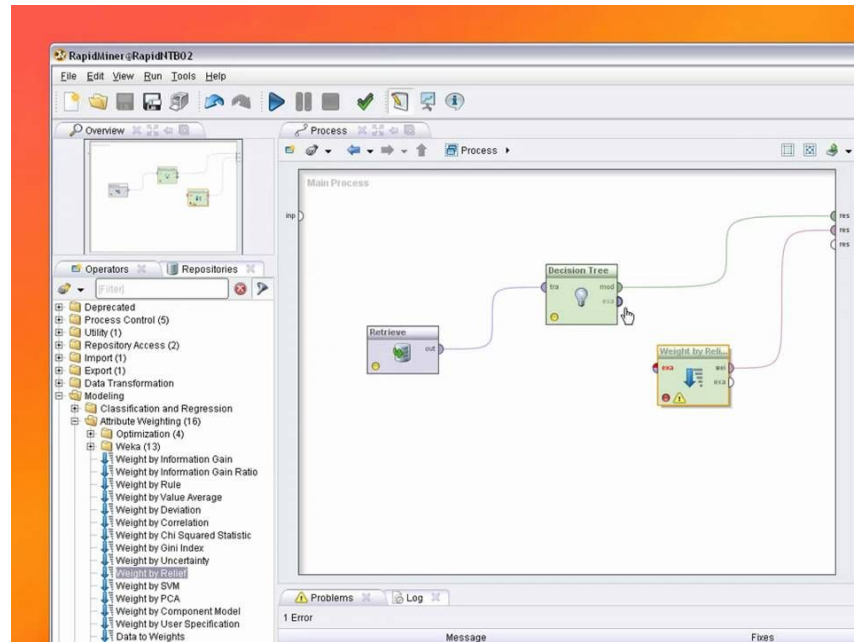


Adatbányászati módszerek

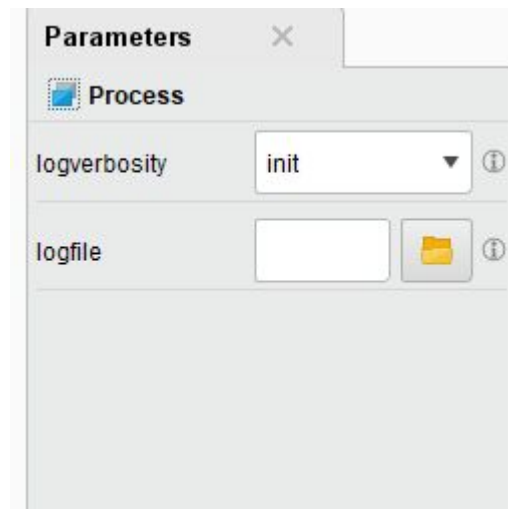
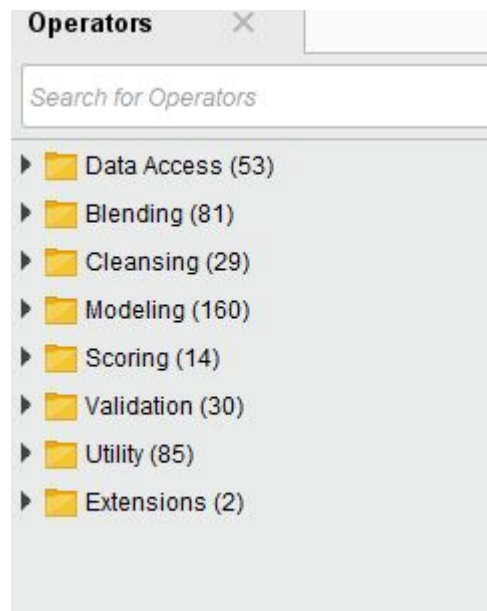
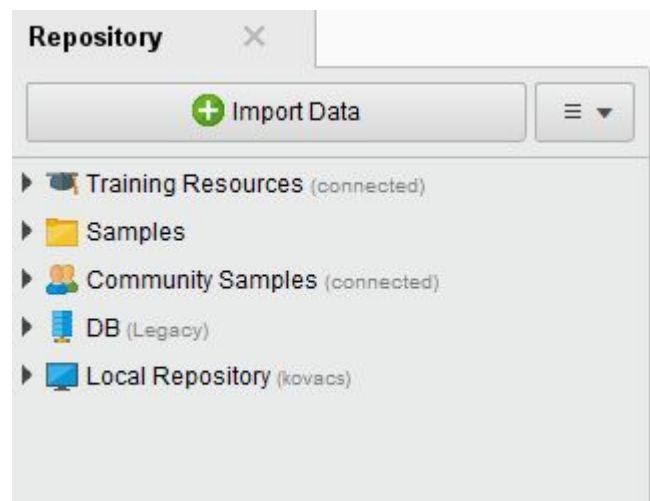
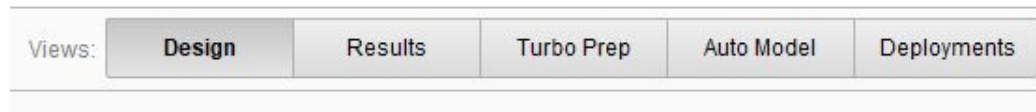
Kovács László, ME

rapidMiner adatelemző rendszer

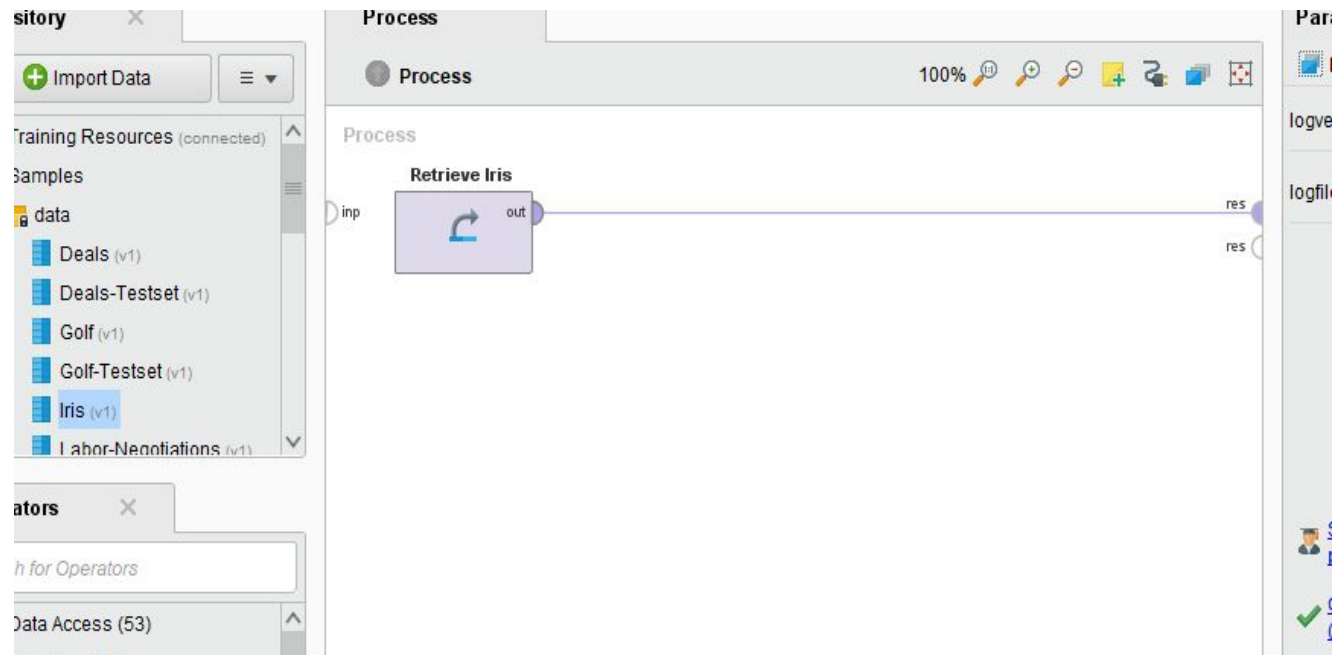
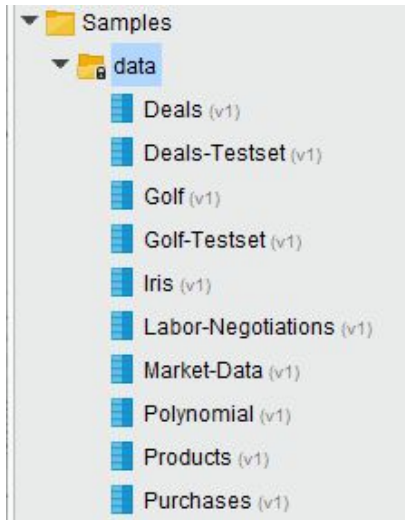
- Ingyenes adatelemző keretrendszer
- Kódírás nélküli programozás
- Drag and drop alapú fejlesztés
- Adatfolyam összeállítás
- Műveletek paraméterezése
- Heterogén adatforrások támogatása
- Adatok transzformációja, szűrése
- Adatelemzések támogatása
- Eredmények vizualizációja
- Adatok export / import



