

**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 (prediction).

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 -

- Choose the best feature to split the data (using Gini Index, Entropy / Information Gain) (on Variance Reduction)

- Create decision nodes and branches.

- Repeat recursively until stopping conditions (pure nodes or max depth.)

e.g - if ( $\text{age} \leq 15$ )

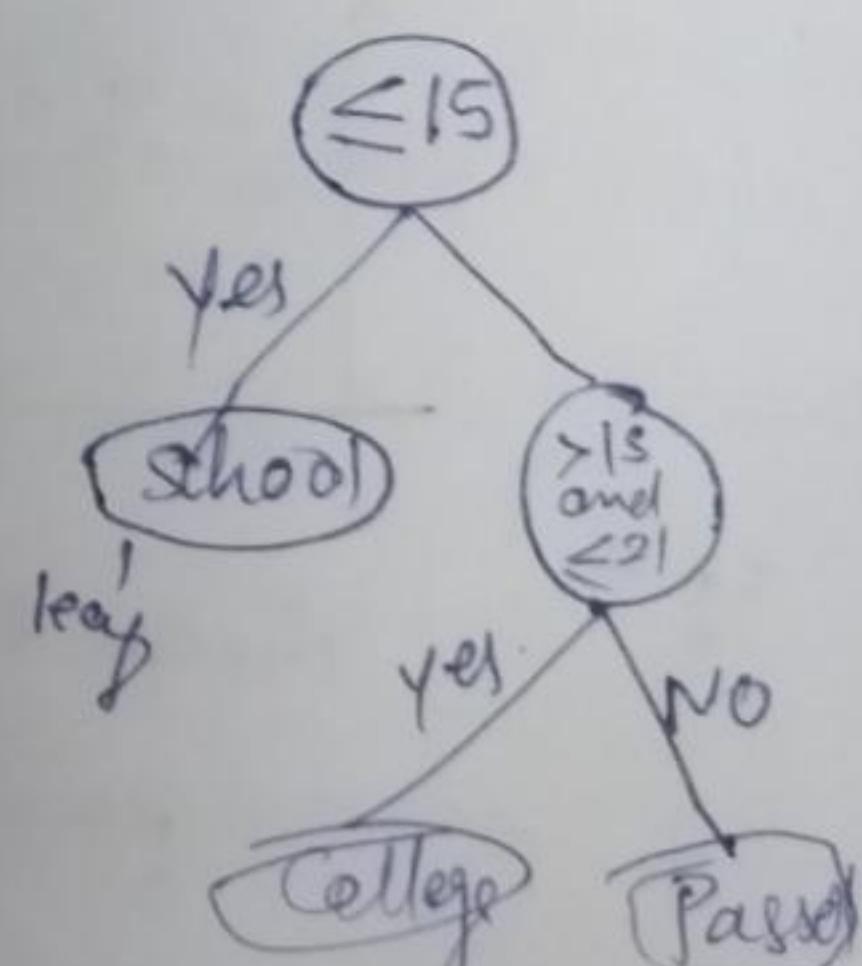
Print("School")

else if ( $\text{age} > 15$  and  $\text{age} \leq 21$ ):

Print ("College")

else

Print ("CollegePassed")



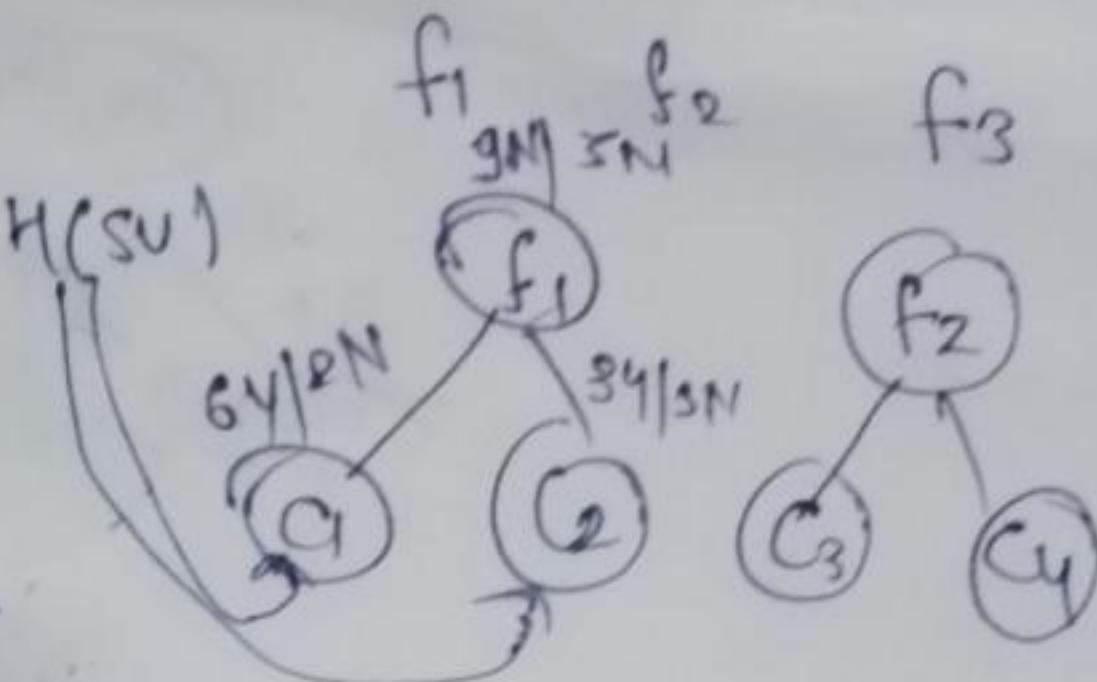
Information gain -

$$\text{Gain}(S, f_1) = H(S) - \frac{\log_2 |S|}{\text{Eval}} H(S|f_1)$$

$$H(S) = -P_1 \log_2 P_1 - P_2 \log_2 P_2$$

$$= \frac{9}{14} \log_2 \left(\frac{9}{14}\right) + \frac{5}{14} \log_2 \frac{5}{14}$$

$$\approx 0.094$$



$$H(c_1) = -\frac{6}{8} \log\left(\frac{6}{8}\right) - \frac{2}{8} \log\left(\frac{2}{8}\right)$$

$$H(c_2) = 0.81$$

$$\text{Gain}(S, f_1) \leq 0.094 - \left[ \frac{8}{14} \times 0.81 + \frac{6}{14} \times 0 \right]$$

$$\leq 0.049$$

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

> greater than 0.049  
that's why split well