

## BAB IV HASIL DAN PEMBAHASAN

### 1.1 Hasil Pengujian dan Pembahasan

Dalam jurnal Prosiding Seminar Nasional Darmajaya mengutip (Tinggi et al., 2022) ada beberapa tahapan untuk membangun pohon keputusan algoritma C4.5 yaitu :

1. Memerlukan pelatihan data, dapat diambil dari data historis yang telah terjadi dan dikelompokkan kedalam kelas-kelas tertentu.
2. Tentukan akar pohon dengan menghitung nilai gain tertinggi setiap atribut atau nilai indeks entropy terendah. Sebelumnya, nilai indeks entropy di hitung menggunakan rumus :

$$Entropy(S) = \sum_{i=1}^n -P_i * \log_2 P_i$$

3. Nilai gain dengan rumus

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

4. Untuk menghitung gain ratio perlu diketahui suatu tren baru yang di sebut *Split Information* dengan rumus :

$$Entropy(S) = \sum_{t=1}^i \frac{S_1}{S} \log_2 \frac{S_1}{S}$$

5. Selanjutnya menghitung gain ratio

$$Gainratio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)}$$

6. Ulangi langkah 2 sampai semua *record* terpecahkan. Proses pohon keputusan berakhir ketika :
  - a Semua tupel dalam catatan simpul m adalah kelas yang sama
  - b Atribut dalam dataset tidak di bagi lagi
  - c Cabang kosong tidak memiliki catatan

Tahap perhitungan matematika algoritma c4.5 ini menggunakan rumus diatas. Pada penelitian ini :

#### 4.1.1 Pengujian Hitung Matematika

Tahap pertama dalam penelitian ini adalah menghitung hasil proses *training* data yang sudah dilakukan sebelumnya, dengan jumlah data 1613 terdapat nilai optimal 1026 kolom dan suboptimal 578 kolom.

Langkah pertama pada penelitian ini yaitu proses perhitungan nilai *entropy*, untuk mendapatkan *entropy* masing-masing atribut penulis mencari nilai *entropy* untuk keseluruhan data (*class* = "optimal" dan *class* = "suboptimal") terlebih dahulu.

Jumlah *class* target (*class*) = 2 ("optimal" dan "suboptimal")

Jumlah optimal = 1.026

Jumlah suboptimal = 578

Total keseluruhan = 1.613

Berdasarkan rumus perhitungan *entropy* yang diperoleh dari :

$$\begin{aligned} Entropy(class) &= (((-1026) / 1613 \times \text{Log}_2(1026/1613)) + (((-578) / 1613 \times \\ &\quad \text{Log}_2(578/1613)))) \\ &= 0,945887 \end{aligned}$$

Atribut	Jumlah Total	Optimal	Sub Optimal	Entropy	Info Gain	Split Info	Gain Rasio
Class	1613	1026	587	0,897852			

Tabel 4.1 Nilai *entropy* total 1-1

Hasil perhitungan *entropy* total dari seluruh atribut terdapat dalam tabel di atas. Selanjutnya, penulis melanjutkan dengan melakukan perhitungan nilai *entropy*, gain, split info, dan gain rasio dari atribut kadar oksigen, suhu air, pH air, populasi, dan protein. Langkah ini dilakukan untuk mengidentifikasi atribut yang paling signifikan dalam mempengaruhi pembentukan pohon keputusan.

1. Menghitung Nilai *Entropy*
  - a. *Entropy* (Kadar Oksigen)

*Entropy* (Baik)

$$= (((-686) / 1101 \times \text{Log}_2(686/1101)) + (((-415) / 1101 \times \text{Log}_2(415/1101)))) \\ = 0,95584$$

*Entropy* (Buruk)

$$= (((-344) / 512 \times \text{Log}_2(344/512)) + (((-168) / 512 \times \text{Log}_2(168/512)))) \\ = 0,91299$$

*Entropy* (Total)

$$= (1101 \times 0,95584 + 512 \times 0,91299) \\ = 0,94224$$

b. *Entropy* (Suhu Air)

*Entropy* (<25)

$$= (((-284) / 476 \times \text{Log}_2(284/476)) + (((-192) / 476 \times \text{Log}_2(192/476)))) \\ = 0,97288$$

*Entropy* (>29)

$$= (((-742) / 1137 \times \text{Log}_2(742/1137)) + (((-395) / 1137 \times \text{Log}_2(395/1137)))) \\ = 0,93172$$

*Entropy* (Total)

$$= (476 \times 0,97288 + 1137 \times 0,93172) \\ = 0,94387$$

c. *Entropy* (pH Air)

*Entropy* (<0,013)

$$= (((-321) / 504 \times \text{Log}_2(321/504)) + (((-183) / 504 \times \text{Log}_2(183/504))))$$

$$= 0,94522$$

*Entropy (>0,019)*

$$= (((-706) / 1109 \times \text{Log}_2(706/1109)) + (((-403) / 1109 \times \text{Log}_2(403/1109))))$$

$$= 0,94546$$

*Entropy (Total)*

$$= (504 \times 0,94522 + 1109 \times 0,94546)$$

$$= 0,94538$$

d. *Entropy (Populasi)*

*Entropy (10ribu)*

$$= (((-192) / 323 \times \text{Log}_2(192/323)) + (((-131) / 323 \times \text{Log}_2(131/323))))$$

$$= 0,97411$$

*Entropy (>20 ribu)*

$$= (((-646) / 965 \times \text{Log}_2(646/965)) + (((-319) / 965 \times \text{Log}_2(319/965))))$$

$$= 0,91550$$

*Entropy (>50 ribu)*

$$= (((-188) / 325 \times \text{Log}_2(188/325)) + (((-137) / 325 \times \text{Log}_2(137/325))))$$

