

S.No	outlook	temperature	Humidity	windy	Play Tennis
1	Sunny	Hot	High	weak	No
2	Sunny	Hot	High	strong	No
3	overcast	Hot	High	weak	Yes
4	Rainy	mild	High	weak	Yes
5	Rainy	Cool	normal	weak	Yes
6	Rainy	Cool	normal	strong	No
7	overcast	Cool	normal	strong	Yes
8	Sunny	mild	High	weak	No
9	Sunny	Cool	normal	weak	Yes
10	Rainy	mild	normal	weak	Yes
11	Sunny	mild	normal	strong	Yes
12	overcast	mild	High	strong	Yes
13	overcast	Hot	normal	weak	Yes
14	Rainy	mild	High	strong	No

\*  $P = 9$

\*  $N = 5$

\*  $\text{total} = 14$

\* 
$$\text{Entropy} = -\frac{P}{P+N} \log_2 \left( \frac{P}{P+N} \right) - \frac{n}{P+N} \log_2 \left( \frac{n}{P+N} \right)$$

\* 
$$\text{Entropy}(S) = -\frac{9}{9+5} \log_2 \left( \frac{9}{9+5} \right) - \frac{5}{9+5} \log_2 \left( \frac{5}{9+5} \right) = \underline{\underline{0.940}}$$

\* For each attribute = (outlook):

\* Calculate entropy for each values [Sunny, Rainy, overcast]



(2)

outlook	Play Tennis
Sunny	No
Sunny	No
Sunny	No
Sunny	Yes
Sunny	Yes

outlook	Play Tennis
Rainy	Yes
Rainy	Yes
Rainy	No
Rainy	Yes
Rainy	No

outlook	Play Tennis
overcast	Yes
overcast	Yes
overcast	Yes
overcast	Yes

outlook	P	n	Entropy
Sunny	2	3	0.971
Rainy	3	2	0.971
overcast	4	0	0

→ (1)  
→ (2)  
→ (3)

$$* 1 - \text{entropy}_{\text{sunny}} = \frac{-2}{5} * \log_2\left(\frac{2}{5}\right) - \frac{3}{5} * \log_2\left(\frac{3}{5}\right) = 0.971$$

$$* 2 - \text{entropy}_{\text{rainy}} = \frac{-3}{5} * \log_2\left(\frac{3}{5}\right) - \frac{2}{5} * \log_2\left(\frac{2}{5}\right) = 0.971$$

$$* 3 - \text{entropy}_{\text{overcast}} = \frac{-4}{4} * \log_2\left(\frac{4}{4}\right) - \frac{0}{4} * \log_2\left(\frac{0}{4}\right) = 0$$

\* Average information Entropy =

$$* I(\text{outlook}) = \frac{P_{\text{sunny}} + n_{\text{sunny}}}{P+n} \text{Entropy}(\text{outlook} = \text{sunny}) + \frac{P_{\text{rainy}} + n_{\text{rainy}}}{P+n} \text{Entropy}(\text{outlook} = \text{Rainy}) + \frac{P_{\text{over}} + n_{\text{over}}}{P+n} \text{Entropy}(\text{outlook} = \text{overcast})$$



(3)

$$I(\text{outlook}) = \frac{3+2}{9+5} * 0.971 + \frac{2+3}{9+5} * 0.971 + \frac{4+0}{9+5} * 0 = 0.693$$

$$* \text{Gain} = \text{Entropy}(S) - I(\text{attribute})$$

$$* \text{entropy}(S) = 0.940$$

$$* \text{gain} = 0.940 - 0.693 = \underline{0.247}$$

Temp	Play	Temp	Play	Temp	Play
Hot	No	mild	yes	Cool	yes
Hot	No	mild	no	Cool	no
Hot	yes	mild	yes	Cool	yes
Hot	yes	mild	yes	Cool	yes
		mild	yes		
		mild	no		

