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

Total Feature → 4 [outlook, temperature, humidity, wind]

*Finding the Root Feature 📑

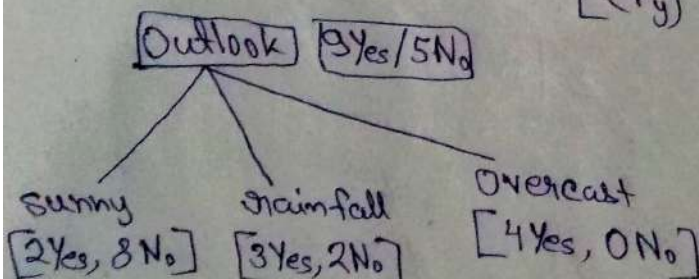I. need to findout the purity of the feature.

Purity Formulas are — ① entropy → $-\sum_{i=1}^{m} P_i \times \log_2(P_i)$

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

Here my classes are two, Yes and No, SO formulas will be :—

① entropy → $P_y \log_2(P_y) - P_n \log_2(P_n)$

② gini-impurity → $1 - [(P_y)^2 + (P_n)^2]$

Outlook $\boxed{9 \text{Yes}/5\text{No}}$

Sunny $\qquad$ rainfall $\qquad$ Overcast
[2Yes, 8No] $\quad$ [3Yes, 2No] $\quad$ [4Yes, 0No]

Sunny

$H(s) = -\frac{2}{5} \log_2(\frac{2}{5}) - \frac{3}{5} \log_2(\frac{3}{5})$

$= -0.4 [\log(2) - \log(5)] - 0.6 [\log(3) - \log(5)]$

$= -0.4 [1 - 2.322] - 0.6 [1.585 - 2.322]$

$= 0.5288 + 0.4422 = 0.97$

rainfall 3Yes/2No

entropy $\{H(s) = -\frac{3}{5}\log_2(\frac{3}{5}) - \frac{2}{5}\log_2\frac{2}{5}$

$= -0.6[\log_2(3) - \log_2(5)] - 0.4[\log_2(2) - \log_2(5)]$

$= 0.97$

overcast 4 Yes / 0 No

$H(s) = -\frac{4}{4}\log_2(\frac{4}{4}) - \frac{0}{4}\log_2(\frac{0}{4}) \Rightarrow 0$
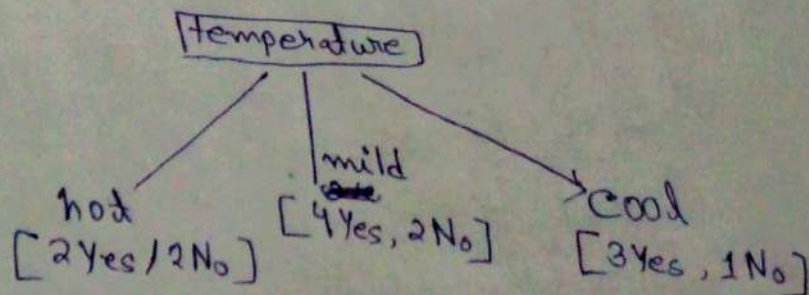
entropy of "Outlook" Feature is :- [9Yes, 5No]

$H(s) = -\frac{9}{14}\log(\frac{9}{14}) - \frac{5}{14}\log(\frac{5}{14})$

$= -\frac{9}{14}[\log(9) - \log(14)] - \frac{5}{14}[\log(5) - \log(14)]$

$= -0.642[3.17 - 3.807] - 0.357[2.322 - 3.807]$

≈ 0.93

Information Gain of outlook :-

Formula of Gain $= H(s) - \sum\limits_{value} \frac{|Sv|}{|S|} H(Sv)$

IG of outlook :-

= entropy of (outlook) $-\frac{5}{4}\times$ entropy of (sunny) $-\frac{4}{14}\times$ entropy of (overcast) $-\frac{5}{14}\times$ entropy of (rainfall)

$= 0.93 - \frac{5}{14}\times 0.97 - \frac{4}{14}\times 0 - \frac{5}{14}\times 0.97$

$= 0.246$

```
              Temperature
          /       |        \
       hot       mild       cool
  [2Yes/2No]  [4Yes, 2No]  [3Yes, 1No]
```

$H(s) = -\frac{1}{2}[\log_2(1) - \log_2(2)] - \frac{1}{2}[\log_2(1) - \log_2(2)]$

