

**LAPORAN PRAKTIKUM**  
**DATA WAREHOUSE DAN DATA MINING**



Oleh:

Nama : Nafiah Nurrahmah

NIM : L200170114

Kelas : D

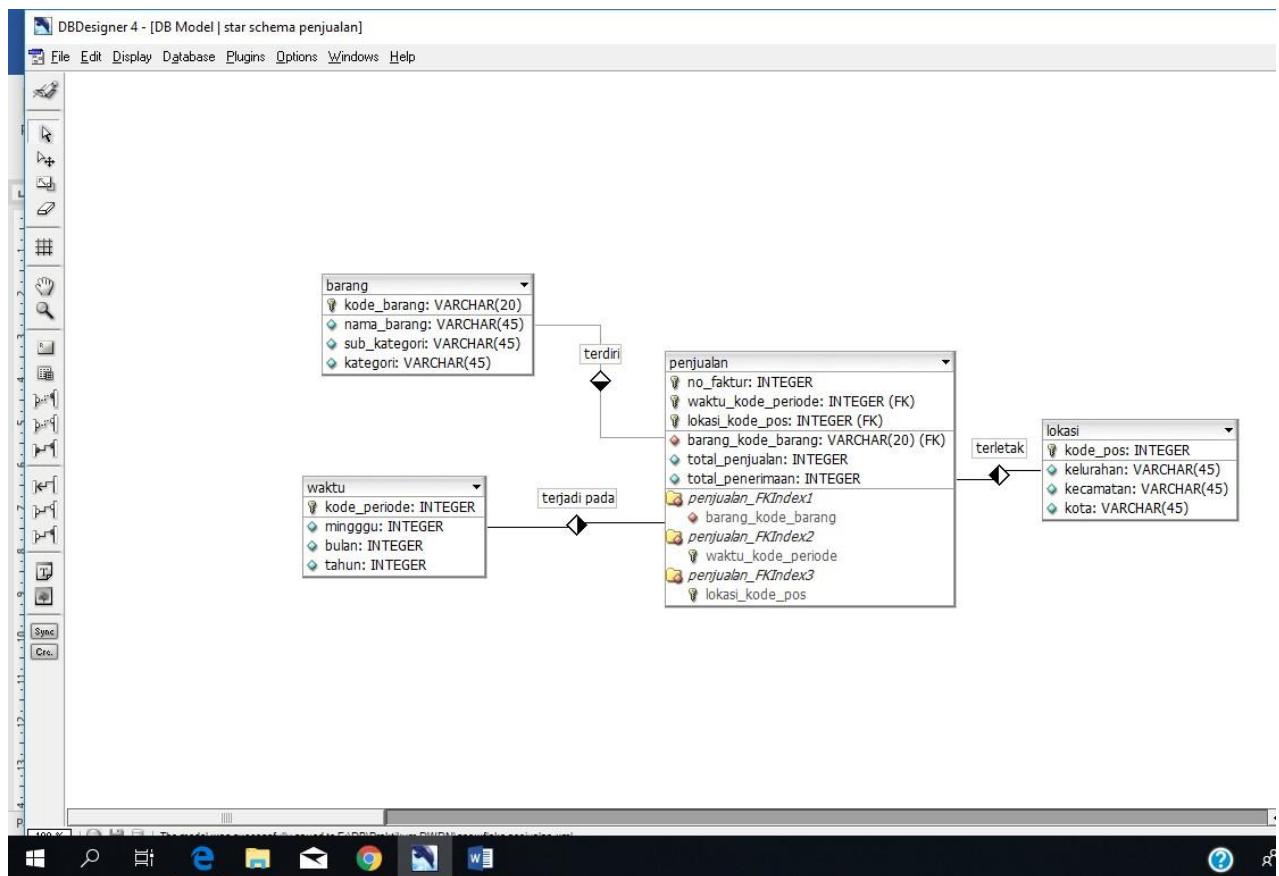
**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS KOMUNIKASI DAN INFORMATIKA**  
**UNIVERSITAS MUHAMMADIYAH SURAKARTA**  
**SURAKARTA**  
**2019**

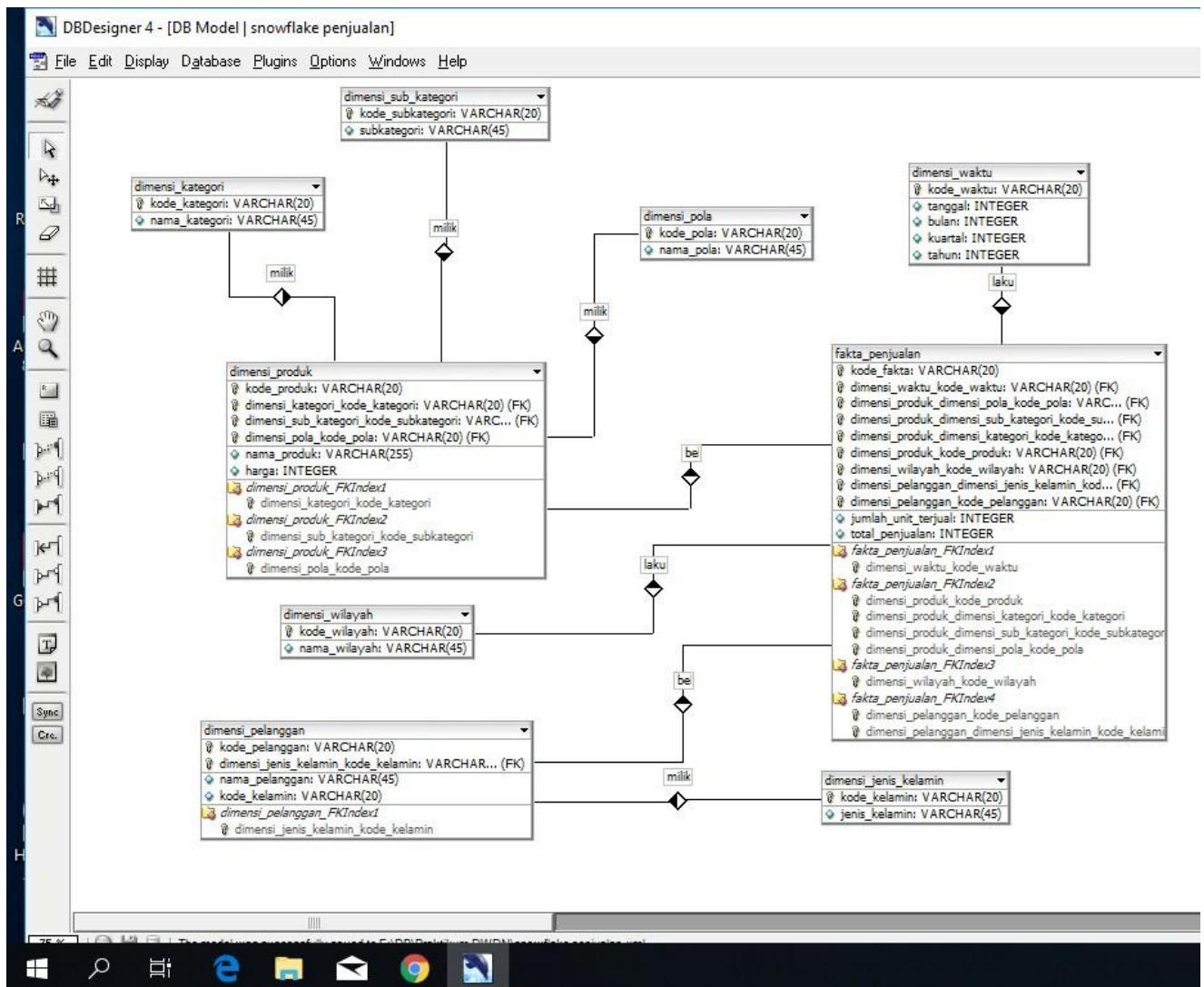
## MODUL 1

### Perancangan Star Schema dan Snowflake

#### Tugas

1. Rancanglah diagram *Snowflake schema* berdasarkan gambar di bawah dengan menggunakan DBDesigner seperti gambar berikut! Simpan dengan nama file "snowflake penjualan.xml" ke dalam folder "Praktikum Data Warehouse"



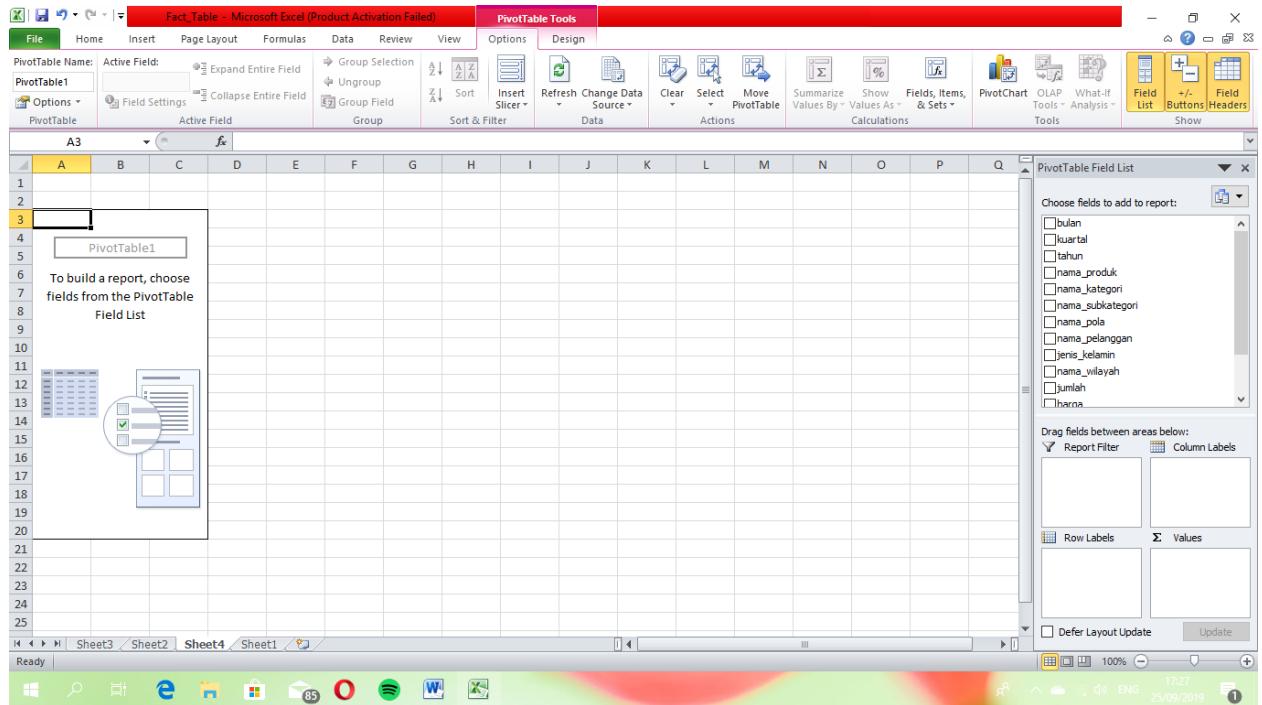


## MODUL 5

### Pivot Table dan Chart

Kegiatan 1 : Membuat Pivot Table

1. Buka file yang akan dibuat pivot tablenya
2. Tekan tombol CTRL+SHIFT+\*
3. Klik tab Insert, pilih menu Pivot Table|Insert Pivot Table
4. Pilih New Worksheet, klik OK



5. Susunlah layout field dengan urutan:
  - a. Field nama\_kategori ke kotak Row Labels
  - b. Field tahun ke kotak Column Labels
  - c. Field jumlah ke kotak Values

## Kegiatan 2 : Menambahkan Tipe Summary Baru

1. Tambahkan field jumlah kembali ke kotak Value, sehingga muncul field baru dengan nama Sum of jumlah 2
2. Klik tombol panah ke bawah pada field Sum of jumlah2, pilih Value Field Settings
3. Ubah Sum menjadi Count, klik OK

## Kegiatan 3 : Calculated Field

1. Letakkan kursor pada area Pivot Table. Pada menu Pivot Table Tools|Options klik button Formulas dan pilih Calculated Field

2. Pada kotak dialog Calculated Field masukkan nilai berikut kemudian klik OK  
Nama : Pendapatan  
Formula: jumlah\*harga
  3. Field baru Sum of Pendapatan akan muncul pada Pivot Table

## Kegiatan 4 : Operasi Roll Up dan Drill Down

1. Beri tanda cek dan letakkan field nama\_subkategori dan nama\_produk ke Row Tables
  2. Beri tanda cek dan letakkan field kuartal dan bulan ke Column Tables

The screenshot shows a Microsoft Excel window with the title "Fact\_Table - Microsoft Excel (Product Activation Failed)". The ribbon is visible at the top with tabs like File, Home, Insert, Page Layout, Formulas, Data, Review, View, Options, and Design. The "PivotTable Tools" tab is selected, showing various options for managing the PivotTable structure.

The PivotTable is named "PivotTable1" and has "PivotTable" selected as the Active Field. The Active Field is currently set to "nama\_kategori".

The PivotTable displays the following data:

|                   |          | 2010     | 2011      | 2012      | Grand Total |
|-------------------|----------|----------|-----------|-----------|-------------|
| Sum of Pendapatan |          |          |           |           |             |
| Bahan             | 500000   | 960000   | 2120000   | 15045000  |             |
| Batik             | 0        | 0        | 150000    | 150000    |             |
| Bolero            | 0        | 225000   | 0         | 225000    |             |
| Celana            | 935000   | 0        | 935000    | 3740000   |             |
| Hem               | 500000   | 4960000  | 12360000  | 64770000  |             |
| Jam               | 0        | 0        | 3520000   | 3520000   |             |
| Jarik             | 0        | 450000   | 16000     | 1374000   |             |
| Kaos              | 0        | 60000    | 420000    | 1350000   |             |
| Rok               | 0        | 0        | 225000    | 225000    |             |
| Sarimbit          | 0        | 150000   | 0         | 150000    |             |
| Grand Total       | 15065000 | 29400000 | 362607000 | 815698000 |             |

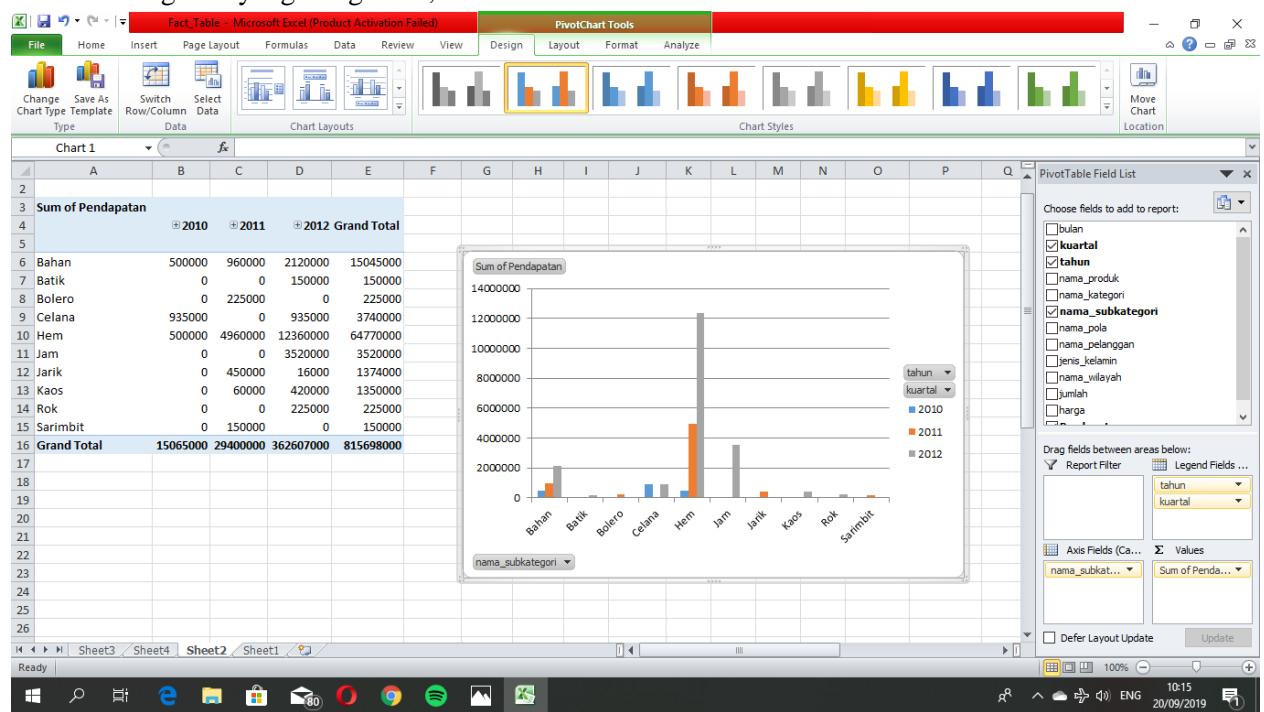
The PivotTable Field List pane on the right side lists fields available for report items:

- checkbox  **kuartal**
- checkbox  **tahun**
- checkbox  **nama\_produk**
- checkbox  **nama\_kategori**
- checkbox  **nama\_subkategori**
- checkbox  **nama\_pola**
- checkbox  **nama\_pelanggan**
- checkbox  **jenis\_kelamin**
- checkbox  **nama\_wilayah**
- checkbox  **jumlah**
- checkbox  **harga**
- checkbox  **Pendapatan**

Below the field list, there are sections for Report Filter, Column Labels, Row Labels, and Values, with "tahun" and "bulan" selected for Column Labels, and "nama\_kategori", "nama\_subkateg...", and "nama\_produk" selected for Row Labels. The Values section shows "Sum of Pendapatan".

## Kegiatan 5 : Pivot Chart

1. Arahkan kursor pada area pivot table
2. Pada menu Option, klik Pivot Chart
3. Pilih bentuk grafik yang diinginkan, klik OK



## Tugas

1. Tambahkan 2 buah field yaitu:
  - a. PPN

Screenshot of Microsoft Excel showing the creation of a calculated field in a PivotTable.

The PivotTable structure is as follows:

|               |                  |                   |            |                          | Total Sum of jumlah | Total Count of jumlah2 | Total Sum of Pendapatan | Total Sum of PPN |
|---------------|------------------|-------------------|------------|--------------------------|---------------------|------------------------|-------------------------|------------------|
| Sum of jumlah | Count of jumlah2 | Sum of Pendapatan | Sum of PPN | Sum of Total Penghasilan |                     |                        |                         |                  |
| 8             | 2                | 2120000           | 212000     | 1908000                  | 17                  | 4                      | 15045000                |                  |
| 1             | 1                | 150000            | 15000      | 135000                   | 1                   | 1                      | 150000                  |                  |
|               |                  | 0                 | 0          | 0                        | 1                   | 1                      | 225000                  |                  |
| 17            | 1                |                   |            |                          | 34                  | 2                      | 3740000                 |                  |
| 4             | 2                |                   |            |                          | 17                  | 5                      | 64770000                |                  |
| 44            | 1                |                   |            |                          | 44                  | 1                      | 3520000                 |                  |
| 4             | 1                |                   |            |                          | 6                   | 2                      | 1374000                 |                  |
| 14            | 1                |                   |            |                          | 15                  | 2                      | 1350000                 |                  |
| 1             | 1                |                   |            |                          | 1                   | 1                      | 225000                  |                  |
|               |                  |                   |            |                          | 1                   | 1                      | 150000                  |                  |
|               |                  |                   |            |                          | 137                 | 20                     | 815698000               |                  |

The "Insert Calculated Field" dialog box is open, showing the formula `=10% *Pendapatan`.

### b. Total penghasilan

Screenshot of Microsoft Excel showing the creation of another calculated field in a PivotTable.

The PivotTable structure is identical to the previous one, with the same data and calculated fields.

The "Insert Calculated Field" dialog box is open, showing the formula `=Pendapatan +PPN`.

## 2. Buatlah pivot table dan pivot chart

Fact\_Table - Microsoft Excel (Product Activation Failed)

PivotTable Tools

PivotTable Name: Active Field: Count of jumlah2 Options Design

PivotTable1 Active Field Group Selection Group Ungroup Group Field Group Sort & Filter Refresh Data Clear Select Move PivotTable Summarize Values By Show Fields, Items, & Sets Calculations PivotChart OLAP What-If Tools Analysis Field List Buttons Field Headers Tools

H15 fx 7

2

3 Total Sum of jumlah Total Count of jumlah2 Total Sum of Pendapatan Total Sum of PPN Total Sum of Total Penghasilan

4 Sum of PPN Sum of Total Penghasilan

5 212000 1908000 17 4 15045000 1504500 13540500

6 15000 135000 1 1 150000 15000 135000

7 0 0 1 1 225000 22500 202500

8 93500 841500 34 2 3740000 374000 3366000

9 1236000 11124000 17 5 64770000 6477000 58293000

10 352000 3168000 44 1 3520000 352000 3168000

11 1600 14400 6 2 1374000 137400 1236600

12 42000 378000 15 2 1350000 135000 1215000

13 22500 202500 1 1 225000 22500 202500

14 0 0 1 1 150000 15000 135000

15 36260700 326346300 137 20 815698000 81569800 734128200

16

17

18

19

20

21

22

23

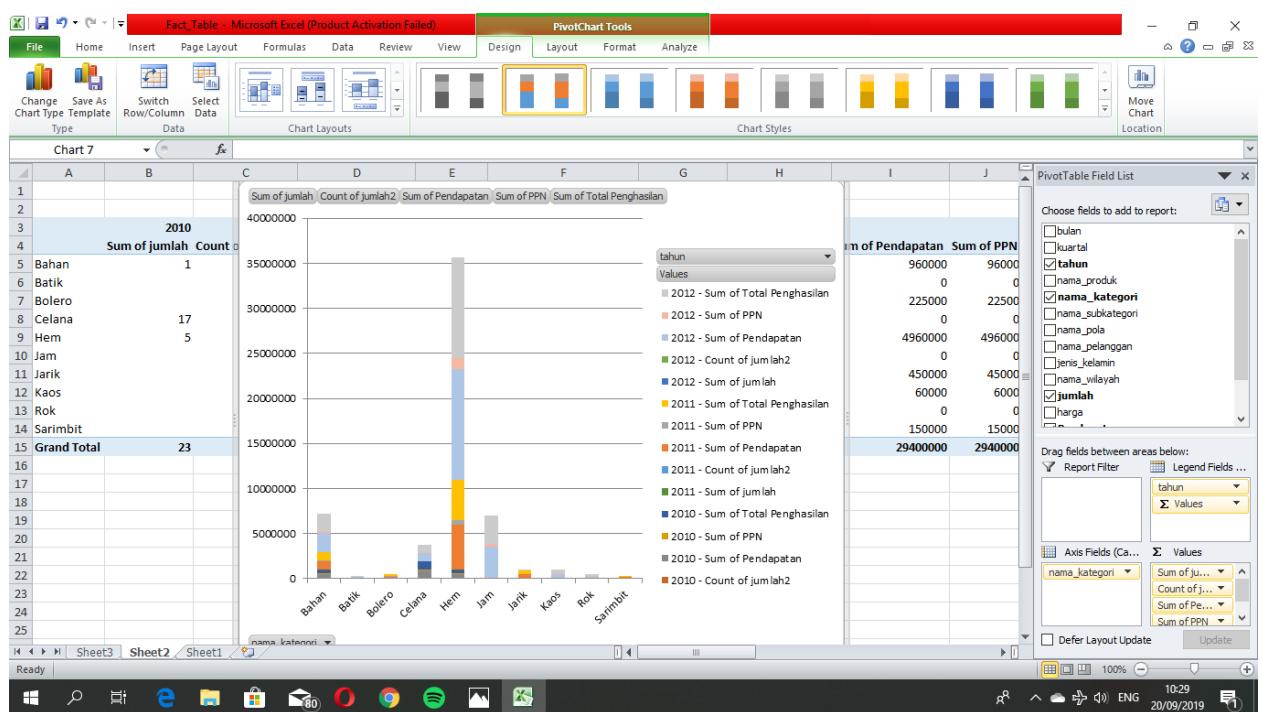
24

25

26

Sheet3 Sheet2 Sheet1 Ready

10:25 20/09/2019



Dari table dan chart diatas, dapat diketahui bahwa Hem yang memberi nilai penghasilan terbanyak selama 3 tahun.

