

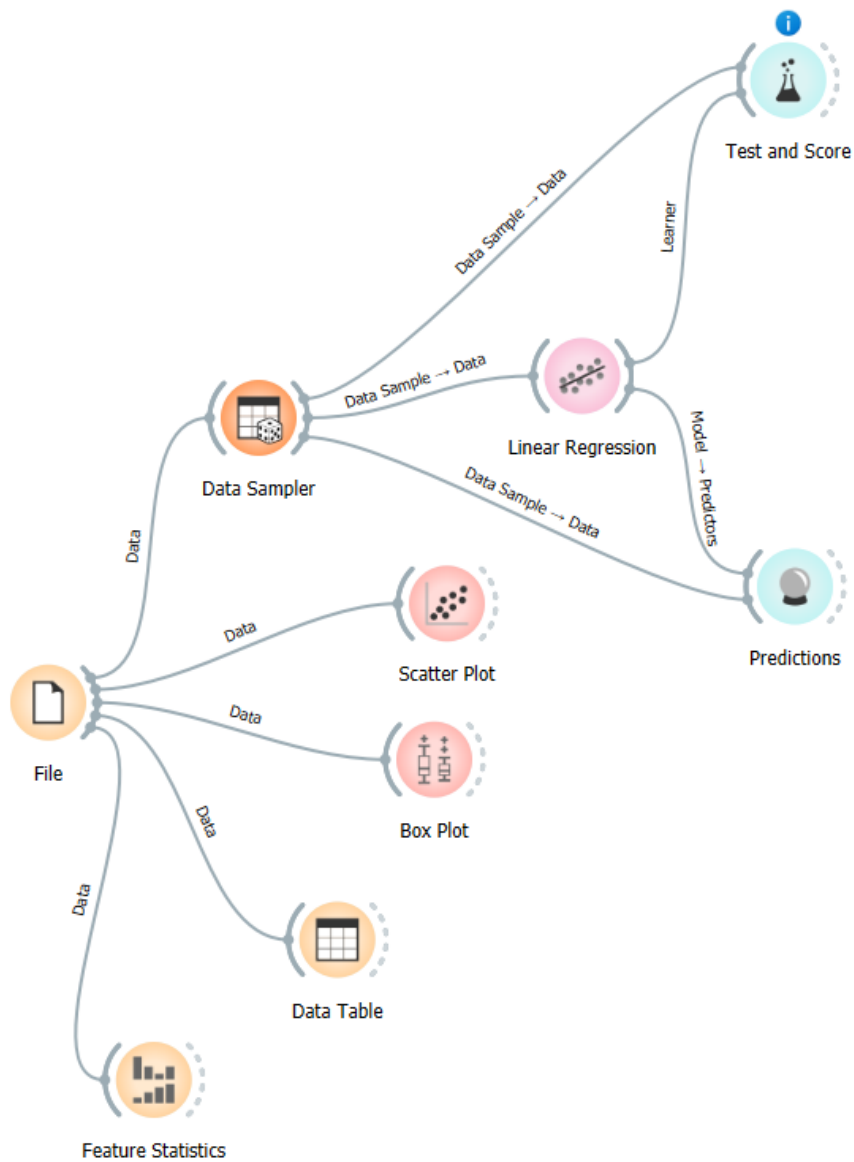
ORANGE DATA MINING

SALARY DATASET

MACHINE LEARNING WEEK 1

M Rakan Bagus

1103213162



File Edit View Widget Window Options Help

Filter...

Data

- File
- CSV File Import
- Datasets
- SQL Table
- Data Table
- Print Data
- Data Info
- Rank
- Edit Domain
- Color
- Feature Statistics
- Save Data

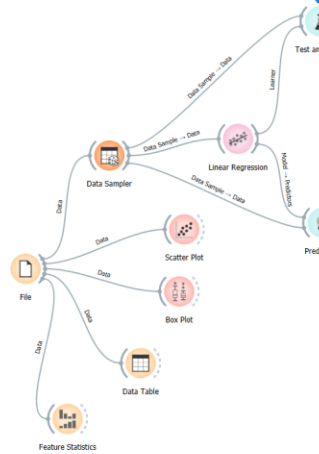
Transform

- Visualize
- Model
- Evaluate
- Unsupervised

Feature Statistics

Show basic statistics for data features.

More...