$= -0.5[0 - 1] - 0.5[0 - 1]$

$= 1$

mild

$= -\frac{4}{6}\log_2(\frac{4}{6}) - \frac{2}{6}\log_2(\frac{2}{6})$

$= -0.66[\log_2(4) - \log_2(6)] - 0.33[\log_2(2) - \log_2(6)]$

$= -0.66(2 - 2.585) - 0.33(1 - 2.585)$

$= 0.91$

cool [3Yes, 1 No]

$H(s) = -\frac{3}{4} \log_2(\frac{3}{4}) - \frac{1}{4} \log_2(\frac{1}{4})$
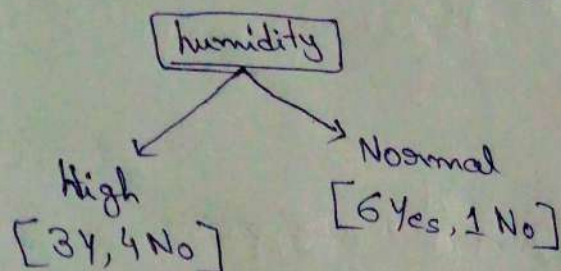
$= -0.75 [1.585 - 2] - 0.25 [0 - 2]$

$= 0.811$

entropy of "temperature" :- Here we have 9 Yes and 5 No
So the entropy of "temperature" is 0.93 same as outlook, humidity and wind.

so, IG of "temperature" = $0.93 - \frac{4}{14} \times 1 - \frac{6}{14} \times 0.91 - \frac{4}{14} \times 0.811$

$= 0.93 - 0.2857 - 0.39 - 0.2317$

$= 0.0226$

humidity

High   Normal
[3Y, 4No]   [6Yes, 1No]

high ⊞ $H(s) = -\frac{3}{7} \log_2(\frac{3}{7}) - \frac{4}{7} \log_2(\frac{4}{7})$

$= -0.42 [1.585 - 2.8074] - 0.57 [2 - 2.8074]$

$= 0.973$

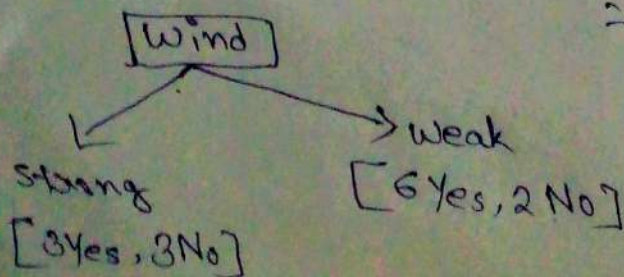normal $H(s) = -\frac{6}{7} [\log_2(6) - \log_2(7)] - \frac{1}{7} [\log_2(1) - \log_2(7)]$

$= -0.85 \times [2.585 - 2.8074] - 0.14 [0 - 2.8074]$

$= 0.58$

entropy of "humidity" :- Here we have 9 Yes and 5 No, so, the entropy of "humidity" is 0.93 same as outlook, ~~humidit~~ temperature, and wind.

so, IG of "humidity" = $0.93 - \frac{7}{14} \times 0.973 - \frac{7}{14} \times 0.587$

$= 0.15$

Wind

Strong   weak
[3Yes, 3No]   [6Yes, 2No]

<u>strong</u>

are there.
$$H(S) = 1 \text{ because } 50\% \text{ of both Yes and No}$$

<u>weak</u>

$$= -\frac{6}{8}[\log_2(6) - \log_2(8)] - \frac{2}{8}[\log_2(2) - \log_2(8)]$$
$$= 0.75[2.585 - 3] - 0.25[1 - 3]$$
$$= 0.8112$$

entropy of "wind" is 0.93 same as outlook, temperature and humidity.

