**Random Forest:** It’s most used technique of Bagging.

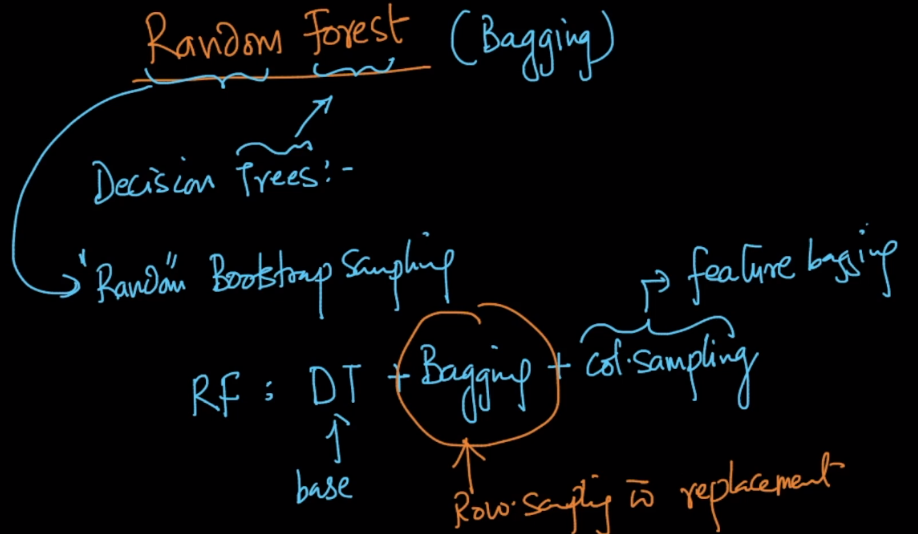
It make use of ‘n’ numbers of decision trees as base learners.

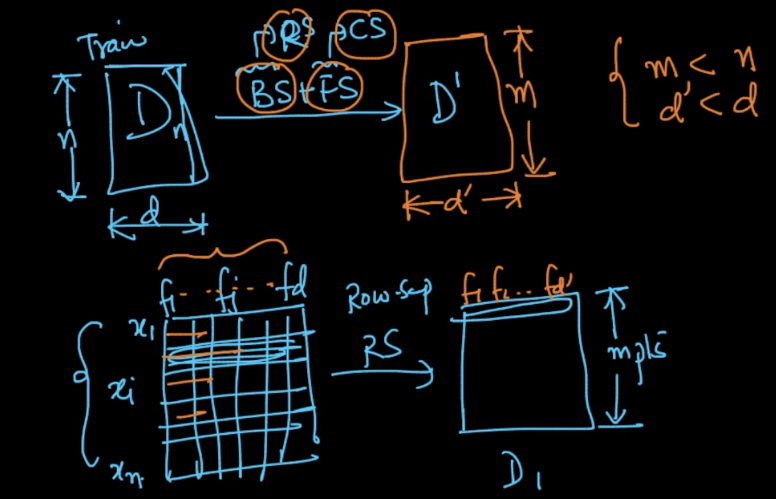
So Random Forest = DT + bagging + column sampling.

