

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

Data: 213 instances.
Predictors: 1
Task: Classification

Restore Original Order

Show

☒ Predicted class
☒ Predicted probabilities for:

0
1

☒ Draw distribution bars

Data View

☒ Show full dataset

Output

☒ Original data
☒ Predictions
☒ Probabilities

5 Logistic Regression

1	<u>0.01 : 0.99</u> → 1	1	56	female	asymptomatic	134
2	<u>0.01 : 0.99</u> → 1	1	59	male	asymptomatic	164
3	<u>0.04 : 0.96</u> → 1	1	54	male	asymptomatic	110
4	<u>0.99 : 0.01</u> → 0	0	54	female	non-anginal	108
5	<u>0.96 : 0.04</u> → 0	0	59	male	non-anginal	150
6	<u>0.88 : 0.12</u> → 0	0	45	male	asymptomatic	115
7	<u>0.06 : 0.94</u> → 1	1	61	male	asymptomatic	140
8	<u>0.99 : 0.01</u> → 0	0	46	female	atypical ang	105
9	<u>0.92 : 0.08</u> → 0	0	41	male	non-anginal	130
10	<u>0.98 : 0.02</u> → 0	0	56	male	atypical ang	120
11	<u>0.03 : 0.97</u> → 1	0	64	male	asymptomatic	128
12	<u>0.11 : 0.89</u> → 1	1	53	male	asymptomatic	140
13	<u>0.38 : 0.62</u> → 1	1	49	male	non-anginal	118
14	<u>0.09 : 0.91</u> → 1	1	77	male	asymptomatic	125
15	<u>0.94 : 0.06</u> → 0	0	44	female	non-anginal	118
16	<u>0.01 : 0.99</u> → 1	1	54	male	asymptomatic	124
17	<u>0.97 : 0.03</u> → 0	0	44	male	non-anginal	140
18	<u>0.44 : 0.56</u> → 1	1	64	male	non-anginal	125
19	<u>0.07 : 0.93</u> → 1	1	40	male	asymptomatic	110
20	<u>0.97 : 0.03</u> → 0	0	60	female	non-anginal	120
21	<u>0.05 : 0.95</u> → 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 **Datasets** 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 **Tree** 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.

The screenshot displays the Orange Data Mining software interface. On the left, a workflow canvas shows two 'Datasets' widgets (one for training, one for prediction) connected to a 'Tree' widget, which is then connected to a 'Predictions' widget. Below the workflow, the 'Datasets (predict)' widget is open, showing a list of datasets with columns for Title, Size, Instances, and Variables. The 'Attrition - Predict' dataset is selected, and its description is shown below the list.

The 'Predictions' widget is open on the right, displaying the results of the classification task. It shows the predicted class and predicted probabilities for each instance. The 'Info' tab is active, showing the task as 'Classification' and the number of instances and predictors. The 'Show' tab is also active, showing the predicted class and predicted probabilities for each instance. The 'Data View' tab is active, showing the full dataset with columns for Name, Age, BusinessTravel, Department, and Disposition.

	Tree	Name	Age	BusinessTravel	Department	Disposition
1	0.82 : 0.18 → No	John	27.0	Travel_Rarely	Research & ...	1
2	0.88 : 0.12 → No	Rachel	50.0	Travel_Rarely	Sales	2
3	0.82 : 0.18 → No	Veronica	35.0	Travel_Frequ...	Research & ...	3

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 Regression** and 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**.

The diagram shows a workflow starting with a **File** widget connected to a **Data Sampler** widget. The **Data Sampler** widget has two outputs: **Data Sample → Data** leading to a **Preprocess** widget, and **Remaining Data → Data** leading directly to the **Predictions** widget. The **Preprocess** widget is connected to a **Logistic Regression** widget, which then connects to the **Predictions** widget. The **Predictions** widget also receives input from the **Logistic Regression** widget.

The **Data Sampler** widget settings are as follows:

- Information:** 303 instances in input dataset. Outputting 213 instances.
- Sampling Type:**
 - ☒ Fixed proportion of data: 70 %
 - ☐ Fixed sample size: Instances: 1
 - ☐ Cross validation: Number of folds: 10, Selected fold: 1
 - ☐ Bootstrap
- Options:**
 - ☐ Replicable (deterministic) sampling
 - ☐ Stratify sample (when possible)
- Sample Data** button

The **Preprocess** widget settings are as follows:

- Preprocessors:**
 - ☒ Impute Missing Values
 - ☐ Discretize Continuous Variables
 - ☐ Continuize Discrete Variables
 - ☐ Select Relevant Features
 - ☐ Select Random Features
 - ☐ Normalize Features
 - ☐ Randomize
 - ☐ Principal Component Analysis
 - ☐ CUR Matrix Decomposition
- Output:** ☒ Send Automatically

The **Impute Missing Values** sub-widget settings are as follows:

- ☒ Average/Most frequent
- ☐ Replace with random value
- ☐ Remove rows with missing values.

The **Predictions** widget settings are as follows:

- Info:** Data: 90 instances. Predictors: 1. Task: Classification. [Restore Original Order](#)
- Show:**
 - ☒ Predicted class
 - ☒ Predicted probabilities for: 0, 1
 - ☒ Draw distribution bars
- Data View:** ☒ Show full dataset

	Logistic Regression	diameter narrowing	age	gender
1	0.01 : 0.99 → 1	1	53	male
2	0.01 : 0.99 → 1	1	60	male
3	0.39 : 0.61 → 1	0	69	male
4	0.84 : 0.16 → 0	0	51	male
5	0.97 : 0.03 → 0	0	42	male
6	0.73 : 0.27 → 0	0	61	male
7	0.67 : 0.33 → 0	1	58	female
8	0.84 : 0.16 → 0	0	69	female
9	0.53 : 0.47 → 0	0	42	female
10	0.92 : 0.08 → 0	0	52	male
11	0.02 : 0.98 → 1	1	58	male
12	0.98 : 0.02 → 0	0	63	female
13	0.04 : 0.96 → 1	1	54	male
14	0.07 : 0.93 → 1	1	40	male
15	0.99 : 0.01 → 0	0	46	female

<input checked="" type="checkbox"/> Show full dataset	16	<u>0.97 : 0.03</u> → 0	0	38	male
Output	17	<u>0.09 : 0.91</u> → 1	1	58	male
	18	<u>0.01 : 0.99</u> → 1	1	60	male
	19	<u>0.75 : 0.25</u> → 0	0	59	male
	20	<u>0.05 : 0.95</u> → 1	1	50	male
	21	<u>0.01 : 0.99</u> → 1	1	63	male