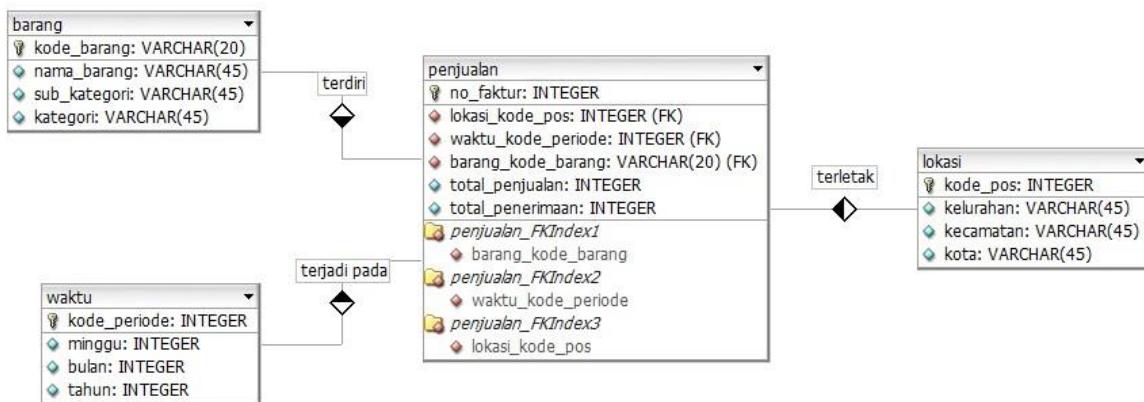


Nama : Yussynta Dewi Aprilya Putri

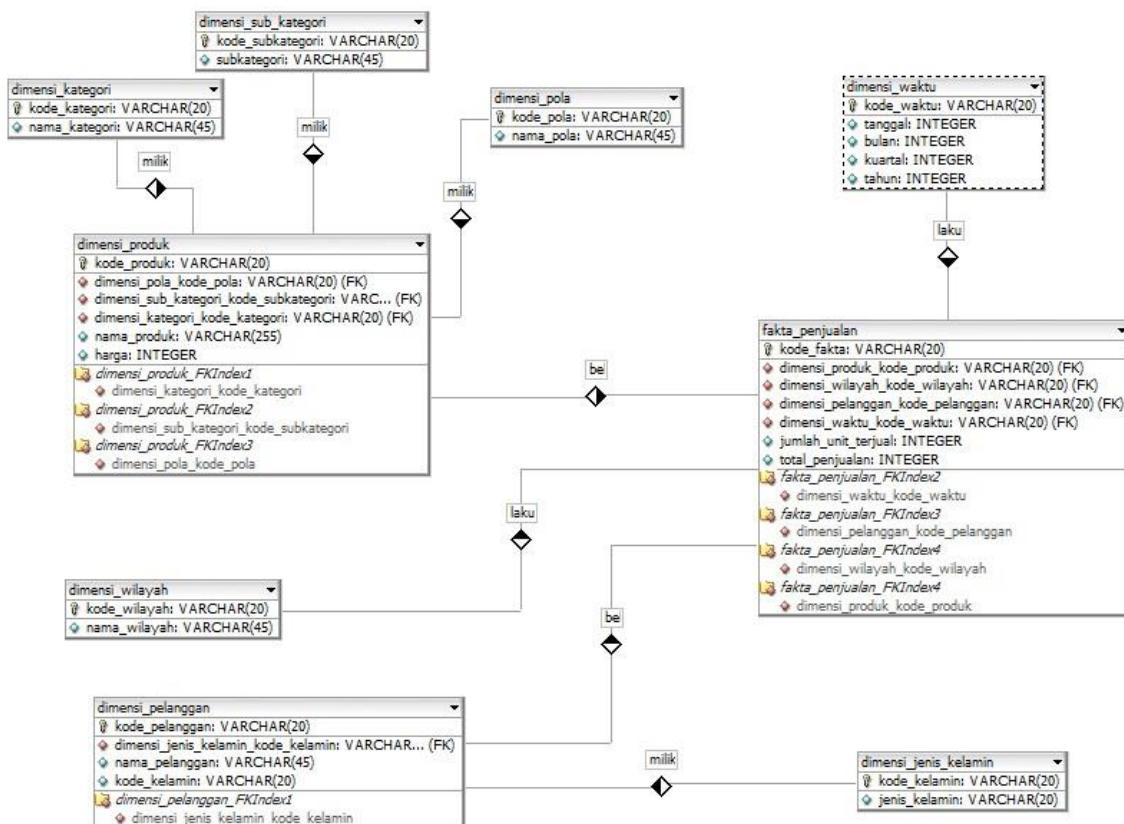
NIM : L200170084

Kelas : D

MODUL01



Tugas



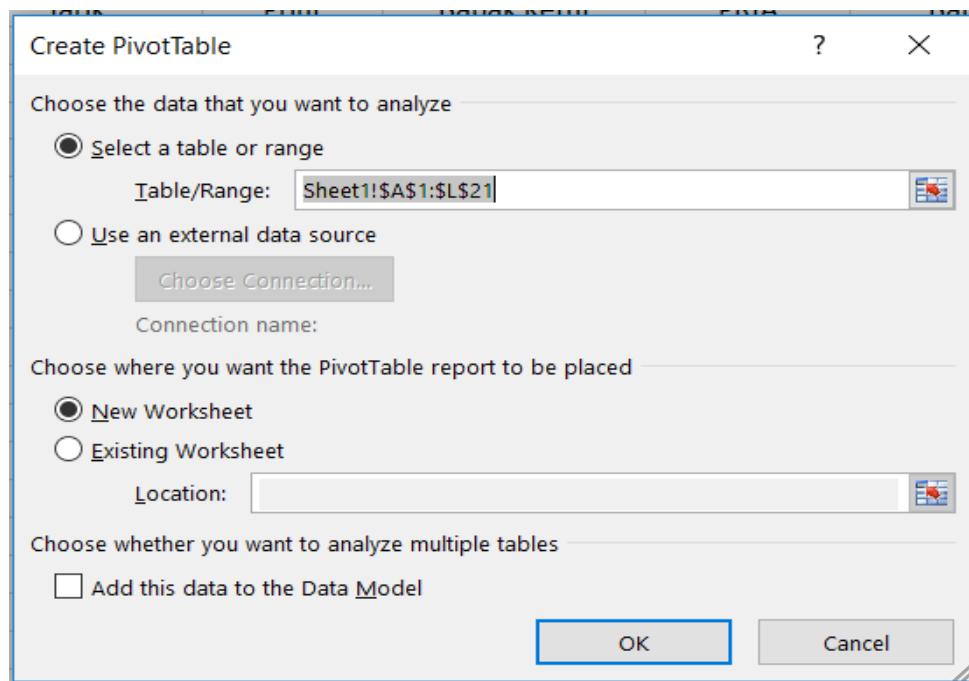
MODUL05

Kegiatan 1: Membuat Pivot Table

1. Buka sheet Fact_Table, dimana datanya terlihat seperti pada gambar berikut :

1	C	D	E	F	G	H	I	J	K	L
	tahun	nama_produk	nama_kategori	nama_subkategori	nama_pola	nama_pelanggan	jenis_kelamin	nama_wilayah	jumlah	harga
2	2011	Jarik Standar Print Sogan	Standar	Jarik	Print	Bapak Ketut	PRIA	Bali	2	225000
3	2012	Kaos Batik Cap Lukis	Batik	Kaos	Cap	Ibu Harini	WANITA	Jawa Timur	14	30000
4	2012	Jarik Standar Tulis Sarimbit	Standar	Jarik	Tulis	Ibu Harini	WANITA	Jawa Timur	4	40000
5	2011	Hem Katun Print kaung	Katun	Hem	Print	Ibu Harini	WANITA	Jawa Timur	3	70000
6	2012	Batik Standar CapTumpal	Standar	Batik	Cap	Bapak Heru	PRIA	Jawa Timur	1	150000
7	2012	Hem Katun Print Kelengan	Katun	Hem	Print	Bapak Totok	PRIA	Jawa Timur	3	299000
8	2011	Bolero Standar Cap Sido Mukti	Standar	Bolero	Cap	Ibu Hatamah	WANITA	Jawa Timur	1	225000
9	2011	Sarimbit Standar Print Lukis	Standar	Sarimbit	Print	Ibu Hatamah	WANITA	Jawa Timur	1	150000
10	2011	Kaos Katon Print Bola	Katun	Kaos	Print	Bapak Imron	PRIA	Jawa Barat	1	60000
11	2012	Celana Standar Cap Warna	Standar	Celana	Cap	Ibu Hadi Sukarni	WANITA	Jawa Barat	17	55000
12	2010	Celana Standar Cap Warna	Standar	Celana	Print	Ibu Hadi Sukarni	WANITA	Jawa Barat	17	55000
13	2011	Bahan Standar Cap Lasem	Standar	Bahan	Cap	Ibu Siti Arya	WANITA	Jawa Barat	8	120000
14	2012	Rok Batik Print kombinasi	Batik	Rok	Print	Ibu Siti Arya	WANITA	Jawa Barat	1	225000
15	2012	Jam Standar Print Lukis	Standar	Jam	Print	Ibu Siti Arya	WANITA	Jawa Barat	44	80000
16	2012	Hem Standar Cap Tumpal	Standar	Hem	Cap	Ibu Aini Kasmaji	WANITA	Jawa Tengah	1	100000
17	2012	Bahan Lawasan Tulis Toilet	Lawasan	Bahan	Tulis	Ibu Niken	WANITA	Jawa Tengah	1	130000
18	2011	Hem Standar Tulis Madura	Standar	Hem	Tulis	Ibu Atik	WANITA	Jawa Tengah	5	550000
19	2012	Bahan Standar Cap Garis	Standar	Bahan	Cap	Ibu Tyas	WANITA	Jawa Tengah	7	135000
20	2010	Bahan Beludru Cap Mahkota	Beludru	Bahan	Cap	Ibu Tyas	WANITA	Jawa Tengah	1	500000
21	2010	Hem Sutra Print Rama	Sutra	Hem	Print	Ibu Tyas	WANITA	Jawa Tengah	5	100000

2. Pilih range data A1:L21 atau tekan tombol CTRL + SHIFT + *.
3. Klik tab Insert pada Ribbon, pilih menu PivotTable | Insert PivotTable.
4. Pada dialog Create PivotTable yang muncul, pilih New Worksheet, klik tombol OK.



5. Sheet baru akan muncul disertai suatu kotak / placeholder PivotTable (PivotTable Box).

The screenshot shows the Microsoft Excel interface with a new sheet named 'Sheet2'. A PivotTable placeholder is present in cell A3. The PivotTable Fields pane is open on the right side, showing a list of fields to add to the report. The 'ROWS' area has 'nama_sub...' selected, and the 'VALUES' area has 'Sum of ju...' selected.

6. Susunlah layout field dengan urutan berikut :
- Field `nama_subkategori` ke kotak Row Labels.
 - Field `tahun` ke kotak Column Labels.
 - Field `jumlah` ke kotak Values.

The screenshot shows the PivotTable Fields pane with the following layout configuration:

- ROWS:** `nama_sub...`
- COLUMNS:** `tahun`
- VALUES:** `Sum of ju...`

7. Perhatikan hasil pengaturan ini pada area PivotTable. Area ini akan berisi suatu tabel dengan grouping field **nama_subkategori** pada bagian baris, field **tahun** pada kolom. Sedangkan nilai total jumlah_unit ditempatkan pada cell-cell hasil perpotongan item grouping baris dan kolom tersebut.

	Sum of jumlah	Column Labels			
	Row Labels	2010	2011	2012	Grand Total
5	Bahan	1	8	8	17
6	Batik			1	1
7	Bolero		1		1
8	Celana	17		17	34
9	Hem	5	8	4	17
10	Jam			44	44
11	Jarik		2	4	6
12	Kaos	1	14		15
13	Rok			1	1
14	Sarimbit	1			1
15	Grand Total	23	21	93	137

Salah satu contoh perpotongan adalah total jumlah yang terjual dengan kategori **Jam** selama tahun **2012**, adalah sebesar **44** unit.

Kegiatan 2 : Menambahkan Tipe Summary Baru

1. Masih bekerja menggunakan file “Fakta_Penjualan.xls” pada **Kegiatan 1** dengan Sheet1 PivotTable.
2. Tambahkan field **jumlah** kembali ke kotak **Value** dengan cara klik dan drag, sehingga muncul field baru dengan nama **Sum of jumlah2**.

PivotTable Fields

Choose fields to add to report:

- kuartal
- tahun**
- nama_produk
- nama_kategori
- nama_subkategori**
- nama_pola
- nama_pelanggan
- jenis_kelamin
- nama_wilayah
- jumlah**
- harga

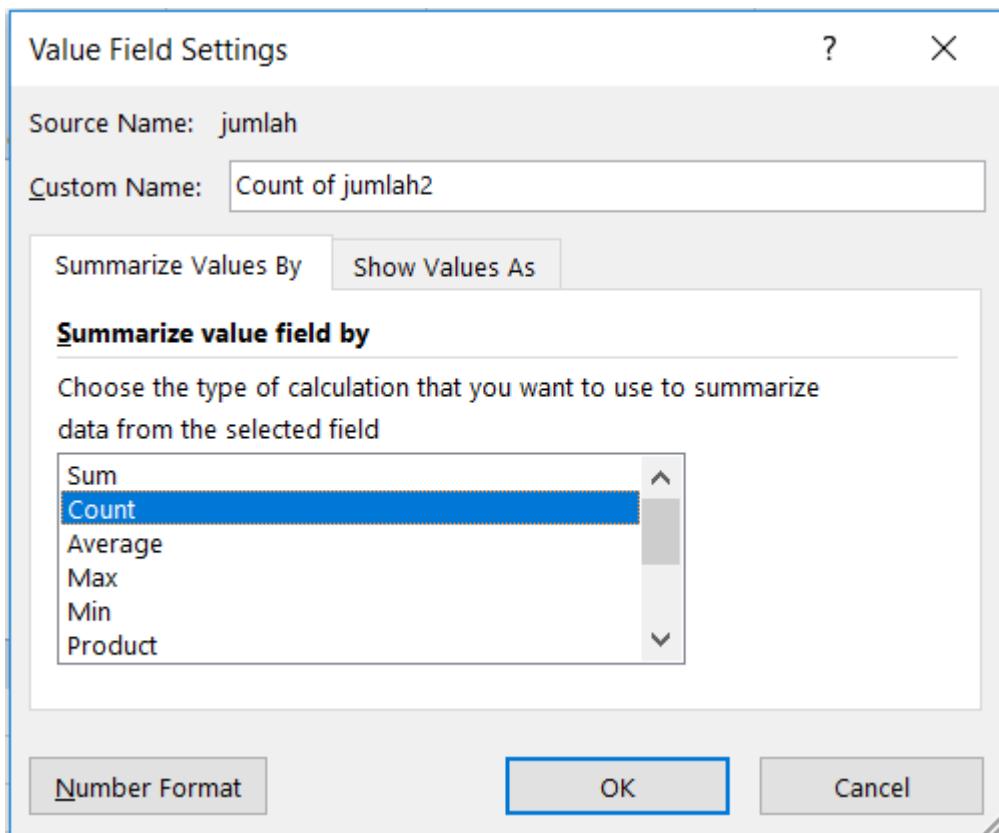
Drag fields between areas below:

FILTERS	COLUMNS
	tahun Σ Values
ROWS	VALUES
nama_subkate...	Sum of jumlah Sum of jumlah2

3. Dan akan diperoleh tambahan satu kolom perhitungan baru yang sama dengan hasil sebelumnya pada masing-masing tahun.

3	Column Labels	2010	2011	2012	Total Sum of jumlah	Total Sum of jumlah2
4	Row Labels	Sum of jumlah	Sum of jumlah2	Sum of jumlah	Sum of jumlah2	Sum of jumlah
5	Bahan	1	1	8	8	8
6	Batik				1	1
7	Bolero		1	1		1
8	Celana	17	17		17	34
9	Hem	5	5	8	4	17
10	Jam				44	44
11	Jarik		2	2	4	6
12	Kaos		1	1	14	15
13	Rok				1	1
14	Sarimbit		1	1		1
15	Grand Total	23	23	21	93	137
16						

4. Kembali ke area Values, dan klik tombol panah ke bawah pada field Sum of jumlah2. Pilih item Value Field Settings.
5. Pada dialog Value Field Settings, ubah Sum menjadi Count. Perhatikan nama field akan berubah menjadi Count of jumlah2.



6. Klik tombol **OK**
7. Pada area PivotTable, didapatkan dua summary yaitu:
 - a. nilai jumlah unit penjualan yang terjadi (**sum**).
 - b. jumlah transaksi yang terjadi (**count**).

	Column Labels		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	17
Batik					1	1	1
Bolero			1	1			1
Celana	17	1			17	1	34
Hem	5	1	8	2	4	2	17
Jam					44	1	44
Jarik		2	1	4	1		6
Kaos		1	1	14	1		15
Rok					1	1	1
Sarimbit		1	1		1		1
Grand Total	23	3	21	7	93	10	137
							20

Kegiatan 3 : Calculated Field dan Calculated Item di Pivot Table

1. Buka Sheet1 dalam file Fakta_Penjualan.xls, dan letakkan kursor ke area PivotTable.
2. Pada menu ribbon PivotTable Tools | Options, klik button **Field, Item & Sets** dan pilih Calculated Field.
3. Pada kotak dialog **Insert Calculated Field** yang muncul, masukkan nilai berikut kemudian klik tombol **OK**.
 - a. Name : Pendapatan
 - b. Formula : = jumlah * harga (Pilih field **jumlah** kemudian klik Insert Field kemudian ketikkan tanda "*" dan masukkan field **harga**)



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

		2012		Total Sum of jumlah	Total Count of jumlah2	Total Sum of Pendapatan
	Row Labels	Sum of jumlah	Count of jumlah2	Sum of Pendapatan		
6	Bahan	8	2	2120000	17	4 15045000
7	Batik	1	1	150000	1	1 150000
8	Bolero			0	1	1 225000
9	Celana	17	1	935000	34	2 3740000
10	Hem	4	2	1596000	17	5 19023000
11	Jam	44	1	3520000	44	1 3520000
12	Jarik	4	1	160000	6	2 1590000
13	Kaos	14	1	420000	15	2 1350000
14	Rok	1	1	225000	1	1 225000
15	Sarimbit			0	1	1 150000
16	Grand Total	93	10	115692000	137	20 451963000

Kegiatan 4 : Operasi Roll Up dan Drill Down

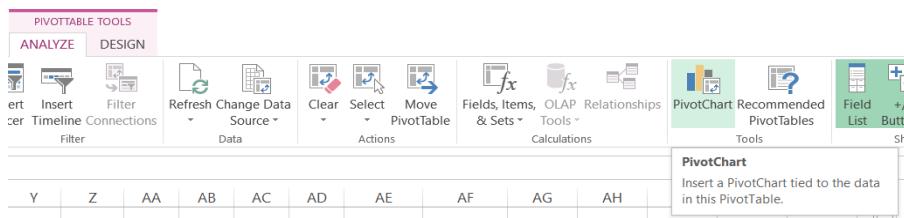
1. Buka Sheet1 (hasil pivot table) dan letakkan kursor pada area pivot table.
2. Pada kotak PivotTable Field List, hilangkan tanda cek pada field **jumlah** (field ini sementara tidak digunakan), dan beri tanda cek pada field (kolom) yang akan ditampilkan ke dalam cube.
3. Beri tanda cek dan letakkan field-field berikut pada kotak Row Labels atau Column Labels sesuai dengan kebutuhan tampilan cube. Urutan field dalam kotak ini menentukan urutan rincian kategori data. Field yang terletak pada urutan teratas merupakan field dengan kategori paling umum, sedangkan field yang terletak pada urutan terbawah adalah field dengan kategori paling spesifik (paling rinci).
4. Misalkan pada Row Labels akan ditampilkan data berdasarkan urutan **nama_kategori**, **nama_subkategori**, dan **nama_produk**. Beri tanda cek pada field tersebut (bisa drag and drop) dan letakkan pada kotak Row Labels.

