Entropy for the whole data-set

Out of 14 instances n(Yes) => 8, n(NO) => 6 } Total => 14

Entropy (S) => $-p(yes)\log_2[p(yes)] - p(NO)\log_2[p(NO)]$

$\Rightarrow -\left[\frac{8}{14}\log_2\left(\frac{8}{14}\right) + \frac{6}{14}\log_2\left(\frac{6}{14}\right)\right]$

$\Rightarrow -\left[\frac{4}{7}\log_2\left(\frac{4}{7}\right) + \frac{3}{7}\log_2\left(\frac{3}{7}\right)\right]$

$\Rightarrow -[-0.13887 - 0.15770]$

$\Rightarrow -[-0.46134 + (-0.525)]$

Entropy (S) => 0.98514

Class	Stay in hostel	Total values	p(yes)	p(NO)
8	yes = 2, no = 1	3	2/3	1/3
9	yes = 2, no = 1	3	2/3	1/3
10	yes = 1, no = 3	4	1/4	3/4
11	yes = 1, no = 4	4	3/4	1/4
		n => 14		

Which feature is chosen as Root Node?

Entropy (class = 8) => $-p(yes)\log_2[p(yes)] - p(NO)\log_2[p(NO)]$

$\Rightarrow -\left[\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right]$

$\Rightarrow -[-0.38997 + (-0.52832)]$

Entropy (class = 8) => 0.918

Entropy (class = 9) $\Rightarrow -p(yes)\log_2[p(yes)] - p(NO)\log_2[p(NO)]$

$\Rightarrow -[2/3\log_2(2/3) + 1/3\log_2(1/3)]$

Entropy [class=9] $\Rightarrow$ 0.918

Entropy [class = 10] $\Rightarrow -p(NO)\log_2[p(yes)] - p(NO)\log_2[p(NO)]$

$\Rightarrow -[1/4\log_2(1/4) + 3/4\log_2(3/4)]$

$\Rightarrow -[-1/2 - 0.31127]$

Entropy [class=10] $\Rightarrow$ 0.811

Entropy [class = 11] $\Rightarrow -p(yes)\log_2(p(yes)) - p(NO)\log_2(p(NO))$

$\Rightarrow -[3/4\log_2(3/4) + 1/4\log_2(1/4)]$

$\Rightarrow -[0.75\log_2(0.75) + 0.25\log_2(0.25)]$

Entropy [class = 11] $\Rightarrow$ 0.811

Information gain from the 'class' feature:

$I(class) \Rightarrow \left(\frac{3}{14} * 0.918\right) + \left(\frac{3}{14} * 0.918\right) + \left(\frac{4}{14} * 0.811\right) + \left(\frac{4}{14} * 0.81\right)$

$I(class) \Rightarrow 0.8574$

Information Gain from the 'class' feature:

Information Gain (class) = Entropy (class) - Information (class)

$= 0.985 - 0.857$

Information Gain (class) = 0.1276

Information Gain for the 'Gender' feature

Gender	Stay in Hostel	n	p(yes)	p(no)
Male	Yes = 5 No = 3	8	5/8	3/8
Female	Yes = 3 No = 3	6	3/6	3/6
		14		

Entropy (Gender) $\Rightarrow -p(yes)\log_2[p(yes)] - p(NO)\log_2[p(NO)]$

$\Rightarrow -\left[8/14 \log_2(8/14) + 6/14 \log_2(6/14)\right]$

$\Rightarrow -\left[-0.46134 - 0.5238\right]$

Entropy (Gender) $\Rightarrow 0.98514$

14

Entropy (Gender = 'Male')

$\Rightarrow -p(yes)\log_2[p(yes)] - p(NO)\log_2[p(NO)]$

$\Rightarrow -\left[\frac{5}{8}\log_2\left(\frac{5}{8}\right) + \frac{3}{8}\log_2\left(\frac{3}{8}\right)\right]$

$\Rightarrow -[-0.42379 - 0.53063]$

$\Rightarrow 0.95442$

Entropy [Gender = 'Male'] $\Rightarrow$ 0.95442

Entropy [Gender = 'Female'] $\Rightarrow -p(yes)\log_2\left(p(yes)\right) - p(NO)\log_2\left(p(NO)\right)$

$\Rightarrow -\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$

Entropy [Gender = 'Female'] $\Rightarrow$ 1

Information from 'Gender' feature:

Information (Gender) $\Rightarrow \left(\frac{8}{14} * 0.953 + \frac{6}{14} * 1\right)$

Information (Gender) $\Rightarrow$ 0.972

Information Gain from 'Gender' feature:

Information Gain (Gender) $\Rightarrow$ Entropy (Gender) − Information (Gender)

$\Rightarrow 0.98514 - 0.972$

Information Gain Gender $\Rightarrow$ 0.012