**Bagging:** It also called as Bootstrap Aggregation.

What it does is suppose we have **n** data points in training data, then we do sample with replacement, where each sample has **m** data points, where m < n.

**Sample with replacement:** data point can be repeated in samples with small probability.

Now we provide each sample to different models, here each model have different samples, that means all models have almost different data points.

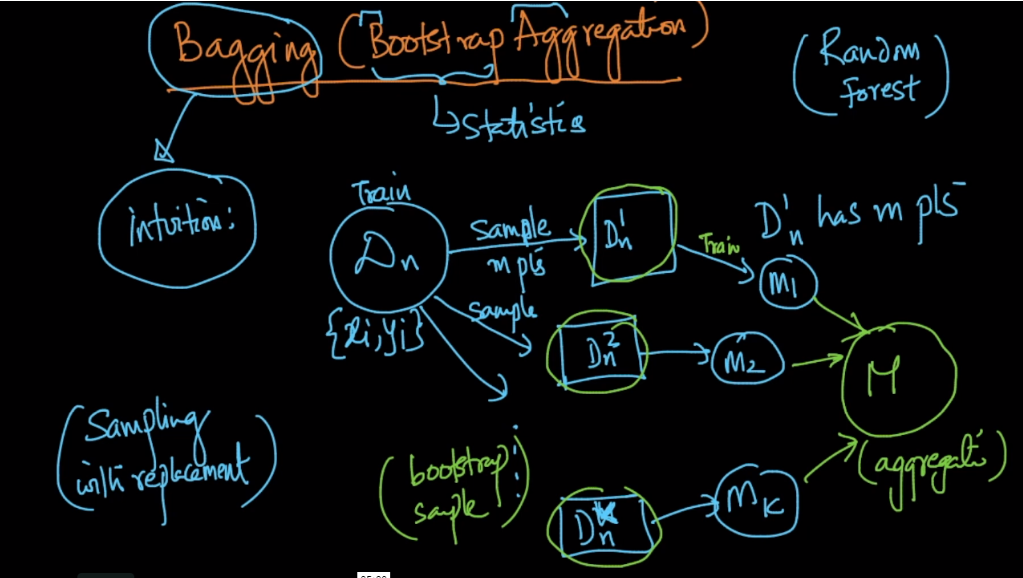
Suppose we trained K models using K samples, then we do aggregation of all the models output as:

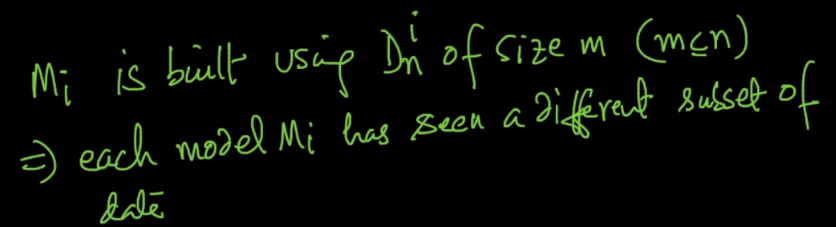
**For classification**:

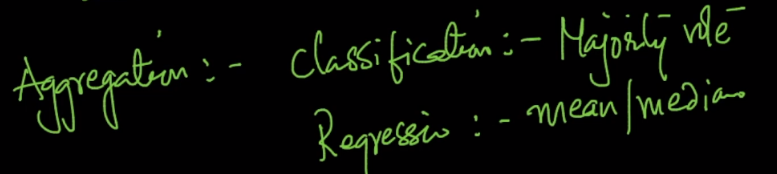
Suppose m1 gives 1, m2 gives 0, …….., mk gives 1. So At aggregation stage we do majority voting, whichever class has highest vote will be chosen as output.

