Statistcal Machine Learning

Ensemble Learning

Ensemble Learning

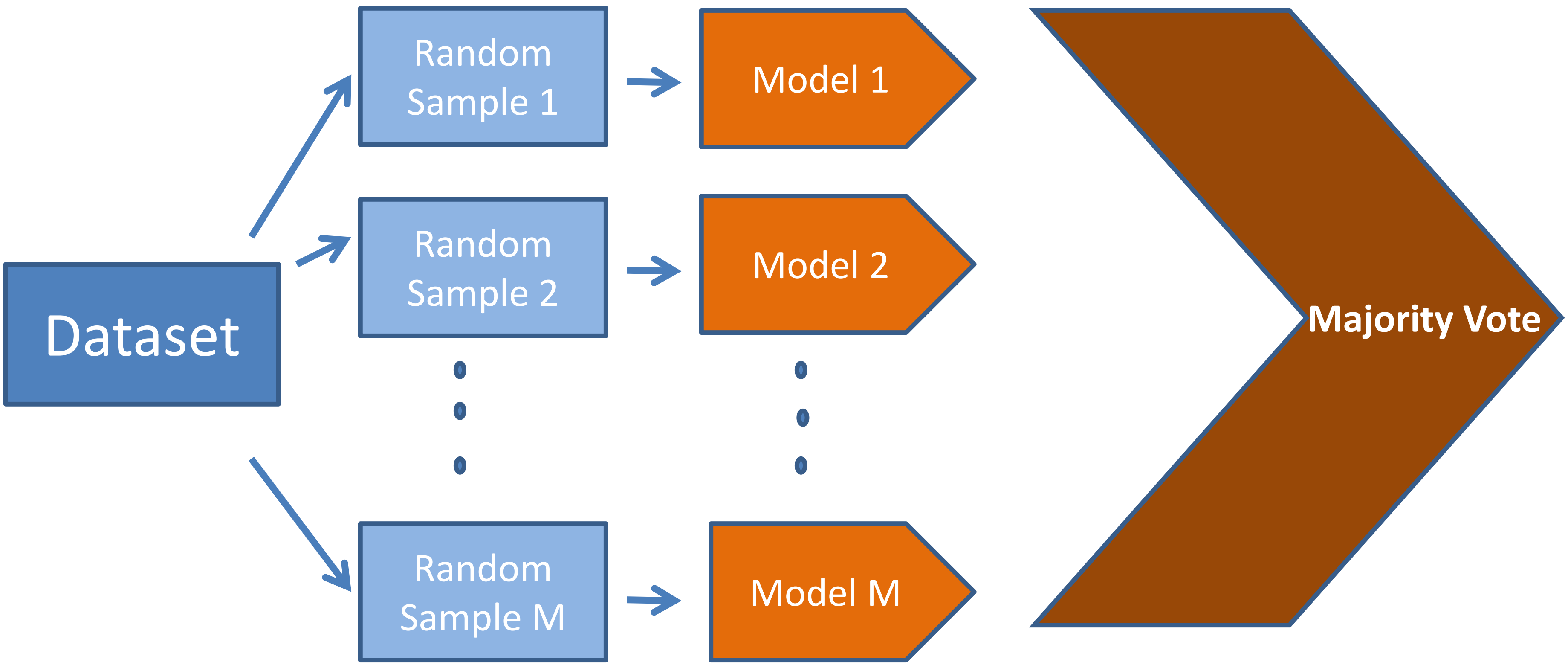
Idea: train several (many) predictive models and combine the results to make a final predictive model.

A few popular methods:

- Bagging
- Boosting
- Stacking
- Random Forests

Bagging

Bootstrap **Agg**regation:



