

BAB IV

HASIL DAN PEMBAHASAN

Dalam bab ini penulis akan membahas mengenai hasil dari penelitian, data tersebut akan dihitung menggunakan feature selection dengan menggunakan algoritma Decision Tree C4.5 dan Algoritma Naïve Bayes kemudian akan diuji menggunakan Split Validation menggunakan algoritma Decision Tree C. 45 dan Algoritma Naïve Bayes Dataset dapat dilihat pada gambar berikut:

4.1 Pengumpulan Data

Data akan diseleksi untuk menentukan variabel apa saja yang akan diambil agar tidak terjadi kesamaan dan perulangan yang tidak diperlukan dalam pengolahan teknik data mining, dalam penanganan kasus yang terjadi pada dataset dalam penseleksian data dapat dilakukan dengan cara mengimport data ke dalam program rapidminer 9.10, data tersebut tidak akan dapat diolah karena masih terdapat beberapa error data dan program akan secara otomatis memberi tahu bahwa dataset perlu dilakukan preprocessing. Setelah data error tersebut ditemukan kita dapat menyeleksi dengan melakukan filter untuk menghapus data tersebut sehingga data dapat diproses kedalam system.

Table 4. 1 Dataset Penerima Bedah Rumah

Jenis Dinding	Jenis Lantai	Bahan Bakar Masak	fasilitas MCK	Tanggungan	keterangan
semen	semen	gas	ada	Kurang dari lima	tidak
semen	semen	gas	ada	Kurang dari lima	tidak
semen	semen	gas	ada	Kurang dari lima	tidak
semen	semen	gas	ada	Kurang dari lima	tidak
kayu	semen	kayu	tidak ada	Kurang dari lima	ya
semen	semen	gas	ada	Kurang dari lima	tidak
kayu	semen	kayu	tidak ada	Lebih dari Lima	ya
kayu	semen	kayu	tidak ada	Kurang dari lima	ya
semen	semen	gas	ada	Kurang dari lima	tidak
semen	semen	gas	ada	Kurang dari lima	tidak
semen	semen	gas	ada	Lebih dari Lima	tidak
semen	semen	gas	ada	Kurang dari lima	tidak
semen	semen	gas	ada	Kurang dari lima	tidak
semen	semen	gas	ada	Kurang dari lima	tidak
semen	semen	gas	ada	Kurang dari lima	tidak
semen	semen	gas	ada	Lebih dari Lima	tidak

Table 4. 2 Hasil perhitungan entropi dan gain

	Jumlah (S)	A (Si)	B (Si)	Entrophy	Gain
1.1 Fasilitas MCK Tidak Ada	118	117	1	0,07050176	
					0,706
KAYU	105	104	1	0,077619511	
SEMEN	13	13	0	0	
					0,708
SEMEN	74	74	0	0	
TANAH	44	43	1	0,156491063	
					0,706
GAS	6	6	0	0	
KAYU	112	111	1	0,073603483	
					0,706
KURANG DARI LIMA	104	103	1	0,078232465	
LEBIH DARI LIMA	14	14	0	0	

Table 4. 3 Hasil perhitungan entropi dan gain

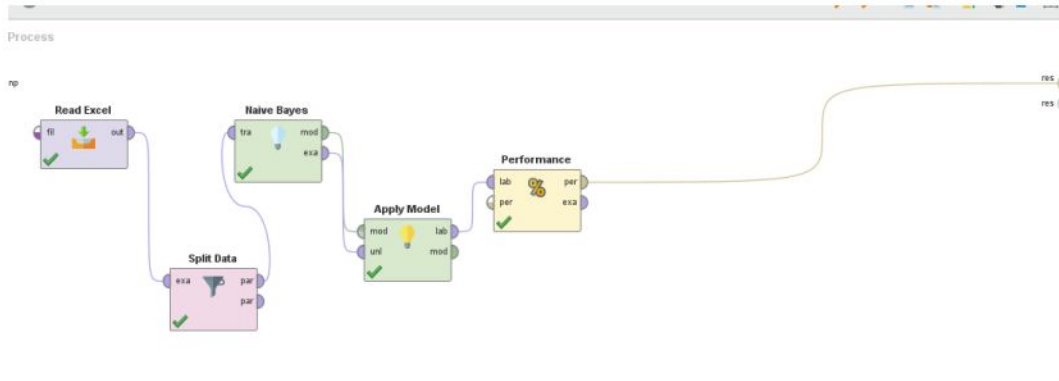
		Jumlah (S)	A (Si)	B (Si)	Entrophy	Gain
Total	1.1 Fasilitas MCK Tidak Ada	118	117	1	0,07050176	
Jenis Dinding						0,706
	KAYU	105	104	1	0,077619511	
	SEMEN	13	13	0	0	
Jenis Lantai						0,708