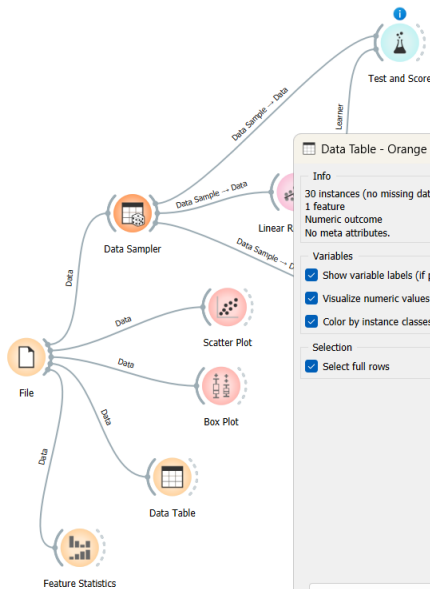


Feature Statistics - Orange

Name	Distribution	Mean	Mode	Median	Dispersion
YearsExperience		5.41333	3.3	4.8	0.515429
Salary		76004.0	37732.0	65238.0	0.355

Color: ■ Salary

Send Automatically



Data Table - Orange

Info

- 30 instances (no missing data)
- 1 feature
- Numeric outcome
- No meta attributes.

Variables

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

Selection

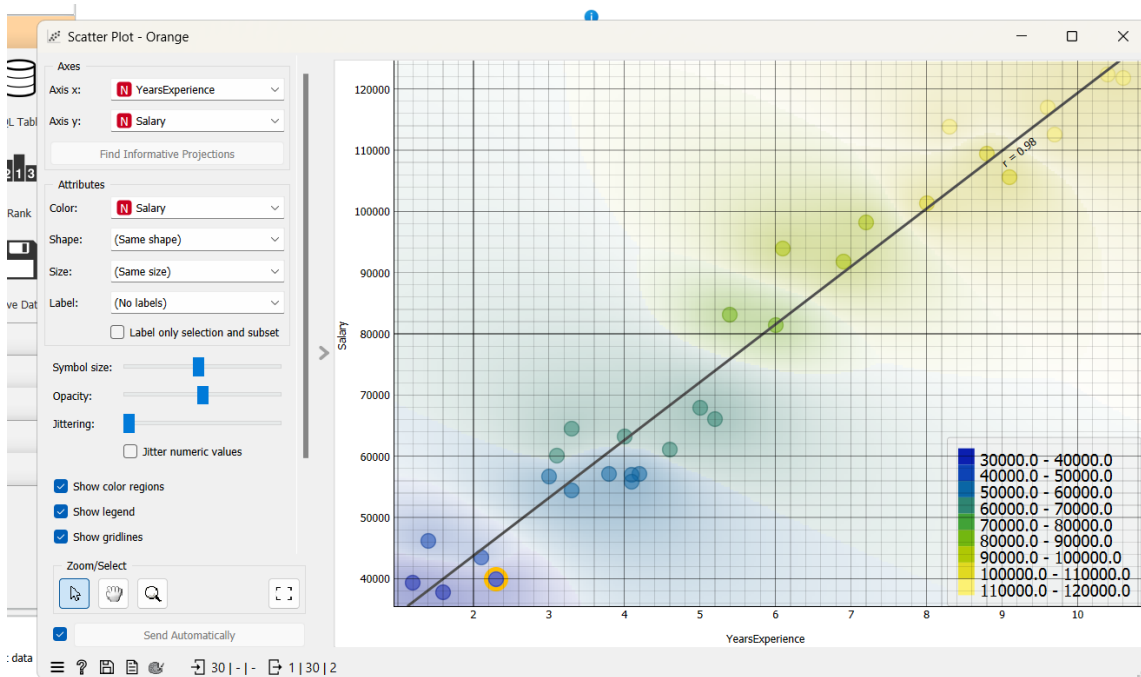
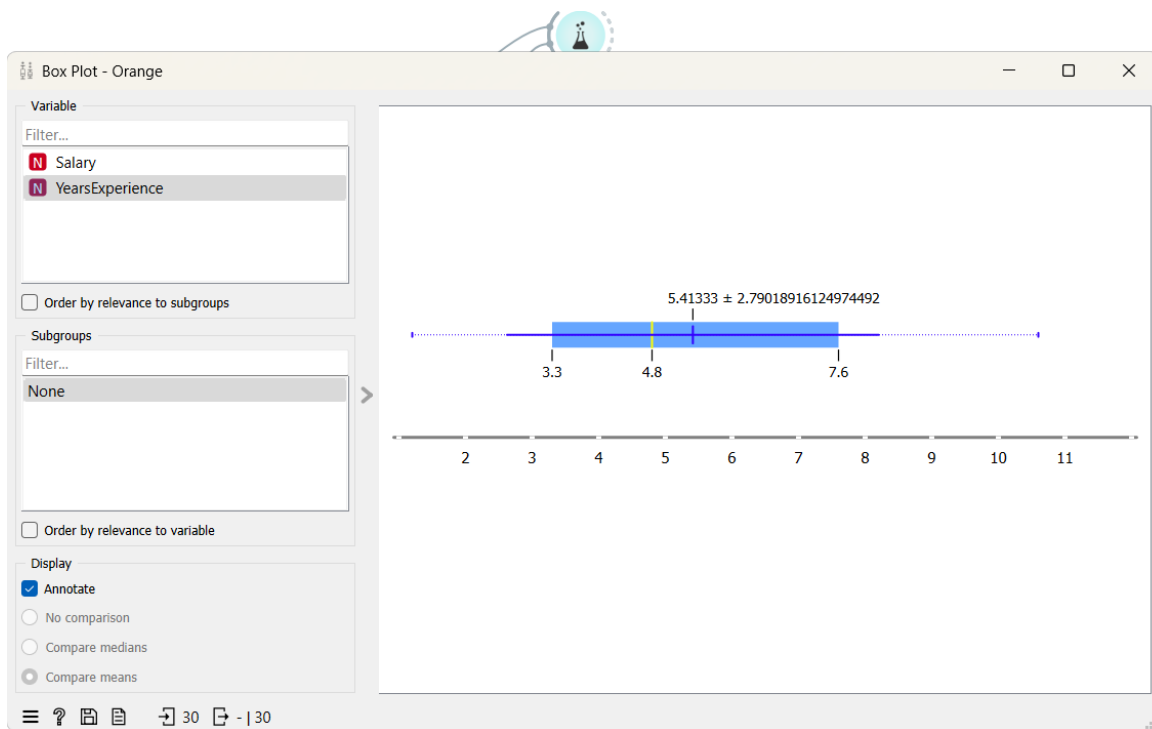
- ☒ Select full rows

