

BAB IV

HASIL DAN PEMBAHASAN

4.1 Persiapan Data

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 Nive bayes yang kemudian akan diuji menggunakan Split Validation menggunakan algoritma Decision Tree C.45 dan Algoritma Nive bayes Dataset yang digunakan merupakan dataset publik yang berasal dari <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>

dataset publik heart disease yang di peroleh dari Dataset dapat dilihat pada tableberikut:

Tabel 4.1 Dataset Gagal Jantung

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
40	M	ATA	140	289	0	Normal	172	N	0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1	Flat	1
37	M	ATA	130	283	0	ST	98	N	0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0	Up	0
39	M	NAP	120	339	0	Normal	170	N	0	Up	0
45	F	ATA	130	237	0	Normal	170	N	0	Up	0
54	M	ATA	110	208	0	Normal	142	N	0	Up	0
37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
48	F	ATA	120	284	0	Normal	120	N	0	Up	0
37	F	NAP	130	211	0	Normal	142	N	0	Up	0
58	M	ATA	136	164	0	ST	99	Y	2	Flat	1
39	M	ATA	120	204	0	Normal	145	N	0	Up	0
49	M	ASY	140	234	0	Normal	140	Y	1	Flat	1
42	F	NAP	115	211	0	ST	137	N	0	Up	0
54	F	ATA	120	273	0	Normal	150	N	1.5	Flat	0
38	M	ASY	110	196	0	Normal	166	N	0	Flat	1
43	F	ATA	120	201	0	Normal	165	N	0	Up	0
60	M	ASY	100	248	0	Normal	125	N	1	Flat	1

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 publi, dalam penseleksian data dapat dilakukan dengan cara mengimport data ke dalam program rapidminer 10.1, 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 sistem

Age	Sex	ChestPainT...	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAn...	Oldpeak	ST_Slope
40	M	ATA	140	289	0	Normal	172	N	0	Up
49	F	NAP	160	180	0	Normal	156	N	1	Flat
37	M	ATA	130	283	0	ST	98	N	0	Up
48	F	ASY	138	214	0	Normal	108	Y	1.500	Flat
54	M	NAP	150	195	0	Normal	122	N	0	Up
39	M	NAP	120	339	0	Normal	170	N	0	Up
45	F	ATA	130	237	0	Normal	170	N	0	Up
54	M	ATA	110	208	0	Normal	142	N	0	Up
37	M	ASY	140	207	0	Normal	130	Y	1.500	Flat
48	F	ATA	120	284	0	Normal	120	N	0	Up
37	F	NAP	130	211	0	Normal	142	N	0	Up
58	M	ATA	136	164	0	ST	99	Y	2	Flat
39	M	ATA	120	204	0	Normal	145	N	0	Up

ExampleSet (918 examples, 1 special attribute, 11 regular attributes)

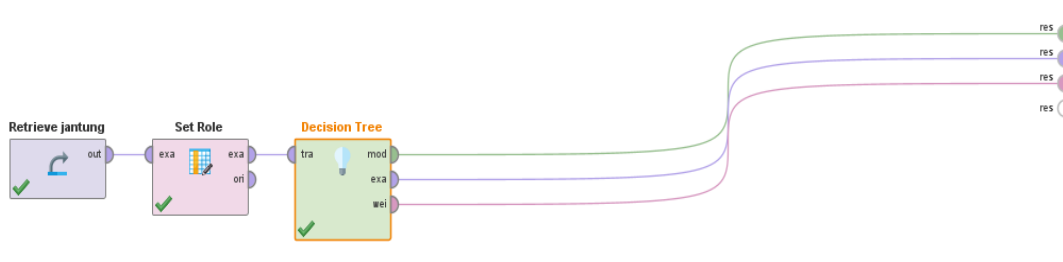
Gambar 4.2 Dataset Gagal Jantung

4.2 Hasil Pengujian

Proses data mining dilakukan dengan bantuan perangkat lunak datamining yaitu RapidMiner versi 10.1 Metode yang digunakan yaitu *feature selection* dengan menggunakan *Algoritma Decision Tree C4.5* dan *Algoritma Naïve Bayes*. Hasil dari data keseluruhan 918 record dari dataset gagal jantung dengan 12 Atribut dan 1 atribut sebagai Class.

1. Penelitian Menggunakan Algoritma Decision Tree C4.5

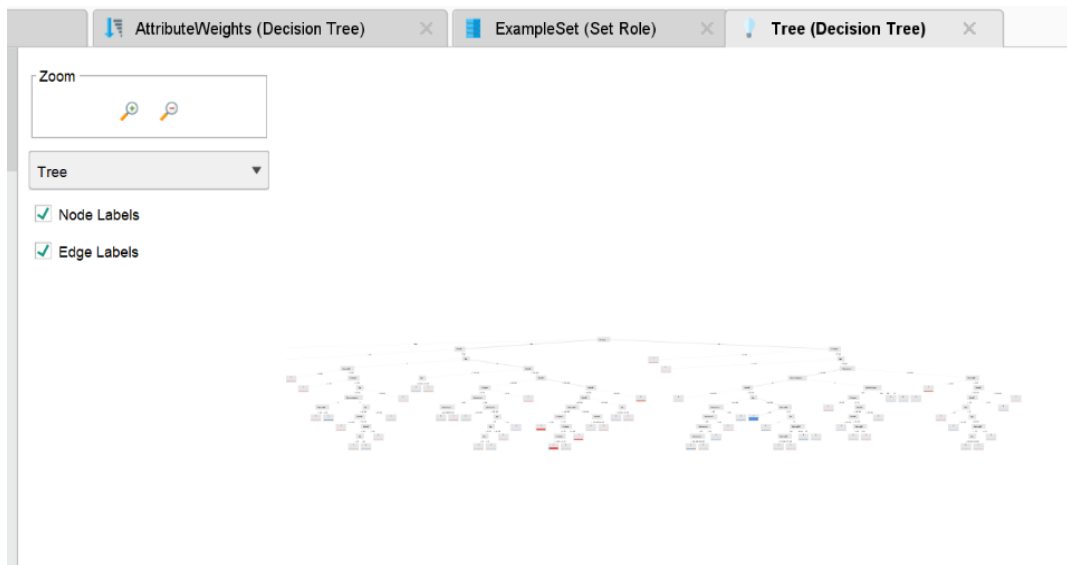
Pada algoritma *decision tree*, record yang sudah di import ke rapid miner akan digunakan untuk menentukan pola pohon keputusan, hasil dari pola pohon keputusan dapat dilihat pada gambar berikut.



