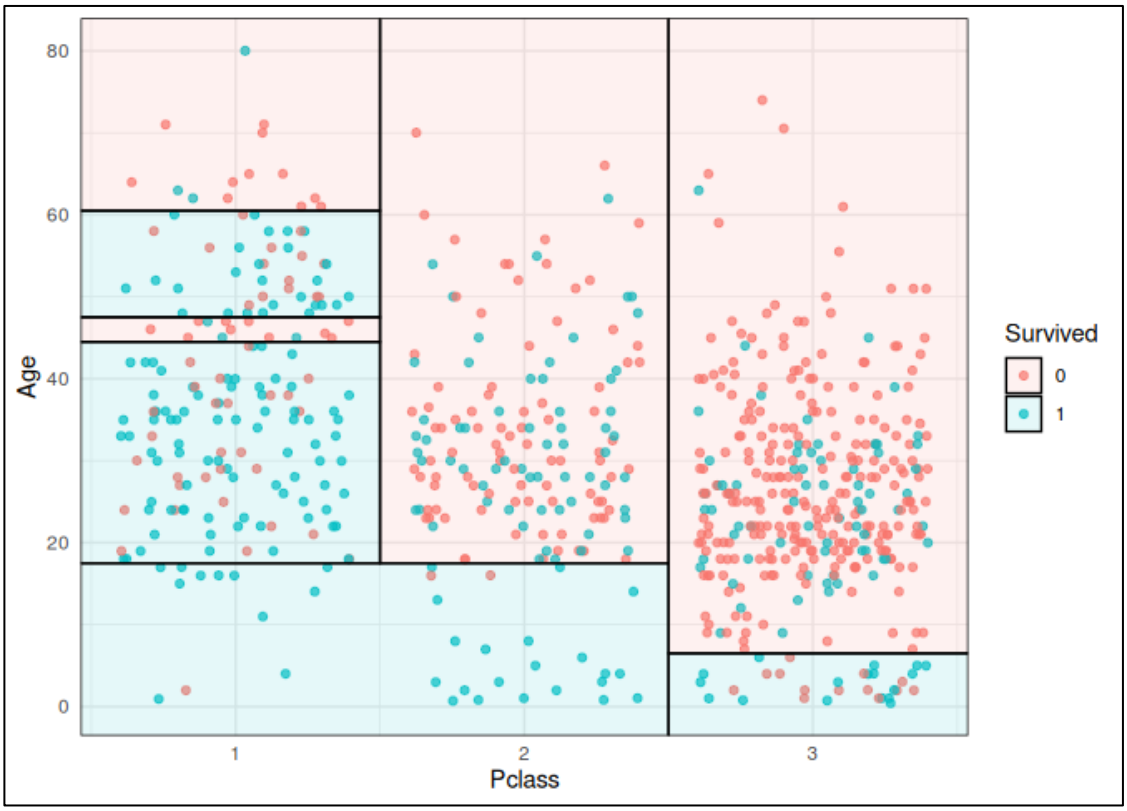
I. Decision Tree (DT)

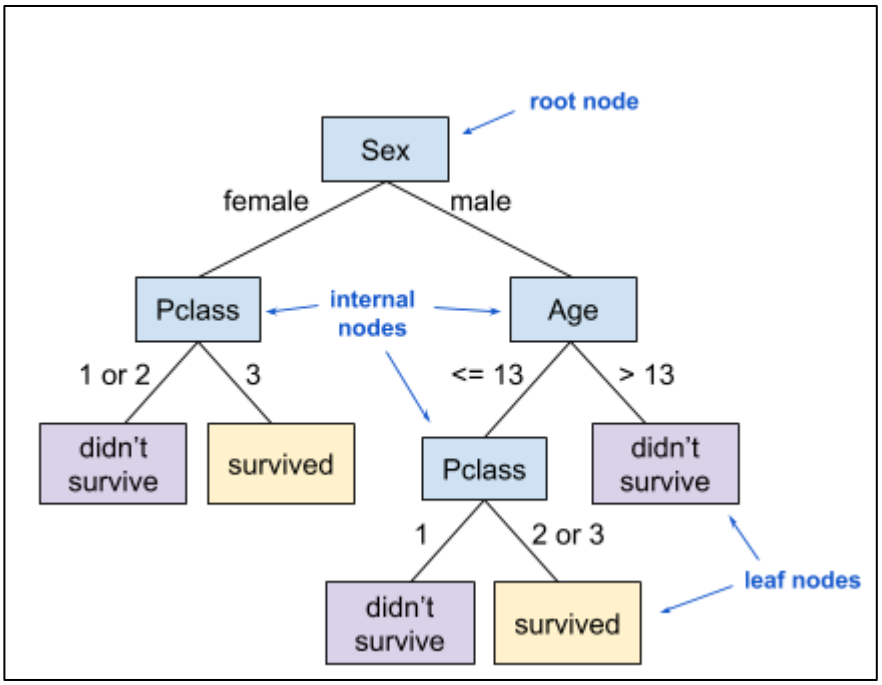
1. Definition:

Decision Tree is a nonparametric algorithm



2. Structure:

There is no rules to use every feature or what order we use the features



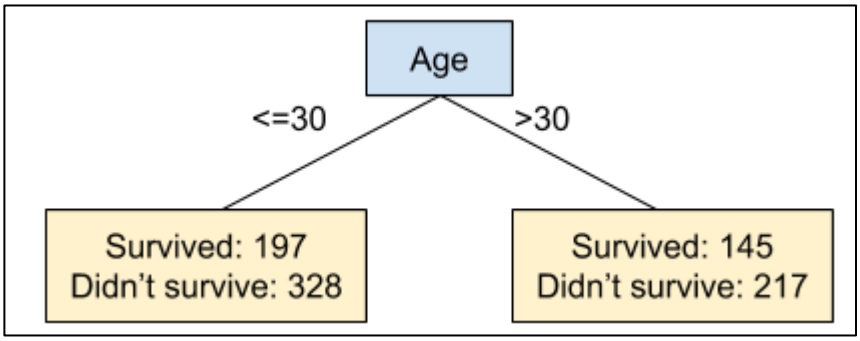
Alternative terms: parent node, child node

3. Information Gain:

Information Gain value is from 0 to 1 where 0 is the information gain of a useless split and 1 is the information gain of a perfect split

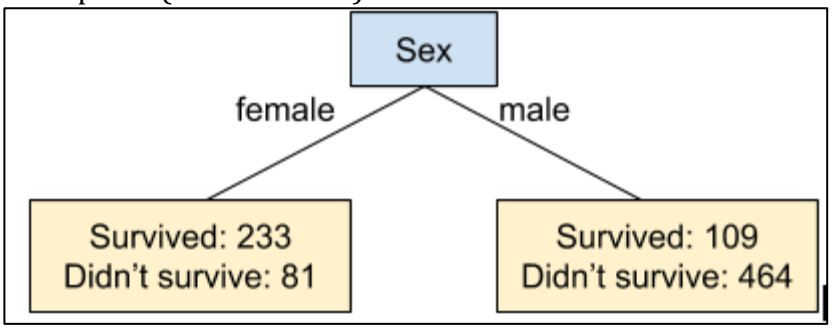
$$\text{Information Gain} = H(\text{S}) - \frac{|A|}{|\text{S}|}H(A) - \frac{|B|}{|\text{S}|}H(B)$$

Example 1: (Gini measure)



- Information gain = 0.4738 - 525/887 * 0.4689 - 362/887 * 0.4802 = 0.0003

Example 2: (Gini measure)



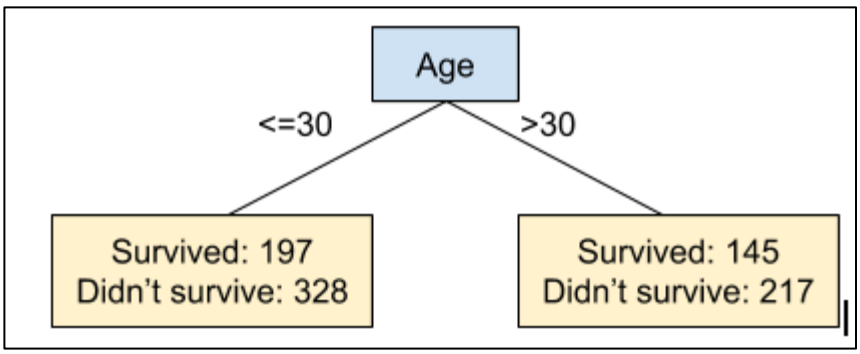
- Information gain = 0.4738 - 314/887 * 0.3828 - 573/887 * 0.3081 = 0.1393

