

LAMPIRAN

3.2 Tahapan Pemodelan Data

tahap awal data yang ada akan dilakukan pemodelan dengan algoritma Forward Selection, yaitu dengan menggunakan model-model basis yang telah ditraining untuk melakukan klasifikasi pada testing dataset

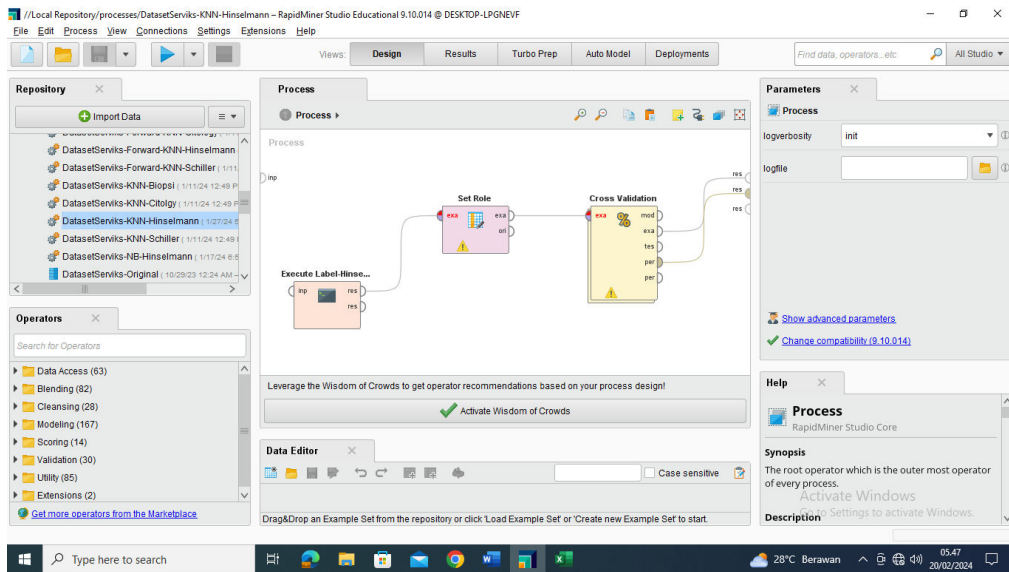
1. Read Dataset

TIPE DATA PADA TIAP KOLOM PADA DATASET					
Kolom	Tipe Data	Kolom	Tipe Data	Kolom	Tipe Data
<i>Age</i>	<i>Integer</i>	<i>STDs (number)</i>	<i>Boolean</i>	<i>STDs: HPV</i>	<i>Boolean</i>
<i>Number of sexual partners</i>	<i>Integer</i>	<i>STDs: condylomatosis</i>	<i>Boolean</i>	<i>STDs: Number of diagnosis</i>	<i>Integer</i>
<i>First sexual intercourse</i>	<i>Integer</i>	<i>STDs: cervical condylomatosis</i>	<i>Boolean</i>	<i>STDs: Time since first diagnosis</i>	<i>Integer</i>
<i>Number of pregnancies</i>	<i>Integer</i>	<i>STDs: vaginal condylomatosis</i>	<i>Boolean</i>	<i>STDs: Time since last diagnosis</i>	<i>Integer</i>
<i>Smokes</i>	<i>Boolean</i>	<i>STDs: vulvo-perineal condylomatosis</i>	<i>Boolean</i>	<i>Dx: Cancer</i>	<i>Boolean</i>
<i>Smokes (years)</i>	<i>Integer</i>	<i>STDs: syphilis</i>	<i>Boolean</i>	<i>Dx: CIN</i>	<i>Boolean</i>
<i>Smokes (packs/year)</i>	<i>Boolean</i>	<i>STDs: pelvic inflammatory disease</i>	<i>Boolean</i>	<i>Dx: HPV</i>	<i>Boolean</i>
<i>Hormonal Contraceptives</i>	<i>Boolean</i>	<i>STDs: genital herpes</i>	<i>Boolean</i>	<i>Dx</i>	<i>Boolean</i>
<i>Hormonal Contraceptives (years)</i>	<i>Integer</i>	<i>STDs: molluscum contagiosum</i>	<i>Boolean</i>	<i>Hinselmann (Label)</i>	<i>Boolean</i>
<i>IUD</i>	<i>Boolean</i>	<i>STDs: AIDS</i>	<i>Boolean</i>	<i>Schiller (Label)</i>	<i>Boolean</i>
<i>IUD (years)</i>	<i>Integer</i>	<i>STDs: HIV</i>	<i>Boolean</i>	<i>Citology (Label)</i>	<i>Boolean</i>
<i>STDs</i>	<i>Boolean</i>	<i>STDs: Hepatitis B</i>	<i>Boolean</i>	<i>Biopsy (Label)</i>	<i>Boolean</i>

