Question-2

*) Decision tree using Gini Indexing

	Weather	Temperature	Wind Level	Golf for running
1)	Sunny ✓	High ✓	Low ✓	No
2)	Sunny ✓	Medium ←	Medium ←	Yes
3)	Cloudy ←	High ✓	Medium ←	Yes
4)	Cloudy ←	Medium ←	High .	Yes
5)	Rainy .	High ✓	Low ✓	No
6)	Rainy .	~~Medium~~ High ✓	Medium ←	No
7)	Sunny ✓	Low	High .	No

Gini Impurity for the 'weather' feature:

$$G(\text{class} = \text{'sunny'}) \Rightarrow 1 - \left[(P(\text{yes}))^2 + (P(\text{NO}))^2 \right]$$

$$\Rightarrow 1 - \left[(1/3)^2 + (2/3)^2 \right]$$

$$\Rightarrow 1 - \left[1/9 + 4/9 \right] \Rightarrow 1 - \left[5/9 \right] \Rightarrow 4/9$$

$$G(\text{class} = \text{'Cloudy'}) \Rightarrow 1 - \left[(P(\text{yes}))^2 + (P(\text{NO}))^2 \right]$$

$$\Rightarrow 1 - \left[(2/2)^2 + (0)^2 \right] \Rightarrow 1 - \left[(1)^2 \right]$$

$$\Rightarrow 0$$

$$G(\text{class} = \text{'Rainy'}) \Rightarrow 1 - \left[(P(\text{yes}))^2 + (P(\text{NO}))^2 \right]$$

$$\Rightarrow 1 - \left[0 + (2/2)^2 \right] \Rightarrow 0$$

Weighted sum of gini impurity for the 'weather' feature.

$\Rightarrow 3/7 * (4/9) + 2/7 * 0 + 2/7 * 0$

$\Rightarrow 4/21 \Rightarrow 0.1904$

Gini Impurity for the 'temperature' feature:

$G(\text{class} = \text{'High'}) \Rightarrow 1 - [(P(yes))^2 + (P(NO))^2]$

$\Rightarrow 1 - [(1/4)^2 + (3/4)^2] \Rightarrow 1 - [1/16 + 9/16] \Rightarrow 1 - 4/16$

$\Rightarrow 1 - 1/4 \Rightarrow 3/4$

$G(\text{class} = \text{'Medium'}) \Rightarrow 1 - [(P(yes))^2 + (P(NO))^2]$

$\Rightarrow 1 - [(2/2)^2 + 0]$

$\Rightarrow 0$

$G(\text{class} = \text{'Low'}) \Rightarrow 1 - [(P(yes))^2 + (P(NO))^2]$

$\Rightarrow 1 - [0^2 + (1/1)^2] \Rightarrow 1 - 1$

$\Rightarrow 0$

Weighted sum of Gini Impurity for the 'temperature' feature.

$\Rightarrow 3/7 * 3/4 + 2/7 * 0 + 1/7 * 0$

$\Rightarrow 3/7 \Rightarrow 0.4285$

Gini Impurity for the feature Wind level.

*)

$G_I(\text{class} = \text{'Low'}) \Rightarrow 1 - \left[0 + (2/2)^2\right] \Rightarrow 0$

$G_I(\text{class} = \text{'Medium'}) \Rightarrow 1 - \left[(2/3)^2 + (1/3)^2\right] \Rightarrow 1 - \left[4/9 + 1/9\right]$

$\Rightarrow 1 - 5/9 \Rightarrow 4/9$

$G_I(\text{class} = \text{'High'}) \Rightarrow 1 - \left[(1/2)^2 + (1/2)^2\right]$

$\Rightarrow 1 - \left[1/4 + 1/4\right] \Rightarrow 1 - 1/2 \Rightarrow 1/2$

Weighted sum

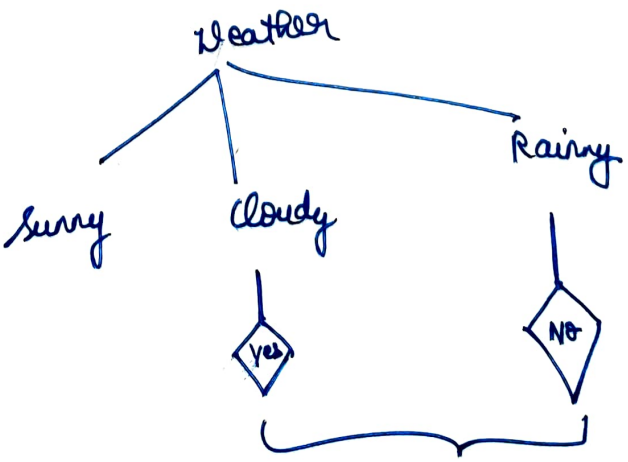
$\Rightarrow 2/7 * 0 + (3/7 * 4/9) + (2/7 * 1/2)$

$\Rightarrow (1/7 * 4/3) + 1/7 \Rightarrow \frac{4}{21} + \frac{1}{7} \Rightarrow 1/7\left(\frac{4}{3} + 1\right] \Rightarrow 1/3\left(\frac{7}{3}\right) \Rightarrow \frac{7}{9}$

$\Rightarrow 0.777$

weather	temperature	wind
0.1904 (Minimum)	0.4285	0.777

*) Splitting on the root Node is Weather since it has least gini Index value



Requires No further splitting

Temperature	Wind level	Running
High	Low	No
Medium	Medium	Yes
Low	High	No

Temperature

$Gi(class = 'high') \Rightarrow 1 - 1 \Rightarrow 0$

$Gi(class = 'medium') \Rightarrow 1 - 1 \Rightarrow 0$

Hence it is clear that the decision tree requires No splitting