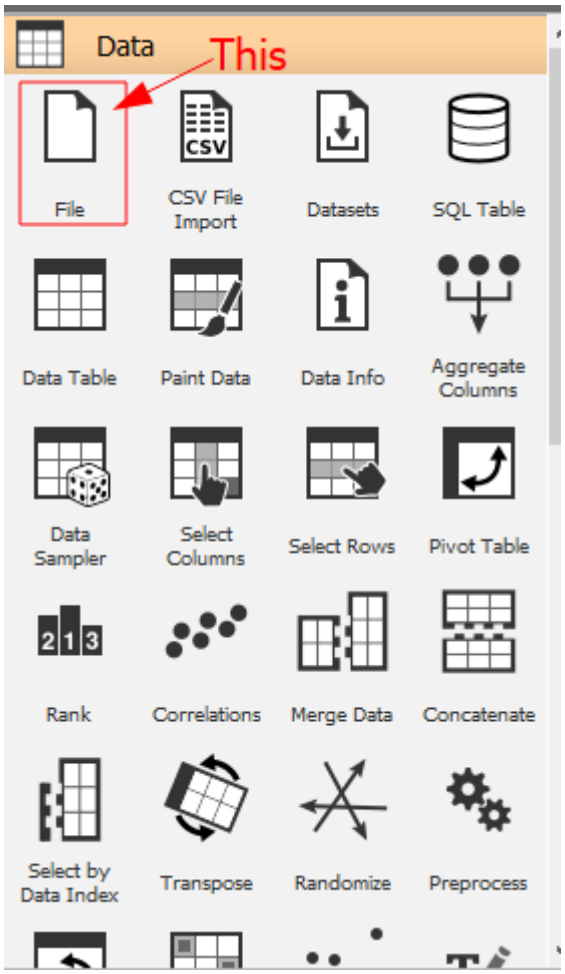
Orange Practicals

Practical 1

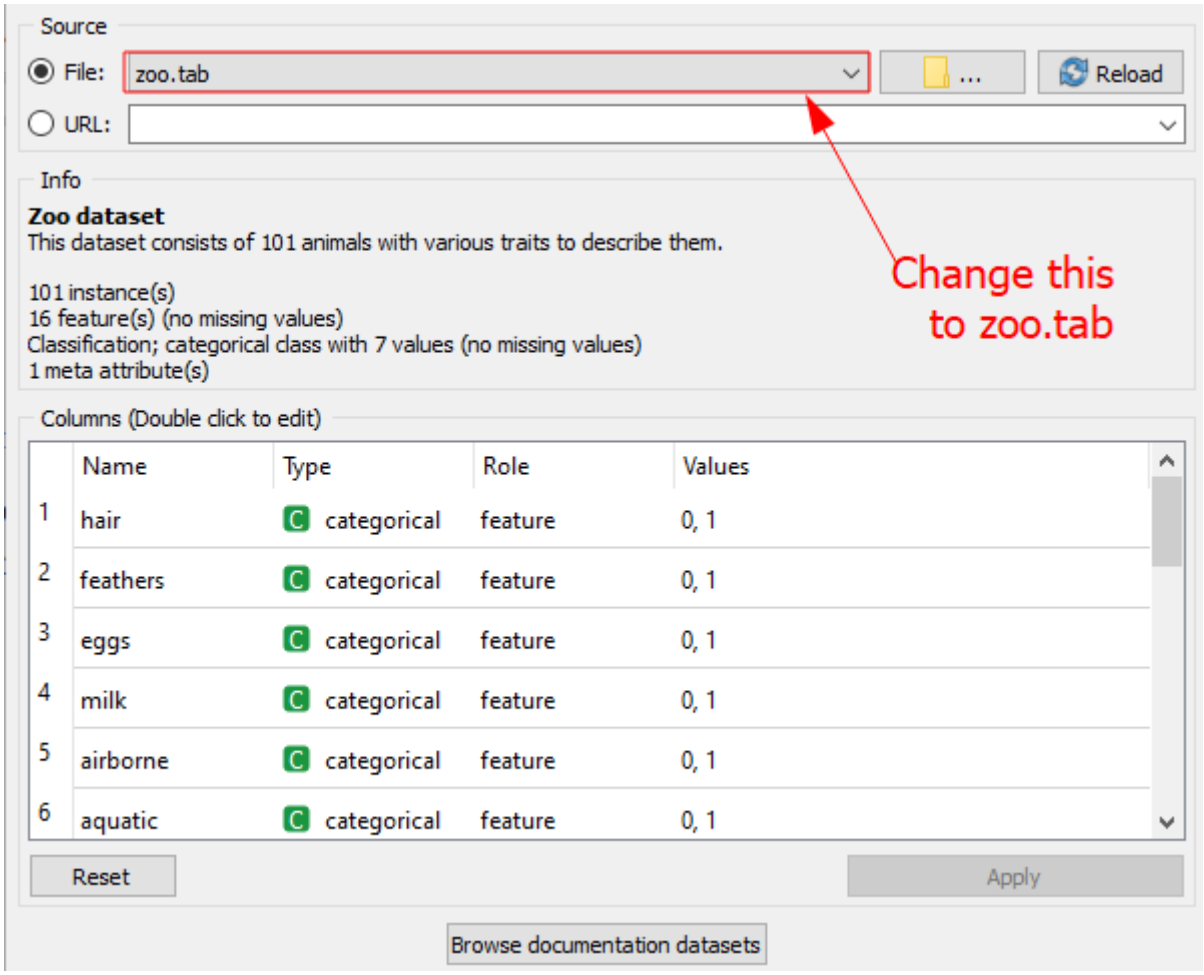
Aim: Classification using orange.

Procedure:

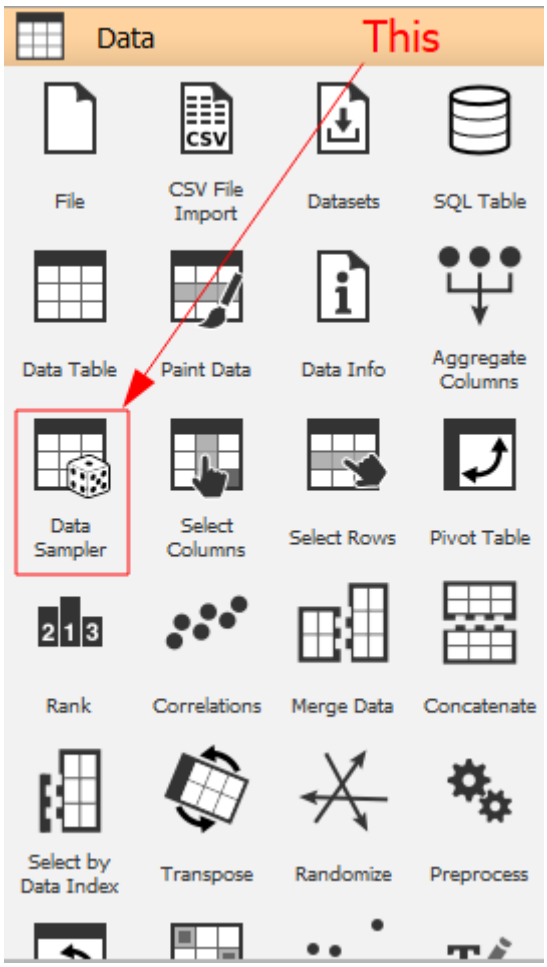
- Drag and drop a `File` widget from the `Data` section found in the left panel to the workspace.

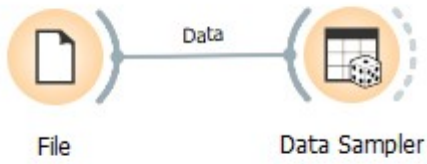


➢ Double click the `File` widget and choose the `zoo.tab` dataset.

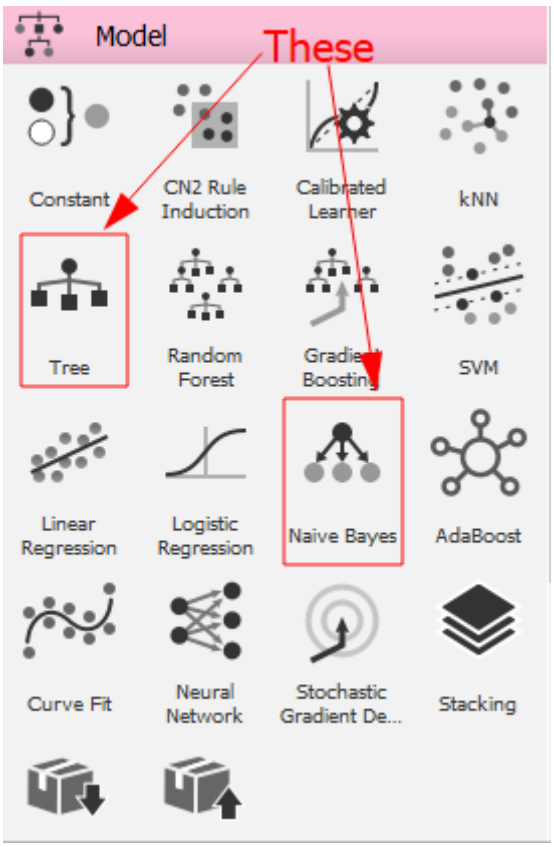


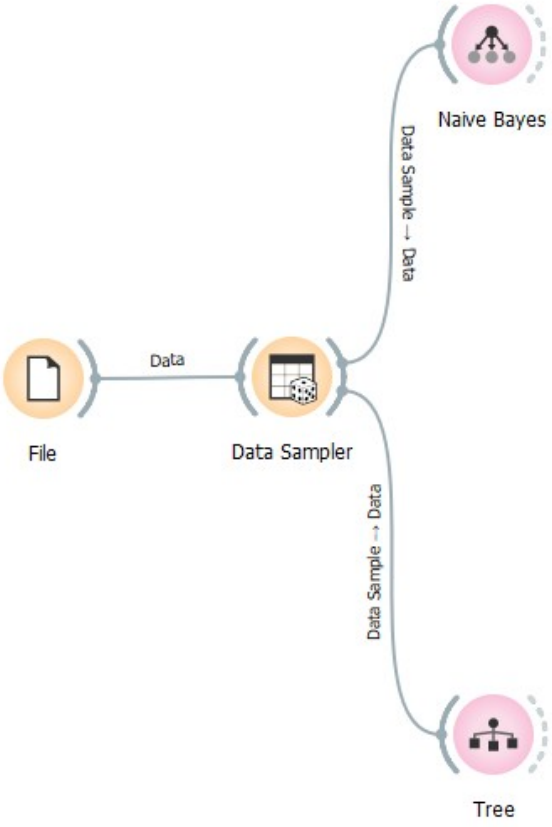
➢ Drag and drop a `Data Sampler` widget onto the workspace and connect it to the `File` widget. This widget is found in the `Data` section.



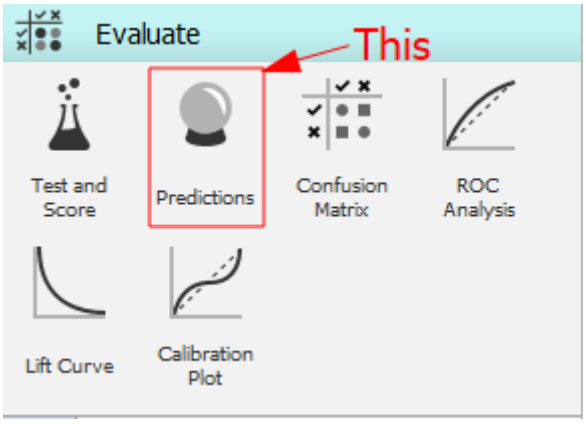


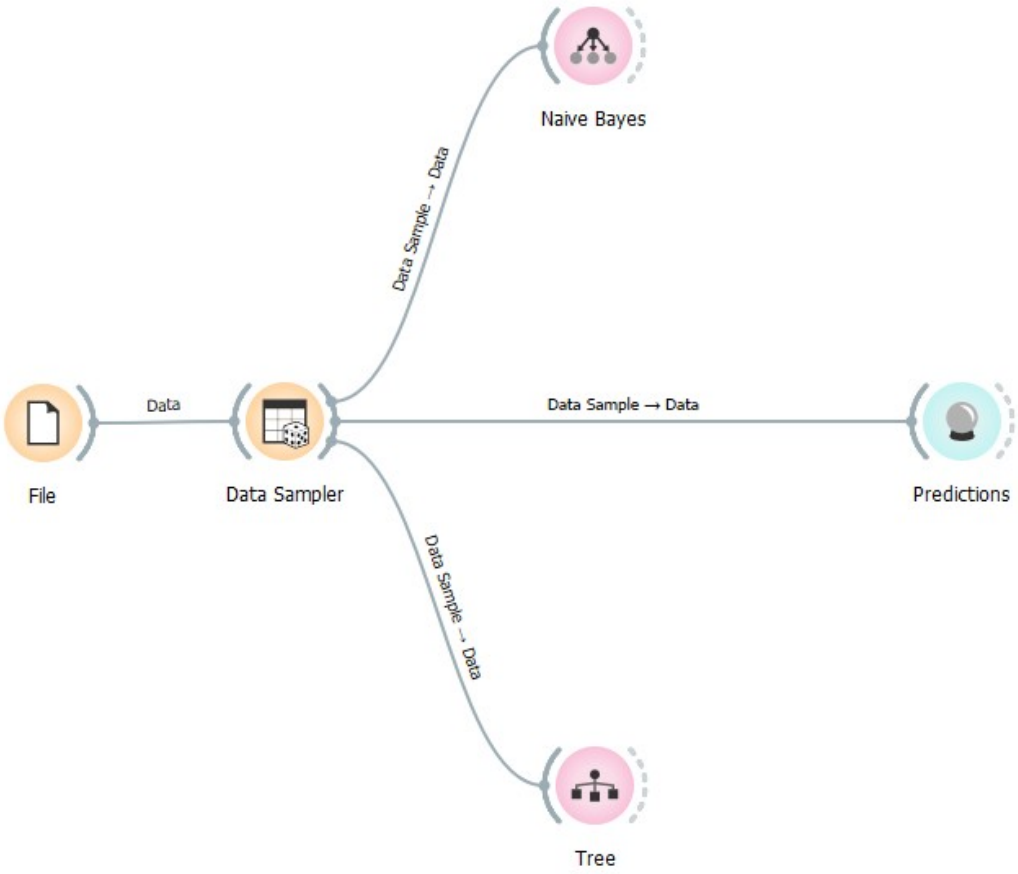
- We choose the `Naive Bayes`' and Classification `Tree` models to classify our data. Drag and drop these widgets to the workspace. They are found under the `Models` section.


Model
These
Constant
CN2 Rule Induction
Calibrated Learner
kNN
Tree
Random Forest
SVM
Linear Regression
Logistic Regression
Naive Bayes
AdaBoost
Curve Fit
Neural Network
Stochastic Gradient De...
Stacking

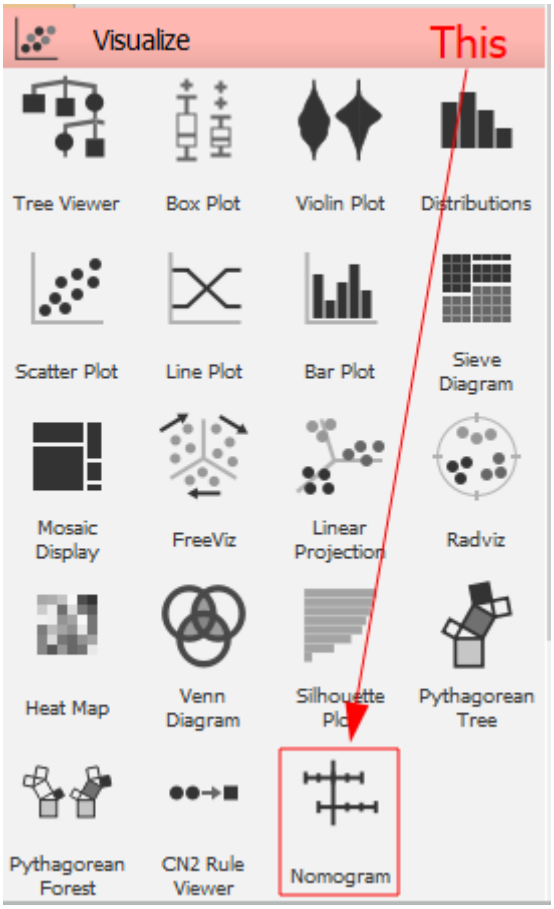


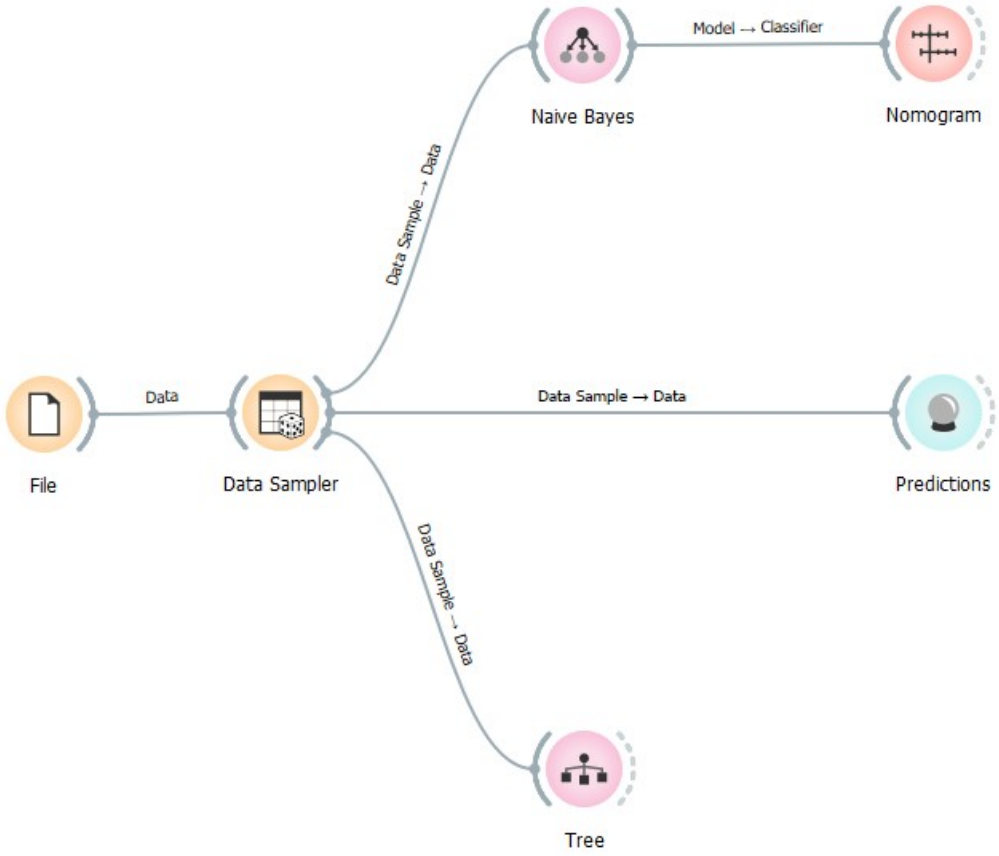
- To perform predictions on the data, we drag and drop the `Predictions` widget onto the workspace.This widget is found in the `Evaluate` section.



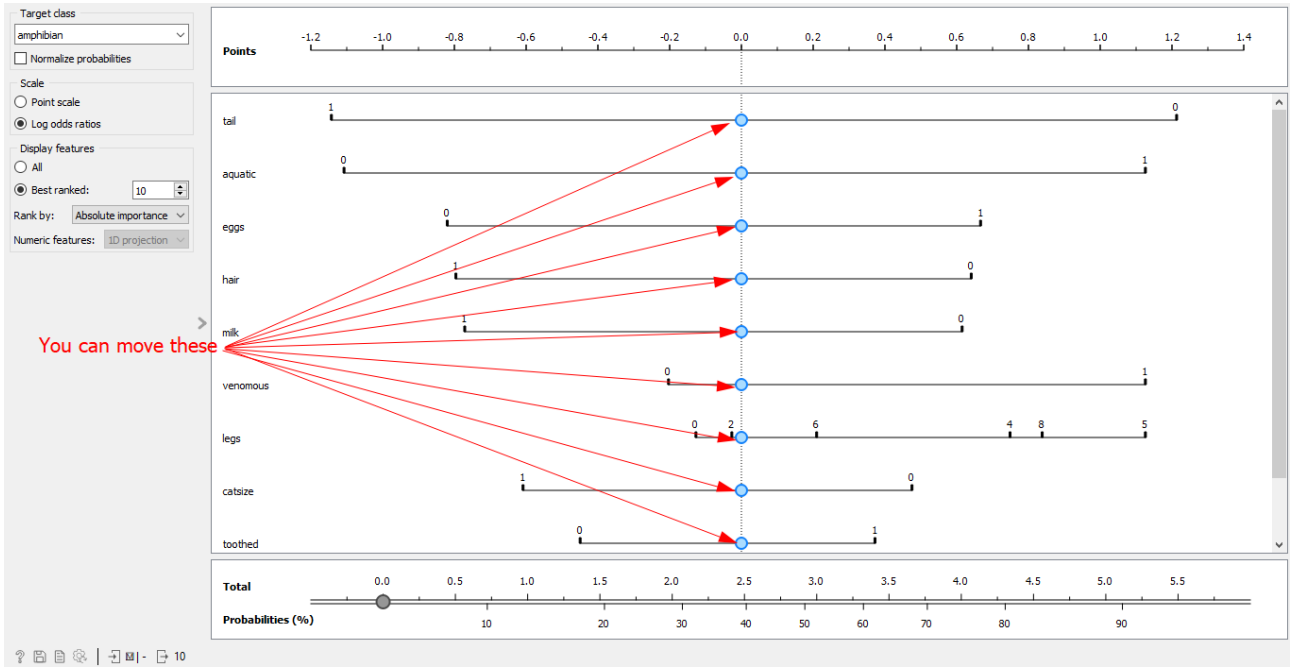


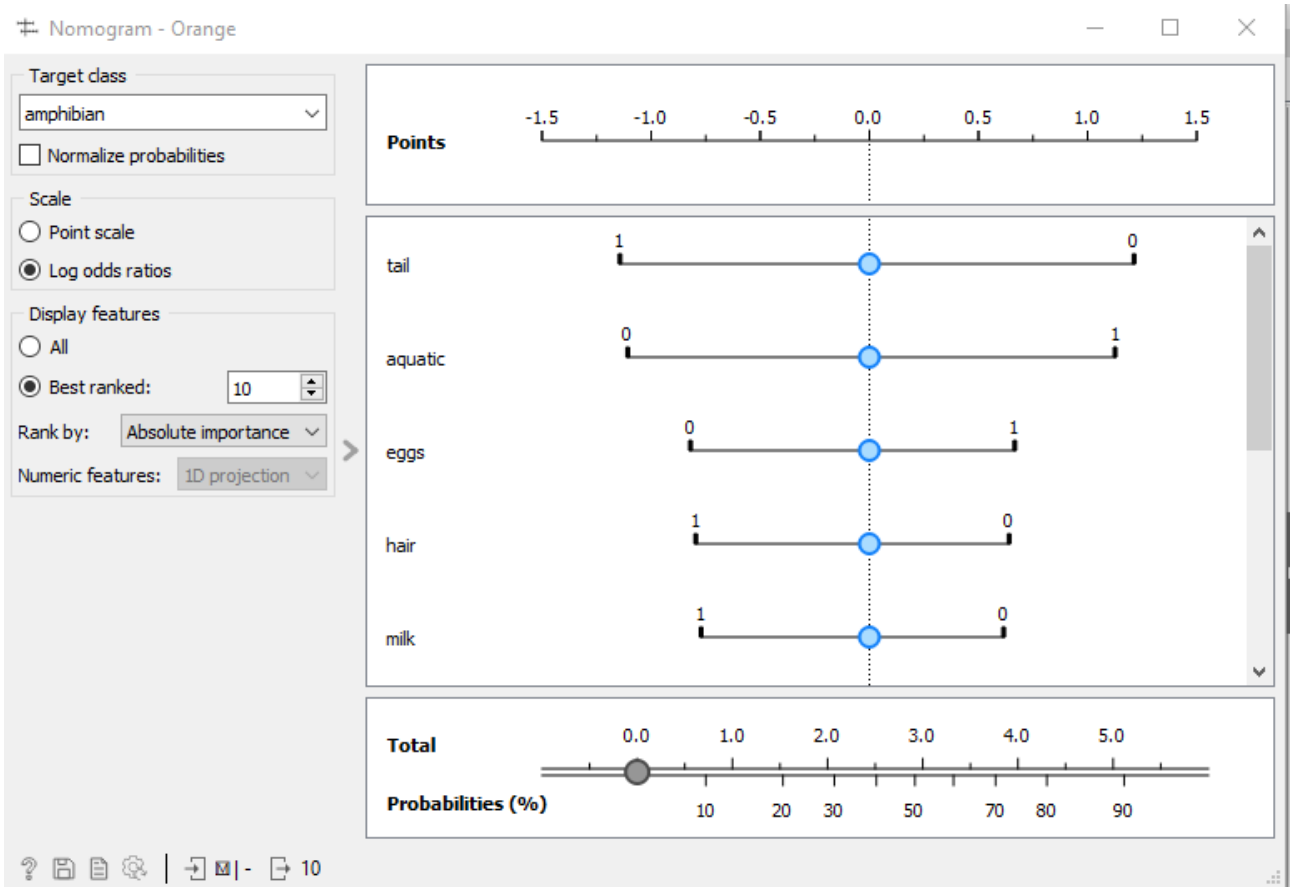
- A `Nomogram` is useful to view data from a `Naive Bayes`' model. Drag and drop this widget onto the workspace. It can be found in the `Visualise` section.

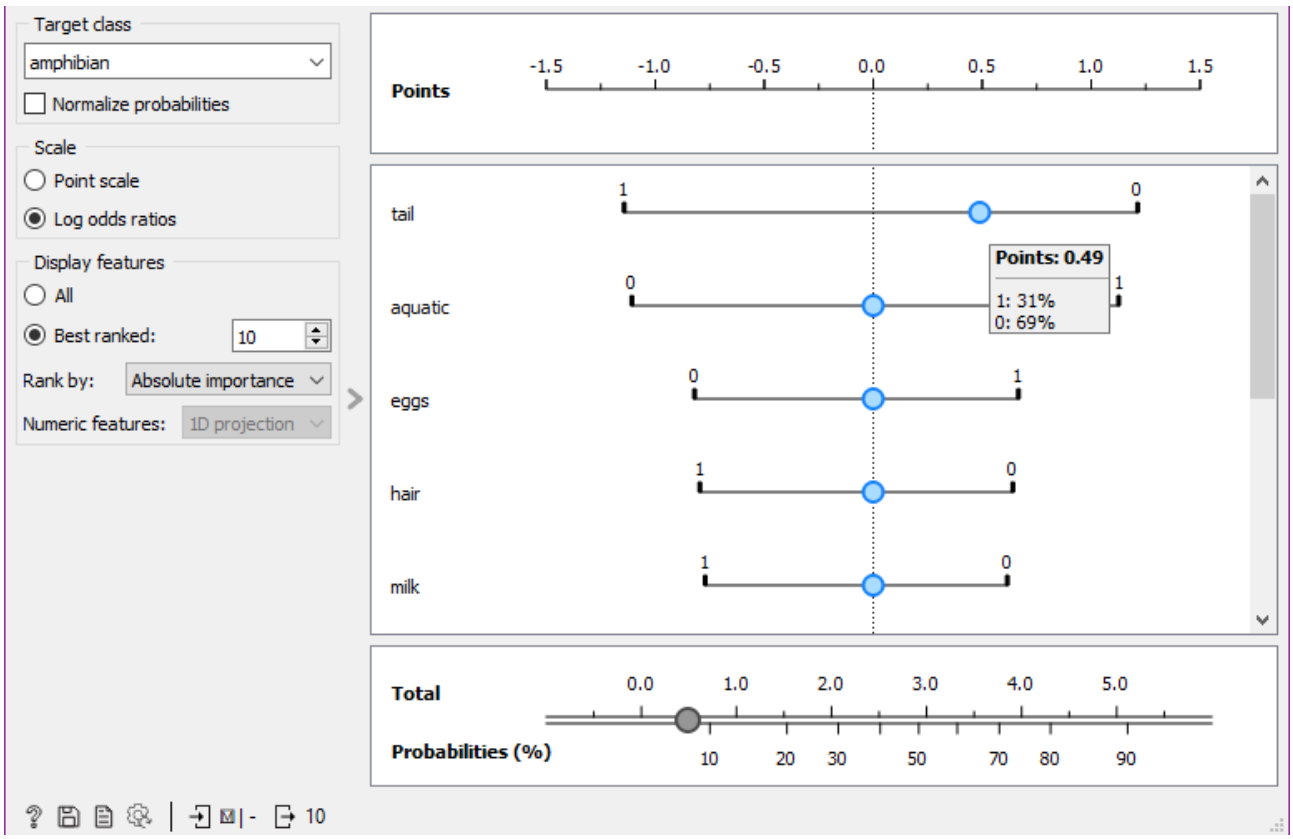

Visualize
This
Tree Viewer
Box Plot
Violin Plot
Distributions
Scatter Plot
Line Plot
Bar Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Silhouette Plo
Pythagorean Tree
Pythagorean Forest
CN2 Rule Viewer
Nomogram



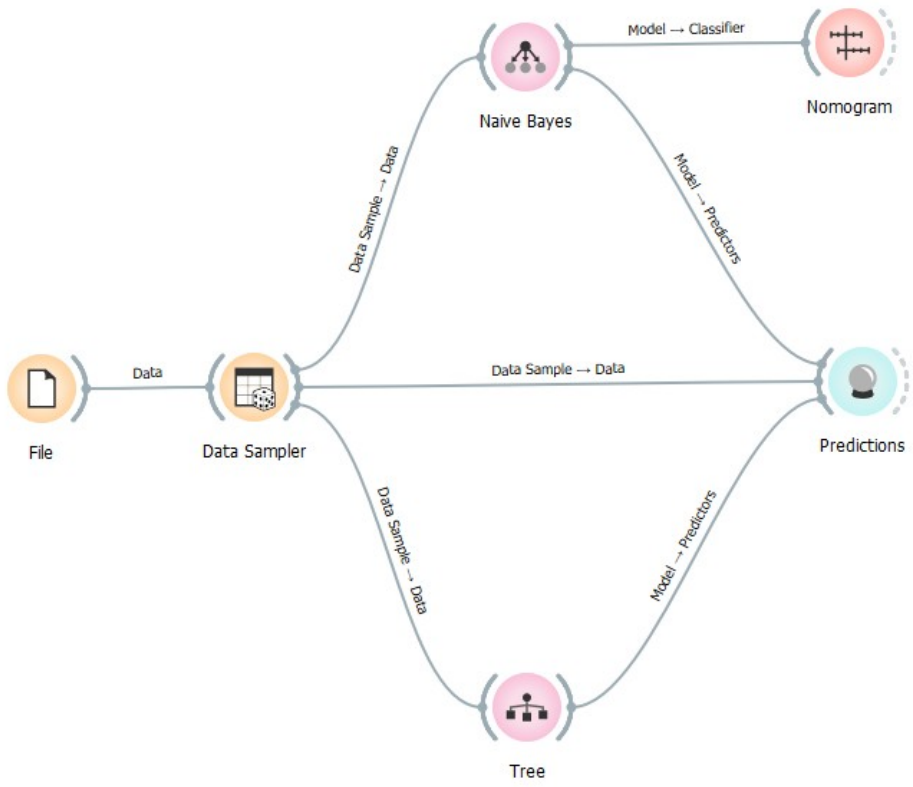
- You can move the points in the `Nomogram` to see the probabilities of a particular class. Here 1 indicates favourable probability while 0 indicates unfavourable probability.


