

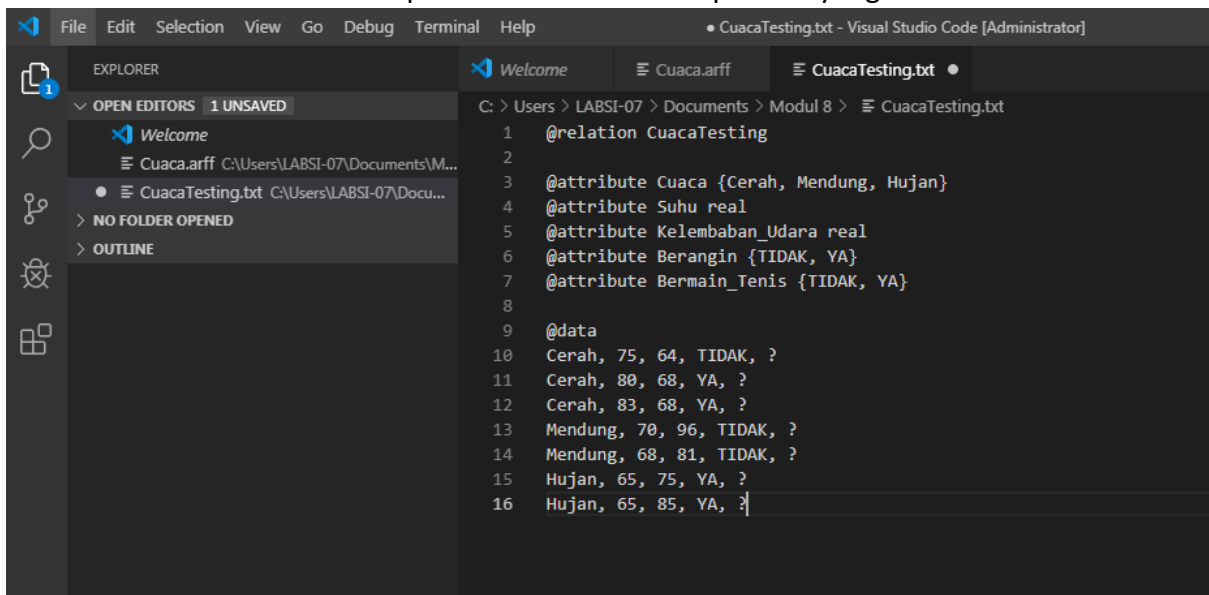
Kelas : C

KLASIFIKASI : NAÏVE BAYES

➤ Implementasi Naïve Bayes dengan Weka

[illegible]

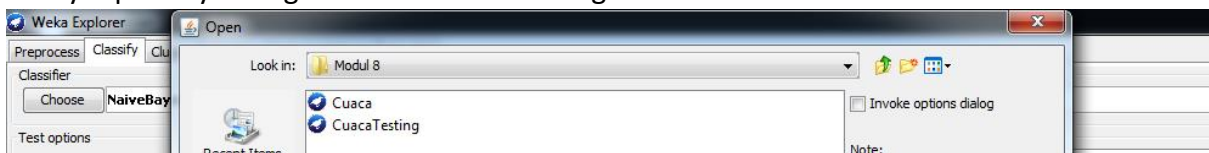
2. Data Testing dengan format ARFF dari tabel 8.1 sebagai data uji yang akan di prediksi memiliki variable-variable independen dan variabel dependen yang sama



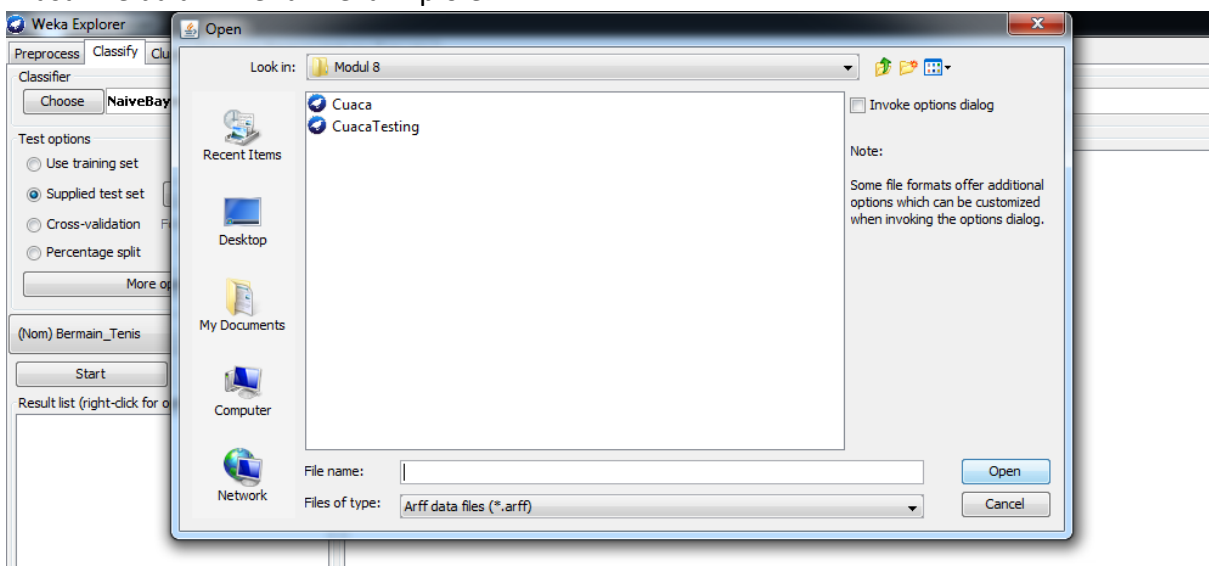
The screenshot shows the Visual Studio Code editor with the file 'CuacaTesting.txt' open. The file content is as follows:

```
@relation CuacaTesting
@attribute Cuaca {Cerah, Mendung, Hujan}
@attribute Suhu real
@attribute Kelembaban_Udara real
@attribute Berangin {TIDAK, YA}
@attribute Bermain_Tenis {TIDAK, YA}
@data
Cerah, 75, 64, TIDAK, ?
Cerah, 80, 68, YA, ?
Cerah, 83, 68, YA, ?
Mendung, 70, 96, TIDAK, ?
Mendung, 68, 81, TIDAK, ?
Hujan, 65, 75, YA, ?
Hujan, 65, 85, YA, ?
```