Gambar 4.3 Proses penghitungan variable menggunakan *rapidminer*

attribute	weight
ST_Slope	0.012
RestingBP	0.163
Resting...	0.011
MaxHR	0.193
Exercise...	0.025
FastingBS	0.034
Sex	0.012
Oldpeak	0.074
ChestPa...	0.015
Age	0.370
Choleste...	0.091

Gambar 4.4 Atribut Weigh



Gambar 4.5 Pohon Keputusan

Bentuk *Tree* dari perhitungan *Decision Tree*

ST_Slope = Down

```
|   MaxHR > 184.500: 0 {0=2, 1=0}
|
|   MaxHR ≤ 184.500
|
|   |   ExerciseAngina = N
|   |   |   RestingBP > 117
|   |   |   |   Age > 66.500: 1 {0=0, 1=2}
|   |   |   |   Age ≤ 66.500
|   |   |   |   |   Age > 60
|   |   |   |   |   |   Age > 62.500: 0 {0=3, 1=0}
|   |   |   |   |   |   Age ≤ 62.500: 1 {0=0, 1=3}
|   |   |   |   |   |   Age ≤ 60: 0 {0=6, 1=0}
|   |   |   |   |   |   RestingBP ≤ 117: 1 {0=0, 1=6}
|   |   |   |   |   ExerciseAngina = Y: 1 {0=3, 1=38}
```

ST_Slope = Flat

```
|   MaxHR > 183: 0 {0=2, 1=0}
|
|   MaxHR ≤ 183
|
|   |   Sex = F
|   |   |   FastingBS > 0.500: 1 {0=0, 1=10}
|   |   |   FastingBS ≤ 0.500
|   |   |   |   Oldpeak > 2.300: 1 {0=0, 1=3}
|   |   |   |   Oldpeak ≤ 2.300
|   |   |   |   |   Age > 38.500
|   |   |   |   |   |   ExerciseAngina = N
|   |   |   |   |   |   |   RestingBP > 148: 1 {0=0, 1=3}
|   |   |   |   |   |   |   RestingBP ≤ 148: 0 {0=28, 1=3}
|   |   |   |   |   |   |   ExerciseAngina = Y
|   |   |   |   |   |   |   Age > 46.500
```

```

| | | | | | | | | MaxHR > 142.500: 1 {0=0, 1=6}
| | | | | | | | | MaxHR ≤ 142.500
| | | | | | | | | MaxHR > 109
| | | | | | | | | Age > 62: 1 {0=0, 1=3}
| | | | | | | | | Age ≤ 62: 0 {0=6, 1=5}
| | | | | | | | | MaxHR ≤ 109: 1 {0=0, 1=4}
| | | | | | | | | Age ≤ 46.500: 0 {0=2, 1=0}
| | | | | | | | | Age ≤ 38.500: 1 {0=0, 1=2}
| | | Sex = M
| | | | MaxHR > 176.500
| | | | | Age > 40.500: 0 {0=3, 1=0}
| | | | | Age ≤ 40.500: 1 {0=0, 1=2}
| | | | MaxHR ≤ 176.500
| | | | | MaxHR > 150.500
| | | | | | Oldpeak > 0.300
| | | | | | | Cholesterol > 256
| | | | | | | | Cholesterol > 297.500: 0 {0=1, 1=1}
| | | | | | | | Cholesterol ≤ 297.500: 1 {0=1, 1=10}
| | | | | | | | Cholesterol ≤ 256
| | | | | | | | | Cholesterol > 233.500: 0 {0=4, 1=0}
| | | | | | | | | Cholesterol ≤ 233.500
| | | | | | | | | Age > 46
| | | | | | | | | Age > 55.500
| | | | | | | | | | Age > 57: 1 {0=2, 1=4}
| | | | | | | | | | Age ≤ 57: 0 {0=2, 1=0}
| | | | | | | | | | Age ≤ 55.500: 1 {0=0, 1=3}
| | | | | | | | | | Age ≤ 46: 0 {0=3, 1=0}
| | | | | | | | | | Oldpeak ≤ 0.300: 1 {0=0, 1=13}

```

```

| | | | MaxHR ≤ 150.500
| | | | | MaxHR > 95.500
| | | | | | MaxHR > 97.500
| | | | | | | RestingBP > 104.500
| | | | | | | | Oldpeak > 1.850: 1 {0=1, 1=85}
| | | | | | | | Oldpeak ≤ 1.850
| | | | | | | | | Oldpeak > 0.050
| | | | | | | | | | Oldpeak > 0.250: 1 {0=17,
1=117}
| | | | | | | | | | Oldpeak ≤ 0.250: 0 {0=2,
1=1}
| | | | | | | | | | Oldpeak ≤ 0.050: 1 {0=1, 1=62}
| | | | | | | | RestingBP ≤ 104.500
| | | | | | | | MaxHR > 138.500: 0 {0=2, 1=0}
| | | | | | | | MaxHR ≤ 138.500: 1 {0=0, 1=6}
| | | | | | | | MaxHR ≤ 97.500
| | | | | | | | Age > 55.500: 0 {0=2, 1=0}
| | | | | | | | Age ≤ 55.500: 1 {0=0, 1=3}
| | | | | | | | MaxHR ≤ 95.500: 1 {0=0, 1=35}
ST_Slope = Up
| Oldpeak > 2.350: 1 {0=0, 1=8}
| Oldpeak ≤ 2.350
| | Age > 75.500: 1 {0=0, 1=2}
| | Age ≤ 75.500
| | | Cholesterol > 42.500
| | | | ExerciseAngina = N
| | | | | MaxHR > 193.500: 0 {0=1, 1=1}
| | | | | MaxHR ≤ 193.500