Target class
amphibian
Normalize probabilities
Scale
Point scale
Log odds ratios
Display features
All
Best ranked: 10
Rank by: Absolute importance
Numeric features: 1D projection
Points
tail
aquatic
eggs
hair
milk
venomous
legs
catsize
toothed
You can move these
Total
Probabilities (%)

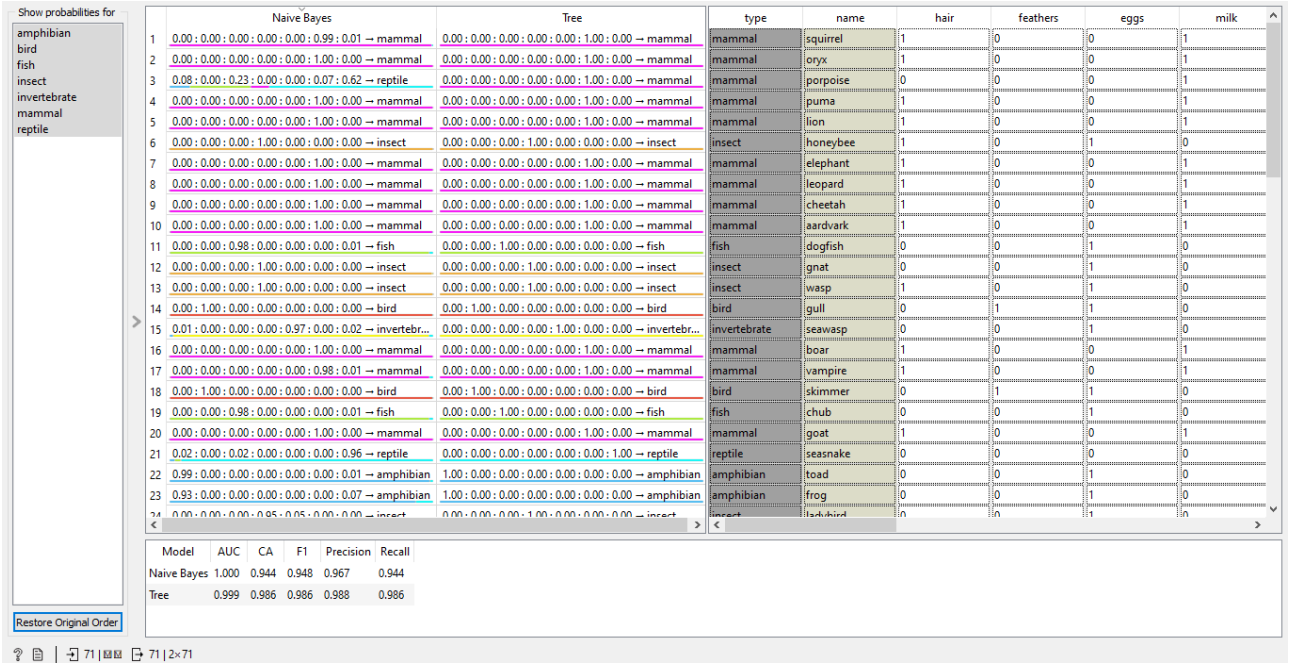

Nomogram - Orange
Target class
amphibian
Normalize probabilities
Scale
Point scale
Log odds ratios
Display features
All
Best ranked: 10
Rank by: Absolute importance
Numeric features: 1D projection
Points
tail
aquatic
eggs
hair
milk
Total
Probabilities (%)



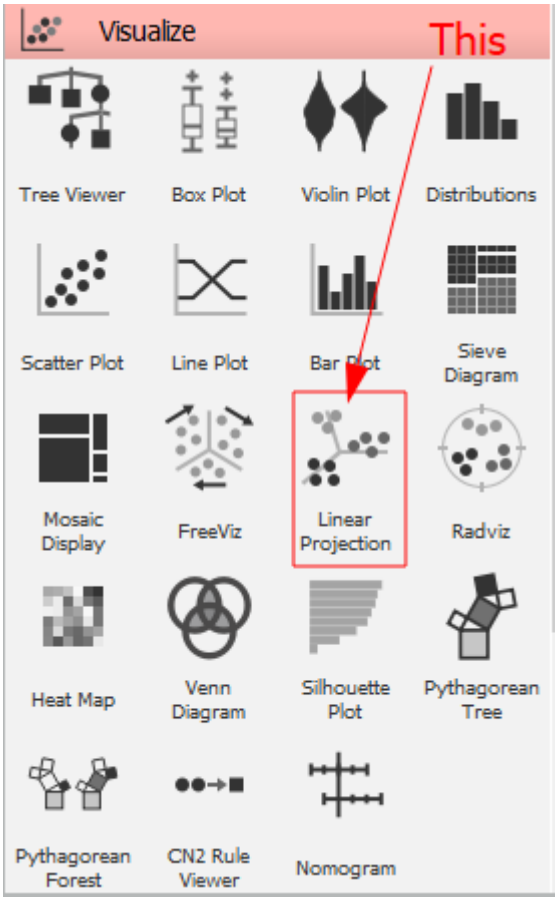
➢ Connect the `Naive Bayes`' widget and `Tree` widget to the `Predictions` widget to perform predictions.

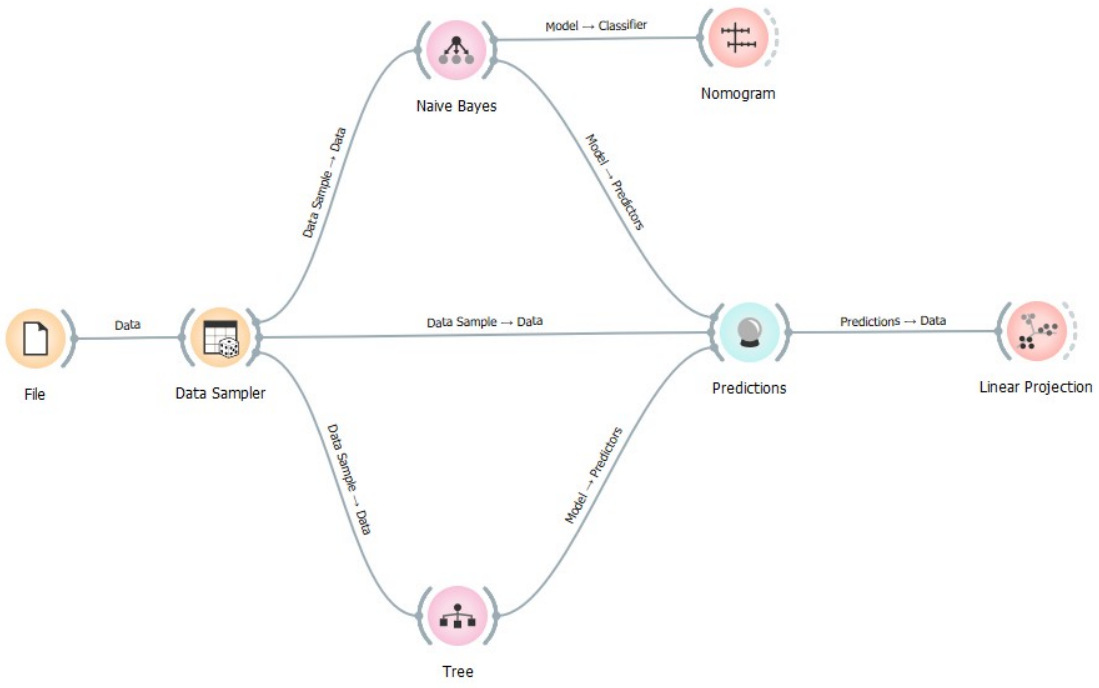


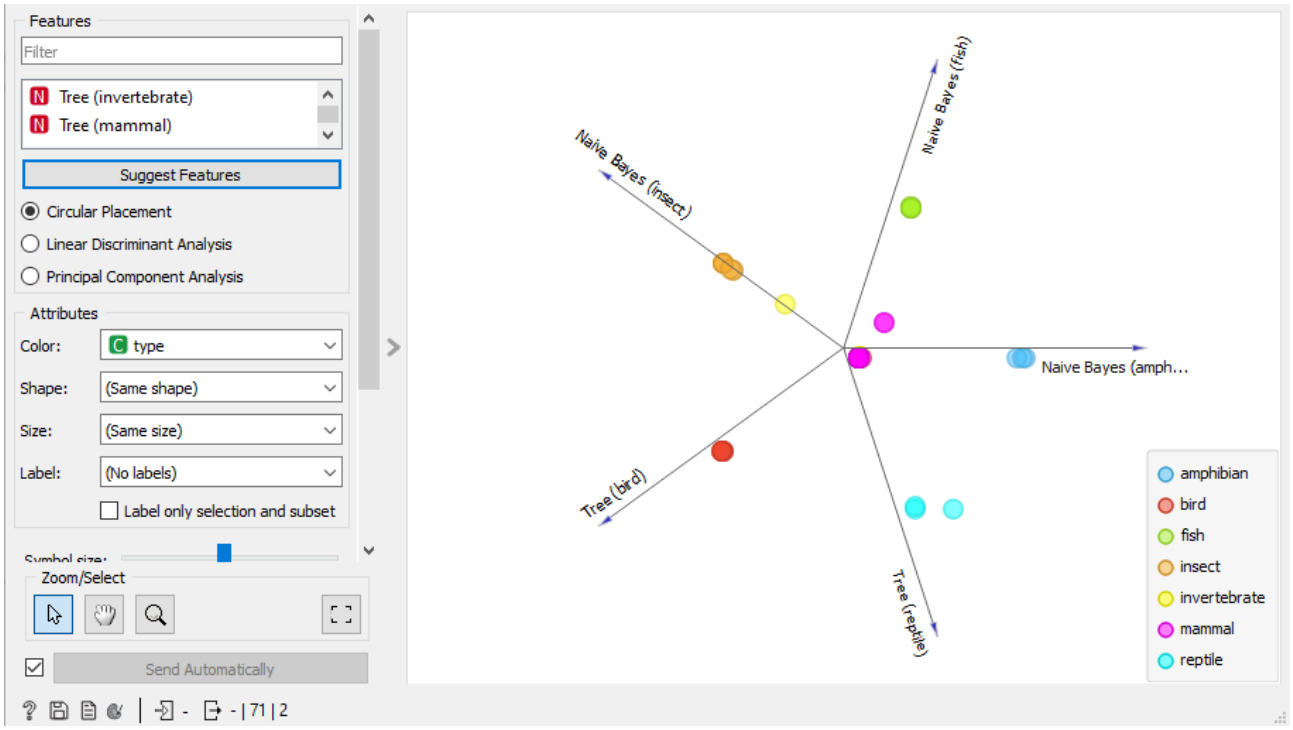
➢ Double click the **Predictions** widget to see the predictions.



➢ We will use the `Linear Projection` widget to visualize the data. Drag and drop this widget from the `Visualise` section.




Model → Classifier
Naive Bayes
Nomogram
Data Sample → Data
Model → Predictors
Data
Data Sample → Data
Predictions → Data
File
Data Sampler
Predictions
Linear Projection
Data Sample → Data
Model → Predictors
Tree

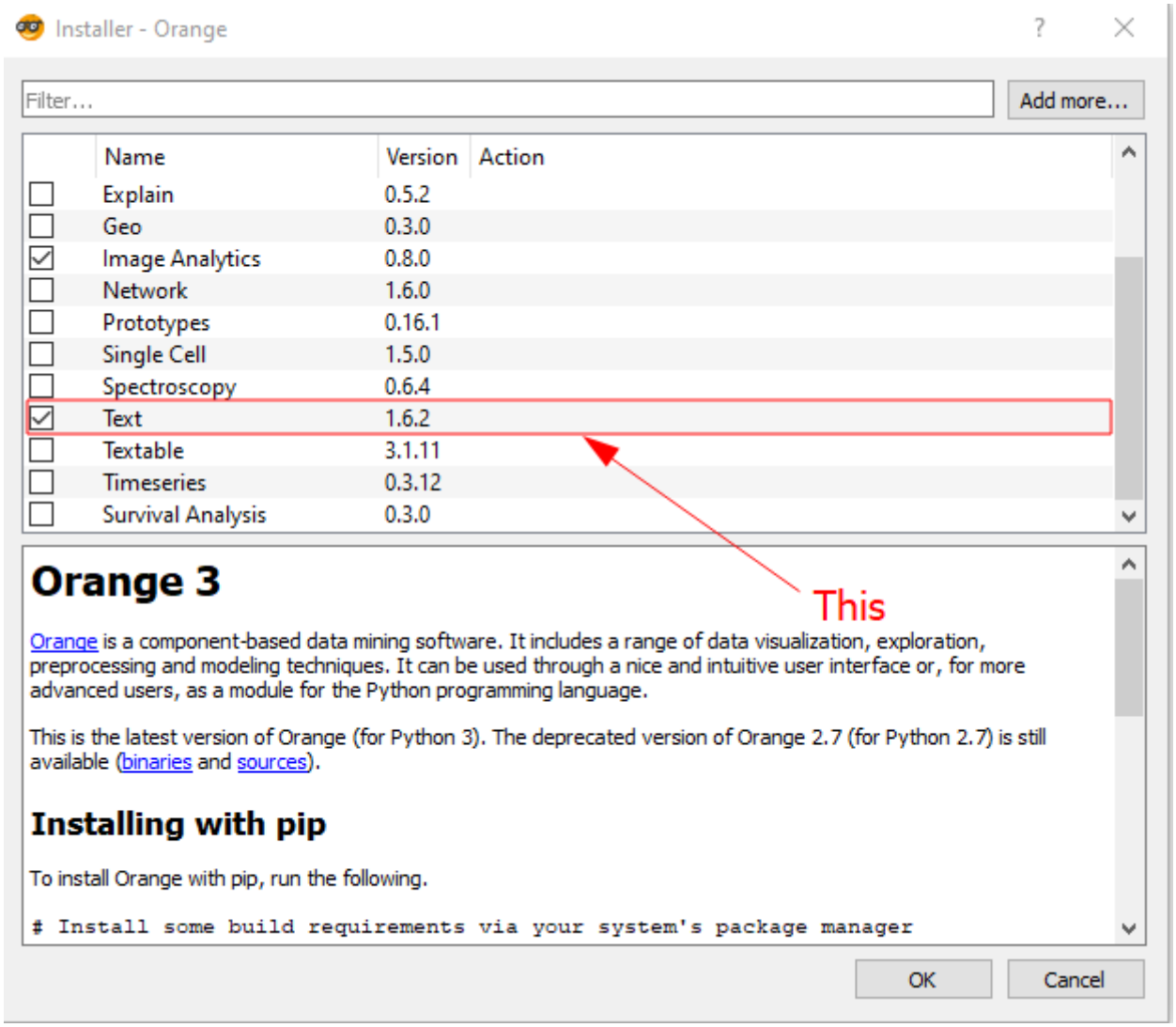

Features
Filter
Tree (invertebrate)
Tree (mammal)
Suggest Features
Circular Placement
Linear Discriminant Analysis
Principal Component Analysis
Attributes
Color: type
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Zoom/Select
Send Automatically
Naive Bayes (fish)
Naive Bayes (insect)
Naive Bayes (amph...
Tree (bird)
Tree (reptile)
amphibian
bird
fish
insect
invertebrate
mammal
reptile

Practical 2

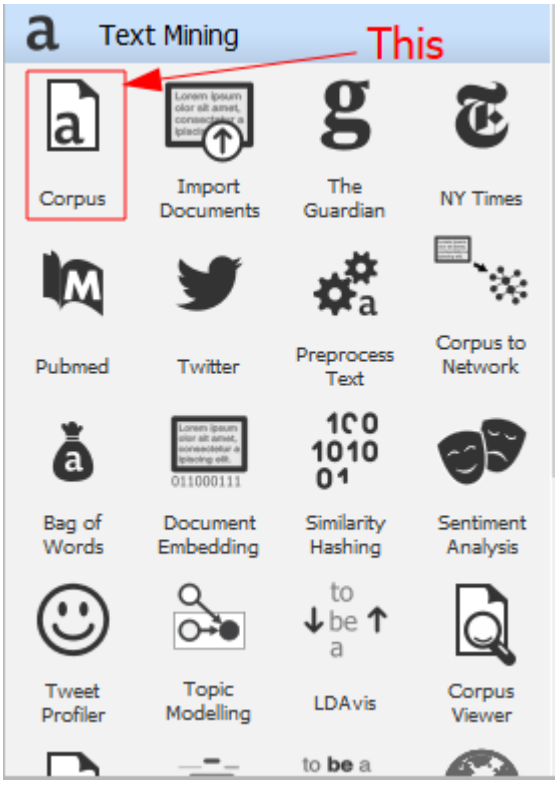
Aim: Text classification using Orange to determine whether the story has magic or is related to animals

Procedure:

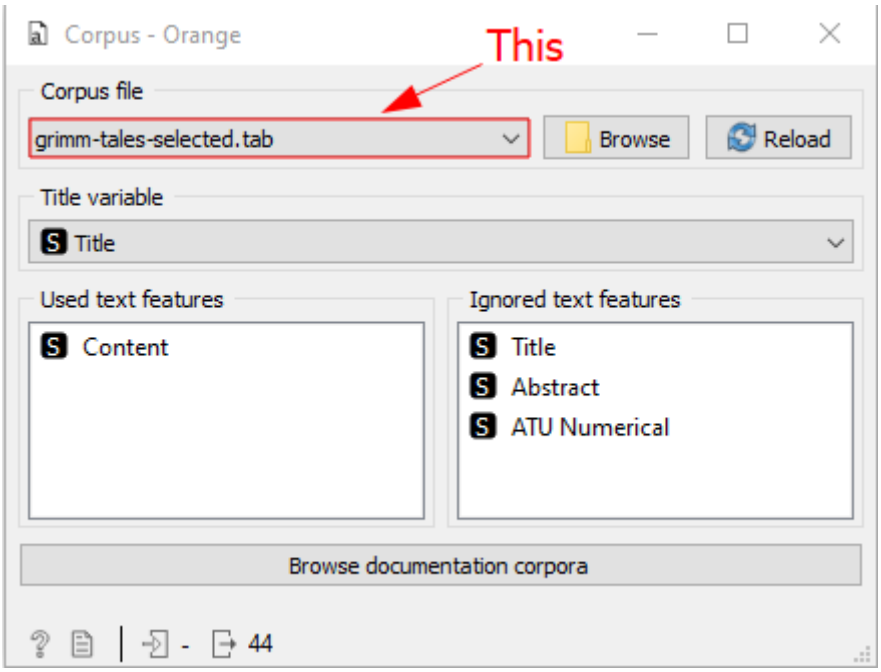
- Before starting, we need to install the `Text` add-on. Install it by navigating to Options > Add ons…. You will be prompted to restart Orange after the installation completes. If it is already installed, skip this step.



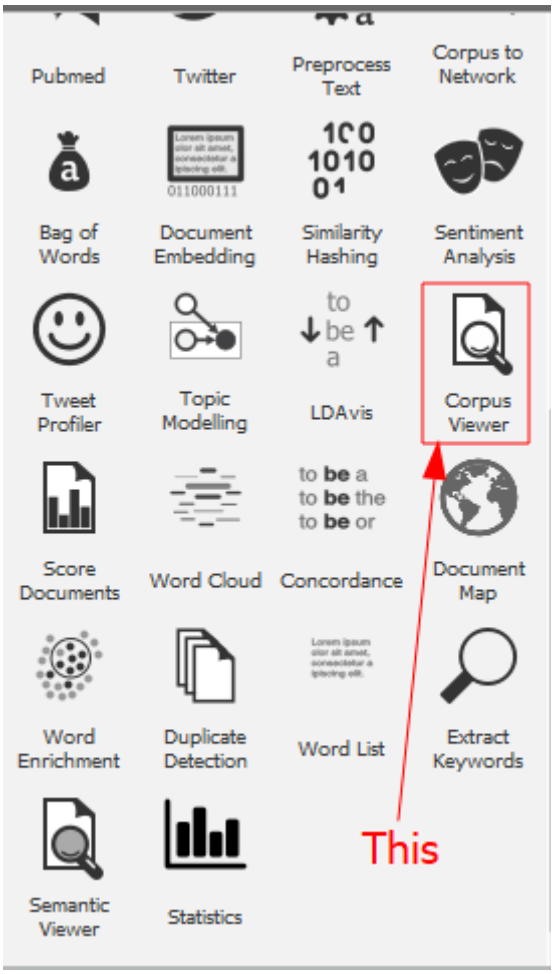
➢ Drag and drop a `Corpus` widget to the workspace. It can be found in the newly added `Text Mining` section.

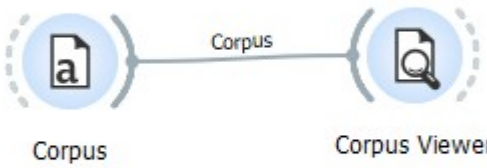


➢ Double click the **Corpus** widget and select the `grimm-tales-selected.tab` corpus file.

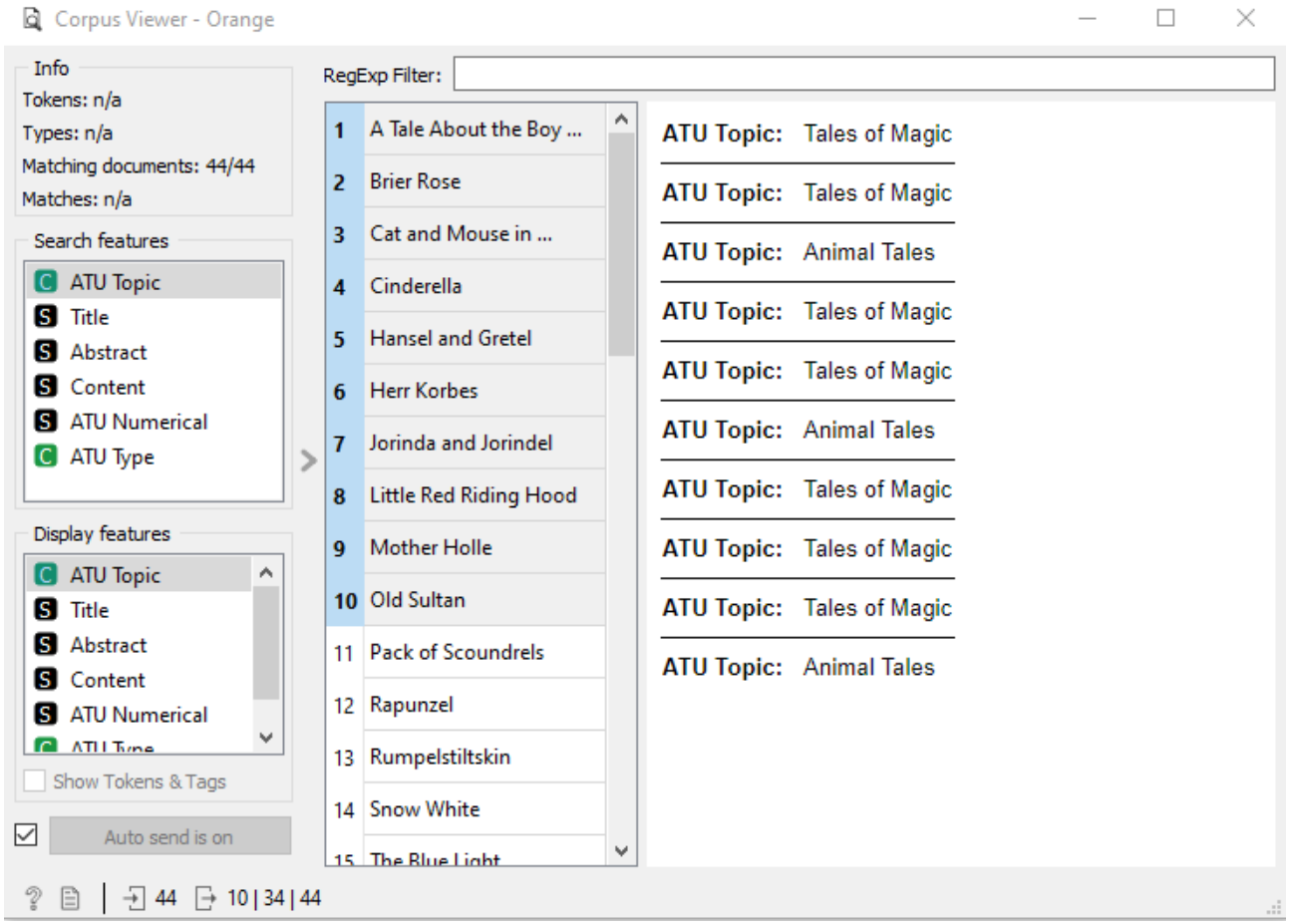


➢ Drag and drop a `Corpus Viewer` widget and connect it to the `Corpus` widget. The `Corpus Viewer` widget can be found in the `Text Mining` section.

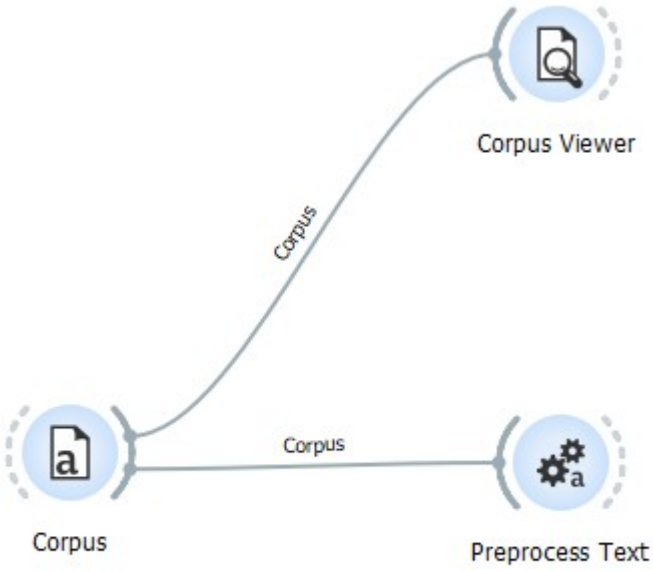




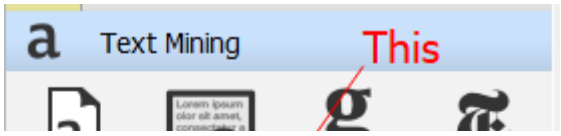
➢ Double click the `Corpus Viewer` widget to visualize the corpus in a tabular format. Select the first 10 entries to use as training set.

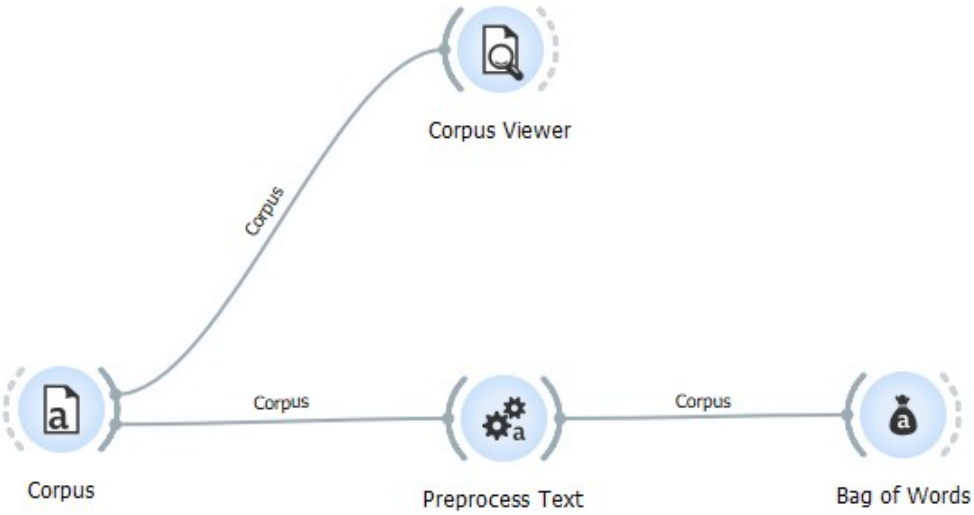


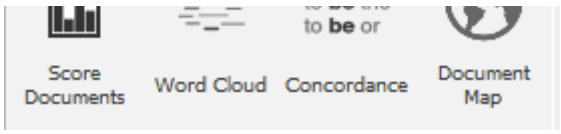
➢ Add a `Preprocess Text` widget to the workspace. This widget can also be found in the `Text Mining` section.



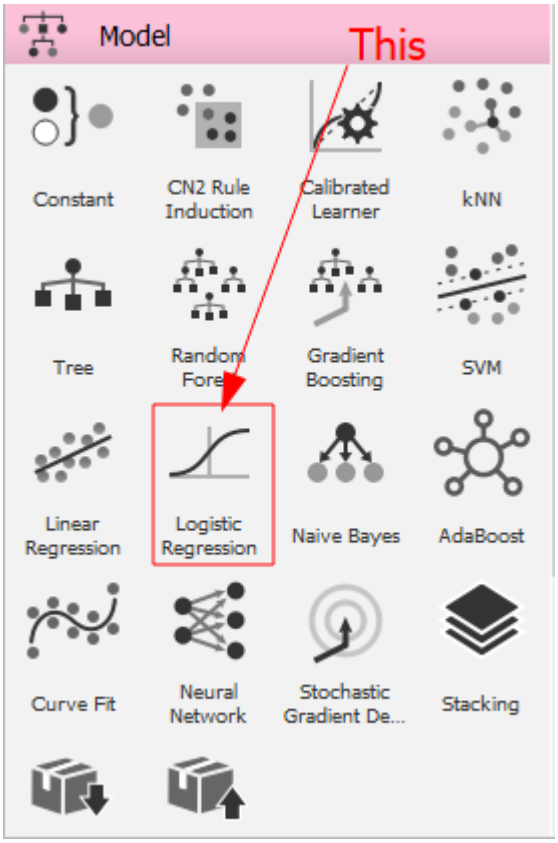
- Add a `Bag Of Words` widget to the workspace and connect it to the `Preprocess Text` widget. This widget is also found in the `Text Mining` section.

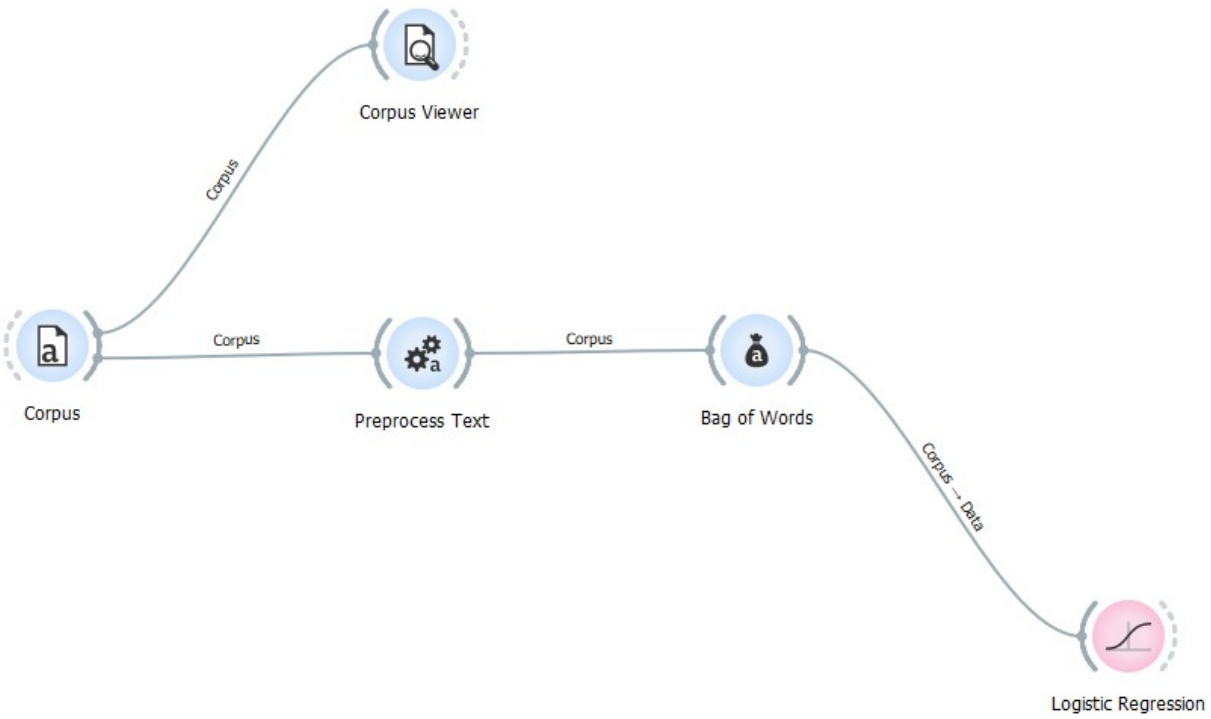




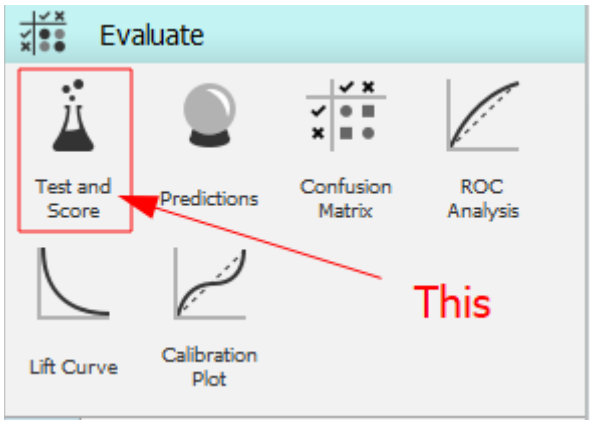


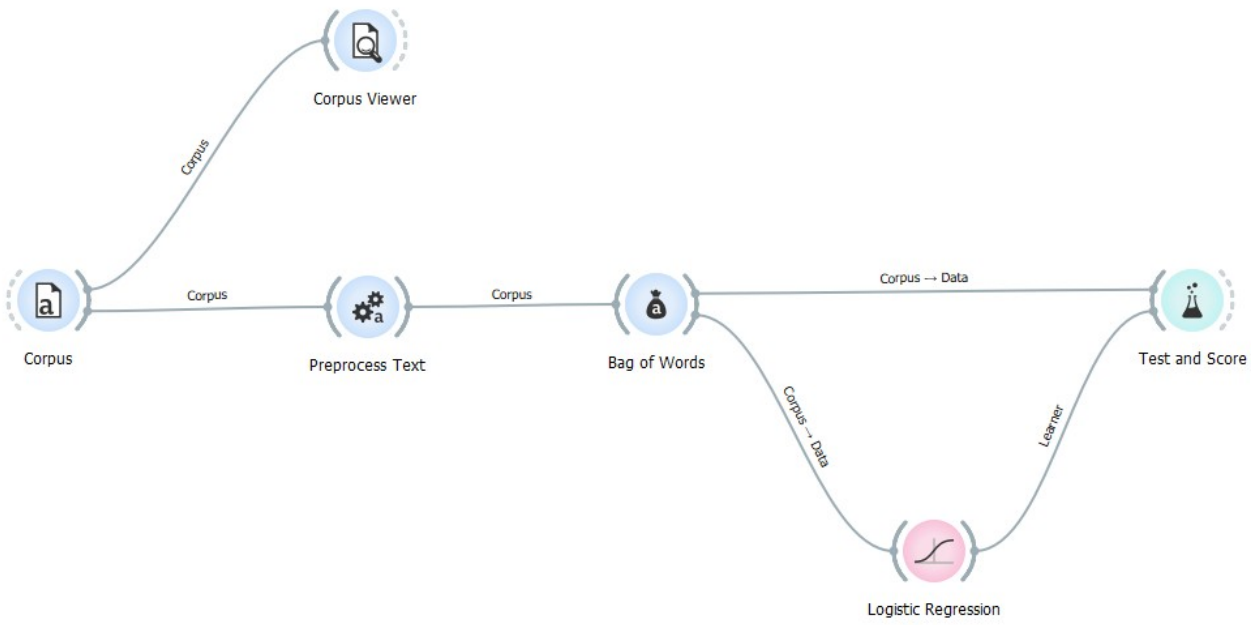
➢ We will use `Logistic Regression` to model our data. Drag and drop this widget from the `Model` section onto the workspace.



