Entropy = $-(p(a) \times \log 2(p(a)) - (p(b) \times \log 2(p(b)))$

• Training example: 9 yes / 5 no

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	NO
D2	Sunny	High	Strong	NO
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	NO
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	NO
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	NO

~~D15~~

P (playing tennis) $= \frac{9}{14} = 0.64$

P (not playing tennis) $= \frac{5}{14} = 0.35$

Entropy $= -[p(a) * \log_2(p(a))]$

Entropy $= -\left[\frac{9}{14} * \log_2\left(\frac{9}{14}\right)\right] - \left[\frac{5}{14} * \log_2\left(\frac{5}{14}\right)\right]$

Entropy = 0.940

Dividing the outlook
[Sunny, overcast, Rain]

Sunny:

Sunny	High	Weak	NO
Sunny	High	Strong	NO
Sunny	High	Weak	NO
Sunny	Normal	Weak	Yes
Sunny	Normal	Strong	Yes

p(Yes) = 2/5 = 0.4
p(NO) = 3/5 = 0.6

Entropy [Sunny] = $-0.4 \times \log_2(0.4) - 0.6 \times \log_2(0.6)$
Entropy [Sunny] = 0.97

Overcast:

Overcast	High	Weak	Yes
Overcast	Normal	Strong	Yes
Overcast	High	Strong	Yes
Overcast	Normal	Weak	Yes

p(yes) = 4/4 = 1.0
p(NO) = 0/4 = 0

Entropy [overcast] = $-1 \times \log_2(1) - 0 \times \log_2(0)$
Entropy [overcast] = 0

Rain:

Rain	High	Weak	Yes
Rain	Normal	Weak	Yes
Rain	Normal	Strong	NO
Rain	Normal	Weak	Yes
Rain	High	Strong	NO

$p(\text{yes}) = 3/5 = 0.6$
$p(\text{NO}) = 2/5 = 0.4$

$\text{Entropy}[\text{Rain}] = -0.6 \times \log_2[0.6) - 0.4 \times \log_2[0.4]$
$\text{Entropy}[\text{Rain}] = 0.97$

Entropy [Outlook] = [(No. of sunny days) / (total days) × (Entropy [Sunny])
+ [(No. of overcast days) / total days) × (Entropy [Overcast]
+ [(No. of rain days) / total days) × (Entropy [Rain])

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

$= 0.69$

Information Gain of Outlook $= 0.940 - 0.69$
$= 0.246$

Attribute : Humidity

Dividing the Humidity [High, Normal]

$p(yes) = 9/14$
$p(no) = 5/14$

$$\text{Entropy [Humidity]} = \frac{-9}{14} \times \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \times \log_2\left(\frac{5}{14}\right)$$

Entropy [Humidity] = 0.94

$S_{High} = [3^+, 4^-]$

$$\text{Entropy } [S_{High}] = -\frac{3}{7} \times \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \times \log_2\left(\frac{4}{7}\right)$$

Entropy [S_{High}] = 0.9852

$S_{Normal} = [6^+, 1^-]$

$$\text{Entropy } [S_{Normal}] = \frac{-6}{7} \times \log_2\left(\frac{6}{7}\right) - \frac{1}{7} \times \log\left(\frac{1}{7}\right)$$

Entropy [S_{Normal}] = 0.5916

Information Gain [Sunny - Humidity]

$$= 0.94 - \frac{7}{14} \times 0.9852 - \frac{7}{14} \times 0.5916$$

$= 0.1516$

Attribute : Wind

Dividing the wind [strong, weak]

$p(yes) = 9/14$
$p(no) = 5/14$

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

$$Entropy[Wind] = 0.94$$

$$S_{strong} = [3^+, 3^-]$$

$$Entropy[S_{strong}] = -\frac{3}{6} \times \log_2\left(\frac{1}{2}\right) - \frac{3}{6} \times \log_2\left(\frac{1}{2}\right)$$

$$Entropy[S_{strong}] = 1$$

$$S_{weak} = [6^+, 2^-]$$

$$Entropy[S_{weak}] = -\frac{6}{8} \times \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \times \log_2\left(\frac{2}{8}\right)$$

$$Entropy[S_{weak}] = 0.8113$$

Information Gain [Sunny - Wind]

