

# **LAPORAN PRAKTIKUM DATA WAREHOUSING DAN DATA MINING**

## **PERTEMUAN 6**

### **“ALGORITMA KLASIFIKASI: K-NEAREST NEIGHBOR(K-NN)”**



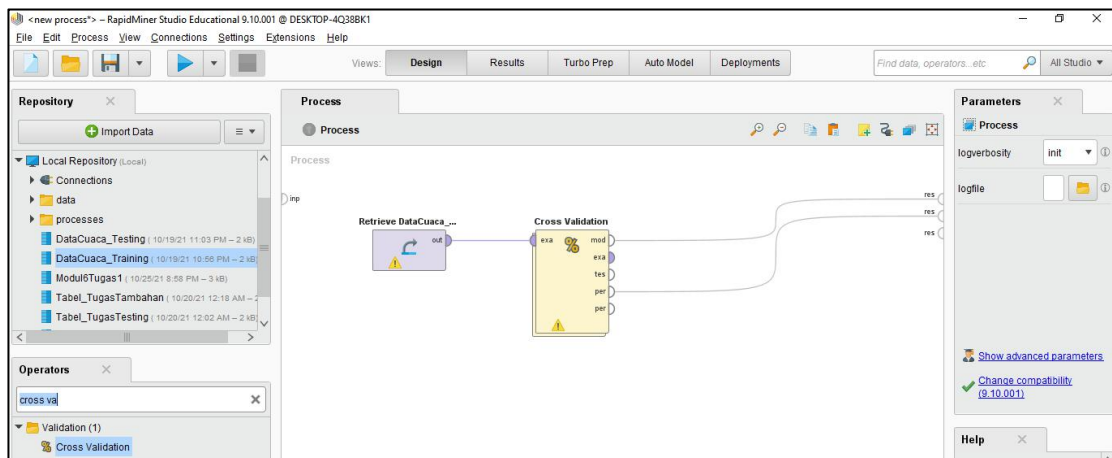
**Oleh:**

**NAMA : Daffa Putra Alwansyah**  
**NIM : L200190031**  
**KELAS : B**  
**PRODI : INFORMATIKA**

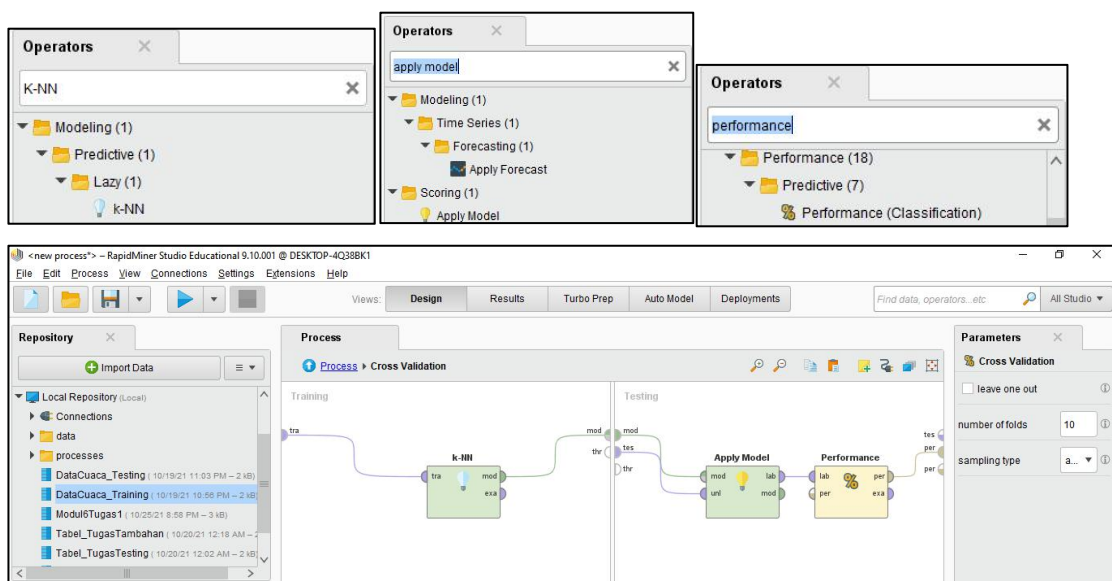
**Fakultas Komunikasi dan Informatika**  
**Universitas Muhammadiyah Surakarta**

