

TEORI KE-2
MACHINE LEARNING
DECISION TREE

Dosen Pengampu : Nur Rosyid Muhtadai S.Kom., M.T



Disusun Oleh:

Muhammad Krisnanda Vilovan Saputra

(3323600010)

Sains Data Terapan A

POLITEKNIK ELEKTRONIKA NEGERI SURABAYA
DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER
PROGRAM STUDI SAINS DATA TERAPAN

Nomor 1.

WAKTU	PAKET	FREKWEKSI	PRIORITAS	GANGGUAN
PENDEK	BESAR	SEDANG	RENDAH	GANGGUAN
PENDEK	KECIL	RENDAH	TINGGI	GANGGUAN
PANJANG	BESAR	SEDANG	TINGGI	NORMAL
PANJANG	KECIL	TINGGI	RENDAH	NORMAL
PENDEK	BESAR	TINGGI	TINGGI	GANGGUAN
PANJANG	KECIL	RENDAH	TINGGI	GANGGUAN
PANJANG	KECIL	TINGGI	RENDAH	GANGGUAN
PANJANG	KECIL	SEDANG	RENDAH	NORMAL
PANJANG	BESAR	TINGGI	TINGGI	NORMAL
PANJANG	KECIL	SEDANG	RENDAH	GANGGUAN
PENDEK	BESAR	SEDANG	TINGGI	NORMAL
PANJANG	BESAR	RENDAH	TINGGI	NORMAL

Menentukan Node Awal

A. Menentukan entropy dari atribut ["Waktu"]

$$\text{Entropy}(S) = -P \log_2 P + -P \cdot \log_2 P.$$

Waktu	Gangguan	Total
Pendek	Gangguan (-)	3
Pendek	Normal (+)	1
Panjang	Gangguan (-)	3
Panjang	Normal (+)	5
Total		12

- $Q_1 (\text{untuk waktu} = \text{pendek}) = -\frac{3}{4} \cdot \log_2 \frac{3}{4} - \frac{1}{4} \cdot \log_2 \frac{1}{4} = \underline{0,81127}$
- $Q_2 (\text{untuk waktu} = \text{panjang}) = -\frac{3}{8} \cdot \log_2 \frac{3}{8} - \frac{5}{8} \cdot \log_2 \frac{5}{8} = \underline{0,9544}$
- $\text{Entropy Waktu} = \frac{4}{12} \cdot (0,81127) + \frac{8}{12} \cdot (0,9544) = \underline{0,9067}$

B. Menentukan entropy dari atribut ["Paket"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Paket	Gangguan	Total
Kecil	Gangguan (-)	2
Kecil	Normal (+)	4
Besar	Gangguan (-)	4
Besar	Normal (+)	2
Total		12

- Q_1 (untuk paket = kecil) = $-\frac{2}{6} \cdot \log_2 \frac{2}{6} - \frac{4}{6} \cdot \log_2 \frac{4}{6} = \underline{0,9182}$
- Q_2 (untuk paket = besar) = $-\frac{4}{6} \cdot \log_2 \frac{4}{6} - \frac{2}{6} \cdot \log_2 \frac{2}{6} = \underline{0,9182}$
- Entropy Paket = $\frac{6}{12} \cdot (0,9182) + \frac{6}{12} \cdot (0,9182) = \underline{0,9182}$

C. Menentukan entropy dari atribut ["Frekwensi"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Frekwensi	Gangguan	Total
Rendah	Gangguan (-)	2
Rendah	Normal (+)	1
Sedang	Gangguan (-)	2
Sedang	Normal (+)	3
Tinggi	Gangguan (-)	2
Tinggi	Normal (+)	2
Total		12

- Q_1 (untuk frekwensi= rendah) = $-\frac{2}{3} \cdot \log_2 \frac{2}{3} - \frac{1}{3} \cdot \log_2 \frac{1}{3} = \underline{0,9182}$
- Q_2 (untuk frekwensi = sedang) = $-\frac{2}{5} \cdot \log_2 \frac{2}{5} - \frac{3}{5} \cdot \log_2 \frac{3}{5} = \underline{0,9709}$
- Q_3 (untuk frekwensi = tinggi) = $-\frac{2}{4} \cdot \log_2 \frac{2}{4} - \frac{2}{4} \cdot \log_2 \frac{2}{4} = \underline{1}$
- Entropy Frewensi = $\frac{3}{12} \cdot (0,9182) + \frac{5}{12} \cdot (0,9709) + \frac{4}{12} \cdot (1) = \underline{0,9674}$

D. Menentukan entropy dari atribut ["Prioritas"]

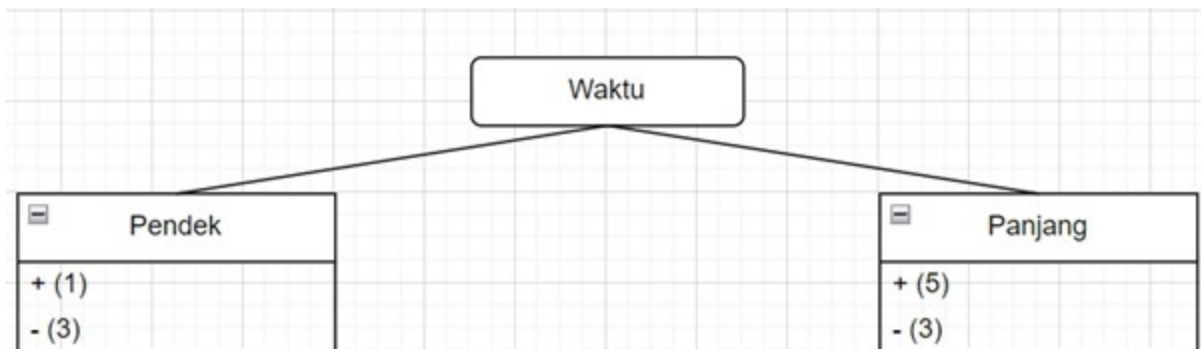
$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Prioritas	Gangguan	Total
Rendah	Gangguan (-)	3
Rendah	Normal (+)	2
Tinggi	Gangguan (-)	3
Tinggi	Normal (+)	4
Total		12