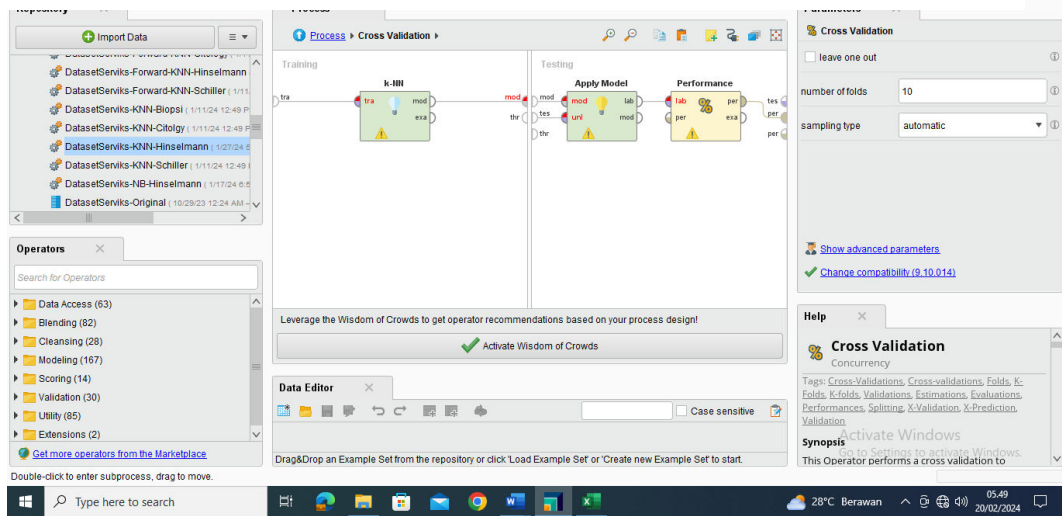
Gambar 3.2 dataset berdasarkan label

Algoritma	Forward & KNN
Jenis Data	Publick
Alamat Dataset	https://www.kaggle.com/datasets/loveall/cervical-cancer-risk-classification

