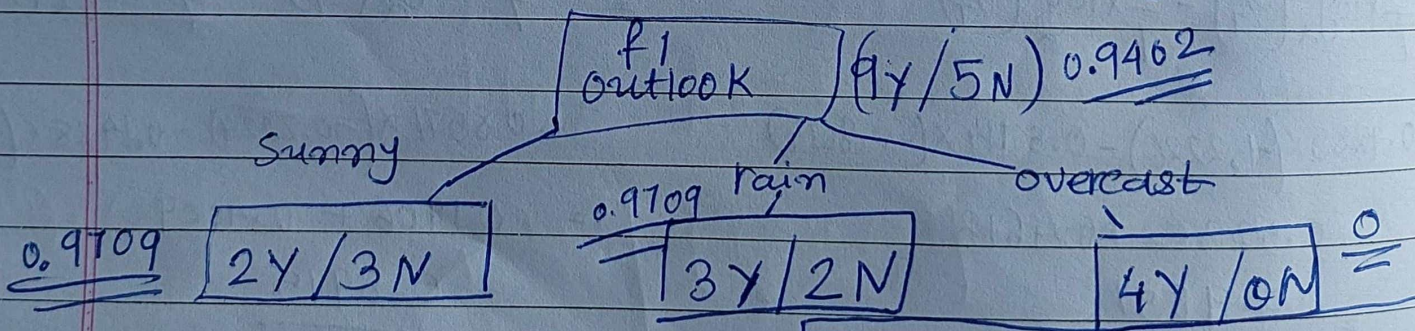


Darshita Paaghadal

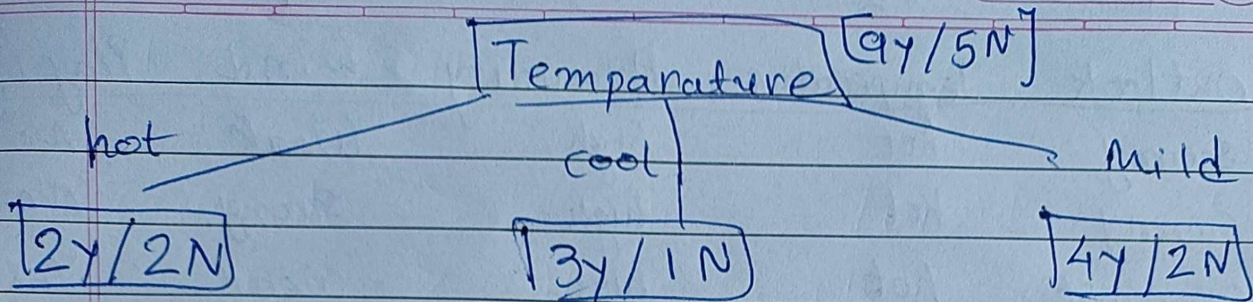
DATE: / /

PAGE NO: P4

f_1	f_2	f_3	f_4	
outlook	Temperature	Humidity	Wind	Play (Yes/No)
Sunny	hot	high	Weak	No
Sunny	hot	high	Strong	No
Overcast	hot	high	Weak	Yes
Rain	Mild	high	Weak	Yes
Rain	cool	Normal	Weak	Yes
Rain	cool	Normal	Strong	No
Overcast	cool	Normal	Strong	Yes
Sunny	Mild	high	Weak	No
Sunny	cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	high	Strong	Yes
Overcast	hot	Normal	Weak	Yes
Rain	Mild	high	Strong	No



$\frac{9}{14}$



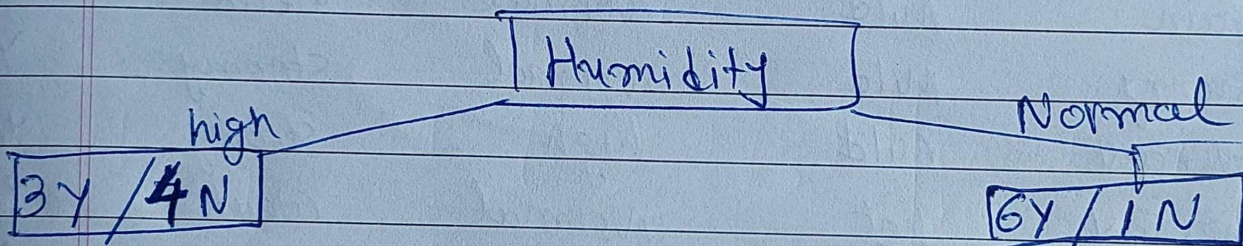
Entropy = 1

$$= -\frac{3}{4} \times \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \times \log_2\left(\frac{1}{4}\right) = -\frac{4}{6} \times \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \times \log_2\left(\frac{2}{6}\right)$$

$$= -0.75 \times (-0.41) - 0.25 \times (-2) = 0.6667 \times (-0.5848) - 0.33 \times (-1.5851)$$

$$= 0.3075 + 0.5 = 0.3899 + 0.5230$$

$$= \underline{\underline{0.8075}} = \underline{\underline{0.9129}}$$



$$= -\frac{3}{7} \times \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \times \log_2\left(\frac{4}{7}\right) = -\frac{6}{7} \times \log_2\left(\frac{6}{7}\right) - \frac{1}{7} \times \log_2\left(\frac{1}{7}\right)$$

