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

The widget receives a dataset and one or more predictors (predictive models, not learning algorithms - see the example below). It outputs the data and the predictions.

Info ①		6 Logistic Regression	diameter narrowing	age	gender	chest pain	rest SBF
Data: 213 instances. Predictors: 1 Task: Classification Restore Original Order	1	0.01 : 0.99 → 1	1	56	female	asymptomatic	134
	2	0.01 : 0.99 → 1	1	59	male	asymptomatic	164
	3	0.04 : 0.96 → 1	1	54	male	asymptomatic	110
	4	0.99 : 0.01 → 0	0	54	female	non-anginal	108
✓ Predicted class ✓ Predicted probabilities for: 0 1	5	0.96 : 0.04 → 0	0	59	male	non-anginal	150
	6	0.88 : 0.12 → 0	0	45	male	asymptomatic	115
	7	0.06 : 0.94 → 1	1	61	male	asymptomatic	140
	8	0.99 : 0.01 → 0	0	46	female	atypical ang	105
	9	0.92 : 0.08 → 0	0	41	male	non-anginal	130
	10	0.98 : 0.02 → 0	0	56	male	atypical ang	120
	11	0.03 : 0.97 → 1	0	64	male	asymptomatic	128
	12	<u>0.11 : 0.89 → 1</u>	1	53	male	asymptomatic	140
✓ Draw distribution bars Data View ✓ Show full dataset	13	0.38 : 0.62 → 1	1	49	male	non-anginal	118
	14	<u>0.09</u> : 0.91 → 1	1	77	male	asymptomatic	125
	15	0.94 : 0.06 → 0	0	44	female	non-anginal	118
	16	0.01 : 0.99 → 1	1	54	male	asymptomatic	124
Output 4	17	0.97 : 0.03 → 0	0	44	male	non-anginal	140
✓ Original data✓ Predictions✓ Probabilities	18	0.44 : 0.56 → 1	1	64	male	non-anginal	125
	19	<u>0.07 : 0.93 → 1</u>	1	40	male	asymptomatic	110
	20	0.97 : 0.03 → 0	0	60	female	non-anginal	120
	21	0.05 ⋅ 0.95 → 1	1	43	male	asymptomatic	132

- 1. Information on the input, namely the number of instances to predict, the number of predictors and the task (classification or regression). If you have sorted the data table by attribute and you wish to see the original view, press *Restore Original Order*.
- 2. You can select the options for classification. If *Predicted class* is ticked, the view provides information on predicted class. If *Predicted probabilities for* is ticked, the view provides information on probabilities predicted by the classifier(s). You can also select the predicted class displayed in the view. The option *Draw distribution bars* provides a visualization of probabilities.
- 3. By ticking the Show full dataset, you can view the entire data table (otherwise only class variable will be shown).
- 4. Select the desired output.
- 5. Predictions.

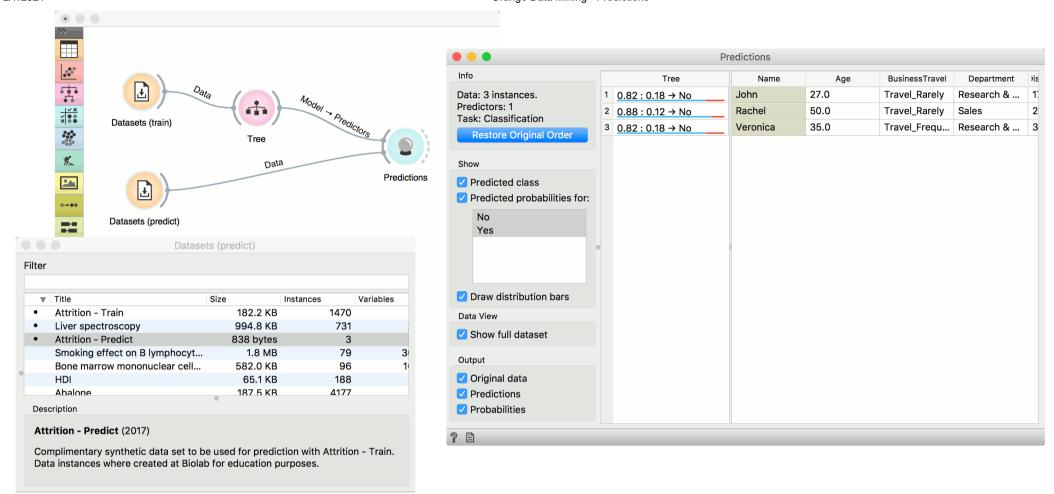
The widget show the probabilities and final decisions of predictive models. The output of the widget is another dataset, where predictions are appended as new meta attributes. You can select which features you wish to output (original data, predictions, probabilities). The result can be observed in a Data Table. If the predicted data includes true class values, the result of prediction can also be observed in a Confusion Matrix.