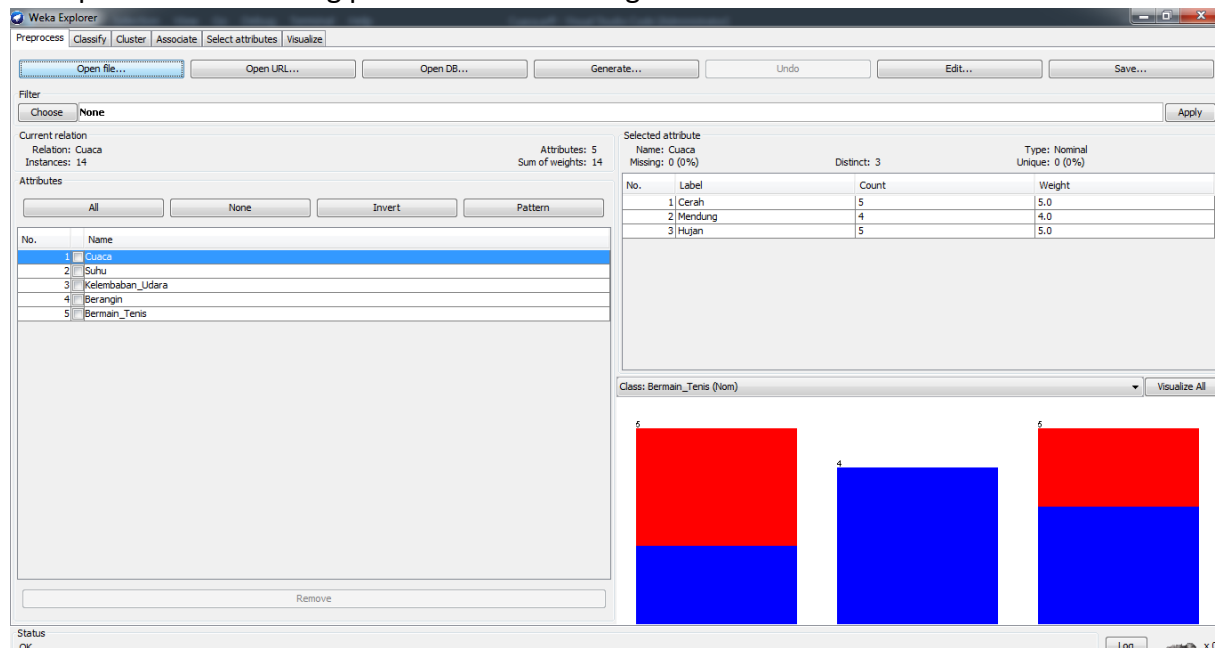
3. Menyimpannya dengan format CuacaTesting.arff



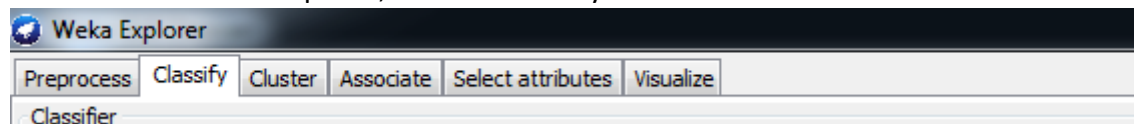
4. Masuk ke dalam menu Weka Explorer



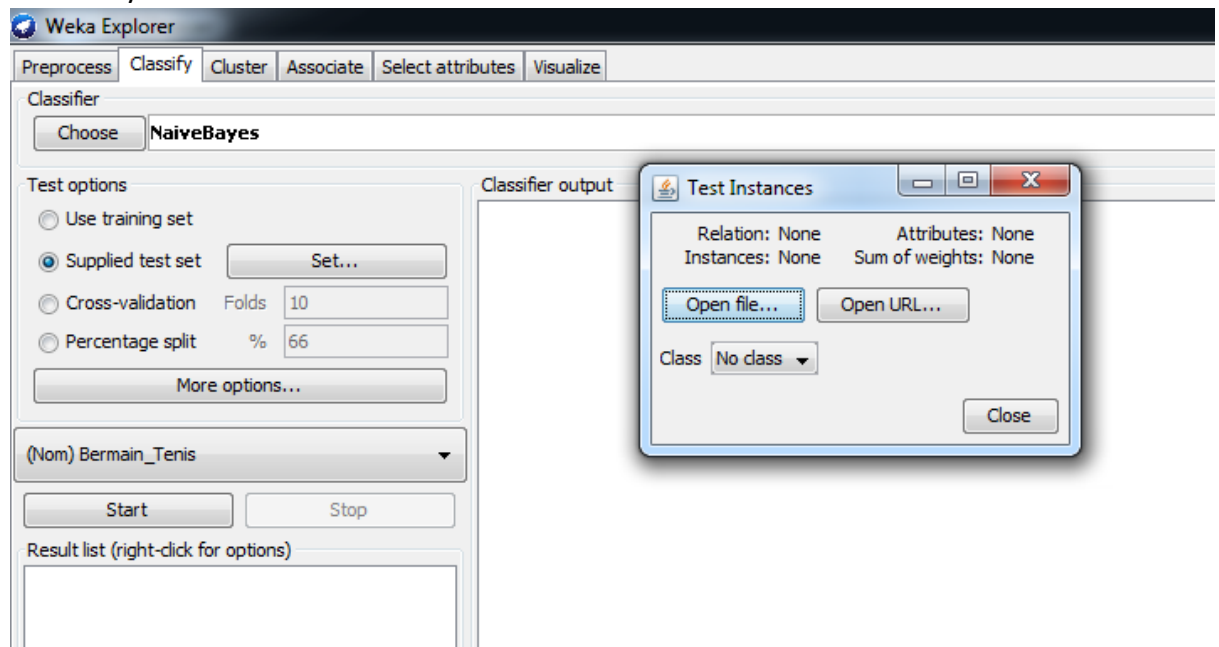
- Buka kembali file Cuaca.arff, file ini akan di gunakan sebagai data pelatihan untuk memprediksi data testing pada file CuacaTesting.arff



- Pada window Weka Explorer, Pilih tab Classify



- Sehingga muncul jendela Weka Explorer pada tab Classify, lalu pilih metode/algorithm Naive Bayes

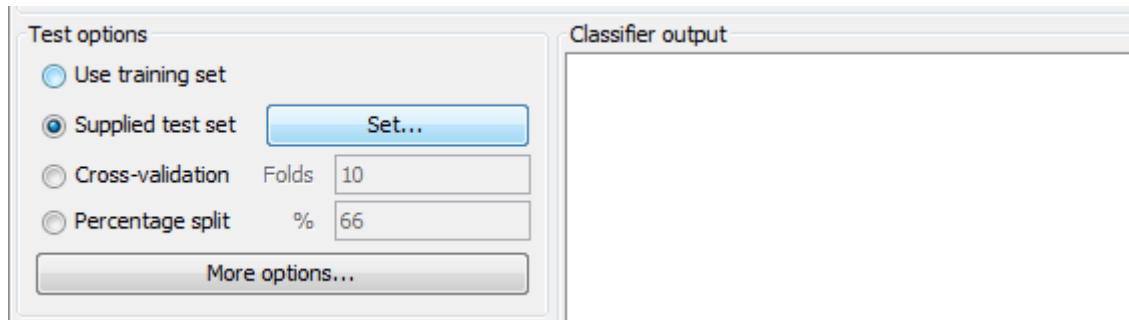


8. File CuacaTesting.arff di tentukan sebagai data testing pada kegiatan ini.

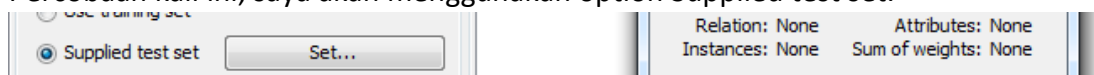


9. Pada menu test Option terdapat 4 pilihan pengujian, yaitu:

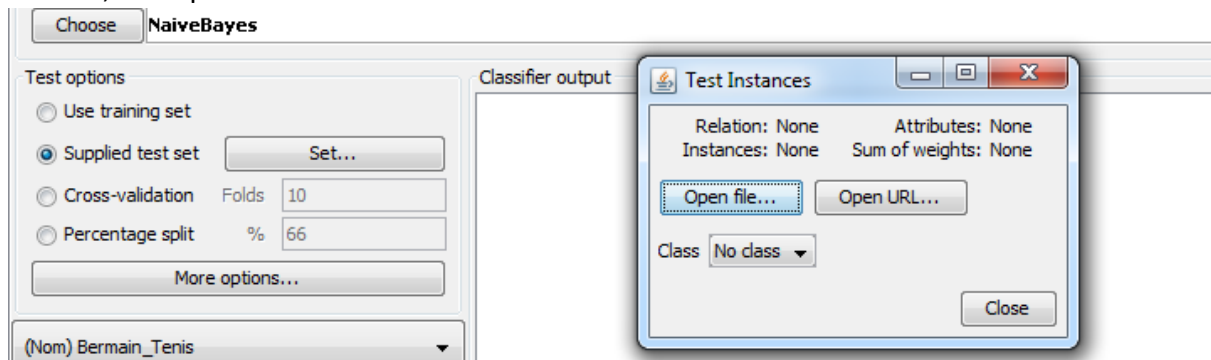
- Use training set
- Supplied test set
- Cross-validation
- Percentage split