Bagging

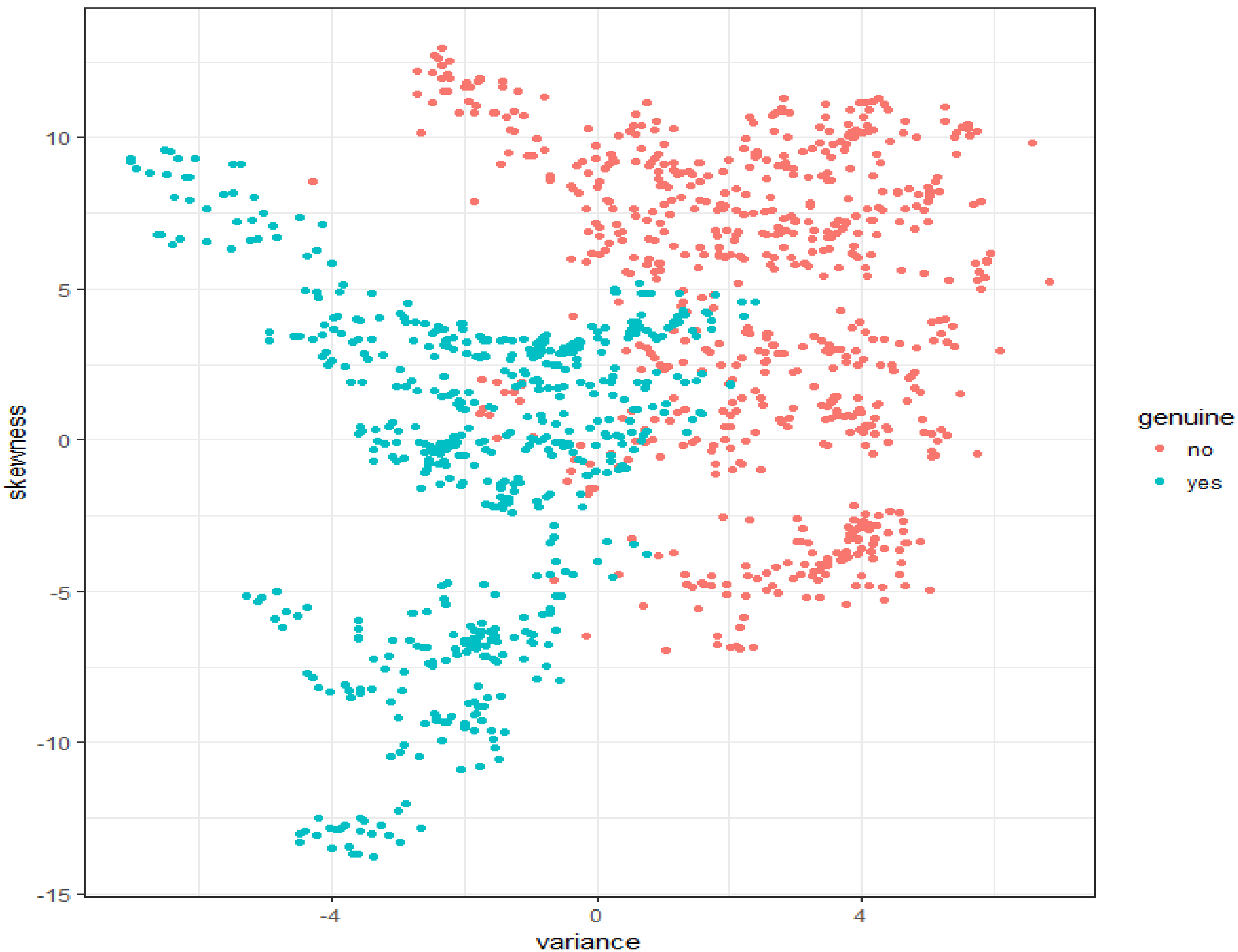
Choose a type of model (e.g. neural net)

1. Obtain a random sample (with replacement) from the training dataset, both of size n
2. Train a model using the random sample

Repeat steps 1-2 M times.

Majority vote/average of the predictions from M models for the final prediction model.

Example: Bank Notes



Training and Test sets

```
> str(ss)
'data.frame':   1235 obs. of  6 variables:
 $ variance: num  4.546 3.866 3.457 0.329 3.591 ...
 $ skewness: num  8.17 -2.64 9.52 -4.46 3.01 ...
 $ kurtosis: num  -2.459 1.924 -4.011 4.572 0.729 ...
 $ entropy : num  -1.462 0.106 -3.594 -0.989 0.564 ...
 $ type    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ genuine : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
> str(tt)
'data.frame':   137 obs. of  6 variables:
 $ variance: num  3.0948 3.0864 3.8999 -1.2424 -0.00129 ...
 $ skewness: num  8.732 -2.584 1.734 -1.718 0.139 ...
 $ kurtosis: num  -2.901 2.231 1.601 -0.526 -0.197 ...
 $ entropy : num  -0.96682 0.30947 0.96765 -0.21036 0.00818 ...
 $ type    : int  0 0 0 1 1 1 1 1 1 0 ...
 $ genuine : Factor w/ 2 levels "no","yes": 1 1 1 2 2 2 2 2 2 1 ...
```

