

**José Alberto Silva de Sá
Gilberto Ribeiro de Queiroz
Rafael Duarte Coelho dos Santos**

Introdução à Programação Visual

orange



**São José dos Campos
2018**

**José Alberto Silva de Sá
Gilberto Ribeiro de Queiroz
Rafael Duarte Coelho dos Santos**

**Introdução à
Programação Visual**

orange



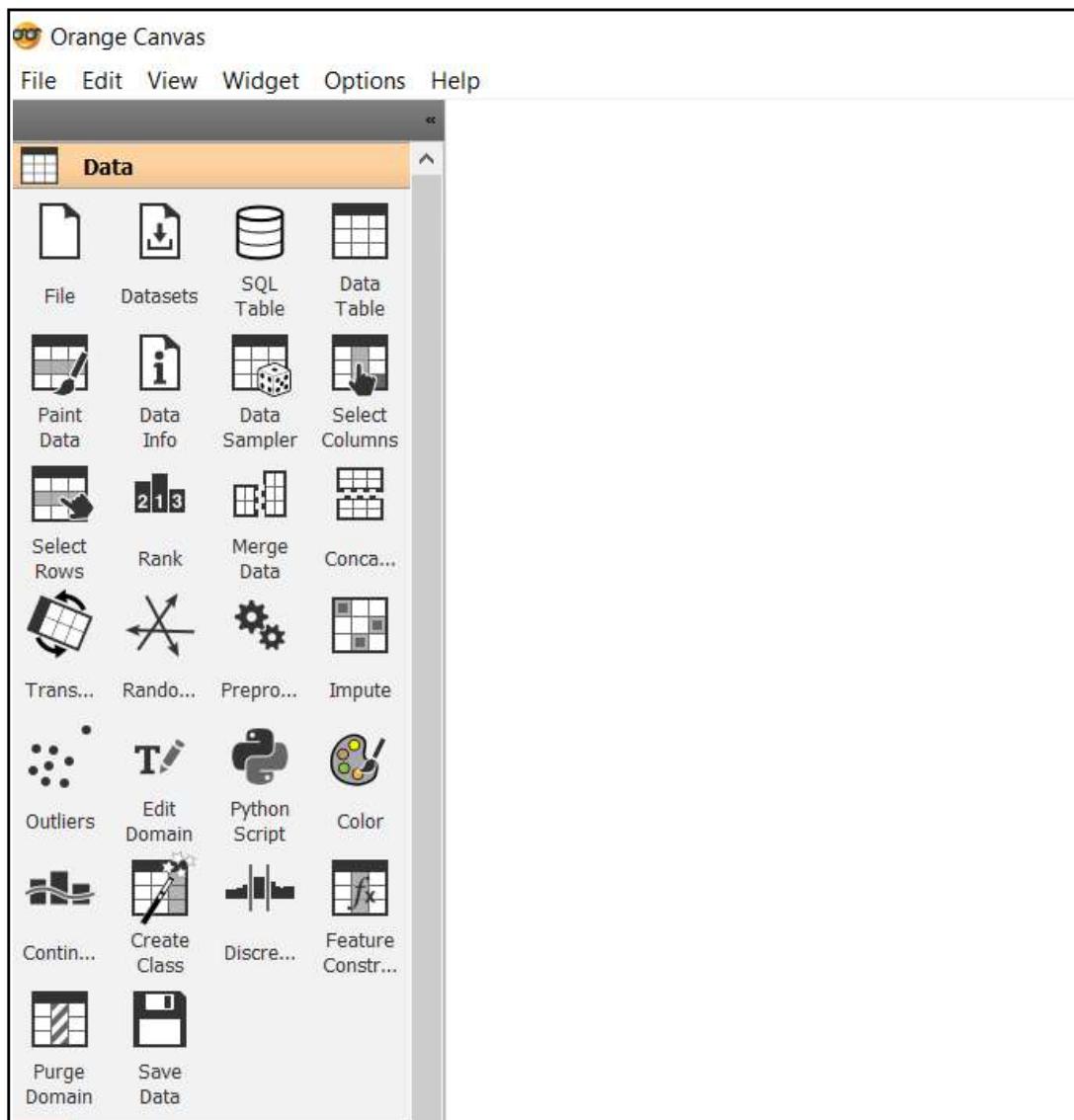
**São José dos Campos
2018**

APRESENTAÇÃO

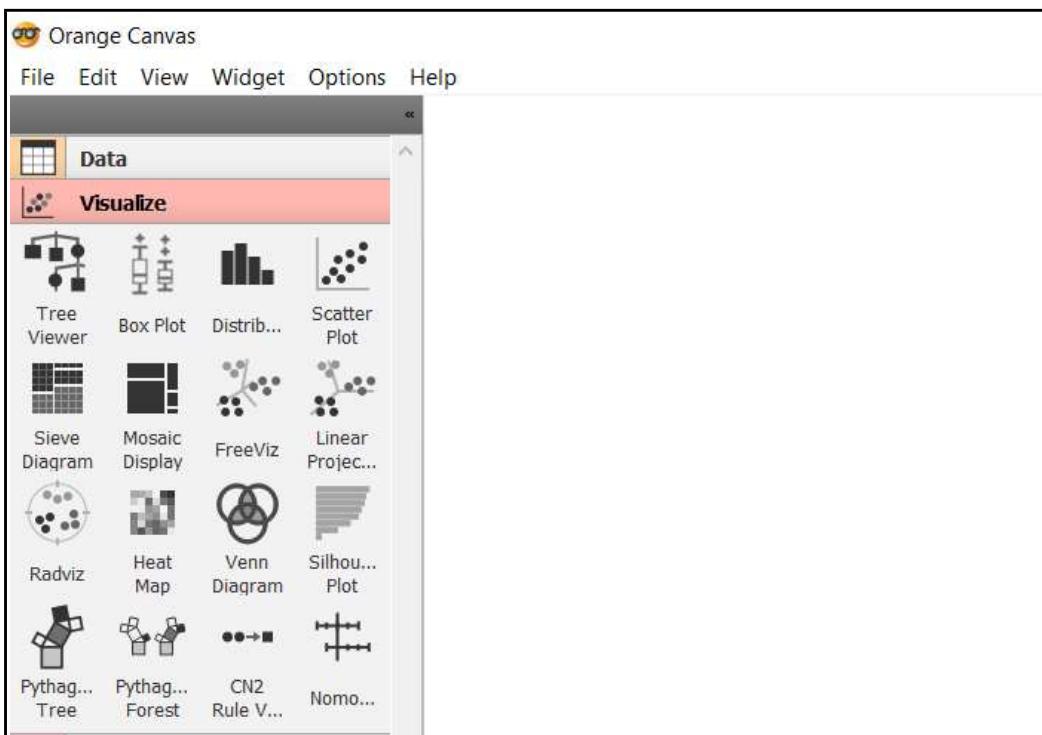
Orange é um *Software Open Source* de programação visual baseado em componentes (*component-based visual programming*) utilizado para a análise de dados e aprendizado de máquina. Ele possui um ambiente interativo (interface gráfica), onde os componentes do Orange (chamados **widgets**) podem realizar desde simples visualização, seleção e pré-processamento de dados até avaliação de algoritmos de aprendizado (classificadores e regressores).

Na interface gráfica de programação visual são criados os **workflows** dos *widgets*. Os *widgets* podem ser pré-definidos ou desenvolvidos pelos próprios usuários. Usuários experientes podem utilizar o Orange como um biblioteca Python de análise de dados. Existe uma grande quantidade de *widgets*, normalmente divididos nas categorias:

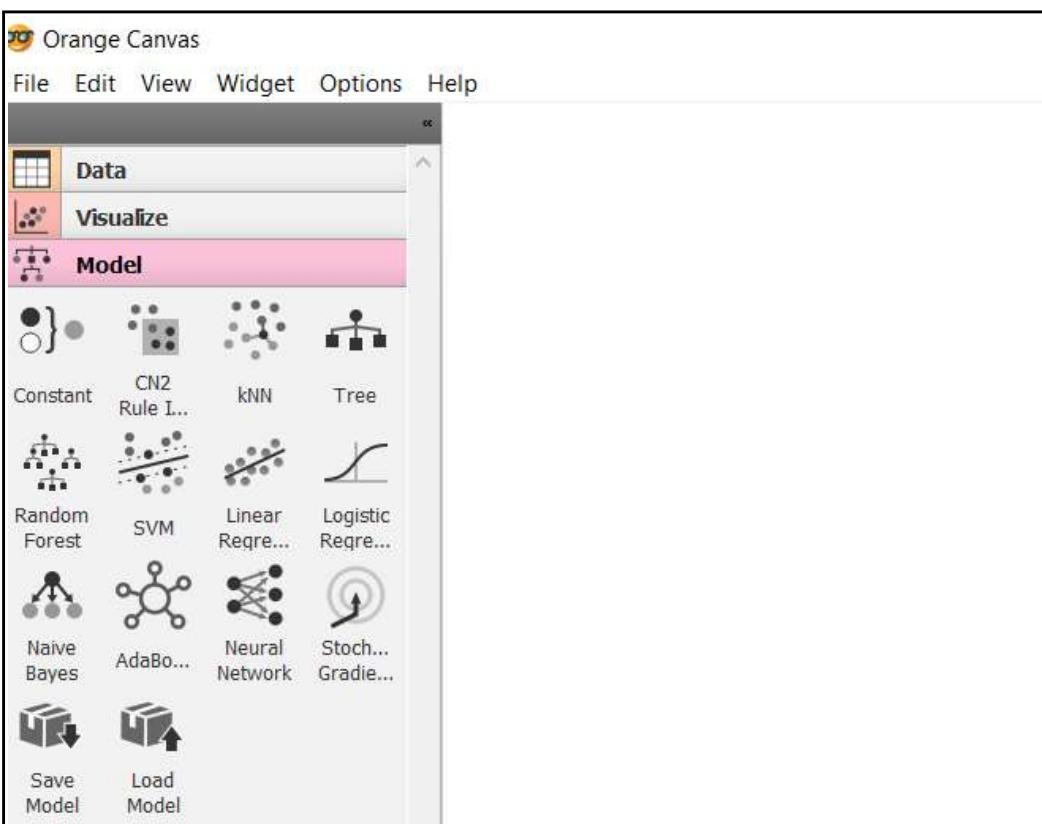
a) Dados (Data): Widgets para realizar operações básicas com os dados.



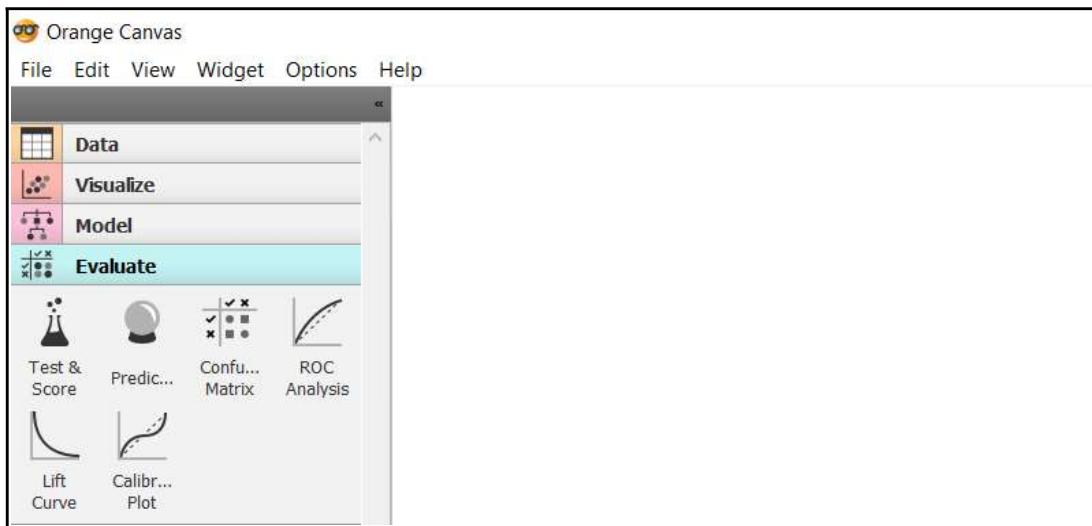
b) Visualização (Visualize): Widgets utilizados para a visualização dos dados.



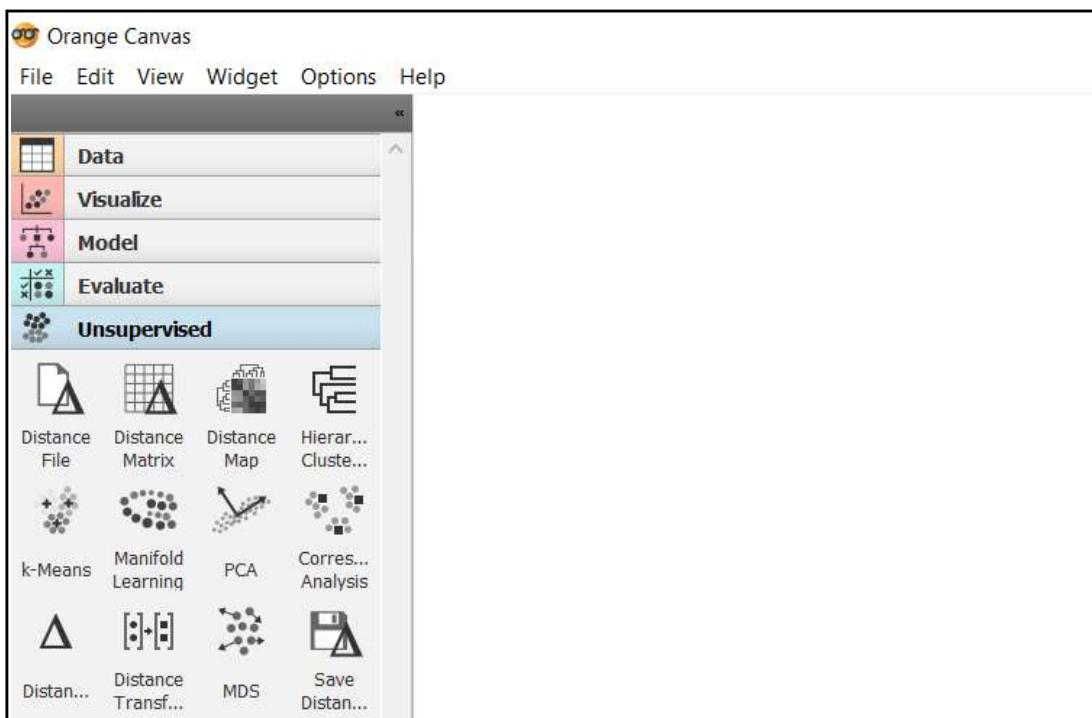
c) Modelos de Aprendizagem Supervisionada (Model): Widgets utilizados para desenvolver classificadores ou regressores por aprendizagem supervisionada.



d) Avaliação (Evaluate): Widgets utilizados para avaliar a acurácia de modelos.



e) Modelos de Aprendizagem Não Supervisionada (Unsupervised): Widgets utilizados para desenvolver agrupamentos de dados.



O site oficial (<https://blog.biolab.si/>) fornece suporte com extensa documentação, tutoriais, blog e vídeos sobre a utilização do aplicativo nos assuntos relacionados a *Data Science*. A instalação do Orange pode ser realizada nos ambientes:



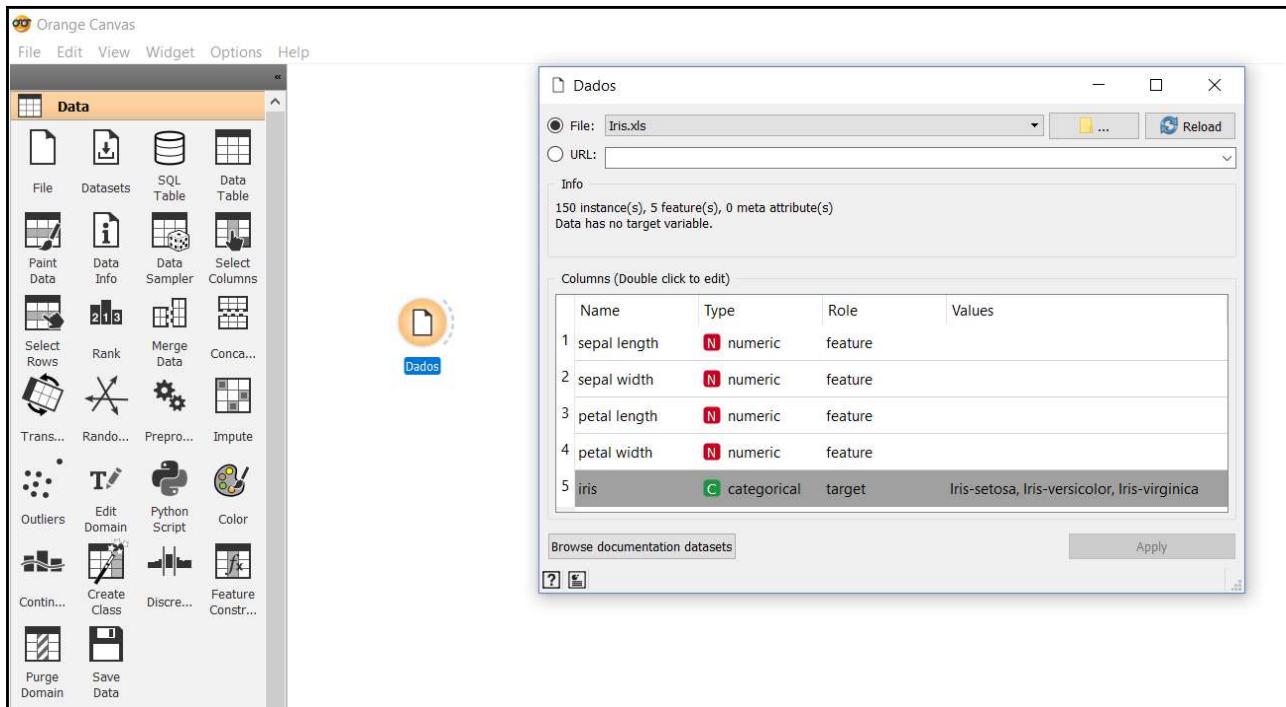
1 LEITURA DE DADOS E OPERAÇÕES BÁSICAS

Neste item iremos demonstrar como se realiza a leitura e as operações básicas de manipulação de dados no Software Orange Data Mining.

1.1 File Widget

O Orange possui formato próprio de dados (.tab), porém também pode manipular arquivos de dados nativos do Excel (.xlsx ou .xls), delimitados por vírgulas ou tabulações. O conjunto de dados geralmente é apresentado em uma tabela, com instâncias de dados (amostras) dispostas em linhas e seus atributos em colunas. Os atributos podem ser de tipos diferentes (contínuos, discretos, dentre outros) podendo ter diferentes finalidades (recursos de entrada, meta-atributos e classe).

Veja o exemplo:



Etapas:

- 1) Clicar (BE) e Arrastar o “File Widget” para a Área de Trabalho.
- 2) Clicar (BD) e Renomear o “File Widget” para “Dados”.
- 3) Clicar (BD) e Abrir o arquivo Iris.xls.
- 4) Se necessário, clicar sobre os elementos de “Name”, “Type”, “Role” e “Values” para realizar alterações. Por exemplo, você pode mudar o elemento “target” por “feature” clicando sobre ele e realizando a modificação.

1.2 Data Table Widget

Podemos ver todos os dados mediante o *Data Table Widget* (uma tabela). Veja o exemplo:

	iris	sepal length	sepal width	petal length	petal width
1	Iris-setosa	5.1	3.5	1.4	0.2
2	Iris-setosa	4.9	3.0	1.4	0.2
3	Iris-setosa	4.7	3.2	1.3	0.2
4	Iris-setosa	4.6	3.1	1.5	0.2
5	Iris-setosa	5.0	3.6	1.4	0.2
6	Iris-setosa	5.4	3.9	1.7	0.4
7	Iris-setosa	4.6	3.4	1.4	0.3
8	Iris-setosa	5.0	3.4	1.5	0.2
9	Iris-setosa	4.4	2.9	1.4	0.2
10	Iris-setosa	4.9	3.1	1.5	0.1
11	Iris-setosa	5.4	3.7	1.5	0.2
12	Iris-setosa	4.8	3.4	1.6	0.2
13	Iris-setosa	4.8	3.0	1.4	0.1
14	Iris-setosa	4.3	3.0	1.1	0.1
15	Iris-setosa	5.8	4.0	1.2	0.2

Etapas:

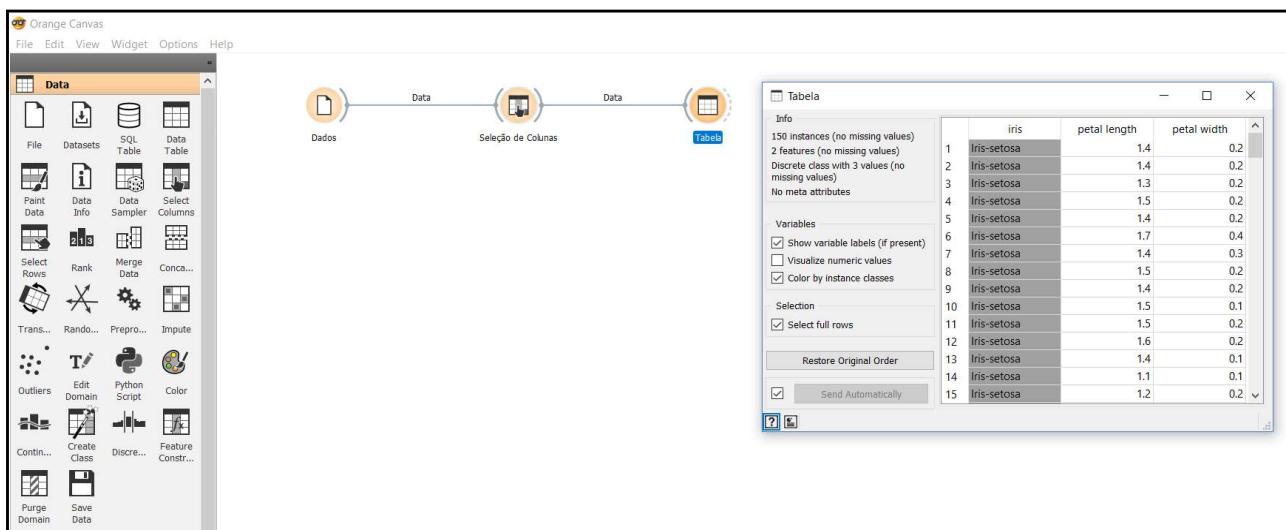
- 1) Clicar (BE) e Arrastar o “Data Table Widget” para a Área de Trabalho.
- 2) Clicar (BD) e Renomear o “Data Table Widget” para “Tabela”.
- 3) Ligar (BE) o “Dados Widget” com o “Tabela Widget”.
- 4) Clicar (BD) e Abrir a Tabela.

1.3 Select Columns Widget

Você pode criar subconjuntos dos dados originais com o *Select Columns Widget*.

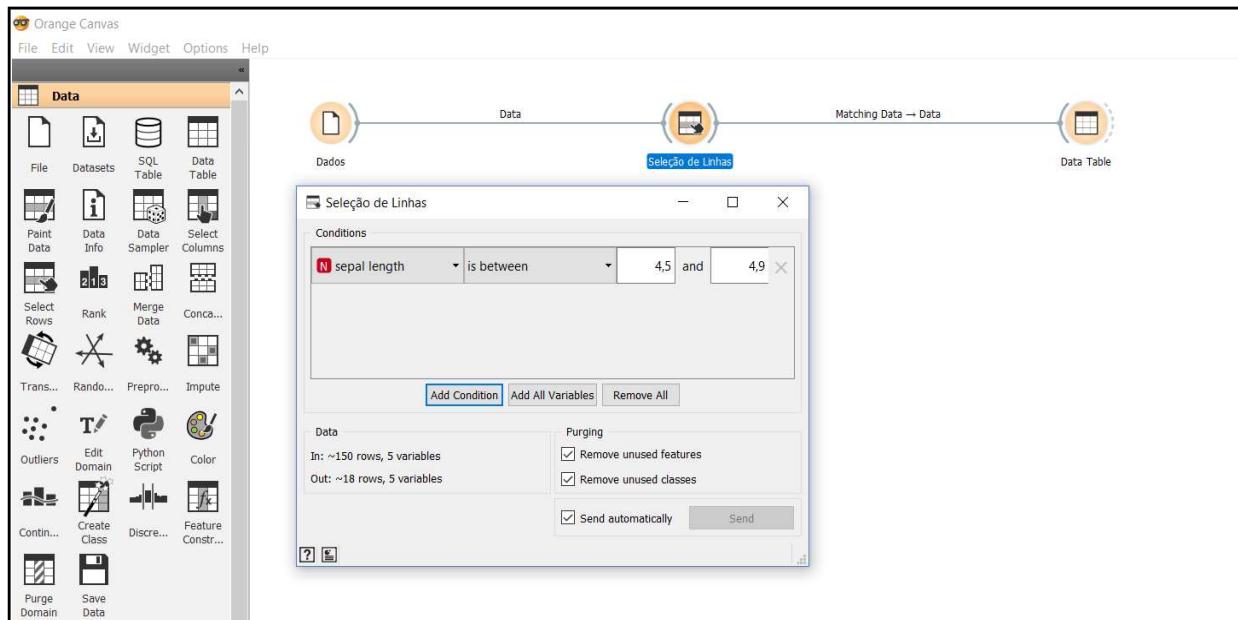
Etapas:

- 1) Clicar (BE) e Arrastar o “Select Columns Widget” para a Área de Trabalho.
- 2) Clicar (BD) e Renomear o “Select Columns Widget” para “Seleção de Colunas”.
- 3) Ligar (BE) o “Seleção de Colunas” com os “Dados” e com “Tabela”.
- 4) Clicar (BD) e Abrir a “Seleção de Colunas”.
- 5) Filtrar (Filter) apenas os atributos que sairão da massa original. Neste exemplo, foram retirados da massa original os atributos “sepal length” e “sepal width”.
- 6) O subconjunto formado pode ser visualizados mediante o “Tabela Widget”, ou seja, o subconjunto possui os atributos “petal length” e “petal width” e classe-alvo (target) “iris”.



