MODULE 8

CLASSIFICATION: NAÏVE BAYES

Sort by :

Hamzah L200154013

\

Informatics Study Program

Faculty of Communication and Informatics

Muhammadiyah University of Surakarta

Praktikum Steps

Naïve Bayes Implementations using Weka.

1. Prepare **Cuaca.arff** file in module 7, the file will be used as a training data.

Cuaca.arff

C: > Users > Hamzah > Documents > Hamzah Backup > Kuliah >

```
@relation Cuaca

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

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

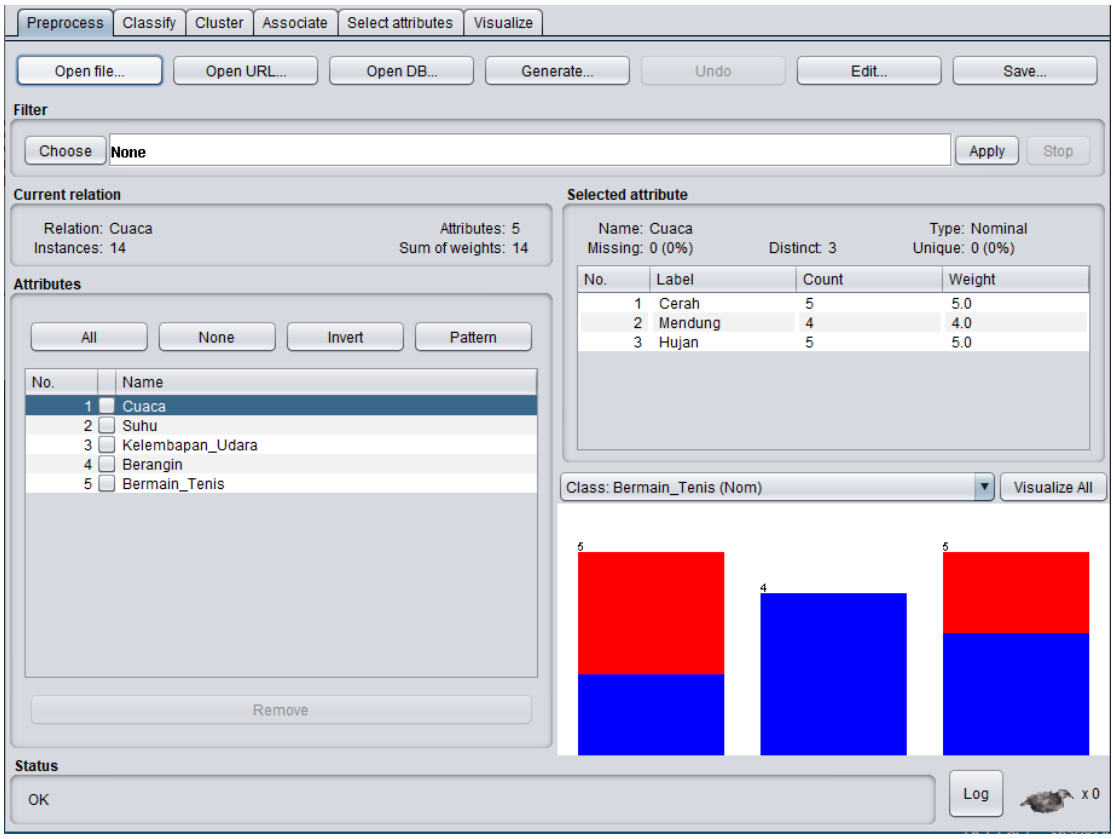
2. Creating a test data with the arff format as a test data as shown below, save it with the name **CuacaTesting.arff**.

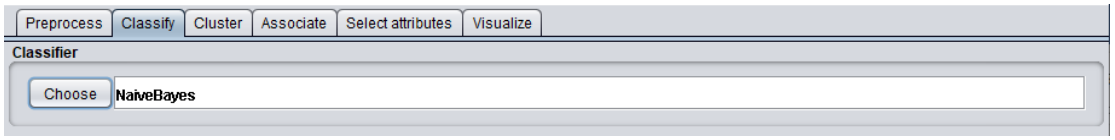
CuacaTesting.arff

CuacaTesting.arff

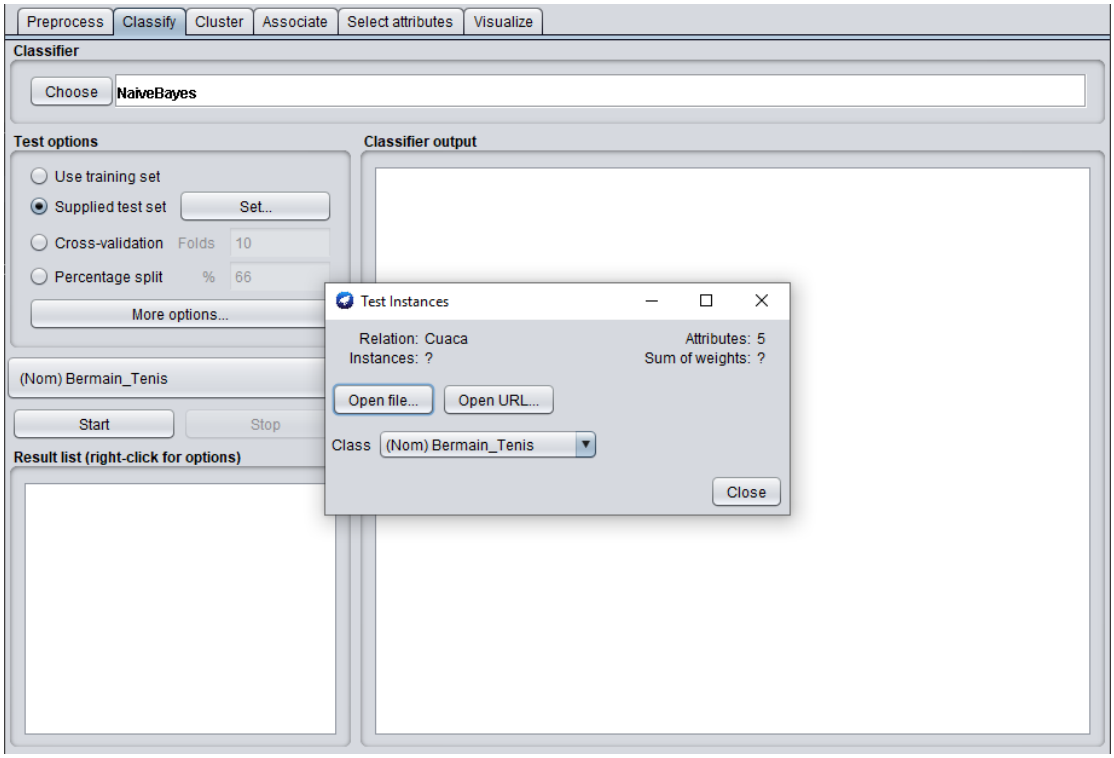
```
@relation Cuaca

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

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

3. Enter the **Cuaca.arff** file in the preprocessing tab, select the classify tab then in the classifier box click the choose button to select the Naïve Bayes method / algorithm.

