Ensemble models

By Shivam Sharma
Software engineer (AI&ML)

INTRODUCTION

Ensemble models are a powerful technique in machine learning that combine the predictions of multiple individual models to make more accurate and robust predictions. They are particularly effective when dealing with complex and challenging problems.

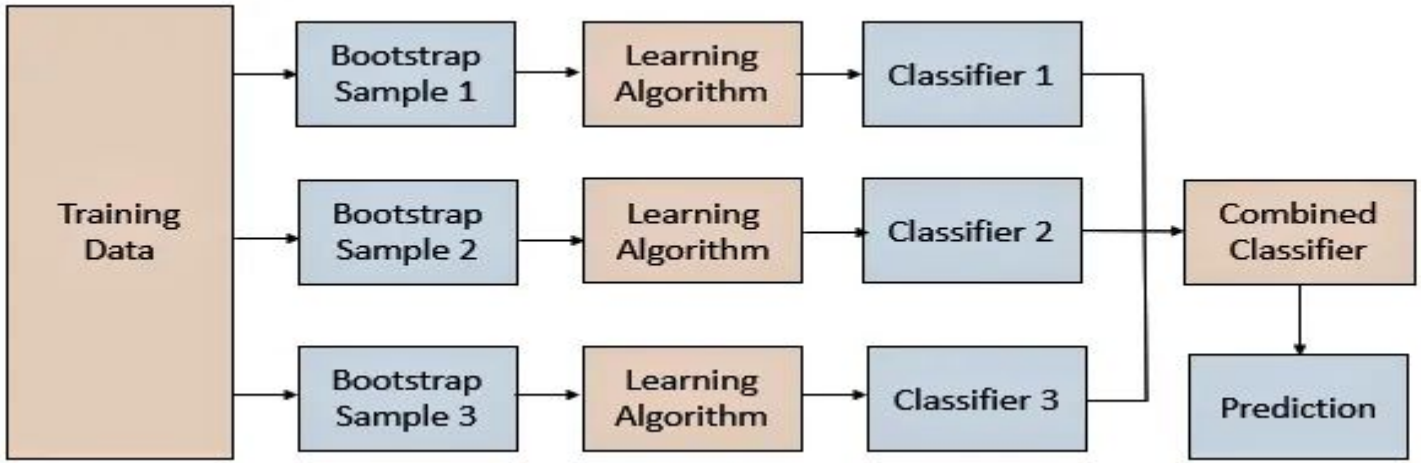
Types of Ensemble Models

- Bagging: Uses bootstrap sampling to create subsets of the data, trains multiple models on these subsets, and averages their predictions.

- Boosting: Trains models sequentially, where each subsequent model focuses on the examples misclassified by the previous models, giving more weight to those examples.

