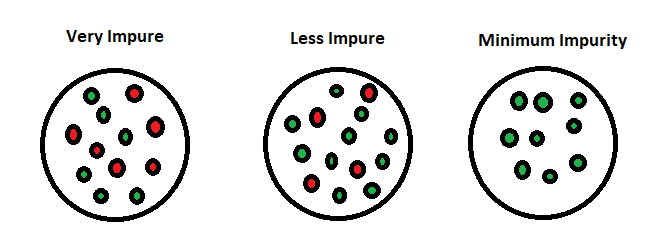
**Entropy and Information Gain to Build Decision Trees in Machine Learning**

A decision tree is a supervised learning algorithm used for both classification and regression problems. Simply put, it takes the form of a tree with branches representing the potential answers to a given question. There are metrics used to train decision trees. One of them is information gain.

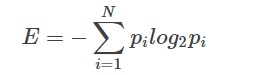
### Entropy

Entropy is an information theory metric that measures the impurity or uncertainty in a group of observations. It determines how a decision tree chooses to split data. The image below gives a better description of the purity of a set.

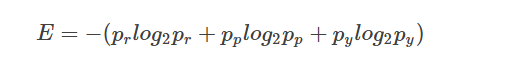


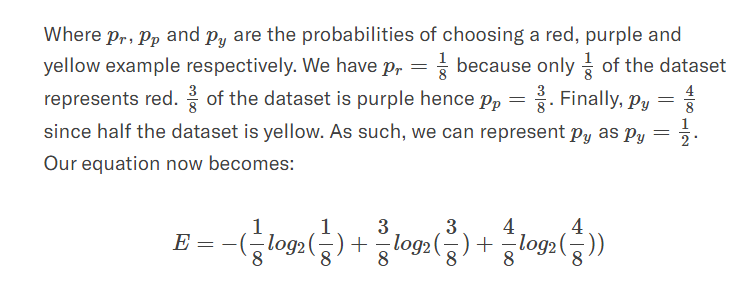
[Source](https://towardsdatascience.com/what-is-entropy-and-information-gain-how-are-they-used-to-construct-decision-trees-aadf7d0728f0)

Consider a dataset with N classes. The entropy may be calculated using the formula below:



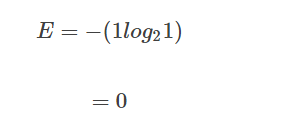
pi is the probability of randomly selecting an example in class i. Let’s have an example to better our understanding of entropy and its calculation. Let’s have a dataset made up of three colors; red, purple, and yellow. If we have one red, three purple, and four yellow observations in our set, our equation becomes:



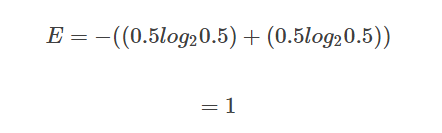


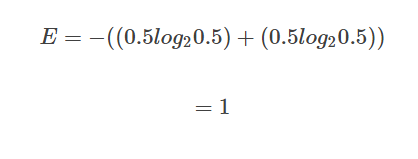
Our entropy would be: 1.41

You might wonder, what happens when all observations belong to the same class? In such a case, the entropy will always be zero.



Such a dataset has no impurity. This implies that such a dataset would not be useful for learning. However, if we have a dataset with say, two classes, half made up of yellow and the other half being purple, the entropy will be one.





This kind of dataset is good for learning.