

BAB IV

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan jumlah data sebanyak 425 record data Kelulusan Mahasiswa dengan 12 Atribut yaitu NIM, Nilai Ijazah SMA, Jurusan, Jenis Kelamin, Usia, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, IPK, Masa Studi, dan Kategori Kelulusan . Dalam penelitian ini menggunakan Software *RapidMiner*.

4.1 Praproses

Tahap pembersihan data merupakan awal dari proses Data Mining. Proses pembersihan data mencakup antara lain memeriksa data yang tidak konsisten, data dengan missing value dan redundant data. Seluruh atribut pada dua kelompok data (tabel) dibersihkan karena hal tersebut merupakan syarat awal untuk proses data mining yang akan menghasilkan dataset yang bersih dan siap digunakan pada tahap mining data. Dikatakan missing value jika pada salah satu atribut nilai record tersebut hilang maka record yang dimaksud akan dihapus, karena record tersebut dinilai kehilangan data atau missing value. Apabila dalam dataset yang sama terdapat lebih dari satu record yang berisi nilai yang sama, maka record yang dimaksud juga harus dihapus karena tidak akan memberi informasi yang berarti jika dipertahankan. Tahap ini tidak hanya membersihkan data yang mengandung missing value saja akan tetapi terhadap data yang tidak konsisten juga dilakukan.

Data pada penelitian ini merupakan data yang sudah konsisten. Karena dua kelompok data (tabel) diambil seluruhnya tidak ada data yang dicleaning, maka jumlah atribut dan record pada kelompok data (tabel) adalah tetap. Pada tahap ini data sudah bersih dan siap untuk digunakan pada tahap selanjutnya yaitu integrasi data.

4.2 Pengelompokkan Data

Data yang digunakan bersumber dari BAAK Universitas Nahdlatul Ulama Lampung, jenjang strata 1 tahun angkatan 2022-2023 yang telah lulus sebanyak 425 orang . Data dibagi menjadi 70% data training dan 30% data testing. Atribut yang menjadi parameter sebanyak 12 atribut, dimana 11 atribut predictor dan 1 atribut hasil.

Tabel 4.1 Sampel Data

NIM	Nilai Ijazah SMA	Jurusan	JK	Usia	IPS 1	IPS 2	IPS 3	IPS 4	IPK	Masa Studi	Kategori
1988203012	85	Pendidikan Bahasa Inggris	P	24	3.80	3.79	3.77	3.76	3.77	4 th	Pujian
1988203015	81	Pendidikan Bahasa Inggris	P	25	3.74	3.75	3.72	3.69	3.70	4 th	Pujian
1988203021	80	Pendidikan Bahasa Inggris	P	24	3.74	3.73	3.72	3.69	3.71	4 th	Pujian
1820201030	78	Teknik Elektro	L	27	3.10	2.90	2.85	2.91	2.94	4,5 th	Memuaskan
1820201032	77	Teknik Elektro	L	27	3.10	2.90	2.80	2.79	2.80	4,5 th	Memuaskan
1820201033	81	Teknik Elektro	L	28	2.90	2.85	2.79	2.77	2.75	4,5 th	Memuaskan
1854201121	84	Agribisnis	L	25	3.65	3.60	3.55	3.50	3.54	4,5 th	Pujian
1854201125	82	Agribisnis	P	27	3.45	3.40	3.30	3.25	3.26	4,5 th	Sangat Memuaskan
1854201127	85	Agribisnis	P	27	3.70	3.65	3.60	3.61	3.62	4,5 th	Pujian

