

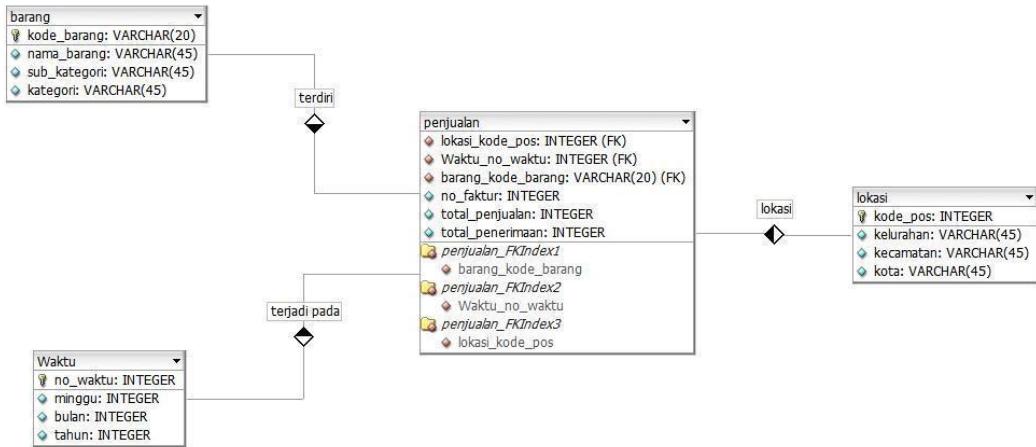
Nama : Fischella Angieta C.

NIM : L200170077

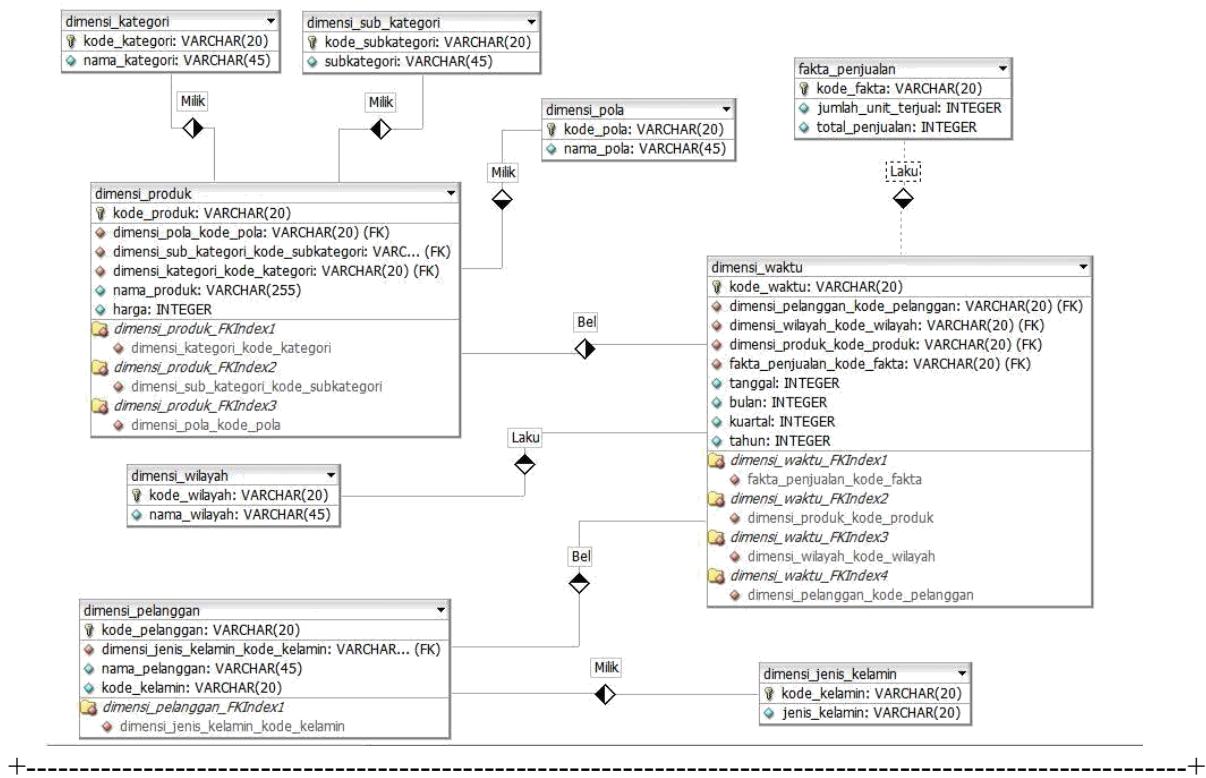
Kelas : C

MODUL I PERANCANGAN STAR SCHEMA DAN SNOWFLAKE

D. Langkah-langkah Praktikum



E. Tugas



MODUL V

PIVOT TABLE DAN CHART

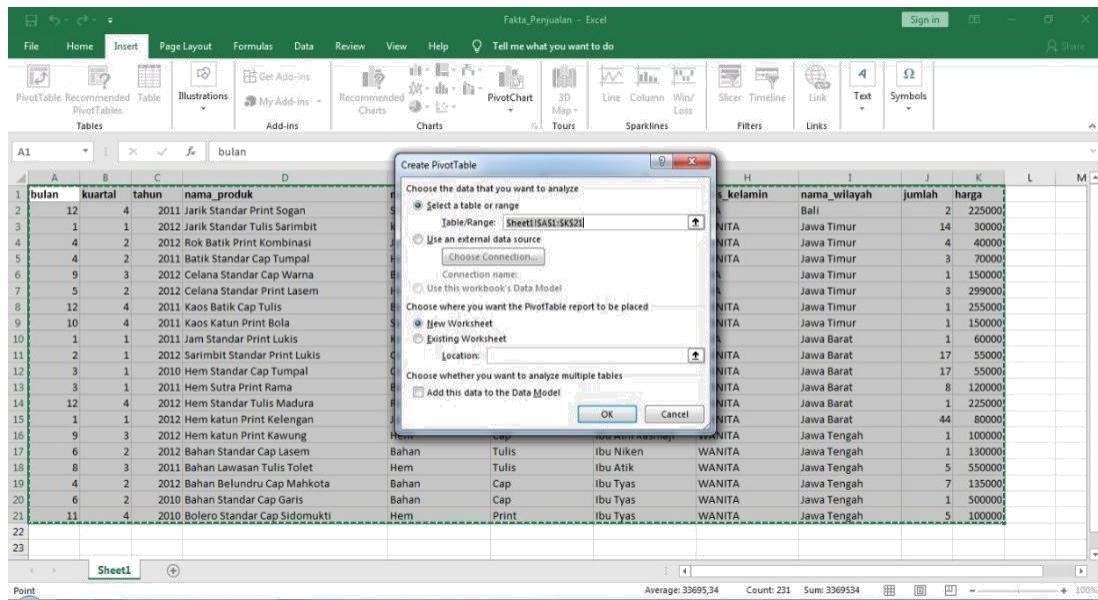
1. Langkah-langkah Praktikum D. 1. Kegiatan 1 : Membuat Pivot Table

1. Membuat data Excel dan menyimpannya dengan nama ‘Fakta_Penjualan.xls’.

The screenshot shows an Excel spreadsheet titled "Fakta_Penjualan - Excel". The data is organized into columns representing dimensions and facts:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	bulan	kuartal	tahun	nama_produk	nama_subkategori	nama_pola	nama_pelanggan	jenis_kelamin	nama_wilayah	jumlah	harga		
2	12	4	2011	Jarik Standar Print Sogan	Standar	Print	Bapak Ketut	PRIA	Bali	2	225000		
3	1		2012	Jarik Standar Tulis Sarimbit	kaos	Cap	Ibu Harini	WANITA	Jawa Timur	14	30000		
4	4	2	2012	Rok Batik Print Kombinasi	Jarik	Tulis	Ibu Harini	WANITA	Jawa Timur	4	40000		
5	4	2	2011	Batik Standar Cap Tumpal	Hem	Print	Ibu Harini	WANITA	Jawa Timur	3	70000		
6	9	3	2012	Celana Standar Cap Warna	Batik	Cap	Bapak Heru	PRIA	Jawa Timur	1	150000		
7	5	2	2012	Celana Standar Print Lasem	Hem	Print	Bapak Totok	PRIA	Jawa Timur	3	299000		
8	12	4	2011	Kaos Batik Cap Tulis	Bolero	Cap	Ibu Hatamah	WANITA	Jawa Timur	1	255000		
9	10	4	2011	Kaos Katun Print Bola	Sarimbit	Print	Ibu Hatamah	WANITA	Jawa Timur	1	150000		
10	1		2011	Jam Standar Print Lukis	Kaos	Print	Bapak Imron	PRIA	Jawa Barat	1	60000		
11	2		2012	Sarimbit Standar Print Lukis	Celana	Cap	Ibu Hadi Sukarni	WANITA	Jawa Barat	17	55000		
12	3	1	2010	Hem Standar Cap Tumpal	Celana	Print	Ibu Hadi Sukarni	WANITA	Jawa Barat	17	55000		
13	3	1	2012	Hem Sutra Print Rama	Bahan	Cap	Ibu Siti Arya	WANITA	Jawa Barat	8	120000		
14	12	4	2012	Hem Standar Tulis Madura	Rok	Print	Ibu Siti Arya	WANITA	Jawa Barat	1	225000		
15	1	1	2012	Hem Katun Print Kelengan	Jam	Print	Ibu Siti Arya	WANITA	Jawa Barat	44	80000		
16	9	3	2012	Hem Katun Print Kawung	Hem	Cap	Ibu Aini Kasmaji	WANITA	Jawa Tengah	1	100000		
17	6	2	2012	Bahan Standar Cap Lasem	Bahan	Tulis	Ibu Niken	WANITA	Jawa Tengah	1	130000		
18	8	3	2011	Bahan Lawasan Tulis Tolet	Hem	Tulis	Ibu Atik	WANITA	Jawa Tengah	5	550000		
19	4	2	2012	Bahan Beludru Cap Mahkota	Bahan	Cap	Ibu Tyas	WANITA	Jawa Tengah	7	135000		
20	6	2	2010	Bahan Standar Cap Garis	Bahan	Cap	Ibu Tyas	WANITA	Jawa Tengah	1	500000		
21	11	4	2010	Bolero Standar Cap Sidomukti	Hem	Print	Ibu Tyas	WANITA	Jawa Tengah	5	100000		

2. Membuat Pivot Table dengan menggunakan data tersebut.



3. Menampilkan data Pivot Table dengan mencoba kombinasi **tahun**, **nama_subkategori**, dan **jumlah**.

D. 2. Kegiatan 2 : Menambahkan Tipe Summary Baru

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

Fakta_Penjualan - Excel

PivotTable Tools

Analyze Design

PivotTable Name: PivotTable1

Active Field: Active Field

PivotTable Fields

Choose fields to add to report:

Search: jumlah

Drag fields between areas below:

Filters

Columns: tahun

Rows: nama_subkat...

Values: Sum of jumlah, Count of jumlah2

PivotTable Fields

Choose fields to add to report:

Search: jumlah

Drag fields between areas below:

Filters

Columns: tahun

Rows: nama_subkat...

Values: Sum of jumlah, Count of jumlah2

	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 Bahan	1	1	8	1	8	2	17	4
7 Batik					1	1	1	1
8 Bolero				1	1		1	1
9 Celana	17	1			17	1	34	2
10 Hem	5	1	8	2	4	2	17	5
11 Jam					44	1	44	1
12 Jarik					4	1	4	1
13 kaos			1	1	14	1	15	2
14 Rok					1	1	1	1
15 Sarimbit			1	1			1	1
16 Standar		2	1			2	1	1
17 Grand Total	23	3	21	7	93	10	137	20

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

Fakta_Penjualan - Excel

PivotTable Tools

Analyze Design

PivotTable Fields

Choose fields to add to report:

Search: jumlah

Drag fields between areas below:

Filters

Columns: tahun

Rows: nama_subkat...

Values: Sum of jumlah, Sum of jumlah2

Value Field Settings

Source Name: jumlah

Custom Name: Count of jumlah2

Summarize Values By: Show Values As

Summarize value field by: Count

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

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

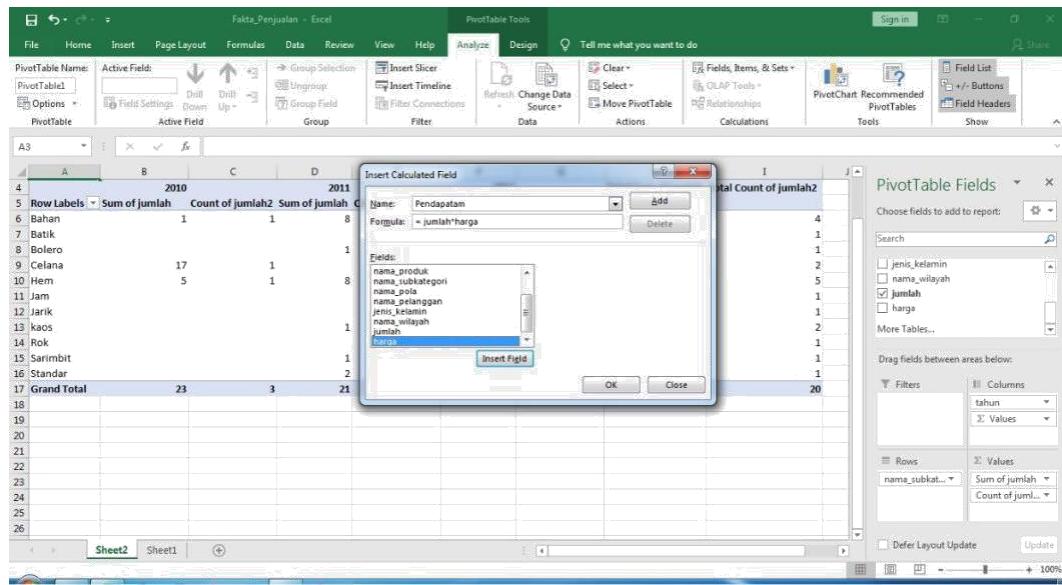
Number Format: OK Cancel

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

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

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

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



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

Row Labels	2010			2011			2012	
	Sum of jumlah	Count of jumlah2	Sum of Pendapatan	Sum of jumlah	Count of jumlah2	Sum of Pendapatan	Sum of jumlah	
Bahan	1	1	500000	8	1	960000	8	
Batik			0			0	1	
Bolero			0	1	1	255000		
Celana	17	1	935000			0	17	
Hem	5	1	500000	8	2	4960000	4	
Jam			0			0	44	
Jarik			0			0	4	
kaos			0	1	1	60000	14	
Rok			0			0	1	
Sarimbit			0	1	1	150000		
Standar			0	2	1	450000		
Grand Total	23	3	15065000	21	7	30030000	93	
			Total Sum of jumlah	Total Count of jumlah2	Total Sum of Pendapatan			
			Count of jumlah2	Sum of Pendapatan				
	2	1	2120000	17	4	15045000		
	1	1	150000	1	1	150000		
			0	1	1	255000		
	1	1	935000	34	2	3740000		
	2	2	1596000	17	5	19023000		
	1	1	3520000	44	1	3520000		
	1	1	160000	4	1	160000		
	1	1	420000	15	2	1350000		
	1	1	225000	1	1	225000		
			0	1	1	150000		
			0	2	1	450000		
	10	115692000		137	20	456073000		

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

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

The screenshot shows a PivotTable in Microsoft Excel with the following data structure:

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

- Menambahkan field pada kotak **Column (kuartal dan bulan)** dan **Rows (nama_produk dan nama_pelanggan)**, untuk menambah field-field yang bisa diperinci dan diringkas sesuai urutan kategori data yang lebih spesifik.

Roll Up

The screenshot shows the same PivotTable after adding fields to the columns and rows:

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

Drop Down

Fakta_Penjualan - Excel

PivotTable Tools

File Home Insert Page Layout Formulas Data Review View Help Analyze Design Tell me what you want to do

PivotTable Name: Active Field: **nama_subkateg** Group Selection Insert Slicer
PivotTable Options Drill Down Drill Up Ungroup Group Field Filter Connections
PivotTable Active Field Group Filter Refresh Change Data Source Select Move PivotTable
PivotTable Tools PivotChart Recommended PivotTables
Field List / Buttons Field Headers Show

A53 Standar

Sum of Pendapatan

	B	C	D	E	F	G	H	I	J	K	L	M	N
Row Labels	2010	2011	2012	Grand Total									
3 Bahan	500000	960000	2120000	15045000									
4 Bahan Beludru Cap Mahkota	0	0	945000	945000									
5 Ibu Tyas	0	0	945000	945000									
6 Bahan Standar Cap Garis	500000	0	0	500000									
7 Ibu Tyas	500000	0	0	500000									
8 Bahan Standar Cap Lasem	0	0	130000	130000									
9 Ibu Niken	0	0	130000	130000									
10 Hem Sutra Print Rama	0	960000	0	960000									
11 Ibu Siti Arya	0	960000	0	960000									
12 Batik	0	0	150000	150000									
13 Celana Standar Cap Warna	0	0	150000	150000									
14 Bapak Heru	0	0	150000	150000									
15 Bolero	0	255000	0	255000									
16 Kaos Batik Cap Tullis	0	255000	0	255000									
17 Ibu Hatamah	0	255000	0	255000									
18 Celana	935000	0	935000	3740000									
19 Hem Standar Cap Tumpal	935000	0	0	935000									
20 Ibu Hadi Sukarni	935000	0	0	935000									
21 Sarimbit Standar Print Lukis	0	0	935000	935000									
22 Ibu Hadi Sukarni	0	0	935000	935000									
23 Hem	500000	4960000	1596000	19023000									

PivotTable Fields

Choose fields to add to report:

Search

bulan
kuartal
 tahun
 nama_produk
 nama_subkateg

Drag fields between areas below:

Filters Columns tahun

Rows nama_subkateg Sum of Pendapatan

Values nama_produk nama_pelan...

Defer Layout Update

Sheet2 Sheet1

5. Tugas

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

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

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

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

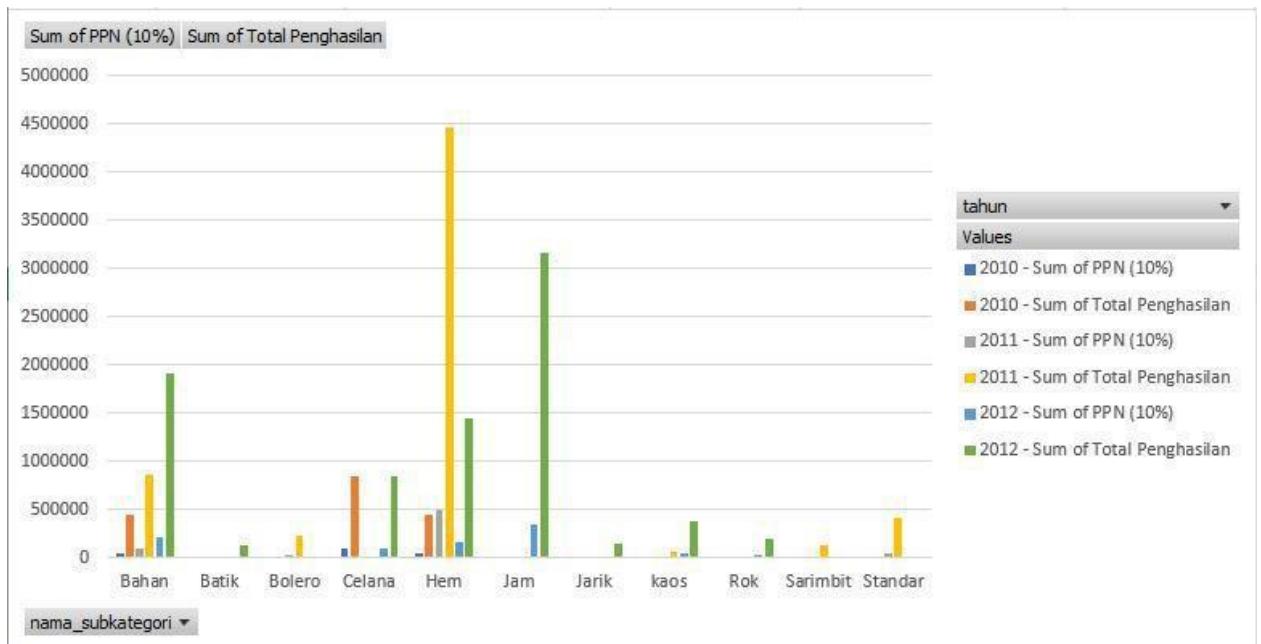
Total Sum of Pendapatan	Total Sum of PPN (10%)	Total Sum of Total Penghasilan	Total Sum of jumlah	Total Sum of kuartal
1504500	1504500	13540500	17	7
150000	15000	135000	1	3
255000	25500	229500	1	4
3740000	374000	3366000	34	2
19023000	1902300	17120700	17	14
3520000	352000	3168000	44	1
160000	16000	144000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

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

PivotTable

Row Labels	Column Labels		2010		2011		2012		Total Sum of PPN (10%)	Total Sum of Total Penghasilan
	Sum of PPN (10%)	Sum of Total Penghasilan	Sum of PPN (10%)	Sum of Total Penghasilan	Sum of PPN (10%)	Sum of Total Penghasilan	Sum of PPN (10%)	Sum of Total Penghasilan		
Bahan	50000	450000	96000	864000	212000	1908000	1504500	13540500		
Batik	0	0	0	0	15000	135000	15000	135000		
Bolero	0	0	25500	229500	0	0	25500	229500		
Celana	93500	841500	0	0	93500	841500	374000	3266000		
Hem	50000	450000	496000	4464000	159600	1436400	1902300	17120700		
Jam	0	0	0	0	352000	3168000	352000	3168000		
Jarik	0	0	0	0	16000	144000	16000	144000		
kaos	0	0	6000	54000	42000	378000	135000	1215000		
Rok	0	0	0	0	22500	202500	22500	202500		
Sarimbit	0	0	15000	135000	0	0	15000	135000		
Standar	0	0	45000	405000	0	0	45000	405000		
Grand Total	1506500	13558500	3003000	27027000	11569200	104122800	45607300	410465700		

