

LAPORAN AKHIR MODUL KESELURUHAN

MODUL 1

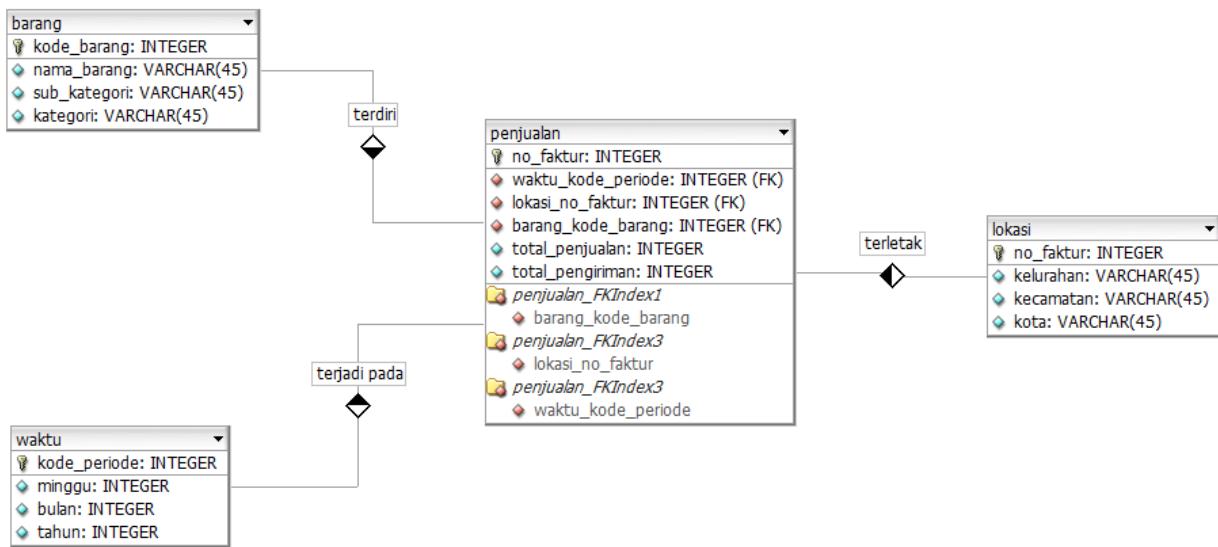
Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

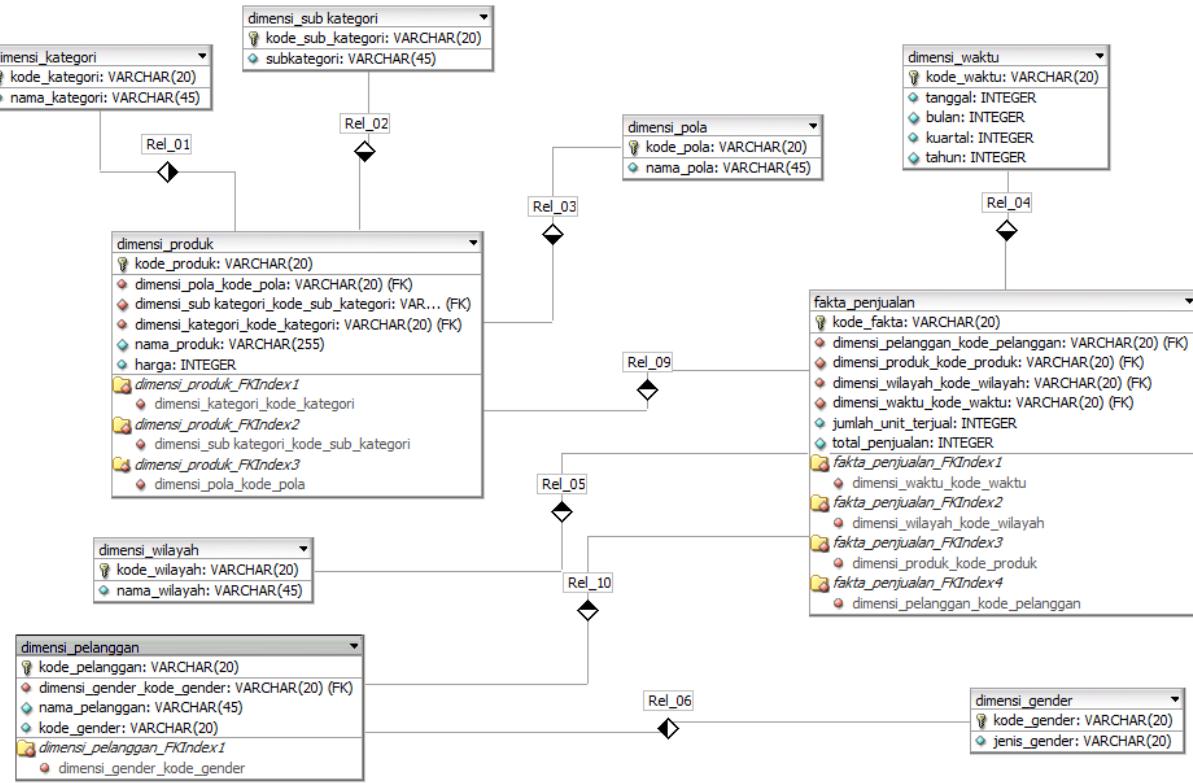
Langkah Langkah Praktikum

“star chema penjualan”



Tugas

“snowflake penjualan”



Nama : Oki Kus Mahesa

NIM : L200170064

Kelas : C

MODUL 5

Langkah-langkah Praktikum

• Kegiatan 1

1. Membuat table data **Fakta Penjualan**

2. Membuat Pivot Table

The screenshot shows a Microsoft Excel window with the title "Fakta_Penjualan - Excel". The ribbon tabs are visible at the top. A "Create PivotTable" dialog box is open in the foreground, centered over the worksheet area. The dialog box has three main sections: "Choose the data that you want to analyze", "Choose where you want the PivotTable report to be placed", and "Choose whether you want to analyze multiple tables". The "Table/Range" field in the first section is set to "Sheet1!\$A\$1:\$K\$21". The "OK" button is highlighted in blue.

3. Menampilkan data tahun, nama_subkategori, dan jumlah.

The screenshot shows a Microsoft Excel window with the title "Fakta_Penjualan - Excel". The ribbon tabs are visible at the top, and the "PivotTable Tools" tab is selected. The "Analyze" tab is active. On the left, the "PivotTable Fields" pane is open, showing a list of fields: bulan, kuartal, tahun, nama_produk, and nama_subkategori. The "tahun" field is checked. In the center, a PivotTable is displayed with data grouped by year (2010, 2011, 2012) and subcategory (Bahan, Batik, Bolero, Celana, Hem, Jam, Jarik, kaos, Rok, Sarimbit, Standar). The "Grand Total" row shows the sum of the values for each year.

- Kegiatan 2**

1. Menambahkan field jumlah ke kotak value

The screenshot shows a Microsoft Excel spreadsheet titled "Fakta_Penjualan - Excel". The PivotTable Fields pane on the right indicates that the "jumlah" field has been added to the "Values" section. The main table area displays sales data for various items (Bahan, Batik, Bolero, Celana, Hem, Jam, Jarik, kaos, Rok, Sarimbit, Standar) across three years (2010, 2011, 2012). The table includes columns for the sum of sales and the count of sales.

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

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

The screenshot shows the "Value Field Settings" dialog box open in Microsoft Excel. The "Source Name" is set to "jumlah" and the "Custom Name" is "Count of jumlah2". The "Summarize Values By" option is set to "Show Values As". The "Summarize value field by" dropdown is set to "Count". The main table area displays sales data for various items (Bahan, Batik, Bolero, Celana, Hem, Jam, Jarik, kaos, Rok, Sarimbit, Standar) across three years (2010, 2011, 2012). The table includes columns for the sum of sales and the count of sales.

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

- Kegiatan 3

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

The screenshot shows a Microsoft Excel window with a PivotTable named 'PivotTable1'. The PivotTable has 'Bahan' in Row Labels, '2010' and '2011' in Columns, and 'jumlah' in Values. A calculated field 'Pendapatan' is being created with the formula '= jumlah*harga'. The PivotTable Fields pane on the right shows fields like 'jumlah' and 'harga' selected. Below the PivotTable is a report table with columns: 'Count of jumlah2', 'Sum of Pendapatan', 'Total Sum of jumlah', 'Total Count of jumlah2', and 'Total Sum of Pendapatan'. The data in the report table is as follows:

| Count of jumlah2 | Sum of Pendapatan | Total Sum of jumlah | Total Count of jumlah2 | Total Sum of Pendapatan |
|------------------|-------------------|---------------------|------------------------|-------------------------|
| 2 | 2120000 | 17 | 4 | 15045000 |
| 1 | 150000 | 1 | 1 | 150000 |
| | 0 | 1 | 1 | 255000 |
| 1 | 935000 | 34 | 2 | 3740000 |
| 2 | 1596000 | 17 | 5 | 19023000 |
| 1 | 3520000 | 44 | 1 | 3520000 |
| 1 | 160000 | 4 | 1 | 160000 |
| 1 | 420000 | 15 | 2 | 1350000 |
| 1 | 225000 | 1 | 1 | 225000 |
| | 0 | 1 | 1 | 150000 |
| | 0 | 2 | 1 | 450000 |
| 10 | 115692000 | 137 | 20 | 456073000 |

2. Setelah meng-klik tombol '**OK**' maka akan muncul field baru dengan nama

'Pendapatan', cukup centang field tersebut dan data akan otomatis keluar.

| | 2010 | | 2011 | | 2012 | | |
|--------------------|---------------|-----------------|-------------------|---------------|-----------------|-------------------|---------------|
| Row Labels | Sum of jumlah | Count of jumlah | Sum of Pendapatan | Sum of jumlah | Count of jumlah | 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 |

• Kegiatan 4

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

The screenshot shows a Microsoft Excel spreadsheet titled "Fakta_Penjualan - Excel". The PivotTable Tools ribbon tab is selected. The PivotTable Fields pane on the right side of the screen lists fields: "tahun" under Columns, "nama_subkat..." under Rows, and "Sum of Pendapatan" under Values. The main data area shows a table with columns for "tahun" (2010, 2011, 2012, Grand Total) and rows for various subcategories. The data includes values such as 500000 for Bahan in 2010, 960000 for Bahan in 2011, and so on.

| | 2010 | 2011 | 2012 | Grand Total |
|--------------------|-----------------|-----------------|------------------|------------------|
| Bahan | 500000 | 960000 | 2120000 | 15045000 |
| Batik | 0 | 0 | 150000 | 150000 |
| Bolero | 0 | 255000 | 0 | 255000 |
| Celana | 935000 | 0 | 935000 | 3740000 |
| Hem | 500000 | 4960000 | 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 |

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

Roll Up

Screenshot of Microsoft Excel showing a PivotTable named "Fakta_Penjualan" with the following data:

| | | 2010 | 2011 | 2012 | Grand Total |
|-------------------|---------------|----------|-----------|-----------|-------------|
| Sum of Pendapatan | Column Labels | | | | |
| | | + 2010 | + 2011 | + 2012 | Grand Total |
| 6 Row Labels | | | | | |
| 7 Bahan | 500000 | 960000 | 2120000 | 15045000 | |
| 8 Batik | 0 | 0 | 150000 | 150000 | |
| 9 Bolero | 0 | 255000 | 0 | 255000 | |
| 10 Celana | 935000 | 0 | 935000 | 3740000 | |
| 11 Hem | 500000 | 4960000 | 1596000 | 19023000 | |
| 12 Jam | 0 | 0 | 3520000 | 3520000 | |
| 13 Jarik | 0 | 0 | 160000 | 160000 | |
| 14 kaos | 0 | 60000 | 420000 | 1350000 | |
| 15 Rok | 0 | 0 | 225000 | 225000 | |
| 16 Sarimbit | 0 | 150000 | 0 | 150000 | |
| 17 Standar | 0 | 450000 | 0 | 450000 | |
| 18 Grand Total | 15065000 | 30030000 | 115692000 | 456073000 | |

The PivotTable Fields pane on the right shows fields: bulan, kuartal, tahun, nama_produk, and nama_subkategori.

Drop Down

Fakta_Perjualan - Excel

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

Cut Copy Format Painter Clipboard

Font Alignment Number Styles Cells Editing

A36 : Jam

PivotTable Tools

PivotTable Fields

Choose fields to add to report:

Search

bulan
 kuartal
 tahun
 nama_produk
 nama_subkатегори

Drag fields between areas below:

Filters Columns tahun

Rows Values Sum of Penda...

Defer Layout Update

100%

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|-----------------------------|-----------------|-----------------|------------------|------------------|---|---|---|---|---|---|---|---|---|
| 49 | Ibu Siti Arya | 0 | 0 | 225000 | 225000 | | | | | | | | | |
| 50 | └ Sarimbit | 0 | 150000 | 0 | 150000 | | | | | | | | | |
| 51 | └ Kaos Katun Print Bola | 0 | 150000 | 0 | 150000 | | | | | | | | | |
| 52 | └ Ibu Hatamah | 0 | 150000 | 0 | 150000 | | | | | | | | | |
| 53 | └ Standar | 0 | 450000 | 0 | 450000 | | | | | | | | | |
| 54 | └ Jarik Standar Print Sogan | 0 | 450000 | 0 | 450000 | | | | | | | | | |
| 55 | └ Bapak Ketut | 0 | 450000 | 0 | 450000 | | | | | | | | | |
| 56 | Grand Total | 15065000 | 30030000 | 115692000 | 456073000 | | | | | | | | | |
| 57 | | | | | | | | | | | | | | |
| 58 | | | | | | | | | | | | | | |
| 59 | | | | | | | | | | | | | | |
| 60 | | | | | | | | | | | | | | |
| 61 | | | | | | | | | | | | | | |
| 62 | | | | | | | | | | | | | | |
| 63 | | | | | | | | | | | | | | |
| 64 | | | | | | | | | | | | | | |
| 65 | | | | | | | | | | | | | | |
| 66 | | | | | | | | | | | | | | |
| 67 | | | | | | | | | | | | | | |
| 68 | | | | | | | | | | | | | | |
| 69 | | | | | | | | | | | | | | |
| 70 | | | | | | | | | | | | | | |
| 71 | | | | | | | | | | | | | | |

Sheet2 Sheet1 +

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.

| | Column Labels | | | | | | | |
|--------------------|-------------------|------------------|--------------------------|---------------|----------------|-------------------|------------------|--|
| | 2010 | | | 2011 | | | | |
| 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

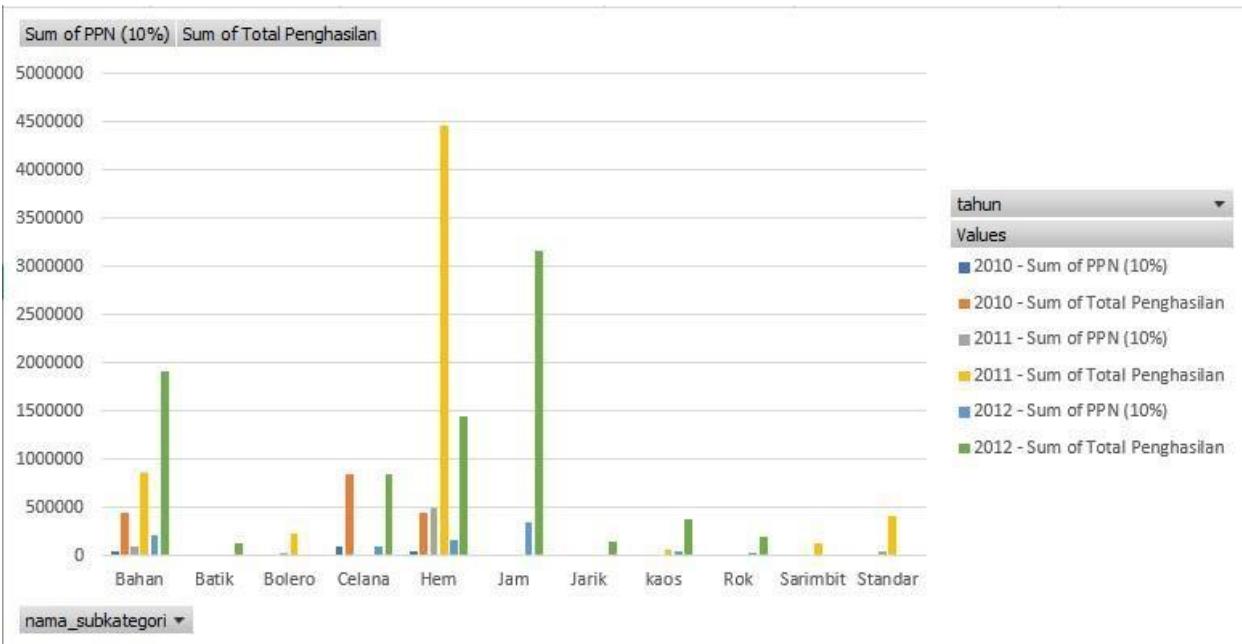
| | | | | |
|------------------|-----------------|------------------|------------|-----------|
| 15045000 | 1504500 | 13540500 | 17 | 7 |
| 150000 | 15000 | 135000 | 1 | 3 |
| 255000 | 25500 | 229500 | 1 | 4 |
| 3740000 | 374000 | 3366000 | 34 | 2 |
| 19023000 | 1902300 | 17120700 | 17 | 14 |
| 3520000 | 352000 | 3168000 | 44 | 1 |
| 160000 | 16000 | 144000 | 4 | 2 |
| 1350000 | 135000 | 1215000 | 15 | 2 |
| 225000 | 22500 | 202500 | 1 | 4 |
| 150000 | 15000 | 135000 | 1 | 4 |
| 450000 | 45000 | 405000 | 2 | 4 |
| 456073000 | 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

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

PivotChart



NAMA : OKI KUS MAHESA

NIM : L200170064

KELAS : C

Modul 6

Pengenalan Data Mining

6.5 Tugas

1. Dengan Menggunakan Ms.Excel, buatlah tabel berikut.

| | A | B | C | D | E | F | |
|----|-------------|--------|--------------|------------|---------|------------|--|
| 1 | Jurusan_SMA | Gender | Asal_Sekolah | Rerata_SKS | Asisten | Lama_Studi | |
| 2 | IPS | WANITA | SURAKARTA | 18 | TIDAK | TERLAMBAT | |
| 3 | IPA | PRIA | SURAKARTA | 19 | YA | TEPAT | |
| 4 | LAIN | PRIA | SURAKARTA | 19 | TIDAK | TERLAMBAT | |
| 5 | IPA | PRIA | LUAR | 17 | TIDAK | TERLAMBAT | |
| 6 | IPA | WANITA | SURAKARTA | 17 | TIDAK | TEPAT | |
| 7 | IPA | WANITA | LUAR | 18 | YA | TEPAT | |
| 8 | IPA | PRIA | SURAKARTA | 18 | TIDAK | TERLAMBAT | |
| 9 | IPA | PRIA | SURAKARTA | 19 | TIDAK | TEPAT | |
| 10 | IPS | PRIA | LUAR | 18 | TIDAK | TERLAMBAT | |
| 11 | LAIN | WANITA | SURAKARTA | 18 | TIDAK | TEPAT | |
| 12 | IPA | WANITA | SURAKARTA | 19 | TIDAK | TEPAT | |
| 13 | IPS | PRIA | SURAKARTA | 20 | TIDAK | TEPAT | |
| 14 | IPS | PRIA | SURAKARTA | 19 | TIDAK | TEPAT | |
| 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 | |
| 22 | | | | | | | |

2a.) Dengan menggunakan formula dalam Ms.Excel, carilah masing masing kelas IPA, IPS dan LAIN

- Jumlah Data Kelas IPA

| | | | =COUNTIF(A2:A21,"IPA") |
|-------|---------|------------|------------------------|
| E | F | G | H |
| SKS | Asisten | Lama_Studi | Jumlah Data IPA |
| TIDAK | | TERLAMBAT | |
| YA | | TEPAT | 9 |

- Jumlah Data Kelas IPS

| | | | =COUNTIF(A2:A21,"IPS") |
|-------|---------|------------|------------------------|
| E | F | G | I |
| S | Asisten | Lama_Studi | Jumlah Data IPS |
| TIDAK | | TERLAMBAT | |
| YA | | TEPAT | 7 |

- Jumlah Data kelas LAIN

| | | | =COUNTIF(A2:A21,"LAIN") |
|-------|---------|------------|-------------------------|
| E | F | G | J |
| S | Asisten | Lama_Studi | Jumlah Data Lain |
| TIDAK | | TERLAMBAT | |
| YA | | TEPAT | 4 |

2b.) Pada atribut Lama Studi, berapa jumlah data masing-masing kelas TEPAT, TERLAMBAT

- Jumlah data Lama Studi TEPAT

| |
|--------------------------|
| =COUNTIF(F2:F21,"TEPAT") |
| Lama Studi Tepat 13 |

- Jumlah data Lama Studi TERLAMBAT

| |
|------------------------------|
| =COUNTIF(F2:F21,"TERLAMBAT") |
| Lama Studi Terlambat 7 |

3c.) Pada atribut Rearata SKS, berapa nilai MAX, MIN, MEAN, dan Standar Deviasi

- Rearata SKS nilai Max

| f _x | =MAX(D2:D21) |
|----------------|-----------------------|
| D | E F G H |
| | Rearata SKS MAX 23 |

- Rerata SKS nilai Min

| f _x | =MIN(D2:D21) |
|----------------|----------------------|
| D E F G H I | K R |
| | Rerata SKS Min 16 |

- Rerata SKS nilai Mean

| f _x | =AVERAGE(D2:D21) |
|----------------|--------------------------|
| D E F G H I J | K S |
| | Rerata SKS Mean 18.95 |

- Rerata SKS nilai Standar Deviasi

A screenshot of a Microsoft Excel spreadsheet. The formula `=STDEV(D2:D21)` is entered in cell I1. The result, `1.66938375`, is displayed in the cell. The cell is highlighted with a green border. The formula bar at the top shows the formula. The columns are labeled D, E, F, G, H, I, J, K, L.

2d.) Pada tabel tersebut, berapakah jumlah data gabungan untuk kelas pada atribut Jurusan SMA = IPA, Gender = Pria, Asisten = Ya, Lama Studi = Tepat

- IPA, PRIA, YA, TEPAT

A screenshot of a Microsoft Excel spreadsheet. The formula `=COUNTIFS(A2:A21,"IPA";B2:B21,"PRIA";E2:E21,"YA";F2:F21,"TEPAT")` is entered in cell H11. The result, `3`, is displayed in the cell. The cell is highlighted with a green border. The formula bar at the top shows the formula. The columns are labeled A, B, C, D, E, F, G, H.

Nama : Oki Kus Mahesa

Nim : L200170064

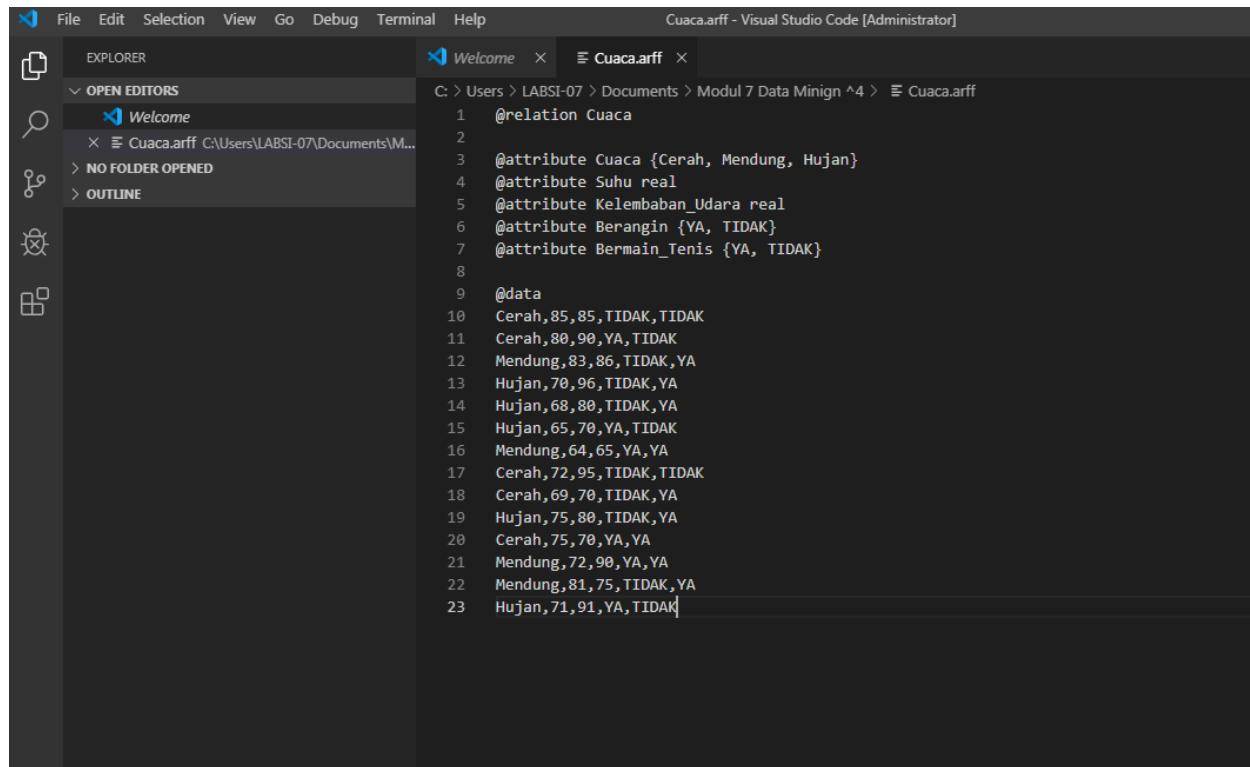
Kelas : C

Modul 7

Data Pre Processing

- Langkah – langkah Praktikum :

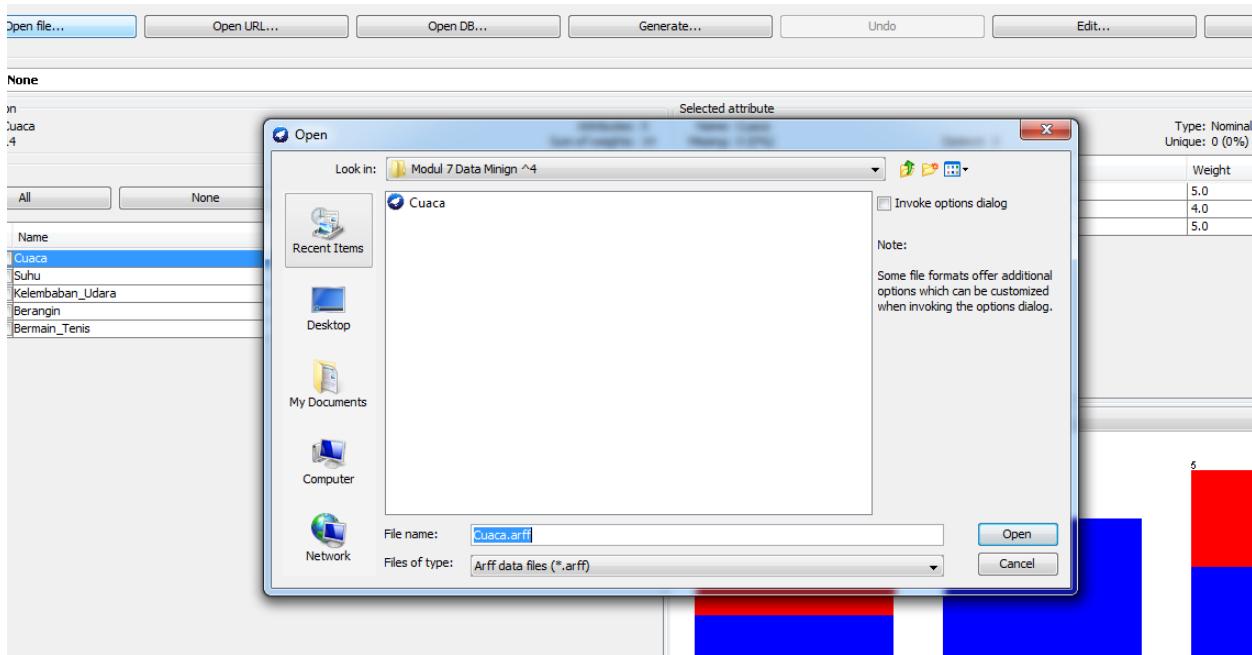
1. Membuat Syntax weka untuk di export ke weka Table untuk Cuaca



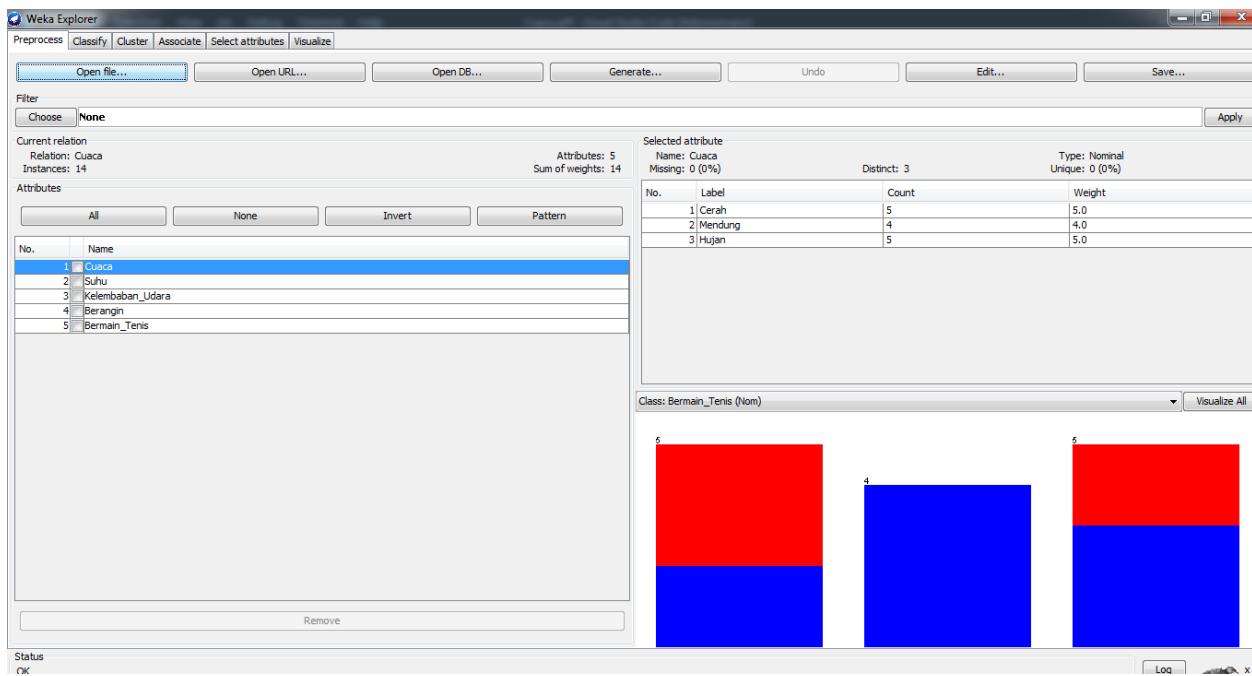
The screenshot shows a Visual Studio Code interface with a dark theme. The top menu bar includes File, Edit, Selection, View, Go, Debug, Terminal, and Help. A tab bar at the top right shows 'Cuaca.arff - Visual Studio Code [Administrator]'. The left sidebar has icons for Explorer, Search, and Outline, with 'OPEN EDITORS' expanded to show 'Welcome' and 'Cuaca.arff'. The main editor area displays the following ARFF code:

```
C: > Users > LABSI-07 > Documents > Modul 7 Data Minign ^4 > 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,78,96,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|
```

2. Open() file Cuaca yang di ketik ke weka



3. Data Di dari table Cuaca di sajikan di weka software



4. Informasi Atribut (X1) Cuaca

Selected attribute

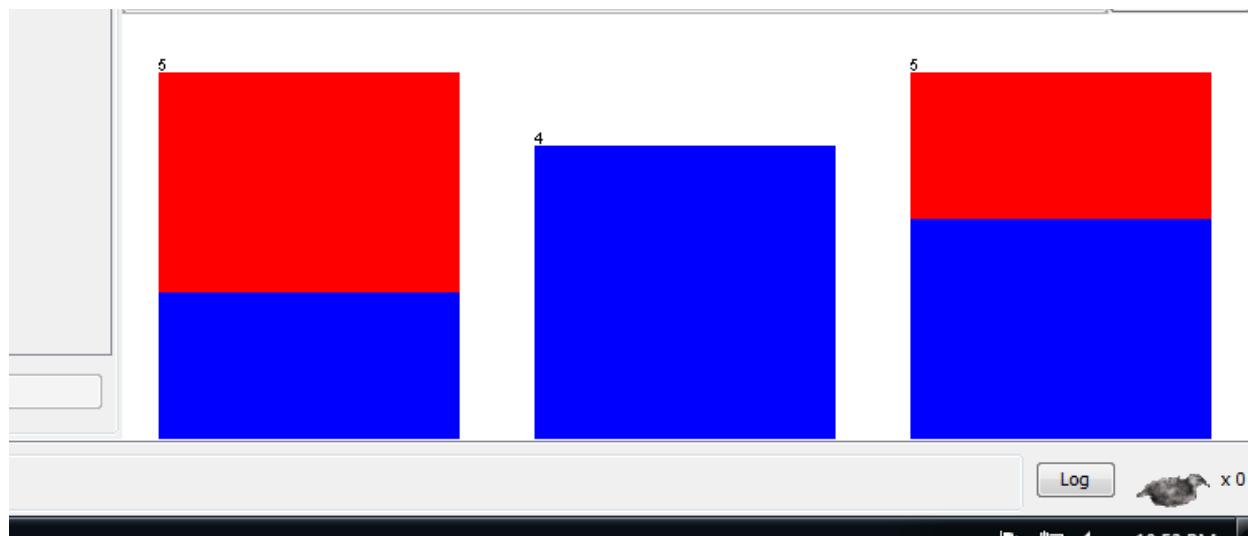
| | | | |
|-------------------------------------|--------------------------------|-------------|---------------------------------|
| Attributes: 5 Sum of weights: 14 | Name: Cuaca Missing: 0 (0%) | Distinct: 3 | Type: Nominal Unique: 0 (0%) |
|-------------------------------------|--------------------------------|-------------|---------------------------------|

