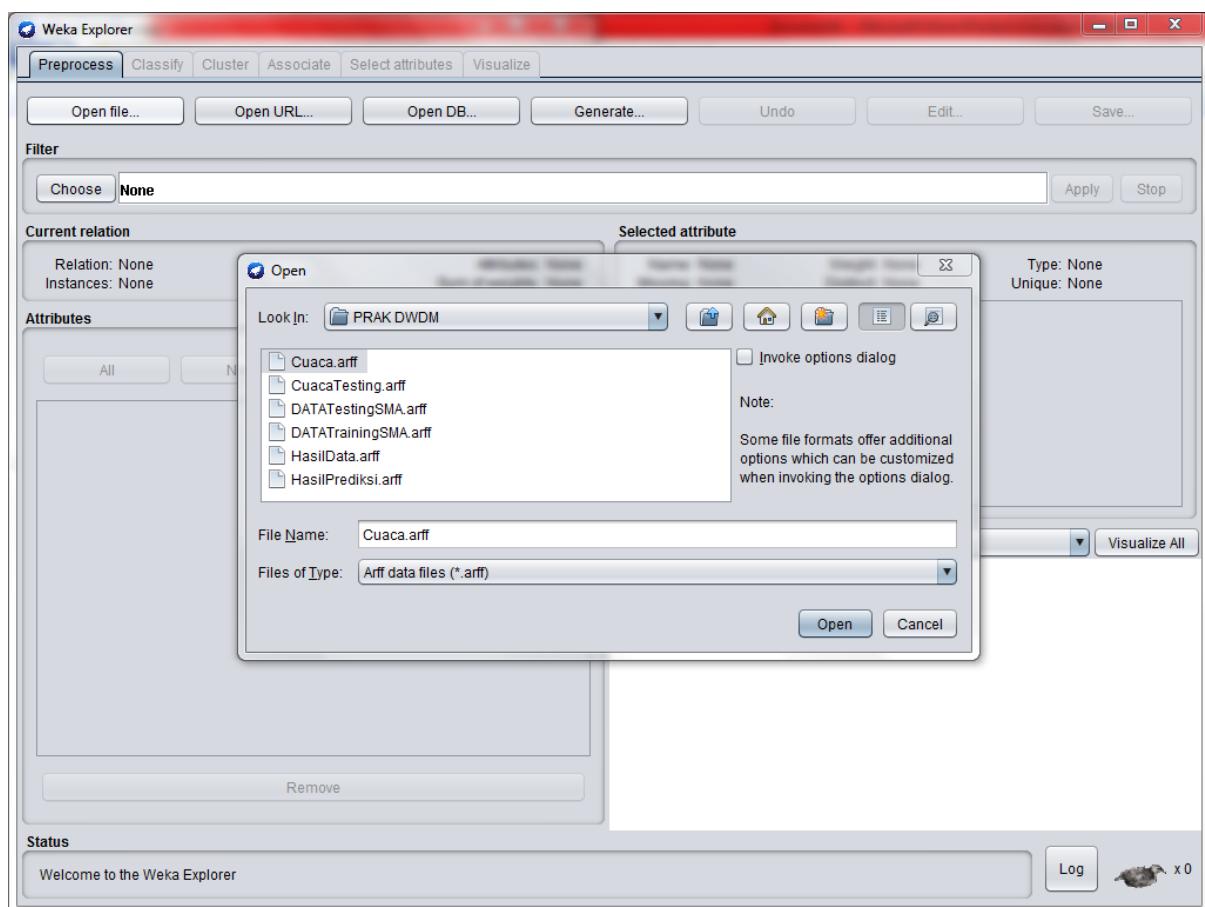


LAPORAN PRAKTIKUM

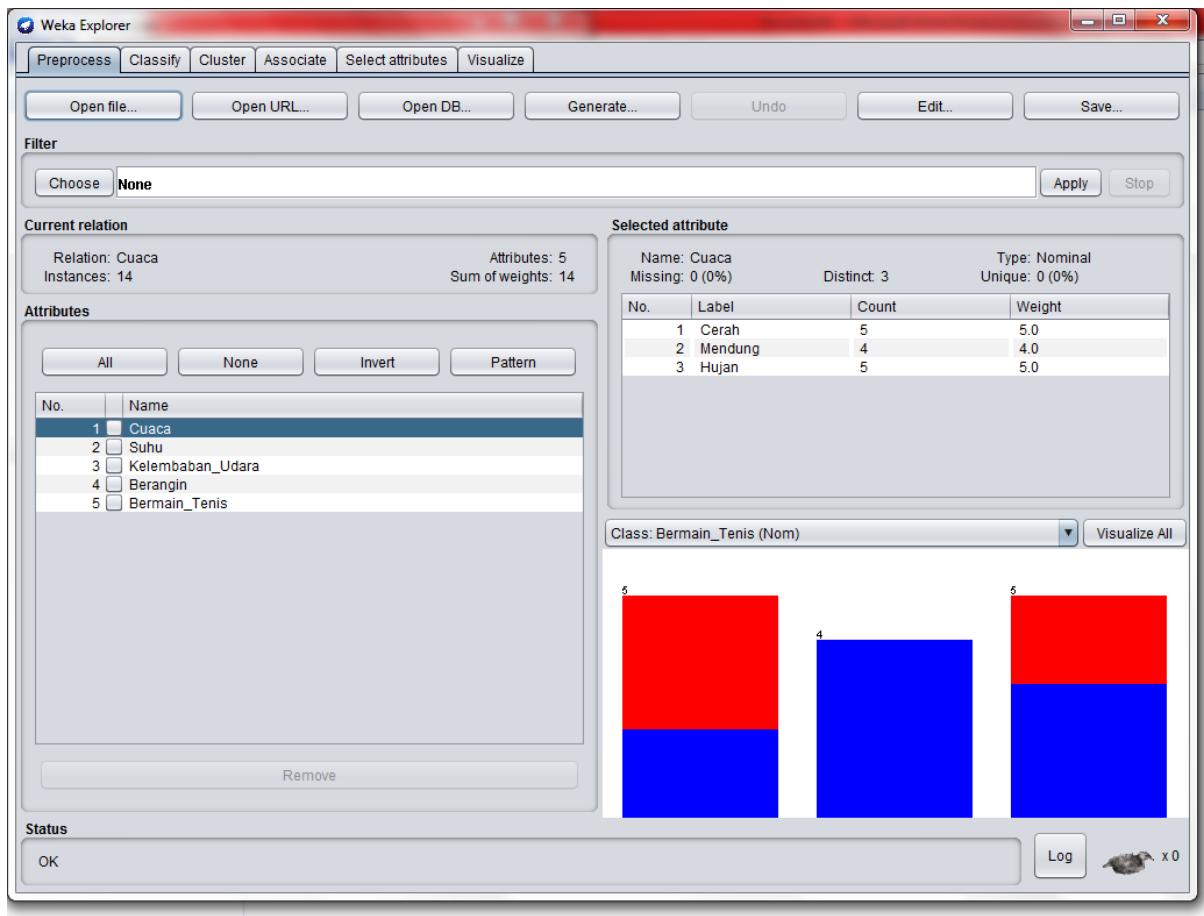
DATA WAREHOUSING DAN DATA MINING

NAMA : Ilham Athur B.W
KELAS : B
NIM : L200170023

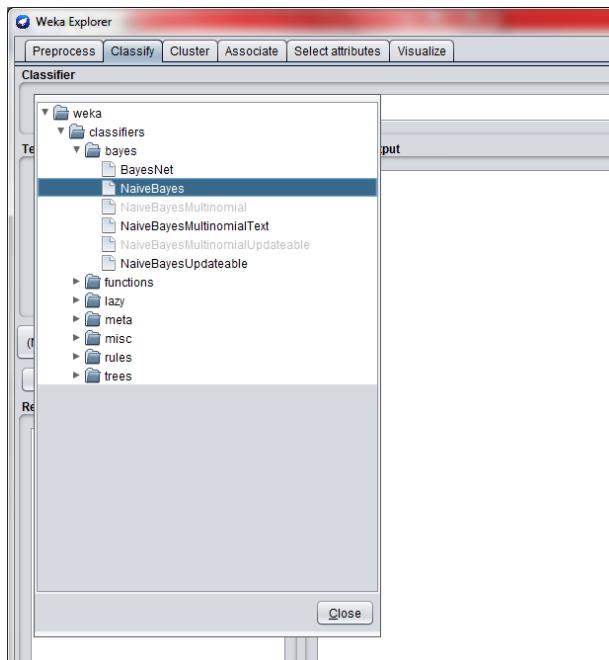
Persiapkan file **Cuaca.arff** Lalu buka **WEKA** Dan open file **Cuaca.arff** tadi

