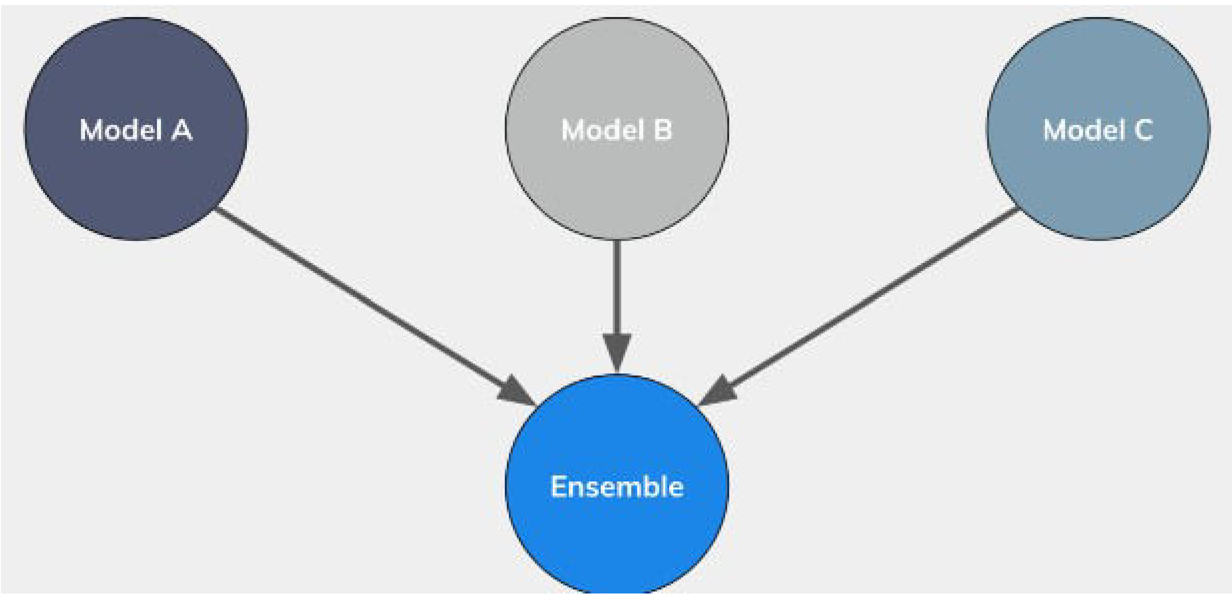
Ensemble Techniques

A Brief Introduction To Ensemble Techniques

Let’s say an individual wants to buy a new phone, it is highly unlikely that he/she will go to a shop and directly buy the phone that the seller recommends. Rather here is what individuals do. They will visit various physical stores, as well as online shopping portals. Then they will browse through various models, look for their ratings from a huge customer base, or ask for recommendations of their friends, colleagues etc. So the main idea is that before we take any decision, we look at various factors, measure the pros and cons and then take decisions.

This idea or working style lays down the framework / basic working principle of Ensemble Techniques. Ensemble modelling techniques combine the results of various models to enhance performance. This can be accomplished in a variety of ways. The integrated models considerably improve the results' accuracy. Due of this, ensemble approaches in machine learning have gained prominence.

The below diagram accurately sums up what Ensemble Techniques are.



Working Mechanism Of Ensemble Techniques

Before going directly into how Ensemble Techniques work, let us first understand some terms which will help to clear our understanding.

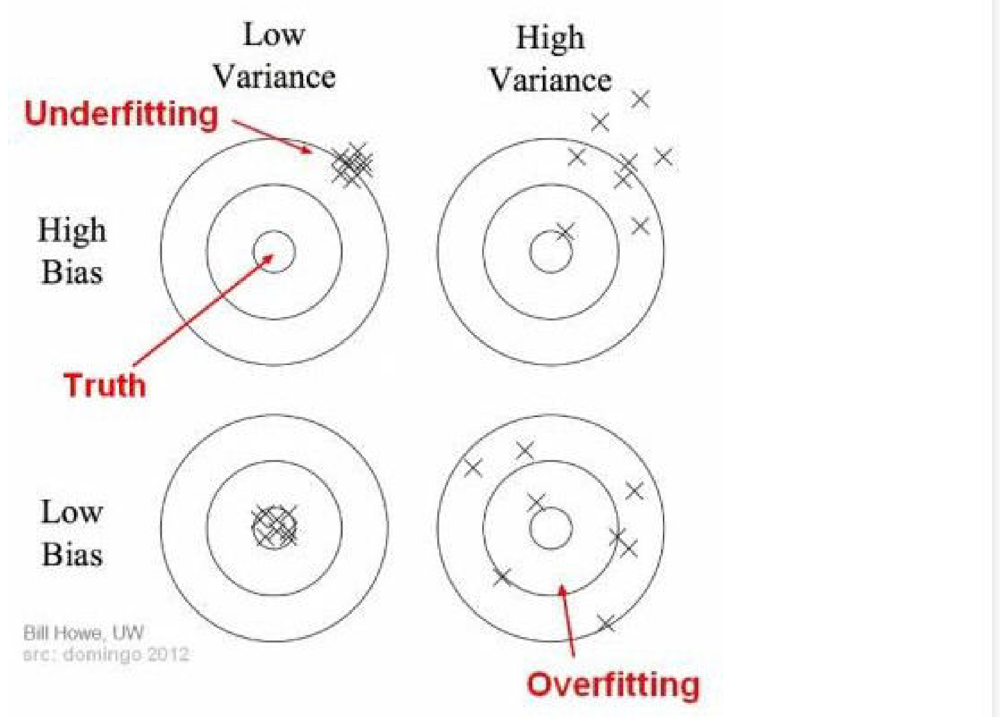
1. Errors – The below attached diagram clearly explains what errors are, in a simple manner.