PivotChart



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

+-----+

MODUL VI

PENGENALAN DATA MINING

6.5 Tugas (Dikerjakan Saat Ini)

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

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

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

- Tabel Kanan Paling Atas

E. Pada attribut **Lama_Studi**, berapa jumlah data masing-masing kelas **Tepat**, **Terlambat**?

- Tabel Kanan Nomor 2 dari Atas

F. Pada attribute **Rerata_SKS**, berapa kali Max, Min, Mean, dan Standard Deviation?

- Tabel Kanan Nomor 2 dari Bawah

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

- Tabel Kanan Paling Bawah

+-----+

MODUL VII

DATA PREPROCESSING

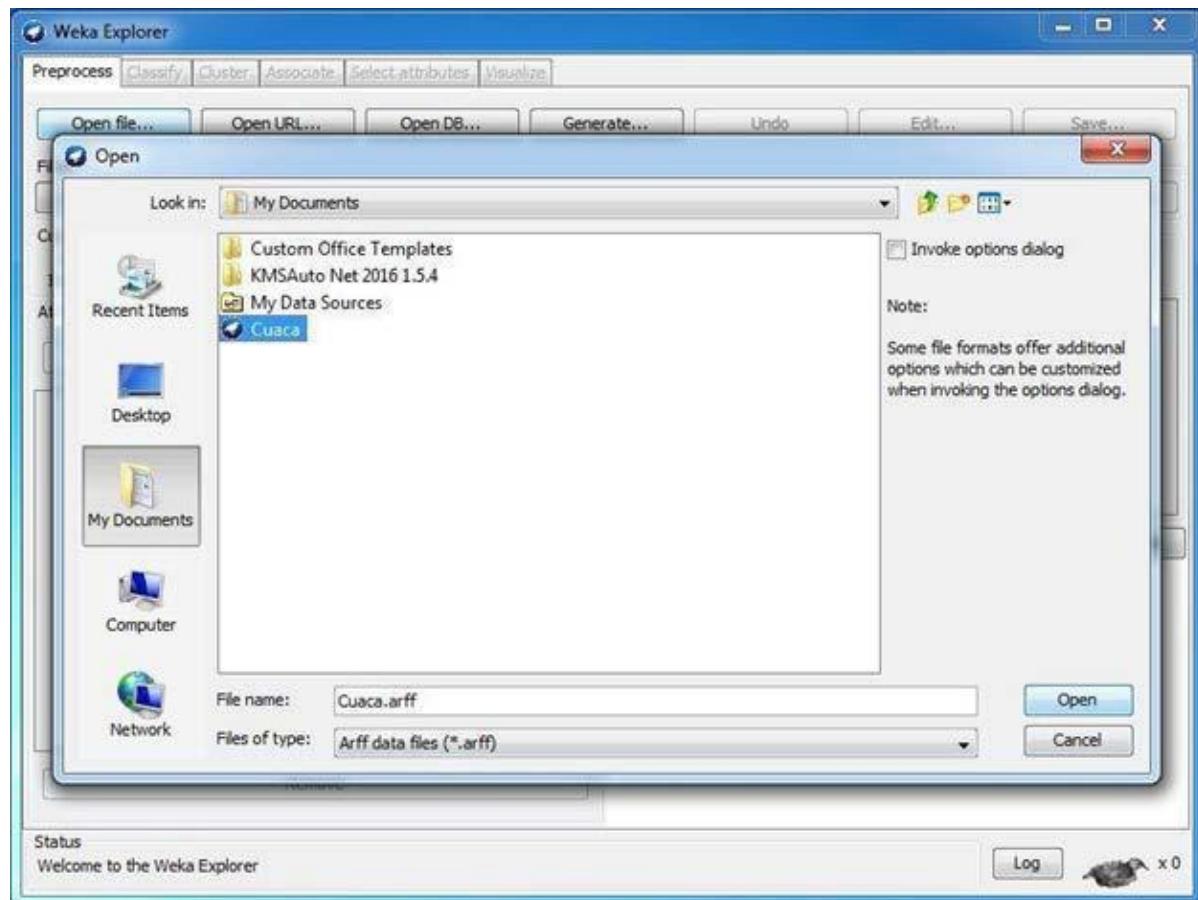
7.4 Langkah-langkah Praktikum

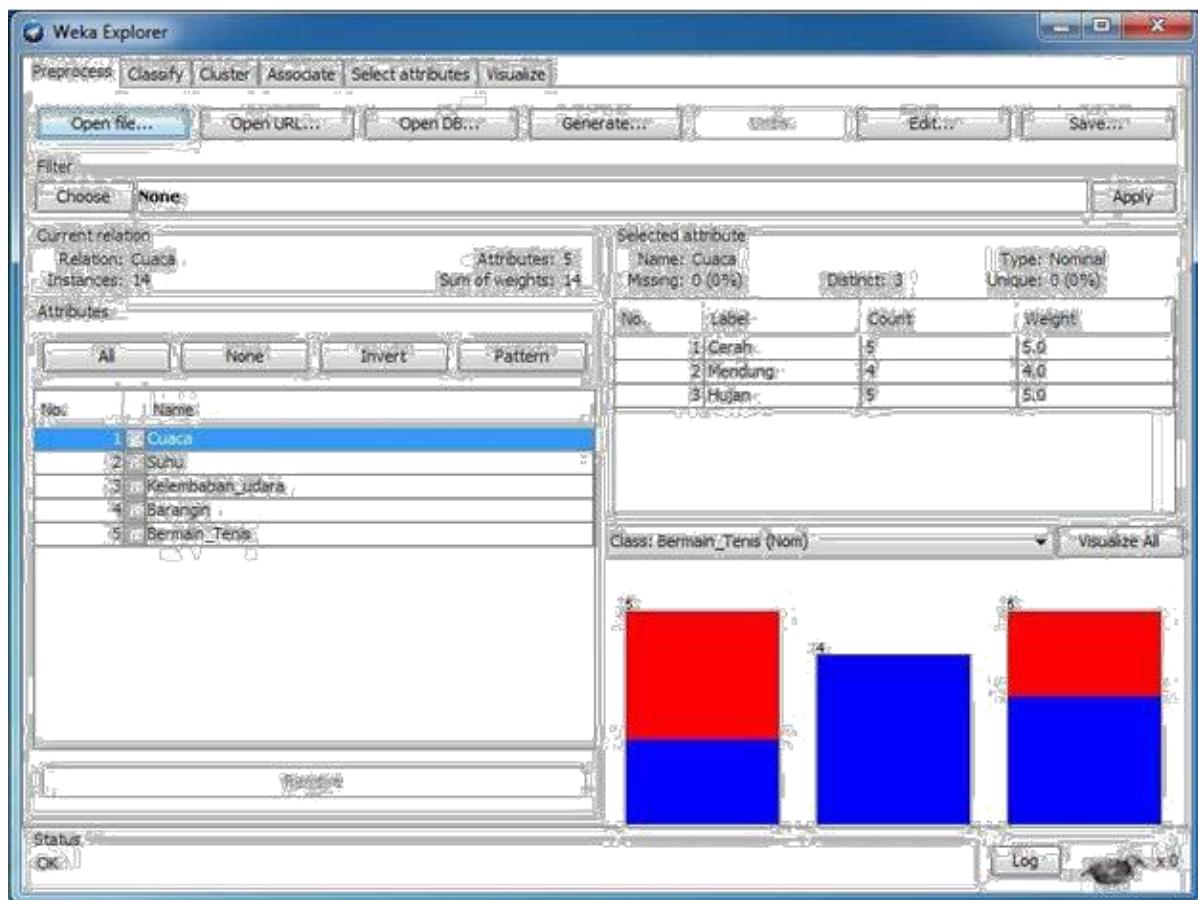
7.4.1 Membuat Format Data ARFF (Attribut-Relation File Format)

```
≡ Cuaca.arff ×

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

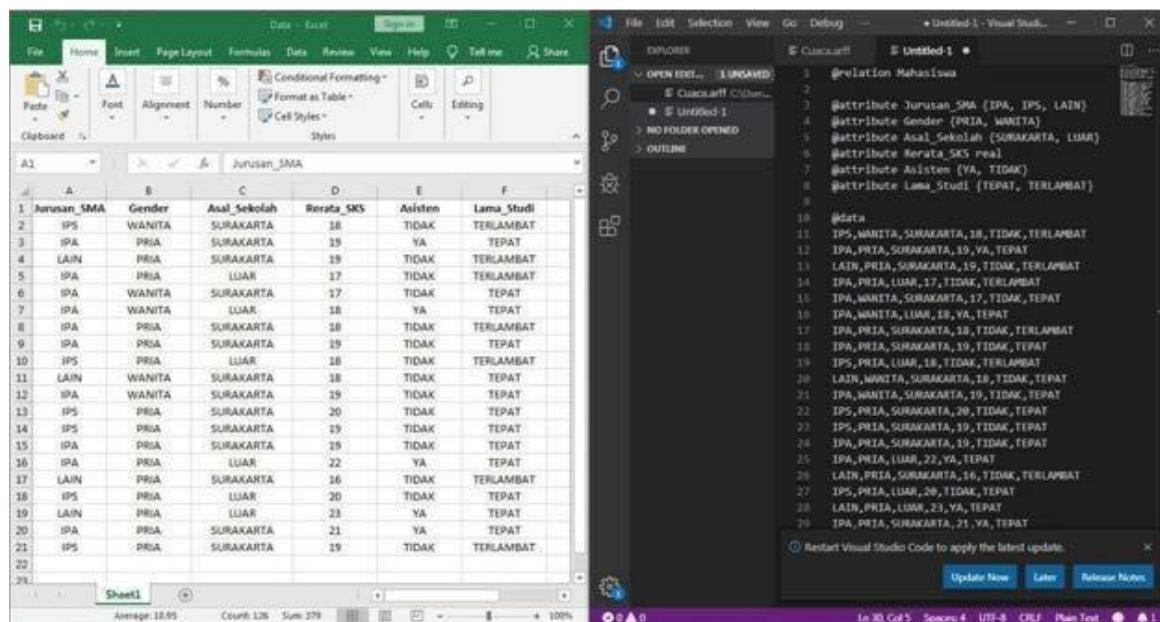
7.4.2 Menggunakan File ARFF dengan WEKA





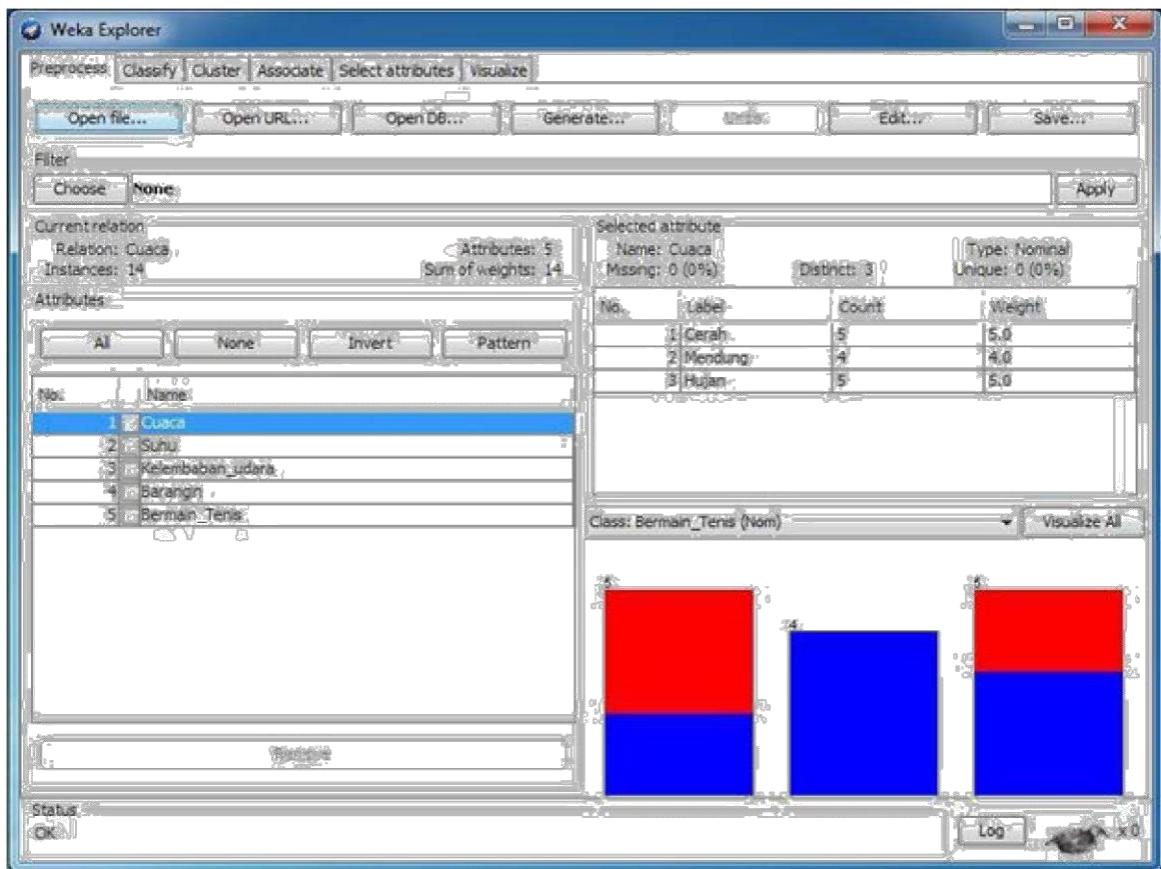
7.5. Tugas

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



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

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



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

4. Berapa jumlah atribut yang bertipe binomial dan polynomial?

Binomial : 4 (Gender, Asal_Sekolah, Asisten, Lama_Studi)

Polynomial : 1 (Jurusان_SMA)

5. Berapa jumlah atribut yang bertipe real?

Real : 1 (Rerata_SKS)

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

Maximum : 23

Minimum : 16

Mean : 18.95

Standard Deviation : 1.66938375

MODUL VII KLASIFIKASI : NAÏVE BAYES

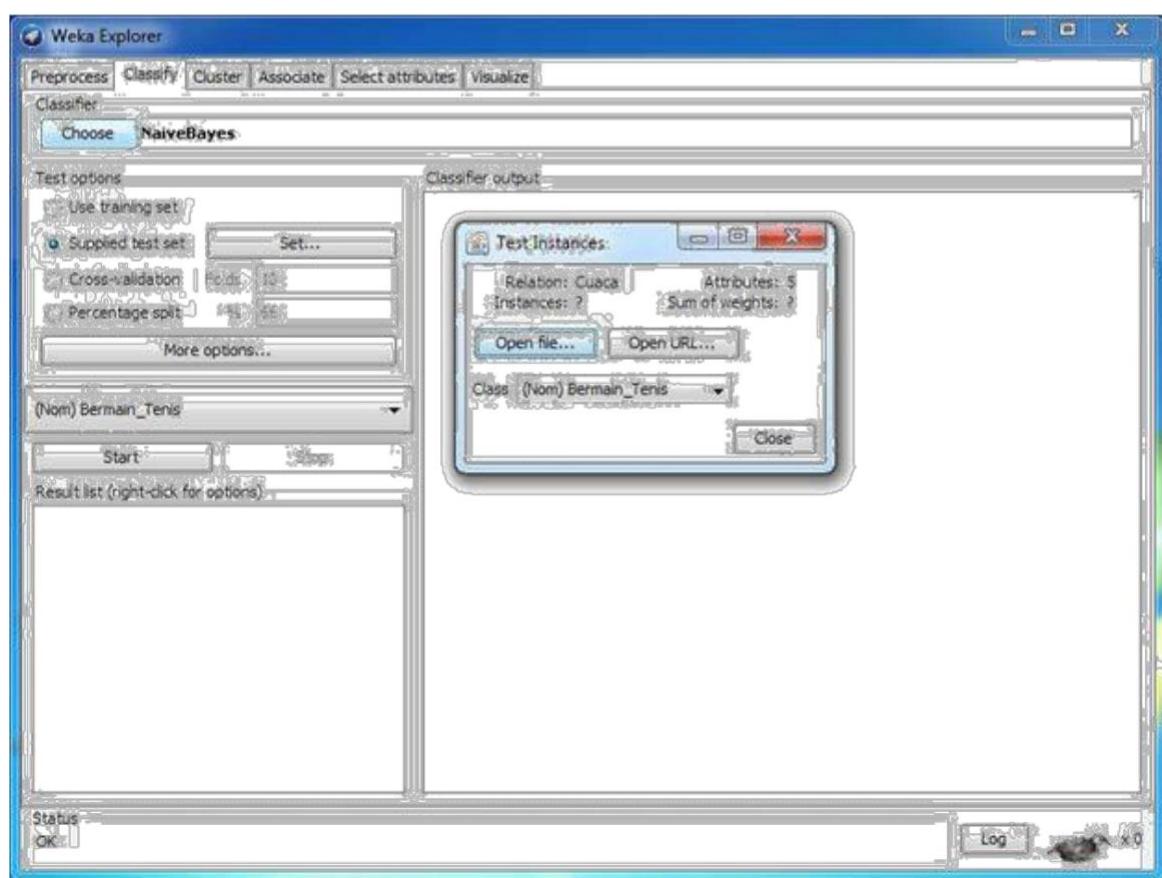
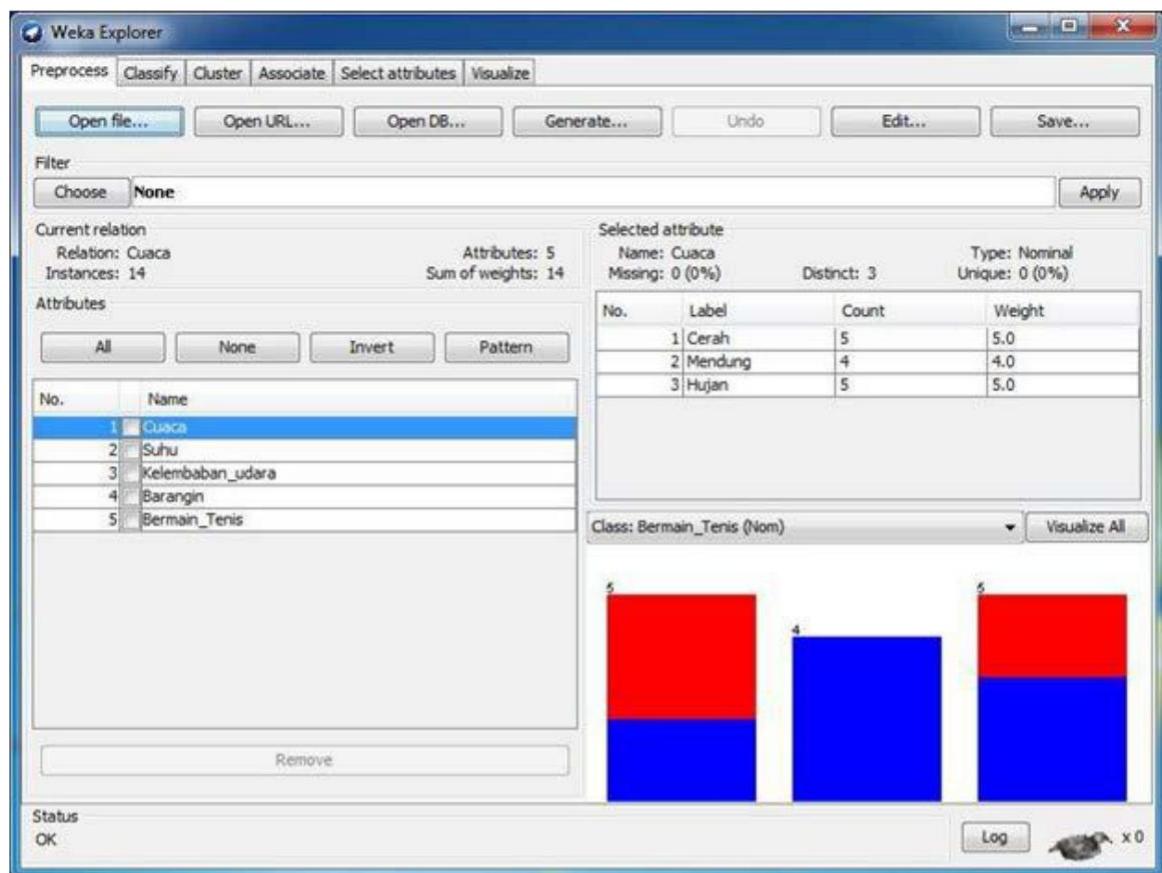
8.4 Langkah-langkah Praktikum

8.4.1 Implementasi Naïve Bayes dengan Weka

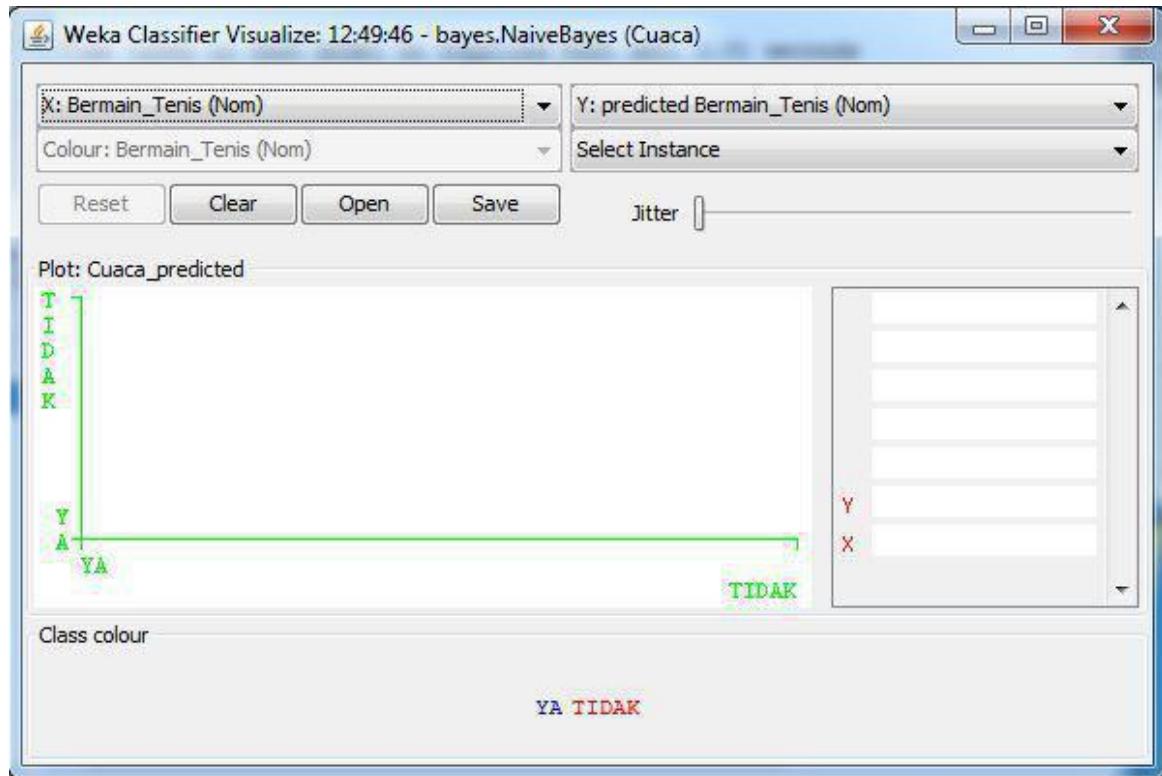
1. Membuat table Data Testing Cuaca yang ada dimodul ke bentuk ARFF dan simpan dengan nama **CuacaTesting.arff**.

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

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



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



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

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

File Edit View

HasilPrediksi.arff

Relasi: Cuaca_predicted

No.	1: Cuaca_Nominal	2: Suhu_Numeric	3: Kelembaban_Jatara_Numeric	4: Berangin_Nominal	5: prediction margin_Numeric	6: predicted Bernmain_Tenis_Nominal	7: Bernmain_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	67,0	YA	-0,676866	TIDAK	
4.	Mendung	70,0	96,0	TIDAK	0,428523	YA	
5.	Mendung	68,0	81,0	TIDAK	0,833996	YA	
6.	Hujan	93,0	75,0	YA	-0,262211	YA	
7.	Hujan	64,0	85,0	YA	-0,160143	TIDAK	

8.4.2 Implementasi Naïve Bayes dengan RapidMiner

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

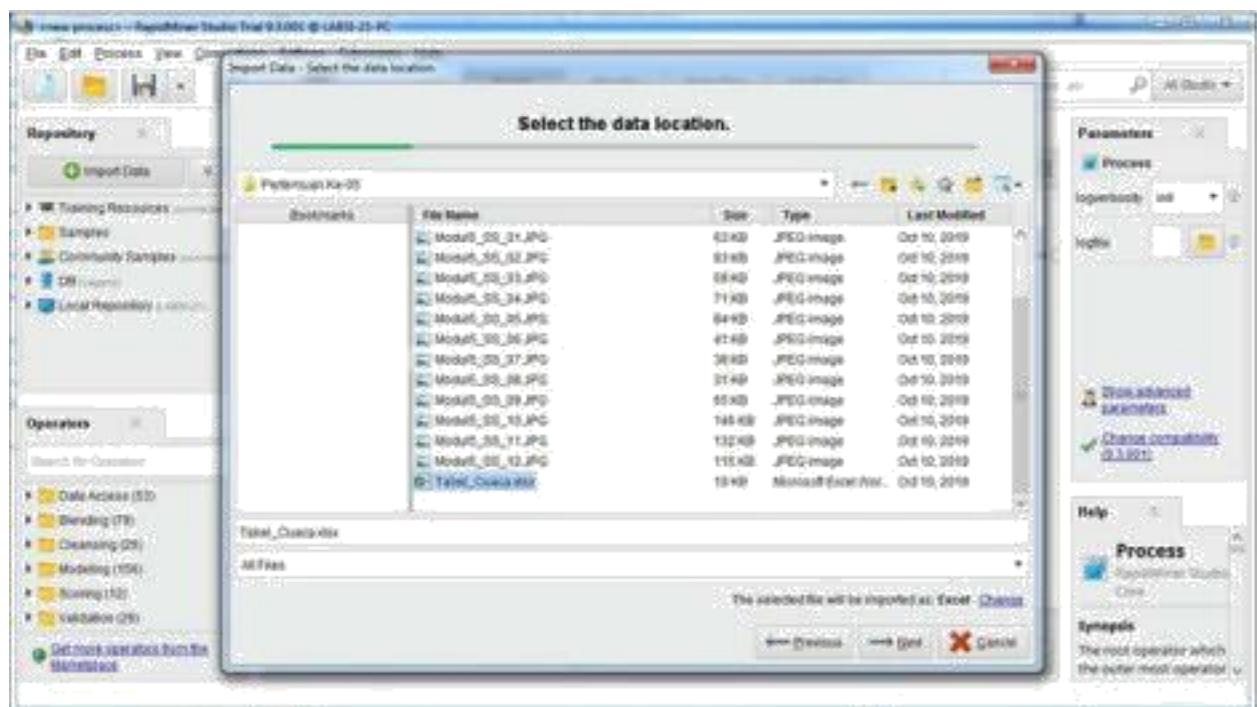
Screenshot of Microsoft Excel showing a dataset titled "Tabel_Cuaca - Excel". The data consists of 16 rows and 5 columns. The columns are labeled: Cuaca, Suhu, Kelembaban_udara, Berangin, and Bermain_tenis. The data entries are as follows:

	Cuaca	Suhu	Kelembaban_udara	Berangin	Bermain_tenis
2	Cerah	83	85 TIDAK		TIDAK
3	Cerah	80	90 YA		TIDAK
4	Mendung	83	88 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
16					
17					
18					
19					
20					
21					
22					
23					

Screenshot of Microsoft Excel showing the same dataset as the first screenshot. The data consists of 8 rows and 5 columns. The columns are labeled: Cuaca, Suhu, Kelembaban_udara, Berangin, and Bermain_tenis. The data entries are as follows:

	Cuaca	Suhu	Kelembaban_udara	Berangin	Bermain_tenis
2	Cerah	75	85 TIDAK		
3	Cerah	80	88 YA		
4	Cerah	83	87 YA		
5	Mendung	70	96 TIDAK		
6	Mendung	68	81 TIDAK		
7	Hujan	65	75 YA		
8	Hujan	64	85 YA		
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					

- Menjalankan **Rapid Miner** dan Meng-import file excel yang sudah disimpan tadi (nama file **Tabel_Cuaca.xls**).



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

Import Data - Select the cells to import.

Select the cells to import.

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

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

← Previous → Next 

Import Data - Format your columns.

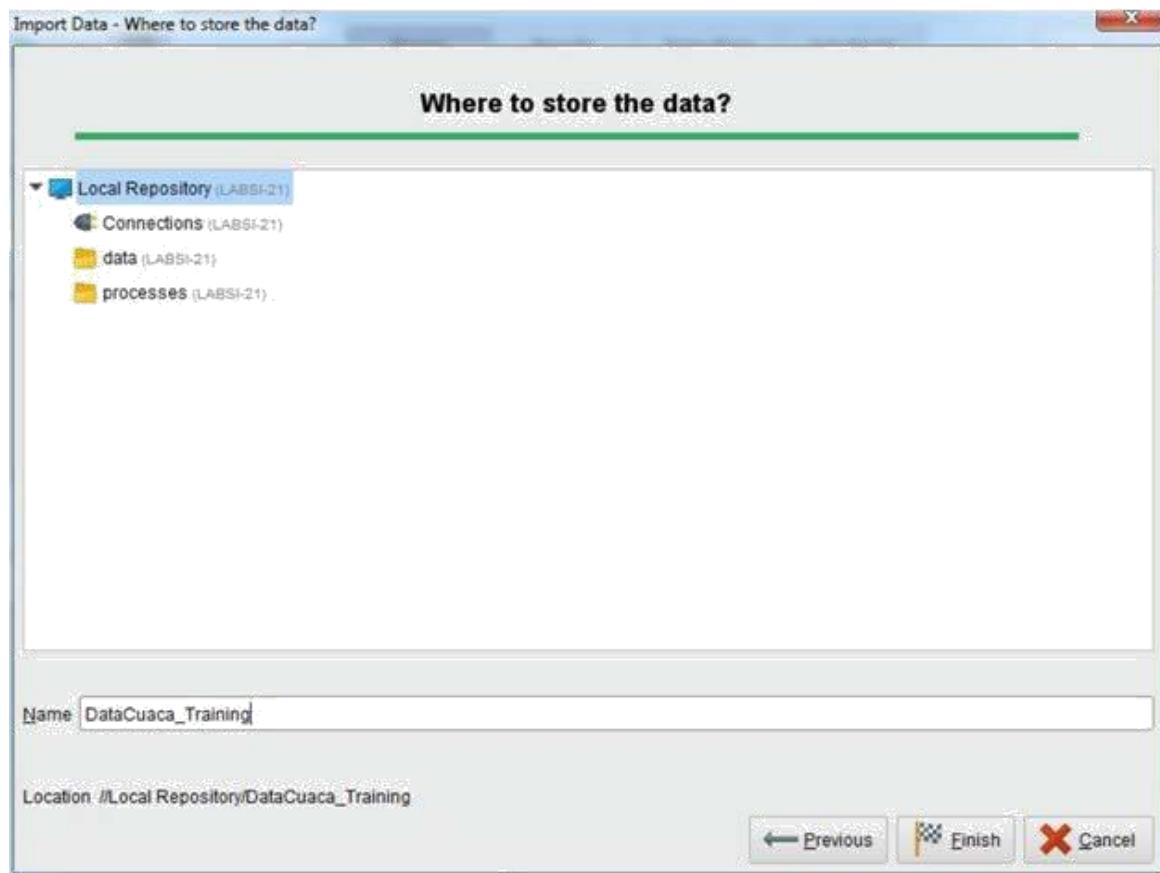
Format your columns.

Replace errors with missing values

	Cuaca polynominal	Suhu integer	Kelembaban_udara integer	Berangin polynominal	Bermain_tenis binominal label
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
8	Cerah	72	95	TIDAK	TIDAK
9	Cerah	69	70	TIDAK	YA
10	Hujan	75	80	TIDAK	YA
11	Cerah	75	70	YA	YA
12	Mendung	72	90	YA	YA
13	Mendung	81	75	TIDAK	YA

 no problems

← Previous → Next 



<new process> - RapidMiner Studio Trial 9.3.001 @ LABSI-21-PC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Find data (operations... All Studio)

Result History ExampleSet (//Local Repository/DataCuaca_Training)

Data

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

Row No.	Bermula_te...	Cuaca	Suhu	Ketebalan...	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	66	YA
8	TIDAK	Cerah	72	96	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

Repository

- Training Resources
- Samples
- Community Samples
- DB Scripts
- Local Repository (LABSI-21)
 - Connections (LABSI-21)
 - data (LABSI-21)
 - processes (LABSI-21)
 - DataCuaca_Training (LABSI-21)

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

The screenshot shows the RapidMiner Studio interface with the 'Design' tab selected. On the left, there's a sidebar with icons for Data, Statistics, Visualizations, and Annotations. The main area displays a table titled 'ExampleSet' with 14 rows and 6 columns. The columns are labeled: Row No., Bermula_te..., Cuaca, Suhu, Ketebalan..., and Berangin. The data in the table is as follows:

Row No.	Bermula_te...	Cuaca	Suhu	Ketebalan...	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	66	YA
8	TIDAK	Cerah	72	96	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

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

Import Data - Select the cells to import.

Select the cells to import.

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

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

← Previous → Next X Cancel

Import Data - Format your columns.

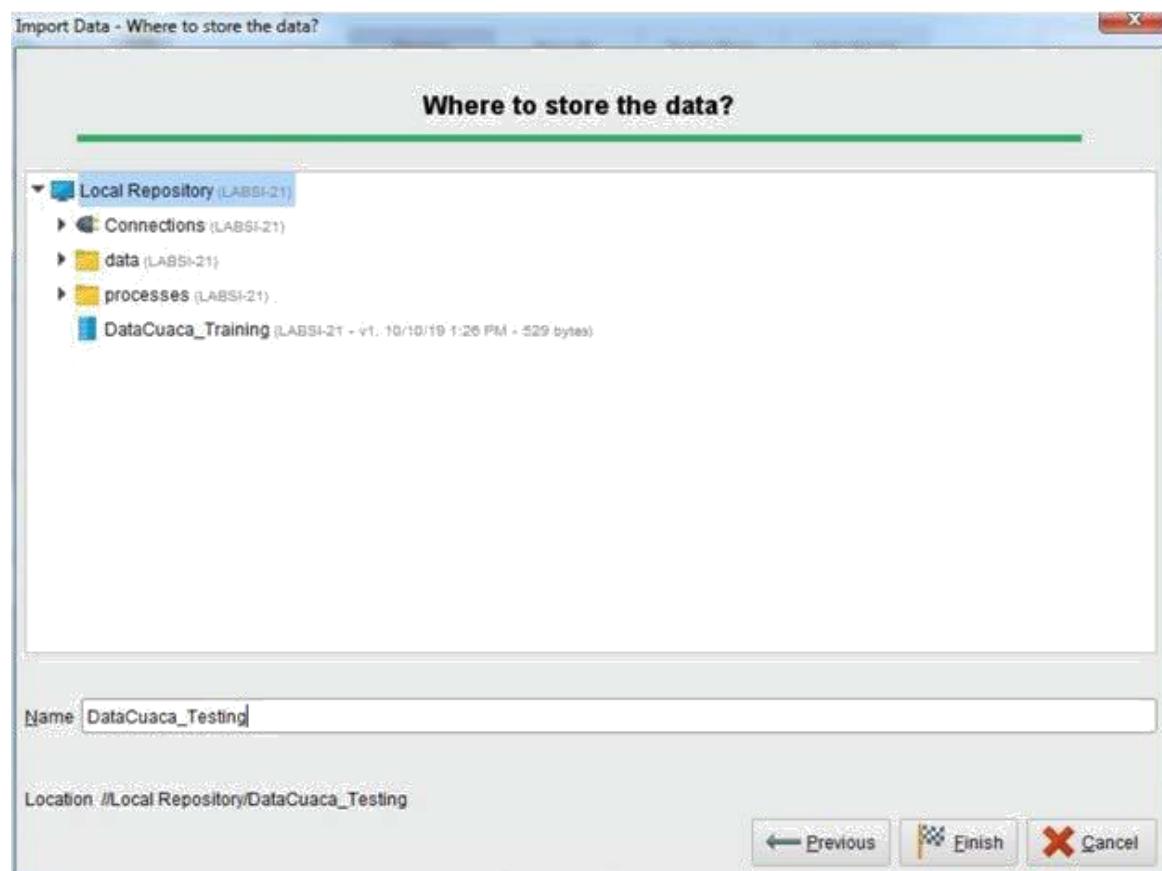
Format your columns.

Replace errors with missing values.

	Cuaca polynominal	Suhu integer	Kelembaban_udara integer	Berangin binomial
1	Cerah	75	65	TIDAK
2	Cerah	80	68	YA
3	Cerah	83	67	YA
4	Mendung	79	96	TIDAK
5	Mendung	88	81	TIDAK
6	Hujan	65	75	YA
7	Hujan	64	85	YA

 no problems

 Previous  Next  Cancel



<new process> - RapidMiner Studio Trial 9.3.001 @ (LABSI-21-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_Training)

ExampleSet (/Local Repository/DataCuaca_Testing)

Open in: Turbo Prep Auto Model Filter (7/7 examples): 00

Row No.	Cuaca	Suhu	Kehembaban...	Berangin
1	Cerah	75	95	TIDAK
2	Cerah	80	88	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

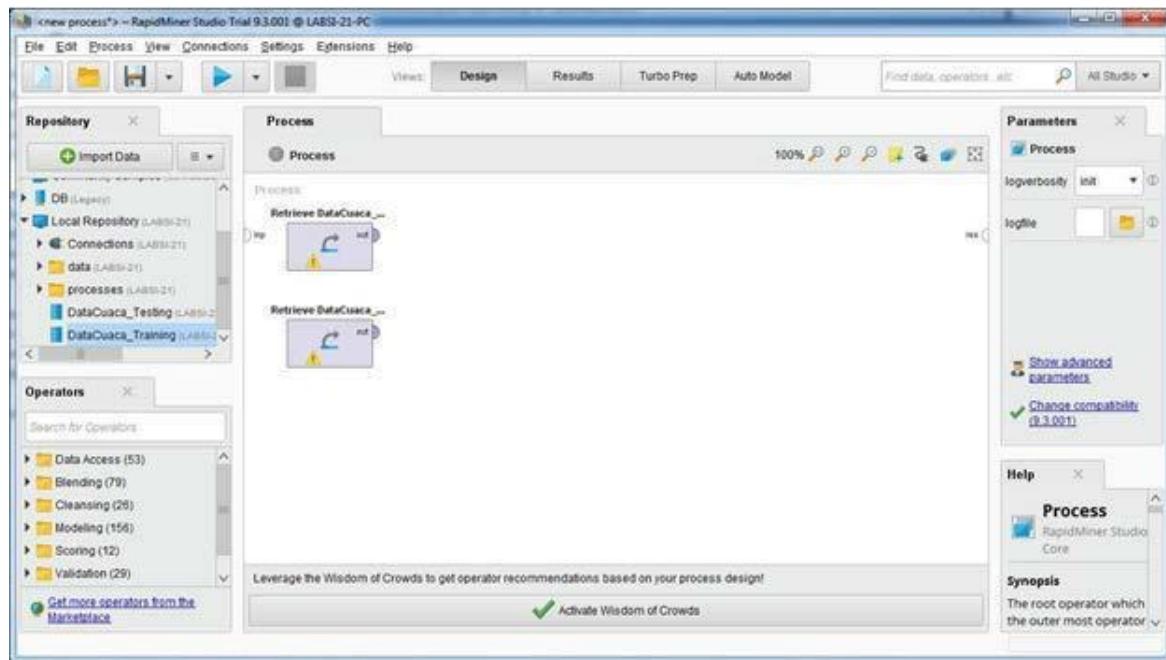
Annotations

Repository

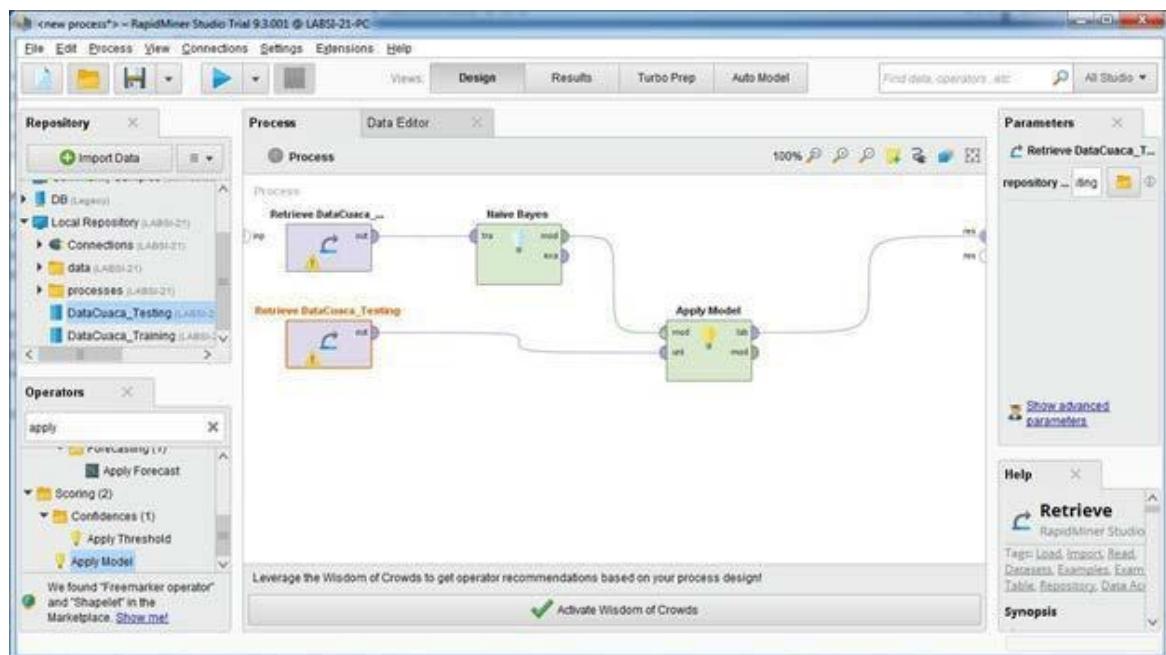
- Training Resources
- Samples
- Community Samples
- DB Sources
- Local Repository (LABSI-21)
 - Connections (LABSI-21)
 - data (LABSI-21)
 - processes (LABSI-21)
 - DataCuaca_Testing (LABSI-21 - v1)
 - DataCuaca_Training (LABSI-21 - v1)

ExampleSet (7 examples, 0 special attributes, 4 regular attributes)

5. Kembali ke halaman utama **Rapid Miner** dan membuat desain Naïve Bayes, dengan mengdrag **DataCuaca_Training** dan **DataCuaca_Testing** ke dalam jendela **Process View**.



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



The screenshot shows the RapidMiner Studio interface with the 'Statistics' view selected. The main area displays a table of 7 examples with the following data:

Row No.	prediction(Bermain_tenis)	confidence(TIDAK)	confidence(YA)	Cuaca	Suhu	Kelembaban_udara	Berangin
1	YA	0.154	0.846	Cerah	75	85	70%
2	YA	0.498	0.502	Cerah	80	88	79%
3	TIDAK	0.856	0.144	Condong	85	85	79%
4	YA	0.019	0.981	Mendung	70	88	70%
5	YA	0.007	0.993	Mendung	70	85	70%
6	YA	0.371	0.629	Hujan	65	78	79%
7	TIDAK	0.568	0.432	Hujan	65	85	79%

7. Klik Statistics untuk melihat data lebih detail.

The screenshot shows the detailed statistics for each attribute in the 'Statistics' view. The table includes the following data:

Name	Type	Missing	Statistics	Values
Prediction	Binomial	0	Least: TIDAK (2)	Most: YA (5)
confidence(TIDAK)	Real	0	Min: 0.007	Max: 0.856
confidence(YA)	Real	0	Min: 0.144	Max: 0.993
Cuaca	Polynomial	0	Least: Mendung (2)	Most: Cerah (3)
Suhu	Integer	0	Min: 64	Max: 83
Kelembaban_udara	Integer	0	Min: 65	Max: 96
Berangin	Binomial	0	Least: TIDAK (3)	Most: YA (4)

8.5 Tugas

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

The ARFF code in the Notepad window:

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

@data
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,?
```

