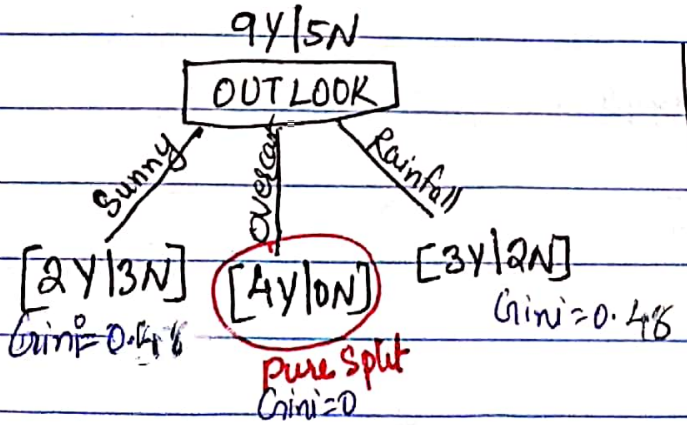
Data Set

	Outlook	Temperature	humidity	wind	Decision
1	Sunny	hot	high	weak	No
2	Sunny	hot	high	Strong	No
3	Overcast	hot	high	weak	Yes
4	Rainfall	mild	high	weak	Yes
5	Rainfall	cool	normal	weak	Yes
6	Rainfall	cool	normal	Strong	No
7	Overcast	cool	normal	strong	Yes
8	Sunny	mild	high	weak	No
9	Sunny	cool	normal	weak	Yes
10	Rainfall	mild	normal	weak	Yes
11	Sunny	mild	normal	Strong	Yes
12	Overcast	mild	high	strong	Yes
13	Overcast	hot	normal	weak	Yes
14	Rainfall	mild	high	Strong	No

Decide the Root node

try → outlook, Temperature, humidity and wind.



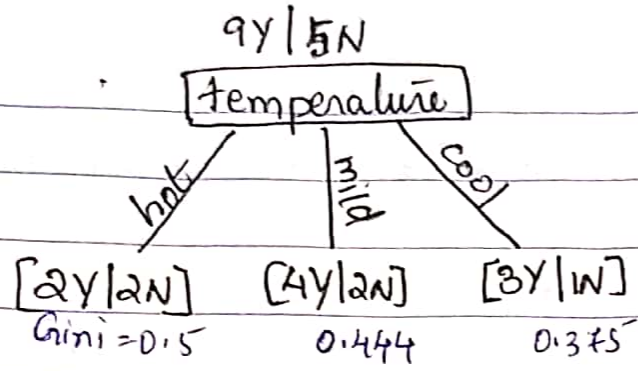
$Gini_{(S)} = 1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2 = 0.48.$

$Gini_{(R)} = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 = 0.48$

weighted avg of Gini $= \frac{5}{14} \times 0.48 + 0 + \frac{5}{14} \times 0.48$

$= 0.$

$= 0.343$



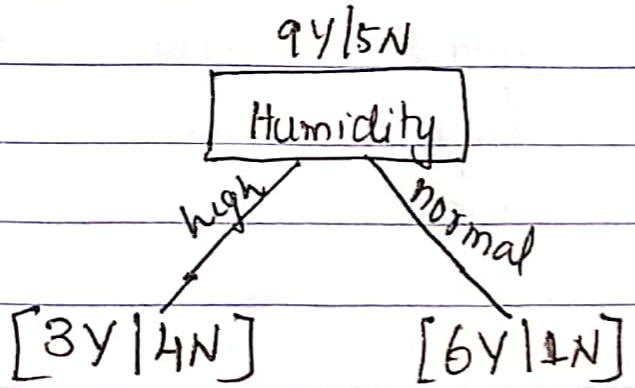
wt avg $= \frac{4}{14} \times 0.5 + \frac{6}{14} \times 0.444 + \frac{4}{14} \times 0.375$

