

# Introduction to Orange

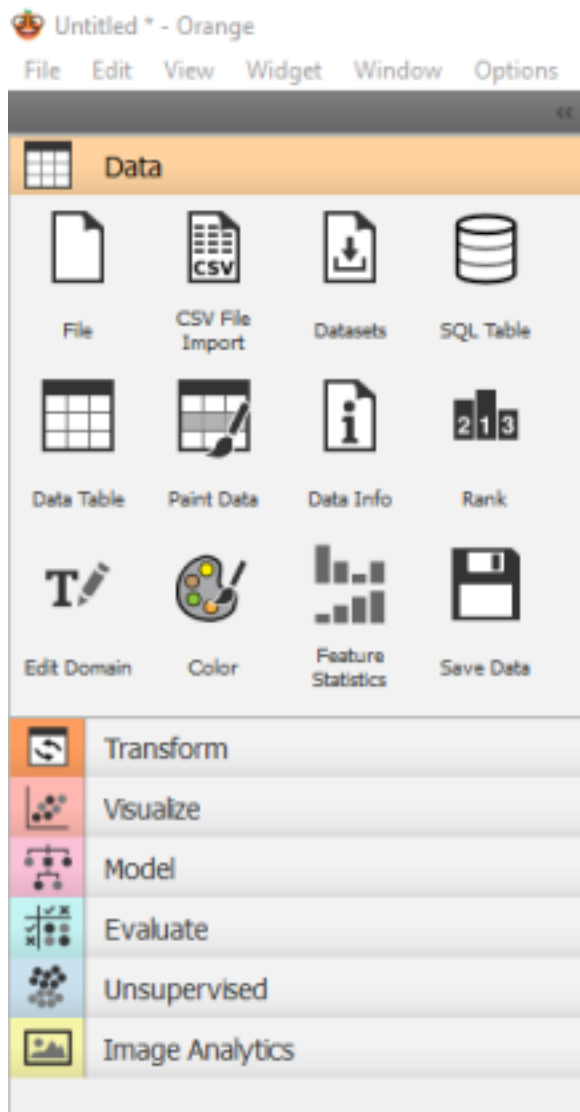
<https://orangedatamining.com/>

Orange is a **powerful platform to perform data analysis and visualization, see data flow and become more** productive.

Open source machine learning and data visualization. Build data analysis workflows visually, with a large, diverse toolbox.

## Purpose

Orange provides a dynamic domain for developers, analysts, and data mining specialists. Python is a new generation scripting language and coding environment, where our data mining scripts can be simply but dynamic. Orange uses an element-based approach for quick prototyping.



# Data

## File

Reads attribute-value data from an input file.

## CSV File Import

Import a data table from a CSV formatted file.

## Datasets

Load a dataset from an online repository.

## SQL Table

Reads data from an SQL database.

## Data Table

Displays attribute-value data in a spreadsheet.

## Data Info

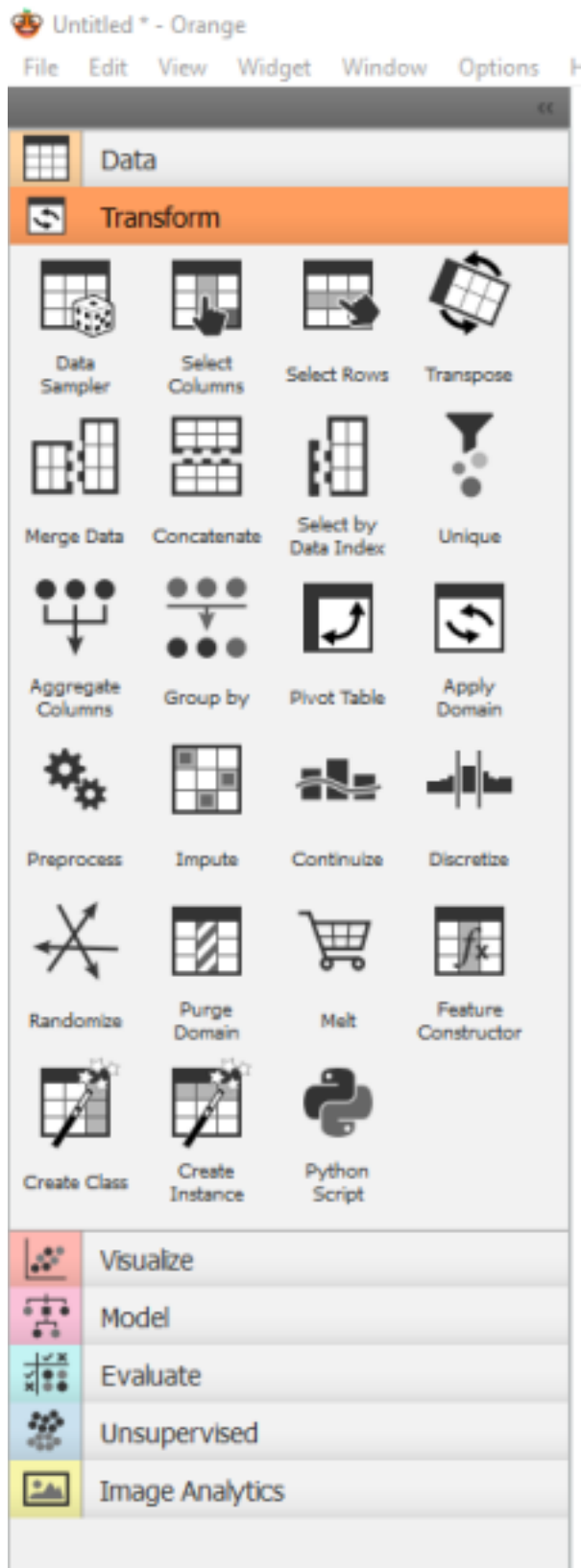
Displays information on a selected dataset.

