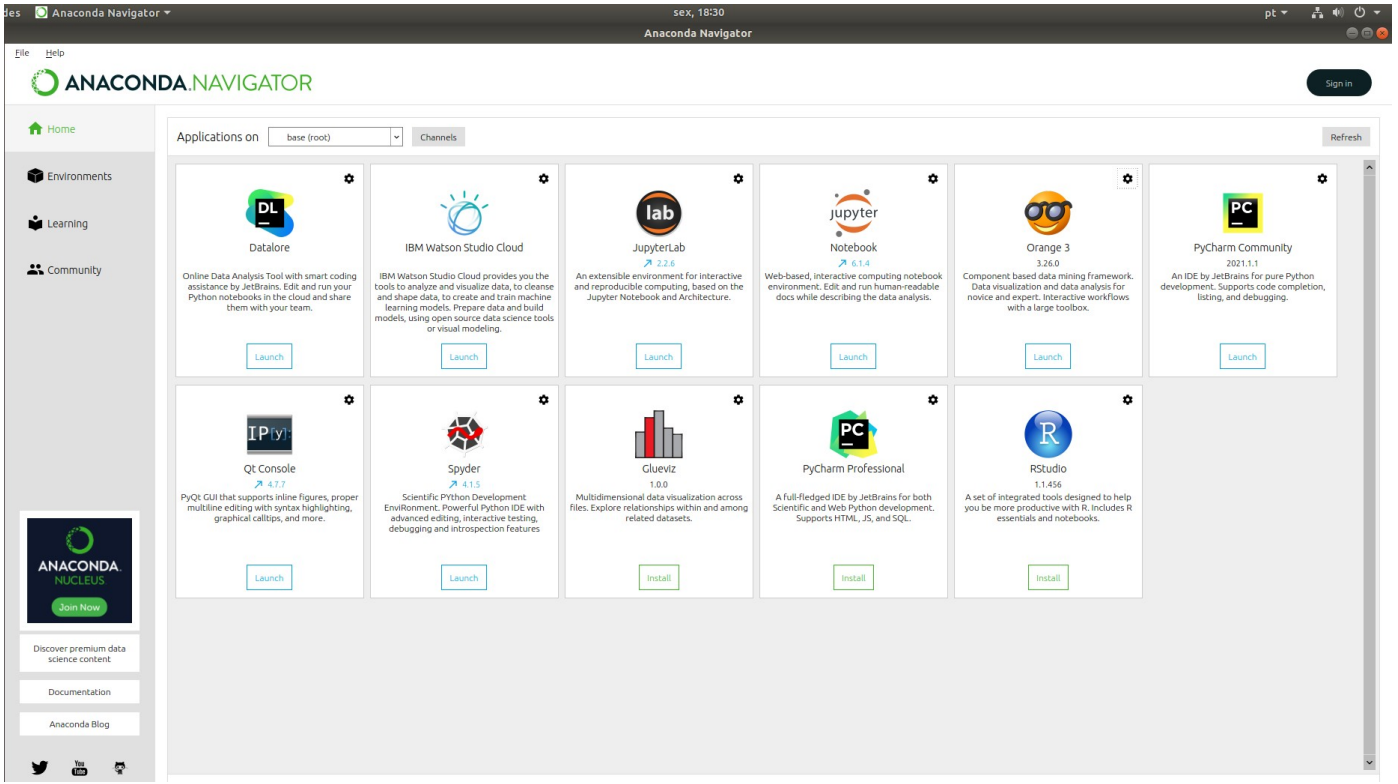
Orange
Data Mining Tool

Nádia Félix

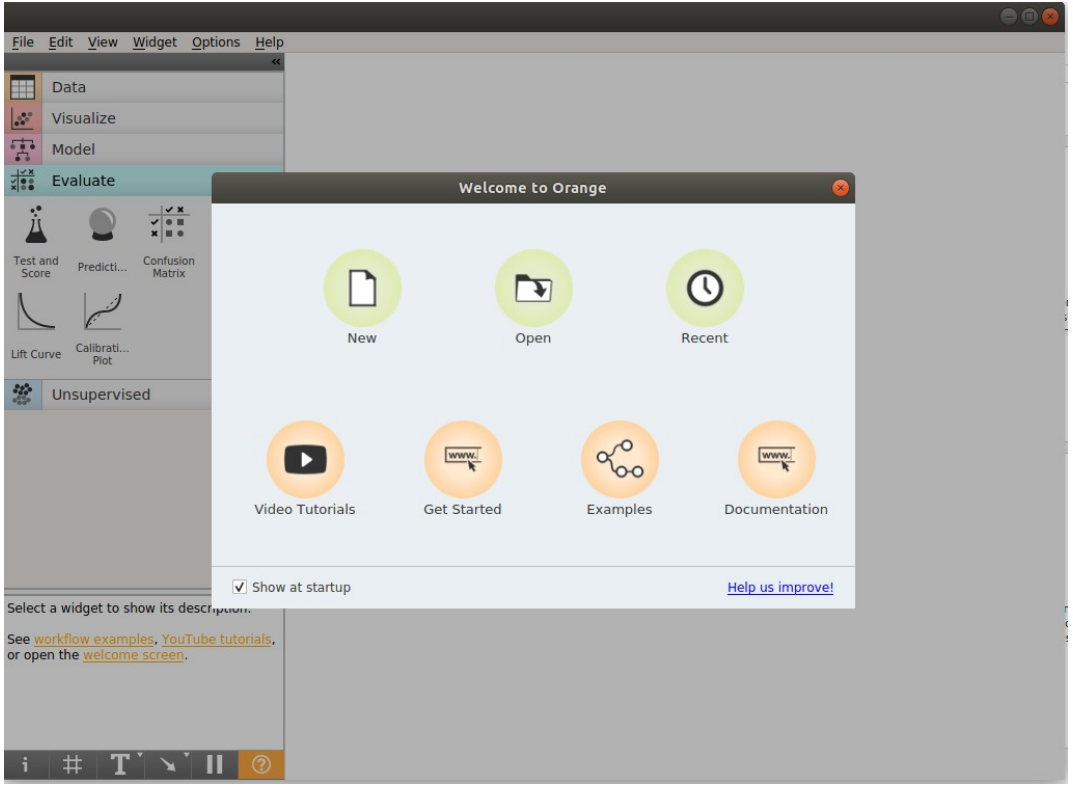
Por que Orange?

- Open Source
- Component based
- No programming
- Data visualization
- Platform independent software
- Allows clustering and classification
- Data mining through visual programming
- and python scripting
- Por ser baixado em: https://orangedatamining.com/

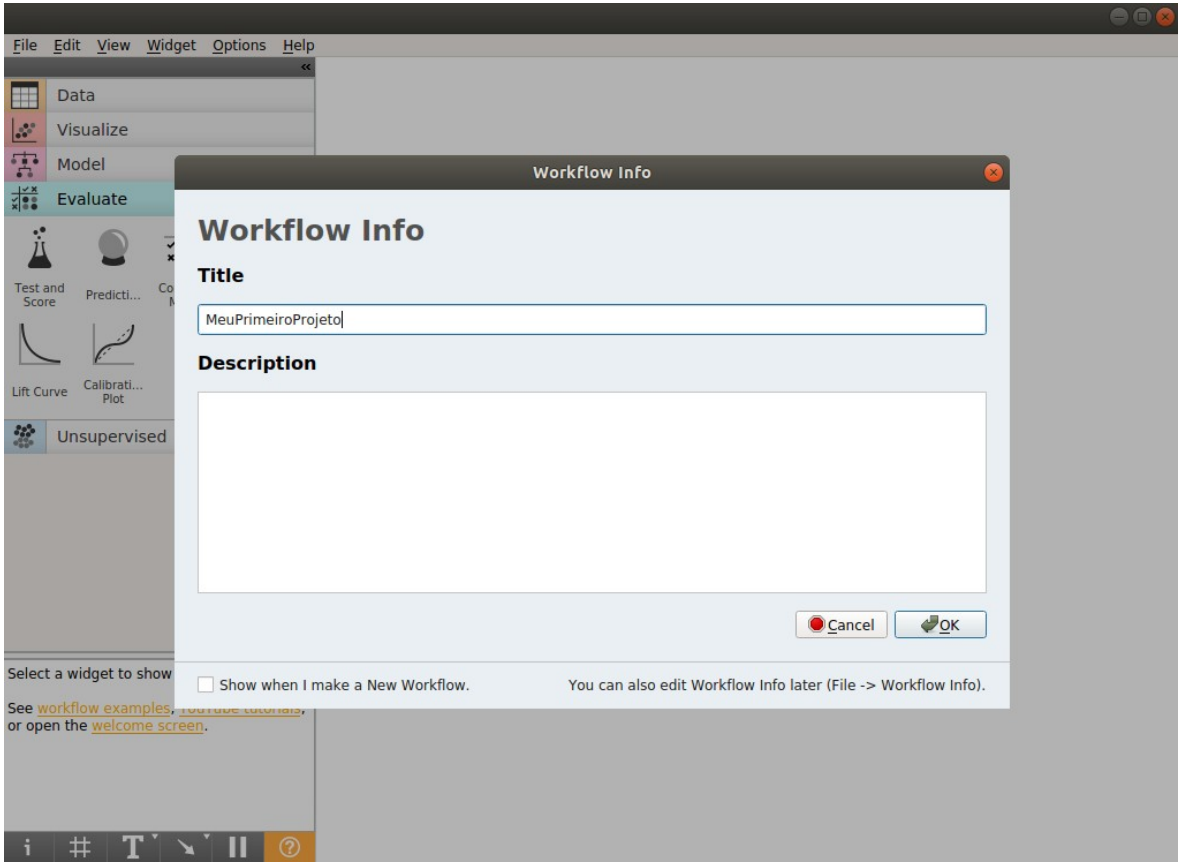
Getting Started With ORANGE!!