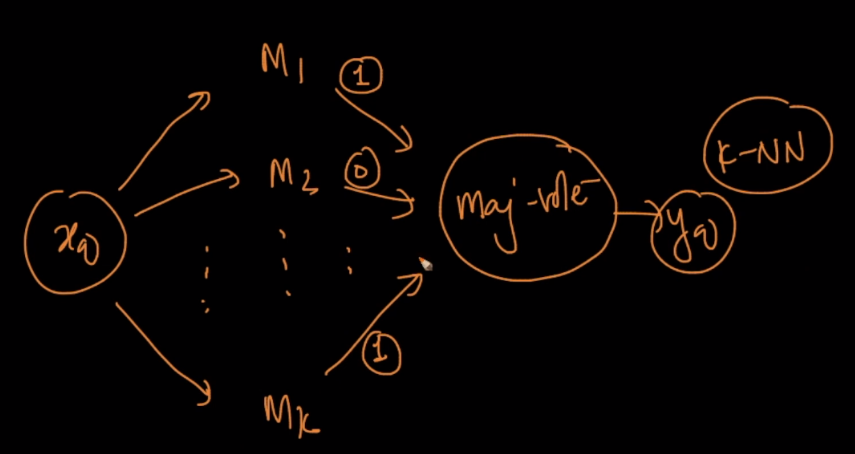
**For Regression:**

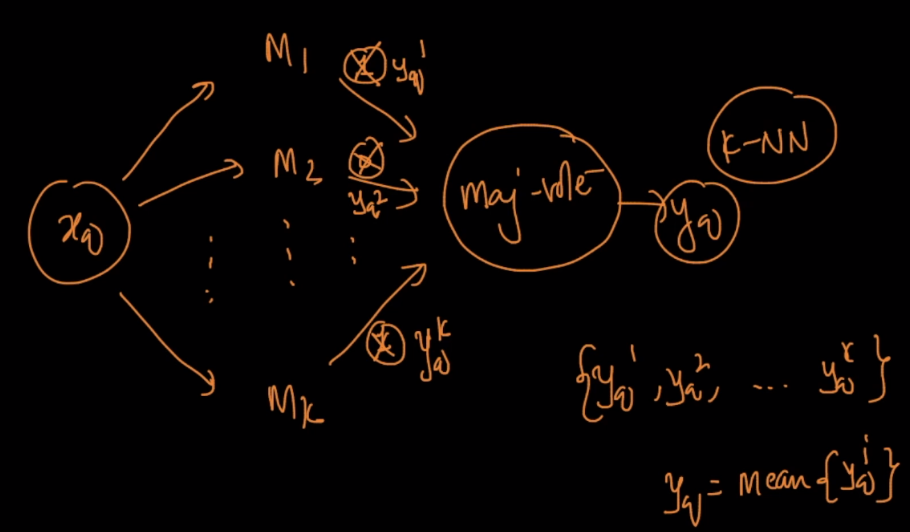
Suppose m1 gives y1, m2 gives y2, ……, mk gives yk. So At aggregation stage we either find mean or median of output generated by all the models, and that value would the end result.









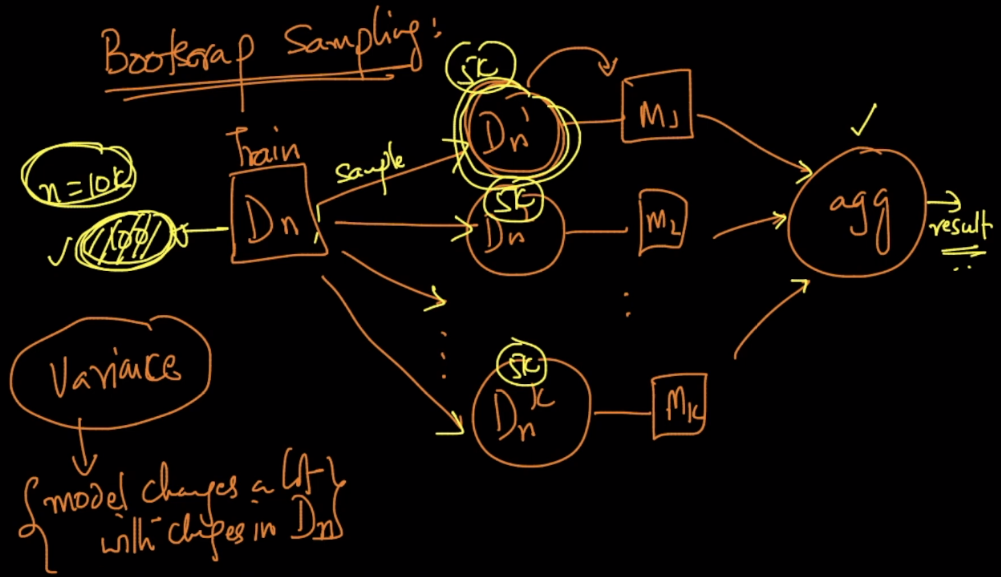


**Why it’s used:**

First recall Variance in model error, variance is how much my model changes if I change training data.