2. KNN

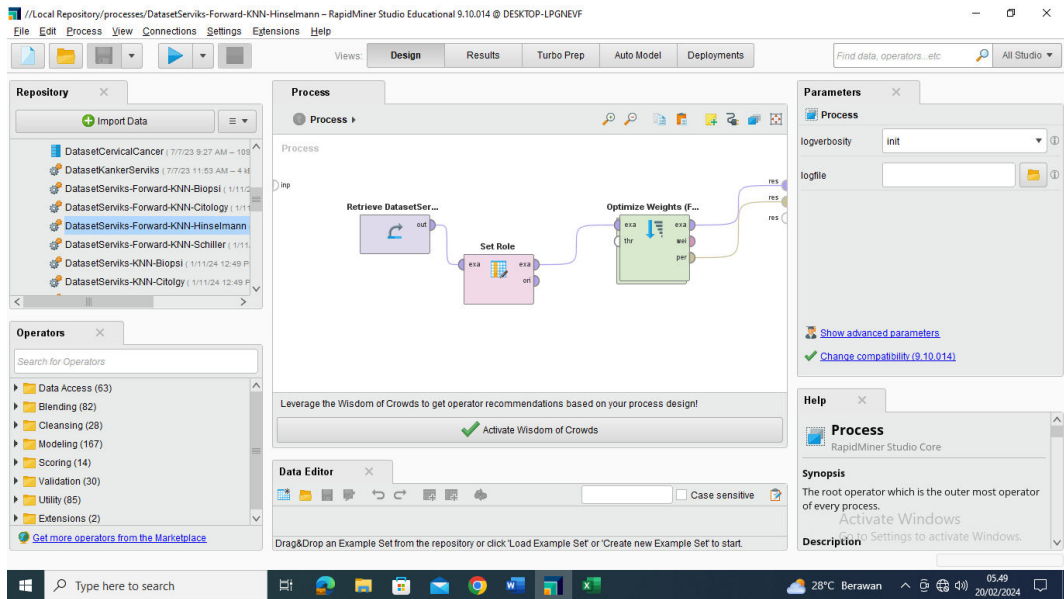


Gambar 3.3 Pengujian Algoritma KNN dengan Rapidminer

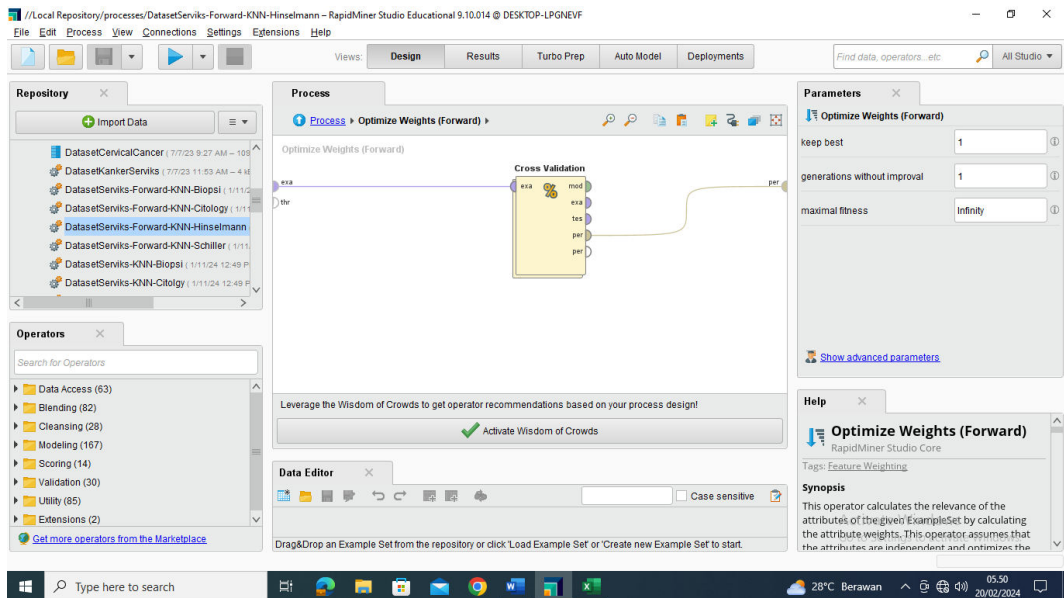


Gambar 3.3 Pengujian Algoritma KNN dengan Rapidminer

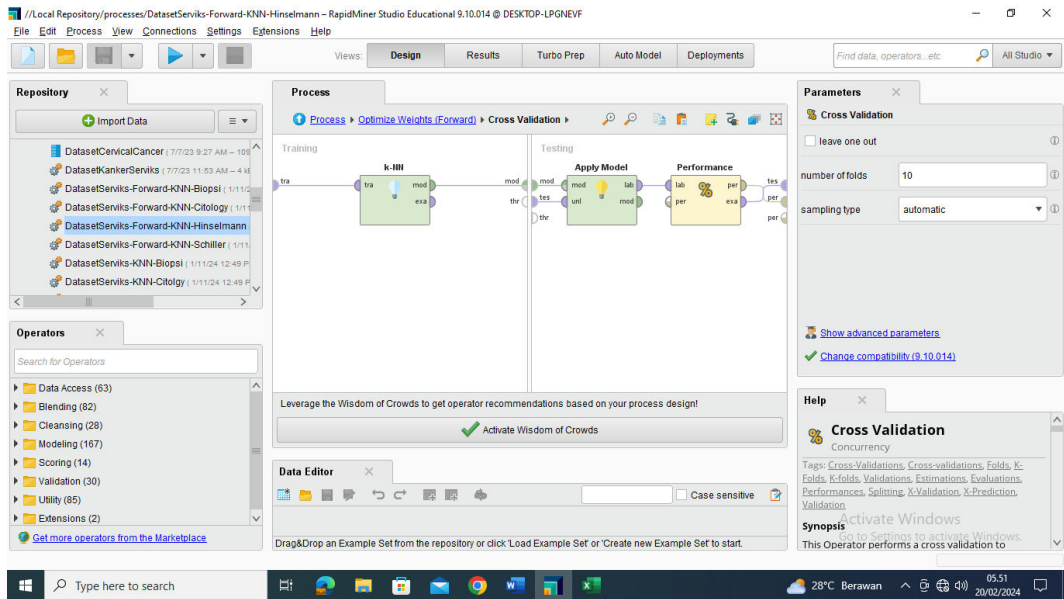
3. Forward Selection



Gambar 3.4 Pengujian Algoritma Forward-KNN dengan Rapidminer-1



Gambar 3.5 Pengujian Algoritma Forward-KNN dengan Rapidminer-2



Gambar 3.6 Pengujian Algoritma Forward-KNN dengan Rapidminer-3

4. Evaluasi

Eksperimen dengan menerapkan teknik Forward Selection, yang menggabungkan beberapa model base yang beragam untuk meningkatkan akurasi prediksi, selama eksperimen ini. Hasil dari eksperimen ini yang merupakan informasi kunci nuntuk evaluasi kualitas prediksi kanker serviks. Dapat dilihat pada gambar dibawah ini:

Criterion

- accuracy
- precision
- recall
- AUC (optimistic)
- AUC
- AUC (pessimistic)

Table View Plot View

accuracy: 97.35% +/- 1.59% (micro average: 97.35%)

	true 0	true 1	class precision
pred. 0	788	15	98.13%
