Hesti Putri Utami

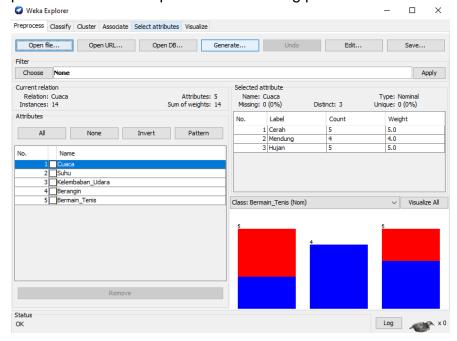
L200170009

Kelas A

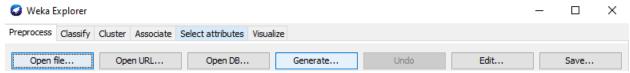
## **MODUL 8. KLASIFIKASI: NAÏVE BAYES**

Langkah-langkah praktikum

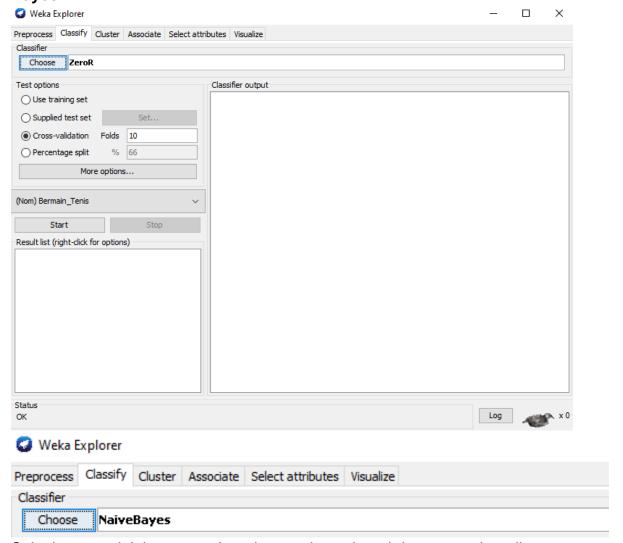
- Implementasi Naïve Bayes dengan Weka Langkah-langkah menggunakan algoritma naïve bayes dengan Weka sebagai berikut :
  - 1. Persiapkan file **Cuaca.arff** dari hasil percobaan kegiatan 7.4.1 pada Modul 7. File ini akan kita gunakan sebagai data training.
  - 2. Buatlah sebuah data testing dengan format **ARFF** dari tabel 8.1 sebagai data uji yang akan di prediksi dengan memiliki variabel-variabel independen dan variabel dependen yang sama. Dengan ketentuan variabel dependen diisi dengan tanda Tanya (?). Asumsi bahwa kita belum mengetahui nilai / kelas dari variabel tersebut. Nilai kelas yang akan kita prediksi dengan menggunakan algoritma Naïve Bayes.
  - 3. Simpan dengan nama CuacaTesting.arff
  - 4. Jika telah selesai membuat Buka aplikasi Weka, masuk dalam menu Weka Explorer.
  - 5. Buka kembali file **Cuaca.arff** dari hasil kegiatan 7.4.1 pada Modul 7 dengan menggunakan Weka Explorer. File ini akan kita gunakan sebagai data pelatihan untuk mempresiksi data testing pada file **CuacaTesting.arff**.



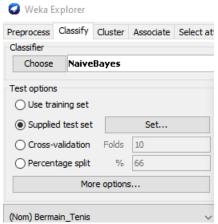
6. Masih pada jendela Weka Explorer, pilih tab Classify.



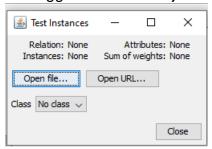
 Sehingga akan muncul jendela Weka Explorer pada tab Classify. Pada kotak Classifier klik tombol Choose untuk memilih metode / algoritma Naïve Bayes.



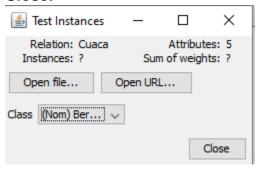
- 8. Selanjutnya adalah menentukan data testing sebagai data yang akan di prediksi variabel dependennya. File **CuacaTesting.arff** ditentukan sebagai data testing pada kegiatan ini.
- 9. Pada menu Test Options terdapat 4 pilihan pengujian.
- 10. Pada percobaan kali ini, kita akan menggunakan pilihan **Supplied test set.** Klik tombol Set untuk menentukan file ARFF sebagai data uji.