Temperature	P	n	entropy
Hot	2	2	1
Mild	4	2	0.918
Cool	3	1	0.811

$$* I(\text{Temperature}) = \frac{P_{\text{hot}} + n_{\text{hot}}}{P+n} \text{entropy}(\text{hot}) +$$

$$\frac{P_{\text{mild}} + n_{\text{mild}}}{P+n} \text{entropy}(\text{Temp} = \text{mild}) + \frac{P_{\text{cool}} + n_{\text{cool}}}{P+n} \text{entropy}(\text{Cool})$$

$$* I(\text{Temp}) = \frac{2+2}{9+5} * 1 + \frac{4+2}{9+5} * 0.918 + \frac{3+1}{9+5} * 0.811 \Rightarrow 0.911$$

(4)

$$* \text{gain} = \text{entropy}(S) - I(\text{attribute})$$

$$* \text{gain} = 0.940 - 0.911 = \underline{\underline{0.029}}$$

←

\* Humidity → high / normal

Hum	Play
normal	yes
~	no
~	yes
~	yes
~	yes

Hum	Play
High	no
~	no
~	yes
~	yes
~	no
~	yes
~	no

Hum	p	n	entropy
High	3	4	0.985
normal	6	1	0.591

$$* \text{Average} \Rightarrow I = \frac{3+4}{9+5} * 0.985 + \frac{6+1}{9+5} * 0.591 = \underline{\underline{0.788}}$$

$$* \text{gain} = 0.940 - 0.788 = \underline{\underline{0.152}}$$



(5)

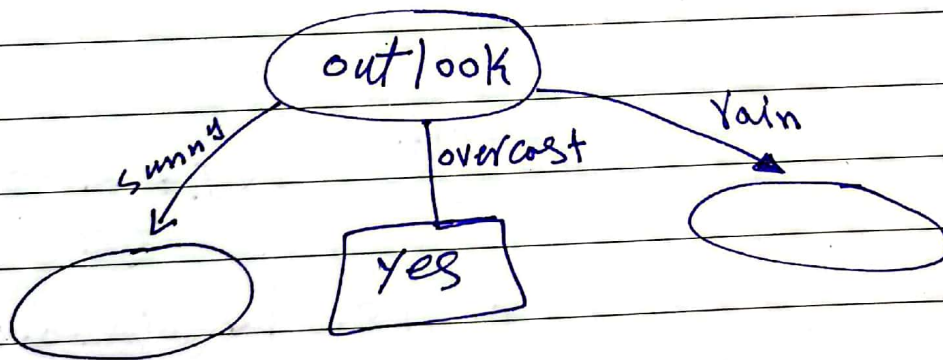
\* Windy  $\rightarrow$  Strong & Weak

windy	P	n	entropy
strong	3	3	1
weak	6	2	0.811

\* Average =  $I = \frac{3+3}{9+5} \times 1 + \frac{6+2}{9+5} \times 0.811 = \underline{\underline{0.892}}$

\* gain =  $0.940 - 0.892 = \underline{\underline{0.048}}$

\* Highest gain attribute  $\rightarrow$  outlook  $\rightarrow$  gain = 0.247  
 \* root node  $\Rightarrow$  outlook  
 Temperature  $\rightarrow \sim = 0.029$   
 Humidity  $\rightarrow \sim = 0.152$   
 $\sim \Rightarrow \sim = 0.048$



(6)

* outlook	Temp	Hum	windy	play
Sunny	Hot	High	Weak	No
~	Hot	~	Strong	No
~	mild	~	Weak	No
~	Cool	normal	~	Yes
~	mild	~	Strong	Yes

\* P = 2

\* N = 3

\* Total = 5

$$* \text{entropy (Sunny)} = \frac{-2}{5} \log_2 \left( \frac{2}{5} \right) - \frac{3}{2+3} \log_2 \left( \frac{3}{5} \right) = 0.971$$