4. Gini impurity:

Gini Impurity value is between 0 and 0.5 where 0.5 is completely impure (50% survived and 50% didn't survive) and 0 is completely pure (100% in the same class)

$$\text{gini} = 2 \times p \times (1 - p)$$

Example:



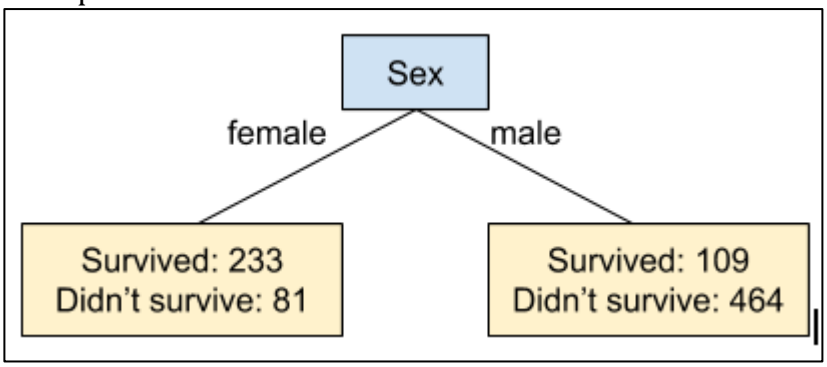
Left side:

- Percent of passengers who survived = 197/(197+328) = 0.3752
- Percent of passengers who didn't survive = 1 - 0.375 = 0.6248
- Gini impurity = 2 * 0.3752 * 0.6248 = 0.4689

Right side:

- Gini impurity = 2 * 145/(145+217) * 217/(145+217) = 0.4802

Example 2:



Left side:

- Gini impurity = 2 * 233/(233+81) * 81/(233+81) = 0.3828

Right side:

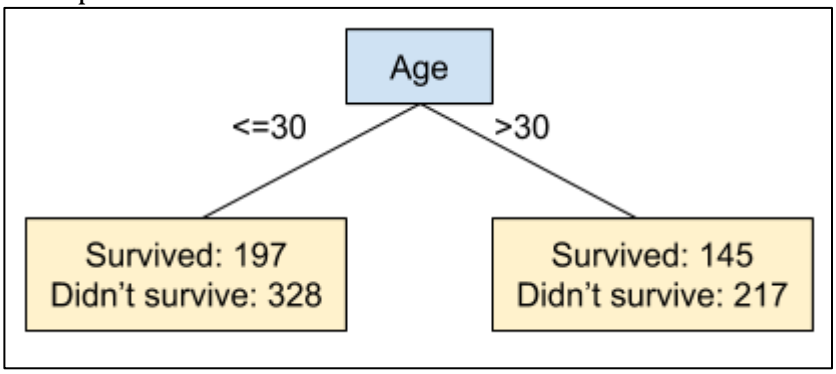
- Gini impurity = 2 * 109/(109+464) * 464/(109+464) = 0.3081

5. Entropy:

Entropy is another measure of purity. It will be a value between 0 and 1 where 1 is completely impure (50% survived and 50% didn't survive) and 0 is completely pure (100% the same class)

$$\text{entropy} = -\left[p \log_2 p + (1-p) \log_2 (1-p)\right]$$

Example:



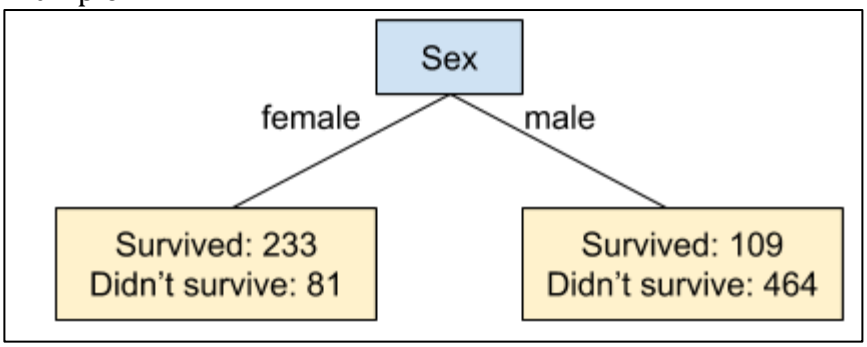
Left side:

- Entropy = 0.9546

Right side:

- Entropy = 0.9713

Example 2:



Left side:

- Entropy = 0.8237

Right side:

- Entropy = 0.7019

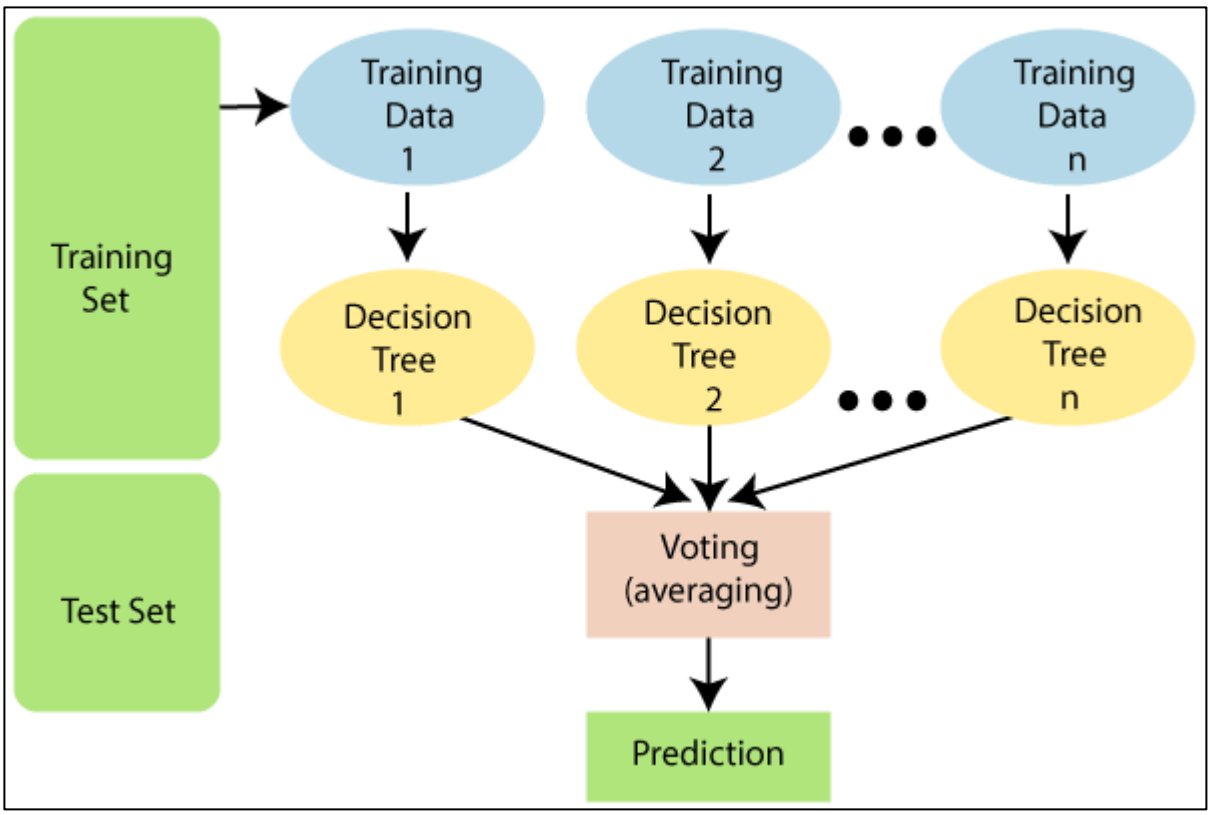
II. Random Forest (RF)

