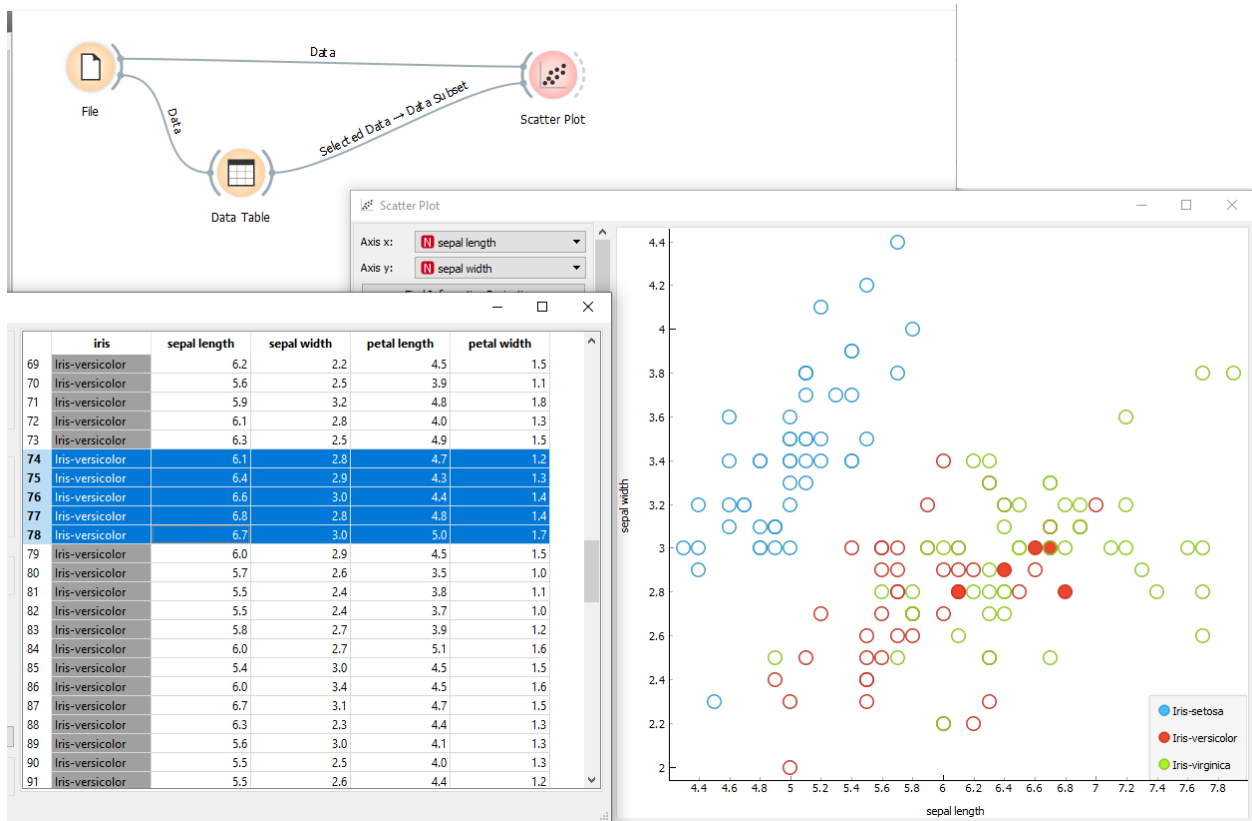


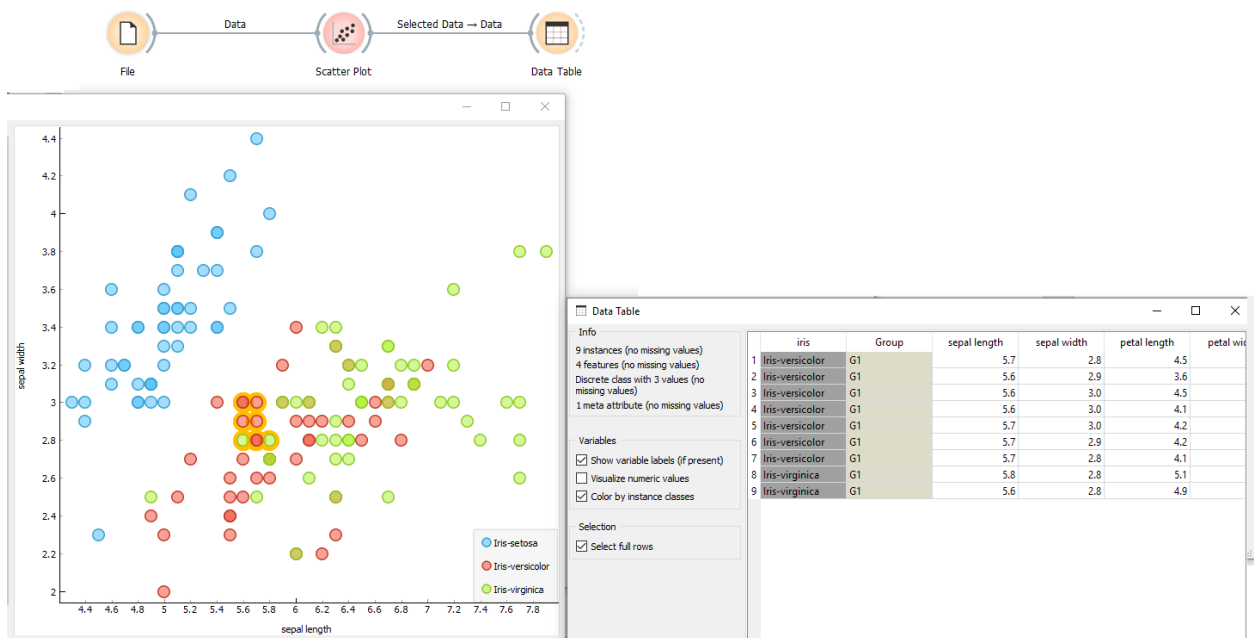
# ORANGE WORKFLOWS

## INTERACTIVE VISUALIZATION IN ORANGE



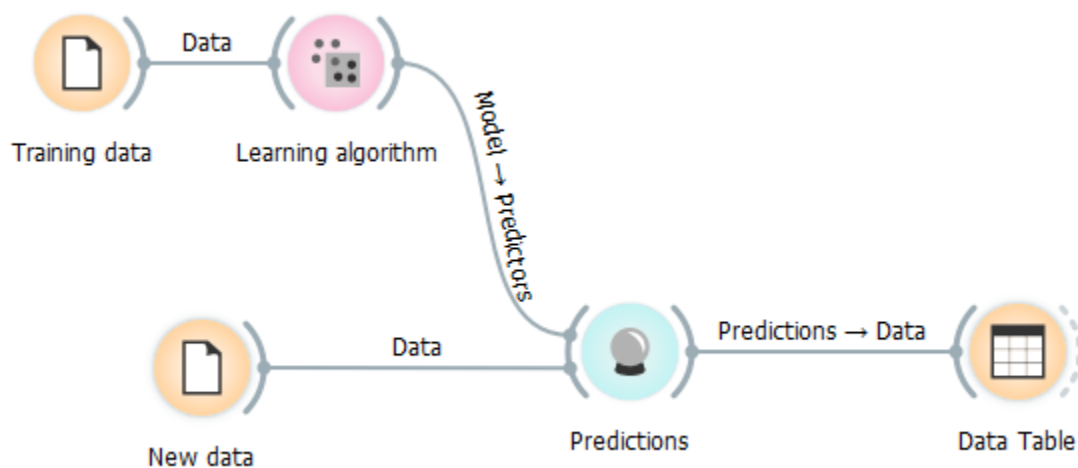
- The widgets File, Data Table and Scatter Plot are connected to form a visual program.
- The selected examples in the Data Table widget are displayed as full circles in the Scatterplot.
- Note: Scatter Plot has two inputs: Data and Data subset and they need to be connected correctly.

Interactive visualization in Orange 2



- The same widgets composed into a different visual program.
- The selected examples in Scatter Plot are shown in Data Table.