Pattern

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

Apply

- Grafik Atribut Cuaca



5. Informasi Atribut (X2) Suhu

Attributes: 5
Sum of weights: 14

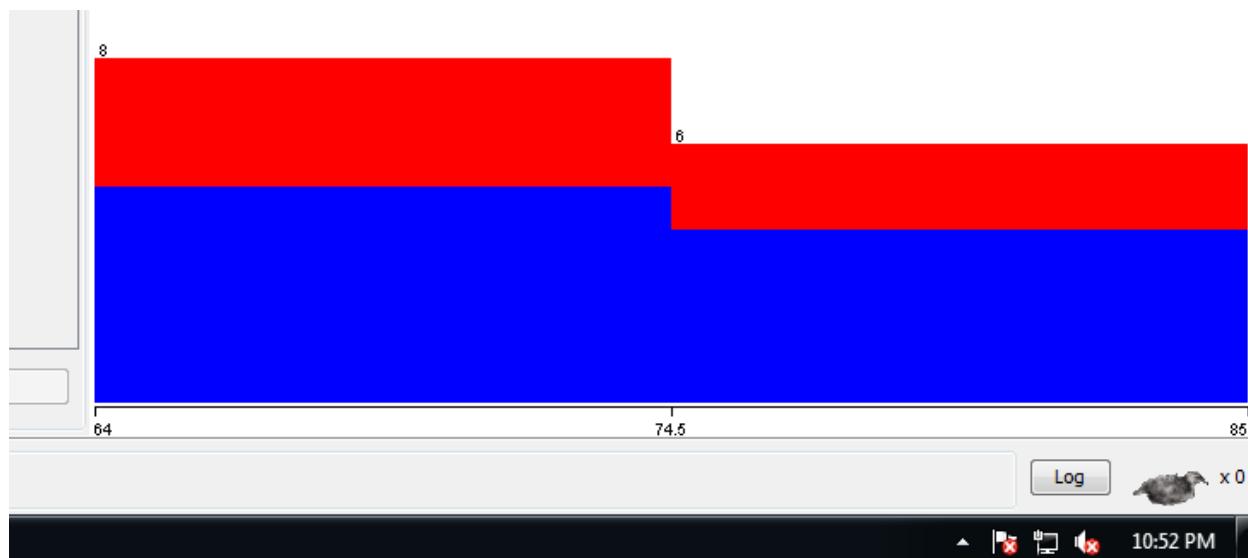
Selected attribute
Name: Suhu
Missing: 0 (0%)
Distinct: 12
Type: Numeric
Unique: 10 (71%)

| Statistic | Value |
|-----------|--------|
| Minimum | 64 |
| Maximum | 85 |
| Mean | 73.571 |
| StdDev | 6.572 |

Class: Bermain_Tenis (Nom)

Apply

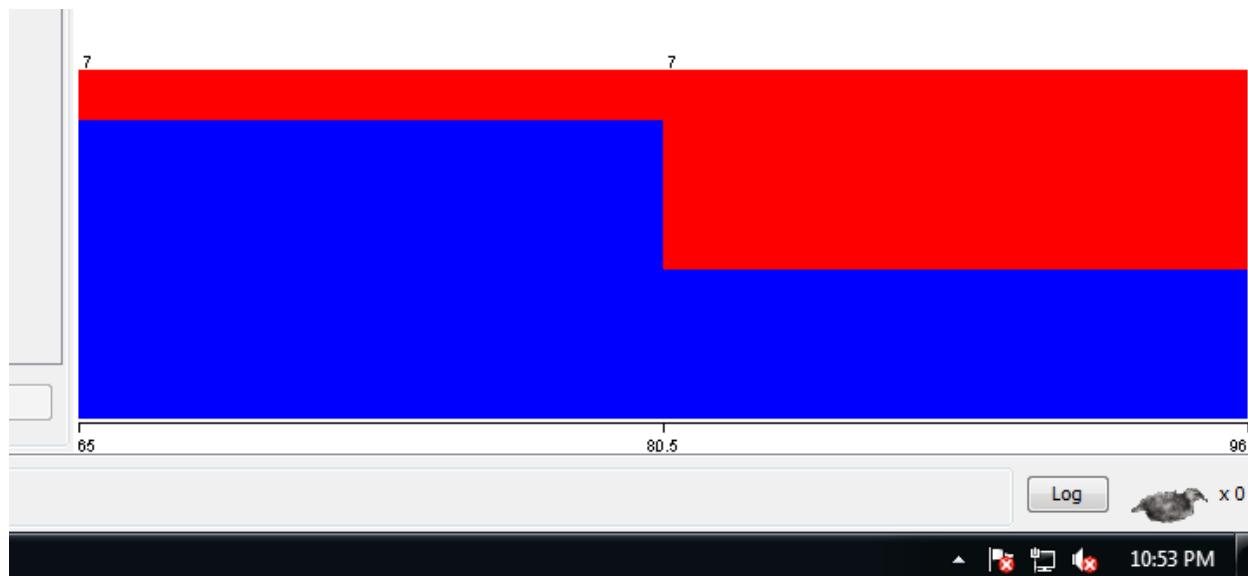
- Grafik Atribut Suhu



6. Informasi Atribut (X3) Kelembaban_Udara

| Selected attribute | | Type: Numeric |
|--------------------|------------------------|-----------------|
| Attributes: 5 | Name: Kelembaban_Udara | Distinct: 10 |
| Sum of weights: 14 | Missing: 0 (0%) | Unique: 7 (50%) |
| Statistic | | Value |
| Minimum | 65 | |
| Maximum | 96 | |
| Mean | 81.643 | |
| StdDev | 10.285 | |

- Grafik Atribut Kelembaban_Udara



7. Informasi Atribut (X5) Berangin

General... Options Edit... Save...

Selected attribute
Name: Berangin
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

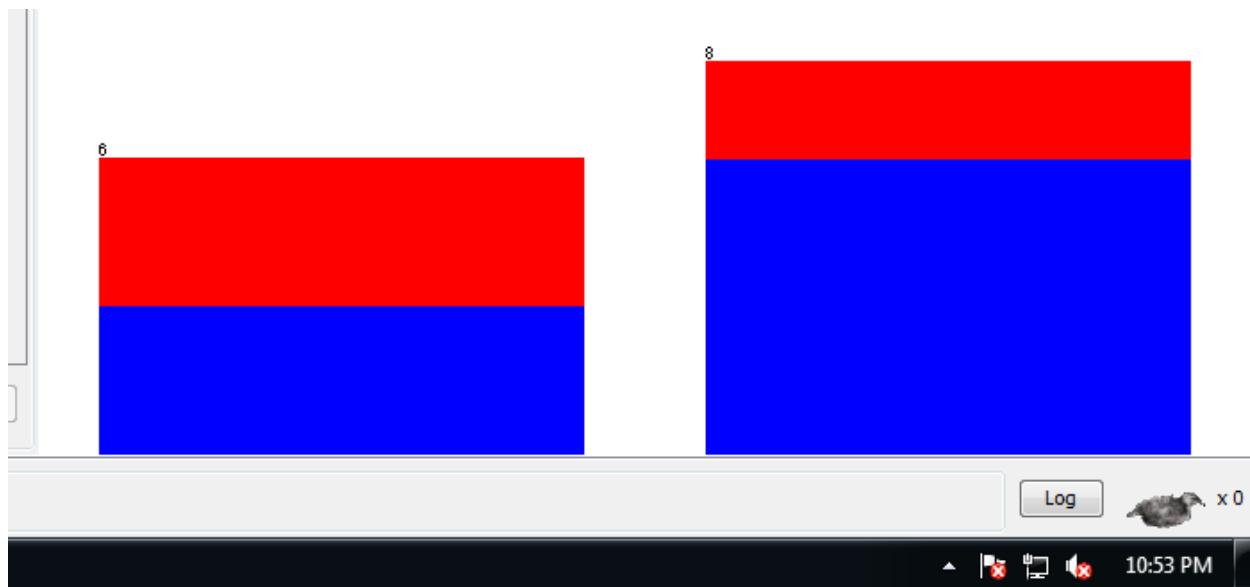
Attributes: 5
Sum of weights: 14

| No. | Label | Count | Weight |
|-----|-------|-------|--------|
| 1 | YA | 6 | 6.0 |
| 2 | TIDAK | 8 | 8.0 |

Pattern

Class: Bermain_Tenis (Nom) Visualize All

- Grafik Atribut Berangin



8. Infromasi Atribut (Y) Bermain_Tenis

Attributes: 5
Sum of weights: 14

Selected attribute
Name: Bermain_Tenis
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

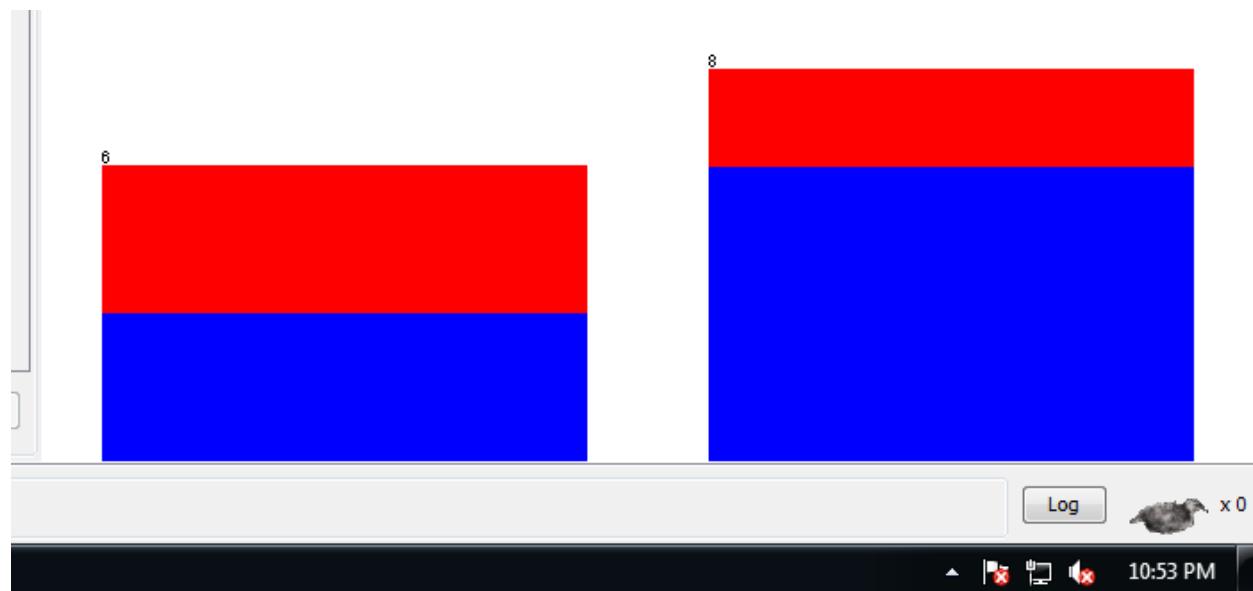
| No. | Label | Count | Weight |
|-----|-------|-------|--------|
| 1 | YA | 9 | 9.0 |
| 2 | TIDAK | 5 | 5.0 |

Pattern

Class: Bermain_Tenis (Nom) Visualize All

9 

- Grafik Atribut (Y) Bermain_Tenis



Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

Modul 7

Data Pre Processing

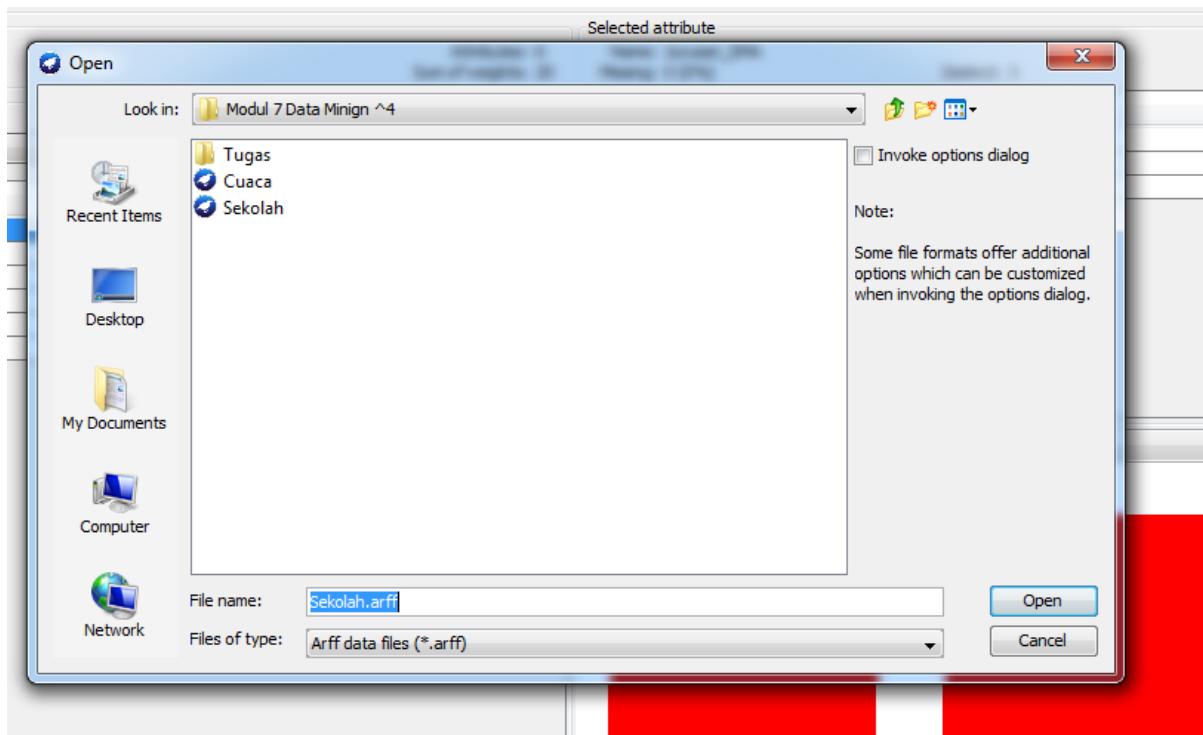
Tugas :

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

```
C: > Users > LABSI-07 > Documents > Modul 7 Data Minign ^4 > Sekolah.arff
1 @relation Sekolah
2
3 @attribute Jurusan_SMA {IPS, IPA, LAIN}
4 @attribute Gender {WANITA, PRIA}
5 @attribute Asal_Sekolah {SURAKARTA, LUAR}
6 @attribute Rerata_SKS real
7 @attribute Asisten {TIDAK, YA}
8 @attribute Lama_Studi {TERLAMBAT, TEPAT}
9
10 @data
11 IPS, WANITA, SURAKARTA, 18, TIDAK, TERLAMBAT
12 IPA, PRIA, SURAKARTA, 19, YA, TEPAT
13 LAIN, PRIA, SURAKARTA, 19, TIDAK, TERLAMBAT
14 IPA, PRIA, LUAR, 17, TIDAK, TERLAMBAT
15 IPA, WANITA, SURAKARTA, 17, TIDAK, TEPAT
16 IPA, WANITA, LUAR, 18, YA, TEPAT
17 IPA, PRIA, SURAKARTA, 18, TIDAK, TERLAMBAT
18 IPA, PRIA, SURAKARTA, 19, TIDAK, TEPAT
19 IPS, PRIA, LUAR, 18, TIDAK, TERLAMBAT
20 LAIN, WANITA, SURAKARTA, 18, TIDAK, TEPAT
21 IPA, WANITA, SURAKARTA, 19, TIDAK, TEPAT
22 IPS, PRIA, SURAKARTA, 20, TIDAK, TEPAT
23 IPS, PRIA, SURAKARTA, 19, TIDAK, TEPAT
24 IPA, PRIA, SURAKARTA, 19, TIDAK, TEPAT
25 IPA, PRIA, LUAR, 22, YA, TEPAT
26 LAIN, PRIA, SURAKARTA, 16, TIDAK, TERLAMBAT
27 IPS, PRIA, LUAR, 20, TIDAK, TEPAT
28 LAIN, PRIA, LUAR, 23, YA, TEPAT
29 IPA, PRIA, SURAKARTA, 21, YA, TEPAT
30 IPS, PRIA, SURAKARTA, 19, TIDAK, TERLAMBAT
31
```

2. Perlihatkan file ARFF dan juga gambar grafik untuk setiap data yang ditampilkan dalam Weka berdasarkan file ARFF!

- **Membuka Data yang telah di buat di microsoft visual dengan menggunakan data tabel Sekolah pada tugas 1 di modul 6**



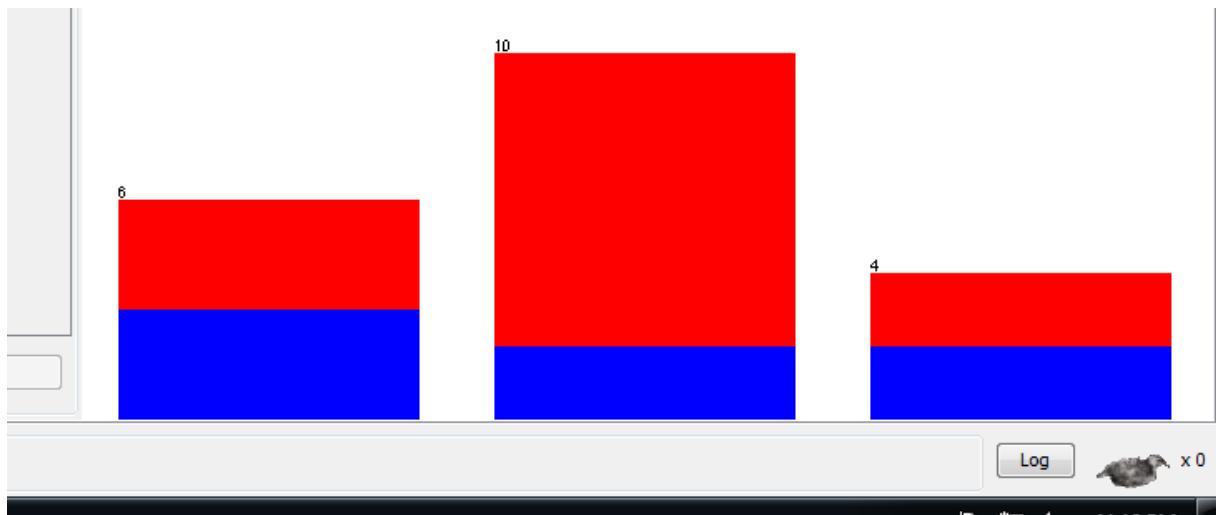
- **Informasi Berupa (LABEL yang Berisi = IPS, IPA, LAIN) dengan menampilkan Count Data dan Weight Data dari Attribut Jurusan_SMA**

Attributes: 6
Sum of weights: 20

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

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

- Grafik Attribut Jurusan_SMA

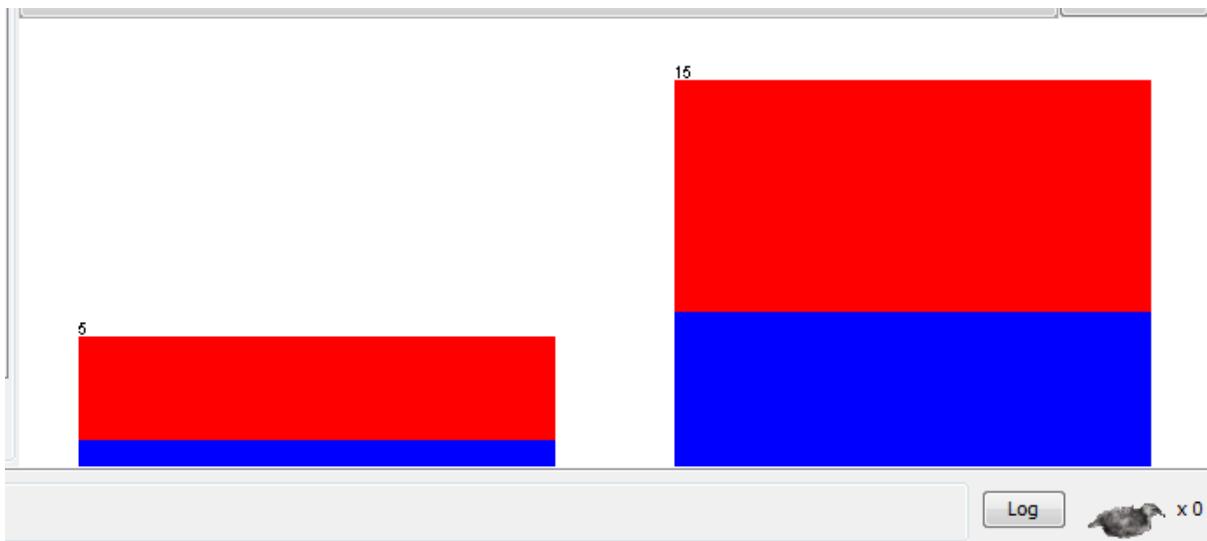


- Informasi Berupa (LABEL yang Berisi = WANITA, PRIA) dengan menampilkan Count Data dan Weight Data dari Attribut Gender

Selected attribute

| | | | | |
|-------------------------------------|---------------------------------|-------------|---------------------------------|--------|
| Attributes: 6 Sum of weights: 20 | Name: Gender Missing: 0 (0%) | Distinct: 2 | Type: Nominal Unique: 0 (0%) | |
| Pattern | Apply | | | |
| | No. | Label | Count | Weight |
| | 1 | WANITA | 5 | 5.0 |
| | 2 | PRIA | 15 | 15.0 |

- Grafik Attribut Gender



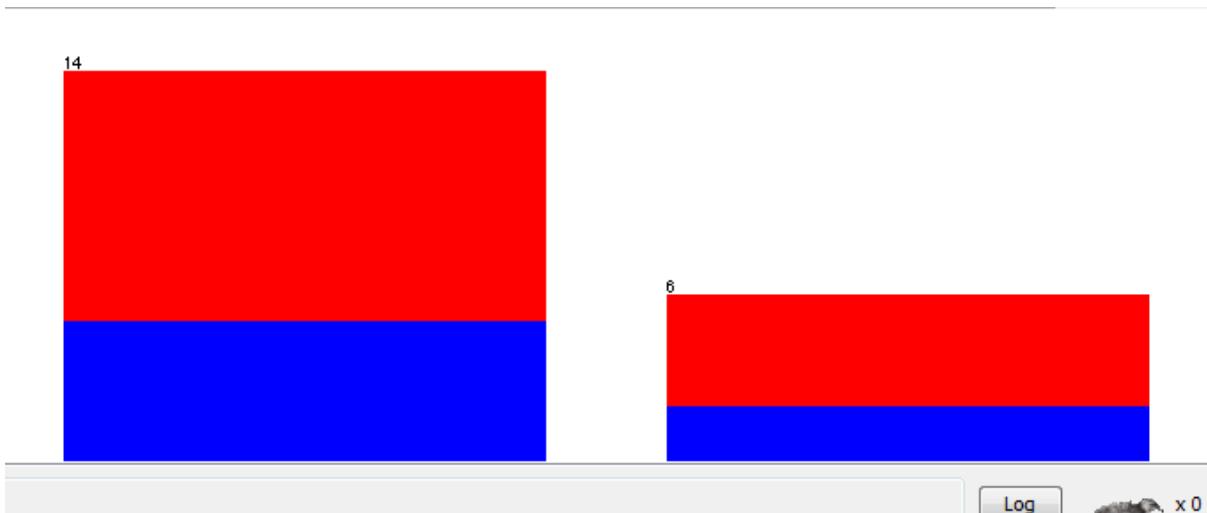
- Informasi Berupa (LABEL yang Berisi = SURAKARTA, LAIN) dengan menampilkan Count Data dan Weight Data dari Attribut Asal_Sekolah

Attributes: 6
Sum of weights: 20

Selected attribute
Name: Asal_Sekolah
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

| No. | Label | Count | Weight |
|-----|-----------|-------|--------|
| 1 | SURAKARTA | 14 | 14.0 |
| 2 | LUAR | 6 | 6.0 |

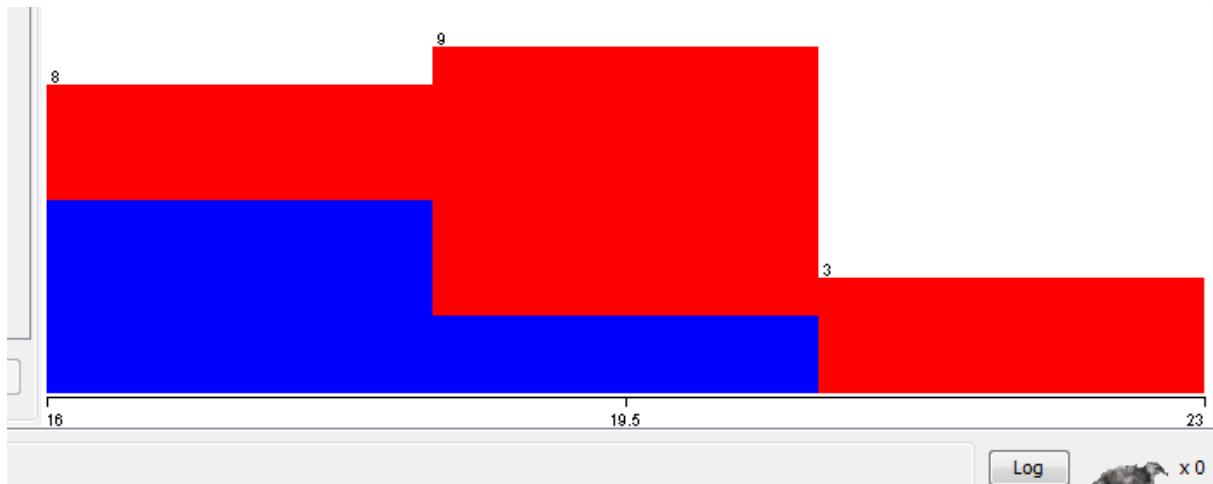
- Grafik Attribut Asal_Sekolah



- Informasi Berupa (LABEL yang Berisi nilai = *real* dengan menampilkan Count Data dan Weight Data dari Attribut Rerata_SKS

| | | Selected attribute | | Apply |
|-----------------|----|--------------------|---------------|-----------------|
| | | Name: Rerata_SKS | Type: Numeric | |
| | | Missing: 0 (0%) | Distinct: 8 | Unique: 4 (20%) |
| Attributes: | 6 | Statistic | Value | |
| Sum of weights: | 20 | Minimum | 16 | |
| | | Maximum | 23 | |
| | | Mean | 18.95 | |
| | | StdDev | 1.669 | |

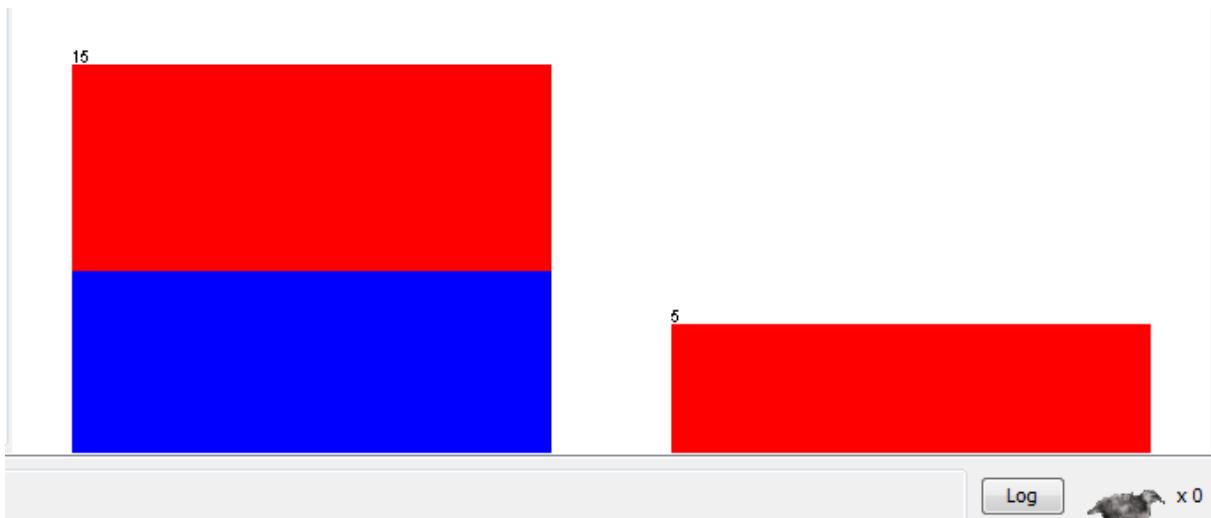
- Grafik Attribut Rerata_SKS



- **Informasi Berupa (LABEL yang Berisi = TIDAK, YA) dengan menampilkan Count Data dan Weight Data dari Attribut Asisten**

| Selected attribute | | | |
|-------------------------------------|----------------------------------|-------------|---------------------------------|
| Attributes: 6 Sum of weights: 20 | Name: Asisten Missing: 0 (0%) | Distinct: 2 | Type: Nominal Unique: 0 (0%) |
| No. | Label | Count | Weight |
| 1 | TIDAK | 15 | 15.0 |
| 2 | YA | 5 | 5.0 |

- **Grafik Attribut Asisten**

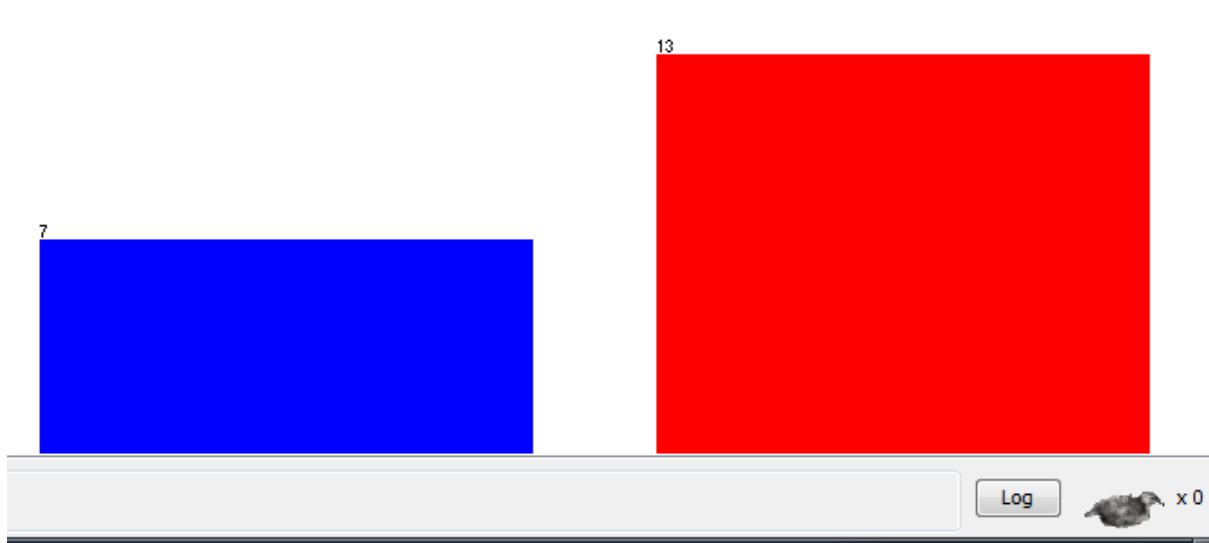


- **Informasi Berupa (LABEL yang Berisi = TERLAMBAT, TEPAT) dengan menampilkan Count Data dan Weight Data dari Attribut Lama_Studi**