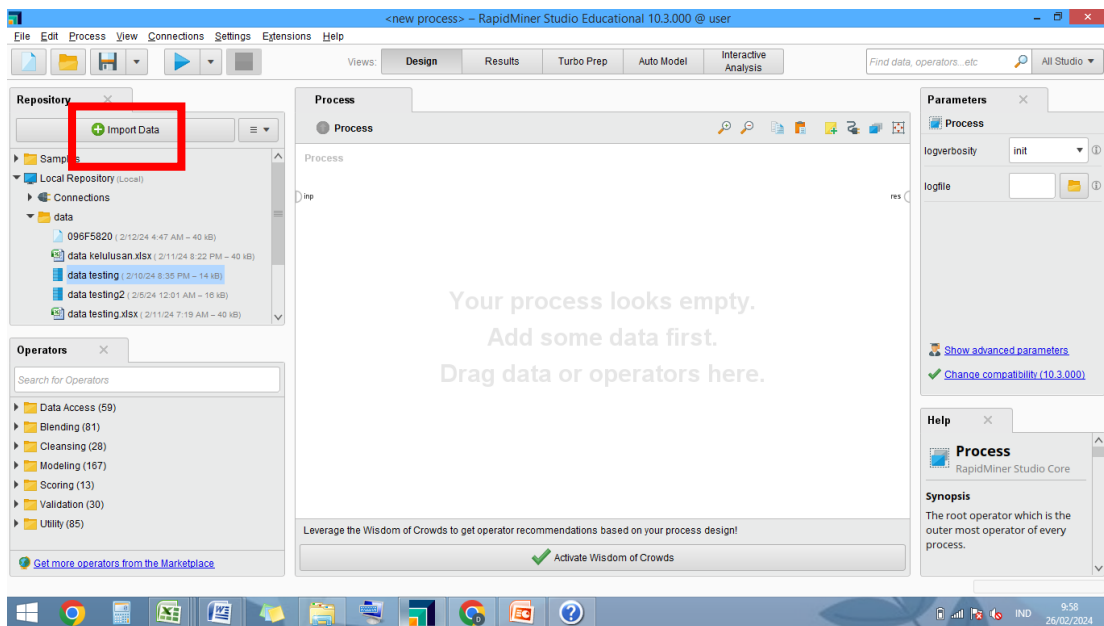
Masing-masing data dikelompokkan menjadi 3(tiga) kategori yaitu Memuaskan, Sangat Memuaskan dan Pujian dengan ketentuan sebagai berikut :

Tabel 4.2 Kategori Kelulusan

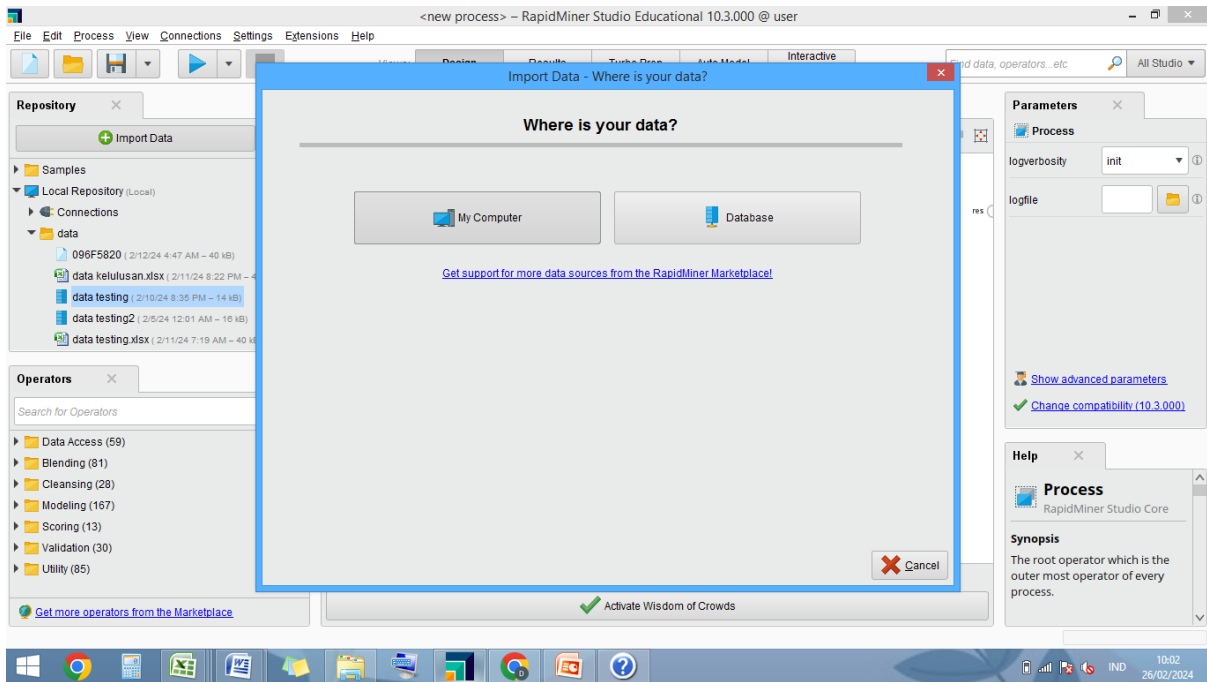
Kategori Kelulusan	Jumlah IPK
Memuaskan	2,76 – 3,00
Sangat Memuaskan	3,01 – 3,50
Pujian	> 3,50