	SEMEN	74	74	0	0	
	TANAH	44	43	1	0,156491063	
Bahan Bakar Masak						0,706
	GAS	6	6	0	0	
	KAYU	112	111	1	0,073603483	
Jumlah Tanggungan						0,706
	KURANG DARI LIMA	104	103	1	0,078232465	
	LEBIH DARI LIMA	14	14	0	0	

Table 4. 4 Hasil perhitungan entropi dan gain

		Jumlah (S)	A (Si)	B (Si)	Entrophy	Gain
Total	2.1 Jenis Lantai Tanah	44	43	1	0,156491063	
Jenis Dinding						0,708
	KAYU	44	43	1	0,156491063	
	SEMEN	0	0	0	0	
Bahan Bakar Masak						0,708
	GAS	0	0	0	0	
	KAYU	44	43	1	0,156491063	
Jumlah Tanggungan						0,708
	KURANG DARI LIMA	43	42	1	0,159350063	
	LEBIH DARI LIMA	1	1	0	0	

4.2 Menggunakan Algoritma Naïve Bayes Pengujian Klasifikasi dengan Algoritma Naive Bayes



Gambar 4. 1 Proses menggunakan Algoritma Naïve Bayes Split Validation

Hasil dari proses *dataset* bedah rumah didapatkan nilai *Accuracy* 82,2 ,16%
Seperti gambar dibawah ini

accuracy: 82.21%

	true tidak	true ya	class precision
pred. tidak	314	56	84.86%
pred. ya	18	28	60.87%
class recall	94.58%	33.33%	

Gambar 4. 2 Nilai accuracy Algoritma Naïve Bayes. Split Validation

$$\begin{aligned}
 \text{akurasi} &= \frac{(TN + TP)}{(TN + FN + FP + TP)} \\
 &= \frac{(28 + 314)}{(28 + 18 + 56 + 314)} \\
 &= \frac{342}{416} = 0,8221
 \end{aligned}$$

$$\text{akurasi} = 82,2\%$$

Hasil dari proses *dataset* bedah rumah didapatkan nilai *Precision* 84 ,66%
Seperti gambar dibawah ini

precision: 60.87% (positive class: ya)

	true tidak	true ya	class precision
pred. tidak	314	56	84.86%
pred. ya	18	28	60.87%
class recall	94.58%	33.33%	

Gambar 4. 3 Nilai Precision Algoritma Naïve Bayes. Split Validation

$$\begin{aligned}
 precision &= \frac{TP}{(TP + FP)} \\
 &= \frac{314}{(314 + 56)} \\
 &= \frac{314}{370} = 0,84864 \\
 \mathbf{precision} &= \mathbf{84,86\%}
 \end{aligned}$$

Hasil dari proses *dataset* bedah rumah didapatkan nilai *Reccal* 94,58% Seperti gambar dibawah ini

