Machine Learning Algorithms

① Decision Tree Classifier → [ID3 ✓
                              CART ✓

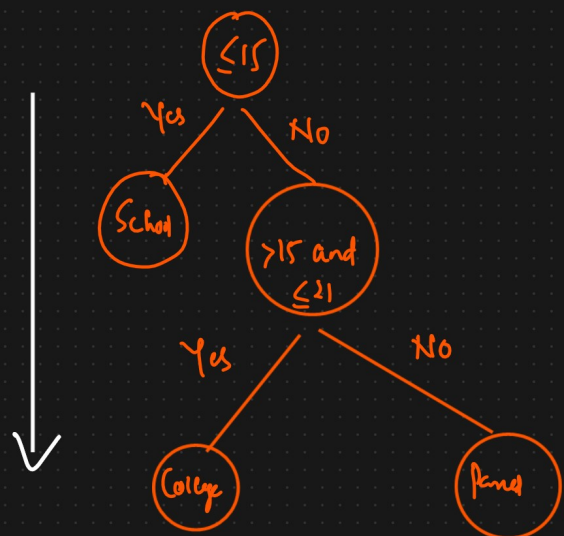a) Entropy And Gini Index → Purity Split

b) Information Gain → feature Decision Tree Split

age = 14

```
if (age ≤ 15):
    Print("The person is in school)

elif (age >15 and age ≤ 21):
    Print (" College)

else
    Print ("Passed College)
```
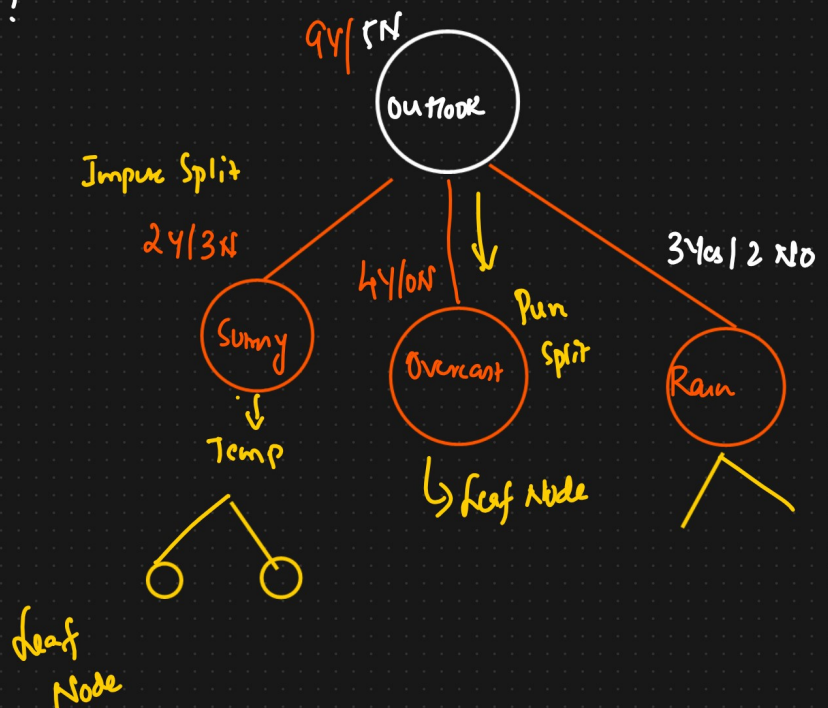

Tennis ??

Outlook	Temp	Humidity	Wind	Play
Sunny ✓	Hot	High	False	No ✓
Sunny	Hot	High	True	No ✓
Overcast ✓	Hot	High	Weak	Yes ✓
Rain ✓	Mild	High	Weak	Yes ✓
Rain	Cool	Normal	False	Yes ✓
Rain	Cool	Normal	True	No ✓
Overcast	Cool	Normal	True	Yes ✓
Sunny	Mild	High	False	No ✓
Sunny	Cool	Normal	False	Yes ✓
Rain	Mild	Normal	False	Yes ✓
Sunny	Mild	Normal	True	Yes ✓
Overcast	Mild	High	True	Yes ✓
Overcast	Hot	Normal	False	Yes ✓
Rain	Mild	High	True	No ✓

9Y / 5N

Outdoor

Impure Split

2Y/3N   4Y/0N   3 Yes / 2 No

Sunny   Overcast   Pure Split   Rain
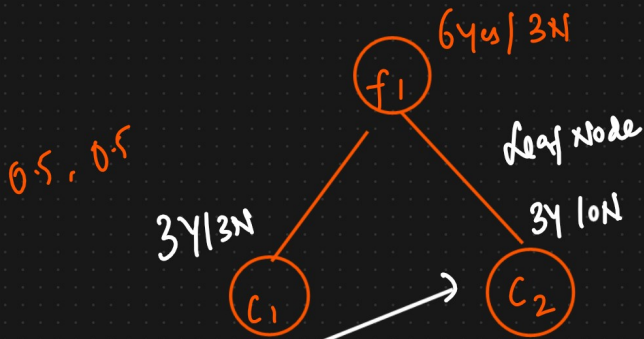
Temp   ↳ Leaf Node

Leaf
Node

① Purity → PureSplit ??