rapidMiner: funkció modulok





rapidMiner: adatforrások



rapidMiner: adatforrások elemzése

Result History

ExampleSet (Retrieve Iris) X

Open in  Turbo Prep  Auto Model Filter (150 / 150 examples):

Data

Statistics

Visualizations

Simple Charts

Advanced Charts

Row No.	id	label	a1	a2
1	id_1	Iris-setosa	5.100	3.500
2	id_2	Iris-setosa	4.900	3
3	id_3	Iris-setosa	4.700	3.200
4	id_4	Iris-setosa	4.600	3.100
5	id_5	Iris-setosa	5.000	3.600
6	id_6	Iris-setosa	5.400	3.900
7	id_7	Iris-setosa	4.500	3.000
8	id_8	Iris-setosa	4.800	3.400
9	id_9	Iris-setosa	4.900	3.500
10	id_10	Iris-setosa	4.700	3.200
11	id_11	Iris-setosa	4.600	3.100
12	id_12	Iris-setosa	5.000	3.600
13	id_13	Iris-setosa	5.400	3.900
14	id_14	Iris-setosa	4.300	3

ExampleSet (Retrieve Iris) X

Name | Type | Missing | Statistics

Filter (6 / 6 attributes): Search for Attributes

Name	Type	Missing	Least	Most	Values
id	Nominal	0	id_99 (1)	id_1 (1)	id_1 (1), id_10 (1), ...[148 more]
label	Nominal	0	Iris-virginica (50)	Iris-setosa (50)	Iris-setosa (50), Iris-versicolor (50)
a1	Real	0	Min 4.300	Max 7.900	Average 5.843
a2	Real	0	Min 2	Max 4.400	Average 3.054

rapidMiner: adatforrások elemzése

Plot

Plot 1

Plot type

Scatter 3D

X-Axis column

label

Value columns

a3

Y Axis

-

Color

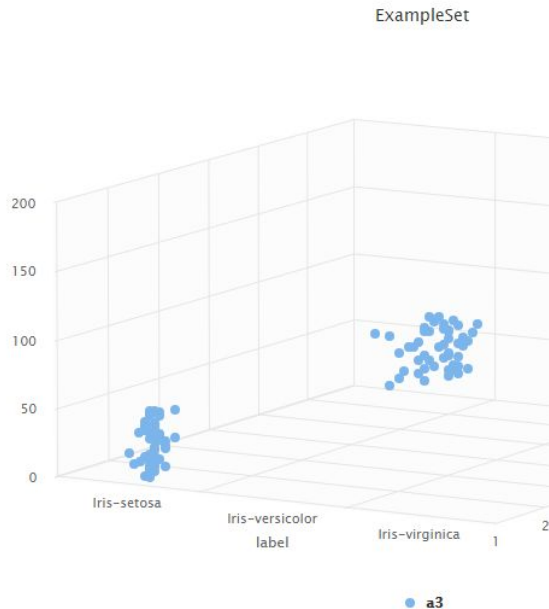
-

Plot style

[Add new plot](#)

General

X-Axis



Charts

Line

Step Line

Spline

Area

Step Area

Spline Area

Scatter

Scatter Matrix

Scatter 3D

Bar (Column)

Bar (Horizontal)

Streamgraph

Histogram

Boxplot

Bell Curve

Heatmap

Treemap

Sunburst

Pie

Funnel

Pyramid

Packed Bubble

Parliament

Pareto

Range (Column)

Range (Error Bar)

Range (Line)

Range (Step)

Range (Spline)

Vector

Sankey

Chord

Deviation

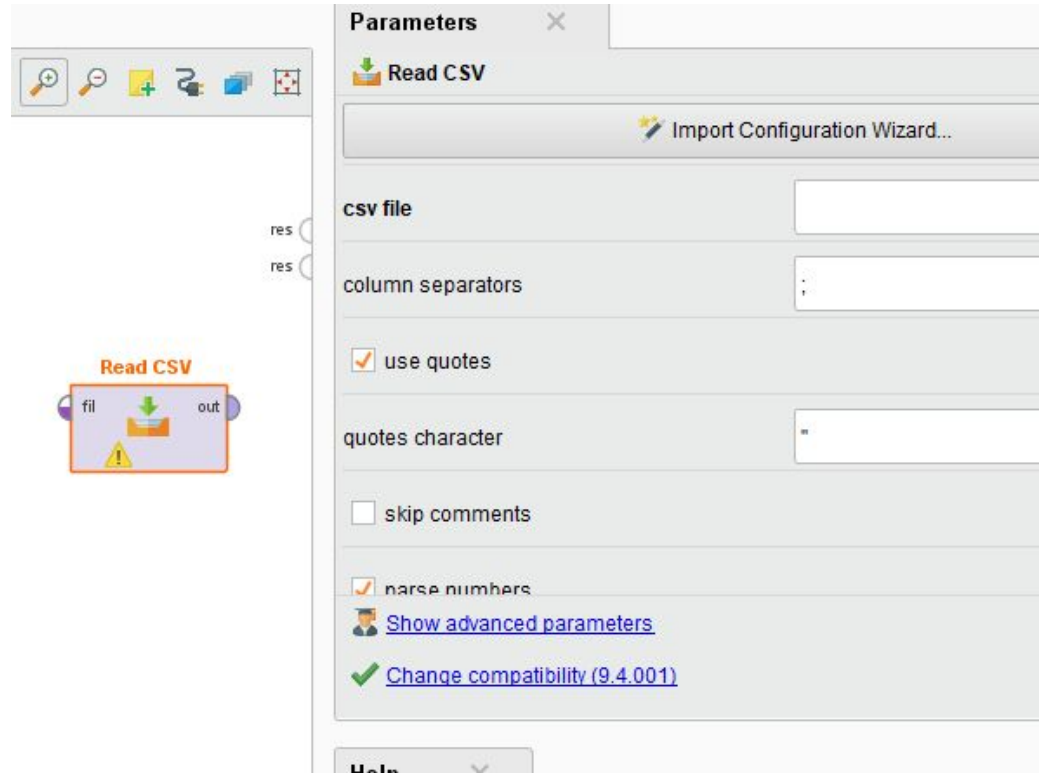
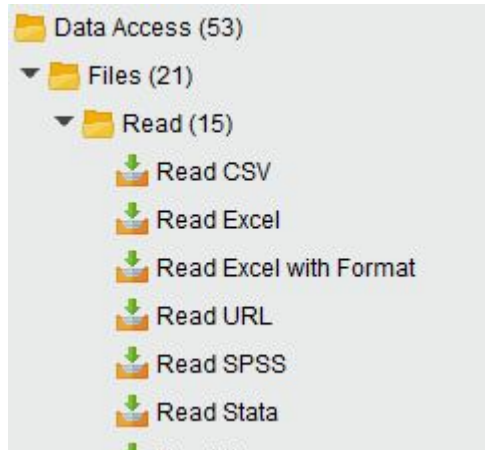
Parallel Coordinates

Andrews Curves

Wordcloud

Maps

rapidMiner: külső adatforrások olvasása



rapidMiner: adatszerkezet előfeldolgozás

The screenshot displays the RapidMiner software interface, which is used for data preprocessing. The main window is titled "Process" and shows a workflow diagram with two nodes: "Retrieve Iris" and "Rename". The "Rename" node is highlighted with a red border and a yellow warning icon, indicating a configuration issue. The "Rename" node's configuration panel is open on the right, showing fields for "old name", "new name", and "rename additional attributes".