pred. 1	7	20	74.07%
class recall	99.12%	57.14%	

Gambar 3.7 Hasil Akurasi Forward-KNN

4.1.1 Hasil dari Proses *Import Dataset*

untuk melihat hasil dari dataset yang telah di import maka hubungkan operator “Read Excel” dengan Res (Result). Gambar 4.1 merupakan gambar dataset teratas yang muncul Ketika menjalankan operator “Read Excel”

hasil yang tampilan adalah keseluruhan data dari dataset yang telah di dapatkan.

Selain itu dapat di lihat juga statistic dari dataset yang di jalankan. Gambar 4.2 adalah analisis statistic dari dataset yang di jalankan:

Row No.	Age	Number of s...	First sexual ...	Num of preg...	Smokes	Smokes (ye...	Smokes (pa...	Hormonal C...	Horr
1	18	4	15	1	0	0	0	0	0
2	15	1	14	1	0	0	0	0	0
3	34	1	?	1	0	0	0	0	0
4	52	5	16	4	1	37	37	1	3
5	46	3	21	4	0	0	0	1	15
6	42	3	23	2	0	0	0	0	0
7	51	3	17	6	1	34	3.4	0	0
8	26	1	26	3	0	0	0	1	2
9	45	1	20	5	0	0	0	0	0
10	44	3	15	?	1	1	2.8	0	0
11	44	3	26	4	0	0	0	1	2
12	27	1	17	3	0	0	0	1	8
13	45	4	14	6	0	0	0	1	10
14	44	2	25	2	0	0	0	1	5