2. Melakukan Prediksi dengan memanfaatkan Software WEKA.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter

Choose None Apply Stop

Current relation

Relation: Mahasiswa Instances: 20 Attributes: 6 Sum of weights: 20

Selected attribute

Name: Jurusan_SMA	Type: Nominal
Missing: 0 (0%)	Distinct: 3
Unique: 0 (0%)	

No.	Label	Count	Weight
1	IPA	10	10.0
2	IPS	6	6.0
3	LAIN	4	4.0

Attributes

All None Invert Pattern

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

Remove

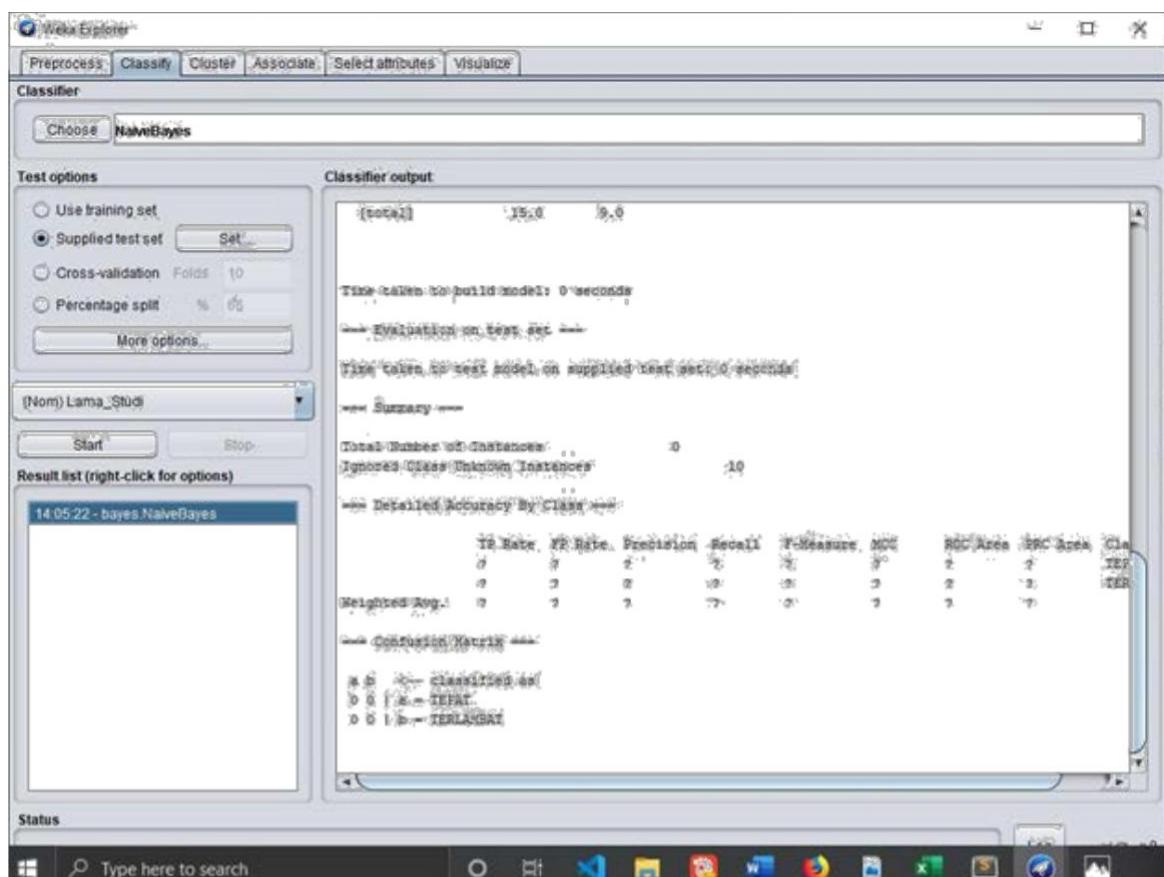
Status

Type here to search

Class: Lama_Studi (Nom)

Visualize All

Three stacked bar charts showing the distribution of 'Lama_Studi' for each 'Jurusan_SMA' category (IPA, IPS, LAIN).



ARFF-Viewer - F:\Tugas\Praktikum Data Warehousing dan Data Mining\Pertemuan Ke-05\HasilPredksi.arff						
File Edit View						
HasilPredksi.arff						
Relation: Cuaca_predicted						
No.	1: Cuaca Nominal	2: Suhu Numeric	3: Kelembaban_udara Numeric	4: Barang 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		-0,087978 YA	
3.	Cerah	83,0	87,0 YA		-0,678366 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		-0,160343 TIDAK	

3. Melakukan Prediksi dengan menggunakan Software RapidMiner Studio.

Import Data - Select the cells to import.

Select the cells to import.

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

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

← Previous → Next  Cancel

Import Data - Format your columns

Format your columns.

Replace errors with missing values

	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten	Lama_Studi
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 → | 

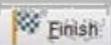
Import Data - Where to store the data?

Where to store the data?

- Local Repository (Yudha Ayudya Nandha)
 - Connections (Yudha Ayudya Nandha)
 - DATA (Yudha Ayudya Nandha)
 - processes (Yudha Ayudya Nandha)

Name: Tabel_Tugas_Training_Skipipap

Location: /Local Repository/Tabel_Tugas_Training_Skipipap

← Previous |  Finish | 

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Testing Cell range: AE 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 → X Cancel

Import Data - Format your columns.

