Decision Tree Example

Author - Kanav Bansal
You can find me on **Linkedin** -
www.linkedin.com/in/kanavbansal
Or http://www.thataiguy.com

Consider this binary classification data set:

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

→ X_q Rain Hot Normal Weak ?

Given X_q → Outlook = 'Rain' & Temp = 'Hot' &
Humidity = 'Normal' & Wind = 'Weak'

$$IG(Outlook, Y) = 0.94 - \boxed{0.693} = \boxed{0.246}$$
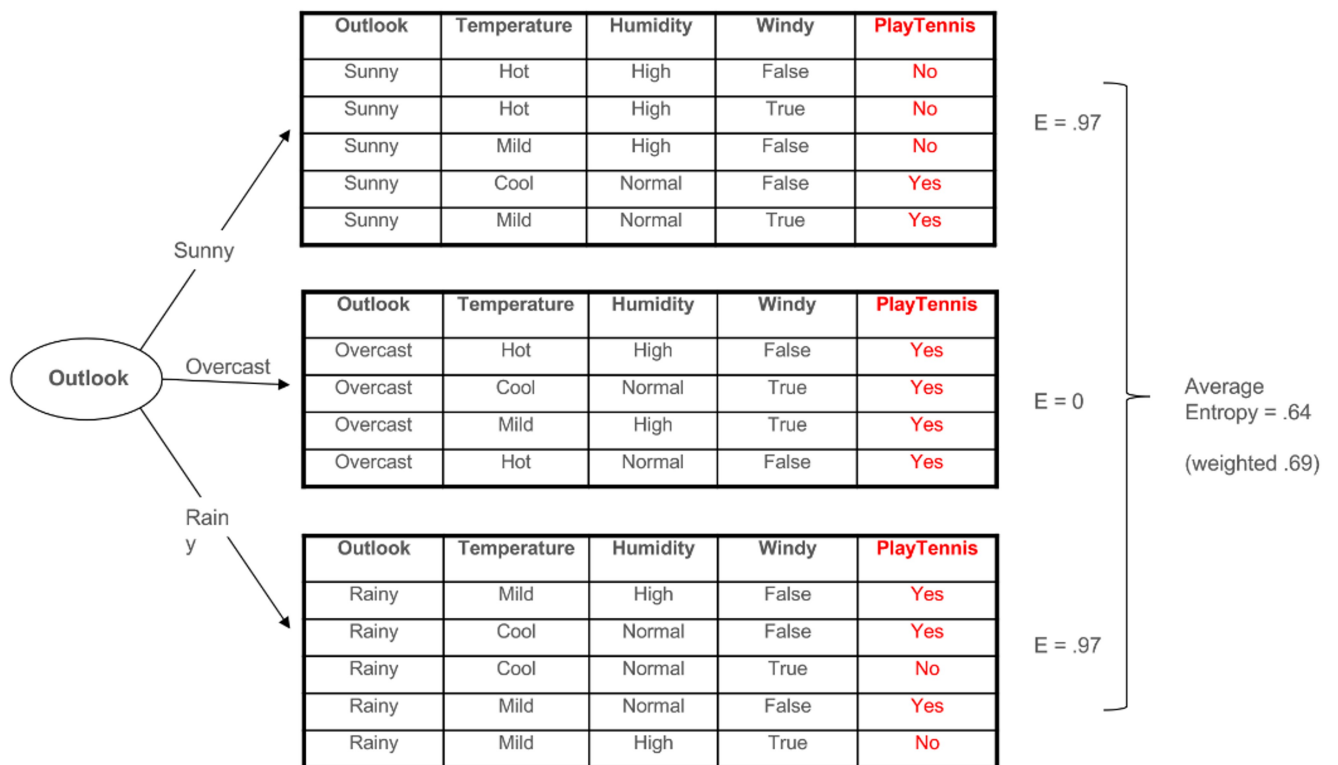
$$IG(Temp, Y) = 0.94 - \boxed{0.91} = \boxed{0.029}$$

$$IG(Humidity, Y) = 0.94 - \boxed{0.789} = \boxed{0.151}$$

$$IG(Wind, Y) = 0.94 - \boxed{0.892} = \boxed{0.048}$$

minimize weighted entropy

maximize information gain

Partitioning the Data Set

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	True	Yes

E = .97

Outlook	Temperature	Humidity	Windy	PlayTennis
Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes

E = 0

Outlook	Temperature	Humidity	Windy	PlayTennis
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Rainy	Mild	Normal	False	Yes
Rainy	Mild	High	True	No

E = .97

Outlook

Sunny

Overcast

Rainy

Average Entropy = .64

(weighted .69)

Create a decision tree model that can predict the **Hours Played** given that we know the **Outlook, Temperature, Humidity and Wind**.

Day	Outlook	Temperature	Humidity	Wind	HoursPlayed
1	Rainy	Hot	High	FALSE	25
2	Rainy	Hot	High	TRUE	30
3	Overcast	Hot	High	FALSE	46
4	Sunny	Mild	High	FALSE	45
5	Sunny	Cool	Normal	FALSE	52
6	Sunny	Cool	Normal	TRUE	23
7	Overcast	Cool	Normal	TRUE	43
8	Rainy	Mild	High	FALSE	35
9	Rainy	Cool	Normal	FALSE	38
10	Sunny	Mild	Normal	FALSE	46
11	Rainy	Mild	Normal	TRUE	48
12	Overcast	Mild	High	TRUE	52
13	Overcast	Hot	Normal	FALSE	44
14	Sunny	Mild	High	TRUE	30