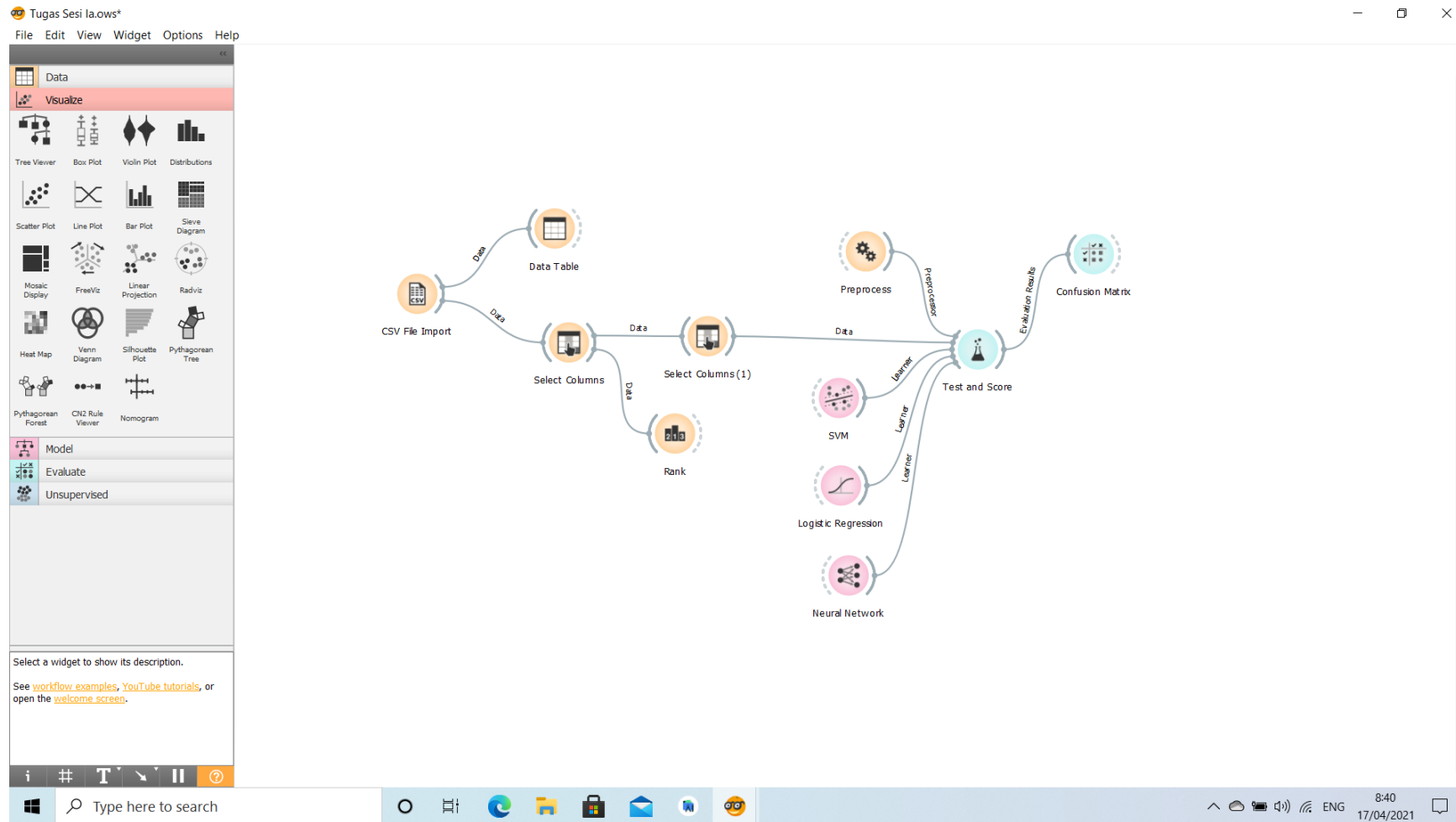


TUGAS SESI 1

A. Permodelan “Breast Cancer wisconsin”



Rank

Scoring Methods

- ☐ Information Gain
- ☒ Information Gain Ratio
- ☐ Gini Decrease
- ☒ ANOVA
- ☐ χ^2
- ☐ ReliefF
- ☐ FCBF