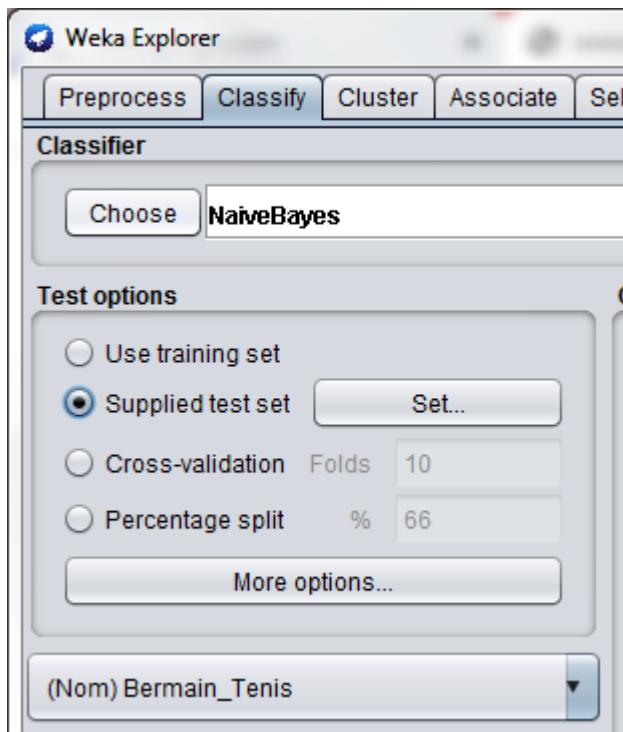


Dan hasilnya

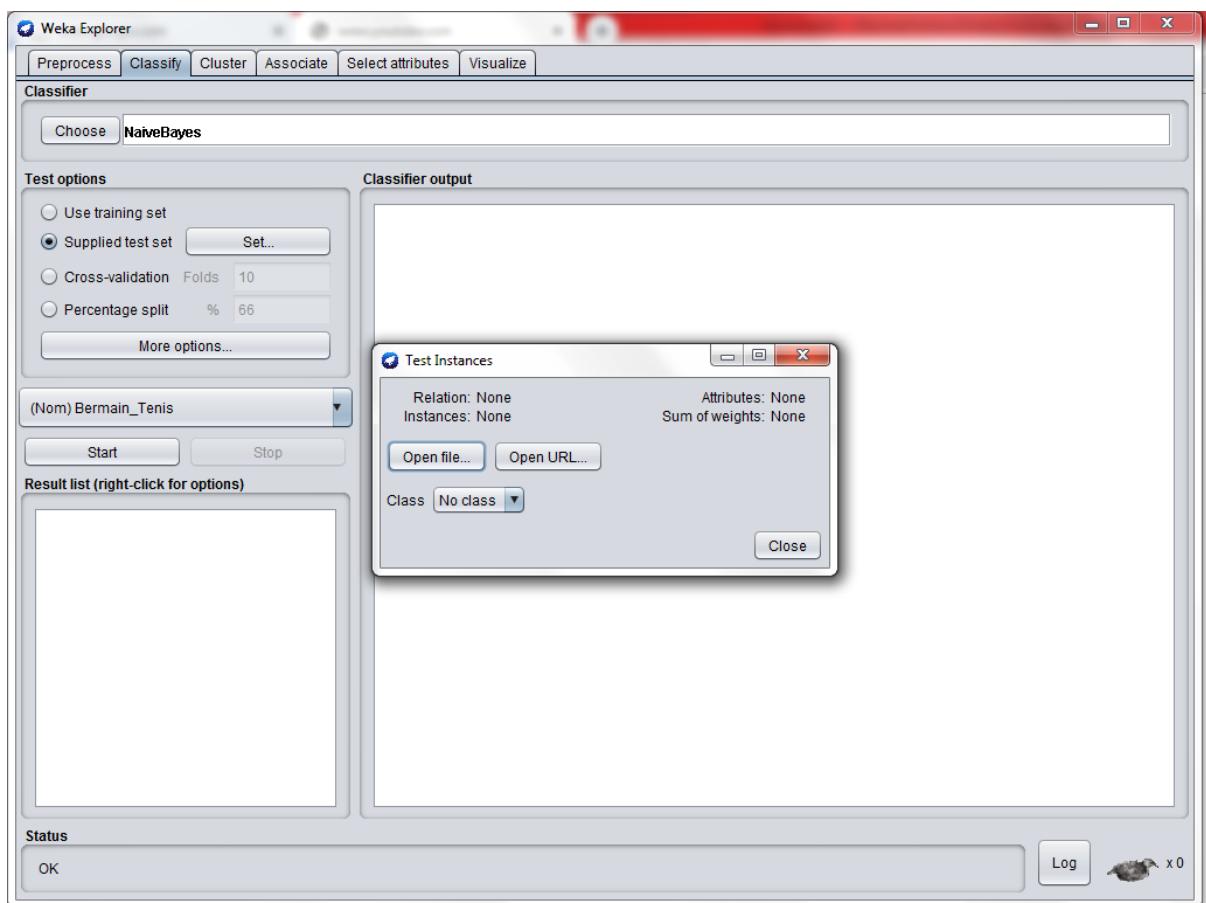


Lalu klik **CLASSIFY** dan klik choose dan pilih **naive bayes** dan lalu klik **SUPPLIED TEST SET** dan kita klik **SET**

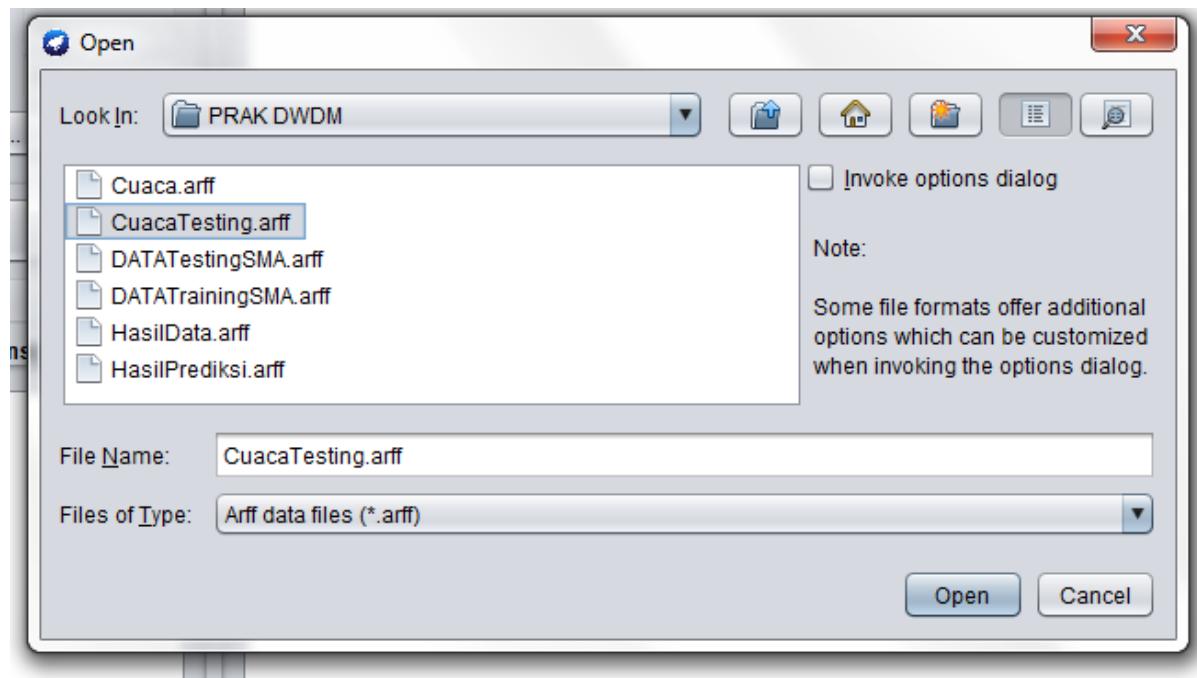




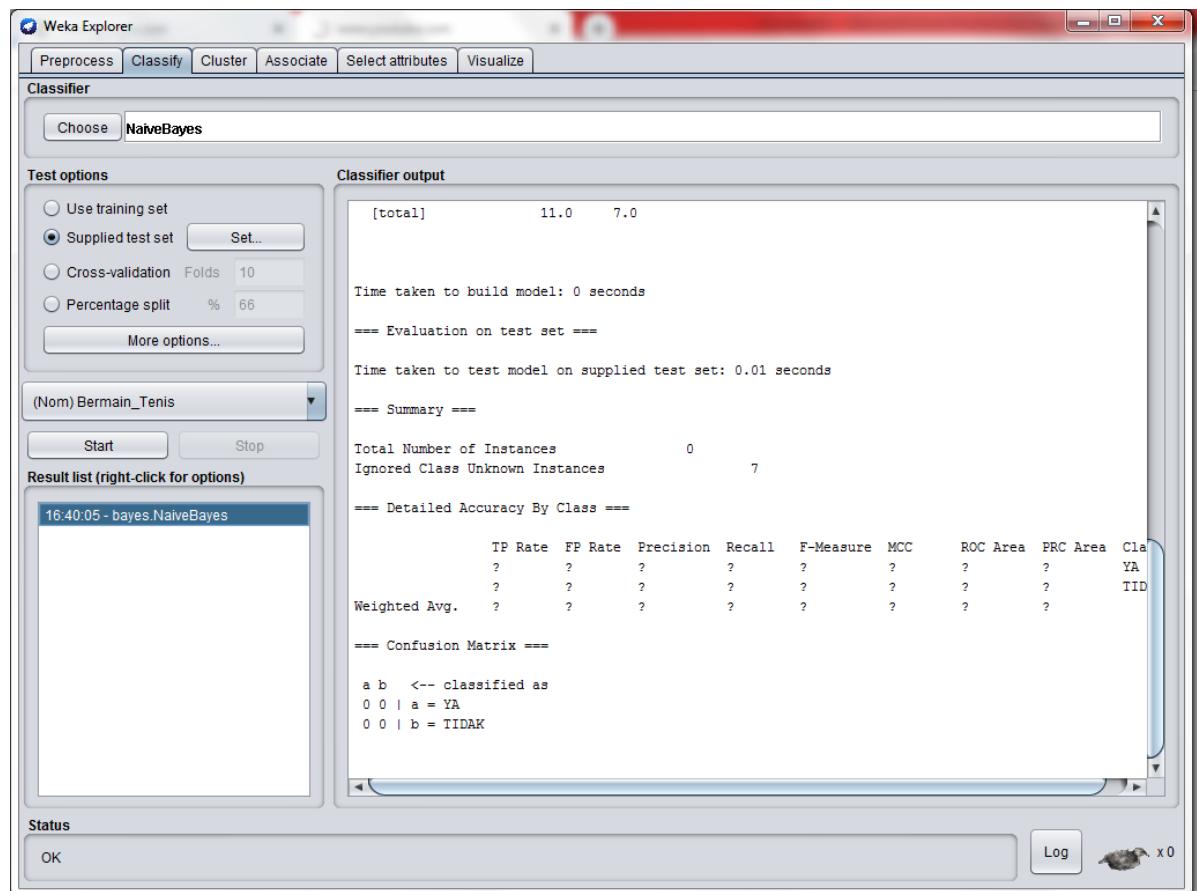
Lalu klik open file dan cari data yang akan dipakai



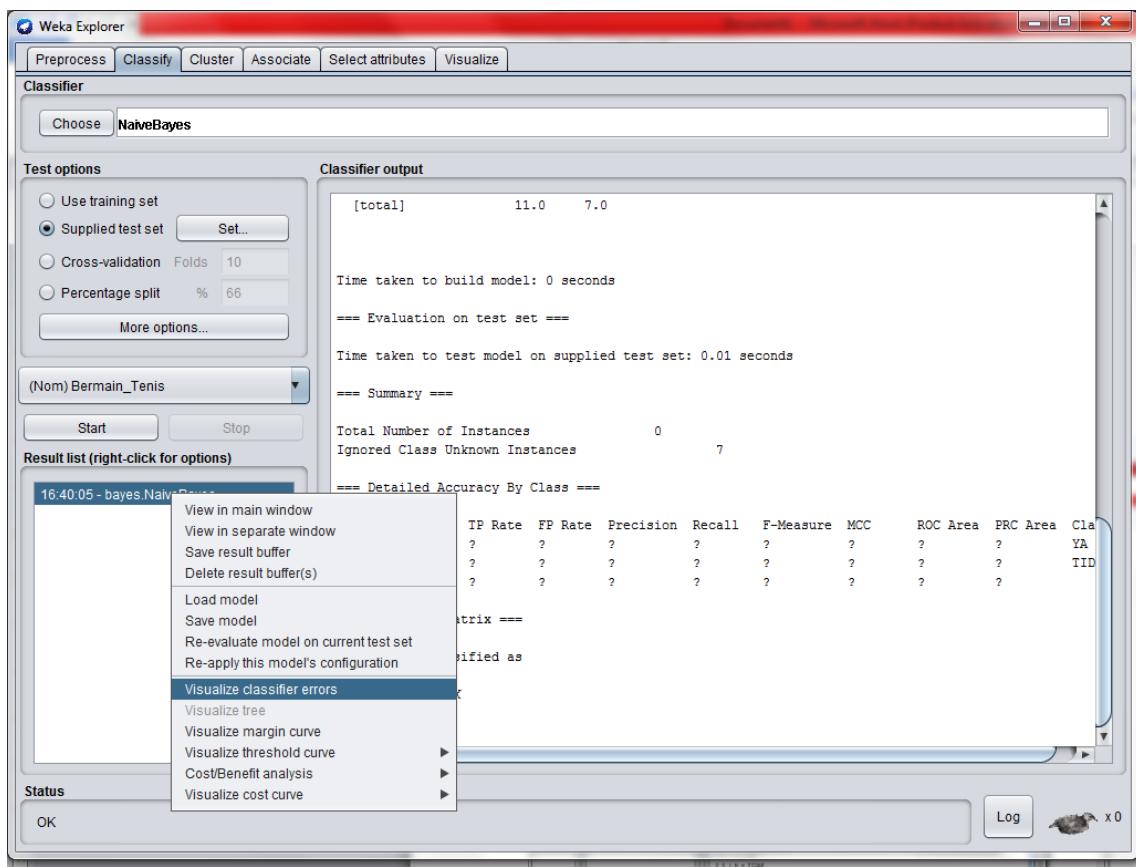
Kali ini akan memakai data **Cuacatesting.arff** lalu klik open



