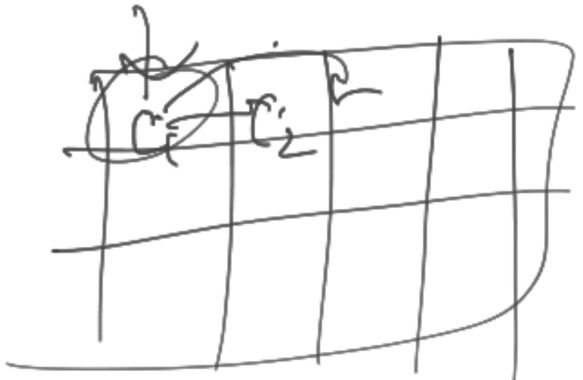
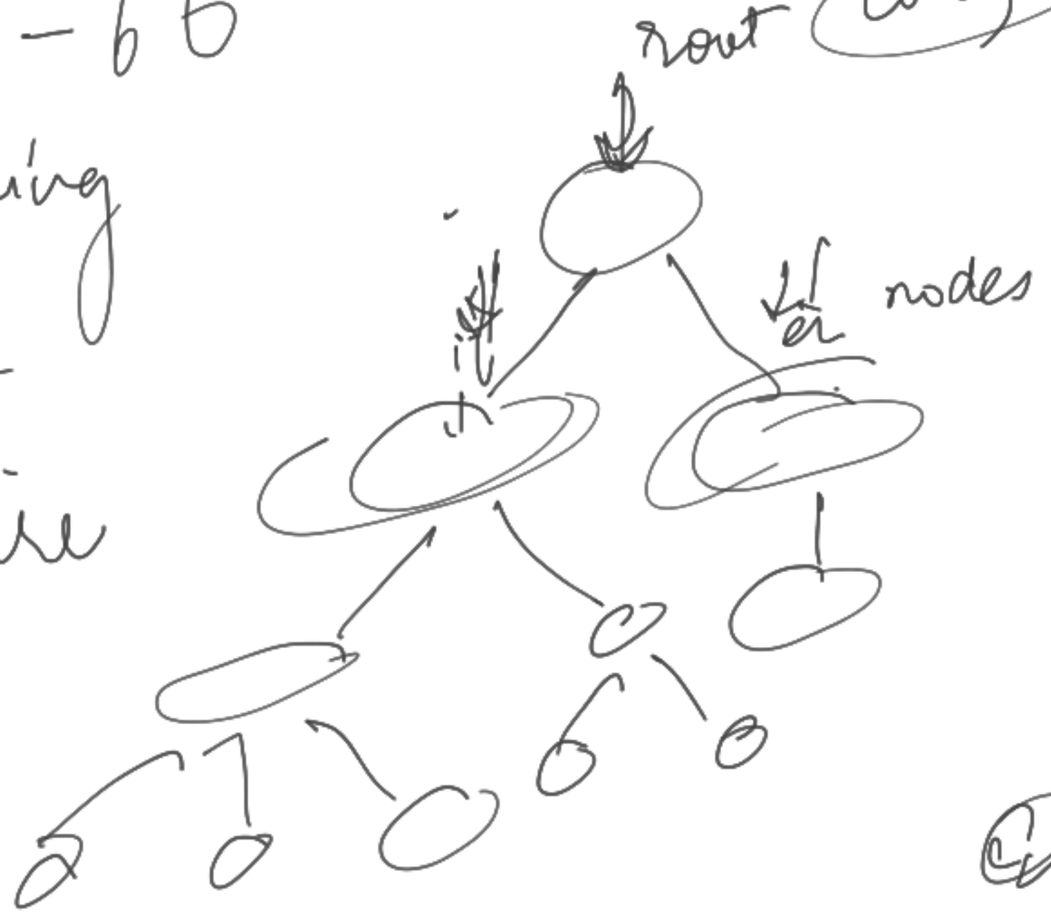
