DS Assignment #3

Charlotte Seane

Attribute	Information Gain
Sky	0.811
Airtemp	0.811
Humidity	0.122
Wind	0.000
Water	0.122
Forecast	0.311

Calculations:

Using $Entropy = -P_{\oplus}\log_2 P_{\oplus} - P_{\ominus}\log_2 P_{\ominus}$ & $Entropy(S,A) = Entropy(S) - \sum_{v \in value(A)} \frac{|S_v|}{|S|} Entropy(S_v)$

In this case, "Yes" is the positive, "No" is the negative

Our data set shows 3/4 "Yes" and 1/4 "No". Therefore we get

$Entropy(S) = -3/4 \log_2(3/4) - 1/4 \log_2(1/4) = 0.811$

For Gain(S,Sky): Values(Sky) = {Sunny, Rainy}

$S = [3+, 1-]$ $S_{Sunny} = [3+, 0-]$ $S_{Rainy} = [0+, 1-]$

So $Gain(S,Sky) = 0.811 - 3/4\, Entropy(S_{Sunny}) - 1/4\, Entropy(S_{Rainy})$

$= 0.811 - 3/4(-1\log_2(1) - 0\cdot\log_2(0)) - 1/4(0\cdot\log_2(0) - 1\log_2(1))$

$= \boxed{0.811}$

For Gain(S, Airtemp): Values(Airtemp) = {Warm, Cold}

$S = [3+, 1-]$ $S_{Warm} = [3+, 0-]$ $S_{Cold} = [0+, 1-]$

So $Gain(S, Airtemp) = \boxed{0.811}$ (because calculation is same as Gain(S, Sky))

For Gain(S, Humidity): Values(Humidity) = {Normal, High}

$S = [3+, 1-]$ $S_{Normal} = [2+, 0-]$ $S_{High} = [2+, 1-]$

So $Gain(S, Humidity) = 0.811 - 1/4\,(0) - 3/4(-2/3\log_2(2/3) - 1/3\log_2(1/3))$

$= \boxed{0.122}$ ← Entropy = 0 since 1 positive, 0 negatives

For Gain(S, Wind): Values(Wind) = {Strong}

$S = [3+, 1-]$ $S_{Strong} = [3+, 1-]$

So, $Gain(S, Wind) = 0.811 - 1 \cdot Entropy(S_{Strong}) = 0.811 - (-3/4\log_2(3/4) - 1/4\log_2(1/4))$

$= 0.811 - 0.811 = \boxed{0}$