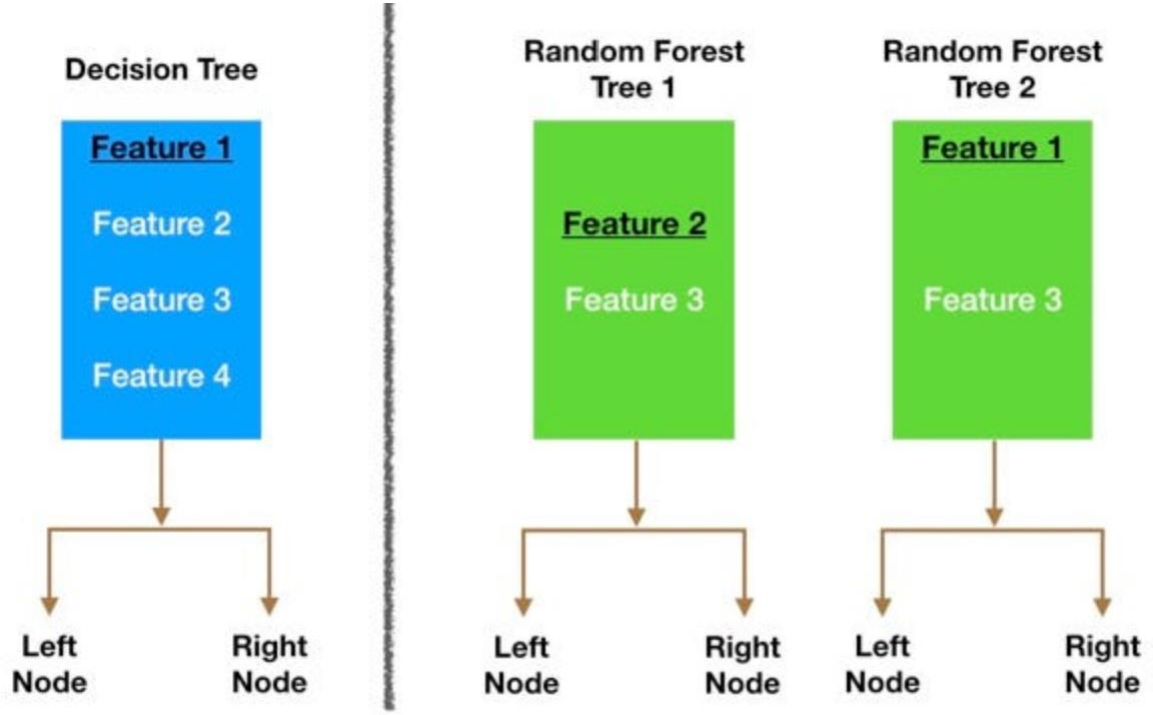
Random Forest

Presented by : **Eng. Doaa Fahmy**

The Random Forest Classifier

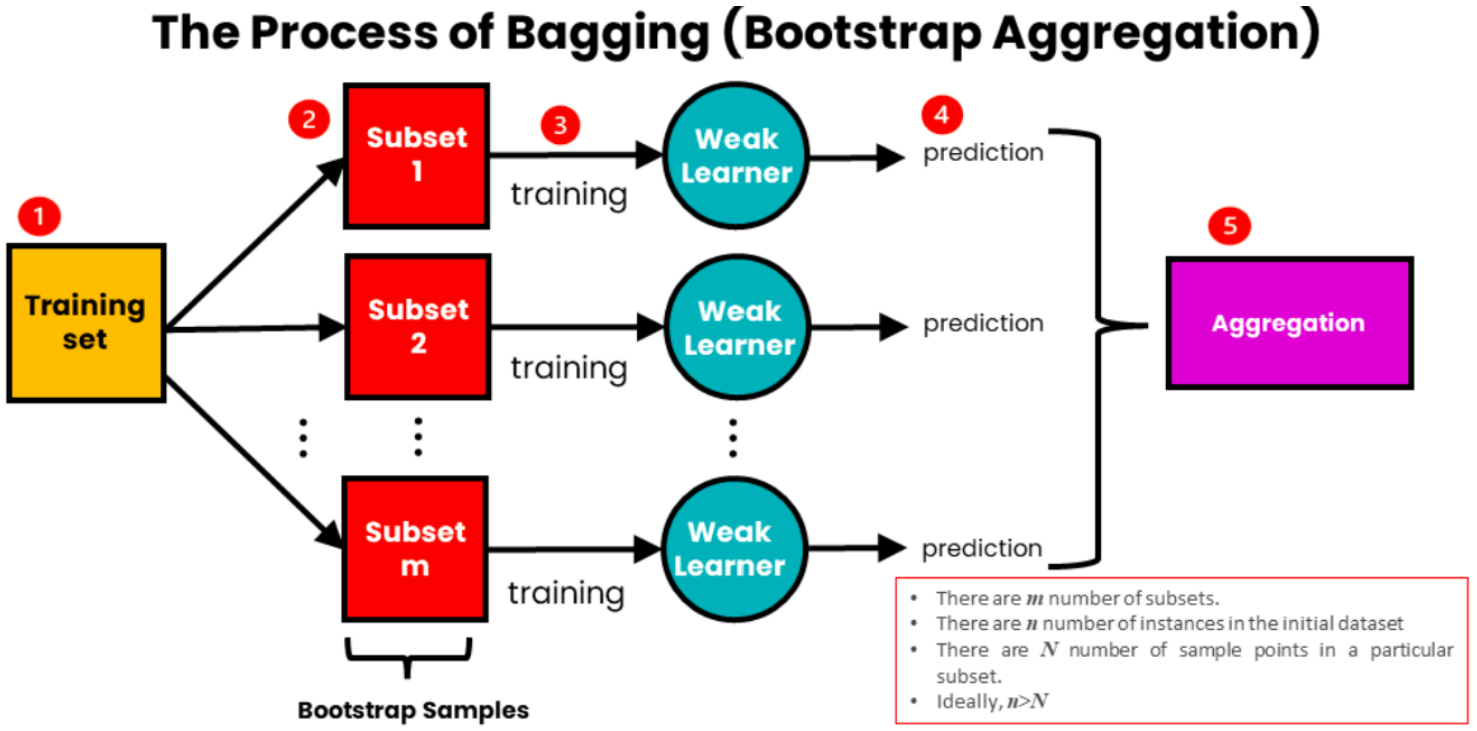
Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction (see figure below).



Node splitting in a random forest model is based on a random subset of features for each tree.

Bagging

bagging, also known as Bootstrap Aggregation, serves as the ensemble technique in the Random Forest algorithm. Here are the steps involved



