

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

$$H(S) = -P_{C_1} \log_2 P_{C_1} - P_{C_2} \log_2 P_{C_2} - P_{C_3} \log_2 P_{C_3}$$

↑

my formula will be expanded

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

Gini impurity

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

↓↓

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

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