## Kegiatan 4.1 (Menentukan Nilai k dan Tingkat Akurasi Algoritma K-NN)

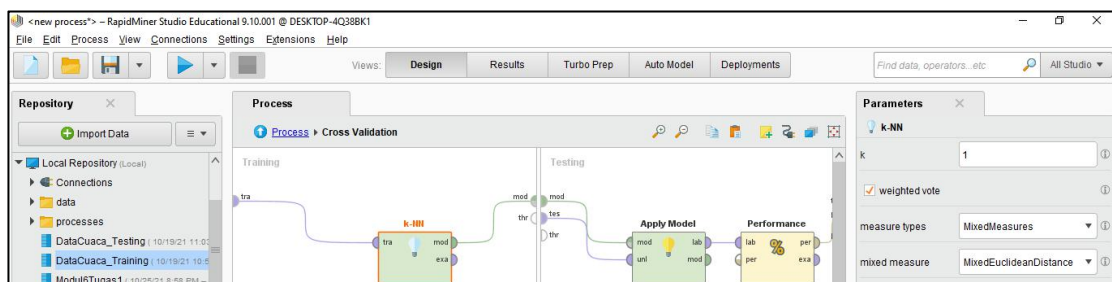
1. Menggunakan file DataCuaca\_Training dan operator Cross Validation, lalu menghubungkan port sesuai dimodul.



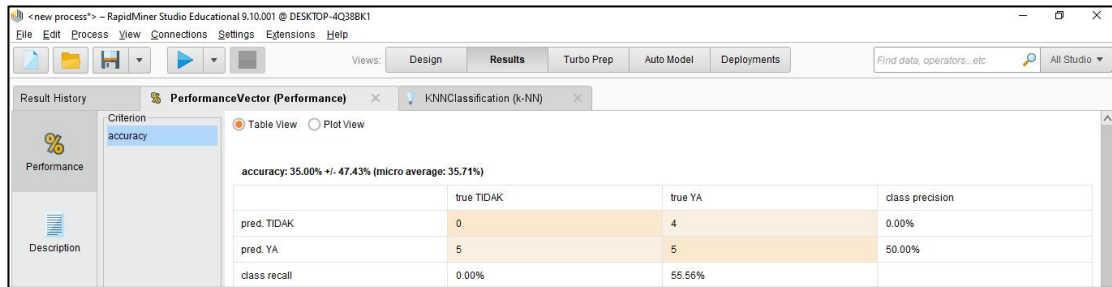
2. Klik 2x pada Cross Validation, masukan 'K-NN', 'Apply Model' dan 'Performance', lalu sambungkan port sesuai modul.



3. Ubah parameter pada k-NN menjadi 1, setelah itu klik process dan klik RUN.



4. Hasil proses klasifikasi dengan algoritma K-NN pada bagian:  
PerformanceVector (Performance).



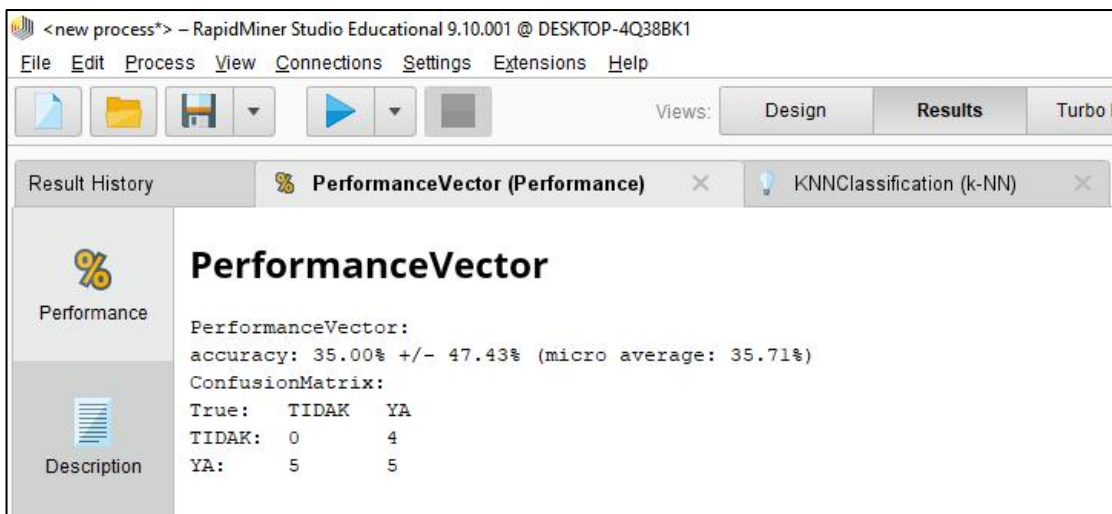
PerformanceVector (Performance)