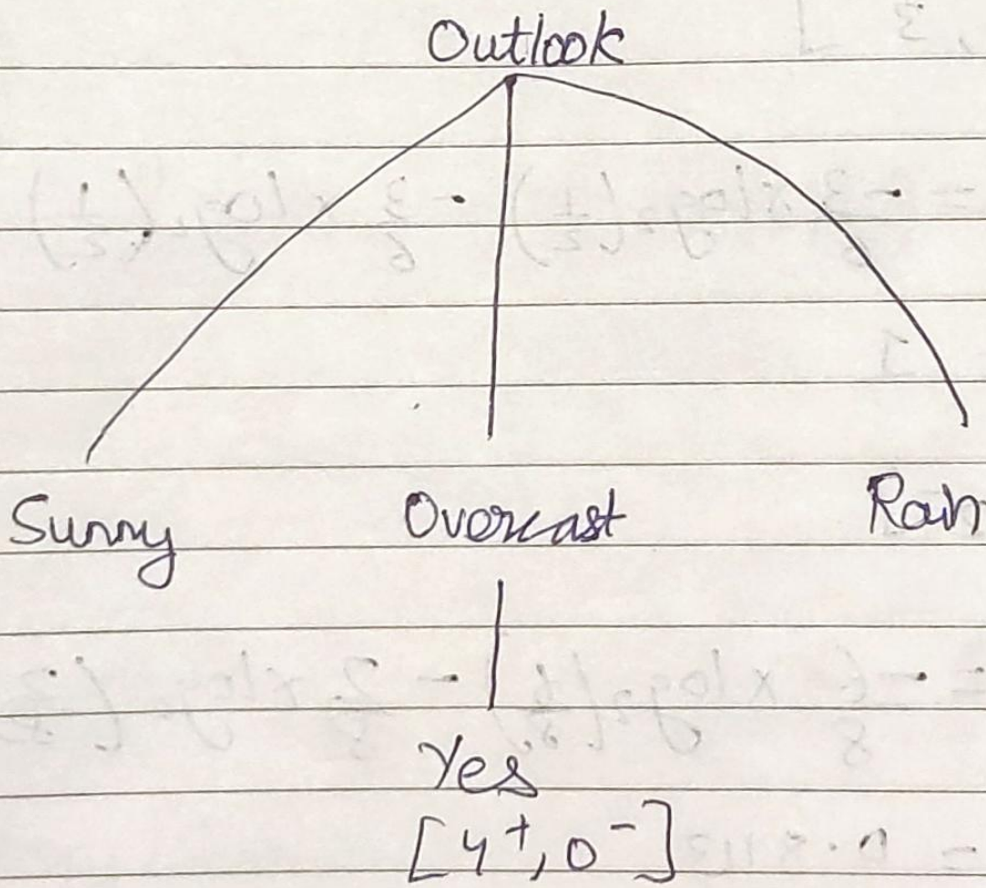
$$= 0.94 - \frac{6}{14} \times 1 - \frac{8}{14} \times 0.8113$$

$$= 0.0478$$

Information Gain of Outlook = 0.2464
Information Gain of Humidity = 0.1516
Information Gain of Wind = 0.0478.

The Information Gain of Outlook is highest among Humidity & Wind.

So, the Root node will be Outlook.



Attribute: Humidity in Sunny - Outlook
[High, ~~Low~~ Normal]

High	Weak	No
High	Strong	No
High	Weak	No
~~Low~~ Normal	Weak	Yes
~~Low~~ Normal	Strong	Yes

$p(yes) = 9/14$

$p(no) = 5/14$

Entropy [Sunny] $= 0.97$

$S_{High} = [0^+, 3^-]$

Entropy $[S_{High}] = -\frac{0}{3} \times \log_2\left(\frac{0}{3}\right) - 1 \times \log_2(1)$

Entropy $[S_{High}] = 0$

$S_{Normal} = [2^+, 0^-]$

Entropy $[S_{Normal}] = -\frac{2}{2} \times \log_2(1) - \frac{0}{2} \times \log_2\left(\frac{0}{2}\right)$

Entropy $[S_{Normal}] = 0$

Information Gain of [Sunny, Humidity] $= 0.97 - 0 - 0 = 0.97$

Attribute : Wind in Sunny outlook
[Weak, Strong]

Entrophy [sunny] = 0.97

S_{weak} = [1^+, 2^-]

$$\text{Entropy [Sweak]} = -\frac{1}{3} \times \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \times \log_2\left(\frac{2}{3}\right)$$

Entropy [Sweak] = 0.9183

S_{strong} = [1^+, 1^-]

Entropy [Sstrong] = 1

Information Gain of Sunny - Wind

$$= 0.97 - \frac{3}{5} \times 0.9183 - 1 \times \frac{2}{5}$$

= ~~[illegible]~~ 0.0192

Compairing the Information Gain of Sunny - Humidity & Sunny - Wind we found that Sunny - Humidity is greater than Sunny - Wind.

The Sunny - Humidity will be next node in the decision tree.