- Q_1 (untuk prioritas= rendah) $= -\frac{3}{5} \cdot \log_2 \frac{3}{5} - \frac{2}{5} \cdot \log_2 \frac{2}{5} = \underline{0,9709}$
- Q_2 (untuk prioritas= tinggi) $= -\frac{3}{7} \cdot \log_2 \frac{3}{7} - \frac{4}{7} \cdot \log_2 \frac{4}{7} = \underline{0,985573}$
 $= \underline{0,986}$
- Entropy Frewensi $= \frac{5}{12} \cdot (0,9709) + \frac{7}{12} \cdot (0,986) = \underline{0,97937} = \underline{0,9794}$

Dapat disimpulkan bahwa node awal dari tree tersebut adalah ["Waktu"]

Hasil Tree untuk Node Awal



Menentukan Leaf Node

Data training untuk waktu = pendek

Paket	Frekwensi	Prioritas	Gangguan
Besar	Sedang	Rendah	Gangguan (-)
Kecil	Rendah	Tinggi	Gangguan (-)
Besar	Tinggi	Tinggi	Gangguan (-)
Besar	Sedang	Tinggi	Normal (+)

A. Menentukan entropy dari atribut ["Paket"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Paket	Gangguan	Total
Kecil	Gangguan (-)	1
Kecil	Normal (+)	0
Besar	Gangguan (-)	2
Besar	Normal (+)	1
Total		4

- Q_1 (untuk paket = kecil) = $-\frac{1}{1} \cdot \log_2 \frac{1}{1} - \frac{0}{1} \cdot \log_2 \frac{0}{1} = 0$
- Q_2 (untuk paket = besar) = $-\frac{2}{3} \cdot \log_2 \frac{2}{3} - \frac{1}{3} \cdot \log_2 \frac{1}{3} = 0,9183$
- Entropy Paket = $\frac{1}{4} \cdot (0) + \frac{3}{4} \cdot (0,9183) = 0,6887 = 0,69$

B. Menentukan entropy dari atribut ["Frekwensi"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Frekwensi	Gangguan	Total
Rendah	Gangguan (-)	1
Rendah	Normal (+)	0
Sedang	Gangguan (-)	1
Sedang	Normal (+)	1
Tinggi	Gangguan (-)	1
Tinggi	Normal (+)	0
Total		4

- Q_1 (untuk frekwensi= rendah) = $-\frac{1}{1} \cdot \log_2 \frac{1}{1} - \frac{0}{1} \cdot \log_2 \frac{0}{1} = 0$
- Q_2 (untuk frekwensi = sedang) = $-\frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{1}{2} \cdot \log_2 \frac{1}{2} = 1$
- Q_3 (untuk frekwensi = tinggi) = $-\frac{1}{1} \cdot \log_2 \frac{1}{1} - \frac{0}{1} \cdot \log_2 \frac{0}{1} = 0$
- Entropy Frekwensi = $\frac{1}{4} \cdot (0) + \frac{2}{4} \cdot (1) + \frac{1}{4} \cdot (0) = 0,5$

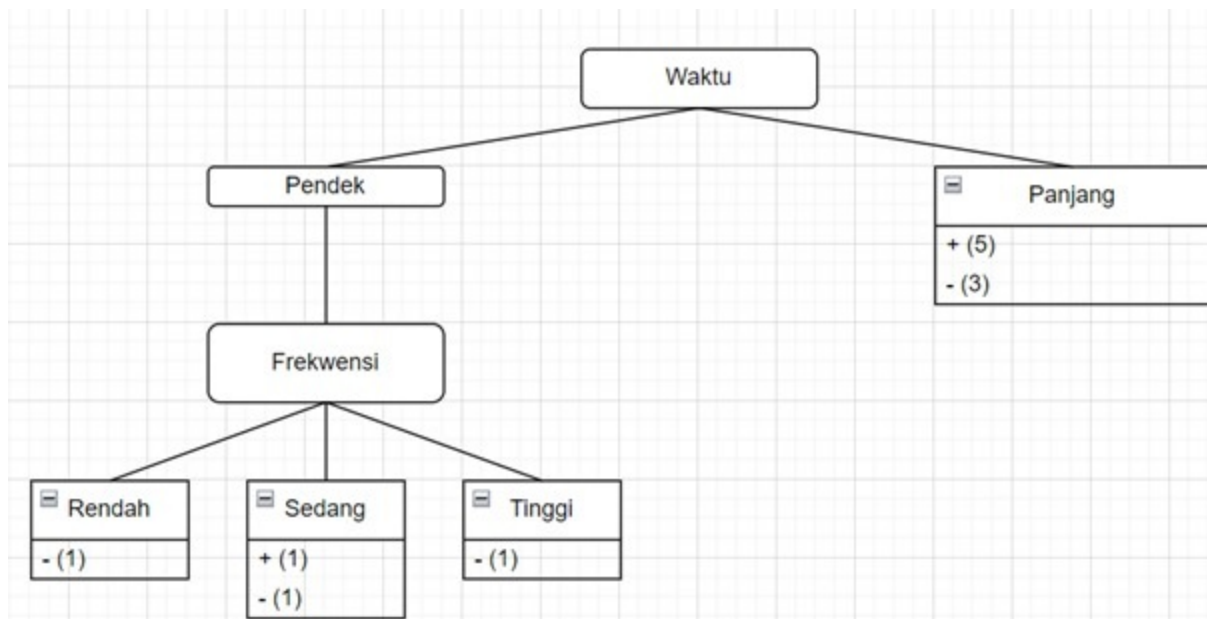
C. Menentukan entropy dari atribut ["Prioritas"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Prioritas	Gangguan	Total
Rendah	Gangguan (-)	1
Rendah	Normal (+)	0
Tinggi	Gangguan (-)	2
Tinggi	Normal (+)	1
Total		4

- Q_1 (untuk prioritas= rendah) = $-\frac{1}{1} \cdot \log_2 \frac{1}{1} - \frac{0}{1} \cdot \log_2 \frac{0}{1} = 0$
- Q_2 (untuk prioritas = sedang) = $-\frac{2}{3} \cdot \log_2 \frac{2}{3} - \frac{1}{3} \cdot \log_2 \frac{1}{3} = 0,92$
- Entropy Prioritas = $\frac{1}{4} \cdot (0) + \frac{3}{4} \cdot (0,92) = 0,69$

