# Select Columns

Manual selection of data attributes and composition of data domain.

**Inputs**

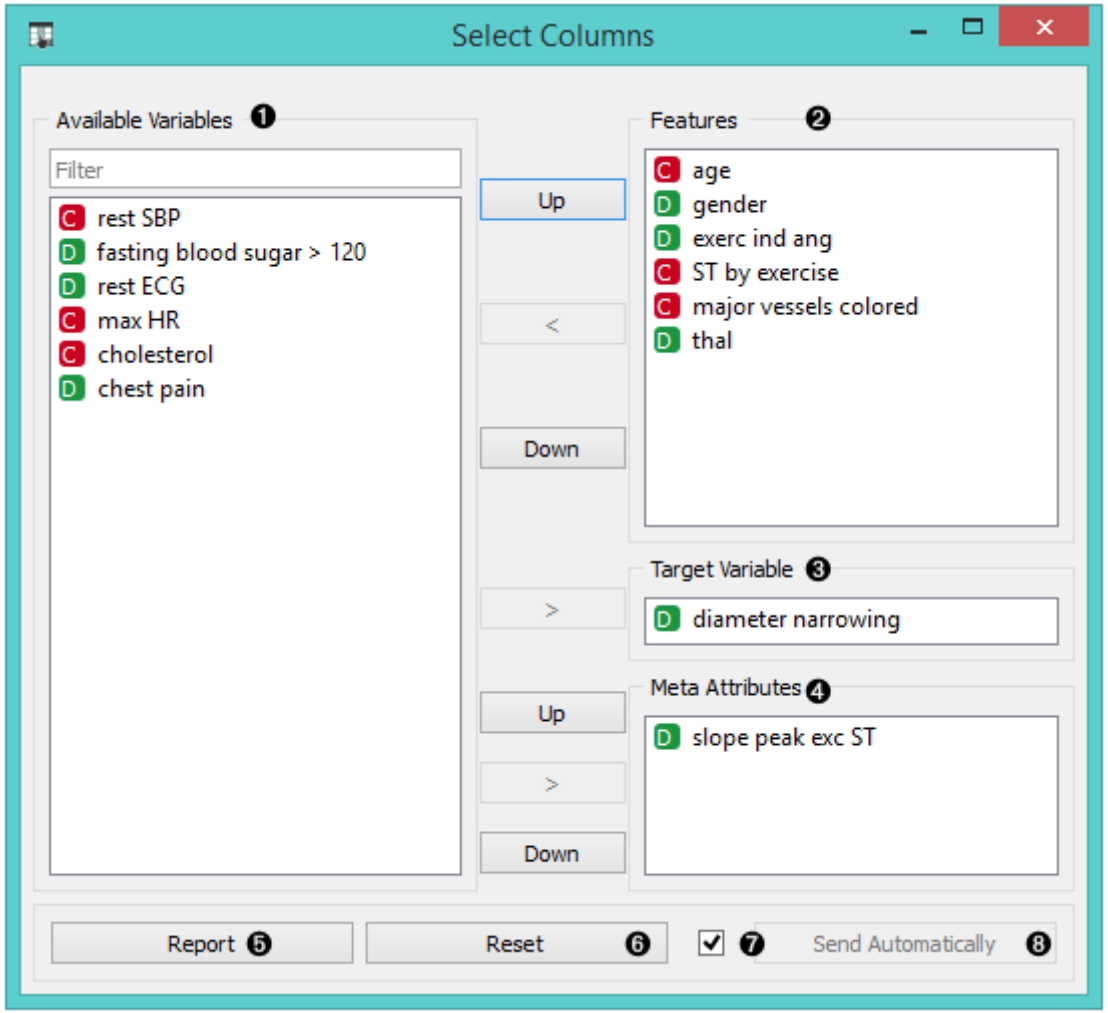
- Data: input dataset

**Outputs**

- Data: dataset with columns as set in the widget

The **Select Columns** widget is used to manually compose your data domain. The user can decide which attributes will be used and how. Orange distinguishes between ordinary attributes, (optional) class attributes and meta attributes. For instance, for building a classification model, the domain would be composed of a set of attributes and a discrete class attribute. Meta attributes are not used in modeling, but several widgets can use them as instance labels.

