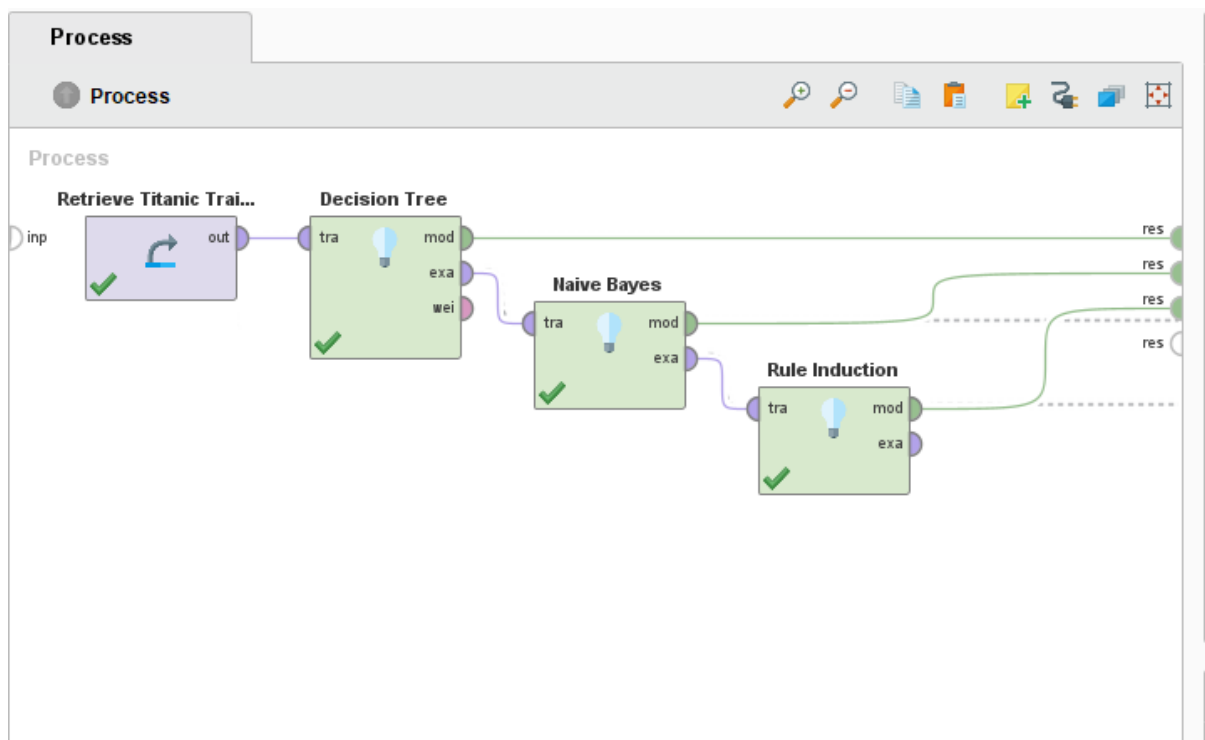


UT2 PD2

Ejercicio 1:

A) Modeling:



En este caso utilizamos el Titanic Training Data, ya que no tiene valores faltantes. Colocamos el operador Decision Tree para predecir en este caso la supervivencia de los pasajeros en el Titanic.

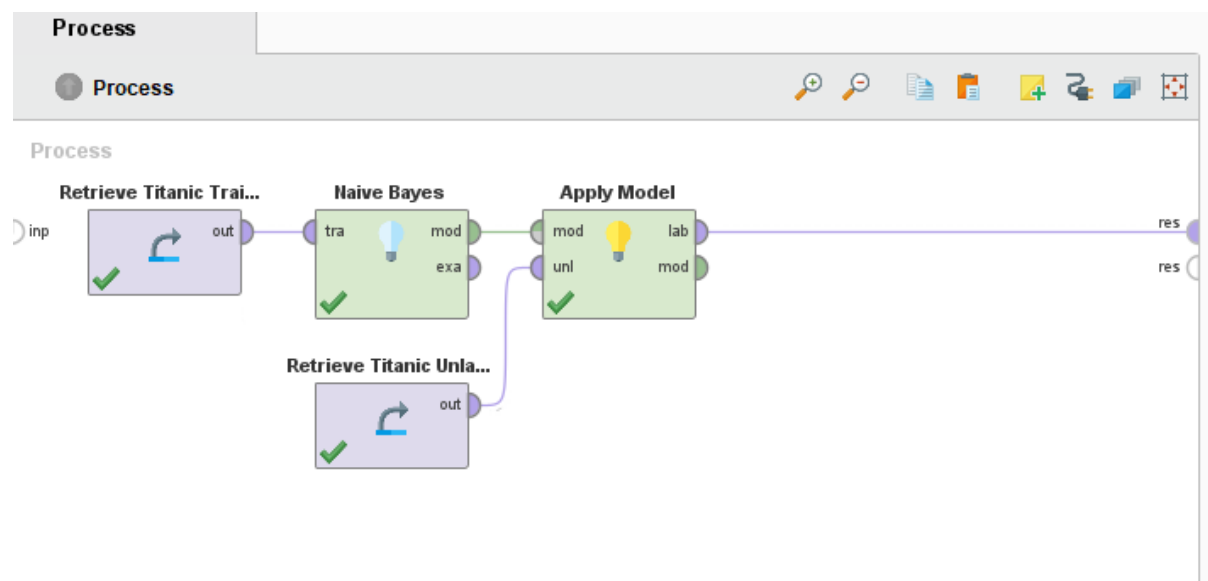
Luego colocamos el operador Naive Bayes que es un algoritmo de aprendizaje supervisado, que en este caso va a darnos la probabilidad de supervivencia de los pasajeros del Titanic.

