

Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



Task :

Analisis entropy dan information gain dari dataset tennis. Kemudian gambarkan bagaimana decision treenya (gambar tangan dan simulasi weka), dan buat decision rulennya.

Jawab :

1. Isi dataset Tennis

outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

Features: **Outlook, Temperature, Humidity, Wind**

Label: **play** (output)

Class: **Yes, No** (nilai untuk label)

2. Menghitung entropy

outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

Jumlah baris data = 14

Yes (+) = 9

No (-) = 5

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

Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



$$\text{Entropy (S)} = -p_{(\text{yes})} \log_2 p_{(\text{yes})} - p_{(\text{no})} \log_2 p_{(\text{no})}$$

$$\text{Entropy (S)} = - (9/14) \log_2 (9/14) - (5/14) \log_2 (5/14)$$

$$\text{Entropy (S)} = - (-0.41) - (-0.53)$$

$$\text{Entropy (S)} = \mathbf{0.94}$$

Nilai kategori untuk masing-masing fitur

a. Outlook (Sunny, Overcast, Rainy)

outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

$$\text{Jumlah Sunny} = 5$$

$$\text{Yes (+)} = 2$$

$$\text{No (-)} = 3$$

$$\text{Jumlah Rain} = 5$$

$$\text{Yes (+)} = 3$$

$$\text{No (-)} = 2$$

$$\text{Jumlah Overcast} = 4$$

$$\text{Yes (+)} = 4$$

$$\text{No (-)} = 0$$

$$\text{Entropy (Outlook=Sunny)} = -p_{(\text{yes})} \log_2 p_{(\text{yes})} - p_{(\text{no})} \log_2 p_{(\text{no})}$$

$$\text{Entropy (Outlook=Sunny)} = - (2/5) \log_2 (2/5) - (3/5) \log_2 (3/5)$$

$$\text{Entropy (Outlook=Sunny)} = \mathbf{0.971}$$

$$\text{Entropy (Outlook=Rain)} = -p_{(\text{yes})} \log_2 p_{(\text{yes})} - p_{(\text{no})} \log_2 p_{(\text{no})}$$

$$\text{Entropy (Outlook=Rain)} = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5)$$

$$\text{Entropy (Outlook=Rain)} = \mathbf{0.971}$$

$$\text{Entropy (Outlook=Overcast)} = -p_{(\text{yes})} \log_2 p_{(\text{yes})} - p_{(\text{no})} \log_2 p_{(\text{no})}$$

$$\text{Entropy (Outlook=Overcast)} = - (4/4) \log_2 (4/4) - (0/4) \log_2 (0/4)$$

$$\text{Entropy (Outlook=Overcast)} = \mathbf{0}$$

Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



b. Temperature (Hot, Mild, Cool)

outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

Jumlah Hot	= 4
Yes (+)	= 2
No (-)	= 2
Jumlah Mild	= 6
Yes (+)	= 4
No (-)	= 2
Jumlah Cool	= 4
Yes (+)	= 3
No (-)	= 1

Entropy (Temperature=Hot) = $-p_{(yes)} \log_2 p_{(yes)} - p_{(no)} \log_2 p_{(no)}$
Entropy (Temperature=Hot) = $-(2/4) \log_2 (2/4) - (2/4) \log_2 (2/4)$
Entropy (Temperature=Hot) = **1**

Entropy (Temperature=Mild) = $-p_{(yes)} \log_2 p_{(yes)} - p_{(no)} \log_2 p_{(no)}$
Entropy (Temperature=Mild) = $-(4/6) \log_2 (4/6) - (2/6) \log_2 (2/6)$
Entropy (Temperature=Mild) = **0.918**

Entropy (Temperature=Cool) = $-p_{(yes)} \log_2 p_{(yes)} - p_{(no)} \log_2 p_{(no)}$
Entropy (Temperature=Cool) = $-(3/4) \log_2 (3/4) - (1/4) \log_2 (1/4)$
Entropy (Temperature=Cool) = **0.811**

c. Humidity (High, Normal)

outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

Jumlah High	= 7
Yes (+)	= 3
No (-)	= 4
Jumlah Normal	= 7
Yes (+)	= 6
No (-)	= 1

Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



$$\text{Entropy (Humidity=High)} = -p_{(\text{yes})} \log_2 p_{(\text{yes})} - p_{(\text{no})} \log_2 p_{(\text{no})}$$