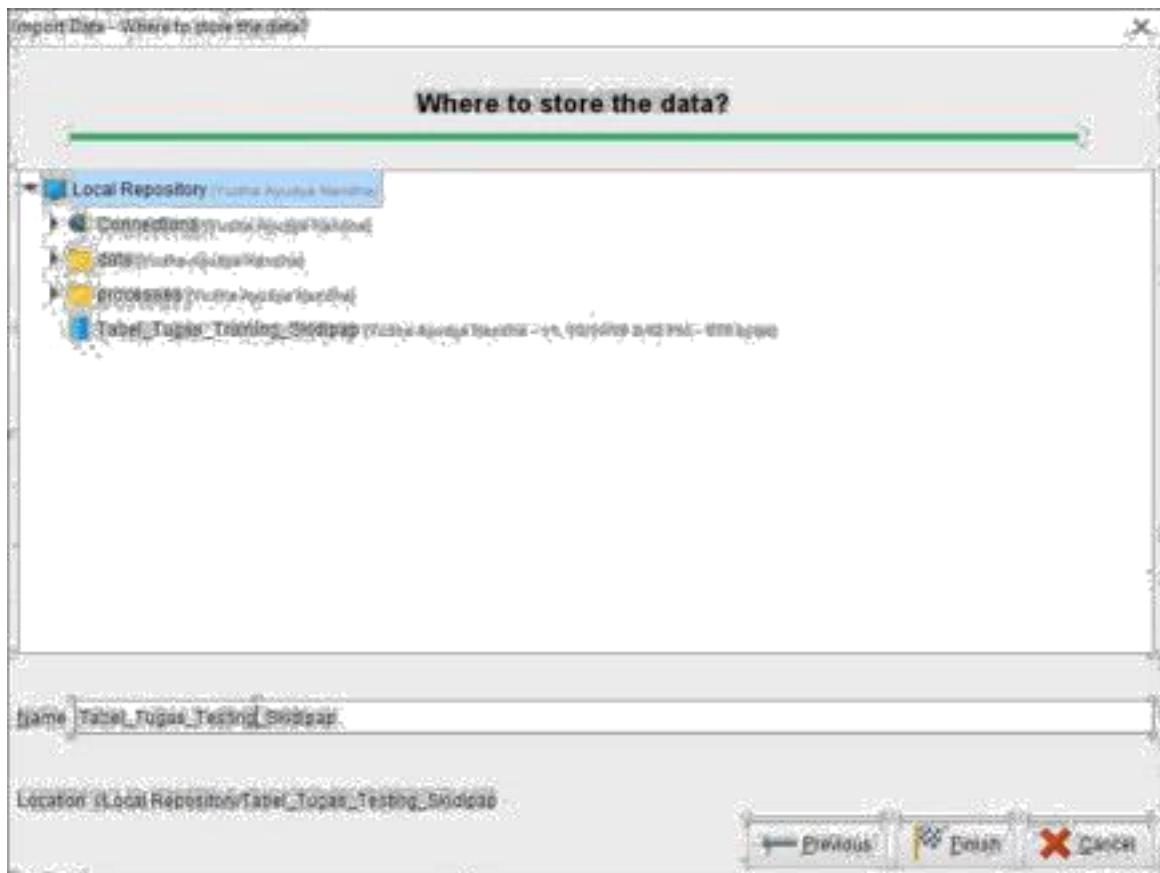
Format your columns.

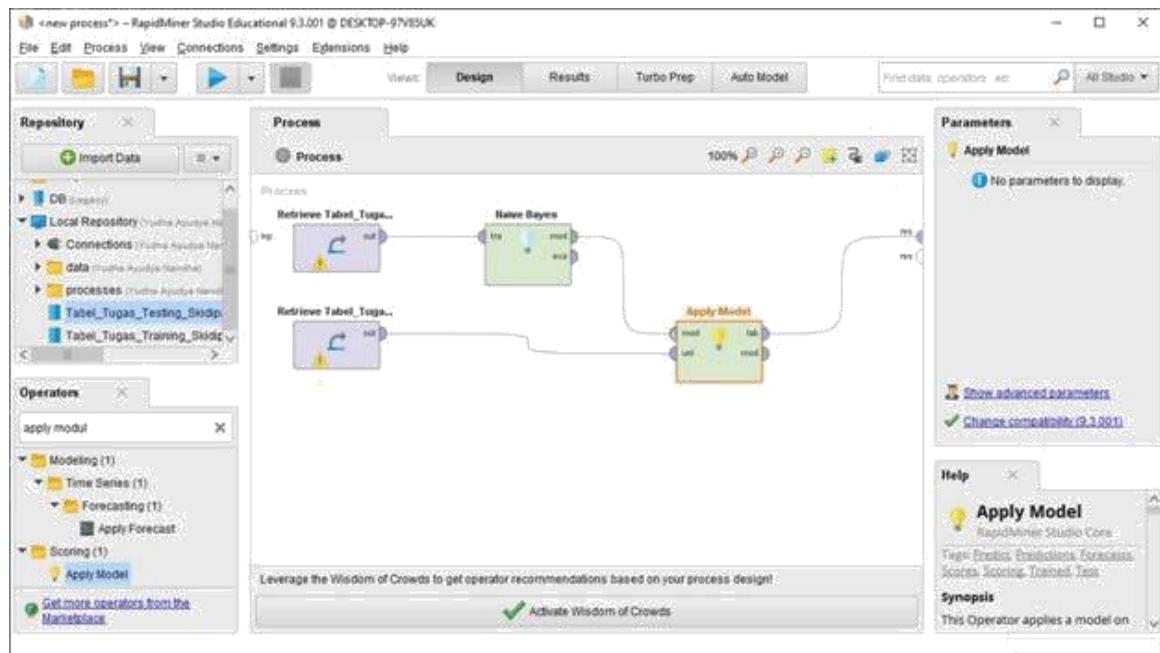
Replace errors with missing values

	Jurusan_SMA polynomial	Gender polynomial	Asal_Sekolah polynomial	Rerata_SKS Integer	Asisten binomial
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 → X Cancel





Row No.	prediction...	confidence...	confidence...	Jurasan_SMA	Gender	Anak_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	0.668	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	0.188	IPS	PRIA	LUAR	18	TIDAK
10	TERLAMBAT	0.648	0.352	LAIN	WANITA	SURAKARTA	18	TIDAK

ExampleSet (10 examples, 3 special attributes, 6 regular attributes)

4. Nilai Rerata hasil dari Prediksi RapidMiner untuk kolom Lama_Studie :

- Rerata untuk TEPAT : 0.476
- Rerata untuk TERLAMBAT : 0.524

5. Dari hasil prediksi dari Software Rapidminer Studio, dapat disimpulkan bahwa 3 Mahasiswa akan lulus TEPAT waktu, dan 7 Mahasiswa akan lulus TERLAMBAT.

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

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

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

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

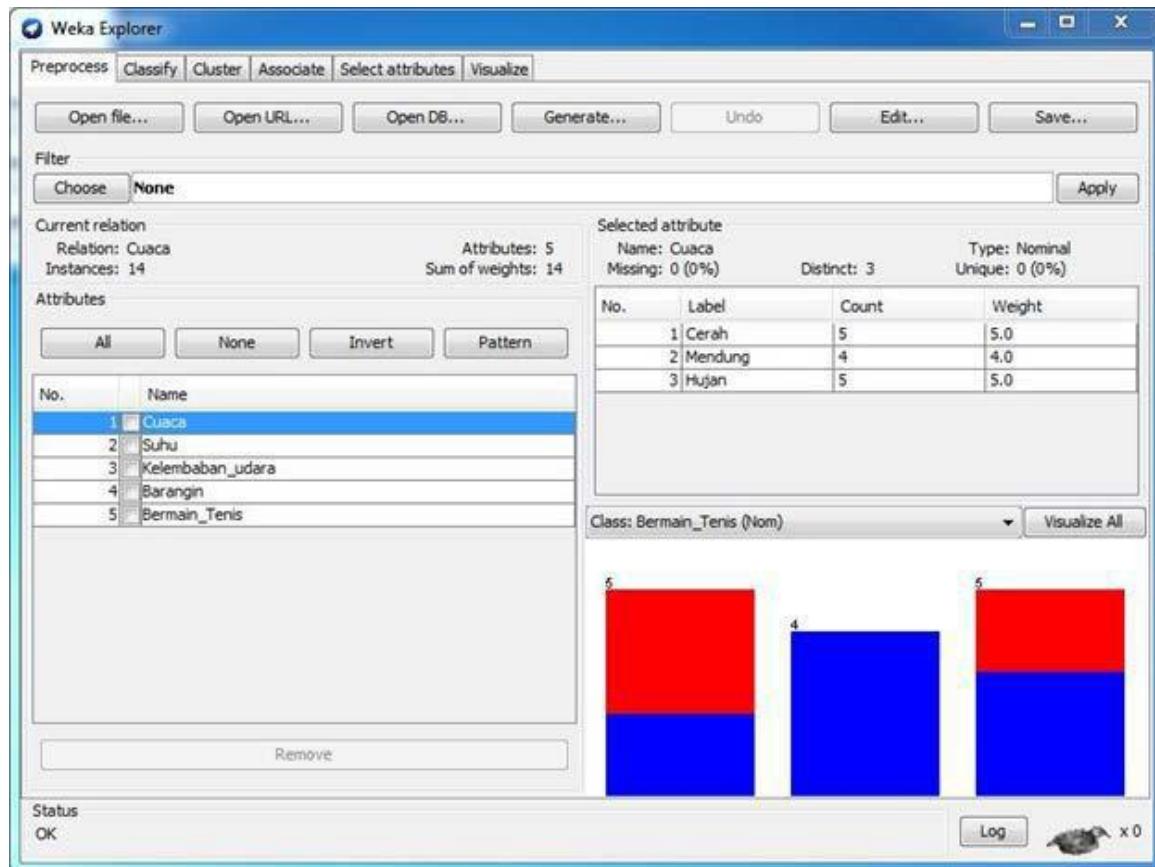
+-----+

MODUL IX KLASIFIKASI : DECISION TREE

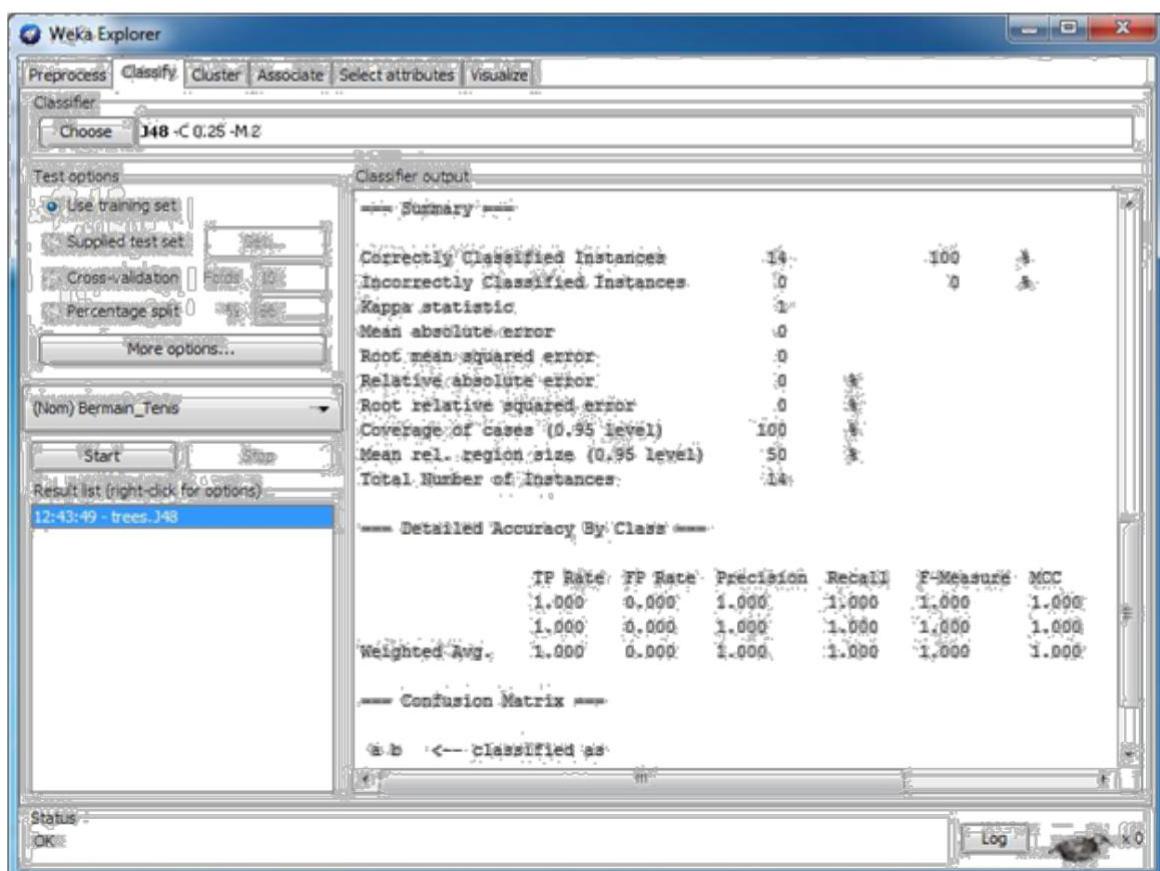
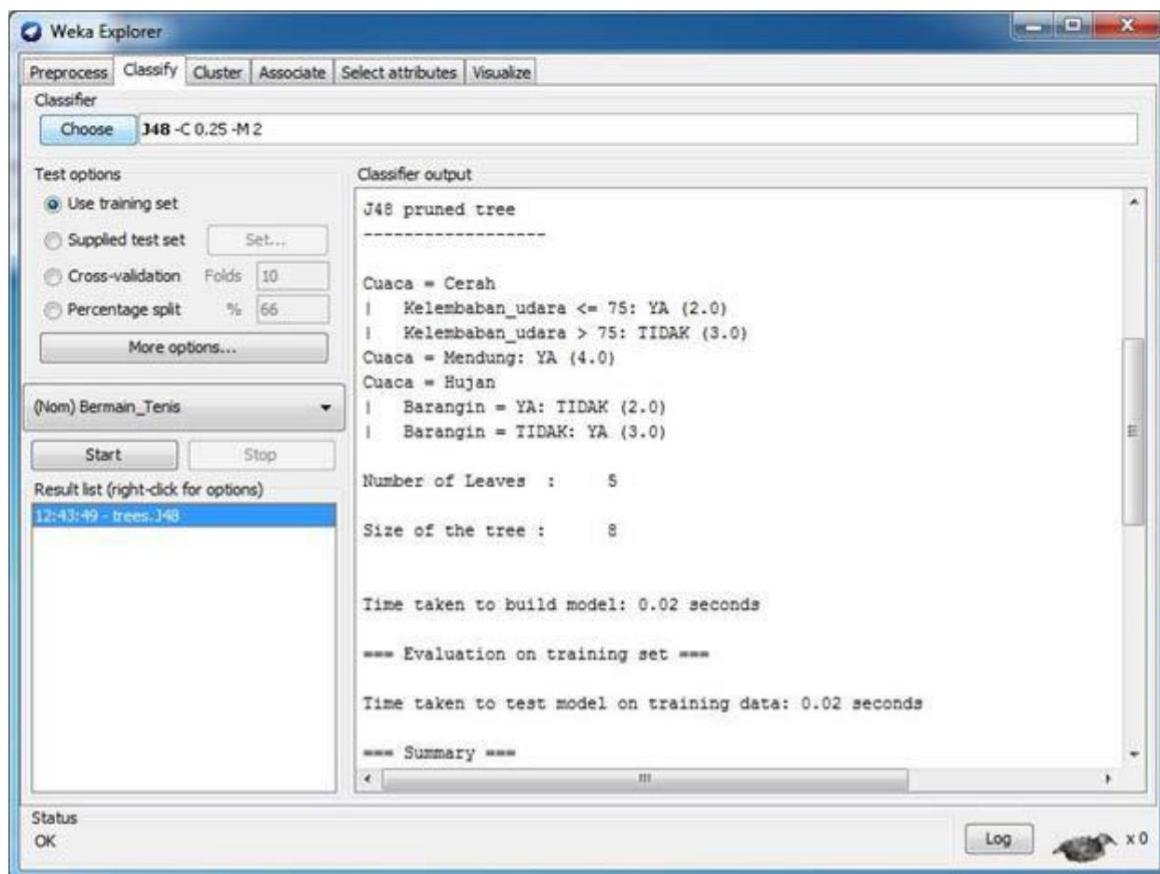
9.4 Langkah-langkah Praktikum

9.4.1 Pohon Keputusan Menggunakan WEKA

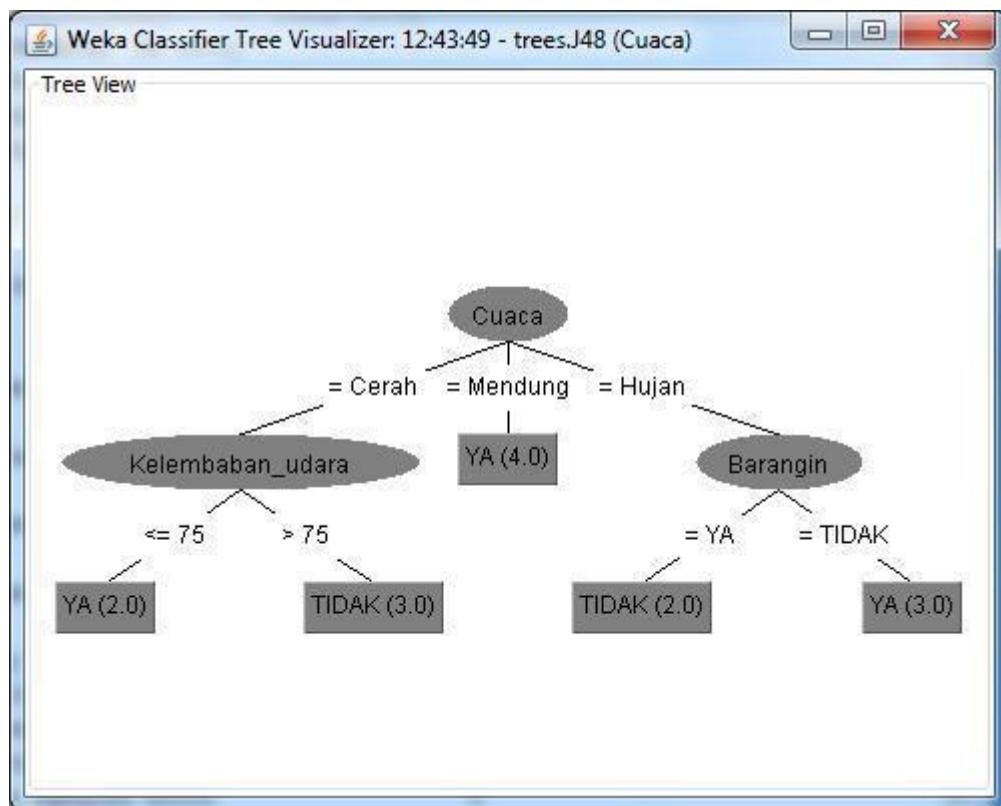
1. Buka file **Cuaca.arff**, dengan Weka Explore.