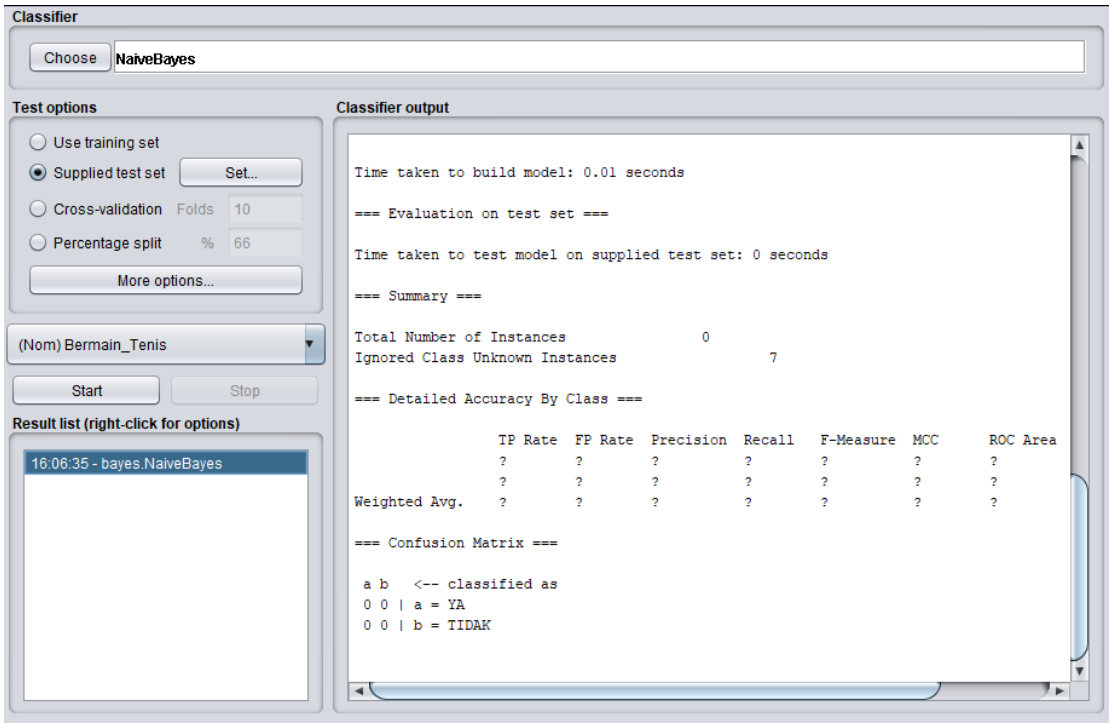




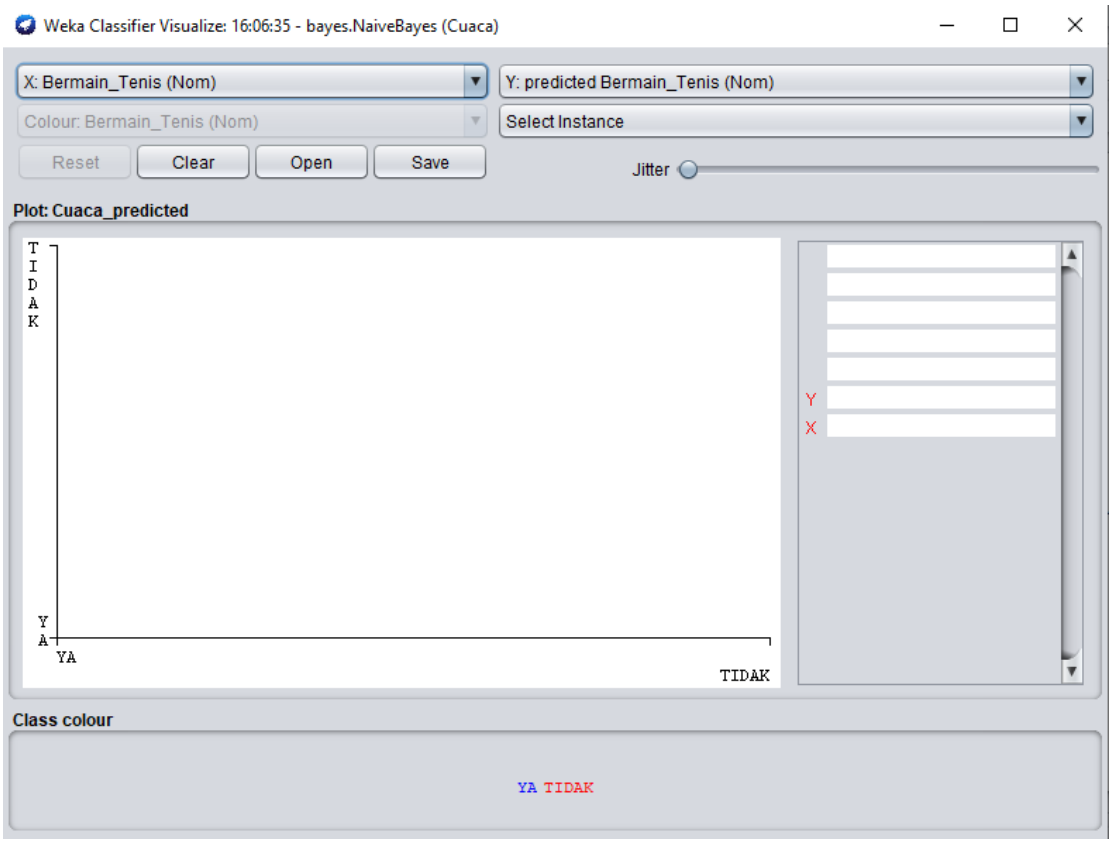
4. On the test option select the **Supplied test set**, then click on the box **set ...**
5. The test instance window will appear. Click open file, enter the file **CuacaTesting.arff**.



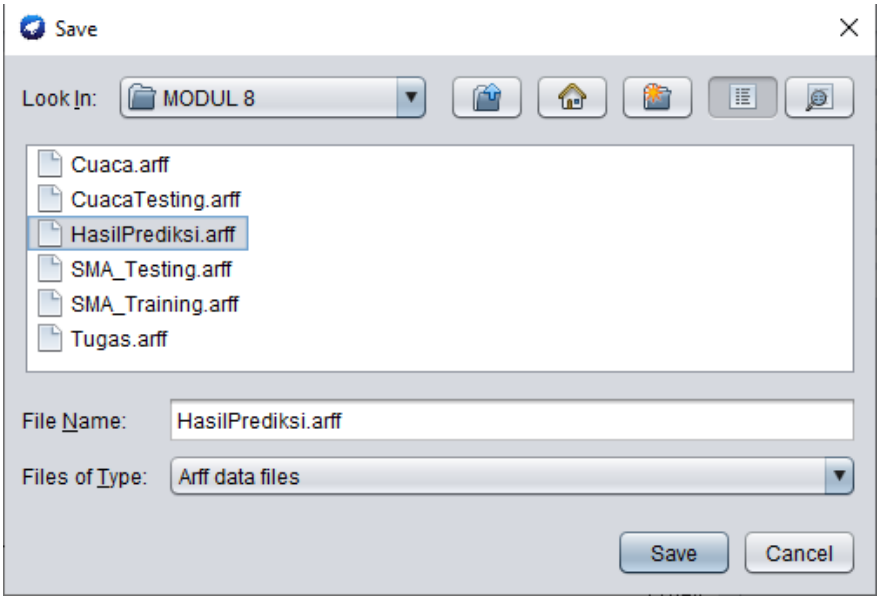
6. Click Start to start the Naïve Bayes process.