$$IG \text{ of "wind"} = 0.93 - \frac{6}{14} \times 1 - \frac{8}{14} \times 0.8112$$
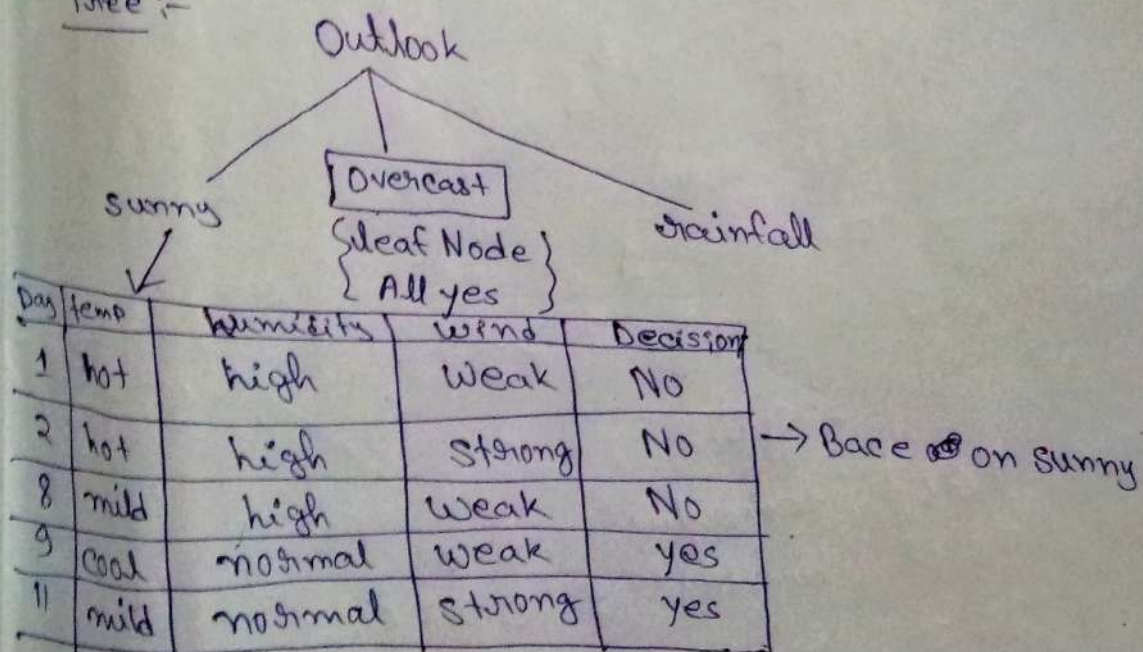$$= 0.038$$

IG of (Outlook) = 0.246 (high)

IG of (temperature) = 0.0226   ∴ Here outlook is highest

IG of (humidity) = 0.15   value.

IG of (wind) = 0.038

<u>Tree :-</u>

Day	temp	humidity	wind	Decision
1	hot	high	weak	No
2	hot	high	strong	No
8	mild	high	weak	No
9	cool	normal	weak	yes
11	mild	normal	strong	yes

→ Base on sunny

In this case, the entropy of sunny $(2Y, 3No) = 0.971$

①entropy of "hot" is 0; because there is only 2 No. Hot = 0

② There is one No and One Yes, so the entropy of mild is = 1

③ There is only one Yes, so the entropy of cool is = 0

Now, the entropy of "Temperature" is :-

$$= 0.97 - \frac{2}{5} \times 0 - \frac{2}{5} \times 1 - \frac{1}{5} \times 0$$
$$= 0.97 - 0.4$$
$$= 0.57$$

humidity

high   Normal
[0 yes, 3 No] [2 Yes, 0 No]

Here, ① the entropy of "high" is = 0

~~also~~

② the entropy of "normal" is = 0

.:. IG e.

Now, the Information Gain of "humidity" $= 0.97 - \frac{3}{5} \times 0 - \frac{2}{5} \times 0$
$$= 0.97$$

Wind

① entropy of "weak" $= -\frac{1}{3}\left[\log\left(\frac{1}{2}\right) - \log_2 (3)\right] - \frac{2}{3}$