↳ Entropy ✓ }
↳ Gini Impurity } ✓

② Information Gain ⟶ How the features are selected? ✓

① **Entropy**

$$H(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

② **Gini Impurity**

$$G.I = 1 - \sum_{i=1}^{n} (p)^2$$

6Yes/ 3N

(f1)

0.5, 0.5

3Y/3N

Leaf Node

3Y 10N

(C1) ⟶ (C2)

$$H(S) = -\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3}$$

$$= -1 \log_2 1 = 0 \Rightarrow \text{Pure Split}$$

$$H(C_1) = -\frac{3}{6} \log_2 (3/6) - \frac{3}{6} \log_2 (3/6)$$

$$= \boxed{1} \Rightarrow \text{Impure Split}$$

Entropy

$H(S)$

1

0.5

Gini Impurity

0   0.5   1 P^+/P^-

$P^+ = 0.5$

$P^- = 1 - 0.5$
$= 0.5$

Impure Split

5Y/3N

3Y 10N
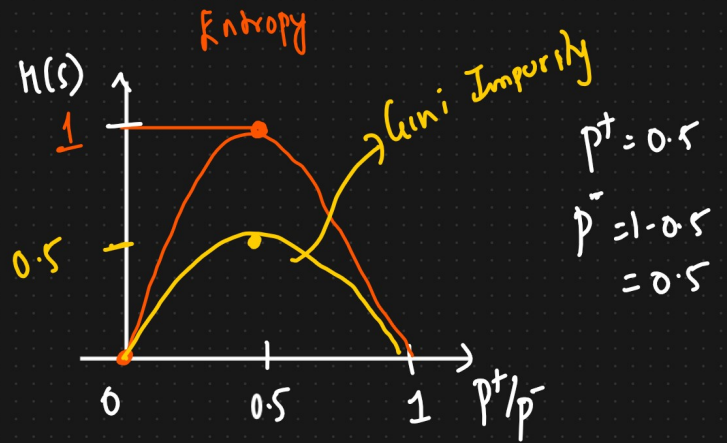
② **Gini Impurity**

$$G.I = 1 - \sum_{i=1}^{n} (p)^2$$

$$= 1 - ((P_+)^2 + (P_-)^2)$$

$$= 1 - ((\tfrac{1}{2})^2 + (\tfrac{1}{2})^2)$$

$$= 1 - \frac{1}{2} = \boxed{0.5}$$

$$= 1 - ((1)^2 + 0)$$

$$= 0$$

$f_1 \qquad f_2 \qquad f_3 \qquad$ { Information Gain }

Information Gain

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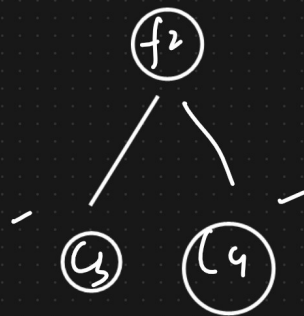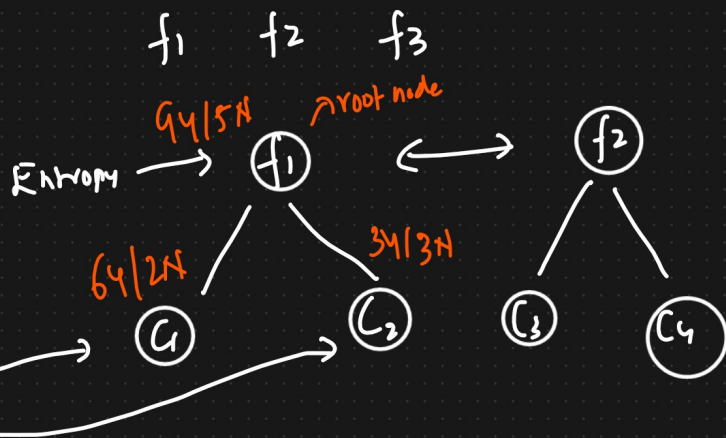
→ Entropy of Root Node

$$H(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

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

$$\approx 0.94$$

$$\text{Gain}(S, f_1) = 0.94 - \left[\frac{8}{14} \times 0.81 + \frac{6}{14} \times 1 \right]$$

$$= 0.049$$

$$\text{Gain}(S, f_2) = 0.052$$

$f_1 \qquad f_2 \qquad f_3$

Entropy →

$$H(r_1) = -\frac{6}{8} \log_2 \left(\frac{6}{8}\right) - \frac{2}{8} \log_2 (2/8)$$

$$= 0.81$$

$$H(2) = 1$$