4.3 Implementasi pada Software RapidMiner

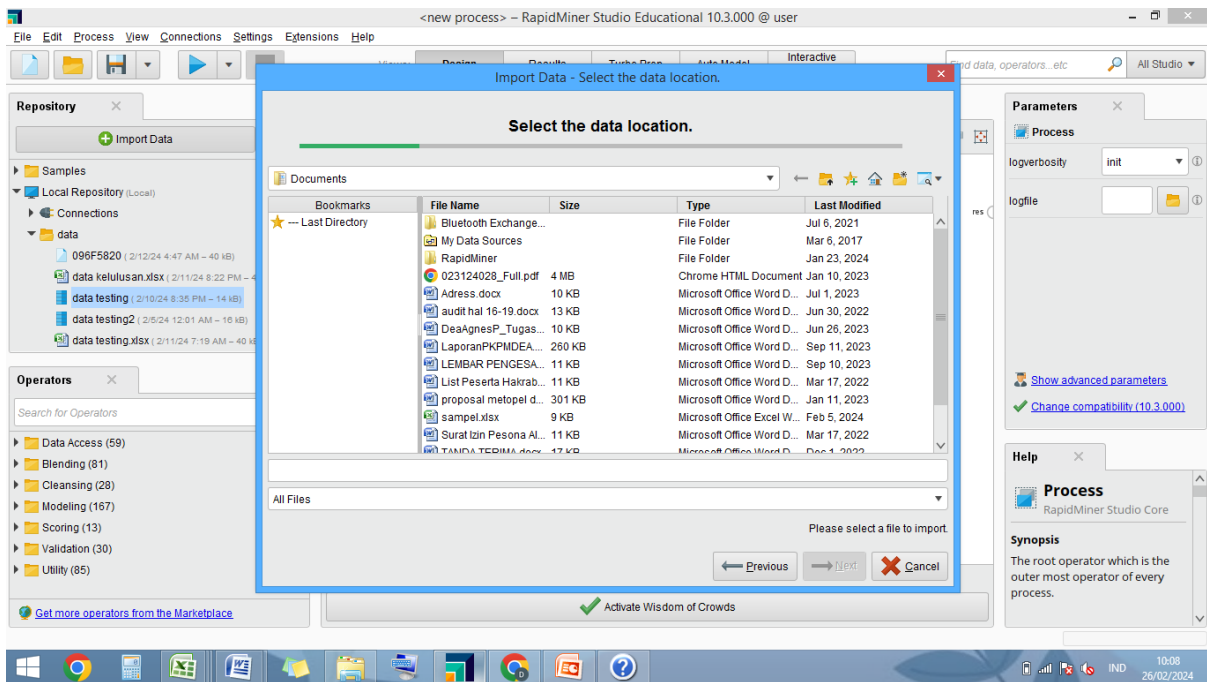
1. Tahap Import Data



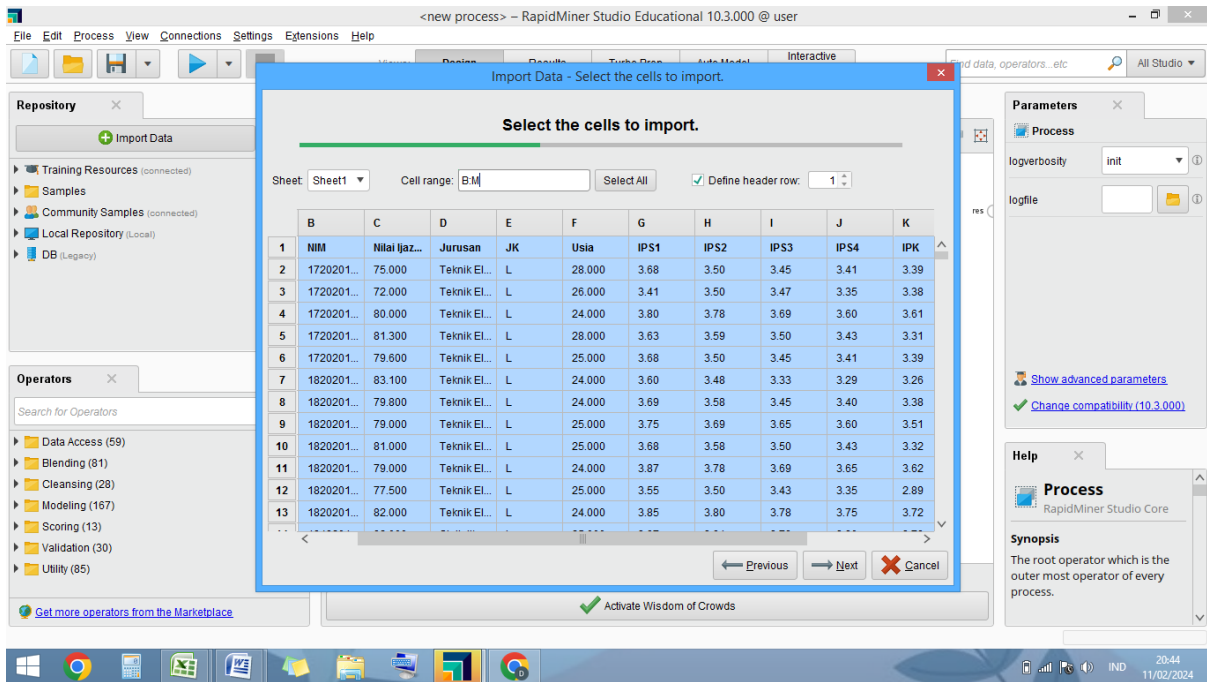
Gambar 4.1 Import Data



Gambar 4.2 Penyimpanan Data

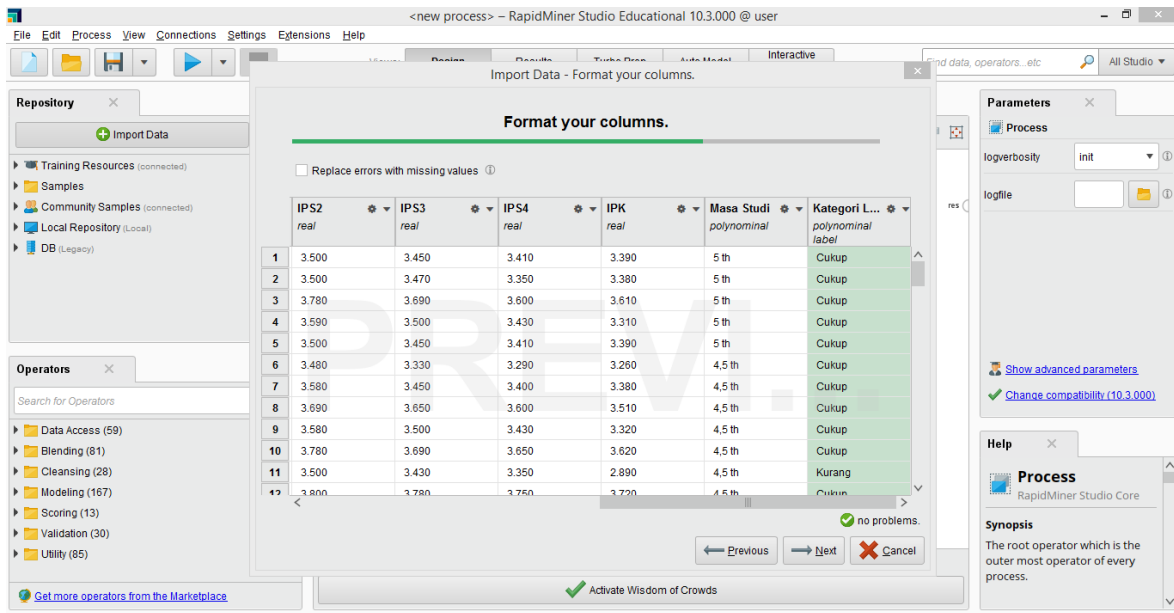


Gambar 4.3 Select Data Location



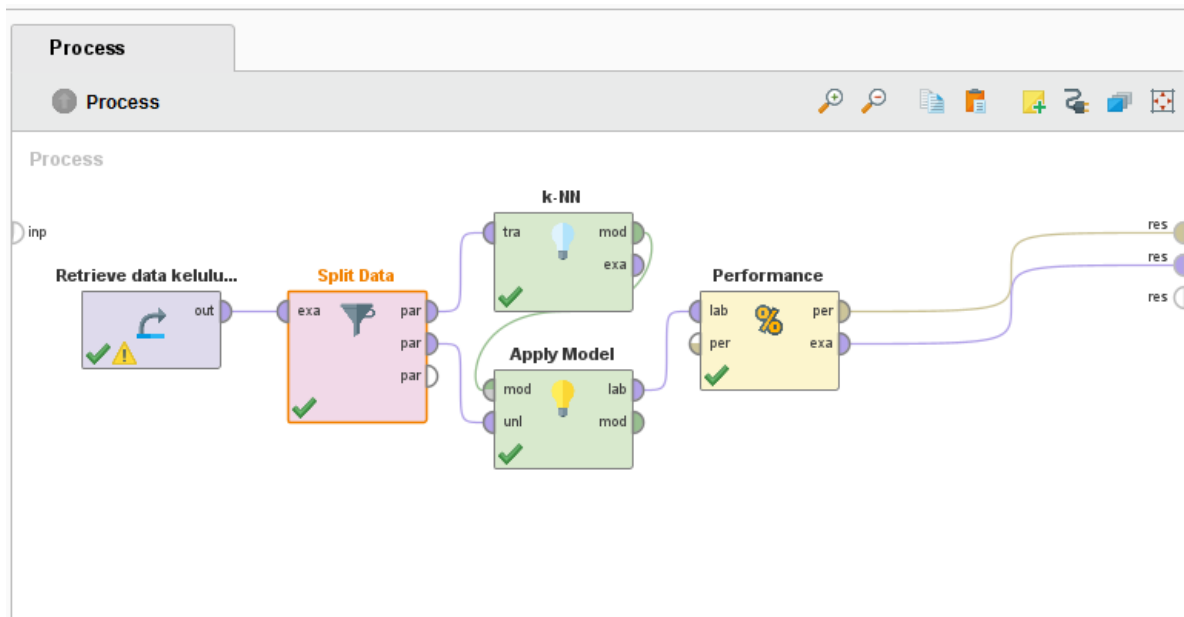
Gambar 4.4 Tampilan data yang sudah di Import

Setelah data di import kemudian atribut “Kategori Kelulusan” diberikan peran/role sebagai “Label”



Gambar 4.5 Penamaan Atribut dan Label