Selected attribute

| Attributes: 6 Sum of weights: 20 | Name: Lama_Studi Missing: 0 (0%) | Distinct: 2 | Type: Nominal Unique: 0 (0%) | | | | | | | | | | | | |
|-------------------------------------|---|-------------|---------------------------------|-----|-------|-------|--------|---|-----------|---|-----|---|-------|----|------|
| Pattern | <table border="1"> <thead> <tr> <th>No.</th> <th>Label</th> <th>Count</th> <th>Weight</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>TERLAMBAT</td> <td>7</td> <td>7.0</td> </tr> <tr> <td>2</td> <td>TEPAT</td> <td>13</td> <td>13.0</td> </tr> </tbody> </table> | | | No. | Label | Count | Weight | 1 | TERLAMBAT | 7 | 7.0 | 2 | TEPAT | 13 | 13.0 |
| No. | Label | Count | Weight | | | | | | | | | | | | |
| 1 | TERLAMBAT | 7 | 7.0 | | | | | | | | | | | | |
| 2 | TEPAT | 13 | 13.0 | | | | | | | | | | | | |

- **Grafik Attribut Lama_Studi**



4. Berapa jumlah attribut yang bertipe binomial dan polynomial?

- **Jumlah attribut yang bertipe binomial :**
Jumlah attribut ada 4 dari keseluruhan 6 attribut
Yaitu :
 1. Gender
 2. Asal_Sekolah
 3. Asisten
 4. Lama_Studi
- **Jumlah attribut yang bertipe polynomial :**
Jumlah attribut ada 2 dari keseluruhan 6 attribut
Yaitu :
 1. Jurusan
 2. Rerata_SKS

5. Berapa jumlah attribut yang bertipe *real* ?

- **Jumlah attribut yang bertipe real :**
Jumlah attribut ada 1 dari keseluruhan 6 attribut
Yaitu :
 1. Rerata_SKS

6. Pada attribut Rerata_SKS , berapakah besarnya nilai Maximum, Minimum, Mean, Standar Deviasi ?

Selected attribute

Name: Rerata_SKS
Missing: 0 (0%)

Distinct: 8

Type: Numeric
Unique: 4 (20%)

| Statistic | Value |
|-----------|-------|
| Minimum | 16 |
| Maximum | 23 |
| Mean | 18.95 |
| StdDev | 1.669 |

Maximum = 16

Minimum = 16

Mean = 18.95

StdDev = 1.669

Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

MODUL 8

KLASIFIKASI : NAÏVE BAYES

- **Langkah-langkah Praktikum**

- **Implementasi Naïve Bayes dengan Weka**

1. Persiapkan file Cuaca.arff , yang akan di gunakan sebagai Data Training

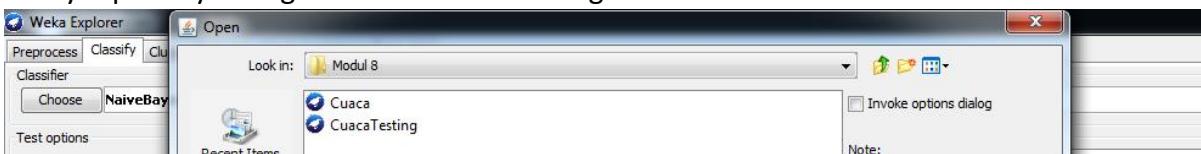
Tabel_Cuaca - Excel

| | A | B | C | D | E | F | G | H | I |
|----|---------|------|------------------|----------|---------------|---|---|---|---|
| 1 | Cuaca | Suhu | Kelembaban_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 | Tidak | Tidak | | | | |
| 16 | | | | | | | | | |
| 17 | | | | | | | | | |
| 18 | | | | | | | | | |
| 19 | | | | | | | | | |
| 20 | | | | | | | | | |
| 21 | | | | | | | | | |
| 22 | | | | | | | | | |

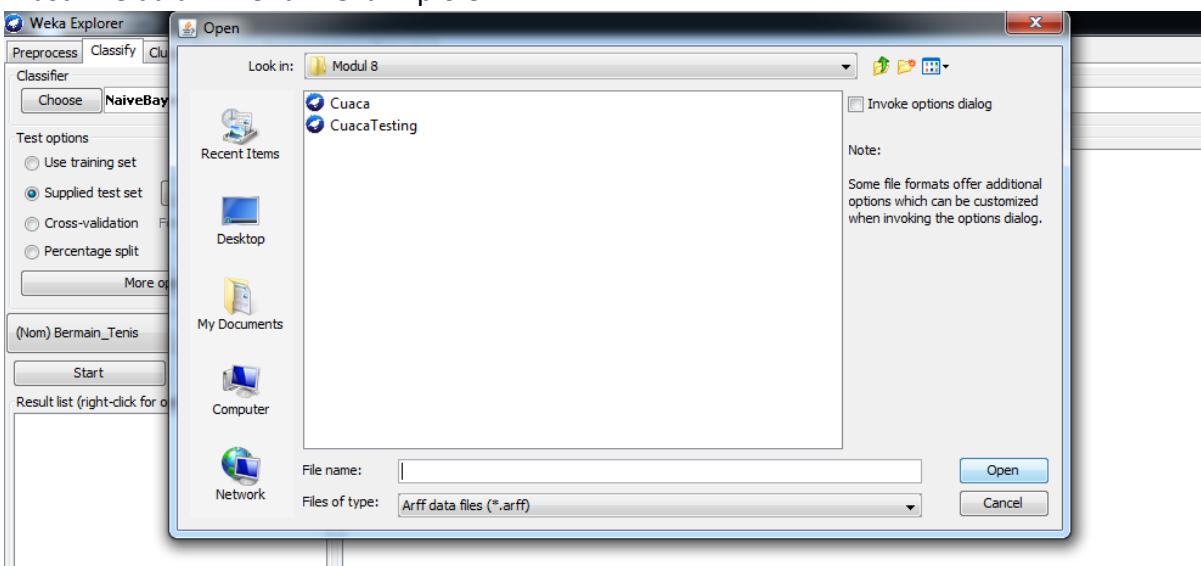
2. Data Testing dengan format ARFF dari tabel 8.1 sebagai data uji yang akan di prediksi memiliki variable-variable independen dan variabel dependen yang sama

```
C: > Users > LABSI-07 > Documents > Modul 8 > CuacaTesting.txt
1 @relation CuacaTesting
2
3 @attribute Cuaca {Cerah, Mendung, Hujan}
4 @attribute Suhu real
5 @attribute Kelembaban_Udara real
6 @attribute Berangin {TIDAK, YA}
7 @attribute Bermain_Tenis {TIDAK, YA}
8
9 @data
10 Cerah, 75, 64, TIDAK, ?
11 Cerah, 80, 68, YA, ?
12 Cerah, 83, 68, YA, ?
13 Mendung, 70, 96, TIDAK, ?
14 Mendung, 68, 81, TIDAK, ?
15 Hujan, 65, 75, YA, ?
16 Hujan, 65, 85, YA, ?
```

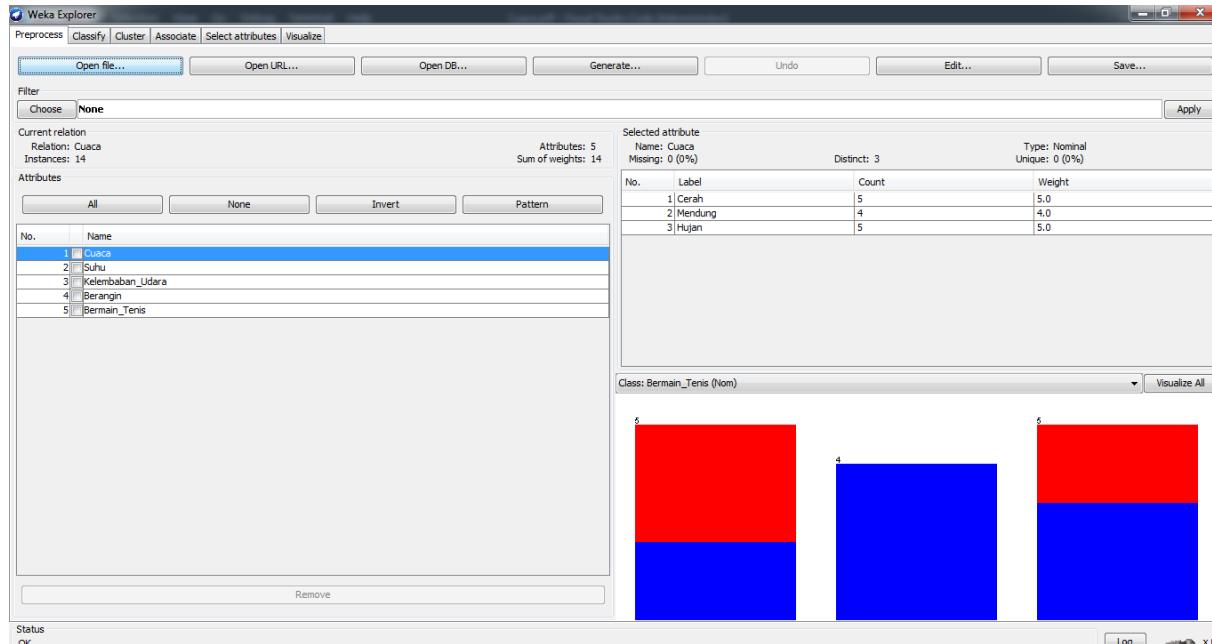
3. Menyimpannya dengan format CuacaTesting.arff



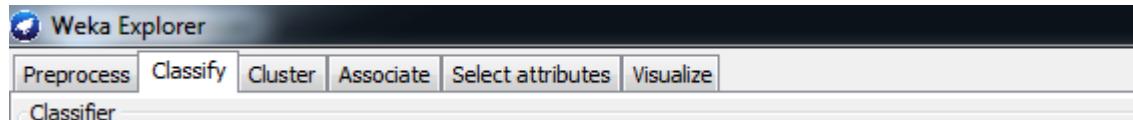
4. Masuk ke dalam menu Weka Explorer



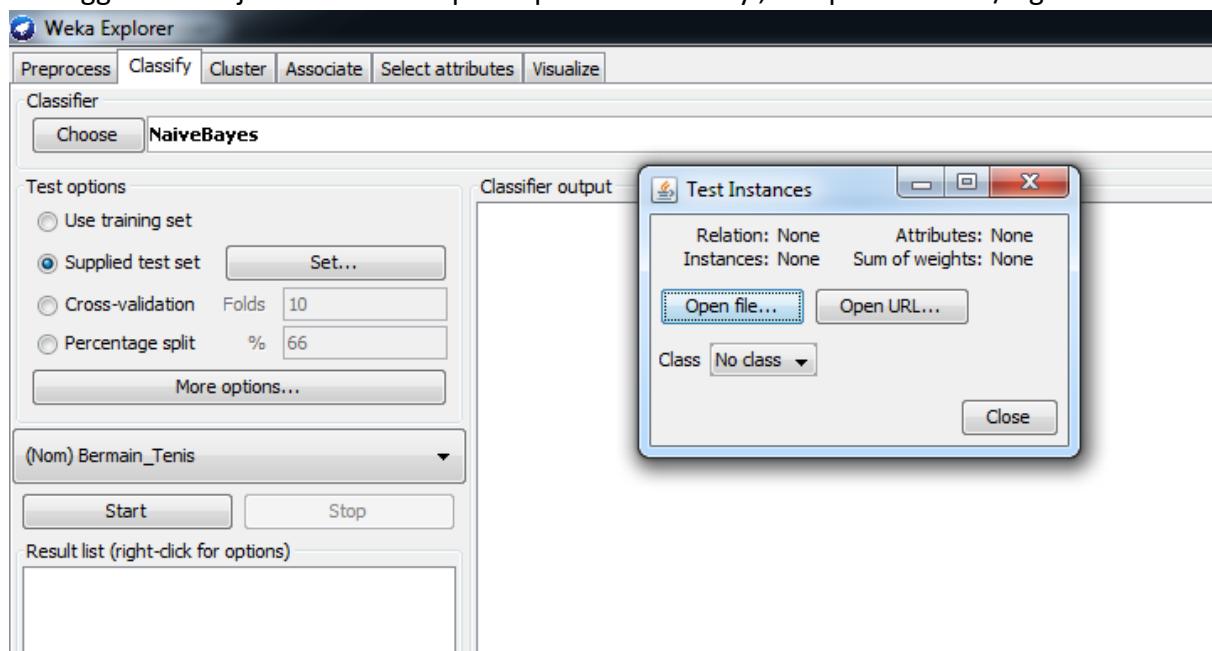
5. Buka kembali file Cuaca.arff, file ini akan di gunakan sebagai data pelatihan untuk memprediksi data testing pada file CuacaTesting.arff



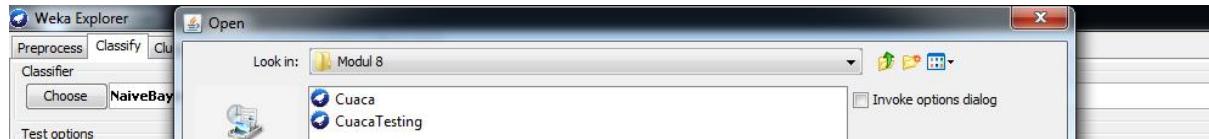
6. Pada window Weka Exploler, Pilih tab Classify



7. Sehingga muncul jendela Weka Exploler pad atab Classify , lalu pilih metode/algoritma Naive Bayes

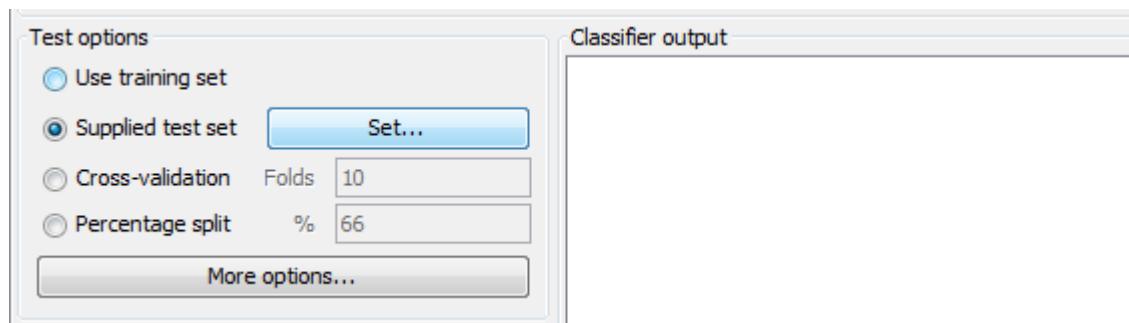


8. File CuacaTesting.arff di tentukan sebagai data testing pada kegiatan ini.



9. Pada menu test Option terdapat 4 pilihan pengujian, yaitu:

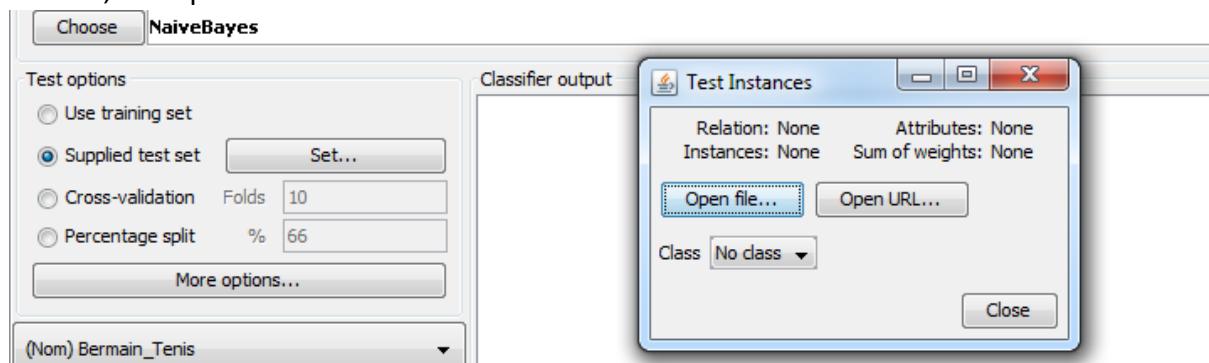
- Use training set
- Supplied test set
- Cross-validation
- Percentage split



10. Percobaan kali ini, saya akan menggunakan option Supplied test set.



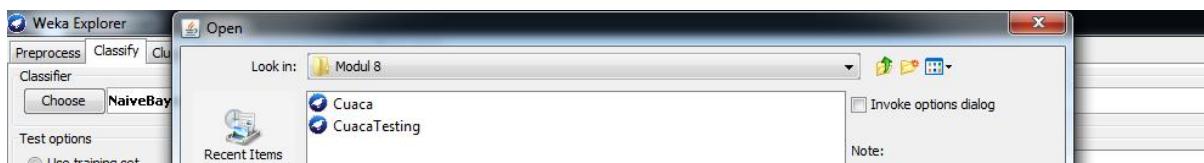
11. Klik Set, lalu open file



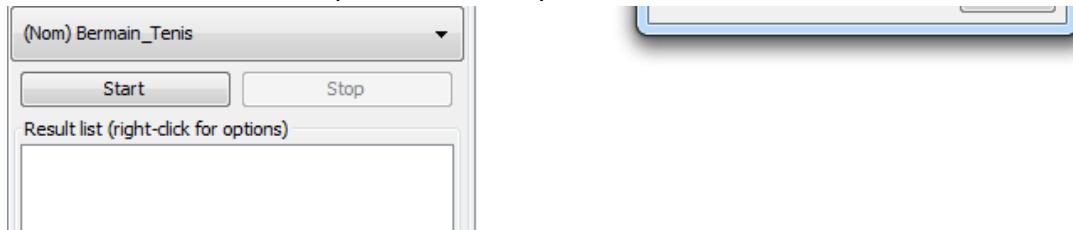
12. Pilih file CuacaTesting.arff sebagai data untuk di uji



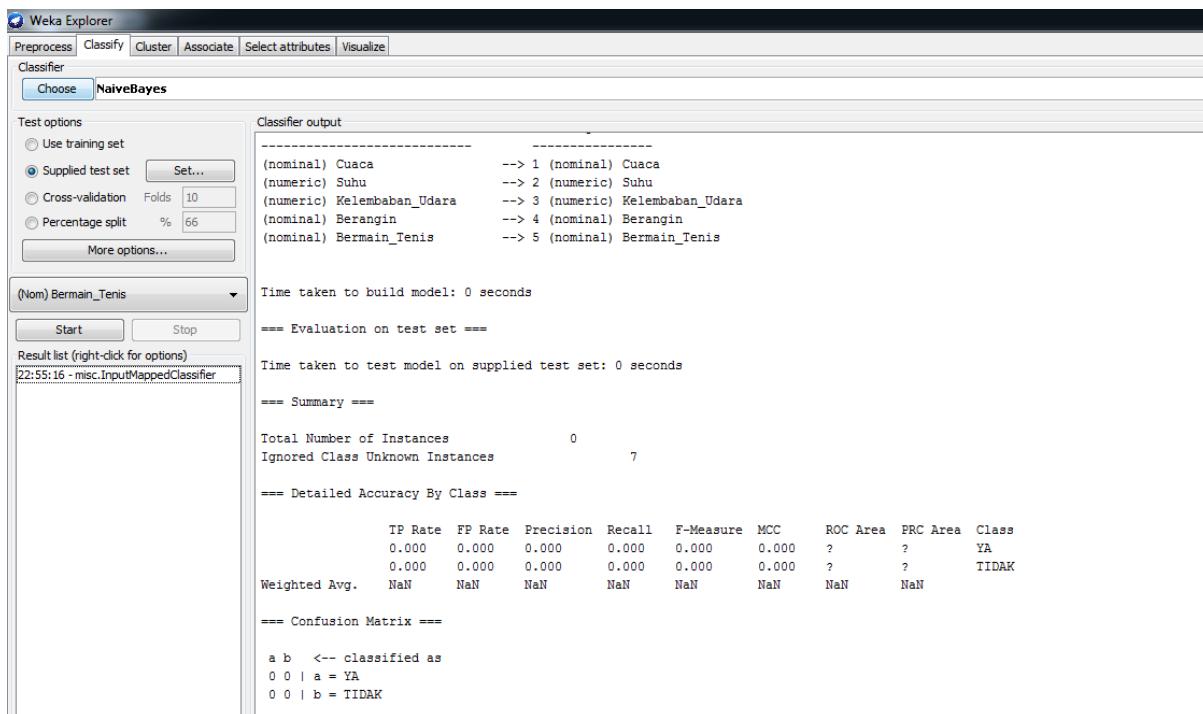
13. File CuacaTesting.arff akan diset sebagai data uji dengan variable predictor (class) adalah Bermain_Tenis.



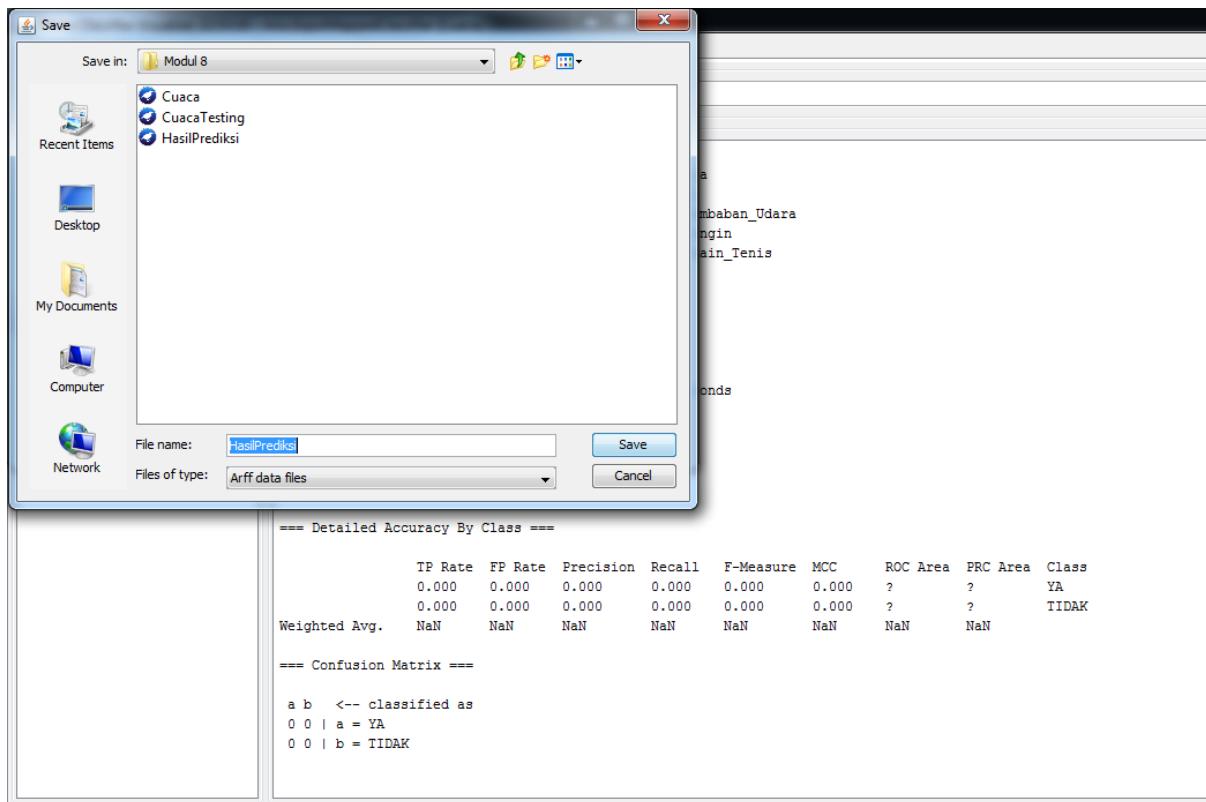
14. Klik Start untuk memulai proses naive bayes



17. Untuk melihat hasil prdiksi kita melakukan proses Classifier Errors, dengan memilih kotak result list, dengan memilih option Visualize classifier errors



18. Pada weka classifier visualize window, abaikan hasil apapun , lalu klik save, dan simpan dengan nama file HasilPrediksi.arff



19. Tutup semua weka window, lalu pilih menu Weka GUI Chooser, lalu pilih menu Tools – ArffViewer.
20. Lalu Open file HasilPrediksi.arff yang telah di simpan, dan lihatlah hasil prdiksi yang telah di ketahui pada kolom predicted Bermain_Tenis Nominal.

| No. | Cuaca | Suhu | Kelembaban_Udara | Berangin | prediction margin | predicted Bermain_Tenis |
|-----|---------|---------|------------------|----------|-------------------|-------------------------|
| | Nominal | Numeric | Numeric | Nominal | Numeric | Nominal |
| 1 | Cerah | 75.0 | 64.0 | TIDAK | 0.762765 | YA |
| 2 | Cerah | 80.0 | 68.0 | YA | 0.087878 | YA |
| 3 | Cerah | 83.0 | 68.0 | YA | -0.083834 | 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 | 65.0 | 85.0 | YA | -0.160143 | TIDAK |

➤ Implementasi Naïve Bayes dengan Rapid Miner

1. Persiapkan file Tabel_Cuaca.xls yang terdiri dari 2 sheet.
2. Sheet1 di gunakan sebagai data training dan Sheet2 di gunakan sebagai data uji.
3. Masing masing tabel memiliki atribut yaitu:
 - a. Cuaca(X1)
 - b. Suhu(X2)
 - c. Kelembaban_udara(X3)
 - d. Berangin(X4)
 - e. Bermain_Tenis(Y), sebagai variable predictor
- Tabel Data Training pada Sheet1

Tabel_Cuaca - Excel

The screenshot shows a Microsoft Excel spreadsheet titled "Tabel_Cuaca - Excel". The ribbon menu is visible at the top, with "Home" selected. The table structure is as follows:

| | A | B | C | D | E | F | G | H | I |
|----|---------|------|------------------|----------|---------------|---|---|---|---|
| 1 | Cuaca | Suhu | Kelembaban_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 | Tidak | Tidak | | | | |
| 16 | | | | | | | | | |
| 17 | | | | | | | | | |
| 18 | | | | | | | | | |

- Tabel Data Testing pada Sheet2

| | A | B | C | D | E | F | G | H |
|----|---------|------|------------------|----------|---|---|---|---|
| 1 | Cuaca | Suhu | Kelembaban_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 | | | | |
| 9 | | | | | | | | |
| 10 | | | | | | | | |

4. Buka Aplikasi RapidMiner, dan membuat repositories baru yang berfungsi sebagai lokasi penyimpanan terpusat untuk data dan proses analisa.

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

- Repository:** Shows available training resources, samples, and local repositories.
- Operators:** A list of operator categories including Data Access, Blending, Cleansing, Modeling, Scoring, and Validation.
- Process:** The main workspace where processes are designed. It currently displays a message: "Your process looks empty. Add some data first. Drag data or operators here."
- Parameters:** A panel for setting process parameters like logverbosity and logfile.
- Help:** A panel providing information about the current process and RapidMiner Studio Core.

5. Import Data, pilih file Tabel_Cuaca.xls

6. Perhatikan Sel excel untuk menyesuaikan banyaknya data yang akan di gunakan, Select the cells to import

| | A | B | C | D | E |
|----|---------|--------|------------------|----------|---------------|
| 1 | Cuaca | Suhu | Kelembaban_Udara | Berangin | Bерmain_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 | 98.000 | 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 | 92.000 | Tidak | Tidak |

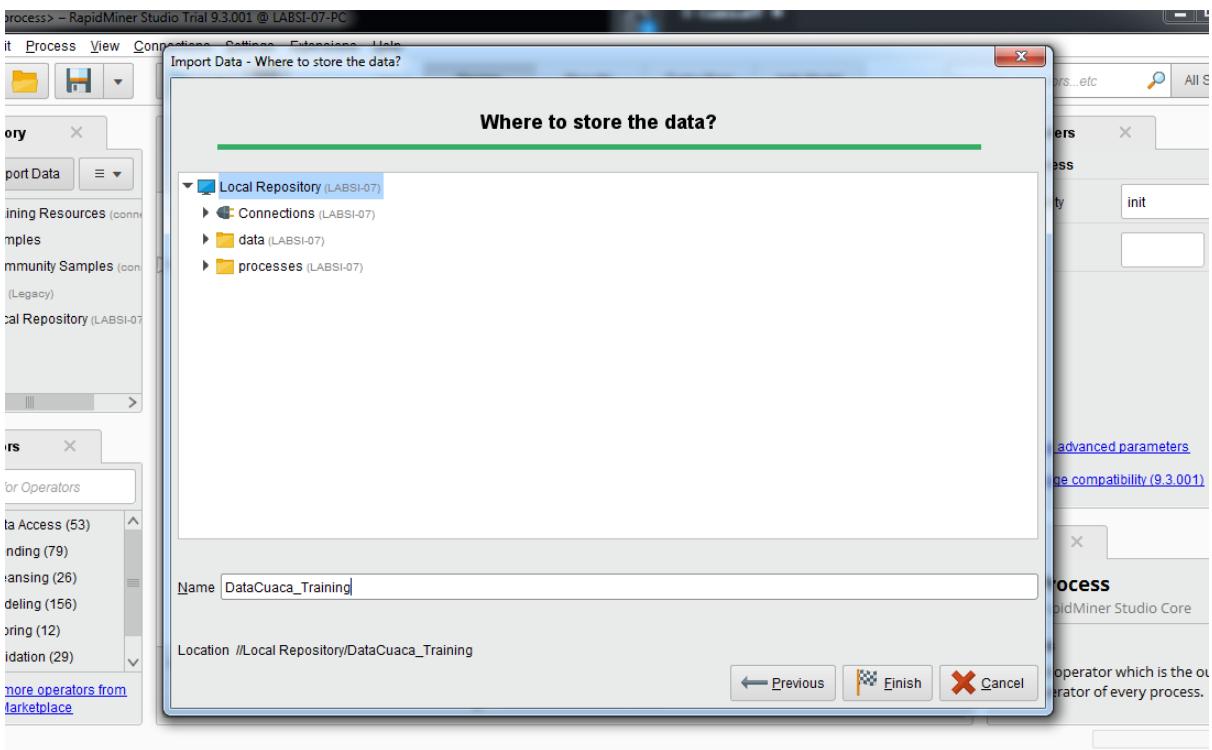
7. Langkah Format your column ubah kolom Bermain Tenis dengan tipe data binomial karena hanya ada dua keputusan yaitu (YA & TIDAK)

| | Cuaca | Suhu | Kelembaban_Udara | Berangin | Bерmain_Tenis |
|----|---------|------|------------------|----------|---------------|
| 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 |
| 14 | Hujan | 71 | 91 | Tidak | Tidak |

8. Ubah juga menjadi label dan Change Role sebagai option untuk tipe binomial

| | Cuaca polynomial | Suhu integer | Kelembaban_U... integer | Berangin polynomial | Bermain_Tenis binomial label |
|----|---------------------|-----------------|----------------------------|------------------------|------------------------------------|
| 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 |
| 14 | Hujan | 71 | 91 | Tidak | Tidak |

9. Simpan dengan nama DataCuaca_Training, finish

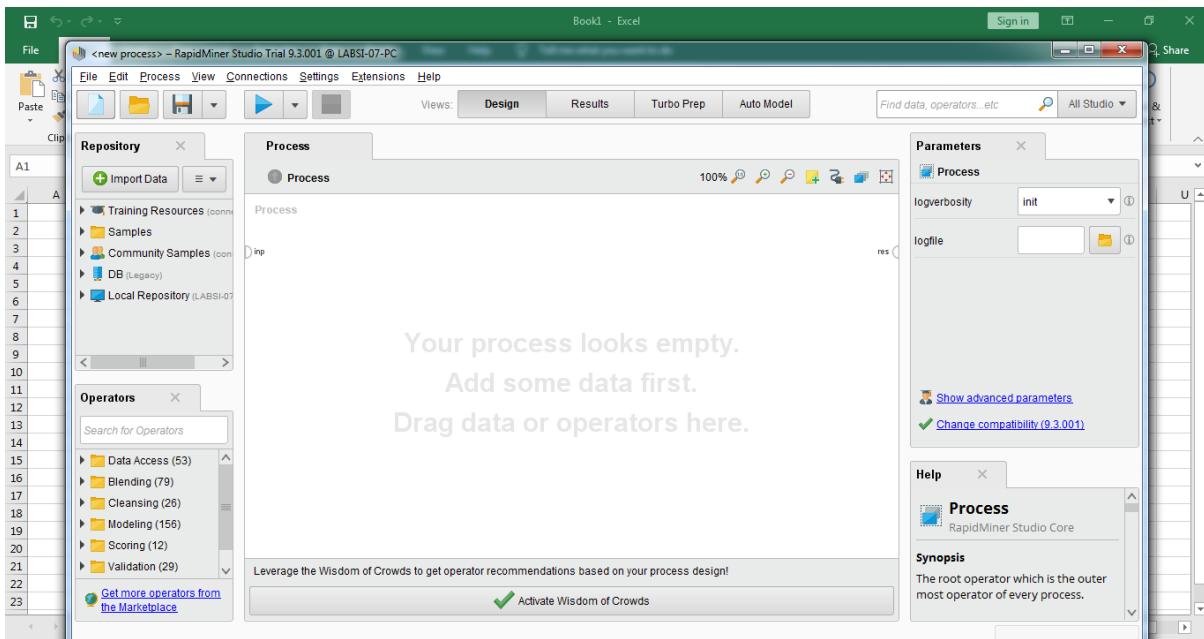


10. Hasil import file Tabel_Cuaca.xls pada Sheet1 akan menampilkan Exempel Set;

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

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

11. Kembali ke jendela Design Prospective



12. Sekarang adalah, kita mengambil data testing yang di import dari file Tabel_Cuaca.xls pada sheet2 (Testing), di langkah ini pastikan data terpilih ada 4 variabel, dan tanpa ada label.