$$= 0,98216$$

*Entropy (Total)*

$$= (323 \times 0,97411 + 965 \times 0,91550 + 325 \times 0,98216)$$

$$= 0,94067$$

e. *Entropy (Protein)*

*Entropy (<28)*

$$= (((-206) / 313 \times \text{Log}_2(206/313)) + (((-107) / 313 \times \text{Log}_2(107/313))))$$

$$= 0,92658$$

*Entropy (>29)*

$$= (((-820) / 1300 \times \text{Log}_2(820/1300)) + (((-480) / 1300 \times \text{Log}_2(480/1300))))$$

$$= 0,95007$$

*Entropy (Total)*

$$= (313 \times 0,92658 + 1300 \times 0,95007)$$

$$= 0,94551$$

Kemudian, langkah selanjutnya adalah menghitung nilai gain pada setiap atribut seperti kadar oksigen, suhu air, pH air, populasi, dan protein. Perhitungan tersebut dapat dijalankan dengan langkah-langkah sebagai berikut:

## 2. Menghitung Nilai *Gain*

### a. Menghitung nilai *gain* kadar oksigen

$$\text{Entropy Total} = (0,94588 - 0,94224)$$

$$= 0,00364$$

### b. Menghitung nilai *gain* suhu air

$$\text{Entropy Total} = (0,94588 - 0,97288)$$

$$= 0,00201$$

### c. Menghitung nilai *gain* ph air

$$\text{Entropy Total} = (0,94588 - 0,94538)$$

$$= 0,0005$$

### d. Menghitung nilai *gain* populasi

$$\begin{aligned} \text{Entropy Total} &= (0,94588 - 0,94067) \\ &= 0,00521 \end{aligned}$$

e. Menghitung nilai *gain* protein

$$\begin{aligned} \text{Entropy Total} &= (0,94588 - 0,94551) \\ &= 0,00037 \end{aligned}$$

Langkah berikutnya setelah menghitung nilai *gain* terhitung, penulis menghitung nilai *split info* dari atribut, kadar oksigen, suhu air, ph air, populasi, nilai protein.

### 3. Menghitung Nilai *Split Info*

a. Nilai *split info* kadar oksigen

$$\begin{aligned} \text{Entropy} &= -(1101/1613) \times \text{Log}_2 (1101/1613) - (512/1613) \times \text{Log}_2 (512/1613) \\ &= 0,90155 \end{aligned}$$

b. Nilai *split info* suhu air

$$\begin{aligned} \text{Entropy} &= -(476/1613) \times \text{Log}_2 (476/1613) - (1137/1613) \times \text{Log}_2 (1137/1613) \\ &= 0,87522 \end{aligned}$$

c. Nilai *split info* ph air

$$\begin{aligned} \text{Entropy} &= -(504/1613) \times \text{Log}_2 (504/1613) - (1109/1613) \times \text{Log}_2 (1109/1613) \\ &= 0,89599 \end{aligned}$$

d. Nilai *split info* populasi

$$\begin{aligned} \text{Entropy} &= -(323/1613) \times \text{Log}_2 (323/1613) - (965/1613) \times \text{Log}_2 (965/1613) \times \\ &\quad \text{Log}_2 (325/1613) - (325/1613) \\ &= 0,908 \end{aligned}$$

e. Nilai *split info* protein

$$\begin{aligned} \text{Entropy} &= -(313/1613) \times \text{Log}_2(313/1613) - (1300/1613) \times \text{Log}_2(1300/1613) \\ &= 0,70986 \end{aligned}$$

Setelah menghitung nilai *split info* diketahui, langkah selanjutnya yaitu menghitung nilai *gain rasio*. dari atribut, kadar oksigen, suhu air, ph air, populasi, nilai protein.

4. Menghitung nilai *gain rasio*

a. *Gain rasio* kadar oksigen

$$\begin{aligned} \text{Gainrasio (S, kadar oksigen)} &= 0,00364 / 0,90155 \\ &= 0,004040 \end{aligned}$$

b. *Gain rasio* suhu air

$$\begin{aligned} \text{Gainrasio (S, suhu air)} &= 0,00201 / 0,87522 \\ &= 0,00230 \end{aligned}$$

c. *Gain rasio* ph air

$$\begin{aligned} \text{Gainrasio (S, A)} &= 0,0005 / 0,89599 \\ &= 0,00055 \end{aligned}$$

d. *Gain rasio* populasi

$$\begin{aligned} \text{Gainrasio (S, ph air)} &= 0,00521 / 0,908 \\ &= 0,00574 \end{aligned}$$

e. *Gain rasio* protein

$$\begin{aligned} \text{Gainrasio(S, protein)} &= 0,00037 / 0,70986 \\ &= 0,00051 \end{aligned}$$

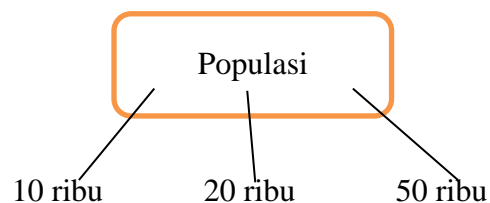
Setelah nilai split info dan gain rasio di hitung, peneliti membuat tabel hasil perhitungan dari seluh data atribut kadar oksigen, ph air, suhu air, populasi, dan protein yang sudah di hitung bernilai *entropy*, *gain*, *split info*, *gain rasio* bisa di lihat pada tabel di bawah ini :

Atribut	Jumlah Total	Optimal	Sub Optimal	Entropy	Info Gain	Split Info	Gain Rasio
<i>Class</i>	1613	1026	587	0,94588			
Kadar Oksigen				0,94224	0,0002	0,9015	0,00404
<25	1101	686	415	0,95584			
>29	512	344	168	0,91299			
Suhu Air				0,94387	0,0129	0,8752	0,00230
<0,013	476	284	192	0,97288			
>0,019	1137	742	395	0,93172			
Ph air				0,94538	0,0005	0,8959	0,00055
<0,013	504	321	183	0,94522			
>0,019	1109	706	403	0,94546			
Populasi				0,94067	0,0291	0,8959	0,00574
10rb	323	192	131	0,97411			
20rb	965	646	319	0,91550			
50rb	325	188	137	0,98216			
Protein				0,94551	0,0004	0,7098	0,00051
<28	313	206	107	0,92658			
>30	1300	820	480	0,95007			

Tabel 4.2 Hasil perhitungan atribut 1

Hasil perhitungan *entropy*, *Info gain*, *split info* dan *gain rasio* sudah diketahui dengan *gain rasio* tertinggi adalah populasi, yaitu 0,00574. Dengan demikian populasi dijadikan *node* akar. Dari hasil tersebut dapat di gambarkan keputusan sementara seperti gambar 4.1.