Una vez que vemos los resultados podemos ver claramente que la cantidad de familiares que estaban en el barco es más importante que la clase en la que

estaban las mujeres. Esto sucede sólo con las mujeres porque los hombres tenían menos chances de sobrevivir ("Mujeres y niños primero")

B) Scoring:

En este caso utilizamos también el DataSet de Titanic Training



Luego de usar el Operador Naive Bayes tenemos que usar el operador Apply Model para crear predicciones sobre la supervivencia de los pasajeros en el Titanic.

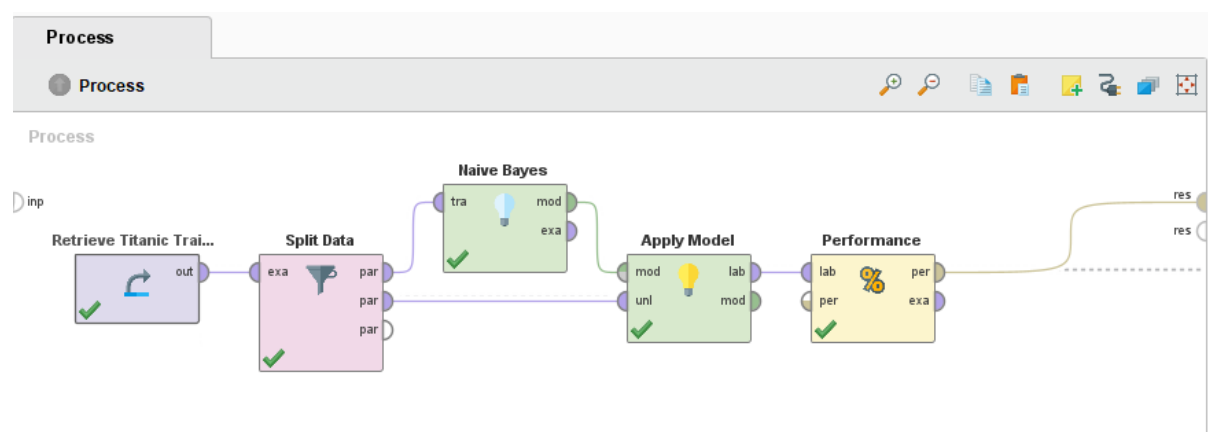
En la tabla que nos devuelve podemos ver la predicción en si la persona sobrevive o no, y en las columnas siguientes podemos ver el porcentaje de la predicción por ejemplo:

Row No.	prediction(Survived)	confidence(Yes)	confidence(No)	Age	Passenger ...	Sex	No of Sibling
1	Yes	0.987	0.013	0.917	First	Male	1
2	Yes	0.714	0.286	53	First	Female	2
3	Yes	0.542	0.458	71	First	Male	0
4	Yes	1.000	0.000	47	First	Male	1

En la primera fila tenemos una persona que se predice que va a sobrevivir y la confiabilidad de ese dato es del 98% y 1,3% de confiabilidad de no.

C) Test Splits and Validation

Este tutorial se utilizara para ver si el modelo utilizado acertó en los casos mostrados. En este caso utilizamos el operador Split Data que toma un conjunto de ejemplos en las particiones que nosotrs decidimos en este caso lo hicimos 30%, 70% cual fernet. En este caso el 70% se convertira en nuestros datos de entrenamiento mientras que el 30 van a ser los datos para comparar las predicciones de nuestro modelo.



Podemos ver la tabla luego con los resultados:

Nos muestran 80,36% de aciertos.

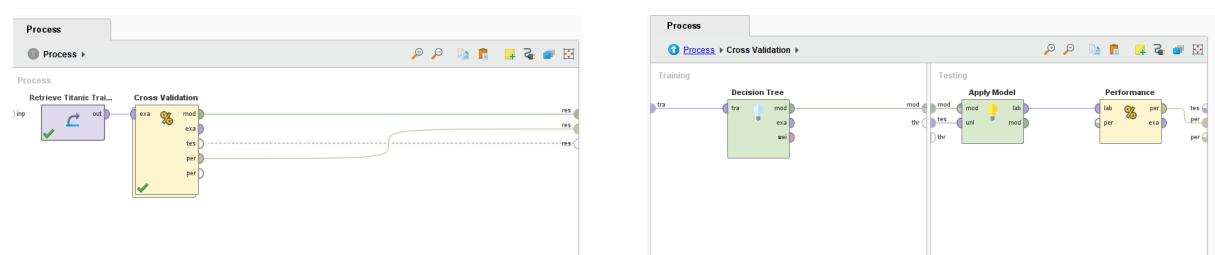
Table View Plot View

accuracy: 80.36%

	true Yes	true No	class precision
pred. Yes	76	25	75.25%
pred. No	29	145	83.33%
class recall	72.38%	85.29%	

D) Cross Validation

En este tutorial podemos calcular la precision de los modelos. Para eso utilizamos el Dataset de Titanic training.



El modelo que nos devolverá:

The screenshot shows the 'PerformanceVector (Performance)' window with the 'Table View' selected. The table displays the following performance metrics:

	true Yes	true No	class precision
pred. Yes	253	84	75.07%
pred. No	96	483	83.42%
class recall	72.49%	85.19%	

En este caso tiene una precision de 80,35% de aciertos

E) Visual Model Comprasion

En este tutorial utilizamos el Titanic Training y luego el Compare ROCs. En este caso comparamos 3 tipos de modelos de predicción que son: Rule

induction, Decision Tree y ruel
induction:

Como resultado tenemos una grafica
de los tres modelos, siendo el peor el
de Naive Bayes. Pero esto es según el
conjunto de datos que tengamos.

