

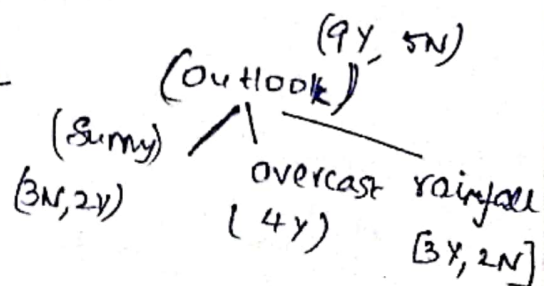
# DECISION TREE

Eg

DAY	outlook	Temperature	humidity	wind	Decision
1	Sunny	hot	high	weak	N
2	Sunny	hot	High	Strong	N
3	overcast	hot	High	weak	Y
4	rainfall	mild	High	weak	Y
5	rainfall	cool	normal	weak	Y
6	rainfall	cool	normal	Strong	N
7	overcast	cool	normal	Strong	Y
8	Sunny	mild	High	weak	N
9	Sunny	cool	normal	weak	Y
10	rainfall	mild	normal	weak	Y
11	Sunny	mild	normal	strong	Y
12	overcast	mild	High	strong	Y
13	overcast	Hot	normal	weak	Y
14	rainfall	mild	High	strong	N

Entropy =  $-\sum_{i=1}^n p_i \log_2(p_i)$  (Y, N) = 2

① outlook  $E_{(outlook)} =$



$$H(S) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.971$$

$$H(overcast) = \frac{4}{4} (\log_2 4/4) = 0$$

$$H(rainfall) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.971$$

$$H(\text{Outlook}) = 40.971 \times 0.971 - = = \frac{5}{14} (0.971) + 0 + \frac{5}{14} (0.971)$$

$$= -\frac{9}{14} \log(9/14) - \frac{5}{14} \log(5/14) = 0.693$$

$$= 0.940$$

$$H(\text{Temperature}) =$$

$$H(\text{Hot}) = -\frac{2}{4} (\log(1/2)) - \frac{1}{2} (\log(1/2)) = 1$$

(2N, 2Y)

$$H(\text{mild}) = -\frac{4}{6} (\log(4/6)) - \frac{2}{6} \log(2/6) = 0.918$$

(4Y, 2N)

$$H(\text{Cool}) = -\frac{1}{4} \log(1/4) - \frac{3}{4} \log(3/4) = 0.8825$$

(1N, 3Y)

$$H(\text{Temperature}) = \frac{4}{14} (1) + \frac{6}{14} (0.918) = 0.912$$

~~0.68~~

$$H(\text{Humidity}) =$$

$$H(\text{High}) = -\frac{3}{7} (\log(3/7)) - \frac{4}{7} \log(4/7) = 0.985$$

(3Y, 4N)

$$H(\text{normal}) = -\frac{1}{7} \log(1/7) - \frac{6}{7} (\log(6/7)) = 0.591$$

(1N, 6Y)

$$H(\text{Humidity}) = \frac{7}{14} (0.985) + \frac{7}{14} (0.591) = 0.788$$

$$H(\text{wind}) =$$

$$H(\text{strong}) = 1$$

(3Y, 3N)

$$H(\text{weak}) = -\frac{1}{8} (\log(1/8)) - \frac{7}{8} \log(7/8) = 0.5435$$

(1N, 7Y)

$$H(\text{wind}) = \frac{6}{14} + \frac{0.5435 \times 8}{14} = 0.7391$$

$H(\text{outlook})$  is the least  $\Rightarrow$  outlook has less randomness. & High Information gain  
So, root node would be outlook



calculation for decision node (Sunny)

Sunny

Day	outlook	Temp	Humidity	wind	Decision
1	Sunny	hot	High	weak	N
2	Sunny	hot	High	strong	N
8	Sunny	mild	High	weak	N
9	Sunny	cool	normal	weak	Y
11	Sunny	mild	normal	strong	Y

$$\textcircled{1} H(S) = 0.971$$

(3N, 2Y)

$$\underline{H_S(\text{Temp})} :-$$

$$H_S(\text{Hot}) = \frac{2}{2} \log 1 = 0$$

(0Y, 2N)

$$H_S(\text{mild}) = \frac{1}{2} (\log(1/2) + \frac{1}{2} \log(1/2)) = 1$$

(1N, 1Y)

$$H_S(\text{cool}) = 0$$

(1Y, 0N)

$$H_S(\text{Temp}) = \frac{2}{5} = 0.4$$

$$\begin{aligned} I.G &= 0.971 - 0.4 \\ &= 0.571 \end{aligned}$$

Hs(Humidity) (3N, 2Y) :-

$$Hs(\text{High}) = \frac{3}{3} \log(1) = 0 \quad I.G = 0.912 - 0$$

$$Hs(\text{normal}) = \frac{2}{2} (0) = 0 \quad = 0.912$$

$$Hs(\text{Humidity}) = 0$$

Hs(wind) :-

$$Hs(\text{weak}) = \frac{1}{3} \log(1/3) - \frac{2}{3} \log(2/3) = +0.918$$

$$Hs(\text{strong}) = 0$$

$$Hs(\text{wind}) = \frac{3}{5} (0.918) = 0.551$$

$$I.G = 0.912 - 0.551 = 0.361$$

$$Hs(\text{Humidity}) < Hs(\text{temp}) < Hs(\text{wind})$$

$$I.G(\text{Humidity}) > I.G(\text{Temp}) > I.G(\text{wind})$$

Calculation for node (Rain) :-

Day	outlook	Temp	Humidity	wind	Decision
4		mild	High	weak	Y
5	Rain Jazz	cool	normal	weak	Y
6		cool	normal	strong	N
10		mild	normal	weak	Y
14		mild	High	strong	N

$$H(R) = \frac{3}{5} \log(3/5) - \frac{2}{5} \log(2/5) =$$



$H_R$  (Temp) :-

$$H_R (\text{mild}) = \frac{1}{3} \log(1/3) - \frac{2}{3} \log(2/3) = 0.918$$

(2Y, 1N)

$$H_R (\text{cool}) = 0$$

(1Y, 1N)

$$H_R (\text{Temp}) = \frac{3}{5} (0.918) = 0.551; \quad I_G =$$

$H_R$  (Humidity) :-

$$H_R (\text{high}) = 0$$

(1Y, 1N)

$$H_R (\text{normal}) = 0.918$$

(2Y, 1N)

$$H_R (\text{Humidity}) = 0.551$$

$H_R$  (wind) =

$$H_R (\text{weak}) = 1$$

(3Y)

$$H_R (\text{strong}) = 1$$

(2N)

$$H_R (\text{wind}) = \frac{2}{5} = 0.4$$

$$\checkmark \quad H_R (\text{wind}) < H_R (\text{Humidity}), H_R (\text{Temp})$$

$$I_G (\text{wind}) > I_G (\text{Humidity, temp})$$

