

## **1. Decision Tree**

- a. Before we look at the entire forest, let's start by looking at the smaller details instead!
- b. Trees are made by maximizing information gain and minimizing entropy which is the uncertainty of a data point. When we are uncertain of a data point, we require more internal nodes.

## **2. Problems with Decision Trees**

- a. Pruning: Reducing the number of internal nodes
- b. A single decision tree is good but if there's changes to the values in the input variables, the decision tree will like change, due to its hierarchical nature which causes the variance is high

## **3. Random Forest (RF)**

- a. Since its sampling with replacement, some observations may be repeated. Each observation have the same probability of being picked from the original data.
- b. not all variables are chosen!

## **4. Bagging**

- a. if every tree used the same variables or all of the variables, then each tree would pretty much be the same!

## **5. OOB Samples**

- a. These can be used to assess the performance of the decision tree because RF uses the trained model of each tree to predict the OOB samples.

## **6. Coding RF**

- a. these are just a few things you can change to optimize your random forest, but there are many other adjustments you can add to the function!