

LAPORAN PRAKTIKUM DATA WAREHOUSING DAN DATA MINING

MODUL 9 “KLASIFIKASI: DECISION TREE”



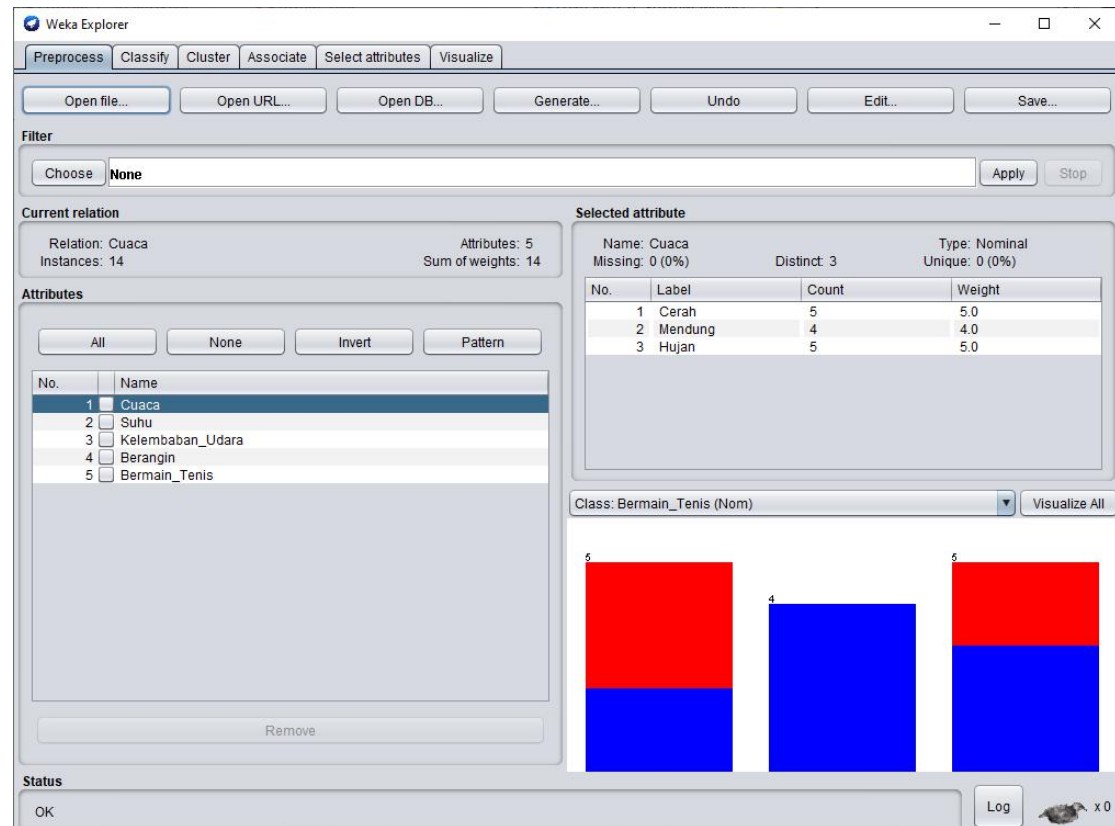
Oleh:

NAMA	: Daffa Putra Alwansyah
NIM	: L200190031
KELAS	: B
PRODI	: INFORMATIKA

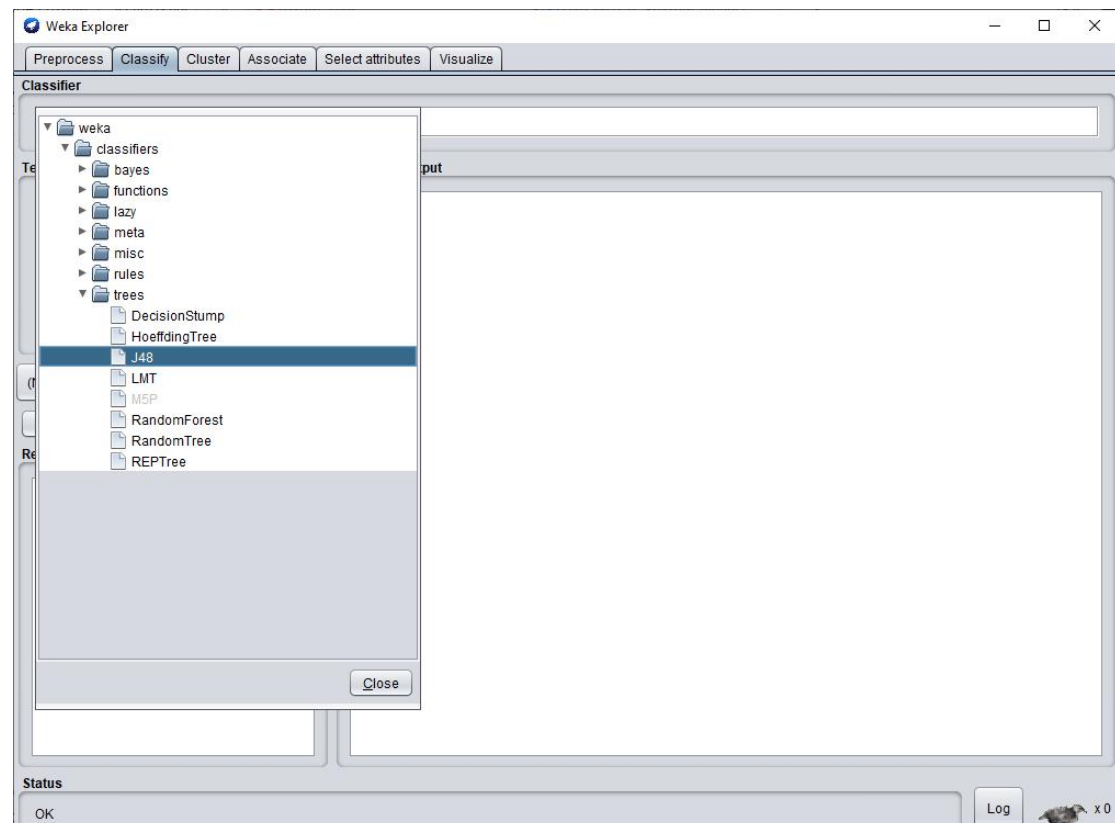
**Fakultas Komunikasi dan Informatika
Universitas Muhammadiyah Surakarta**

