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

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1
Gain = 0.029			

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1
Gain = 0.152			

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3
Gain = 0.048			

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

G(PlayGolf, Outlook) = **E**(PlayGolf) – **E**(PlayGolf, Outlook)

= 0.940 – 0.693 = 0.247

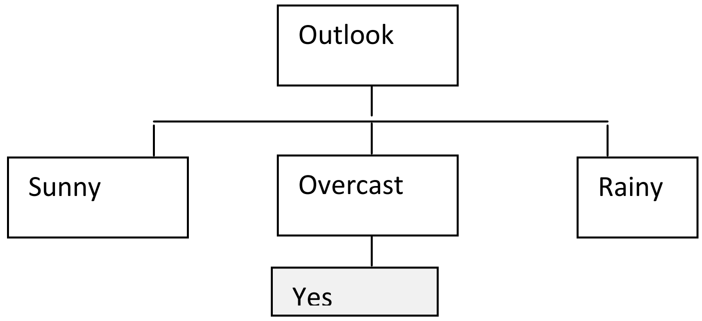


Table 1: Splitting on outlook=sunny

Outlook	temperature	Humidity	Windy	Play golf
Sunny	Mild	High	false	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Sunny	Mild	Normal	False	Yes
Sunny	Mild	High	True	no

Table 2: Splitting on outlook=overcast

Outlook	temperature	Humidity	Windy	Play golf
Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes

Table 3: Splitting on outlook=Rainy

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

Since the entropy of sunny branch is more than 0 therefore it needs more splitting. so, consider the subset of tuples given in table 1, the first column of outlook is discarded and colored yellow and is given for the sake of convenience. In order to determine the splitting attribute, we consider the subset of tuples given in table 1 and determine entropy of each attribute against the target attribute (PlayGolf) based on frequency table.

Note: Entropy and information gain will be determined for each attribute each time, when a new subset (here shown in table) is obtained after splitting.

Playgolf	
Yes	No
3	2

Entropy(Playgolf)=E(3,2)=E(3/5,2/5)= -3/5log3/5 - 2/5 log 2/5 = 0.970951

Temperature		PlayGolf		
		Yes	NO	
	cool	1 (1/2)	1 (1/2)	2 (2/5)
	mild	2 (2/3)	1 (1/3)	3 (3/5)

Entropy (PlayGolf, Temperature) = P(cool)E(1,1) + P(mild) E(2,1)= 0.950978

Gain(T, X)= Entropy(T) - Entropy (T,X)

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

=0.970951-0.950978

Gain (PlayGolf, Temperature) = 0.019973

Humidity		PlayGolf		
		Yes	NO	
	High	1 (1/2)	1 (1/2)	2 (2/5)
	normal	2 (2/3)	1 (1/3)	3 (3/5)

Entropy (PlayGolf, humidity) = P(high)E(1,1) + P(normal) E(2,1)= 0.950978

Gain(T, X)= Entropy(T) - Entropy (T,X)

Gain (PlayGolf, humidity) = Entropy (PlayGolf) - Entropy (PlayGolf, humidity)

=0.970951-0.950978

Gain (PlayGolf, humidity) = 0.019973

windy		PlayGolf		
		Yes	NO	
	true	0 (0/2)	2 (2/2)	2 (2/5)
	false	3 (3/3)	0 (0/3)	3 (3/5)

Entropy (PlayGolf, windy) = P(true)E(0,2) + P(false) E(3,0)= 0

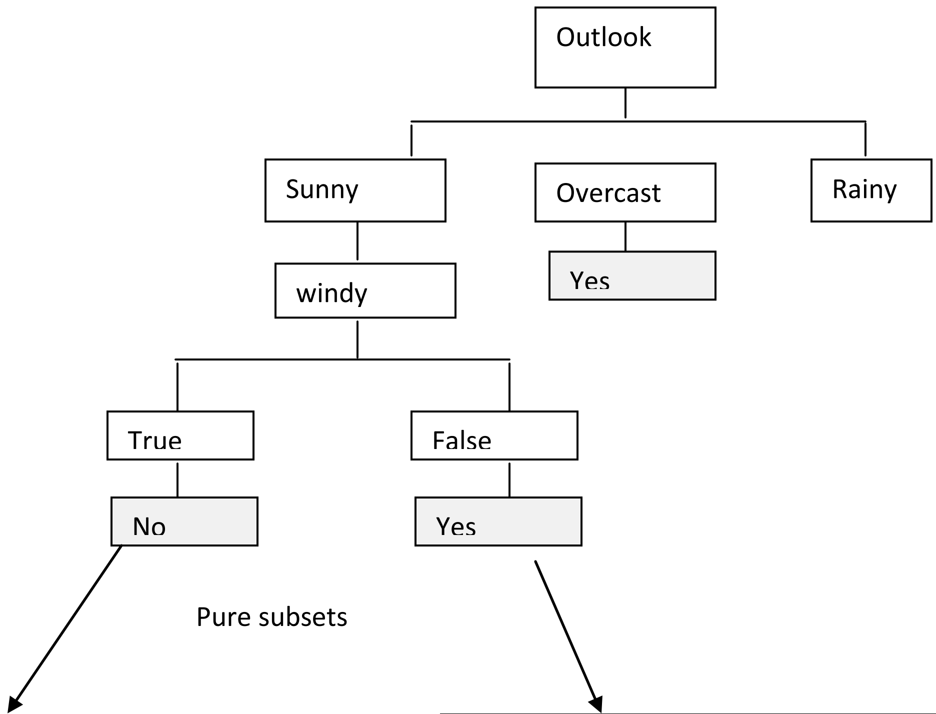
Gain(T, X)= Entropy(T) - Entropy (T,X)

Gain (PlayGolf, windy) = Entropy (PlayGolf) - Entropy (PlayGolf, windy)

=0.970951-0

Gain (PlayGolf,windy) = 0.970951

Since the information gain of windy attribute is more than information gain of other attributes, therefore, we further split the sunny branch by windy attribute.