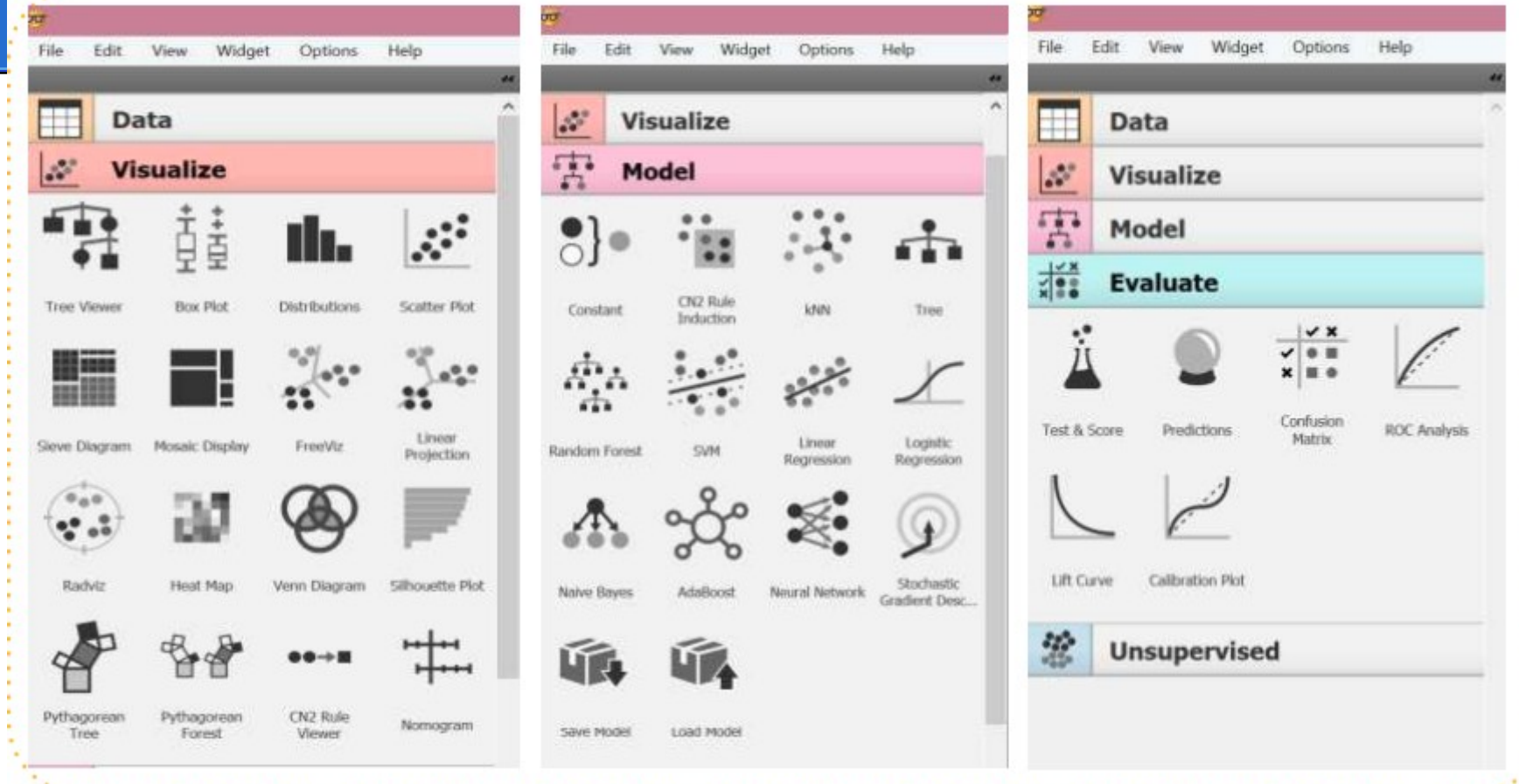


Getting Started With ORANGE!!



Getting Started With ORANGE!!




File Edit View Widget Options Help
Data
Visualize
Tree Viewer
Box Plot
Distributions
Scatter Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Silhouette Plot
Pythagorean Tree
Pythagorean Forest
CN2 Rule Viewer
Nomogram
File Edit View Widget Options Help
Visualize
Model
Constant
CN2 Rule Induction
kNN
Tree
Random Forest
SVM
Linear Regression
Logistic Regression
Naive Bayes
AdaBoost
Neural Network
Stochastic Gradient Desc...
Save Model
Load Model
File Edit View Widget Options Help
Data
Visualize
Model
Evaluate
Test & Score
Predictions
Confusion Matrix
ROC Analysis
Lift Curve
Calibration Plot
Unsupervised

