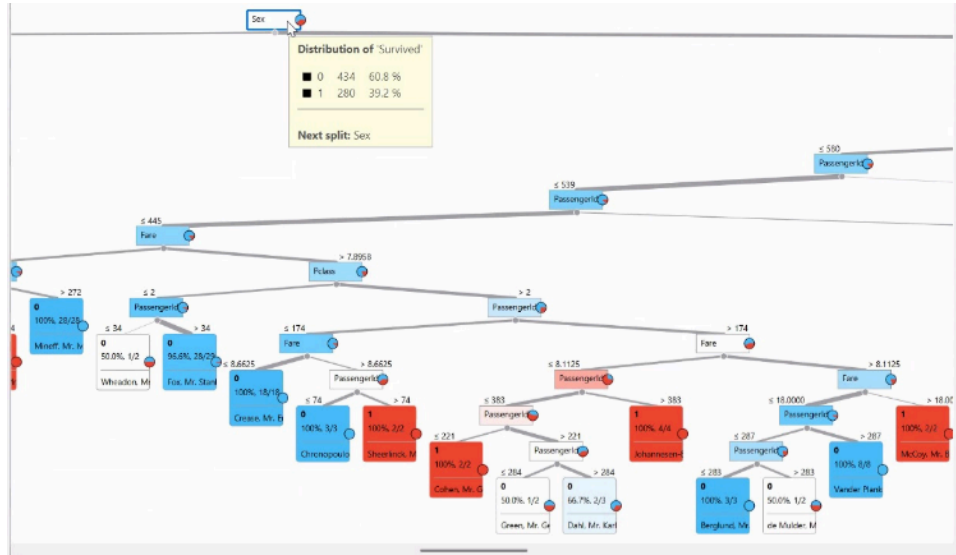


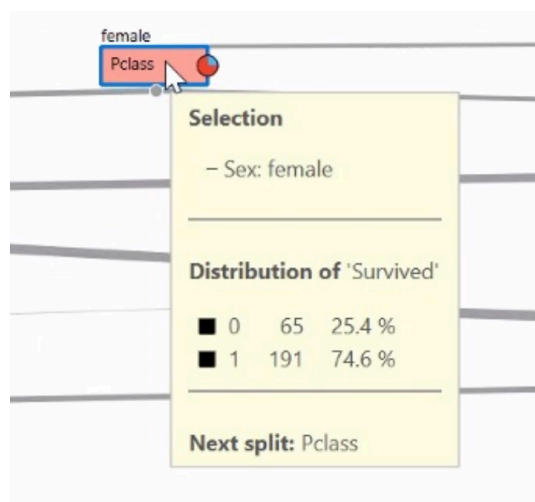
Análisis dataset en ORANGE

Como primera medida de división entre los datos que nos brinda el dataset encontramos el "Sexo", ya sea femenino (color: rojo) o masculino (color: azul). Como lo vemos en orange:



Además, entre las variables analizadas, también se observa la edad como un posible factor de influencia, ya que dentro del grupo masculino y femenino existen diferencias claras en la tasa de supervivencia según el rango etario. Por ejemplo, los pasajeros más jóvenes presentan una mayor probabilidad de sobrevivir, lo que refuerza la idea de que el modelo toma en cuenta múltiples atributos además del género.

La segunda división que encontramos es la "Supervivencia", la cual en el género femenino la podemos ver de esta manera:



En este punto, el árbol de decisión nos permite entender visualmente cómo se dividen los datos y qué características resultan más relevantes para la predicción.

Lo mismo lo podemos ver si estamos analizando el género masculino.

Si profundizamos en los datos que nos brinda la tabla de predicciones, podemos ver que en la primera línea el modelo predijo con una probabilidad del 95% que el pasajero no sobreviviría, pero en realidad sí lo hizo.

Y analizando con exactitud el modelo, el índice CA es de un 75%. Eso significa que el modelo acertó el 75% de las veces.

Predictions - Orange

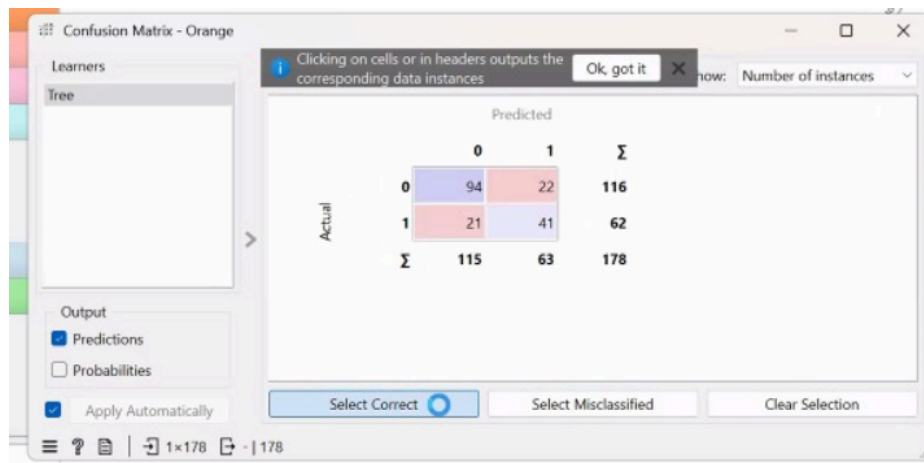
Show probabilities for: Classes in data ☒ Show classification errors Restore Original Order