1.4 Select Rows Widget

Você também pode eliminar instâncias (amostras) de um conjunto de dados mediante o *Select Rows Widget*. Veja o exemplo:



Como consequência temos:

The screenshot shows the Orange Data Mining software interface. On the left is the 'Data' tab of the palette, containing icons for various data manipulation and visualization tools. A flow diagram is visible in the center: 'Dados' (Data) connects to 'Seleção de Linhas' (Selection), which then connects to a 'Data Table' widget. The 'Data Table' widget displays the Iris dataset. The table has 150 rows and 5 columns: 'iris' (class), 'sepal length', 'sepal width', 'petal length', and 'petal width'. The first 150 rows represent 'Iris-setosa', row 151 represents 'Iris-versicolor', and row 152 represents 'Iris-virginica'. The 'Data Table' window also includes sections for 'Info', 'Variables', and 'Selection'.

Atenção!

Você pode acrescentar, na massa original de dados, prefixos aos atributos usando o símbolo hash "#":

Informações do Atributo (Roles):

- c: Classe (class attribute)
- m: Meta Dado (meta attribute)
- i: Ignorar o Atributo (ignore the attribute)
- w: Pesos do Atributo (instance weights)

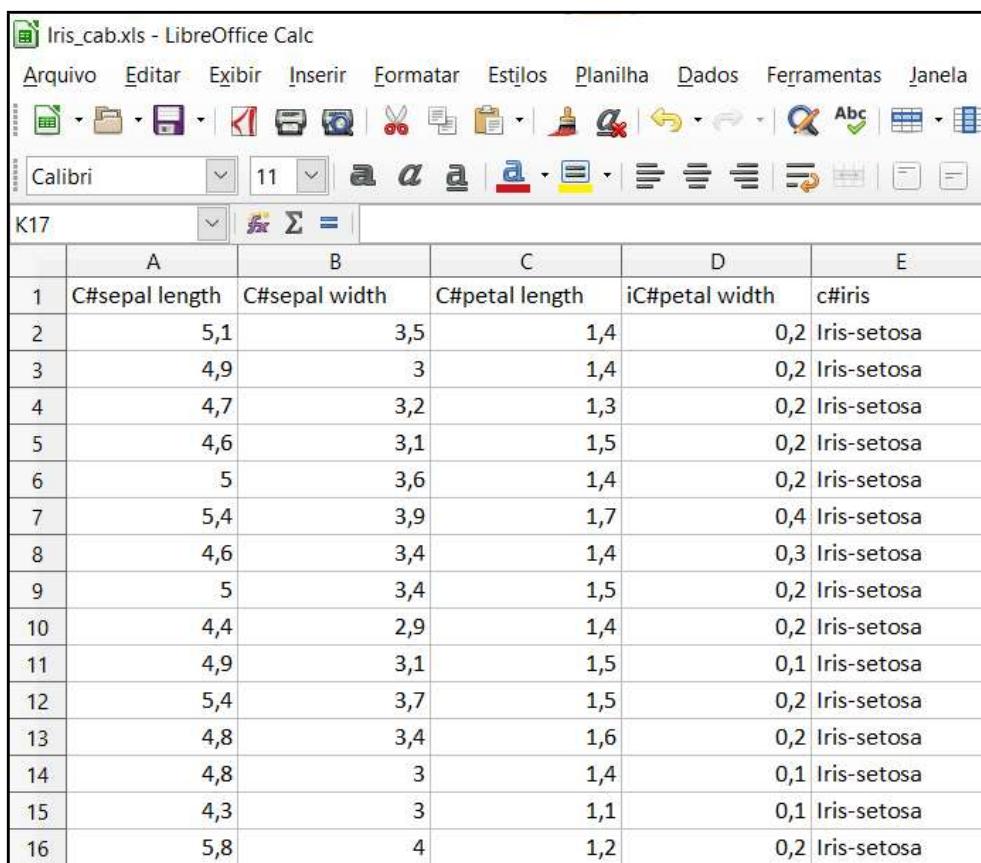
Tipo do Atributo (Type):

- C: Contínuo (Continuous)
- D: Discreto (Discrete)
- T: Tempo (Time)
- S: Texto (String)

Observações:

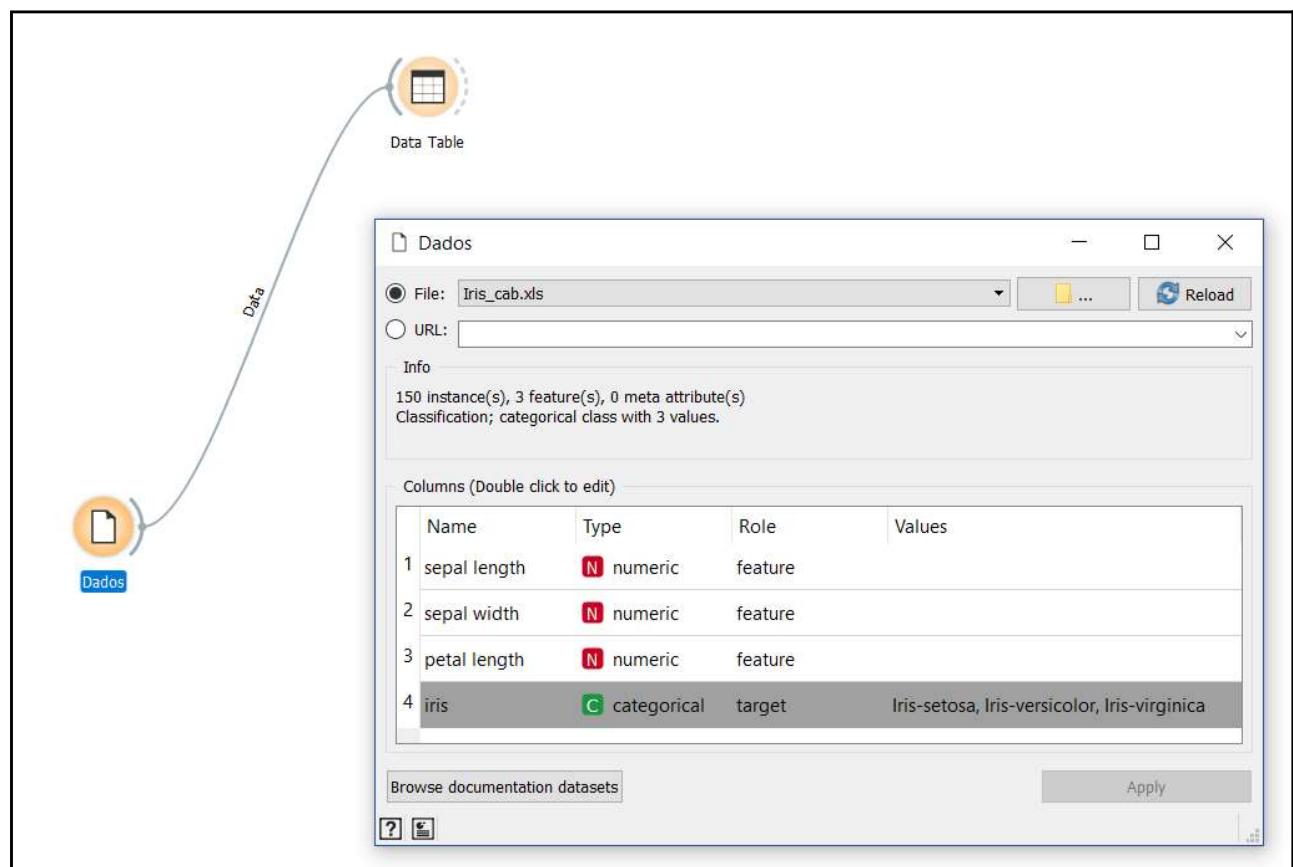
- Caso não haja informações do Atributo ele será um Atributo Preditivo.
- A informação "i" retira o Atributo da Tabela de Dados (Data Table).

Como consequência temos:



The screenshot shows a LibreOffice Calc spreadsheet titled "Iris_cab.xls - LibreOffice Calc". The menu bar includes Arquivo, Editar, Exibir, Inserir, Formatar, Estilos, Planilha, Dados, Ferramentas, and Janela. The toolbar includes icons for file operations, text styles, and tables. The formula bar shows "K17" and the formula "Σ =". The data starts at row 1 with columns A through E. Column A contains row numbers from 1 to 16. Columns B through D contain numerical values. Column E contains categorical labels: Iris-setosa for rows 1-15 and Iris-virginica for row 16.

	A	B	C	D	E
1	C#sepal length	C#sepal width	C#petal length	iC#petal width	c#iris
2	5,1	3,5	1,4	0,2	Iris-setosa
3	4,9	3	1,4	0,2	Iris-setosa
4	4,7	3,2	1,3	0,2	Iris-setosa
5	4,6	3,1	1,5	0,2	Iris-setosa
6	5	3,6	1,4	0,2	Iris-setosa
7	5,4	3,9	1,7	0,4	Iris-setosa
8	4,6	3,4	1,4	0,3	Iris-setosa
9	5	3,4	1,5	0,2	Iris-setosa
10	4,4	2,9	1,4	0,2	Iris-setosa
11	4,9	3,1	1,5	0,1	Iris-setosa
12	5,4	3,7	1,5	0,2	Iris-setosa
13	4,8	3,4	1,6	0,2	Iris-setosa
14	4,8	3	1,4	0,1	Iris-setosa
15	4,3	3	1,1	0,1	Iris-setosa
16	5,8	4	1,2	0,2	Iris-virginica



The screenshot shows the KNIME Data Table node interface. On the left, there is a 'Dados' icon with a 'Data' connection line pointing to the node. The node itself has a 'Data Table' icon. The interface includes an 'Info' section with details about the dataset: 150 instances (no missing values), 3 features (no missing values), Discrete class with 3 values (no missing values), and No meta attributes. It also includes sections for 'Variables' (with checkboxes for 'Show variable labels (if present)', 'Visualize numeric values', and 'Color by instance classes'), 'Selection' (with a checkbox for 'Select full rows'), and buttons for 'Restore Original Order' and 'Send Automatically'. The main area displays a table with 19 rows of data:

	iris	sepal length	sepal width	petal length
1	Iris-setosa	5.1	3.5	1.4
2	Iris-setosa	4.9	3.0	1.4
3	Iris-setosa	4.7	3.2	1.3
4	Iris-setosa	4.6	3.1	1.5
5	Iris-setosa	5.0	3.6	1.4
6	Iris-setosa	5.4	3.9	1.7
7	Iris-setosa	4.6	3.4	1.4
8	Iris-setosa	5.0	3.4	1.5
9	Iris-setosa	4.4	2.9	1.4
10	Iris-setosa	4.9	3.1	1.5
11	Iris-setosa	5.4	3.7	1.5
12	Iris-setosa	4.8	3.4	1.6
13	Iris-setosa	4.8	3.0	1.4
14	Iris-setosa	4.3	3.0	1.1
15	Iris-setosa	5.8	4.0	1.2
16	Iris-setosa	5.7	4.4	1.5
17	Iris-setosa	5.4	3.9	1.3
18	Iris-setosa	5.1	3.5	1.4
19	Iris-setosa	5.7	3.8	1.7

Você também pode formatar usando um cabeçalho de três linhas:

The screenshot shows a LibreOffice Calc spreadsheet titled 'Iris_cab_3_lin.xls'. The data starts with a three-line header: 'sepal length', 'sepal width', and 'petal length'. Rows 2 through 14 contain numerical data. Row 15 contains the categorical 'iris' column. The KNIME interface is shown below the spreadsheet, with a 'Dados' icon connected to the 'Dados' node. The 'Dados' node interface shows the file is 'Iris_cab_3_lin.xls', with 'File' selected. It displays the same 150 instances, 3 features, and 1 meta attribute as the KNIME Data Table node.

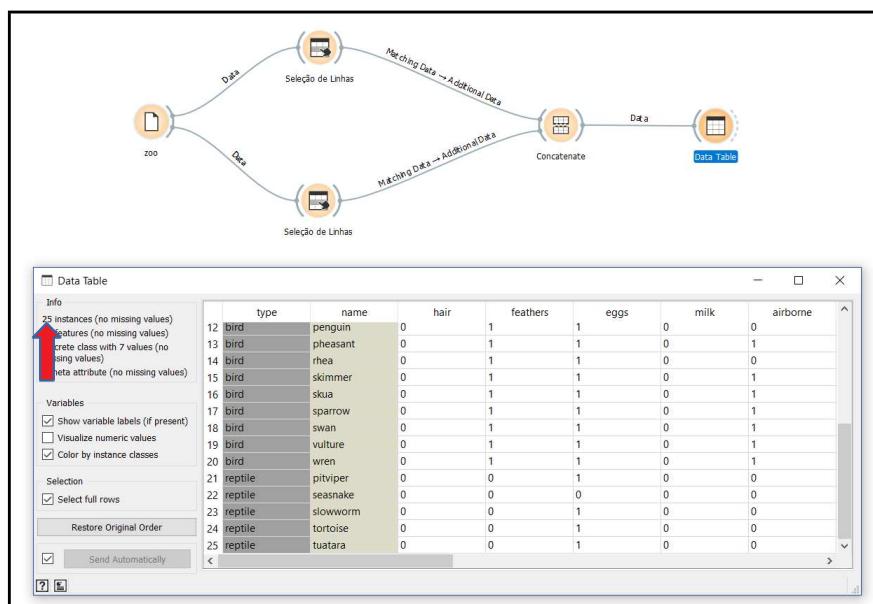
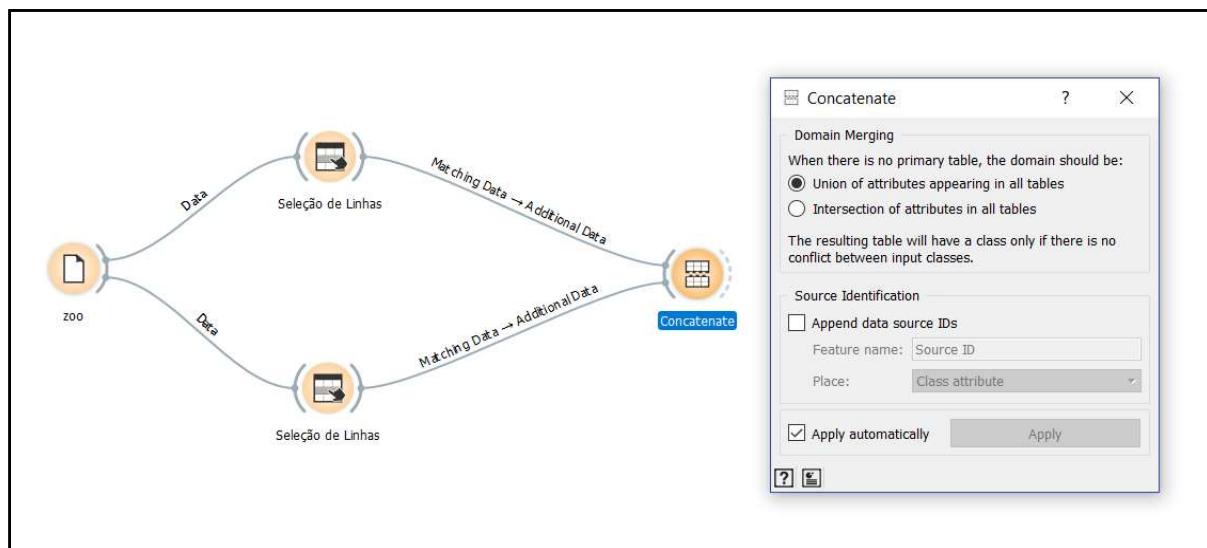
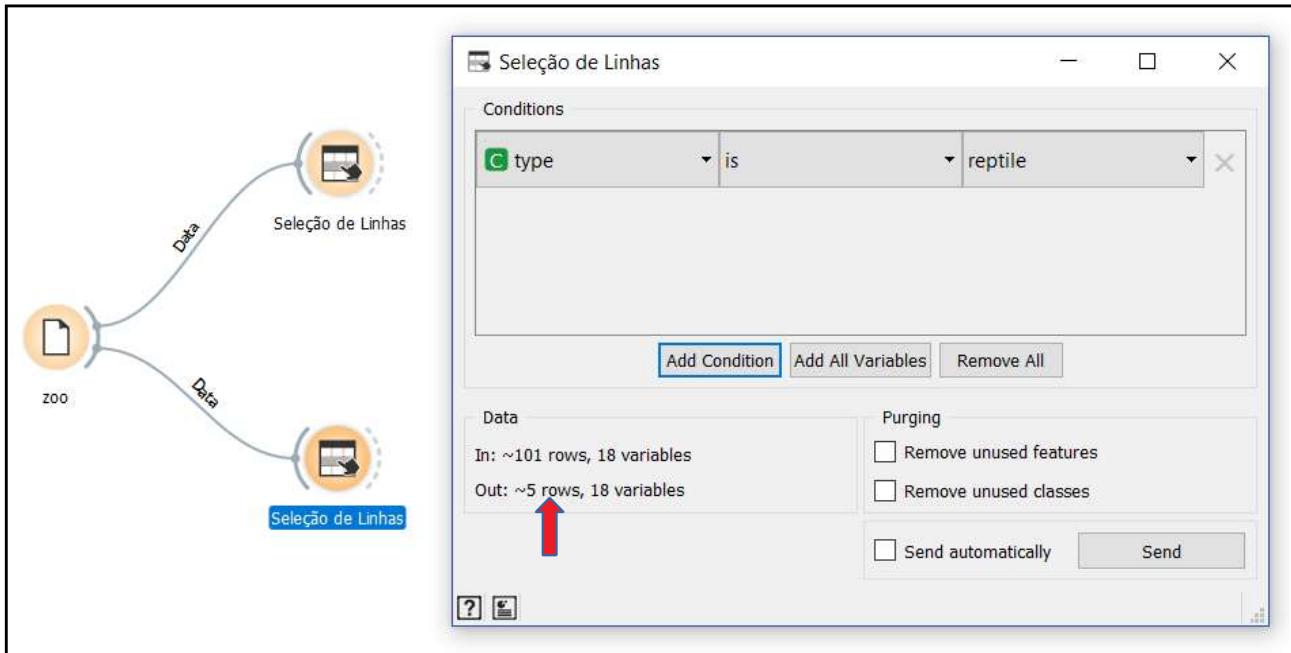
1.5 Transpose Widget

Podemos transpor uma tabela de dados com *Transpose Widget*.

	Feature name	Feature 001 Iris-setosa	Feature 002 Iris-setosa	Feature 003 Iris-setosa	Feature 004 Iris-setosa	Feature 005 Iris-setosa	Feature 006 Iris-setosa	Feature 007 Iris-setosa
1	sepal length	5.100	4.900	4.700	4.600	5.000	5.400	4.600
2	sepal width	3.500	3.000	3.200	3.100	3.600	3.900	3.400
3	petal length	1.400	1.400	1.300	1.500	1.400	1.700	1.400
4	petal width	0.200	0.200	0.200	0.200	0.200	0.400	0.300

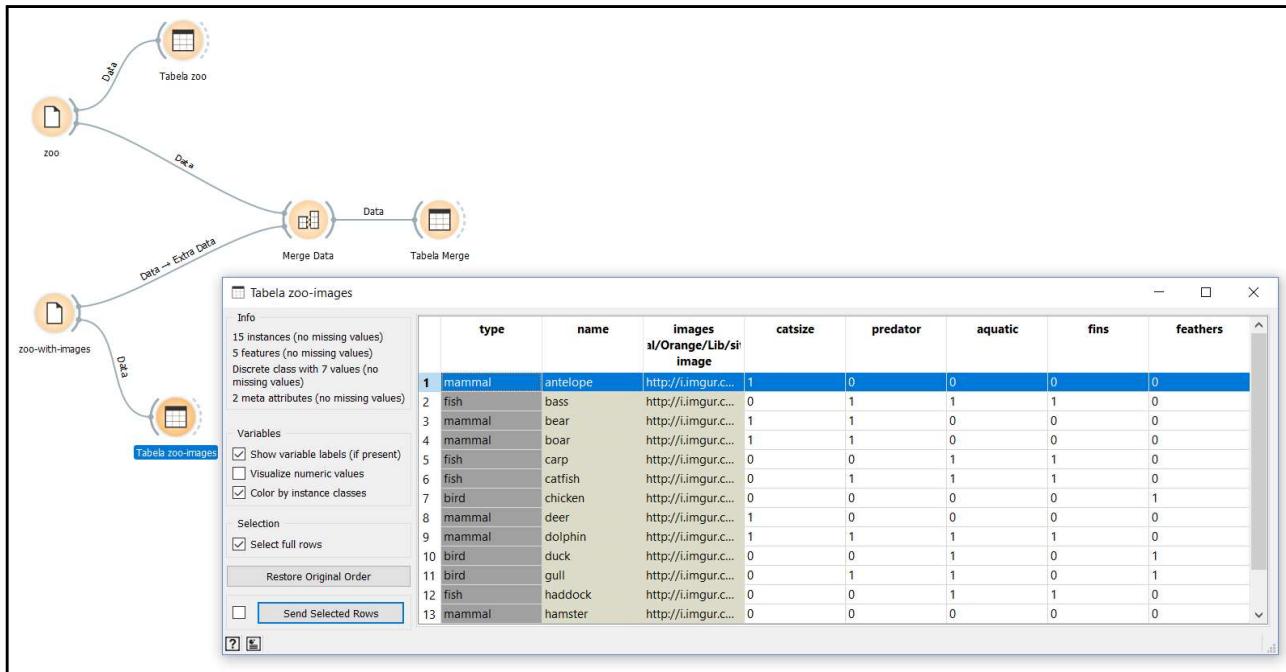
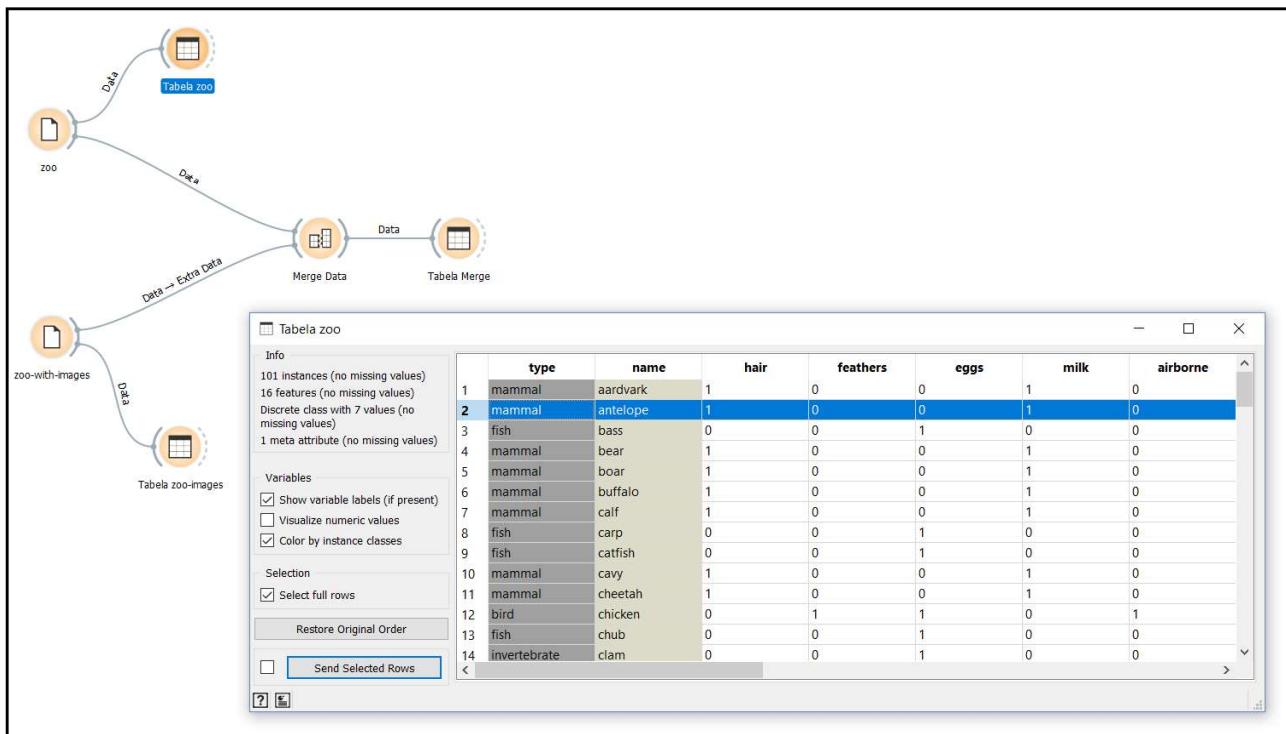
1.6 Concatenate Widget

Podemos concatenar dados de múltiplas fontes.