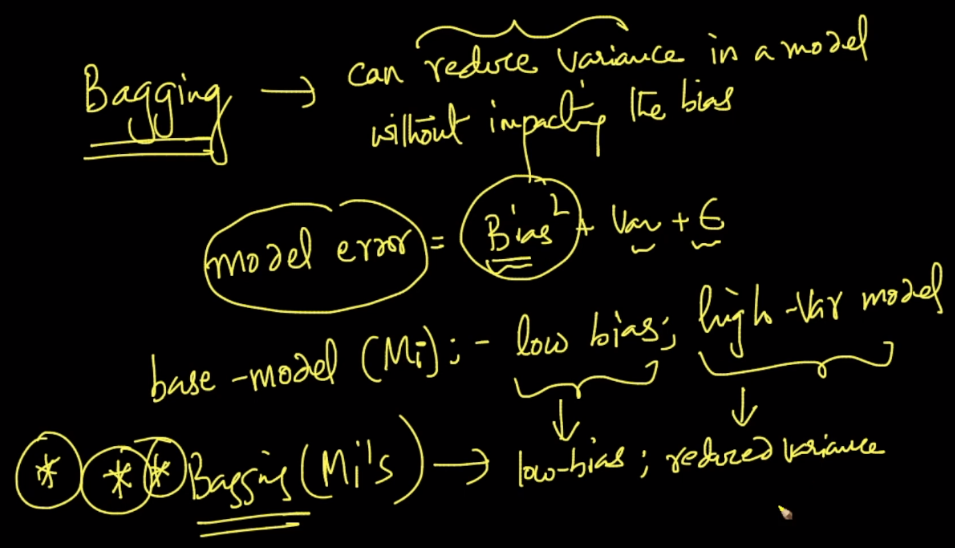
So suppose we have made changes in training data, now these changed data will be available in some of the samples only, and not all the samples.

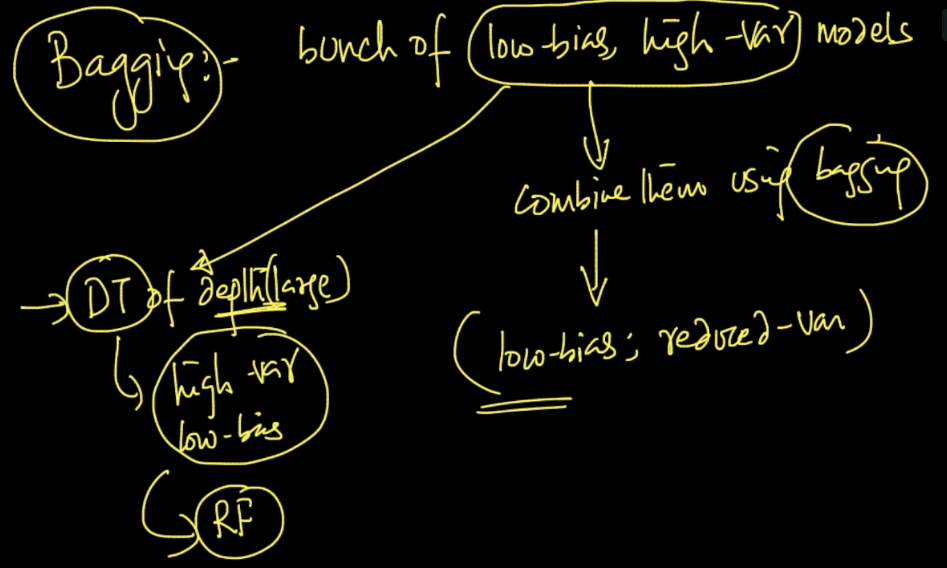
Since changed data is available in some of the samples, so only few models will face this variance problem, and not all, and since in Bagging we aggregate results of all the models so eventually the variance will be averaged or reduced, Note: variance will not be completely eliminated.



So by considering this property of bagging that it reduces variance, we can say that bagging can be used with **low bias and high variance models**, because of this we get very less error, as bias is already less and variance will also become less.

Exqample: DT with large depth have high variance and low bias, so DT with large depth can be used in bagging technique





**Comments:**

