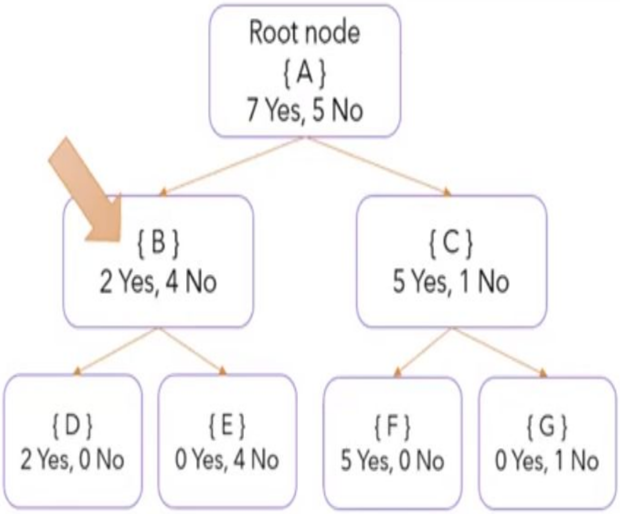
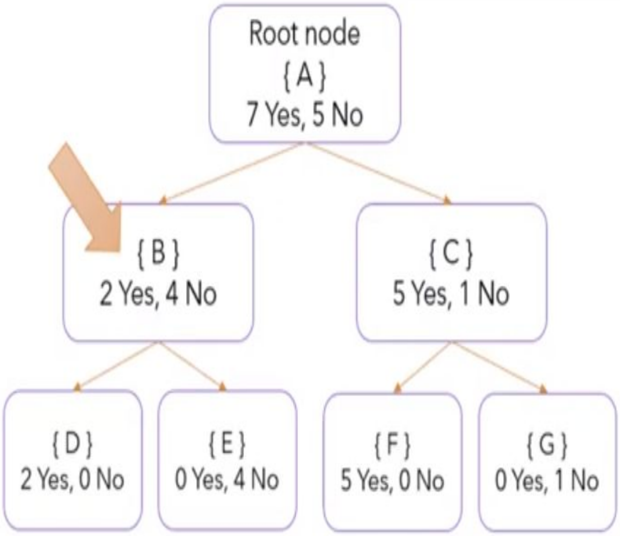


Entropy

$$Entropy(S) = \sum_{i=1}^{c} -p_i \, log_2(p_i)$$

$p(Yes) = \frac{P(Yes)}{P(Yes)+P(No)}$, for "Yes" component of node A

$$Entropy(B) = -\frac{2}{2+4}\, log_2 \left(\frac{2}{2+4}\right) - \frac{4}{4+2}\, log_2 \left(\frac{4}{4+2}\right)$$



Entropy

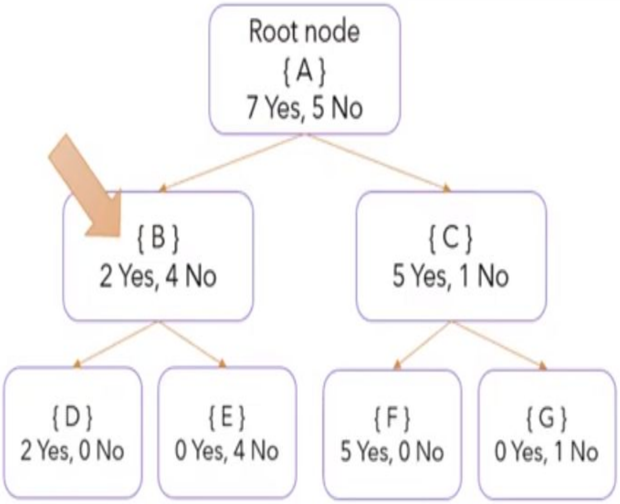
$$Entropy(S) = \sum_{i=1}^{c} -p_i \, log_2(p_i)$$

$p(Yes) = \frac{P(Yes)}{P(Yes)+P(No)}$, for "Yes" component of node A

$$Entropy(B) = -\frac{2}{2+4}\, log_2\left(\frac{2}{2+4}\right) - \frac{4}{4+2}\, log_2\left(\frac{4}{4+2}\right)$$

$$= -\frac{2}{6}\, log_2\left(\frac{2}{6}\right) - \frac{4}{6}\, log_2\left(\frac{4}{6}\right)$$

$$= 0.92\ bits$$



Entropy

$$Entropy(S) = \sum_{i=1}^{c} -p_i \, log_2(p_i)$$

$p(Yes) = \frac{P(Yes)}{P(Yes)+P(No)}$, for "Yes" component of node A

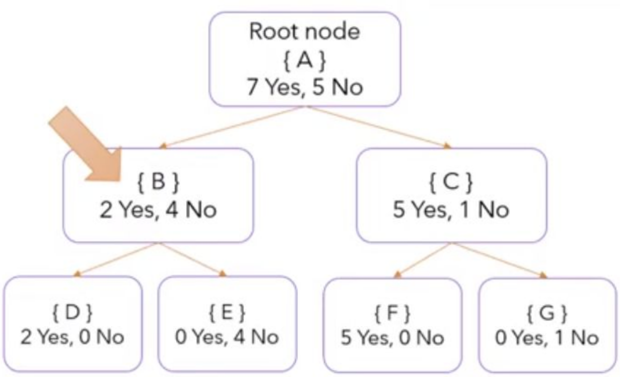
$$Entropy(B) = -\frac{2}{2+4} \, log_2 \left(\frac{2}{2+4}\right) - \frac{4}{4+2} \, log_2 \left(\frac{4}{4+2}\right)$$