# **MODUL 6**

## **PENGENALAN DATA MINING**

1. Dengan menggunakan formula dalam Ms. Excel, carilah:
    - a) Pada atribut Jurusan\_SMA, berapa jumlah data masing-masing kelas IPA, IPS dan LAIN? (gunakan formula = COUNTIF)
    - b) Pada atribut Lama\_Studi, berapa jumlah data masing-masing kelas TEPAT, TERLAMBAT?
    - c) Pada atribut Rerata\_SKS, berapa nilai Max, Min, Mean, dan *Standard Deviation*?
    - d) Pada tabel tersebut, berapakah jumlah data gabungan untuk kelas pada atribut Jurusan\_SMA = IPA, Gender = PRIA, Asisten = YA, Lama\_studi = TEPAT? (gunakan formula = COUNTIFS)

|    | A           | B      | C            | D          | E       | F          | G  | H               | I | J        | K |
|----|-------------|--------|--------------|------------|---------|------------|----|-----------------|---|----------|---|
| 1  | Jurusan_SMA | Gender | Asal_Sekolah | Rerata_SKS | Asisten | Lama_Studi |    |                 |   |          |   |
| 2  | IPS         | WANITA | SURAKARTA    | 18         | TIDAK   | TERLAMBAT  | a. | SMA IPA         |   | 6        |   |
| 3  | IPA         | PRIA   | SURAKARTA    | 19         | YA      | TEPAT      |    | SMA IPS         |   | 10       |   |
| 4  | LAIN        | PRIA   | SURAKARTA    | 19         | TIDAK   | TERLAMBAT  |    | SMA LAIN        |   | 4        |   |
| 5  | IPA         | PRIA   | LUAR         | 17         | TIDAK   | TERLAMBAT  |    |                 |   |          |   |
| 6  | IPA         | WANITA | SURAKARTA    | 17         | TIDAK   | TEPAT      | b. | TEPAT           |   | 13       |   |
| 7  | IPA         | WANITA | LUAR         | 18         | YA      | TEPAT      |    | TERLAMBAT       |   | 7        |   |
| 8  | IPA         | PRIA   | SURAKARTA    | 18         | TIDAK   | TERLAMBAT  |    |                 |   |          |   |
| 9  | IPA         | PRIA   | SURAKARTA    | 19         | TIDAK   | TEPAT      | c. | Rerata          |   | 18.95    |   |
| 10 | IPS         | PRIA   | LUAR         | 18         | TIDAK   | TERLAMBAT  |    | Max             |   | 23       |   |
| 11 | LAIN        | WANITA | SURAKARTA    | 18         | TIDAK   | TEPAT      |    | Min             |   | 16       |   |
| 12 | IPA         | WANITA | SURAKARTA    | 19         | TIDAK   | TEPAT      |    | Standar Deviasi |   | 1.669384 |   |
| 13 | IPS         | PRIA   | SURAKARTA    | 20         | TIDAK   | TEPAT      |    |                 |   |          |   |
| 14 | IPS         | PRIA   | SURAKARTA    | 19         | TIDAK   | TEPAT      | d. | data gabungan   |   | 3        |   |
| 15 | IPA         | PRIA   | SURAKARTA    | 19         | TIDAK   | TEPAT      |    |                 |   |          |   |
| 16 | IPA         | PRIA   | LUAR         | 22         | YA      | TEPAT      |    |                 |   |          |   |
| 17 | LAIN        | PRIA   | SURAKARTA    | 16         | TIDAK   | TERLAMBAT  |    |                 |   |          |   |
| 18 | IPS         | PRIA   | LUAR         | 20         | TIDAK   | TEPAT      |    |                 |   |          |   |
| 19 | LAIN        | PRIA   | LUAR         | 23         | YA      | TEPAT      |    |                 |   |          |   |
| 20 | IPA         | PRIA   | SURAKARTA    | 21         | YA      | TEPAT      |    |                 |   |          |   |
| 21 | IPS         | PRIA   | SURAKARTA    | 19         | TIDAK   | TERLAMBAT  |    |                 |   |          |   |

## MODUL 7

### DATA PREPROCESSING

#### Percobaan

1. Buat tabel Cuaca

| Cuaca   | Suhu | Kelembapan Udara | Berangin | Bermain_Tenis |
|---------|------|------------------|----------|---------------|
| Cerah   | 85   | 85               | TIDAK    | TIDAK         |
| Cerah   | 80   | 90               | YA       | TIDAK         |
| Mendung | 83   | 86               | TIDAK    | YA            |
| Hujan   | 70   | 96               | TIDAK    | YA            |
| Hujan   | 68   | 80               | TIDAK    | YA            |
| Hujan   | 65   | 70               | YA       | TIDAK         |
| Mendung | 64   | 65               | YA       | YA            |
| Cerah   | 72   | 95               | TIDAK    | TIDAK         |
| Cerah   | 69   | 70               | TIDAK    | YA            |
| Hujan   | 75   | 80               | TIDAK    | YA            |
| Cerah   | 75   | 70               | YA       | YA            |
| Mendung | 72   | 90               | YA       | YA            |
| Mendung | 81   | 75               | TIDAK    | YA            |
| Hujan   | 71   | 91               | YA       | TIDAK         |

2. Buka notepad, dan buat relasi, atribut dan data. Save dengan format Cuaca.arff

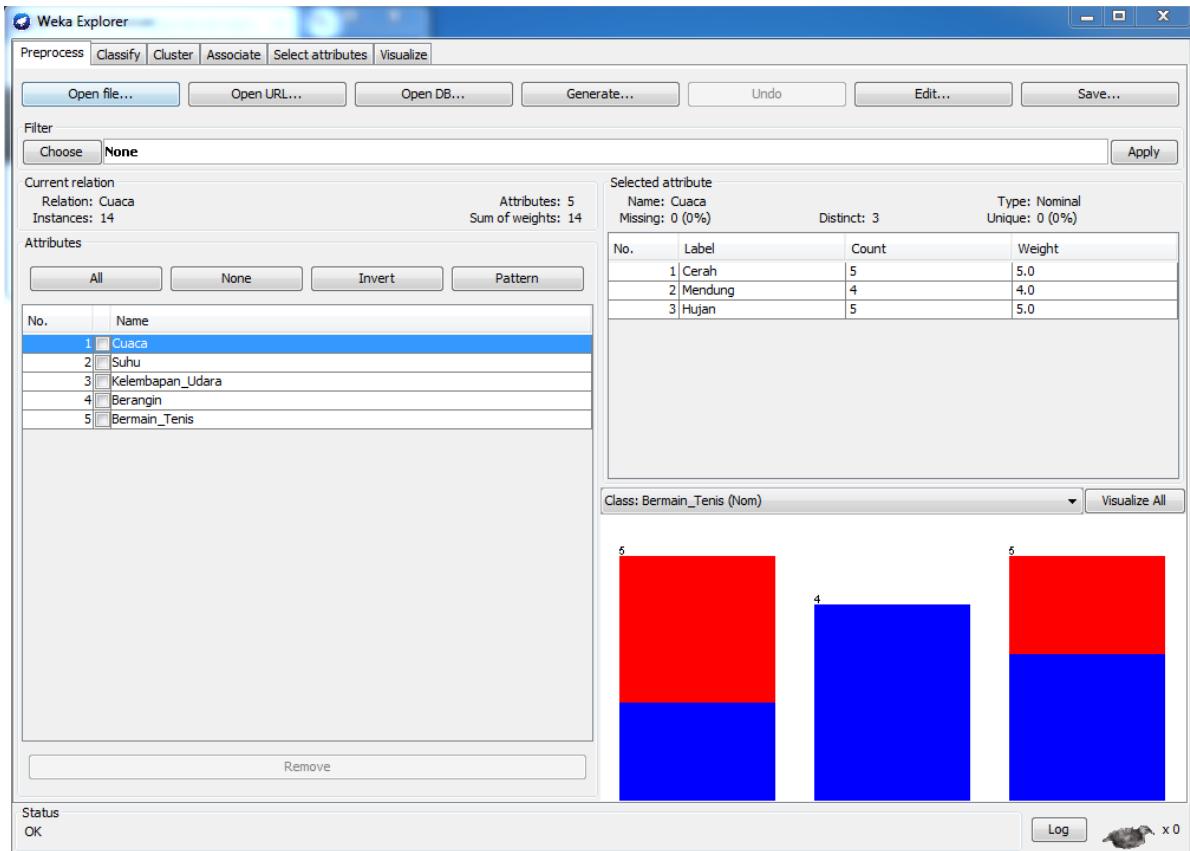
```
relation Cuaca

@attribute Cuaca {Cerah, Mendung, Hujan}
@attribute Suhu real
@attribute Kelembapan_Udara real
@attribute Berangin {YA, TIDAK}
@attribute Bermain_Tenis {YA, TIDAK}

@data
Cerah,85,85,TIDAK,TIDAK
Cerah,80,90,YA,TIDAK
Mendung,83,86,TIDAK,YA
Hujan,70,96,TIDAK,YA
Hujan,68,80,TIDAK,YA
Hujan,65,70,YA,TIDAK
Mendung,64,65,YA,YA
Cerah,72,95,TIDAK,TIDAK
Cerah,69,70,TIDAK,YA
Hujan,75,80,TIDAK,YA
Cerah,75,70,YA,YA
Mendung,72,90,YA,YA
Mendung,81,75,TIDAK,YA
Hujan,71,91,YA,TIDAK
```

3. Buka aplikasi Weka. Pilih menu Explorer, lalu pilih Open file.

Pilih file Cuaca.arff sehingga akan muncul grafik statistik masing-masing atribut pada tabel Cuaca



## TUGAS

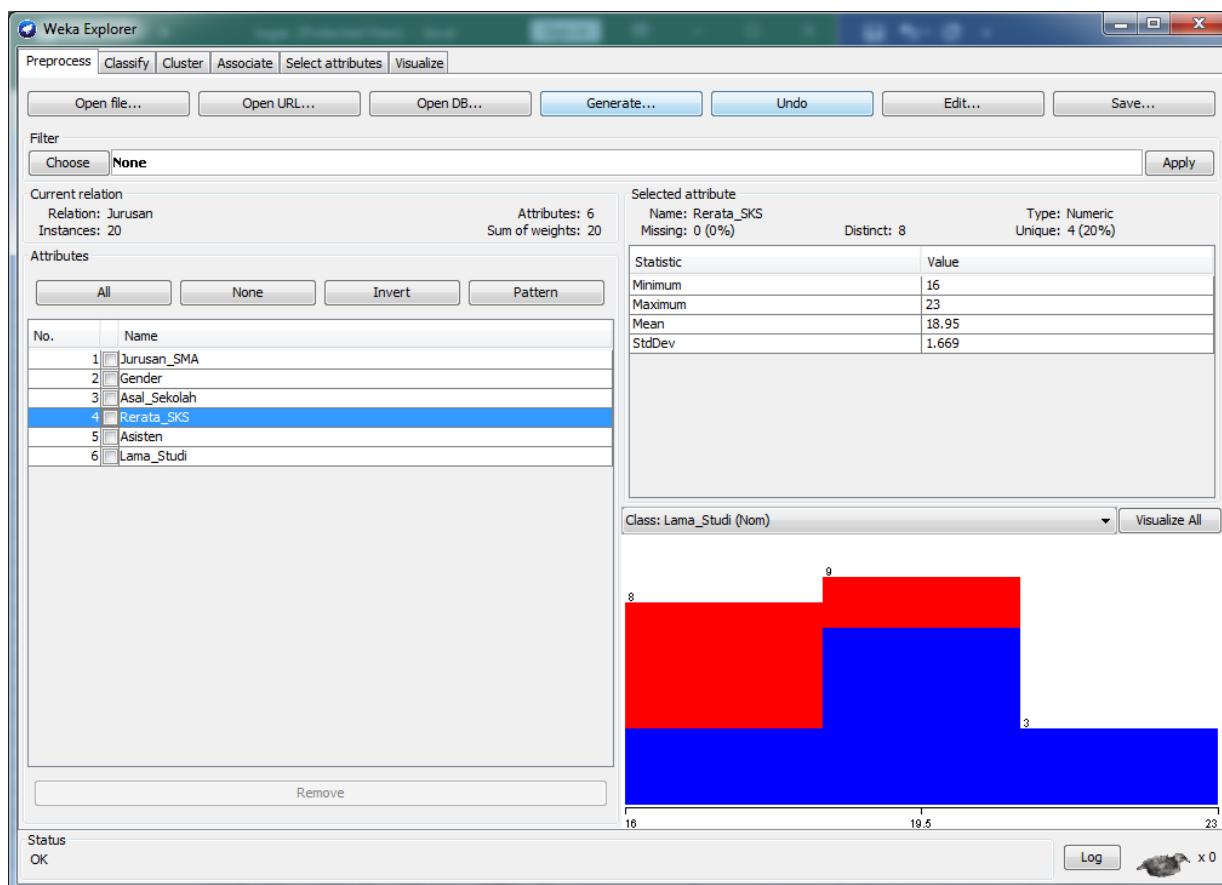
1. Buat file arff berdasarkan tugas pada Modul 6

```
Jurusan - Notepad
File Edit Format View Help
@relation Jurusan

@attribute Jurusan_SMA {IPA, IPS, LAIN}
@attribute Gender {WANITA, PRIA}
@attribute Asal_Sekolah {SURAKARTA, LUAR}
@attribute Rerata_SKS real
@attribute Asisten {YA, TIDAK}
@attribute Lama_Studi {TEPAT, TERLAMBAT}

@data
IPS,WANITA,SURAKARTA,18,TIDAK,TERLAMBAT
IPA,PRIA,SURAKARTA,19,YA,TEPAT
LAIN,PRIA,SURAKARTA,19,TIDAK,TERLAMBAT
IPA,PRIA,LUAR,17,TIDAK,TERLAMBAT
IPA,WANITA,SURAKARTA,17,TIDAK,TEPAT
IPA,WANITA,LUAR,18,YA,TEPAT
IPA,PRIA,SURAKARTA,18,TIDAK,TERLAMBAT
IPA,PRIA,SURAKARTA,19,TIDAK,TEPAT
IPS,PRIA,LUAR,18,TIDAK,TERLAMBAT
LAIN,WANITA,SURAKARTA,18,TIDAK,TEPAT
IPA,WANITA,SURAKARTA,19,TIDAK,TEPAT
IPS,PRIA,SURAKARTA,20,TIDAK,TEPAT
IPS,PRIA,SURAKARTA,19,TIDAK,TEPAT
IPA,PRIA,SURAKARTA,19,TIDAK,TEPAT
IPA,PRIA,LUAR,22,YA,TEPAT
LAIN,PRIA,SURAKARTA,16,TIDAK,TERLAMBAT
IPS,PRIA,LUAR,20,TIDAK,TEPAT
LAIN,PRIA,LUAR,23,YA,TEPAT
IPA,PRIA,SURAKARTA,21,YA,TEPAT
IPS,PRIA,SURAKARTA,19,TIDAK,TERLAMBAT
```

2. Buka file tersebut melalui Weka



3.

- Berapa jumlah atribut yang bertipe binomial?  
Ada 4 atribut yang bertipe binomial yaitu Gender, Asal\_Sekolah, Asisten dan Lama\_Studi.
- Berapa jumlah atribut yang bertipe polynomial?  
Ada 1 atribut yang bertipe polynomial yaitu Jurusan\_SMA.

4. Berapa jumlah atribut yang bertipe real?

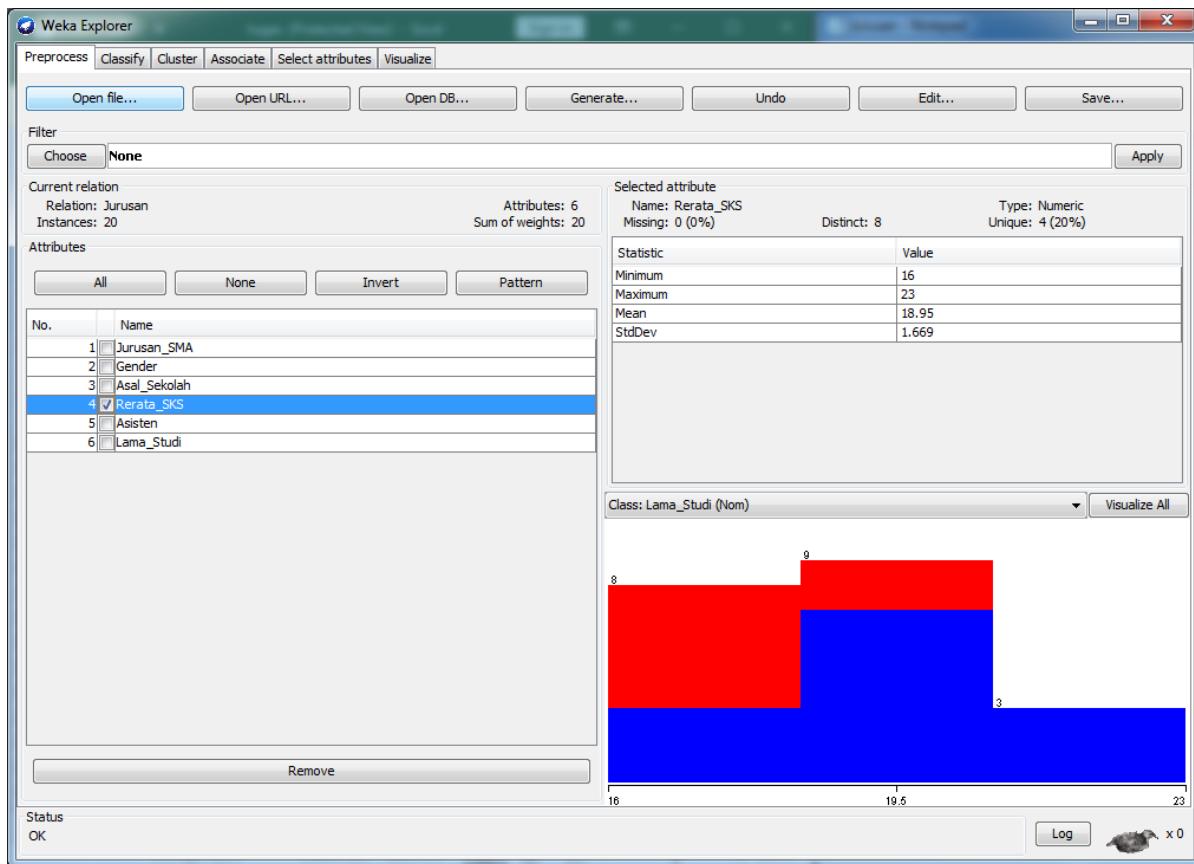
Ada 1 atribut yang bertipe real yaitu Rerata\_SKS

5. Pada atribut Rerata\_SKS, berapakah nilai Maximum, Minimum, Mean dan StdDev? Maximum = 23

Minimum = 16

Mean = 18.95

StdDev = 1.669



## MODUL 8

### NAÏVE BAYES

#### PERCOBAAN

```
cuaca.arff × cuacaTesting.arff
d: > KAMPUS > SMT 5 > DWDM > modul 8 > cuaca.arff
1 @relation Cuaca
2
3 @attribute Cuaca {Cerah, Mendung, Hujan}
4 @attribute Suhu real
5 @attribute Kelembapan_Udara real
6 @attribute Berangin {YA, TIDAK}
7 @attribute Bermain_Tenis {YA, TIDAK}
8
9 @data
10 Cerah,85,85,TIDAK,TIDAK
11 Cerah,80,90,YA,TIDAK
12 Mendung,83,86,TIDAK,YA
13 Hujan,70,96,TIDAK,YA
14 Hujan,68,80,TIDAK,YA
15 Hujan,65,70,YA,TIDAK
16 Mendung,64,65,YA,YA
17 Cerah,72,95,TIDAK,TIDAK
18 Cerah,69,70,TIDAK,YA
19 Hujan,75,80,TIDAK,YA
20 Cerah,75,70,YA,YA
21 Mendung,72,90,YA,YA
22 Mendung,81,75,TIDAK,YA
23 Hujan,71,91,YA,TIDAK
24
```

```

≡ cuaca.arff      ≡ cuacaTesting.arff ×

d: > KAMPUS > SMT 5 > DWDM > modul 8 > ≡ cuacaTesting.arff

1  @relation Cuaca
2
3  @attribute Cuaca {Cerah, Mendung, Hujan}
4  @attribute Suhu real
5  @attribute Kelembapan_Udara real
6  @attribute Berangin {YA, TIDAK}
7  @attribute Bermain_Tenis {YA, TIDAK}
8
9  @data
10 Cerah,75,65,TIDAK,?
11 Cerah,80,68,YA,?
12 Cerah,83,87,YA,?
13 Mendung,70,96,TIDAK,?
14 Mendung,68,81,TIDAK,?
15 Hujan,65,75,YA,?
16 Hujan,64,85,YA,?

