**Bias Variance tradeoff:**

There are 3 hyperparameters in Random Forest:

1. Number of base models
2. Column splitting rate
3. Row splitting rate

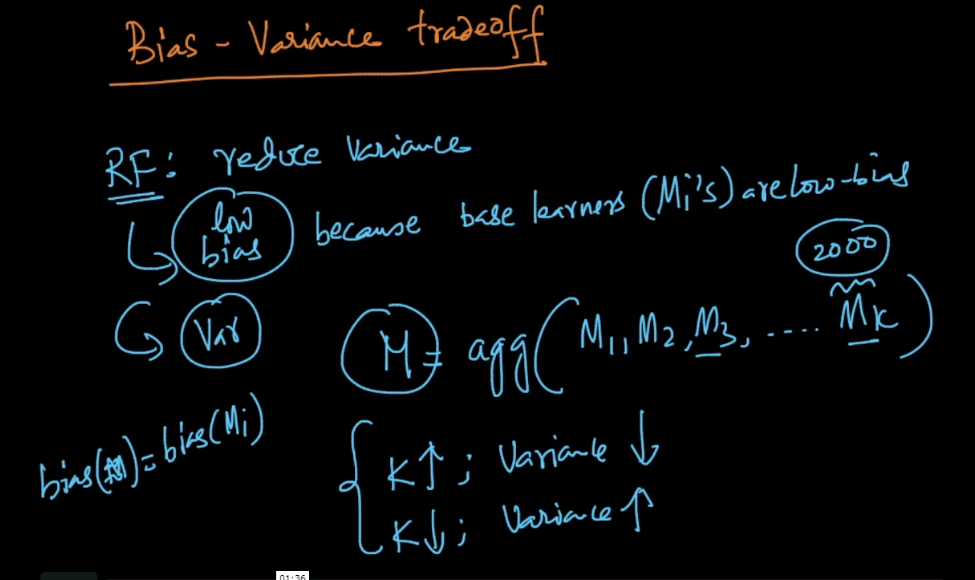
**Number of base models:**

As we know that Random Forest reduces Variances. So typically we pick base models which have low bias and high variance.

Now in RF or ensemble models with bagging, as we increase number of base models ie k, the less will be the variance, because if no of models are more, then very few models will be impacted by change in training dataset.

Hence if k increases variance decreases

If k decreases, variance increases.



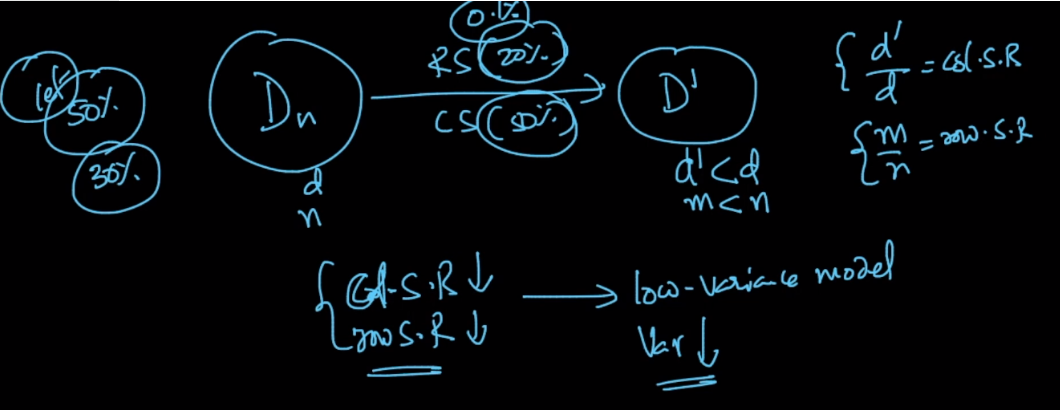
**Column splitting Rate & Row splitting rate:**