```



```

| | | | | | | Oldpeak ≤ 0.700
| | | | | | | MaxHR > 128.500
| | | | | | | MaxHR > 162.500: 0 {0=3, 1=0}
| | | | | | | MaxHR ≤ 162.500
| | | | | | | RestingBP > 145: 0 {0=2, 1=0}
| | | | | | | RestingBP ≤ 145: 1 {0=0, 1=5}
| | | | | | | MaxHR ≤ 128.500: 0 {0=3, 1=0}
| | | | | | | ChestPainType = ATA: 0 {0=6, 1=0}
| | | | | | | ChestPainType = NAP: 0 {0=7, 1=0}
| | | | | | | ChestPainType = TA: 0 {0=2, 1=0}
| | | | Cholesterol ≤ 42.500
| | | | | FastingBS > 0.500: 1 {0=0, 1=20}
| | | | | FastingBS ≤ 0.500
| | | | | | MaxHR > 102.500
| | | | | | | MaxHR > 119
| | | | | | | Age > 67: 1 {0=0, 1=2}
| | | | | | | Age ≤ 67
| | | | | | | | Age > 58.500: 0 {0=6, 1=0}
| | | | | | | | Age ≤ 58.500
| | | | | | | | | RestingBP > 113
| | | | | | | | | | Age > 52.500: 0 {0=1, 1=1}
| | | | | | | | | | Age ≤ 52.500: 1 {0=0, 1=3}
| | | | | | | | | | RestingBP ≤ 113: 0 {0=2, 1=0}
| | | | | | | | | | MaxHR ≤ 119: 0 {0=4, 1=0}
| | | | | | | | | | MaxHR ≤ 102.500: 1 {0=0, 1=3}

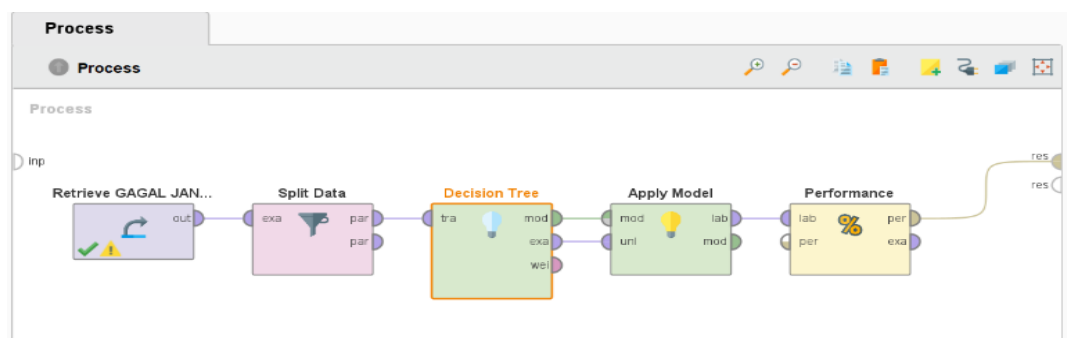
```

Mencari *entropy* dan *gain* untuk menentukan *node* root awal, diketahui dari proses diatas bahwa dengan menggunakan tools rapidminer node root awal adalah Disini kita akan melakukan pengukuran nilai *entropy* dan *gain* dengan cara menghitung

manual. Proses ini adalah untuk pembuktian apakah tools rapidminer berjan dengan baik.

2. Penelitian Menggunakan Algoritma Decision Tree C4.5 Split Validation

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 70:30. dapat dilihat pada gambar berikut:



Gambar 4.6 Proses menggunakan Algoritma C4.5 Split Validation

PerformanceVector (Performance)

Criterion: Table View Plot View

accuracy: 94.40%

	true 0	true 1	class precision
pred. 0	270	19	93.43%
pred. 1	17	337	95.20%
class recall	94.08%	94.66%	

Gambar 4.7 Nilai accuracy Algoritma Decision Tree C4.5 Split Validation

$$\text{Akurasi} = \frac{(TN+TP)}{(TN+FN+FP+TP)} \quad (4.1)$$

$$\text{Akurasi} = \frac{(270+337)}{(270+337+17)} = \frac{607}{624} = 94,40\% \quad (4.2)$$

precision: 95.20% (positive class: 1)			
	true 0	true 1	class precision
pred. 0	270	19	93.43%
pred. 1	17	337	95.20%
class recall	94.08%	94.66%	

Gambar 4.8. Precision Algoritma Decision Tree C4.5 Split Validation

$$precision = \frac{TP}{(TP+FP)} \quad (4.3)$$

$$:p(1) = \frac{270}{(270+19)} = \frac{270}{289} = 93,43\%$$

$$:p(0) = \frac{337}{(337+17)} = \frac{337}{354} = 95,20\%$$

recall: 94.66% (positive class: 1)			
	true 0	true 1	class precision
pred. 0	270	19	93.43%
pred. 1	17	337	95.20%
class recall	94.08%	94.66%	

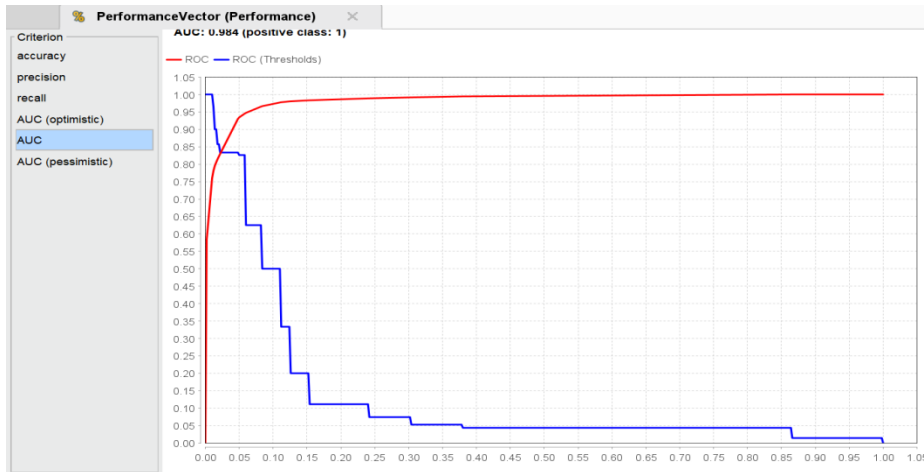
Gambar 4.9 Confusion Matrix Hasil klasifikasi gagal jantung menggunakan Algoritma Decision Tree pada Rapidminer 10.1