Import Data - Format your columns.

Format your columns.

Replace errors with missing values

| Cuaca | Suhu | Kelembaban_Udara | Berangin |
|------------|---------|------------------|------------|
| polynomial | integer | integer | polynomial |
| 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 |

no problems.

← Previous Next → Cancel

13. Simpan dengan nama DataCuaca_Training.

Import Data - Where to store the data?

Where to store the data?

Local Repository (LABSI-07)

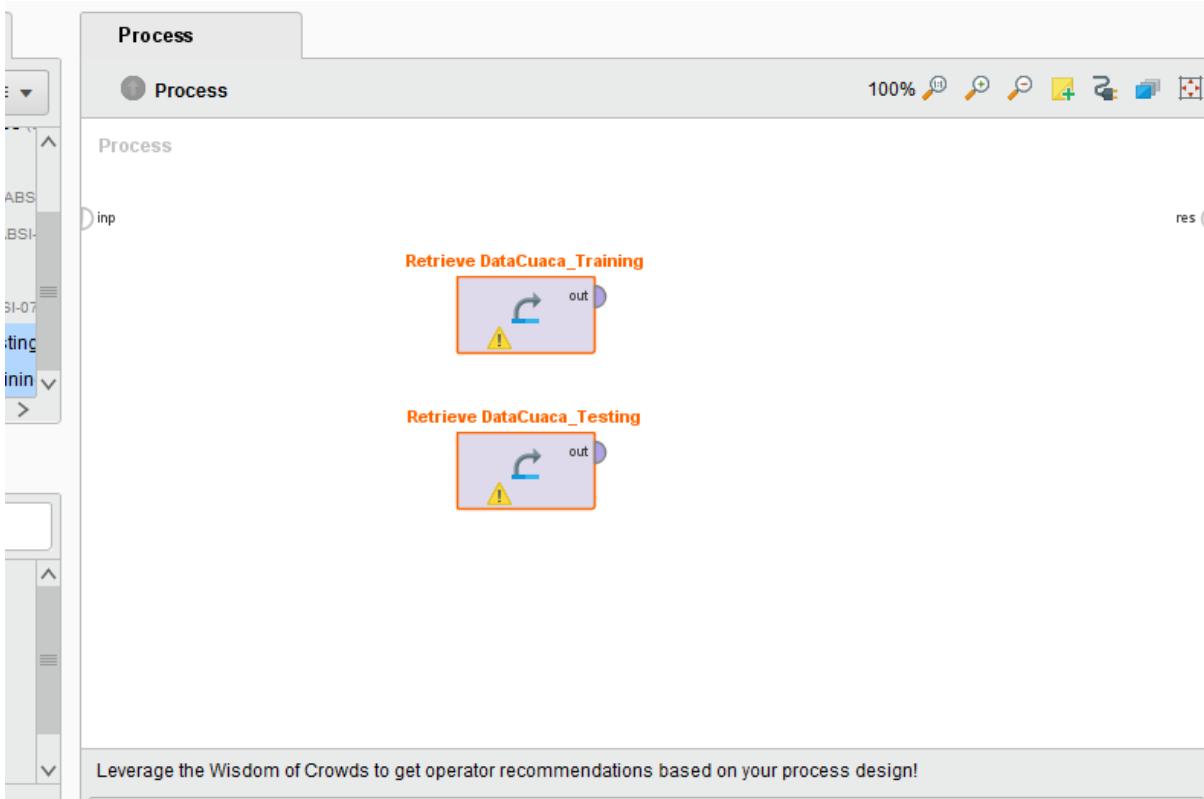
- Connections (LABSI-07)
- data (LABSI-07)
- processes (LABSI-07)
- DataCuaca_Training (LABSI-07 - v1, 10/9/19 11:19 PM - 529 bytes)

Name

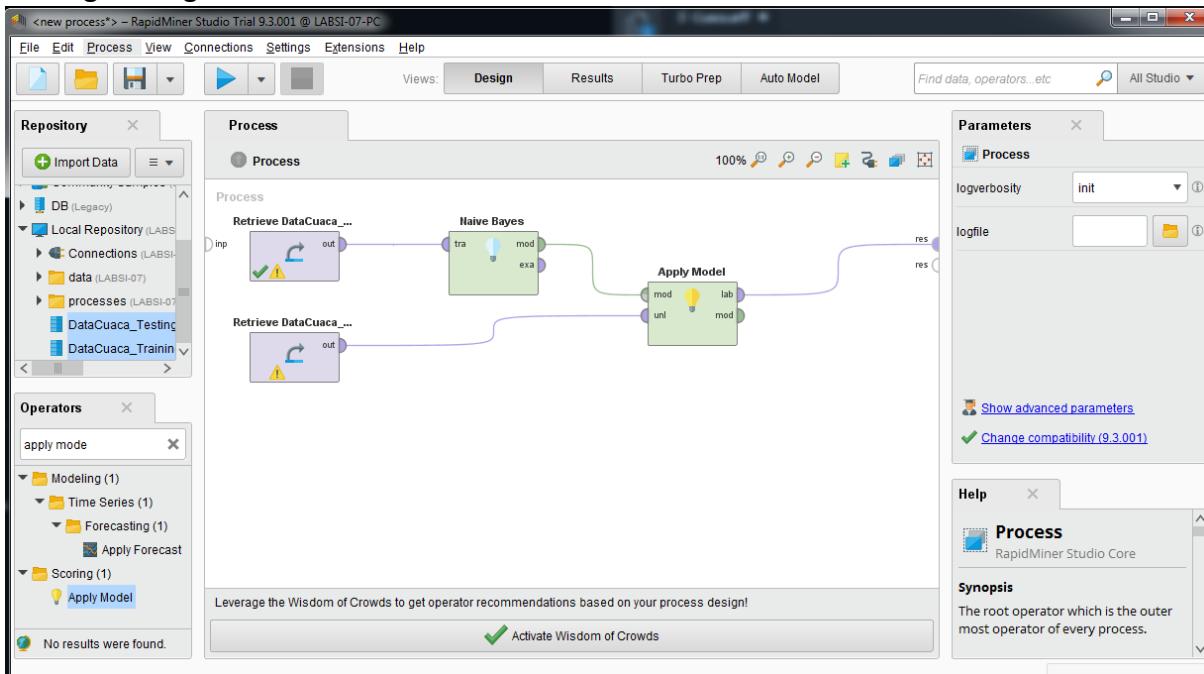
Location /Local Repository/DataCuaca_Testing

← Previous Finish Cancel

14. Selanjutnya adalah membuat desain Naive Bayes, drag and drop DataCuaca_Training dan DataCuaca_Testing ke process window view.



15. Masukan juga operatator Naive Bayes dan Apply Model ke dalam proces view, hubungkan konksi masing masing .



16. Run proses naive Bayes.
17. Hasil proses klasifikasi naive bayes, pada tab Data, dapat di lihat hasil prediksi terhadap data testing serta tingkat confidence nilai kelas pada masing-masing data

ExampleSet (/Local Repository/DataCuaca_Testing) ExampleSet (/Local Repository/DataCuaca_Training) ExampleSet (Apply Model)

| Row No. | prediction(Bern...) | confidence(...) | confidence(...) | Cuaca | Suhu | Kelembaban... | Berangin |
|---------|---------------------|-----------------|-----------------|---------|------|---------------|----------|
| 1 | Ya | 0.213 | 0.787 | Cerah | 75 | 65 | Tidak |
| 2 | Ya | 0.401 | 0.599 | Cerah | 80 | 68 | Ya |
| 3 | Tidak | 0.801 | 0.199 | Cerah | 83 | 87 | Ya |
| 4 | Ya | 0.027 | 0.973 | Mendung | 70 | 96 | Tidak |
| 5 | Ya | 0.011 | 0.989 | Mendung | 68 | 81 | Tidak |
| 6 | Ya | 0.285 | 0.715 | Hujan | 65 | 75 | Ya |
| 7 | Ya | 0.470 | 0.530 | Hujan | 64 | 85 | Ya |

Filter (7 / 7 examples): all

Repository

- Training Resources (connected)
 - Samples
 - Community Samples (connected)
 - DB (Legacy)
 - Local Repository (LABSI-07)
 - Connections (LABSI-07)
 - data (LABSI-07)
 - processes (LABSI-07)
 - DataCuaca_Testing (LABSI-07 - v1, 103 examples)
 - DataCuaca_Training (LABSI-07 - v1, 103 examples)

Pada tab Statistics distribusi kelas pada variable Y (Bermain_Tenis) rerata nilai confidence sebesar 0,353 untuk nilai TIDAK dan 0,647 untuk nilai YA

ExampleSet (/Local Repository/DataCuaca_Testing) ExampleSet (/Local Repository/DataCuaca_Training) ExampleSet (Apply Model)

| Name | Type | Missing | Statistics | Filter (7 / 7 attributes): Search for Attributes |
|--|------------|---------|---|--|
| prediction(Bermain_Tenis) | Binominal | 0 | Least: Tidak (1) Most: Ya (6) Values: Ya (6), Tidak (1) | |
| Confidence_Tidak confidence(Tidak) | Real | 0 | Min: 0.011 Max: 0.801 Average: 0.315 | |
| Confidence_Ya confidence(Ya) | Real | 0 | Min: 0.199 Max: 0.989 Average: 0.685 | |
| Cuaca | Polynomial | 0 | Least: Mendung (2) Most: Cerah (3) Values: Cerah (3), Hujan (1), Mendung (2), Tidak (1) | |
| Suhu | Integer | 0 | Min: 64 Max: 83 Average: 72.143 | |
| Kelembaban_Udara | Integer | 0 | Min: 65 Max: 96 Average: 79.571 | |
| Berangin | Polynomial | 0 | Least: Tidak (3) Most: Ya (4) Values: Ya (4), Tidak (3) | |

Showing attributes 1 - 7 Examples: 7 Special Attributes: 3 Regular Attributes: 4

Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

MODUL 8

KLASIFIKASI : NAÏVE BAYES

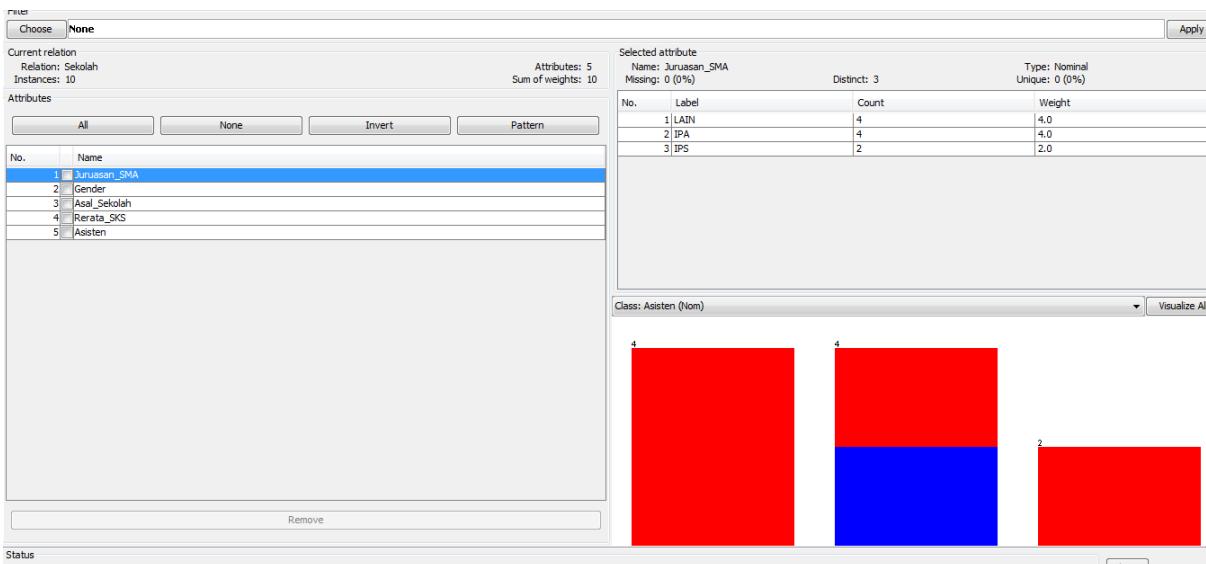
- **TUGAS**

1. Berdasarkan tabel berikut, buatlah file dalam format.xls dan format.arff, Data ini akan di gunakan sebagai DataTesting

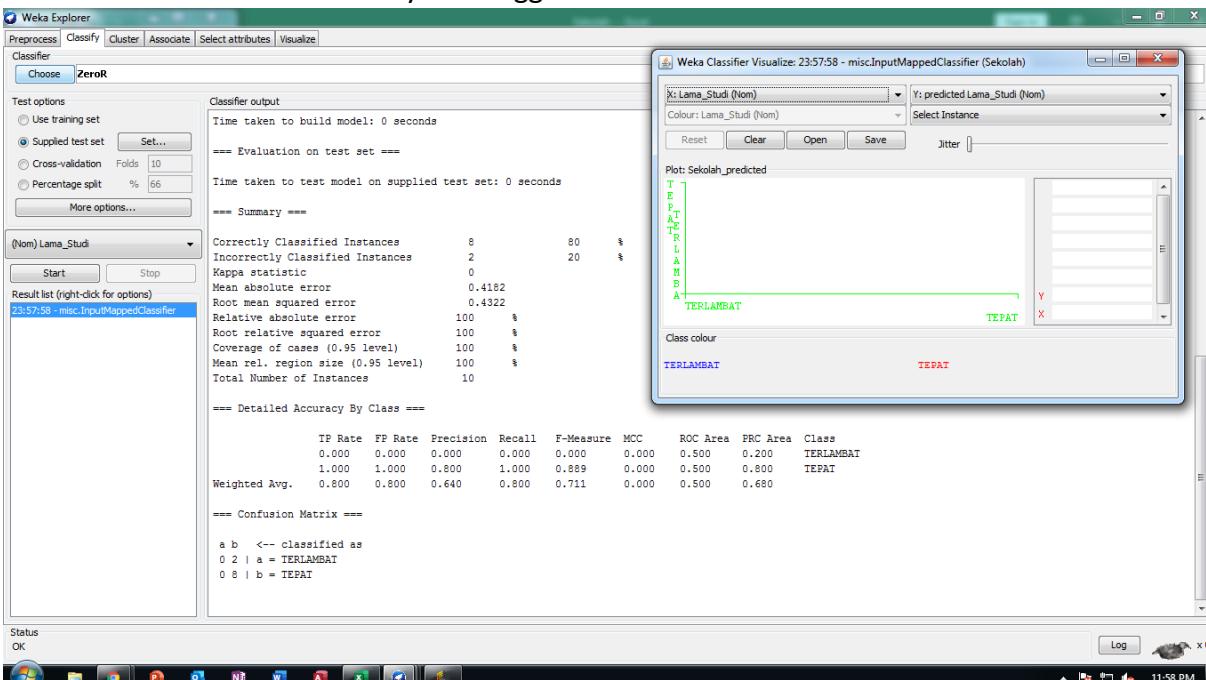
The screenshot shows two windows side-by-side. On the left is Microsoft Excel with a spreadsheet titled 'data testing - Excel'. The spreadsheet has columns labeled 'Jurusan_SMA' (values: LAIN, IPA, LAIN, IPS, LAIN, IPA, IPA, IPA, IPS, LAIN), 'Gender' (values: WANITA, PRIA, PRIA, PRIA, WANITA, WANITA, PRIA, PRIA, PRIA, WANITA), and 'Asal_Sekolah' (values: SURAKARTA, SURAKARTA, SURAKARTA, LUAR, SURAKARTA, LUAR, SURAKARTA, SURAKARTA, LUAR, SURAKARTA). Row 12 is empty. On the right is Visual Studio Code with a file named 'Cuaca.arff' open. The code defines a relation 'Sekolah' with attributes 'Juruasan_SMA', 'Gender', and 'Asal_Sekolah'. It includes a header section with '@relation Sekolah', attribute definitions like '@attribute Jurusan_SMA {LAIN, IPA, IPS}', and a data section starting with '@data' and listing 19 data points. The status bar at the bottom of VS Code shows 'Ln 1, Col 1' and '11:50 PM'.

| | Jurusan_SMA | Gender | Asal_Sekolah |
|----|-------------|--------|--------------|
| 1 | LAIN | WANITA | SURAKARTA |
| 2 | IPA | PRIA | SURAKARTA |
| 3 | LAIN | PRIA | SURAKARTA |
| 4 | IPS | PRIA | LUAR |
| 5 | LAIN | WANITA | SURAKARTA |
| 6 | IPA | WANITA | LUAR |
| 7 | IPA | PRIA | SURAKARTA |
| 8 | IPA | PRIA | SURAKARTA |
| 9 | IPS | PRIA | LUAR |
| 10 | LAIN | WANITA | SURAKARTA |
| 11 | | | |
| 12 | | | |
| 13 | | | |
| 14 | | | |
| 15 | | | |
| 16 | | | |
| 17 | | | |
| 18 | | | |
| 19 | | | |
| 20 | | | |
| 21 | | | |
| 22 | | | |
| 23 | | | |

2. Gunakan file arff pada tugas no 1 dalam modul 7 sebagai DataTraining, lakukan prdiksi terhadap DataTesting.arff menggunakan weka



Hasil Prediksi metode Naive Bayes menggunakan Weka



Hasil di prediction menggunakan weka

ARFF-Viewer - C:\Users\Person\Documents\modul 8\tugas\hasilprediksi.arff

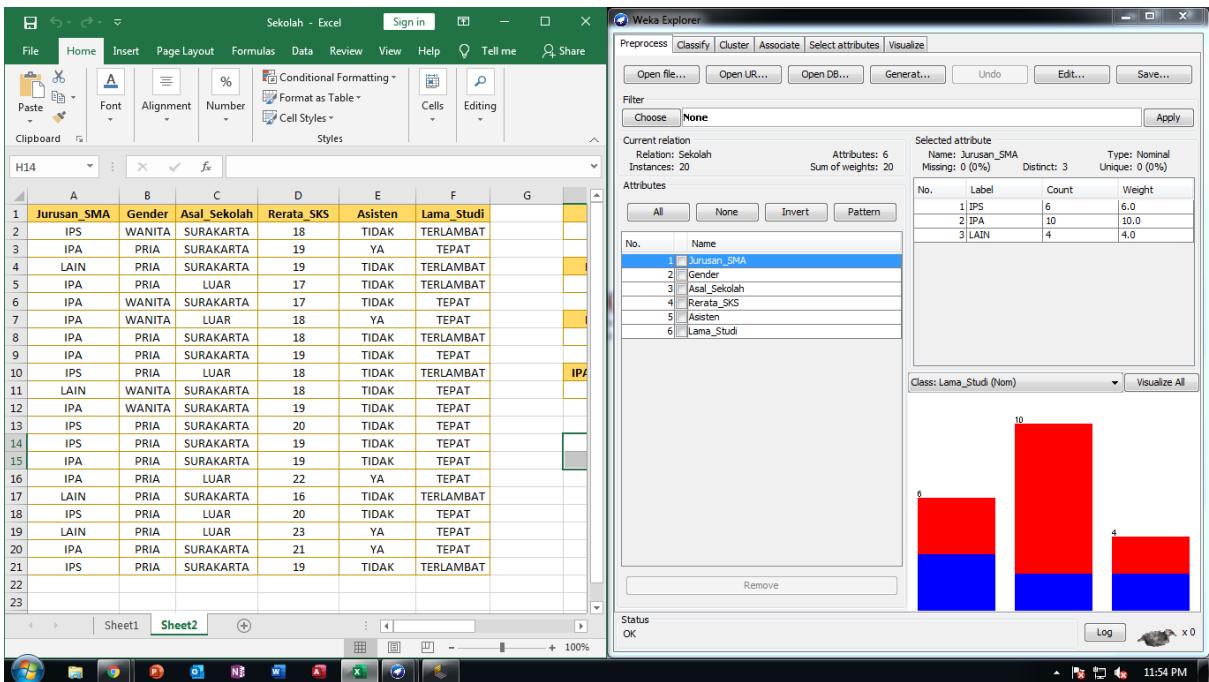
File Edit View

hasilprediksi.arff

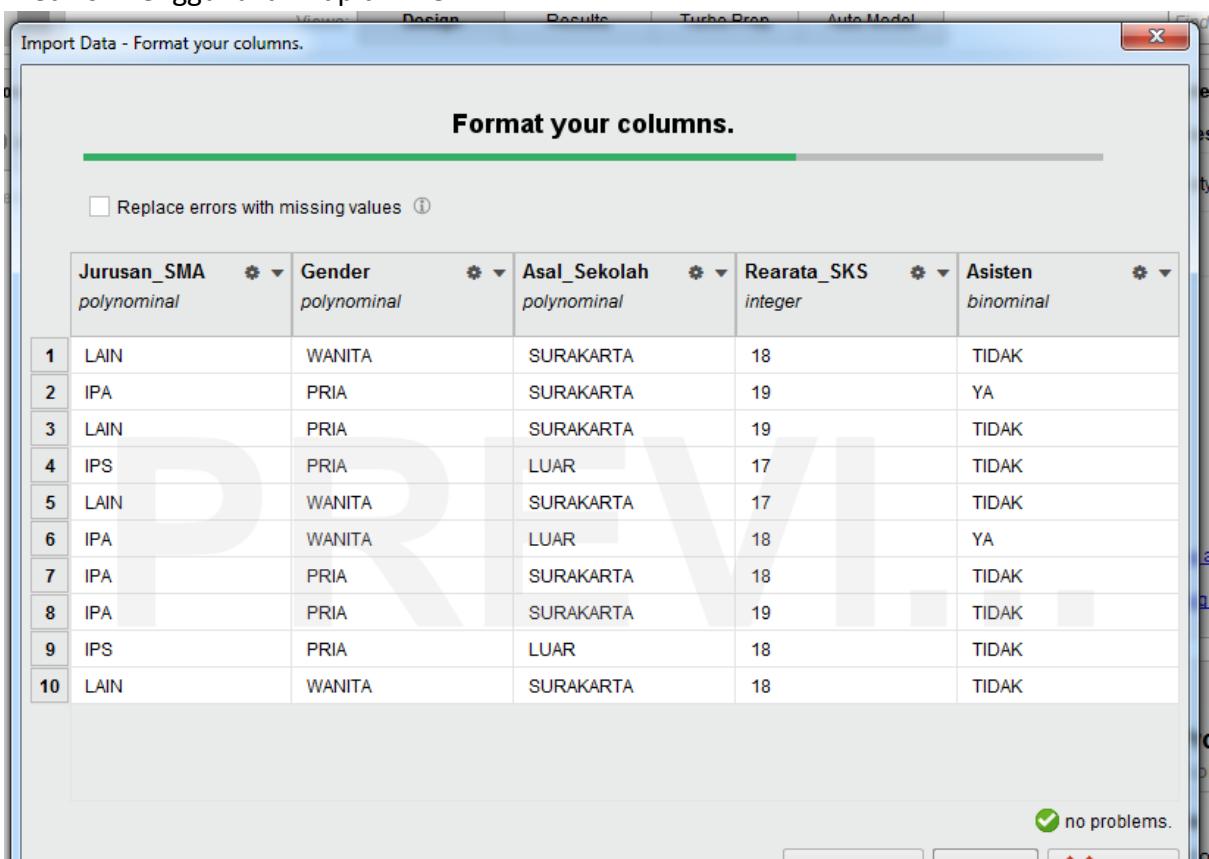
Relation: Sekolah_predicted

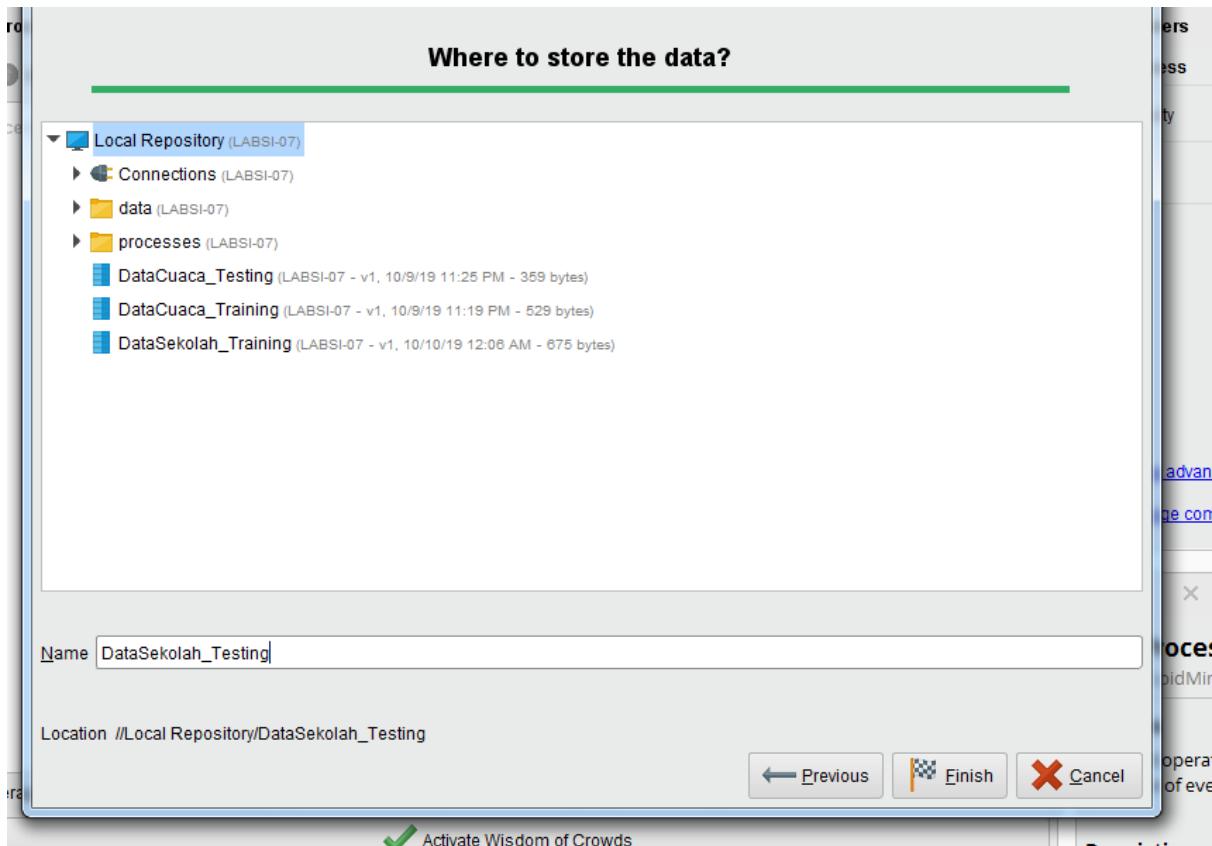
| No. | 1: Jurusan_SMA Nominal | 2: Gender Nominal | 3: Asal_Sekolah Nominal | 4: Rerata_SKS Numeric | 5: prediction margin Numeric | 6: predicted Asisten Nominal | 7: Asisten Nominal |
|-----|------------------------|-------------------|-------------------------|-----------------------|------------------------------|------------------------------|--------------------|
| 1 | LAIN | WANITA | SURAKARTA | 18.0 | 0.785601 | TIDAK | TIDAK |
| 2 | IPA | PRIA | SURAKARTA | 19.0 | -0.033585 | TIDAK | YA |
| 3 | LAIN | PRIA | SURAKARTA | 19.0 | 0.684916 | TIDAK | TIDAK |
| 4 | IPS | PRIA | LUAR | 17.0 | 0.93308 | TIDAK | TIDAK |
| 5 | LAIN | WANITA | SURAKARTA | 17.0 | 0.973646 | TIDAK | TIDAK |
| 6 | IPA | WANITA | LUAR | 18.0 | 0.166955 | YA | YA |
| 7 | IPA | PRIA | SURAKARTA | 18.0 | 0.42833 | TIDAK | TIDAK |
| 8 | IPA | PRIA | SURAKARTA | 19.0 | 0.033585 | TIDAK | TIDAK |
| 9 | IPS | PRIA | LUAR | 18.0 | 0.525209 | TIDAK | TIDAK |
| 10 | LAIN | WANITA | SURAKARTA | 18.0 | 0.785601 | TIDAK | TIDAK |

3. Gunakan file excel pad tugas no1 dalam modul 6 sebagai data training, lakukan prediksi terhadap, data testing (excel) di atas menggunakan RapidMiner!
DataTraining.xls dan DataTraining.arff.



Prediksi menggunakan RapidMiner.





4. Hasil percobaan Tugas no 3, berapakah nilai rerata confidence untuk atribut Lama_Sudi, dengan nilai TEPAT? Dan berapakah nilai rerata confidence utnuk atribut lama_studi dengan nilai TERLAMBAT?

The screenshot shows the KNIME interface with the title "ExampleSet (Apply Model)". The table contains 10 rows of data:

| Ro.. | prediction(L...) | confidence(TERLAMBAT) | confidence(TEPAT) | 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 |

ExampleSet (10 examples, 3 special attributes, 5 regular attributes)

- Nilai Rerata Confidence untuk attribute Lama_Studi dengan nilai TEPAT:



- Nilai Rerata Confidence untuk attribute Lama_Studi dengan nilai TERLAMBAT:



5. Dari hasil percobaan tugas no 3 berapa jumlah banyaknya orang yang akan lulus TEPAT, dan berapa jumlah banyaknya orang yang akan lulus TERLAMBAT

- Jumlah banyaknya Orang yang akan lulus TEPAT :



- Jumlah banyaknya Orang yang akan lulus TERLAMBAT :



6. Prediksikan 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

- Memilih Nama dewi untuk di prediksikan

Views Design Results Turbo Prep Auto Model Deployments Find

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Testing Cell range: A1:F2 Select All Define header row: 1

| | A | B | C | D | E | F |
|----|-------------|--------|--------------|------------|---------|------|
| 1 | Jurusan_SMA | Gender | Asal_Sekolah | Rerata_SKS | Asisten | Nama |
| 2 | IPA | WANITA | LUAR | 18.000 | TIDAK | Dewi |
| 3 | IPA | PRIA | SURAKARTA | 19.000 | YA | |
| 4 | LAIN | PRIA | SURAKARTA | 19.000 | TIDAK | |
| 5 | IPS | PRIA | LUAR | 17.000 | TIDAK | |
| 6 | LAIN | WANITA | SURAKARTA | 17.000 | TIDAK | |
| 7 | IPA | WANITA | LUAR | 18.000 | YA | |
| 8 | IPA | PRIA | SURAKARTA | 18.000 | TIDAK | |
| 9 | IPA | PRIA | SURAKARTA | 19.000 | TIDAK | |
| 10 | IPS | PRIA | LUAR | 18.000 | TIDAK | |
| 11 | LAIN | WANITA | SURAKARTA | 18.000 | TIDAK | |

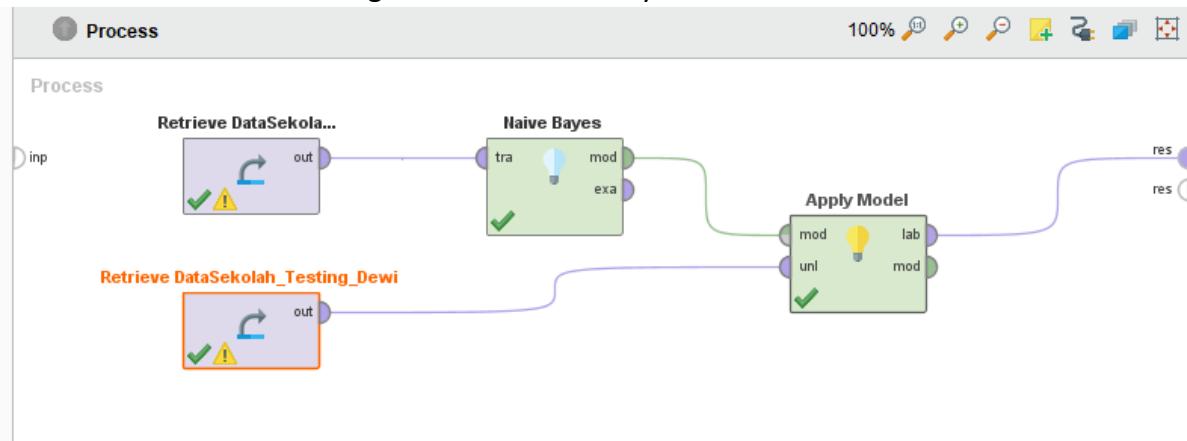
- Tidak ada yang di ubah dan di kasih label karena ini data testing

Format your columns.

