Machine Learning
ICT-4261

By-
Dr. Jesmin Akhter
Professor
Institute of Information Technology
Jahangirnagar University

Contents

The course will mainly cover the following topics:

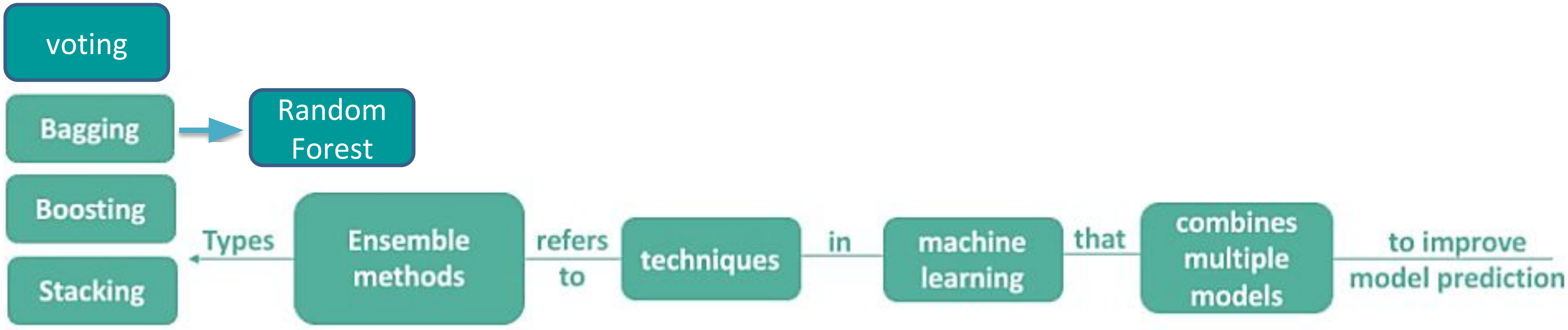
Outline

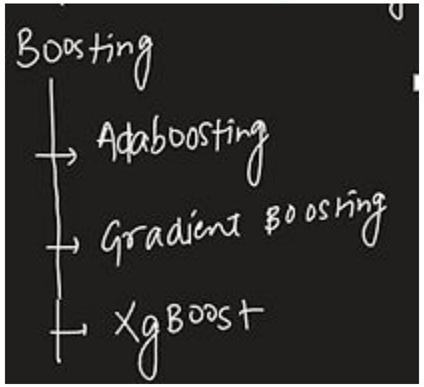
✔ Ensemble Learning

Ensemble Learning

- ✔ Ensemble methods refer to the techniques used in machine learning to combine multiple base models to achieve better predictive performance. Rather than relying on a single model, ensemble methods aim to leverage the strengths of multiple models to create a more accurate and robust predictor.
- ✔ Ensemble methods are widely used in various applications, including image classification, natural language processing, and financial market prediction. By combining multiple models, ensemble methods can help to **reduce the risk of overfitting, improve the accuracy** and stability of predictions, and provide a more robust and reliable predictor.
- ✔ The term “model” to describe the output of the algorithm that trained with data. This model is then used for making predictions. This algorithm can be any machine learning algorithm such as logistic regression, SVM, decision tree, etc. These models, when used as inputs of ensemble methods, are called ”base models,” and the end result is an ensemble model.
- ✔ It combines low performing classifiers (also called as weak learners or base learner) and combine individual model prediction for the final prediction.
- ✔ On the basis of type of base learners, ensemble methods can be categorized as homogeneous and heterogeneous ensemble methods. If base learners are same, then it is a homogeneous ensemble method. If base learners are different then it is a heterogeneous ensemble method.

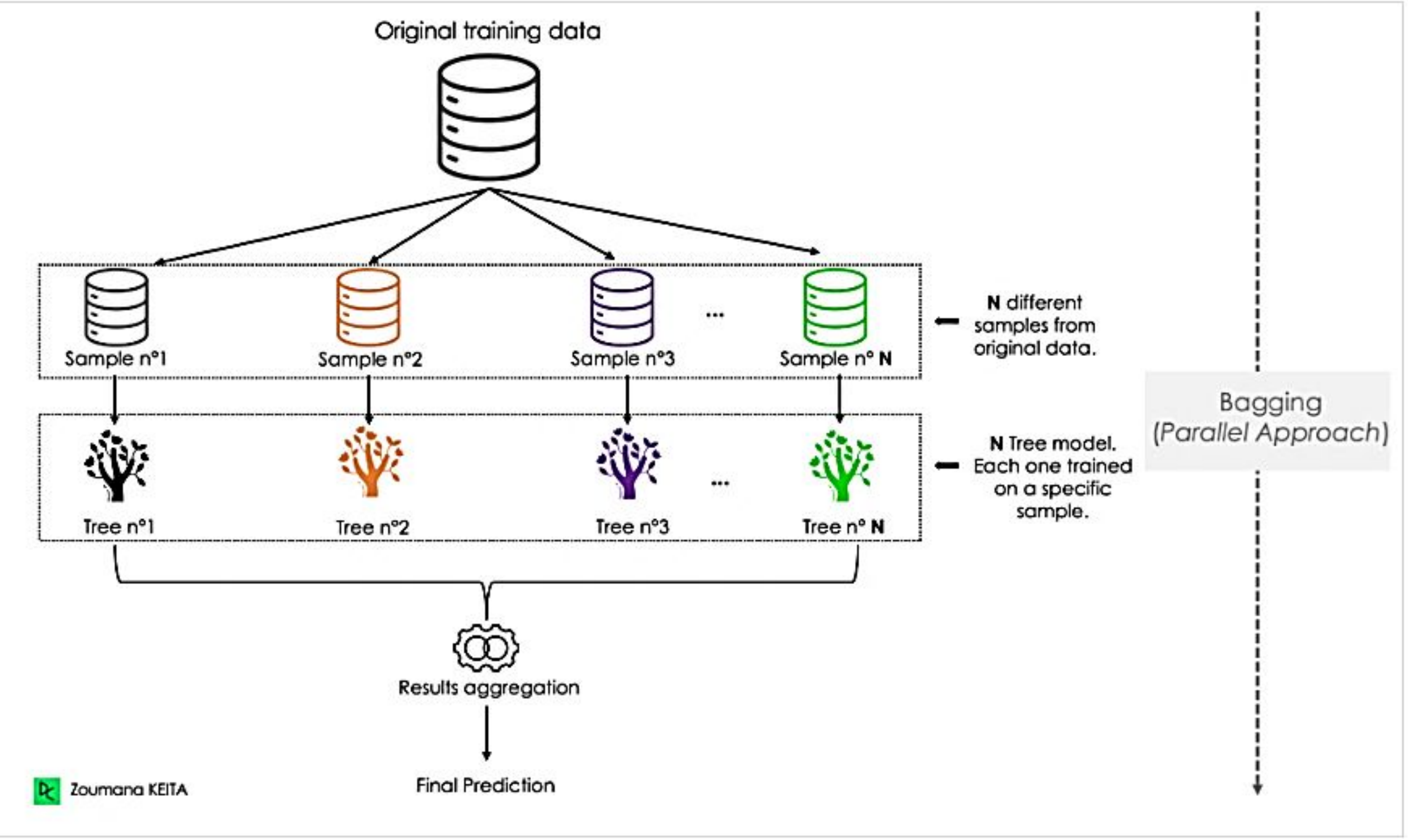
Main Types of Ensemble Methods



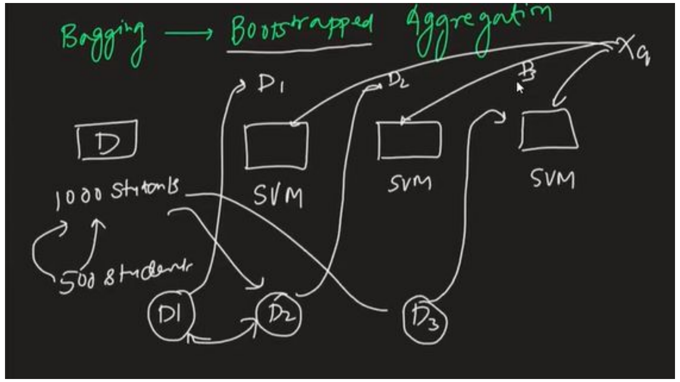


Bagging

- ✔ Bagging, short for Bootstrap Aggregating, is a machine learning ensemble technique used to improve the accuracy and stability of a model. It generates multiple subsets of the training data by random sampling with replacement and then training a model on each subset. Finally, the individual models are combined by taking their predictions' average (for regression) or majority vote (for classification). Random Forest is a special case of Bagging.



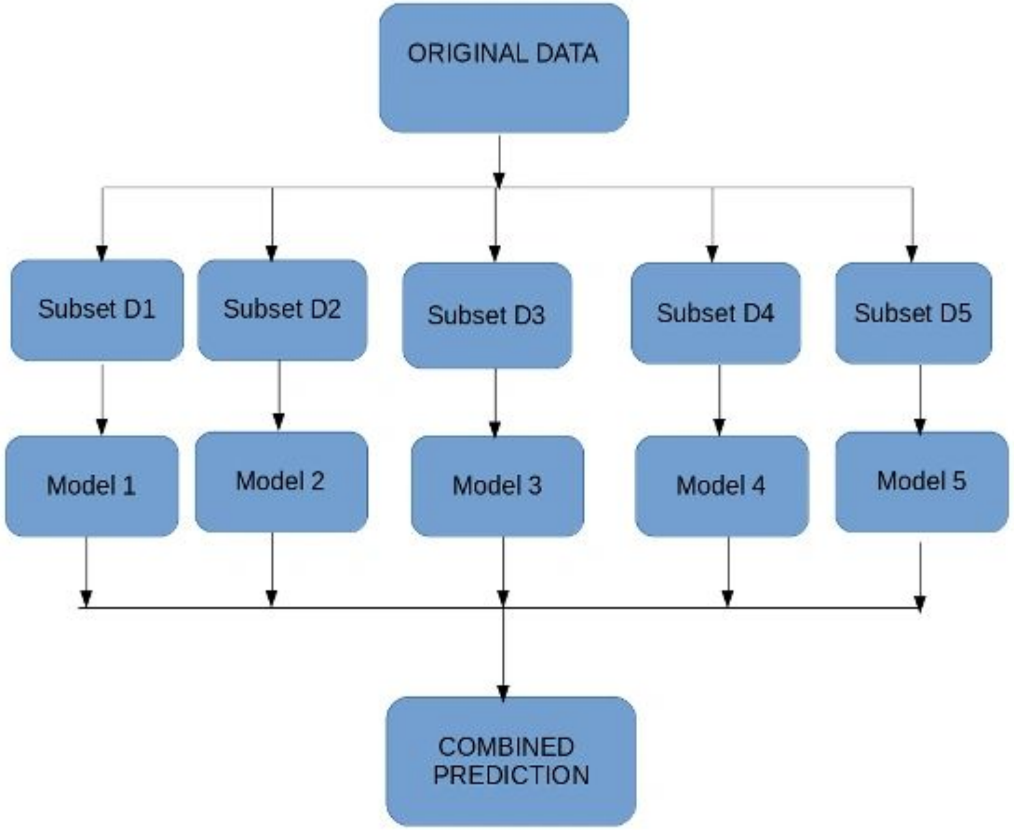
How Does Bagging Work?



- ✔ Multiple subsets are created randomly from the original dataset, selecting observations with replacement. The size of subsets created for bagging may be less than the original set

- ✔ A base model (weak model) is created on each of these subsets.
- ✔ The models run in parallel and are independent of each other. so they have different perspectives on the problem.

- ✔ The predictions from the individual models are combined by taking the average (for regression) or majority vote (for classification) of their predictions. This produces the final prediction of the bagged model.



Bagging

- ✔ By combining multiple models, bagging helps reduce the model's variance and can prevent overfitting by introducing diversity into the training process. It is commonly used with **decision trees** but can also be applied to other models.
- ✔ One important aspect of bagging is that it requires the base model to have high variance but low bias(DT). Bagging can then be used to reduce the variance of the model and improve its generalization performance on new, unseen data.

Example

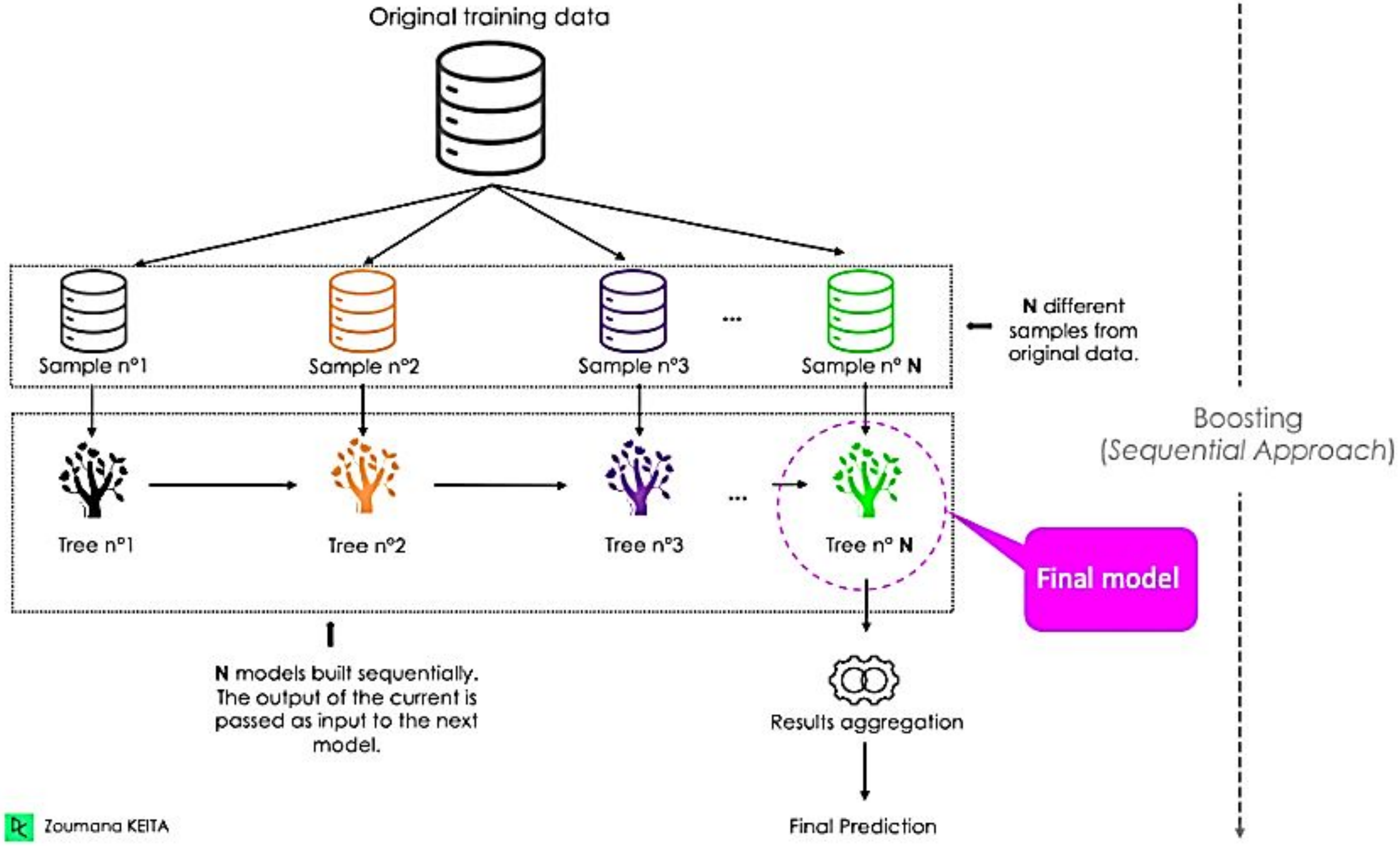
- ✔ Suppose a data scientist is working on a project to predict whether a customer will purchase a certain product based on their **demographic** information and browsing history on a website. The data set consists of 10,000 customers and 50 features, and the data scientist plans to use a decision tree algorithm to build the model.
- ✔ First, the data set is divided into subsets with 1,000 customers. Then, 25 features are randomly selected for each subset, and a decision tree is trained on that subset using only those 25 features.
- ✔ Once all the decision trees are trained, their predictions are combined by taking the majority vote for each customer. If, for example, seven out of ten decision trees predict that a customer will purchase the product, then the bagged model will predict that the customer will make a purchase.

Advantages and Disadvantages

- ✔ Bagging presents several key advantages and disadvantages when used for classification or regression tasks.
- ✔ **Advantages**
  - The modeling process is straightforward and does not require any deep mathematical concepts, and can handle missing values.
  - The scikit-learn package makes it easy to implement the underlying logic.
  - Bagging helps to reduce the model's **variance** and can prevent overfitting by introducing diversity into the training process.
- ✔ **Disadvantages**
  - Bagging is computationally expensive due to the use of several models.
  - The averaging involved across predictions makes it difficult to interpret the final result.

Boosting

✔ Boosting improves machine learning performance by sequentially correcting errors and combining weak learners into strong predictors. Each model iteratively focuses attention on the observations that are misclassified by its predecessors. Models are added until either the complete training data set is predicted correctly or the maximum number of models are added. We've outlined the general process below:

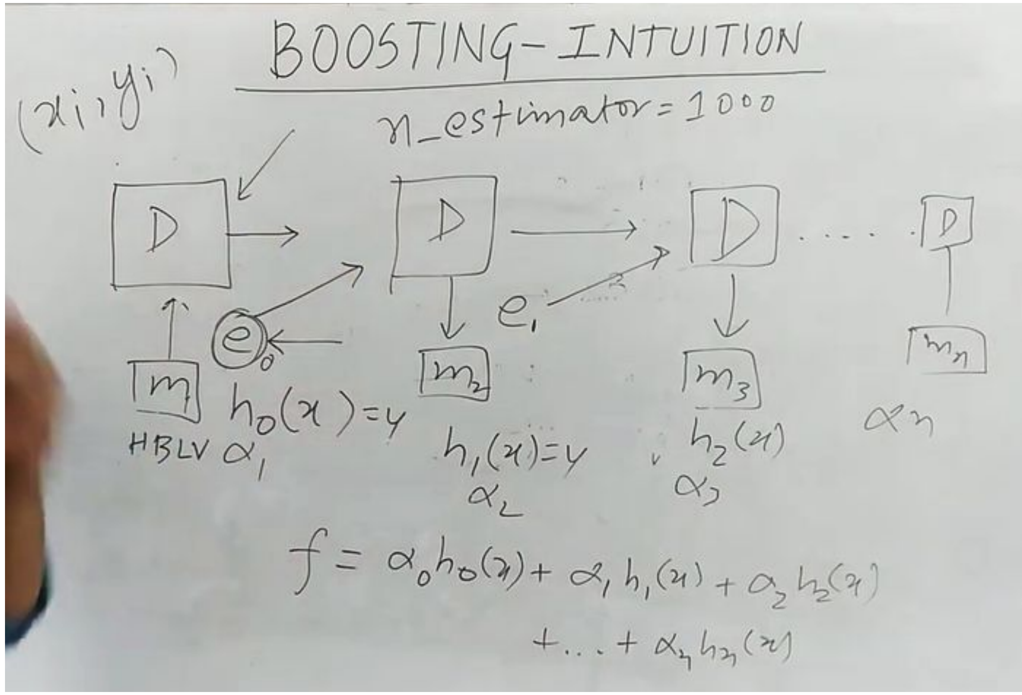


Bagging → LB HV → LB LV

Boosting → HB LV → LB LV

How Boosting Works

- ✔ 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. 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 from the predicted values.
- ✔ The observations which are incorrectly predicted, are given higher weights.
- ✔ Another model is created and predictions are made on the dataset.
- ✔ The final prediction is a weighted sum of the individual models' predictions, with higher weights given to more accurate models.
- ✔ 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.
- ✔ Boosting algorithms like AdaBoost, Gradient Boosting, and XGBoost are popular because they improve model performance.



Boosting

- ✔ **Adaptive boosting** (AdaBoost) weights model errors. That is, when creating a new iteration of a dataset for training the next learner, AdaBoost adds weights to the previous learner's misclassified samples, causing the next learner to prioritize those misclassified samples.

- ✔ **Gradient boosting** uses residual errors when training new learners. Rather than weight misclassified samples, gradient boosting uses residual errors from a previous model to set target predictions for the next model. In this way, it attempts to close the gap of error left by one model.

Advantages and Disadvantages

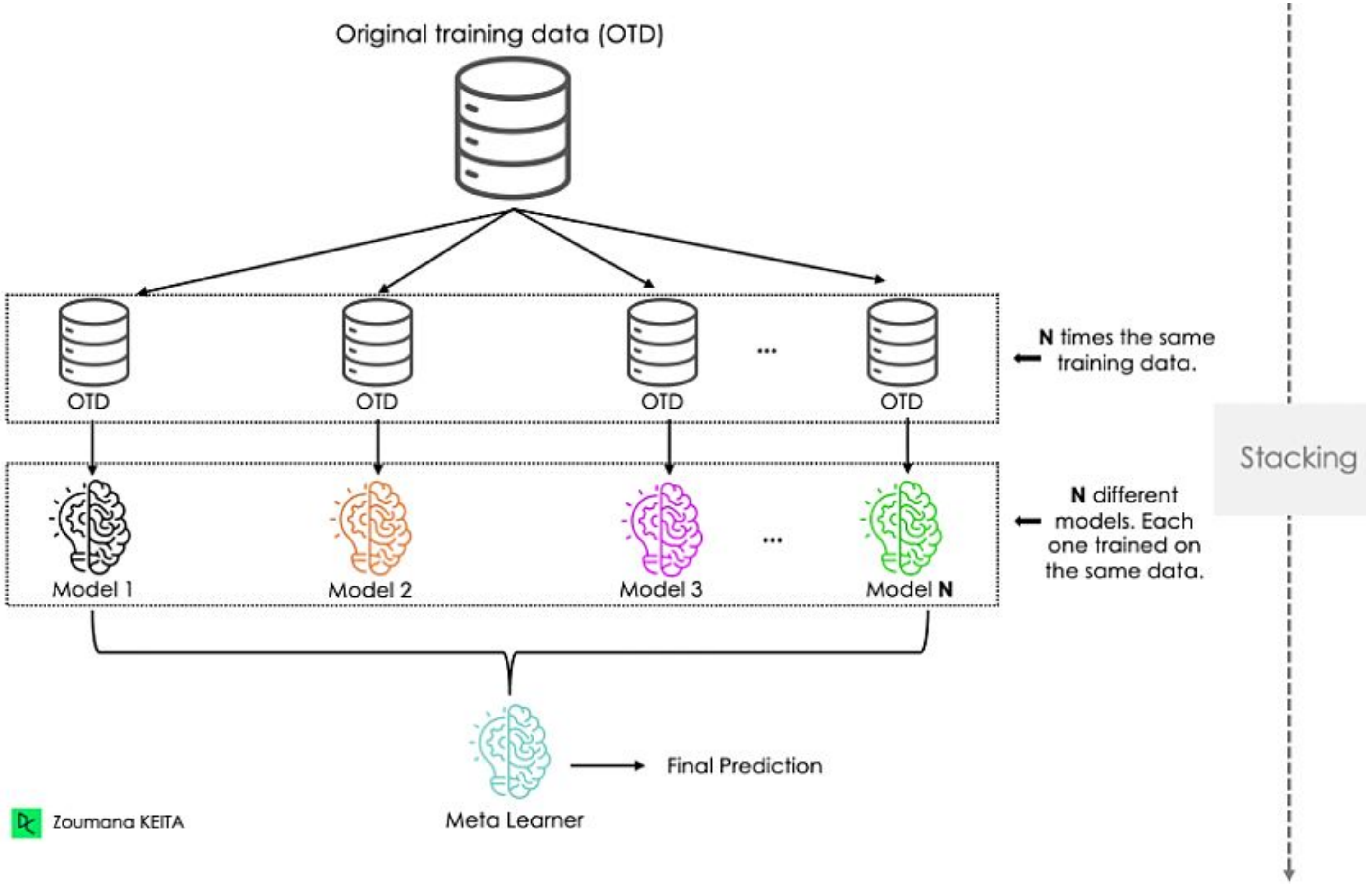
- ✔ **Advantages**
  - Boosting algorithms efficiently reduces bias.
  - This process reduces the dimensionality of the data, hence reducing the computation time.
- ✔ **Disadvantages**
  - The sequential approach of boosting makes the next models correct the mistakes of its predecessor. This makes the overall model vulnerable to outliers.
  - Boosting is not scalable for the same reason as the sequential aspect.

Boosting vs Bagging

Boosting	Bagging
In Boosting we combine predictions that belong to different types	Bagging is a method of combining the same type of prediction
The main aim of boosting is to **decrease bias**, not variance	The main aim of bagging is to **decrease variance** not bias
At every successive layer Models are weighted according to their performance.	All the models have the same weightage
New Models are influenced by the accuracy of previous Models	All the models are independent of each other
Works best with models that have low variance but **high bias**	Works best with models that have **high variance** but low bias

Stacking

✔ The predictions from the base learners are stacked together and are used as the input to train the meta learner to produce more robust predictions. The meta learner is then used to make final predictions. Bagging and boosting typically use homogeneous base learners, whereas stacking tends to include heterogeneous ones.



Advantages and Disadvantages

Advantages

- ✔ It leverages the strengths of multiple high-performing models for both classification and regression tasks.
- ✔ Similar to other ensemble models, stacking helps build a more accurate model than individual models used alone.

Disadvantages

- ✔ Using complex basic models can increase the risk of overfitting.
- ✔ Different levels of a model’s training can make the stacking architecture complex to implement.

Voting

- ✔ In voting, multiple models are trained independently, and their predictions are combined to make a final prediction. It is commonly used for classification tasks, although it can also be applied to regression problems.

- ✔ There are three different types of voting classifiers:
  - **Hard Voting:** In hard voting, the predicted output class is a class with the **highest majority of votes**, i.e., the class with the highest probability of being predicted by each classifier. For example, let's say classifiers predicted the output classes as (Cat, Dog, Dog). As the classifiers predicted class "dog" a maximum number of times, we will proceed with Dog as our final prediction.
  - **Soft Voting:** In this, the **average probabilities of the classes** determine which one will be the final prediction. For example, let's say the probabilities of the class being a "dog" is (0.30, 0.47, 0.53) and a "cat" is (0.20, 0.32, 0.40). So, the average for a class dog is 0.4333, and the cat is 0.3067, from this, we can confirm our final prediction to be a dog as it has the highest average probability.
  - **Weighted voting** assigns different weights to the predictions of individual models based on their performance or reliability. Each model’s prediction is multiplied by its corresponding weight, and the final prediction is obtained by summing or averaging the weighted predictions.

Advantages and Disadvantages

- ✔ **Advantages**
  - Voting architecture is simple to implement compared to stacking, and blending. Also, it does not require complex fine-tuning.
  - Using multiple base learners in the voting makes it less susceptible to the influence of individual models, which contributes to making more stable and reliable predictions.
- ✔ **Disadvantages**
  - It can be difficult to deal with models' prediction conflict, which makes it hard to make the final decision in a meaningful way.
  - Adding more models to the ensemble voting model does not necessarily improve the final performance.

Sample Questions

- ✔ **1. Is bagging random forest?**
  - Random Forest is an ensemble method that uses bagging as its main component. It generates multiple decision trees on different subsets of the training data and combines their predictions using averaging or majority voting.
- ✔ **2. How does bagging reduce variance?**
  - Bagging reduces variance by introducing diversity in the training process. By generating multiple subsets of the training data and training a separate model on each subset using the same learning algorithm, Bagging helps reduce individual samples' impact on the final model. This helps to prevent overfitting and improve the model's generalization performance by reducing the impact of outliers or noisy samples on the final prediction.
- ✔ **3. Does bagging increase bias?**
  - Bagging typically does not increase bias in a model, as it uses the same learning algorithm on each subset of the training data. However, it can reduce variance and overfitting, which may lead to a slightly increased bias-variance trade-off.

Thank You