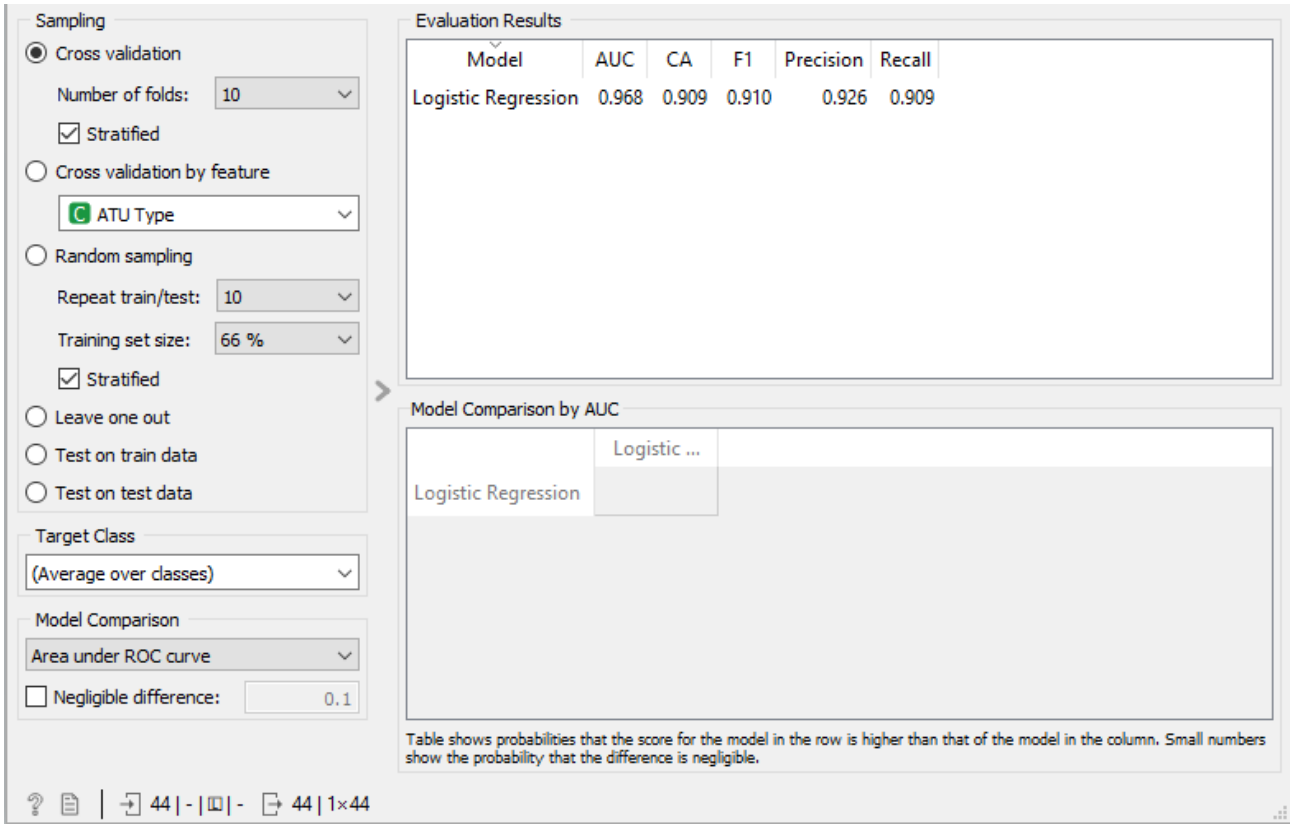


➢ We will use the `Test and Score` widget to check our model. Drag and drop this widget from the `Evaluate` section onto the workspace and connect it to the `Logistic Regression` model and `Bag of Words` widget.

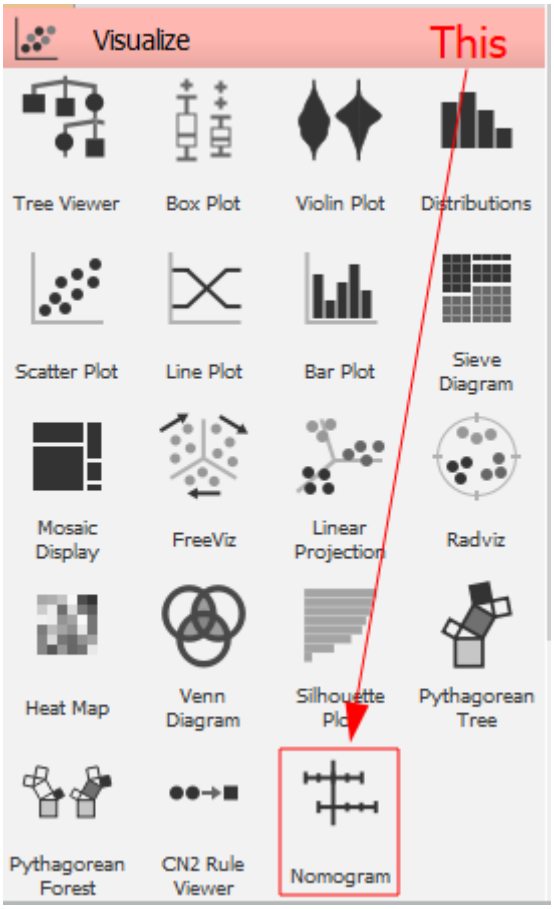


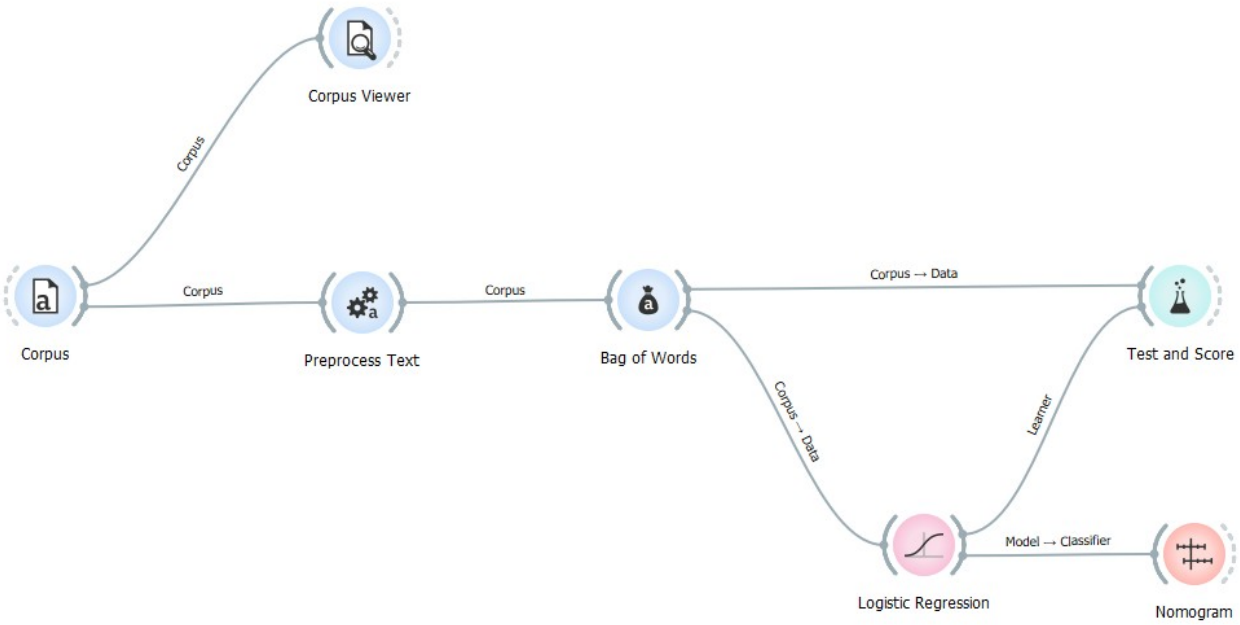


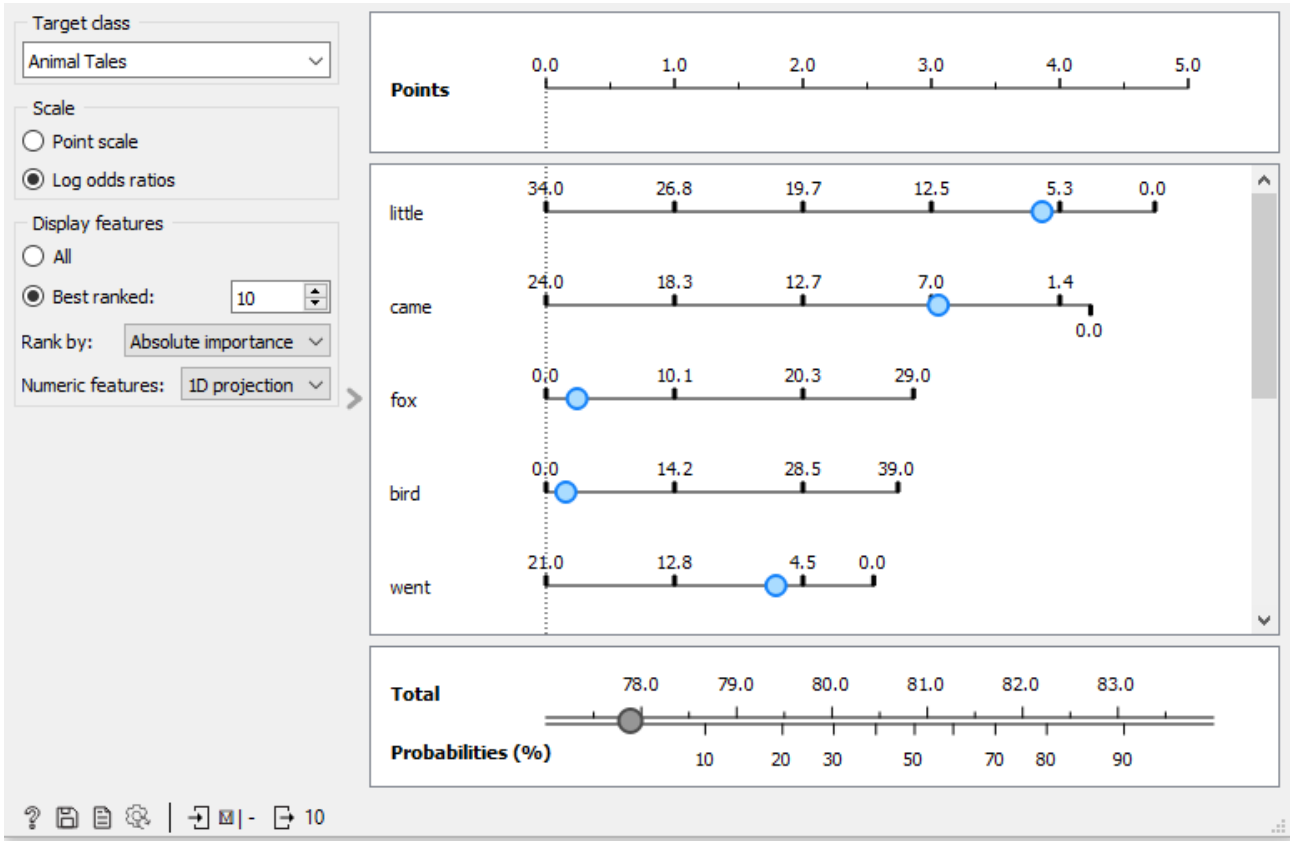
➢ Double-click the `Test and Score` widget after it finishes processing. It will provide you with data such as the accuracy of the model etc.



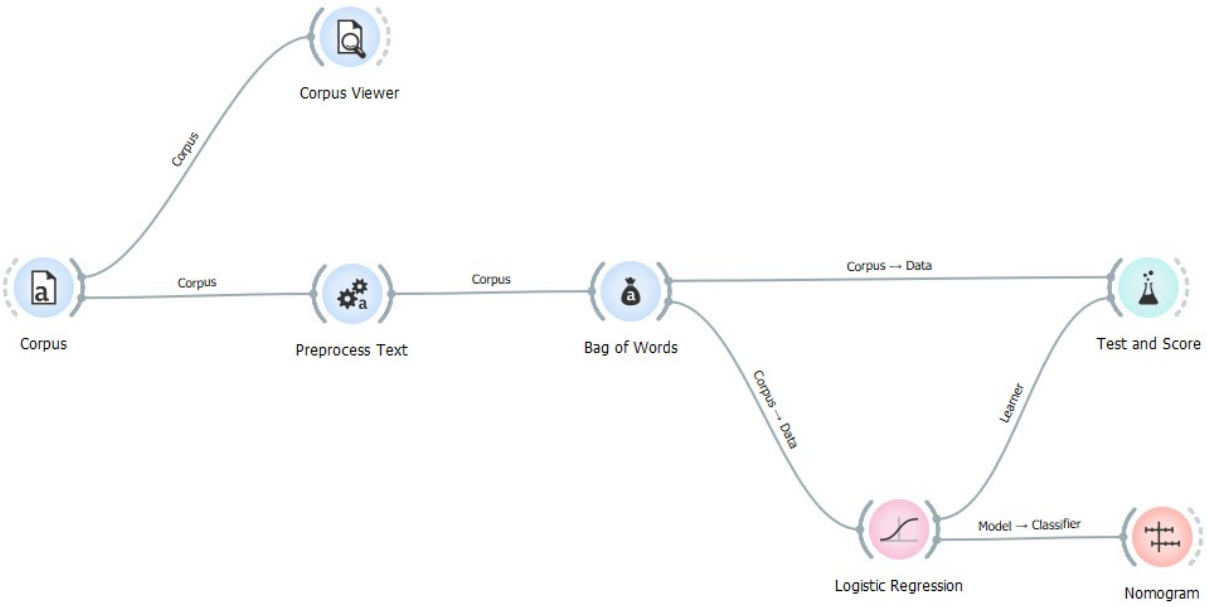
➢ We will use a `Nomogram` to visualise our model. Drag and drop it from the `Visualise` section onto the workspace and connect it to the `Linear Regression` model.


Visualize
This
Tree Viewer
Box Plot
Violin Plot
Distributions
Scatter Plot
Line Plot
Bar Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Silhouette Plo
Pythagorean Tree
Pythagorean Forest
CN2 Rule Viewer
Nomogram

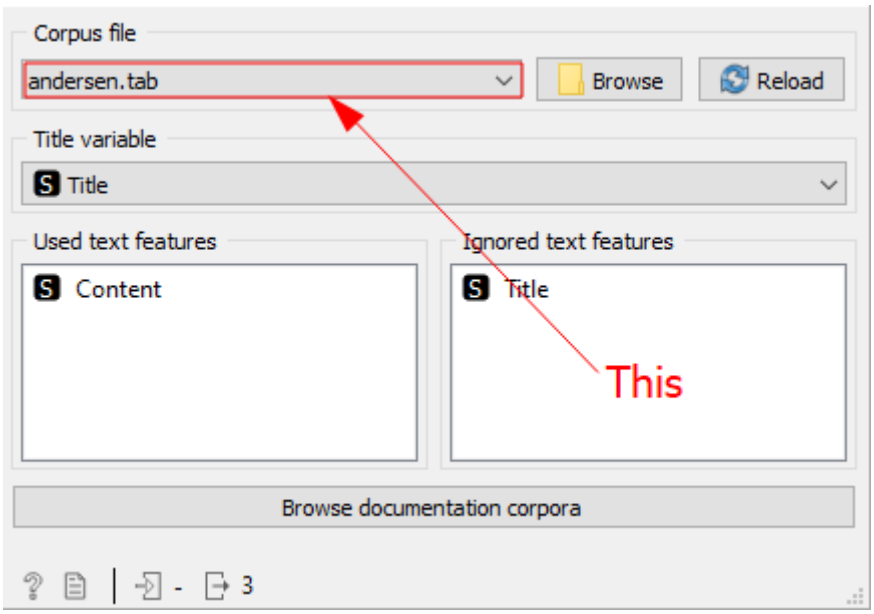

Corpus Viewer
Corpus
Corpus
Corpus
Corpus
Preprocess Text
Bag of Words
Corpus → Data
Test and Score
Corpus → Data
Learner
Model → Classifier
Logistic Regression
Nomogram



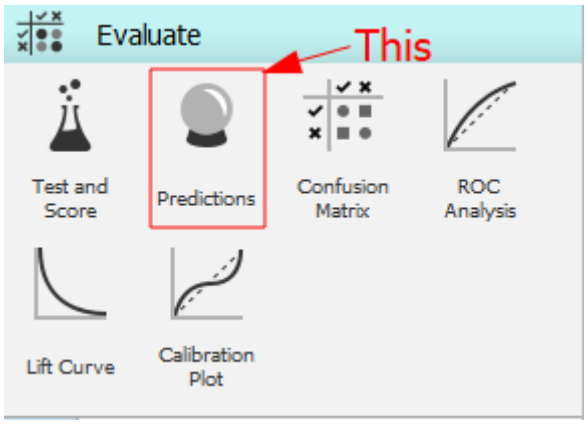
➢ To see whether the model works as intended, we create a new `Corpus` widget and set the file to `andersen.tab`.

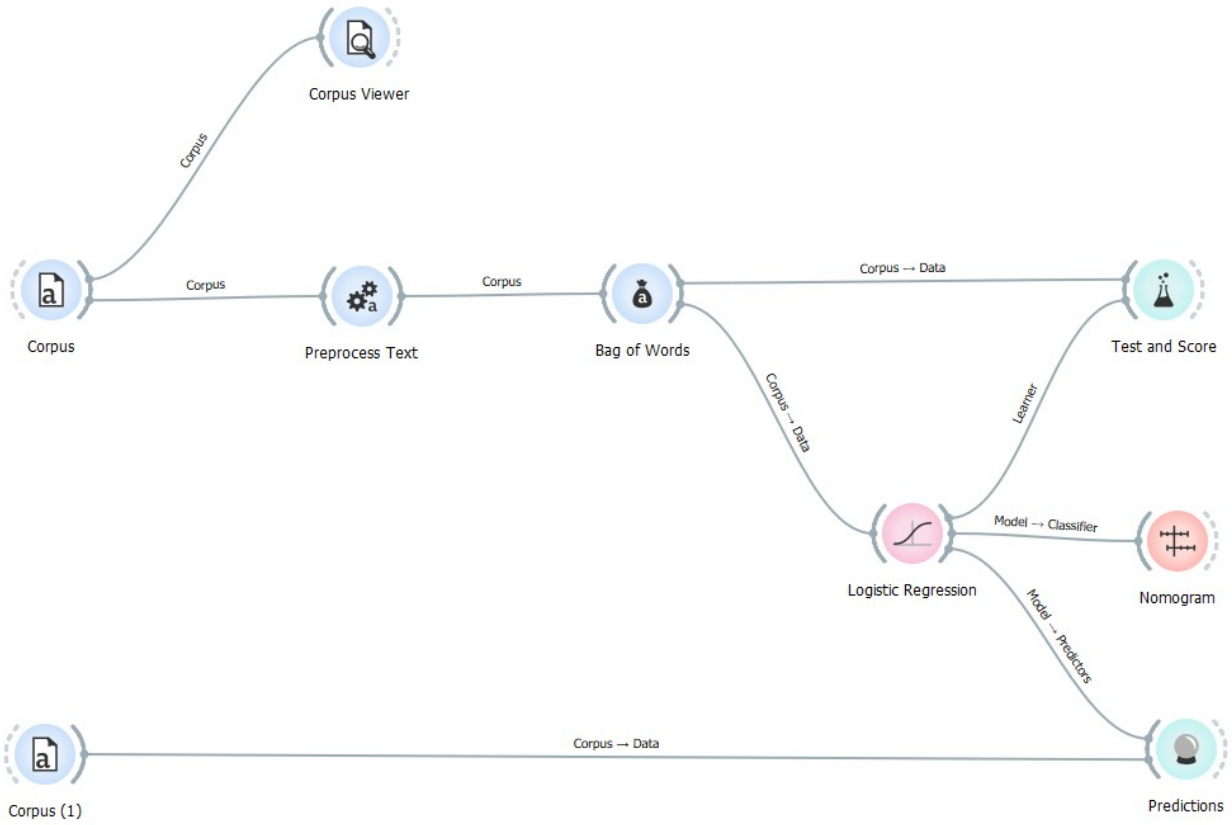

Corpus Viewer
Corpus
Corpus
Corpus
Corpus
Preprocess Text
Bag of Words
Corpus → Data
Test and Score
Corpus → Data
Learner
Model → Classifier
Logistic Regression
Nomogram

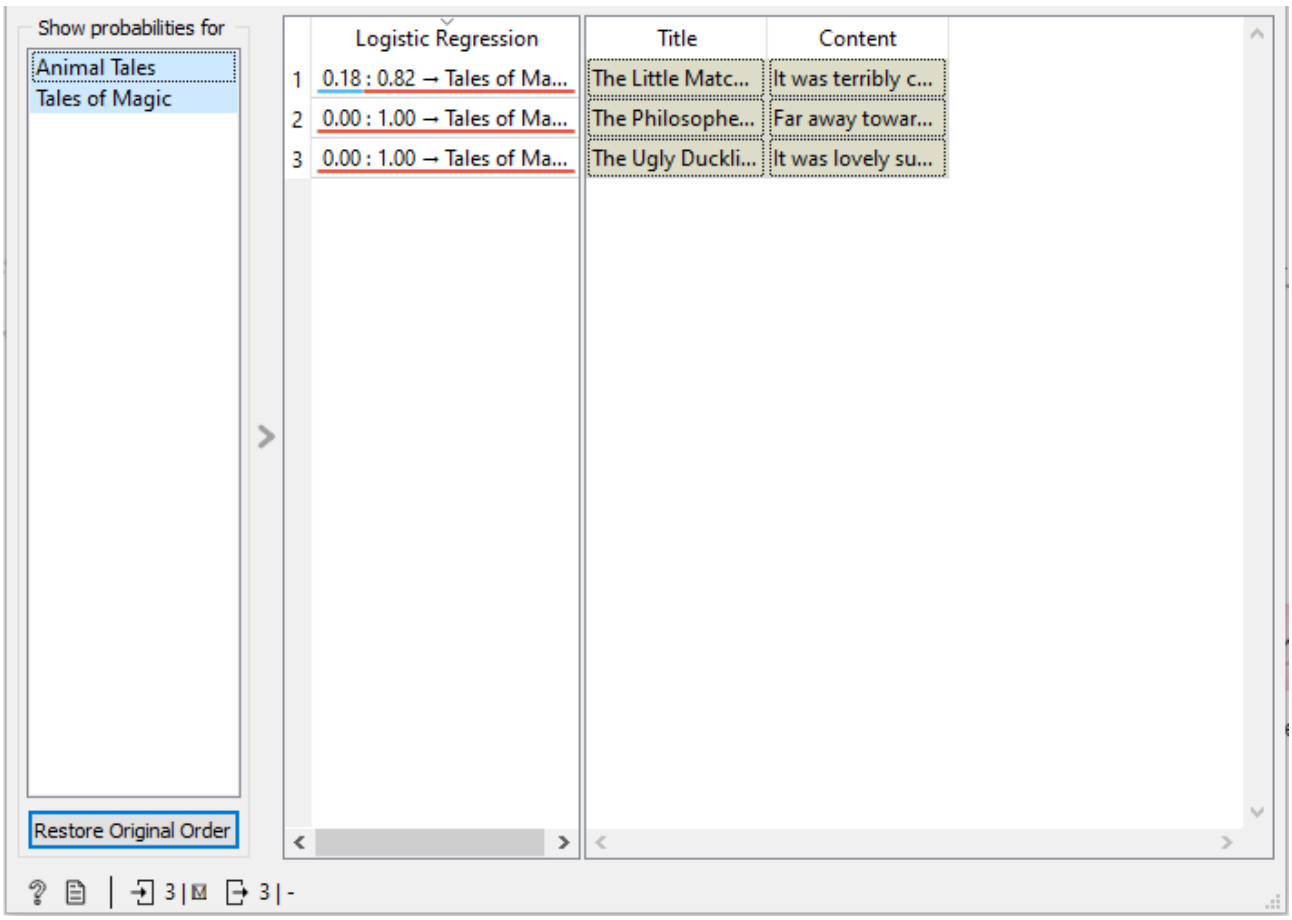
Corpus (1)


Corpus file
andersen.tab
Browse
Reload
Title variable
Title
Used text features
Content
Ignored text features
Title
This
Browse documentation corpora
3

➢ Now drag and drop a `Predictions` widget from the `Evaluate` section onto the workspace. Connect it to the model as well as the new `Corpus` widget to visualise the results.





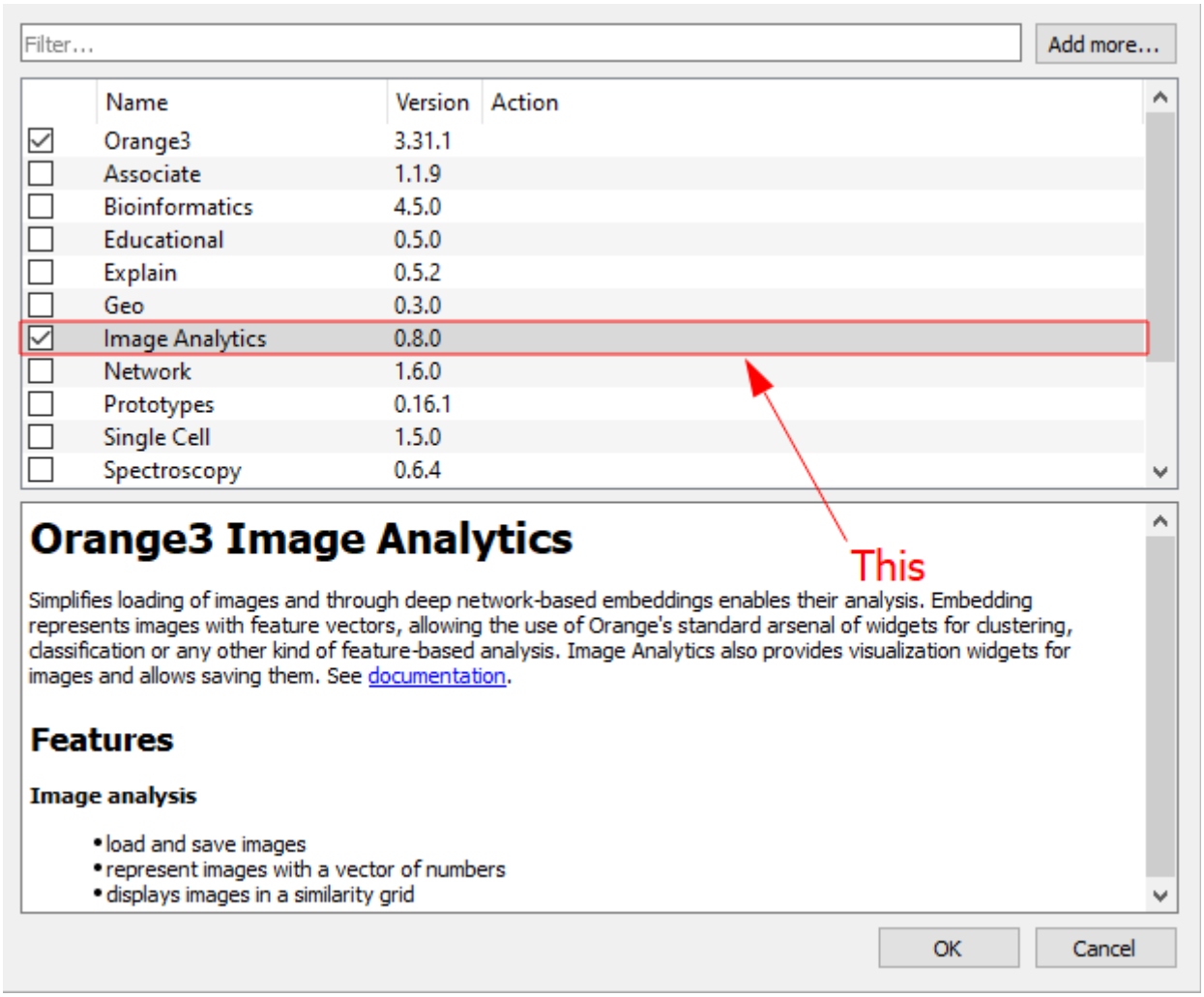

Show probabilities for
Animal Tales
Tales of Magic
Restore Original Order
Logistic Regression
1 0.18 : 0.82 → Tales of Ma...
2 0.00 : 1.00 → Tales of Ma...
3 0.00 : 1.00 → Tales of Ma...
Title
Content
The Little Matc...
It was terribly c...
The Philosophe...
Far away towar...
The Ugly Duckli...
It was lovely su...
3 | M
3 | -

Practical 3

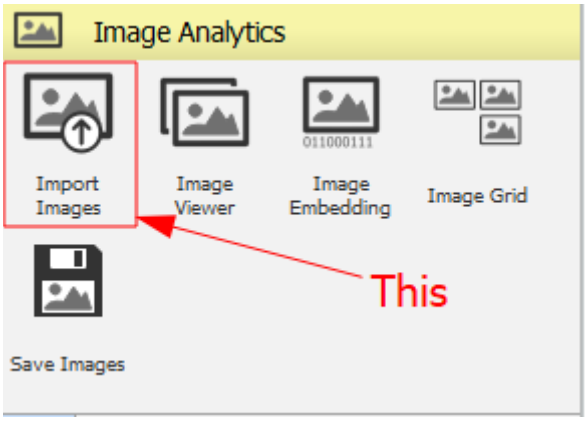
Aim: Image classification using orange.

Procedure:

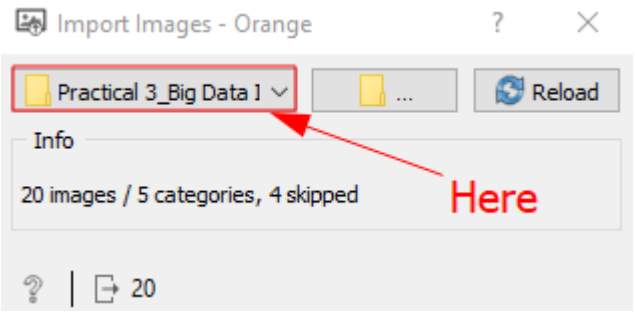
- Before starting, we need to install the `ImageProcessing` add-on. Install it by navigating to Options > Add ons…. You will be prompted to restart Orange after the installation completes. If it is already installed, skip this step.



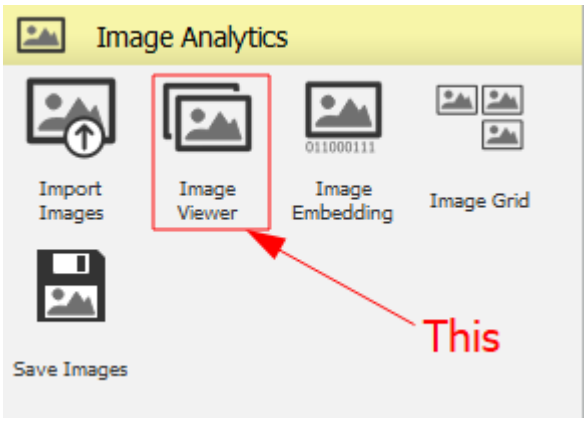
➢ Add a `Import Images` widget from the newly added `Image Analytics` section to the workspace.

