**LAPORAN PRAKTIKUM DATA WAREHOUSING DAN DATA**

**MINING**

**MODUL 9**

**“KLASIFIKASI: DECISION TREE”**



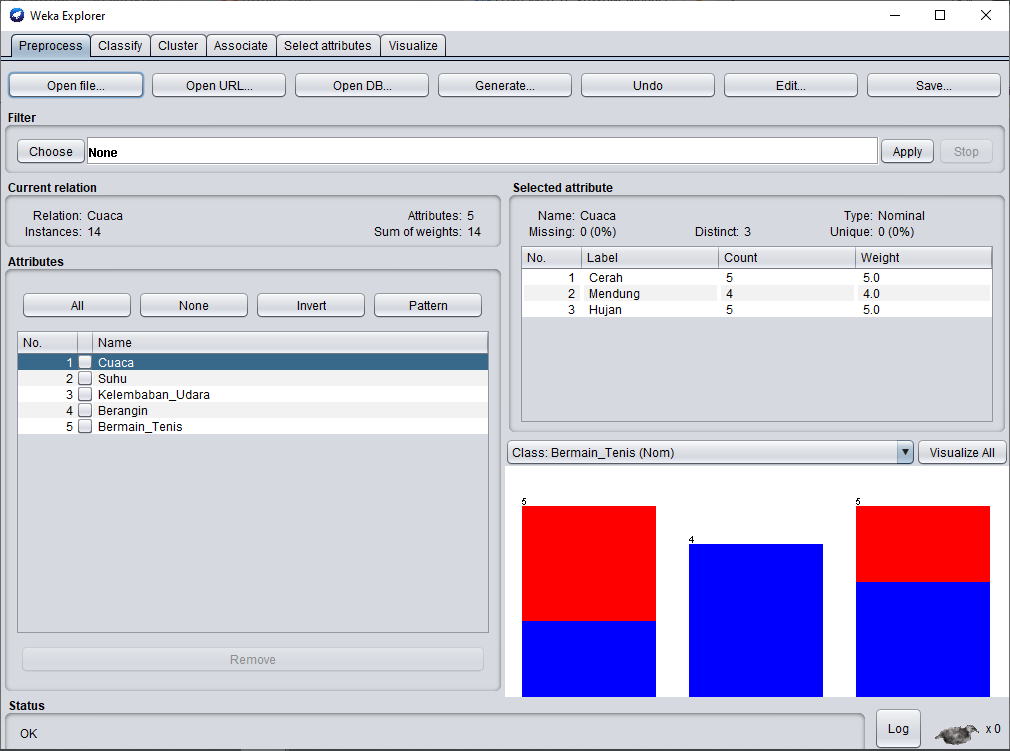
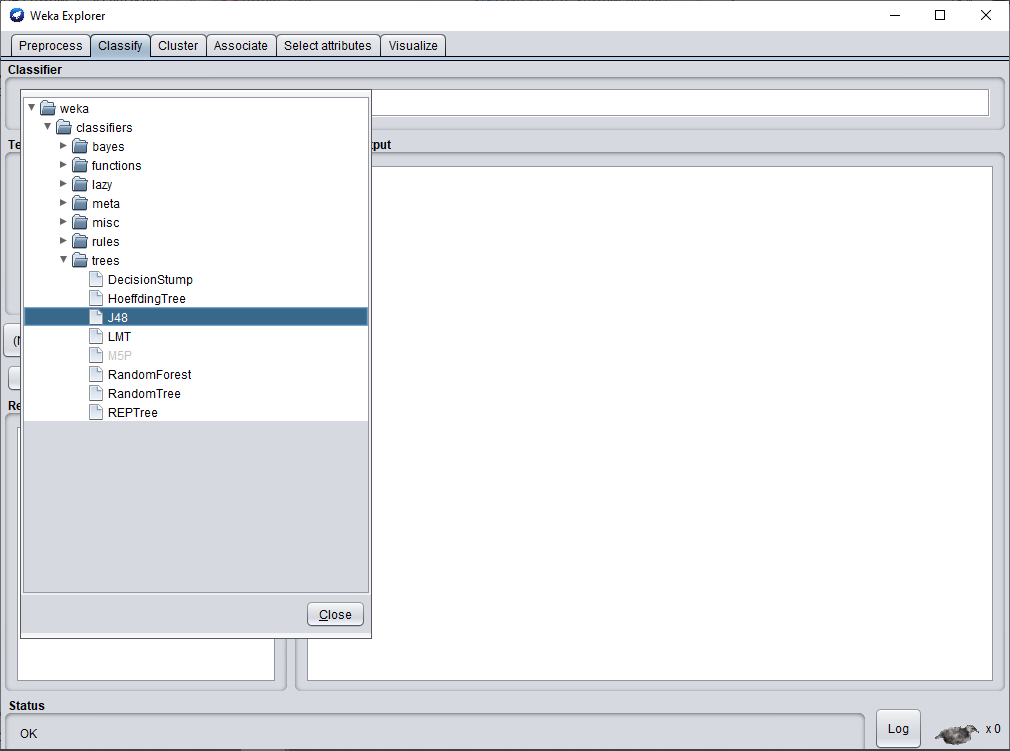
**Oleh:**

