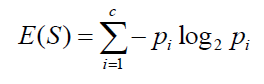
**What is Information Gain?**

Information Gain is used to determine which feature/attribute gives us the maximum information about a class. It is based on the concept of entropy, which is the degree of uncertainty, impurity or disorder. It aims to reduce the level of entropy starting from the root node to the leave nodes.

**Formula for Entropy**

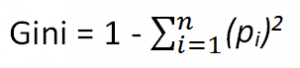


‘*p*’, denotes the probability and E(S) denotes the entropy. Entropy is not preferred due to the ‘log’ function as it increases the computational complexity.

## What is Gini Index?

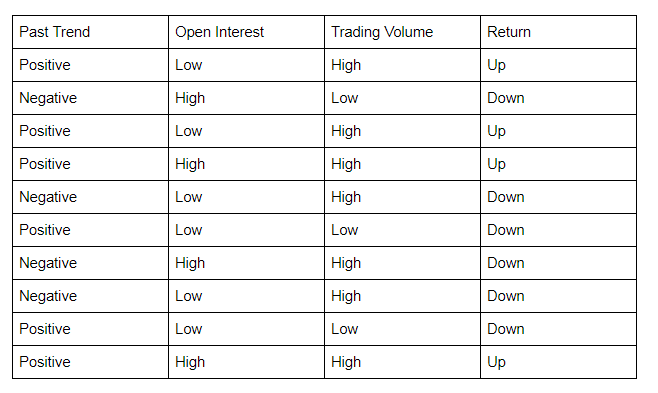
Gini index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen. But what is actually meant by ‘impurity’? If all the elements belong to a single class, then it can be called pure. The degree of Gini index varies between 0 and 1, where 0 denotes that all elements belong to a certain class or if there exists only one class, and 1 denotes that the elements are randomly distributed across various classes. A Gini Index of 0.5 denotes equally distributed elements into some classes.

**Formula for Gini Index**

****

**where piis the probability of an object being classified to a particular class.**

**While building the**[**decision tree**](https://quantra.quantinsti.com/course/decision-trees-analysis-trading-ernest-chan)**, we would prefer choosing the attribute/feature with the least Gini index as the root node.**



## Example of Gini Index

**Let’s start by calculating the Gini Index for ‘Past Trend’.**

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

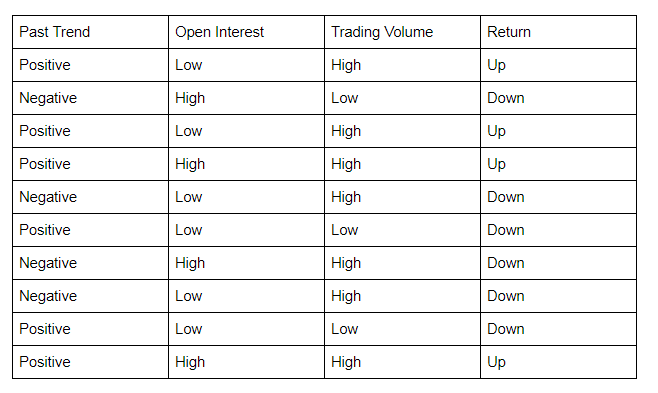
If (Past Trend = Negative & Return = Up), probability = 0

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



**Calculation of Gini Index for Open Interest**

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

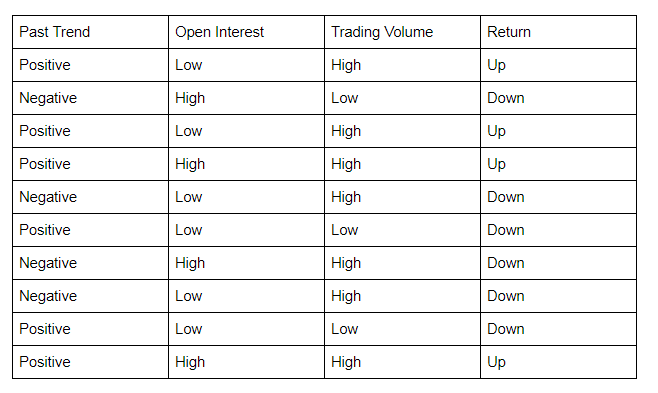
If (Open Interest = Low & Return = Up), probability = 2/6

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



**Calculation of Gini Index for Trading Volume**

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

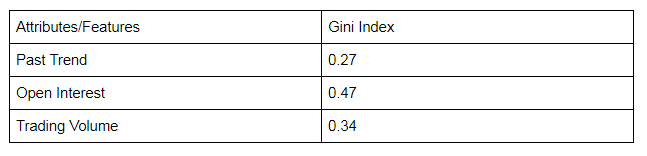
If (Trading Volume = Low & Return = Up), probability = 0

