

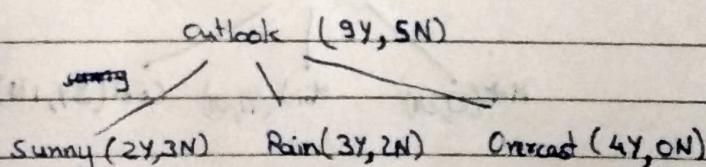
## Task on 6-11-2022 Decision Tree

Day	Outlook	Temp	Humid	Wind	Decision (Play)
1	Sunny	Hot	High	Weak	Yes No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rainfall	Mild	High	Weak	Yes
5	Rainfall	Cool	Normal	Weak	Yes
6	Rainfall	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rainfall	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rainfall	Mild	High	Strong	No

Total  $\rightarrow$  9Y & 5N

Choosing root feature by the information gain.

1.



$$H(S) = - \sum P_i \log_2 (P_i)$$

$$= - P_s \log_2 (P_s) - P_R \log_2 (P_R)$$

$$= - \frac{2}{14} \log_2 \left( \frac{2}{14} \right) - \frac{5}{14} \log_2 \left( \frac{5}{14} \right)$$

$$H(S) = 0.94$$

$$H(S)_{\text{sunny}} = -\frac{2}{5} \log_2 \left(\frac{2}{5}\right) = \frac{3}{5} \log_2 \left(\frac{3}{5}\right)$$

$$= 0.971$$

$$A(S)_{\text{rainfall}} = -\frac{3}{5} \log_2 \left(\frac{3}{5}\right) = \frac{2}{5} \log_2 \left(\frac{2}{5}\right)$$

$$= 0.971$$

$$H(S)_{\text{overcast}} = -\frac{1}{4} \log_2 \left(\frac{1}{4}\right) = \left(\frac{1}{4}\right) \log_2 \left(\frac{1}{4}\right)$$

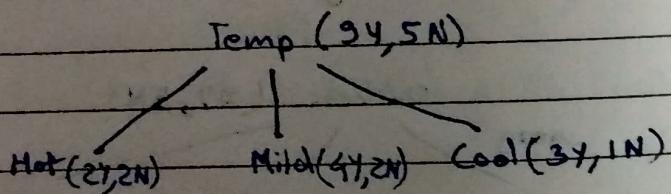
$$= 0$$

$$\therefore \text{Gain}(S, \text{outlook}) = H(S) - \frac{\sum S_v}{|S|} \times H(S_v)$$

$$= 0.94 - \left[ \frac{5}{14} \times 0.971 + \frac{5}{14} \times 0.971 + \frac{4}{14} \times 0 \right]$$

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

2.



$$H(S) = 0.974$$

$$H(S)_{\text{hot}} = -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) = \frac{2}{4} \log_2 \left(\frac{2}{4}\right)$$

$$= 1$$

$$H(S)_{\text{mild}} = -\frac{1}{6} \log_2 \left(\frac{1}{6}\right) = \frac{2}{6} \log_2 \left(\frac{2}{6}\right)$$

$$= 0.918$$

$$H(S)_{old} = -\frac{3}{4} \log_2 \left(\frac{3}{4}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right)$$

$$= 0.811$$

$$Gain(S, Temp) = 0.94 - \left[ \frac{5}{14} \times 1 + \frac{6}{14} \times 0.918 + \frac{4}{14} \times 0.811 \right]$$

$$Gain(S, Temp) = 0.029$$

3. ~~Humid~~

Humid (3y, 5N)

High (3y, 4N)