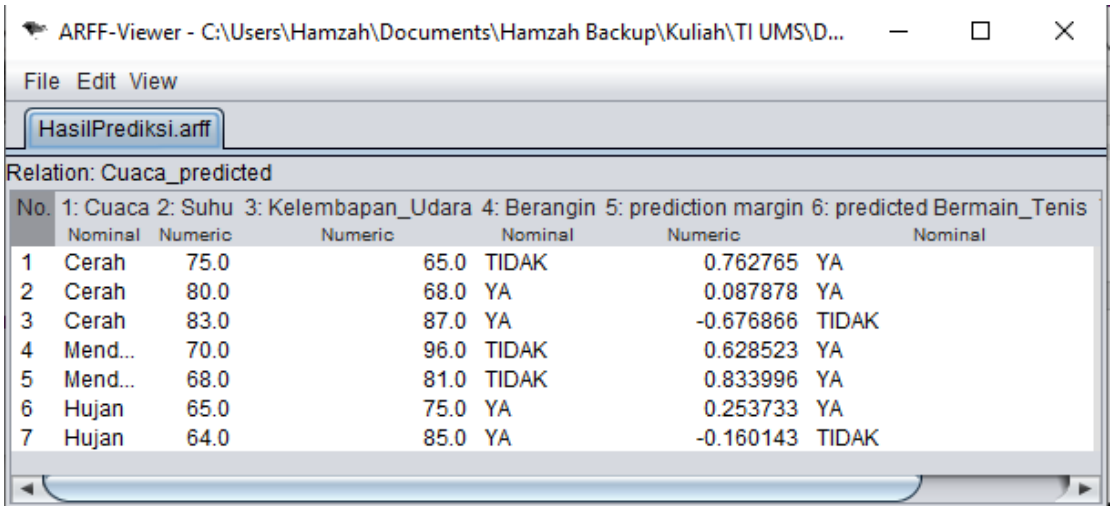
7. Right-click on the results of the process in the result list box, then select Visualize classifier errors.



8. Then save the file with the name **HasilPrediksi.arff**.



9. Close all windows including weka explorer and return to weka gui chooser. Select the tools menu - **ArffViewer**.
10. Open the **HasilPrediksi.arff** file, then the tennis prediction data will come out.



No.	1: Cuaca	2: Suhu	3: Kelembapan_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

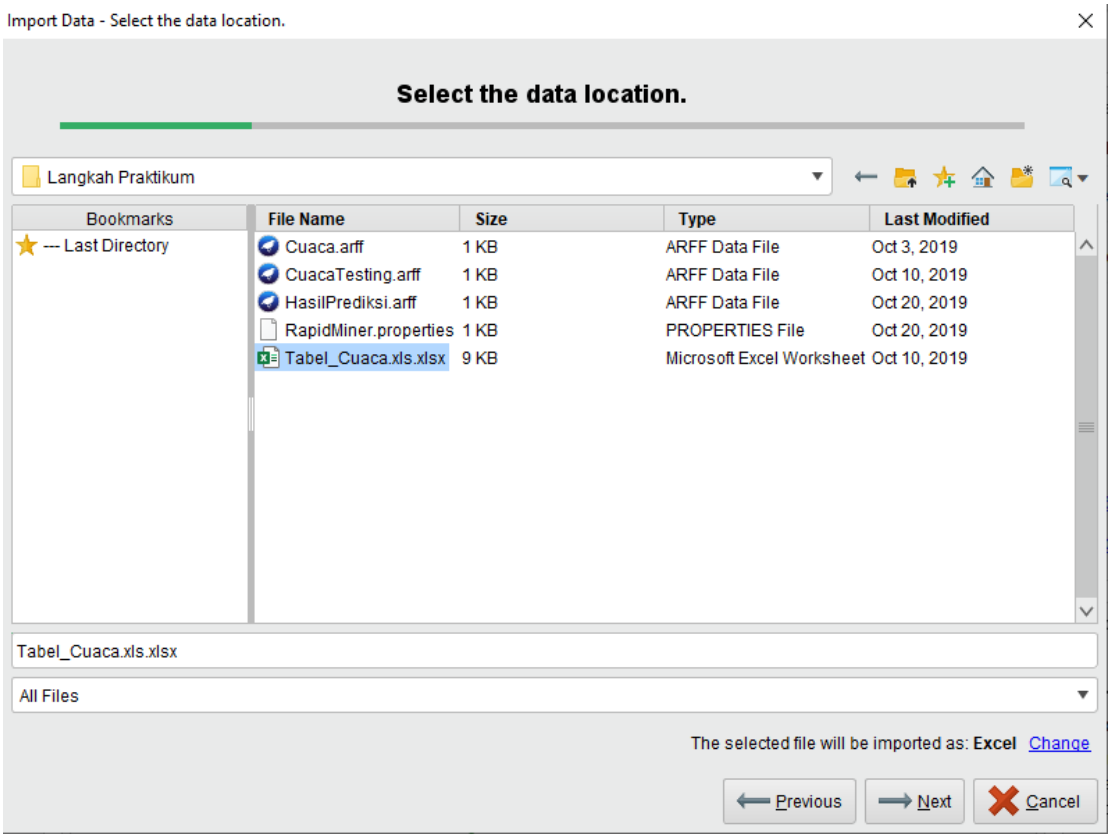
Naïve Bayes Implementation using RapidMiner.

1. Prepare file **TabelCuaca.xls** consisting of 2 sheets, sheet1 as training data and sheet2 as data testing.

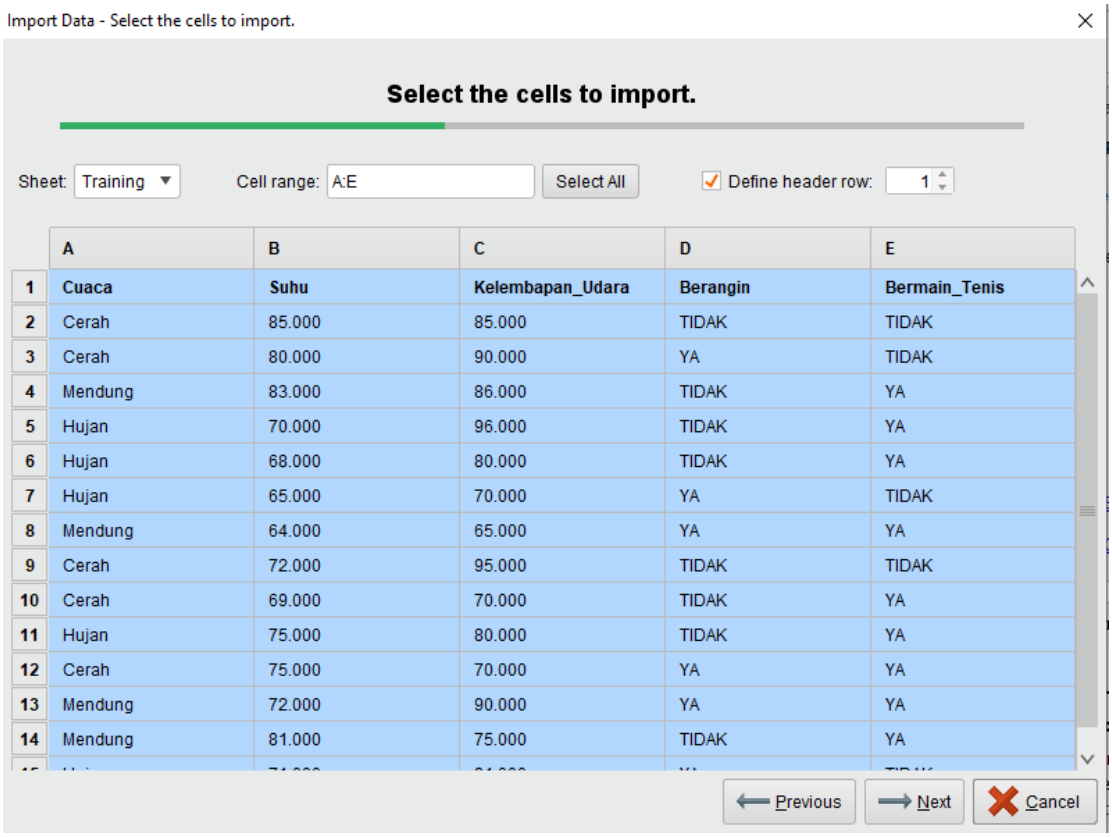
	A	B	C	D	E
1	Cuaca	Suhu	Kelembapan_Udara	Berangin	Bermain_Tenis
2	Cerah	85	85	TIDAK	TIDAK
3	Cerah	80	90	YA	TIDAK
4	Mendung	83	86	TIDAK	YA
5	Hujan	70	96	TIDAK	YA
6	Hujan	68	80	TIDAK	YA
7	Hujan	65	70	YA	TIDAK
8	Mendung	64	65	YA	YA
9	Cerah	72	95	TIDAK	TIDAK
10	Cerah	69	70	TIDAK	YA
11	Hujan	75	80	TIDAK	YA
12	Cerah	75	70	YA	YA
13	Mendung	72	90	YA	YA
14	Mendung	81	75	TIDAK	YA
15	Hujan	71	91	YA	TIDAK
16					
17					
18					
19					
20					
21					
22					
23					