$$= -\frac{2}{6} \, log_2 \left(\frac{2}{6}\right) - \frac{4}{6} \, log_2 \left(\frac{4}{6}\right)$$

$$= 0.92 \; bits$$

$$Entropy(C) = -\frac{5}{6} \, log_2 \left(\frac{5}{6}\right) - \frac{1}{6} \, log_2 \left(\frac{1}{6}\right)$$

$$= 0.65 \text{ bits}$$


Root node
{ A }
7 Yes, 5 No
{ B }
2 Yes, 4 No
{ C }
5 Yes, 1 No
{ D }
2 Yes, 0 No
{ E }
0 Yes, 4 No
{ F }
5 Yes, 0 No
{ G }
0 Yes, 1 No

Entropy

$$Entropy(S) = \sum_{i=1}^{c} -p_i \, log_2(p_i)$$

$p(Yes) = \frac{P(Yes)}{P(Yes)+P(No)}$, for "Yes" component of nod

$$Entropy(B) = -\frac{2}{2+4} \, log_2 \left(\frac{2}{2+4}\right) - \frac{4}{4+2} \, log_2 \left(\frac{4}{4+2}\right)$$

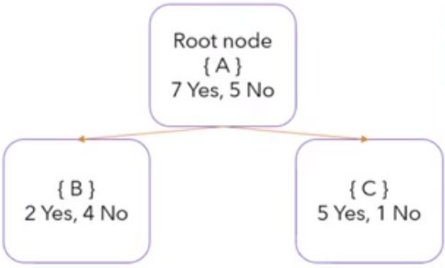
$$= -\frac{2}{6} \, log_2 \left(\frac{2}{6}\right) - \frac{4}{6} \, log_2 \left(\frac{4}{6}\right)$$

$$= 0.92 \; bits$$

$$Entropy(C) = -\frac{5}{6} \, log_2 \left(\frac{5}{6}\right) - \frac{1}{6} \, log_2 \left(\frac{1}{6}\right)$$

$$= 0.65 \text{ bits}$$

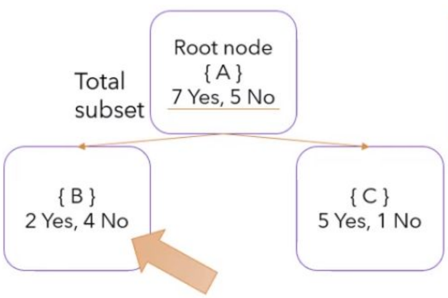
$$\boldsymbol{Entropy\ ranges\ from\ 0\ to\ 1, where\ 0\ consider\ as\ pure\ sub}$$



Information Gain

Information gain calculates the reduction in entropy or surprise from transforming a dataset in some way. Specifically, this metrics measure the **quality of a split.** Information gain can also be used for feature selection, by evaluating the gain of each variable in the context of the target variable

$$Gain(S, A) = Entropy(S) - \sum_{v \in values(A)} \frac{|Sv|}{|S|} Entropy(S_v)$$



$$Entropy(S) = -\frac{7}{12}\, log_2\left(\frac{7}{12}\right) - \frac{5}{12}\, log_2\left(\frac{5}{12}\right)$$

= 0.97 bits

$Entropy(B)$ = 0.92 bits
$Entropy(C)$ = 0.65 bits

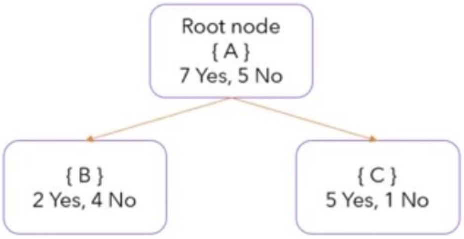
Information Gain

Information gain calculates the reduction in entrop
or surprise from transforming a dataset in some wa
Specifically, this metrics measure the **quality of a split.** Information gain can also be used for feature selection, by evaluating the gain of each variable in the context of the target variable

$$Gain(S, A) = Entropy(S) - \sum_{v \in values(A)} \frac{|Sv|}{|S|} Entropy(S$$

$$Gain(S, A) = 0.97 - \frac{|Sv|}{|S|}(0.92) - \frac{|Sv|}{|S|}(0.65)$$

$$Gain(S, A) = 0.97 - \frac{6}{12}(0.92) - \frac{6}{12}(0.65)$$



$Entropy(C)$ = 0.65 bits

$$GI = 1 - [(\frac{5}{5+1})^2 + (\frac{1}{1+5})^2]$$

$$= 1 - [(\frac{5}{6})^2 + \left(\frac{1}{6}\right)^2]$$

$$= 0.3$$

Gini Impurity

Gini Impurity is a measurement of the likelihood of an incorrect classification of a new instance of a random variable, if that new instance were randomly classified according to the distribution of class labels from the data set.

$$GI = 1 - \sum_{i=1}^{n} p^2 \Rightarrow GI = 1 - [p(Yes)^2 + p(No)^2]$$