Análise exploratória dos dados – Exploratory DATA analysis

Dataset: Heart Disease

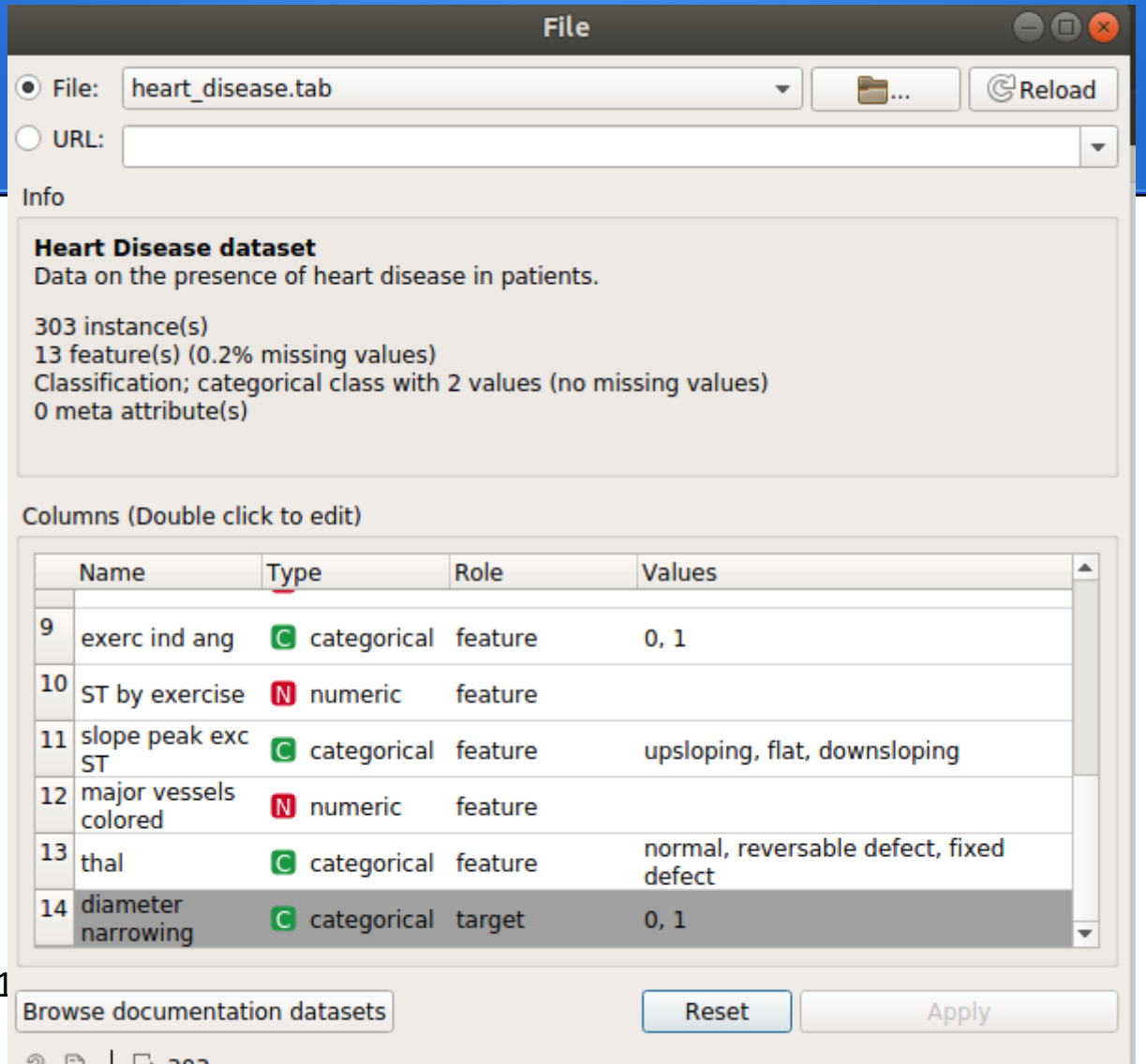
ATTRIBUTES

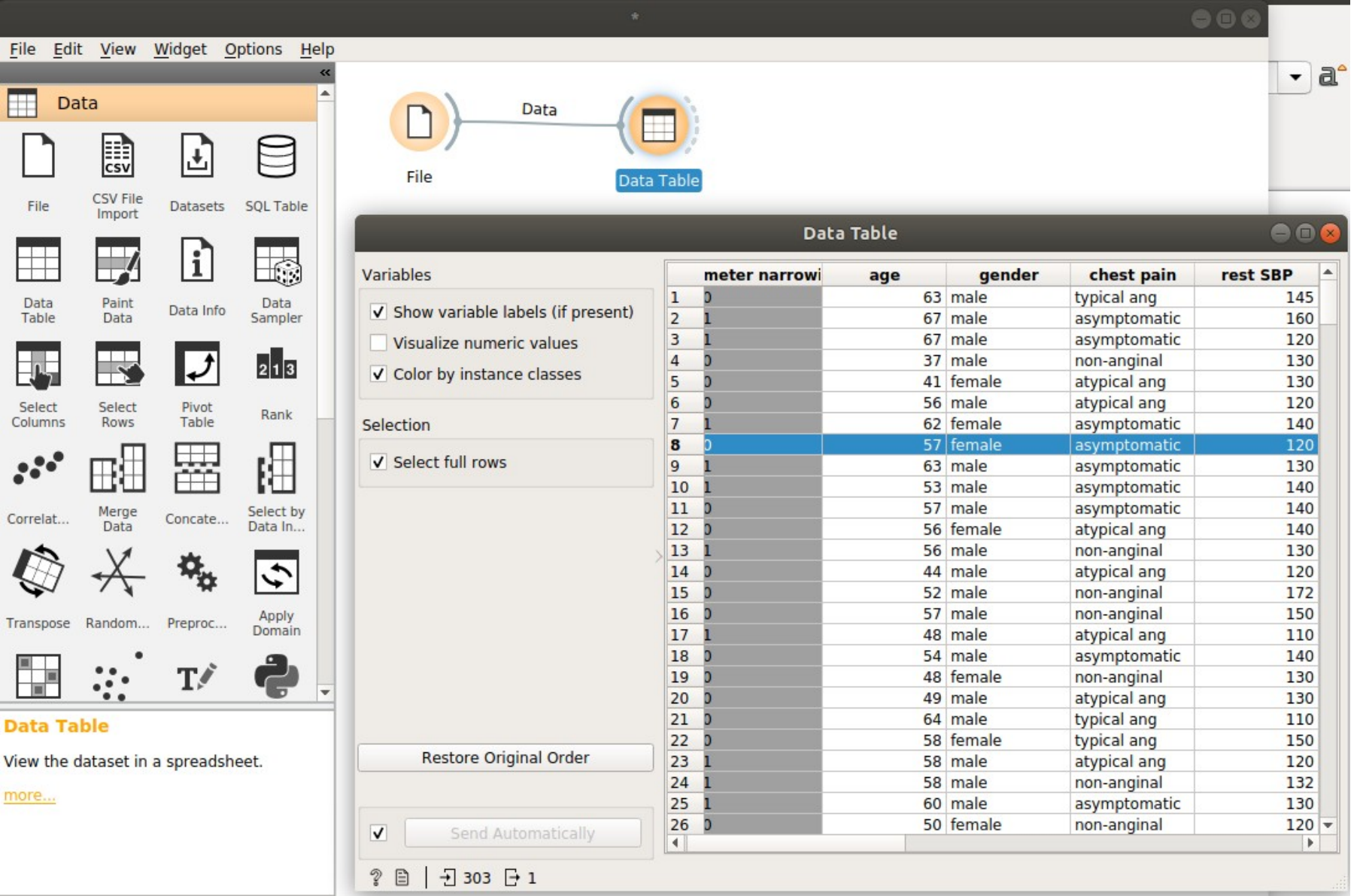
- Narrowing diameter
- Cholesterol
- Chest pain (dor no peito)
- Rest ECG
- Fasting blood sugar
- Max HR (Freq. Cardíaca)
- Age,gender and more

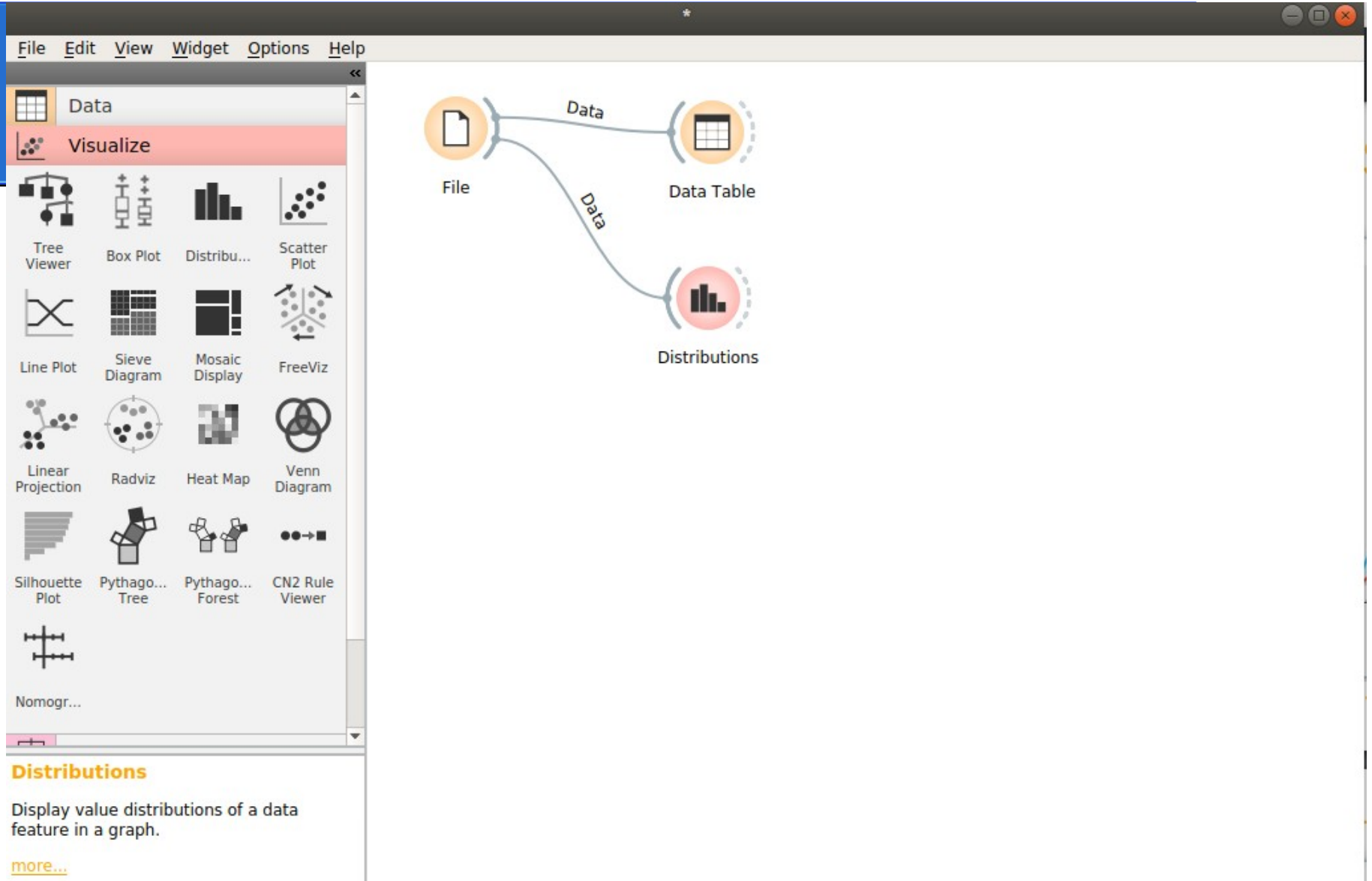
- Has 303 instances
- 13 attributes
- Categorical class with 2 values (0,1)
- In .csv format
- datasets of Orange.

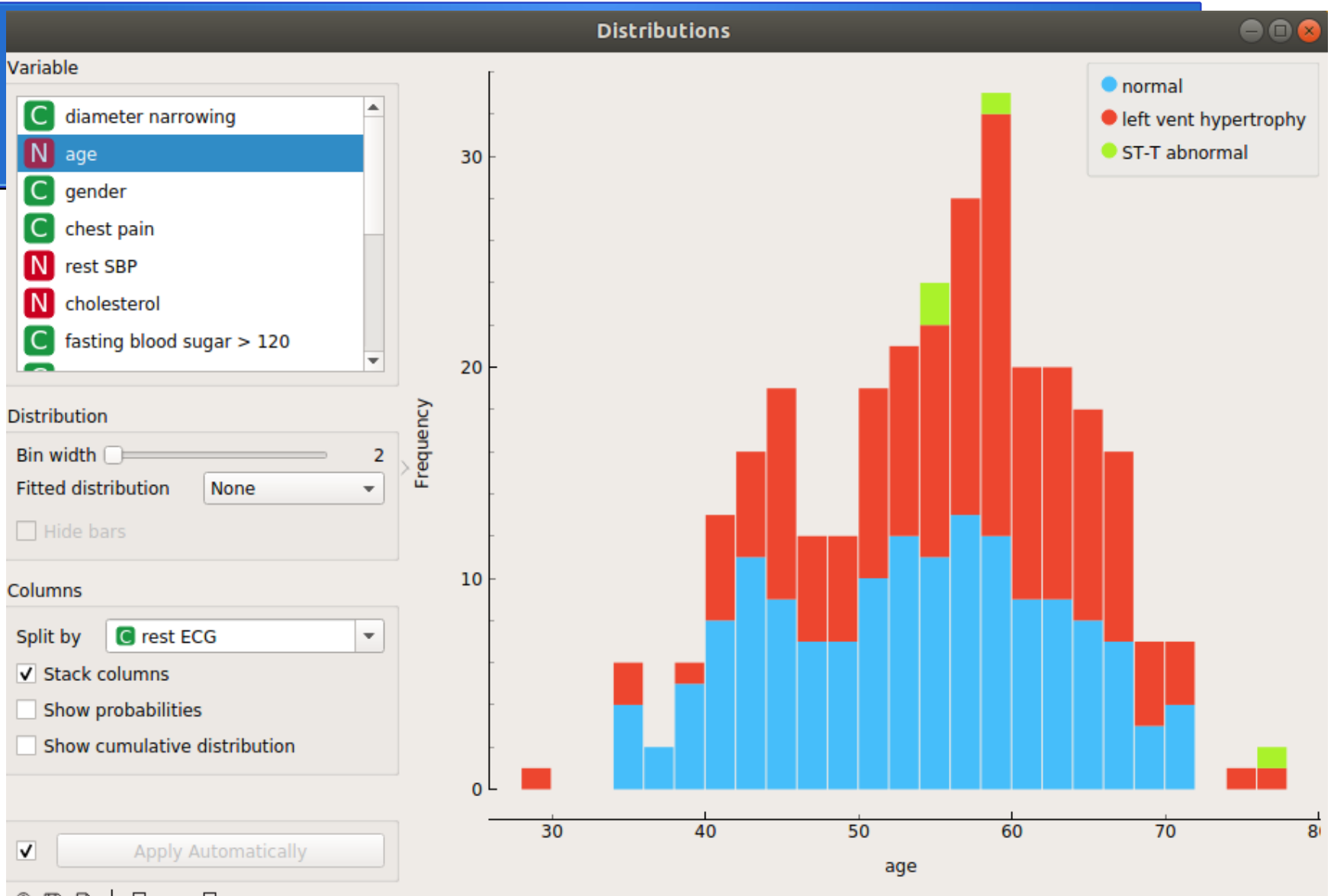
Verificar correlações entre os dados

- Idade: as doenças cardíacas aumentam com a idade acima de 65 anos
- Sexo: a doença cardíaca é a principal causa de morte tanto para homens como para mulheres