	A	B	C	D	E
1	Cuaca	Suhu	Kelembapan_Udara	Berangin	Bermain_Tenis
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	

2. Open rapidminer application then press add data button, select **TabelCuaca.xls**.

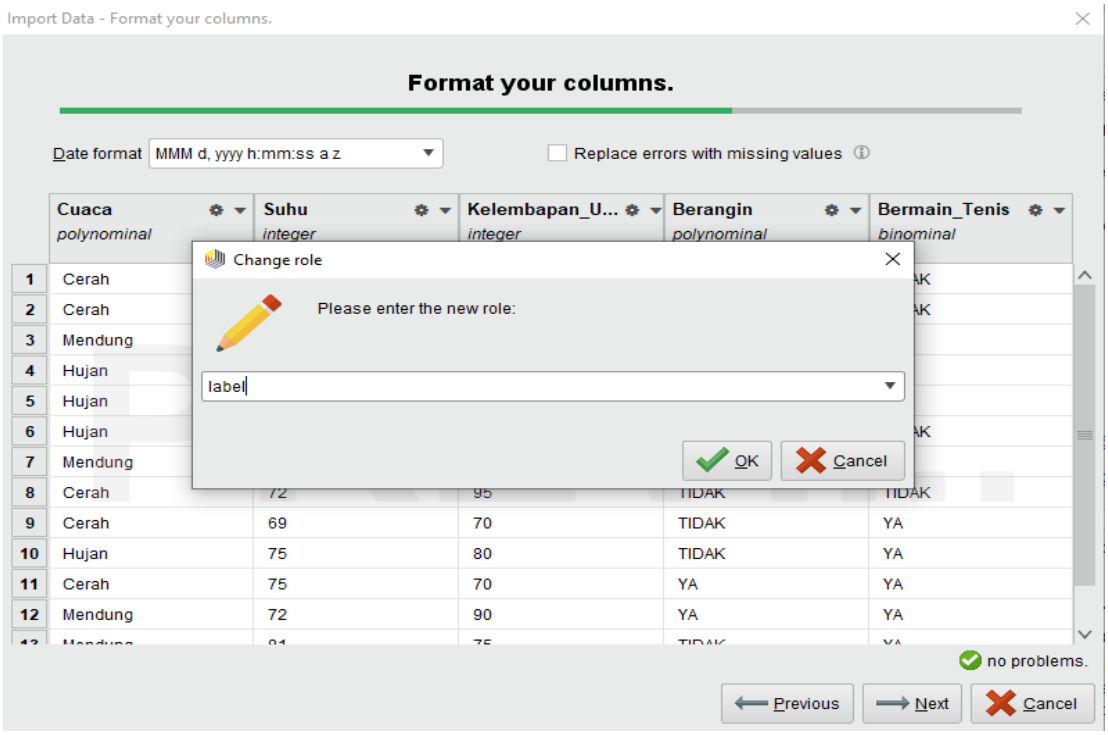


3. First we choose sheet1 / training data, block all the tables inside the training data, then click next.



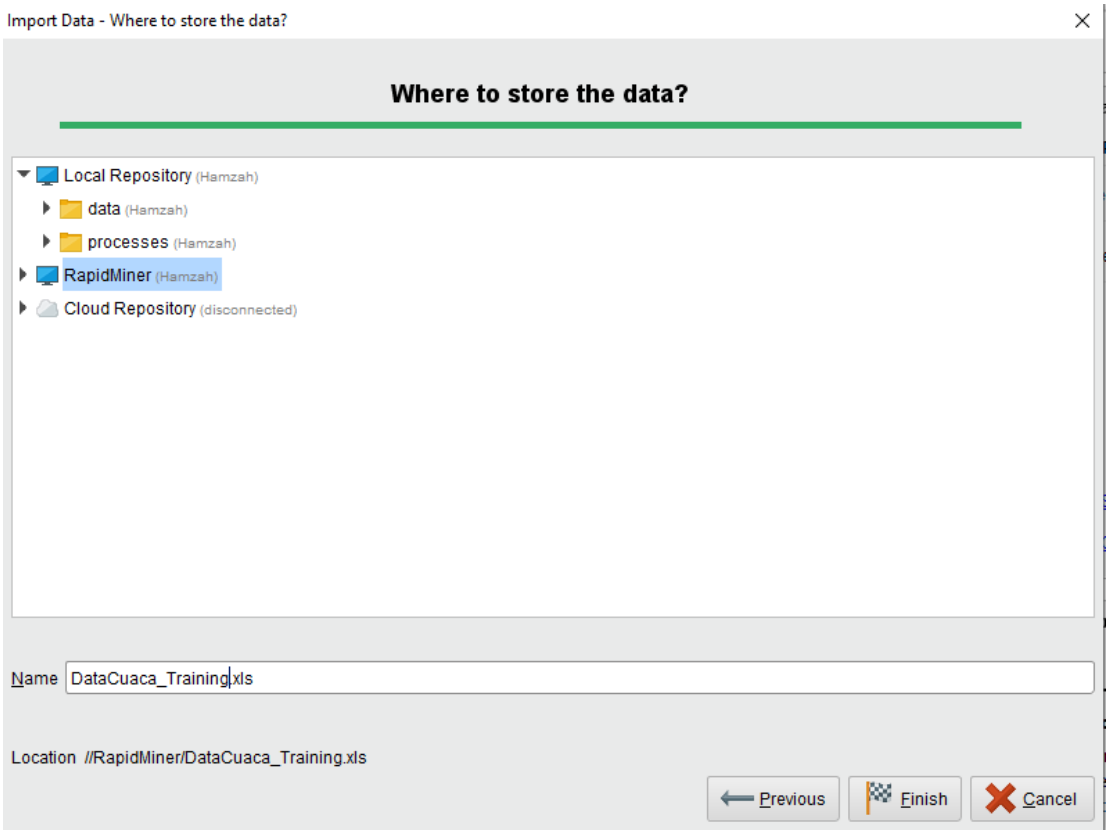
	A	B	C	D	E
1	Cuaca	Suhu	Kelembapan_Udara	Berangin	Bermain_Tenis
2	Cerah	85.000	85.000	TIDAK	TIDAK
3	Cerah	80.000	90.000	YA	TIDAK
4	Mendung	83.000	86.000	TIDAK	YA
5	Hujan	70.000	96.000	TIDAK	YA
6	Hujan	68.000	80.000	TIDAK	YA
7	Hujan	65.000	70.000	YA	TIDAK
8	Mendung	64.000	65.000	YA	YA
9	Cerah	72.000	95.000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

4. Change the type and role in the **bermain_tenis** column with binomial types (because there are only 2 data) and role with labels, then click next.