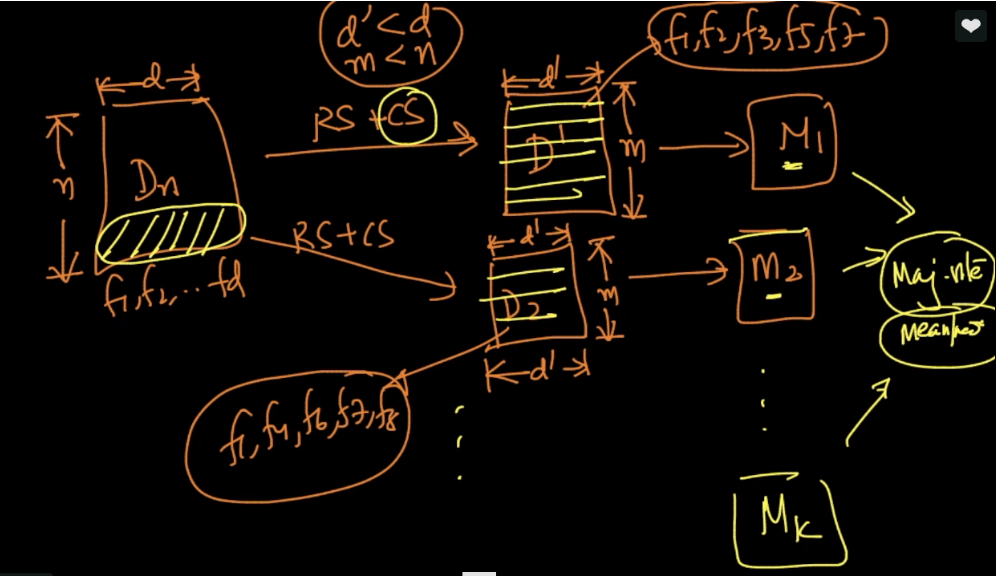
**Column sampling:** pick ‘m’ features from ‘n’ features randomly for each base learner where m < n.

So Random Forest do following things:

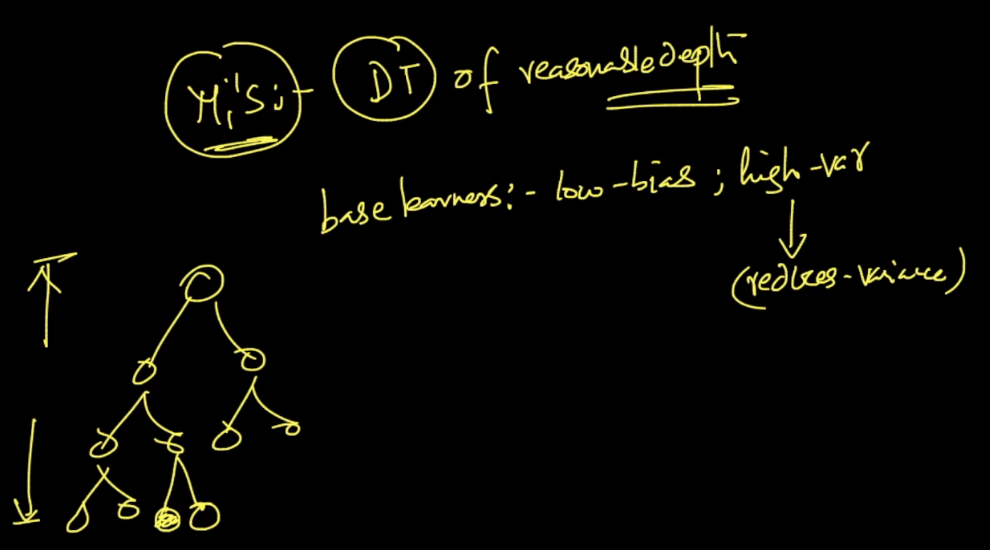
* Pick d’ rows from d rows and m columns from n columns randomly. And create k such samples
* Apply each sample from k samples to a separate DT.
* And eventually use results of all DT to yield result of new query point.







Since in ensemble models we take models with low bias and high variance. SO we allow DT to grow fully or will take large depth, Since Ensemble model reduces variance



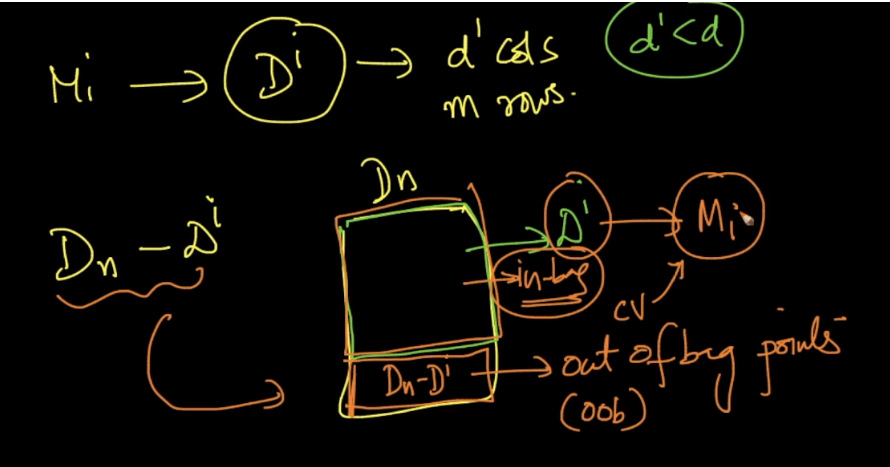
**Out of Bag points and Out of bag error:**