Table View Plot View

accuracy: 35.00% +/- 47.43% (micro average: 35.71%)

	true TIDAK	true YA	class precision
pred. TIDAK	0	4	0.00%
pred. YA	5	5	50.00%
class recall	0.00%	55.56%	

#### Mode Text View

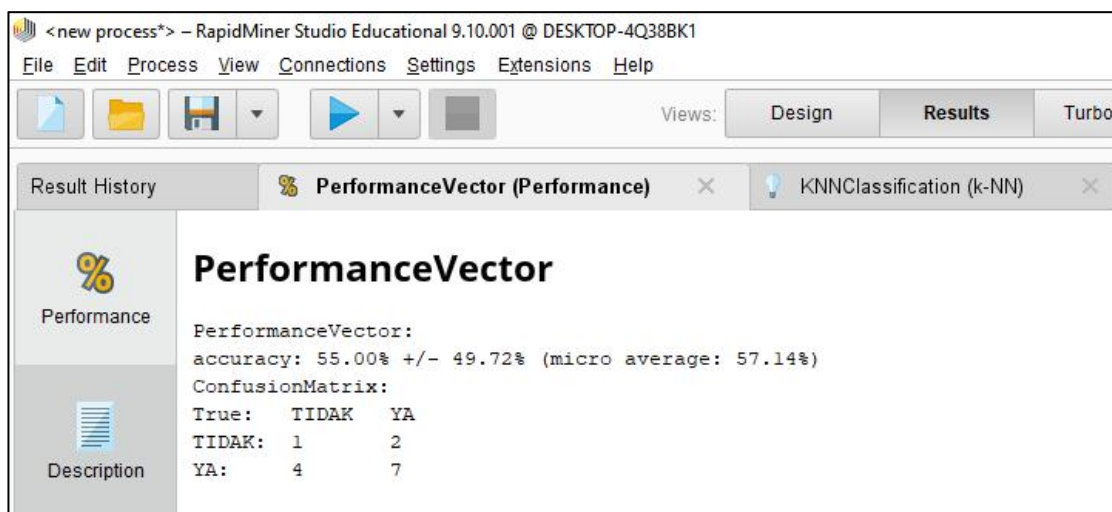


PerformanceVector

PerformanceVector:  
accuracy: 35.00% +/- 47.43% (micro average: 35.71%)  
ConfusionMatrix:  
True: TIDAK YA  
TIDAK: 0 4  
YA: 5 5

- Tingkat akurasi pada percobaan ini mencapai 35% yang menunjukkan angka keakuratannya yang sangat rendah.

5. Dikarenakan tingkat akurasinya rendah, maka ubah parameter k pada K-NN menjadi angka 6.



PerformanceVector

PerformanceVector:  
accuracy: 55.00% +/- 49.72% (micro average: 57.14%)  
ConfusionMatrix:  
True: TIDAK YA  
TIDAK: 1 2  
YA: 4 7

- Hasil tingkat akurasi yang diperoleh sebesar 55% dengan Mode Text View

6. Ubah nilai K dengan angka sembarang dari 1 - 14 (saya menggunakan 7).

The screenshot shows the RapidMiner Studio interface with the 'Results' view selected. The 'PerformanceVector (Performance)' tab is active, displaying the following data:

**PerformanceVector**

PerformanceVector:  
accuracy: 65.00% +/- 41.16% (micro average: 64.29%)

ConfusionMatrix:  
True: TIDAK YA  
TIDAK: 0 0  
YA: 5 9

- Ternyata, tingkat akurasi maksimal yang diperoleh sebesar 65% dan ini terjadi pada saat nilai  $k = 4, 7, 9, 11, 12, 13$  dan  $14$ .