Bagging

Selection of Subset: Bagging starts by choosing a random sample, or subset, from the entire dataset.

1.Bootstrap Sampling: Each model from these samples, called Bootstrap Samples, which we take from the original data with replacement. This process is known as row sampling.

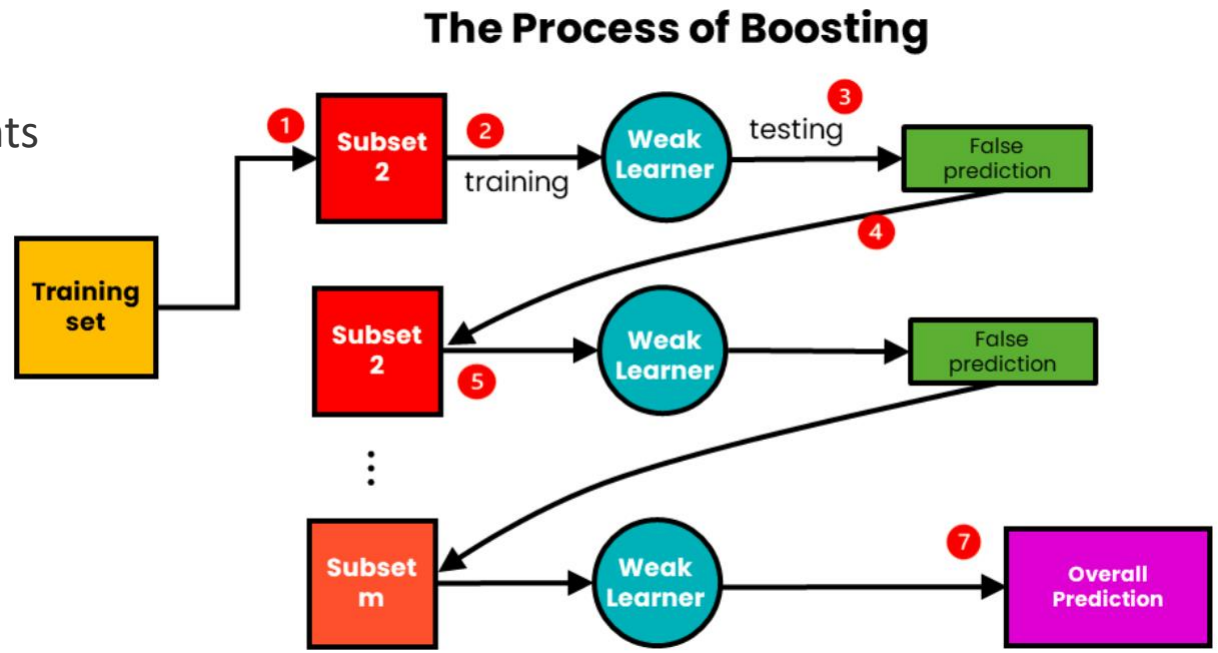
2.Bootstrapping: The step of row sampling with replacement is referred to as bootstrapping.