$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\hat{f}(x) - E[\hat{f}(x)]\right]^2 + \sigma_e^2$$

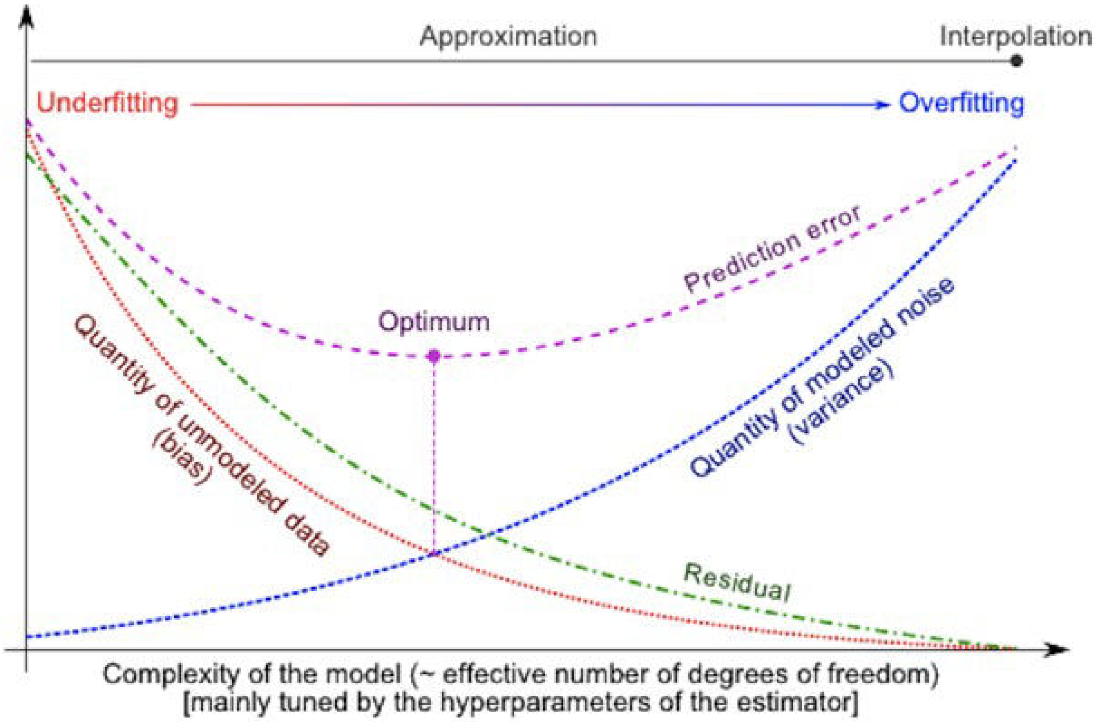
$$Err(x) = \mathrm{Bias}^2 + \mathrm{Variance} + \mathrm{Irreducible\ Error}$$

