A Step by Step ID3 Decision Tree Example

Step 1: Compute the Entropy for data set-Entropy(S)

Step 2: For Every Attribute/Feature:

1. Calculate entropy for all other values Entropy (A)
2. Take Average Information Entropy for the current Attribute
3. Calculate Gain for the current attribute.

Step 3. Pick the highest gain attribute.

Step 4. Repeat until we get the tree we desired.

Make a Decision tree that predicts whether tennis will be played on the day?

Data set

For instance, the following table informs about decision making factors to play tennis at outside for previous 14 days.

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Step 1: Compute the Entropy for data set-Entropy(S)

Positive - P

Negative-N

P=9

N=5

Total= 14

S= [9+, 5-]

$$Entropy(S) = -P_+ \, log_2 \, P_+ - P\_log_2 \, P\_$$

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

$$Entropy(S) = 0.94$$

Step 2: For Every Attribute/Feature:

1. Calculate entropy for all other values Entropy(A)

2. Take Average Information Entropy for the current Attribute

3. Calculate Gain for the current attribute.

Now, we have four features to make decision and they are:

- **Outlook**
- **Temperature**
- **Windy**
- **Humidity**

Attribute: Outlook (Sunny, Overcast, Rain)

S=Outlook

$$\boldsymbol{Entropy(S) = -P_+ \, log_2 \, P_+ - P\_log_2 \, P\_}$$

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

$$Entropy(S) = 0.94$$

S_{sunny}=[2+,3-]

$$Entropy(S_{\text{sunny}}) = -P_+ log_2 P_+ - P_- log_2 P_-$$

$$Entropy(S_{\text{sunny}}) = -\left(\frac{2}{5}\right) log_2 \left(\frac{2}{5}\right) - \left(\frac{3}{5}\right) log_2 \left(\frac{3}{5}\right)$$

$$Entropy(S_{\text{sunny}}) = 0.971$$

$S_{overcast}$=[4+,0-]

$$Entropy(S_{\text{overcast}}) = -P_+ log_2 P_+ - P_- log_2 P_-$$

$$Entropy(S_{\text{overcast}}) = -\left(\frac{4}{4}\right) log_2 \left(\frac{4}{4}\right) - \left(\frac{0}{4}\right) log_2 \left(\frac{0}{4}\right)$$

$$Entropy(S_{\text{overcast}}) = 0$$

S_{rain}=[3+,2-]

$$Entropy(S_{\text{rain}}) = -P_+ log_2 P_+ - P_- log_2 P_-$$

$$Entropy(S_{\text{rain}}) = -\left(\frac{3}{5}\right) log_2 \left(\frac{3}{5}\right) - \left(\frac{2}{5}\right) log_2 \left(\frac{2}{5}\right)$$

$$Entropy(S_{\text{rain}}) = 0.971$$

$$Gain(S, Outlook) = Entropy(S) - \sum_{V \in (\text{Sunny,Overcast,Rain})} \frac{|S_V|}{|S|} Entropy(V)$$

$$Gain(S, Outlook) = Entropy(S) - \left[\frac{9}{14} Entropy(S_{\text{sunny}}) + \frac{4}{14} Entropy(S_{\text{overcast}}) + \frac{5}{14} Entropy(S_{\text{rain}})\right]$$

$$Gain(S, Outlook) = 0.94 - \left[\frac{9}{14} * 0.971 + \frac{4}{14} * 0 + \frac{5}{14} * 0.971\right]$$

$$Gain(S, Outlook) = 0.2464$$

Attribute: Temp (Hot,Mild,Cool)

S=Temp

$$Entropy(S) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

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

$$Entropy(S) = 0.94$$

S_{Hot}=[2+,2-]

$$Entropy(S_{\text{Hot}}) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

$$Entropy(S_{\text{Hot}}) = -\left(\frac{2}{4}\right) log_2 \left(\frac{2}{4}\right) - \left(\frac{2}{4}\right) log_2 \left(\frac{2}{4}\right)$$

$$Entropy(S_{\text{Hot}}) = 1.0$$

S_{Mild}=[4+,2-]

$$Entropy(S_{\text{Mild}}) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

$$Entropy(S_{\text{Mild}}) = -\left(\frac{4}{6}\right) log_2 \left(\frac{4}{6}\right) - \left(\frac{2}{6}\right) log_2 \left(\frac{2}{6}\right)$$

$$Entropy(S_{\text{Mild}}) = 0.9183$$

S_{Cool}=[3+,1-]

$$Entropy(S_{\text{cool}}) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

$$Entropy(S_{\text{cool}}) = -\left(\frac{2}{5}\right) log_2 \left(\frac{2}{5}\right) - \left(\frac{3}{5}\right) log_2 \left(\frac{3}{5}\right)$$

$$Entropy(S_{\text{cool}}) = 0.8113$$

$$Gain(S, Temp) = Entropy(S) - \sum_{V \in (\text{Hot,Mild,Cool})} \frac{|S_V|}{|S|} Entropy(V)$$

$$Gain(S, Temp) = Entropy(S) - \left[\frac{4}{14}Entropy(S_{Hot}) + \frac{6}{14}Entropy(S_{Mild}) + \frac{4}{14}Entropy(S_{Cool})\right]$$

$$Gain(S, Temp) = 0.94 - \left[\frac{4}{14} * 1.0 + \frac{6}{14} * 0.9183 + \frac{4}{14} * 0.8113\right]$$

$$Gain(S, Temp) = 0.0289$$

Attribute: Humidity (High, Normal)