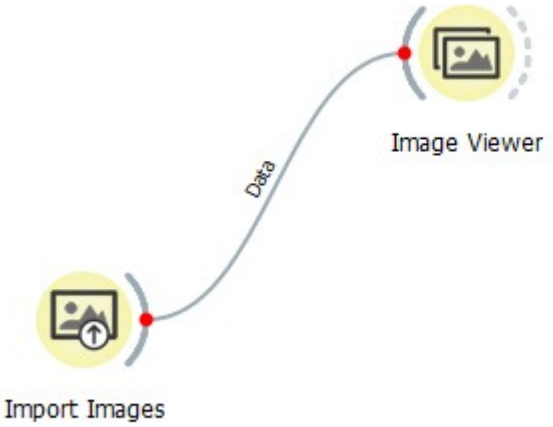


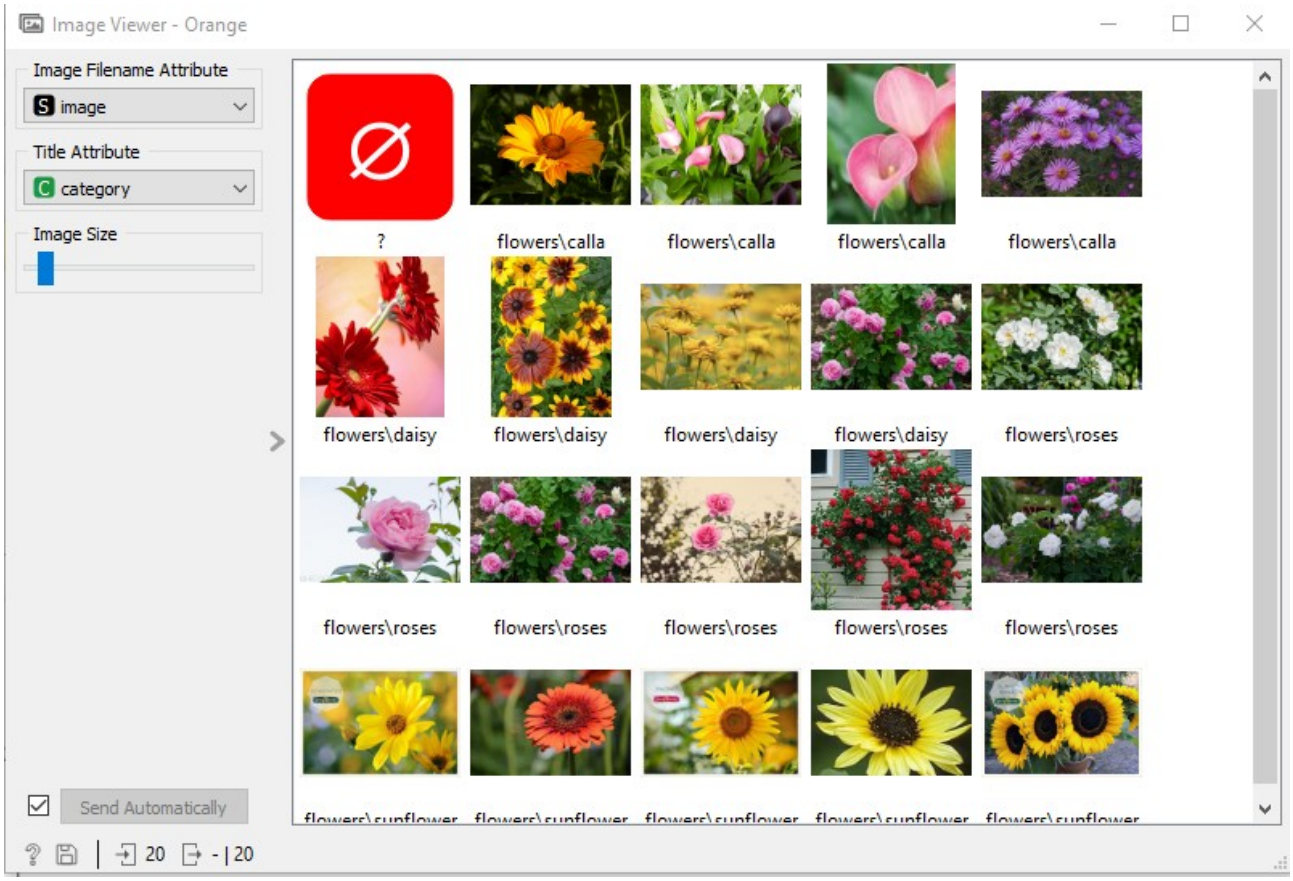
➢ Load the directory containing the images in the newly added `Import Images` widget by double clicking the widget.



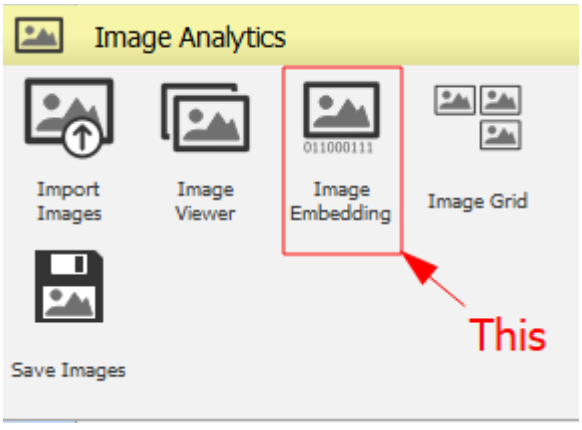
➢ We can view these images using the `Image Viewer` widget found in the `Image Analytics` section. Drag and drop this widget onto the workspace and connect it to the `Import Images` widget.

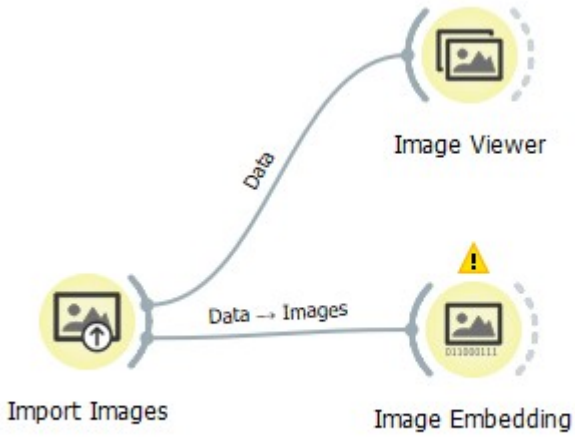




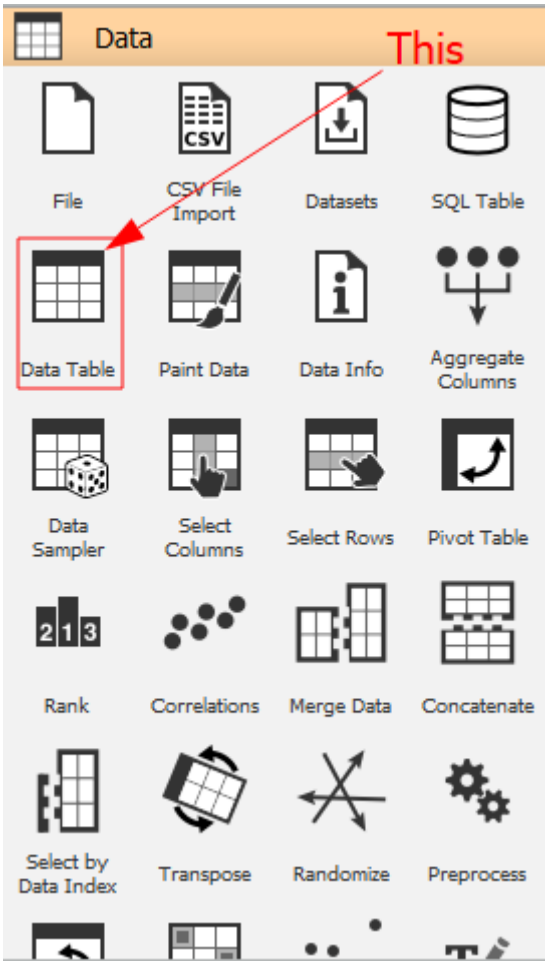


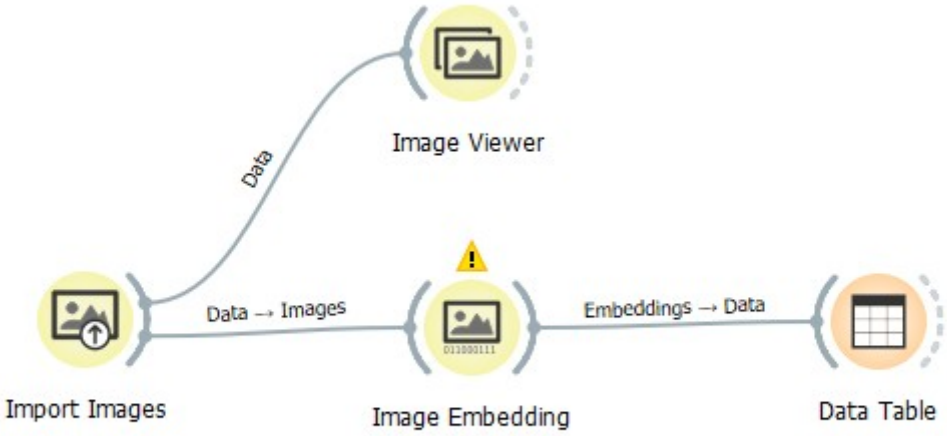
- As our models can only process numbers, we need to convert the images into numerical data. This is where `Image Embedding` comes into play. Drag and drop a `Image Embedding` widget from the `Image Analytics` section onto the workspace and connect it to the `Import Images` widget.





- We use a `Data Table` to visualise the tabular data generated by the `Image Embedding` widget. Drag and drop a `Data Table` widget from the `Data` section onto the workspace and connect it to the Image Embedding widget.





Info
18 instances (no missing data)
1000 features
Target with 5 values
5 meta attributes

Variables
- [x] Show variable labels (if present)
- [ ] Visualize numeric values
- [x] Color by instance classes

Selection
- [x] Select full rows

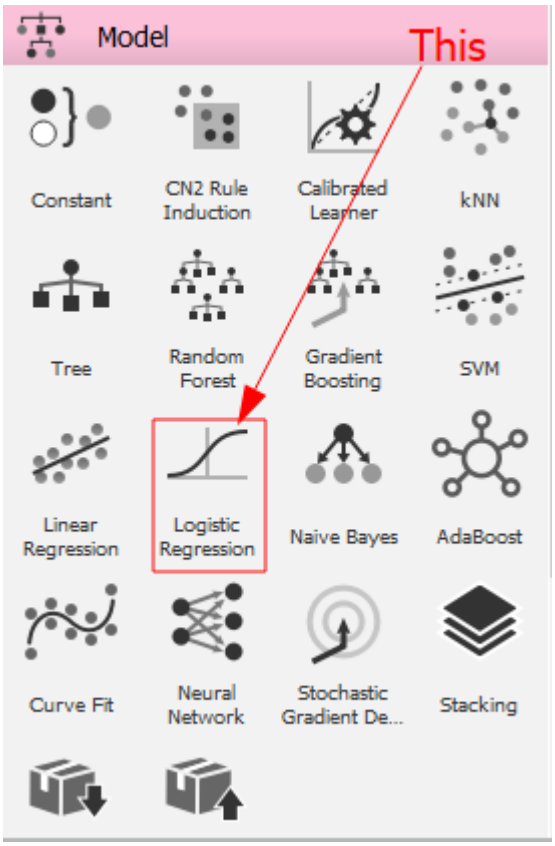
Restore Original Order

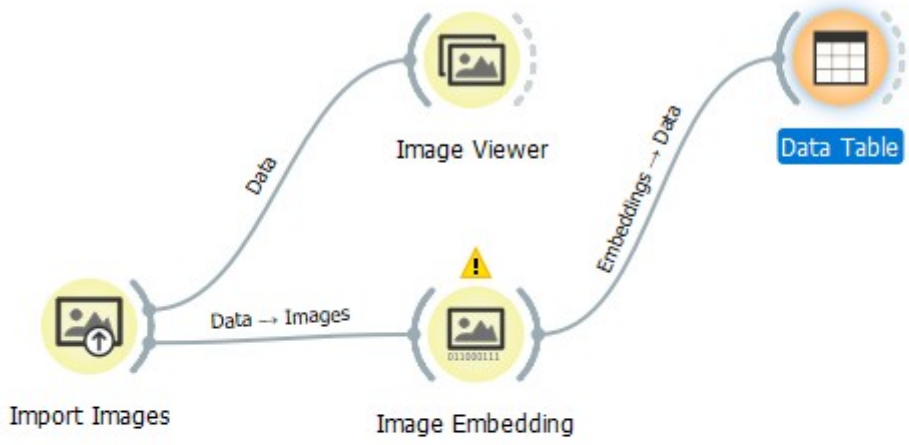
- [x] Send Automatically

18 | 18 | 18

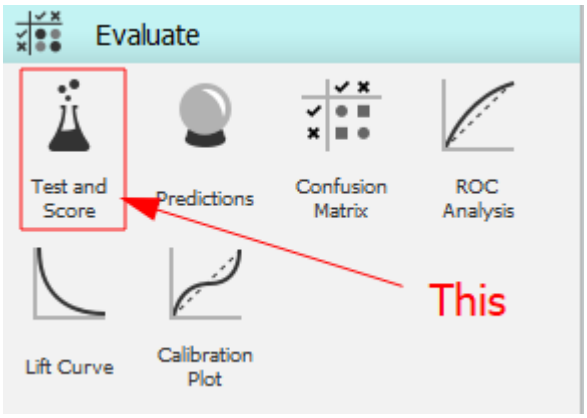
hidden origin	category	image name	image _Big Data Image An image	size	width
1	flowers\calla	calla	flowers\calla\c...	3429626	2560
2	flowers\calla	calla1	flowers\calla\c...	29712	400
3	flowers\calla	calla2	flowers\calla\c...	2890	105
4	flowers\daisy	daisy1	flowers\daisy\d...	39660	400
5	flowers\daisy	daisy4	flowers\daisy\d...	71325	375
6	flowers\daisy	daisy6	flowers\daisy\d...	3206304	2084
7	flowers\daisy	daisy8	flowers\daisy\d...	526653	933
8	flowers\roses	rose1	flowers\roses\r...	535368	933
9	flowers\roses	rose2	flowers\roses\r...	354357	933
10	flowers\roses	rose3	flowers\roses\r...	526653	933
11	flowers\roses	rose4	flowers\roses\r...	429411	933
12	flowers\roses	rose6	flowers\roses\r...	841733	700
13	flowers\roses	rose7	flowers\roses\r...	523616	933
14	flowers\sunflo...	sun1	flowers\sunflo...	6649	170
15	flowers\sunflo...	sun2	flowers\sunflo...	5481	176
16	flowers\sunflo...	sun3	flowers\sunflo...	7061	170
17	flowers\sunflo...	sun4	flowers\sunflo...	5808	177
18	flowers\sunflo...	sun5	flowers\sunflo...	222011	800

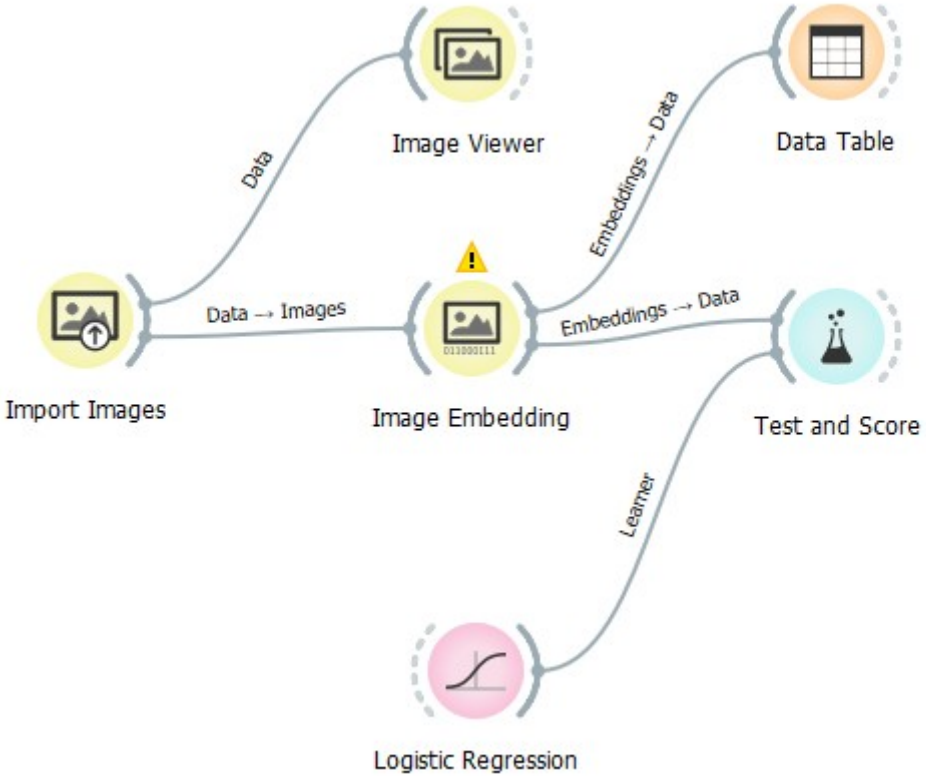
➢ We will use `Logistic Regression` as our model. Drag and drop a `Logistic Regression` widget from the `Model` section onto the workspace.





- Drag and drop a `Test and Score` widget from the `Evaluate` section onto the workspace. This widget will allow us to verify our model. Connect this widget to the `Image Embedding` and `Logistic Regression` widgets.





Sampling

- Cross validation
  - Number of folds: 3
  - Stratified
- Cross validation by feature
- Random sampling
  - Repeat train/test: 10
  - Training set size: 66 %
  - Stratified
- Leave one out
- Test on train data
- Test on test data

Target Class

(Average over classes)

Model Comparison

Area under ROC curve

Negligible difference: 0.1

Evaluation Results

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.718	0.556	0.519	0.543	0.556

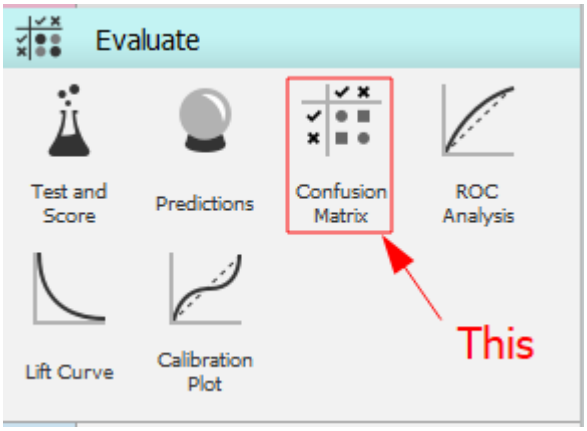
Model Comparison by AUC

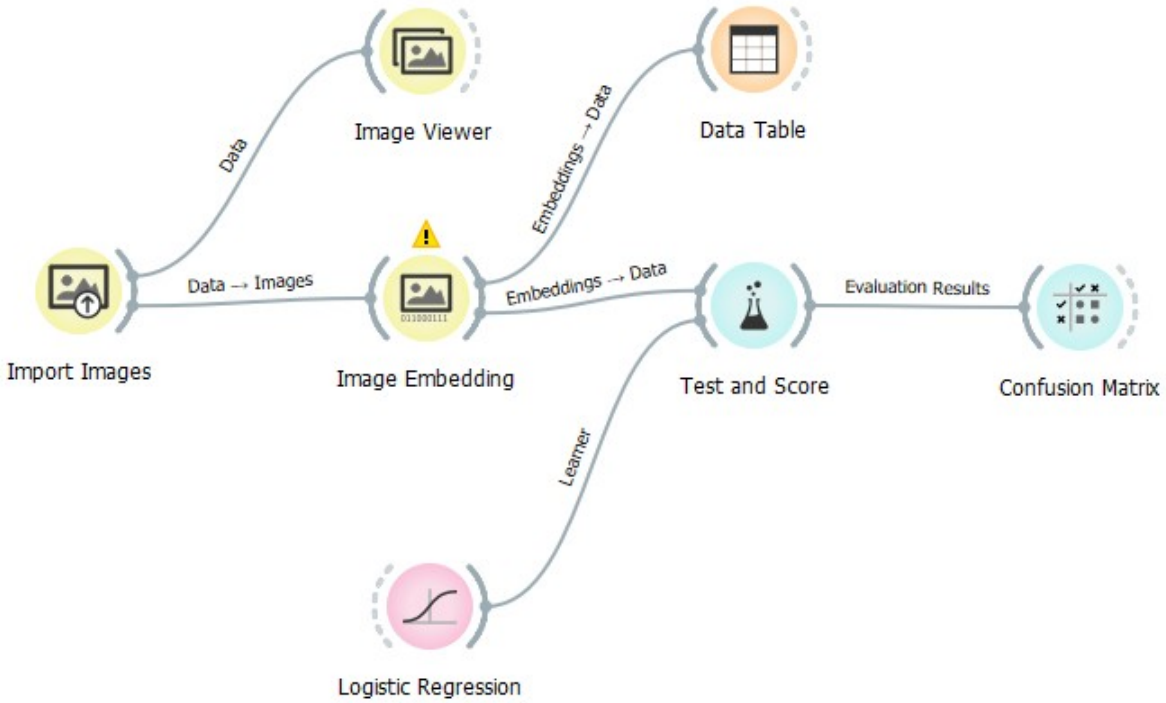
	Logistic ...
Logistic Regression	

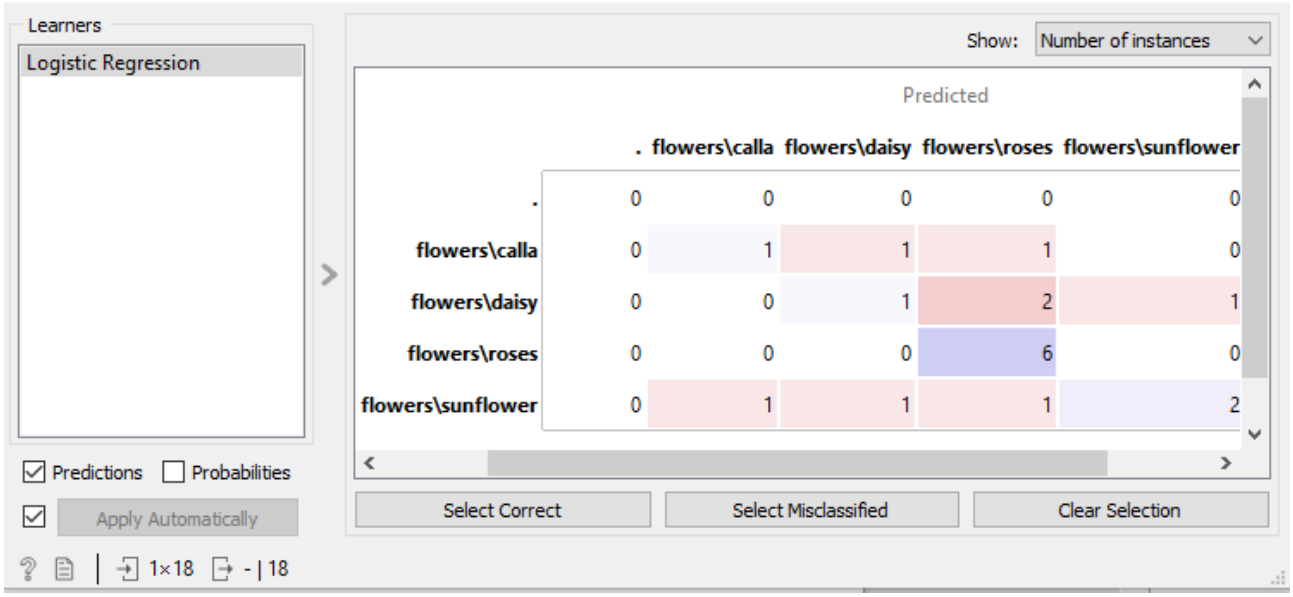
Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

18 | - | - 18 | 1×18

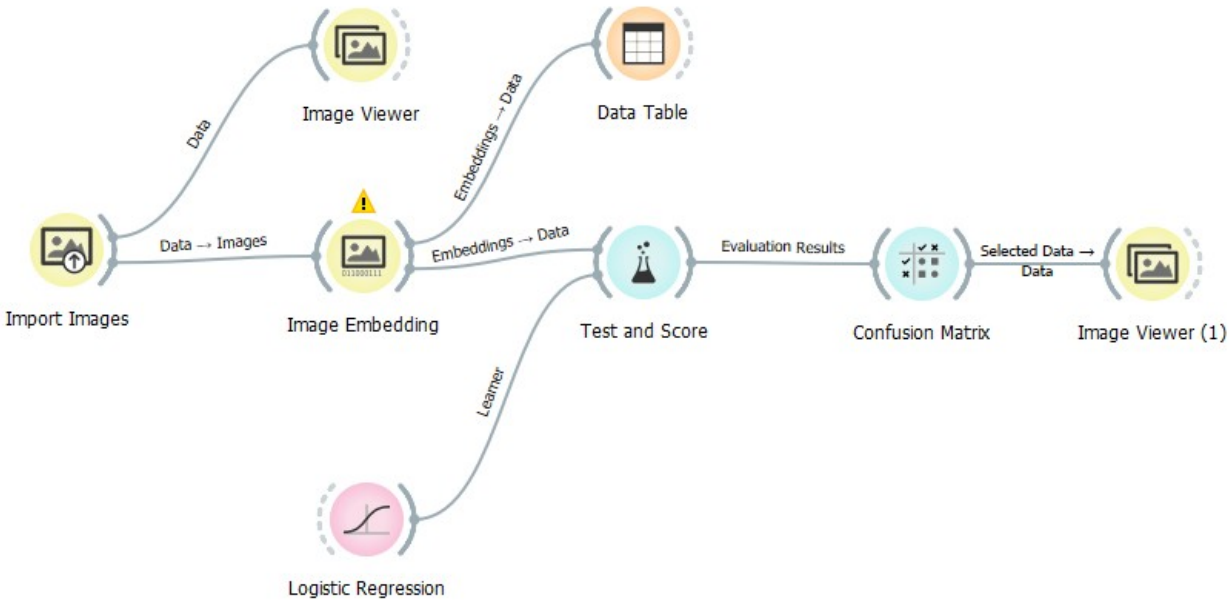
➢ We now drag and drop a `Confusion Matrix` widget from the `Evaluate` section onto the workspace and connect it to the `Test and Score` widget.

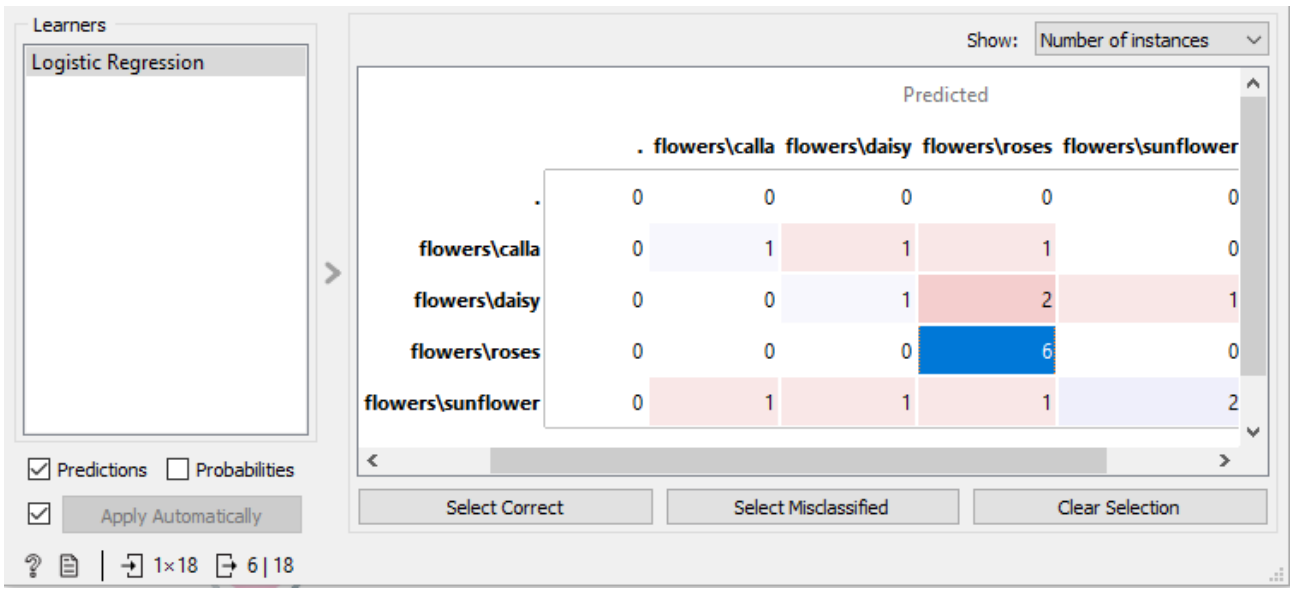






➢ Drag and drop another `Image Viewer` widget to view the selected cell(s) from the `Confusion Matrix`.




Learners
Logistic Regression
Show:
Number of instances
Predicted
. flowers\calla flowers\daisy flowers\roses flowers\sunflower
. 0 0 0 0 0
flowers\calla 0 1 1 1 0
flowers\daisy 0 0 1 2 1
flowers\roses 0 0 0 6 0
flowers\sunflower 0 1 1 1 2
Predictions
Probabilities
Apply Automatically
1×18
6 | 18
Select Correct
Select Misclassified
Clear Selection

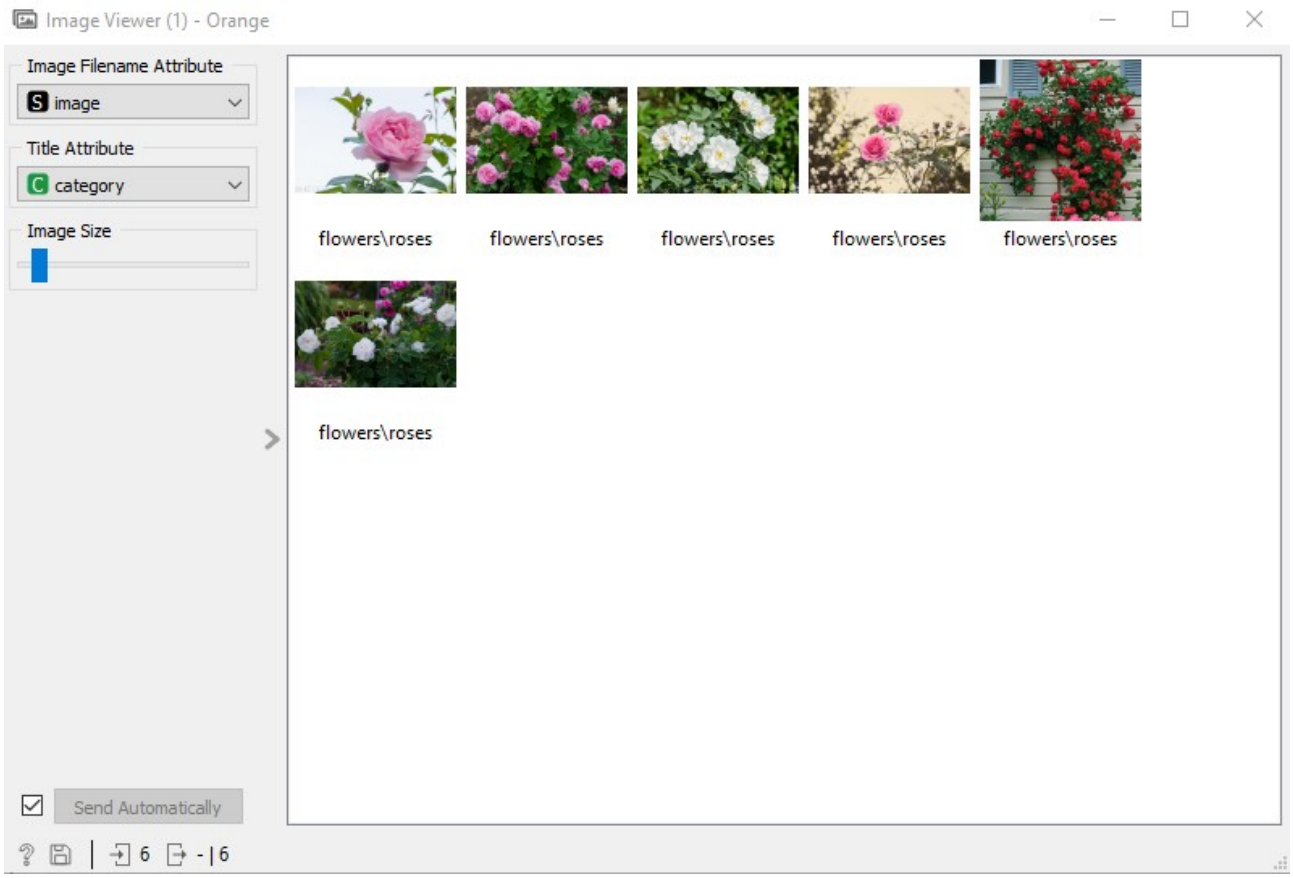


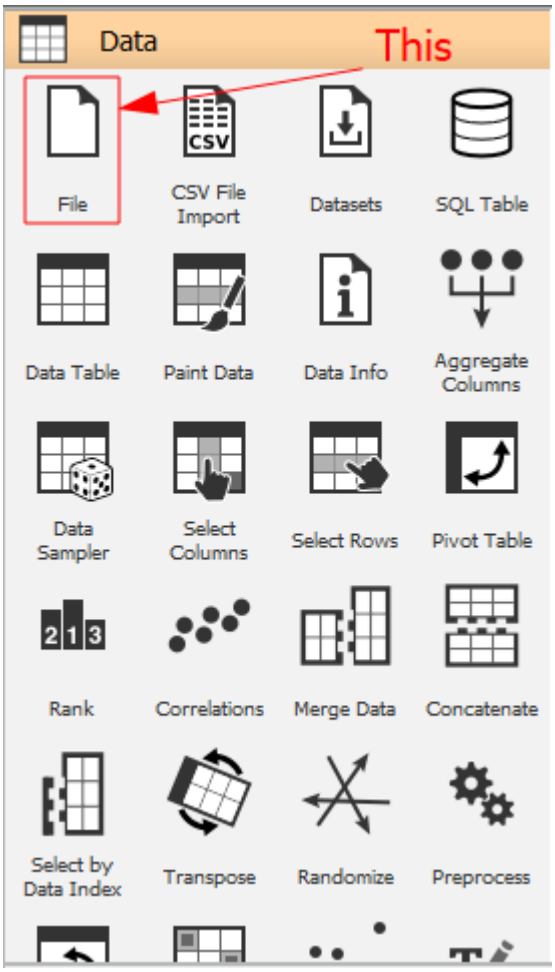
Image Viewer (1) - Orange
Image Filename Attribute
image
Title Attribute
category
Image Size
flowers\roses
flowers\roses
flowers\roses
flowers\roses
flowers\roses
flowers\roses
Send Automatically
6
- | 6

Practical 4

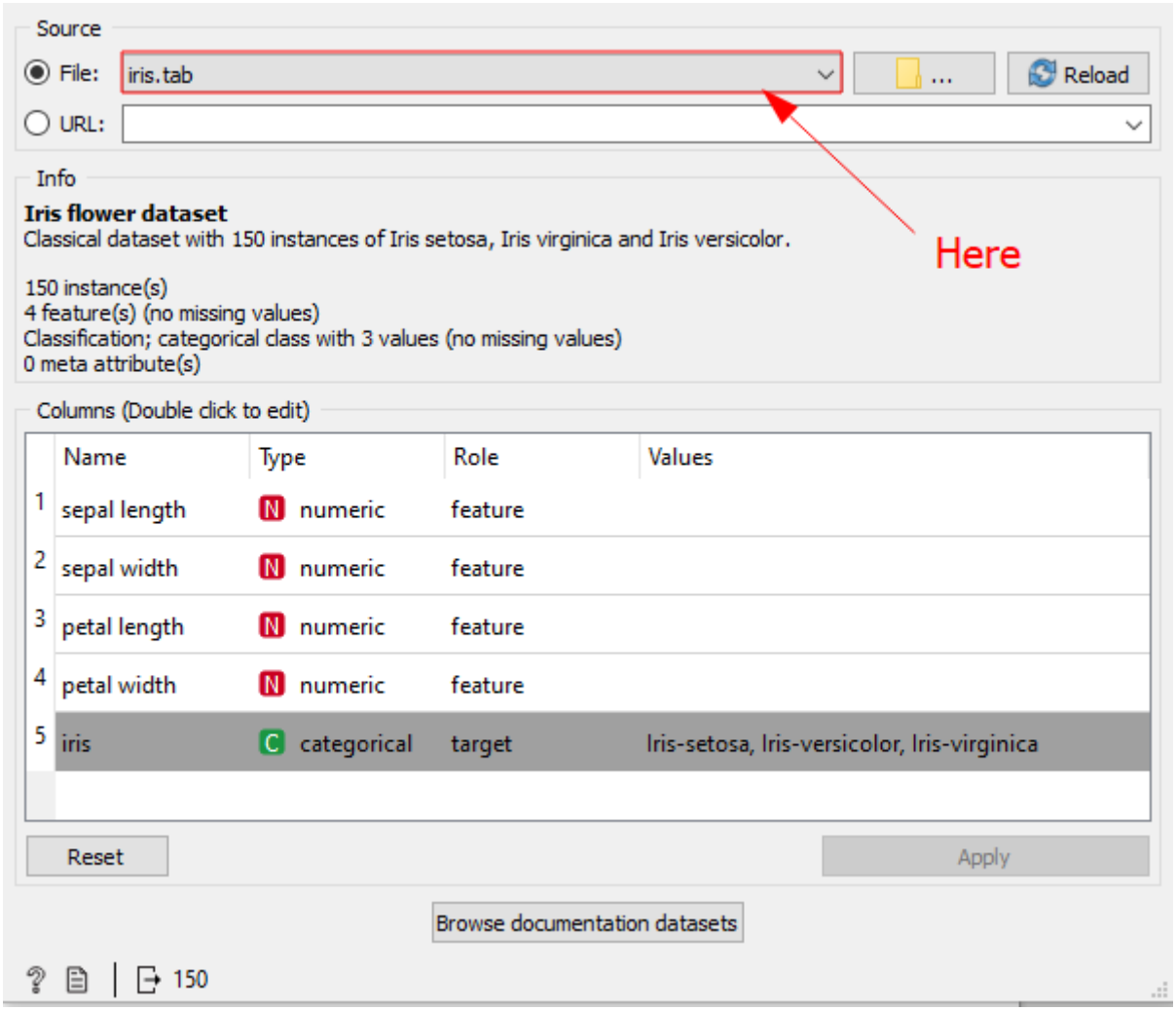
Aim: Hierarchical clustering using orange.