3.Independent Model Training: We train each model independently on its corresponding Bootstrap. This training process generates results for each model.

4.Majority Voting: The final output by combining the results of all models through majority voting. We select the most commonly predicted outcome among the models.

Boosting

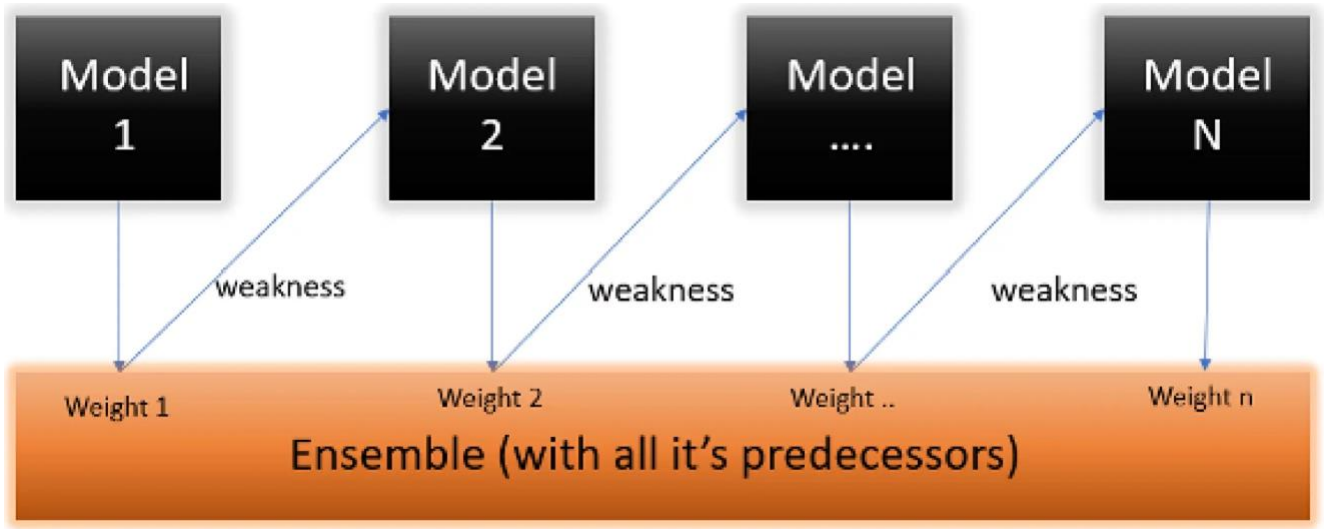
What this algorithm does is that it builds a model and gives equal weights to all the data points. It then assigns higher weights to points that are wrongly classified. Now all the points with higher weights are given more importance in the next model. It will keep training models until and unless a lower error is received.



- AdaBoost
- Gradient Boosting Machines (GBM)
- XGBoost
- LightGBM

Boosting has been proved in the literature to produce accurate results since it is a resilient strategy that quickly prevents overfitting. It is, however, sensitive to outliers because each classifier is required to correct the errors in the predecessors, and it is also difficult to scale up because each estimator bases its correctness on the previous predictors.