4.4 Desain Perhitungan Algoritma K-NN



Gambar 4.6 Desain Perhitungan K-NN

4.5 Hasil Perhitungan K-Nearest Neighbors

Dari 425 yang diinput dan dibagi menjadi 70% data training dan 30% data testing, diproseslah data testing sejumlah 127. Perhitungan menggunakan algoritma K-Nearest Neighbors menghasilkan tingkat akurasi sebesar 87,40%.

ExampleSet (Apply Model) PerformanceVector (Performance)

Table View Plot View

accuracy: 87.40%

	true Cukup	true Kurang	true Memuaskan	class precision
pred. Cukup	54	3	1	93.10%
pred. Kurang	1	4	0	80.00%
pred. Memuaskan	11	0	53	82.81%
class recall	81.82%	57.14%	98.15%	

Gambar 4.7 Hasil Akurasi Algoritma K-NN

- Predikat Memuaskan

$$\text{Sensitivitas (Precision)} = \frac{TP}{TP+FN} * 100\% = \frac{54}{54+3+1} * 100\% = 93,10\%$$

$$\text{Spesifitas (Recall)} = \frac{TN}{TN+FP} * 100\% = \frac{54}{54+1+11} * 100\% = 81,82\%$$

- Predikat Sangat Memuaskan

$$\text{Sensitivitas (Precision)} = \frac{TP}{TP+FN} * 100\% = \frac{4}{4+1} * 100\% = 80,00\%$$