Gambar Fehler! Kein Text mit angegebener Formatvorlage im Dokument..1 Sampel dataset teratas dari seluruh atribut

Name	Type	Missing	Min	Max	Average
Age	Integer	0	13	84	26.821
Number of sexual partners	Integer	26	1	28	2.528
First sexual intercourse	Integer	7	10	32	16.995
Num of pregnancies	Integer	56	0	11	2.276
Smokes	Binominal	13	Negative 0	Positive 1	Values 0 (722), 1 (123)
Smokes (years)	Integer	13	0	37	1.217
Smokes (packs/year)	Nominal	13	Negative 7.6 (1)	Positive 0 (722)	Values 0 (722), 0.5132
Hormonal Contraceptives	Binominal	108	Negative 0	Positive 1	Values 1 (481), 0 (269)

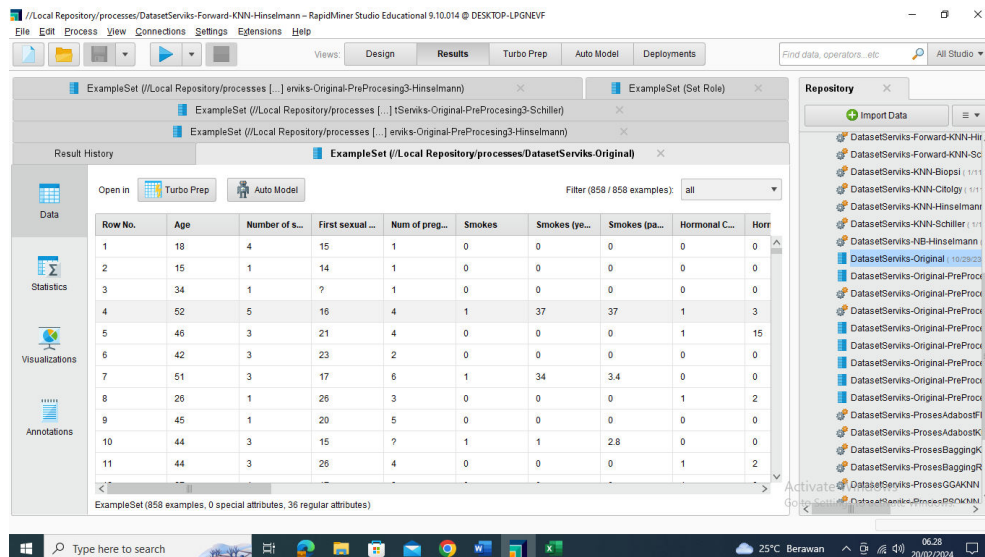
hasil analisa statistik yang muncul dapat berupa *values, least, most, type data*, dan

Gambar Fehler! Kein Text mit angegebener Formatvorlage im Dokument..2 analisa statistik dari keseluruhan dataset

total *missing values*

4.1.2 Hasil dari seleksi atribut

Untuk melihat hasil dari tahap Feature Selection yang telah di lakukan, koneksikan operator “Select Attribute” dengan Res (Result). Gambar 4.3 adalah hasil dari dataset yang ditampilkan dan telah di lakukan tahap pemilihan atribut:



Output Gambar Fehler! Kein Text mit angegebener Formatvorlage im Dokument..3 Sampel data teratas sesuai dengan atribut yang sudah di pilih, dan atribut filter n proses

Analisa nantinya.