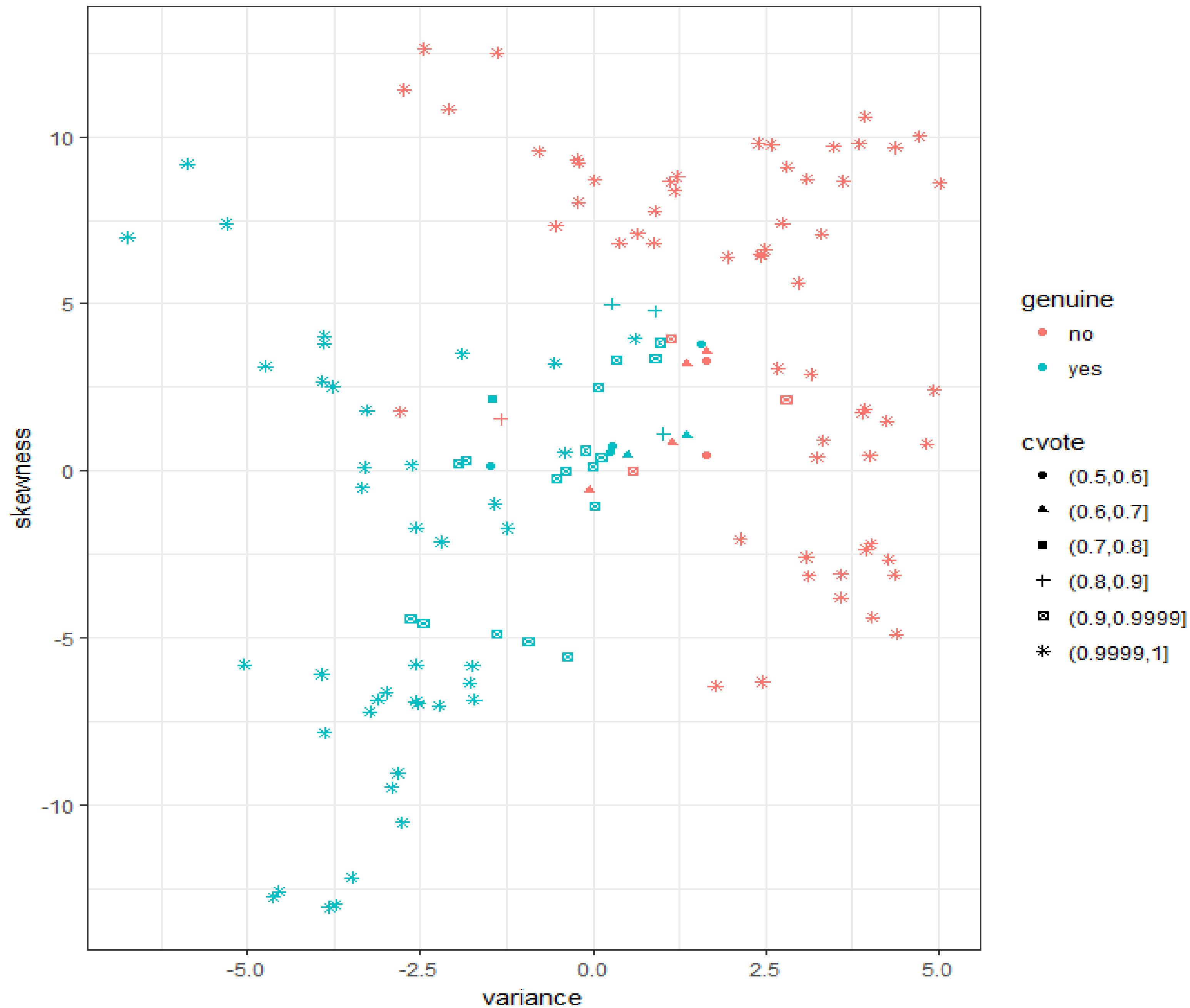
Bagging kNN in R: bnn

```
> library(FNN)
> out <- ownn(ss[,1:2], tt[1:2], cl=ss[,6], testcl=tt[,6])
> out$accuracy
```

```
 knn      ownn      bnn
0.9343066 0.9343066 0.9416058
```

Number of neighbors k is chosen by 5-fold CV

Results of Majority Voting



Bias-Variance Tradeoff & Bagging

Bagging is

- probably worthwhile when a method tends to produce highly variable model fits across different training sets (i.e., high variance).