\* Humidity  $\rightarrow$  high & normal

Hum	P	n	entropy
high	0	3	0
normal	2	0	0

\* I = 0

\* gain = 0.971

windy  $\rightarrow$

windy	P	n	entropy
strong	1	1	1
weak	1	2	0.918

\* I = 0.951

\* gain = 0.020

Temperature  $\rightarrow$

Temp	P	n	entropy
Cool	1	0	0
mild	0	2	0
mild	1	1	1

I = 0.4

gain = 0.571

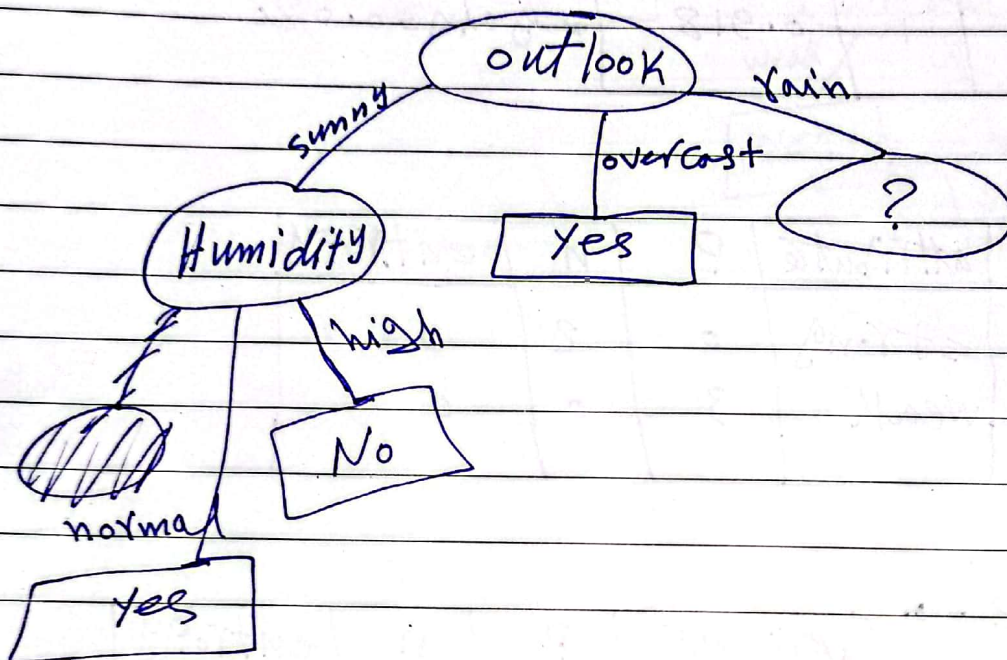


(4)

(2)

\* Highest Gain  $\Rightarrow$  Attributes  $\rightarrow$  Gain  
 Temperature  $\rightarrow$  0.571  
 Humidity  $\rightarrow$  0.971  
 Windy  $\rightarrow$  0.02

\* next node  $\Rightarrow$  Humidity



outlook	Temp	hum	windy	play
Rainy	mild	high	weak	yes
~	Cool	normal	~	~
~	~	~	Strong	no
~	mild	~	weak	yes
~	~	high	Strong	no

\* P = 3      \* n = 2      \* total = 5

\* entropy (Rainy) = 0.971

(8)

\* Humidity →

Hum	P	n	entropy
high	1	1	1
normal	2	1	0.918

\*  $I = 0.951$

\*  $\text{gain} = 0.020$

\* Windy →

attribute	P	n	entropy
strong	0	2	0
weak	3	0	0

\*  $I = 0$   
→

\*  $\text{gain} = 0.971$   
→

\* Temperature →

attribute	P	n	entropy
Cool	1	1	1
mild	2	1	0.918

\*  $I = 0.951$   
→

\*  $\text{gain} = 0.020$   
→

\* highest gain → Hum = 0.02

Windy = 0.971

Temp = 0.02

\* next node ⇒ Windy  
→



9