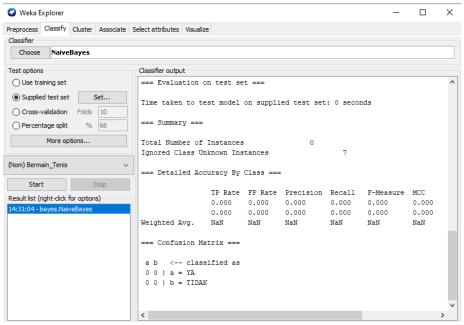
11. Sehingga akan muncul jendela Test Instance. Klik Open file.



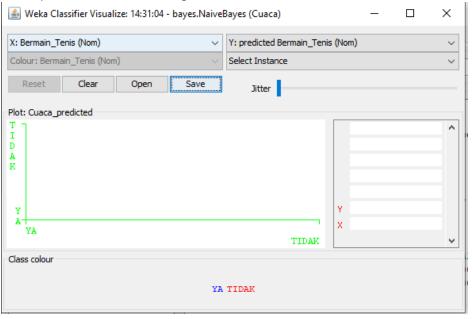
- 12. Pilih file CuacaTesting.arff sebagai data uji. Klik Open.
- 13. File **CuacaTesting.arff** akan di set sebagai data uji pada jendela Test Instance dengan variabel predictor (Class) adalah Bermain\_Tenis. Klik **Close.**



14. Klik Start untuk memulai proses naïve bayes.



- 15. Jika muncul jendela pesan **Classifier Panel**, kita abaikan dengan mengklik Yes. Sehingga algoritma naïve bayes akan diproses.
- 16. Karena pada percobaan ini kita memproses data uji yang belum diketahui nilai kelas dari variabel dependen yang diajukan, maka kita abaikan nilai-nilai yang ditampilkan dalam jendela **Classifier Output**.
- 17. Untuk melihat hasil prediksi terhadap data uji, yang perlu kita lakukan berikutnya adalah dengan melihat nilai **Classifier Errors**. Klik kanan pada hasil proses dalam kotak **result list**. Pilih menu **Visualize classifier errors**.
- Pada jendela Weka Classifier Visualize, abaikan hasil apapun yang ditampilkan. Klik Sav dengan nama file HasilPrediksi.arff.



- 19. Tutup semua jendela termasuk Weka Explorer dan kembali ke **Weka GUI Chooser**. Pilih menu **Tools ArffViewer**.
- 20. Jendela Arff-Viewer akan ditampilkan. Buka menu **File Open.** Tunjukan pada file **HasilPrediksi.arff** yang telah anda simpan pada langkah ke 18. Lihatlah, hasil prediksi telah di ketahui pada kolom **predicted**

### **Bermain\_Tenis Nominal.**

lo.	1: Cuaca Nominal	2: Suhu Numeric	3: Kelembaban_Udara Numeric	4: Berangin Nominal	5: prediction margin Numeric	6: predicted Bermain_Tenis Nominal	7: Bermain_Tenis Nominal	
1	Cerah	75.0	65.0	TIDAK	0.762765	YA		
2	Cerah	80.0	68.0	YA	0.087878	YA		
3	Cerah	83.0	87.0	YA	-0.676866	TIDAK		
4	Mendung	70.0	96.0	TIDAK	0.628523	YA		
5	Mendung	68.0	81.0	TIDAK	0.833996	YA		
5	Hujan	65.0	75.0	YA	0.253733	YA		
7	Hujan	64.0	85.0	YA	-0.160143	TIDAK		

Implementasi Naïve Bayes dengan RapidMiner

Langkah-Langkah:

- 1. Persipakan file **Tabel\_Cuaca.xls** yang terdiri dari 2 sheet.
- 2. Sheet1 digunakan sebagai data training, dan sheet2 digunakan sebagai data uji.
- 3. Masing-masing tabel memiliki attribute yang sama, yaitu :
  - a. Cuaca (X1)
  - b. Suhu (X2)
  - c. Kelembaban\_udara (X3)
  - d. Berangin (X4)
  - e. Bermain\_Tenis (Y), sebagai variabel presictor.

