

Decision tree - A DT is Supervised learning algorithm used for classification and regression. It splits the dataset into subset based on feature value, forming a tree like structure with nodes (question/conditions) and leaves (final decision/predictions)

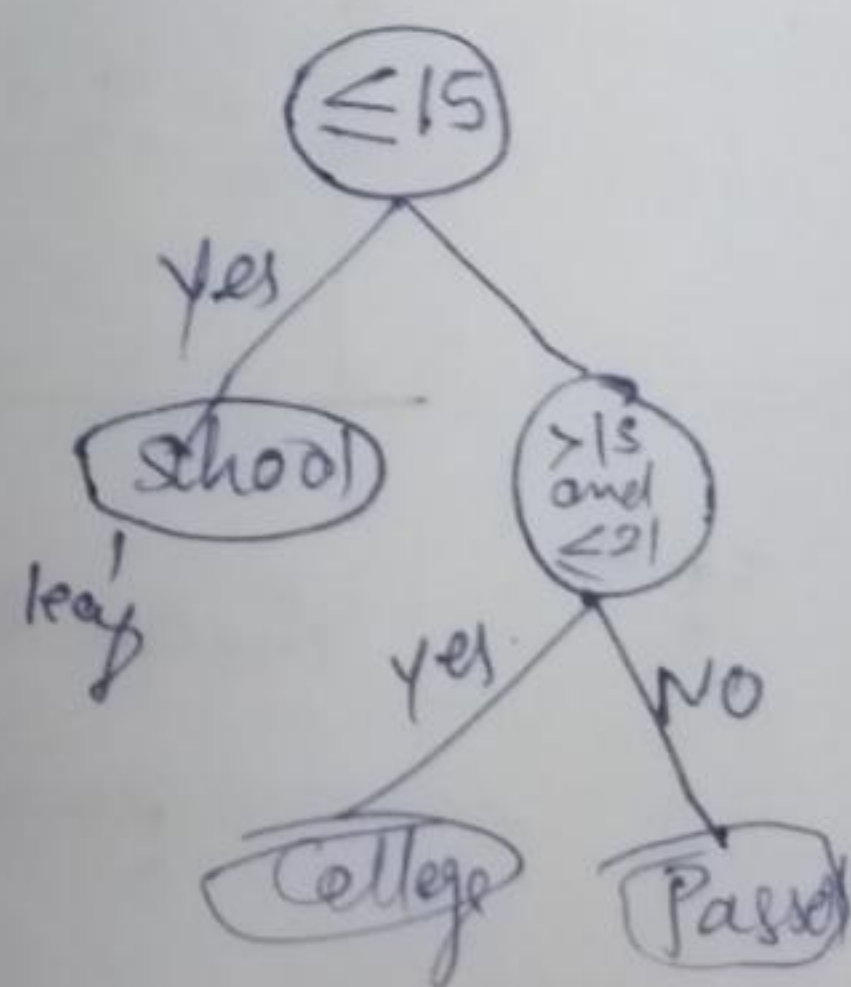
Key terms -

- Root Node - The first node (entire dataset)
- Decision Node - Splitting point based on a feature
- Leaf Node - final Output (Class or value)
- Branches - Path from root to leaf

How it works -

1. Choose the best features to split the data (using Gini Index, Entropy/Information Gain) or Variance Reduction)
2. Create decision nodes and branches.
3. Repeat recursively until Stopping Conditions (pure nodes or max depth.)

eg - if (age ≤ 15)
 Print ("School")
- elif (age > 15 and age ≤ 21):
 Print ("College")
else
 Print ("College Passed")



Information Gain -

$$\text{Gain}(S, f_1) = H(S) - \sum_{v \in \text{val}} \frac{|S_v|}{|S|} H(S_v)$$

$$\begin{aligned} H(S) &= -p + \log_2 p + -p - \log_2 p - \\ &= -\frac{9}{14} \log_2 \left(\frac{9}{14}\right) - \frac{5}{14} \log_2 \frac{5}{14} \\ &\approx 0.94 \end{aligned}$$

$$\begin{aligned} H(C_1) &= -\frac{6}{8} \log_2 \left(\frac{6}{8}\right) - \frac{2}{8} \log_2 \left(\frac{2}{8}\right) \\ H(C_2) &= 0.81 \end{aligned}$$

$$\begin{aligned} \text{Gain}(S, f_1) &\leq 0.94 - \left[\frac{8}{14} \times 0.81 + \frac{6}{14} \times 1 \right] \\ &\leq 0.049 \end{aligned}$$

$$\text{Gain}(S, f_1) \leq 0.052$$

Greater than 0.049
That's why split with