Langkah-Langkah Praktikum

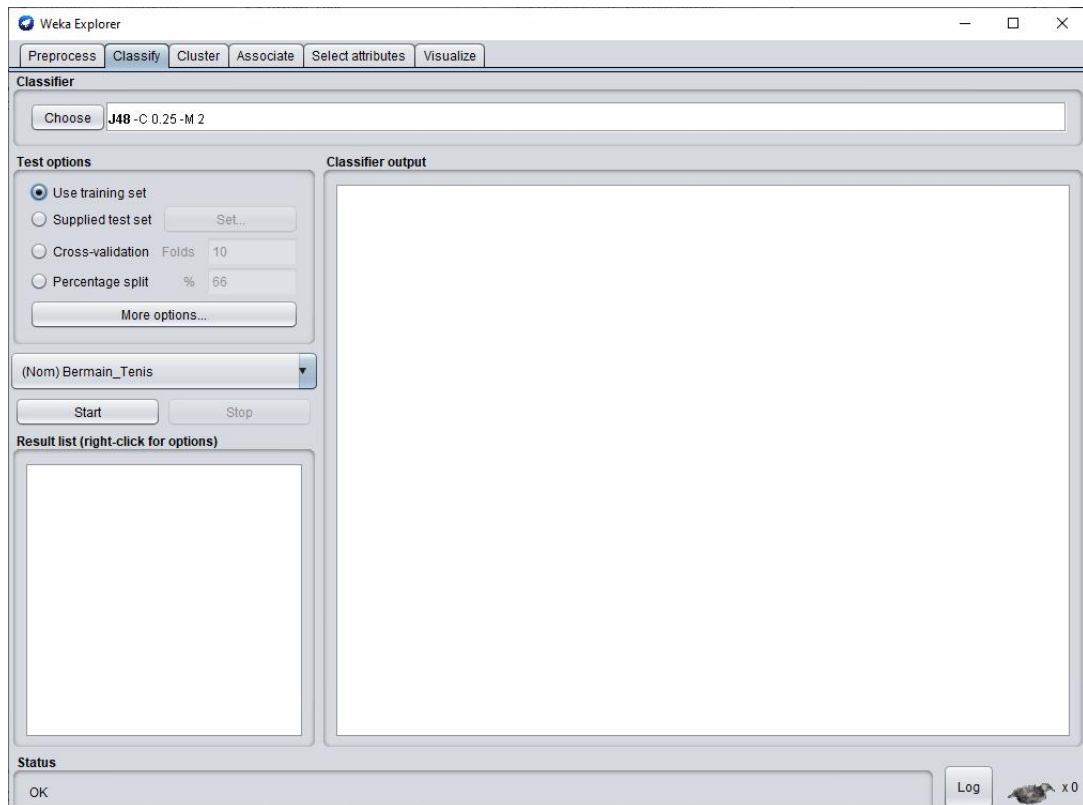
1. Buka kembali Cuaca.arff dengan Weka Explorer.



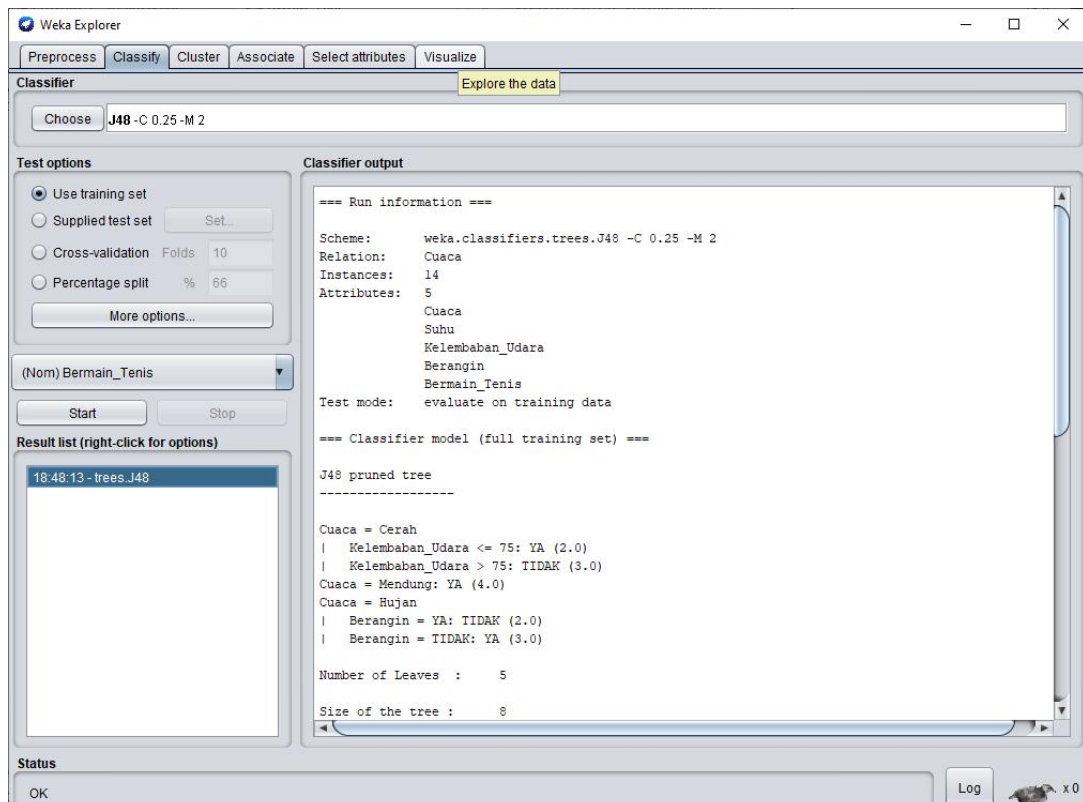
2. Pilih tab Classify, tekan tombol Choose - Trees - J48.



