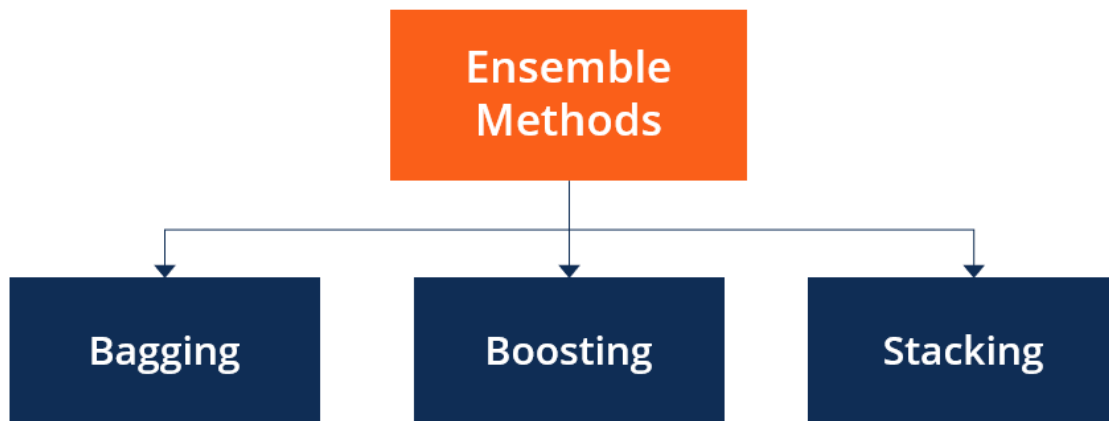# Ensemble Techniques Methods

What are Ensemble Methods?

Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning.

Summary

- Ensemble methods aim at improving predictability in models by combining several models to make one very reliable model.
- The most popular ensemble methods are boosting, bagging, and stacking.
- Ensemble methods are ideal for regression and classification, where they reduce bias and variance to boost the accuracy of models.

Main Types of Ensemble Methods

1. Bagging (Bootstrap Aggregating)

- **Concept**: Bagging involves creating multiple subsets of the original dataset through a process called bootstrapping. Bootstrapping randomly samples the data with replacement, creating diverse training sets for each model. Then, individual models (often decision trees) are trained on these subsets in parallel. The final prediction is obtained by averaging (for regression) or voting (for classification) the predictions of all individual models.

- **Benefits** The primary benefits of bagging include reducing overfitting, improving stability, and increasing accuracy. The diversity of training sets prevents models from relying on the same patterns, leading to better generalization.

- **Example**: Random Forest is a popular bagging algorithm that builds a collection of decision trees.

BAGGING Algorithm
### Bootstrap Aggrigating
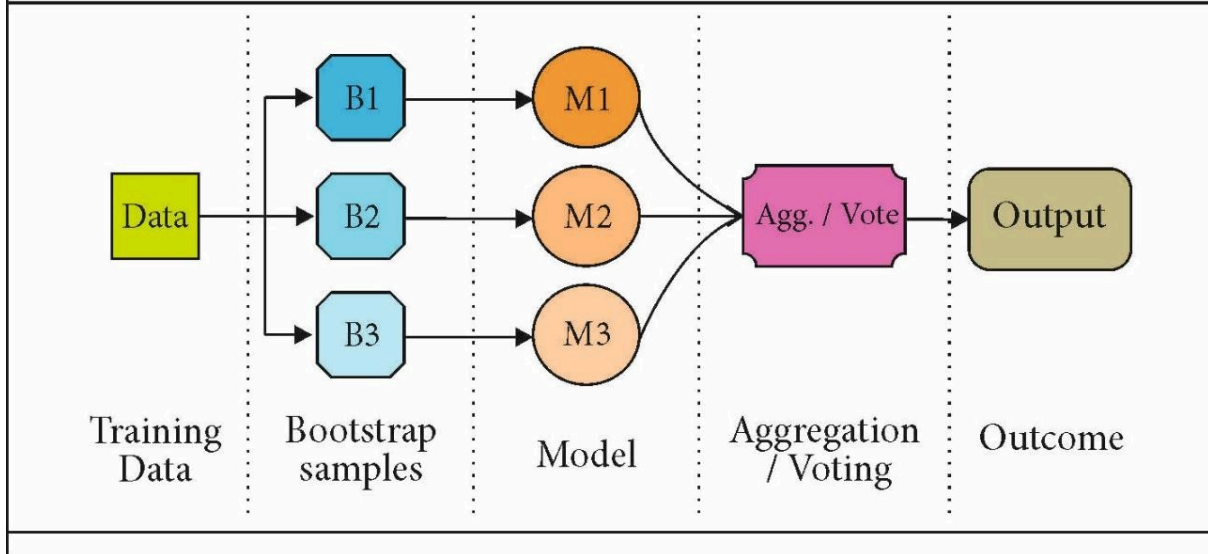
2. **Boosting**