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



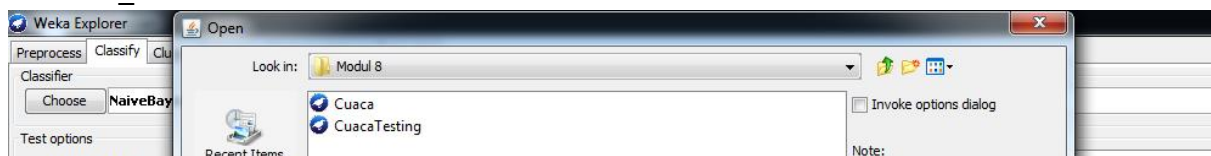
11. Klik Set, lalu open file



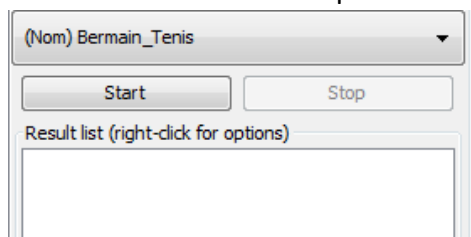
12. Pilih file CuacaTesting.arff sebagai data untuk di uji



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 prediksi kita melakukan proses Classifier Errors, dengan memilih kotak result list, dengan memilih option Visualize classifier errors

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☒ Supplied test set Set...

☐ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) Bermain_Tenis

Start Stop

Result list (right-click for options)

22:55:16 - misc.InputMappedClassifier

Classifier output

```
-----
(nominal) Cuaca      --> 1 (nominal) Cuaca
(numeric) Suhu       --> 2 (numeric) Suhu
(numeric) Kelembaban_Udara --> 3 (numeric) Kelembaban_Udara
(nominal) Berangin   --> 4 (nominal) Berangin
(nominal) Bermain_Tenis --> 5 (nominal) Bermain_Tenis

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

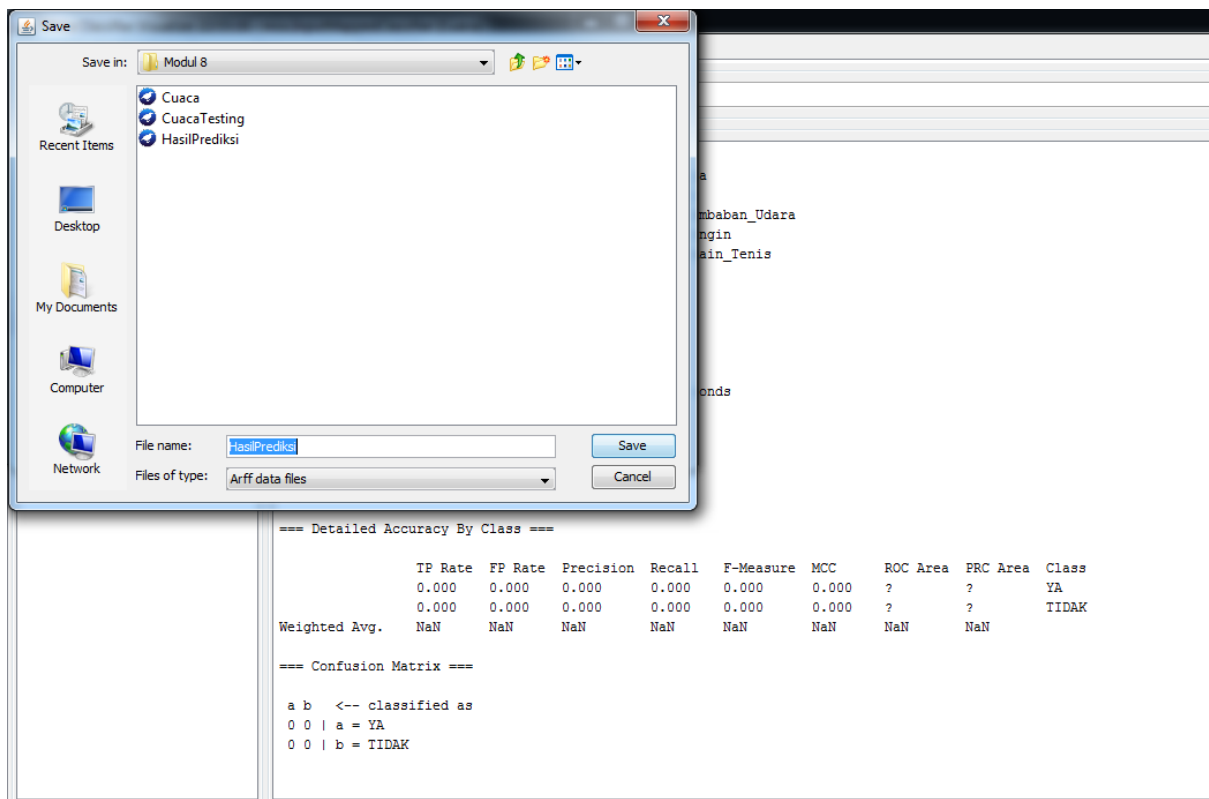
Total Number of Instances      0
Ignored Class Unknown Instances      7

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      0.000    0.000    0.000    0.000    0.000    0.000    ?        ?        YA
      0.000    0.000    0.000    0.000    0.000    0.000    ?        ?        TIDAK
Weighted Avg.   NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
=== Confusion Matrix ===

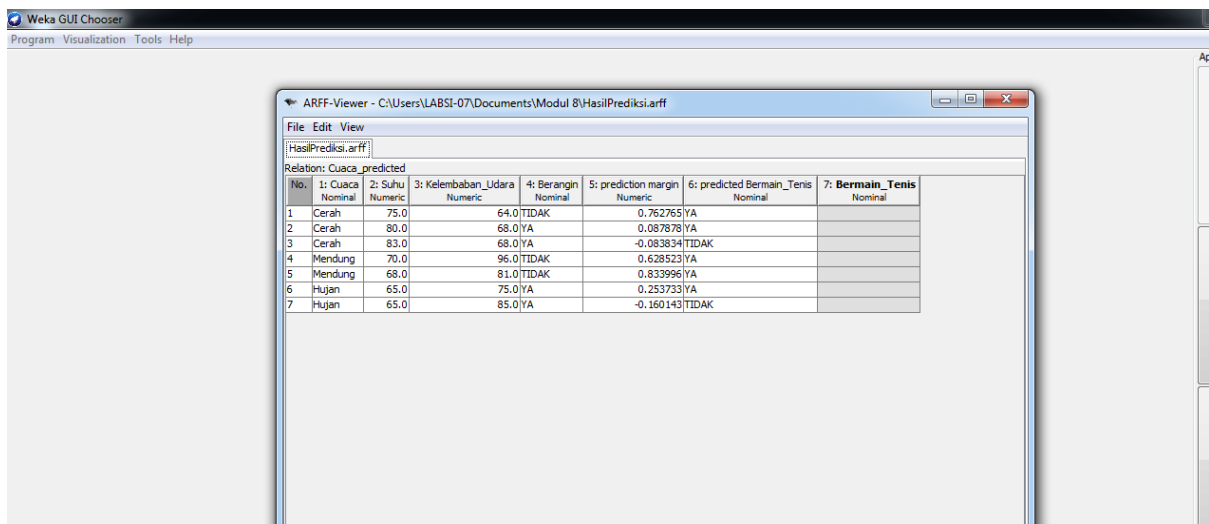
a b  <-- classified as
0 0 | a = YA
0 0 | b = TIDAK
```

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.

