

LAPORAN PRAKTIKUM  
DATA WAREHOUSING DAN DATA MINING



Disusun oleh:  
Sidiq Zainudin [L200170058]

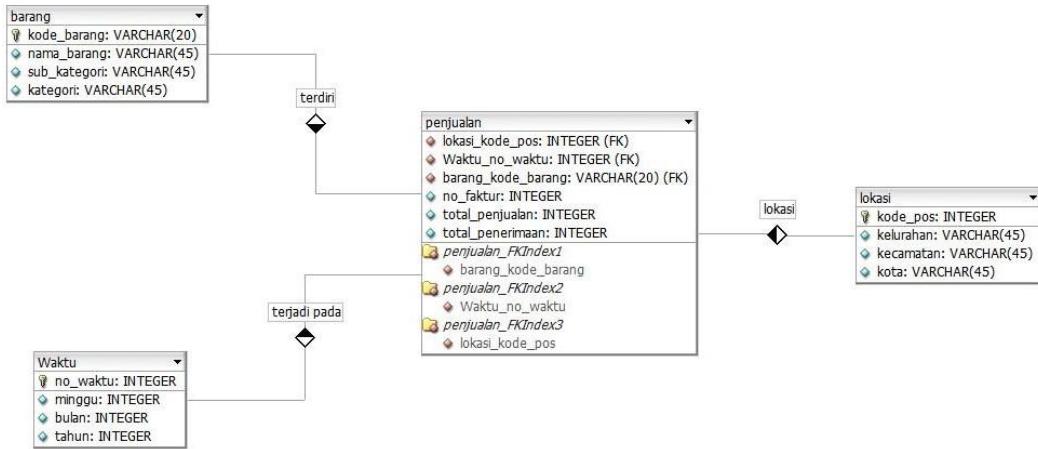
Dosen Pengampu:  
Yudi Wahyu Wibowo, S.T.,M.Eng

PRORAM STUDI INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA  
2019

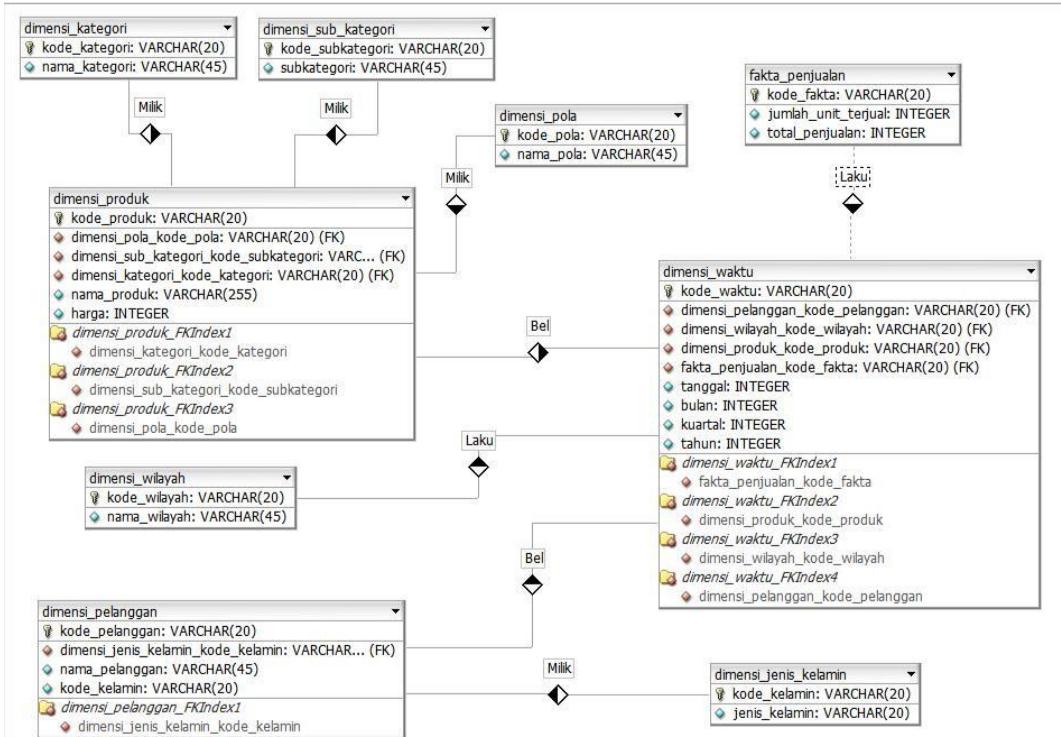
# MODUL I

## PERANCANGAN STAR SCHEMA DAN SNOWFLAKE

### D. Langkah-langkah Praktikum



### E. Tugas



## **MODUL V**

# **PIVOT TABLE DAN CHART**

## **1. Langkah-langkah Praktikum**

#### **D. 1. Kegiatan 1 : Membuat Pivot Table**

1. Membuat data Excel dan menyimpannya dengan nama ‘**Fakta\_Penjualan.xls**’.

## 2. Membuat Pivot Table dengan menggunakan data tersebut.

The screenshot shows the 'Create PivotTable' dialog box in Microsoft Excel. The 'Select a table or range' option is selected, with the range set to 'Sheet1!\$A\$1:\$K\$21'. The 'New Worksheet' option is also selected. The background displays a table with columns labeled 'bulan', 'kuartal', 'tahun', and 'nama\_produk'.

## 3. Menampilkan data Pivot Table dengan mencoba kombinasi tahun, nama\_subkategori, dan jumlah.

The screenshot shows a PivotTable named 'PivotTable1' on 'Sheet2'. The PivotTable displays data grouped by year ('2010', '2011', '2012') and subcategory ('Bahan', 'Batik', etc.). The 'Sum of jumlah' field is used as the value. The PivotTable Tools ribbon tab is active, showing various analysis tools like Group Selection, Insert Slicer, and Refresh Data.

## D. 2. Kegiatan 2 : Menambahkan Tipe Summary Baru

- Menambahkan field **jumlah** ke kotak **Value** (karena sudah ada field jumlah sebelumnya, maka field dengan nama yang sama di kotak yang sama akan memiliki tambahan karakter berupa angka di akhiran kata, pada kasus ini **jumlah2** atau **sum of jumlah2**).

Fakta\_Penjualan - Excel

PivotTable Tools

PivotTable Name: Active Field: PivotTable1

PivotTable Options: Active Field: PivotTable

PivotTable Fields

Choose fields to add to report:

- jenis\_kelamin
- nama\_wilayah
- jumlah
- harga
- Pendapatan

Drag fields between areas below:

Filters

Columns: tahun, Σ Values

Rows: nama\_subkat..., Σ Values: Sum of jumlah, Count of jumlah2

Defer Layout Update

	2010	2011	2012	Total Sum of jumlah	Total Count of jumlah2	
Row Labels	Sum of jumlah	Count of jumlah2	Sum of jumlah	Count of jumlah2	Sum of jumlah	Count of jumlah2
Bahan	1	1	8	1	8	2
Batik					1	1
Bolero				1	1	
Celana	17	1			17	1
Hem	5	1	8	2	4	2
Jam					44	1
Jarik					4	1
kaos				1	14	1
Rok					1	1
Sarimbit				1	1	
Standar			2	1		2
Grand Total	23	3	21	7	93	10
					137	20

- Mengubah jumlah nilai penjualan yang terjadi (**sum**) pada **jumlah2**, menjadi jumlah transaksi yang terjadi (**count**).

Fakta\_Penjualan - Excel

PivotTable Tools

Value Field Settings

Source Name: jumlah

Custom Name: Count of jumlah2

Summarize Values By: Show Values As

Summarize value field by: Count

Choose the type of calculation that you want to use to summarize data from the selected field:

- Sum
- Count
- Average
- Max
- Min
- Product

Number Format: OK Cancel

PivotTable Fields

Choose fields to add to report:

- jenis\_kelamin
- nama\_wilayah
- jumlah
- harga
- Pendapatan

Drag fields between areas below:

Filters

Columns: tahun, Σ Values

Rows: nama\_subkat..., Σ Values: Sum of jumlah, Count of jumlah2

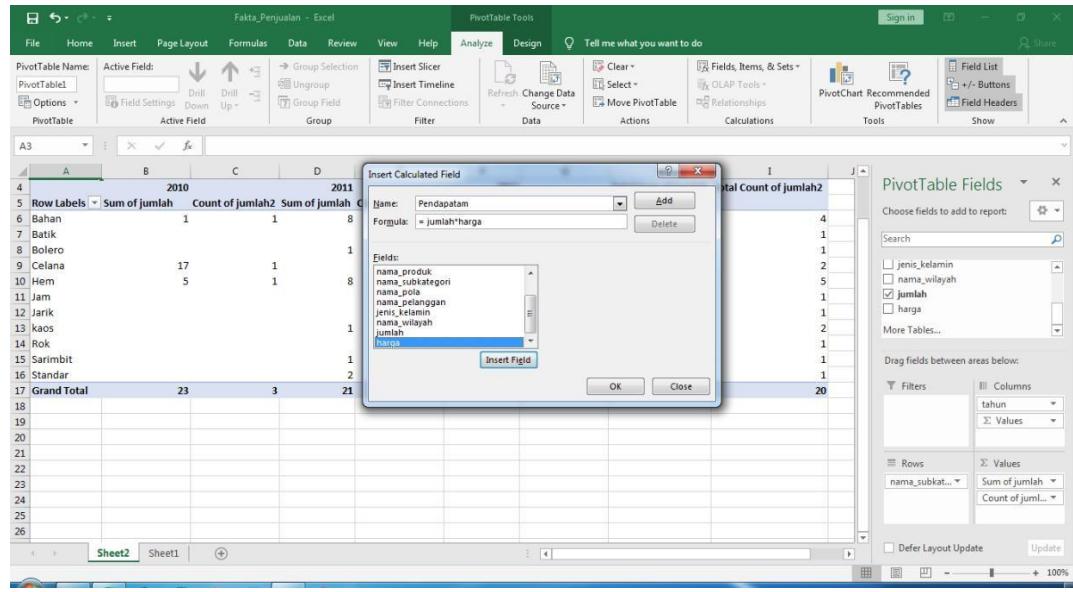
Defer Layout Update

	2010	2011	2012	Total Sum of jumlah2
Row Labels	Sum of jumlah	Sum of jumlah2	Sum of jumlah	Sum of jumlah2
Bahan	1	1	8	17
Batik				1
Bolero				1
Celana	17	17		34
Hem	5	5	8	17
Jam				44
Jarik				4
kaos			1	15
Rok				1
Sarimbit			1	1
Standar			2	2
Grand Total	23	23	21	137

## D. 3. Kegiatan 3 : Calculated Field dan Calculated Item di Pivot Table

### 1. Calculated Field (Menambah field / kolom baru pada daftar field yang ada )

1. Membuat field baru dengan menggunakan tab **Formulas -> Calculated Field**, lalu membuat field baru dengan nama **Pendapatan** dengan formula, **jumlah \* harga**.



2. Setelah meng-klik tombol ‘OK’ maka akan muncul file baru dengan nama ‘Pendapatan’, cukup centang field tersebut dan data akan otomatis keluar.

	2010		2011		2012		
Row Labels	Sum of jumlah	Count of jumlah2	Sum of Pendapatan	Sum of jumlah	Count of jumlah2	Sum of Pendapatan	Sum of jumlah
Bahan	1	1	500000	8	1	960000	8
Batik			0			0	1
Bolero			0	1	1	255000	
Celana	17	1	935000			0	17
Hem	5	1	500000	8	2	4960000	4
Jam			0			0	44
Jarik			0			0	4
kaos			0	1	1	60000	14
Rok			0			0	1
Sarimbit			0	1	1	150000	
Standar			0	2	1	450000	
<b>Grand Total</b>	<b>23</b>	<b>3</b>	<b>15065000</b>	<b>21</b>	<b>7</b>	<b>30030000</b>	<b>93</b>

		Total Sum of jumlah	Total Count of jumlah2	Total Sum of Pendapatan
Count of jumlah2	Sum of Pendapatan			
2	2120000		17	15045000
1	150000		1	150000
	0		1	255000
1	935000		34	3740000
2	1596000		17	19023000
1	3520000		44	3520000
1	160000		4	160000
1	420000		15	1350000
1	225000		1	225000
	0		1	150000
	0		2	450000
<b>10</b>	<b>115692000</b>		<b>137</b>	<b>456073000</b>

#### D. 4. Kegiatan 4 : Operasi Roll Up dan Drill Down

- Menghilangkan beberapa field, sehingga tinggal tersisa **tahun** (Columns), **nama\_subkategori** (Rows), dan **Sum of Pendapatan** (Value).

The screenshot shows a Microsoft Excel spreadsheet titled "Fakta\_Penjualan - Excel". A PivotTable is being used to analyze data. The PivotTable Fields pane on the right indicates that the "Pendapatan" field is currently selected. The main table area displays data for various items (Bahan, Batik, Bolero, Celana, Hem, Jam, Jarik, kaos, Rok, Sarimbit, Standar) across three years (2010, 2011, 2012) and a Grand Total. The columns represent the years, and the rows represent the items. The values in the cells are the sum of pendapatan for each item per year.

	2010	2011	2012	Grand Total
Bahan	500000	960000	2120000	15045000
Batik	0	0	150000	150000
Bolero	0	255000	0	255000
Celana	935000	0	935000	3740000
Hem	500000	4960000	15960000	19023000
Jam	0	0	3520000	3520000
Jarik	0	0	160000	160000
kaos	0	60000	420000	1350000
Rok	0	0	225000	225000
Sarimbit	0	150000	0	150000
Standar	0	450000	0	450000
<b>Grand Total</b>	<b>15065000</b>	<b>30030000</b>	<b>115692000</b>	<b>456073000</b>

2. Menambahkan field pada kotak **Column** (**kuartal** dan **bulan**) dan **Rows** (**nama\_produk** dan **nama\_pelanggan**), untuk menambah field-field yang bisa diperinci dan diringkas sesuai urutan kategori data yang lebih spesifik.

## Roll Up

	2010	2011	2012	Grand Total
<b>Bahan</b>	500000	960000	2120000	15045000
<b>Batik</b>	0	0	150000	150000
<b>Bolero</b>	0	255000	0	255000
<b>Celana</b>	935000	0	935000	3740000
<b>Hem</b>	500000	4960000	1596000	19023000
<b>Jam</b>	0	0	3520000	3520000
<b>Jarik</b>	0	0	160000	160000
<b>Kaos</b>	0	60000	420000	1350000
<b>Rok</b>	0	0	225000	225000
<b>Sarimbit</b>	0	150000	0	150000
<b>Standar</b>	0	450000	0	450000
<b>Grand Total</b>	15065000	30030000	115692000	456073000

## Drop Down

	2010	2011	2012	Grand Total
<b>Bahan</b>	500000	960000	2120000	15045000
<b>Bahan Belundru Cap Mahkota</b>	0	0	945000	945000
Ibu Tyas	0	0	945000	945000
<b>Bahan Standar Cap Garis</b>	500000	0	0	500000
Ibu Tyas	500000	0	0	500000
<b>Bahan Standar Cap Lasem</b>	0	0	130000	130000
Ibu Niken	0	0	130000	130000
<b>Hem Sutra Print Rama</b>	0	960000	0	960000
Ibu Siti Arya	0	960000	0	960000
<b>Batik</b>	0	0	150000	150000
<b>Celana Standar Cap Warna</b>	0	0	150000	150000
Bapak Heru	0	0	150000	150000
<b>Bolero</b>	0	255000	0	255000
<b>Kaos Batik Cap Tulis</b>	0	255000	0	255000
Ibu Hatamah	0	255000	0	255000
<b>Celana</b>	935000	0	935000	3740000
<b>Hem Standar Cap Tumpal</b>	935000	0	0	935000
Ibu Hadi Sukarni	935000	0	0	935000
<b>Sarimbit Standar Print Lukis</b>	0	0	935000	935000
Ibu Hadi Sukarni	0	0	935000	935000
<b>Hem</b>	500000	4960000	1596000	19023000

Falta\_Perjalanan - Excel

PivotTable Tools

Choose fields to add to report:

- bulan
- kuartal
- tahun
- nama\_produk
- nama\_subkategori

Drag fields between areas below:

Filters

Columns: tahun

Rows: nama\_subkategori

Values: Sum of Penda...

Cells

Defer Layout Update

Jam

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
26	=Bahan Lawasan Tulis Tolet		0	2750000	0	2750000								
27	Ibu Atik		0	2750000	0	2750000								
28	=Batik Standar Cap Tumpal		0	210000	0	210000								
29	Ibu Harini		0	210000	0	210000								
30	=Bolero Standar Cap Sidomukti	500000	0	0	0	500000								
31	Ibu Tyas	500000	0	0	0	500000								
32	=Celana Standar Print Lasem	0	0	897000	0	897000								
33	Bapak Totok	0	0	897000	0	897000								
34	=Hem katun Print Kawung	0	0	100000	0	100000								
35	Ibu Aini Kasmaji	0	0	100000	0	100000								
36	<b>Jam</b>	0	0	3520000	0	3520000								
37	=Hem katun Print Kelengan	0	0	3520000	0	3520000								
38	Ibu Siti Arya	0	0	3520000	0	3520000								
39	<b>Jarik</b>	0	0	160000	0	160000								
40	=Rok Batik Print Kombinasi	0	0	160000	0	160000								
41	Ibu Harini	0	0	160000	0	160000								
42	<b>Kaos</b>	0	60000	420000	0	1350000								
43	=Jarn Standar Print Lukis	0	60000	0	0	60000								
44	Bapak Imron	0	60000	0	0	60000								
45	=Jarik Standar Tulis Sarimbint	0	0	420000	0	420000								
46	Ibu Harini	0	0	420000	0	420000								
47	<b>Rok</b>	0	0	225000	0	225000								
48	=Hem Standar Tulis Madura	0	0	225000	0	225000								

Falta\_Perjalanan - Excel

PivotTable Tools

Choose fields to add to report:

- bulan
- kuartal
- tahun
- nama\_produk
- nama\_subkategori

Drag fields between areas below:

Filters

Columns: tahun

Rows: nama\_subkategori

Values: Sum of Penda...

Cells

Defer Layout Update

Ibu Siti Arya

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
49	Ibu Siti Arya		0	0	225000	0	225000							
50	<b>Sarimbint</b>	0	150000	0	0	150000								
51	=Kaos Katun Print Bola	0	150000	0	0	150000								
52	Ibu Hatamah	0	150000	0	0	150000								
53	<b>Standar</b>	0	450000	0	0	450000								
54	=Jarik Standar Print Sogan	0	450000	0	0	450000								
55	Bapak Ketut	0	450000	0	0	450000								
56	<b>Grand Total</b>	15065000	30030000	0	115692000	0	456073000							

## 5. Tugas

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

1. **PPN** (Pajak Pertambahan Nilai) sebesar 10% dari tiap pendapatan pada Pivot Table.
2. **Total Penghasilan** yang dihitung dari pendapatan dikurangi dengan PPN tersebut.

Row Labels	Column Labels		2010		2011		
	Sum of Pendapatan	Sum of PPN (10%)	Sum of Total Penghasilan	Sum of jumlah	Sum of kuartal	Sum of Pendapatan	Sum of PPN (10%)
Bahan	500000	50000	450000	1	2	960000	96000
Batik	0	0	0			0	0
Bolero	0	0	0			255000	25500
Celana	935000	93500	841500	17	1	0	0
Hem	500000	50000	450000	5	4	4960000	496000
Jam	0	0	0			0	0
Jarik	0	0	0			0	0
kaos	0	0	0			60000	6000
Rok	0	0	0			0	0
Sarimbit	0	0	0			150000	15000
Standar	0	0	0			450000	45000
<b>Grand Total</b>	<b>15065000</b>	<b>1506500</b>	<b>13558500</b>	<b>23</b>	<b>7</b>	<b>30030000</b>	<b>3003000</b>

2012							
Sum of Total Penghasilan	Sum of jumlah	Sum of kuartal	Sum of Pendapatan	Sum of PPN (10%)	Sum of Total Penghasilan	Sum of jumlah	Sum of kuartal
864000	8	1	2120000	212000	1908000	8	4
0			150000	15000	135000	1	3
229500	1	4	0	0	0		
0			935000	93500	841500	17	1
4464000	8	5	1596000	159600	1436400	4	5
0			3520000	352000	3168000	44	1
0			160000	16000	144000	4	2
54000	1	1	420000	42000	378000	14	1
0			225000	22500	202500	1	4
135000	1	4	0	0	0		
405000	2	4	0	0	0		
<b>27027000</b>	<b>21</b>	<b>19</b>	<b>115692000</b>	<b>11569200</b>	<b>104122800</b>	<b>93</b>	<b>21</b>

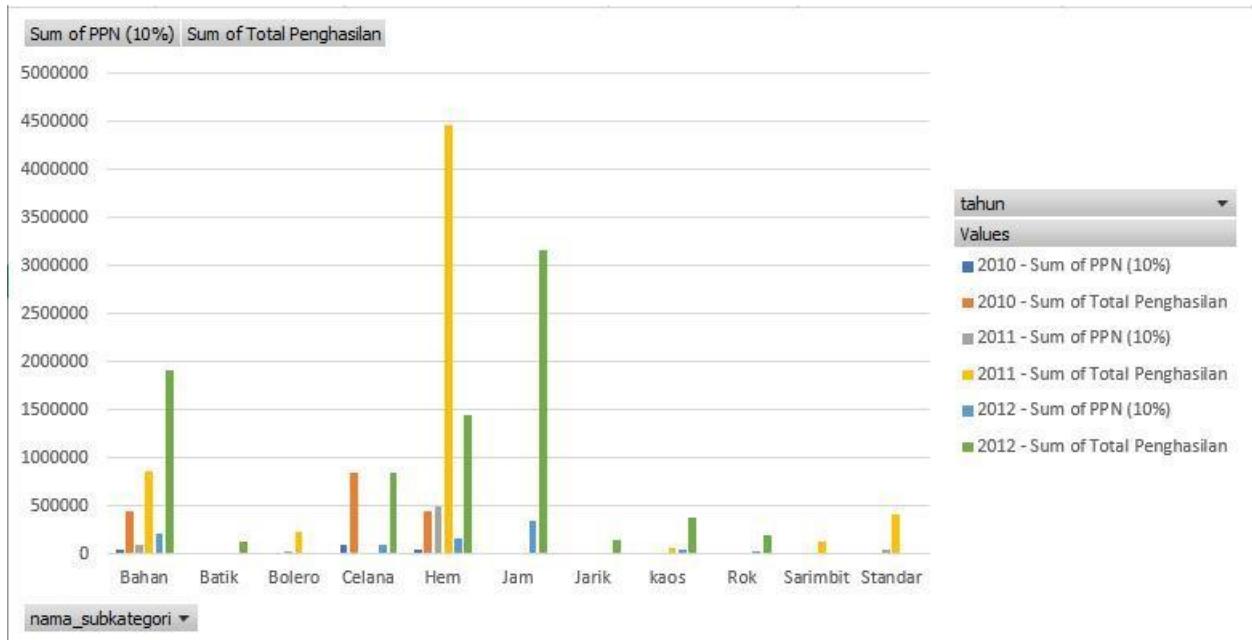
Total Sum of Pendapatan	Total Sum of PPN (10%)	Total Sum of Total Penghasilan	Total Sum of jumlah	Total Sum of kuartal
15045000	1504500	13540500	17	7
150000	15000	135000	1	3
255000	25500	229500	1	4
3740000	374000	3366000	34	2
19023000	1902300	17120700	17	14
3520000	352000	3168000	44	1
160000	16000	144000	4	2
1350000	135000	1215000	15	2
225000	22500	202500	1	4
150000	15000	135000	1	4
450000	45000	405000	2	4
<b>456073000</b>	<b>45607300</b>	<b>410465700</b>	<b>137</b>	<b>47</b>

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

### PivotTable

	Column Labels	2010	2011	2012	Total Sum of PPN (10%)	Total Sum of Total Penghasilan
Row Labels		Sum of PPN (10%)	Sum of Total Penghasilan	Sum of PPN (10%)	Sum of Total Penghasilan	
Bahan	50000	450000	96000	864000	212000	1908000
Batik	0	0	0	0	15000	135000
Bolero	0	0	25500	229500	0	25500
Celana	93500	841500	0	0	93500	841500
Hem	50000	450000	496000	4464000	159600	1436400
Jam	0	0	0	0	352000	3168000
Jarik	0	0	0	0	16000	144000
kaos	0	0	6000	54000	42000	378000
Rok	0	0	0	0	22500	22500
Sarimbit	0	0	15000	135000	0	0
Standar	0	0	45000	405000	0	45000
<b>Grand Total</b>	<b>1506500</b>	<b>13558500</b>	<b>3003000</b>	<b>27027000</b>	<b>11569200</b>	<b>104122800</b>
						<b>45607300</b>
						<b>410465700</b>

### PivotChart



Dapat dilihat dari data di atas bahwa, kategori produk dengan nilai penghasilan terbanyak pada tahun 2010 – 2012 adalah **Hem**.

+.....+.....+

## MODUL VI

### PENGENALAN DATA MINING

#### 6.5 Tugas (Dikerjakan Saat Ini)

2. Dengan menggunakan formula dalam Ms. Excel, carilah :

Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten	Lama_Studi	JURUSAN_SMA	TOTAL
IPS	WANITA	SURAKARTA	18	TIDAK	TERLAMBAT	IPA	10
IPA	PRIA	SURAKARTA	19	YA	TEPAT	IPS	6
LAIN	PRIA	SURAKARTA	19	TIDAK	TERLAMBAT	LAIN	4
IPA	PRIA	LUAR	17	TIDAK	TERLAMBAT		
IPA	WANITA	SURAKARTA	17	TIDAK	TEPAT	LAMA_STUDI	TOTAL
IPA	WANITA	LUAR	18	YA	TEPAT	TERLAMBAT	13
IPA	PRIA	SURAKARTA	18	TIDAK	TERLAMBAT	TEPAT	7
IPA	PRIA	SURAKARTA	19	TIDAK	TEPAT	CARI	JUMLAH
IPS	PRIA	LUAR	18	TIDAK	TERLAMBAT	MAX	23
LAIN	WANITA	SURAKARTA	18	TIDAK	TEPAT	MIN	16
IPA	WANITA	SURAKARTA	19	TIDAK	TEPAT	MEAN	18,95
IPS	PRIA	SURAKARTA	20	TIDAK	TEPAT	STANDAR DEVIASI	1,66938375
IPS	PRIA	SURAKARTA	19	TIDAK	TEPAT		
IPA	PRIA	SURAKARTA	19	TIDAK	TEPAT	DATA	TOTAL
IPA	PRIA	LUAR	22	YA	TEPAT	JURUSAN_SMA	10
LAIN	PRIA	SURAKARTA	16	TIDAK	TERLAMBAT	GENDER	15
IPS	PRIA	LUAR	20	TIDAK	TEPAT	ASISTEN	5
LAIN	PRIA	LUAR	23	YA	TEPAT	LAMA_STUDI	13
IPA	PRIA	SURAKARTA	21	YA	TEPAT	JUMLAH	43
IPS	PRIA	SURAKARTA	19	TIDAK	TERLAMBAT		

D. Pada attribut **Jurusian\_SMA**, berapa jumlah data masing-masing kelas **IPA**, **IPS**, dan **LAIN**? (Gunakan formula = COUNTIF)

- Tabel Kanan Paling Atas

E. Pada attribut **Lama\_Studi**, berapa jumlah data masing-masing kelas **Tepat**, **Terlambat**?

- Tabel Kanan Nomor 2 dari Atas

F. Pada attribute **Rerata\_SKS**, berapa kali Max, Min, Mean, dan Standard Deviation?

- Tabel Kanan Nomor 2 dari Bawah

G. Pada tabel tersebut, berapakah jumlah data gabungan untuk kelas pada attribut **Jurusan\_Studi = IPA**, **Gender = PRIA**, **Asisten = Ya**, **Lama\_Studi = TEPAT**? (Gunakan formula = COUNTIF)

- Tabel Kanan Paling Bawah

+ ..... +

## MODUL VII

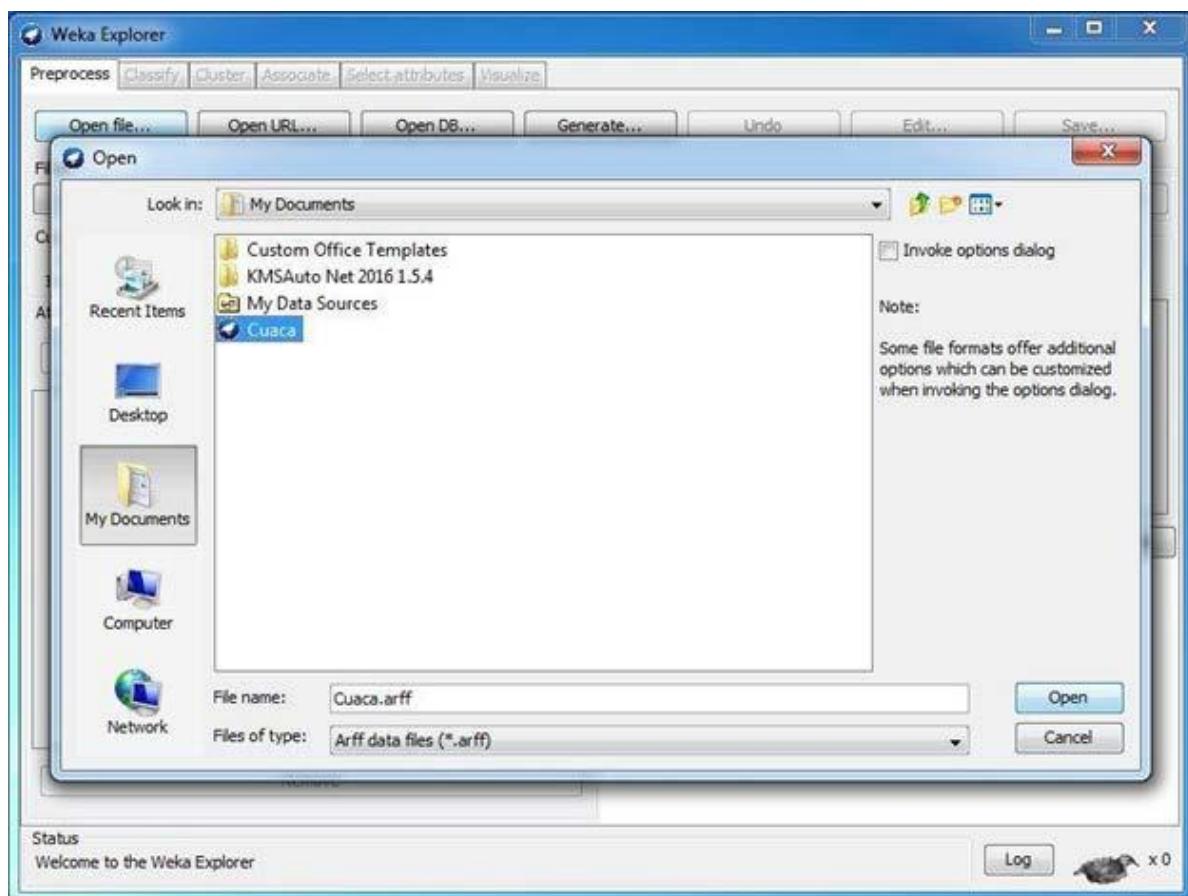
### DATA PREPROCESSING

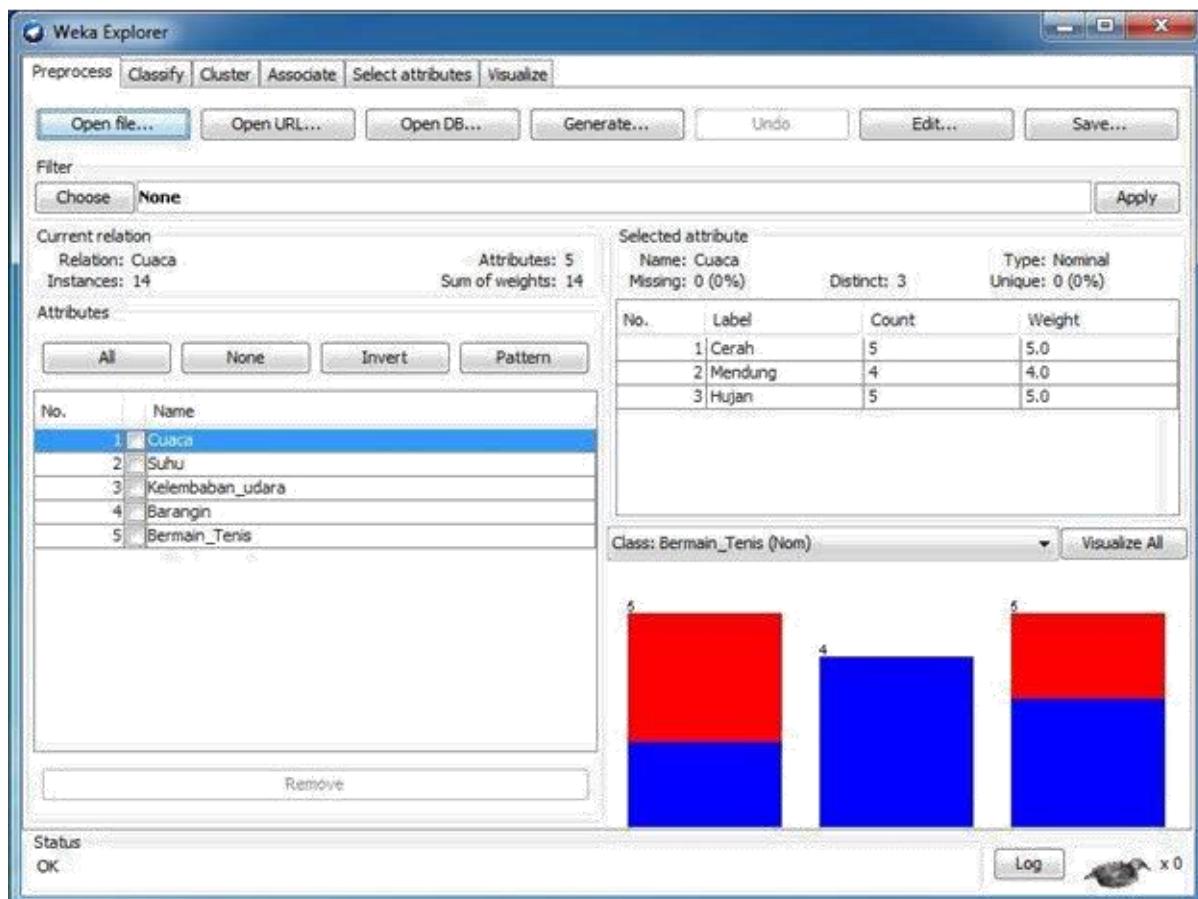
#### Langkah-langkah Praktikum

##### Membuat Format Data ARFF (Attribut-Relation File Format)

```
≡ Cuaca.arff ×  
C: > Users > LABSI-21 > Documents > ≡ Cuaca.arff  
1  @relation Cuaca  
2  
3  @attribute Cuaca {Cerah, Mendung, Hujan}  
4  @attribute Suhu real  
5  @attribute Kelembaban_udara real  
6  @attribute Barangin {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
```

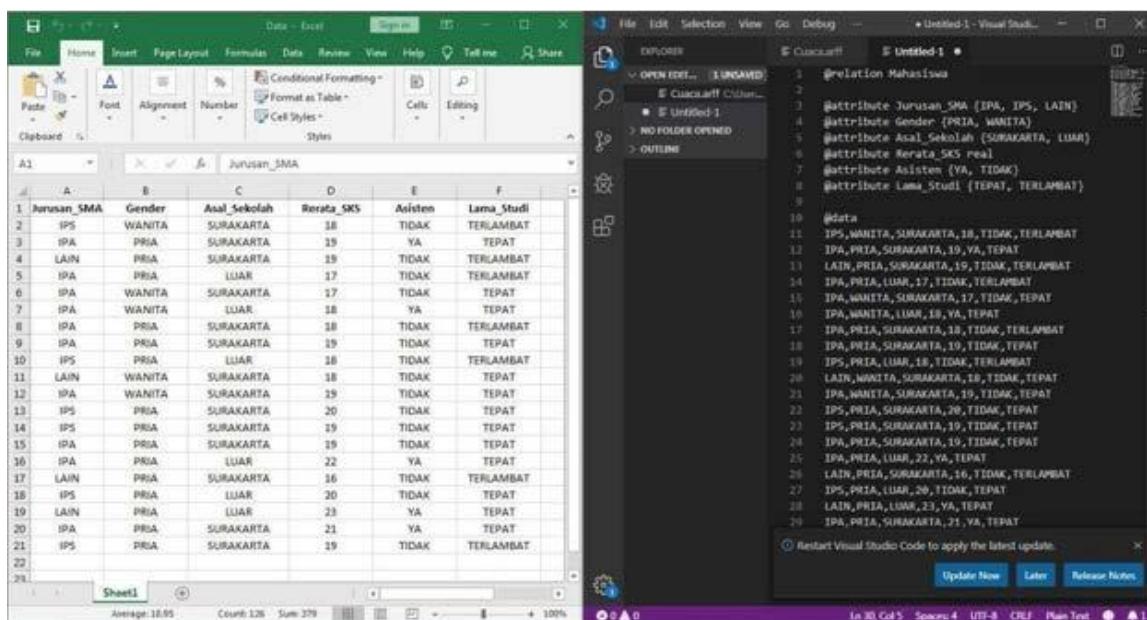
## Menggunakan File ARFF dengan WEKA





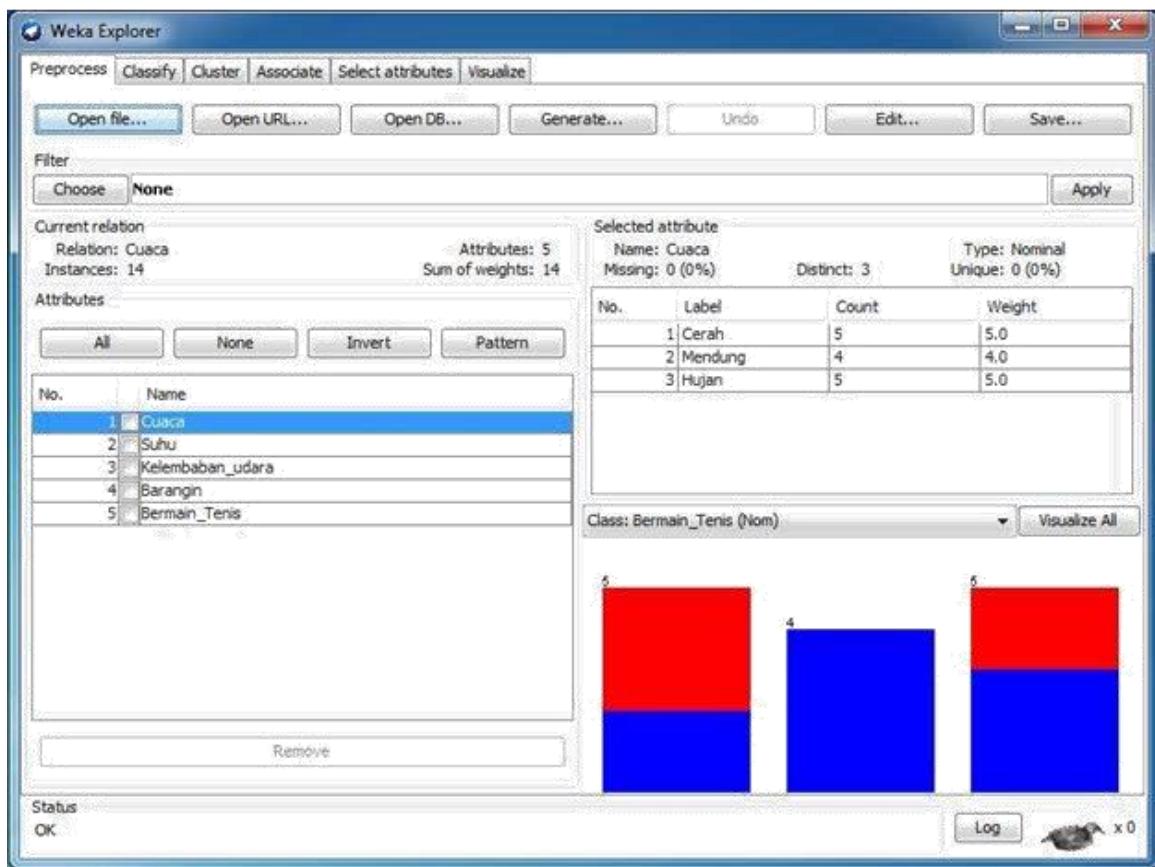
## 7.5. Tugas

- Buat file ARFF berdasarkan tugas pada Modul 6 soal nomor 1!



2. Perlihatkan file ARFF dan juga gambar grafik untuk setiap data yang ditampilkan dalam Weka berdasarkan file ARFF anda kepada dosen dan sistem untuk dinilai!

```
≡ Mahasiswa.arff ×  
C: > Users > LABSI-21 > Documents > ≡ Mahasiswa.arff  
1  @relation Mahasiswa  
2  
3  @attribute Jurusan_SMA {IPA, IPS, LAIN}  
4  @attribute Gender {PRIA, WANITA}  
5  @attribute Asal_Sekolah {SURAKARTA, LUAR}  
6  @attribute Rerata_SKS real  
7  @attribute Asisten {YA, TIDAK}  
8  @attribute Lama_Studi {TEPAT, TERLAMBAT}  
9  
10 @data  
11  IPS,WANITA,SURAKARTA,18,TIDAK,TERLAMBAT  
12  IPA,PRIA,SURAKARTA,19,YA,TEPAT  
13  LAIN,PRIA,SURAKARTA,19,TIDAK,TERLAMBAT  
14  IPA,PRIA,LUAR,17,TIDAK,TERLAMBAT  
15  IPA,WANITA,SURAKARTA,17,TIDAK,TEPAT  
16  IPA,WANITA,LUAR,18,YA,TEPAT  
17  IPA,PRIA,SURAKARTA,18,TIDAK,TERLAMBAT  
18  IPA,PRIA,SURAKARTA,19,TIDAK,TEPAT  
19  IPS,PRIA,LUAR,18,TIDAK,TERLAMBAT  
20  LAIN,WANITA,SURAKARTA,18,TIDAK,TEPAT  
21  IPA,WANITA,SURAKARTA,19,TIDAK,TEPAT  
22  IPS,PRIA,SURAKARTA,20,TIDAK,TEPAT  
23  IPS,PRIA,SURAKARTA,19,TIDAK,TEPAT  
24  IPA,PRIA,SURAKARTA,19,TIDAK,TEPAT  
25  IPA,PRIA,LUAR,22,YA,TEPAT  
26  LAIN,PRIA,SURAKARTA,16,TIDAK,TERLAMBAT  
27  IPS,PRIA,LUAR,20,TIDAK,TEPAT  
28  LAIN,PRIA,LUAR,23,YA,TEPAT  
29  IPA,PRIA,SURAKARTA,21,YA,TEPAT  
30  IPS,PRIA,SURAKARTA,19,TIDAK,TERLAMBAT
```



3. Tulis di atas selembar kerta HVS (mintalah kepada asisten)

4. Berapa jumlah atribut yang bertipe binomial dan polynomial?

- └ Binomial : 4 (Gender, Asal\_Sekolah, Asisten, Lama\_Studi)
- └ Polynomial : 1 (Jurusan\_SMA)

5. Berapa jumlah atribut yang bertipe real?

- └ Real : 1 (Rerata\_SKS)

6. Pada atribut Rerata\_SKS, berapakah besarnya nilai Maximum, Minimum, Mean, StdDev (Standard Deviation)?

- └ Maximum : 23
- └ Minimum : 16
- └ Mean : 18.95
- └ Standard Deviation : 1.66938375

+ ..... +

## MODUL VII

### KLASIFIKASI : NAÏVE BAYES

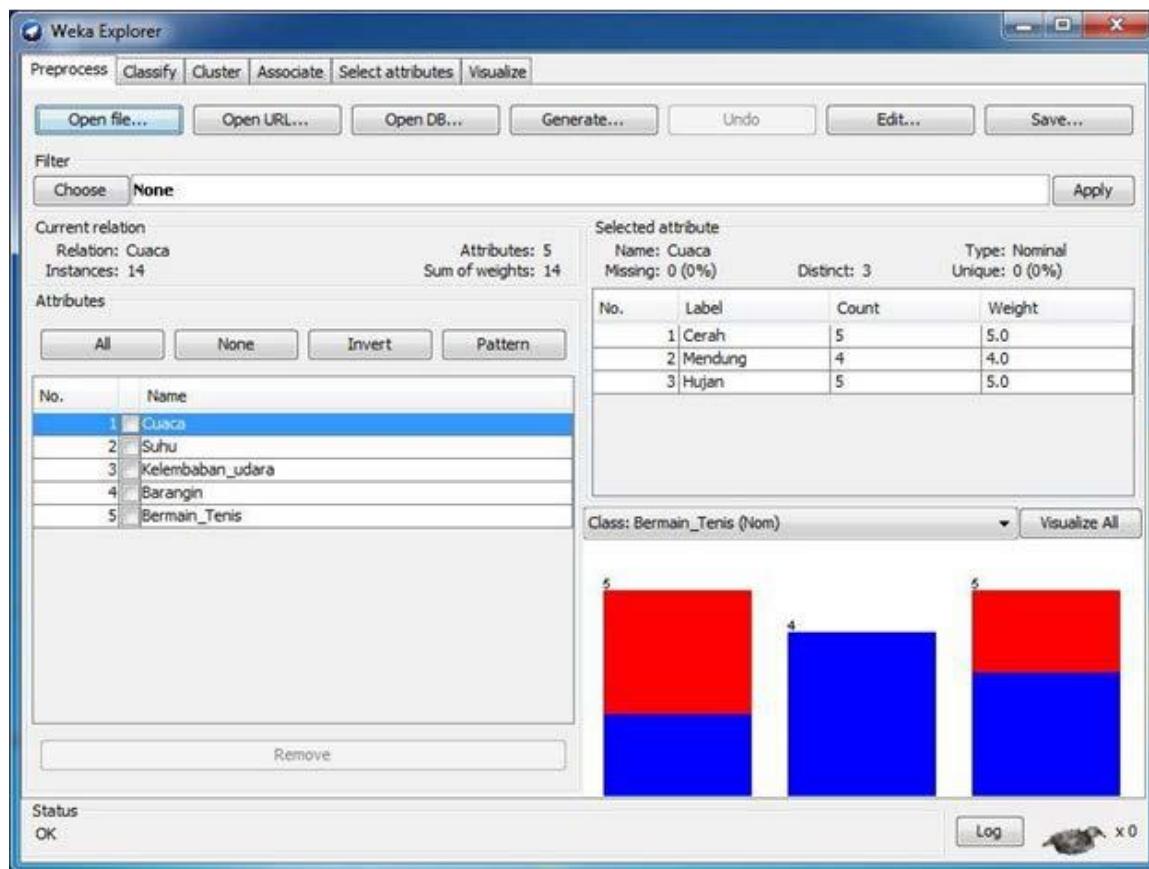
#### **Langkah-langkah Praktikum**

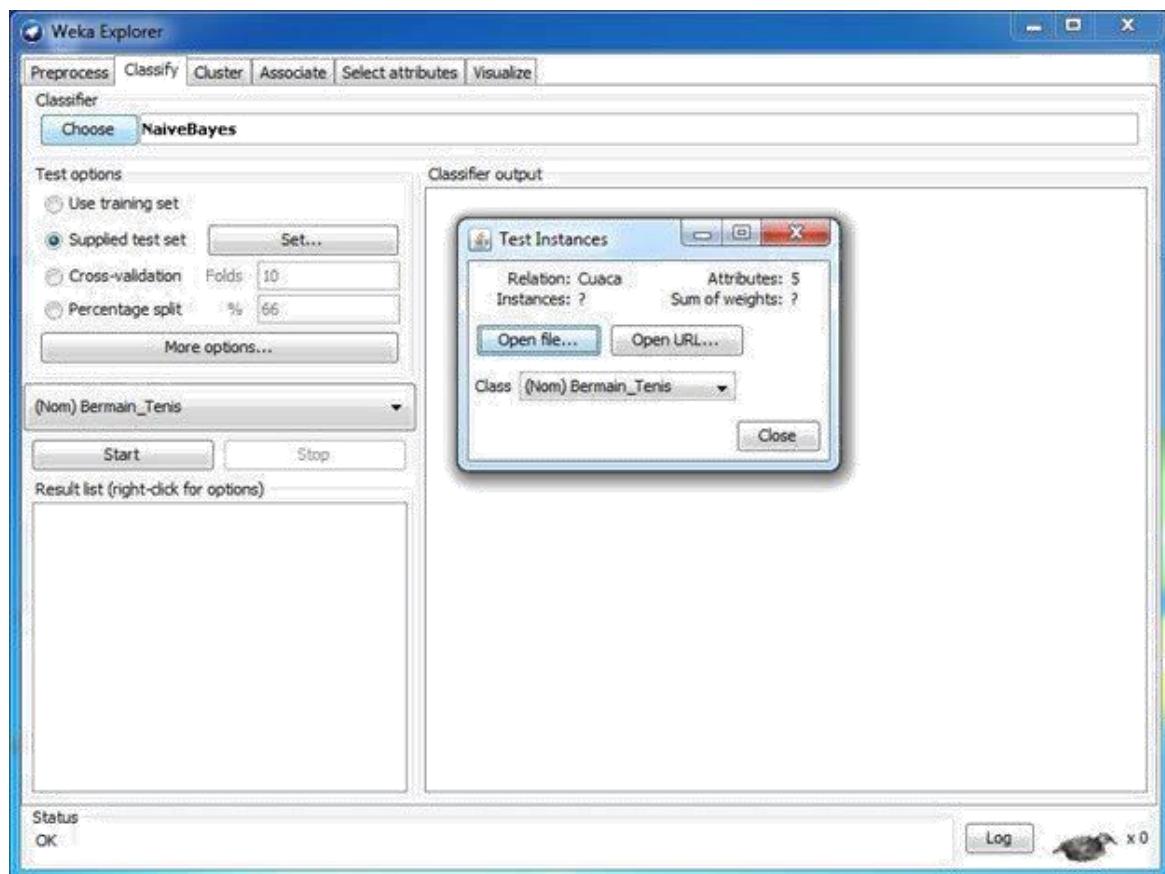
##### **Implementasi Naïve Bayes dengan Weka**

1. Membuat table Data Testing Cuaca yang ada dimodul ke bentuk ARFF dan simpan dengan nama **CuacaTesting.arff**.

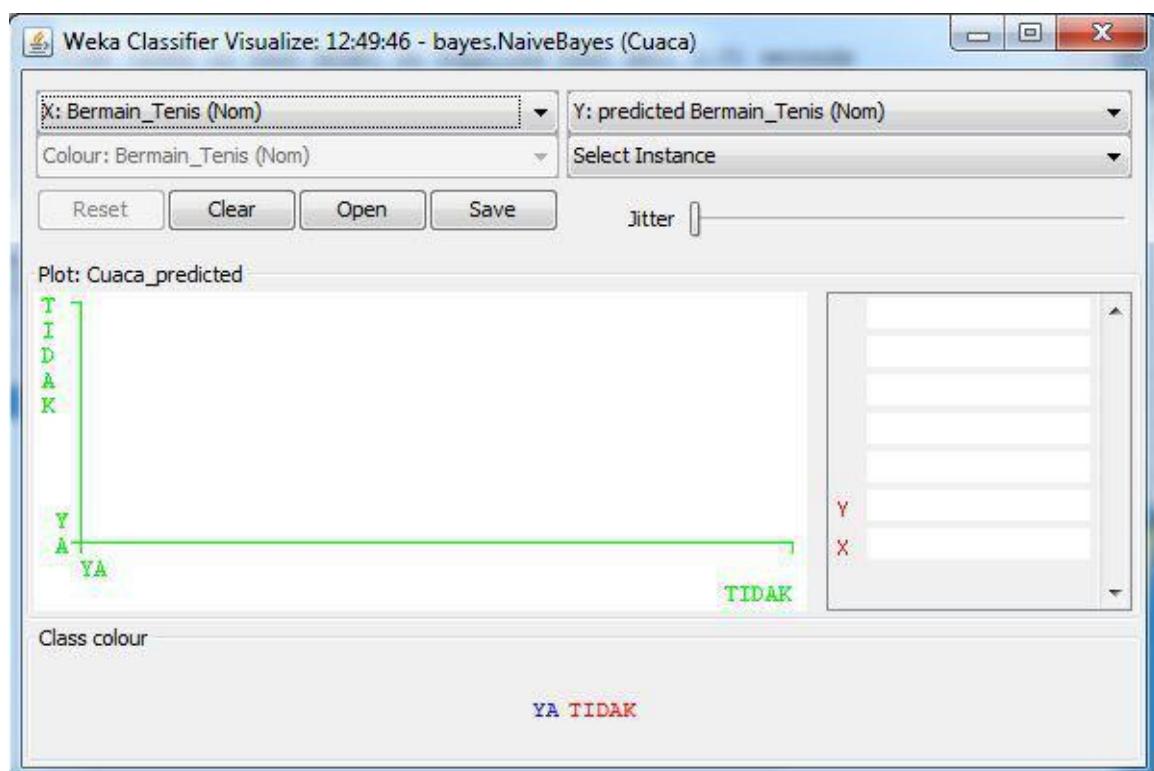
```
1  @relation Cuaca
2
3  @attribute Cuaca {Cerah, Mendung, Hujan}
4  @attribute Suhu real
5  @attribute Kelembaban_udara real
6  @attribute Barangin {YA, TIDAK}
7  @attribute Bermain_Tenis {YA, TIDAK}
8
9  @data
10 Cerah,75,65,TIDAK,?
11 Cerah,80,68,YA,?
12 Cerah,83,87,YA,?
13 Mendung,70,96,TIDAK,?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

2. Membuka file **Cuaca.arff** pada tab **Preprocess**, lalu pada tab **Classify** pilih mode **Naïve Bayes**, klik **Supplied test set**, klik **set**, klik **Open file...** lalu pilih file **CuacaTesting.arff**, **Close**, lalu klik **Start**.





3. Klik kanan pada **Result list**, dan pilih opsi **Classifier Visualize**, biarkan apa yang ditampilkan dan klik **save**.



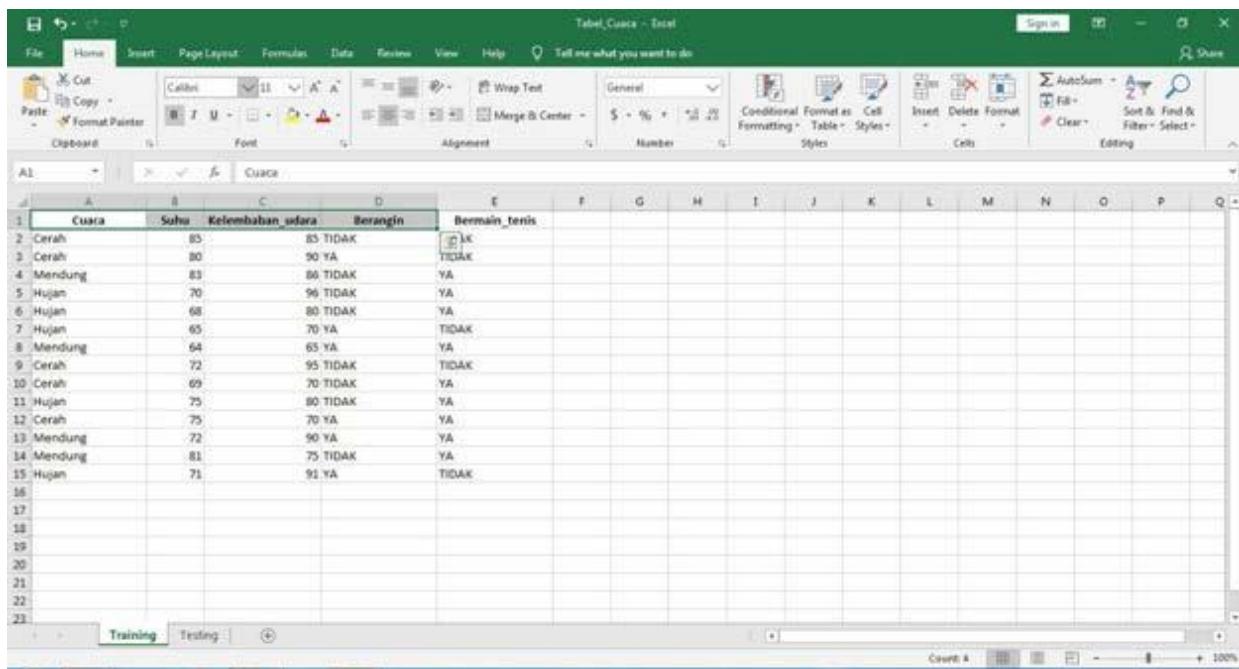
4. Kembali ke GUI utama **WEKA** dan pilih opsi **Tool => ArffViewer => File => Open =>** pilih file hasil dari **Classifier Visualize (HasilPrediksi.arff)** untuk menampilkan hasil prediksi yang sudah dibuat.

The screenshot shows the ARFF-Viewer application window with the title "ARFF-Viewer - F:\Tugas\Praktikum Data Warehousing dan Data Mining\Pertemuan Ke-05\HasilPrediksi.arff". The menu bar includes "File", "Edit", and "View". The main area displays a table titled "Relation: Cuaca\_predicted". The table has 7 columns with the following headers: No., 1: Cuaca Nominal, 2: Suhu Numeric, 3: Kelembaban\_udara Numeric, 4: Barangin Nominal, 5: prediction margin Numeric, 6: predicted Bermain\_Tenis Nominal, and 7: Bermain\_Tenis Nominal. The table contains 7 rows of data. Row 6, which corresponds to "Hujan" with values 65.0, 75.0, 75.0, 0.253733, and YA, is highlighted with a blue background. The other rows show various weather conditions like Cerah, Mendung, and Hujan with their respective predictions.

No.	1: Cuaca Nominal	2: Suhu Numeric	3: Kelembaban_udara Numeric	4: Barangin Nominal	5: prediction margin Numeric	6: predicted Bermain_Tenis Nominal	7: Bermain_Tenis 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	Mendung	70.0	96.0	TIDAK	0.628523	YA	
5	Mendung	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	

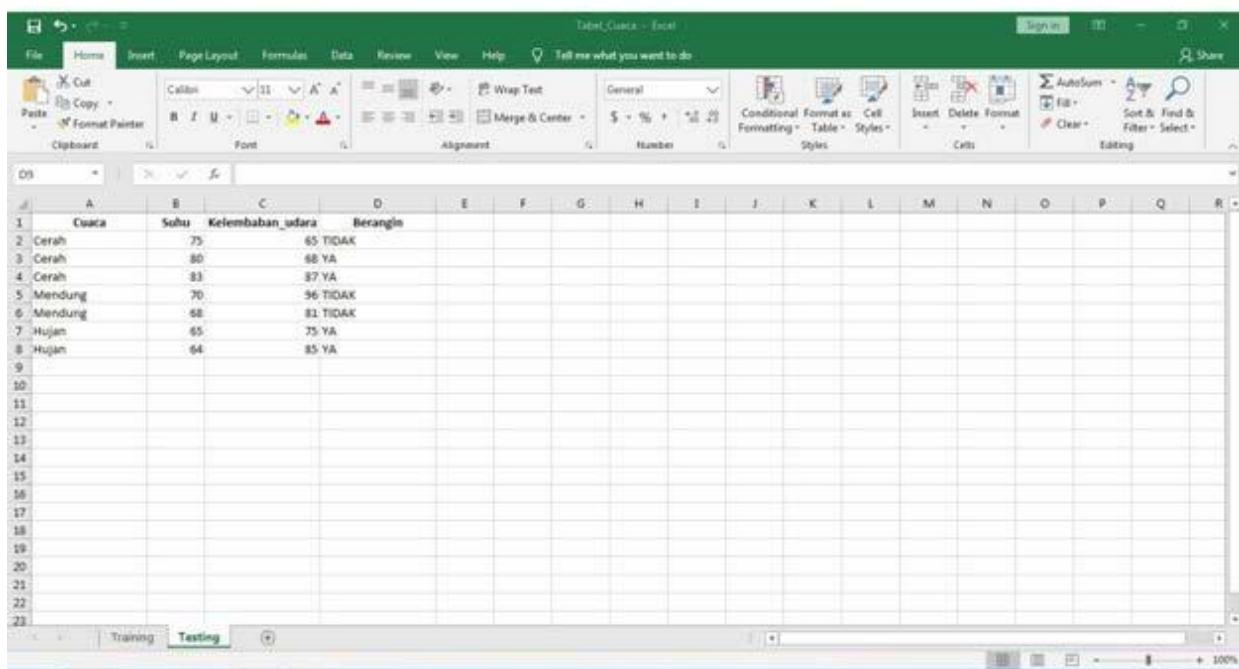
## Implementasi Naïve Bayes dengan RapidMiner

1. Mempersiapkan data dengan Microsoft Excel dengan membuat table **Training** dan **Testing** dengan tema **Cuaca**.



The screenshot shows the Microsoft Excel interface with the title bar "Tabel\_Cuaca - Excel". The ribbon menu is visible with tabs like File, Home, Insert, Page Layout, Formulas, Data, Review, View, and Help. The "Home" tab is selected. The main area displays a table titled "Cuaca" with columns: A (Index), B (Cuaca), C (Suhu), D (Kelembaban\_udara), and E (Berangin). Column E contains categorical values: TIDAK, YA, TIDAK. The table has 17 rows, starting from row 2. The status bar at the bottom shows "Cuaca!4".

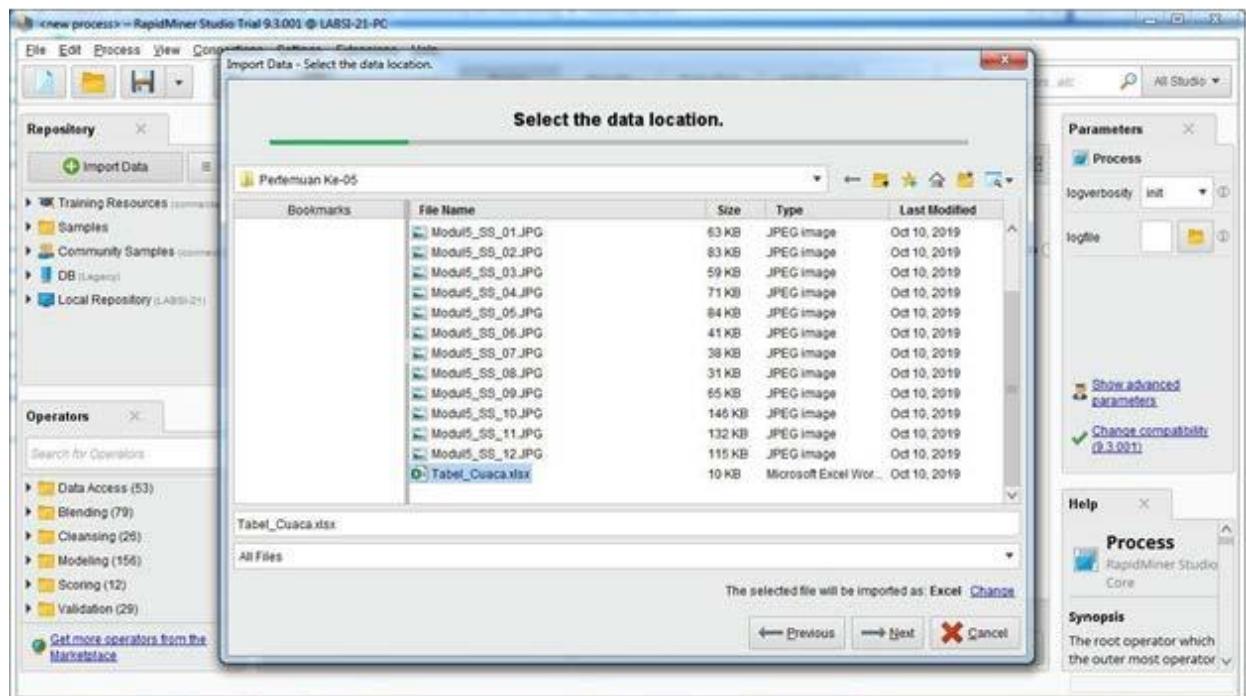
A	B	C	D	E
1	Cuaca	Suhu	Kelembaban_udara	Berangin
2	Cerah	85	85	TIDAK
3	Cerah	80	90	YA
4	Mendung	83	86	TIDAK
5	Hujan	70	96	TIDAK
6	Hujan	68	80	YA
7	Hujan	65	70	YA
8	Mendung	64	65	YA
9	Cerah	72	95	TIDAK
10	Cerah	69	70	TIDAK
11	Hujan	75	80	TIDAK
12	Cerah	75	70	YA
13	Mendung	72	90	YA
14	Mendung	81	75	TIDAK
15	Hujan	71	91	YA
16				TIDAK
17				
18				
19				
20				
21				
22				
23				



The screenshot shows the Microsoft Excel interface with the title bar "Tabel\_Cuaca - Excel". The ribbon menu is visible with tabs like File, Home, Insert, Page Layout, Formulas, Data, Review, View, and Help. The "Home" tab is selected. The main area displays a table titled "Cuaca" with columns: A (Index), B (Cuaca), C (Suhu), D (Kelembaban\_udara), and E (Berangin). Column E contains categorical values: TIDAK, YA, TIDAK. The table has 17 rows, starting from row 2. The status bar at the bottom shows "Cuaca!4".

A	B	C	D	E
1	Cuaca	Suhu	Kelembaban_udara	Berangin
2	Cerah	75	85	TIDAK
3	Cerah	80	68	YA
4	Cerah	83	87	YA
5	Mendung	70	96	TIDAK
6	Mendung	68	81	TIDAK
7	Hujan	65	75	YA
8	Hujan	64	85	YA
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				

2. Menjalankan **Rapid Miner** dan Meng-import file excel yang sudah disimpan tadi (nama file **Tabel\_Cuaca.xls**).



3. Setelah itu pilih Sheet ke Training, klik Next lalu ubah tipe table **Bermain\_Tenis** data dari **polynomial** dengan **binomial** dan mengganti **Role** dengan label, setelah itu simpan data ke Local Repository.

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Training ▾ Cell range: A:E Select All  Define header row: 1

	A	B	C	D	E
1	Cuaca	Suhu	Kelembaban_udara	Berangin	Bermain_tenis
2	Cerah	85.000	85.000	TIDAK	TIDAK
3	Cerah	80.000	90.000	YA	TIDAK
4	Mendung	83.000	86.000	TIDAK	YA
5	Hujan	70.000	96.000	TIDAK	YA
6	Hujan	68.000	80.000	TIDAK	YA
7	Hujan	65.000	70.000	YA	TIDAK
8	Mendung	64.000	65.000	YA	YA
9	Cerah	72.000	95.000	TIDAK	TIDAK
10	Cerah	69.000	70.000	TIDAK	YA
11	Hujan	75.000	80.000	TIDAK	YA
12	Cerah	75.000	70.000	YA	YA
13	Mendung	72.000	90.000	YA	YA
14	Mendung	81.000	75.000	TIDAK	YA
15	Rata-rata	74.000	84.000	YA	TIDAK

← Previous      → Next      X Cancel

Import Data - Format your columns.

### Format your columns.

Replace errors with missing values (1)

	Cuaca	Suhu	Kelembaban_udara	Berangin	Bermain_tenis
1	Cerah	85	88	TIDAK	TIDAK
2	Cerah	80	90	YA	TIDAK
3	Mendung	83	88	TIDAK	YA
4	Hujan	70	36	YILO#	YA
5	Hujan		80	TIDAK	YA
6	Hujan	65	70		TIDAK YA
8	Cerah	72	95	TIOM	TIDAK
9	Cerah	89	70	TOOK	YA
10	Hujan	28	80	TiOa	YA
11	Cerah	7?	70	F4	YA
12	Mendung	72	90	YA	YA
...					

 no problems.

## Where to store the data?

---

\* Location Repository

Location /Local Repository/DataCuaca\_Training

 Previous |  Finish |  Cancel

4. Lakukan langkah yang sama pada **Sheet Testing**, dengan perbedaan tidak ada tabel yang diubah **Role** menjadi **label**.

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Testing ▾ Cell range: A:D Select All  Define header row: 1

	A	B	C	D
1	Cuaca	Suhu	Kelembaban_udara	Berangin
2	Cerah	75.000	65.000	TIDAK
3	Cerah	80.000	68.000	YA
4	Cerah	83.000	87.000	YA
5	Mendung	70.000	96.000	TIDAK
6	Mendung	68.000	81.000	TIDAK
7	Hujan	65.000	75.000	YA
8	Hujan	64.000	85.000	YA

← Previous    → Next     Cancel

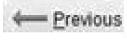
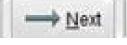
Import Data - Format your columns.

### Format your columns.

Replace errors with missing values (1)

Cuaca	Suhu	Kelembaban_udara	Berangin	dinominal
1 Cerah	75	65	TIDAK	
2 Cerah	80	68	YA	
S . Car th	83	8?	YA	
4 Mendung		96	Ti0w	
5 Mendung	68	81	TID*ñ	

 no problems.

## Where to store the data?

---

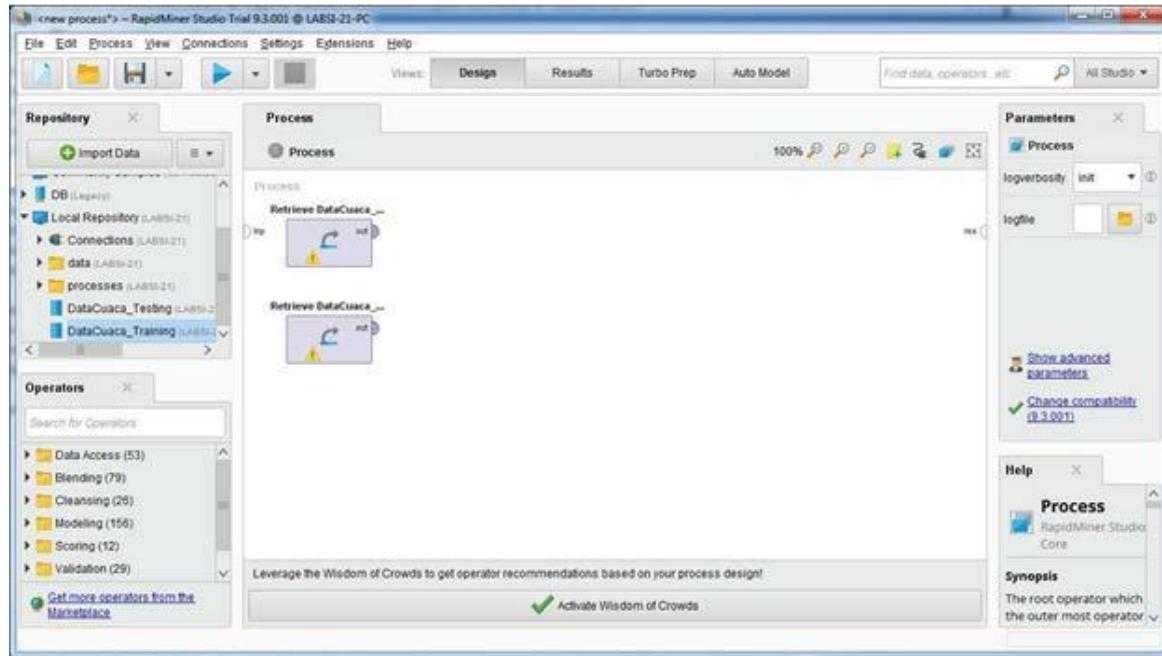
\* Location Repository\*

Location: Local Repository/DataCuaca\_Testing

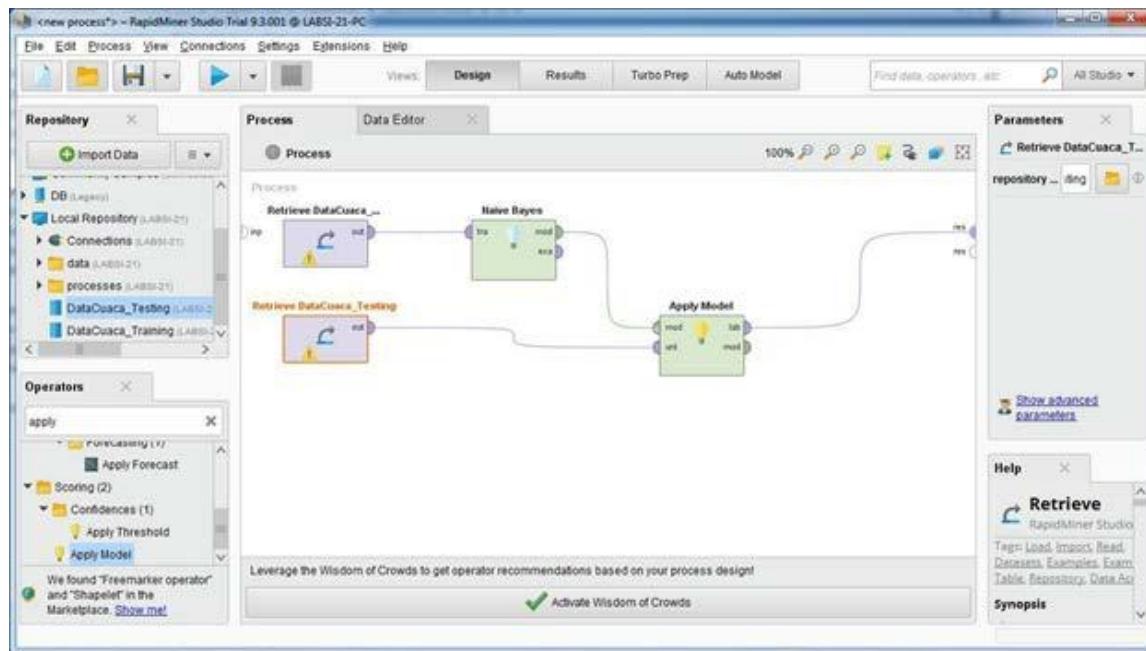
 Previous ||  Finish ||  Cancel

Kelimbahan

5. Kembali ke halaman utama **Rapid Miner** dan membuat desain **Naïve Bayes**, dengan mengdrag **DataCuaca\_Training** dan **DataCuaca\_Testing** ke dalam jendela **Process View**.



6. Cari operator **Naïve Bayes** dan **Apply Model** (bisa dicari di jendela **Operator**) dan drag ke jendela **Process**, lalu sambungkan masing-masing node, lalu klik simbol **Run**.



The screenshot shows the RapidMiner Studio interface with the 'Data' tab selected. On the left, there's a sidebar with icons for Data, Statistics, Visualizations, and Annotations. The main area displays a table titled 'ExampleSet (7 examples; 3 special attributes, 4 regular attributes)'. The table has columns: Row No., prediction(B...), confidence(T..., confidence(Y..., Cuaca, Suhu, Kelembaban\_..., Berangin. The data rows are:

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

At the bottom, it says 'ExampleSet (7 examples; 3 special attributes, 4 regular attributes)'. The right side shows the 'Repository' pane with 'Local Repository' expanded, showing 'Connections', 'data', and 'processes'.

7. Klik Statistics untuk melihat data lebih detail.

The screenshot shows the RapidMiner Studio interface with the 'Statistics' tab selected. The left sidebar shows the 'Statistics' icon is highlighted. The main area displays a table of attribute statistics. The columns are: Name, Type, Missing, Statistics, Filter (7 / 7 attributes), and a search bar. The attributes listed are:

Name	Type	Missing	Statistics	Filter (7 / 7 attributes)
<b>prediction(Berman_tenis)</b>	Binomial	0	Min: TIDAK (2)   Max: YA (5)   Average: YA (5)	
<b>confidence(TIDAK)</b>	Real	0	Min: 0.007   Max: 0.856   Average: 0.353	
<b>confidence(YA)</b>	Real	0	Min: 0.144   Max: 0.993   Average: 0.647	
<b>Cuaca</b>	Polynomial	0	Min: Mendung (2)   Max: Cerah (3)   Average: Cerah (2)	
<b>Suhu</b>	Integer	0	Min: 64   Max: 83   Average: 72.143	
<b>Kelembaban_udara</b>	Integer	0	Min: 65   Max: 96   Average: 79.571	
<b>Berangin</b>	Binomial	0	Min: TIDAK (3)   Max: YA (4)   Average: YA (4)	

At the bottom, it says 'Showing attributes 1 - 7' and 'Examples: 7 Special Attributes: 3 Regular Attributes: 4'. The right side shows the 'Repository' pane with 'Local Repository' expanded, showing 'Connections', 'data', and 'processes'.

## Tugas

1. Menulis ulang table baru yang ada di Modul Tugas dengan format Excel dan Arff dari data mahasiswa (Tugas Modul 6), sebagai data Testing untuk tabel Mahasiswa.

The screenshot shows two tabs in Microsoft Excel. The left tab, 'Tabel\_Tugas.arff', contains the following ARFF file content:

```
2
3 @attribute Jurusan_SMA {IPA, IPS, LAIN}
4 @attribute Gender {PRIA, WANITA}
5 @attribute Asal_Sekolah {SURAKARTA, LUAR}
6 @attribute Rerata_SKS real
7 @attribute Asisten {YA, TIDAK}
8 @attribute Lama_Studi {TEPAT, TERLAMBAT}
9
10 #@data
11 LAIN,WANITA,SURAKARTA,18,TIDAK,?
12 IPA,PRIA,SURAKARTA,19,YA,?
13 LAIN,PRIA,SURAKARTA,19,TIDAK,?
14 IPS,PRIA,LUAR,17,TIDAK,?
15 LAIN,WANITA,SURAKARTA,17,TIDAK,?
16 IPA,WANITA,LUAR,18,YA,?
17 IPA,PRIA,SURAKARTA,18,TIDAK,?
18 IPA,PRIA,SURAKARTA,19,TIDAK,?
19 IPS,PRIA,LUAR,18,TIDAK,?
20 LAIN,WANITA,SURAKARTA,18,TIDAK,?
```

The right tab, 'Mahasiswa.xlsx', contains a table named 'Jurusan\_SMA' with the following data:

A	B	C	D	E	
#	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten
1	LAIN	WANITA	SURAKARTA	18	TIDAK
2	IPA	PRIA	SURAKARTA	19	YA
3	LAIN	PRIA	SURAKARTA	19	TIDAK
4	IPS	PRIA	LUAR	17	TIDAK
5	LAIN	WANITA	SURAKARTA	17	TIDAK
6	IPA	WANITA	LUAR	18	YA
7	IPA	PRIA	SURAKARTA	18	TIDAK
8	IPA	PRIA	SURAKARTA	19	TIDAK
9	IPS	PRIA	LUAR	18	TIDAK
10	LAIN	WANITA	SURAKARTA	18	TIDAK
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

2. Melakukan Prediksi dengan memanfaatkan Software WEKA.

The screenshot shows the Weka Explorer interface. The 'Selected attribute' panel displays the following information for 'Jurusan\_SMA':

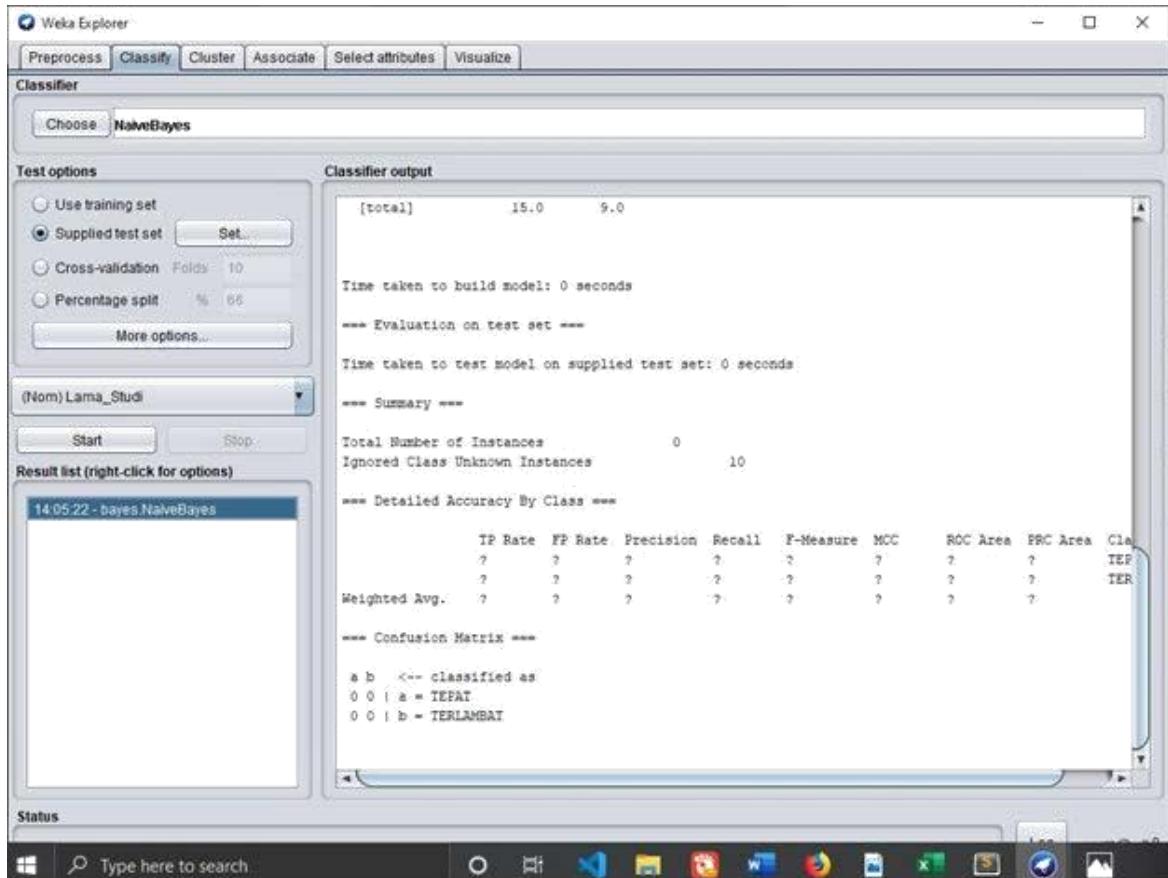
- Name: Jurusan\_SMA
- Type: Nominal
- Distinct: 3
- Unique: 0 (0%)

The 'Attributes' panel lists the following attributes:

- All
- None
- Invert
- Pattern

No.	Name
1	Jurusan_SMA
2	Gender
3	Asal_Sekolah
4	Rerata_SKS
5	Asisten
6	Lama_Studi

The 'Status' bar at the bottom shows the Windows taskbar with icons for File Explorer, Task View, Start, and other applications.



ARFF-Viewer - F:\Tugas\Praktikum Data Warehousing dan Data Mining\Pertemuan Ke-05\HasilPrediksi.arff

Relation: Cuaca_predicted							
No.	1: Cuaca Nominal	2: Suhu Numeric	3: Kelembaban_udara Numeric	4: Barangin Nominal	5: prediction margin Numeric	6: predicted Bermain_Tenis Nominal	7: Bermain_Tenis Nominal
1	Cerah	75.0	65.0	TIDAK	0.762765	YA	
2	Cerah	80.0	68.0	YA	0.087978	YA	
3	Cerah	83.0	87.0	YA	-0.676866	TIDAK	
4	Mendung	70.0	96.0	TIDAK	0.628523	YA	
5	Mendung	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	

3. Melakukan Prediksi dengan menggunakan Software RapidMiner Studio.

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Training ▾ Cell range: A:F Select All  Define header row: 1

	A	B	C	D	E	F
1	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten	Lama_Studi
2	IPS	WANITA	SURAKARTA	18.000	TIDAK	TERLAMBAT
3	IPA	PRIA	SURAKARTA	19.000	YA	TEPAT
4	LAIN	PRIA	SURAKARTA	19.000	TIDAK	TERLAMBAT
5	IPA	PRIA	LUAR	17.000	TIDAK	TERLAMBAT
6	IPA	WANITA	SURAKARTA	17.000	TIDAK	TEPAT
7	IPA	WANITA	LUAR	18.000	YA	TEPAT
8	IPA	PRIA	SURAKARTA	18.000	TIDAK	TERLAMBAT
9	IPA	PRIA	SURAKARTA	19.000	TIDAK	TEPAT
10	IPS	PRIA	LUAR	18.000	TIDAK	TERLAMBAT
11	LAIN	WANITA	SURAKARTA	18.000	TIDAK	TEPAT
12	IPA	WANITA	SURAKARTA	19.000	TIDAK	TEPAT
13	IPS	PRIA	SURAKARTA	20.000	TIDAK	TEPAT
14	IPS	PRIA	SURAKARTA	19.000	TIDAK	TEPAT
15	IPS	PRIA	SURAKARTA	18.000	TIDAK	TEPAT

← Previous      → Next      

## Formed your columns.

Qepiace errors with missing values

Jurusan SMA O • Gender o • Asal Sakolah O • Rata SKS 0 • Akisten 0 - Lama\_Studi 0 •

1	IPS	WANITA	SURAKARTA	18	TIDAK	TERLAMBAT	A
2	IPA	PRIA	SURAKARTA	19	YA	TEPAT	
3	LAIN	PRIA	SURAKARTA	19	TIDAK	TERLAMBAT	
4	IPA	PRIA	LUAR	17	TIDAK	TERLAMBAT	
5	IPA	WANITA	SURAKARTA	17	TIDAK	TEPAT	
6	IPA	WANITA	LUAR	18	YA	YEPAT	
7	IPA	PRIA	SURAKARTA	18	TIDAK	TERLAMBAT	
8	IPA	PRIA	SURAKARTA	19	TIDAK	TEPAT	
9	IPS	PRIA	LUAR	18	TIDAK	TERLAMBAT	
10	LAIN	WANITA	SURAKARTA	18	TIDAK	TEPAT	
11	IPA	WANITA	SURAKARTA	19	TIDAK	TEPAT	
12	IPS	PRIA	SURAKARTA	20	TIDAK	TEPAT	

z-car  
@ no problems.

Previous ... Next Cancel

Import Data - Where to store the data? X

### Where to store the data?

- (@ Le0M RepOsBQ
  - B Connected on s

Name | Tabel\_Tugas\_Training\_Skipipap

Location local E:\pos4or\Tabel\_7ugas\_Training\_Skipipap

 Previous  Finish  Cancel

## Select the cells to import.

ghgg1. Testing •	Cell range: A:E	Select All	Done	Cancel
A	B	C	D	E
2 LAIN	WANITA	SURAKARTA	18.g%"	TIDAK
3 IPA	PRIA	SURAKARTA	19.000	YA
4 LAIN	PRIA	SURAKARTA	-1g.é"00	TIDAK
5 IPS	PRIA	LUAR	GOX	TIDAK
6 LAIN	WANITA	SURAKARTA	17.000	TIDAK
7 IPA	WANITA	LUAR	18.000	YA
8 IPA	PRIA	SURAKARTA	18.000	TIDAK
9 "-IPA	PRIA	SURAKARTA	19.000	TIDAK
10 "-IPS"	PRIA	LUAR	18.000	TIDAK
11 -UIN	WANITA	SURATMftTA"	18.000	TIDAK

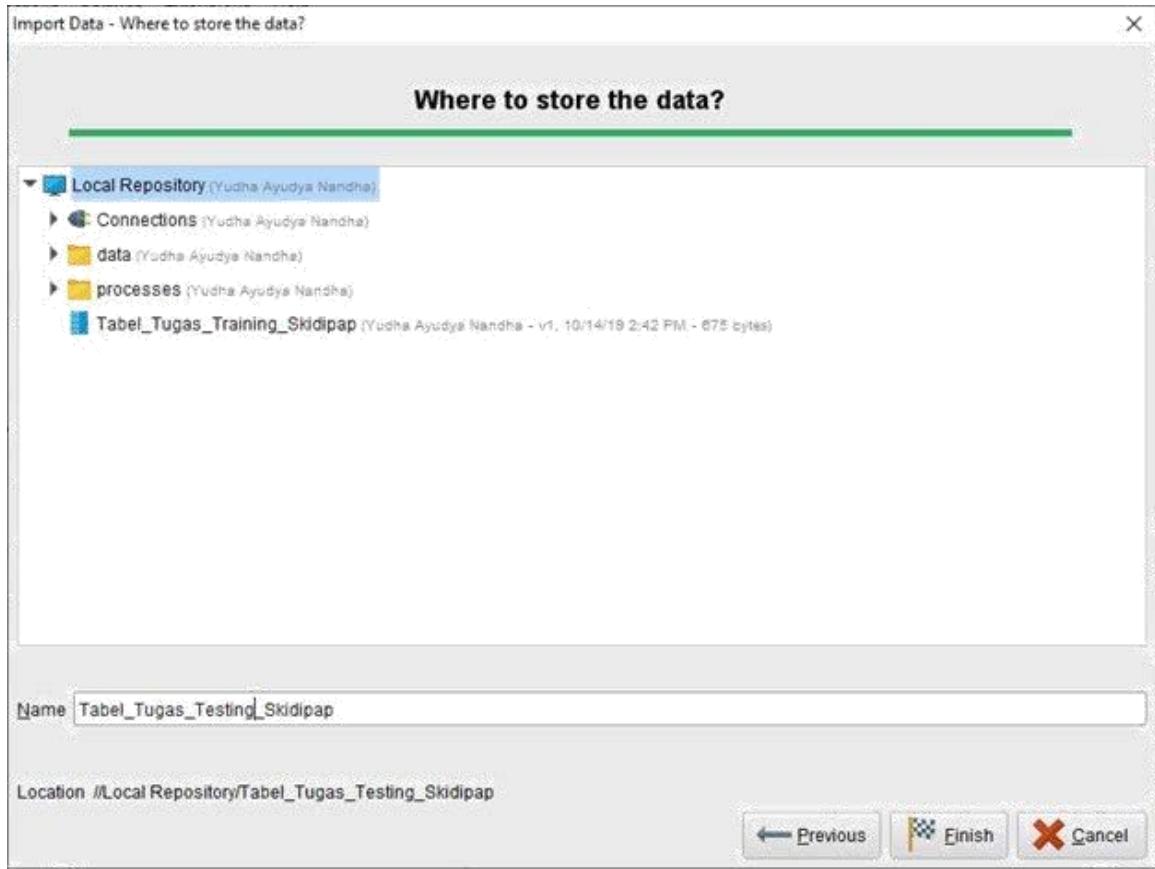
0--•• FTBñOuS -0] Wed QdNCM

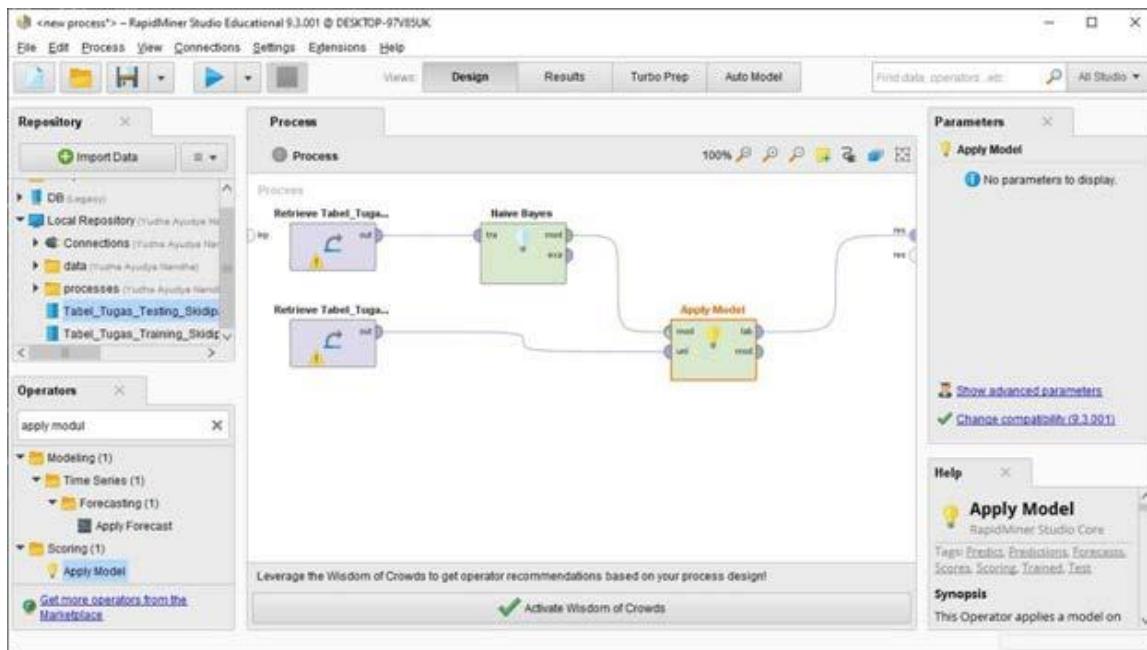
## Format your columns.

Juru6an_SMA	Gandar	As.al_Sekolah	Rarata SKS	Asia£an
1 MN	WANITA	SURAKART4	IB	TIO4K
2 IPA	PRIA	9UR Tn	19	YA
3 UAT	PRIA	SURAKARTA	19	TIDAK
4 IPS	PRIA	LU W	17	TIDAK
5 LAN	WANITA	9U RTC	47	TIDAK
6 IP.4	WANITA	LUAR	18	YA
7 iPa	PRIA	9URAKARTA	1g	TIDAK
8 IP.4	PRIA	SUR'E7RTA	19	TIDAK
9 iPS	PRIA	LU AP	18	TIDAK
10 MN	WANITA	9UPAJGIRTC	18	TIDAK

F161Df0b18ITI8.

Previous    ...+ jjext    Cancel





**Data**

Row No.	prediction	confidence	confidencetab	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten
1	TERLAMBAT	0.648	0.352	LAIN	WANITA	SURAKARTA	18	TIDAK
2	TEPAT	0.005	0.995	IPA	PRIA	SURAKARTA	19	YA
3	TERLAMBAT	0.650	0.350	LAIN	PRIA	SURAKARTA	19	TIDAK
4	TERLAMBAT	0.660	0.132	IPS	PRIA	LUAR	17	TIDAK
5	TERLAMBAT	0.738	0.262	LAIN	WANITA	SURAKARTA	17	TIDAK
6	TEPAT	0.005	0.995	IPA	WANITA	LUAR	18	YA
7	TERLAMBAT	0.547	0.453	IPA	PRIA	SURAKARTA	18	TIDAK
8	TEPAT	0.321	0.679	IPA	PRIA	SURAKARTA	19	TIDAK
9	TERLAMBAT	0.811	0.189	IPS	PRIA	LUAR	18	TIDAK
10	TERLAMBAT	0.548	0.352	LAIN	WANITA	SURAKARTA	18	TIDAK

**Repository**

- Import Data
- Samples
- DB (empty)
- Local Repository (Yudha Ayudya Handoko)
  - Connections (Yudha Ayudya Handoko)
  - data (Yudha Ayudya Handoko)
  - processes (Yudha Ayudya Handoko)
    - Tabel\_Tugas\_Testing\_Skipip (Yudha Ayudya Handoko)
    - Tabel\_Tugas\_Training\_Skipip (Yudha Ayudya Handoko)

4. Nilai Rerata hasil dari Prediksi RapidMiner untuk kolom Lama\_Studie :
  - Rerata untuk TEPAT : 0.476
  - Rerata untuk TERLAMBAT : 0.524
5. Dari hasil prediksi dari Software Rapidminer Studio, dapat disimpulkan bahwa 3 Mahasiswa akan lulus TEPAT waktu, dan 7 Mahasiswa akan lulus TERLAMBAT.

6. Dewi adalah Mahasiswa WANITA, berasal dari jurusan IPA saat bangku SMA, bersekolah di LUAR Surakarta, mengambil SKS rata-rata sebanyak 18 tiap semester, dan TIDAK pernah menjadi asisten selama kuliah.

1	TEPAT	0.298	0.702	IPA	WANITA	LUAR	18	TIDAK
---	-------	-------	-------	-----	--------	------	----	-------

7. Jono adalah Mahasiswa PRIA, tidak berasal dari jurusan IPA maupun IPS, bersekolah di SURAKARTA, mengambil SKS rata-rata sebanyak 17 tiap semester, dan PERNAH menjadi asisten selama kuliah.

2	TEPAT	0.076	0.924	LAIN	PRIA	SURAKARTA	17	YA
---	-------	-------	-------	------	------	-----------	----	----

+.....+

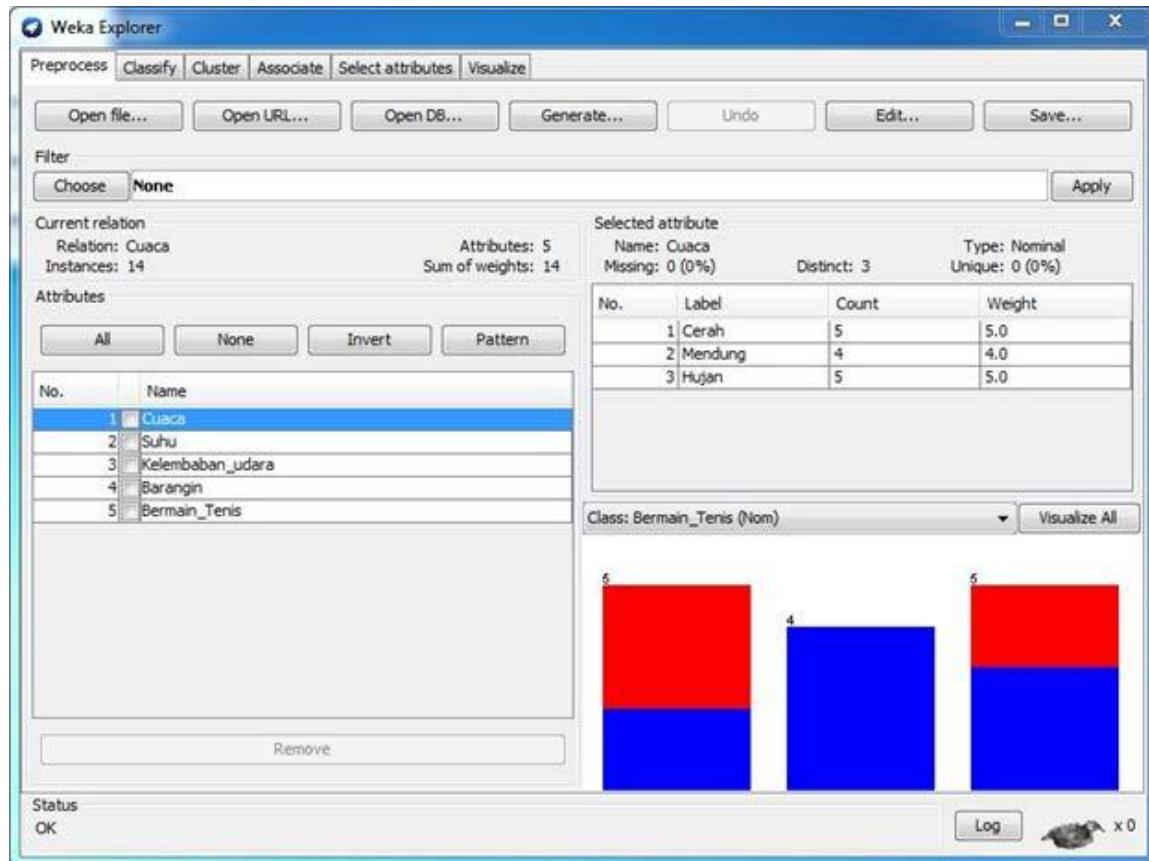
# MODUL IX

## KLASIFIKASI : DECISION TREE

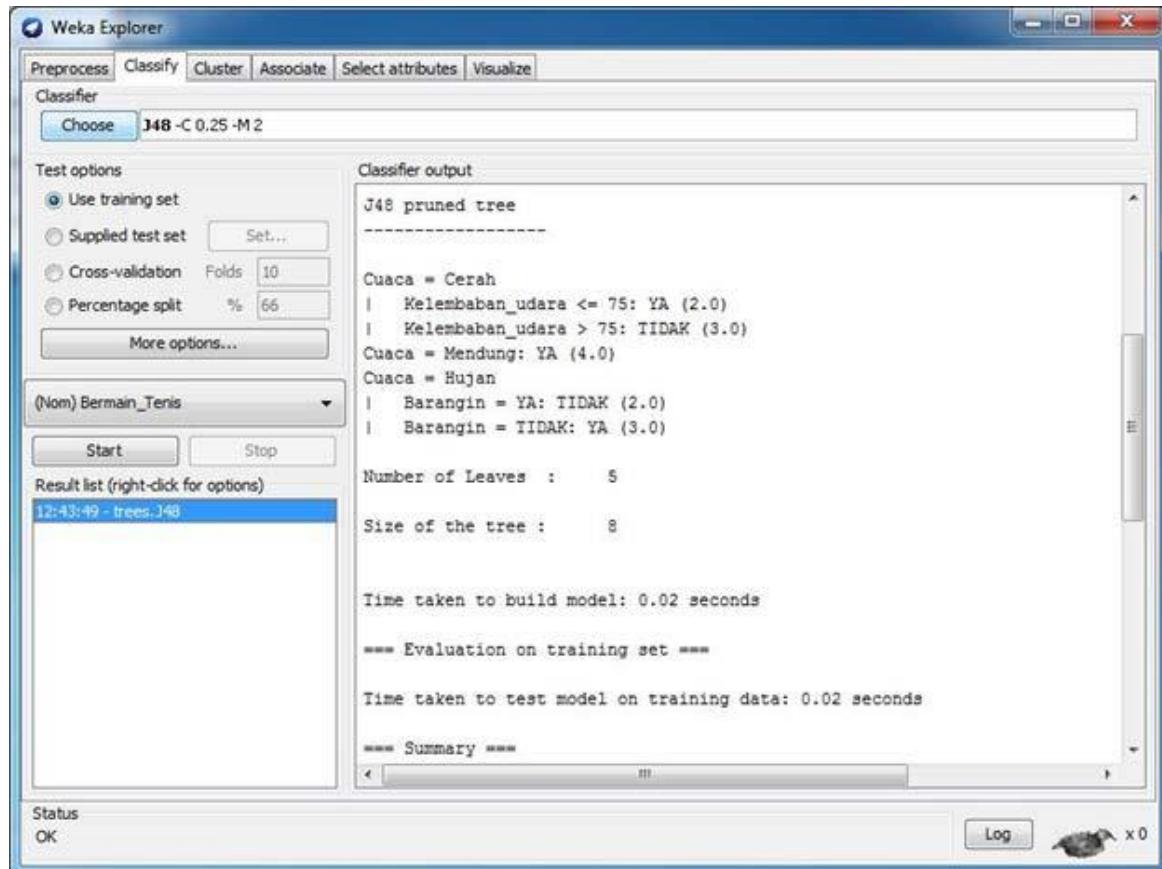
### Langkah-langkah Praktikum

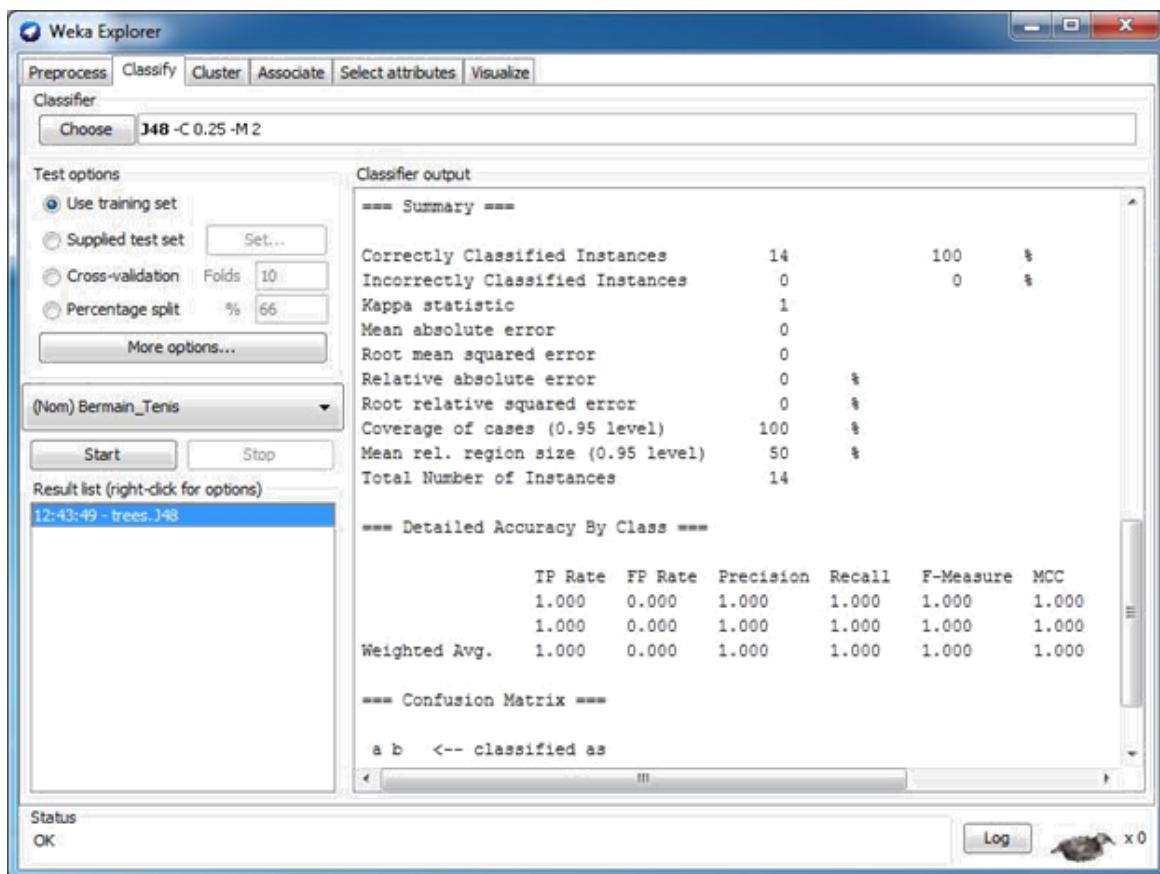
#### Pohon Keputusan Menggunakan WEKA

1. Buka file **Cuaca.arff**, dengan Weka Explore.

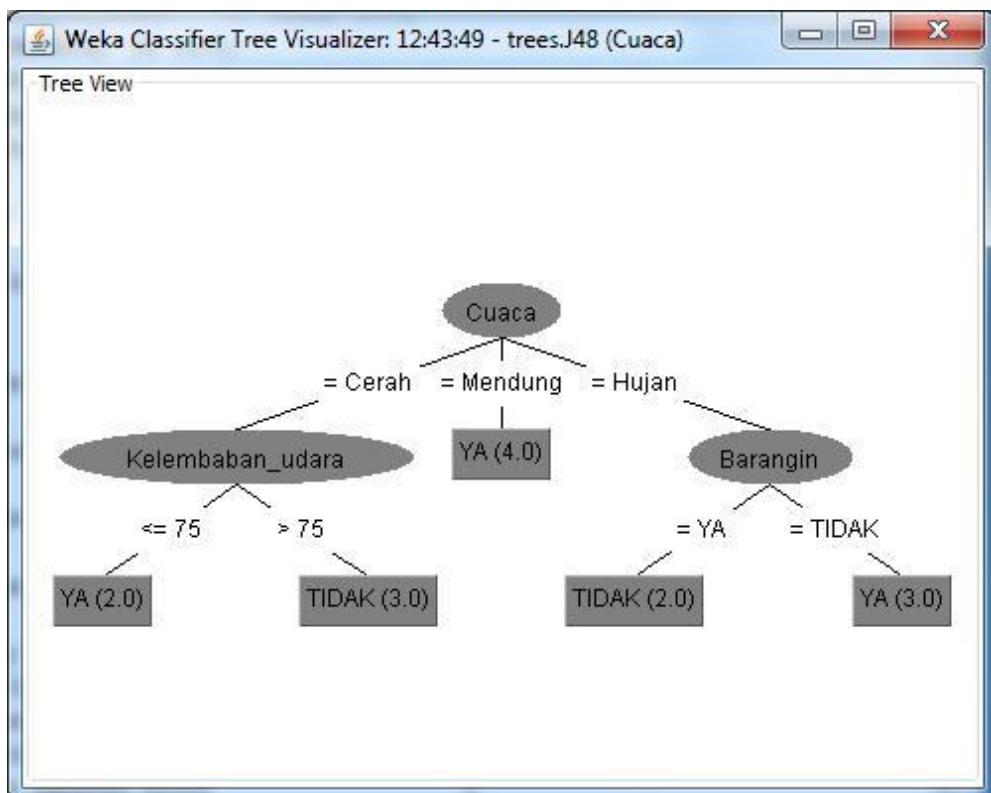


2. Klik tab **Classify** dan tekan tombol **Choose**, pilih **Trees => J48**, pilih **Use training set** pada **Test options** dan pastikan atribut dependen pada **Bermain\_Tenis** dan klik **Start** (symbol segitiga).



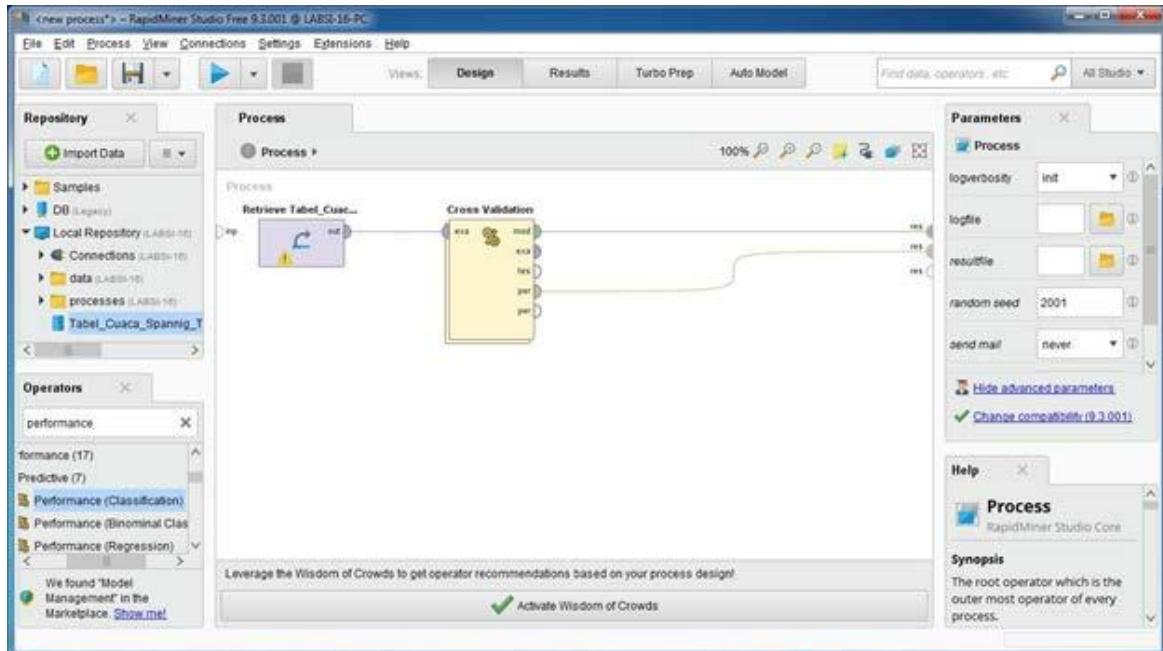


3. Klik kanan pada **Result list** dan pilih **Visual tree**, untuk melihat **Pohon Keputusannya**.

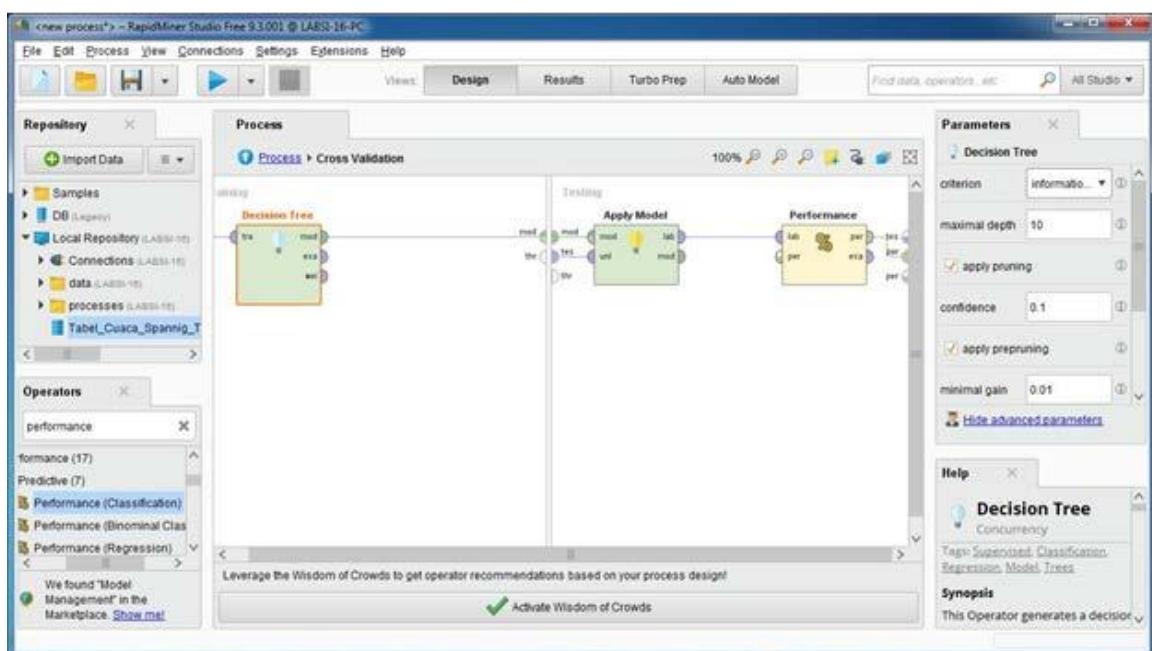


## Pohon Keputusan Menggunakan RapidMiner

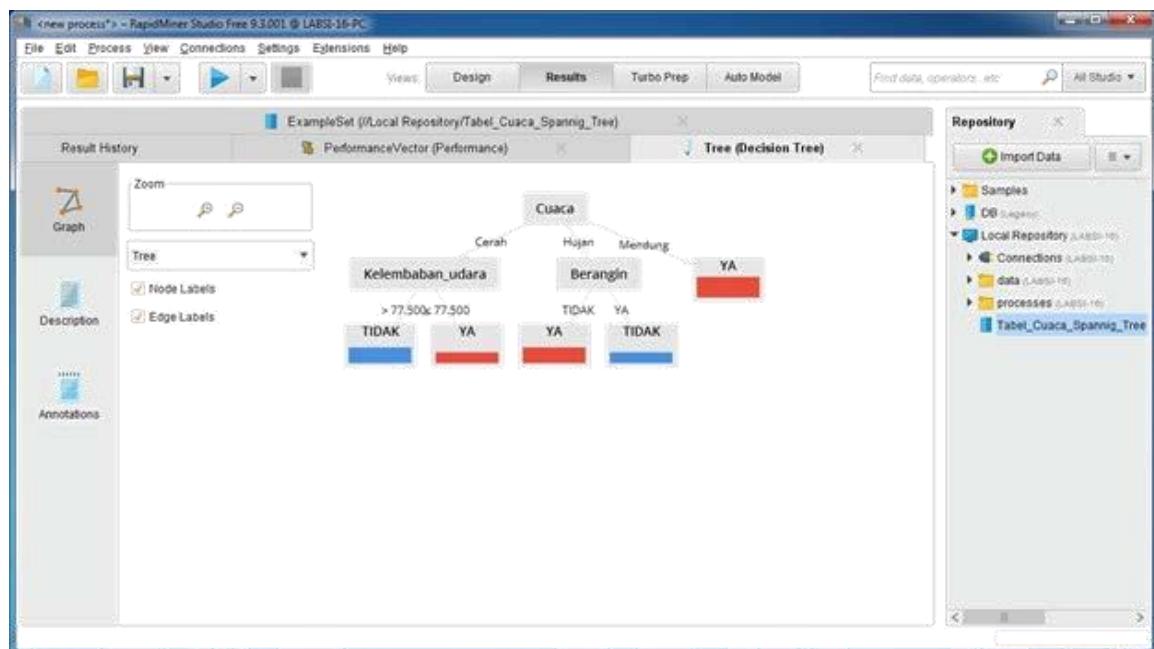
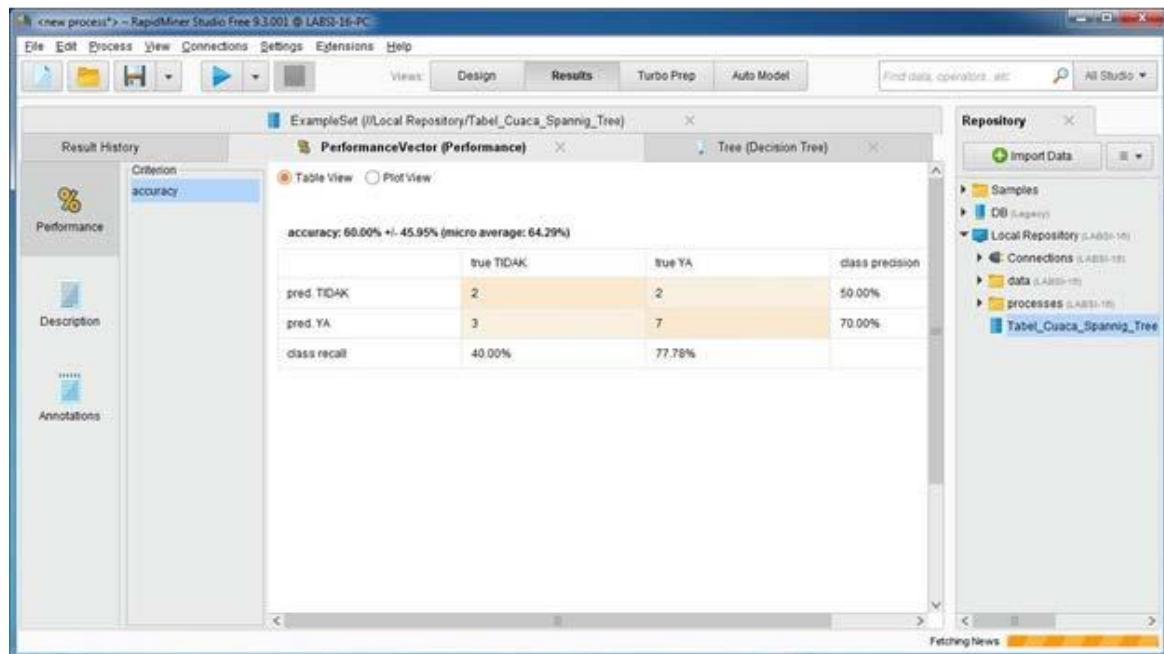
1. Buka **RapidMiner** dan klik **New Process => Blank** pada halaman perspective RapidMiner dan buat data Training, masukkan data training tersebut ke kolom **Process** lalu cari pada kolom **Operators** ketikkan **cross validation**.



2. Klik 2X pada **Cross Validation** dan masukkan **decision**, **apply model**, dan **performance**, lalu arahkan sesuai instruksi buku.



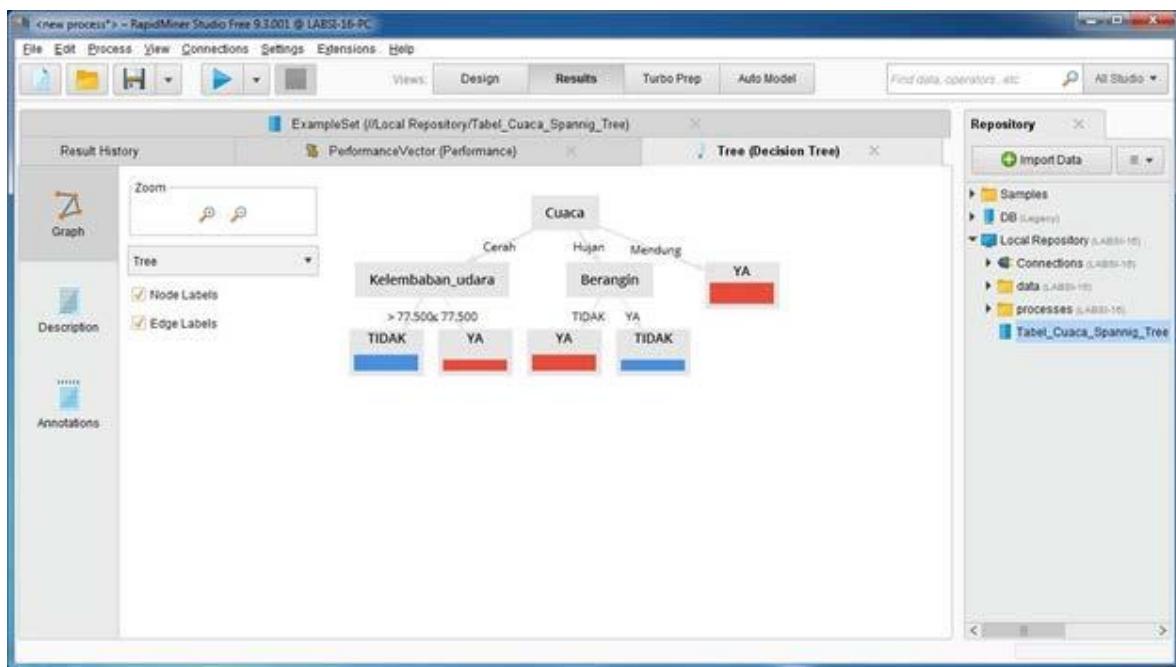
3. Klik Start (Simbol segitiga) lalu lihat hasilnya.



## Tugas

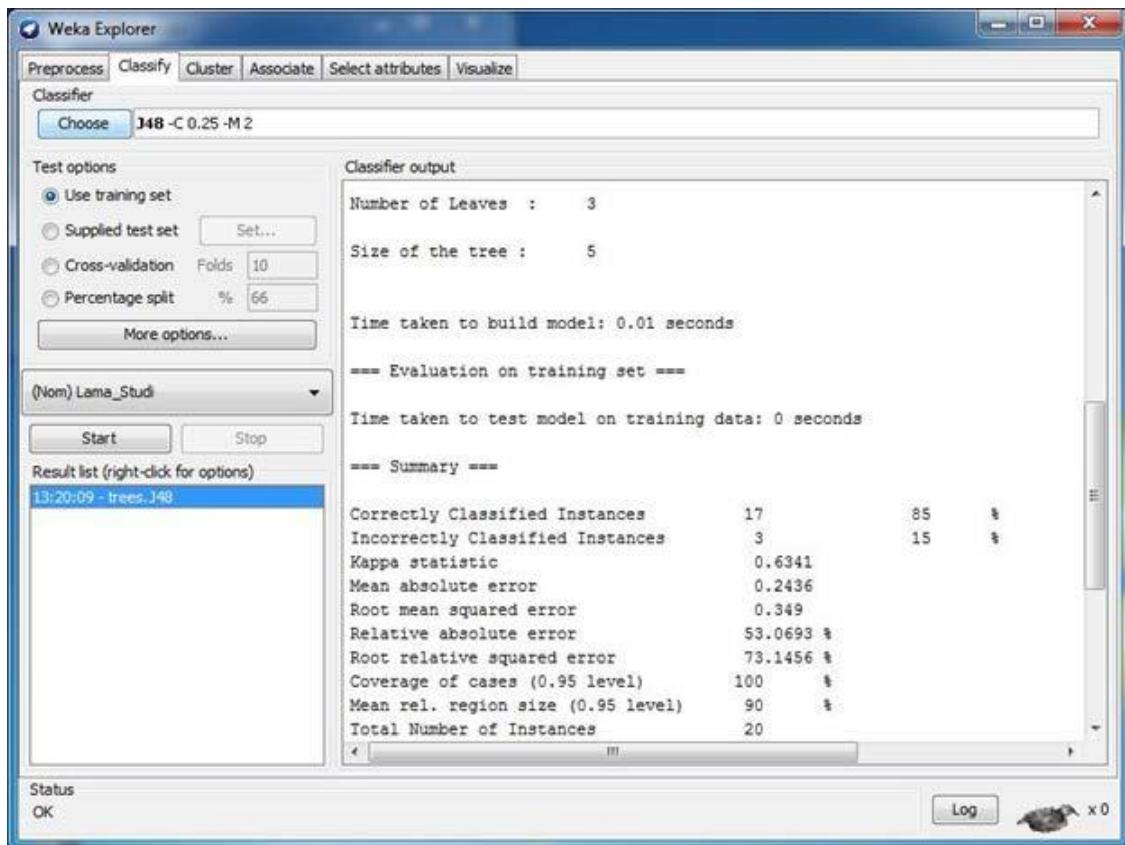
1. Berdasarkan pohon keputusan pada kegiatan 9.4.2 (menggunakan RapidMiner), isikan nilai kelas atribut Bermain\_Tenis pada tabel Testing berikut :

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



2. Gunakan file ARFF yang dikerjakan pada Tugas nomor 1 dalam Modul 7 sebagai data training.

a) Buatlah dan cetaklah pohon keputusan berdasarkan data tersebut!

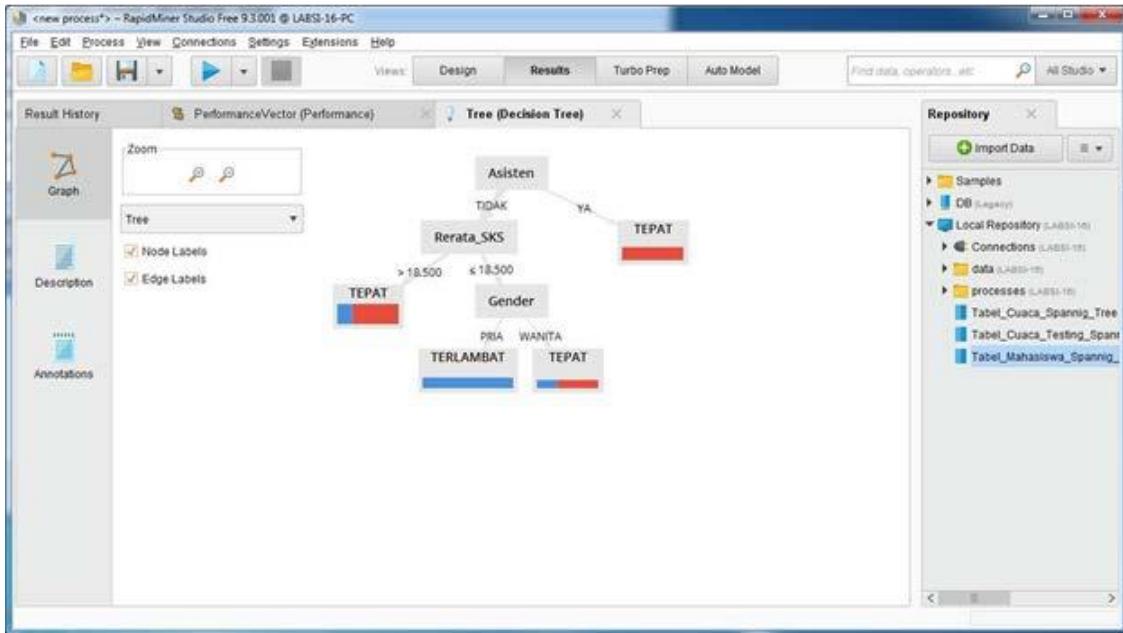


b) Carilah nilai-nilai parameter berikut :

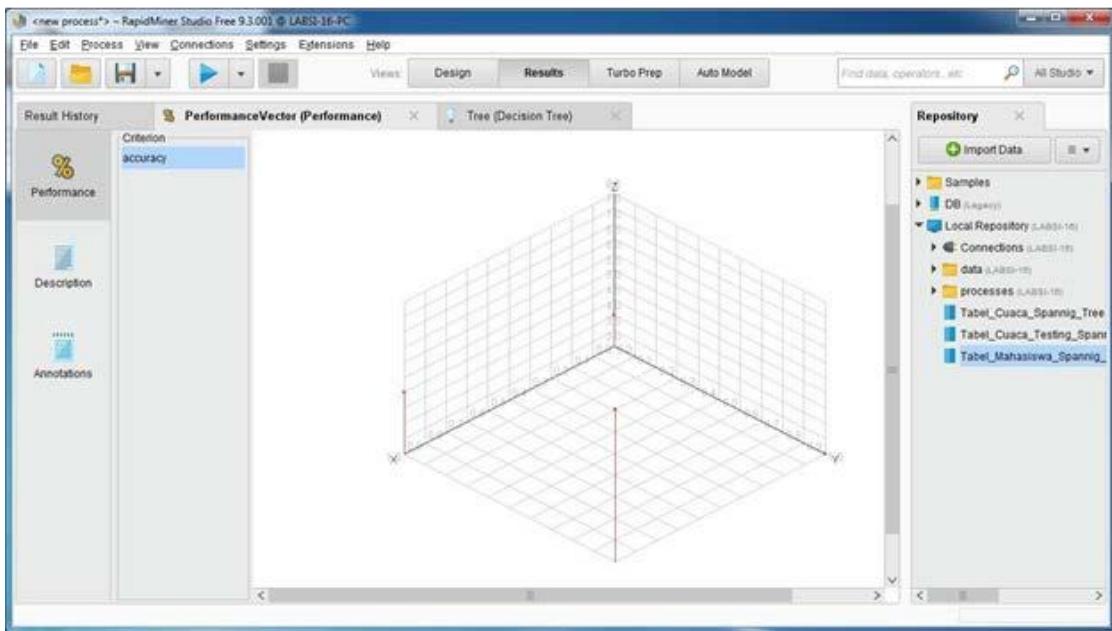
- i. Jumlah simpul daun pada pohon keputusan = **3**
- ii. Jumlah simpul keseluruhan pada pohon keputusan = **5**
- iii. Waktu yang dibutuhkan untuk proses pelatihan = **0**
- iv. Tingkat ketepatan klasifikasi = **85%**
- v. Tingkat ketidaktepatan klasifikasi = **15%**

3. Gunakan file Excel yang dikerjakan pada Tugas nomor 1 dalam Modul 6 sebagai data training.

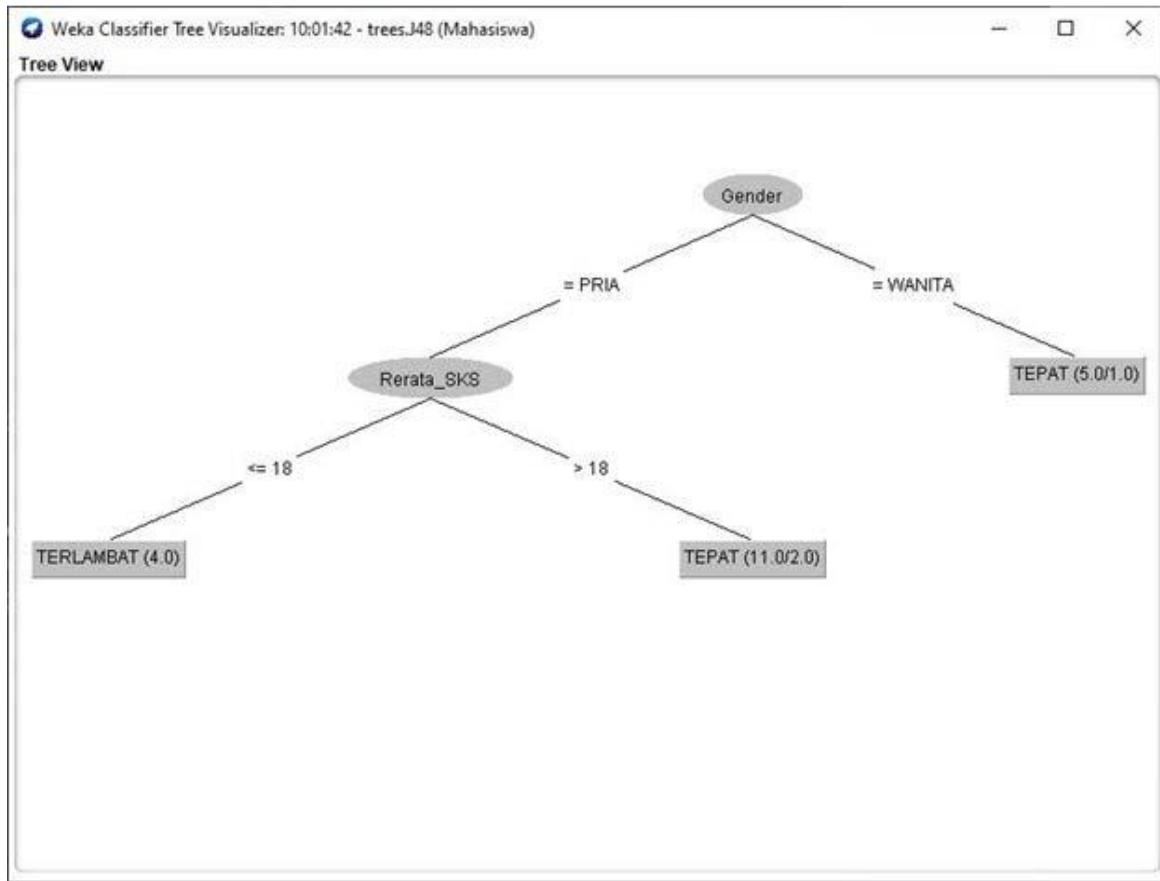
a) Buatlah dan cetaklah pohon keputusan berdasarkan data tersebut!



b) Cetaklah Perspektif Plot View dengan model Scatter. X-Axis = Gender, Y-Axis = Asisten, dan Color Column = Lama\_Studi. Nilai Jitter bisa diubah-ubah untuk memperoleh pola penyebaran yang lebih jelas.



4. Cetaklah Perspektif Plot View dengan model Scatter. X-Axis = Gender, Y-Axis = Asisten, dan Color Column = Lama\_Studi. Nilai Jitter bisa diubah-ubah untuk memperoleh pola penyebaran yang lebih jelas.



Klasifikasi yang terbentuk yaitu,

- a) Mahasiswa akan lulus TEPAT waktu jika,
  - i. PRIA = Rerata\_SKS > 18 (Nilai attribute lain diabaikan).
  - ii. WANITA = Semua akan lulus TEPAT waktu (Nilai attribute lain diabaikan).
- b) Mahasiswa akan lulus TERLAMBAT jika,
  - i. PRIA = Rerata\_SKS <= 18 (Nilai attribute lain diabaikan).
  - ii. WANITA = Tidak ada yang akan lulus TERLAMBAT (Nilai attribute lain diabaikan).

+-----+  
+-----+

## MODUL X

### CLUSTERING : K-MEANS

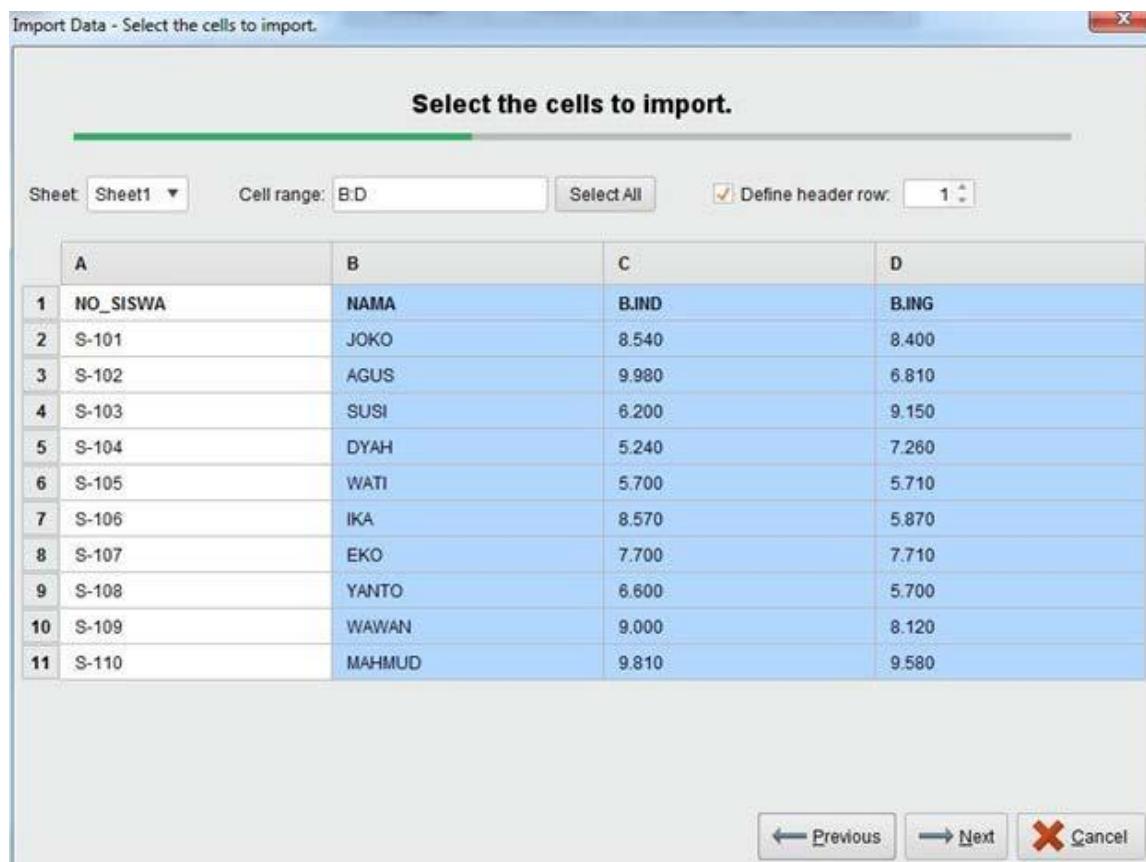
#### Langkah-langkah Praktikum

##### Algoritma K-Means Menggunakan RapidMiner

0. Membuat Tabel data data mahasiswa dan menyimpannya dengan nama **Tabel\_NilaiUjian.xls**.

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

1. Menggunakan data excel tersebut dan masukkan pada **RapidMiner**.



Import Data - Format your columns.



## Format your columns.

Replace empty cells with missing values

	NAMA	B.IND	B.ING
1	JOKO	8.540	8.400
2	AGUS	P980	5.810
3	SUSI	6.500	9.150
4	DYAH	E.2A0	7.2Gg
5	WATI	1700	5710
6	IMA	7.700	7.710
7	EKO	6.60g	\$700
8	YANTO	9.000	8120
9	WAWAN	9.910	9.580
10	IdAfimUD		

no problems.

## Where to store the data?

Local Repository (LABSI-16)

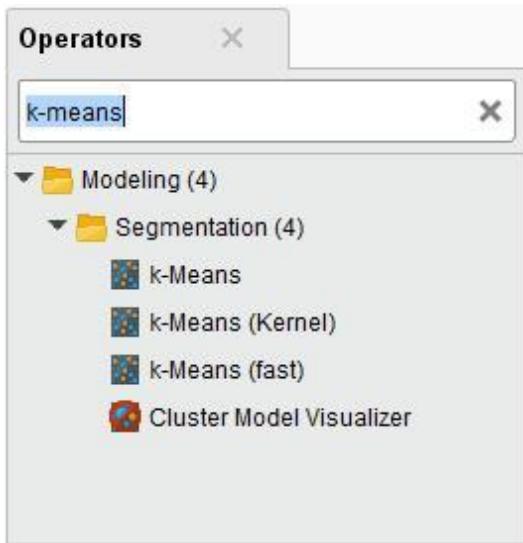
# 41° 20' \* 0C00° 5':'

Name: Data\_NilaiUjian

Location: /Local Repository/Data\_NilaiUjian

Previous Finish Cancel

3. Masukkan data tersebut ke area proses, serta cari di **K-Means** dan **SVD** pada kolom **Operator** dan masukkan ke dalam area proses.



## Operators

sod, <  
" Cleansing"; 1  
- D mensicnakr. Reduction <1:  
¿nguiar .'alue Deco uosilfon

\*d tTo results were 'ouno

Parameters  
PTOCESS

Data\_Net

---

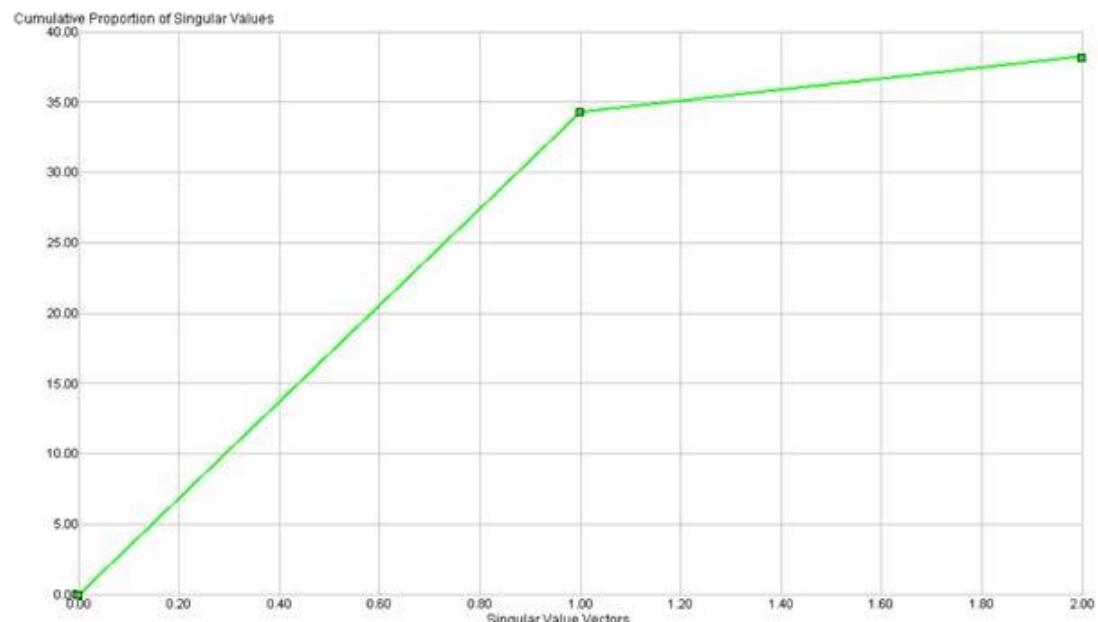
---

4. Setelah jalankan **Run**, hasilnya berupa :

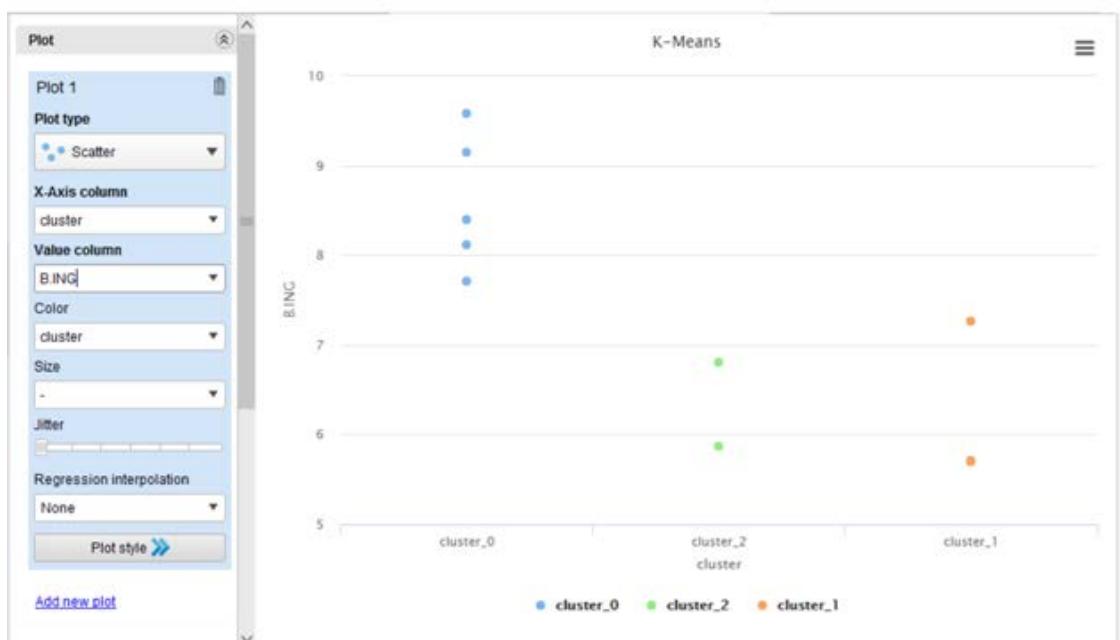
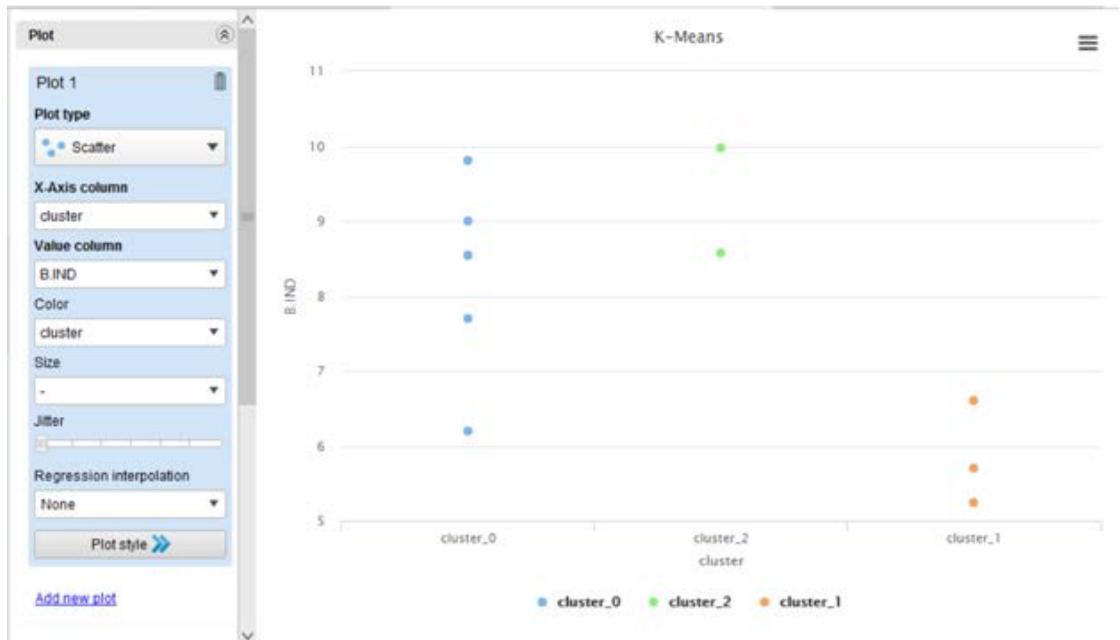
- a) SVD (Singular Value Decomposition)

Component	Singular Value	Proportion of Singular Values	Cumulative Singular Values	Cumulative Proportion of Sin...
SVD 1	34.340	0.898	34.340	0.898
SVD 2	3.906	0.102	38.246	1.000

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



b) ExampleSet (K-Means)



c) ExampleSet (SVD)

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

d) Cluster Model (Clustering)

## Cluster Model

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



### Interpretasi Hasil Algoritma K-Means

Row No.	NAMA	cluster	B.IND	B.ING
1	JOKO	cluster_0	8.540	8.400
2	AGUS	cluster_2	9.980	6.810
3	SUSI	cluster_0	6.200	9.150
4	DYAH	cluster_1	5.240	7.260
5	WATI	cluster_1	5.700	5.710
6	IKA	cluster_2	8.570	5.870
7	EKO	cluster_0	7.700	7.710
8	YANTO	cluster_1	6.600	5.700
9	WAWAN	cluster_0	9	8.120
10	MAHMUD	cluster_0	9.810	9.580

## Tugas

Dalam sebuah kelas terdapat 30 siswa telah menempuh ujian 4 mata pelajaran, yaitu Bahasa Indonesia, Bahasa Inggris, Matematika, dan IPA seperti dalam Tabel Data Nilai Ujian berikut.

D. Buatlah table berikut dengan menggunakan Microsoft Excel!

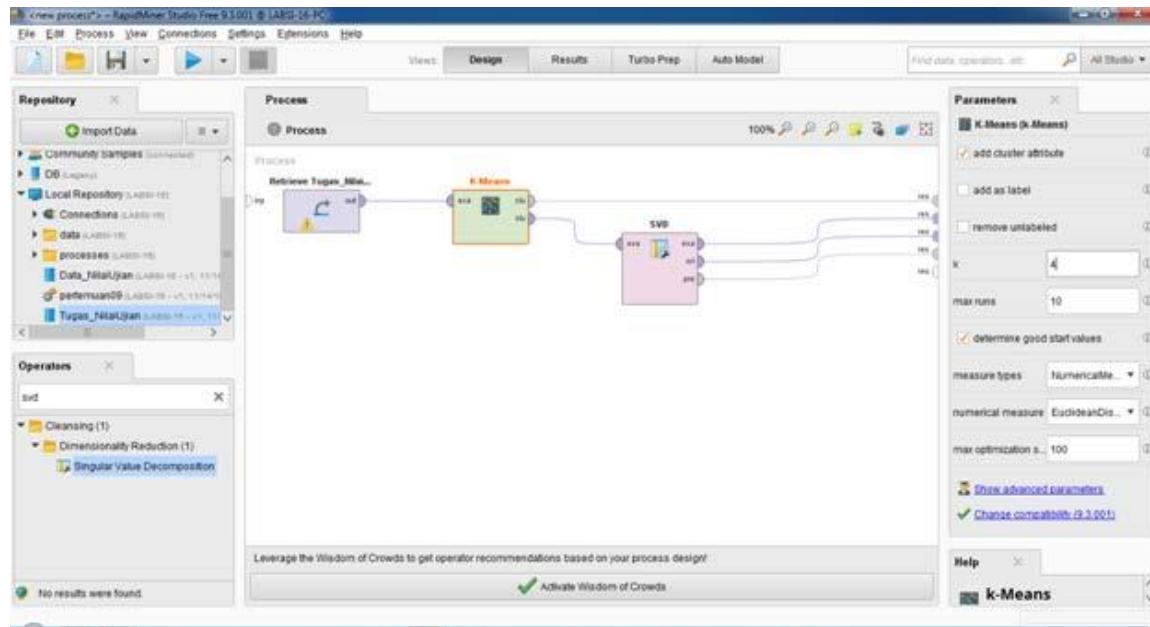
NO_SISWA	NAMA	B.IND	B.ING	MTK	IPA
S-101	JOKO	6.95	8.89	7.43	5.91
S-102	AGUS	9.06	8.31	6.89	8.86
S-103	SUSI	8.02	7.57	9.20	8.02
S-104	DYAH	8.86	9.90	9.85	8.14
S-105	WATI	7.45	6.45	8.40	8.78
S-106	IKA	9.51	9.72	7.70	8.42
S-107	EKO	8.08	8.19	9.12	5.32
S-108	YANTO	8.34	5.75	7.42	5.74
S-109	WAWAN	6.15	8.36	8.39	9.04
S-110	MAHMUD	9.99	6.36	6.87	5.57
S-111	BUDI	7.02	6.08	6.16	5.48
S-112	SANTI	9.52	9.35	6.59	9.02
S-113	DIAN	8.40	7.38	7.35	9.79
S-114	DANI	6.04	6.31	8.11	8.98
S-115	AHMAD	7.79	7.10	8.55	8.09
S-116	BAYU	6.69	7.95	5.01	6.34
S-117	RISA	7.96	8.48	7.26	9.88
S-118	RANI	8.14	6.09	7.77	7.49
S-119	YANI	5.38	7.59	6.63	9.55
S-120	RATIH	8.36	7.40	5.02	8.75
S-121	INDAH	7.07	8.48	9.42	5.06
S-122	JONO	5.07	7.83	5.31	5.86
S-123	SARAH	6.46	8.48	7.61	6.82
S-124	RAMA	8.65	7.66	9.89	6.84
S-125	BAMBANG	7.34	8.17	8.13	6.65
S-126	HADI	5.06	6.55	9.57	6.13

Untuk mengisi daftar nilai dalam table, gunakan formula berikut pada salah satu sel.  
Kemudian bias di copy-paste ke sel yang lain.

$$=5+\text{RAND}()*5$$

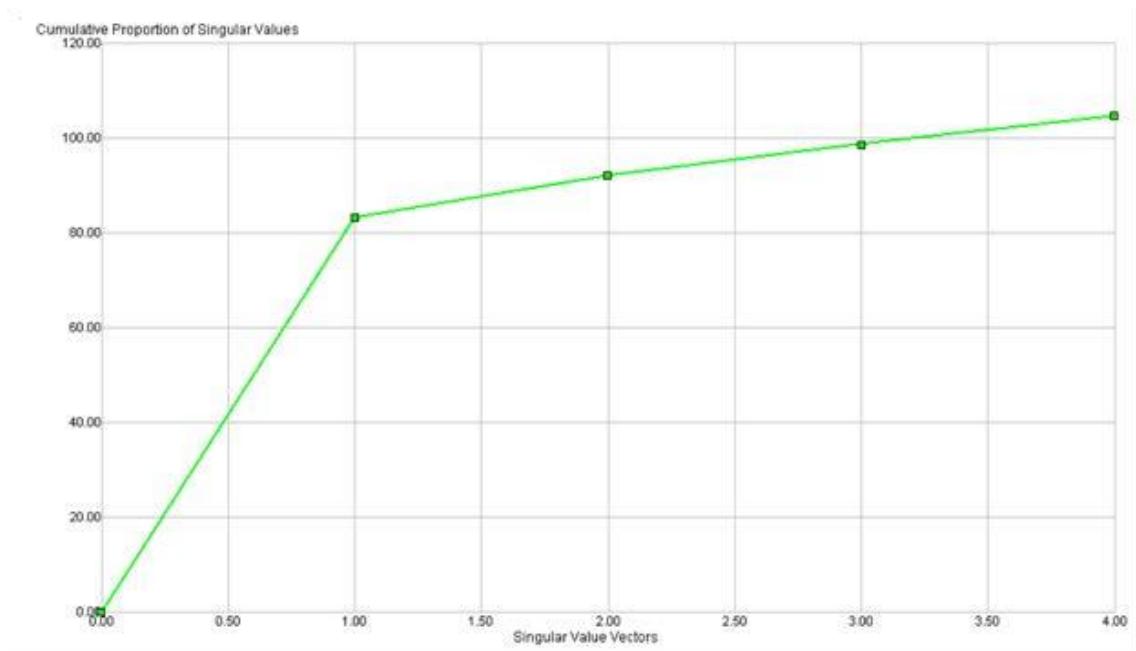
(Catatan : setiap mahasiswa pasti akan memiliki data yang berlainan, sehingga hasilnya juga berbeda)

2. Lakukan kembali kegiatan 10.4.1 dan 10.4.2 pada modul 10 isi secara lengkap menggunakan data yang terdapat pada table **Tabel Data Nilai Ujian 30 Siswa** tersebut, dengan ketentuan jumlah Cluster = 4. Catat dan tulis semua hasilnya pada lembar jawaban anda, untuk gambar bias di copy-paste.

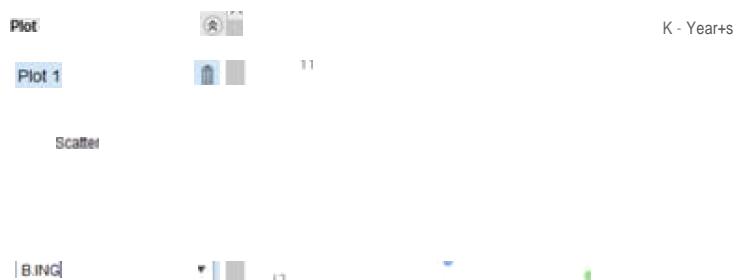


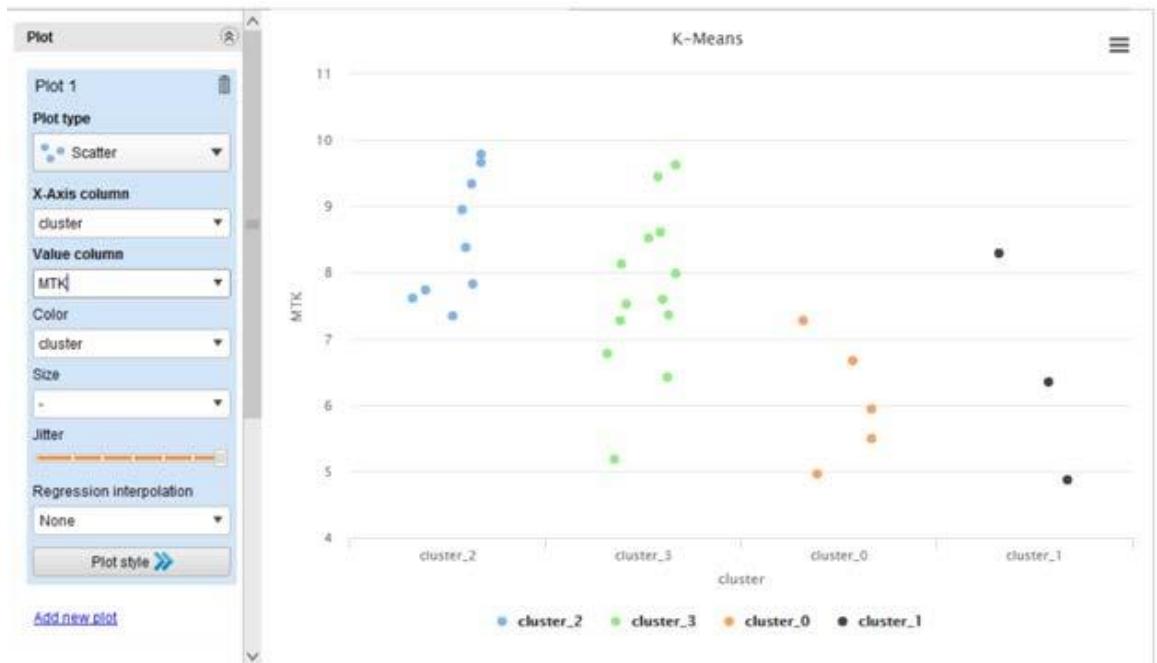
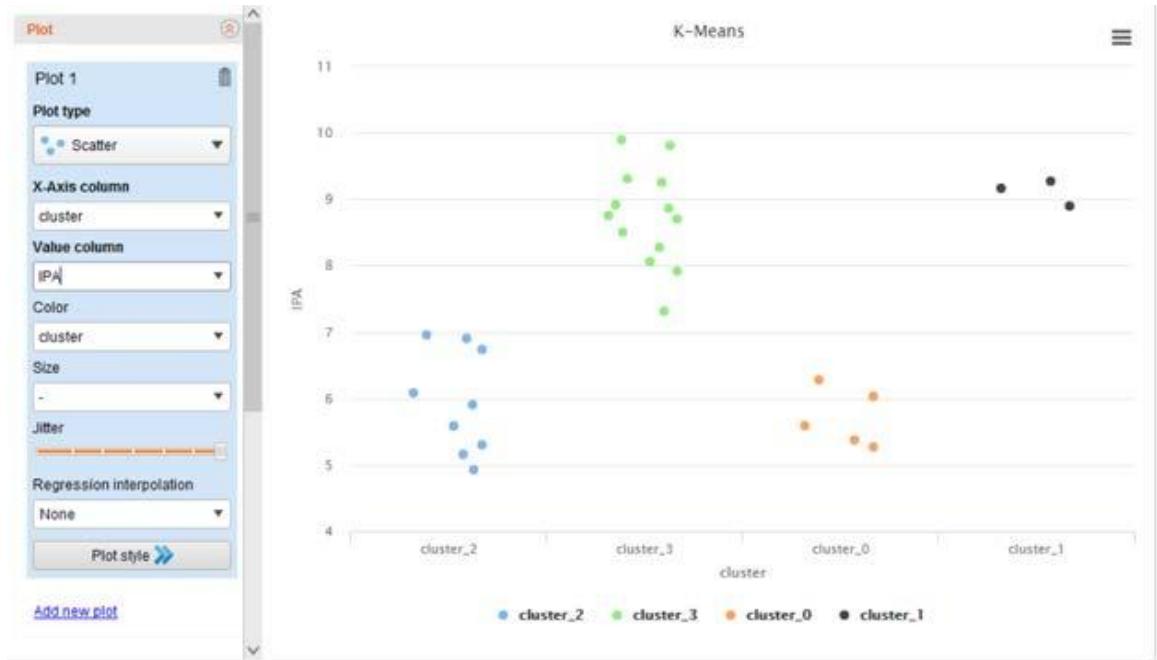
Component	Singular Value	Proportion of Singular Values	Cumulative Singular Values	Cumulative Proportion of Sing...
SVD 1	83.180	0.794	83.180	0.794
SVD 2	8.870	0.085	92.050	0.879
SVD 3	6.571	0.063	98.621	0.942
SVD 4	6.085	0.058	104.706	1.000

Attribute	SVD Vector 1	SVD Vector 2	SVD Vector 3
B.IND	0.499	-0.061	0.860
B.ING	0.506	0.271	-0.195
MTK	0.501	0.558	-0.309
IPA	0.494	-0.782	-0.356



None





 Plot



ExampleSet

Scatter

svd\_1



## Cluster Model

Clumcer 0: 5 items  
Cluster 1: 3 items  
Clumcer 2: 9 items  
Cluscer 3: 13 icems  
Total number of items: 30

Node Labels

Edge Labels

root set

0

1	<b>JOKO</b>	<b>cluster_2</b>	#sss	8886	7 428	5. 1?7
2	<b>AGUS</b>	<b>cluster_3</b>	90W	8.310	6.8B7	8 86
3	<b>SUSI</b>	duster_3	8 023	7566	9 202	8 020
4	<b>DYAH</b>	duster_3	<b>B.862</b>	9.80g	B88	<b>8.139</b>
5	WATT	<b>cluster_3</b>	7 4d9	6.452	8.400	8 .w9
6	<b>IKA</b>	duster_3	<b>g 514</b>	9 724	7 704	<b>8.424</b>
7	<b>EKO</b>	du8t8t_2	B084	8.187	0.119	5.320
8	<b>YANTO</b>	<b>cluster_0</b>	8.3a0	5 751	7a23	<b>5.738</b>
9	<b>WAWAN</b>	<b>cluster_3</b>	<b>6.153</b>	8 363	8 B92	<b>9.038</b>
40	<b>MAHMUD</b>	dx%_0	9 983	6.362	6.873	5.573
11	<b>BUDI</b>	duster_0	7 025	6 079	G 160	5m8a
12	<b>SANTI</b>	<b>cluster_3</b>	9 521	9 350	6. g7	9.Q17
13	<b>DIAN</b>		8 405	7.383	7.346	<b>9.785</b>
14	<b>DANI</b>	duster_J	60 5	6 315	<b>8.112</b>	8.985

Row No.	NAMA	cluster	B.IND	B.JNG	MTK	IPA
15	AHMAD	cluster_3	7.789	7.100	8.550	8.095
16	BAYU	cluster_0	6.693	7.950	5.011	6.336
17	RISA	cluster_3	7.961	8.481	7.264	9.885
18	RANI	cluster_3	8.145	6.094	7.771	7.485
19	YANI	cluster_1	5.384	7.586	6.626	9.553
20	RATIH	cluster_3	8.358	7.401	5.018	8.753
21	INDAH	cluster_2	7.070	8.480	9.424	5.064
22	JONO	cluster_0	5.075	7.826	5.312	5.857
23	SARAH	cluster_2	6.457	8.477	7.608	6.823
24	RAMA	cluster_2	8.651	7.657	9.892	6.835
25	BAMBANG	cluster_2	7.344	8.168	8.133	6.654
26	HADI	cluster_2	5.063	6.548	9.565	6.128
27	NANA	cluster_3	9.866	6.801	7.515	9.303
28	FEBRI	cluster_1	5.855	5.234	5.020	9.049
29	DENI	cluster_2	6.818	8.346	8.095	5.204
30	TONI	cluster_2	5.800	9.033	7.065	5.305

3. Tuliskan masing-masing nama siswa yang terdapat dalam kelompok Cluster 0, Cluster 1, Cluster 2, dan Cluster 3.

➤ Cluster 0

- YANTO
- MAHMUD
- BUDI
- BAYU
- JONO

➤ Cluster 1

- DANI
- YANI
- FEBRI

➤ Cluster 2

- JOKO
- EKO
- INDAH
- SARAH
- RAMA
- BAMBANG
- HADI
- DENI
- TONI

➤ Cluster 3

- AGUS
- SUSI
- DYAH
- WATI
- IKA
- WAWAN
- SANTI
- DIAN
- AHMAD
- RISA
- RANI
- RATIH
- NANA

..... + +

# MODUL XI

## INDUKSI DAN ATURAN ASOSIASI

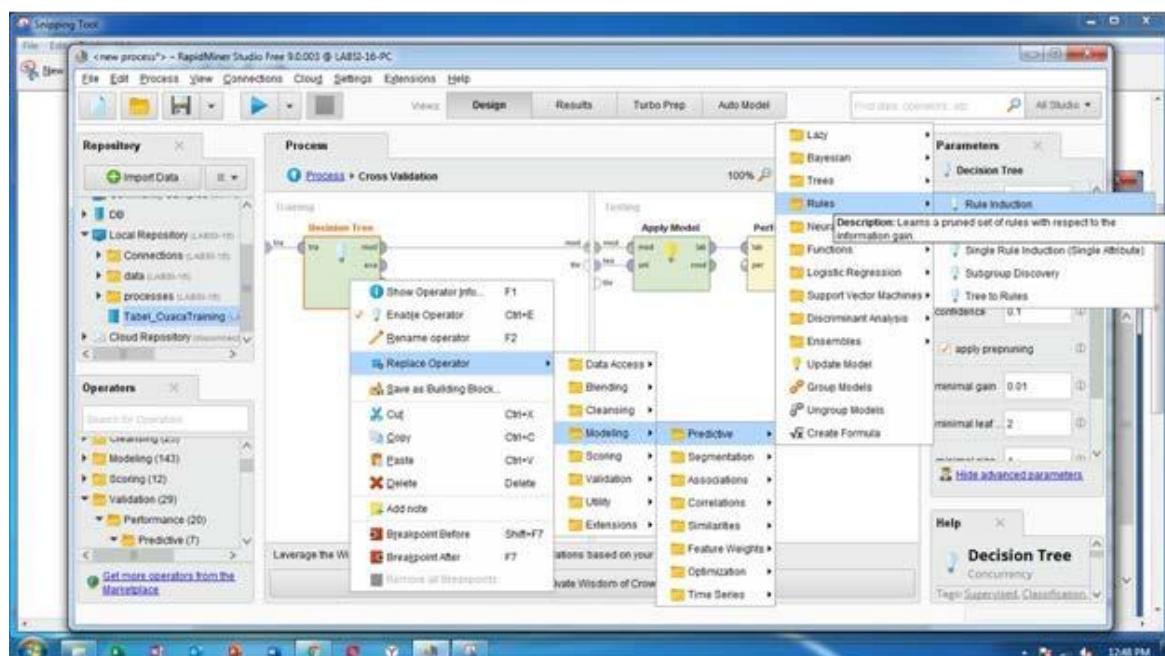
### Langkah-langkah Praktikum

#### Induksi Aturan Data Cuaca

1. Jalankan RapidMiner dan proses dat table cuaca menjadi decision tree.



2. Kembali ke Process -> Cross Validation, klik kanan pada operator Decision Tree, pilih Replace Operator -> Modeling -> Predictive -> Rules -> Rule Induction.



3. Klik **Run**, sehingga mendapatkan **RuleModel (Rule Induction)** dan **Performance Vector (Performance)**.

## RuleModel

```
if Kelembaban_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)

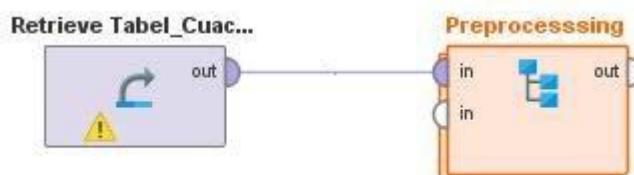
correct: 12 out of 13 training examples.
```

accuracy: 65.00% +/- 45.00% (micro average: 71.43%)

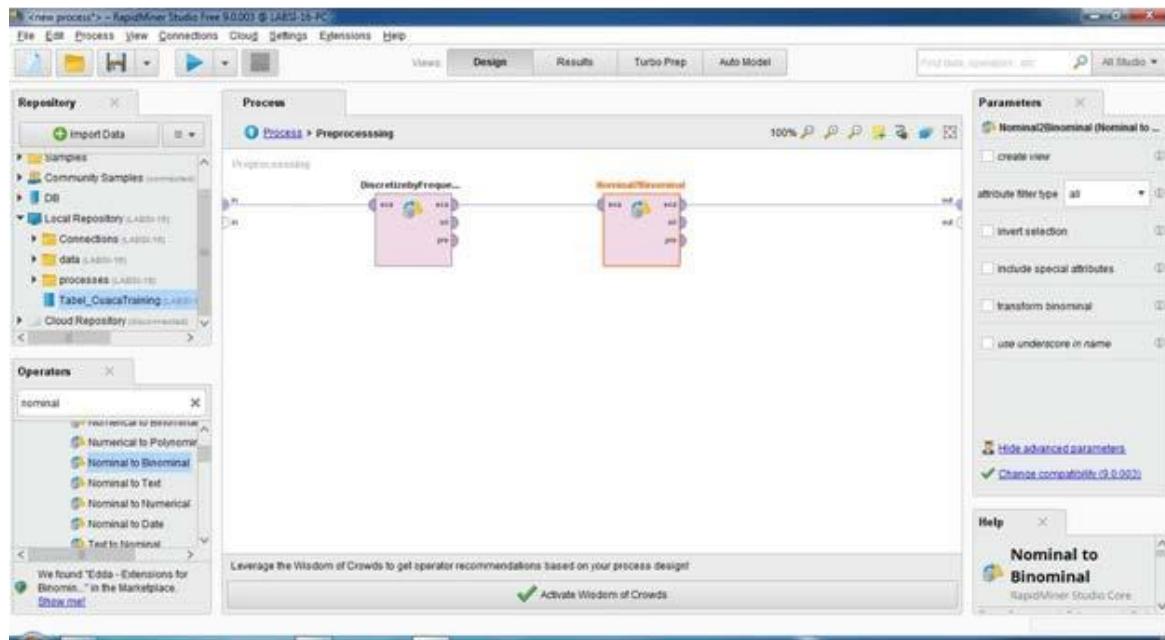
	true TIDAK	true YA	class precision
pred. TIDAK	2	1	66.67%
pred. YA	3	8	72.73%
class recall	40.00%	88.89%	

## Aturan Asosiasi Data Cuaca

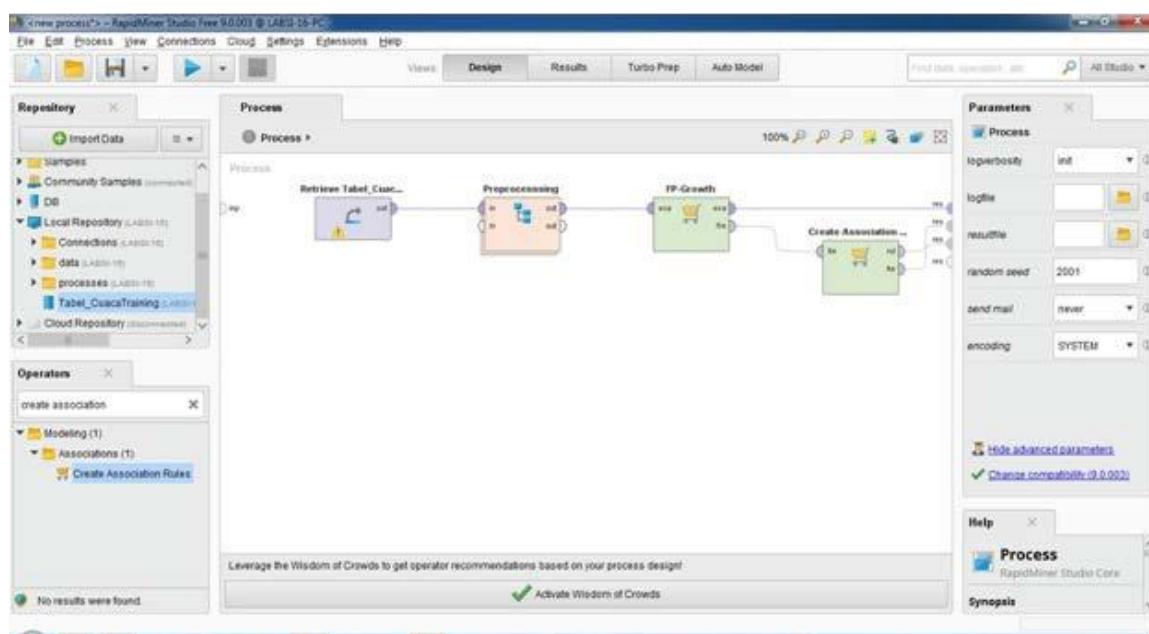
- Buat **New Process** dengan menekan tombol **New**, kembali gunakan data **Cuaca Training** dan drag ke repository **Process View**, tambahkan operator **subprocess** dan ubah Namanya menjadi **Subprocess**.



2. Klik 2X pada **Preprocessing** untuk masuk ke halaman **Nested Chain**, lalu tambahkan operator **Discretize by Frequency (DiscretizebyFrequency)** dan **Nominal to Binomial (Nominal2Binomial)** dan hubungkan keduanya.



3. Kembali ke **Main Process**, da tambhkan operator **FP-Growth** dan **Create Association Rules**, untuk **FP-Growth** ubah nilai parameter pada min support = 0.1.



4. Klik **Run** dan tampilkan :

➤ **Frequent Item Set (FP-Growth)**

No. of Sets: 26	Size	Support	Item 1	Item 2	Item 3	Item 4
Total Max. Size: 4						
Min. Size: <input type="text" value="1"/>	1	0.500	Kelembaban_udara			
Max. Size: <input type="text" value="4"/>	1	0.429	Berangin			
Contains Item:	1	0.429	Suhu			
<input type="text"/>	1	0.357	Cuaca = Cerah			
<input type="text"/>	1	0.357	Cuaca = Hujan			
<input type="text"/>	1	0.286	Cuaca = Mendung			
<input type="text"/>	2	0.214	Kelembaban_udara	Berangin		
<input type="text"/>	2	0.214	Kelembaban_udara	Suhu		
<input type="text"/>	2	0.214	Kelembaban_udara	Cuaca = Cerah		
<input type="text"/>	2	0.143	Kelembaban_udara	Cuaca = Hujan		
<input type="text"/>	2	0.143	Kelembaban_udara	Cuaca = Mendung		
<input type="text"/>	2	0.143	Berangin	Suhu		
<input type="text"/>	2	0.143	Berangin	Cuaca = Cerah		
<input type="text"/>	2	0.143	Berangin	Cuaca = Hujan		
<input type="text"/>	2	0.143	Berangin	Cuaca = Mendung		

No. of Sets: 26	Size	Support	Item 1	Item 2	Item 3	Item 4
Total Max. Size: 4						
Min. Size: <input type="text" value="1"/>	2	0.143	Kelembaban_udara	Cuaca = Mendung		
Max. Size: <input type="text" value="4"/>	2	0.143	Berangin	Suhu		
Contains Item:	2	0.143	Berangin	Cuaca = Cerah		
<input type="text"/>	2	0.143	Berangin	Cuaca = Hujan		
<input type="text"/>	2	0.143	Berangin	Cuaca = Mendung		
<input type="text"/>	2	0.214	Suhu	Cuaca = Cerah		
<input type="text"/>	2	0.071	Suhu	Cuaca = Hujan		
<input type="text"/>	2	0.143	Suhu	Cuaca = Mendung		
<input type="text"/>	3	0.071	Kelembaban_udara	Berangin	Suhu	
<input type="text"/>	3	0.071	Kelembaban_udara	Berangin	Cuaca = Cerah	
<input type="text"/>	3	0.071	Kelembaban_udara	Berangin	Cuaca = Hujan	
<input type="text"/>	3	0.071	Kelembaban_udara	Berangin	Cuaca = Mendung	
<input type="text"/>	3	0.143	Kelembaban_udara	Suhu	Cuaca = Cerah	
<input type="text"/>	3	0.071	Kelembaban_udara	Suhu	Cuaca = Mendung	
<input type="text"/>	3	0.143	Berangin	Suhu	Cuaca = Cerah	
<input type="text"/>	4	0.071	Kelembaban_udara	Berangin	Suhu	Cuaca = Cerah

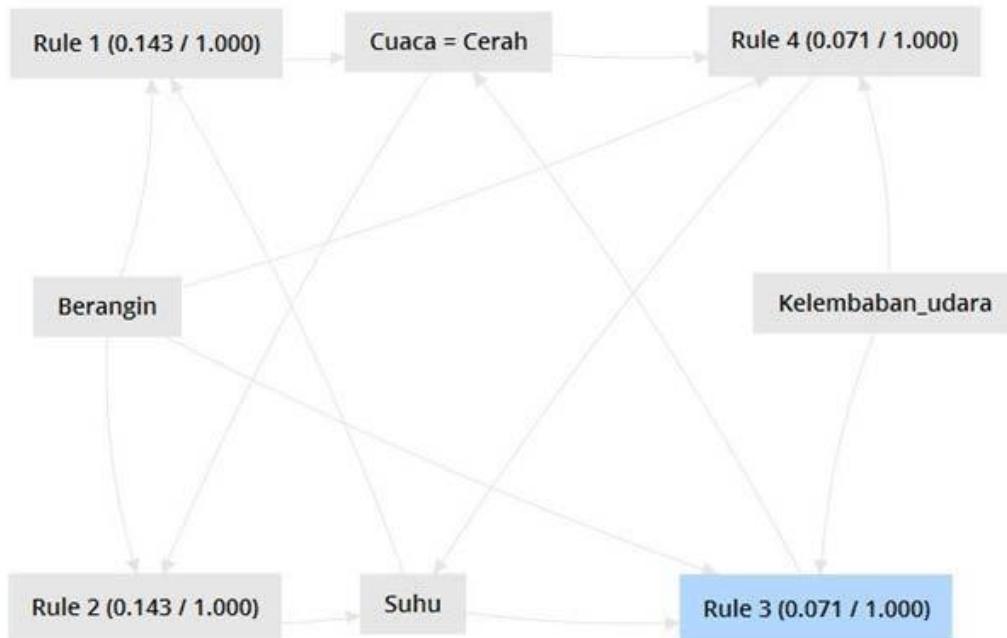
➤ Association Rules (Create Association Rules)

a) Table Mode

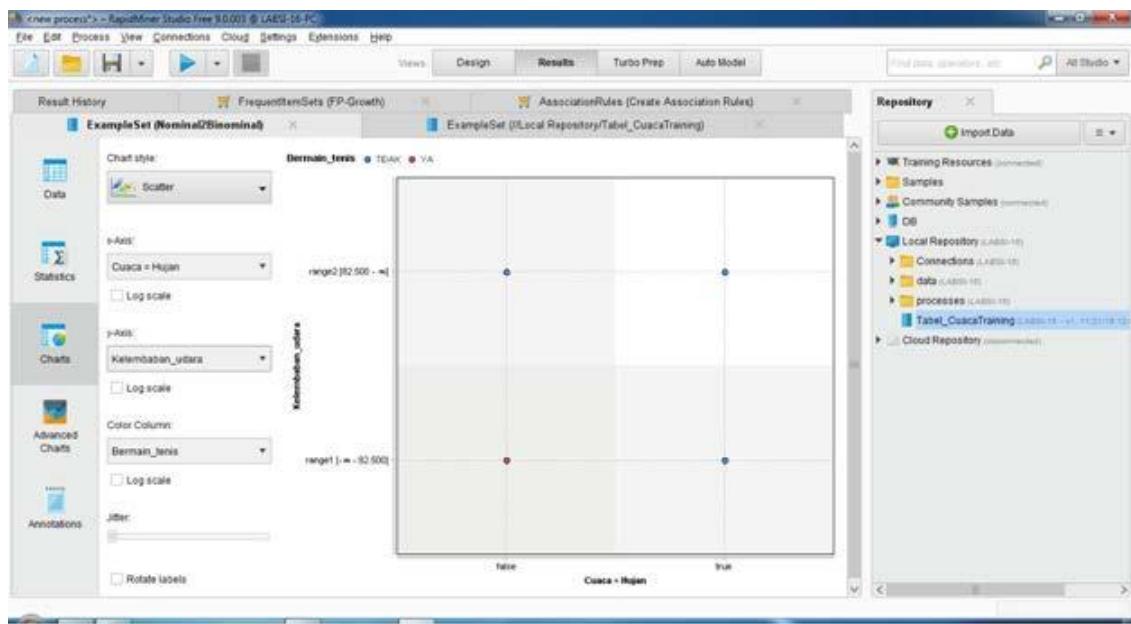
No.	Premises	Conclusion
1	Berangin, Suhu	Cuaca = Cerah
2	Berangin, Cuaca = Cerah	Suhu
3	Kelembaban_udara, Berangin, Suhu	Cuaca = Cerah
4	Kelembaban_udara, Berangin, Cuaca = Cerah	Suhu

Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
0.143	1	1	-0.143	0.092	2.800	$\infty$
0.143	1	1	-0.143	0.082	2.333	$\infty$
0.071	1	1	-0.071	0.046	2.800	$\infty$
0.071	1	1	-0.071	0.041	2.333	$\infty$

b) Graph View



## ➤ ExampleSet (Nominal2Binomial)



## Tugas

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

## 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.

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
- b) Number of bins = 3

Carilah masing-masing nilai berikut berdasarkan number of bins-nya:

- a) Jumlah set aturan asosiasi dan total max size yang terbentuk berdasarkan FP-Growth (table view)! Gambarkan tabelnya!

➤ Number of bins = 2

No. of Sets: 15	Size	Support	Item 1	Item 2	Item 3	Item 4
Total Max. Size: 4	1	0.750	Gender			
Min. Size: <input type="text" value="1"/>	1	0.300	Asal_Sekolah			
Max. Size: <input type="text" value="4"/>	1	0.250	Asisten			
Contains Item:	1	0.250	Rerata_SKS			
	2	0.250	Gender	Asal_Sekolah		
<input type="button" value="Update View"/>	2	0.200	Gender	Asisten		
	2	0.250	Gender	Rerata_SKS		
	2	0.150	Asal_Sekolah	Asisten		
	2	0.150	Asal_Sekolah	Rerata_SKS		
	2	0.150	Asisten	Rerata_SKS		
	3	0.100	Gender	Asal_Sekolah	Asisten	
	3	0.150	Gender	Asal_Sekolah	Rerata_SKS	
	3	0.150	Gender	Asisten	Rerata_SKS	
	3	0.100	Asal_Sekolah	Asisten	Rerata_SKS	

➤ Number of bins = 3

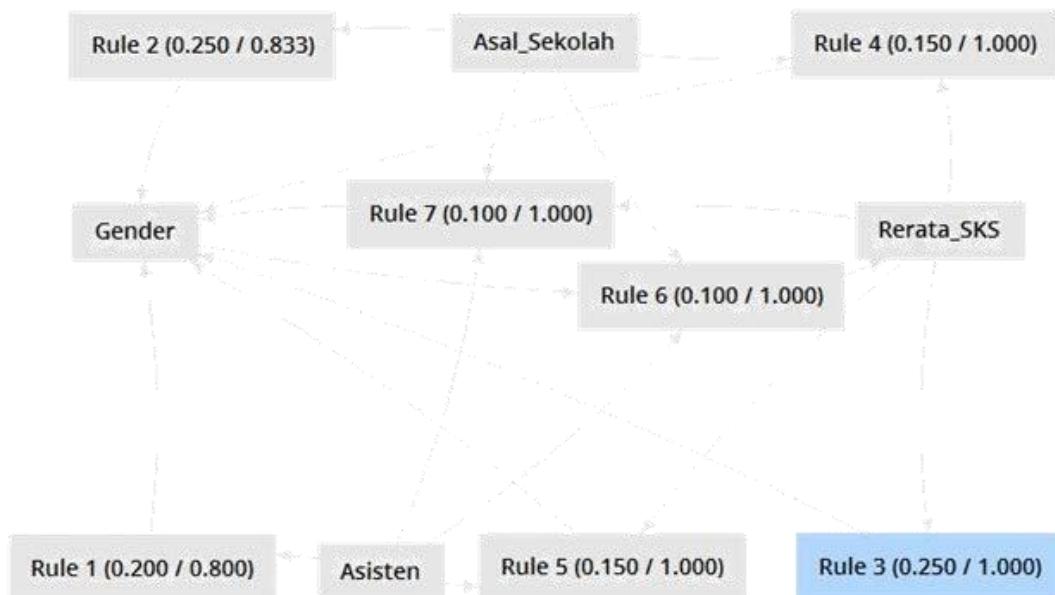
No. of Sets: 7	Size	Support	Item 1	Item 2	Item 3
Total Max. Size: 3	1	0.750	Gender		
Min. Size: <input type="text" value="1"/>	1	0.300	Asal_Sekolah		
Max. Size: <input type="text" value="3"/>	1	0.250	Asisten		
Contains Item:	2	0.250	Gender	Asal_Sekolah	
	2	0.200	Gender	Asisten	
<input type="button" value="Update View"/>	2	0.150	Asal_Sekolah	Asisten	
	3	0.100	Gender	Asal_Sekolah	Asisten

b) Jumlah data pasangan premis dan kesimpulan pada Association Rules (Create Association Rules)! Gambarkan tabelnya! Gambarkan pula grafik yang terbentuk!

➤ Number of bins = 2

No.	Premises	Conclusion
2	Asal_Sekolah	Gender
3	Rerata_SKS	Gender
4	Asal_Sekolah, Rerata_SKS	Gender
5	Asisten, Rerata_SKS	Gender
6	Gender, Asal_Sekolah, Asisten	Rerata_SKS
7	Asal_Sekolah, Asisten, Rerata_SKS	Gender

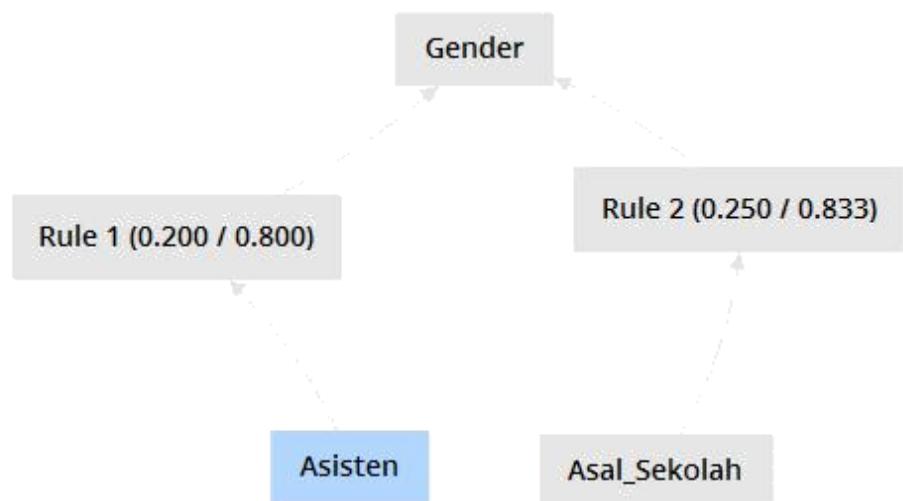
Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
0.250	0.833	0.962	-0.350	0.025	1.111	1.500
0.250	1	1	-0.250	0.062	1.333	$\infty$
0.150	1	1	-0.150	0.038	1.333	$\infty$
0.150	1	1	-0.150	0.038	1.333	$\infty$
0.100	1	1	-0.100	0.075	4	$\infty$
0.100	1	1	-0.100	0.025	1.333	$\infty$



- Number of bins = 3

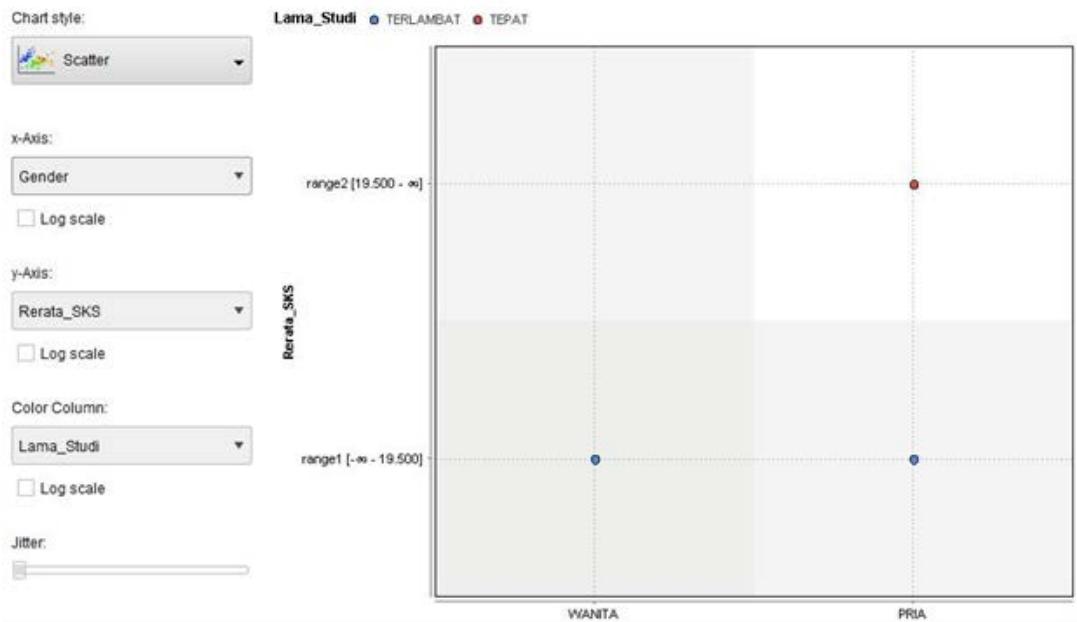
No.	Premises	Conclusion
2	Asal_Sekolah	Gender

Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
0.250	0.833	0.962	-0.350	0.025	1.111	1.500

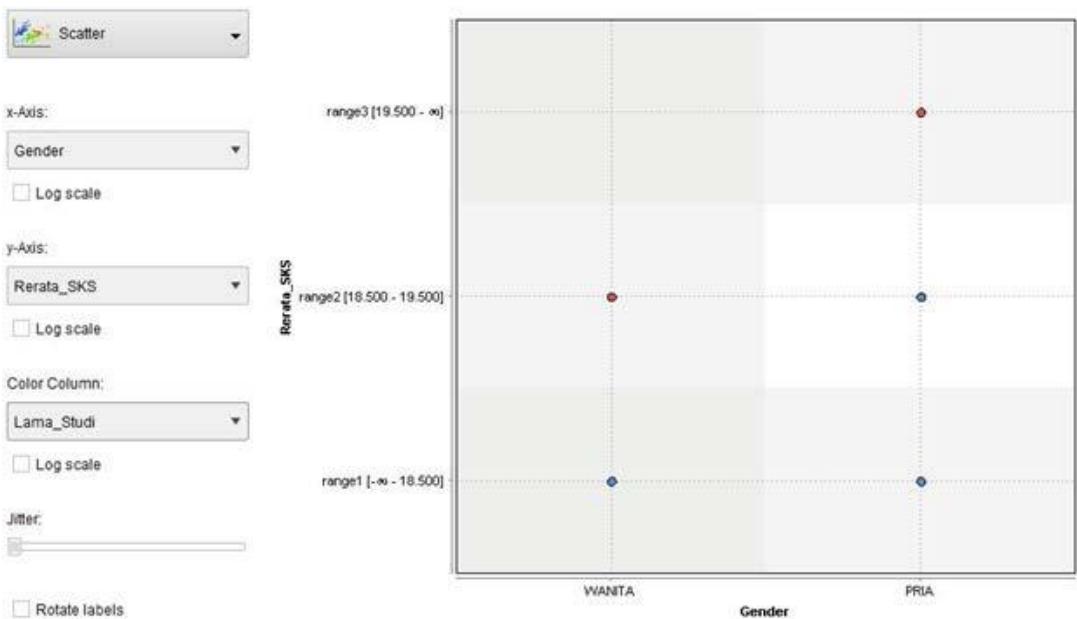


c) Gambarkan grafik chart pola distribusi data pada ExampleSet yang terbentuk!

➤ Number of bins = 2



➤ Number of bins = 3



+ ..... +

## **MODUL XII**

### **REGRESI LINIER SEDERHANA**

#### **Langkah-langkah Praktikum**

##### **Contoh Kasus :**

Dalam sebuah kelas yang memiliki 10 siswa dilakukan sebuah survei terhadap lama belajar seorang siswa dan nilai hasil ujiannya. Data siswa tersebut akan kita gunakan sebagai dasar perhitungan untuk memprediksi nilai ujian terhadap siswa lain berdasarkan lama belajarnya.

#### **Induksi Aturan Data Cuaca**

##### **Hipotesis :**

Bagaimana

1. Membuat data pada Microsoft Excel dengan nama **Tabel\_LamaBelajardanNilaiUjian.xls**.

<b>No_Siswa</b>	<b>Nama</b>	<b>Lama Belajar (Jam)</b>	<b>Nilai</b>
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. Mengimport data table **Tabel\_LamaBelajardanNilaiUjian.xls** ke **RapidMiner** yang akan digunakan sebagai regresi Linear, dengan memberikan pengaturan tipe dan jenis data pada atribut,
  - NO\_SISWA : polynomial, id
  - NAMA : pilih Exclude column
  - LAMA JAM BELAJAR : integer
  - NILAI : integer, label

Simpan dengan nama **Data\_LamaBelajar dan NilaiUjian**.

Import Data - Format your columns.

**Format your columns.**

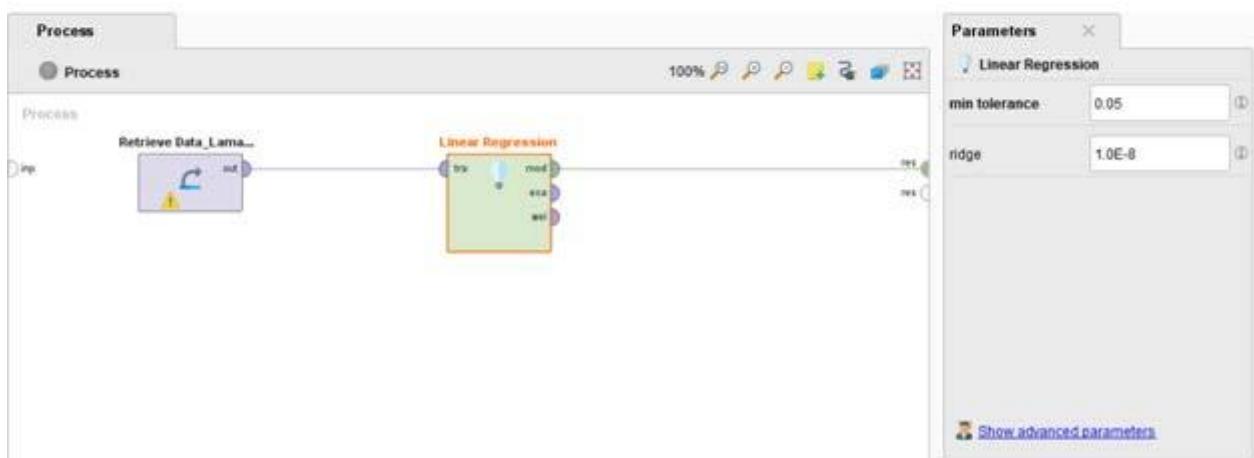
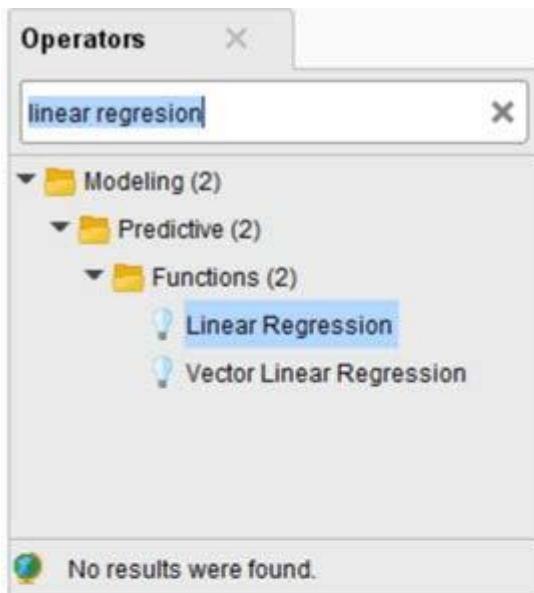
Replace errors with missing values ⓘ

No_Siswa <i>polynominal id</i>	Nama <i>polynominal</i>	Lama Belajar (Jam) <i>integer</i>	Nilai <i>integer 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	BA	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

 no problems.

3. Masukkan **Data\_LamaBelajardanNilaiUjian** ke dalam area process, dan tambahkan **Linier Regresion**, lalu klik pada operator **Linear Regresion** dan tentukan parameter **min tolerance = 0.05** (Batas toleransi sebesar 5%).



4. Jalankan proses dengan meng-klik **Run**.

➤ Table View (Mencari besarnya nilai t-hitung)

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Lama Belajar (J... (Intercept)	21.608 492.769	7.645 96.909	0.707 ?	1 ?	2.827 5.085	0.022 0.001	** ***

- Text View (Mencari model regresi)

## LinearRegression

```
21.608 * Lama Belajar (Jam)
+ 492.769
```

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

#### Hipotesis :

Bagaimana memprediksi nilai ujian siswa berdasarkan lama belajarnya menggunakan model regresi linier yang telah dihasilkan menggunakan RapidMiner.

2. Membuat table data baru di **Tabel\_LamaBelajardanNilaiUjian.xls** Pada **Sheet 2**.

No_Siswa	Nama	Lama Belajar (Jam)
S-111	BUDI	12
S-112	SANTI	13
S-113	DIAN	14
S-114	DANI	11
S-115	AHMAD	5
S-116	BAYU	13
S-117	RISA	9
S-118	RANI	10
S-119	YANI	10
S-120	RATIH	9

3. Mengimport data tersebut ke **RapidMiner** yang akan digunakan sebagai regresi Linear (Sama seperti percobaan **12.4.1**), dengan memberikan pengaturan tipe dan jenis data pada atribut,

- NO\_SISWA : polynomial, id
- NAMA : pilih Exclude column
- LAMA JAM BELAJAR : integer

Simpan dengan nama **Data\_PrediksiNilaiUjian**

---

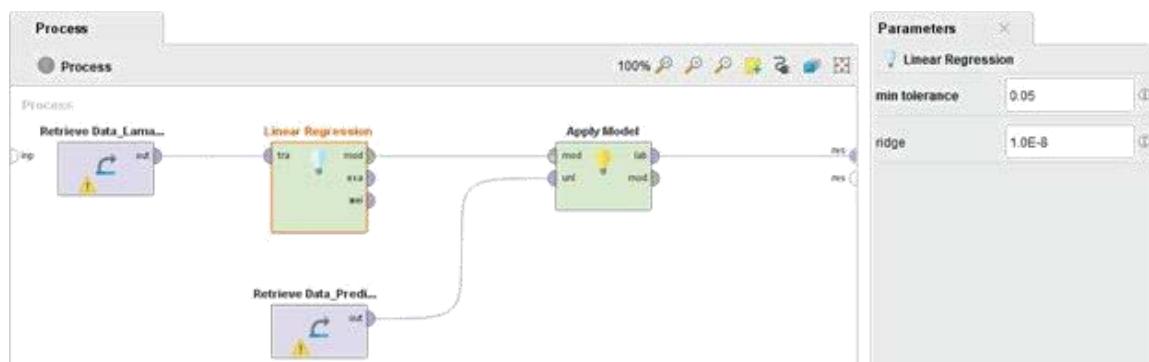
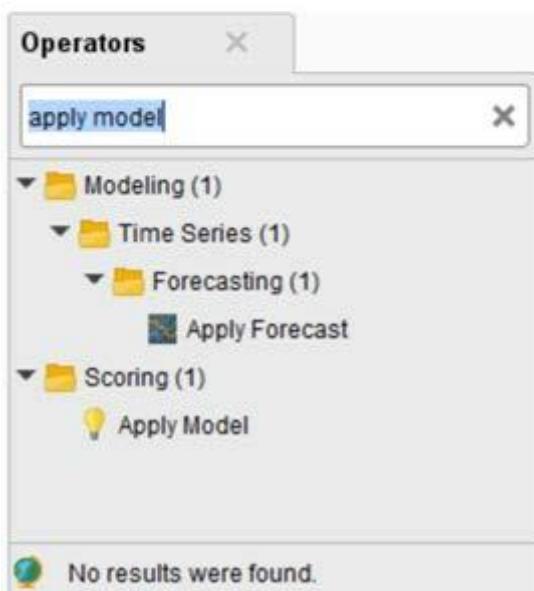
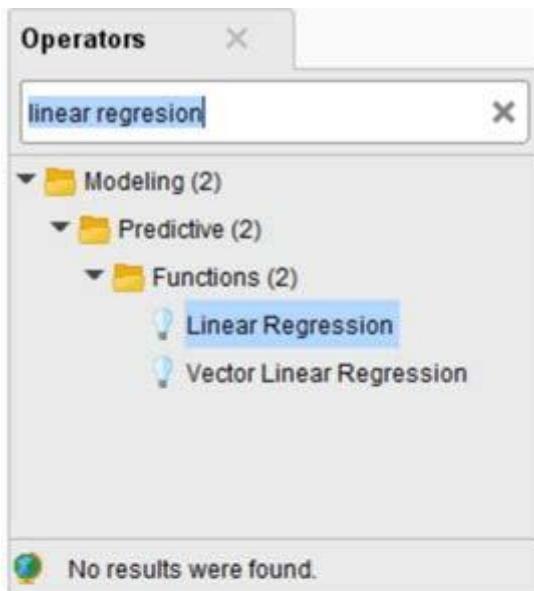
## Fermat your columns.

---

Replace errors with missing values

No_Siswa polynomial		0 * La r\lo Boda ar {Jom) 'clear	e -
1 S'1T1		12	
g g-1Tg			
Z S-1t3			
4 S 111		11	
~i S-11?			
6 5-11†		13	
8 5-11		*3	
9 S'1T9		10	
10 S-120		0	

3. Masukkan **Data\_PrediksiNilaiUjian** ke dalam area process, dan tambahkan **Linier Regresion** dan **Apply Model**, lalu klik pada operator Linear Regresion dan tentukan parameter **min tolerance = 0.05** (Batas toleransi sebesar 5%).

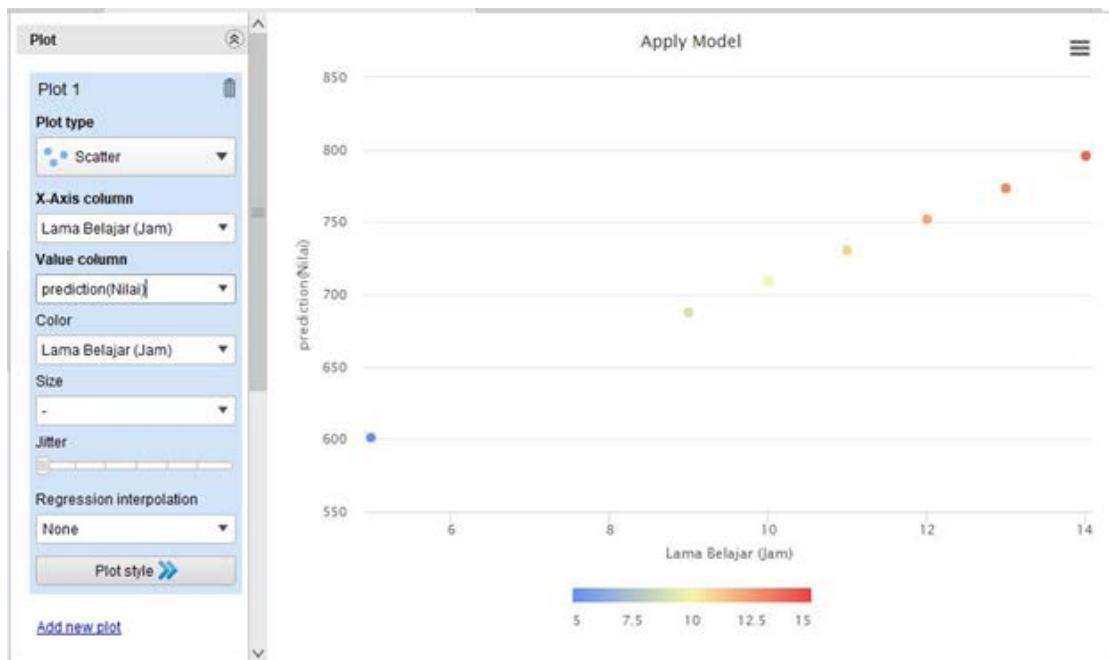


4. Jalankan proses dengan meng-klik tombol **Run**.

➤ Data View (hasil Prediksi nilai ujian)

Row No.	No_Siswa	prediction(Nilai)	Lama Belaja...
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

➤ Chart View (Scatter Plot)



## Pembuktian Model Regresi

Melakukan pembuktian hasil prediksi menggunakan **RapidMiner** dengan hasil prediksi menggunakan model regresi yang telah dihasilkan pada kegiatan 12.4.1

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

Memasukkan nilai variable  $X_1$  ke dalam model regresi :

1. No. SISWA = S-112, NAMA = SANTI, memiliki nilai  $X_1 = 13$   
Jam. Sehingga,  $Y = (21,608 \times 13) + 492,769 = 773,672$
2. No. SISWA = S-113, NAMA = DIAN, memiliki nilai  $X_1 = 14$   
Jam. Sehingga,  $Y = (21,608 \times 14) + 492,769 = 795,28$
3. No. SISWA = S-114, NAMA = DANI, memiliki nilai  $X_1 = 11$   
Jam. Sehingga,  $Y = (21,608 \times 11) + 492,769 = 730,456$
4. No. SISWA = S-116, NAMA = BAYU, memiliki nilai  $X_1 = 13$   
Jam. Sehingga,  $Y = (21,608 \times 13) + 492,769 = 773,672$
5. No. SISWA = S-112, NAMA = SANTI, memiliki nilai  $X_1 = 13$   
Jam. Sehingga,  $Y = (21,608 \times 13) + 492,769 = 773,672$
6. No. SISWA = S-113, NAMA = DIAN, memiliki nilai  $X_1 = 14$   
Jam. Sehingga,  $Y = (21,608 \times 14) + 492,769 = 795,28$
7. No. SISWA = S-114, NAMA = DANI, memiliki nilai  $X_1 = 11$   
Jam. Sehingga,  $Y = (21,608 \times 11) + 492,769 = 730,456$
8. No. SISWA = S-116, NAMA = BAYU, memiliki nilai  $X_1 = 13$   
Jam. Sehingga,  $Y = (21,608 \times 13) + 492,769 = 773,672$

No_Siswa	Nama	Lama Belajar (Jam)		
S-111	BUDI	12		752,064
S-112	SANTI	13		773,672
S-113	DIAN	14		795,28
S-114	DANI	11		730,456
S-115	AHMAD	5		600,808
S-116	BAYU	13		773,672
S-117	RISA	9		687,24
S-118	RANI	10		708,848
S-119	YANI	10		708,848
S-120	RATIH	9		687,24

### Tugas

#### Kasus :

Dalam sebuah survei terhadap 15 kepala keluarga telah diperoleh variable pendapatan rata-rata perbulan, jumlah anggota keluarga yang tinggal serumah, dan daya beli rata-rata perbulan.

#### Hipotesis :

Bagaimana model regresi linier yang terbentuk, dan lakukan prediksi terhadap 10 data yang belum diketahui nilai daya belinya.

1. Buatlah table berikut dengan menggunakan Microsoft Excel.

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

2. Buatlah proses Regresi Linier Sederhana menggunakan RapidMiner dengan ketentuan sebagai berikut.
  - a) Variable bebas (X) = Pendapatan (X<sub>1</sub>), Jumlah Anggota Keluarga (X<sub>2</sub>)
  - b) Variable terikat (Y) = Daya Beli
  - c) Toleransi yang digunakan = 5%

## Pendapatan (X<sub>1</sub>)

Import Data - Format your columns.

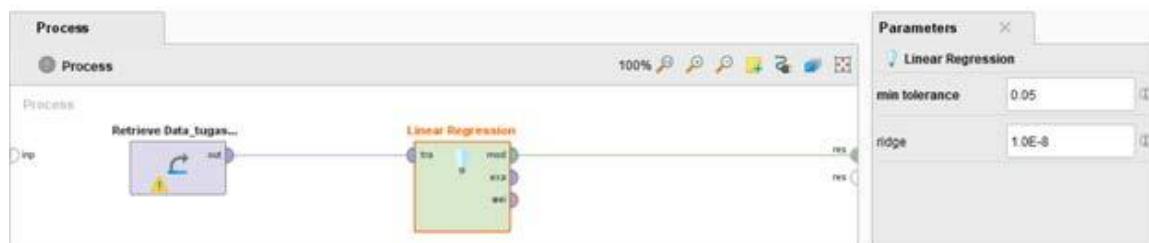
**Format your columns.**

Replace errors with missing values ①

NO. RESPONDEN	polynomial <i>id</i>	PENDAPATAN (RUPIAH)	integer	JUMLAH ANGGOTA K...	integer	DAYA BELI (RUPIAH)	integer <i>label</i>
1	1	1000000	6		834000		
2	2	1400000	7		1200000		
3	3	200000	3		134000		
4	4	1400000	6		1167000		
5	5	500000	3		334000		
6	6	1700000	5		1360000		
7	7	400000	3		267000		
8	8	1900000	5		1520000		
9	9	300000	3		200000		
10	10	500000	4		375000		
11	11	700000	7		600000		
12	12	1900000	3		1267000		
13	13	800000	4		600000		

✓ no problems.

◀ Previous ▶ Next ✖ Cancel



Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
PENDAPATAN (RUPIAH)	0.788	0.038	0.985	1	20.731	0.000	***
(Intercept)	-7467.579	44843.156	?	?	-0.167	0.870	

## LinearRegression

$$0.788 * \text{PENDAPATAN (RUPIAH)} \\ - 7467.579$$

## Jumlah Anggota Keluarga (X<sub>2</sub>)

Import Data - Format your columns.

**Format your columns.**

Replace errors with missing values ⓘ

NO. RESPONDEN	PENDAPATAN (RUPIAH)	JUMLAH ANGGOTA K...	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	800000

no problems.

← Previous    → Next    X Cancel



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

## LinearRegression