## Rank

Ranking of attributes in classification or regression datasets.

## Feature Statistics

Show basic statistics for data features.



## Transform

### Data Sampler

Selects a subset of data instances from an input dataset.

## **Inputs**

- Data: input dataset

## **Outputs**

- Data Sample: sampled data instances
- Remaining Data: out-of-sample data

### 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

### Select Rows

Selects data instances based on conditions over data features.

## **Inputs**

- Data: input dataset

## **Outputs**

- Matching Data: instances that match the conditions ■  
Non-Matching Data: instances that do not match the conditions
- Data: data with an additional column showing whether a instance is selected

### Select by Data Index

Match instances by index from data subset.

## **Inputs**

- Data: reference data set
- Data Subset: subset to match

## **Outputs**

- Matching data: subset from reference data set that matches indices from subset data
- Unmatched data: subset from reference data set that does not match indices from subset data
- Annotated data: reference data set with an additional column defining matches

## Preprocess

Preprocesses data with selected methods.

## **Inputs**

- Data: input dataset

## **Outputs**

- Preprocessor: preprocessing method
- Preprocessed Data: data preprocessed with selected methods

## Feature Constructor

Add new features to your dataset.

## **Inputs**

- Data: input dataset

## **Outputs**

- Data: dataset with additional features

# MODEL

## Tree

A tree algorithm with forward pruning.

### Inputs

- Data: input dataset
- Preprocessor: preprocessing method(s)

### Outputs

- Learner: decision tree learning algorithm
- Model: trained model

## Linear Regression

A linear regression algorithm with optional L1 (LASSO), L2 (ridge) or L1L2 (elastic net) regularization.

### Inputs

- Data: input dataset
- Preprocessor: preprocessing method(s)

### Outputs

- Learner: linear regression learning algorithm
- Model: trained model
- Coefficients: linear regression coefficients

## Logistic Regression

The logistic regression classification algorithm with LASSO (L1) or ridge (L2) regularization.

### Inputs

- Data: input dataset
- Preprocessor: preprocessing method(s)

### Outputs

- Learner: logistic regression learning algorithm
- Model: trained model
- Coefficients: logistic regression coefficients

## Naive Bayes

A fast and simple probabilistic classifier based on Bayes' theorem with the assumption of feature independence.

### Inputs

- Data: input dataset
- Preprocessor: preprocessing method(s)

### Outputs

- Learner: naive bayes learning algorithm
- Model: trained model



## Evaluate

### Test and Score

Tests learning algorithms on data.

### Inputs

- Data: input dataset

- Test Data: separate data for testing ■
- Learner: learning algorithm(s)

## **Outputs**

- Evaluation Results: results of testing classification algorithms

## Predictions

Shows models' predictions on the data.

## **Inputs**

- Data: input dataset
- Predictors: predictors to be used on the data

## **Outputs**

- Predictions: data with added predictions
- Evaluation Results: results of testing classification algorithms

## Confusion Matrix

Shows proportions between the predicted and actual class.


