Outlook	Temperature	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
sunny	Mild	High	False	Yes
sunny	Cool	Normal	False	Yes
sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
sunny	Mild	High	True	No

Frequency table of classes

Play Golf (14)	
Yes	No
9	5
p(Y)=9/14=0.64	p(N)=5/14=0.36

Calculating Entropy for the classes (Play Golf)

E(5,9)=--p(Y)*log2(p(Y))-p(N)*log2(p(N))=

-0.64*log2(0.64)-0.36*log2(0.36)=-0.64*-0.6439-0.36*-1.474=0.412+0.531=0.94

Calculate Entropy for Other Attributes After Split

E(PlayGolf, Outloook)=SUM(p©*E©) where e belongs to outlook c={ Rainy , Overcast, sunny**}**

E(PlayGolf, Outloook)=p(Rainy)E(Rainy)+ **p(**Overcast)E(Overcast)+ **p(**sunny)E(sunny)

		PlayGolf(14)		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5

p(Rainy)=5/14 **p(**Overcast)=4/14 p(sunny)=5/14

$$E(Sunny) = E(3,2)$$

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

$$= -(0.60\ \log_2 0.60) - (0.40 \log_2 0.40)$$

$$= -(0.60 * 0.737) - (0.40 * 0.529)$$

$$= 0.971$$

$$E(Overcast) = E(4,0)$$

$$= -\left(\frac{4}{4}\log_2\frac{4}{4}\right) - \left(\frac{0}{4}\log_2\frac{0}{4}\right)$$

$$= -(0) - (0)$$

$$= 0$$

$$E(Rainy) = E(2,3)$$

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

$$= -(0.40\ \log_2 0.40) - (0.6 \log_2 0.60)$$

$$= 0.971$$

$$E(PlayGolf, Outlook) = \frac{5}{14}E(3,2) + \frac{4}{14}E(4,0) + \frac{5}{14}E(2,3)$$

$$= \frac{5}{14}0.971 + \frac{4}{14}0.0 + \frac{5}{14}0.971$$

$$= 0.357 * 0.971 + 0.0 + 0.357 * 0.971$$

$$= 0.693$$

E(PlayGolf, Temperature) Calculation

		PlayGolf(14)		
		Yes	No	
Temperature	Hot	2	2	4
	Cold	3	1	4
	Mild	4	2	6

E(PlayGolf, Temperature) = P(Hot) E(2,2) + P(Cold) E(3,1) + P(Mild) E(4,2)

E (PlayGolf, Temperature) = **4/14 * E(Hot) + 4/14 * E(Cold) + 6/14 * E(Mild)**

E (PlayGolf, Temperature) = **4/14 * E(2, 2) + 4/14 * E(3, 1) + 6/14 * E(4, 2)**

E (PlayGolf, Temperature) = 4/14 * -(2/4 log 2/4) - (2/4 log 2/4)

+ 4/14 * -(3/4 log 3/4) - (1/4 log 1/4)

+ 6/14 * -(4/6 log 4/6) - (2/6 log 2/6)

E (PlayGolf, Temperature) = 5/14 * 1.0

+ 4/14 * 1.811

+ 5/14 * 0.918

= 0.911

E(PlayGolf, Humidity) Calculation

		PlayGolf(14)		
		Yes	No	
Humidity	High	3	4	7
	Normal	6	1	7

E (PlayGolf, Humidity) = **7/14 * E(High) + 7/14 * E(Normal)**

E (PlayGolf, Humidity) = **7/14 * E(3, 2) + 7/14 * E(4, 0)**

E (PlayGolf, Humidity) = 7/14 * -(3/7 log 3/7) - (4/7 log 4/7)

+ 7/14 * -(6/7 log 6/7) - (1/7 log 1/7)

E (PlayGolf, Humidity) = 7/14 * 0.985

+ 7/14 * 0.592

= 0.788

E(PlayGolf, Windy) Calculation

		PlayGolf(14)		
		Yes	No	
Windy	TRUE	3	3	6
	FALSE	6	2	8

E (PlayGolf, Windy) = 6/14 * E(True) + 8/14 * E(False)