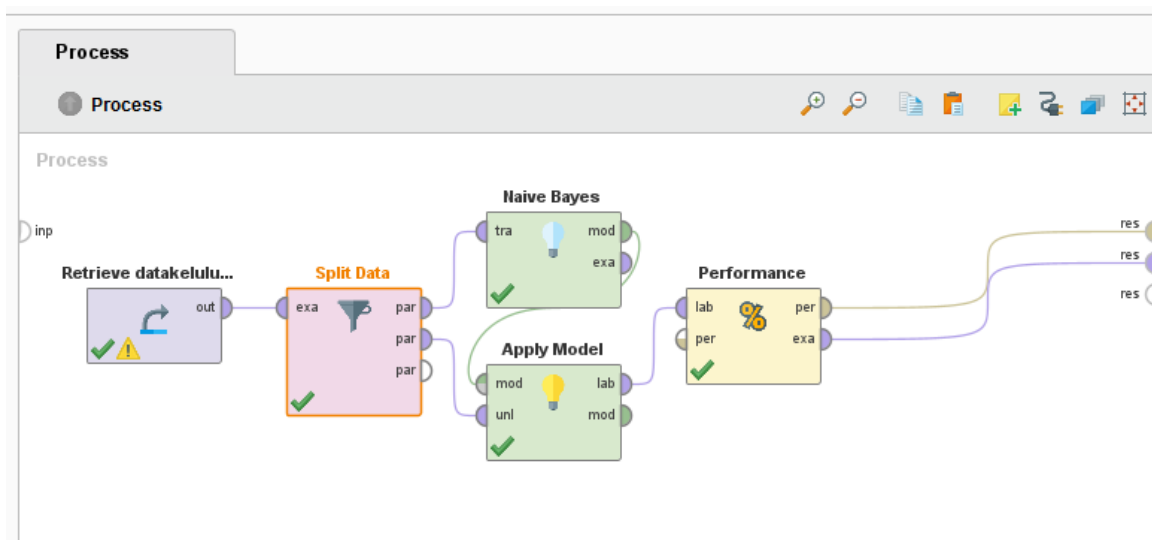
$$\text{Spesifitas (Recall)} = \frac{TN}{TN+FP} * 100\% = \frac{4}{4+3} * 100\% = 57,14\%$$

- Predikat Pujian

$$\text{Sensitivitas (Precision)} = \frac{TP}{TP+FN} * 100\% = \frac{53}{53+11} * 100\% = 82,81\%$$

$$\text{Spesifitas (Recall)} = \frac{TN}{TN+FP} * 100\% = \frac{53}{53+1} * 100\% = 98,15\%$$

4.6 Perhitungan Akurasi Algoritma Naive Bayes



Gambar 4.8 Desain Perhitungan Akurasi Naive Bayes

4.7 Hasil Perhitungan Akurasi Naive Bayes

Dari 425 yang diinput dan dibagi menjadi 70% data training dan 30% data testing, diproseslah data testing sejumlah 127. Perhitungan menggunakan algoritma Naive Bayes menghasilkan tingkat akurasi sebesar 99,21%.

	true Cukup	true Kurang	true Memuaskan	class precision
pred. Cukup	66	1	0	98.51%
pred. Kurang	0	6	0	100.00%
pred. Memuaskan	0	0	54	100.00%
class recall	100.00%	85.71%	100.00%	

Gambar 4.9 Hasil Akurasi Algoritma Naive Bayes

- Predikat Memuaskan

$$\text{Sensitivitas (Precision)} = \frac{TP}{TP+FN} * 100\% = \frac{66}{66+1} * 100\% = 98,51\%$$

$$\text{Spesifitas (Recall)} = \frac{TN}{TN+FP} * 100\% = \frac{66}{66+0} * 100\% = 100\%$$

- Predikat Sangat Memuaskan

$$\text{Sensitivitas (Precision)} = \frac{TP}{TP+FN} * 100\% = \frac{6}{6+0} * 100\% = 100,00\%$$