Select Attributes

- ☐ None
- ☐ All
- ☒ Manual
- ☐ Best ranked: 5

☒ Send Automatically

	#	Gai...tio	ANOVA
<input checked="" type="checkbox"/> radius_worst		0.310	860.782
<input checked="" type="checkbox"/> perimeter_worst		0.310	897.944
<input checked="" type="checkbox"/> area_worst		0.308	661.600
<input checked="" type="checkbox"/> concave points_worst		0.293	964.385
<input checked="" type="checkbox"/> concave points_mean		0.276	861.676
<input checked="" type="checkbox"/> perimeter_mean		0.249	697.235
<input checked="" type="checkbox"/> area_mean		0.247	573.061
<input checked="" type="checkbox"/> radius_mean		0.246	646.981
<input checked="" type="checkbox"/> concavity_mean		0.232	533.793
<input checked="" type="checkbox"/> area_se		0.225	243.652
<input checked="" type="checkbox"/> concavity_worst		0.223	436.692
<input checked="" type="checkbox"/> perimeter_se		0.157	253.897
<input checked="" type="checkbox"/> compactness_worst		0.148	304.341
<input checked="" type="checkbox"/> compactness_mean		0.146	313.233
<input checked="" type="checkbox"/> radius_se		0.140	268.840
<input checked="" type="checkbox"/> concavity_se		0.093	39.014
<input checked="" type="checkbox"/> concave points_se		0.090	113.263

Select Columns (1)

Ignored

Filter

- ☒ texture_mean
- ☒ area_mean
- ☒ smoothness_mean
- ☒ radius_mean
- ☒ perimeter_mean
- ☒ compactness_mean
- ☒ concave points_mean
- ☒ symmetry_mean
- ☒ fractal_dimension_mean
- ☒ radius_se
- ☒ texture_se
- ☒ perimeter_se
- ☒ area_se
- ☒ smoothness_se
- ☒ compactness_se
- ☒ concavity_se
- ☒ concave points_se
- ☒ symmetry_se
- ☒ fractal_dimension_se

Features

Filter

- ☒ radius_worst
- ☒ perimeter_worst
- ☒ area_worst
- ☒ concave points_worst
- ☒ concavity_mean

Target

☒ diagnosis

Metas

☒ Send Automatically

Reset ☐ Ignore new variables by default

569 569

Test and Score

Sampling

☒ 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

Target Class

(Average over classes)