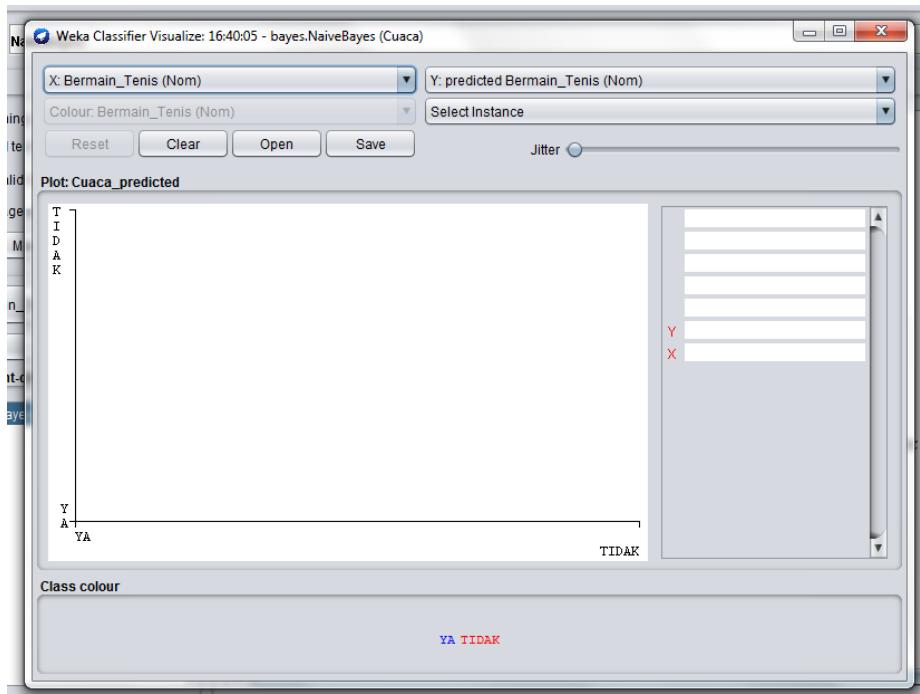
Dan hasilnya

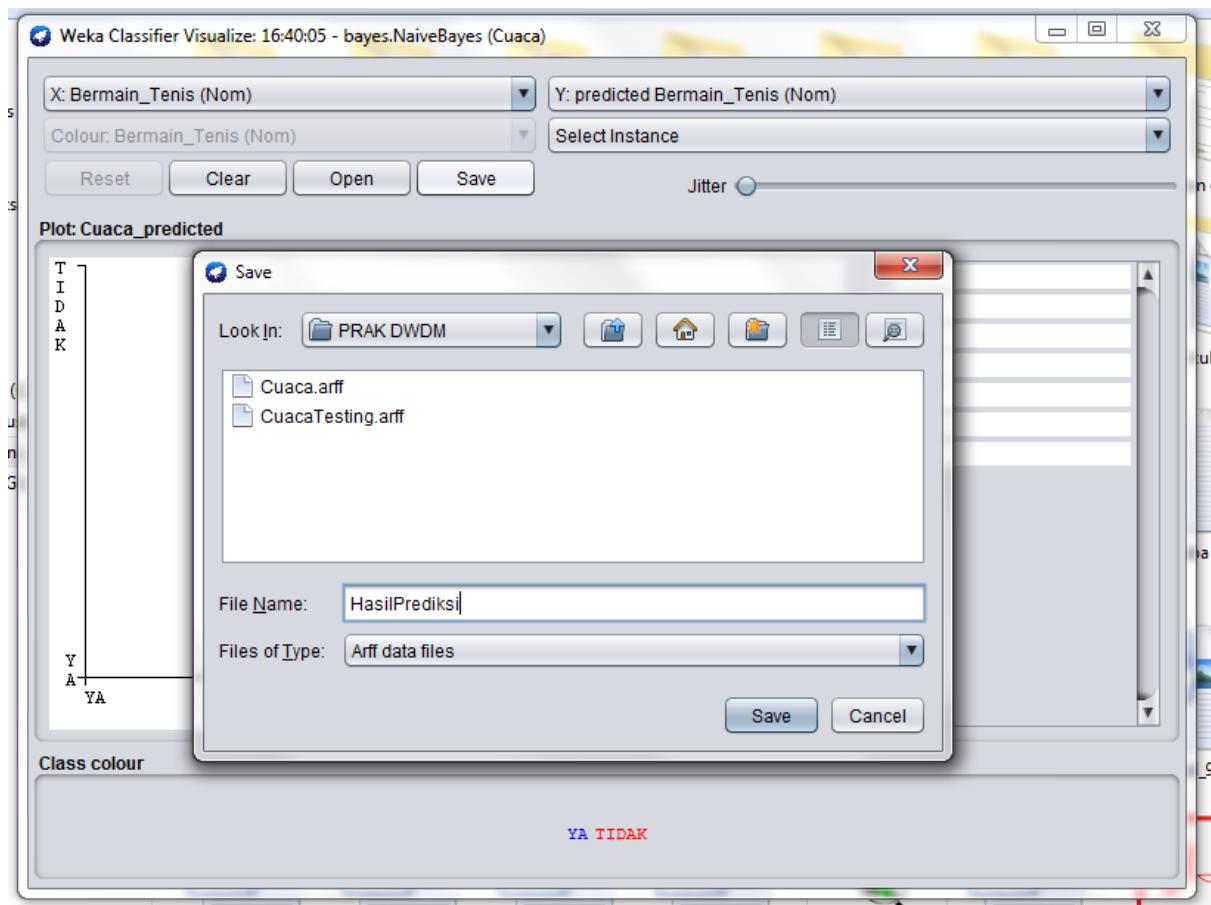


Lalu kita pilih **VISUALIZE CLASSIFIER ERRORS**



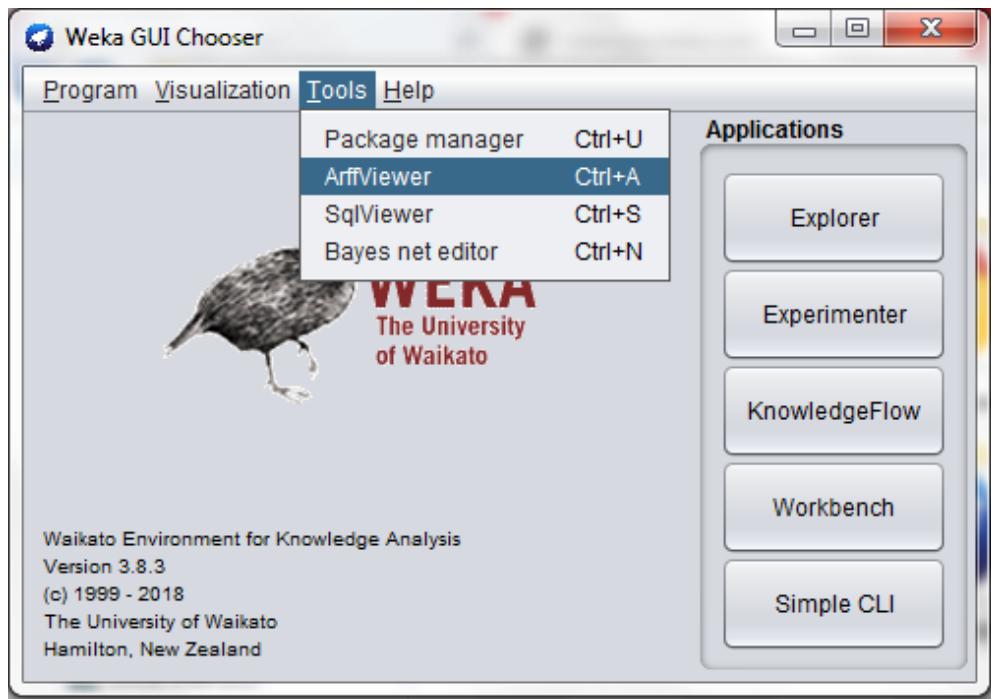
Lalu klik save dan simpan filenya dengan nama **HasilPrediksi.arff**





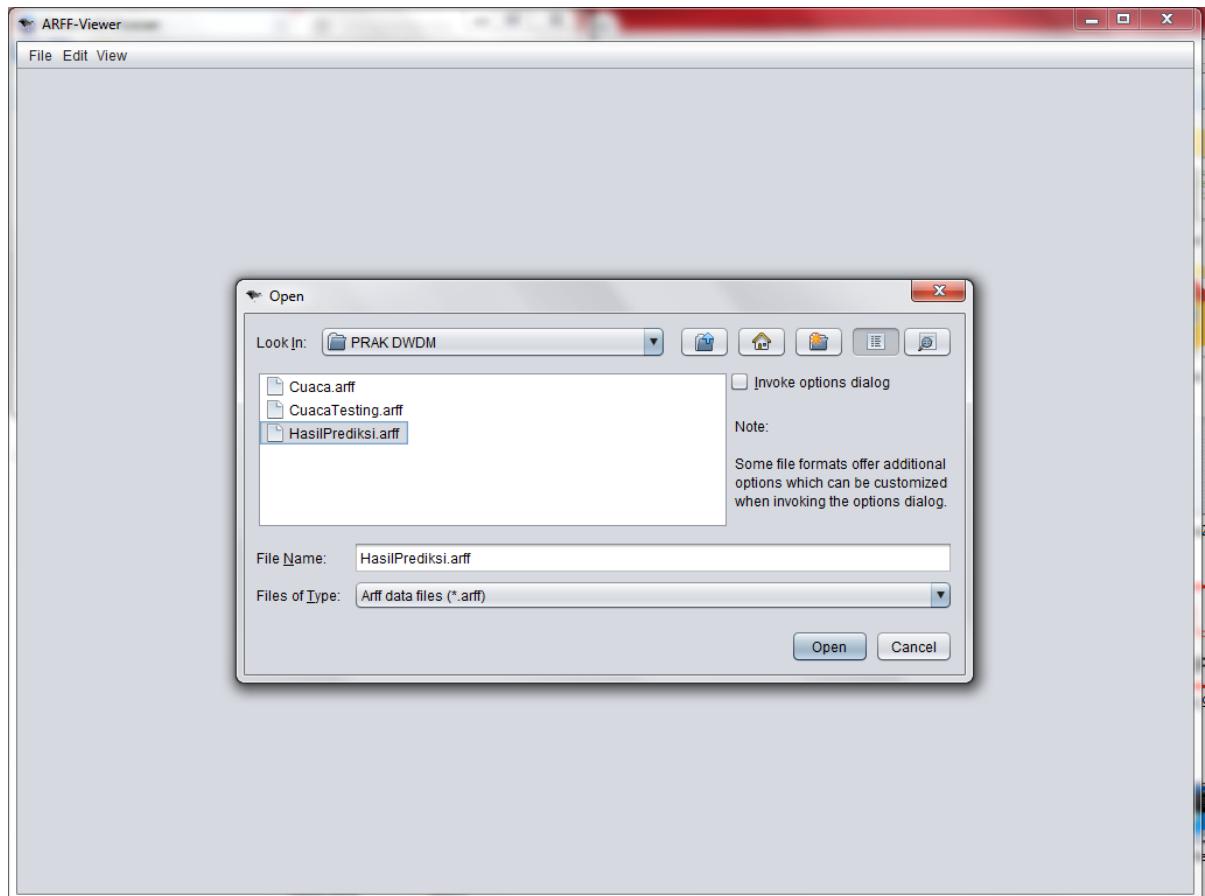
CLOSE semua laman WEKA sampai WEKA EXPLORER

Dan kita pilih **ArffViewers**



Maka tampilannya akan seperti ini

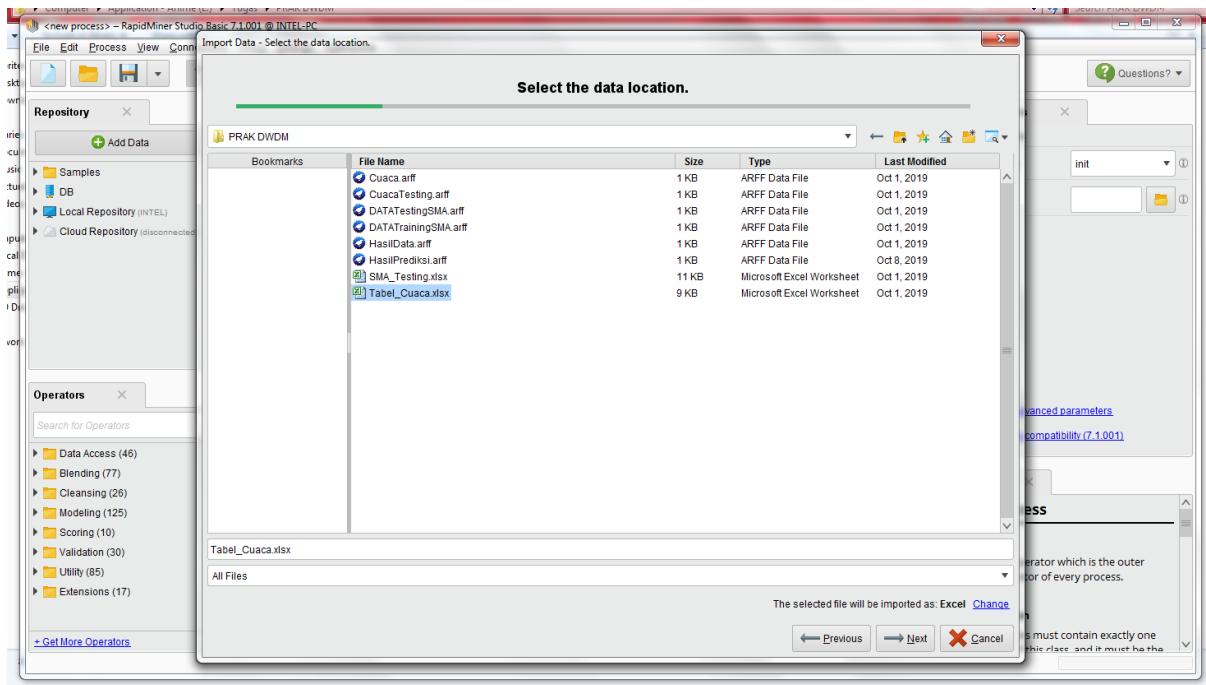
Dan kita open file data yang tadi kita simpan yaitu **HasilPrediksi.arff** lalu klik open