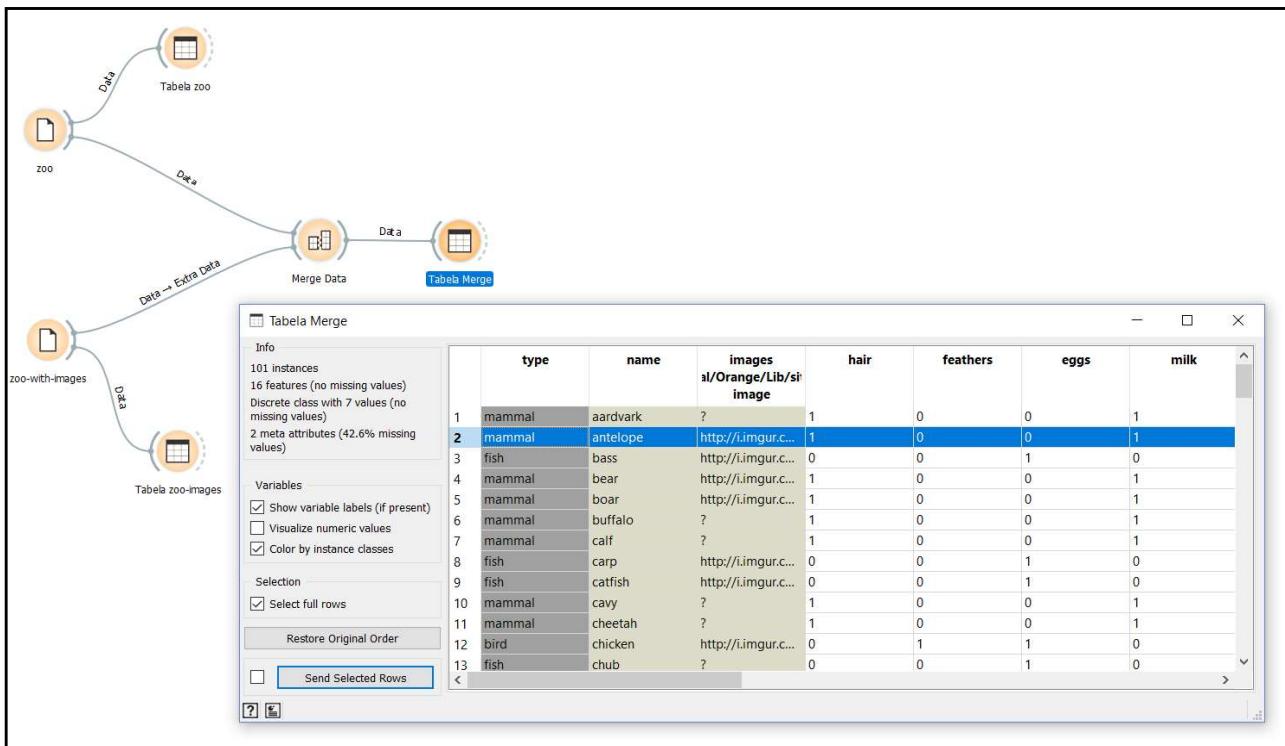
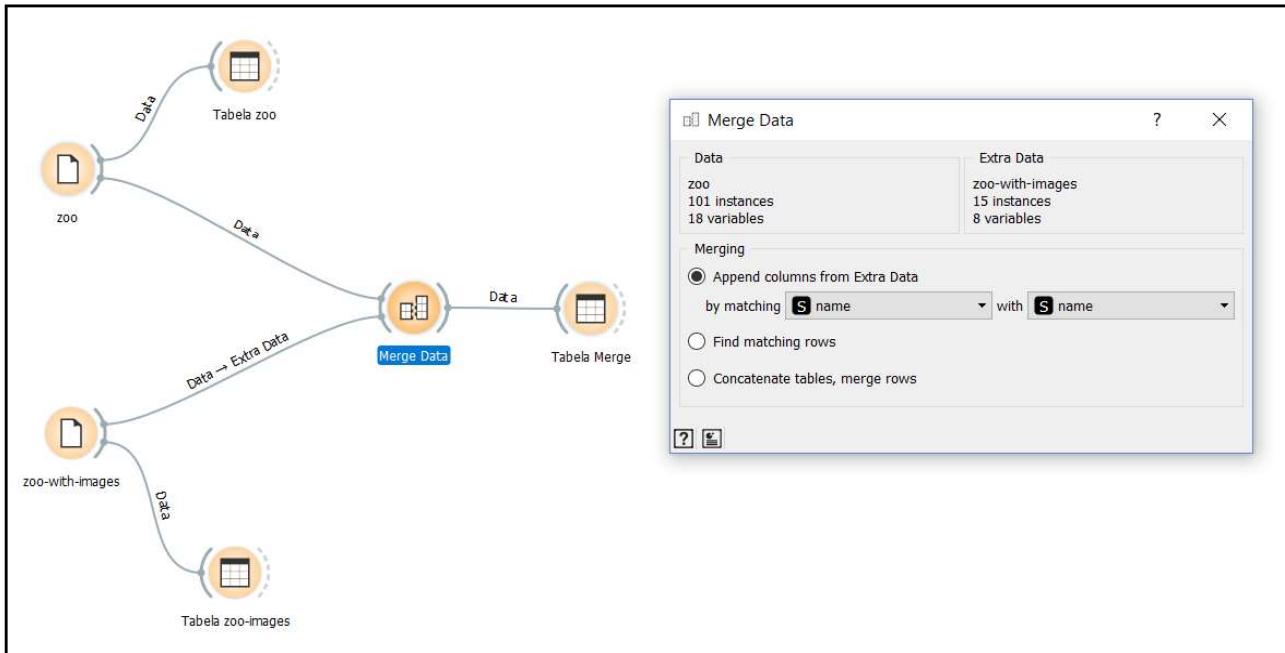
1.7 Merge Widget

Podemos mesclar dois conjuntos de dados, com base nos valores dos atributos selecionados. No exemplo abaixo, deseja-se mesclar o arquivo “zoo.tab” com “zoo-with-images.tab”. Ambos os arquivos compartilham nomes de atributos comuns.



Atenção!

“Extra Data” não sobrepõe valores. Os valores da tabela principal possuem preferência.



1.8 Discretize Widget

Podemos discretizar os valores contínuos de atributos, ou seja, dividir ou particionar um todo em partes com menor complexidade, com a finalidade de facilitar a análise dos dados. No *workflow* abaixo, mostramos o conjunto de “Dados Íris” com atributos contínuos (como no arquivo de dados original) e com atributos discretizados.

Top Screenshot: Discretize Dialog Box

The 'Discretize' dialog box shows the following settings:

- Default Discretization:** Entropy-MDL discretization (selected)
- Num. of intervals:** 3
- Individual Attribute Settings:**
 - sepal length: 5.55, 6.15
 - sepal width: 2.95, 3.35
 - petal length: 2.45, 4.75
 - petal width: 0.80, 1.75
- Buttons:** Apply automatically (unchecked), Apply

Middle Screenshot: Data Flow and Tabela Original/Tabela após Discretize

The workflow diagram shows the 'Iris' dataset being loaded into 'Tabela Original'. This is then processed by the 'Discretize' widget, resulting in 'Tabela após Discretize'.

Bottom Screenshot: Tabela Original and Tabela após Discretize

Tabela Original:

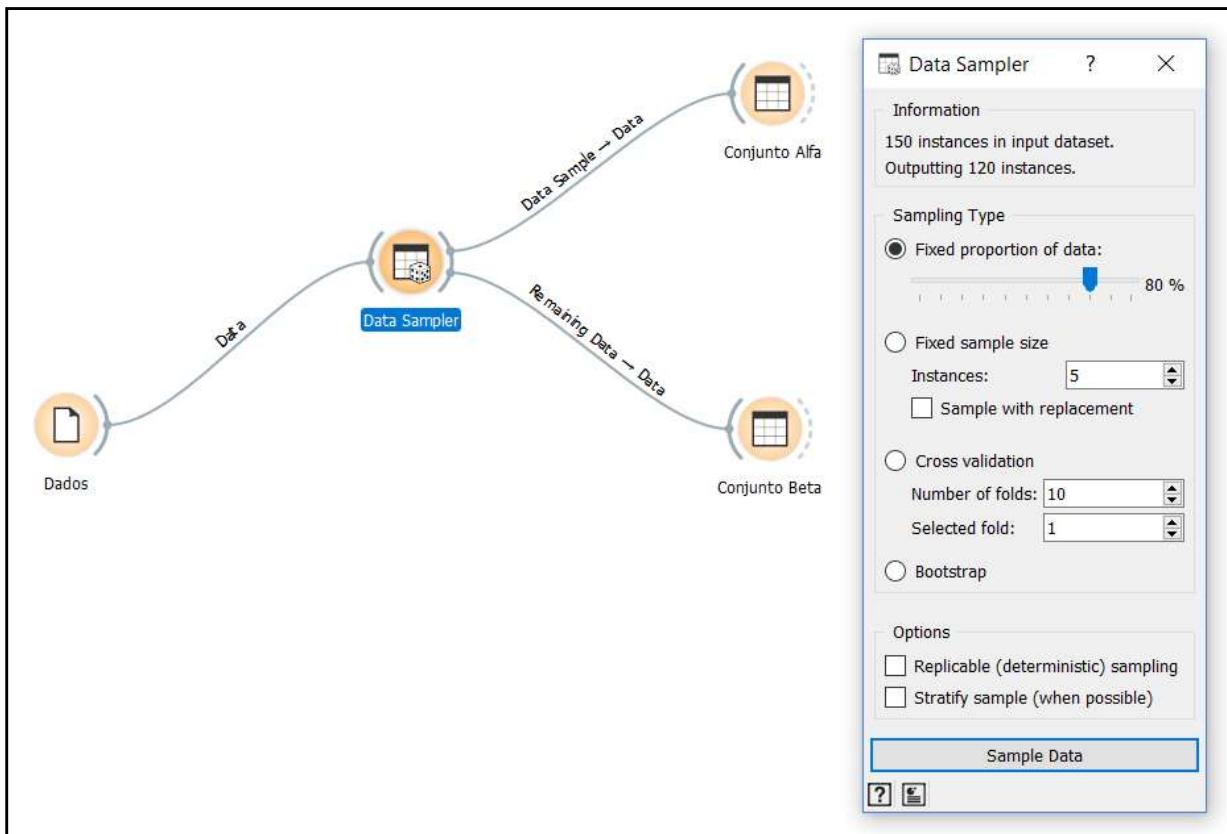
	iris	sepal length	sepal width	petal length	petal width
1	Iris-setosa	5.1	3.5	1.4	0.2
2	Iris-setosa	4.9	3.0	1.4	0.2
3	Iris-setosa	4.7	3.2	1.3	0.2
4	Iris-setosa	4.6	3.1	1.5	0.2
5	Iris-setosa	5.0	3.6	1.4	0.2
6	Iris-setosa	5.4	3.9	1.7	0.4
7	Iris-setosa	4.6	3.4	1.4	0.3
8	Iris-setosa	5.0	3.4	1.5	0.2
9	Iris-setosa	4.4	2.9	1.4	0.2
10	Iris-setosa	4.9	3.1	1.5	0.1
11	Iris-setosa	5.4	3.7	1.5	0.2
12	Iris-setosa	4.8	3.4	1.6	0.2
13	Iris-setosa	4.8	3.0	1.4	0.1
14	Iris-setosa	4.3	3.0	1.1	0.1
15	Iris-setosa	5.8	4.0	1.2	0.2

Tabela após Discretize:

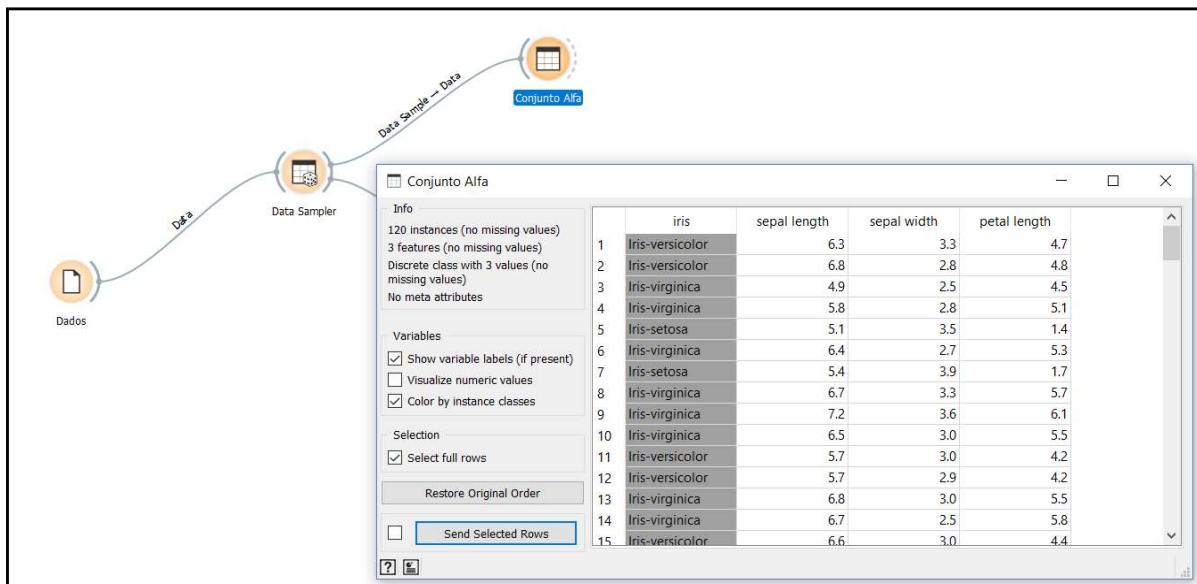
	iris	sepal length	sepal width	petal length	petal width
1	Iris-setosa	< 5.6	≥ 3.4	< 2.4	< 0.8
2	Iris-setosa	< 5.6	3 - 3.4	< 2.4	< 0.8
3	Iris-setosa	< 5.6	3 - 3.4	< 2.4	< 0.8
4	Iris-setosa	< 5.6	3 - 3.4	< 2.4	< 0.8
5	Iris-setosa	< 5.6	≥ 3.4	< 2.4	< 0.8
6	Iris-setosa	< 5.6	≥ 3.4	< 2.4	< 0.8
7	Iris-setosa	< 5.6	≥ 3.4	< 2.4	< 0.8
8	Iris-setosa	< 5.6	≥ 3.4	< 2.4	< 0.8
9	Iris-setosa	< 5.6	< 3	< 2.4	< 0.8
10	Iris-setosa	< 5.6	3 - 3.4	< 2.4	< 0.8
11	Iris-setosa	< 5.6	≥ 3.4	< 2.4	< 0.8
12	Iris-setosa	< 5.6	≥ 3.4	< 2.4	< 0.8
13	Iris-setosa	< 5.6	3 - 3.4	< 2.4	< 0.8
14	Iris-setosa	< 5.6	3 - 3.4	< 2.4	< 0.8
15	Iris-setosa	5.6 - 6.7	≥ 3.4	> 2.4	< 0.8

1.9 Data Sampler Widget

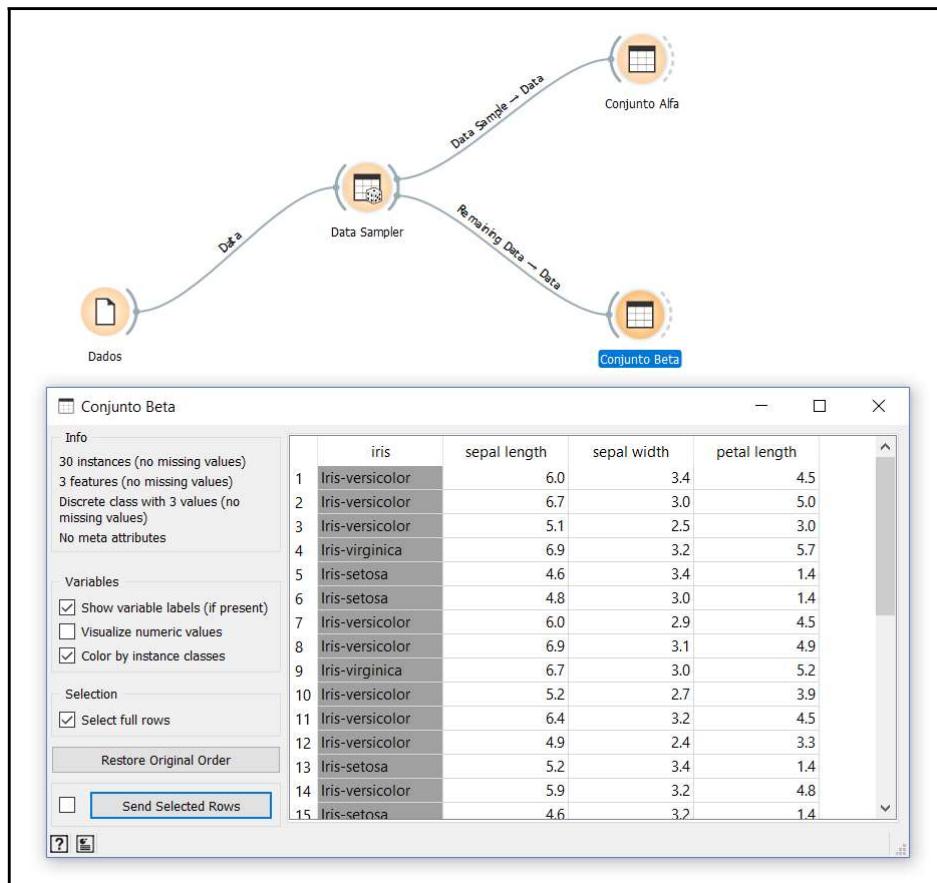
Podemos selecionar conjuntos de amostras, de forma aleatória, com *Data Sampler Widget*. Isto é muito útil para formar conjuntos de dados para o treinamento e testes de classificadores e regressores.



Aqui solicitamos que 80% dos dados, de forma aleatória, formem o Conjunto Alfa (120 Amostras).

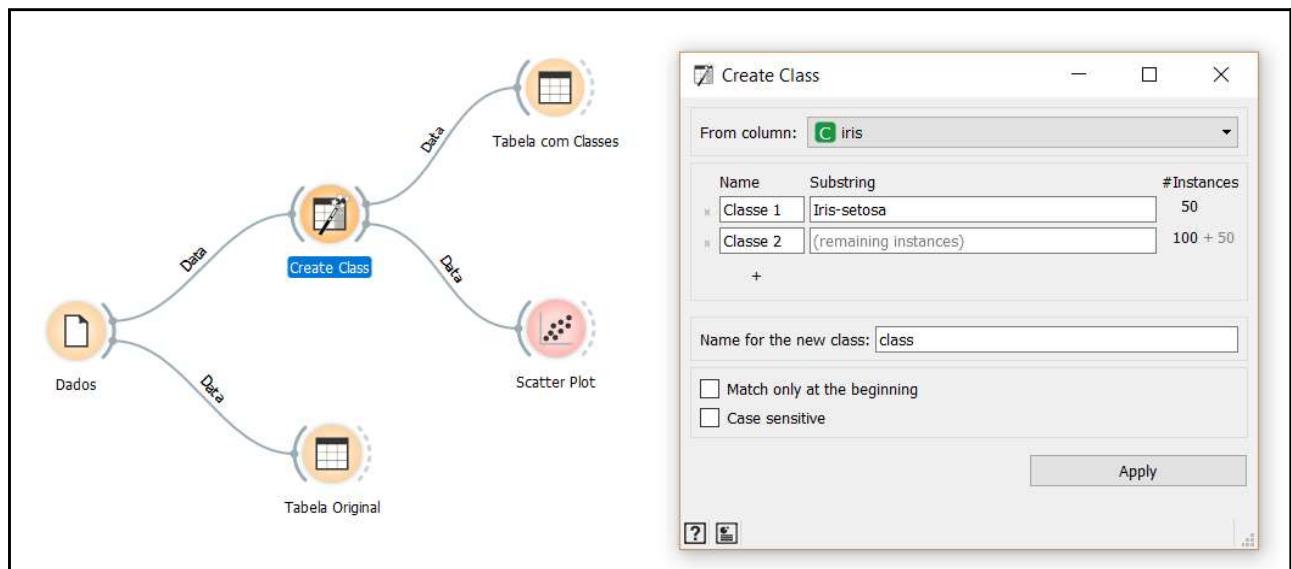


Os dados restantes formam o Conjunto Beta (30 Amostras).

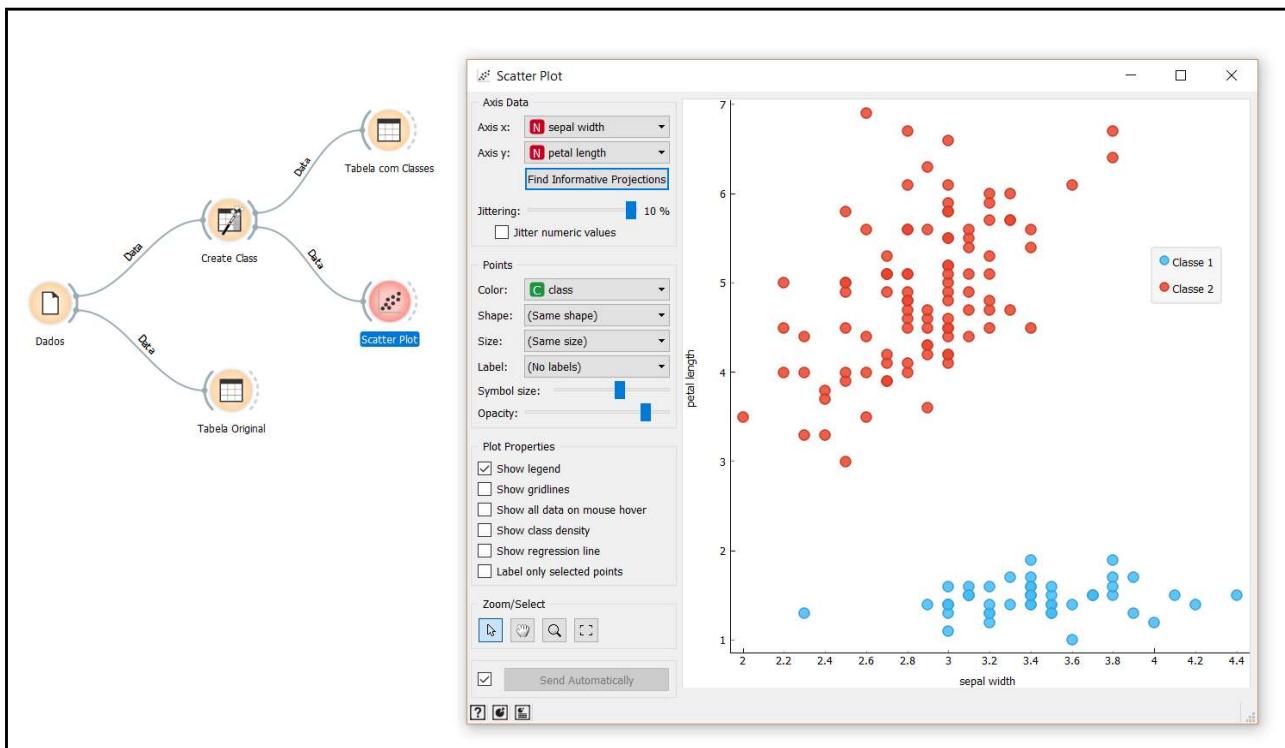


1.10 Create Class Widget

Podemos criar/agrupar classes com *Create Class Widget*. Para o exemplo a seguir, vamos criar duas classes (Classe 1 e Classe 2) para os “Dados Íris”.