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

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

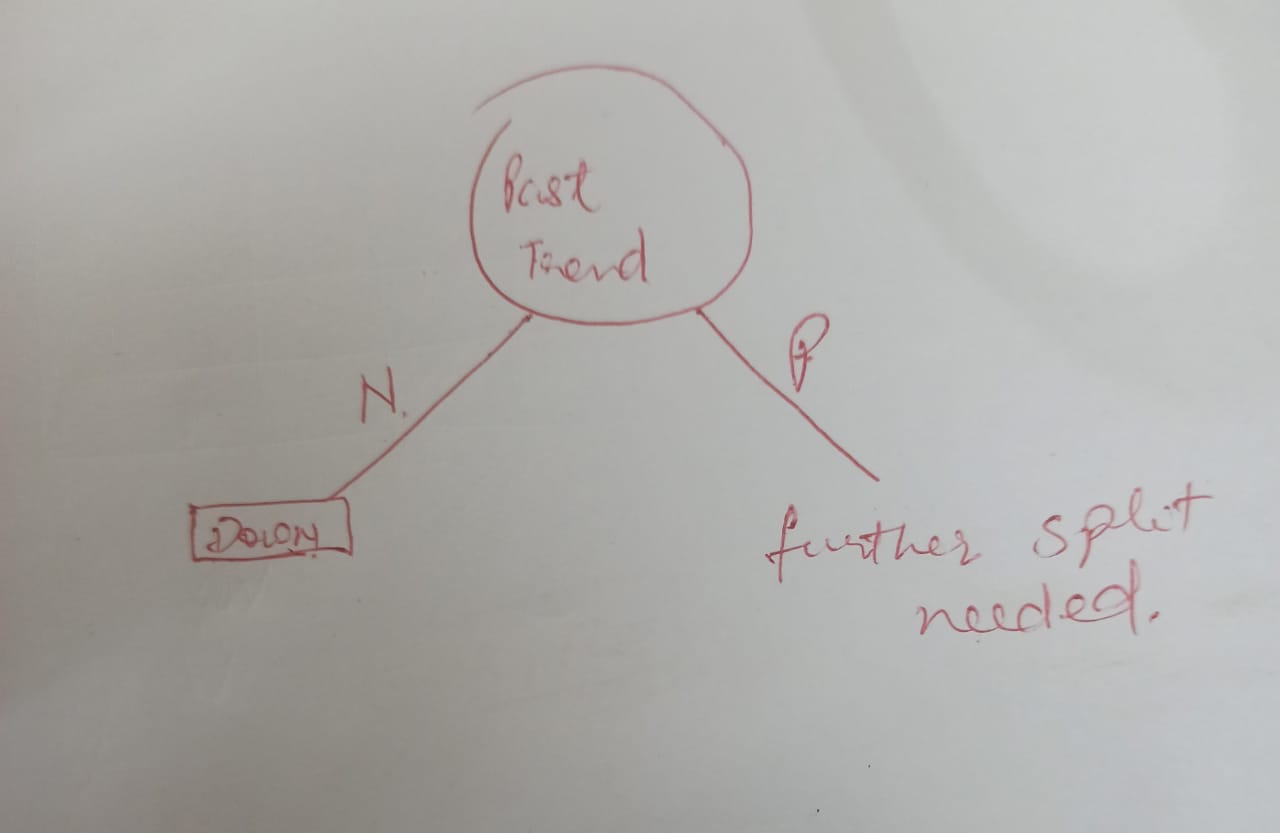
Weighted sum of the Gini Indices can be calculated as follows:

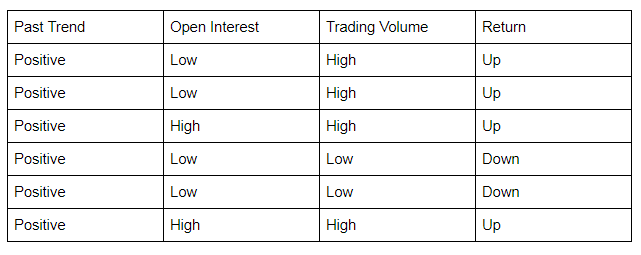
Gini Index for Trading Volume = (7/10)0.49 + (3/10)0 = 0.34



From the above table, we observe that ‘Past Trend’ has the lowest Gini Index and hence it will be chosen as the root node for how decision tree works.

We will repeat the same procedure to determine the sub-nodes or branches of the decision tree.



We will calculate the Gini Index for the ‘Positive’ branch of Past Trend as follows:

**Calculation of Gini Index of Open Interest for Positive Past Trend**

P(Open Interest=High): 2/6

P(Open Interest=Low): 4/6

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

If (Open Interest = High & Return = Down), probability = 0

Gini index = 1 - (sq(2/2) + sq(0)) = 0

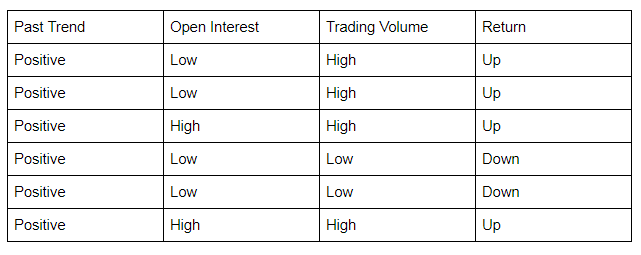
If (Open Interest = Low & Return = Up), probability = 2/4

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

Gini index = 1 - (sq(0) + sq(2/4)) = 0.50

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

Gini Index for Open Interest = (2/6)0 + (4/6)0.50 = 0.33



**Calculation of Gini Index for Trading Volume**

P(Trading Volume=High): 4/6

P(Trading Volume=Low): 2/6

If (Trading Volume = High & Return = Up), probability = 4/4

If (Trading Volume = High & Return = Down), probability = 0

Gini index = 1 - (sq(4/4) + sq(0)) = 0

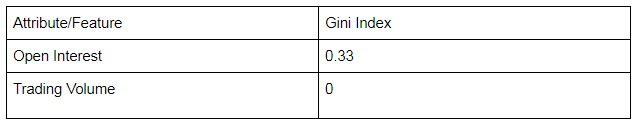
If (Trading Volume = Low & Return = Up), probability = 0

If (Trading Volume = Low & Return = Down), probability = 2/2

Gini index = 1 - (sq(0) + sq(2/2)) = 0

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

Gini Index for Trading Volume = (4/6)0 + (2/6)0 = 0



We will split the node further using the ‘Trading Volume’ feature, as it has the minimum Gini index.