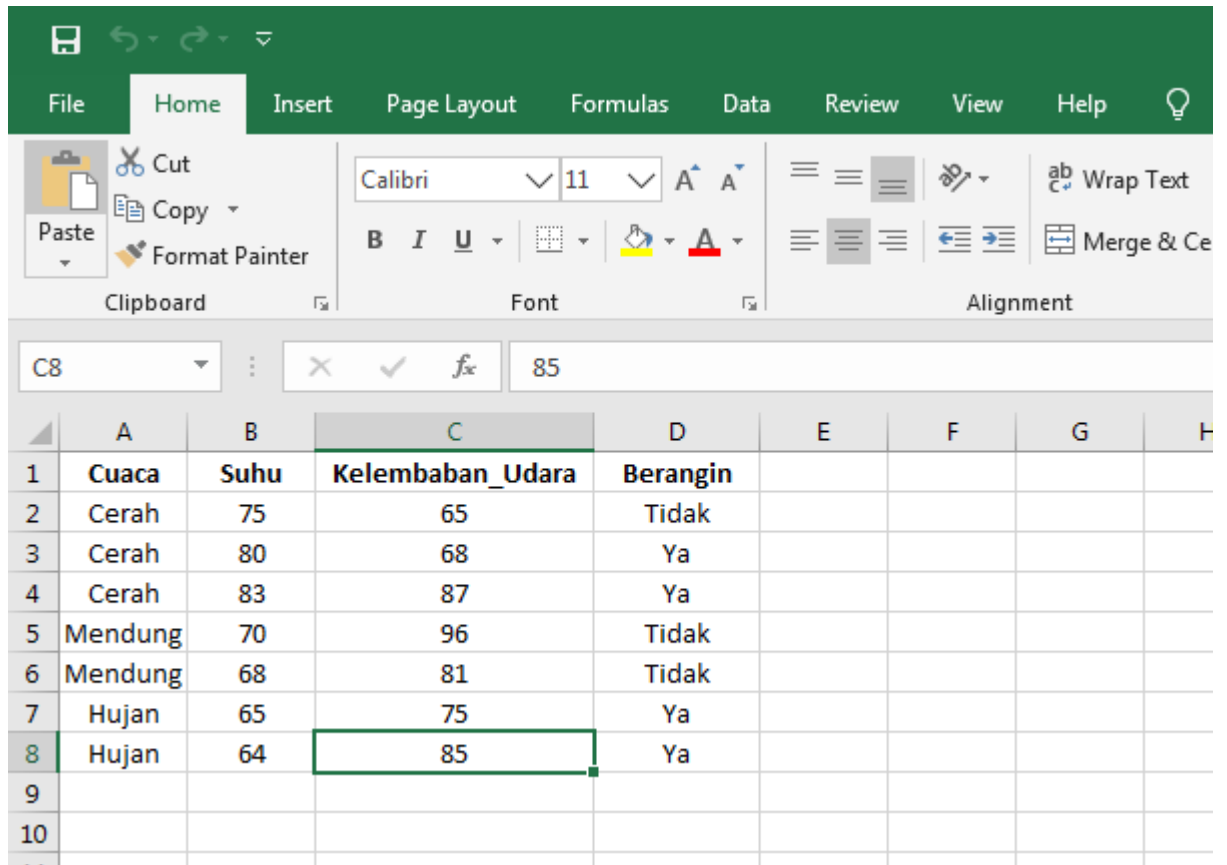


- **Implementasi Naïve Bayes dengan Rapid Miner**

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

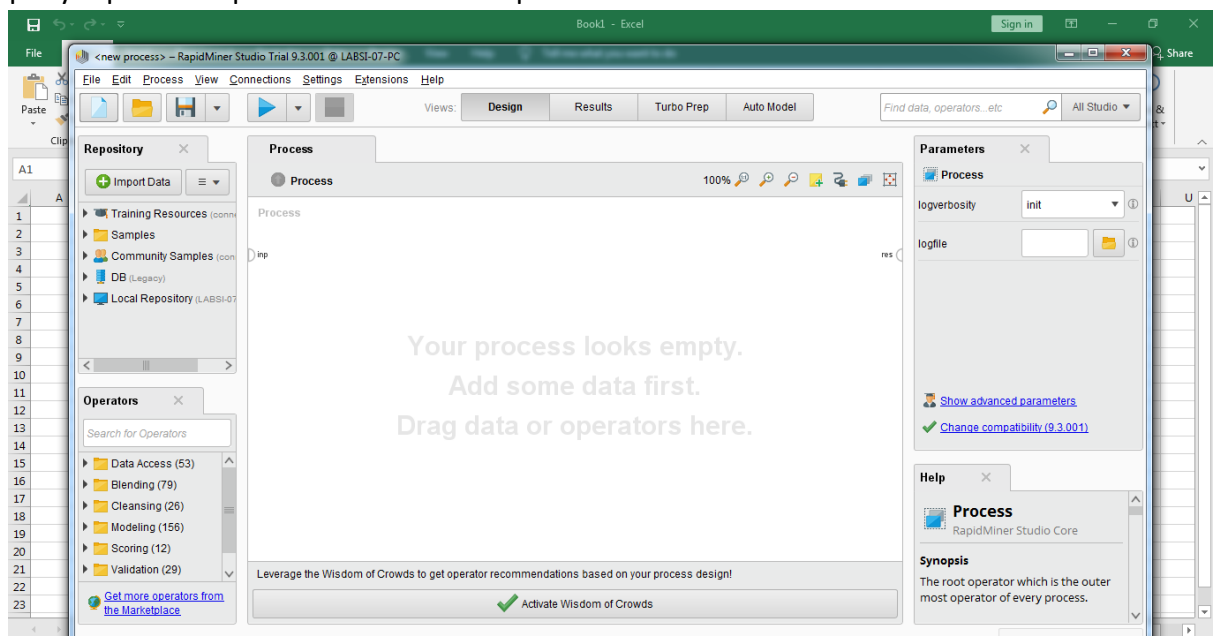
[illegible]