$$\text{Entropy (Humidity=High)} = - (3/7) \log_2 (3/7) - (4/7) \log_2 (4/7)$$

$$\text{Entropy (Humidity=High)} = \mathbf{0.985}$$

$$\text{Entropy (Humidity=Normal)} = -p_{(\text{yes})} \log_2 p_{(\text{yes})} - p_{(\text{no})} \log_2 p_{(\text{no})}$$

$$\text{Entropy (Humidity=Normal)} = - (6/7) \log_2 (6/7) - (1/7) \log_2 (1/7)$$

$$\text{Entropy (Humidity=Normal)} = \mathbf{0.591}$$

d. Wind (Strong, Weak)

outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

$$\text{Jumlah Weak} = 8$$

$$\text{Yes (+)} = 6$$

$$\text{No (-)} = 2$$

$$\text{Jumlah Strong} = 6$$

$$\text{Yes (+)} = 3$$

$$\text{No (-)} = 3$$

$$\text{Entropy (Wind=Weak)} = -p_{(\text{yes})} \log_2 p_{(\text{yes})} - p_{(\text{no})} \log_2 p_{(\text{no})}$$

$$\text{Entropy (Wind=Weak)} = - (6/8) \log_2 (6/8) - (2/8) \log_2 (2/8)$$

$$\text{Entropy (Wind=Weak)} = \mathbf{0.811}$$

$$\text{Entropy (Wind=Strong)} = -p_{(\text{yes})} \log_2 p_{(\text{yes})} - p_{(\text{no})} \log_2 p_{(\text{no})}$$

$$\text{Entropy (Wind=Strong)} = - (3/6) \log_2 (3/6) - (3/6) \log_2 (3/6)$$

$$\text{Entropy (Wind=Strong)} = \mathbf{1}$$

3. Menghitung information gain

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{v \in \text{values(A)}} |S_v|/|S| \text{Entropy}(S_v)$$

a. Outlook (Sunny, Overcast, Rainy)

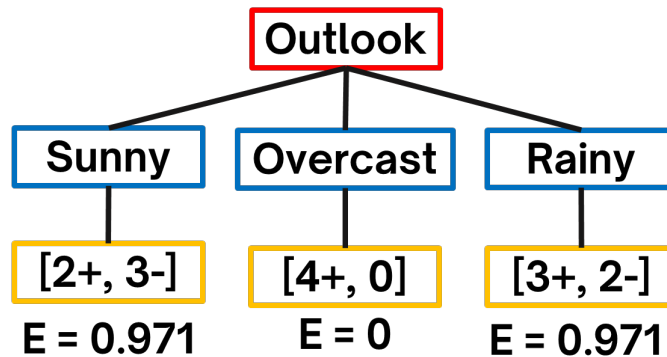
$$\text{Entropy(S)} = 0.94$$

$$\text{Entropy (Outlook=Sunny)} = 0.971$$

$$\text{Entropy (Outlook=Rainy)} = 0.971$$

$$\text{Entropy (Outlook=Overcast)} = 0$$

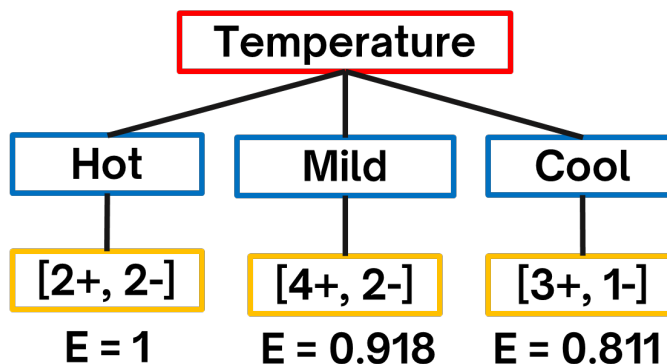
Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



$$\text{Gain}(S, \text{Outlook}) = 0.94 - (5/14) \cdot 0.971 - (5/14) \cdot 0.971 - (4/14) \cdot 0$$