Attribute : Humidity in Rain - Outlook
{High, Normal}

High	Weak	Yes
Normal	Weak	Yes
Normal	Strong	No
Normal	Weak	Yes
High	Strong	No

Entropy [Rain-Outlook] = 0.97

$R_{High} = [1^+, 1^-]$

$$\text{Entropy}[R_{High}] = -\frac{1}{2} \times \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \times \log_2\left(\frac{1}{2}\right)$$

Entropy [R_{High}] = 1

$R_{Normal} = [2^+, 1^-]$

$$\text{Entropy}[R_{Normal}] = -\frac{2}{3} \times \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \times \log_2\left(\frac{1}{3}\right)$$

Entropy [R_{Normal}] = 0.9183

Information Gain of Rain - Humidity

$$= 0.97 - \frac{2}{5} \times 1 - \frac{3}{5} \times 0.9183$$

$$= 0.202$$

Attribute: Wind in Rain-Outlook

~~[High, No]~~ [Strong, Weak]

Entropy [Rain-Wind] = 0.97

$R_{strong} = [0^+, 2^-]$

$\text{Entropy}[R_{strong}] = \frac{0}{2} \times \log_2\left(\frac{0}{2}\right) - 1 \times \log_2(1)$

$\text{Entropy}[R_{strong}] = 0$

$R_{weak} = [3^+, 0^-]$

$\text{Entropy}[R_{weak}] = 1 \times \log_2(1) - 0$

$\text{Entropy}[R_{weak}] = 0$

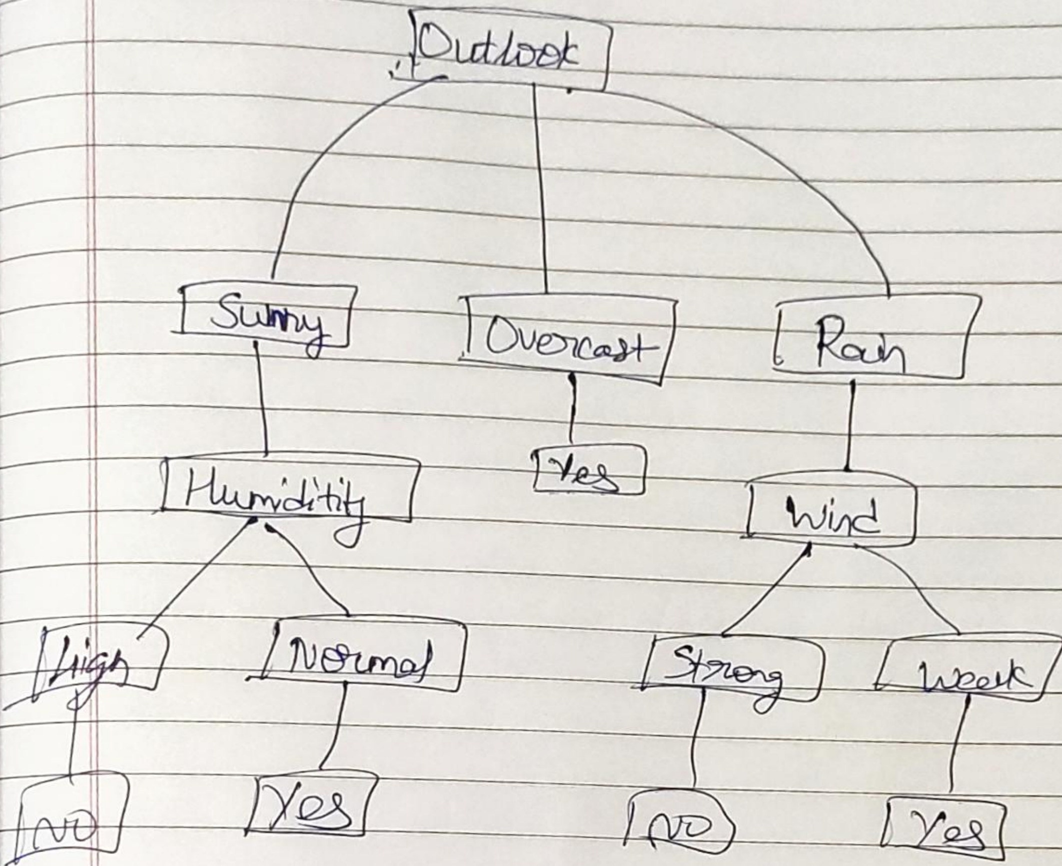
Information Gain of Rain-Wind:

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

$= 0.97$

Comparing the Information Gain of Rain-Humidity & Rain-Wind, we found that Rain-Wind is greater than Rain-Hummidity.

∴ The Rain-Wind will be the next node in the decision tree.



To Find:

D15 Rain High Weak ?

On D15, the Rain was there with high humidity & Weak wind, that we found that John will play tennis "Yes".