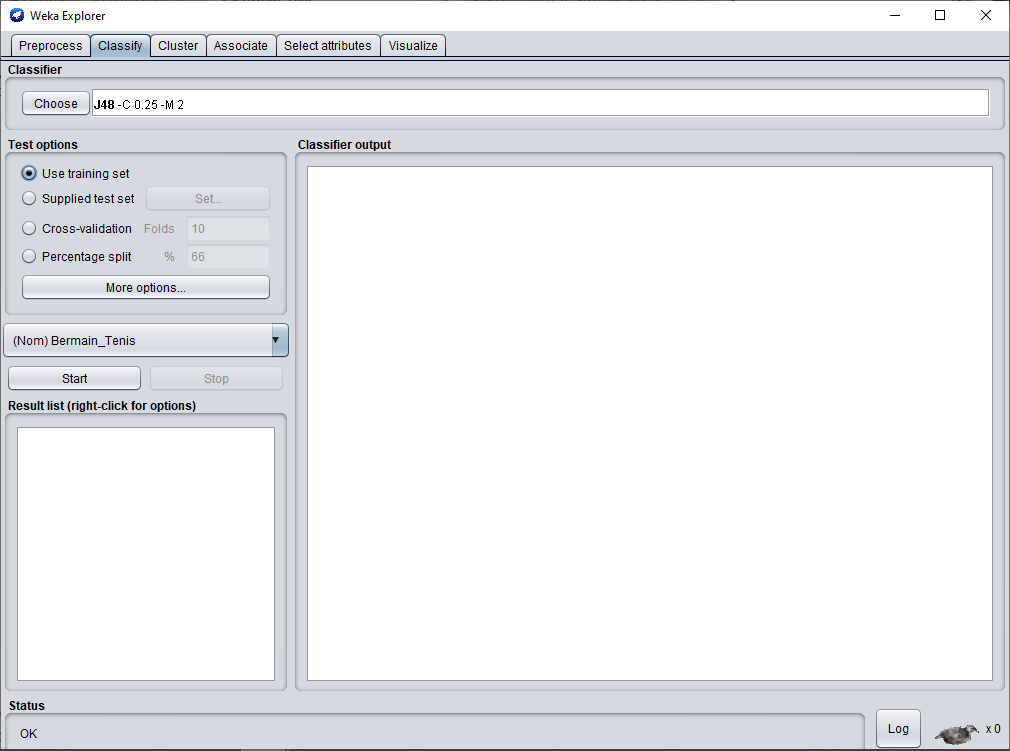
**NAMA : Daffa Putra Alwansyah NIM : L200190031**

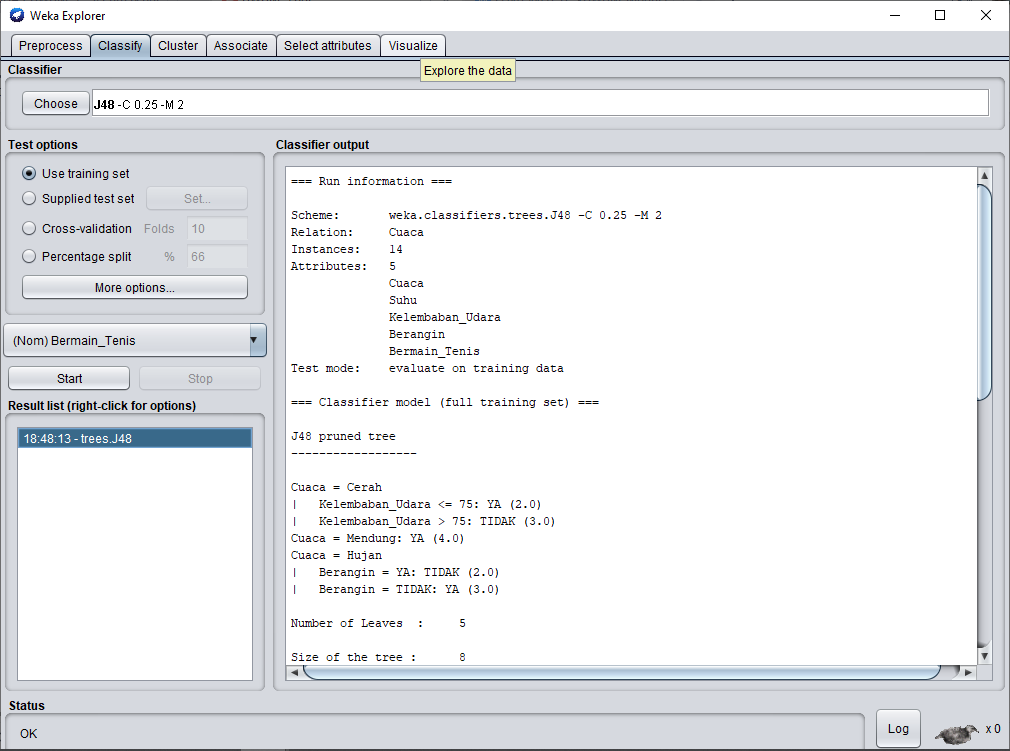
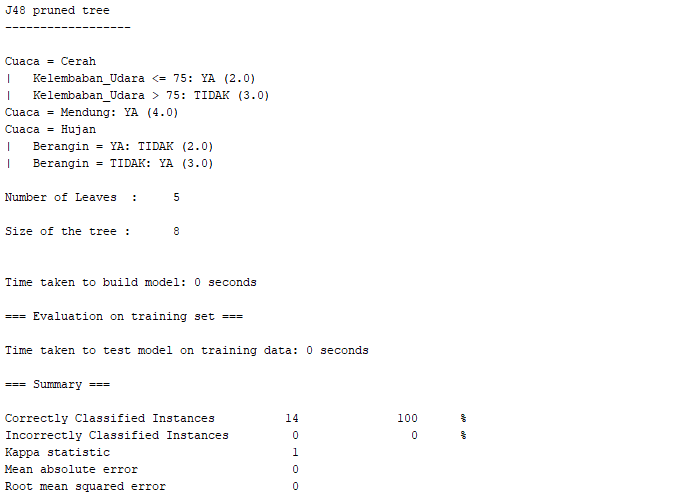
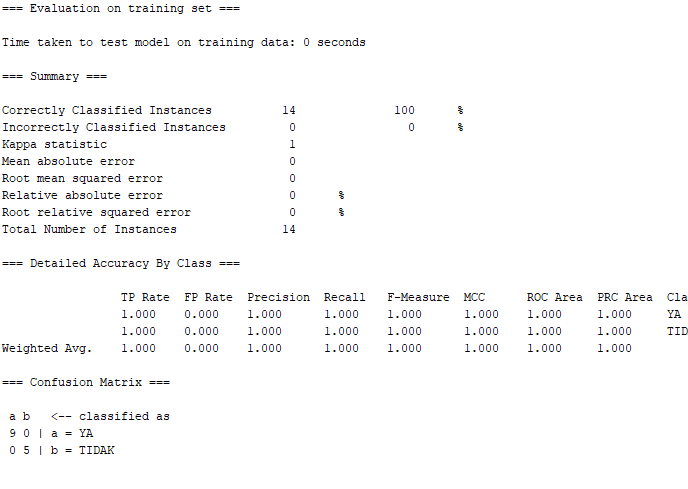
**KELAS : B**