$= 0.202$

$$Gini_{(H)} = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.5$$

$$Gini_{(M)} = 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 = 0.444$$

$$Gini_{(C)} = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0.375$$

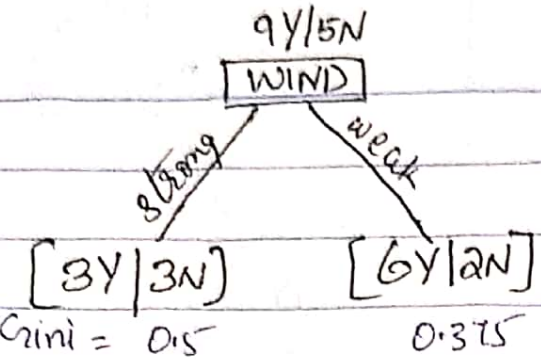


wt-avg Gini $= \frac{7}{14} \times 0.49 + \frac{7}{14} \times 0.245$

$= 0.183$

$$Gini_{(H)} = 1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2 = 0.49$$

$$Gini_{(N)} = 1 - \left(\frac{6}{7}\right)^2 - \left(\frac{1}{7}\right)^2 = 0.245$$



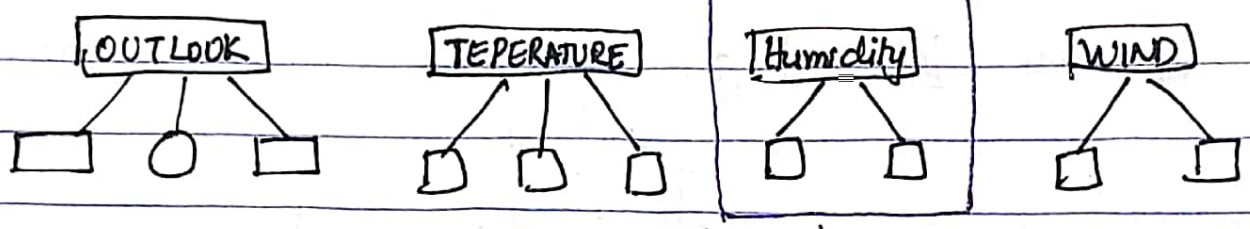
Wt. avg Gini $= \frac{6}{14} \times 0.5 + \frac{8}{14} \times 0.375$

$= 0.295$

$Gini(S) = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0.5$

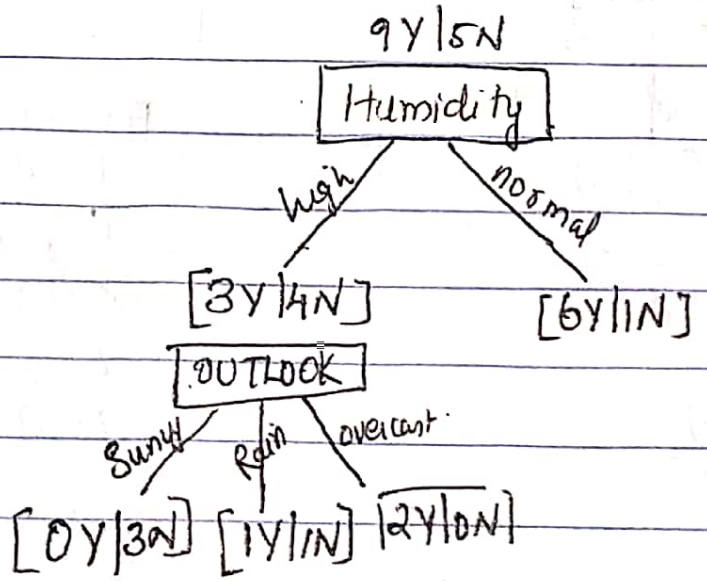
$Gini(W) = 1 - \left(\frac{6}{8}\right)^2 - \left(\frac{2}{8}\right)^2 = 0.375$