S= Humidity

$$Entropy(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

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

$$Entropy(S) = 0.94$$

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

$$Entropy(S_{High}) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

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

$$Entropy(S_{High}) = 0.9852$$

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

$$Entropy(S_{Normal}) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

$$Entropy(S_{Normal}) = -\left(\frac{6}{7}\right) \log_2 \left(\frac{6}{7}\right) - \left(\frac{1}{7}\right) \log_2 \left(\frac{1}{7}\right)$$

$$Entropy(S_{Normal}) = 0.5916$$

$$Gain(S, Humidity) = Entropy(S) - \left[\frac{7}{14}Entropy(S_{High}) + \frac{7}{14}Entropy(S_{Normal})\right]$$

$$Gain(S, Humidity) = 0.94 - \left[\frac{7}{14} * 0.9852 + \frac{7}{14} * 0.5916\right]$$

$$Gain(S, Humidity) = 0.1516$$

Attribute: Wind (Strong, Weak)

S= Wind

$$Entropy(S) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

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

$$Entropy(S) = 0.94$$

S_{Strong} = [3+,3-]

$$Entropy(S_{Strong}) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

$$Entropy(S_{Strong}) = -\left(\frac{3}{6}\right) log_2 \left(\frac{3}{6}\right) - \left(\frac{3}{6}\right) log_2 \left(\frac{3}{6}\right)$$

$$Entropy(S_{Strong}) = 1.0$$

S_{Weak} =[6+,2-]

$$Entropy(S_{Strong}) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

$$Entropy(S_{Strong}) = -\left(\frac{6}{8}\right) log_2 \left(\frac{6}{8}\right) - \left(\frac{2}{8}\right) log_2 \left(\frac{2}{8}\right)$$

$$Entropy(S_{Strong}) = 0.8113$$

$$Gain(S, Wind) = Entropy(S) - \left[\frac{6}{14} Entropy(S_{Strong}) + \frac{6}{14} Entropy(S_{Weak})\right]$$

$$Gain(S, Wind) = 0.94 - \left[\frac{6}{14} * 1.0 + \frac{6}{14} * 0.8113\right]$$

$$Gain(S, Wind) = 0.0478$$

Step 3 : Pick the highest gain attribute.

