Classification Models

- In supervised learning
  1. Logistic Regression
  2. Practical Implementation of Logistic Regression
  3. Support Vector Machine (SVM)
  4. Practical Implementation of SVM
  5. K-Nearest Neighbor (KNN)
  6. A Numerical of KNN
  7. Practical Implementation of KNN

VIT UNIVERSITY
VELLORE INSTITUTE OF TECHNOLOGY

VIT®
BHOPAL
www.vitbhopal.ac.in

Ensemble Learning

(An Introduction of Ensemble Learning)

Dr. Virendra Singh Kushwah

Assistant Professor Grade-II

School of Computing Science and Engineering

Virendra.Kushwah@vitbhopal.ac.in

7415869616

Ensemble Methods in Machine Learning

VIT UNIVERSITY
VELLORE INSTITUTE OF TECHNOLOGY
VIT BHOPAL
www.vitbhopal.ac.in

- Ensemble = “a group of items viewed as a whole rather than individually”

- Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. To better understand this definition lets take a step back into ultimate goal of machine learning and model building.

- Let's understand the concept of ensemble learning with an example. Suppose you are a movie director and you have created a short movie on a very important and interesting topic. Now, you want to take preliminary feedback (ratings) on the movie before making it public. What are the possible ways by which you can do that?

- A: You may ask one of your friends to rate the movie for you.

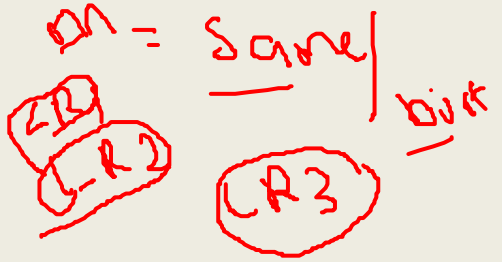
- Now it's entirely possible that the person you have chosen loves you very much and doesn't want to break your heart by providing a 1-star rating to the horrible work you have created.

- B: Another way could be by asking 5 colleagues of yours to rate the movie.

- This should provide a better idea of the movie. This method may provide honest ratings for your movie. But a problem still exists. These 5 people may not be “Subject Matter Experts” on the topic of your movie. Sure, they might understand the cinematography, the shots, or the audio, but at the same time may not be the best judges of dark humour.

- C: How about asking 50 people to rate the movie?

- Some of which can be your friends, some of them can be your colleagues and some may even be total strangers.

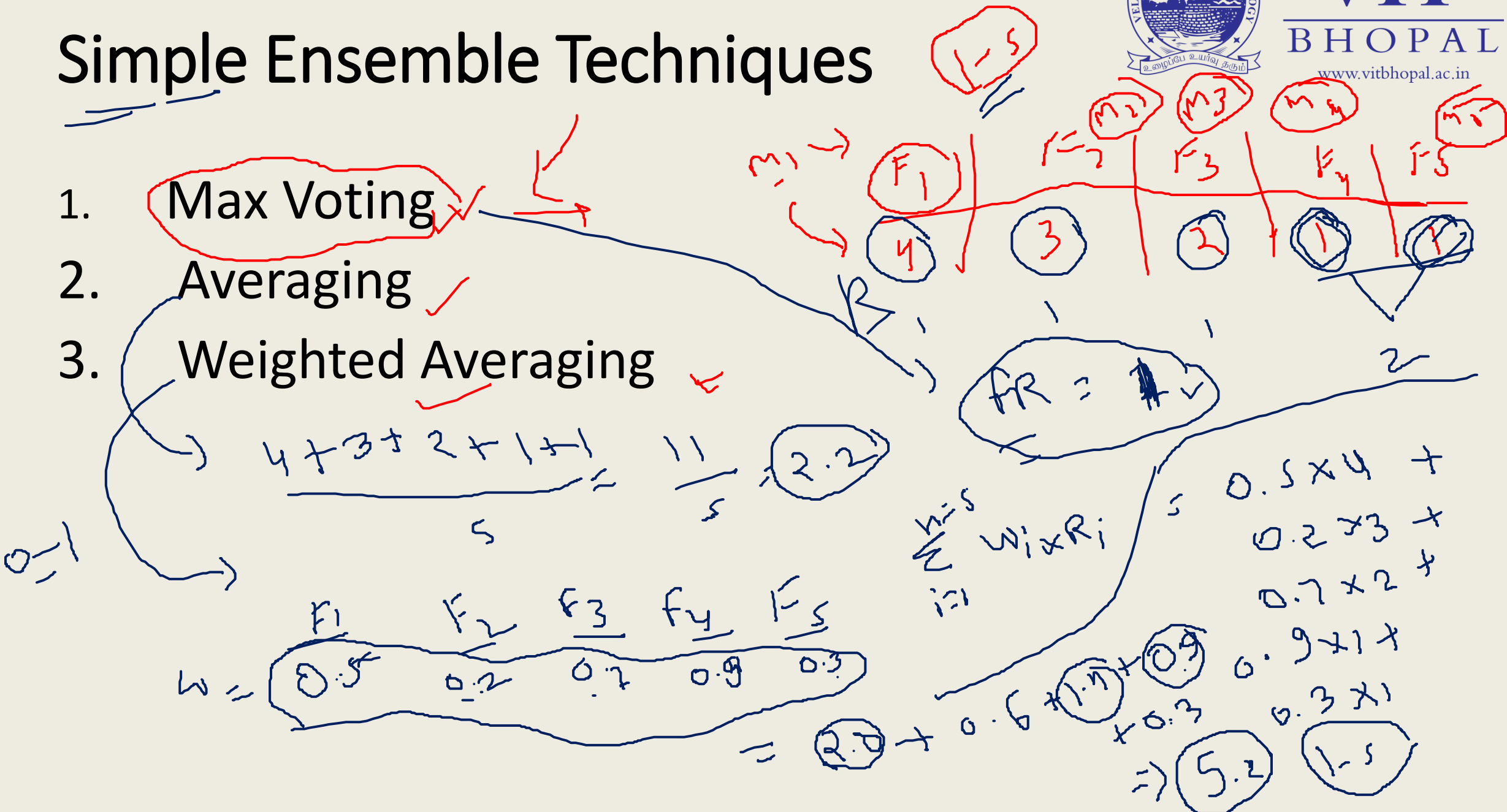


- The responses, in this case, would be more generalized and diversified since now you have people with different sets of skills. And as it turns out – this is a better approach to get honest ratings than the previous cases we saw.

- With these examples, you can infer that a diverse group of people are likely to make better decisions as compared to individuals. Similar is true for a diverse set of models in comparison to single models. **This diversification in Machine Learning is achieved by a technique called Ensemble Learning.**

Simple Ensemble Techniques

1. Max Voting
2. Averaging
3. Weighted Averaging



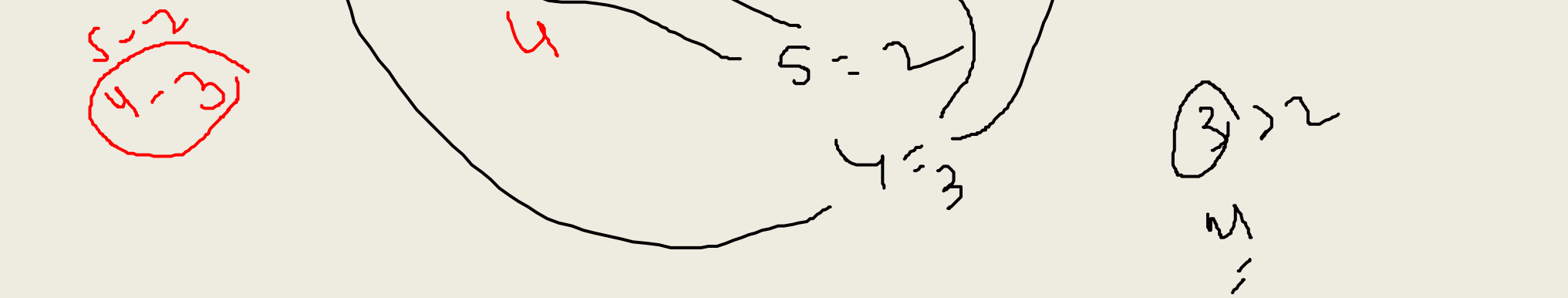
Max Voting

- The max voting method is generally used for classification problems. In this technique, multiple models are used to make predictions for each data point. The predictions by each model are considered as a ‘vote’. The predictions which we get from the majority of the models are used as the final prediction.

- For example, when you asked 5 of your colleagues to rate your movie (out of 5); we’ll assume three of them rated it as 4 while two of them gave it a 5. Since the majority gave a rating of 4, the final rating will be taken as 4. You can consider this as taking the mode of all the predictions.

The result of max voting would be something like this:

Colleague 1	Colleague 2	Colleague 3	Colleague 4	Colleague 5	Final rating
5	4	5	4	4	4



Averaging

- Similar to the max voting technique, multiple predictions are made for each data point in averaging. In this method, we take an average of predictions from all the models and use it to make the final prediction. Averaging can be used for making predictions in regression problems or while calculating probabilities for classification problems.

- For example, in the below case, the averaging method would take the average of all the values.

- i.e. (5+4+5+4+4)/5 = 4.4

VIT UNIVERSITY
VELLORE INSTITUTE OF TECHNOLOGY
VIT®
BHOPAL
www.vitbhopal.ac.in

Colleague 1	Colleague 2	Colleague 3	Colleague 4	Colleague 5	Final rating
5	4	5	4	4	4.4

Weighted Average

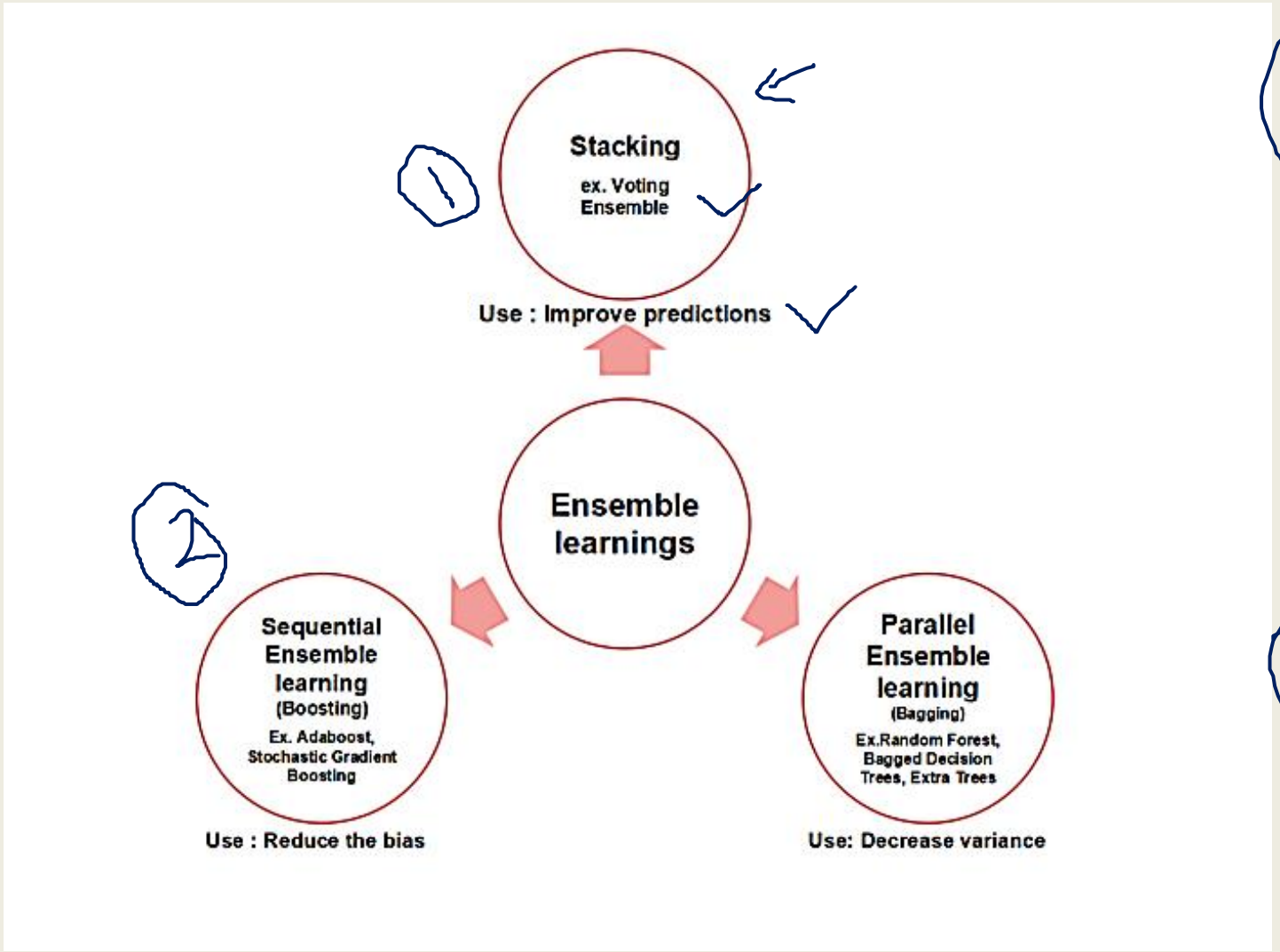
- This is an extension of the averaging method. All models are assigned different weights defining the importance of each model for prediction. For instance, if two of your colleagues are critics, while others have no prior experience in this field, then the answers by these two friends are given more importance as compared to the other people.

- The result is calculated as [(5*0.23) + (4*0.23) + (5*0.18) + (4*0.18) + (4*0.18)] = 4.41.

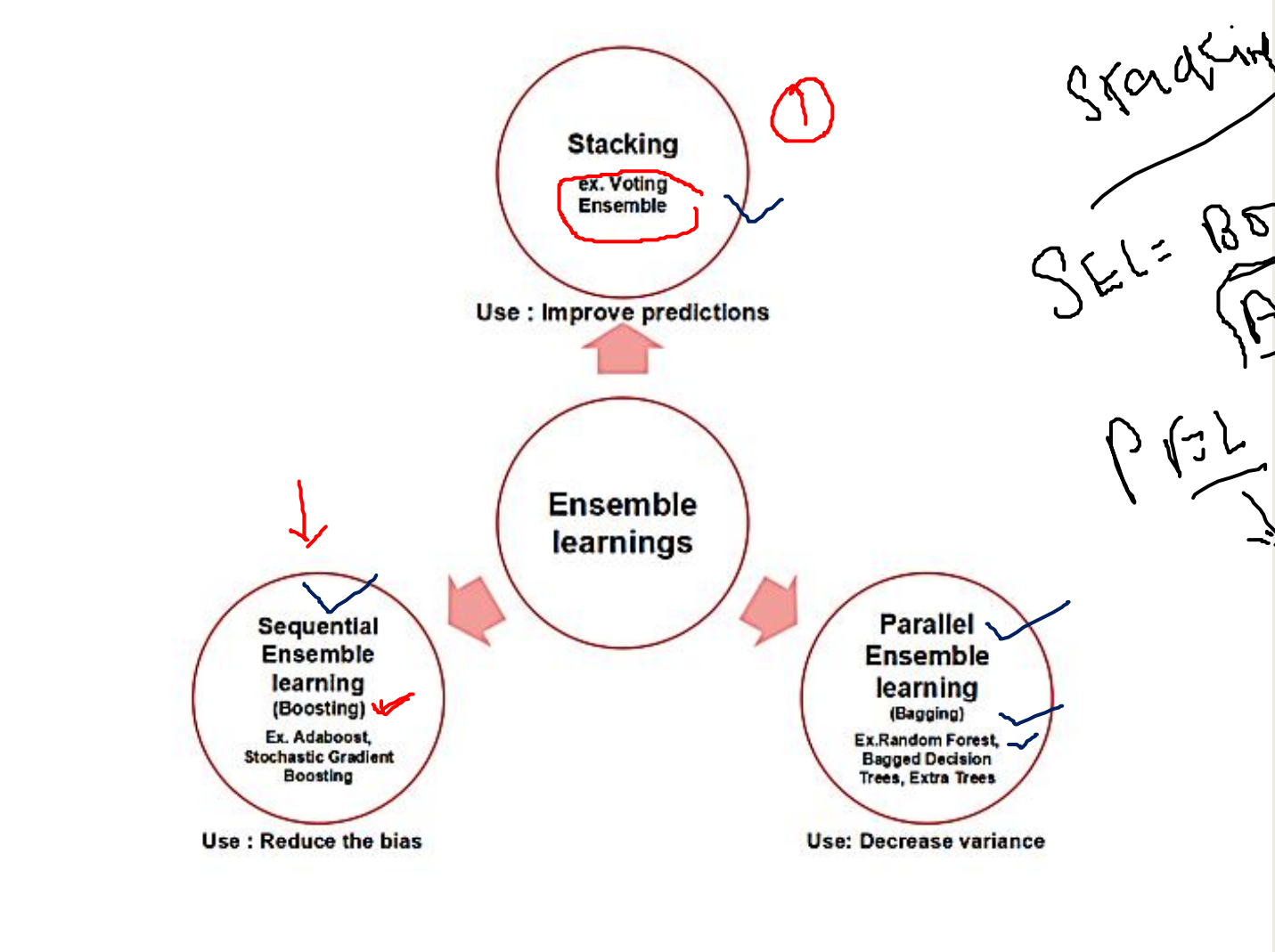
$$= \sum_{i=1} w_i x R_i$$

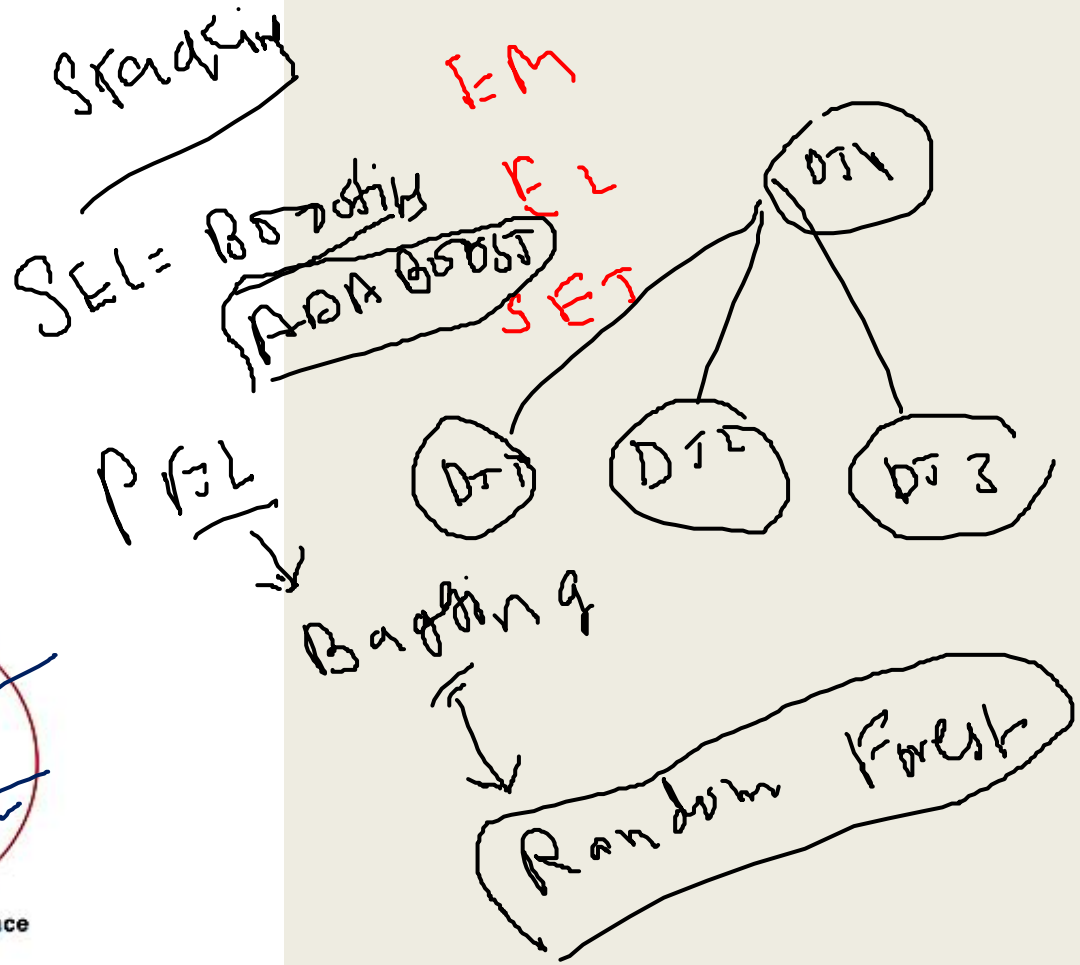
Colleague 1	Colleague 2	Colleague 3	Colleague 4	Colleague 5	Final rating	
weight	0.23	0.23	0.18	0.18	0.18	
rating	5	4	5	4	4	4.41

Ensemble learning Types



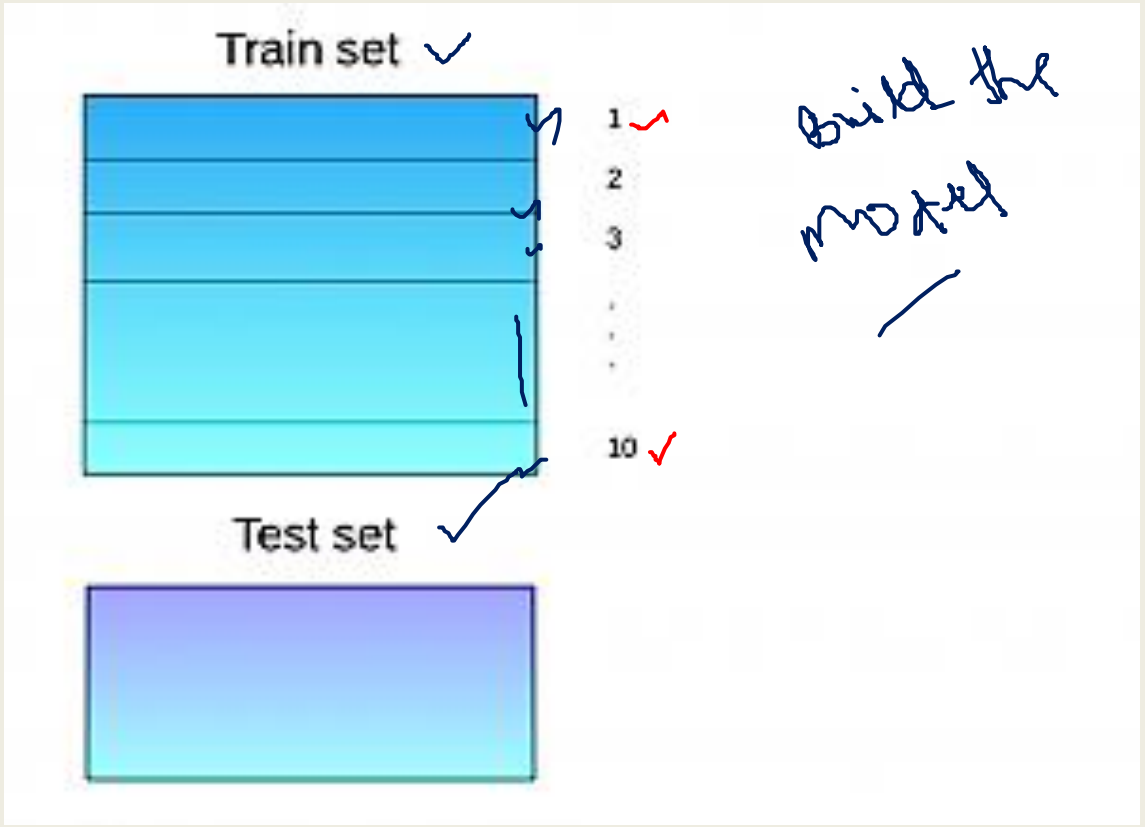
Ensemble learning Types





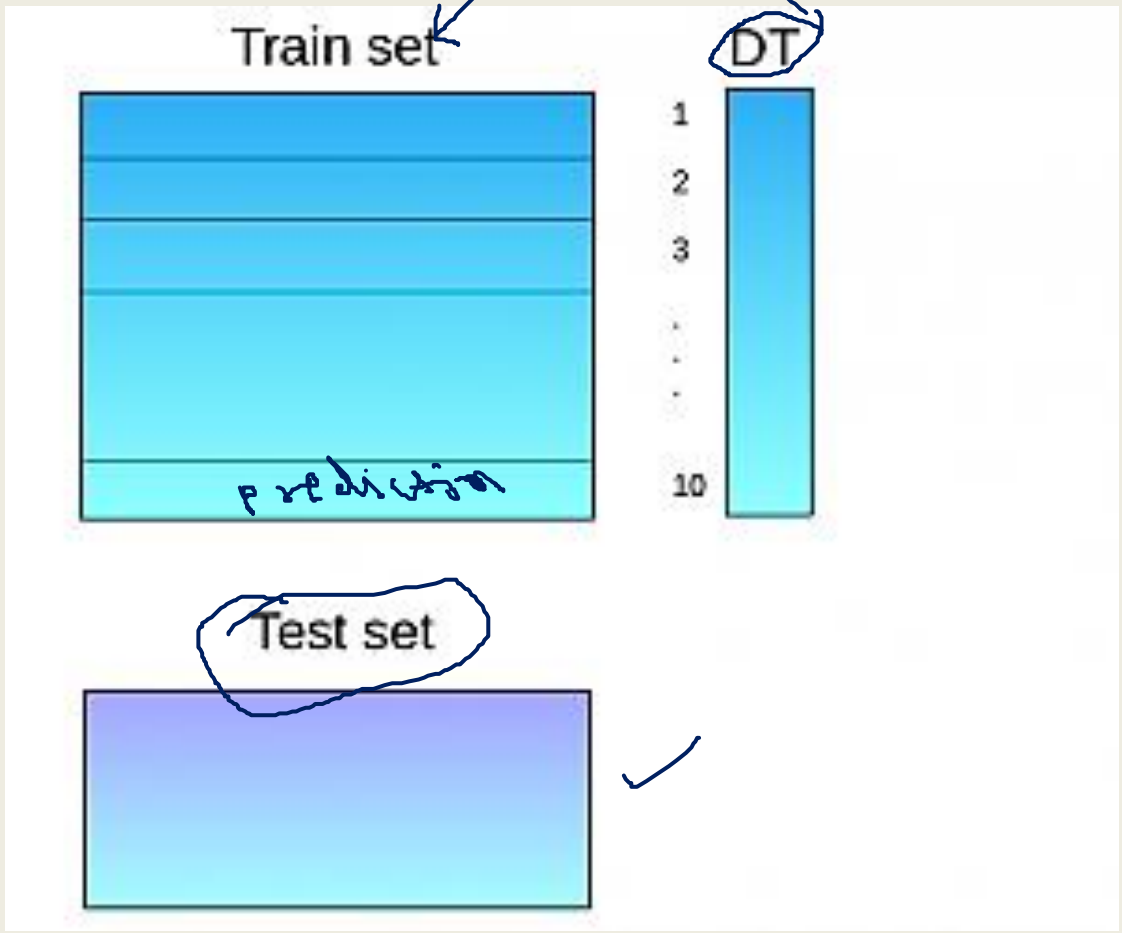
Stacking

- Stacking is an ensemble learning technique that uses predictions from multiple models (for example decision tree, knn or svm) to build a new model. This model is used for making predictions on the test set. Below is a step-wise explanation for a simple stacked ensemble:

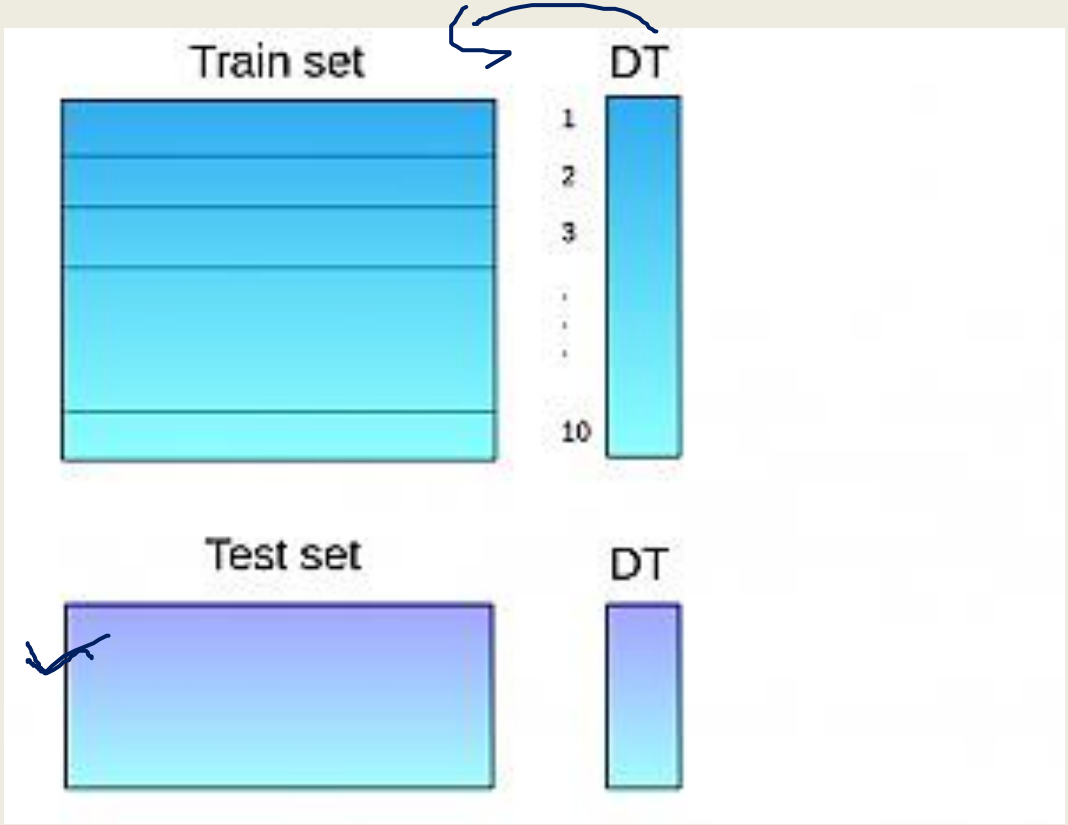


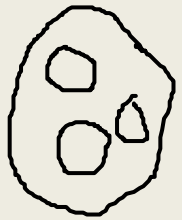
- The train set is split into 10 parts.

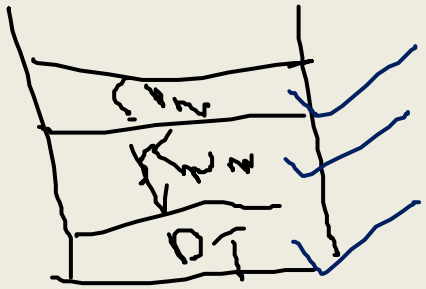
- A base model (suppose a decision tree) is fitted on 9 parts and predictions are made for the 10th part. This is done for each part of the train set.



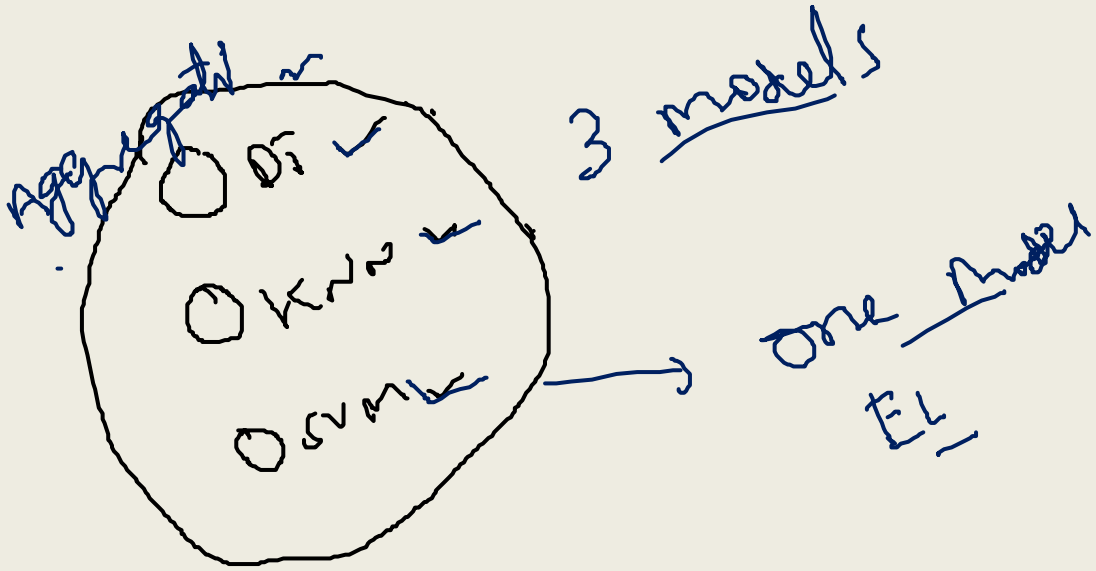
- The base model (in this case, decision tree) is then fitted on the whole train dataset.
- Using this model, predictions are made on the test set.

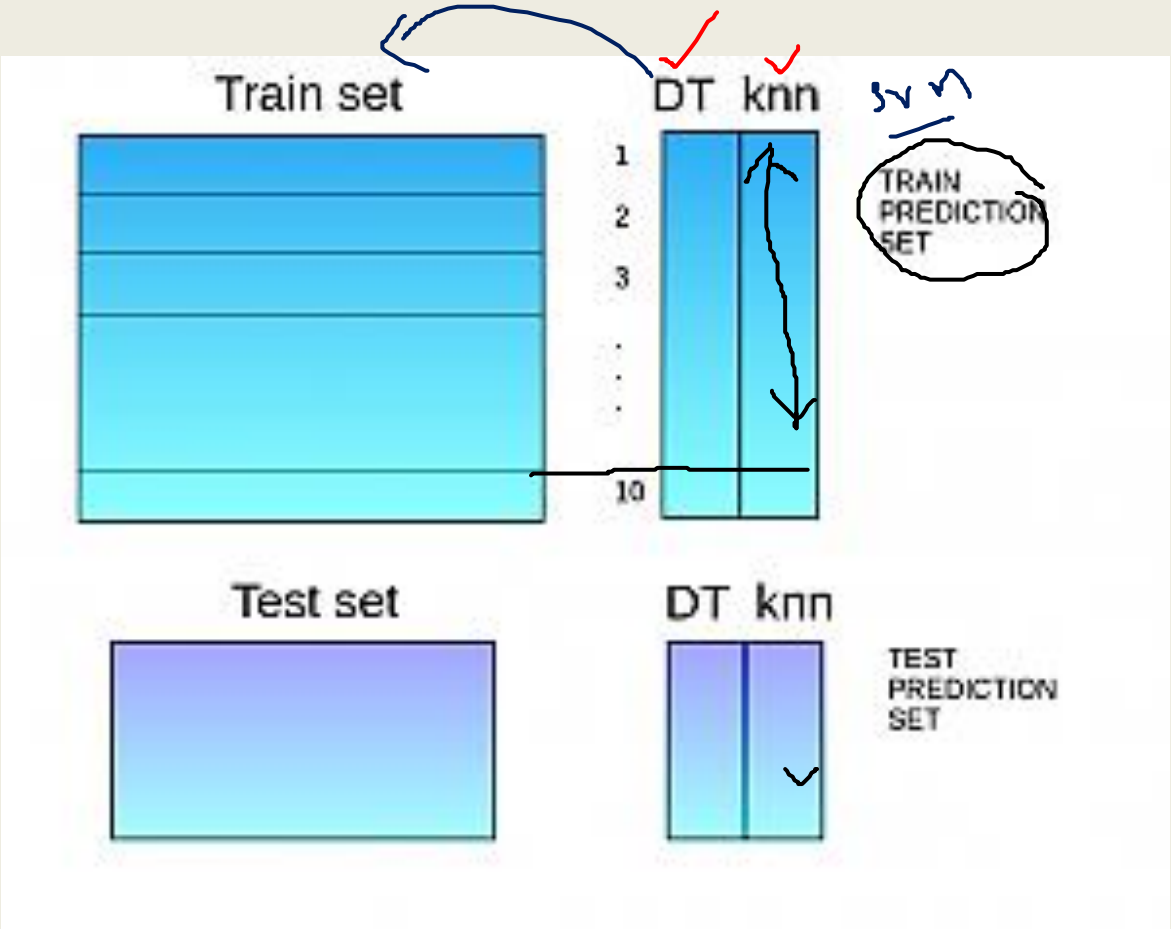


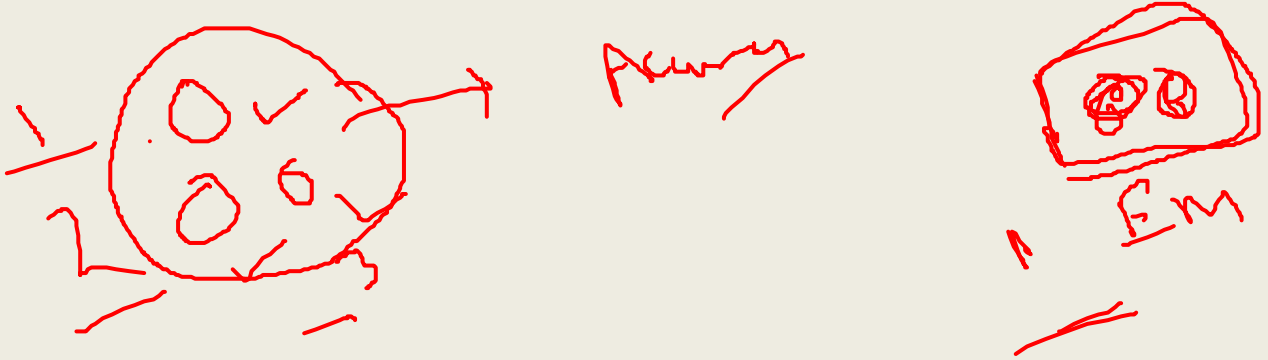




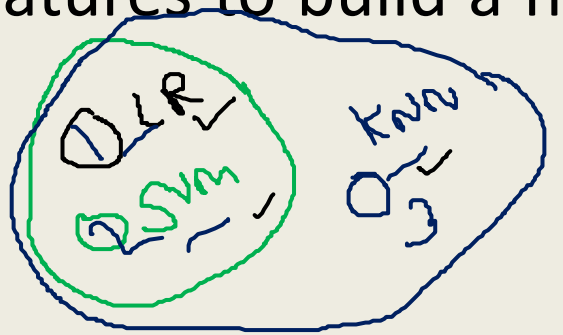
- Steps 2 to 4 are repeated for another base model (say knn) resulting in another set of predictions for the train set and test set.



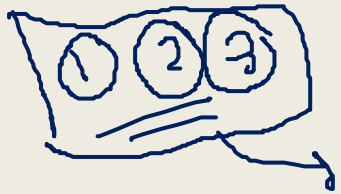




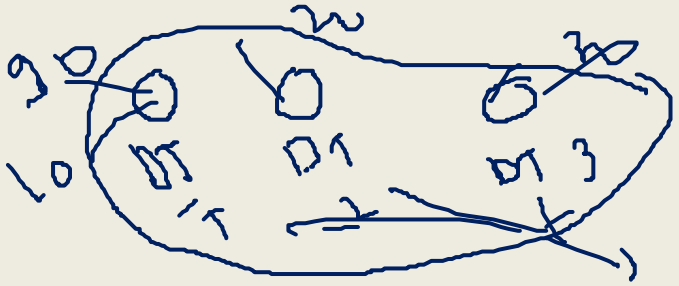
- The predictions from the train set are used as features to build a new model.

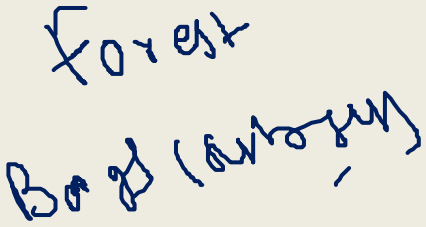


- This model is used to make final predictions on the test prediction set.



Bagging



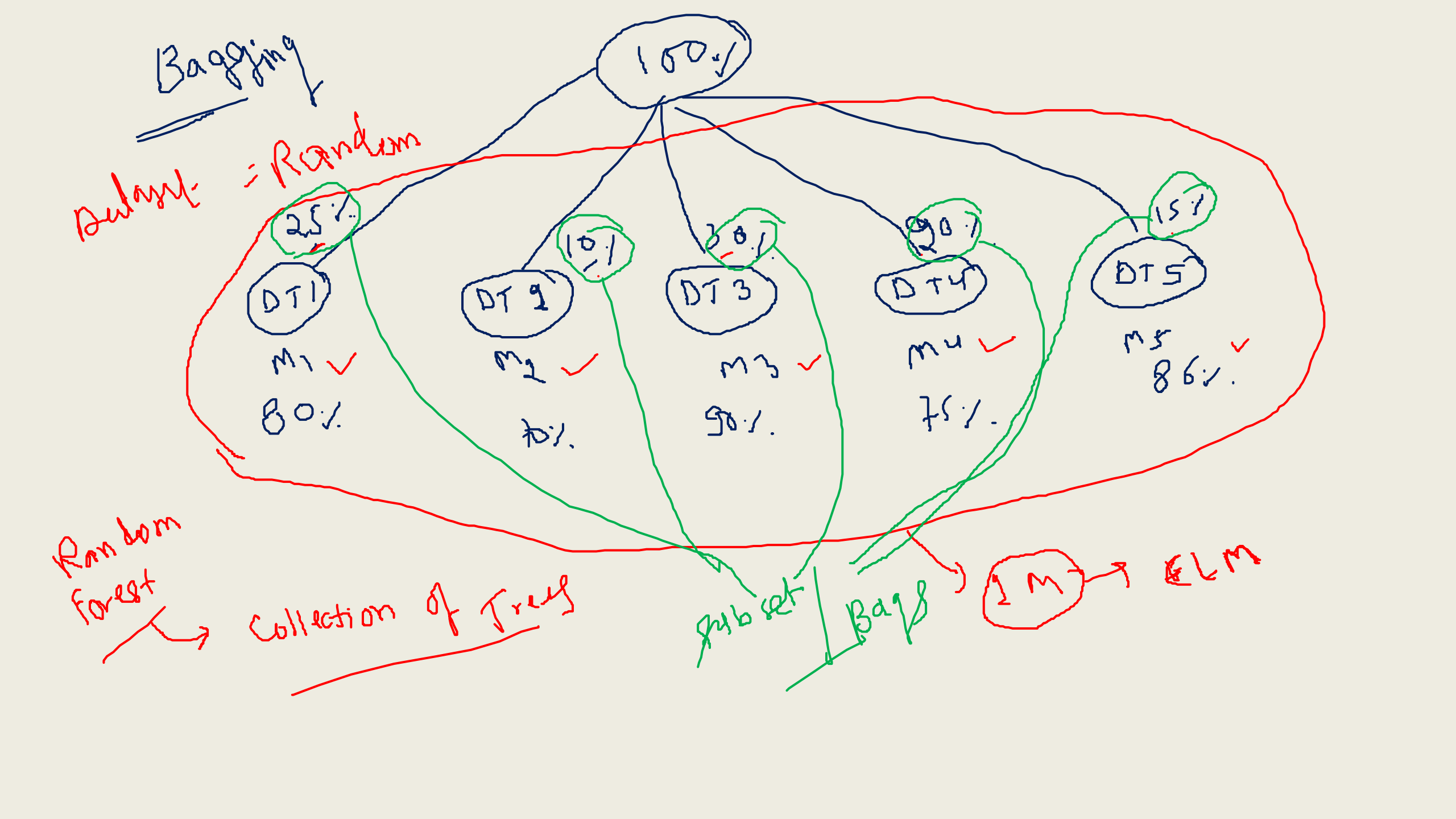


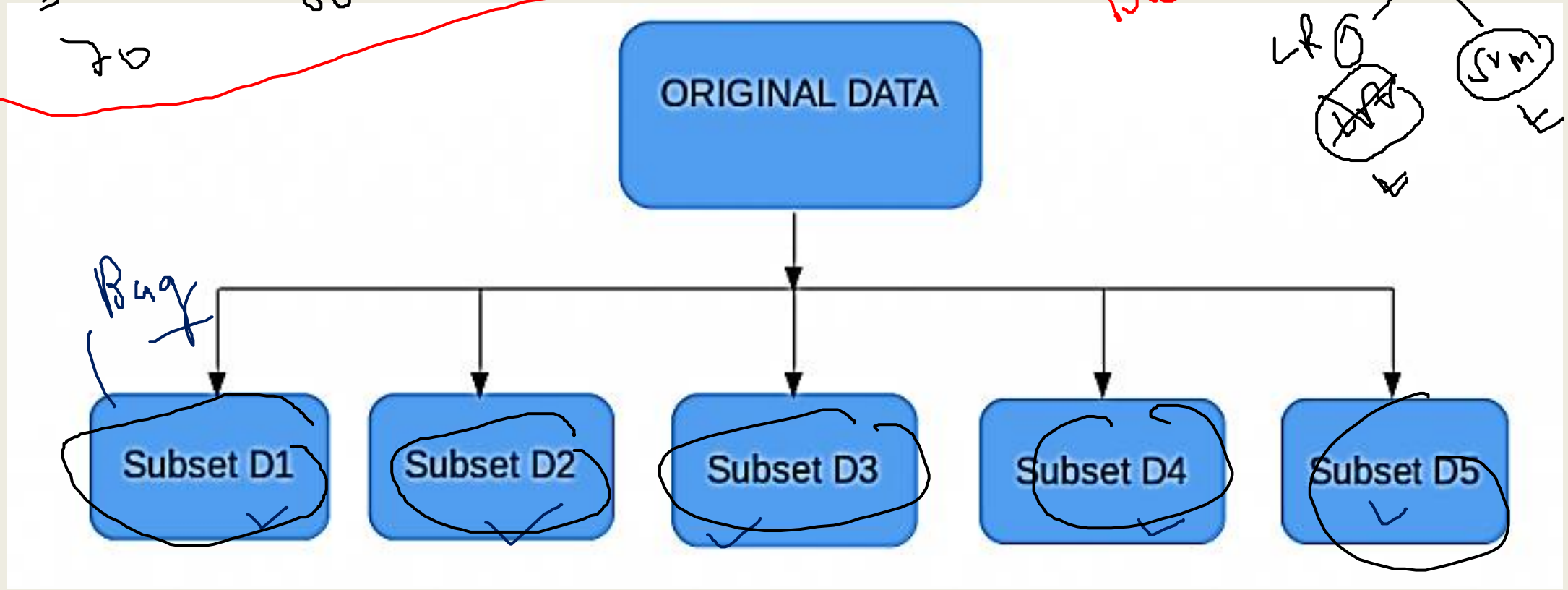
VIT UNIVERSITY
VELLORE INSTITUTE OF TECHNOLOGY
VIT BHOPAL
www.vitbhopal.ac.in

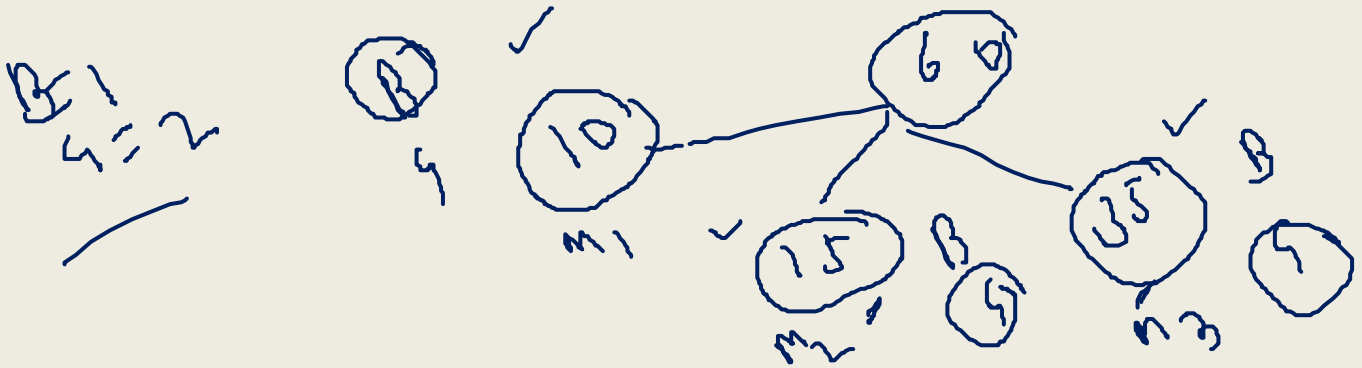
- The idea behind bagging is combining the results of multiple models (for instance, all decision trees) to get a generalized result. Here's a question: If you create all the models on the same set of data and combine it, will it be useful? There is a high chance that these models will give the same result since they are getting the same input. So how can we solve this problem? One of the techniques is bootstrapping.

- Bootstrapping is a sampling technique in which we create subsets of observations from the original dataset, with replacement. The size of the subsets is the same as the size of the original set.

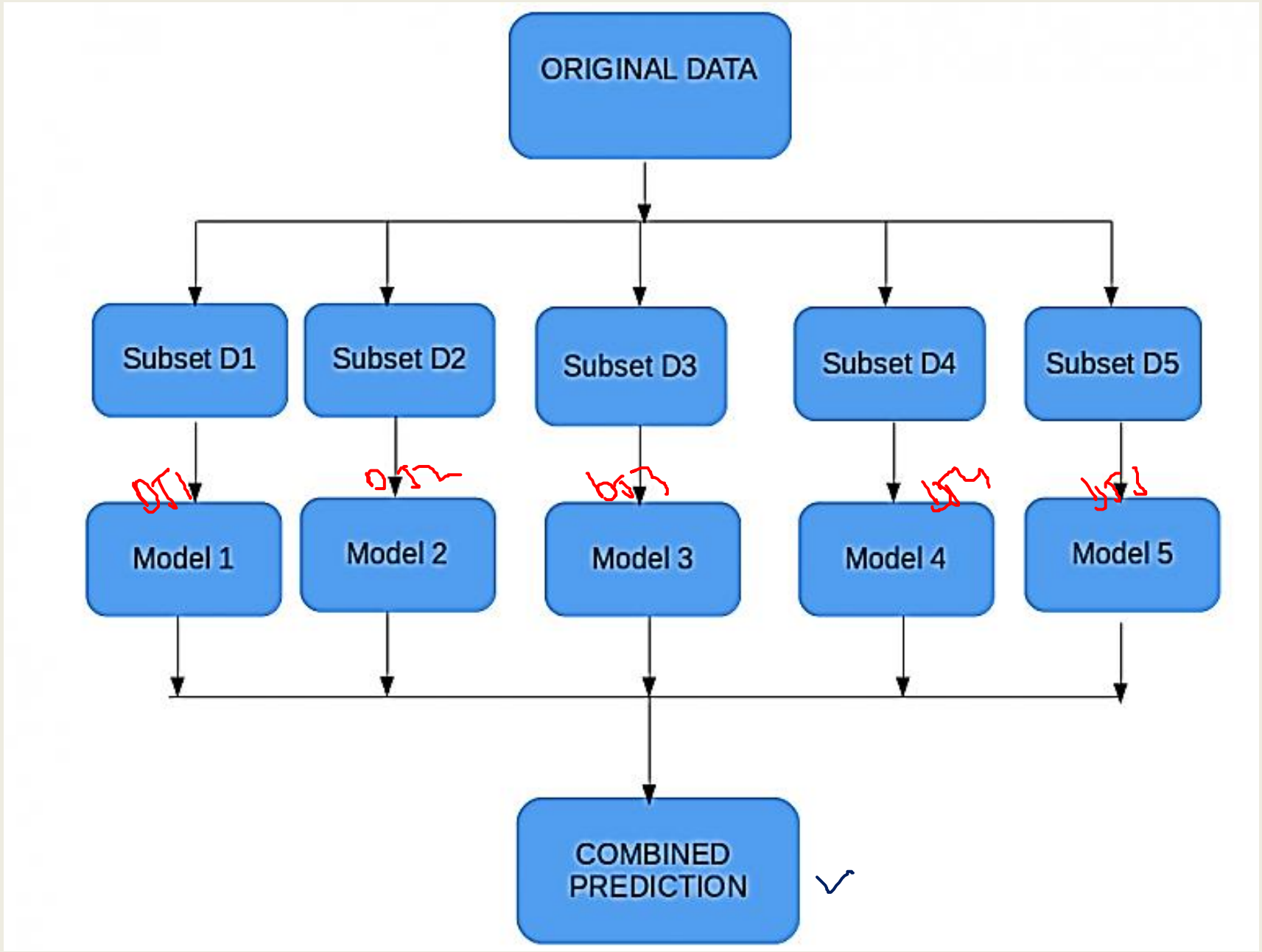
- **Bagging (or Bootstrap Aggregating) technique uses these subsets (bags) to get a fair idea of the distribution (complete set). The size of subsets created for bagging may be less than the original set.**

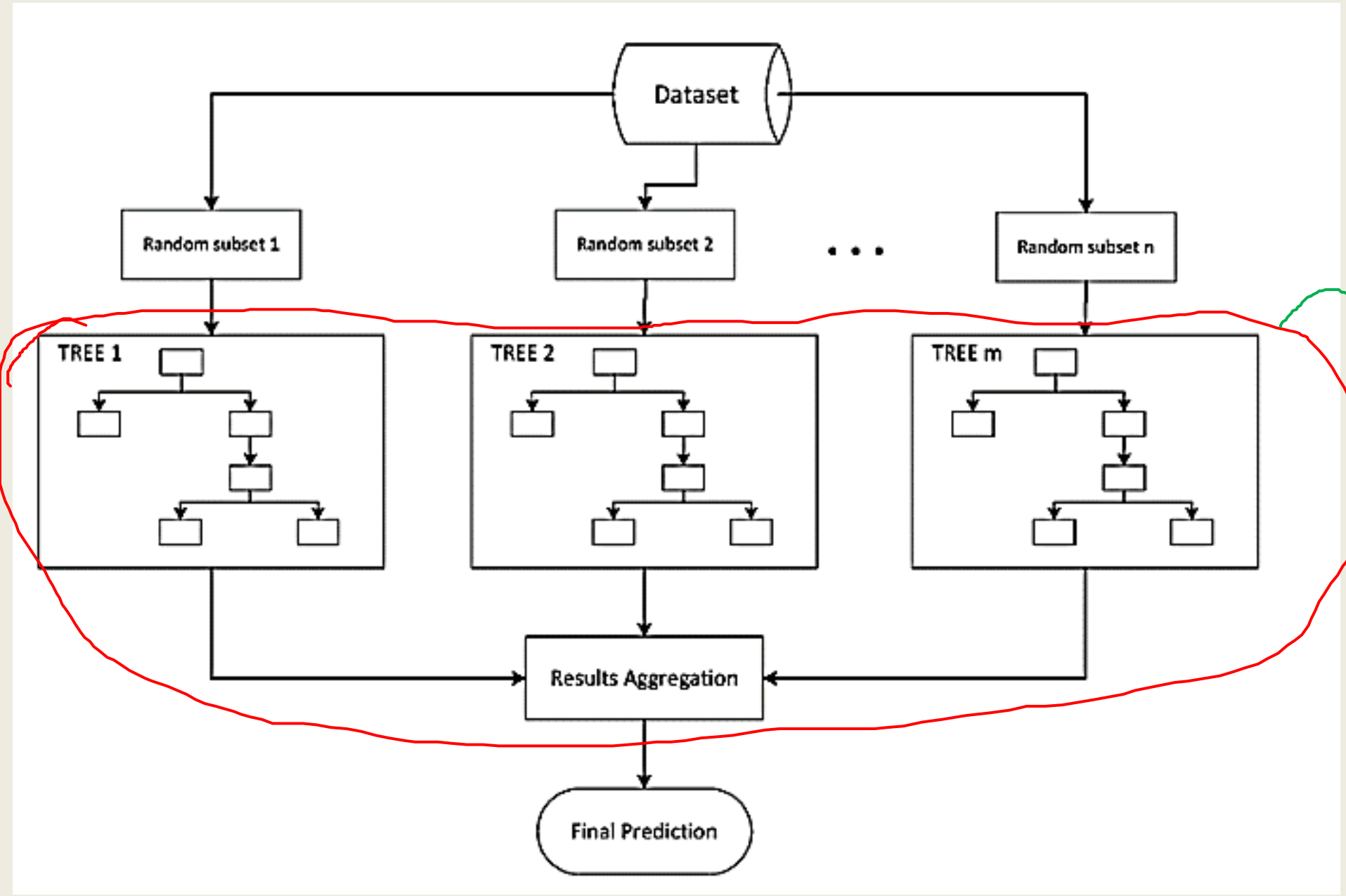

Bagging
Random
100%
25%
10%
20%
90%
15%
DT1
DT 2
DT 3
DT 4
DT 5
M1
80%
M2
70%
M3
90%
M4
75%
M5
86%
Random forest
Collection of Trees
subset
Bags
M
ELM


ORIGINAL DATA
Subset D1
Subset D2
Subset D3
Subset D4
Subset D5



- Multiple subsets are created from the original dataset, selecting observations with replacement.
- A base model (weak model) is created on each of these subsets.
- The models run in parallel and are independent of each other.
- The final predictions are determined by combining the predictions from all the models.


ORIGINAL DATA
Subset D1
Subset D2
Subset D3
Subset D4
Subset D5
Model 1
Model 2
Model 3
Model 4
Model 5
COMBINED PREDICTION



Random
Forest
ML algo

Boosting

- Before we go further, here's another question for you: If a data point is incorrectly predicted by the first model, and then the next (probably all models), will combining the predictions provide better results? Such situations are taken care of by boosting.

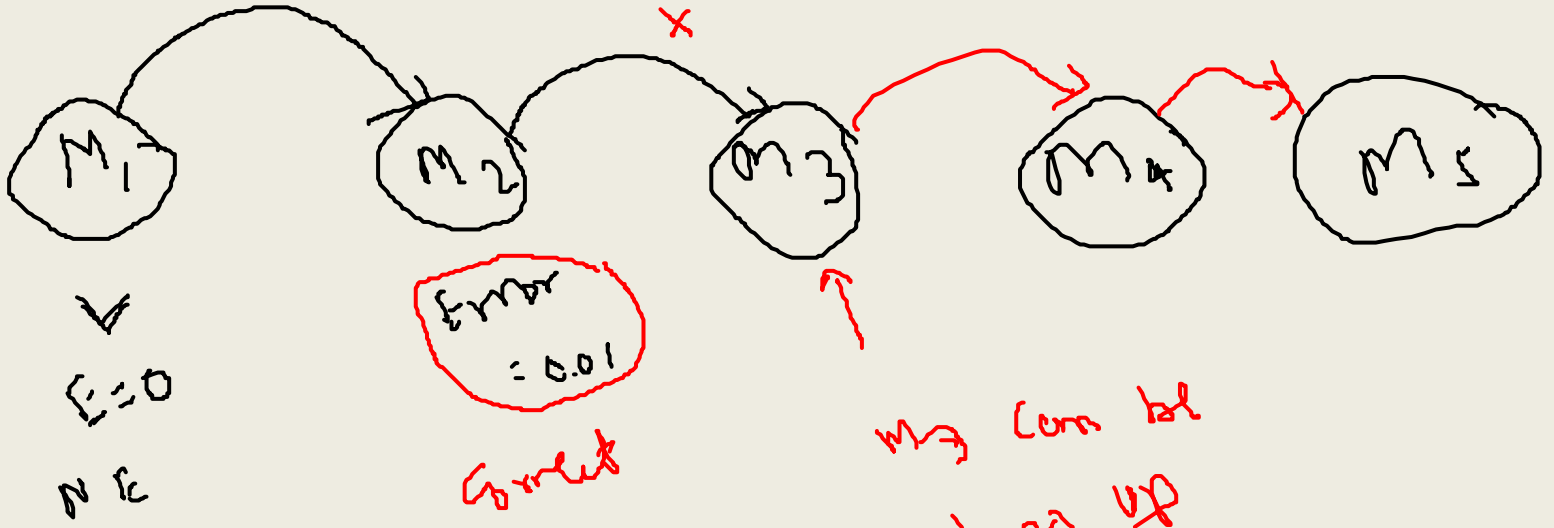
- Boosting is a sequential process, where each subsequent model attempts to correct the errors of the previous model. The succeeding models are dependent on the previous model.

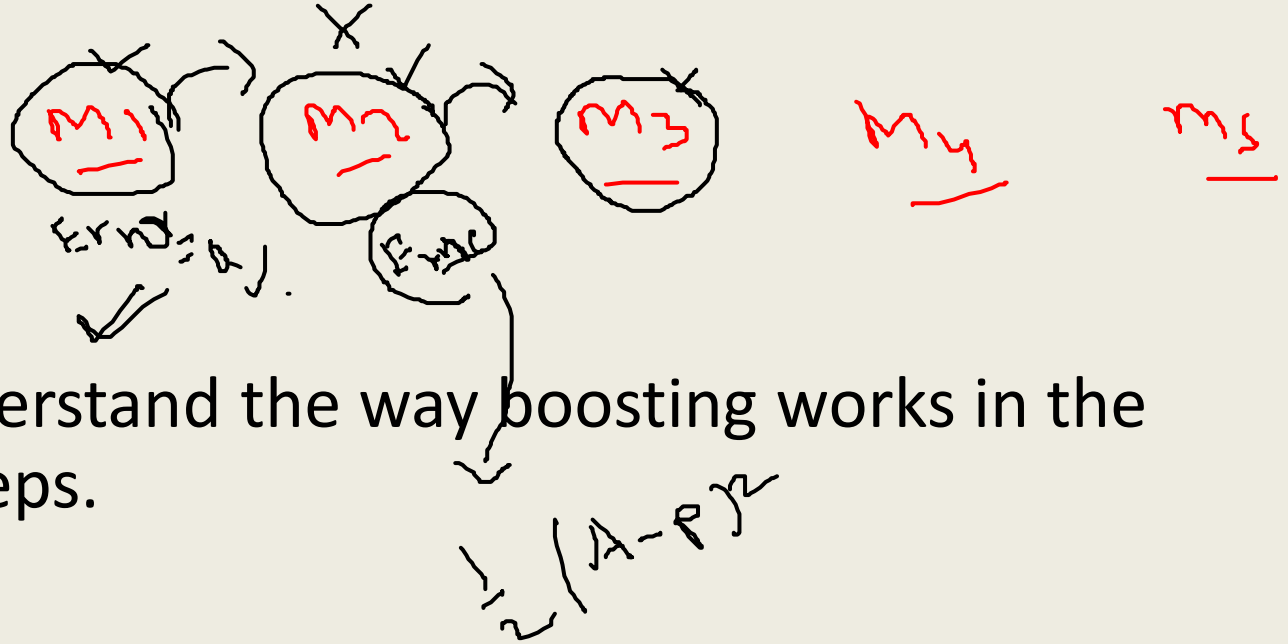
VIT UNIVERSITY
VELLORE INSTITUTE OF TECHNOLOGY
VIT®
BHOPAL
www.vitbhopal.ac.in

$E = \frac{1}{2}(A-P)^2$

Boosting

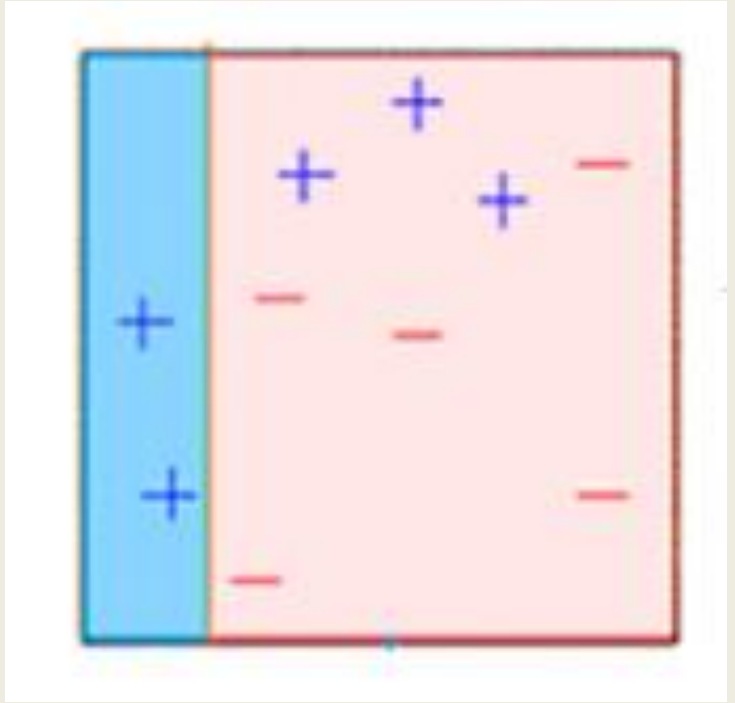
ADABOOST
(Adaptive Boosting)



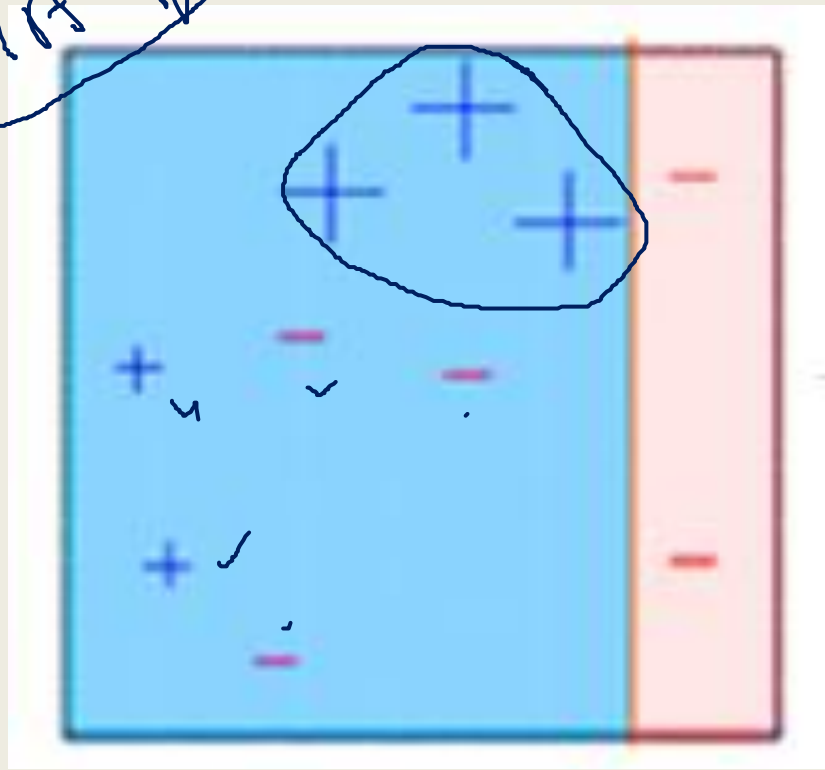


- Let’s understand the way boosting works in the below steps.

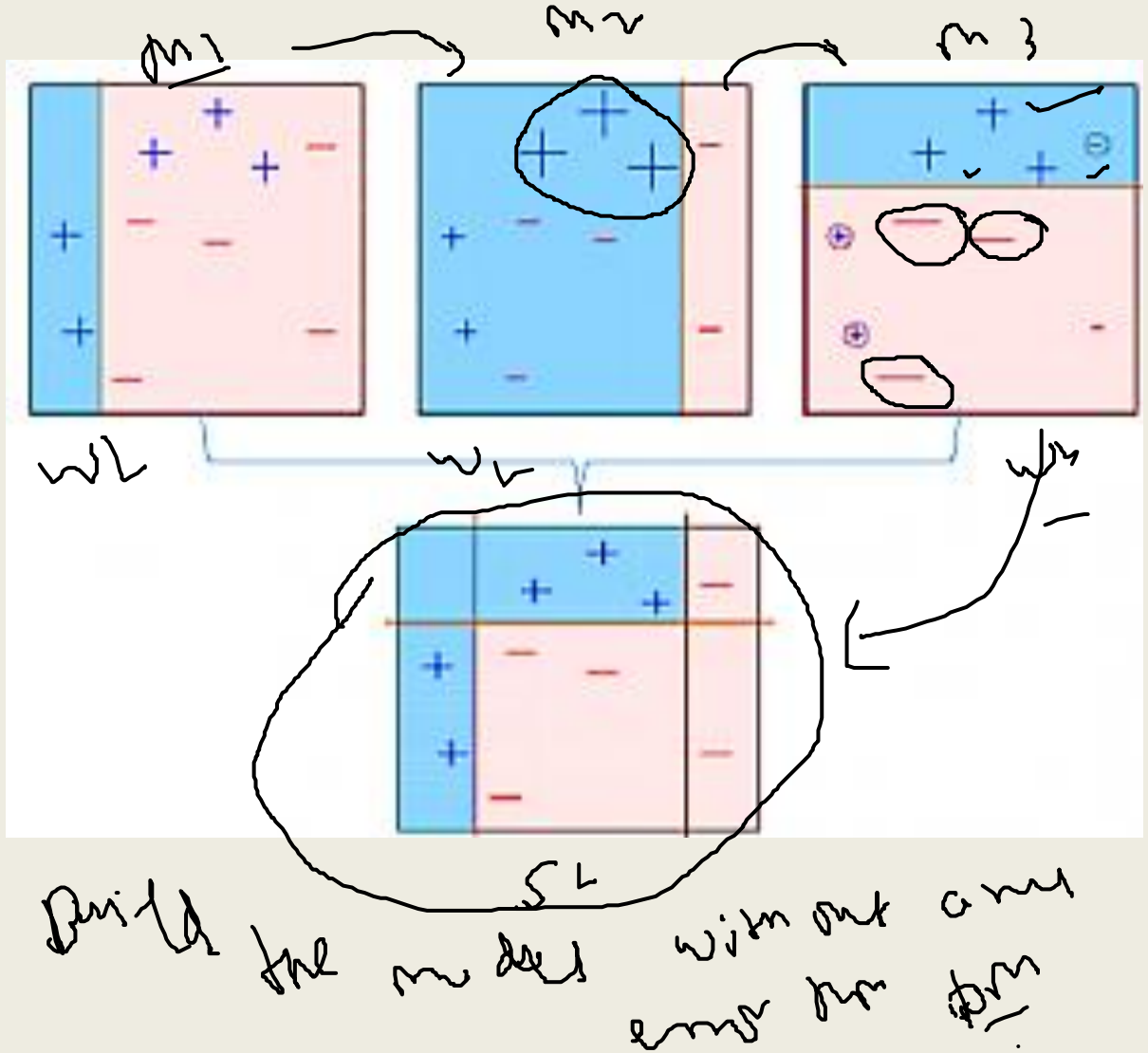
- A subset is created from the original dataset.
- Initially, all data points are given equal weights.
- A base model is created on this subset.
- This model is used to make predictions on the whole dataset.



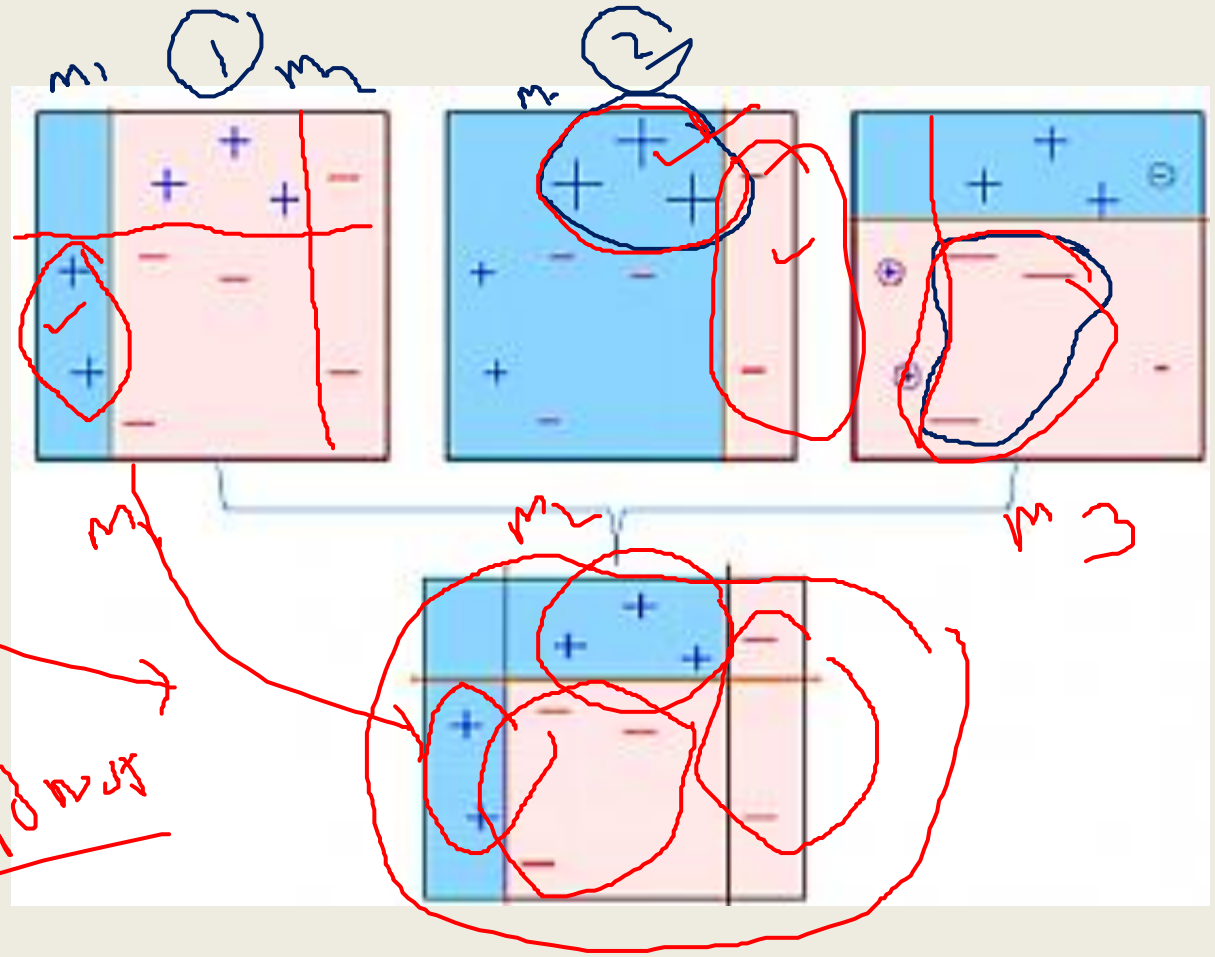
- Errors are calculated using the actual values and predicted values.
- The observations which are incorrectly predicted, are given higher weights.
- (Here, the three misclassified blue-plus points will be given higher weights)
- Another model is created and predictions are made on the dataset.
- (This model tries to correct the errors from the previous model)



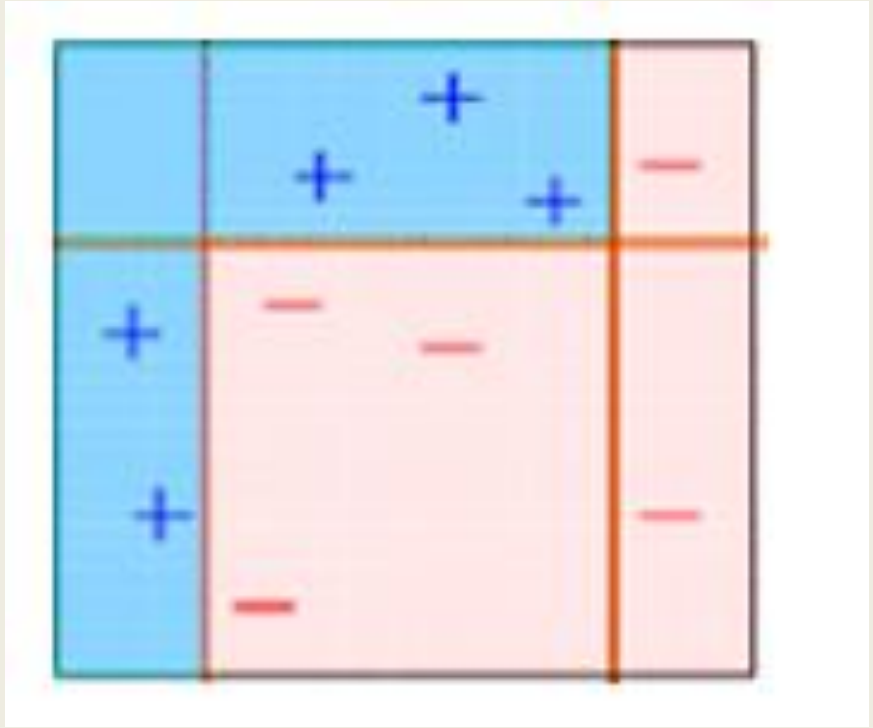
- Similarly, multiple models are created, each correcting the errors of the previous model.
- The final model (strong learner) is the weighted mean of all the models (weak learners)



- Similarly, multiple models are created, each correcting the errors of the previous model.
- The final model (strong learner) is the weighted mean of all the models (weak learners)



- Thus, the boosting algorithm combines a number of weak learners to form a strong learner. The individual models would not perform well on the entire dataset, but they work well for some part of the dataset. Thus, each model actually boosts the performance of the ensemble.

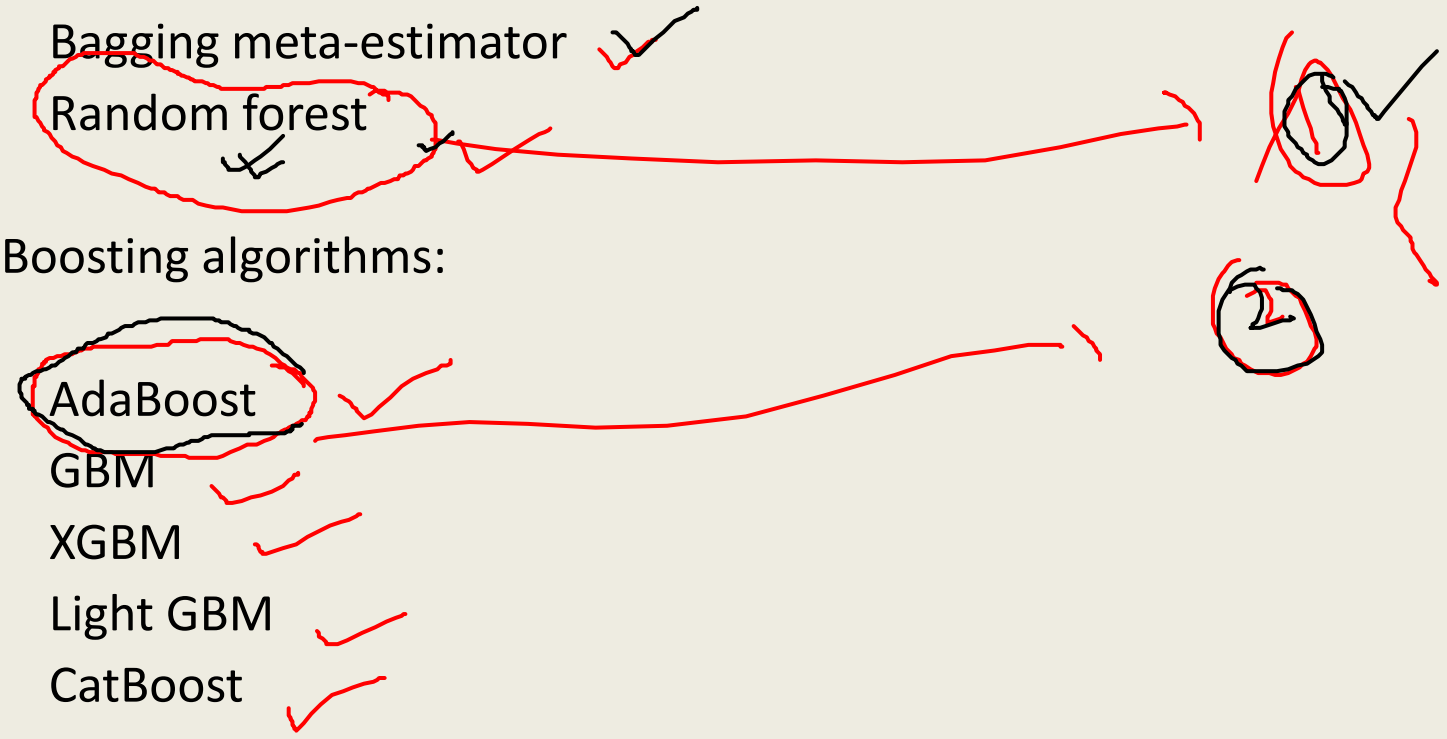


Algorithms based on Bagging and Boosting

VIT UNIVERSITY
VELLORE INSTITUTE OF TECHNOLOGY
VIT BHOPAL
www.vitbhopal.ac.in

- Bagging and Boosting are two of the most commonly used techniques in machine learning. In this section, we will look at them in detail. Following are the algorithms we will be focusing on:

- Bagging algorithms:
  - Bagging meta-estimator
  - Random forest
- Boosting algorithms:
  - AdaBoost
  - GBM
  - XGBM
  - Light GBM
  - CatBoost



VIT®

BHOPAL

www.vitbhopal.ac.in