Out-of Bag evaluation:

→ In Bagging, there might be some data which are never sampled at all. The remaining data which are not sampled are called out-of bag instances.

→ Since model never trains over these datas, they can be used for evaluating the accuracy of model by using these data for prediction.

---

(4) Random Forest:

→ Decision trees have low bias but high variance. And we know bagging technique is a very good solution for decreasing the variance for a decision tree.

→ Instead of using a bagging model with underlying model as a decision tree, we can also use Random Forest which is more convenient & well optimized for decision trees.

→ The main issue of bagging is that there is not much independence among sampled datasets. (There is correlation).

→ The adv. of RF over bagging is that RF makes a tweak in working algorithm of bagging model to decrease the correlation in trees. The idea is to introduce more randomness while creating trees which will help in reducing correlation.

working of RF model

① Different samples are collected from training dataset using bootstrapping.

② on each sample we train our tree model & we allow the trees to grow depth.

③ once the trees all formed, prediction is made by the random forest by aggregating the predictions of all the model. For regression model, the mean of all predictions is the final & for classification mode, the mode of all the predictions is consider the final.

Adv & Disadv of Random Forest

① It can be used for both regression & classification problems.

② since base model is a tree, handling of missing values is easy.

③ It gives very accurate result with very low variance.

) Results of a random forest are hard to interpret in comparison with DT.

) High computational time than other respective models.

) should be used when accuracy is optimal priority,