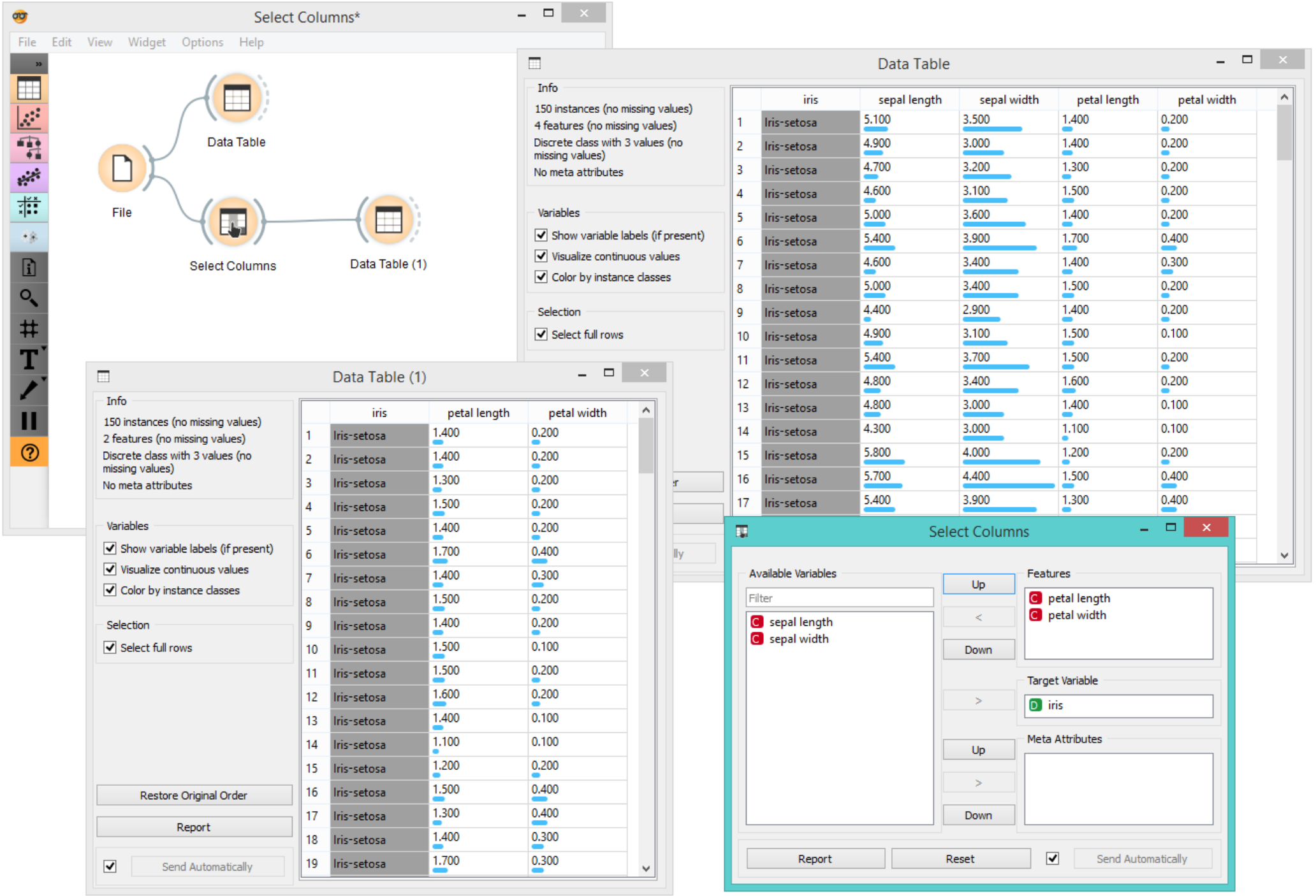
Orange attributes have a type and are either discrete, continuous or a character string. The attribute type is marked with a symbol appearing before the name of the attribute (D, C, S, respectively).