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

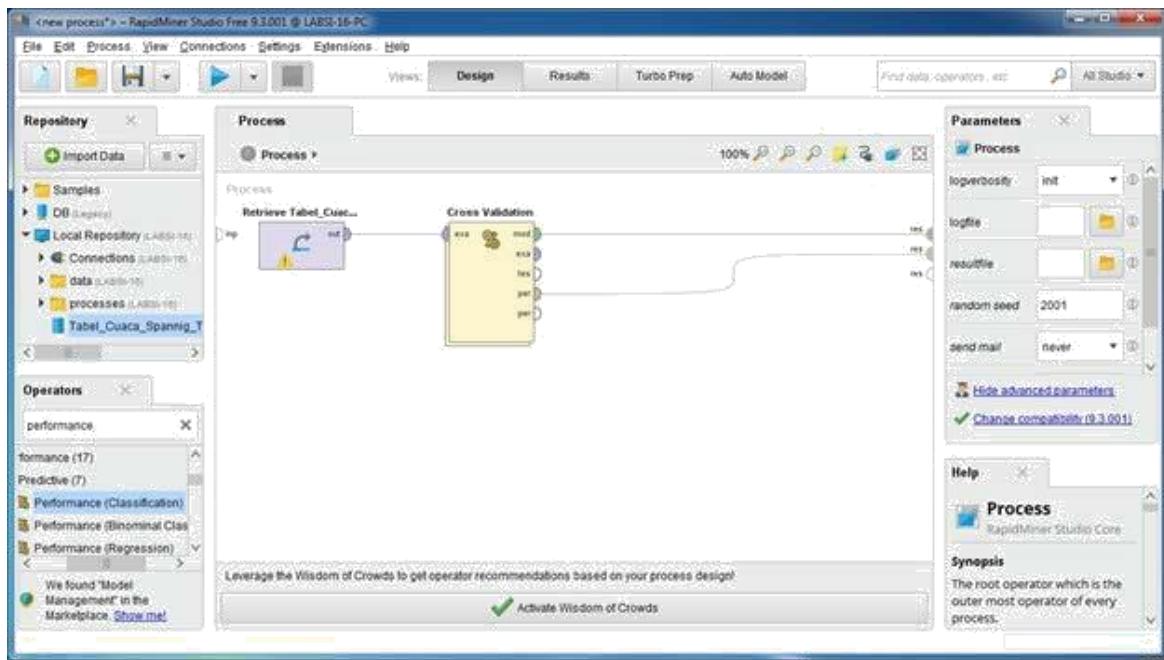


3. Klik kanan pada **Result list** dan pilih **Visual tree**, untuk melihat **Pohon Keputusannya**.

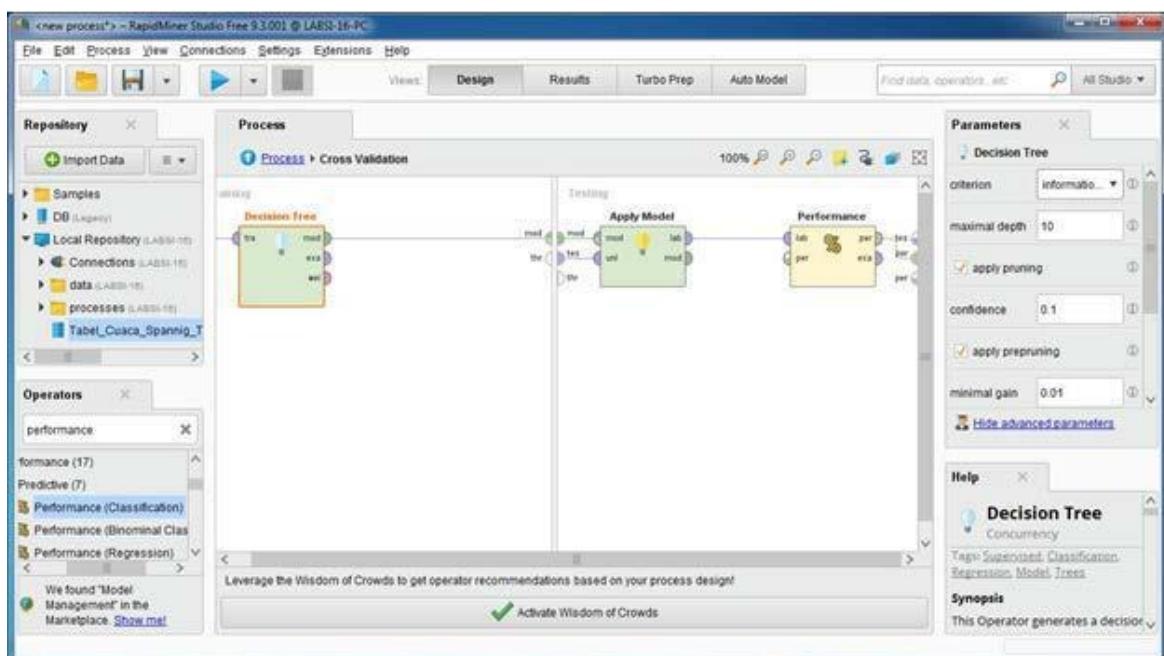


9.4.2 Pohon Keputusan Menggunakan RapidMiner

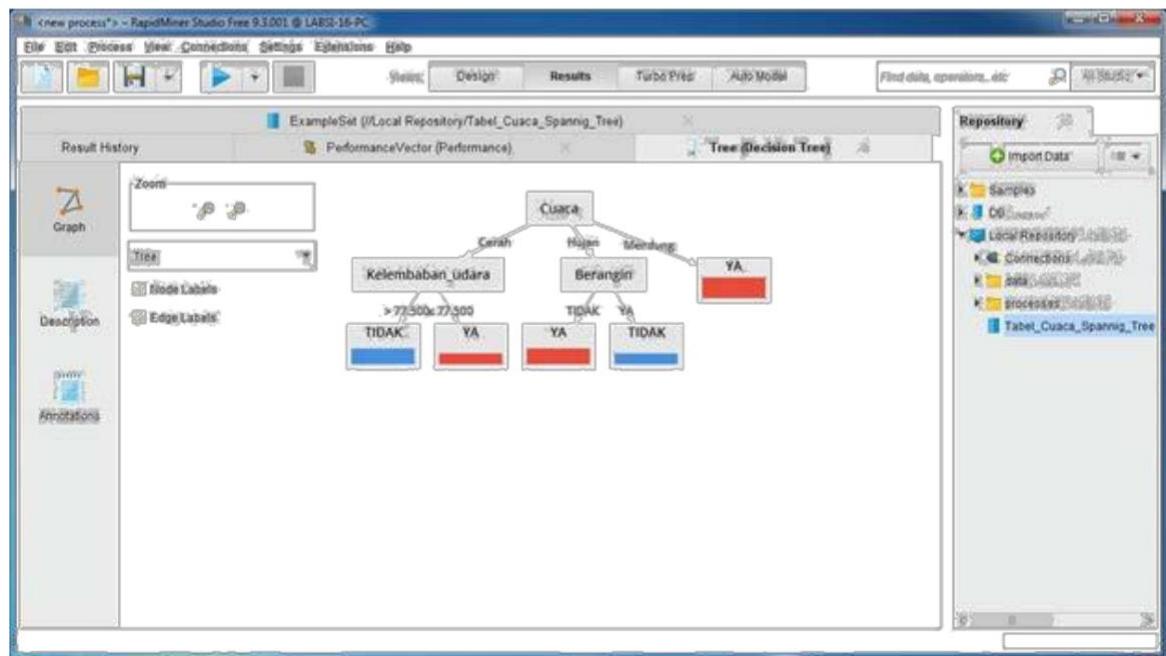
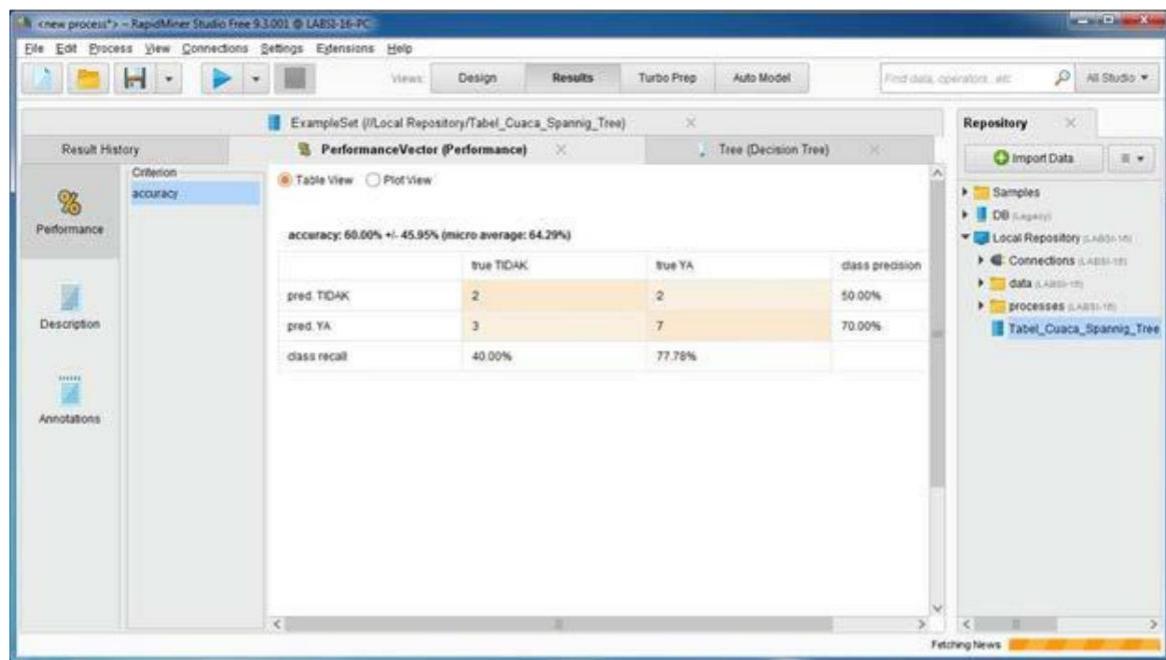
1. Buka **RapidMiner** dan klik **New Process** => **Blank** pada halaman perspective RapidMiner dan buat data Training, masukkan data training tersebut ke kolom **Process** lalu cari pada kolom **Operators** ketikkan **cross validation**.



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



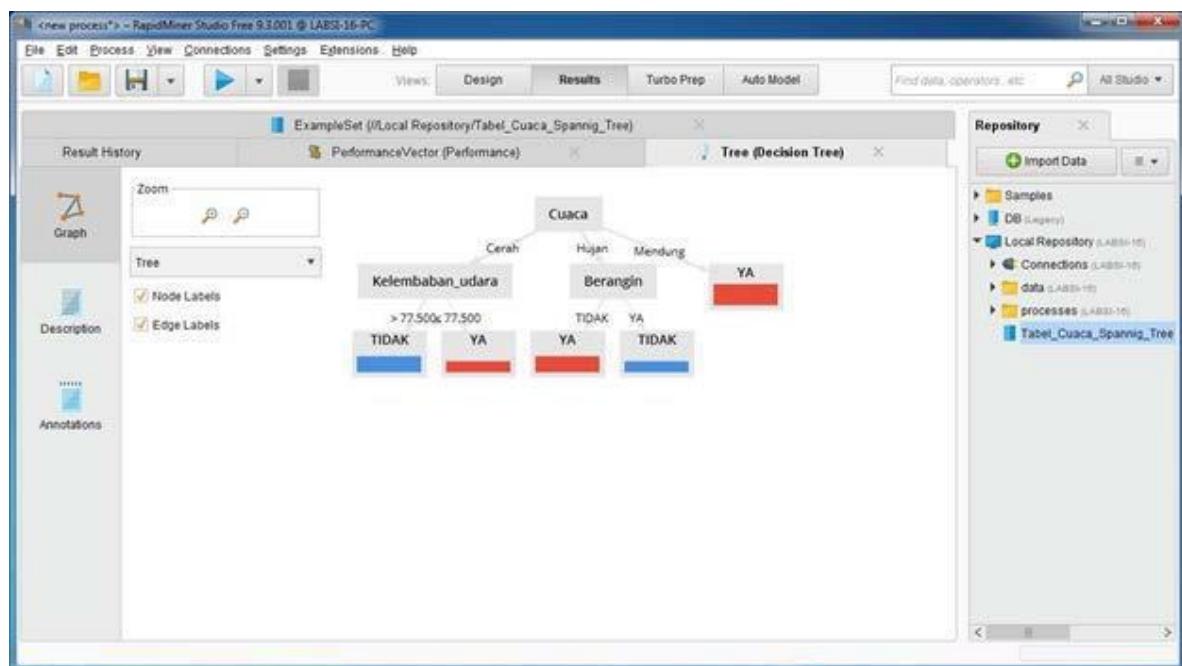
3. Klik **Start (Simbol segitiga)** lalu lihat hasilnya.



9.5 Tugas

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

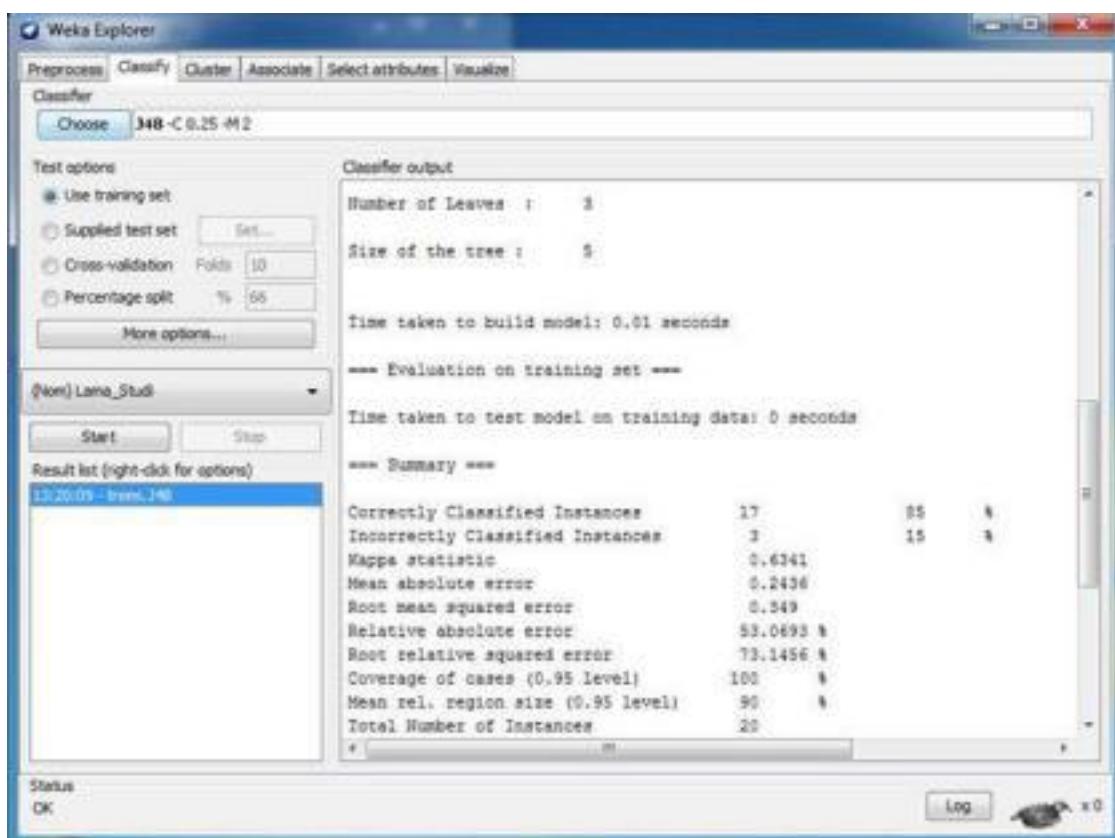
Cuaca	Suhu	Kelembapan_udara	Berangin	Bermain_Tenis
Cerah	75	65	TIDAK	YA
Cerah	80	68	YA	YA
Cerah	83	87	YA	TIDAK
Mendung	70	96	TIDAK	YA
Mendung	68	81	TIDAK	YA
Hujan	65	75	TIDAK	YA
Hujan	64	85	YA	TIDAK



2.

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

a) Buatlah dan cetaklah pohon keputusan berdasarkan data tersebut!

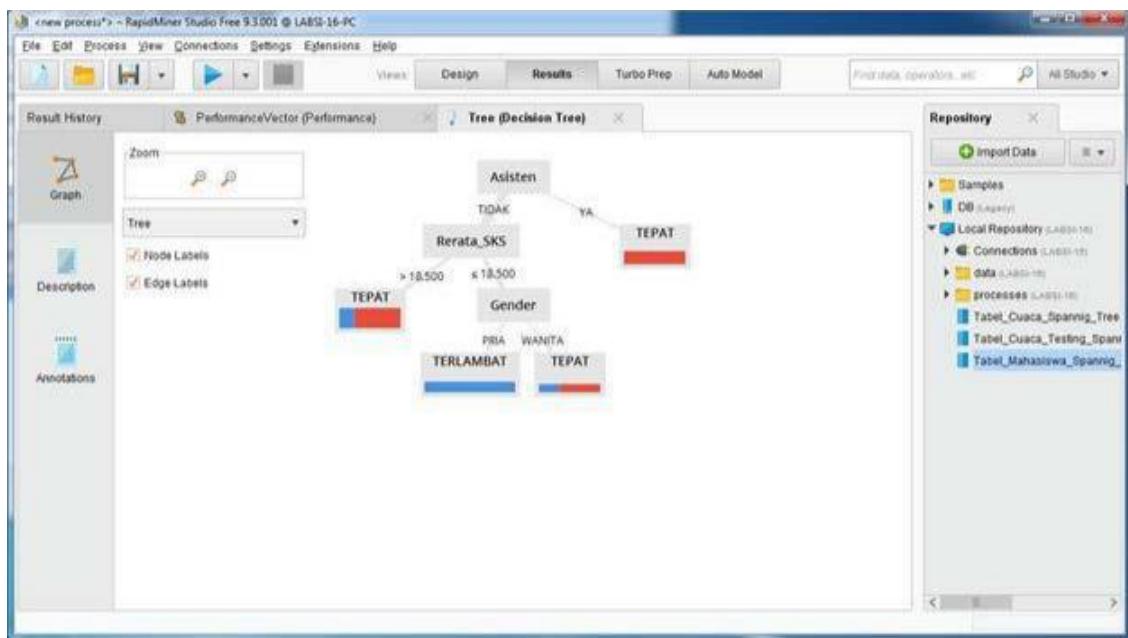


b) Carilah nilai-nilai parameter berikut :

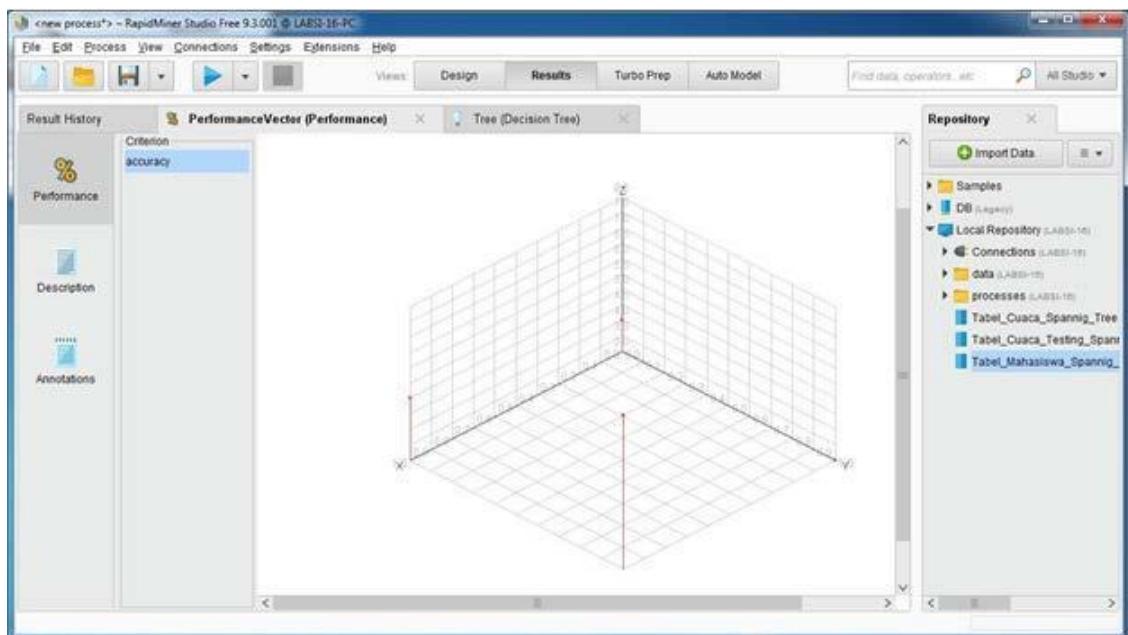
- i. Jumlah simpul daun pada pohon keputusan = **3** ii.
Jumlah simpul keseluruhan pada pohon keputusan = **5** iii.
Waktu yang dibutuhkan untuk proses pelatihan = **0**
- iv. Tingkat ketepatan klasifikasi = **85%** v.
Tingkat ketidaktepatan klasifikasi = **15%** Gunakan file Excel yang dikerjakan pada Tugas nomor 1 dalam Modul 6 sebagai data training.

a) Buatlah dan cetaklah pohon keputusan berdasarkan data tersebut!

3.

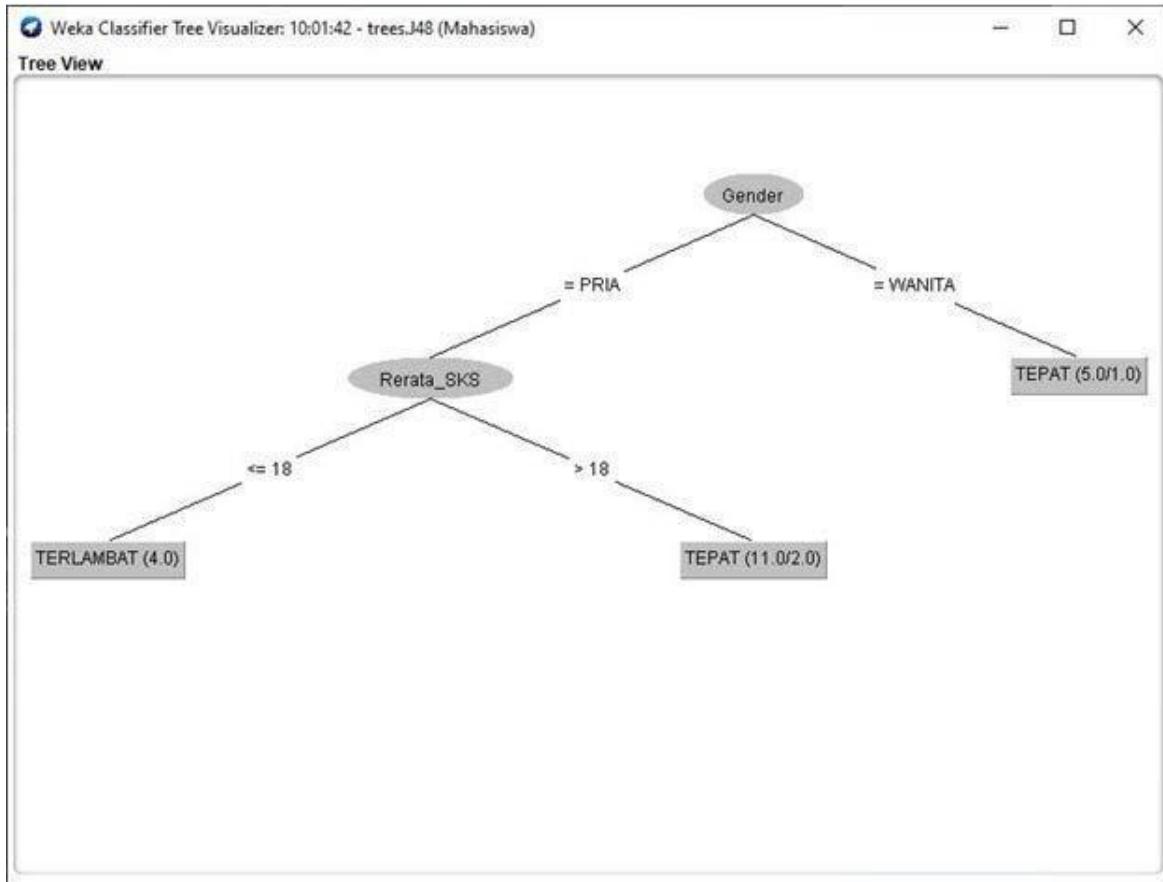


- b) Cetaklah Perspektif Plot View dengan model Scatter. X-Axis = Gender, Y-Axis = Asisten, dan Color Column = Lama_Studi. Nilai Jitter bisa diubah-ubah untuk memperoleh pola penyebaran yang lebih jelas.



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

4.



Klasifikasi yang terbentuk yaitu,

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

+-----+

MODUL X CLUSTERING : K-MEANS

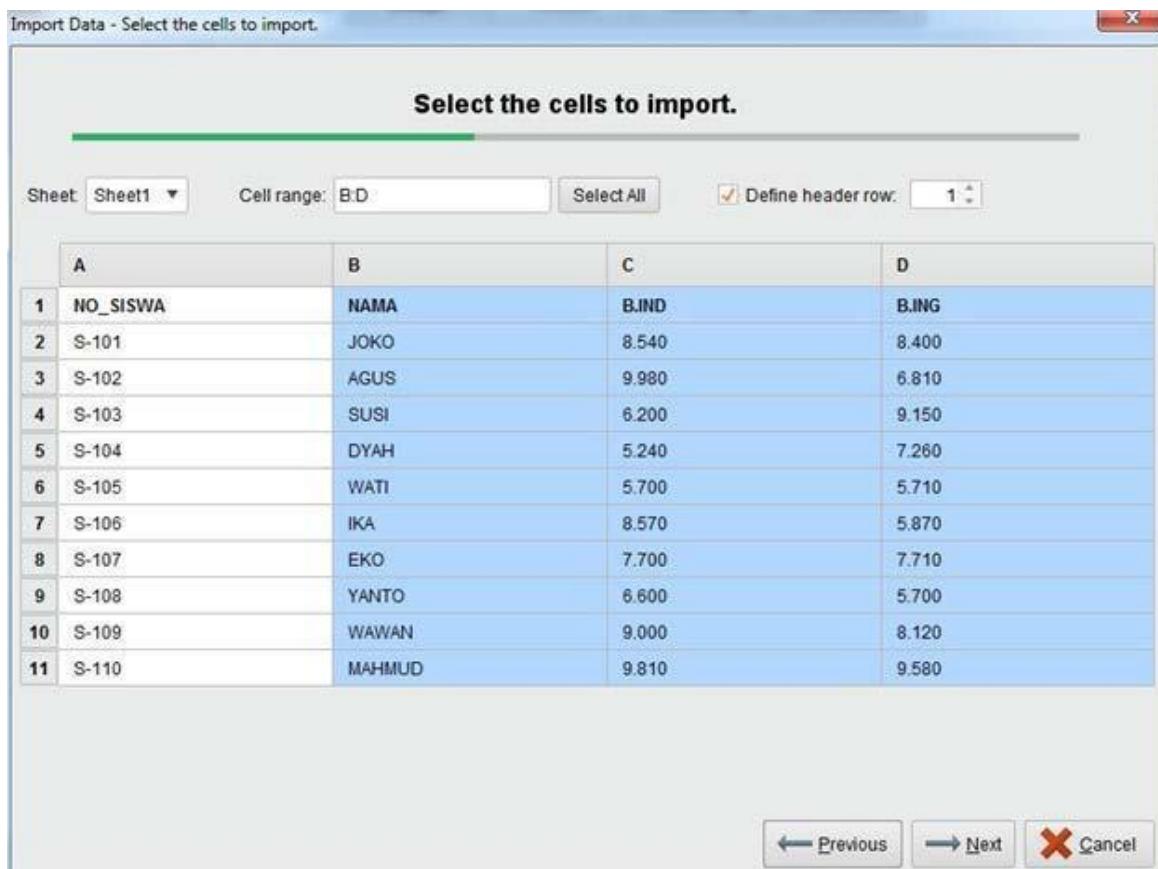
10.4 Langkah-langkah Praktikum

10.4.1 Algoritma K-Means Menggunakan RapidMiner

0. Membuat Tabel data data mahasiswa dan menyimpannya dengan nama Tabel_NilaiUjian.xls.

NO_SISWA	NAMA	B.IND	B.ING
S-101	JOKO	8.54	8.4
S-102	AGUS	9.98	6.81
S-103	SUSI	6.2	9.15
S-104	DYAH	5.24	7.26
S-105	WATI	5.7	5.71
S-106	IKA	8.57	5.87
S-107	EKO	7.7	7.71
S-108	YANTO	6.6	5.7
S-109	WAWAN	9	8.12
S-110	MAHMUD	9.81	9.58

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



Import Data - Format your columns.

Format your columns.

Replace 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

Import Data - Where to store the data?

Where to store the data?