Hasil Tree untuk Leaf Waktu = Pendek



D. Menentukan entropy dari atribut [“Frekwensi = Sedang” & “Waktu = Pendek”]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Paket	Gangguan	Total
Besar	Gangguan (-)	1
Besar	Normal (+)	1
Total		2

$$\bullet \quad Q_1 (\text{untuk paket}) = -\frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{1}{2} \cdot \log_2 \frac{1}{2} = 1$$

$$\bullet \quad \text{Entropy Paket} = \frac{2}{2} \cdot (1) = 1$$

E. Menentukan entropy dari atribut [“Frekwensi = Sedang” & “Waktu = Pendek”]

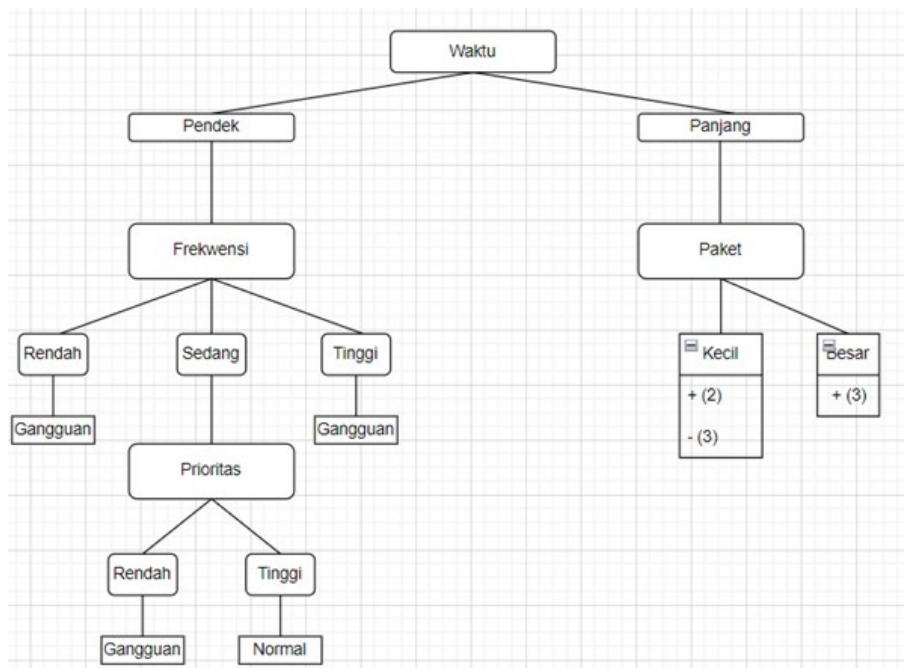
$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Prioritas	Gangguan	Total
Rendah	Gangguan (-)	1
Rendah	Normal (+)	0
Tinggi	Gangguan (-)	0
Tinggi	Normal (+)	1
Total		2

$$\bullet \quad Q_1 (\text{untuk prioritas = rendah}) = -\frac{1}{1} \cdot \log_2 \frac{1}{1} - \frac{0}{1} \cdot \log_2 \frac{0}{1} = 0$$

$$\bullet \quad Q_2 (\text{untuk prioritas = tinggi}) = -\frac{0}{1} \cdot \log_2 \frac{0}{1} - \frac{1}{1} \cdot \log_2 \frac{1}{1} = 0$$

$$\bullet \quad \text{Entropy Prioritas} = 0$$



Data training untuk waktu = panjang

Paket	Frekwensi	Prioritas	Gangguan
Besar	Sedang	Tinggi	Normal
Kecil	Tinggi	Rendah	Normal
Kecil	Rendah	Tinggi	Gangguan
Kecil	Tinggi	Rendah	Gangguan
Kecil	Sedang	Rendah	Normal
Besar	Tinggi	Tinggi	Normal
Kecil	Sedang	Rendah	Gangguan
Besar	Rendah	Tinggi	Normal

A. Menentukan entropy dari atribut ["Paket"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Paket	Gangguan	Total
Kecil	Gangguan (-)	3
Kecil	Normal (+)	2
Besar	Gangguan (-)	0
Besar	Normal (+)	3
Total		8

- Q_1 (untuk paket = kecil) = $-\frac{3}{5} \cdot \log_2 \frac{3}{5} - \frac{2}{5} \cdot \log_2 \frac{2}{5} = \underline{0,9709}$
- Q_2 (untuk paket = besar) = $-\frac{0}{3} \cdot \log_2 \frac{0}{3} - \frac{3}{3} \cdot \log_2 \frac{3}{3} = \underline{0}$
- Entropy Paket = $\frac{5}{8} \cdot (0,9709) + \frac{3}{8} \cdot (0) = \underline{0,6068}$

B. Menentukan entropy dari atribut ["Frekwensi"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Paket	Gangguan	Total
Rendah	Gangguan (-)	1
Rendah	Normal (+)	1
Sedang	Gangguan (-)	1
Sedang	Normal (+)	2
Tinggi	Gangguan (-)	1
Tinggi	Normal (+)	2
Total		8

- Q_1 (untuk frekwensi= rendah) = $-\frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{1}{2} \cdot \log_2 \frac{1}{2} = \underline{1}$
- Q_2 (untuk frekwensi= sedang) = $-\frac{1}{3} \cdot \log_2 \frac{1}{3} - \frac{2}{3} \cdot \log_2 \frac{2}{3} = \underline{0,9183}$
- Q_3 (untuk frekwensi= tinggi) = $-\frac{1}{3} \cdot \log_2 \frac{1}{3} - \frac{2}{3} \cdot \log_2 \frac{2}{3} = \underline{0,9183}$

- Entropy Frekwensi = $\frac{2}{8} \cdot (1) + \frac{3}{8} \cdot (0,9183) + \frac{3}{8} \cdot (0,9183) = \underline{0,9392}$