Procedure:

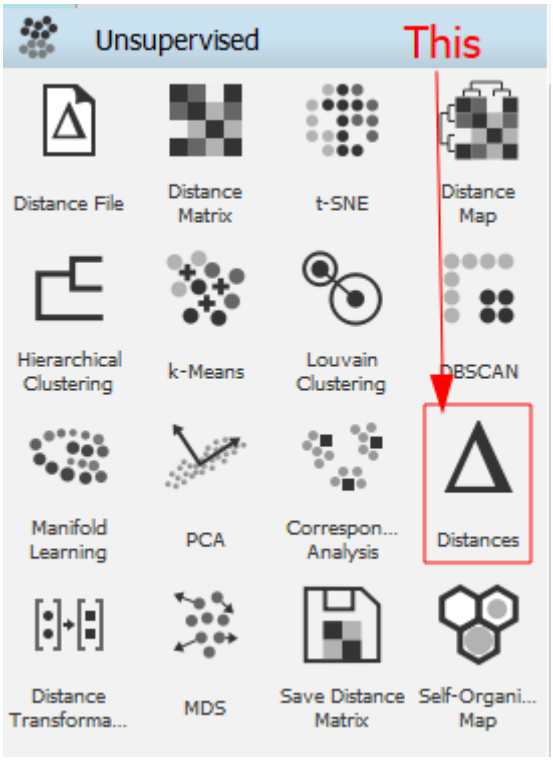
- Drag and drop a `File` widget from the `Data` section onto the workspace.



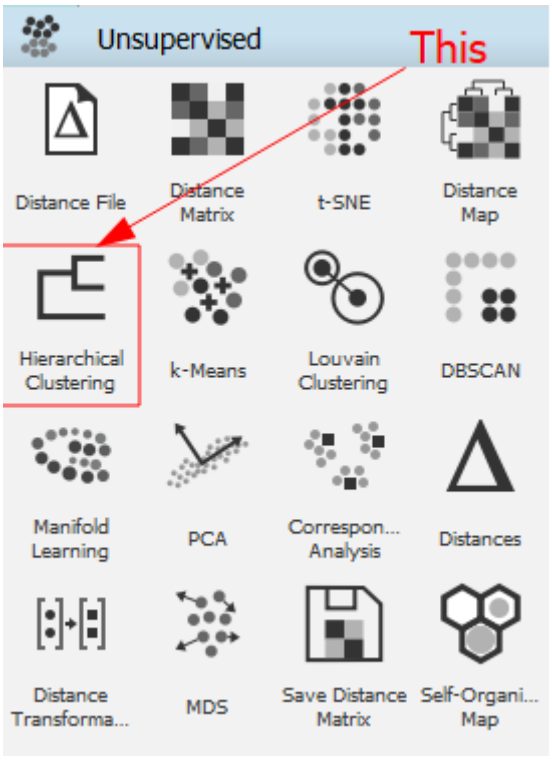
➢ Double click the `File` widget and set the name to `iris.tab`.



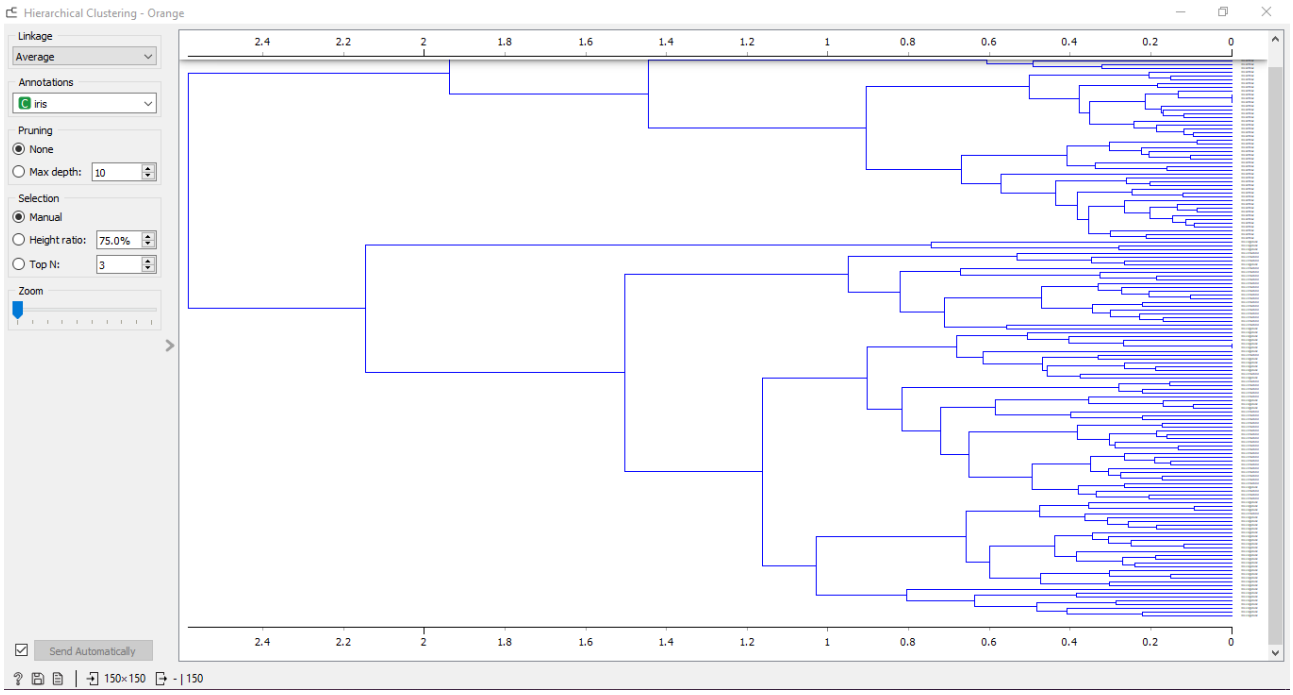
➢ Drag and drop the `Distances` widget from the `Unsupervised` section onto the workspace and connect it to the File widget.

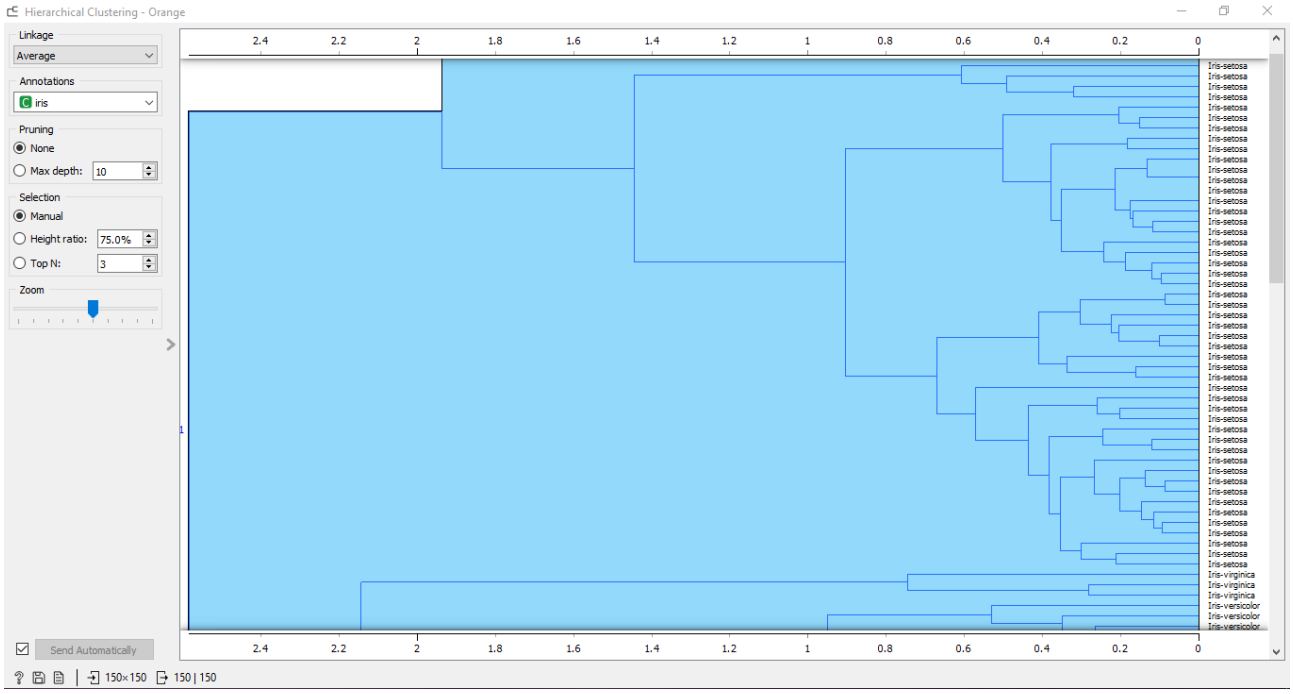


➢ Drag and drop a `Hierarchical Clustering` widget from the `Unsupervised` section to the workspace and connect it to the `Distances` widget.

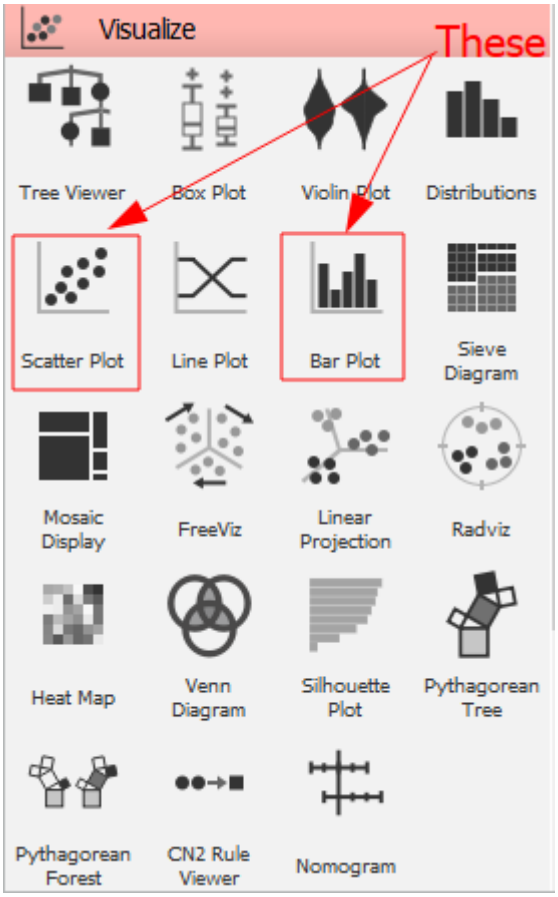


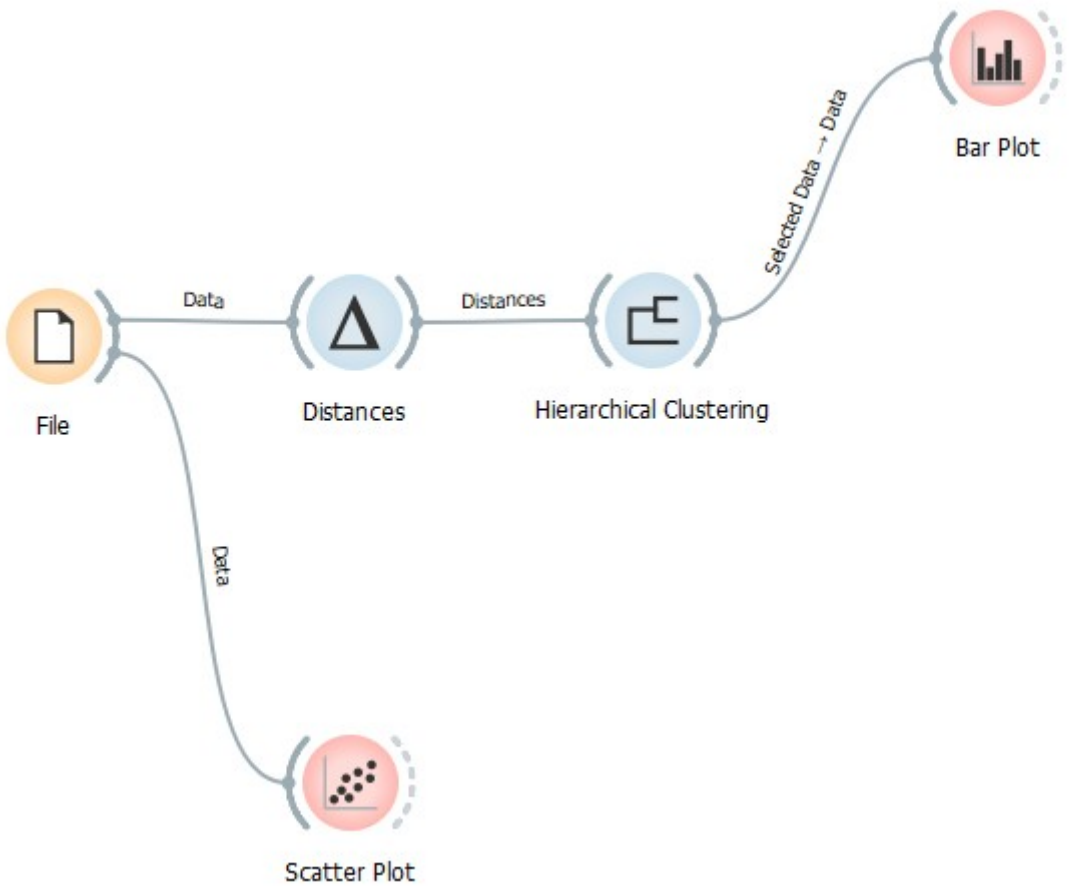
➢ Double click the `Hierarchical Clustering` widget to view the dendogram. Select any sub-cluster as per convenience.

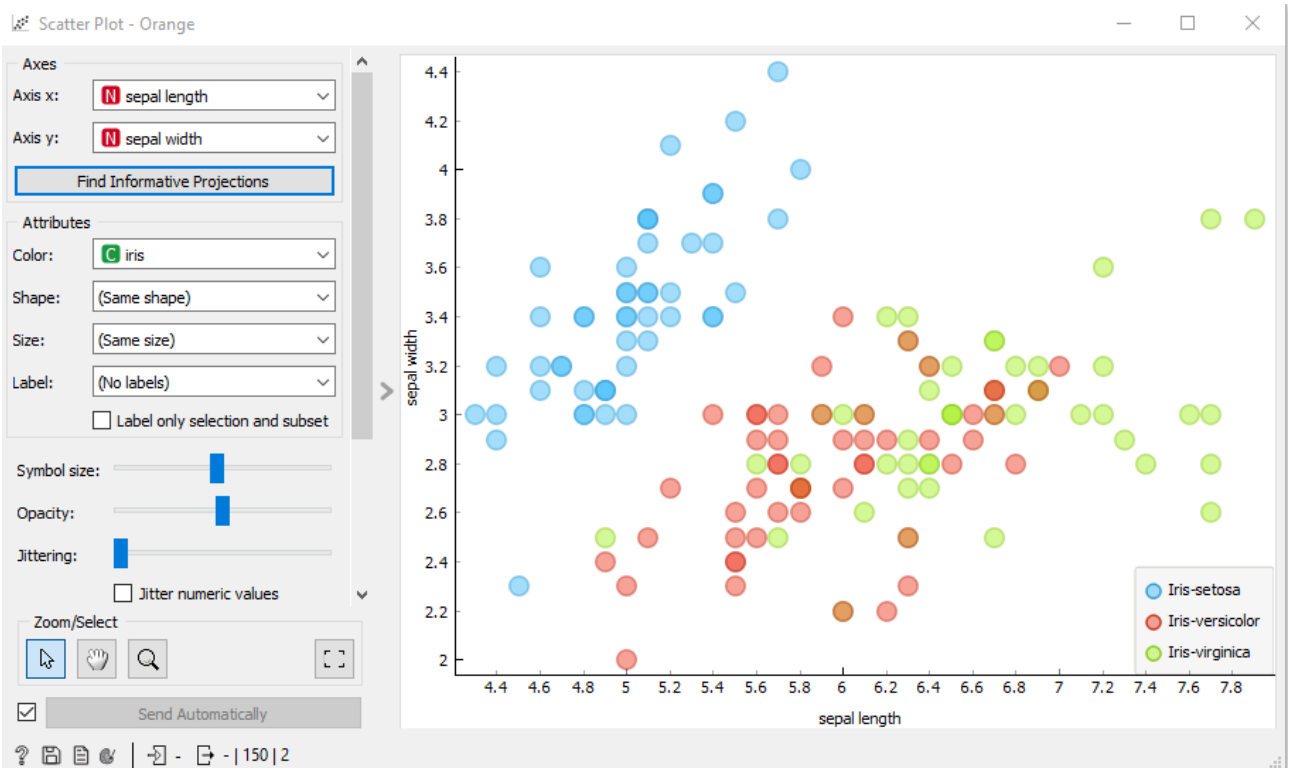


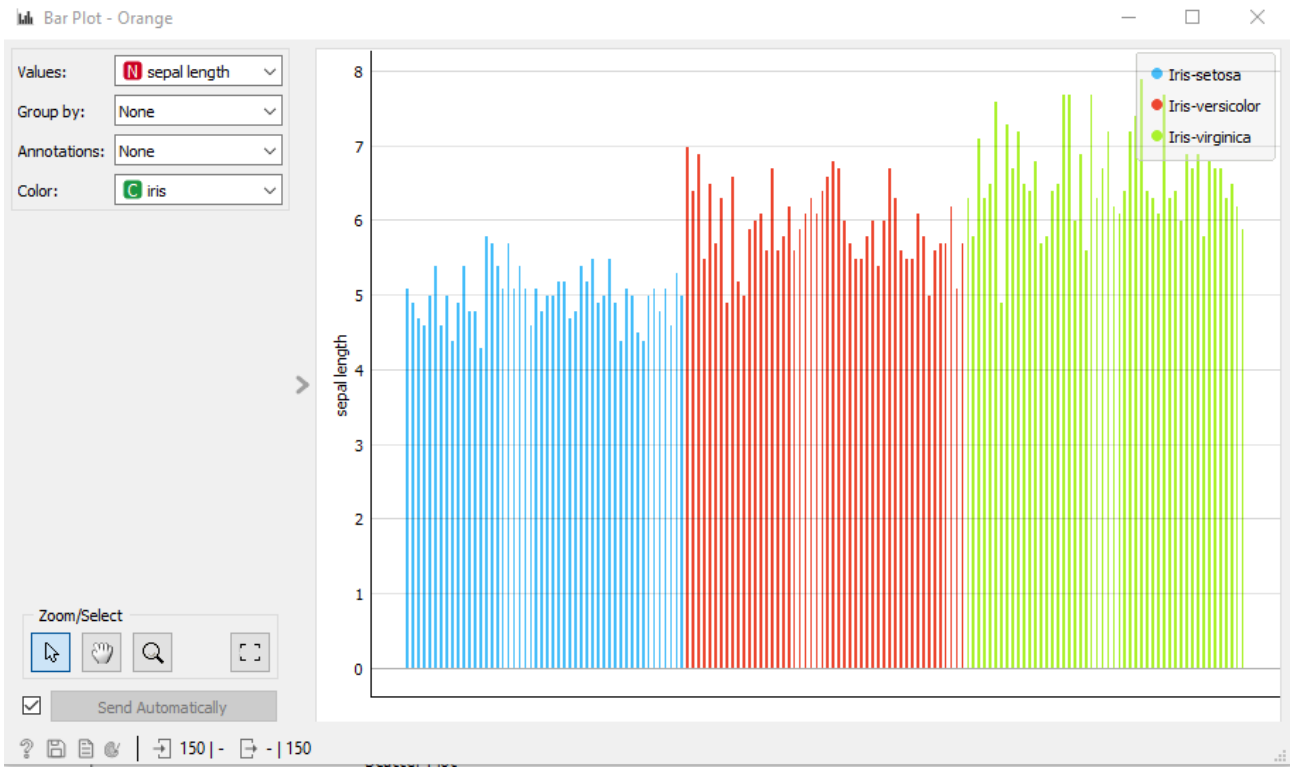


➢ Choose any visualization method of your choice. This example assumes a `Bar Plot` for the sub-cluster and a `Scatter Plot` for the source dataset. You can find both of these widgets in the `Visualise` section. Connect the `Bar Plot` to the `Hierarchical Clustering` widget and `Scatter Plot` to the `File` widget.




Bar Plot
Selected Data → Data
Data
Distances
File
Distances
Hierarchical Clustering
Data
Scatter Plot


Scatter Plot - Orange
Axes
Axis x: sepal length
Axis y: sepal width
Find Informative Projections
Attributes
Color: iris
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Jitter numeric values
Zoom/Select
Send Automatically
- | 150 | 2
sepal width
sepal length
Iris-setosa
Iris-versicolor
Iris-virginica

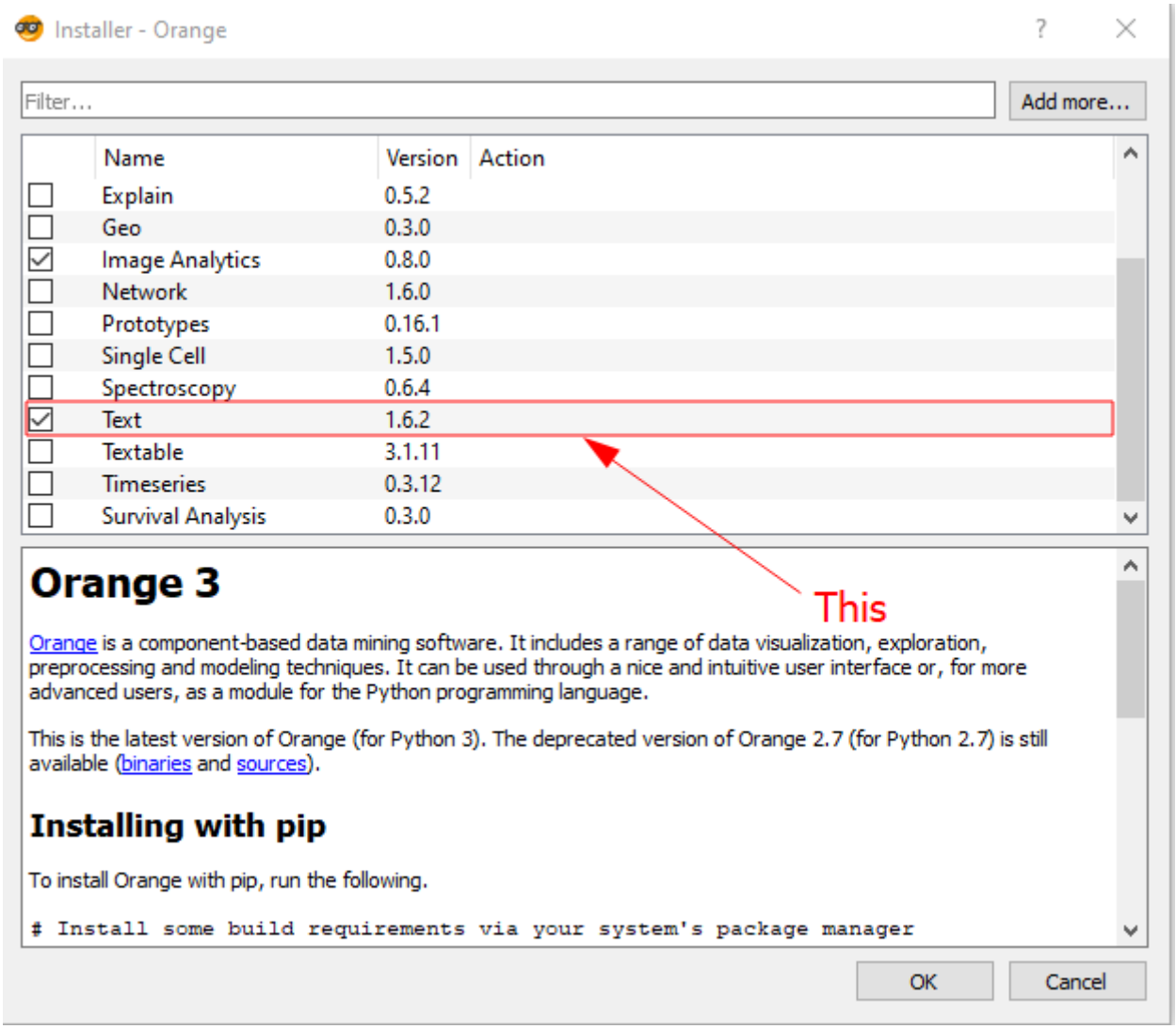

Bar Plot - Orange
Values:
sepal length
Group by:
None
Annotations:
None
Color:
iris
Zoom/Select
Send Automatically
150 | -
- | 150
sepal length
0
1
2
3
4
5
6
7
8
Iris-setosa
Iris-versicolor
Iris-virginica

Practical 5

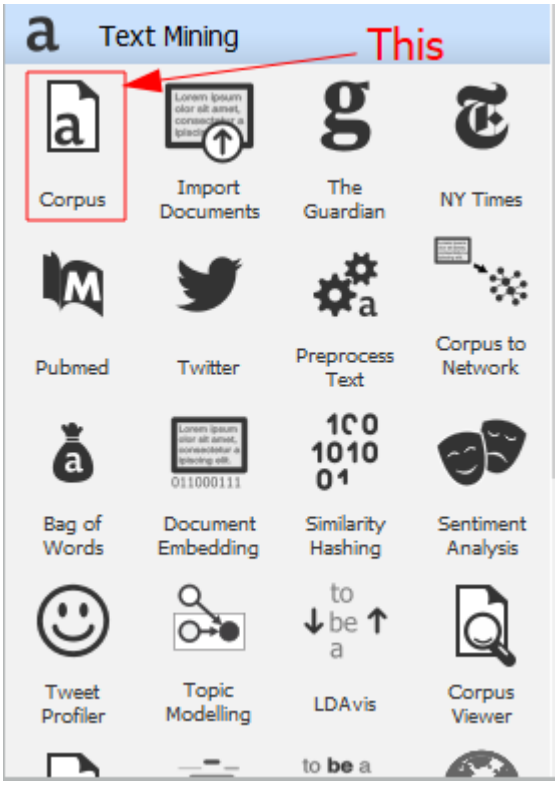
Aim: Hierarchical clustering using orange.

Procedure:

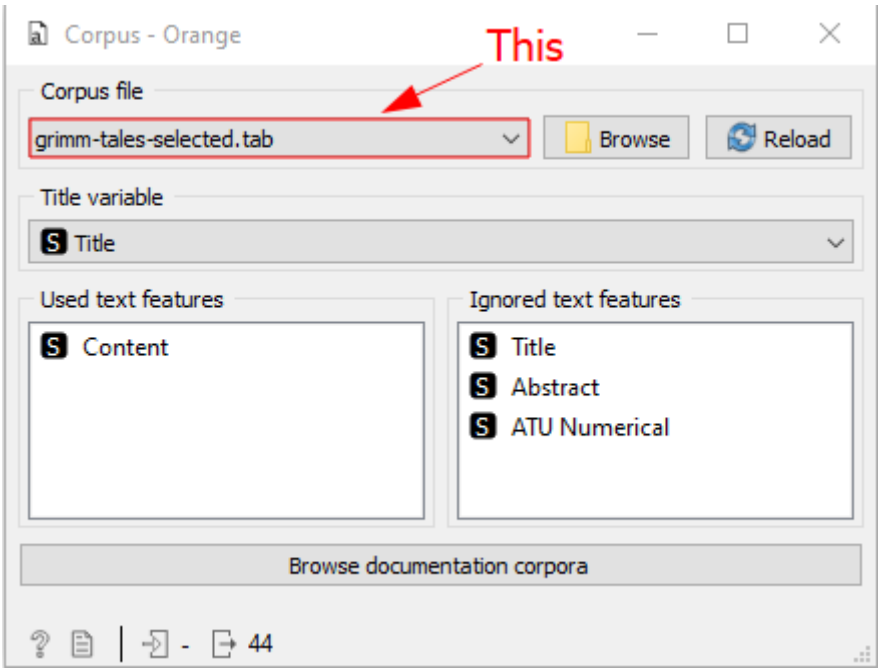
- Before starting, we need to install the `Text` add-on. Install it by navigating to Options > Add ons…. You will be prompted to restart Orange after the installation completes. If it is already installed, skip this step.