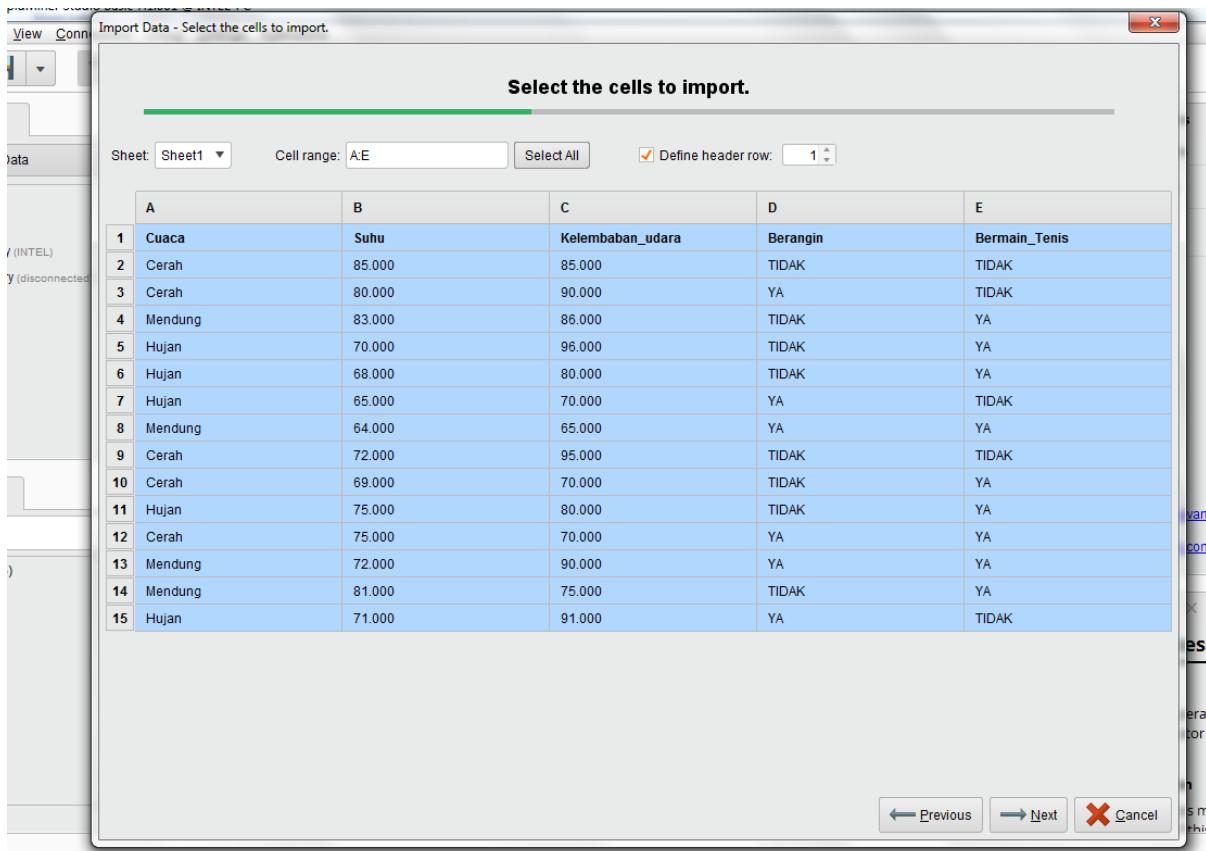
Dan hasilnya

Relation: Cuaca_predicted						
No.	1: Cuaca	2: Suhu	3: Kelembaban_Udara	4: Berangin	5: prediction margin	6: predicted Bermain_Tenis
	Nominal	Numeric	Numeric	Nominal	Numeric	Nominal
1	Cerah	75.0	65.0	TIDAK	0.762765	YA
2	Cerah	80.0	68.0	YA	0.087878	YA
3	Cerah	83.0	87.0	YA	-0.676866	TIDAK
4	Mend...	70.0	96.0	TIDAK	0.628523	YA
5	Mend...	68.0	81.0	TIDAK	0.833996	YA
6	Hujan	65.0	75.0	YA	0.253733	YA
7	Hujan	64.0	85.0	YA	-0.160143	TIDAK

Buka **RAPID MINER STUDIO** lalu Import data **TabelCuaca.xls**



Lalu klik next



Lalu klik next

Maka hasilnya akan seperti ini dan langsung kita ganti tipe data **binomial** di tabel **bermain_tenis**

Import Data - Format your columns.

Format your columns.

Date format: MMM d, yyyy h:mm:ss a z Replace errors with missing values

	Cuaca polynomial	Suhu integer	Kelembaban_udara integer	Berangin polynomial	Bermain_Tenis polynomial
1	Cerah	85	85	TIDAK	TIDAK
2	Cerah	80	90	YA	TIDAK
3	Mendung	83	86	TIDAK	YA
4	Hujan	70	96	TIDAK	YA
5	Hujan	68	80	TIDAK	YA
6	Hujan	65	70	YA	TIDAK
7	Mendung	64	65	YA	YA
8	Cerah	72	95	TIDAK	TIDAK
9	Cerah	69	70	TIDAK	YA
10	Hujan	75	80	TIDAK	YA
11	Cerah	75	70	YA	YA
12	Mendung	72	90	YA	YA
13	Mendung	81	75	TIDAK	YA
14	Hujan	71	91	YA	TIDAK

Change Type: **binominal**

Change Role

Rename column

Exclude column

Advanced parameters

compatibility (7.1.001)

no problems.

Previous Next Cancel

Lalu diubah juga **Change Role** menjadi **TABLE**

Import Data - Format your columns.

Format your columns.

Date format: MMM d, yyyy h:mm:ss a z Replace errors with missing values

	Cuaca polynomial	Suhu integer	Kelembaban_udara integer	Berangin polynomial	Bermain_Tenis binominal
1	Cerah	85	85	TIDAK	TIDAK
2	Cerah	80	90	YA	TIDAK
3	Mendung	83	86	TIDAK	YA
4	Hujan	70	96	YA	YA
5	Hujan	68	80	TIDAK	YA
6	Hujan	65	70	YA	TIDAK
7	Mendung	64	65	YA	YA
8	Cerah	72	95	TIDAK	YA
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	YA	TIDAK

Please enter the new role:

label
label
id
weight

no problems.

Previous Next Cancel

Maka hasilnya akan seperti ini

Import Data - Format your columns.

Format your columns.

Date format: MMM d, yyyy h:mm:ss a z Replace errors with missing values

	Cuaca polynominal	Suhu integer	Kelembaban_udara integer	Berangin polynominal	Bermain_Tenis binomial label
1	Cerah	85	85	TIDAK	TIDAK
2	Cerah	80	90	YA	TIDAK
3	Mendung	83	86	TIDAK	YA
4	Hujan	70	96	TIDAK	YA
5	Hujan	68	80	TIDAK	YA
6	Hujan	65	70	YA	TIDAK
7	Mendung	64	65	YA	YA
8	Cerah	72	95	TIDAK	TIDAK
9	Cerah	69	70	TIDAK	YA
10	Hujan	75	80	TIDAK	YA
11	Cerah	75	70	YA	YA
12	Mendung	72	90	YA	YA
13	Mendung	81	75	TIDAK	YA
14	Hujan	71	91	YA	TIDAK

no problems.

Lalu klik next dan kita save dengan nama file **DataCuaca_Training**

Studio Basic 7.1.001 @ INTEL-PC

Import Data - Where to store the data?

Where to store the data?

- Local Repository (INTEL)
 - data (INTEL)
 - processes (INTEL)
- Cloud Repository (disconnected)

Name: DataCuaca_Training

Location: //Local Repository/DataCuaca_Training

Hasil importnya akan seperti ini

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

Row No.	Bermain_Terbatas	Cuaca	Suhu	Kelembaban_udara	Berangin
1	TIDAK	Cerah	85	85	TIDAK
2	TIDAK	Cerah	80	90	YA
3	YA	Mendung	83	86	TIDAK
4	YA	Hujan	70	96	TIDAK
5	YA	Hujan	68	80	TIDAK
6	TIDAK	Hujan	65	70	YA
7	YA	Mendung	64	65	YA
8	TIDAK	Cerah	72	95	TIDAK
9	YA	Cerah	69	70	TIDAK
10	YA	Hujan	75	80	TIDAK
11	YA	Cerah	75	70	YA
12	YA	Mendung	72	90	YA
13	YA	Mendung	81	75	TIDAK
14	TIDAK	Hujan	71	91	YA

Lalu klik jendela design dan kita lakukan langkah yang sama namun yang kita import adalah yang Sheet2

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Sheet2 Cell range: A1 Define header row: 1

A	B	C	D
1 Cuaca	Suhu	Kelembaban_udara	Berangin
2 Cerah	75,000	65,000	TIDAK
3 Cerah	80,000	68,000	YA
4 Cerah	83,000	87,000	YA
5 Mendung	70,000	96,000	TIDAK
6 Mendung	68,000	81,000	TIDAK
7 Hujan	65,000	75,000	YA
8 Hujan	64,000	85,000	YA

Pada langkah ini jangan dirubah **tipenya** dan **jenis role nya**

Import Data - Format your columns.

Format your columns.