	Tree	error	Survived	Name	Age	Ticket	Cabin	Passengerid	Pclass	Sex	SibSp	Parch	Fare
1	0.95 : 0.05 → 0	0.955	1	Barah, Mr. Han...	20	2663	P	763	3	male	0	0	7.2292
2	1.00 : 0.00 → 0	1.000	1	Harper, Mr. He...	40	PC 17572	D33	846	1	male	1	0	76.7292
3	0.00 : 1.00 → 1	0.000	1	Hamalainen, M...	18.67	250649	F	756	2	male	1	1	14.5000
4	0.01 : 0.99 → 1	0.012	1	Duff Gordon, La...	48	111755	A16	557	1	female	1	0	39.6000
5	1.00 : 0.00 → 0	0.000	0	Svensson, Mr. J...	74	347060	P	852	3	male	0	0	7.7750
6	0.01 : 0.99 → 1	0.012	1	Silvey, Mrs. Will...	29	13507	E44	578	1	female	1	0	55.9000
7	1.00 : 0.00 → 0	1.000	1	Backstrom, Mrs...	33	3101278	F	86	3	female	3	0	15.8500
8	0.97 : 0.03 → 0	0.034	0	Coleridge, Mr. ...	29	W.C. 14263	P	343	2	male	0	0	10.5000
9	0.67 : 0.33 → 0	0.333	0	Thayer, Mr. Joh...	49	17421	C68	699	1	male	1	1	110.8833
10	1.00 : 0.00 → 0	0.000	0	Sage, Master. T...	?	CA. 2343	P	160	3	male	0	?	69.5500
11	1.00 : 0.00 → 0	0.000	0	Kassem, Mr. Far...	?	2700	P	525	3	male	0	0	7.2292
12	1.00 : 0.00 → 0	0.000	0	Holverson, Mr. ...	42	113789	P	136	1	male	1	0	52.0000
13	0.01 : 0.99 → 1	0.012	1	Crosby, Miss. H...	36	Wk/P 5735	B22	541	1	female	0	2	71.0000
14	1.00 : 0.00 → 0	0.000	0	Van der hoef, ...	61	111240	B19	171	1	male	0	0	33.5000
15	0.00 : 1.00 → 1	1.000	0	Hegarty, Miss. ...	18	365226	P	655	3	female	0	0	6.7500
16	0.95 : 0.05 → 0	0.045	0	Eklund, Mr. Han...	16	347074	P	765	3	male	0	0	7.7750
17	0.95 : 0.05 → 0	0.045	0	Carlsson, Mr. A...	28	350042	P	757	3	male	0	0	7.7958
18	0.50 : 0.50 → 0	0.500	1	Marechal, Mr. P...	?	11774	C47	840	1	male	0	0	29.7000
19	1.00 : 0.00 → 0	0.000	0	Shenmy, Mr. Ch...	?	374913	P	96	3	male	0	0	8.0500
20	0.00 : 1.00 → 1	1.000	0	Simmons, Mr. J...	?	SOTON/OQ 39...	P	564	3	male	0	0	8.0500

Show performance scores Target class: (Average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.720	0.758	0.759	0.759	0.758	0.470

También en el modelo podemos ver falsos positivos, donde la predicción decía que iba a sobrevivir y no sobrevivió, en este caso 22 personas tiraron un falso positivo. En esta variable, lo que deberíamos pensar es ¿Qué es peor que te diga que va a morir y no se murió? Que no va a morir y se muere? Entendemos que lo pero es falso positivo, ya que genera una esperanza de sobrevivir y no sobrevive.



También es interesante analizar el equilibrio del modelo, es decir, si clasifica de manera similar a ambos grupos (sobrevivientes y no sobrevivientes). En este caso, se observa que el modelo tiende a acertar más en la predicción de quienes no sobrevivieron, lo cual podría deberse a un desbalance en los datos originales.

El índice de precisión fue de 0,759, lo que muestra que el modelo predice bastante bien, aunque todavía se podría mejorar ajustando algunos valores o probando otros tipos de modelos. Aun así, los resultados son buenos y ayudan a entender qué factores aumentan las chances de sobrevivir.

Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.720	0.758	0.759	0.759	0.758	0.470

178 | 1 | 178 | 1x178

En resumen, con este análisis pudimos ver cómo distintas variables influyen en la supervivencia de los pasajeros. El modelo mostró que el género fue una de las más importantes, ya que las mujeres tuvieron una tasa de supervivencia mucho mayor. También se nota que la edad y la clase en la que viajaban tuvieron su peso en el resultado.