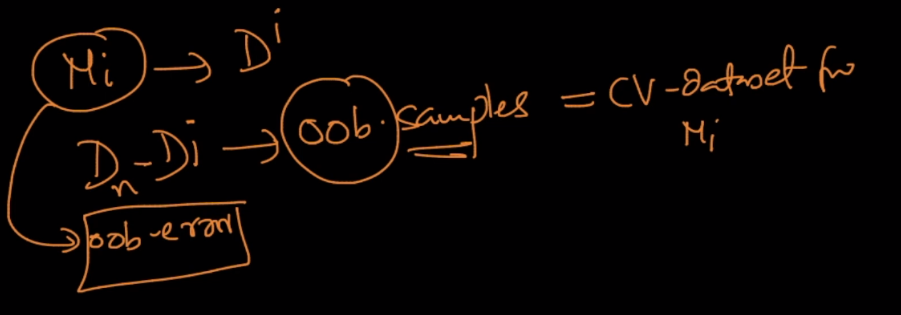
Since here for each base learners we are taking sample containing only d’ rows.

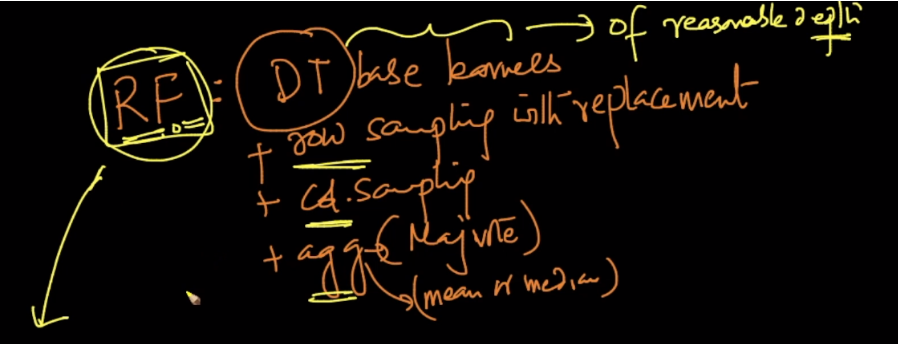
The remaining rows/datapoints ie d-d’ will be called as out of bag points(oob).

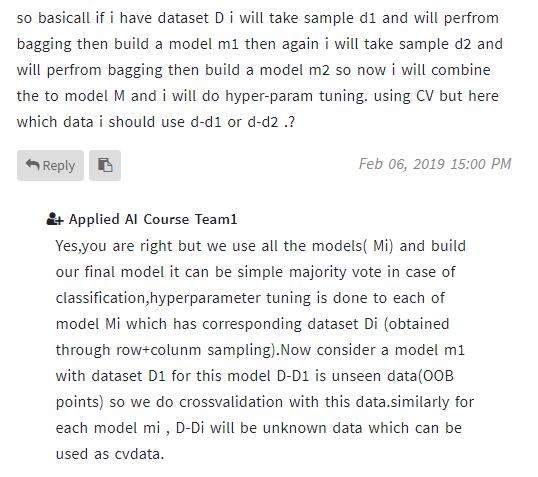
So this OOB points can be used as CV data for that particular base learner.

And the error we get from this oob cv data is called oob error.

