

TUGAS ORANGE WEEK 2

Nabila Putri Rihan

1103213055

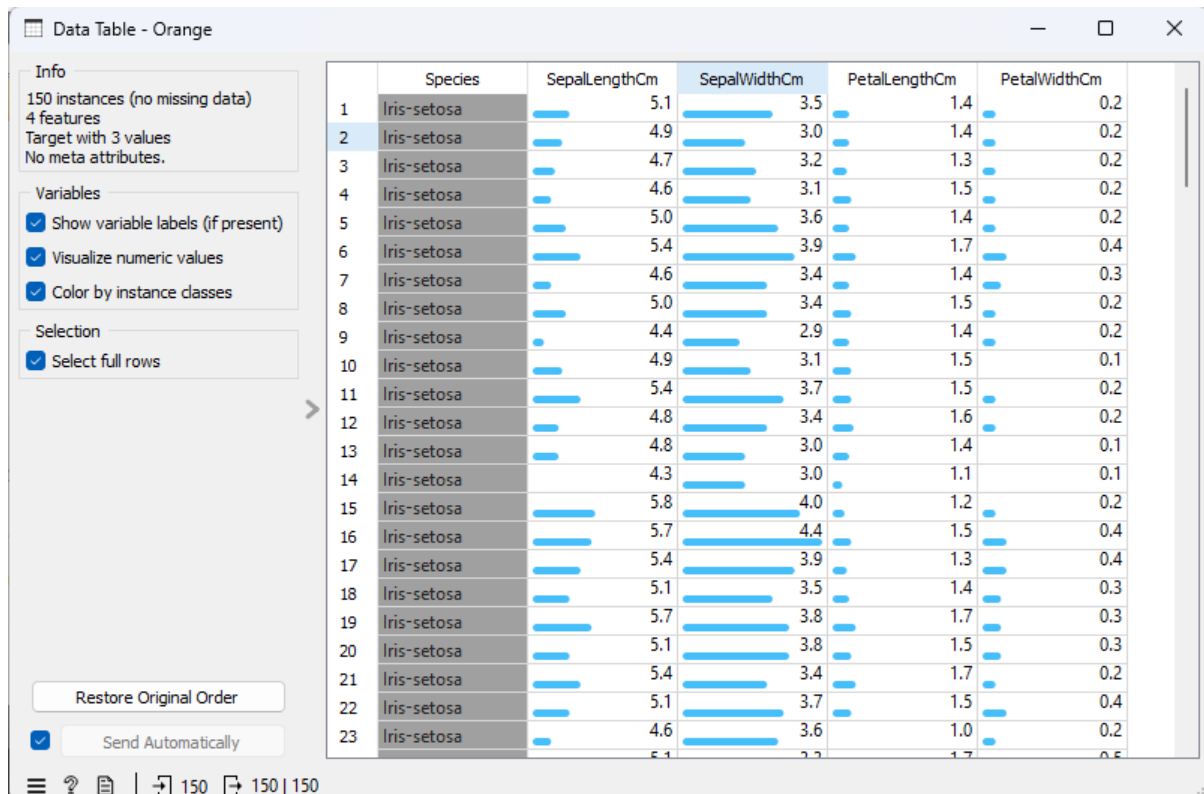
1. Memasukkan dataset Iris ke file orange

The screenshot shows the 'File' widget in the Orange Data Mining software. The 'Source' section has 'File' selected with the path 'Downloads\Iris.csv'. The 'File Type' is set to 'Automatically detect type'. The 'Info' section displays: 150 instances, 6 features (no missing values), Data has no target variable, and 0 meta attributes. The 'Columns' table is as follows:

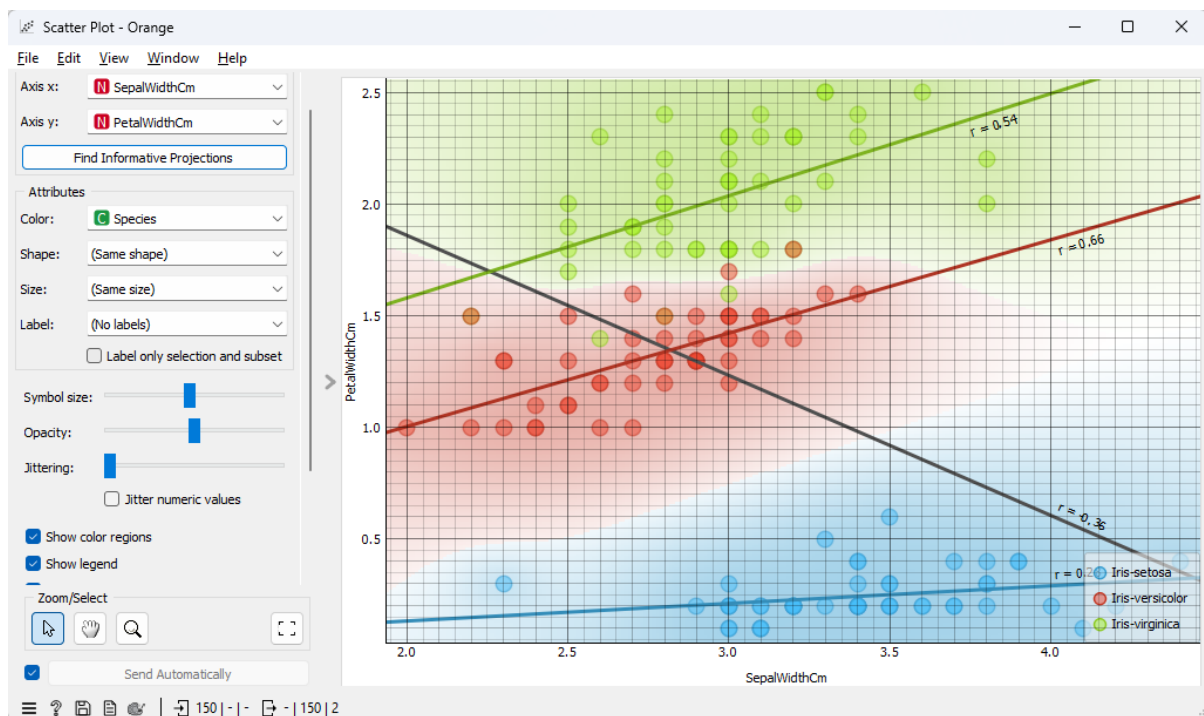
	Name	Type	Role	Values
1	Id	N numeric	skip	
2	SepalLengthCm	N numeric	feature	
3	SepalWidthCm	N numeric	feature	
4	PetalLengthCm	N numeric	feature	
5	PetalWidthCm	N numeric	feature	
6	Species	C categorical	target	Iris-setosa, Iris-versicolor, Iris-virginica

Buttons at the bottom include 'Reset', 'Apply', and 'Browse documentation datasets'. The status bar at the bottom left shows icons for file operations and a count of 150.