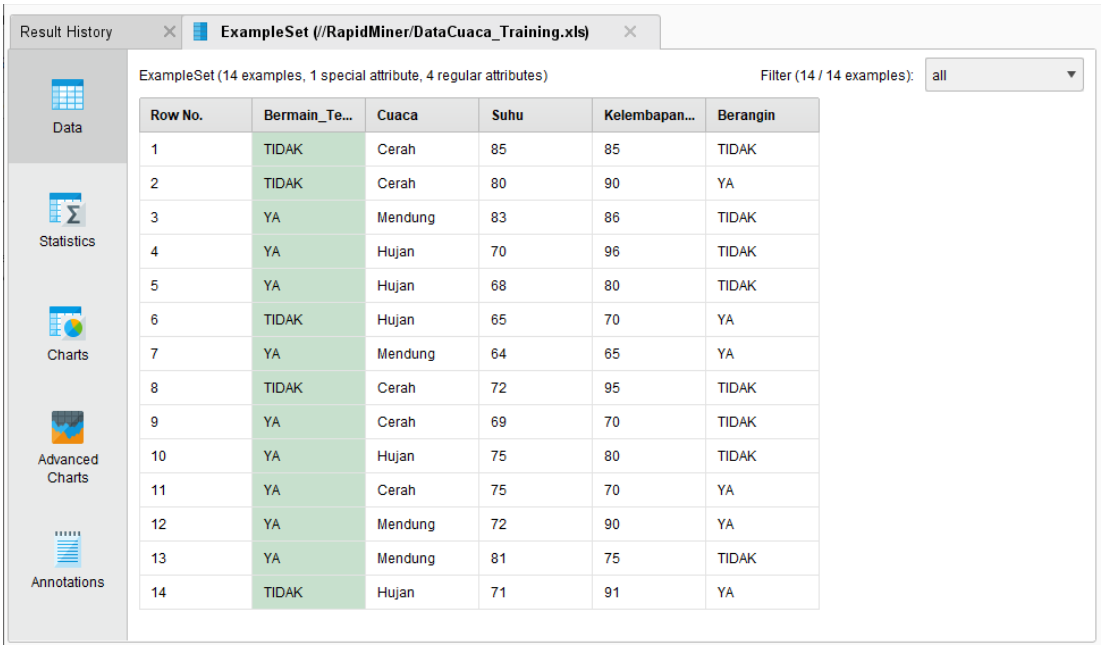


	Cuaca *polynominal*	Suhu *integer*	Kelembapan_U... *integer*	Berangin *polynominal*	Bermain_Tenis *binominal*
1	Cerah				
2	Cerah				
3	Mendung				
4	Hujan				
5	Hujan				
6	Hujan				
7	Mendung				
8	Cerah	72	95	TIDAK	TIDAK
9	Cerah	69	70	TIDAK	YA
10	Hujan	75	80	TIDAK	YA
11	Cerah	75	70	YA	YA
12	Mendung	72	90	YA	YA

