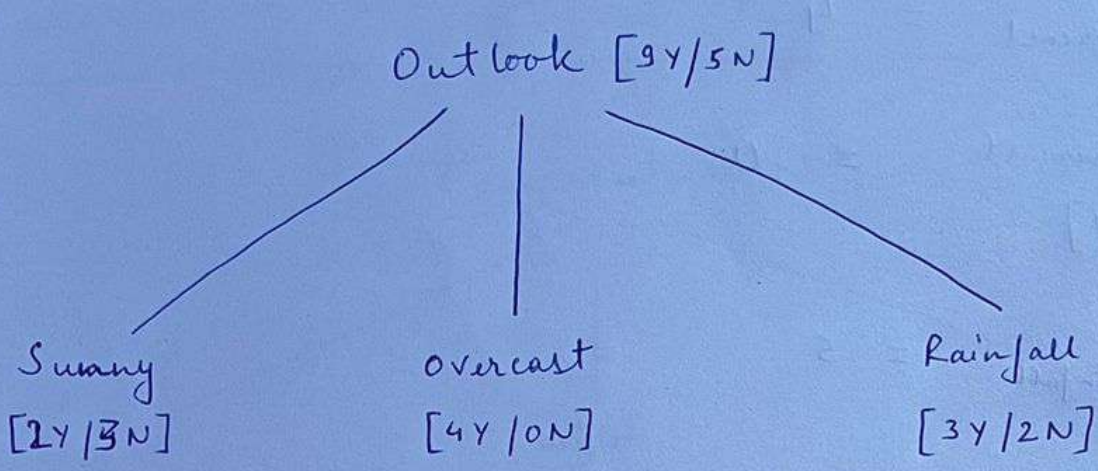
Day	Outlook	Temperature	Humidity	Wind	Decision
1	Sunny	Hot	High	weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	yes
4	Rainfall	Mild	High	Weak.	yes
5	Rainfall	Cool	Normal	Weak	yes
6	Rainfall	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	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



Information gain

$$= \boxed{H(s) - \sum \frac{|S_v|}{|S|} \times H(S_v)}$$

$H(s)_{\text{outlook}}$ = Root feature Entropy of outlook [9Y/5N]

$$= -P_Y \times \log_2(P_Y) - P_N \times \log_2(P_N)$$

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

$$= 0.94$$

$|S_v|_{\text{sunny}} = 5$, $|S| = 14$

$H(S_v)_{\text{sunny}}$ [2Y/3N]

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

$$= -\frac{2}{5} \times (-1.32) - \frac{3}{5} \times (-0.737)$$

$$= .528 + .442$$

$$= 0.97$$

$|S_v|_{overcast} = 4$

$H(S_v)_{overcast} = 0$

[4Y/0N]

$|S_v|_{rainfall} = 5$

$H(S_v)_{rainfall} = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)$

[3Y/2N]

$= 0.97$

Putting all above values in the I.G formulae $= H(S) - \sum \frac{|S_v|}{|S|} \times H(S_v)$

$= .94 - \left[\frac{5}{14} \times .97 + \frac{4}{14} \times 0 + \frac{5}{14} \times .97\right]$