	Sum of Pendapatan						2012 Total									
	2012															
	1		1 Total		2		2 Total		3		3 Total		4		4 Total	
	Row Labels	1	2	4	5	6	9	12	13	14	15	16	17	18	19	
3	Bahan	0	0	0	945000	0	130000	2120000	0	0	0	0	0	2120000	0	
4	Beludru	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	Bahan Beludru Cap Mahkota	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	Lawasan	0	0	0	0	0	130000	130000	0	0	0	0	0	130000	0	
7	Bahan Lawasan Tulis Toilet	0	0	0	0	0	130000	130000	0	0	0	0	0	130000	0	
8	Standar	0	0	0	945000	0	0	945000	0	0	0	0	0	945000	0	
9	Bahan Standar Cap Garis	0	0	0	945000	0	0	945000	0	0	0	0	0	945000	0	
10	Bahan Standar Cap Lasem	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	Batik	0	0	0	0	0	0	150000	150000	0	0	0	0	150000	0	
12	Bolero	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	Celana	0	935000	935000	0	0	0	0	0	0	0	0	0	935000	0	
14	Hem	0	0	0	0	897000	0	897000	100000	100000	0	0	0	1596000	0	
15	Jam	3520000	0	3520000	0	0	0	0	0	0	0	0	0	3520000	0	
16	Jarik	0	0	0	160000	0	0	160000	0	0	0	0	0	160000	0	
17	Kaos	420000	0	420000	0	0	0	0	0	0	0	0	0	420000	0	
18	Rok	0	0	0	0	0	0	0	0	0	0	0	0	225000	225000	
19	Sarimbit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
20	Grand Total	6380000	935000	12375000	1925000	897000	130000	9060000	500000	500000	225000	225000	115692000			

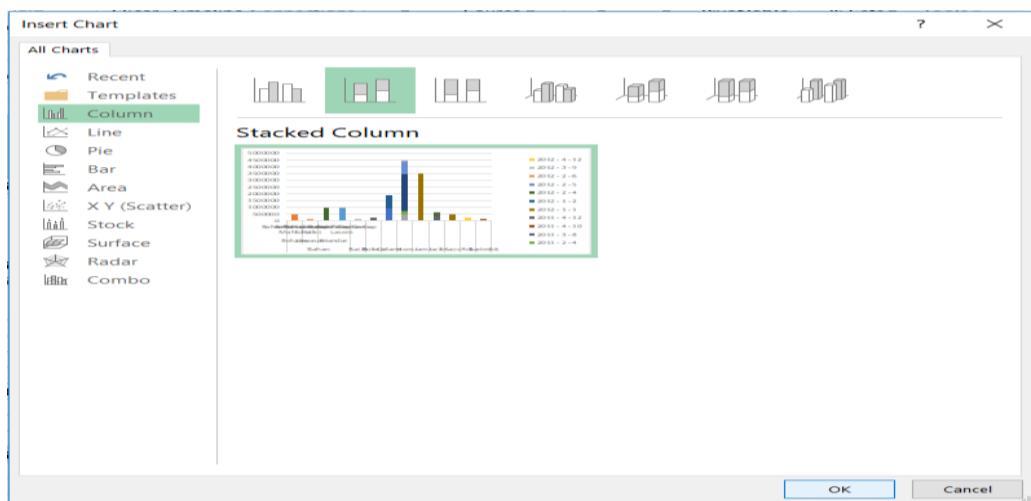
5. Klik tanda [-] untuk melakukan operasi Roll Up dan klik tanda [+] untuk melakukan operasi Drill Down.

Kegiatan 5 : Menggunakan Pivot Chart

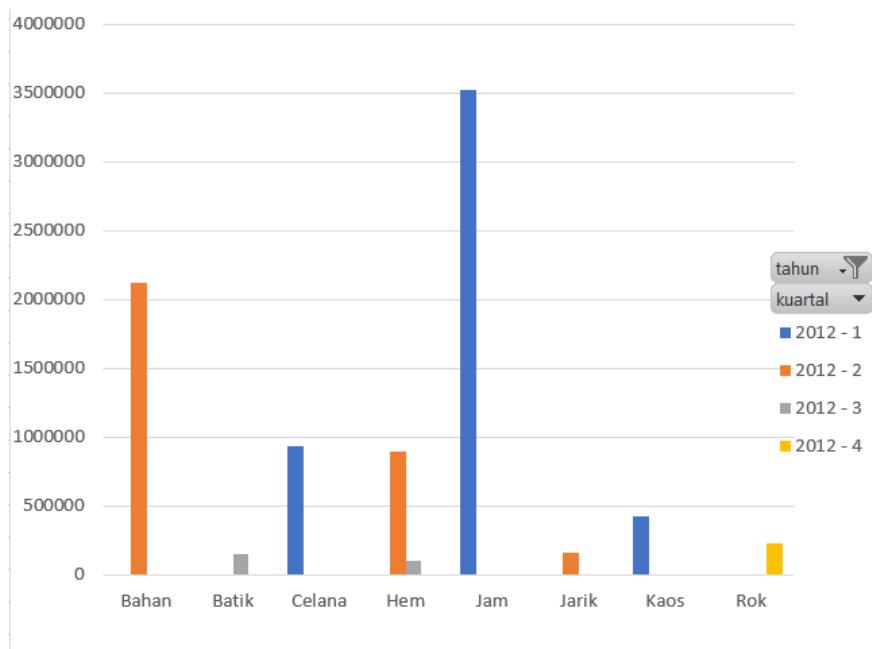
- Arahkan kursor pada area pivot table dalam Sheet1 (Hasil PivotTable).
- Pada menu Analyze, klik PivotChart.



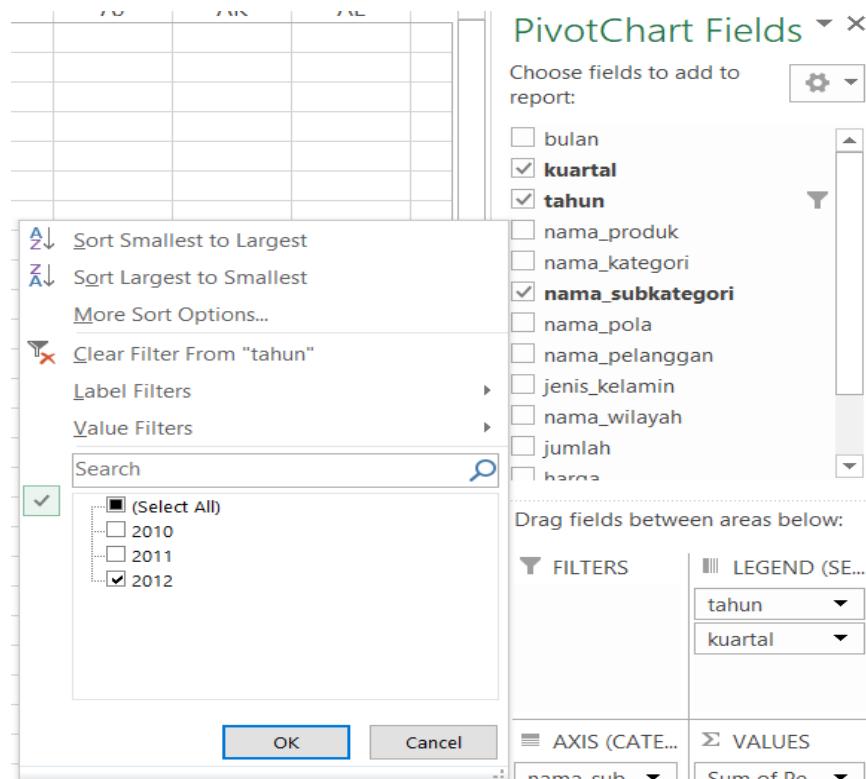
- Pada jendela Insert Chart, pilih bentuk grafik yang diinginkan.



4. Jika grafik terlalu rinci, maka bisa dibuat secara lebih umum dengan menghilangkan kembali tanda cek pada field dalam PivotTable Field List. Misalkan hilangkan tanda cek pada field **nama_produk**, **nama_kategori**, dan **bulan**.

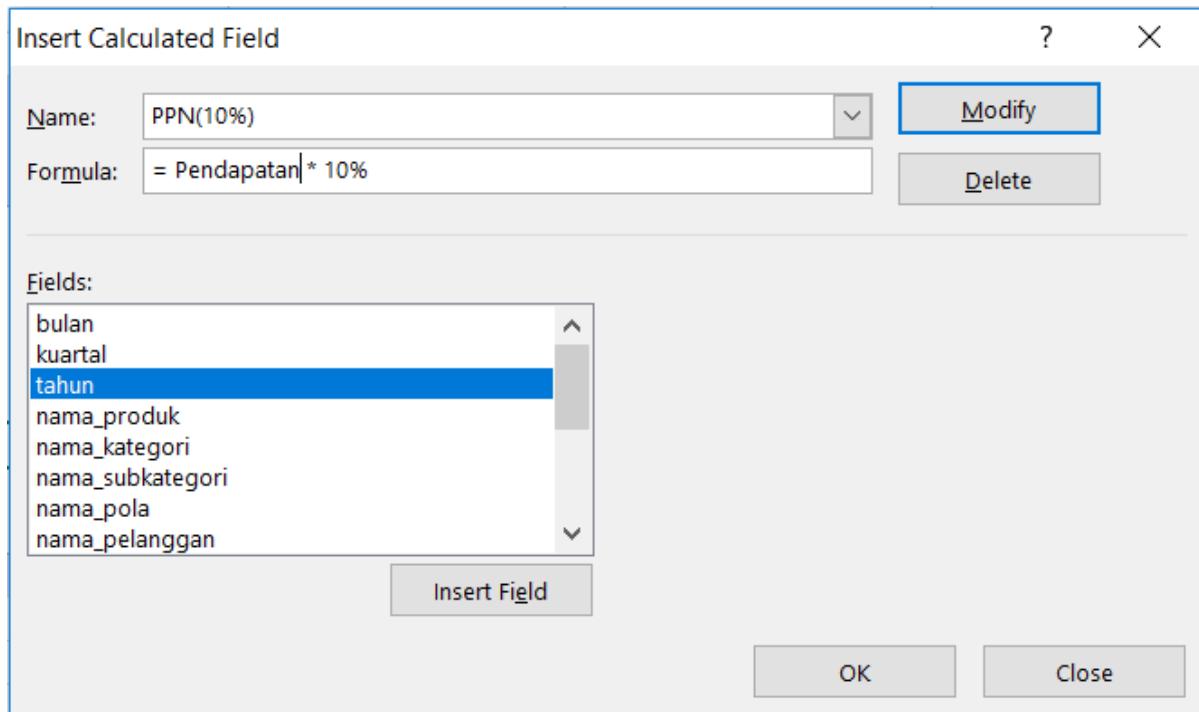


5. Dengan melihat grafik PivotChart, pola transaksi dari kuartal pertama hingga kuartal keempat dapat dilihat dengan mudah apakah terjadi kenaikan, penurunan atau stabil untuk masing-masing kategori produk.
6. Jendela PivotChart Filter Pane berfungsi untuk menyaring (filter) data-data khusus yang akan ditampilkan saja.

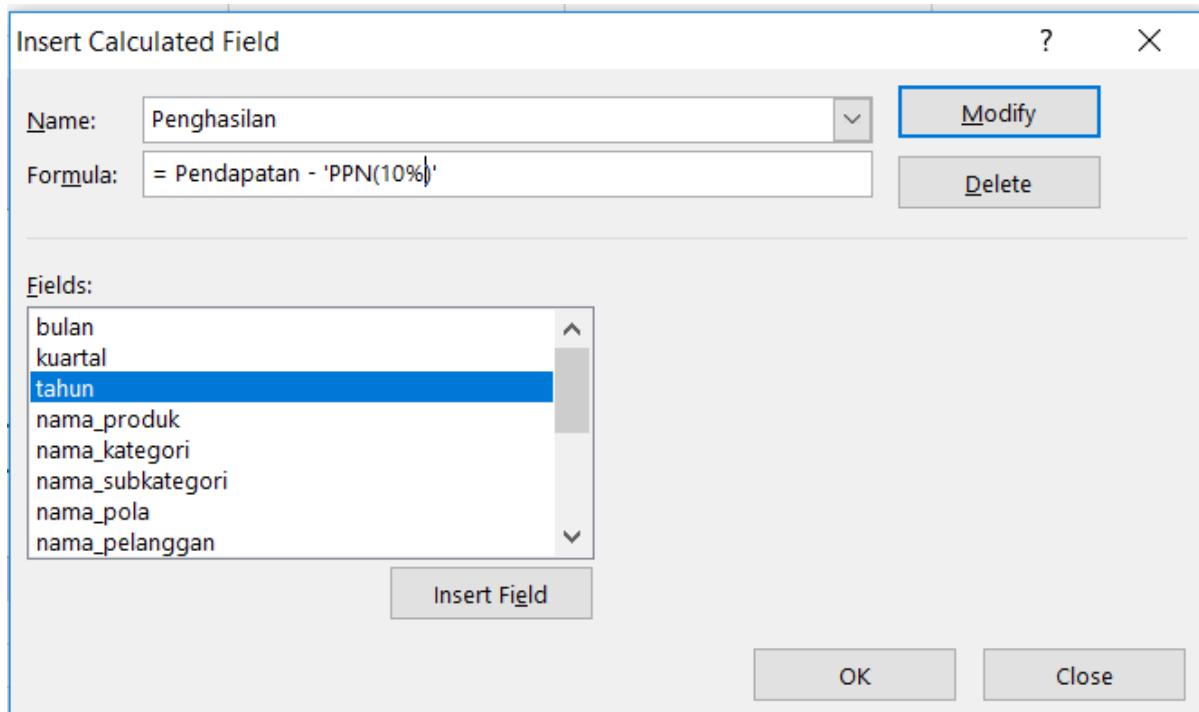


Tugas

1. Dengan menggunakan PivotTable pada file Fakta_Penjualan.xls tambahkan 2 buah field, yaitu :
 - a. PPN (Pajak Pertambahan Nilai) sebesar 10% dari tiap pendapatan pada Pivot Table.



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



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

Pivot Table

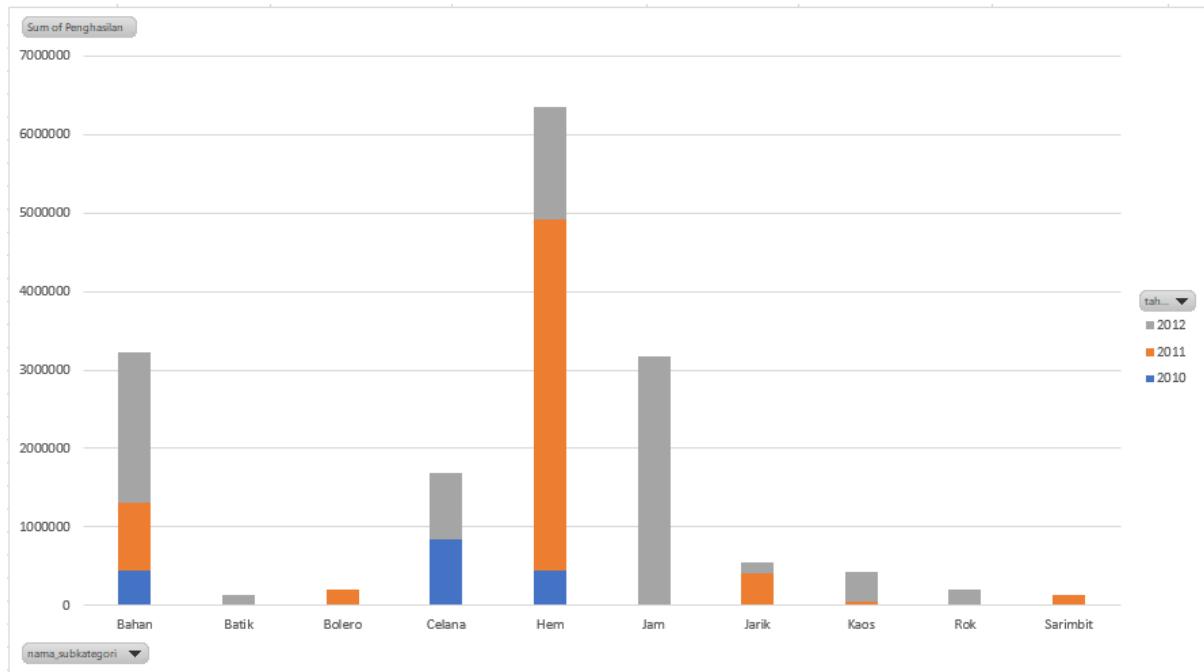
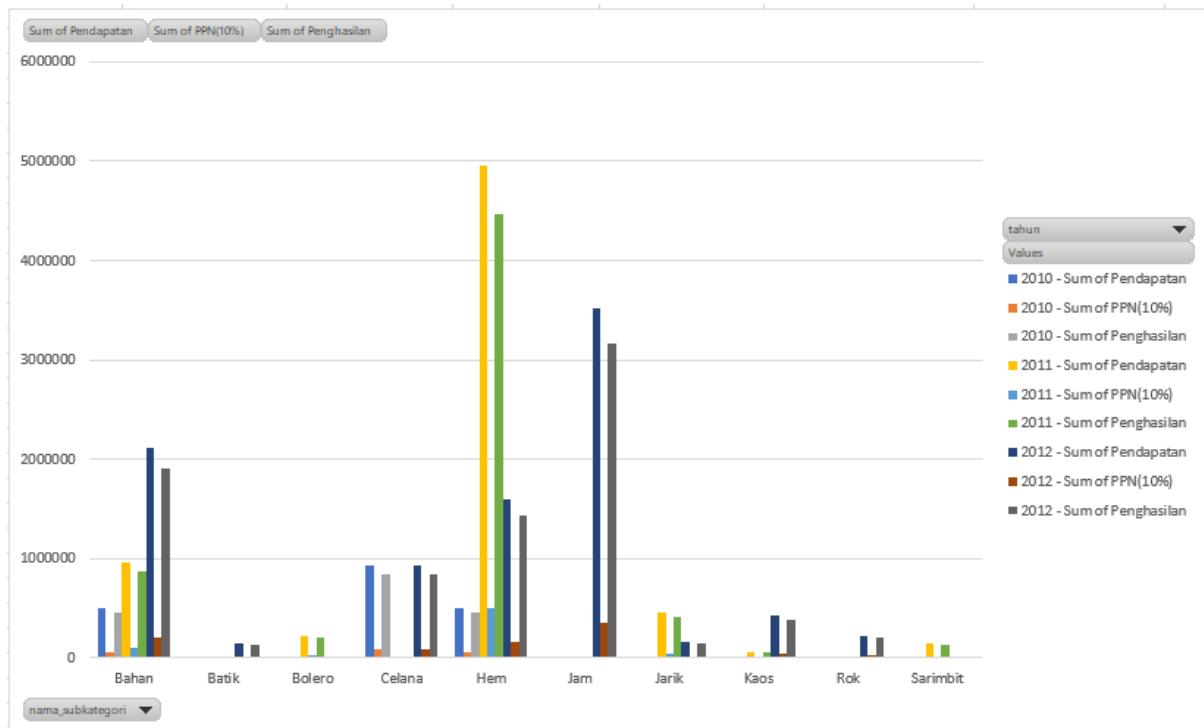
	Column Labels			Total Sum of Pendapatan	Total Sum of PPN(10%)	Total Sum of Penghasilan
	2010					
Row Labels	Sum of Pendapatan	Sum of PPN(10%)	Sum of Penghasilan			
Bahan	500000	50000	450000	500000	50000	450000
Celana	935000	93500	841500	935000	93500	841500
Hem	500000	50000	450000	500000	50000	450000
Grand Total	15065000	1506500	13558500	15065000	1506500	13558500

	Column Labels			Total Sum of Pendapatan	Total Sum of PPN(10%)	Total Sum of Penghasilan
	2011					
Row Labels	Sum of Pendapatan	Sum of PPN(10%)	Sum of Penghasilan			
Bahan	960000	96000	864000	960000	96000	864000
Bolero	225000	22500	202500	225000	22500	202500
Hem	4960000	496000	4464000	4960000	496000	4464000
Jarik	450000	45000	405000	450000	45000	405000
Kaos	60000	6000	54000	60000	6000	54000
Sarimbit	150000	15000	135000	150000	15000	135000
Grand Total	29400000	2940000	26460000	29400000	2940000	26460000

	Column Labels			Total Sum of Pendapatan	Total Sum of PPN(10%)	Total Sum of Penghasilan
	2012					
Row Labels	Sum of Pendapatan	Sum of PPN(10%)	Sum of Penghasilan			
Bahan	2120000	212000	1908000	2120000	212000	1908000
Batik	150000	15000	135000	150000	15000	135000
Celana	935000	93500	841500	935000	93500	841500
Hem	1596000	159600	1436400	1596000	159600	1436400
Jam	3520000	352000	3168000	3520000	352000	3168000
Jarik	160000	16000	144000	160000	16000	144000
Kaos	420000	42000	378000	420000	42000	378000
Rok	225000	22500	202500	225000	22500	202500
Grand Total	115692000	11569200	104122800	115692000	11569200	104122800

	2010			2011			2012			Total Sum of Pendapatan	Total Sum of PPN(10%)	Total Sum of Penghasilan
Row Labels	Sum of Pendapatan	Sum of PPN(10%)	Sum of Penghasilan	Sum of Pendapatan	Sum of PPN(10%)	Sum of Penghasilan	Sum of Pendapatan	Sum of PPN(10%)	Sum of Penghasilan			
Bahan	500000	50000	450000	960000	96000	864000	2120000	212000	1908000	1504500	1504500	13540500
Batik	0	0	0	0	0	0	150000	15000	135000	150000	15000	135000
Bolero	0	0	0	225000	22500	202500	0	0	0	225000	22500	202500
Celana	935000	93500	841500	0	0	0	935000	93500	841500	3740000	374000	3366000
Hem	500000	50000	450000	4960000	496000	4464000	1596000	1596000	1436400	19023000	1902300	17120700
Jam	0	0	0	0	0	0	3520000	3520000	3168000	3520000	352000	3168000
Jarik	0	0	0	450000	45000	405000	160000	16000	144000	1590000	159000	1431000
Kaos	0	0	0	60000	6000	54000	420000	42000	378000	1350000	135000	1215000
Rok	0	0	0	0	0	0	225000	225000	202500	225000	22500	202500
Sarimbit	0	0	0	150000	15000	135000	0	0	0	150000	15000	135000
Grand Total	15065000	1506500	13558500	29400000	2940000	26460000	115692000	11569200	104122800	45196300	45196300	406766700

Pivot Chart



Subkategori produk yang memberikan nilai penghasilan terbanyak selama 3 tahun tersebut adalah **Hem**

