

ORANGE DATA MINING

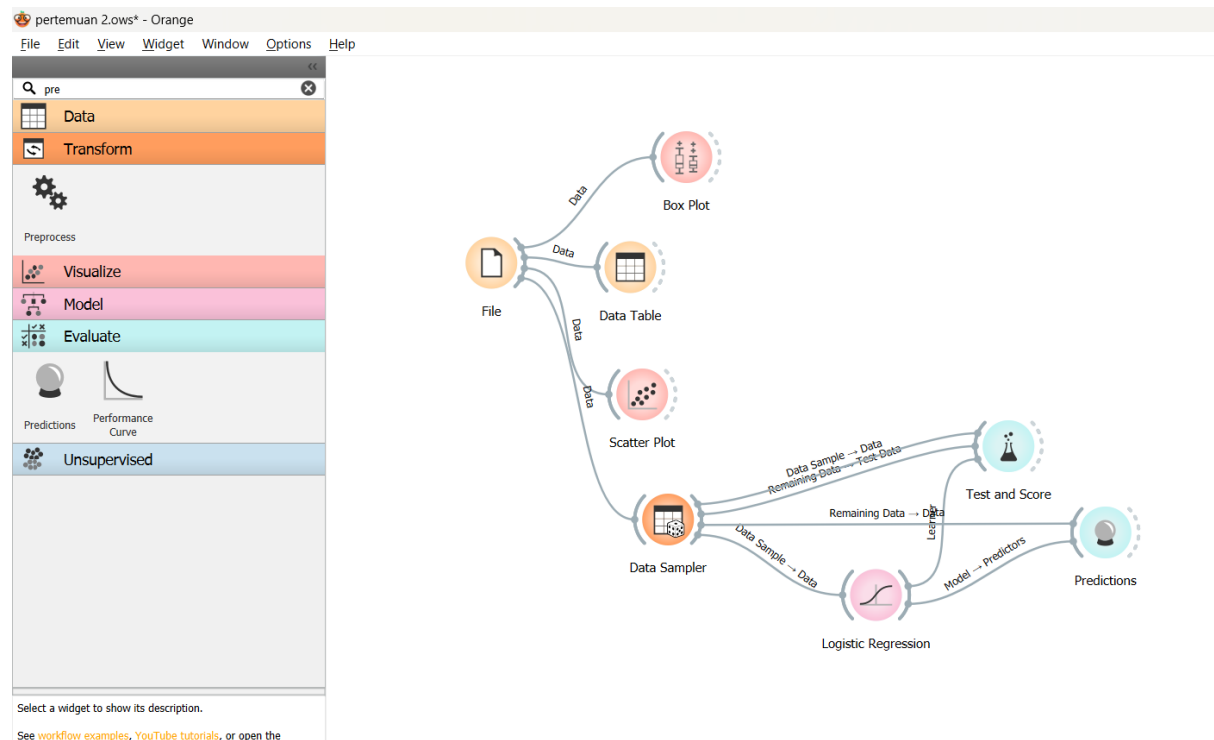
SALARY DATASET

MACHINE LEARNING WEEK 2

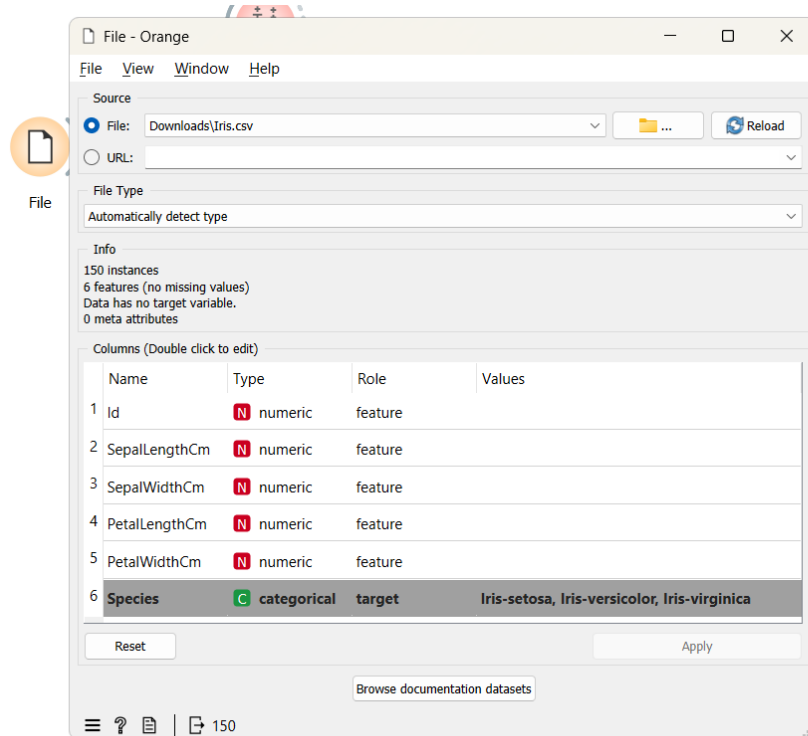
M Rakan Bagus

1103213162

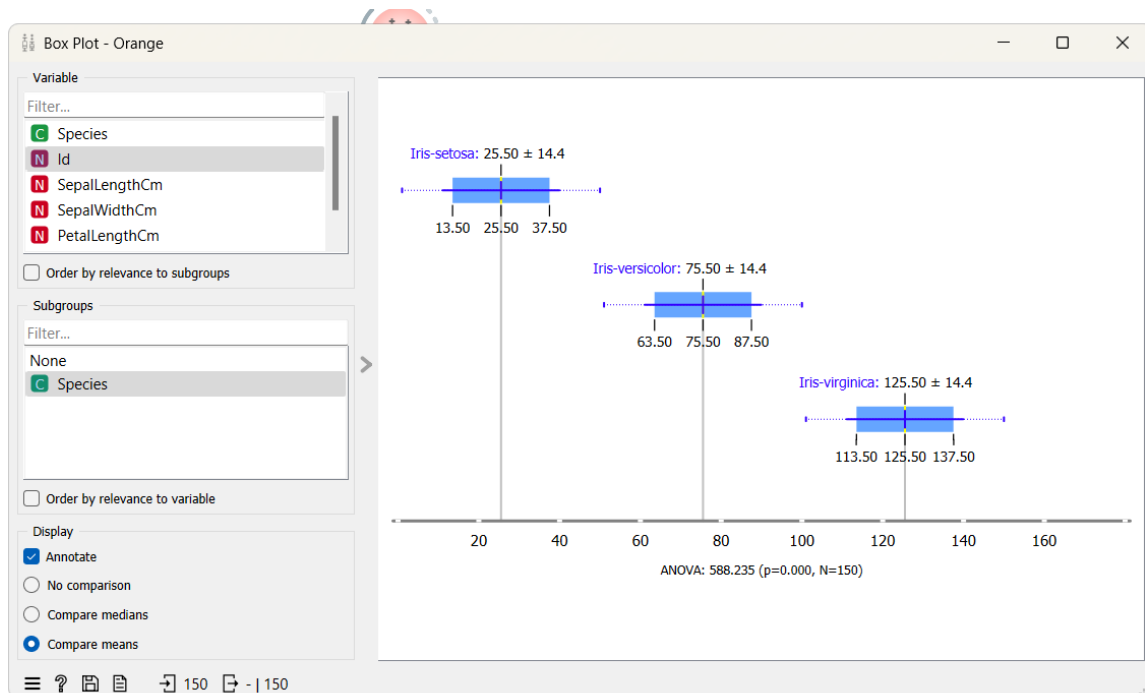
1. File Mining



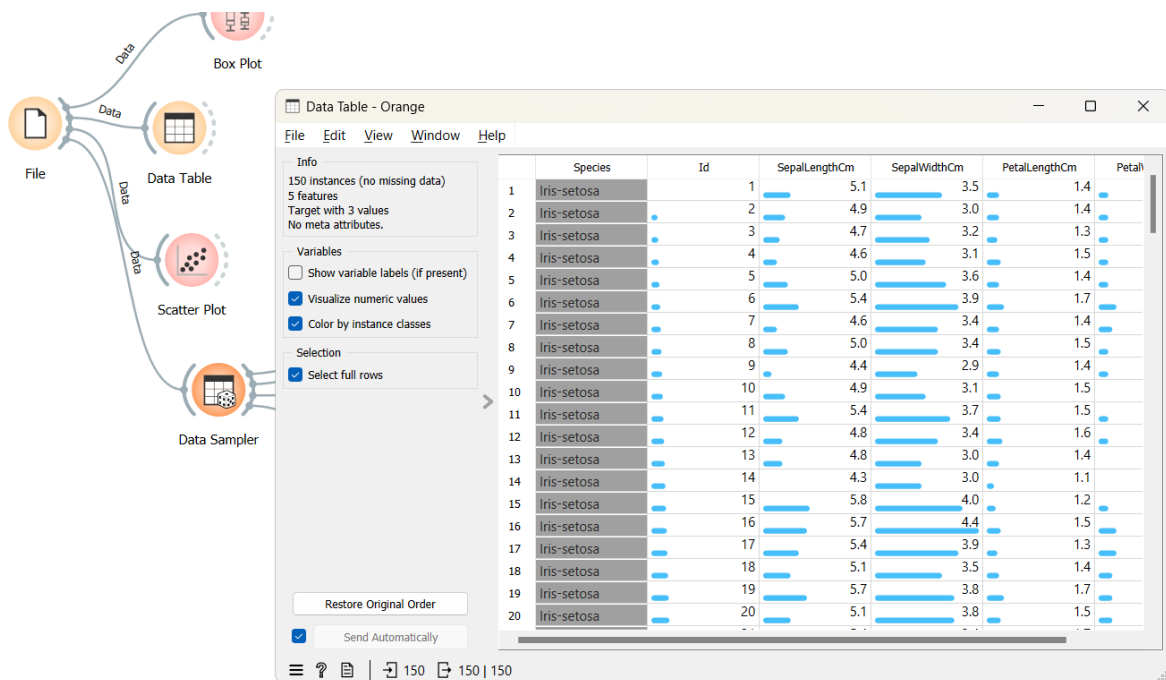
2. File



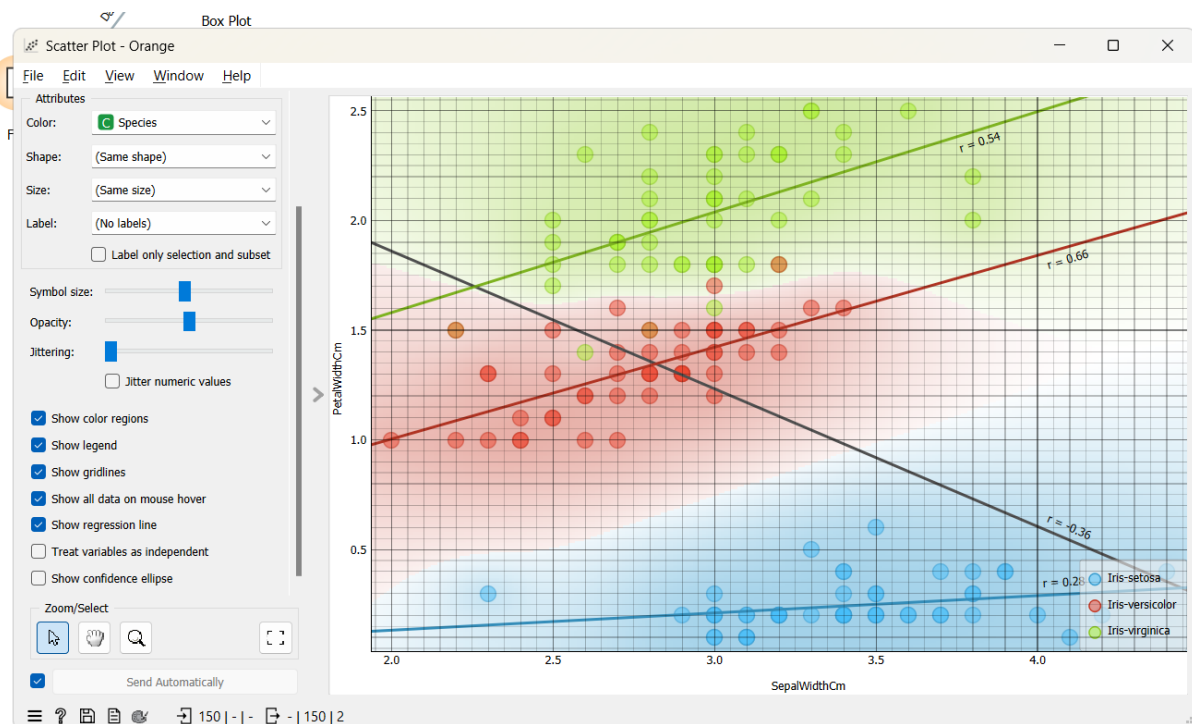
3. Box Plot



4. Data Table



5. Scatter Plot



6. Data Sampler

The screenshot shows the Orange Data Sampler widget in a workflow. The widget is connected to a Logistic Regression widget. The workflow includes a 'Data Sample' widget, a 'Test Data' widget, and a 'Data' widget. The 'Data Sampler' widget is connected to the 'Data Sample' widget, and the 'Test Data' widget is connected to the 'Data' widget. The 'Data' widget is connected to the 'Logistic Regression' widget. The 'Data Sampler' widget is also connected to the 'Data' widget. The 'Data Sampler' widget is connected to the 'Data' widget. The 'Data Sampler' widget is connected to the 'Data' widget.

Data Sampler - Orange

