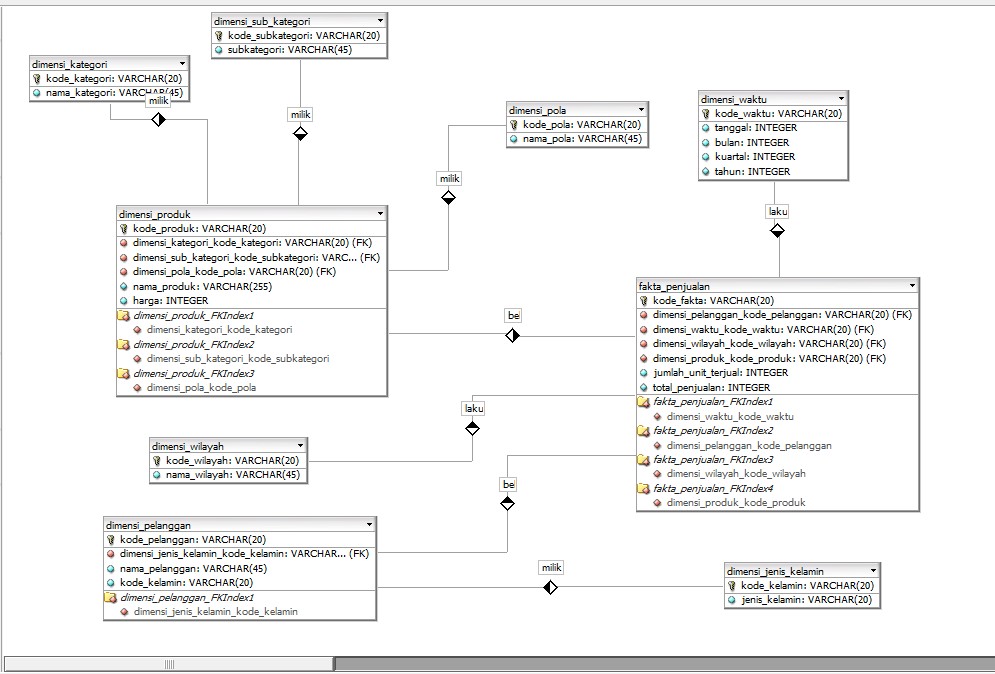
**LAPORAN PRAKTIKUM**

**DATA WAREHOUSE & DATA MINING**

Azzahra Salsabilla; L200170130; E

**MODUL 1**

**(Perancangan *Star Schema* dan *Snowflake*)**

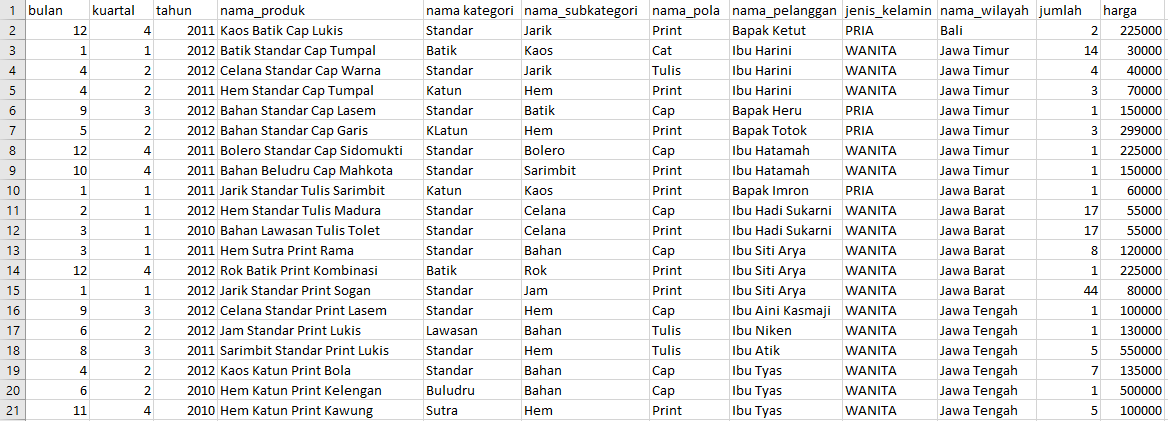


**MODUL 5**

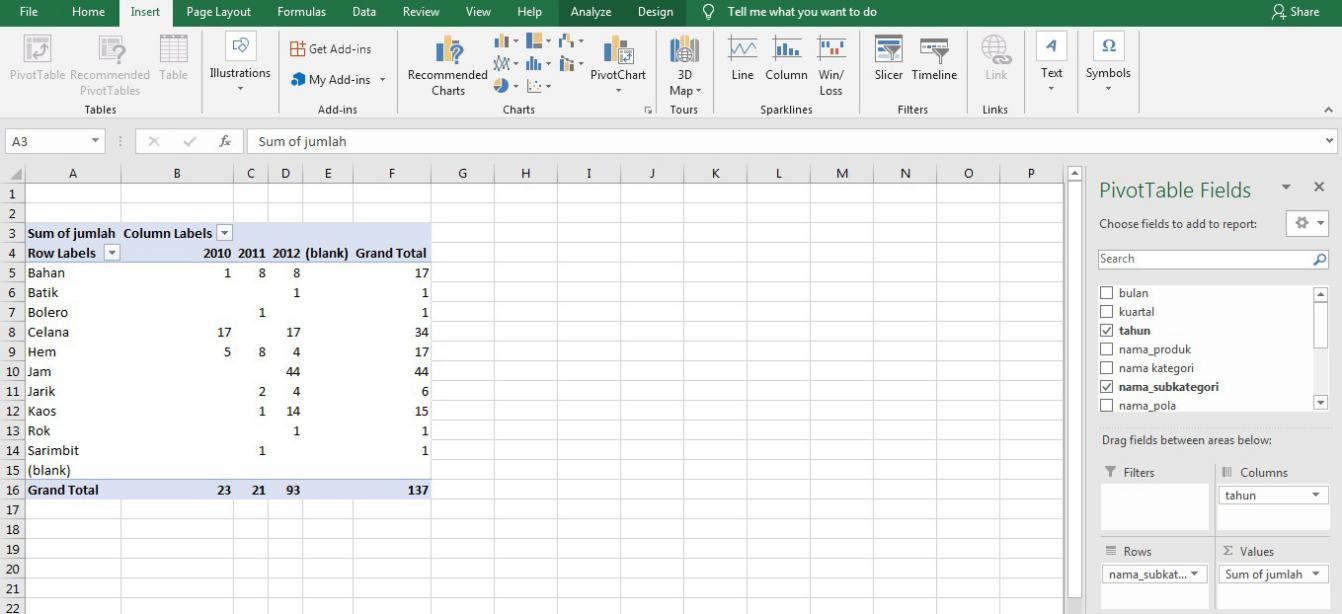
**(Pivot *Table* dan *Chart*)**

**KEGIATAN**

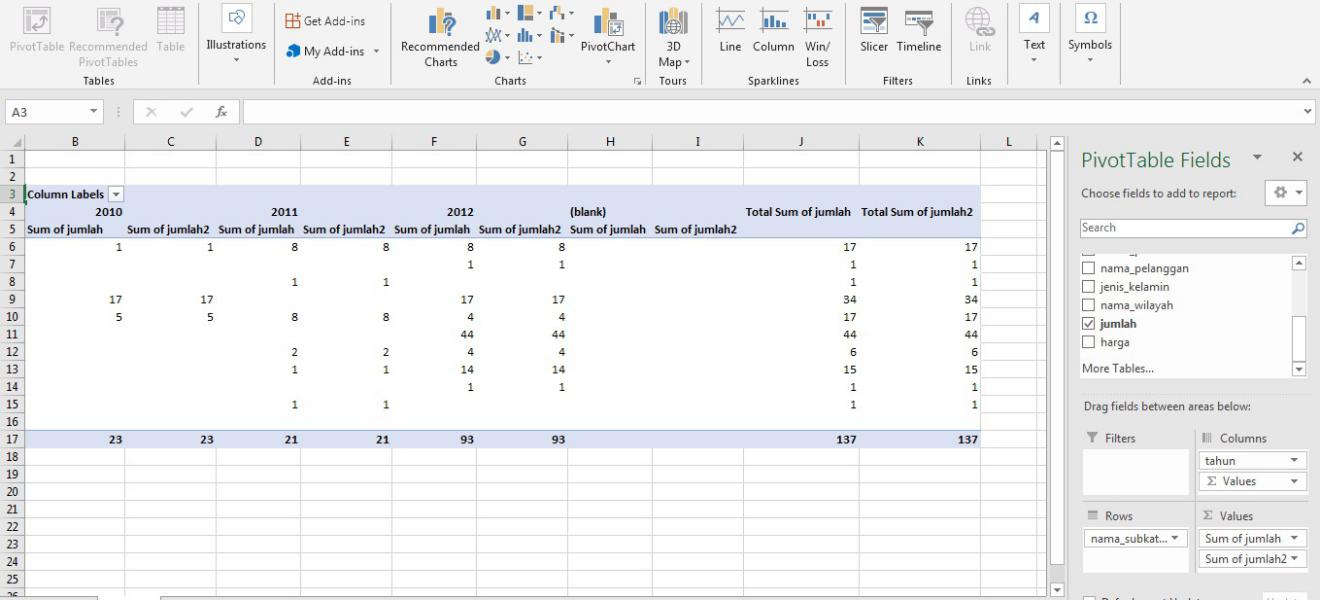
Membuat pivot table kemudian menyimpannya dengan nama ‘Fakta\_Penjualan.xls’.



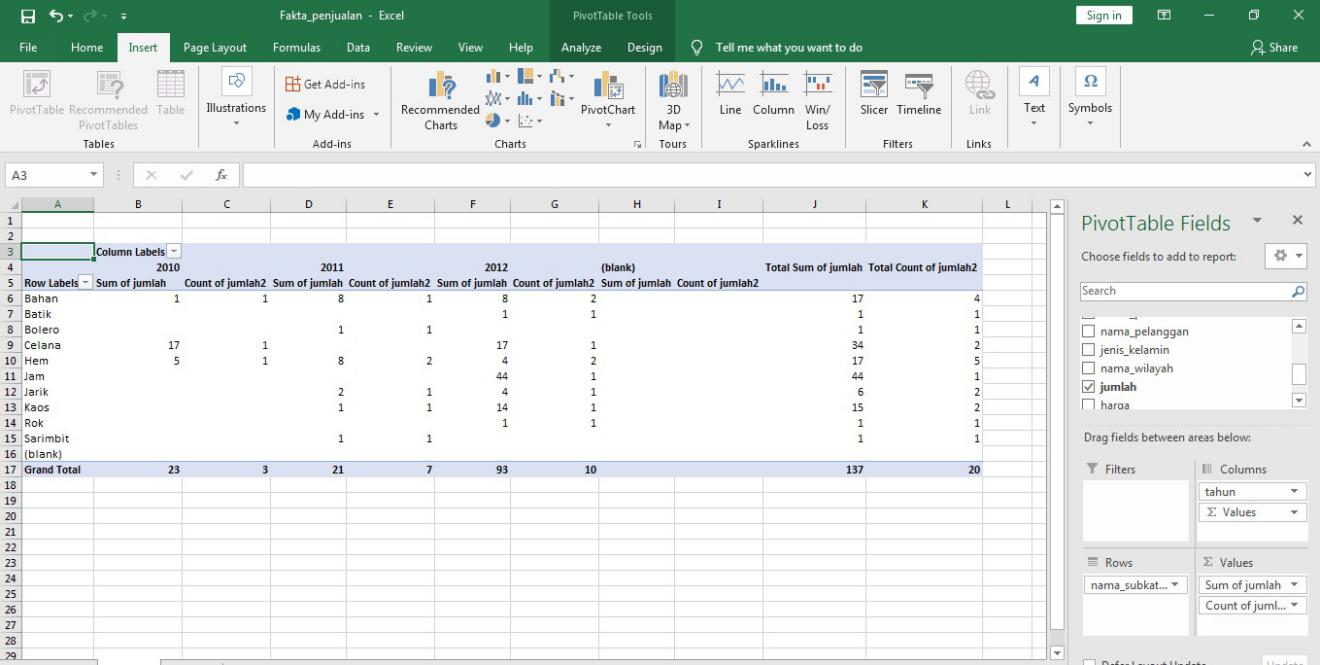
Hasil pivot table yang menampilkan nama\_subkategori, tahun dan jumlah.



Menambahkan tipe summary baru, tambahkan field jumlah kembali ke kotak Value dengan cara klik dan drag, sehingga muncul field baru dengan nama Sum of jumlah2.



Kemudian mengubah nama field sum of jumlah2 menjadi count of jumlah 2 dengan klik sum of jumlah 2 pada PivotTableField dan pilih dialog Value Field Settings.



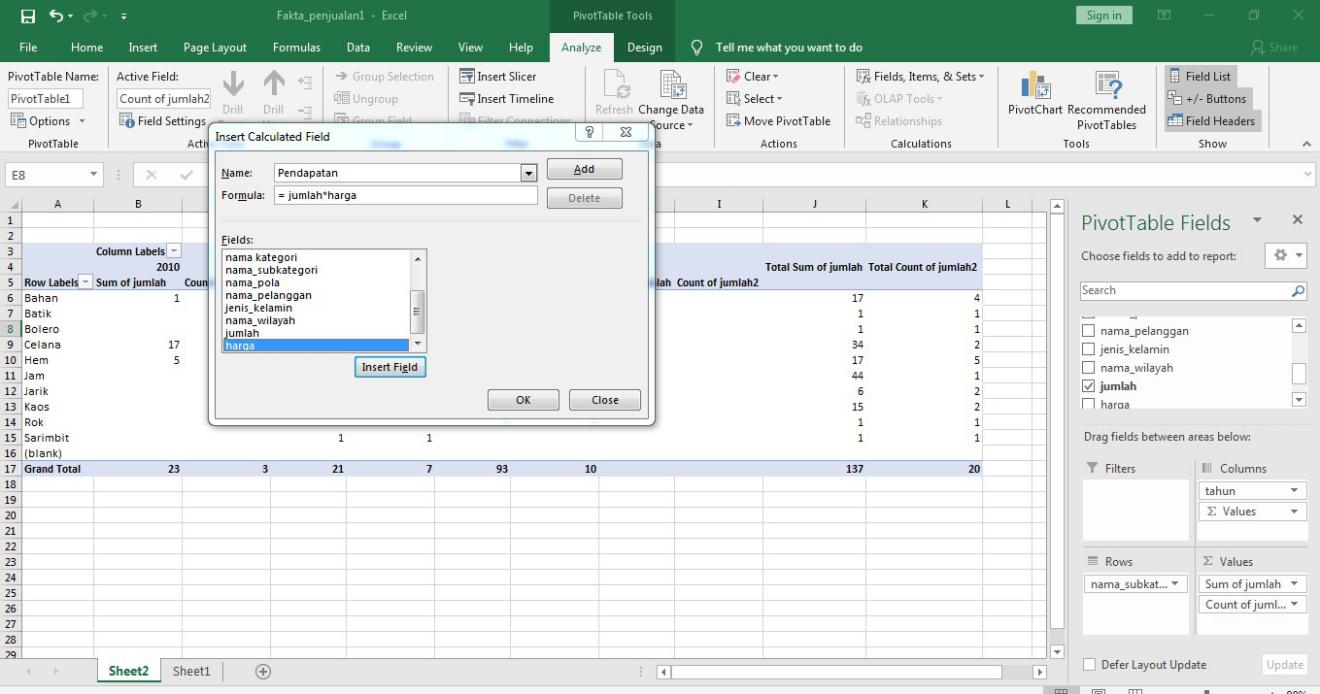
Calculated Field

Caranya klik analyze → pilih fields, items&sets→ calculated field

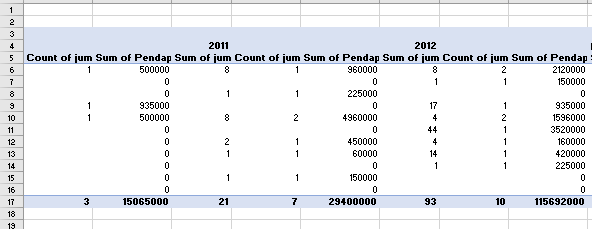
Pada kotak dialog Insert Calculated Field yang muncul, masukkan nilai berikut kemudian klik tombol OK.

Name : Pendapatan

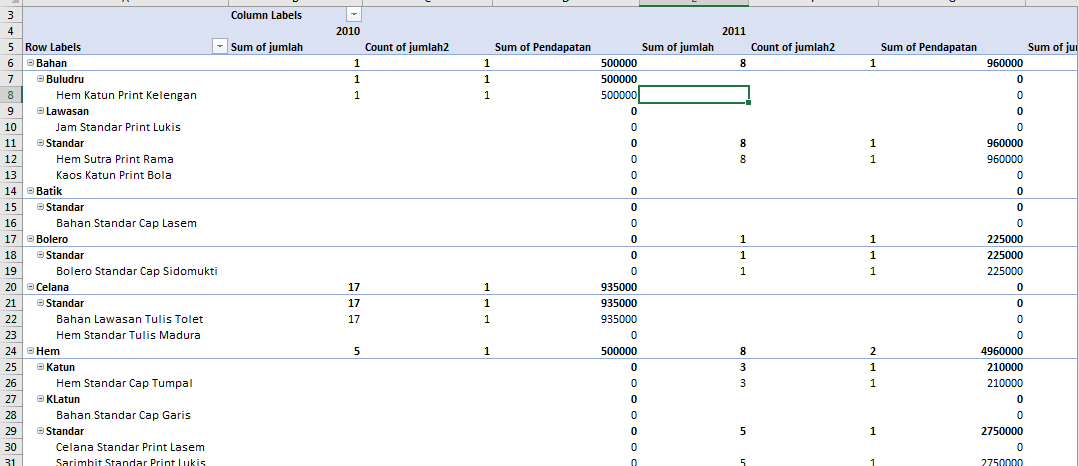
Formula : = jumlah \* harga (Pilih field jumlah kemudian klik Insert Field kemudian ketikkan tanda “\*” dan masukkan field harga)



Field baru, "Sum of Pendapatan" akan muncul pada Pivot Table



Operasi Roll Up dan Drill Down



**MODUL 5**

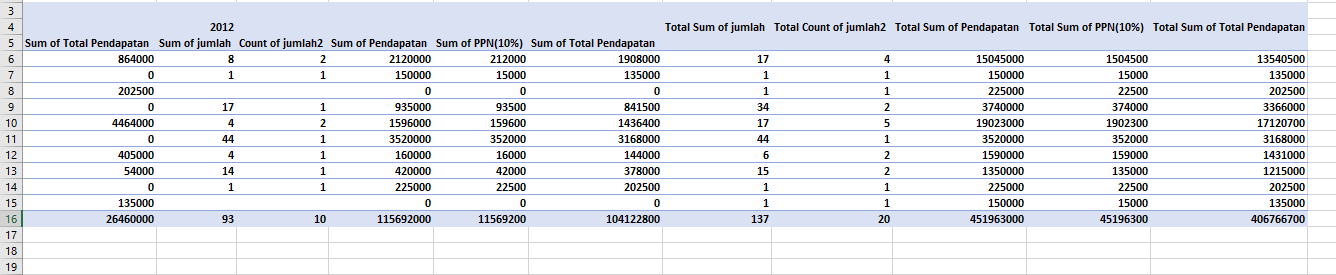
**(Pivot *Table* dan *Chart*)**

**TUGAS**

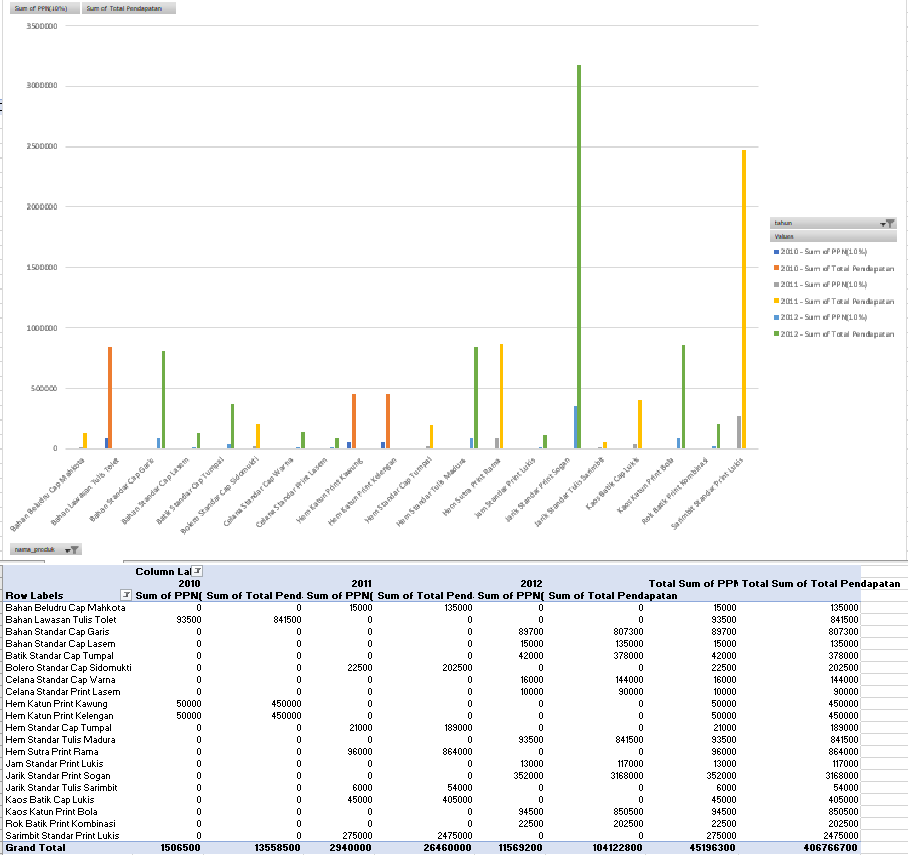
Dengan menggunakan PivotTable pada file Fakta\_Penjualan.xls tambahkan 2 buah field, yaitu :

- PPN (Pajak Pertambahan Nilai) sebesar 10% dari tiap pendapatan pada Pivot Table.