4.1.3 Hasil dari pemberian label pada atribut

Untuk melihat hasil dari atribut yang sudah di beri label maka koneksikan operator “Set Role” ke Res (Result). Gambar 4.4 adalah tampilan dataset teratas yang sudah diberi label:

Row No.	Hinselmann	Age	Number of s...	First sexual ...	Num of preg...	Smokes	Smokes (ye...	Smokes (pa...	Horr
1	0	18	4	15	1	0	0	0	0
2	0	15	1	14	1	0	0	0	0
3	0	34	1	17	1	0	0	0	0
4	0	52	5	16	4	1	37	37	1
5	0	46	3	21	4	0	0	0	1
6	0	42	3	23	2	0	0	0	0
7	1	51	3	17	6	1	34	3.4	0
8	0	26	1	26	3	0	0	0	1
9	0	45	1	20	5	0	0	0	0
10	0	44	3	15	2	1	1	2.8	0
11	0	44	3	26	4	0	0	0	1
12	0	27	1	17	3	0	0	0	1
13	0	45	4	14	6	0	0	0	1

Gambar Fehler! Kein Text mit angegebener Formatvorlage im Dokument..4 Sampel dataset teratas yang sudah diberi label

Output yang muncul adalah keseluruhan atribut yang sudah melalui tahap seleksi atribut dan ada 1 atribut yang sudah di beri label (berwarna hijau)

4.1.4 Hasil *modelling* dari proses *training* dan *testing* dataset

Hasil dari proses *training* dan *testing* yang di lakukan di dalam rapid miner adalah sebuah model yang bertujuan agar rapid miner dapat memunculkan output berupa *performance* dari masing-masing algoritma. Gambar 4.5 adalah hasil *modelling* proses *training* dan *testing* dataset dengan menggunakan algoritma KNN:

Gambar Fehler! Kein Text mit angegebener Formatvorlage im Dokument..5 Hasil modelling dengan menggunakan algoritma KNN-1

The screenshot displays the RapidMiner Studio Educational 9.10.014 interface. The main workspace shows a process design with the following operators: 'Execute Label Hints', 'Set Role', and 'Cross Validation'. The 'Parameters' panel on the right shows the following settings:

- logverbosity: init
- logfile: [empty field]

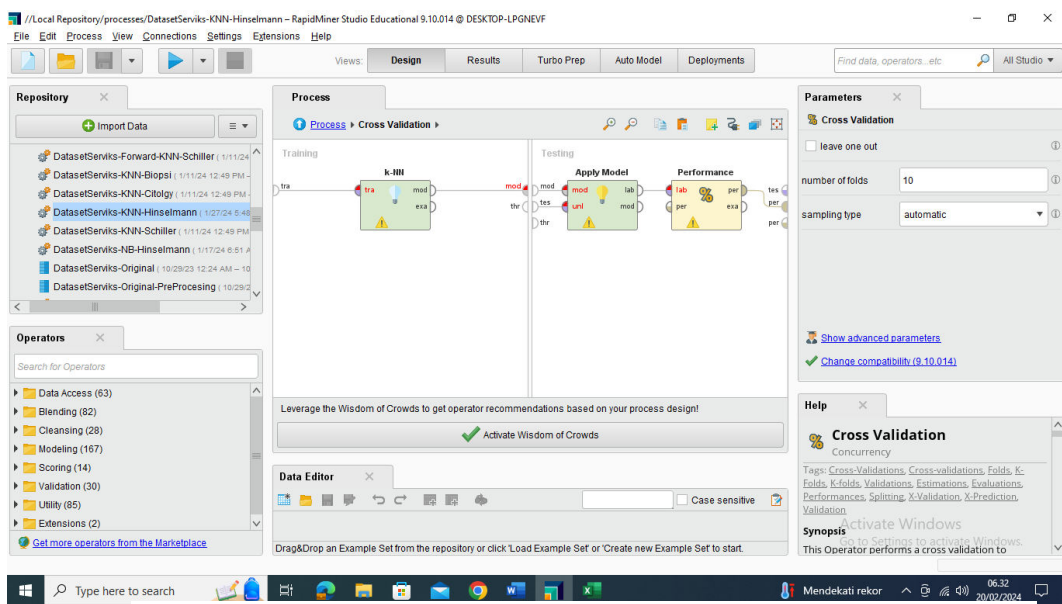
The 'Help' panel shows the following information for the 'Process' operator:

- Synopsis:** The root operator which is the outer most operator of every process.
- Description:** Activate Windows. Go to Settings to activate Windows.