Gain (S, Outlook) = 0.246

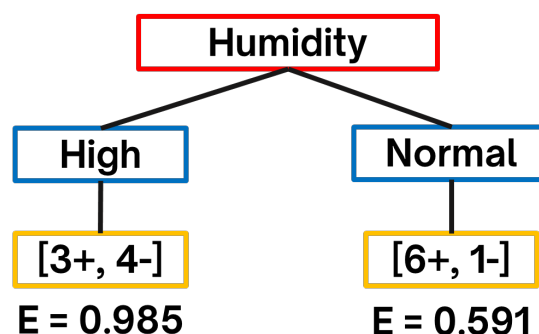
- b. Temperature (Hot, Mild, Cool)
- Entropy(S) = 0.94
- Entropy (Temperature=Hot) = 1
- Entropy (Temperature=Mild) = 0.918
- Entropy (Temperature=Cold) = 0.811



$$\text{Gain}(S, \text{Temperature}) = 0.94 - (4/14) \cdot 1 - (6/14) \cdot 0.918 - (4/14) \cdot 0.811$$

Gain (S, Temperature) = 0.029

- c. Humidity (High, Normal)
- Entropy(S) = 0.94
- Entropy (Humidity=High) = 0.985
- Entropy (Humidity=Normal) = 0.591



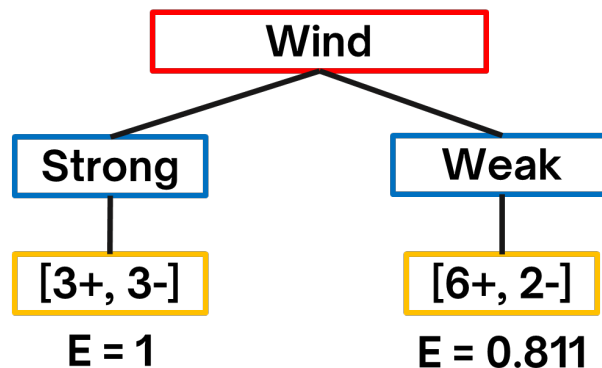
Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



$$\begin{aligned}\text{Gain (S, Humidity)} &= 0.94 - (7/14) * 0.985 - (7/14) * 0.591 \\ \text{Gain (S, Humidity)} &= \mathbf{0.152}\end{aligned}$$

d. Wind (Strong, Weak)

$$\begin{aligned}\text{Entropy(S)} &= 0.94 \\ \text{Entropy (Wind= Strong)} &= \mathbf{1} \\ \text{Entropy (Wind = Weak)} &= \mathbf{0.811}\end{aligned}$$



$$\begin{aligned}\text{Gain (S, Wind)} &= 0.94 - (8/14) * 0.811 - (6/14) * 1 \\ \text{Gain (S, Wind)} &= \mathbf{0.048}\end{aligned}$$

Information Gain masing-masing fitur :

$$\text{Gain (S, Outlook)} = \mathbf{0.246}$$

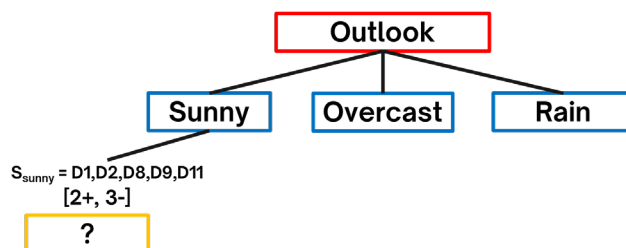
$$\text{Gain (S, Temperature)} = 0.029$$

$$\text{Gain (S, Humidity)} = 0.152$$

$$\text{Gain (S, Wind)} = 0.048$$

(Nilai information gain terbesar akan menjadi root/pangkal dari decision treenya, yaitu Outlook)

4. Membuat Decision Tree (Algoritma ID3)



$$\text{Entropy (Outlook=Sunny)} = 0.971$$

Uji untuk mengisi simpul dengan mencari nilai gain pada masing-masing feature :

Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

$$\begin{aligned}
\text{Gain}(S_{\text{sunny}}, \text{Temperature}) &= \text{Entropy}(\text{Sunny}) - I(\text{Temperature}) \\
&= \text{Entropy}(\text{Sunny}) - \frac{2}{5} \cdot \text{Entropy}(\text{Temperature}=\text{Hot}) - \frac{2}{5} \cdot \text{Entropy}(\text{Temperature}=\text{Mild}) - \frac{1}{5} \cdot \text{Entropy}(\text{Temperature}=\text{Cool}) \\
&= (-\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5}) - \frac{2}{5} \cdot (-\frac{0}{2} \log_2 \frac{0}{2} - \frac{0}{2} \log_2 \frac{0}{2}) - \frac{2}{5} \cdot (-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2}) - \frac{1}{5} \cdot (-\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{0} \log_2 \frac{0}{0}) \\
&= 0.970 - \frac{2}{5} \cdot 0.0 - \frac{2}{5} \cdot 1.0 - \frac{1}{5} \cdot 0.0 \\
&= 0.570
\end{aligned}$$

outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

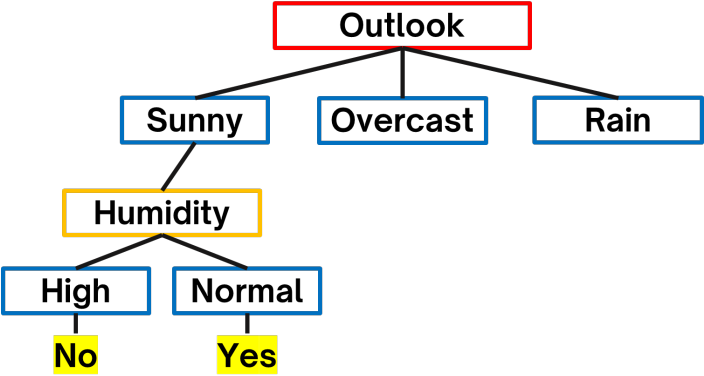
$$\begin{aligned}
\text{Gain}(S_{\text{sunny}}, \text{Humidity}) &= \text{Entropy}(\text{Sunny}) - I(\text{Humidity}) \\
&= \text{Entropy}(\text{Sunny}) - \frac{3}{5} \cdot \text{Entropy}(\text{Humidity}=\text{High}) - \frac{2}{5} \cdot \text{Entropy}(\text{Humidity}=\text{Normal}) \\
&= (-\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5}) - \frac{3}{5} \cdot (-\frac{0}{3} \log_2 \frac{0}{3} - \frac{3}{3} \log_2 \frac{3}{3}) - \frac{2}{5} \cdot (-\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{0} \log_2 \frac{0}{0}) \\
&= 0.970 - \frac{3}{5} \cdot 0.0 - \frac{2}{5} \cdot 0.0 \\
&= 0.970
\end{aligned}$$

Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5

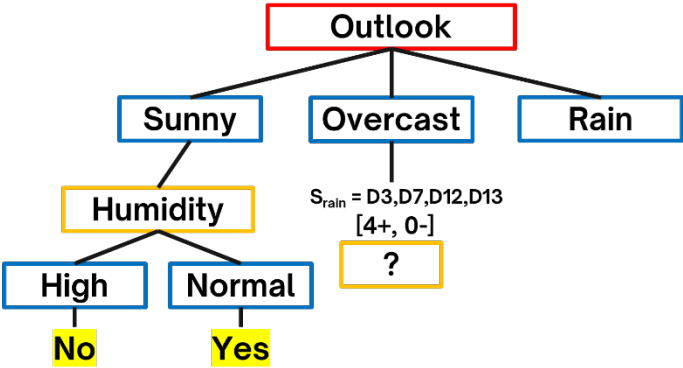


outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

$$\begin{aligned}
 \text{Gain}(S_{\text{sunny}}, \text{Wind}) &= \text{Entropy}(\text{Sunny}) - I(\text{Wind}) \\
 &= \text{Entropy}(\text{Sunny}) - 2/5 * \text{Entropy}(\text{Wind}=\text{Strong}) - 3/5 * \text{Entropy}(\text{Wind}=\text{Weak}) \\
 &= (- (2/5) \log_2 (2/5) - (3/5) \log_2 (3/5)) - 2/5 * (- (1/2) \log_2 (1/2) - (1/2) \log_2 (1/2)) \\
 &\quad - 3/5 * (- (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3)) \\
 &= 0.970 - (2/5)1.0 - (3/5)0.918 \\
 &= 0.0192
 \end{aligned}$$