Gini → 0.343 0.202 0.183 0.295



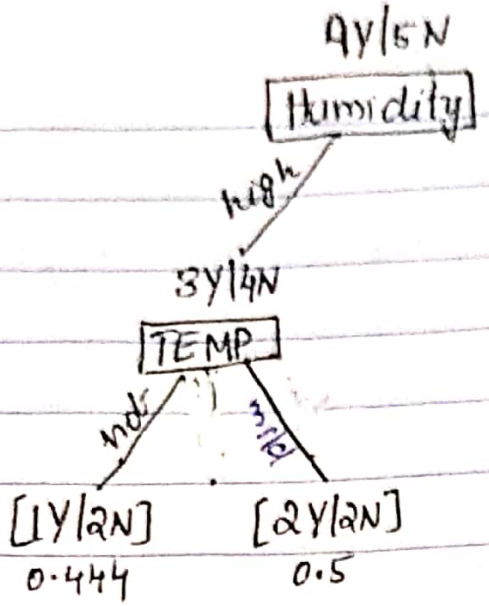
→ it has minimum Gini Impurity

→ So select Humidity as root



Gini 0 0.5 0

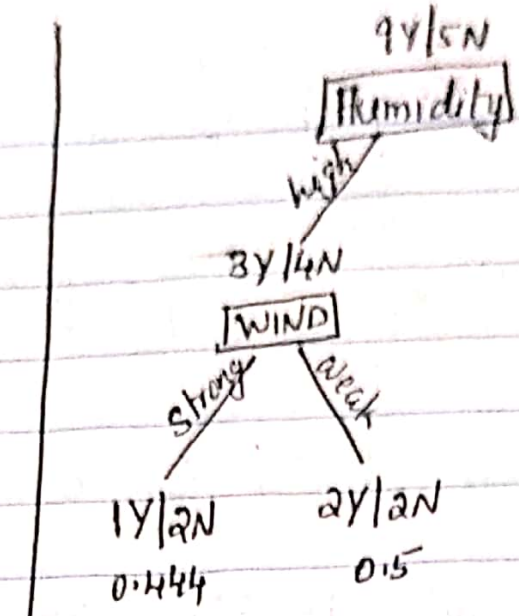
Wt-avg Gini $\frac{2}{7} \times 0.5 = 0.143$



Gini 0.444 0.5

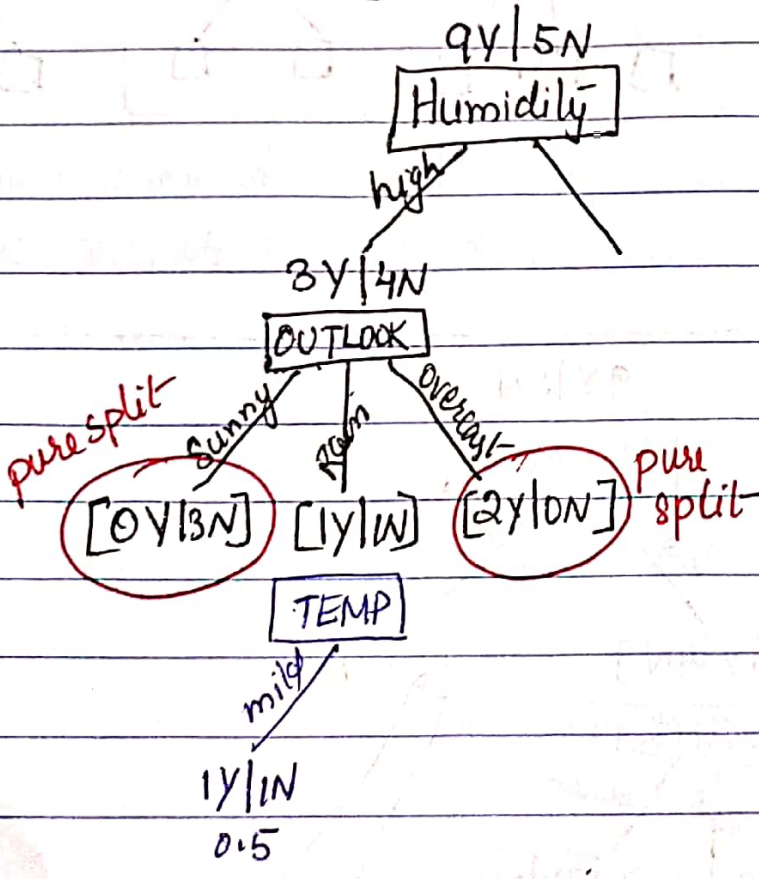
wt. avg $= \frac{3}{7} \times 0.444 + \frac{4}{7} \times 0.5$

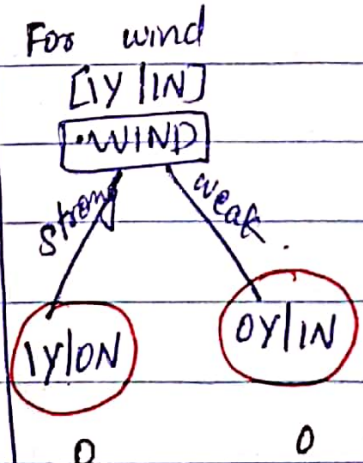
$= 0.476$.