Replace errors with missing values ⓘ

| Jurusan_SMA | Gender | Asal_Sekolah | Rerata_SKS | Asisten | Nama |
|-------------|------------|--------------|------------|------------|------------|
| polynomial | polynomial | polynomial | integer | polynomial | polynomial |
| 1 IPA | WANITA | LUAR | 18 | TIDAK | Dewi |

- Melakukan model dengan metode naive bayes



- Hasil ketepatan Lama_Studi Dewi adalah (TEPAT) tidak TERLAMBAT

The screenshot shows the RapidMiner interface with three tabs at the top: 'ExampleSet (/Local Repository/DataSekolah_Testing_MilikDewi)', 'ExampleSet (Apply Model)', and 'ExampleSet (/Local Repository/DataSekolah_Training)'. The 'ExampleSet (Apply Model)' tab is active. On the left, there's a sidebar with icons for Data, Statistics, Visualizations, and Help. The main area shows a table with one row. The columns are: Row No., prediction(Lama_Studi), confidence(TERLAMBAT), confidence(TEPAT), Jurusan_SMA, Gender, Asal_Sekolah, Rerata_SKS, Asisten, and Nama. The row data is: 1, TEPAT, 0.298, 0.702, IPA, WANITA, LUAR, 18, TIDAK, Dewi.

7. Prediksikan ketepatan studi si Jono, jika jono adalah PRIA jurusan LAIN pada SMA, asal sekolah dari SURAKARTA, mengambil sks rata rata sebanyak 17 SKS, dan pernah menjadi Asisten
- Memilih nama jono untuk di prediksikan

The screenshot shows a 'Select the cells to import' dialog box. At the top, it says 'Import Data - Select the cells to import.' Below that is a title 'Select the cells to import.' The dialog has a 'Sheet' dropdown set to 'Testing', a 'Cell range' input field containing 'A1:F2', a 'Select All' button, and a checked 'Define header row' checkbox with a value of '1'. The main area is a table with columns labeled A through F and rows numbered 1 through 5. The first row contains column headers: Jurusan_SMA, Gender, Asal_Sekolah, Rerata_SKS, Asisten, and Nama. The second row contains data: Lain, PRIA, SURAKARTA, 17.000, YA, JONO. The third row contains: PRIA, SURAKARTA, 19.000, YA. The fourth row contains: LAIN, PRIA, SURAKARTA, 19.000, TIDAK. The fifth row contains: IPS, PRIA, LUAR, 17.000, TIDAK.

| | A | B | C | D | E | F |
|---|-------------|--------|--------------|------------|---------|------|
| 1 | Jurusan_SMA | Gender | Asal_Sekolah | Rerata_SKS | Asisten | Nama |
| 2 | Lain | PRIA | SURAKARTA | 17.000 | YA | JONO |
| 3 | | PRIA | SURAKARTA | 19.000 | YA | |
| 4 | LAIN | PRIA | SURAKARTA | 19.000 | TIDAK | |
| 5 | IPS | PRIA | LUAR | 17.000 | TIDAK | |

- Tidak ada yang di ubah dan di kasih label, karena ini data testing

Import Data - Format your columns.

Format your columns.

Replace errors with missing values ⓘ

| Jurusan_SMA * | Gender * | Asal_Sekolah * | Rerata_SKS * | Asisten * | Nama * |
|---------------|------------|----------------|--------------|------------|------------|
| polynomial | polynomial | polynomial | integer | polynomial | polynomial |
| 1 Lain | PRIA | SURAKARTA | 17 | YA | JONO |

Views: Design Results Turbo Prep Auto Model Deployments

Import Data - Where to store the data?

Where to store the data?

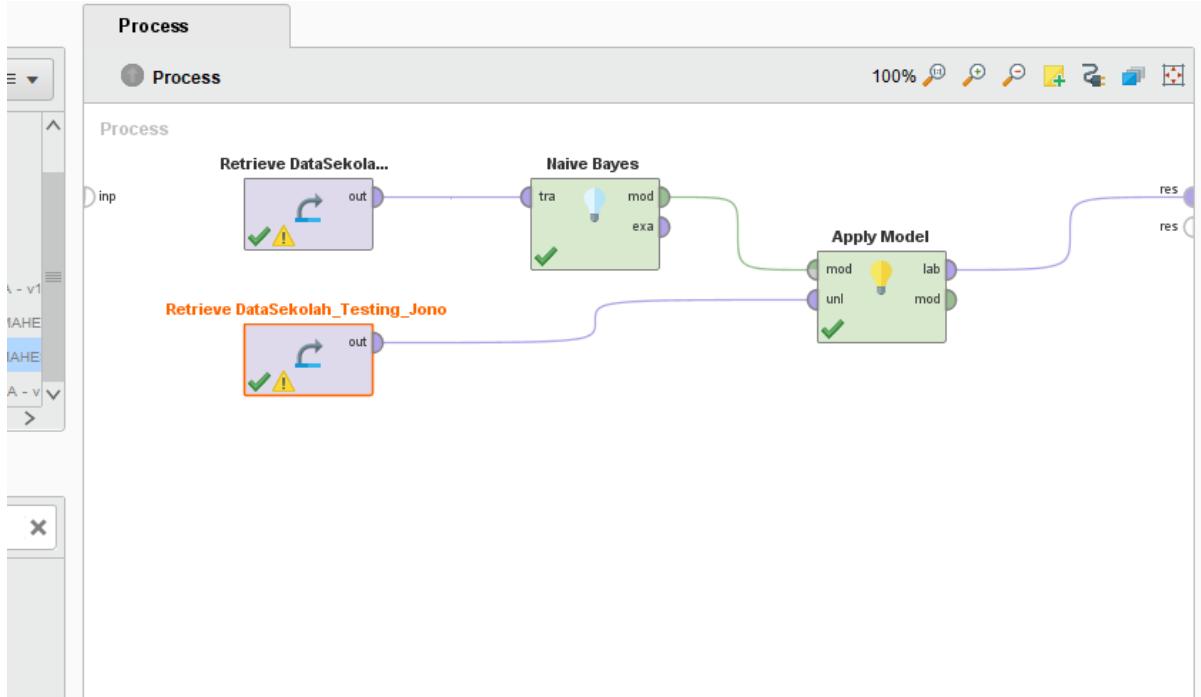
▼ Local Repository (OKI KUS MAHESA)

- ▶ **data** (OKI KUS MAHESA)
- ▶ **processes** (OKI KUS MAHESA)
- DataSekolah_Testing** (OKI KUS MAHESA - v1, 10/11/19 9:45 PM - 485 bytes)
- DataSekolah_Testing_Dewi** (OKI KUS MAHESA - v1, 10/11/19 10:11 PM - 415 bytes)
- DataSekolah_Training** (OKI KUS MAHESA - v1, 10/11/19 9:44 PM - 675 bytes)

Name

Location //Local Repository/DataSekolah_Testing_Jono

- Melakukan model dengan metode naive bayes



- Hasil ketepatan Lama_Studi Jono adalah (TEPAT) tidak TERLAMBAT

The screenshot shows the KNIME result history window with the following details:

- ExampleSet** entries:
 - //Local Repository/DataSekolah_Testing
 - //Local Repository/DataSekolah_Testing
 - //Local Repository/DataSekolah_Testing_MilikDewi
 - //Local Repository/DataSekolah_Testing_Dewi
 - //Local Repository/DataSekolah_Training
 - //Local Repository/DataSekolah_Training
 - //Local Repository/DataSekolah_Training
 - //Local Repository/DataSekolah_Training
 - //Local Repository/DataSekolah_Testing_Dewi
 - //Local Repository/DataSekolah_Testing_Jono
- Result History**: Shows the result of the **ExampleSet (Apply Model)**.
- Data View** table:

| Row No. | prediction(Lama_Studi) | confidence(TERLAMBAT) | confidence(TEPAT) | Jurusan_SMA | Gender | Asal_Sekolah | Rerata_SKS | Asisten | Nama |
|---------|------------------------|-----------------------|-------------------|-------------|--------|--------------|------------|---------|------|
| 1 | TEPAT | 0.076 | 0.924 | Lain | PRIA | SURAKARTA | 17 | YA | JONO |

Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

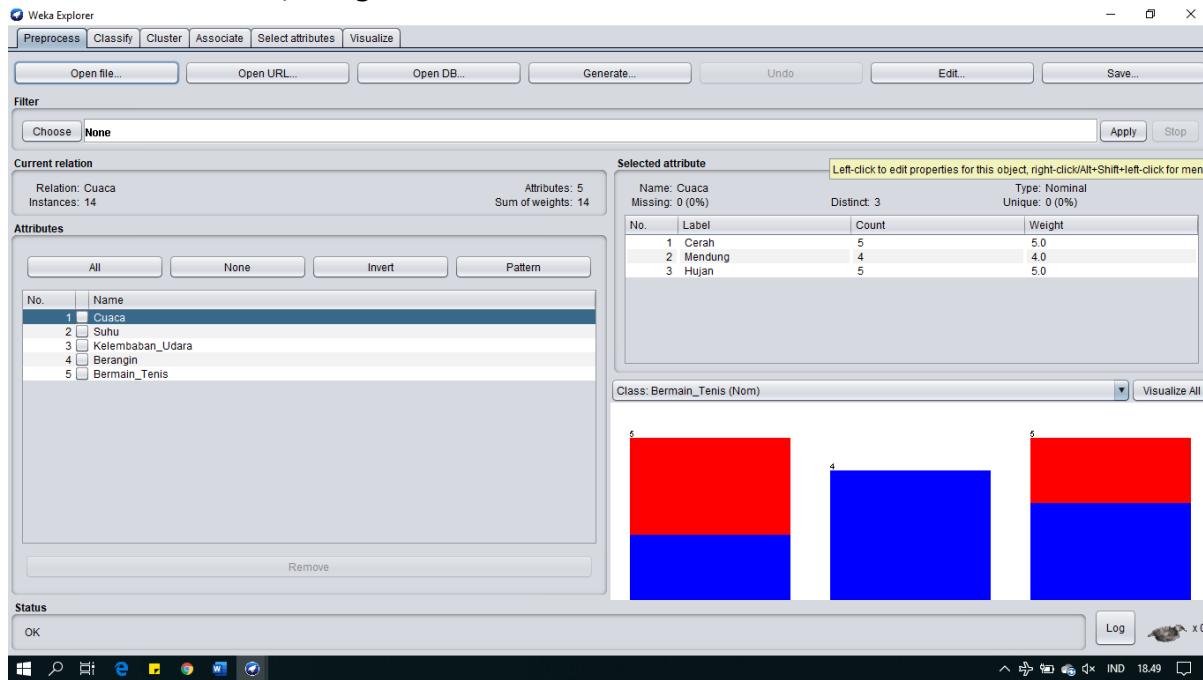
Modul 9

KLASIFIKASI : DECISION TREE

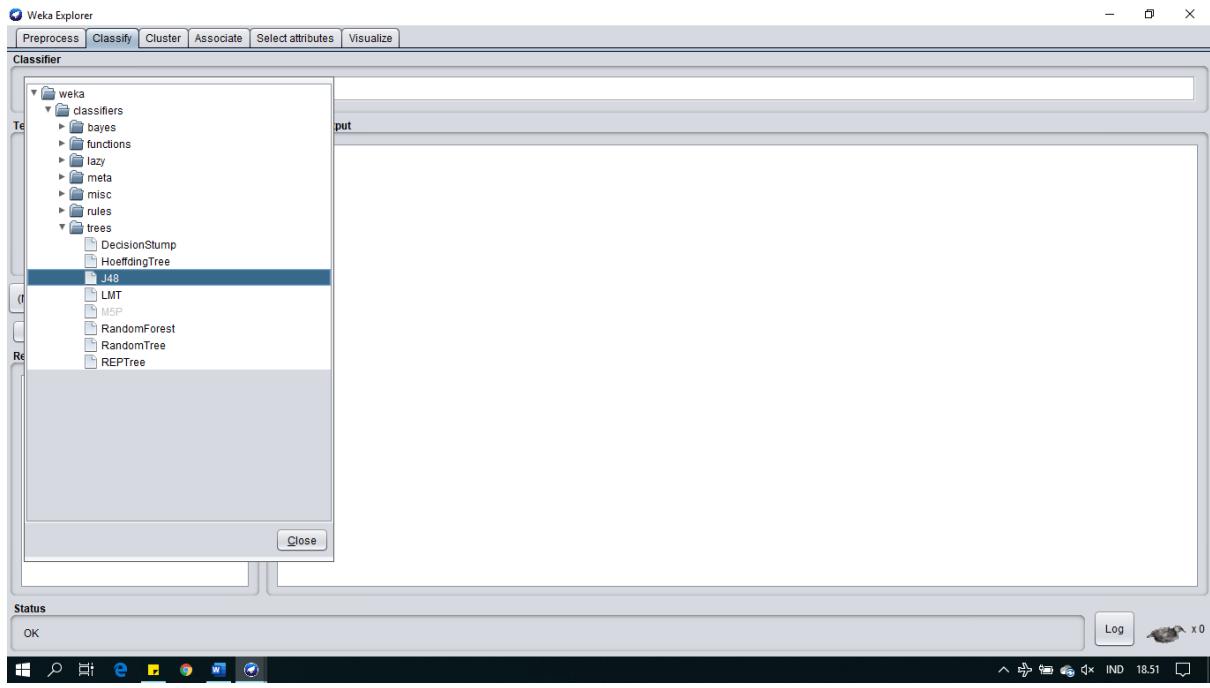
❖ Langkah – Langkah Praktikum

- Pohon Keputusan Menggunakan Weka**

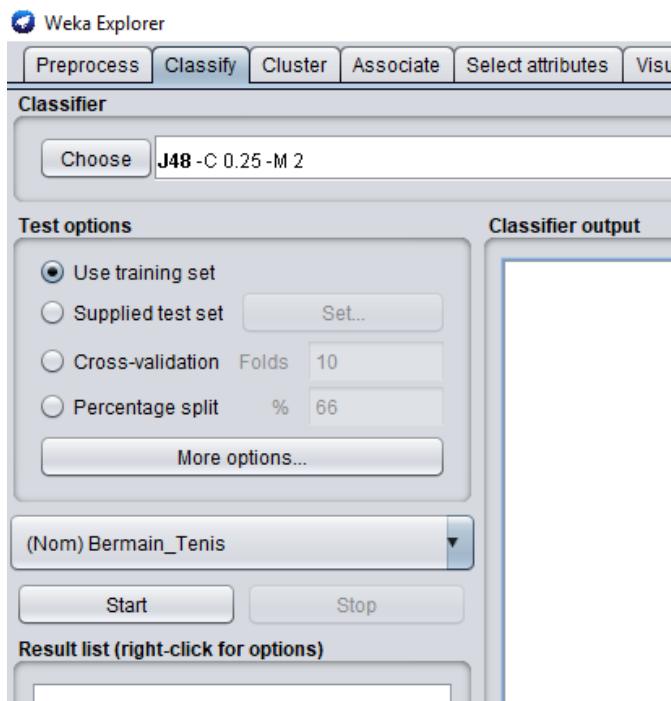
1. Jalankan aplikasi Weka-Exploler
2. Buka file Cuaca.arff, dengan weka



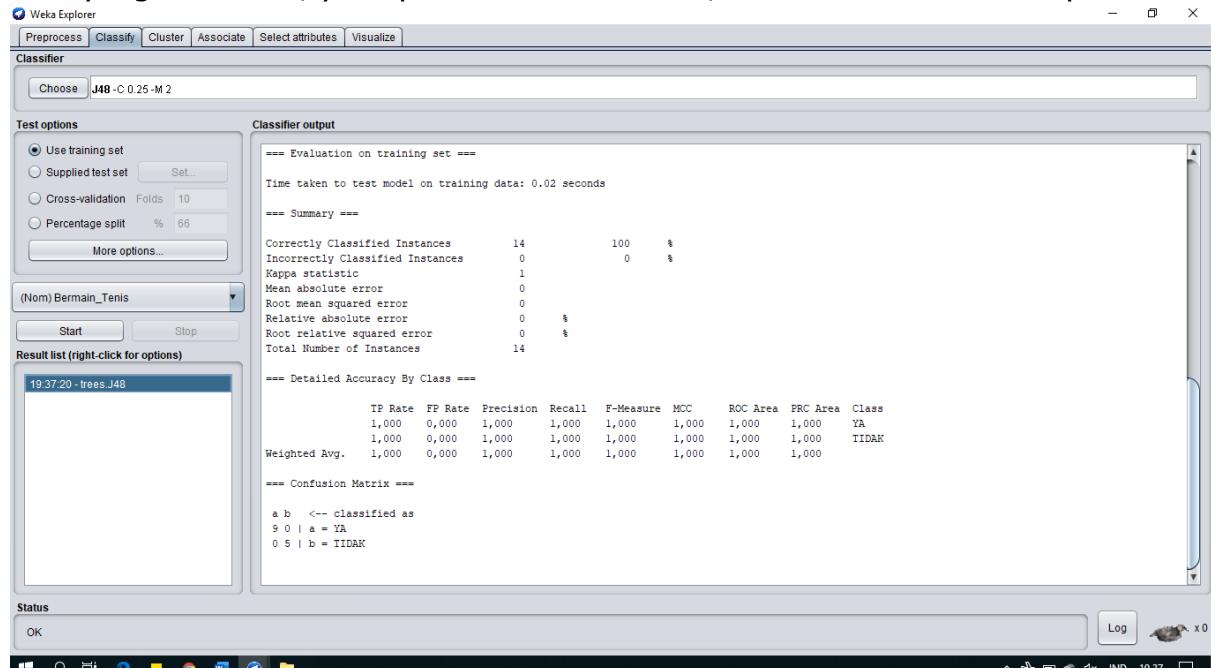
3. Buka tab Classify.
4. Pilih Trees-J48, karena kita akan menggunakan algoritma pohon keputusan J48



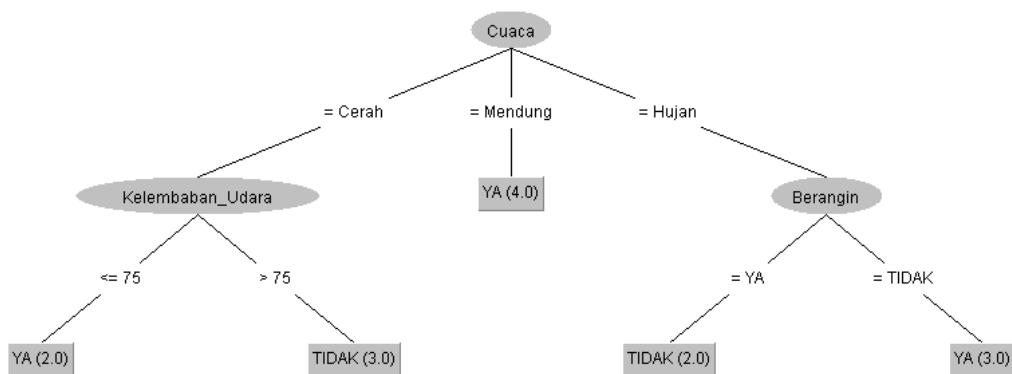
5. Pilih Use Training set data Cuaca.arff kita hanya menggunakan sebagai data training untuk klasifikasi.



6. Setelah Proses selesai maka perhitungan klasifikasi menggunakan algoritma J48, ada 2 hasil yang di berikan, yaitu pada kolom Result list, dan kolom Classifier output.



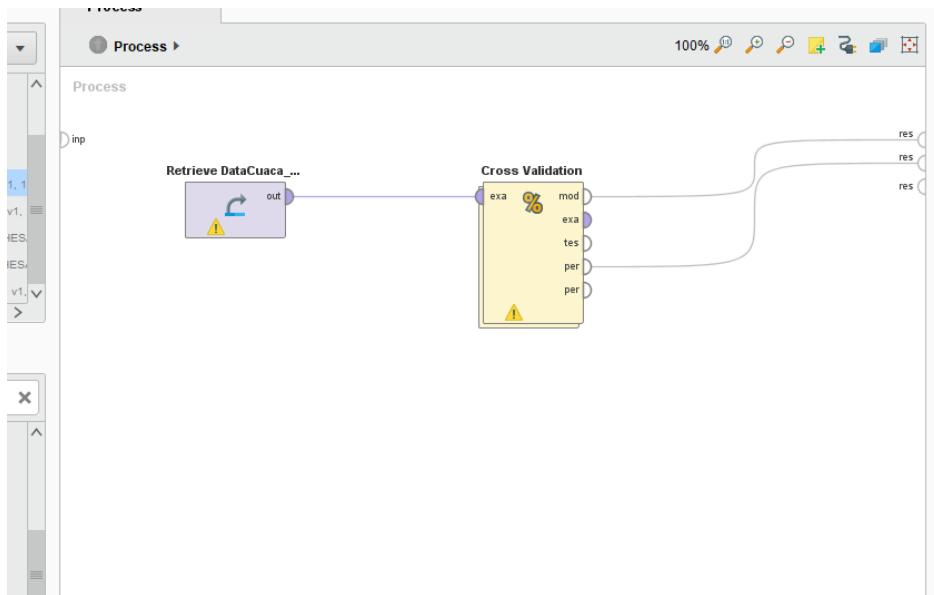
7. Berdasarkan hasil tersebut dapat di ketahui sebagai berikut:
- Jumlah simpul daun pada pohon keputusan = 5
 - Jumlah simpul keseluruhan pada kolom keputusan = 8
 - Waktu yang di butuhkan untuk proses pelatihan = 0,02 detik
 - Tingkat ketetapan klasifikasi = 100%
 - Tingkat ketidaktepatan klasifikasi = 0%
8. Hasil skema pohon keputusan



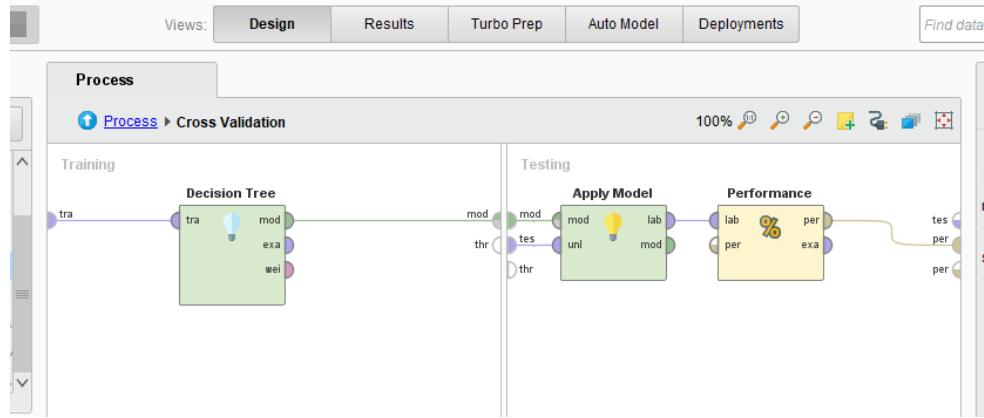
9. Berdasarkan pohon keputusan tersebut, dapat di lihat jenis-jenis simpul yang ada, berikut:
 - a) Simpu akar = Cuaca
 - b) Simpul internal = Kelembaban_udara, dan Berangin
 - c) Simpul daun = YA, TIDAK
10. Klasifikasi yang berbentuk yaitu:
 - a) Seseorang akan bermain tenis (YA) jika kondisi sebagai berikut:
 - i. Cuaca = Cerah, Kelembaban_udara ≤ 75 , (nilai atribut lain di abaikan)
 - ii. Cuaca = Mendung, (kondisi lain di abaikan)
 - iii. Cuaca = Hujan, Berangin=TIDAK, (nilai atribut lain di abaikan)
 - b) Seseorang tidak akan bermain tenis (TIDAK) jika kondisi sebagai berikut:
 - i. Cuaca = Cerah, Kelembaban_udara > 75 , (nilai atribut lain abaikan)
 - ii. Cuaca = Hujan, berangin = ya, (nilai atribut lain di abaikan)

• Pohon Keputusan Menggunakan RapidMiner

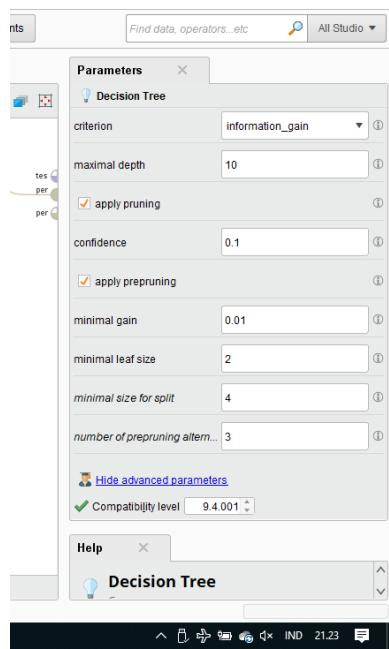
1. Menggunakan DataCuaca_Training. Untuk dilakukan oprasi klsifikasi decision tree, dan di area procces view drag oprator Cross Validation untuk membuat Apply Model dan Performance



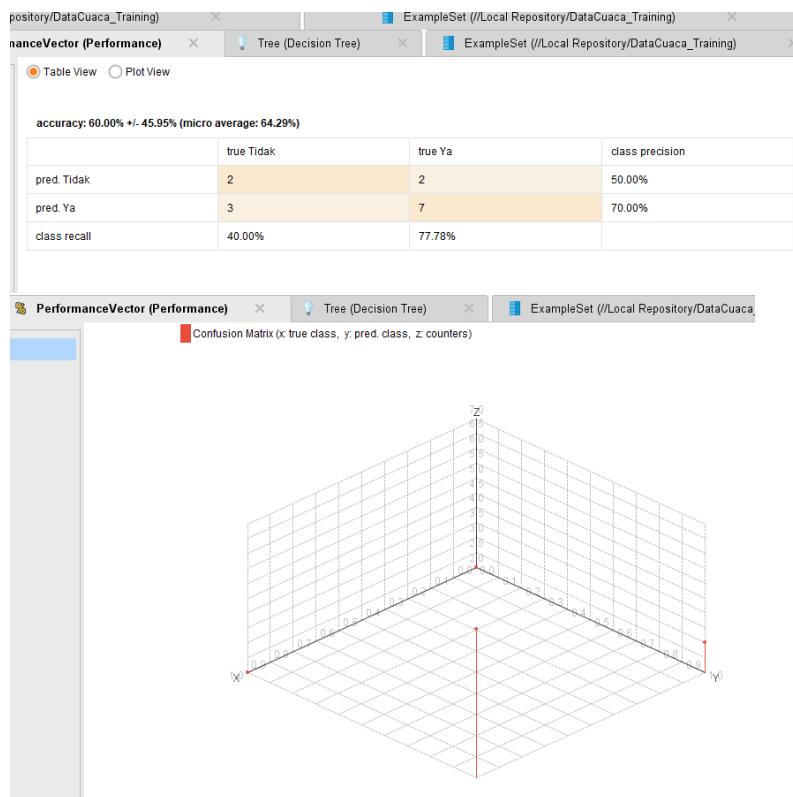
2. Setting configurssi Cross Validation dengan klik 2x, lalu setting Oprattor Decision Tree dalam area Training, oprattor Apply Model dan Performance dalam area Testing



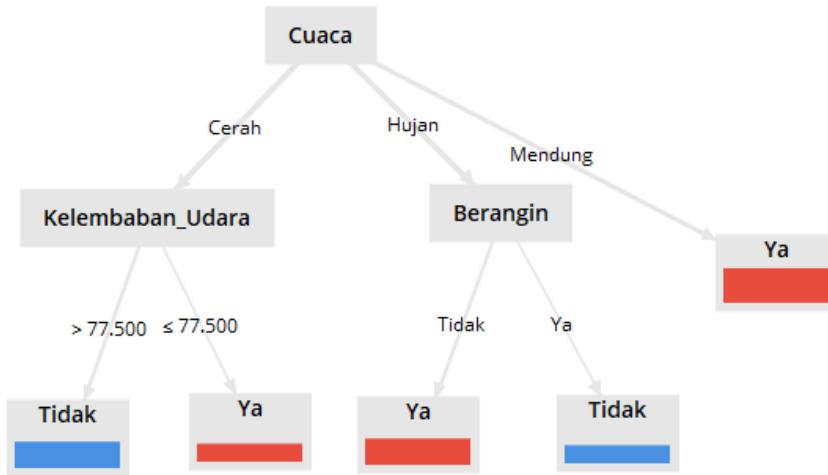
3. Klik 1x Decision Tree Opration pad area Training, dan pastikan kriteria yang di gunakan adalah Information Gain (j48) pada kolom Parameter.



4. Hasil klasifikasi pohon keputusan., yaitu ada 2 hasil yang menjadi output dari proses klasifikasi pohon keputusan:
- PerformanceVector (Performance), menunjukkan tingkat akurasi, presisi, recal. Dalam bentuk tabel dan plot view.



b) Tree (Decision Tree) menunjukkan hasil pohon keputusan dari proses klasifikasi.