$= .94 - \left[0.357 \times 0.97 + 0.357 \times 0.97\right]$

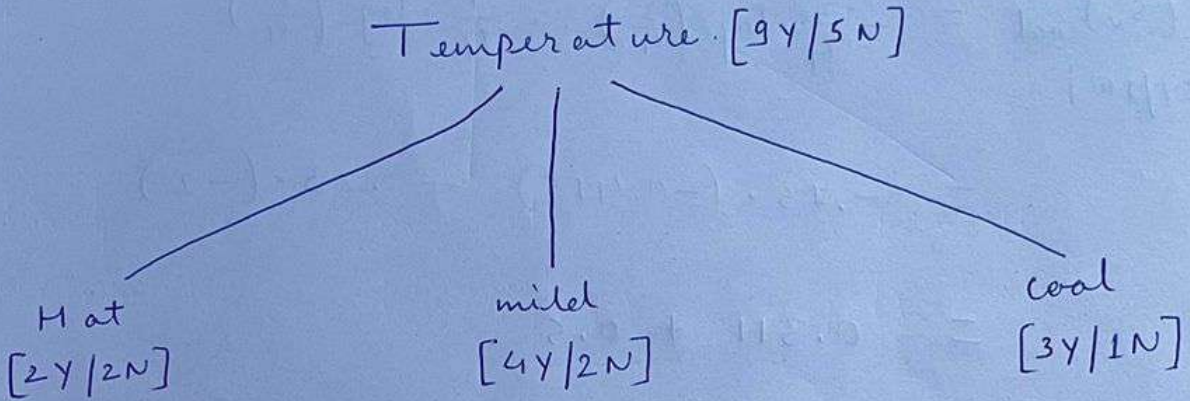
$= .94 - \left[0.346 + 0.346\right]$

$= .94 - .692$

$= 0.248$

$\Rightarrow$ $I.G_{outlook} = 0.248$

Calculating I.G for temperature.



$H(S)_{Temp.}$ for [9Y/5N] is same as of out look.

$|S_v|_{Hot} = 4$ | $|S| = 14$

$H(S_v)_{Hot} = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)$

$[2Y/2N] = 1$

$|S_v|_{mild} = 6$

$H(S_v)_{mild} = -\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)$

$[4Y/2N]$

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

$= -0.66 \times (-0.584) - 0.33 \times (-1.584)$

$= 0.385 + 0.522$

$= 0.907$

$|S_v|_{coal} = 4$

$H(S_v)_{coal} = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)$

[3Y/1N]

$= -.75 \times (-0.415) - .25 \times (-2)$

$= 0.311 + 0.5$

$= 0.811$

Putting above value in I.G formulae

$I.G_{Temp} = .94 - \left[\frac{4}{14} \times 1 + \frac{6}{14} \times .907 + \frac{4}{14} \times .811\right]$

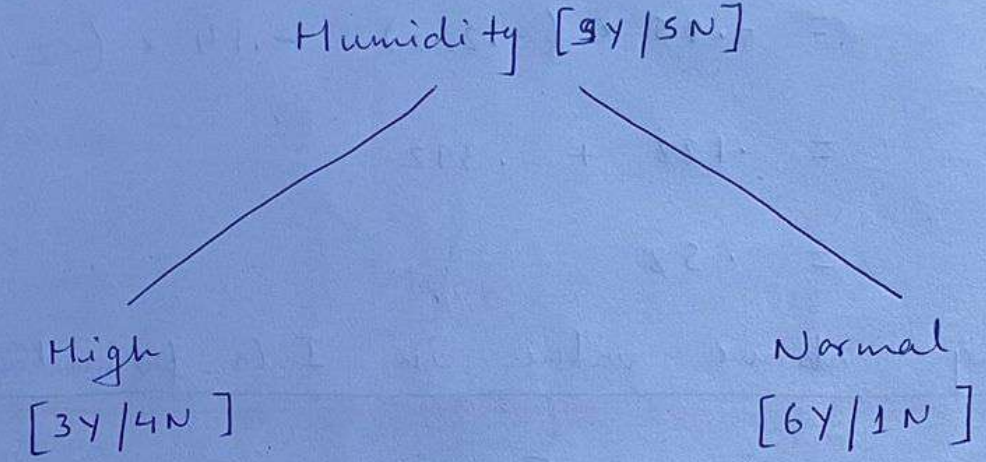
$= .94 - [.285 + .388 + .231]$

$= .94 - .904$

$= 0.036$

$\Rightarrow$ $I.G_{Temp} = 0.036$

Calculating I.G for Humidity.



$H(S)_{\text{Humidity}} = .94$
[9Y/5N]

$|S_v|_{\text{High}} = 7$

$H(S_v)_{\text{High}} = -\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right)$

$= -.428 \times (-1.22) \quad -.571 \times (-.807)$

$= .522 + .460$

$= .982$

$|S_v|_{\text{Normal}} = 7$

$$H(S_v)_{Normal} = -\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right)$$

$$= -.85 \times (-.222) \quad -.14 \times (-2.8)$$

$$= .188 + .392$$

$$= .58$$

Putting above value in I.G formula

$$I.G_{Humidity} = .94 - \left[\frac{7}{14} \times .982 + \frac{7}{14} \times .58\right]$$

$$= .94 - \left[.491 + .29\right]$$

$$= .94 - .781$$

$$= .159$$

$\Rightarrow$ **$I.G_{Humidity} = .159$**

Calculating I.G for wind.;

wind [9Y/5N]

- Weak [6Y/2N]
- Strong [3Y/3N]