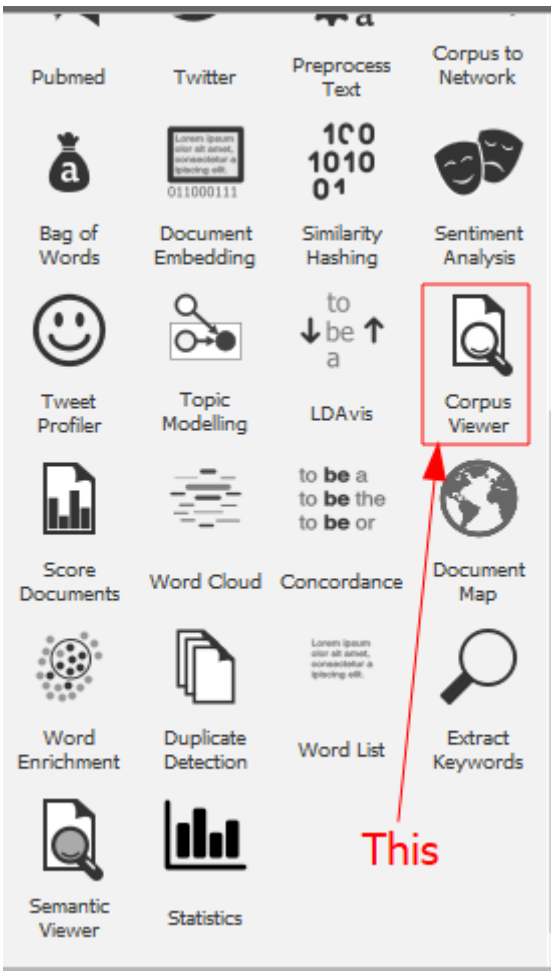
➢ Drag and drop a `Corpus` widget to the workspace. It can be found in the newly added `Text Mining` section.

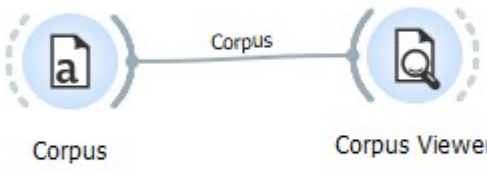


➢ Double click the `Corpus` widget and select the `grimm-tales-selected.tab` corpus file.

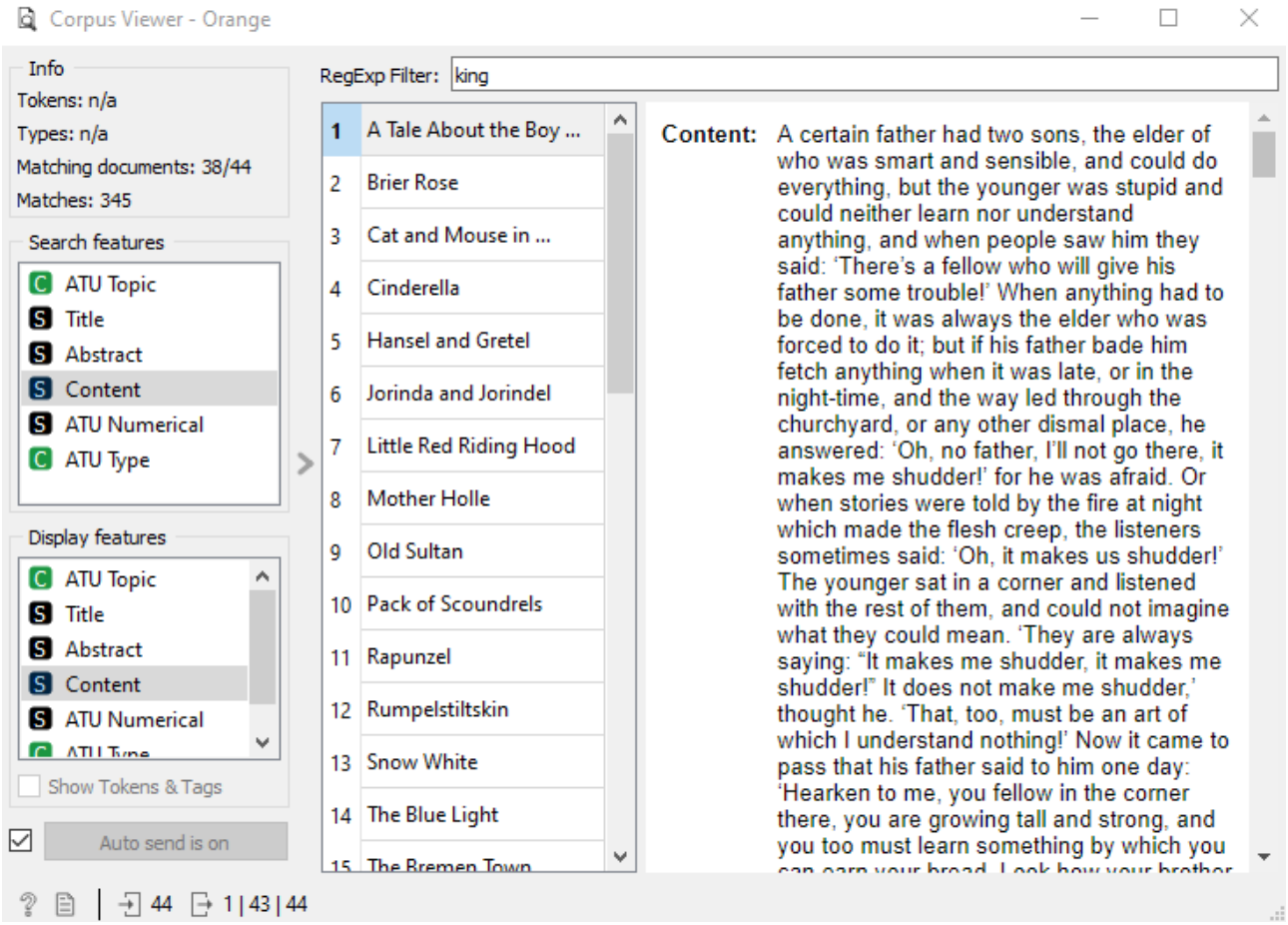


➢ Drag and drop a `Corpus Viewer` widget and connect it to the `Corpus` widget. The `Corpus Viewer` widget can be found in the `Text Mining` section.


Pubmed
Twitter
Preprocess Text
Corpus to Network
Bag of Words
Document Embedding
Similarity Hashing
Sentiment Analysis
Tweet Profiler
Topic Modelling
LDAvis
Corpus Viewer
Score Documents
Word Cloud
Concordance
Document Map
Word Enrichment
Duplicate Detection
Word List
Extract Keywords
Semantic Viewer
Statistics
This

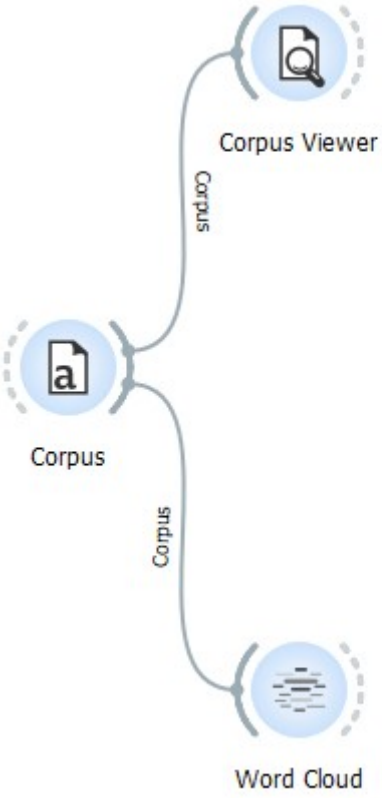

Corpus
Corpus
Corpus Viewer

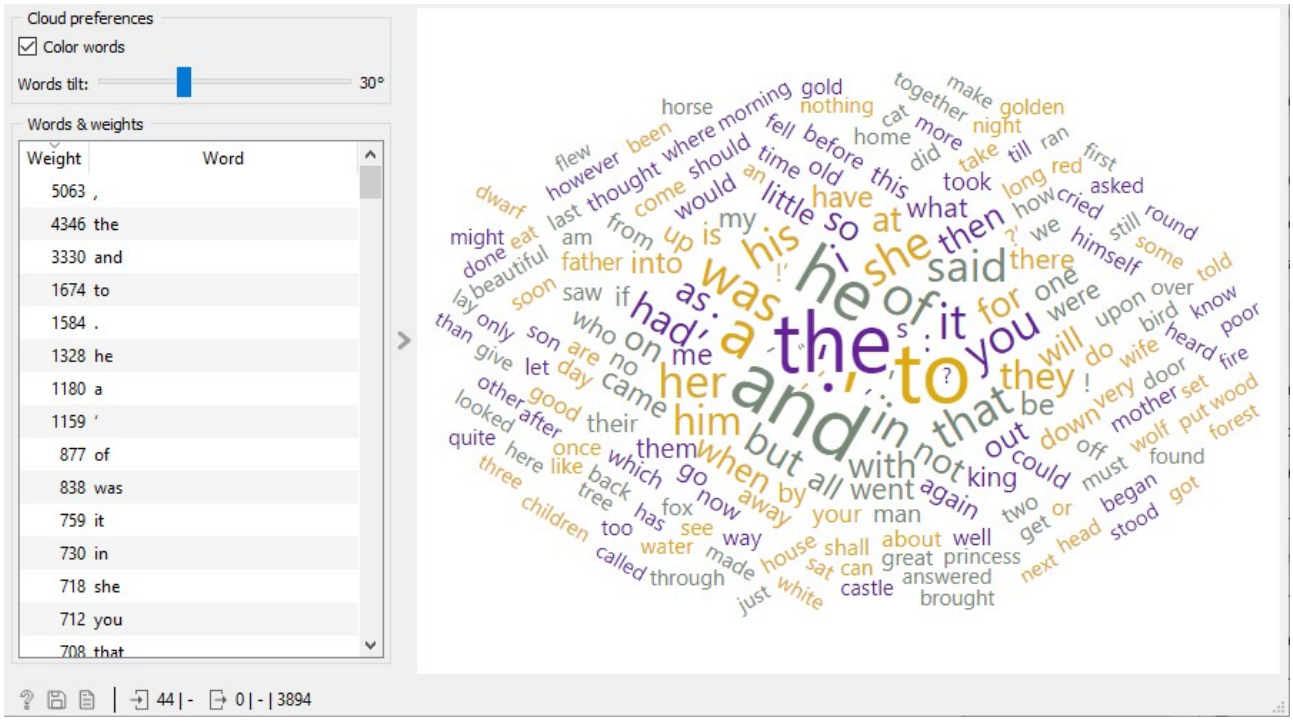
- Double-clicking the `Corpus Viewer` allows us to peek through the corpus and also allows us to filter text.



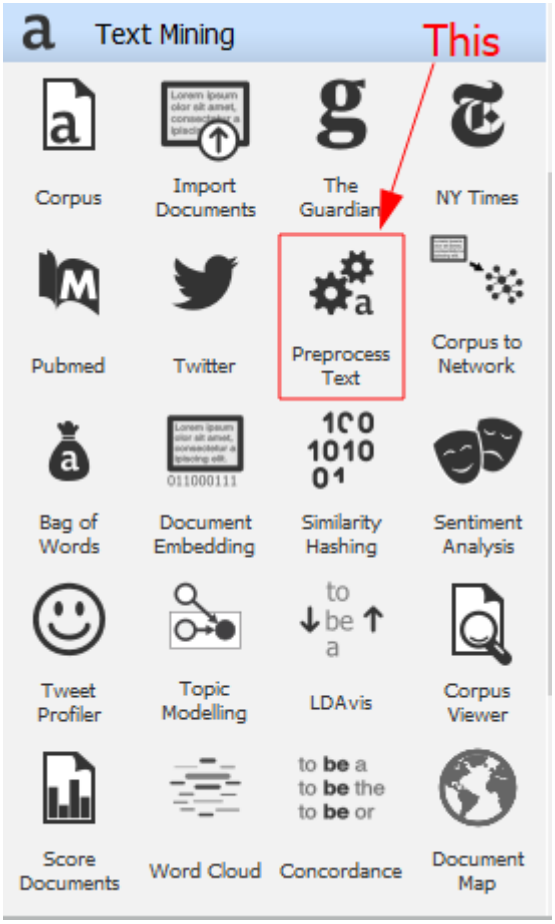
- Another method of visualizing data is though Word Clouds. Drag and drop a `Word Cloud` widget from the `Text Mining` section and connect it to the `Corpus` widget.

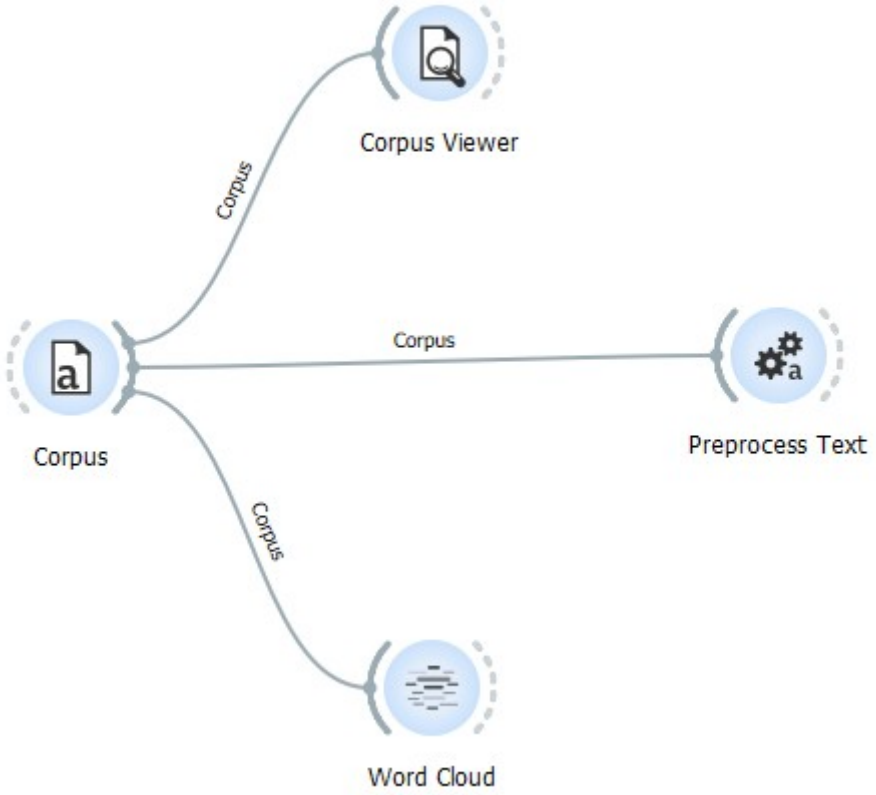
Text Mining
This
Corpus
Import Documents
The Guardian
NY Times
Pubmed
Twitter
Preprocess Text
Corpus to Network
Bag of Words
Document Embedding
Similarity Hashing
Sentiment Analysis
Tweet Profiler
Topic Modelling
LDAvis
Corpus Viewer
Score Documents
Word Cloud
to be a
to be the
to be or
Concordance
Document Map


Corpus Viewer
Corpus
Corpus
Corpus
Word Cloud

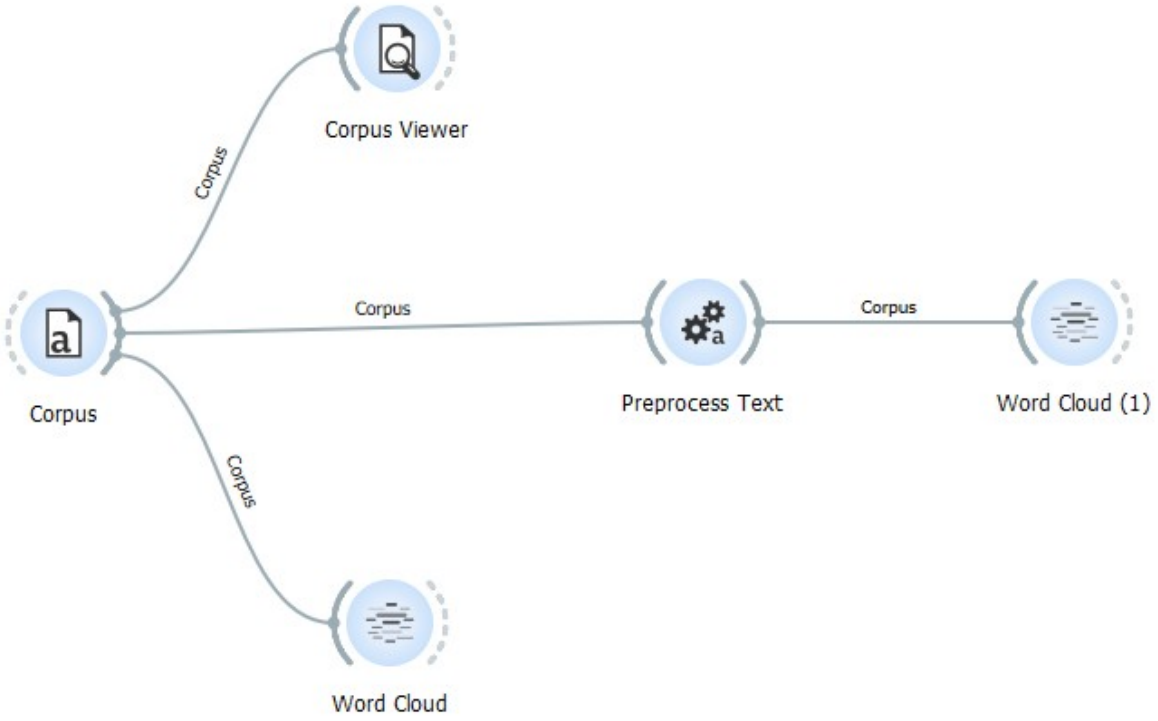

Cloud preferences
Color words
Words tilt:
30°
Words & weights
Weight
Word
5063 ,
4346 the
3330 and
1674 to
1584 .
1328 he
1180 a
1159 '
877 of
838 was
759 it
730 in
718 she
712 you
708 that
44 | -
0 | - | 3894

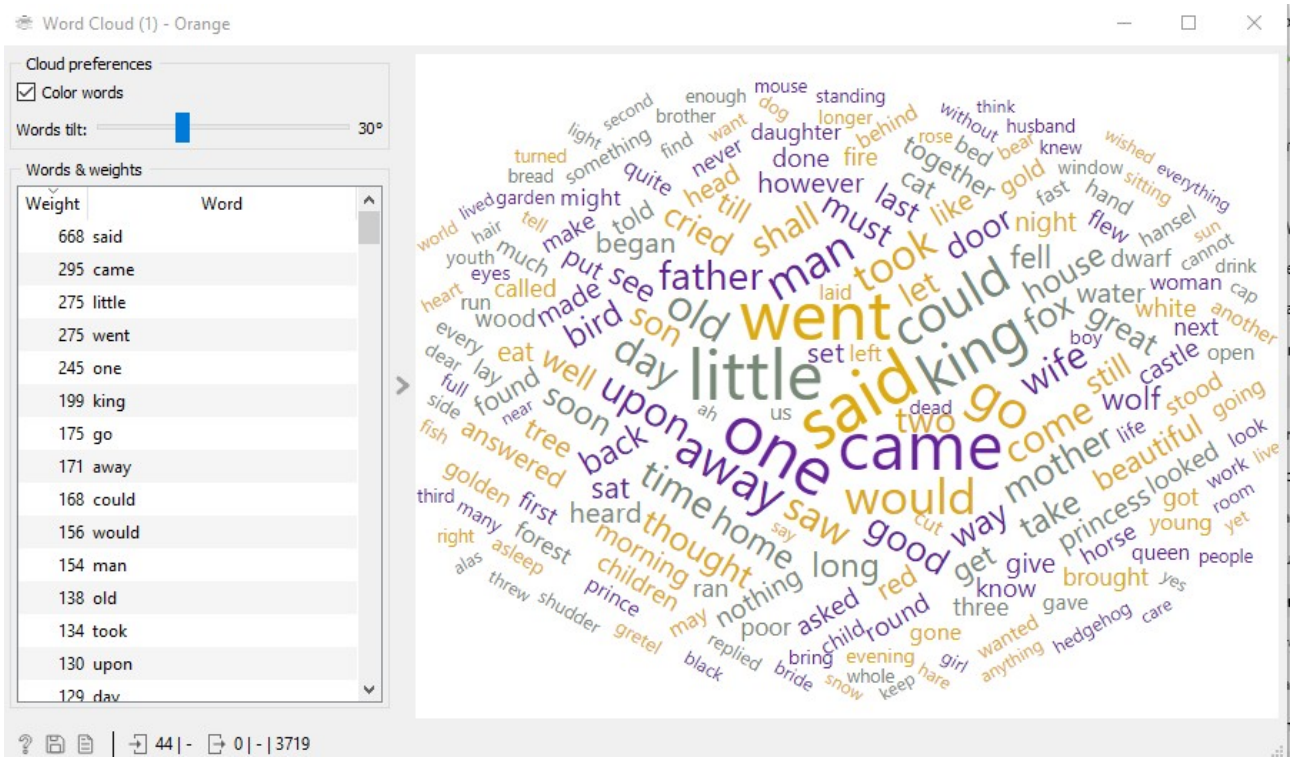
➢ We see that a lot of punctuation marks and uninformative words have made their way into the `Word Cloud`. In order to eliminate this, we have to use the `Preprocess Text` widget. Drag and drop a `Preprocess Text` widget from the `Text Mining` section onto the workspace and connect it to the Corpus widget.





- We now add another `Word Cloud` widget to see our updated cloud.


Corpus Viewer
Corpus
Corpus
Corpus
Corpus
Preprocess Text
Word Cloud (1)
Corpus
Word Cloud

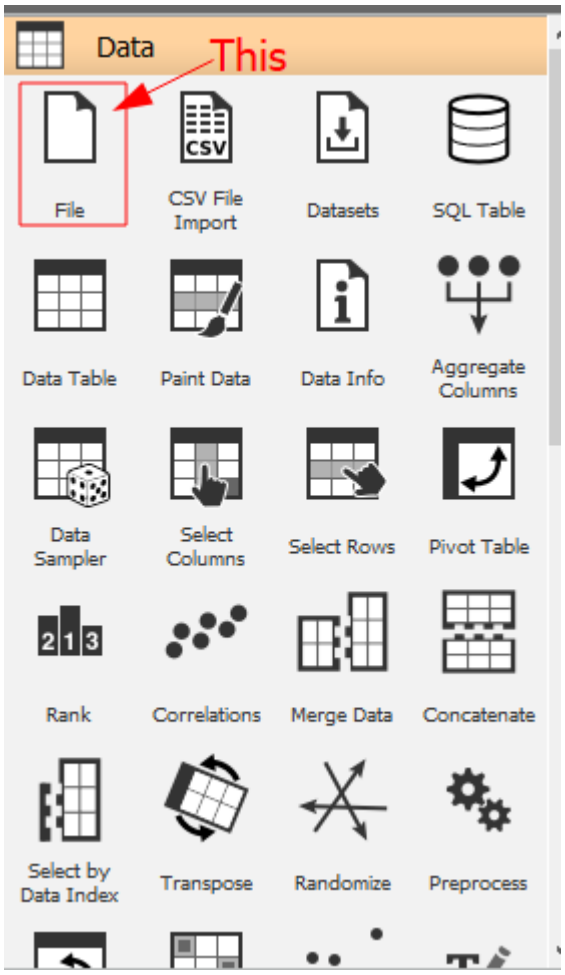

Word Cloud (1) - Orange
Cloud preferences
Color words
Words tilt:
30°
Words & weights
Weight
Word
668 said
295 came
275 little
275 went
245 one
199 king
175 go
171 away
168 could
156 would
154 man
138 old
134 took
130 upon
129 day
44 | -
0 | - | 3719

Practical 6

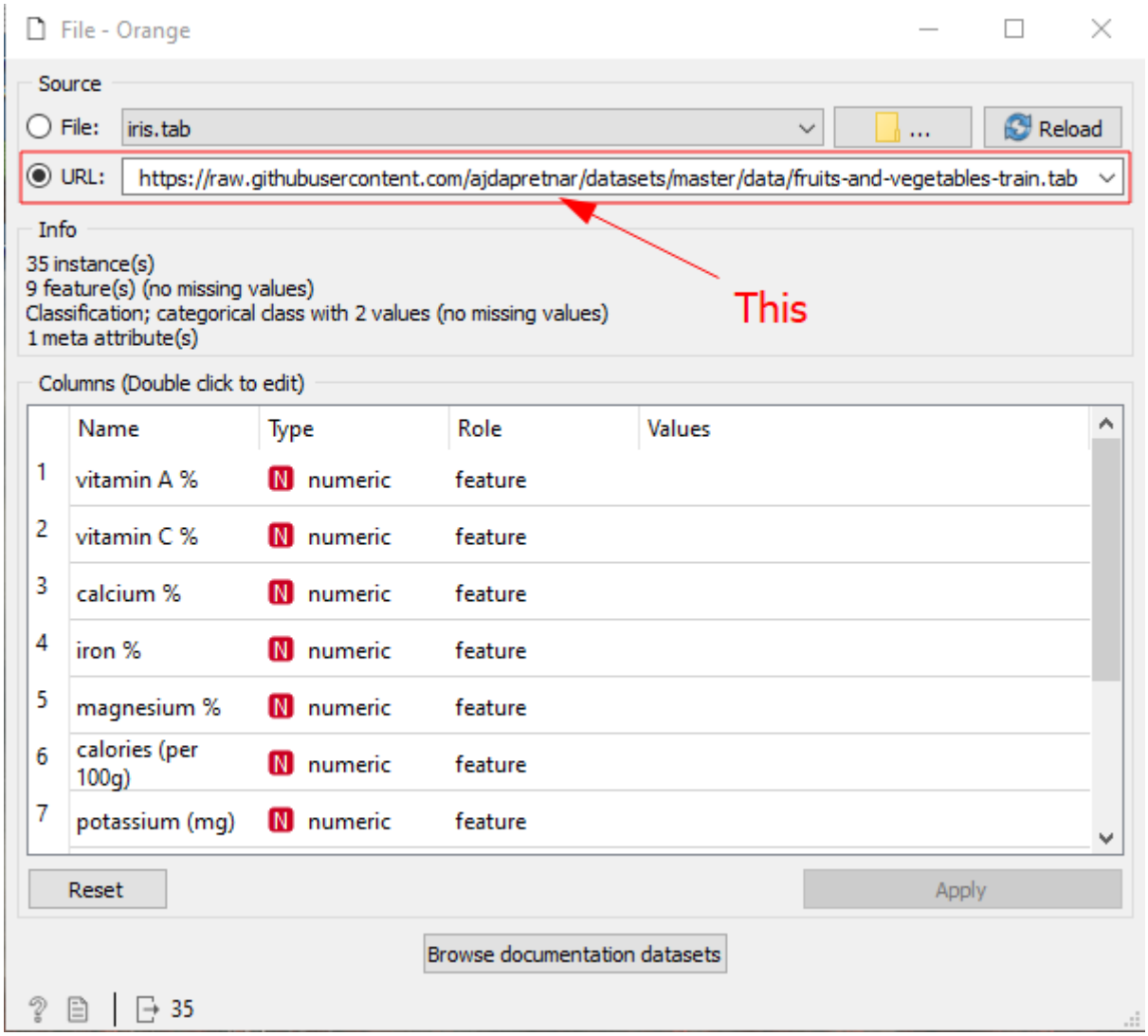
Aim: Predictions using orange.

Procedure:

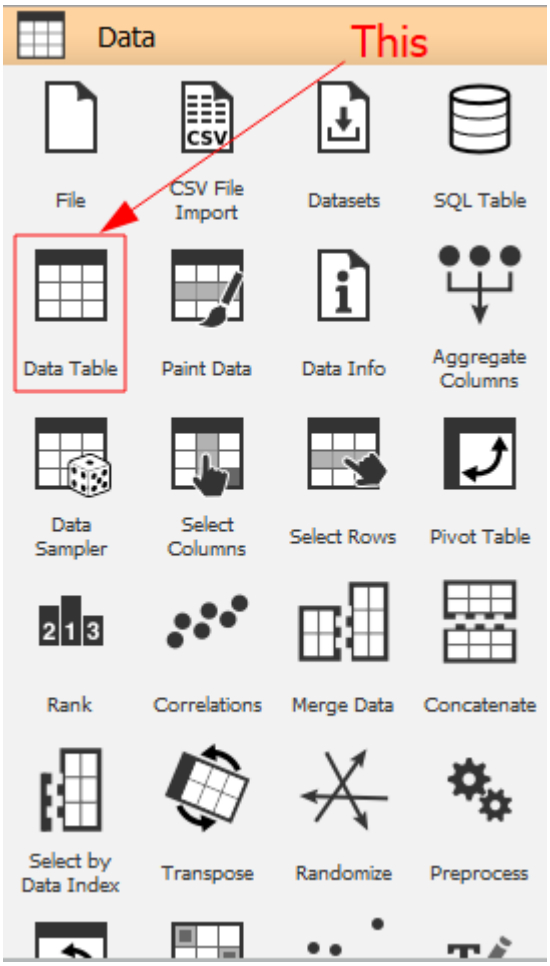
- Drag and drop a new `File` widget from the `Data` section to the workspace.

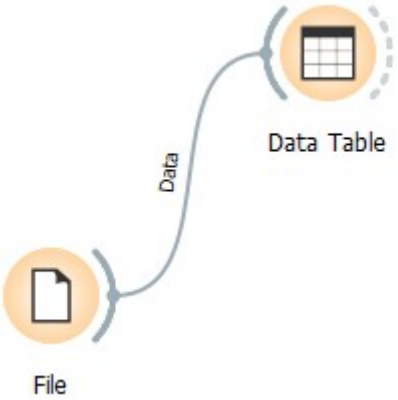


➢ Double click the `File` widget, select the URL radiobutton and enter the following URL.

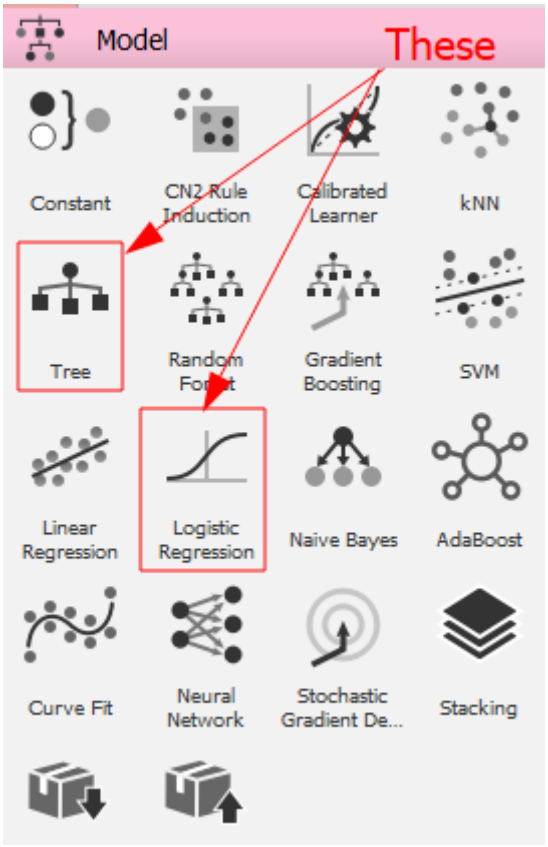


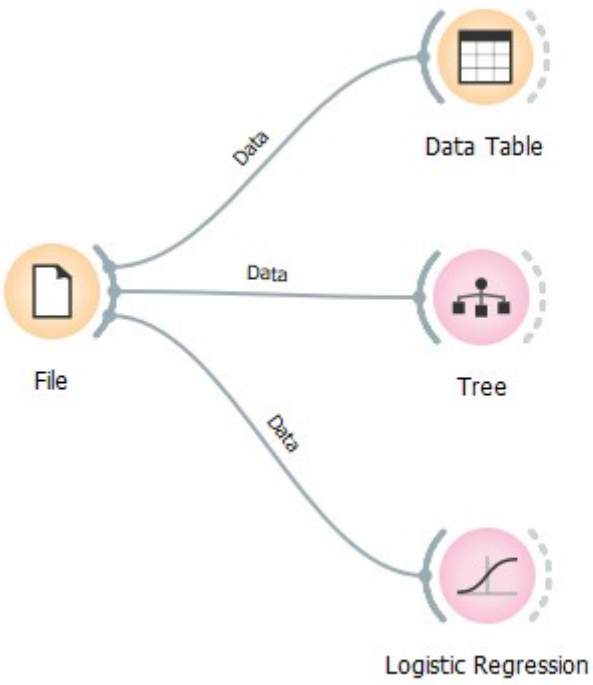
➢ We can visualise the data loaded in the `File` widget using the `Data Table`. Drag and drop a `Data Table` widget from the `Data` section onto the workspace and connect it to the file widget.