5. Berdasarkan pohon keputusan jenis-jenis simpul sebagai berikut:
 - a) Simpul akar = Cuaca
 - b) Simpul internal = Kelembaban_udara, dan Berangin
 - c) Simpul Daun = Ya, Tidak
6. Klasifikasi yang terbentuk:
 - a) Seseorang akan bermain tenis (YA) jika kondisi sebagai berikut:

- i. Cuaca = Cerah, Kelembaban_udara <= 77,5. (nilai atribut lain di abaikan)
 - ii. Cuaca = Mendung, (kondisi lain di abaikan)
 - iii. Cuaca = Hujan, Berangin=TIDAK, (nilai atribut lain di abaikan)
- b) Seseorang tidak akan bermain tenis (TIDAK) jika kondisi sebagai berikut:
- i. Cuaca = Cerah, Kelembaban_udara > 77,5. (nilai atribut lain abaikan)
 - ii. Cuaca = Hujan, berangin = Ya, (nilai atribut lain di abaikan)

Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

Modul 9

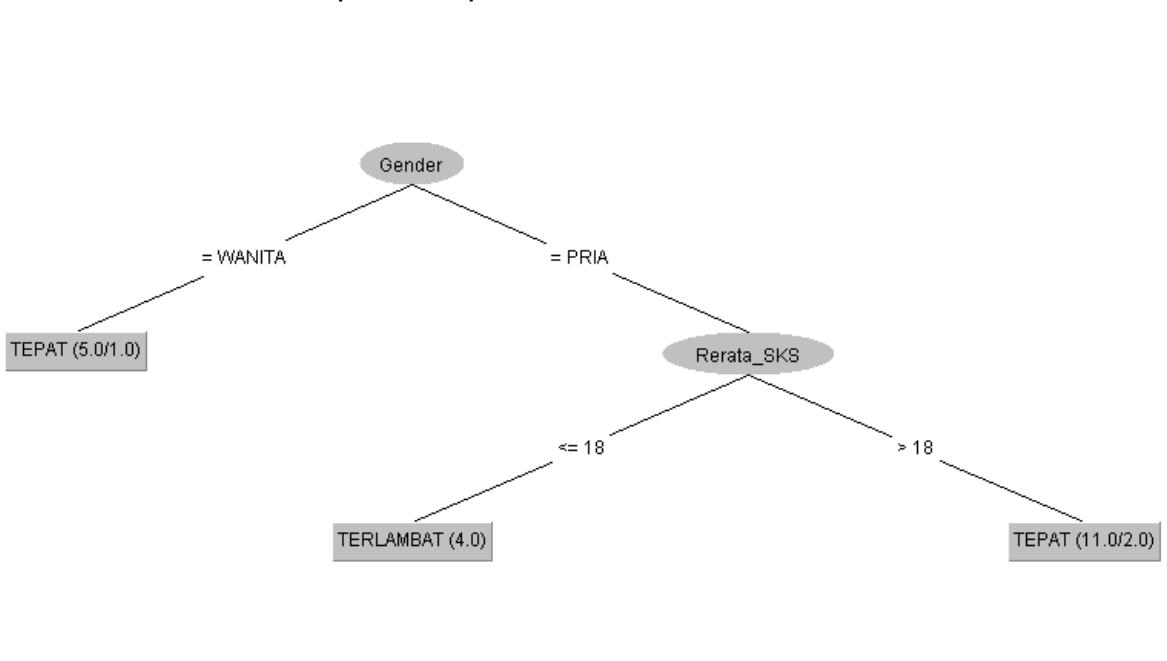
KLASIFIKASI : DECISION TREE

❖ Tugas

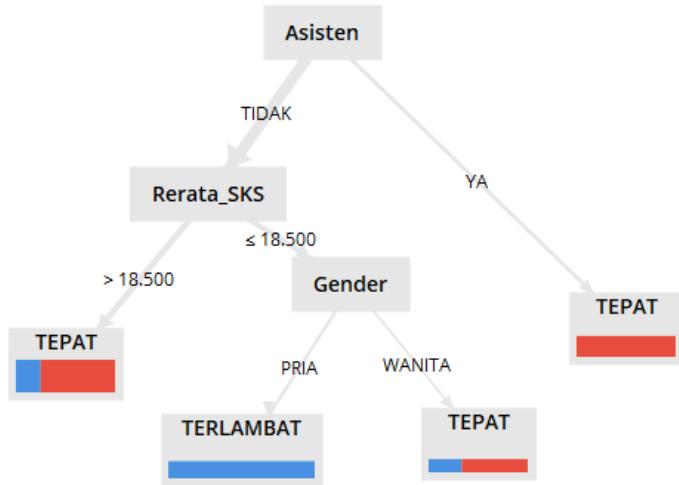
1. Berdasarkan pohon keputusan pada kegiatan 9.4.2 (menggunakan RapidMiner), nilai kelas pada atribut Bermain_Tenis pada tabel Testing sebagai berikut:

| 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 |

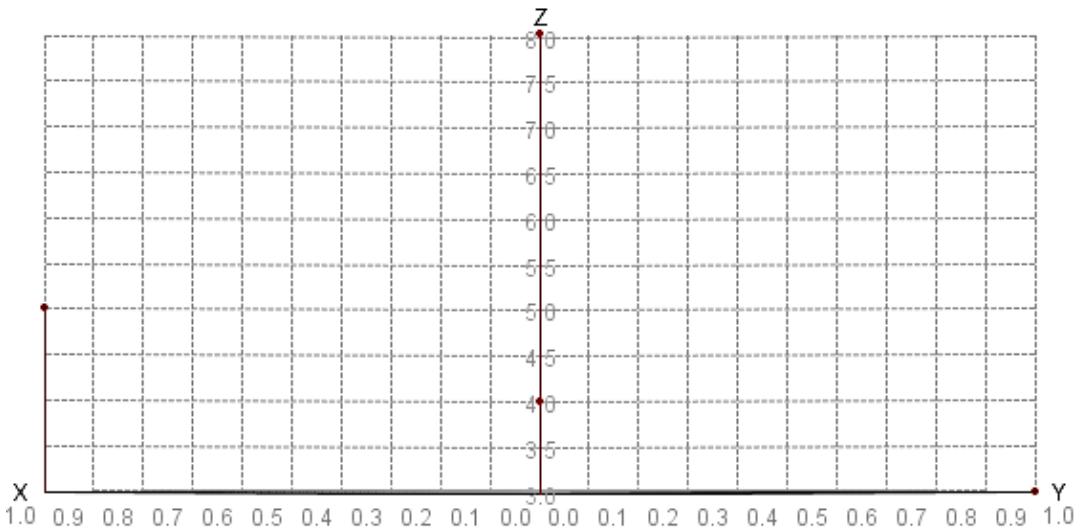
2. Gunakan file ARFF yang di kerjakan pada tugas nomor 1 dalam Modul 7 sebagai data trainig.
 - a) Buatlah dan cetaklah pohon keputusan berdasarkan data tersebut:



- b) Nilai-nilai parameter berikut:
- Jumlah simpul daun pada pohon keputusan = 3
 - Jumlah simpul keseluruhan pada pohon keputusan = 5
 - Waktu yang di butuhkan untuk proses pelatihan = 0.02 detik
 - Tingkat ketepatan klasifikasi = 85%
 - Tingkat ketidaktepatan klasifikasi = 15%
3. Gunakan file excel yang di kerjakan pada tugas nomor 1 dalam Modul 6 sebagai data training.
- a) Buatlah dan cetaklah pohon keputusan berdasarkan data tersebut:



- b) Cetaklah prespektif 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. Berdasarkan pohon keputusan dari soal nomor2, tentukan klasifikasi yang terbentuk berdasarkan kondisinya sesuai dengan simpul-simpulnya:
- Lama_Studi yang akan di capai (TEPAT) jika kondisi sebagai berikut:
 - Sekolah : Gender = Wanita, (nilai atribut lain di abaikan)
 - Sekolah : Gender = Pria, Rerata_SKS > 18, (nilai atribut lain di abaikan)
 - Lama_Studi yang akan di capai (TERLAMBAT) jika kondisi sebagai berikut:
 - Sekolah : Gender = Pria, Rerata_SKS <= 18, (nilai atribut lain di abaikan)

Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

Modul 10

K-MEANS Data Clustering

❖ Langkah – Langkah Percobaan Praktikum

1. Data Tabel_NilaiUjian.xls

| | A | B | C | D |
|----|----------|--------|-------|-------|
| 1 | NO_SISWA | NAMA | B.IND | B.ING |
| 2 | S-101 | JOKO | 8,54 | 8,4 |
| 3 | S-102 | AGUS | 9,98 | 6,81 |
| 4 | S-103 | SUSI | 6,2 | 9,15 |
| 5 | S-104 | DYAH | 5,24 | 7,26 |
| 6 | S-105 | WATI | 5,7 | 5,71 |
| 7 | S-106 | IKA | 8,57 | 5,87 |
| 8 | S-107 | EKO | 7,7 | 7,71 |
| 9 | S-108 | YANTO | 6,6 | 5,7 |
| 10 | S-109 | WAWAN | 9 | 8,12 |
| 11 | S-110 | MAHMUD | 9,81 | 9,58 |
| 12 | | | | |

2. Penggunaan Data pada Table_NilaiUjian.xls untuk proses Clustering, dan hanya menggunakan 3 kolom yaitu

- Nama Siswa (NAMA)
- Nilai Bahasa Indonesia (B.IND)
- Nilai Bahasa Inggris (B.ING)

Import Data - Select the cells to import.

X

Select the cells to import.

Sheet: k-means ▾ Cell range: B:D Select All Define header row: 1

| | A | B | C | D |
|----|----------|--------|-------|-------|
| 1 | NO_SISWA | NAMA | B.IND | B.ING |
| 2 | S-101 | JOKO | 8.540 | 8.400 |
| 3 | S-102 | AGUS | 9.980 | 6.810 |
| 4 | S-103 | SUSI | 6.200 | 9.150 |
| 5 | S-104 | DYAH | 5.240 | 7.260 |
| 6 | S-105 | WATI | 5.700 | 5.710 |
| 7 | S-106 | IKA | 8.570 | 5.870 |
| 8 | S-107 | EKO | 7.700 | 7.710 |
| 9 | S-108 | YANTO | 6.600 | 5.700 |
| 10 | S-109 | WAWAN | 9.000 | 8.120 |
| 11 | S-110 | MAHMUD | 9.810 | 9.580 |

← Previous

→ Next

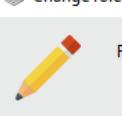
X Cancel

3. Mengubah Column NAMA menjadi id dengan label

Format your columns.

Replace errors with missing values (i)

| | NAMA polynominal | B.IND real | B.ING real |
|----|---------------------|---------------|---------------|
| 1 | JOKO | | |
| 2 | AGUS | | |
| 3 | SUSI | | |
| 4 | DYAH | | |
| 5 | WATI | | |
| 6 | IKA | | |
| 7 | EKO | | |
| 8 | YANTO | 0.000 | 0.700 |
| 9 | WAWAN | 9.000 | 8.120 |
| 10 | MAHMUD | 9.810 | 9.580 |

 Change role

Please enter the new role:

OK Cancel

 no problems.

Previous Next Cancel

4. Save data yang sudah di berikan label untuk proses Clustering

Import Data - Format your columns.

Format your columns.

Replace errors with missing values ⓘ

| | NAMA <i>polynomial_id</i> | B.IND <i>real</i> | B.ING <i>real</i> |
|----|------------------------------|----------------------|----------------------|
| 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 | 7.700 5.700 |
| 9 | WAWAN | 9.000 | 8.120 |
| 10 | MAHMUD | 9.810 | 9.580 |

✓ no problems.

← Previous Next → ✖ Cancel

Save data

Import Data - Where to store the data?

Where to store the data?

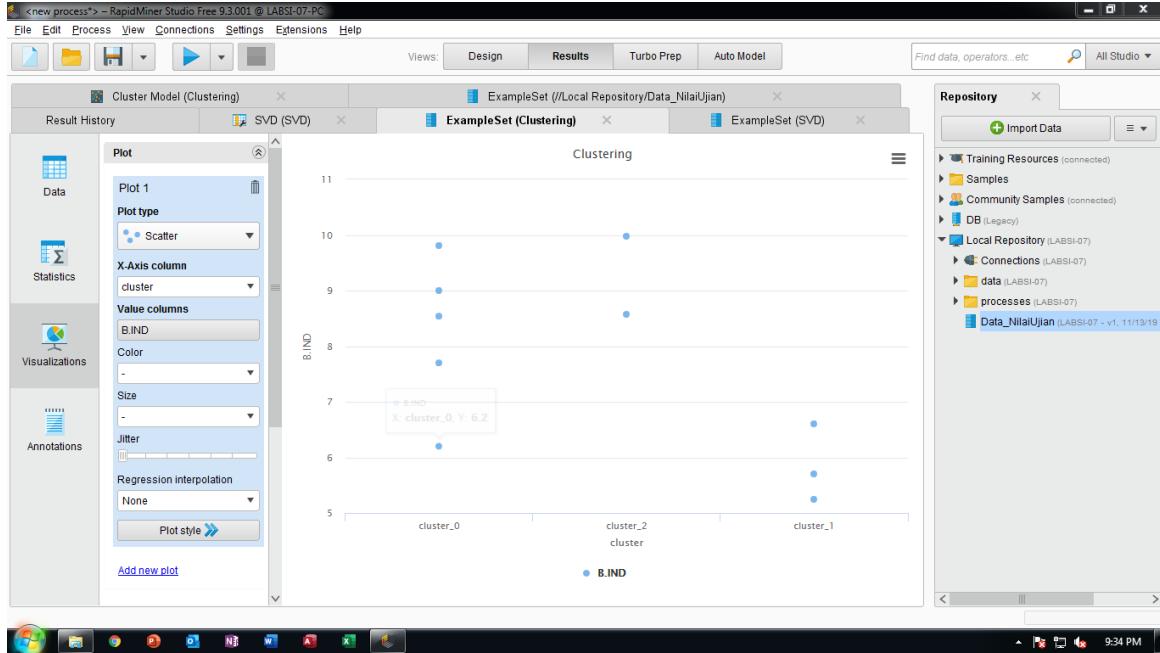
- ▼ Local Repository (OKI KUS MAHESA)
 - ▶ data (OKI KUS MAHESA)
 - ▶ processes (OKI KUS MAHESA)
 - DataCuaca_Training (OKI KUS MAHESA - v1, 11/7/19 9:44 PM - 527 bytes)
 - DataSekolah_Testing (OKI KUS MAHESA - v1, 10/11/19 9:45 PM - 485 bytes)
 - DataSekolah_Testing_Dewi (OKI KUS MAHESA - v1, 10/11/19 10:11 PM - 415 bytes)
 - DataSekolah_Testing_Jono (OKI KUS MAHESA - v1, 10/11/19 10:28 PM - 416 bytes)
 - DataSekolah_Training (OKI KUS MAHESA - v1, 10/11/19 9:44 PM - 675 bytes)

Name

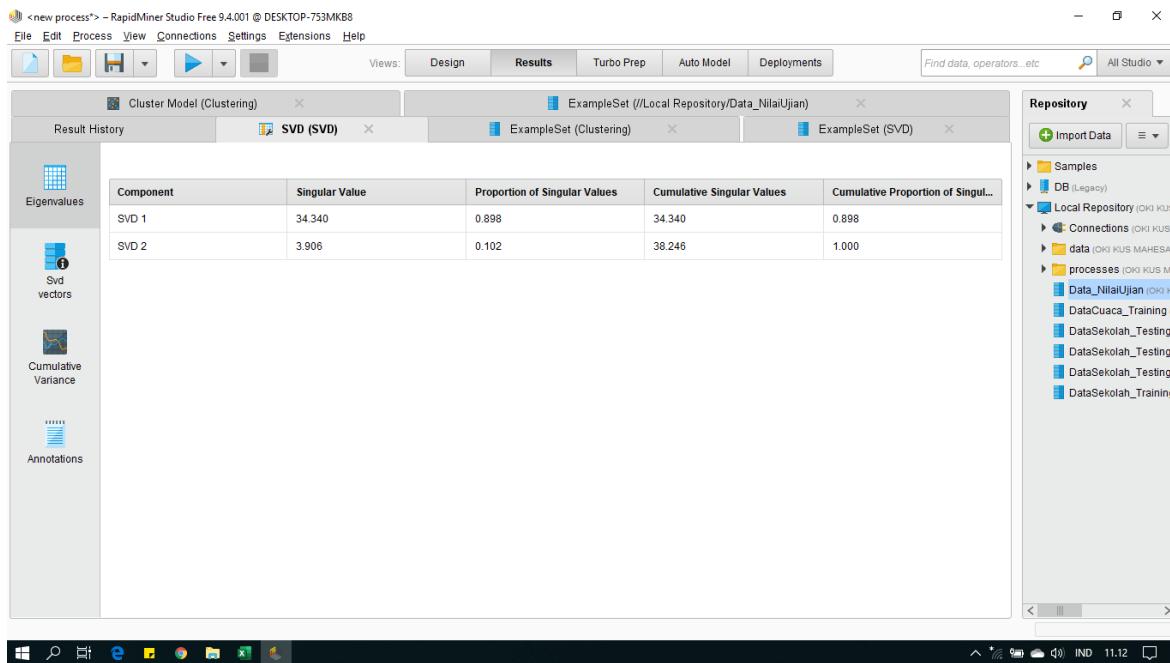
Location //Local Repository/Data_NilaiUjian

← Previous Finish ✓ ✖ Cancel

5. Hasil Proses Yang DI Lakukan dengan keluaran nilai Plot Simpul Clustering



6. Data Penyebaran SVD



7. Exempel Set K-MEANS dengan mode pengurutan cluster(0,1,2) untuk menentukan kelompok siswa Cluster yang di calonkan untuk maju ke olimpiade mata pelajaran berdasarkan nilai tertinggi ujian.

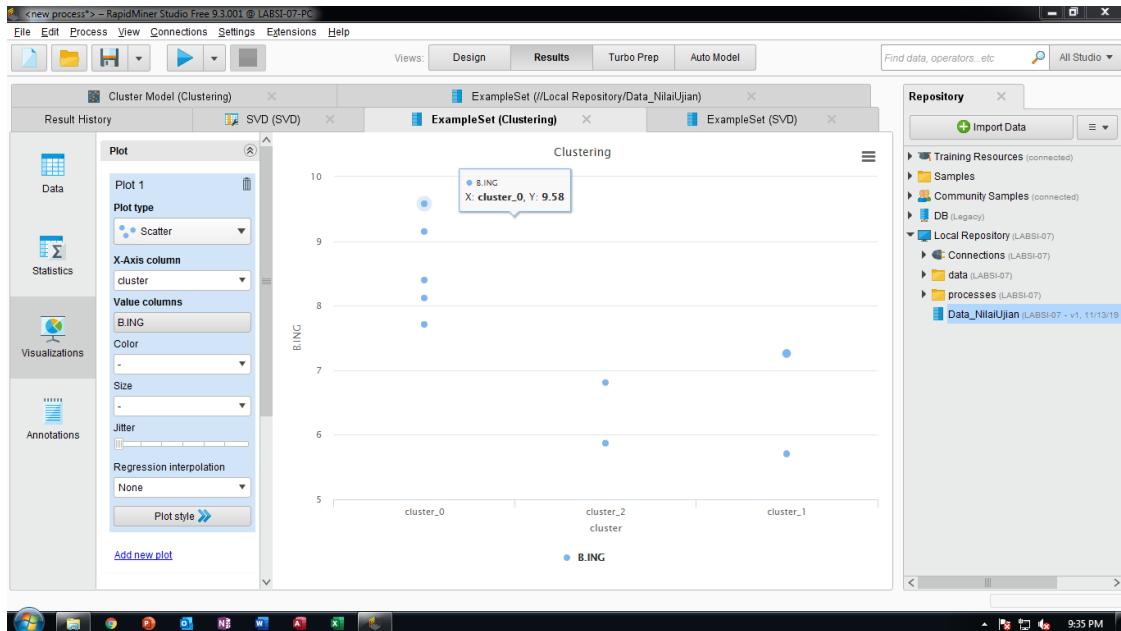
The screenshot shows the RapidMiner Studio interface with the following details:

- Title Bar:** <new process> - RapidMiner Studio Free 9.3.001 @ LABSI-07-PC
- Menu Bar:** File Edit Process View Connections Settings Extensions Help
- Toolbar:** Standard icons for Open, Save, Run, Stop, etc.
- Views:** Design, Results, Turbo Prep, Auto Model
- Search Bar:** Find data, operators...etc
- Repository:** Local Repository (LABSI-07) containing Training Resources, Samples, Community Samples, DB (Legacy), and Data_NilaiUjian.
- Data View:** Shows a table with columns Row No., NAMA, cluster ↑, BJND, and BJNG. The data is as follows:

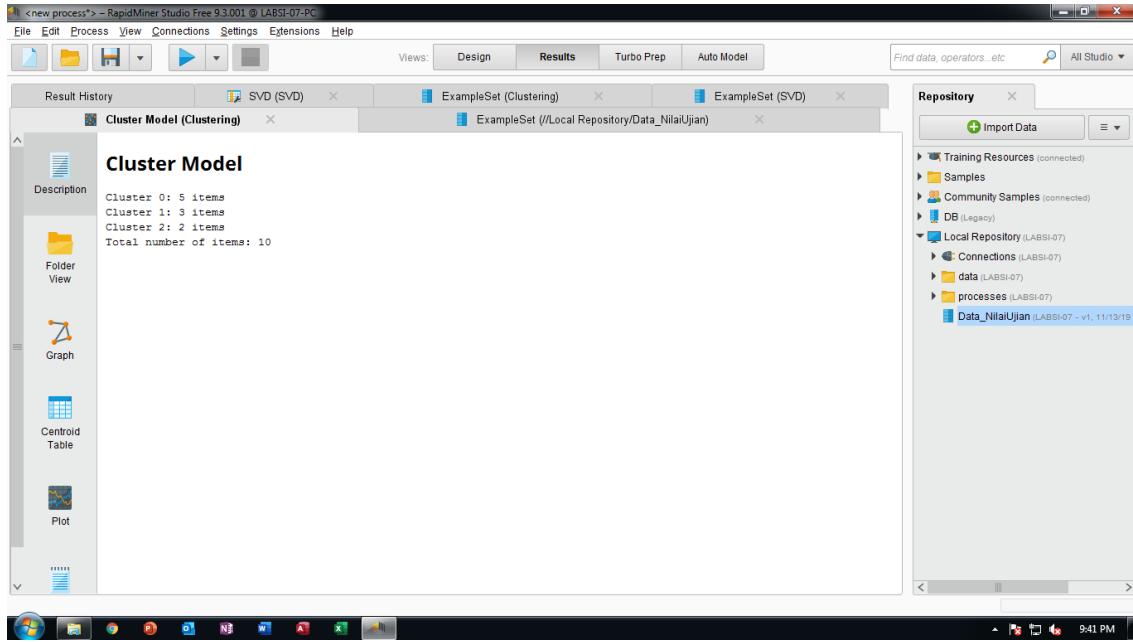
| Row No. | NAMA | cluster ↑ | BJND | BJNG |
|---------|--------|-----------|-------|-------|
| 1 | JOKO | cluster_0 | 8.540 | 8.400 |
| 3 | SUSI | cluster_0 | 6.200 | 9.150 |
| 7 | EKO | cluster_0 | 7.700 | 7.710 |
| 9 | WAWAN | cluster_0 | 9 | 8.120 |
| 10 | MAHMUD | cluster_0 | 9.810 | 9.580 |
| 4 | 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 |

- Message:** ExampleSet (10 examples, 2 special attributes, 2 regular attributes)

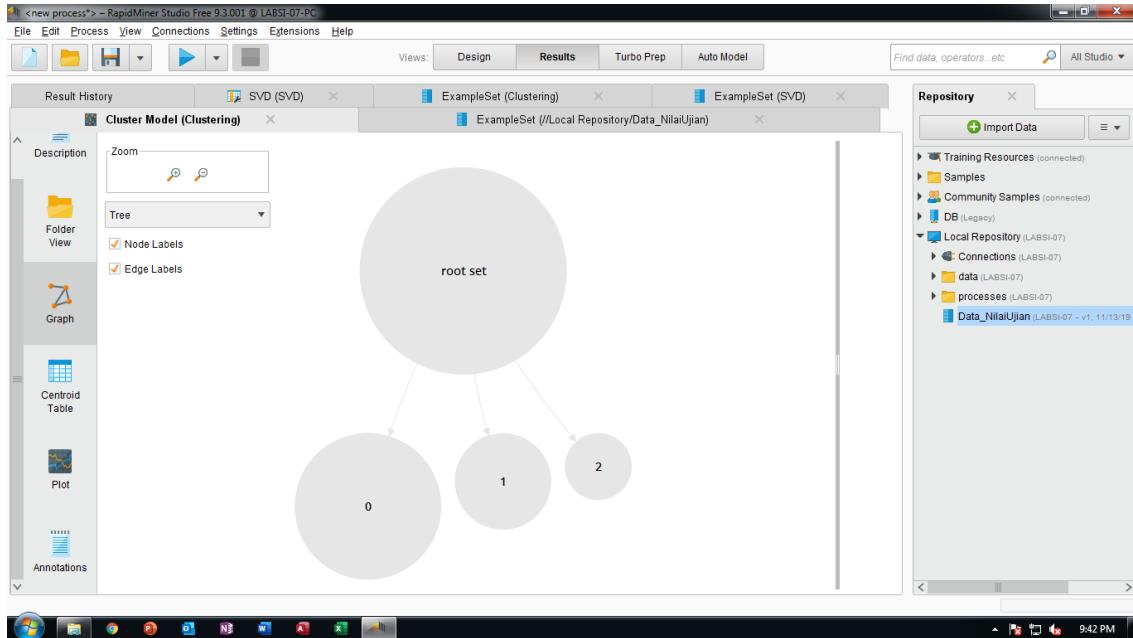
8. Exempel Set K-MEANS dengan mode plot view menggunakan grafik Scetter untuk menentukan kelompok siswa Cluster yang di calonkan untuk maju ke olimpiade mata pelajaran berdasarkan nilai tertinggi ujian.



9. Cluster Model , ini adalah total items yang di hasilkan



10. Graph Hasil Model



11. Hasil akhir dari metode K-MEANS Clustering, disini kita bisa perhatikan, pengelompokan data yang sekarang yaitu dengan pengurutan Cluseter(0,1,2)

| | A | B | C | D | E |
|----|----------|--------|-----------|-------|-------|
| 1 | NO_SISWA | NAMA | CLUSTER | B.IND | B.ING |
| 2 | S-101 | JOKO | cluster_0 | 8.54 | 8.4 |
| 3 | S-103 | SUSI | cluster_0 | 6.2 | 9.15 |
| 4 | S-107 | EKO | cluster_0 | 7.7 | 7.71 |
| 5 | S-109 | WAWAN | cluster_0 | 9 | 8.12 |
| 6 | S-110 | MAHMUD | cluster_0 | 9.81 | 9.58 |
| 7 | S-104 | DYAH | cluster_1 | 5.24 | 7.26 |
| 8 | S-105 | WATI | cluster_1 | 5.7 | 5.71 |
| 9 | S-108 | YANTO | cluster_1 | 6.6 | 5.7 |
| 10 | S-102 | AGUS | cluster_2 | 9.98 | 6.81 |
| 11 | S-106 | IKA | cluster_2 | 8.57 | 5.87 |

Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

Modul 10

K -MEANS Data Clustering

❖ Tugas Praktikum

1. TABEL DATA NILAI UJIAN 30 SISWA , melakukan rumus random untuk mendapatkan nilai random percobaan dengan $=5+RAND()*5$, dan lakukan 2 angka di belakang koma dengan, formula number

| | A | B | C | D | E | F |
|----|----------|--------|-------------|-------|-----|-----|
| 1 | NO_SISWA | NAMA | B.IND | B.ING | MTK | IPA |
| 2 | S-101 | JOKO | =5+RAND()*5 | | | |
| 3 | S-102 | AGUS | | | | |
| 4 | S-103 | SUSI | | | | |
| 5 | S-104 | DYAH | | | | |
| 6 | S-105 | WATI | | | | |
| 7 | S-106 | IKA | | | | |
| 8 | S-107 | EKO | | | | |
| 9 | S-108 | YANTO | | | | |
| 10 | S-109 | WAWAN | | | | |
| 11 | S-110 | MAHMUD | | | | |
| 12 | S-111 | BUDI | | | | |
| 13 | S-112 | SANTI | | | | |
| 14 | S-113 | DIAN | | | | |

2. Hasil sesudah melakukan number random method

| NO_SISWA | NAMA | B.IND | B.ING | MTK | IPA |
|-----------------|-------------|--------------|--------------|------------|------------|
| S-101 | JOKO | 6.90 | 6.99 | 8.20 | 5.32 |
| S-102 | AGUS | 8.18 | 6.04 | 8.76 | 5.18 |
| S-103 | SUSI | 6.22 | 6.17 | 9.46 | 8.49 |
| S-104 | DYAH | 9.74 | 8.49 | 6.81 | 7.32 |
| S-105 | WATI | 5.17 | 6.88 | 9.52 | 5.80 |
| S-106 | IKA | 7.75 | 10.00 | 5.40 | 6.35 |
| S-107 | EKO | 8.00 | 5.14 | 9.24 | 7.49 |
| S-108 | YANTO | 9.28 | 8.90 | 7.92 | 8.03 |
| S-109 | WAWAN | 9.11 | 5.93 | 8.65 | 6.56 |
| S-110 | MAHMUD | 8.04 | 5.10 | 9.02 | 5.78 |
| S-111 | BUDI | 8.94 | 8.96 | 9.44 | 9.79 |
| S-112 | SANTI | 9.35 | 7.08 | 9.54 | 7.96 |
| S-113 | DIAN | 6.57 | 7.09 | 8.80 | 8.00 |
| S-114 | DANI | 6.41 | 7.46 | 8.29 | 9.49 |
| S-115 | AHMAD | 7.37 | 9.68 | 7.35 | 8.99 |
| S-116 | BAYU | 5.49 | 8.05 | 7.83 | 5.68 |
| S-117 | RISA | 7.85 | 7.47 | 6.99 | 8.34 |
| S-118 | RANI | 8.40 | 8.25 | 7.92 | 6.67 |
| S-119 | YANI | 5.19 | 7.49 | 5.24 | 7.40 |
| S-120 | RATIH | 5.90 | 7.07 | 8.58 | 6.98 |
| S-121 | INDAH | 7.85 | 9.53 | 8.83 | 6.38 |
| S-122 | JONO | 6.28 | 9.08 | 6.75 | 8.03 |
| S-123 | SARAH | 9.33 | 9.72 | 5.07 | 8.03 |
| S-124 | RAMA | 8.55 | 8.00 | 5.20 | 8.48 |
| S-125 | BAMBANG | 7.23 | 5.83 | 9.70 | 5.55 |
| S-126 | HADI | 7.99 | 6.82 | 6.85 | 7.61 |
| S-127 | NANA | 7.62 | 5.86 | 6.45 | 6.59 |
| S-128 | FEBRI | 8.35 | 7.49 | 9.48 | 9.45 |
| S-129 | DENI | 8.02 | 7.91 | 7.67 | 5.69 |
| S-130 | TONI | 6.77 | 9.78 | 8.59 | 6.97 |

3. Melakukan label di nama dengan label id dan type polynomial

Import Data - Format your columns.

Format your columns.

Replace errors with missing values ⓘ

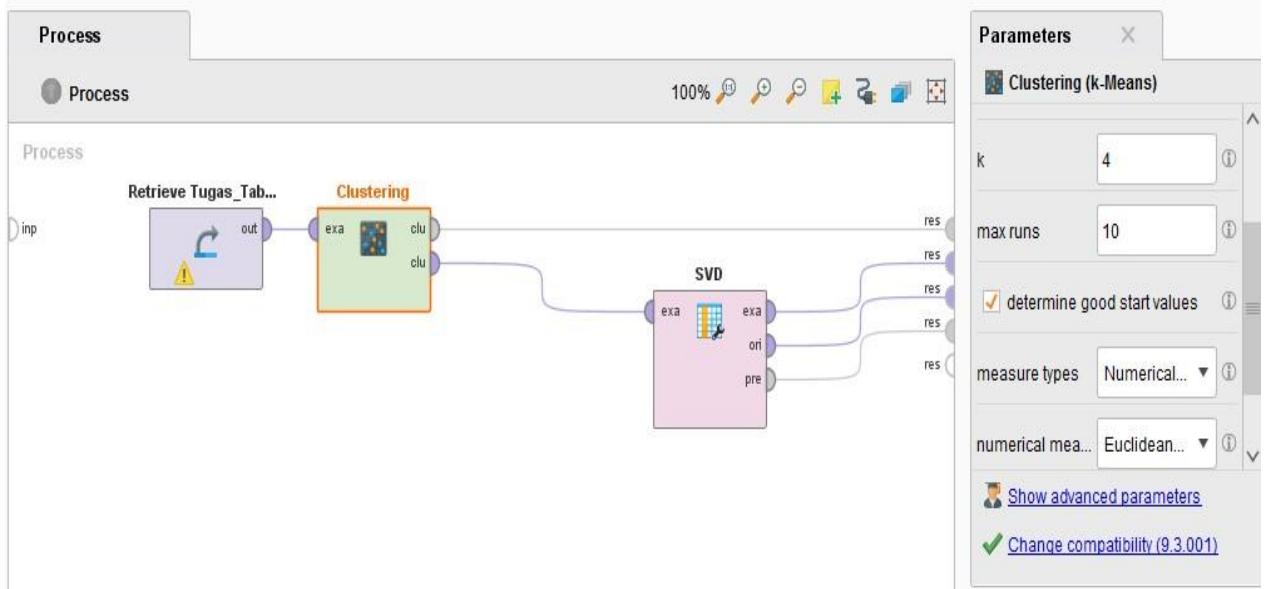
| | NAMA polynomial | B.IND real | B.ING real | MTK real | IPA real |
|----|--------------------|---------------|---------------|-------------|-------------|
| 1 | JOKO | | | | |
| 2 | AGUS | | | | |
| 3 | SUSI | | | | |
| 4 | DYAH | | | | |
| 5 | WATI | | | | |
| 6 | IKA | | | | |
| 7 | EKO | | | | |
| 8 | YANTO | | | | |
| 9 | WAWAN | 9.109 | 5.935 | 8.650 | 6.561 |
| 10 | MAHMUD | 8.035 | 5.104 | 9.016 | 5.782 |
| 11 | BUDI | 8.939 | 8.955 | 9.443 | 9.786 |
| 12 | SANTI | 9.351 | 7.078 | 9.537 | 7.961 |
| 13 | DIAN | 6.570 | 7.094 | 8.802 | 8.003 |

Change role

Please enter the new role:

no problems.