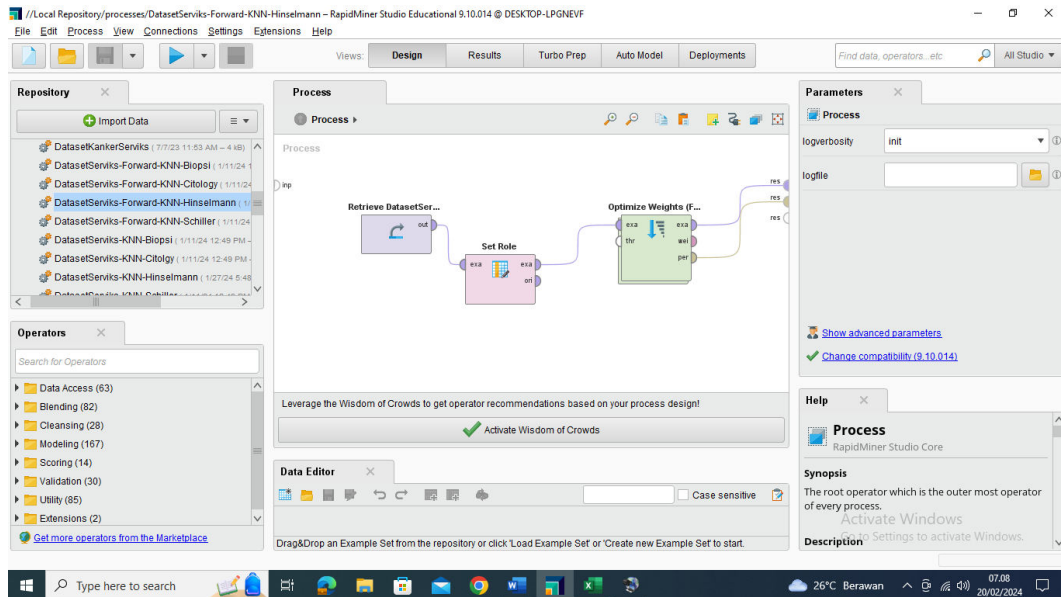
The bottom of the screenshot shows the Windows taskbar with the system clock at 06:30 on 20/02/2024.

Ketika proses *preparation* data dan *preprocessing* data telah dilakukan dan menjalankan model *training* dari algoritma KNN ini, maka output yang muncul adalah *confusion matrix* yang dapat di Analisa nilai performa algoritma KNN untuk mengklasifikasi data berupa akurasi