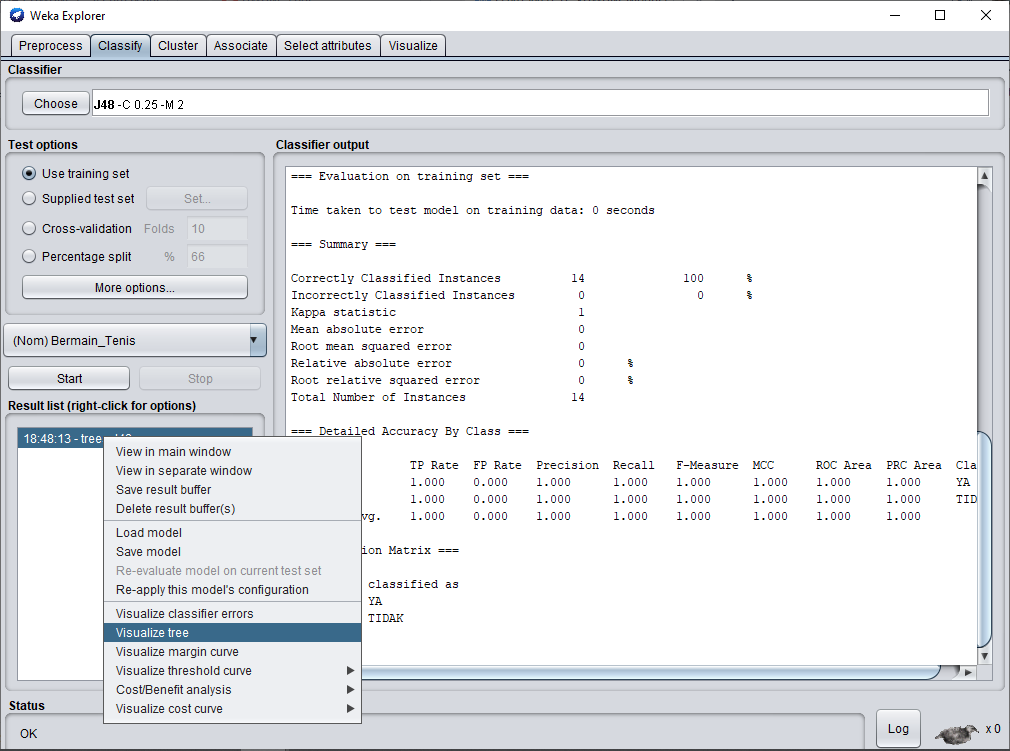
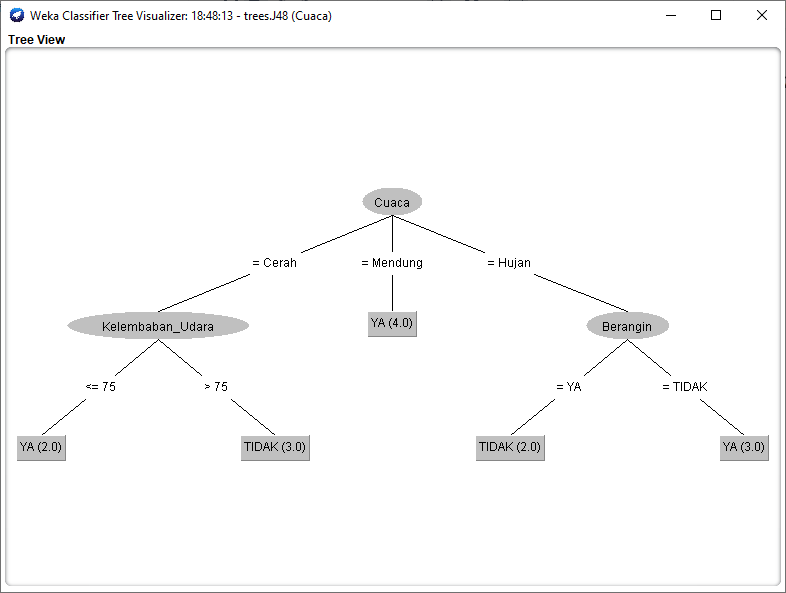
**PRODI : INFORMATIKA**

**Fakultas Komunikasi dan Informatika   
Universitas Muhammadiyah Surakarta**

**Langkah-Langkah Praktikum**1. Buka kembali Cuaca.arff dengan Weka Exploler.  
  
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 ouput.  
  
5. Mengamati kolom Classifier ouput.  
  


6. Untuk melihata 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 terbentuuk yaitu :

a. Seseorang akan bermain tenis (YA) jika kondisi sebagai berikut:

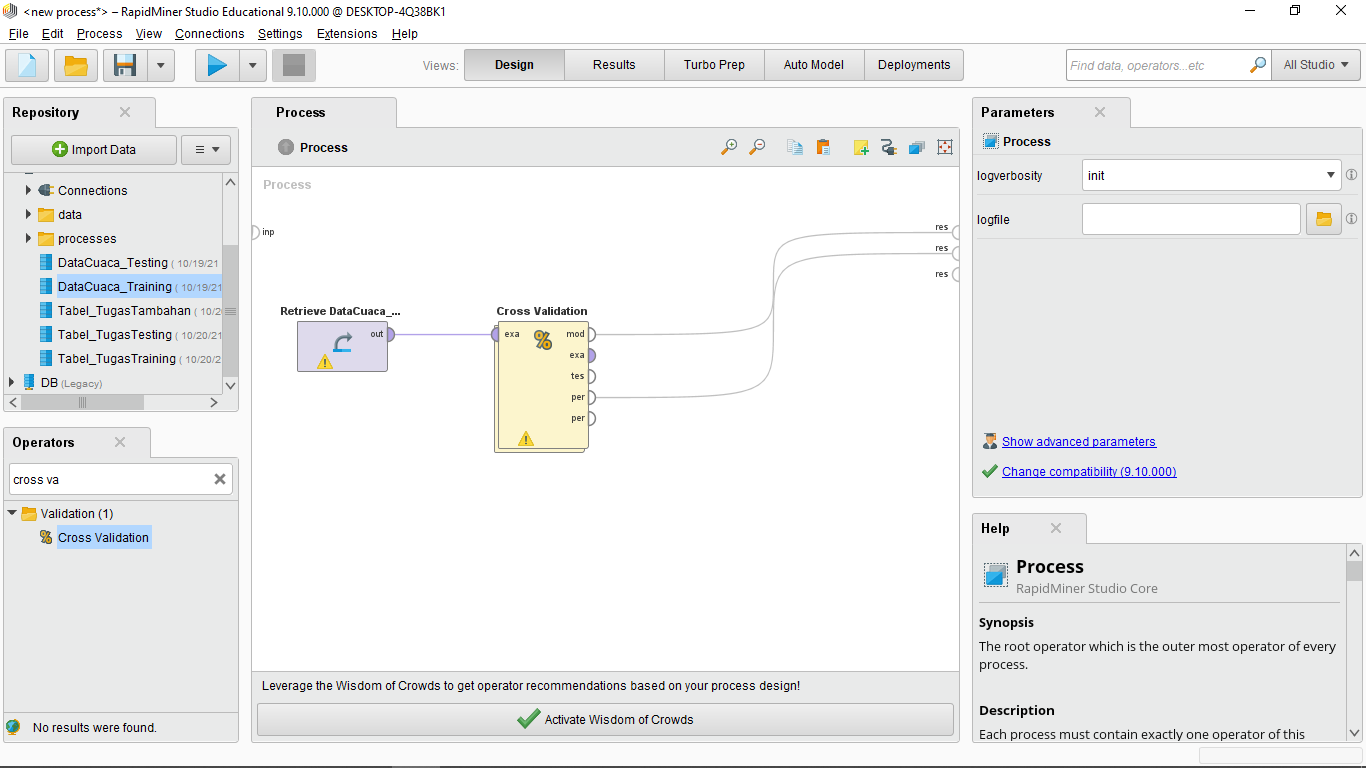
i. Cuaca = Cerah, Kelembaban\_udara > 75, (nilai atribut lain diabaikan)