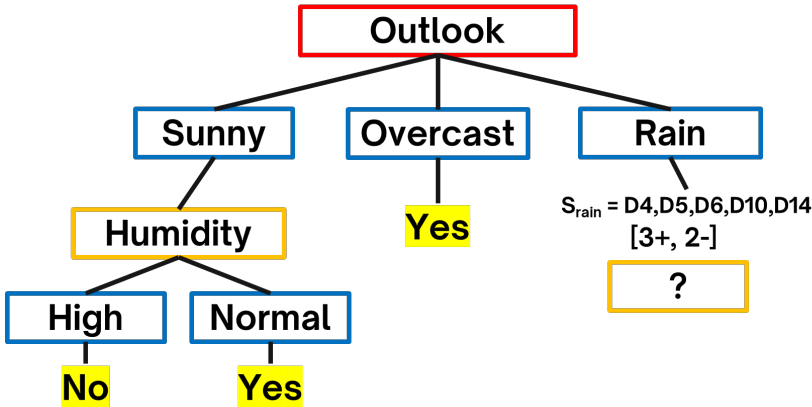
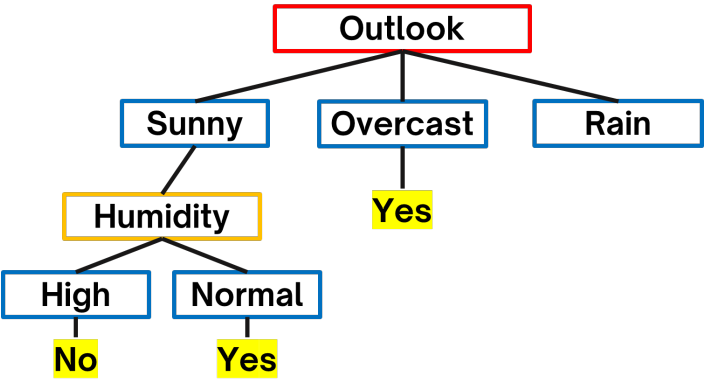


$$\begin{aligned}
 \text{Gain}(S_{\text{sunny}}, \text{Temperature}) &= 0.570 \\
 \text{Gain}(S_{\text{sunny}}, \text{Humidity}) &= 0.970 \\
 \text{Gain}(S_{\text{sunny}}, \text{Wind}) &= 0.019
 \end{aligned}$$



Entropy (Outlook=Overcast) = 0
 Sehingga langsung ke play (yes/no)

Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



Entropy (Outlook=Rainy) = 0.971
 Uji untuk mengisi simpul dengan mencari nilai gain pada masing-masing feature :

outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

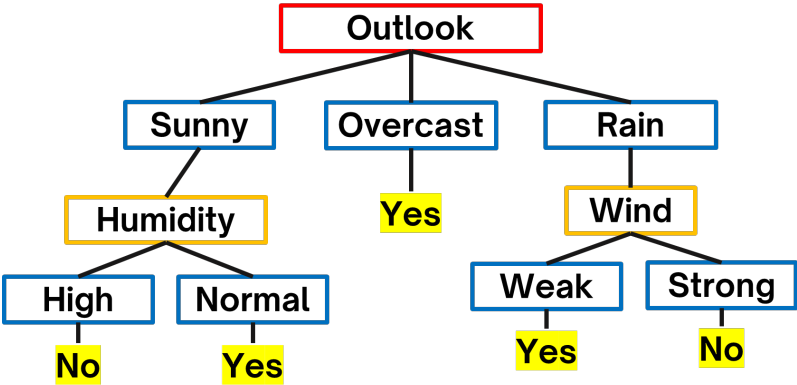
$$\begin{aligned}
 \text{Gain}(S_{\text{rain}}, \text{Temperature}) &= \text{Entropy}(\text{Rain}) - I(\text{Temperature}) \\
 &= \text{Entropy}(\text{Rain}) - 0/5 \cdot \text{Entropy}(\text{Temperature}=\text{Hot}) - \\
 &\quad 3/5 \cdot \text{Entropy}(\text{Temperature}=\text{Mild}) - 2/5 \cdot \text{Entropy}(\text{Temperature}=\text{Cool}) \\
 &= (- (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5)) - 0/5 \cdot (- (0/0) \log_2 (0/0) - (0/0) \log_2 \\
 &\quad (0/0)) - 3/5 \cdot (- (2/3) \log_2 (2/3) - (1/3) \log_2 (1/3)) - 2/5 \cdot (- (1/2) \log_2 (1/2) - (1/2) \\
 &\quad \log_2 (1/2)) \\
 &= 0.970 - (0/5)0.0 - (3/5)0.918 - (2/5)1 \\
 &= 0.0192
 \end{aligned}$$

Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



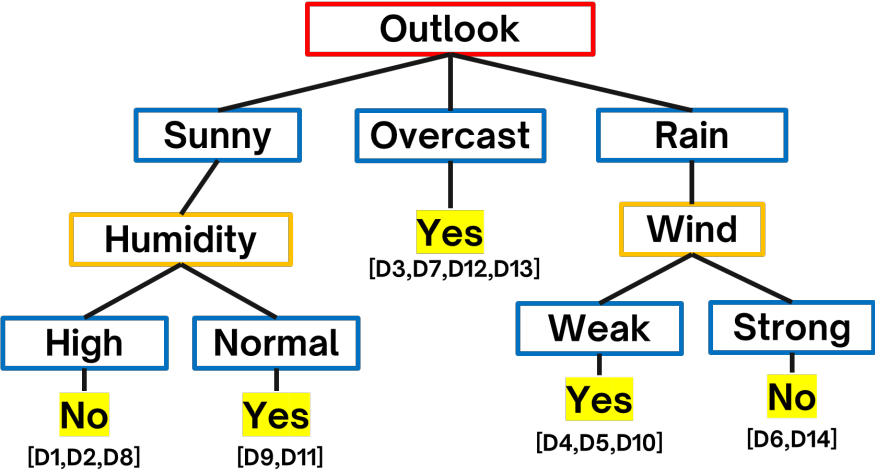
outlook	temperature	humidity	wind	play
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no

$$\begin{aligned}
 \text{Gain}(S_{\text{rain}}, \text{Wind}) &= \text{Entropy}(\text{Rain}) - I(\text{Wind}) \\
 &= \text{Entropy}(\text{Rain}) - 2/5 * \text{Entropy}(\text{Wind}=\text{Strong}) - 3/5 * \text{Entropy}(\text{Wind}=\text{Weak}) \\
 &= (- (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5)) - 2/5 * (- (0/0) \log_2 (0/0) - (0/0) \log_2 (0/0)) - 3/5 * (- (3/3) \log_2 (3/3) - (0/3) \log_2 (0/3)) \\
 &= 0.970 - (2/5)0.0 - (3/5)0.0 \\
 &= 0.970
 \end{aligned}$$



$$\begin{aligned}
 \text{Gain}(S_{\text{sunny}}, \text{Temperature}) &= 0.0192 \\
 \text{Gain}(S_{\text{sunny}}, \text{Wind}) &= 0.970
 \end{aligned}$$