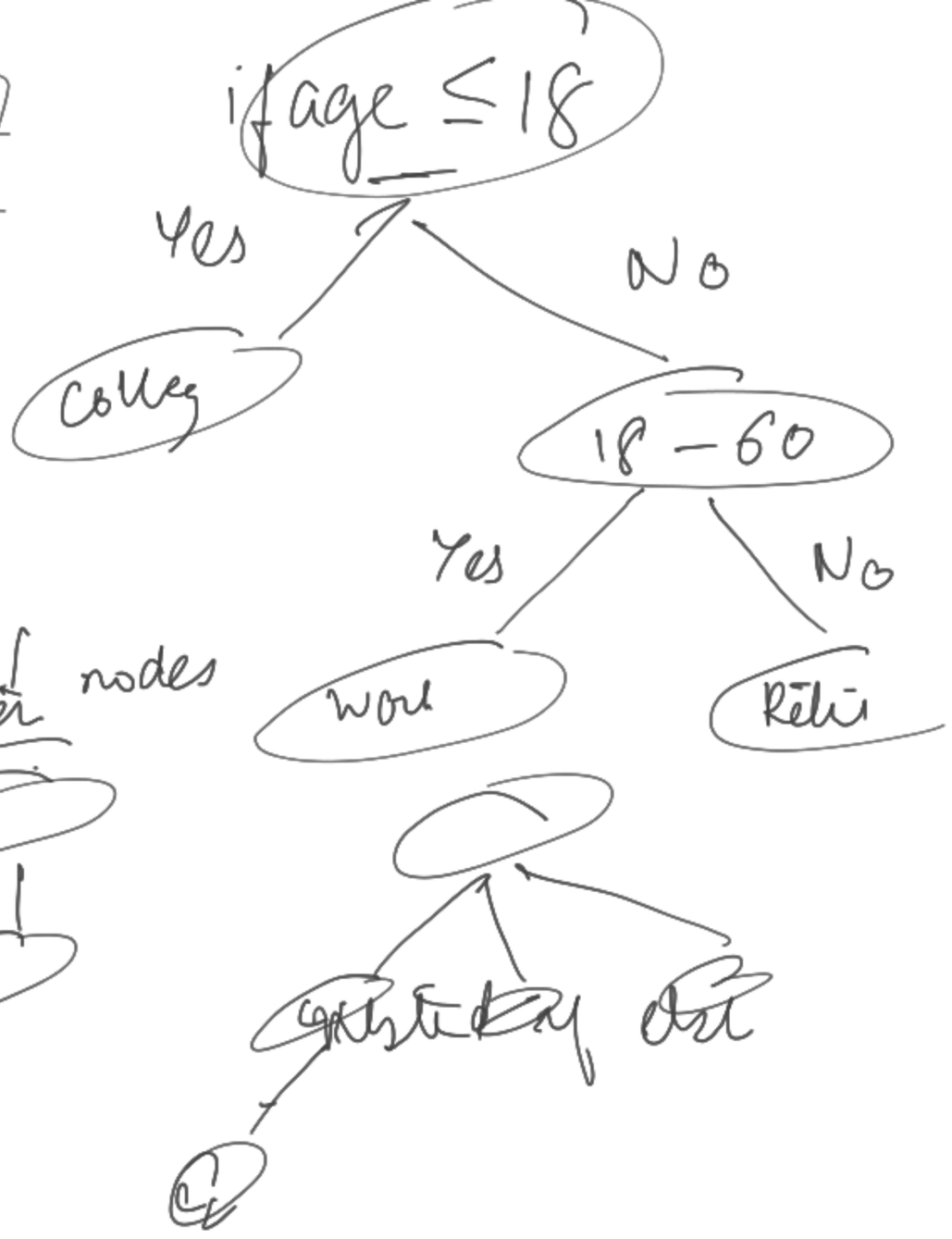
≤ 18 :
college