- Tabel Data Testing pada Sheet2



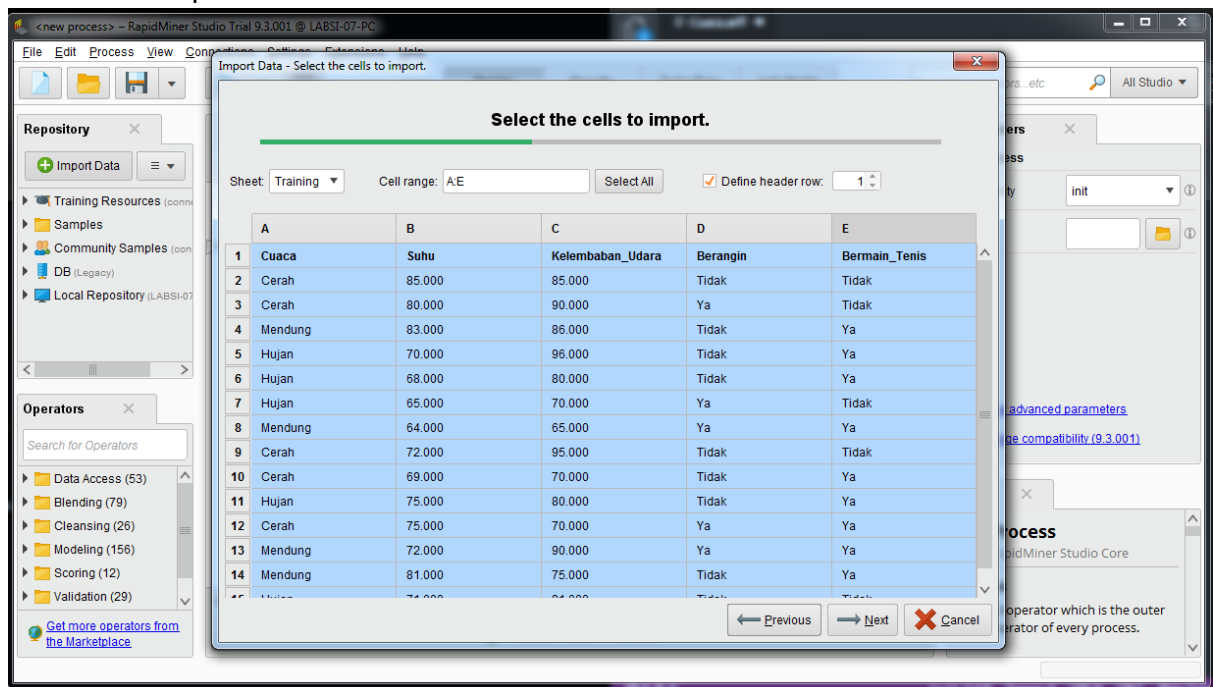
	A	B	C	D	E	F	G	H
1	Cuaca	Suhu	Kelembaban_Udara	Berangin				
2	Cerah	75	65	Tidak				
3	Cerah	80	68	Ya				
4	Cerah	83	87	Ya				
5	Mendung	70	96	Tidak				
6	Mendung	68	81	Tidak				
7	Hujan	65	75	Ya				
8	Hujan	64	85	Ya				
9								
10								

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

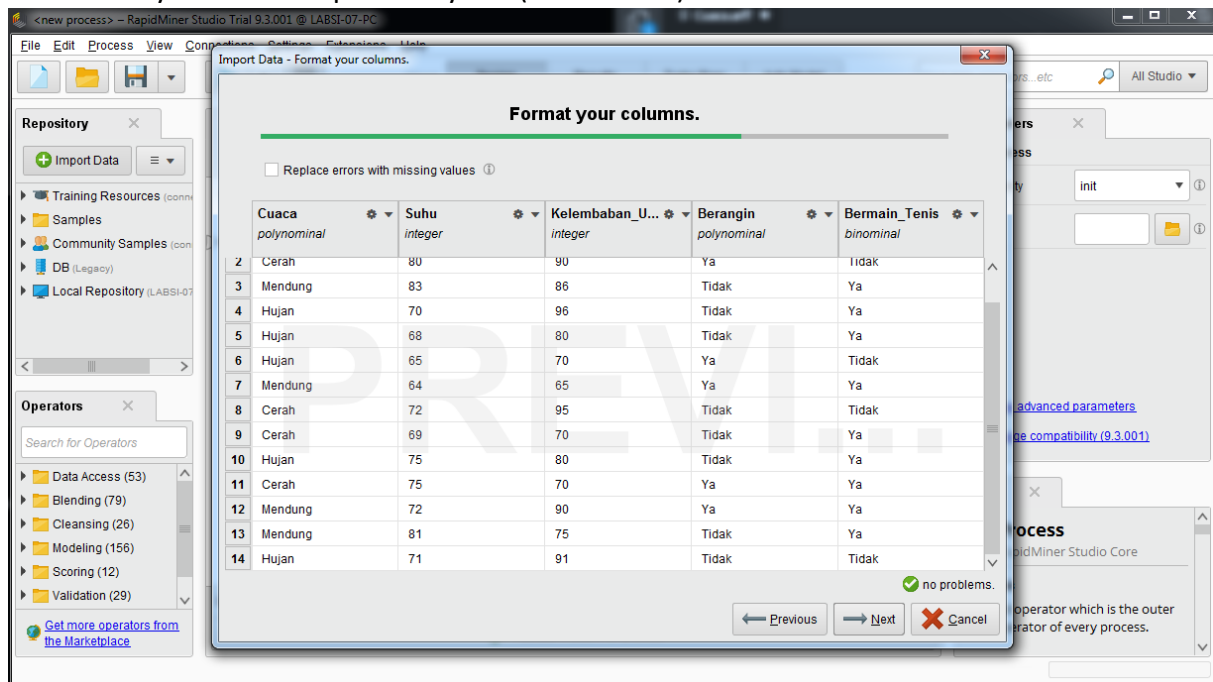


5. Import Data, pilih file Tabel_Cuaca.xls

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



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



- Import Data - Format your columns.

Format your columns.

☐ Replace errors with missing values ⓘ

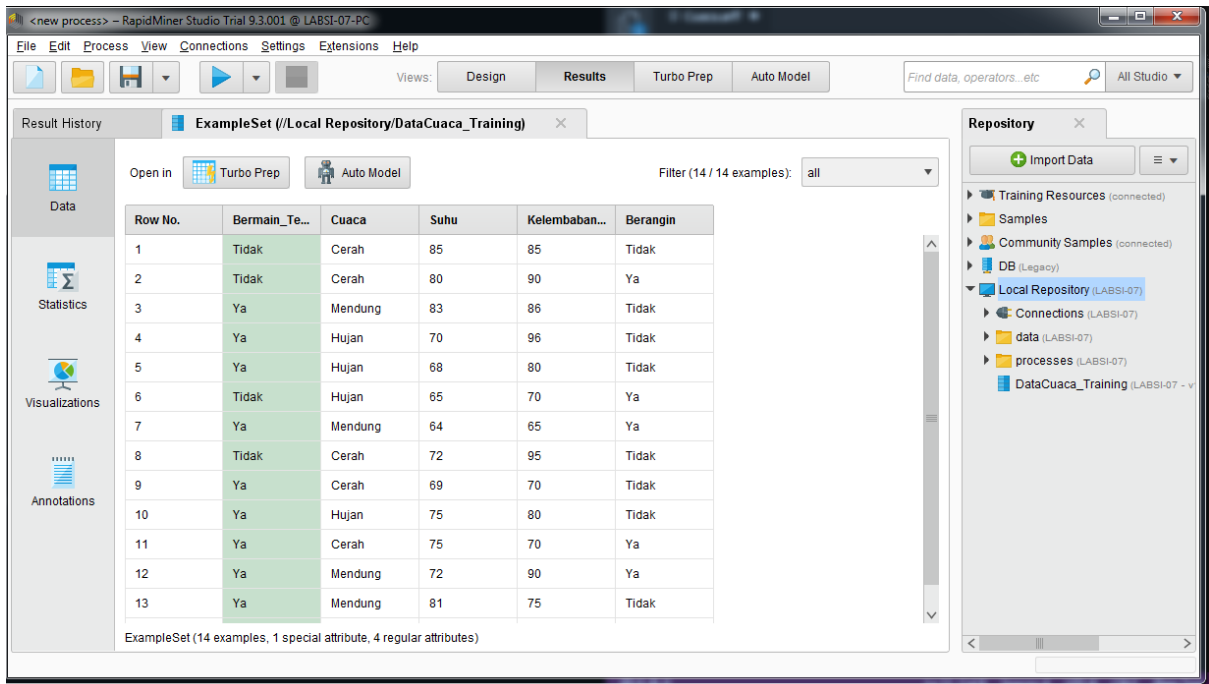
	Cuaca <i>polynomial</i>	Suhu <i>integer</i>	Kelembaban_U... <i>integer</i>	Berangin <i>polynomial</i>	Bermain_Tenis <i>binomial label</i>
2	Cerah	80	90	Ya	Tidak
3	Mendung	83	86	Tidak	Ya
4	Hujan	70	96	Tidak	Ya
5	Hujan	68	80	Tidak	Ya
6	Hujan	65	70	Ya	Tidak
7	Mendung	64	65	Ya	Ya
8	Cerah	72	95	Tidak	Tidak
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

no problems.

