Day	Outlook	Temperature	Humidity	Wind	Play Cricket
01	Sunny	Hot	High	Weak	No
02	Sunny	Hot	High	Strong	No
03	Overcast	Hot	High	Weak	yes
04	Rainy	Mild	High	Weak	Yes
05	Rainy	Cool	Normal	~~Weak~~ Strong	yes
06	Rainy	Cool	Normal	Strong	No
07	Overcast	Cool	Normal	~~Strong Weak~~	Yes
08	Sunny	Mild	High	Weak	No
09	Sunny	Cool	Normal	Weak	yes
010	Rainy	Mild	Normal	Weak	Yes
011	Sunny	Mild	Normal	Strong	Yes
012	Overcast	Mild	High	Strong	yes
013	Overcast	Hot	Normal	Weak	yes
014	Rainy	Mild	High	Strong	No

Make a decision tree that predict whether cricket will be played or not on that day.

1) Entropy of total dataset =

$$P = 9 \quad\quad N = 5$$
$$(yes) \quad\quad\quad (No)$$

Step 2) Entropy of Outlook :-

- Sunny $\quad\quad P = 2 \quad\quad\quad N = 3$
  $\quad\quad\quad\quad (yes) \quad\quad\quad\quad (No)$

$$Entropy (Sunny) = -\left[\frac{2}{2+3} \log_2\left(\frac{2}{2+3}\right) + \frac{3}{2+3} \log_2\left(\frac{3}{2+3}\right)\right]$$

$$= 0.96$$

Archana Patnaik

- Rainy $P_{(yes)} = 3$ $N_{(No)} = 2$

Entropy (Rainy) $= -\left[\dfrac{3}{2+3}\log_2\left(\dfrac{3}{2+3}\right) + \dfrac{2}{2+3}\log_2\left(\dfrac{2}{2+3}\right)\right]$

$= 0.96$

- Overcast $P_{(yes)} = 4$ $N_{(No)} = 0$

Entropy (Overcast) $= -\left[\dfrac{4}{4}\log_2\left(\dfrac{4}{4}\right) + \dfrac{0}{4}\log_2\left(\dfrac{0}{4}\right)\right] = 0$

- $I(\text{outlook}) = \dfrac{2+3}{14} \times 0.96 + \dfrac{3+2}{14} \times 0.96 + \dfrac{4+0}{14} \times 0$

$= 0.69$

- Information Gain $=$ Entropy of total dataset $- I(\text{outlook})$

$= 0.94 - 0.69 = 0.25$

(2.2)   Entropy of temperature

- Hot $P = 2 \ (yes)$ $N = 2 \ (NO)$

$E(Hot) = -\left[\dfrac{2}{2+2}\log_2\left(\dfrac{2}{2+2}\right) + \dfrac{2}{2+2}\log_2\left(\dfrac{2}{2+2}\right)\right] = 1$

- Mild $P = 4 \ (yes)$ $N = 2 \ (NO)$

$E(Mild) = -\left[\dfrac{4}{4+2}\log_2\left(\dfrac{4}{4+2}\right) + \dfrac{2}{4+2}\log_2\left(\dfrac{2}{4+2}\right)\right] = 0.9$

- Cool $P = 3 \ (Yes)$ $N = 1 \ (No)$

$E(cool) = -\left[\dfrac{3}{3+1}\log_2\left(\dfrac{3}{3+1}\right) + \dfrac{1}{3+1}\log_2\left(\dfrac{1}{3+1}\right)\right] = 0.81$

- $I(temp) = \dfrac{2+2}{14} \times 1 + \dfrac{4+2}{14} \times 0.91 + \dfrac{3+1}{14} \times 0.81 = 0.91$

- Information Gain $= 0.94 - 0.91 = 0.03$

2.3) Entropy of Humidity

- **High.** $P = 3 \cdot (Yes)$ $N = 4\,(No)$

$$E(High) = -\left[\frac{3}{3+4}\log_2\left(\frac{3}{3+4}\right) + \frac{4}{3+4}\log_2\left(\frac{4}{3+4}\right)\right] = 0.98$$

- **Normal** $P = 6\,(Yes)$ $N = 1\,(No)$

$$E(Normal) = -\left[\frac{6}{6+1}\log_2\left(\frac{6}{6+1}\right) + \frac{1}{6+1}\log_2\left(\frac{1}{6+1}\right)\right] = 0.59$$

- $I(Humidity) = \dfrac{3+4}{14} \times 0.59 + \dfrac{6+1}{14} \times 0.98 = 0.78$