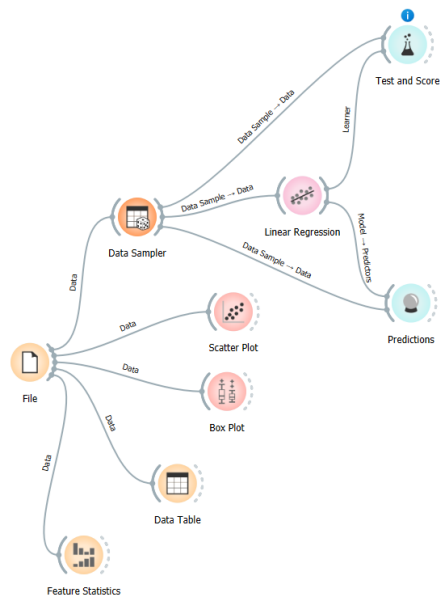
Restore Original Order

☒ Send Automatically

	Salary	YearsExperience
1	39344.0	1.2
2	46206.0	1.4
3	37732.0	1.6
4	43526.0	2.1
5	39892.0	2.3
6	56643.0	3
7	60151.0	3.1
8	54446.0	3.3
9	64446.0	3.3
10	57190.0	3.8
11	63219.0	4
12	55795.0	4.1
13	56958.0	4.1
14	57082.0	4.2
15	61112.0	4.6
16	67939.0	5
17	66030.0	5.2
18	83089.0	5.4
19	81364.0	6
20	93941.0	6.1
21	91739.0	6.9
22	98274.0	7.2

30 | 30





Predictions - Orange

Shown regression error: Difference

	Linear Regression	error	Salary	YearsExperience
1	116172.4	3536.4	112636.0	9.7
2	71666.8	3727.8	67939.0	5
3	102915.4	-1089...	113813.0	8.3
4	75454.5	-7634.5	83089.0	5.4
5	55569.1	-8876.9	64446.0	3.3
6	60303.7	3113.7	57190.0	3.8
7	122800.9	408.9	122392.0	10.4
8	107650.0	-1782.0	109432.0	8.8
9	63144.5	6186.5	56958.0	4.1
10	35683.6	-3660.4	39344.0	1.2
11	46099.8	6207.8	39892.0	2.3
12	73560.7	7530.7	66030.0	5.2
13	52728.3	-3914.7	56642.0	1.3

☒ Show performance scores

Model	MSE	RMSE	MAE	MAPE	R2
Linear Regression	33216979.733	5763.417	4990.261	0.078	0.957

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

Model	MSE	RMSE	MAE	MAPE	R2
Linear Regression	41...	6445...	568...	0.091	0.946

Compare models by: Mean square error

☐ Negligible diff.: 0.1

Linear Regression	Linear Re...
Linear 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.

Stratification is ignored for regression