- Total Penghasilan yang dihitung dari pendapatan dikurangi dengan PPN tersebut.



Buatlah PivotTable dan PivotChart untuk melihat PPN dan Total Penghasilan tersebut selama tahun 2010 – 2012. Kategori produk apakah yang memberikan nilai penghasilan terbanyak selama 3 tahun tersebut?



**MODUL 6**

**(Pengenalan *Data* *Mining*)**



Menghitung jumlah data yang terdapat dalam kolom Jurusan\_SMA

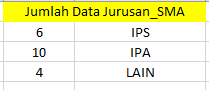
IPS : =COUNTIF(A1:A21,"IPS")

IPA : =COUNTIF(A1:A21,"IPA")

LAIN : =COUNTIF(A1:A21,"LAIN")

Fungsi COUNTIF : menghitung banyaknya variable IPA/IPS/LAIN pada kolom Jurusan\_SMA

Hasil :



Menghitung jumlah data yang terdapat dalam kolom Lama\_Studi

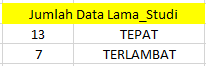
TEPAT : =COUNTIF(F1:F21,"TEPAT")

TERLAMBAT : =COUNTIF(F1:F21,"TERLAMBAT")

Fungsi COUNTIF : menghitung banyaknya variable TEPAT/TERLAMBAT pada

kolom Lama\_Studi

Hasil :



Menghitung nilai MAX, MIN, MEAN, Standard Deviation yang terdapat dalam kolom Rerata\_SKS

MAX : =MAX(D2:D21)

Fungsi MAX : menghitung nilai maksimum pada kolom Rerata\_SKS

MIN : =MIN(D2:D21)

Fungsi MIN : menghitung nilai minimum pada kolom Rerata\_SKS

MEAN : =AVERAGE(D2:D21)

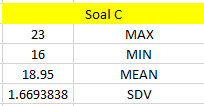
Fungsi MEAN : menghitung nilai rata-rata pada kolom Rerata\_SKS

Standard Deviation : =STDEV(D2:D21)

Fungsi STDEV : mengukur jumlah variasi atau sebaran sejumlah nilai data pada

kolom Rerata\_SKS

Hasil :



Menghitung julmah data gabungan dari kelas atribut Jurusan\_SMA = IPA, Gender = PRIA, Asisten = YA, Lama\_Studi = TEPAT.

Formula : =COUNTIFS(A2:A21,"IPA",B2:B21,"PRIA",E2:E21,"YA",

F2:F21,"TEPAT")

Fungsi COUNTIFS : menghitung jumlah data atau sel yang memenuhi banyak kriteria

Hasil :



**MODUL 7**

**(*Data Preprocessing*)**

**KEGIATAN**

Ketentuan variable/atribut :

Variabel Y = Bermain\_Tenis

Variable X1 = Cuaca

Variable X2 = Suhu

Variable X3 = Kelembaban\_Udara

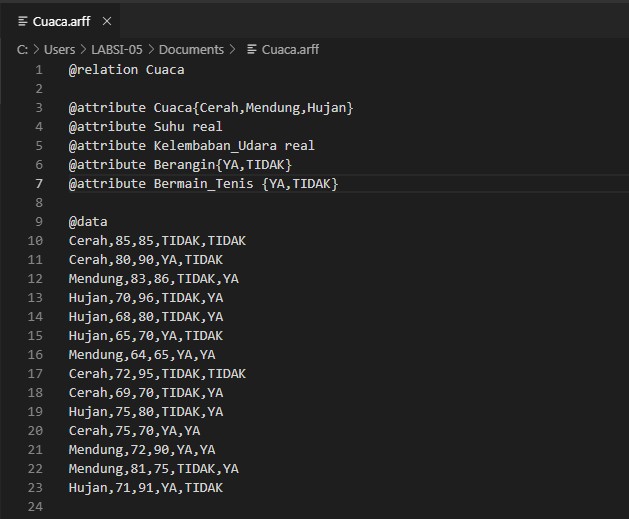
Variable X4 = Berangin

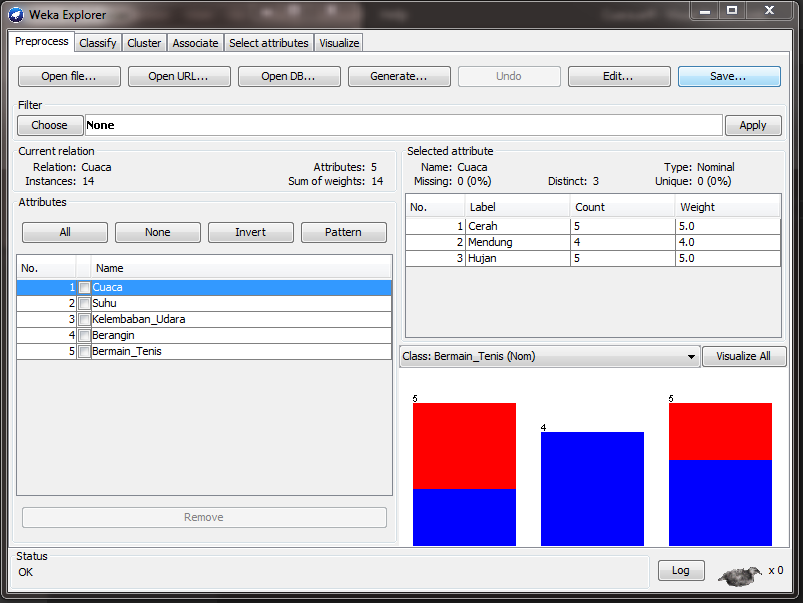
Penentuan Data :

Binomial : jika nilai data hanya ada 2 kemungkinan

Polynomial : jika nilai data lebih dari 2 kemungkinan, namum tidak terlalu banyak

Real : jika kemungkinan nilai data sangat beragam dan jumlahnya banyak

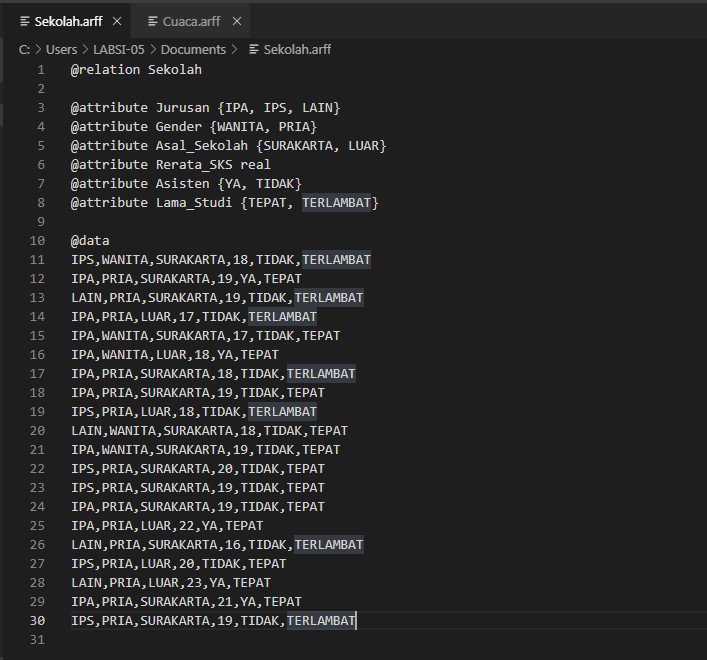


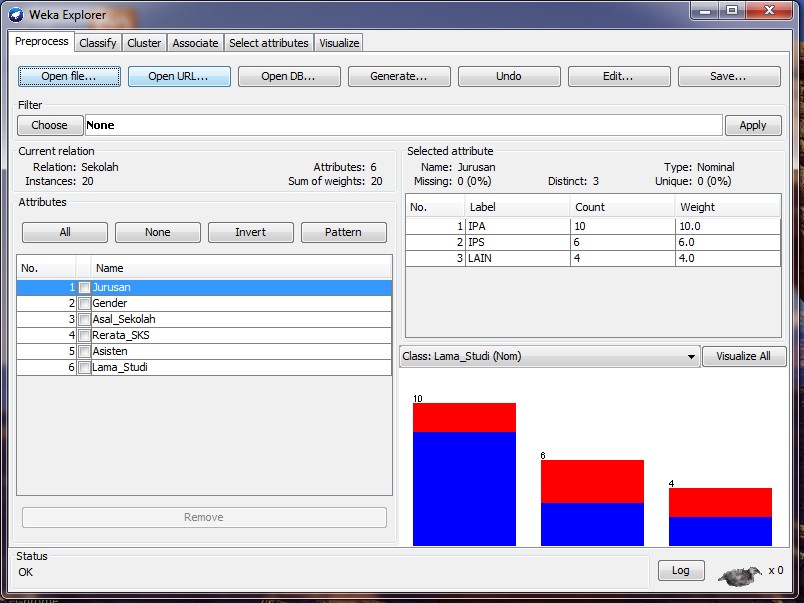


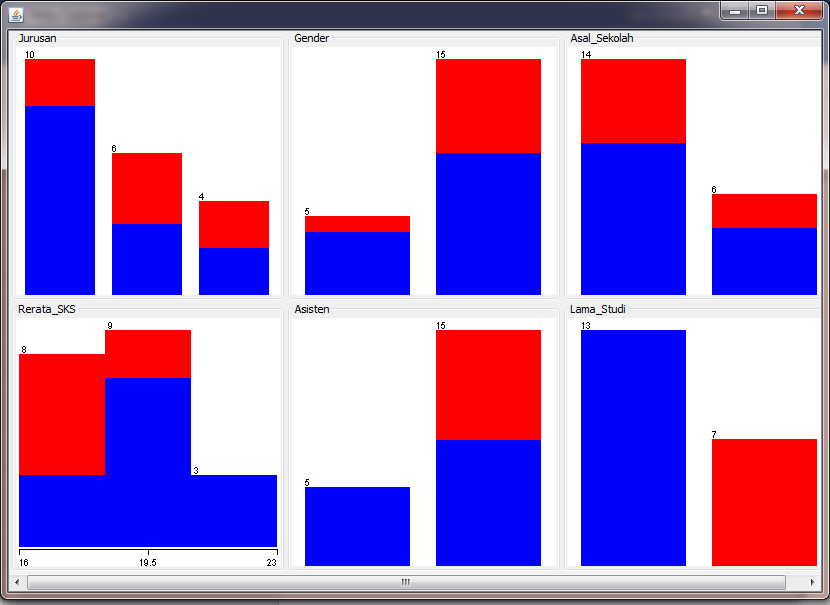
**MODUL 7**

**(*Data Preprocessing*)**

**TUGAS**







Jumlah atribut bertipe binomial yaitu 4

- Gender {WANITA, PRIA}

- Asal\_Sekolah {SURAKARTA, LUAR}

- Asisten {YA, TIDAK}

- Lama\_Studi {TEPAT, TERLAMBAT}

Jumlah atribut bertipe polynomial yaitu 1

- Jurusan {IPA, IPS, LAIN}

Jumlah atribut bertipe real yaitu 1

- Rerata\_SKS real

Pada atribut Rerata\_SKS, besarnya nilai :

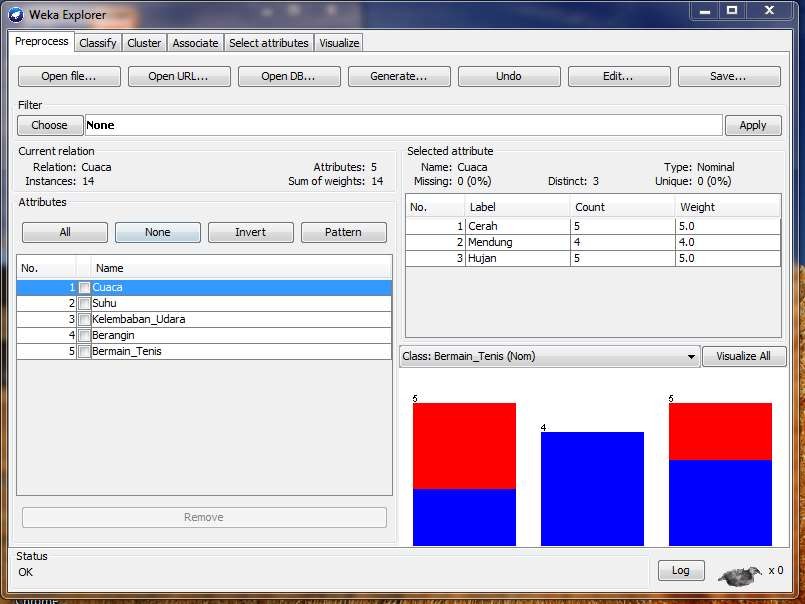
1. Maximum : 23
2. Minimum : 16
3. Mean : 18.95
4. StdDev : 1.6693838