On the left side, the "Import Data" panel lists several datasets: "Deals-Testset (v1)", "Golf (v1)", "Golf-Testset (v1)", and "Iris (v1)". Below this, the "Operators" panel is visible, showing a search bar and a list of operators categorized by function. The "Names & Roles" category is expanded, showing the "Rename" operator selected.

Below the "Operators" panel, a context menu is open, displaying a list of operators under the "Types (16)" category. The operators listed are:

- Numerical to Binominal
- Numerical to Polynominal
- Numerical to Real
- Numerical to Date
- Real to Integer
- Nominal to Binominal

On the right side, another context menu is open, displaying a list of operators under the "Selection (7)" category. The operators listed are:

- Select Attributes
- Select by Weights
- Select by Random
- Remove Attribute Range
- Remove Useless Attribute
- Remove Correlated Attribute

Below the "Selection (7)" category, a list of operators is visible, including:

- Generate Attributes
- Generate ID
- Generate Empty Attribute
- Generate Copy
- Generate Concatenation
- Generate Aggregation
- Generate Absolutes

rapidMiner: adattisztítási előfeldolgozás

The screenshot displays the RapidMiner software interface, specifically the workflow editor and the configuration panel for the 'Discretize by Binning' process.

Workflow Editor:

- The **Process** view shows a sequence of two processes: **Retrieve Iris** (purple box) and **Discretize** (pink box).
- The **Retrieve Iris** process has an **inp** port and an **out** port.
- The **Discretize** process has three input ports labeled **exa**, **ori**, and **pre**.

Left Panel (Process List):

- Cleansing (29)**
 - Normalization (3)
 - Binning (5)
 - Missing (8)
 - Duplicates (1)
 - Outliers (4)
 - Dimensionality Reduction (6)
 - Statistics
 - Quality Measures
- Discretize by Binning** (highlighted)
- Discretize by Frequency**
- Discretize by User Specification**
- Discretize by Entropy**
- Missing (8)**

Right Panel (Configuration):

- attribute filter type** (dropdown menu)
- ☐ invert selection
- ☐ include special attributes
- number of bins** (input field with value 6)
- [Show advanced parameters](#)
- [Change compatibility \(9.4.001\)](#)

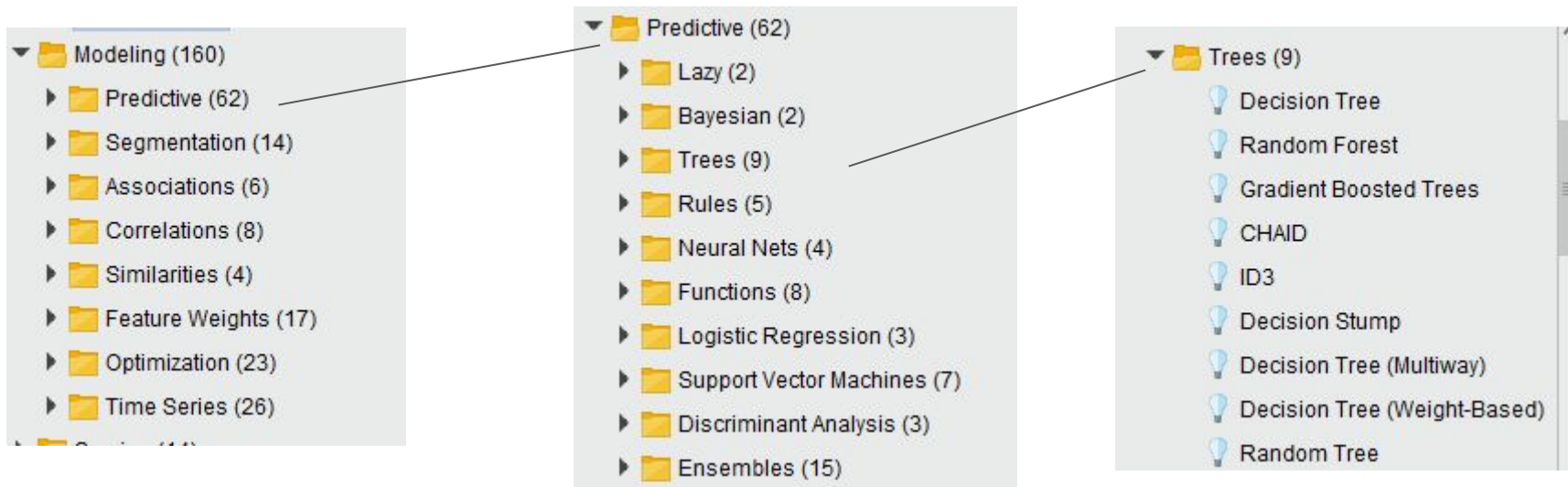
Bottom Panel (Help):

- Discretize by Binning** (title)
- RapidMiner Studio Core

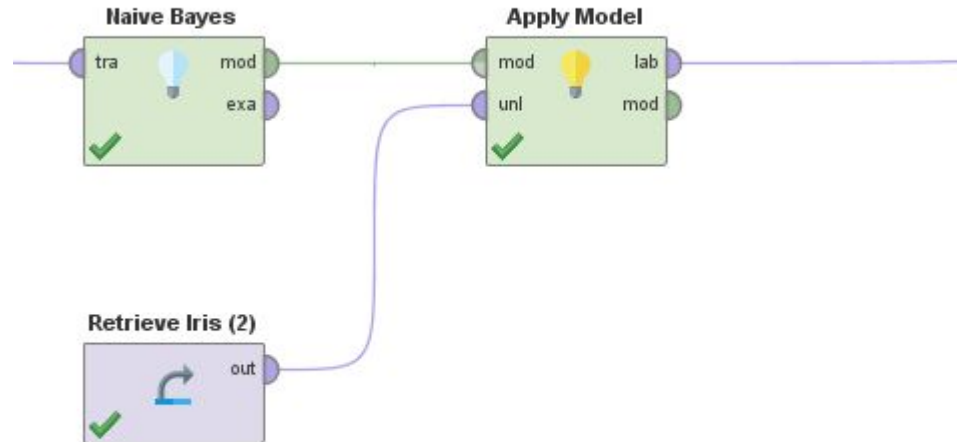
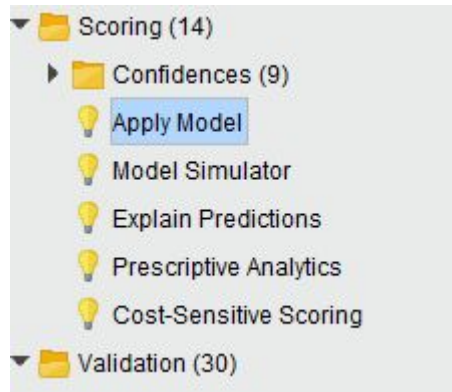
Bottom Center Panel (Missing (8) list):

- ☐ Replace Missing Values
- ☐ Impute Missing Values
- ☐ Declare Missing Value
- ☐ Replace Infinite Values
- ☐ Remove Unused Values
- ☐ Fill Data Gaps
- ☐ Replace All Missings
- ☐ Handle Unknown Values

