Generate a Decision Tree

- Measure Impurity :

$$Entropy = \sum_j -p_j log_2 p_j \quad Gini\ Index = 1 - \sum_j p_j^2$$

- Information Gain :

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Variance :

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

S^2 = sample variance

x_i = the value of the one observation

$\bar{x}$ = the mean value of all observations

n = the number of observations