File View Window Help

Sampling Type

- ☒ Fixed proportion of data: 85 %
- ☐ Fixed sample size
- ☐ Cross validation
- ☐ Bootstrap

Instances: 1

☐ Sample with replacement

Number of subsets: 10

Unused subset: 1

Options

- ☒ Replicable (deterministic) sampling
- ☐ Stratify sample (when possible)

Sample Data

150 128

7. Logistic Regression

The screenshot shows the Orange Logistic Regression widget settings dialog. The dialog is titled 'Logistic Regression - Ora...'. It has a 'Name' field with the value 'Logistic Regression'. The 'Regularization type' is set to 'Ridge (L2)'. The 'Strength' is set to 'C=1' on a scale from 'Weak' to 'Strong'. There is a checkbox for 'Balance class distribution' which is unchecked. There is a checkbox for 'Apply Automatically' which is checked. The dialog also shows a 'Test' button and a 'Model' button.

Logistic Regression - Ora...

Name: Logistic Regression

Regularization type: Ridge (L2)

Strength: Weak Strong C=1

☐ Balance class distribution


☒ Apply Automatically

Test

Model

128 6

8. Test & Score


Test and Score - Orange

—
□
✕

☒ Cross validation

Number of folds: 5

☒ 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

Evaluation results for target (None, show average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	1.000	1.000	1.000	1.000	1.000	1.000

Compare models by: Area under ROC curve
☐ Negligible diff.: 0.1

Logistic Regression	Logistic ...
---------------------	--------------

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.

≡
?
📄

🔄 128 | 22
🖼️ 128 | 1×128

9. Predictions

The screenshot shows the Orange Data Mining interface with the "Predictions - Orange" window active. The window displays the results of a Logistic Regression model's predictions on a dataset.

Window Title: Predictions - Orange

File View Window Help

Show probabilities for: Classes in data ☒ Show classification errors [Restore Original Order](#)

	Logistic Regression	error	Species	Id	SepalLengthCm	SepalWidth
1	1.00 : 0.00 : 0.00 → Iris-setosa	0.000	Iris-setosa	22	5.1	3.7
