

ORANGE DATA MINING

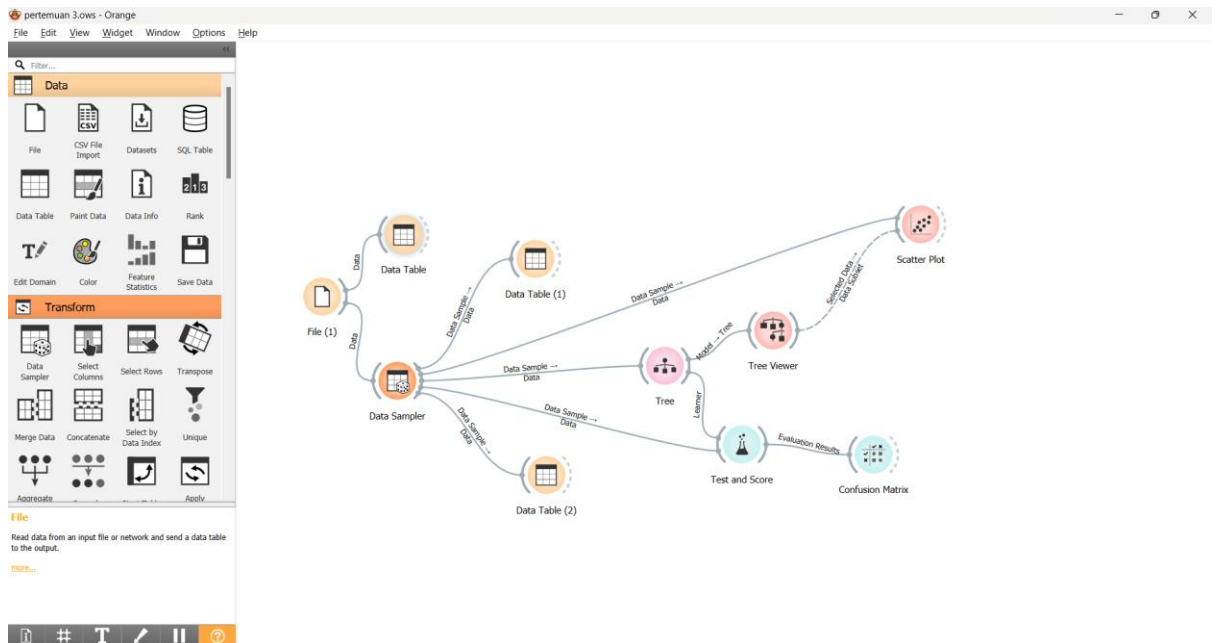
SALARY DATASET

MACHINE LEARNING WEEK 3

M Rakan Bagus

1103213162

1. FILE MINING



2. FILE DATASET DIABETES

The screenshot shows the 'File (1) - Orange' dialog box. The 'Source' section has 'File: Downloads\diabetes_dataset.csv' selected. The 'File Type' section has 'Automatically detect type' selected. The 'Info' section displays the following information:

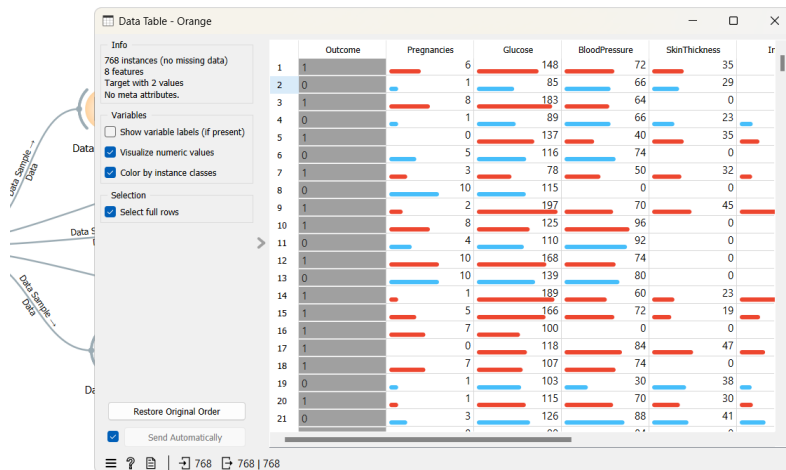
- 768 instances
- 9 features (no missing values)
- Data has no target variable.
- 0 meta attributes

The 'Columns (Double click to edit)' section contains a table with the following data:

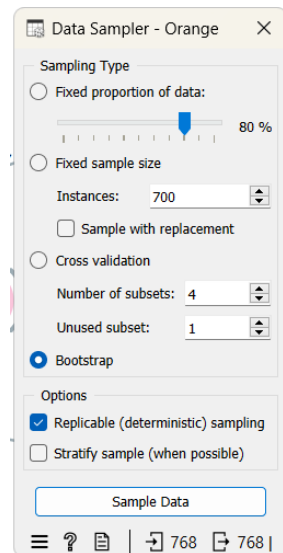
	Name	Type	Role	Values
1	Pregnancies	numeric	feature	
2	Glucose	numeric	feature	
3	BloodPressure	numeric	feature	
4	SkinThickness	numeric	feature	
5	Insulin	numeric	feature	
6	BMI	numeric	feature	
7	DiabetesPedigr...	numeric	feature	

At the bottom of the dialog box, there are 'Reset' and 'Apply' buttons, and a link to 'Browse documentation datasets'.