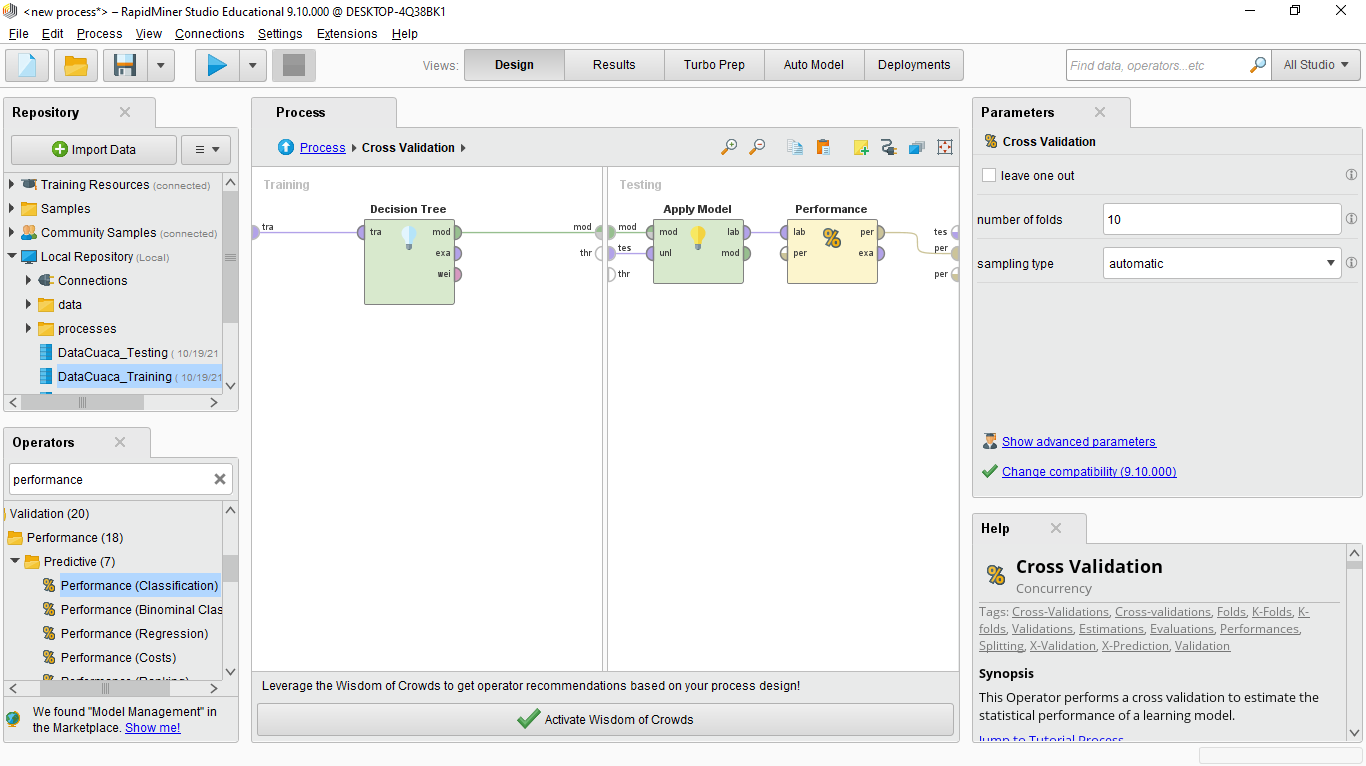
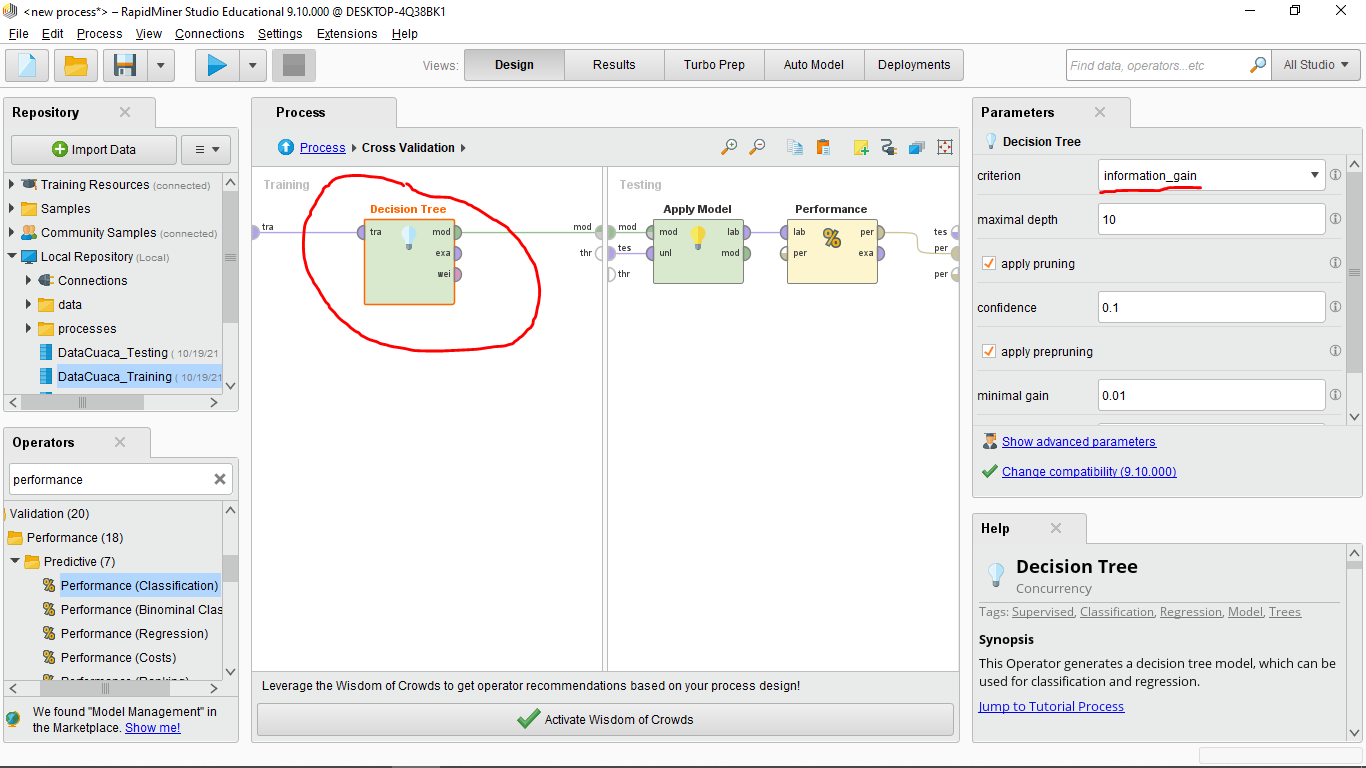
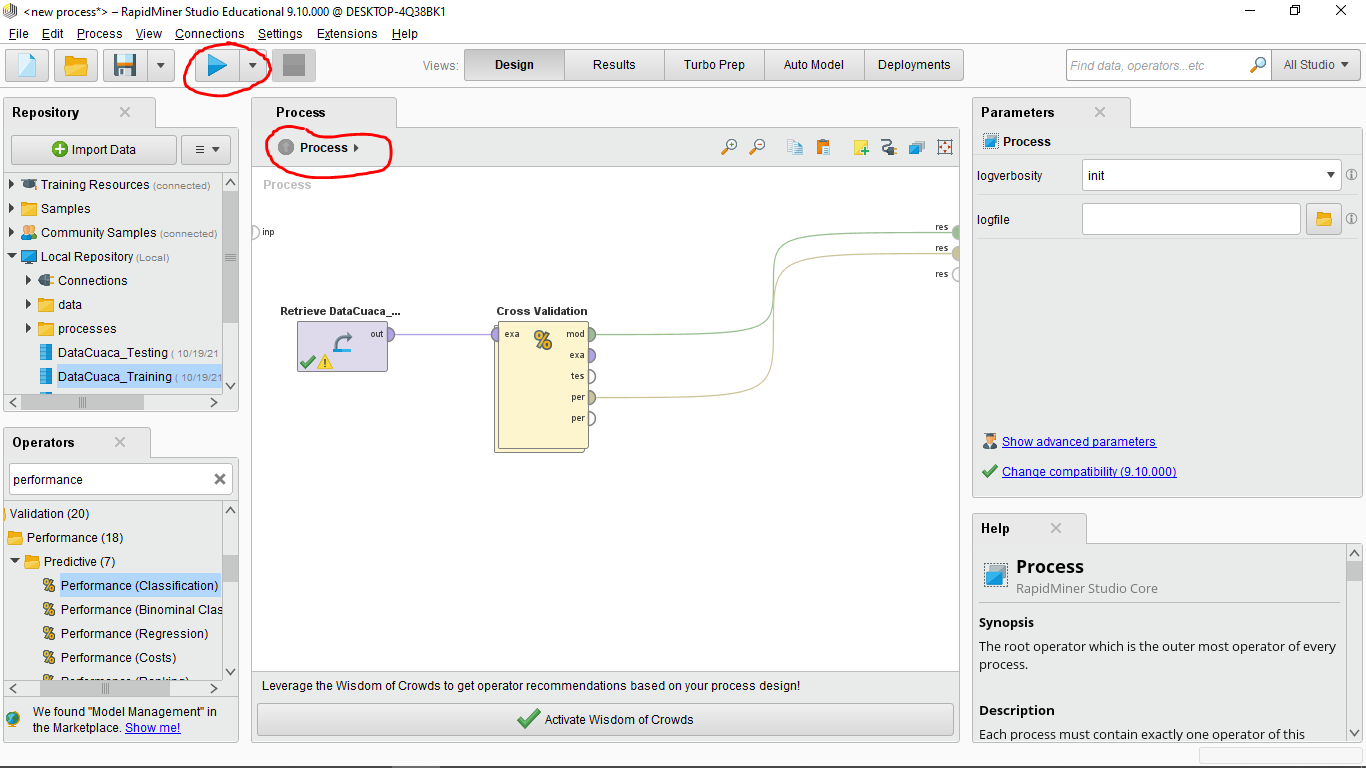