Previous Next Cancel

-
- The screenshot shows the 'Import Data' dialog in RapidMiner. The dialog has a title bar 'Import Data - Where to store the data?'. The main content area is titled 'Where to store the data?' and displays a tree view of the 'Local Repository (LABSI-07)'. The tree shows three sub-items: 'Connections (LABSI-07)', 'data (LABSI-07)', and 'processes (LABSI-07)'. Below the tree, the 'Name' field is filled with 'DataCuaca_Training' and the 'Location' is '/Local Repository/DataCuaca_Training'. At the bottom, there are three buttons: 'Previous', 'Finish', and 'Cancel'.

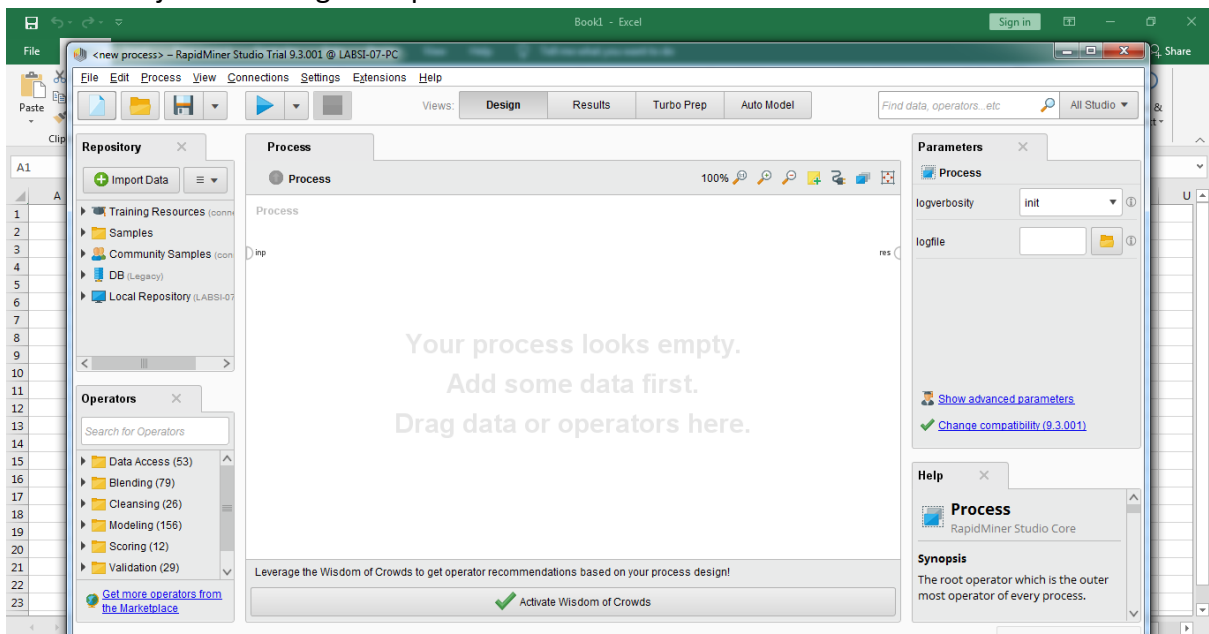
10. Hasil import file Tabel_Cuaca.xls pada Sheet1 akan menampilkan Exmempel Set;



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 Prespective



Repository

- Training Resources (connected)
- Samples
- Community Samples (connected)
- DB (Legacy)
- Local Repository (LABSI-07)
 - Connections (LABSI-07)
 - data (LABSI-07)
 - processes (LABSI-07)
 - DataCuaca_Training (LABSI-07 - v)

Operators

- Data Access (53)
- Blending (79)
- Cleansing (26)
- Modeling (156)
- Scoring (12)
- Validation (29)

Process

Your process looks empty.
Add some data first.
Drag data or operators here.

Parameters

Process

logverbosity: init

logfile: [File Icon]

Show advanced parameters

Change compatibility (9.3.001)