MODUL06

1	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten	Lama_Studi		
2	IPS	WANITA	SURAKARTA	18	TIDAK	TERLAMBAT	Jml IPA	10
3	IPA	PRIA	SURAKARTA	19	YA	TEPAT	Jml IPS	6
4	LAIN	PRIA	SURAKARTA	19	TIDAK	TERLAMBAT	Jml LAIN	4
5	IPA	PRIA	LUAR	17	TIDAK	TERLAMBAT		
6	IPA	WANITA	SURAKARTA	17	TIDAK	TEPAT	Jml TEPAT	13
7	IPA	WANITA	LUAR	18	YA	TEPAT	Jml TERLAMBAT	7
8	IPA	PRIA	SURAKARTA	18	TIDAK	TERLAMBAT		
9	IPA	PRIA	SURAKARTA	19	TIDAK	TEPAT	MAX	23
10	IPS	PRIA	LUAR	18	TIDAK	TERLAMBAT	MIN	16
11	LAIN	WANITA	SURAKARTA	18	TIDAK	TEPAT	MEAN	18.95
12	IPA	WANITA	SURAKARTA	19	TIDAK	TEPAT	Standar Deviation	1.669384
13	IPS	PRIA	SURAKARTA	20	TIDAK	TEPAT		
14	IPS	PRIA	SURAKARTA	19	TIDAK	TEPAT	GABUNGAN	3
15	IPA	PRIA	SURAKARTA	19	TIDAK	TEPAT		
16	IPA	PRIA	LUAR	22	YA	TEPAT		
17	LAIN	PRIA	SURAKARTA	16	TIDAK	TERLAMBAT		
18	IPS	PRIA	LUAR	20	TIDAK	TEPAT		
19	LAIN	PRIA	LUAR	23	YA	TEPAT		
20	IPA	PRIA	SURAKARTA	21	YA	TEPAT		
21	IPS	PRIA	SURAKARTA	19	TIDAK	TERLAMBAT		

MODUL07

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter Choose None Apply

Current relation
Relation: Cuaca Attributes: 5 Instances: 14 Sum of weights: 14

Attributes

All None Invert Pattern

No.	Name
1	Cuaca
2	Suhu
3	Kelembapan_Udara
4	Berangin
5	Bermain_Tenis

Remove

Selected attribute
Name: Cuaca Type: Nominal
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

No.	Label	Count	Weight
1	Cerah	5	5.0
2	Mendung	4	4.0
3	Hujan	5	5.0

Class: Bermain_Tenis (Nom) Visualize All

Status OK Log x 0

Cuaca - Notepad

File Edit Format View Help

```
@relation Cuaca

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

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

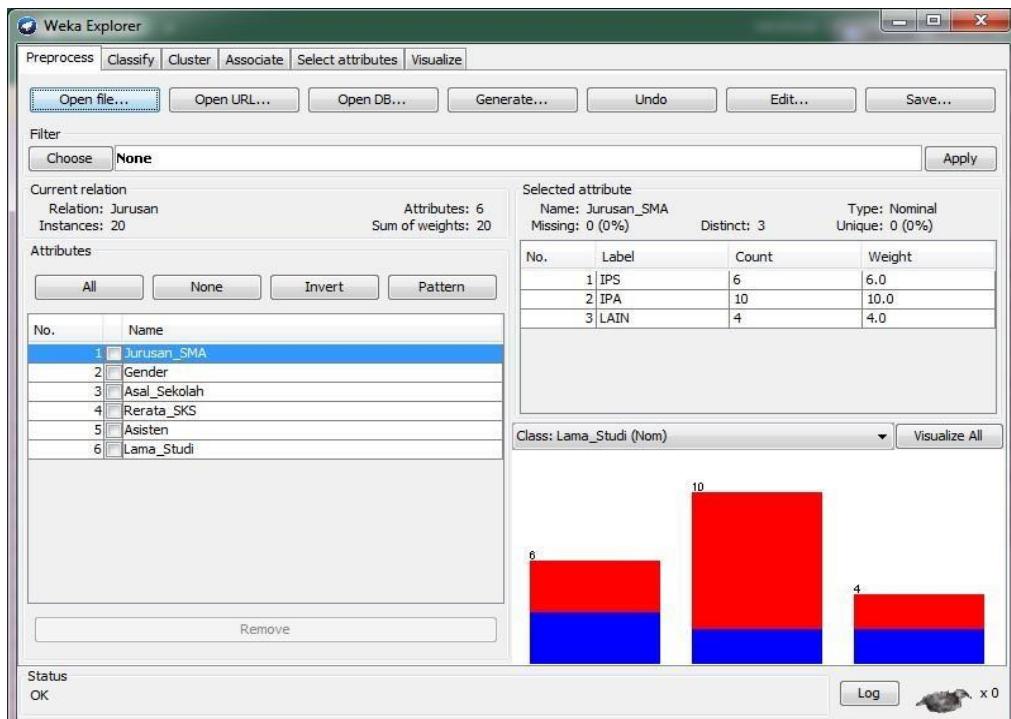
Tugas

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

```
Jurusan - Notepad
File Edit Format View Help
@relation Jurusan
@attribute Jurusan_SMA {IPS, IPA, LAIN}
@attribute Gender {WANITA, PRIA}
@attribute Asal_Sekolah {SURAKARTA, LUAR}
@attribute Rerata_SKS real
@attribute Asisten {YA, TIDAK}
@attribute Lama_Studi{TERLAMBAT, TEPAT}

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

- Perlihatkan file ARFF dan juga gambar grafik untuk setiap data yang ditampilkan dalam WEKA berdasarkan file ARFF



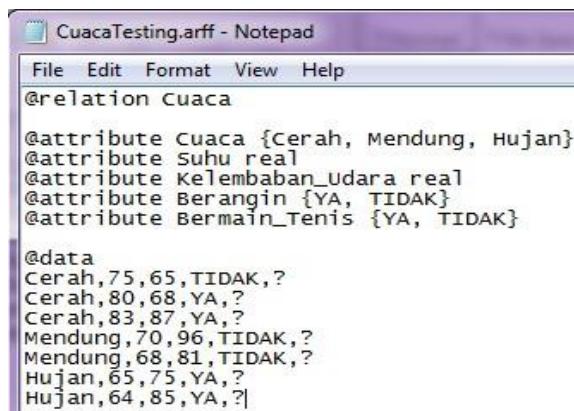
- Berapa jumlah atribut yang bertipe binomial dan polynomial
 - Ada 4 atribut yang bertipe binomial yaitu Gender, Asal_Sekolah, Asisten, dan Lama_Studi
 - Ada 1 atribut yang bertipe Polynomial yaitu Jurusan_SMA 4.
- Ada 1 atribut bertipe real yaitu Rerata_SKS
- Pada atribut Rerata_SKS :
 - Nilai Maximum : 23
 - Nilai Minimum : 16
 - Mean : 18,95
 - StdDev : 1,669384

MODUL08

- Persiapkan file Cuaca.arff :

```
1 @relation Cuaca
2
3 @attribute Cuaca {Cerah, Mendung, Hujan}
4 @attribute Suhu real
5 @attribute Kelembaban_Udara real
6 @attribute Berangin {YA, TIDAK}
7 @attribute Bermain_Tenis {YA, TIDAK}
8
9 @data
10 Cerah,85,85,TIDAK,TIDAK
11 Cerah,80,90,YA,TIDAK
12 Mendung,83,86,TIDAK,YA
13 Hujan,70,96,TIDAK,YA
14 Hujan,68,80,TIDAK,YA
15 Hujan,65,70,YA,TIDAK
16 Mendung,64,65,YA,YA
17 Cerah,72,95,TIDAK,TIDAK
18 Cerah,69,70,TIDAK,YA
19 Hujan,75,80,TIDAK,YA
20 Cerah,75,70,YA,YA
21 Mendung,72,90,YA,YA
22 Mendung,81,75,TIDAK,YA
23 Hujan,71,91,YA,TIDAK
```

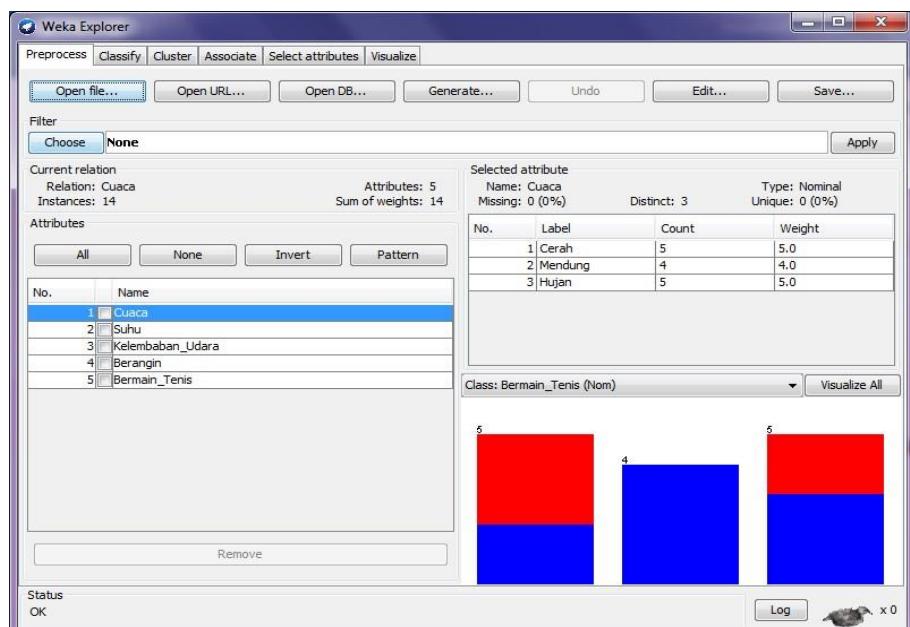
- Buatlah seperti pada gambar dan save dengan nama CuacaTesting.arff :



```
CuacaTesting.arff - Notepad
File Edit Format View Help
@relation Cuaca
@attribute Cuaca {Cerah, Mendung, Hujan}
@attribute Suhu real
@attribute Kelembaban_udara real
@attribute Berangin {YA, TIDAK}
@attribute Bermain_Tenis {YA, TIDAK}

@data
Cerah,75,65,TIDAK,?
Cerah,80,68,YA,?
Cerah,83,87,YA,?
Mendung,70,96,TIDAK,?
Mendung,68,81,TIDAK,?
Hujan,65,75,YA,?
Hujan,64,85,YA,?]
```

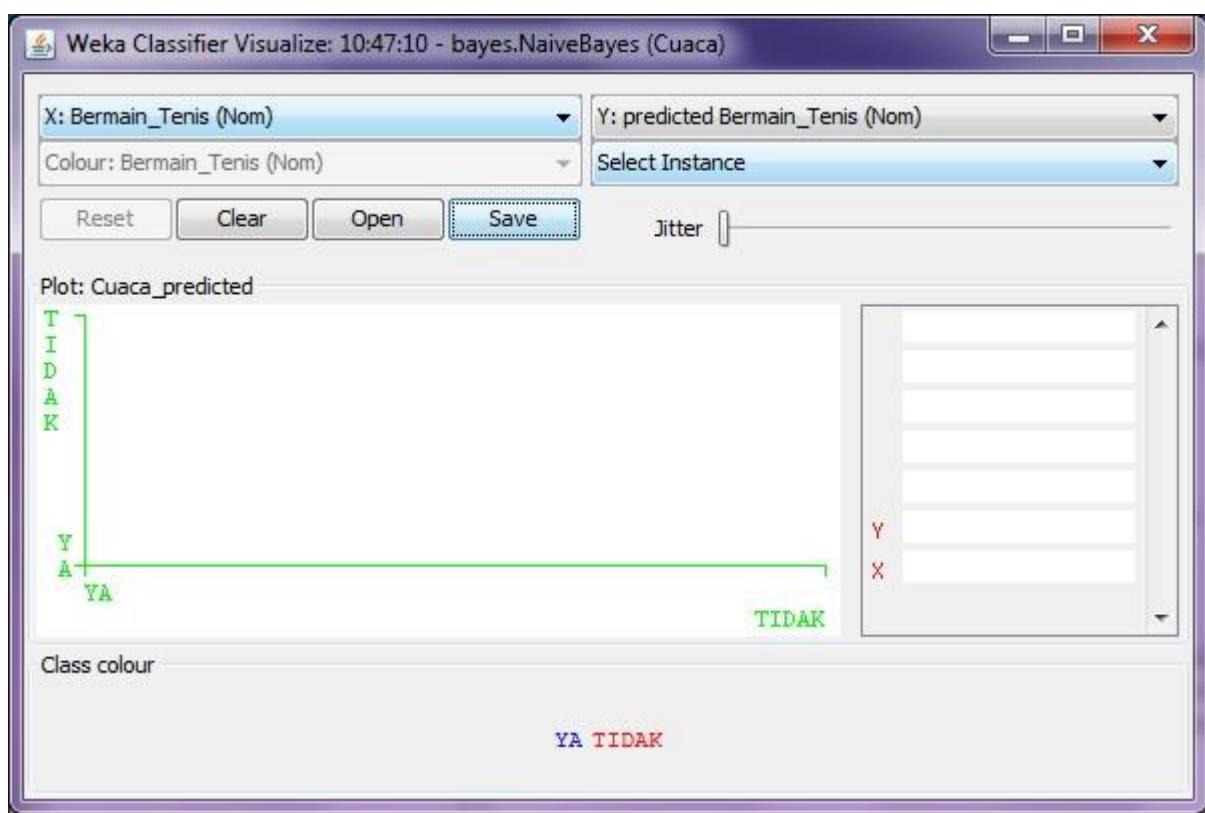
- Buka kembali file Cuaca.arff dengan menggunakan Weka Explorer. Buka Weka Explorer – Klik Open File – Pilih Cuaca.arff



4. Masih pada Weka Explorer, pilih tab **Classify**
5. Sehingga akan muncul jendela Weka Explorer pada tab Classify, pada kotak **Classifier** klik tombol **Choose** untuk memilih metode/algoritma **Naïve Bayes**
6. Klik tombol Set untuk menentukan file ARFF sebagai data uji
7. Sehingga akan muncul jendela Test Instance, klik **Open file . . .**



8. Pilih file **CuacaTesting.arff**
9. Kemudian klik **Close**
10. Klik Start untuk memulai proses naïve bayes
11. Klik kanan pada hasil proses dalam kotak **result list**. Pilih **Visual classifier errors**
12. Klik **Save**. Simpan dengan nama file **HasilPrediksi.arff**



13. Tutup semua jendela termasuk Weka Explorer dan kembali ke Weka GUI Chooser. Pilih menu **Tools-ArffViewer**
14. Buka menu File-Open. Tunjukkan pada file HasilPrediksi.arff yang telah Anda simpan

ARFF-Viewer - C:\Users\LABSI-15\Documents\HasilProduksi.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 RapidMiner

- Persiapkan file **Tabel_Cuaca.xls** yang terdiri dari 2 sheet.
- Tabel data training** pada Sheet1

	A	B	C	D	E
1	Cuaca	Suhu	Kelembapan Udara	Berangin	Bermain_Tenis
2	Cerah	85	85	TIDAK	TIDAK
3	Cerah	80	90	YA	TIDAK
4	Mendung	83	86	TIDAK	YA
5	Hujan	70	96	TIDAK	YA
6	Hujan	68	80	TIDAK	YA
7	Hujan	65	70	YA	TIDAK
8	Mendung	64	65	YA	YA
9	Cerah	72	95	TIDAK	TIDAK
10	Cerah	69	70	TIDAK	YA
11	Hujan	75	80	TIDAK	YA
12	Cerah	75	70	YA	YA
13	Mendung	72	90	YA	YA
14	Mendung	81	75	TIDAK	YA
15	Hujan	71	91	YA	TIDAK

< > Training Testing +

- Tabel data uji pada Sheet2 tanpa ada variabel **Bermain_Tenis**

	A	B	C	D
1	Cuaca	Suhu	Kelembapan_udara	Berangin
2	Cerah		75	65 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

< > Training Testing +

- Buka aplikasi Rapid Miner
- Klik **Import Data**. Arahkan direktori tempat penyimpanan file pada langkah **Select data location**, kemudian pilih file yang akan digunakan dan klik **Next**.
- Pastikan sel Excel sesuai di langkah **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	F
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	Hujan	71.000	91.000	YA	TIDAK	

Ac
Previous Next Cancel

7. Pada langkah **Format your columns** ubah kolom **Bermain_Tenis** dengan tipe data **binomial** karena hanya ada dua keputusan (YA dan TIDAK)

Format your columns.

Replace errors with missing values

	Cuaca	Suhu	Kelembaban_Udara	Berangin	Bermain_Tenis
1	Cerah	85	85	TIDAK	TIDAK
2	Cerah	80	90	YA	TIDAK
3	Mendung	83	86	TIDAK	YA
4	Hujan	70	96	TIDAK	YA
5	Hujan	68	80	TIDAK	YA
6	Hujan	65	70	YA	TIDAK
7	Mendung	64	65	YA	YA

Change Type binomial

8. Ubah pula sebagai **label** pada **Change Role**

Bermain_Tenis
binomial
label
TIDAK

9. Simpan dengan nama **DataCuaca_Training** dilanjutkan klik tombol **Finish** 10. Hasil import file **Tabel_Cuaca.xls** pada Sheet1 akan di tampilkan.

ory ExampleSet (/Local Repository/DataCuaca_Training) X

Open in Turbo Prep Auto Model Filter (14 / 14 examples): all

Row No.	Bermain_Tenis	Cuaca	Suhu	Kelembaban_Udara	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	TIDAK	Hujan	71	91	YA

11. Kembali ke jendela Design Perspective dengan shortcut tombol **F8**
12. Lakukan hal yang sama untuk data testing yang diambil dari **Tabel_Cuaca.xls** pada Sheet2(Testing) dengan mengulang dari langkah 5

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

13. Simpan dengan nama **DataCuaca_Testing**.

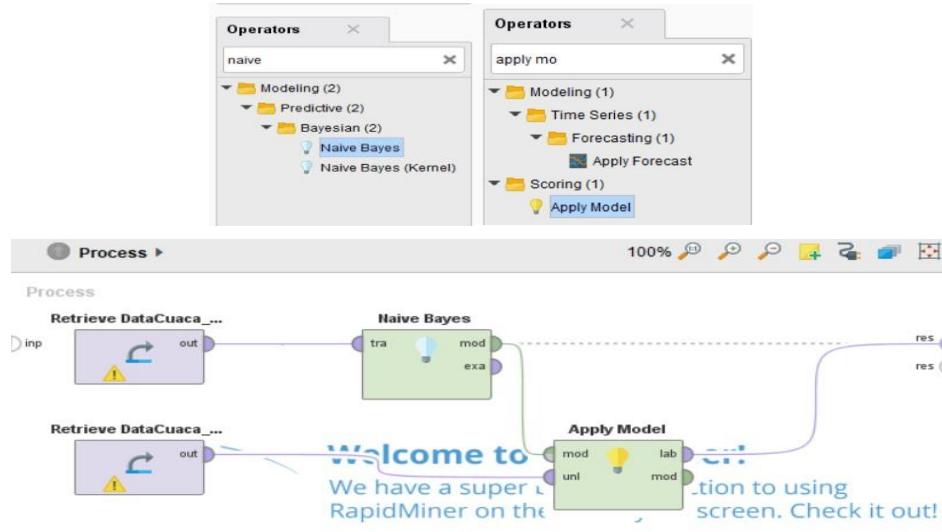
itory ExampleSet (/Local Repository/DataCuaca_Testing) X

Open in Turbo Prep Auto Model Filter (7 / 7 examples):

Row No.	Cuaca	Suhu	Kelembaban_Udara	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

14. Drag **DataCuaca_Training** dan **DataCuaca_Testing** kedalam jendela Process View

15. Masukkan juga operator **Naive Bayes** dan **Apply Model** ke dalam Proses View. Hubungkan koneksi masing-masing data terhadap operator seperti gambar :



16. Jalankan proses naive bayes dengan menekan tombol **Run** (atau menekan tombol F11).

17. Perhatikan hasil proses klasifikasi naive bayes. Pada tab **Data**, dapat dilihat hasil prediksi serta tingkat confidence nilai kelas pada masing-masing data.