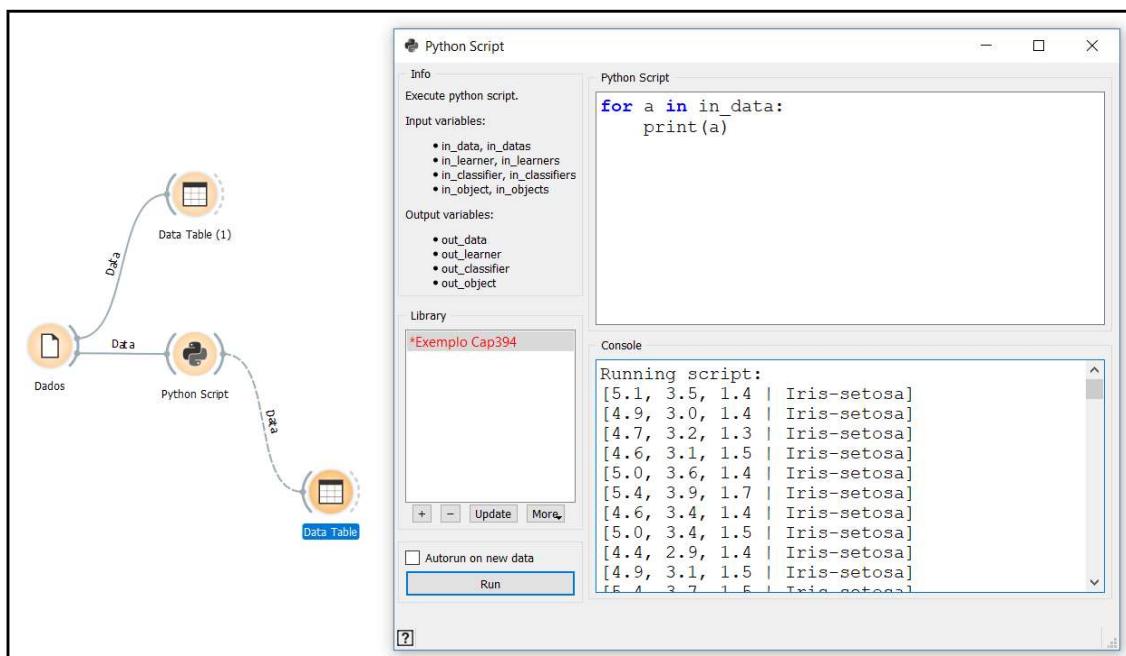


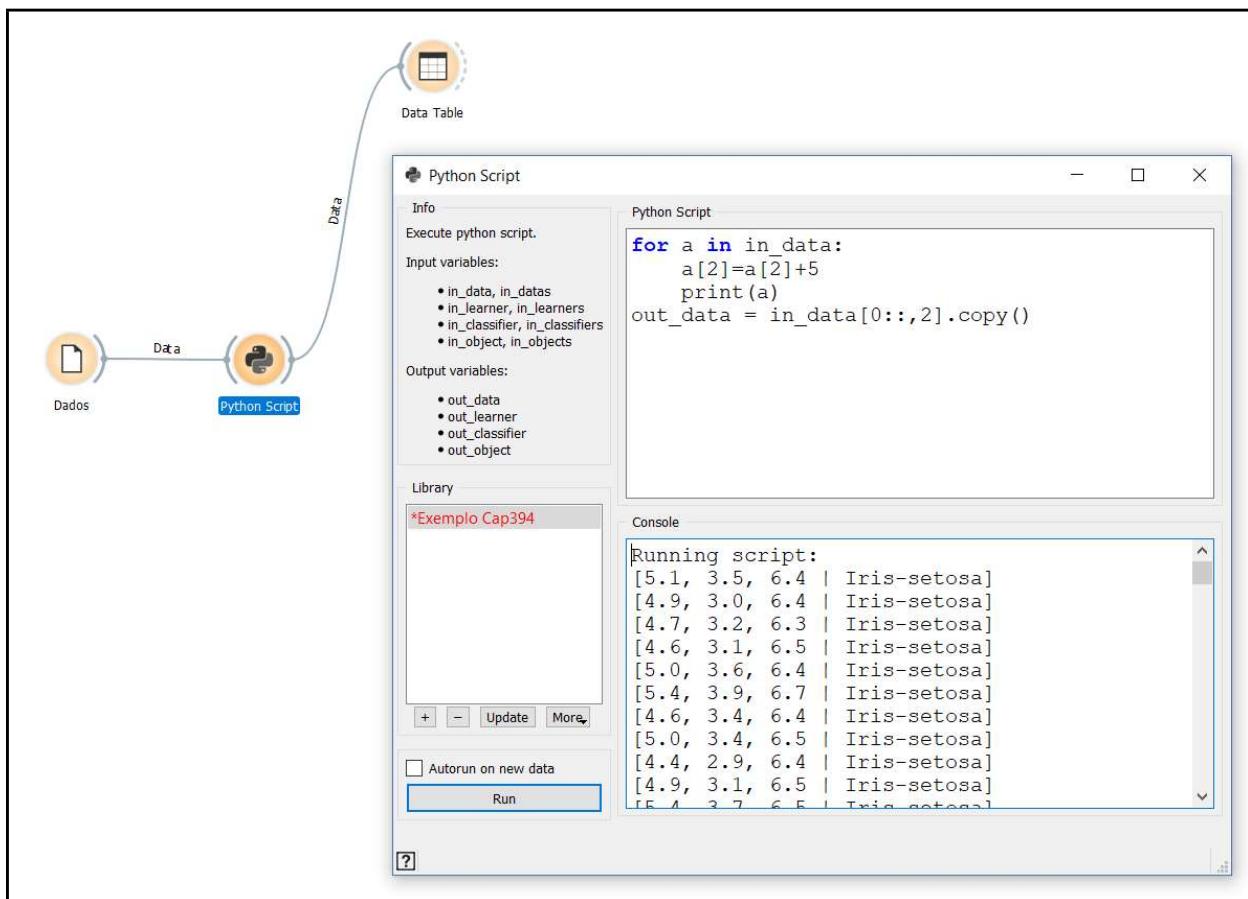
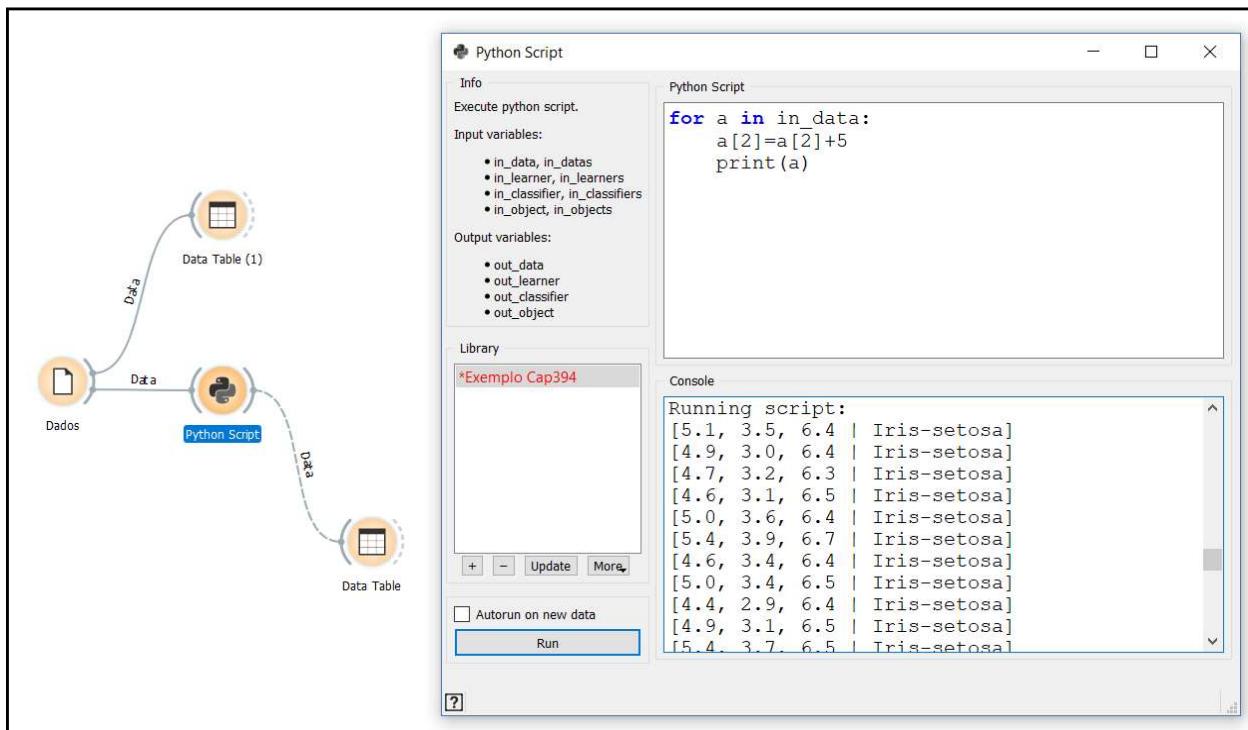
As amostras “Íris-Setosa” formam a Classe 1. As demais amostras (“Íris-Versicolor” e “Íris-Virgínica”) formam a Classe 2.

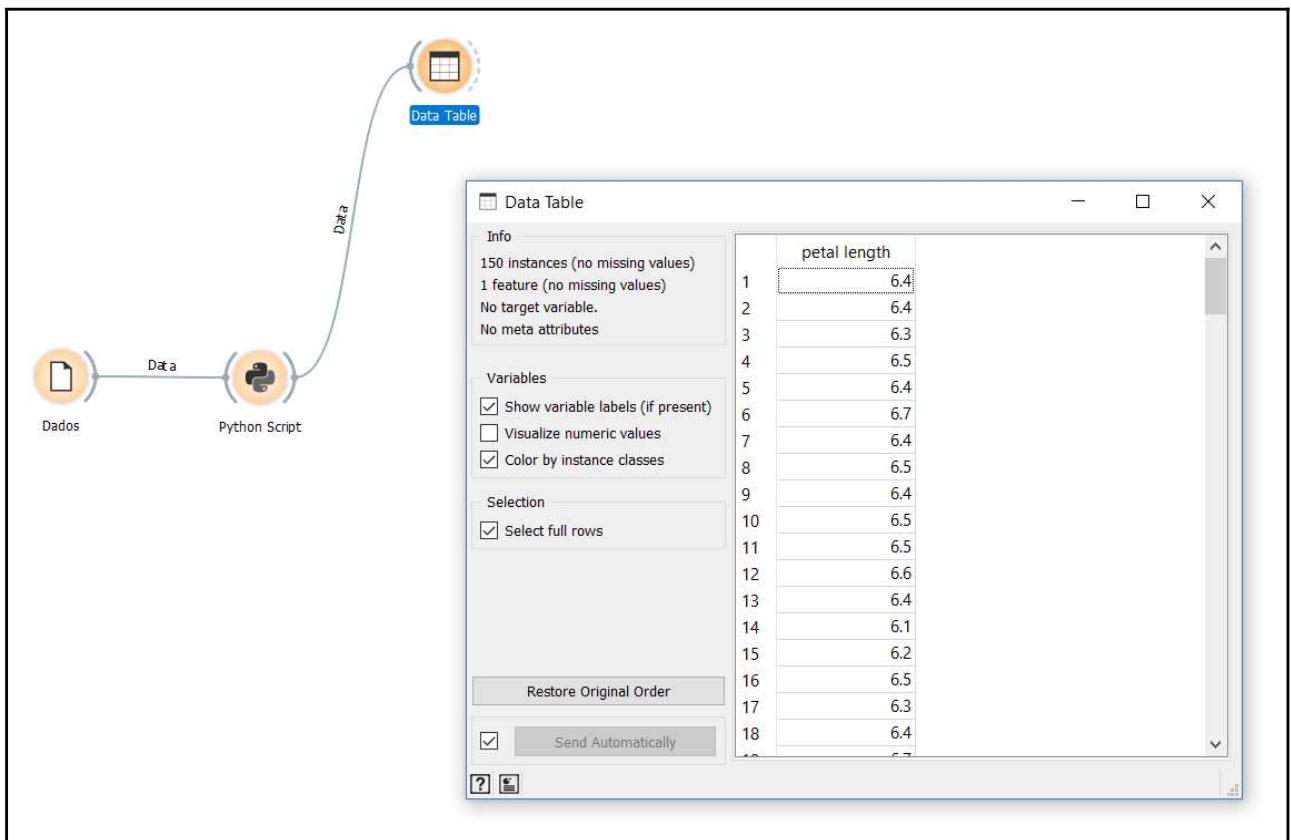


1.11 Python Script Widget

Você pode estende funcionalidades mediante *Python Script Widget*. Por exemplo, digamos que você precisa obter todos os valores do Atributo “petal length” acrescentadas de um valor constante $\alpha=5$.

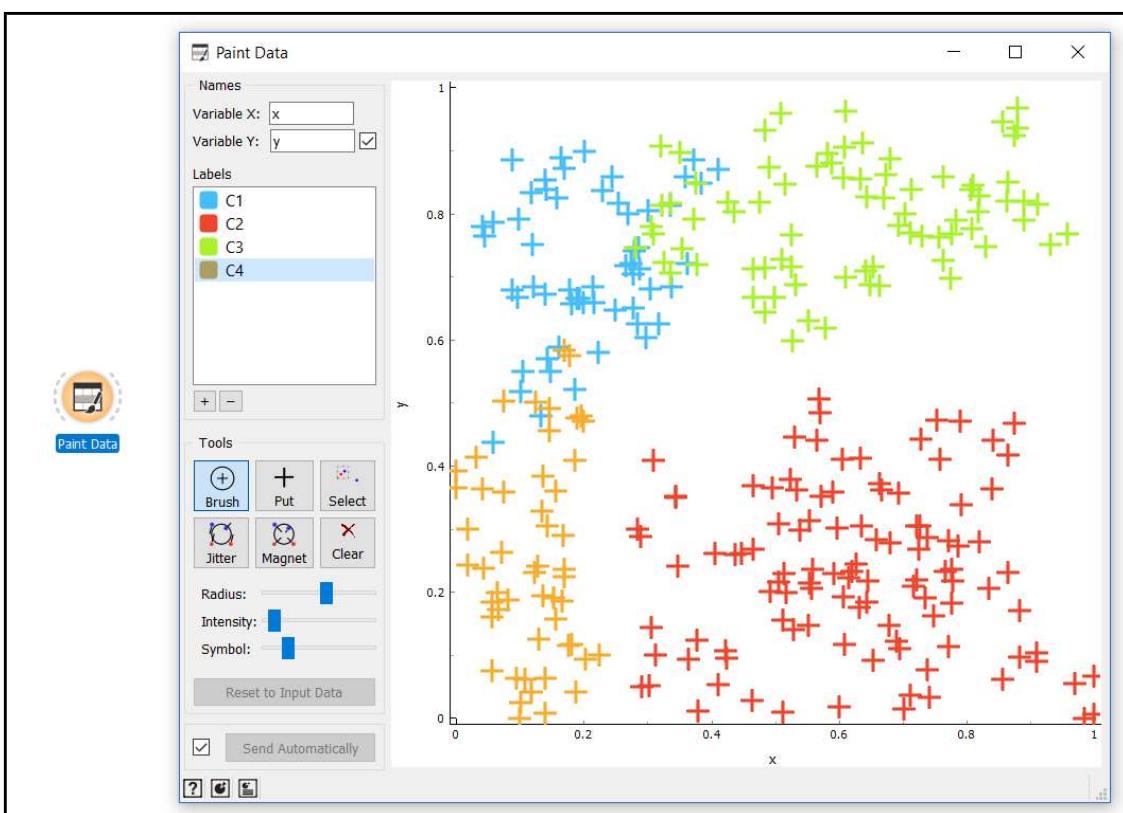






1.12 Paint Widget

Você pode criar uma amostragem aleatória para testes com o *Paint Widget*.



Data Table

Info
293 instances (no missing values)
2 features (no missing values)
Discrete class with 4 values (no missing values)
No meta attributes

Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

Selection
 Select full rows

Restore Original Order
 Send Selected Rows

	Class	x	y
235	C3	0.730	0.765
236	C3	0.654	0.717
237	C3	0.712	0.770
238	C3	0.763	0.726
239	C3	0.958	0.769
240	C3	0.890	0.821
241	C3	0.912	0.815
242	C3	0.931	0.751
243	C4	0.042	0.363
244	C4	0.019	0.243
245	C4	0.081	0.187
246	C4	0.064	0.200
247	C4	0.056	0.160
248	C4	0.140	0.064
249	C4	0.109	0.061
...
249	C4	0.110	0.061

1.13 Rank Widget

Você pode classificar atributos utilizando o *Rank Widget*.

Rank

Scoring Methods

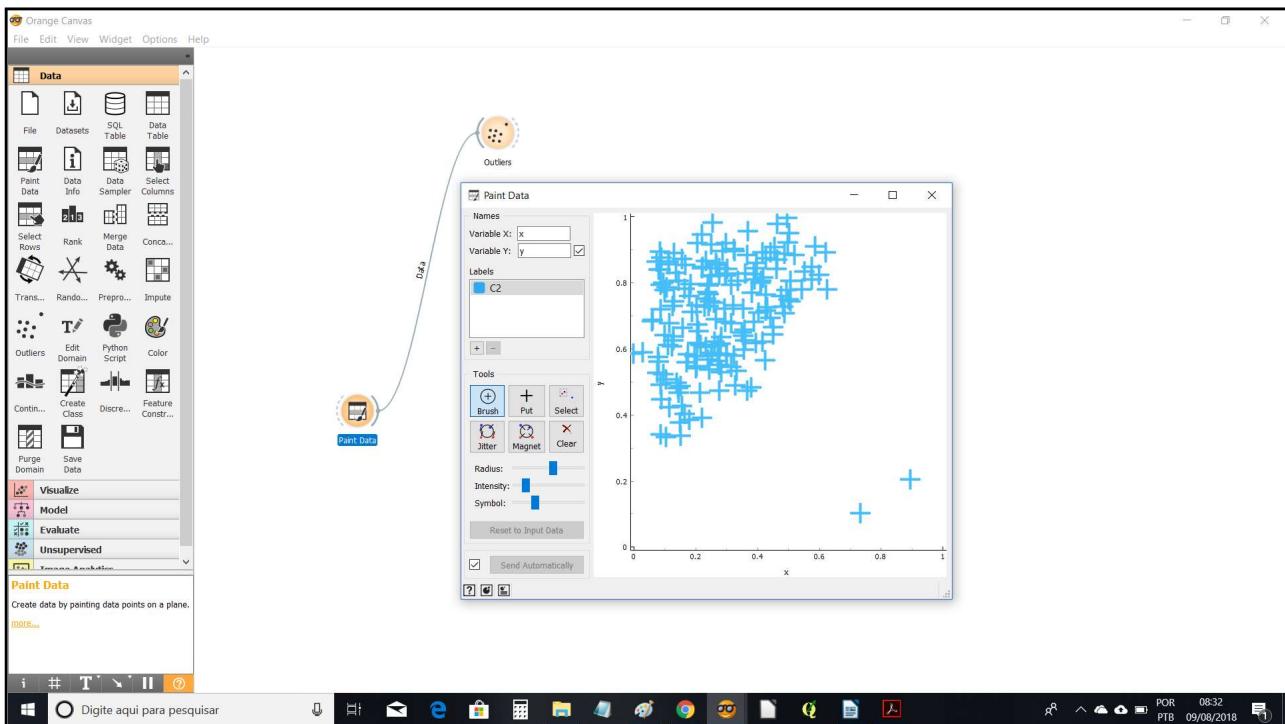
- Information Gain
- Information Gain Ratio
- Gini Decrease
- ANOVA
- χ^2
- ReliefF
- FCBF

Select Attributes

- None
- All
- Manual
- Best ranked: 1

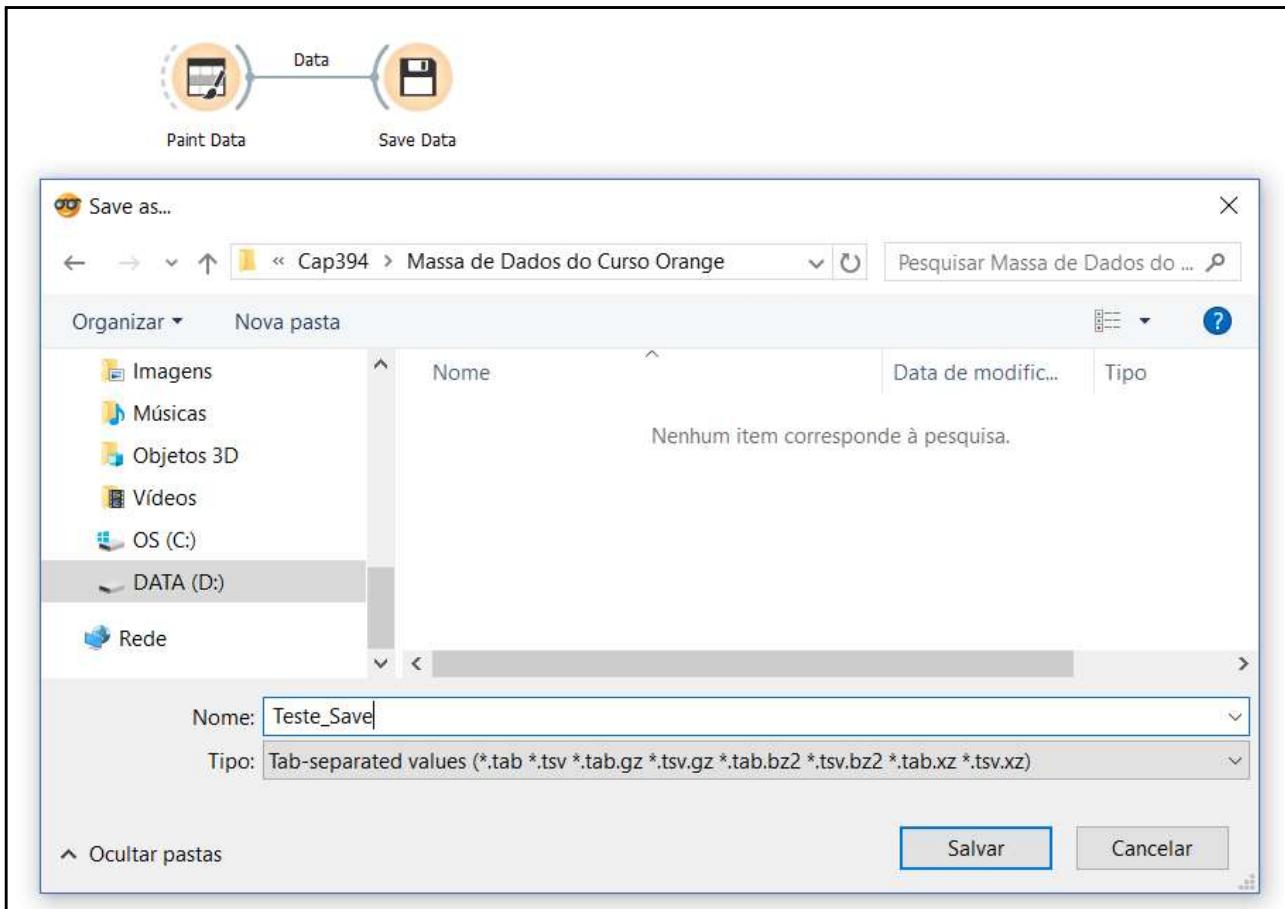
Send Automatically

#	Info. gain	Gain ratio	Gini	ANOVA	χ^2
N petal length	1.086	0.544	0.423	1179.034	98.946
N petal width	1.059	0.532	0.407	959.324	94.162
N sepal length	0.624	0.313	0.247	119.265	79.243
N sepal width	0.361	0.183	0.154	47.364	50.082



1.14 Save Widget

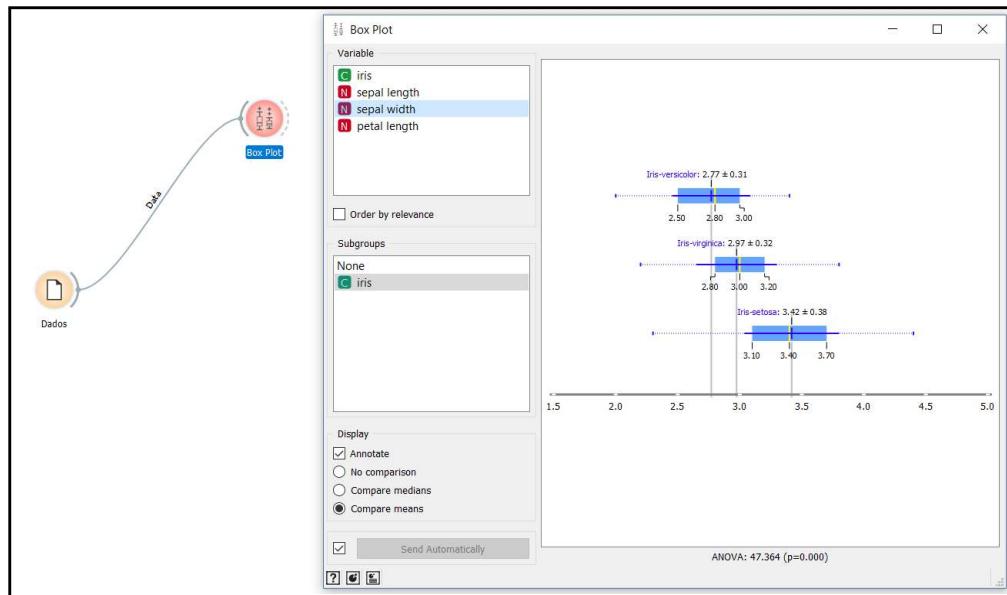
Você pode salvar dados com o *Save Widget*.



2 VISUALIZAÇÃO DE DADOS

Você pode realizar a visualização dos dados com várias ferramentas do Orange.