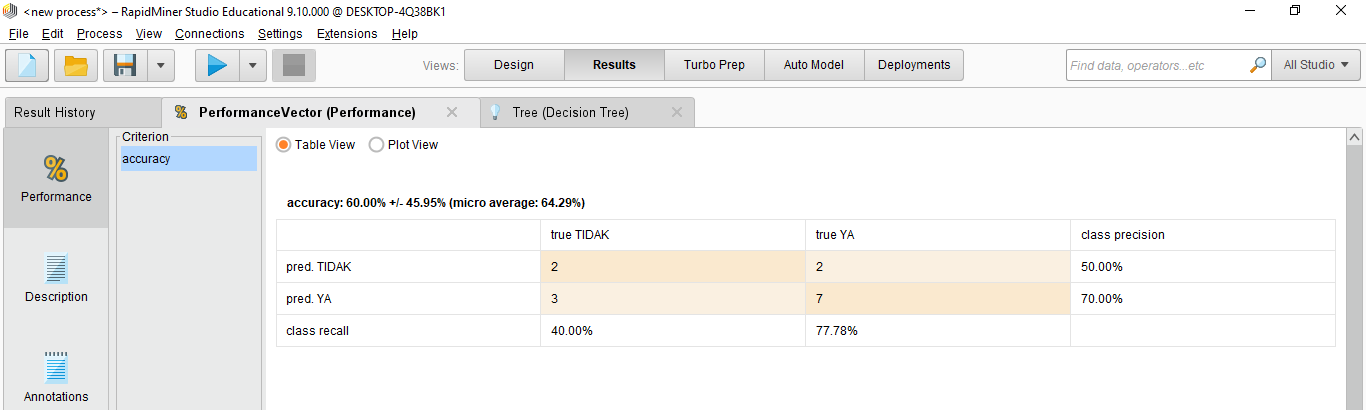
ii. Cuaca = Mendung, (kondisi lain diabaikan)

