Decision trees are often used while implementing machine learning algorithms. The hierarchical structure of a decision tree leads us to the final outcome by traversing through the nodes of the tree. Each node consists of an attribute or feature which is further split into more nodes as we move down the tree. But how do we decide:

- Which attribute/feature should be placed at the root node?
- Which features will act as internal nodes or leaf nodes?

To decide this, and how to split the tree, we use splitting measures like Gini Index, Information Gain, etc

What is Gini Index?

Gini Index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

But what is actually meant by 'impurity'?
If all the elements belong to a single class, then it can be called pure. The degree of Gini Index varies between 0 and 1,

where,
'0' denotes that all elements belong to a certain class or there exists only one class (pure), and
'1' denotes that the elements are randomly distributed across various classes (impure).

A Gini Index of '0.5 'denotes equally distributed elements into some classes.

Example of Gini Index

let us calculate Gini Index for past trend, open interest, trading volume and return in the following manner with the example data:

Past Trend	Open Interest	Trading Volume	Return
Positive	Low	High	Up
Negative	High	Low	Down
Positive	Low	High	Up
Positive	High	High	Up
Negative	Low	High	Down
Positive	Low	Low	Down
Negative	High	High	Down
Negative	Low	High	Down
Positive	Low	Low	Down
Positive	High	High	Up

Calculation of Gini Index

We will now calculate the Gini Index with the following -

- Calculating the Gini Index for past trend
- Calculating the Gini Index for open interest

Calculating the Gini Index for past trend

Since the past trend is positive 6 number of times out of 10 and negative 4 number of times, the calculation will be as follows:

P(Past Trend=Positive): 6/10

P(Past Trend=Negative): 4/10

- If (Past Trend = Positive & Return = Up), probability = 4/6
- If (Past Trend = Positive & Return = Down), probability = 2/6

Gini Index = 1 - ((4/6)^2 + (2/6)^2) = 0.45

- If (Past Trend = Negative & Return = Up), probability = 0
- If (Past Trend = Negative & Return = Down), probability = 4/4

Gini Index = 1 - ((0)^2 + (4/4)^2) = 0

- Weighted sum of the Gini Indices can be calculated as follows:

Gini Index for Past Trend = (6/10)0.45 + (4/10)0 = 0.27

Calculating the Gini Index for open interest

Coming to open interest, the open interest is high 4 times and low 6 times out of total 10 times and is calculated as follows:

P(Open Interest=High): 4/10

P(Open Interest=Low): 6/10

- If (Open Interest = High & Return = Up), probability = 2/4
- If (Open Interest = High & Return = Down), probability = 2/4

Gini Index = 1 - ((2/4)^2 + (2/4)^2) = 0.5

- If (Open Interest = Low & Return = Up), probability = 2/6
- If (Open Interest = Low & Return = Down), probability = 4/6

Gini Index = 1 - ((2/6)^2 + (4/6)^2) = 0.45

- Weighted sum of the Gini Indices can be calculated as follows:

Gini Index for Open Interest = (4/10)0.5 + (6/10)0.45 = 0.47

Calculating the Gini Index for trading volume

Trading volume is 7 times high and 3 times low and is calculated as follows:

P(Trading Volume=High): 7/10

P(Trading Volume=Low): 3/10

- If (Trading Volume = High & Return = Up), probability = 4/7
- If (Trading Volume = High & Return = Down), probability = 3/7

Gini Index = 1 - ((4/7)^2 + (3/7)^2) = 0.49

- If (Trading Volume = Low & Return = Up), probability = 0
- If (Trading Volume = Low & Return = Down), probability = 3/3

Gini Index = 1 - ((0)^2 + (1)^2) = 0

- Weighted sum of the Gini Indices can be calculated as follows:

Gini Index for Trading Volume = (7/10)0.49 + (3/10)0 = 0.34

Gini Index attributes or features

Attributes/Features	Gini Index
Past Trend	0.27
Open Interest	0.47
Trading Volume	0.34