Row No.	prediction(B...)	confidence(...)	confidence(...)	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

Pada tab **Statistics**, dapat dilihat bahwa distribusi nilai kelas pada variabel Y (Bermain_Tenis) rerata nilai confidence sebesar 0,353 untuk nilai TIDAK, dan 0,647 untuk nilai YA

	Name	Type	Missing	Statistics	Filter (7 / 7 attributes):
Data	prediction(Bermain_Tenis)	Binominal	0	Least TIDAK (2) Most YA (5)	Values YA (5), TIDAK (2)
Statistics	confidence(TIDAK)	Real	0	Min 0.007 Max 0.856	Average 0.353
Visualizations	confidence(YA)	Real	0	Min 0.144 Max 0.993	Average 0.647
Annotations	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)

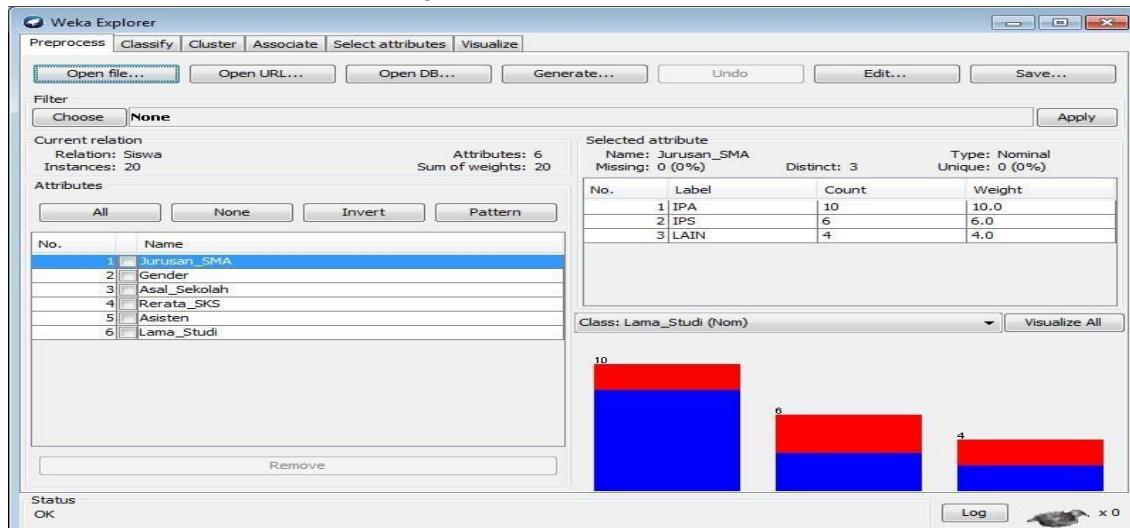
Tugas

- Berdasarkan tabel berikut, buatlah file dalam format Excel (.xls) dan format ARFF (.arff) ! Data ini akan digunakan sebagai **Data Testing**.

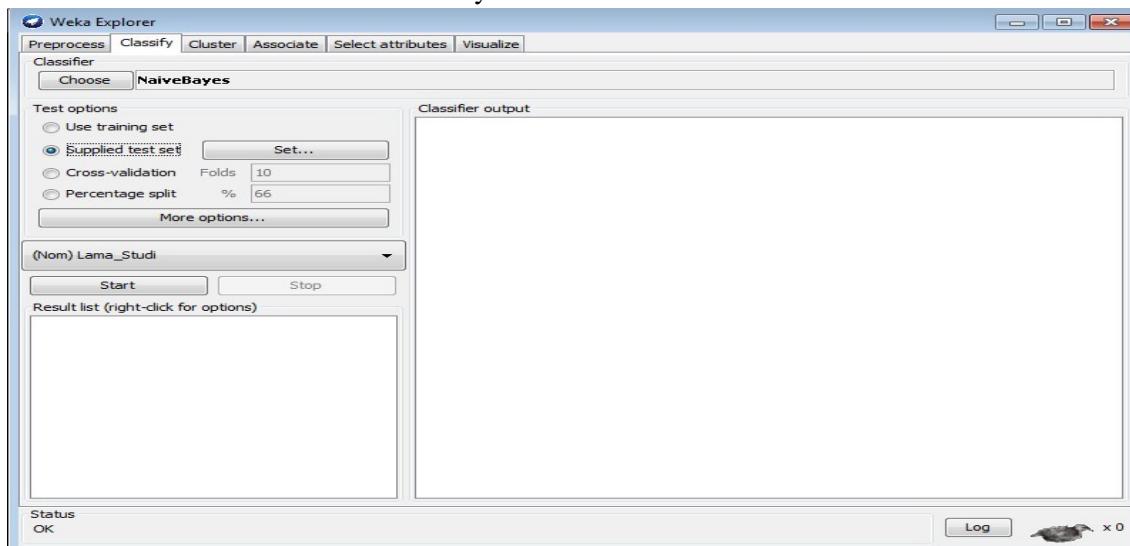
1	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten
2	LAIN	WANITA	SURAKARTA	18	TIDAK
3	IPA	PRIA	SURAKARTA	19	YA
4	LAIN	PRIA	SURAKARTA	19	TIDAK
5	IPS	PRIA	LUAR	17	TIDAK
6	LAIN	WANITA	SURAKARTA	17	TIDAK
7	IPA	WANITA	LUAR	18	YA
8	IPA	PRIA	SURAKARTA	18	TIDAK
9	IPA	PRIA	SURAKARTA	19	TIDAK
10	IPS	PRIA	LUAR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

- Gunakan file ARFF yang dikerjakan pada Tugas nomor 1 dalam Modul 7 sebagai data training. Lakukan prediksi terhadap data testing (ARFF) di atas menggunakan WEKA!

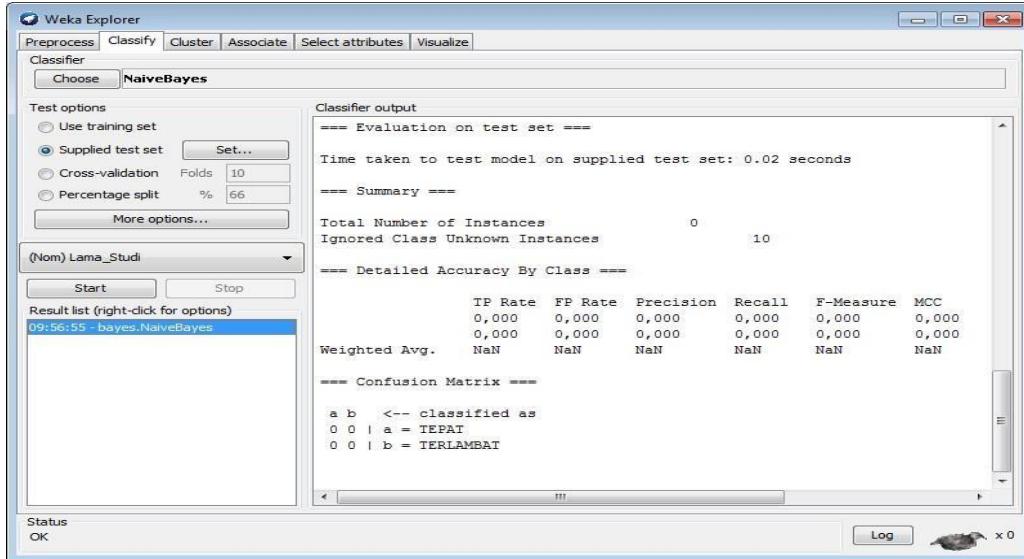
a. Membuka file Training



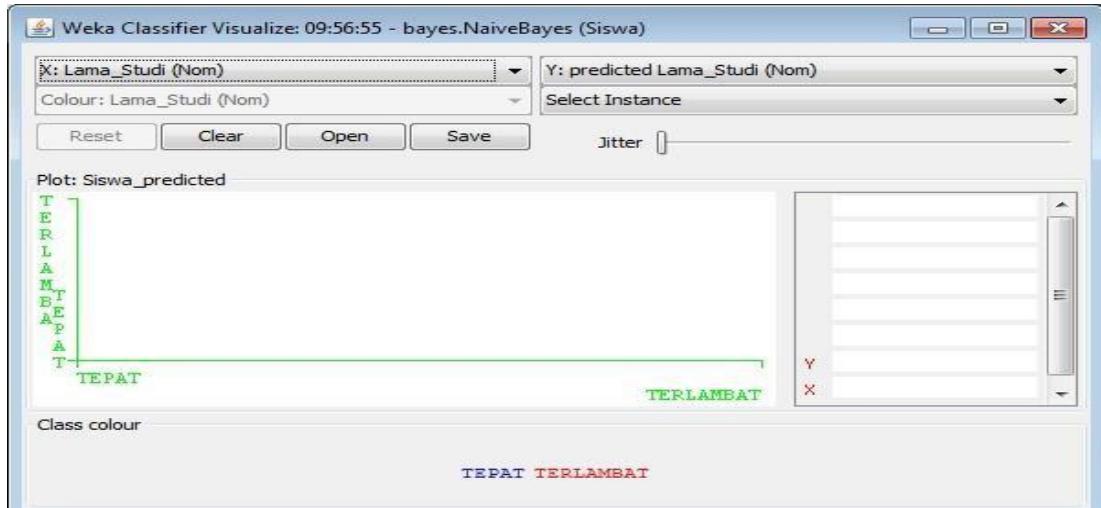
b. Memilih metode Naive Bayes



c. Membuka file Testing lalu klik kanan pilih Visualize classifier errors



d. Menyimpan file hasil



e. Hasil Prediksi

ARFF-Viewer - D:\dwdm\modul 8\tugas\HasiTugas.arff

Relation: Siswa_predicted							
No.	1: Jurusan_SMA Nominal	2: Gender Nominal	3: Asal_Sekolah Nominal	4: Rerata_SKS Numeric	5: Asisten Nominal	6: prediction margin Numeric	7: predicted Lama_Studi 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

3. Gunakan file Excel yang dikerjakan pada Tugas nomor 1 dalam Modul 6 sebagai data training. Lakukan prediksi terhadap data testing (Excel) di atas menggunakan RapidMiner!
- a. Data Training

Select the cells to import.

Sheet: Sheet1 ▾ 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	IPA	PRIA	SURAKARTA	18.000	TIDAK	TEPAT

← Previous → Next ✖ Cancel

Format your columns.

Replace errors with missing values ⓘ

	Jurusan_SMA * polynominal	Gender * polynominal	Asal_Sekolah * polynominal	Rerata_SKS * integer	Asisten * binominal	Lama_Studi * binominal label
1	IPS	WANITA	SURAKARTA	18	TIDAK	TERLAMBAT
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	TEPAT
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
13	IPS	PRIA	SURAKARTA	19	TIDAK	TEPAT

✓ no problems. ← Previous → Next ✖ Cancel

<new process> - RapidMiner Studio Educational 9.3.001 @ asus-PC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Find data, operators...etc All Studio

Result History ExampleSet (/Local Repository/DataCuaca_Testing) ExampleSet (/Local Repository/DataCuaca_Training) ExampleSet (/Local Repository/DataSekolah_Training) ExampleSet (Apply Model)

Data Statistics Visualizations Annotations

Open in Turbo Prep Auto Model Filter (20 / 20 examples): all

Row No.	Lama_Studi	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten
1	TERLAMBAT	IPS	WANITA	SURAKARTA	18	TIDAK
2	TEPAT	IPA	PRIA	SURAKARTA	19	YA
3	TERLAMBAT	LAIN	PRIA	SURAKARTA	19	TIDAK
4	TERLAMBAT	IPA	PRIA	LUAR	17	TIDAK
5	TEPAT	IPA	WANITA	SURAKARTA	17	TIDAK
6	TEPAT	IPA	WANITA	LUAR	18	YA
7	TERLAMBAT	IPA	PRIA	SURAKARTA	18	TIDAK
8	TEPAT	IPA	PRIA	SURAKARTA	19	TIDAK
9	TERLAMBAT	IPS	PRIA	LUAR	18	TIDAK
10	TEPAT	LAIN	WANITA	SURAKARTA	18	TIDAK
11	TEPAT	IPA	WANITA	SURAKARTA	19	TIDAK
12	TEPAT	IPS	PRIA	SURAKARTA	20	TIDAK
13	TEPAT	IPS	PRIA	SURAKARTA	19	TIDAK
14	TEPAT	IPA	PRIA	SURAKARTA	19	TIDAK

ExampleSet (20 examples, 1 special attribute, 5 regular attributes)

Repository

- Import Data
- Samples
- DB Legacy
- Local Repository (asus)
 - Connections (asus)
 - data (asus)
 - processes (asus)
 - DataCuaca_Testing (asus - v1, 10/5/19 9:25 AM)
 - DataCuaca_Training (asus - v1, 10/5/19 9:25 AM)
 - DataSekolah_Training (asus - v1, 10/5/19 10:00 AM)
 - HasilCuaca (asus - v1, 10/5/19 10:00 AM)

b. Data Testing

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Sheet2 Cell range: A:E Select All Define header row: 1

A	B	C	D	E
1	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS
2	LAIN	WANITA	SURAKARTA	18.000
3	IPA	PRIA	SURAKARTA	19.000
4	LAIN	PRIA	SURAKARTA	19.000
5	IPS	PRIA	LUAR	17.000
6	LAIN	WANITA	SURAKARTA	17.000
7	IPA	WANITA	LUAR	18.000
8	IPA	PRIA	SURAKARTA	18.000
9	IPA	PRIA	SURAKARTA	19.000
10	IPS	PRIA	LUAR	18.000
11	LAIN	WANITA	SURAKARTA	18.000

Previous Next Cancel

Import Data - Format your columns.

Format your columns.

Replace errors with missing values

Jurusan_SMA polynominal	Gender polynominal	Asal_Sekolah polynominal	Rerata_SKS integer	Asisten polynominal
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

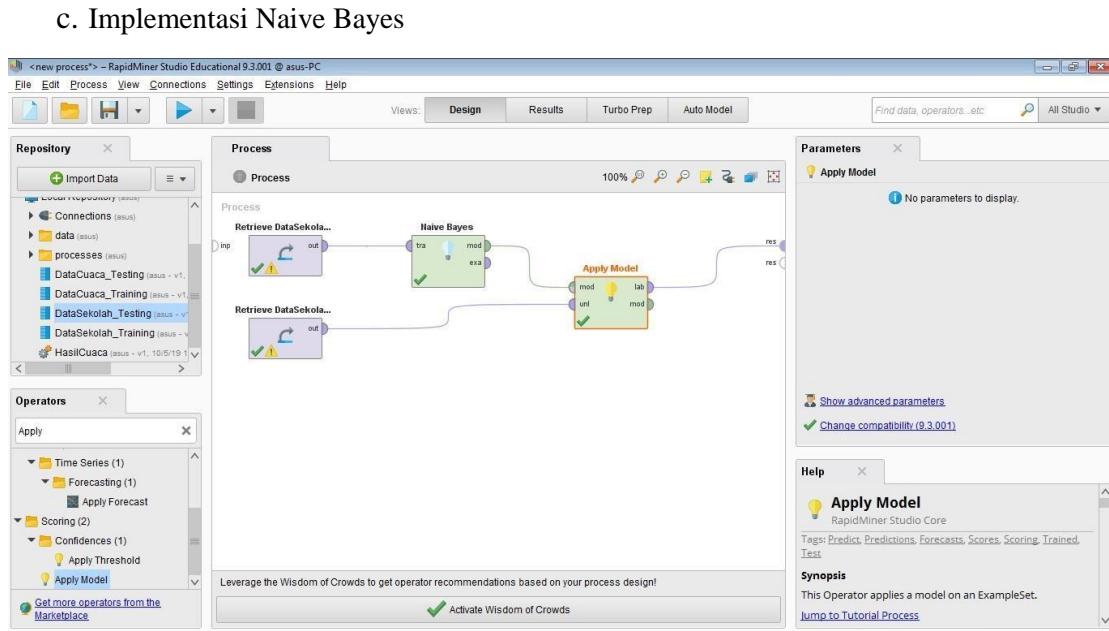
no problems.

Previous Next Cancel

c. Implementasi Naive Bayes

The screenshot shows the RapidMiner Studio interface. The top menu bar includes File, Edit, Process, View, Connections, Settings, Extensions, Help, Views: Design, Results, Turbo Prep, Auto Model, Find data, operators... etc., and All Studio. The main area has tabs for Data, Statistics, Visualizations, and Annotations. The Data tab displays a table with columns: Row No., Jurusan_SMA, Gender, Asal_Sekolah, Rerata_SKS, and Asisten. The table contains 10 rows of student information. The Repository tab shows a process flow: 'Retrieve DataSekolah...' feeds into a 'Naive Bayes' operator, which then feeds into an 'Apply Model' operator. The 'Apply Model' operator has inputs 'mod', 'tra', 'res', and 'out'. The 'Repository' sidebar lists various datasets and processes.

Row No.	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	IPS	PRIA	LUAR	19	TIDAK
9	IPS	PRIA	LUAR	18	TIDAK
10	LAIN	WANITA	SURAKARTA	18	TIDAK



d. Hasil

The screenshot shows the RapidMiner Studio interface with the Data tab active. It displays a table titled 'ExampleSet (Apply Model)' with columns: prediction(L...), confidence(tra...), confidence(mod...), Jurusan_SMA, Gender, Asal_Sekolah, Rerata_SKS, and Asisten. The table contains 10 rows of predicted values. The Repository sidebar lists various datasets and processes.

Row No.	prediction(L...)	confidence(tra...)	confidence(mod...)	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	0.868	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	0.648	0.352	LAIN	WANITA	SURAKARTA	18	TIDAK

4. Dari hasil percobaan Tugas nomor 3 di atas, berapakah nilai rerata confidence untuk atribut Lama_Studi dengan nilai TEPAT? Berapakah nilai rerata confidence untuk atribut Lama_Studi dengan nilai TERLAMBAT?

confidence(TERLAMBAT)	Real	0	Min 0.005	Max 0.868	Average 0.524
confidence(TEPAT)	Real	0	Min 0.132	Max 0.995	Average 0.476

Jadi nilai rerata confidence untuk atribut Lama_Studi dengan nilai TEPAT adalah **0,524** dan nilai rerata confidence untuk atribut Lama_Studi dengan nilai TERLAMBAT adalah **0,476**

5. Dari hasil percobaan Tugas nomor 3 di atas, berapa orang yang akan lulus TEPAT, dan berapa orang yang akan lulus TERLAMBAT?

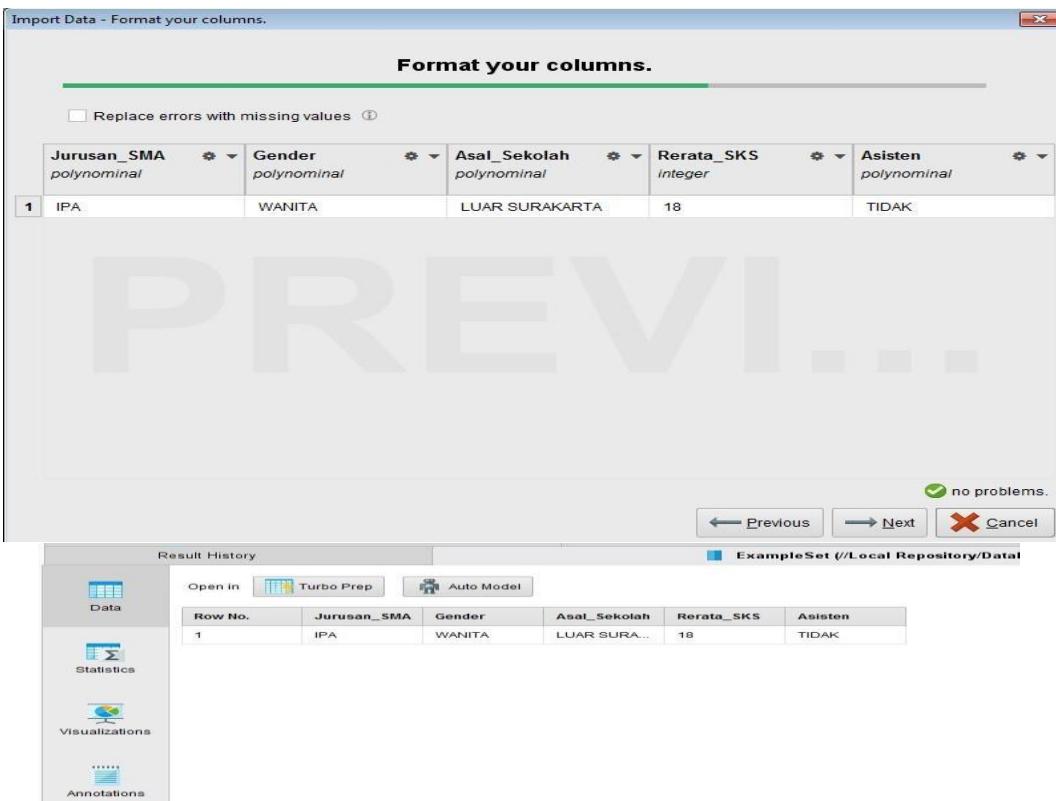
Prediction prediction(Lama_Studi)	Binomial	0	Least TEPAT (3)	Most TERLAMBAT (7)	Values TERLAMBAT (7), TEPAT (3)
--	----------	---	-----------------	--------------------	---------------------------------

Jadi yang lulus dengan tepat sebanyak **3** orang sedangkan yang lulus terlambat sebanyak **7** orang

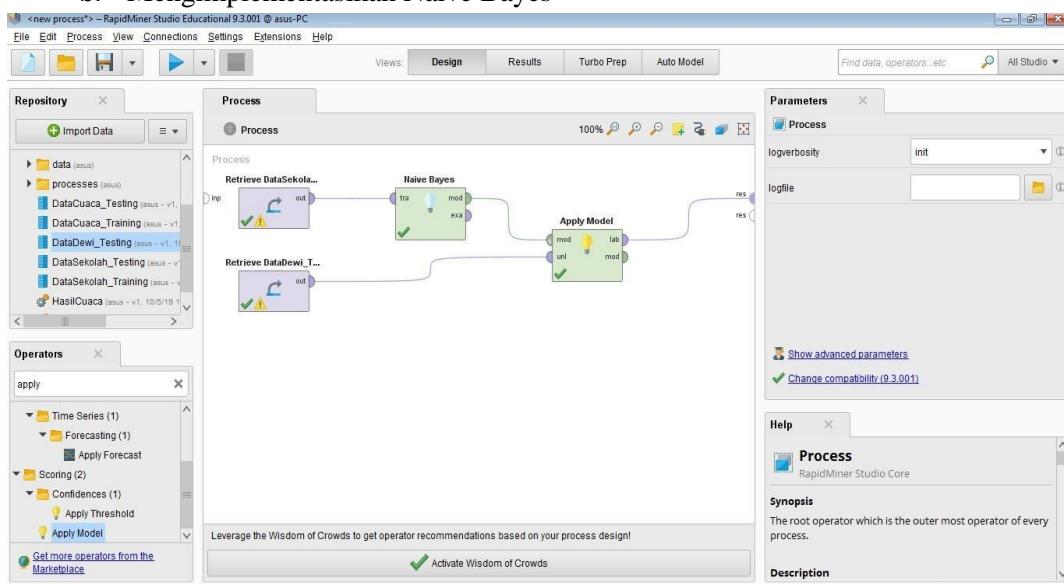
6. Prediksi ketepatan lama studi si Dewi, jika Dewi adalah seorang WANITA yang berasal dari jurusan IPA pada saat SMA, asal sekolah dari LUAR SURAKARTA, mengambil SKS dengan rata-rata sebanyak 18 SKS tiap semester, dan tidak pernah menjadi Asisten selama kuliah.