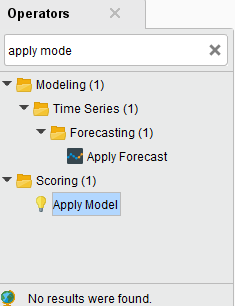
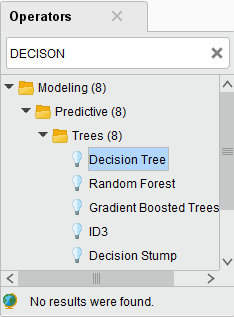
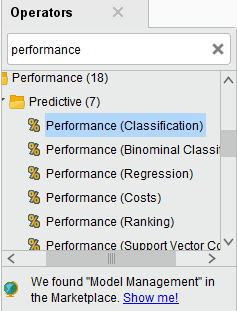
iii. Cuaca = Hujan, Berangin = TIDAK, (nilai atribut lain diabaikan)

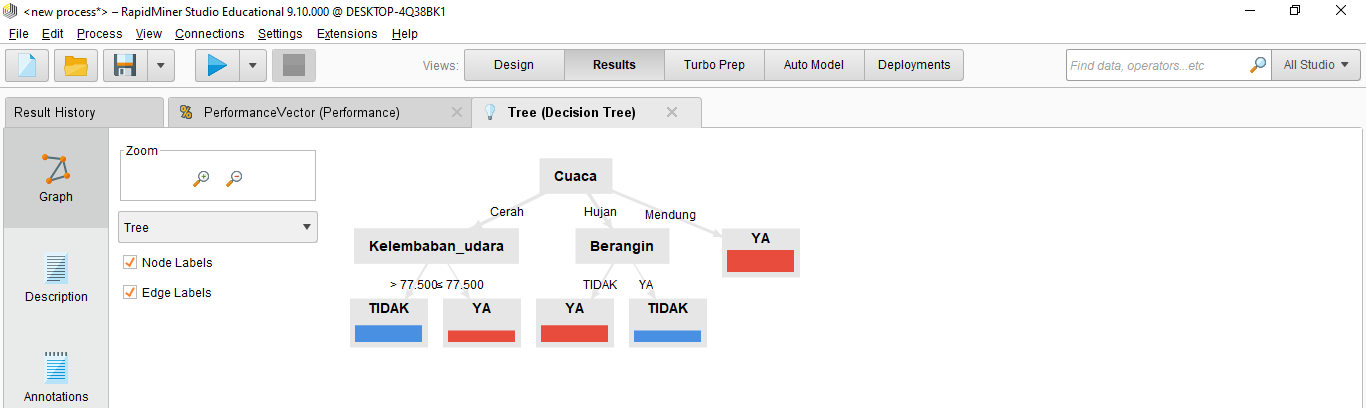
b. Seseorang tidak akan bermain tenis (TIDAK) jika kondisi sebagai berikut:

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 pohoh keputusan:  
 a. PerformanceVector (Permormance), menunjukan tingkat akurasi, presisi, recal dan lain-lain dalam bentuk tabel dan plot view.  




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:

a. Simpul akar = Cuaca

b. Simpul Internal = Kelembaban\_udara, dan Berangin

c. Simpul daun = YA, TIDAK

8. Klasifikasi yang terbentuk yaitu :

a. Seseorang akan bermain tenis (YA) jika kondisi sebagai berikut:

i. Cuaca = Cerah, Kelembaban\_udara <= 77,5 (nilai atribut lain diabaikan).

ii. Cuaca = Mendung. (nilai atribut lain diabaikan).

iii. Cuaca = Hujan, Berangin = TIDAK (nilai atribut lain diabaikan).

b. Seseorang tidak akan bermain tenis (TIDAK) jika kondisi sebagai berikut:

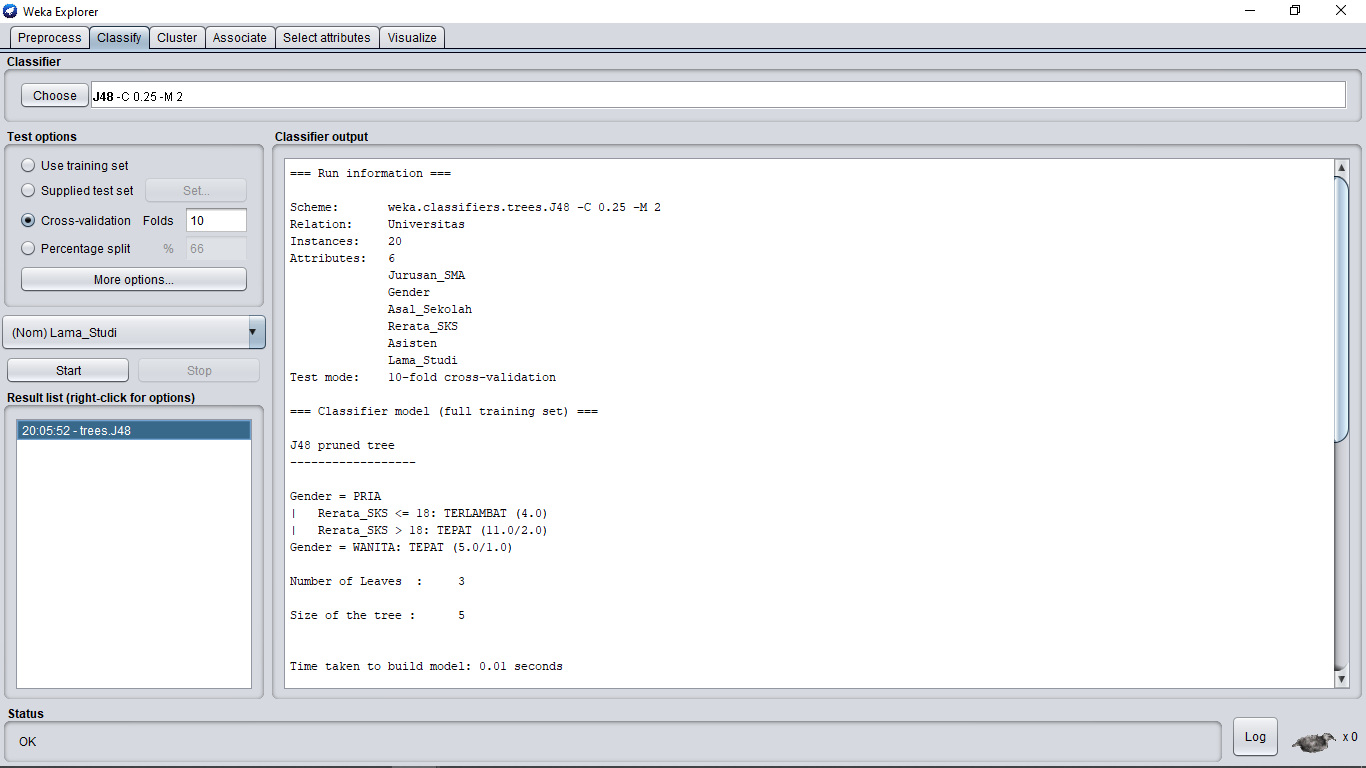
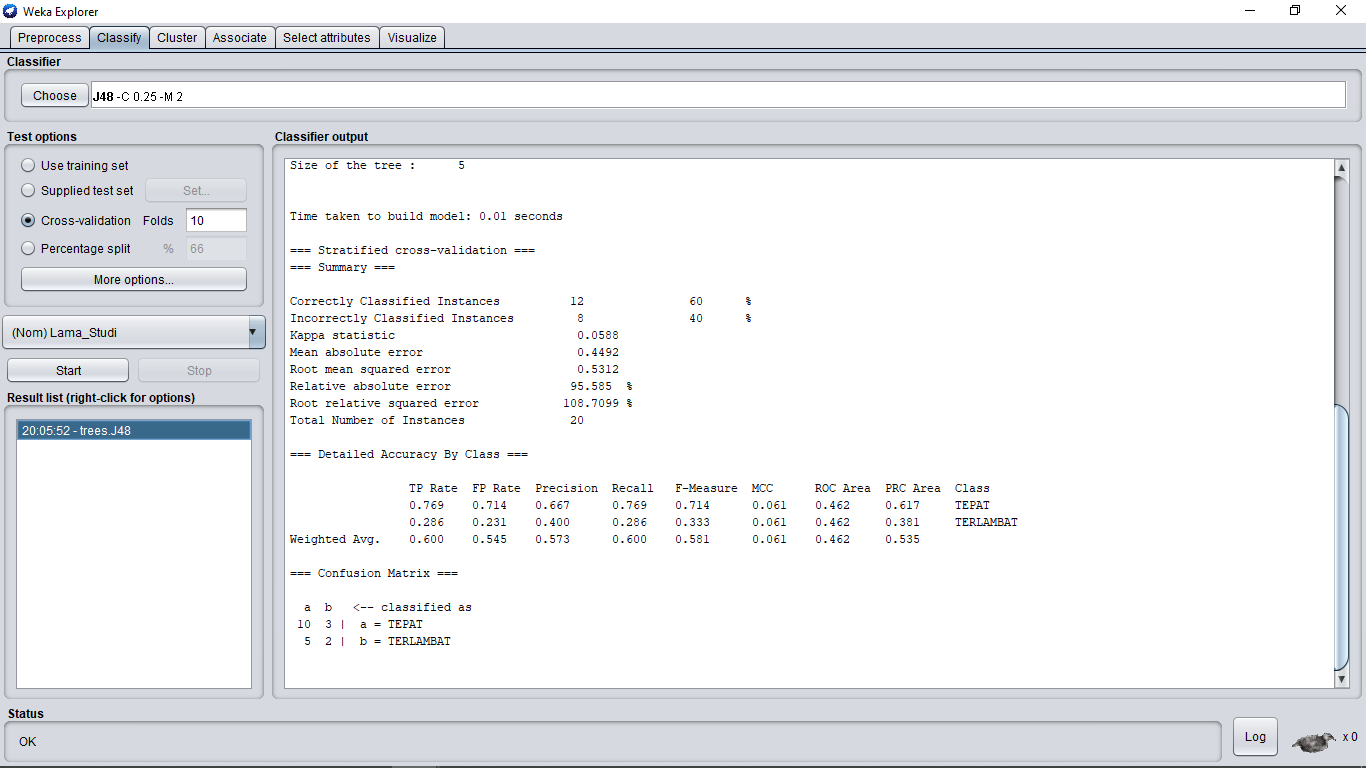
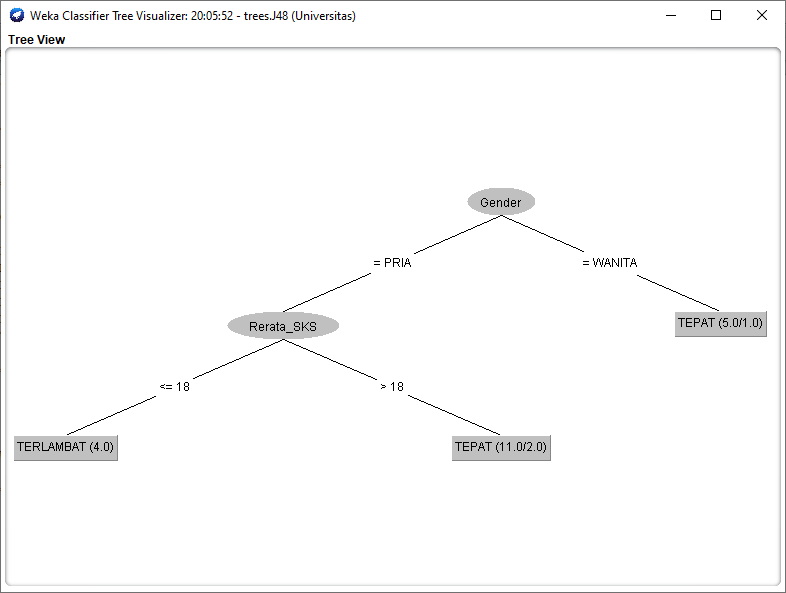
i. Cuaca = Cerah, Kelembaban\_udara > 77,5 (nilai atribut lain diabaikan)

ii. 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 |

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

a. Pohon keputusan berdasarkan data tersebut!  
  