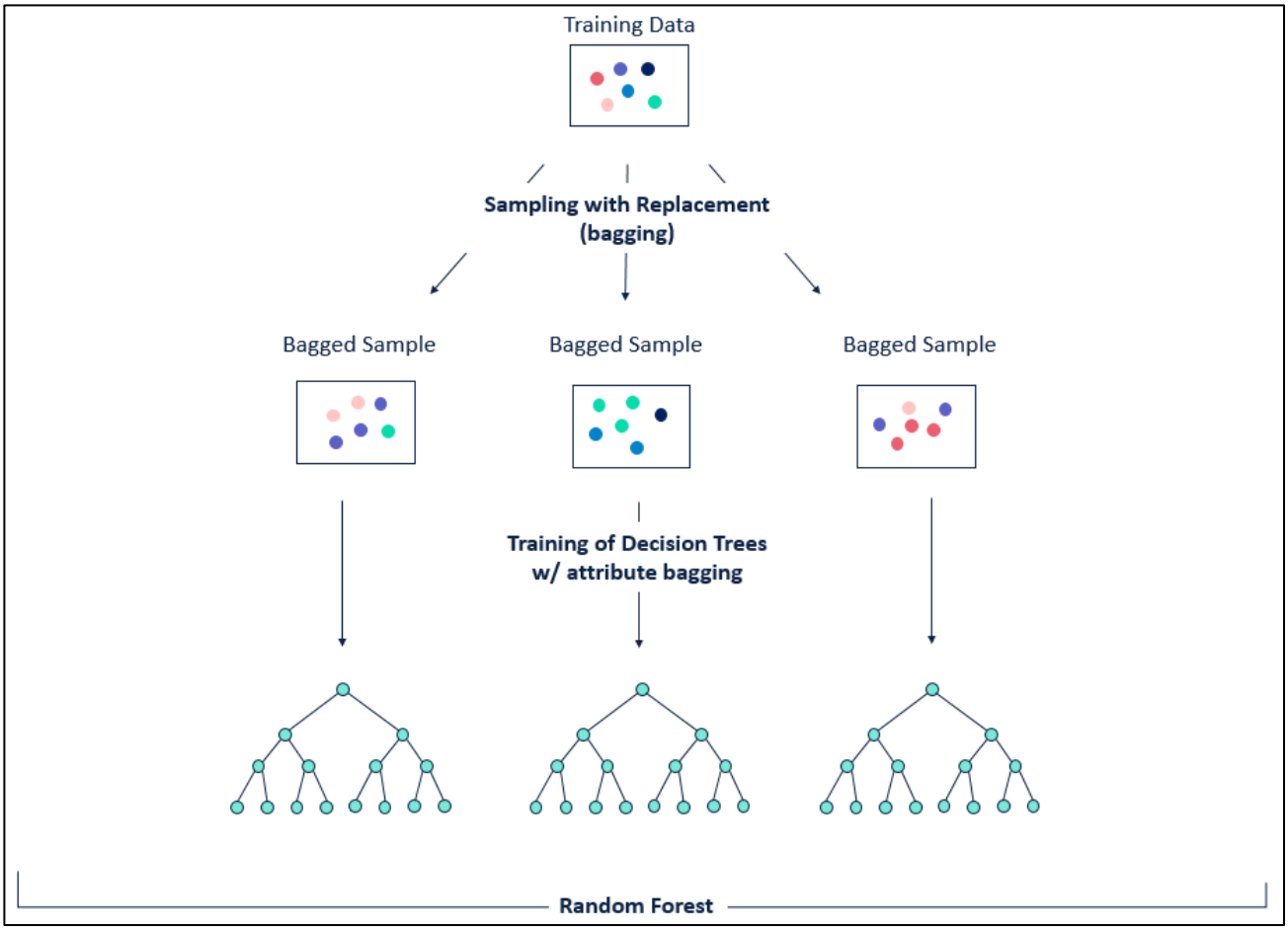
1. Definition:

Random Forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.

