsos positivos/negativos; si detectamos un sesgo en clases (por ejemplo más no sobrevivientes) podemos usar muestreo estratificado o validación cruzada, probar otros modelos (Random Forest, Logistic Regression) o afinar el preprocesado (imputación distinta, nuevas variables derivadas) para mejorar el rendimiento.



A partir del árbol de decisión obtenido, se pueden identificar los principales factores que influyen en la probabilidad de supervivencia de los pasajeros del Titanic. La primera variable que utiliza el modelo para dividir los datos es el **género (Sex)**, lo que indica que esta es la característica más determinante. En el árbol se observa que la mayoría de las mujeres (color rojo, que representa “sobrevivió”) tienen una probabilidad mucho mayor de sobrevivir, mientras que los hombres (color azul, “no sobrevivió”) presentan tasas de supervivencia más bajas.

Entre las mujeres, el **precio del pasaje (Fare)** y la **clase del boleto (Pclass)** son variables importantes que explican las diferencias de supervivencia: quienes pagaron tarifas más altas o viajaban en primera clase tienen una mayor probabilidad de haber sobrevivido. Esto sugiere una fuerte relación entre el nivel socioeconómico y las posibilidades de rescate durante el naufragio.

Por otro lado, entre los hombres, el modelo muestra que variables como **Fare**, **SibSp** (número de hermanos o pareja a bordo) y **Parch** (número de padres o hijos a bordo) también influyen en la predicción, aunque con menor peso. Los hombres que viajaban solos y con boletos de bajo costo presentan las tasas más altas de mortalidad.

En este árbol de decisión puede verse que la primera variable que el modelo utiliza para dividir los datos es nuevamente el **sexo (Sex)**, confirmando que esta característica es el principal factor para predecir la supervivencia. En este caso, el árbol muestra principalmente el comportamiento de los **hombres (Sex = male)**, ya que la mayoría de los nodos son de color azul, lo que representa pasajeros que *no sobrevivieron*.

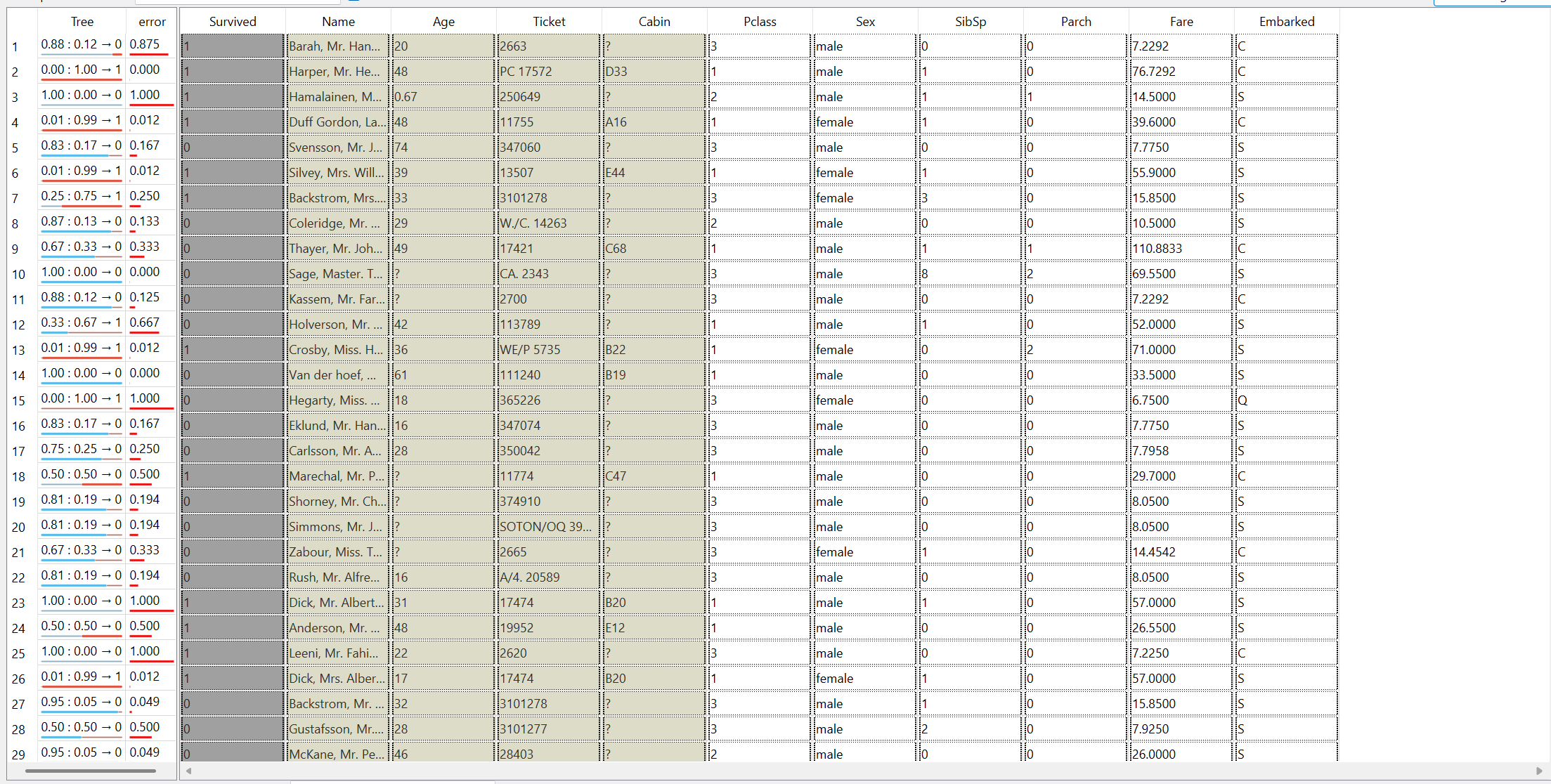
Dentro del grupo masculino, las siguientes variables que más influyen son el **precio del pasaje (Fare)**, la **clase (Pclass)** y la cantidad de familiares a bordo (**SibSp**). Los hombres con tarifas más bajas (por debajo de aproximadamente 8 libras) y que viajaban solos fueron los que presentaron las mayores tasas de mortalidad. Esto coincide con el hecho de que los pasajeros de tercera clase y con menos recursos tenían menos posibilidades de acceso a los botes salvavidas.