E (PlayGolf, Windy) = 6/14 * E(3, 3) + 8/14 * E(6, 2)

E (PlayGolf, Windy) = 6/14 * -(3/6 log 3/6) - (3/6 log 3/6)

\+ 8/14 * -(6/8 log 6/8) - (2/8 log 2/8)

E (PlayGolf, Windy) = 6/14 * 1.0

\+ 8/14 * 0.811

= 0.892

1. **E(PlayGolf, Outloook) = 0.693**
2. **E(PlayGolf, Temperature) = 0.911**
3. **E(PlayGolf, Humidity) = 0.788**
4. **E(PlayGolf,Windy) = 0.892**

Calculating Information Gain for Each Split

Gain(S,T) = Entropy(S) – Entropy(S,T)

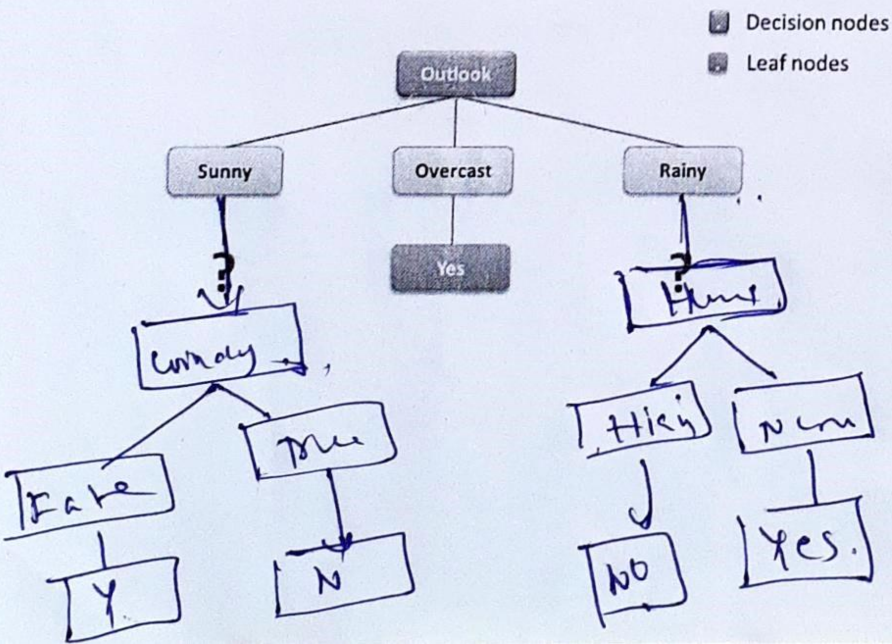
Gain(PlayGolf, Outlook) = Entropy(PlayGolf) – Entropy(PlayGolf, Outlook) = 0.94 – 0.693 = **0.247**

Gain(PlayGolf, Temperature) = Entropy(PlayGolf) – Entropy(PlayGolf, Temparature)= 0.94 – 0.911 = **0.029**

Gain(PlayGolf, Humidity) = Entropy(PlayGolf) – Entropy(PlayGolf, Humidity)= 0.94 – 0.788 = **0.152**

Gain(PlayGolf, Windy) = Entropy(PlayGolf) – Entropy(PlayGolf, Windy)= 0.94 – 0.892 = **0.048**

Now that we have all the information gain, we then split the tree based on the attribute with the highest information gain. From our calculation, the highest information gain comes from Outlook. Therefore the split will look like this:



Now that we have the first stage of the decison tree, we see that we have one leaf node. But we still need to split the tree further.

To do that, we need to also split the original table to create sub tables.
This sub tables are given in below.

