C4.5 Algorithm Example:-

Given Data Set:-

Sl.No	Weather	Mood	Study
1	Sunny	Happy	Yes
2	Rainy	sad	No
3	Sunny	Sad	Yes
4	~~Happy~~ Rainy	Happy	Yes

(1) Calculate The Entropy of the dataset (D) for target attribute "Study".

$$\text{Entropy}(D) = -\sum_{i=1}^{n} P_i \times \log_2(P_i)$$

$$= -P(\text{yes}) \cdot \log_2(P(\text{yes})) - P(\text{No}) \cdot \log_2(P(\text{No}))$$

$$= -3/4 \times \log_2(3/4) - 1/4 \times \log_2(1/4)$$

$$\simeq 0.811$$

(2) Calculate Gain Ration of each ~~Attribute~~ = Weather

(a) ~~Attribute = Weather~~

$$\text{Entropy}(\text{Sunny}) = -2/2 \log(2/2) - 0/2 \log(0/2)$$

$$= 0$$

$$\text{Entropy}(\text{Rainy}) = -1/2 \log(1/2) - 1/2 \log(1/2)$$

$$= 1$$

$$\text{Avg Entropy}(\text{weather}) = 2/4 \times 0 + 2/4 \times 1 = 0.5$$

$$\text{Information Gain}(\text{Weather}) = E(D) - E(\text{weather})$$

$$= 0.811 - 0.5$$

$$= 0.311$$

~~Split Rati~~

$$\text{Split Info [weather]} = -\sum_{i=1}^{m} \frac{|D_j|}{|D|} \cdot \log \frac{|D_j|}{|D|}$$

$$= -2/4 \times \log(2/4) - 2/4 \cdot \log(2/4)$$

$$= 0.5 + 0.5 = 1$$

$$\text{Gain Ratio [weather]} = \frac{\text{Gain [weather]}}{\text{SplitInfo [weather]}}$$

$$= \frac{0.311}{1} = 0.311$$

(3) Calculate Gain Ration of ~~Attribute~~ = Mood.

$$\text{Entropy (Happy)} = -2/2 \log(2/2) - 0/2 \log(0/2)$$

$$= 0$$

$$\text{Entropy (Sad)} = -1/2 \log(1/2) - 1/2 \log(1/2)$$

$$= 1$$

$$\text{Avg Entropy (Mood)} = 2/4 \times 0 + 2/4 \times 1$$

$$= 0.5$$

$$\text{Information Gain (Mood)} = E[D] - E[\text{mood}]$$

$$= 0.811 - 0.5 = 0.311$$

$$\text{SplitInfo [Mood]} = -\sum_{i=1}^{m} \frac{|D_j|}{|D|} \log \frac{|D_j|}{|D|}$$

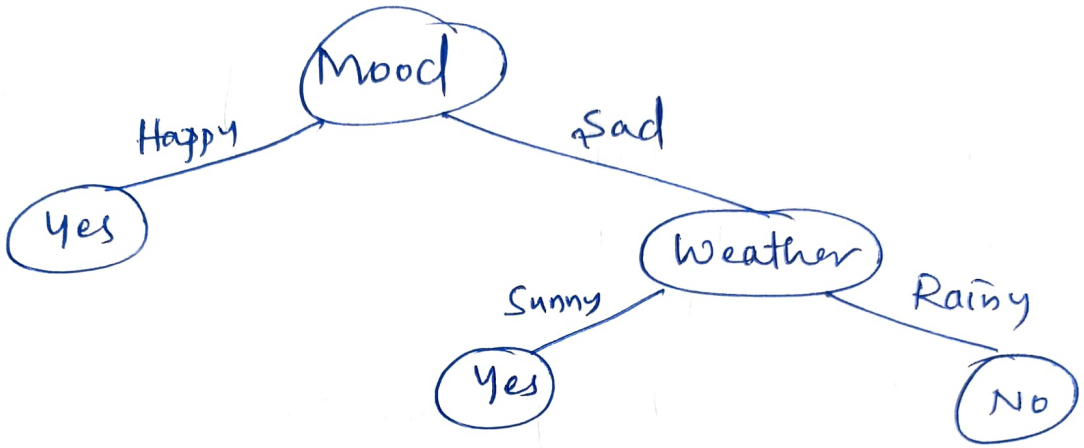
$$= -2/4 \times \log(2/4) - 2/4 \log(2/4)$$

$$= 0.5 + 0.5 = 1$$

$$\text{Gain Ratio [Mood]} = \frac{\text{Gain [Mood]}}{\text{SplitInfo [mood]}} = \frac{0.311}{1} = 0.311$$

(4) Since Gain Ration of Weather & Mood attribute are same. The attribute "Mood" is choosen as splitting attribute.

The final Decision Tree constructed for C4.5 algorithm as:



———— x ————