```

Hasil prediksi dari Weka

ARFF-Viewer - D:\KAMPUS\SMT 5\DWDM\modul 8\HasilPrediksi.arff

File Edit View

HasilPrediksi.arff

Relation: Cuaca\_predicted

No. 1: Cuaca 2: Suhu 3: Kelembapan\_Udara 4: Berangin 5: prediction margin 6: predicted Bermain\_Tenis 7: Bermain\_Tenis

|   | 1: Cuaca | 2: Suhu | 3: Kelembapan_Udara | 4: Berangin | 5: prediction margin | 6: predicted Bermain_Tenis | 7: Bermain_Tenis |
|---|----------|---------|---------------------|-------------|----------------------|----------------------------|------------------|
|   | Nominal  | Numeric | Numeric             | Nominal     | Numeric              | Nominal                    | Nominal          |
| 1 | Cerah    | 75.0    | 65.0                | TIDAK       | 0.762765             | YA                         |                  |
| 2 | Cerah    | 80.0    | 68.0                | YA          | 0.087878             | YA                         |                  |
| 3 | Cerah    | 83.0    | 87.0                | YA          | -0.676866            | TIDAK                      |                  |
| 4 | Mend...  | 70.0    | 96.0                | TIDAK       | 0.628523             | YA                         |                  |
| 5 | Mend...  | 68.0    | 81.0                | TIDAK       | 0.833996             | YA                         |                  |
| 6 | Hujan    | 65.0    | 75.0                | YA          | 0.253733             | YA                         |                  |
| 7 | Hujan    | 64.0    | 85.0                | YA          | -0.160143            | TIDAK                      |                  |

## Hasil prediksi dari Rapid Miner

 <new process\*> – RapidMiner 5.3.000 @ DESKTOP-CK4GJGD

File Edit Process Tools View Help

ExampleSet (/Local Repository/DataCuaca\_Testing) X ExampleSet (/Local Repository/D  
Result Overview X ExampleSet (Retrieve DataCuaca\_Testing)

Data View Meta Data View Plot View Advanced Charts Annotations

ExampleSet (7 examples, 3 special attributes, 4 regular attributes) View File

| Row No. | confidence(... confidence(... | confidence(... | prediction(Bermain_Tenis) | Cuaca   | Suhu | Kelembapa... | Berangin |
|---------|-------------------------------|----------------|---------------------------|---------|------|--------------|----------|
| 1       | 0.154                         | 0.846          | YA                        | Cerah   | 75   | 65           | TIDAK    |
| 2       | 0.498                         | 0.502          | YA                        | Cerah   | 80   | 68           | YA       |
| 3       | 0.856                         | 0.144          | TIDAK                     | Cerah   | 83   | 87           | YA       |
| 4       | 0.019                         | 0.981          | YA                        | Mendung | 70   | 96           | TIDAK    |
| 5       | 0.007                         | 0.993          | YA                        | Mendung | 68   | 81           | TIDAK    |
| 6       | 0.371                         | 0.629          | YA                        | Hujan   | 65   | 75           | YA       |
| 7       | 0.568                         | 0.432          | TIDAK                     | Hujan   | 64   | 85           | YA       |

## TUGAS

Hasil prediksi Jurusan melalui Weka

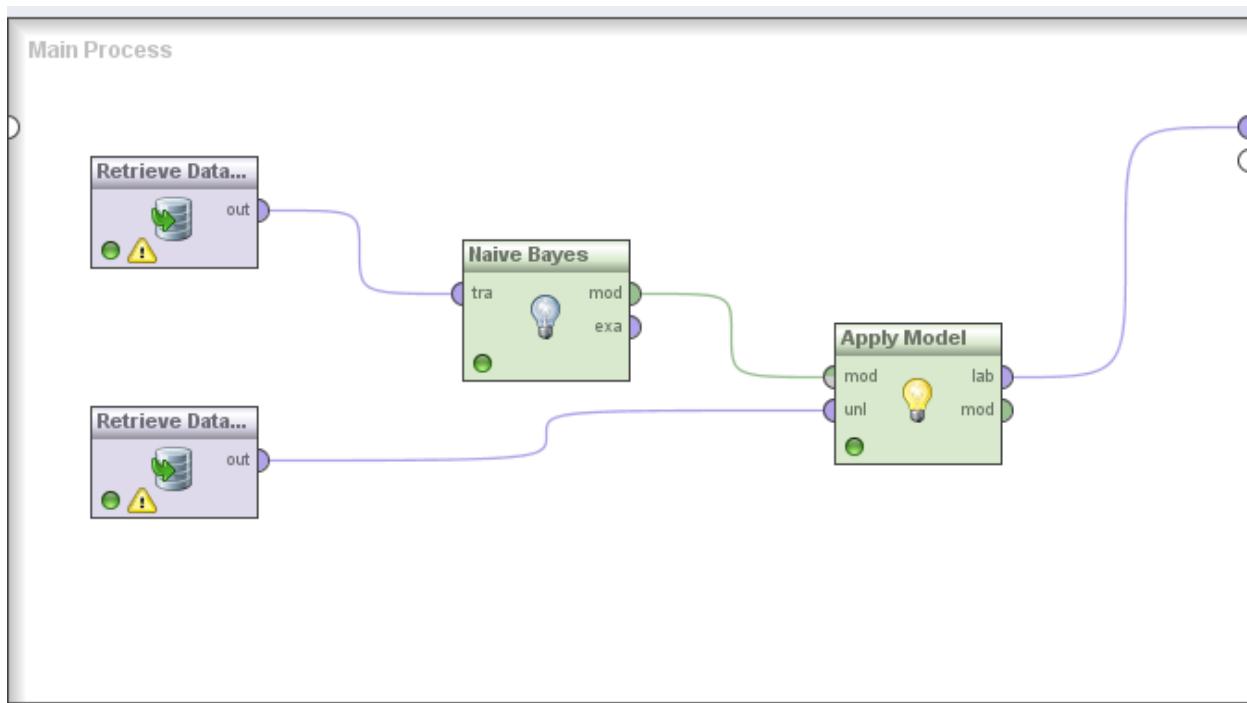
ARFF-Viewer - D:\KAMPUS\SMT 5\DWDM\modul 8\hasilPrediksiJurusan.arff

File Edit View

hasilPrediksiJurusan.arff

Relation: Jurusan\_predicted

| No. | 1: Jurusan_SMA | 2: Gender | 3: Asal_Sekolah | 4: Rerata_SKS | 5: Asisten | 6: prediction margin | 7: predicted Lama_Studi | 8: Lama_Studi |
|-----|----------------|-----------|-----------------|---------------|------------|----------------------|-------------------------|---------------|
|     | Nominal        | Nominal   | Nominal         | Numeric       | Nominal    | Numeric              | Nominal                 | Nominal       |
| 1   | LAIN           | WANITA    | SURAKARTA       | 18.0          | TIDAK      | -0.375862            | TERLAMBAT               |               |
| 2   | IPA            | PRIA      | SURAKARTA       | 19.0          | YA         | 0.836469             | TEPAT                   |               |
| 3   | LAIN           | PRIA      | SURAKARTA       | 19.0          | TIDAK      | -0.175169            | TERLAMBAT               |               |
| 4   | IPS            | PRIA      | LUAR            | 17.0          | TIDAK      | -0.713206            | TERLAMBAT               |               |
| 5   | LAIN           | WANITA    | SURAKARTA       | 17.0          | TIDAK      | -0.546846            | TERLAMBAT               |               |
| 6   | IPA            | WANITA    | LUAR            | 18.0          | YA         | 0.757815             | TEPAT                   |               |
| 7   | IPA            | PRIA      | SURAKARTA       | 18.0          | TIDAK      | -0.125076            | TERLAMBAT               |               |
| 8   | IPA            | PRIA      | SURAKARTA       | 19.0          | TIDAK      | 0.356012             | TEPAT                   |               |
| 9   | IPS            | PRIA      | LUAR            | 18.0          | TIDAK      | -0.588286            | TERLAMBAT               |               |
| 10  | LAIN           | WANITA    | SURAKARTA       | 18.0          | TIDAK      | -0.375862            | TERLAMBAT               |               |



Hasil prediksi jurusan melalui Rapid Miner

| Row No. | confidence(...) | confidence(...) | prediction(Lama_Studi) | Jurusan_S... | Gender | Asal_Sekol... | Rerata_SKS | Asisten | Lama_Studi |
|---------|-----------------|-----------------|------------------------|--------------|--------|---------------|------------|---------|------------|
| 1       | 0.648           | 0.352           | TERLAMBAT              | LAIN         | WANITA | SURAKARTA     | 18         | TIDAK   | ?          |
| 2       | 0.005           | 0.995           | TEPAT                  | IPA          | PRIA   | SURAKARTA     | 19         | YA      | ?          |
| 3       | 0.650           | 0.350           | TERLAMBAT              | LAIN         | PRIA   | SURAKARTA     | 19         | TIDAK   | ?          |
| 4       | 0.868           | 0.132           | TERLAMBAT              | IPS          | PRIA   | LUAR          | 17         | TIDAK   | ?          |
| 5       | 0.738           | 0.262           | TERLAMBAT              | LAIN         | WANITA | SURAKARTA     | 17         | TIDAK   | ?          |
| 6       | 0.005           | 0.995           | TEPAT                  | IPA          | WANITA | LUAR          | 18         | YA      | ?          |
| 7       | 0.547           | 0.453           | TERLAMBAT              | IPA          | PRIA   | SURAKARTA     | 18         | TIDAK   | ?          |
| 8       | 0.321           | 0.679           | TEPAT                  | IPA          | PRIA   | SURAKARTA     | 19         | TIDAK   | ?          |
| 9       | 0.811           | 0.189           | TERLAMBAT              | IPS          | PRIA   | LUAR          | 18         | TIDAK   | ?          |
| 10      | 0.648           | 0.352           | TERLAMBAT              | LAIN         | WANITA | SURAKARTA     | 18         | TIDAK   | ?          |

Nilai rata-rata confidence

|                      |                       |      |                       |                 |
|----------------------|-----------------------|------|-----------------------|-----------------|
| confidence_TERLAMBAT | confidence(TERLAMBAT) | real | avg = 0.524 +/- 0.312 | [0.005 ; 0.868] |
| confidence_TEPAT     | confidence(TEPAT)     | real | avg = 0.476 +/- 0.312 | [0.132 ; 0.995] |

Dari data berikut diperoleh 7 orang akan lulus TEPAT dan 3 orang akan lulus TERLAMBAT

|            |                        |          |   |                          |
|------------|------------------------|----------|---|--------------------------|
| prediction | prediction(Lama_Studi) | binomial | mode = TERLAMBAT (7), least = TEPAT (3) | TERLAMBAT (7), TEPAT (3) |
|------------|------------------------|----------|---|--------------------------|

Hasil data Dewi

| Row No. | confidence(...) | confidence(...) | prediction(Lama_Studi) | Jurusan_S... | Gender | Asal_Sekol... | Rerata_SKS | Asisten | Lama_Studi |
|---------|-----------------|-----------------|------------------------|--------------|--------|---------------|------------|---------|------------|
| 1       | 0.298           | 0.702           | TEPAT                  | IPA          | WANITA | LUAR          | 18         | TIDAK   | ?          |

## Hasil data Joni

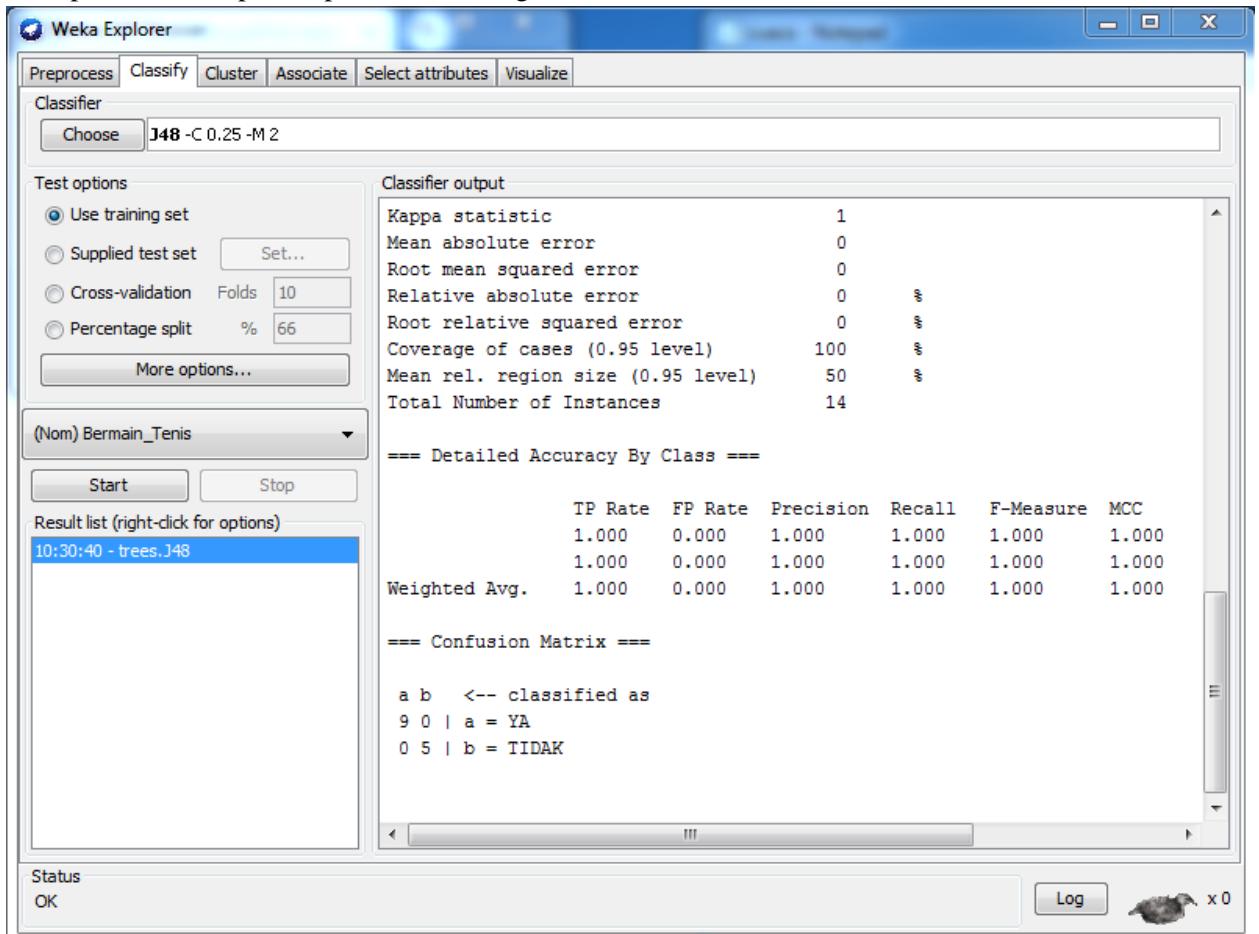
| Row No. | confidence(...confidence(...prediction(L... | Jurusan_S... | Gender | Asal_Sekol... | Rerata_SKS | Asisten   | Lama_Studi |
|---------|---|--------------|--------|---------------|------------|-----------|------------|
| 1       | 0.076                                       | 0.924        | TEPAT  | LAIN          | PRIA       | SURAKARTA | 17         |

## MODUL 9

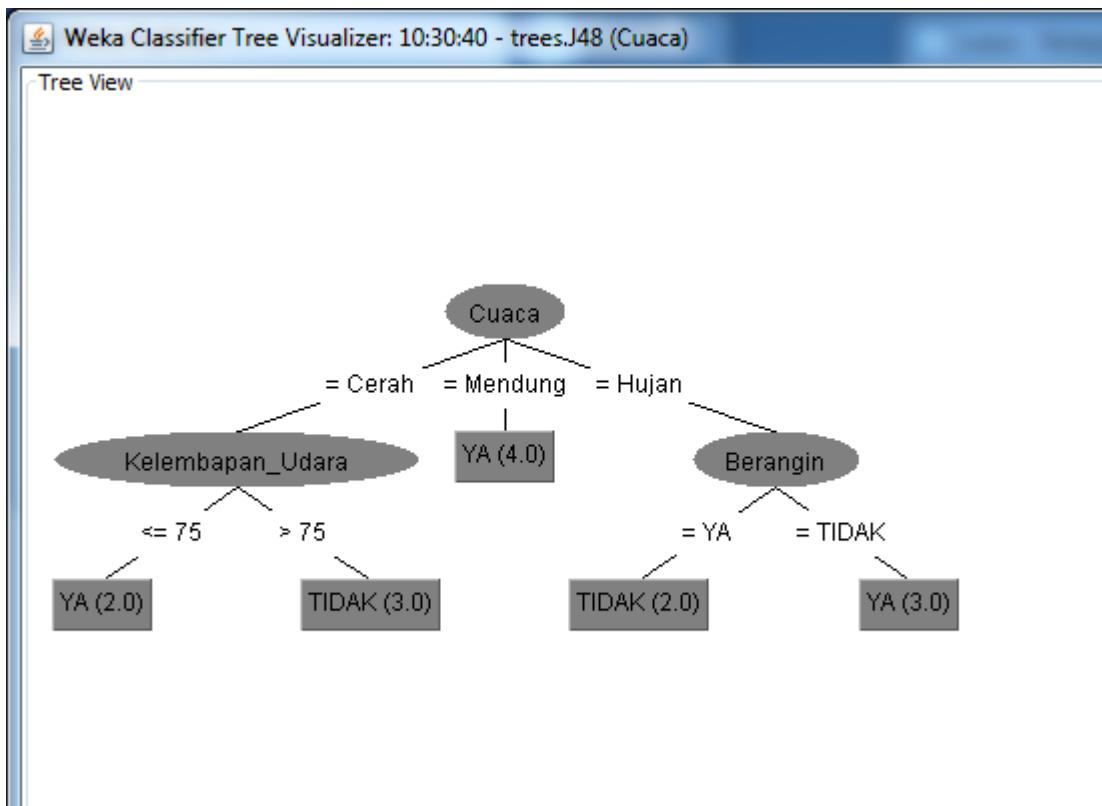
### DECISION TREE

#### 9.4.1 Pohon Keputusan Menggunakan Weka Explorer

1. Buka file Cuaca.arff dari Weka Explorer
2. Buka tab Classify, tekan tombol Choose-Trees-J48
3. Pada pilihan Test Options, pilih Use training set.