  Code : https://github.com/shivam-nsit19/ensemle_learning/tree/main

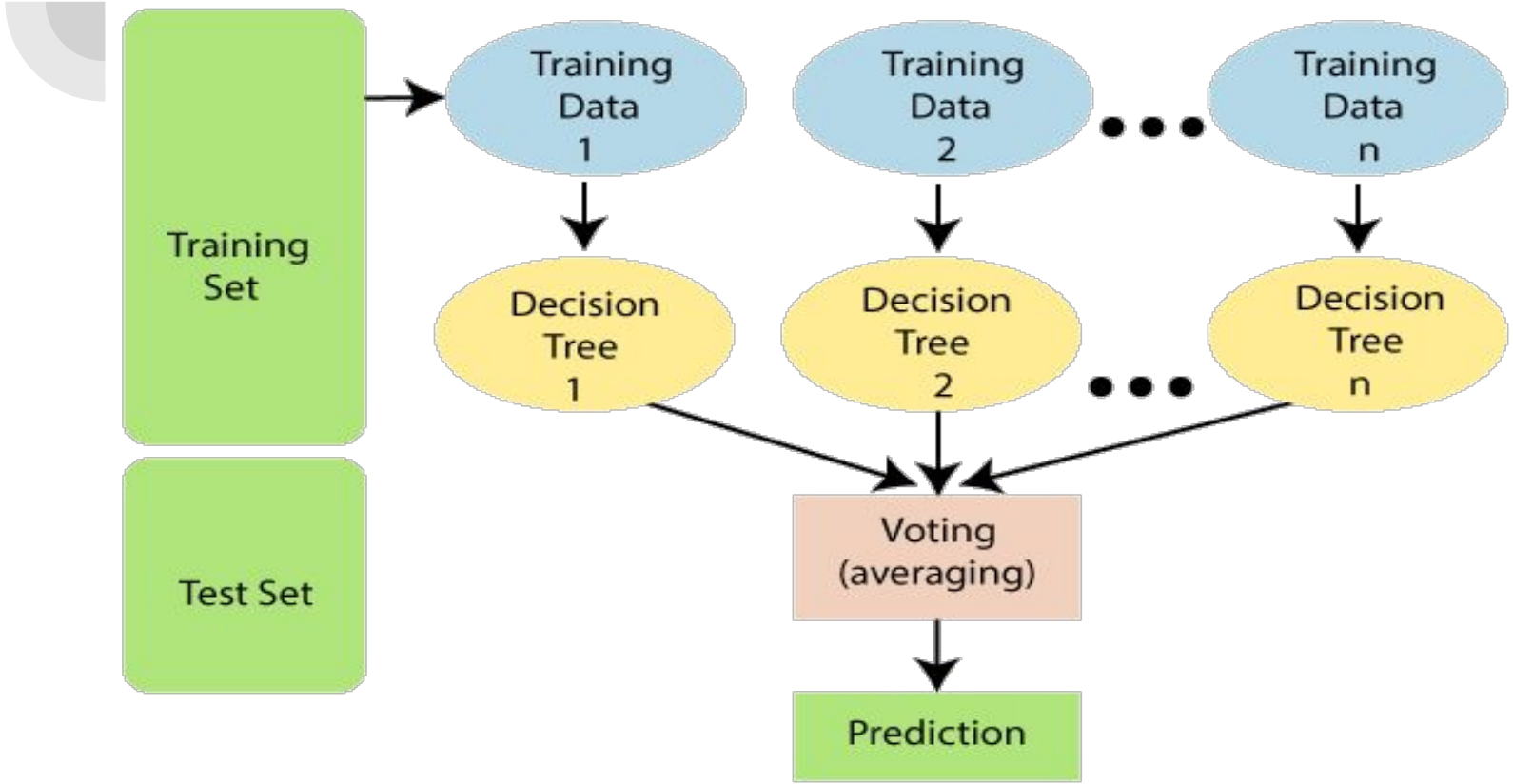
BAGGING (BOOTSTRAP AGGREGATION)