a. Mengimport data Dewi

A	B	C	D	E
Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten
IPA	WANITA	LUAR SURAKARTA	18.000	TIDAK

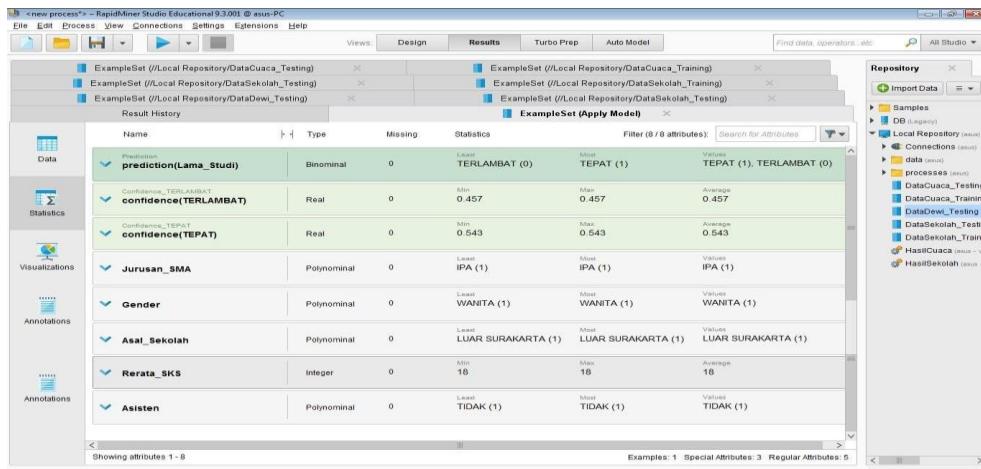


b. Mengimplementasikan Naive Bayes



c. Hasil

The screenshot shows the results table in RapidMiner Studio. The table has nine columns: Row No., prediction(L...), confidence(...), confidence(...), Jurusan_SMA, Gender, Asal_Sekolah, Rerata_SKS, and Asisten. The single row contains the value 'TEPAT' in the 'prediction' column, and '0.457' and '0.543' in the 'confidence' columns. The other columns show the input data: IPA, WANITA, LUAR SURAKARTA, 18, and TIDAK. The top of the table includes buttons for 'Open in' (Turbo Prep, Auto Model) and a 'Filter' dropdown set to 'all'.



Berdasarkan hasil di atas maka prediksi untuk Dewi dia akan lulus dengan tepat

7. Prekdiksiakan ketepatan lama studi si Jono, jika Jono adalah seorang PRIA yang berasal dari jurusan IPA pada saat SMA, asal sekolah dari SURAKARTA, mengambil SKS dengan rata-rata sebanyak 17 SKS tiap semester, dan pernah menjadi Asisten selama kuliah.

a. Mengimport Data Jono

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Sheet4 Cell range: A:E Select All Define header row: 1

A	B	C	D	E	
1	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten
2	LAIN	PRIA	SURAKARTA	17.000	YA

← Previous → Next ✖ Cancel

<new process> – RapidMiner Studio Educational 9.3.001 @ asus-PC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Find data, operators... etc. All Studio

Repository Import Data Samples DB (Legacy) Local Repository (ausu) Connections (ausu) data (ausu) processes (ausu) DataCuaca_Testing DataCuaca_Training DataDewi_Testing DataSekolah_Testing DataSekolah_Training HasilCuaca (ausu - v1) HasilDewi (ausu - v1) HasilSekolah (ausu)

Result History ExampleSet (/Local Repository/DataCuaca_Testing) ExampleSet (/Local Repository/DataCuaca_Training) ExampleSet (/Local Repository/DataSekolah_Testing) ExampleSet (/Local Repository/DataDewi_Testing) ExampleSet (/Local Repository/DataSekolah_Training) ExampleSet (Apply Mode)

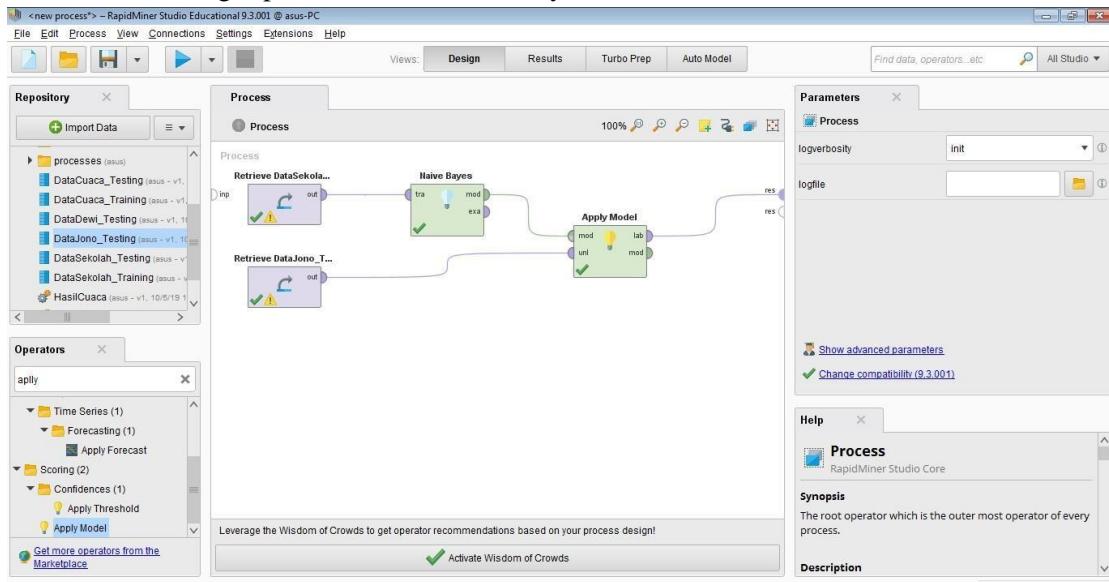
Data Statistics Visualizations Annotations

Open in Turbo Prep Auto Model Filter (1 / 1 examples): all

Row No.	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten
1	LAIN	PRIA	SURAKARTA	17	YA

ExampleSet (1 example, 0 special attributes, 5 regular attributes)

b. Mengimplementasikan Naive Bayes



c. Hasil

The screenshot shows the RapidMiner Studio interface with the following components:

- Data:** Shows a table with one row of data: Row No. 1, prediction(TEPAT), confidence(0.076), confidence(0.924), Jurusan_SMA (LAIN), Gender (PRIA), Asal_Sekolah (SURAKARTA), Rerata_SKS (17), and Asisten (YA).
- Statistics:** Shows the 'ExampleSet (Apply Model)' results table with columns: Name, Type, Missing, Statistics, Filter (8 / 8 attributes), and Search for Attributes.
- Annotations:** Shows the 'prediction(Lama_Studi)', 'confidence(TERLAMBAT)', 'confidence(TEPAT)', 'Jurusan_SMA', 'Gender', 'Asal_Sekolah', 'Rerata_SKS', and 'Asisten' attributes with their respective statistics.
- Repository:** Shows the local repository with various samples and datasets.

Row No.	prediction(TEPAT)	confidence(0.076)	confidence(0.924)	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten
1	TEPAT	0.076	0.924	LAIN	PRIA	SURAKARTA	17	YA

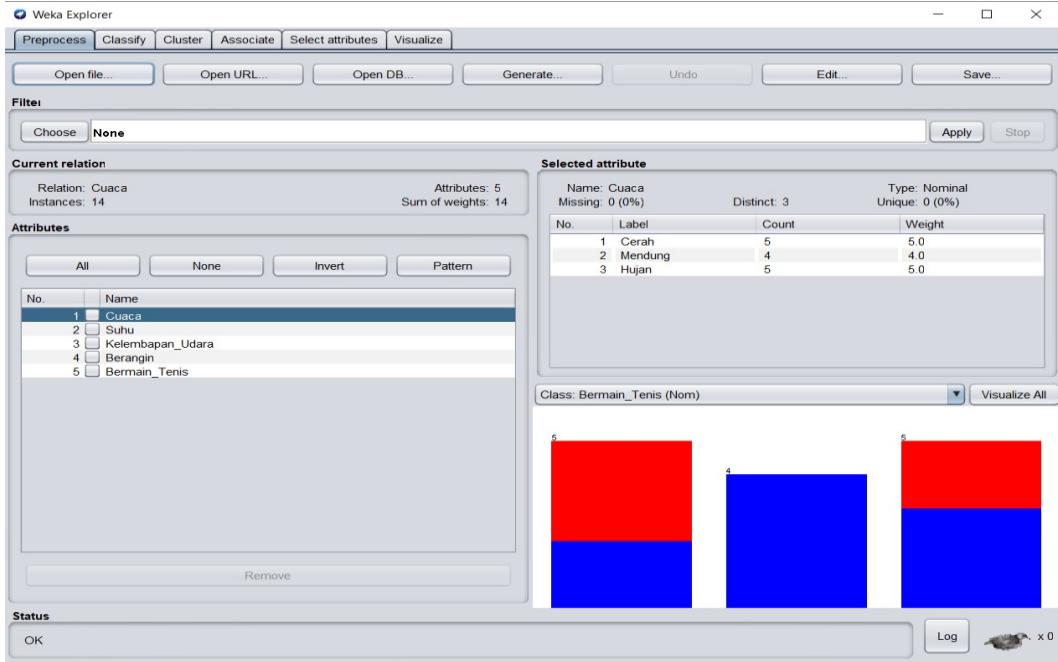
Berdasarkan hasil di atas maka prediksi untuk Jono dia akan lulus dengan tepat

MODUL09

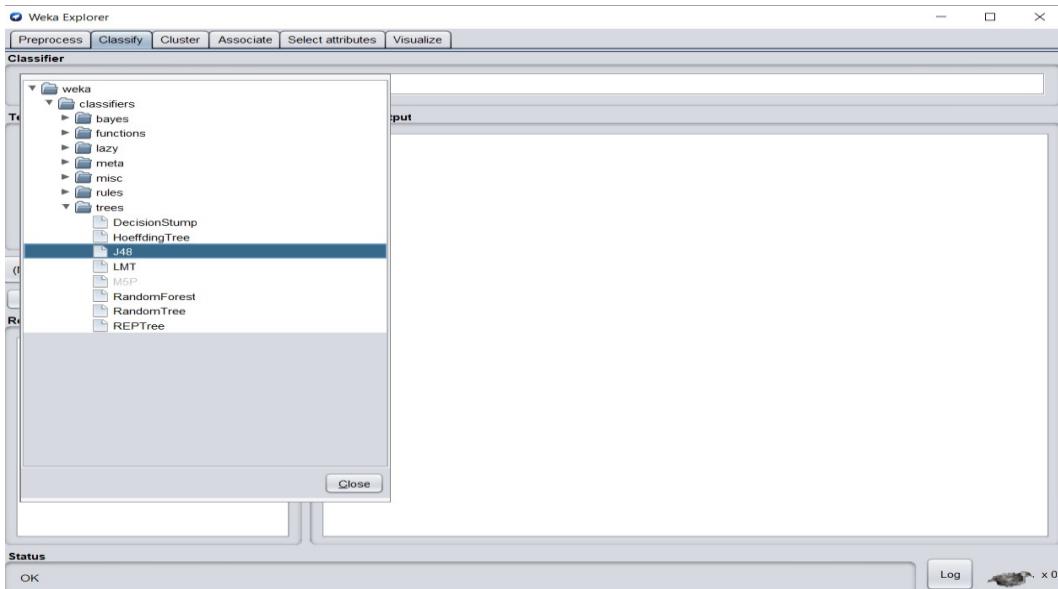
Langkah-Langkah Praktikum

9.1. Pohon Keputusan Menggunakan Weka

1. Jalankan aplikasi Weka Explorer
2. Buka file Cuaca.arff, dengan Weka Explorer

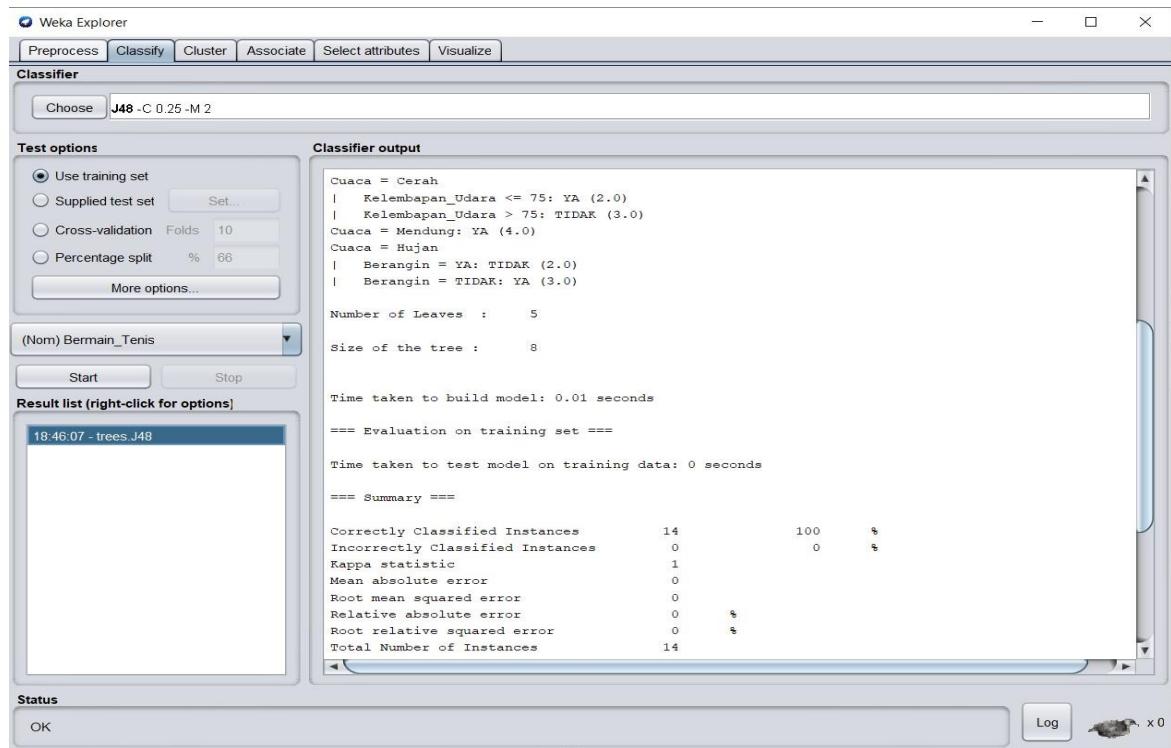


3. Setelah buka file tersebut, buka tab **Classify**.
4. Tekan tombol **Choose – Trees – J84**.



5. Pada pilihan **Test Options**, pilih **Use training set**.
6. Pastikan pada pilihan atribut dependen adalah **Bermain_Tenis**. Kemudian klik **Start**.

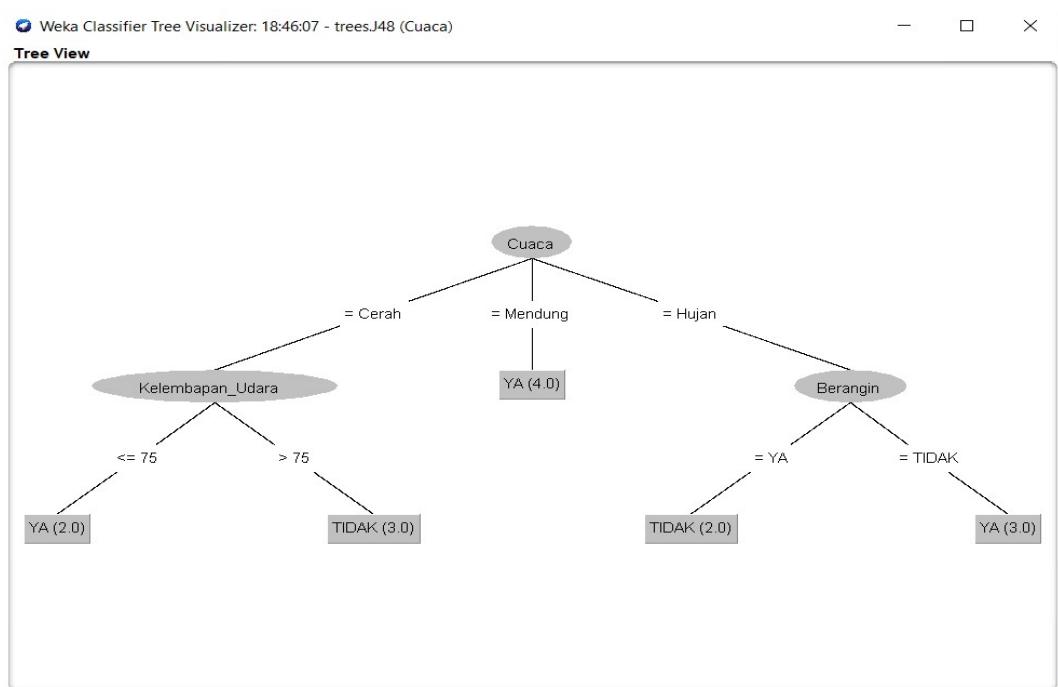
7. Setelah proses selesai. Ada 2 hasil yang diberikan, yaitu pada kolom **Result list** dan kolom **Classifier output**.



8. Berdasarkan hasil tersebut dapat diketahui sebagai berikut :

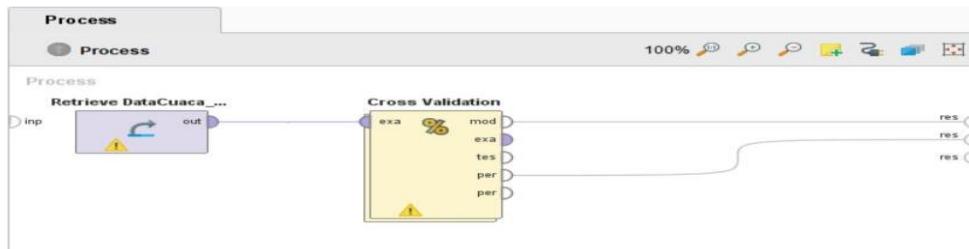
- Jumlah simpul daun pada pohon keputusan = 5
- Jumlah simpul keseluruhan pada pohon keputusan = 8
- Waktu yang dibutuhkan untuk proses pelatihan = 0,01 detik
- Tingkat ketepatan klasifikasi = 100%
- Tingkat ketidaktepatan klasifikasi = 0%

9. Kembali ke kolom **Result List**. Klik kanan pada hasil **trees.J48 – Visualize tree**.

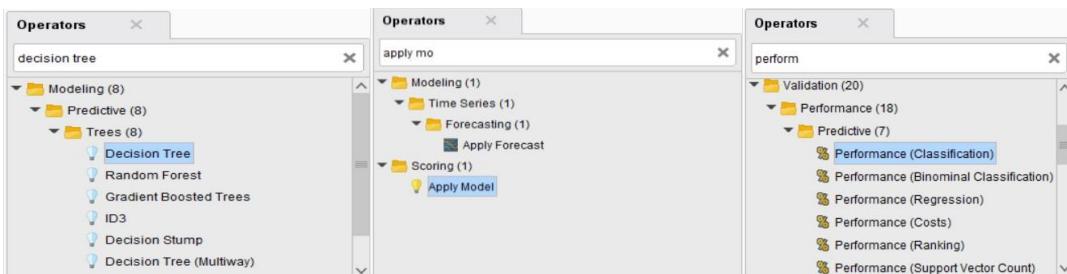


9.2. Pohon Keputusan Menggunakan RapidMiner

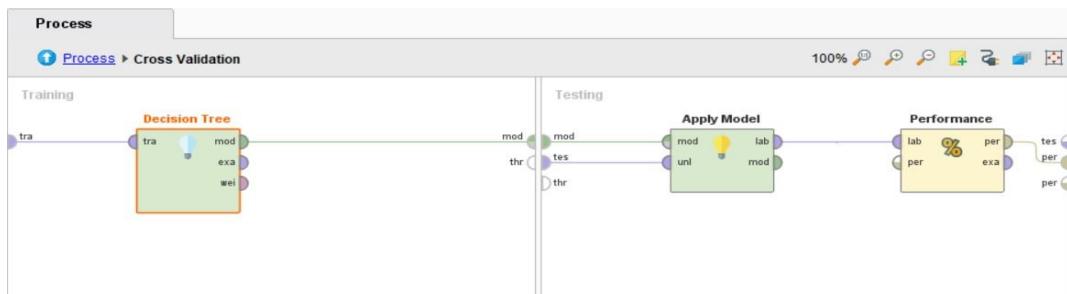
1. Buka aplikasi RapidMiner
2. Klik New Process (Ctrl + N) → Blank pada halaman prespektive RapidMiner.
3. Pada area Process View, kita mendesain model proses pohon keputusan menggunakan data **DataCuaca_Training**.
4. Drag **DataCuaca_Training** dan **Cross Validation** ke area Process View.