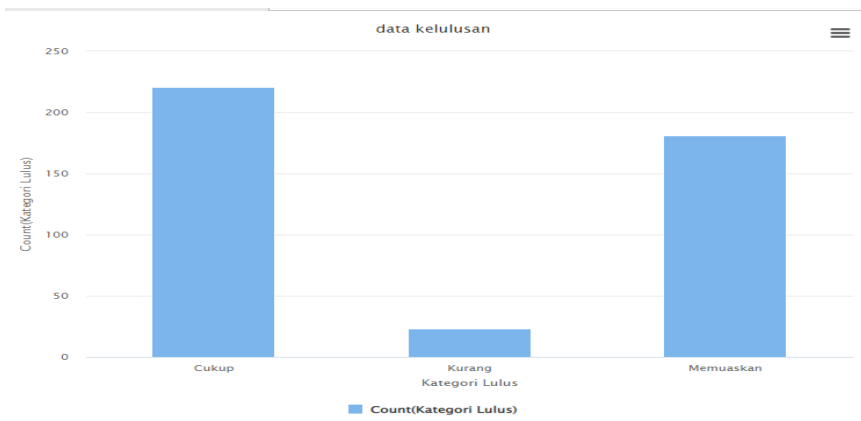
$$\text{Spesifitas (Recall)} = \frac{TN}{TN+FP} * 100\% = \frac{6}{6+1} * 100\% = 85,71\%$$

- Predikat Pujian

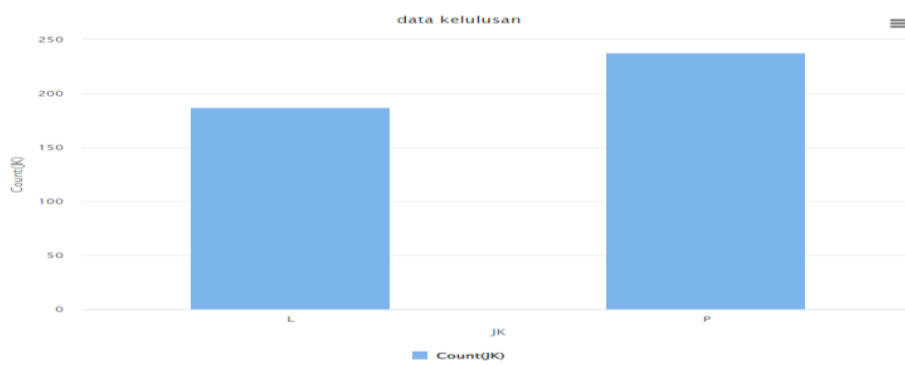
$$\text{Sensitivitas (Precision)} = \frac{TP}{TP+FN} * 100\% = \frac{54}{54+0} * 100\% = 100,00\%$$

$$\text{Spesifitas (Recall)} = \frac{TN}{TN+FP} * 100\% = \frac{54}{54+0} * 100\% = 100,00\%$$