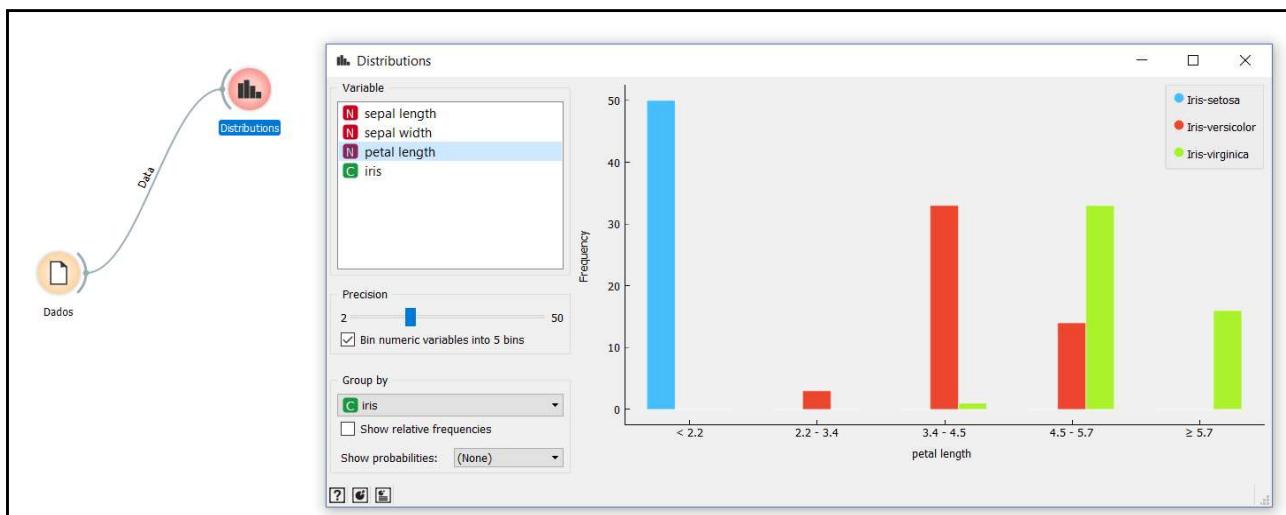
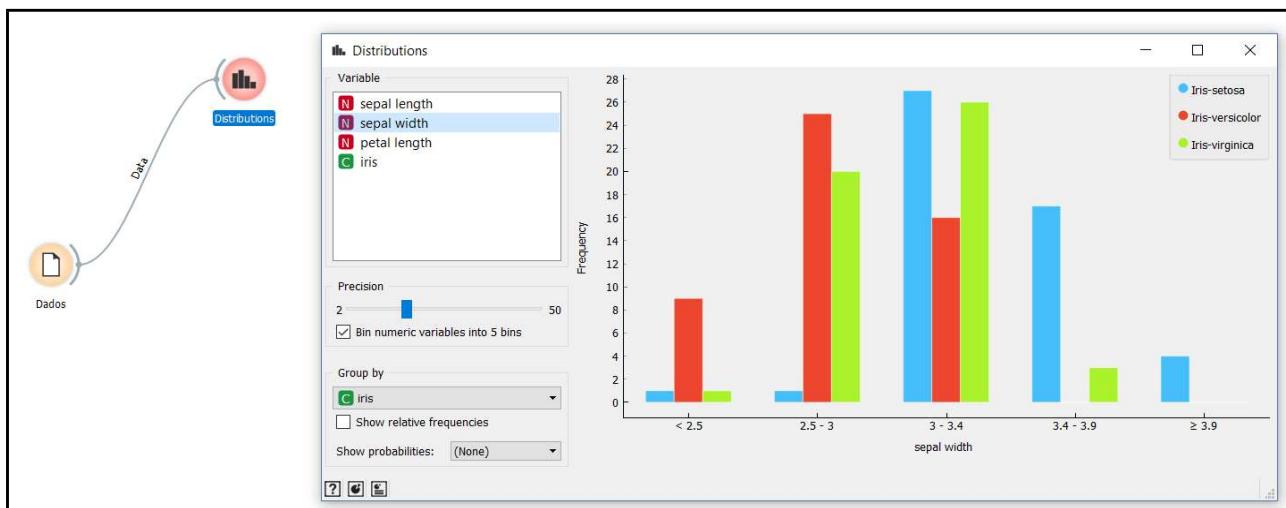
2.1 Box Plot Widget



Para atributos contínuos, o Box Plot Widget exibe:

- A média (a linha vertical azul escura)
- Valores limite para o desvio padrão da média. A área destacada em azul é o desvio padrão inteiro da média.
- A mediana (linha vertical amarela). A fina linha azul representa a área entre o primeiro (25%) e o terceiro (75%) quartil, enquanto a linha pontilhada fina representa o intervalo inteiro de valores (do menor para o maior valor no conjunto de dados para o parâmetro selecionado).

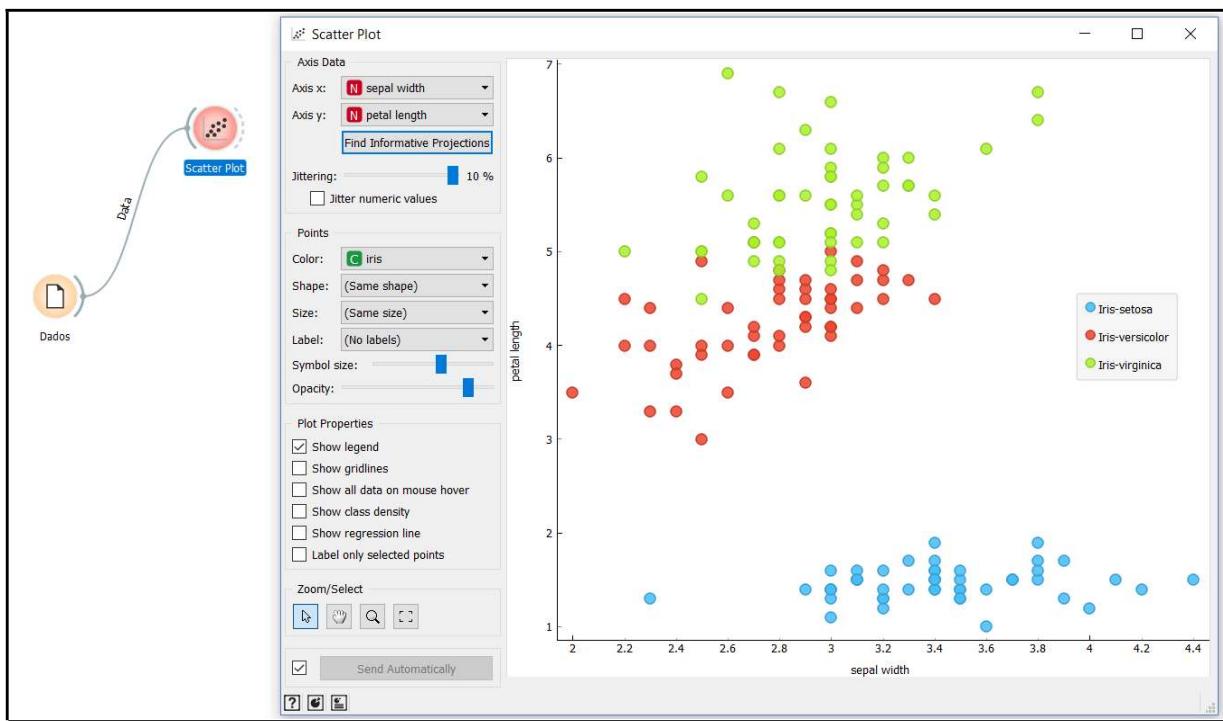
2.2 Distributions Widget



Atenção!

Observe que *Distributions Widget* permite visualizar que o atributo “petal length” possui valores bem distintos para a classe “Íris-Setosa” em relação as demais classes (Dados Íris).

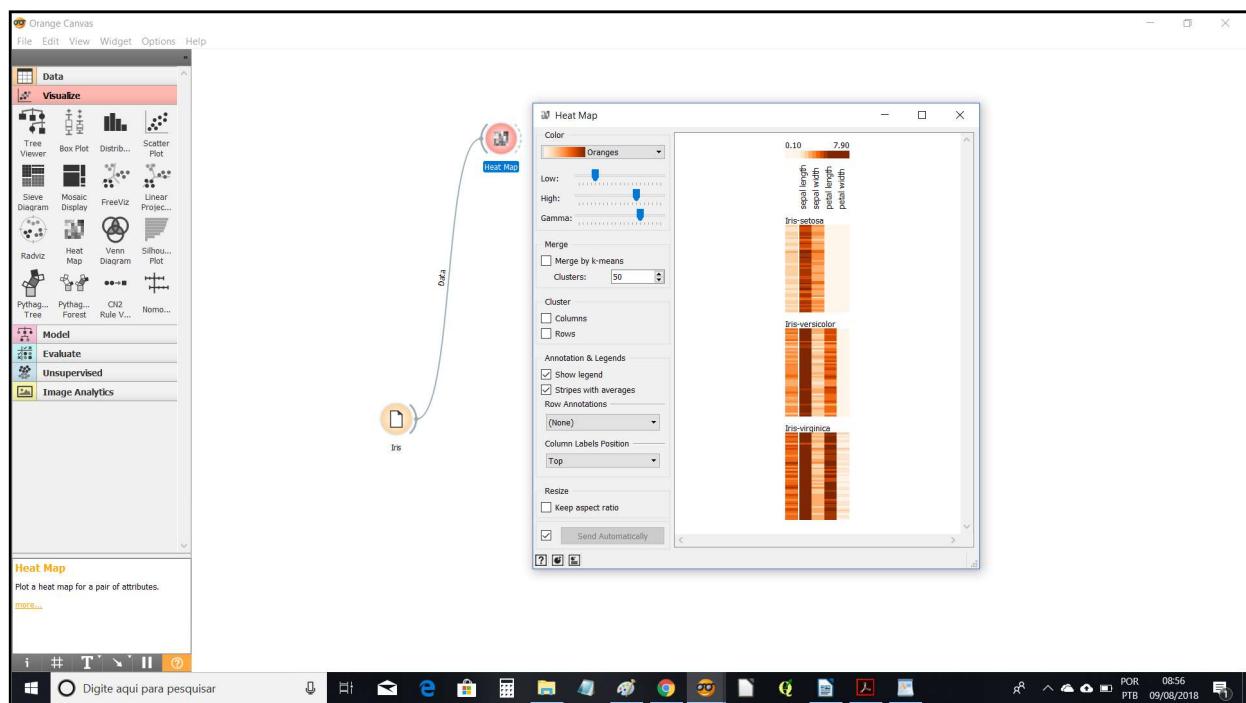
2.3 Scatter Plot Widget



Podemos observar de forma semelhante a distinção entre os atributos com *Scatter Plot Widget*.

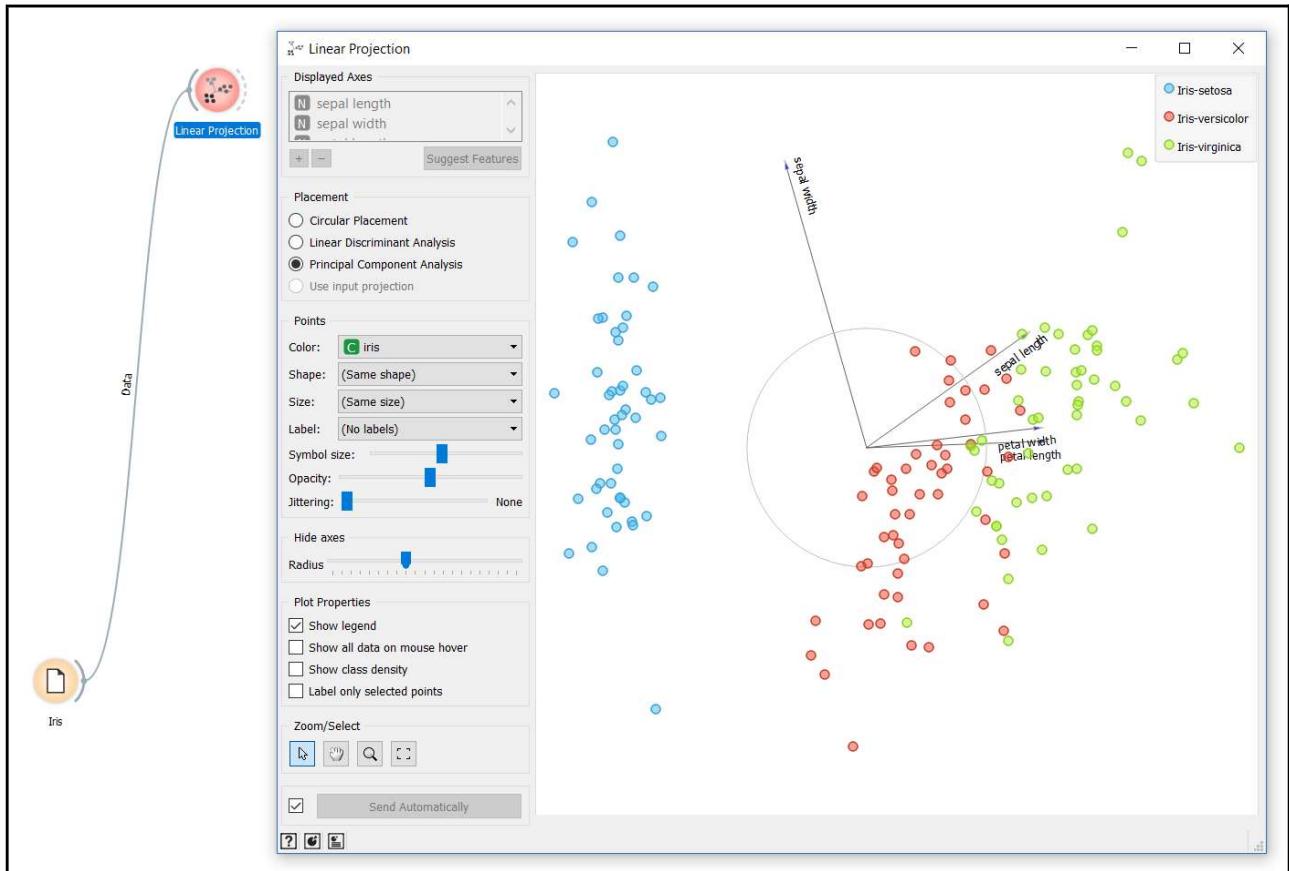
2.4 Heat Map Widget

Podemos observar diferenças significativas nos atributos com o *Heat Map Widget*.



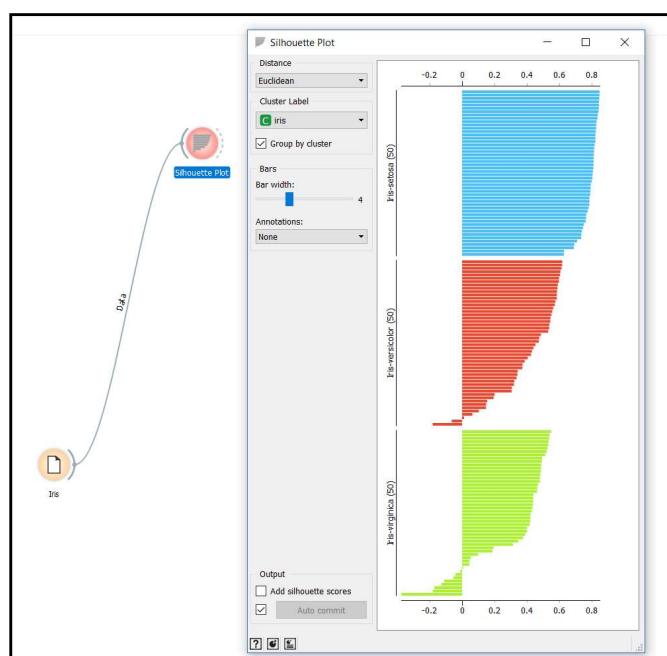
2.5 Linear Projection Widget

Podemos realizar projeção linear para análise exploratória de dados com o *Linear Projection Widget*.



2.6 Silhouette Plot Widget

Podemos analisar a consistência em agrupamentos com o *Silhouette Plot Widget*.

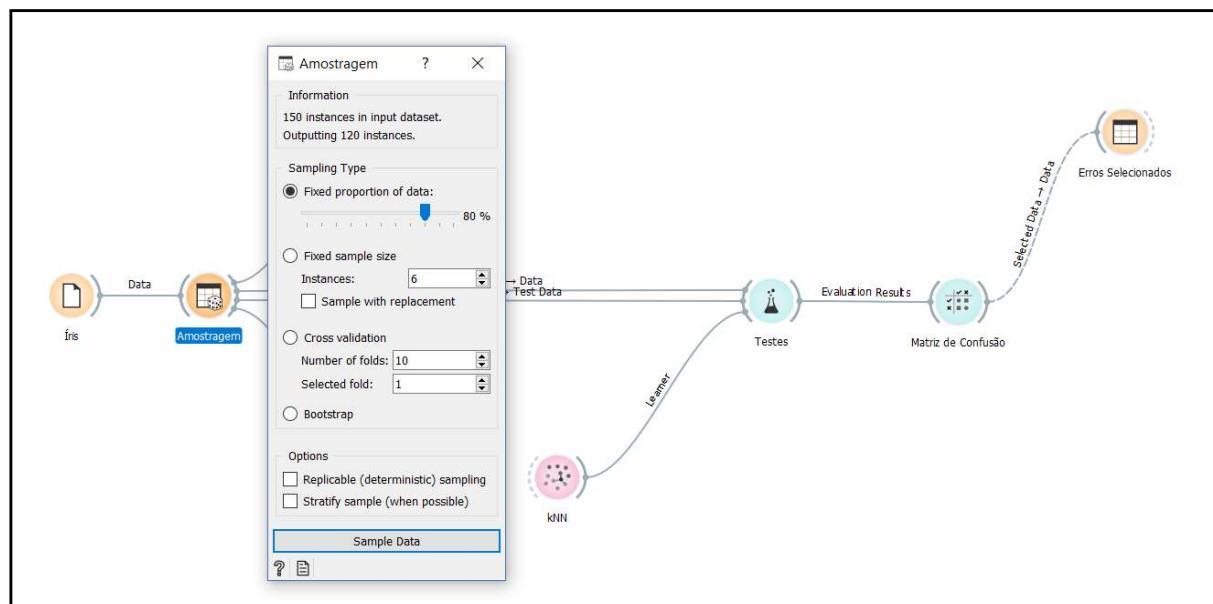
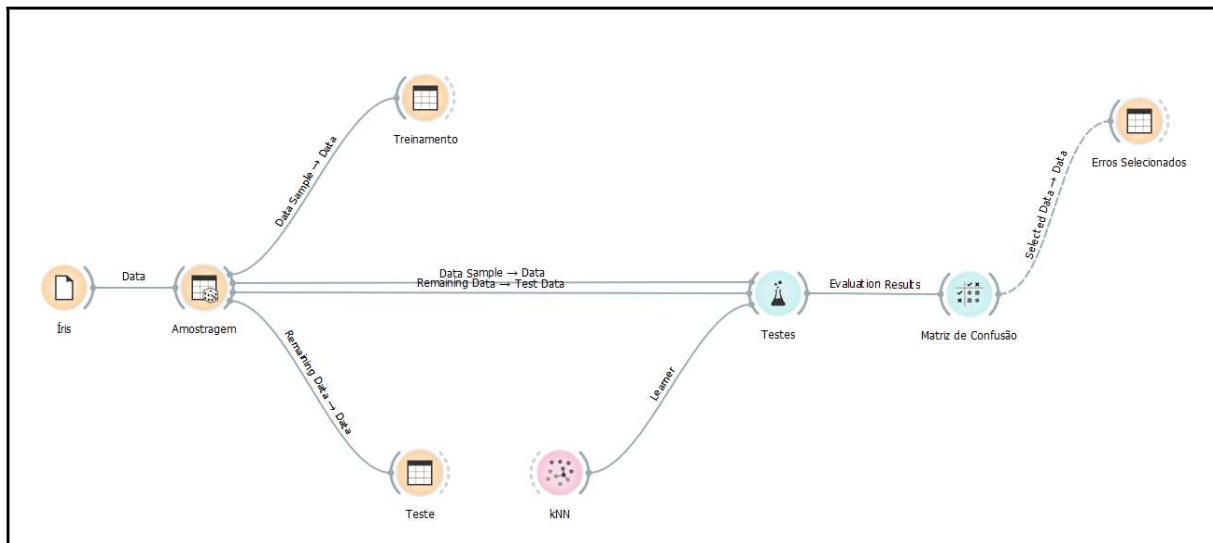


3 APRENDIZAGEM SUPERVISIONADA

A Aprendizagem Supervisionada (*Supervised Learning*) é o termo usado sempre que o algoritmo é “treinado” sobre um conjunto de dados pré-definido (elemento supervisor), ou seja, com base em dados pré-definidos, o programa pode realizar previsões (classificações ou regressões) quando receber novos dados. Você pode utilizar vários métodos de classificação e regressão com os *widgets* de Orange.

3.1 k-NN Widget

Como este *widget* você pode prever de acordo com as instâncias de treinamento mais próximas. Por exemplo, pretende-se criar um Classificador k-NN (com $k = 5$ vizinhos mais próximos e métrica = distância euclidiana) para os “Dados Íris”. Os dados serão divididos, aleatoriamente, em: dados de treinamento (80%) e dados de teste (20%).



Treinamento

Info
120 instances (no missing values)
4 features (no missing values)
Discrete class with 3 values (no missing values)
No meta attributes

Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

Selection
 Select full rows

Send Automatically

	iris	sepal length	sepal width	petal length	petal width
1	Iris-versicolor	5.6	2.9	3.6	1.3
2	Iris-virginica	6.7	2.5	5.8	1.8
3	Iris-setosa	4.4	3.0	1.3	0.2
4	Iris-versicolor	5.8	2.7	4.1	1.0
5	Iris-versicolor	5.5	2.4	3.7	1.0
6	Iris-versicolor	5.6	2.5	3.9	1.1
7	Iris-setosa	4.6	3.6	1.0	0.2
8	Iris-versicolor	7.0	3.2	4.7	1.4
9	Iris-versicolor	6.7	3.1	4.7	1.5
10	Iris-setosa	5.0	3.6	1.4	0.2
11	Iris-virginica	6.3	2.5	5.0	1.9
12	Iris-setosa	5.4	3.9	1.3	0.4
13	Iris-virginica	7.7	3.0	6.1	2.3
14	Iris-versicolor	5.4	3.0	4.5	1.5
15	Iris-versicolor	5.6	2.7	4.2	1.3

Teste

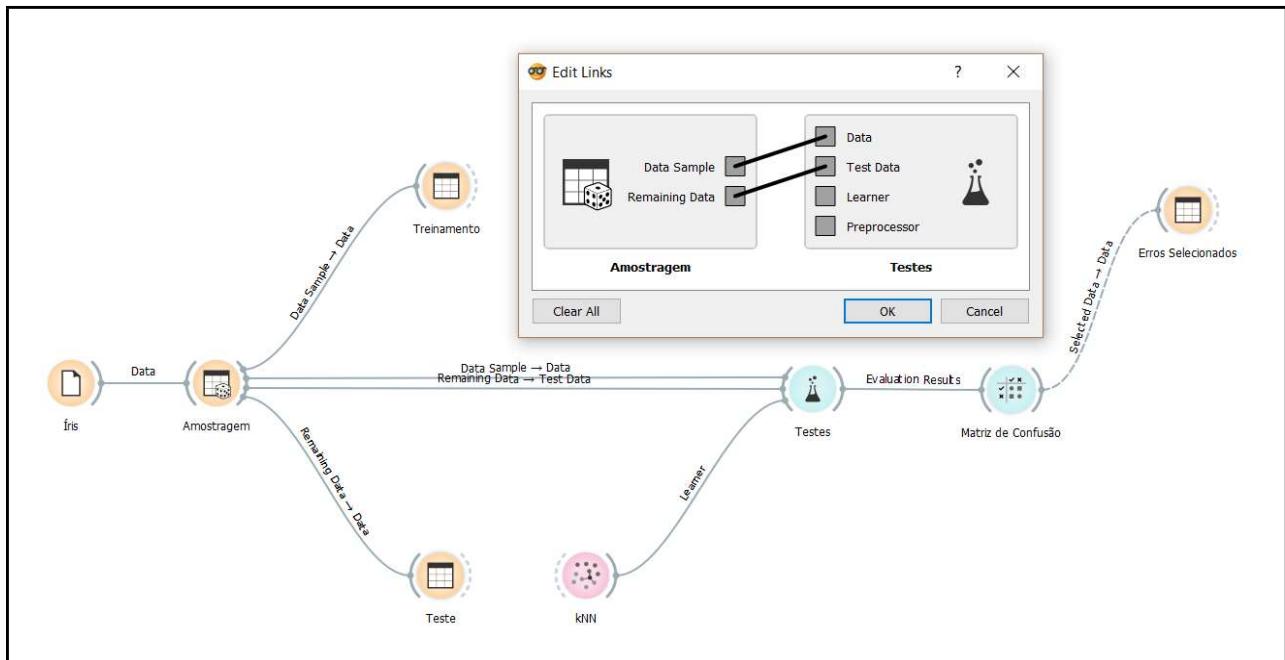
Info
30 instances (no missing values)
4 features (no missing values)
Discrete class with 3 values (no missing values)
No meta attributes