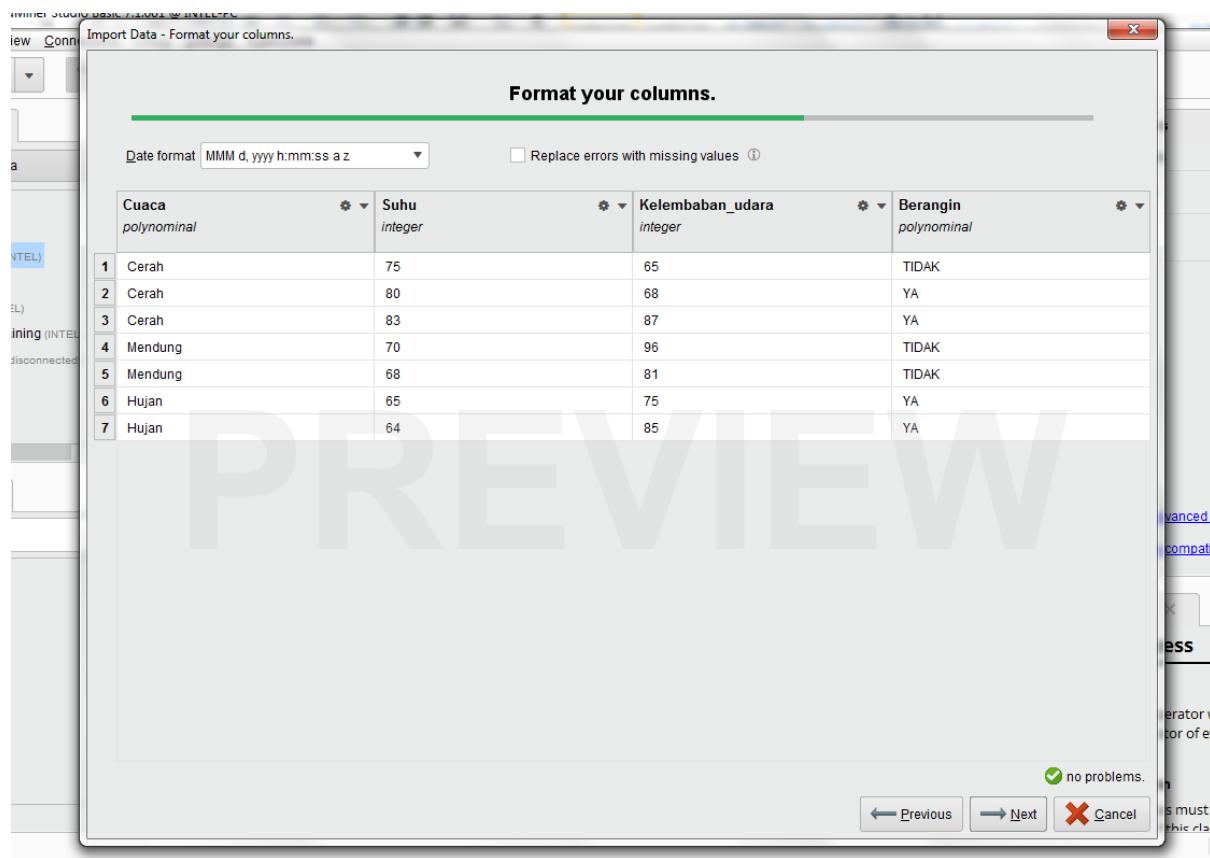
Date format: MMM d, yyyy h:mm:ss a z

Replace errors with missing values

	Cuaca polynomial	Suhu integer	Kelembaban_udara integer	Berangin 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



Dan kita save dengan nama file **DataCuaca_Testing**

Import Data - Where to store the data?

Where to store the data?

Local Repository (INTEL)

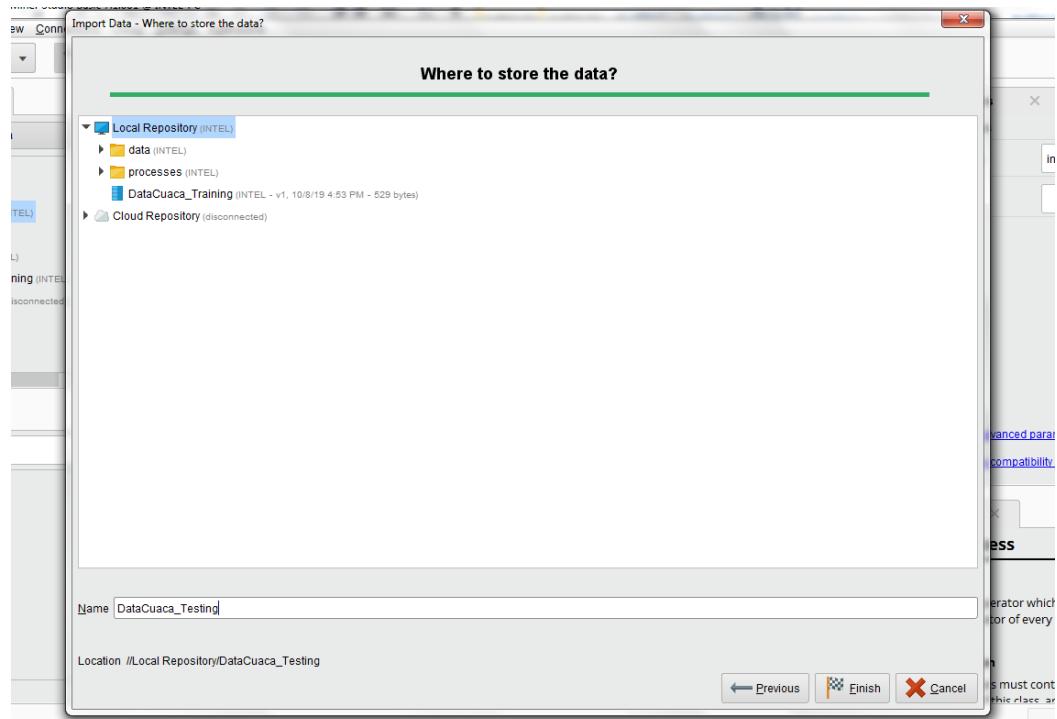
- data (INTEL)
- processes (INTEL)
- DataCuaca_Training (INTEL - v1, 10/8/19 4:53 PM - 529 bytes)

Cloud Repository (disconnected)

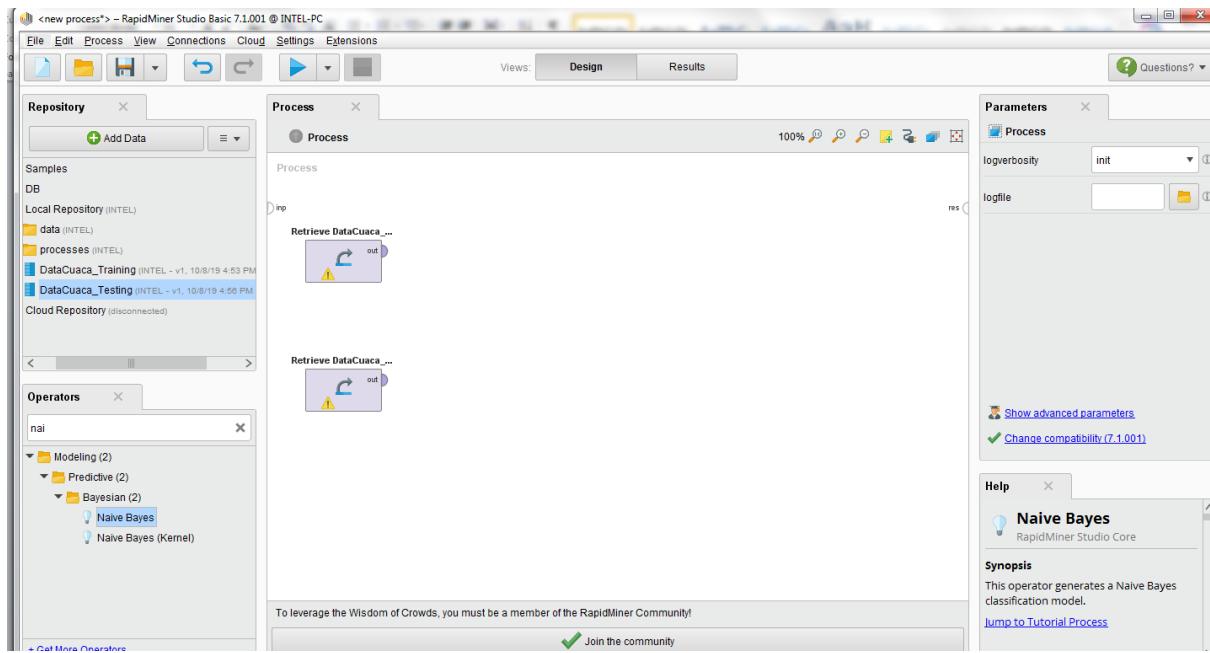
Name: DataCuaca_Testing

Location: //Local Repository/DataCuaca_Testing

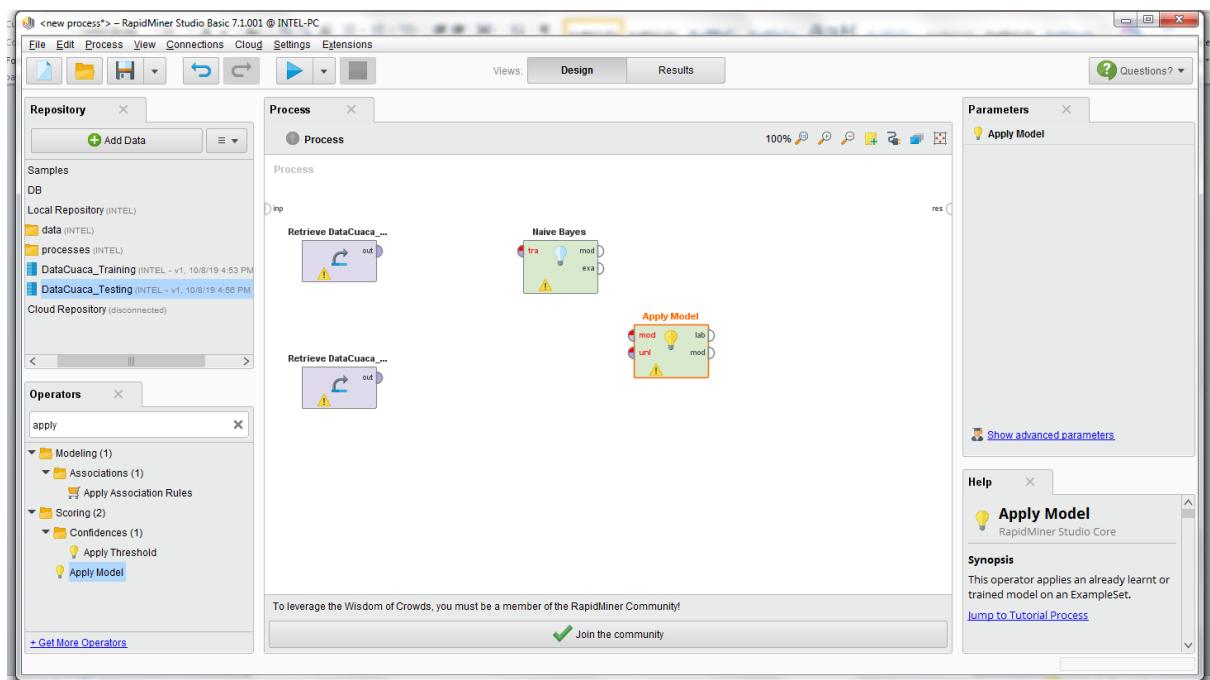
Previous Finish Cancel



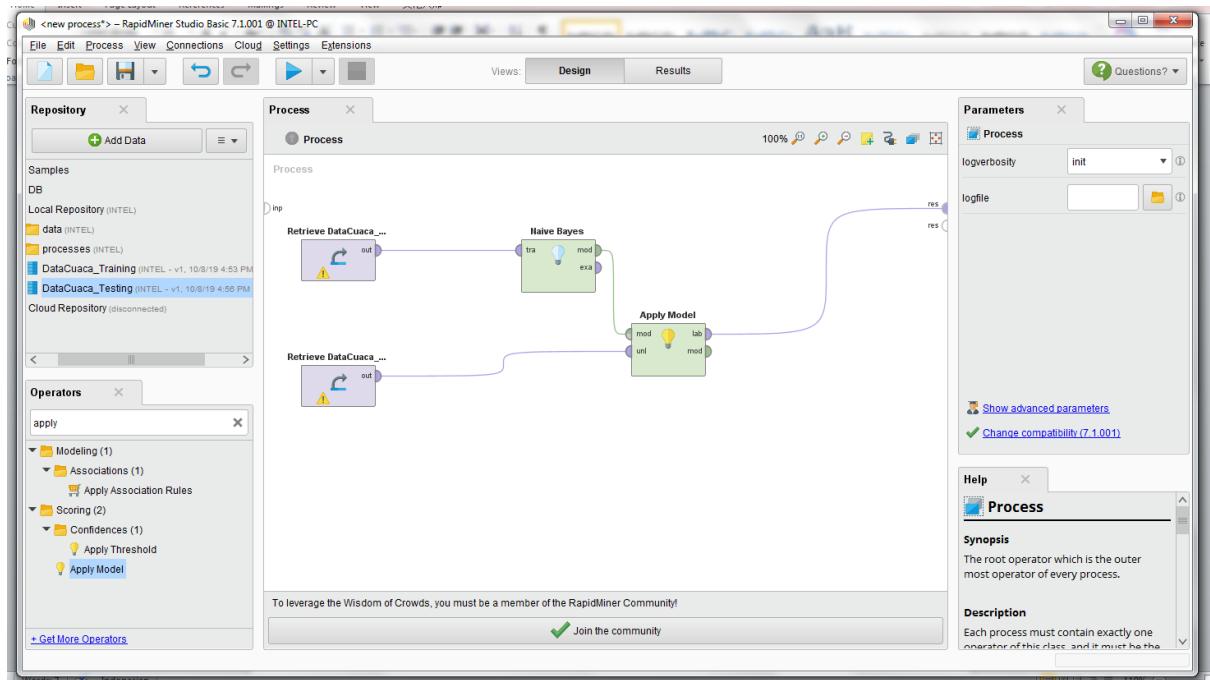
Dan kalau sudah kita kembali ke halaman design dan masukkan **DataCuaca_Testing** dan **DataCuaca_Training**



Lalu masukkan Naive Bayes dan Apply Model



Dan kita susun rangakaianya seperti ini



Lalu kita jalankan dan hasilnya akan seperti ini

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

- Data**: Displays two ExampleSets: "ExampleSet (/Local Repository/DataCuaca_Testing)" and "ExampleSet (Retrieve DataCuaca_Testing)".
- Statistics**: Shows the content of the ExampleSets:

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

- Annotations**: A panel on the left side of the interface.