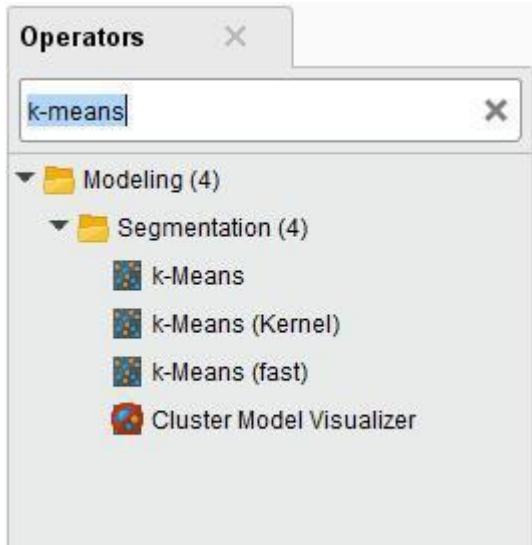
Local Repository (LABSI-16)

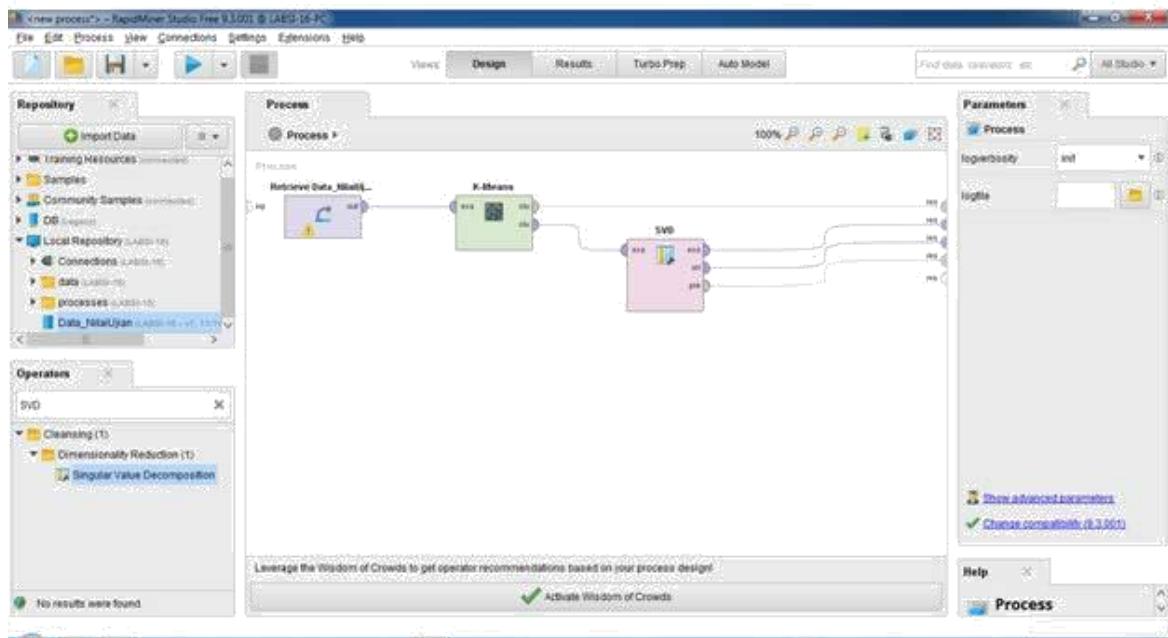
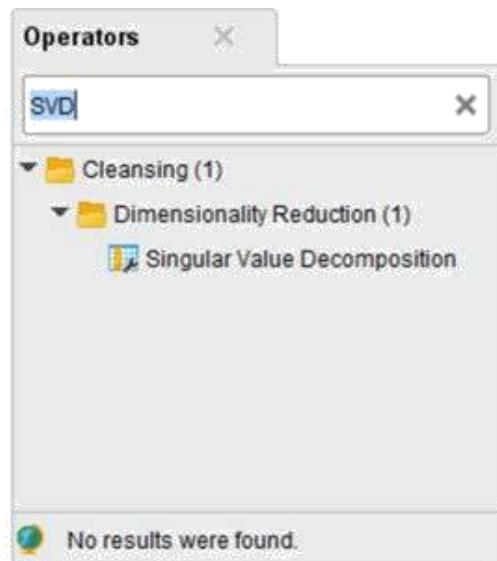
- Connections (LABSI-16)
- data (LABSI-16)
- processes (LABSI-16)

Name: Data_NilaiUjian

Location: /Local Repository/Data_NilaiUjian

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



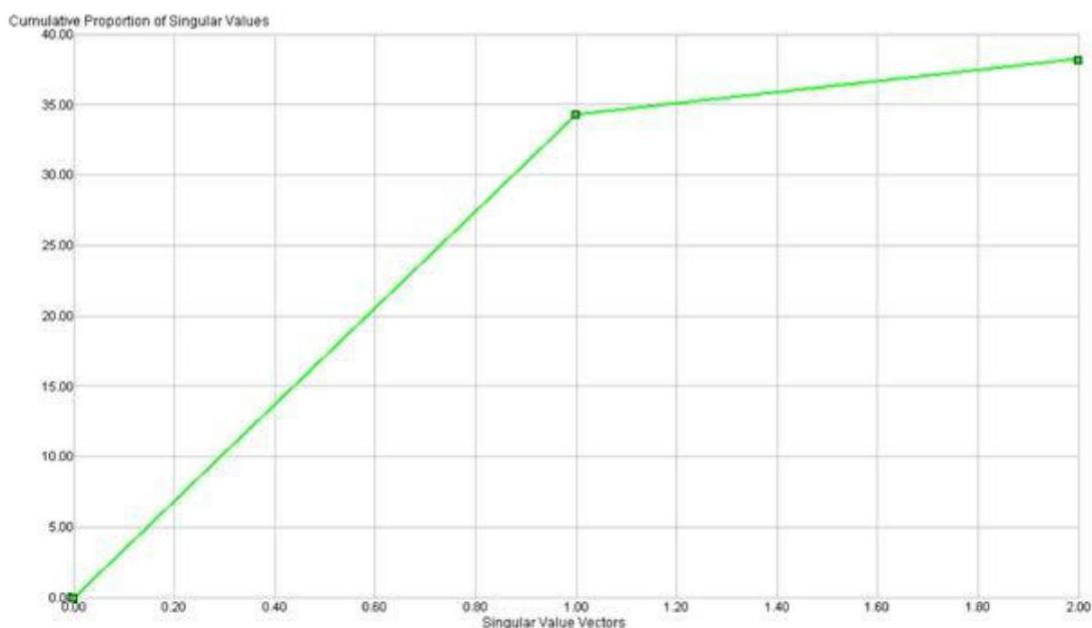


4. Setelah jalankan **Run**, hasilnya berupa :

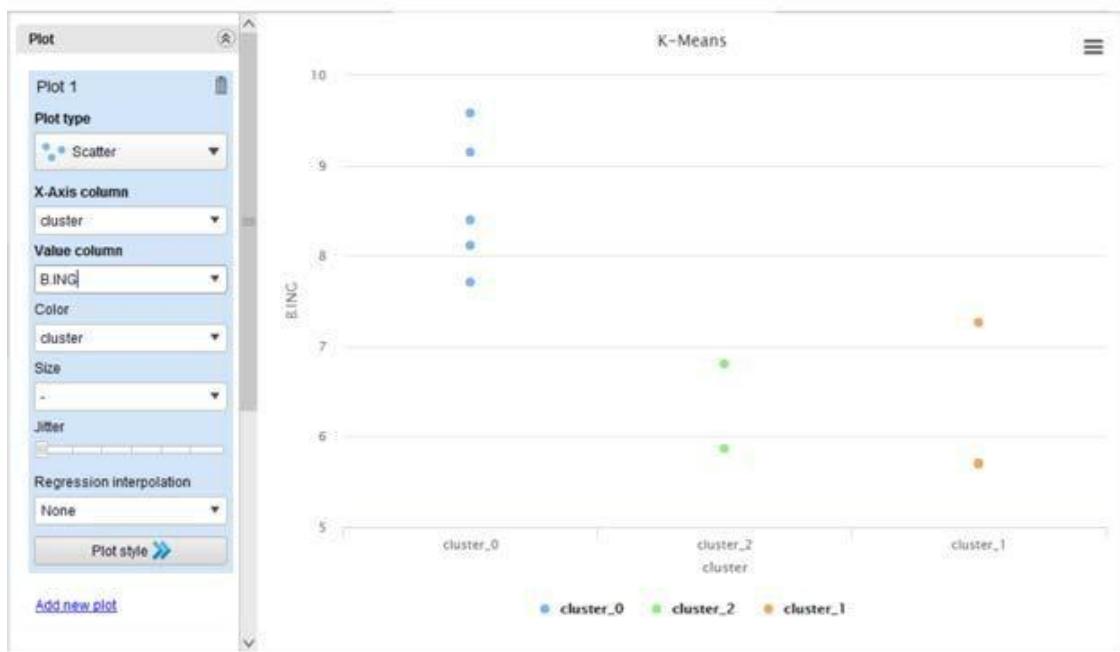
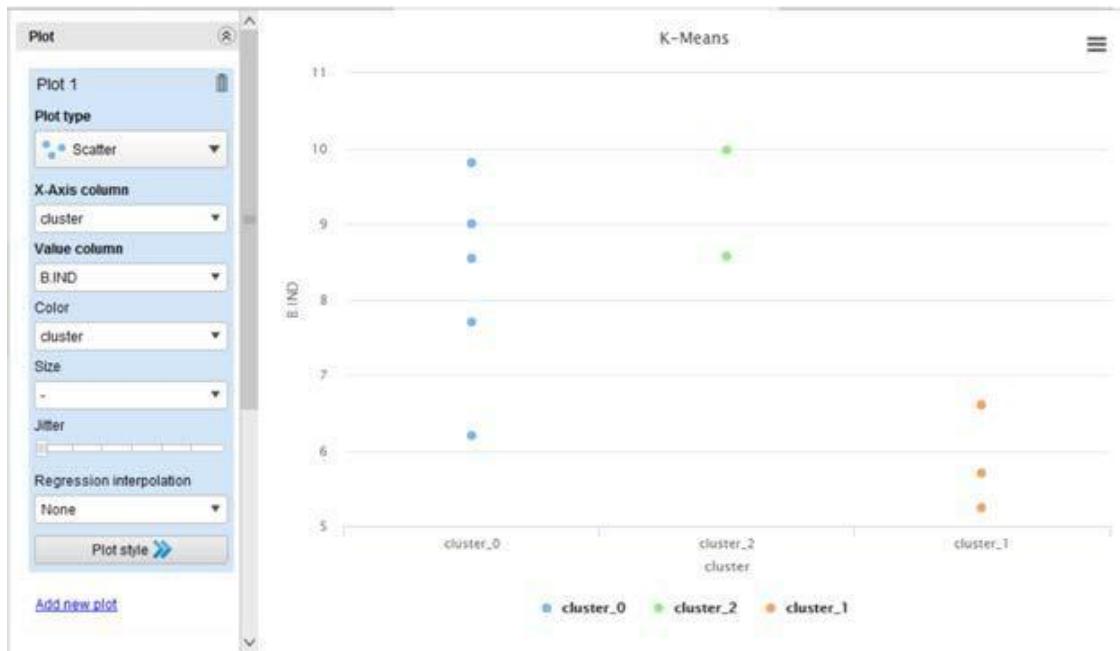
a) SVD (Singular Value Decomposition)

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

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



b) ExampleSet (K-Means)



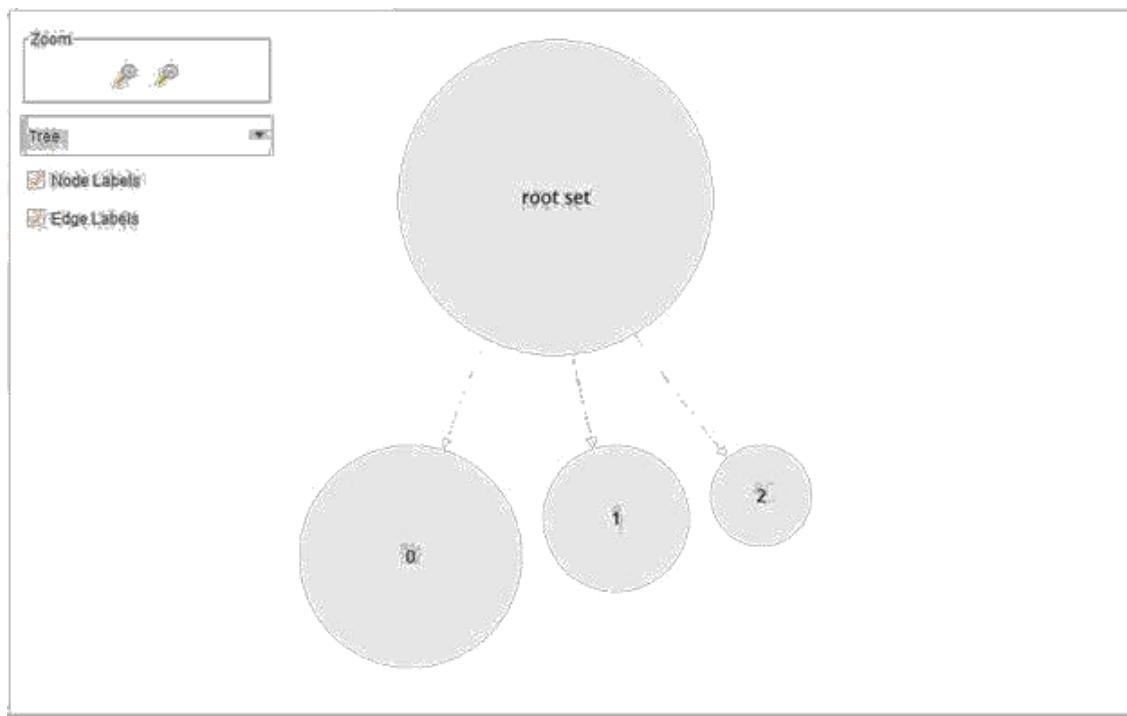
c) ExampleSet (SVD)

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

d) Cluster Model (Clustering)

Cluster Model

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



10.4.2 Interpretasi Hasil Algoritma K-Means

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

10.5 Tugas

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

D. Buatlah table berikut dengan menggunakan Microsoft Excel!

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

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

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

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

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

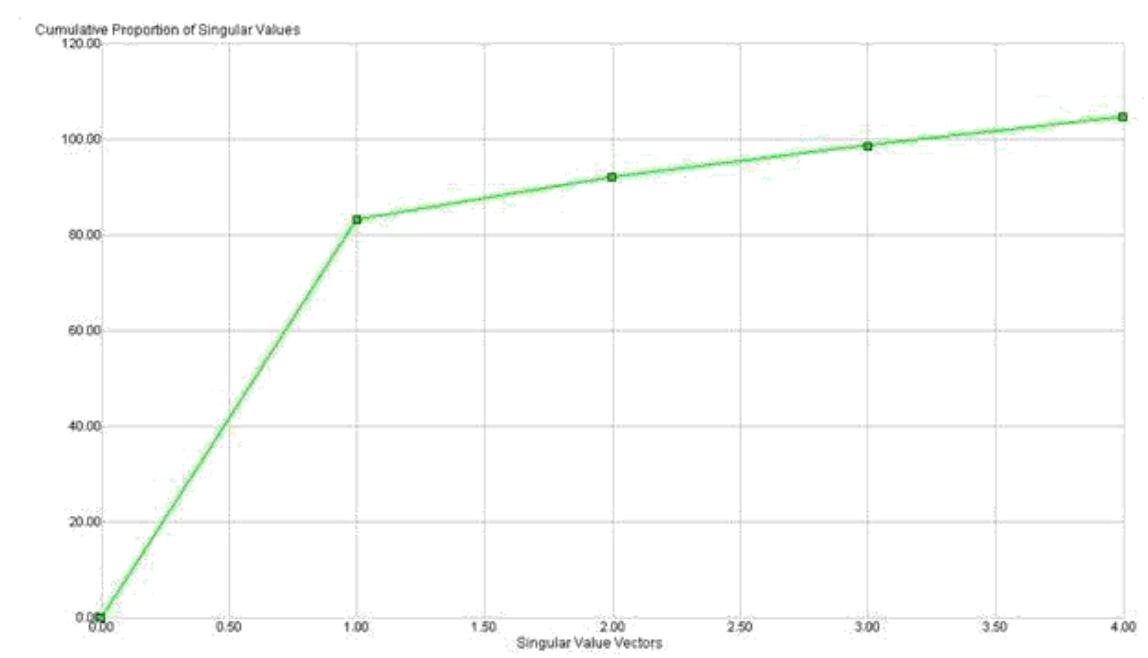
The screenshot shows the KNIME Analytics Platform interface with the following details:

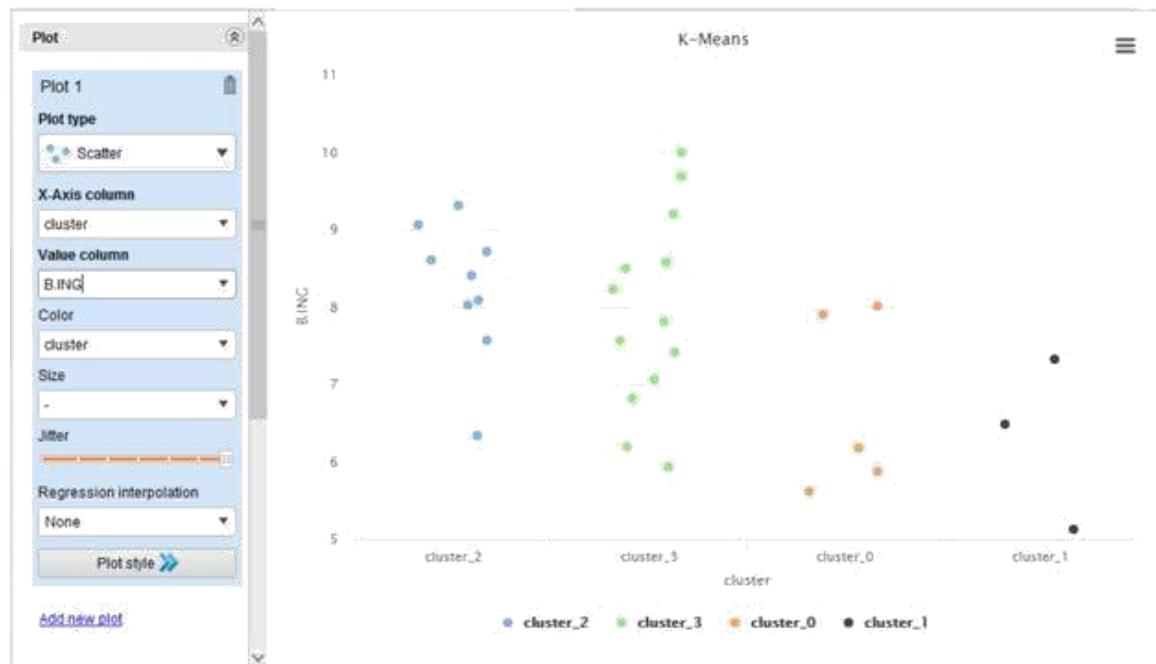
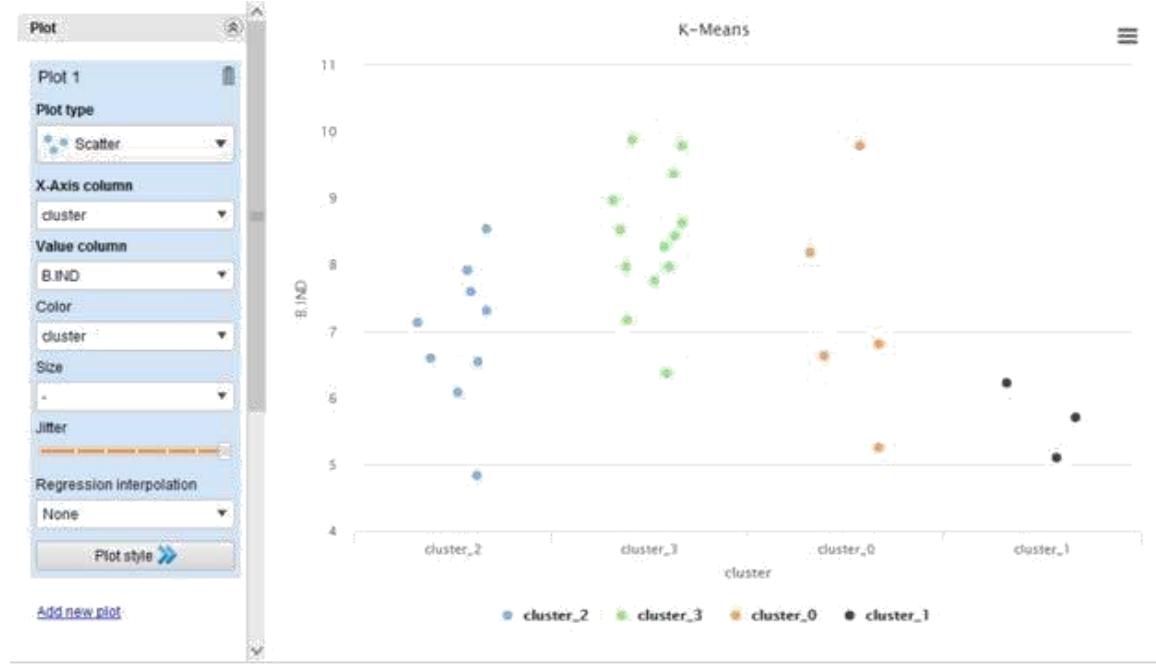
- Repository:** Local Repository contains a process named "Tabel_Nilai_ujian_30_Siswa".
- Operations:** Dimensionality Reduction (1) includes Singular Value Decomposition (SVD).
- Process:**
 - Start node → "Retrieved Train Data" node (blue)
 - "Retrieved Train Data" → "SVD" node (green)
 - "SVD" node → "k-Means" node (orange)
 - "k-Means" node → End node (grey)
- Parameters:**
 - Number of Clusters: 4
 - Add cluster attribute: checked
 - Add as label: checked
 - Remove unlabeled: checked
 - Response: 10
 - Determine good cluster size: checked
 - Measure type: Euclidean
 - Normalisation: Euclidean
 - Max iteration: 100
 - Use advanced strategies: checked
 - Cluster compatibility (0.1-0.5): checked
- Help:** k-Means