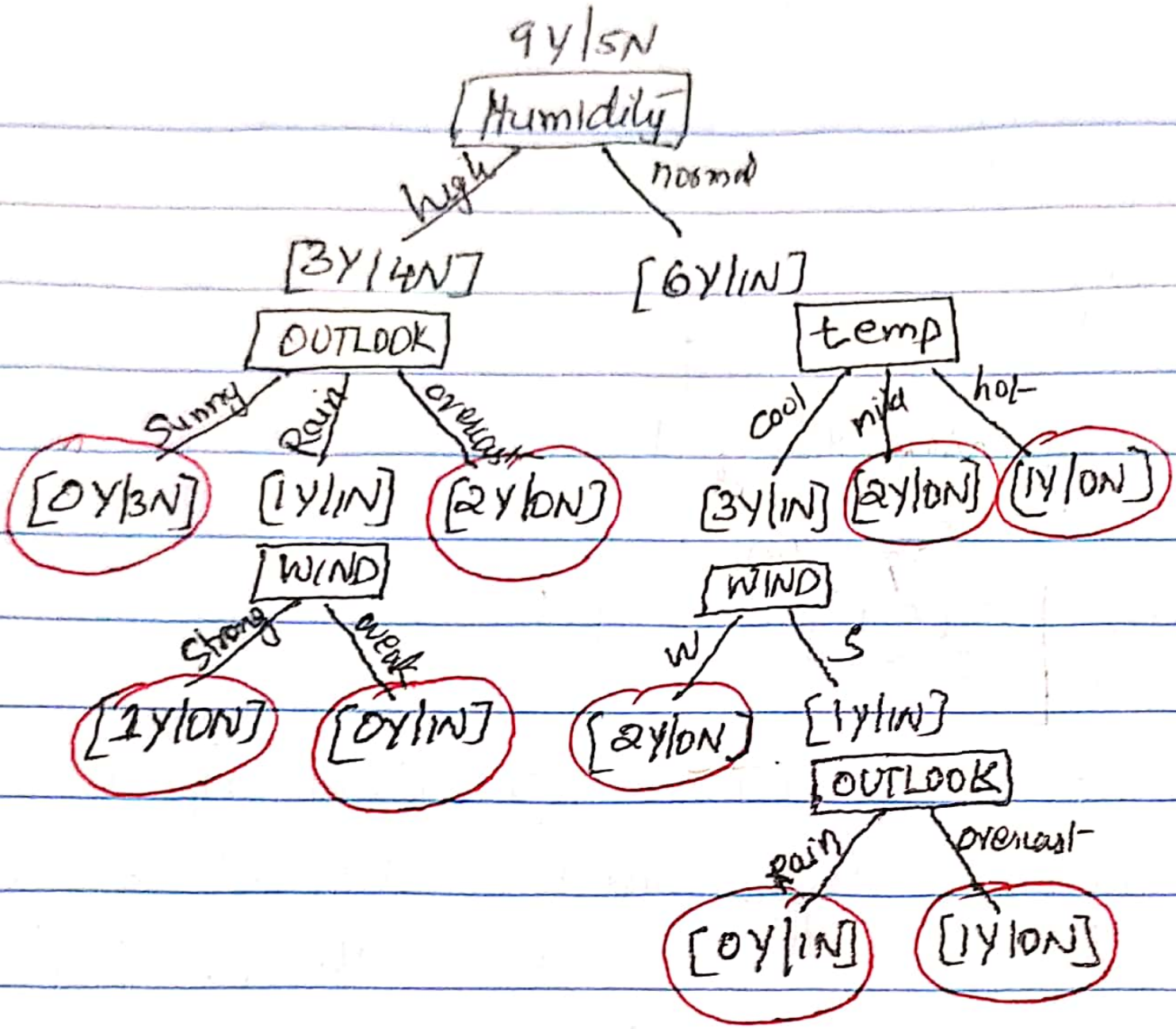
wt. avg = 0.476

Since OUTLOOK gives minimum Gini we choose it





Since WIND has minimum Gini (impurity), choose it.



Entropy Vs Gini impurity

Entropy	Gini Impurity
(1) Value ranges between 0-1	(1) Value ranges between 0-0.5
(2) $H(S) = -\sum_j P_j \log(P_j)$	(2) Gini imp $= 1 - \sum_j (P_j)^2$
(3) The compatation is complex since it includes logarithms	(3) Gini imp computation is straight forward and faster than H(S)
(4) The results obtained is better	(4) Results obtained is not as good as of entropy
(5) Training is time consuming	(5) Training is faster than H(S) method.