Mid

(2)- Decision Tree :-

(a) :- Find Entropy
(b) :- Entropy of Feature
(c) :- Information Gain

(a) :- $\text{Entropy} = -P(\text{Yes}) \log_2 P(\text{Yes}) - P(\text{No}) \log_2 P(\text{No})$

(b) :- $E(\text{Feature}) = E(F) \cdot P(\text{Yes}) + E(F) P(\text{No})$

(c) :- $G(\text{Gain}) = E(T) - E(F)$

Wind	Humidity	Clouds	Rain	
10	10	No	No	
20	60	Heavy	Yes	No = 3
15	50	Medium	Yes	Yes = 4
25	(70)	Medium	Yes	
20	10	No	No	T = 7
(40)	20	Medium	No	
15	45	Heavy	Yes	

Find Root-Node = (?)

Find Entropies

Step #01:- Calculate Entropy of Rain (Target):-

$E(R) = ?$

$E(R) = E(4,3) \rightarrow E(\text{Yes}, \text{No})$

$E(4,3) = E(R)$

$$E(R) = -\left(\frac{4}{7}\right)\log_2\left(\frac{4}{7}\right) - \left(\frac{3}{7}\right)\log_2\left(\frac{3}{7}\right)$$

$$\boxed{E(R) = 0.9852}$$

$E(W) = ?$

		Yes	No	
Wind	$W \geq 20$	2	2	(4)
	$W < 20$	2	1	(3)
				(7)

$$E(W) = \left(\frac{4}{7}\right)E(W \geq 20) + \left(\frac{3}{7}\right)E(W < 20)$$

$$= \left(\frac{4}{7}\right)E(2,2) + \left(\frac{3}{7}\right)E(2,1)$$

$$= \left(\frac{4}{7}\right)[(1)] + \left(\frac{3}{7}\right)\left[-\left(\frac{2}{3}\right)\log_2\left(\frac{2}{3}\right) - \left(\frac{1}{3}\right)\log_2\left(\frac{1}{3}\right)\right]$$

$$\boxed{E(W) = 0.9649}$$

$E(Hy) = P$

		Yes	No	
Humidity	$H \geq 35$	4	0	(4)
	$H < 35$	0	3	(3)
				7

$$E(H) = \left(\frac{4}{7}\right) E(H \geq 35) + \left(\frac{3}{7}\right) E(H < 35)$$

$$= \left(\frac{4}{7}\right) E(4,0) + \left(\frac{3}{7}\right) E(0,3)$$

$$= \left(\frac{4}{7}\right)(0) + \left(\frac{3}{7}\right)(0)$$

$$\boxed{E(H) = 0}$$

$E(C) = P$

		Yes	No	
Clouds	No	0	2	(2)
	Heavy	2	0	(2)
	Medium ~~Yes~~	2	1	(3)
				7

$$E(C) = \left(\frac{2}{7}\right) E(No) + \left(\frac{2}{7}\right) E(Heavy) + \left(\frac{3}{7}\right) E(Medium)$$

$$E(C) = \left(\frac{2}{7}\right) E(0,2) + \left(\frac{2}{7}\right) E(2,0) + \left(\frac{3}{7}\right) E(2,1)$$

$$= \left(\frac{2}{7}\right)(0) + \left(\frac{2}{7}\right)(0) + \left(\frac{3}{7}\right)\left[-\left(\frac{2}{3}\right)\log_2\left(\frac{2}{3}\right) - \left(\frac{1}{3}\right)\log_2\left(\frac{1}{3}\right)\right]$$

$$\boxed{E(C) = 0.3935}$$

Step #02 :- Information Gain :-

$$G(W) = E(R) - E(W)$$

$$= 0.9852 - 0.9649$$

$$\boxed{G(\text{Wind}) = 0.0203}$$

$$G(H) = E(R) - E(H)$$

$$= 0.9852 - 0$$

$$\boxed{G(H) = 0.9852} \longrightarrow \text{(Highest)}$$

$$G(C) = E(R) - E(C)$$

$$= 0.9852 - 0.3935$$

$$\boxed{G(\text{Clouds}) = 0.5917}$$

Hence,

The Gain of (Humidity) is greater than other Features :-

Root - Node is :-