C. Menentukan entropy dari atribut ["Prioritas"]

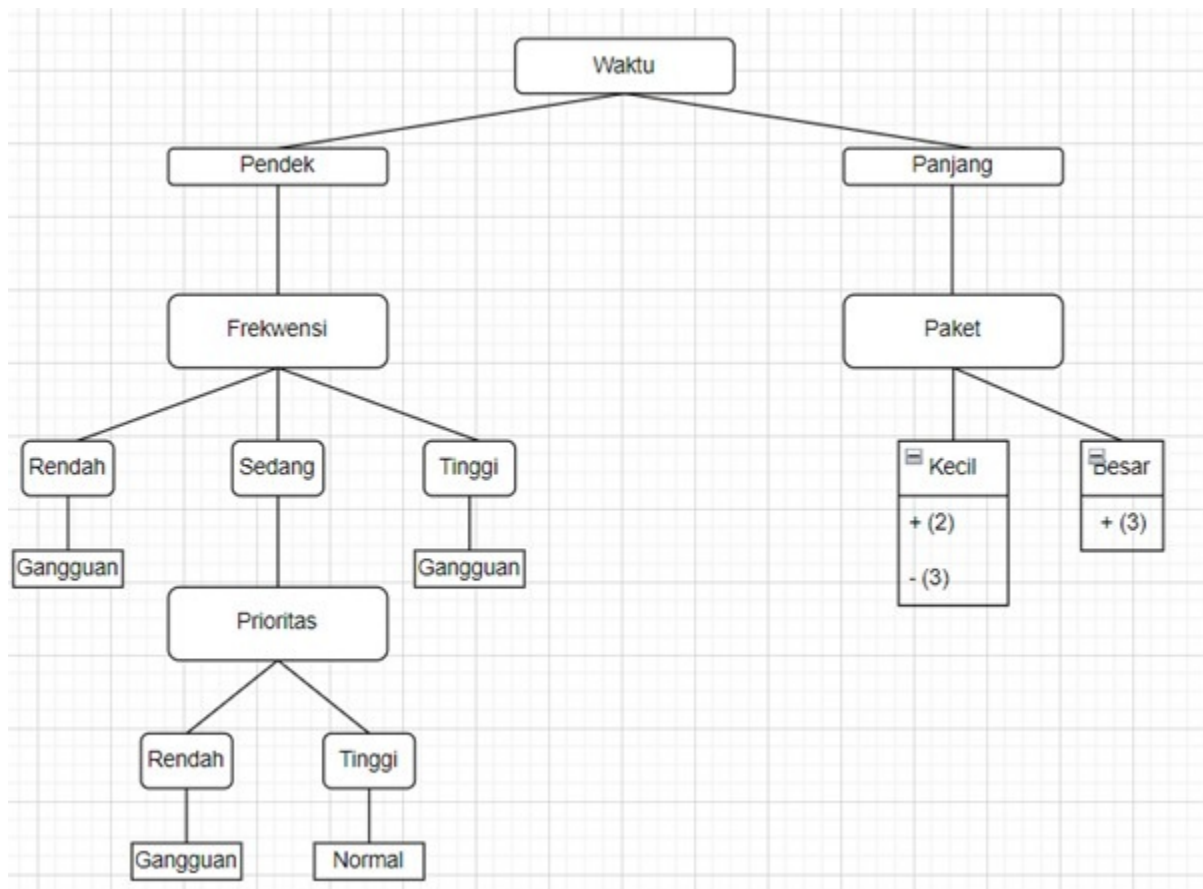
$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Prioritas	Gangguan	Total
Rendah	Gangguan (-)	2
Rendah	Normal (+)	2
Tinggi	Gangguan (-)	1
Tinggi	Normal (+)	3
Total		8

- Q_1 (untuk prioritas = rendah) = $-\frac{2}{4} \cdot \log_2 \frac{2}{4} - \frac{2}{4} \cdot \log_2 \frac{2}{4} = \underline{1}$

- Q_2 (untuk prioritas = tinggi) = $-\frac{1}{4} \cdot \log_2 \frac{1}{4} - \frac{3}{4} \cdot \log_2 \frac{3}{4} = \underline{0,81125}$
 $= \underline{0,8113}$

- Entropy Prioritas = $\frac{4}{8} \cdot (1) + \frac{4}{8} \cdot (0,8113) = \underline{0,90565} = \underline{0,9057}$



D. Menentukan entropy dari atribut [“Waktu = Panjang & Paket = Kecil”]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Frekwensi	Gangguan	Total
Rendah	Gangguan (-)	1
Rendah	Normal (+)	0
Sedang	Gangguan (-)	1
Sedang	Normal (+)	1
Tinggi	Gangguan (-)	1
Tinggi	Normal (+)	1
Total		5