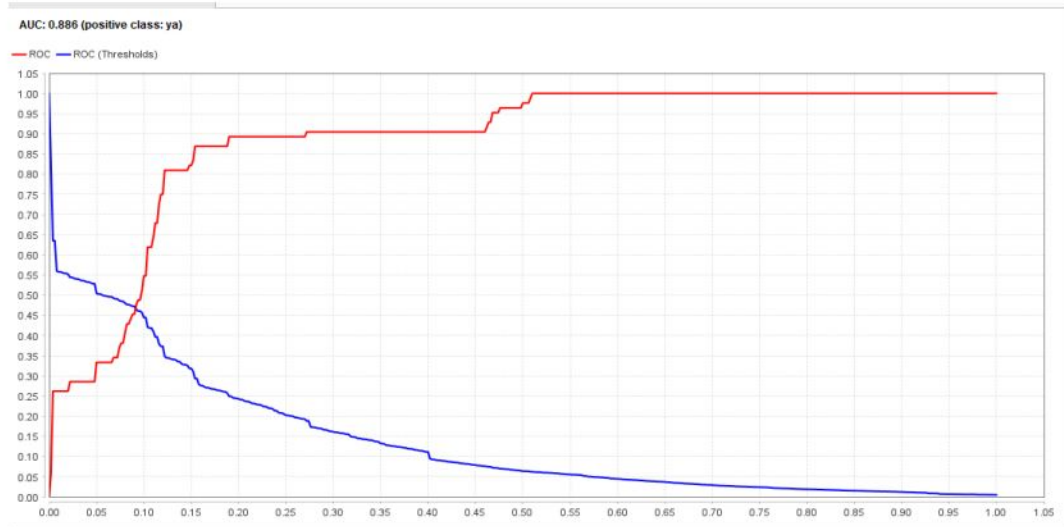
recall: 33.33% (positive class: ya)

	true tidak	true ya	class precision
pred. tidak	314	56	84.86%
pred. ya	18	28	60.87%
class recall	94.58%	33.33%	

Gambar 4. 4 Confusion Matrix Hasil klasifikasi Algoritma Naïve Bayes. Split Validation

$$\begin{aligned}
 precision &= \frac{TP}{(TP + FP)} \\
 &= \frac{314}{(314 + 18)} \\
 &= \frac{314}{332} = 0,9457 \\
 \mathbf{precision} &= \mathbf{94,58\%}
 \end{aligned}$$

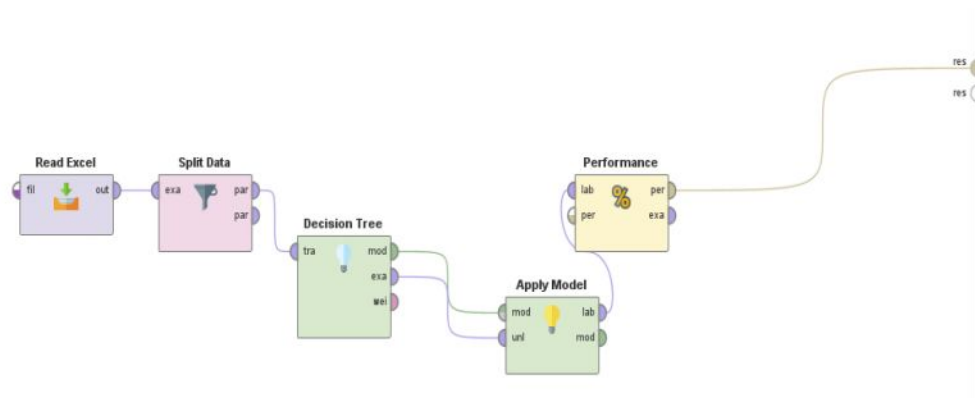
Berikut adalah hasil kurva AUC yaitu mendapatkan nilai 0,886 seperti gambar dibawah ini:



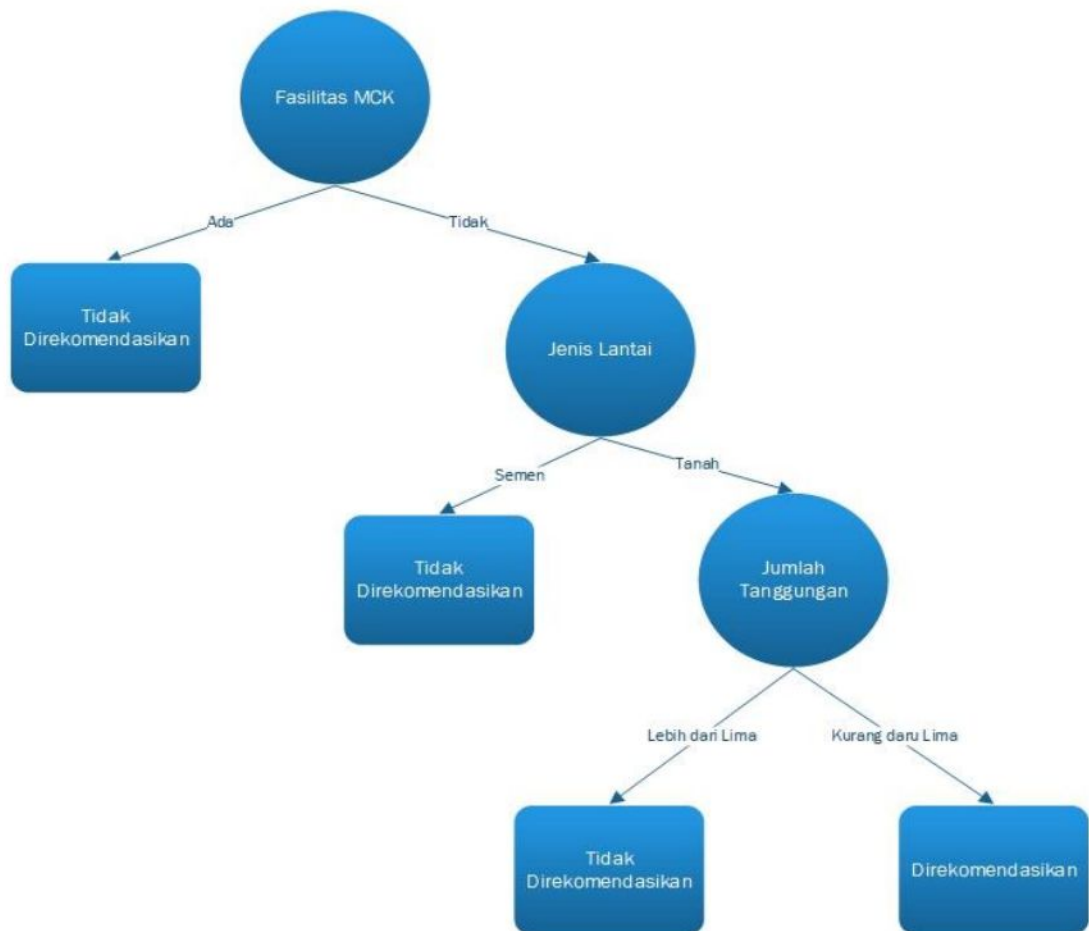
Gambar 4. 5 Nilai dan Curve AUC Algoritma Naïve Bayes Split Validation

4.3 Menggunakan Algoritma Decision Tree C.45

Pada algoritma *decision tree*, record yang sudah di import ke rapid miner akan digunakan untuk menentukan pola pohon keputusan, hasil dari pola pohon keputusan dapat dilihat pada gambar berikut.

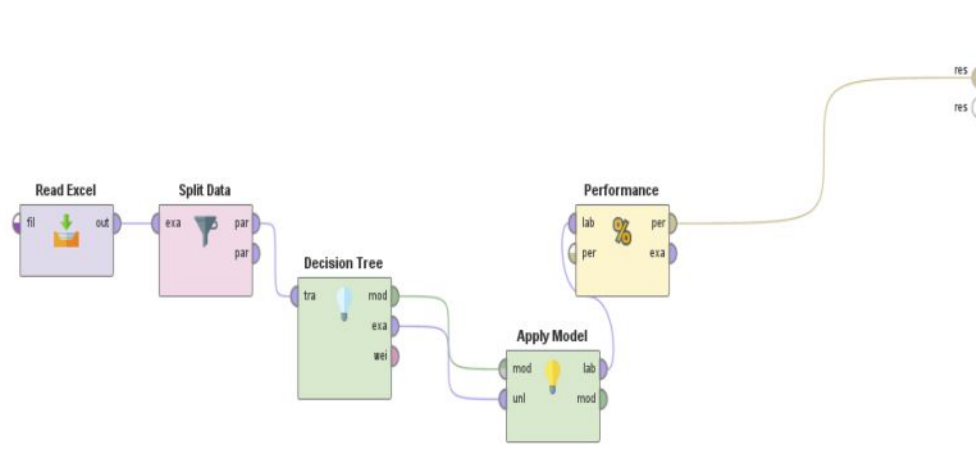


Gambar 4. 6 Proses import data ke rapidminer



Gambar 4. 7 Pola Pohon Keputusan

Penerapan data pada Rapidminer untuk klasifikasi dengan menggunakan split validation dengan nilai akurasi, precision, confusion matrix atau nilai recall dan nilai AUC dengan pembagian Data training dan testing sebesar 80:20. dapat dilihat pada gambar berikut