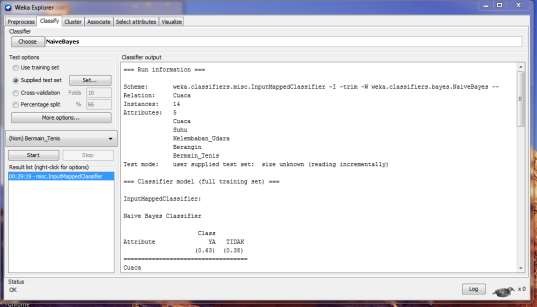
**MODUL 8**

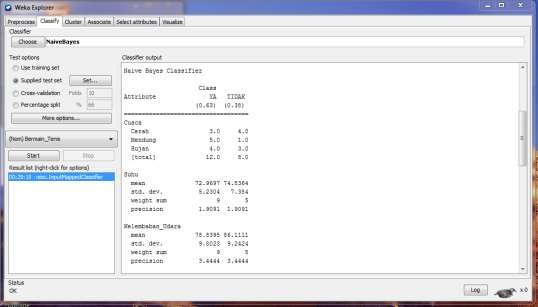
**(Klasifikasi : Naïve Bayes)**

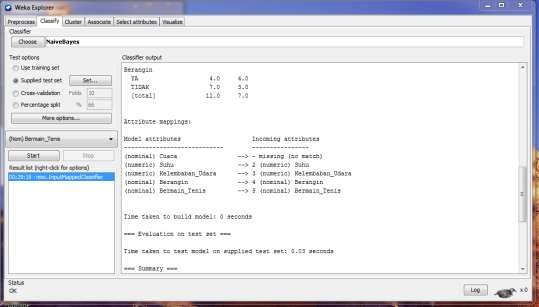
**KEGIATAN**

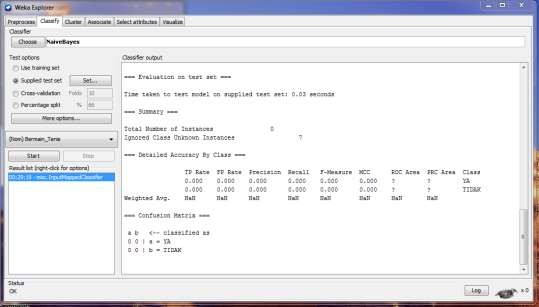
Implementasi Naïve Bayes dengan Weka

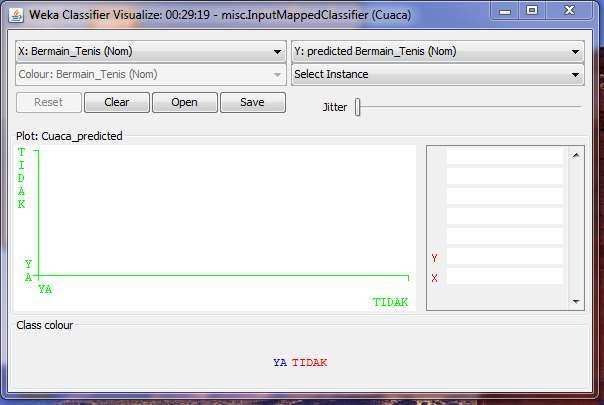


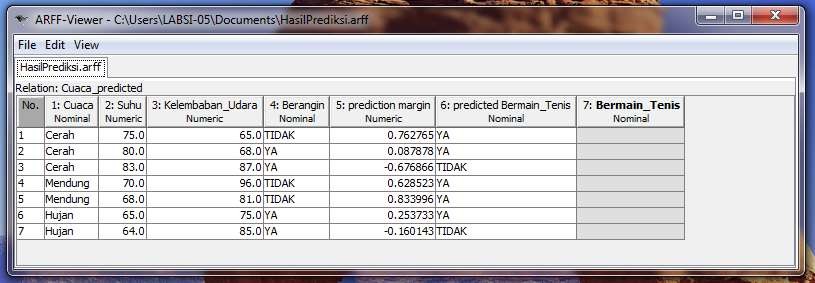


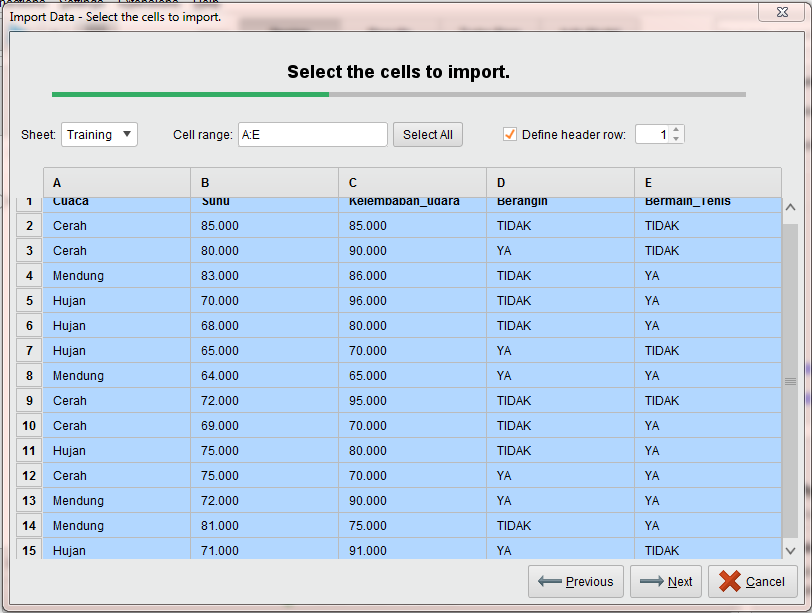


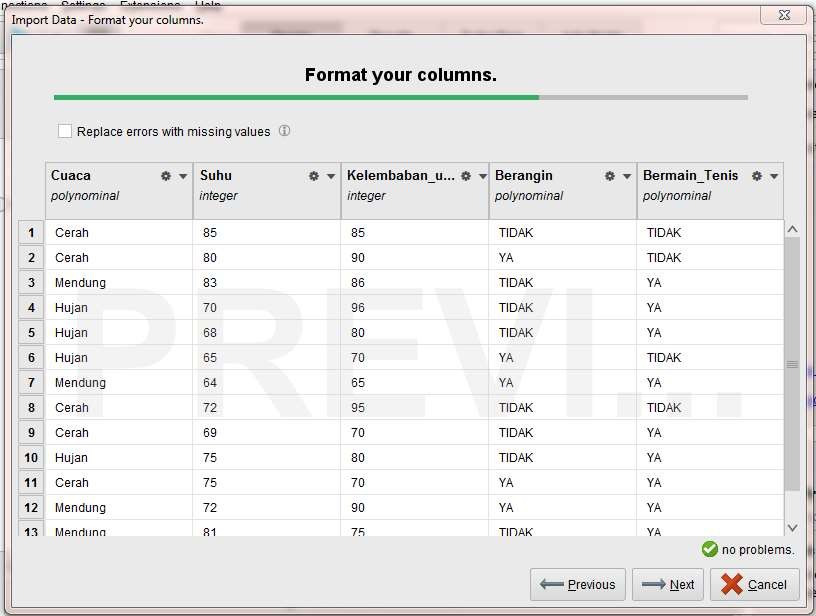


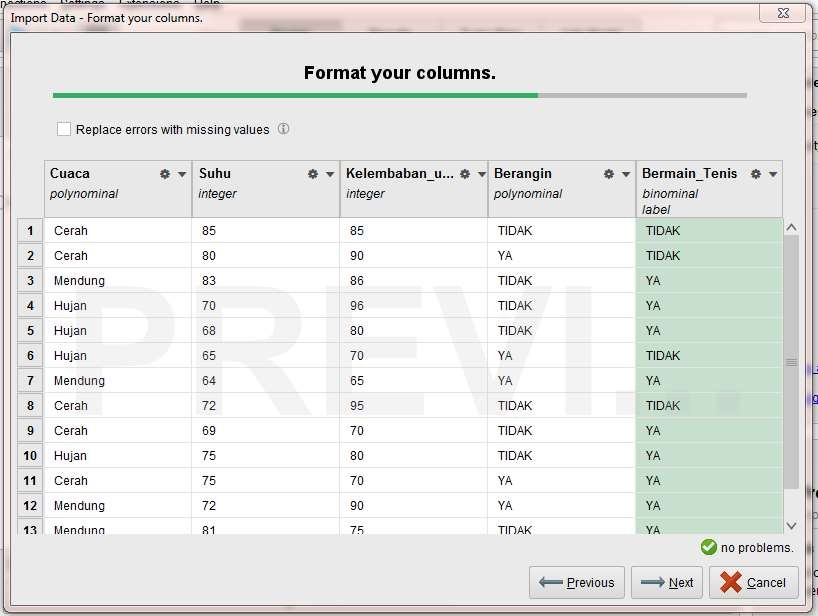


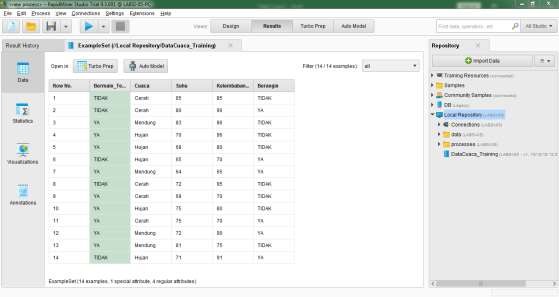


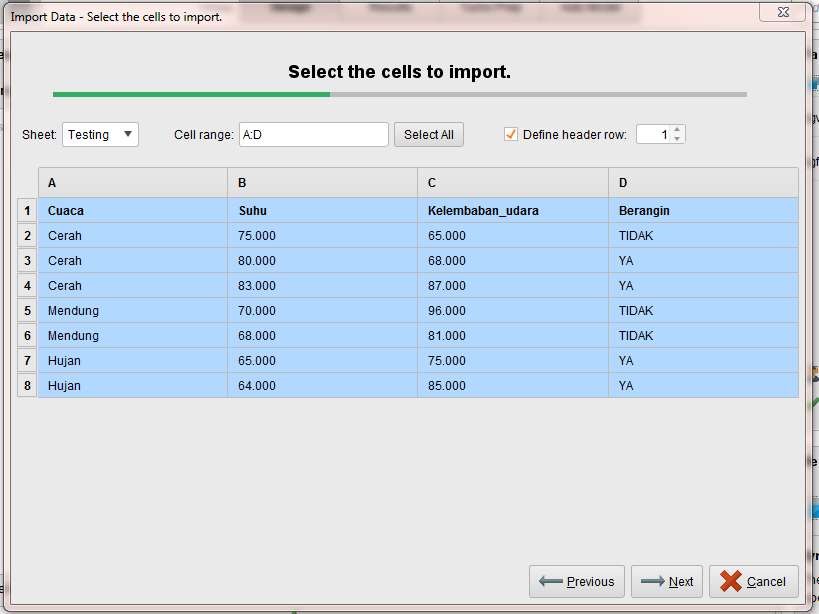


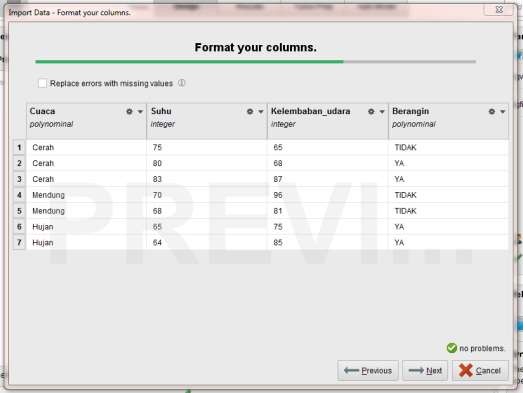


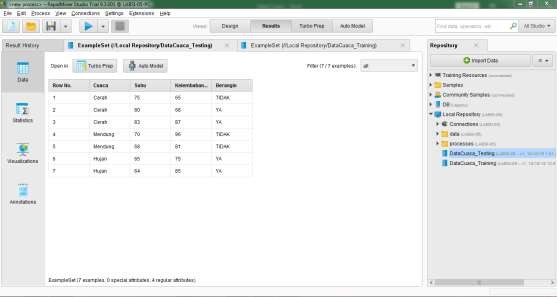


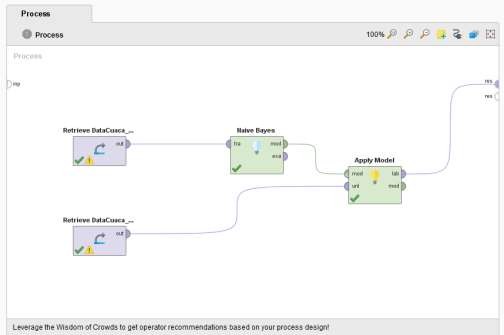


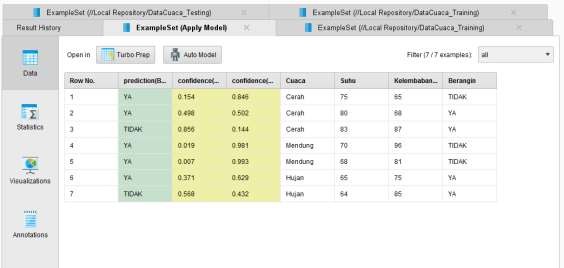












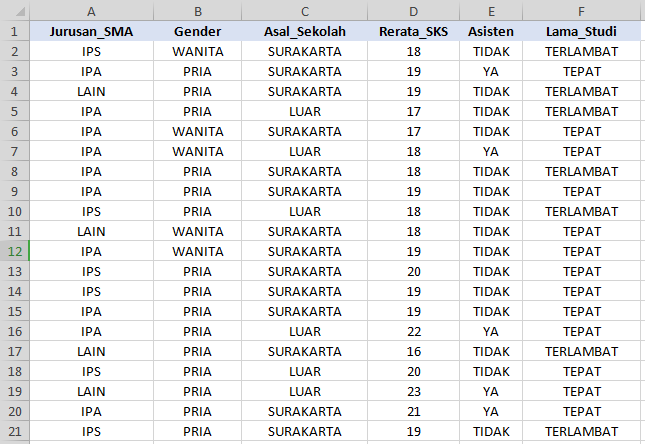


**MODUL 8**

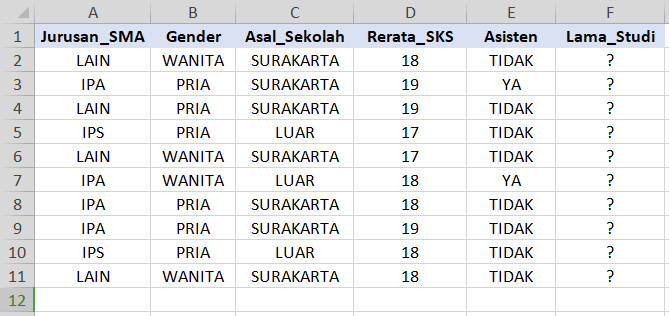
**(Klasifikasi : Naïve Bayes)**

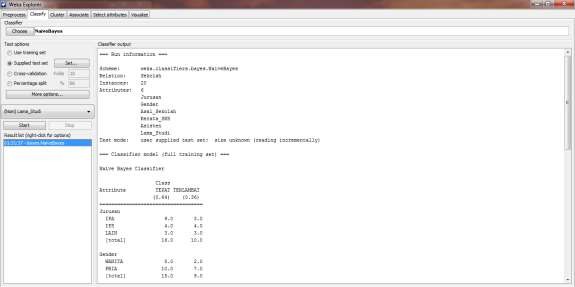
**TUGAS**

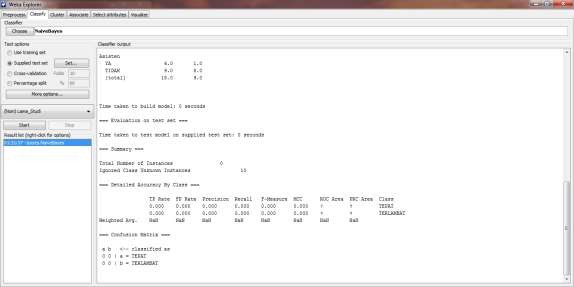
Data Training

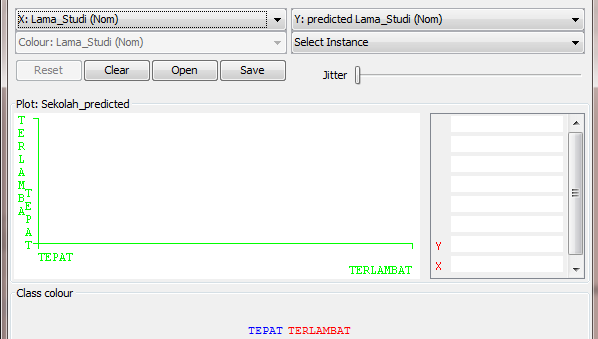


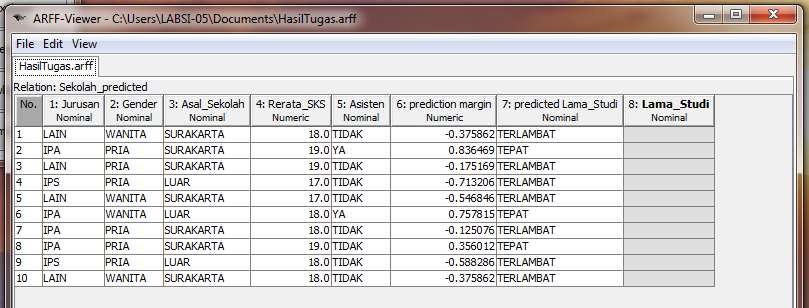
Data Testing

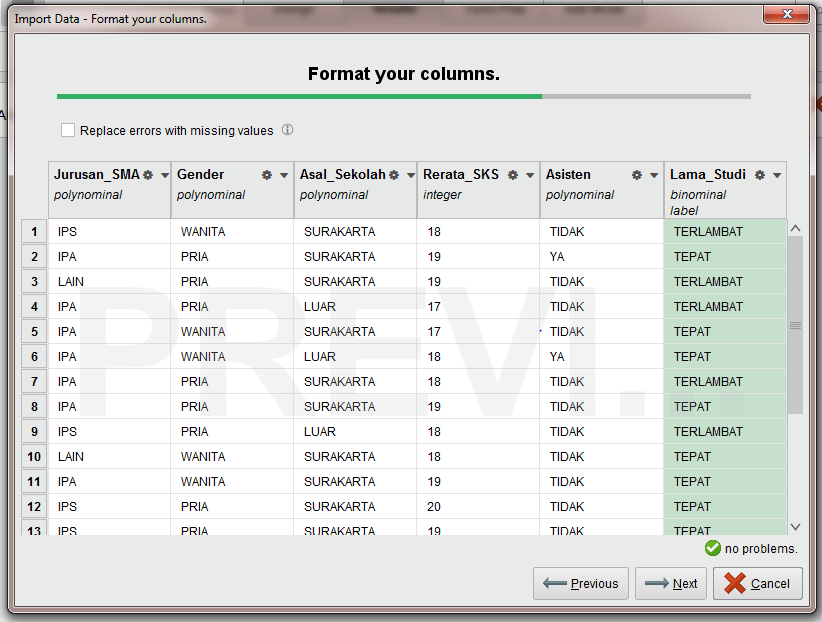


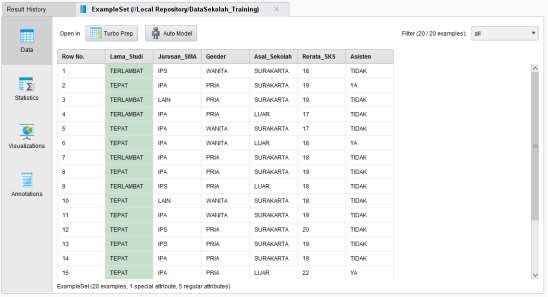


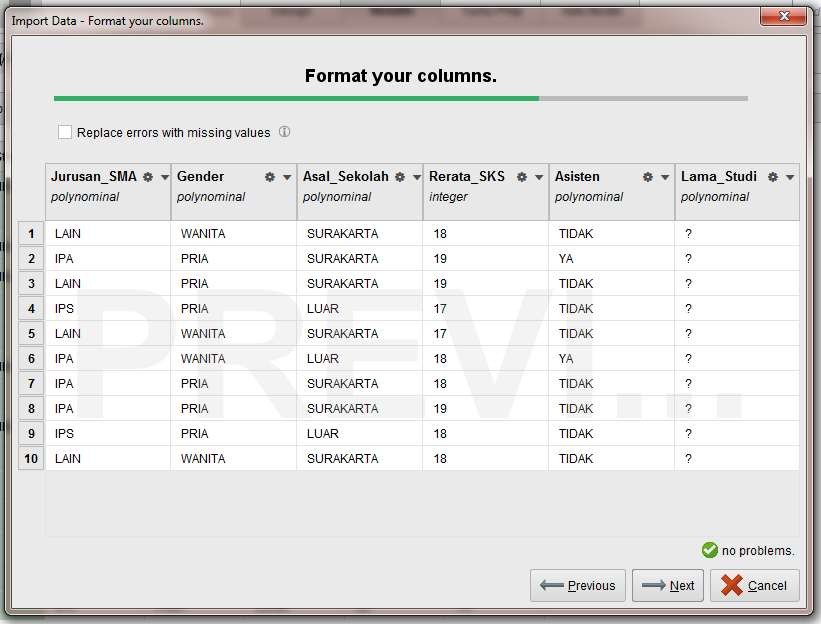


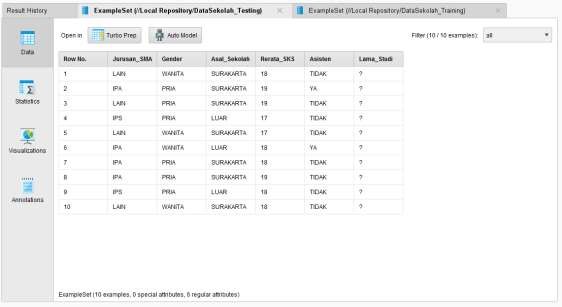


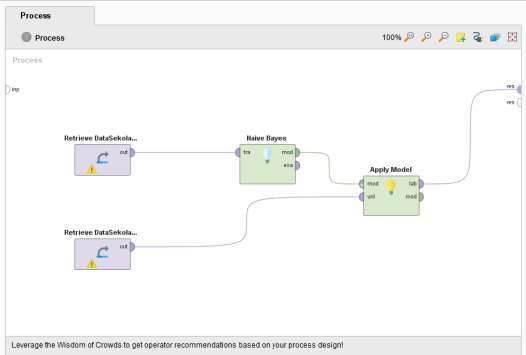


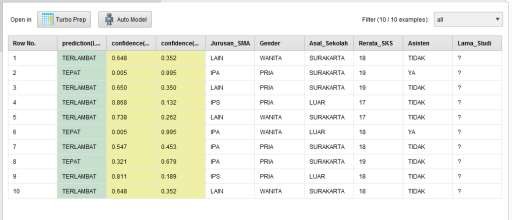












Rerata confidence Lama\_studi dengan nilai TEPAT



Rerata : 0.648 + 0.005 + 0.650 + 0.868 + 0.738 + 0.005 + 0.547 + 0.321 + 0.811 + 0.648

10 : 0.5241

Rerata confidence Lama\_studi dengan nilai TERLAMBAT



Rerata : 0.352 + 0.995 + 0.350 + 0.132 + 0.262 + 0.995 + 0.453 + 0.679 + 0.189 + 0.352

10 : 0.4759

Jumlah yang akan lulus TEPAT : 3 Orang



Jumlah yang akan lulus TERLAMBAT : 7 Orang



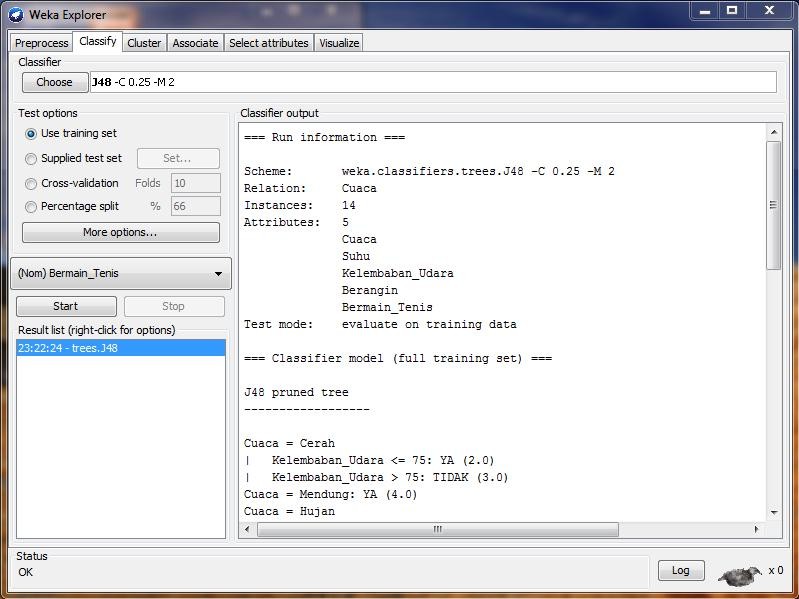
* Ketepatan lama studi Dewi, apabila Dewi adalah seorang WANITA dari jurusan IPA, asal sekolah LUAR SURAKARTA yang mengambil SKS dengan rata-rata 18 SKS tiap semester dan tidak pernah menjadi Asisten selama kuliah.
* Ketepatan lama studi Jono, apabila Jono adalah seorang PRIA dari jurusan selain IPA dan IPS, asal sekolah SURAKARTA yang mengambil SKS dengan rata-rata 17 SKS tiap semester dan pernah menjadi Asisten selama kuliah.
* Dewi dan Jono akan lulus TEPAT