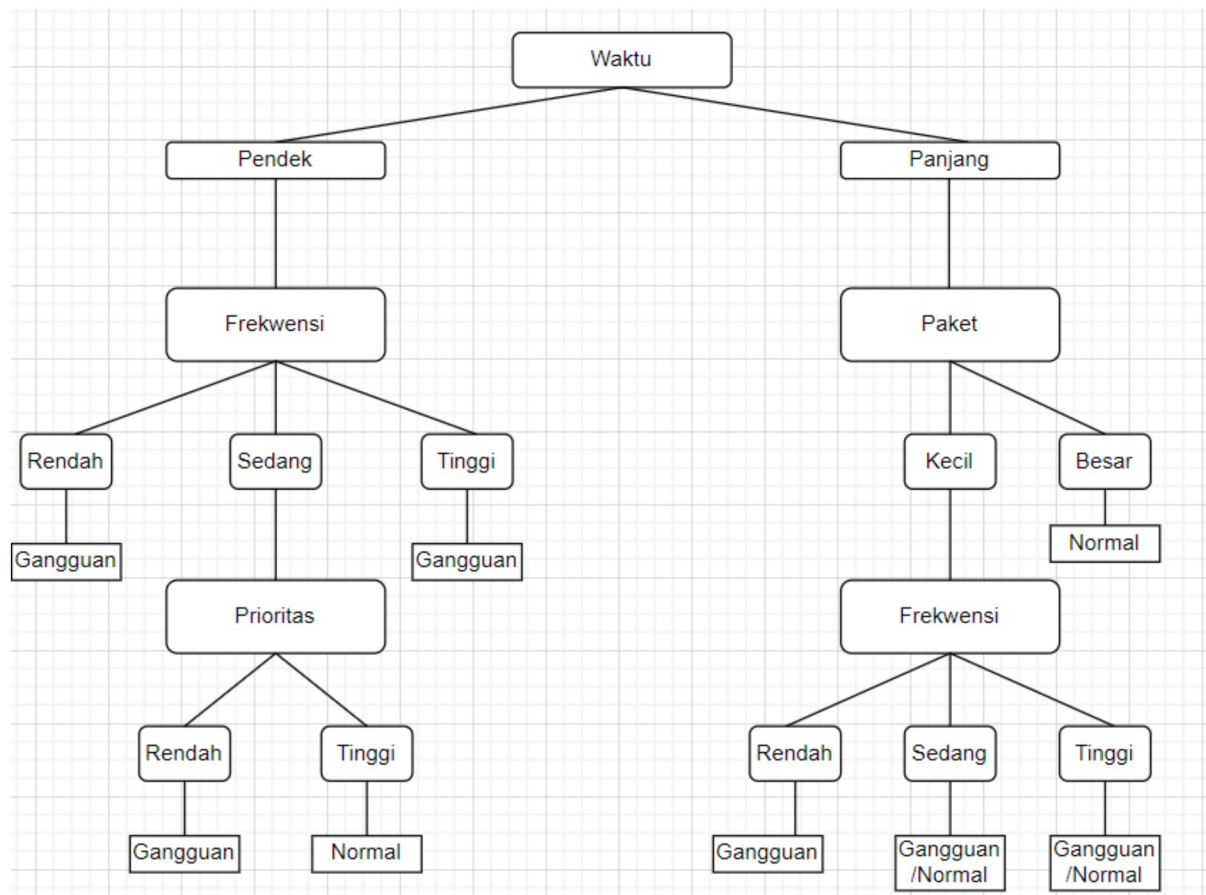
- Q_1 (untuk frekwensi= rendah) = $-\frac{1}{1} \cdot \log_2 \frac{1}{1} - \frac{0}{1} \cdot \log_2 \frac{0}{1} = \underline{0}$
- Q_2 (untuk frekwensi= sedang) = $-\frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{1}{2} \cdot \log_2 \frac{1}{2} = \underline{1}$
- Q_3 (untuk frekwensi= tinggi) = $-\frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{1}{2} \cdot \log_2 \frac{1}{2} = \underline{1}$
- Entropy Frekwensi = $\frac{1}{5} \cdot (0) + \frac{2}{5} \cdot (1) + \frac{2}{5} \cdot (1) = \underline{0,8}$

E. Menentukan entropy dari atribut [“Waktu = Panjang & Paket = Kecil”]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Prioritas	Gangguan	Total
Rendah	Gangguan (-)	2
Rendah	Normal (+)	2
Tinggi	Gangguan (-)	0
Tinggi	Normal (+)	1
Total		5

- Q_1 (untuk prioritas = rendah) = $-\frac{2}{4} \cdot \log_2 \frac{2}{4} - \frac{2}{4} \cdot \log_2 \frac{2}{4} = \underline{1}$
- Q_2 (untuk prioritas= tinggi) = $-\frac{0}{1} \cdot \log_2 \frac{0}{1} - \frac{1}{1} \cdot \log_2 \frac{1}{1} = \underline{0}$
- Entropy prioritas = $\frac{4}{5} \cdot (1) + \frac{1}{5} \cdot (0) = \underline{0,8}$



Rule Decision Tree

1. R1 = IF Waktu = pendek ^ Frekwensi = rendah THEN Gangguan = gangguan.
2. R2 = IF Waktu = pendek ^ Frekwensi = sedang ^ Prioritas = rendah THEN Gangguan = gangguan.
3. R3 = IF Waktu = pendek ^ Frekwensi = sedang ^ Prioritas = tinggi THEN Gangguan = normal.
4. R4 = IF Waktu = pendek ^ Frekwensi = Tinggi THEN Gangguan = gangguan.
5. R5 = IF Waktu = panjang ^ Paket = kecil ^ Frekwensi = rendah THEN Gangguan = gangguan.
6. R6 = IF Waktu = panjang ^ Paket = kecil ^ Frekwensi = sedang THEN Gangguan = gangguan.
7. R7 = IF Waktu = panjang ^ Paket = kecil ^ Frekwensi = tinggi THEN Gangguan = gangguan.
8. R8 = IF Waktu = Panjang ^ Paket = besar THEN Gangguan = normal.

Menentukan Error

WAKTU	PAKET	FREKWEKSI	PRIORITAS	GANGGUAN
PENDEK	BESAR	SEDANG	RENDAH	GANGGUAN
PENDEK	KECIL	RENDAH	TINGGI	GANGGUAN
PANJANG	BESAR	SEDANG	TINGGI	NORMAL
PANJANG	KECIL	TINGGI	RENDAH	NORMAL
PENDEK	BESAR	TINGGI	TINGGI	GANGGUAN
PANJANG	KECIL	RENDAH	TINGGI	GANGGUAN
PANJANG	KECIL	TINGGI	RENDAH	GANGGUAN
PANJANG	KECIL	SEDANG	RENDAH	NORMAL
PANJANG	BESAR	TINGGI	TINGGI	NORMAL
PANJANG	KECIL	SEDANG	RENDAH	GANGGUAN
PENDEK	BESAR	SEDANG	TINGGI	NORMAL
PANJANG	BESAR	RENDAH	TINGGI	NORMAL

Error pada data :

- Baris ke 4
- Baris ke 8

Error ada 2 dari 12 data maka : $\frac{2}{12} \times 100\% = 16,67\%$

Nomor 2.