Normal (6y, 1N)

$$H(S) = 0.94$$

$$H(S)_{high} = -\frac{3}{7} \log_2 \left(\frac{3}{7}\right) - \frac{4}{7} \log_2 \left(\frac{4}{7}\right)$$

$$= 0.985$$

$$H(S)_{normal} = -\frac{6}{7} \log_2 \left(\frac{6}{7}\right) - \frac{1}{7} \log_2 \left(\frac{1}{7}\right)$$

$$= 0.592$$

$$Gain(S, Humid) = 0.94 - \left[ \frac{7}{14} \times 0.985 + \frac{7}{14} \times 0.592 \right]$$

$$= 0.151$$

Wind (3y, 5N)

weak (6y, 2N)

strong (3y, 3N)

$$H(S) = 0.94$$

$$H(S)_{weak} = -\frac{6}{8} \log_2 \left(\frac{6}{8}\right) - \frac{2}{8} \log_2 \left(\frac{2}{8}\right)$$

$$= 0.811$$

$$H(S)_{strong} = -\frac{3}{6} \log_2 \left(\frac{3}{6}\right) - \frac{3}{6} \log_2 \left(\frac{3}{6}\right)$$

$$= 1$$

$$\text{Gain}(S, \text{Wind}) = 0.94 - \left[ \frac{8}{14} \times 0.811 + \frac{6}{14} \times 1 \right]$$

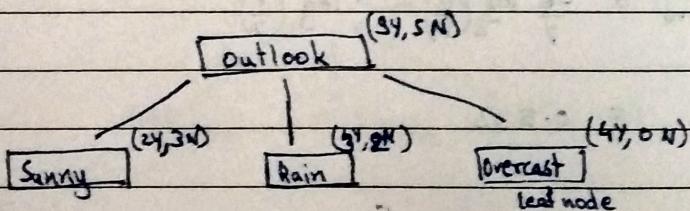
$$\text{Gain}(S, \text{Wind}) = 0.048$$

$$\text{Gain}(S, \text{Outlook}) = 0.296 \quad \leftarrow \text{Highest I.G.} \therefore \text{our root node}$$

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

$$\text{Gain}(S, \text{Humid}) = 0.151$$

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



For decision node (Sunny)

$$H(S) = 0.971$$

Table for sunny outlook

Temp	Humid	Wind	Decision
Hot	High	Weak	No
Hot	High	Strong	No
Mild	High	Weak	No
(Cool)	Normal	Weak	Yes
Mild	Normal	Strong	Yes

✓ Temp

Temp

1.

Temp (2Y, 3N)

Hot (0Y, 2N) Mild (1Y, 1N) Cool (1Y, 0N)

$$H(S) = 0.971$$

$$H(S)_{\text{hot}} = 0 - \frac{2}{2} \log_2 \left( \frac{2}{2} \right) = 0$$

$$H(S)_{\text{mild}} = -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) = 1$$

$$H(S)_{\text{cool}} = -\frac{1}{1} \log_2 \left( \frac{1}{1} \right) = 0 = 0$$

$$\text{Gain}(S, \text{Temp}) = 0.971 - \left[ \frac{2}{5} \times 0 + \frac{2}{5} \times 1 + \frac{1}{5} \times 0 \right] \\ = 0.571$$

2.

Humid (2Y, 3N)

High (0, 3N) ~~fixed~~

Normal (2Y, 1N)

$$H(S) = 0.971$$

$$H(S)_{\text{high}} = -0 - \frac{3}{3} \log_2 \left( \frac{3}{3} \right) = 0$$

$$H(S)_{\text{normal}} = -\frac{2}{2} \log_2 \left( \frac{2}{2} \right) = 0 = 0$$

$$\text{Gain}(S, \text{Humid}) = 0.971 - [0 + 0] = 0.971$$

3.

Wind (2Y, 3N)

Weak (2Y, 2N)

Strong (1Y, 1N)

$$H(S) = 0.971$$

$$H(S)_{\text{weak}} = -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) = 0.918$$

$$H(S)_{\text{strong}} = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$\text{Gain}(S, \text{Wind}) = 0.971 - \left[ \frac{3}{5} \times 0.918 + \frac{2}{5} \times 1 \right]$$

$$\text{Gain}(S, \text{Wind}) = 0.0192$$

$$\text{Gain}(S, \text{Temp}) = 0.571$$

$$\text{Gain}(S, \text{Humid}) = 0.971 \rightarrow \text{Highest I.G (our decision node for sunny)}$$

$$\text{Gain}(S, \text{Wind}) = 0.0192$$

For decision node (rain)

Table for rain outlook

$$H(S) = 0.971$$

Temp	Humid	Wind	Decision
Mild	High	Weak	Yes
(60)	Normal	Weak	Yes
Cool	Normal	Strong	No
Mild	Normal	Weak	Yes
Mild	High	Strong	No

1.