$$: Recall = \frac{TP}{(TP+FN)} \quad (4.4)$$

$$R(0) = \frac{270}{(270+17)} = \frac{270}{287} = 94,08\%$$

$$R(1) = \frac{337}{(337+19)} = \frac{337}{356} = 94,66\%$$

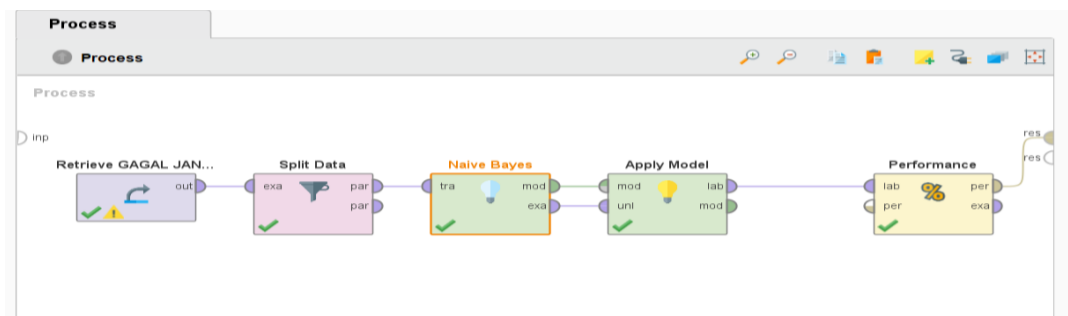
Berikut adalah hasil kurva AUC yaitu mendapatkan nilai 0,984 seperti gambar dibawah ini.



Gambar 4.10 Nilai dan Curve AUC Algoritma Decision Tree C. 45 Split Validation

3. Penelitian Menggunakan Algoritma Naïve Bayes Split Validation

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 70:30. dapat dilihat pada gambar berikut



Gambar 4.11. Proses menggunakan Algoritma Naïve Bayes Split Validation

PerformanceVector (Performance)

Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

Table View | Plot View

Changes to a table showing the confusion matrix.

accuracy: 85.54%

	true 0	true 1	class precision
pred. 0	239	45	84.15%
pred. 1	48	311	86.63%
class recall	83.28%	87.36%	

Gambar 4.12 Nilai accuracy Algoritma Naïve Byes Split Validation

$$\begin{aligned}
 \text{akurasi} &= \frac{(\text{TN} + \text{TP})}{(\text{TN} + \text{FN} + \text{FP} + \text{TP})} \\
 &= \frac{(239 + 311)}{(239 + 45 + 48 + 311)} \\
 &= \frac{550}{643} = 0,855365
 \end{aligned}$$

Didapatkan akurasi = 85,54%

	true 0	true 1	class precision
pred. 0	239	45	84.15%
pred. 1	48	311	86.63%
class recall	83.28%	87.36%	

Gambar 4.13. Precision Algoritma Naïve Byes Split Validation

$$\begin{aligned}
 \text{precision} &= \frac{\text{TP}}{(\text{TP} + \text{FP})} \\
 &= \frac{311}{(311 + 48)} \\
 &= \frac{311}{284} = 0,841549
 \end{aligned}$$

precision = 84,15%

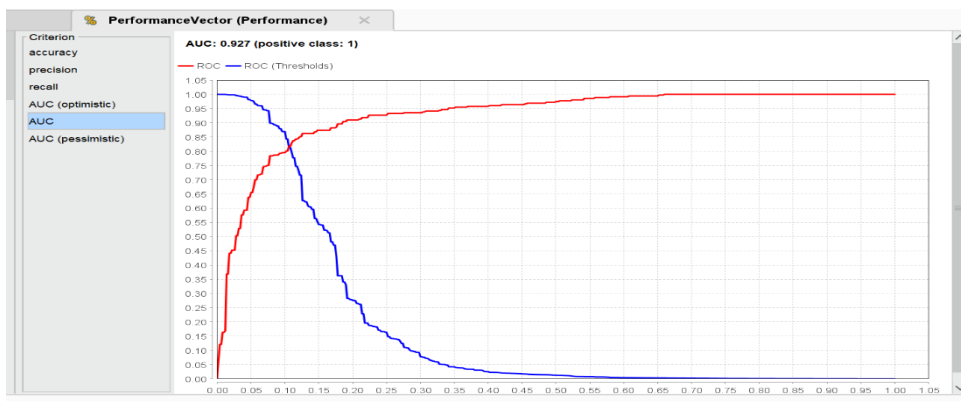
	true 0	true 1	class precision
pred. 0	239	45	84.15%
pred. 1	48	311	86.63%
class recall	83.28%	87.36%	

Gambar 4.14 Confusion Matrix Hasil klasifikasi gagal jantung menggunakan Algoritma Naïve Bayes pada Rapidminer 10.1

$$\begin{aligned}
 precision &= \frac{TP}{(TP + FP)} \\
 &= \frac{311}{(311 + 48)} \\
 &= \frac{311}{287} = 0,83275
 \end{aligned}$$

precision = 83,28%

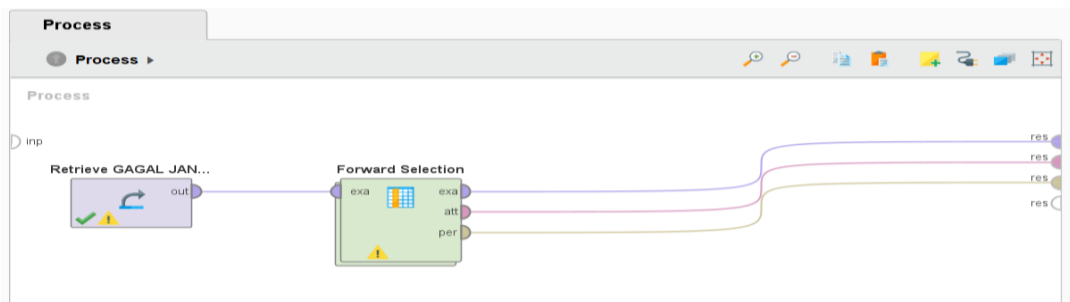
Berikut adalah hasil kurva AUC yaitu mendapatkan nilai 0,830 seperti gambar dibawah ini:



Gambar 4.15. Nilai dan Curve AUC Algoritma Naïve Bayes Split Validation

4.3 Penelitian Forward Selections Decision Tree C4.5 dan Algoritma Naïve Bayes

Proses untuk klasifikasi dengan menggunakan *Forward Selection* mencari Nilai *Accuracy*, *Precision* dan *Recall* Algoritma *Decision Tree* Seperti gambar dibawah ini.



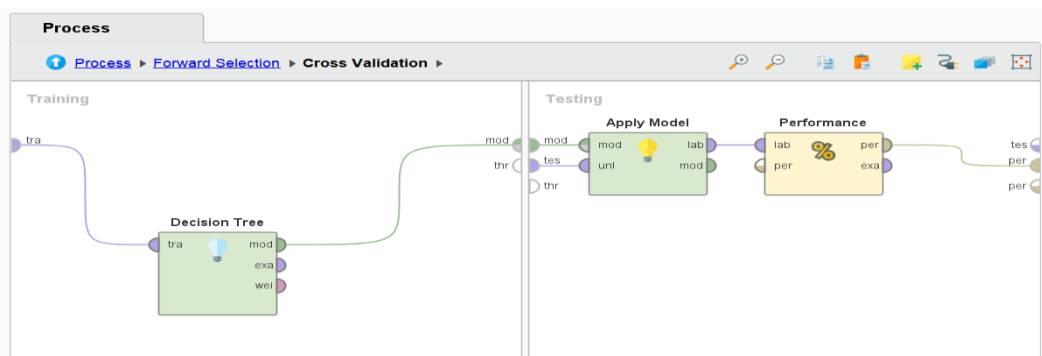
Gambar 4.16 Proses Forward Selections Algoritma C4.5

Proses untuk klasifikasi dengan menggunakan *Forward Selection dan Cross Validation* mencari Nilai *Accuracy, Precesion dan Reccall* Algoritma *Decision Tree* Seperti gambar dibawah ini.



Gambar 4.17 Proses Forward Selections Algoritma C4.5

Proses untuk klasifikasi dengan menggunakan *Forward Selection. Algoritma Decision Tree, Aply Model dan Performance Cross Validation* mencari Nilai *Accuracy, Precesion dan Reccall* Seperti gambar dibawah ini



Gambar 4.18 Proses Forward Selections Algoritma C4.5

Result History

PerformanceVector (Performance)

Table View Plot View

accuracy: 85.73% +/- 3.41% (micro average: 85.73%)

	true 0	true 1	class precision
pred. 0	340	61	84.79%
pred. 1	70	447	86.46%
class recall	82.93%	87.99%	

Gambar 4.18 Akurasi Algoritma C4.5 Forward Selections

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

$$\text{Akurasi} = \frac{(340+447)}{(340+447+61+70)} = \frac{787}{918} = 85,73\%$$

Result History PerformanceVector (Performance) X

Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

precision: 86.57% +/- 3.78% (micro average: 86.46%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	340	61	84.79%
pred. 1	70	447	86.46%
class recall	82.93%	87.99%	

Gambar 4.19 Precision Algoritma C.4.5. Forward Selections

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

$$:p(1) = \frac{340}{(340+61)} = \frac{340}{401} = 84,79\%$$

$$:p(0) = \frac{447}{(447+70)} = \frac{447}{517} = 86,46\%$$

Result History PerformanceVector (Performance) X

Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

recall: 87.99% +/- 3.64% (micro average: 87.99%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	340	61	84.79%
pred. 1	70	447	86.46%
class recall	82.93%	87.99%	

Gambar 4.20 Reccal Algoritma C4.5 Forward Selections

$$: \text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

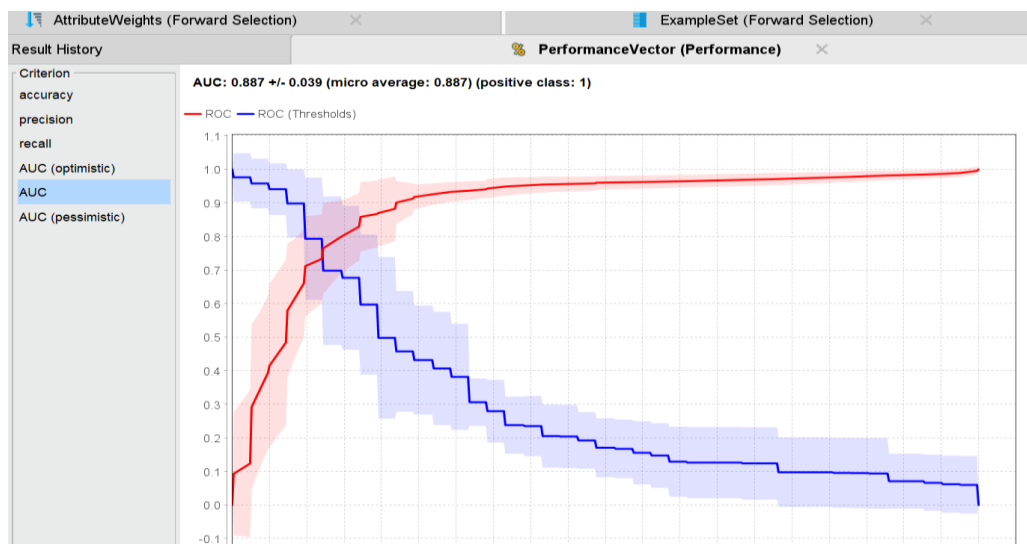
$$R(1) = \frac{340}{(340+70)} = \frac{340}{410} = 82,92\%$$

$$R(0) = \frac{447}{(447+70)} = \frac{447}{517} = 86,46\%$$

Perhitungan F=Measure nilai Tengah recall

$$F1 \text{ Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

dalam kasus di atas, $F1 \text{ Score} = 2 * (85,57\% * 87,99\%) / (85,57\% + 87,99\%) = 15.059/173,56 = 86,54\%$