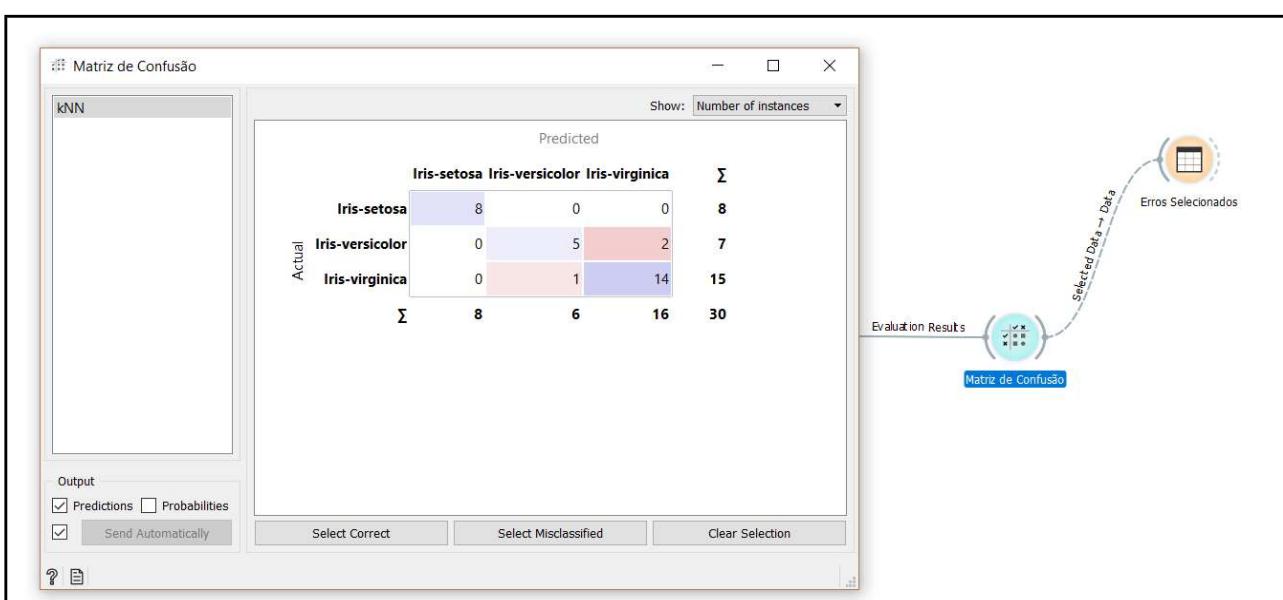
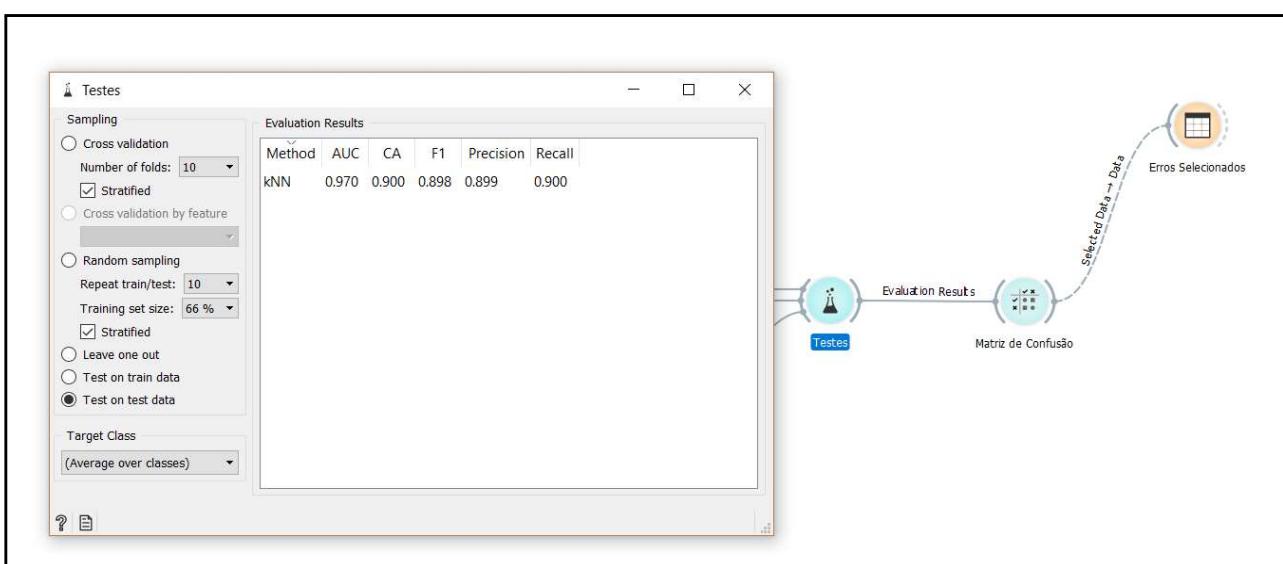
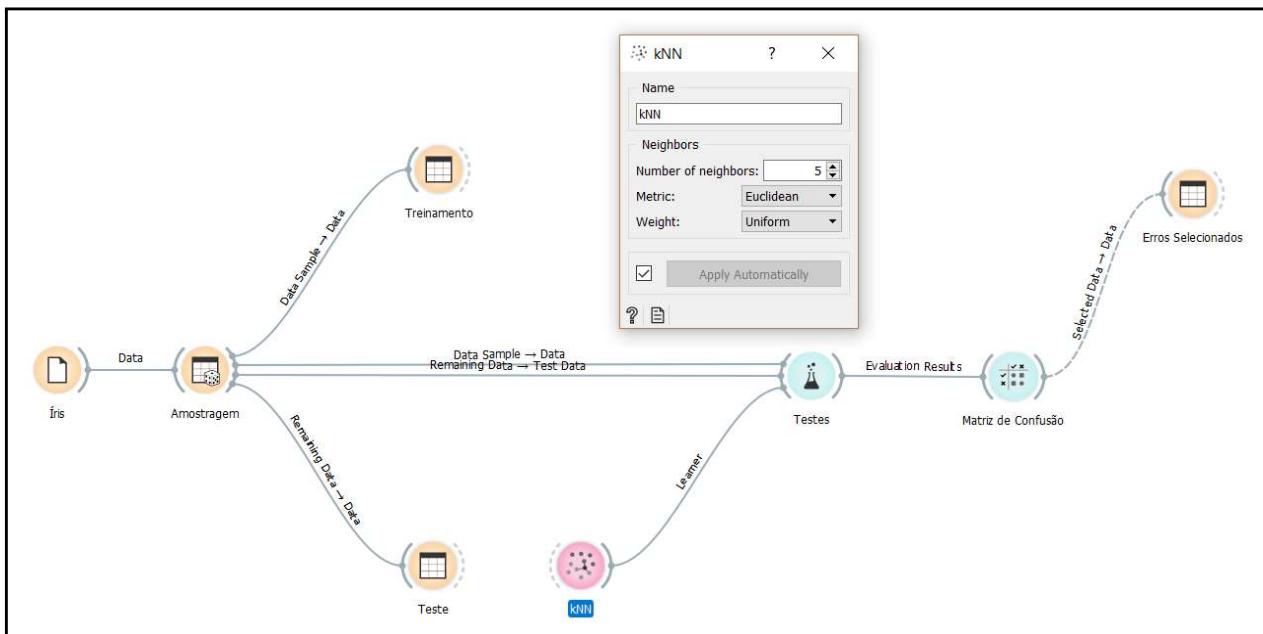
Variables
 Show variable labels (if present)
 Visualize numeric values
 Color by instance classes

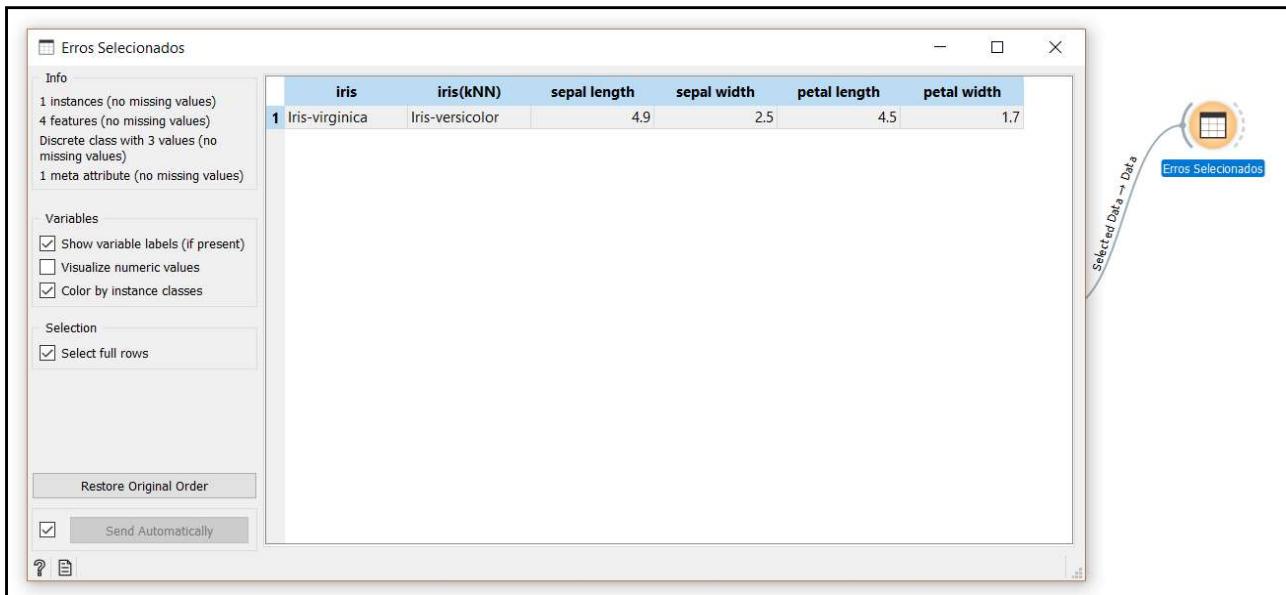
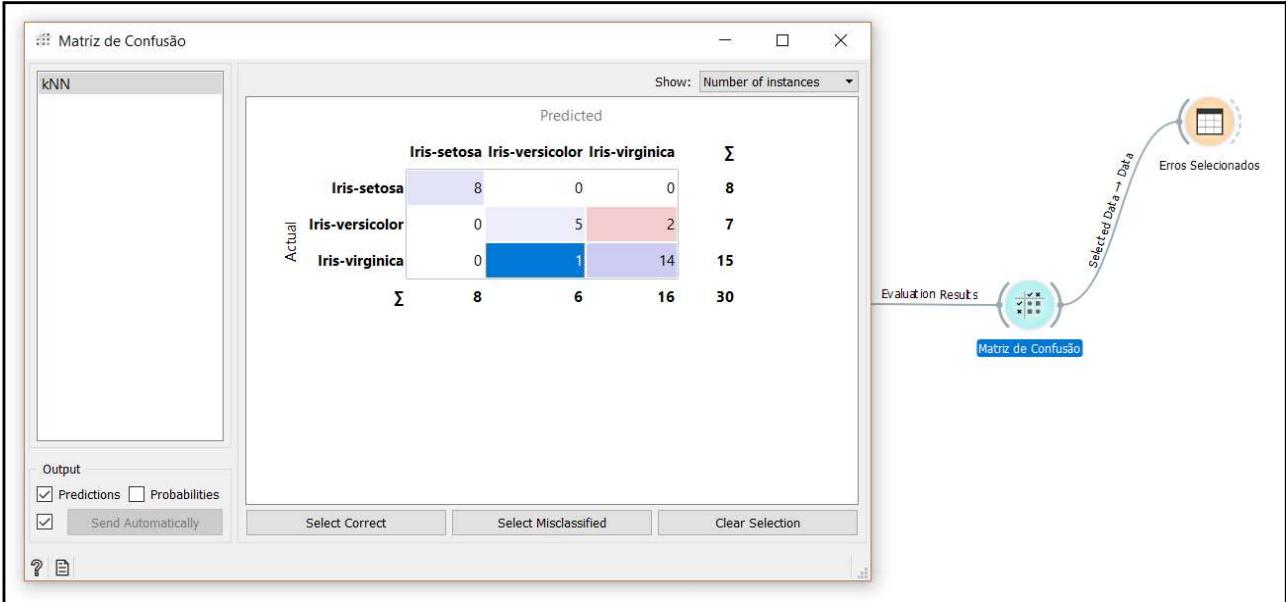
Selection
 Select full rows

Send Automatically

	iris	sepal length	sepal width	petal length	petal width
1	Iris-versicolor	6.5	2.8	4.6	1.5
2	Iris-virginica	6.4	2.8	5.6	2.1
3	Iris-setosa	4.9	3.0	1.4	0.2
4	Iris-setosa	5.4	3.4	1.5	0.4
5	Iris-virginica	6.7	3.3	5.7	2.5
6	Iris-virginica	6.9	3.1	5.4	2.1
7	Iris-virginica	6.7	3.0	5.2	2.3
8	Iris-virginica	4.9	2.5	4.5	1.7
9	Iris-versicolor	6.1	3.0	4.6	1.4
10	Iris-setosa	4.4	2.9	1.4	0.2
11	Iris-versicolor	6.3	2.5	4.9	1.5
12	Iris-virginica	7.1	3.0	5.9	2.1
13	Iris-setosa	5.7	3.8	1.7	0.3
14	Iris-virginica	6.2	3.4	5.4	2.3
15	Iris-versicolor	6.4	2.9	4.3	1.3

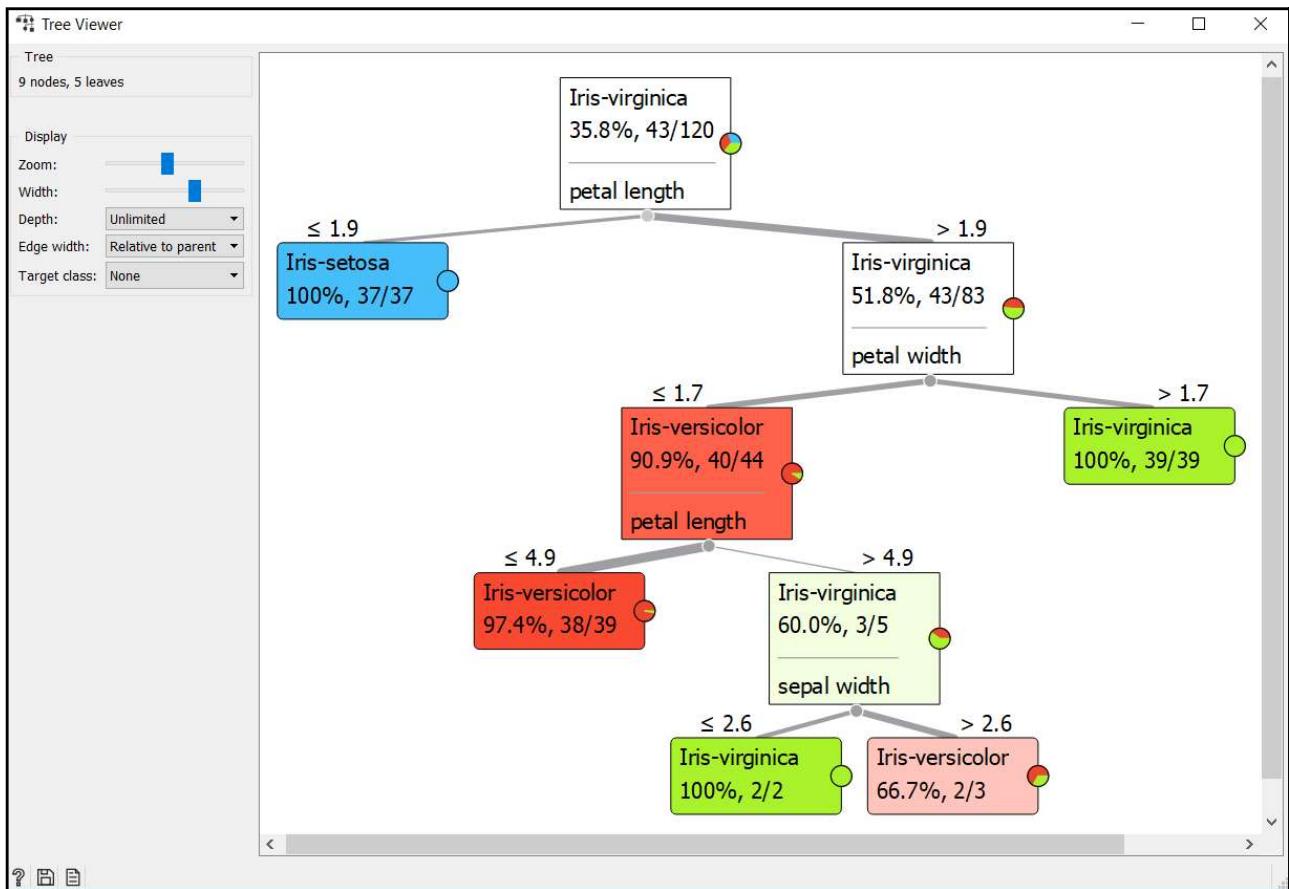
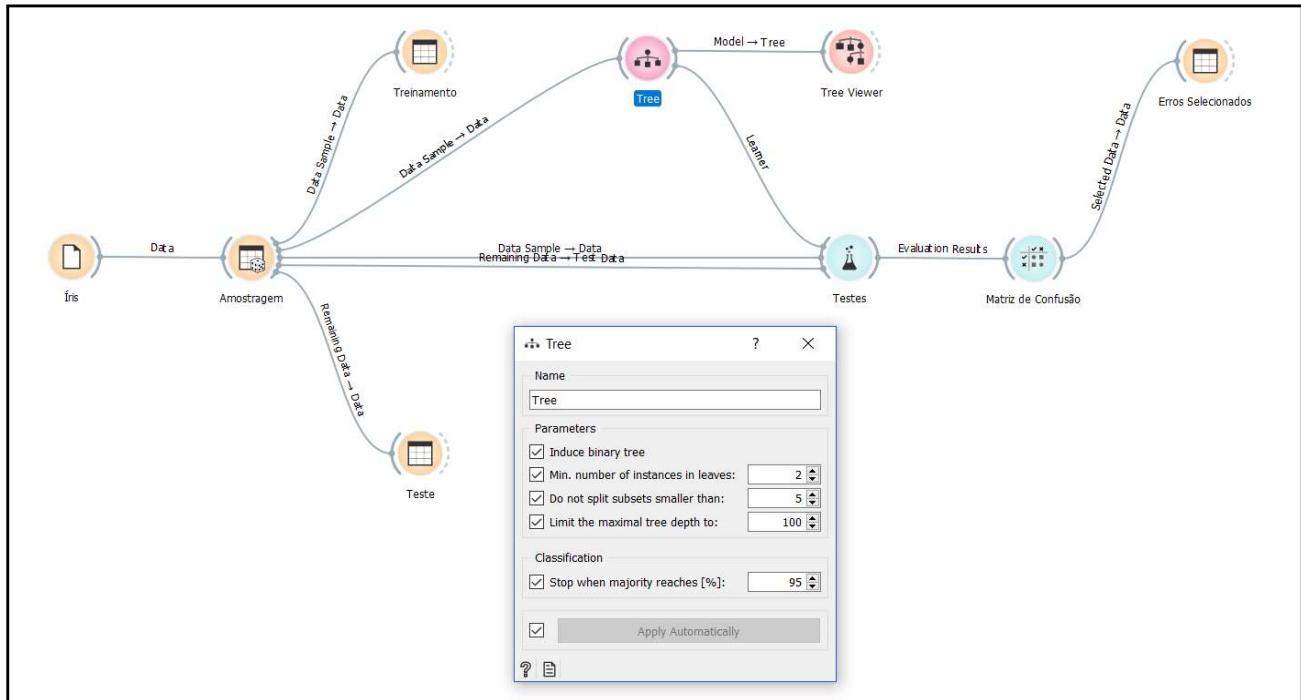


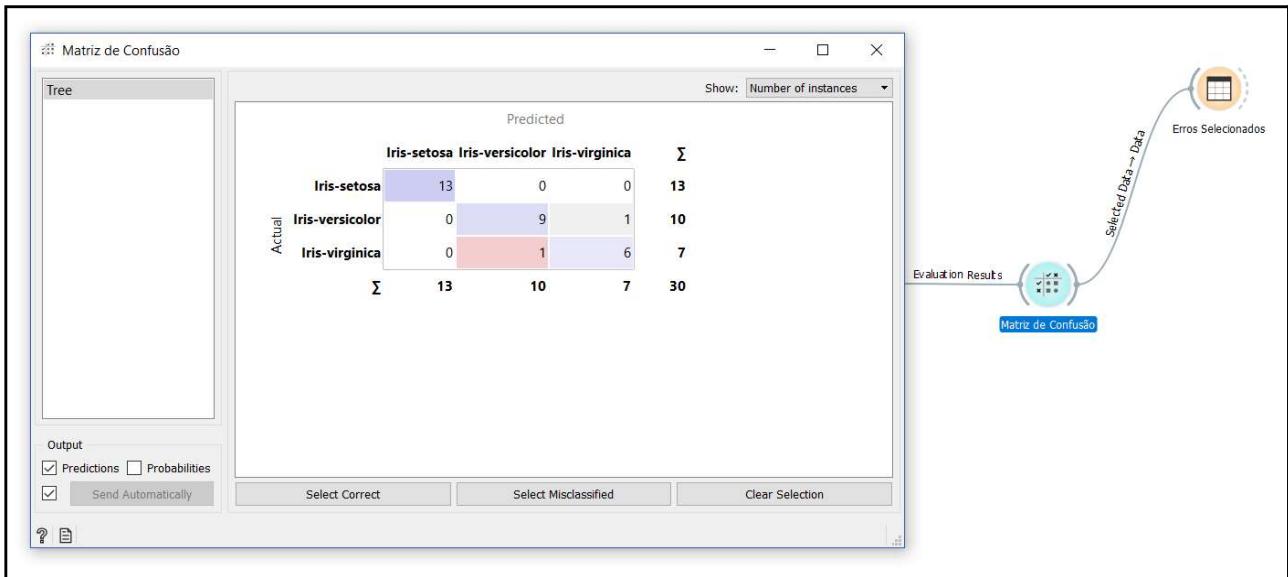
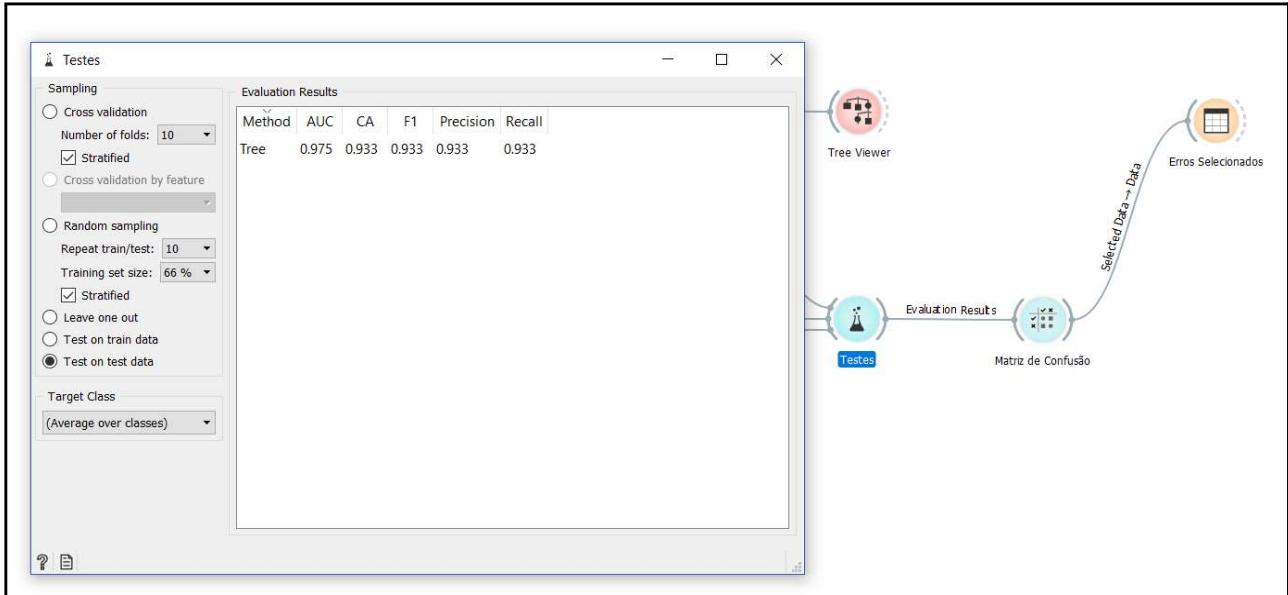




3.2 Tree Widget

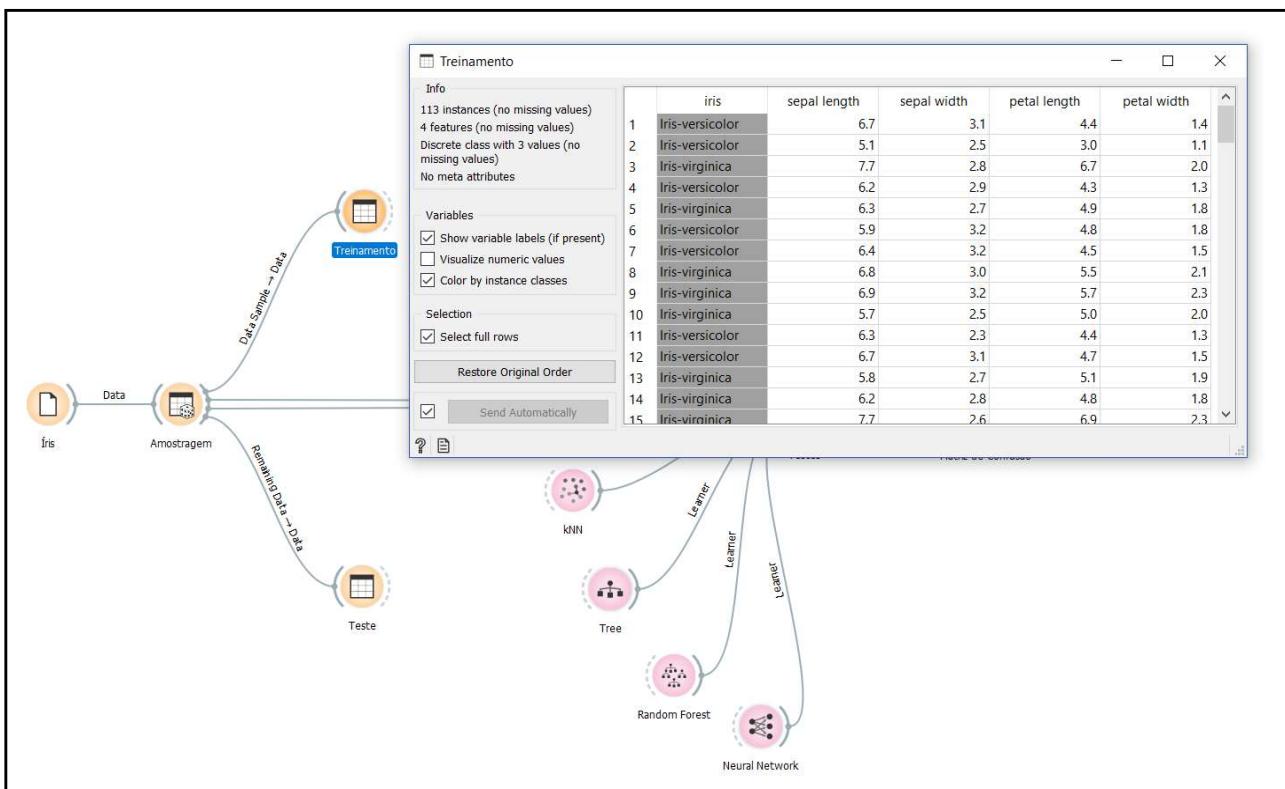
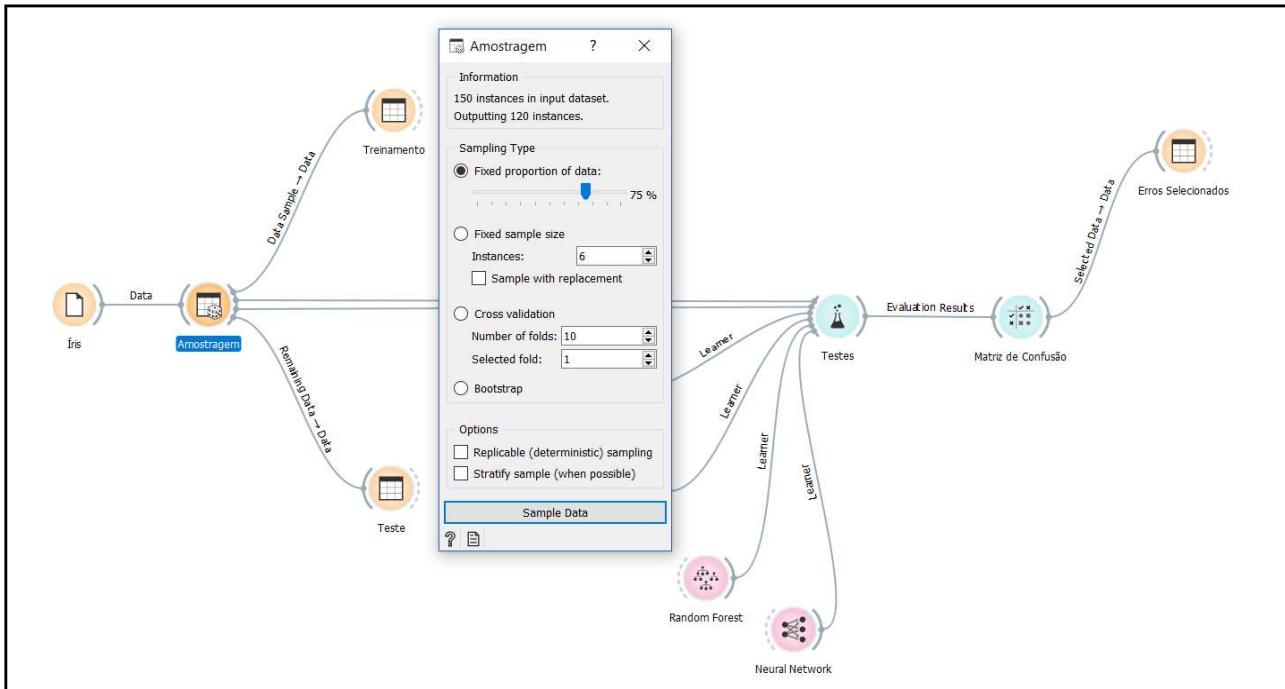
Tree (Árvore de Decisão) é um algoritmo que divide os dados em nós por pureza de classe. É um precursor da Random Forest. No Orange, o Tree Widget pode operar com conjuntos de dados discretos e contínuos. Por exemplo, pretende-se criar um Classificador Tree (utilizando as configurações padrões da Versão Orange 3.15) com os “Dados Íris”. Os dados, novamente, serão divididos de forma aleatória em: dados de treinamento (80%) e dados de teste (20%).