Temp (34, 2N)

/ \

Mild (24, 1M) Cool (14, 1W) Hot (08, 0N)

$$H(S) > 0.971$$

$$H(S)_{\text{hot}} = 0$$

$$H(S)_{\text{mild}} = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.918$$

$$H(S)_{\text{cold}} = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 0$$

$$\text{Gain}(S, \text{Temp}) = 0.971 - \left[ 0 + \frac{3}{5} \times 0.918 + 0 \right] = 0.4192$$

2. Humid

Humid (3Y, 2N)

High (1Y, 1N)

Normal (2Y, 1N)

$$H(S) = 0.971$$

$$H(S)_{\text{High}} = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$H(S)_{\text{Normal}} = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.918$$

$$\text{Gain}(S, \text{Humid}) = 0.971 - \left[ \frac{2}{5} \times 1 + \frac{3}{5} \times 0.918 \right] = 0.02$$

3. Wind

Wind (3Y, 2N)

Weak (3Y, 0N)

Strong (0Y, 2N)

$$H(S) = 0.971$$

$$H(S)_{\text{Weak}} = -\frac{3}{3} \log_2\left(\frac{3}{3}\right) - 0 = 0$$

$$H(S)_{\text{Strong}} = 0 - \frac{2}{2} \log_2\left(\frac{2}{2}\right) = 0$$

$$\text{Gain}(S, \text{Wind}) = 0.971 - [0 + 0] = 0.971$$

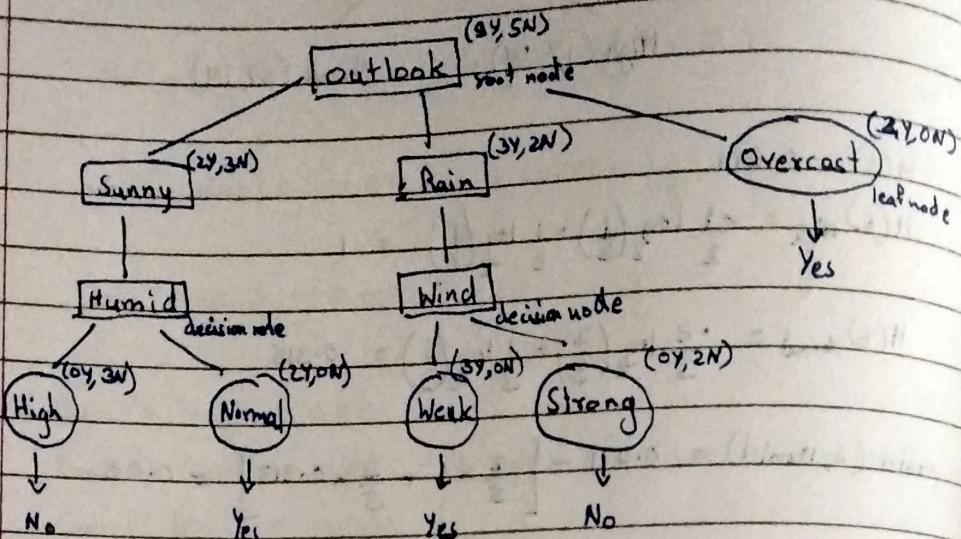
$$\text{Gain}(S, \text{Temp}) = 0.4192$$

$$\text{Gain}(S, \text{Humid}) = 0.02$$

$$\text{Gain}(S, \text{Wind}) = 0.971 \rightarrow \text{Highest I.G. (our decision node for rain)}$$

For overcast no decision node possible hence a leaf node

## Final Decision Tree



## Difference between Entropy &amp; Gini Impurity (Decision Tree)

Parameter	Entropy	Gini Impurity
Formula	$-\sum_{i=1}^n p_i \log(p_i)$	$1 - \sum_{i=1}^n p_i^2$
Values	0 to 1	0 to 0.5
Geographical rep. wrt probability		
Used in	ID3	CART