7. Mengubah Kembali nilai  $k = 4$  dan memberikan tanda cek pada pilihan weighted vote.

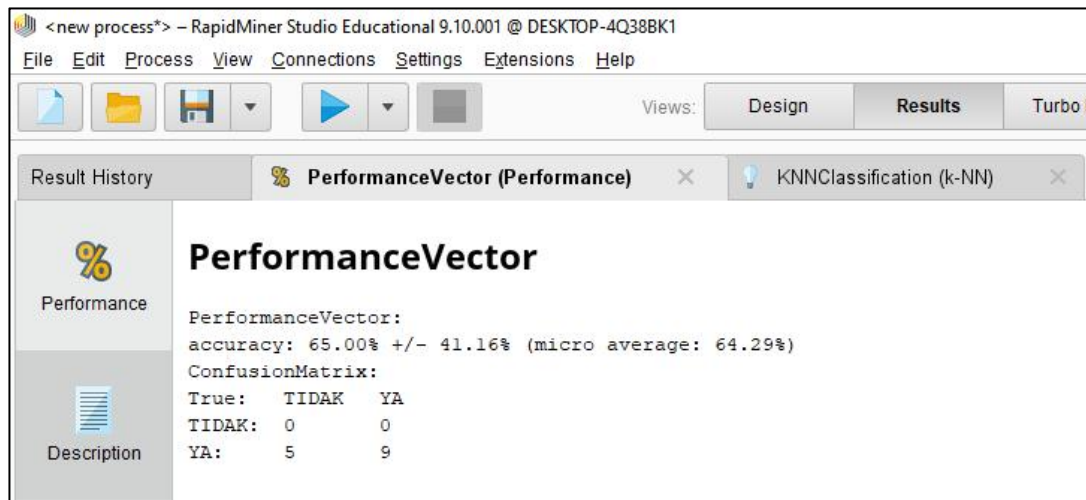
The screenshot shows the RapidMiner Studio interface with the 'Results' view selected. The 'PerformanceVector (Performance)' tab is active, displaying the following data:

**PerformanceVector**

PerformanceVector:  
accuracy: 50.00% +/- 47.14% (micro average: 50.00%)

ConfusionMatrix:  
True: TIDAK YA  
TIDAK: 1 3  
YA: 4 6

8. Mengubah Kembali nilai  $k = 7$  dan memberikan tanda cek pada pilihan weighted vote.



PerformanceVector

PerformanceVector:

accuracy: 65.00% +/- 41.16% (micro average: 64.29%)

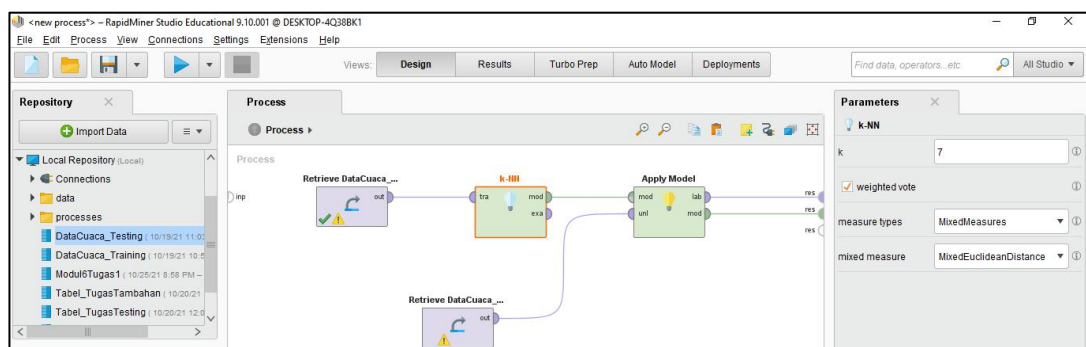
ConfusionMatrix:

True:	TIDAK	YA
TIDAK:	0	0
YA:	5	9

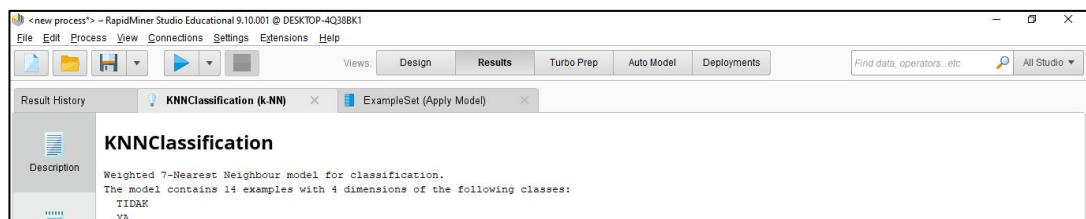
- Nilai  $k = 4$  dengan menambahkan bobot weight vote tingkat akurasi berkurang menjadi 50%, Nilai  $k = 7$  tingkat akurasi nya tetap sama sebesar 65% meskipun diberi bobot weighted vote. Sehingga dapat disimpulkan bahwa nilai  $k$  yang akan digunakan untuk melakukan prediksi terhadap data testing adalah  $k = 7$ .

## KEGIATAN 4.2 (Prediksi dengan Algoritma K-NN)

1. Masukkan DataCuaca\_Testing & DataCuca\_Training, lalu masukan operator K-NN beserta Apply Model, setelah itu ubah nilai  $k = 7$  pada K-NN. Hubungkan port sesuai modul, klik process dan tombol RUN.



2. Berikut hasil prosesnya:



KNNClassification

Weighted 7-Nearest Neighbour model for classification.

The model contains 14 examples with 4 dimensions of the following classes:

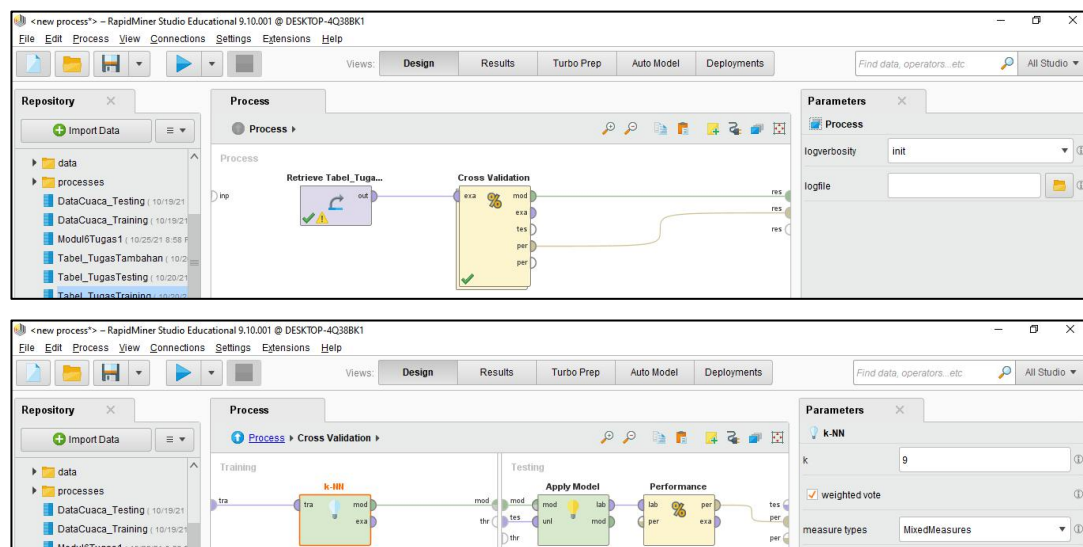
Class	Count
TIDAK	5
YA	9



Row No.	prediction(B...	confidence(...	confidence(...	Cuaca	Suhu	Kelembaban...	Berangin
1	YA	0.143	0.857	Cerah	75	65	TIDAK
2	YA	0.137	0.863	Cerah	80	68	YA
3	YA	0.441	0.559	Cerah	83	87	YA
4	YA	0.445	0.555	Mendung	70	96	TIDAK
5	YA	0.277	0.723	Mendung	68	81	TIDAK
6	YA	0.154	0.846	Hujan	65	75	YA
7	YA	0.420	0.580	Hujan	64	85	YA

## TUGAS!

1. Menggunakan DataStudi\_Training (nama di saya beda) pada respository sebagai data training, Kemudian mencari nilai k yang memiliki tingkat akurasi tertinggi dengan menggunakan algoritma k-NN dari data training tersebut.