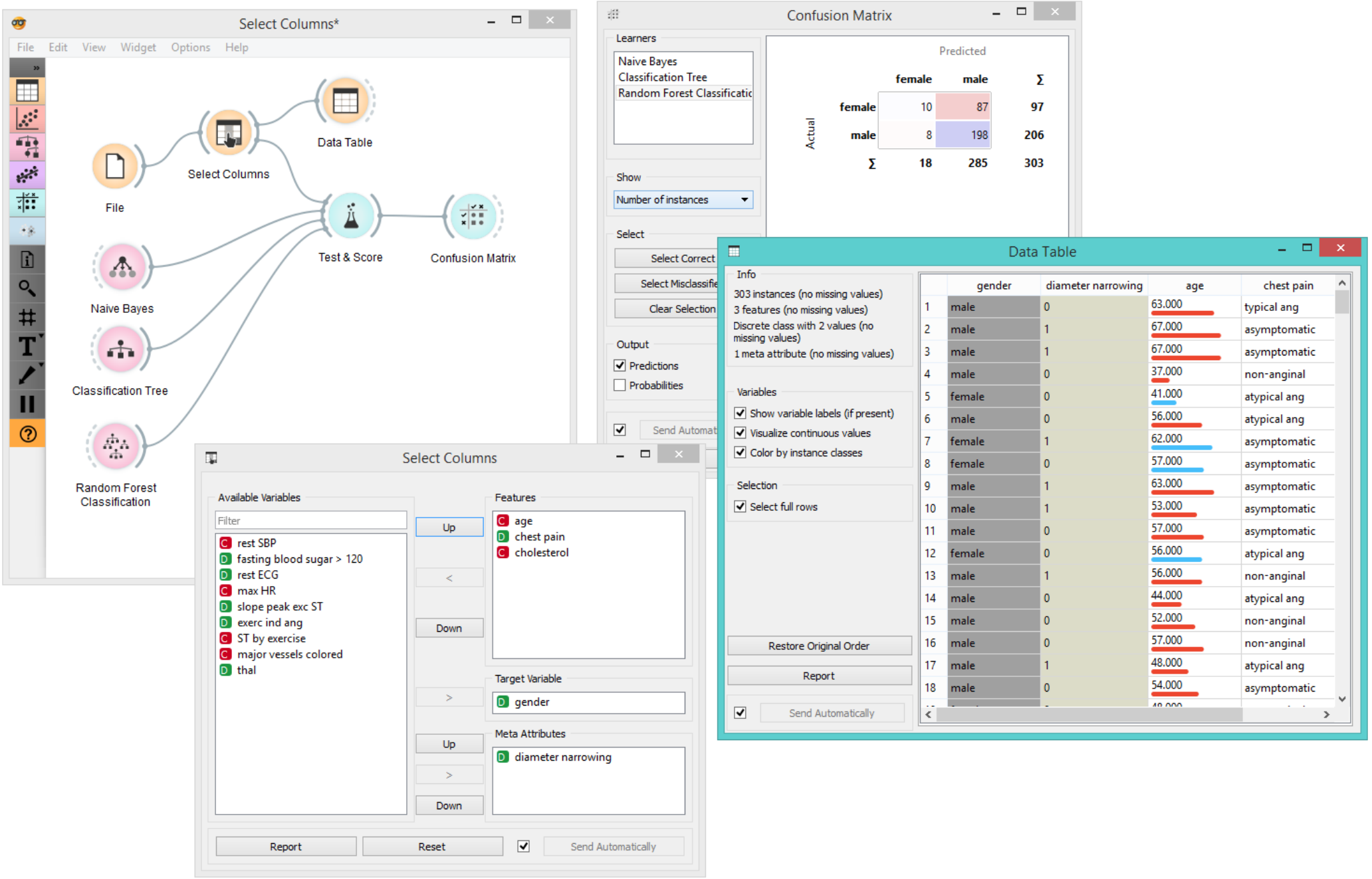

1. Left-out data attributes that will not be in the output data file
2. Data attributes in the new data file
3. Target variable. If none, the new dataset will be without a target variable.
4. Meta attributes of the new data file. These attributes are included in the dataset but are, for most methods, not considered in the analysis.
5. Produce a report.
6. Reset the domain composition to that of the input data file.
7. Tick if you wish to auto-apply changes of the data domain.
8. Apply changes of the data domain and send the new data file to the output channel of the widget.

# Examples

In the workflow below, the *Iris* data from the File widget is fed into the **Select Columns** widget, where we select to output only two attributes (namely petal width and petal length). We view both the original dataset and the dataset with selected columns in the Data Table widget.