File
File:
heart_disease.tab
...
Reload
URL:
Info
Heart Disease dataset
Data on the presence of heart disease in patients.
303 instance(s)
13 feature(s) (0.2% missing values)
Classification; categorical class with 2 values (no missing values)
0 meta attribute(s)
Columns (Double click to edit)
Name
Type
Role
Values
9 exerc ind ang C categorical feature 0, 1
10 ST by exercise N numeric feature
11 slope peak exc ST C categorical feature upsloping, flat, downsloping
12 major vessels colored N numeric feature
13 thal C categorical feature normal, reversable defect, fixed defect
14 diameter narrowing C categorical target 0, 1
Browse documentation datasets
Reset
Apply

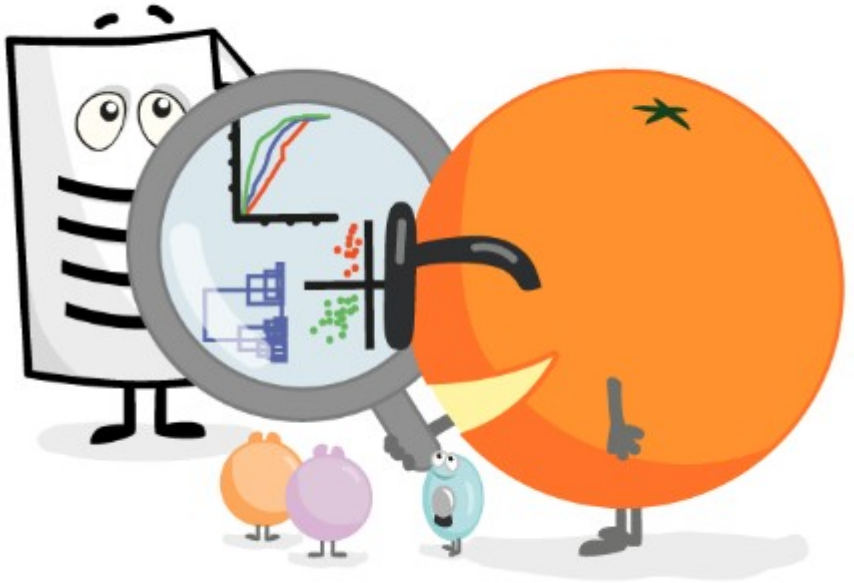

File Edit View Widget Options Help
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlat...
Merge Data
Concate...
Select by Data In...
Transpose
Random...
Preproc...
Apply Domain
Data Table
View the dataset in a spreadsheet.
more...
Data
File
Data Table
Data Table
Variables
Show variable labels (if present)
Visualize numeric values
Color by instance classes
Selection
Select full rows
Restore Original Order
Send Automatically
303
1
meter narrowi
age
gender
chest pain
rest SBP
1 0 63 male typical ang 145
2 1 67 male asymptomatic 160
3 1 67 male asymptomatic 120
4 0 37 male non-anginal 130
5 0 41 female atypical ang 130
6 0 56 male atypical ang 120
7 1 62 female asymptomatic 140
8 0 57 female asymptomatic 120
9 1 63 male asymptomatic 130
10 1 53 male asymptomatic 140
11 0 57 male asymptomatic 140
12 0 56 female atypical ang 140
13 1 56 male non-anginal 130
14 0 44 male atypical ang 120
15 0 52 male non-anginal 172
16 0 57 male non-anginal 150
17 1 48 male atypical ang 110
18 0 54 male asymptomatic 140
19 0 48 female non-anginal 130
20 0 49 male atypical ang 130
21 0 64 male typical ang 110
22 0 58 female typical ang 150
23 1 58 male atypical ang 120
24 1 58 male non-anginal 132
25 1 60 male asymptomatic 130
26 0 50 female non-anginal 120

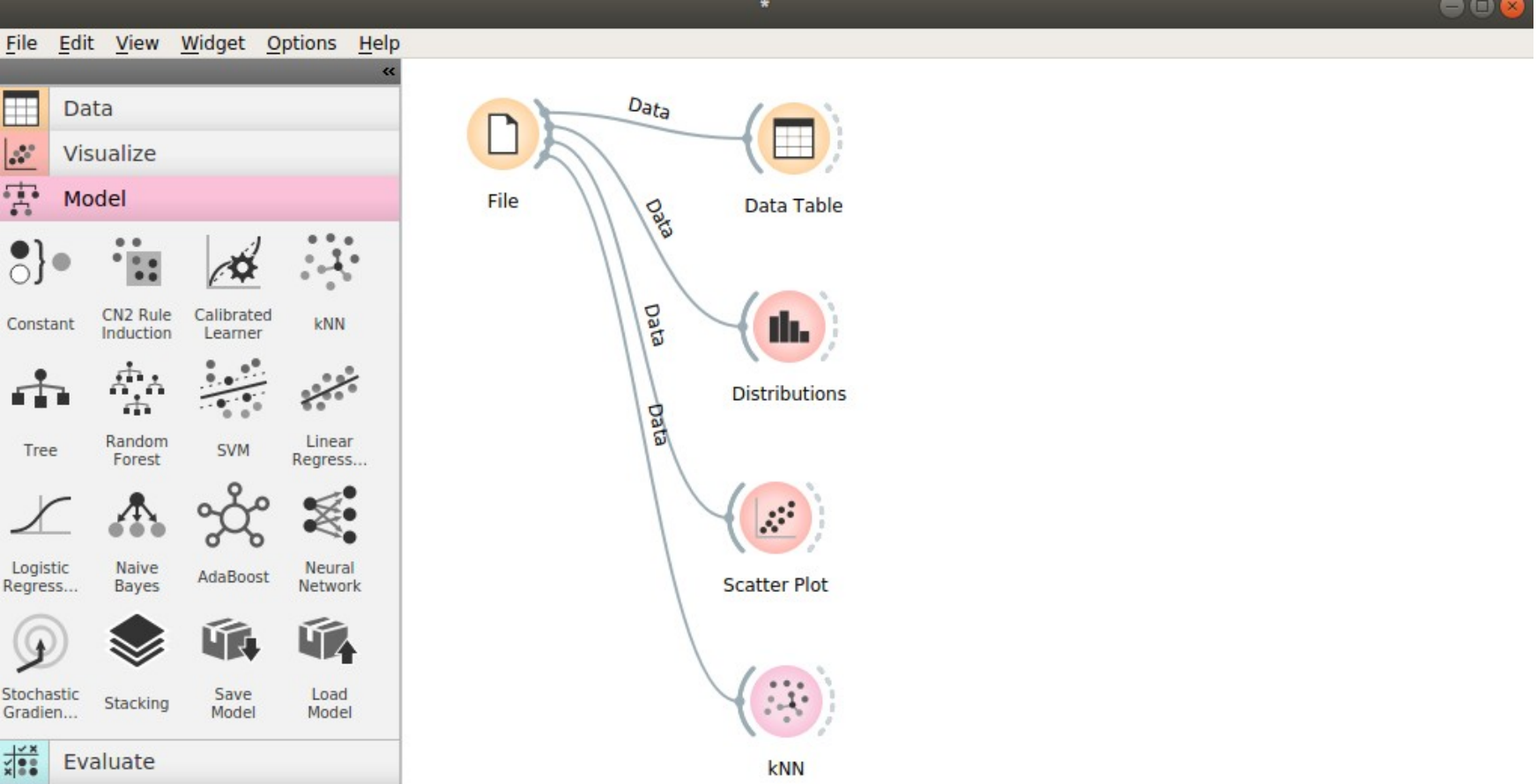

*
File Edit View Widget Options Help
Data
Visualize
Tree Viewer
Box Plot
Distribu...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Silhouette Plot
Pythago... Tree
Pythago... Forest
CN2 Rule Viewer
Nomogr...
Distributions
Display value distributions of a data feature in a graph.
more...
File
Data
Data Table
Data
Distributions