Help

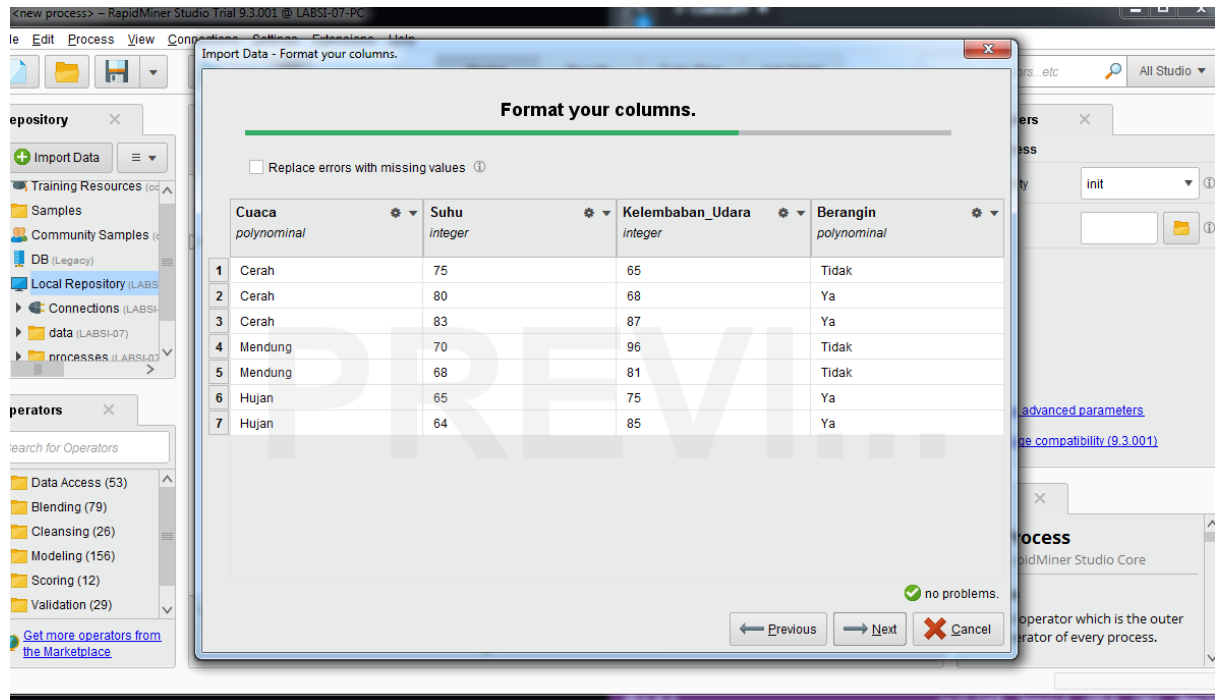
Process

RapidMiner Studio Core

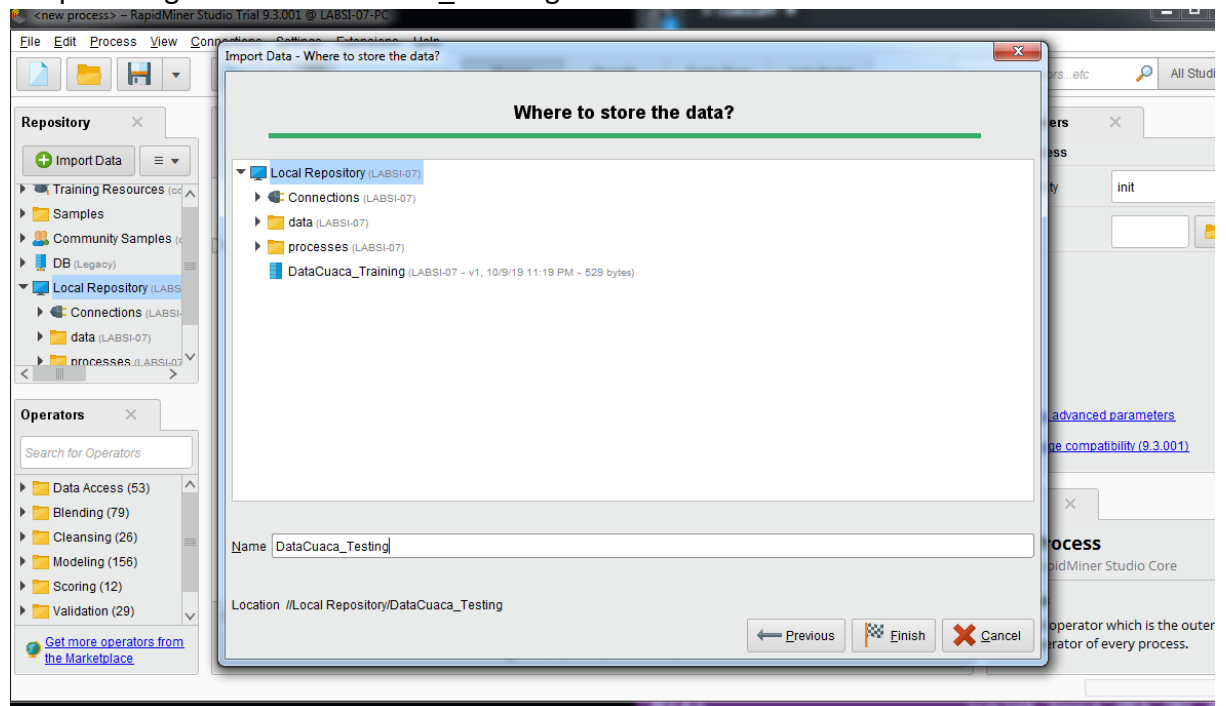
Synopsis

The root operator which is the outer most operator of every process.

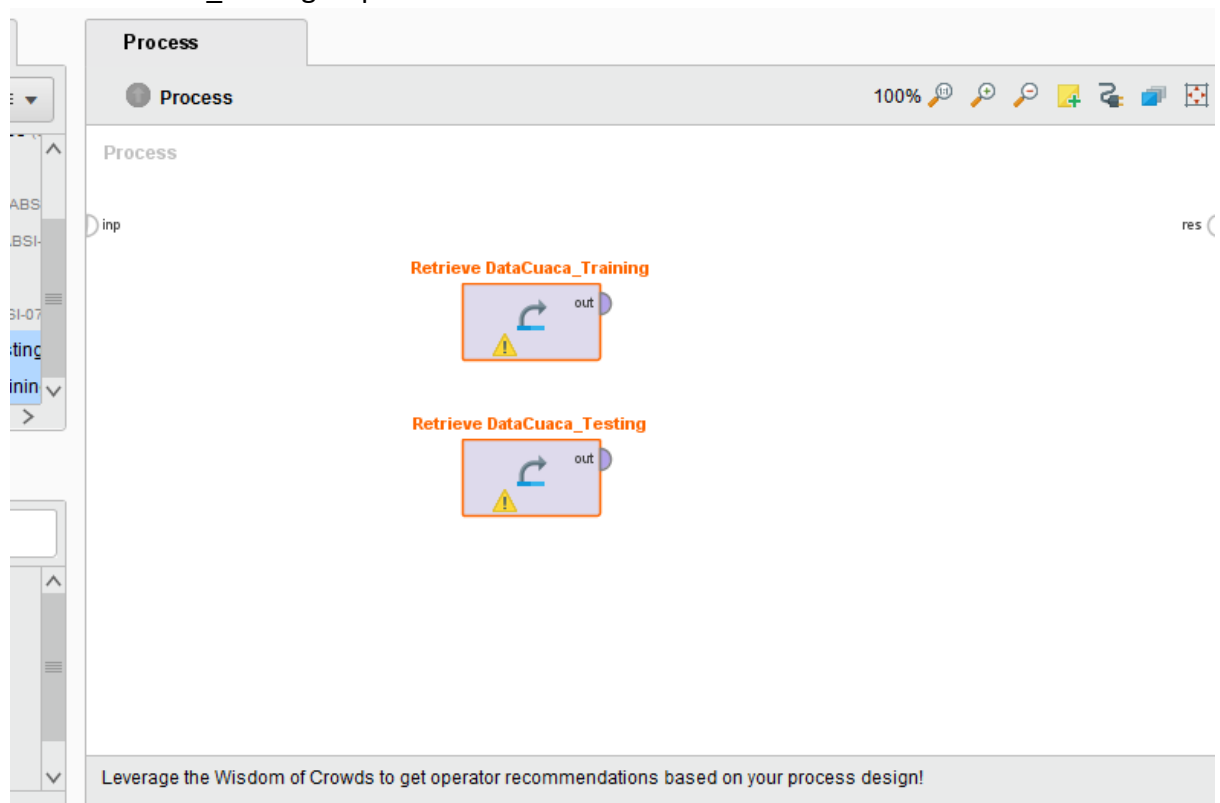
12. Sekarang adalah, kita mengambil data testing yang di import dari file Tabel_Cuaca.xls pada sheet2 (Testing), di langkah ini pastikan data terpilih ada 4 variabel, dan tanpa ada label.



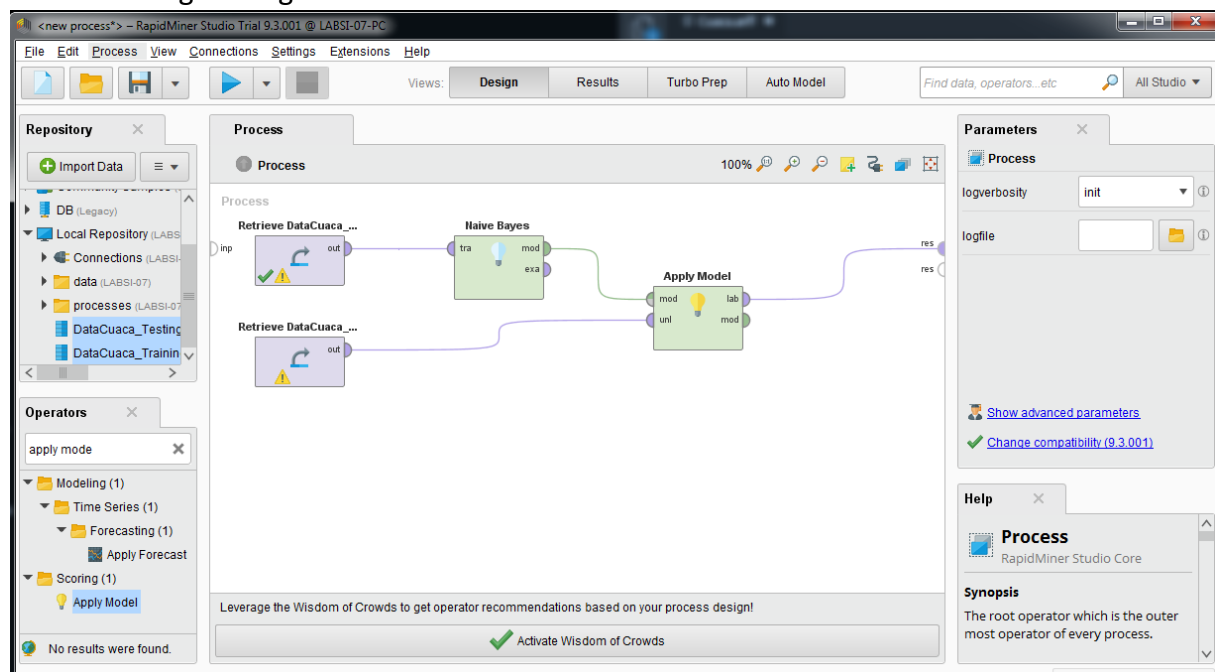
13. Simpan dengan nama DataCuaca_Training.