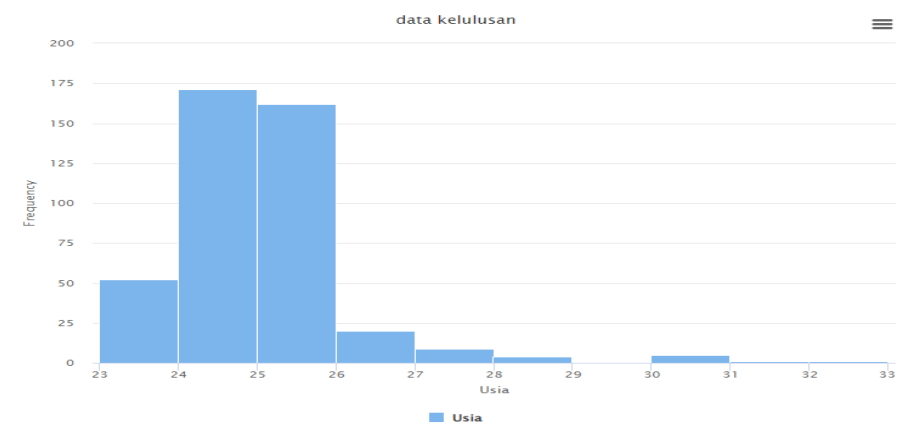
4.8 Visualisasi Menggunakan RapidMiner



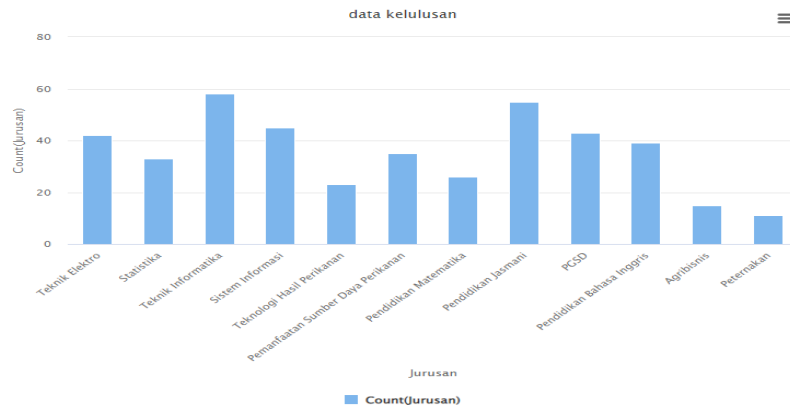
Gambar 4.10 Visualisasi Kategori Kelulusan



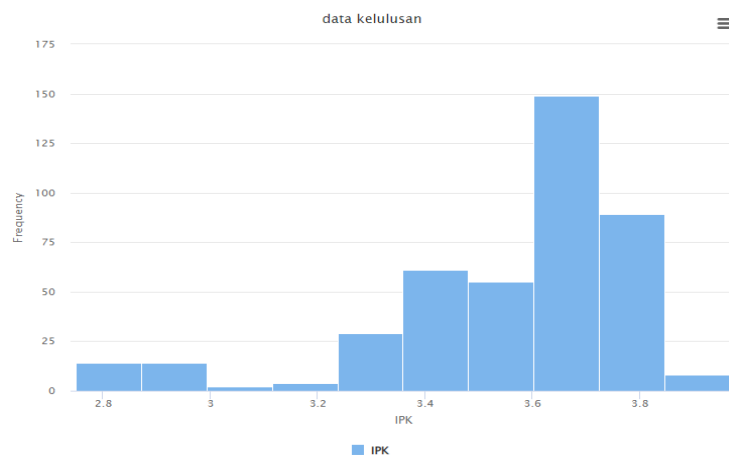
Gambar 4.11 Visualisasi Jenis Kelamin



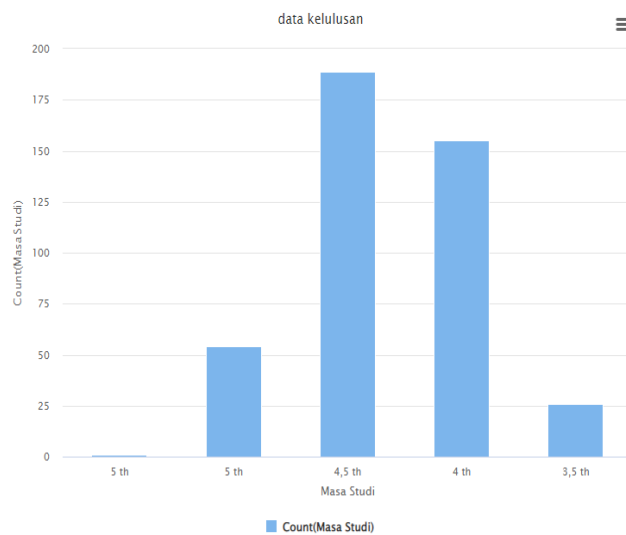
Gambar 4.12 Visualisasi Usia



Gambar 4.13 Visualisasi Jurusan



Gambar 4.14 Visualisasi IPK



Gambar 4.15 Visualisasi Masa Studi

4.9 Pembahasan

Dataset diolah menggunakan teknik klasifikasi Naive Bayes dan K-nearest neighbour. Pada outputnya ditemukan pola-pola baru pada data yang menghubungkan pola data yang ada dengan data baru, sehingga jenis pengujian yang digunakan adalah training set.

Proses klasifikasi dipengaruhi oleh atribut-atribut terpilih yang membantu menentukan kelompok memuaskan, adil, dan miskin. Dari 127 data uji, hasil klasifikasinya adalah 64 data berada pada kelas Pujian, 58 data berada pada kelas Sangat Memuaskan, dan 5 data berada pada kelas Memuaskan.

Nilai Presentase keakuratan menunjukkan keefektifan dataset Kelulusan Mahasiswa yang diterapkan ke dalam metode Naive Bayes mencapai 99,21% dan metode K-NN mencapai 84,70%.