Gambar 4.1. Subpohon hasil perhitungan node 1





Langkah selanjutnya. Penulis mengulang perhitungan atribut populasi nilai cabang 10 ribu, untuk mengorientasikan pada populasi 10 ribu, penulis memulai perhitungan *entropy*, *gain*, *split info*, *gain rasio* dari atribut tersebut.

Jumlah *class* target (*class*) = 2 (“optimal” dan “suboptimal”)

Jumlah optimal = 192

Jumlah suboptimal = 131

Total keseluruhan = 323

$$\text{Entropy total} = (((-192) / 323 \times \text{Log}_2(192/323)) + (((-131) / 323 \times \text{Log}_2(131/323))))$$

$$= 0,97411$$

1. Menghitung Nilai *Entropy* Cabang 10 ribu

a. *Entropy* (Kadar Oksigen)

$$\text{Entropy}(\text{Baik}) = (((-58)/96 \times \text{Log}_2(58/96)) + (((-38) / 96 \times \text{Log}_2(38/96))))$$

$$= 0,96846$$

$$\text{Entropy}(\text{Buruk}) = (((-137)/227 \times \text{Log}_2(137/227)) + (((-77) / 227 \times \text{Log}_2(77/227))))$$

$$= 0,97633$$

$$\text{Entropy}(\text{Total}) = (96 \times 0,96846 + 227 \times 0,97633)$$

$$= 0,97399$$

b. *Entropy* (Suhu Air)

$$\text{Entropy}(<25) = (((-55) / 109 \times \text{Log}_2(55 / 109)) + (((-54) / 109 \times \text{Log}_2(54 / 109))))$$

$$= 0,99993$$

$$\text{Entropy}(>29) = (((-137) / 214 \times \text{Log}_2(137 / 214)) + (((-77) / 135 \times \text{Log}_2(77 / 135))))$$

$$= 0,94252$$

$$\text{Entropy (Total)} = (109 \times 0,99993 + 214 \times 0,94252)$$

$$= 0,96190$$

c. *Entropy* (pH Air)

$$\text{Entropy (<0,013)} = (((-137) / 226 \times \text{Log}_2(137 / 226)) + (((-89) / 226 \times \text{Log}_2(89 / 226))))$$

$$= 0,96721$$

$$\text{Entropy (>0,019)} = (((-55) / 97 \times \text{Log}_2(55 / 97)) + (((-26) / 97 \times \text{Log}_2(26 / 97))))$$

$$= 0,98700$$

$$\text{Entropy (Total)} = (227 \times 0,96721 + 97 \times 0,98700)$$

$$= 0,94506$$

d. *Entropy* (protein)

$$\text{Entropy (<28)} = (((-45) / 71 \times \text{Log}_2(45 / 71)) + (((-26) / 71 \times \text{Log}_2(26 / 71))))$$

$$= 0,94770$$

$$\text{Entropy (>30)} = (((-147) / 252 \times \text{Log}_2(147 / 252)) + (((-105) / 252 \times \text{Log}_2(105 / 252))))$$

$$= 0,97986$$

$$\text{Entropy (Total)} = (9 \times 0,99107 + 30 \times 0,94945)$$

$$= 0,97279$$

2. Mengitung *Gain*

a. Menghitung nilai *gain* kadar oksigen

$$\text{Entropy Total} = (0,97411 - 0,97399)$$

$$= 0,00012$$

b. Menghitung nilai *gain* suhu air

$$\begin{aligned} \text{Entropy Total} &= (0,97411 - 0,96190) \\ &= 0,01222 \end{aligned}$$

c. Menghitung nilai *gain* ph air

$$\begin{aligned} \text{Entropy Total} &= (0,97411 - 0,97315) \\ &= 0,00096 \end{aligned}$$

d. Menghitung nilai *gain* ph air

$$\begin{aligned} \text{Entropy Total} &= (0,97411 - 0,97279) \\ &= 0,00132 \end{aligned}$$

3. Menghitung *split info*

a. Nilai *split info* kadar oksigen

$$\begin{aligned} \text{Entropy} &= -(96/323) \times \text{Log}_2(96/323) - (227/323) \times \text{Log}_2(227/323) \\ &= 0,87786 \end{aligned}$$

b. Nilai *split info* suhu air

$$\begin{aligned} \text{Entropy} &= -(109/323) \times \text{Log}_2(109/323) - (214/323) \times \text{Log}_2(214/323) \\ &= 0,92237 \end{aligned}$$

c. Nilai *split info* ph air

$$\begin{aligned} \text{Entropy} &= -(226/323) \times \text{Log}_2(226/323) - (97/323) \times \text{Log}_2(97/323) \\ &= 0,88167 \end{aligned}$$

d. Nilai *split info* protein

$$\begin{aligned} \text{Entropy} &= -(71/323) \times \text{Log}_2(71/323) - (252/323) \times \text{Log}_2(252/323) \\ &= 0,00173 \end{aligned}$$