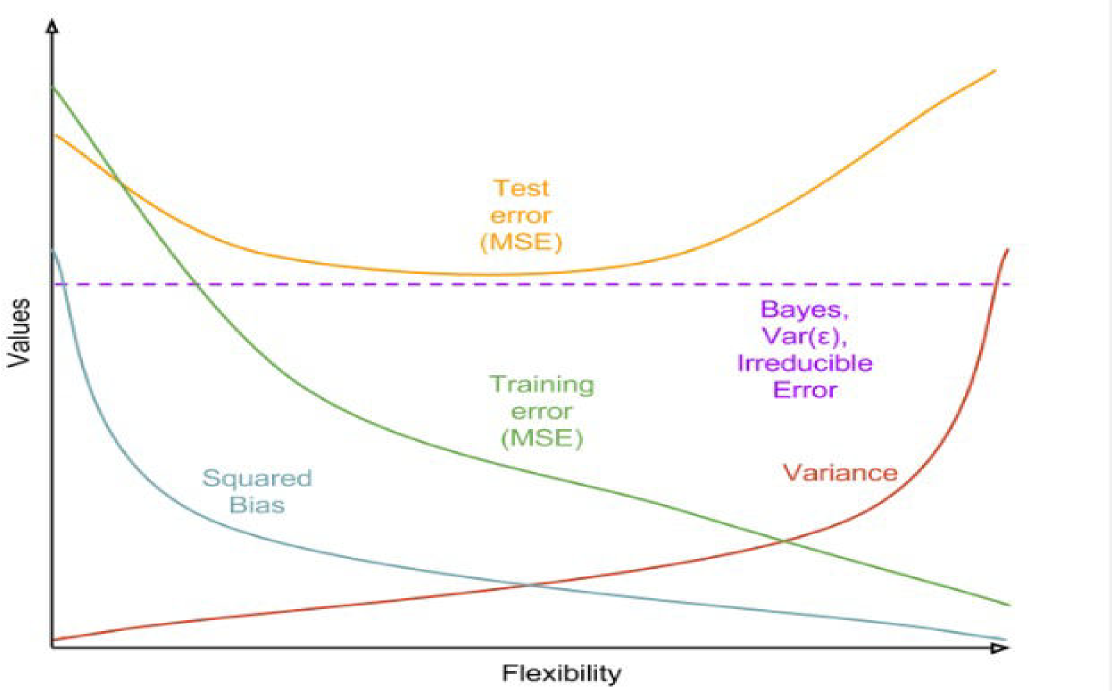
2. Bias Error – Helps to identify how much on average, the actual values differ from predicted values. A low bias error means that we have an over performing model and vice versa.

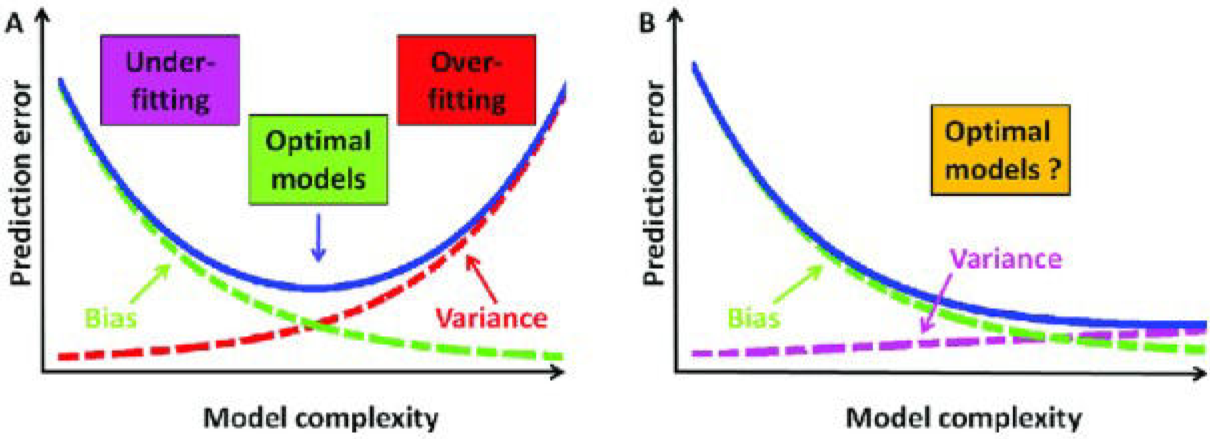
3. Variance - This quantifies how prediction made on same observations differ from each other. A high variance model will over-fit on your training population and perform badly on any observation beyond training.



Thus, any model should try to strike a balance between these two types of errors. This trade off analysis is carried out using Ensemble Techniques.







A group of predictors is called an ensemble; thus, this technique is called Ensemble Learning, and an Ensemble Learning algorithm is called an Ensemble method.