Gambar 4. 8 Proses menggunakan Algoritma Decision Tree C. 45 Split Validation

Pada gambar 4.8 data yang sudah disiapkan diterapkan pada aplikasi rapidminer klasifikasi dengan melakukan eksperimen menggunakan split data yang dapat membagi data menjadi data training dan data testing karena data yang digunakan adalah data sekunder dan algoritma yang digunakan adalah decision tree. Hasil eksperimen dapat kita lihat pada gambar 4.5

Hasil dari proses *dataset* bedah rumah didapatkan nilai *Accuracy* 96,88% Seperti gambar dibawah ini

accuracy: 96.88%

	true tidak	true ya	class precision
pred. tidak	320	1	99.69%
pred. ya	12	83	87.37%
class recall	96.39%	98.81%	

Gambar 4. 9 Nilai accuracy Algoritma Decision Tree C. 45 Split Validation

Menghitung akurasi:

Rumus:

$$\text{Akurasi} = \frac{(\text{TN} + \text{TP})}{(\text{TN} + \text{FN} + \text{FP} + \text{TP})}$$

$$\text{Akurasi} = \frac{(320+83)}{(320+83+1+12)} = \frac{403}{416} = 96,88\%$$

Hasil dari proses *dataset* bedah rumah didapatkan nilai *Precision* Seperti gambar dibawah ini

precision: 87.37% (positive class: ya)

	true tidak	true ya	class precision
pred. tidak	320	1	99.69%
pred. ya	12	83	87.37%
class recall	96.39%	98.81%	

Gambar 4. 10 Precision Algoritma Decision Tree C. 45 Split Validation

Menghitung Precision, Rumus

$$P:\text{precision} = \frac{\text{TP}}{(\text{TP}+\text{FP})}$$

$$:p(1) = \frac{320}{(320+1)} = \frac{320}{321} = 99,69\% \quad :p(0) = \frac{83}{(83+12)} = \frac{83}{95} = 87,37\%$$

Hasil dari proses *dataset* bedah rumah didapatkan nilai *Recall* Seperti gambar dibawah ini

recall: 98.81% (positive class: ya)

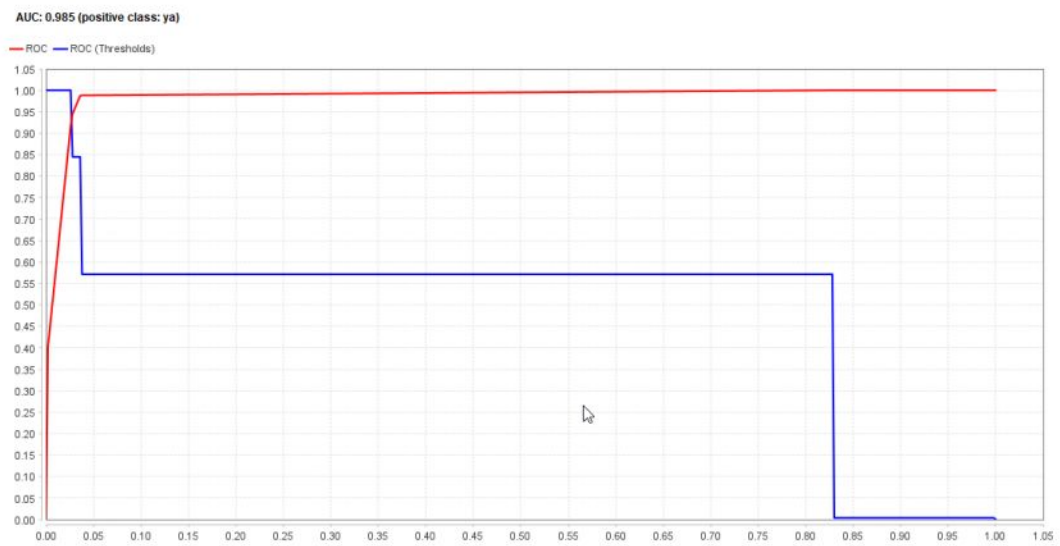
	true tidak	true ya	class precision
pred. tidak	320	1	99.69%
pred. ya	12	83	87.37%
class recall	96.39%	98.81%	