4. Menghitung *Gain Rasio*

a. *Gain rasio* kadar oksigen

$$\text{Gainrasio}(S, \text{kadar oksigen}) = 0,00012 / 0,87786$$

$$= 0,000132$$

b. *Gain rasio* suhu air

$$\begin{aligned} \text{Gainrasio}(S, \text{ suhu air}) &= 0,01222 / 0,92237 \\ &= 0,01324 \end{aligned}$$

c. *Gain rasio* ph air

$$\begin{aligned} \text{Gainrasio}(S, \text{ ph air}) &= 0,00096 / 0,88167 \\ &= 0,00109 \end{aligned}$$

d. *Gain rasio* protein

$$\begin{aligned} \text{Gainrasio}(S, \text{ protein}) &= 0,00132 / 0,7598 \\ &= 0,00173 \end{aligned}$$

Berikut adalah catatan hasil seluruh perhitungan cabang 10 ribu yang meliputi perhitungan *entropy*, *info gain*, *split info*, *gain rasio* yang tercatat pada tabel di bawah ini :

Atribut	Jumlah Total	Optimal	Sub Optimal	<i>Entropy</i>	<i>Info Gain</i>	<i>Split Info</i>	<i>Gain Rasio</i>
<i>Class</i>	323	192	131	0,97471			
Kadar Oksigen							
<25	1101	686	415	0,97399	0,00012	0,8778	0,00013
>29	512	344	168	0,96846			
Suhu Air							
<0,013	476	284	192	0,96190	0,01222	0,9223	0,01324
>0,019	1137	742	395	0,99993			
Ph air							
<0,013	504	321	183	0,97315	0,00096	0,8816	0,00109
>0,019	1109	706	403	0,96721			
Protein							
<28	313	206	107	0,97279	0,0132	0,7598	0,00173
>30	1300	820	480	0,94770			
				0,97986			

Tabel 4.3 Hasil hitung populasi 10 ribu 1

Berdasarkan hasil perhitungan yang tertera dalam tabel di atas, diperoleh kesimpulan bahwa atribut suhu air menunjukkan gain rasio dengan nilai tertinggi.

Oleh karena itu, atribut suhu air dipilih sebagai akar berikutnya di bawah cabang populasi dengan nilai 10 ribu. Selanjutnya, dilakukan langkah evaluasi untuk menentukan apakah atribut suhu air memiliki nilai yang kosong atau terisi. Apabila terdapat nilai kosong, maka penulis akan melanjutkan perhitungan dengan memanggil kembali fungsi C4.5 untuk dilakukan pemilihan atribut yang tepat.

Tindakan selanjutnya adalah melakukan perhitungan untuk cabang atribut suhu air dengan nilai <25 dan >29, ini dilakukan karena fungsi C4.5 masih dalam tahap berjalan. Proses ini akan memberikan gambaran yang lebih mendalam terhadap pembentukan pohon keputusan dan memastikan bahwa analisis berlangsung dengan baik.

Jumlah <i>class</i> target ( <i>class</i> )	= 2 (“optimal” dan “suboptimal”)
Jumlah optimal	= 55
Jumlah suboptimal	= 54
Total keseluruhan	= 109

$$Entropy = (((-55) / 109 \times \text{Log}_2(55/109) + (((-54) / 109 \times \text{Log}_2(54/109)))) \\ = 0,99993$$

#### 1. Menghitung Nilai *Entropy* Cabang <25

##### a. *Entropy* (Kadar Oksigen)

$$Entropy(\text{Baik}) = (((-30) / 72 \times \text{Log}_2(30/72)) + (((-42) / 72 \times \text{Log}_2(42/72)))) \\ = 0,96798$$

$$Entropy(\text{Buruk}) = (((-24) / 37 \times \text{Log}_2(24/37)) + (((-24) / 37 \times \text{Log}_2(24/37)))) \\ = 0,9352$$

$$Entropy(\text{Total}) = (72 \times 0,96798 + 6 \times 0,9352) \\ = 0,96472$$

b. *Entropy* (pH Air)

$$\begin{aligned} \text{Entropy (<0,013)} &= (((-9) / 24 \times \text{Log}_2(9/24)) + (((-15) / 24 \times \text{Log}_2(15/24)))) \\ &= 0,95443 \end{aligned}$$

$$\begin{aligned} \text{Entropy (>0,019)} &= (((-40)/85 \times \text{Log}_2(40/85)) + (((-45) / 85 \times \text{Log}_2(45/85)))) \\ &= 0,99750 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Total)} &= (24 \times 0,95443 + 85 \times 0,99750) \\ &= 0,98801 \end{aligned}$$

c. *Entropy* (Protein)

$$\begin{aligned} \text{Entropy (<28)} &= (((-12) / 22 \times \text{Log}_2(12/22)) + (((-10) / 22 \times \text{Log}_2(10/22)))) \\ &= 0,99403 \end{aligned}$$

$$\begin{aligned} \text{Entropy (>30)} &= (((-43) / 87 \times \text{Log}_2(43/87)) + (((-44) / 87 \times \text{Log}_2(44/87)))) \\ &= 0,99990 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Total)} &= (22 \times 0,99403 + 87 \times 0,99990) \\ &= 0,99871 \end{aligned}$$