Model Comparison

Area under ROC curve

☐ Negligible difference: 0.1

569
 569

Evaluation Results

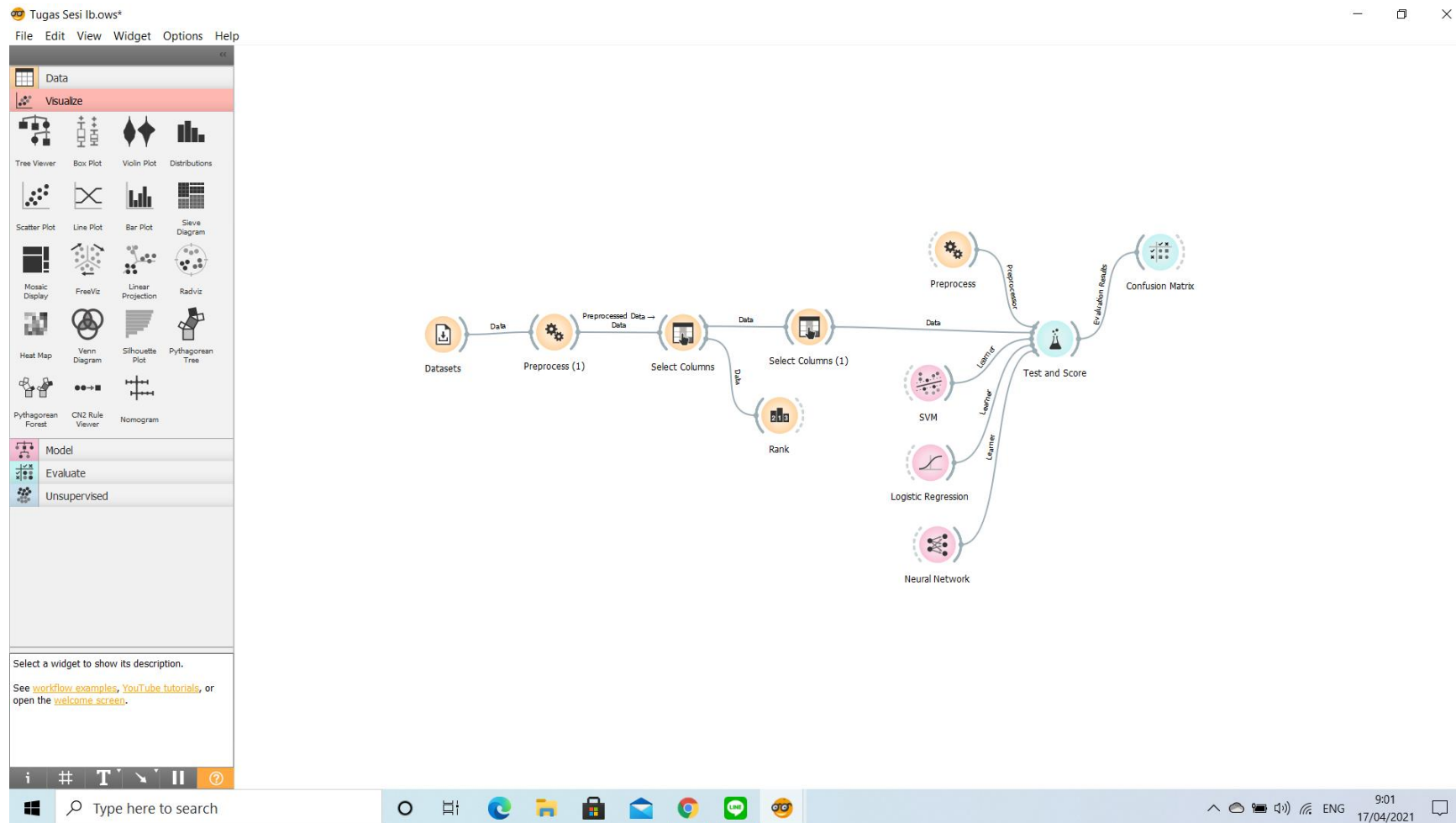
Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.987	0.953	0.953	0.953	0.953
Neural Network	0.985	0.942	0.942	0.942	0.942
SVM	0.983	0.947	0.947	0.948	0.947

Model Comparison by AUC

	Logistic Re...	Neural Net...	SVM
Logistic Regression		0.691	0.937
Neural Network	0.309		0.899
SVM	0.063	0.101	

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.

B. Permodelan “Heart Disease”



Rank

Scoring Methods

- ☐ Information Gain
- ☒ Information Gain Ratio
- ☐ Gini Decrease
- ☐ ANOVA
- ☒ χ^2
- ☐ ReliefF
- ☐ FCBF

Select Attributes

- ☐ None
- ☐ All
- ☒ Manual
- ☐ Best ranked: 5

☒ Send Automatically

	#	Gai...tio	χ^2
N major vessels colored		0.117	82.731
C chest pain	4	0.113	59.872
N ST by exercise		0.073	49.211
C thal	3	0.168	47.790
N max HR		0.063	41.391
C exerc ind ang	2	0.145	35.508
C slope peak exc ST	3	0.084	20.819
N aqe		0.031	15.324
C qender	2	0.064	7.444
N cholesterol		0.008	5.314
C rest ECG	3	0.022	5.095
N rest SBP		0.008	2.426
C fasting blo...sugar > 120	2	0.000	0.003

Select Columns (1)

Ignored

Filter

- N** age
- C** gender
- N** rest SBP
- N** cholesterol
- C** fasting blood sugar > 120
- C** rest ECG
- C** slope peak exc ST
- N** max HR
- C** exerc ind ang

Features

Filter

- N** major vessels colored
- C** chest pain
- N** ST by exercise
- C** thal

Target

C diameter narrowing

Metas

Reset ☐ Ignore new variables by default ☒ Send Automatically

? | 297 | 297

Test and Score

Sampling

☒ 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

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	0.899	0.818	0.817	0.819	0.818
Neural Network	0.891	0.838	0.838	0.838	0.838
SVM	0.875	0.822	0.819	0.827	0.822

Model Comparison by AUC

	SVM	Neural Net...	Logistic Re...
SVM		0.634	0.485
Neural Network	0.366		0.249
Logistic Regression	0.515	0.751	

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.

? | 297 | 297