Lalu kita klik statistik dan hasilnya akan seperti ini

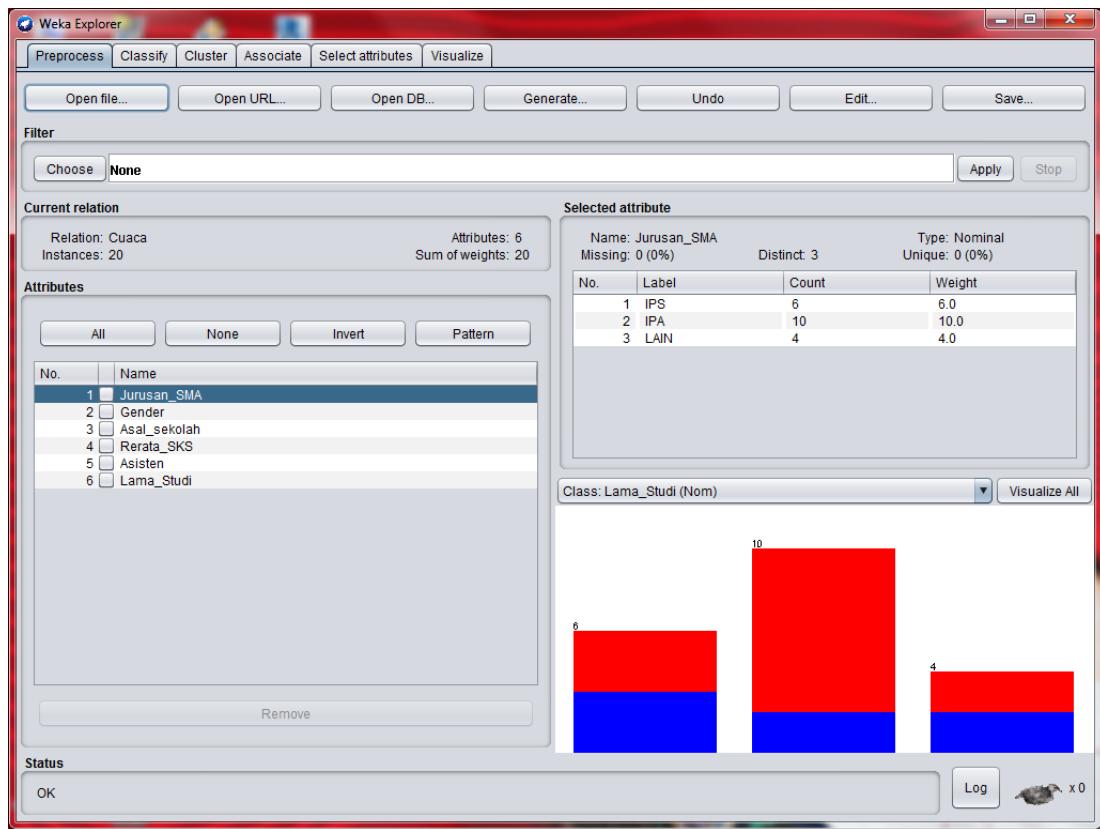
The screenshot shows the RapidMiner Studio Basic 7.1.001 interface. The main window title is "<new process> - RapidMiner Studio Basic 7.1.001 @ INTEL-PC". The menu bar includes File, Edit, Process, View, Connections, Cloud, Settings, and Extensions. The toolbar contains icons for opening files, saving, running processes, and other operations. The left sidebar has tabs for Data, Statistics, Charts, Advanced Charts, and Annotations. The central area is titled "ExampleSet (/Local Repository/DataCuaca_Testing)" and shows a table of attributes:

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

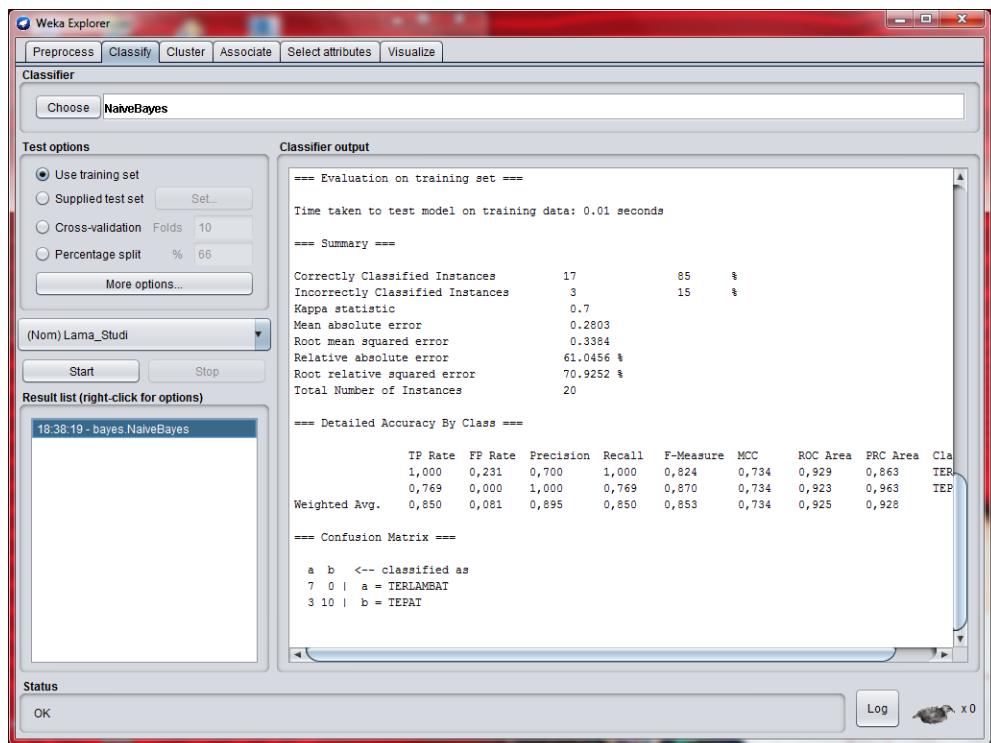
At the bottom, it says "Showing attributes 1 - 7 Examples: 7 Special Attributes: 3 Regular Attributes: 4". The right sidebar is titled "Repository" and shows a tree structure with "Samples", "DB", and "Local Repository (INTEL)" containing "data (INTEL)", "processes (INTEL)", "DataCuaca_Training (INTEL - v1, 10/8/19 4:45 PM)", and "DataCuaca_Testing (INTEL - v1, 10/8/19 4:45 PM)". It also shows a "Cloud Repository (disconnected)".

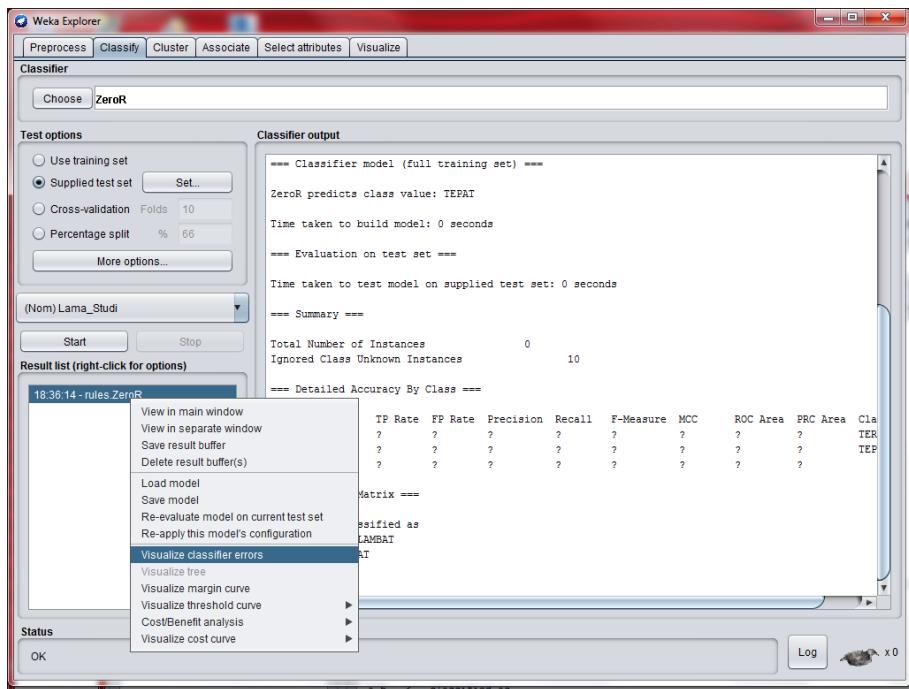
TUGAS

Buka WEKA dna open file data SMA

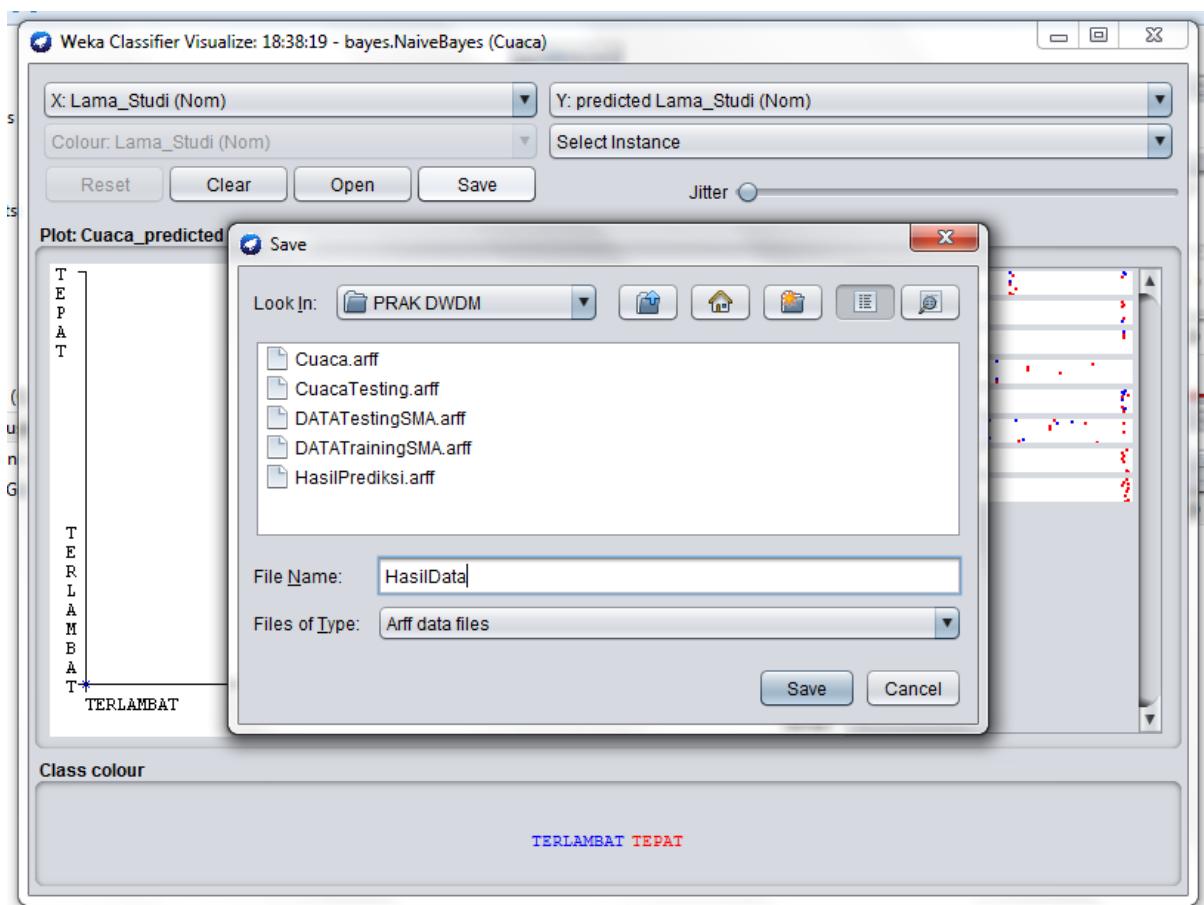


lalu klik Classify dan lakukan seperti latihan diatas





Klik Start dan simpan dengan nama **HasilData**



CLOSE semua laman WEKA sampai WEKA EXPLORER

Dan kita pilih **ArffViewers** Maka tampilannya akan seperti ini

Dan kita open file data yang tadi kita simpan yaitu **HasilData.arff** lalu klik open