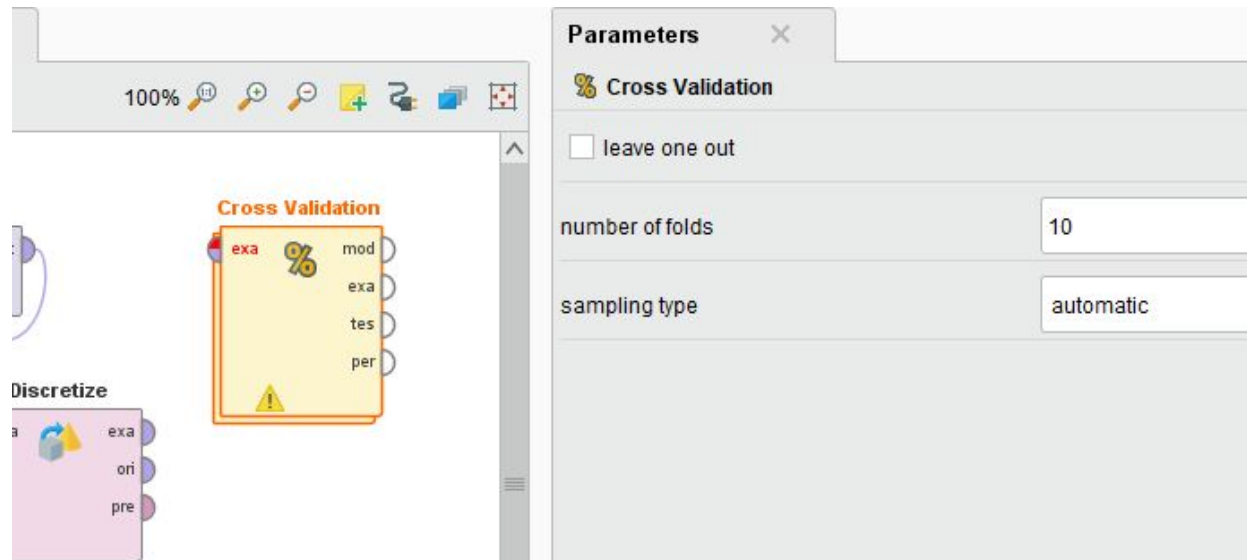
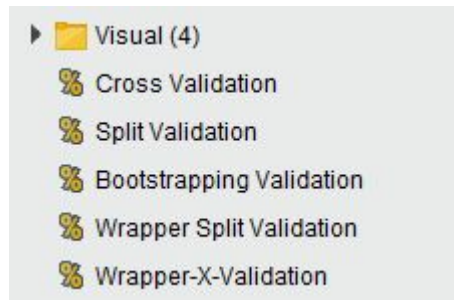
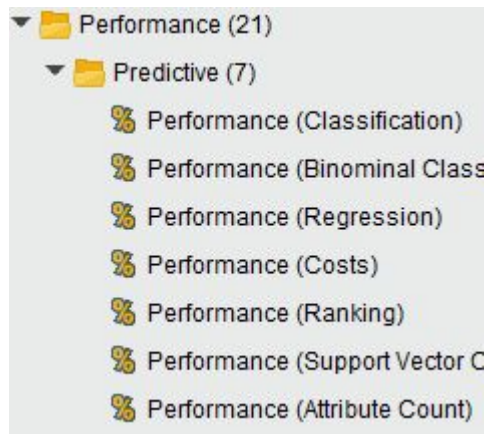
rapidMiner: elemzési módszerek