5. Save the data with the name **DataCuaca_Training**, click finish.

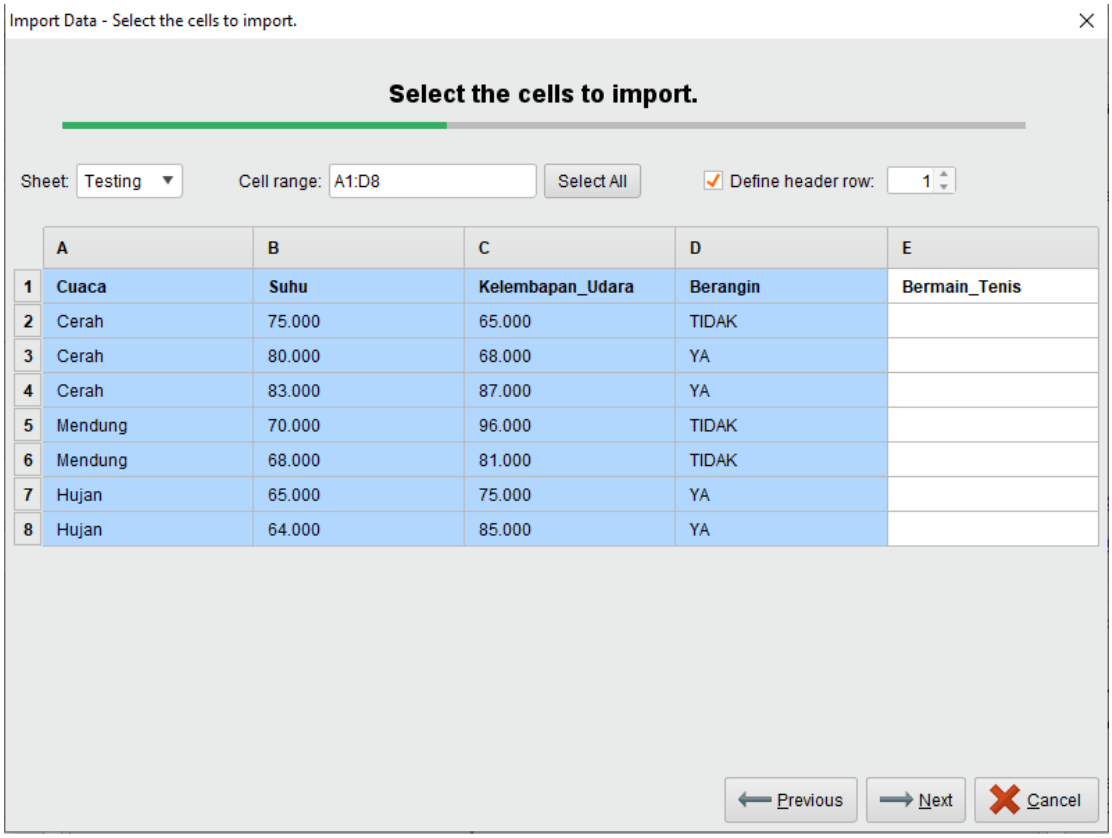


6. The result of **DataCuaca_Training** will appear like the picture below.

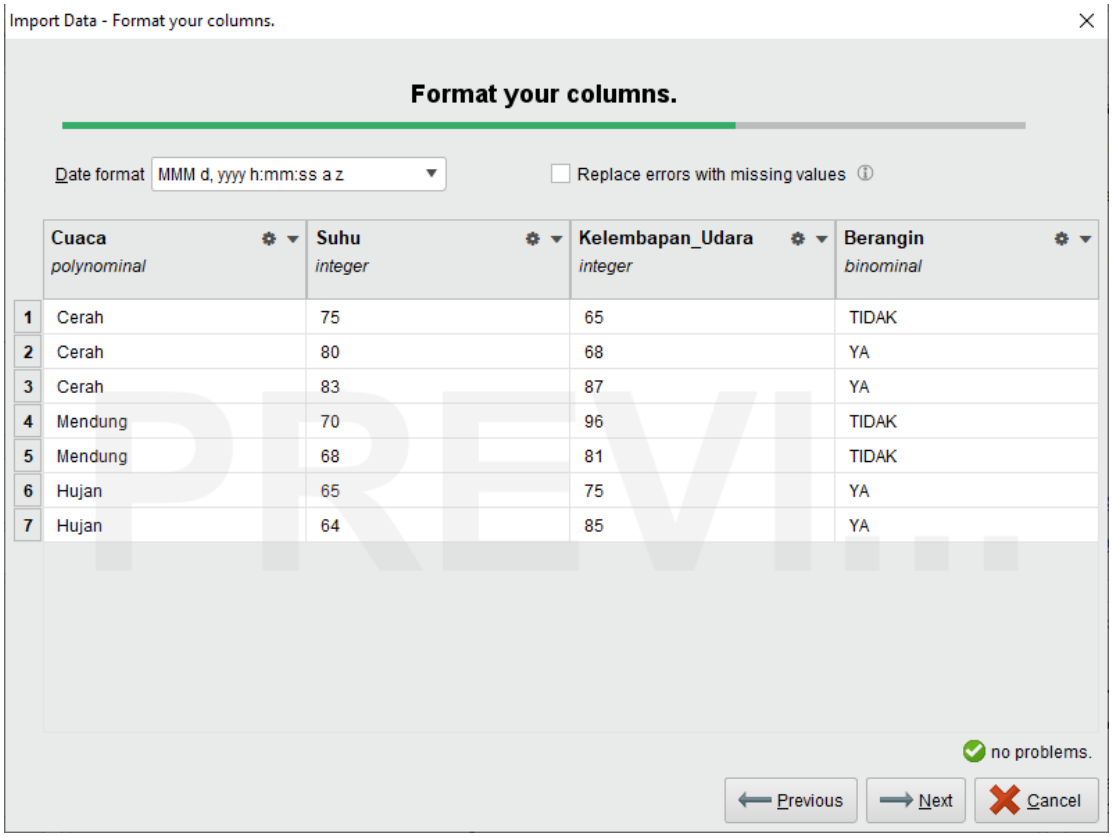


Row No.	Bermain_Te...	Cuaca	Suhu	Kelembapan...	Berangin
1	TIDAK	Cerah	85	85	TIDAK
2	TIDAK	Cerah	80	90	YA
3	YA	Mendung	83	86	TIDAK
4	YA	Hujan	70	96	TIDAK
5	YA	Hujan	68	80	TIDAK
6	TIDAK	Hujan	65	70	YA
7	YA	Mendung	64	65	YA
8	TIDAK	Cerah	72	95	TIDAK
9	YA	Cerah	69	70	TIDAK
10	YA	Hujan	75	80	TIDAK
11	YA	Cerah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

7. Return to the design tab, select add data again to enter the testing data, the process is the same as the previous step (but there is no change in role type for data testing).

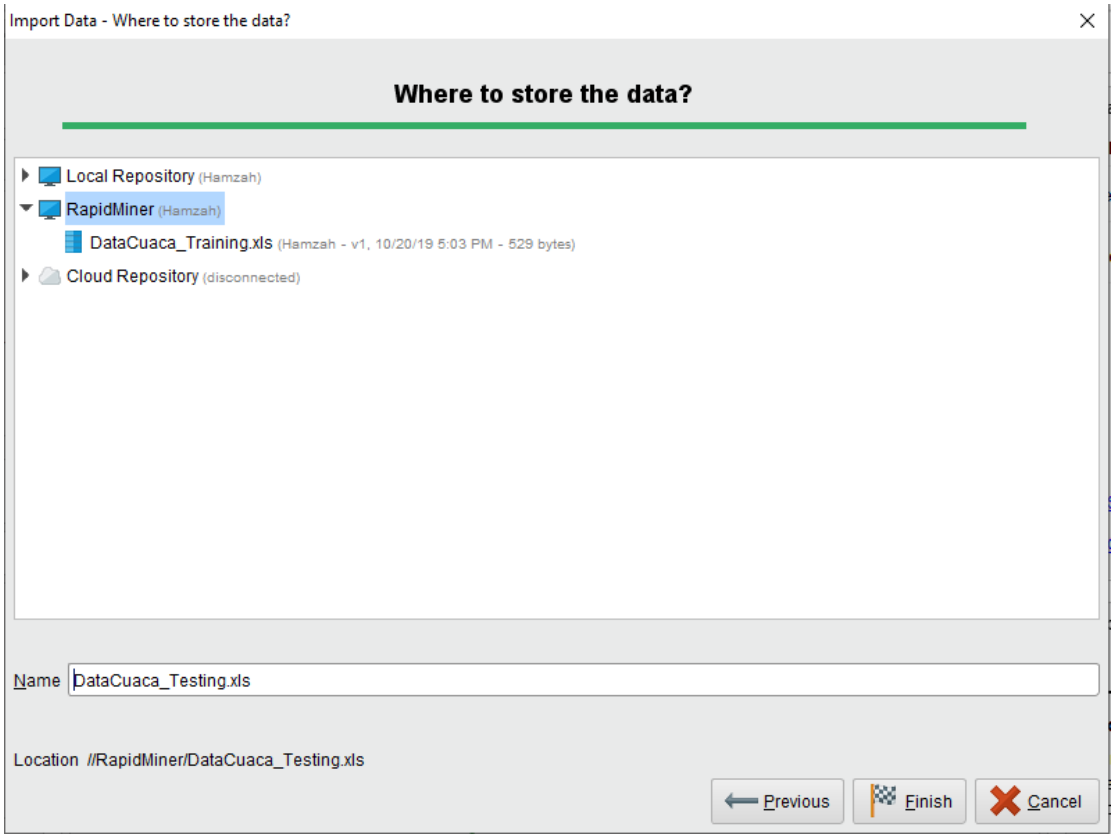


	A	B	C	D	E
1	Cuaca	Suhu	Kelembapan_Udara	Berangin	Bermain_Tenis
2	Cerah	75.000	65.000	TIDAK	
3	Cerah	80.000	68.000	YA	
4	Cerah	83.000	87.000	YA	
5	Mendung	70.000	96.000	TIDAK	
6	Mendung	68.000	81.000	TIDAK	
7	Hujan	65.000	75.000	YA	
8	Hujan	64.000	85.000	YA	

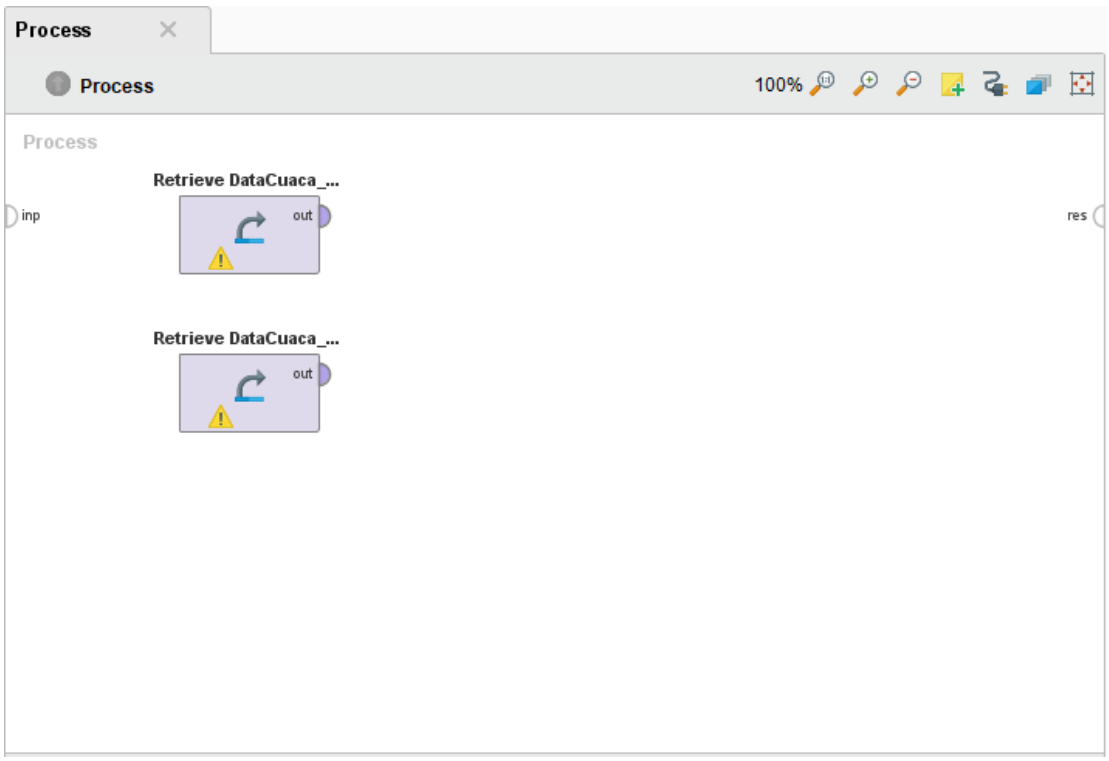


	Cuaca *polynominal*	Suhu *integer*	Kelembapan_Udara *integer*	Berangin *binominal*
1	Cerah	75	65	TIDAK
2	Cerah	80	68	YA
3	Cerah	83	87	YA
4	Mendung	70	96	TIDAK
5	Mendung	68	81	TIDAK
6	Hujan	65	75	YA
7	Hujan	64	85	YA