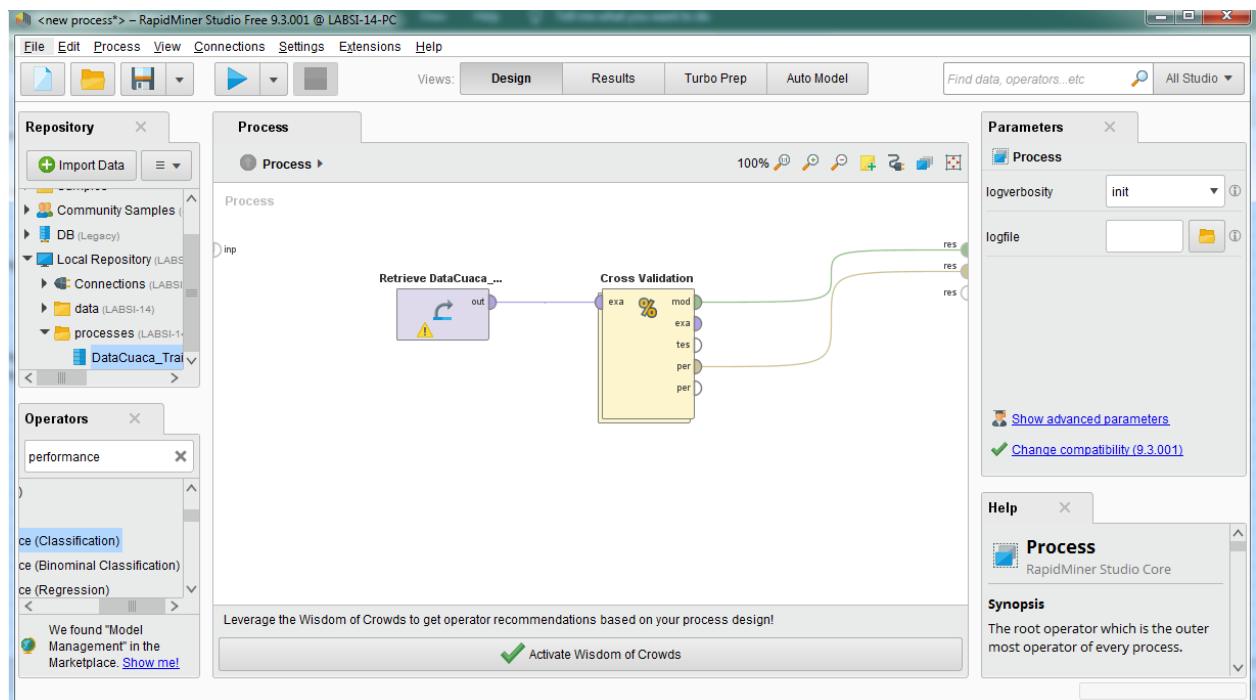


4. Klik kanan pada hasil trees.J48 – Visualize tree

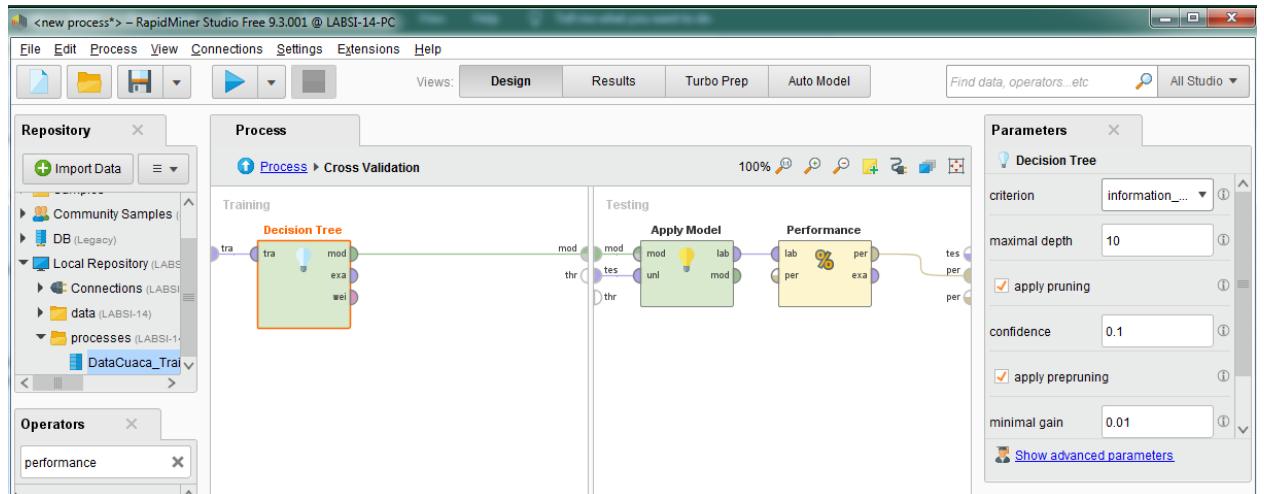


#### 9.4.2 Pohon Keputusan Menggunakan Rapid Miner

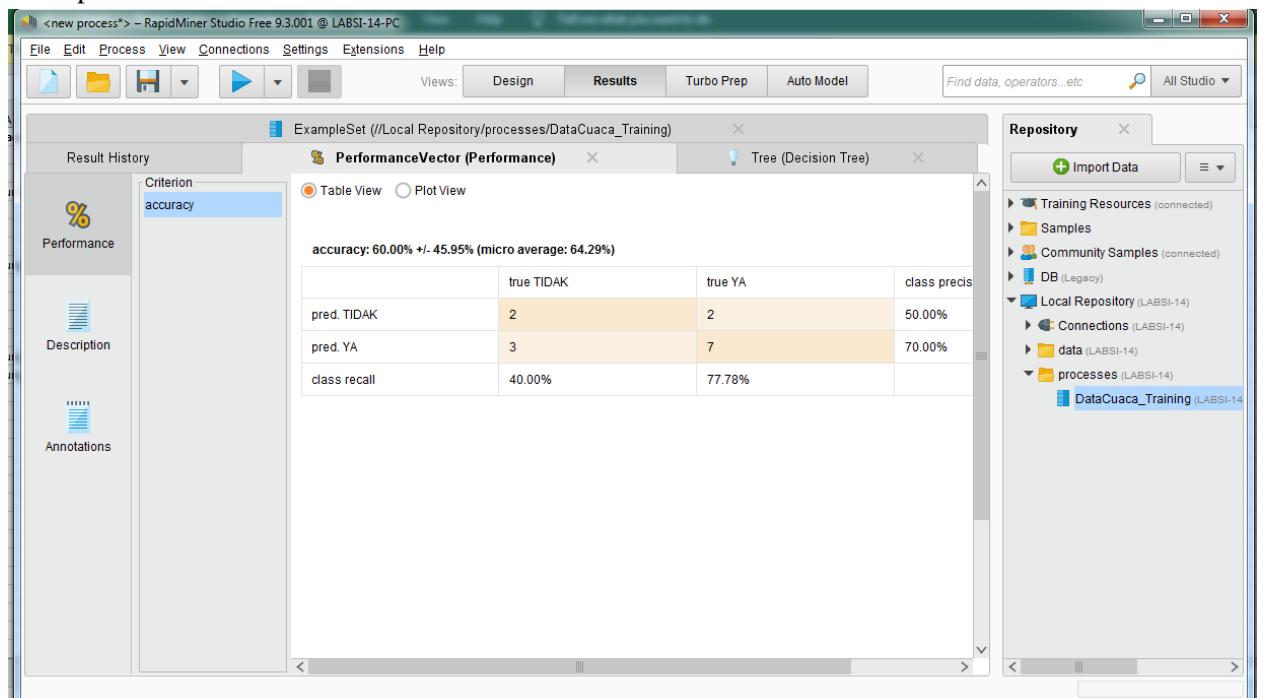
1. Gunakan DataCuaca\_Training yang telah dibuat pada kegiatan 8.4.2 dalam modul 8.
2. Drag DataCuaca\_Training ke area Process View. Drag juga Cross Validation. Hubungkan port output data training ke port input Cross Validation serta port output Cross Validation dengan port input Result.



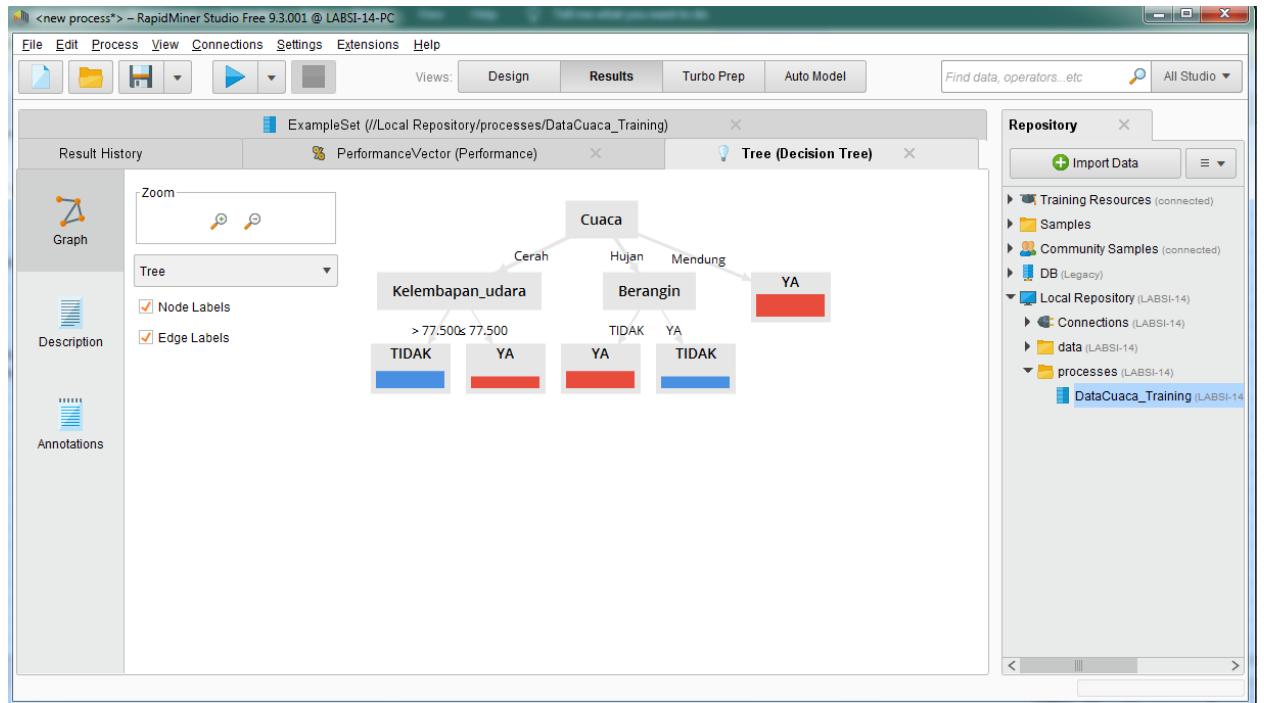
3. Klik ganda operator Cross Validation sehingga masuk ke jendela Process – Validation
4. Masukkan operator Decision Tree dalam area Training, operator Apply Model dan Performance dalam area Testing
5. Hubungkan port input dan output, dan pastikan Criterion yang dipakai adalah Information Gain



6. Jalankan proses dengan menekan tombol Run
7. Hasil performance vector



8. Hasil decision tree

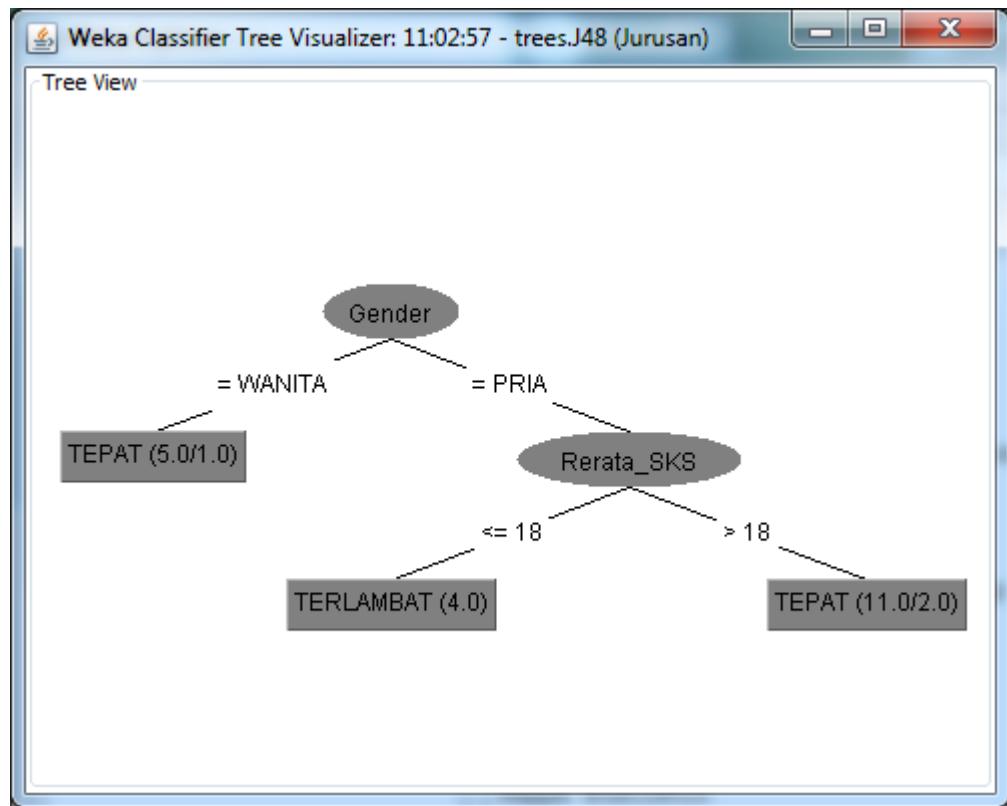


## 9.5 Tugas

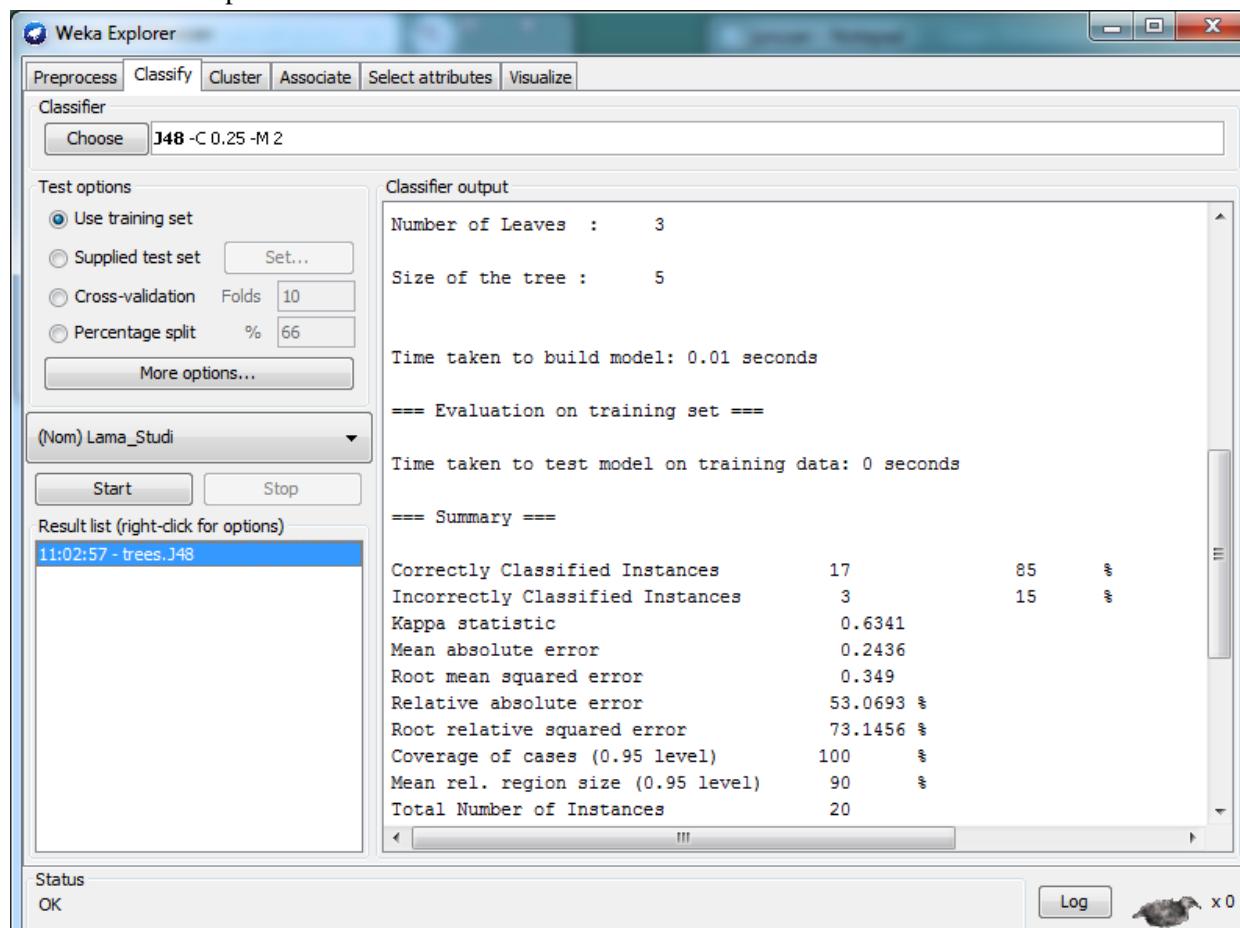
1. Isilah berdasarkan pohon keputusan pada kegiatan 9.4.2

| 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               | YA       | YA            |
| Hujan   | 64   | 85               | YA       | TIDAK         |

2. Gunakan file arff yang digunakan pada Tugas nomor 1 dalam modul 7 sebagai data training.
  - a. Buatlah dan cetaklah pohon keputusan

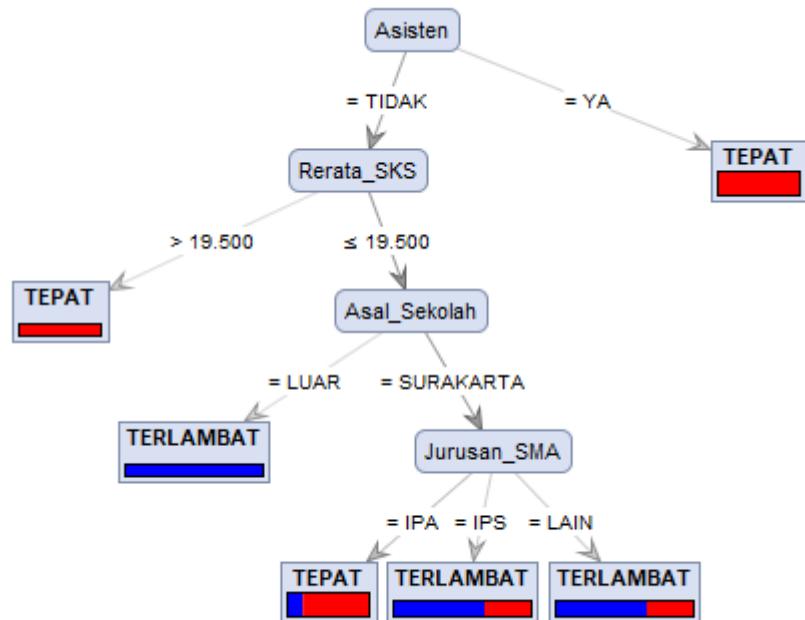


- b. Carilah nilai-nilai parameter berikut:

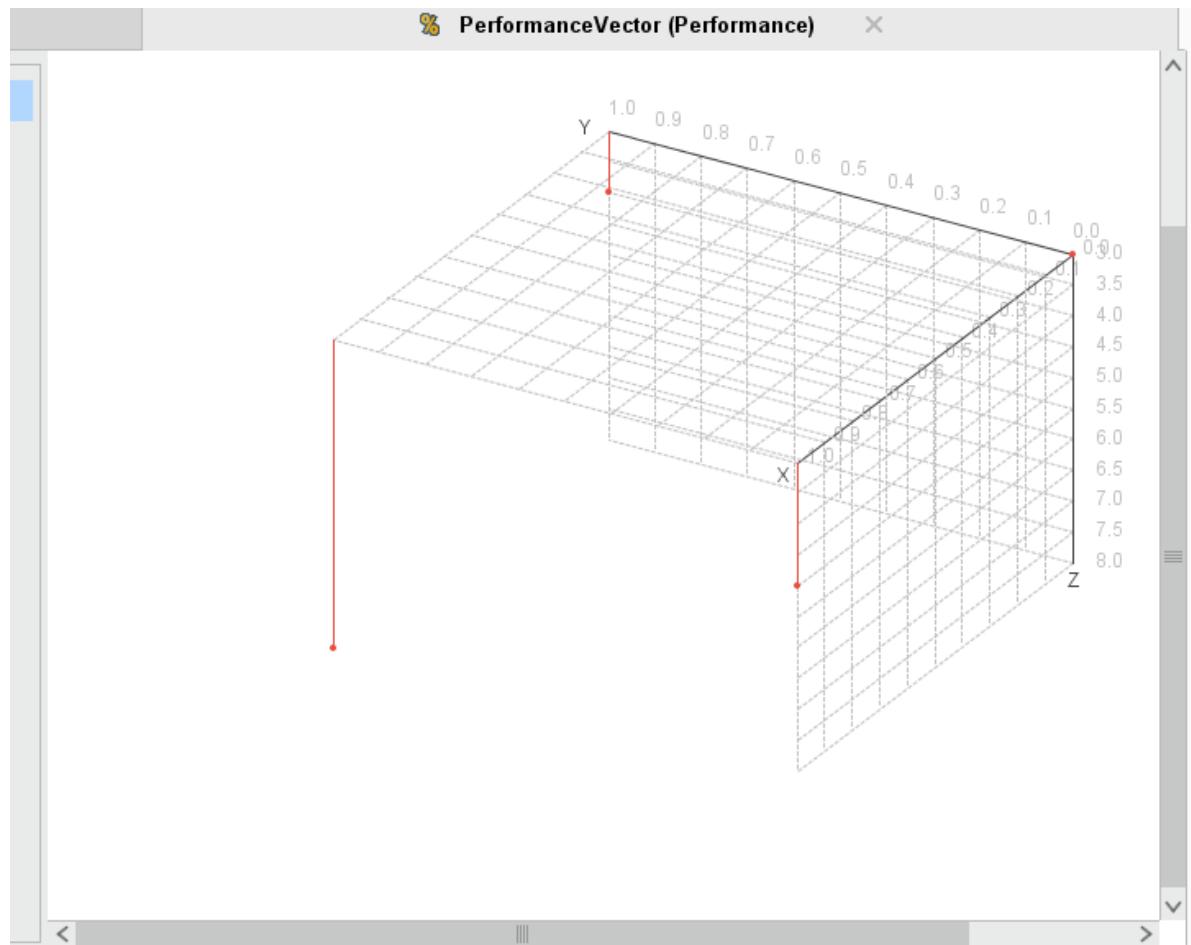


- Jumlah simpul daun pada pohon keputusan = 3
- Jumlah simpul keseluruhan pada pohon keputusan = 5

- iii. Waktu yang dibutuhkan untuk proses pelatihan = **0** detik
  - iv. Tingkat ketepatan klasifikasi = **85%**
  - v. Tingkat ketidaktepatan klasifikasi = **15%**
3. Gunakan file excel yang dikerjakan pada Tugas nomor 1 dalam modul 6 sebagai data training
- Buatlah dan cetaklah pohon keputusan



- Cetaklah Perspektif Plot View



4. Berdasarkan pohon keputusan dari soal nomor 2, tentukan klasifikasi yang terbentuk berdasarkan kondisinya sesuai dengan simpul-simpulnya

Klasifikasi yang terbentuk yaitu:

- Seseorang akan lulus tepat waktu (TEPAT) jika kondisi sebagai berikut:
  - Asisten = YA (nilai atribut lain diabaikan)
  - Asisten = TIDAK, Rerata\_SKS > 19.5
  - Asisten = TIDAK, Rerata\_SKS <= 19.5, Asal\_Sekolah = SURAKARTA, Jurusan\_SMA = IPA
- Seseorang akan lulus tidak tepat waktu (TERLAMBAT) jika kondisi sebagai berikut:
  - Asisten = TIDAK, Rerata\_SKS <= 19.5 , Asal\_Sekolah = LUAR
  - Asisten = TIDAK, Rerata\_SKS <=19.5, Jurusan\_SMA = IPS/LAIN

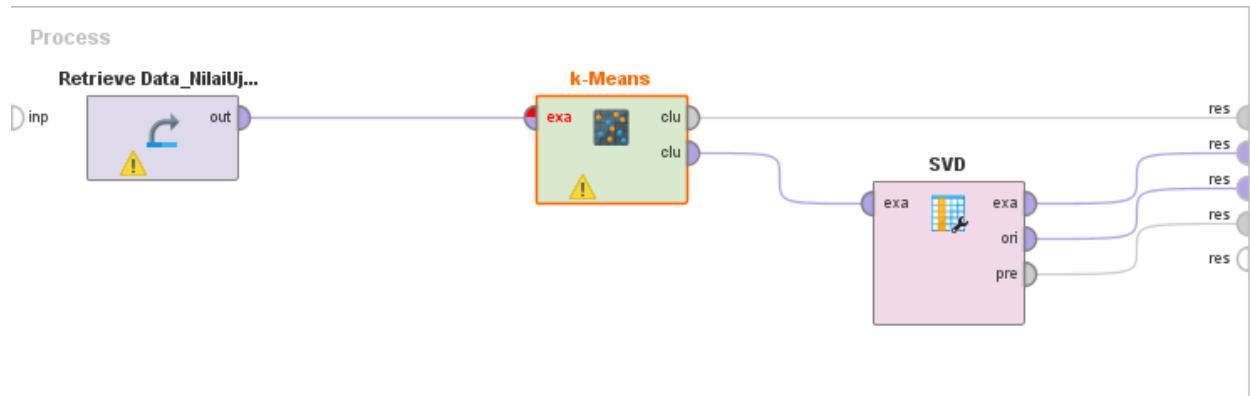
## MODUL 10

### CLUSTERING: ALGORITMA K-MEANS

- Simpan data berikut dengan nama Tabel\_NilaiUjian

| NO_SISWA | NAMA   | B.IND | B.ING |
|----------|--------|-------|-------|
| S-101    | JOKO   | 8,54  | 8,4   |
| S-102    | AGUS   | 9,98  | 6,81  |
| S-103    | SUSI   | 6,2   | 9,15  |
| S-104    | DYAH   | 5,24  | 7,26  |
| S-105    | WATI   | 5,7   | 5,71  |
| S-106    | IKA    | 8,57  | 5,87  |
| S-107    | EKO    | 7,7   | 7,71  |
| S-108    | YANTO  | 6,6   | 5,7   |
| S-109    | WAWAN  | 9     | 8,12  |
| S-110    | MAHMUD | 9,81  | 9,58  |