rapidMiner: modell alkalmazása



rapidMiner: módszerek értékelése



rapidMiner: Iris elemzése NB osztályozóval

Plot

Plot 1

Plot type

Scatter 3D

X-Axis column

a1

Value column

a2

Y Axis

-

Color

label

Plot style

Add new plot

General

X-Axis

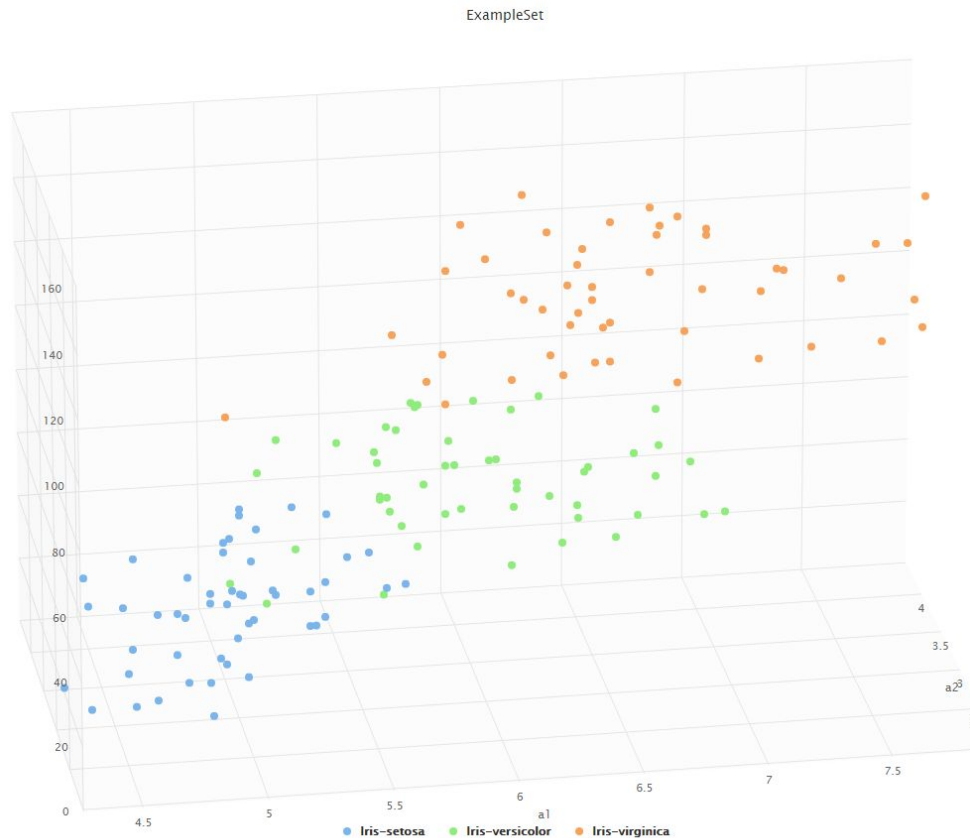
Y-Axis

Z-Axis

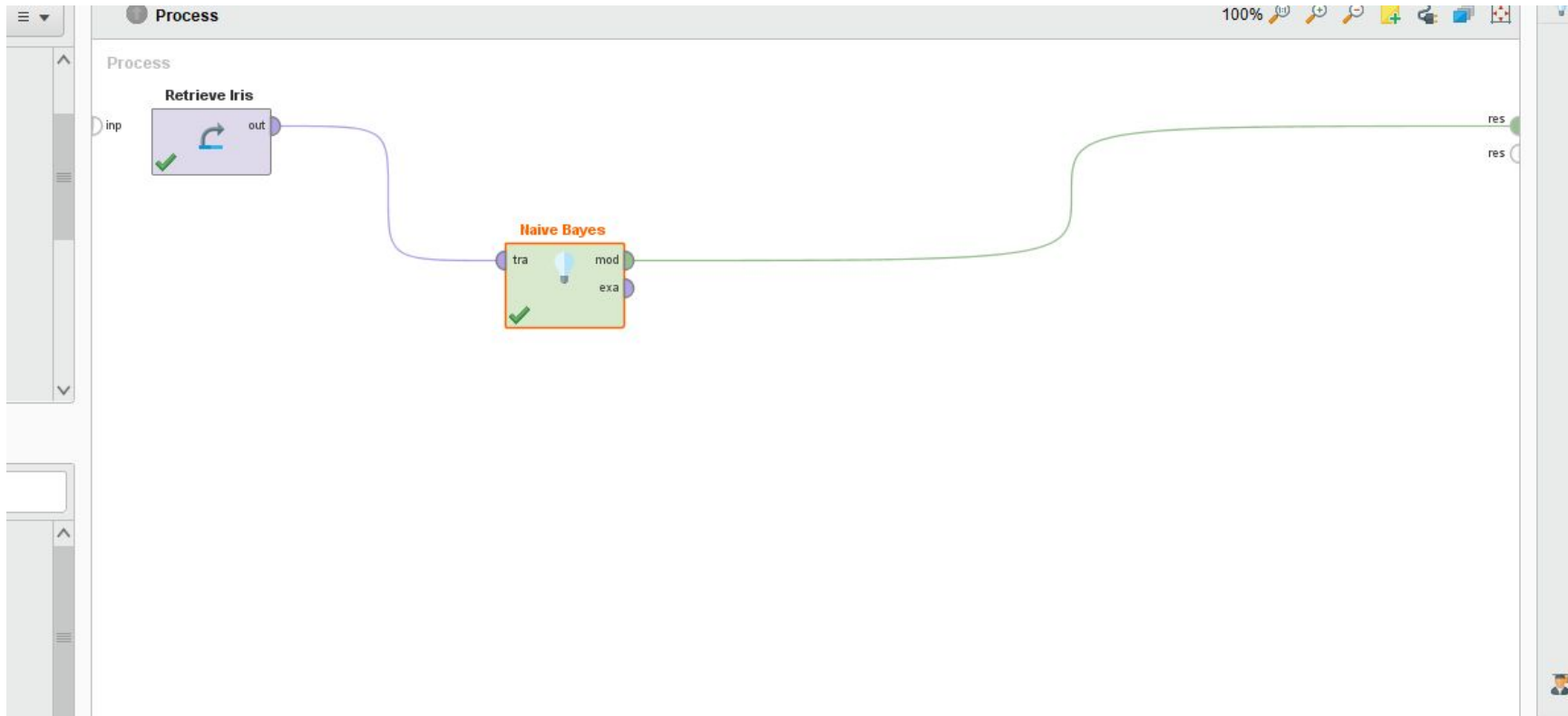
Title

Legend

Tooltip



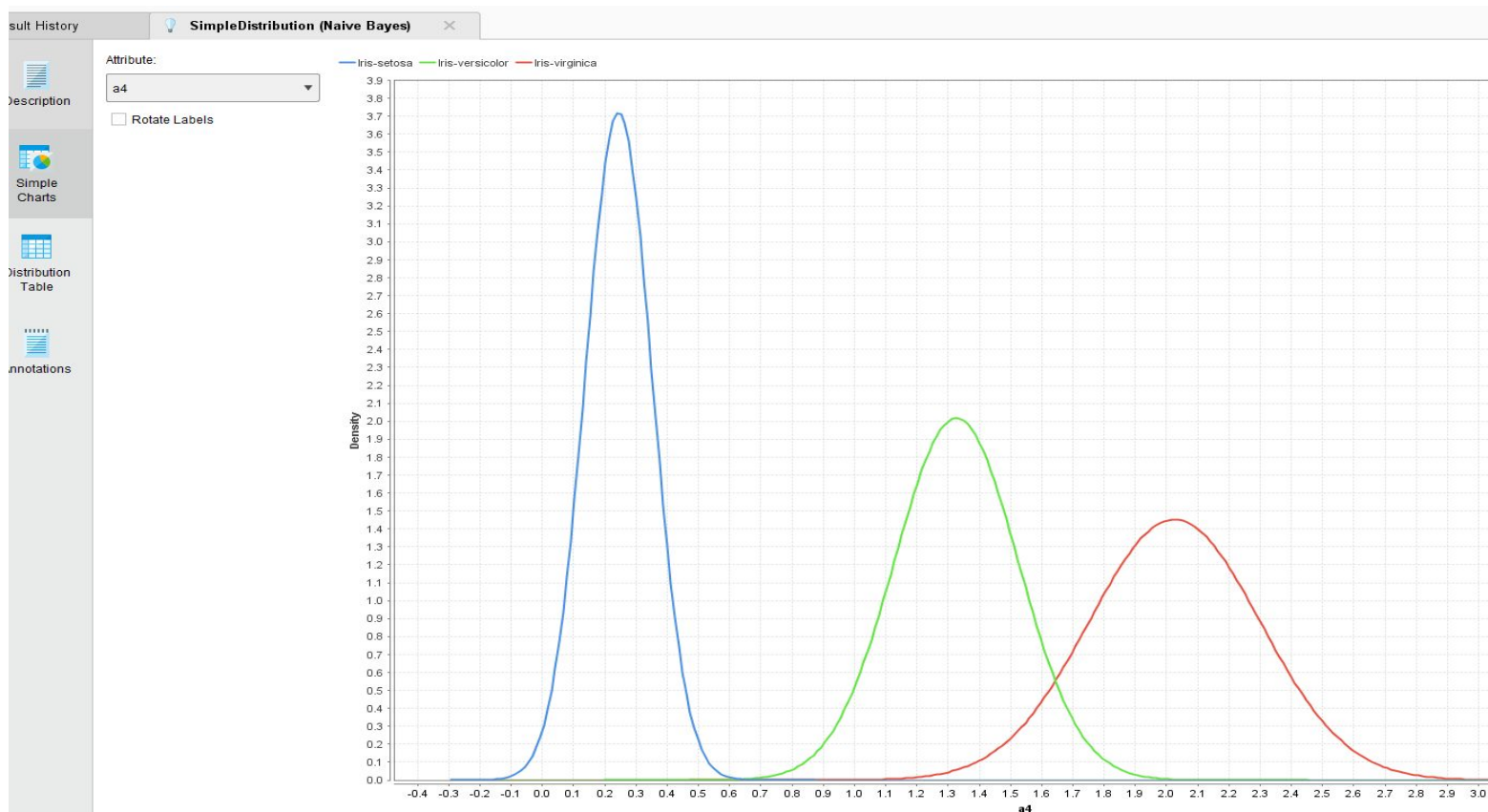
rapidMiner: Iris elemzése NB osztályozóval



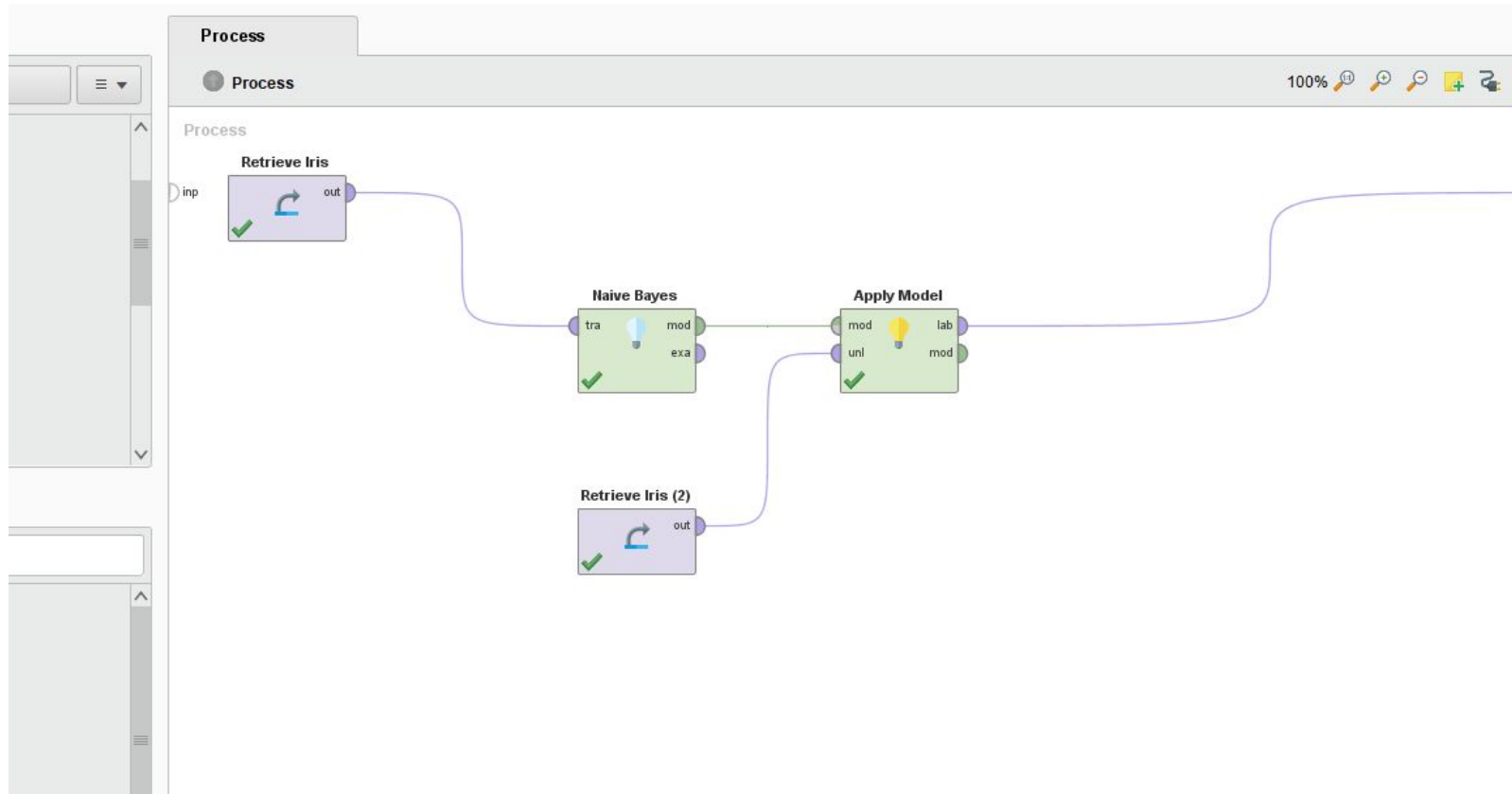
rapidMiner: Iris elemzése NB osztályozóval