$H(S)_{wind} = .94$

[9Y | 5N]

$|S_v|_{weak} = 8$

$H(S_v)_{weak} = -\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right)$

[6Y | 2N]

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

$= .311 + .05$

$= .811$

$|S_v|_{strong} = 6$

$H(S_v)_{strong} = 1$

[3Y | 3N]

Putting above values in I.G formulae

$I.G_{wind} = .94 - \left[\frac{8}{14} \times .811 + \frac{6}{14} \times 1\right]$

$= .94 - \left[.571 \times .811 + .428\right]$

$= .94 - \left[.463 + .428\right]$

$= .94 - .891$

$= .049$

$\Rightarrow$ $I.G_{wind} = 0.049$

Comparing information gain of each feature:

$I.G_{outlook} = 0.248$ | $I.G_{Temp} = 0.036$

$I.G_{Humidity} = 0.159$ | $I.G_{wind} = 0.049$

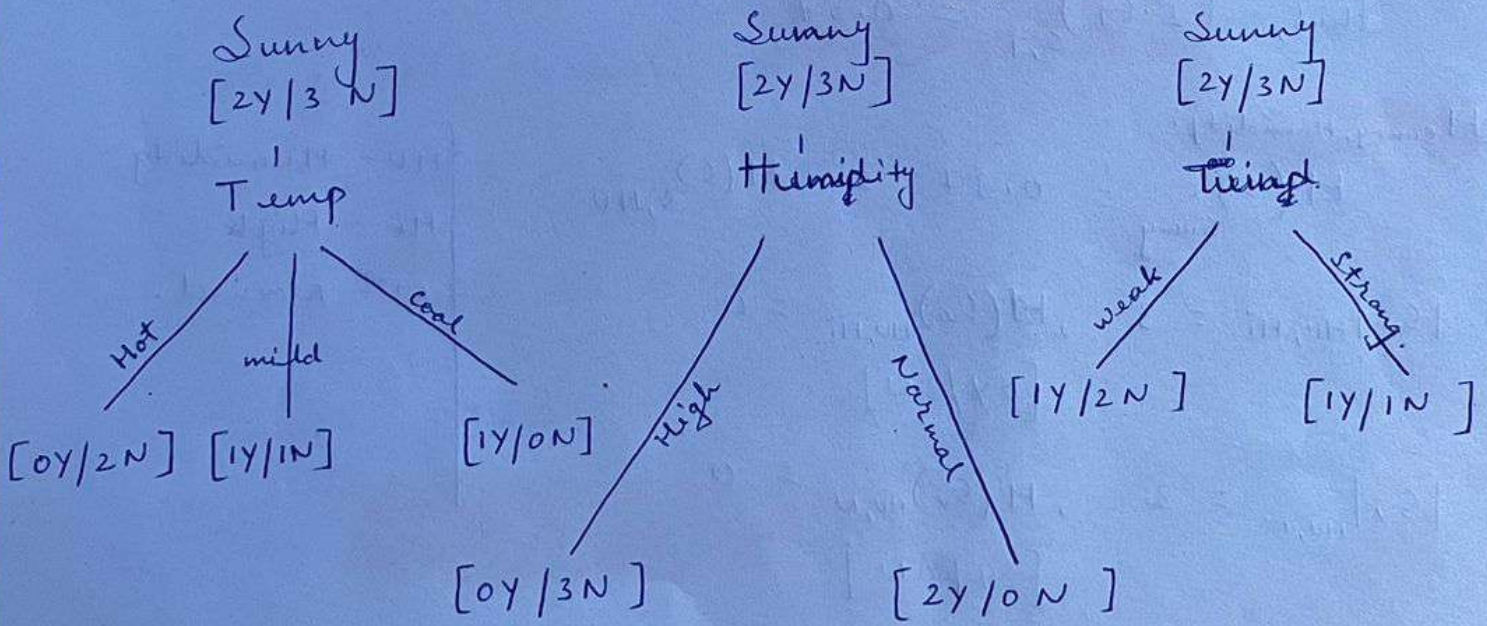
From above we have:

$I.G_{outlook} > I.G_{Humidity} > I.G_{wind} > I.G_{Temp}$.

$\Rightarrow$ Outlook will provide best info.

Step 2: Building 2nd decision Tree.

Considering Sunny from the Outlook and building Temp, Humidity and wind Tree out of it.