Gambar 4.21 AUC Algoritma C4.5 Forward Selections

PerformanceVector

PerformanceVector:

accuracy: 85.73% +/- 3.41% (micro average: 85.73%)

ConfusionMatrix:

True:	0	1
0:	340	61
1:	70	447

precision: 86.57% +/- 3.78% (micro average: 86.46%) (positive class: 1)

ConfusionMatrix:

True: 0 1
 0: 340 61
 1: 70 447

recall: 87.99% +/- 3.64% (micro average: 87.99%) (positive class: 1)

ConfusionMatrix:

True: 0 1
 0: 340 61
 1: 70 447

AUC (optimistic): 0.908 +/- 0.034 (micro average: 0.908) (positive class: 1)

AUC: 0.887 +/- 0.039 (micro average: 0.887) (positive class: 1)

AUC (pessimistic): 0.867 +/- 0.047 (micro average: 0.867) (positive class: 1)

Row No.	HeartDisease	ST_Slope	Cholesterol	ChestPainT...	Sex	ExerciseAn...
1	0	Up	289	ATA	M	N
2	1	Flat	180	NAP	F	N
3	0	Up	283	ATA	M	N
4	1	Flat	214	ASY	F	Y
5	0	Up	195	NAP	M	N
6	0	Up	339	NAP	M	N
7	0	Up	237	ATA	F	N
8	0	Up	208	ATA	M	N
9	1	Flat	207	ASY	M	Y
10	0	Up	284	ATA	F	N
11	0	Up	211	NAP	F	N
12	1	Flat	164	ATA	M	Y

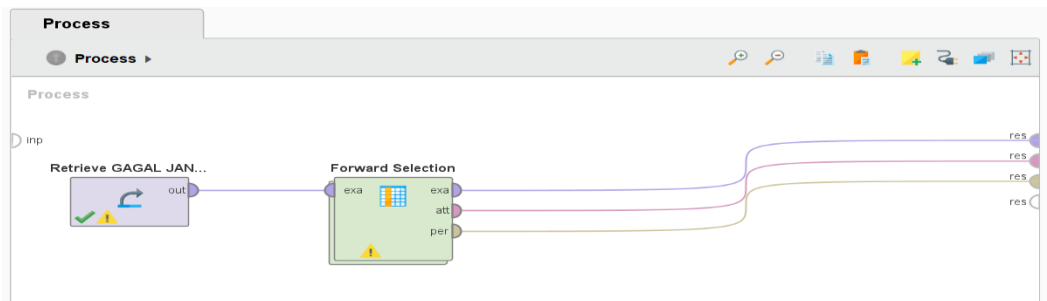
ExampleSet (918 examples, 1 special attribute, 5 regular attributes)

Gambar 4.22 Atribut Forward Selection

attribute	weight
Age	0
Sex	1
ChestPa...	1
RestingBP	0
Choleste...	1
FastingBS	0
Resting...	0
MaxHR	0
Exercise...	1
Oldpeak	0
ST_Slope	1

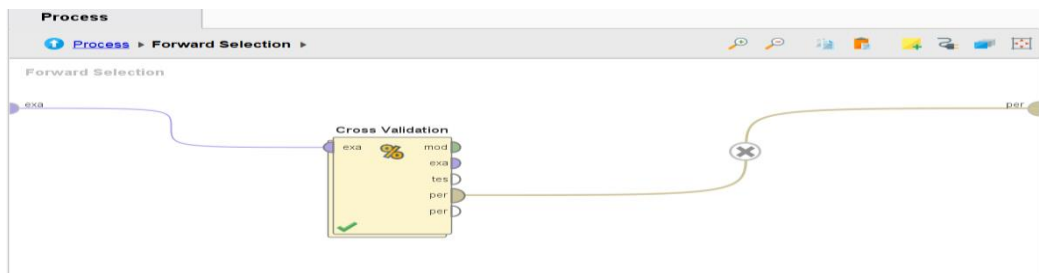
Gambar 4.23 Atribut Weights

Proses untuk klasifikasi dengan menggunakan *Forward Selection* mencari Nilai *Accuracy*, *Precesion* dan *Reccall* dengan algoritama *Naïve Bayes* Seperti gambar dibawah ini.



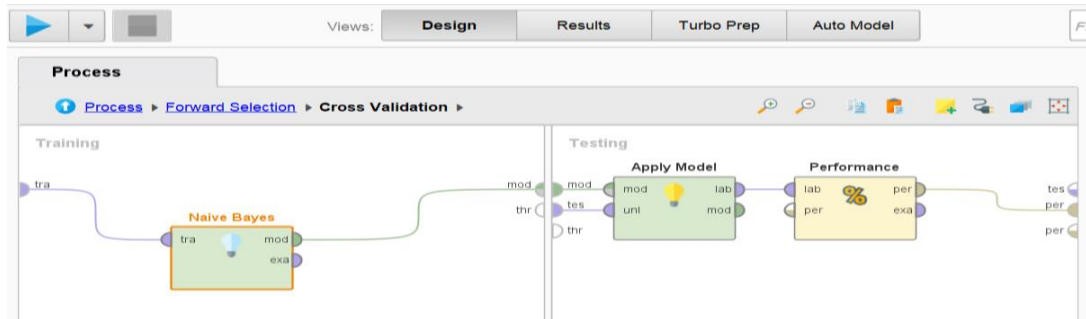
Gambar 4.24 Proses Forward Selections Algoritma Naïve Bayes

Proses untuk klasifikasi dengan menggunakan *Forward Selection* dan *Cross Validation* mencari Nilai *Accuracy*, *Precesion* dan *Reccall* Algoritma *Naïve Bayes* Seperti gambar dibawah ini