3. Pilih Use Training Set, pastikan pilihan atribut dependen adalah Bermain_Tenis, Data Cuaca.arff digunakan sebagai data training untuk menciptakan klasifikasi.



4. Setelah proses selesai, maka akan ditampilkan 2 hasil perhitungan klasifikasi menggunakan algoritma J48. Yaitu kolom Result List dan kolom Classifier output.



5. Mengamati kolom Classifier output.

J48 pruned tree

Cuaca = Cerah

| Kelembaban_Udara <= 75: YA (2.0)

| Kelembaban_Udara > 75: TIDAK (3.0)

Cuaca = Mendung: YA (4.0)

Cuaca = Hujan

| Berangin = YA: TIDAK (2.0)

| Berangin = TIDAK: YA (3.0)

Number of Leaves : 5

Size of the tree : 8

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	14	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	14	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	14		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	YA
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	TIDAK
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

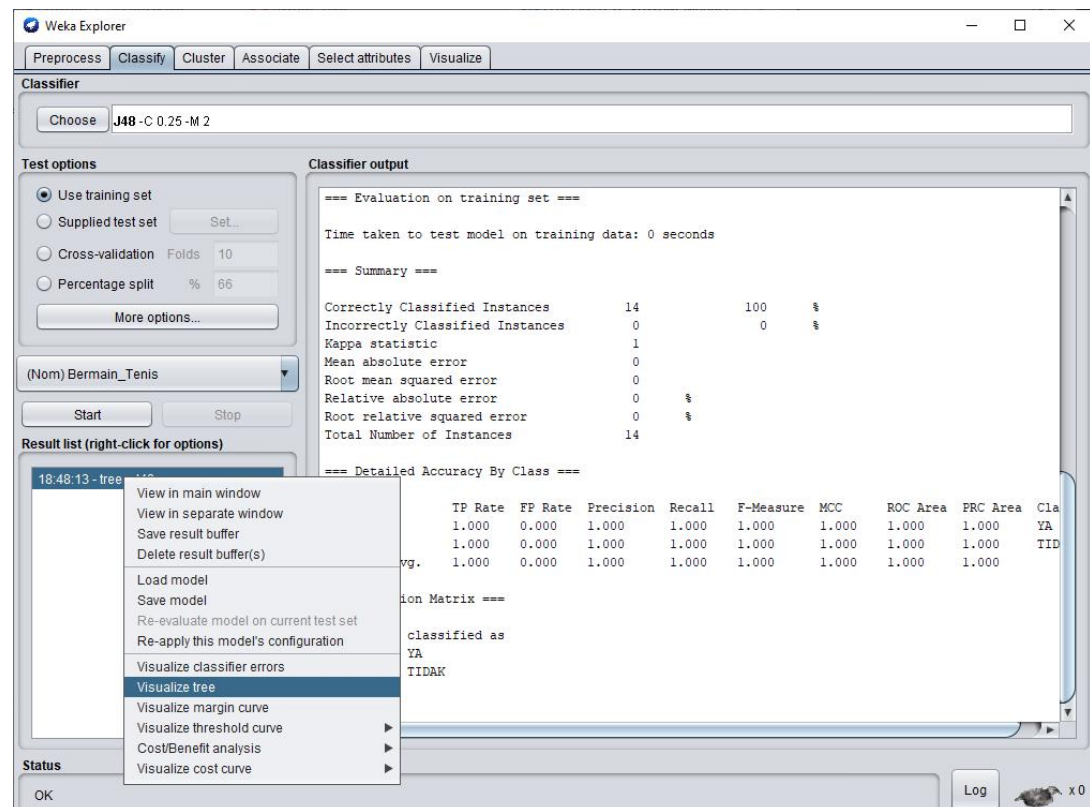
=== Confusion Matrix ===