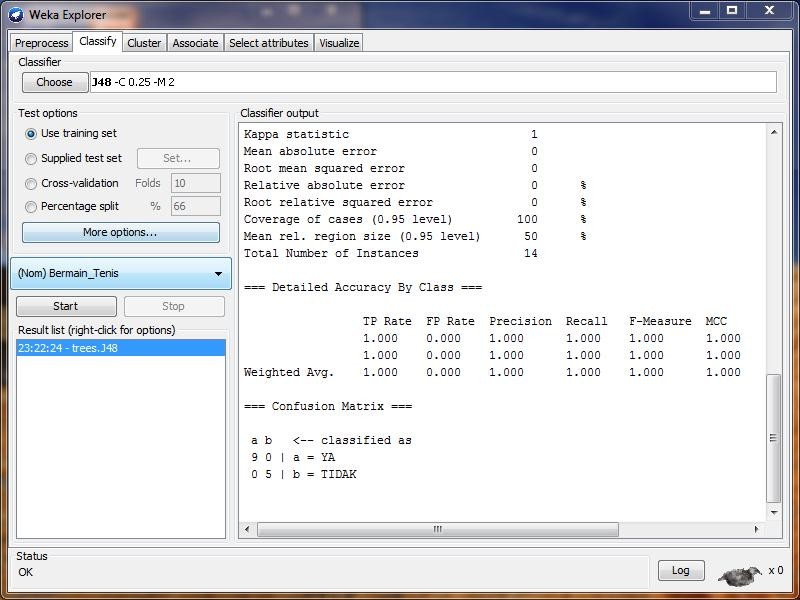
**MODUL 9**

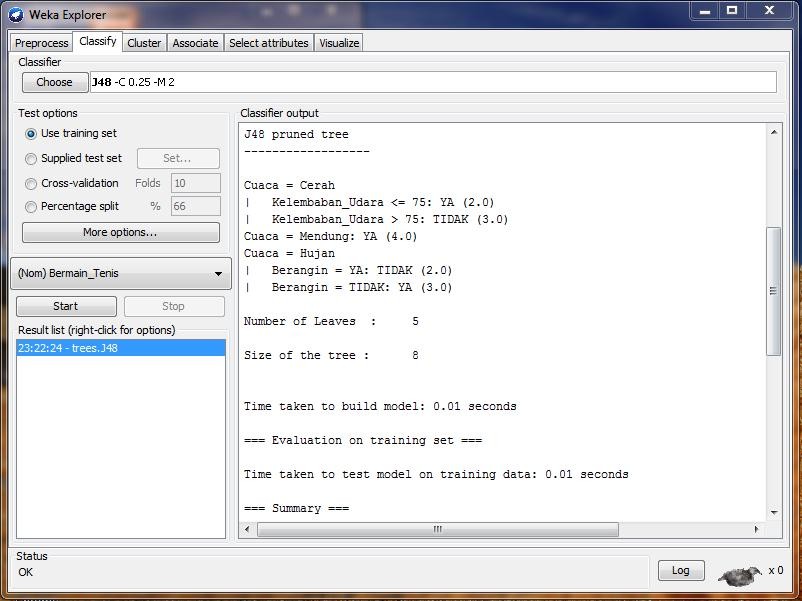
**(Klasifikasi : Decision Tree)**

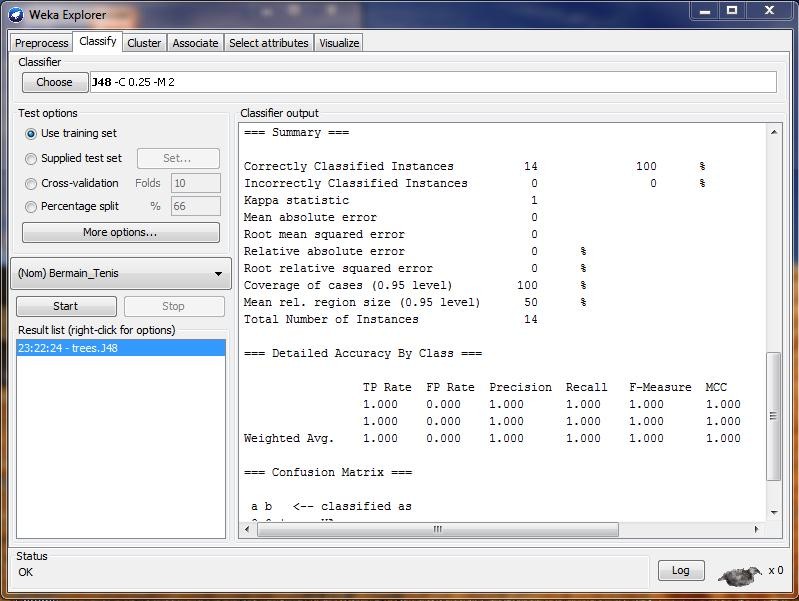
**KEGIATAN**

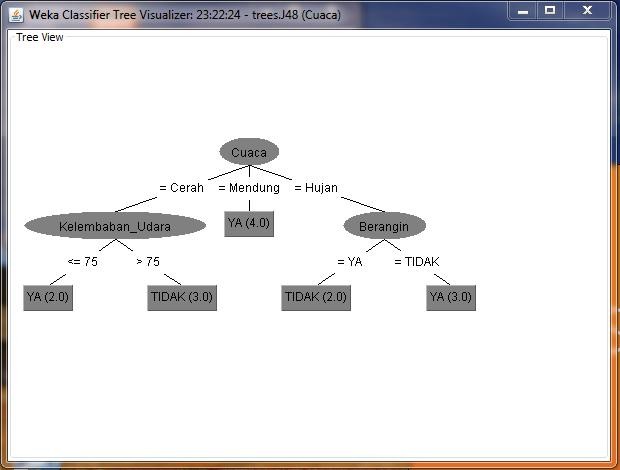
Pohon Keputusan menggunakan Weka











Berdasarkan pohon keputusan tesebut, dapat dilihat jenis-jenis simpul yang ada sebagai berikut :

Simpul Akar = Cuaca

Simpul internal = Kelembapan\_udara, dan Berangin

Simpul Daun = YA, TIDAK

Klasifikasi yang terbentuk yaitu :

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

Cuaca = Cerah, Kelembapan\_udara <= 75. (nilai atribut lain diabaikan)

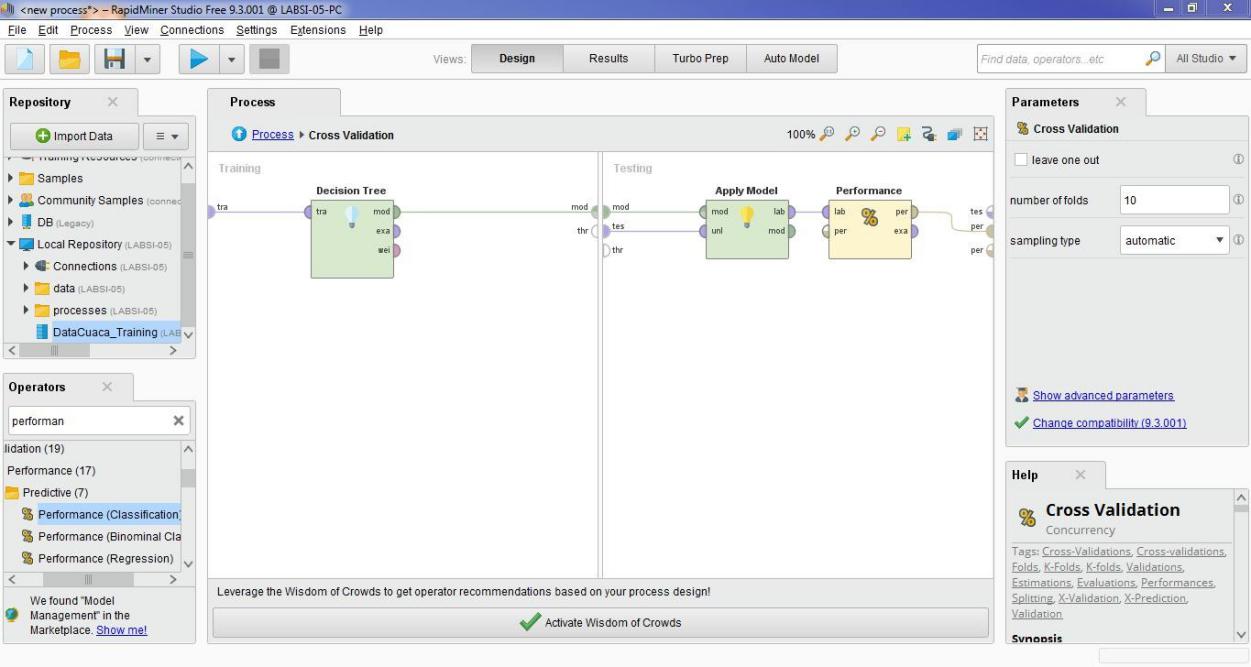
Cuaca = Mendung. (kondisi lain diabaikan)

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

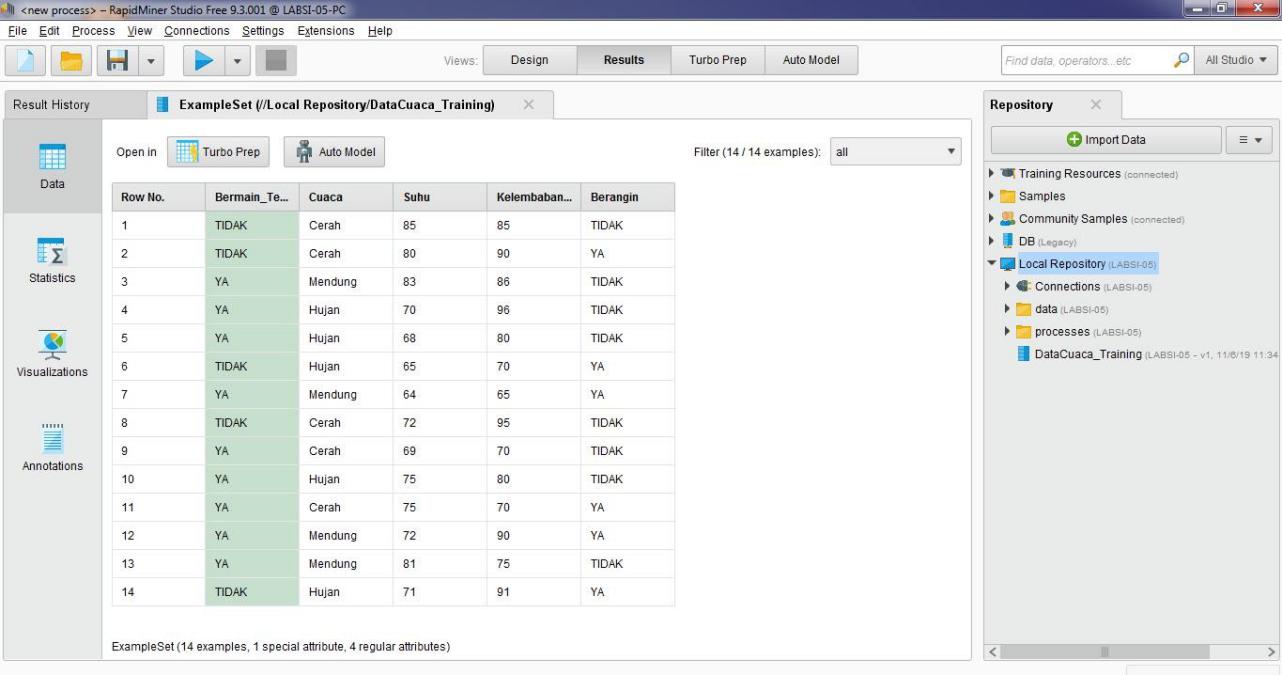
Seseorang tidak akan bermain tenis (TIDAK) jika kondisi sebagai berikut :

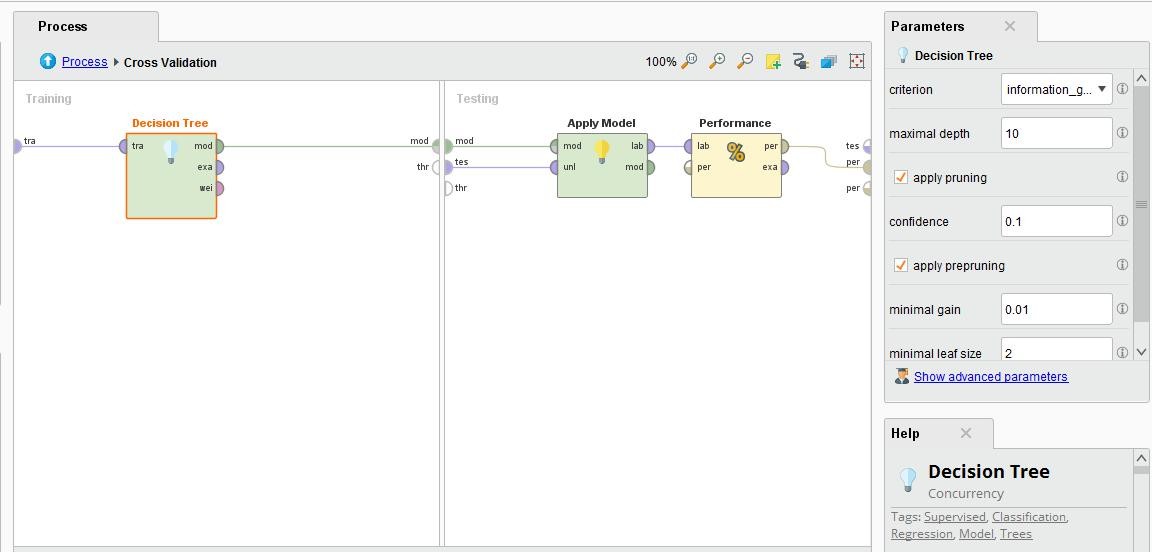
Cuaca = Cerah, Kelembapan\_udara > 75. (nilai atribut lain diabaikan)

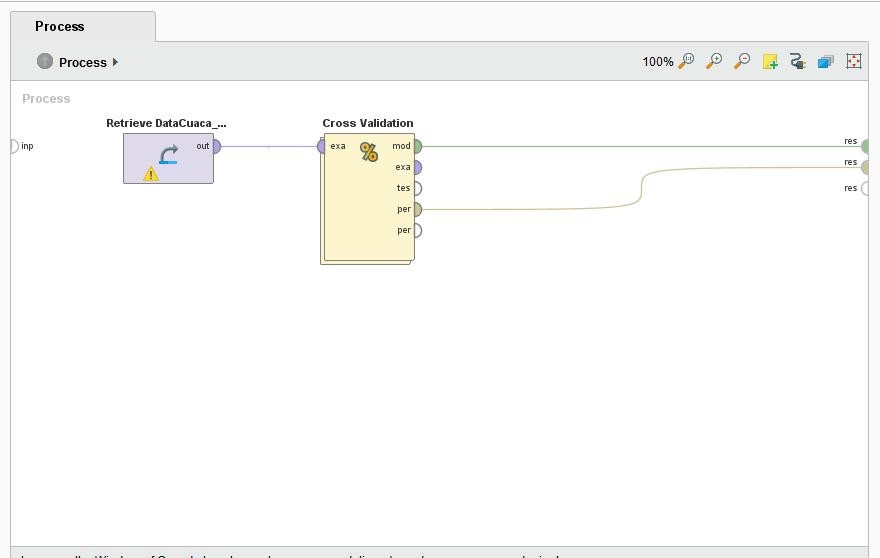
Cuaca = Hujan, Berangin = YA. (nilai atribut lain diabaikan)

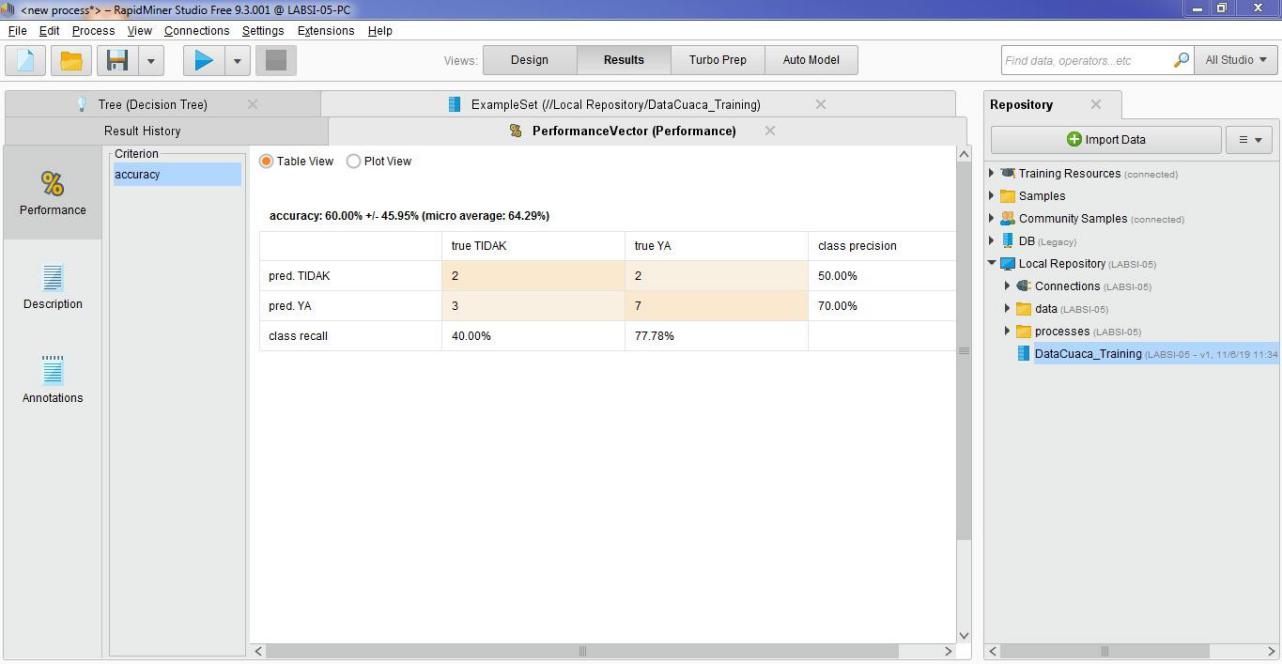


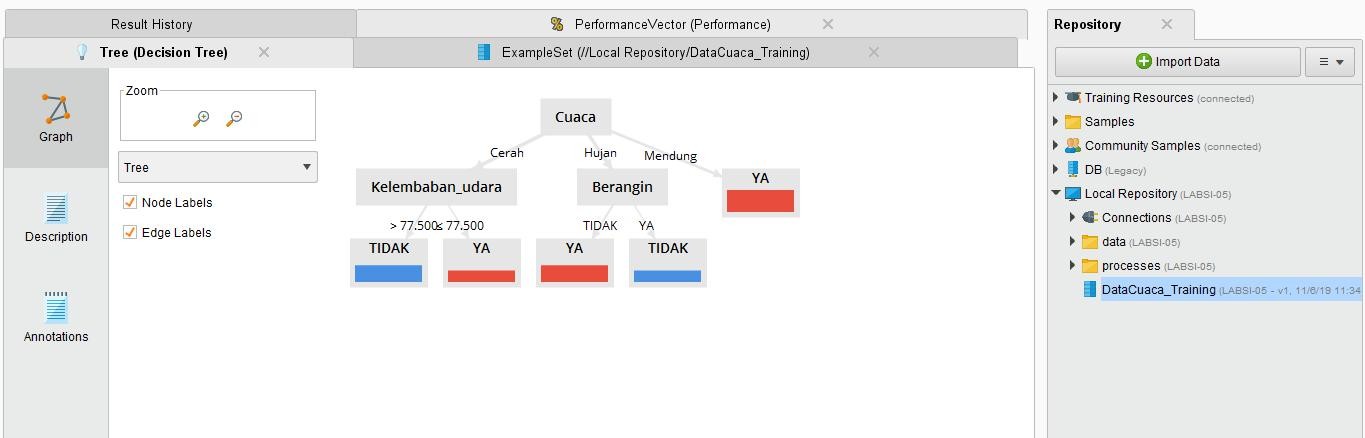
Pohon Keputusan menggunakan Rapid Miner











Berdasarkan pohon keputusan tesebut, dapat dilihat jenis-jenis simpul yang ada sebagai berikut :

Simpul Akar = Cuaca

Simpul internal = Kelembapan\_udara, dan Berangin

Simpul Daun = YA, TIDAK

Klasifikasi yang terbentuk yaitu :

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

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

Cuaca = Mendung. (kondisi lain diabaikan)

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

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

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

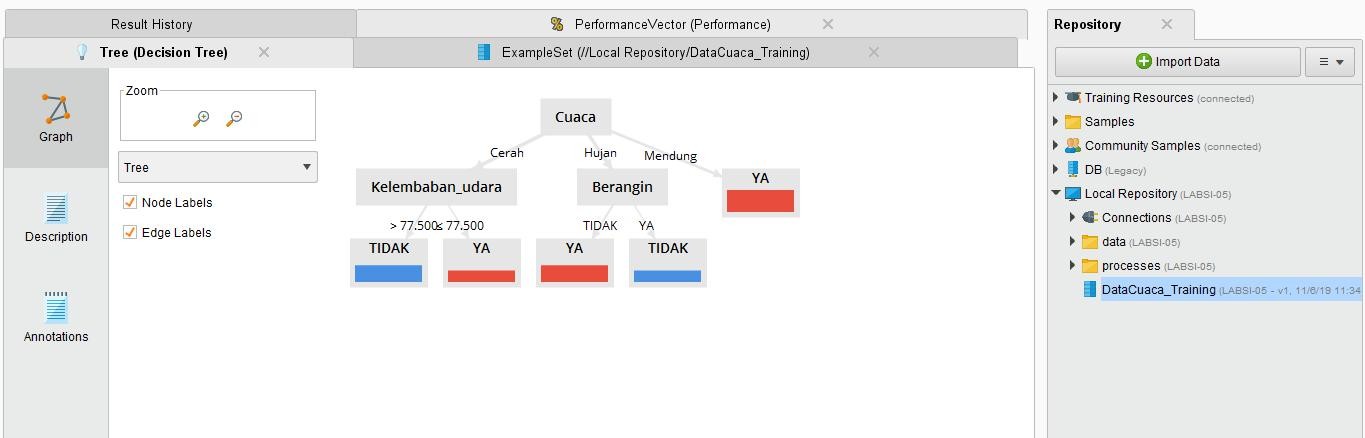
Cuaca = Hujan, Berangin = YA. (nilai atribut lain diabaikan)