if 18 - 60
working

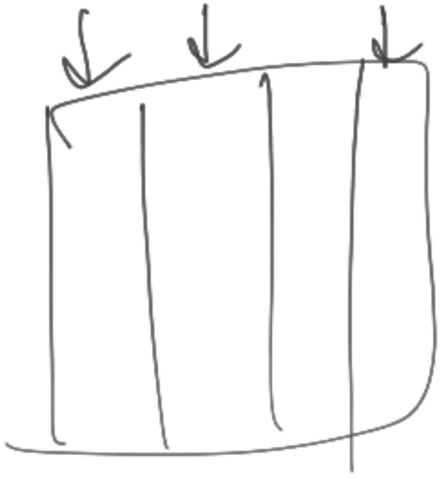
60 +
retire







Impure split :



pure split

Entropy

Gini Impurity

measures of Impurity

Entropy $H(S)$ 4+ 4-

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

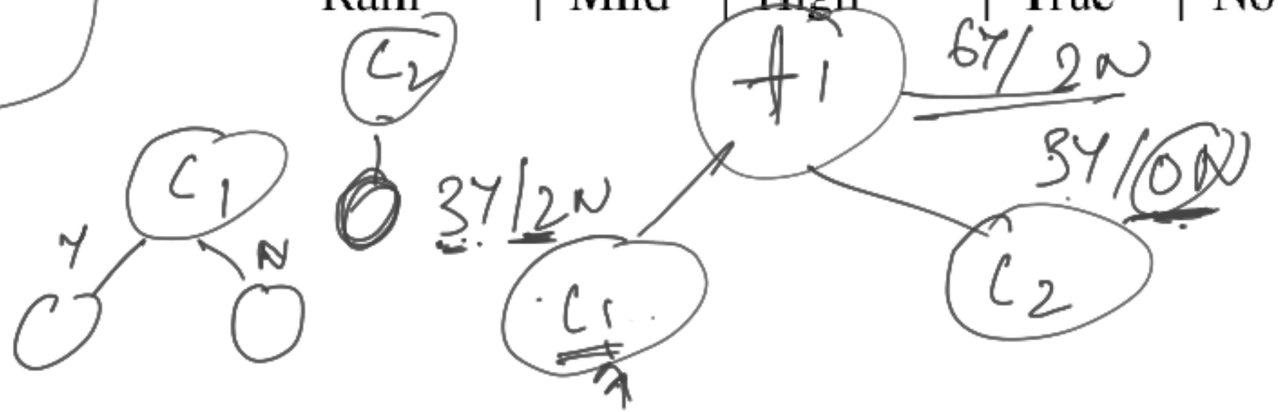
$= -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right)$

$= 0.97$

$C_2 = \left[-\frac{3}{3}\log_2\left(\frac{3}{3}\right)\right] - 0$

$= -\log_2(1) = 0$

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



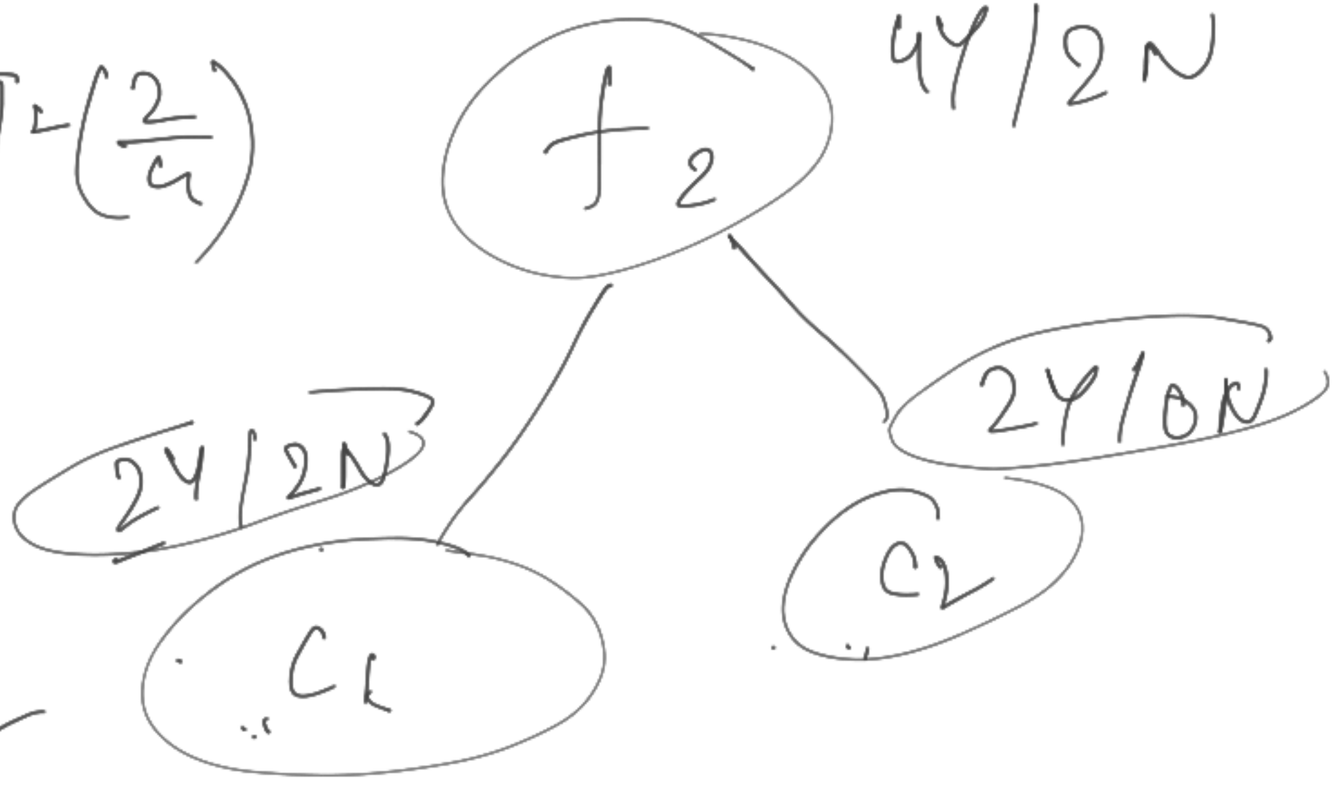
$$H(S) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)$$