No.	1: Jurusan_SMA	2: Gender	3: Asal_sekolah	4: Rerata_SKS	5: Asisten	6: prediction margin	7: predicted Lama_Studi	8: Lama_Studi
	Nominal	Nominal	Nominal	Numeric	Nominal	Numeric	Nominal	Nominal
1	IPS	WANITA	SURAKARTA	18.0	TIDAK	0.375862	TERLAMBAT	TERLAMBAT
2	IPA	PRIA	SURAKARTA	19.0	YA	0.836469	TEPAT	TEPAT
3	LAIN	PRIA	SURAKARTA	19.0	TIDAK	0.175169	TERLAMBAT	TERLAMBAT
4	IPA	PRIA	LUAR	17.0	TIDAK	0.331376	TERLAMBAT	TERLAMBAT
5	IPA	WANITA	SURAKARTA	17.0	TIDAK	-0.064476	TERLAMBAT	TEPAT
6	IPA	WANITA	LUAR	18.0	YA	0.757815	TEPAT	TEPAT
7	IPA	PRIA	SURAKARTA	18.0	TIDAK	0.125076	TERLAMBAT	TERLAMBAT
8	IPA	PRIA	SURAKARTA	19.0	TIDAK	0.356012	TEPAT	TEPAT
9	IPS	PRIA	LUAR	18.0	TIDAK	0.588286	TERLAMBAT	TERLAMBAT
10	LAIN	WANITA	SURAKARTA	18.0	TIDAK	-0.375862	TERLAMBAT	TEPAT
11	IPA	WANITA	SURAKARTA	19.0	TIDAK	0.573095	TEPAT	TEPAT
12	IPS	PRIA	SURAKARTA	20.0	TIDAK	0.53916	TEPAT	TEPAT
13	IPS	PRIA	SURAKARTA	19.0	TIDAK	-0.175169	TERLAMBAT	TEPAT
14	IPA	PRIA	SURAKARTA	19.0	TIDAK	0.356012	TEPAT	TEPAT
15	IPA	PRIA	LUAR	22.0	YA	0.999709	TEPAT	TEPAT
16	LAIN	PRIA	SURAKARTA	16.0	TIDAK	0.681399	TERLAMBAT	TERLAMBAT
17	IPS	PRIA	LUAR	20.0	TIDAK	0.53916	TEPAT	TEPAT
18	LAIN	PRIA	LUAR	23.0	YA	0.999968	TEPAT	TEPAT
19	IPA	PRIA	SURAKARTA	21.0	YA	0.995581	TEPAT	TEPAT
20	IPS	PRIA	SURAKARTA	19.0	TIDAK	0.175169	TERLAMBAT	TERLAMBAT

Buka **RAPID MINER STUDIO** lalu import data **SMA** dan lakukan hal yang sama seperti latihan

The screenshot shows the RapidMiner Studio interface with the 'Import Data' dialog open. The dialog displays a preview of the data from 'Sheet2' with a cell range of A:F. The data consists of 21 rows and 6 columns. The columns are labeled: Jurusan_SMA, Gender, Asal_Sekolah, Rerata_Sekolah, Asisten, and Lama_Studi. The 'Select All' and 'Define header row' checkboxes are checked. The header row value is set to 1. The background shows the repository tree and process editor.

Dan data dirubah dengan binomial dan dirubah ke label

Import Data - Format your columns.

Format your columns.

Date format: MMM d, yyyy h:mm:ss a z Replace errors with missing values

Cuaca polynominal	Suhu integer	Kelembaban_udara integer	Berangin polynominal	Bermain_Tenis binomial label
1 Cerah	85	85	TIDAK	TIDAK
2 Cerah	80	90	YA	TIDAK
3 Mendung	83	86	TIDAK	YA
4 Hujan	70	96	TIDAK	YA
5 Hujan	68	80	TIDAK	YA
6 Hujan	65	70	YA	TIDAK
7 Mendung	64	65	YA	YA
8 Cerah	72	95	TIDAK	TIDAK
9 Cerah	69	70	TIDAK	YA
10 Hujan	75	80	TIDAK	YA
11 Cerah	75	70	YA	YA
12 Mendung	72	90	YA	YA
13 Mendung	81	75	TIDAK	YA
14 Hujan	71	91	YA	TIDAK

no problems. [Previous](#) [Next](#) [Cancel](#)

Lalu next

<new process> - RapidMiner Studio Basic 7.1.001 @ INTEL-PC

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

ExampleSet (/Local Repository/DataCuaca_Testing) ExampleSet (/Local Repository/DataCuaca_Training) ExampleSet (Retrieve DataCuaca_Testing)

Result History ExampleSet (/Local Repository/DataSMA_Training) Filter (20 / 20 examples): all

Data Statistics Charts Advanced Charts Annotations

Repository

- Samples
- DB
- Local Repository (INTEL)
 - data (INTEL)
 - processes (INTEL)
 - DataCuaca_Training (INTEL - v1, 10/8/19 4:45 PM)
 - DataCuaca_Testing (INTEL - v1, 10/8/19 4:45 PM)
 - DataSMA_Training (INTEL - v1, 10/8/19 7:00 PM)
- Cloud Repository (disconnected)

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

Row No.	Lama_Studi	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_Sek...	Asisten
1	TERLAMBAT	IPS	WANITA	SURAKARTA	18	TIDAK
2	TEPAT	IPA	PRIA	SURAKARTA	19	YA
3	TERLAMBAT	LAIN	PRIA	SURAKARTA	19	TIDAK
4	TERLAMBAT	IPA	PRIA	LUAR	17	TIDAK
5	TEPAT	IPA	WANITA	SURAKARTA	17	TIDAK
6	TEPAT	IPA	WANITA	LUAR	18	YA
7	TERLAMBAT	IPA	PRIA	SURAKARTA	18	TIDAK
8	TEPAT	IPA	PRIA	SURAKARTA	19	TIDAK
9	TERLAMBAT	IPS	PRIA	LUAR	18	TIDAK
10	TEPAT	LAIN	WANITA	SURAKARTA	18	TIDAK
11	TEPAT	IPA	WANITA	SURAKARTA	19	TIDAK
12	TEPAT	IPS	PRIA	SURAKARTA	20	TIDAK
13	TEPAT	IPS	PRIA	SURAKARTA	19	TIDAK
14	TEPAT	IPA	PRIA	SURAKARTA	19	TIDAK
15	TEPAT	IPA	PRIA	LUAR	22	YA
16	TERLAMBAT	LAIN	PRIA	SURAKARTA	16	TIDAK
17	TEPAT	IPS	PRIA	LUAR	20	TIDAK
18	TEPAT	LAIN	PRIA	LUAR	23	YA

Lalu buka kembali data SMA tadi namun diganti ke Sheet1

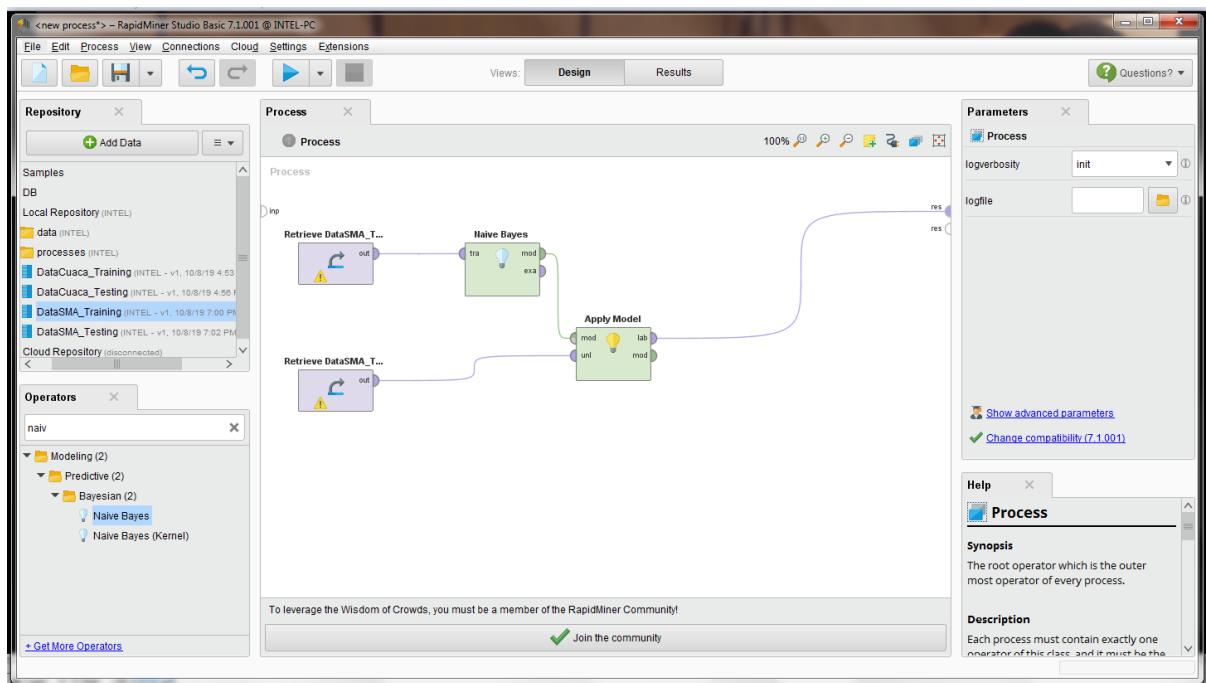
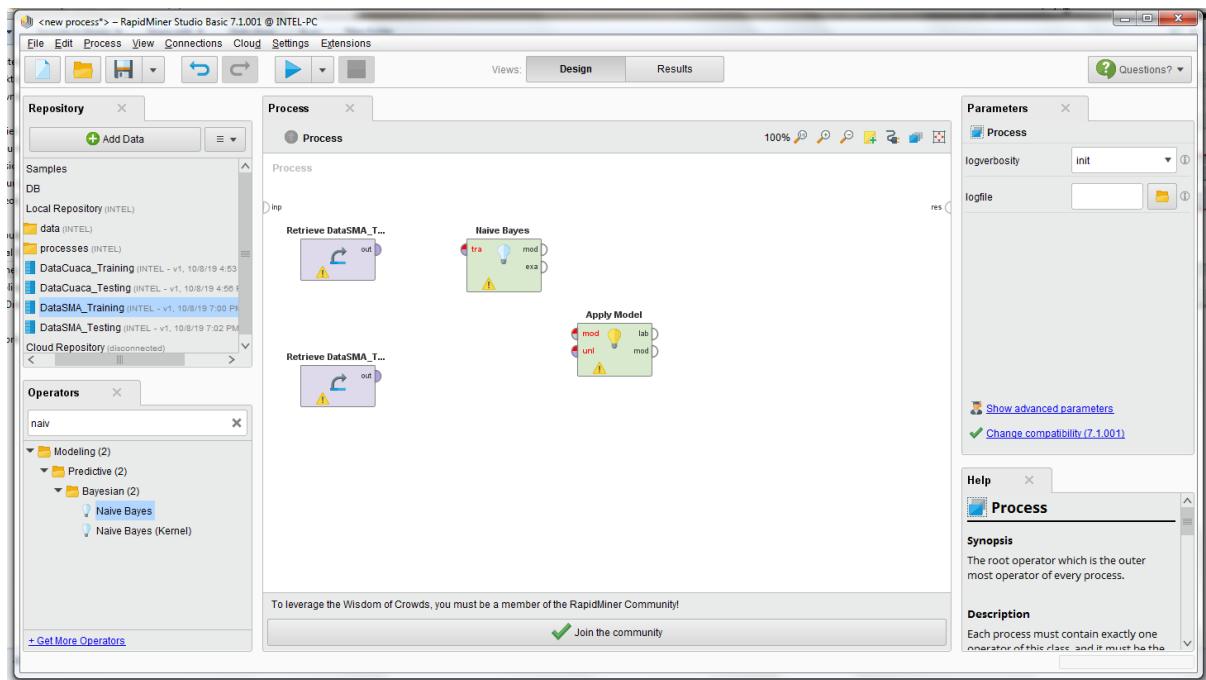
Screenshot of RapidMiner Studio Basic 7.1.001 showing the 'Import Data - Select the cells to import' dialog. The dialog displays a table with 11 rows and 5 columns (A-E). The columns are labeled: Jurusan_SMA, Gender, Asal_Sekolah, Rerata_Sekolah, and Asisten.