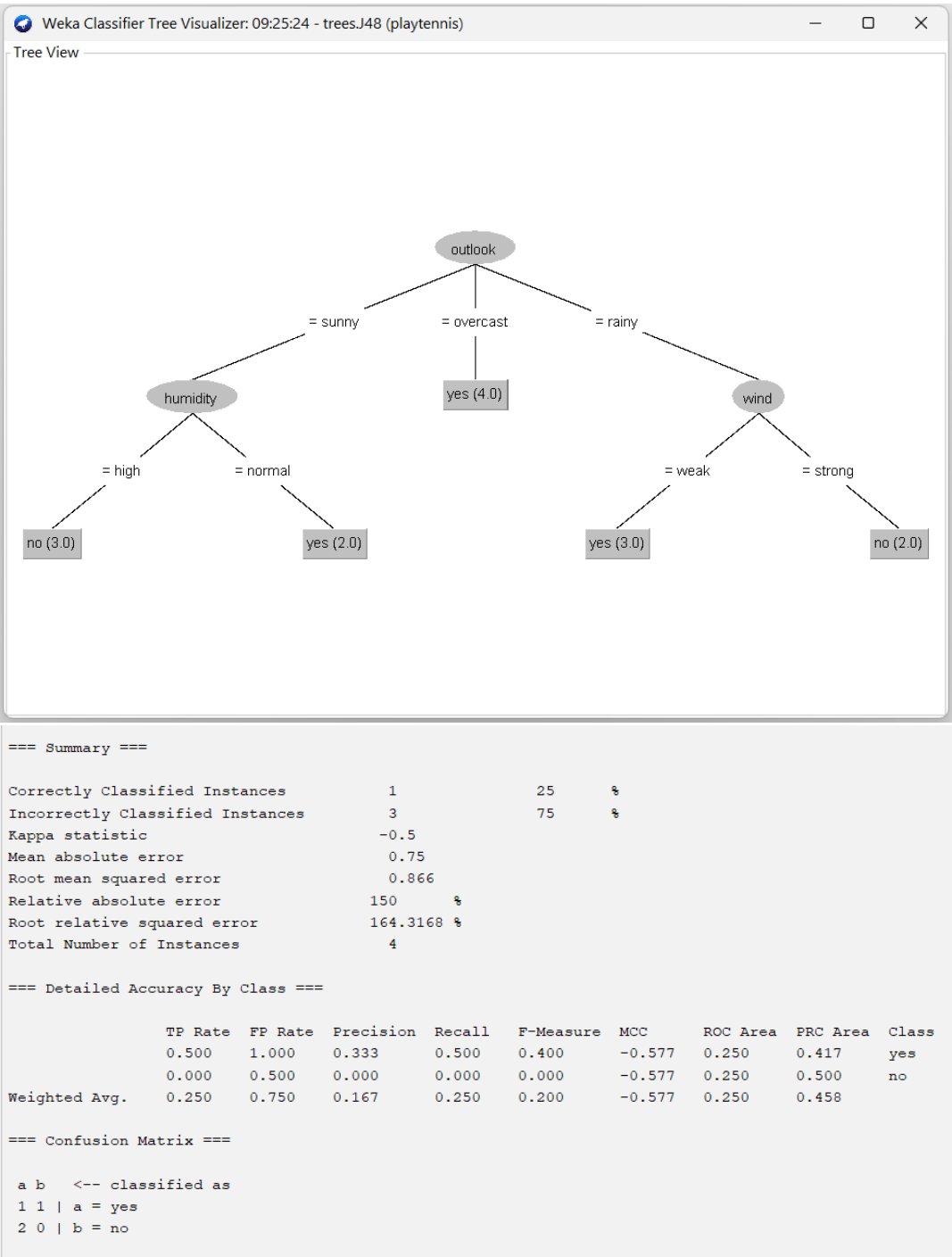
Hasil Decision Tree :



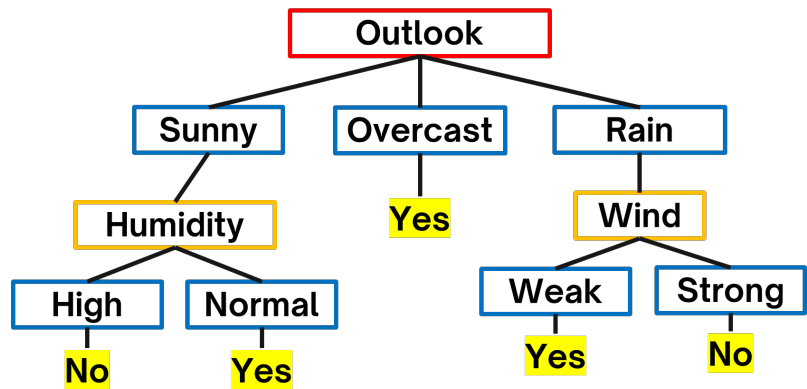
Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



Simulasi Weka :



Name	KAIRA MILANI FITRIA
NPM	2221210065
Subject	Artificial Intelligence and Pattern Recognition
Assignment	5



Decision Rule:

Rule 1:

IF (Outlook=Sunny) AND (Humidity=Normal) THEN (Play=Yes)

Outlook = Sunny ^ Humidity=Normal

Rule 2:

IF (Outlook=Sunny) AND (Humidity=High) THEN (Play=No)

Outlook = Sunny v Humidity=High

Rule 3:

IF (Outlook=Overcast) THEN (Play=Yes)

Outlook=Overcast

Rule 4:

IF (Outlook=Rain) AND (Wind=Weak) THEN (Play=Yes)

Outlook=Rain ^ Wind=Weak

Rule 5:

IF (Outlook=Rain) AND (Wind=Strong) THEN (Play=No)

Outlook=Rain v Wind=Strong

Decision rule secara keseluruhan adalah:

(Outlook = Sunny ^ Humidity=Normal) v (Outlook=Overcast)

v (Outlook=Rain ^ Wind=Weak)