

Nama : Muhammad Huda Fatoni

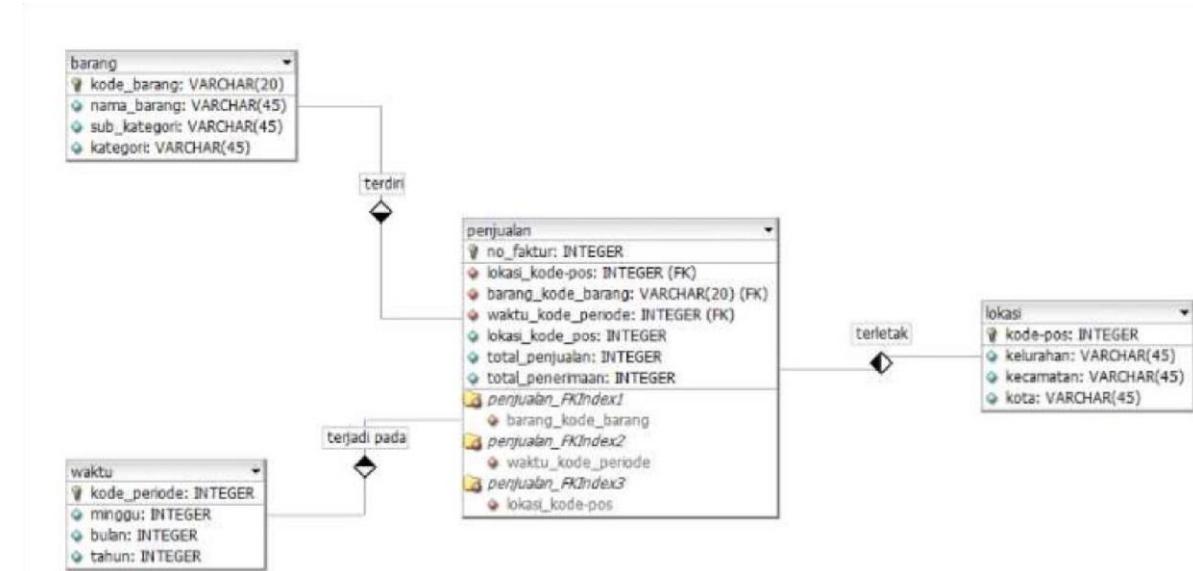
NIM : L200170062

Kelas : C

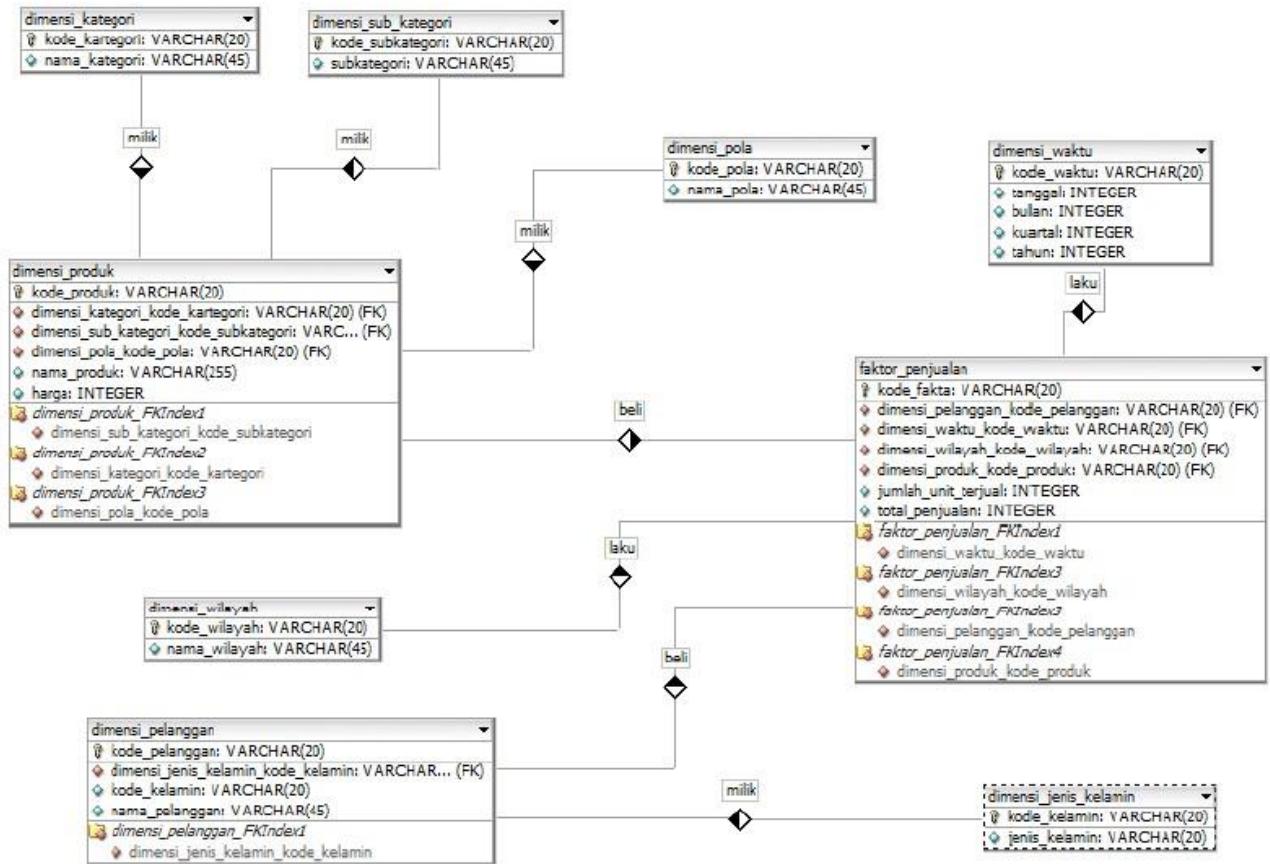
**LAPORAN PRAKTIKUM**  
**DATA WAREHOUSING & DATA MINING**

**MODUL 1**

**Latihan**



## Tugas



## Latihan

### Kegiatan 1

#### Fact Table

The screenshot shows a Microsoft Excel spreadsheet titled "Sheet1". The data is presented in a table with 21 rows and 13 columns. The columns are labeled A through M. The first few rows of data are as follows:

bulan	kuartal	tahun	nama_produk	nama_kategori	nama_subkategori	nama_pola	nama_pelanggan	jenis_kelamin	nama_wilayah	jumlah	harga
2	12	4	2011 Jarik Standar	Standar	Jarik	Print	Bapak Ketut	PRIA	Bali	2	225000
3	1	1	2012 Kaos Batik	Batik	Kaos	Cap	Ibu Harini	WANITA	Jawa Timur	14	30000
4	4	2	2012 Jarik Standar	Standar	Jarik	Tulis	Ibu Harini	WANITA	Jawa Timur	4	40000
5	4	2	2011 Helm Katun	Katun	Helm	Print	Ibu Harini	WANITA	Jawa Timur	3	70000
6	9	3	2012 Batik Standar	Standar	Batik	Cap	Bapak Heru	PRIA	Jawa Timur	1	150000
7	5	2	2012 Hem Katun	Katun	Hem	Print	Bapak Totok	PRIA	Jawa Timur	3	299000
8	12	4	2011 Bolero Standar	Standar	Bolero	Cap	Ibu Hatamah	WANITA	Jawa Timur	1	225000
9	10	4	2011 Sarimbit Standar	Standar	Sarimbit	Print	Ibu Hatamah	WANITA	Jawa Timur	1	150000
10	1	1	2011 Kaos Katun	Katun	Kaos	Print	Bapak Imron	PRIA	Jawa Barat	1	60000
11	2	1	2012 Celana Standar	Standar	Celana	Cap	Ibu Hadi Sukarni	WANITA	Jawa Barat	17	55000
12	3	1	2010 Celana Standar	Standar	Celana	Print	Ibu Hadi Sukarni	WANITA	Jawa Barat	17	55000
13	3	1	2011 Bahan Standar	Standar	Bahan	Cap	Ibu Siti Arya	WANITA	Jawa Barat	8	120000
14	12	4	2012 Rok Batik	Batik	Rok	Print	Ibu Siti Arya	WANITA	Jawa Barat	1	225000
15	1	1	2012 Jam Standar	Standar	Jam	Print	Ibu Siti Arya	WANITA	Jawa Barat	44	80000
16	9	3	2012 Hem Standar	Standar	Hem	Cap	Ibu Aini Kasmaji	WANITA	Jawa Tengah	1	100000
17	6	2	2012 Bahan Standar	Lawasan	Bahan	Tulis	Ibu Niken	WANITA	Jawa Tengah	1	130000
18	8	3	2011 Hem Standar	Standar	Hem	Tulis	Ibu Atik	WANITA	Jawa Tengah	5	550000
19	4	2	2012 Bahan Standar	Standar	Bahan	Cap	Ibu Tyas	WANITA	Jawa Tengah	7	135000
20	6	2	2010 Bahan Beludru	Beludru	Bahan	Cap	Ibu Tyas	WANITA	Jawa Tengah	1	500000
21	11	4	2010 Hem Sutra	Sutra	Hem	Print	Ibu Tyas	WANITA	Jawa Tengah	5	100000

Membuat pivot table

PivotTable Fields ribbon tab selected.

Choose fields to add to report:

- bulan
- kuartal
- tahun
- nama\_produk
- nama\_kategori

Drag fields between areas below:

Filters	Columns
Rows	Σ Values

Defer Layout Update      Update

Menyusun layout field dengan grouping field (subkategori, tahun, jumlah)

PivotTable Fields ribbon tab selected.

Choose fields to add to report:

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

More Tables...

Drag fields between areas below:

Filters	Columns
tahun	
Rows	Σ Values
nama_kategori	Sum of jumlah

Defer Layout Update      Update



## Kegiatan 2

Menambah tipe summary baru (sum of jumlah 2)

A3	B	C	D	E	F	G	H	I	J
1									
2									
3	Column Labels								
4	2010	2011	2012		Total Sum of jumlah	Total Sum of jumlah2			
5	Row Labels	Sum of jumlah	Sum of jumlah2	Sum of jumlah	Sum of jumlah2	Sum of jumlah	Sum of jumlah2		
6	Batik				15	15	15	15	
7	Beludru	1	1				1	1	
8	Katun			4	4	3	3	7	7
9	Lawasan					1	1	1	1
10	Standar	17	17	17	17	74	74	108	108
11	Sutra	5	5					5	5
12	Grand Total	23	23	21	21	93	93	137	137
13									
14									
15									
16									

Pada dialog Value Field Settings, mengubah Sum menjadi Count

A3	B	C	D	E	F	G	H	I	J
1									
2									
3	Column Labels								
4	2010	2011	2012		Total Sum of jumlah	Total Count of jumlah2			
5	Row Labels	Sum of jumlah	Count of jumlah2	Sum of jumlah	Count of jumlah2	Sum of jumlah	Count of jumlah2		
6	Batik				15	2	15	2	
7	Beludru	1	1				1	1	
8	Katun			4	2	3	1	7	3
9	Lawasan					1	1	1	1
10	Standar	17	1	17	5	74	6	108	12
11	Sutra	5	1					5	1
12	Grand Total	23	3	21	7	93	10	137	20
13									
14									
15									
16									

Insert calculated field

### Kegiatan 3

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable. The PivotTable Fields pane is open, showing 'Choose fields to add to report'. The 'Values' section lists 'Sum of jumlah' and 'Count of jumlah2'. The 'Columns' section lists 'Tahun' and 'Values'. The 'Fields' section lists various categories like Batik, Beludru, Katun, etc., and years 2010 and 2011. A 'Grand Total' row is also present. A 'Sheet2' tab is selected at the bottom.

Field baru “Sum of Pendapatan” muncul pada pivot table

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4									
5									
6	Batik				0		0		
7	Beludru	1	1	500000		0	15	2	
8	Katun			0	4	2	520000	1	
9	Lawasan			0			0	1	
10	Standar	17	1	935000	17	5	21590000	6	
11	Sutra	5	1	500000			0		
12	<b>Grand Total</b>	<b>23</b>	<b>3</b>	<b>15065000</b>	<b>21</b>	<b>7</b>	<b>29400000</b>	<b>93</b>	<b>10</b>
13									
14									
15									
16									

Melakukan operasi roll up dan drill down (nama\_kategori, nama\_subkategori, nama\_produk, dan tahun)

## Kegiatan 4

Sum of Pendapatan		Column Labels										
		⊕ 2010	⊕ 2011	⊖ 2012	⊕ 1	⊖ 2	2 Total		⊕ 3	⊕ 4	2012 Total	Grand Total
Row Labels					4	5	6					
⊕ Batik		0	0	420000	0	0	0	0	0	225000	3825000	3825000
⊕ Beludru		500000	0	0	0	0	0	0	0	0	0	500000
⊖ Katun		0	520000	0	0	897000	0	897000	0	0	897000	3003000
⊕ Helm		0	210000	0	0	0	0	0	0	0	0	210000
⊕ Hem		0	0	0	0	897000	0	897000	0	0	897000	897000
⊕ Kaos		0	60000	0	0	0	0	0	0	0	0	60000
⊕ Lawasan		0	0	0	0	0	130000	130000	0	0	130000	130000
⊕ Standar		935000	21590000	8235000	1925000	0	0	1925000	500000	0	41440000	203580000
⊕ Sutra		500000	0	0	0	0	0	0	0	0	0	500000
Grand Total		15065000	29400000	12375000	1925000	897000	130000	9060000	500000	225000	115692000	451963000

## Tugas

### Tugas 1

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

	A	B	C	D	E	F	G
1							
2							
3	Column Labels						
4		2012			Total Sum of Pendapatan	Total Sum of PPN (10%)	Total Sum of Total Penghasilan
5	Row Labels	Sum of Pendapatan	Sum of PPN (10%)	Sum of Total Penghasilan			
6	Bahan	212000	21200	1908000	2120000	212000	1908000
7	Batik	150000	15000	135000	150000	15000	135000
8	Celana	935000	93500	841500	935000	93500	841500
9	Hem	1596000	159600	1436400	1596000	159600	1436400
10	Jam	3520000	352000	3168000	3520000	352000	3168000
11	Jarik	160000	16000	144000	160000	16000	144000
12	Kaos	420000	42000	378000	420000	42000	378000
13	Rok	225000	22500	202500	225000	22500	202500
14	Grand Total	115692000	11569200	104122800	115692000	11569200	104122800
15							
16							
17							

### Tugas 2

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

Hem memiliki total penghasilan sebesar 2.475.000

	2010			2011			2012	
	Sum of Pendapatan	Sum of PPN (10%)	Sum of Total Penghasilan	Sum of Pendapatan	Sum of PPN (10%)	Sum of Total Penghasilan	Sum of Pendapatan	Sum of PPN (10%)
	500000	50000	450000	960000	96000	864000	2120000	212000
	0	0	0	0	0	0	150000	15000
	0	0	0	225000	22500	202500	0	0
	935000	93500	841500	0	0	0	935000	93500
	0	0	0	210000	21000	189000	0	0
	500000	50000	450000	2750000	275000	2475000	1596000	159600
	0	0	0	0	0	0	3520000	352000
	0	0	0	450000	45000	405000	160000	16000
	0	0	0	60000	6000	54000	420000	42000
	0	0	0	0	0	0	225000	22500
	0	0	0	150000	15000	135000	0	0
	15065000	1506500	13558500	29400000	2940000	26460000	11569200	11569200

## MODUL 6

### Tugas

#### 1. Membuat table

Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten	Lama_Studi
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. Mencari menggunakan formula

a. Jumlah data kelas IPA : =COUNTIF(A2:A21,A3)

Jumlah data kelas IPS : =COUNTIF(A2:A21,A2)

Jumlah data kelas LAIN : =COUNTIF(A2:A21,A4)

b. Jumlah kelas Tepat : =COUNTIF(F2:F21,F3)

Jumlah kelas Terlambat : =COUNTIF(F2:F21,F2)

c. Atribut Rerata\_SKS

Nilai Max : =MIN(D2:D21)

Nilai Min : =MAX(D2:D21)

Nilai Mean : =AVERAGE(D2:D21)

Nilai Standar Deviation : =STDEV.S(D2:D21)

d. Data gabungan Jurusan\_SMA (IPA), Gender (PRIA), Asisten (YA),

Lama\_Studi (TEPAT)

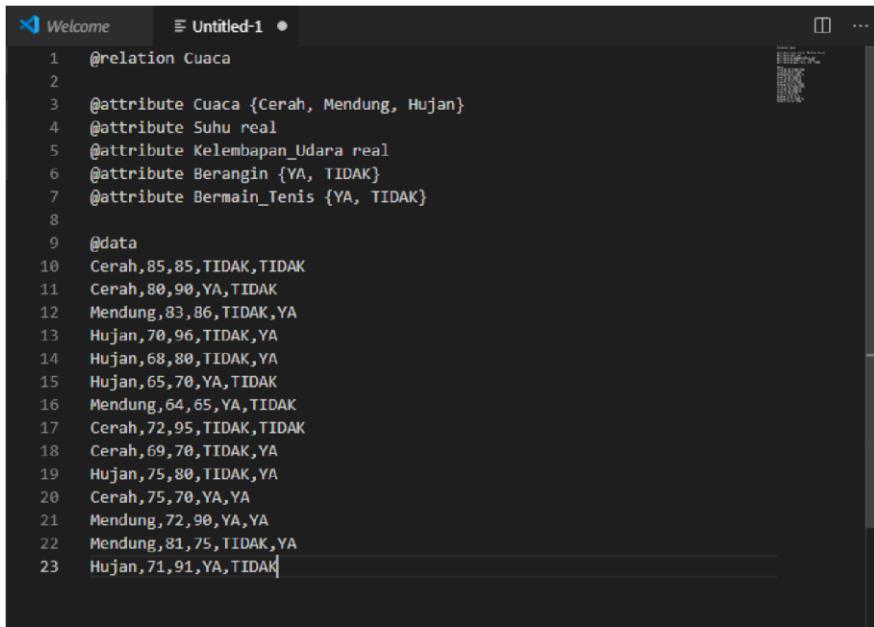
=COUNTIFS(A2:A21,A3,B2:B21,B3,E2:E21,E3,F2:F21,F3)

Jumlah Data IPA	Jumlah Data IPS	Jumlah Data LAIN		
10	6	4		
Lama_Studi (TEPAT)	Lama_Studi (TERLAMBAT)			
13	7			
ATRIBUT	MIN	MAX	MEAN	STANDAR DEVIATION
Rerata_SKS	16	23	18.95	1.66938375
IPA, PRIA, ASISTEN, TEPAT				
3				

## MODUL 7

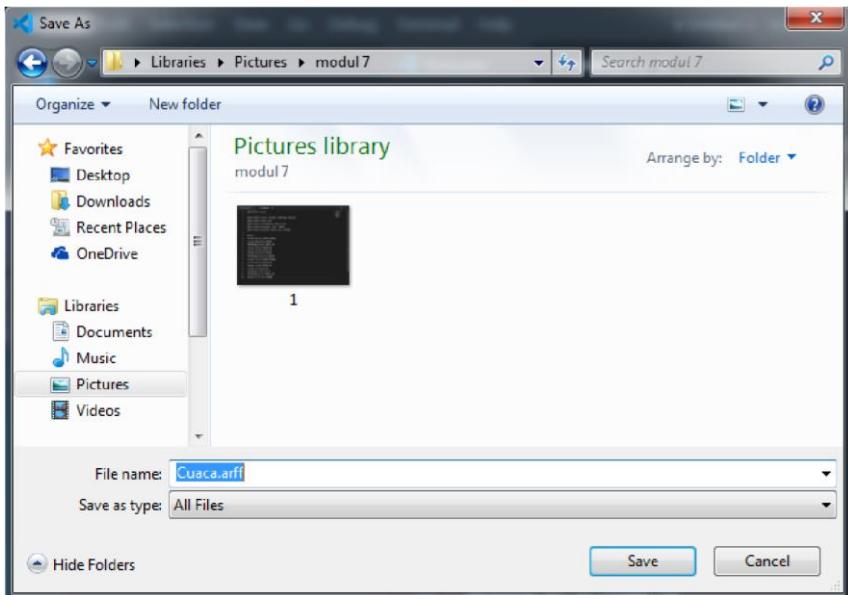
### Latihan

Mengetik nilai data sesuai di modul

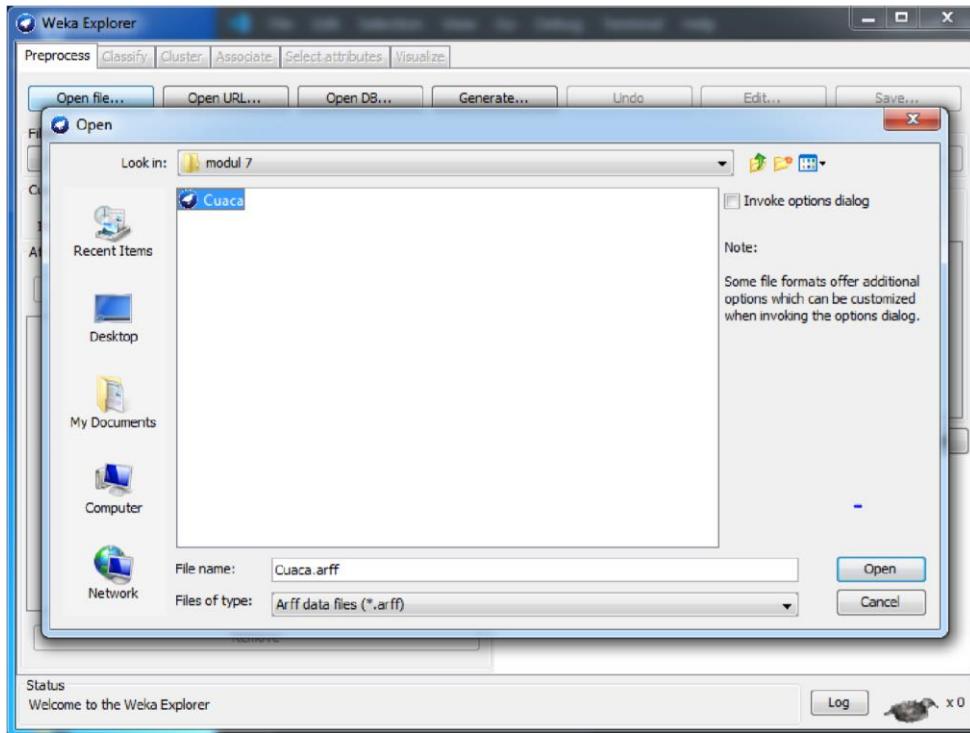


```
1 @relation Cuaca
2
3 @attribute Cuaca {Cerah, Mendung, Hujan}
4 @attribute Suhu real
5 @attribute Kelembapan_Udara real
6 @attribute Berangin {YA, TIDAK}
7 @attribute Bermain_Tenis {YA, TIDAK}
8
9 @data
10 Cerah,85,85,TIDAK,TIDAK
11 Cerah,80,90,YA,TIDAK
12 Mendung,83,86,TIDAK,YA
13 Hujan,70,96,TIDAK,YA
14 Hujan,68,80,TIDAK,YA
15 Hujan,65,70,YA,TIDAK
16 Mendung,64,65,YA,TIDAK
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
```

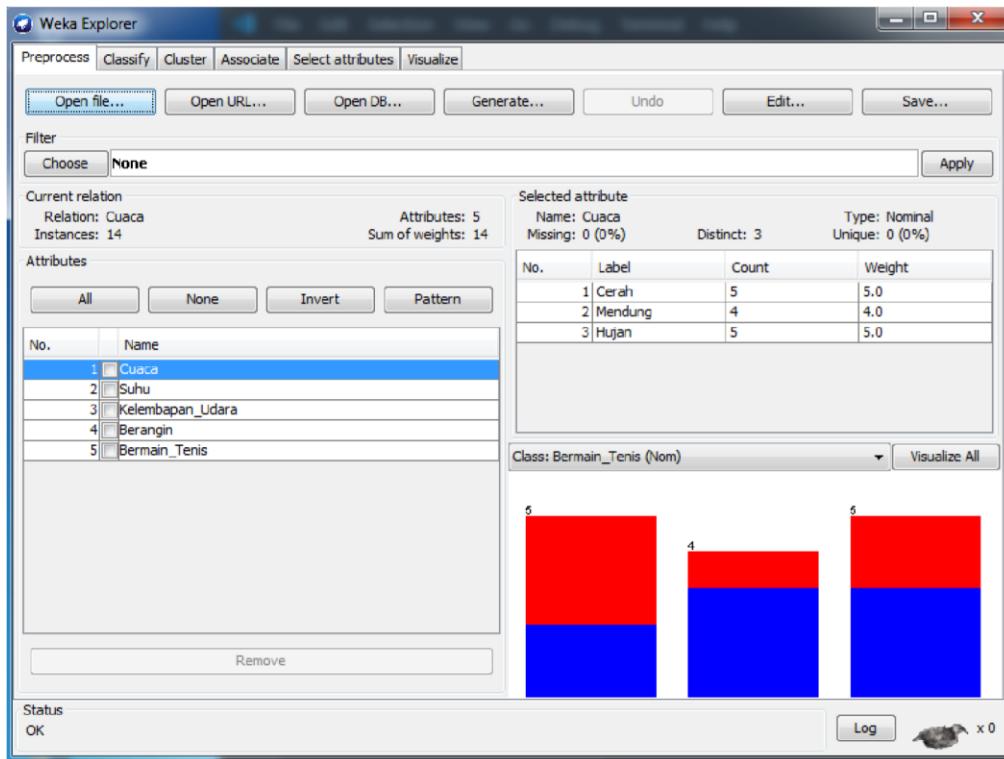
Simpan file dengan nama Cuaca.arff

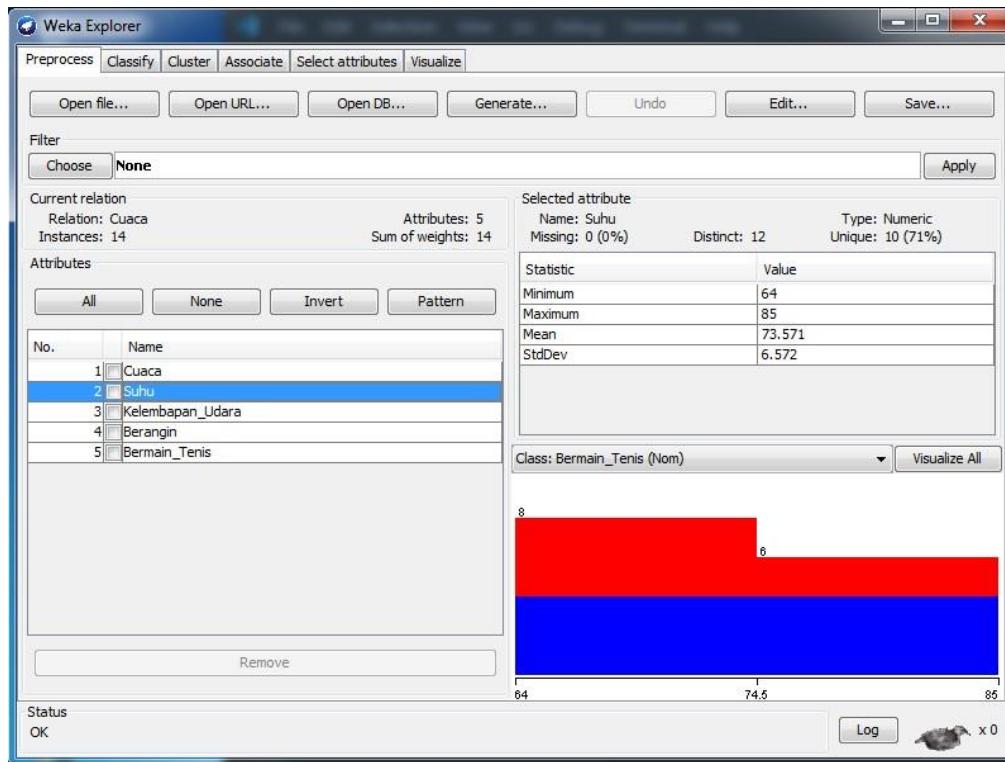


Membuka file Cuaca.arff dengan WEKA



Dalam Weka Explorer akan tampak grafik statistik masing-masing atribut pada table Cuaca