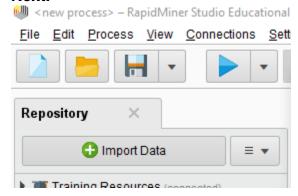
Tabel data Training pada Sheet1

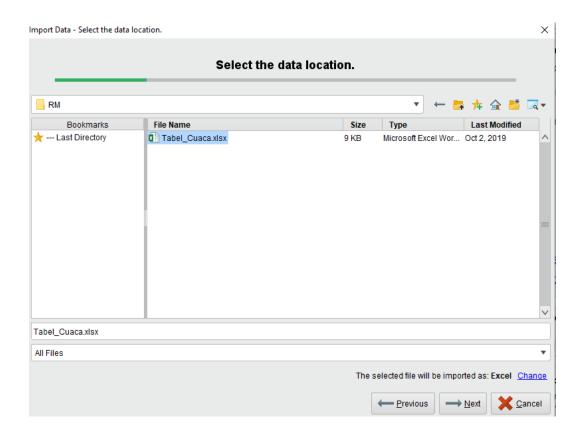
	Α	В	С	D	E
1	Cuaca	Suhu	Kelembab	Berangin	Bermain_Tenis
2	Cerah	85	85	TIDAK	TIDAK
3	Cerah	80	90	YA	TIDAK
4	Mendung	83	86	TIDAK	YA
5	Hujan	70	96	TIDAK	YA
6	Hujan	68	80	TIDAK	YA
7	Hujan	65	70	YA	TIDAK
8	Mendung	64	65	YA	YA
9	Cerah	72	95	TIDAK	TIDAK
10	Cerah	69	70	TIDAK	YA
11	Hujan	75	80	TIDAK	YA
12	Cerah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

Tabel data uji pda Sheet2 tanpa ada variabel Bermain\_Tenis.

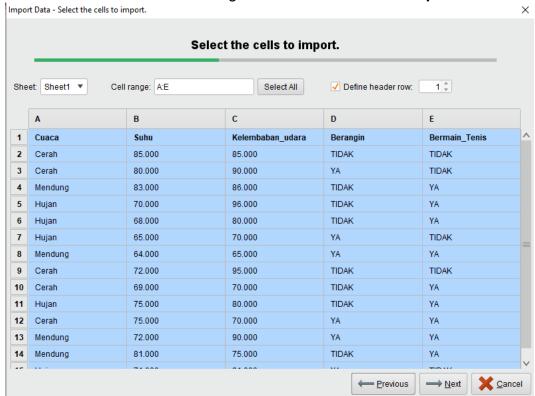
	Α	В	С	D
1	Cuaca	Suhu	Kelembab	Berangin
2	Cerah	75	65	TIDAK
3	Cerah	80	68	YA
4	Cerah	83	87	YA
5	Mendung	70	96	TIDAK
6	Mendung	68	81	TIDAK
7	Hujan	65	75	YA
8	Hujan	64	85	YA

- 4. Buka aplikasi **RapidMiner.** Menjalankan RapidMiner untuk pertama kali, terlebih dahulu membuat repositori baru. Repositori ini berfungsi sebagai lokasi penyimpanann terpusat untuk data dan proses analisa.
- 5. Klik **Import Data**. Arahkan direktori tempat penyimpanan file pada langkah **Select the data location**, kemudian pilih file yang akan digunakan dan klik **Next.**



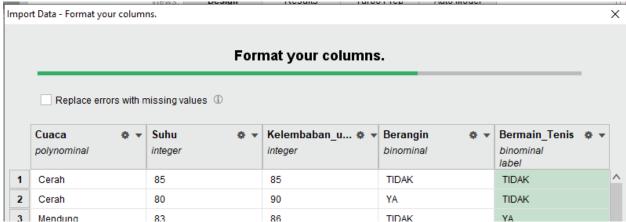


6. Pastikan sel Excel sesuai di langkah Select the cells to import.



7. Pada langkah **Format your colums** ubah kolom **Bermain\_Tenis** dengan tipe data **binomial** karena hanya ada dua keputusan (YA atau TIDAK).

8. Ubah pula sebagai label pada Change Role.



9. Simpan dengan nama DataCuaca\_Training dilanjutkan klik tombol Finish.

<u>N</u> ame	Tabel_CuacaTraining

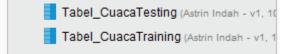
10. Hasil import file **Tabel\_Cuaca.xls** pada Sheet1 akan ditampilkan.



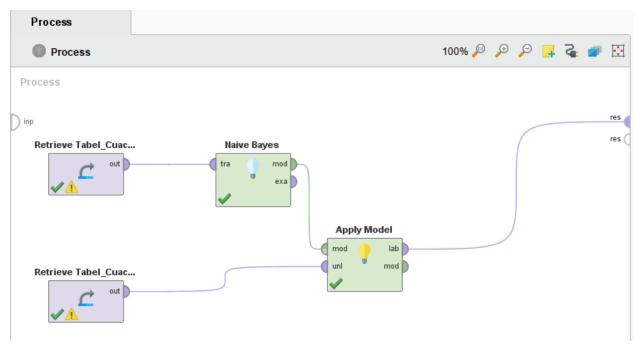
- 11. Kembali ke jendela Design Perspectve dengan shortcut tomblo F8.
- 12. Lakukan hal yang sama untuk data testing yang diambil dari **TabelCuaca.xls** PADA Sheet2 (Testing) dengan mengulang langkah 5.



