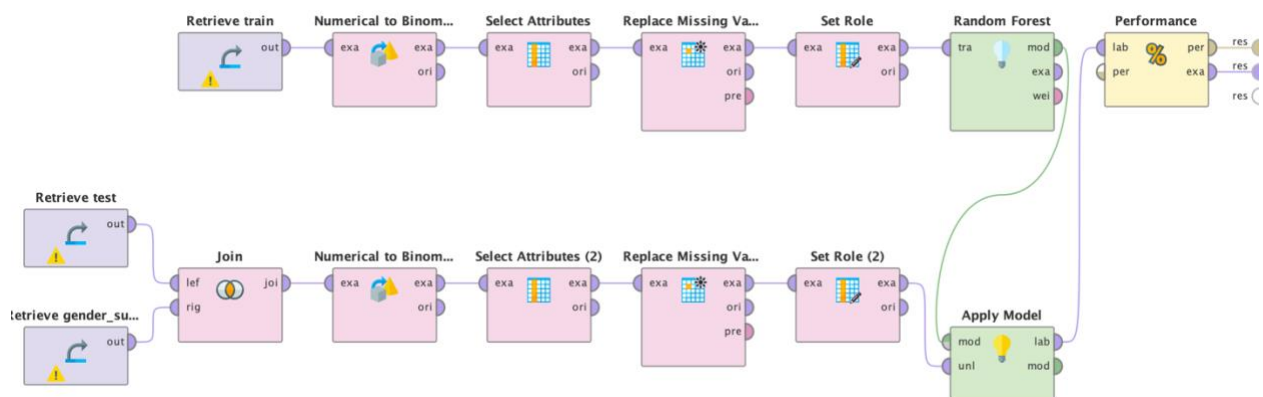


# TITANIC ASSIGNMENT

## Process:

- Import the Train.csv dataset.
- Clean the data by removing unnecessary columns by transforming categorical variables into numerical variables(Numerical to binomial), selecting attributes, handling missing values(Replace missing values), and set label using set role operator.
- Train a machine learning model on the training set using the "Random forest" operator and evaluate its performance using the "Performance(Classification)" operator through apply model operator, Here I have selected Random Forest because it gives a high accuracy comparing any other classification operator.
- Test the model on the train set and generate predictions for the "Survived" column and do the same for test set by importing test.csv and merging it with gender submission data file using "Join" operator.
- Merge the predictions with the "PassengerId" column from the test data and the "PassengerId" column from the gender submission file.



accuracy: 92.58%

	true false	true true	class precision
pred. false	253	18	93.36%
pred. true	13	134	91.16%
class recall	95.11%	88.16%	

Row No.	Survived	prediction(...)	confidence(...)	confidence(...)	Pclass	Sex	Age	SibSp	Parch
1	false	false	0.897	0.103	3	male	34.500	0	0
2	true	false	0.857	0.143	3	female	47	1	0
3	false	false	0.897	0.103	2	male	62	0	0
4	false	false	0.897	0.103	3	male	27	0	0
5	true	true	0.409	0.591	3	female	22	1	1
6	false	false	0.897	0.103	3	male	14	0	0
7	true	true	0.409	0.591	3	female	30	0	0
8	false	false	0.897	0.103	2	male	26	1	1
9	true	true	0.409	0.591	3	female	18	0	0
10	false	false	0.897	0.103	3	male	21	2	0
11	false	false	0.897	0.103	3	male	30.273	0	0
12	false	false	1	0	1	male	46	0	0
13	true	true	0	1	1	female	23	1	0
14	false	false	0.897	0.103	2	male	63	1	0
15	true	true	0	1	1	female	47	1	0
16	true	true	0.090	0.910	2	female	24	1	0
17	false	false	0.897	0.103	2	male	35	0	0
18	false	false	0.897	0.103	3	male	21	0	0
19	true	true	0.409	0.591	3	female	27	1	0