USIA	KELAMIN	MEROKOK	OLAHRAGA	JANTUNG
TUA	PRIA	TIDAK	YA	TIDAK
TUA	PRIA	YA	YA	TIDAK
MUDA	PRIA	YA	TIDAK	TIDAK
TUA	PRIA	TIDAK	TIDAK	TIDAK
MUDA	WANITA	TIDAK	TIDAK	YA
MUDA	PRIA	TIDAK	YA	YA
MUDA	PRIA	TIDAK	YA	TIDAK
TUA	WANITA	TIDAK	TIDAK	YA
MUDA	PRIA	YA	TIDAK	TIDAK
TUA	PRIA	YA	TIDAK	TIDAK
MUDA	PRIA	YA	YA	YA
TUA	PRIA	YA	TIDAK	TIDAK
MUDA	PRIA	TIDAK	TIDAK	TIDAK
TUA	PRIA	TIDAK	YA	TIDAK
MUDA	PRIA	YA	TIDAK	TIDAK

Menentukan Node Awal

A. Menentukan entropy dari atribut [“Usia”]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Waktu	Jantung	Total
Muda	Ya (+)	3
Muda	Tidak (-)	5
Tua	Ya (+)	1
Tua	Tidak (-)	6
Total		15

- $Q_1 (\text{untuk usia} = \text{muda}) = -\frac{3}{8} \cdot \log_2 \frac{3}{8} - \frac{5}{8} \cdot \log_2 \frac{5}{8} = \underline{0,9551}$
- $Q_2 (\text{untuk uis a= tua}) = -\frac{1}{7} \cdot \log_2 \frac{1}{7} - \frac{6}{7} \cdot \log_2 \frac{6}{7} = \underline{0,63}$
- $\text{Entropy Usia} = \frac{8}{15} \cdot (0,9551) + \frac{7}{15} \cdot (0,63) = \underline{0,8027}$

B. Menentukan entropy dari atribut ["Kelamin"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Kelamin	Jantung	Total
Pria	Ya (+)	2
Pria	Tidak (-)	11
Wanita	Ya (+)	2
Wanita	Tidak (-)	0
Total		15

- $Q_1 (\text{untuk kelamin} = \text{pria}) = -\frac{2}{13} \cdot \log_2 \frac{2}{13} - \frac{11}{13} \cdot \log_2 \frac{11}{13} = \underline{0,61}$
- $Q_2 (\text{untuk kelamin} = \text{wanita}) = -\frac{2}{2} \cdot \log_2 \frac{2}{2} - \frac{0}{2} \cdot \log_2 \frac{0}{2} = \underline{0}$
- $\text{Entropy Kelamin} = \frac{8}{15} \cdot (0,9551) + \frac{7}{15} \cdot (0,63) = \underline{0,52}$

C. Menentukan entropy dari atribut ["Merokok"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Merokok	Jantung	Total
Ya	Ya (+)	1
Ya	Tidak (-)	6
Tidak	Ya (+)	3
Tidak	Tidak (-)	5
Total		15

- $Q_1 (\text{untuk merokok} = \text{ya}) = -\frac{1}{7} \cdot \log_2 \frac{1}{7} - \frac{6}{7} \cdot \log_2 \frac{6}{7} = \underline{0,586}$
- $Q_2 (\text{untuk merokok} = \text{tidak}) = -\frac{3}{8} \cdot \log_2 \frac{3}{8} - \frac{5}{8} \cdot \log_2 \frac{5}{8} = \underline{0,9551}$
- $\text{Entropy Merokok} = \frac{7}{15} \cdot (0,586) + \frac{8}{15} \cdot (0,9551) = \underline{0,7828}$

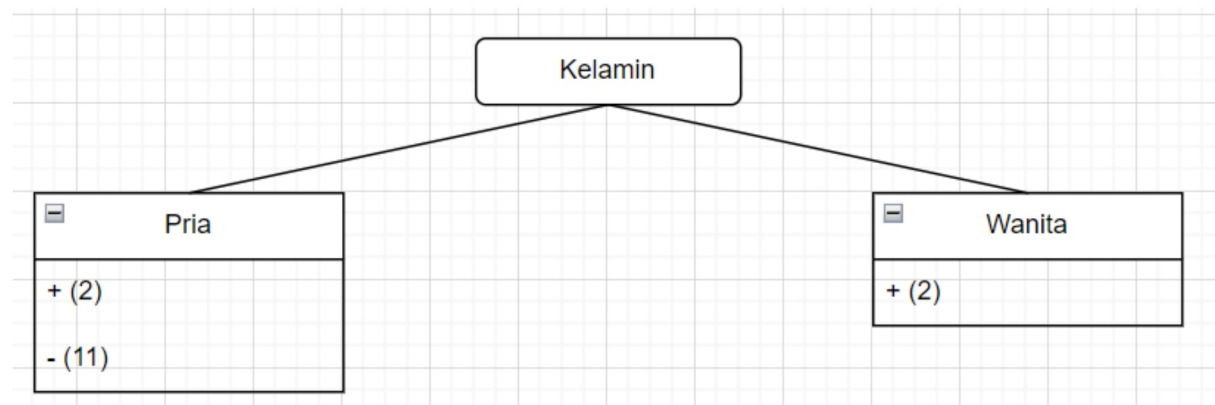
D. Menentukan entropy dari atribut ["Olahraga"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Olahraga	Jantung	Total
Ya	Ya (+)	2
Ya	Tidak (-)	4
Tidak	Ya (+)	2
Tidak	Tidak (-)	7
Total		15

- Q_1 (untuk olahraga = ya) = $-\frac{2}{6} \cdot \log_2 \frac{2}{6} - \frac{4}{6} \cdot \log_2 \frac{4}{6} = \underline{0,9183}$
- Q_2 (untuk olahraga = tidak) = $-\frac{2}{9} \cdot \log_2 \frac{2}{9} - \frac{7}{9} \cdot \log_2 \frac{7}{9} = \underline{0,7608}$
- Entropy Olahraga = $\frac{6}{15} \cdot (0,9183) + \frac{9}{15} \cdot (0,7608) = \underline{0,8238}$

Dapat disimpulkan bahwa node awal dari tree tersebut adalah [**"Kelamin"**]