**Modul 9**

**(Klasifikasi : Decision Tree)**

**TUGAS**

Berdasarkan pohon keputusan pada kegiatan 9.4.2 (menggunakan Rapid Miner), nilai kelas atribut Bermain\_Tenis pada tabel Testing sebagai berikut :



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cuaca | Suhu | Kelembapan\_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 | YA |
| Hujan | 64 | 85 | YA | TIDAK |

Berdasarkan poho keputusan dari data Training pada Modul 7

Jumlah simpul daun pada pohon keputusan = 4

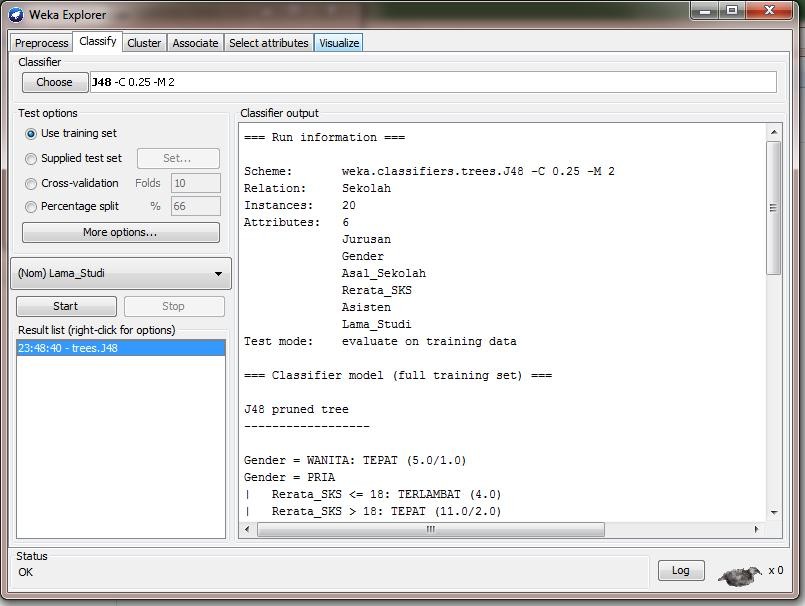
Jumlah simpul keseluruhan pada pohon keputusan = 7

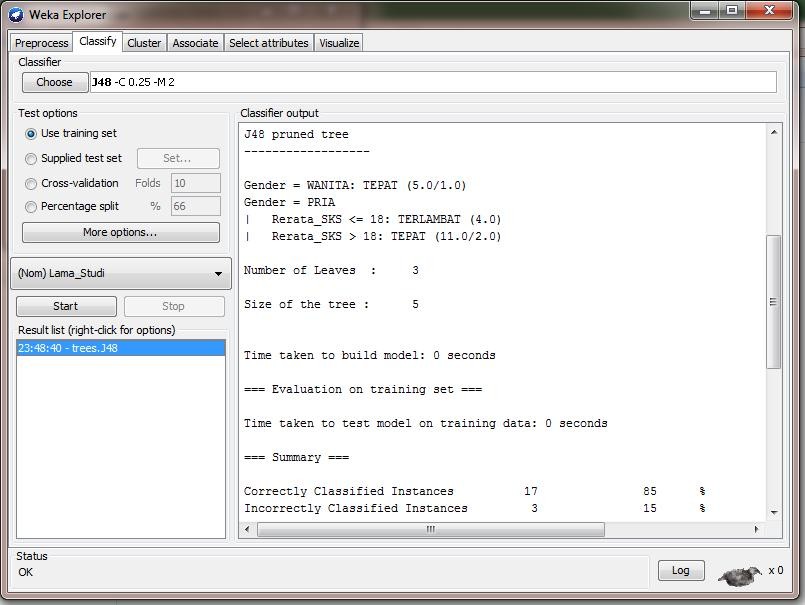
Waktu yang dibutuhkan untuk proses pelatihan = 0 detik

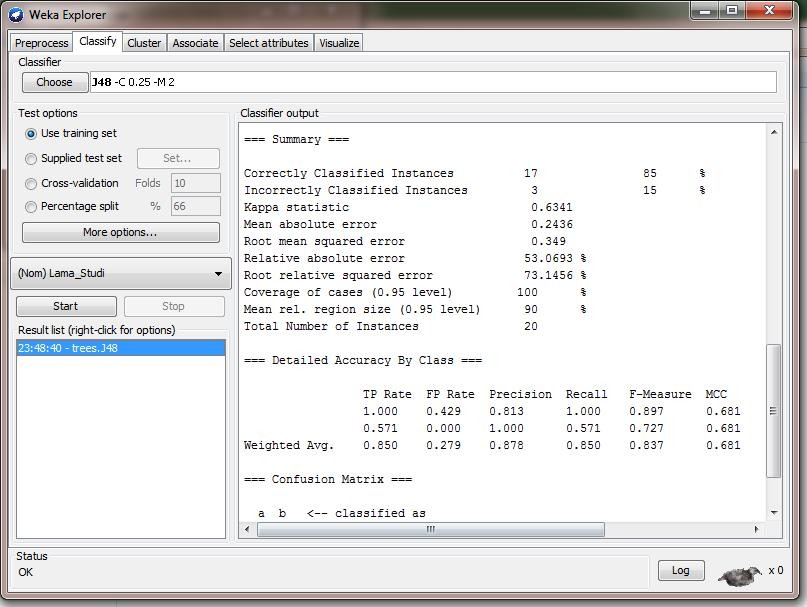
Tingkat ketepatan klasifikasi = 85%

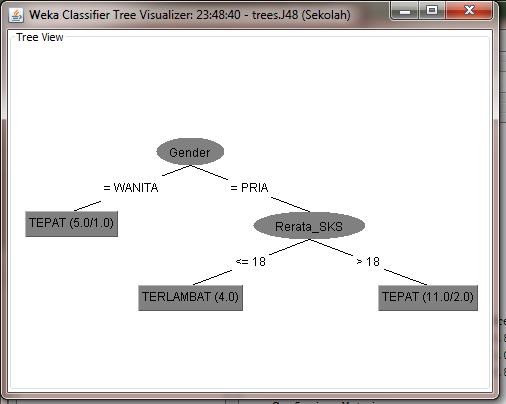
Tingkat ketidaktepatan klasifikasi = 15%

Pohon keputusan berdasarkan data Sekolah.arff

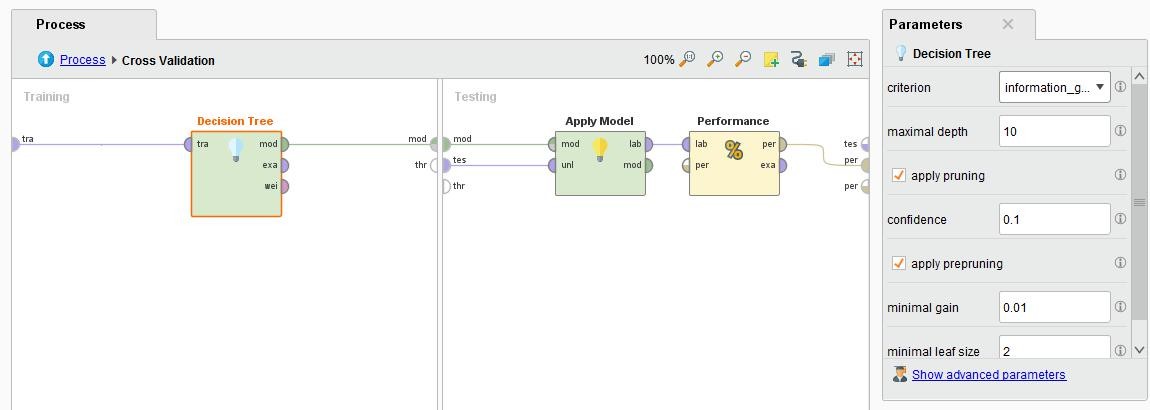


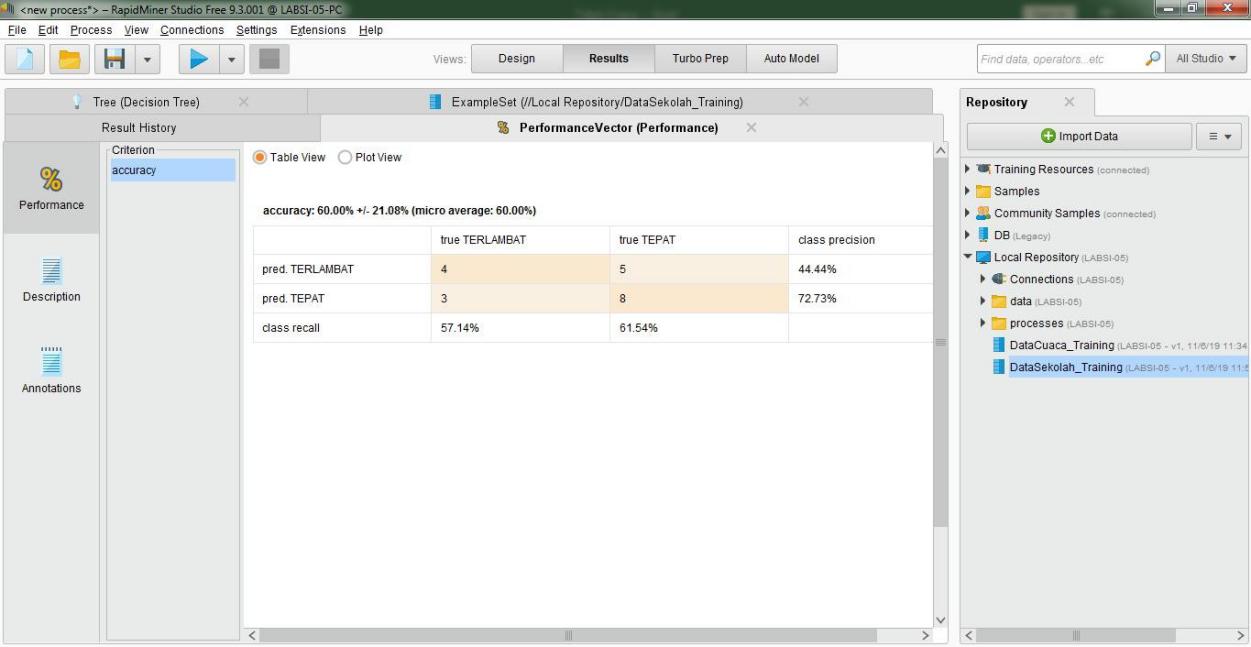


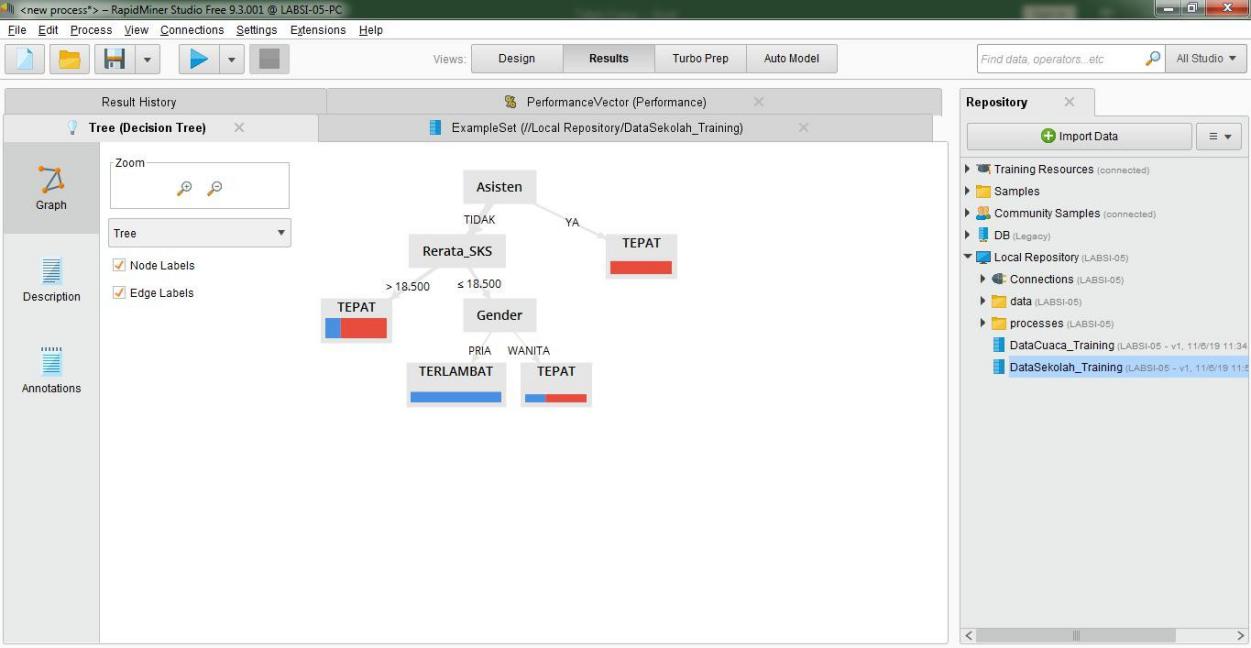


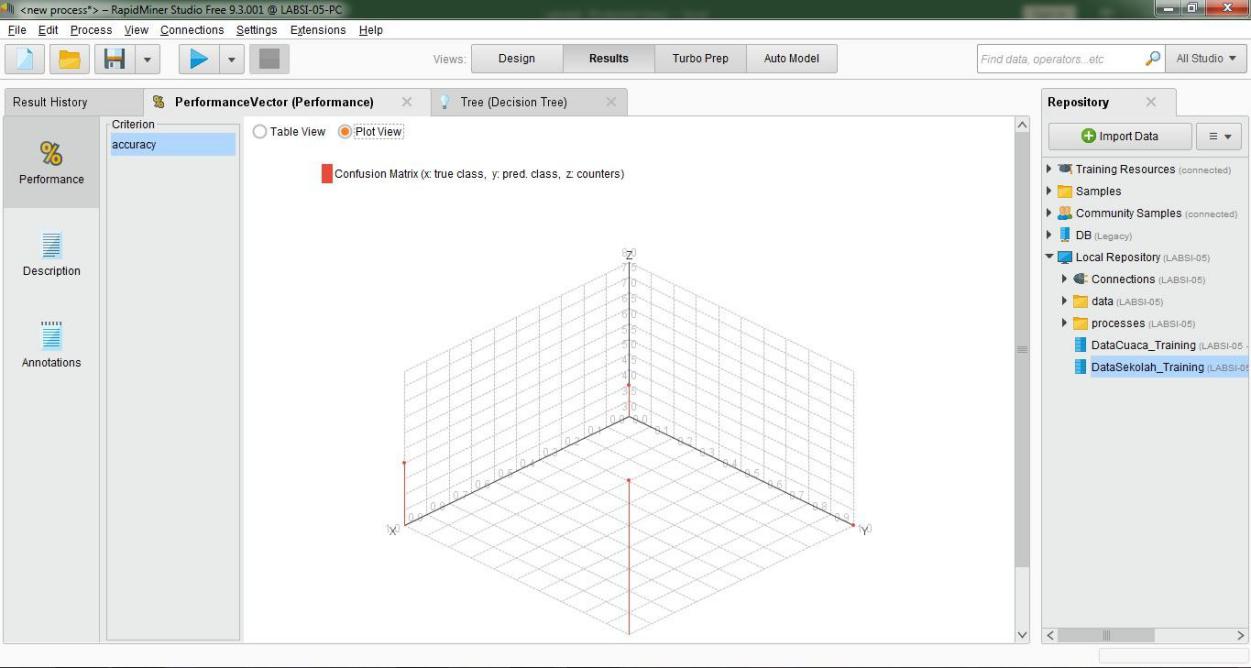












Jumlah simpul daun pada pohon keputusan = 3

Jumlah simpul keseluruhan pada pohon keputusan = 5

Waktu yang dibutuhkan untuk proses pelatihan = 0 detik

Tingkat ketepatan klasifikasi = 85%

Tingkat ketidaktepatan klasifikasi = 15%

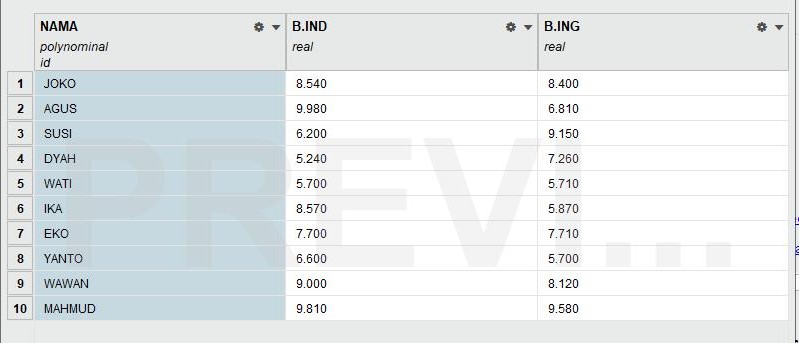
**MODUL 10**

**(Clustering : K-Means)**

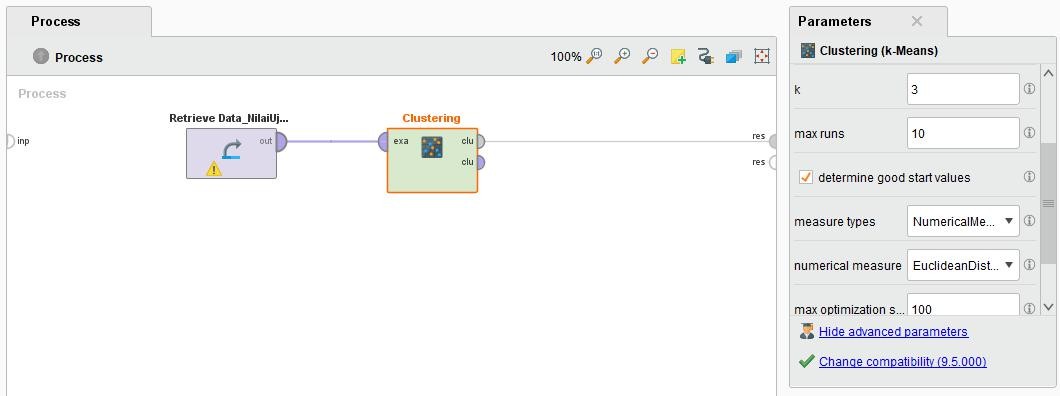
**KEGIATAN**

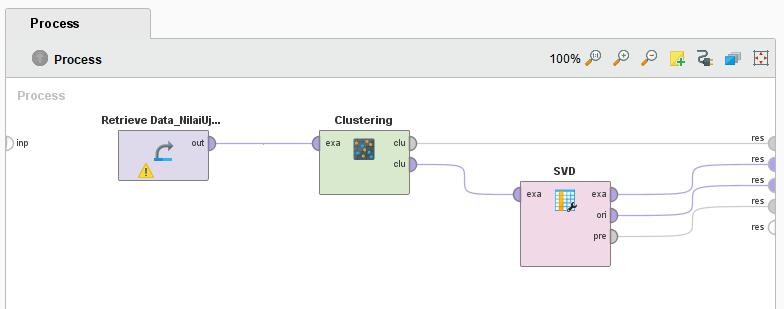
Tabel data nilai siswa





Area Design Process

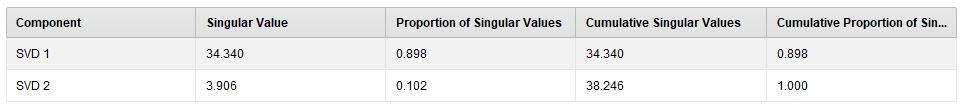




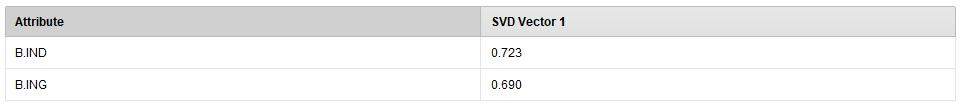
Hasil proses Clustering dengan Algoritma k-Means

SVD (Singular Value Decomposition)