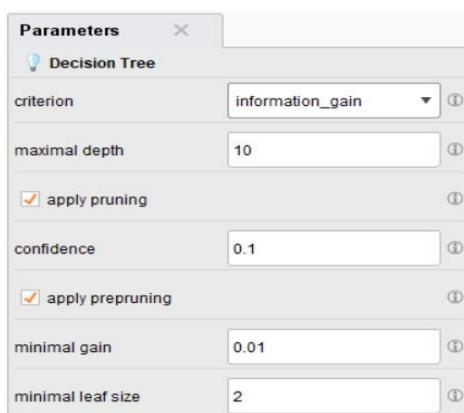
5. Klik ganda operator **Cross Validation** yang terdapat pada Process View.
6. Masukkan operator **Decision Tree** dalam area Training, operator **Apply Model** dan **Performance** dalam area Testing.



7. Hubungkan port input dan output masing-masing operator.



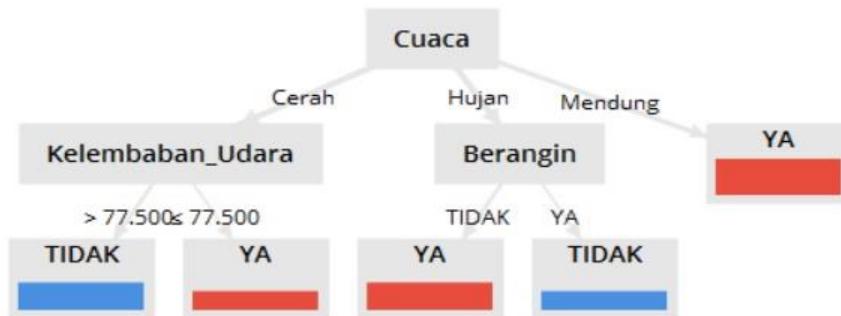
8. Klik Decision Tree pada area Training dan pastikan kriteria yang di pakai adalah Information Gain.



9. Tekan tombol Process yang terletak di atas area view untuk kembali ke desain awal. 10. Jalankan proses dengan menekan tombol **Run** (atau F11) 11. Berikut 2 hasil proses klasifikasi pohon keputusan :
- PerformanceVector (Performance)

	true TIDAK	true YA	class precision
pred. TIDAK	2	2	50.00%
pred. YA	3	7	70.00%
class recall	40.00%	77.78%	

- Tree (Decision Tree)

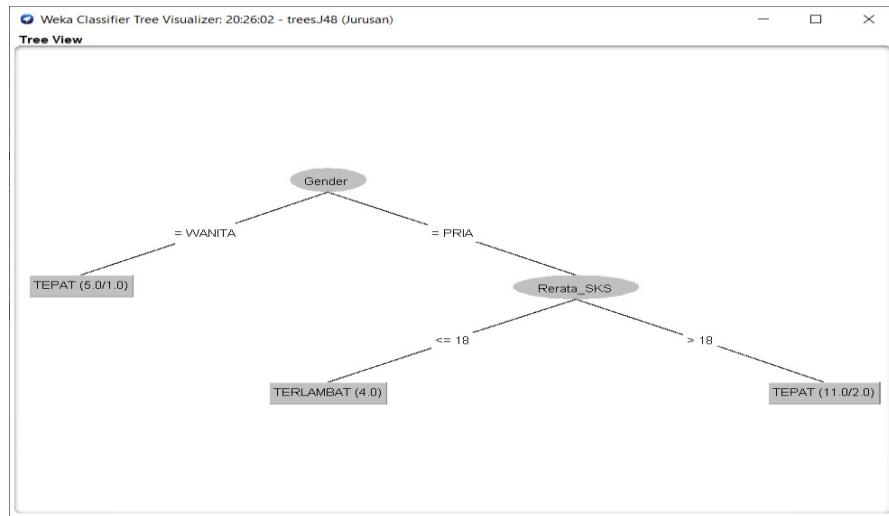


Tugas

- Berdasarkan pohon keputusan pada kegiatan 9.2 (menggunakan RapidMiner), isikan nilai kelas atribut Bermain_Tenis pada tabel Testing berikut :

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

- Gunakan file ARFF yang di kerjakan pada Tugas nomor1 dalam Modul 7 sebagai data training.
 - Buatlah dan cetaklah pohon keputusan berdasarkan data tersebut !



b. Carilah nilai-nilai parameter berikut :

```

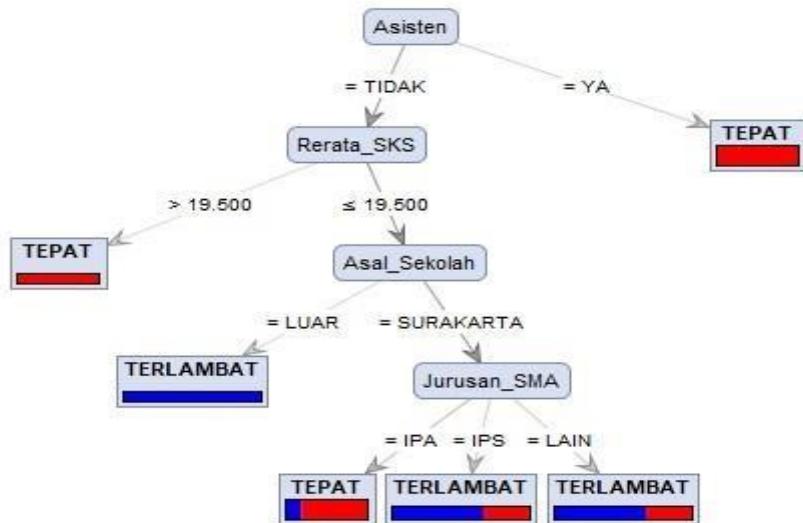
Classifier output
=====
J48 pruned tree
-----
Gender = WANITA: TEPAT (5.0/1.0)
Gender = PRIA
|   Rerata_SKS <= 18: TERLAMBAT (4.0)
|   Rerata_SKS > 18: TEPAT (11.0/2.0)

Number of Leaves :      3
Size of the tree :      5

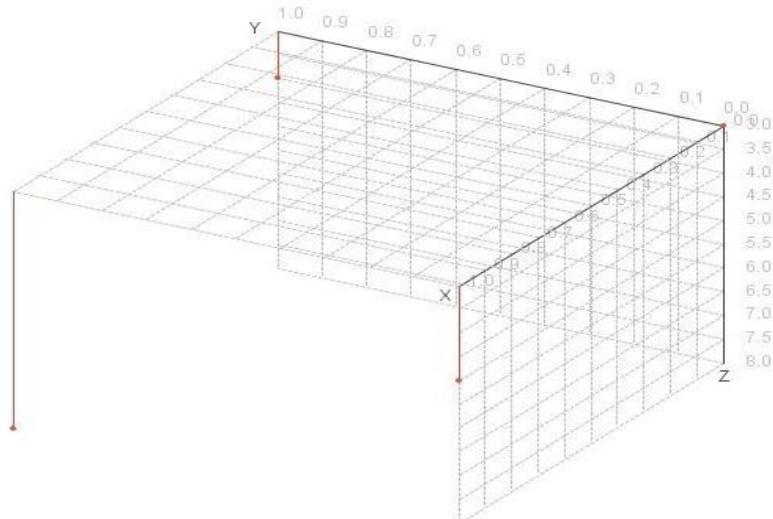
Time taken to build model: 0 seconds
==== Evaluation on training set ====
Time taken to test model on training data: 0 seconds
==== Summary ====
Correctly Classified Instances      17          85      %
Incorrectly Classified Instances     3           15      %
Kappa statistic                      0.6341
Mean absolute error                  0.2436
Root mean squared error              0.349
Relative absolute error              53.0693 %
Root relative squared error         73.1456 %
Total Number of Instances           20


```

- 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 = **15%**
3. Gunakan file Exel 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



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

- Seseorang akan lulus tepat waktu (TEPAT) jika kondisi sebagai berikut:
 - Asisten = YA (nilai atribut lain diabaikan)
 - Asisten = TIDAK, Rerata_SKS > 19.5
 - Asisten = TIDAK, Rerata_SKS <= 19.5, Asal_Sekolah = SURAKARTA, Jurusan_SMA = IPA
- Seseorang akan lulus tidak tepat waktu (TERLAMBAT) jika kondisi sebagai berikut:
 - Asisten = TIDAK, Rerata_SKS <= 19.5 , Asal_Sekolah = LUAR
 - Asisten = TIDAK, Rerata_SKS <=19.5, Jurusan_SMA = IPS/LAIN

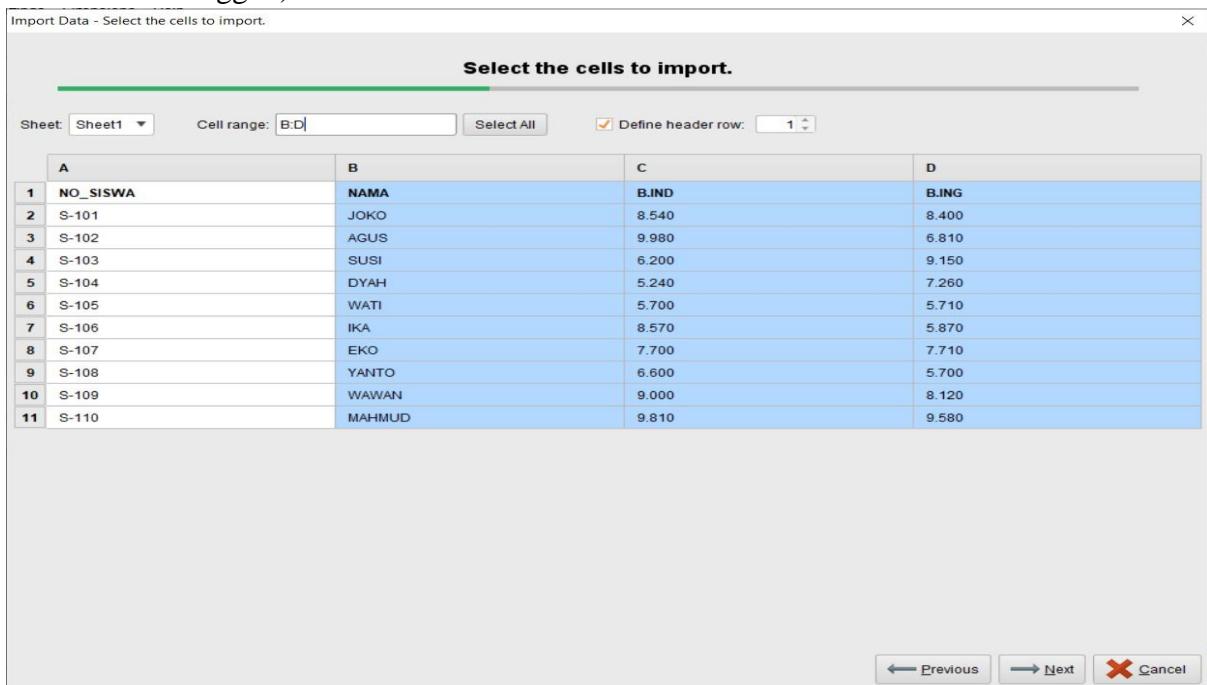
MODUL10

10.1. Algoritma K-Means Menggunakan Rapid Miner

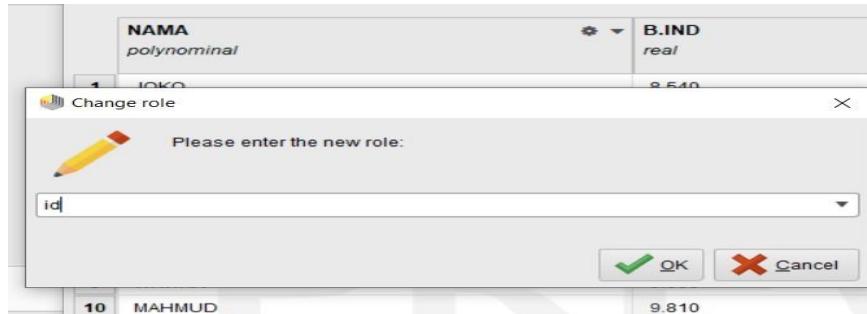
- Buatlah tabel data nilai ujian siswa dengan nama **Tabel_NilaiUjian.xls**.

A	B	C	D
1	NO_SISWA	NAMA	B.IND
2	S-101	JOKO	8,54
3	S-102	AGUS	9,98
4	S-103	SUSI	6,2
5	S-104	DYAH	5,24
6	S-105	WATI	5,7
7	S-106	IKA	8,57
8	S-107	EKO	7,7
9	S-108	YANTO	6,6
10	S-109	WAWAN	9
11	S-110	MAHMUD	9,81

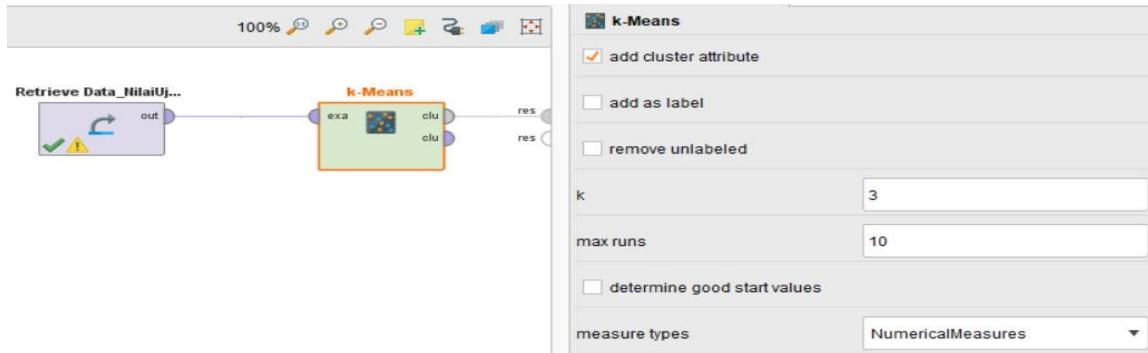
- Jalankan aplikasi RapidMiner
- Gunakan file **Tabel_NilaiUjian.xls** sebagai data yang digunakan dalam proses Clustering. Kita hanya menggunakan 3 kolom (nama siswa,nilai bahasa idonesia, nilai bahasa inggris).



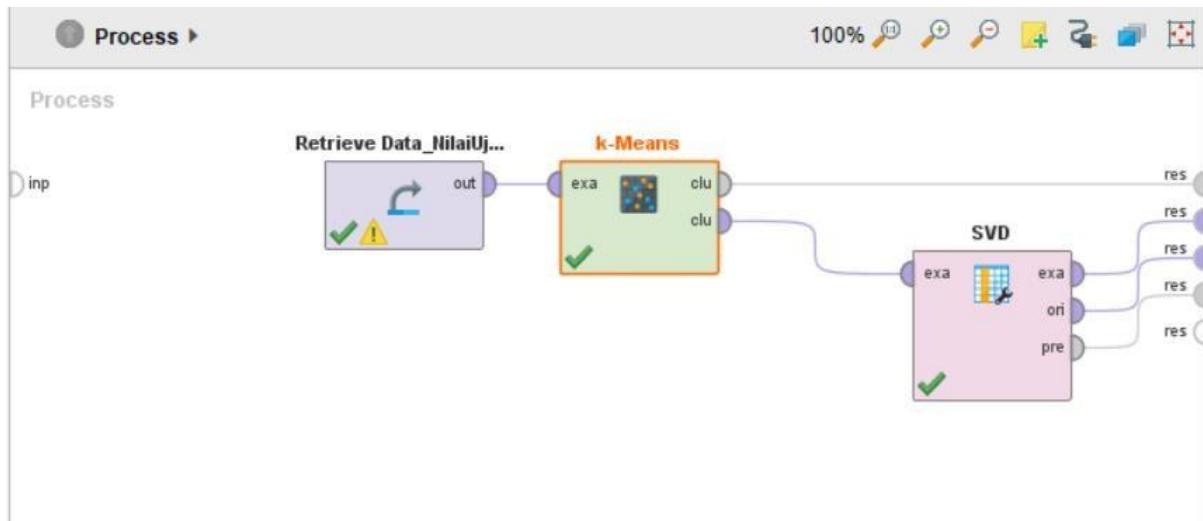
- Ubah kolom NAMA menjadi id, dengan cara klik **Change role >> id >> OK**.



5. Beri nama **Data_NilaiUjian** dan masukkan pada repositories. Kemudian klik **Finish**.
6. Gunakan **Data_NilaiUjian** ini dan masukkan ke dalam area process.
7. Tambahkan operator **k-Means**. Ubah nama operator ini menjadi k-Means. Hubungkan seperti gambar di bawah. Ubah parameter k=3 pada operator ini.



8. Tambahkan pula operator **SDV(Singular Value Decomposition)**.



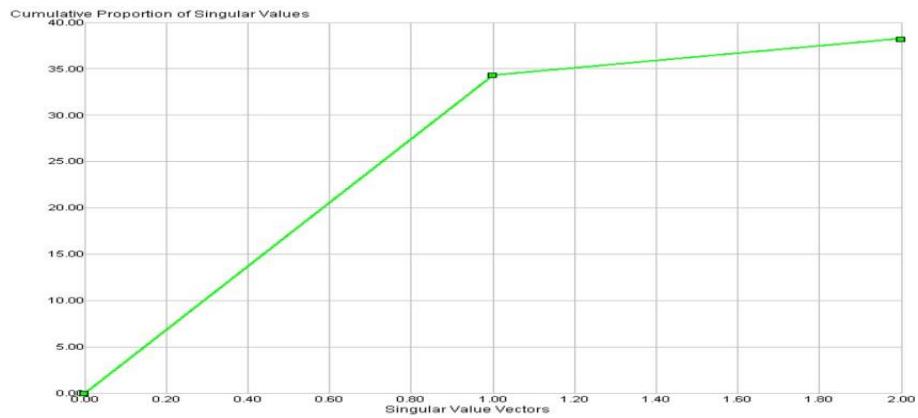
9. Jalankan dengan menekan tombol Run (F11).
10. Berikut hasil proses Clustering dengan algoritma k-Means :
 - a. **SDV(Singular Value Decomposition) i. Nilai Eigenvalue**

Component	Singular Value	Proportion of Singul...	Cumulative Singular ...	Cumulative Proportio...
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

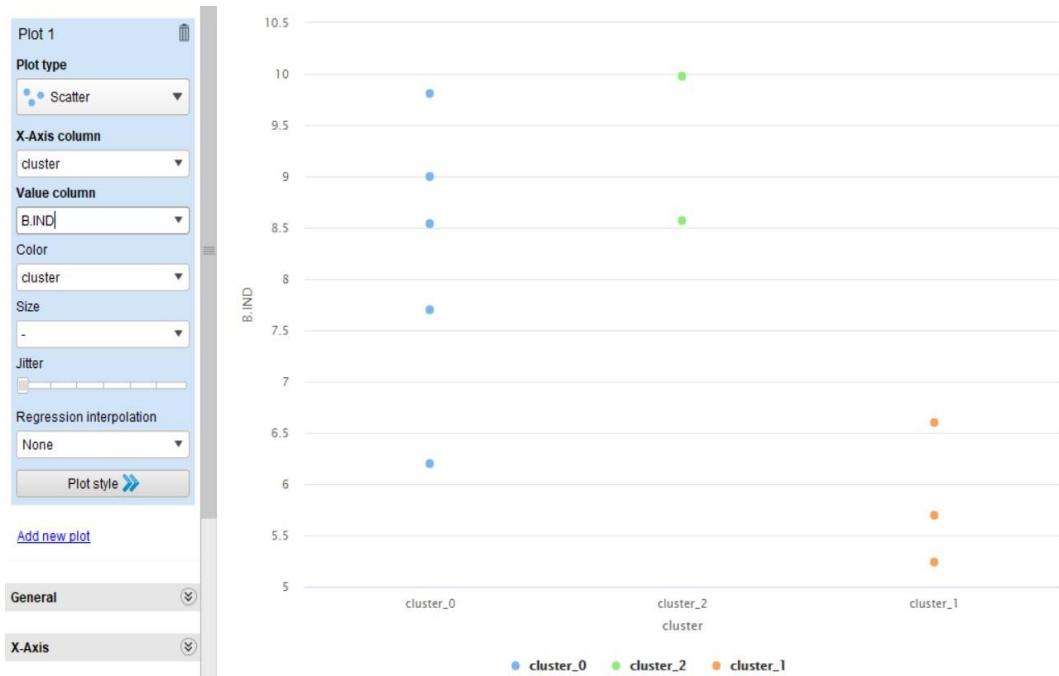


b. Example (k-Means)

