

Entropy Vs Gini Impurity

Entropy

$$H(S) = -P + \log_2 P + \\ - P - \log_2 P -$$

Let say I have 3 o/p

o/p = 3 categories

$$H(S) = -P_{c1} + \log_2 P_{c1} \\ - P_{c2} \log_2 P_{c2} - \\ P_{c3} \log P_{c3}$$



my formula will be expanded

Whenever dataset is small we can use entropy as the entropy uses Log.

Gini impurity

$$GI = 1 - \sum_{i=1}^n (P)^2$$



For Gini also this formula will get expanded based on the categories

Whenever the dataset is large we can make use of Gini Impurity.