Calculating I.G one by one for each:

$H_{sunny, Temp}$:

$H(s)_{sunny} = .97 = H(s)_{(s,T)}$

[2Y/3N]

$|S_v|_{T,H} = 2$

$H(S_v)_{T,H} = 0$

[0Y/2N]

$|S_v|_{T,M} = 2$

$H(S_v)_{T,M} = 1$

[1Y/1N]

$|S_v|_{T,C} = 1$

$H(S_v)_{T,C} = 0$

[1Y/0N]

S = Sunny
T - Temp.
H - Hot
M - Mild
C - Cool.
HU - Humidity

Putting all values to find I.G

$= .97 - \left[\frac{2}{5} \times 0 + \frac{2}{5} \times 1 + \frac{1}{5} \times 0\right]$

$= .97 - .4$

$= 0.57$

Thus, $(I.G)_{S,T} = 0.57$

$H_{sunny, Humidity}$:

$H(s)_{sunny} = 0.97 = H(s)_{s,HU}$

$|S_v|_{HU,Hi} = 3$, $H(S_v)_{HU,Hi} = 0$

[0Y/3N]

$|S_v|_{HU,N} = 2$, $H(S_v)_{HU,N} = 0$

[2Y/0N]

HU - Humidity
Hi - High
N - Normal.

Putting all values to find I.G

$$= .97 - \left[\frac{3}{5} \times 0 + \frac{2}{5} \times 0\right]$$

$$= .97$$

Thus, $(I.G)_{S,HU} = 0.97$

$H_{sunny,wind}$:

$$H(S) = 0.97 = H_{S,W}$$

$$|S_v|_{W,We} = 3$$

$$H(S_v)_{W,We} = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)$$

$$[1Y/2N] = .528 + .386$$

$$= .914$$

$$|S_v|_{W,S} = 2$$

$$H(S_v)_{W,S} = 0$$

W - wind
We - weak
S - Strong

Putting all values to find I.G

$$= .97 - \left[\frac{3}{5} \times .914\right]$$

$$= .97 - .548$$

$$= .422$$

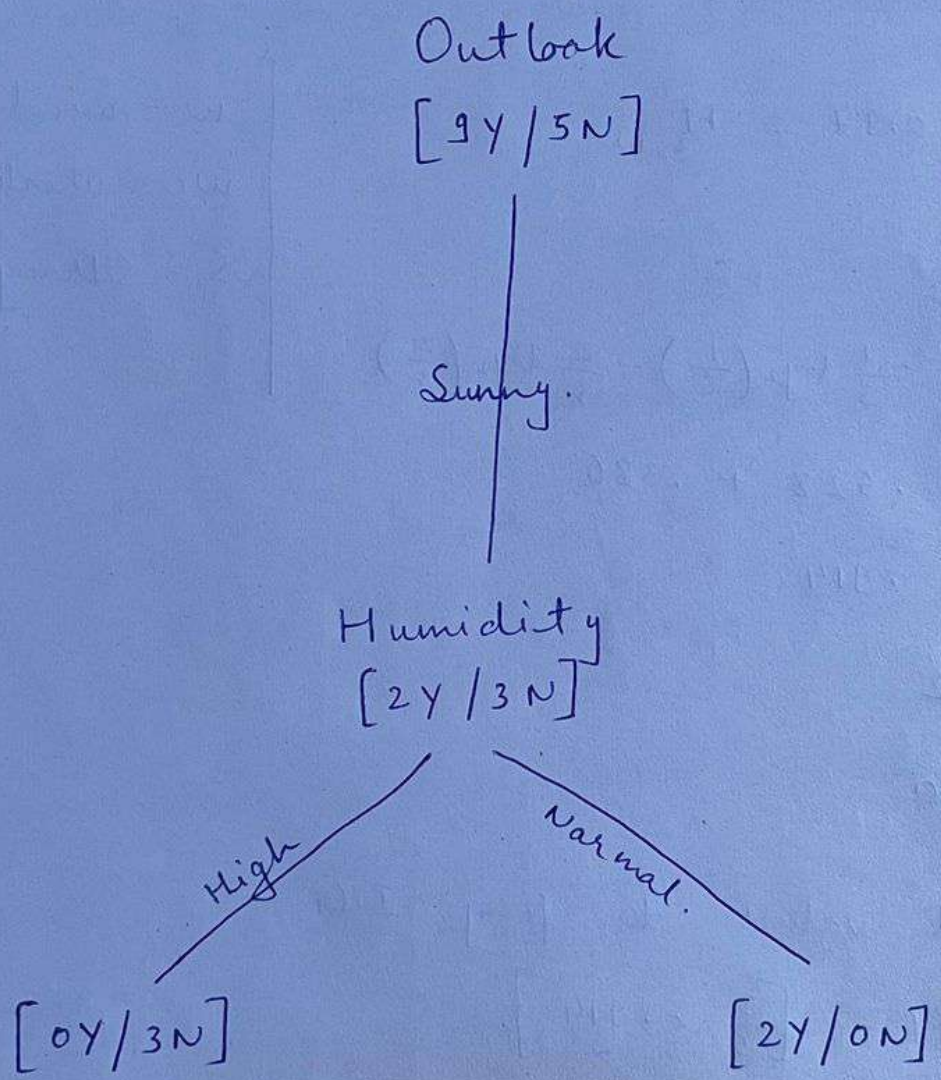
Thus, $(I.G)_{S,W} = .422$

From above $(I \cdot G)_{S,Hu} = .97$ is max of all

from $(I \cdot G)_{S,T}$, $(I \cdot G)_{S,W}$

Thus, the left side final Decision Tree be :

–



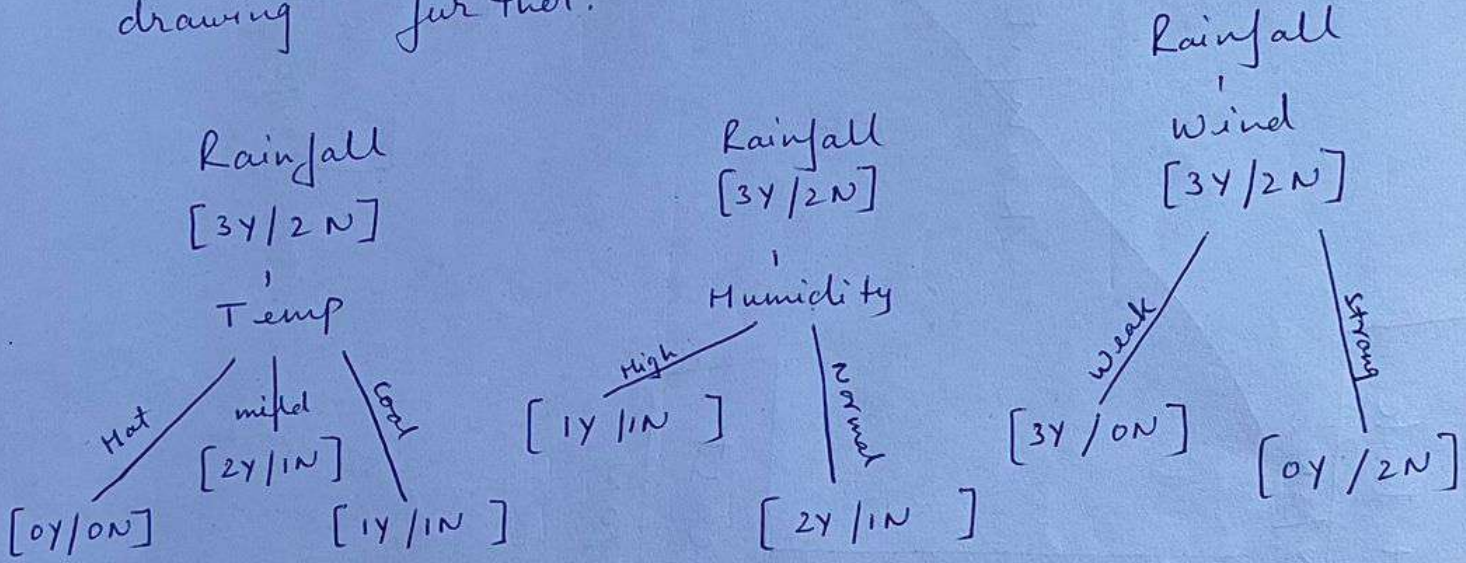
Thus, we have considered Humidity as further tree from Sunny which is having Pure node for both High and Normal.

Step 3: Buiding 3rd decision Tree.

Considering Overcast from the outlook and which is Pure node so we need not to buid it further.

Step 4: Buiding 4th decision Tree.