2. Mengitung *Gain*a. Menghitung nilai *gain* kadar oksigen

$$\begin{aligned} \text{Entropy Total} &= (0,99993 - 0,96472) \\ &= 0,03520 \end{aligned}$$

b. Menghitung nilai *gain* ph air

$$\begin{aligned} \text{Entropy Total} &= (0,99993 - 0,98801) \\ &= 0,01191 \end{aligned}$$

c. Menghitung nilai *gain* protein

$$\text{Entropy Total} = (0,99993 - 0,99871)$$

$$= 0,00122$$

5. Menghitung *split info*

a. Nilai *split info* kadar oksigen

$$\begin{aligned} \text{Entropy} &= -(72/109) \times \text{Log}_2 (72/109) - (37/109) \times \text{Log}_2 (37/109) \\ &= 0,92429 \end{aligned}$$

b. Nilai *split info* ph air

$$\begin{aligned} \text{Entropy} &= -(24/109) \times \text{Log}_2 (24/109) - (85/109) \times \text{Log}_2 (85/109) \\ &= 0,7605 \end{aligned}$$

c. Nilai *split info* protein

$$\begin{aligned} \text{Entropy} &= -(22/109) \times \text{Log}_2 (22/109) - (87/109) \times \text{Log}_2 (87/109) \\ &= 0,72558 \end{aligned}$$

6. Menghitung *Gain Rasio*

a. *Gain rasio* kadar oksigen

$$\begin{aligned} \text{Gainrasio}(S, \text{ kadar oksigen}) &= 0,03520 / 0,92429 \\ &= 0,03809 \end{aligned}$$

b. *Gain rasio* suhu air

$$\begin{aligned} \text{Gainrasio}(S, \text{ suhu air}) &= 0,01191 / 0,7065 \\ &= 0,01567 \end{aligned}$$

c. *Gain rasio* protein

$$\begin{aligned} \text{Gainrasio}(S, \text{ protein}) &= 0,00122 / 0,72558 \\ &= 0,00168 \end{aligned}$$

Hasil perhitungan dari atribur suhu cabang <25, penulis mencatat dala tabel di bawah ini

Atribut	Jumlah Total	Optimal	Sub Optimal	Entropy	Info Gain	Split Info	Gain Rasio
Class	10	4	6	0,99993			
<b>Kadar Oksigen</b>				0,96472	0,0352	0,9242	0,03809

Baik	72	30	42	0,97986			
Buruk	37	13	24	0,93526			
<b>Ph Air</b>				0,98801	0,0119	0,7065	0,01567
<0,013	24	9	15	0,95443			
>0,019	85	40	45	0,99750			
<b>Protein</b>				0,99871	0,0012	0,7255	0,00168
<28	22	12	10	0,99403			
>30	87	43	44	0,99990			

Tabel 4.4 Hasil perhitungan suhu &lt;25 1

Setelah itu, langkah selanjutnya adalah menghitung nilai dari lebih dari 29 cabang pada atribut suhu. Dalam perhitungan ini, jumlah kelas target (class) yang terdiri dari "optimal" dan "suboptimal" adalah sebanyak 2.

$$\text{Jumlah optimal} = 137$$

$$\text{Jumlah suboptimal} = 77$$

$$\text{Total keseluruhan} = 214$$

$$\text{Entropy} = (((-19) / 214 \times \text{Log}_2 (19/214) + (((-10) / 214 \times \text{Log}_2 (10/214))))$$

$$= 0,94253$$

#### 1. Menghitung Nilai Entropy Cabang <29

##### a. Entropy (Kadar Oksigen)

$$\text{Entropy(Baik)} = (((-95)/154 \times \text{Log}_2 (95/154)) + (((-42) / 154 \times \text{Log}_2(95/154))))$$

$$= 0,96021$$

$$\text{Entropy (Buruk)} = (((-42) / 60 \times \text{Log}_2(42/60)) + (((-18) /60 \times \text{Log}_2(18/60))))$$

$$= 0,88129$$

$$\text{Entropy (Total)} = (154 \times 0,9380 + 60 \times 0,8812)$$

$$= 0,9380$$

##### b. Entropy (pH Air)

$$\text{Entropy (<0,013)} = (((-43) /72 \times \text{Log}_2 (43/72)) + (((-29)/72 \times \text{Log}_2 (29/72))))$$



$$= 0,9380$$

$$\begin{aligned} \text{Entropy (>0,019)} &= (((-92)/142 \times \text{Log}_2(92/142)) + (((-48)/142 \times \text{Log}_2(48/142)))) \\ &= 0,9346 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Total)} &= (72 \times 0,9725 + 142 \times 0,9346) \\ &= 0,9473 \end{aligned}$$

c. *Entropy* (Protein)

$$\begin{aligned} \text{Entropy (<28)} &= (((-33)/49 \times \text{Log}_2(33/49)) + (((-16)/49 \times \text{Log}_2(16/49)))) \\ &= 0,9113 \end{aligned}$$

$$\begin{aligned} \text{Entropy (>30)} &= (((-104)/165 \times \text{Log}_2(104/165)) + (((-61)/165 \times \text{Log}_2(61/165)))) \\ &= 0,9504 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Total)} &= (49 \times 0,7219 + 165 \times 0,9504) \\ &= 0,8849 \end{aligned}$$

2. Mengitung *Gain*

a. Menghitung nilai *gain* kadar oksigen

$$\begin{aligned} \text{Entropy Total} &= (0,9425 - 0,9153) \\ &= 0,0044 \end{aligned}$$

b. Menghitung nilai *gain* ph air

$$\begin{aligned} \text{Entropy Total} &= (0,9425 - 0,9473) \\ &= -0,0049 \end{aligned}$$

c. Menghitung nilai *gain* protein

$$\begin{aligned} \text{Entropy Total} &= (0,9425 - 0,9414) \\ &= 0,0010 \end{aligned}$$

3. Untuk menghitung *gain rasio* dibutuhkan perhitungan. Untuk menghitung *split info* sebagai berikut :

a. Nilai *split info* kadar oksigen

$$\begin{aligned} \text{Entropy} &= -(137/29) \times \text{Log}_2(20/29) - (9/29) \times \text{Log}_2(9/29) \\ &= 0,8936 \end{aligned}$$

b. Nilai *split info* ph air

$$\begin{aligned} \text{Entropy} &= -(137/29) \times \text{Log}_2(13/29) - (16/29) \times \text{Log}_2(16/29) \\ &= 0,9923 \end{aligned}$$

c. Nilai *split info* protein

$$\begin{aligned} \text{Entropy} &= -(4/29) \times \text{Log}_2(4/29) - (25/29) \times \text{Log}_2(25/29) \\ &= 0,5788 \end{aligned}$$