  
  
  
  
  
  
  
  
  
 b. Carilah nilai-nilai parameter !

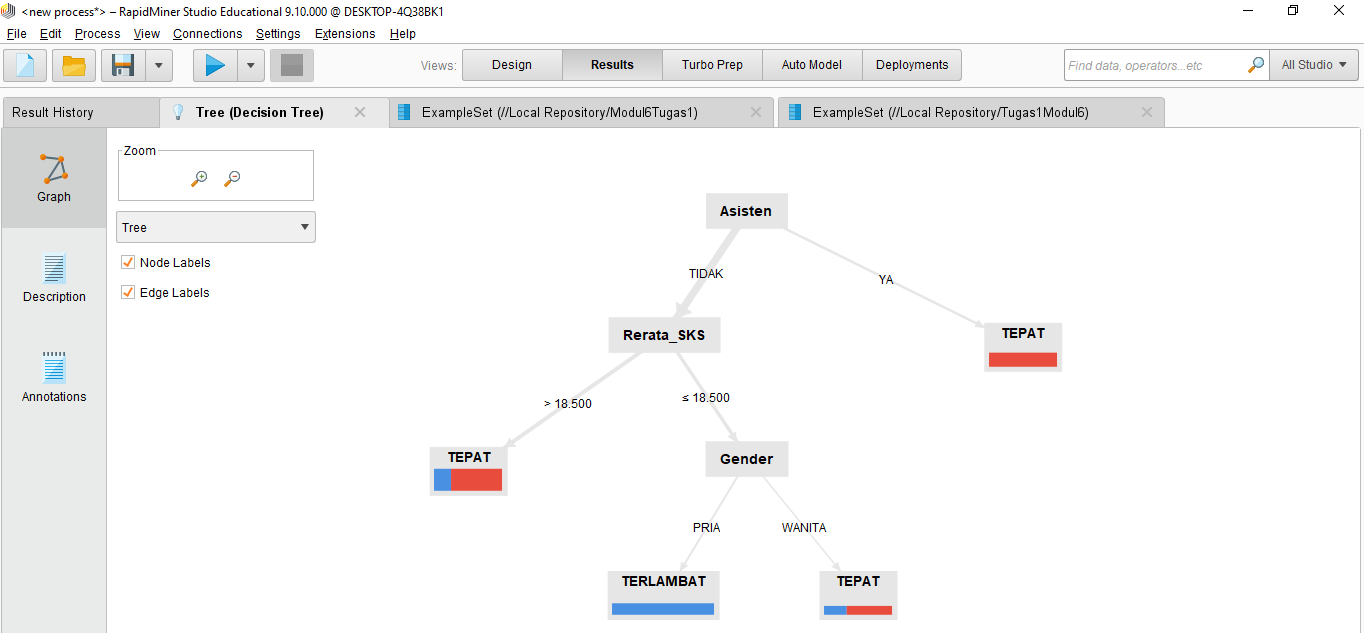
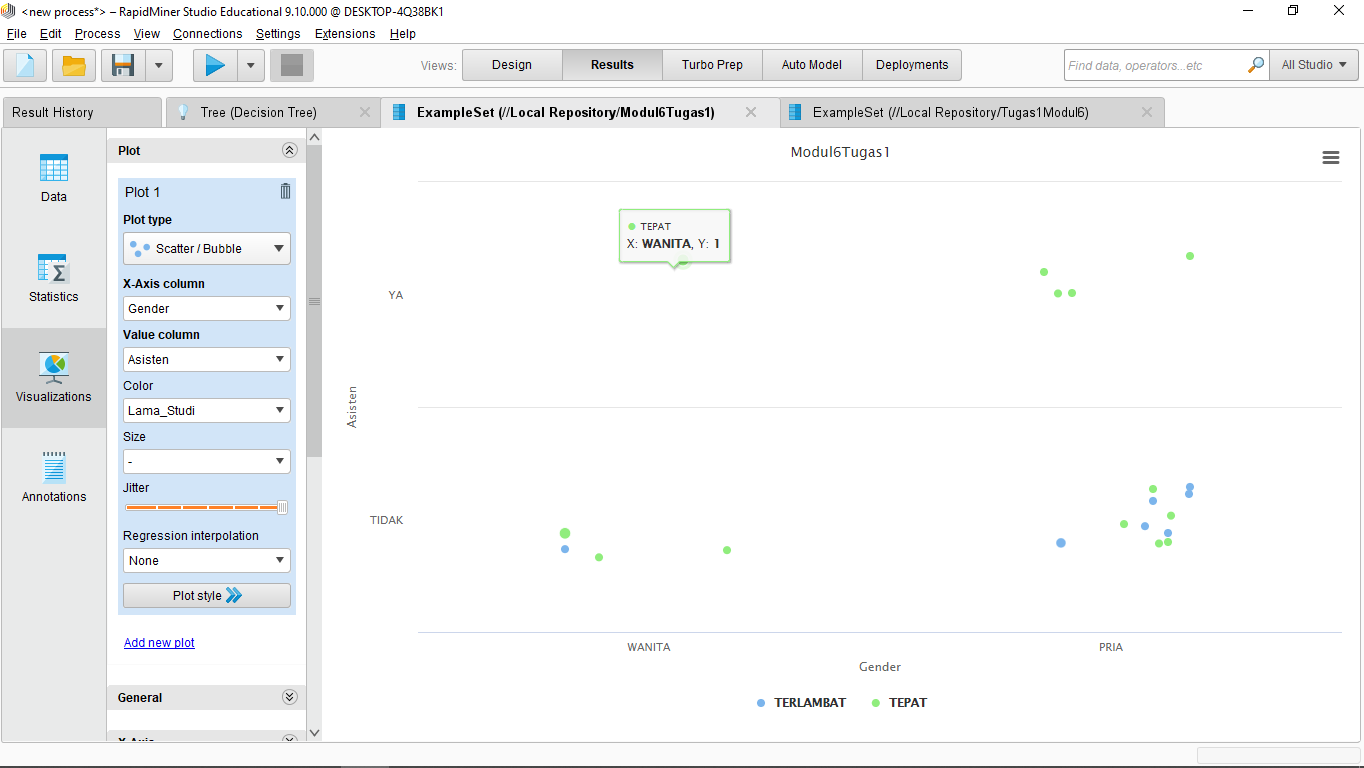
i. Jumlah simpul daun pada pohon keputusan = 3

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 diubahubah 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, 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, Rerata\_SKS <= 18, (nilai atribut lain diabaikan).