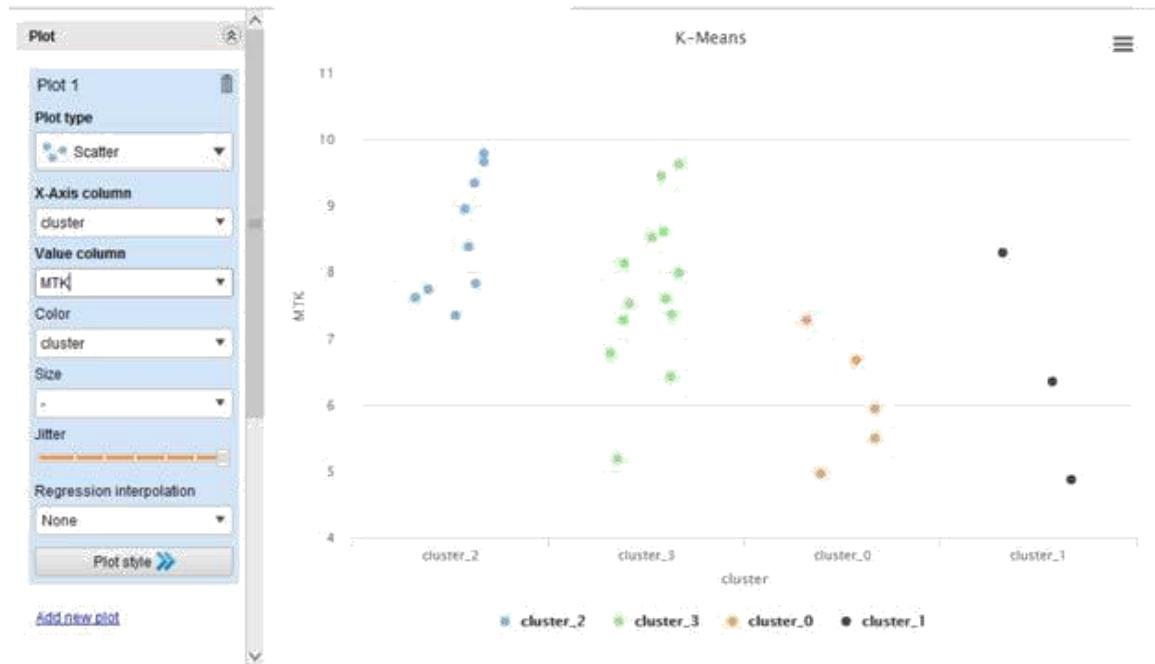
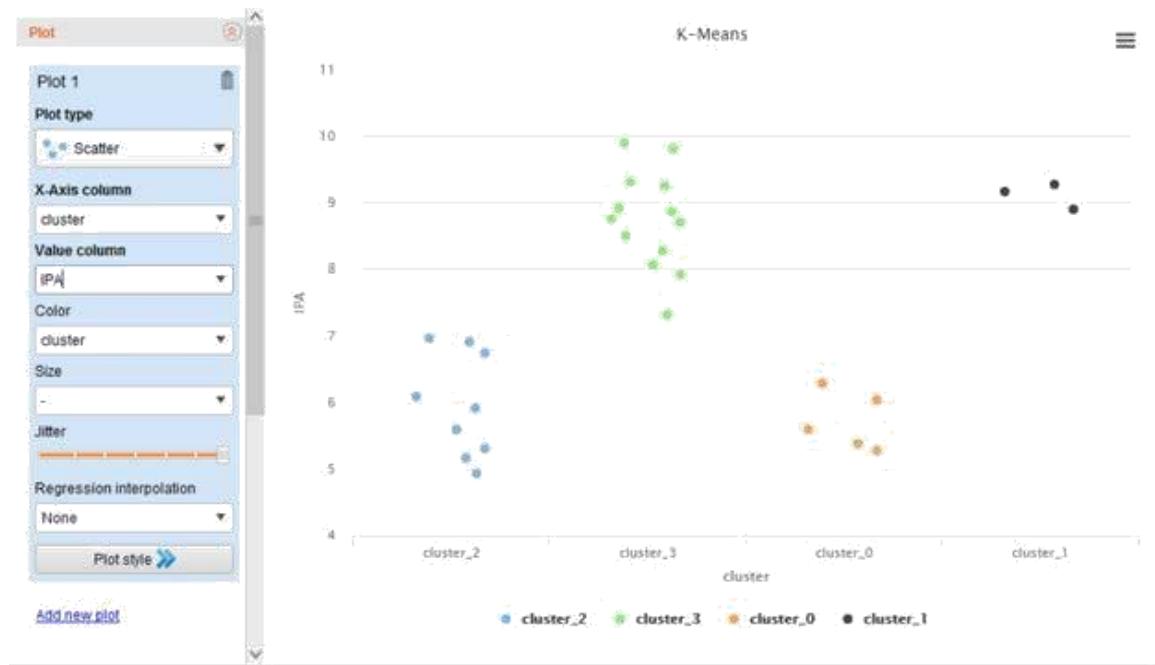
Below the process view, there is a table titled "Component" showing the results of the Singular Value Decomposition:

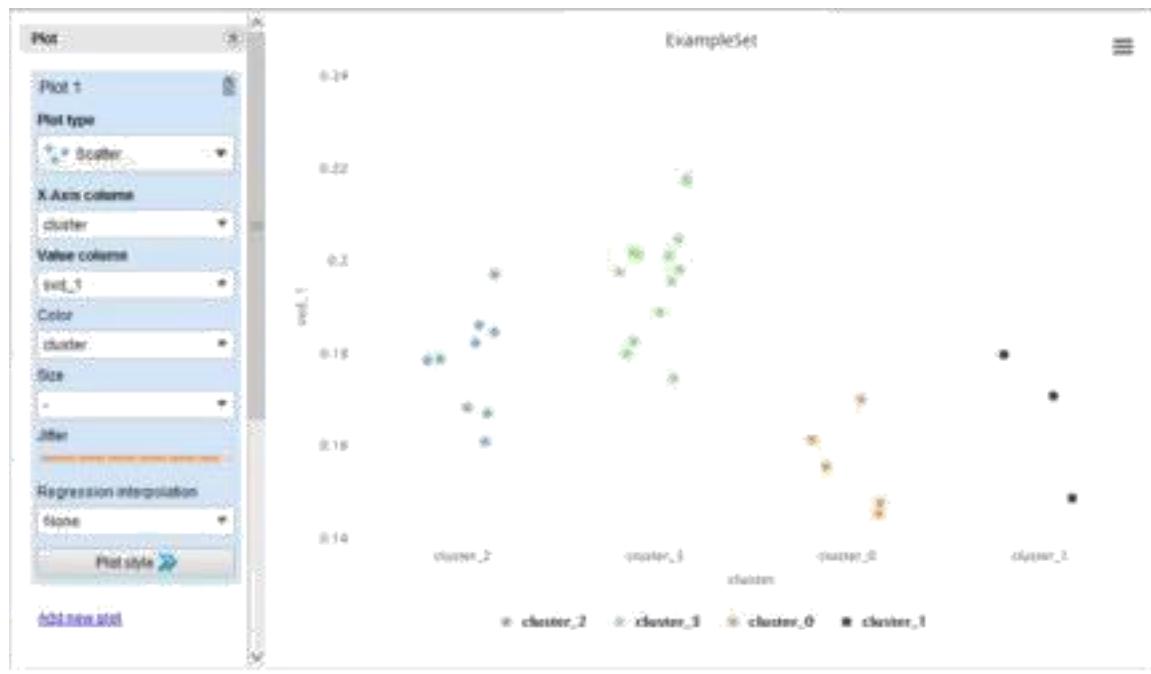
Component	Singular Value	Proportion of Singular Values	Cumulative Singular Values	Cumulative Proportion of Sing...
SVD 1	83.180	0.794	83.180	0.794
SVD 2	8.870	0.065	92.050	0.879
SVD 3	6.571	0.063	98.621	0.942
SVD 4	6.085	0.058	104.706	1.000

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









Cluster Model

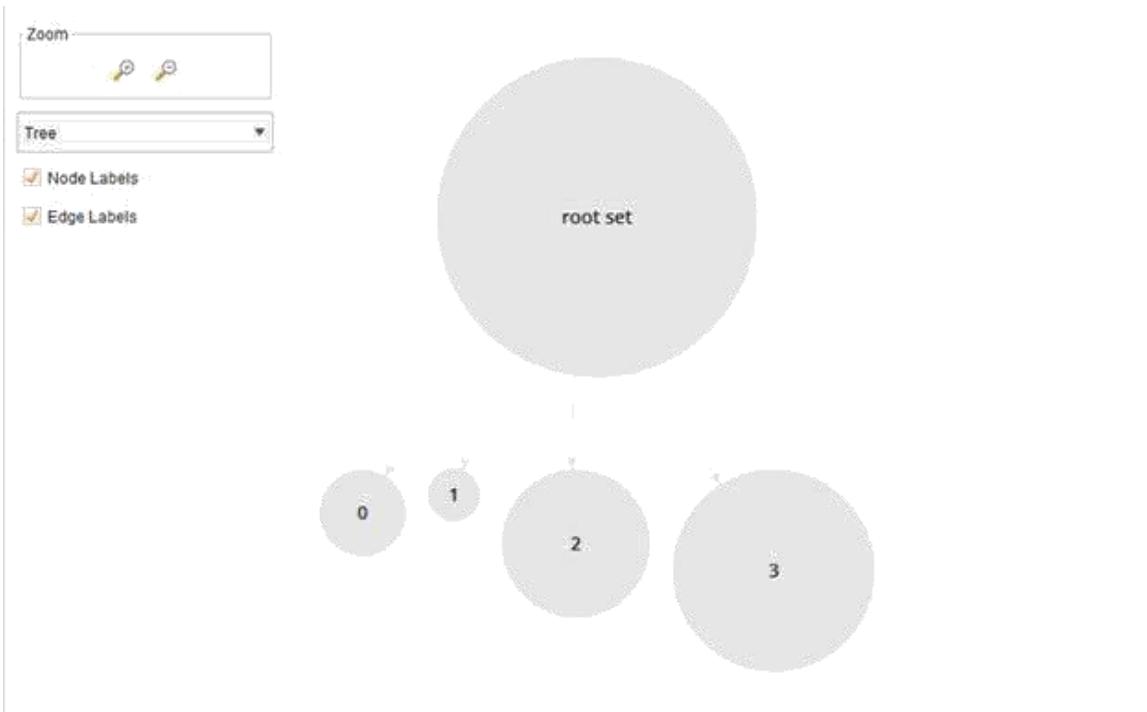
Cluster 0: 5 items

Cluster 1: 3 items

Cluster 2: 9 items

Cluster 3: 13 items

Total number of items: 30



Row No.	NAMA	cluster	B.IND	BJING	MTK	IPA
1	JOKO	cluster_2	6.955	8.886	7.428	5.907
2	AGUS	cluster_3	9.064	8.310	6.887	8.861
3	SUSI	cluster_3	8.023	7.566	9.202	8.020
4	DYAH	cluster_3	8.862	9.904	9.848	8.139
5	WATI	cluster_3	7.449	6.452	8.400	8.779
6	IKA	cluster_3	9.514	9.724	7.704	8.424
7	EKO	cluster_2	8.084	8.187	9.119	5.320
8	YANTO	cluster_0	8.340	5.751	7.423	5.738
9	WAWAN	cluster_3	6.153	8.363	8.392	9.038
10	MAHMUD	cluster_0	9.993	6.362	6.873	5.573
11	BUDI	cluster_0	7.025	6.079	6.160	5.481
12	SANTI	cluster_3	9.521	9.350	6.587	9.017
13	DIAN	cluster_3	8.405	7.383	7.346	9.785
14	DANI	cluster_1	6.045	6.315	8.112	8.985

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

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

➤ Cluster 0

- YANTO
- MAHMUD
- BUDI
- BAYU
- JONO

➤ Cluster 1

- DANI
- YANI
- FEBRI

➤ Cluster 2

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

➤ Cluster 3

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

MODUL XI INDUKSI DAN ATURAN ASOSIASI

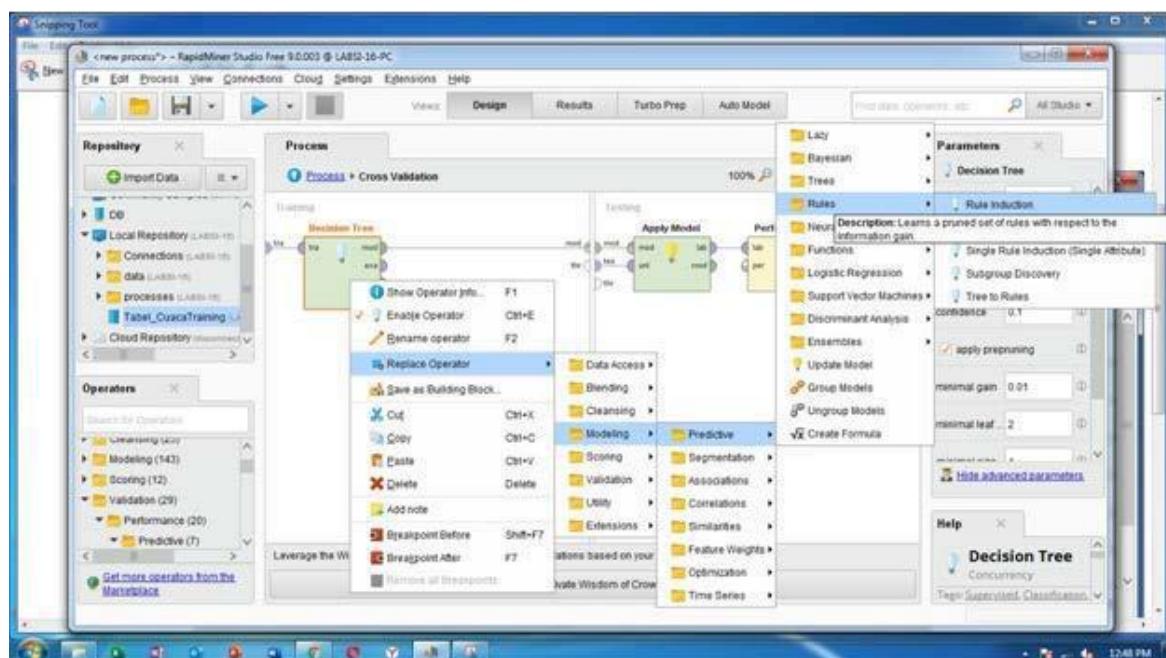
11.4 Langkah-langkah Praktikum

11.4.1 Induksi Aturan Data Cuaca

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



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



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

RuleModel

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

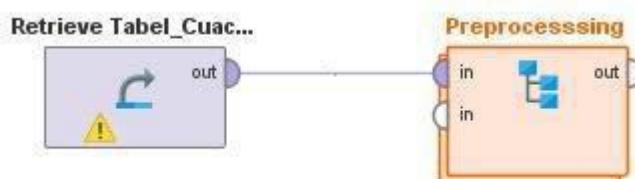
correct: 12 out of 13 training examples.

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

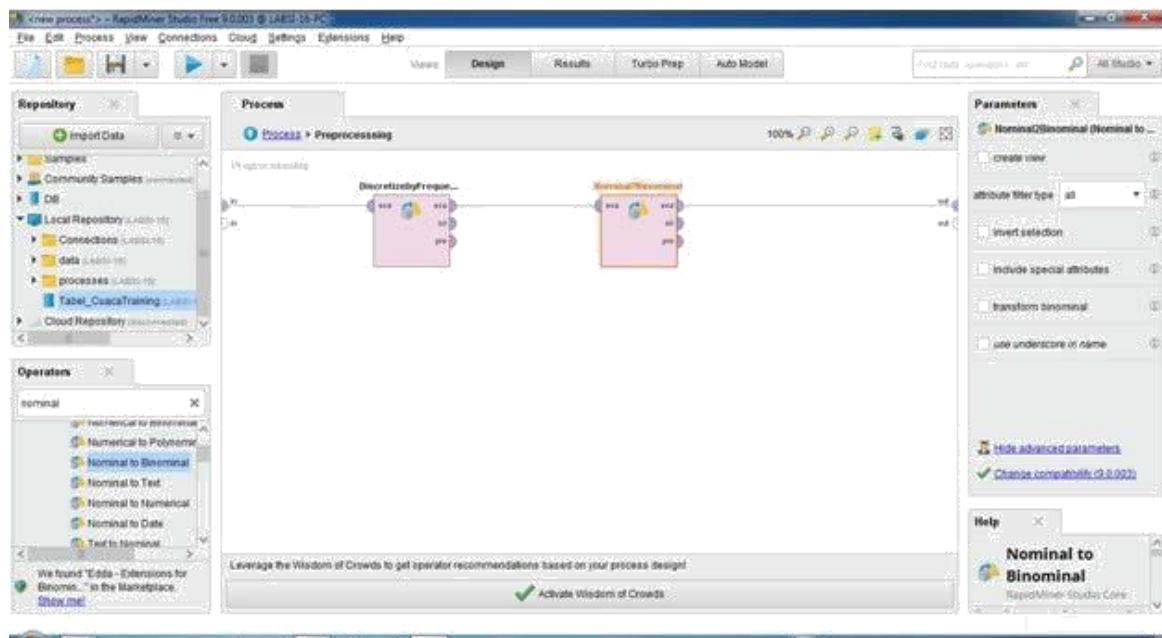
	true TIDAK	true YA	class precision
pred. TIDAK	2	1	66.67%
pred. YA	3	8	72.73%
class recall	40.00%	88.89%	

11.4.2 Aturan Asosiasi Data Cuaca

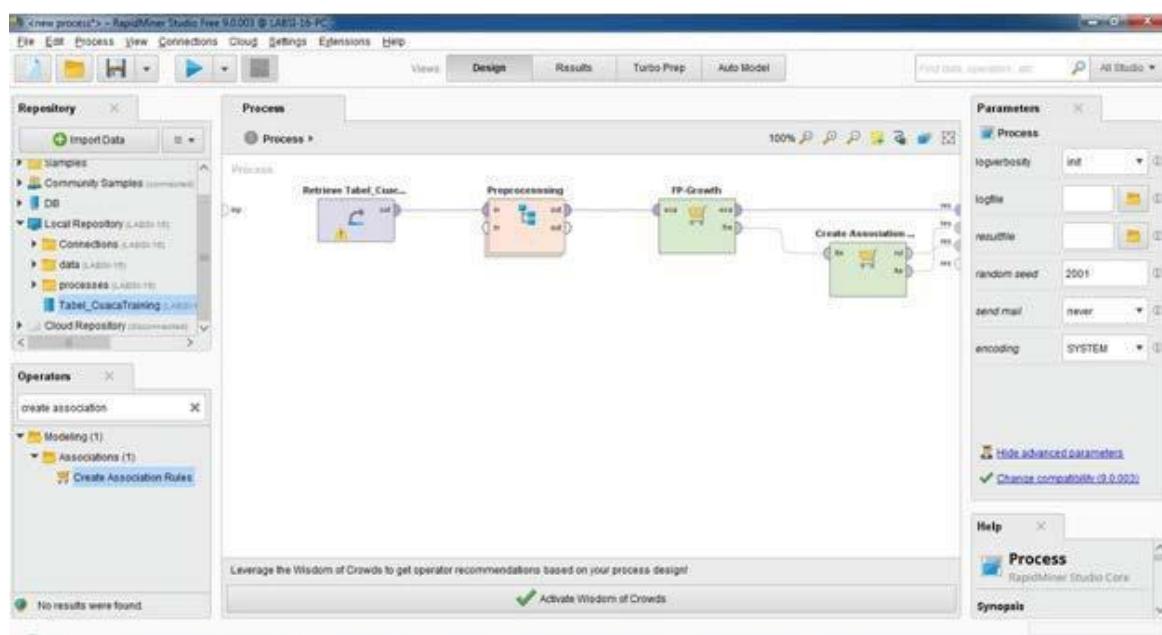
- Buat **New Process** dengan menekan tombol **New**, kembali gunakan data **Cuaca Training** dan drag ke repository **Process View**, tambahkan operator **subprocess** dan ubah Namanya menjadi **Subprocess**.



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



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



4. Klik **Run** dan tampilkan :

➤ Frequent Item Set (FP-Growth)

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

No. of Sets: 26	Size	Support	Item 1	Item 2	Item 3	Item 4
Total Max. Size: 4	2	0.143	Kelembaban_udara	Cuaca = Mendung		
Min. Size: 1	2	0.143	Berangin	Suhu		
Max. Size: 4	2	0.143	Berangin	Cuaca = Cerah		
Contains Item:	2	0.143	Berangin	Cuaca = Hujan		
	2	0.143	Berangin	Cuaca = Mendung		
Update View	2	0.214	Suhu	Cuaca = Cerah		
	2	0.071	Suhu	Cuaca = Hujan		
	2	0.143	Suhu	Cuaca = Mendung		
	3	0.071	Kelembaban_udara	Berangin	Suhu	
	3	0.071	Kelembaban_udara	Berangin	Cuaca = Cerah	
	3	0.071	Kelembaban_udara	Berangin	Cuaca = Hujan	
	3	0.071	Kelembaban_udara	Berangin	Cuaca = Mendung	
	3	0.143	Kelembaban_udara	Suhu	Cuaca = Cerah	
	3	0.071	Kelembaban_udara	Suhu	Cuaca = Mendung	
	3	0.143	Berangin	Suhu	Cuaca = Cerah	
	4	0.071	Kelembaban_udara	Berangin	Suhu	Cuaca = Cerah

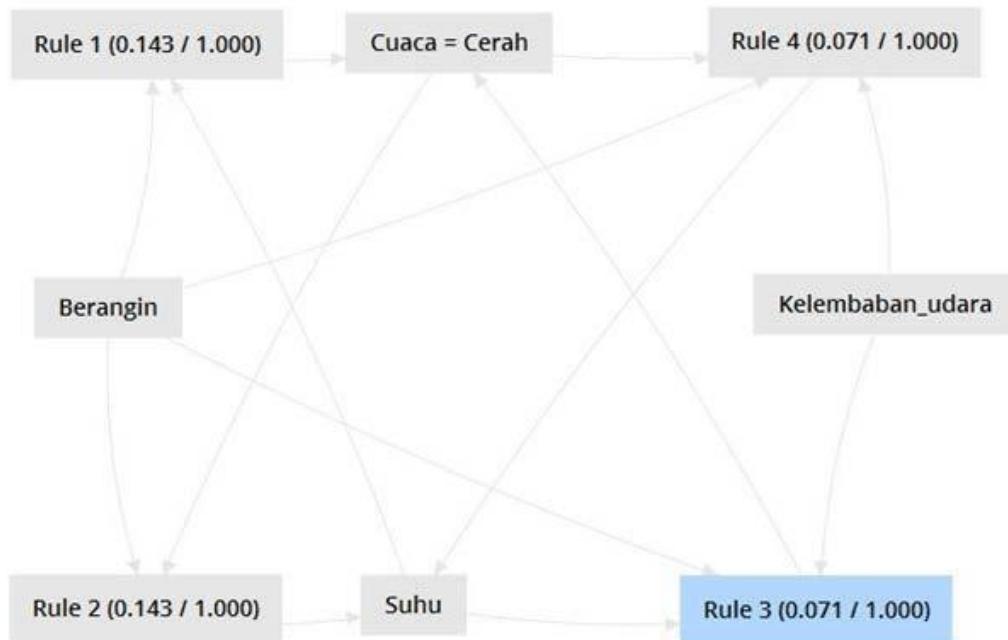
➤ **Association Rules (Create Association Rules)**

a) **Table Mode**

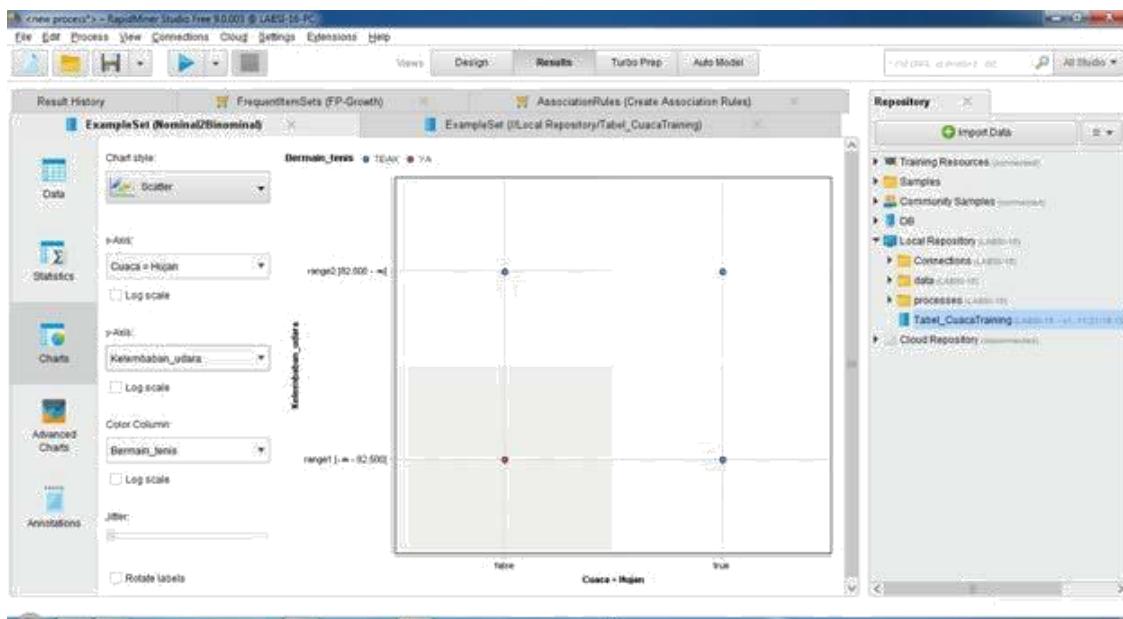
No.	Premises	Conclusion
1	Berangin, Suhu	Cuaca = Cerah
2	Berangin, Cuaca = Cerah	Suhu
3	Kelembaban_udara, Berangin, Suhu	Cuaca = Cerah
4	Kelembaban_udara, Berangin, Cuaca = Cerah	Suhu

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

b) Graph View



➤ ExampleSet (Nominal2Binomial)



11.5 Tugas

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

RuleModel

```

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

correct: 17 out of 19 training examples.