- Import file tersebut ke dalam Rapid Miner
- Ubah kolom NAMA menjadi id
- Simpan dengan nama Data\_NilaiUjian dan masukkan pada resitories kemudian klik Finish
- Gunakan Data\_NilaiUjian ini dan masukkan ke dalam area process
- Tambahkan operator k-Means dan SVD (Singular Value Decomposition)
- Hubungkan output-outputnya sesuai gambar berikut kemudian jalankan proses dengan menekan tombol Run



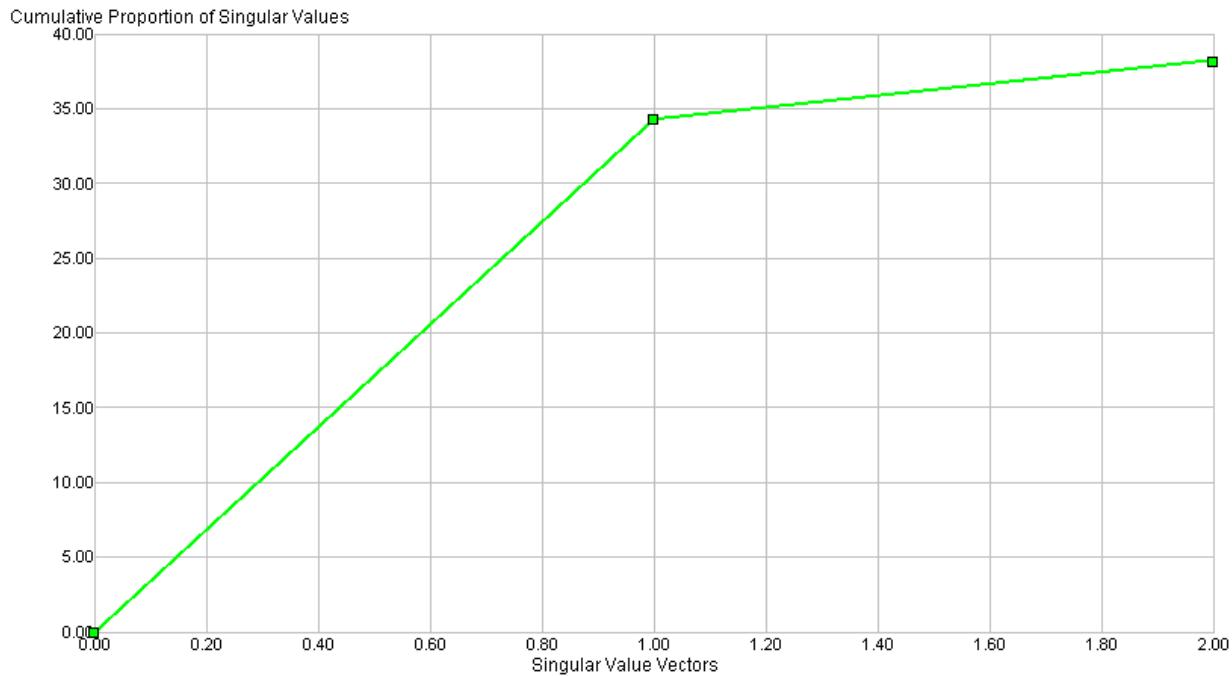
- Berikut hasil proses Clustering dengan algoritma k-Means
  - SVD
    - Nilai Eigenvalue

| Component | Singular Value | Proportion of Singular V... | Cumulative Singular Val... | Cumulative Proportion o... |
|-----------|----------------|-----------------------------|----------------------------|----------------------------|
| SVD 1     | 34.340         | 0.898                       | 34.340                     | 0.898                      |
| SVD 2     | 3.906          | 0.102                       | 38.246                     | 1.000                      |

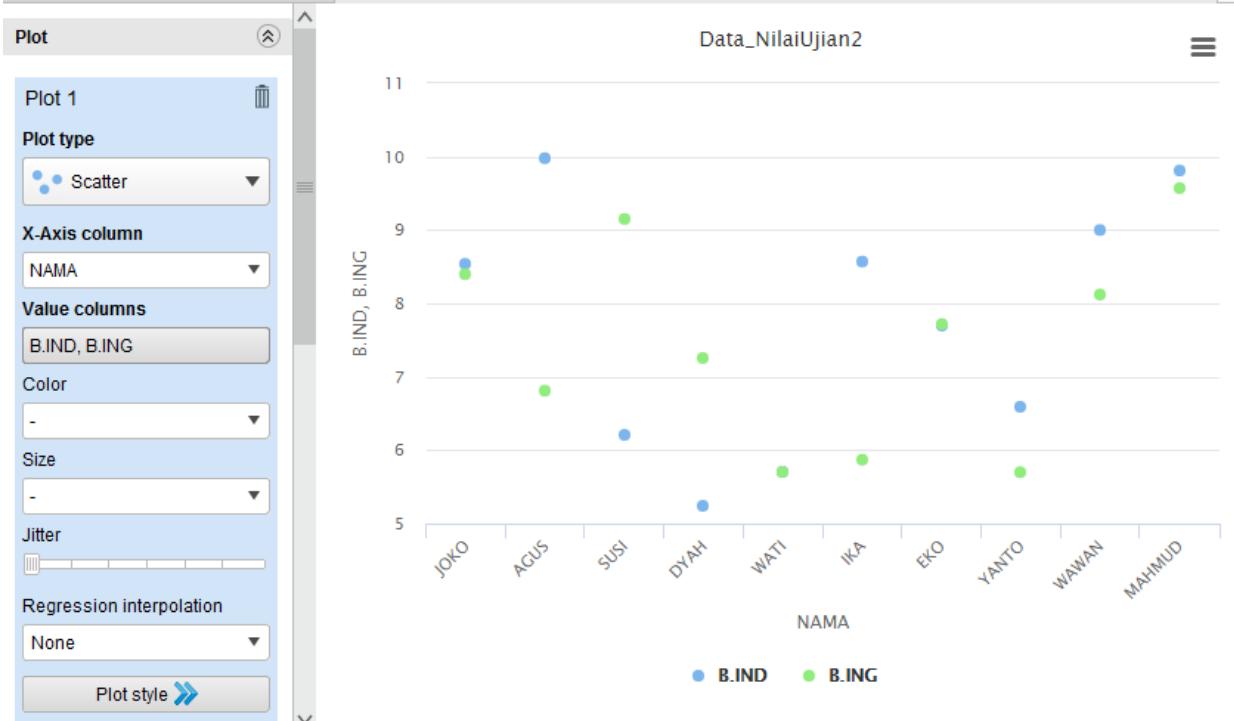
- Nilai SVD Vectors

| Attribute | SVD Vector 1 |
|-----------|--------------|
| B.IND     | 0.723        |
| B.ING     | 0.690        |

### iii. Nilai Cumulative variance



### b. ExampleSet (k-Means)



### c. ExampleSet (SVD)

Open in [!\[\]\(c5fe8f57e7a8541b205b3906c0f60920\_img.jpg\) Turbo Prep](#) [!\[\]\(94cb933bcc5c0c3c8f7137d27346ac23\_img.jpg\) Auto Model](#)

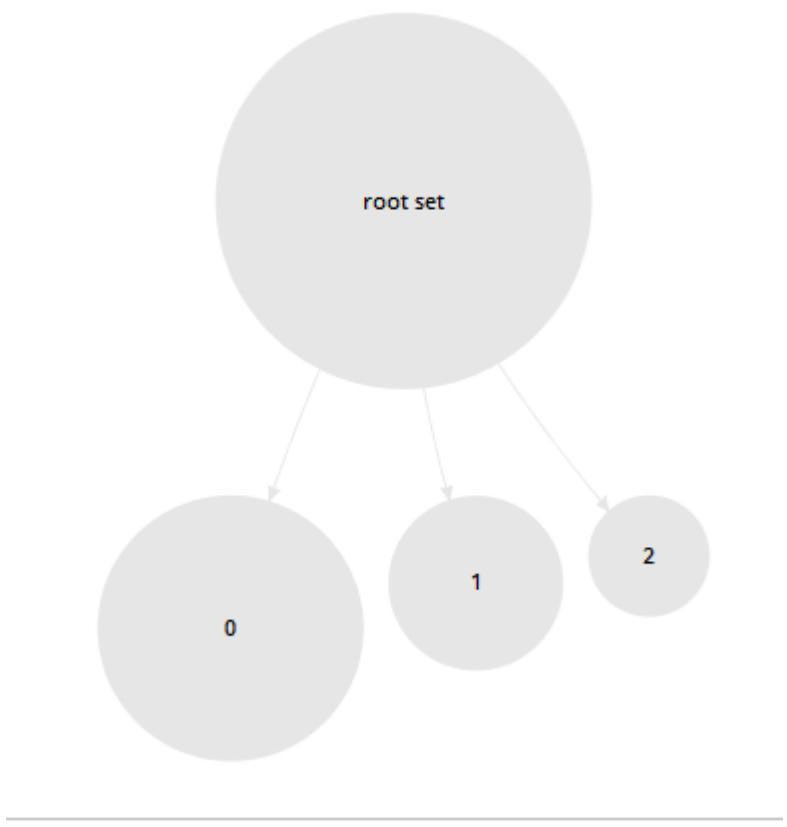
| Row No. | NAMA   | cluster ↑ | svd_1 |
|---------|--------|-----------|-------|
| 1       | JOKO   | cluster_0 | 0.349 |
| 3       | SUSI   | cluster_0 | 0.315 |
| 7       | EKO    | cluster_0 | 0.317 |
| 9       | WAWAN  | cluster_0 | 0.353 |
| 10      | MAHMUD | cluster_0 | 0.399 |
| 4       | DYAH   | cluster_1 | 0.256 |
| 5       | WATI   | cluster_1 | 0.235 |
| 8       | YANTO  | cluster_1 | 0.254 |
| 2       | AGUS   | cluster_2 | 0.347 |
| 6       | IKA    | cluster_2 | 0.299 |

- d. Cluster Model (Clustering)
  - i. Description

## Cluster Model

```
Cluster 0: 5 items
Cluster 1: 3 items
Cluster 2: 2 items
Total number of items: 10
```

- ii. Graph

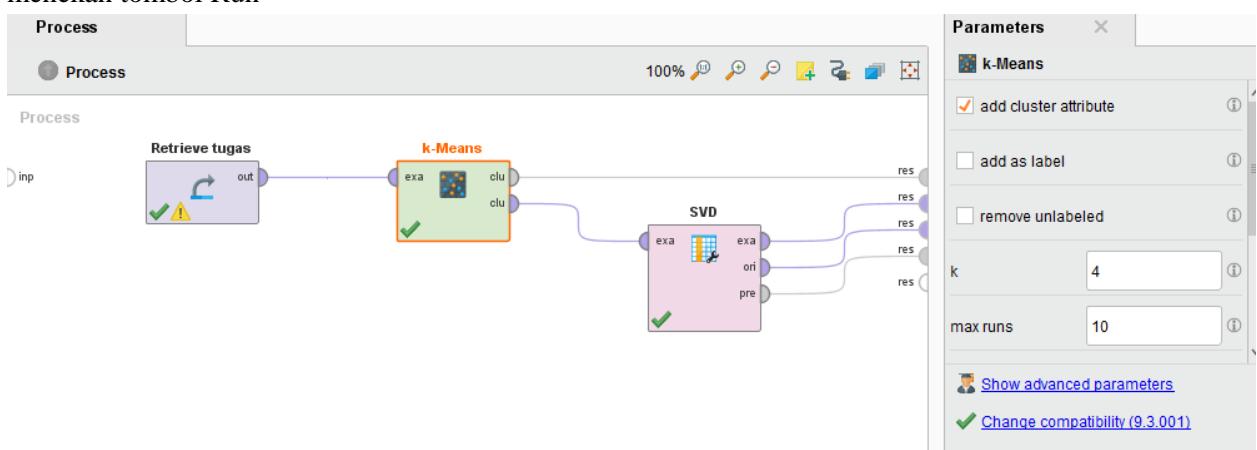


## TUGAS

- Simpan data berikut

| NO SISWA | NAMA    | B.IND | B.ING | MTK  | IPA  |
|----------|---------|-------|-------|------|------|
| S-101    | JOKO    | 8,03  | 6,28  | 8,37 | 7,87 |
| S-102    | AGUS    | 9,01  | 9,87  | 9,54 | 6,63 |
| S-103    | SUSI    | 6,92  | 9,67  | 9,04 | 6,13 |
| S-104    | DYAH    | 9,43  | 8,51  | 8,52 | 5,39 |
| S-105    | WATI    | 6,17  | 6,93  | 7,01 | 8,37 |
| S-106    | IKA     | 5,48  | 6,96  | 5,66 | 9,33 |
| S-107    | EKO     | 7,92  | 8,58  | 8,92 | 5,60 |
| S-108    | YANTO   | 9,57  | 5,61  | 5,93 | 8,05 |
| S-109    | WAWAN   | 7,33  | 5,92  | 6,25 | 8,24 |
| S-110    | MAHMUD  | 6,84  | 9,92  | 5,83 | 7,20 |
| S-111    | BUDI    | 9,81  | 5,87  | 6,75 | 5,85 |
| S-112    | SANTI   | 8,60  | 8,93  | 5,50 | 8,11 |
| S-113    | DIAN    | 7,24  | 8,11  | 8,13 | 8,83 |
| S-114    | DANI    | 5,62  | 7,53  | 7,15 | 5,74 |
| S-115    | AHMAD   | 8,97  | 7,00  | 9,12 | 5,54 |
| S-116    | BAYU    | 9,86  | 5,95  | 7,30 | 9,38 |
| S-117    | RISA    | 9,10  | 7,02  | 9,14 | 8,17 |
| S-118    | RANI    | 8,38  | 6,71  | 9,01 | 6,42 |
| S-119    | YANI    | 9,29  | 5,33  | 8,52 | 8,22 |
| S-120    | RATIH   | 6,62  | 8,69  | 9,34 | 5,39 |
| S-121    | INDAH   | 9,26  | 6,48  | 9,67 | 9,39 |
| S-122    | JONO    | 8,08  | 5,54  | 8,61 | 9,54 |
| S-123    | SARAH   | 8,91  | 6,47  | 8,21 | 5,28 |
| S-124    | RAMA    | 9,32  | 7,49  | 6,04 | 5,41 |
| S-125    | BAMBANG | 6,81  | 5,21  | 5,82 | 5,78 |
| S-126    | HADI    | 6,62  | 6,37  | 5,77 | 7,97 |
| S-127    | NANA    | 7,05  | 7,70  | 9,48 | 5,02 |
| S-128    | FEBRI   | 7,02  | 9,95  | 8,73 | 9,61 |
| S-129    | DENI    | 5,33  | 6,42  | 9,79 | 7,27 |
| S-130    | TONI    | 7,84  | 9,28  | 9,29 | 5,94 |

- Import file tersebut ke dalam Rapid Miner
- Ubah kolom NAMA menjadi id
- Simpan dengan nama tugas dan masukkan pada repositories kemudian klik Finish
- Gunakan tugas ini dan masukkan ke dalam area process
- Tambahkan operator k-Means dan SVD (Singular Value Decomposition)
- Hubungkan output-outputnya sesuai gambar berikut kemudian jalankan proses dengan menekan tombol Run



- Berikut hasil proses Clustering dengan algoritma k-Means
  - SVD

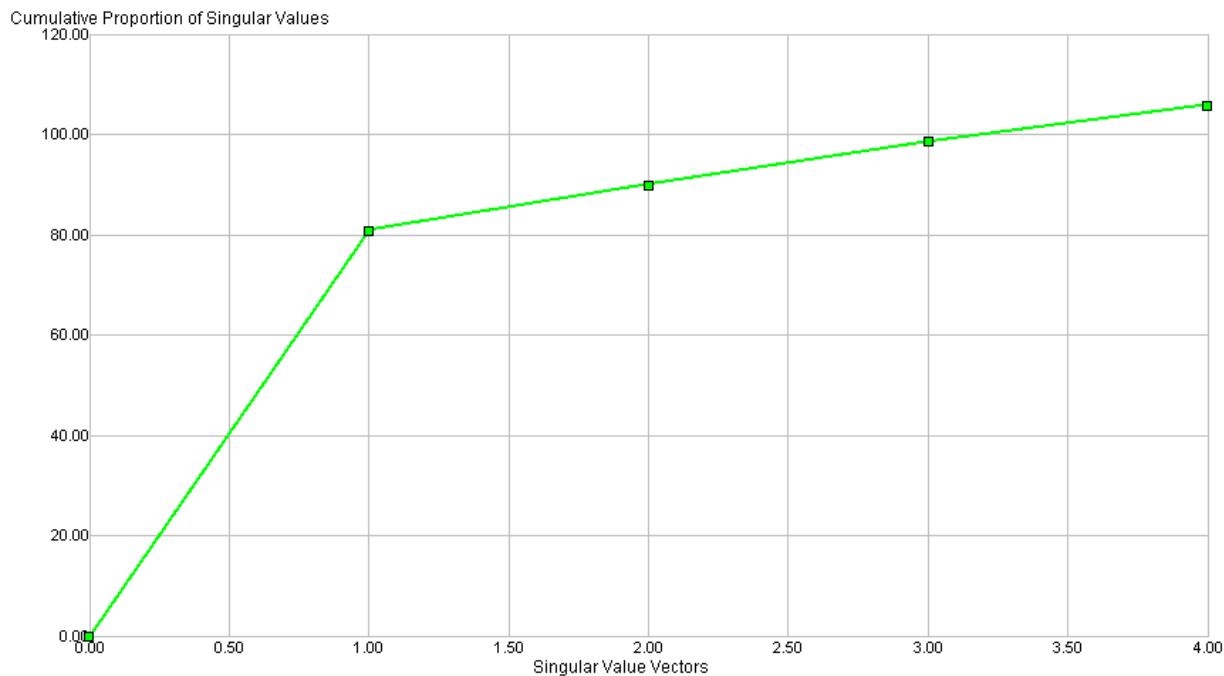
i. Nilai Eigenvalue

| Component | Singular Value | Proportion of Singular Val... | Cumulative Singular Values | Cumulative Proportion of ... |
|-----------|----------------|-------------------------------|----------------------------|------------------------------|
| SVD 1     | 81.027         | 0.764                         | 81.027                     | 0.764                        |
| SVD 2     | 9.100          | 0.086                         | 90.127                     | 0.850                        |
| SVD 3     | 8.665          | 0.082                         | 98.792                     | 0.932                        |
| SVD 4     | 7.208          | 0.068                         | 106.000                    | 1.000                        |

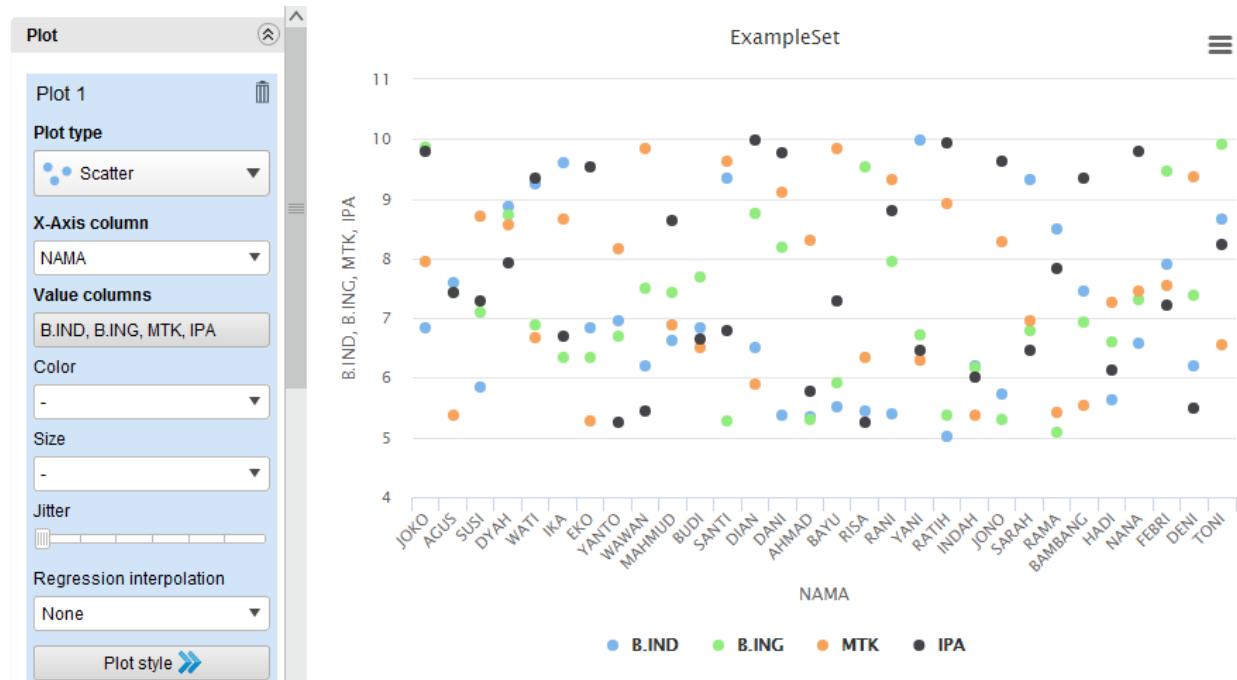
ii. Nilai SVD Vectors

| Attribute | SVD Vector 1 | SVD Vector 2 | SVD Vector 3 |
|-----------|--------------|--------------|--------------|
| B.IND     | 0.478        | 0.528        | 0.619        |
| B.ING     | 0.489        | 0.115        | -0.014       |
| MTK       | 0.510        | -0.816       | 0.208        |
| IPA       | 0.522        | 0.205        | -0.757       |

iii. Nilai Cumulative variance



b. ExampleSet (k-Means)



c. ExampleSet (SVD)

| Row No. | NAMA  | cluster ↑ | svd_1 |
|---------|-------|-----------|-------|
| 3       | SUSI  | cluster_0 | 0.179 |
| 14      | DANI  | cluster_0 | 0.201 |
| 16      | BAYU  | cluster_0 | 0.177 |
| 18      | RANI  | cluster_0 | 0.195 |
| 20      | RATIH | cluster_0 | 0.182 |
| 22      | JONO  | cluster_0 | 0.180 |
| 8       | YANTO | cluster_1 | 0.167 |
| 9       | WAWAN | cluster_1 | 0.179 |
| 11      | BUDI  | cluster_1 | 0.171 |
| 15      | AHMAD | cluster_1 | 0.153 |
| 17      | RISA  | cluster_1 | 0.163 |
| 21      | INDAH | cluster_1 | 0.146 |

| Row No. | NAMA    | cluster ↑ | svd_1 |
|---------|---------|-----------|-------|
| 26      | HADI    | cluster_1 | 0.158 |
| 29      | DENI    | cluster_1 | 0.175 |
| 1       | JOKO    | cluster_2 | 0.213 |
| 2       | AGUS    | cluster_2 | 0.171 |
| 7       | EKO     | cluster_2 | 0.173 |
| 10      | MAHMUD  | cluster_2 | 0.183 |
| 13      | DIAN    | cluster_2 | 0.193 |
| 25      | BAMBANG | cluster_2 | 0.181 |
| 27      | NANA    | cluster_2 | 0.193 |
| 28      | FEBRI   | cluster_2 | 0.198 |
| 30      | TONI    | cluster_2 | 0.205 |
| 4       | DYAH    | cluster_3 | 0.210 |

|    |       |           |       |
|----|-------|-----------|-------|
| 5  | WATI  | cluster_3 | 0.198 |
| 6  | IKA   | cluster_3 | 0.193 |
| 12 | SANTI | cluster_3 | 0.191 |
| 19 | YANI  | cluster_3 | 0.180 |
| 23 | SARAH | cluster_3 | 0.181 |
| 24 | RAMA  | cluster_3 | 0.165 |

d. Cluster Model (Clustering)

i. Description

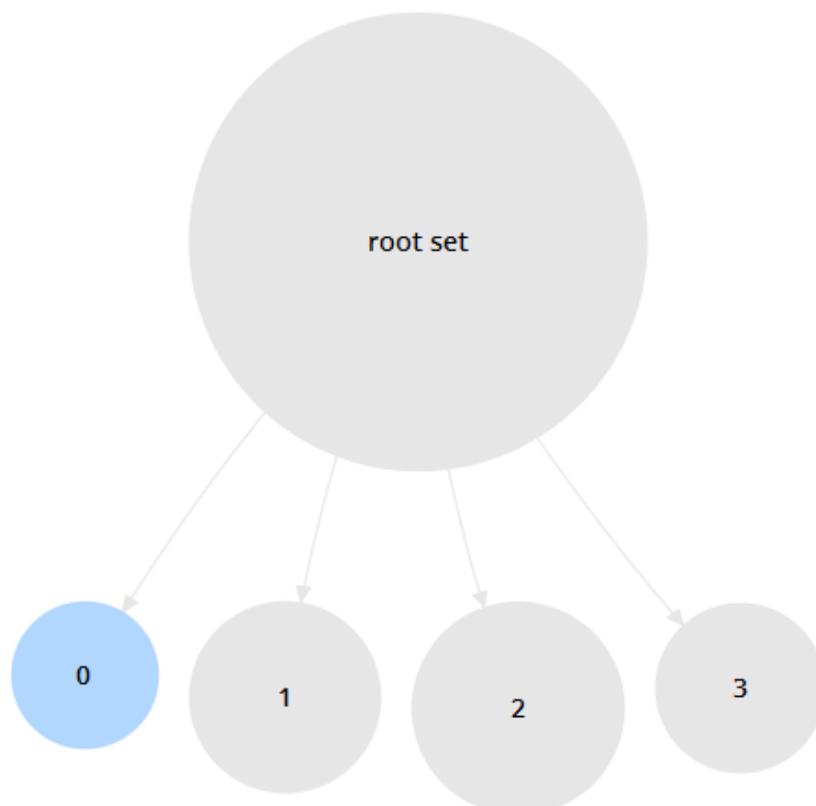
## Cluster Model

```

Cluster 0: 6 items
Cluster 1: 8 items
Cluster 2: 9 items
Cluster 3: 7 items
Total number of items: 30