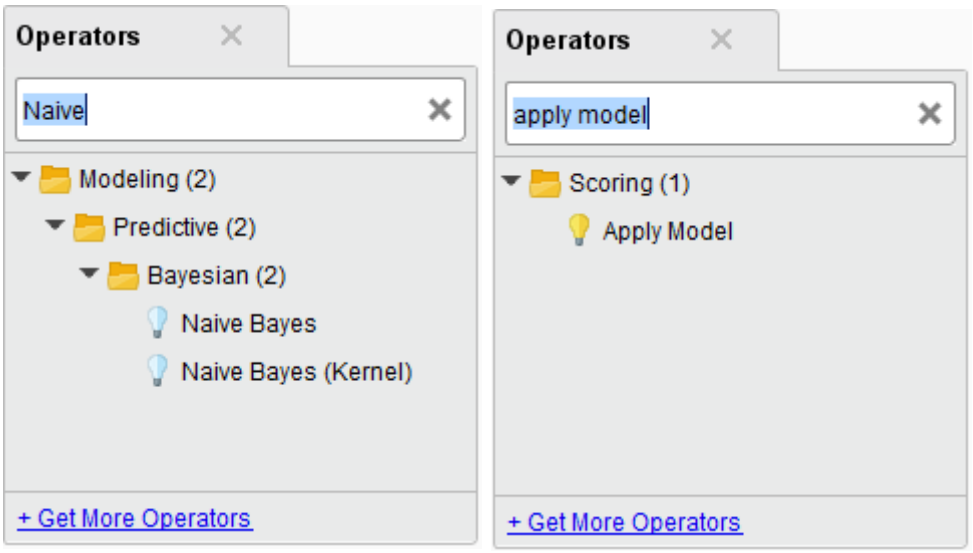
8. Save with the name **DataCuaca_Testing**.

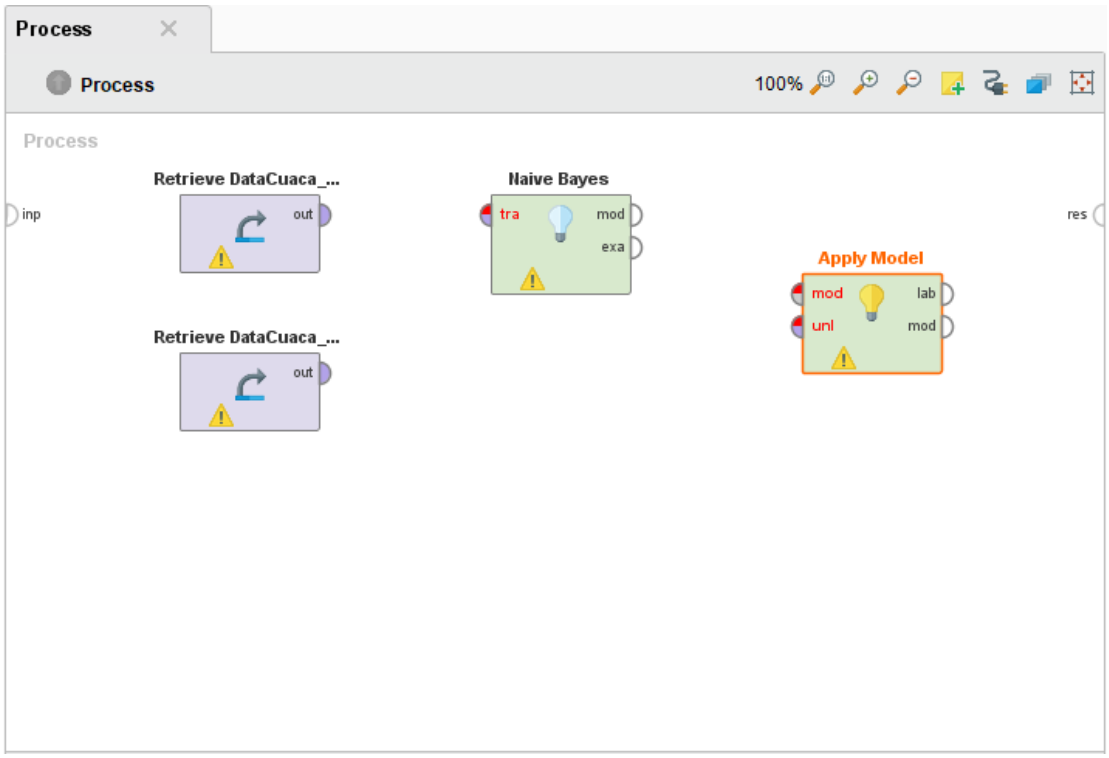


9. The next step is to create a Naïve Bayes design. Drag **DataCuaca_Training** and **DataCuaca_Testing**, into the process view window.

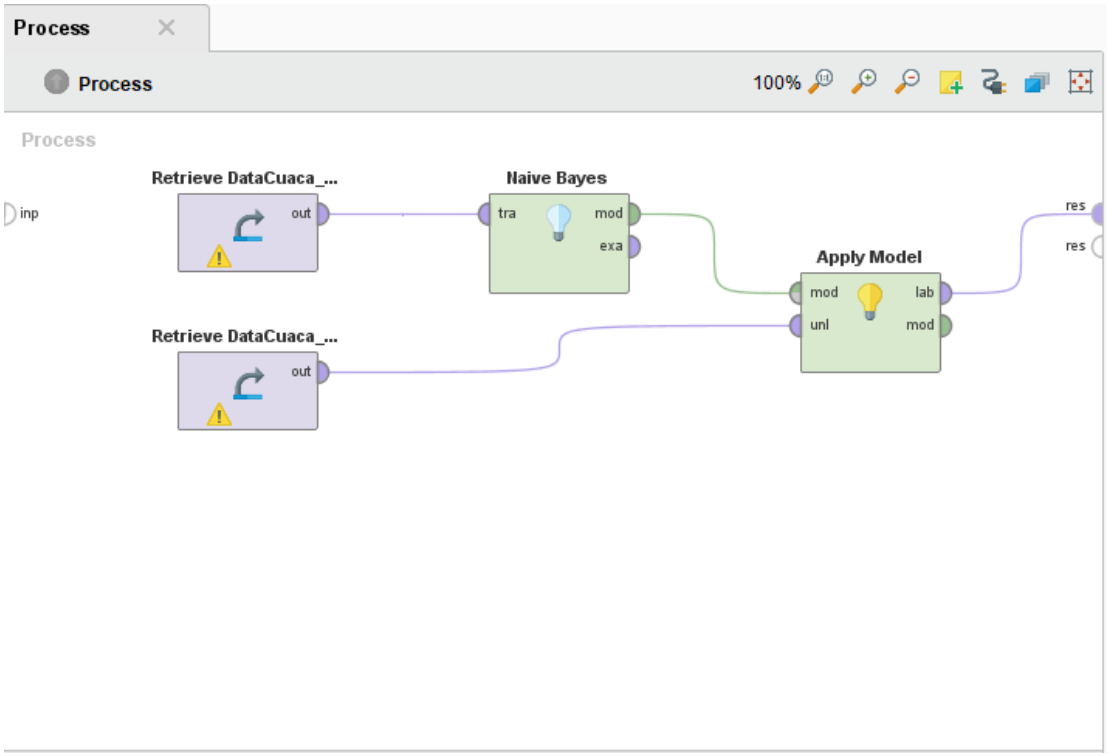


10. In the operator column, enter **Naïve Bayes** and **Apply Model** (type both in the search column and it will appear by itself). Drag both of them into the process view window too.