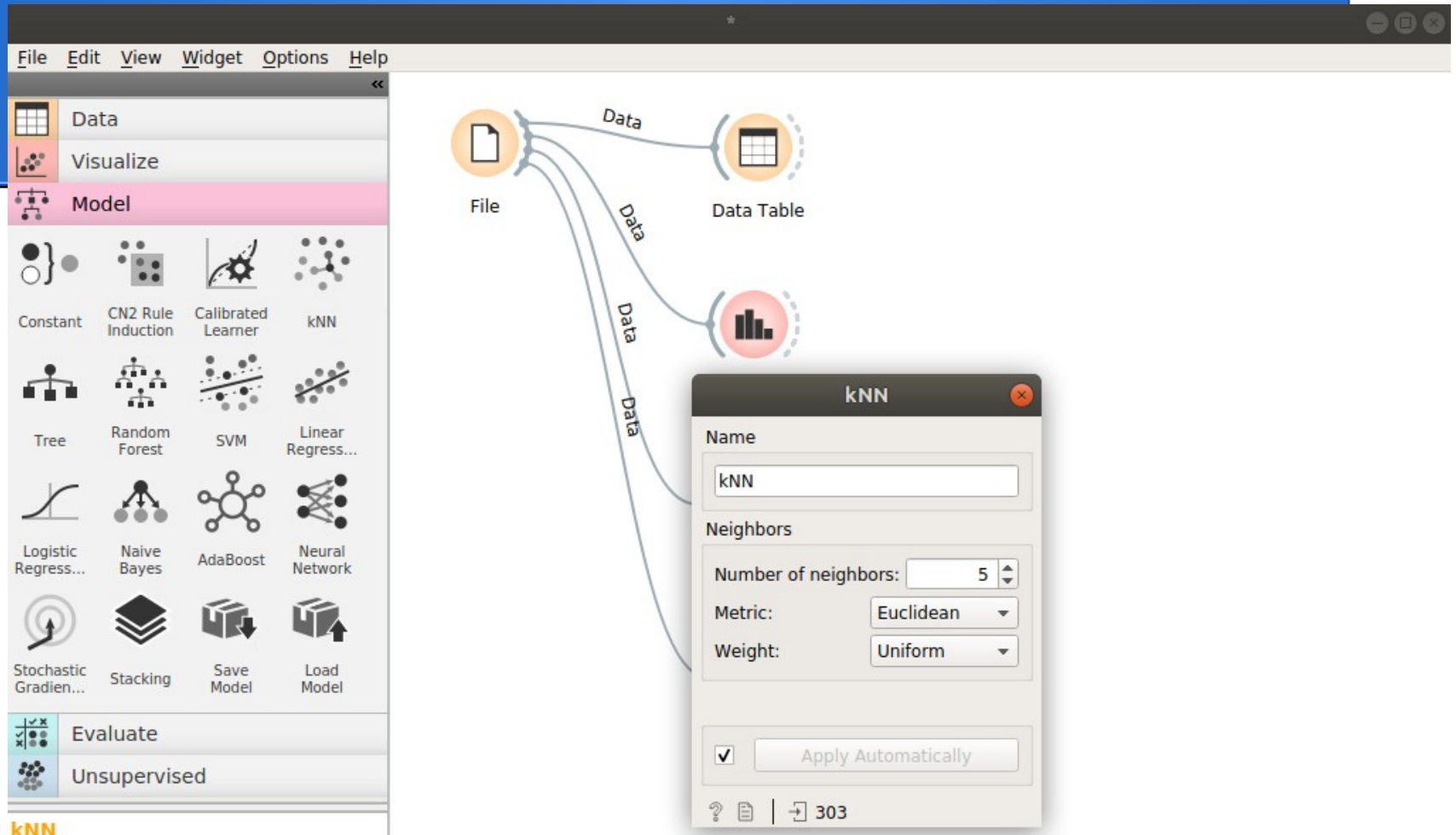

Distributions
Variable
diameter narrowing
age
gender
chest pain
rest SBP
cholesterol
fasting blood sugar > 120
Distribution
Bin width
2
Fitted distribution
None
Hide bars
Columns
Split by
rest ECG
Stack columns
Show probabilities
Show cumulative distribution
Apply Automatically
normal
left vent hypertrophy
ST-T abnormal
Frequency
0
10
20
30
30
40
50
60
70
age

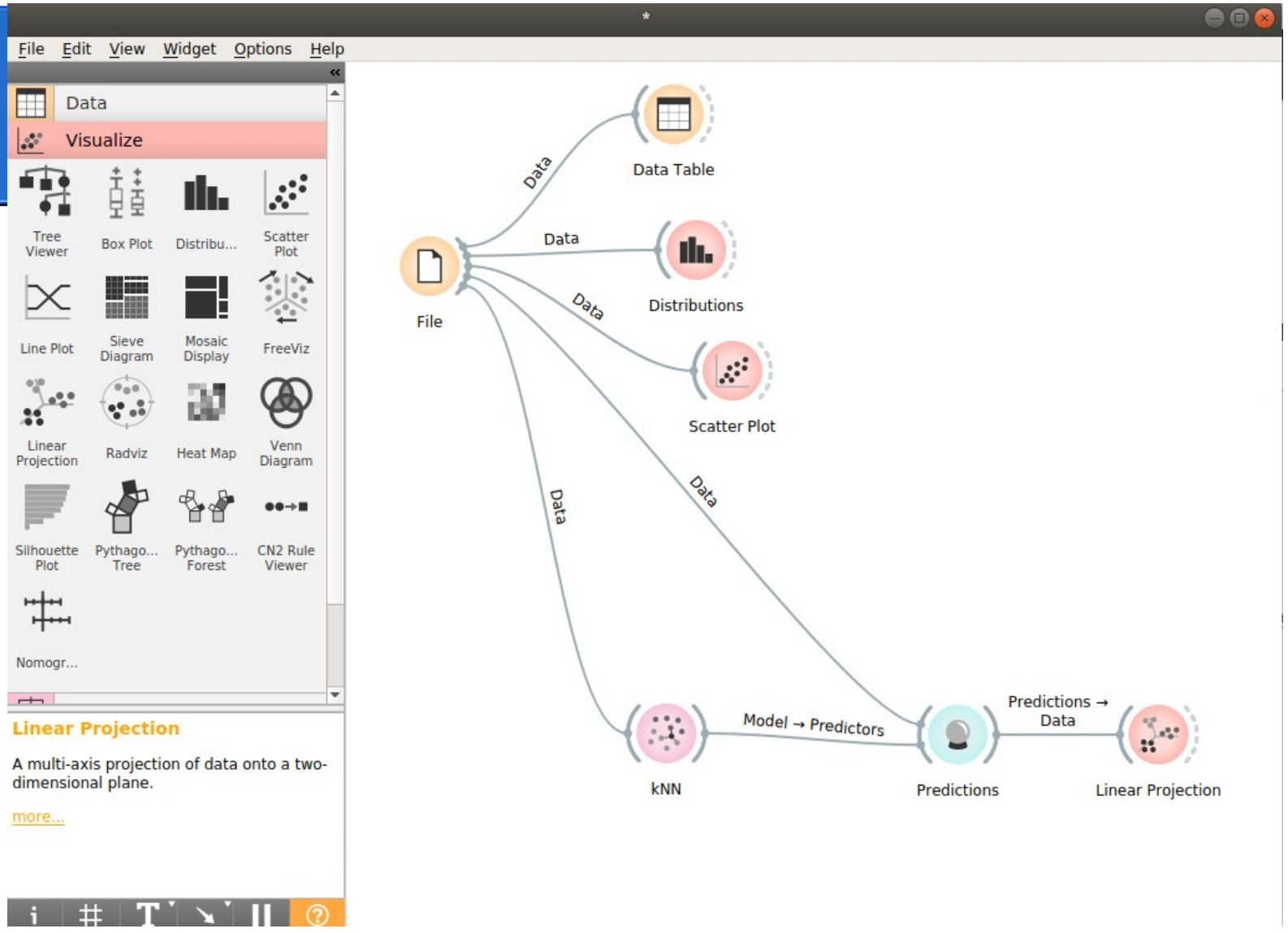
Algorithms:

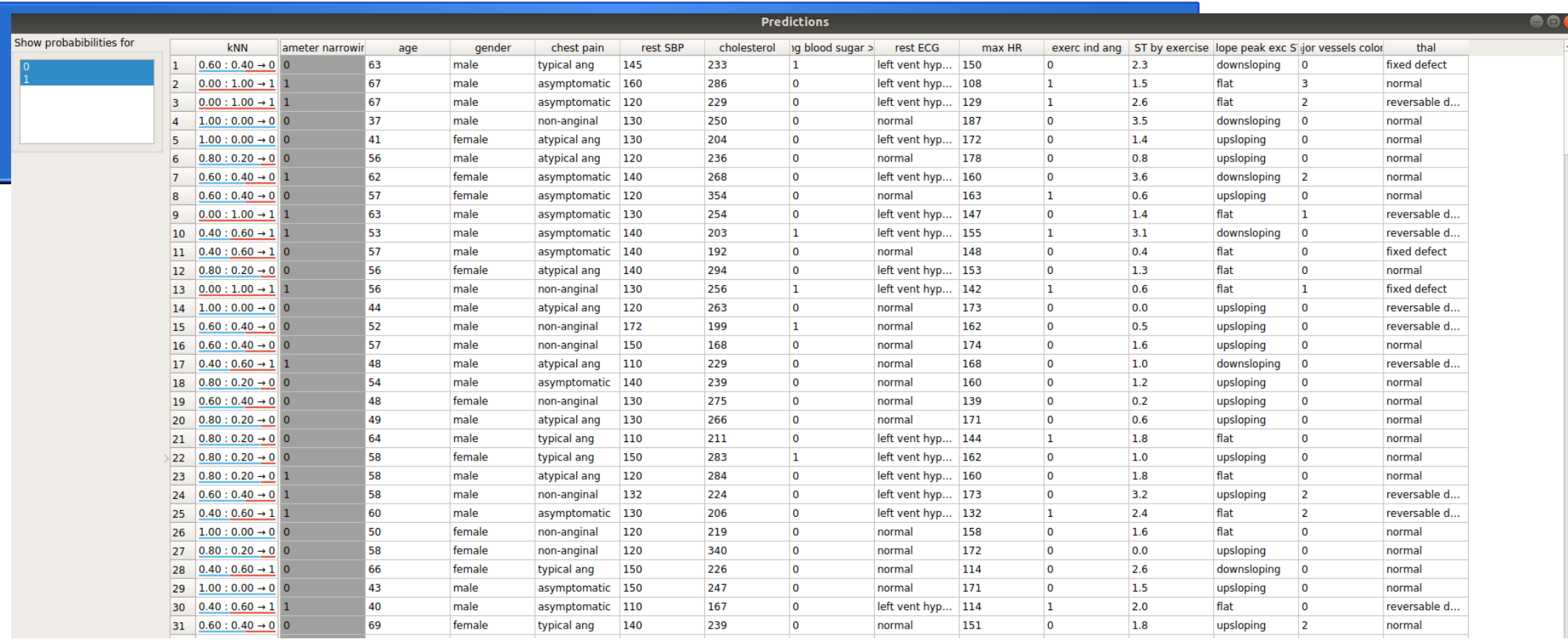
- KNN
- Naïve Bayes
- Decision Tree
- SVM