```

ii. Graph



### Cluster 0

| Row No. | Nama |
|---------|------|
| 3       | Susi |

|    |       |
|----|-------|
| 14 | Dani  |
| 16 | Bayu  |
| 18 | Rani  |
| 20 | Ratih |
| 22 | Jono  |

### Cluster 1

| Row No. | Nama  |
|---------|-------|
| 8       | Yanto |
| 9       | Wawan |
| 11      | Budi  |
| 15      | Ahmad |
| 17      | Risa  |
| 21      | Indah |
| 26      | Hadi  |
| 29      | Deni  |

### Cluster 2

| Row No. | Nama    |
|---------|---------|
| 1       | Joko    |
| 2       | Agus    |
| 7       | Eko     |
| 10      | Mahmud  |
| 13      | Dian    |
| 25      | Bambang |
| 27      | Nana    |
| 28      | Febri   |
| 30      | Toni    |

### Cluster 3

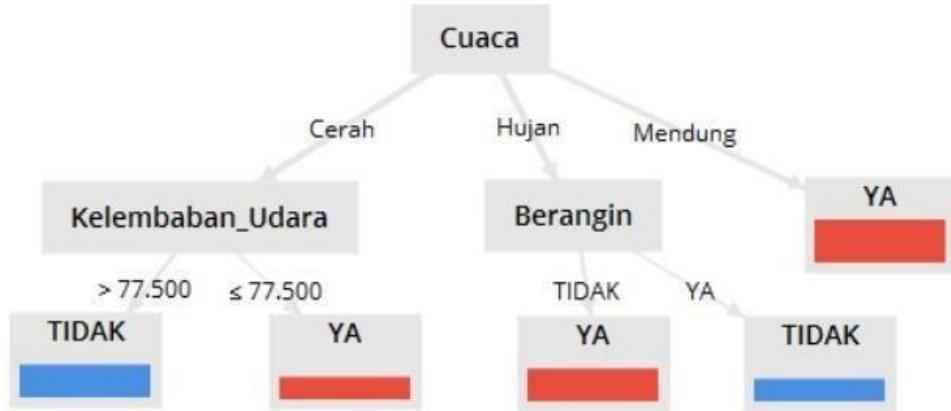
| Row No. | Nama  |
|---------|-------|
| 4       | Dyah  |
| 5       | Wati  |
| 6       | Ika   |
| 12      | Santi |
| 19      | Yani  |
| 23      | Sarah |
| 24      | rama  |

## MODUL 11

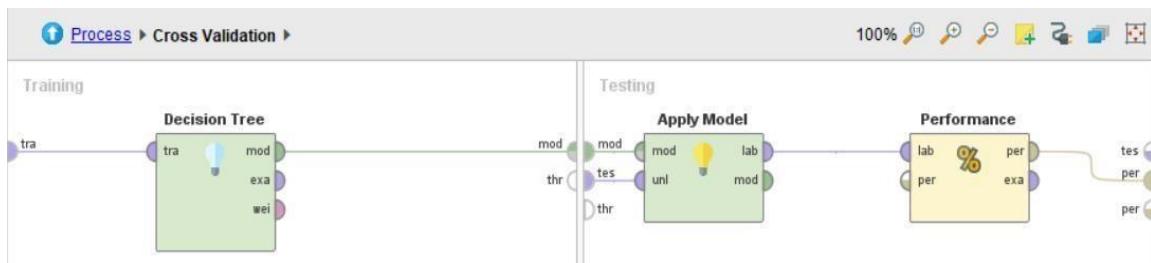
### INDUKSI DAN ATURAN ASOSIASI

#### 11.1. Induksi Aturan Data Cuaca

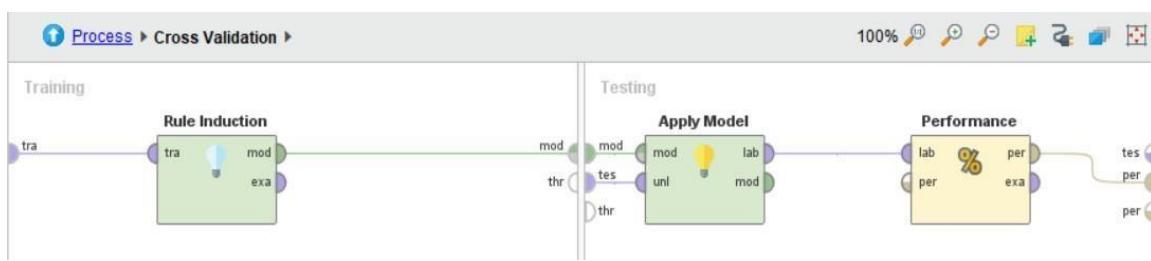
1. Jalankan aplikasi RapidMiner.
2. Gunakan model proses dari praktikum data mining **Modul 9 Kegiatan 9.4.2**.
3. Dari hasil tersebut menghasilkan sebuah pohon keputusan seperti pada gambar berikut



4. Kembali ke **Process – Cross Validation**.



5. Klik kanan operator **Decision Tree** pada area **Training**.
6. Pilih **Replace Operator** → **Modelling** → **Predictive** → **Rules** → **Rule Induction**.



7. Jalankan dengan menekan tombol **Run** (atau F11).

- Sehingga akan diperoleh sebuah induksi yang disebut sebagai **RuleModel (Rule Induction)**.

## RuleModel

```

if Kelembapan_udara ≤ 82.500 then YA (1 / 6)
if Cuaca = Cerah then TIDAK (3 / 0)
if Cuaca = Mendung then YA (0 / 2)
if Suhu ≤ 70.500 then YA (0 / 1)
else TIDAK (0 / 0)

correct: 12 out of 13 training examples.

```

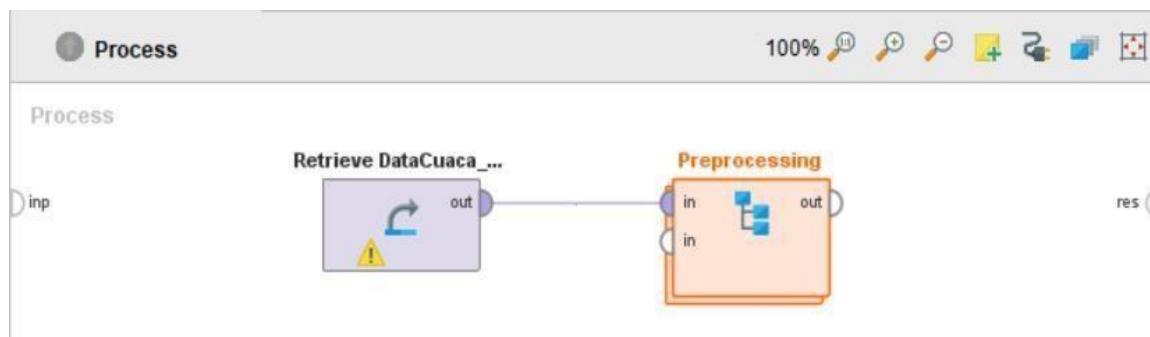
- Model Rule Induction ini juga bisa ditunjukkan hasil Performance Vector (Performance).

accuracy: 65.00% +/- 47.43% (micro average: 71.43%)

|              | true TIDAK | true YA | class precision |
|--------------|------------|---------|-----------------|
| pred. TIDAK  | 2          | 1       | 66.67%          |
| pred. YA     | 3          | 8       | 72.73%          |
| class recall | 40.00%     | 88.89%  |                 |

### 11.2. Aturan Asosiasi Data Cuaca

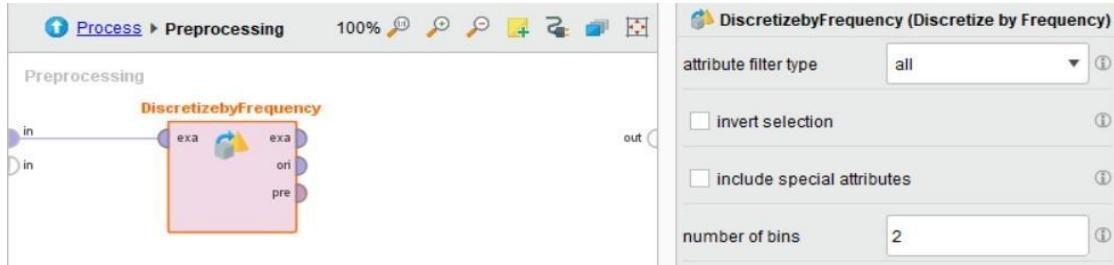
- Masih dengan menggunakan RapidMiner, pilih **New Process**.
- Gunakan **DataCuaca\_Training** dan drag dari *repository* ke area **Process View**.
- Tambahkan operator **Utility** → **Subprocess** ke dalam area. Ubah nama operator menjadi **Preprocessing**. Hubungkan port output Retrieve dengan port input Preprocessing.



- Klik ganda operator **Preprocessing** sehingga masuk pada area Nested Chain.
- Pada Nested Chain ini tambahkan operator-operator berikut :

a. **Cleansing → Binning → Discretize by Frequency**

ubah nama operator ini menjadi “**DiscretizebyFrequency**” dan biarkan nilai parameter **number of bins** (jumlah interval) = 2. Hubungkan seperti pada gambar.



b. **Blending → Attribute → Types → Nominal to Binominal**

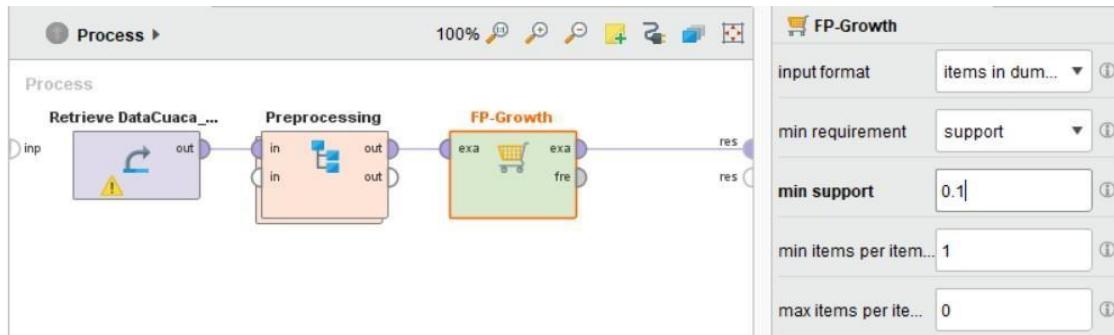
Ubah nama pada operator ini menjadi “**Nominal2Binominal**”, dan hubungkan seperti pada gambar.



6. Kembali ke *main process*, tambahkan 2 buah operator:

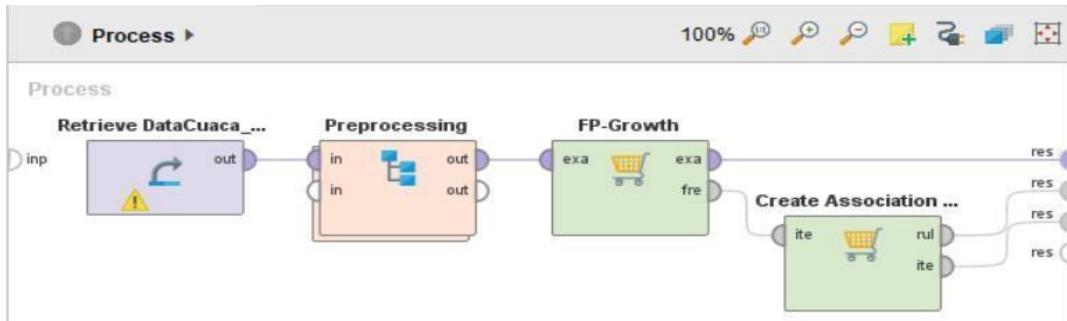
a. **Modelling → Association → FP-Growth**

Ubah nilai parameter pada min support = 0.1, hubungkan port seperti berikut.



b. **Modelling → Association → Create Association Rules**

Hubungkan seperti pada gambar.



7. Jalankan proses dengan menekan tombol **Run** (atau F11). Dapat dilihat pada hasil berikut :

a. Frequent Item Set (FP-Growth)

| No. of Sets: 26    | Size | Support | Item 1           | Item 2          | Item 3 | Item 4 |
|--------------------|------|---------|------------------|-----------------|--------|--------|
| Total Max. Size: 4 | 1    | 0.500   | Kelembapan_udara |                 |        |        |
| Min. Size: 1       | 1    | 0.429   | Berangin         |                 |        |        |
| Max. Size: 4       | 1    | 0.429   | Suhu             |                 |        |        |
| Contains Item:     | 1    | 0.357   | Cuaca = Cerah    |                 |        |        |
|                    | 1    | 0.357   | Cuaca = Hujan    |                 |        |        |
|                    | 1    | 0.286   | Cuaca = Mendung  |                 |        |        |
|                    | 2    | 0.214   | Kelembapan_udara | Berangin        |        |        |
|                    | 2    | 0.214   | Kelembapan_udara | Suhu            |        |        |
|                    | 2    | 0.214   | Kelembapan_udara | Cuaca = Cerah   |        |        |
|                    | 2    | 0.143   | Kelembapan_udara | Cuaca = Hujan   |        |        |
|                    | 2    | 0.143   | Kelembapan_udara | Cuaca = Mendung |        |        |
|                    | 2    | 0.143   | Berangin         | Suhu            |        |        |
|                    | 2    | 0.143   | Berangin         | Cuaca = Cerah   |        |        |
|                    | 2    | 0.143   | Berangin         | Cuaca = Hujan   |        |        |
|                    | 2    | 0.143   | Berangin         | Cuaca = Mendung |        |        |
|                    | 2    | 0.214   | Suhu             | Cuaca = Cerah   |        |        |
|                    | 2    | 0.071   | Suhu             | Cuaca = Hujan   |        |        |
|                    | 2    | 0.143   | Suhu             | Cuaca = Mendung |        |        |

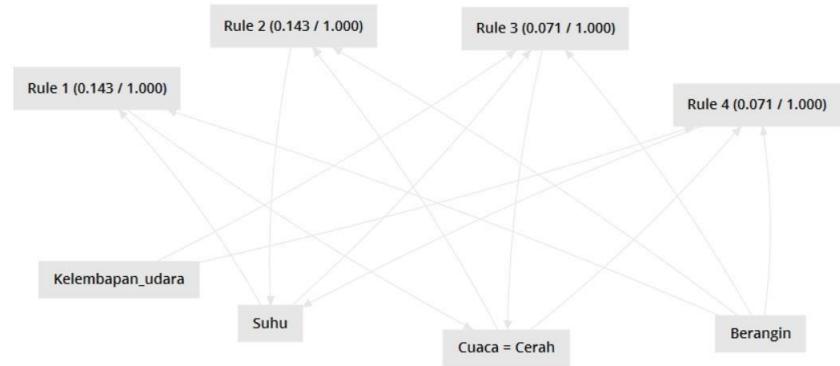
Dapat dilihat bahwa jumlah aturan asosiasi yang terbentuk adalah 28 set, dan jumlah total max size = 4, yang terdiri dari 4 buah itemset.

b. Association Rules (Create Assosiation Rules)

i. Table View

| No. | Premises                                  | Conclusion    | Support | Confidence | LaPlace | Gain   | p-s   | Lift  | Conviction |
|-----|---|---------------|---------|------------|---------|--------|-------|-------|------------|
| 1   | Berangin, Suhu                            | Cuaca = Cerah | 0.143   | 1          | 1       | -0.143 | 0.092 | 2.800 | $\infty$   |
| 2   | Berangin, Cuaca = Cerah                   | Suhu          | 0.143   | 1          | 1       | -0.143 | 0.082 | 2.333 | $\infty$   |
| 3   | Kelembapan_udara, Berangin, Suhu          | Cuaca = Cerah | 0.071   | 1          | 1       | -0.071 | 0.046 | 2.800 | $\infty$   |
| 4   | Kelembapan_udara, Berangin, Cuaca = Cerah | Suhu          | 0.071   | 1          | 1       | -0.071 | 0.041 | 2.333 | $\infty$   |

ii. Graph View



c. Exampleset (Nominal2Binomial) → Charts View



## Tugas

1. Dengan menggunakan data Lama Studi (format Excel) pada **Tugas Modul 6 Soal No.1**, carilah pola hubungan berdasarkan *Induction RuleI* (Rule Model), dan nilai *performance vector!*

- a. *Induction Rule* (Rule Model)

### RuleModel

```
if Rerata_SKS > 18.500 then TEPAT (2 / 10)
if Gender = PRIA then TERLAMBAT (4 / 0)
if Jurusan_SMA = IPA then TEPAT (0 / 2)
if Jurusan_SMA = IPS then TERLAMBAT (1 / 0)
else TEPAT (0 / 0)

correct: 17 out of 19 training examples.
```

- b. Nilai *Performance Vektor*

accuracy: 65.00% +/- 32.02% (micro average: 65.00%)

|                 | true TERLAMBAT | true TEPAT | class precision |
|-----------------|----------------|------------|-----------------|
| pred. TERLAMBAT | 4              | 4          | 50.00%          |
| pred. TEPAT     | 3              | 9          | 75.00%          |
| class recall    | 57.14%         | 69.23%     |                 |

2. Masih dengan menggunakan data *training* yang sama dengan ketentuan bahwa pada operator *Discretize by Frequency* memiliki nilai :

- a. *Number of bins* = 2

- Frequent Item Set (FP-Growth)

| No. of Sets: 55    | Size | Support | Item 1             | Item 2             | Item 3 | Item 4 | Item 5 |
|--------------------|------|---------|--------------------|--------------------|--------|--------|--------|
| Total Max. Size: 5 | 1    | 0.750   | Gender             |                    |        |        |        |
| Min. Size: 1       | 1    | 0.500   | Jurusan_SMA = IPA  |                    |        |        |        |
| Max. Size: 5       | 1    | 0.300   | Asal_Sekolah       |                    |        |        |        |
| Contains Item:     | 1    | 0.300   | Jurusan_SMA = IPS  |                    |        |        |        |
|                    | 1    | 0.250   | Asisten            |                    |        |        |        |
|                    | 1    | 0.250   | Rerata_SKS         |                    |        |        |        |
|                    | 1    | 0.200   | Jurusan_SMA = LAIN |                    |        |        |        |
|                    | 2    | 0.350   | Gender             | Jurusan_SMA = IPA  |        |        |        |
|                    | 2    | 0.250   | Gender             | Asal_Sekolah       |        |        |        |
|                    | 2    | 0.250   | Gender             | Jurusan_SMA = IPS  |        |        |        |
|                    | 2    | 0.200   | Gender             | Asisten            |        |        |        |
|                    | 2    | 0.250   | Gender             | Rerata_SKS         |        |        |        |
|                    | 2    | 0.150   | Gender             | Jurusan_SMA = LAIN |        |        |        |
|                    | 2    | 0.150   | Jurusan_SMA = IPA  | Asal_Sekolah       |        |        |        |
|                    | 2    | 0.200   | Jurusan_SMA = IPA  | Asisten            |        |        |        |
|                    | 2    | 0.100   | Jurusan_SMA = IPA  | Rerata_SKS         |        |        |        |
|                    | 2    | 0.100   | Asal_Sekolah       | Jurusan_SMA = IPS  |        |        |        |
|                    | 2    | 0.150   | Asal_Sekolah       | Asisten            |        |        |        |
|                    | 2    | 0.150   | Asal_Sekolah       | Rerata_SKS         |        |        |        |

Dapat dilihat bahwa jumlah aturan asosiasi yang terbentuk adalah 55 set dan jumlah total max size = 5, yang terdiri dari 5 buah itemset.