$$= 1$$

$$H(S) = -\frac{2}{2}\log_2\left(\frac{2}{2}\right)$$

$$\therefore -1\log(1) = 0$$

entropy $\Rightarrow$ 0 − 1



0.78 0.55

1 0.4 0.2

Gini Index

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

$$= 1 - \left[(p_+)^2 + (p_-)^2 \right]$$

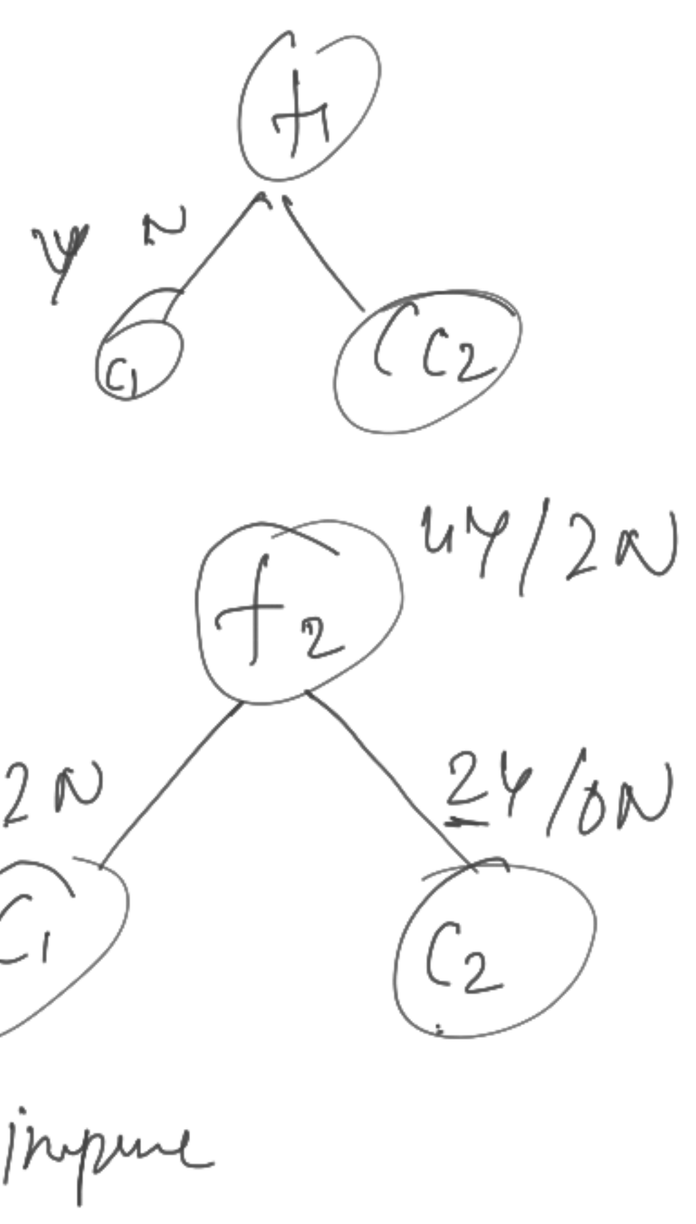
$$G.I_{C_1} = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2 \right]$$