Outlook	Temperature	Humidity	Windy	Play Golf
Sunny	Mild	Normal	FALSE	Yes
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Sunny	Mild	High	TRUE	No

Overcast	Hot	High	FALSE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Overcast	Cool	Normal	TRUE	Yes

Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes

Outlook	Temperature	Humidity	Windy	Play Golf
Sunny	Mild	Normal	FALSE	Yes
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Sunny	Mild	High	TRUE	No

Frequency table of classes

Play Golf (5)	
Yes	No
3	2
$P(Y) = \frac{3}{5} = 0.6$	$P(N) = \frac{2}{5} = 0.4$

Calculating Entropy for the classes (play Golf)

$$E(2,3) = -P(Y) * \log_2(P(Y)) - P(N) * \log_2(P(N))$$
$$= -0.6 * \log_2(0.6) - 0.4 * \log_2(0.4)$$
$$= -0.6 * (-0.737) - 0.4 * (-1.322)$$
$$= 0.970$$

Calculate Entropy for other Attributes After Split

$E(\text{PlayGolf}, \text{Temperature}) = \text{SUM}(P(c) * E(c))$ where c belongs to temperature $c = \{\text{Mild}, \text{cool}\}$

$$E(\text{PlayGolf}, \text{Temperature}) = P(\text{Mild})\,E(\text{Mild}) + P(\text{cool})\,E(\text{cool})$$

		Play Golf (5)		
		Yes	No	
Temperature	Mild	2	1	3
	Cool	1	1	2

$P(\text{Mild}) = \frac{3}{5}$ $\quad P(\text{Cool}) = \frac{2}{5}$

$E(\text{Mild}) = E(1,2) = -P(Y) * \log_2(P(Y)) - P(N) * \log_2(P(N))$

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

$= 0.9182$

$E(\text{cool}) = E(1,1) = -P(Y) * \log_2(P(Y)) - P(N) * \log_2(P(N))$

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

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

$E(\text{Play Golf, Temperature}) = P(\text{Mild})\, E(\text{Mild}) + P(\text{cool})\, E(\text{cool})$

$= \left(\frac{3}{5}\right)(0.9182) + \left(\frac{2}{5}\right)(1)$

$= 0.95092$

		Play Golf (5)		
		Yes	No	
Humidity	Normal	2	1	3
	High	1	1	2

$P(\text{Normal}) = \frac{3}{5}$ $\quad P(\text{High}) = \frac{2}{5}$

$E(\text{Normal}) = -P(Y) * \log_2(P(Y)) - P(N) * \log_2(P(N))$

$E(1,2) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.9182$

$$E(\text{High}) = E(1,1) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

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

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

$$E(\text{PlayGolf, Humidity}) = P(\text{Normal})\,E(\text{Normal}) + P(\text{High})\,E(\text{High})$$

$$= \left(\frac{3}{5}\right)(0.9182) + \left(\frac{2}{5}\right)(1)$$

$$= 0.95092$$

		Play Golf (5)		
		Yes	No	
Windy	FALSE	3	0	3
	TRUE	0	2	2

$P(\text{FALSE}) = \frac{3}{5}$ $\quad P(\text{TRUE}) = \frac{2}{5}$

$$E(\text{FALSE}) = E(0,3) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

$$= -\left(\frac{3}{3}\right)\log_2\left(\frac{3}{3}\right) - (0)\log_2(0)$$

$$= 0$$

$$E(\text{TRUE}) = E(2,0) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

$$= -(0)\log_2(0) - \left(\frac{2}{2}\right)\log_2\left(\frac{2}{2}\right)$$

$$= 0$$

$$E(\text{PlayGolf, Windy}) = P(\text{FALSE})\,E(\text{FALSE}) + P(\text{TRUE})\,E(\text{TRUE})$$

$$= \left(\frac{3}{5}\right)(0) + \left(\frac{2}{5}\right)(0)$$

$$= 0$$

calculating Information Gain for Each split.

Gain(S, T) = Entropy(S) − Entropy(S, T)

Gain(playGolf, ~~outlook~~ Temperature) = Entropy(Play Golf) − Entropy(PlayGolf, ~~outlook~~ Temperature)

= 0.97 − 0.95 = 0.02

Gain(play Golf, Humidity) = Entropy(PlayGolf) − Entropy(playGolf, Humidity)

= 0.97 − 0.95 = 0.02

Gain(play Golf, Windy) = Entropy(playGolf) − Entropy(PlayGolf, Windy)

= 0.97 − 0 = 0.97

Outlook	Temperature	Humidity	Windy	Play Golf
Rainy	Hot	High	FALSE	NO
Rainy	Hot	High	TRUE	No
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes

frequency table of classes.