```

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

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

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

- a) Number of bins = 2
- b) Number of bins = 3

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

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

➤ Number of bins = 2

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

➤ Number of bins = 3

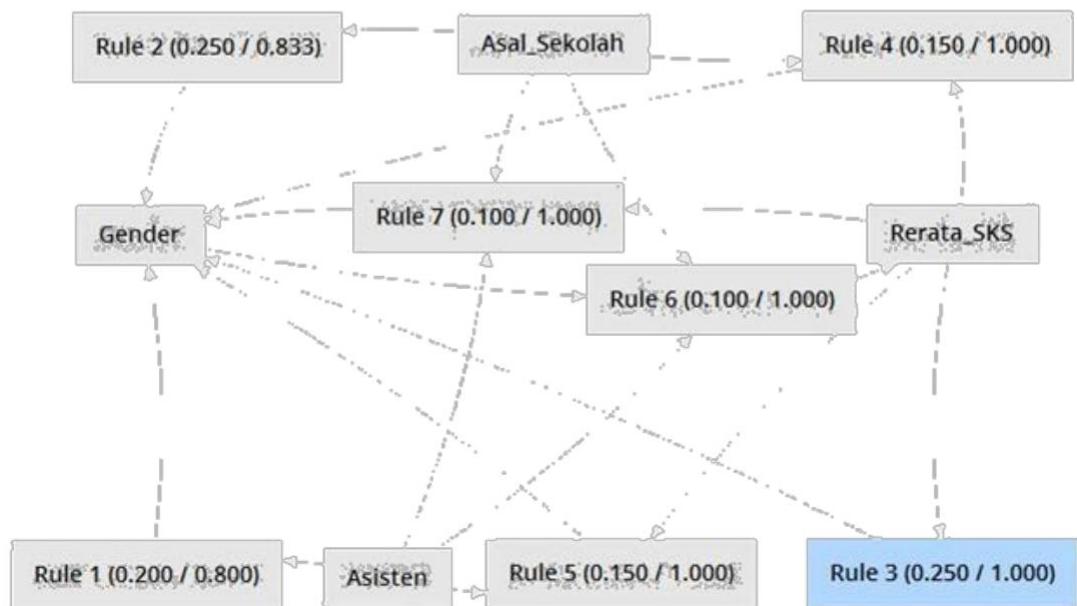
No. of Sets: 7	Size	Support	Item 1	Item 2	Item 3
Total Max. Size: 3	1	0.750	Gender		
Min. Size: <input type="text" value="1"/>	1	0.300	Asal_Sekolah		
Max. Size: <input type="text" value="3"/>	1	0.250	Asisten		
Contains Item:	2	0.250	Gender	Asal_Sekolah	
	2	0.200	Gender	Asisten	
<input type="button" value="Update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

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

➤ Number of bins = 2

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

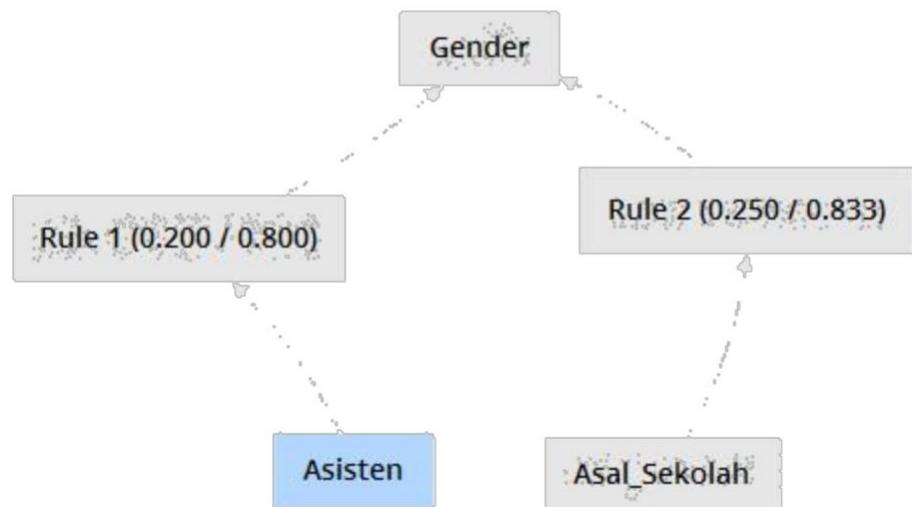
Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
0.250	0.833	0.962	-0.350	0.025	1.111	1.500
0.250	1	1	-0.250	0.062	1.333	∞
0.150	1	1	-0.150	0.038	1.333	∞
0.150	1	1	-0.150	0.038	1.333	∞
0.100	1	1	-0.100	0.075	4	∞
0.100	1	1	-0.100	0.025	1.333	∞



➤ Number of bins = 3

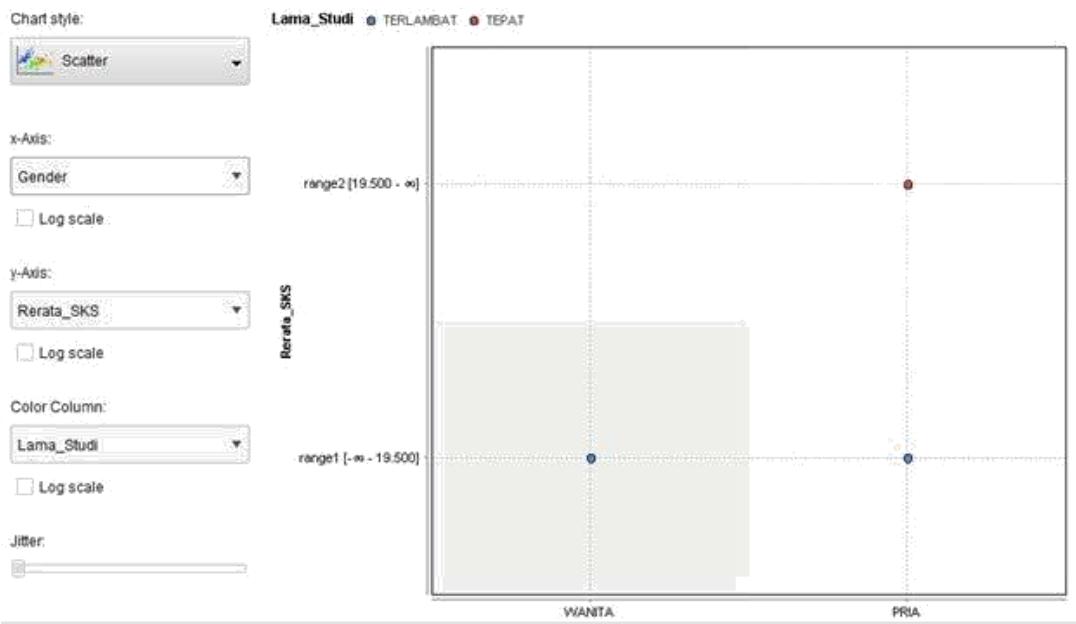
No.	Premises	Conclusion
2	Asal_Sekolah	Gender

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

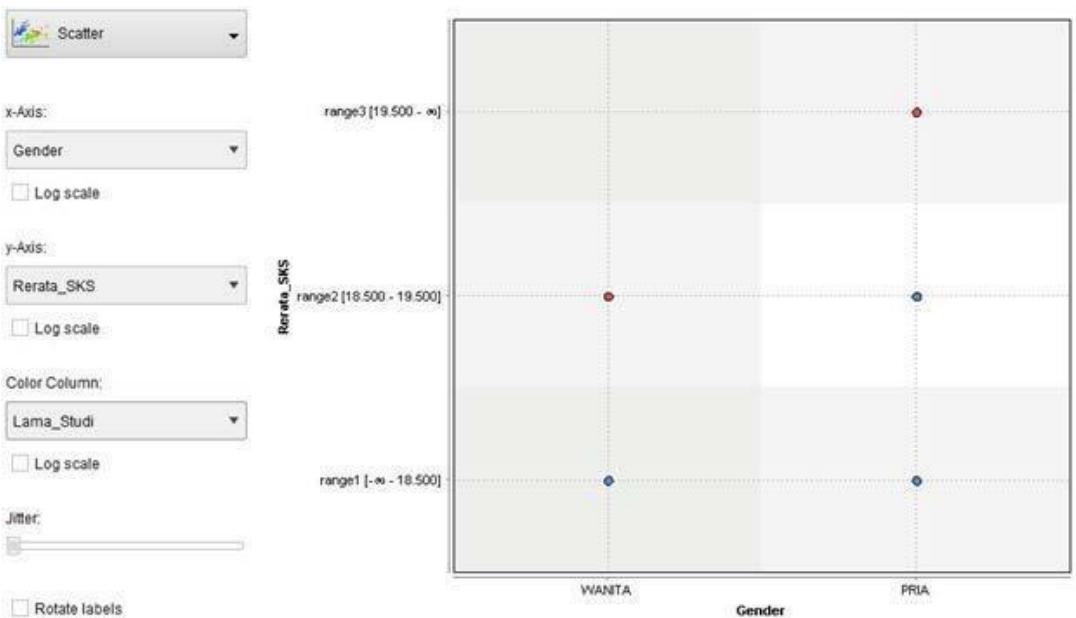


c) Gambarkan grafik chart pola distribusi data pada ExampleSet yang terbentuk!

- Number of bins = 2



Number of bins = 3



+-----+

MODUL XII REGRESI LINIER SEDERHANA

12.4 Langkah-langkah Praktikum Contoh

Kasus :

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

12.4.1 Induksi Aturan Data Cuaca

Hipotesis : Bagaimana

1. Membuat data pada Microsoft Excel 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. Mengimport data table **Tabel_LamaBelajardanNilaiUjian.xls** ke **RapidMiner** yang akan digunakan sebagai regresi Linear, dengan memberikan pengaturan tipe dan jenis data pada atribut,

- NO_SISWA : polynomial, id
- NAMA : pilih Exclude column
- LAMA JAM BELAJAR : integer
- NILAI : integer, label

Simpan dengan nama **Data_LamaBelajardanNilaiUjian**.

Import Data - Format your columns.

Format your columns.

Replace errors with missing values ⓘ

No_Siswa <i>polynominal id</i>	Nama <i>polynominal</i>	Lama Belajar (Jam) <i>integer</i>	Nilai <i>integer label</i>
1 S-101	JOKO	15	783
2 S-102	AGUS	18	877
3 S-103	SUSI	7	505
4 S-104	DIYAH	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	WAHAN	8	667
10 S-110	MAHMUD	15	723

 no problems.

 Previous  Next  Cancel

3. Masukkan **Data_LamaBelajar dan NilaiUjian** 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%).

Operators

linear regression

- Modeling (2)
 - Predictive (2)
 - Functions (2)
 - Linear Regression
 - Vector Linear Regression

No results were found.

Process

Process

Previews

Read CSV File

Linear Regression

Parameters

Linear Regression

pre tolerance: 0.05

cope: 1.0E-8

Show advanced parameters

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

➤ Table View (Mencari besarnya nilai t-hitung)

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

➤ Text View (Mencari model regresi)

Linear Regression

$$21.608 * \text{Lama Belajar (Jam)} + 492.769$$

12.4.2 Mencari Nilai t dan Model Regresi Linier Menggunakan RapidMiner

Hipotesis :

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

2. Membuat table data baru di **Tabel_LamaBelajardanNilaiUjian.xls** Pada Sheet 2.

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

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

- NO_SISWA : polynomial, id
- NAMA : pilih Exclude column
- LAMA JAM BELAJAR : integer

Simpan dengan nama **Data_PrediksiNilaiUjian**

Import Data - Format your columns.

Format your columns.

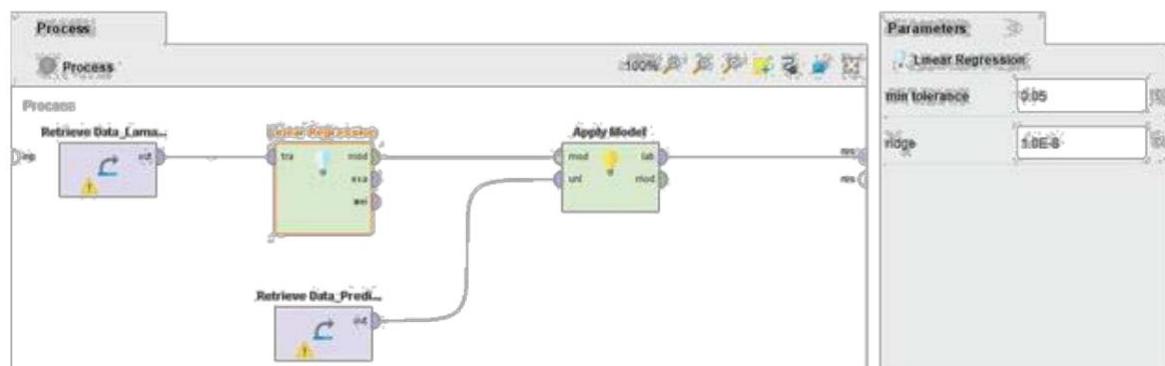
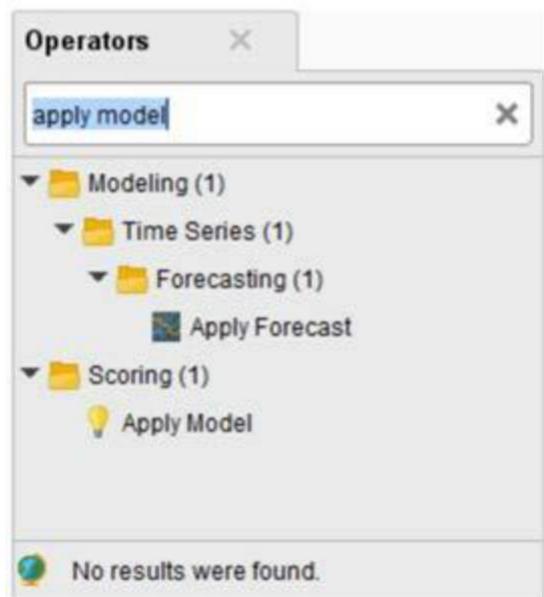
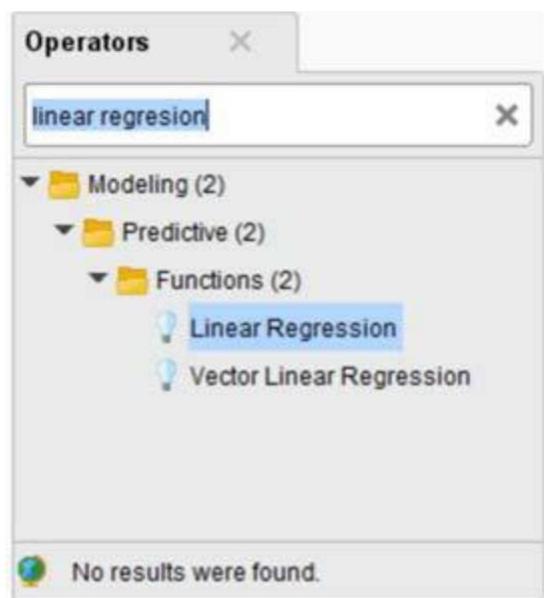
Replace errors with missing values ①

No_Siswa <i>polynominal</i> <i>id</i>	Nama <i>polynominal</i>	Lama Belajar (Jam) <i>integer</i>
1 S-111	BUDI	12
2 S-112	SINTI	13
3 S-113	DIAN	14
4 S-114	DANI	11
5 S-115	AHMAD	5
6 S-116	DAYU	13
7 S-117	RISA	9
8 S-118	RANI	10
9 S-119	YANI	10
10 S-120	RATHI	9

 no problems.

 Previous  Next  Cancel

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

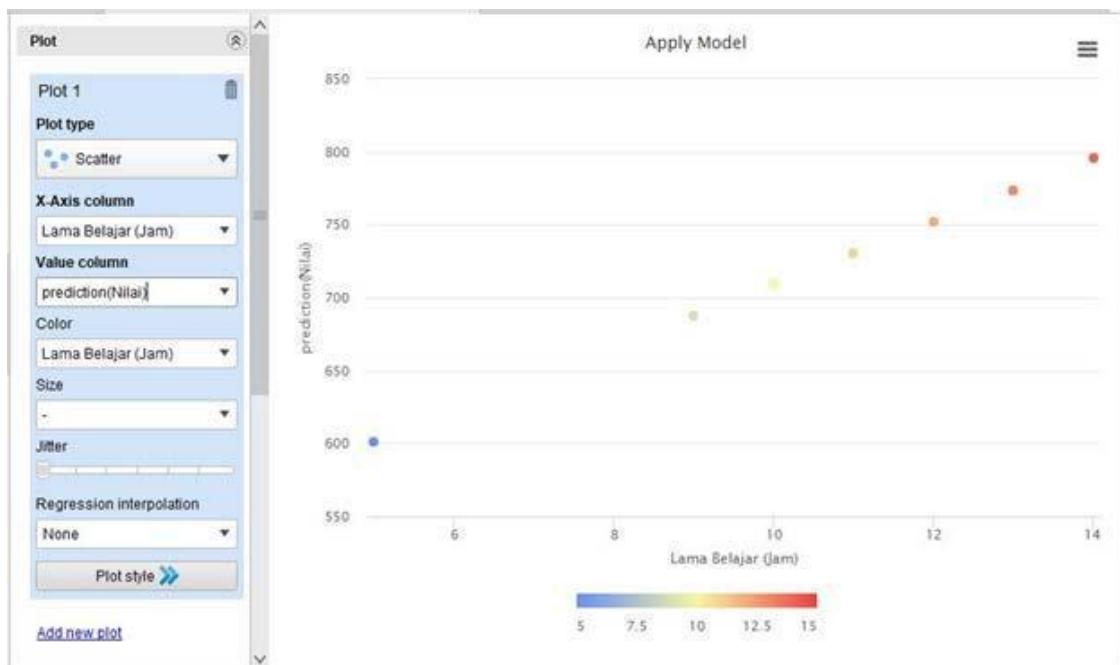


4. Jalankan proses dengan meng-klik tombol Run.

➤ Data View (hasil Prediksi nilai ujian)

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

➤ Chart View (Scatter Plot)



12.4.3 Pembuktian Model Regresi

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

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

Memasukkan nilai variable X_1 ke dalam model regresi :

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

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

12.5 Tugas Kasus

:

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

Hipotesis :

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

1. Buatlah table berikut dengan menggunakan Microsoft Excel.

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

2. Buatlah proses Regresi Linier Sederhana menggunakan RapidMiner dengan ketentuan sebagai berikut.
 - a) Variable bebas (X) = Pendapatan (X_1), Jumlah Anggota Keluarga (X_2)
 - b) Variable terikat (Y) = Daya Beli
 - c) Toleransi yang digunakan = 5%

Pendapatan (X_1)

Import Data - Format your columns.

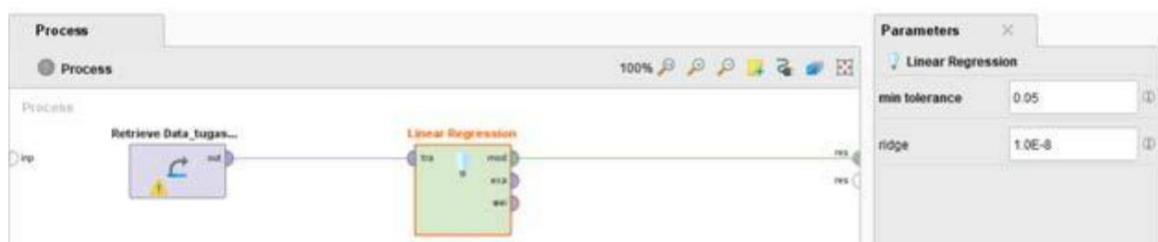
Format your columns.

Replace errors with missing values ⓘ

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

✓ no problems.

← Previous Next → X Cancel



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

LinearRegression

$$Y = 0.788 \times \text{PENDAPATAN (RUPIAH)} + -7467.579$$

Jumlah Anggota Keluarga (X_2)

Import Data - Format your columns.

Format your columns.

Replace errors with missing values ⓘ

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

no problems.

← Previous → Next Cancel



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)	99652.830	347817.155	?	?	0.284	0.781	

LinearRegression

$$151688.679 * \text{JUMLAH ANGGOTA KELUARGA} + 99652.830$$

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

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

$$X_1 \Rightarrow Y = 0,788 X 1.000.000 - 7.463,579 = 780.536,421 X_2 \\ \Rightarrow Y = 151.688,679 X 6 + 98.652,830 = 811.479,244$$

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

4. Tuliskan model persamaan regresi linier sederhana yang terbentuk!

$$X_1 \Rightarrow Y = 0,788 X \text{ Pendapatan (Rupiah)} - 7.463,579 \\ X_2 \Rightarrow Y = 151.688,679 X \text{ Jumlah Anggota Keluarga} + 98.652,830$$

5. Gunakan data testing untuk menjawab perintah berikut :

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

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

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

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

Pendapatan (X_1)

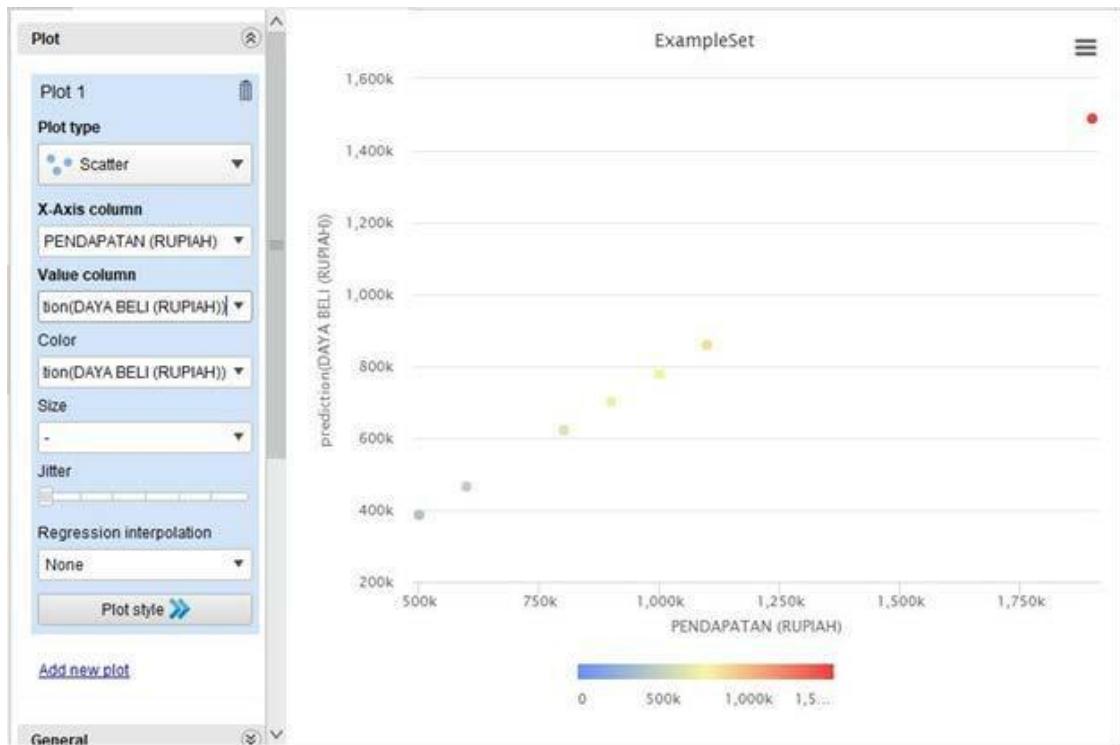
Row No.	NO. RESPON...	prediction(D...)	PENDAPATA...
1	1	701500.958	900000
2	2	622726.676	800000
3	3	386403.830	500000
4	4	1489243.776	1900000
5	5	465178.112	600000
6	6	622726.676	800000
7	7	780275.239	1000000
8	8	859049.521	1100000
9	9	780275.239	1000000
10	10	386403.830	500000

Jumlah Anggota Keluarga (X_2)

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

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

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



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