Select Columns*
File Edit View Widget Options Help
Data Table
File
Select Columns
Data Table (1)
Data Table
Info
150 instances (no missing values)
4 features (no missing values)
Discrete class with 3 values (no missing values)
No meta attributes
Variables
Show variable labels (if present)
Visualize continuous values
Color by instance classes
Selection
Select full rows
iris sepal length sepal width petal length petal width
1 Iris-setosa 5.100 3.500 1.400 0.200
2 Iris-setosa 4.900 3.000 1.400 0.200
3 Iris-setosa 4.700 3.200 1.300 0.200
4 Iris-setosa 4.600 3.100 1.500 0.200
5 Iris-setosa 5.000 3.600 1.400 0.200
6 Iris-setosa 5.400 3.900 1.700 0.400
7 Iris-setosa 4.600 3.400 1.400 0.300
8 Iris-setosa 5.000 3.400 1.500 0.200
9 Iris-setosa 4.400 2.900 1.400 0.200
10 Iris-setosa 4.900 3.100 1.500 0.100
11 Iris-setosa 5.400 3.700 1.500 0.200
12 Iris-setosa 4.800 3.400 1.600 0.200
13 Iris-setosa 4.800 3.000 1.400 0.100
14 Iris-setosa 4.300 3.000 1.100 0.100
15 Iris-setosa 5.800 4.000 1.200 0.200
16 Iris-setosa 5.700 4.400 1.500 0.400
17 Iris-setosa 5.400 3.900 1.300 0.400
Data Table (1)
Info
150 instances (no missing values)
2 features (no missing values)
Discrete class with 3 values (no missing values)
No meta attributes
Variables
Show variable labels (if present)
Visualize continuous values
Color by instance classes
Selection
Select full rows
Restore Original Order
Report
Send Automatically
iris petal length petal width
1 Iris-setosa 1.400 0.200
2 Iris-setosa 1.400 0.200
3 Iris-setosa 1.300 0.200
4 Iris-setosa 1.500 0.200
5 Iris-setosa 1.400 0.200
6 Iris-setosa 1.700 0.400
7 Iris-setosa 1.400 0.300
8 Iris-setosa 1.500 0.200
9 Iris-setosa 1.400 0.200
10 Iris-setosa 1.500 0.100
11 Iris-setosa 1.500 0.200
12 Iris-setosa 1.600 0.200
13 Iris-setosa 1.400 0.100
14 Iris-setosa 1.100 0.100
15 Iris-setosa 1.200 0.200
16 Iris-setosa 1.500 0.400
17 Iris-setosa 1.300 0.400
18 Iris-setosa 1.400 0.300
19 Iris-setosa 1.700 0.300
Select Columns
Available Variables
Filter
sepal length
sepal width
Up
<
Down
Features
petal length
petal width
Target Variable
iris
Meta Attributes
Up
>
Down
Report
Reset
Send Automatically

For a more complex use of the widget, we composed a workflow to redefine the classification problem in the *heart-disease* dataset. Originally, the task was to predict if the patient has a coronary artery diameter narrowing. We changed the problem to that of gender classification, based on age, chest pain and cholesterol level, and informatively kept the diameter narrowing as a meta attribute.


Select Columns*
File Edit View Widget Options Help
File
Select Columns
Data Table
Test & Score
Confusion Matrix
Naive Bayes
Classification Tree
Random Forest Classification
Confusion Matrix
Learners
Naive Bayes
Classification Tree
Random Forest Classificatic
Show
Number of instances
Select
Select Correct
Select Misclassifie
Clear Selection
Output
Predictions
Probabilities
Send Automat
Predicted
female male Σ
Actual
female 10 87 97
male 8 198 206
Σ 18 285 303
Data Table
Info
303 instances (no missing values)
3 features (no missing values)
Discrete class with 2 values (no missing values)
1 meta attribute (no missing values)
Variables
Show variable labels (if present)
Visualize continuous values
Color by instance classes
Selection
Select full rows
Restore Original Order
Report
Send Automatically
gender diameter narrowing age chest pain
1 male 0 63.000 typical ang
2 male 1 67.000 asymptomatic
3 male 1 67.000 asymptomatic
4 male 0 37.000 non-anginal
5 female 0 41.000 atypical ang
6 male 0 56.000 atypical ang
7 female 1 62.000 asymptomatic
8 female 0 57.000 asymptomatic
9 male 1 63.000 asymptomatic
10 male 1 53.000 asymptomatic
11 male 0 57.000 asymptomatic
12 female 0 56.000 atypical ang
13 male 1 56.000 non-anginal
14 male 0 44.000 atypical ang
15 male 0 52.000 non-anginal
16 male 0 57.000 non-anginal
17 male 1 48.000 atypical ang
18 male 0 54.000 asymptomatic
Select Columns
Available Variables
Filter
rest SBP
fasting blood sugar > 120
rest ECG
max HR
slope peak exc ST
exerc ind ang
ST by exercise
major vessels colored
thal
Up
<
Down
Features
age
chest pain
cholesterol
Target Variable
gender
Meta Attributes
diameter narrowing
Report
Reset
Send Automatically