- Information Gain $= 0.94 - 0.78 = 0.16$

2.4) Entropy of Windy

- **Weak** $P = 6\,(Yes)$ $N = 1\,(No)$

$$E(Weak) = -\left[\frac{6}{6+2}\log_2\left(\frac{6}{6+2}\right) + \frac{2}{6+2}\log_2\left(\frac{2}{6+2}\right)\right] = 0.82$$
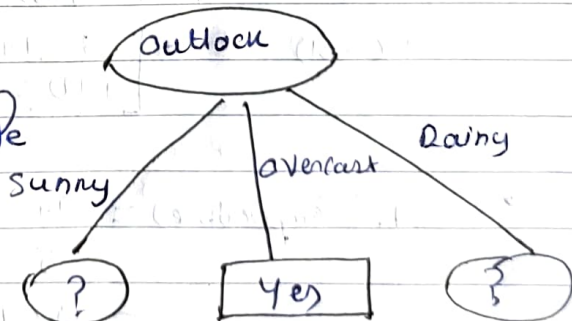
- **Strong** $P = 3\,(Yes)$ $N = 3\,(No)$

$$E(Strong) = -\left[\frac{3}{3+3}\log_2\left(\frac{3}{3+3}\right) + \frac{3}{3+3}\log_2\left(\frac{3}{3+3}\right)\right] = 1$$

- $I(Windy) = \dfrac{6+2}{14} \times 0.82 + \dfrac{3+3}{14} \times 1 = 0.89$

- Information Gain $= 0.94 - 0.89 = 0.05$

Attribute	Gain
✓Outlook	0.25 (Highest gain) Root Node
Temperature	0.03
Humidity	0.16
Windy	0.05

Outlook → Sunny → (?)
Outlook → overcast → [Yes]
Outlook → Rainy → (?)

Step 3) For Sunny :-

SlNo	Outlock	Temperature	Humidity	Windy	Play cricket
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Sunny	Mild	High	Weak	No
4	Sunny	Cool	Normal	Weak	Yes
5	Sunny	Mild	Normal	Strong	Yes

For total dataset $P=2, N=3$

$$E(Total) = -\left[\frac{2}{2+3} \log_2\left(\frac{2}{2+3}\right) + \frac{3}{2+5} \log_2\left(\frac{2}{2+3}\right)\right]$$

$$= 0.97$$

3.1) Entropy of Temperature

• Temperature Hot $\quad P = 0$ (yes) $\quad N = 2$ (No)

$$E(Hot) = -\left[\frac{0}{0+2} \log_2\left(\frac{0}{0+2}\right) + \frac{2}{0+2} \log_2\left(\frac{2}{0+2}\right)\right] = 0$$

• Mild $\quad P = 1$ (Yes) $\quad N = 1$ (No)

$$E(Mild) = -\left[\frac{1}{1+1} \log_2\left(\frac{1}{1+1}\right) + \frac{1}{1+1} \log_2\left(\frac{1}{1+1}\right)\right] = 1$$

• Cool $\quad P = 1$ (Yes) $\quad N = 0$ (No)

$$E(Cool) = -\left[\frac{1}{1+0} \log_2\left(\frac{1}{1+0}\right) + \frac{0}{1+0} \log_2\left(\frac{0}{1+0}\right)\right] = 0$$

$$I(Temperature) = \frac{0+2}{5}(0) + \frac{1+1}{5}(1) + \frac{1+0}{5}(0)$$

$$= 0.4$$

Information gain $= 0.97 - 0.4 = 0.57$

3-2) Entropy of Humidity

- High: $\quad P = 0 \ (yes) \qquad N = 3 \ (No)$

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

- Normal $\qquad P = 2 \ (yes) \qquad N = 0 \ (No)$

$$E(Normal) = - \left[\frac{2}{2+0} \log_2 \left(\frac{2}{2+0} \right) + \frac{0}{2+0} \log_2 \left(\frac{0}{2+0} \right) \right]$$

$$= 0$$

- $I(Humidity) = \dfrac{0+3}{5} \times 0 + \dfrac{2+0}{5} \times 0 = 0$

- Information Gain $= 0.97 - 0 = 0.97$

3-3) Entropy of Windy

- Weak $\qquad P = 1 \ (yes) \qquad N = 2 \ (No)$

$$E(Weak) = - \left[\frac{1}{1+2} \log_2 \left(\frac{1}{1+2} \right) + \frac{2}{1+2} \log_2 \left(\frac{2}{1+2} \right) \right]$$