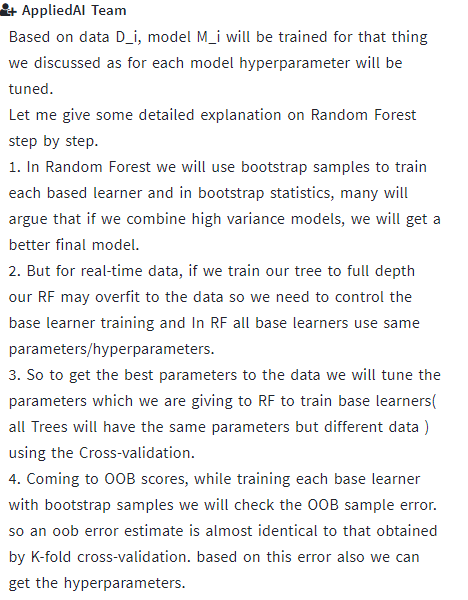




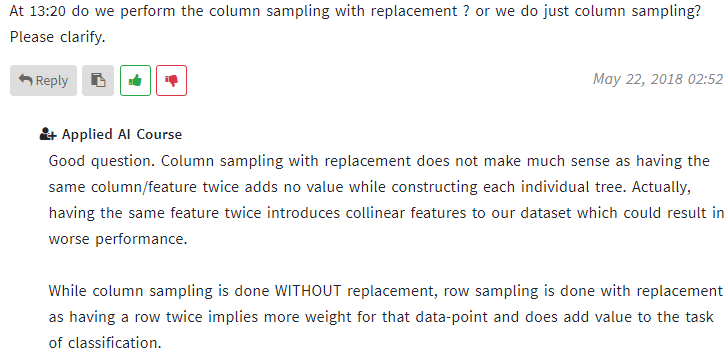


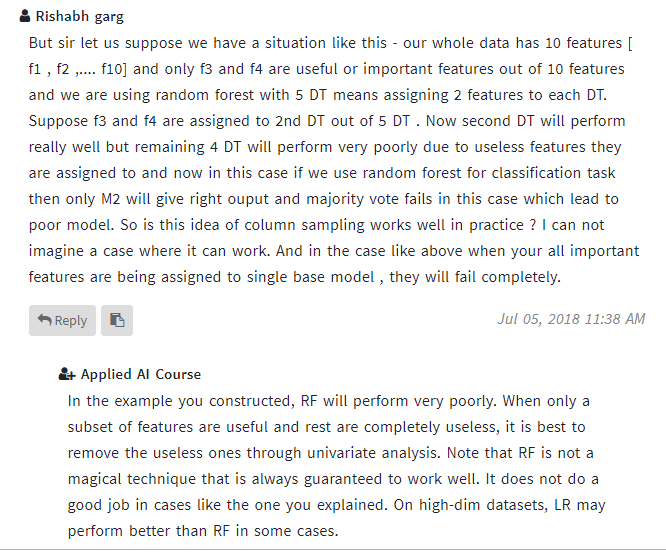


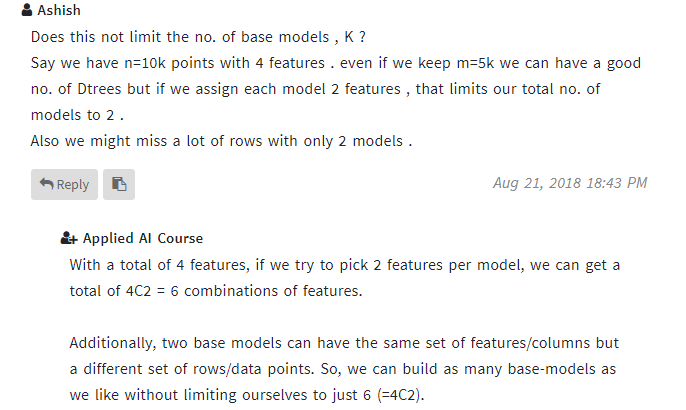
**Why do we do hyperparameter tuning, where in random forest we are allowed to be overfit in model?**



**Comments:**







2)