## **Inputs**


- Evaluation results: results of testing classification algorithms


## **Outputs**


- Selected Data: data subset selected from confusion matrix
- Data: data with the additional information on whether a data instance was selected






Data



Transform



Visualize



Tree Viewer



Box Plot



Violin Plot



Distributions



Scatter Plot



Line Plot



Bar Plot



Sieve Diagram



Mosaic Display



FreeViz



Linear Projection



Radviz



Heat Map



Venn Diagram



Silhouette Plot



Pythagorean Tree



Pythagorean Forest



CN2 Rule Viewer


Nomogram


Model



Evaluate



Unsupervised



Image Analytics


Untitled \* - Orange


File Edit View Widget Window Options H


 Data


 Transform

 Visualize

 Model


  
Constant

  
ON2 Rule Induction


  
Calibrated Learner


  
kNN


  
Tree


  
Random Forest


  
Gradient Boosting


  
SVM


  
Linear Regression


  
Logistic Regression


  
Naive Bayes


  
AdaBoost


  
Curve Fit


  
Neural Network


  
Stochastic Gradient Descent


  
Stacking

  
Save Model

  
Load Model

 Evaluate

 Unsupervised

 Image Analytics



File

File:

...

Reload

URL:

Info

336 instance(s), 7 feature(s), 1 meta attribute(s)  
Classification; discrete class with 8 values.

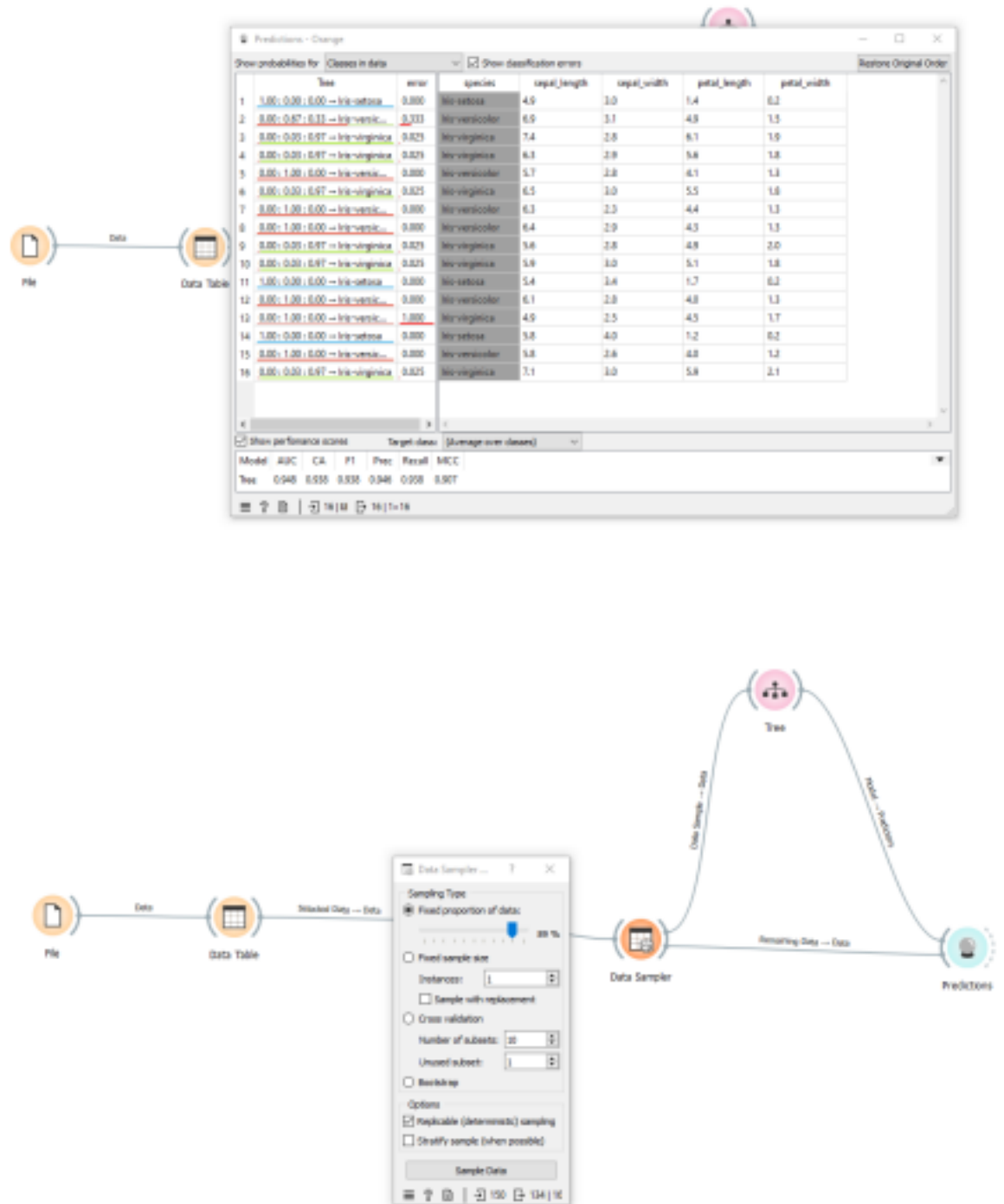
Columns (Double click to edit)

