Rizky Eka Putra Kusnady
1103204195

Random Forests are made from decision trees, Random forests combine the simplicity of decision trees with flexibility resulting in a vast improvement in accuracy. For make a random forest you must do these steps.

Original Dataset

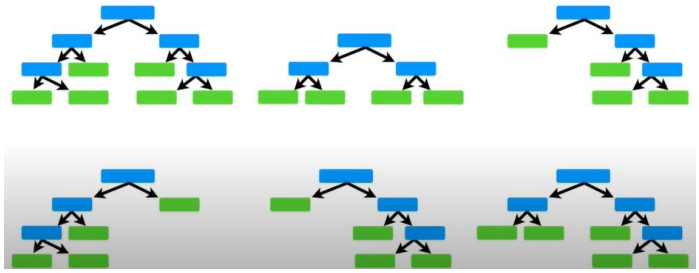
Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No
Yes	No	Yes	167	Yes

Bootstrapped Dataset

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
Yes	Yes	Yes	180	Yes
No	No	No	125	No
Yes	No	Yes	167	Yes
Yes	No	Yes	167	Yes

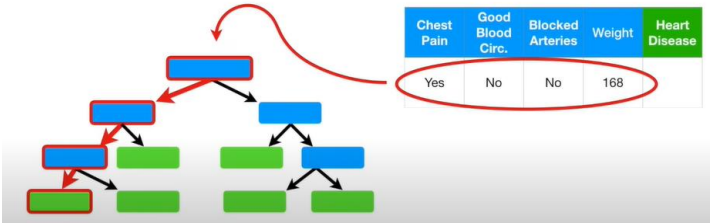
Step1: Create a “bootstrapped” dataset. To create a bootstrapped dataset that is the same size as the original, we just randomly select samples from the original dataset. The important detail is that were allowed to pick the same sample more than once.

Step2: Create a decision tree using the bootstrapped dataset, but only use a random subset of variables (or columns) at each step. From the e.g. considering all 4 variables to figure out how to split the root node is randomly selected.



Step3: After do these two steps, then go back to step 1 and repeat (make a new bootstrapped dataset and build a tree considering a subset of variables at each step.

Using a bootstrapped sample and considering only a subset of variables at each step results in a wide variety of trees, the variety us what makes random forests more effective than individual decision tree. So, in e.g. we want to see the heart disease, we take the data and run it down the first tree that we made.



Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
Yes	No	No	168	

And track of that, and then repeat for all the trees that we made. After running the data down all the trees in the random forest, we see which option received more votes. Bootstrapping the data plus using the aggregate to make a decision is called bagging.