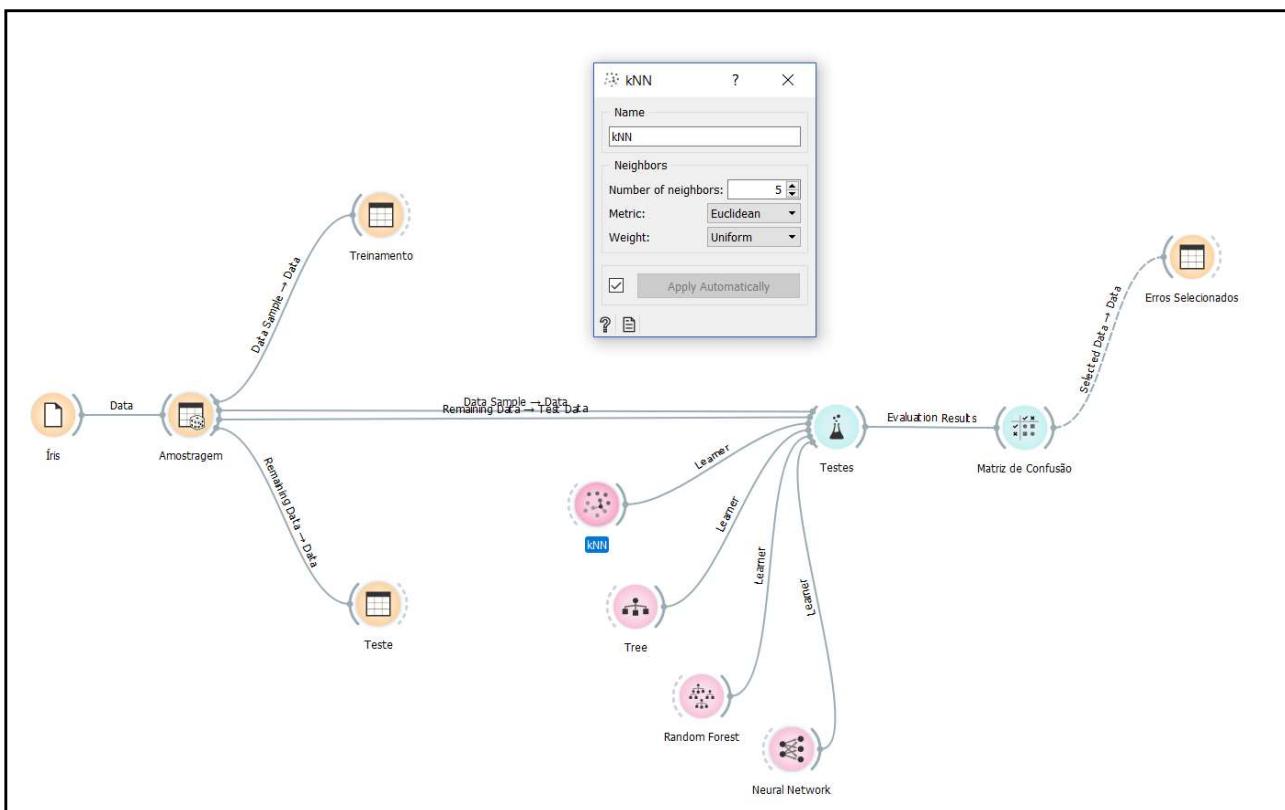
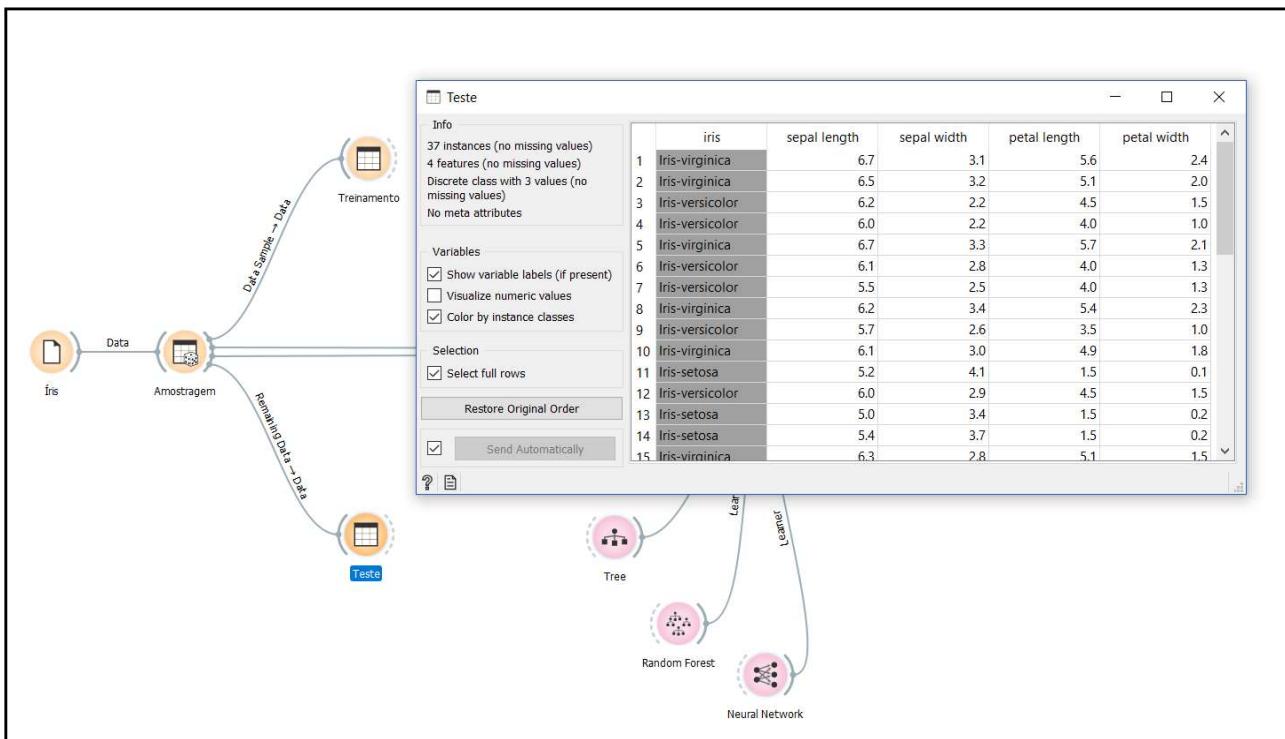


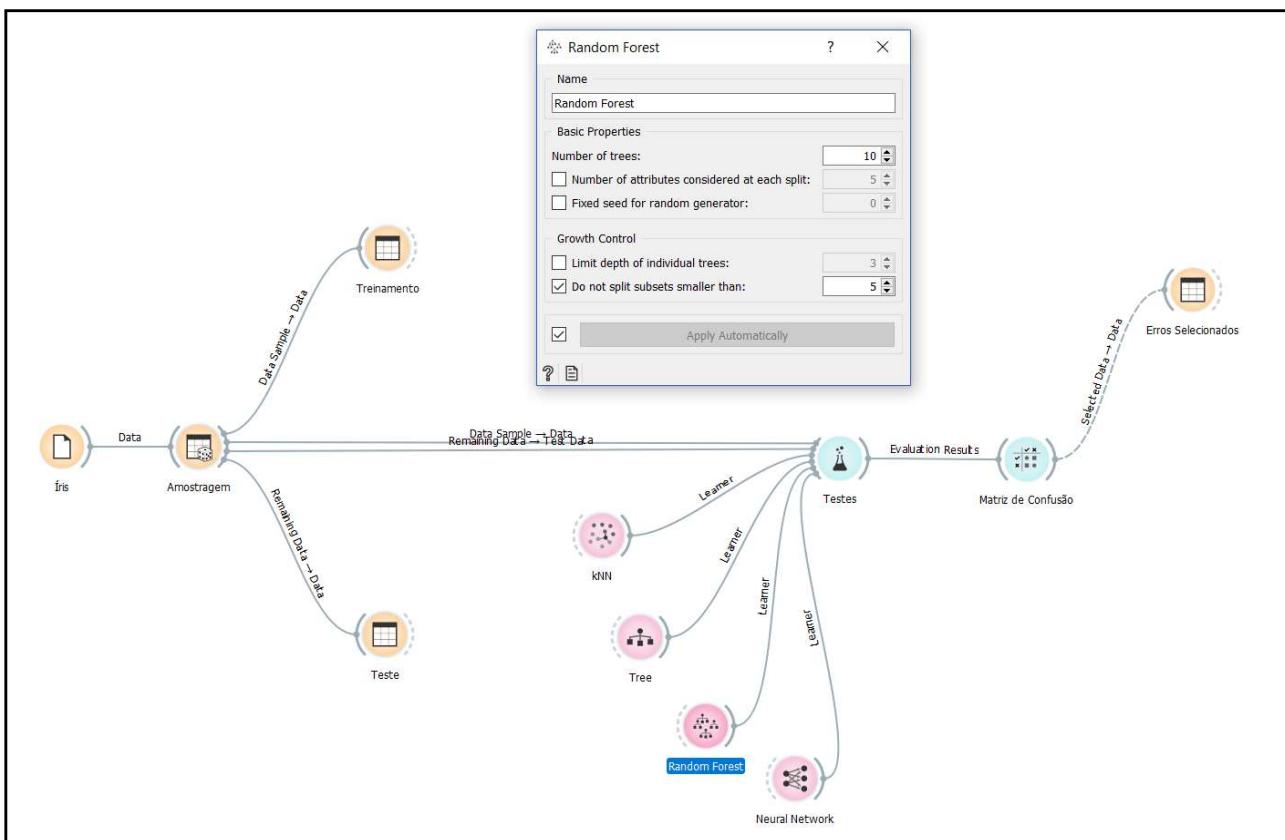
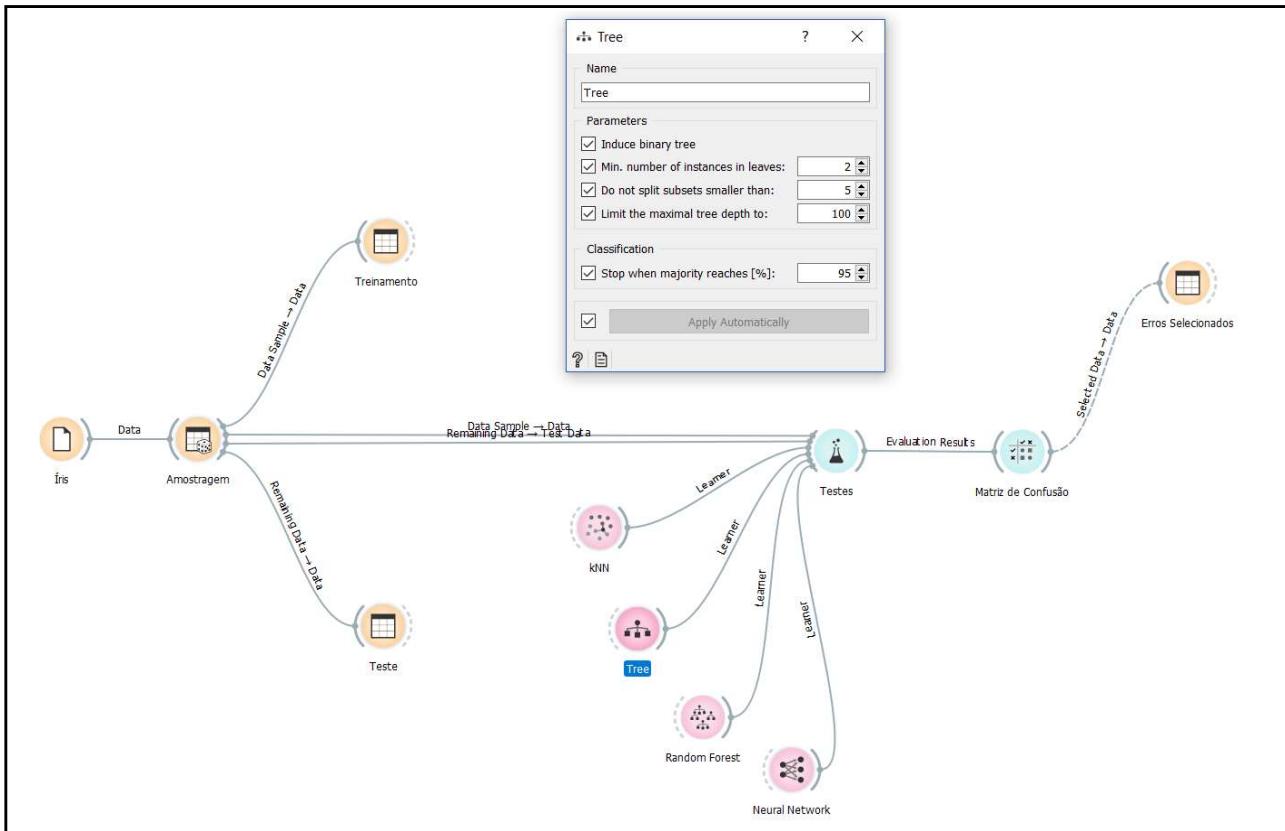


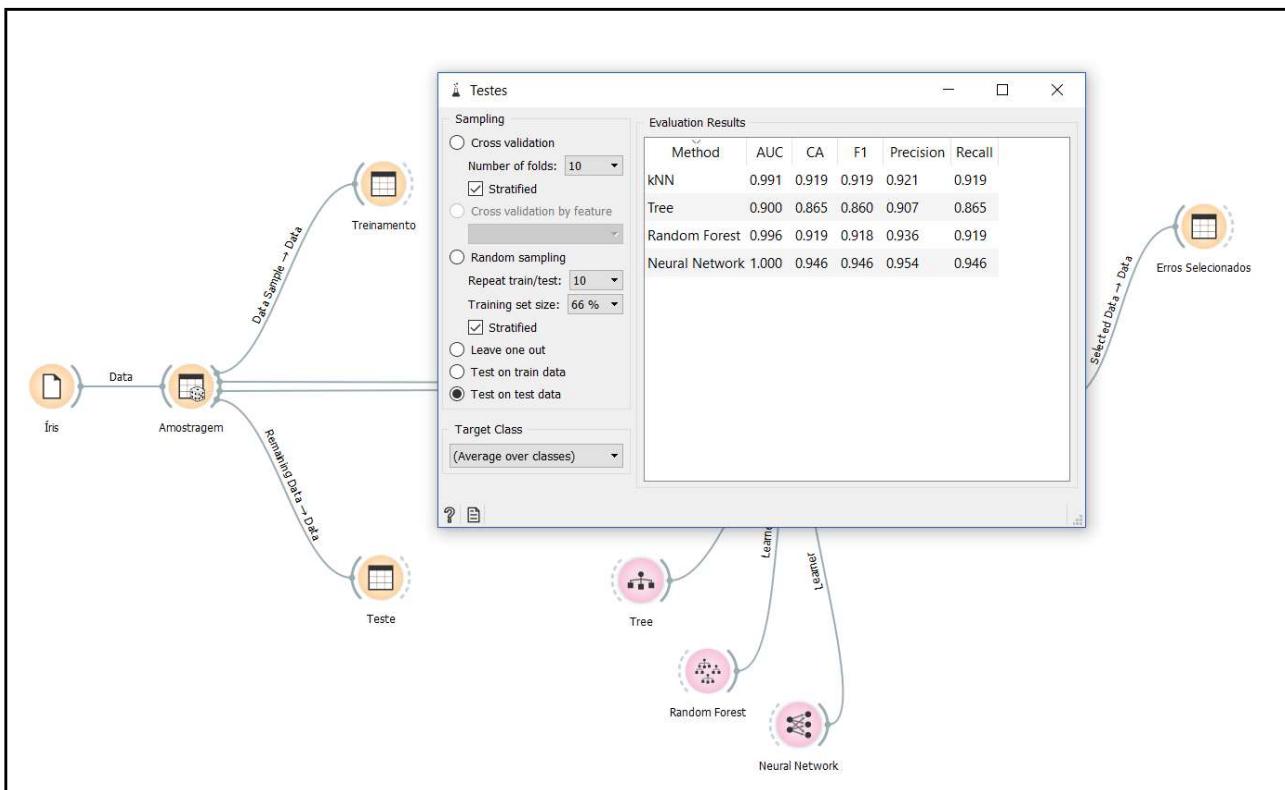
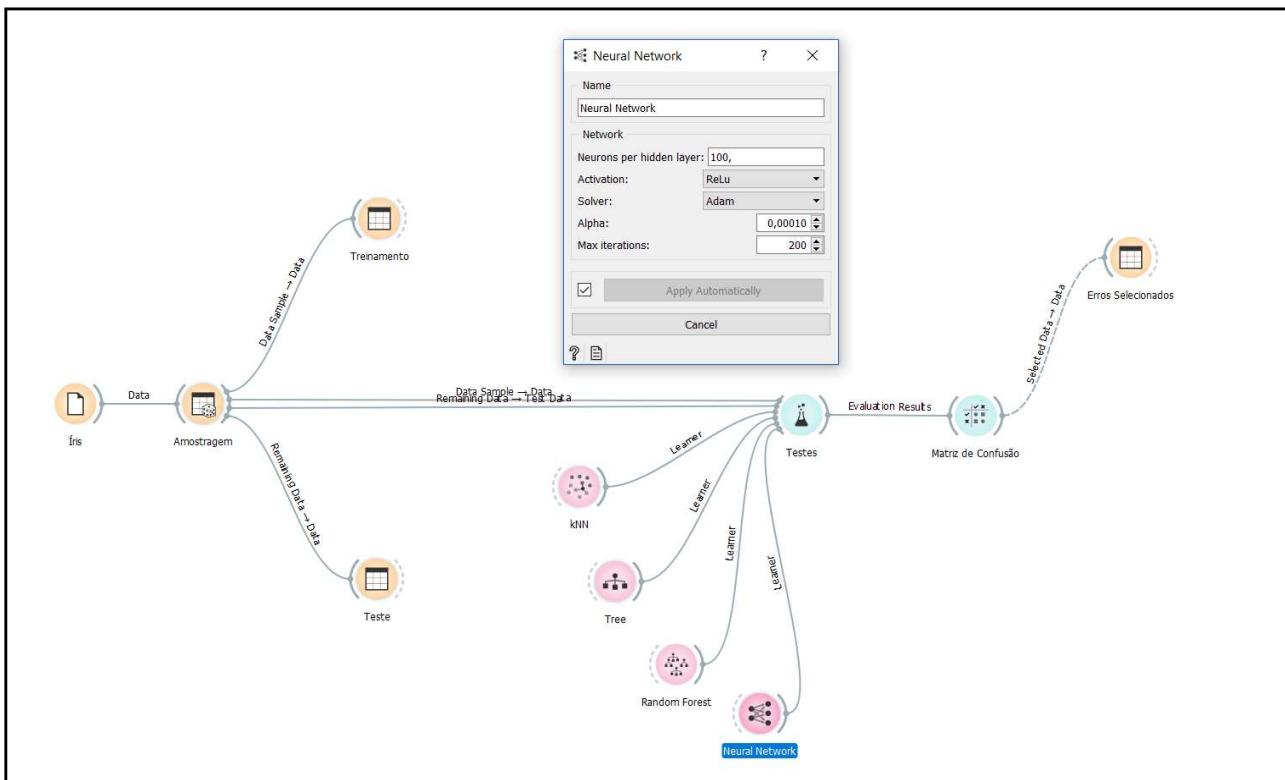
3.3 Comparação entre múltiplos classificadores e regressores