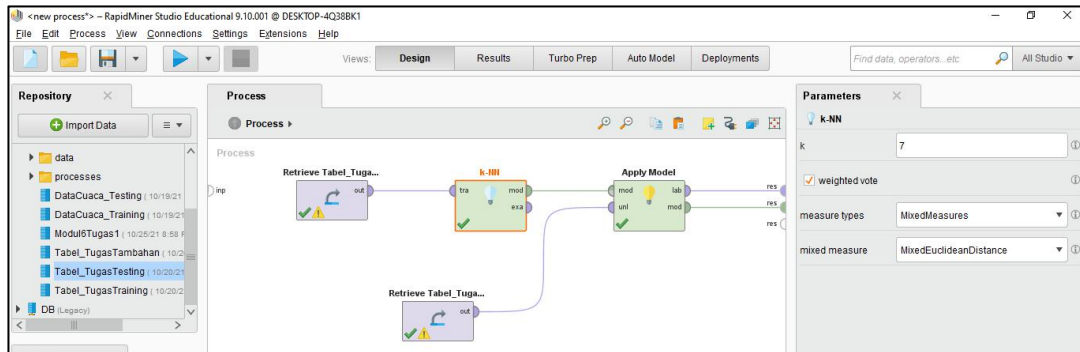


Hasil:

pred. TERLAMBAT	true TERLAMBAT	pred. TEPAT	true TEPAT	class precision
3	0	0	100.00%	
4	13	76.47%		
class recall	42.86%	100.00%		

- Memasukan nilai k = 9 dan memberi weighted vote menghasilkan akurasi sebesar 80 %, angka 1-14 yang merupakan akurasi tertinggi adalah angka 9 dengan menghasilkan akurasi sebesar 80%.

2. Menggunakan file DataStudi\_Testing pada repository sebagai data testing,  
Lakukan prediksi data testing tersebut menggunakan algoritma k-NN.



Hasil:

The screenshot shows the 'Results' view of RapidMiner Studio. The 'Result History' pane at the top lists 'KNNClassification (k-NN)', 'ExampleSet (Apply Model)', and 'ExampleSet (/Local Repository/Tabel\_TugasTraining)'. The 'Description' pane for 'KNNClassification' shows: 'Weighted 7-Nearest Neighbour model for classification. The model contains 20 examples with 5 dimensions of the following classes: TERLAMBAT, TEPAT'.

The screenshot shows the 'Results' view of RapidMiner Studio. The 'Result History' pane at the top lists 'KNNClassification (k-NN)', 'ExampleSet (Apply Model)', and 'ExampleSet (/Local Repository/Tabel\_TugasTraining)'. The 'Data' pane shows a table with 10 rows and 8 columns. The table is filtered to show 10 examples.

Row No.	prediction(L...	confidence(L...	confidence(L...	Jurusan_SMA	Gender	Asal_Sekolah	Rerata_SKS	Asisten
1	TEPAT	0.424	0.576	LAIN	WANITA	SURAKARTA	18	TIDAK
2	TEPAT	0.272	0.728	IPA	PRIA	SURAKARTA	19	YA
3	TEPAT	0.441	0.559	LAIN	PRIA	SURAKARTA	19	TIDAK
4	TERLAMBAT	0.724	0.276	IPS	PRIA	LUAR	17	TIDAK
5	TERLAMBAT	0.565	0.435	LAIN	WANITA	SURAKARTA	17	TIDAK
6	TERLAMBAT	0.556	0.444	IPA	WANITA	LUAR	18	YA
7	TERLAMBAT	0.574	0.426	IPA	PRIA	SURAKARTA	18	TIDAK
8	TEPAT	0.267	0.733	IPA	PRIA	SURAKARTA	19	TIDAK
9	TERLAMBAT	0.726	0.274	IPS	PRIA	LUAR	18	TIDAK
10	TEPAT	0.424	0.576	LAIN	WANITA	SURAKARTA	18	TIDAK