Gambar 4. 11 Confusion Matrix Hasil klasifikasi Bedah Rumah menggunakan

Menghitung Recall, Rumus: $Recall = \frac{TP}{(TP+FN)}$

$$R(1) = \frac{320}{(320+12)} = \frac{320}{332} = 96,39\% \quad R(0) = \frac{83}{(83+1)} = \frac{83}{84} = 98,81\%$$

Berikut adalah hasil kurva AUC yaitu mendapatkan nilai 0,985 seperti gambar dibawah ini



Gambar 4. 12 Nilai dan Curve AUC Algoritma Decision Tree C. 45 Split Validation

4.4 Pengujian Precision, Recall dan Accuracy Naive Bayes dan C.45

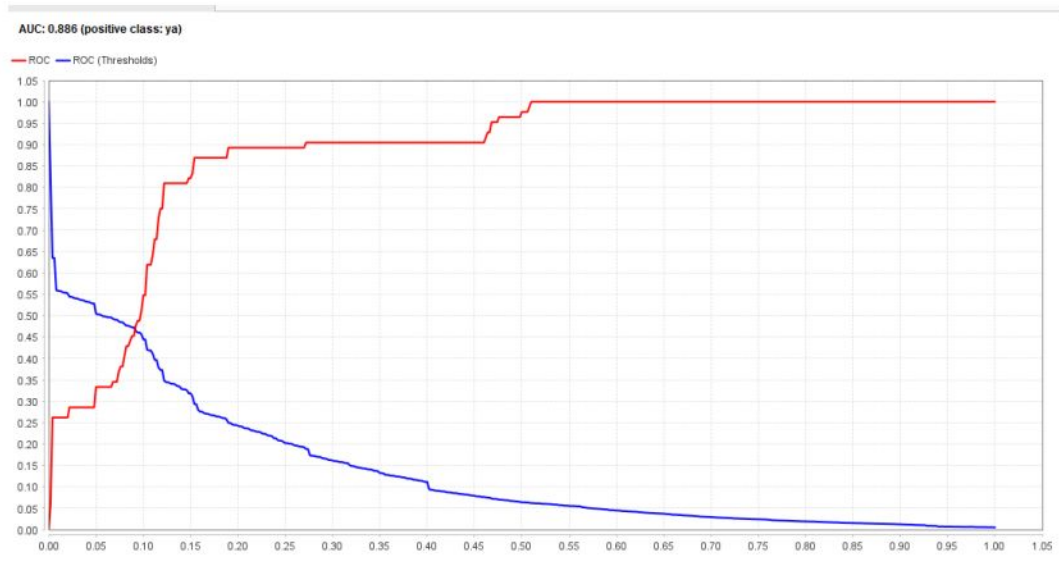
Hasil pengujian precision, recall dan accuracy pada kasus penentuan penerima bantuan rumah. Dari data training membuat nilai precision, recall dan accuracy dari algoritma Naive Bayes dan C.45 semakin meningkat. Dari data tersebut, dapat pula dilihat bahwa nilai precision, recall dan accuracy dari algoritma C.45 lebih tinggi dari Naive Bayes di setiap pembagian data. Hal ini menunjukkan bahwa algoritma C.45 lebih baik daripada algoritma Naive Bayes untuk kasus penerima bantuan bedah rumah.