11. Connect all the designs in the process view window, then click run.



12. The results of the **Naïve Bayes** classification will appears as shown below.

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 more]
Suhu	Integer	0	Min 64	Max 83	Average 72.143
Kelembapan_Udara	Integer	0	Min 65	Max 96	Average 79.571
Berangin	Binominal	0	Least TIDAK (3)	Most YA (4)	Values YA (4), TIDAK (3)

Task

Question.

1. Based on the following table, create files in Excel (.xls) and ARFF (.arff) format! This data will be used as testing data.

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

2. Use the ARFF file worked on Task number 1 in module 7 as training data. Predict the testing data (ARFF) above using WEKA.
3. Use the excel file that was worked out in Task number 1 in module 6 as training data. Predict the testing data (Excel) above using RapidMiner.
4. From the experimental results of Task number 3 above, what is the average confidence value for the Lama_studi attribute with the Right value? What is the average confidence value for the Lama_studi attribute with the Late value?
5. From the results of Experiment Task number 3 above, how many people will pass precisely, and how many people will graduate late.
6. Predict the accuracy study of Dewi, if Dewi is a woman who came from the science department during high school, from schools from outside Surakarta, taking credits with an average of 18 credits per semester, and was never an assistant during college.
7. Predict the accuracy study of Jono, if Jono is a man who comes from the Department of Natural Sciences and Social Sciences at the time of high school, from school from Surakarta, taking credits with an average of 17 credits per semester, and had been an assistant during college.

Answer

1. Data testing in excel and arff.

	A	B	C	D	E
1	**Jurusan_SMA**	**Gender**	**Asal_Sekolah**	**Rerata_SKS**	**Asisten**
2	LAIN	WANITA	SURAKARTA	18	TIDAK
3	IPA	PRIA	SURAKARTA	19	YA
4	LAIN	PRIA	SURAKARTA	19	TIDAK
5	IPS	PRIA	LUAR	17	TIDAK
6	LAIN	WANITA	SURAKARTA	17	TIDAK
7	IPA	WANITA	LUAR	18	YA
8	IPA	PRIA	SURAKARTA	18	TIDAK
9	IPA	PRIA	SURAKARTA	19	TIDAK
10	IPS	PRIA	LUAR	18	TIDAK
11	LAIN	WANITA	SURAKARTA	18	TIDAK
12	IPA	WANITA	LUAR	18	TIDAK
13	LAIN	PRIA	SURAKARTA	17	YA
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					

Data_Training | **Data_Testing**

SMA_Testing.arff

C: > Users > Hamzah > Documents > Hamzah Backup > Kuliah >

```
@relation Jurusan_SMA

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

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

2. Data Training of Modul 7 number 1, and the prediction result using WEKA application.

SMA_Training.arff

C: > Users > Hamzah > Documents > Hamzah Backup > Kuliah > TI U

```
@relation Jurusan_SMA

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

@data
IPS, WANITA, SURAKARTA, 18, TIDAK, TERLAMBAT
IPA, PRIA, SURAKARTA, 19, YA, TEPAT
LAIN, PRIA, SURAKARTA, 19, TIDAK, TERLAMBAT
IPA, PRIA, LUAR, 17, TIDAK, TERLAMBAT
IPA, WANITA, SURAKARTA, 17, TIDAK, TEPAT
IPA, WANITA, LUAR, 18, YA, TEPAT
IPA, PRIA, SURAKARTA, 18, TIDAK, TERLAMBAT
IPA, PRIA, SURAKARTA, 19, TIDAK, TEPAT
IPA, PRIA, LUAR, 18 TIDAK, TERLAMBAT
LAIN, WANITA, SURAKARTA, 18, TIDAK, TEPAT
IPA, WANITA, SURAKARTA, 19, TIDAK, TEPAT
IPS, PRIA, SURAKARTA, 20, TIDAK, TEPAT
IPS, PRIA, SURAKARTA, 19, TIDAK, TEPAT
IPA, PRIA, SURAKARTA, 19, TIDAK, TEPAT
IPA, PRIA, LUAR, 22, YA, TEPAT
LAIN, PRIA, SURAKARTA, 16, TIDAK, TERLAMBAT
IPS, PRIA, LUAR, 20, TIDAK, TEPAT
LAIN, PRIA, LUAR, 23, YA, TEPAT
IPA, PRIA, SURAKARTA, 21, YA, TEPAT
IPS, PRIA, SURAKARTA, 19, TIDAK, TERLAMBAT
```

ARFF-Viewer - C:\Users\Hamzah\Documents\Hamzah Backup\Kuliah\TI UMS\Data Mining & ...

File Edit View

HasilPrediksi.arff

Relation: Jurusan_SMA_predicted

No.	1: Jurusan_SMA	2: Gender	3: Asal_Sekolah	4: Rerata_SKS	5: Asisten	6: prediction margin	7: predicted Lama_Studi
	Nominal	Nominal	Nominal	Numeric	Nominal	Numeric	Nominal
1	LAIN	WANITA	SURAKARTA	18.0	TIDAK	0.375862	TERLAMBAT
2	IPA	PRIA	SURAKARTA	19.0	YA	-0.787744	TEPAT
3	LAIN	PRIA	SURAKARTA	19.0	TIDAK	0.175169	TERLAMBAT
4	IPS	PRIA	LUAR	17.0	TIDAK	0.635052	TERLAMBAT
5	LAIN	WANITA	SURAKARTA	17.0	TIDAK	0.546846	TERLAMBAT
6	IPA	WANITA	LUAR	18.0	YA	-0.689615	TEPAT
7	IPA	PRIA	SURAKARTA	18.0	TIDAK	0.26323	TERLAMBAT
8	IPA	PRIA	SURAKARTA	19.0	TIDAK	-0.224577	TEPAT
9	IPS	PRIA	LUAR	18.0	TIDAK	0.486297	TERLAMBAT
10	LAIN	WANITA	SURAKARTA	18.0	TIDAK	0.375862	TERLAMBAT

3. Data training of Modul 6 number 1 and the prediction result using RapidMiner application.

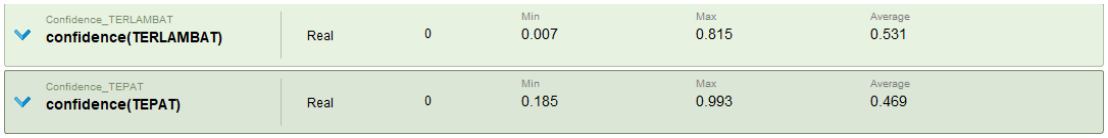
H9

	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	IPA	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						
23						

Data_Training | Data_Testing

Prediction	prediction(Lama_Studi)	Binominal	0	Least: TEPAT (3)	Most: TERLAMBAT (7)	Values: TERLAMBAT (7), TEPAT (3)
Confidence_TERLAMBAT	confidence(TERLAMBAT)	Real	0	Min: 0.007	Max: 0.815	Average: 0.531
Confidence_TEPAT	confidence(TEPAT)	Real	0	Min: 0.185	Max: 0.993	Average: 0.469
	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_SKS	Integer	0	Min: 17	Max: 19	Average: 18.100

4. Average confidence value of the Lama_studi attribute with the Tepat and Late value.



Confidence_TERLAMBAT **confidence(TERLAMBAT)**	Real	0	Min 0.007	Max 0.815	Average 0.531
Confidence_TEPAT **confidence(TEPAT)**	Real	0	Min 0.185	Max 0.993	Average 0.469

5. The number of people that will pass precisely and graduate late.

Prediction **prediction(Lama_Studi)**	Binominal	0	Least TEPAT (3)	Most TERLAMBAT (7)	Values TERLAMBAT (7), TEPAT (3)

6. The accuracy of study for Dewi.

11	TEPAT	0.387	0.613	IPA	WANITA	LUAR	18	TIDAK

7. The accuracy of study for Jono.

12	TEPAT	0.076	0.924	LAIN	PRIA	SURAKARTA	17	YA