Outlook	temperature	Humidity	Windy	Play golf
Sunny	Cool	Normal	True	No
Sunny	Mild	High	True	No

Outlook	temperature	Humidity	Windy	Play golf
Sunny	Mild	High	false	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes

Yellow colored for outlook and windy columns indicates discarded attributes, because these attributes have been considered for splitting.

We can see the non pure (non homogenous)tuples in table 3 that correspond to the branch (outlook=rainy), so this branch needs further splitting. we have only two remaining attributes (temperature, humidity) to test in order to decide the splitting attribute. Let us consider the following set of tuples where outlook=rainy

Outlook	temperature	Humidity	Windy	Play golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Rainy	Mild	High	False	No
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes

Next, determine entropy of each attribute against the target attribute (PlayGolf) based on frequency table.

Playgolf	
Yes	No
2	3

Entropy(Playgolf)=E(2,3)=E(2/5,3/5)= -2/5log2/5 - 3/5 log 3/5 = 0.970951

Humidity		PlayGolf		
		Yes	NO	
	High	0 (0/3)	3 (3/3)	3 (3/5)
	normal	2 (2/2)	0 (0/2)	2 (2/5)

Entropy (PlayGolf, humidity) = P(high)E(0,3) + P(normal) E(2,0)= 0

Gain(T, X)= Entropy(T) - Entropy (T,X)

Gain (PlayGolf, humidity) = Entropy (PlayGolf) - Entropy (PlayGolf, humidity)

=0.970951-0

Gain (PlayGolf, humidity) = 0.970951

Temperature		PlayGolf		
		Yes	NO	
	hot	0 (0/2)	2 (2/2)	2 (2/5)
	mild	2 (2/3)	1 (1/3)	3 (3/5)

Entropy (PlayGolf, Temperature) = P(hot)E(0,2) + P(mild) E(2,1)

= 0 + 3/5*(-2/3*log(2/3,2)- 1/3*log(1/3,2) = 0.550978

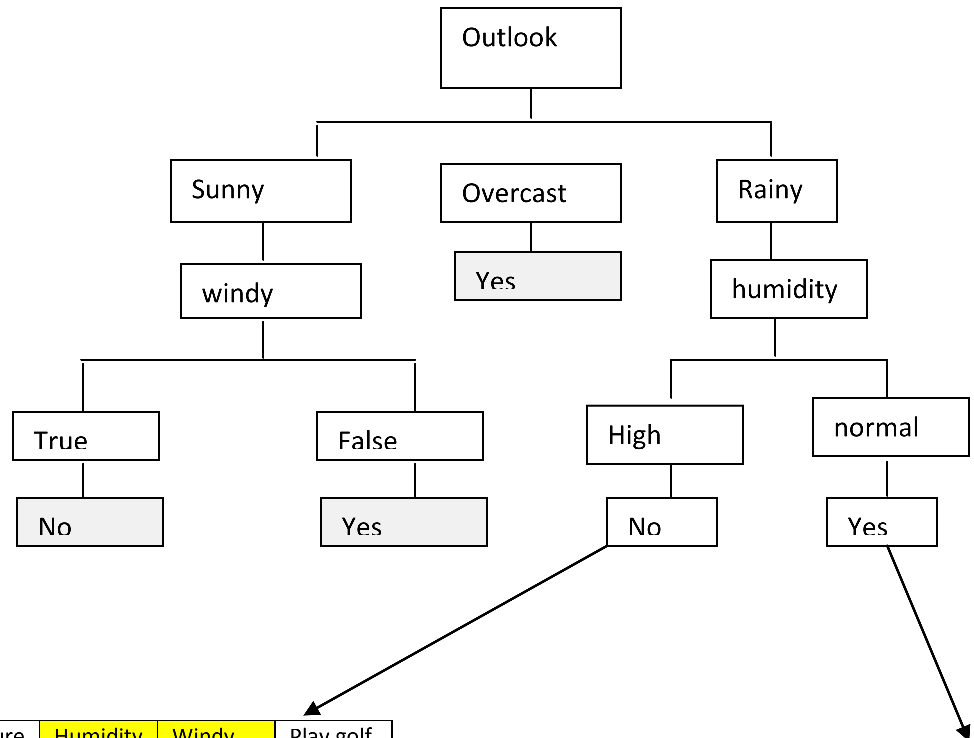
Gain(T, X)= Entropy(T) - Entropy (T,X)

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

=0.970951- 0.550978

Gain (PlayGolf, Temperature) =0.419973

Since the information gain of humidity attribute is more than information gain of temperature therefore, we further split the (outlook=rainy) branch by humidity attribute. so the resultant decision tree will become as follows:



Outlook	temperature	Humidity	Windy	Play golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Rainy	Mild	High	False	No

Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes

After splitting on humidity attribute, the two subsets of tuples obtained are pure (homogenous). the ID3 algorithm will stop since all data has been classified