$$= 0.91$$

- Strong $\qquad P = 1 \ (yes) \qquad N = 1 \ (No)$
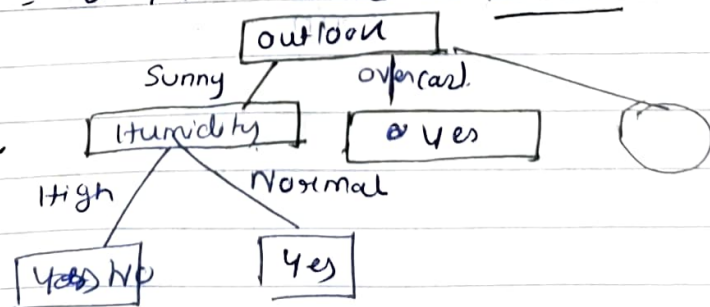
$$E(Strong) = - \left[\frac{1}{1+1} \log_2 \left(\frac{1}{1+1} \right) + \frac{1}{1+1} \log_2 \left(\frac{1}{1+1} \right) \right]$$

$$= 1$$

- $I(Windy) = \dfrac{1+2}{5}(0.91) + \dfrac{1+1}{5}(1) = 0.95$

- Information Gain $= 0.97 - 0.95 = 0.02$

Attribute	Gain
Temperature	0.57
Humidity	0.97 ✓
Windy	0.95

outlook

Sunny / overcast

Humidity | 0 yes

High / Normal

Yes No | Yes

Step 4) A1 For Rainy –:

Sl no	Outlook	Temperature	Humidity	Windy	Play cricket
1.	Rainy	Mild	High	Weak	Yes
2.	Rainy	Cool	Normal	Weak	Yes
3.	Rainy	Cool	Normal	Strong	No
4.	Rainy	Mild	Normal	Weak	Yes
5.	Rainy	Mild	High	Strong	No

For total dataset   P = 3 , N = 2

$$E(Total) = -\left[\frac{3}{3+2}\log_2\left(\frac{3}{3+2}\right) + \frac{2}{3+2}\log_2\left(\frac{2}{3+2}\right)\right]$$

$$= 0.97$$

4·1) Entropy of Humidity

- High   P = 1 (Yes)   N = 1 (No)

$$E(High) = -\left[\frac{1}{1+1}\log_2\left(\frac{1}{1+1}\right) + \frac{1}{1+1}\log_2\left(\frac{1}{1+1}\right)\right] = 1$$

- Normal   P = 2 (Yes)   N = 1 (No)

$$E(Normal) = -\left[\frac{2}{2+1}\log_2\left(\frac{2}{2+1}\right) + \frac{1}{2+1}\log_2\left(\frac{1}{2+1}\right)\right]$$

$$= 0.91$$

- $$I(Humidity) = \frac{1+1}{5} \times 1 + \frac{2+1}{5} \times 0.91 = 0.95$$

- Information $= 0.97 - 0.95 = 0.02$
  Gain

4.2) Entropy of Windy

- Weak $\qquad$ $P = 3\ (Yes),\ N = 0\ (No)$

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

- Strong $\qquad$ $P = 0\ (Yes),\ N = 2\ (No)$

$$E(Strong) = -\left[\frac{0}{0+2}\log\left(\frac{0}{0+2}\right) + \frac{2}{0+2}\log_2\left(\frac{0}{0+2}\right)\right] = 0$$

- $I(Windy) = \dfrac{3+0}{5}\ (0) + \dfrac{0+2}{5}\ (0) = 0$

- Information $\qquad = 0.97 - 0 = 0.97$
  gain

4.3) Entropy of temperature

- Cool $\qquad\qquad P = 1\ (Yes),\ N = 1\ (No)$

$$E(cool) = -\left[\frac{1}{1+1}\log_2\left(\frac{1}{1+1}\right) + \frac{1}{1+1}\log_2\left(\frac{1}{1+1}\right)\right]$$
$$= 1$$

- Mild $\qquad\qquad P = 2\ (Yes),\ N = 1\ (No)$

$$E(Mild) = -\left[\frac{2}{2+1}\log_2\left(\frac{2}{2+1}\right) + \frac{1}{2+1}\log_2\left(\frac{1}{2+1}\right)\right]$$
$$= 0.91$$

- $I(Temperature) = \dfrac{2+1}{5}\ (1) + \dfrac{2+1}{5}\ (0.91)$
  $$= 0.95$$

- Information $\qquad$ gain $= 0.97 - 0.95 = 0.02$

Attribute	Gain
Humidity	0.02
Windy	0.97 ✓
Temperature	0.95

```
                        ┌──────────┐
                        │ Outlook  │
                        └──────────┘
            Sunny        Overcast        Rainy
       ┌──────────┐    ┌────────┐     ┌────────┐
       │ Humidity │    │  Yes   │     │ Windy  │
       └──────────┘    └────────┘     └────────┘
       High    Normal                  Weak    Str
   ┌──────┐  ┌──────┐              ┌──────┐
   │  No  │  │ Yes  │              │ Yes  │
   └──────┘  └──────┘              └──────┘
```