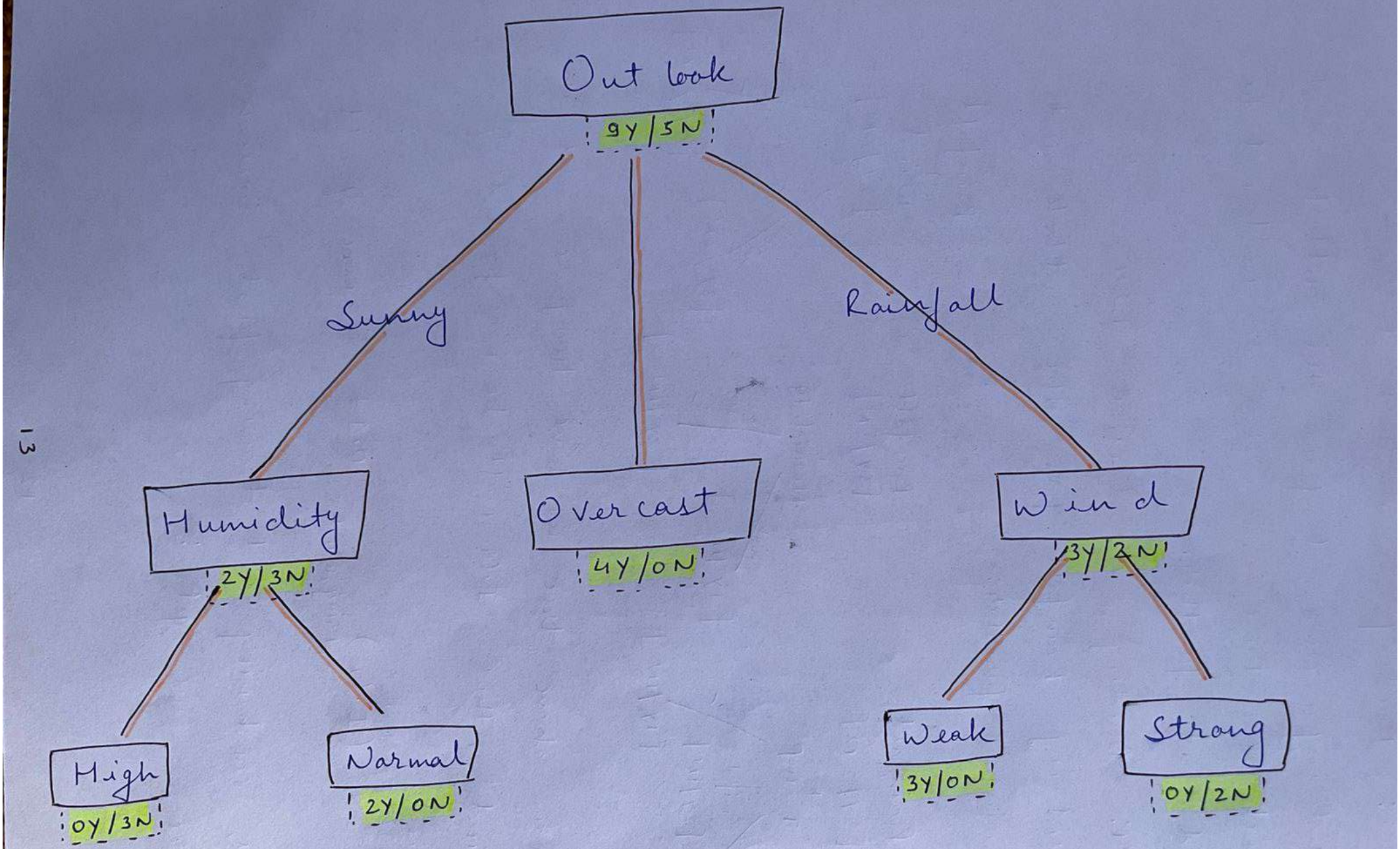
Considering Rainfall from the outlook and drawing further.



from above tree we have two pure nodes in wind, thus I.G will be more for case of wind.

$\Rightarrow (I.G)_{Rainy, wind} = 0.97$ is max of all

Thus, combaning and drawing final decision tree, we get:


Out look
9Y/5N
Sunny
Rainfall
Humidity
2Y/3N
Overcast
4Y/0N
Wind
3Y/2N
High
0Y/3N
Narmal
2Y/0N
Weak
3Y/0N
Strong
0Y/2N