a b <-- classified as

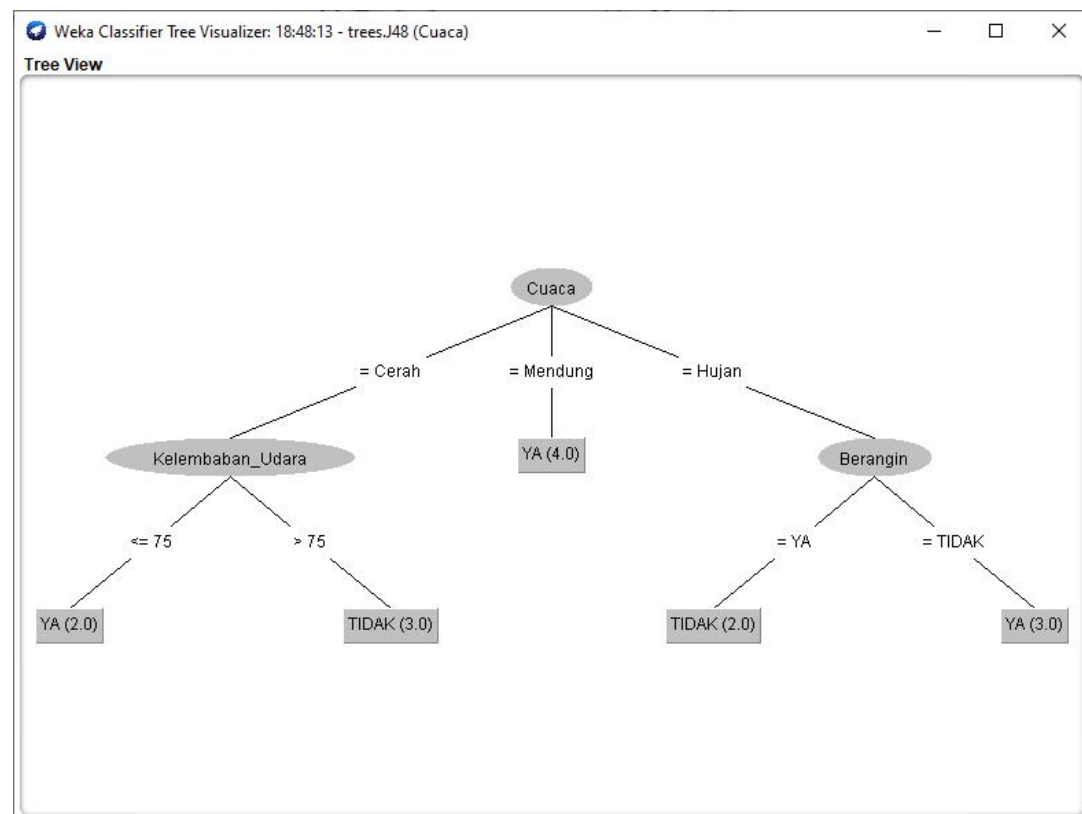
9 0 | a = YA

0 5 | b = TIDAK

6. Untuk melihat skema pohon keputusan klik kanan hasil dari test, lalu pilih Visualize Tree.



7. Hasil dari Visualize Tree.



8. Berdasarkan pohon keputusan tersebut, dapat dilihat jenis-jenis simpul yang ada, sebagai berikut :

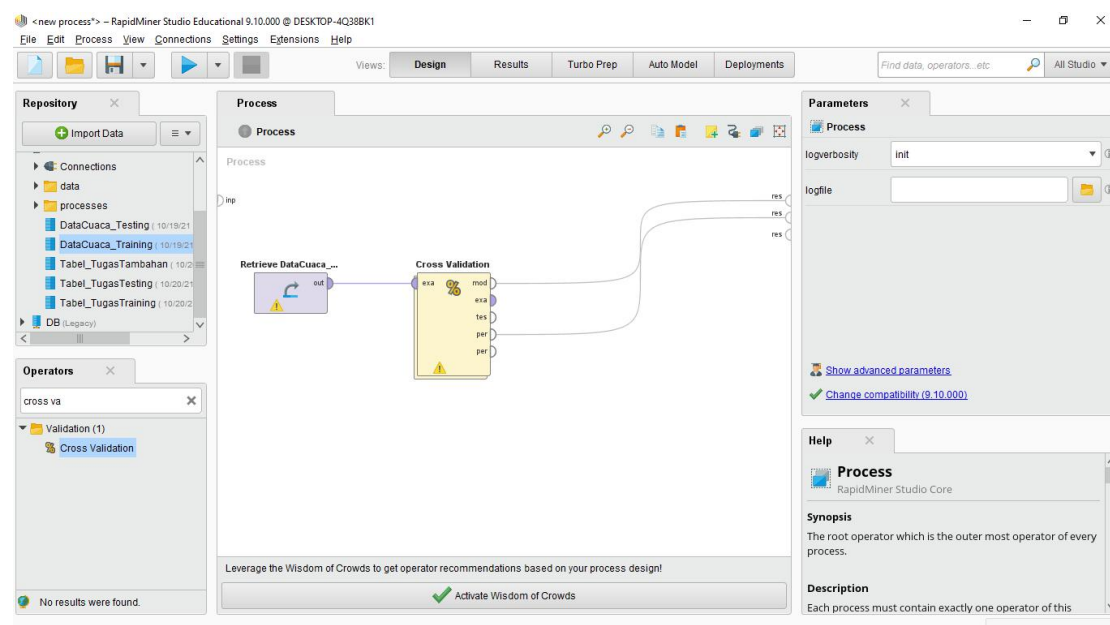
- a. Simpul akar = Cuaca
- b. Simpul internal = Kelembaban_udara, Berangin
- c. Simpul daun = YA, TIDAK

9. Klasifikasi yang terbentuk yaitu :

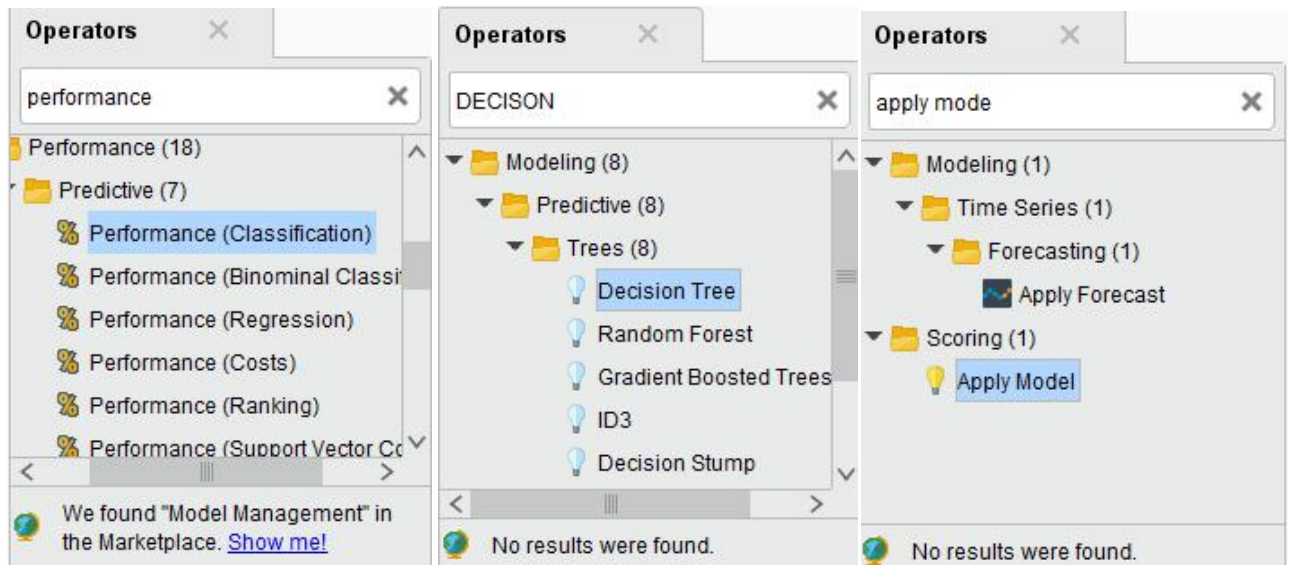
- a. Seseorang akan bermain tenis (YA) jika kondisi sebagai berikut:
 - i. Cuaca = Cerah, Kelembaban_udara > 75, (nilai atribut lain diabaikan)
 - ii. Cuaca = Mendung, (kondisi lain diabaikan)
 - iii. Cuaca = Hujan, Berangin = TIDAK, (nilai atribut lain diabaikan)
- b. Seseorang tidak akan bermain tenis (TIDAK) jika kondisi sebagai berikut:
 - i. Cuaca = Cerah, Kelembaban_udara > 75, (nilai atribut lain diabaikan)
 - ii. Cuaca = Hujan, Berangin = YA, (nilai atribut lain diabaikan).