1	mcg	<span>C</span>	numeric	feature	
2	gvh	<span>C</span>	nominal	feature	
3	lip	<span>D</span>	numeric	feature	0.48, 1.00
4	chg	<span>D</span>	nominal	feature	0.50, 1.00
5	aac	<span>C</span>	numeric	feature	
6	alm1	<span>C</span>	numeric	feature	
7	alm2	<span>C</span>	numeric	feature	
8	localization site	<span>D</span>	nominal	target	cp, im, imL, imS, imU, om, omL, pp
9	name	<span>S</span>	string	meta	

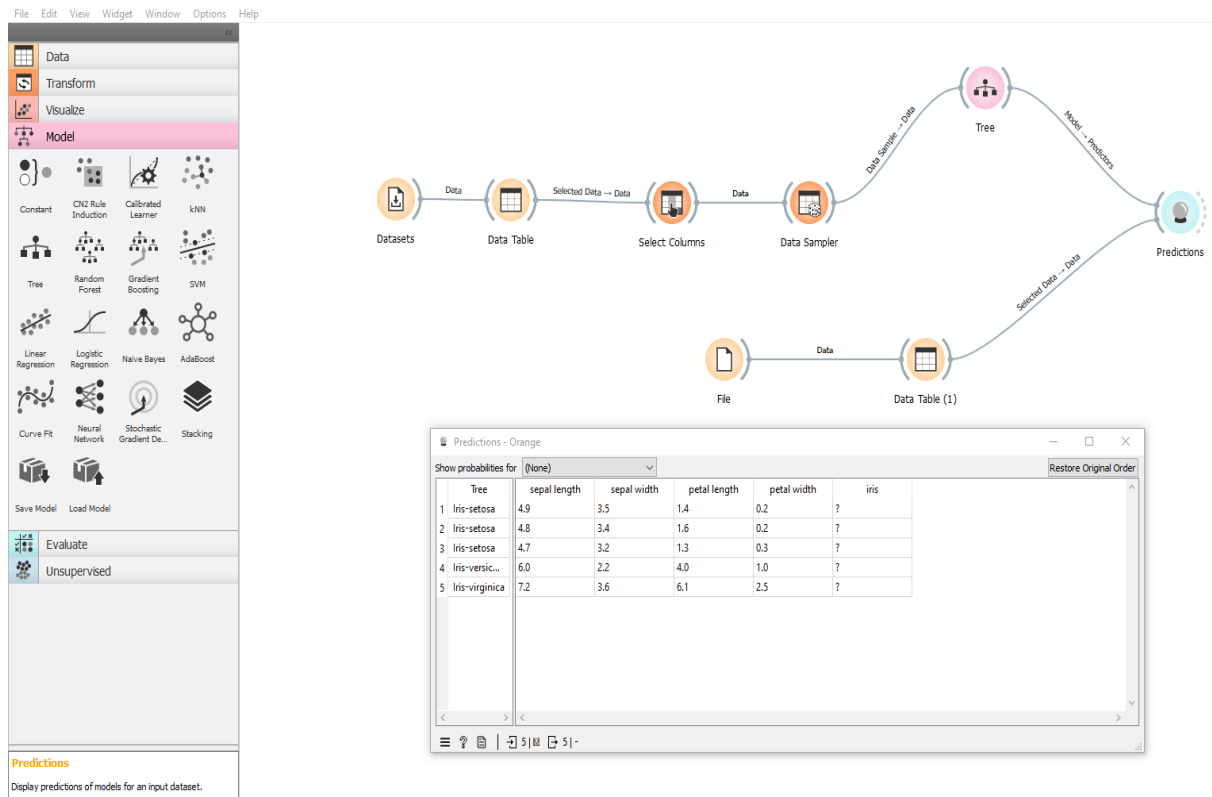
Browse documentation data sets

Report

Apply



Prediction of species in iris dataset



## Model Flow for training dataset:

- Datasets
- Data Table
- Select Columns
- Data Sampler
- Tree
- Predictions

## Model Flow for testing dataset:

- File
- Data Table
- Predictions