Examples

In the first example, we will use *Attrition - Train* data from the <u>Datasets</u> widget. This is a data on attrition of employees. In other words, we wish to know whether a certain employee will resign from the job or not. We will construct a predictive model with the <u>Tree</u> widget and observe probabilities in **Predictions**.

For predictions we need both the training data, which we have loaded in the first **Datasets** widget and the data to predict, which we will load in another **Datasets** widget. We will use *Attrition - Predict* data this time. Connect the second data set to **Predictions**. Now we can see predictions for the three data instances from the second data set.

The Tree model predicts none of the employees will leave the company. You can try other model and see if predictions change. Or test the predictive scores first in the Test & Score widget.



In the second example, we will see how to properly use Preprocess with Predictions or Test & Score.

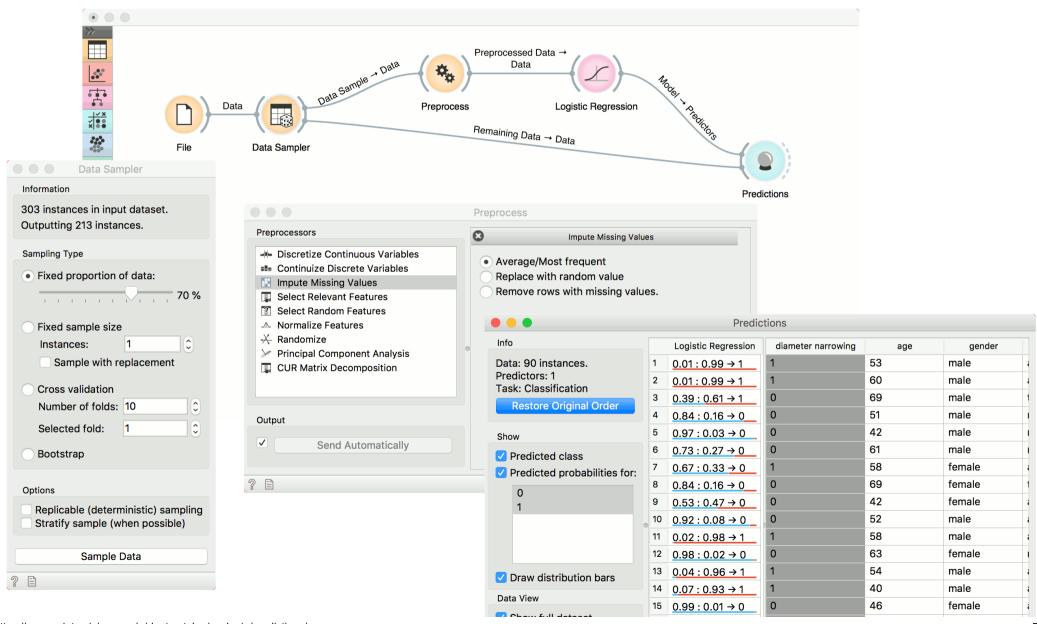
This time we are using the *heart disease.tab* data from the File widget. You can access the data through the dropdown menu. This is a dataset with 303 patients that came to the doctor suffering from a chest pain. After the tests were done, some patients were found to have diameter narrowing and others did not (this is our class variable).

The heart disease data have some missing values and we wish to account for that. First, we will split the data set into train and test data with Data Sampler.

Then we will send the *Data Sample* into Preprocess. We will use *Impute Missing Values*, but you can try any combination of preprocessors on your data. We will send preprocessed data to Logistic Regressionand the constructed model to **Predictions**.

Finally, **Predictions** also needs the data to predict on. We will use the output of **Data Sampler** for prediction, but this time not the **Data Sample**, but the **Remaining Data**, this is the data that wasn't used for training the model.

Notice how we send the remaining data directly to **Predictions** without applying any preprocessing. This is because Orange handles preprocessing on new data internally to prevent any errors in the model construction. The exact same preprocessor that was used on the training data will be used for predictions. The same process applies to **Test & Score**.



Orange Data Mining - Predictions

✓ Snow Tull dataset	16 0.97 : 0.03 →	00	38	male
Output	17 0.09 : 0.91 →	11	58	male
Original dataPredictionsProbabilities	18 0.01 : 0.99 →	11	60	male
	19 0.75 : 0.25 →	00	59	male
	20 0.05 : 0.95 →	11	50	male ;
	21 ∩ ∩1 · ∩ qq →	1	63	male ;
? 🗎				