Pohon keputusan menggunakan RapidMiner

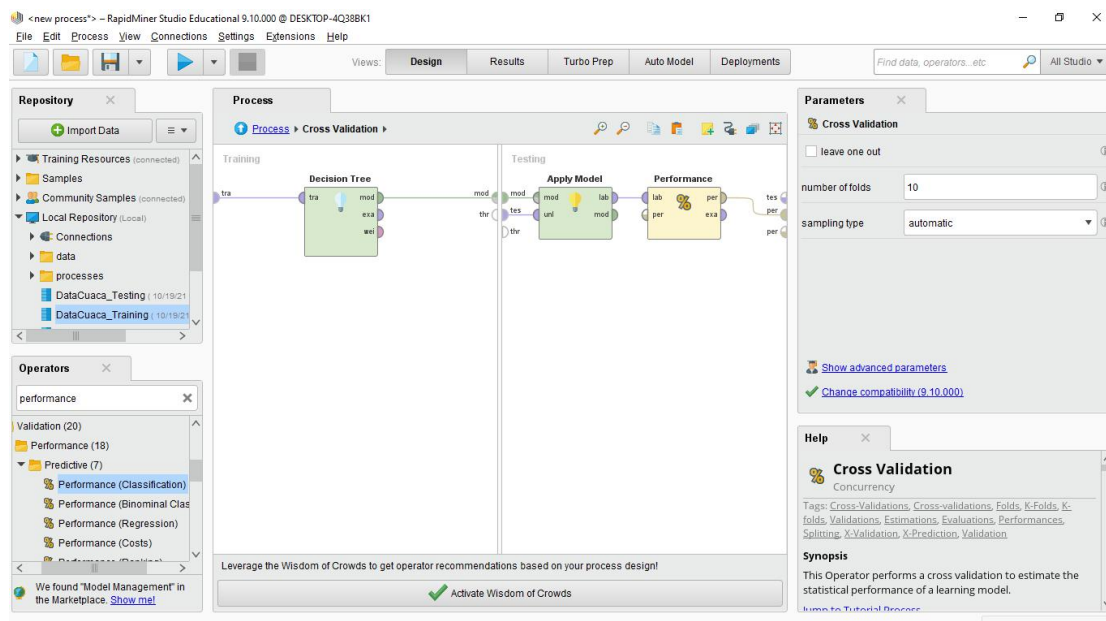
1. Klik New Process > Blank pada halaman. Klik Local Repository, Drag DataCuacaTraining dan Cross Validation pada Operators, klik Cross Validation 2x.



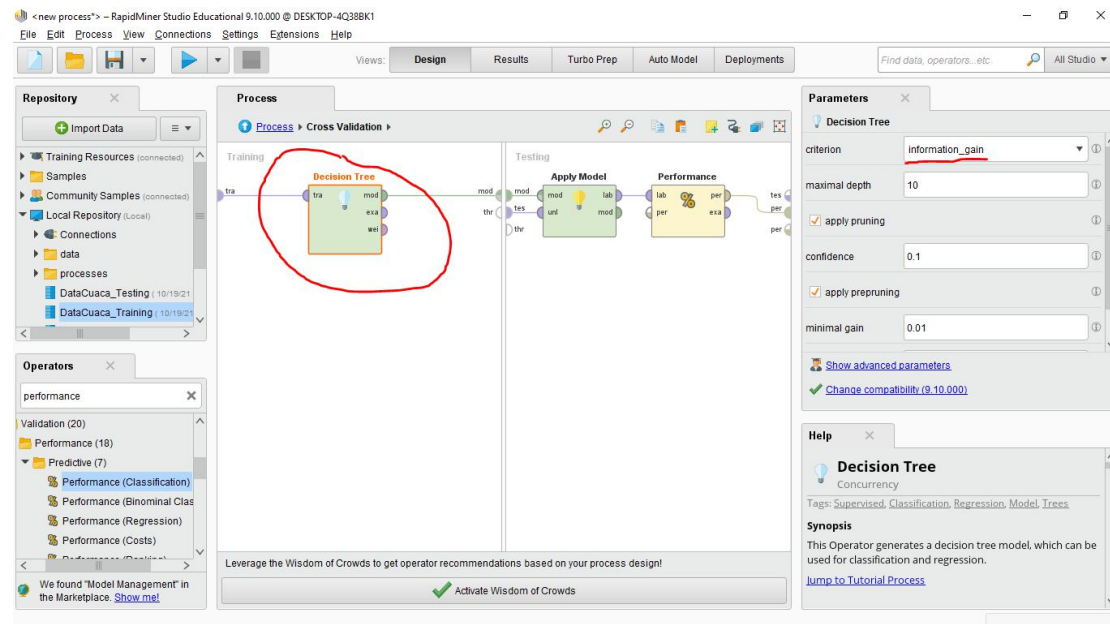
2. Cari di Operators pada gambar berikut.