Play Golf (5)	
Yes	No
2	3
$P(Y) = \frac{2}{5} = 0.4$	$P(N) = \frac{3}{5} = 0.6$

Calculating Entropy for the classes (Play Golf)

$E(3,2) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$

$= -0.4\log_2(0.4) - 0.6\log_2(0.6) = 0.97095$

		Play Golf (5)		
		Yes	No	
Temperature	Hot	0	2	2
	Mild	1	1	2
	cool	1	0	1

$P(\text{Hot}) = \frac{2}{5}$ $P(\text{Mild}) = \frac{2}{5}$ $P(\text{cool}) = \frac{1}{5}$

$$E(\text{Hot}) = E(2,0) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

$$= -\left(\frac{0}{2}\right)\log_2\left(\frac{0}{2}\right) - \left(\frac{2}{2}\right)\log_2\left(\frac{2}{2}\right)$$

$$= 0$$

$$E(\text{Mild}) = E(1,1) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

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

$$= 1$$

$$E(\text{cool}) = E(0,1) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

$$= -\left(\frac{1}{1}\right)\log_2\left(\frac{1}{1}\right) - \left(\frac{0}{1}\right)\log_2\left(\frac{0}{1}\right)$$

$$= 0$$

		Play Golf (5)		
		Yes	No	
Humidity	High	0	3	3
	Normal	2	0	2

$P(\text{High}) = \frac{3}{5}$ $P(\text{Normal}) = \frac{2}{5}$

$$E(\text{High}) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

$$= -0\log_2(0) - \left(\frac{3}{3}\right)\log_2\left(\frac{3}{3}\right)$$

$$= 0$$

$$E(\text{Normal}) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

$$= -\left(\frac{2}{2}\right)\log_2\left(\frac{2}{2}\right) - \left(\frac{0}{2}\right)\log_2\left(\frac{0}{2}\right)$$

$$= 0$$

$$E(\text{PlayGolf}, \text{Humidity}) = P(\text{High})\,E(\text{High}) + P(\text{Normal})\,E(\text{Normal})$$

$$= \left(\frac{3}{5}\right)(0) + \left(\frac{2}{5}\right)(0)$$

$$= 0$$

		Play Golf (5)		
		Yes	No	
Windy	FALSE	1	2	3
	TRUE	1	1	2

$P(\text{FALSE}) = \frac{3}{5}$ $\quad P(\text{TRUE}) = \frac{2}{5}$

$$E(\text{FALSE}) = E(2,1) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

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

$$= 0.9182$$

$$E(\text{TRUE}) = E(1,1) = -P(Y)\log_2(P(Y)) - P(N)\log_2(P(N))$$

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

$$= 1$$

$$E(\text{PlayGolf}, \text{Windy}) = P(\text{FALSE})\,E(\text{FALSE}) + P(\text{TRUE})\,E(\text{TRUE})$$

$$= \left(\frac{3}{5}\right)(0.9182) + \frac{2}{5}(1)$$

$$= 0.15092$$

calculating Information Gain for Each split.

$$Gain(S,T) = Entropy(S) - Entropy(S,T)$$

$$Gain(PlayGolf, Temperature) = Entropy(PlayGolf) - Entropy(PlayGolf, Temperature)$$

$$= 0.97095 - 0.4$$

$$= 0.57095$$

$$Gain(PlayGolf, Humidity) = Entropy(PlayGolf) - Entropy(PlayGolf, Humidity)$$

$$= 0.97095 - 0 = 0.97095$$

$$Gain(PlayGolf, Windy) = Entropy(PlayGolf) - Entropy(PlayGolf, Windy)$$

$$= 0.97095 - 0.15092$$

$$= 0.82003$$