SimpleDistribution (Naive Bayes) ×				
Attribute	Parameter	Iris-setosa	Iris-versicolor	Iris-virginica
a1	mean	5.006	5.936	6.588
a1	standard deviation	0.352	0.516	0.636
a2	mean	3.418	2.770	2.974
a2	standard deviation	0.381	0.314	0.322
a3	mean	1.464	4.260	5.552
a3	standard deviation	0.174	0.470	0.552
a4	mean	0.244	1.326	2.026
a4	standard deviation	0.107	0.198	0.275



rapidMiner: Iris elemzése NB osztályozóval



rapidMiner: Iris elemzése NB osztályozóval

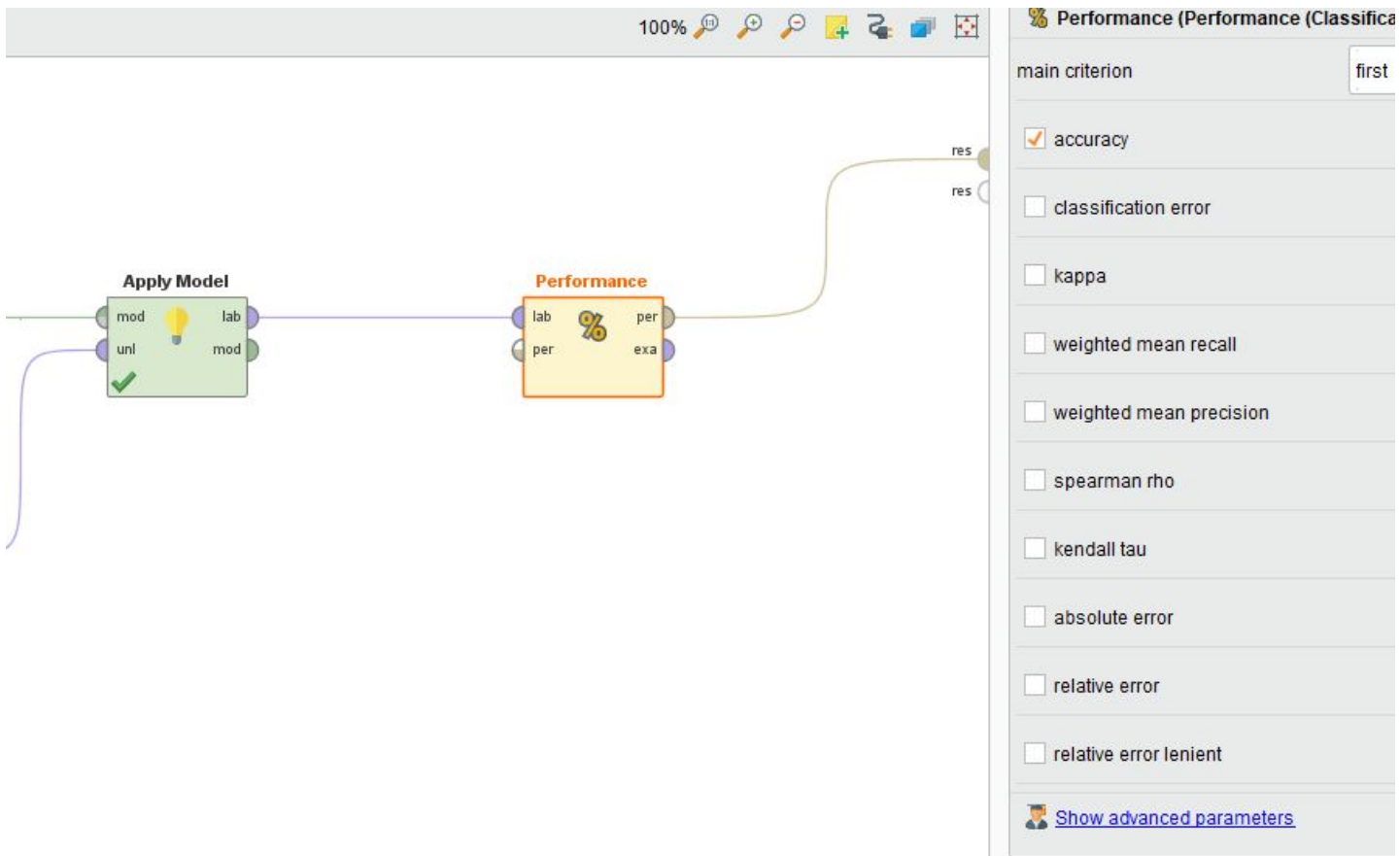


rapidMiner: Iris elemzése NB osztályozóval

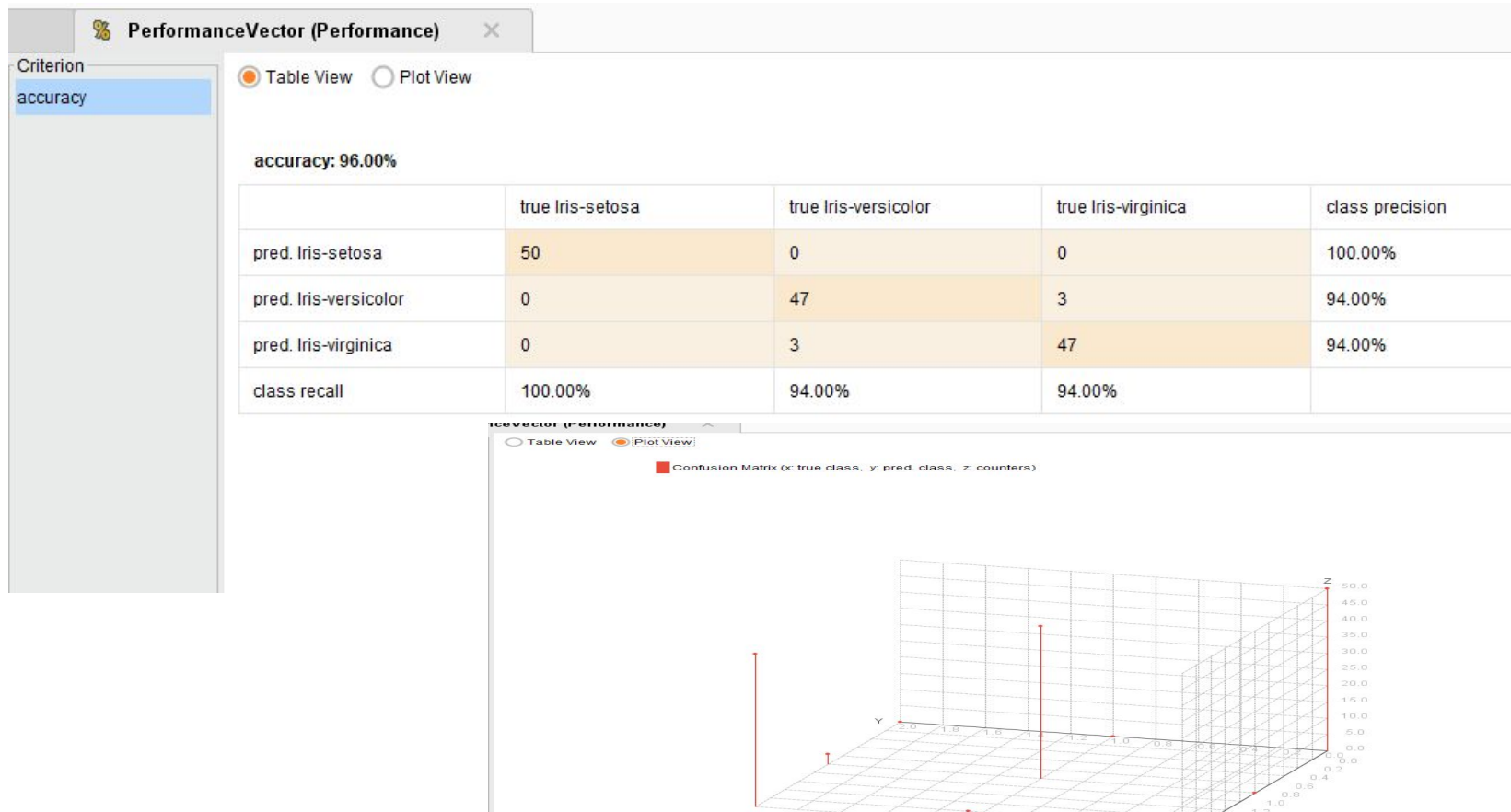
Open in  Turbo Prep  Auto Model

Row No.	id	label	prediction(la...	confidence(l...	confidence(l...	confidence(l...	a1	a2
60	id_60	Iris-versicolor	Iris-versicolor	0	1.000	0.000	5.200	2.700
61	id_61	Iris-versicolor	Iris-versicolor	0	1.000	0.000	5	2
62	id_62	Iris-versicolor	Iris-versicolor	0	0.996	0.004	5.900	3
63	id_63	Iris-versicolor	Iris-versicolor	0	1.000	0.000	6	2.200
64	id_64	Iris-versicolor	Iris-versicolor	0	0.985	0.015	6.100	2.900
65	id_65	Iris-versicolor	Iris-versicolor	0	1.000	0.000	5.600	2.900
66	id_66	Iris-versicolor	Iris-versicolor	0	0.979	0.021	6.700	3.100
67	id_67	Iris-versicolor	Iris-versicolor	0	0.990	0.010	5.600	3
68	id_68	Iris-versicolor	Iris-versicolor	0	1.000	0.000	5.800	2.700
69	id_69	Iris-versicolor	Iris-versicolor	0	0.994	0.006	6.200	2.200
70	id_70	Iris-versicolor	Iris-versicolor	0	1.000	0.000	5.600	2.500
71	id_71	Iris-versicolor	Iris-virginica	0	0.161	0.839	5.900	3.200
72	id_72	Iris-versicolor	Iris-versicolor	0	1.000	0.000	6.100	2.800
73	id_73	Iris-versicolor	Iris-versicolor	0	0.925	0.075	6.300	2.500
74	id_74	Iris-versicolor	Iris-versicolor	0	0.998	0.002	6.100	2.800
75	id_75	Iris-versicolor	Iris-versicolor	0	0.998	0.002	6.400	2.900
76	id_76	Iris-versicolor	Iris-versicolor	0	0.987	0.013	6.600	3
77	id_77	Iris-versicolor	Iris-versicolor	0	0.910	0.090	6.800	2.800
78	id_78	Iris-versicolor	Iris-virginica	0	0.080	0.920	6.700	3

rapidMiner: Iris elemzése NB osztályozóval



rapidMiner: Iris elemzése NB osztályozóval



rapidMiner: Iris elemzése NB osztályozóval

adatok_iris - Notepad

a1;	a2;	a3;	a4
5.1;	3.5;	1.4;	0.2
4.9;	3.0;	1.4;	0.2
4.7;	3.2;	1.3;	0.2
4.6;	3.1;	1.5;	0.2
5.0;	3.6;	1.4;	0.2
5.4;	3.9;	1.7;	0.4
4.6;	3.4;	1.4;	0.3

Parameters

Read CSV

Import Configuration Wizard...

csv file: nkaasztal_2019\adatok_iris.txt

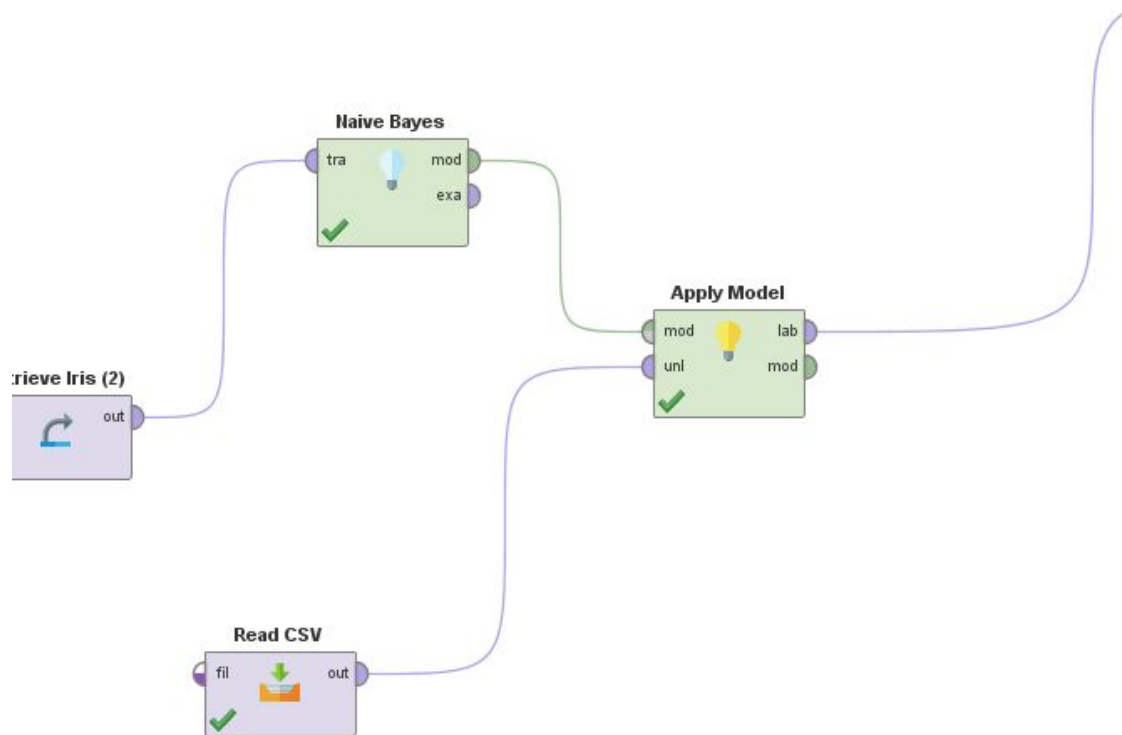
column separators: ;

