Two metrics for choosing how to split a tree. Gini measurement is the probability of a random sample being classified incorrectly if we randomly pick a label according to the distribution in a branch.

Entropy is a measurement of information (or rather lack thereof). You calculate the information gain by making a split. Which is the difference in entripies. This measures how you reduce the uncertainty about the label.

**Gini index vs Entropy**

* Decision tree algorithms use information gain to split a node. Gini index or entropy is the criterion for calculating information gain.

Both gini and entropy are measures of impurity of a node. A node having multiple classes is impure whereas a node having only one class is pure.  Entropy in statistics is analogous to entropy in thermodynamics where it signifies disorder. If there are multiple classes in a node, there is disorder in that node.

Information gain is the entropy of parent node minus sum of weighted entropies of child nodes.   
 Weight of a child node is number of samples in the node/total samples of all child nodes. Similarly information gain is calculated with gini score.

**[Check out my Gigs](https://www.fiverr.com/ankitchouksey" \t "_blank)**