- probably not worthwhile for low variance methods

Bagging vs. Boosting

M random samples are drawn from original dataset

M models fit separately

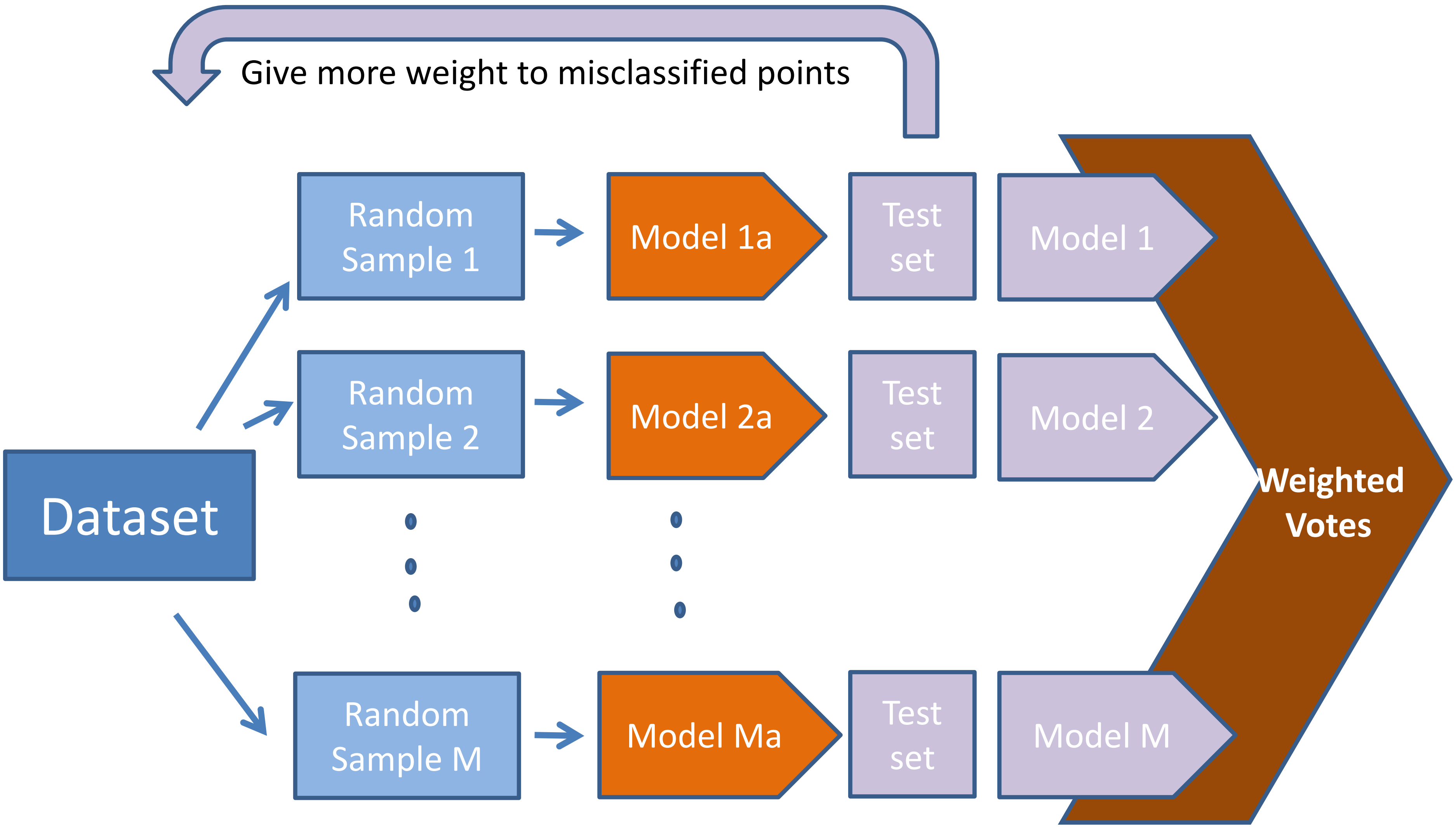
Each model's vote is equally important, i.e. equally weighted.

Initialize: Bagging

Iterate: Draw M random samples, but give points that were misclassified in the previous round more weight (more likely to be in the samples). Fit M models

Each model's vote is weighted by its accuracy to build the final prediction model.