4. Untuk menghitung *Gain Rasio* digunakan persamaan sebagai berikut :

a. *Gain rasio* kadar oksigen

$$\begin{aligned} \text{Gainrasio}(\text{S, kadar oksigen}) &= 0,0044 / 0,8559 \\ &= 0,00519 \end{aligned}$$

b. *Gain rasio* ph air

$$\begin{aligned} \text{Gainrasio}(\text{S, ph air}) &= 0,00291 / 0,9213 \\ &= 0,00316 \end{aligned}$$

c. *Gain rasio* protein

$$\begin{aligned} \text{Gainrasio}(\text{S, protein}) &= 0,0048 / 0,5788 \\ &= 0,0083 \end{aligned}$$

Dari hasil perhitungan yang telah dilakukan terhadap lebih dari 29 cabang atribut suhu air, data hasil perhitungan tersebut akan diisi dalam tabel yang

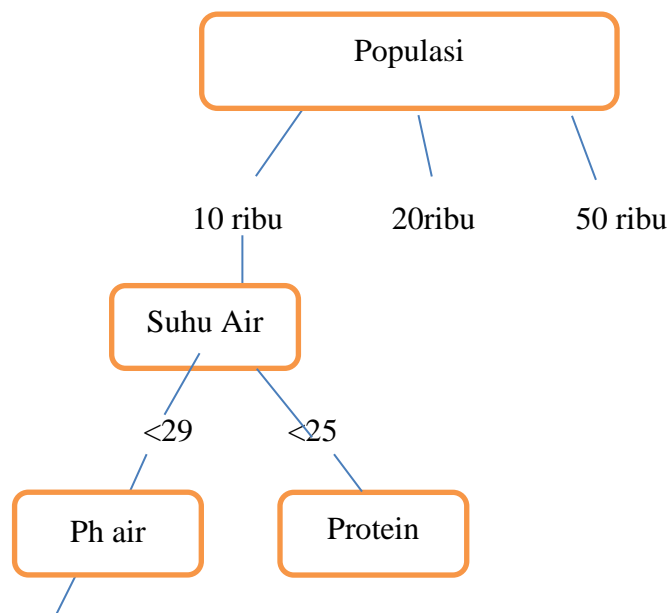
disajikan di bawah ini. Tabel ini akan memberikan informasi secara komprehensif mengenai nilai-nilai perhitungan yang telah diperoleh dari masing-masing cabang atribut suhu air, yang nantinya dapat digunakan untuk analisis lebih lanjut.

Atribut	Jumlah Total	Optimal	Sub Optimal	Entropy	Info Gain	Split Info	Gain Rasio
<b>Class</b>	214	137	77	0,9425			
<b>Kadar Oksigen</b>				0,9380	0,004	0,7219	0,00519
Baik	154	95	59	0,9602	4		
Buruk	60	42	18	0,8812			
<b>Ph Air</b>				0,9396	0,002	0,9213	0,00316
<0,013	72	43	29	0,9725	9		
>0,019	142	94	48	0,9229			
<b>Protein</b>				0,9414	0,001	0,7762	0,00134
<28	49	33	16	0,9113	0		
>30	165	104	61	0,9504			

Tabel 4.6 Hasil nilai suhu >29 1

Hasil perhitungan digambarkan dalam pohon keputusan dibawah ini dari atribut populasi pohon keputusan ini mempertimbangkan atribut populasi sebagai faktor dalam pengambilan keputusan. Pada gambar dibawah ini :