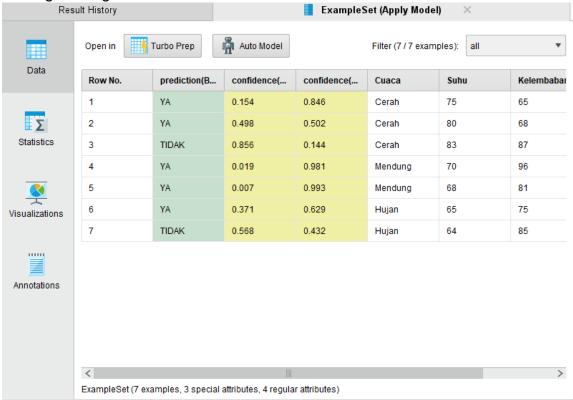
13. Simpan dengan nama DataCuaca\_Testing.



- 14. Langkah selanjutnya adalah membuat desain Naïve Bayes. Drag **DataCuaca\_Training** dan **DataCuaca\_Testing** ke dalam jendela Process View.
- 15. Masukkan juga operator Naïve Bayes dan Apply Model ke dalam Process View. Hubungkan konektor masing-masing data terhadap operator seperti gambar.



- 16. Jalankan proses naïve bayes dengan menekan tombol **Run** / menekan tombol F11.
- 17. Perhatikan hasil proses klasifiasi naïve bayes, Pada tab **Data** dapat dilihat hasil prediksi terhadap data testing secara tingkat confidence nilai kelas pada masing-masing data.



Pada tab **Statistic**, dapat dilihat bahwa distribusi nilai kelas pada variabel Y (Bermain\_Tenis) rerata nilai confidence sebesar 0.353 untuk nilai TIDAK, dan 0,647 untuk nilai YA.

	Name	<b>! !</b>	Туре	Missing	Statistics	Filter (7 / 7 attributes):	Search for Attributes  ▼ ▼
~	Prediction prediction(Bermain_Tenis)		Binominal	0	Least TIDAK (2)	Most YA (5)	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)	∨alues Cerah (3), Hujan (
~	Suhu		Integer	0	Min 64	Max 83	Average 72.143
~	Kelembaban_udara		Integer	0	Min 65	Max 96	Average 79.571
~	Berangin		Binominal	0	Least TIDAK (3)	Most YA (4)	YA (4), TIDAK (3)

18. Bandingkan dengan hasil prediksi menggunakan WEKA. Dapat dilihat bahwa prediksi masing-masing aplikasi menunjukkan hasil yang sama.

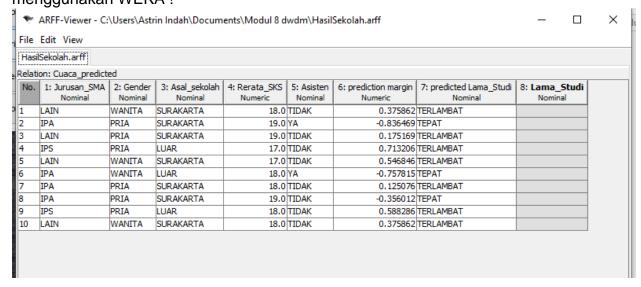
## **TUGAS**

1. Berdasarkan tabel berikut, buatlah dile dalam format Excel (.xls) dan format ARFF (.arff)! Data ini akan digunakan sebagai **data testing.** 

	Α	В	С	D	Е
1	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_Sekolah	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