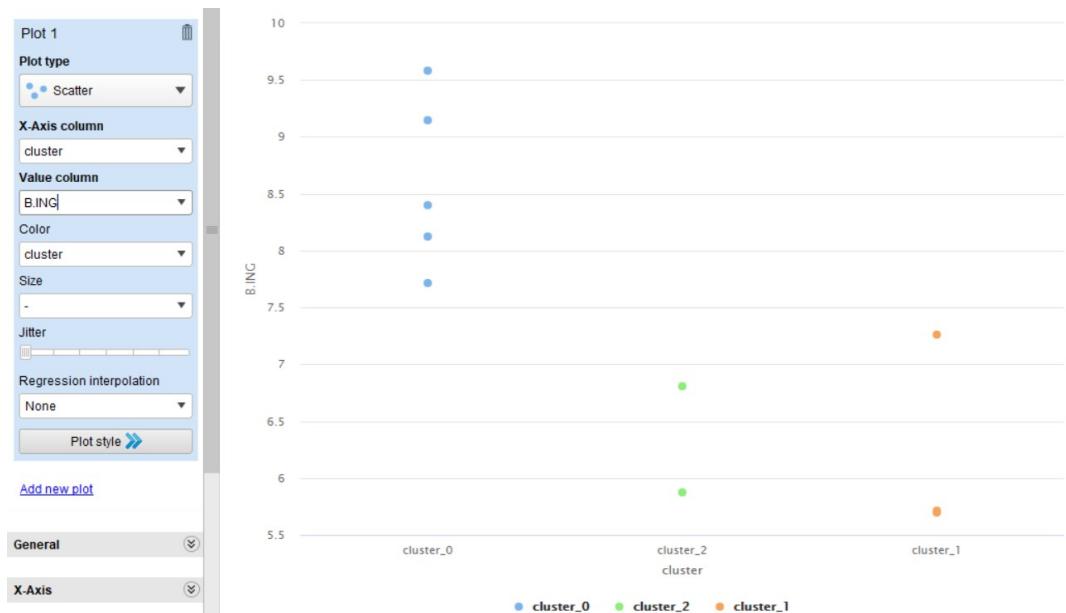
Hasil ini kita lihat dengan mode Plot View menggunakan grafik Scatter untuk menentukan kelompok siswa (cluster) yang dicalonkan untuk maju ke dalam olimpiade mata pelajaran berdasarkan nilai tertinggi ujian. Ketentuan :

Plotter	= Scatter
x-Axis	= cluster
y-Axis	= B.IND, B.ING(diubah-ubah)
Color Column	= cluster
Jitter	detail. = bisa diubah-ubahuntuk melihat distribusi data secara lebih

i. Kelompok siswa bidang B. Indonesia



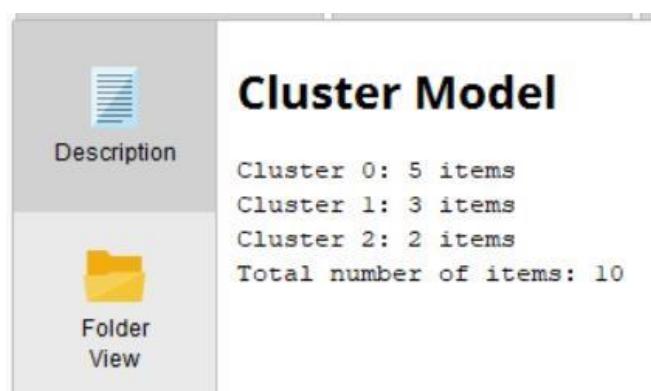
ii. Kelompok siswa bidang B. Inggris



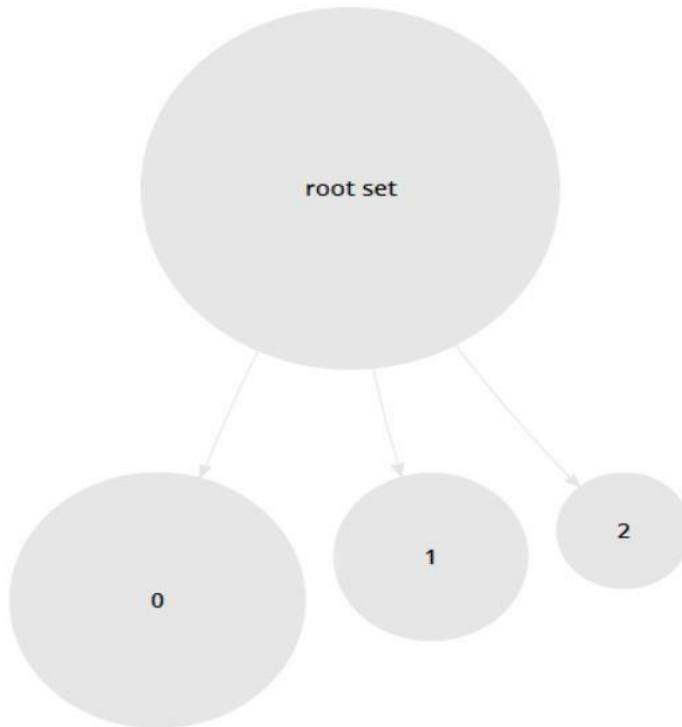
c. ExampleSet (SDV)

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

d. Cluster Model (Clustering) i. Description



ii. Graph



10.2. Interpretasi Hasil Algoritma K-Means

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

Tugas

1. Buatlah tabel berikut dengan menggunakan Microsoft Excel!

1	NO_SISWA	NAMA	B.IND	B.ING
2	S-101	JOKO	6,68	9,21
3	S-102	AGUS	9,02	6,42
4	S-103	SUSI	7,95	6,04
5	S-104	DYAH	6,65	6,89
6	S-105	WATI	7,40	7,59
7	S-106	IKA	8,46	7,63
8	S-107	EKO	9,48	9,16
9	S-108	YANTO	8,17	9,04
10	S-109	WAWAN	5,95	5,80
11	S-110	MAHMUD	6,51	8,94
12	S-111	BUDI	5,25	6,83
13	S-112	SANTI	6,50	7,17
14	S-113	DIAN	7,43	9,31
15	S-114	DANI	8,04	9,88
16	S-115	AHMAD	5,12	6,35
17	S-116	BAYU	9,97	9,71
18	S-117	RISA	7,96	8,98
19	S-118	RANI	7,46	5,29
20	S-119	YANI	6,67	9,15
21	S-120	RATIH	9,84	5,12
22	S-121	INDAH	9,37	6,69
23	S-122	JONO	9,31	9,28
24	S-123	SARAH	7,20	9,08
25	S-124	RAMA	7,99	7,05
26	S-125	BAMBANG	7,57	5,38
27	S-126	HADI	8,61	8,64
28	S-127	NANA	9,99	6,74
29	S-128	FEBRI	7,43	5,55
30	S-129	DENI	8,15	7,51
31	S-130	TONI	6,26	8,05

2. Ketentuan jumlah Cluster = 4.

Hasil :

- a. SDV (Singular Value Decomposition)

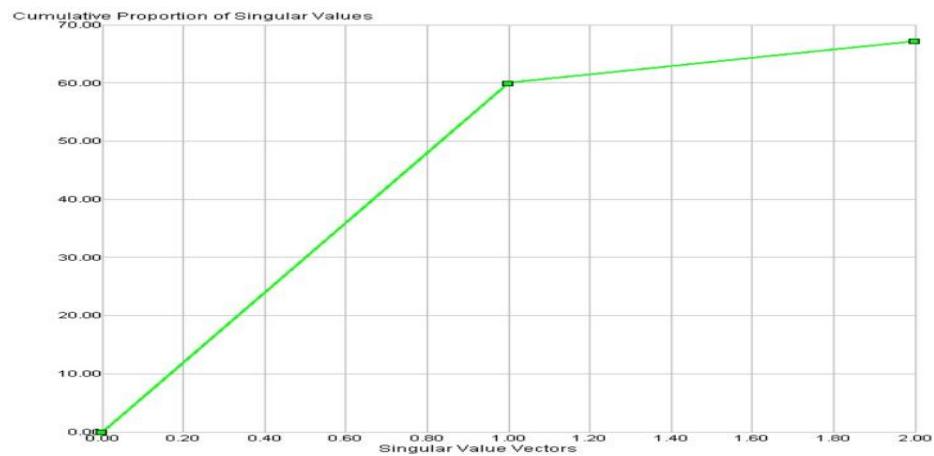
- i. Nilai Eigenvalue

Component	Singular Value	Proportion of Si...	Cumulative Sin...	Cumulative Pro...
SVD 1	60.034	0.893	60.034	0.893
SVD 2	7.199	0.107	67.233	1.000

- ii. Nilai Sdv vectors

Attribute	SVD Vector 1
B.IND	0.712
B.ING	0.702

iii. Nilai Cumulative Variance

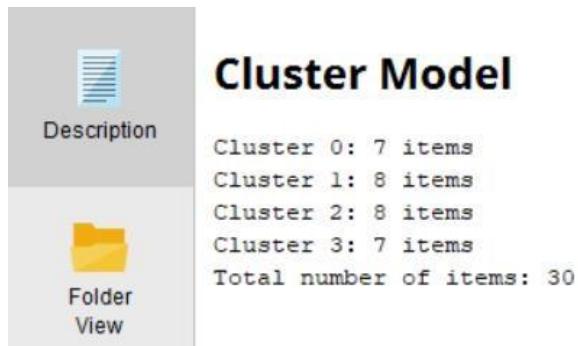


b. ExampleSet (SDV)

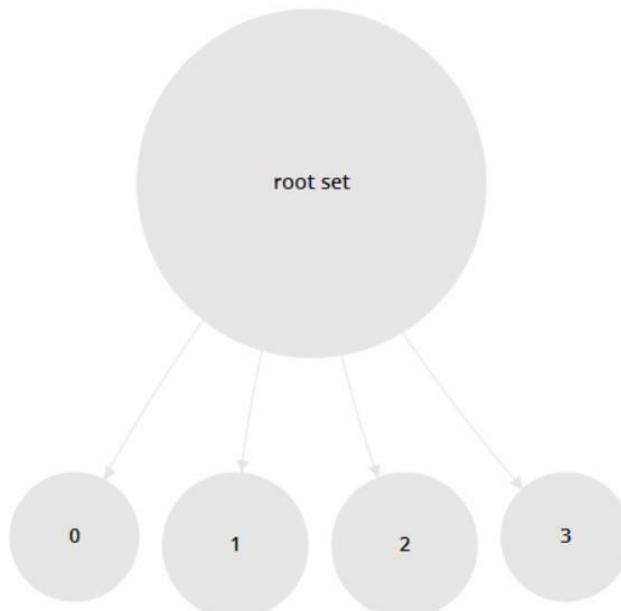
Row No.	NAMA	cluster ↑	svd_1
1	JOKO	cluster_0	0.187
5	WATI	cluster_0	0.177
10	MAHMUD	cluster_0	0.182
13	DIAN	cluster_0	0.197
19	YANI	cluster_0	0.186
23	SARAH	cluster_0	0.192
30	TONI	cluster_0	0.168
4	DYAH	cluster_1	0.159
9	WAWAN	cluster_1	0.138
11	BUDI	cluster_1	0.142
12	SANTI	cluster_1	0.161
15	AHMAD	cluster_1	0.135
18	RANI	cluster_1	0.150
25	BAMBANG	cluster_1	0.153
28	FEBRI	cluster_1	0.153
2	AGUS	cluster_2	0.182
3	SUSI	cluster_2	0.165
6	IKA	cluster_2	0.190
20	RATIH	cluster_2	0.177
21	INDAH	cluster_2	0.189
24	RAMA	cluster_2	0.177
27	NANA	cluster_2	0.197
29	DENI	cluster_2	0.185
7	EKO	cluster_3	0.220
8	YANTO	cluster_3	0.203
14	DANI	cluster_3	0.211
16	BAYU	cluster_3	0.232
17	RISA	cluster_3	0.199
22	JONO	cluster_3	0.219
26	HADI	cluster_3	0.203

c. Description

i. Description



ii. Graph



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

- Cluster 0 : JOKO, WATI, MAHMUD, DIAN, YANI, SARAH, TONI
- Cluster 1 : DYAH, WAWAN, BUDI, SANTI, AHMAD, RANI, BAMBANG, FEBRI
- Cluster 2 : AGUS, SUSI, IKA, RATIH, INDAH, RAMA, NANA, DENI
- Cluster 3 : EKO, YANTO, DANI, BAYU, RISA, JONO, HADI

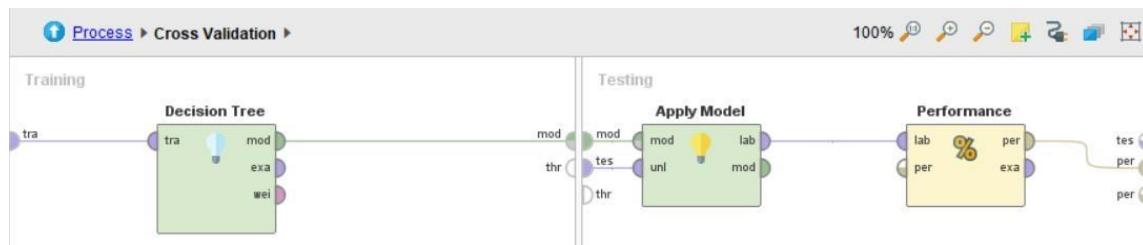
MODUL11

11.1. Induksi Aturan Data Cuaca

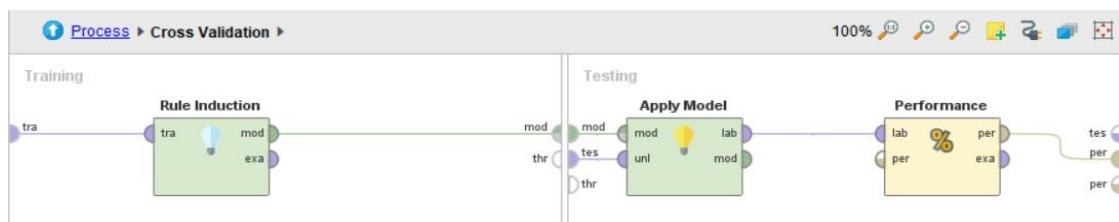
1. Jalankan aplikasi RapidMiner.
2. Gunakan model proses dari praktikum data mining **Modul 9 Kegiatan 9.4.2**.
3. Dari hasil tersebut menghasilkan sebuah pohon keputusan seperti pada gambar berikut



4. Kembali ke Process – Cross Validation.



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



7. Jalankan dengan menekan tombol **Run** (atau F11).
8. Sehingga akan diperoleh sebuah induksi yang disebut sebagai **RuleModel (Rule Induction)**.