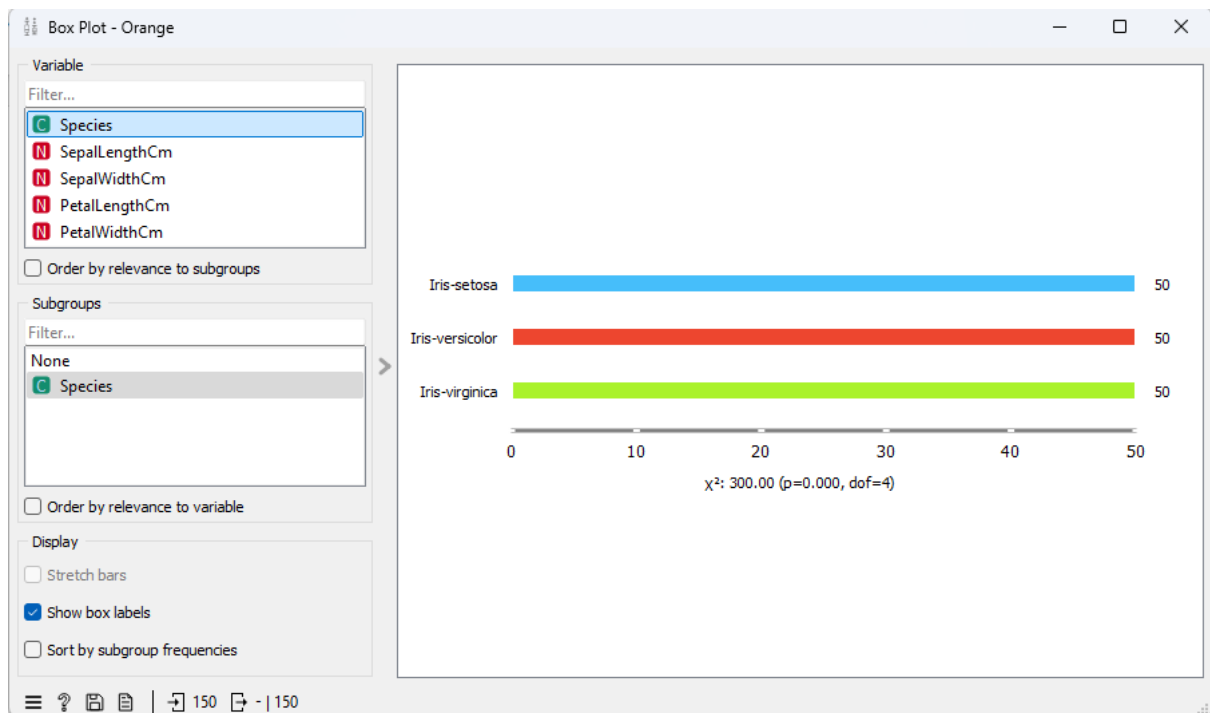
2. Menambah data table



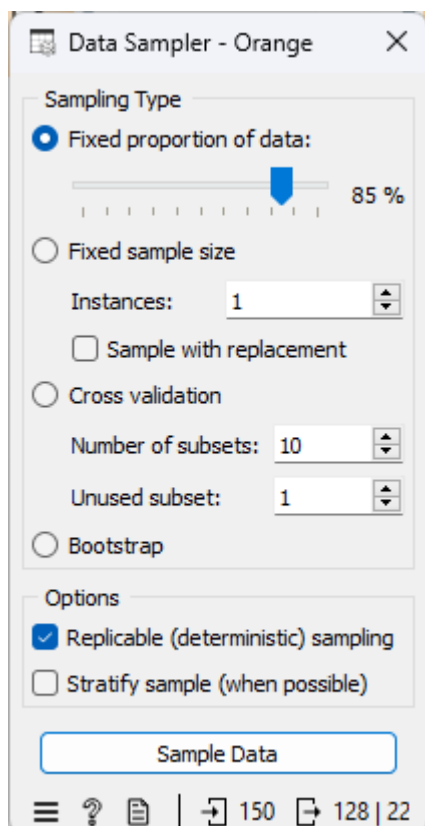
3. Membuat Scatterplot dan mengubah axis x dan y nya



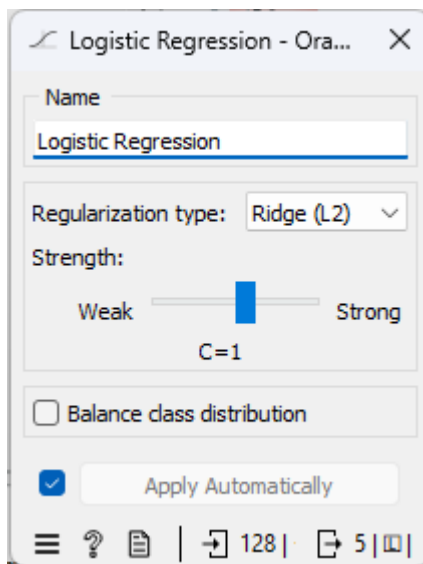
4. Mengganti variable di boxplot menjadi Iris



5. Menyesuaikan data sampler menjadi 85%



6. Menambahkan model Logistic Regression



7. Menguji performa model menggunakan test and 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	0.997	0.953	0.953	0.954	0.953	0.930

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

Model	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.

8. Menambahkan dan memastikan predictions untuk menerapkan model ke data baru

Predictions - Orange

Show probabilities for: Classes in data ☒ Show classification errors Restore Original Order

	Logistic Regression	error	Species	SepalLengthCm	SepalWidthCm
1	0.98 : 0.02 : 0.00 → Iris-setosa	0.020	Iris-setosa	5.1	3.7
2	0.16 : 0.84 : 0.00 → Iris-versicolor	0.160	Iris-versicolor	4.9	2.4
3	0.00 : 0.02 : 0.98 → Iris-virginica	0.016	Iris-virginica	6.7	3.3
4	0.00 : 0.12 : 0.88 → Iris-virginica	0.122	Iris-virginica	7.2	3.0
5	0.97 : 0.03 : 0.00 → Iris-setosa	0.032	Iris-setosa	4.9	3.1
6	0.00 : 0.03 : 0.97 → Iris-virginica	0.025	Iris-virginica	6.7	3.1
7	0.07 : 0.03 : 0.00 → Iris-setosa	0.020	Iris-setosa	4.9	3.0

☒ Show performance scores Target class: (Average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.994	0.955	0.955	0.960	0.955	0.933

22 | 22 | 1×22

9. Menambahkan confusion matrix untuk mengukur performansi model klasifikasi

Confusion Matrix - Orange

Clicking on cells or in headers outputs the corresponding data instances Ok, got it Show: Number of instances

		Predicted			Σ
		Iris-setosa	Iris-versicolor	Iris-virginica	
Actual	Iris-setosa	5	0	0	5
	Iris-versicolor	0	7	0	7
	Iris-virginica	0	1	9	10
Σ		5	8	9	22

☒ Predictions ☐ Probabilities

☒ Apply Automatically

Select Correct Select Misclassified Clear Selection

1×22 | 22