AdaBoosting



The main disadvantage of **Adaboost** is that it needs a quality dataset. Noisy data and outliers have to be avoided before adopting an **Adaboost** algorithm.

Adaboost is being used to classify text and images rather than binary classification problems

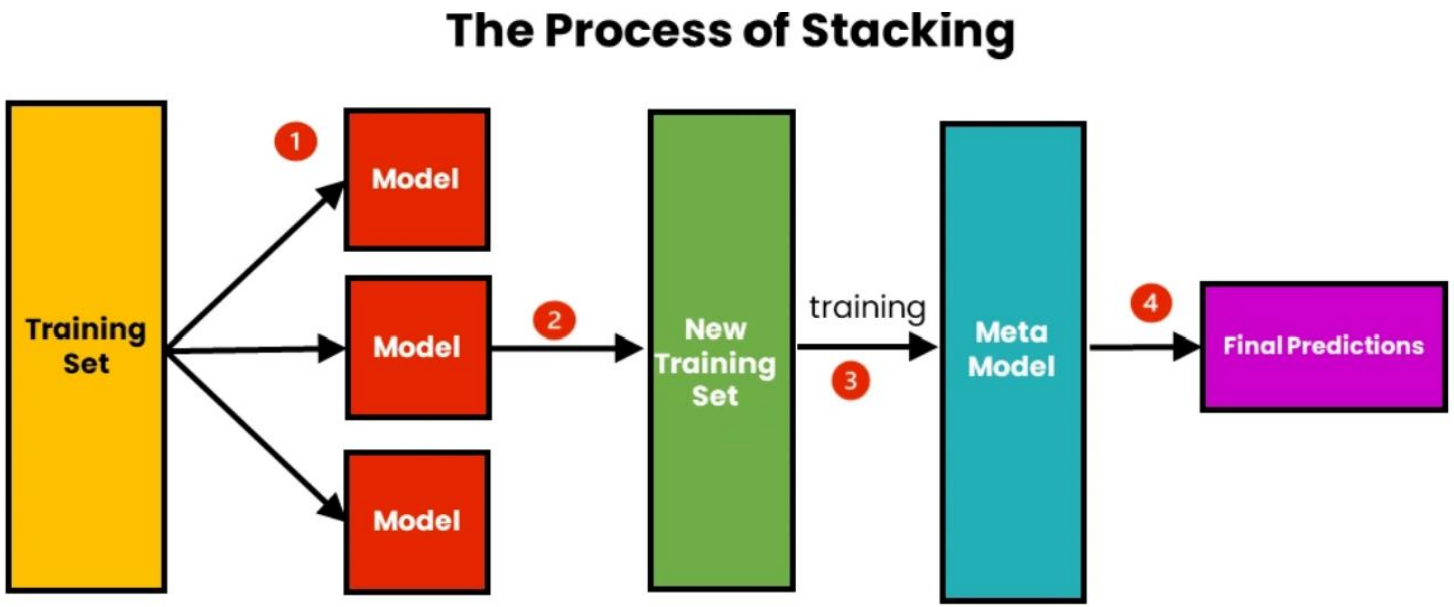
https://www.analyticsvidhya.com/blog/2021/09/adaboost-algorithm-a-complete-guide-for-beginners/

Stacking

We use stacking to improve the prediction accuracy of strong learners. Stacking aims to create a single robust model from multiple heterogeneous strong learners.

Stacking differs from bagging and boosting in machine learning in that:

It combines strong learners
It combines heterogeneous models
It consists of creating a Metamodel.



Stacking

(Stacked Generalization) is an ensemble learning technique that aims to combine multiple models to improve predictive performance. It involves the following steps:

Base Models: Training multiple models (level-0 models) on the same dataset.

Meta-Model: Training a new model (level-1 or meta-model) to combine the predictions of the base models. Using the predictions of the base models as input features for the meta-model.

The logistic regression is the meta-model for the stacking ensemble method. Since we are in a binary classification task, we used a 5-fold cross-validation for training.