RuleModel

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

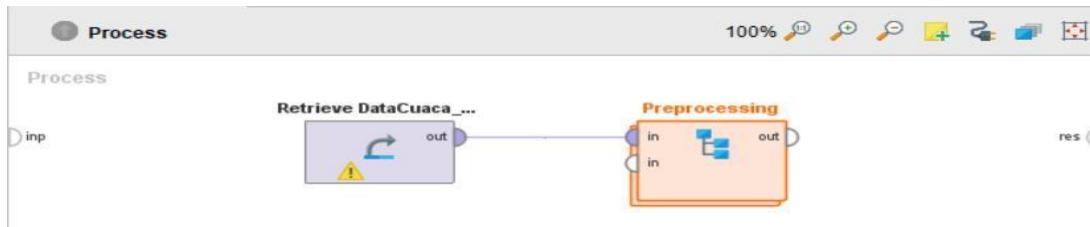
correct: 12 out of 13 training examples.
```

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

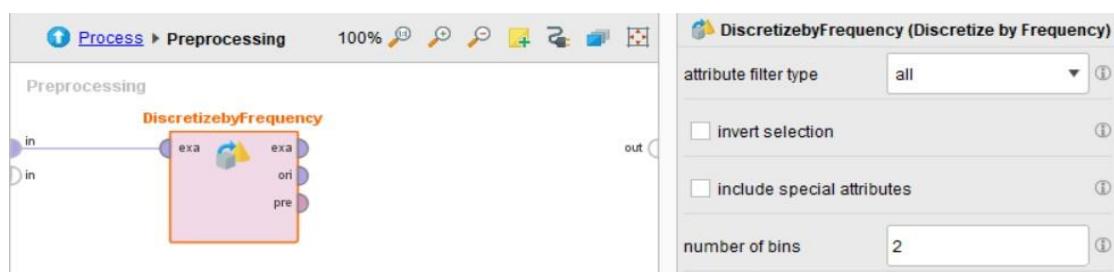
accuracy: 65.00% +/- 47.43% (micro average: 71.43%)			
	true TIDAK	true YA	class precision
pred. TIDAK	2	1	66.67%
pred. YA	3	8	72.73%
class recall	40.00%	88.89%	

11.2. Aturan Asosiasi Data Cuaca

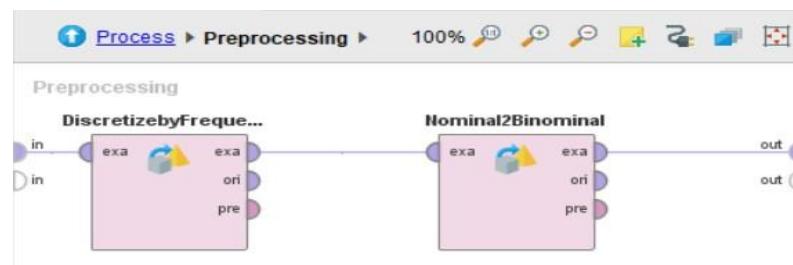
1. Masih dengan menggunakan RapidMiner, pilih **New Process**.
2. Gunakan **DataCuaca_Training** dan drag dari *repository* ke area **Process View**.
3. Tambahkan operator **Utility → Subprocess** ke dalam area. Ubah nama operator menjadi **Preprocessing**. Hubungkan port output Retrieve dengan port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



4. Klik ganda operator **Preprocessing** sehingga masuk pada area Nested Chain.
5. Pada Nested Chain ini tambahkan operator-operator berikut :
 - a. **Cleansing → Binning → Discretize by Frequency** ubah nama operator ini menjadi “**DiscretizebyFrequency**” dan biarkan nilai parameter **number of bins** (jumlah interval) = 2. Hubungkan seperti pada gambar.



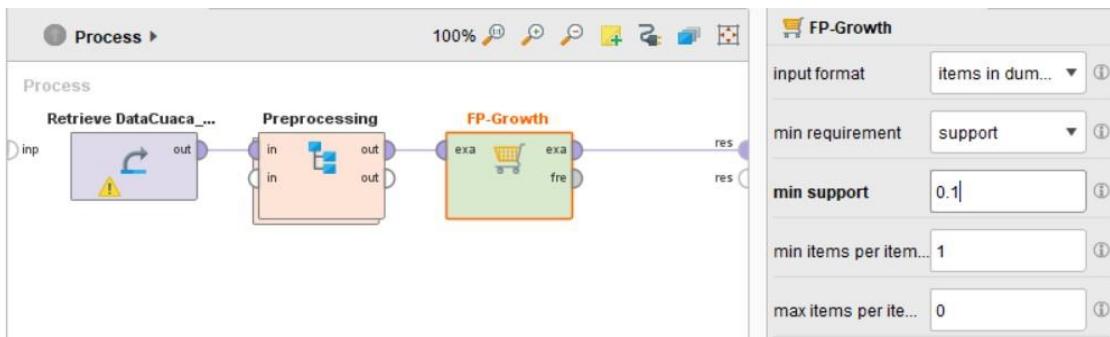
- b. **Blending → Attribute → Types → Nominal to Binomial**
Ubah nama pada operator ini menjadi “**Nominal2Binomial**”, dan hubungkan seperti pada gambar.



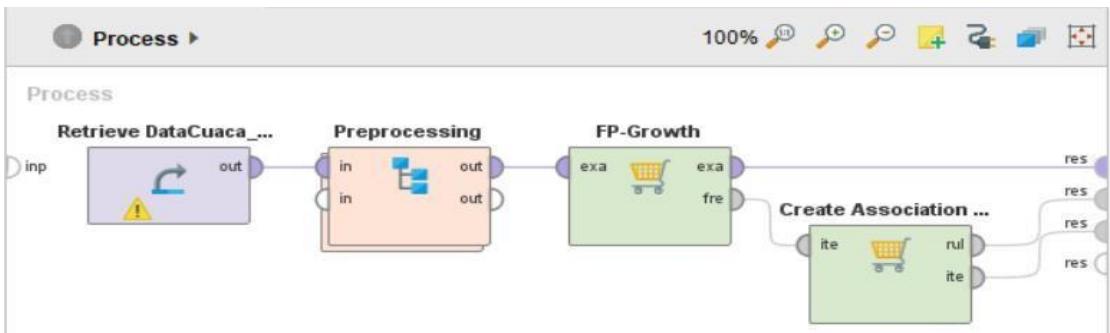
6. Kembali ke *main process*, tambahkan 2 buah operator:

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

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



b. **Modelling → Association → Create Association Rules** Hubungkan seperti pada gambar.



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

a. Frequent Item Set (FP-Growth)

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

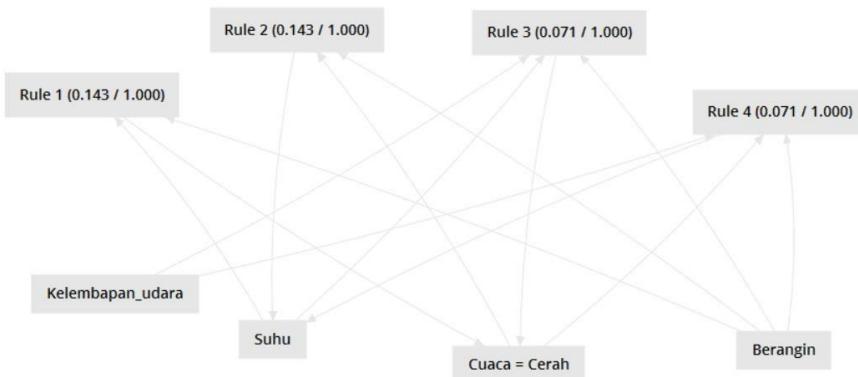
Dapat dilihat bahwa jumlah aturan asosiasi yang terbentuk adalah 28 set, dan jumlah total max size = 4, yang terdiri dari 4 buah itemset.

b. Association Rules (Create Assosiation Rules)

i. Table View

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

ii. Graph View



c. Exampleset (Nominal2Binomial) → Charts View



Tugas

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

- a. *Induction Rule* (Rule Model)

RuleModel

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

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

- b. Nilai *Performance Vektor*

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

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

2. Masih dengan menggunakan data *training* yang sama dengan ketentuan bahwa pada operator *Discretize by Frequency* memiliki nilai : a. *Number of bins = 2*
 - Frequent Item Set (FP-Growth)

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

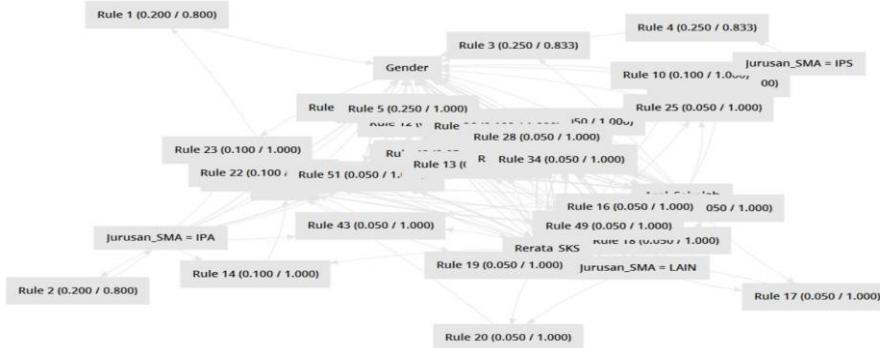
Dapat dilihat bahwa jumlah aturan asosiasi yang terbentuk adalah 55 set dan jumlah total max size = 5, yang terdiri dari 5 buah itemset.

- Association Rules
 - Tabel View

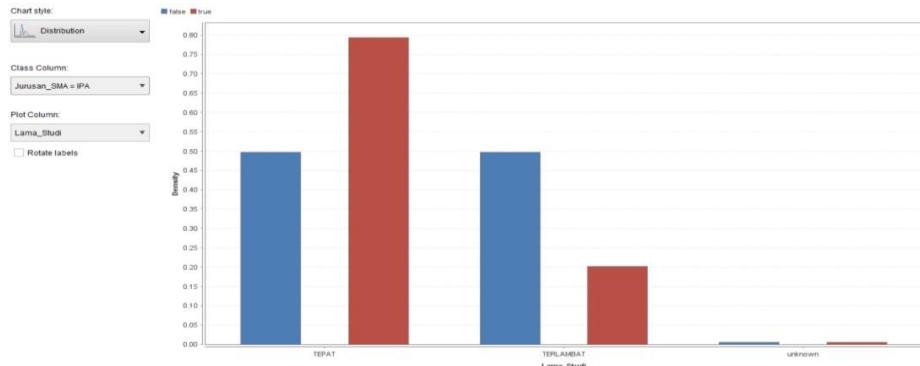
No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s
3	Asal_Sekolah	Gender	0.250	0.833	0.962	-0.350	0.025
4	Jurusan_SMA = IPS	Gender	0.250	0.833	0.962	-0.350	0.025
5	Rerata_SKS	Gender	0.250	1	1	-0.250	0.062
6	Jurusan_SMA = IPA, Rerata_SKS	Gender	0.100	1	1	-0.100	0.025
7	Asal_Sekolah, Jurusan_SMA = IPS	Gender	0.100	1	1	-0.100	0.025
8	Asal_Sekolah, Rerata_SKS	Gender	0.150	1	1	-0.150	0.038
9	Asal_Sekolah, Jurusan_SMA = LAIN	Gender	0.050	1	1	-0.050	0.012
10	Jurusan_SMA = IPS, Rerata_SKS	Gender	0.100	1	1	-0.100	0.025
11	Asisten, Rerata_SKS	Gender	0.150	1	1	-0.150	0.038
12	Asisten, Jurusan_SMA = LAIN	Gender	0.050	1	1	-0.050	0.012
13	Rerata_SKS, Jurusan_SMA = LAIN	Gender	0.050	1	1	-0.050	0.012
14	Jurusan_SMA = IPA, Rerata_SKS	Asisten	0.100	1	1	-0.100	0.075
15	Asal_Sekolah, Jurusan_SMA = LAIN	Asisten	0.050	1	1	-0.050	0.038
16	Asisten, Jurusan_SMA = LAIN	Asal_Sekolah	0.050	1	1	-0.050	0.035
17	Asal_Sekolah, Jurusan_SMA = LAIN	Rerata_SKS	0.050	1	1	-0.050	0.038
18	Rerata_SKS, Jurusan_SMA = LAIN	Asal_Sekolah	0.050	1	1	-0.050	0.035
19	Asisten, Jurusan_SMA = LAIN	Rerata_SKS	0.050	1	1	-0.050	0.038
20	Rerata_SKS, Jurusan_SMA = LAIN	Asisten	0.050	1	1	-0.050	0.038
21	Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS	Gender	0.050	1	1	-0.050	0.012
22	Jurusan_SMA = IPA, Rerata_SKS	Gender, Asisten	0.100	1	1	-0.100	0.080
23	Gender, Jurusan_SMA = IPA, Rerata_SKS	Asisten	0.100	1	1	-0.100	0.075
24	Jurusan_SMA = IPA, Asisten, Rerata_SKS	Gender	0.100	1	1	-0.100	0.025
25	Asal_Sekolah, Jurusan_SMA = IPS, Rerata_SKS	Gender	0.050	1	1	-0.050	0.012
26	Gender, Asal_Sekolah, Asisten	Rerata_SKS	0.100	1	1	-0.100	0.075
27	Asal_Sekolah, Asisten, Rerata_SKS	Gender	0.100	1	1	-0.100	0.025
28	Asal_Sekolah, Jurusan_SMA = LAIN	Gender, Asisten	0.050	1	1	-0.050	0.040
29	Gender, Asal_Sekolah, Jurusan_SMA = LAIN	Asisten	0.050	1	1	-0.050	0.038
30	Asisten, Jurusan_SMA = LAIN	Gender, Asal_Sekolah	0.050	1	1	-0.050	0.038
31	Gender, Asisten, Jurusan_SMA = LAIN	Asal_Sekolah	0.050	1	1	-0.050	0.035
32	Asal_Sekolah, Asisten, Jurusan_SMA = LAIN	Gender	0.050	1	1	-0.050	0.012
33	Asal_Sekolah, Jurusan_SMA = LAIN	Gender, Rerata_SKS	0.050	1	1	-0.050	0.038
34	Gender, Asal_Sekolah, Jurusan_SMA = LAIN	Rerata_SKS	0.050	1	1	-0.050	0.038
35	Rerata_SKS, Jurusan_SMA = LAIN	Gender, Asal_Sekolah	0.050	1	1	-0.050	0.038
36	Gender, Rerata_SKS, Jurusan_SMA = LAIN	Asal_Sekolah	0.050	1	1	-0.050	0.035
37	Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN	Gender	0.050	1	1	-0.050	0.012
38	Asisten, Jurusan_SMA = LAIN	Gender, Rerata_SKS	0.050	1	1	-0.050	0.038
39	Gender, Asisten, Jurusan_SMA = LAIN	Rerata_SKS	0.050	1	1	-0.050	0.038
40	Rerata_SKS, Jurusan_SMA = LAIN	Gender, Asisten	0.050	1	1	-0.050	0.040
41	Gender, Rerata_SKS, Jurusan_SMA = LAIN	Asisten	0.050	1	1	-0.050	0.038
42	Asisten, Rerata_SKS, Jurusan_SMA = LAIN	Gender	0.050	1	1	-0.050	0.012
43	Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS	Asisten	0.050	1	1	-0.050	0.038
44	Asal_Sekolah, Jurusan_SMA = LAIN	Asisten, Rerata_SKS	0.050	1	1	-0.050	0.043
45	Asisten, Jurusan_SMA = LAIN	Asal_Sekolah, Rerata_SKS	0.050	1	1	-0.050	0.043
46	Asal_Sekolah, Asisten, Jurusan_SMA = LAIN	Rerata_SKS	0.050	1	1	-0.050	0.038
47	Rerata_SKS, Jurusan_SMA = LAIN	Asal_Sekolah, Asisten	0.050	1	1	-0.050	0.043
48	Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN	Asisten	0.050	1	1	-0.050	0.038
49	Asisten, Rerata_SKS, Jurusan_SMA = LAIN	Asal_Sekolah	0.050	1	1	-0.050	0.035
50	Gender, Jurusan_SMA = IPA, Asal_Sekolah, Asisten	Rerata_SKS	0.050	1	1	-0.050	0.038
51	Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS	Gender, Asisten	0.050	1	1	-0.050	0.040
52	Gender, Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS	Asisten	0.050	1	1	-0.050	0.038
53	Jurusan_SMA = IPA, Asal_Sekolah, Asisten, Rerata_SKS	Gender	0.050	1	1	-0.050	0.012
54	Asal_Sekolah, Jurusan_SMA = LAIN	Gender, Asisten, Rerata_SKS	0.050	1	1	-0.050	0.043
55	Gender, Asal_Sekolah, Jurusan_SMA = LAIN	Asisten, Rerata_SKS	0.050	1	1	-0.050	0.043
56	Asisten, Jurusan_SMA = LAIN	Gender, Asal_Sekolah	0.050	1	1	-0.050	0.043
57	Gender, Asisten, Jurusan_SMA = LAIN	Asal_Sekolah, Rerata_SKS	0.050	1	1	-0.050	0.043
58	Asal_Sekolah, Asisten, Jurusan_SMA = LAIN	Gender, Rerata_SKS	0.050	1	1	-0.050	0.038
59	Gender, Asal_Sekolah, Asisten, Jurusan_SMA = LAIN	Rerata_SKS	0.050	1	1	-0.050	0.038
60	Rerata_SKS, Jurusan_SMA = LAIN	Gender, Asal_Sekolah	0.050	1	1	-0.050	0.045
61	Gender, Rerata_SKS, Jurusan_SMA = LAIN	Asal_Sekolah, Asisten	0.050	1	1	-0.050	0.043
62	Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN	Gender, Asisten	0.050	1	1	-0.050	0.040
63	Gender, Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN	Asisten	0.050	1	1	-0.050	0.038
64	Asisten, Rerata_SKS, Jurusan_SMA = LAIN	Gender, Asal_Sekolah	0.050	1	1	-0.050	0.038
65	Gender, Asisten, Rerata_SKS, Jurusan_SMA = LAIN	Asal_Sekolah	0.050	1	1	-0.050	0.035
66	Asal_Sekolah, Asisten, Rerata_SKS, Jurusan_SMA = LAIN	Gender	0.050	1	1	-0.050	0.012

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

- Graph View



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



a. Number of bins = 3

- Frequent Item Set (FP-Growth)

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

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

- Association Rules

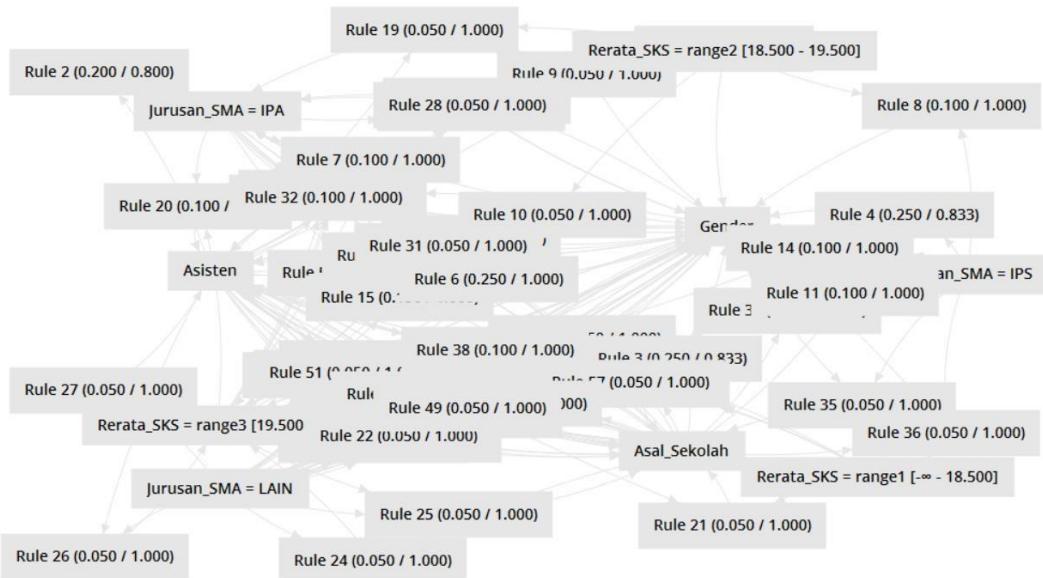
- Tabel View

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

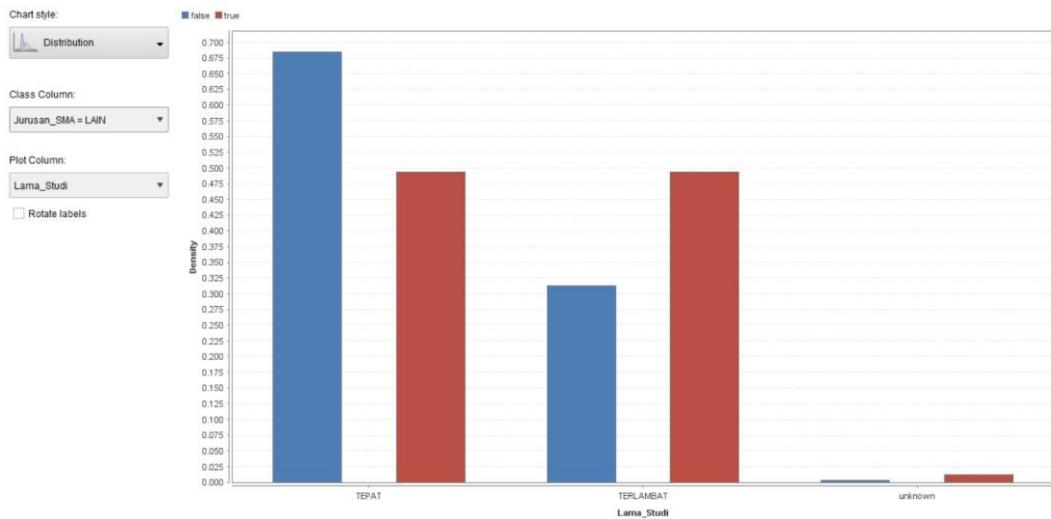
41	Gender, Asal_Sekolah, Jurusan_SMA = LAIN	Asisten	0.050	1	1	-0.050	0.038
42	Asisten, Jurusan_SMA = LAIN	Gender, Asal_Sek...	0.050	1	1	-0.050	0.038
43	Gender, Asisten, Jurusan_SMA = LAIN	Asal_Sekolah	0.050	1	1	-0.050	0.035
44	Asal_Sekolah, Asisten, Jurusan_SMA = LAIN	Gender	0.050	1	1	-0.050	0.012
45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

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

- Graph View



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



MODUL12

Mencari Nilai t-hitung dan Model Regresi Linier

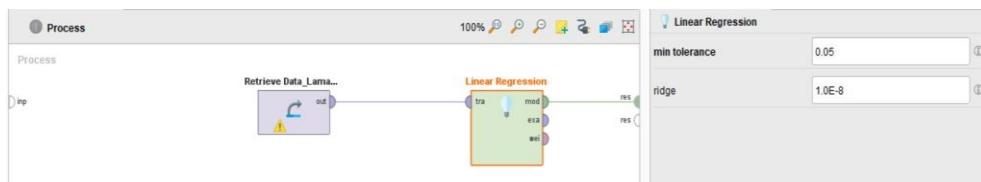
1. Buka Ms.Excel, dan buatlah tabel data siswa. Simpan dengan nama **Tabel_LamaBelajardanNilaiUjian.xls**.

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

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

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

5. Pada langkah ini, beri nama **Data_LamaBelajardanNilaiUjian** dan masukkan pada repository, klik **Finish**.
6. Gunakan **Data_LamaBelajardanNilaiUjian** ini dan masukkan ke dalam area process.
7. Tambahkan operator **Modelling→Predictive→Functions→Linier Regression**.
Hubungkan port sesuai pada gambar.
8. Klik pada operator **Linier Regression**, tentukan parameter **min tolerance = 0.05**.



9. Jalankan proses dengan menekan tombol Run (atau F11).
10. Hasil proses regresi linier :

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

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
LAMA BELAJAR (JAM)	21.608	7.845	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

```

Model regresi linier yang terbentuk :

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

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

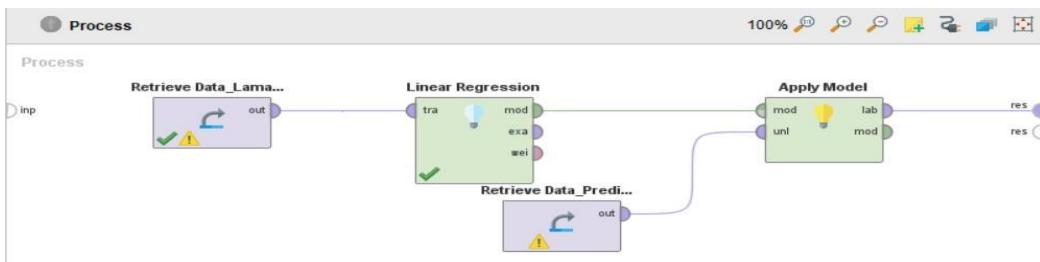
NO_SISWA	NAMA	LAMA BELAJAR (JAM)	NILAI	Nilai Belajar
S-101	JOKO	15	783	816,889
S-102	AGUS	18	877	881,713
S-103	SUSI	7	505	644,025
S-104	DYAH	9	860	687,241
S-105	WATI	15	968	816,889
S-106	IKA	17	793	860,105
S-107	EKO	10	752	708,849
S-108	YANTO	5	571	600,809
S-109	WAWAN	8	667	665,633
S-110	MAHMUD	15	723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

Mencari Nilai t- dan Model Regresi Linier Menggunakan RapidMiner

1. Buka Ms.Excel, dan buatlah tabel data siswa tersebut. Simpan dengan nama **Tabel_PrediksiNilaiUjian.xls**
2. Jalankan **RapidMiner**.
3. Gunakan file **Tabel_PrediksiNilaiUjian.xls** sebagai data testing. Import file ke dalam repository.
4. Ubah tipe data dan jenis masing-masing atribut sebagai berikut :

NO_SISWA polynomial <i>id</i>	NAMA polynomial	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

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

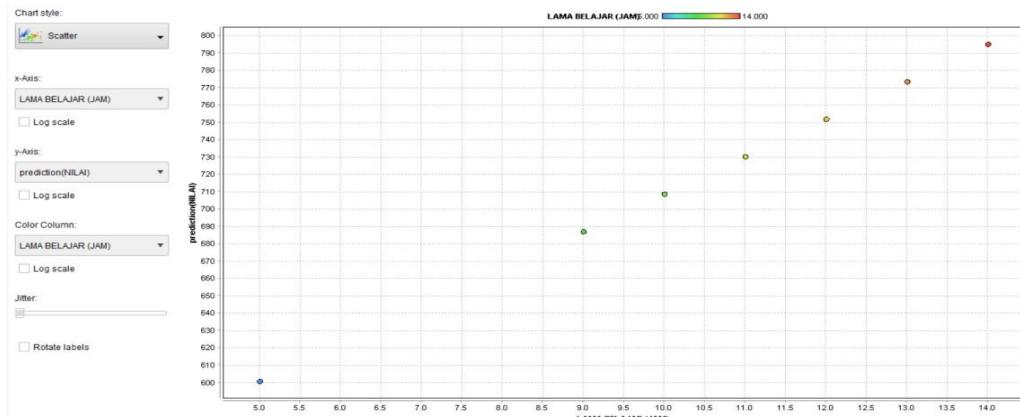


8. Dengan menggunakan parameter yang sama pada operator **Regression Linear**, jalankan proses dengan tombol **Run**.
9. Hasil proses prediksi terhadap data testing menggunakan regresi linear :
 - a. Data View (hasil prediksi nilai ujian)

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

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

- b. Chart View (Scatter Plot)



Pembuktian Model Regresi

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

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

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

Tugas

1. Buatlah tabel dengan menggunakan Microsoft Excel seperti berikut :

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

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

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

Linear Regression

151688.679 * JUMLAH ANGGOTA KELUARGA
 + 98652.830

3. Tentukan apakah variable X1 dan X2 mempengaruhi secara signifikan terhadap nilai variable Y berdasarkan nilai t-stat?

Dari tabel tersebut dapat dilihat nilai t-stat sebesar 2,142. Berdasarkan aturan statistik, variabel X dikatakan mempengaruhi secara signifikan terhadap Y jika nilai t-stat > ttable.

4. Tuliskan model persamaan regresi linier sederhana yang terbentuk!

$$Y = 151688,679 X_2 + 98652,830$$

5. Gunakan data testing untuk menjawab perintah :

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

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

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