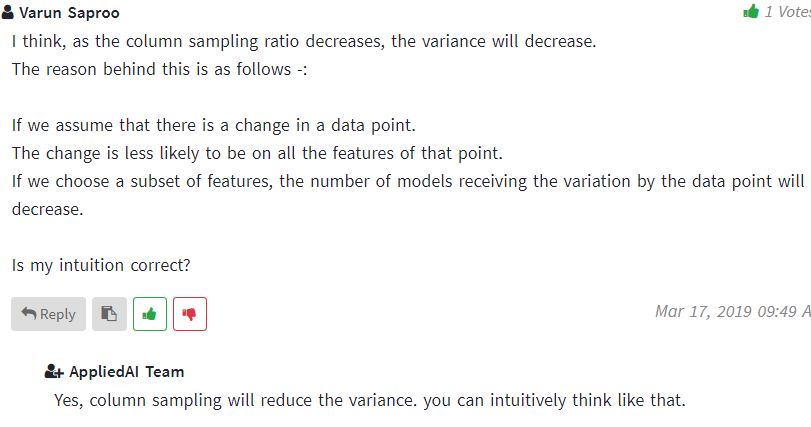
Column splitting rate = sample # of columns / total columns

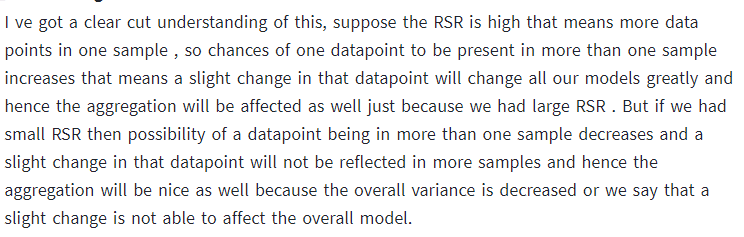
Row splitting rate = sample # or rows / total rows.

Typically if Col S.R or row S.R. decreases, variance will increase. But it doesn’t mean that we will keep it at 0.1 % rate, we keep reasonable value.

We can find it using different values of CSR and RSR.

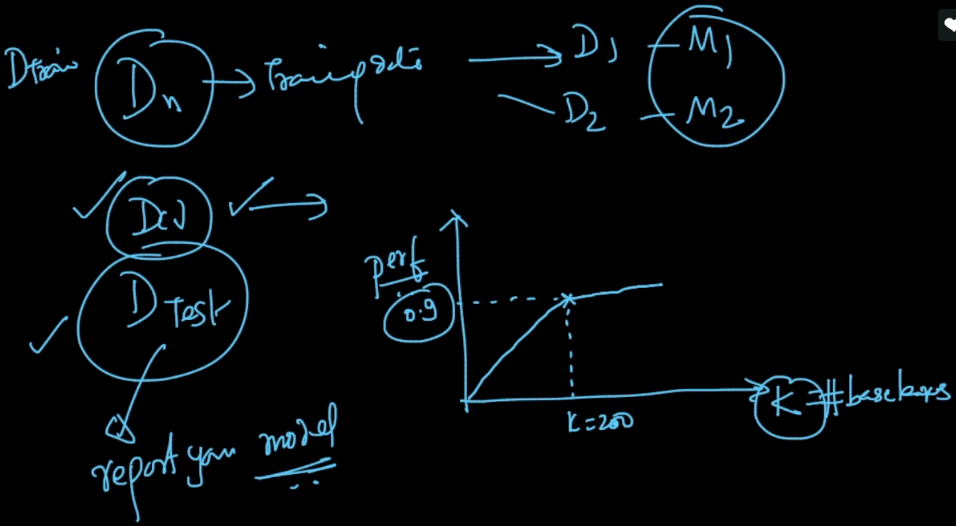






Usually people fix CSR and RSR and try different values of K,

and by doing this, if you plot a graph in which k is on x-axis and performance is on y-axis then you will see that there will be constant performance after certain k value, let’s say after k=200, there is constant performance and then we choose k = 200.



**Comments:**

1)