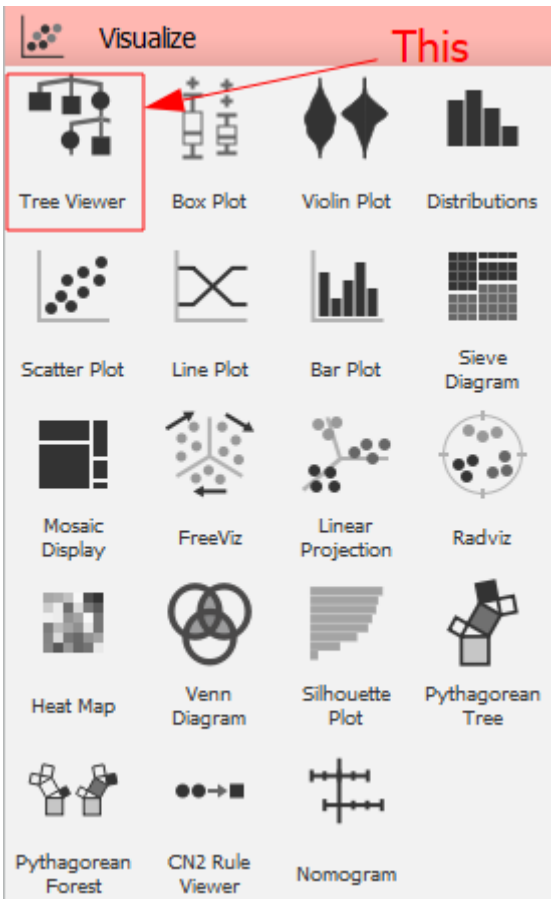


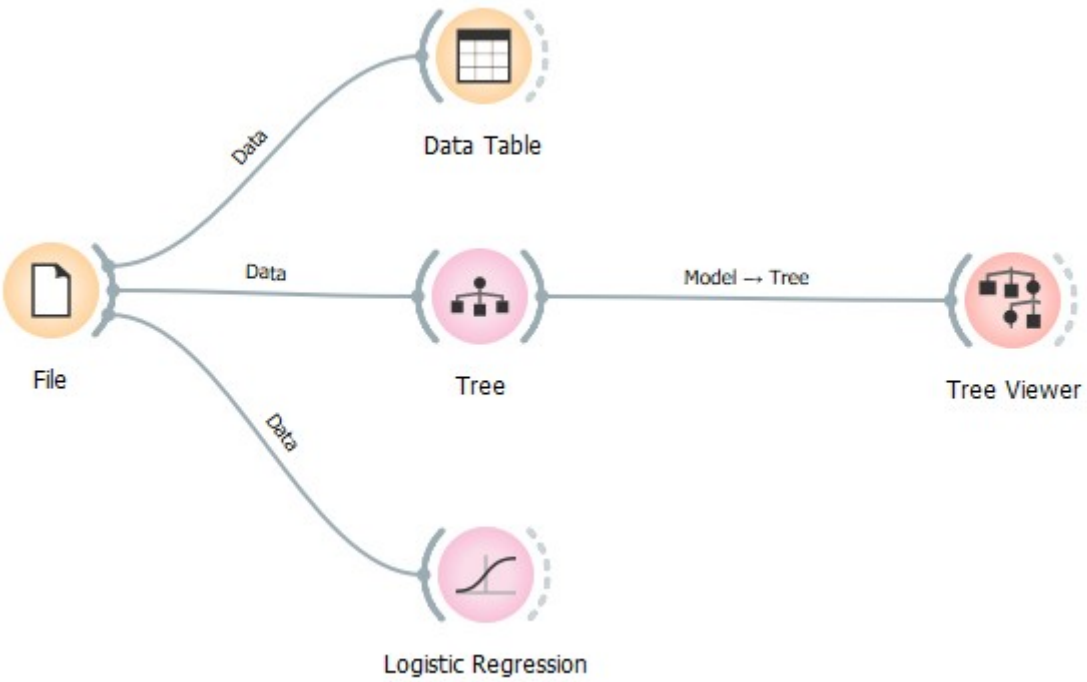
➢ We will use Classification Tree and Logistic Regression models. Drag and drop a `Tree` and a `Logistic Regression` widget from the `Models` section onto the workspace and connect them to the `File` widget.

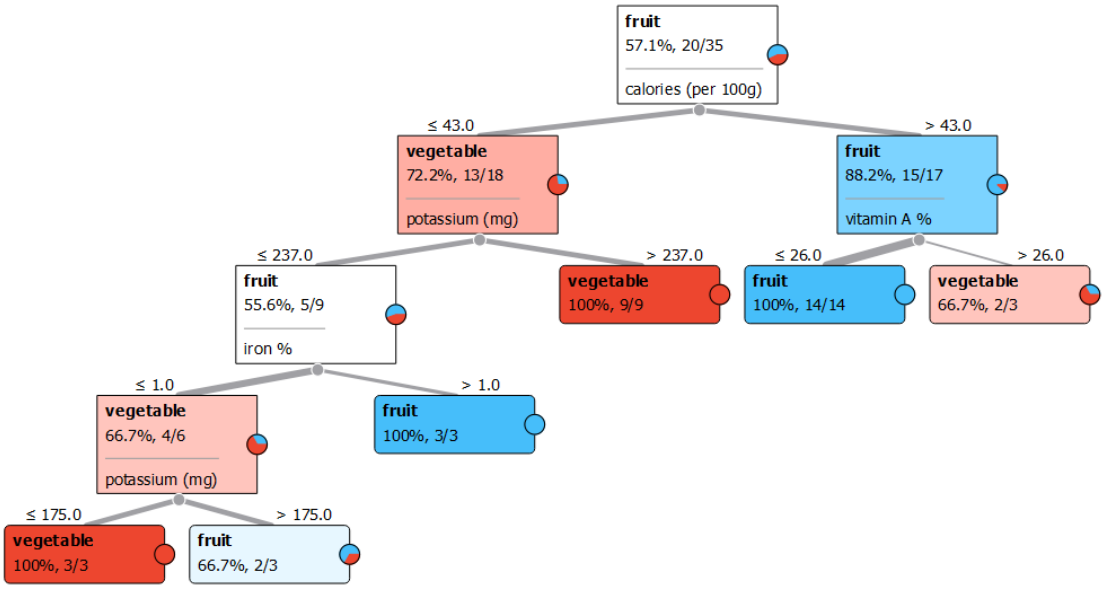




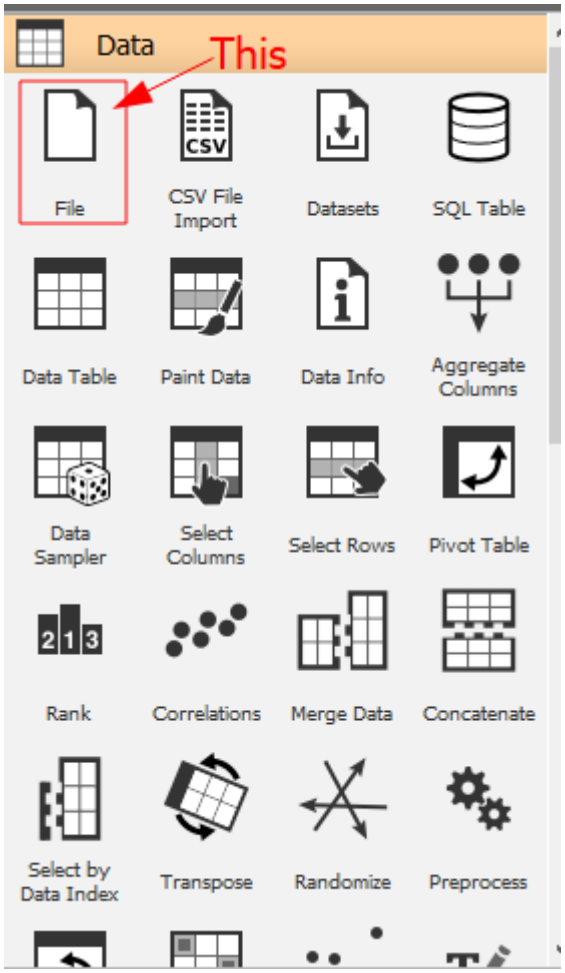
- We use the `Tree Viewer` widget to view the Classification `Tree`. Drag and drop a `Tree Viewer` widget from the `Visualise` section onto the workspace and connect it to the Tree widget.

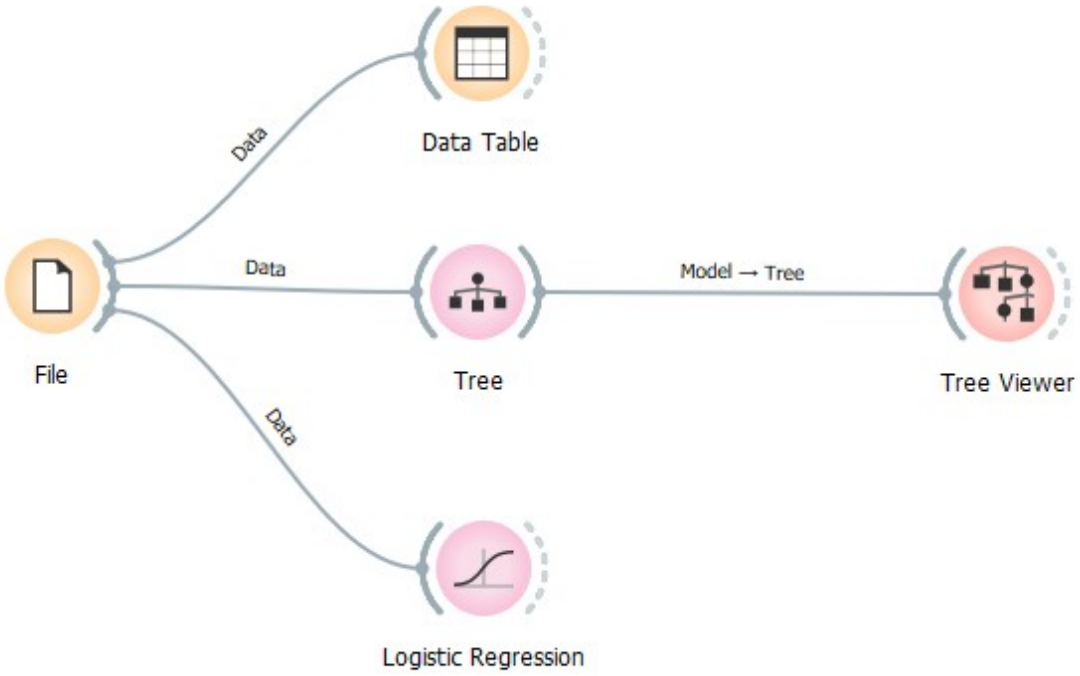

Visualize
This
Tree Viewer
Box Plot
Violin Plot
Distributions
Scatter Plot
Line Plot
Bar Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Silhouette Plot
Pythagorean Tree
Pythagorean Forest
CN2 Rule Viewer
Nomogram


Data
Data Table
Data
Model → Tree
File
Tree
Tree Viewer
Data
Logistic Regression

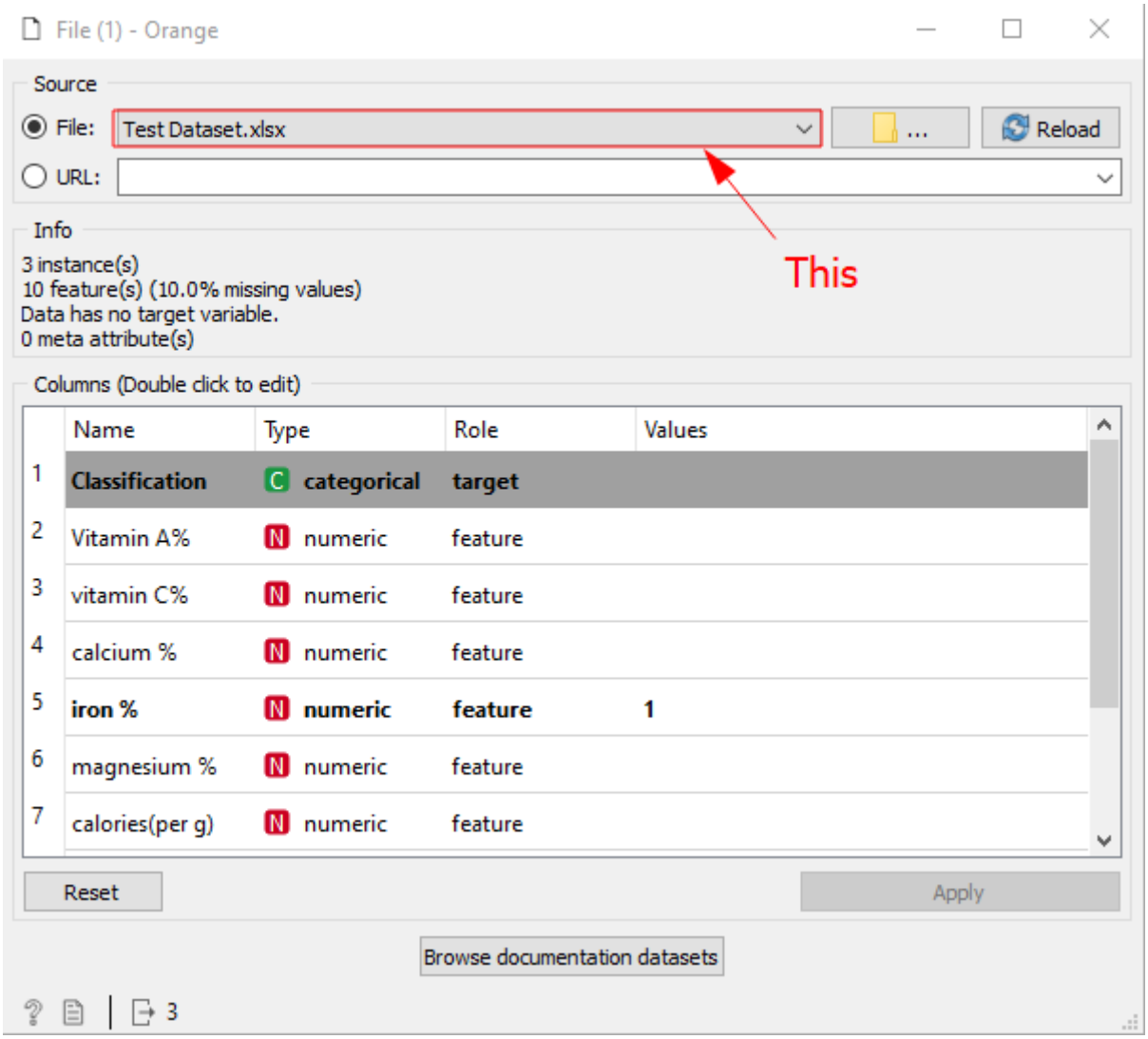


➢ Drag and drop another `File` widget from the `Data` section onto the workspace.

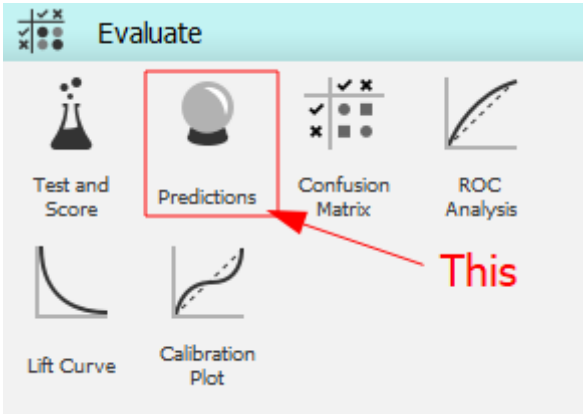


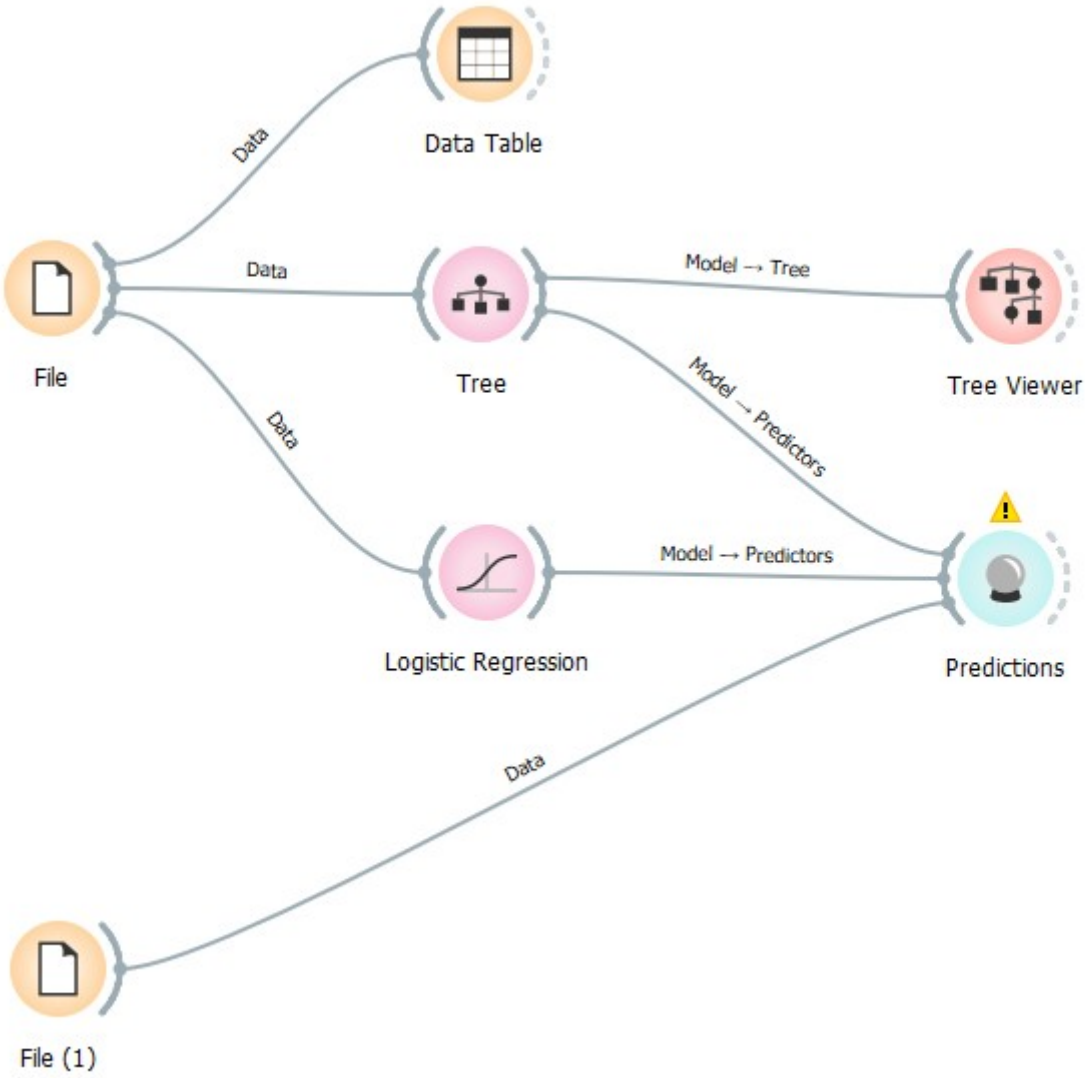


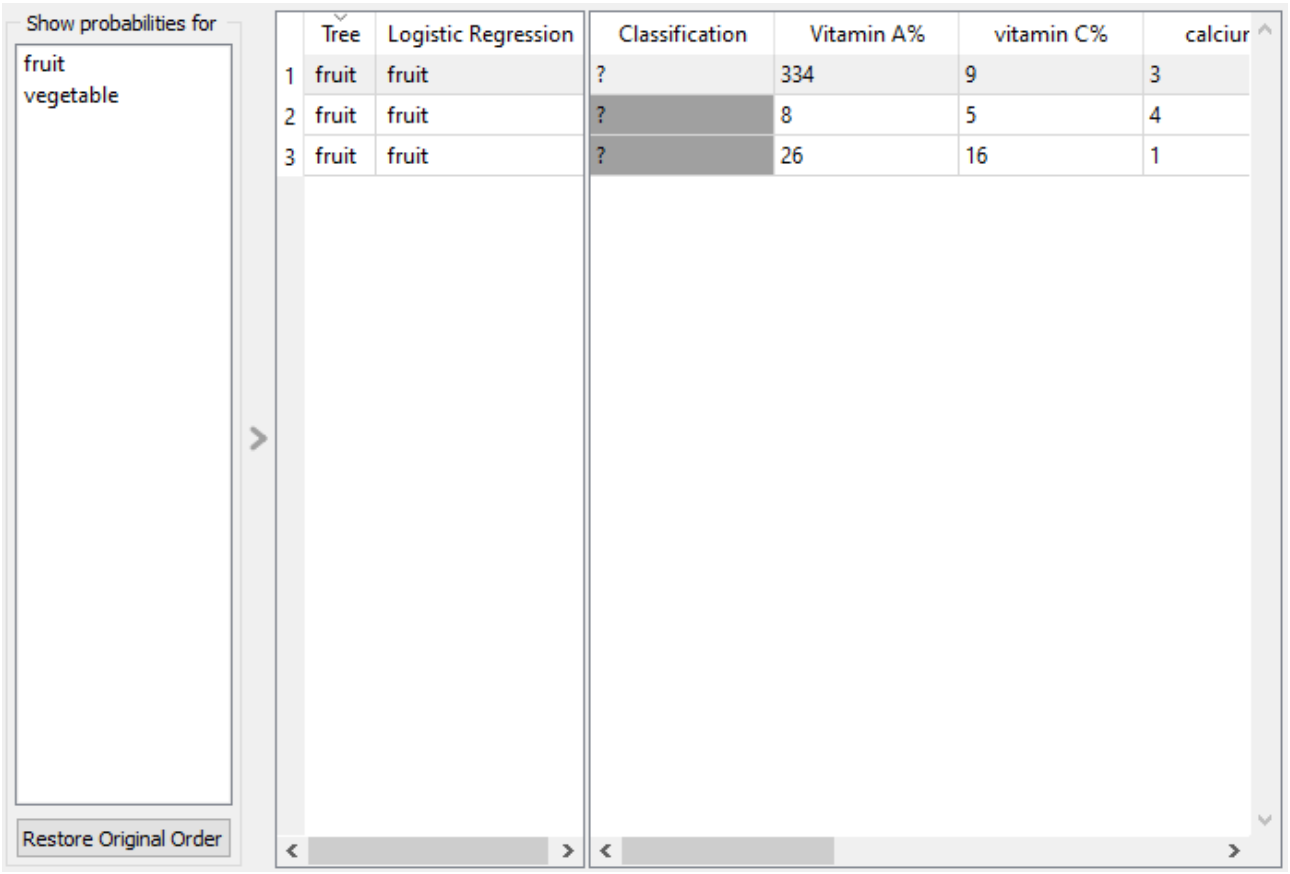
- Double-click the newly created `File` widget and add the `Test Dataset.xlsx` file.



➢ Add a `Predictions` widget from the `Evaluate` section onto the workspace and connect it to the `Tree`, `Logistic Regression` and newly created `File` widget.




Data
Data Table
Data
Model → Tree
File
Tree
Tree Viewer
Data
Model → Predictors
Model → Predictors
Logistic Regression
Predictions
Data
File (1)

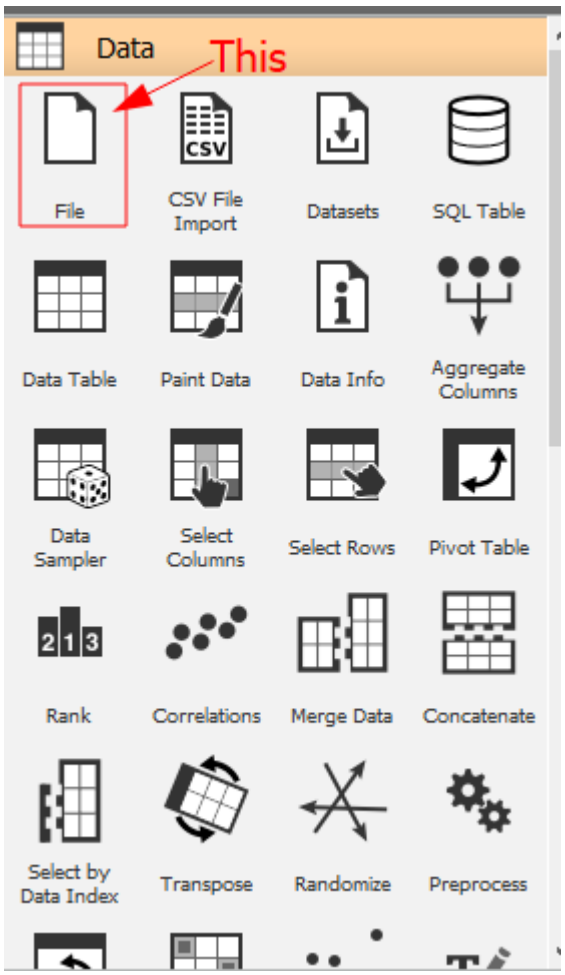

Show probabilities for
fruit
vegetable
Restore Original Order
Tree
Logistic Regression
Classification
Vitamin A%
vitamin C%
calciur
1 fruit fruit ? 334 9 3
2 fruit fruit ? 8 5 4
3 fruit fruit ? 26 16 1

Practical 7

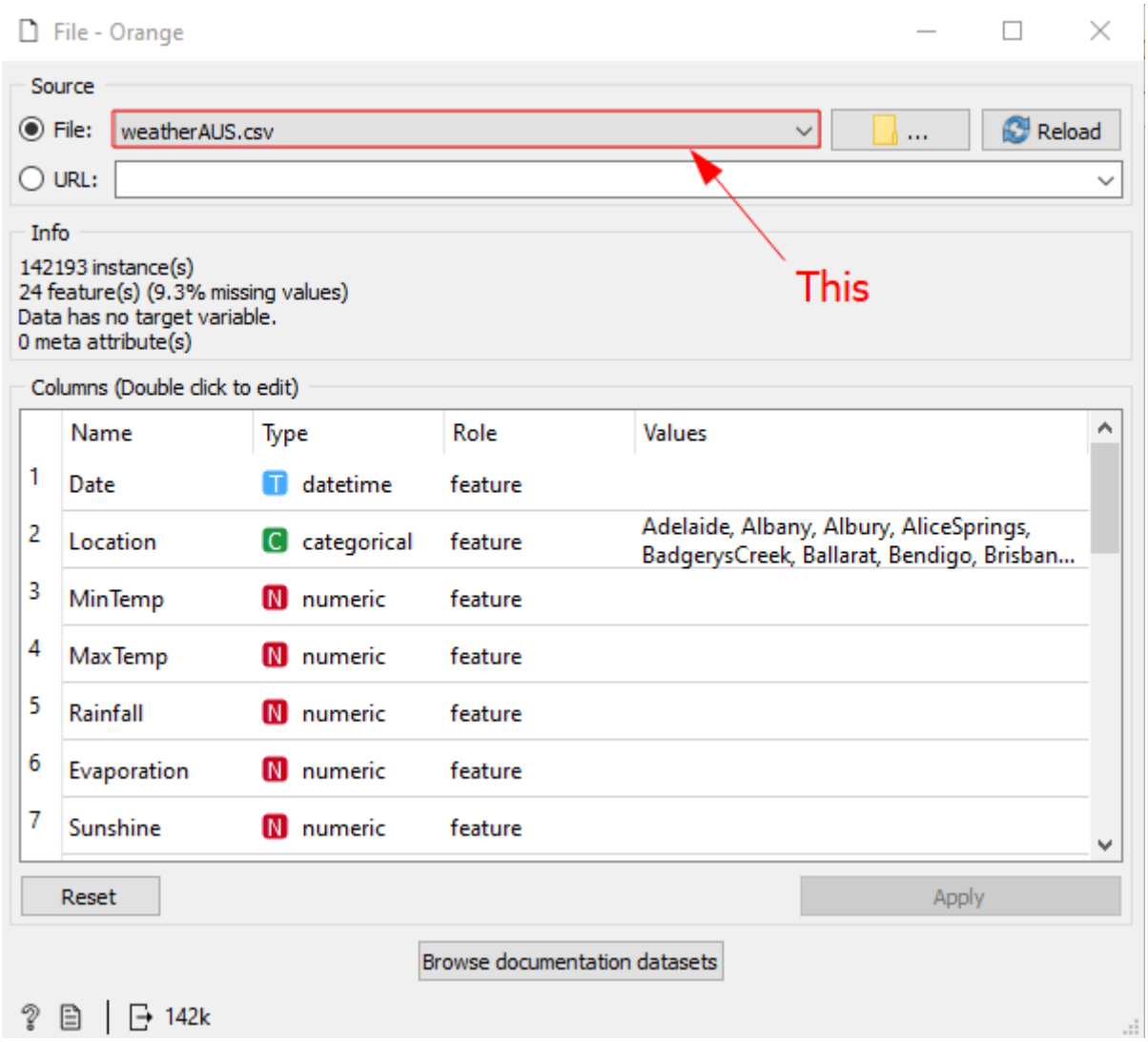
Aim: Predictions using orange.

Procedure:

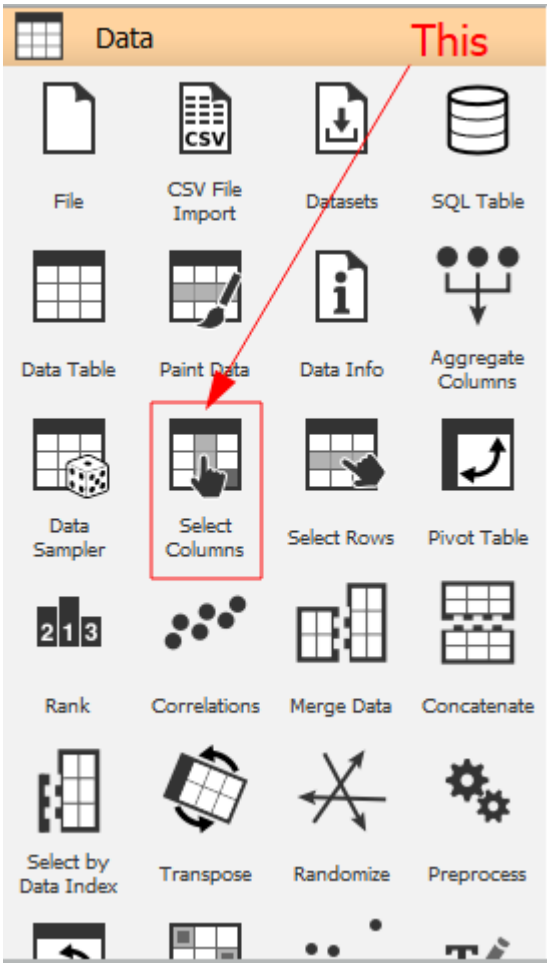
- Download the weatherAUS.csv file from kaggle.
- Drag and drop a File widget from the Data section onto the workspace.



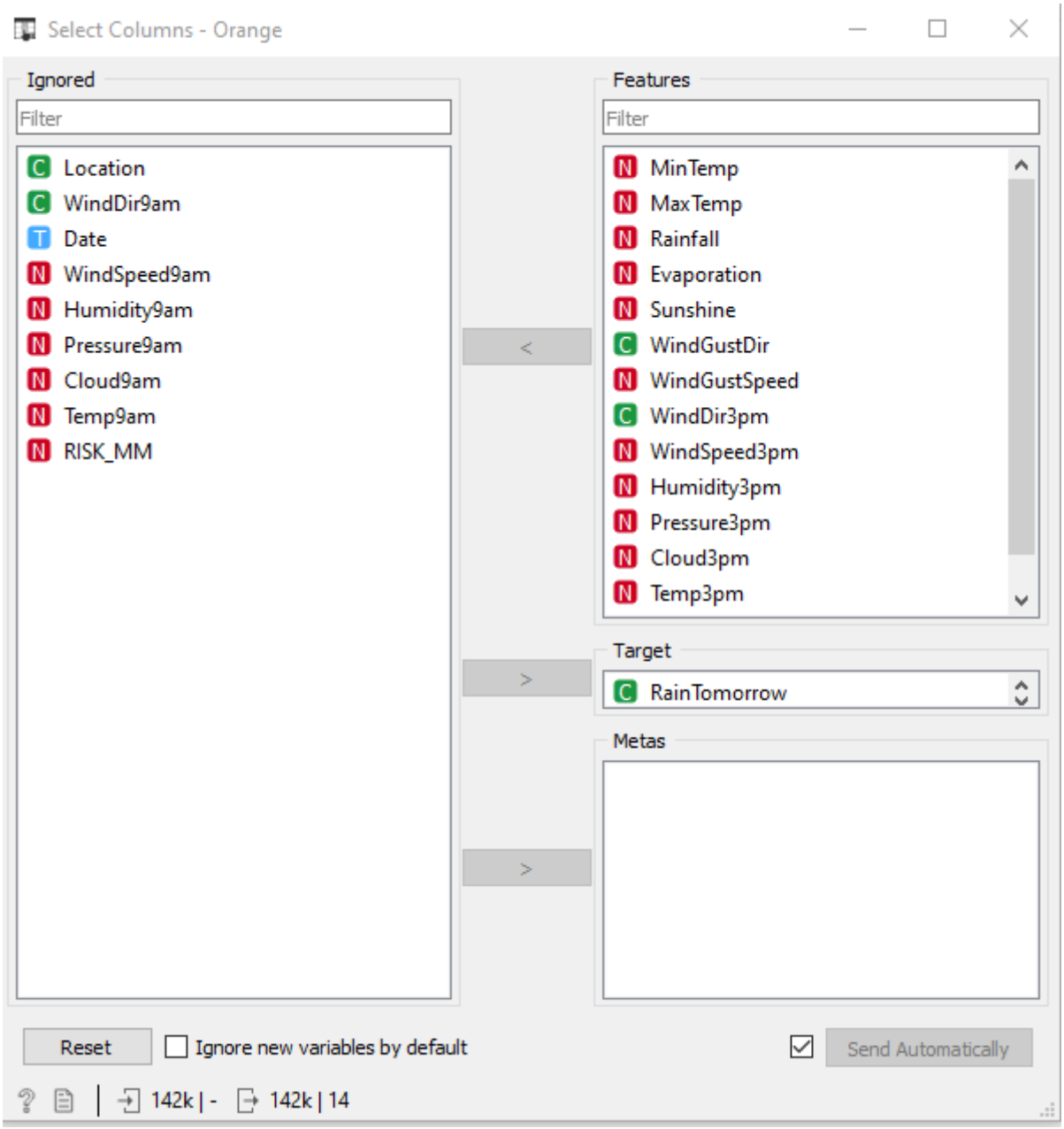
➢ Double-click the File widget and change the File attribute to weatherAUS.csv.



