

## A Decision Tree Terminology

### 2. Splitting

It is a process of dividing a node into two or more sub-nodes.

### 4. Leaf/ Terminal Node

The node that does not split is called a Leaf or Terminal node.

### 5. Pruning

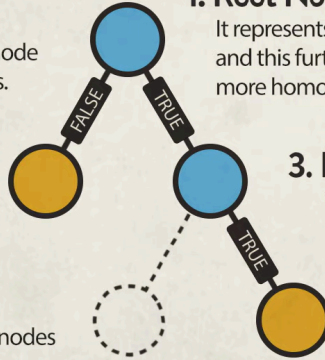
When we remove sub-nodes of a decision node, this process is called pruning.

### 1. Root Node

It represents entire population or sample and this further gets divided into two or more homogeneous sets.

### 3. Decision Node

When a sub-node splits into further sub-nodes, it is called a decision node.



Age	Competition	Type	Profit
old	Yes	S/w	Down
old	No	S/w	Down
old	No	H/w	Down
old	Yes	S/w	Down
mid	Yes	H/w	Up
mid	No	H/w	Up
mid	No	S/w	Up
mid	No	S/w	Up
new	Yes	H/w	Up
new	No	H/w	Up
new	No	S/w	Up

Age	Down	Up
old	3	0
mid	2	2
new	0	3

$$I(\text{old}) = -\left[\frac{3}{3} \log_2\left(\frac{3}{3}\right) + \frac{0}{3} \log_2\left(\frac{0}{3}\right)\right] = 0 \times 3/10 = 0$$

$$I(\text{mid}) = -\left[\frac{2}{4} \log_2\left(\frac{2}{4}\right) + \frac{2}{4} \log_2\left(\frac{2}{4}\right)\right] = 1 \times 4/10 = 0.4$$

$$I(\text{new}) = -\left[\frac{0}{3} \log_2\left(\frac{0}{3}\right) + \frac{3}{3} \log_2\left(\frac{3}{3}\right)\right] = 0 \times 3/10 = 0$$

$$E(\text{Age}) = 0.4$$

$$I.G = -\frac{P}{P+N} \log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N} \log_2\left(\frac{N}{P+N}\right)$$

$$E(A) = \sum_{i=1}^v \frac{P_i + N_i}{P+N} I(P_i N_i)$$

$$\text{Gain} = I.G - E(A)$$

$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2}$$

$$I.G = -\left[\frac{5}{10} \log_2\left(\frac{5}{10}\right) + \frac{5}{10} \log_2\left(\frac{5}{10}\right)\right]$$

$$= -\left[0.5 \times \log_2 2^{-1} + 0.5 \log_2 2^{-1}\right]$$

$$= -\left[0.5 \times (-1 \log_2 2) + 0.5 \times (-1 \log_2 2)\right]$$

$$= -[-0.5 - 0.5] = -[-1]$$

$$I.G = 1$$

$$\text{Gain} = 1 - 0.4$$

$$= 0.6$$

Age	Competition	Type	Profit
old	Yes	S/w	Down
old	No	S/w	Down
old	No	H/w	Down
old	Yes	S/w	Down
mid	Yes	H/w	Down
mid	Yes	H/w	Up
mid	No	H/w	Up
mid	No	S/w	Up
mid	No	S/w	Up
new	Yes	S/w	Up
new	No	H/w	Up
new	No	S/w	Up



Gain(Age)  $\rightarrow 0.60$

Gain(Competition)  $\rightarrow 0.124$

Gain(Type)  $\rightarrow 0$

$$I \cdot G = 1$$

id	$x_0$	$x_1$	$x_2$	$x_3$	$x_4$	$y$
0	4.3	4.9	4.1	4.7	5.5	0
1	3.9	6.1	5.9	5.5	5.9	0
2	2.7	4.8	4.1	5.0	5.6	0
3	6.6	4.4	4.5	3.9	5.9	1
4	6.5	2.9	4.7	4.6	6.1	1
5	2.7	6.7	4.2	5.3	4.8	1

id
2
0
2
4
5
5

id
2
1
3
1
4
4

id
4
1
3
0
0
2

id
3
3
2
5
1
2

$x_0, x_1$

$x_2, x_3$

$x_2, x_4$

$x_1, x_3$

2.8 6.2 4.3 5.3 5.5

Bootstrap + Aggregating  
(Bagging)