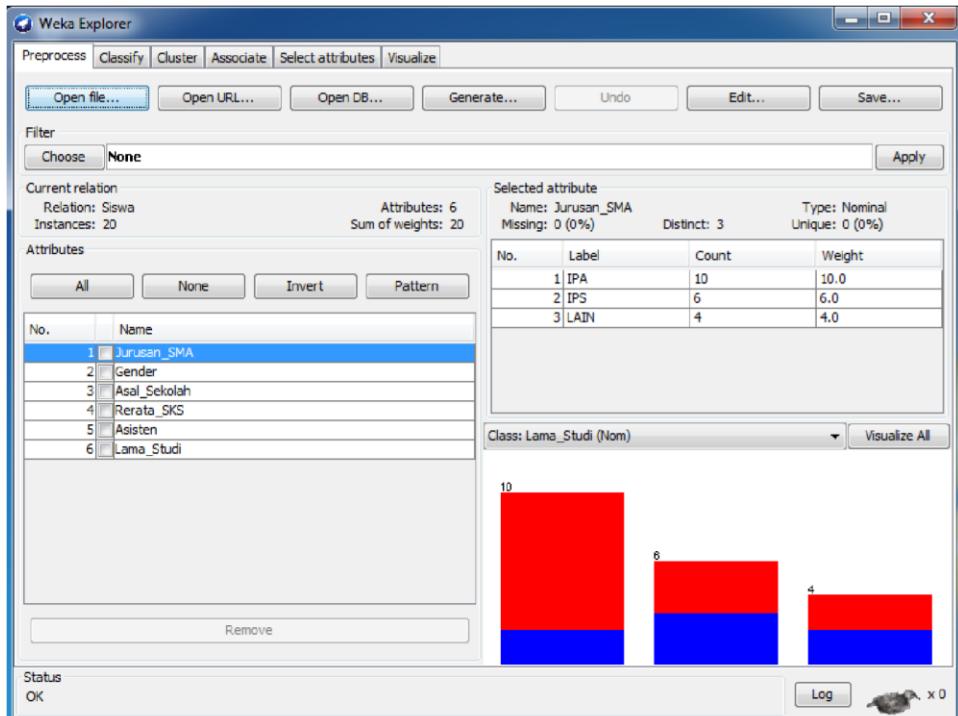


1. Buat file arff berdasarkan tugas modul 6 soal nomor 1

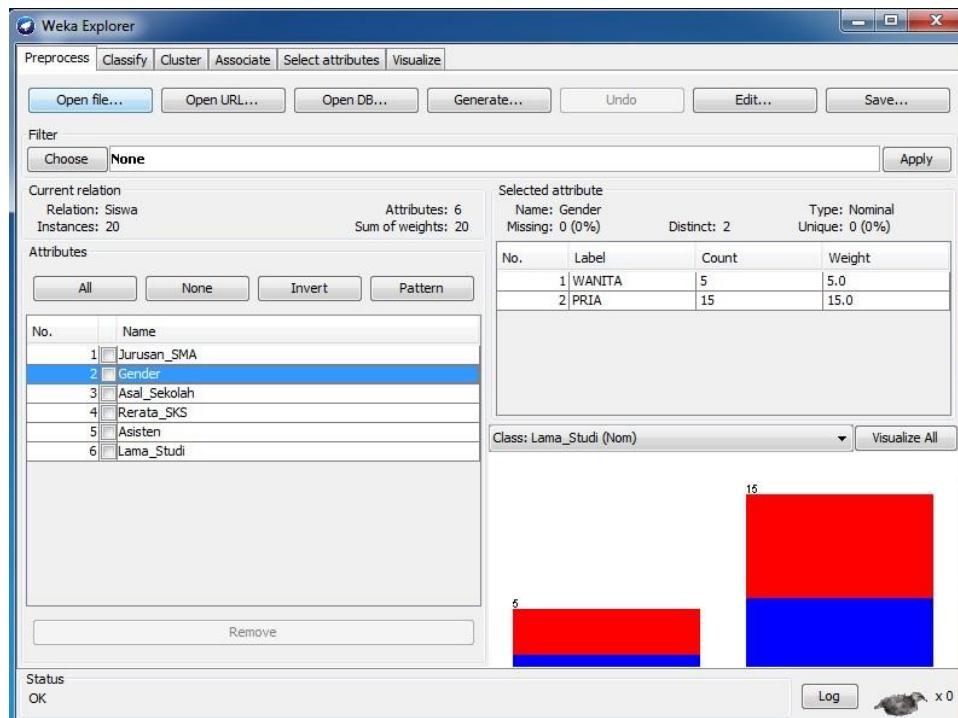
```
 1  @relation Siswa
 2
 3  @attribute Jurusan_SMA {IPA, IPS, LAIN}
 4  @attribute Gender {WANITA, PRIA}
 5  @attribute Asal_Sekolah {SURAKARTA, LUAR}
 6  @attribute Rerata_SKS real
 7  @attribute Asisten {TIDAK, YA}
 8  @attribute Lama_Studi {TERLAMBAT, TEPAT}
 9
10 @data
11 IPS,WANITA,SURAKARTA,18,TIDAK,TERLAMBAT
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
```

2. Gambar grafik setiap data

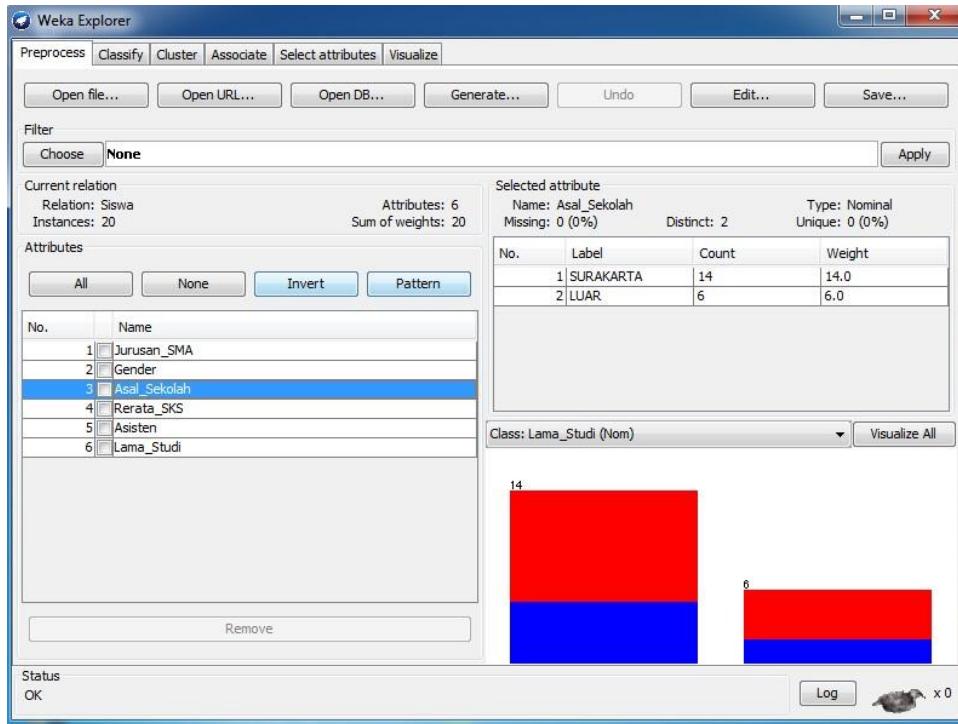
Jurusan\_SMA



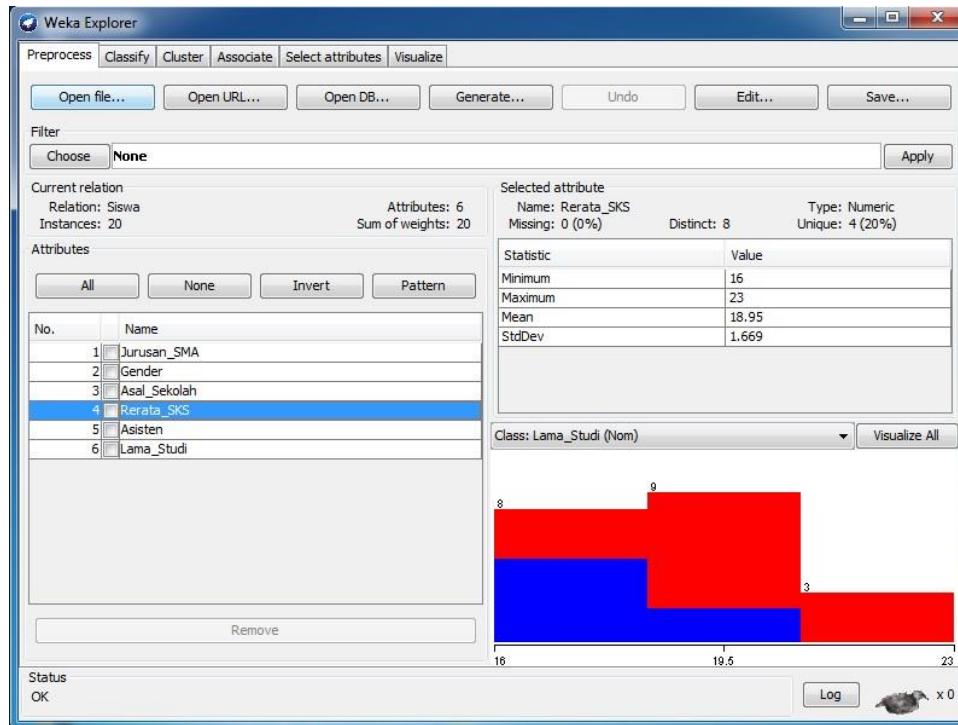
## Gender



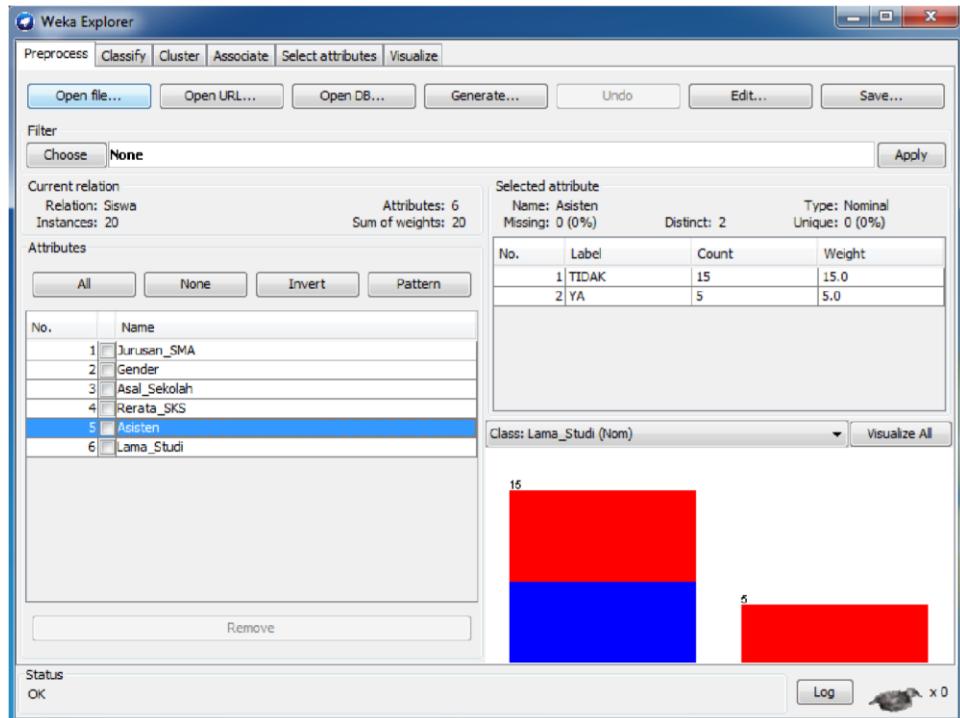
## Asal\_Sekolah



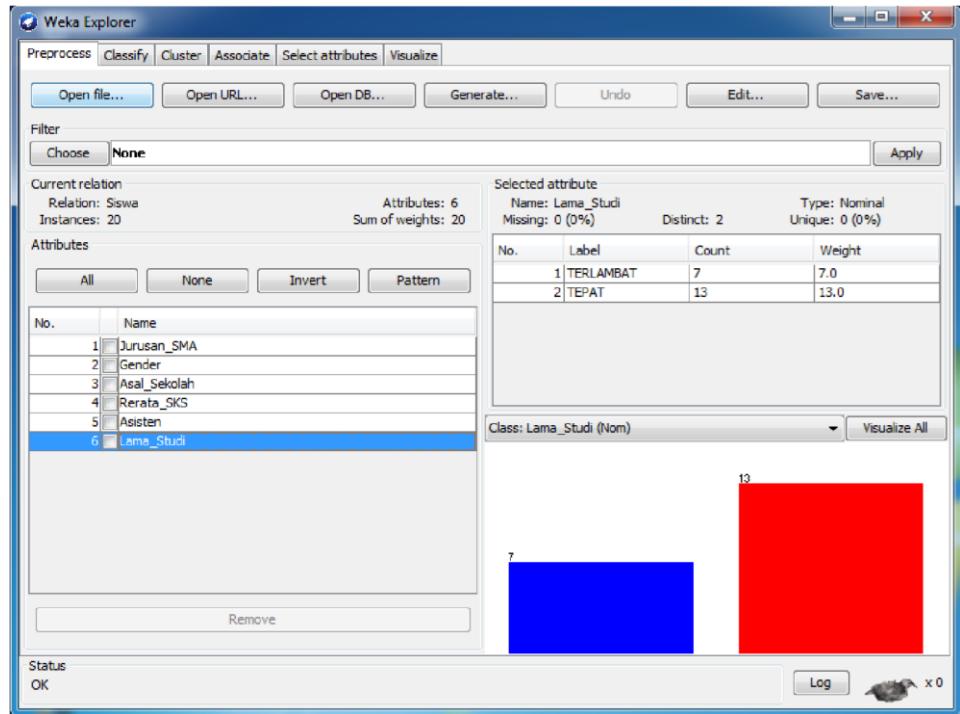
Rerata\_SKS



Asisten



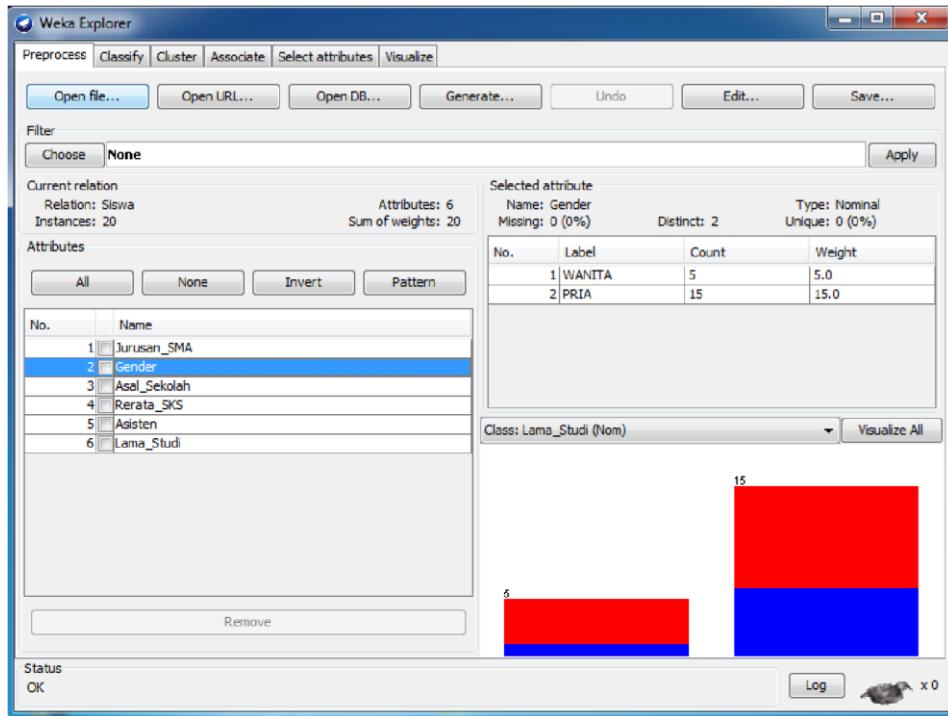
Lama\_Studi



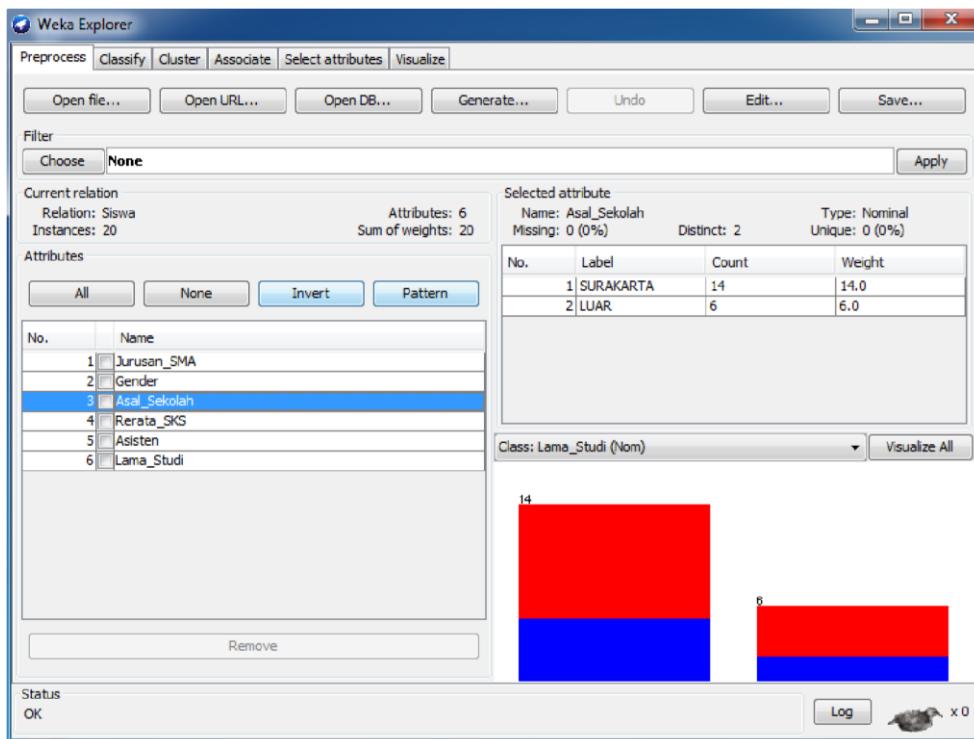
3. —

4. Jumlah atribut yang bertipe binomial dan polynomial **Binomial**

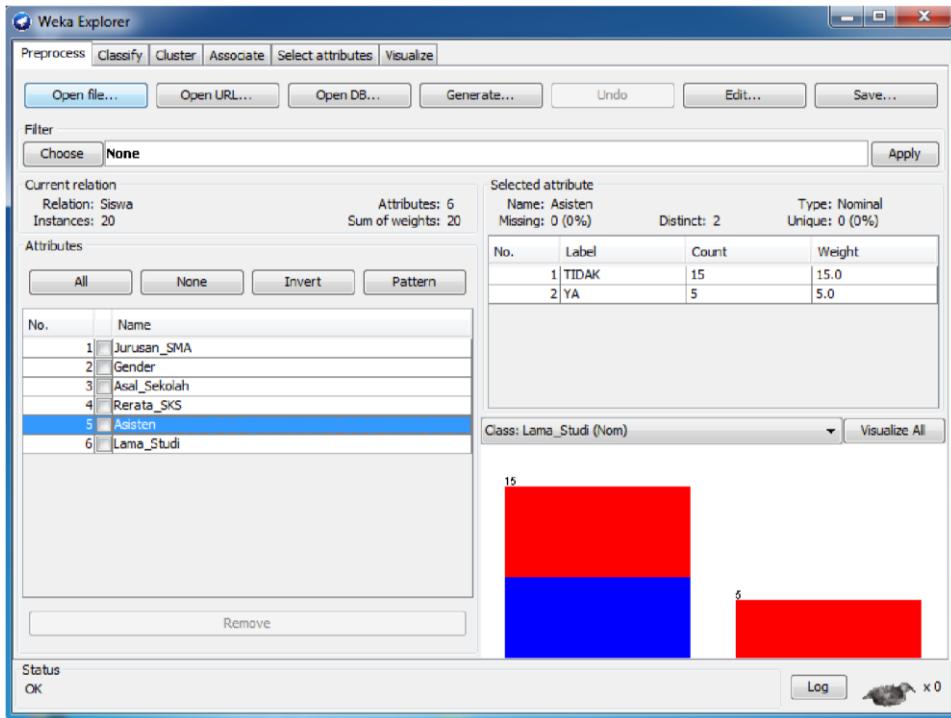
Gender



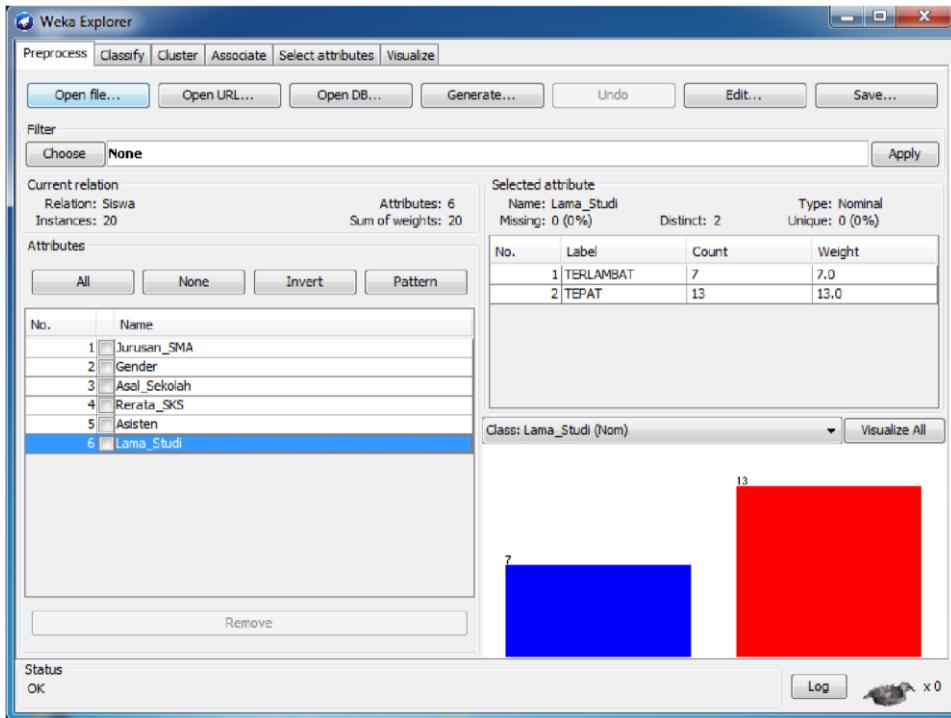
Asal\_Sekolah



Asisten

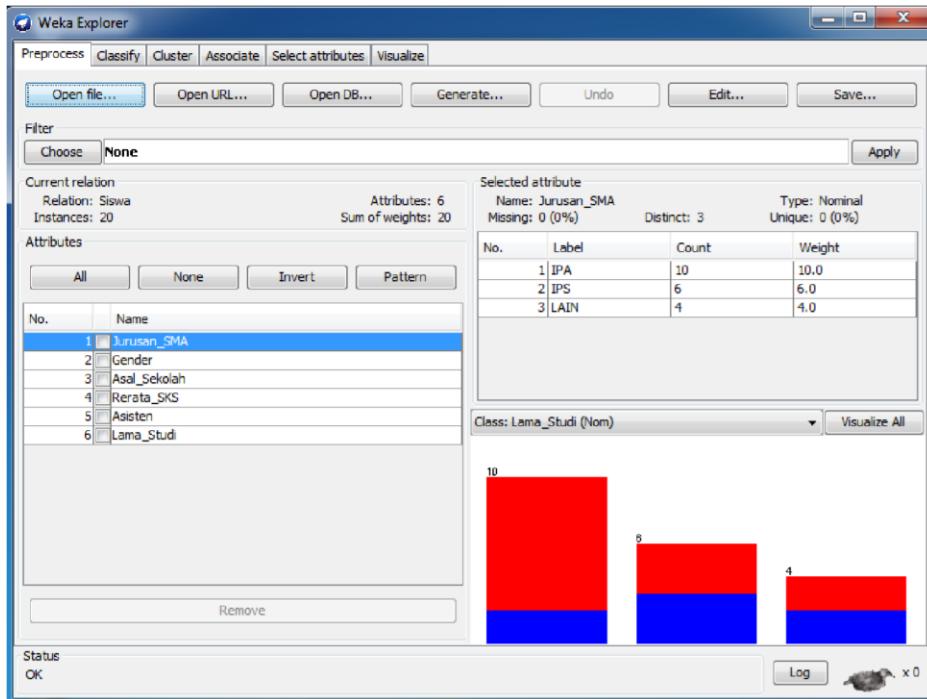


Lama\_Studi



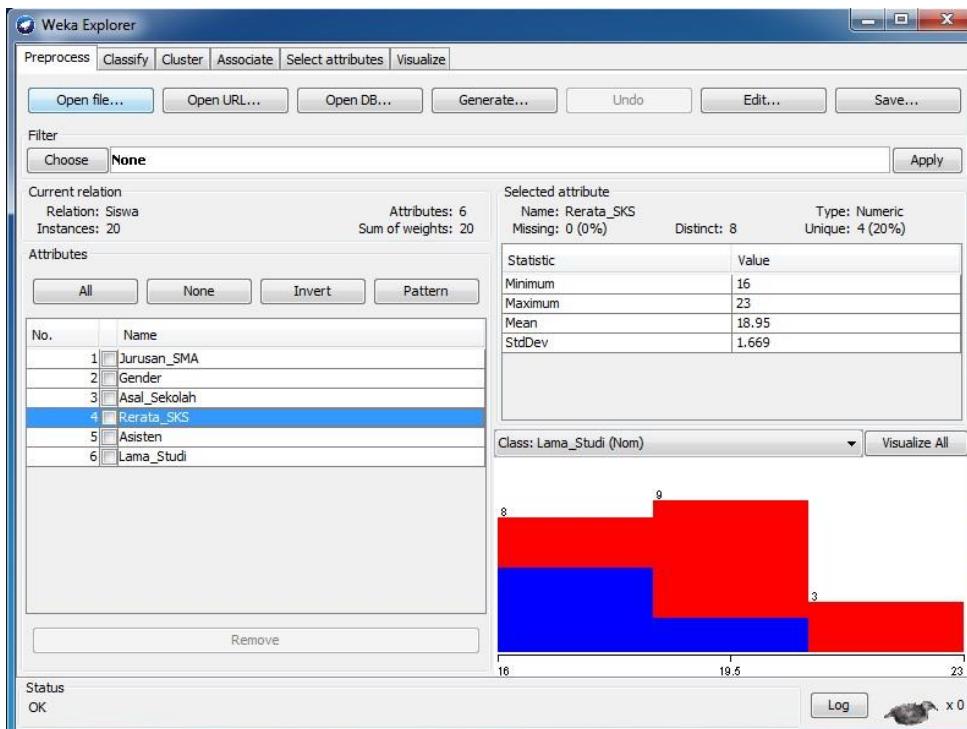
## Polynomial

Jurusian\_SMA



## 5. Atribut yang bertipe real

### Rerata\_SKS



## 6. Pada attribute Rerata\_SKS berapakah besarnya nilai Maximum, Minimun, Mean dan StdDev (Standard Deviation) Besar nilai :

- Max = 16
- Min = 23
- Mean = 18,95
- StdDev = 1,669

## MODUL 8

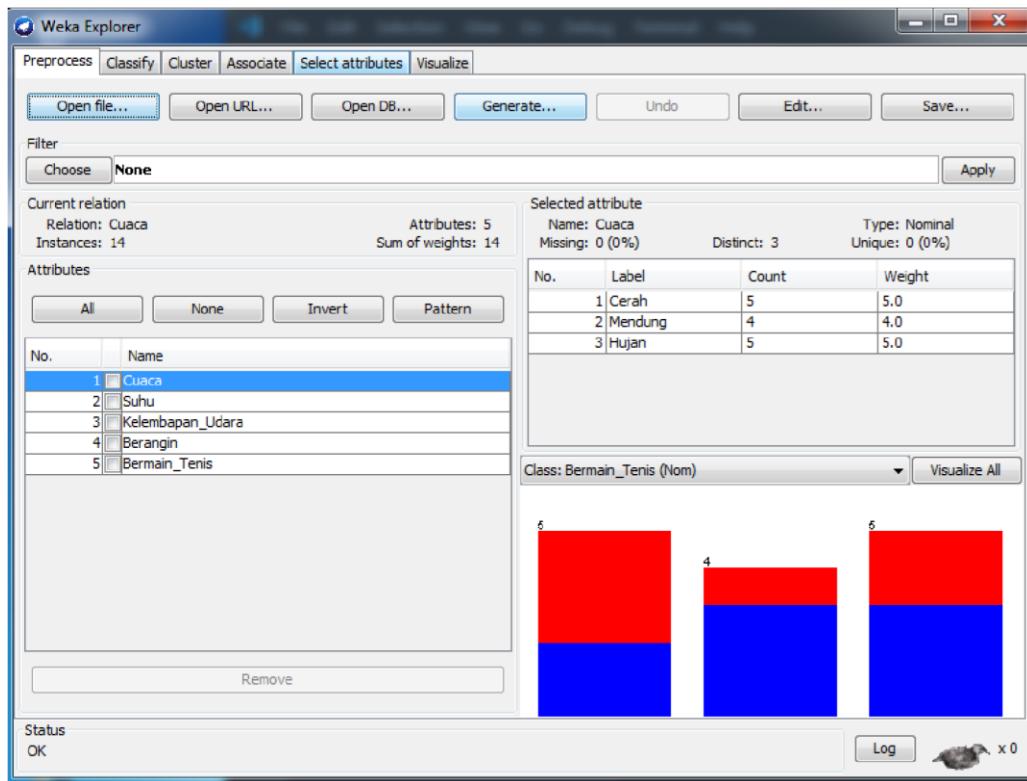
### Latihan

#### Implementasi Naïve Bayes dengan Weka

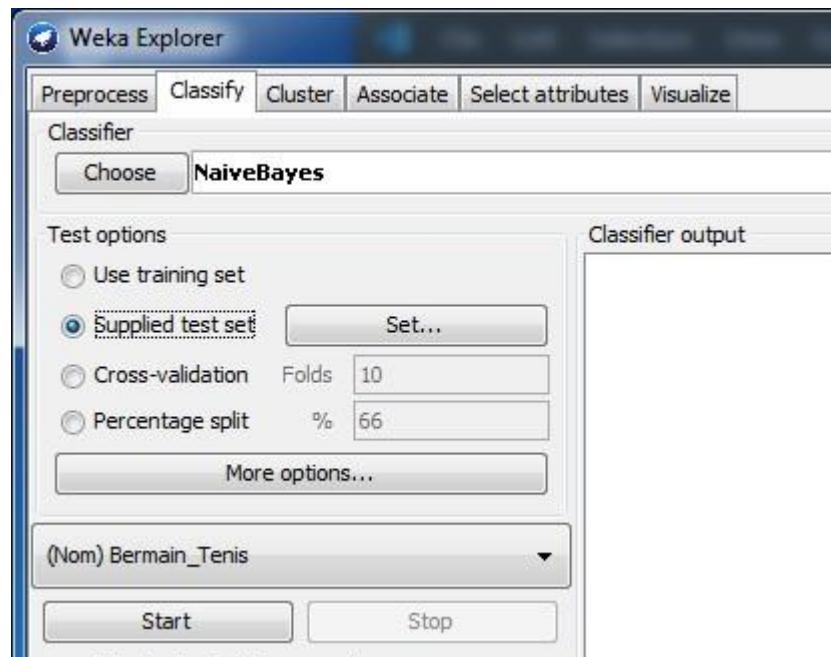
1. Persiapkan file Cuaca.arff dari hasil percobaan modul 7 (sebagai data training).
2. Membuat data testing.
3. Simpan dengan nama CuacaTesting.arff.

```
Welcome CuacaTesting.arff
C: > Users > LABSI-16 > Documents > CuacaTesting.arff
1 @relation CuacaTesting
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,?
```

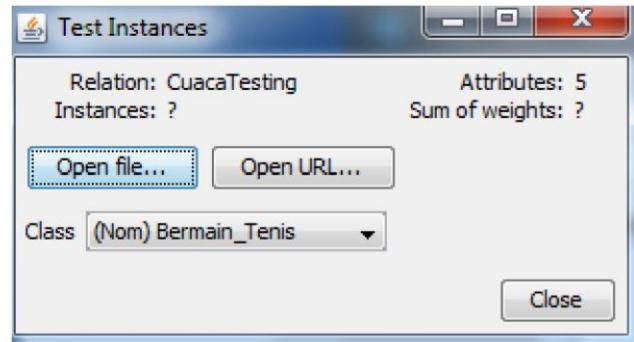
4. Membuka aplikasi WEKA
5. Buka file Cuaca.arff



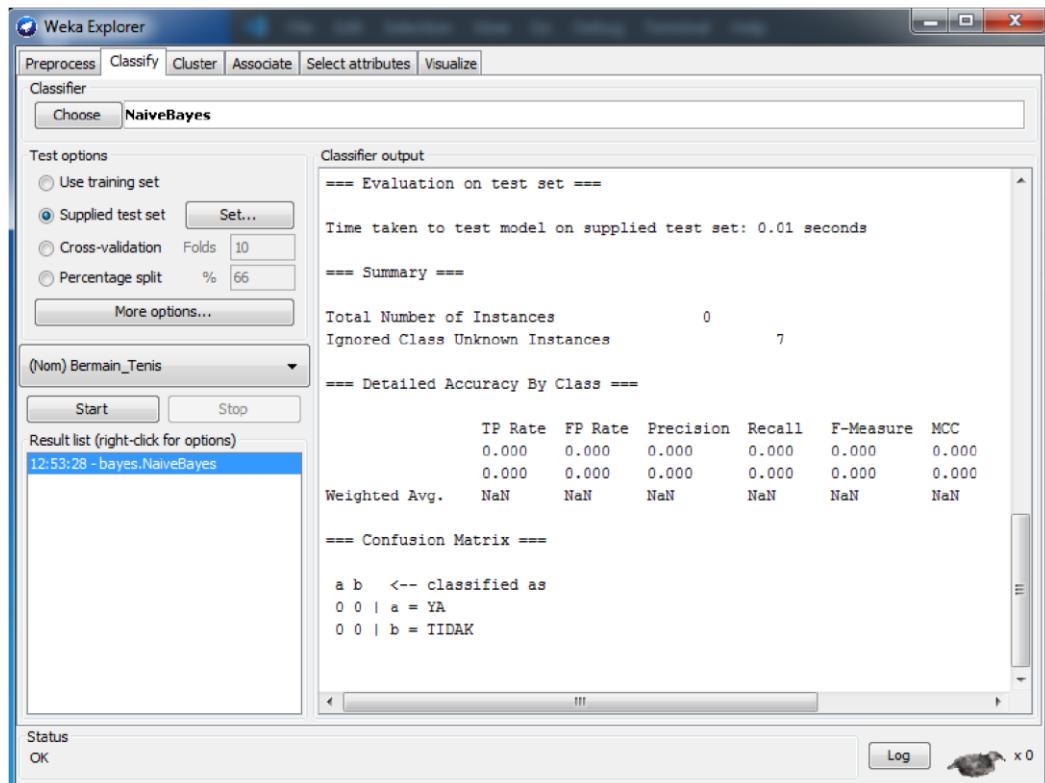
6. Weka Explorer, pilih tab classify, choose naïve bayes



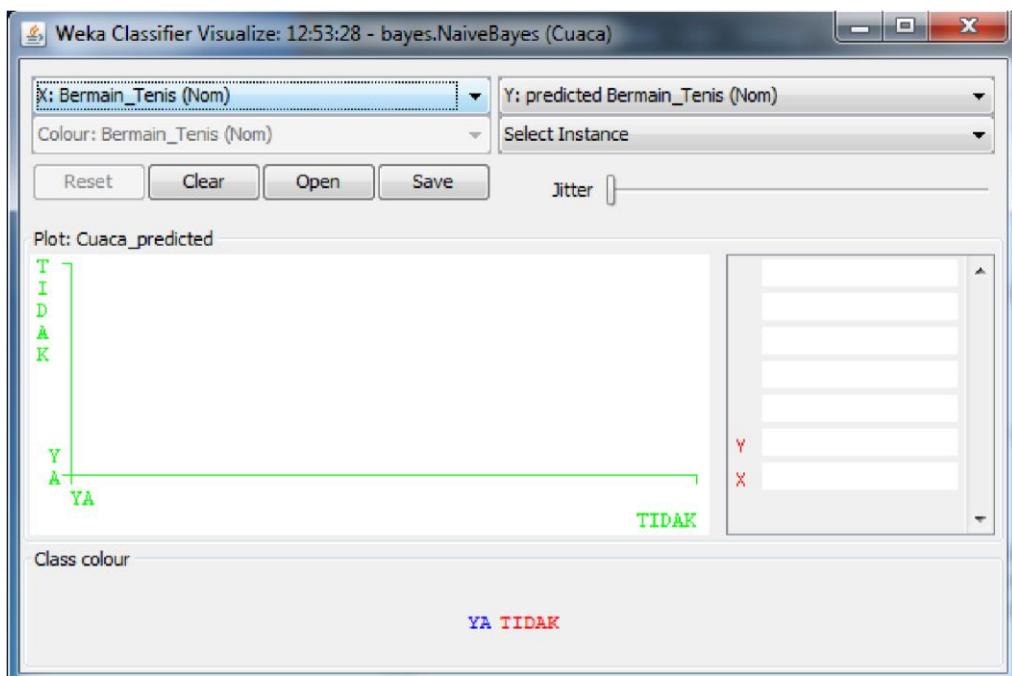
7. Supplied test set, SET, open file CuacaTesting.arff.



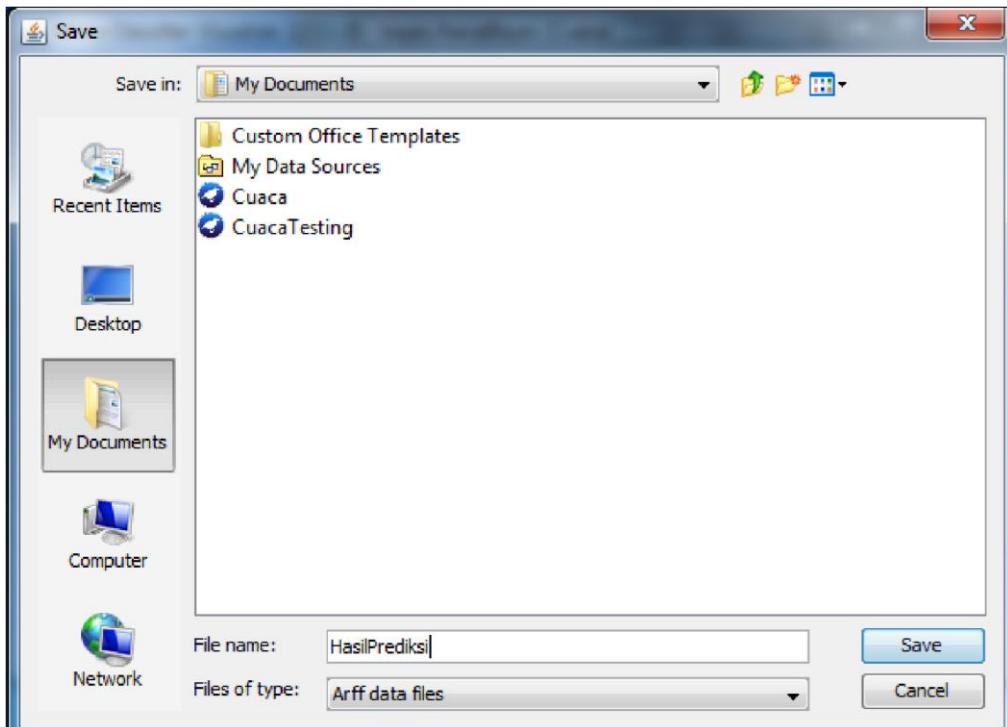
8. Klik Start untuk memulai proses naïve bayes.



9. Klik kanan pada hasil proses dalam kotak result list. Pilih menu Visualize classifier errors.



10. Klik save, simpan dengan nama file HasilPrediksi.arff.



11. Kembali ke Weka GUI Chooser, pilih menu Tools – ArffViewer, buka menu file – open. Tunjukkan pada file HasilPrediksi.arff.

ARFF-Viewer - C:\Users\LABSI-16\Documents\modul 8\HasilPrediksi.arff

Relation: Cuaca_predicted						
No.	1: Cuaca Nominal	2: Suhu Numeric	3: Kelembaban_Udara Numeric	4: Berangin Nominal	5: prediction margin Numeric	6: predicted 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 Rapid Miner

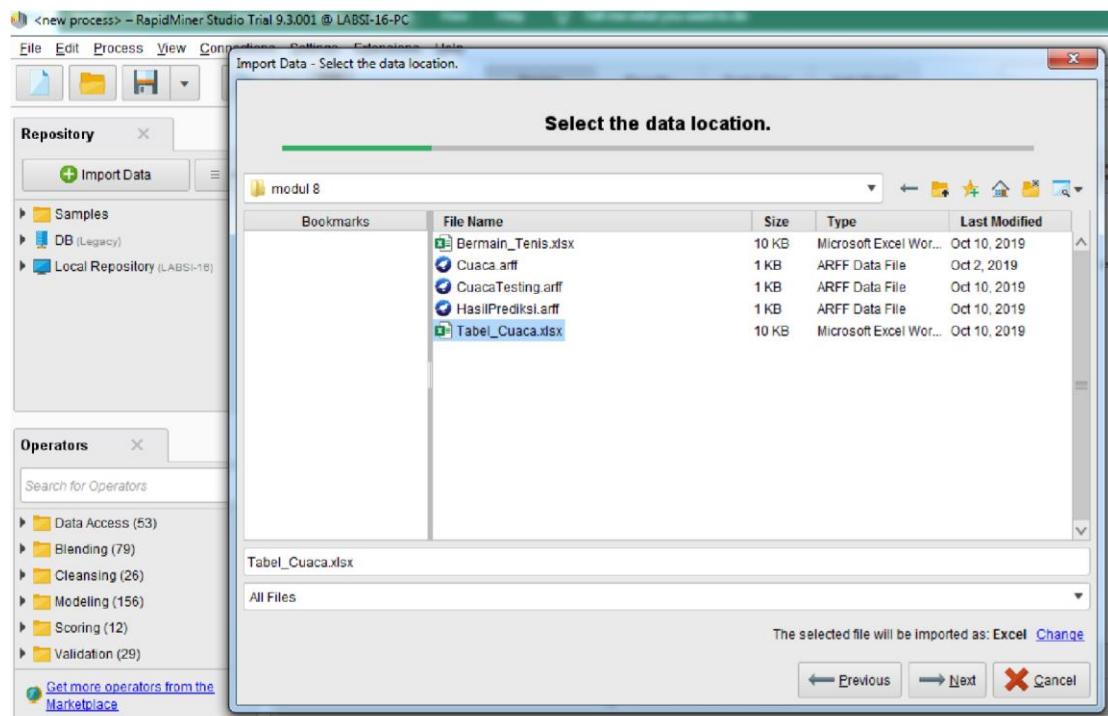
- Persiapkan file Tabel\_Cuaca.xls yang terdiri dari 2 sheet. Tabel data training

	A	B	C	D	E
1	Cuaca	Suhu	Kelembaban_udara	Berangin	Bermain_Tenis
2	Cerah	85		85	TIDAK
3	Cerah	80		90	YA
4	Mendung	83		86	TIDAK
5	Hujan	70		96	TIDAK
6	Hujan	68		80	TIDAK
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

Tabel data uji (testing)

	A	B	C	D
1	Cuaca	Suhu	Kelembaban_udara	Berangin
2	Cerah	75		65
3	Cerah	80		68
4	Cerah	83		87
5	Mendung	70		96
6	Mendung	68		81
7	Hujan	65		75
8	Hujan	64		85

- Buka aplikasi Rapid Miner, klik import data (Tabel\_Cuaca.xlsx).



3. Pastikan sel exce sesuai di langkah select the cells to import.

Select the cells to import.				
Sheet:	Training	Cell range:	A:E	Select All
				<input checked="" type="checkbox"/> Define header row: 1
1	Cuaca	Suhu	Kelembaban_udara	Berangin
2	Cerah	85.000	85.000	TIDAK
3	Cerah	80.000	90.000	YA
4	Mendung	83.000	86.000	TIDAK
5	Hujan	70.000	96.000	TIDAK
6	Hujan	68.000	80.000	TIDAK
7	Hujan	65.000	70.000	YA
8	Mendung	64.000	65.000	YA
9	Cerah	72.000	95.000	TIDAK
10	Cerah	69.000	70.000	YA
11	Hujan	75.000	80.000	TIDAK
12	Cerah	75.000	70.000	YA
13	Mendung	72.000	90.000	YA
14	Mendung	81.000	75.000	TIDAK
				YA
				TIDAK

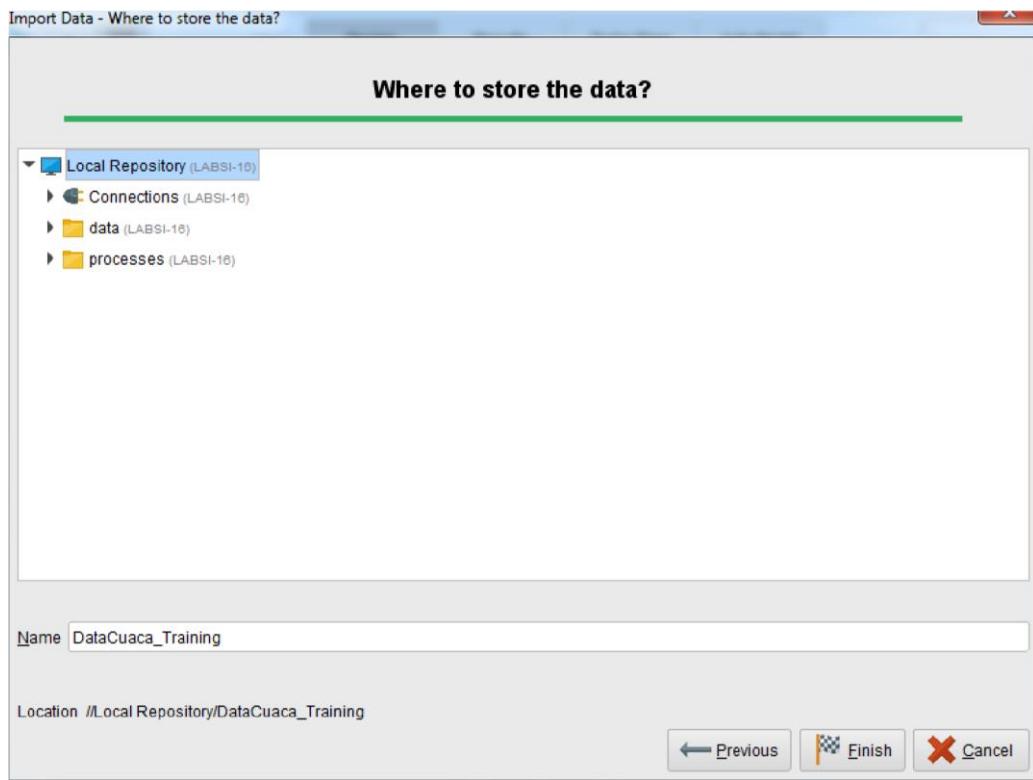
4. Pada langkah Format your coloums ubah kolom Bermain\_Tenis dengan tipe binomial.

Bermain_Tenis
<i>binominal</i>
TIDAK
TIDAK
YA
YA

5. Ubah pula sebagai label pada Change Role.

label
TIDAK
TIDAK
YA
YA
YA
TIDAK
YA

6. Simpan dengan nama DataCuaca\_Training, lalu finish.



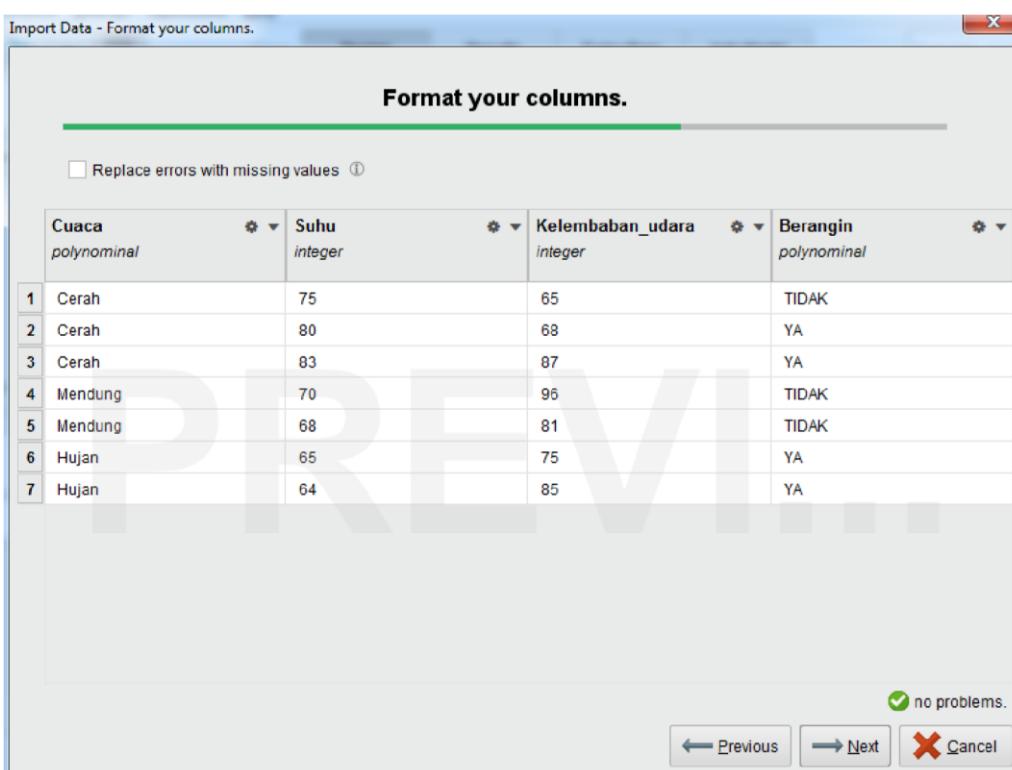
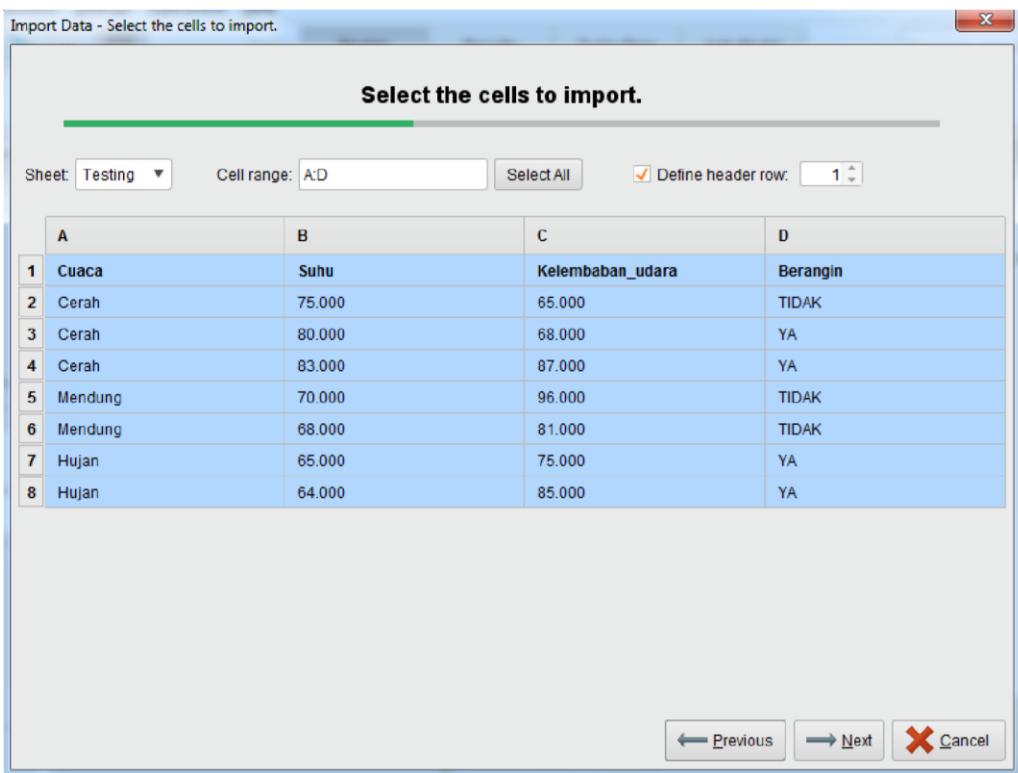
7. Hasil impoert file Tabel\_Cuaca.xls pada sheet1 akan ditampilkan.

Row No.	Bermain_Te...	Cuaca	Suhu	Kelembaban...	Berangin
1	TIDAK	Cerah	85	85	TIDAK
2	TIDAK	Cerah	80	90	YA
3	YA	Mendung	83	86	TIDAK
4	YA	Hujan	70	96	TIDAK
5	YA	Hujan	68	80	TIDAK
6	TIDAK	Hujan	65	70	YA
7	YA	Mendung	64	65	YA
8	TIDAK	Cerah	72	95	TIDAK
9	YA	Cerah	69	70	TIDAK
10	YA	Hujan	75	80	TIDAK
11	YA	Cerah	75	70	YA
12	YA	Mendung	72	90	YA
13	YA	Mendung	81	75	TIDAK
14					

ExampleSet (14 examples, 1 special attribute, 4 regular attributes)

8. Melakukan hal sama untuk data testing yang diambil dari Tabel\_Cuaca.xls pada sheet2 (Testing)

dengan mengulang langkah sebelumnya.

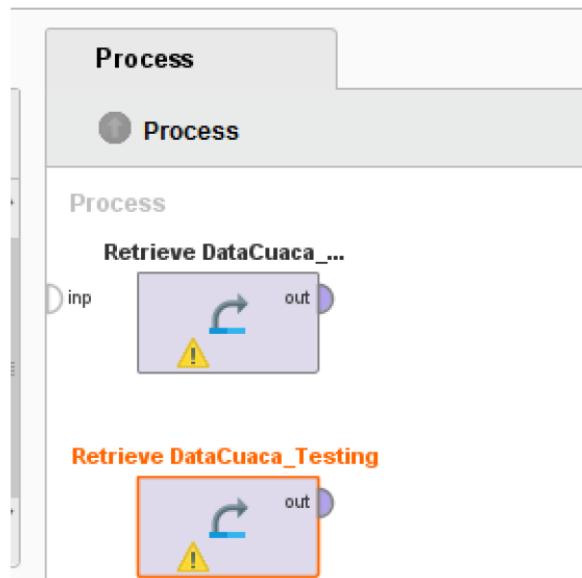


9. Simpan dengan nama DataCuaca\_Testing.

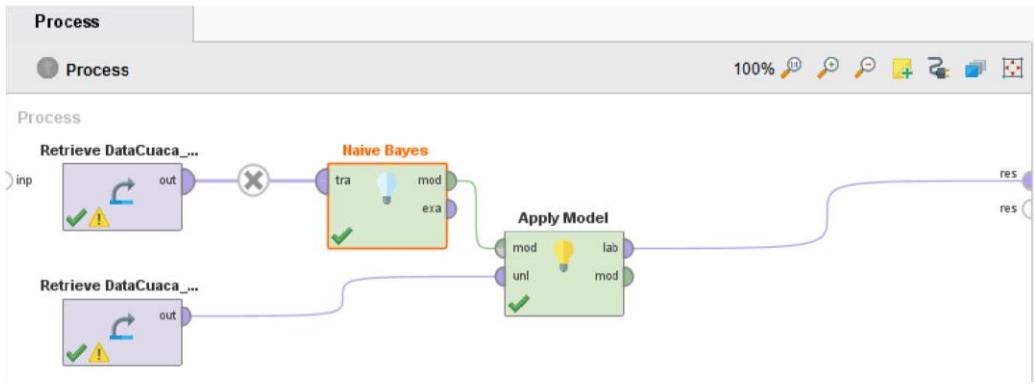
Open in [Turbo Prep](#) [Auto Model](#)

Row No.	Cuaca	Suhu	Kelembaban...	Berangin
1	Cerah	75	65	TIDAK
2	Cerah	80	68	YA
3	Cerah	83	87	YA
4	Mendung	70	96	TIDAK
5	Mendung	68	81	TIDAK
6	Hujan	65	75	YA
7	Hujan	64	85	YA

10. Membuat desain Naïve Bayes, drag DataCuaca\_Training dan DataCuaca\_Testing ke dalam jendela Process View.



11. Masukkan juga operator Naïve Bayes dan Apply Model.



12. Jalankan proses naïve bayes dan perhatikan hasil pada tab Data.

Row No.	prediction(Bermain_Tenis)	confidence(TIDAK)	confidence(YA)	Cuaca	Suhu	Kelembaban_udara	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

13. Pada tab Statistics, dapat dilihat bahwa hasil prediksi sama dengan prediksi menggunakan Weka.

Result History		ExampleSet (Apply Model)				
Name	Type	Missing	Statistics	Filter (7 / 7 attributes): Search for Attributes		
prediction(Bermain_Tenis)	Binomial	0	TIDAK (2) YA (5)	YA (5), TIDAK (2)		
Confidence_TIDAK confidence(TIDAK)	Real	0	Min: 0.007 Max: 0.856	Average: 0.353		
Confidence_YA confidence(YA)	Real	0	Min: 0.144 Max: 0.993	Average: 0.647		
Cuaca	Polynominal	0	Least: Mendung (2) Most: Cerah (3)	Values: Cerah (3), Hujan (2), ...[1]		
Suhu	Integer	0	Min: 64 Max: 83	Average: 72.143		
Kelembaban_udara	Integer	0	Min: 65 Max: 96	Average: 79.571		
Berangin	Polynominal	0	Least: TIDAK (3) Most: YA (4)	Values: YA (4), TIDAK (3)		

## 1. Membuat tabel seperti di modul

The screenshot shows a Microsoft Excel spreadsheet titled "DataTesting.arff" and its corresponding ARFF file content.

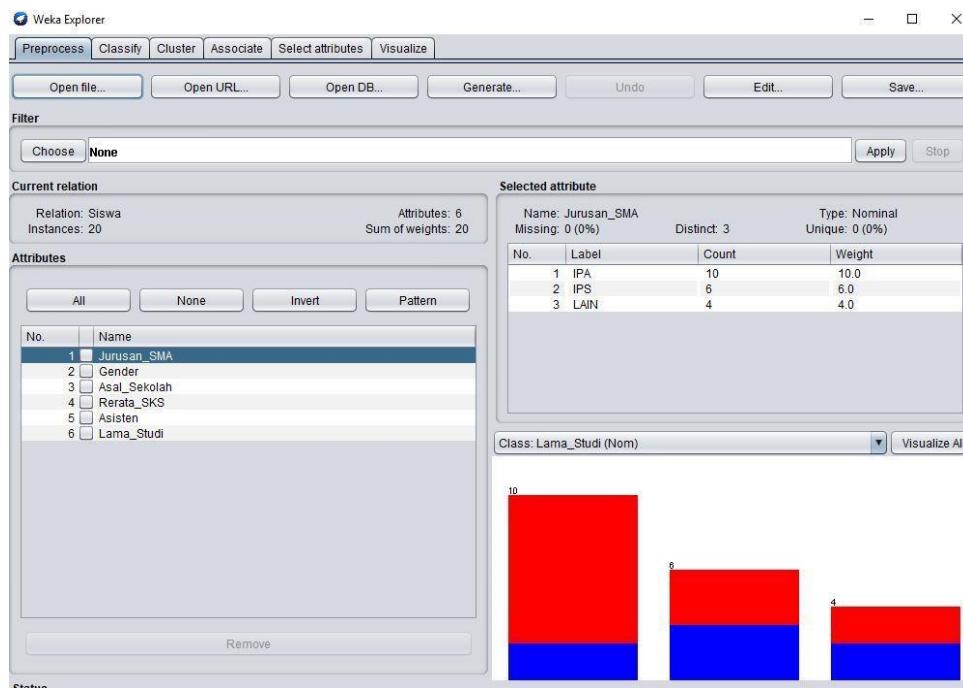
**Excel Spreadsheet:**

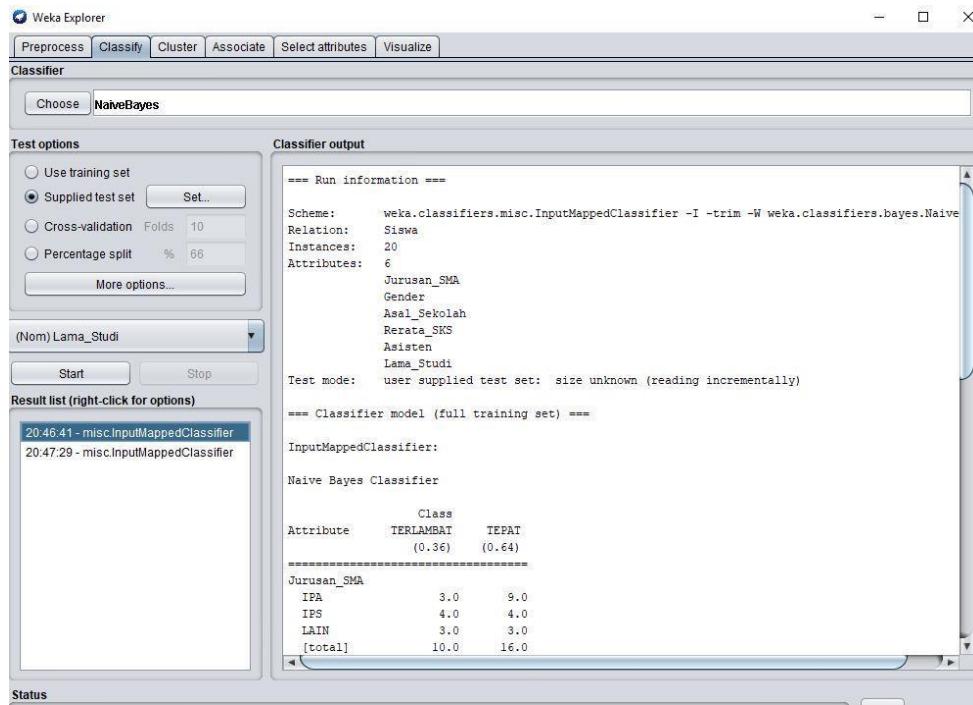
	A	B	C	D	E	F
1	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten	
2	LAIN	WANITA	SURAKARTA	18	TIDAK	
3	IPA	PRIA	SURAKARTA	19	YA	
4	LAIN	PRIA	SURAKARTA	19	TIDAK	
5	IPS	PRIA	LUAR	17	TIDAK	
6	LAIN	WANITA	SURAKARTA	17	TIDAK	
7	IPA	WANITA	LUAR	18	YA	
8	IPA	PRIA	SURAKARTA	18	TIDAK	
9	IPA	PRIA	SURAKARTA	19	TIDAK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

**ARFF File Content:**

```
C:\> Users > acer > Downloads > Documents > Petunjuk Tiap Modul > DataTesting.arff
1 @attribute Jurusan_SMA {IPA, IPS, LAIN}
2 @attribute Gender {PRIA, WANITA}
3 @attribute Asal_Sekolah {SURAKARTA, LUAR}
4 @attribute Rerata_SKS real
5 @attribute Asisten {YA, TIDAK}
6 @attribute Lama_Studi {TEPAT, TERLAMBAT}
7
8 @data
9 LAIN,WANITA,SURAKARTA,18,TIDAK,?
10 IPA,PRIA,SURAKARTA,19,YA,?
11 LAIN,PRIA,SURAKARTA,19,TIDAK,?
12 IPS,PRIA,LUAR,17,TIDAK,?
13 LAIN,WANITA,SURAKARTA,17,TIDAK,?
14 IPA,WANITA,LUAR,18,YA,?
15 IPA,PRIA,SURAKARTA,18,TIDAK,?
16 IPA,PRIA,SURAKARTA,19,TIDAK,?
17 IPS,PRIA,LUAR,18,TIDAK,?
18 LAIN,WANITA,SURAKARTA,18,TIDAK,?
```

## 2. Melakukan prediksi menggunakan WEKA





Status

Relation: Mahasiswa_predicted						
	Nominal	Nominal	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

Repositories

(20 / 20): all

Jurusan_SM	Gender	Asal_Sekolah	Rerata_SKS	Asisten
polyno...	binomi...	binomi...	integer	binomi...
attribute	attribute	attribute	attribute	attribute
LAIN	WANITA	SURAKARTA	18	TIDAK
IPA	PRIA	SURAKARTA	19	YA
LAIN	PRIA	SURAKARTA	19	TIDAK
IPS	PRIA	LUAR	17	TIDAK
LAIN	WANITA	SURAKARTA	17	TIDAK
IPA	WANITA	LUAR	18	YA
IPA	PRIA	SURAKARTA	18	TIDAK
IPA	PRIA	SURAKARTA	19	TIDAK
IPS	PRIA	LUAR	18	TIDAK
LAIN	WANITA	SURAKARTA	18	TIDAK

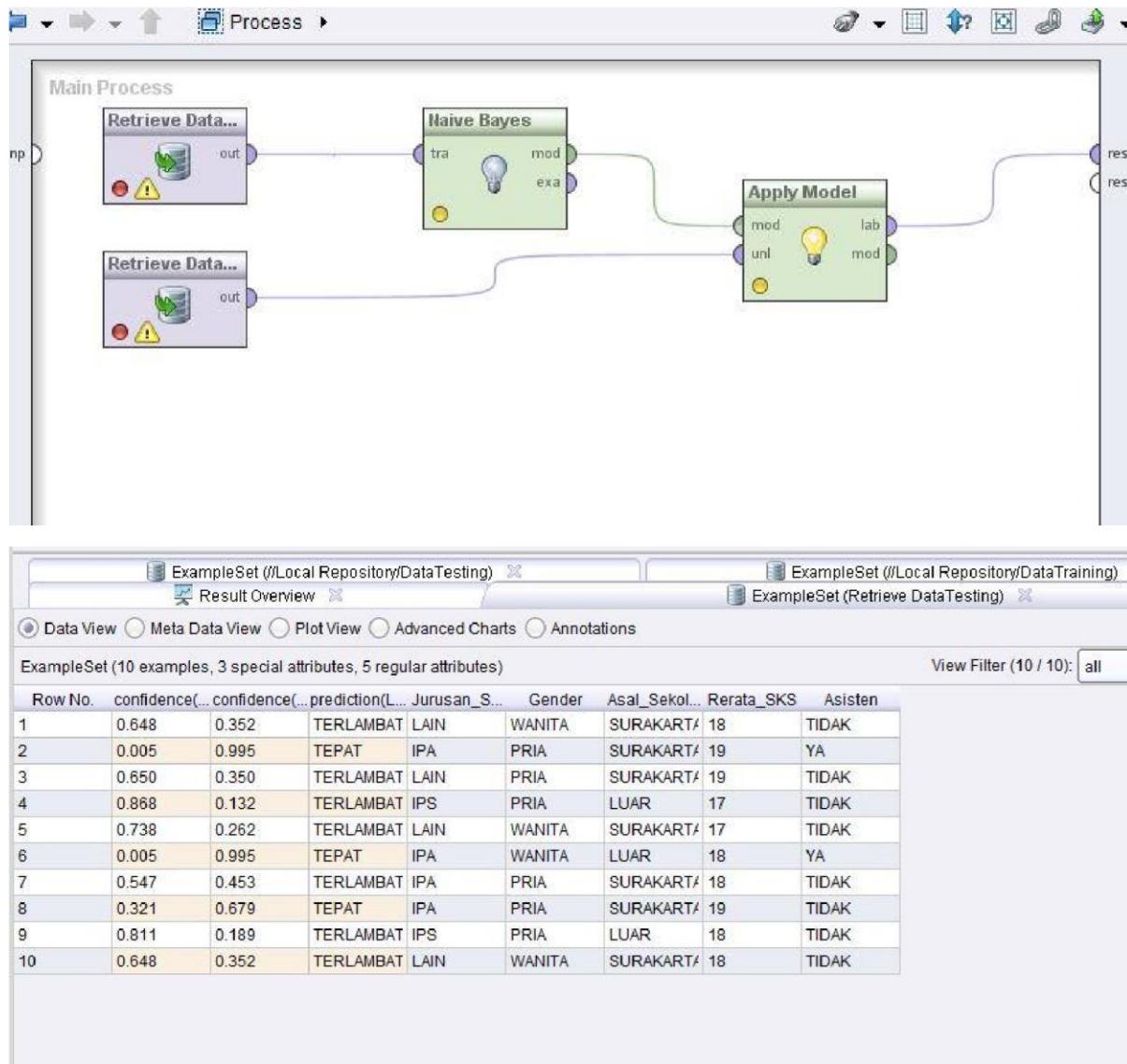
0 errors.

Local Repository

Name DataTesting

Location //Local Repository/DataTesting

Previous Next Cancel OK



4. Nilai Rerata confidence untuk atribut Lama\_Studi:
  - a. Nilai TEPAT : 0.476
  - b. Nilai TERLAMBAT : 0.524
5. Dari hasil percobaan tugas nomor 3, didapatkan hasil yaitu 3 Mahasiswa akan lulus TEPAT waktu dan 7 Mahasiswa akan lulus TERLAMBAT.
6. Dewi adalah Mahasiswa WANITA, berasal dari jurusan IPA pada saat 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 bersal dari jurusan IPA maupun IPS, bersekolah di SURAKARTA, mengambil SKS rata-rata sebanyak 17 tiap semester, dan PERNAH menjadi asisten selama kuliah.

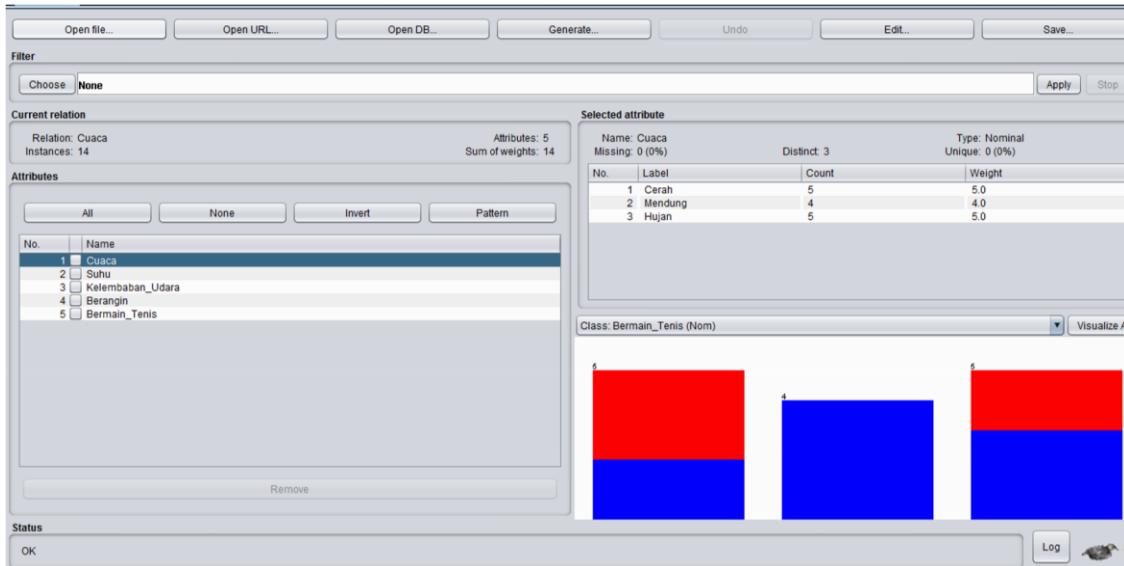
2	TEPAT	0.076	0.924	LAIN	PRIA	SURAKARTA	17	YA
---	-------	-------	-------	------	------	-----------	----	----

## MODUL 9

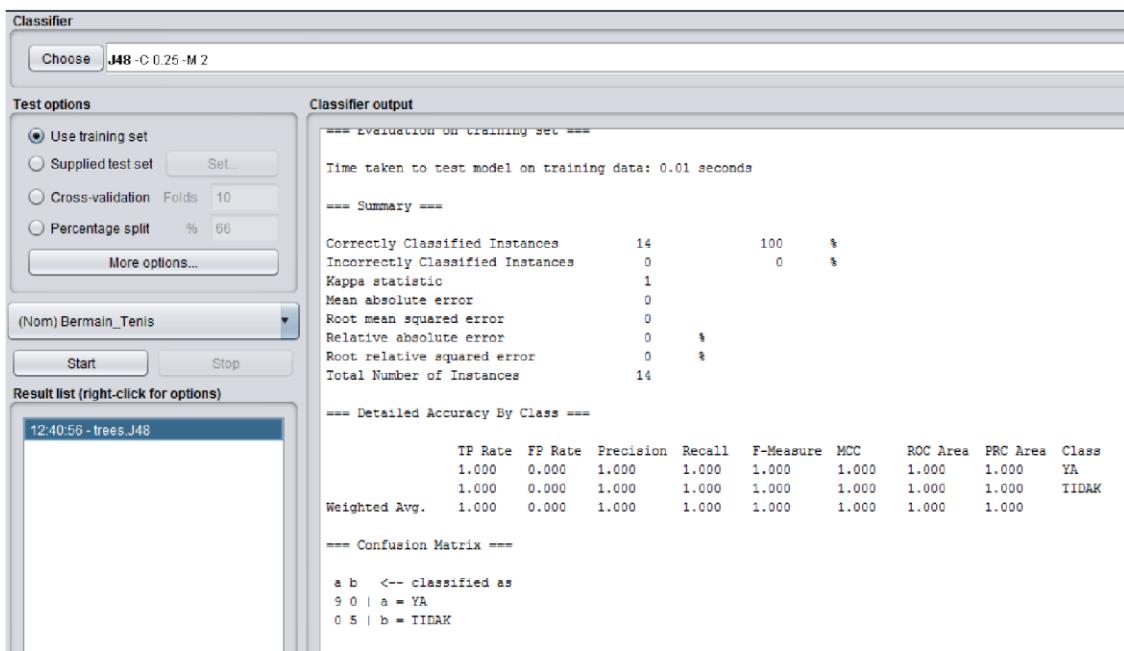
Latihan

#### 9.4.1 Pohon Keputusan Menggunakan Weka

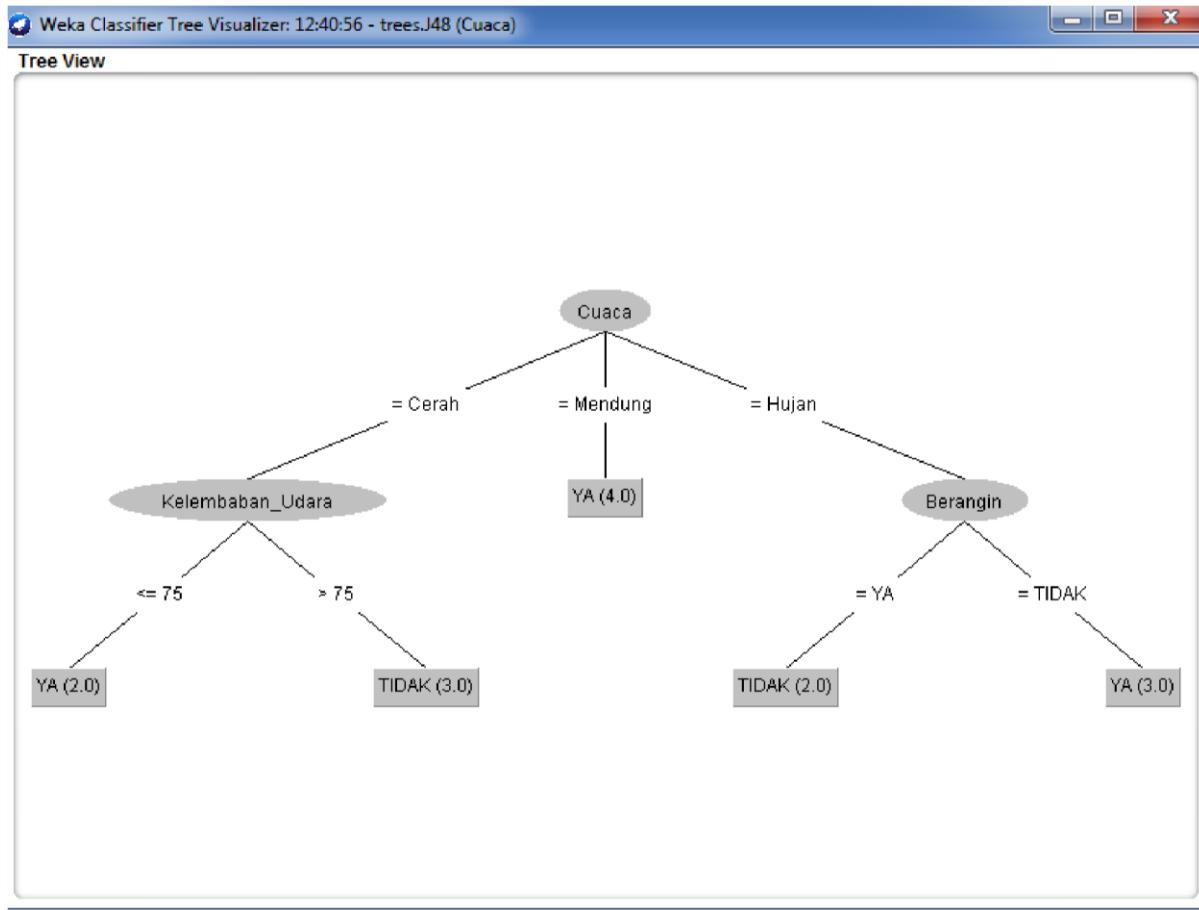
Membuka file cuaca.arff dengan Weka Explorer



Pada tab Classify tekan tombol choose – trees – j48, lalu pada pilihan Test Option pilih Use training set dan atribut dependen adalah Bermain\_Tenis. Kemudian start



Pada Result list, klik kanan pada hasil trees.j48 – visualize tree



#### 9.4.2 Pohon Keputusan Menggunakan RapidMiner

Membuka RapidMiner dan import data bernama DataCuaca\_Training

Import Data - Select the cells to import.

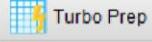
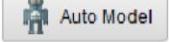
Select the cells to import.

Sheet: Sheet1 ▾ 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

← Previous → Next ✖ Cancel

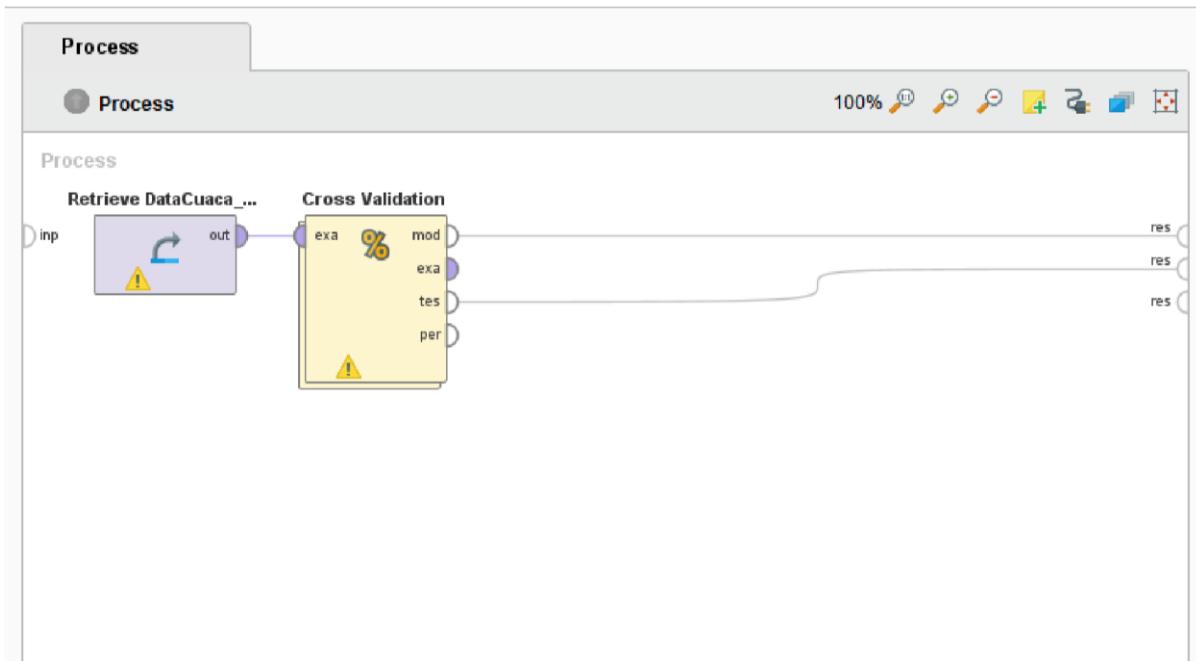
ExampleSet (/Local Repository/DataCuaca\_Training) X

Open in  Turbo Prep  Auto Model Filter (14 / 14 ex)

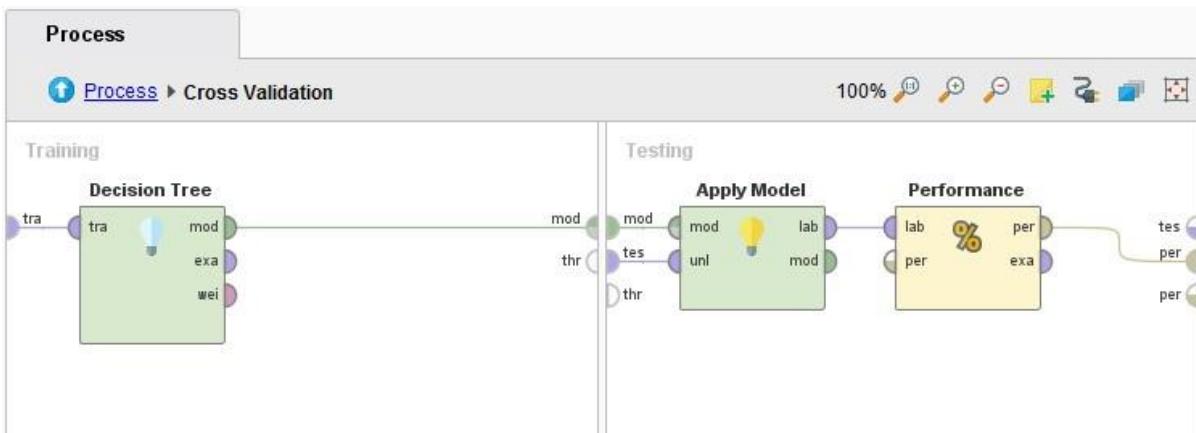
Row No.	Bermain_Te...	Cuaca	Suhu	Kelembaban...	Berangin
1	TIDAK	Cerah	85	85	TIDAK
2	TIDAK	Cerah	80	90	YA
3	YA	Mendung	83	86	TIDAK
4	YA	Hujan	70	96	TIDAK
5	YA	Hujan	68	80	TIDAK
6	TIDAK	Hujan	65	70	YA
7	YA	Mendung	64	65	YA
8	TIDAK	Cerah	72	95	TIDAK
9	YA	Cerah	69	70	TIDAK
10	YA	Hujan	75	80	TIDAK
11	YA	Cerah	75	70	YA
12	YA	Mendung	72	90	YA
13	YA	Mendung	81	75	TIDAK

ExampleSet (14 examples, 1 special attribute, 4 regular attributes)

Drag DataCuaca\_Training ke area process view, dan juga operator Cross Validation

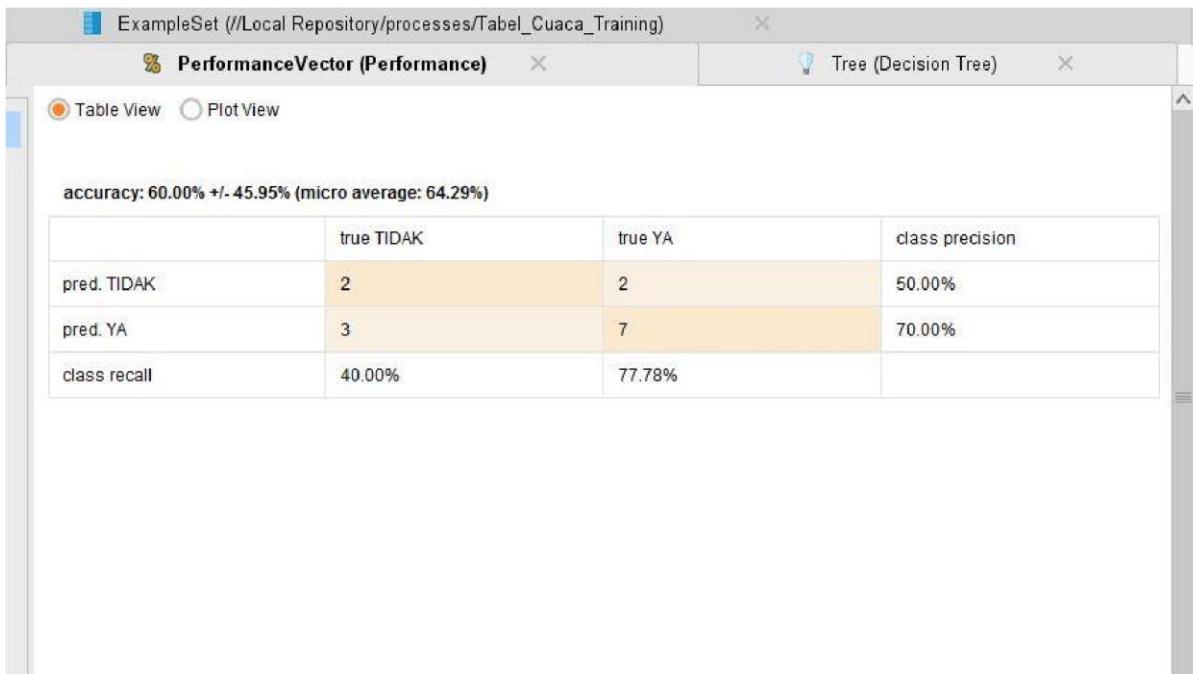


Setelah klik ganda operator Cross Validation, masukkan operator Decision Tree, Apply Model dan Performance ke dalam area training seperti gambar di bawah ini

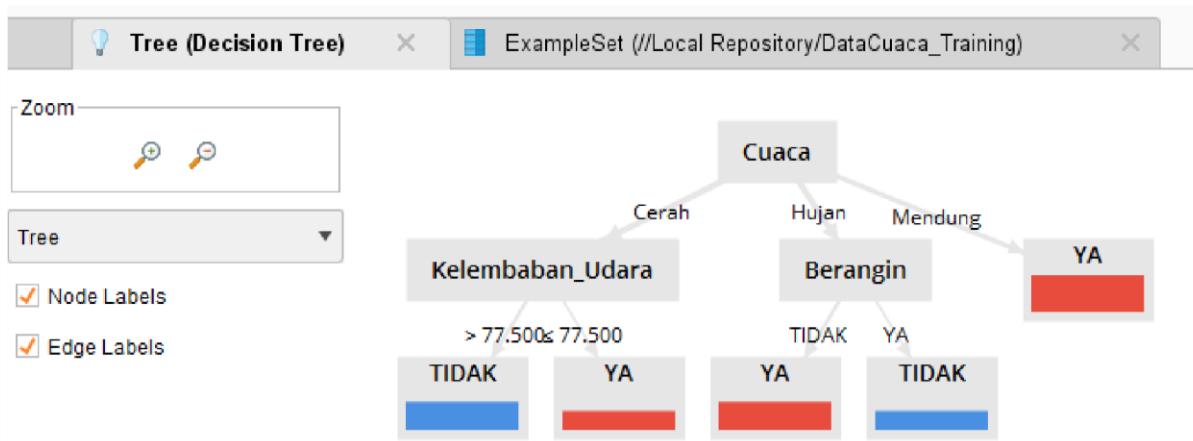


Setelah di Run akan muncul 2 hasil proses klasifikasi pohon keputusan:

## 1. PerformanceVector

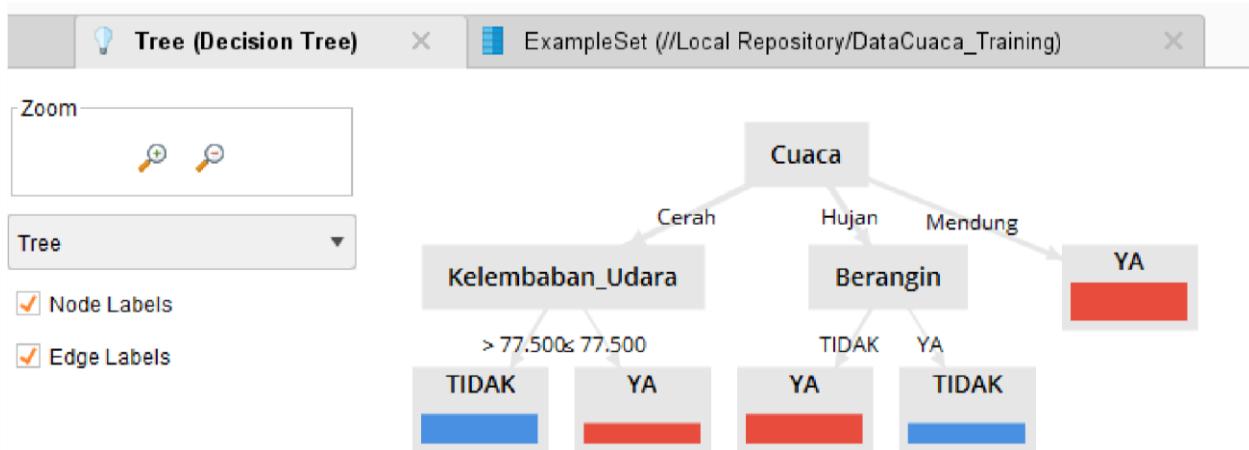


## 2. DecisionTree



Tugas

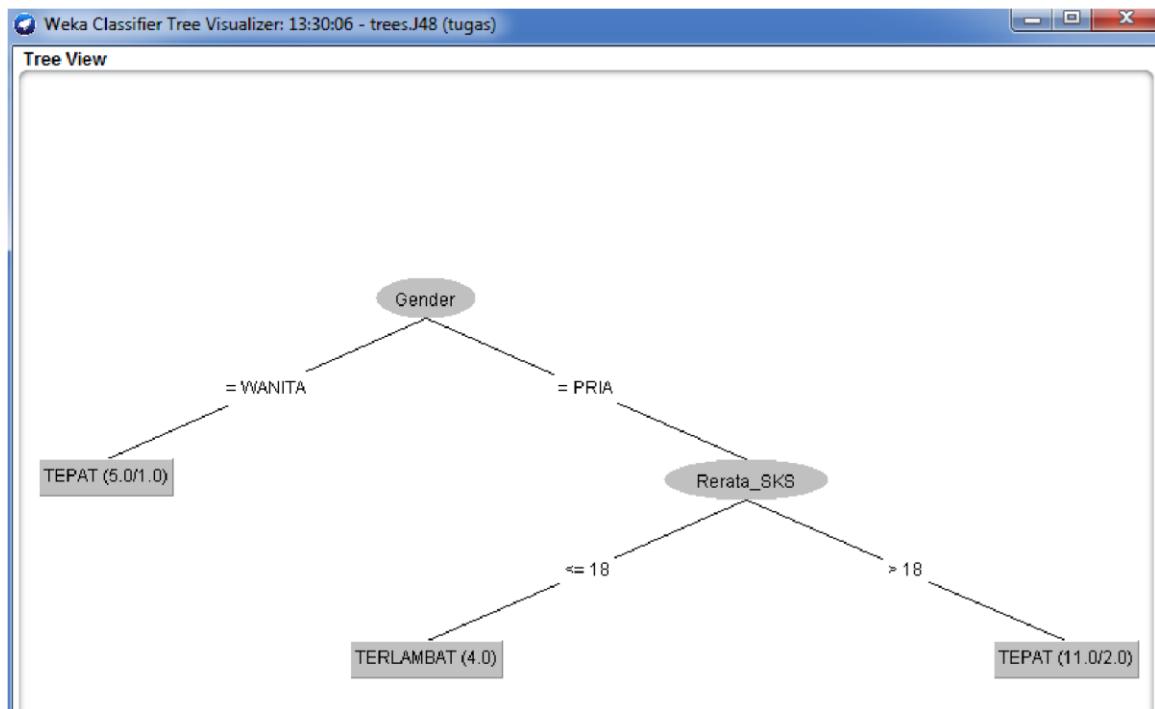
1. Berdasarkan pohon keputusan, kelas atribut Bermain\_Tenis pada table Testing didapatkan hasil berikut



	A	B	C	D	E
1	Cuaca	Suhu	Kelembaban_udara	Berangin	Bermain_Tenis
2	Cerah	75	65	TIDAK	YA
3	Cerah	80	68	YA	YA
4	Cerah	83	87	YA	TIDAK
5	Mendung	70	96	TIDAK	YA
6	Mendung	68	81	TIDAK	YA
7	Hujan	65	75	TIDAK	YA
8	Hujan	64	85	YA	TIDAK

2.

Membuat dan mencetak pohon keputusan berdasarkan file arff pada tugas nomor 1 dalam modul 7

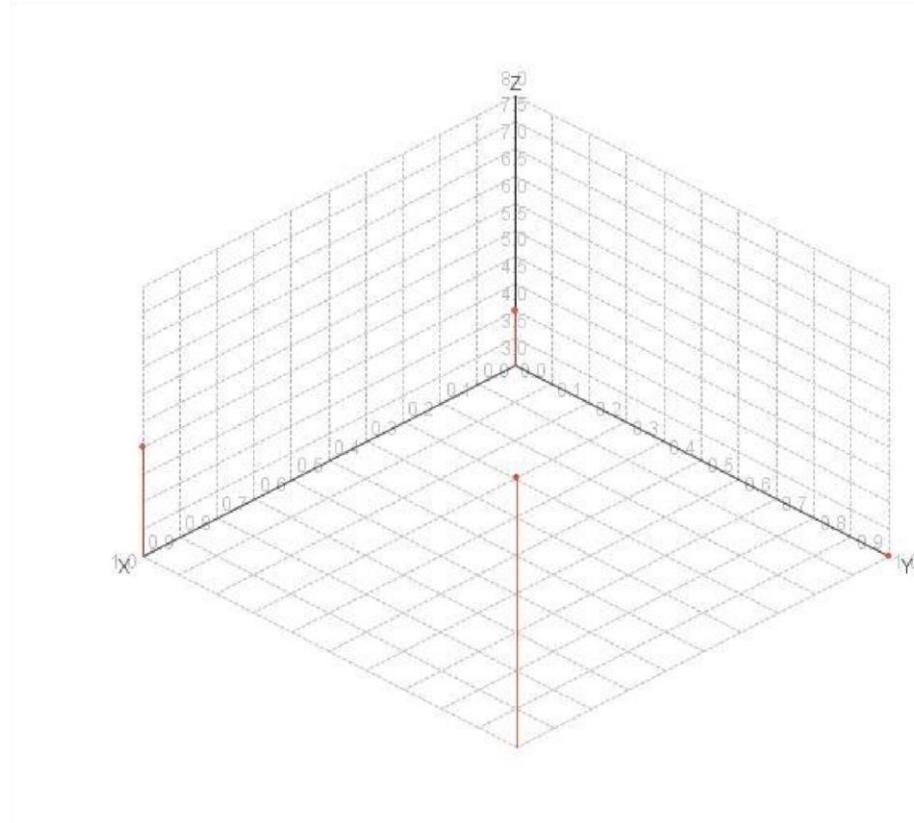
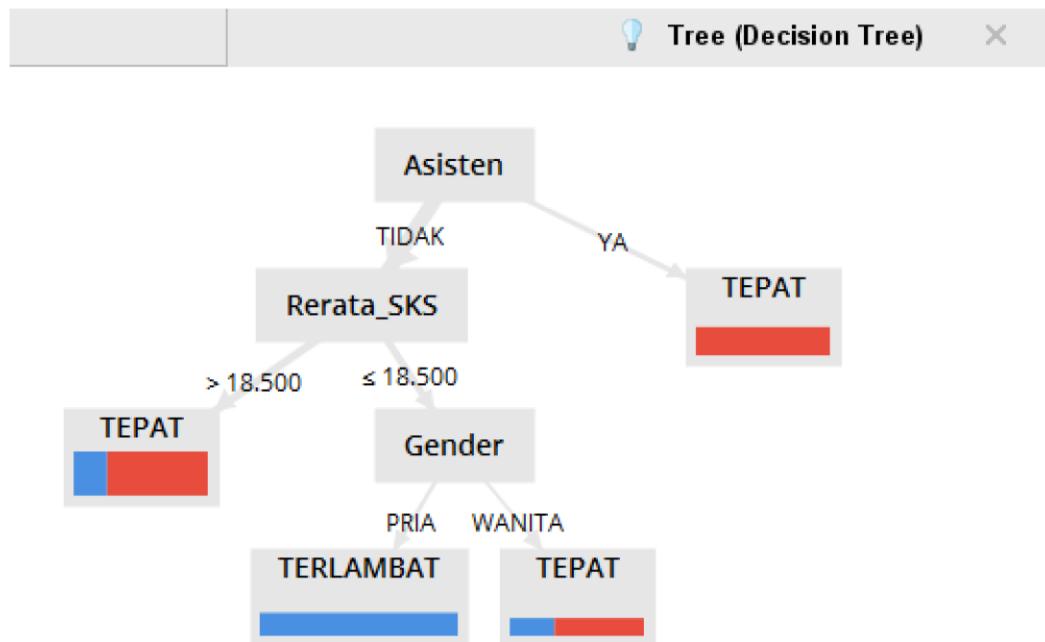


Mencari nilai-nilai parameter:

- i. jumlah simpul daun pada pohon keputusan = 3
- ii. jumlah simpul keseluruhan pada pohon keputusan = 5
- iii. waktu yang dibutuhkan untuk proses pelatihan = 0 detik
- iv. tingkat ketepatan klasifikasi = 85%
- v. tingkat ketidaktepatan klasifikasi = 15%

3.

Membuat dan mencetak pohon keputusan berdasarkan file arff pada tugas nomor 1 dalam modul 6



4.

Klasifikasi kondisi dan simpul yang terbentuk berdasarkan pohon keputusan dari soal nomor 2

- a. simpul akar : Gender (Pria, Wanita)
- b. simpul internal : Rerata\_SKS
- c. simpul daun : Terlambat, Tepat

5.

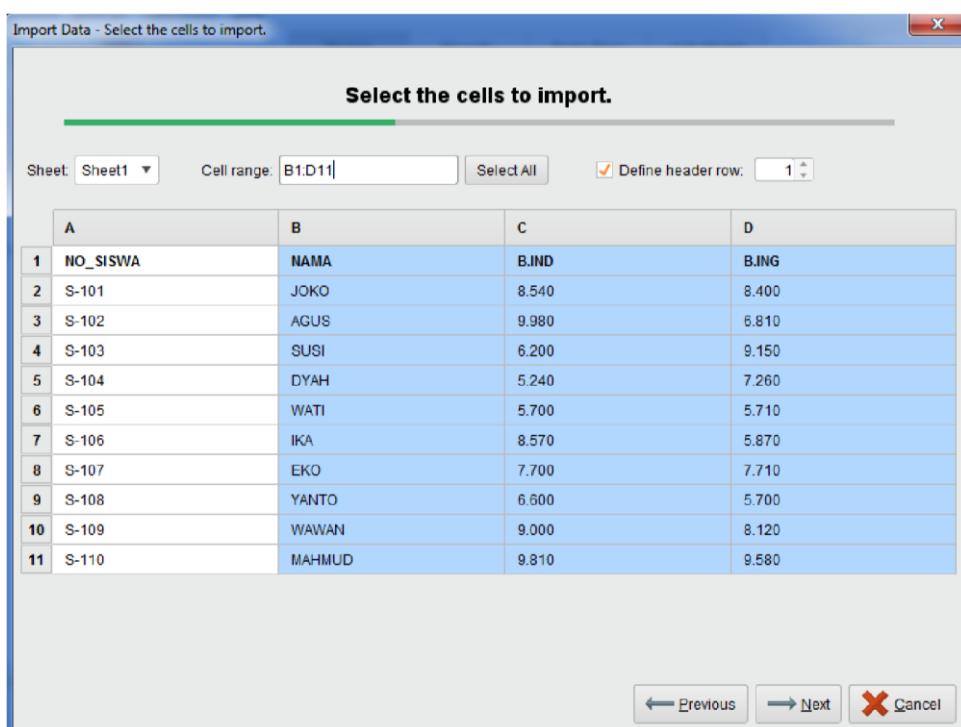
## MODUL 10

### Latihan

Table data nilai siswa

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

Proses clustering menggunakan RapidMiner



Import Data - Format your columns.

### Format your columns.

Replace errors with missing values  ⓘ

	NAMA polynomial id	B.IND real	B.ING real
1	JOKO	8.540	8.400
2	AGUS	9.980	6.810
3	SUSI	6.200	9.150
4	DYAH	5.240	7.260
5	WATI	5.700	5.710
6	IKA	8.570	5.870
7	EKO	7.700	7.710
8	YANTO	6.600	5.700
9	WAWAN	9.000	8.120
10	MAHMUD	9.810	9.580

 no problems.

 Previous  Next  Cancel

Import Data - Where to store the data?

### Where to store the data?

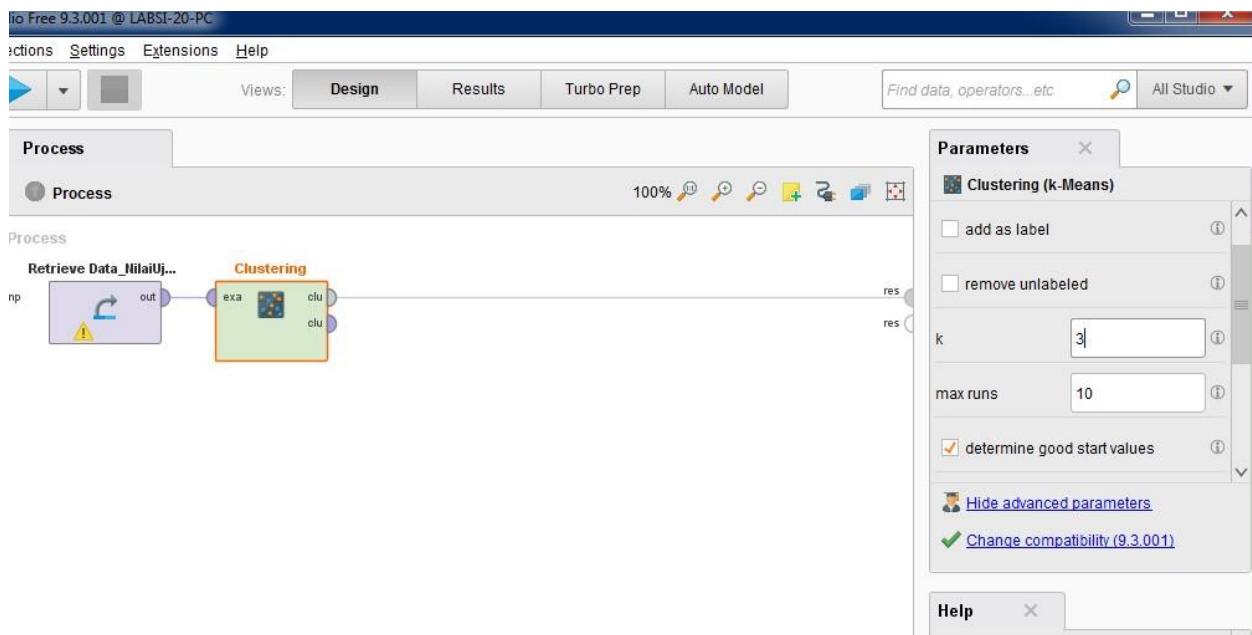
- Local Repository (LABSI-20)
  - Connections (LABSI-20)
  - data (LABSI-20)
  - processes (LABSI-20)

Name

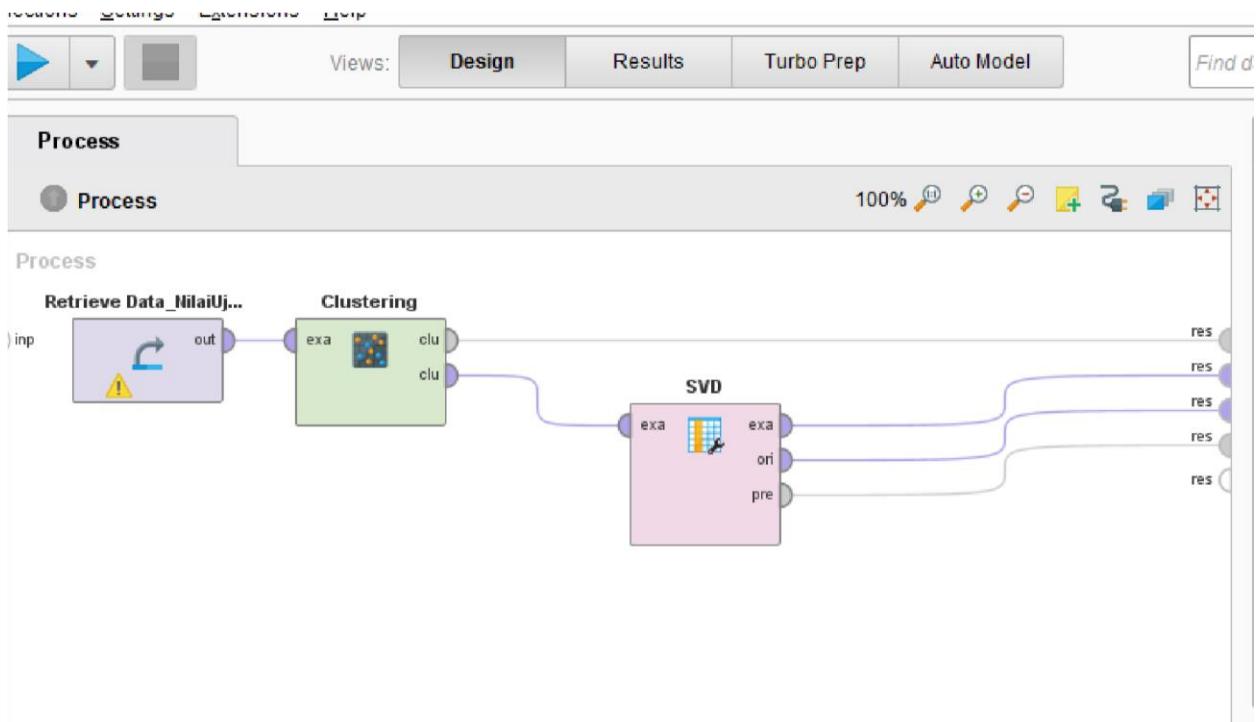
Location //Local Repository/Data\_NilaiUjian

 Previous  Finish  Cancel

## Cluster k-Means ada 3



Menambahkan SVD lalu di Run



Hasil proses Clustering dengan algoritma K-Means

a) SVD

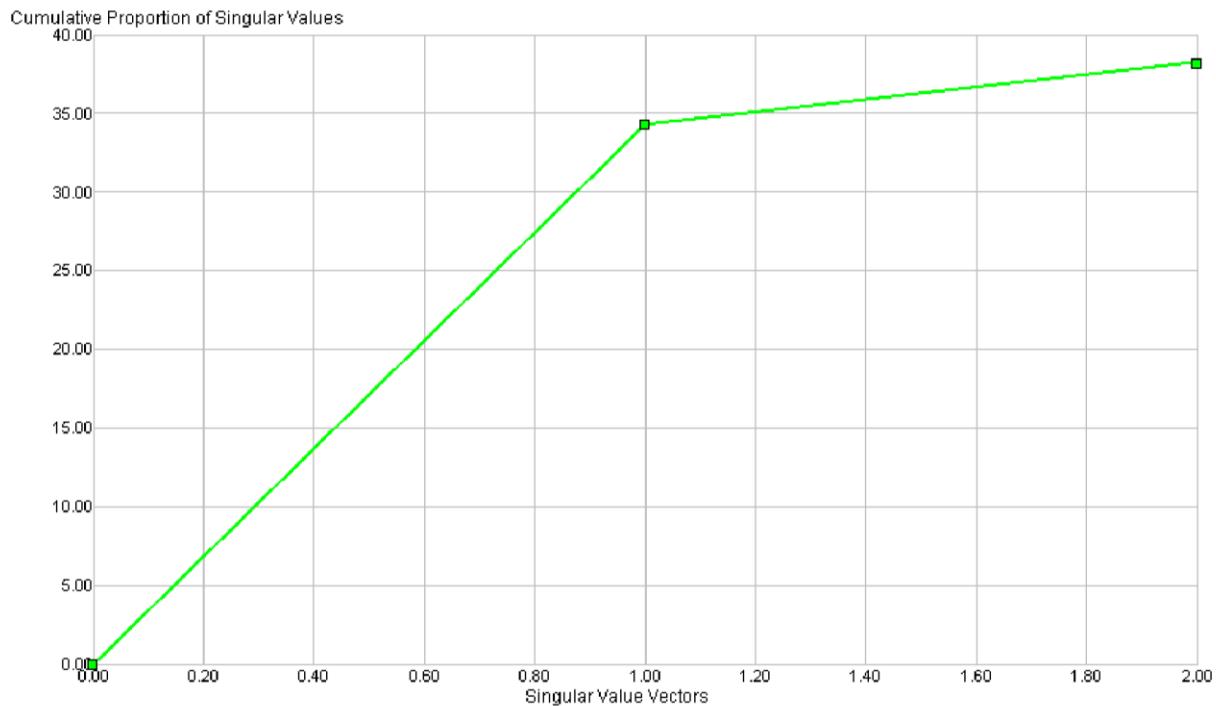
i. Nilai Eigenvalue

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

ii. Nilai SVD vectors

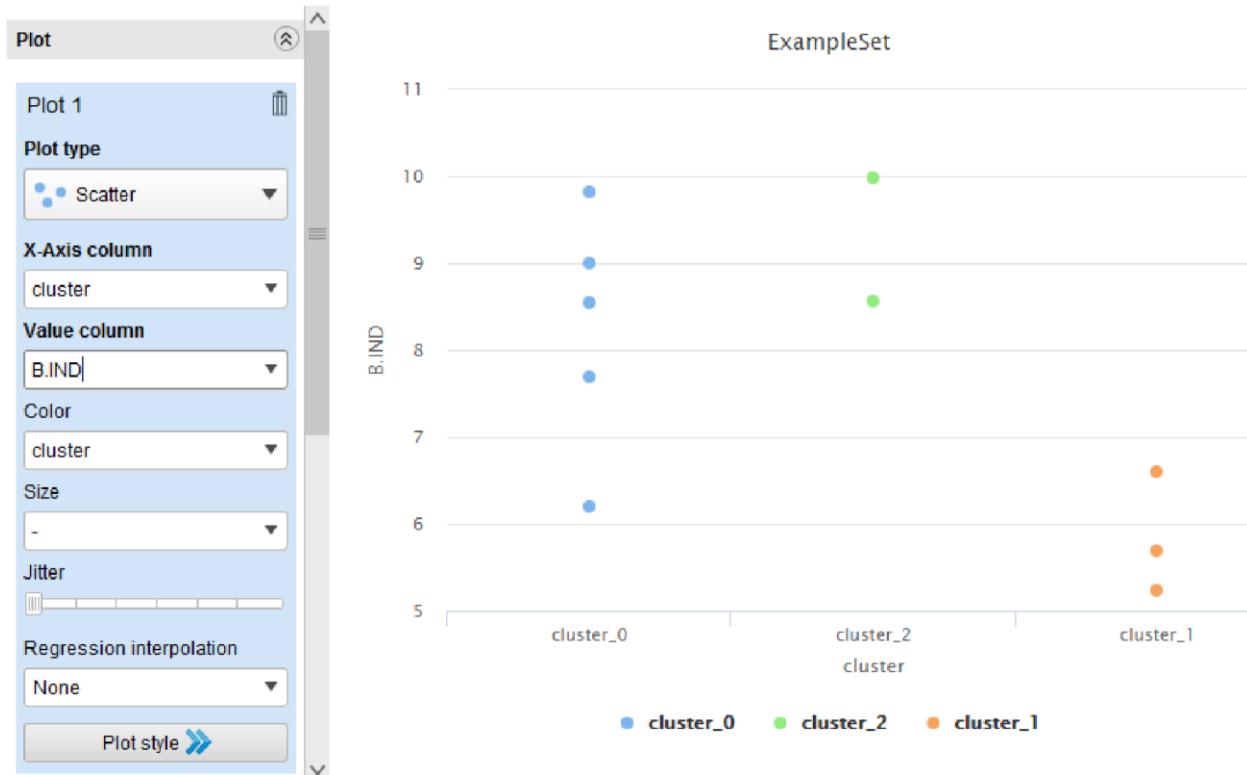
Attribute	SVD Vector 1
B.IND	0.723
B.ING	0.690

iii. Nilai Cumulative variance

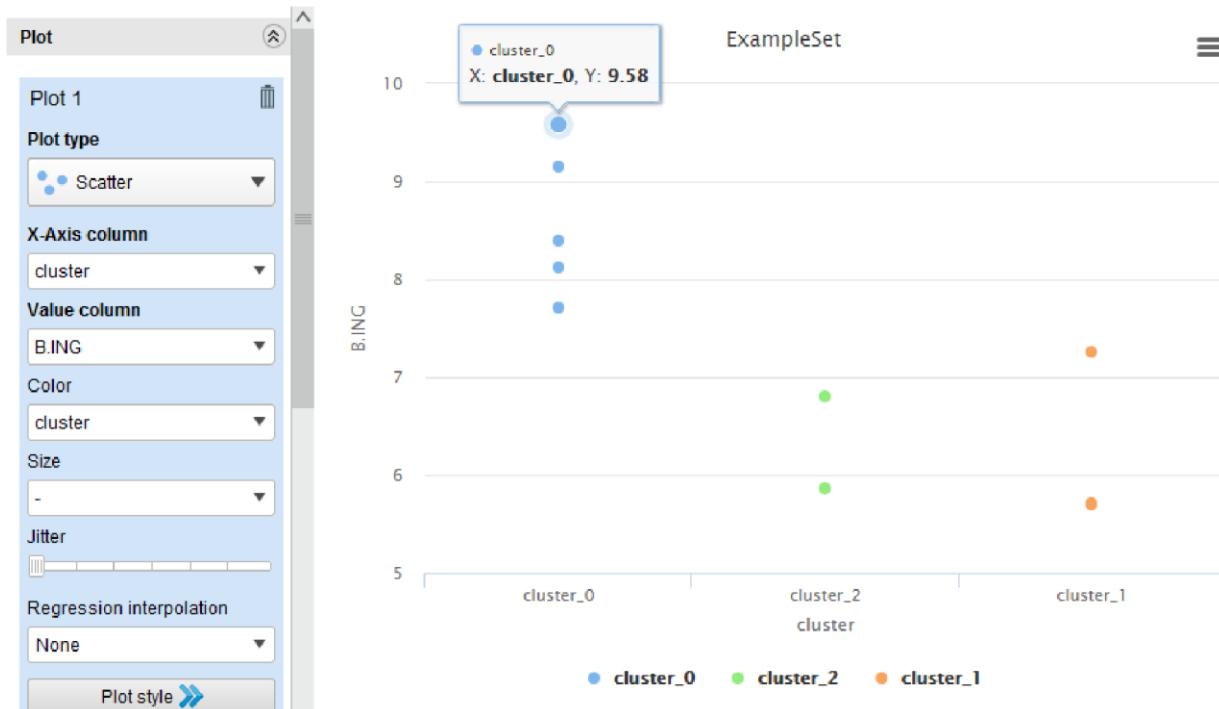


b) ExampleSet (k-Means)

i. Kelompok siswa bidang B.Indonesia



ii. Kelompok siswa bidang B. Inggris



c) ExampleSet (SVD)

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 (Clusering)

i. Description

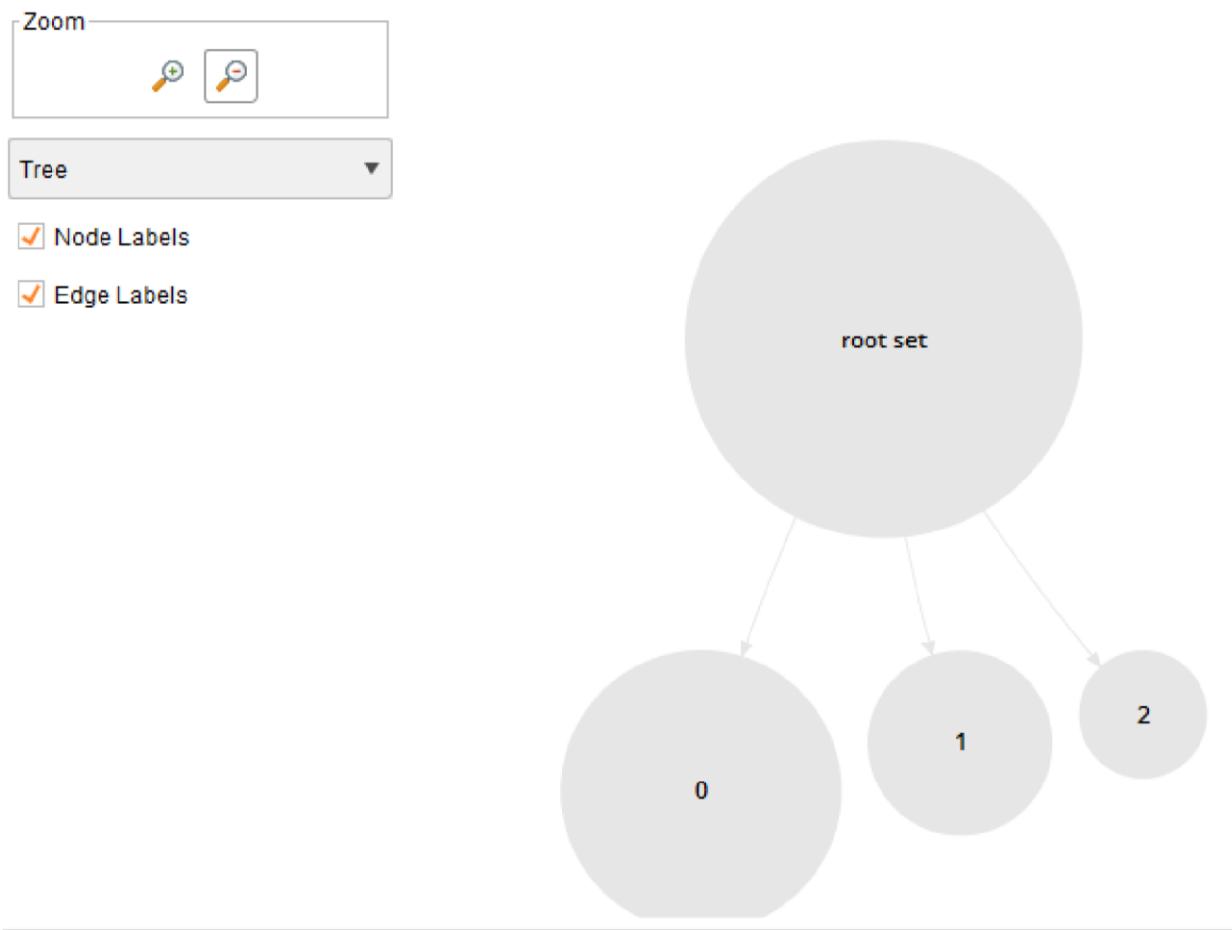
## Cluster Model

```

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

```

ii. Graph



### Interpretasi hasil algoritma K-Means

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

Tugas

Table data nilai ujian

NO_SISWA	NAMA	B.IND	B.ING	MTK	IPA
S-101	JOKO	6.06	7.04	9.41	7.47
S-102	AGUS	6.27	6.14	8.91	5.12
S-103	SUSI	8.68	9.66	6.61	6.57
S-104	DYAH	9.39	9.51	6.83	5.08
S-105	WATI	8.21	9.39	5.81	8.15
S-106	IKA	6.90	5.98	9.90	6.60
S-107	EKO	7.36	6.19	8.76	7.87
S-108	YANTO	5.35	6.51	8.67	9.54
S-109	WAWAN	9.86	5.69	8.89	5.72
S-110	MAHMUD	9.54	6.98	8.90	5.04
S-111	BUDI	6.77	9.99	9.52	6.67
S-112	SANTI	9.57	8.61	8.72	8.93
S-113	DIAN	9.87	6.60	8.45	5.89
S-114	DANI	9.31	7.12	7.90	6.11
S-115	AHMAD	5.02	5.37	5.02	8.96
S-116	BAYU	7.24	8.68	6.28	9.75
S-117	RISA	8.25	7.28	6.16	7.51
S-118	RANI	7.60	9.51	6.58	7.66
S-119	YANI	8.49	7.70	8.45	6.04
S-120	RATIH	7.97	8.43	6.46	8.89
S-121	INDAH	5.27	6.67	9.52	6.85
S-122	JONO	9.92	8.23	7.36	7.83
S-123	SARAH	8.36	5.91	5.35	6.57
S-124	RAMA	6.91	8.62	7.39	6.38
S-125	BAMBANG	9.66	9.00	6.88	8.97
S-126	HADI	5.67	9.48	7.42	6.35
S-127	NANA	7.76	5.29	9.22	5.65
S-128	FEBRI	5.99	6.33	7.86	6.91
S-129	DENI	6.43	5.52	8.75	7.12
S-130	TONI	8.46	5.79	9.69	9.91

## Proses clustering menggunakan RapidMiner

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Sheet1 ▾ Cell range: B:F Select All  Define header row: 1

	A	B	C	D	E	F
1	NO_SISWA	NAMA	B.IND	B.ING	MTK	IPA
2	S-101	JOKO	6.862	9.457	6.296	7.125
3	S-102	AGUS	6.359	5.393	7.296	5.781
4	S-103	SUSI	6.147	7.829	6.751	9.936
5	S-104	DYAH	9.432	5.173	8.262	7.558
6	S-105	WATI	8.074	5.985	8.507	9.882
7	S-106	IKA	6.453	5.654	6.804	9.457
8	S-107	EKO	5.816	9.010	8.119	9.439
9	S-108	YANTO	8.733	9.317	7.920	7.992
10	S-109	WAWAN	8.041	9.719	7.600	6.320
11	S-110	MAHMUD	7.278	7.803	5.026	8.300
12	S-111	BUDI	5.492	7.850	5.256	9.440
13	S-112	SANTI	9.799	5.907	6.544	6.566
14	S-113	DIAN	9.590	5.676	8.094	9.036
15	S-114	DANI	7.000	9.100	8.000	5.000

Previous Next  Cancel

Format your columns.

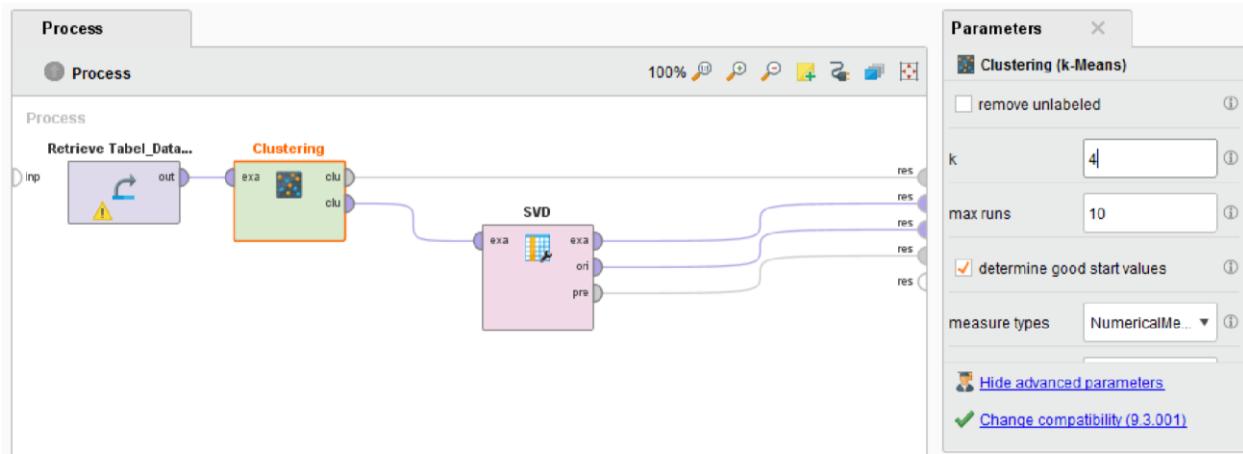
Replace errors with missing values 

	NAMA polynominal id	B.IND real	B.ING real	MTK real	IPA real
1	JOKO	6.862	9.457	6.296	7.125
2	AGUS	6.359	5.393	7.296	5.781
3	SUSI	6.147	7.829	6.751	9.936
4	DYAH	9.432	5.173	8.262	7.558
5	WATI	8.074	5.985	8.507	9.882
6	IKA	6.453	5.654	6.804	9.457
7	EKO	5.816	9.010	8.119	9.439
8	YANTO	8.733	9.317	7.920	7.992
9	WAWAN	8.041	9.719	7.600	6.320
10	MAHMUD	7.278	7.803	5.026	8.300
11	BUDI	5.492	7.850	5.256	9.440
12	SANTI	9.799	5.907	6.544	6.566
13	DIAN	9.590	5.676	8.094	9.036

 no problem

Previous Next  Cancel

Cluster k-Means ada 4 dan menambahkan SVD lalu di Run



Hasil proses Clustering dengan algoritma K-Means

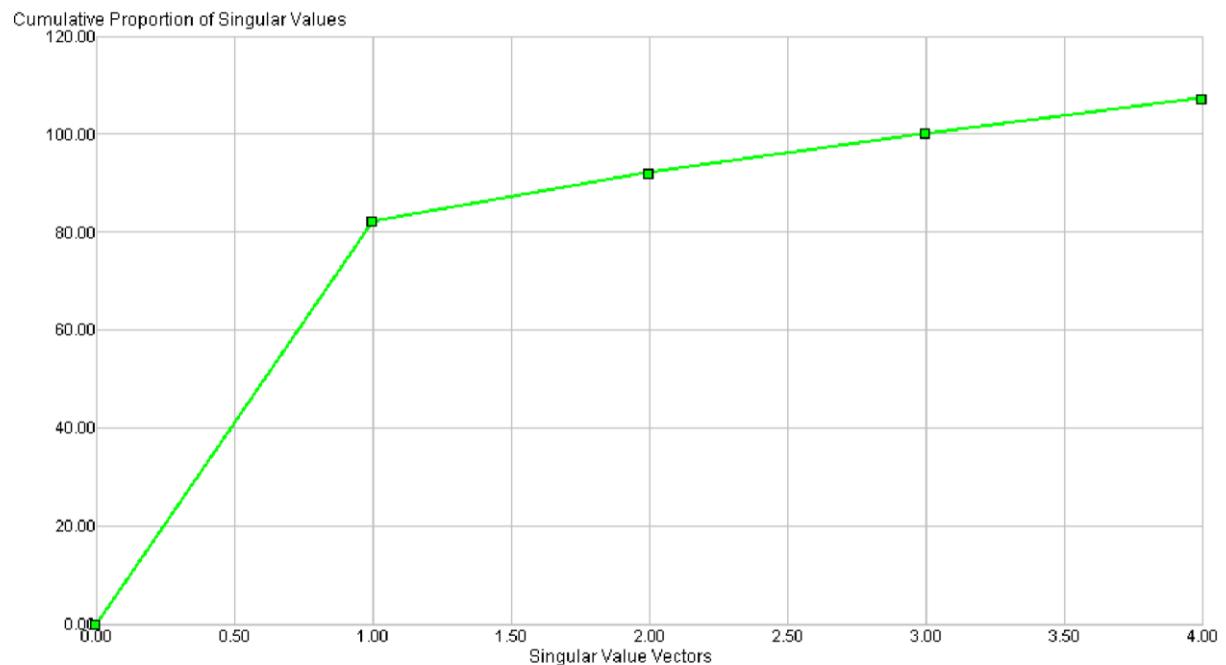
e) SVD iv. Nilai Eigenvalue

Component	Singular Value	Proportion of Singular V...	Cumulative Singular Val...	Cumulative Proportion o...
SVD 1	82.367	0.767	82.367	0.767
SVD 2	9.828	0.092	92.195	0.859
SVD 3	8.128	0.076	100.323	0.934
SVD 4	7.041	0.066	107.364	1.000

v. Nilai SVD vectors

Attribute	SVD Vector 1	SVD Vector 2	SVD Vector 3
B.IND	0.504	-0.584	-0.207
B.ING	0.483	0.477	0.675
MTK	0.494	-0.409	0.246
IPA	0.517	0.514	-0.665

vi. Nilai Cumulative variance



f) ExampleSet (k-Means) iii. Kelompok siswa bidang B.Indonesia



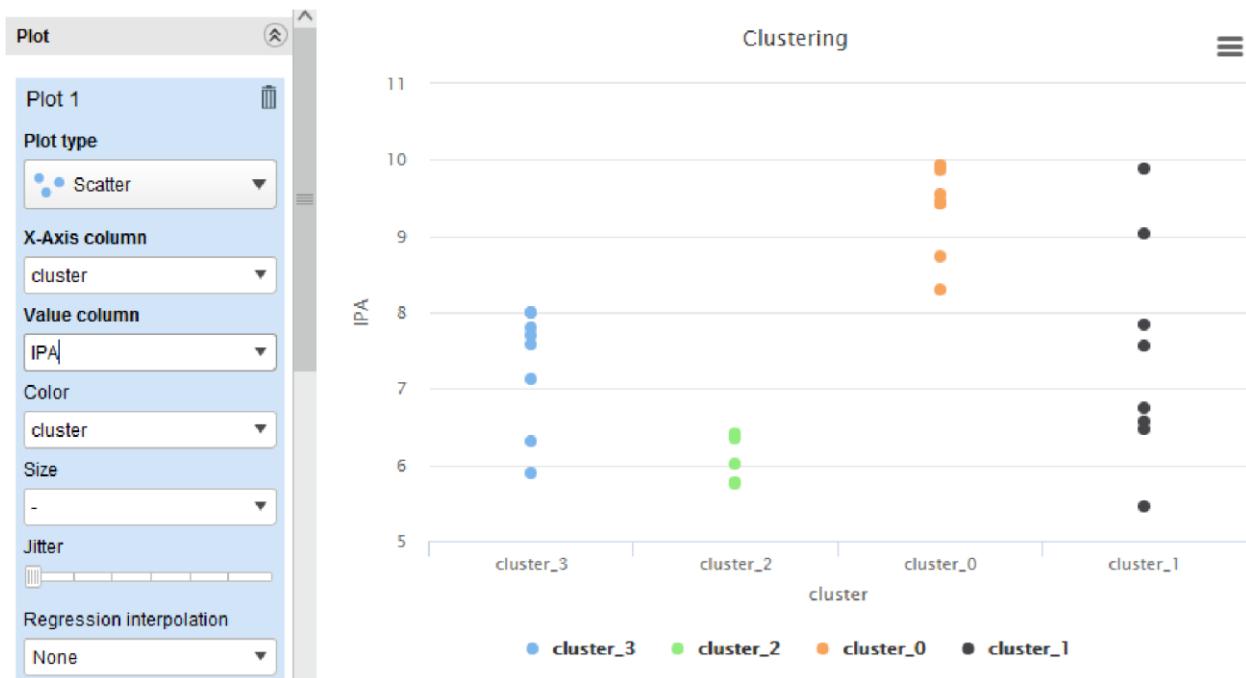
iv. Kelompok siswa bidang B. Inggris



v. Kelompok siswa bidang Matematika



vi. Kelompok siswa bidang IPA



g) ExampleSet (SVD)

Row No.	NAMA	cluster ↑	svd_1
3	SUSI	cluster_0	0.187
6	IKA	cluster_0	0.173
7	EKO	cluster_0	0.197
10	MAHMUD	cluster_0	0.173
11	BUDI	cluster_0	0.171
15	AHMAD	cluster_0	0.192
20	RATIH	cluster_0	0.168
21	INDAH	cluster_0	0.162
23	SARAH	cluster_0	0.186
4	DYAH	cluster_1	0.185
5	WATI	cluster_1	0.198

ExampleSet (30 examples, 2 special attributes, 1 regular attribute)

Row No.	NAMA	cluster ↑	svd_1
12	SANTI	cluster_1	0.175
13	DIAN	cluster_1	0.197
19	YANI	cluster_1	0.186
24	RAMA	cluster_1	0.159
27	NANA	cluster_1	0.181
28	FEBRI	cluster_1	0.191
2	AGUS	cluster_2	0.151
18	RANI	cluster_2	0.157
25	BAMBANG	cluster_2	0.159
26	HADI	cluster_2	0.163
29	DENI	cluster_2	0.186

ExampleSet (30 examples, 2 special attributes, 1 regular attribute)

Row No.	NAMA	cluster ↑	svd_1
25	BAMBANG	cluster_2	0.159
26	HADI	cluster_2	0.163
29	DENI	cluster_2	0.186
1	JOKO	cluster_3	0.180
8	YANTO	cluster_3	0.206
9	WAWAN	cluster_3	0.192
14	DANI	cluster_3	0.189
16	BAYU	cluster_3	0.191
17	RISA	cluster_3	0.195
22	JONO	cluster_3	0.197
30	TONI	cluster_3	0.213

ExampleSet (30 examples, 2 special attributes, 1 regular attribute)

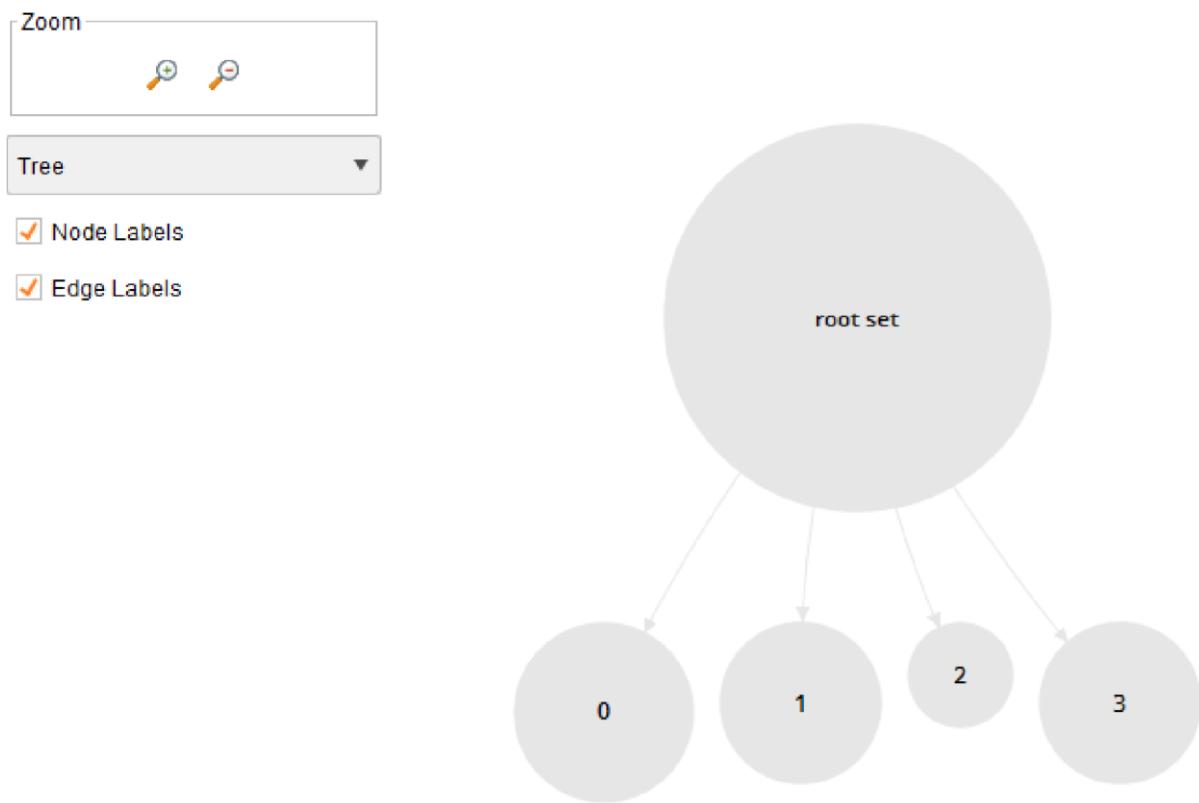
---

h) Cluster Model (Clusering) iii.  
Description

## Cluster Model

```
Cluster 0: 9 items
Cluster 1: 8 items
Cluster 2: 5 items
Cluster 3: 8 items
Total number of items: 30
```

iv. Graph



## Interpretasi hasil algoritma K-Means

Row No.	NAMA	cluster ↑	B.IND	B.ING	MTK	IPA
3	SUSI	cluster_0	6.147	7.829	6.751	9.936
6	IKA	cluster_0	6.453	5.654	6.804	9.457
7	EKO	cluster_0	5.816	9.010	8.119	9.439
10	MAHMUD	cluster_0	7.278	7.803	5.026	8.300
11	BUDI	cluster_0	5.492	7.850	5.256	9.440
15	AHMAD	cluster_0	6.495	7.939	7.290	9.866
20	RATIH	cluster_0	6.376	5.046	6.575	9.547
21	INDAH	cluster_0	7.483	5.043	5.353	8.743
23	SARAH	cluster_0	5.877	9.423	5.373	9.917
4	DYAH	cluster_1	9.432	5.173	8.262	7.558
5	WATI	cluster_1	8.074	5.985	8.507	9.882

ExampleSet (30 examples, 2 special attributes, 4 regular attributes)

Row No.	NAMA	cluster ↑	B.IND	B.ING	MTK	IPA
12	SANTI	cluster_1	9.799	5.907	6.544	6.566
13	DIAN	cluster_1	9.590	5.676	8.094	9.036
19	YANI	cluster_1	8.289	5.171	9.251	7.850
24	RAMA	cluster_1	9.427	6.048	5.292	5.453
27	NANA	cluster_1	9.603	5.399	8.280	6.462
28	FEBRI	cluster_1	9.555	6.680	8.419	6.747
2	AGUS	cluster_2	6.359	5.393	7.296	5.781
18	RANI	cluster_2	5.688	6.358	7.492	6.415
25	BAMBANG	cluster_2	5.579	6.082	8.634	6.004
26	HADI	cluster_2	5.455	6.692	8.955	5.755
29	DENI	cluster_2	6.857	8.083	9.468	6.355

ExampleSet (30 examples, 2 special attributes, 4 regular attributes)

Row No.	NAMA	cluster ↑	B.IND	B.ING	MTK	IPA
25	BAMBANG	cluster_2	5.579	6.082	8.634	6.004
26	HADI	cluster_2	5.455	6.692	8.955	5.755
29	DENI	cluster_2	6.857	8.083	9.468	6.355
1	JOKO	cluster_3	6.862	9.457	6.296	7.125
8	YANTO	cluster_3	8.733	9.317	7.920	7.992
9	WAWAN	cluster_3	8.041	9.719	7.600	6.320
14	DANI	cluster_3	7.836	8.490	9.049	5.886
16	BAYU	cluster_3	9.493	8.816	5.538	7.588
17	RISA	cluster_3	7.805	9.084	7.329	8.004
22	JONO	cluster_3	7.006	9.480	8.285	7.702
30	TONI	cluster_3	9.751	8.126	9.436	7.798

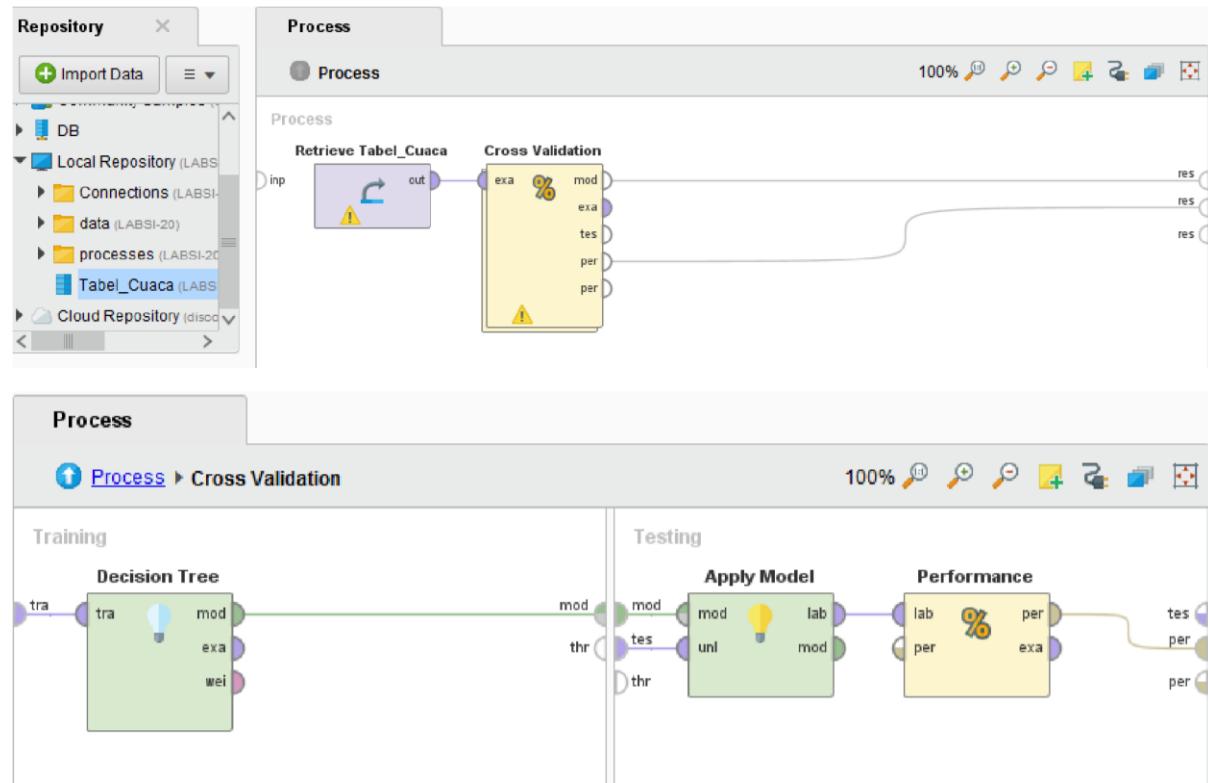
ExampleSet (30 examples, 2 special attributes, 4 regular attributes)

## MODUL 11

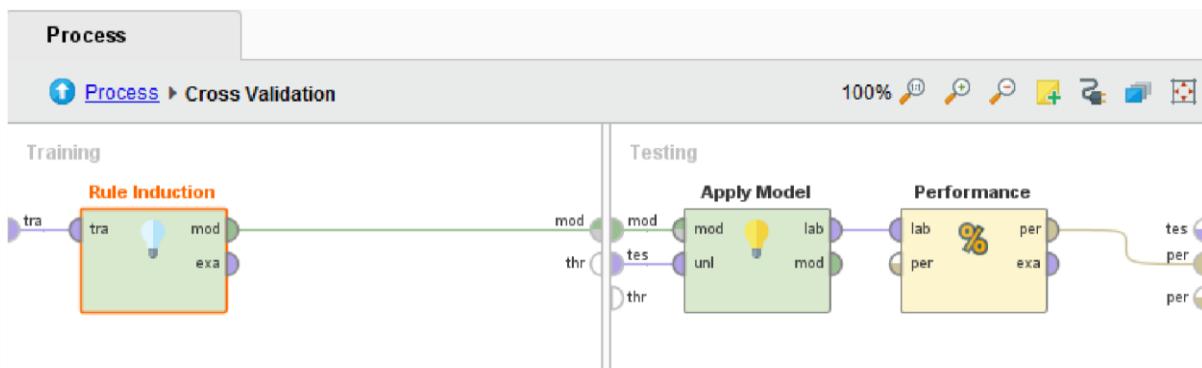
### Latihan

#### 11.4.1. Induksi Aturan Data Cuaca

Menggunakan model proses dari praktikum data mining Modul 9 Kegiatan 9.4.2.



Klik kanan operator Decision Tree pada area Training, pilih Replace Order menjadi Rule Induction



Setelah diRun diperoleh sebuah induksi aturan dari data training yang diberikan disebut sebagai RuleModel (Rule Induction).

## 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)

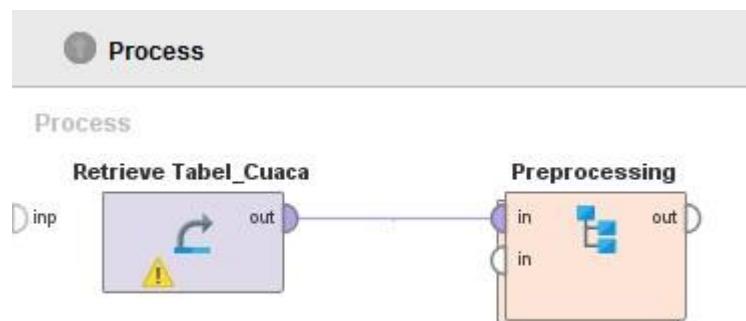
correct: 12 out of 13 training examples.
```

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

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

### 11.4.2. Aturan Asosiasi Data Cuaca

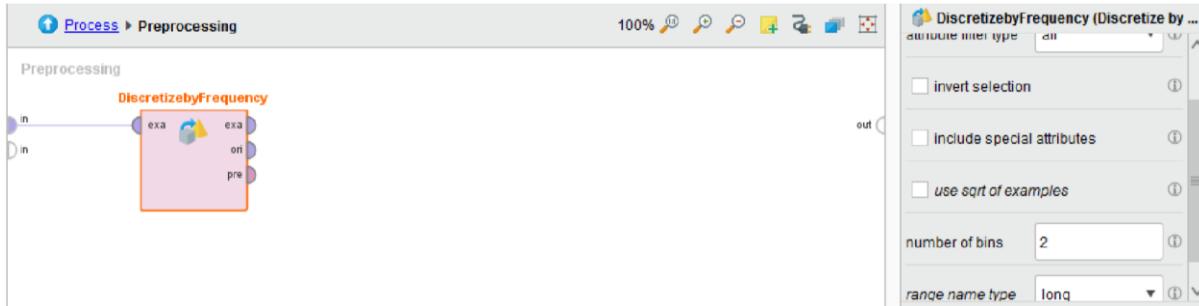
Gunakan DataCuaca\_Training dan drag dari repository ke area Process View tambahkan operator Utility lalu Subprocess ke dalam area. Ubah nama operator ini menjadi Preprocessing.



Klik ganda operator Preprocessing sehingga masuk pada area Nested Chain. Tambahkan operatoroperator berikut:

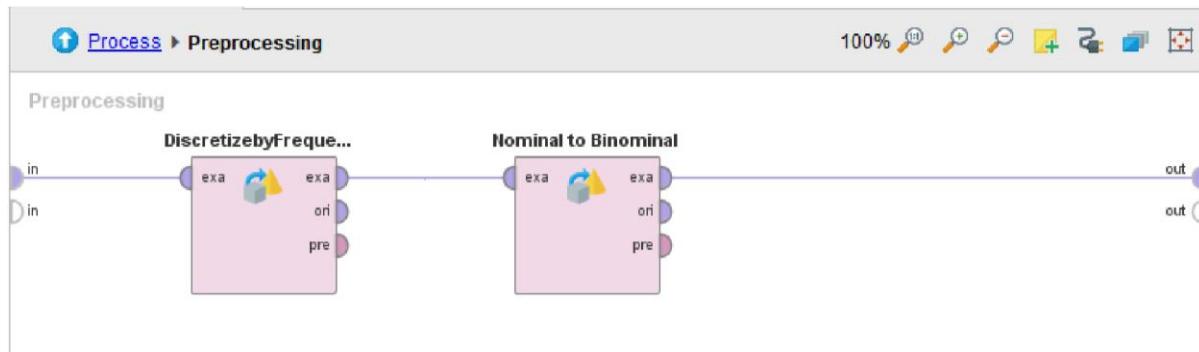
a. Discretize by Frequency

Biarkan nilai parameter number of bins = 2



b. Nominal to Binominal

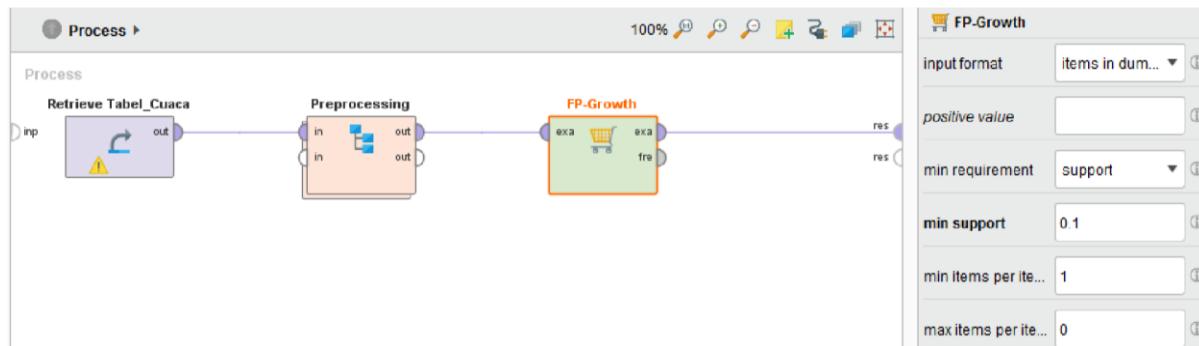
Hubungkan output operator DiscretizebyFrequency dengan masukan examination pada operator ini, dan output examination operator ini dengan panel out.



Kembali ke main process, tambahkan 2 operator:

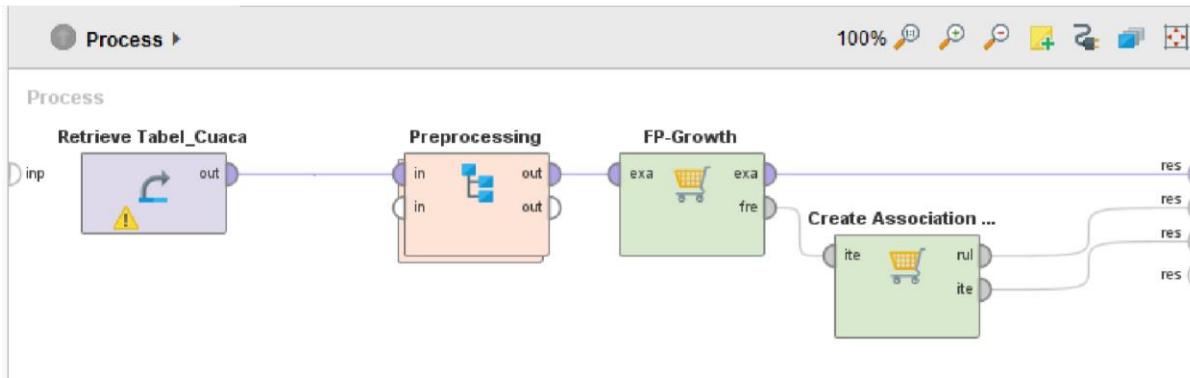
a. FP-Growth

Ubah nilai parameter pada min support = 0.1, hubungkan port out operator Preprocessing dengan port entry exa pada operator ini dan output exa dengan connector res pada work area.



b. Create Association Rules

Hubungkan output fre pada operator FP-Growth dengan masukan ite pada operator ini, output rul



Setelah diRun dapat dilihat hasil-hasil aturan asosiasi sebagai berikut:

a. Frequent Item Set (FP-Growth)

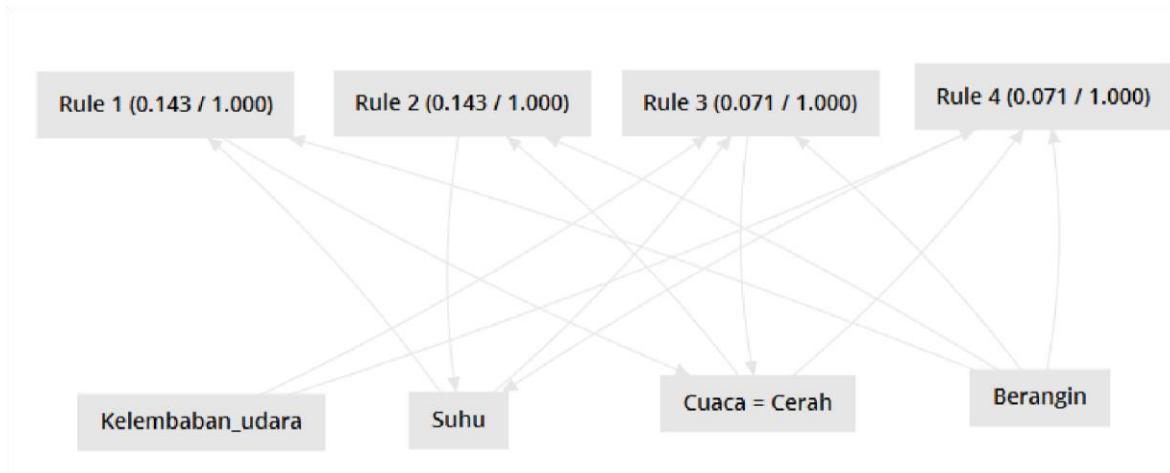
No. of Sets: 26	Size	Support	Item 1	Item 2	Item 3	Item 4
Total Max. Size: 4	1	0.500	Kelembaban_ud...			
Min. Size: 1	1	0.429	Berangin			
Max. Size: 4	1	0.429	Suhu			
Contains Item:	1	0.357	Cuaca = Cerah			
	1	0.357	Cuaca = Hujan			
	1	0.286	Cuaca = Mendu...			
	2	0.214	Kelembaban_ud...	Berangin		
	2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

b. Association Rules (Create Association Rules)

i. Table View

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convic
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	Kelembaban_udara, Berangin, Suhu	Cuaca = Cerah	0.071	1	1	-0.071	0.046	2.800	$\infty$
4	Kelembaban_udara, Berangin, Cuac...	Suhu	0.071	1	1	-0.071	0.041	2.333	$\infty$

ii. Graph View

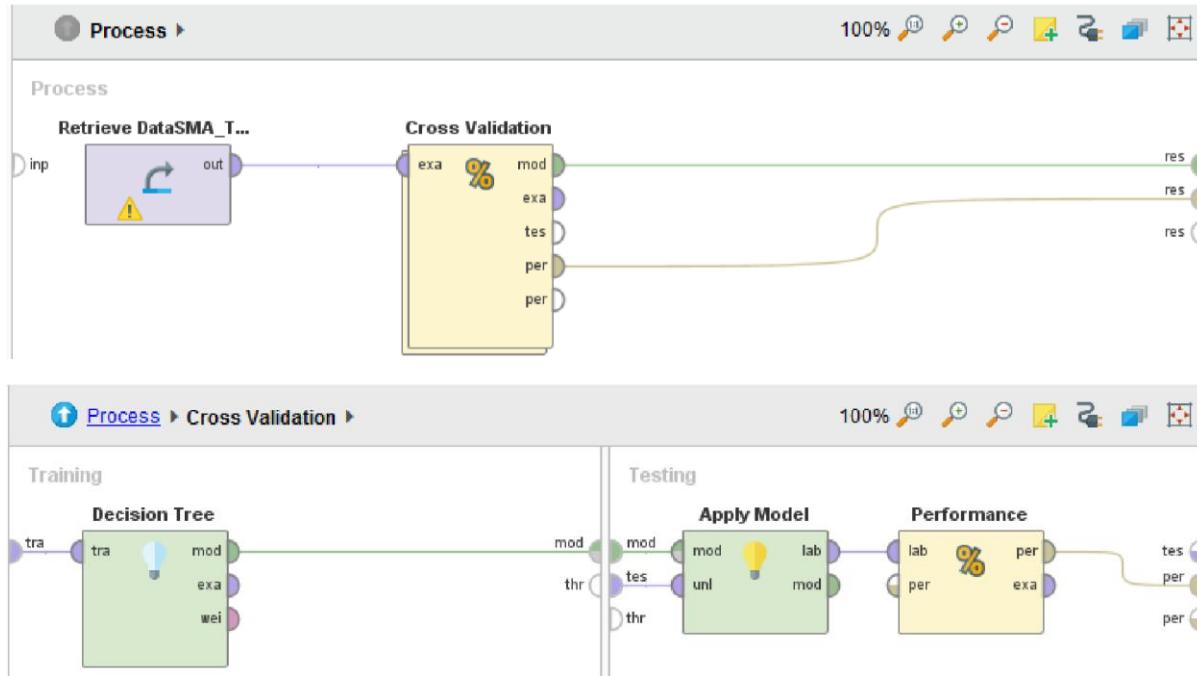


### c. ExampleSet, Chart View

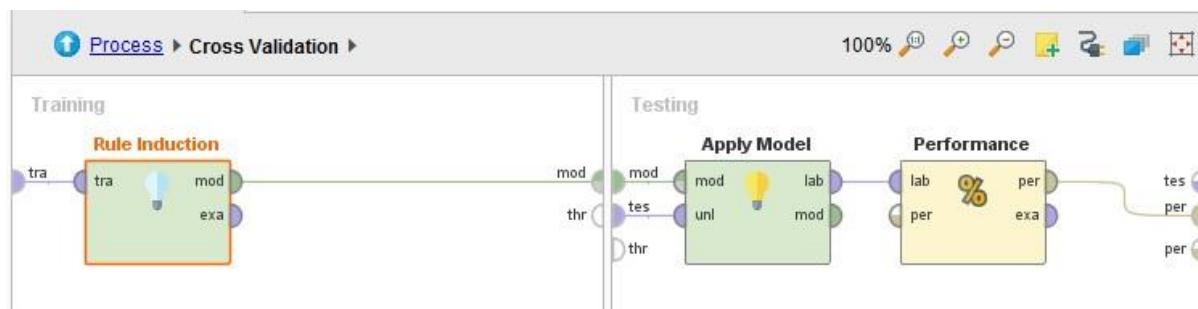


## Induksi Aturan Data Cuaca

Menggunakan data Lama Studi (format excel) pada Tugas Modul 6 Soal No. 1



Klik kanan operator Decision Tree pada area Training, pilih Replace Order menjadi Rule Induction



Setelah diRun diperoleh sebuah induksi aturan dari data training yang diberikan disebut sebagai RuleModel (Rule Induction).

## RuleModel

```
if Rerata_Sekolah > 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)

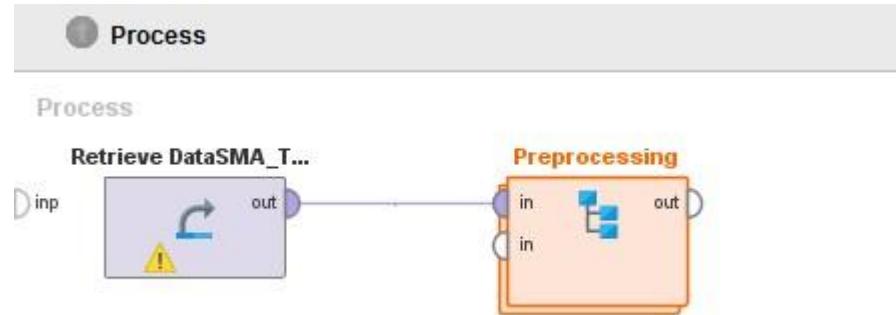
correct: 17 out of 19 training examples.
```

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

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%	

### Aturan Asosiasi Data Cuaca

Gunakan DataCuaca\_Training dan drag dari repository ke area Process View tambahkan operator Utility lalu Subprocess ke dalam area. Ubah nama operator ini menjadi Preprocessing.



Klik ganda operator Preprocessing sehingga masuk pada area Nested Chain. Tambahkan operator operator berikut:

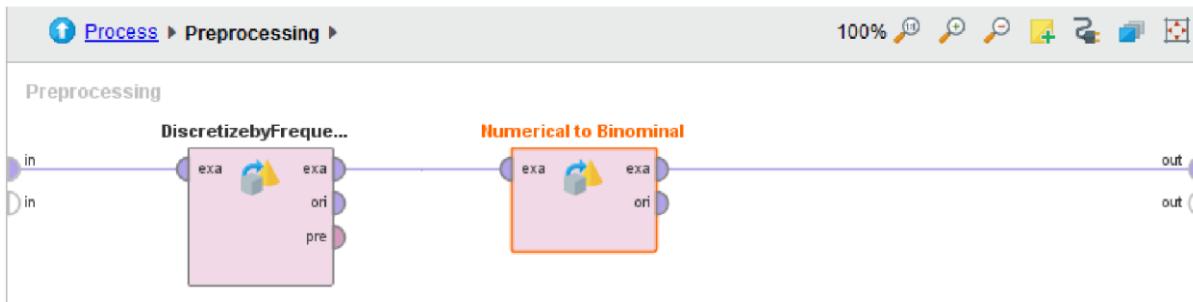
- c. Discretize by Frequency

Biarkan nilai parameter number of bins = 2



#### d. Nominal to Binominal

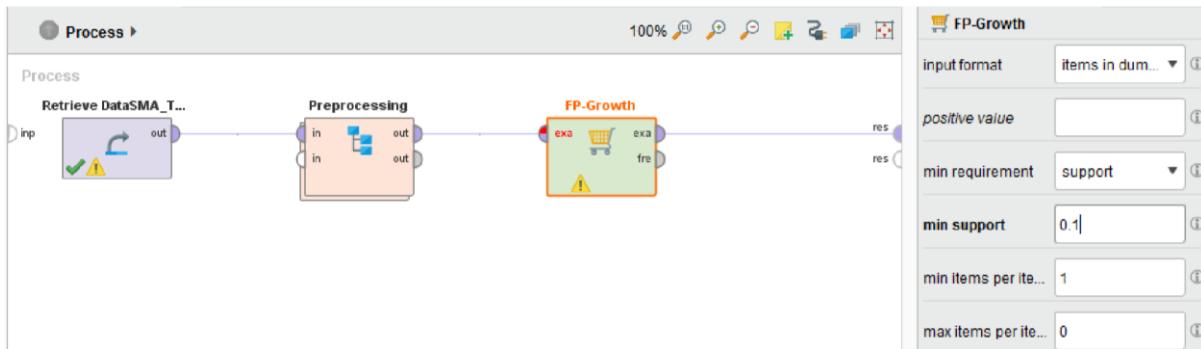
Hubungkan output operator DiscretizebyFrequency dengan masukan examination pada operator ini, dan output examination operator ini dengan panel out.



Kembali ke main process, tambahkan 2 operator:

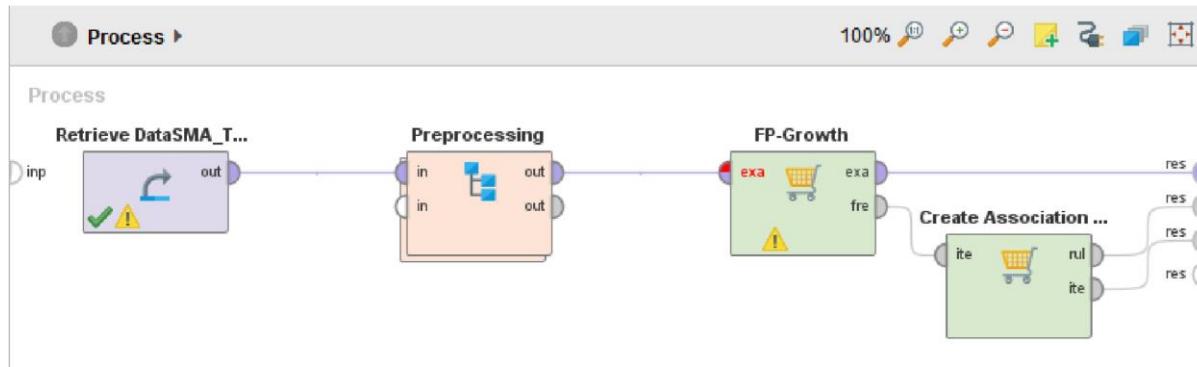
#### c. FP-Growth

Ubah nilai parameter pada min support = 0.1, hubungkan port out operator Preprocessing dengan port entry exa pada operator ini dan output exa dengan connector res pada work area.



#### d. Create Association Rules

Hubungkan output fre pada operator FP-Growth dengan masukan ite pada operator ini, output rul



Setelah diRun dapat dilihat hasil-hasil aturan asosiasi sebagai berikut:

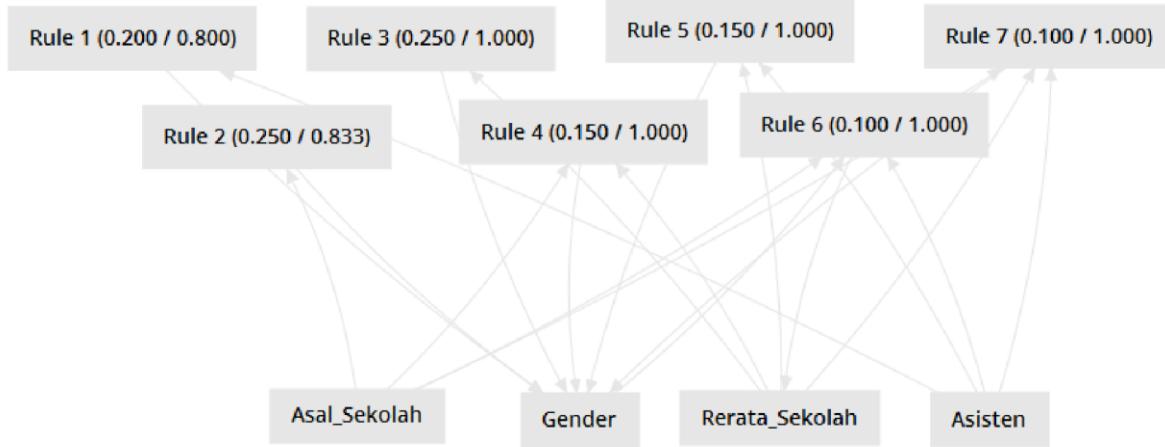
d. Frequent Item Set (FP-Growth)

No. of Sets: 15	Size	Support	Item 1	Item 2	Item 3	Item 4
Total Max. Size: 4	1	0.750	Gender			
Min. Size: 1	1	0.300	Asal_Sekolah			
Max. Size: 4	1	0.250	Asisten			
Contains Item:	1	0.250	Rerata_Sekolah			
	2	0.250	Gender	Asal_Sekolah		
	2	0.200	Gender	Asisten		
	2	0.250	Gender	Rerata_Sekolah		
	2	0.150	Asal_Sekolah	Asisten		
	2	0.150	Asal_Sekolah	Rerata_Sekolah		
	2	0.150	Asisten	Rerata_Sekolah		
	3	0.100	Gender	Asal_Sekolah	Asisten	
	3	0.150	Gender	Asal_Sekolah	Rerata_Sekolah	
	3	0.150	Gender	Asisten	Rerata_Sekolah	
	3	0.100	Asal Sekolah	Asisten	Rerata Sekolah	

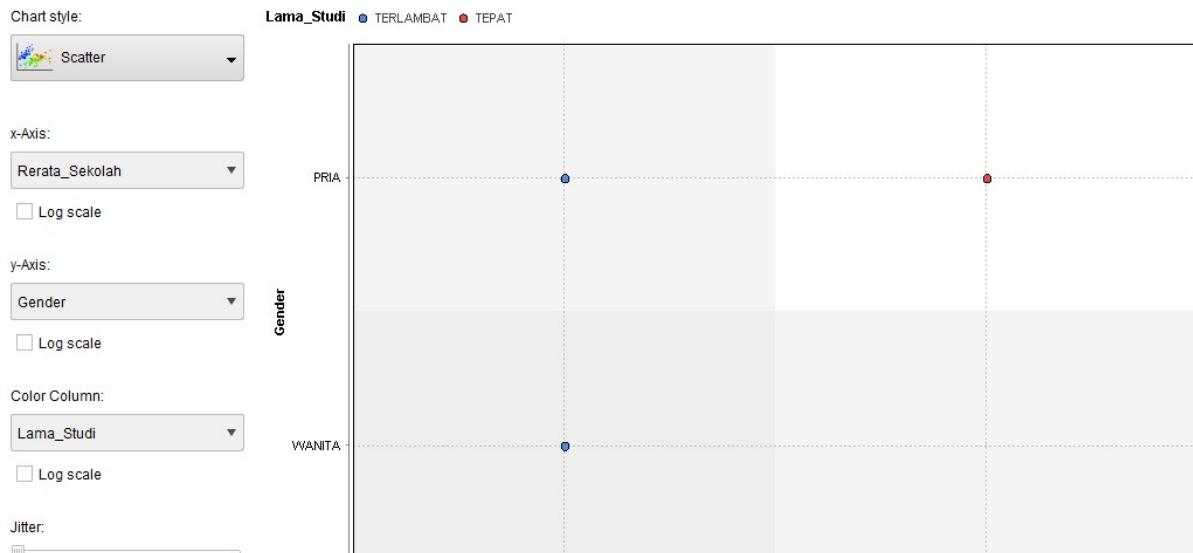
e. Association Rules (Create Association Rules) iii. Table View

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
2	Asal_Sekolah	Gender	0.250	0.833	0.962	-0.350	0.025	1.111	1.500
3	Rerata_Sekolah	Gender	0.250	1	1	-0.250	0.062	1.333	$\infty$
4	Asal_Sekolah, Rerata_Sekolah	Gender	0.150	1	1	-0.150	0.038	1.333	$\infty$
5	Asisten, Rerata_Sekolah	Gender	0.150	1	1	-0.150	0.038	1.333	$\infty$
6	Gender, Asal_Sekolah, Asisten	Rerata_Sekolah	0.100	1	1	-0.100	0.075	4	$\infty$
7	Asal_Sekolah, Asisten, Rerata_Sekolah	Gender	0.100	1	1	-0.100	0.025	1.333	$\infty$

iv. Graph View



#### f. ExampleSet, Chart View



#### b. Numbers of bins = 3

##### a. Frequent Item Set (FP-Growth)

No. of Sets: 7  
Total Max. Size: 3  
Min. Size: 1  
Max. Size: 3  
Contains Item:   
Update View

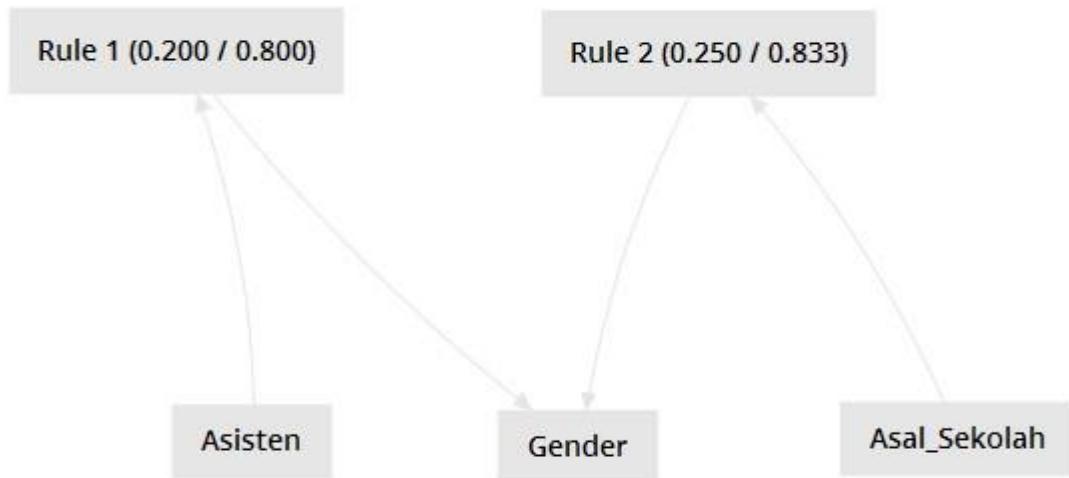
Size	Support	Item 1	Item 2	Item 3
1	0.750	Gender		
1	0.300	Asal_Sekolah		
1	0.250	Asisten		
2	0.250	Gender	Asal_Sekolah	
2	0.200	Gender	Asisten	
2	0.150	Asal_Sekolah	Asisten	
3	0.100	Gender	Asal_Sekolah	Asisten

#### b. Association Rules (Create Association Rules)

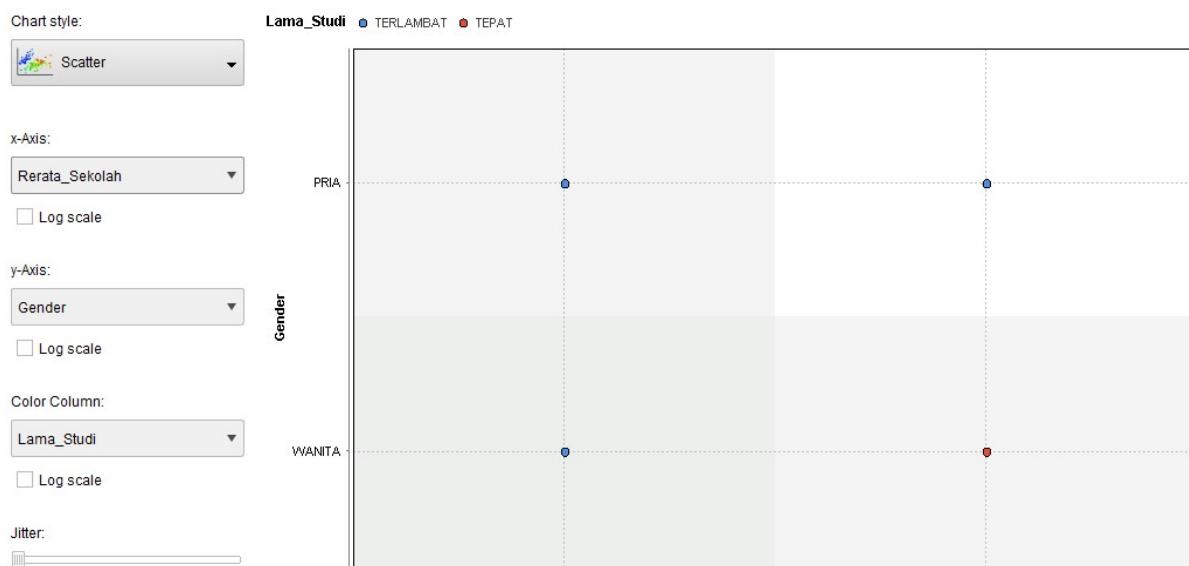
##### i. Table View

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
2	Asal_Sekolah	Gender	0.250	0.833	0.962	-0.350	0.025	1.111	1.500

ii. Graph View



c. ExampleSet, Chart View



#### 12.4.1. Mencari Nilai t-hitung dan Model Regresi Linier Tabel

Data Siswa:

	A	B	C	D
1	NO_SISWA	NAMA	LAMA BELAJAR (JAM)	NILAI
2	S-101	JOKO	15	783
3	S-102	AGUS	18	877
4	S-103	SUSI	7	505
5	S-104	DYAH	9	860
6	S-105	WATI	15	968
7	S-106	IKA	17	793
8	S-107	EKO	10	752
9	S-108	YANTO	5	571
10	S-109	WAWAN	8	667
11	S-110	MAHMUD	15	723
12				

Menggunakan file Tabel\_LamaBelajardanNilaiUjian.xls ke RapidMiner yang akan digunakan sebagai regresi Linear. Beri nama Data\_LamaBelajardanNilaiUjian lalu masukkan repositori.

Import Data - Format your columns.

X

Format your columns.

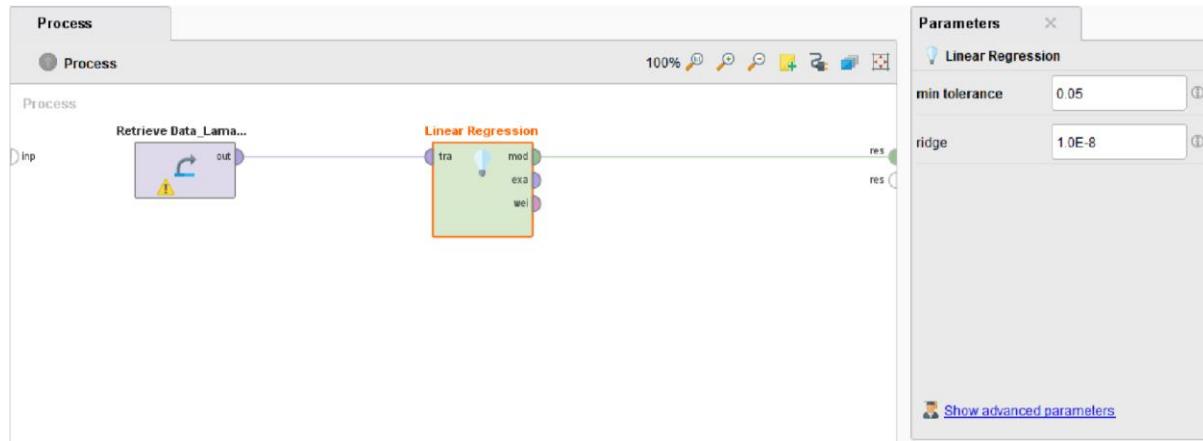
Replace errors with missing values ⓘ

No_Siswa polynomial id	Nama polynomial	Lama Belajar (Jam) integer	Nilai integer label
1	JOKO	15	783
2	AGUS	18	877
3	SUSI	7	505
4	DYAH	9	860
5	WATI	15	968
6	IKA	17	793
7	EKO	10	752
8	YANTO	5	571
9	WAWAN	8	667
10	MAHMUD	15	723

no problems.

Previous Next Cancel

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%). Kemudian jalankan proses dengan menekan tombol Run.



Hasil proses regresi linier:

- a) Table View (Mencari besarnya nilai t-hitung)

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Lama Belajar (J...	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 * Lama Belajar (Jam)
+ 492.769
```

### 12.4.2. Mencari Nilai t dan Model Regresi Linier Menggunakan RapidMiner Tabel

Data Siswa:

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

Menggunakan file Tabel\_PrediksiNilaiUjian ke RapidMiner yang akan digunakan sebagai regresi Linear. Beri nama Data\_PrediksiNilaiUjian dan masukkan pada repositori.

Import Data - Format your columns.

**Format your columns.**

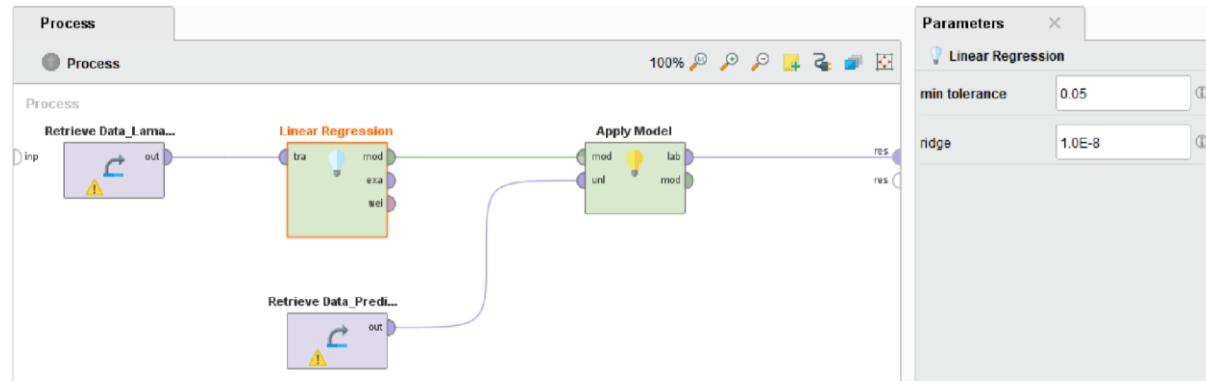
Replace errors with missing values  ⓘ

No_Siswa polynominal id	Nama polynominal	Lama Belajar (Jam) 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

no problems.

Previous Next Cancel

Memasukkan 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%). Kemudian jalankan proses dengan menekan tombol Run.

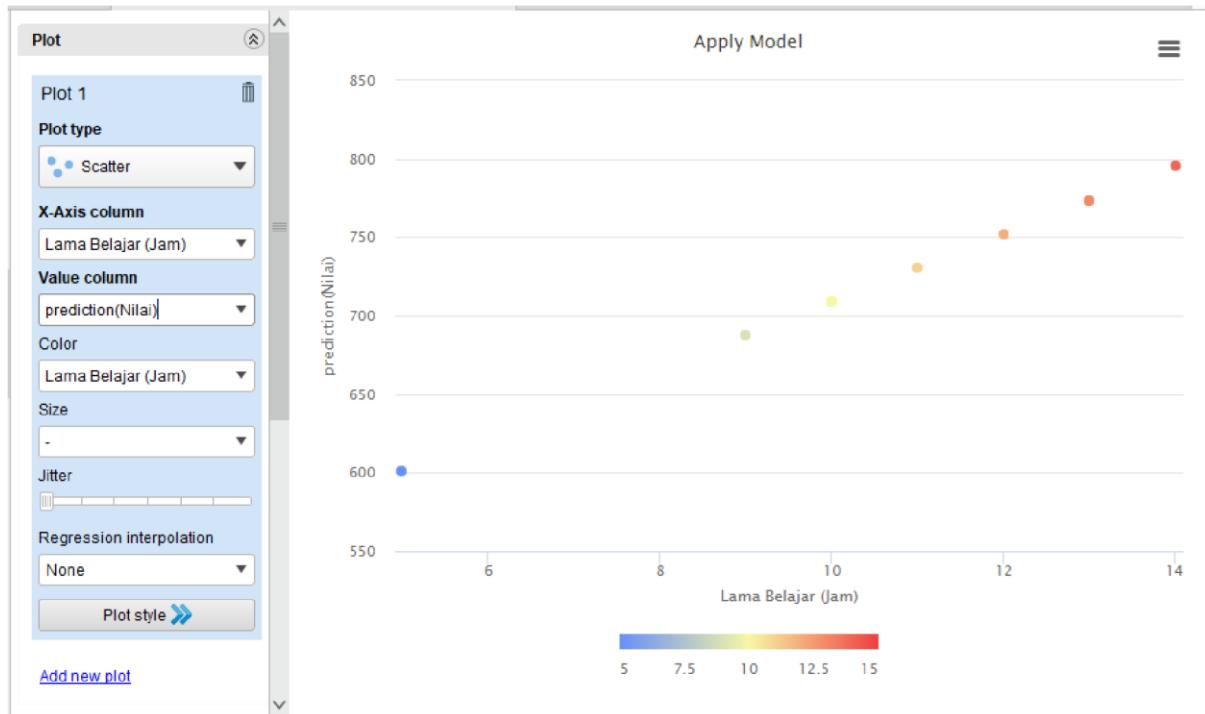


Hasil proses regresi linier:

- 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)



#### 12.4.3. Pembuktian Model Regresi

Tahap ini kita 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 **X1** ke dalam model regresi :

1. No. SISWA = S-112, NAMA = SANTI, memiliki nilai X1 = 13 Jam. Sehingga,  $Y = (21,608 \times 13) + 492,769 = 773,672$
2. No. SISWA = S-113, NAMA = DIAN, memiliki nilai X1 = 14 Jam. Sehingga,  $Y = (21,608 \times 14) + 492,769 = 795,28$
3. No. SISWA = S-114, NAMA = DANI, memiliki nilai X1 = 11 Jam. Sehingga,  $Y = (21,608 \times 11) + 492,769 = 730,456$
4. No. SISWA = S-116, NAMA = BAYU, memiliki nilai X1 = 13 Jam. Sehingga,  $Y = (21,608 \times 13) + 492,769 = 773,672$
5. No. SISWA = S-117, NAMA = RISA, memiliki nilai X1 = 9 Jam. Sehingga,  $Y = (21,608 \times 9) + 492,769 = 687,24$
6. No. SISWA = S-118, NAMA = RANI, memiliki nilai X1 = 10 Jam. Sehingga,  $Y = (21,608 \times 10) + 492,769 = 708,848$
7. No. SISWA = S-119, NAMA = YANI, memiliki nilai X1 = 10 Jam. Sehingga,  $Y = (21,608 \times 10) + 492,769 = 708,848$

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

1. Tabel Hasil Survei 15 Kepala Keluarga

	A	B	C	D
1	NO. RESPONDEN	PENDAPATAN (RUPIAH)	JUMLAH ANGGOTA KELUARGA	DAYA BELI RUPIAH (RUPIAH)
2	1	1.000.000,00	6	834.000,00
3	2	1.400.000,00	7	1.200.000,00
4	3	200.000,00	3	134.000,00
5	4	1.400.000,00	6	1.167.000,00
6	5	500.000,00	3	334.000,00
7	6	1.700.000,00	5	1.360.000,00
8	7	400.000,00	3	267.000,00
9	8	1.900.000,00	5	1.520.000,00
10	9	300.000,00	3	200.000,00
11	10	500.000,00	4	375.000,00
12	11	700.000,00	7	600.000,00
13	12	1.900.000,00	3	1.267.000,00
14	13	800.000,00	4	600.000,00
15	14	1.500.000,00	4	1.125.000,00
16	15	1.300.000,00	7	1.115.000,00

2. Buatlah proses Regresi Linier Sederhana menggunakan RapidMiner dengan ketentuan sebagai berikut.
- Variable bebas (X) = Pendapatan (X1), Jumlah Anggota Keluarga (X2)

- b) Variable terikat (Y) = Daya Beli  
c) Toleransi yang digunakan = 5%

Pendapatan (X1)

Import Data - Format your columns.

**Format your columns.**

Replace errors with missing values ⓘ

NO. RESPONDEN	polynomial id	PENDAPATAN (RUPIAH)	integer	JUMLAH ANGGOTA K...	integer	DAYA BELI (RUPIAH)	integer label
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

Process

Process

Process

Retrieve Data\_tugas...

Linear Regression

Parameters

Linear Regression

min tolerance: 0.05

ridge: 1.0E-8

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 \* PENDAPATAN (RUPIAH)  
- 7467.579

## Jumlah Anggota Keluarga (X2)

Import Data - Format your columns.

**Format your columns.**

Replace errors with missing values [①](#)

NO. RESPONDEN polynomial id	PENDAPATAN (RUPIAH) integer	JUMLAH ANGGOTA K... integer	DAYA BELI (RUPIAH) integer label
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

✓ no problems.

[Previous](#) [Next](#) [Cancel](#)

**Process**

**Parameters**

Linear Regression

min tolerance: 0.05

ridge: 1.0E-8

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 X1 dan X2 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 X1 dan X2 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 :

A	B	C	
1	NO.RESPONDEN	PENDAPATAN (RUPIAH)	JUMLAH ANGGOTA KELUARGA
2	1	900.000	5
3	2	800.000	3
4	3	500.000	2
5	4	1.900.000	6
6	5	600.000	2
7	6	800.000	5
8	7	1.000.000	6
9	8	1.100.000	4
10	9	1.000.000	4
11	10	500.000	3
12			

- 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!

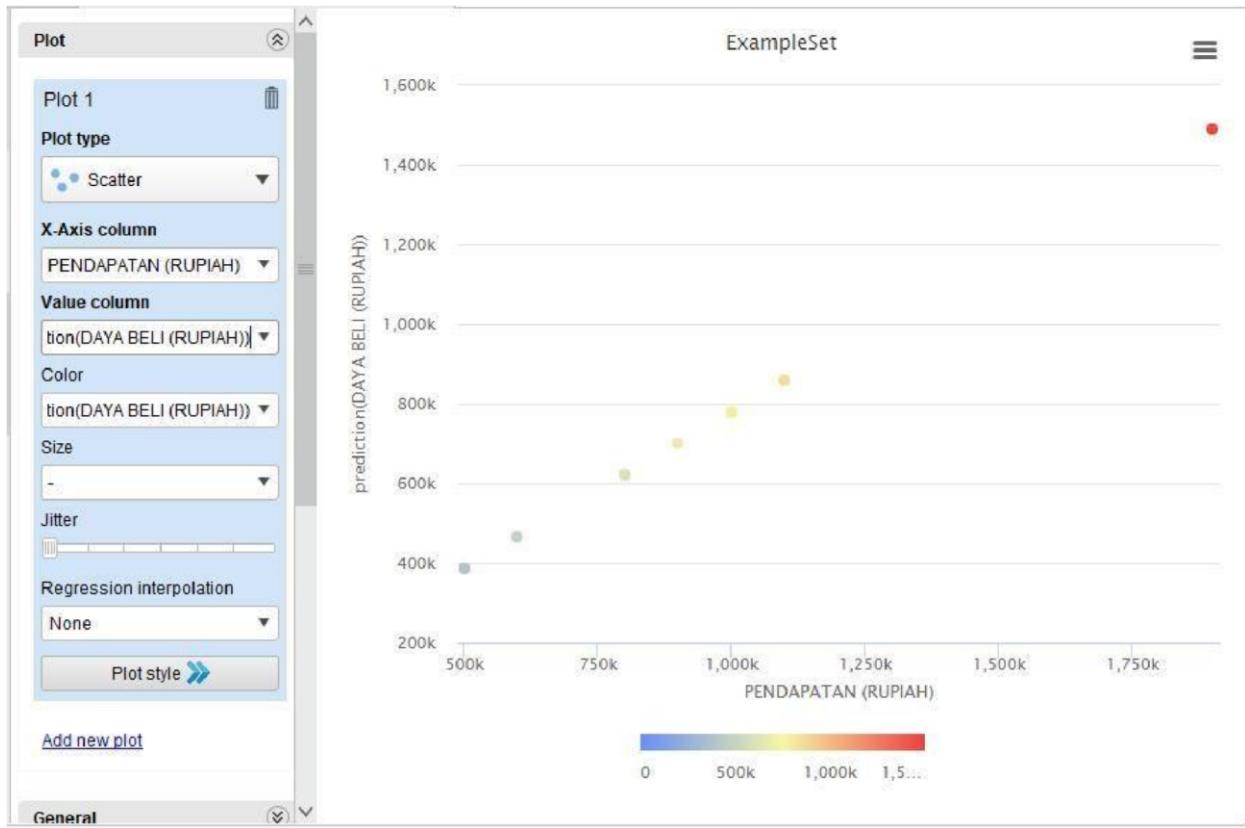
Pendapatan (X1)

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 (X2)

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 :
- 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))