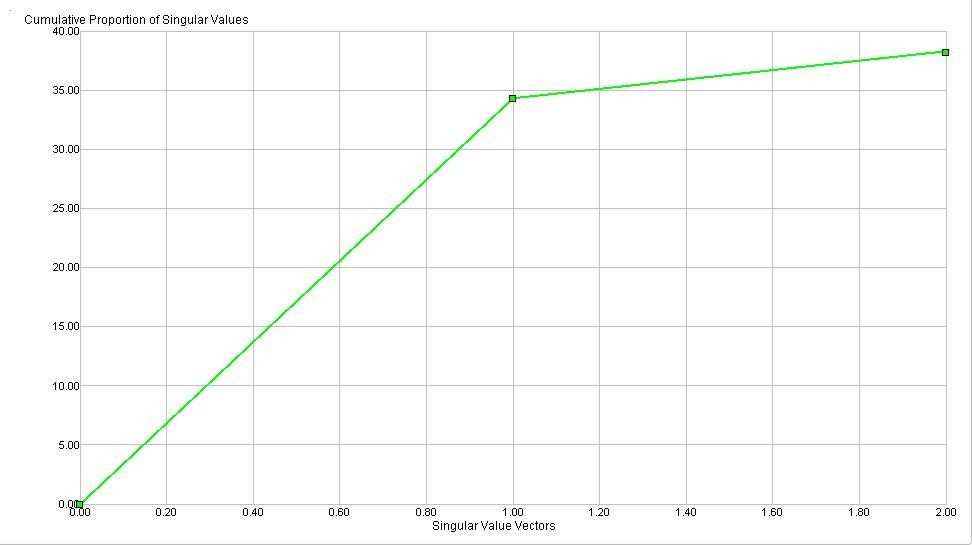
Nilai Eigenvalue



Nilai SVD Vectors



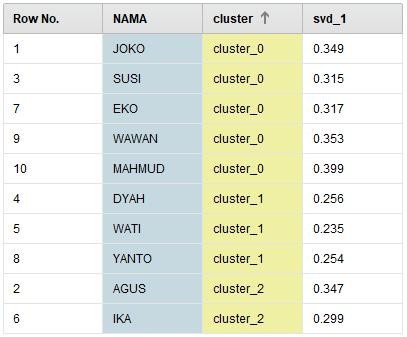
Nilai Cumulative Variance



ExampleSet (k-Means)

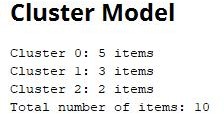


ExampleSet (SVD)

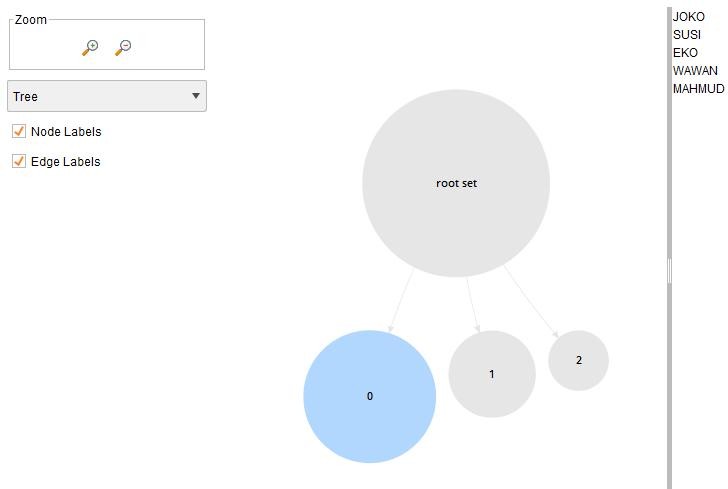


Cluster Model (Clustering)

Description



Graph

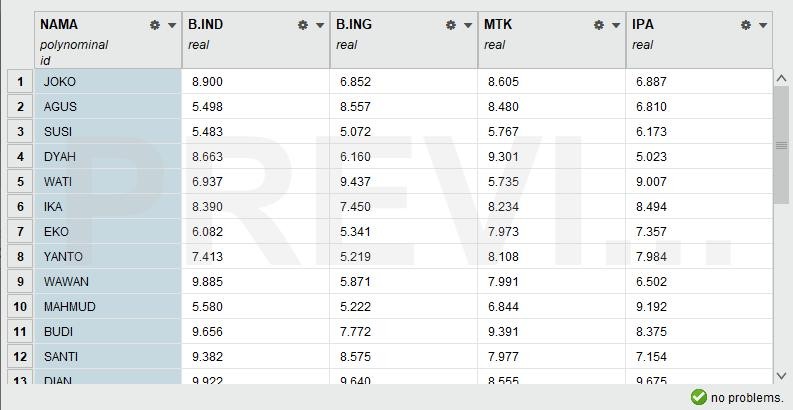


**MODUL 10**

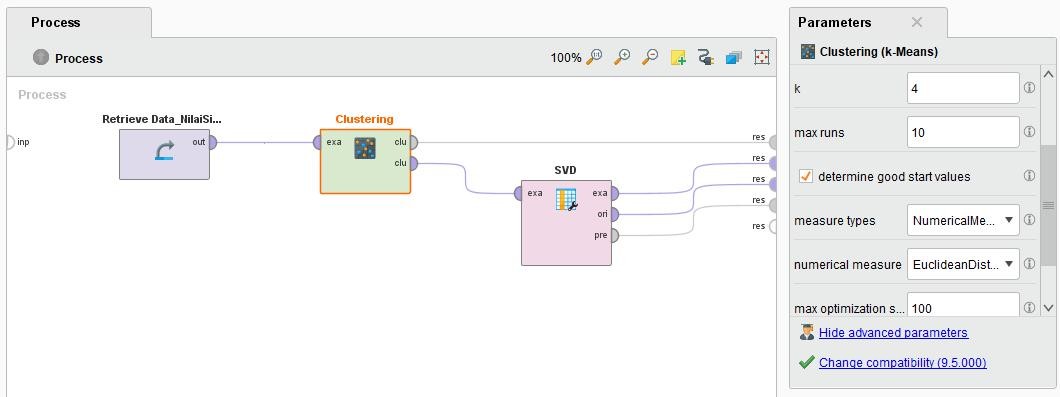
**(Clustering : K-Means)**

**TUGAS**

Tabel data nilai siswa



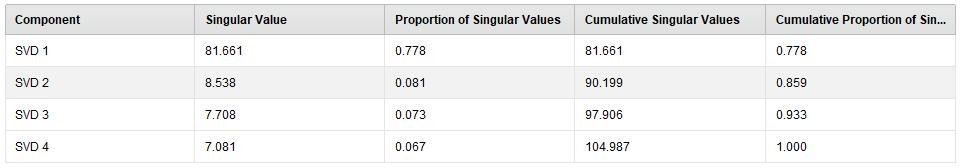
Area Design Process



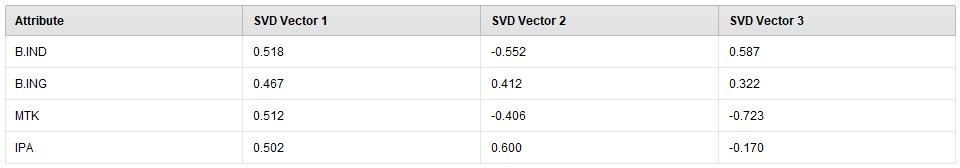
Hasil proses Clustering dengan Algoritma k-Means

SVD (Singular Value Decomposition)

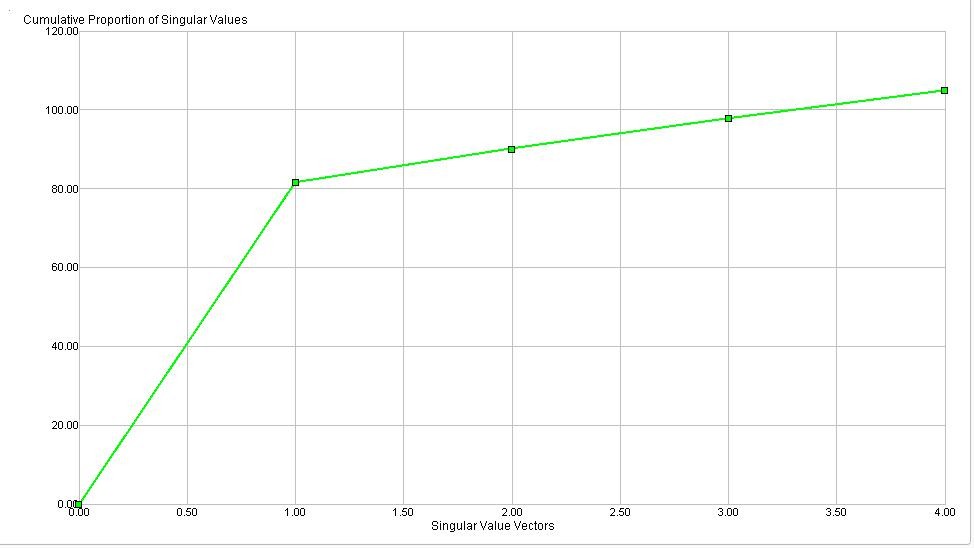
Nilai Eigenvalue



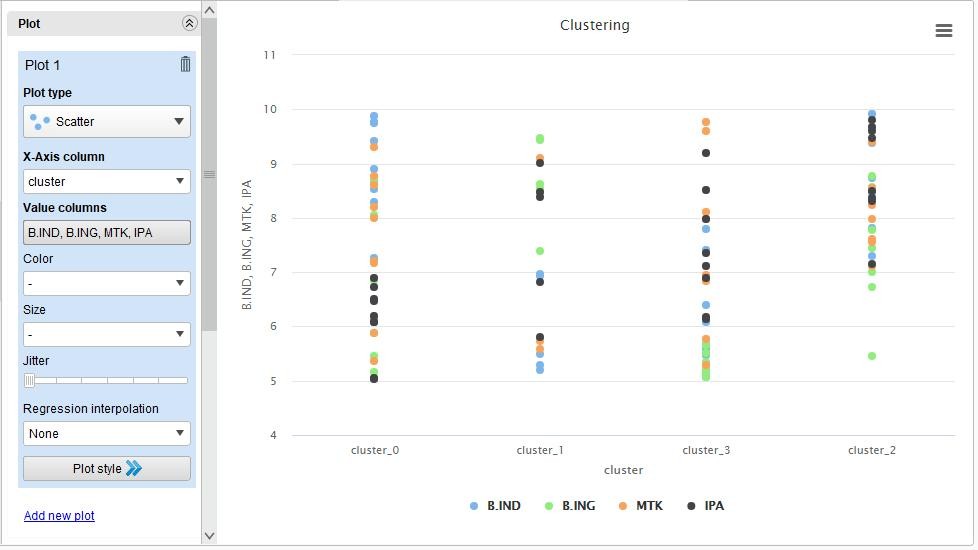
Nilai SVD Vectors



Nilai Cumulative Variance

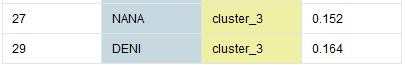


ExampleSet (k-Means)



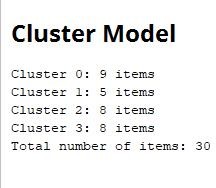
ExampleSet (SVD)





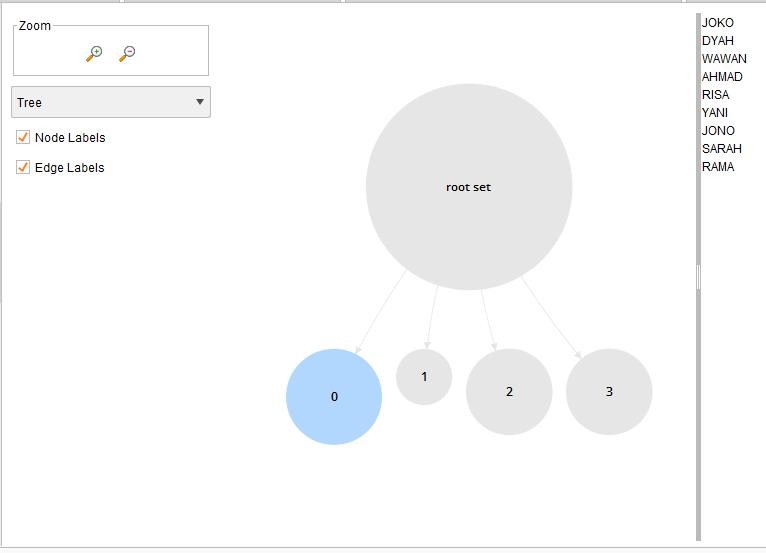
Cluster Model (Clustering)

Description

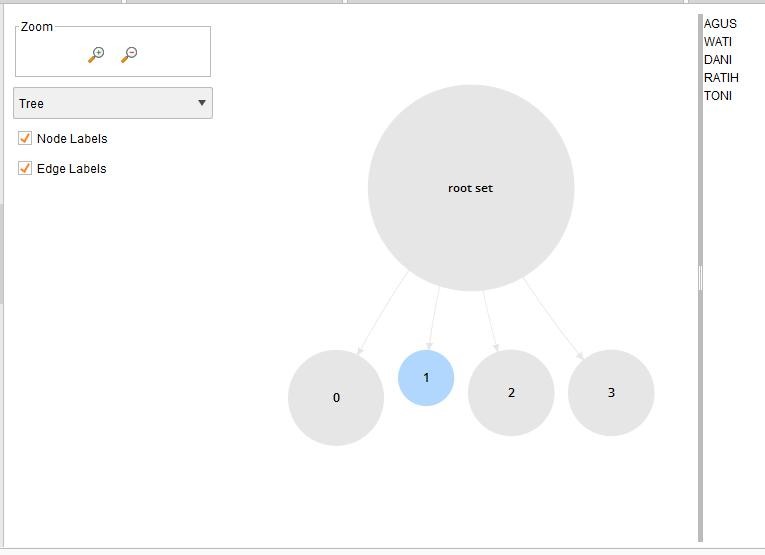


Graph

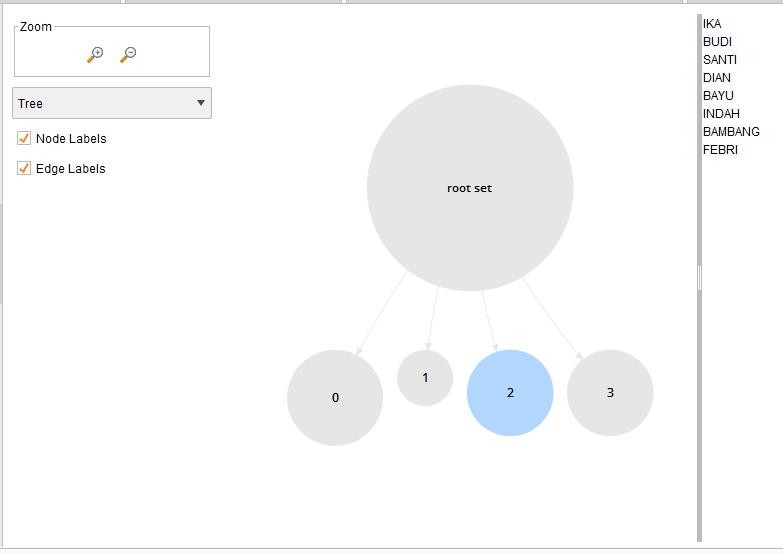
Cluster 0 : JOKO, DYAH, WAWAN, AHMAD, RISA, YANI, JONO, SARAH, RAMA



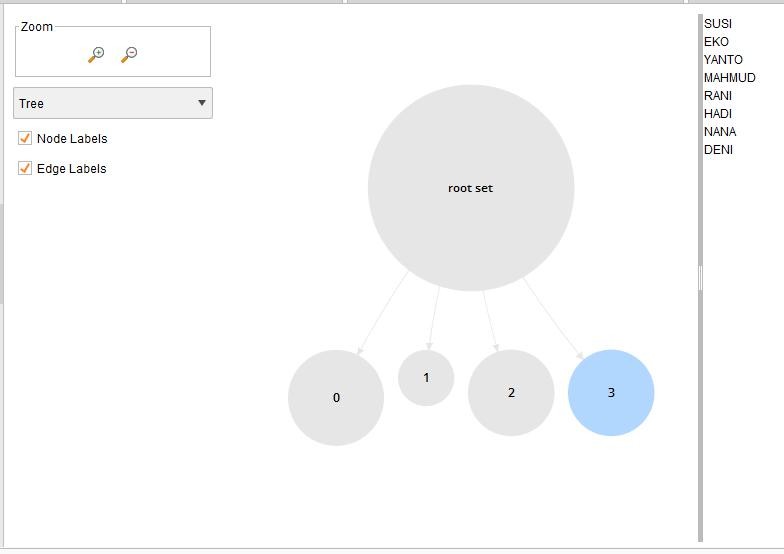
Cluster 1 : AGUS, WATI, DANI, RATIH, TONI



Cluster 2 : IKA, BUDI, SANTI, DIAN, BAYU, INDAH, BAMBANG, FEBRI



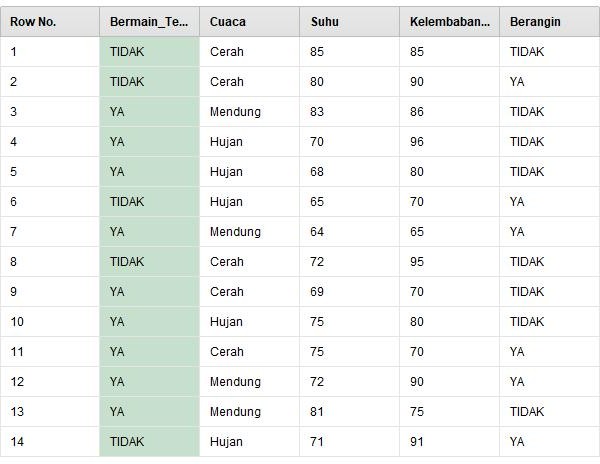
Cluster 3 : SUSI, EKO, YANTO, MAHMUD, RANI, HADI, NANA, DENI