4.5 Evaluasi Model Confusion Matrix, F-measure dan dengan ROC Naïve Bayes

Pada tahap ini akan membahas tentang hasil evaluasi dari eksperimen yang telah dilakukan. Model yang terbentuk akan diuji dengan menggunakan Confusion Matrix untuk mengetahui tingkat akurasi. Confusion Matrix akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi negatif yang benar dan prediksi negatif yang salah. Akurasi akan dihitung dari seluruh prediksi yang benar (baik prediksi positif dan negatif). Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan

Table 4. 5 Nilai Perhitungan Niave Byaes

<i>Naïve Bayes</i>	Nilai (%)
<i>Accuracy</i>	82,2
<i>Sensitivity</i>	84,86
<i>Specitivity</i>	60.87
<i>PPV</i>	94,58
<i>NPV</i>	33,33



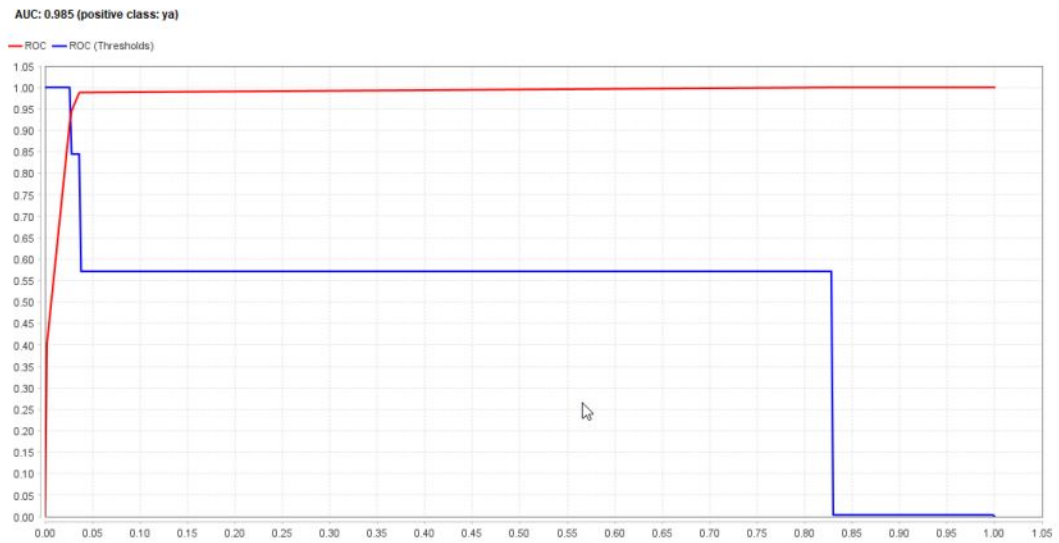
Gambar 4. 13 Nilai dan Curve AUC Algoritma Naïve Bayes Split Validation

4.6 Evaluasi Model Confusion Matrix, F-measure dan Kurva ROC Algoritma C4.5

Pada tahap ini akan membahas tentang hasil evaluasi dari eksperimen yang telah dilakukan. Model yang terbentuk akan diuji dengan menggunakan Confusion Matrix untuk mengetahui tingkat akurasi. Confusion Matrix akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi negatif yang benar dan prediksi negatif yang salah. Akurasi akan dihitung dari seluruh prediksi yang benar (baik prediksi positif dan negatif). Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan.

Table 4. 6 Nilai perhitungan algoritma C 4.5

	Nilai (%)
<i>Accuracy</i>	96,88
<i>Sensitivity</i>	99,69
<i>Specitivity</i>	96,39
<i>PPV</i>	87,37
<i>NPV</i>	98,81



Gambar 4. 14 Nilai AUC dalam grafik ROC algoritma C4.5

4.7 Hasil Perbandingan

Model dengan algoritma *Decision Tree* dan *Naïve Bayes* untuk prediksi kelayakan pemberian bantuan bedah rumah yang diuji tingkat akurasi menghasilkan perbandingan nilai akurasi (*accuracy*) dan nilai AUC. Dengan dataset penerima bantuan bedah rumah sebagai data uji. Algoritma *Decision Tree* mendapatkan akurasi yang paling tinggi yaitu sebesar 98,88%. Sedangkan algoritma *Naïve Bayes* mendapatkan akurasi sebesar 82,2%. Seperti yang ditunjukkan pada tabel 4.4

Table 4. 7 Perbandingan Akurasi

No	Algoritma	Akurasi	AUC
1	C4.5	96,88%	0,985
2	Naïve Bayes	82,2%	0,886