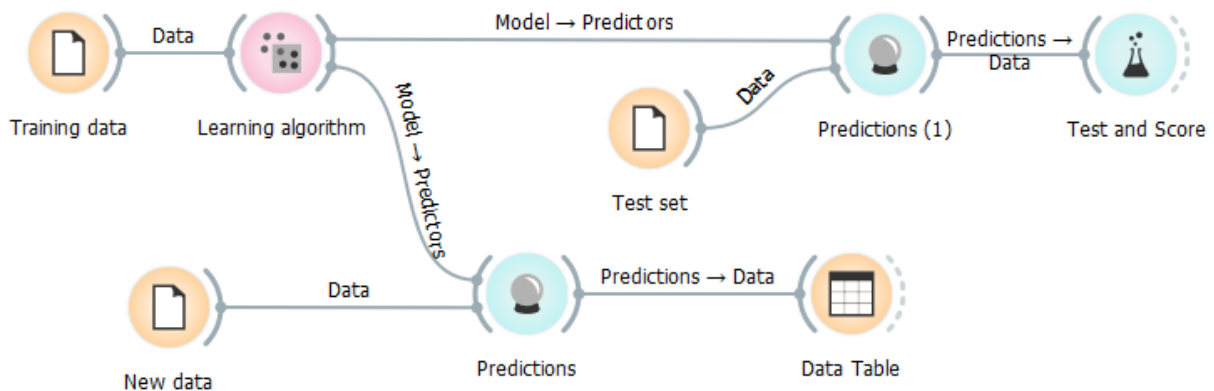
## BASIC CLASSIFICATION SCHEMA IN ORANGE



- We train the model on the train set
- We predict the target for the new instances
- There are several classification algorithms:

- Decision trees
- Naive Bayes classifier
- K nearest neighbors (KNN)
- Artificial neural networks (ANN)
- ....

## CLASSIFICATION WITH EVALUATION



- We train the model on the train set
- We evaluate on the test set
- We classify the new instances

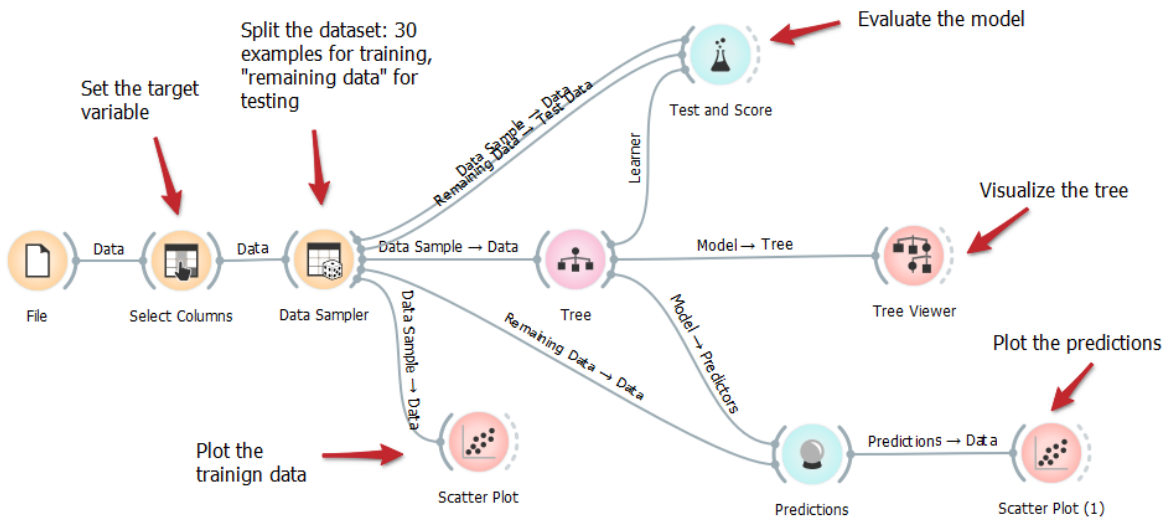
## DECISION TREE LANGUAGE BIAS

- Use a spreadsheet program (e.g. MS Excel) to generate 1000 examples:
  - Attributes A, B and C should have random values
  - Target variable „A>B“, should have value „true“ if A>B else “false”
  - Save the file
- Use Orange trees to predict „A>B“ from the attributes A, B in C
  - Set the target variable
  - Use separate test set for validation
  - Plot the training and classified data in “Scatter Plot”
- How good is your model?
- How does the training set size influence the model performance?

- MS Excel hints:

= RAND()

= IF(A2>B2,"true","false")



- Add evaluation: test on test set
- Observe the performance with different siyes of training set and different sampling random seeds.