2	0.01 : 0.99 : 0.00 → Iris-versicolor	0.007	Iris-versicolor	58	4.9	2.4
3	0.00 : 0.00 : 1.00 → Iris-virginica	0.000	Iris-virginica	145	6.7	3.3
4	0.00 : 0.00 : 1.00 → Iris-virginica	0.000	Iris-virginica	130	7.2	3.0
5	1.00 : 0.00 : 0.00 → Iris-setosa	0.000	Iris-setosa	38	4.9	3.1
6	0.00 : 0.00 : 1.00 → Iris-virginica	0.000	Iris-virginica	141	6.7	3.1
7	1.00 : 0.00 : 0.00 → Iris-setosa	0.000	Iris-setosa	2	4.9	3.0
8	0.02 : 0.98 : 0.00 → Iris-versicolor	0.022	Iris-versicolor	53	6.9	3.1
9	0.00 : 0.00 : 1.00 → Iris-virginica	0.000	Iris-virginica	131	7.4	2.8
10	0.00 : 0.01 : 0.99 → Iris-virginica	0.014	Iris-virginica	104	6.3	2.9
11	0.00 : 0.60 : 0.40 → Iris-versicolor	0.395	Iris-versicolor	100	5.7	2.8

Show performance scores Target class: (Average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	1.000	1.000	1.000	1.000	1.000	1.000

At the bottom of the window, there are icons for help, file operations, and zooming, along with a status bar showing coordinates like "22 | 22 | 1x22".