☒ use quotes

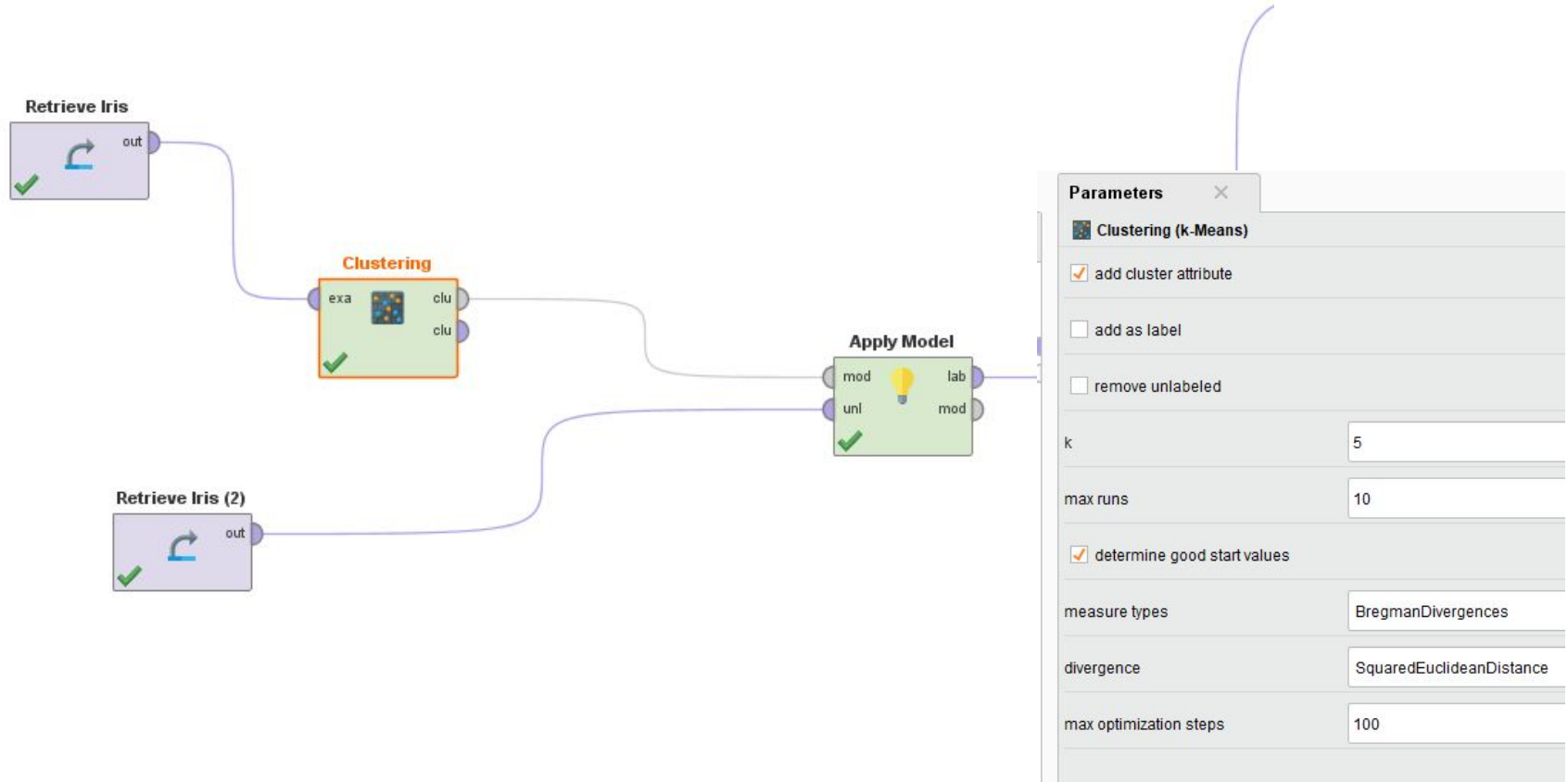
quotes character: "

☐ skip comments

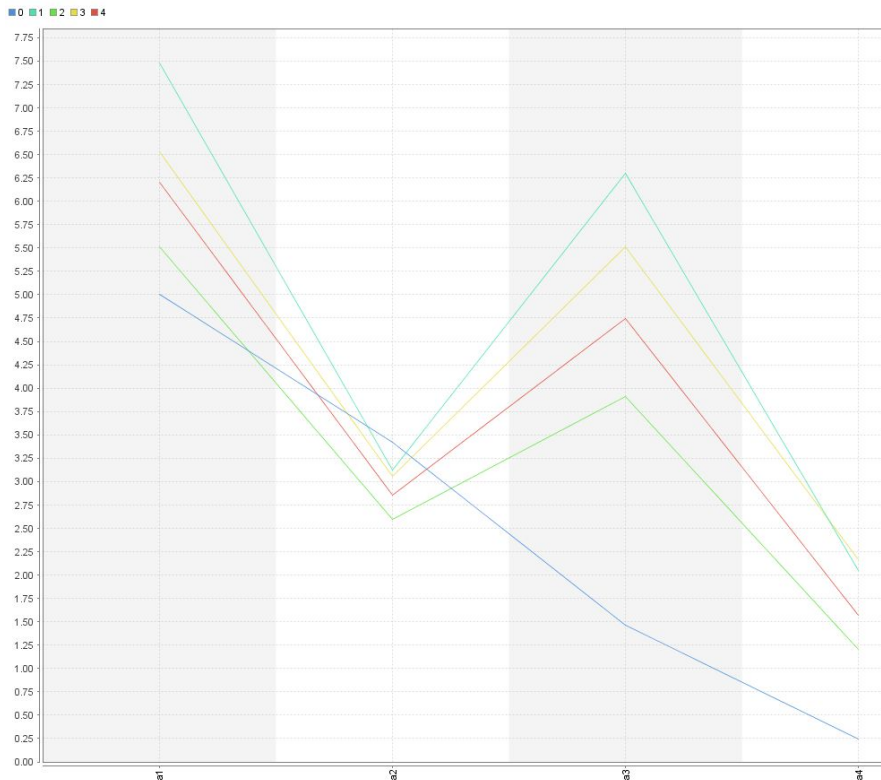
☒ parse numbers



rapidMiner: Iris elemzése k-means klaszterezővel



rapidMiner: Iris elemzése k-means klaszterezővel



ExampleSet (Apply Model) X

Open in [Turbo Prep](#) [Auto Model](#)

Row No.	id	label	cluster	a1	a2	a3
86	id_86	Iris-versicolor	cluster_4	6	3.400	4.500
87	id_87	Iris-versicolor	cluster_4	6.700	3.100	4.700
88	id_88	Iris-versicolor	cluster_4	6.300	2.300	4.400
89	id_89	Iris-versicolor	cluster_2	5.600	3	4.100
90	id_90	Iris-versicolor	cluster_2	5.500	2.500	4
91	id_91	Iris-versicolor	cluster_2	5.500	2.600	4.400
92	id_92	Iris-versicolor	cluster_4	6.100	3	4.600
93	id_93	Iris-versicolor	cluster_2	5.800	2.600	4
94	id_94	Iris-versicolor	cluster_2	5	2.300	3.300
95	id_95	Iris-versicolor	cluster_2	5.600	2.700	4.200
96	id_96	Iris-versicolor	cluster_2	5.700	3	4.200
97	id_97	Iris-versicolor	cluster_2	5.700	2.900	4.200
98	id_98	Iris-versicolor	cluster_4	6.200	2.900	4.300
99	id_99	Iris-versicolor	cluster_2	5.100	2.500	3
100	id_100	Iris-versicolor	cluster_2	5.700	2.800	4.100
101	id_101	Iris-virginica	cluster_3	6.300	3.300	6
102	id_102	Iris-virginica	cluster_4	5.800	2.700	5.100
103	id_103	Iris-virginica	cluster_1	7.100	3	5.900

rapidMiner: Iris elemzése k-means klaszterezővel