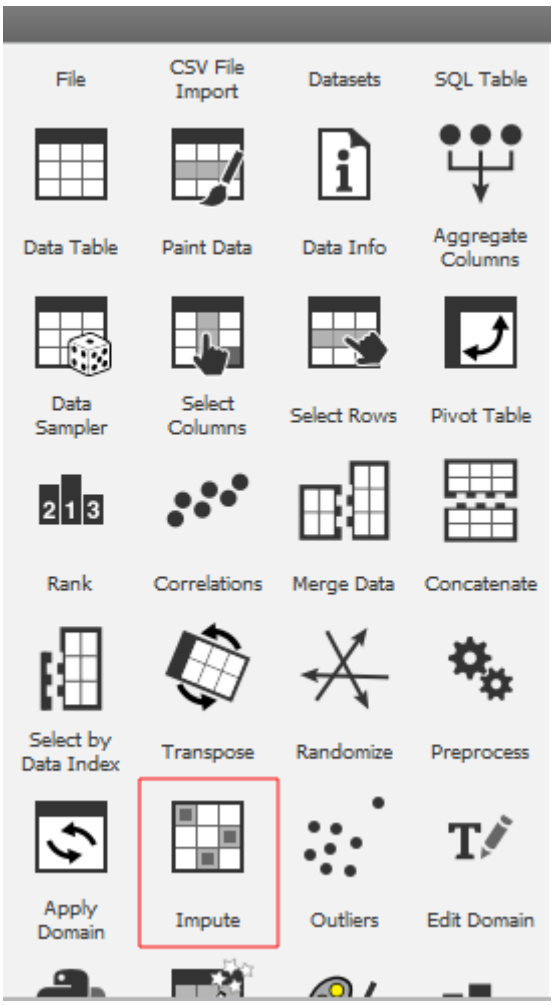
➢ Drag and drop the Select Columns widget from the Data section and connect it to the File widget.



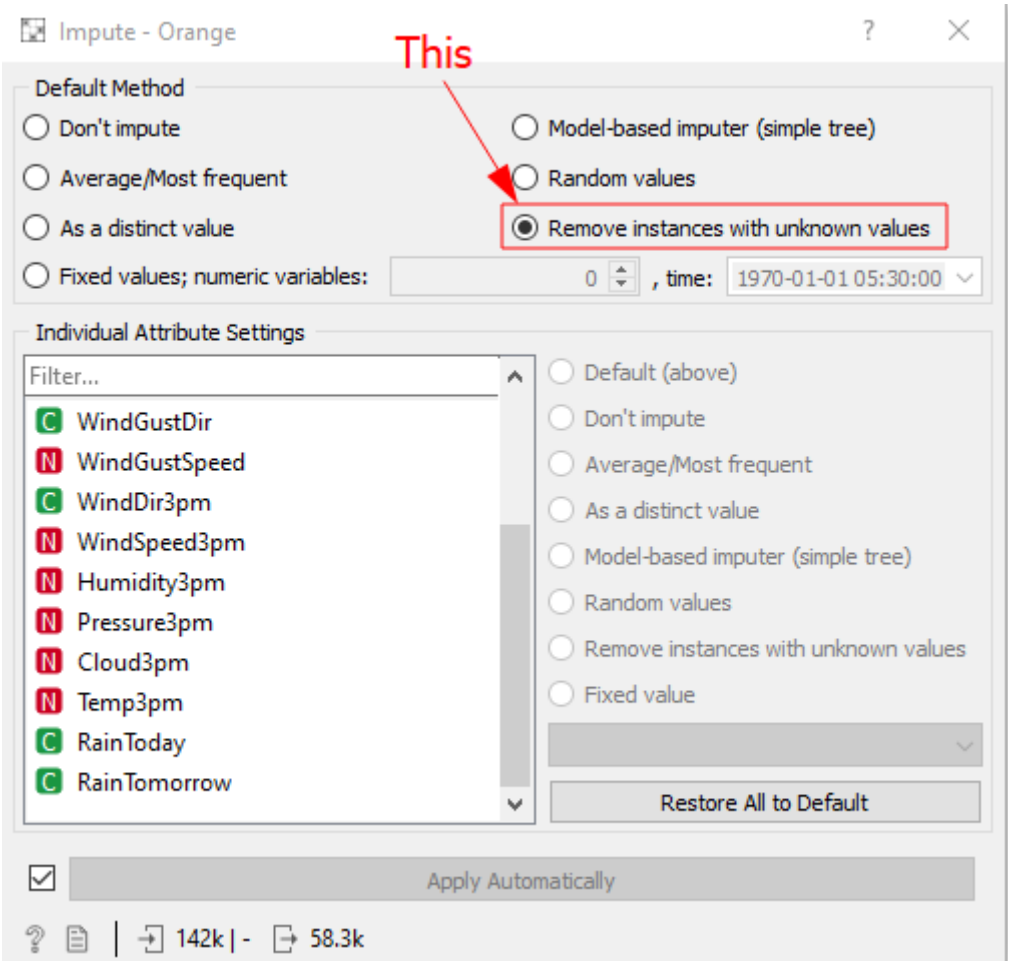
➢ Double-click the Select Columns widgets and perform the following changes:



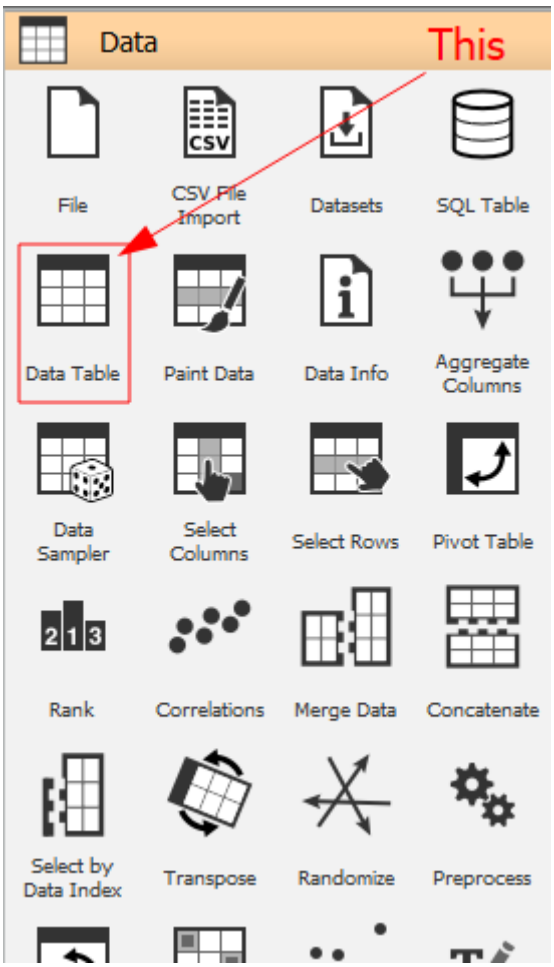
➢ Now, drag and drop a Impute widget from the Data section onto the workspace and connect it to the Select Columns widget.

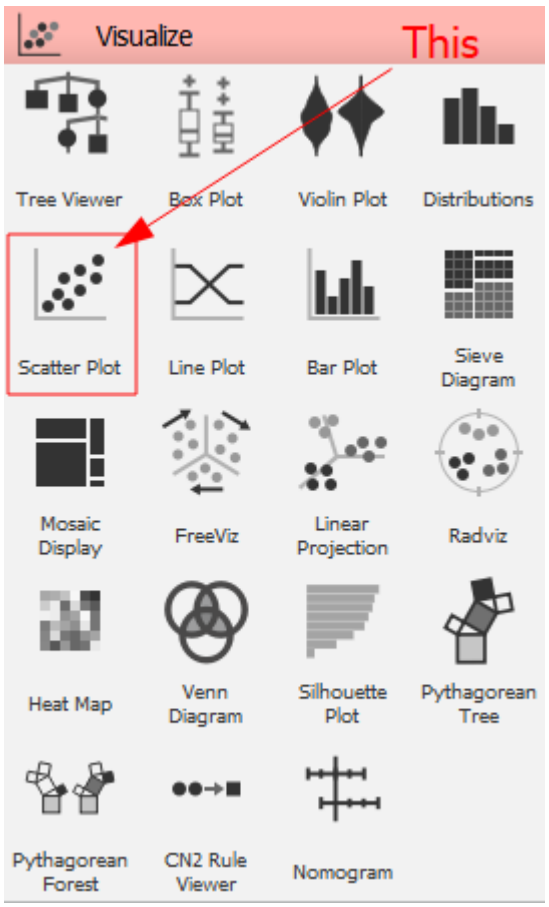


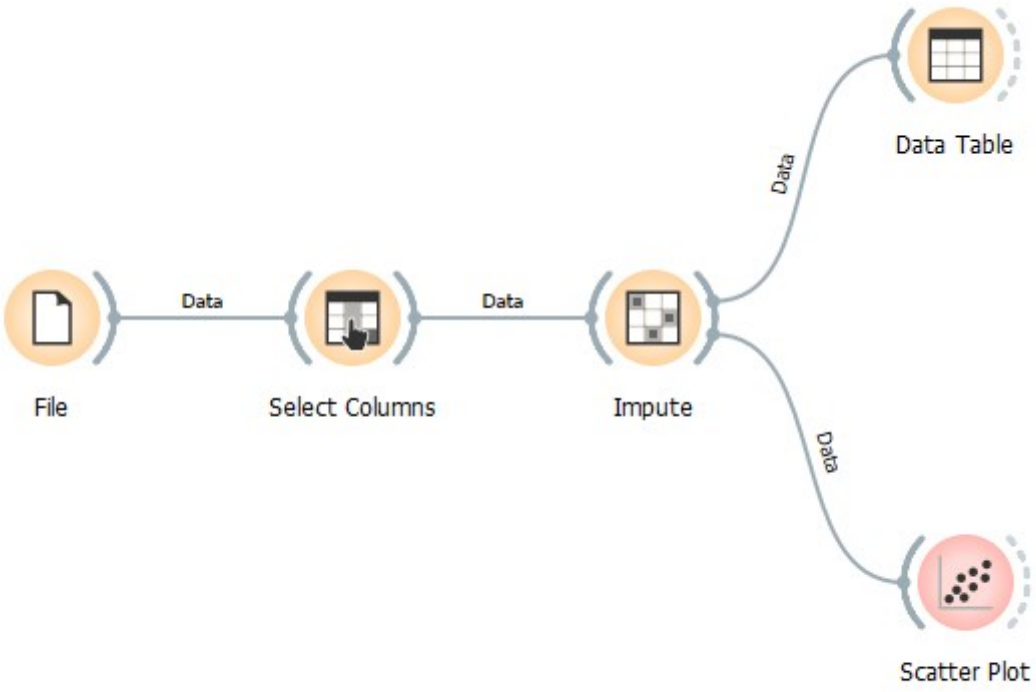
- Double-click the Impute widget and select the Remove instances with unknown values radiobutton.

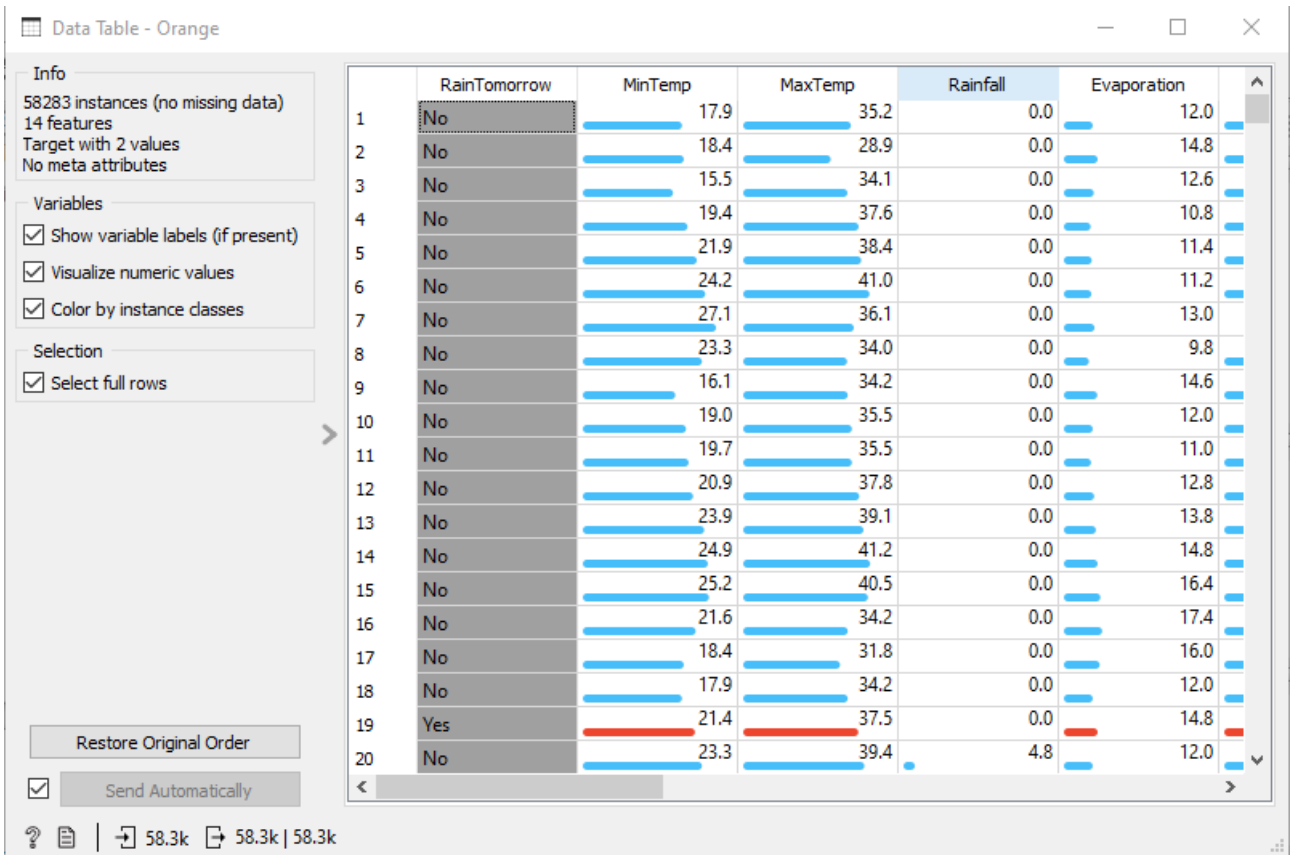


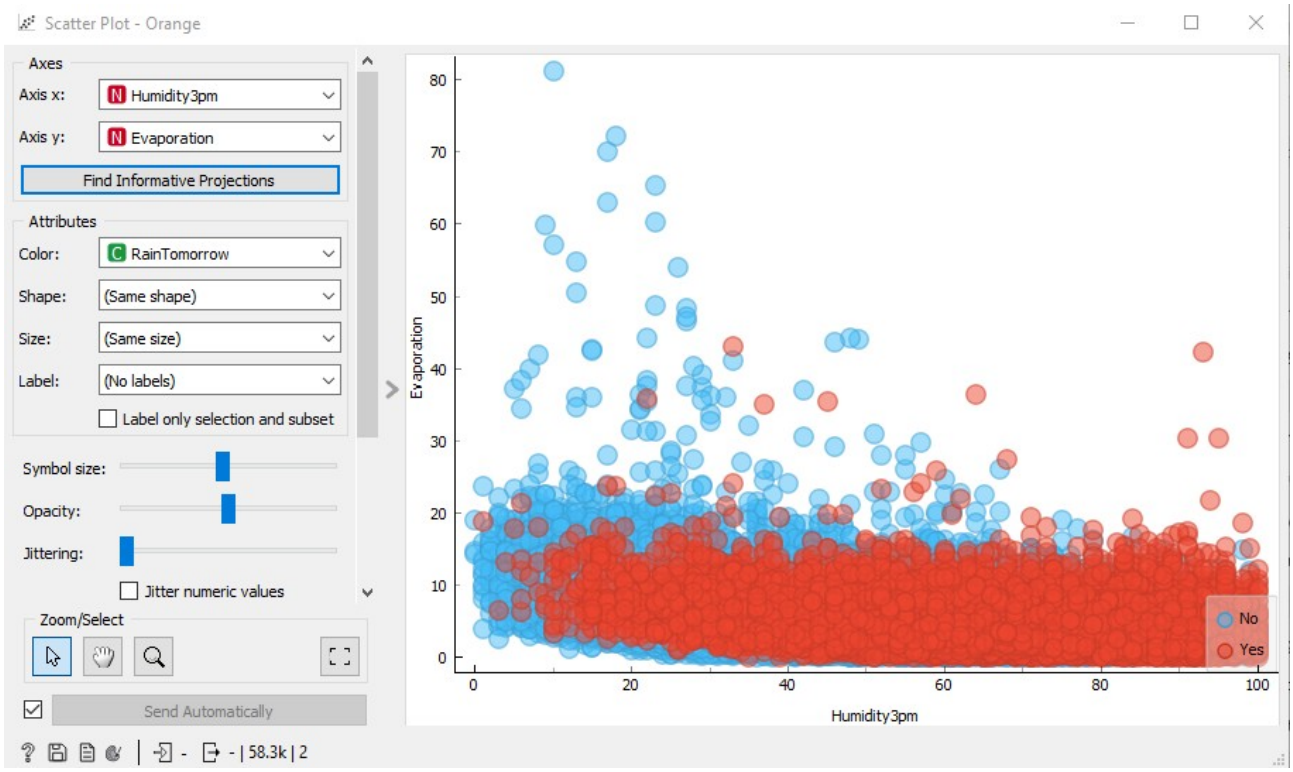
- We will use a Scatter Plot and a Data Table to visualise the results. Drag and drop a Data Table from the Data section and a Scatter Plot from the Visualise section onto the workspace. Connect these two widgets to the Impute widget.


Data
This
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Aggregate Columns
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlations
Merge Data
Concatenate
Select by Data Index
Transpose
Randomize
Preprocess

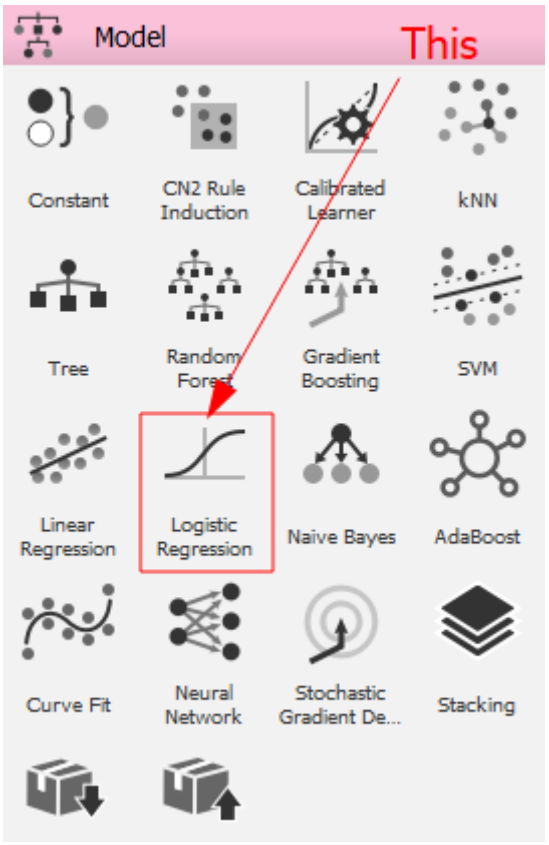

Visualize
This
Tree Viewer
Box Plot
Violin Plot
Distributions
Scatter Plot
Line Plot
Bar Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Silhouette Plot
Pythagorean Tree
Pythagorean Forest
CN2 Rule Viewer
Nomogram

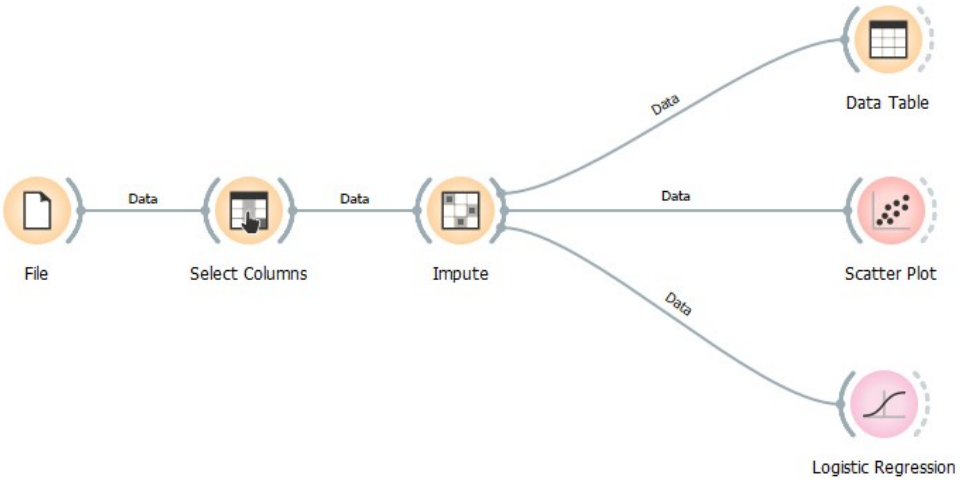

Data Table
Data
Data
Data
Data
File
Select Columns
Impute
Scatter Plot


Data Table - Orange
Info
58283 instances (no missing data)
14 features
Target with 2 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 Automatically
58.3k
58.3k | 58.3k
RainTomorrow
MinTemp
MaxTemp
Rainfall
Evaporation
1 No 17.9 35.2 0.0 12.0
2 No 18.4 28.9 0.0 14.8
3 No 15.5 34.1 0.0 12.6
4 No 19.4 37.6 0.0 10.8
5 No 21.9 38.4 0.0 11.4
6 No 24.2 41.0 0.0 11.2
7 No 27.1 36.1 0.0 13.0
8 No 23.3 34.0 0.0 9.8
9 No 16.1 34.2 0.0 14.6
10 No 19.0 35.5 0.0 12.0
11 No 19.7 35.5 0.0 11.0
12 No 20.9 37.8 0.0 12.8
13 No 23.9 39.1 0.0 13.8
14 No 24.9 41.2 0.0 14.8
15 No 25.2 40.5 0.0 16.4
16 No 21.6 34.2 0.0 17.4
17 No 18.4 31.8 0.0 16.0
18 No 17.9 34.2 0.0 12.0
19 Yes 21.4 37.5 0.0 14.8
20 No 23.3 39.4 4.8 12.0

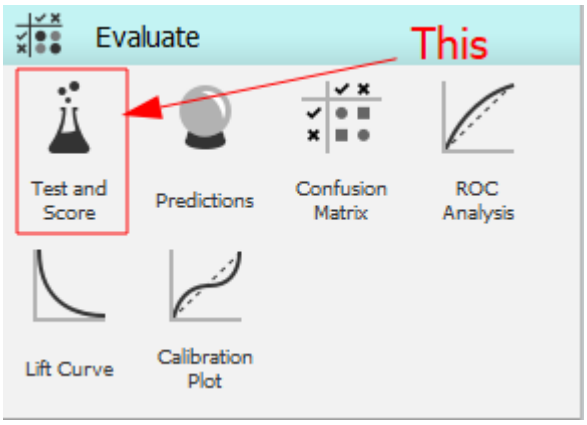

Scatter Plot - Orange
Axes
Axis x: Humidity3pm
Axis y: Evaporation
Find Informative Projections
Attributes
Color: RainTomorrow
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Jitter numeric values
Zoom/Select
Send Automatically
- | 58.3k | 2
Evaporation
Humidity3pm
No
Yes

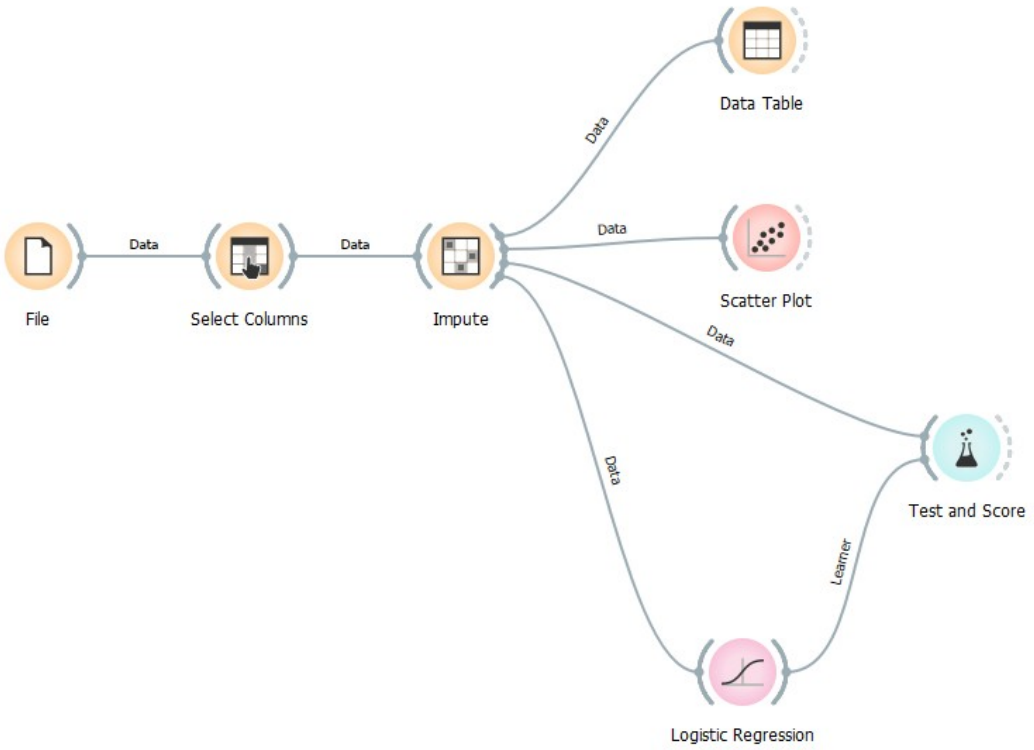
➢ We will use Logistic Regression to model our data. Drag and drop the Logistic Regression widget from the Model section onto the workspace and connect it to the Impute widget.



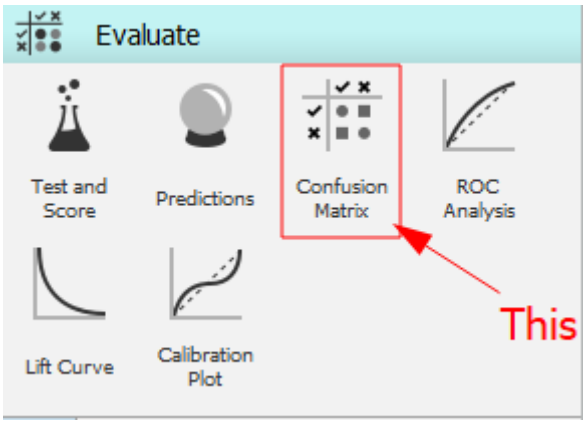


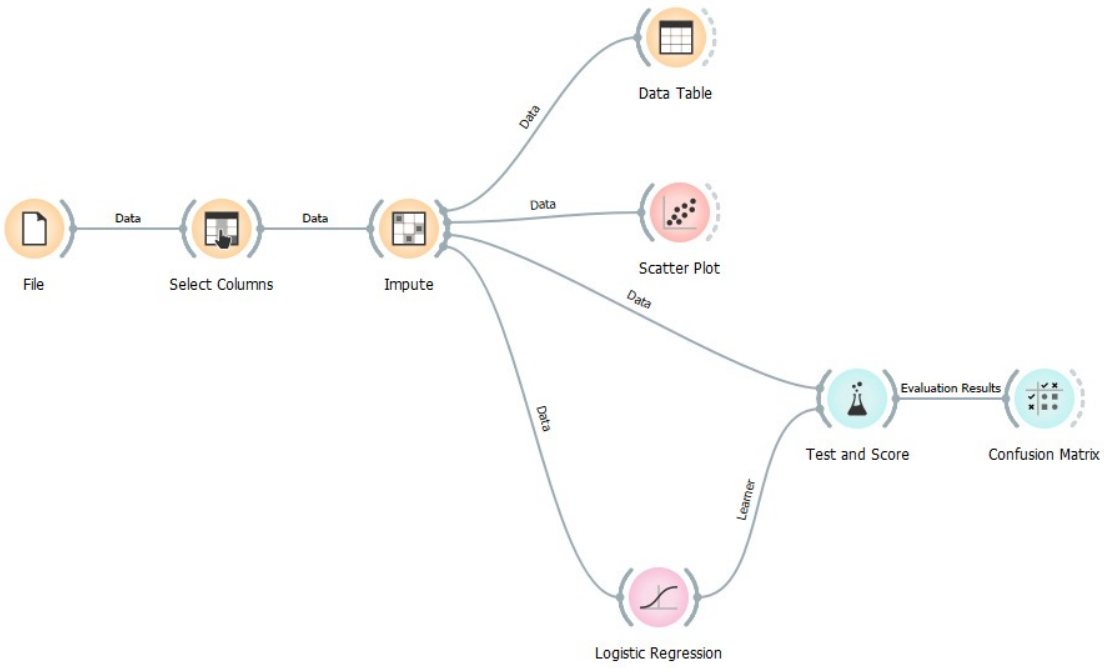
➢ Drag and Drop a Test and Score widget from the Evaluate section onto the workspace and connect it to the Logistic Regression and Impute widgets.

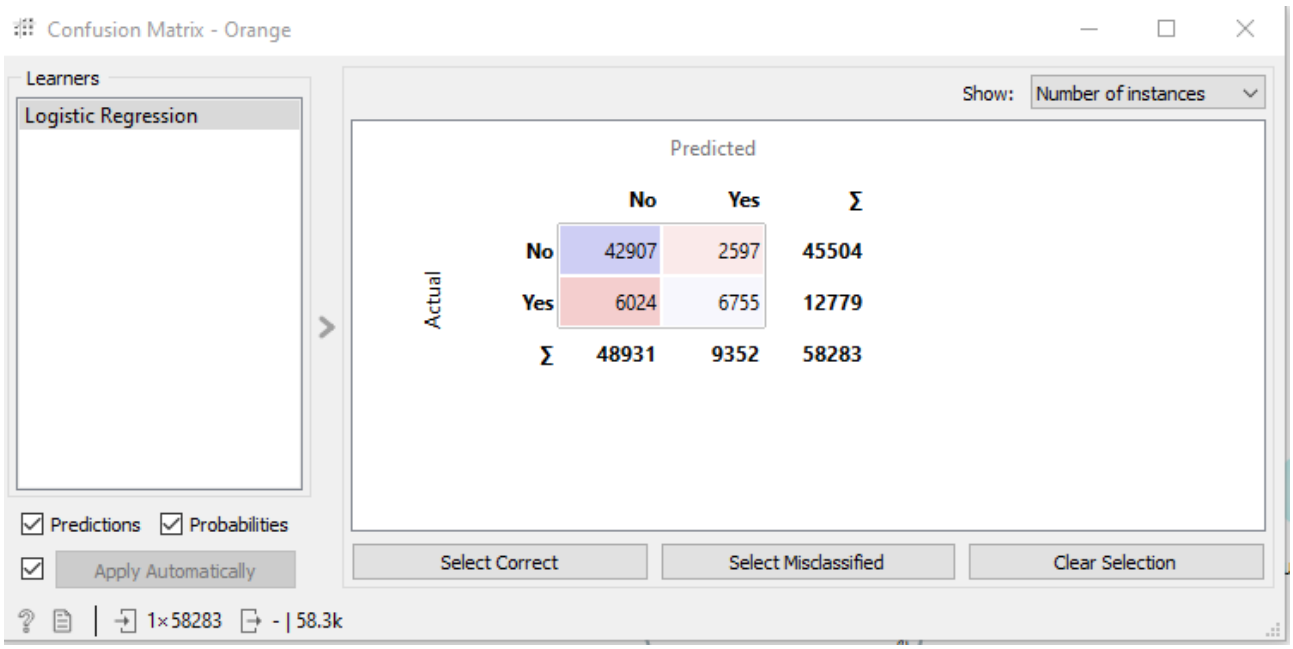




➢ Finally, drag and drop a Confusion Matrix widget from the Evaluate section onto the workspace and connect it to the Test and Score widget.






Confusion Matrix - Orange
Learners
Logistic Regression
Show:
Number of instances
Predicted
No
Yes
Σ
Actual
No
42907
2597
45504
Yes
6024
6755
12779
Σ
48931
9352
58283
Predictions
Probabilities
Apply Automatically
Select Correct
Select Misclassified
Clear Selection
1×58283
- | 58.3k