2. Structure:

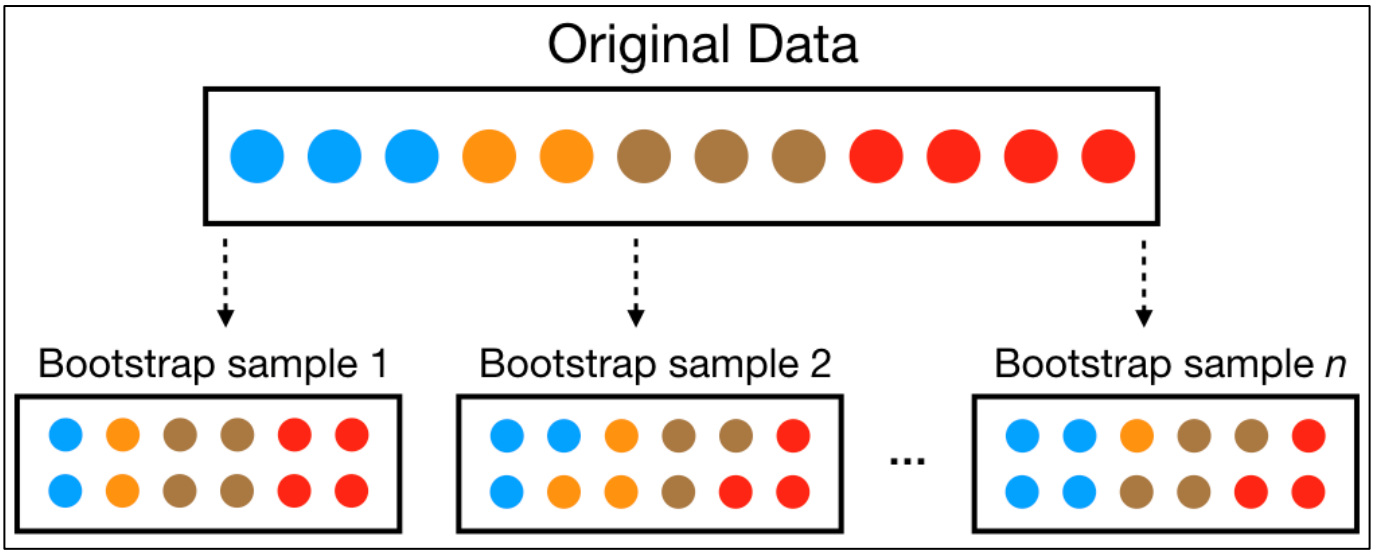
The Random Forest algorithm consists of many decision trees. It is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.


Training Set
Test Set
Training Data 1
Training Data 2
Training Data n
Decision Tree 1
Decision Tree 2
Decision Tree n
Voting (averaging)
Prediction


Training Data
Sampling with Replacement (bagging)
Bagged Sample
Bagged Sample
Bagged Sample
Training of Decision Trees w/ attribute bagging
Random Forest

