Yes

No

Leaf Node

If (Entropy = 0)

Devide the Dataset into Branches

Consider the Input Attribute with Highest Gain Value or Lowest Entroy value as Root Element

Find “Information Gain”

of Each Input Attribute using the Entropy of Target Variable and Entropy of Input Attribute

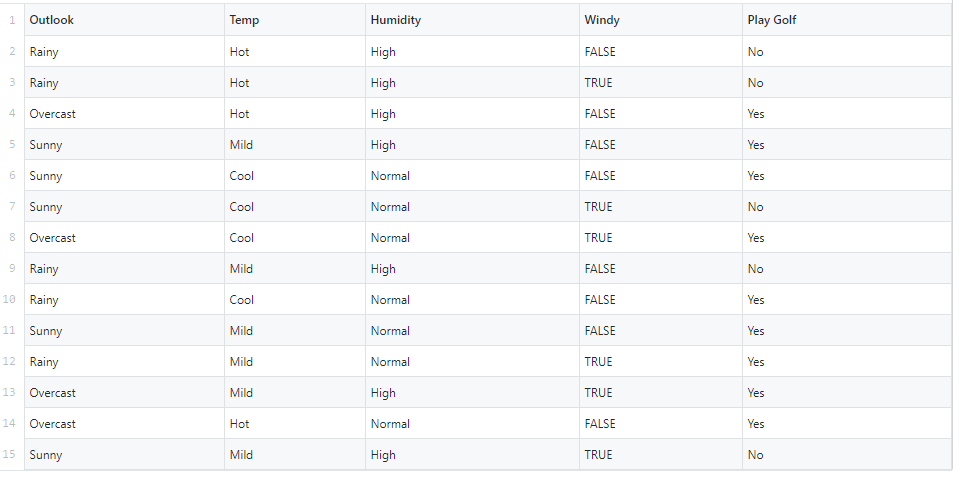
Find “Entropy”

of Each Input Attribute

Find “Entropy” Target Variable

Target Variable

Input Attributes



**Entropy (Play Golf)** = Entropy (5, 9)

= Entropy (0.36, 0.64)

= -(0.36 log2 0.36) - (0.64 log2 0.64)

= 0.94

**Entropy (Play Golf, Outlook)** = P(Sunny)\*E(3, 2) + P(Overcast)\*E(4, 0) + P(Rainy)\*E(2, 3)

= (5/14) \* 0.971 + (4/14) \* 0.0 + (5/14) \* 0.971

= 0.693

**Entropy (Play Golf, Temp)** = P(Hot)\*E(2, 2) + P(Mild)\*E(4, 2) + P(Cool)\*E(3, 1)

= (4/14) \* 1 + (6/14) \* 0.918295 + (4/14) \* 0.811278

= 0.285714 + 0.393555 + 0.231794

= 0.911063

**Entropy (Play Golf, Humidity)** = P(High)\*E(3, 4) + P(Normal)\*E(6, 1)

= (7/14) \* 0.985228 + (7/14) \* 0.591672

= 0.492614 + 0.295836

= 0.78845

**Entropy (Play Golf, Windy)** = P(TRUE)\*E(3, 3) + P(FALSE)\*E(6, 2)

= (6/14) \* 0.1 + (8/14) \* 0.463587

= 0.428571 + 0.463587

= 0.892159

**Gain(Play Golf, Outlook) = Entropy(Play Golf) – Entropy(Play Golf, Outlook)**

= 0.940 – 0.693

= 0.247

**Gain(Play Golf, Temp) = Entropy(Play Golf) – Entropy(Play Golf, Temp)**

= 0.940 – 0.911

= 0.029

**Gain(Play Golf, Humidity) = Entropy(Play Golf) – Entropy(Play Golf, Humidity)**

= 0.940 – 0.788

= 0.152

**Gain(Play Golf, Windy) = Entropy(Play Golf) – Entropy(Play Golf, Windy)**

= 0.940 – 0.892

= 0.048