Gambar 4.25 Proses Forward Selections Algoritma Naïve Bayes

Proses untuk klasifikasi dengan menggunakan *Forward Selection*. Algoritma *Naive Bayes*, *Apply Model* dan *Performance Cross Validation* mencari Nilai *Accuracy*, *Precesion* dan *Reccall* Seperti gambar dibawah ini



Gambar 4.26 Proses Forward Selections Algoritma Naïve Bayes

Result History

PerformanceVector (Performance)

Table View Plot View

accuracy: 86.16% +/- 2.59% (micro average: 86.17%)

	true 0	true 1	class precision
pred. 0	344	61	84.94%
pred. 1	66	447	87.13%
class recall	83.90%	87.99%	

class recall

Gambar 4. 27 Akurasi Agoritma Naïve Bayes. Forward Selctions

$$\begin{aligned}
 \text{akurasi} &= \frac{(TN + TP)}{(TN + FN + FP + TP)} \\
 &= \frac{(447 + 344)}{(447 + 66 + 61 + 344)} \\
 &= \frac{791}{918} = 0,861655
 \end{aligned}$$

$$\text{akurasi} = 86,16\%$$

Result History PerformanceVector (Performance)

Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

precision: 87.27% +/- 3.62% (micro average: 87.13%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	344	61	84.94%
pred. 1	66	447	87.13%
class recall	83.90%	87.99%	

Gambar 4.28 Presions Algoritma Naïve Bayes. Forward Selctions

$$Precision = \frac{TP}{(TP + FP)}$$

$$= \frac{344}{(344 + 61)}$$

$$= \frac{344}{405} = 0,84938$$

$$precision = 84,94\%$$

Result History PerformanceVector (Performance)

Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

recall: 87.99% +/- 2.99% (micro average: 87.99%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	344	61	84.94%
pred. 1	66	447	87.13%
class recall	83.90%	87.99%	

Gambar 4.29 Recall Algoritma Naïve Bayes. Forward Selctions

$$Recall = \frac{TP}{(TP + FP)}$$

$$= \frac{344}{(344 + 66)}$$

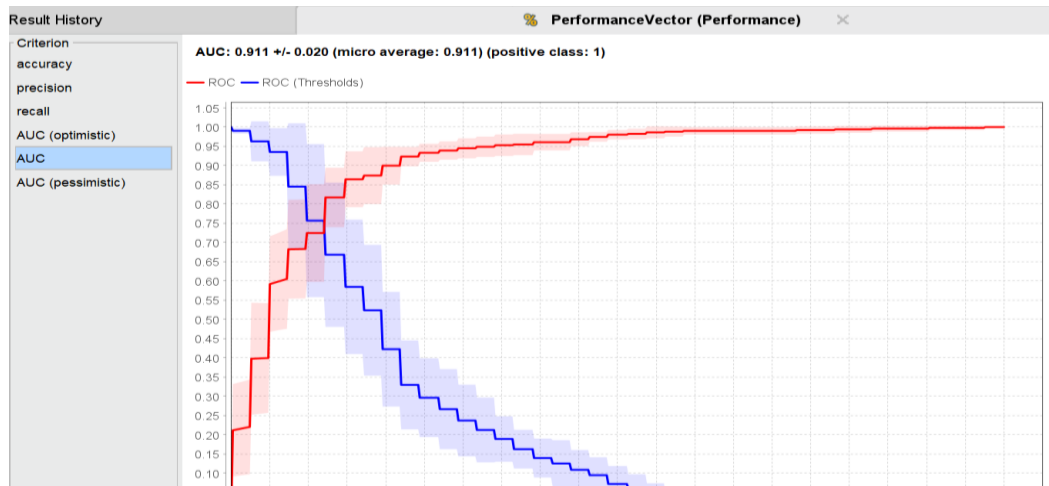
$$= \frac{344}{410} = 0,8390$$

$$Recall = 83,90\%$$

Perhitungan F=Measure nilai Tengah recall

$$F1 \text{ Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

dalam kasus di atas, $F1 \text{ Score} = 2 * (87,27\% * 87,99\%) / (87,27\% + 87,99\%) = 15358/175 = 87,76\%$



Gambar 4.30 AUC Algoritma Naive Bayes

PerformanceVector

PerformanceVector:

accuracy: 86.16% +/- 2.59% (micro average: 86.17%)

ConfusionMatrix:

```
True:  0      1
0:     344    61
1:     66    447
```

precision: 87.27% +/- 3.62% (micro average: 87.13%) (positive class: 1)

ConfusionMatrix:

```
True:  0      1
0:     344    61
1:     66    447
```

recall: 87.99% +/- 2.99% (micro average: 87.99%) (positive class: 1)

ConfusionMatrix:

```

True:  0      1
0:     344   61
1:     66   447

```

AUC (optimistic): 0.912 +/- 0.020 (micro average: 0.912) (positive class: 1)

AUC: 0.911 +/- 0.020 (micro average: 0.911) (positive class: 1)

AUC (pessimistic): 0.911 +/- 0.020 (micro average: 0.911) (positive class: 1)

Row No.	HeartDisease	ST_Slope	Cholesterol	ChestPainT...	ExerciseAn...	RestingECG
1	0	Up	289	ATA	N	Normal
2	1	Flat	180	NAP	N	Normal
3	0	Up	283	ATA	N	ST
4	1	Flat	214	ASY	Y	Normal
5	0	Up	195	NAP	N	Normal
6	0	Up	339	NAP	N	Normal
7	0	Up	237	ATA	N	Normal
8	0	Up	208	ATA	N	Normal
9	1	Flat	207	ASY	Y	Normal
10	0	Up	284	ATA	N	Normal
11	0	Up	211	NAP	N	Normal
12	1	Flat	164	ATA	Y	ST

Gambar 4.31 Data Atribut

attribute	weight
Age	0
Sex	0
ChestPa...	1
RestingBP	0
Choleste...	1
FastingBS	0
Resting...	1
MaxHR	0
Exercise...	1
Oldpeak	0
ST_Slope	1