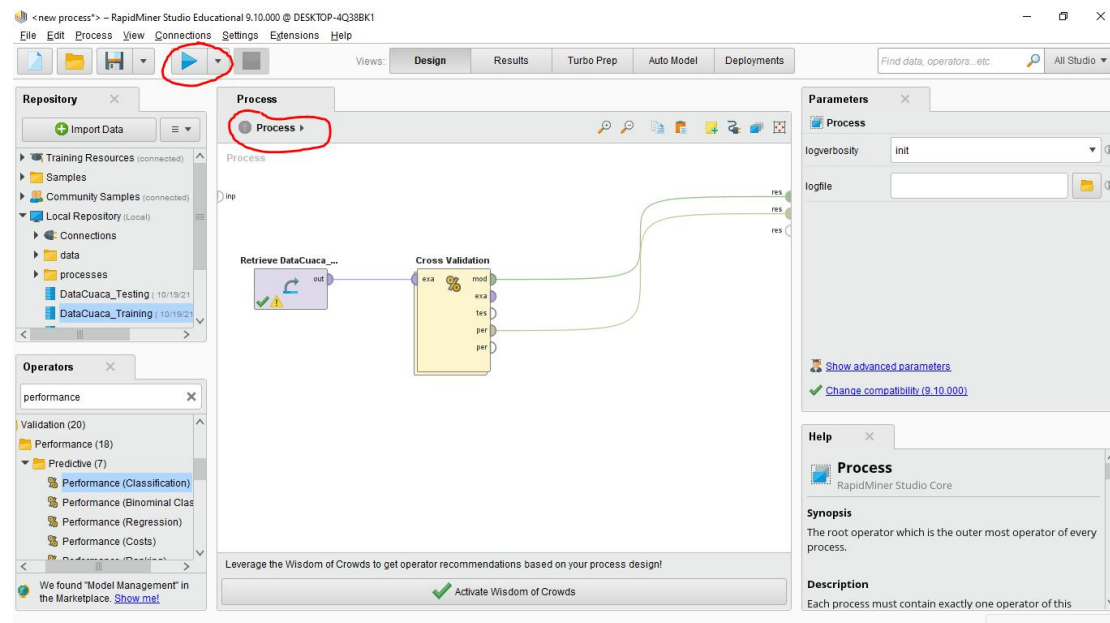
3. Hubungkan kabel sesuai modul.(port input dan ouput masing-masing operator).



4. klik pada Decision Tree, lalu ubah criterion menjadi information_gain, dan pastikan kriteria yang dipakai adalah Information Gain (J48) pada kolom Parameter.

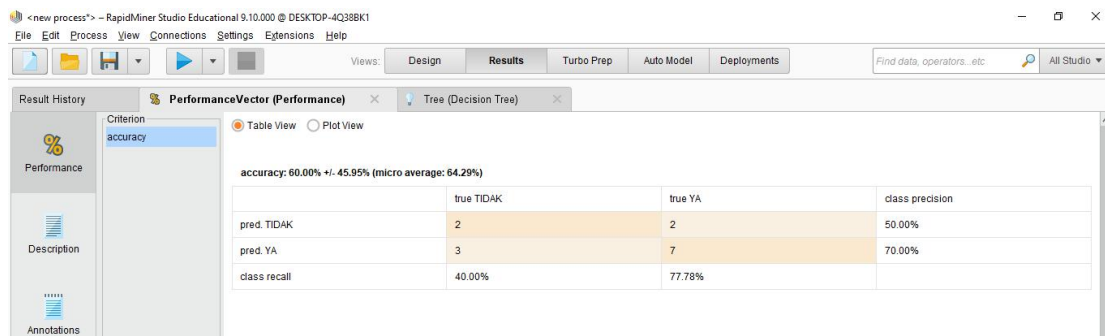


5. Klik Process, lalu klik tanda panah biru(RUN).



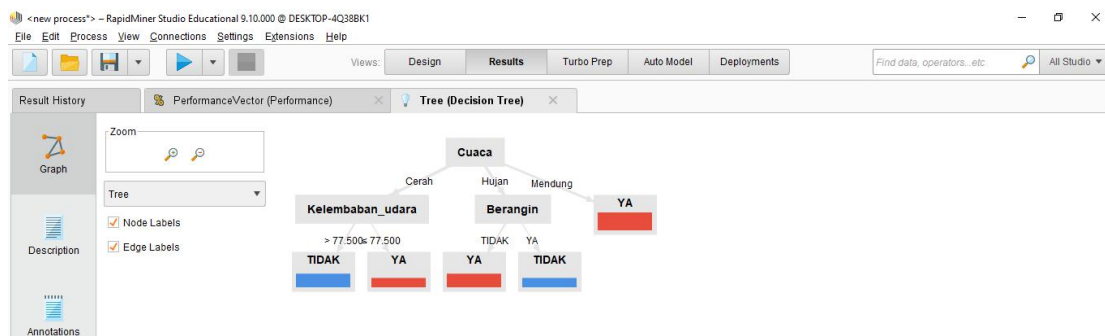
6. Berikut 2 hasil dari proses klasifikasi pohon keputusan:

a. PerformanceVector (Performance), menunjukkan tingkat akurasi, presisi, recal dan lain-lain dalam bentuk tabel dan plot view.



	true TIDAK	true YA	class precision
pred. TIDAK	2	2	50.00%
pred. YA	3	7	70.00%
class recall	40.00%	77.78%	

b. Tree (Decision Tree) menunjukkan hasil pohon keputusan dari proses klasifikasi. Dapat dilihat pohon keputusan yang dihasilkan sama seperti saat menggunakan Weka.