Types Of Ensemble Techniques

Basic Ensemble Techniques

- Max Voting
- Averaging
- Weighted Average

Advanced Ensemble Techniques

- Stacking
- Blending
- Bagging
- Boosting

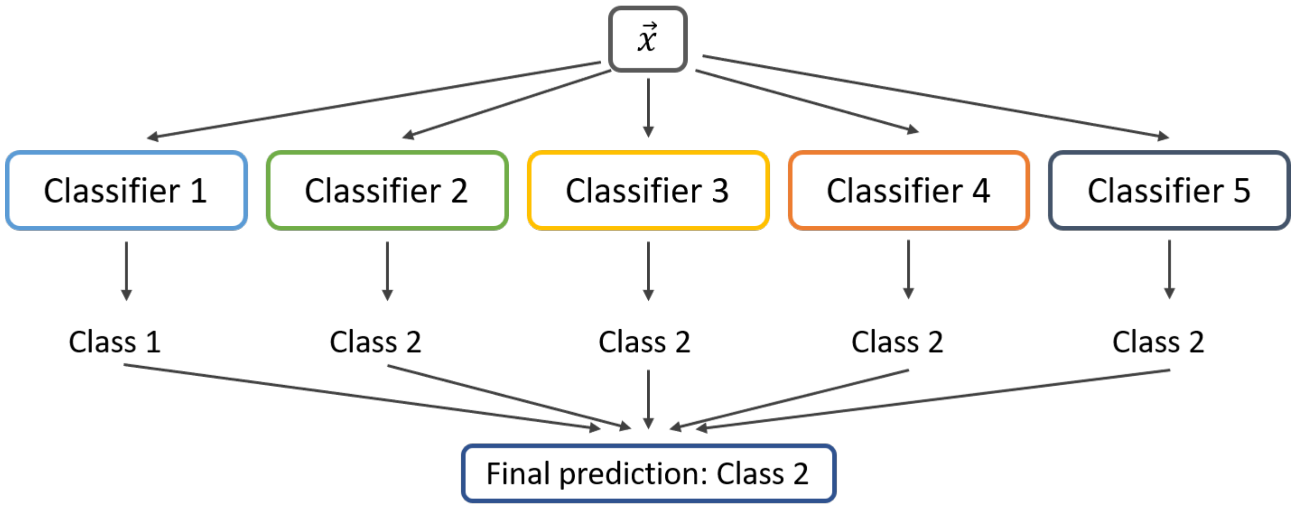
Algorithms based on Bagging and Boosting

- Bagging meta-estimator
- Random Forest
- AdaBoost

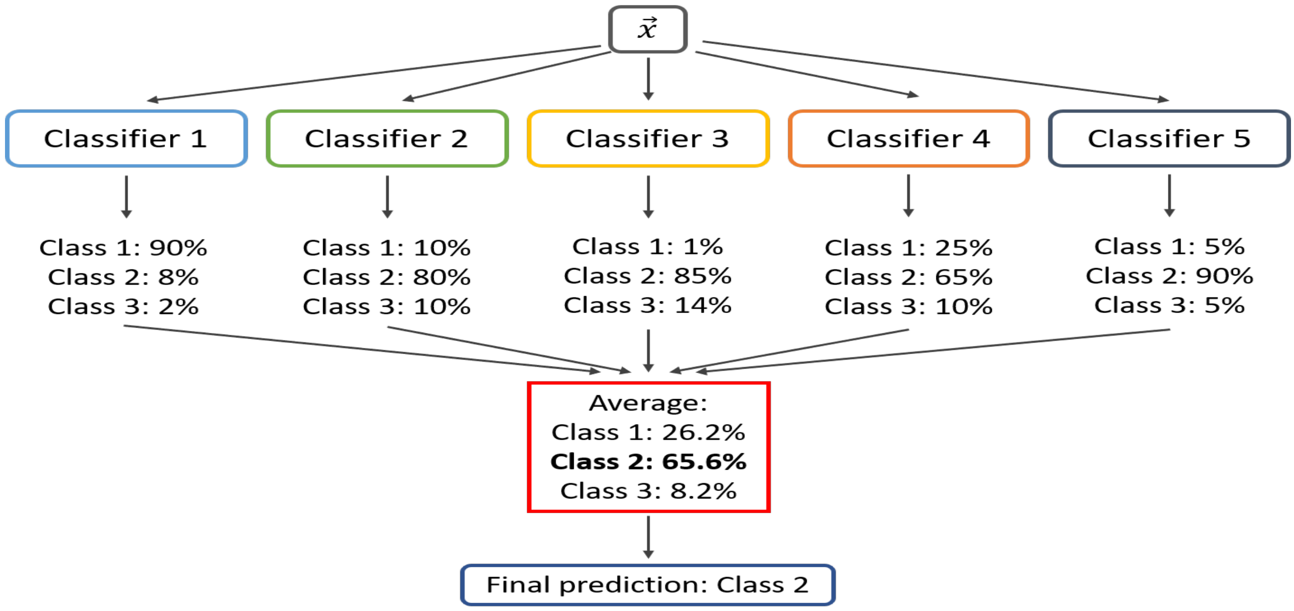
- GBM
- XGB
- Light GBM
- CatBoost

1. Max Voting - For classification issues, the max voting approach is typically utilised. With this method, predictions are made for each data point using a variety of models. Each model's predictions are regarded as a "vote." The majority of the models' forecasts serve as the basis for the final projection.

A voting classifier is a machine learning model that gains experience by training on a collection of several models and forecasts an output (class) based on the class with the highest likelihood of being the output. To predict the output class based on the highest majority of votes, it merely averages the results of each classifier that was passed into the voting classifier.

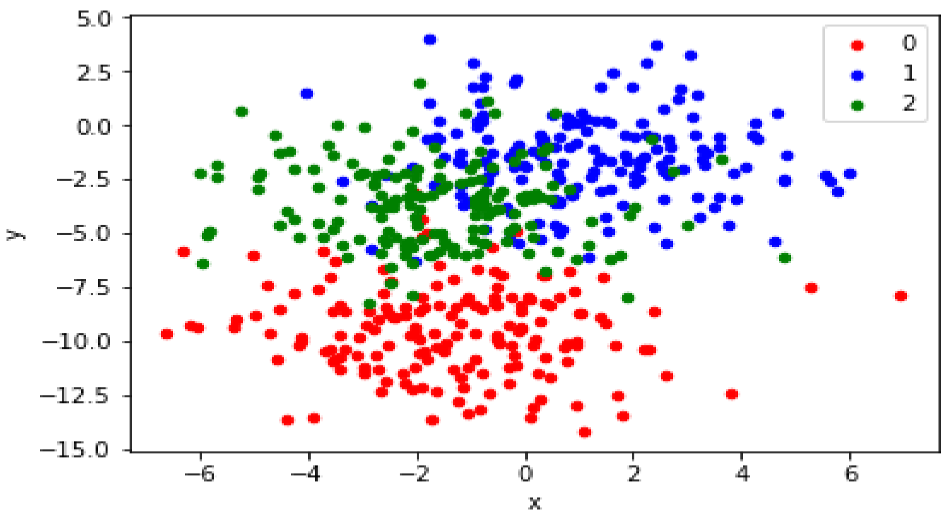


As you can see, the ensemble determines the final prediction is "Class 2" because four classifiers predicted "Class 2" and only one classifier predicted "Class 1". Hard voting is another name for this voting procedure.



If all classifiers in the ensemble have prediction probabilities, we can also use the so-called soft voting rule.

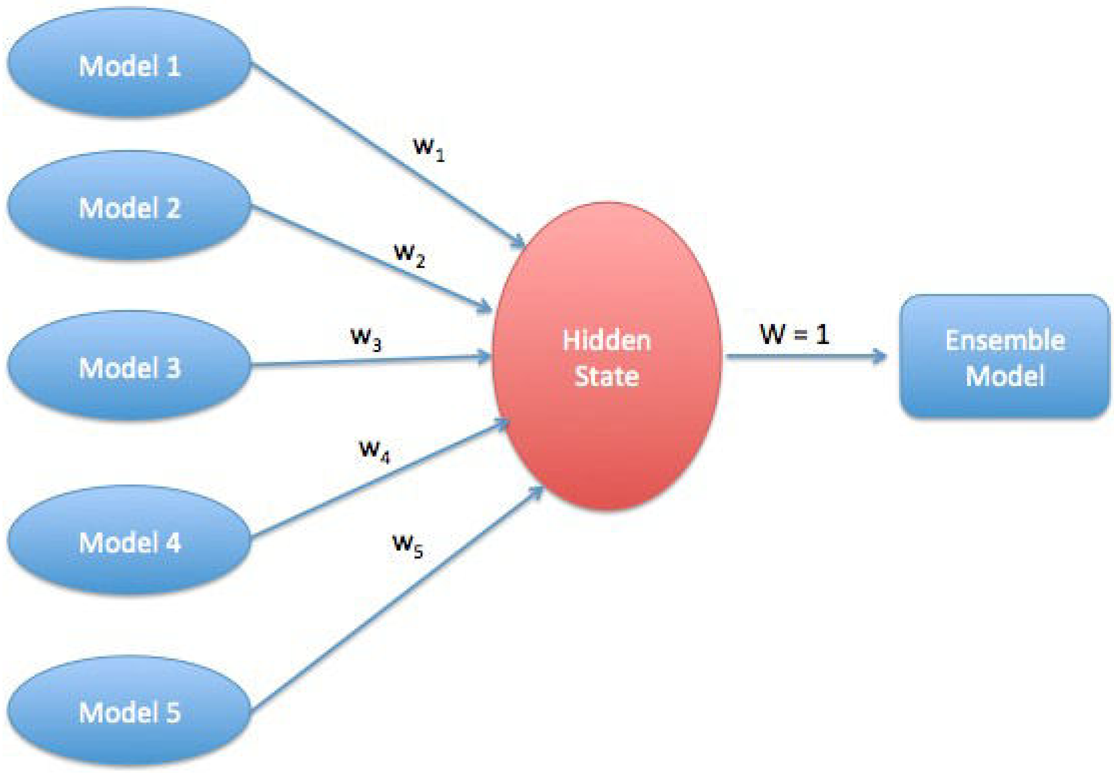
2. Averaging - Multiple forecasts are made for each data point when averaging, similar to the max voting method. In this approach, the final prediction is made by averaging the results of all the models. When computing probabilities for classification problems or making predictions in regression problems, averaging can be applied.



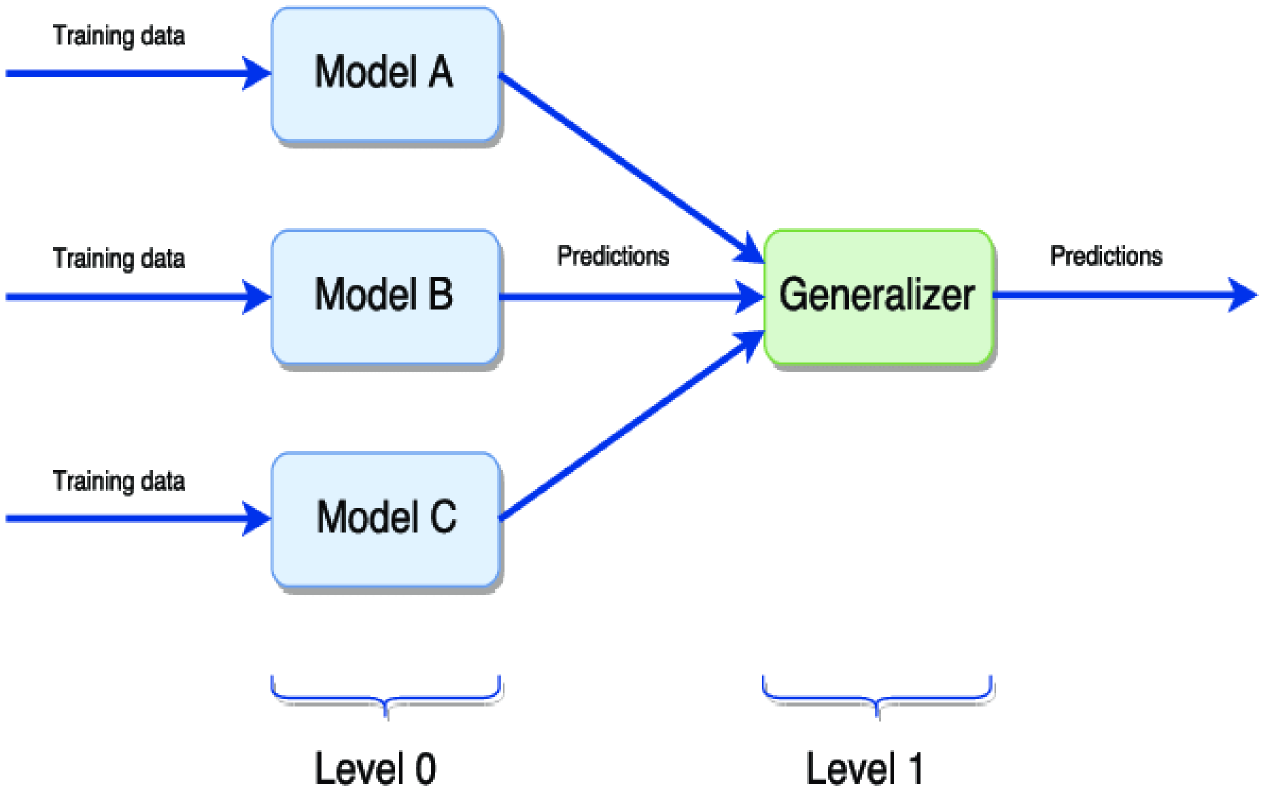
We can observe that the classes are not linearly separable (separable by a line), resulting in numerous confusing locations, as indicated by the standard deviation of 2.0.

So, in conclusion, the average predictions are determined for each occurrence of the test dataset. This method often reduces overfit and creates a smoother regression model.

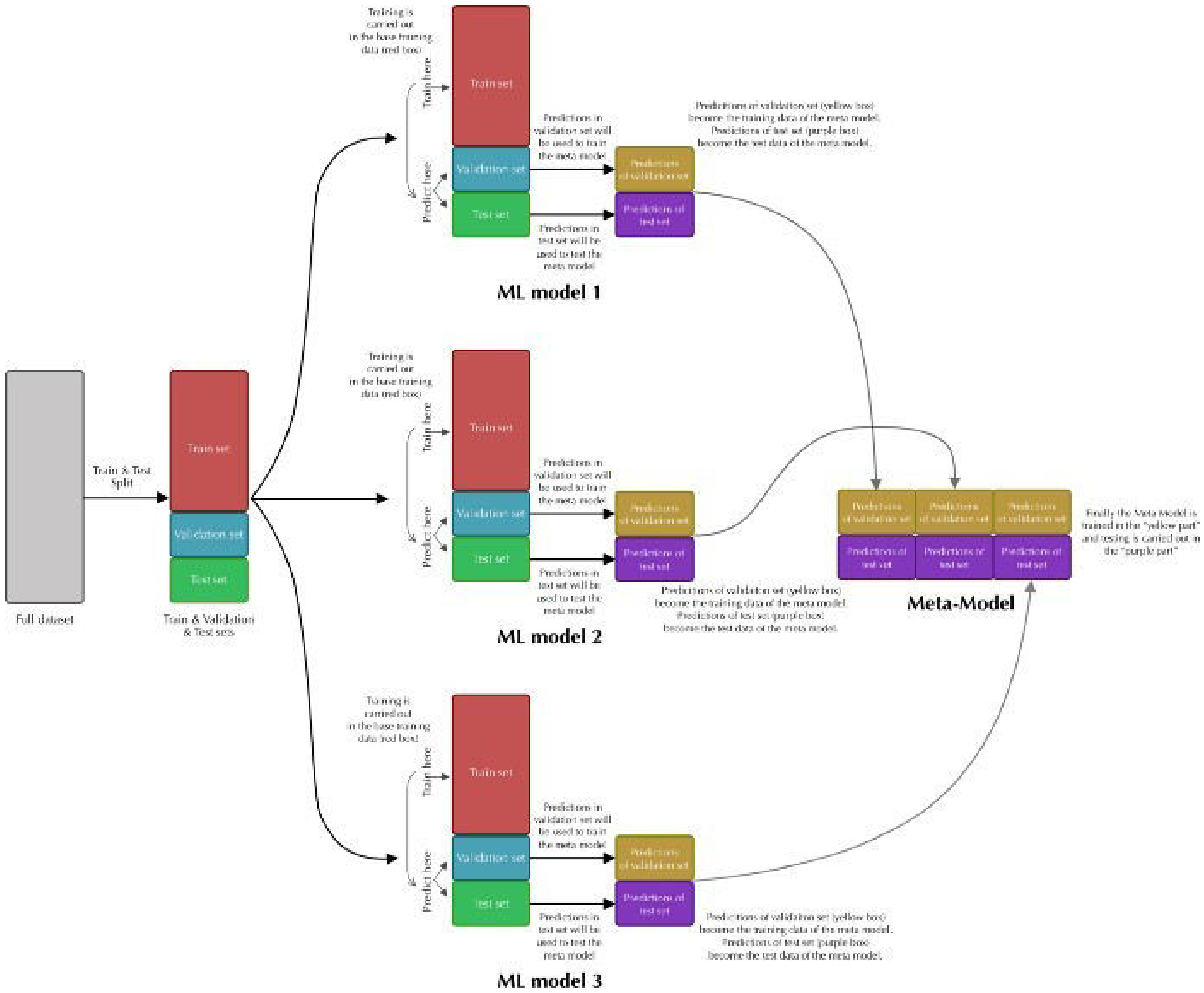
3. This is a development of the averaging approach. Different weights are assigned to each model, indicating the significance of each model for prediction.



4. Stacking - Stacking is an ensemble learning strategy that creates a new model by combining predictions from other models (such as decision trees, knn, or svm). On the test set, predictions are made using this model. Stacking, often referred to as stacked generalisation, is an ensemble technique that employs a meta-classifier or meta-regressor to merge many classification or regression models. The meta-model is trained on the features that are outputs of the base-level models after the base-level models have been trained on the entire training set. Because the base-level frequently consists of various learning methods, stacking ensembles are frequently diverse.



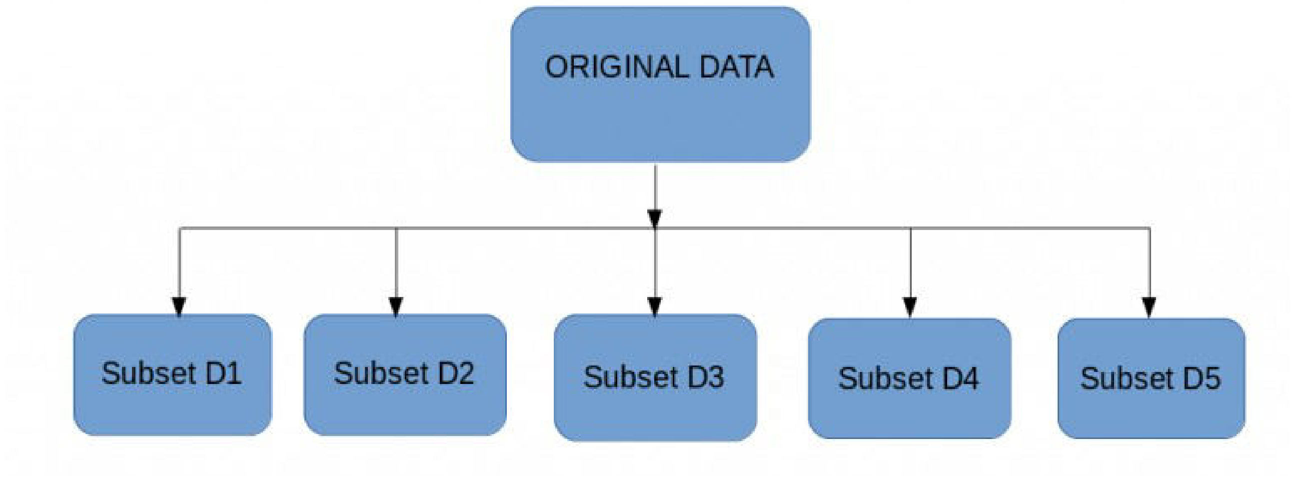
5. Blending - Blending follows the same approach as stacking but uses only a holdout (validation) set from the train set to make predictions. In other words, unlike stacking, the predictions are made on the holdout set only. The holdout set and the predictions are used to build a model which is run on the test set. Here is a detailed explanation of the blending process:



6. Bagging - Combining bootstrapping and aggregating is referred to as bagging. By resampling data from the training set with the same cardinality as the original set, bootstrapping is a technique to assist in lowering the classifier's variance and reducing overfitting. Comparing the developed model to a single individual model, it must be less overfitted.

It is bad for a model to have a high variance because this indicates that the model's success depends on the training set. Therefore, the model can still perform poorly even if extra training data are provided. Additionally, it might not even lessen the variance of our model.

The simplest approach with bagging is to use a couple of small subsamples and bag them, if the ensemble accuracy is much higher than the base models, it’s working; if not, use larger subsamples.Using larger subsamples is not guaranteed to improve your results. In bagging there is a tradeoff between base model accuracy and the gain you get through bagging. The aggregation from bagging may improve the ensemble greatly when you have an unstable model, yet when your base models are more stable — been trained on larger subsamples with higher accuracy — improvements from bagging reduces.

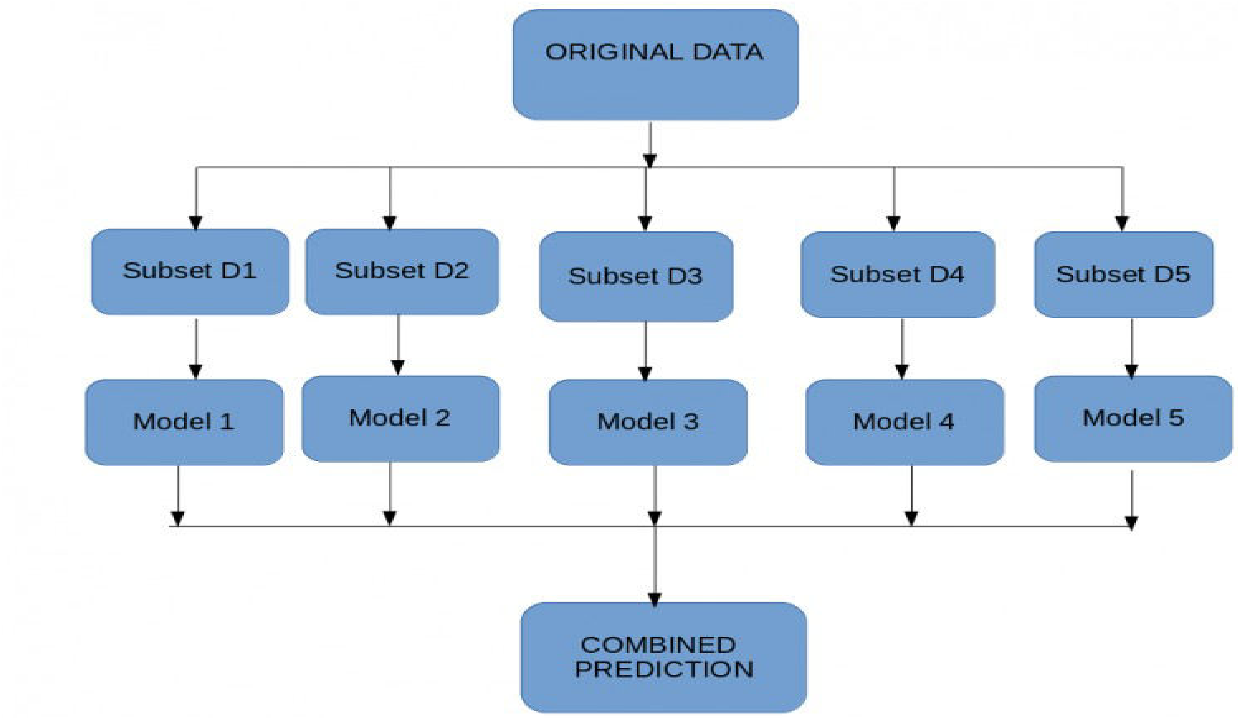


Multiple subsets are created from the original dataset, selecting observations with replacement.

1.A base model (weak model) is created on each of these subsets.

2.The models run in parallel and are independent of each other.

3.The final predictions are determined by combining the predictions from



all the models.

7. Boosting - The fundamental goal of boosting is to successively add new models to the overall ensemble model. Previously, when bagging, we averaged every single model that was produced. This time, a new model is generated with each iteration of boosting, and the new base-learner model is trained (updated) using the mistakes made by the prior learners.

An overall prediction is obtained by adding the result from each weak model that the programme builds. Ensemble modelling from before is being used here. Similar to how gradient descent goes toward the correct values, the now-boosted gradient adjusts the current prediction, nudging it toward the true target. Instead of using each model's specific parameters, the gradient descent optimization is applied to the output of the many models.