Boosting


Give more weight to misclassified points
Dataset
Random Sample 1
Random Sample 2
Random Sample M
Model 1a
Model 2a
Model Ma
Test set
Test set
Test set
Model 1
Model 2
Model M
Weighted Votes

Boosting

- Might reduce both bias and variance

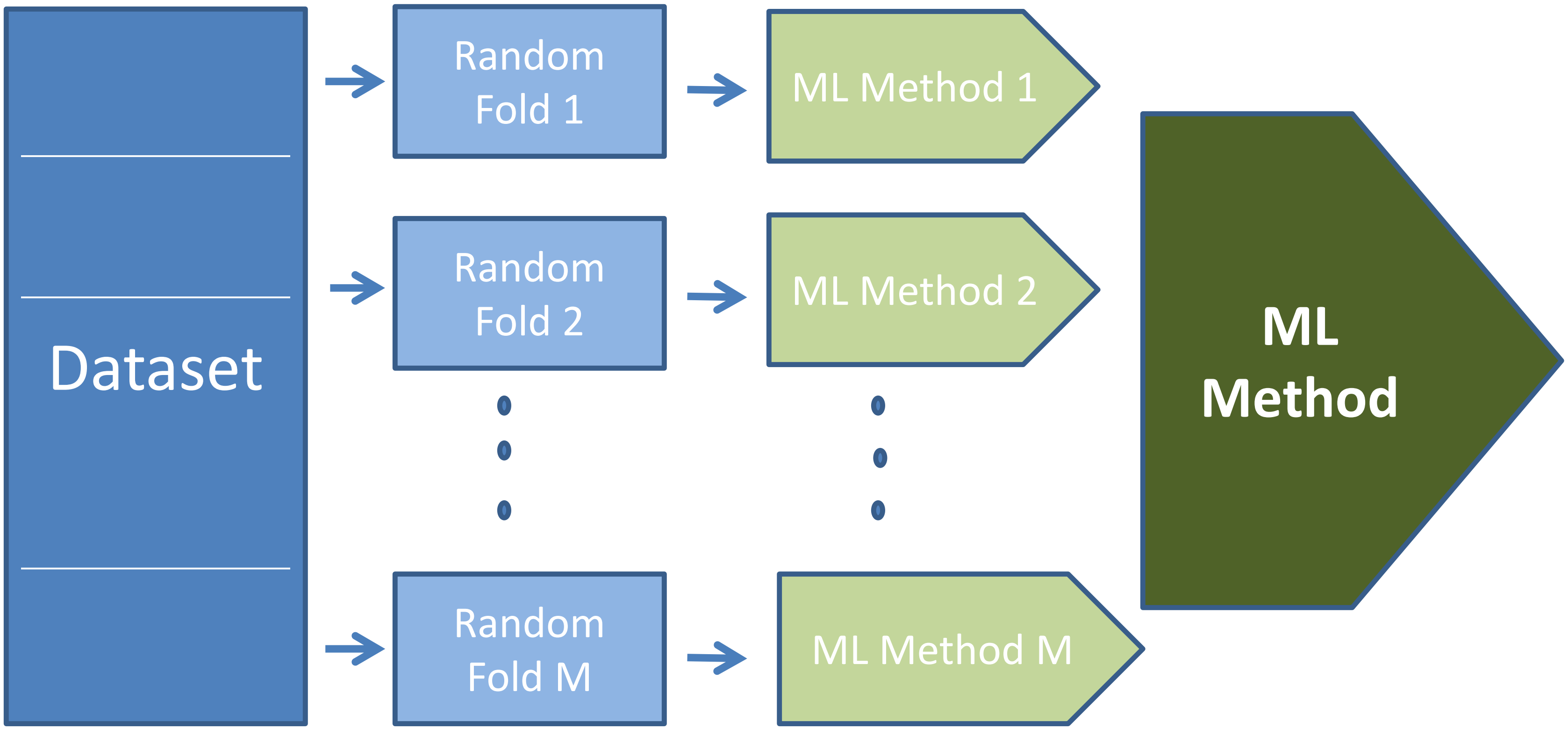
OR

- Might be mislead by a few outliers

Stacking

- Combines different methods of training a model
- Uses the predictions of the models as inputs to a machine learning algorithm to make the final prediction

Stacking



Dataset
Random Fold 1
Random Fold 2
Random Fold M
ML Method 1
ML Method 2
ML Method M
ML Method