7. Berdasarkan pohon keputusan tersebut, dapat dilihat jenis-jenis simpul yang ada, sebagai berikut:

- Simpul akar = Cuaca
- Simpul Internal = Kelembaban_udara, dan Berangin
- Simpul daun = YA, TIDAK

8. Klasifikasi yang terbentuk yaitu :

- Seseorang akan bermain tenis (YA) jika kondisi sebagai berikut:
 - Cuaca = Cerah, Kelembaban_udara $\leq 77,5$ (nilai atribut lain diabaikan).
 - Cuaca = Mendung. (nilai atribut lain diabaikan).
 - Cuaca = Hujan, Berangin = TIDAK (nilai atribut lain diabaikan).
- Seseorang tidak akan bermain tenis (TIDAK) jika kondisi sebagai berikut:
 - Cuaca = Cerah, Kelembaban_udara $> 77,5$ (nilai atribut lain diabaikan)
 - Cuaca = Hujan, Berangin = YA (nilai atribut lain diabaikan)

TUGAS

1. Berdasarkan pohon keputusan pada kegiatan 9.4.2 (Menggunakan RapidMiner).

Isikan nilai kelas ke atribut Bermain_Tenis pada table Testing berikut:

Cuaca	Suhu	Kelembaban_Udara	Berangin	Bermain_Tenis
Cerah	75	65	TIDAK	YA
Cerah	80	68	YA	YA
Cerah	83	87	YA	TIDAK
Mendung	70	96	TIDAK	YA
Mendung	68	81	TIDAK	YA
Hujan	65	75	TIDAK	TIDAK
Hujan	64	85	YA	TIDAK

2. Gunakan file ARFF yang dikerjakan pada tugas nomor 1 dalam modul 7 sebagai data training.

a. Pohon keputusan berdasarkan data tersebut!

The screenshot shows the Weka Explorer interface. The 'Classifier' tab is active, and the 'J48 -C 0.25 -M 2' model is selected. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following information:

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Universitas
Instances:   20
Attributes:  6
  Jurusan_SMA
  Gender
  Asai_Sekolah
  Rerata_SKS
  Asisten
  Lama_Studi

Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

Gender = PRIA
| Rerata_SKS <= 18: TERLAMBAT (4.0)
| Rerata_SKS > 18: TEPAT (11.0/2.0)
Gender = WANITA: TEPAT (5.0/1.0)