Gambar 4.2 Node populasi 1

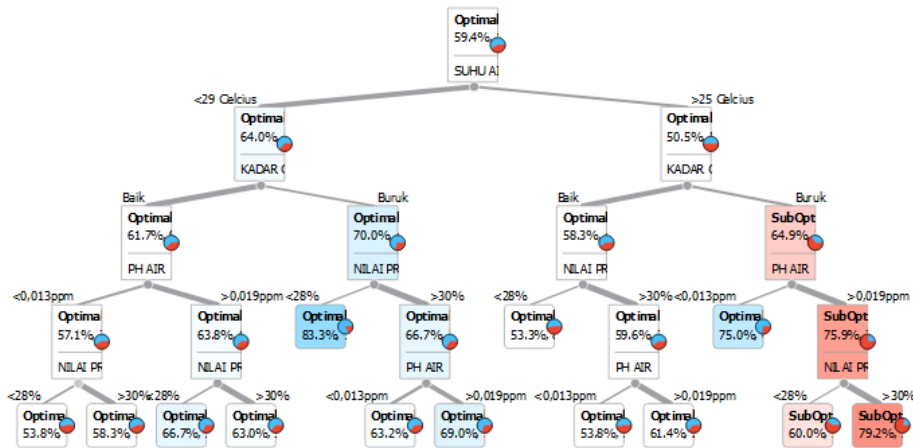




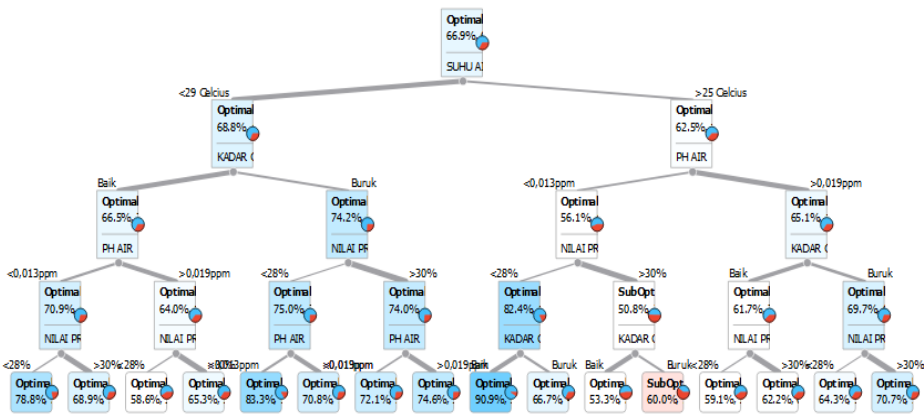
Setelah dilakukan perhitungan untuk mendapatkan nilai gain rasio pada setiap atribut yang dimasukkan ke dalam pohon (dengan kelas yang didefinisikan), langkah selanjutnya adalah menghentikan perhitungan lebih lanjut pada atribut populasi dengan nilai 20 ribu dan 50 ribu. Dalam penelitian ini, perhitungan matematika dihentikan dalam tahap pembentukan pohon keputusan, mengingat bahwa perhitungan tersebut menghasilkan nilai yang identik dengan hasil sebelumnya. Pemutusan perhitungan ini tidak akan berdampak pada hasil akhir algoritma. Perhitungan pada atribut dengan nilai 20 ribu dan 50 ribu dilanjutkan menggunakan aplikasi orange.

Untuk memberikan gambaran lengkap mengenai pohon keputusan yang terbentuk dari hasil pemodelan menggunakan aplikasi (*tools*) Orange, hasil visualisasi dapat ditemukan dalam bentuk gambar yang merepresentasikan pohon keputusan. Visualisasi ini akan memberikan pandangan tentang bagaimana algoritma mengambil keputusan berdasarkan atribut-atribut yang relevan dan perhitungan yang telah dilakukan sebelumnya.

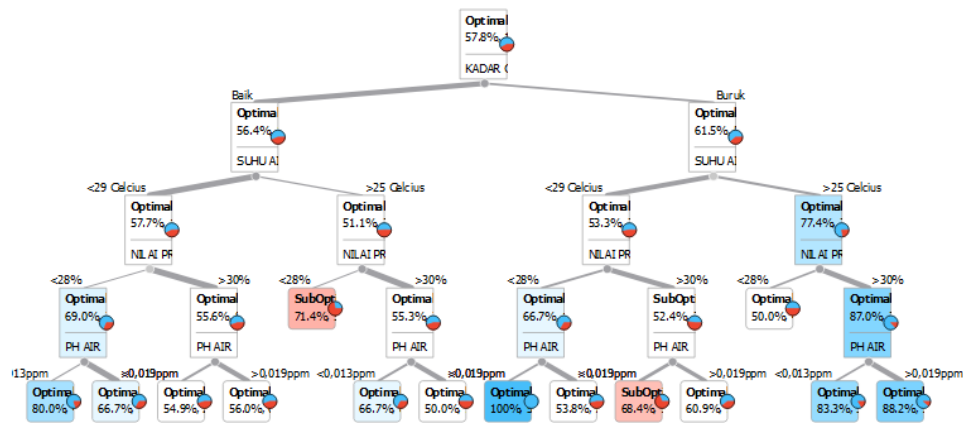
Terkait dengan penelitian, kami telah menghasilkan hasil klasifikasi yang akurat menggunakan aplikasi Orange. Hasil klasifikasi menggunakan persamaan aplikasi orange.