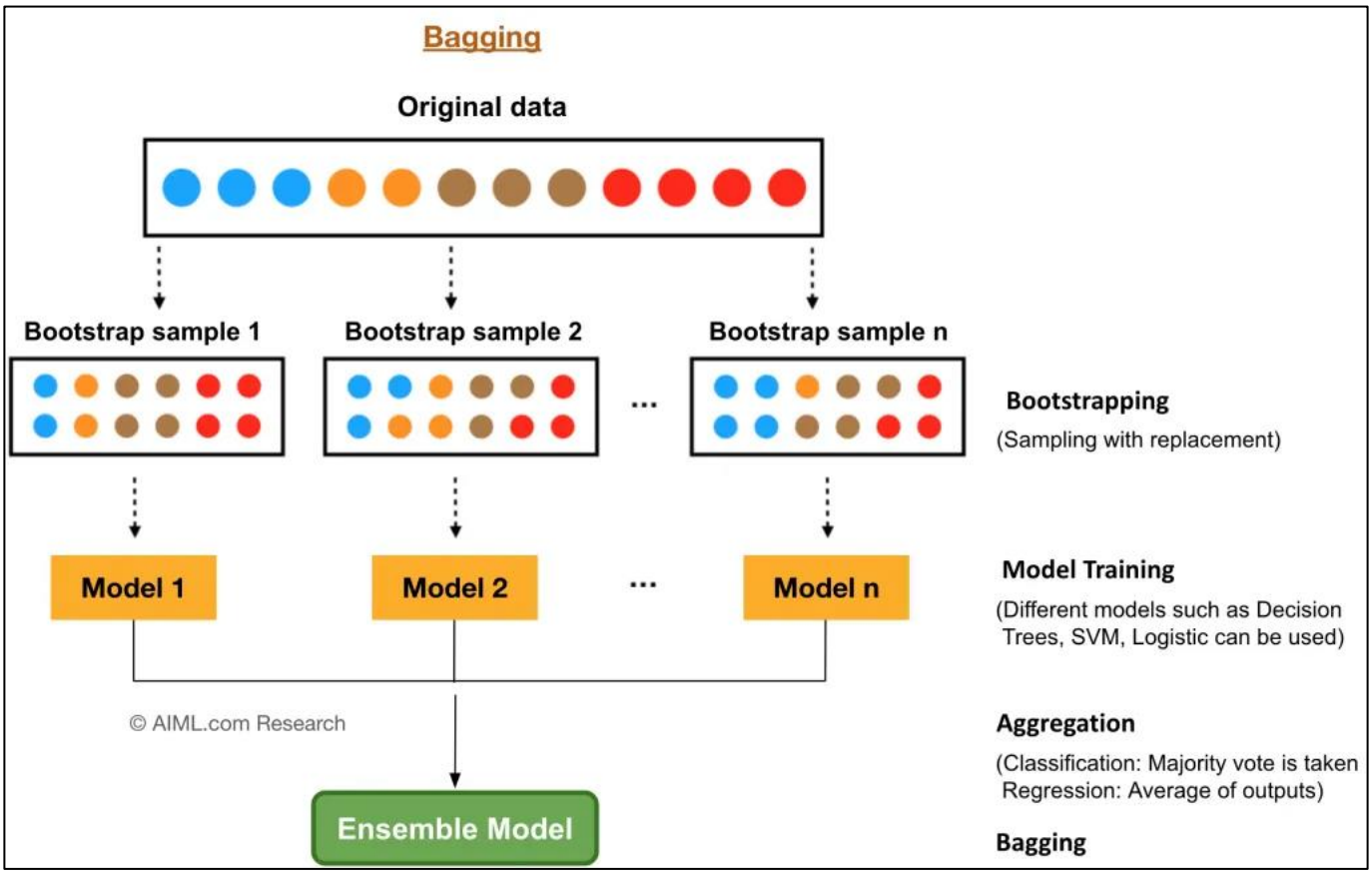
3. Bootstrapping:

- A bootstrapped sample is a random sample of datapoints where we randomly select with replacement datapoints from our original dataset to create a dataset of the same size.
- Randomly selecting with replacement means that we can choose the same datapoint multiple times. This means that in a bootstrapped sample, some datapoints from the original dataset will appear multiple times and some will not appear at all.



4. Bagging (Bootstrap Aggregation):

Bagging (or Bootstrap Aggregation) is a technique for reducing the variance in an individual model by creating an ensemble from multiple models built on bootstrapped samples.



Example:

To bag decision trees, we create multiple (say 10) bootstrapped samples of our training dataset. So, if we have 100 datapoints in our training set, each of the resamples will have 100 datapoints randomly chosen from our training set.

Then, we create a decision tree with each of these 10 resamples (total is 10 decision tree).

To make a prediction, we make a prediction with each of the 10 decision trees and then each decision tree gets a vote. The prediction with the most votes is the final prediction (in classifier problem).

5. Big Note for Random Forest:

Random Forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

III. Genetic Algorithm (GA)

1. Definition:

https://www.geeksforgeeks.org/genetic-algorithms/

https://www.javatpoint.com/genetic-algorithm-in-machine-learning

2. Structure: