Lesson 5: Classification

In one of the previous lessons, we explored the heart disease data. We wanted to predict which persons have clogged arteries — but we did not make any predictions. Let's try it now.



This won't do: the widget Predictions shows the data, but no makes no predictions. It can't. For this, it needs a model. Like this.



The data is fed into the Tree widget, which uses it to infer a predictive model. The Predictions widget now gets the data from the File widget and also a predictive model from the Tree widget. This is something new: in our past workflows, widgets passed only data to each other, but here we have a channel that carries a model.

Info		Tree	diameter narrowing	age	gender	chest p
Data: 303 instances. Predictors: 1 Task: Classification Restore Original Order	1	1.00 : 0.00 → 0	0	63.000	male	typical an
	2	<u>0.00 : 1.00 → 1</u>	1	67.000	male	asymptor
	3	0.04 : 0.96 → 1	1	67.000	male	asymptor
	4	0.96 : 0.04 → 0	0	37.000	male	non-angi
Show	5	0.96 : 0.04 → 0	0	41.000	female	atypical a
 Predicted class Predicted probabilities for: 0 1 Draw distribution bars Data View Show full data set 	6	0.96 : 0.04 → 0	0	56.000	male	atypical a
	7	0.25 : 0.75 → 1	1	62.000	female	asymptor
	8	0.96 : 0.04 → 0	0	57.000	female	asymptor
	9	0.04 : 0.96 → 1	1	63.000	male	asymptor
	10	<u>0.00 : 1.00 → 1</u>	1	53.000	male	asymptor
	• 11	1.00 : 0.00 → 0	0	57.000	male	asymptor
	12	0.96 : 0.04 → 0	0	56.000	female	atypical a
	13	<u>0.00 : 1.00 → 1</u>	1	56.000	male	non-angi
	14	<u>0.25 : 0.75 → 1</u>	0	44.000	male	atypical a
	15	1.00 : 0.00 → 0	0	52.000	male	non-angi
	16	0.96 : 0.04 → 0	0	57.000	male	non-angi
Output	17	<u>0.25 : 0.75 → 1</u>	1	48.000	male	atypical a
 Original data Predictions Probabilities 	18	0.96 : 0.04 → 0	0	54.000	male	asymptor
	19	0.96 : 0.04 → 0	0	48.000	female	non-angi
	20	0.96 : 0.04 → 0	0	49.000	male	atypical a
	21	<u>1.00 : 0.00 → 0</u>	0	64.000	male	typical an
Report	22	1.00 : 0.00 → 0	0	58.000	female	typical ar

The Predictions widget uses the model to make predictions about the data and shows them in the table.

How correct are these predictions? Do we have a good model? How can we tell?

But (and even before answering these critical questions), what is a tree? How does it look like? How does Orange create one? Is this algorithm something we should use? So many questions to answer today!