

Decision Tree: Gini Index

	Trend	Interest	Volume	Return
1	Positive	Low	High	Up
2	Negative	High	Low	Down
3	Positive	Low	High	Up
4	Positive	High	High	Up
5	Negative	Low	High	Down
6	Positive	Low	Low	Down
7	Negative	High	High	Down
8	Negative	Low	High	Down
9	Positive	Low	Low	Down
10	Positive	High	High	Up

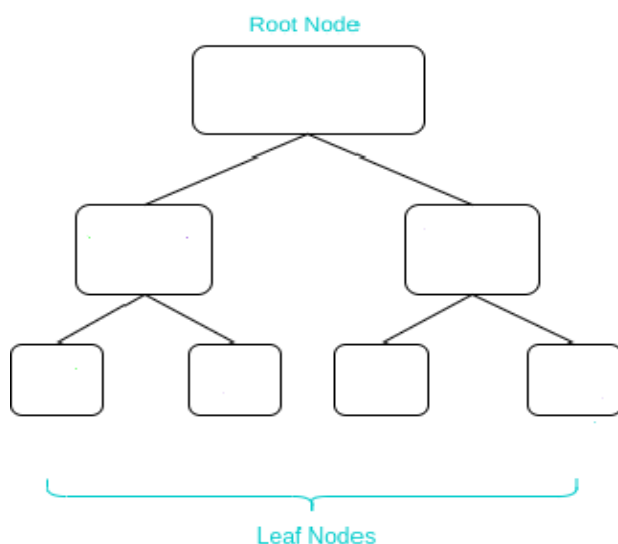
Trend (positive = 6, Negative = 4)

Interest (High= 4, Low= 6)

Volume(High = 7, Low = 3)

The best split is chosen by **minimizing the Gini Impurity**.

$$\text{Gini Index} = 1 - \sum_{i=1}^n (P_i)^2$$



$$\text{Gini Index} = 1 - \sum_{i=1}^n (P_i)^2$$

Formula of Gini Index:

$$\text{Gini} = 1 - \sum_{i=1}^n (P_i)^2$$

Gini Trend:

Gini (Trend = positive) (Return = up + Return = Down)

$$1 - \left[\left(\frac{4}{6} \right)^2 + \left(\frac{2}{6} \right)^2 \right] = 0.45$$

Gini (Trend = Negative)

$$1 - \left[(0)^2 + \left(\frac{4}{4} \right)^2 \right] = 0$$

We will calculate sum of Gini Indexes for the Trend feature:

$$\text{Gini (Trend)} = \left(\frac{6}{10} \times 0.45 \right) + \left(\frac{4}{10} \times 0 \right)$$

$$\boxed{\text{Trend} = 0.27}$$

Gini Interest:

Gini (Interest = High)

$$1 - \left[\left(\frac{2}{4} \right)^2 + \left(\frac{2}{4} \right)^2 \right] = 0.5$$

Gini (Interest = Low)

$$1 - \left[\left(\frac{2}{6} \right)^2 + \left(\frac{4}{6} \right)^2 \right] = 1 - (0.11 + 0.44) = 0.45$$

We will calculate weighted sum of Gini Indexes for the Interest feature:

$$\text{Gini (Interest)} = \left(\frac{4}{10} \times 0.5 \right) + \left(\frac{6}{10} \times 0.45 \right)$$

$$\boxed{\text{Interest} = 0.47}$$

Gini (Volume):

Gini (Volume = High)

$$1 - \left[\left(\frac{4}{7} \right)^2 + \left(\frac{3}{7} \right)^2 \right] = 1 - (0.33 + 0.18) = 0.49$$

Gini (Volume = Low)

$$1 - \left[(0)^2 + \left(\frac{3}{3} \right)^2 \right] = 1 - (0 + 1) = 0$$

Weighted sum of the Gini Indexes for the feature volume:

$$\text{Gini (Volume)} = \left(\frac{7}{10} \times 0.49 \right) + \left(\frac{3}{10} \times 0 \right)$$

$$\boxed{\text{Volume} = 0.34}$$

Attributes/ Features	Gini Index
Trend	0.27
Interest	0.47
Volume	0.34

The Trend has the **minimum Gini Impurity value** so, It will be chosen as the root decision to split our data.

Attributes / features	Gini Index
Trend	0.27
Interest	0.41
volume	0.34

	Trend	Interest	volume	Return
1	Positive	Low	High	up
2	Positive	Low	High	up
3	Positive	High	High	up
4	Positive	Low	Low	Down
5	Positive	Low	Low	Down
6	Positive	High	High	up

Interest	Gini (Interest = High)
	$1 - \left[\left(\frac{2}{2} \right)^2 + (0)^2 \right] = 1 - 1 \Rightarrow 0$
Gini (Interest = Low)	$1 - \left[\left(\frac{2}{4} \right)^2 + \left(\frac{1}{4} \right)^2 \right] = 1 - 0.5 \Rightarrow 0.5$
Gini (Interest)	$\left(\frac{2}{6} \times 0 \right) + \left(\frac{4}{6} \times 0.5 \right) = 0.33$
Gini (volume)	$1 - \left[\left(\frac{4}{4} \right)^2 + (0)^2 \right] = 0$
	$1 - \left[\left(\frac{2}{2} \right)^2 + (0)^2 \right] = 0$
Gini (volume)	$\left(\frac{4}{6} \times 0 \right) + \left(\frac{2}{6} \times 0 \right) = 0$

Features	Gini index
Interest	0.33
volume	0

	Trend	Interest	Volume	Return
1	Positive	Low	High	Up
2	Positive	Low	High	Up
3	Positive	High	High	Up
4	Positive	Low	Low	Down
5	Positive	Low	Low	Down
6	Positive	High	High	Up