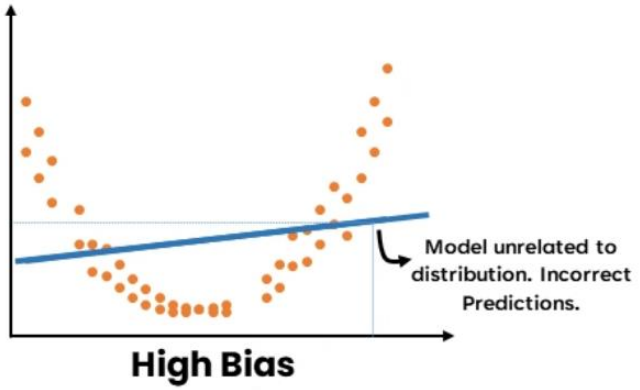
When to use Bagging vs Boosting vs Stacking?

	Bagging	Boosting	Stacking
Purpose	Reduce Variance	Reduce Bias	Improve Accuracy
Base Learner Types	Homogeneous	Homogeneous	Heterogeneous
Base Learner Training	Parallel	Sequential	Meta Model
Aggregation	Max Voting, Averaging	Weighted Averaging	Weighted Averaging

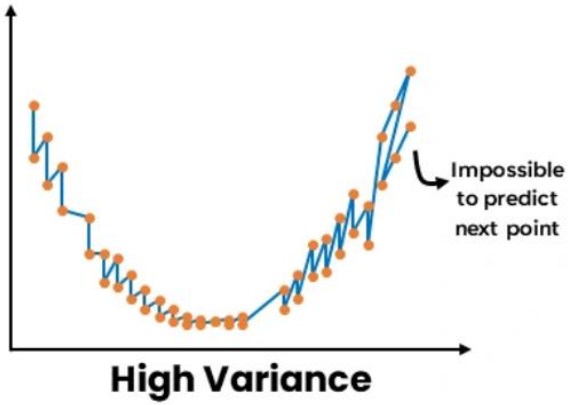
The variance is the difference between the model and the ground truth value, whereas the error is the outcome of sensitivity to tiny perturbations in the training set. Excessive bias might cause an algorithm to miss unique relationships between the intended outputs and the features (underfitting). There is a high variance in the algorithm that models random noise in the training data (overfitting).

High-bias and High-variance Models

•A high-bias model results from not learning data well enough. It is not related to the distribution of the data. Hence, future predictions will be unrelated to the data and thus incorrect.

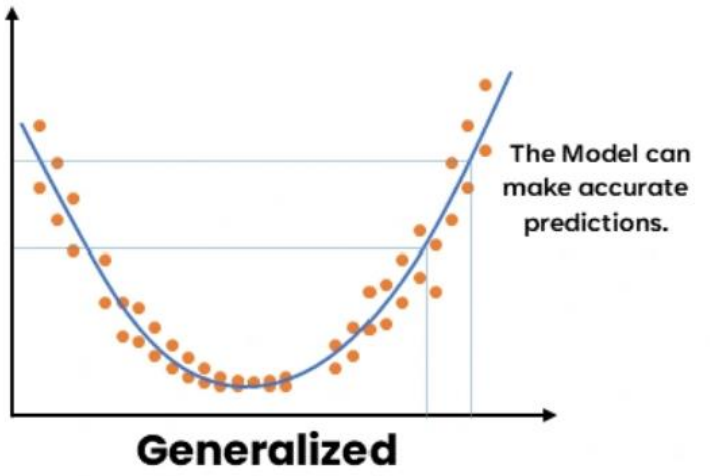


•A high-variance model results from learning the data too well. It varies with each data point, making it impossible to predict the next point accurately.



Thus, both high bias and high variance models cannot be generalized properly. Thus, weak learners will make incorrect generalizations or fail to generalize altogether. Because of this, the predictions of weak learners cannot be relied on by themselves.

As we know from the bias-variance trade-off, an underfit model has high bias and low variance, whereas an overfit model has high variance and low bias. In either case, there is no balance between bias and variance. For there to be a balance, both the bias and variance need to be low. Ensemble learning tries to balance this bias-variance trade-off by reducing either the bias or the variance.



Ensemble learning aims to reduce the bias if we have a weak model with high bias and low variance. This way, the resulting model will be much more balanced, with low bias and variance. Thus, the resulting model will be known as a strong learner. This model will be more generalized than the weak learners. It will thus be able to make accurate predictions.