$$= 0.4285 \times (-1.2226) - 0.5714 \times (-0.8074) = -0.8571 \times (-0.2224) - 0.1428 \times (-2.8079)$$

$$= 0.5238 + 0.4613 = 0.1906 + 0.4009$$

$$= \underline{\underline{0.9851}} = \underline{\underline{0.5915}}$$

Entropy = 1

Wind $\left[\frac{97}{5N} \right] 0.9462$

Strong $\left[\frac{37}{3N} \right]$ weak $\left[\frac{67}{2N} \right]$

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

$$= -0.75(-0.4150) - 0.25(-2.0)$$

$$= 0.3112 + 0.5$$

$$= 0.81125$$

Info. Gain [Wind] = $0.9462 - \left[\frac{6}{14} \times 1 + \frac{8}{14} \times 0.81 \right]$

$$= 0.94 - [0.462 + 0.42] = 0.0485 \Rightarrow \boxed{0.049}$$

Info Gain [humidity] = $0.9462 - \left[\frac{1}{14} \times 0.98 + \frac{7}{14} \times 0.59 \right]$

$$= 0.9462 - [0.49 + 0.295] = \boxed{0.1612}$$

Info Gain [Temperature] = $0.9462 - \left[\frac{4}{14} \times 1 + \frac{4}{14} \times 0.8075 + \frac{6}{14} \times 0.9129 \right]$

$$= 0.9462 - [0.2857 + 0.2307 + 0.3912]$$

$$= 0.9462 - 0.9076 \Rightarrow \boxed{0.0386}$$

Info Gain [Outlook] = $0.9462 - \left[\frac{5}{14} \times 0.9709 + \frac{5}{14} \times 0.9709 + \frac{4}{14} \times 0 \right]$

$$= 0.9462 - [0.3571 \times 0.9709 + 0.3571 \times 0.9709]$$

$$= 0.9462 - [0.34675 + 0.34675] \Rightarrow \boxed{0.2527}$$

DATE: / / PAGE NO.

Info Gain based on Entropy :-

Outlook > humidity > Wind > Temperature
 $0.25 > 0.16 > 0.049 > 0.038$

So Outlook will be root features of my Decision Tree [ID3]

Gini Coeff / Gini Impurity :-

$$= 1 - \sum_{i=1}^n (P_i)^2$$

f₁

$\begin{aligned} \text{Outlook}_{\text{sunny}} &= 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2 \right] \\ &> 1 - [0.16 + 0.36] \\ &= \underline{0.68} \end{aligned}$	$\begin{aligned} \text{Outlook}_{\text{rain}} &= 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2 \right] \\ &= 1 - [0.16 + 0.36] \\ &= \underline{0.68} \end{aligned}$
--	---

f₂

$$\text{Outlook}_{\text{overcast}} = 1 - \left[\left(\frac{4}{4}\right)^2 + \left(\frac{0}{4}\right)^2 \right] = \underline{0}$$

$\begin{aligned} \text{Temperature}_{\text{hot}} &= 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2 \right] \\ &= 1 - 1 = \underline{0} \end{aligned}$	$\begin{aligned} \text{Temperature}_{\text{cool}} &= 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2 \right] \\ &= 1 - [0.56 + 0.06] = \underline{-0.1225} \end{aligned}$
---	--

$$\begin{aligned} \text{Temperature}_{\text{mild}} &= 1 - \left[\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2 \right] \\ &= 1 - [0.44 + 0.1111] \\ &= 1 - 0.55 \\ &= \underline{0.44} \end{aligned}$$

f3

$$\begin{aligned} \text{Humidity} &= 1 - \left[\left(\frac{3}{7} \right)^2 + \left(\frac{4}{7} \right)^2 \right] \\ \text{high} &= 1 - (0.1836 + 0.3265) \\ &= \underline{0.4898} \end{aligned}$$

$$\begin{aligned} \text{Humidity} &= 1 - \left[\left(\frac{6}{7} \right)^2 + \left(\frac{1}{7} \right)^2 \right] \\ \text{Normal} &= 1 - (0.7346 + 0.020) \\ &= \underline{0.2454} \end{aligned}$$

f4

$$\begin{aligned} \text{Wind} &= 1 - \left[\left(\frac{3}{6} \right)^2 + \left(\frac{3}{6} \right)^2 \right] \\ \text{Strong} &= 1 - 1 = \underline{0} \end{aligned}$$

$$\begin{aligned} \text{Wind} &= 1 - \left[\left(\frac{6}{8} \right)^2 + \left(\frac{2}{8} \right)^2 \right] \\ \text{Weak} &= 1 - (0.5625 + 0.0625) \\ &= \underline{0.375} \end{aligned}$$

Impurity \Rightarrow Information Gain \rightarrow Calculation of all Gini

$$f_1 \rightarrow \text{outlook} \rightarrow 0.68 + 0.68 + 0 = 1.36$$

$$f_2 \rightarrow \text{Temperature} \rightarrow 0 + (-0.1225) + 0.44 = 0.3175$$

$$f_3 \rightarrow \text{Humidity} \rightarrow 0.4898 + 0.2454 = 0.7352$$

$$f_4 \rightarrow \text{Wind} \rightarrow 0 + 0.375 = 0.375$$

∴ f2 (Temperature) features has less impurity
So Temperature will be root features
for CART (Decision Tree)