4. Melakukan proses Clustering K-MEANS



5. PENGELOMPOKAN SISWA SESUAI CLUSTER

CLUSTER 0

| NAMA | CLUSTER | B.IND | B.ING | MTK | IPA |
|-------|-----------|-------|-------|------|------|
| DYAH | cluster_0 | 9.74 | 8.49 | 6.81 | 7.32 |
| IKA | cluster_0 | 7.75 | 10.00 | 5.40 | 6.35 |
| YANTO | cluster_0 | 9.28 | 8.90 | 7.92 | 8.03 |
| AHMAD | cluster_0 | 7.37 | 9.68 | 7.35 | 8.99 |
| RISA | cluster_0 | 7.85 | 7.47 | 6.99 | 8.34 |
| RANI | cluster_0 | 8.40 | 8.25 | 7.92 | 6.67 |
| YANI | cluster_0 | 5.19 | 7.49 | 5.24 | 7.40 |
| JONO | cluster_0 | 6.28 | 9.08 | 6.75 | 8.03 |
| SARAH | cluster_0 | 9.33 | 9.72 | 5.07 | 8.03 |
| RAMA | cluster_0 | 8.55 | 8.00 | 5.20 | 8.48 |
| HADI | cluster_0 | 7.99 | 6.82 | 6.85 | 7.61 |

CLUSTER 1

| NAMA | CLUSTER | B.IND | B.ING | MTK | IPA |
|------|-----------|-------|-------|------|------|
| JOKO | cluster_1 | 6.90 | 6.99 | 8.20 | 5.32 |
| WATI | cluster_1 | 5.17 | 6.88 | 9.52 | 5.80 |
| BAYU | cluster_1 | 5.49 | 8.05 | 7.83 | 5.68 |

| | | | | | |
|-------|-----------|------|------|------|------|
| RATIH | cluster_1 | 5.90 | 7.07 | 8.58 | 6.98 |
| INDAH | cluster_1 | 7.85 | 9.53 | 8.83 | 6.38 |
| DENI | cluster_1 | 8.02 | 7.91 | 7.67 | 5.69 |
| TONI | cluster_1 | 6.77 | 9.78 | 8.59 | 6.97 |

CLUSTER 2

| NAMA | CLUSTER | B.IND | B.ING | MTK | IPA |
|-------------|----------------|--------------|--------------|------------|------------|
| SUSI | cluster_2 | 6.22 | 6.17 | 9.46 | 8.49 |
| BUDI | cluster_2 | 8.94 | 8.96 | 9.44 | 9.79 |
| DIAN | cluster_2 | 6.57 | 7.09 | 8.80 | 8.00 |
| DANI | cluster_2 | 6.41 | 7.46 | 8.29 | 9.49 |
| FEBRI | cluster_2 | 8.35 | 7.49 | 9.48 | 9.45 |

CLUSTER 3

| NAMA | CLUSTER | B.IND | B.ING | MTK | IPA |
|---------|-----------|-------|-------|------|------|
| AGUS | cluster_3 | 8.18 | 6.04 | 8.76 | 5.18 |
| EKO | cluster_3 | 8.00 | 5.14 | 9.24 | 7.49 |
| WAWAN | cluster_3 | 9.11 | 5.93 | 8.65 | 6.56 |
| MAHMUD | cluster_3 | 8.04 | 5.10 | 9.02 | 5.78 |
| SANTI | cluster_3 | 9.35 | 7.08 | 9.54 | 7.96 |
| BAMBANG | cluster_3 | 7.23 | 5.83 | 9.70 | 5.55 |
| NANA | cluster_3 | 7.62 | 5.86 | 6.45 | 6.59 |

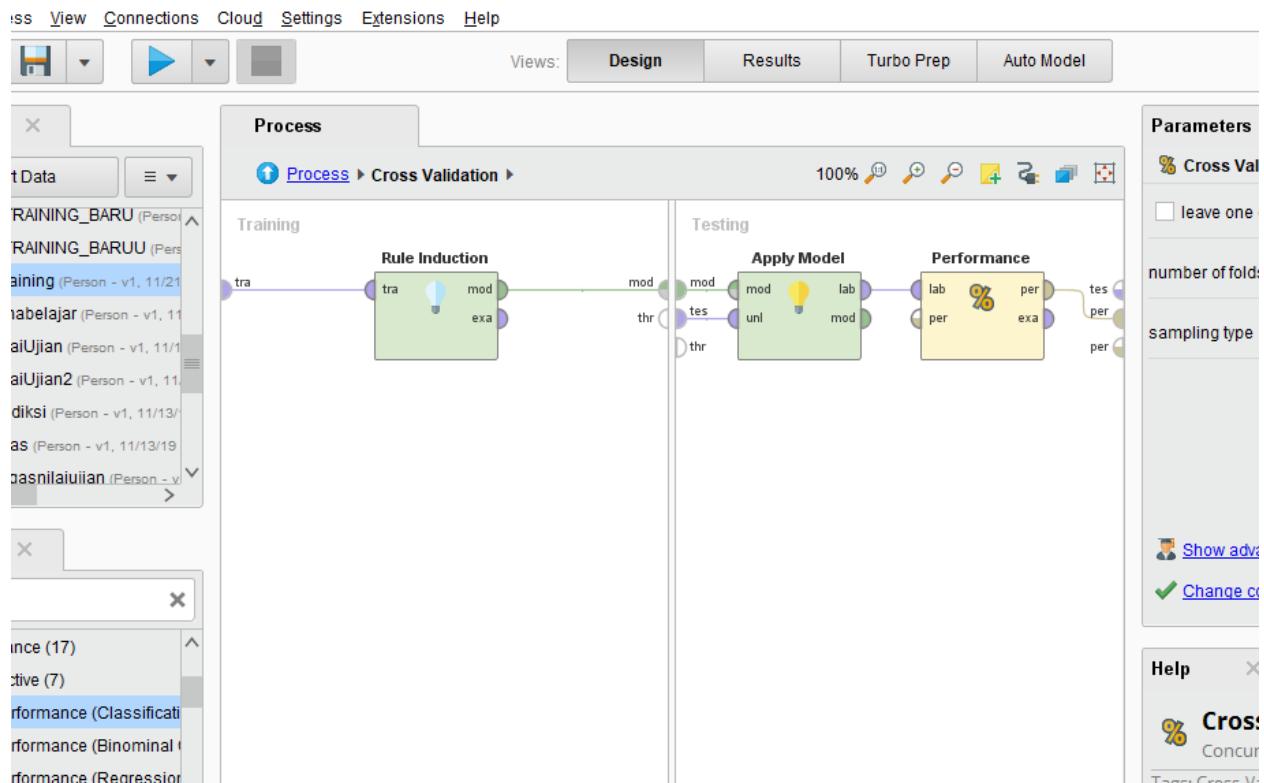
Nama : Oki Kus Mahesa
Nim : L200170064
Kelas : C

Aturan Assosiasi

❖ Langkah-langkah Praktikum

1. Aturan Asosiasi dengan Menggunakan DataCuaca_Traning

- RapidMiner Studio Free 9.0.003 @ DESKTOP-7UM84PR



2. Hasil PerformanceVector (Performance)

The screenshot shows the RapidMiner Studio interface with the 'Results' tab selected. A 'RuleModel (Rule Induction)' process is running on an 'ExampleSet (/Local Repository/CuacaTraining)'. The 'PerformanceVector (Performance)' view is open, showing the following results:

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

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

3. Rule Model kondisi pad kondisi yang ada

The screenshot shows the RapidMiner Studio interface with the 'Results' tab selected. A 'RuleModel (Rule Induction)' process is running on an 'ExampleSet (/Local Repository/CuacaTraining)'. The 'RuleModel' view is open, displaying the following rules:

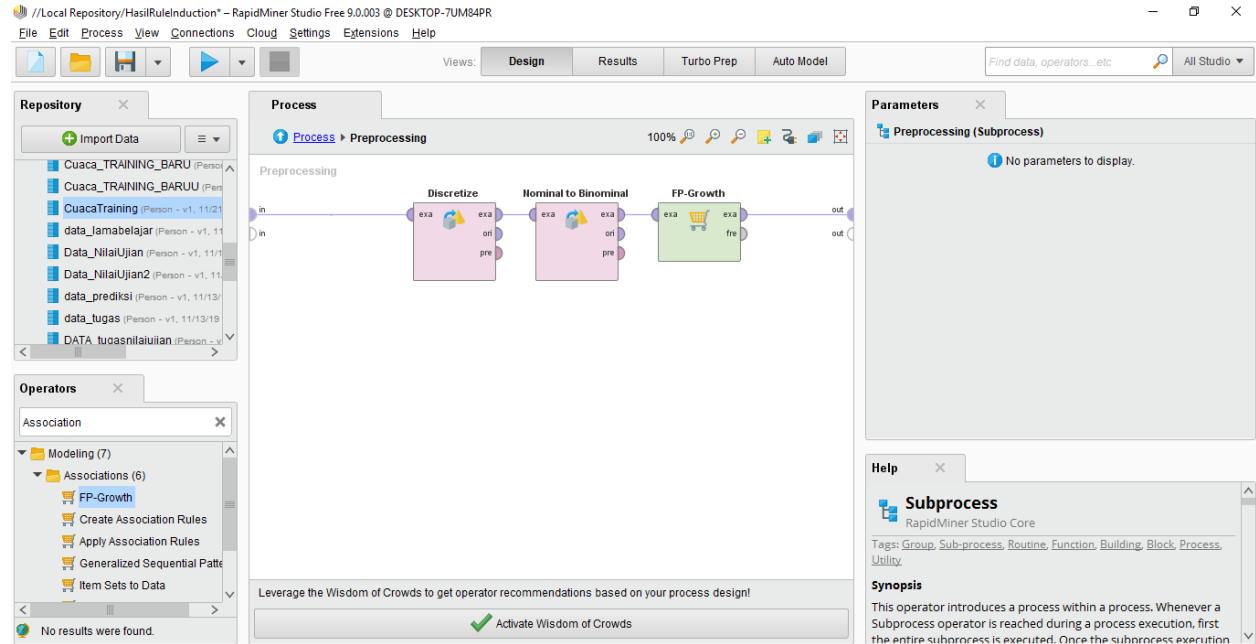
```
if Kelembaban_Udara ≤ 82.500 then Ya (1 / 6)
if Cuaca = Cerah then Tidak (3 / 0)
if Cuaca = Mendung then Ya (0 / 2)
if Suhu ≤ 70.500 then Ya (0 / 1)
else Tidak (0 / 0)

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

The 'Repository' panel on the right lists various datasets and models, including:

- Cuaca333 (Person - v1, 11/21/19)
- Cuaca_TESTING (Person - v1, 11/21/19)
- Cuaca_TRAINING (Person - v1, 11/21/19)
- Cuaca_TRAINING222 (Person - v1, 11/21/19)
- Cuaca_TRAINING_BARU (Person - v1, 11/21/19)
- Cuaca_TRAINING_BARUU (Person - v1, 11/21/19)
- CuacaTraining (Person - v1, 11/21/19)
- data_Jamabelajar (Person - v1, 11/21/19)
- Data_NilaiUjian (Person - v1, 11/21/19)
- Data_NilaiUjian2 (Person - v1, 11/21/19)
- data_prediksi (Person - v1, 11/13/19)
- data_tugas (Person - v1, 11/13/19)
- DATA_tugasnilaiujian (Person - v1, 11/21/19)
- data_tugasprediksi (Person - v1, 11/21/19)
- DataCuaca_Training (Person - v1, 11/21/19)
- L200170107_baru (Person - v1, 11/21/19)
- MahasiswaTraining (Person - v1, 11/21/19)
- Tabel_Cuaca_Testing (Person - v1, 11/21/19)
- Tabel_NilaiUjian (Person - v1, 11/21/19)
- Tugas (Person - v1, 11/21/19 4:39)
- Tugas_NilaiUjian (Person - v1, 11/21/19)

4. proses Preprocessing yang terjadi



Find data, operators...etc All Studio ▾

Parameters

FP-Growth

input format: items in dummy coded colu...

min requirement: support

min support: 0.1

min items per itemset: 1

max items per itemset: 0

max number of itemsets: 1000000

[Show advanced parameters](#)

Help

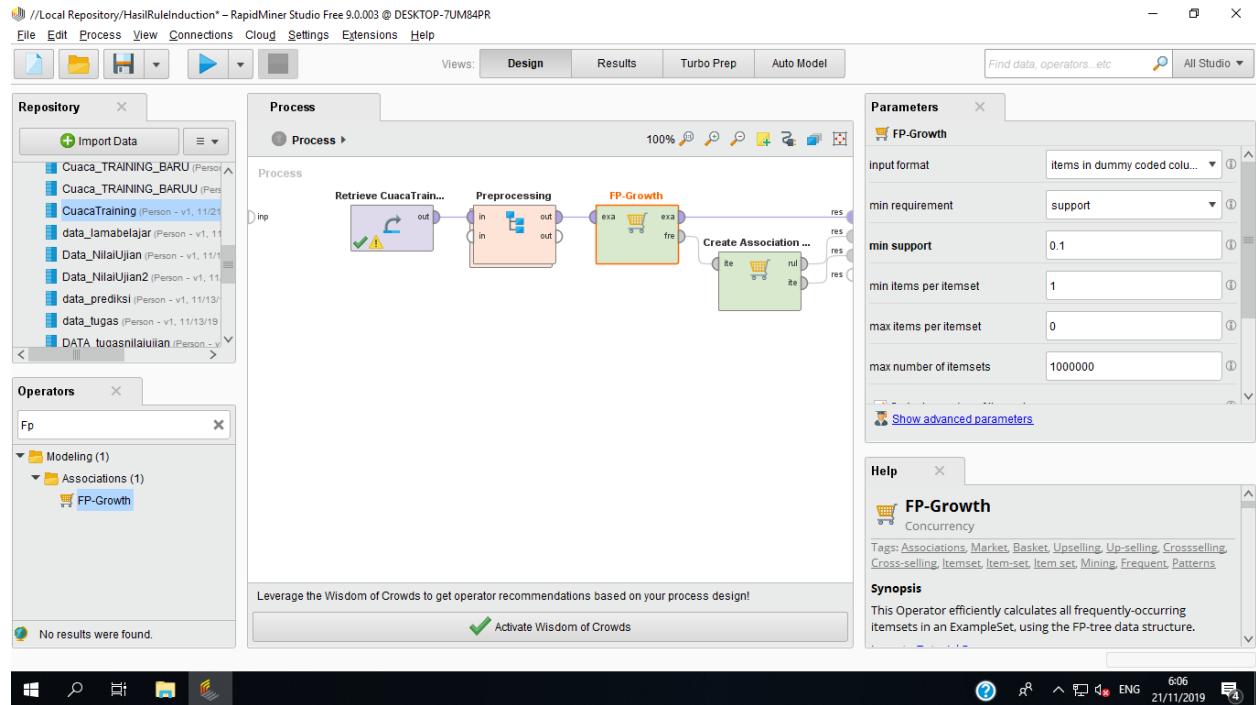
FP-Growth
Concurrency

Tags: [Associations](#), [Market](#), [Basket](#), [Upselling](#), [Up-selling](#), [Crossselling](#), [Cross-selling](#), [Itemset](#), [Item-set](#), [Item set](#), [Mining](#), [Frequent](#), [Patterns](#)

Synopsis

This Operator efficiently calculates all frequently-occurring itemsets in an ExampleSet, using the FP-tree data structure.

5. Proses Association Rules dengan min support = 0.1



6. FrequentItemSets

The screenshot shows the "FrequentItemSets (FP-Growth)" tab in the RapidMiner interface. The table displays the following data:

| | No. of Sets: 25 | Total Max. Size: 4 | Size | Support | Item 1 | Item 2 | Item 3 | Item 4 |
|-------------|-----------------|--------------------|------|---------|------------------|-----------------|-----------------|--------|
| Data | | | 2 | 0.214 | Kelembaban_Udara | Cuaca = Cerah | | |
| Annotations | | | 2 | 0.143 | Kelembaban_Udara | Cuaca = Hujan | | |
| | | | 2 | 0.143 | Kelembaban_Udara | Cuaca = Mendung | | |
| | | | 2 | 0.143 | Suhu | Berangin | | |
| | | | 2 | 0.214 | Suhu | Cuaca = Cerah | | |
| | | | 2 | 0.071 | Suhu | Cuaca = Hujan | | |
| | | | 2 | 0.143 | Suhu | Cuaca = Mendung | | |
| | | | 2 | 0.143 | Berangin | Cuaca = Cerah | | |
| | | | 2 | 0.071 | Berangin | Cuaca = Hujan | | |
| | | | 2 | 0.143 | Berangin | Cuaca = Mendung | | |
| | | | 3 | 0.071 | Kelembaban_Udara | Suhu | Berangin | |
| | | | 3 | 0.143 | Kelembaban_Udara | Suhu | Cuaca = Cerah | |
| | | | 3 | 0.071 | Kelembaban_Udara | Suhu | Cuaca = Mendung | |
| | | | 3 | 0.071 | Kelembaban_Udara | Berangin | Cuaca = Cerah | |
| | | | 3 | 0.071 | Kelembaban_Udara | Berangin | Cuaca = Mendung | |

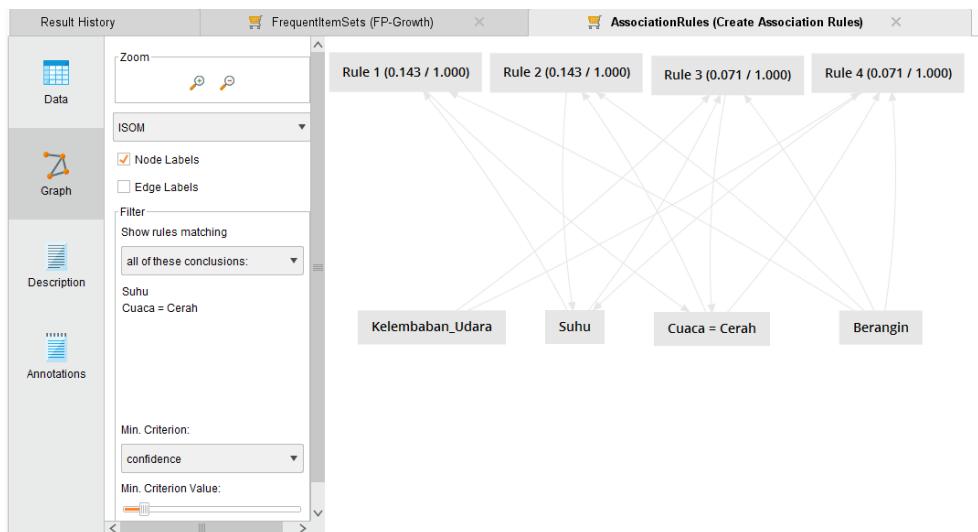
Min Size = 1

Max Size = 4

7. Create Association Rules

| No. | Premises | Conclusion | Support |
|-----|---|---------------|---------|
| 1 | Suhu, Berangin | Cuaca = Cerah | 0.143 |
| 2 | Berangin, Cuaca = Cerah | Suhu | 0.143 |
| 3 | Kelembaban_Udara, Suhu, Berangin | Cuaca = Cerah | 0.071 |
| 4 | Kelembaban_Udara, Berangin, Cuaca = Cerah | Suhu | 0.071 |

Penyebaran Data



Support Confidence LaPlace Gain p-s Lift Convictive\

Berdasarkan Nilai dari penyebaran data saat di prosess dengan gain information

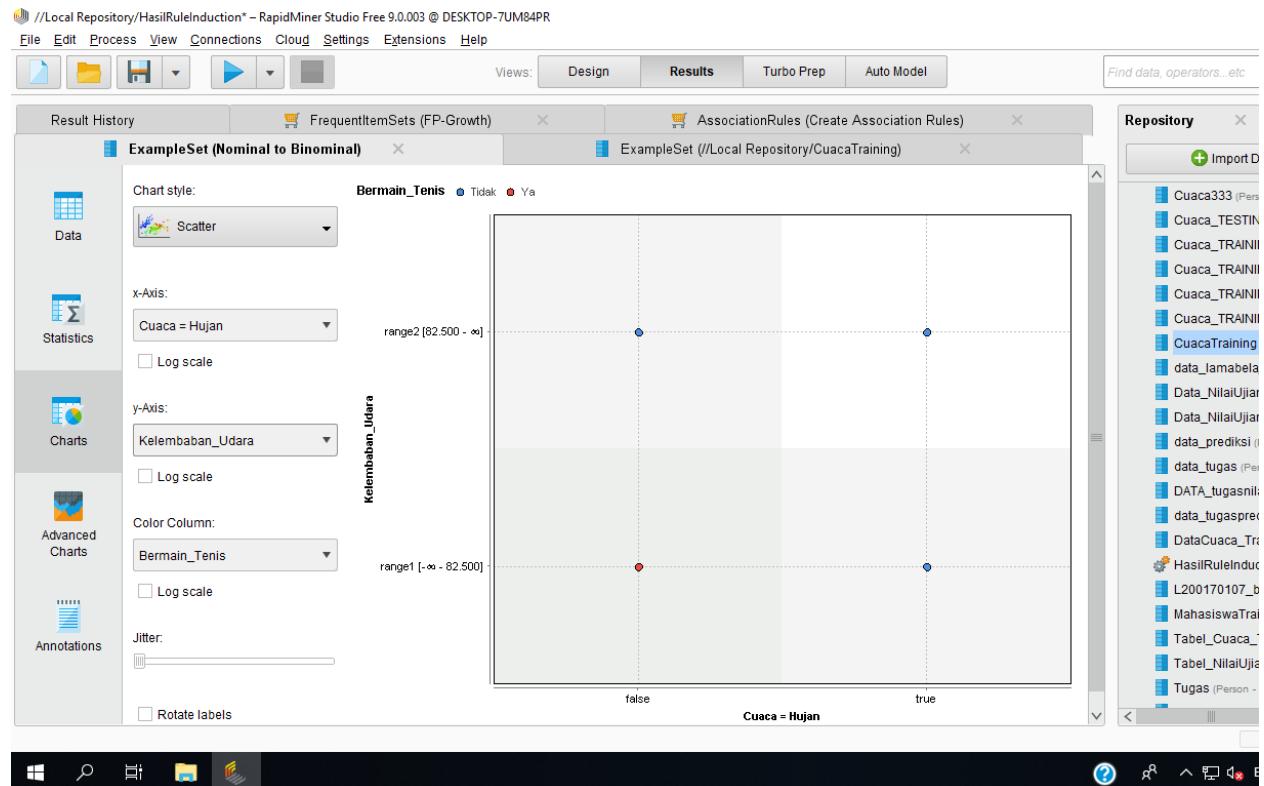
| | Support | Confidence | LaPlace | Gain | p-s | Lift | Convicti... |
|--|---------|------------|---------|--------|-------|-------|-------------|
| | 0.143 | 1 | 1 | -0.143 | 0.092 | 2.800 | ∞ |
| | 0.143 | 1 | 1 | -0.143 | 0.082 | 2.333 | ∞ |
| | 0.071 | 1 | 1 | -0.071 | 0.046 | 2.800 | ∞ |
| | 0.071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

Chart Style Exemple set Nominal to Binominal dengan:"

x-Axis adalah: Cuaca = Hujan

y-Axis adalah : Kelembaban_Udara

Column Index Bermain_Tenis



Nama : Oki Kus Mahesa
Nim : L200170064
Kelas : C

Aturan Assosiasi

❖ Tugas

1. Menggunakan data Lama Studi dengan format DataLamaStudi_Training

The screenshot shows the RapidMiner interface with the following details:

- File Menu:** File, Edit, Process, View, Connections, Cloud, Settings, Extensions, Help.
- Toolbars:** Standard toolbar with icons for Open, Save, Run, Stop, and Help.
- Views:** Design, Results, Turbo Prep, Auto Model.
- Search Bar:** Find data, operators...etc.
- Repository:** Shows a list of available datasets including Cuaca333, Cuaca_TESTING, Cuaca_TRAINING, Cuaca_TRAINING222, Cuaca_TRAINING_BARU, CuacaTraining, data_Jamabelajar, Data_NilaiUjian, Data_NilaiUjian2, Data_pendekripsi, data_tugas, DATA_tugasnilaiujian, data_tugasprediksi, DataCuaca_Training, HasilRuleInduction, L200170107_baru, MahasiswaTraining, Modul6Sekolah_Training, Tabel_Cuaca_Testing, and Tabel_NilaiUjian.
- Current Process:** The process consists of three main steps:
 - ExampleSet (//Local Repository/Modul6Sekolah_Training)**: A data source step.
 - AssociationRules (Create Association Rules)**: An association rule mining step.
 - ExampleSet (Nominal to Binomial)**: A data transformation step.
- Result History:** A table showing frequent item sets (FP-Growth) with the following columns: Size, Support, Item 1, Item 2, Item 3, Item 4, and Item 5. The data is as follows:

| No. of Sets: 55 | Size | Support | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
|--------------------|------|---------|------------------|------------------|--------------|--------|--------|
| Total Max. Size: 5 | 2 | 0.250 | Gender | Rerata_SKS | | | |
| Min. Size: 1 | 2 | 0.150 | Gender | Jurusan_SMA =... | | | |
| Max. Size: 5 | 2 | 0.150 | Jurusan_SMA =... | Asal_Sekolah | | | |
| Contains Item: | 2 | 0.200 | Jurusan_SMA =... | Asisten | | | |
| | 2 | 0.100 | Jurusan_SMA =... | Rerata_SKS | | | |
| | 2 | 0.100 | Asal_Sekolah | Jurusan_SMA =... | | | |
| | 2 | 0.150 | Asal_Sekolah | Asisten | | | |
| | 2 | 0.150 | Asal_Sekolah | Rerata_SKS | | | |
| | 2 | 0.050 | Asal_Sekolah | Jurusan_SMA =... | | | |
| | 2 | 0.100 | Jurusan_SMA =... | Rerata_SKS | | | |
| | 2 | 0.150 | Asisten | Rerata_SKS | | | |
| | 2 | 0.050 | Asisten | Jurusan_SMA =... | | | |
| | 2 | 0.050 | Rerata_SKS | Jurusan_SMA =... | | | |
| | 3 | 0.100 | Gender | Jurusan_SMA =... | Asal_Sekolah | | |

2. Pada Operator Descretize by Frequency memiliki nilai: number of bins=2
- a. Jumlah set aturan assosiasi dan total max size yang terbentuk berdasarkan FP-GROWTH (table view) ! Gambarkan Tabelnya:

| | Support | Confidence | LaPlace | Gain | p-s | Lift | Convicti... |
|--|---------|------------|---------|--------|-------|-------|-------------|
| | 0.250 | 0.833 | 0.962 | -0.350 | 0.025 | 1.111 | 1.500 |
| | 0.250 | 0.833 | 0.962 | -0.350 | 0.025 | 1.111 | 1.500 |
| | 0.250 | 1 | 1 | -0.250 | 0.062 | 1.333 | ∞ |
| | 0.100 | 1 | 1 | -0.100 | 0.025 | 1.333 | ∞ |
| | 0.100 | 1 | 1 | -0.100 | 0.025 | 1.333 | ∞ |
| | 0.150 | 1 | 1 | -0.150 | 0.038 | 1.333 | ∞ |
| | 0.050 | 1 | 1 | -0.050 | 0.012 | 1.333 | ∞ |
| | 0.100 | 1 | 1 | -0.100 | 0.025 | 1.333 | ∞ |
| | 0.150 | 1 | 1 | -0.150 | 0.038 | 1.333 | ∞ |
| | 0.050 | 1 | 1 | -0.050 | 0.012 | 1.333 | ∞ |
| | 0.050 | 1 | 1 | -0.050 | 0.012 | 1.333 | ∞ |
| | 0.100 | 1 | 1 | -0.100 | 0.075 | 4 | ∞ |
| | 0.050 | 1 | 1 | -0.050 | 0.038 | 4 | ∞ |
| | 0.050 | 1 | 1 | -0.050 | 0.035 | 3.333 | ∞ |

Jumlah set aturan assosiasi dan total max size yang terbentuk berdasarkan FP-GROWTH (table-view)

//Local Repository/TugasModul6bins2 – RapidMiner Studio Free 9.0.003 @ DESKTOP-7UM84PR

File Edit Process View Connections Cloud Settings Extensions Help

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

ExampleSet (Nominal to Binominal) ExampleSet (/Local Repository/CuacaTraining)

Result History FrequentItemSets (FP-Growth) AssociationRules (Create Association Rules)

No. of Sets: 55 Total Max. Size: 5

Size Support ↓ Item 1 Item 2 Item 3 Item 4 Item 5

Min. Size: 1 Max. Size: 5 Contains Item: Update View

| Size | Support ↓ | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
|------|-----------|--------------------|--------------------|--------|--------|--------|
| 1 | 0.750 | Gender | | | | |
| 1 | 0.500 | Jurusan_SMA = IPA | | | | |
| 2 | 0.350 | Gender | Jurusan_SMA = IPA | | | |
| 1 | 0.300 | Asal_Sekolah | | | | |
| 1 | 0.300 | Jurusan_SMA = IPS | | | | |
| 1 | 0.250 | Asisten | | | | |
| 1 | 0.250 | Rerata_SKS | | | | |
| 2 | 0.250 | Gender | Asal_Sekolah | | | |
| 2 | 0.250 | Gender | Jurusan_SMA = IPS | | | |
| 2 | 0.250 | Gender | Rerata_SKS | | | |
| 1 | 0.200 | Jurusan_SMA = L... | | | | |
| 2 | 0.200 | Gender | Asisten | | | |
| 2 | 0.200 | Jurusan_SMA = IPA | Asisten | | | |
| 2 | 0.150 | Gender | Jurusan_SMA = L... | | | |
| 2 | 0.150 | Jurusan_SMA = IPA | Asal_Sekolah | | | |

Repository Import Data

- Cuaca333 (Person - v1, 1)
- Cuaca_TESTING (Person)
- Cuaca_TRAINING (Person)
- Cuaca_TRAINING222 (Person)
- Cuaca_TRAINING_BAF (Person)
- Cuaca_TRAINING_BAR (Person)
- CuacaTraining (Person - v1)
- data_lamabelajar (Person)
- Data_NilaiUjian (Person)
- Data_NilaiUjian2 (Person)
- data_prediksi (Person - v1)
- data_tugas (Person - v1, 1)
- DATA_tugasnilaiujian (Person)
- data_tugasprediksi (Person)
- DataCuaca_Training (Person)
- HasilRuleInduction (Person)
- L200170107_bandu (Person)
- MahasiswaTraining (Person)
- ModulSekolah_Trainin (Person)
- Tabel_Cuaca_Testing (Person)
- Tabel_NilaiUjian (Person)

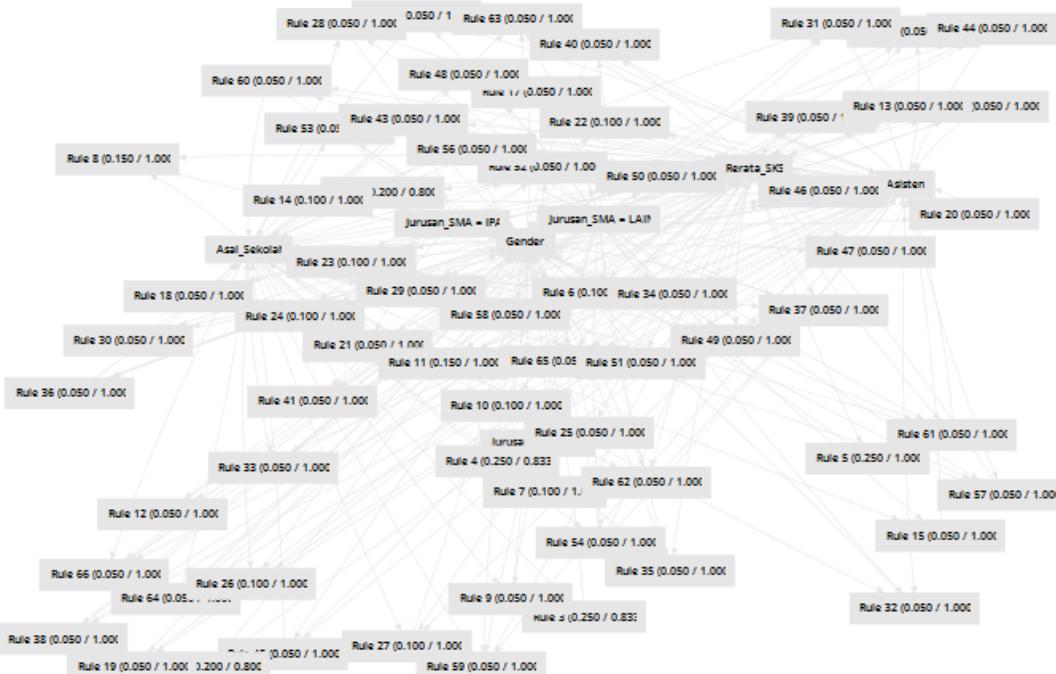
Max Size = 5

Min Size = 1

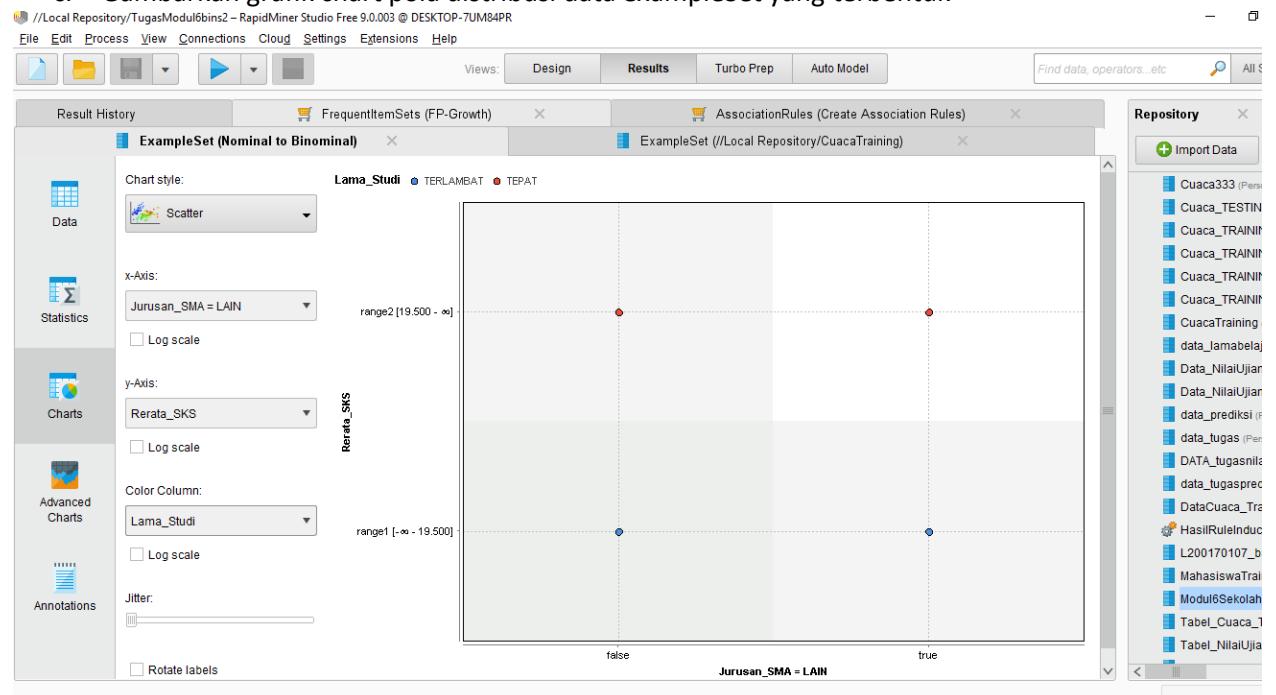
- b. Jumlah data pasangan premis dan kesimpulan pada Association Rules (Create Association Rules)! Gambarkan Tablenya! Gambarkan pula grafik yang terbentuk!

| No. | Premises | Conclusion | Support | Cor |
|-----|--|----------------------------------|---------|-----|
| 44 | Asal_Sekolah, Jurusan_SMA = LAIN | Asisten, Rerata_SKS | 0.050 | 1 |
| 45 | Asisten, Jurusan_SMA = LAIN | Asal_Sekolah, Rerata_SKS | 0.050 | 1 |
| 46 | Asal_Sekolah, Asisten, Jurusan_SMA = LAIN | Rerata_SKS | 0.050 | 1 |
| 47 | Rerata_SKS, Jurusan_SMA = LAIN | Asal_Sekolah, Asisten | 0.050 | 1 |
| 48 | Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN | Asisten | 0.050 | 1 |
| 49 | Asisten, Rerata_SKS, Jurusan_SMA = LAIN | Asal_Sekolah | 0.050 | 1 |
| 50 | Gender, Jurusan_SMA = IPA, Asal_Sekolah, Asisten | Rerata_SKS | 0.050 | 1 |
| 51 | Jurusan_SMA = IPA, Asal_Sekolah, Rerata_SKS | Gender, Asisten | 0.050 | 1 |
| 52 | Gender, Jurusan_SMA = IPA, Asal_Sekolah, Rerat... | Asisten | 0.050 | 1 |
| 53 | Jurusan_SMA = IPA, Asal_Sekolah, Asisten, Rerat... | Gender | 0.050 | 1 |
| 54 | Asal_Sekolah, Jurusan_SMA = LAIN | Gender, Asisten, Rerata_SKS | 0.050 | 1 |
| 55 | Gender, Asal_Sekolah, Jurusan_SMA = LAIN | Asisten, Rerata_SKS | 0.050 | 1 |
| 56 | Asisten, Jurusan_SMA = LAIN | Gender, Asal_Sekolah, Rerata_SKS | 0.050 | 1 |
| 57 | Gender, Asisten, Jurusan_SMA = LAIN | Asal_Sekolah, Rerata_SKS | 0.050 | 1 |
| 58 | Asal_Sekolah, Asisten, Jurusan_SMA = LAIN | Gender, Rerata_SKS | 0.050 | 1 |

Grafik Penyebaran data yang terbentuk



C. Gambarkan grafik chart pola distribusi data exampleSet yang terbentuk



3. Pada Operator Descretize by Frequency memiliki nilai: number of bins=1

a. Jumlah set aturan assosiasi dan total max size yang terbentuk berdasarkan FP-GROWTH (table view) ! Gambarkan Tabelnya:

| Data | | No. of Sets: 55 Total Max. Size: 5 | Size | Support | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
|-------------|--|---------------------------------------|------|---------|--------------------|--------------------|--------|--------|--------|
| Annotations | | | 1 | 0.250 | Asisten | | | | |
| | | Min. Size: 1 | 1 | 0.250 | Rerata_SKS | | | | |
| | | Max. Size: 5 | 1 | 0.200 | Jurusan_SMA = L... | | | | |
| | | Contains Item: | 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 = L... | | | |
| | | | 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 | | | |

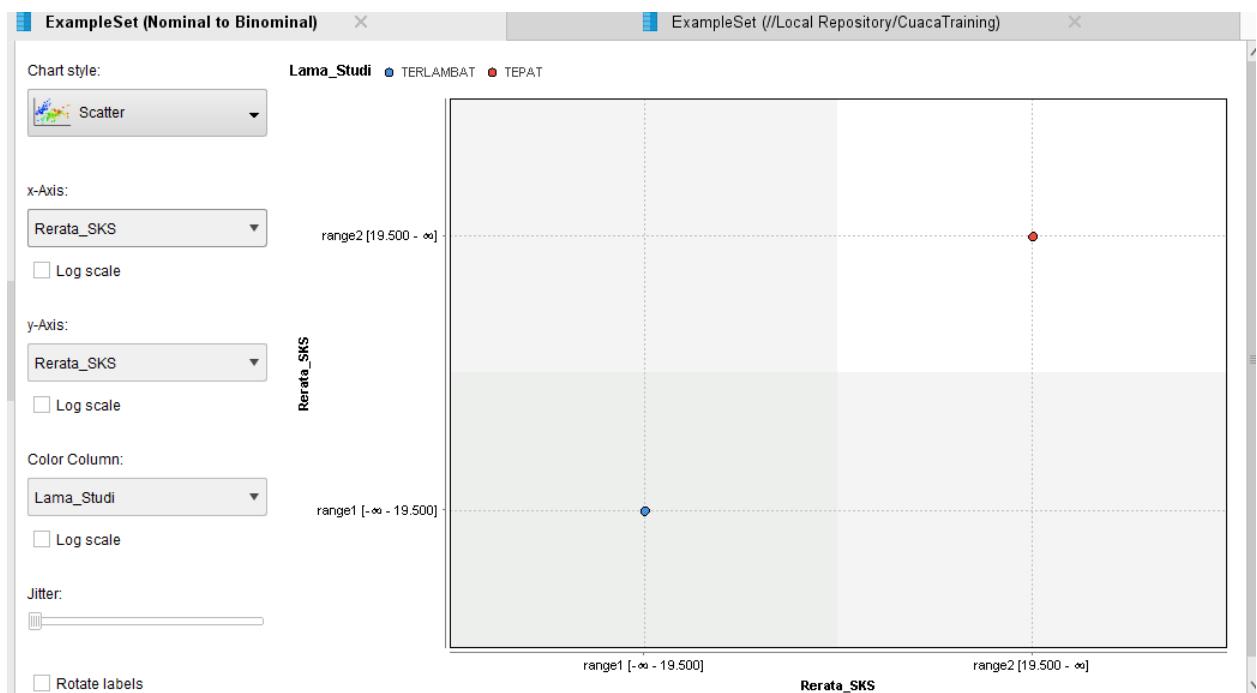
b. Jumlah data pasangan premis dan kesimpulan pada Association Rules (Create Association Rules)! Gambarkan Tablenya! Gambarkan pula grafik yang terbentuk!

| Show rules matching | | No. | Premises | Conclusion | Support | Confidence |
|---------------------------|--|-----|---|----------------------------------|---------|------------|
| all of these conclusions: | | 52 | Gender, Jurusan_SMA = IPA, Asal_Sekolah, Rerata... | Asisten | 0.050 | 1 |
| Gender | | 53 | Jurusan_SMA = IPA, Asal_Sekolah, Asisten, Rerata... | Gender | 0.050 | 1 |
| Jurusan_SMA = IPA | | 54 | Asal_Sekolah, Jurusan_SMA = LAIN | Gender, Asisten, Rerata_SKS | 0.050 | 1 |
| Asal_Sekolah | | 55 | Gender, Asal_Sekolah, Jurusan_SMA = LAIN | Asisten, Rerata_SKS | 0.050 | 1 |
| Asisten | | 56 | Asisten, Jurusan_SMA = LAIN | Gender, Asal_Sekolah, Rerata_SKS | 0.050 | 1 |
| Rerata_SKS | | 57 | Gender, Asisten, Jurusan_SMA = LAIN | Asal_Sekolah, Rerata_SKS | 0.050 | 1 |
| | | 58 | Asal_Sekolah, Asisten, Jurusan_SMA = LAIN | Gender, Rerata_SKS | 0.050 | 1 |
| | | 59 | Gender, Asal_Sekolah, Asisten, Jurusan_SMA = L... | Rerata_SKS | 0.050 | 1 |
| | | 60 | Rerata_SKS, Jurusan_SMA = LAIN | Gender, Asal_Sekolah, Asisten | 0.050 | 1 |
| | | 61 | Gender, Rerata_SKS, Jurusan_SMA = LAIN | Asal_Sekolah, Asisten | 0.050 | 1 |
| | | 62 | Asal_Sekolah, Rerata_SKS, Jurusan_SMA = LAIN | Gender, Asisten | 0.050 | 1 |
| | | 63 | Gender, Asal_Sekolah, Rerata_SKS, Jurusan_SMA =... | Asisten | 0.050 | 1 |
| | | 64 | Asisten, Rerata_SKS, Jurusan_SMA = LAIN | Gender, Asal_Sekolah | 0.050 | 1 |
| | | 65 | Gender, Asisten, Rerata_SKS, Jurusan_SMA = LAIN | Asal_Sekolah | 0.050 | 1 |
| | | 66 | Asal_Sekolah, Asisten, Rerata_SKS, Jurusan_SMA... | Gender | 0.050 | 1 |

Grafik Penyebaran data yang terbentuk



c. Gambarkan grafik chart pola distribusi data exampleSet yang terbentuk



Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

Modul 12

REGRESI LINIER SEDERHANA

❖ Langkah-langkah Praktikum

1. Mencari Nilai t-hitung dan model Regresi Linier

Tabel data siswa dengan format .xls

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

2. Mengubah tipe data dan jenis masing-masing atribut sebagai berikut.

Format your columns.

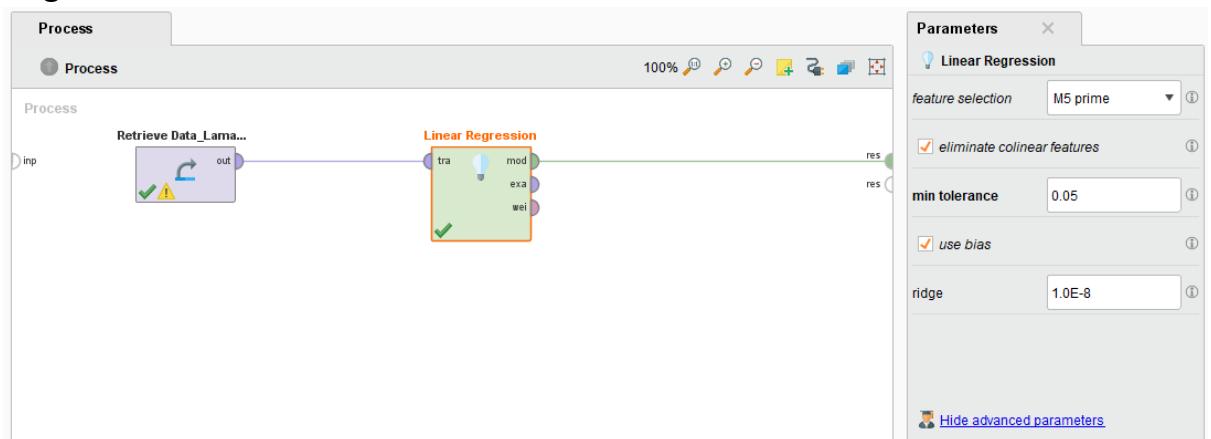
Replace errors with missing values (i)

| NO_SISWA <i>polynomial id</i> | NAMA <i>polynomial</i> | LAMA BELAJAR (JAM) <i>integer</i> | NILAI <i>integer label</i> |
|----------------------------------|---------------------------|--------------------------------------|-------------------------------|
| 1 S-101 | JOKO | 15 | 783 |
| 2 S-102 | AGUS | 18 | 877 |
| 3 S-103 | SUSI | 7 | 505 |
| 4 S-104 | DYAH | 9 | 860 |
| 5 S-105 | WATI | 15 | 968 |
| 6 S-106 | 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 |

✓ no problems.

Previous Next ✖ Cancel

- a. NO_SISWA:*polynomial id*
 - b. NAMA: Exclude column
 - c. LAMA JAM BELAJAR:*integer*
 - d. NILAI:*integer label*
3. Memasukan data pada area proces untuk melakukan Modeling Linier Regression.



Parameter min tolerance = 0.05 yaitu (batas toleransi minimal 5%), jika data input bertipe nominal atau polynomial maka tambahkan oprator blending Nomial/Polynomial to Numerical, karena data input bertipe integer jadinya oprator tersebut tidak digunakan.

4. Hasil dari proses Regresi Linier menggunakan Data Siswa:

- a. Table View (mencari besarnya niai t-hitung)

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

Dari table tersebut dapat terlihat nilai t-statistic (t-hitung) sebesar 2,827, berdasarkan aturan statistik, variable X dikatakan mempengaruhi secara signifikan terhadap Y jika nilai t-hitung > t-table.

- b. Text View (mencari model regresi)

LinearRegression