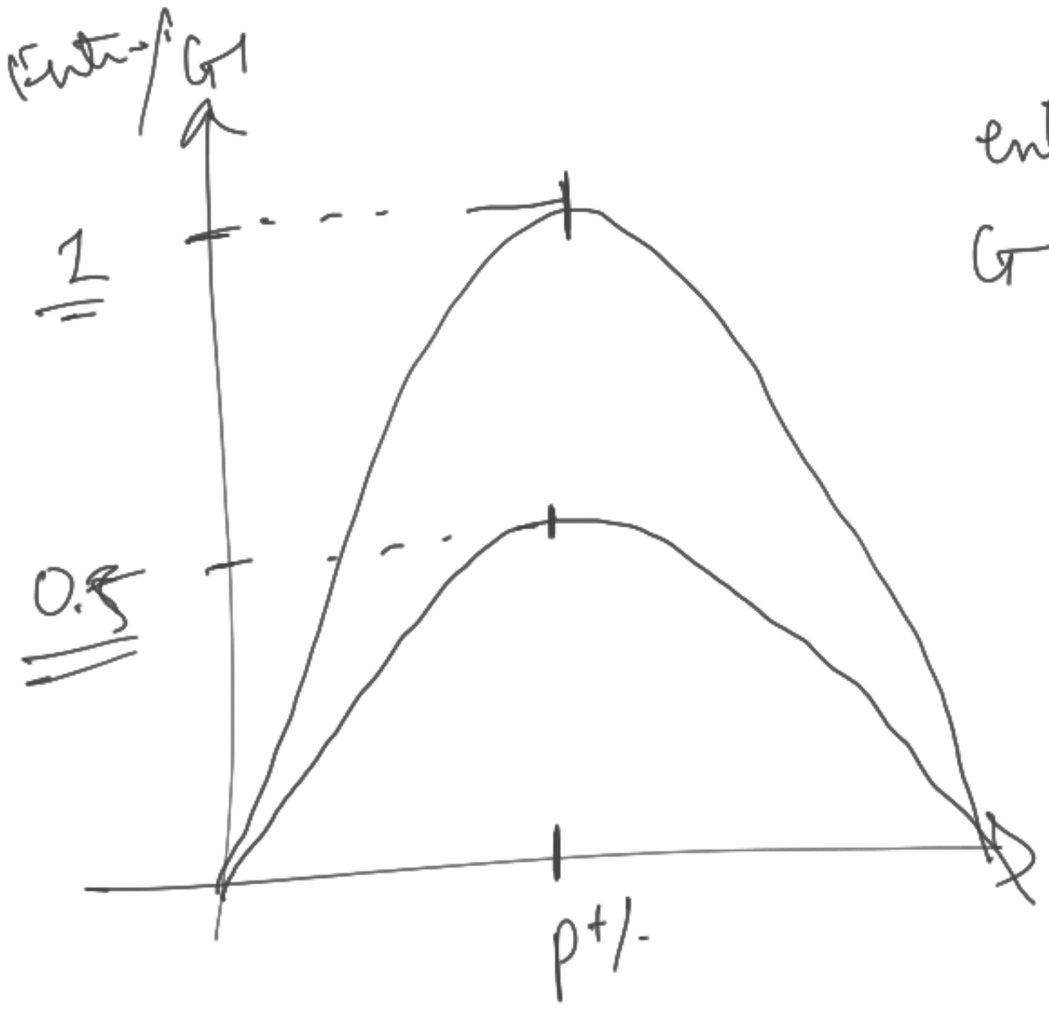
$= 1/4$

$= 0.25 + 0.25$

$= 1 - 0.5 = \boxed{0.5} > 1$

$$GI_{C_2} = 1 - \left[\left(\frac{2}{2}\right)^2 \right] = 0 > \underline{0}$$



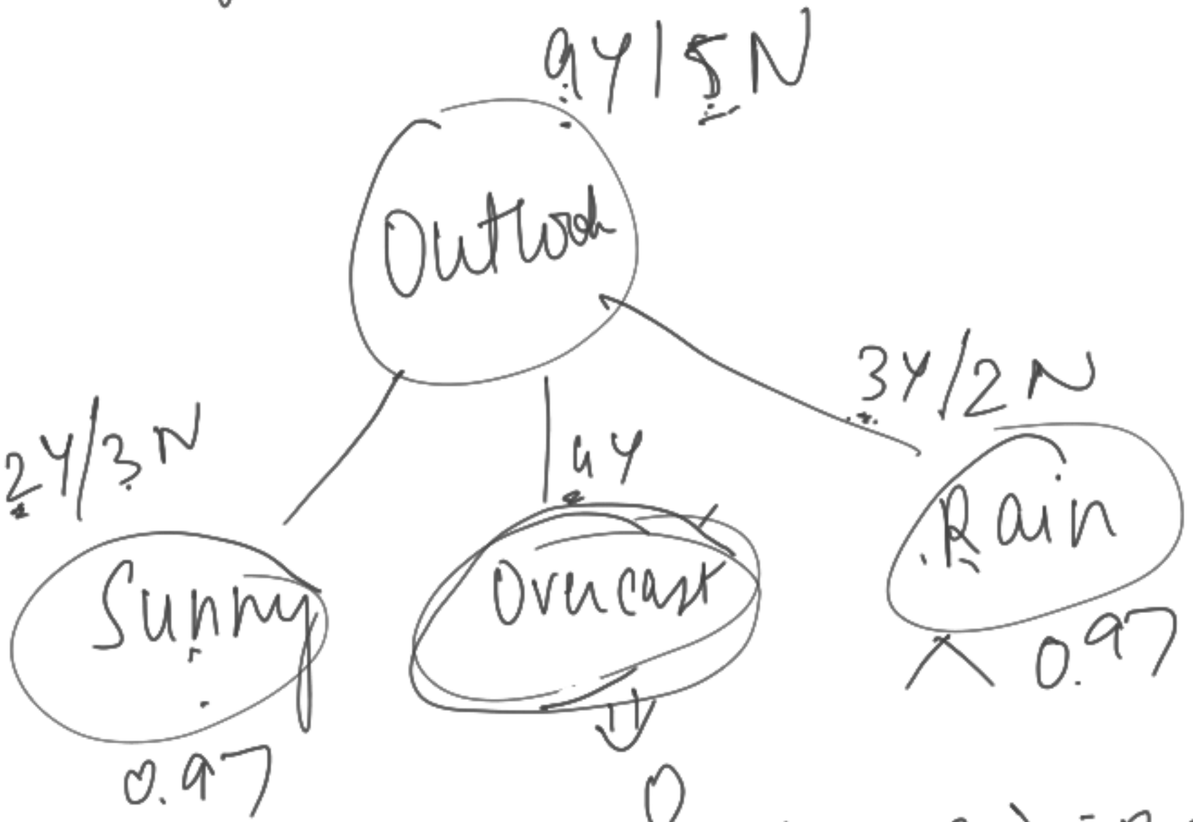


entropy : 0 - 1

GI : 0 - 0.5

Entropy $H(S) =$

$$-p_1 \log_2(p_1) - p_- \log_2(p_-)$$



$$= -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.97$$

$$= -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right)$$

$$= 0.97$$

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

Information Gain : feature selection

$$I.G(S, f_i) = H(S) - \sum_{v \in val} \frac{|S_v|}{|S|} H(S_v)$$