- Association Rules

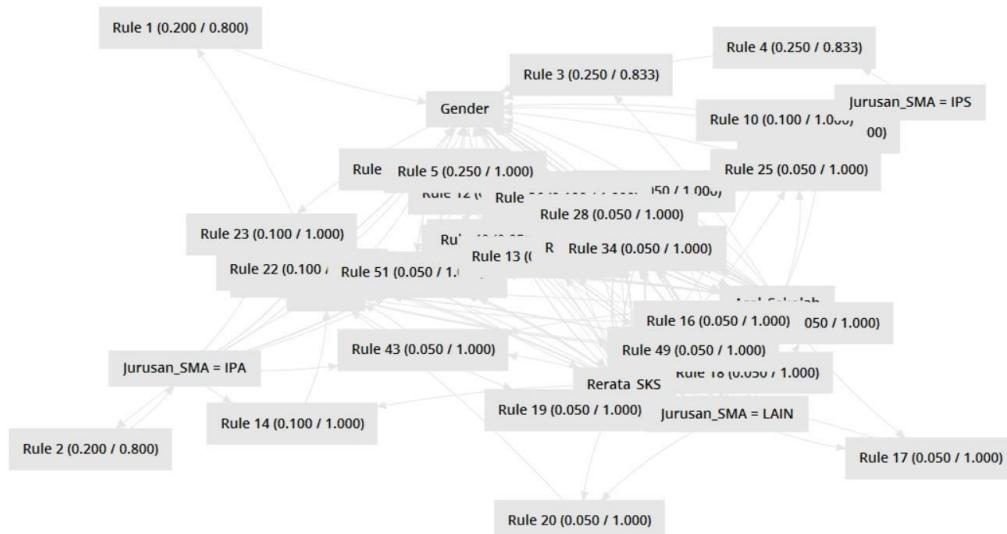
- Tabel View

| No. | Premises                                     | Conclusion           | Support | Confidence | LaPlace | Gain   | p-s   |
|-----|--|----------------------|---------|------------|---------|--------|-------|
| 3   | Asal_Sekolah                                 | Gender               | 0.250   | 0.833      | 0.962   | -0.350 | 0.025 |
| 4   | Jurusan_SMA = IPS                            | Gender               | 0.250   | 0.833      | 0.962   | -0.350 | 0.025 |
| 5   | Rerata_SKS                                   | Gender               | 0.250   | 1          | 1       | -0.250 | 0.062 |
| 6   | Jurusan_SMA = IPA, Rerata_SKS                | Gender               | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 7   | Asal_Sekolah, Jurusan_SMA = IPS              | Gender               | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 8   | Asal_Sekolah, Rerata_SKS                     | Gender               | 0.150   | 1          | 1       | -0.150 | 0.038 |
| 9   | Asal_Sekolah, Jurusan_SMA = LAIN             | Gender               | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 10  | Jurusan_SMA = IPS, Rerata_SKS                | Gender               | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 11  | Asisten, Rerata_SKS                          | Gender               | 0.150   | 1          | 1       | -0.150 | 0.038 |
| 12  | Asisten, Jurusan_SMA = LAIN                  | Gender               | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 13  | Rerata_SKS, Jurusan_SMA = LAIN               | Gender               | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 14  | Jurusan_SMA = IPA, Rerata_SKS                | Asisten              | 0.100   | 1          | 1       | -0.100 | 0.075 |
| 15  | Asal_Sekolah, Jurusan_SMA = LAIN             | Asisten              | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 16  | Asisten, Jurusan_SMA = LAIN                  | Asal_Sekolah         | 0.050   | 1          | 1       | -0.050 | 0.035 |
| 17  | Asal_Sekolah, Jurusan_SMA = LAIN             | Rerata_SKS           | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 18  | Rerata_SKS, Jurusan_SMA = LAIN               | Asal_Sekolah         | 0.050   | 1          | 1       | -0.050 | 0.035 |
| 19  | Asisten, Jurusan_SMA = LAIN                  | Rerata_SKS           | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 20  | Rerata_SKS, Jurusan_SMA = LAIN               | Asisten              | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 21  | Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS  | Gender               | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 22  | Jurusan_SMA = IPA, Rerata_SKS                | Gender, Asisten      | 0.100   | 1          | 1       | -0.100 | 0.080 |
| 23  | Gender, Jurusan_SMA = IPA, Rerata_SKS        | Asisten              | 0.100   | 1          | 1       | -0.100 | 0.075 |
| 24  | Jurusan_SMA = IPA, Asisten, Rerata_SKS       | Gender               | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 25  | Asal_Sekolah, Jurusan_SMA = IPS, Rerata_SKS  | Gender               | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 26  | Gender, Asal_Sekolah, Asisten                | Rerata_SKS           | 0.100   | 1          | 1       | -0.100 | 0.075 |
| 27  | Asal_Sekolah, Asisten, Rerata_SKS            | Gender               | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 28  | Asal_Sekolah, Jurusan_SMA = LAIN             | Gender, Asisten      | 0.050   | 1          | 1       | -0.050 | 0.040 |
| 29  | Gender, Asal_Sekolah, Jurusan_SMA = LAIN     | Asisten              | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 30  | Asisten, Jurusan_SMA = LAIN                  | Gender, Asal_Sekolah | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 31  | Gender, Asisten, Jurusan_SMA = LAIN          | Asal_Sekolah         | 0.050   | 1          | 1       | -0.050 | 0.035 |
| 32  | Asal_Sekolah, Asisten, Jurusan_SMA = LAIN    | Gender               | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 33  | Asal_Sekolah, Jurusan_SMA = LAIN             | Gender, Rerata_SKS   | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 34  | Gender, Asal_Sekolah, Jurusan_SMA = LAIN     | Rerata_SKS           | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 35  | Rerata_SKS, Jurusan_SMA = LAIN               | Gender, Asal_Sekolah | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 36  | Gender, Rerata_SKS, Jurusan_SMA = LAIN       | Asal_Sekolah         | 0.050   | 1          | 1       | -0.050 | 0.035 |
| 37  | Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN | Gender               | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 38  | Asisten, Jurusan_SMA = LAIN                  | Gender, Rerata_SKS   | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 39  | Gender, Asisten, Jurusan_SMA = LAIN          | Rerata_SKS           | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 40  | Rerata_SKS, Jurusan_SMA = LAIN               | Gender, Asisten      | 0.050   | 1          | 1       | -0.050 | 0.040 |

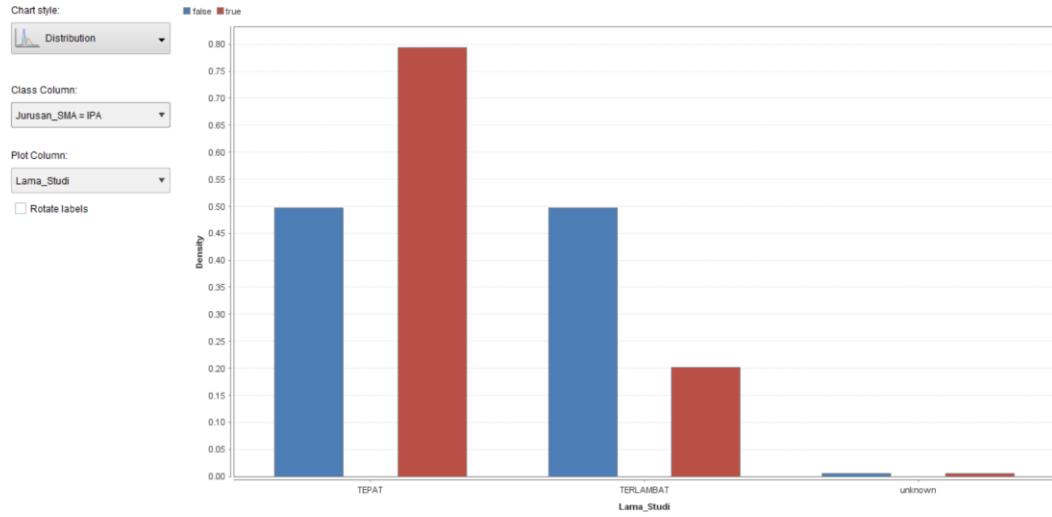
|    |   |                             |       |   |   |        |       |
|----|---|-----------------------------|-------|---|---|--------|-------|
| 41 | Gender, Rerata_SKS, Jurusan_SMA = LAIN                | Asisten                     | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 42 | Asisten, Rerata_SKS, Jurusan_SMA = LAIN               | Gender                      | 0.050 | 1 | 1 | -0.050 | 0.012 |
| 43 | Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS           | Asisten                     | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 44 | Asal_Sekolah, Jurusan_SMA = LAIN                      | Asisten, Rerata_SKS         | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 45 | Asisten, Jurusan_SMA = LAIN                           | Asal_Sekolah, Rerata_SKS    | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 46 | Asal_Sekolah, Asisten, Jurusan_SMA = LAIN             | Rerata_SKS                  | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 47 | Rerata_SKS, Jurusan_SMA = LAIN                        | Asal_Sekolah, Asisten       | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 48 | Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN          | Asisten                     | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 49 | Asisten, Rerata_SKS, Jurusan_SMA = LAIN               | Asal_Sekolah                | 0.050 | 1 | 1 | -0.050 | 0.035 |
| 50 | Gender, Jurusan_SMA = IPA, Asal_Sekolah, Asisten      | Rerata_SKS                  | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 51 | Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS           | Gender, Asisten             | 0.050 | 1 | 1 | -0.050 | 0.040 |
| 52 | Gender, Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS   | Asisten                     | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 53 | Jurusan_SMA = IPA, Asal_Sekolah, Asisten, Rerata_SKS  | Gender                      | 0.050 | 1 | 1 | -0.050 | 0.012 |
| 54 | Asal_Sekolah, Jurusan_SMA = LAIN                      | Gender, Asisten, Rerata_SKS | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 55 | Gender, Asal_Sekolah, Jurusan_SMA = LAIN              | Asisten, Rerata_SKS         | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 56 | Asisten, Jurusan_SMA = LAIN                           | Gender, Asal_Sekolah        | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 57 | Gender, Asisten, Jurusan_SMA = LAIN                   | Asal_Sekolah, Rerata_SKS    | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 58 | Asal_Sekolah, Asisten, Jurusan_SMA = LAIN             | Gender, Rerata_SKS          | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 59 | Gender, Asal_Sekolah, Asisten, Jurusan_SMA = LAIN     | Rerata_SKS                  | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 59 | Gender, Asal_Sekolah, Asisten, Jurusan_SMA = LAIN     | Rerata_SKS                  | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 60 | Rerata_SKS, Jurusan_SMA = LAIN                        | Gender, Asal_Sekolah        | 0.050 | 1 | 1 | -0.050 | 0.045 |
| 61 | Gender, Rerata_SKS, Jurusan_SMA = LAIN                | Asal_Sekolah, Asisten       | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 62 | Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN          | Gender, Asisten             | 0.050 | 1 | 1 | -0.050 | 0.040 |
| 63 | Gender, Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN  | Asisten                     | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 64 | Asisten, Rerata_SKS, Jurusan_SMA = LAIN               | Gender, Asal_Sekolah        | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 65 | Gender, Asisten, Rerata_SKS, Jurusan_SMA = LAIN       | Asal_Sekolah                | 0.050 | 1 | 1 | -0.050 | 0.035 |
| 66 | Asal_Sekolah, Asisten, Rerata_SKS, Jurusan_SMA = LAIN | Gender                      | 0.050 | 1 | 1 | -0.050 | 0.012 |

Pada tabelini dapat dilihat terdapat 66 data pasangan premis-premis dan kesimpulan (Premises and Conclusion) yang menunjukkan hubungan implikasi berdasarkan aturan asosiasi.

#### - Graph View



- Gambarkan grafik chart pola distribusi data pada ExampleSet yang terbentuk!



a. *Number of bins = 3*

- Frequent Item Set (FP-Growth)

| No. of Sets: 85    | Size | Support | Item 1                     | Item 2                     | Item 3 | Item 4 | Item 5 |
|--------------------|------|---------|----------------------------|----------------------------|--------|--------|--------|
| Total Max. Size: 5 |      |         |                            |                            |        |        |        |
| Min. Size: 1       | 1    | 0.750   | Gender                     |                            |        |        |        |
| Max. Size: 5       | 1    | 0.500   | Jurusan_SMA = IPA          |                            |        |        |        |
| Contains item:     | 1    | 0.400   | Rerata_SKS = range1 [-∞... |                            |        |        |        |
|                    | 1    | 0.350   | Rerata_SKS = range2 [18... |                            |        |        |        |
|                    | 1    | 0.300   | Asal_Sekolah               |                            |        |        |        |
|                    | 1    | 0.300   | Jurusan_SMA = IPS          |                            |        |        |        |
|                    | 1    | 0.250   | Asisten                    |                            |        |        |        |
|                    | 1    | 0.250   | Rerata_SKS = range3 [19... |                            |        |        |        |
|                    | 1    | 0.200   | Jurusan_SMA = LAIN         |                            |        |        |        |
|                    | 2    | 0.350   | Gender                     | Jurusan_SMA = IPA          |        |        |        |
|                    | 2    | 0.200   | Gender                     | Rerata_SKS = range1 [-∞... |        |        |        |
|                    | 2    | 0.300   | Gender                     | Rerata_SKS = range2 [18... |        |        |        |
|                    | 2    | 0.250   | Gender                     | Asal_Sekolah               |        |        |        |
|                    | 2    | 0.250   | Gender                     | Jurusan_SMA = IPS          |        |        |        |
|                    | 2    | 0.200   | Gender                     | Asisten                    |        |        |        |
|                    | 2    | 0.250   | Gender                     | Rerata_SKS = range3 [19... |        |        |        |
|                    | 2    | 0.150   | Gender                     | Jurusan_SMA = LAIN         |        |        |        |
|                    | 2    | 0.200   | Jurusan_SMA = IPA          | Rerata_SKS = range1 [-∞... |        |        |        |
|                    | 2    | 0.200   | Jurusan_SMA = IPA          | Rerata_SKS = range2 [18... |        |        |        |

Dapat dilihat bahwa jumlah aturan asosiasi yang terbentuk adalah 85 set dan jumlah total max size = 5, yang terdiri dari 5 buah itemset.

- Association Rules

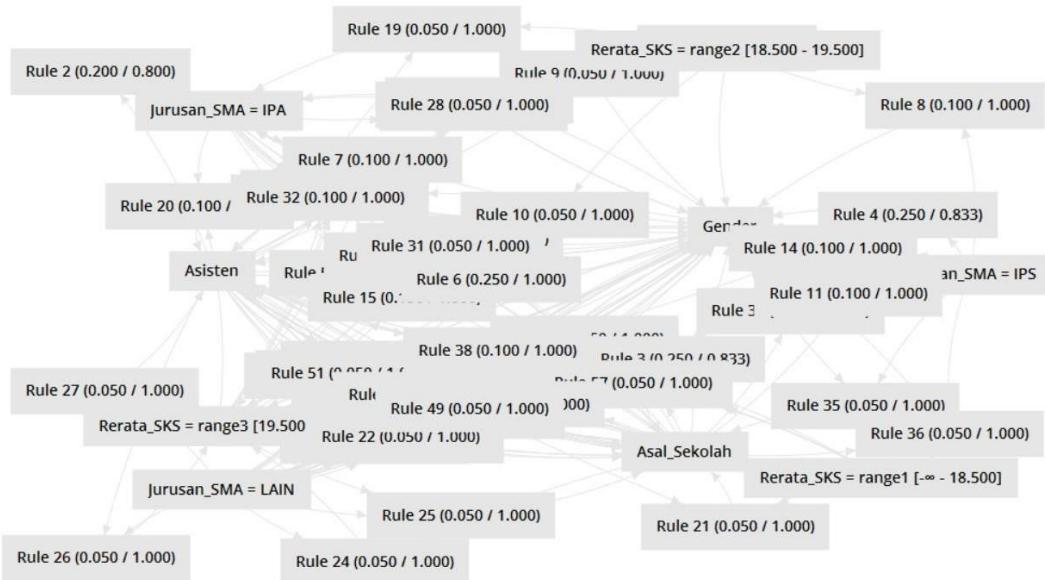
- Tabel View

| No. | Premises  | Conclusion          | Support | Confidence | LaPlace | Gain   | p-s   |
|-----|---|---------------------|---------|------------|---------|--------|-------|
| 3   | Asal_Sekolah  | Gender              | 0.250   | 0.833      | 0.962   | -0.350 | 0.025 |
| 4   | Jurusan_SMA = IPS                                   | Gender              | 0.250   | 0.833      | 0.962   | -0.350 | 0.025 |
| 5   | Rerata_SKS = range2 [18.500 - 19.500]               | Gender              | 0.300   | 0.857      | 0.963   | -0.400 | 0.038 |
| 6   | Rerata_SKS = range3 [19.500 - ∞]                    | Gender              | 0.250   | 1          | 1       | -0.250 | 0.062 |
| 7   | Jurusan_SMA = IPA, Rerata_SKS = range3 [19.50...    | Gender              | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 8   | Rerata_SKS = range2 [18.500 - 19.500], Jurusan_...  | Gender              | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 9   | Rerata_SKS = range2 [18.500 - 19.500], Asisten      | Gender              | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 10  | Rerata_SKS = range2 [18.500 - 19.500], Jurusan_...  | Gender              | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 11  | Asal_Sekolah, Jurusan_SMA = IPS                     | Gender              | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 12  | Asal_Sekolah, Rerata_SKS = range3 [19.500 - ∞]      | Gender              | 0.150   | 1          | 1       | -0.150 | 0.038 |
| 13  | Asal_Sekolah, Jurusan_SMA = LAIN                    | Gender              | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 14  | Jurusan_SMA = IPS, Rerata_SKS = range3 [19.50...    | Gender              | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 15  | Asisten, Rerata_SKS = range3 [19.500 - ∞]           | Gender              | 0.150   | 1          | 1       | -0.150 | 0.038 |
| 16  | Asisten, Jurusan_SMA = LAIN                         | Gender              | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 17  | Rerata_SKS = range3 [19.500 - ∞], Jurusan_SMA ...   | Gender              | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 18  | Rerata_SKS = range1 [-∞ - 18.500], Asisten          | Jurusan_SMA = IPA   | 0.050   | 1          | 1       | -0.050 | 0.025 |
| 19  | Rerata_SKS = range2 [18.500 - 19.500], Asisten      | Jurusan_SMA = IPA   | 0.050   | 1          | 1       | -0.050 | 0.025 |
| 20  | Jurusan_SMA = IPA, Rerata_SKS = range3 [19.50...    | Asisten             | 0.100   | 1          | 1       | -0.100 | 0.075 |
| 21  | Rerata_SKS = range1 [-∞ - 18.500], Asisten          | Asal_Sekolah        | 0.050   | 1          | 1       | -0.050 | 0.035 |
| 22  | Asal_Sekolah, Jurusan_SMA = LAIN                    | Asisten             | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 23  | Asisten, Jurusan_SMA = LAIN                         | Asal_Sekolah        | 0.050   | 1          | 1       | -0.050 | 0.035 |
| 24  | Asal_Sekolah, Jurusan_SMA = LAIN                    | Rerata_SKS = ran... | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 25  | Rerata_SKS = range3 [19.500 - ∞], Jurusan_SMA ...   | Asal_Sekolah        | 0.050   | 1          | 1       | -0.050 | 0.035 |
| 26  | Asisten, Jurusan_SMA = LAIN                         | Rerata_SKS = ran... | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 27  | Rerata_SKS = range3 [19.500 - ∞], Jurusan_SMA ...   | Asisten             | 0.050   | 1          | 1       | -0.050 | 0.038 |
| 28  | Rerata_SKS = range2 [18.500 - 19.500], Asisten      | Gender, Jurusan_... | 0.050   | 1          | 1       | -0.050 | 0.033 |
| 29  | Gender, Rerata_SKS = range2 [18.500 - 19.500], A... | Jurusan_SMA = IPA   | 0.050   | 1          | 1       | -0.050 | 0.025 |
| 30  | Jurusan_SMA = IPA, Rerata_SKS = range2 [18.50...    | Gender              | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 31  | Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS =...    | Gender              | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 32  | Jurusan_SMA = IPA, Rerata_SKS = range3 [19.50...    | Gender, Asisten     | 0.100   | 1          | 1       | -0.100 | 0.080 |
| 33  | Gender, Jurusan_SMA = IPA, Rerata_SKS = range...    | Asisten             | 0.100   | 1          | 1       | -0.100 | 0.075 |
| 34  | Jurusan_SMA = IPA, Asisten, Rerata_SKS = range...   | Gender              | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 35  | Gender, Rerata_SKS = range1 [-∞ - 18.500], Juru...  | Asal_Sekolah        | 0.050   | 1          | 1       | -0.050 | 0.035 |
| 36  | Rerata_SKS = range1 [-∞ - 18.500], Asal_Sekolah...  | Gender              | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 37  | Asal_Sekolah, Jurusan_SMA = IPS, Rerata_SKS =...    | Gender              | 0.050   | 1          | 1       | -0.050 | 0.012 |
| 38  | Gender, Asal_Sekolah, Asisten                       | Rerata_SKS = ran... | 0.100   | 1          | 1       | -0.100 | 0.075 |
| 39  | Asal_Sekolah, Asisten, Rerata_SKS = range3 [19....  | Gender              | 0.100   | 1          | 1       | -0.100 | 0.025 |
| 40  | Asal_Sekolah, Jurusan_SMA = LAIN                    | Gender, Asisten     | 0.050   | 1          | 1       | -0.050 | 0.040 |