A	B	C	D	E
1 Jurusan_SMA	Gender	Asal_Sekolah	Rerata_Sekolah	Asisten
2 LAIN	WANITA	SURAKARTA	18.000	TIDAK
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

Screenshot of RapidMiner Studio Basic 7.1.001 showing the 'Import Data - Format your columns' dialog. The dialog displays the same table as the previous screen, but with column types specified: Jurusan_SMA (polynomial), Gender (polynomial), Asal_Sekolah (polynomial), Rerata_Sekolah (integer), and Asisten (polynomial).

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

Jika sudah semua lalu susunkan lagi seperti latihan diatas



Lalu jalankan

Dan hasilnya

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

Row No.	prediction(Lama_Studi)	confidence(TERLAMBAT)	confidence(TEPAT)	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_Sekolah	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

Dan halaman statistiknya

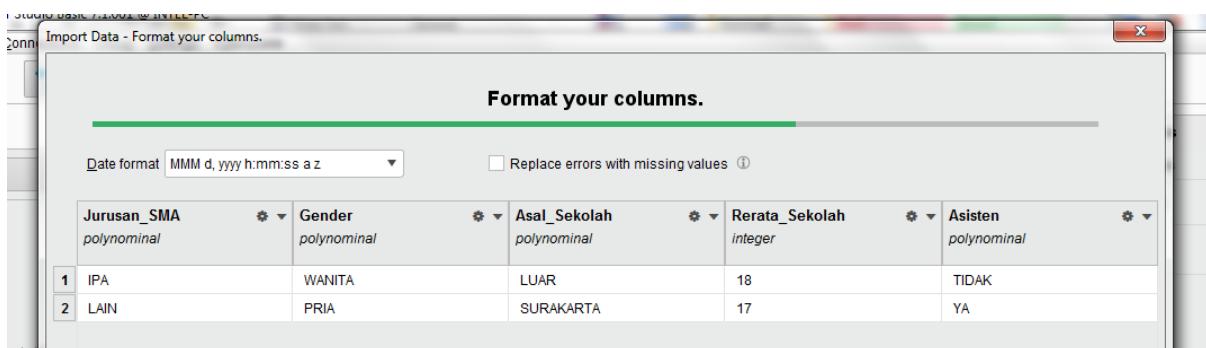
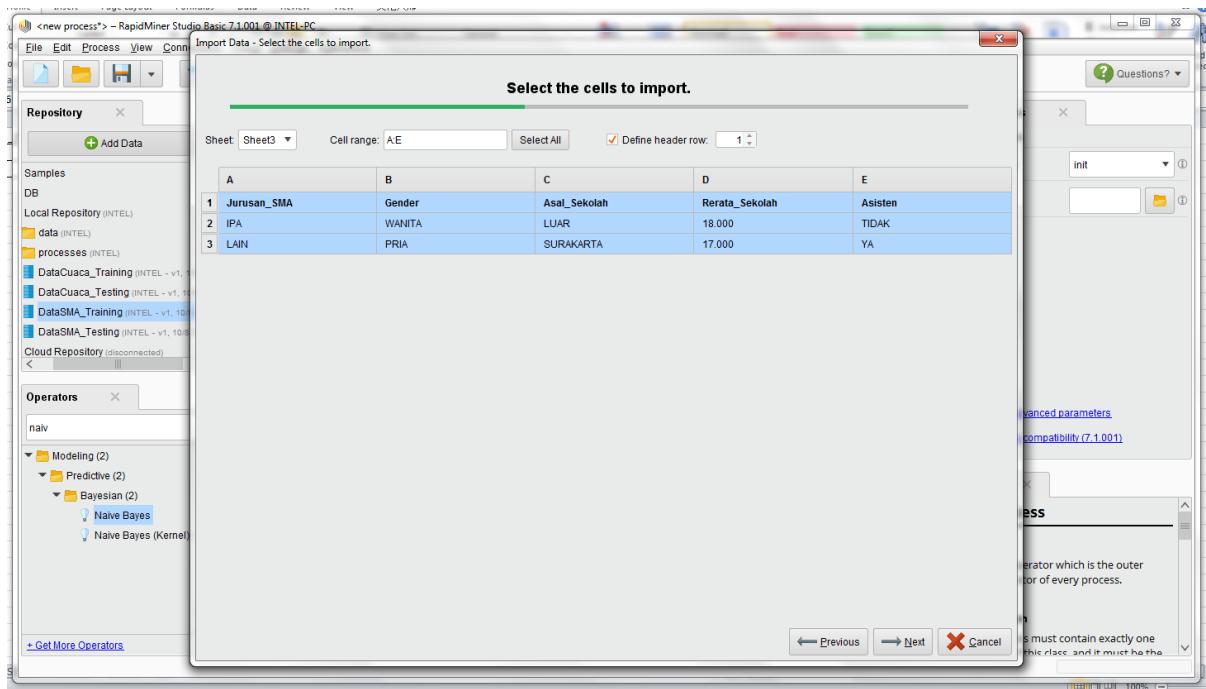
ExampleSet (8 attributes)

Name	Type	Missing	Statistics	Filter (8 / 8 attributes): Search for Attributes
Prediction				
prediction(Lama_Studi)	Binomial	0		Least TEPAT (3) Most TERLAMBAT (7)
Open chart				
Confidence_TERLAMBAT	Real	0	Min 0.005 Max 0.868 Average 0.524	
confidence(TERLAMBAT)	Real	0	Min 0.005 Max 0.868 Average 0.524	
Confidence_TEPAT	Real	0	Min 0.132 Max 0.995 Average 0.476	
confidence(TEPAT)	Real	0	Min 0.132 Max 0.995 Average 0.476	
Jurusan_SMA	Polynominal	0	Least IPS (2) Most IPA (4)	Values IPA (4), LAIN (4), ... [1 more]
Gender	Polynominal	0	Least WANITA (4) Most PRIA (6)	Values PRIA (6), WANITA (4)
Asal_Sekolah	Polynominal	0	Least LUAR (3) Most SURAKARTA (7)	Values SURAKARTA (7), LUAR (3)
Rerata_Sekolah	Integer	0	Min 17 Max 19 Average 18.100	

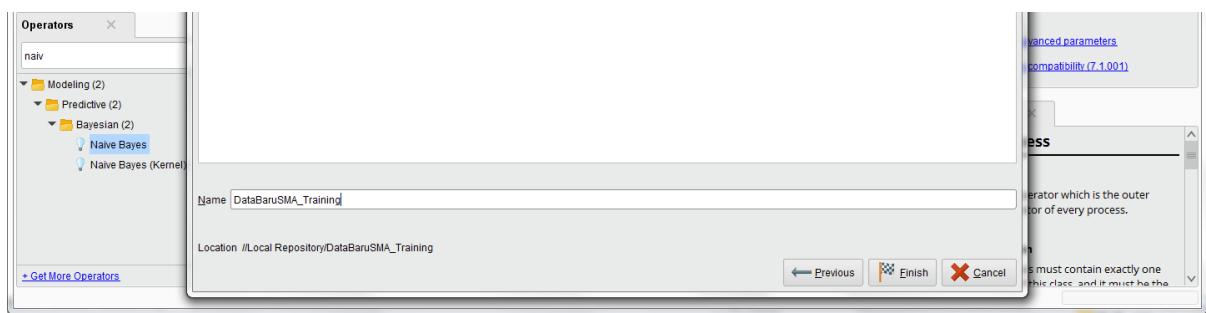
Dan disitu tertera confidence untuk **lama_studi** dengan nilai **tepat**

Dan tertera confidence untuk **lama_studi** dengan nilai **terlambat**

Untuk soal 6 dan 7 kita buat sheet baru dan lakukan sama seperti sebelumnya

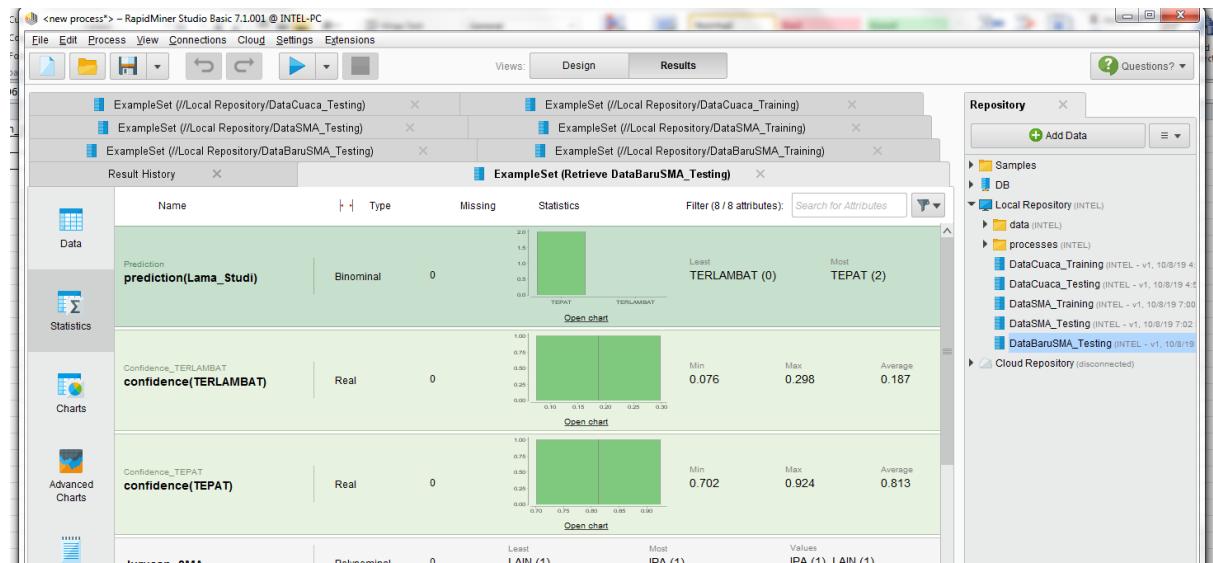


Dan save dengan nama DataBaruSMA_Training



Lalu susunlah dengan sheet1 dengan sheet yang baru

Dan hasilnya



Dan tertera juga hasil prediksi ketetapan dari lama studi dewi dan si Jono