Menentukan Leaf Node

Data training untuk kelamin = pria

Usia	Merokok	Olahtaga	Jantung
Tua	Tidak	Ya	Tidak
Tua	Ya	Ya	Tidak
Muda	Ya	Tidak	Tidak
Tua	Tidak	Tidak	Tidak
Muda	Tidak	Ya	Ya
Muda	Tidak	Ya	Tidak
Muda	Ya	Tidak	Tidak
Tua	Ya	Tidak	Tidak
Muda	Ya	Ya	Ya
Tua	Ya	Tidak	Tidak
Muda	Tidak	Tidak	Tidak
Tua	Ya	Tidak	Ya
Muda	Ya	Tidak	Tidak

E. Menentukan entropy dari atribut ["Usia"]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Usia	Gangguan	Total
Muda	Gangguan (-)	2
Muda	Normal (+)	5
Tua	Gangguan (-)	0
Tua	Normal (+)	6
Total		13

- $Q_1 (\text{untuk usia} = \text{muda}) = -\frac{2}{7} \cdot \log_2 \frac{2}{7} - \frac{5}{7} \cdot \log_2 \frac{5}{7} = \underline{0,86}$
- $Q_2 (\text{untuk usia} = \text{tua}) = -\frac{0}{6} \cdot \log_2 \frac{0}{6} - \frac{6}{6} \cdot \log_2 \frac{6}{6} = \underline{0}$
- $\text{Entropy Usia} = \frac{7}{13} \cdot (0,86) + \frac{6}{13} \cdot (0) = \underline{0,46}$

F. Menentukan entropy dari atribut [“Merokok”]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Merokok	Gangguan	Total
Ya	Ya (+)	1
Ya	Tidak (-)	6
Tidak	Ya (+)	1
Tidak	Tidak (-)	5
Total		13

- Q_1 (untuk merokok = ya) = $-\frac{1}{7} \cdot \log_2 \frac{1}{7} - \frac{6}{7} \cdot \log_2 \frac{6}{7} = \underline{0,586}$
- Q_2 (untuk merokok = tidak) = $-\frac{1}{6} \cdot \log_2 \frac{1}{6} - \frac{5}{6} \cdot \log_2 \frac{5}{6} = \underline{0,6501}$
- Entropy Merokok = $\frac{7}{13} \cdot (0,586) + \frac{6}{13} \cdot (0,6501) = \underline{0,6156}$

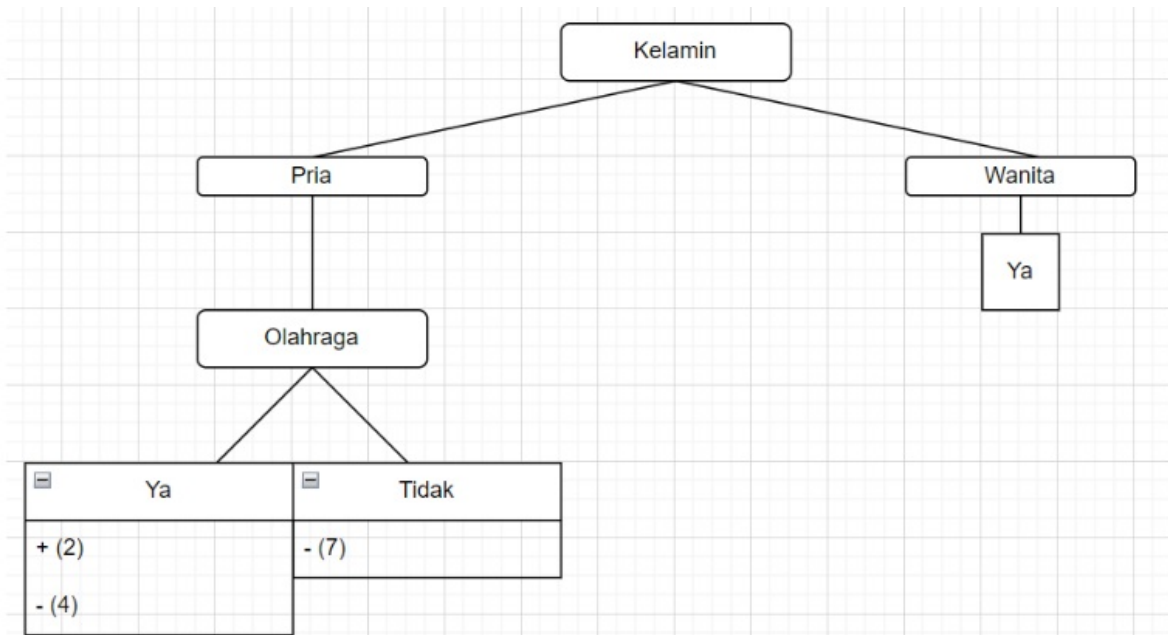
G. Menentukan entropy dari atribut [“Olahraga”]

$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Olahraga	Jantung	Total
Ya	Ya (+)	2
Ya	Tidak (-)	4
Tidak	Ya (+)	0
Tidak	Tidak (-)	7
Total		13

- Q_1 (untuk olahraga = ya) = $-\frac{2}{6} \cdot \log_2 \frac{2}{6} - \frac{4}{6} \cdot \log_2 \frac{4}{6} = \underline{0,9183}$
- Q_2 (untuk olahraga = tidak) = $-\frac{0}{7} \cdot \log_2 \frac{0}{7} - \frac{7}{7} \cdot \log_2 \frac{7}{7} = \underline{0}$
- Entropy Olahraga = $\frac{6}{13} \cdot (0,9183) + \frac{7}{13} \cdot (0) = \underline{0,4246}$

Dapat disimpulkan bahwa node awal dari tree tersebut adalah [“Olahraga”]



H. Menentukan entropy dari atribut ["Olahraga = Ya"]

Entropy(S) = $-P \log_2 P_+ - P_- \log_2 P_-$

Usia	Gangguan	Total
Muda	Gangguan (-)	2
Muda	Normal (+)	1
Tua	Gangguan (-)	0
Tua	Normal (+)	3
Total		6

- Q_1 (untuk usia = muda) = $-\frac{2}{3} \cdot \log_2 \frac{2}{3} - \frac{1}{3} \cdot \log_2 \frac{1}{3} = \underline{0,91}$
- Q_2 (untuk usia = tua) = $-\frac{0}{3} \cdot \log_2 \frac{0}{3} - \frac{3}{3} \cdot \log_2 \frac{3}{3} = \underline{0}$
- Entropy Usia = $\frac{3}{6} \cdot (0,91) + \frac{3}{6} \cdot (0) = \underline{0,455}$