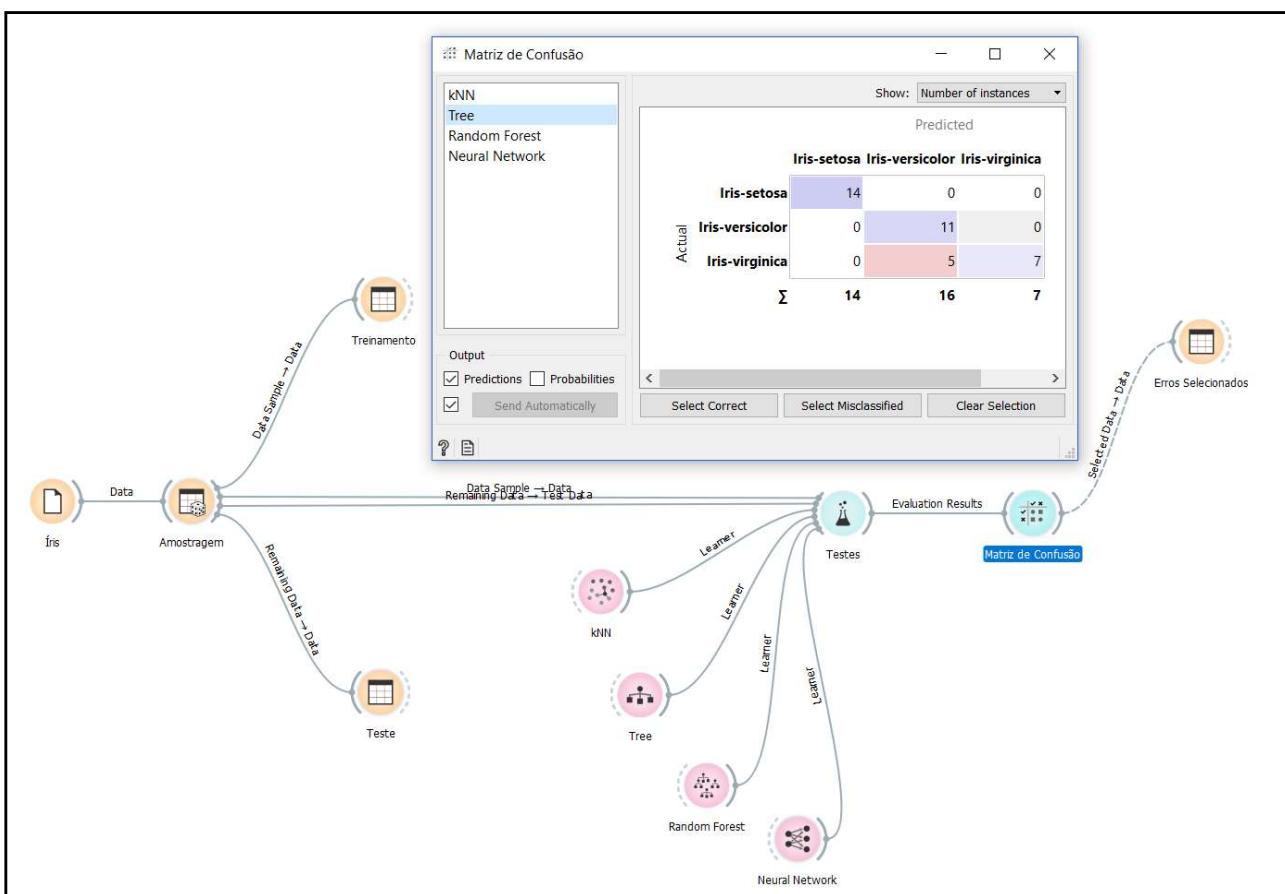
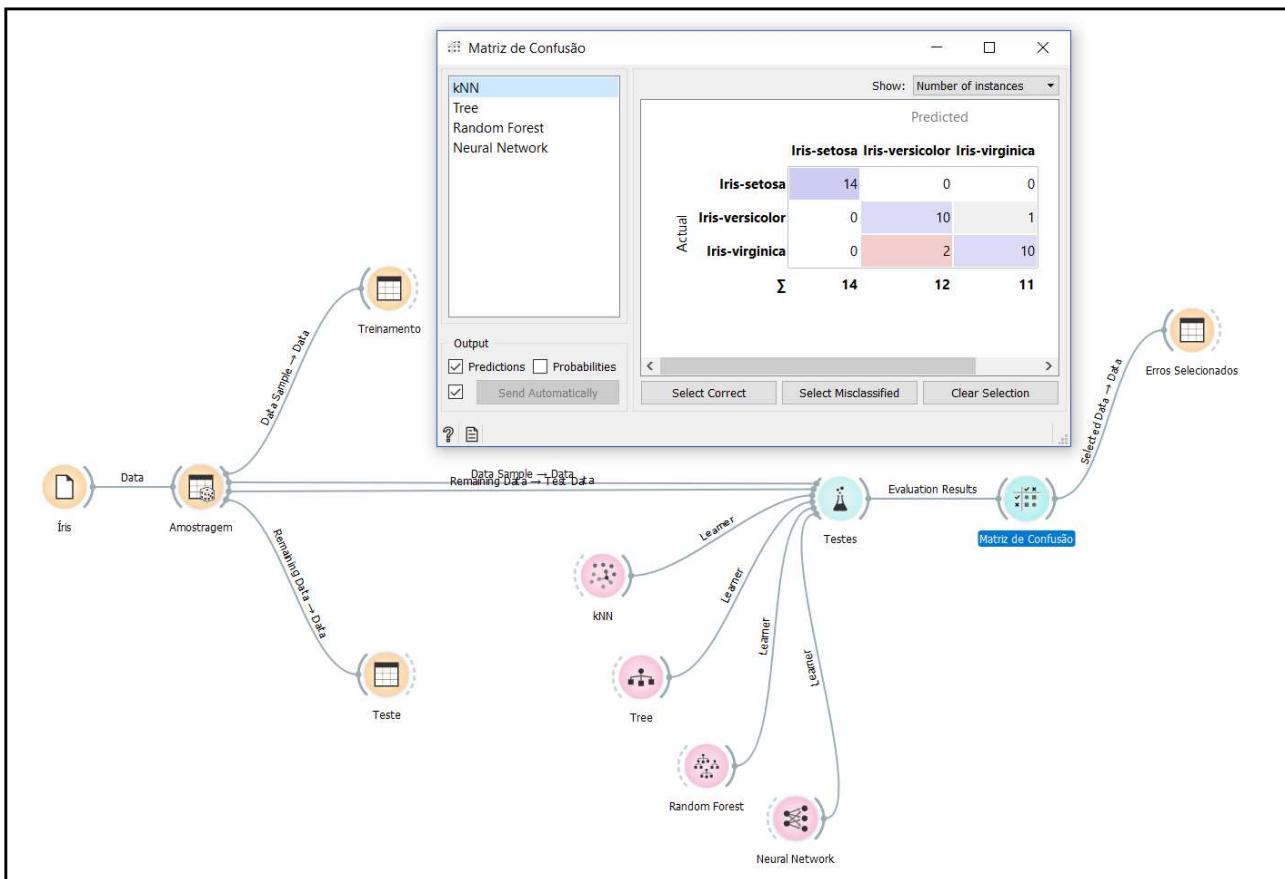
O Software Orange permite a comparação entre múltiplos classificadores e regressores. Por exemplo, pretende-se comparar as acuráncias entre os classificadores: k-NN, Tree, Random Forest, SVM e Neural Network, quando operando sobre os “Dados Íris” (utilizando-se as configurações padrões da Versão Orange 3.15). Os dados serão divididos de forma aleatória em: dados de treinamento (75%) e dados de teste (25%).

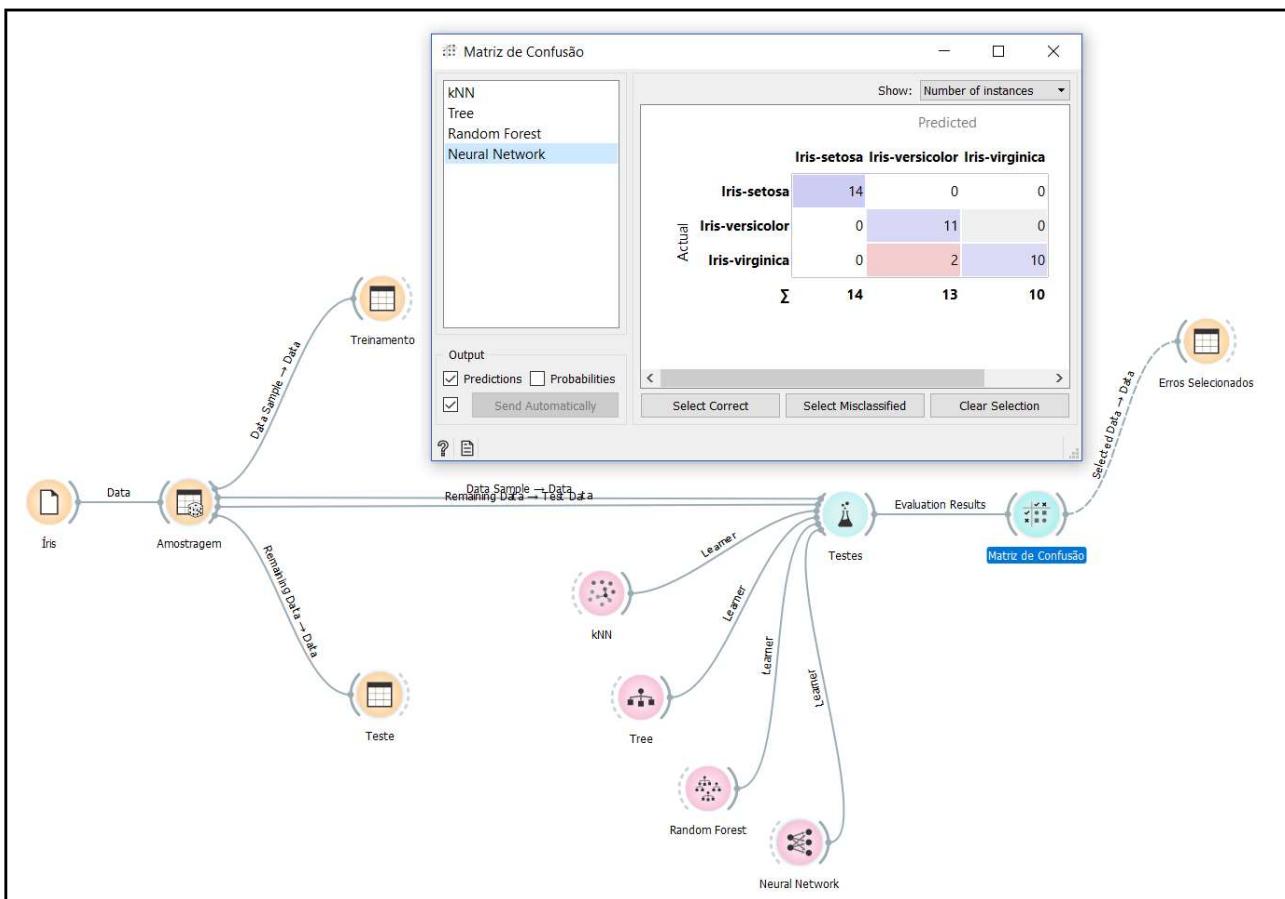
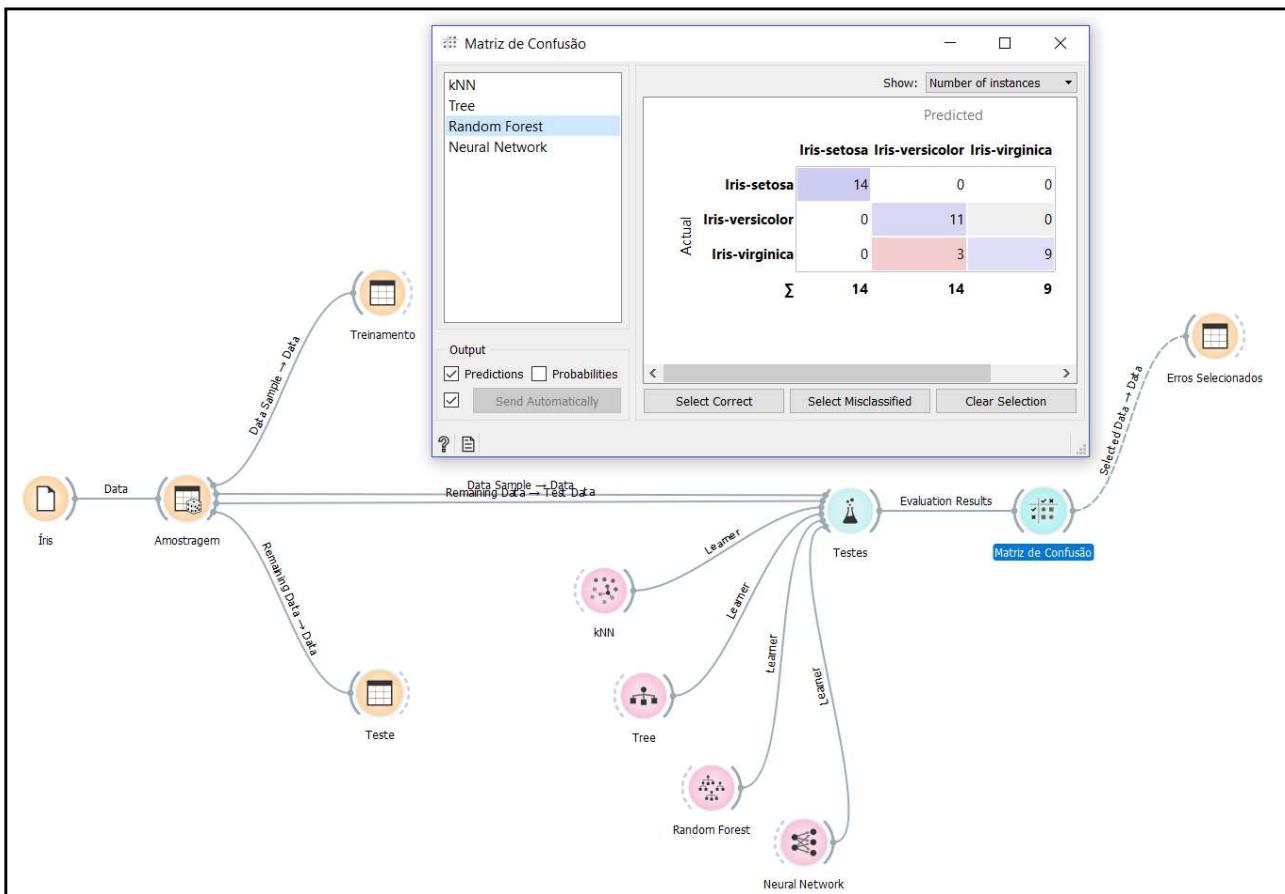










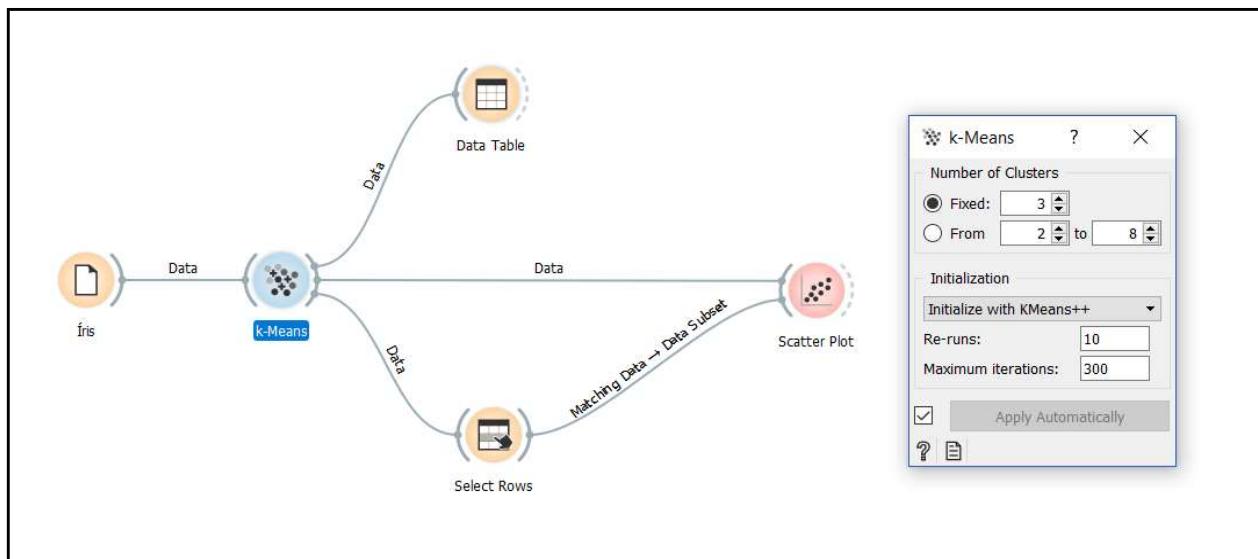


4 APRENDIZAGEM NÃO SUPERVISIONADA

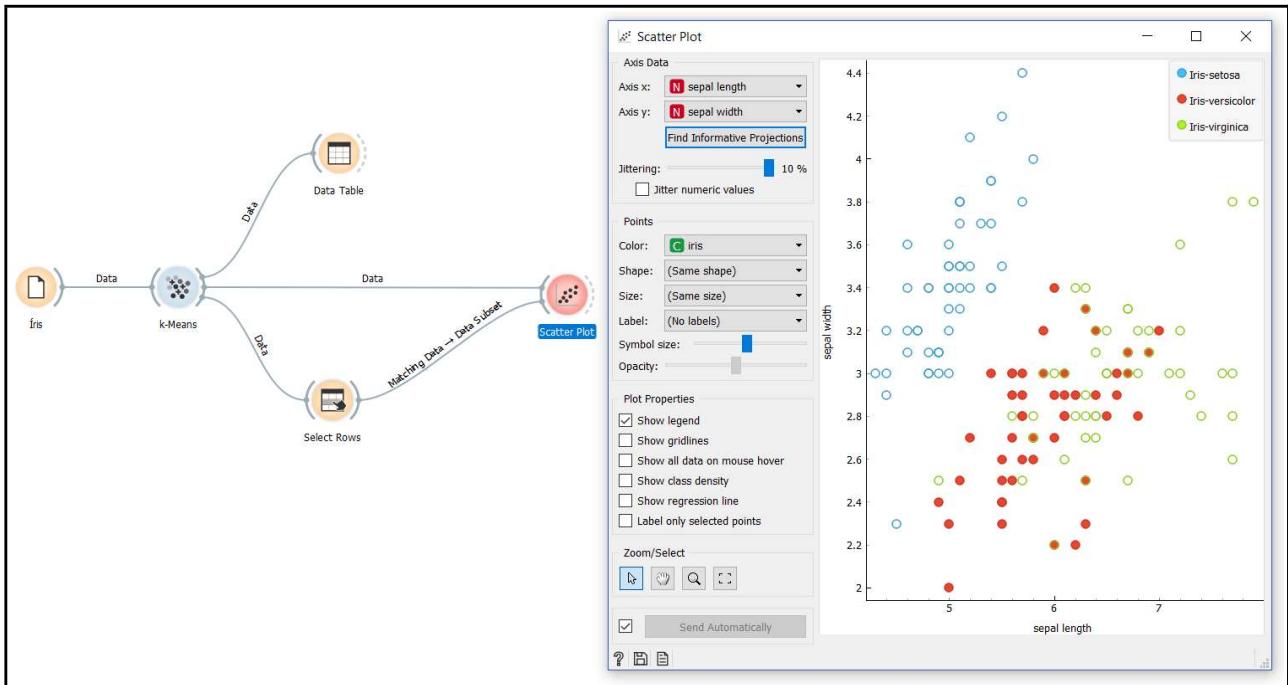
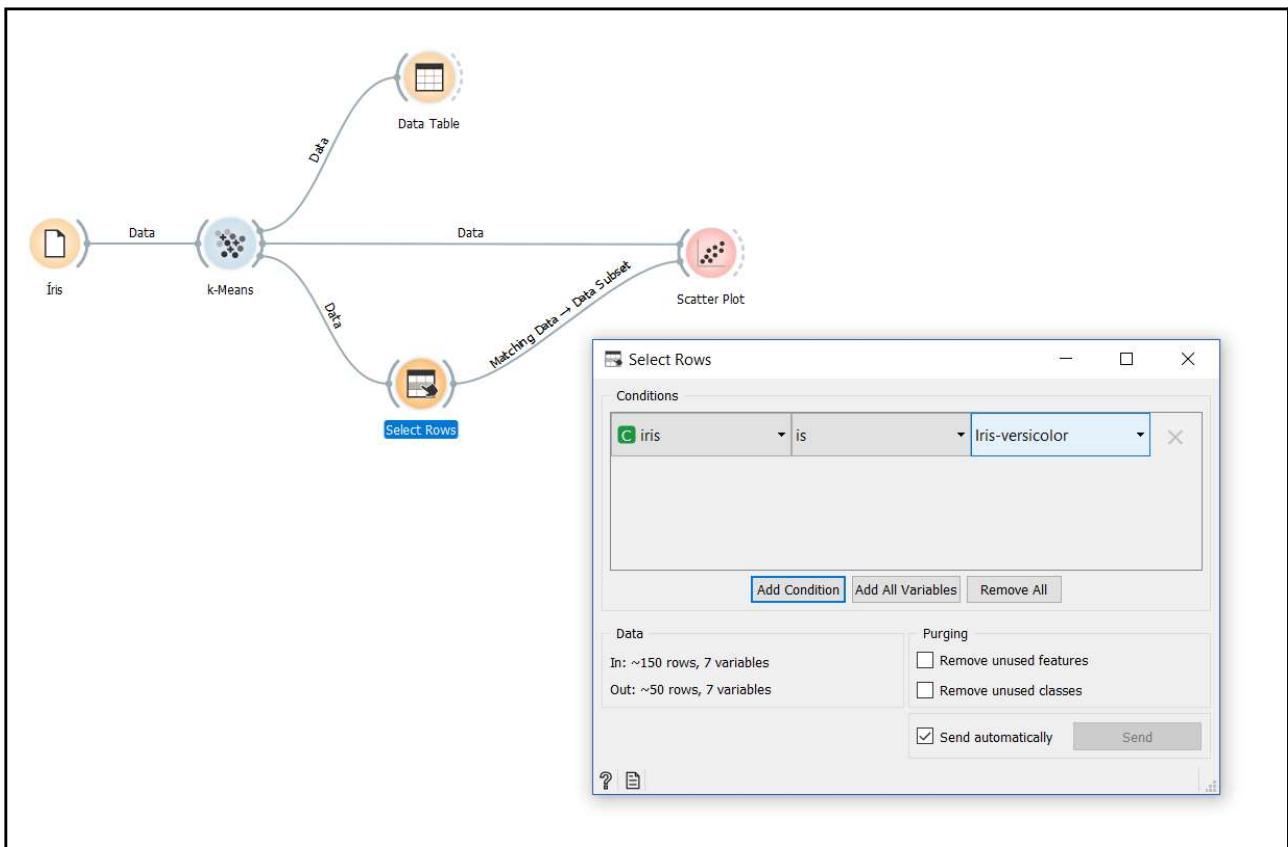
A Aprendizagem Não Supervisionada (*Unsupervised Learning*) é o termo usado quando um algoritmo tem por finalidade o reconhecimento de padrões (agrupamentos) presentes em bases de dados, porém sem a necessidade do algoritmo ser “treinado” sobre um conjunto de dados pré-definido (elemento supervisor). Você pode utilizar vários métodos de agrupamento com os *widgets* de Orange.

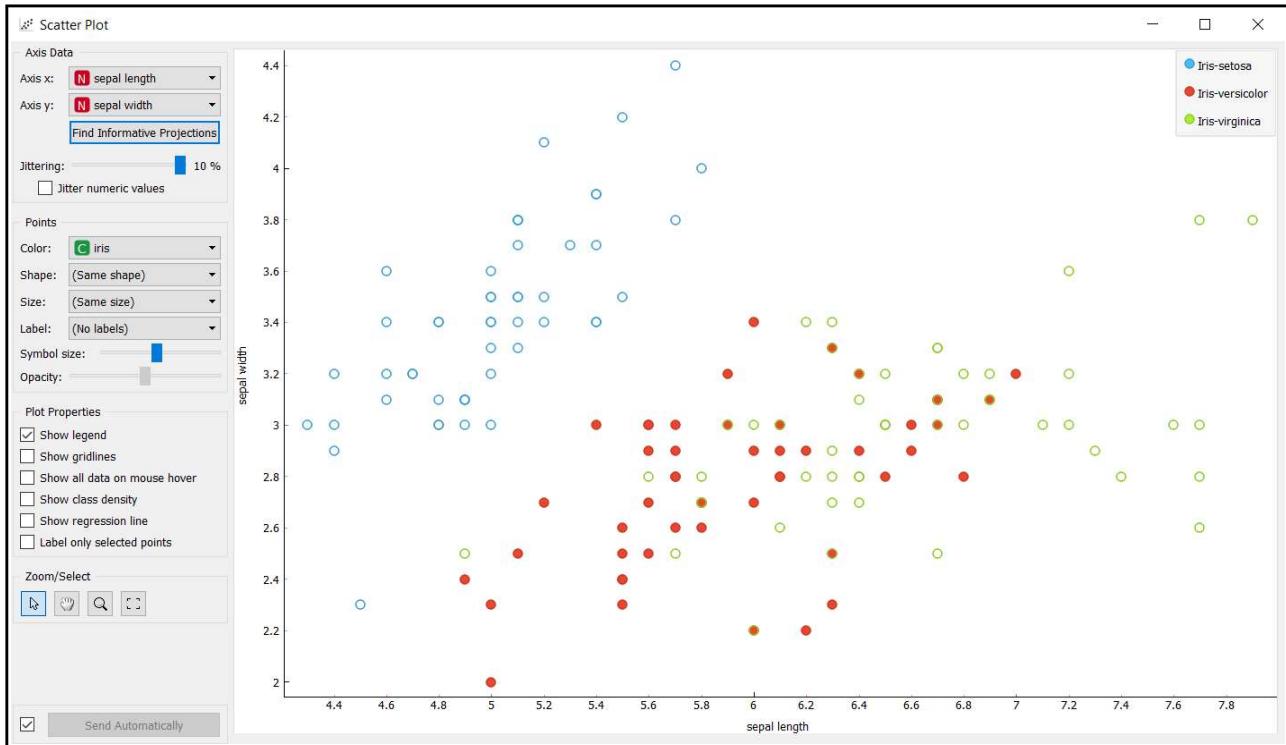
4.1 k-Means Widget

O k-Means Widget aplica o algoritmo de armazenamento em *cluster* denominado k-Means.



	iris	Cluster	Silhouette	sepal length	sepal width	petal length	petal width
92	Iris-versicolor	C1	0.648	5.1	3.0	1.6	1.4
93	Iris-versicolor	C1	0.678	5.8	2.6	4.0	1.2
94	Iris-versicolor	C1	0.590	5.0	2.3	3.3	1.0
95	Iris-versicolor	C1	0.678	5.6	2.7	4.2	1.3
96	Iris-versicolor	C1	0.674	5.7	3.0	4.2	1.2
97	Iris-versicolor	C1	0.677	5.7	2.9	4.2	1.3
98	Iris-versicolor	C1	0.662	6.2	2.9	4.3	1.3
99	Iris-versicolor	C1	0.544	5.1	2.5	3.0	1.1
100	Iris-versicolor	C1	0.679	5.7	2.8	4.1	1.3
101	Iris-virginica	C3	0.647	6.3	3.3	6.0	2.5
102	Iris-virginica	C1	0.573	5.8	2.7	5.1	1.9
103	Iris-virginica	C3	0.675	7.1	3.0	5.9	2.1
104	Iris-virginica	C3	0.610	6.3	2.9	5.6	1.8
105	Iris-virginica	C3	0.662	6.5	3.0	5.8	2.2
106	Iris-virginica	C3	0.659	7.6	3.0	6.6	2.1
107	Iris-virginica	C1	0.639	4.9	2.5	4.5	1.7
108	Iris-virginica	C3	0.662	7.3	2.9	6.3	1.8
109	Iris-virginica	C3	0.632	6.7	2.5	5.8	1.8
110	Iris-virginica	C3	0.663	7.2	3.6	6.1	2.5
111	Iris-virginica	C3	0.581	6.5	3.2	5.1	2.0
112	Iris-virginica	C3	0.572	6.4	2.7	5.3	1.9





4.2 Distances Widget

Calcula as distâncias entre linhas/colunas em um conjunto de dados. É base para outros *widgets* da aprendizagem não supervisionada como por exemplo: **Distance Map widget**, **Distance Matrix widget** e **Hierarchical Clustering widget**.