File Edit View Widget Options Help
Data
Visualize
Model
Constant
CN2 Rule Induction
Calibrated Learner
kNN
Tree
Random Forest
SVM
Linear Regress...
Logistic Regress...
Naive Bayes
AdaBoost
Neural Network
Stochastic Gradien...
Stacking
Save Model
Load Model
Evaluate
Unsupervised
kNN
Predict according to the nearest training instances.
more...
Data
Data
Data
Data
File
Data Table
Distributions
Scatter Plot
kNN


File Edit View Widget Options Help
Data
Visualize
Model
Constant
CN2 Rule Induction
Calibrated Learner
kNN
Tree
Random Forest
SVM
Linear Regress...
Logistic Regress...
Naive Bayes
AdaBoost
Neural Network
Stochastic Gradien...
Stacking
Save Model
Load Model
Evaluate
Unsupervised
kNN
Predict according to the nearest training instances.
more...
File
Data
Data Table
Data
Data
Data
kNN
Name
kNN
Neighbors
Number of neighbors: 5
Metric: Euclidean
Weight: Uniform
Apply Automatically
303


File Edit View Widget Options Help
Data
Visualize
Tree Viewer
Box Plot
Distribu...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Silhouette Plot
Pythago... Tree
Pythago... Forest
CN2 Rule Viewer
Nomogr...
Linear Projection
A multi-axis projection of data onto a two-dimensional plane.
more...
Data
Data Table
Data
Distributions
Data
Scatter Plot
Data
Data
File
kNN
Model → Predictors
Predictions
Predictions → Data
Linear Projection



	kNN	ameter narrowir	age	gender	chest pain	rest SBP	cholesterol	ıg blood sugar >	rest ECG	max HR	exerc ind ang	ST by exercise	lope peak exc S	jor vessels colo	thal
1	0.60 : 0.40 → 0	0	63	male	typical ang	145	233	1	left vent hyp...	150	0	2.3	downsloping	0	fixed defect
2	0.00 : 1.00 → 1	1	67	male	asymptomatic	160	286	0	left vent hyp...	108	1	1.5	flat	3	normal
3	0.00 : 1.00 → 1	1	67	male	asymptomatic	120	229	0	left vent hyp...	129	1	2.6	flat	2	reversable d...
4	1.00 : 0.00 → 0	0	37	male	non-anginal	130	250	0	normal	187	0	3.5	downsloping	0	normal
5	1.00 : 0.00 → 0	0	41	female	atypical ang	130	204	0	left vent hyp...	172	0	1.4	upsloping	0	normal
6	0.80 : 0.20 → 0	0	56	male	atypical ang	120	236	0	normal	178	0	0.8	upsloping	0	normal
7	0.60 : 0.40 → 0	1	62	female	asymptomatic	140	268	0	left vent hyp...	160	0	3.6	downsloping	2	normal
8	0.60 : 0.40 → 0	0	57	female	asymptomatic	120	354	0	normal	163	1	0.6	upsloping	0	normal
9	0.00 : 1.00 → 1	1	63	male	asymptomatic	130	254	0	left vent hyp...	147	0	1.4	flat	1	reversable d...
10	0.40 : 0.60 → 1	1	53	male	asymptomatic	140	203	1	left vent hyp...	155	1	3.1	downsloping	0	reversable d...
11	0.40 : 0.60 → 1	0	57	male	asymptomatic	140	192	0	normal	148	0	0.4	flat	0	fixed defect
12	0.80 : 0.20 → 0	0	56	female	atypical ang	140	294	0	left vent hyp...	153	0	1.3	flat	0	normal
13	0.00 : 1.00 → 1	1	56	male	non-anginal	130	256	1	left vent hyp...	142	1	0.6	flat	1	fixed defect
14	1.00 : 0.00 → 0	0	44	male	atypical ang	120	263	0	normal	173	0	0.0	upsloping	0	reversable d...
15	0.60 : 0.40 → 0	0	52	male	non-anginal	172	199	1	normal	162	0	0.5	upsloping	0	reversable d...
16	0.60 : 0.40 → 0	0	57	male	non-anginal	150	168	0	normal	174	0	1.6	upsloping	0	normal
17	0.40 : 0.60 → 1	1	48	male	atypical ang	110	229	0	normal	168	0	1.0	downsloping	0	reversable d...
18	0.80 : 0.20 → 0	0	54	male	asymptomatic	140	239	0	normal	160	0	1.2	upsloping	0	normal
19	0.60 : 0.40 → 0	0	48	female	non-anginal	130	275	0	normal	139	0	0.2	upsloping	0	normal
20	0.80 : 0.20 → 0	0	49	male	atypical ang	130	266	0	normal	171	0	0.6	upsloping	0	normal
21	0.80 : 0.20 → 0	0	64	male	typical ang	110	211	0	left vent hyp...	144	1	1.8	flat	0	normal
22	0.80 : 0.20 → 0	0	58	female	typical ang	150	283	1	left vent hyp...	162	0	1.0	upsloping	0	normal
23	0.80 : 0.20 → 0	1	58	male	atypical ang	120	284	0	left vent hyp...	160	0	1.8	flat	0	normal
24	0.60 : 0.40 → 0	1	58	male	non-anginal	132	224	0	left vent hyp...	173	0	3.2	upsloping	2	reversable d...
25	0.40 : 0.60 → 1	1	60	male	asymptomatic	130	206	0	left vent hyp...	132	1	2.4	flat	2	reversable d...
26	1.00 : 0.00 → 0	0	50	female	non-anginal	120	219	0	normal	158	0	1.6	flat	0	normal
27	0.80 : 0.20 → 0	0	58	female	non-anginal	120	340	0	normal	172	0	0.0	upsloping	0	normal
28	0.40 : 0.60 → 1	0	66	female	typical ang	150	226	0	normal	114	0	2.6	downsloping	0	normal
29	1.00 : 0.00 → 0	0	43	male	asymptomatic	150	247	0	normal	171	0	1.5	upsloping	0	normal
30	0.40 : 0.60 → 1	1	40	male	asymptomatic	110	167	0	left vent hyp...	114	1	2.0	flat	0	reversable d...
31	0.60 : 0.40 → 0	0	69	female	typical ang	140	239	0	normal	151	0	1.8	upsloping	2	normal