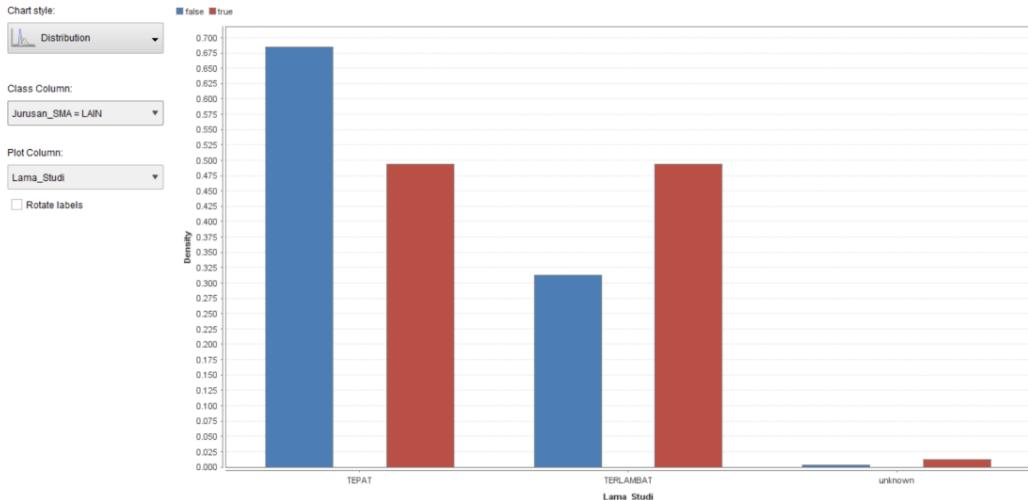
|    |   |                       |       |   |   |        |       |
|----|---|-----------------------|-------|---|---|--------|-------|
| 41 | Gender, Asal_Sekolah, Jurusan_SMA = LAIN            | Asisten               | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 42 | Asisten, Jurusan_SMA = LAIN                         | Gender, Asal_Sek...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 43 | Gender, Asisten, Jurusan_SMA = LAIN                 | Asal_Sekolah          | 0.050 | 1 | 1 | -0.050 | 0.035 |
| 44 | Asal_Sekolah, Asisten, Jurusan_SMA = LAIN           | Gender                | 0.050 | 1 | 1 | -0.050 | 0.012 |
| 45 | Asal_Sekolah, Jurusan_SMA = LAIN                    | Gender, Rerata_S...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 46 | Gender, Asal_Sekolah, Jurusan_SMA = LAIN            | Rerata_SKS = ran...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 47 | Rerata_SKS = range3 [19.500 - ∞], Jurusan_SMA ...   | Gender, Asal_Sek...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 48 | Gender, Rerata_SKS = range3 [19.500 - ∞], Jurus...  | Asal_Sekolah          | 0.050 | 1 | 1 | -0.050 | 0.035 |
| 49 | Asal_Sekolah, Rerata_SKS = range3 [19.500 - ∞],...  | Gender                | 0.050 | 1 | 1 | -0.050 | 0.012 |
| 50 | Asisten, Jurusan_SMA = LAIN                         | Gender, Rerata_S...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 51 | Gender, Asisten, Jurusan_SMA = LAIN                 | Rerata_SKS = ran...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 52 | Rerata_SKS = range3 [19.500 - ∞], Jurusan_SMA ...   | Gender, Asisten       | 0.050 | 1 | 1 | -0.050 | 0.040 |
| 53 | Gender, Rerata_SKS = range3 [19.500 - ∞], Jurus...  | Asisten               | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 54 | Asisten, Rerata_SKS = range3 [19.500 - ∞], Jurus... | Gender                | 0.050 | 1 | 1 | -0.050 | 0.012 |
| 55 | Rerata_SKS = range1 [-∞ - 18.500], Asisten          | Jurusan_SMA = IP...   | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 56 | Jurusan_SMA = IPA, Rerata_SKS = range1 [-∞ - 1...   | Asal_Sekolah          | 0.050 | 1 | 1 | -0.050 | 0.035 |
| 57 | Rerata_SKS = range1 [-∞ - 18.500], Asal_Sekolah...  | Jurusan_SMA = IPA     | 0.050 | 1 | 1 | -0.050 | 0.025 |
| 58 | Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS =...    | Asisten               | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 59 | Asal_Sekolah, Jurusan_SMA = LAIN                    | Asisten, Rerata_S...  | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 60 | Asisten, Jurusan_SMA = LAIN                         | Asal_Sekolah, Rer...  | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 61 | Asal_Sekolah, Asisten, Jurusan_SMA = LAIN           | Rerata_SKS = ran...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 62 | Rerata_SKS = range3 [19.500 - ∞], Jurusan_SMA ...   | Asal_Sekolah, Asi...  | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 63 | Asal_Sekolah, Rerata_SKS = range3 [19.500 - ∞],...  | Asisten               | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 64 | Asisten, Rerata_SKS = range3 [19.500 - ∞], Jurus... | Asal_Sekolah          | 0.050 | 1 | 1 | -0.050 | 0.035 |
| 65 | Gender, Jurusan_SMA = IPA, Asal_Sekolah, Asisten    | Rerata_SKS = ran...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 66 | Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS =...    | Gender, Asisten       | 0.050 | 1 | 1 | -0.050 | 0.040 |
| 67 | Gender, Jurusan_SMA = IPA, Asal_Sekolah, Rerat...   | Asisten               | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 68 | Jurusan_SMA = IPA, Asal_Sekolah, Asisten, Rerat...  | Gender                | 0.050 | 1 | 1 | -0.050 | 0.012 |
| 69 | Asal_Sekolah, Jurusan_SMA = LAIN                    | Gender, Asisten, R... | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 70 | Gender, Asal_Sekolah, Jurusan_SMA = LAIN            | Asisten, Rerata_S...  | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 71 | Asisten, Jurusan_SMA = LAIN                         | Gender, Asal_Sek...   | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 72 | Gender, Asisten, Jurusan_SMA = LAIN                 | Asal_Sekolah, Rer...  | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 73 | Asal_Sekolah, Asisten, Jurusan_SMA = LAIN           | Gender, Rerata_S...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 74 | Gender, Asal_Sekolah, Asisten, Jurusan_SMA = L...   | Rerata_SKS = ran...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 75 | Rerata_SKS = range3 [19.500 - ∞], Jurusan_SMA ...   | Gender, Asal_Sek...   | 0.050 | 1 | 1 | -0.050 | 0.045 |
| 76 | Gender, Rerata_SKS = range3 [19.500 - ∞], Jurus...  | Asal_Sekolah, Asi...  | 0.050 | 1 | 1 | -0.050 | 0.043 |
| 77 | Asal_Sekolah, Rerata_SKS = range3 [19.500 - ∞],...  | Gender, Asisten       | 0.050 | 1 | 1 | -0.050 | 0.040 |
| 78 | Gender, Asal_Sekolah, Rerata_SKS = range3 [19....   | Asisten               | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 79 | Asisten, Rerata_SKS = range3 [19.500 - ∞], Jurus... | Gender, Asal_Sek...   | 0.050 | 1 | 1 | -0.050 | 0.038 |
| 80 | Gender, Asisten, Rerata_SKS = range3 [19.500 - ...  | Asal_Sekolah          | 0.050 | 1 | 1 | -0.050 | 0.035 |
| 81 | Asal_Sekolah, Asisten, Rerata_SKS = range3 [19....  | Gender                | 0.050 | 1 | 1 | -0.050 | 0.012 |

Pada tabelini dapat dilihat terdapat 81 data pasangan premis-premis dan kesimpulan (Premises dan Conclusion) yang menunjukkan hubungan implikasi berdasarkan aturan asosiasi.

- Graph View



- Gambarkan grafik chart pola distribusi data pada ExampleSet yang terbentuk!



**MODUL12**  
**REGRESI LINEAR SEDERHANA**

**Mencari Nilai t-hitung dan Model Regresi Linier**

1. Buka Ms.Excel, dan buatlah tabel data siswa. Simpan dengan nama **Tabel\_LamaBelajardanNilaiUjian.xls**.

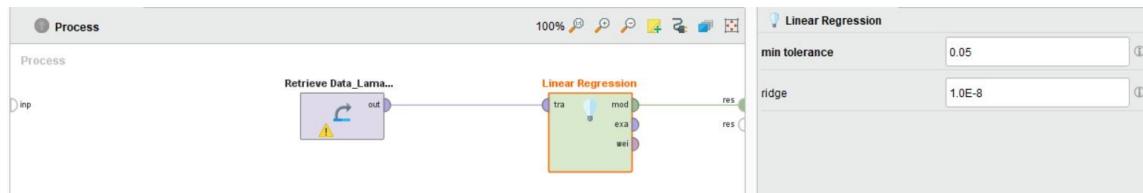
| NO_SISWA | NAMA   | LAMA BELAJAR (JAM) | NILAI |
|----------|--------|--------------------|-------|
| S-101    | JOKO   | 15                 | 783   |
| S-102    | AGUS   | 18                 | 877   |
| S-103    | SUSI   | 7                  | 505   |
| S-104    | DYAH   | 9                  | 860   |
| S-105    | WATI   | 15                 | 968   |
| S-106    | IKA    | 17                 | 793   |
| S-107    | EKO    | 10                 | 752   |
| S-108    | YANTO  | 5                  | 571   |
| S-109    | WAWAN  | 8                  | 667   |
| S-110    | MAHMUD | 15                 | 723   |

2. Jalankan RapidMiner.
3. Gunakan file **Tabel\_LamaBelajardanNilaiUjian.xls** sebagai data yang digunakan dalam proses Regresi Linier. Import file ini ke dalam repositori.
4. Ubahlah tipe data dan jenis masing-masing atribut sebagai berikut :

| NO_SISWA<br><i>polynominal<br/>id</i> | NAMA<br><i>polynominal</i> | LAMA BELAJAR (JAM)<br><i>integer</i> | NILAI<br><i>integer<br/>label</i> |
|---------------------------------------|----------------------------|--------------------------------------|-----------------------------------|
| 1 S-101                               | JOKO                       | 15                                   | 783                               |
| 2 S-102                               | AGUS                       | 18                                   | 877                               |
| 3 S-103                               | SUSI                       | 7                                    | 505                               |
| 4 S-104                               | DYAH                       | 9                                    | 860                               |
| 5 S-105                               | WATI                       | 15                                   | 968                               |
| 6 S-106                               | IKA                        | 17                                   | 793                               |
| 7 S-107                               | EKO                        | 10                                   | 752                               |
| 8 S-108                               | YANTO                      | 5                                    | 571                               |
| 9 S-109                               | WAWAN                      | 8                                    | 667                               |
| 10 S-110                              | MAHMUD                     | 15                                   | 723                               |

5. Pada langkah ini, beri nama **Data\_LamaBelajardanNilaiUjian** dan masukkan pada repository, klik **Finish**.
6. Gunakan Data\_LamaBelajardanNilaiUjian ini dan masukkan ke dalam area process.
7. Tambahkan operator **Modelling→Predictive→Functions→Linier Regression**. Hubungkan port sesuai pada gambar.

8. Klik pada operator **Linier Regression**, tentukan paramete **min tolerance = 0.05**.



9. Jalankan proses dengan menekan tombol Run (atau F11).

10. Hasil proses regresi linier :

- a. Tabel View (mencari besarnya nilai t-hitung)

| Attribute          | Coefficient | Std. Error | Std. Coefficient | Tolerance | t-Stat | p-Value | Code |
|--------------------|-------------|------------|------------------|-----------|--------|---------|------|
| LAMA BELAJAR (JAM) | 21.608      | 7.645      | 0.707            | 1         | 2.827  | 0.022   | **   |
| (Intercept)        | 492.769     | 96.909     | ?                | ?         | 5.085  | 0.001   | ***  |

- b. Text View (mencari model regresi)

## LinearRegression

$$21.608 * \text{LAMA BELAJAR (JAM)} \\ + 492.769$$

Model regresi linier yang terbentuk :

$$Y = 21,608 X_1 + 492,769$$

Dengan model tersebut, dapat dicari Nilai Ujian (Y) dengan memasukkan nilai Lama Belajar pada variabel X1.

| NO_SISWA | NAMA   | LAMA BELAJAR (JAM) | NILAI | Nilai Belajar |
|----------|--------|--------------------|-------|---------------|
| S-101    | JOKO   | 15                 | 783   | 816,889       |
| S-102    | AGUS   | 18                 | 877   | 881,713       |
| S-103    | SUSI   | 7                  | 505   | 644,025       |
| S-104    | DYAH   | 9                  | 860   | 687,241       |
| S-105    | WATI   | 15                 | 968   | 816,889       |
| S-106    | IKA    | 17                 | 793   | 860,105       |
| S-107    | EKO    | 10                 | 752   | 708,849       |
| S-108    | YANTO  | 5                  | 571   | 600,809       |
| S-109    | WAWAN  | 8                  | 667   | 665,633       |
| S-110    | MAHMUD | 15                 | 723   | 816,889       |

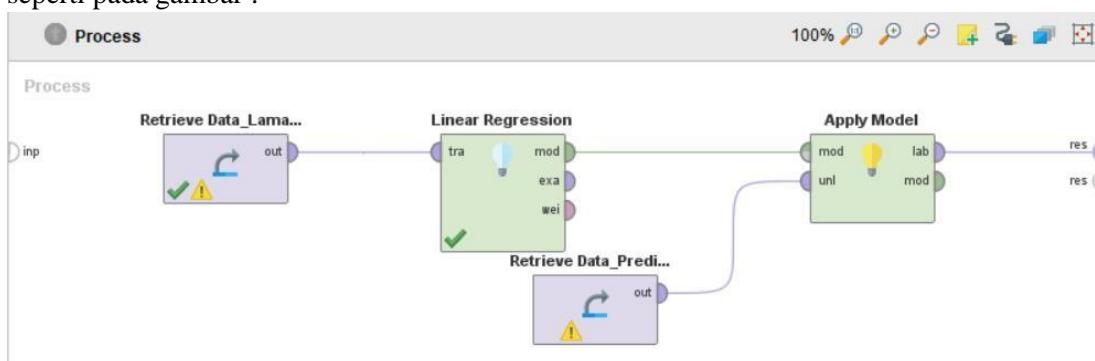
### Mencari Nilai t- dan Model Regresi Linier Menggunakan RapidMiner

- Buka Ms.Excel, dan buatlah tabel data siswa tersebut. Simpan dengan nama **Tabel\_PrediksiNilaiUjian.xls**
- Jalankan **RapidMiner**.

3. Gunakan file **Tabel\_PrediksiNilaiUjian.xls** sebagai data testing. Import file ke dalam repository.
4. Ubah tipe data dan jenis masing-masing atribut sebagai berikut :

| NO_SISWA<br>polynominal<br>id | NAMA<br>polynominal | LAMA BELAJAR (JAM)<br>integer |
|-------------------------------|---------------------|-------------------------------|
| 1 S-111                       | BUDI                | 12                            |
| 2 S-112                       | SANTI               | 13                            |
| 3 S-113                       | DIAN                | 14                            |
| 4 S-114                       | DANI                | 11                            |
| 5 S-115                       | AHMAD               | 5                             |
| 6 S-116                       | BAYU                | 13                            |
| 7 S-117                       | RISA                | 9                             |
| 8 S-118                       | RANI                | 10                            |
| 9 S-119                       | YANI                | 10                            |
| 10 S-120                      | RATIH               | 9                             |

5. Pada langkah ini, beri nama **Data\_PrediksiNilaiUjian** dan masukkan pada repository, klik **Finish**.
6. Tetap menggunakan proses pada kegiatan sebelumnya, masukkan **Data\_PrediksiNilaiUjian** ke dalam area process.
7. Tambahkan **Scoring→Apply Model**, letakkan setelah Linier Regression. Hubungkan port seperti pada gambar :

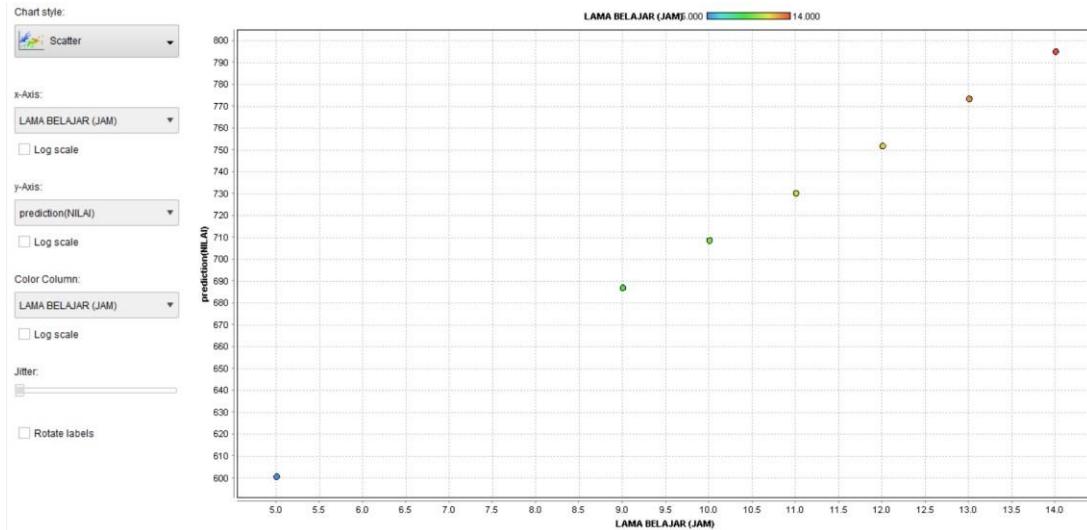


8. Dengan menggunakan parameter yang sama pada operator **Regression Linear**, jalankan proses dengan tombol **Run**.
9. Hasil proses prediksi terhadap data testing menggunakan regresi linear :
  - a. Data View (hasil prediksi nilai ujian)

ExampleSet (10 examples, 2 special attributes, 1 regular attribute)

| Row No. | NO_SISWA | prediction(N...) | LAMA BELA... |
|---------|----------|------------------|--------------|
| 1       | S-111    | 752.061          | 12           |
| 2       | S-112    | 773.668          | 13           |
| 3       | S-113    | 795.276          | 14           |
| 4       | S-114    | 730.453          | 11           |
| 5       | S-115    | 600.807          | 5            |
| 6       | S-116    | 773.668          | 13           |
| 7       | S-117    | 687.238          | 9            |
| 8       | S-118    | 708.845          | 10           |
| 9       | S-119    | 708.845          | 10           |
| 10      | S-120    | 687.238          | 9            |

b. Chart View (Scatter Plot)



Pembuktian Model Regresi Y

$$= 21,608 X_1 + 492,769$$

| NO_SISWA | NAMA  | LAMA BELAJAR (JAM) | Prediction (NILAI) | Prediction (NILAI) |
|----------|-------|--------------------|--------------------|--------------------|
|          |       |                    | Tabel              | Model Regresi      |
| S-111    | BUDI  | 12                 | 752,061            | 752,065            |
| S-112    | SANTI | 13                 | 773,668            | 773,673            |
| S-113    | DIAN  | 14                 | 795,276            | 795,281            |
| S-114    | DANI  | 11                 | 730,453            | 730,457            |
| S-115    | AHMAD | 5                  | 600,807            | 600,809            |
| S-116    | BAYU  | 13                 | 773,668            | 773,673            |
| S-117    | RISA  | 9                  | 687,238            | 687,241            |
| S-118    | RANI  | 10                 | 708,845            | 708,849            |
| S-119    | YANI  | 10                 | 708,845            | 708,849            |
| S-120    | RATIH | 9                  | 687,238            | 687,241            |

Dapat dilihat bahwa hasil prediksi menggunakan RapidMiner menghasilkan nilai yang sama dengan menggunakan Model Persamaan Regresi Linier.

## Tugas

- Buatlah tabel dengan menggunakan Microsoft Excel seperti berikut :

| NO RESPONDEN | PENDAPATAN (RUPIAH) | JUMLAH ANGGOTA KELUARGA | DAYA BELI (RUPIAH) |
|--------------|---------------------|-------------------------|--------------------|
| 1            | 1000000             | 6                       | 834000             |
| 2            | 1400000             | 7                       | 1200000            |
| 3            | 200000              | 3                       | 134000             |
| 4            | 1400000             | 6                       | 1167000            |
| 5            | 500000              | 3                       | 334000             |
| 6            | 1700000             | 5                       | 1360000            |
| 7            | 400000              | 3                       | 267000             |
| 8            | 1900000             | 5                       | 1520000            |
| 9            | 300000              | 3                       | 200000             |
| 10           | 500000              | 4                       | 375000             |
| 11           | 700000              | 7                       | 600000             |
| 12           | 1900000             | 3                       | 1267000            |
| 13           | 800000              | 4                       | 600000             |
| 14           | 1500000             | 4                       | 1125000            |
| 15           | 1300000             | 7                       | 1115000            |

- Buatlah proses Regresi Linier Sederhana menggunakan RapidMiner dengan ketentuan :
  - Variable Bebas(X) = Pendapatan(X1), Jumlah Anggota Keluarga(X2)
  - Variable Terikat(Y) = Daya Beli
  - Toleransi yang digunakan = 5%

| Attribute             | Coefficient | Std. Error | Std. Coefficient | Tolerance | t-Stat | p-Value | Code |
|-----------------------|-------------|------------|------------------|-----------|--------|---------|------|
| JUMLAH ANGGOTA KEL... | 151688.679  | 70801.481  | 0.511            | ?         | 2.142  | 0.052   | *    |
| (Intercept)           | 98652.830   | 347817.155 | ?                | ?         | 0.284  | 0.781   |      |

## Linear Regression

$$151688.679 * \text{JUMLAH ANGGOTA KELUARGA} + 98652.830$$

- Tentukan apakah variable X1 dan X2 mempengaruhi secara signifikan terhadap nilai variable Y berdasarkan nilai t-stat?  
Dari tabel tersebut dapat dilihat nilai t-stat sebesar 2,142. Berdasarkan aturan statistik, variabel X dikatakan mempengaruhi secara signifikan terhadap Y jika nilai t-stat > t-table.
- Tuliskan model persamaan regresi linier sederhana yang terbentuk!  
**Y = 151688,679 X2 + 98652,830**

5. Gunakan data testing untuk menjawab perintah :
    - a. Lakukan prediksi Daya Beli (Y) dengan menggunakan Model Persamaan Regresi Linier dari hasil pertanyaan nomor 4!
    - b. Lakukan prediksi Daya Beli (Y) menggunakan RapidMiner!

| NO. RESPONDEN | PENDAPATAN (RUPIAH) | JUMLAH ANGGOTA KELUARGA | Prediction (NILAI) |             |
|---------------|---------------------|-------------------------|--------------------|-------------|
|               |                     |                         | Model Regresi      | Tabel       |
| 1             | 900000              | 5                       | 857096,225         | 857096,226  |
| 2             | 800000              | 3                       | 553718,867         | 553718,868  |
| 3             | 500000              | 2                       | 402030,188         | 402030,189  |
| 4             | 1900000             | 6                       | 1008784,904        | 1008784,906 |
| 5             | 600000              | 2                       | 402030,188         | 402030,189  |
| 6             | 800000              | 5                       | 857096,225         | 857096,226  |
| 7             | 1000000             | 6                       | 1008784,904        | 1008784,906 |
| 8             | 1100000             | 4                       | 705407,546         | 705407,547  |
| 9             | 1000000             | 4                       | 705407,546         | 705407,547  |
| 10            | 500000              | 3                       | 553718,867         | 553718,868  |

6. Gambarkan pola sebaran data menggunakan Plot View (Scatter)