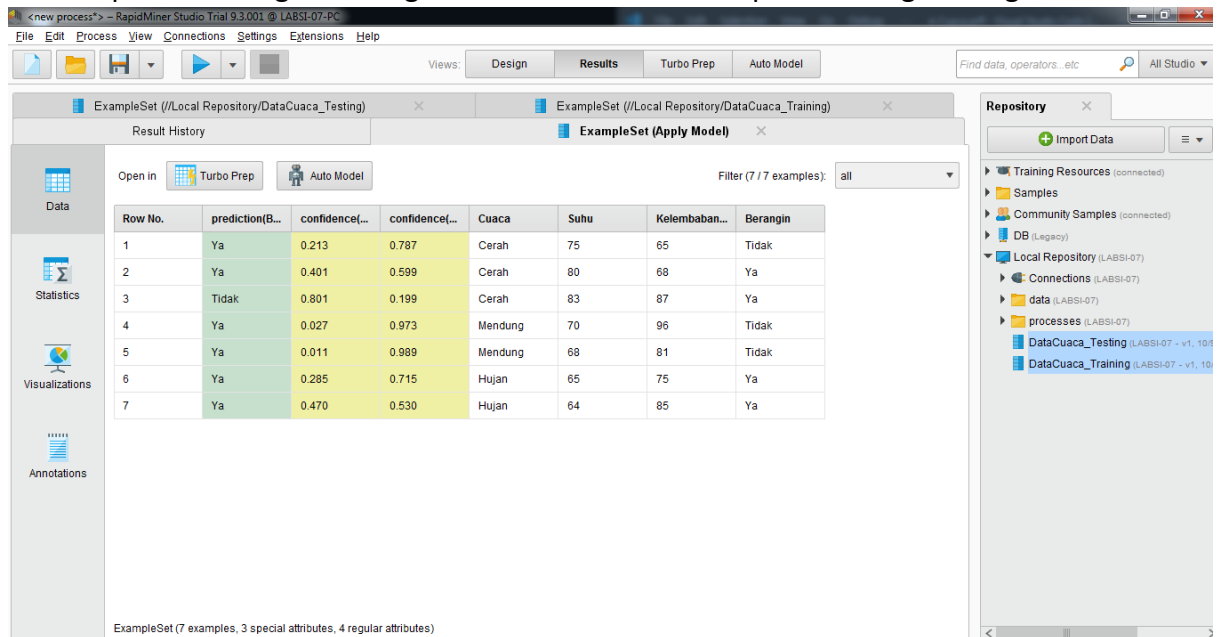
14. Selanjutnya adalah membuat desain Naive Bayes, drag and drop DataCuaca_Training dan DataCuaca_Testing ke process window view.



15. Masukkan juga oprator Naive Bayes dan Apply Model ke dalam procces view, hubungkan koneksi masing masing .

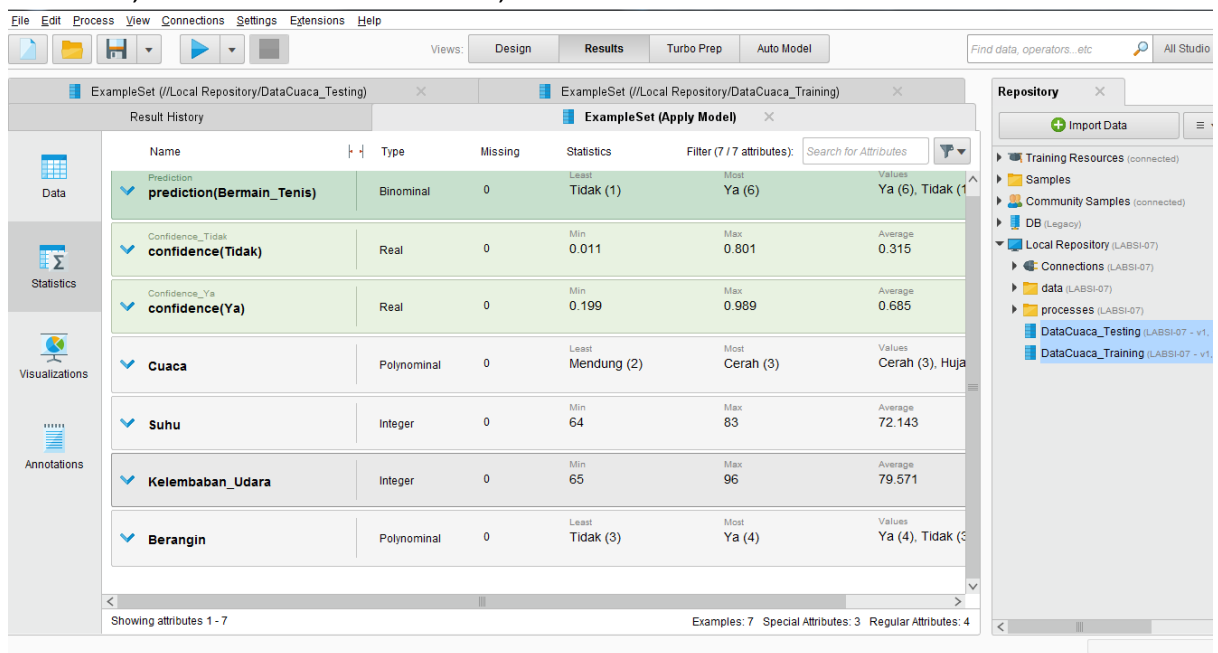


16. Run proses naive Bayes.
17. Hasil proses kalsifikasi naive bayes, pada tab Data, dapat di lihata hasil prediksi terhadap data testing serta tingkat confidence nilai kelas pada masing-masing data



Row No.	prediction(B...	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

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



Name	Type	Missing	Statistics	Filter (7 / 7 attributes)
prediction(Bermain_Tenis)	Binomial	0	Least Tidak (1)	Most Ya (6)
confidence(Tidak)	Real	0	Min 0.011	Max 0.801
confidence(Ya)	Real	0	Min 0.199	Max 0.989
Cuaca	Polynomial	0	Least Mendung (2)	Most Cerah (3)
Suhu	Integer	0	Min 64	Max 83
Kelembaban_Udara	Integer	0	Min 65	Max 96
Berangin	Polynomial	0	Least Tidak (3)	Most Ya (4)