- **Concept**: Boosting focuses on training models sequentially, where each new model attempts to correct the errors made by the previous ones. The models are weighted based on their performance, and the final prediction is a weighted sum of the predictions from all models.
- **Examples**:
  - **AdaBoost**: Adjusts the weights of incorrectly classified instances, giving them more importance in the next iteration.
  - **Gradient Boosting**: Builds models sequentially, each new model tries to reduce the residual error from the previous model. XGBoost and LightGBM are advanced implementations of this concept.

The Process of Boosting
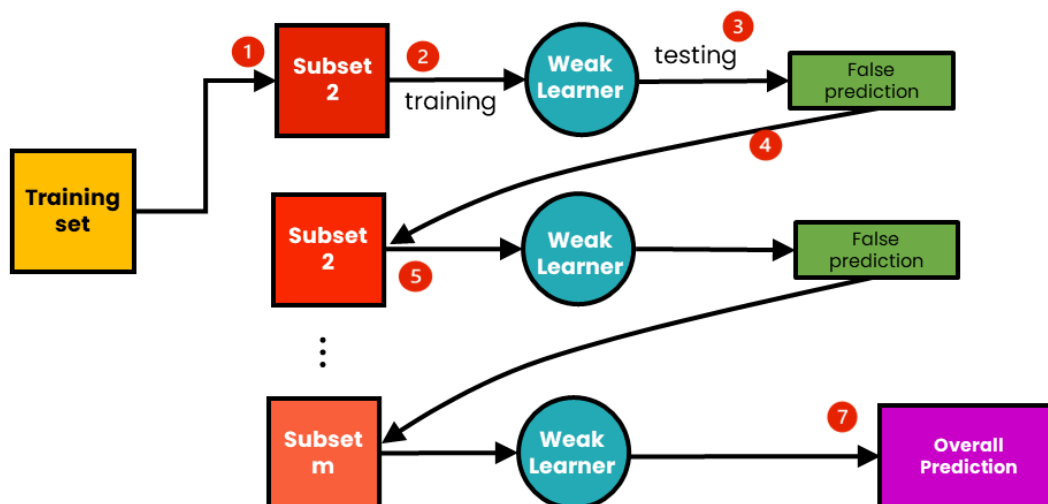
3. **Stacking (Stacked Generalization)**

- **Concept**: Stacking involves training multiple base models and then using their predictions as input features to a meta-model (a higher-level model), which makes the final prediction. Unlike bagging and boosting, stacking allows combining different types of models.
- **Example**: Combining a logistic regression model, a decision tree, and a support vector machine, with a meta-model like a neural network.