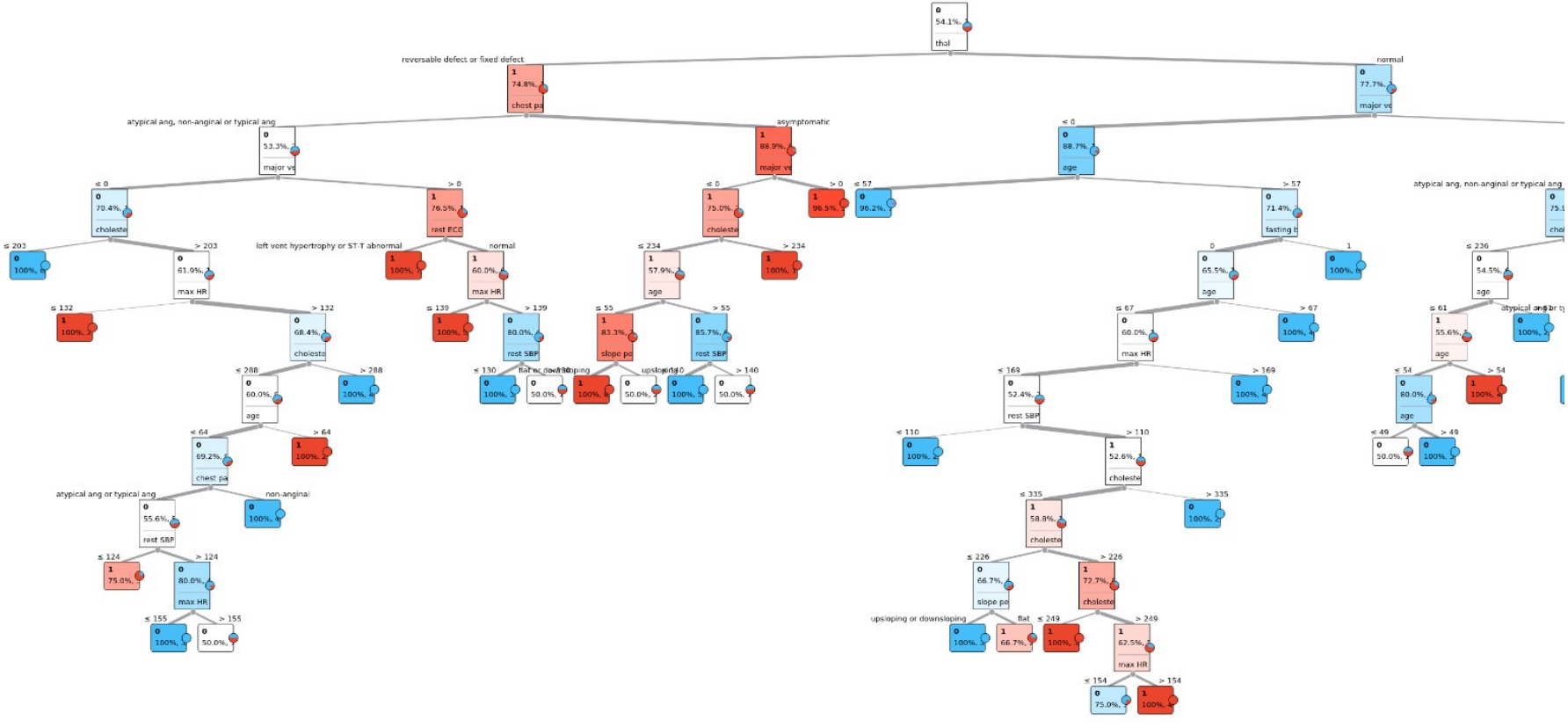
ree Viewer
ree Viewer

Back to Last Scheme

Save | Print

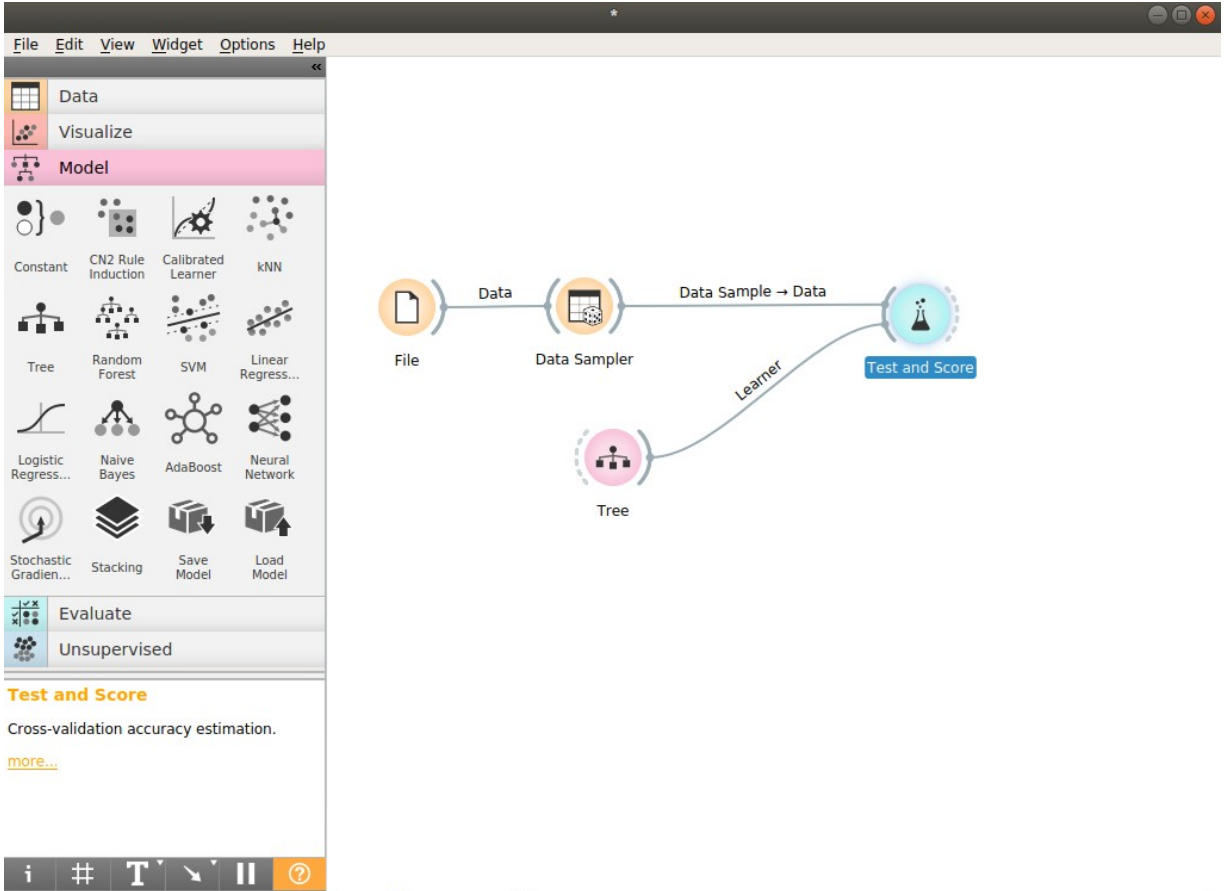
Tree Viewer sex jun 11 21, 20:15:27

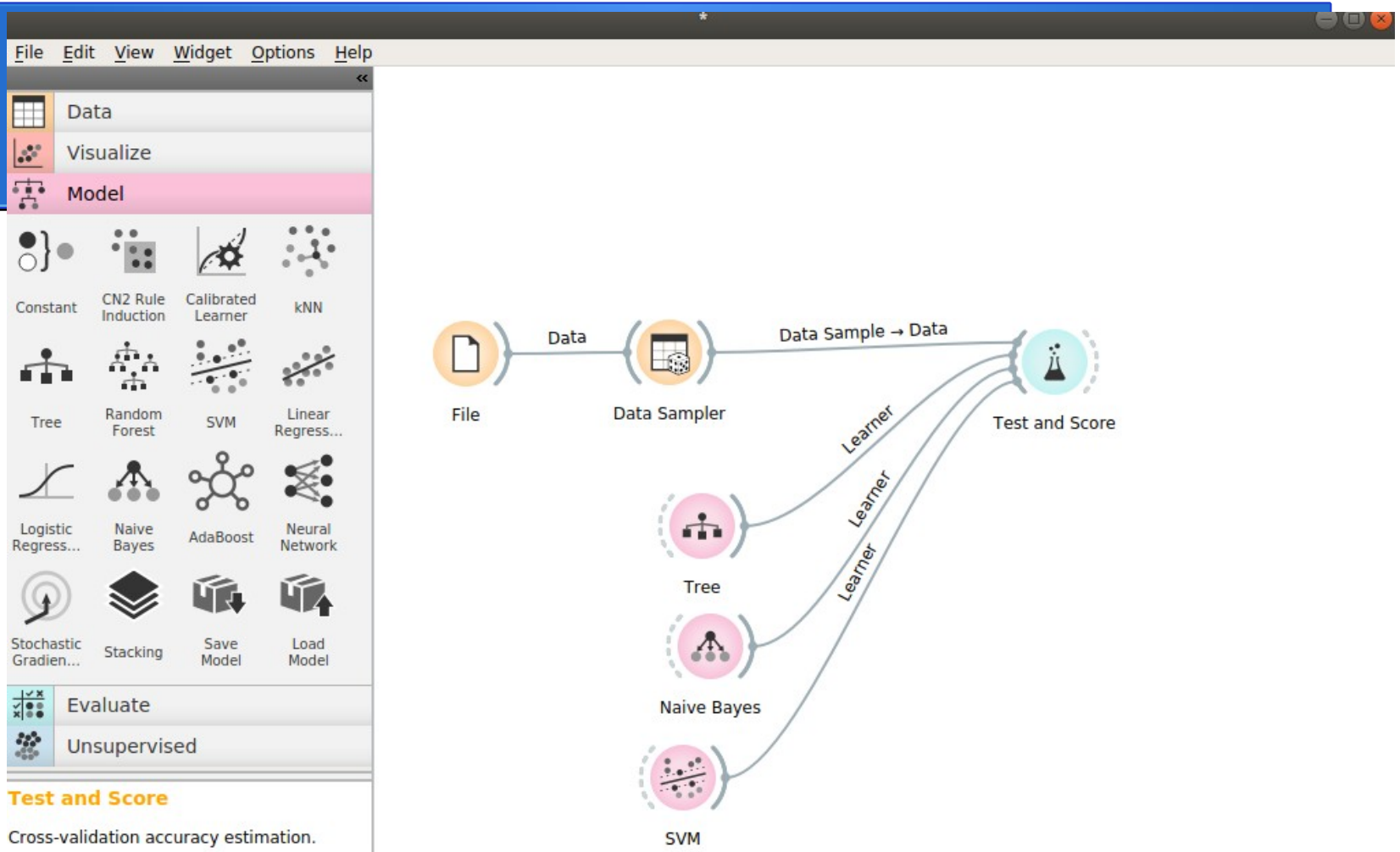
Tree size: 77 nodes, 39 leaves
Edge widths: Relative to parent
Target class: None