Number of Leaves :    3
Size of the tree :    5

Time taken to build model: 0.01 seconds
  
```

The 'Result list' on the left shows the selected model '20:05:52 - trees.J48'.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) Lama_Studi

Start Stop

Result list (right-click for options)

20:05:52 - trees.J48

Classifier output

Size of the tree : 5

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	12	60	%
Incorrectly Classified Instances	8	40	%
Kappa statistic	0.0588		
Mean absolute error	0.4492		
Root mean squared error	0.5312		
Relative absolute error	95.585 %		
Root relative squared error	108.7099 %		
Total Number of Instances	20		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.769	0.714	0.667	0.769	0.714	0.061	0.462	0.617	TEPAT
	0.286	0.231	0.400	0.286	0.333	0.061	0.462	0.381	TERLAMBAT
Weighted Avg.	0.600	0.545	0.573	0.600	0.581	0.061	0.462	0.535	

=== Confusion Matrix ===

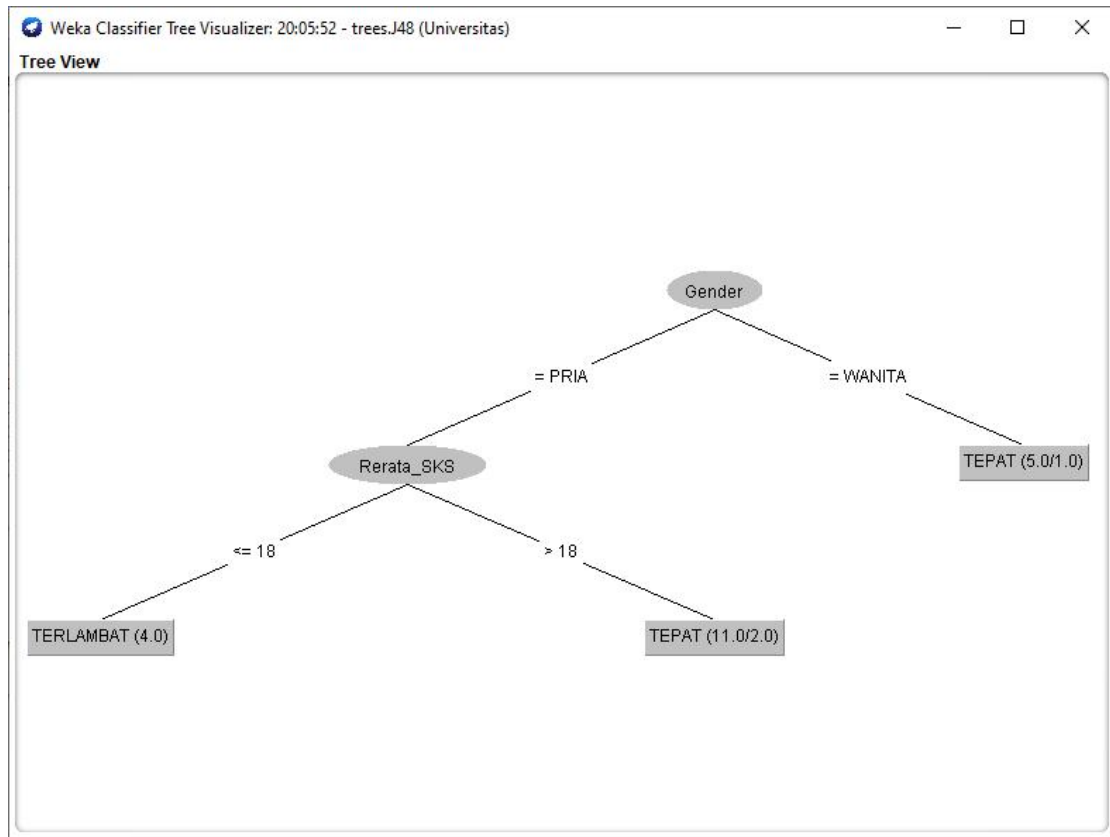
a b <-- classified as

10 3 | a = TEPAT

5 2 | b = TERLAMBAT

Status

OK Log x 0

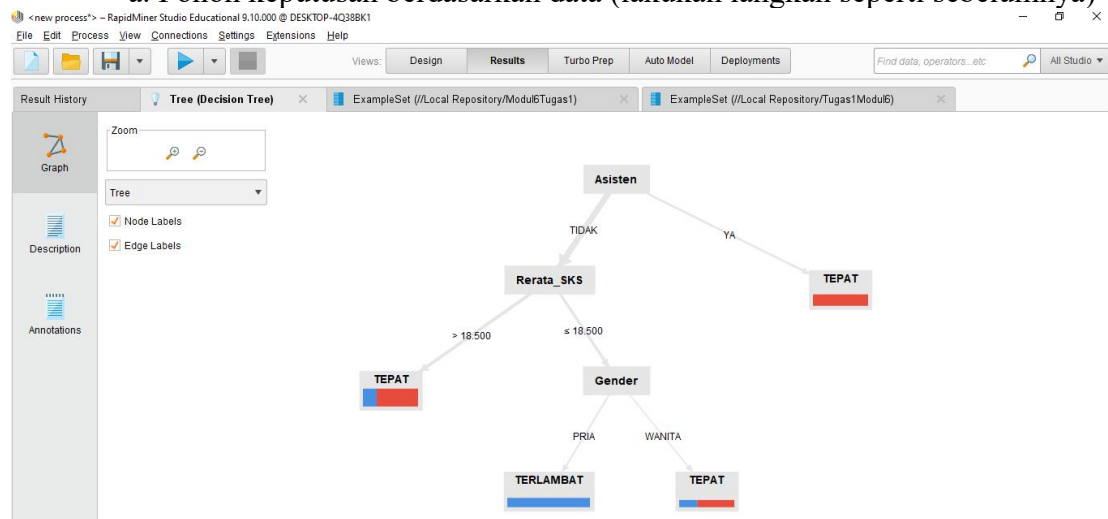


b. Carilah nilai-nilai parameter !

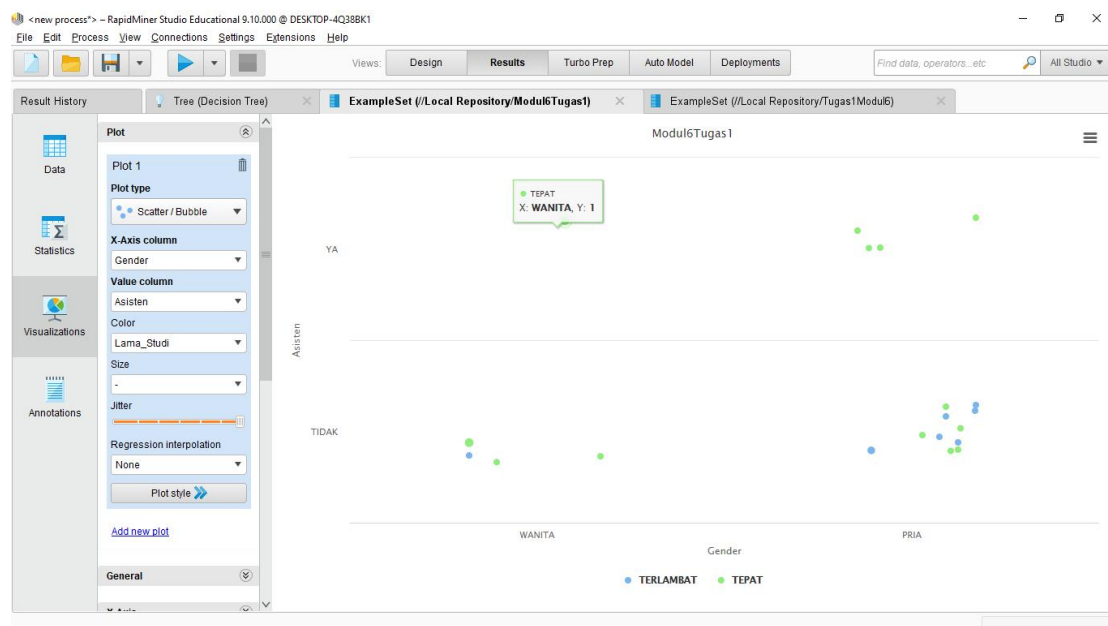
- i. Jumlah simpul daun pada pohon keputusan = 3
- ii. Jumlah simpul keseluruhan pada pohon keputusan = 5
- iii. Waktu yang dibutuhkan untuk proses pelatihan = 0,01 seconds
- iv. Tingkat ketepatan klasifikasi = 60%
- v. Tingkat ketidaktepatan klasifikasi = 40%

3. Gunakan file Excel yang dikerjakan pada tugas nomor 1 dalam modul 6 sebagai data training :

a. Pohon keputusan berdasarkan data (lakukan langkah seperti sebelumnya)



b. Cetaklah Prespektif Plot View dengan model Scatter. X-Axis = Gender, Yaxis = Asisten, dan Color Column = Lama_Studi. Nilai Jitter bisa diubah ubah untuk memperoleh pola penyebaran yang lebih jelas.



4. Berdasarkan pohon keputusan dari soal nomor 2, tentukan klasifikasi yang terbentuk berdasarkan kondisinya sesuai dengan simpul-simpulnya
- a. Seseorang akan memiliki Lama studi (TEPAT) jika kondisi sebagai berikut :
 - i. Gender = Pria, $\text{Rerata_SKS} > 18$ (nilai atribut lain diabaikan).
 - ii. Gender = Wanita (nilai atribut lain diabaikan).
 - b. Seseorang tidak akan memiliki Lama Studi (TERLAMBAT) jika kondisi sebagai berikut :
 - i. Gender = Pria, $\text{Rerata_SKS} \leq 18$, (nilai atribut lain diabaikan).