I. Menentukan entropy dari atribut ["Merokok"]

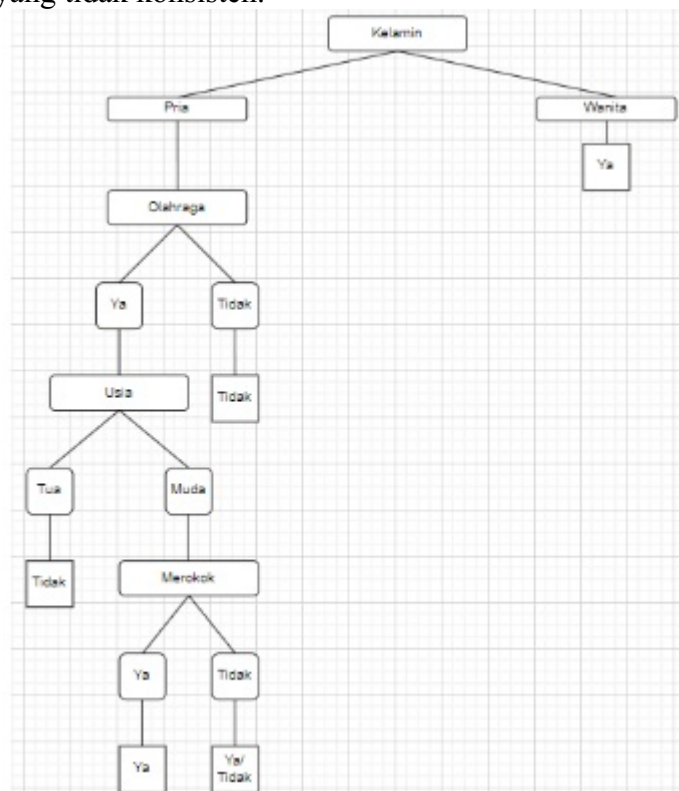
$$\text{Entropy}(S) = -P \log_2 P_+ - P_- \log_2 P_-$$

Merokok	Gangguan	Total
Ya	Ya (+)	1
Ya	Tidak (-)	1
Tidak	Ya (+)	1
Tidak	Tidak (-)	3
Total		6

- Q_1 (untuk merokok = ya) = $-\frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{1}{2} \cdot \log_2 \frac{1}{2} = 1$
- Q_2 (untuk merokok= tidak) = $-\frac{1}{4} \cdot \log_2 \frac{1}{4} - \frac{3}{4} \cdot \log_2 \frac{3}{4} = 0,81$
- Entropy Merokok = $\frac{2}{6} \cdot (1) + \frac{4}{6} \cdot (0,81) = 0,87$

Hasil Entropy terkecilnya adalah Atribut Usia

Dikarenakan hanya tersisa 1 atribut maka langsung dimasukkan, karena memiliki data yang tidak konsisten.



Rule Decision Tree

1. R1 = IF Kelamin = wanita THEN Jantung = ya.
2. R2 = IF Kelamin = pria ^ Olahraga = tidak THEN Jantung = tidak.
3. R3 = IF Kelamin = Pria ^ Olahraga = ya ^ Usia = tua THEN Jantung = tidak.
4. R4 = IF Kelamin = Pria ^ Olahraga = ya ^ Usia = muda ^ Merokok = ya THEN
5. Jantung = ya.
6. R5 = IF Kelamin = Pria ^ Olahraga = ya ^ Usia = muda ^ Merokok = tidak
7. THEN Jantung = tidakl

Menentukan Error

USIA	KELAMIN	MEROKOK	OLAHRAGA	JANTUNG
TUA	PRIA	TIDAK	YA	TIDAK
TUA	PRIA	YA	YA	TIDAK
MUDA	PRIA	YA	TIDAK	TIDAK
TUA	PRIA	TIDAK	TIDAK	TIDAK
MUDA	WANITA	TIDAK	TIDAK	YA
MUDA	PRIA	TIDAK	YA	YA
MUDA	PRIA	TIDAK	YA	TIDAK
TUA	WANITA	TIDAK	TIDAK	YA
MUDA	PRIA	YA	TIDAK	TIDAK
TUA	PRIA	YA	TIDAK	TIDAK
MUDA	PRIA	YA	YA	YA
TUA	PRIA	YA	TIDAK	TIDAK
MUDA	PRIA	TIDAK	TIDAK	TIDAK
TUA	PRIA	TIDAK	YA	TIDAK
MUDA	PRIA	YA	TIDAK	TIDAK

Error pada data :

1. Baris ke 6

- Error ada 1 dari 15 data : $\frac{1}{15} \times 100\% = 6,67\%$

Nomor 3.

Nama	Cuaca	Angin	Temperatur	Main
Ali	cerah	keras	panas	tidak
Budi	cerah	lambat	panas	ya
Heri	berawan	keras	sedang	tidak
Irma	hujan	keras	dingin	tidak
Diman	cerah	lambat	dingin	ya

Entropy Data Cuaca :

$$P+ = \frac{2}{5}$$

$$P- = \frac{3}{5}$$

$$S = -\frac{2}{5} \cdot \log_2 \frac{2}{5} - \frac{3}{5} \cdot \log_2 \frac{3}{5} = \underline{0,97}$$

Nomor 4.

Nama	Usia	Berat	Kelamin	Hipertensi
Ali	muda	overweight	pria	ya
Edi	muda	underweight	pria	tidak
Annie	muda	average	wanita	tidak
Budiman	tua	overweight	pria	tidak
Herman	tua	overweight	pria	ya
Didi	muda	underweight	pria	tidak
Rina	tua	overweight	wanita	ya
Gatot	tua	average	pria	tidak

Entropy Data Hipertensi:

$$P+ = \frac{3}{8}$$

$$P- = \frac{5}{8}$$

$$S = -\frac{3}{8} \cdot \log_2 \frac{3}{8} - \frac{5}{8} \cdot \log_2 \frac{5}{8} = \underline{0,954}$$