BOOSTING

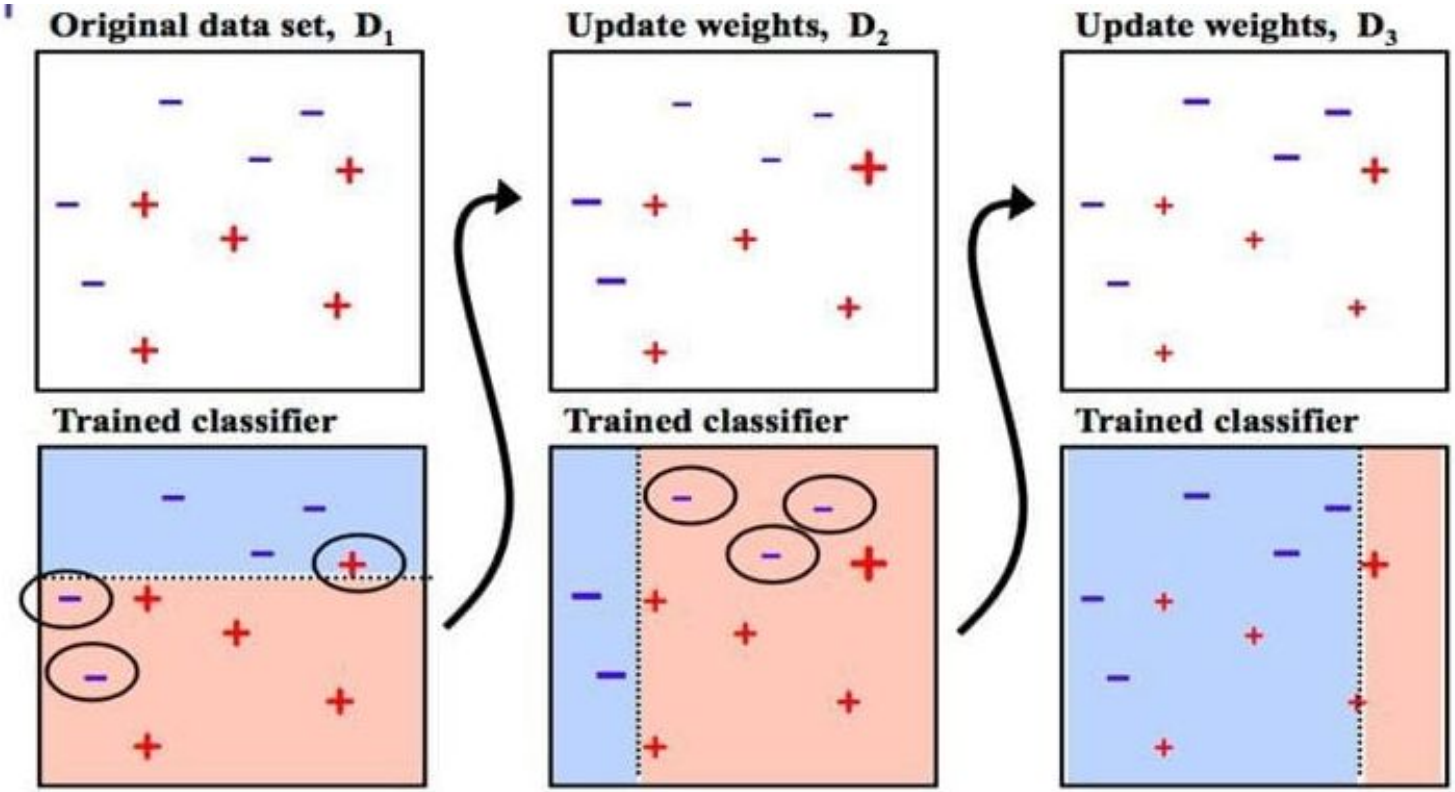
- Random Forest

Random Forest is an ensemble learning algorithm that combines multiple decision trees. It builds a forest of trees where each tree is trained on a random subset of features and votes to make predictions. It reduces overfitting and improves prediction accuracy through averaging or voting.


Training Set
Test Set
Training Data 1
Training Data 2
Training Data n
Decision Tree 1
Decision Tree 2
Decision Tree n
Voting (averaging)
Prediction