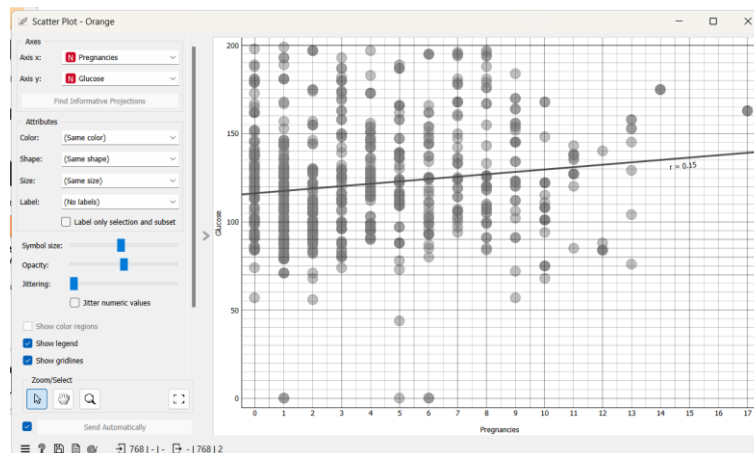
3. DATA TABLE



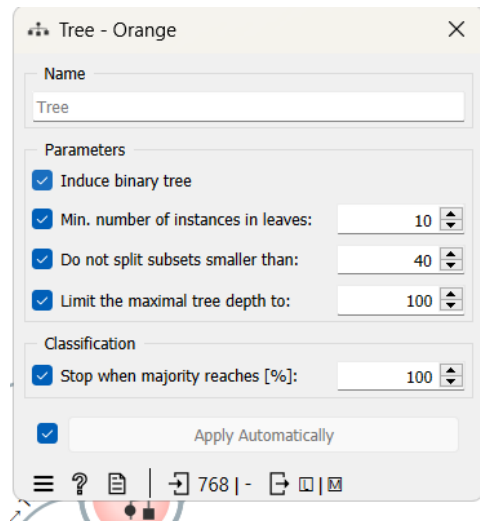
4. DATA SAMPLE



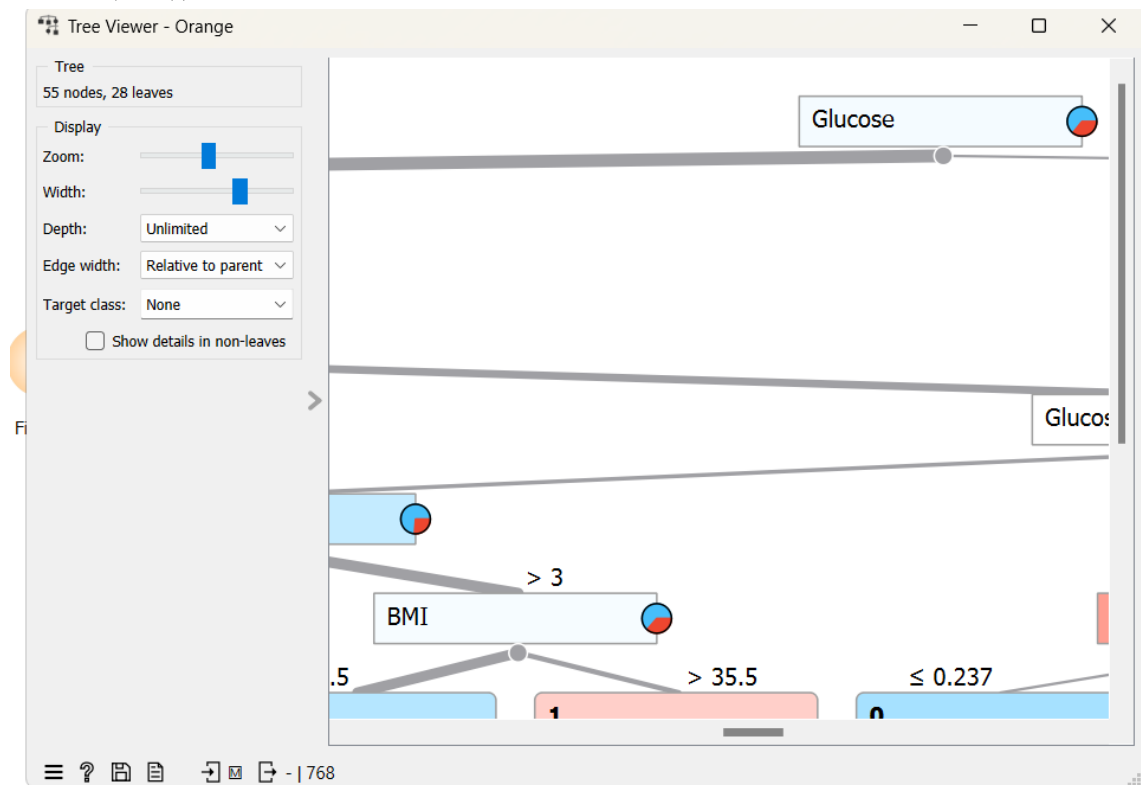
5. SCATTER PLOT



6. TREE



7. TREE VIEWER



8. TEST & SCORE 85%

Test and Score - Orange

☐ Cross validation

Number of folds: 5

☒ Stratified

☐ Cross validation by feature

☐ Random sampling

Repeat train/test: 100

Training set size: 95 %

☒ 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
Tree	0.938	0.855	0.857	0.861	0.855	0.696

Compare models by: Area under ROC curve

☐ Negligible diff.: 0.1

	Tree
Tree	

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.

768 | 1x768

9. CONFUSION MATRIX

Confusion Matrix - Orange

Clicking on cells or in headers outputs the corresponding data instances

Ok, got it

Show: Number of instances

	Predicted		
	0	1	Σ
0	422	71	493
1	40	235	275
Σ	462	306	768

Output

☒ Predictions

☐ Probabilities

☒ Apply Automatically

Select Correct

Select Misclassified

Clear Selection

1x768 | 768