$$Gain(S, Outlook) = 0.2464$$

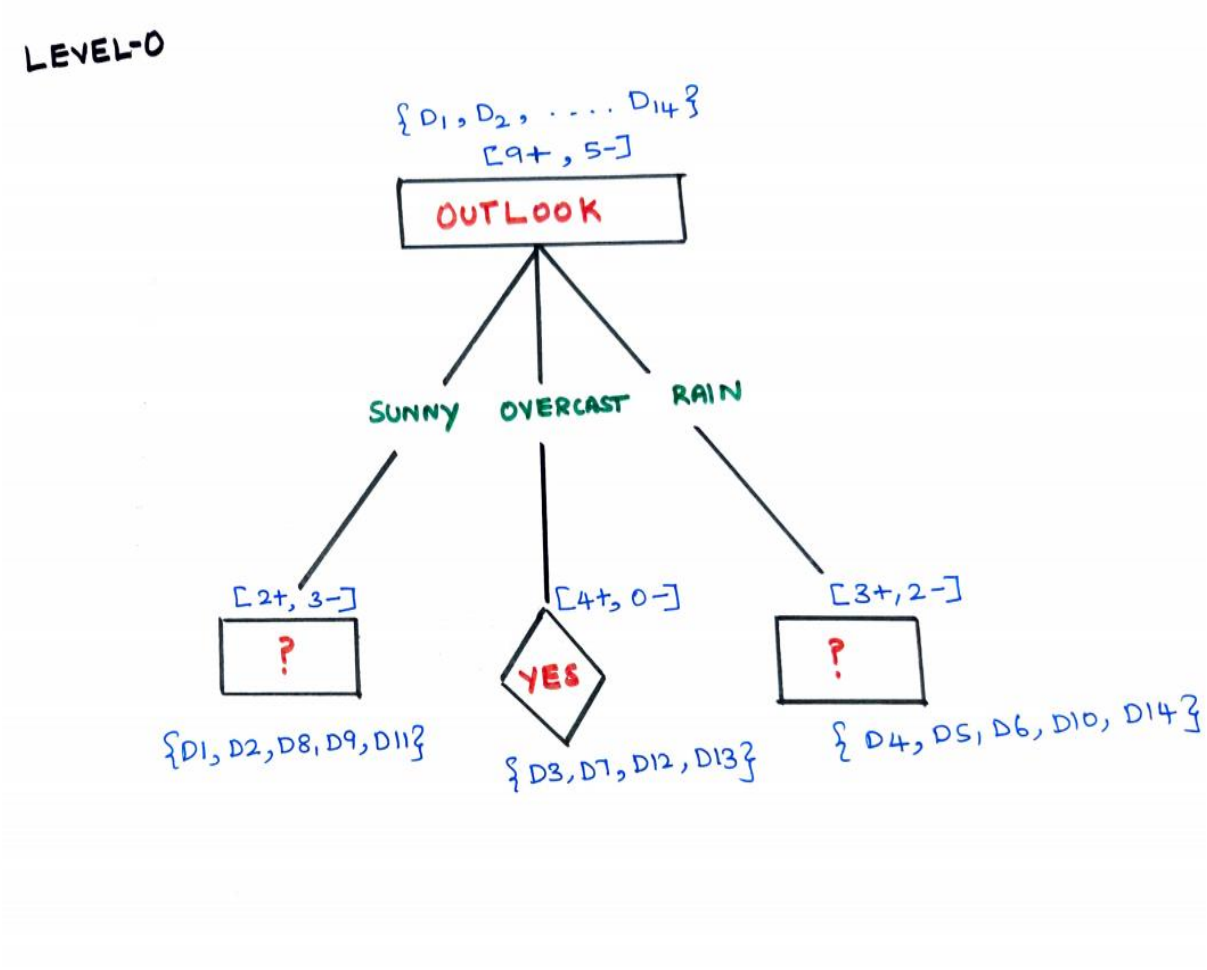
$$Gain(S, Temp) = 0.0289$$

$$Gain(S, Humidity) = 0.1516$$

$$Gain(S, Wind) = 0.0478$$

Level- 0 Root Node:

$$Gain(S, Outlook) = 0.2464$$



Step 4. Repeat until we get the tree we desired.

Now, we have Three features to make decision and they are:

- **Temperature**

- **Windy**
- **Humidity**

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

S=S_{Sunny}=[2+,3-]

$$Entropy(S) = -P_+ \, log_2 \, P_+ - P\_log_2 \, P\_$$

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

$$Entropy(S) = 0.97$$

S_{Hot}=[0+,2-]

$$Entropy(S_{Hot}) = 0.0$$

S_{Mild}=[1+,1-]

$$Entropy(S_{Mild}) = 1.0$$

S_{cool}=[1+,0-]

$$Entropy(S_{cool}) = 0.0$$

$$Gain(S_{Sunny}, Temp) = 0.570$$

S=S_{Sunny}=[2+,3-]

$$Entropy(S) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

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

$$Entropy(S) = 0.97$$

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

$$Entropy(S_{High}) = 0.0$$

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

$$Entropy(S_{Normal}) = 0.0$$

$$Gain(S_{Sunny}, Humidity) = 0.97$$

S=S_{Sunny}=[2+,3-]

$$Entropy(S) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

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

$$Entropy(S) = 0.97$$

S_{Strong} = [1+,1-]

$$Entropy(S_{Strong}) = 1.0$$

S_{Weak} =[1+,2-]

$$Entropy(S_{Strong}) = 0.9183$$

$$Gain(S_{Sunny}, Wind) = 0.0912$$

$$Gain(S_{Sunny}, Temp) = 0.570$$

$$Gain(S_{Sunny}, Humidity) = 0.97$$

$$Gain(S_{Sunny}, Wind) = 0.0912$$

$Gain(S_{Sunny}, Humidity) = 0.97$

Day	Outlook	Temp.	Humidity	Wind	Decision
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
10	Rain	Mild	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

S=S_{Rain}=[3+,2-]

$$Entropy(S) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

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

$$Entropy(S) = 0.97$$

S_{Hot}=[0+,0-]

$$Entropy(S_{Hot}) = 0.0$$

S_{Mild}=[2+,1-]

$$Entropy(S_{Mild}) = 0.9183$$

S_{cool}=[1+,1-]

$$Entropy(S_{cool}) = 1.0$$

$Gain(S_{Sunny}, Temp) = 0.0192$

S=S_{Rain}=[3+,2-]

$$Entropy(S) = -P_+ \, log_2 \, P_+ - P_- log_2 \, P_-$$

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

$$Entropy(S) = 0.97$$

S_{High}=[1+,1-]

$$Entropy(S_{\text{High}}) = 1.0$$

S_{Normal}=[2+,1-]

$$Entropy(S_{Normal}) = 0.0$$

$$Gain(S_{Sunny}, Humidity) = 0.0192$$

S=S_{Rain}=[3+,2-]

$$Entropy(S) = -P_+ log_2 P_+ - P_- log_2 P_-$$

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

$$Entropy(S) = 0.97$$

S_{Strong} = [0+,2-]

$$Entropy(S_{\text{Strong}}) = 0.0$$

S_{Weak} =[3+,0-]

$$Entropy(S_{\text{Strong}}) = 0.0$$

$$Gain(S_{Sunny}, Wind) = 0.97$$

$$Gain(S_{Sunny}, Temp) = 0.0192$$

$$Gain(S_{Sunny}, Humidity) = 0.0192$$

$$Gain(S_{Sunny}, Wind) = 0.97$$