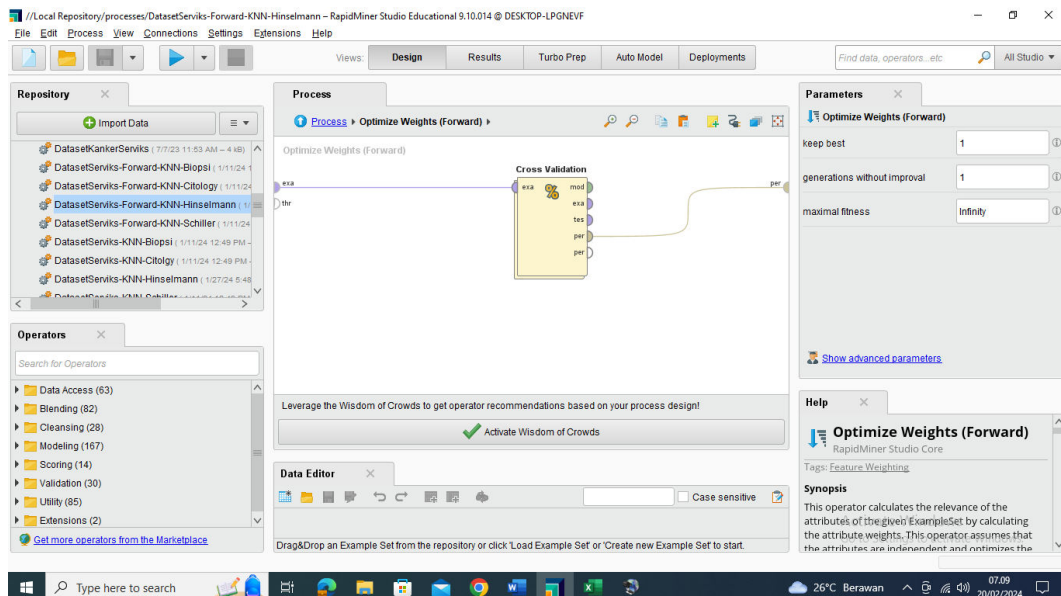
Gambar 4.6 adalah hasil modelling proses training dan testing dataset dengan menggunakan algoritma KNN:



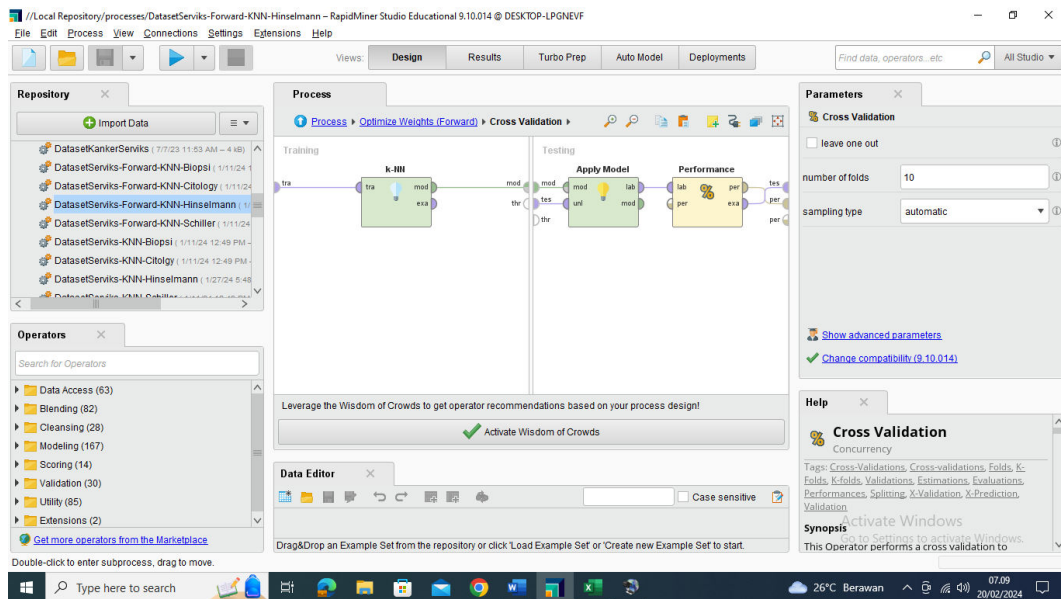
Gambar Fehler! Kein Text mit angegebener Formatvorlage im Dokument..6 Hasil modelling dengan menggunakan algoritma KNN



Gambar Fehler! Kein Text mit angegebener Formatvorlage im Dokument..7
 Hasil modelling dengan menggunakan algoritma Forward-KNN-1



Gambar Fehler! Kein Text mit angegebener Formatvorlage im Dokument..8
 Hasil modelling dengan menggunakan algoritma Forward-KNN-2



Gambar Fehler! Kein Text mit angegebener Formatvorlage im Dokument..9
Hasil modelling dengan menggunakan algoritma Forward-KNN-2