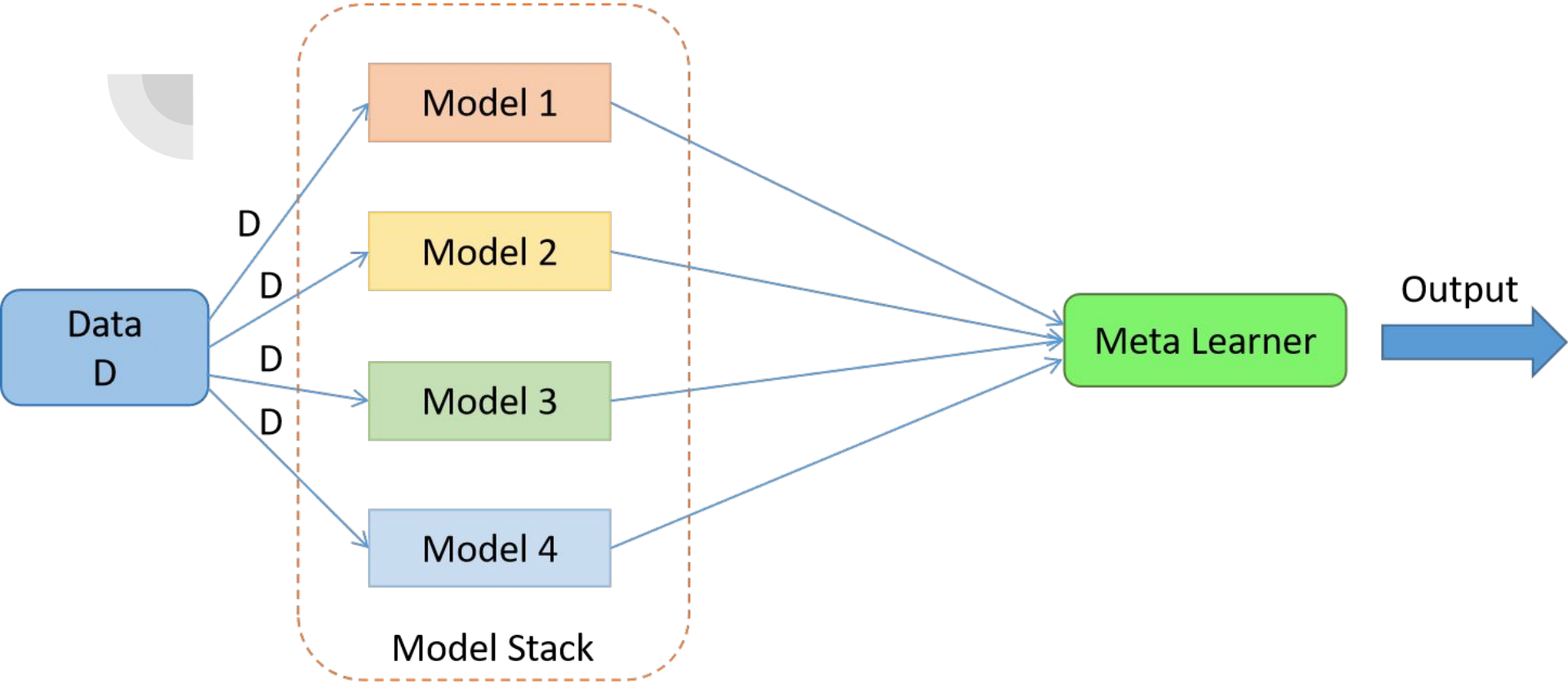
- ADAboost

AdaBoost is a boosting algorithm that combines multiple weak classifiers to create a strong classifier. It assigns higher weights to misclassified instances in each iteration, allowing subsequent models to focus on those examples. By iteratively refining the model, AdaBoost improves classification accuracy and handles complex relationships in the data.


Original data set, D_1
Update weights, D_2
Update weights, D_3
Trained classifier
Trained classifier
Trained classifier

Stacking

Stacking, also known as stacked generalization, is an ensemble learning technique that combines multiple base learners to make predictions. Unlike other ensemble methods that use simple aggregation of base learner predictions, stacking introduces a meta-learner


Data
D
D
D
D
D
Model 1
Model 2
Model 3
Model 4
Model Stack
Meta Learner
Output

Real world use cases

Finance and Investment:

- Ensemble models are used for portfolio optimization, asset pricing, credit risk assessment, fraud detection, and stock market prediction.
- By combining multiple models or strategies, financial institutions can make more accurate predictions, manage risk effectively, and improve investment decisions.

Healthcare and Biomedicine:

- Ensemble models are applied to medical image analysis, disease diagnosis, drug discovery, and patient prognosis.
- Combining multiple models can enhance the accuracy and reliability of medical diagnoses, improve treatment planning, and aid in identifying potential drug candidates.