```
21.608 * LAMA BELAJAR (JAM)
+ 492.769
```

Hasil pada text view terlihat sebuah persamaan berikut:

$$= 21.608 * \text{Lama Jam Belajar} + 492,769$$

Maka dapat dibuat sebuah model persamaan regresi linier sederhana untuk mencari nilai variable Y (Nilai Ujian) berdasarkan variable X_1 (Lama Belajar).

5. Mencari Nilai t dan Model Regresi Linier Menggunakan Rapid Miner

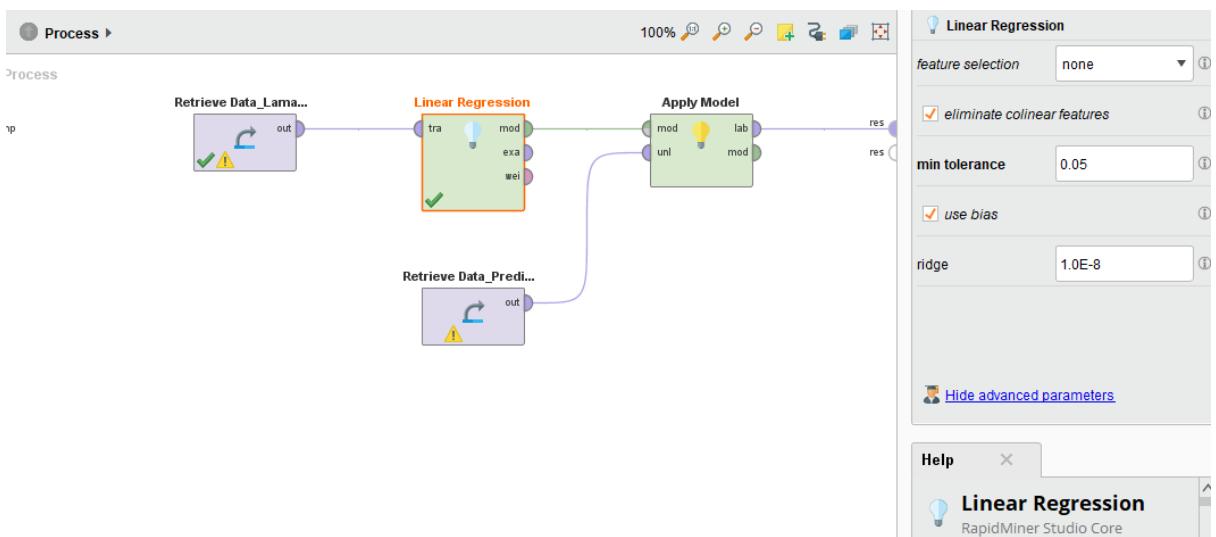
Tabel prediksi data siswa dengan nama **Tabel_PrediksiNilaiUjian.xls**

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

Ubah tipe data dan jenis masing-masing atribut sebagai berikut:

| | NO_SISWA polynomial ID | 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 |

- NO_SISWA: polynomial, id
 - Nama:pilih exclude column
 - LAMA JAM BELAJAR : integer
6. Tambahkan oprator Apply Model, dan dengan menggunakan parameter yang sama pada oprator Regression Linier, berikut dashboard Proses terhadap data testing menggunakan regresi liner:

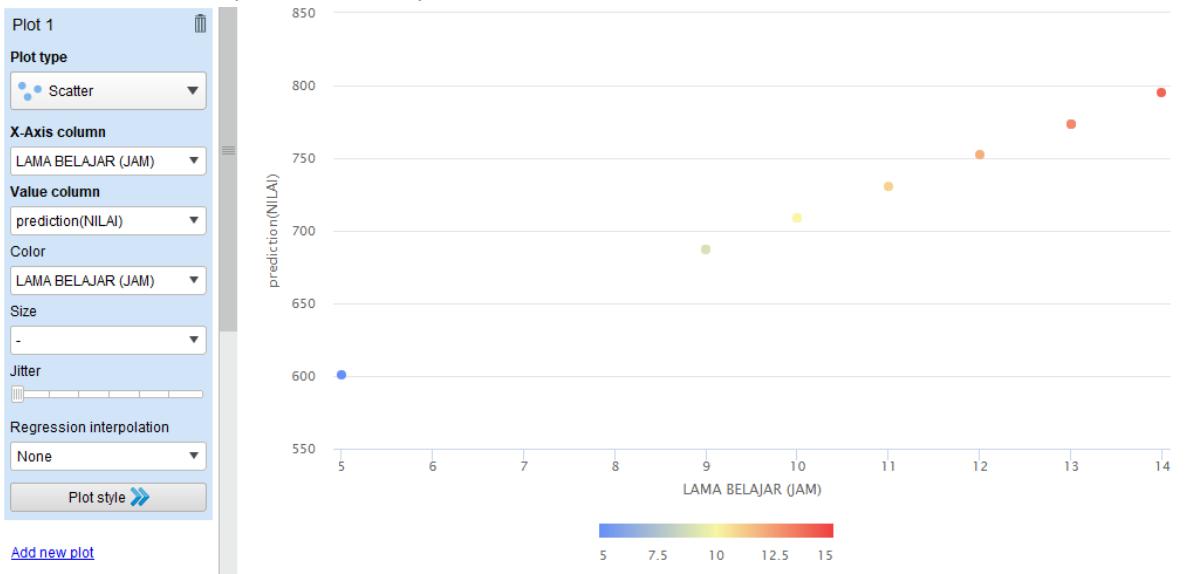


a. Data View (hasil prediksi nilai ujian)

| 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 |

Hasil nilai ujian terhadap 10 siswa lainnya.

b. Chart View (Scatter Plot)



7. Membuktikan Model Regresi

Pada tahap ini membuktikan hasil prediksi menggunakan RapidMiner dengan hasil prediksi menggunakan model regresi yang telah dihasilkan.

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

| | A | B | C | D | E |
|----|----------|-------|--------------------|--------------------|----------------------------------|
| 1 | NO_SISWA | NAMA | LAMA BELAJAR (JAM) | Prediction (NILAI) | Prediction (NILAI) Model Regresi |
| 2 | S-111 | BUDI | 12 | 752.061 | |
| 3 | S-112 | SANTI | 13 | 773.668 | |
| 4 | S-113 | DIAN | 14 | 795.276 | |
| 5 | S-114 | DANI | 11 | 730.453 | |
| 6 | S-115 | AHMAD | 5 | 600.807 | |
| 7 | S-116 | BAYU | 13 | 773.668 | |
| 8 | S-117 | RISA | 9 | 687.238 | |
| 9 | S-118 | RANI | 10 | 708.845 | |
| 10 | S-119 | YANI | 10 | 708.845 | |
| 11 | S-120 | RATIH | 9 | 687.238 | |
| 12 | | | | | |

Memasukan nilai variable X1 ke dalam model regresi :

1. No.Siswa = S-111, Nama = Budi, memiliki nilai X1 =12 Jam Sehingga, Y = $(21,608 \times 12) + 492,769 = 752,065$

2. No.Siswa = S-115, Nama = Ahmad, memiliki nilai X1 = 5 Jam Sehingga, Y = $(21,608 \times 5) + 492,769 = 600,809$
3. No.Siswa = S-120, Nama = Ratih, memiliki nilai X1 = 9 Jam Sehingga, Y = $(21,608 \times 9) + 492,769 = 687,241$

Berikut Tabel hasil pembuktianya :

Menggunakan ms.exel untuk melakukan pembuktian dengan formula
“=(21,603*columnsel)+492,769.

| A | B | C | D | E |
|----|----------|-------|--------------------|----------------------------------|
| 1 | NO_SISWA | NAMA | LAMA BELAJAR (JAM) | Prediction (NILAI) Model Regresi |
| 2 | S-111 | BUDI | 12 | 752.061 |
| 3 | S-112 | SANTI | 13 | 773.668 |
| 4 | S-113 | DIAN | 14 | 795.276 |
| 5 | S-114 | DANI | 11 | 730.453 |
| 6 | S-115 | AHMAD | 5 | 600.807 |
| 7 | S-116 | BAYU | 13 | 773.668 |
| 8 | S-117 | RISA | 9 | 687.238 |
| 9 | S-118 | RANI | 10 | 708.845 |
| 10 | S-119 | YANI | 10 | 708.845 |
| 11 | S-120 | RATIH | 9 | 687.238 |

Dapat dilihat bahwa hasil prediksi menggunakan rapid miner menghasilkan nilai yang sama dengan menggunakan model persamaan regresi linier.

Nama : Oki Kus Mahesa

Nim : L200170064

Kelas : C

Modul 12

REGRESSI LINIER

- Tugas
1. Kasus: Dalam sebuah Kepala Keluarga telah diperoleh variable pendapatan rata-rata perbulan, jumlah anggota keluarga yang tinggal serumah, dan daya beli rata-rata perbulan.
 2. Bagaimana model Regresi Linier yang terbentuk, dan lakukan prediksi terhadap 10 data yang belum diketahui nilai daya belinya.
 3. Tabel Hasil Survei 15 Kepala Keluarga

| A | B | C | D |
|--------------|---------------------|-------------------------|--------------------|
| NO.RESPONDEN | PENDAPATAN (RUPIAH) | JUMLAH ANGGOTA KELUARGA | DAYA BELI (RUPIAH) |
| 1 | 1,000,000 | 6 | 834,000 |
| 2 | 1,400,000 | 7 | 1,200,000 |
| 3 | 200,000 | 3 | 134,000 |
| 4 | 1,400,000 | 6 | 1,167,000 |
| 5 | 500,000 | 3 | 334,000 |
| 6 | 1,700,000 | 5 | 1,360,000 |
| 7 | 400,000 | 3 | 267,000 |
| 8 | 1,900,000 | 5 | 1,520,000 |
| 9 | 300,000 | 3 | 200,000 |
| 10 | 500,000 | 4 | 375,000 |
| 11 | 700,000 | 7 | 600,000 |
| 12 | 1,900,000 | 3 | 1,267,000 |
| 13 | 800,000 | 4 | 600,000 |
| 14 | 1,500,000 | 4 | 125,000 |
| 15 | 1,300,000 | 7 | 1,115,000 |

4. Buatlah proses Regresi Linier Sederhana menggunakan RapidMiner dengan ketentuan sebagai berikut.
 - a. Variable bebas (X) = Pendapatan (X1), Jumlah Anggota Keluarga (X2)
 - b. Variable terikat (Y) = Daya Beli

c. Toleransi yang digunakan = 5%

| NO.RESPONDEN <i>integer id</i> | PENDAPATAN (RUPIAH) <i>integer</i> | JUMLAH ANGGOTA K... <i>integer</i> | DAYA BELI (RUPIAH) <i>integer label</i> |
|---------------------------------------|---------------------------------------|---------------------------------------|--|
| 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.

5. Tentukan apakah variable x1 dan x2 mempengaruhi secara signifikan terhadap nilai Y berdasarkan besarnya nilai t-stat?

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

| Attribute | Coefficient | Std. Error | Std. Coefficient | Tolerance | t-Stat | p-Value | Code |
|------------------|-------------|------------|------------------|-----------|--------|---------|------|
| PENDAPATAN (R... | 0.605 | 0.131 | 0.723 | 0.857 | 4.625 | 0.001 | *** |
| JUMLAH ANGGO... | 85541.767 | 48568.994 | 0.275 | 0.857 | 1.761 | 0.104 | |
| (Intercept) | -284291.899 | 226294.413 | ? | ? | -1.256 | 0.233 | |

Dari tabel tersebut dapat dilihat nilai t-statistic (t-hitung) sebesar:

X1 = Pendapatan = 4.625

X2 = Jumlah Anggota Keluarga = 1.761

Berdasarkan aturan statistik, variable X1 dan X2 dikatakan mempengaruhi secara signifikan terhadap variable Y jika nilai t-hitung > t-table,

Yang artinya bahwa “Pendapatan(x1) dan Jumlah Anggota Keluarga(x2) dapat Mempengaruhi Daya Beli(y).

b. Text View (mencari model regresi)

LinearRegression

```
0.605 * PENDAPATAN (RUPIAH)
+ 85541.767 * JUMLAH ANGGOTA KELUARGA
- 284291.899
```

Berdasarkan hasil pada Text View terlihat sebuah persamaan linier berikut:

$$= 0.605 * \text{PENDAPATAN (RUPIAH)} + 85541.767 * \text{JUMLAH ANGGOTA KELUARGA} - 284291.899$$

Maka dapat dibuat sebuah model persamaan regresi linier sederhana untuk mencari nilai variable y (Daya Beli) berdasarkan variable x_1 (Pendapatan) dan variable x_2 (Jumlah Anggota Keluarga).

Berikut model regresi linier yang terbentuk:

$$Y = 0.605 x_1 + 85541.767 x_2 - 284291.899$$

Dengan model tersebut, dapat dilihat dicari Daya Beli(Y) dengan memasukan nilai Pendapatan pada variable x_1 dan pada variable x_2

6. Gunakan data testing untuk menjawab perintah berikut:

- Lakukan prediksi Daya Beli(Y) dengan menggunakan Model Persamaan Regresi linier dari hasil pertanyaan nomor 5!
 - Lakukan prediksi Daya Beli (Y) menggunakan RapidMiner!
- Data Testing yang digunakan untuk prediksi :

| A | B | C |
|--------------|---------------------|-------------------------|
| NO RESPONDEN | PENDAPATAN (RUPIAH) | JUMLAH ANGGOTA KELUARGA |
| 1 | 900,000 | 5 |
| 2 | 800,000 | 3 |
| 3 | 500,000 | 2 |
| 4 | 1,900,000 | 6 |
| 5 | 600,000 | 2 |
| 6 | 800,000 | 5 |
| 7 | 1,000,000 | 6 |
| 8 | 1,100,000 | 4 |
| 9 | 1,000,000 | 4 |
| 10 | 500,000 | 3 |
| 11 | | |

- Mengubah Tipe Data dan jenis masing-masing atribut sebagai berikut:
 - a. NO.RESPONDEN: integer,id
 - b. PENDAPATAN: integer
 - c. JUMLAH ANGGOTA KELUARGA: integer

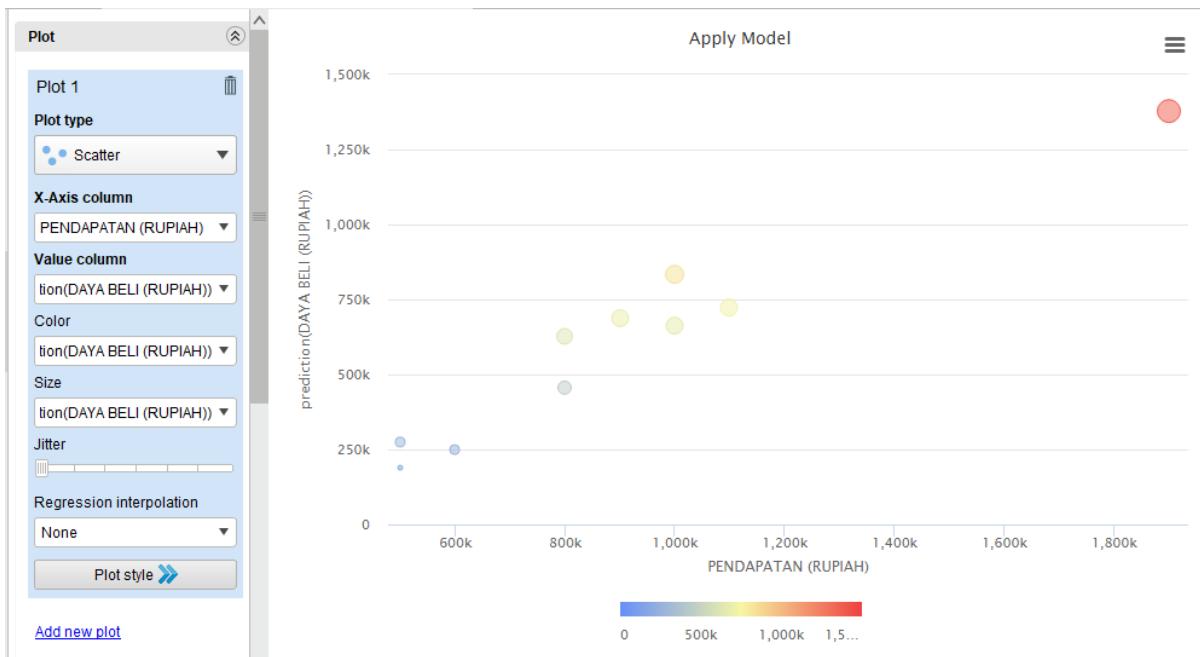
| | NO.RESPONDEN <i>integer id</i> | PENDAPATAN (RUPIAH) <i>integer</i> | JUMLAH ANGGOTA KELUARGA <i>integer</i> |
|----|---------------------------------------|---------------------------------------|---|
| 1 | 1 | 900000 | 5 |
| 2 | 2 | 800000 | 3 |
| 3 | 3 | 500000 | 2 |
| 4 | 4 | 1900000 | 6 |
| 5 | 5 | 600000 | 2 |
| 6 | 6 | 800000 | 5 |
| 7 | 7 | 1000000 | 6 |
| 8 | 8 | 1100000 | 4 |
| 9 | 9 | 1000000 | 4 |
| 10 | 10 | 500000 | 3 |

- Data View (Hasil Prediksi Daya Beli)

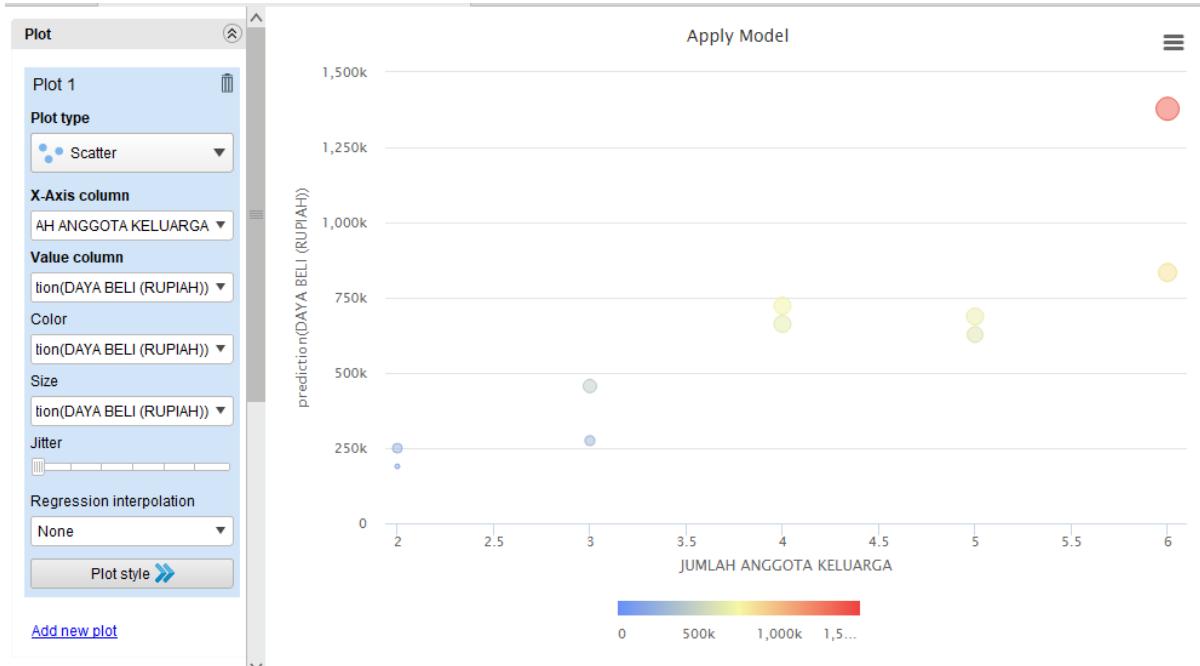
| NO.RESPONDEN | PENDAPATAN (RUPIAH) | JUMLAH ANGGOTA KELUARGA | <code>prediction(DAYA BELI (RUPIAH))</code> |
|--------------|---------------------|-------------------------|---|
| 1 | 900000 | 5 | 687740.118 |
| 2 | 800000 | 3 | 456176.230 |
| 3 | 500000 | 2 | 189193.402 |
| 4 | 1900000 | 6 | 1378085.419 |
| 5 | 600000 | 2 | 249673.756 |
| 6 | 800000 | 5 | 627259.765 |
| 7 | 1000000 | 6 | 833762.239 |
| 8 | 1100000 | 4 | 723159.057 |
| 9 | 1000000 | 4 | 662678.704 |
| 10 | 500000 | 3 | 274735.170 |

Hasil prediksi Daya Beli terhadap 10 Kepala Keluarga lainnya.

7. Pola sebaran data menggunakan Plot View (Scatter) dengan ketentuan:
- 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))



Berdasarkan Plot View dapat dilihat bahwa:

- Semakin Besar Jumlah Pendapatan(X1) Maka Jumlah Daya Beli(Y) akan semakin tinggi.
- Semakin Banyak Jumlah naggota Keluarga(X2) Maka Jumlah Daya Beli(Y) akan semakin tinggi