Gambar 4.32 Atribut Weights

Tabel 4. 2 Perbandingan Akurasi

No	Algoritma	Akurasi	AUC	Recall	Precision
1	C4.5	94,40%	0,984	94,66%	95,20%
2	C4.5 + Forward Selctions	85,73%	0,887	87,36%	86,63%
3	Naïve Byes	85,54%	0,927	87,36%	86,63%
4	Naïve Byes + Forward Selctions	86,16%	0,911	87,13%	87,99%
5	Perhitungan F- Measure C4.5	86,54%			
6	Perhitungan F- Measure Naïve Bayes	87,76%			

4.4 Evaluasi

Dari hasil pengujian pada table 4.2 diatas perbandingan dua algoritma sebelum menggunakan forward selection dapatkan disimpulkan sebagai berikut:

A. Algoritam Decision tree

1. Perbandingan Nilai Akurasi tidak menggunakan forward selection diperoleh dengan menggunakan Algoritma Decision tree C4.5 Nilai Akurasi 94,40% dan algoritma Naïve Byes didapatkan Nilai Akurasi 85,5%. akurasi yang terbaik yaitu Algoritma Decision tree Nilai akurasi yang tinggi menunjukkan bahwa model tersebut mampu memprediksi atau mengklasifikasikan data dengan benar pada sebagian besar data.
2. Perbandingan Area Under Curve (AUC) tidak menggunakan forward selection diperoleh dengan menggunakan Algoritma Decision tree C4.5 Nilai 0,984 dan

algoritma Naïve Byes didapatkan 0,927 Area Under Curve (AUC) yang terbaik yaitu Algoritma Decision tree C4.5 AUC mengukur kinerja suatu model pada tingkat klasifikasi biner, dengan memperhitungkan tingkat false positive rate (FPR) dan true positive rate (TPR) yang diberikan oleh model. Nilai AUC yang tinggi menunjukkan bahwa model tersebut mampu membedakan antara kelas positif dan negatif dengan baik, dan memiliki tingkat FPR dan TPR yang seimbang. Nilai AUC yang mendekati 1 menunjukkan performa model yang sangat baik, sedangkan nilai AUC yang mendekati 0 menunjukkan performa model yang buruk.

3. Perbandingan Recall tidak menggunakan forward selection diperoleh dengan menggunakan Algoritma Decision tree C4.5 Nilai 94,66% dan algoritma Naïve Byes didapatkan 87,36% Recall yang terbaik yaitu Algoritma Decision tree Nilai Recall yang tinggi menunjukkan bahwa model tersebut mampu mengidentifikasi kelas positif dengan baik dan memiliki sedikit kesalahan dalam mengidentifikasi kelas negatif sebagai kelas positif.
4. Perbandingan Precision tidak menggunakan forward selection diperoleh dengan menggunakan Algoritma Decision tree C4.5 Nilai 95,20% dan algoritma Naïve Byes didapatkan 86,63% Precision yang terbaik yaitu Algoritma Decision Tree Precision yang tinggi menunjukkan bahwa model tersebut memiliki sedikit kesalahan dalam memprediksi kelas negatif sebagai kelas positif dan memprediksikan kelas positif dengan benar. Precision sangat penting dalam aplikasi yang membutuhkan prediksi kelas positif yang akurat.

B. Perbandingan dengan menggunakan Forward Selection

1. Perbandingan Nilai Akurasi menggunakan forward selection diperoleh dengan menggunakan Algoritma Decision tree C4.5 Nilai Akurasi 85,73% dan algoritma Naïve Byes didapatkan Nilai Akurasi 86,16%. akurasi yang terbaik yaitu Algoritma Naïve Byes karena dilakukan forward selectionn terdapat pengurangan atribut yang terbuang memiliki nilai kontribusi tinggi.
2. Perbandingan Area Under Curve (AUC) menggunakan forward selection diperoleh dengan menggunakan Algoritma Decision tree C4.5 Nilai **0,887** dan

algoritma Naïve Byes didapatkan 0,911 Area Under Curve (AUC) yang terbaik yaitu Algoritma Naïve Bayes termasuk dalam kategori termasuk kategori good classification.

3. AUC mengukur kinerja suatu model pada tingkat klasifikasi biner, dengan memperhitungkan tingkat false positive rate (FPR) dan true positive rate (TPR) yang diberikan oleh model. Nilai AUC yang tinggi menunjukkan bahwa model tersebut mampu membedakan antara kelas positif dan negatif dengan baik, dan memiliki tingkat FPR dan TPR yang seimbang. Nilai AUC yang mendekati 1 menunjukkan performa model yang sangat baik, sedangkan nilai AUC yang mendekati 0 menunjukkan performa model yang buruk.
4. Perbandingan Recall menggunakan forward selection diperoleh dengan menggunakan Algoritma Decision tree C4.5 Nilai **87,27%** dan algoritma Naïve Byes didapatkan 87,13% Recall yang terbaik yaitu Algoritma Decision tree Nilai Recall yang tinggi menunjukkan bahwa model tersebut mampu mengidentifikasi kelas positif dengan baik dan memiliki sedikit kesalahan dalam mengidentifikasi kelas negatif sebagai kelas positif.
5. Perbandingan **Precissio** menggunakan forward selection diperoleh dengan menggunakan Algoritma Decision tree C4.5 Nilai **87,99%** dan algoritma Naïve Byes didapatkan 87,97% Precissio kedua algoritma memiliki nilai yang sama. Precision yang tinggi menunjukkan bahwa model tersebut memiliki sedikit kesalahan dalam memprediksi kelas negatif sebagai kelas positif dan memprediksikan kelas positif dengan benar. Precision sangat penting dalam aplikasi yang membutuhkan prediksi kelas positif yang akurat.

C. Perhitungan Nilai Tengah

- 1) Perbandingan **perhitungan nilai** diperoleh dengan menggunakan Algoritma Decision tree C4.5 Nilai **86,54%** dan algoritma Naïve Byes didapatkan **87,76%** perhitungan nilai tengah yang terbaik yaitu algoritam naïve bayes mengindikasikan bahwa model klasifikasi precision dan recall yang baik.