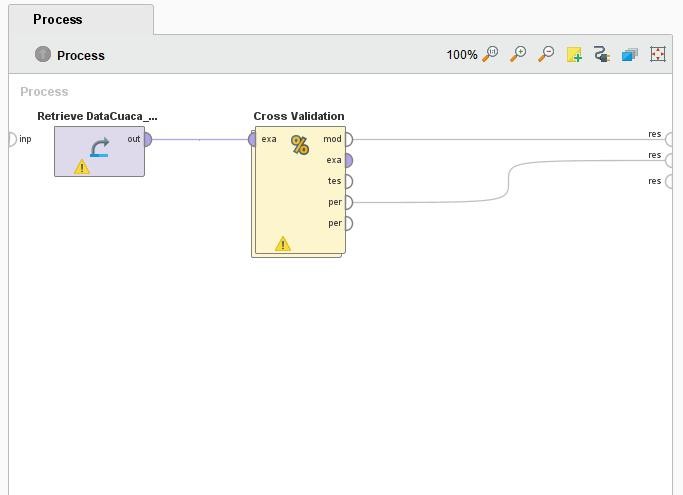


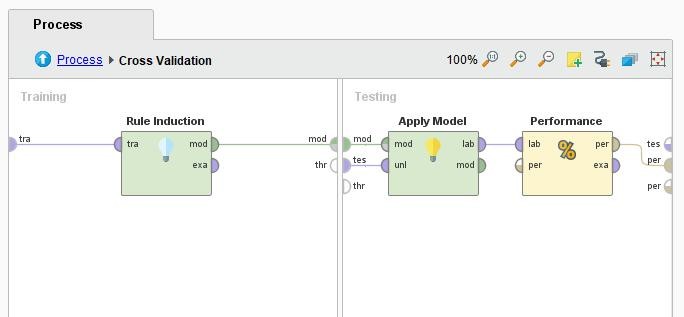
**MODUL 11**

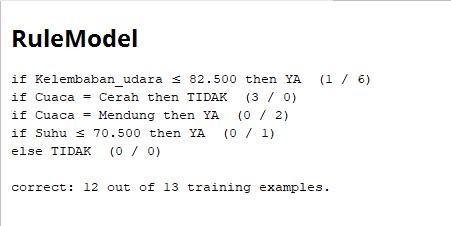
**(Induksi dan Aturan Asosiasi)**

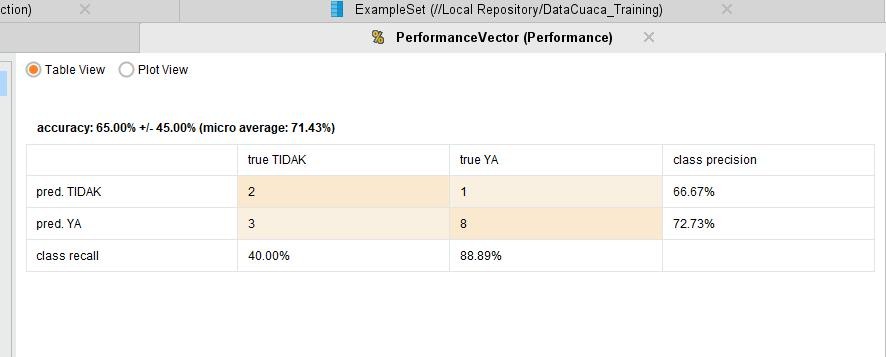
**KEGIATAN**

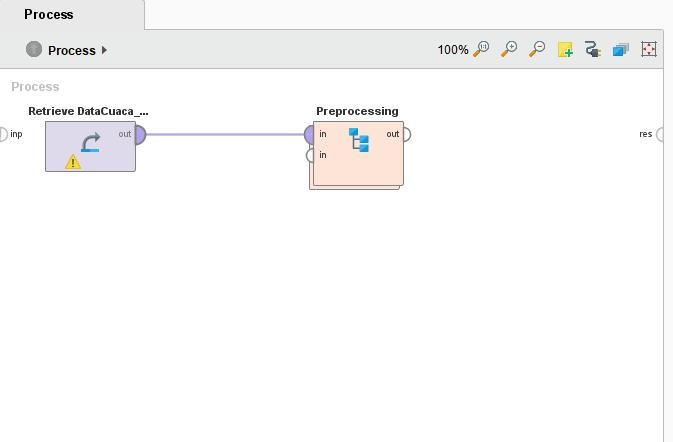


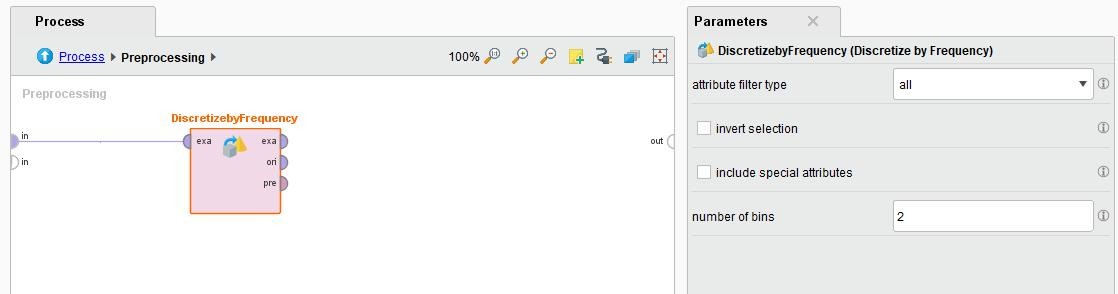


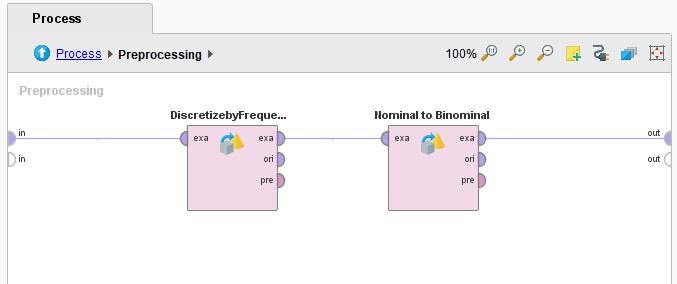


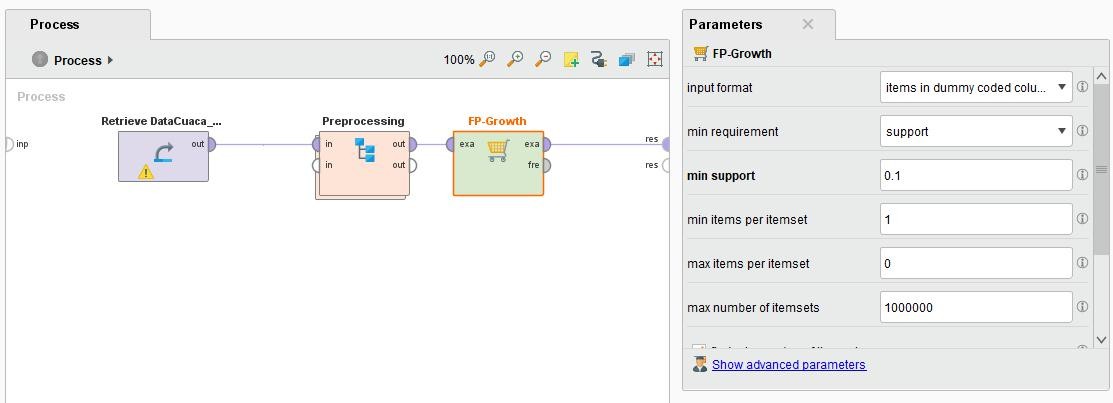


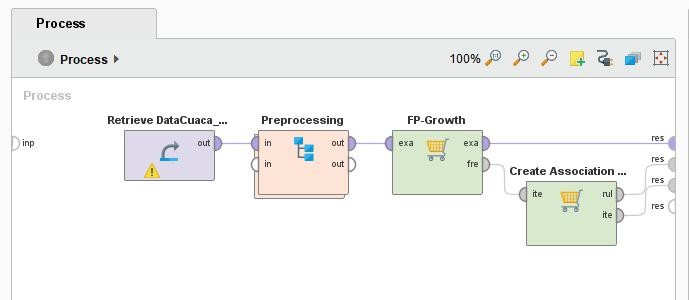




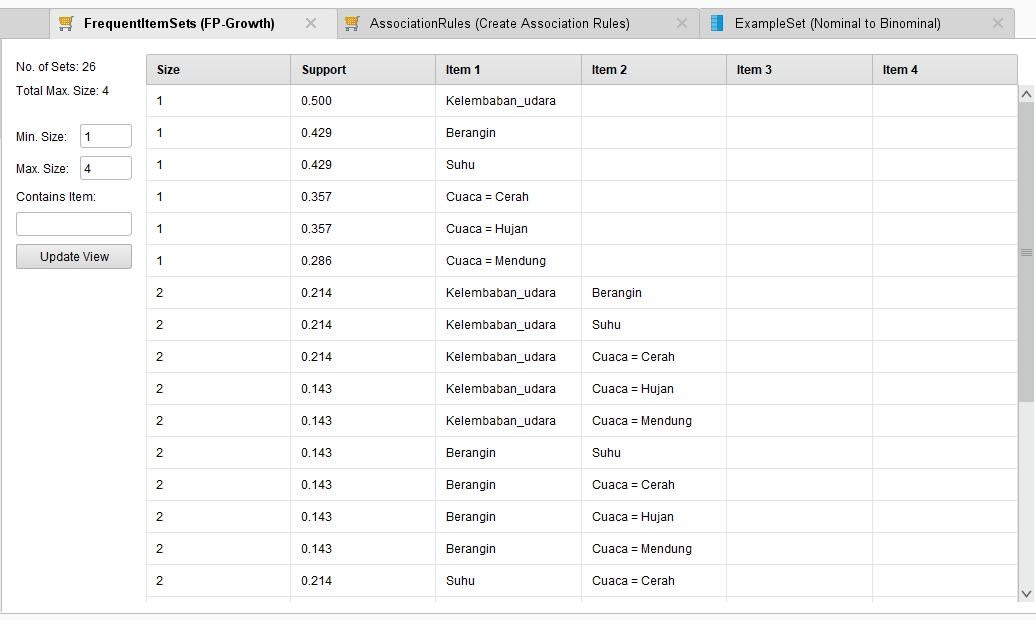


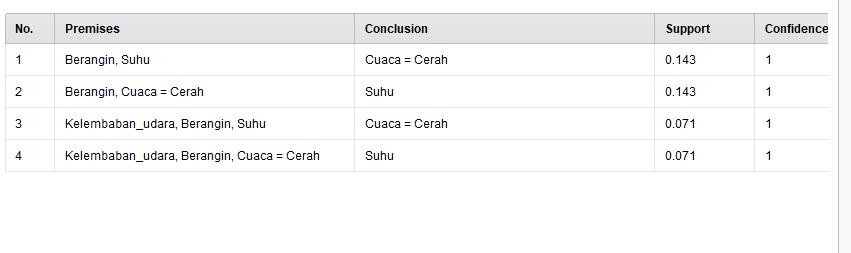


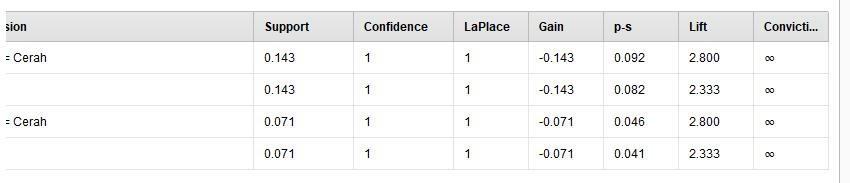


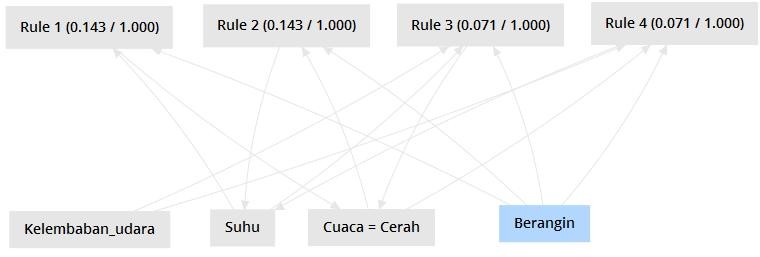


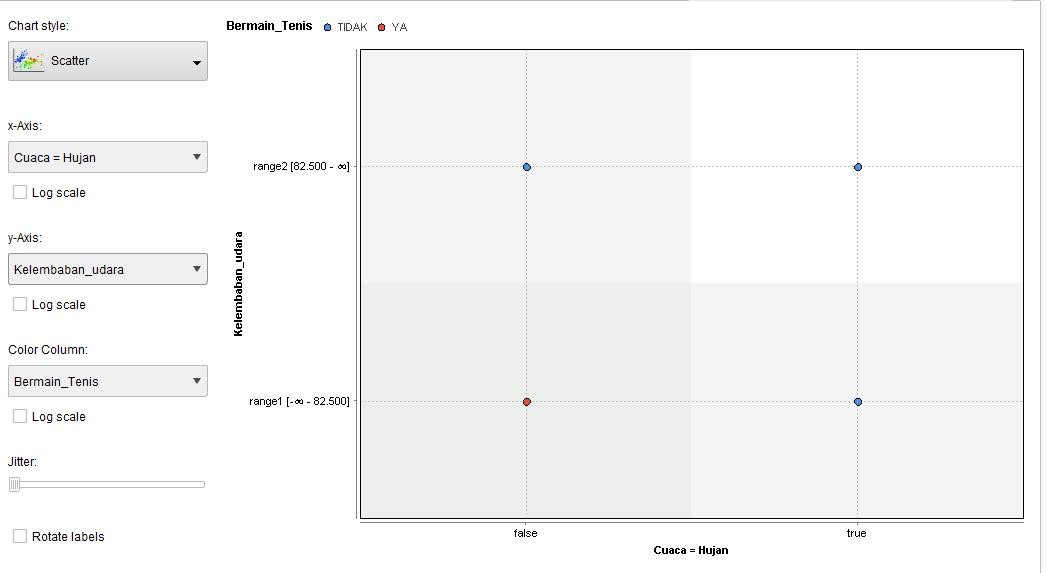
A. FREQUEN









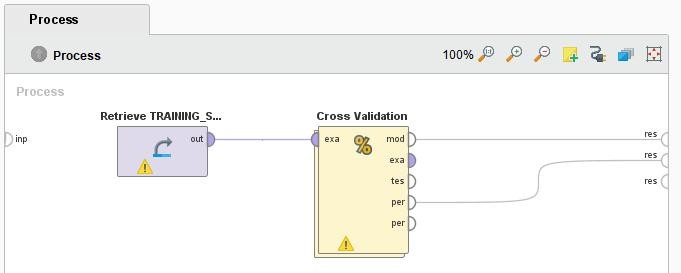


**MODUL 11**

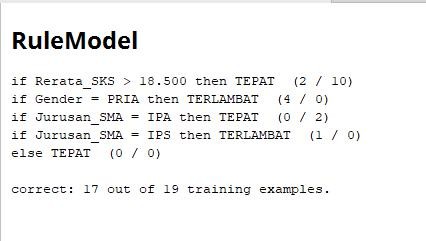
**(Induksi dan Aturan Asosiasi)**

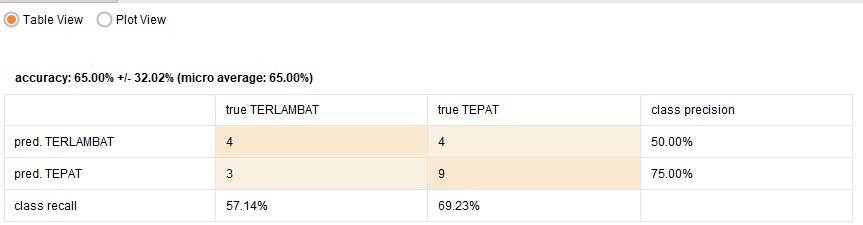
**TUGAS**

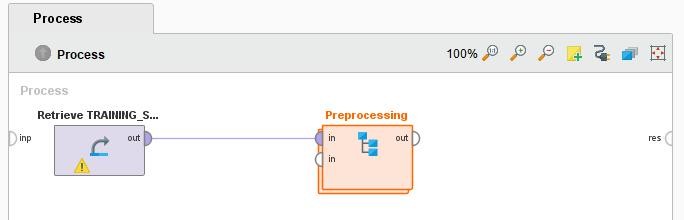


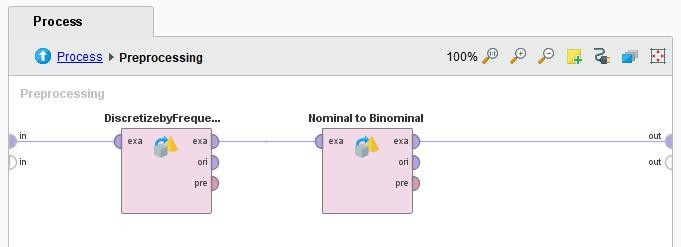


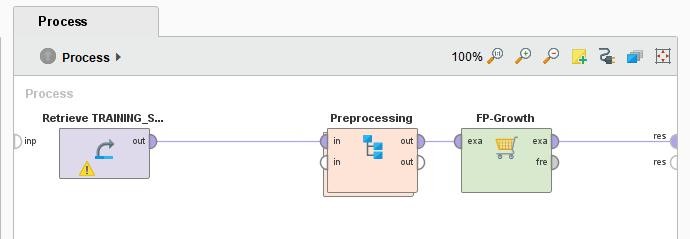


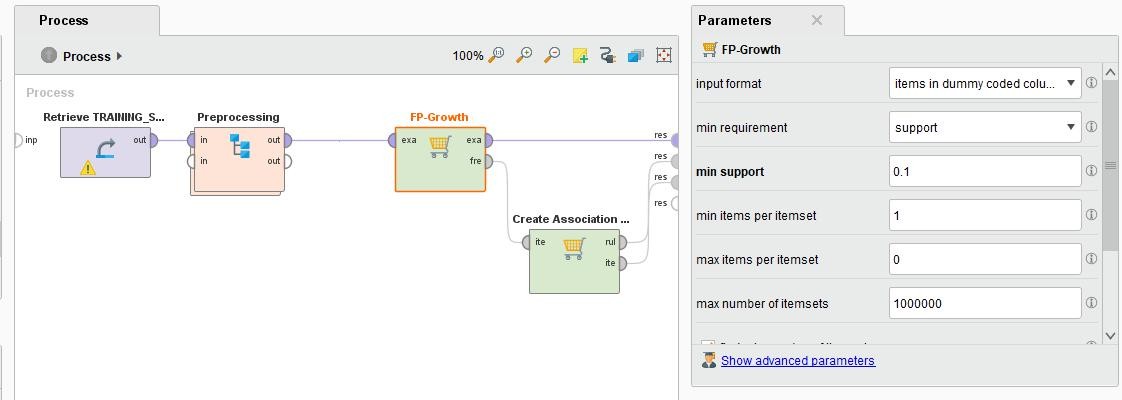


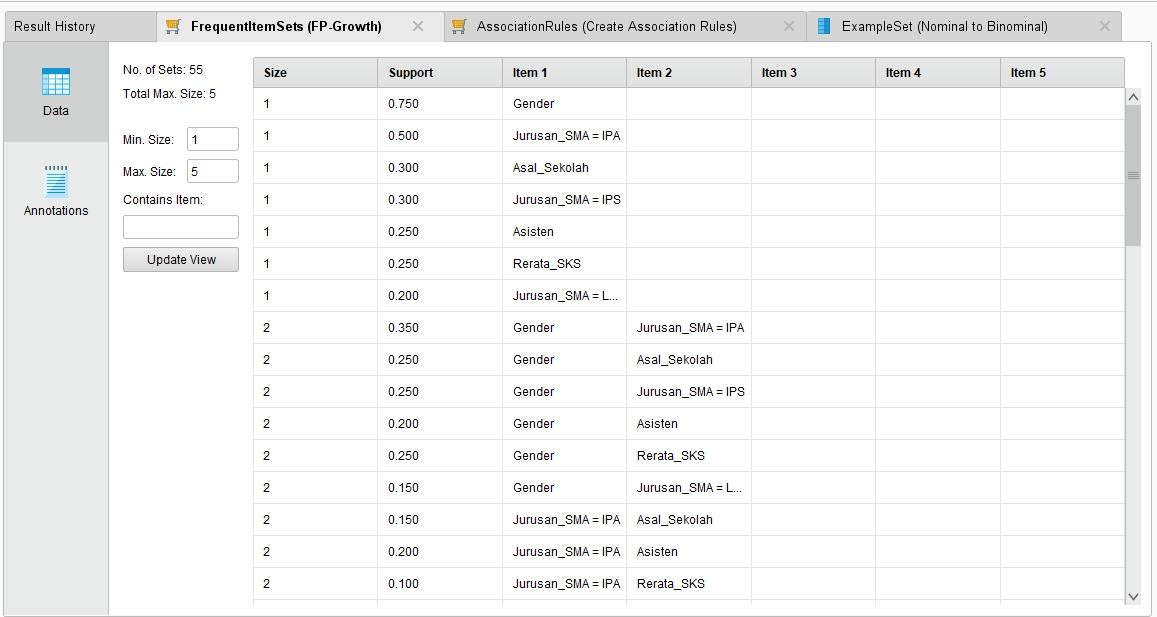