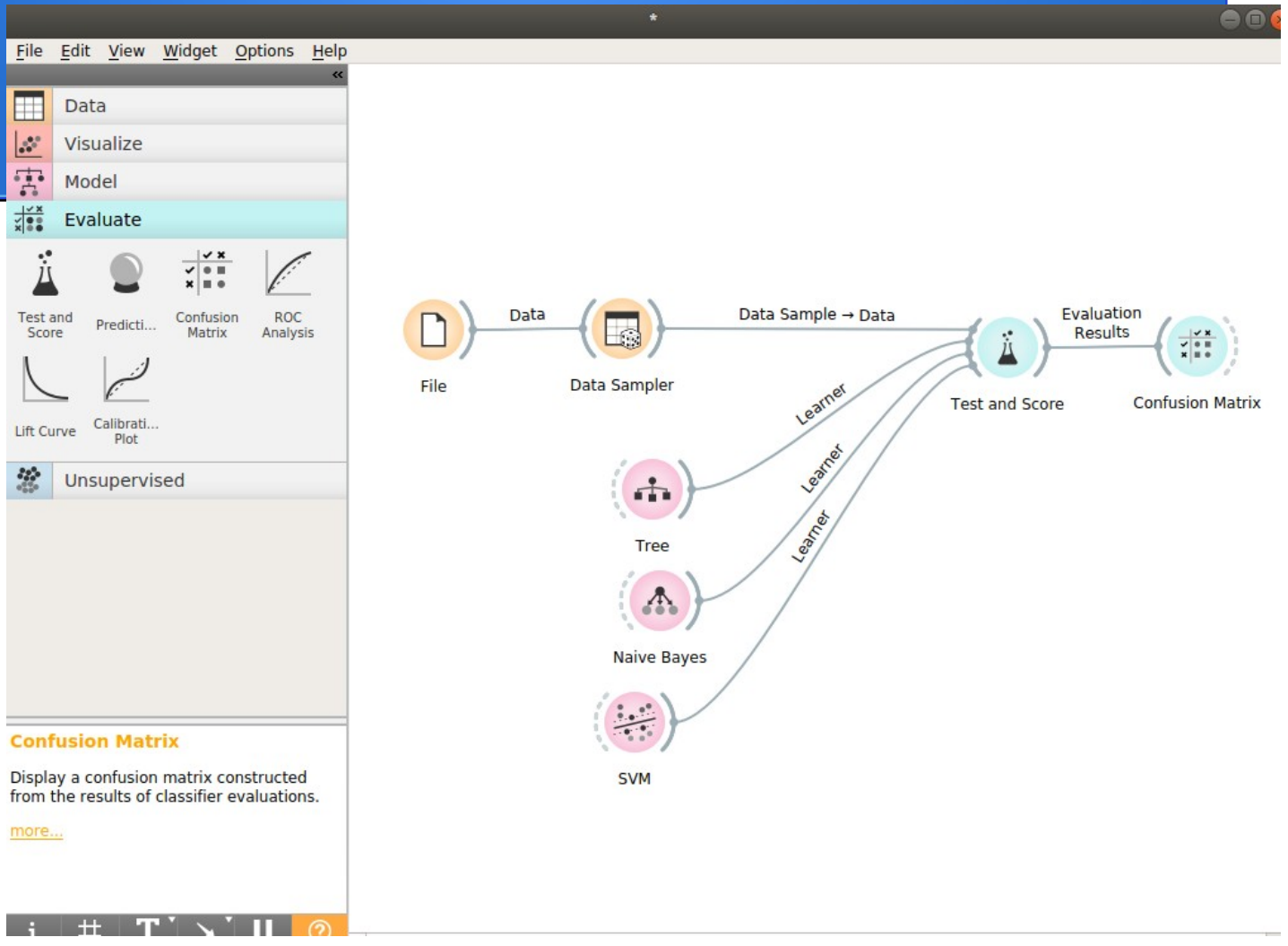


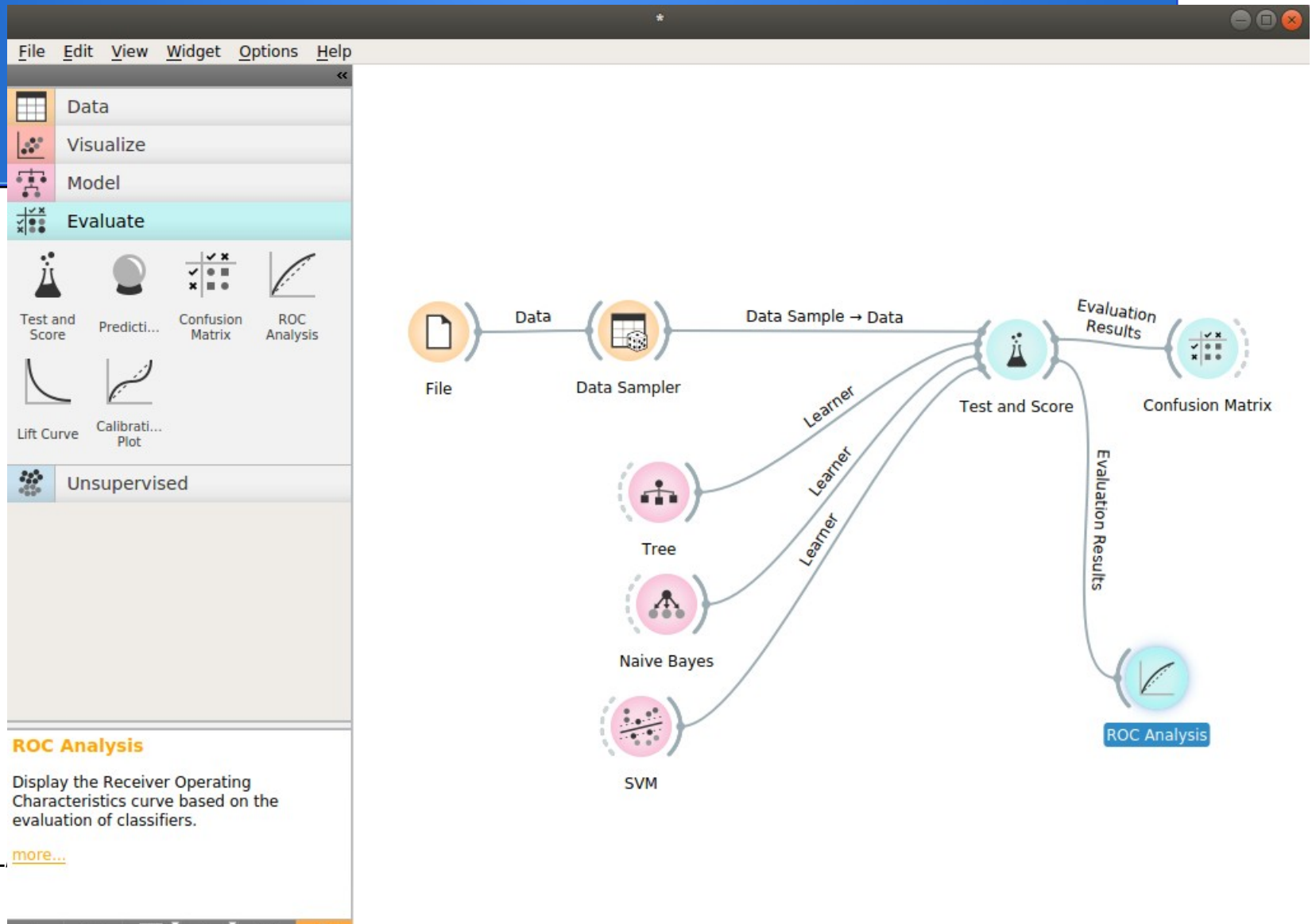
Write a comment...

Validação Cruzada

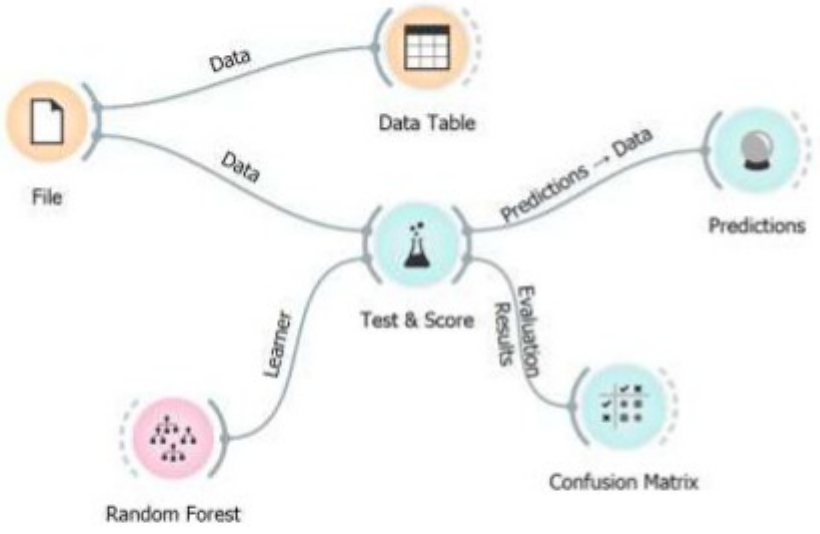



File
Edit
View
Widget
Options
Help
Data
Visualize
Model
Constant
CN2 Rule Induction
Calibrated Learner
kNN
Tree
Random Forest
SVM
Linear Regress...
Logistic Regress...
Naive Bayes
AdaBoost
Neural Network
Stochastic Gradien...
Stacking
Save Model
Load Model
Evaluate
Unsupervised
Test and Score
Cross-validation accuracy estimation.
more...
Data
Data Sample → Data
Learner
Learner
Learner
File
Data Sampler
Test and Score
Tree
Naive Bayes
SVM


*
File Edit View Widget Options Help
Data
Visualize
Model
Evaluate
Test and Score
Predicti...
Confusion Matrix
ROC Analysis
Lift Curve
Calibrati... Plot
Unsupervised
Confusion Matrix
Display a confusion matrix constructed from the results of classifier evaluations.
more...
File
Data
Data Sampler
Data Sample → Data
Test and Score
Evaluation Results
Confusion Matrix
Learner
Tree
Learner
Naive Bayes
Learner
SVM


*
File Edit View Widget Options Help
Data
Visualize
Model
Evaluate
Test and Score
Predicti...
Confusion Matrix
ROC Analysis
Lift Curve
Calibrati... Plot
Unsupervised
ROC Analysis
Display the Receiver Operating Characteristics curve based on the evaluation of classifiers.
more...
File
Data
Data Sampler
Data Sample → Data
Test and Score
Evaluation Results
Confusion Matrix
Learner
Tree
Learner
Naive Bayes
Learner
SVM
Evaluation Results
ROC Analysis


Orange Canvas
File Edit View Widget Options Help
Data
File
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Rank
Merge Data
Concatenate


Data
Data Table
File
Data
Predictions → Data
Predictions
Test & Score
Learner
Evaluation Results
Random Forest
Confusion Matrix

Referências

https://orange.biolab.si/getting-started/
https://orangedatamining.com/