**Random Forest Regression Notes**

* In this section we will focus on Random Forest on Regression Trees rather than classification trees.
* Random Forest is a version of Ensemble Learning.
* Ensemble Learning is when you take multiple algorithms, or the same algorithm multiple times, and you put them together to make something much more powerful than the original.

Steps for Random Forest Regression –

1. Pick a random k data point from your training set.
2. Build a decision tree for these k data points.
3. Choose the number of Ntree you want to build and repeat steps 1 and 2.
4. For a new data point, make each one of your Ntree predict the value of Y for the data point in question, and assign the new data point the average across all of the predicted Y values.

* So, this way you are not predicting based on one tree but predicting the new value of Y based on a forest of trees.
* And that improves the accuracy of your prediction because you are taking the average of many predictions.
* Therefore, even if one of the trees wasn’t built perfectly, if you are using it by itself, you’d get a bad prediction. But because you are using the average it is less likely to happen and you will get a better prediction.
* And secondly, Ensemble algorithms are stable like this, because any changes in your dataset could really impact one tree but for them to impact a forest of trees is much harder.