(A) Bagging   (B) Boosting   (C) Stacking
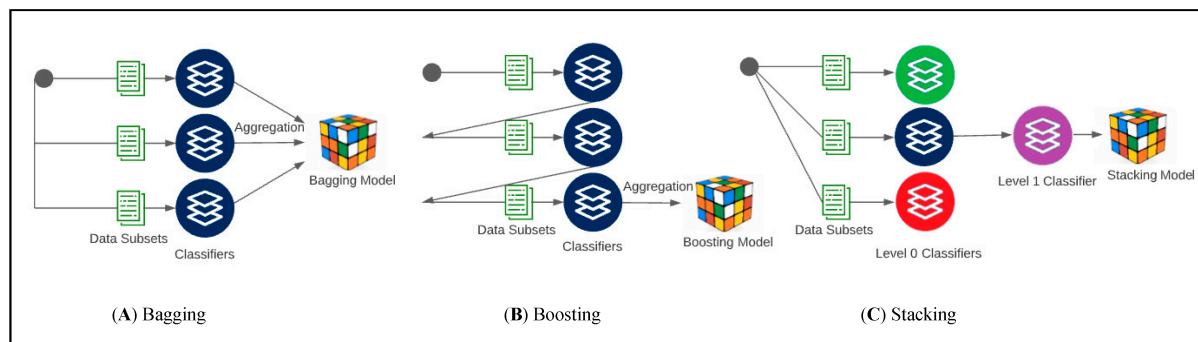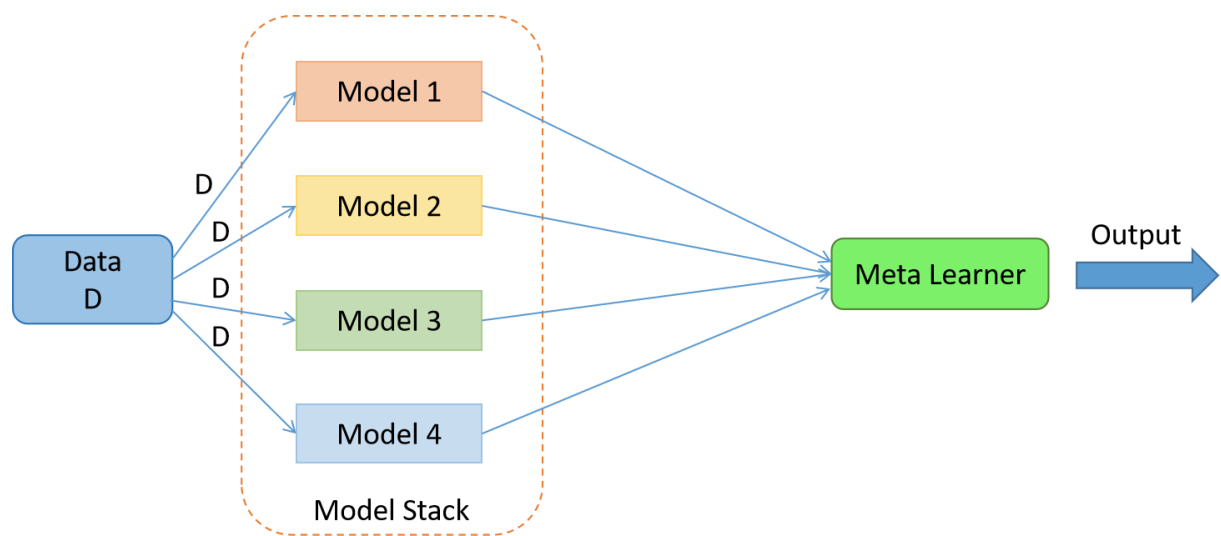
Advantages of Ensemble Methods:

- **Improved Accuracy**: By combining the predictions of multiple models, ensemble methods often result in better performance.
- **Reduced Overfitting**: Especially in bagging, the variance of the predictions is reduced, which helps in lowering the risk of overfitting.
- **Robustness**: Ensemble methods are typically more robust to outliers and noise in the data.

Disadvantages of Ensemble Methods:

- **Complexity**: They can be more computationally expensive and complex to implement, especially when combining different types of models.
- **Interpretability**: The resulting model is often less interpretable compared to a single model.