A medida que aumenta el valor del pasaje, el modelo detecta algunos casos de hombres que sí sobrevivieron, aunque son excepcionales y no forman patrones amplios. Las ramas con **Fare** más alto o con acompañantes (**SibSp > 0**) muestran pequeñas proporciones de supervivencia (nodos rojos), pero siguen siendo minoría.

En resumen, este gráfico complementa el anterior mostrando el lado opuesto:

* **Ser hombre**,
* **viajar solo**,
* **y tener un pasaje de bajo costo**  
  son los factores más fuertemente asociados con la *no supervivencia*.

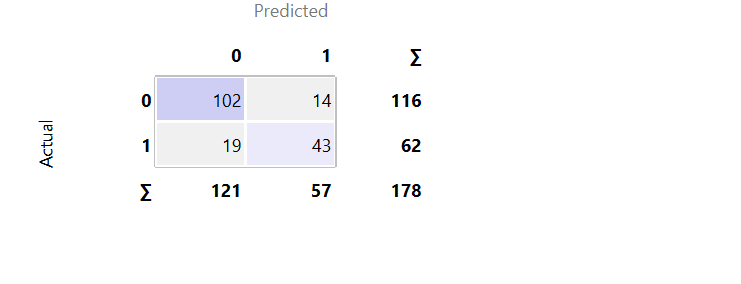
Esto refuerza la conclusión general del modelo: las mujeres y los pasajeros de clases más altas tuvieron más oportunidades de sobrevivir, mientras que los hombres de tercera clase fueron los más afectados.



En general, puede verse que el modelo acierta correctamente en una gran parte de los casos (errores bajos o iguales a 0). Sin embargo, hay algunos ejemplos donde el modelo se equivoca, especialmente con pasajeros masculinos de tercera clase y tarifas intermedias, lo que sugiere que estos perfiles son más difíciles de clasificar correctamente.

Al analizar los resultados, se nota que los pasajeros con mayor probabilidad de supervivencia (predicción = 1) suelen ser **mujeres**, **niños** o personas que **pagaron tarifas más altas** (clases 1 y 2), mientras que las predicciones negativas (predicción = 0) se concentran en **hombres adultos** de **clase 3** con tarifas bajas.

En conclusión, el modelo logra capturar adecuadamente los patrones principales del naufragio —el sesgo por género y clase social—, aunque su precisión podría mejorarse si se probaran otros algoritmos más complejos (como Random Forest o Logistic Regression) o si se ajustaran los parámetros del árbol de decisión para reducir el sobreajuste y los errores en casos ambiguos.

El modelo tiene un **buen rendimiento general (≈81%)**, aunque **predice mejor quiénes no sobrevivieron** que quiénes sí.  
Esto se ve en la alta cantidad de verdaderos negativos (102) frente a los verdaderos positivos (43).  
El árbol de decisión logró captar el patrón principal del desastre (la mayoría no sobrevivió), pero aún **confunde algunos casos de supervivientes**, posiblemente por variables que se solapan (como hombres jóvenes o pasajeros de segunda clase con tarifas medias).