Gambar 4.4 Node atribut populasi 10 ribu 1



Gambar 4.5 Node atribut populasi 20 ribu 1

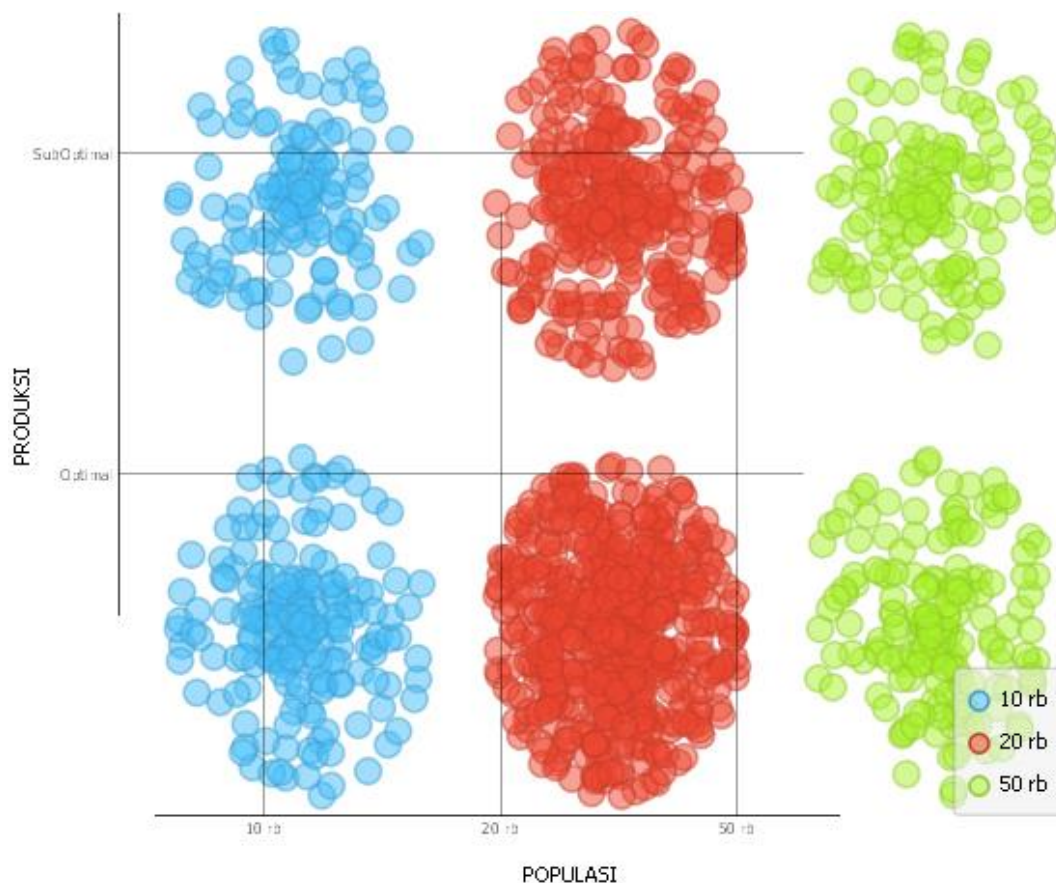


Gambar 4.6 Node atribut populasi 50 ribu 1

## 4.1.2 Pembahasan

### A Scan Plot

Terdapat grafik scanplot yang menggambarkan sebaran data dengan jelas. Titik-titik tersebar merata di seluruh bidang grafik, menunjukkan korelasi yang lemah antara variabel populasi dan produksi. Namun, ada kelompok titik yang terletak lebih dekat di sebelah kiri grafik, menyarankan adanya beberapa data outlier yang perlu dipelajari lebih lanjut. Dengan menggunakan warna biru untuk titik-titik data, grafik ini memberikan pandangan yang intuitif tentang distribusi dan pola hubungan antara dua variabel tersebut.



Gambar 4.7 Scan Plot Hasil produksi 1

#### 4.1.3 Evaluasi Model

Dalam bagian ini, peneliti akan membahas evaluasi model dari pohon keputusan yang telah diimplementasikan. Evaluasi dilakukan untuk memahami sejauh mana model mampu menghasilkan prediksi yang akurat berdasarkan metrik yang digunakan data mining dan analisa prediktif.

Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0,624	0,655	0,559	0,667	0,655	0,156

Tabel 4.6 Hasil evaluasi model 1 1

Pada tabel evaluasi beberapa model berdasarkan metrik evaluasi klasifikasi. Setiap metrik ini memberikan wawasan tentang kinerja model dalam memprediksi klasifikasi data.:

1. AUC (Area Under the Curve) Ini adalah area di bawah kurva ROC (Receiver Operating Characteristic). Kurva ROC membandingkan antara tingkat True Positive Rate (Recall) dan False Positive Rate pada berbagai ambang batas prediksi. AUC adalah ukuran yang bagus untuk menilai sejauh mana model Anda dapat membedakan antara kelas positif dan negatif. Semakin tinggi nilai AUC, semakin baik kemampuan model dalam memprediksi kelas dengan benar.
2. CA (Classification Accuracy) Ini adalah rasio prediksi yang benar (baik True Positive maupun True Negative) terhadap total jumlah data. Ini adalah ukuran dasar yang mengukur seberapa akurat model dalam melakukan prediksi secara keseluruhan.
3. F1-Score F1-score adalah harmonik rata-rata antara precision (presisi) dan recall. Precision adalah rasio True Positive terhadap total prediksi positif model, sementara recall adalah rasio True Positive terhadap total data aktual yang sebenarnya positif. F1-score memberikan keseimbangan antara precision dan recall. Nilai F1-score yang lebih tinggi menunjukkan model yang memiliki keseimbangan yang baik antara akurasi dan kemampuan untuk menemukan data positif.
4. Precision mengukur sejauh mana prediksi positif dari model adalah benar. Ini dihitung sebagai rasio True Positive terhadap total prediksi positif model. Precision yang lebih tinggi menunjukkan bahwa model jarang memberikan hasil positif yang salah.
5. Recall (juga dikenal sebagai True Positive Rate) mengukur sejauh mana model berhasil mendeteksi semua contoh positif yang sebenarnya. Ini dihitung sebagai rasio True Positive terhadap total jumlah data aktual yang positif.
6. MCC (Matthews Correlation Coefficient) MCC adalah ukuran yang menggabungkan informasi dari matriks konfusi, yaitu True Positive, True Negative, False Positive, dan False Negative. MCC menghasilkan nilai di antara -1 dan 1, di mana 1 menunjukkan prediksi sempurna, 0 menunjukkan prediksi acak, dan -1 menunjukkan ketidakcocokan sempurna antara prediksi dan hasil sebenarnya.



Dari nilai-nilai yang Anda berikan, model "Tree" memiliki AUC sekitar 0.624, CA sekitar 0.655, F1-score sekitar 0.559, precision sekitar 0.667, recall sekitar 0.655, dan MCC sekitar 0.156. Anda dapat menginterpretasikan bahwa model ini memiliki kinerja yang cukup baik dalam beberapa aspek, seperti recall yang tinggi, namun F1-score dan MCC tampaknya tidak sebanding dengan nilai-nilai lainnya.

## 1.2 Implementasi Sistem

Implementasi program merupakan hasil pembangunan sistem atau tahap pengkodean yang disesuaikan dengan tahapan rancangan sistem, implementasi digunakan untuk mengetahui tingkat keberhasilan sistem yang telah dibangun, berikut adalah hasil implementasi program :

```
# Input atribut
kadar_oksigen = 7.5
ph = 6.8
suhu = 28
populasi = 150
protein = 30

# Pohon keputusan
if kadar_oksigen >= 7 and kadar_oksigen <= 8.0:
    if ph >= 0.013 and ph <= 0.019:
        if suhu >= 25 and suhu <= 29:
            if populasi >= 100 and populasi <= 20:
                if protein >= 25 and protein <= 35:
                    produksi = "Optimal"
                else:
                    produksi = "Suboptimal"
            else:
                produksi = "Suboptimal"
        else:
            produksi = "Suboptimal"
    else:
        produksi = "Suboptimal"
else:
    produksi = "Suboptimal"

print("Prediksi produksi:", produksi)
```