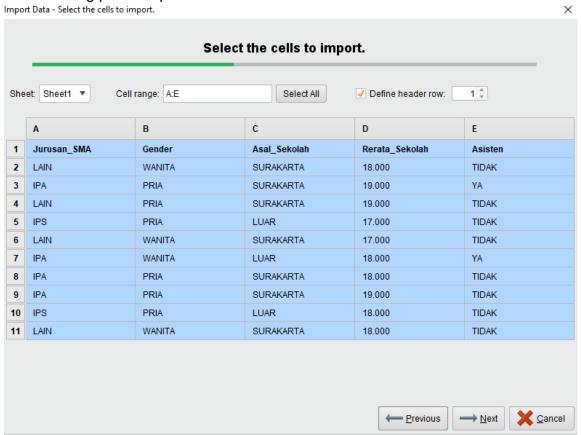
Dat	a Training	q				
	Α	В	С	D	E	F
1	Jurusan_S	Gender	Asal_Seko	Rerata_Se	Asisten	Lama_Studi
2	IPS	WANITA	SURAKART	18	TIDAK	TERLAMBAT
3	IPA	PRIA	SURAKART	19	YA	TEPAT
4	LAIN	PRIA	SURAKART	19	TIDAK	TERLAMBAT
5	IPA	PRIA	LUAR	17	TIDAK	TERLAMBAT
6	IPA	WANITA	SURAKART	17	TIDAK	TEPAT
7	IPA	WANITA	LUAR	18	YA	TEPAT
8	IPA	PRIA	SURAKART	18	TIDAK	TERLAMBAT
9	IPA	PRIA	SURAKART	19	TIDAK	TEPAT
10	IPS	PRIA	LUAR	18	TIDAK	TERLAMBAT
11	LAIN	WANITA	SURAKART	18	TIDAK	TEPAT
12	IPA	WANITA	SURAKART	19	TIDAK	TEPAT
13	IPS	PRIA	SURAKART	20	TIDAK	TEPAT
14	IPS	PRIA	SURAKART	19	TIDAK	TEPAT
15	IPA	PRIA	SURAKART	19	TIDAK	TEPAT
16	IPA	PRIA	LUAR	22	YA	TEPAT
17	LAIN	PRIA	SURAKART	16	TIDAK	TERLAMBAT
18	IPS	PRIA	LUAR	20	TIDAK	TEPAT
19	LAIN	PRIA	LUAR	23	YA	TEPAT
20	IPA	PRIA	SURAKART	21	YA	TEPAT
21	IPS	PRIA	SURAKART	19	TIDAK	TERLAMBAT

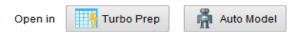
2. Gunakan file ARFF yang dikerjakan pada Tugas nomor 1 dalam Modul 7 sebagai data training. Lakukan prediksi terhadap data testing (ARFF) di atas menggunakan WEKA!



3. Gunakan file Excel yang dikerjakan pada Tugas nomor 1 dalam Modul 6 sebagai data training. Lakukan prediksi terhadap data testing (Excel) di atas menggunakan RapidMiner.

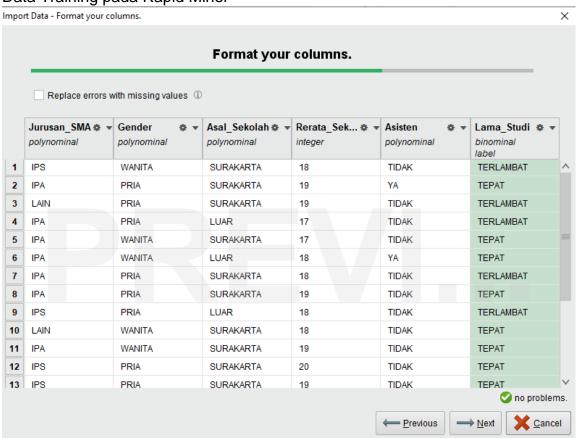
# Data Testing pada Rapid Miner





Row No.	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_Sek	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

## Data Training pada Rapid Miner





Filter (20 / 20 exam)

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

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

#### Hasil Prediksi

Row No.	prediction(L	confidence(	confidence(	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_Sek	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

- 4. Dari hasil percobaan Tugas nomor 3 di atas, berapakh nilai rerata confidence untuk atribut Lama\_studi dengan nilai TEPAT? Berapakah nilai rerata confidence untuk atribut Lama\_Studi dengan nuali TERLAMBAT?
- Dari hasil percobaan Tugas nomor 3 diatas, berapa orang yang akan lulus TEPAT, dan berapa orang yang akan lulus TERLAMBAT? Tambahkan 2 kondisi berikut pada testing.

1 Jurusan S Gender Asal Sekolah Rerata Sekolah	
Todadal_Sector	Asisten
2 IPA WANITA LUAR 18	TIDAK
3 LAIN PRIA SURAKARTA 17	YA

- 6. Prediksikan ketepatan lama studi si 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.
- 7. Prediksikan ketepatan lama studi Jono , jika Jono adalah seorang PRIA yang berasal dari jurusan selain IPA dan IPS pada saat SMA, asal sekolah dari SURAKARTA, mengambil SKS dengan rata-rata sebanyak 17 SKS tiap semester, dan pernah menjadi Asisten selama Sekolah.