$$\left[\log_2(2) - \log_2\left(\frac{3}{2}\right)\right]$$

$$= -0.333 \left[0 - 1.585\right] - 0.667\left[1 - 1.585\right]$$
$$= 0.9184$$

② entropy of "strong" = 1

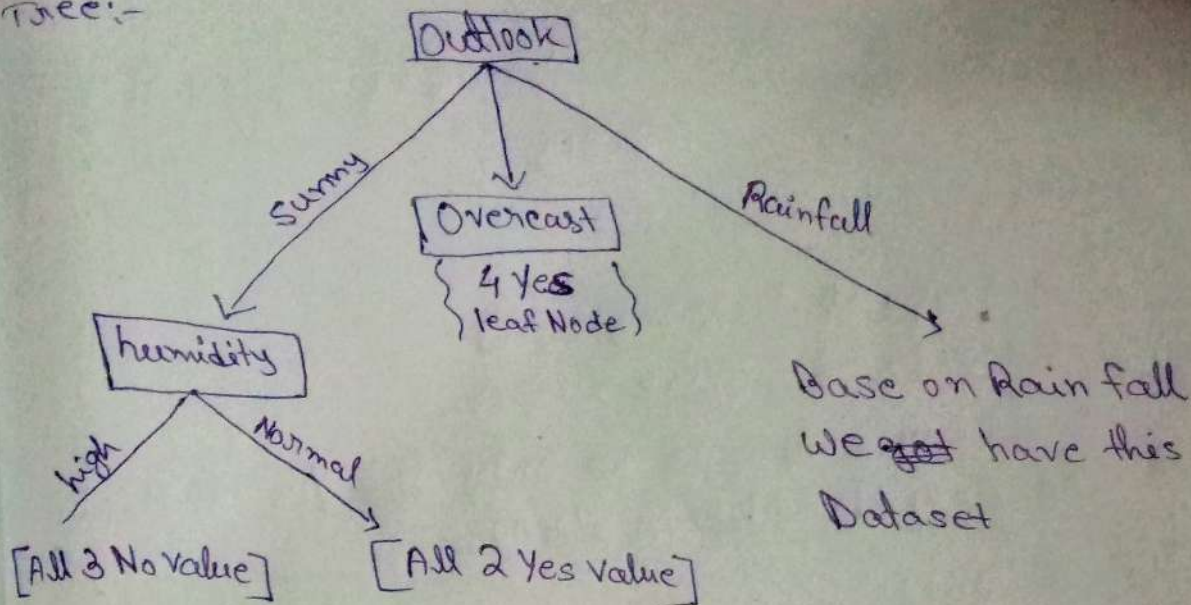IG of "wind" $= 0.97 - \frac{3}{5} \times 0.9184 - \frac{2}{5} \times 1$

$$= 0.019$$

Here, IG of "temperature" = 0.57

IG of "humidity" = 0.97 (high)

IG of "wind" = 0.019

Tree:

Tree:-

Outlook
- Sunny → humidity
- Overcast → { 4 Yes leaf Node }
- Rainfall →

humidity
- high → [All 3 No value]
- Normal → [All 2 Yes value]

Base on Rain fall we ~~get~~ have this Dataset

Day	Temp	humidity	Wind	Decision
4	mild	High	Weak	Yes
5	cool	normal	weak	Yes
6	cool	~~high~~ normal	strong	No
10	mild	normal	weak	Yes
14	mild	high	Strong	No

Entropy of Rainfall [3 Yes 2 No] is = 0.97

Ⓐ Temperature :-

① hot :- entropy of hot is 0, because there is no "hot" value present in this data

② mild :- [2 Yes, 1 No]

entropy of "mild" $= -\frac{2}{3}\left[\log_2(2) - \log_2 3\right] - \frac{1}{3}\left[\log_2(1) - \log_2(3)\right]$

$= 0.9184$

③ cool :- 50% of both Yes and No, so the value is 1.

IG of "temperature" respect of Rainfall :-

$= 0.97 - \frac{0}{5} \times 0 - \frac{3}{5} \times 0.9184 - \frac{2}{5} \times 1$

$= 0.192$

Ⓑ humidity :-

① entropy of "high" = 1

② entropy of "normal" $= -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right)$

$= 0.9184$

So, IG of "humidity" $= 0.97 - \frac{2}{5} \times 1 - \frac{3}{5} \times 0.9184$

$$= 0.019$$

© wind :-

① entropy of "strong" $= 0$

② entropy of "weak" $= 0$

∴ Because there is only Yes values in weak and only No values in strong.

IG. of "wind" $= 0.97 - \frac{2}{5} \times 0 - \frac{3}{5} \times 0 = 0.97$

IG of "Temperature" $= 0.0192$

IG of "humidity" $= 0.0192$

∴ IG of "wind" $= 0.97$ (high)

Decision Tree :-

Outlook

Sunny → Humidity [3 Yes, 0 No]

Overcast (leaf Node)

Rainfall → Wind

strong → 2 No Values

weak → 3 Yes values