```
151688.679 * JUMLAH ANGGOTA KELUARGA
+ 98652.830
```

3. Tentukan apakah variable  $X_1$  dan  $X_2$  mempengaruhi secara signifikan terhadap nilai variable Y berdasarkan besarnya nilai t-stat?

Semisal ambil contoh data **NO. RESPONDEN = 1 :**

$$X_1 \Rightarrow Y = 0,788 X 1.000.000 - 7.463,579 = 780.536,421$$

$$X_2 \Rightarrow Y = 151.688,679 X 6 + 98.652,830 = 811.479,244$$

Jadi, melihat dari hasil yang di dapat variable  $X_1$  dan  $X_2$  memang mempengaruhi secara signifikan terhadap nilai Y.

4. Tuliskan model persamaan regresi linier sederhana yang terbentuk!

$$X_1 \Rightarrow Y = 0,788 X \text{ Pendapatan (Rupiah)} - 7.463,579$$

$$X_2 \Rightarrow Y = 151.688,679 X \text{ Jumlah Anggota Keluarga} + 98.652,830$$

5. Gunakan data testing untuk menjawab perintah berikut :

NO. RESPONDEN	PENDAPATAN (RUPIAH)	JUMLAH ANGGOTA KELUARGA
1	900000	5
2	800000	3
3	500000	2
4	1900000	6
5	600000	2
6	800000	5
7	1000000	6
8	1100000	4
9	1000000	4
10	500000	3

- a) Lakukan prediksi Daya Beli (Y) dengan menggunakan Model Persamaan Regresi Linier dari hasil pertanyaan nomor 4!

NO. RESPONDEN	PENDAPATAN (RUPIAH)	JUMLAH ANGGOTA KELUARGA	X1	X2
1	900000	5	701731,421	857096,225
2	800000	3	622931,421	553718,867
3	500000	2	386531,421	402030,188
4	1900000	6	1489731,421	1008784,904
5	600000	2	465331,421	402030,188
6	800000	5	622931,421	857096,225
7	1000000	6	780531,421	1008784,904
8	1100000	4	859331,421	705407,546
9	1000000	4	780531,421	705407,546
10	500000	3	386531,421	553718,867

b) Lakukan prediksi Daya Beli (Y) menggunakan RapidMiner!

Pendapatan ( $X_1$ )

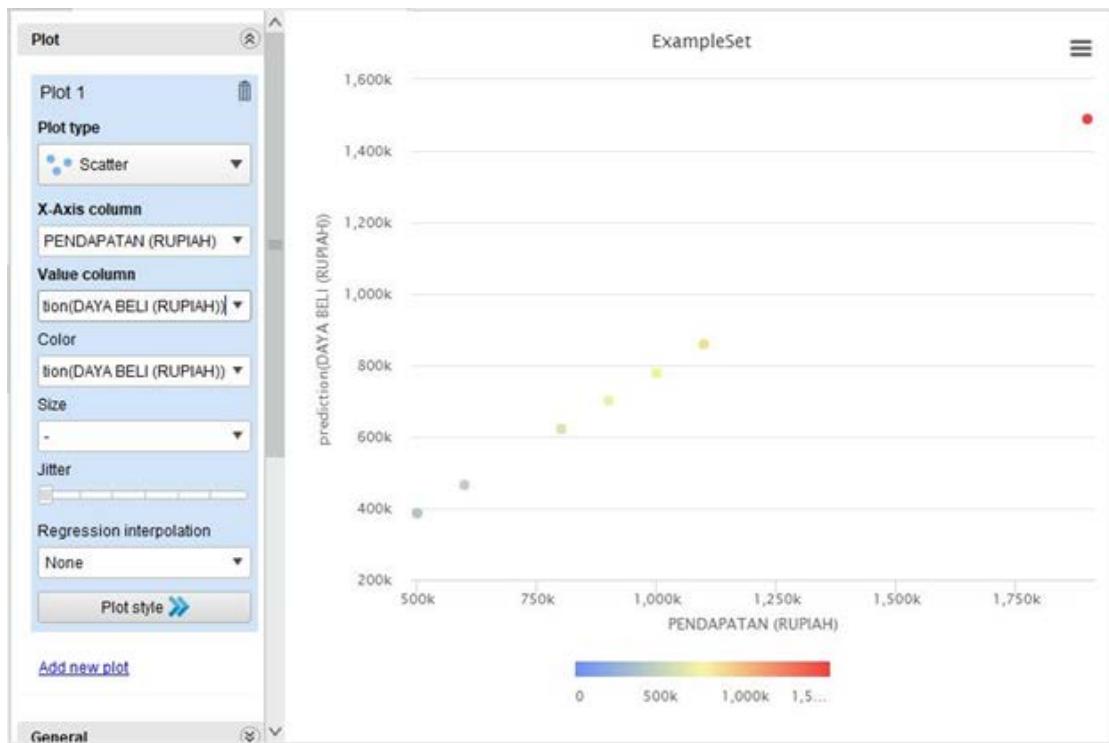
Row No.	NO. RESPON...	prediction(D...)	PENDAPATA...
1	1	701500.958	900000
2	2	622726.676	800000
3	3	386403.830	500000
4	4	1489243.776	1900000
5	5	465178.112	600000
6	6	622726.676	800000
7	7	780275.239	1000000
8	8	859049.521	1100000
9	9	780275.239	1000000
10	10	386403.830	500000

Jumlah Anggota Keluarga ( $X_2$ )

Row No.	NO. RESPON...	prediction(D...)	JUMLAH AN...
1	1	857096.226	5
2	2	553718.868	3
3	3	402030.189	2
4	4	1008784.906	6
5	5	402030.189	2
6	6	857096.226	5
7	7	1008784.906	6
8	8	705407.547	4
9	9	705407.547	4
10	10	553718.868	3

6. Gambarkan pola sebaran data menggunakan Plot View (Scatter) dengan ketentuan berikut :

- a) x-Axis = Pendapatan (Rupiah),  
y-Axis = Prediction (Daya Beli (Rupiah)),  
Color Column = Prediction (Daya Beli (Rupiah))



- b) x-Axis = Jumlah Anggota Keluarga,  
y-Axis = Prediction (Daya Beli (Rupiah)),  
Color Column = Prediction (Daya Beli (Rupiah))