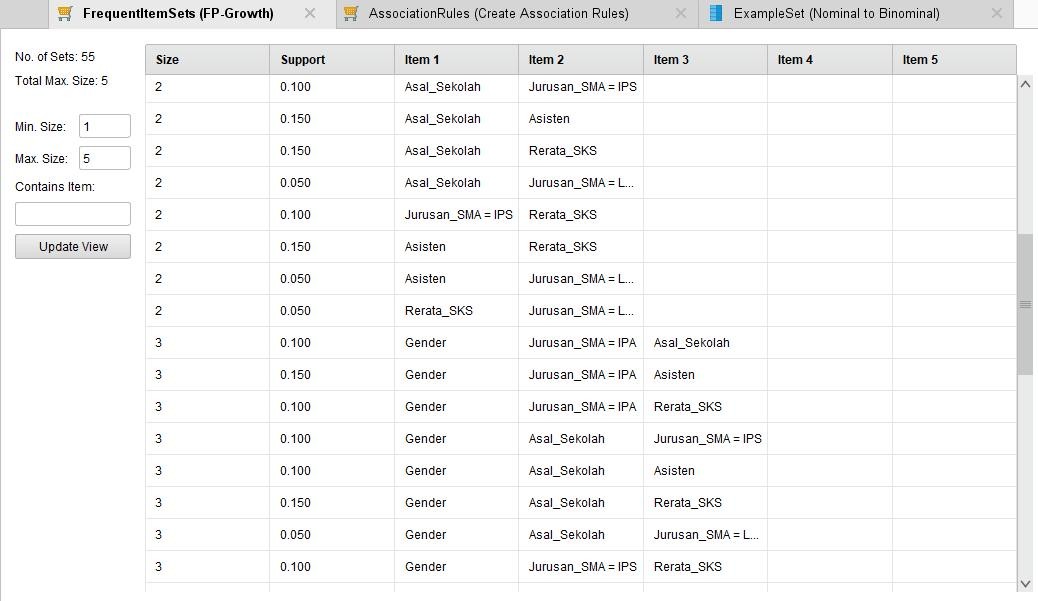


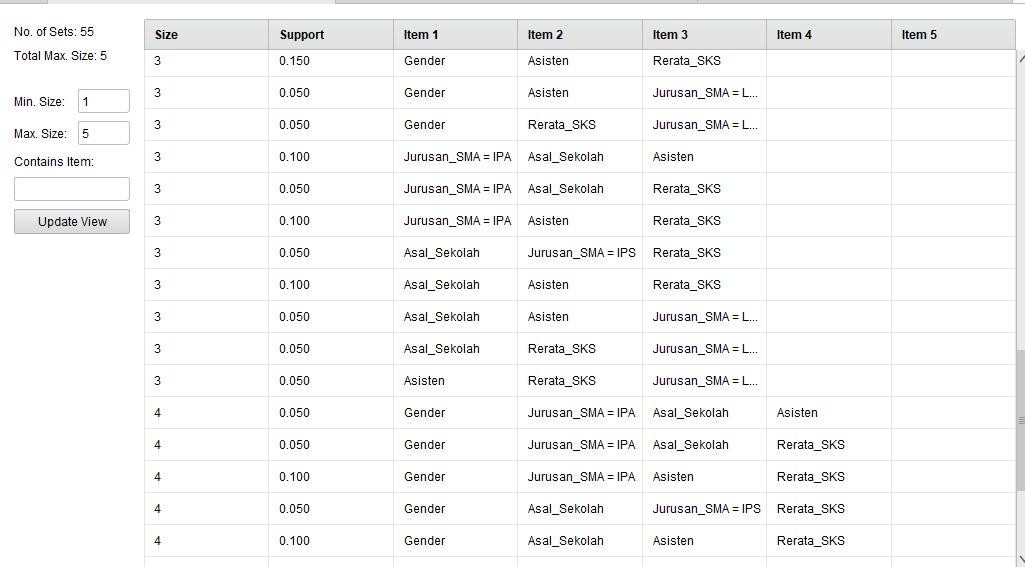




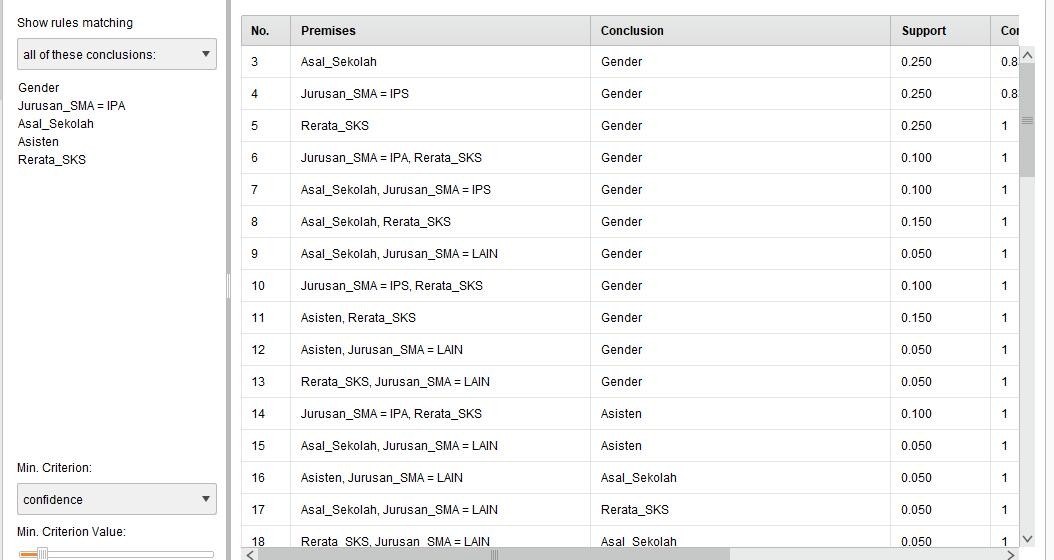




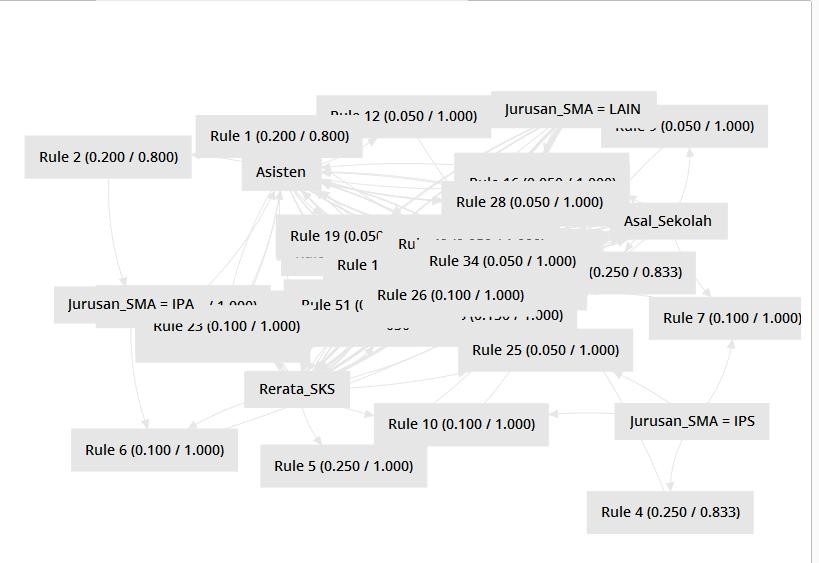


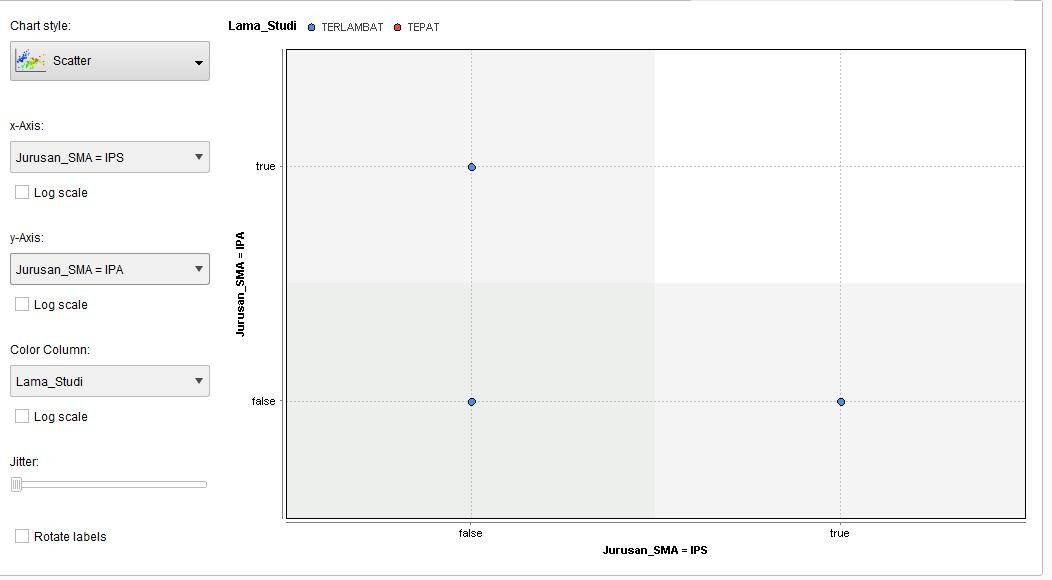










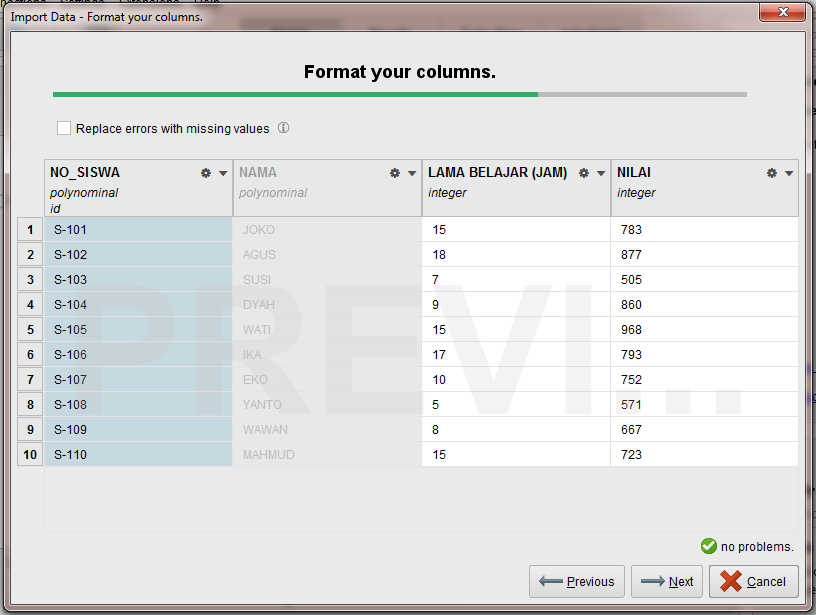


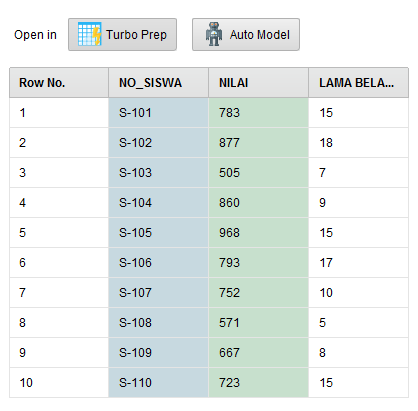
**MODUL 12**

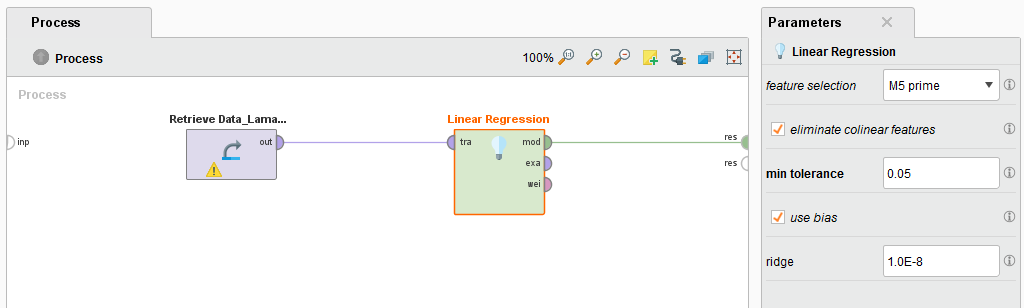
**(Regresi Linier Sederhana)**

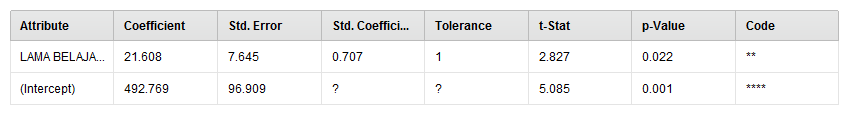
**KEGIATAN**

Kegiatan 1

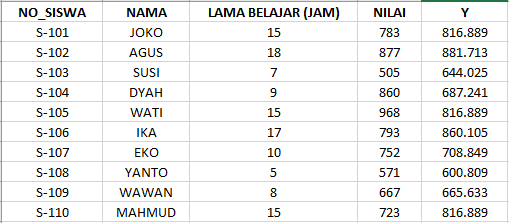




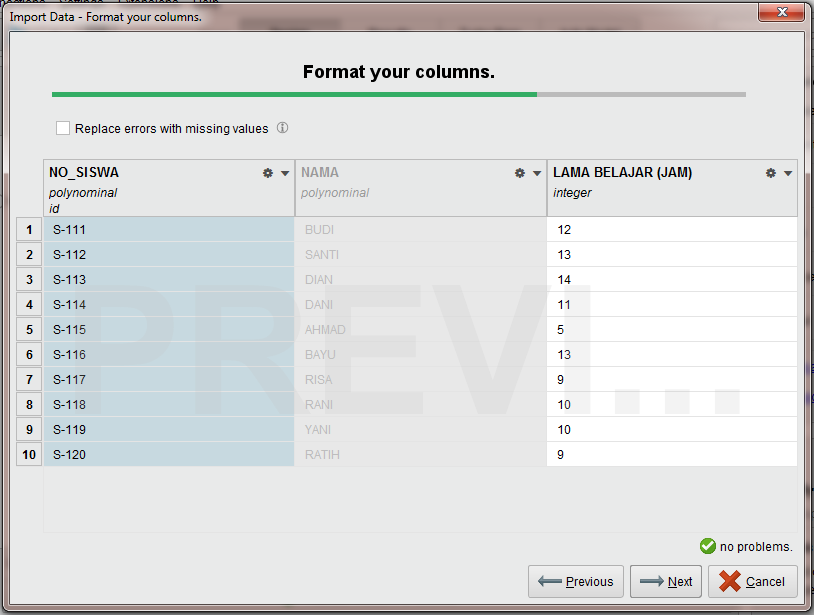


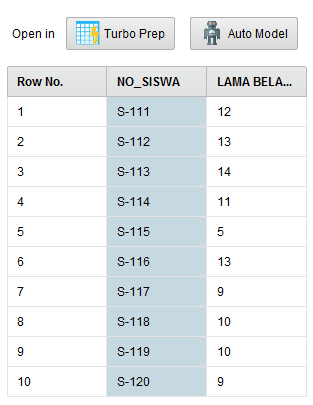


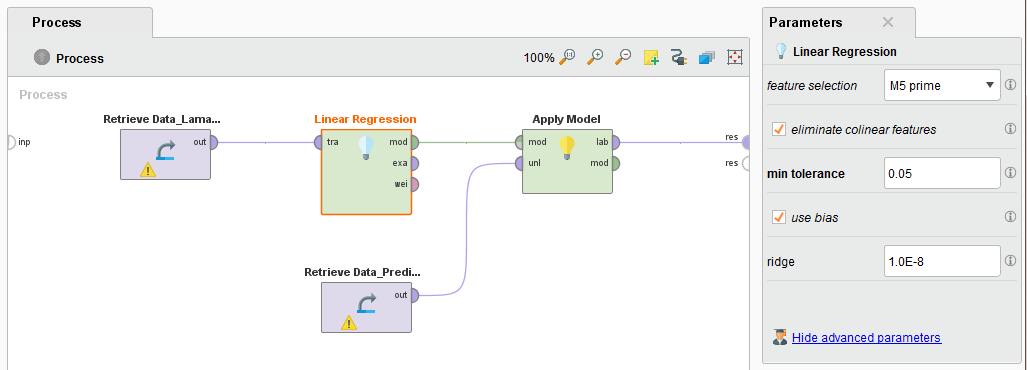


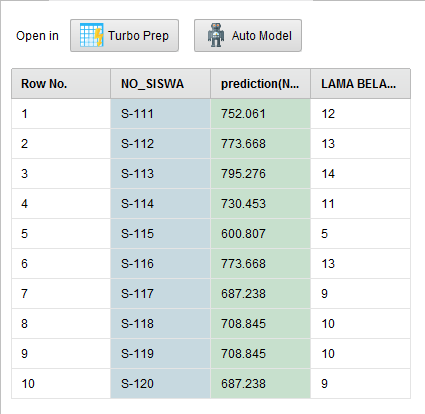


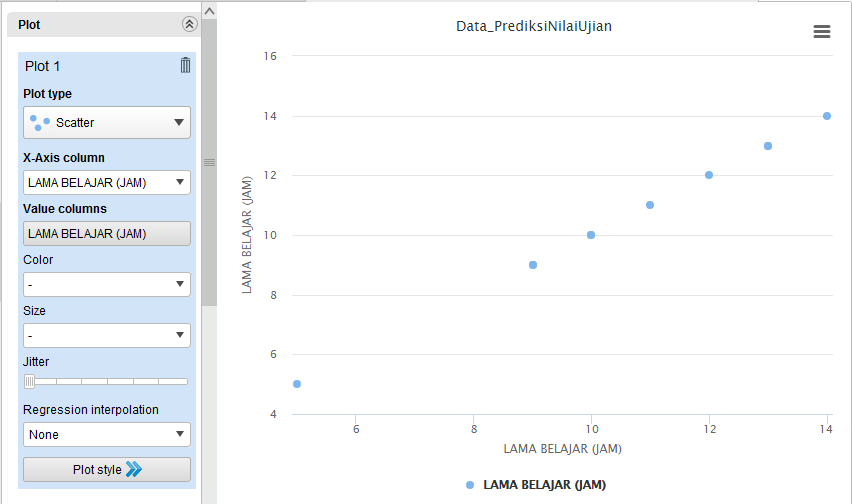
Kegiatan 2

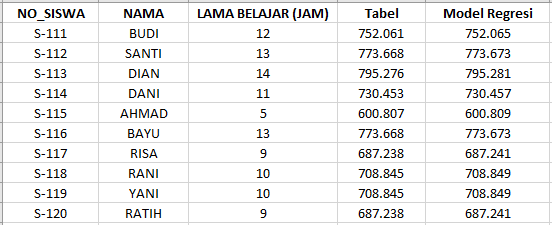










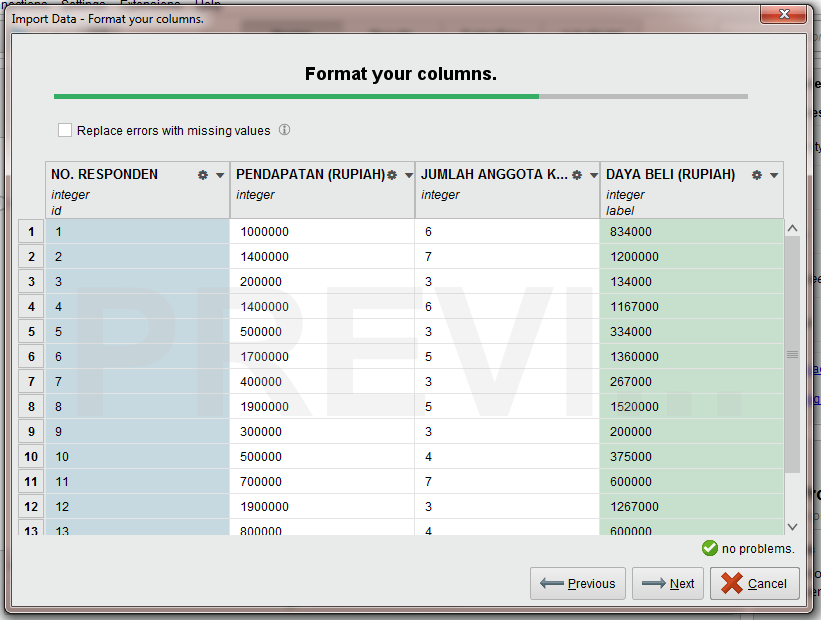


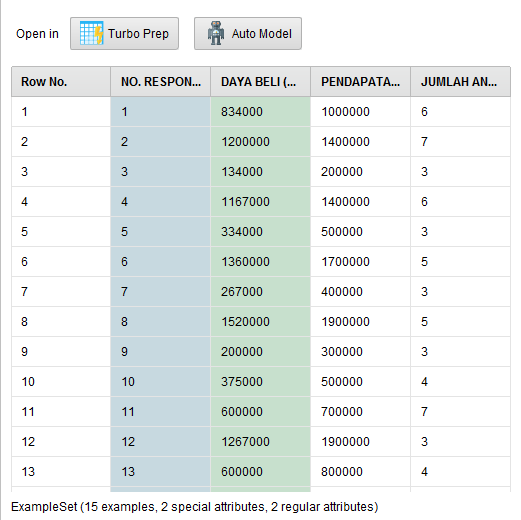
**MODUL 12**

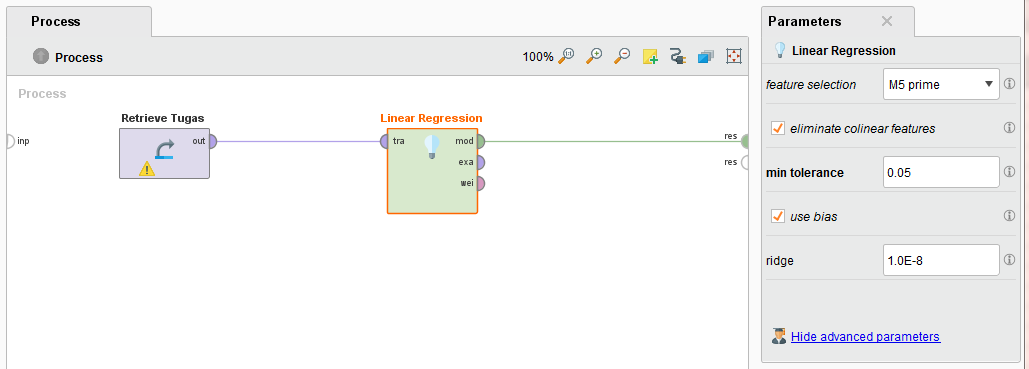
**(Regresi Linier Sederhana)**

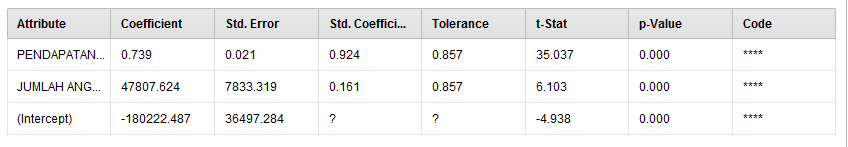
**TUGAS**

Tugas 1













Tugas 2

