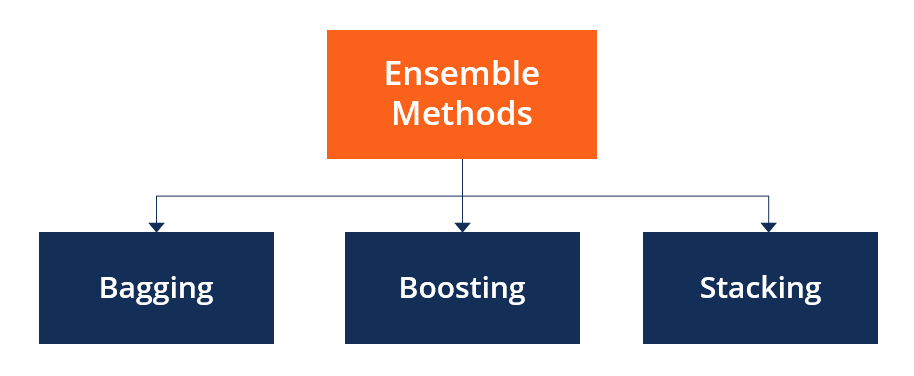
**Ensemble Methods**

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](https://corporatefinanceinstitute.com/course/learn-python/).



**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.

**Categories of Ensemble Methods**

Ensemble methods fall into two broad categories, i.e., sequential ensemble techniques and parallel ensemble techniques. **Sequential ensemble techniques** generate base learners in a sequence, e.g., Adaptive Boosting (AdaBoost). The sequential generation of base learners promotes the dependence between the base learners. The performance of the model is then improved by assigning higher weights to previously misrepresented learners.

In **parallel ensemble techniques**, base learners are generated in a parallel format, e.g., [random forest](https://corporatefinanceinstitute.com/resources/data-science/random-forest/). Parallel methods utilize the parallel generation of base learners to encourage independence between the base learners. The independence of base learners significantly reduces the error due to the application of averages.

The majority of ensemble techniques apply a single algorithm in base learning, which results in homogeneity in all base learners. Homogenous base learners refer to base learners of the same type, with similar qualities. Other methods apply heterogeneous base learners, giving rise to heterogeneous ensembles. Heterogeneous base learners are learners of distinct types.

**Main Types of Ensemble Methods**

**1. Bagging**

Bagging, the short form for bootstrap aggregating, is mainly applied in classification and [regression](https://corporatefinanceinstitute.com/resources/data-science/regression-analysis/). It increases the accuracy of models through decision trees, which reduces variance to a large extent. The reduction of variance increases accuracy, eliminating overfitting, which is a challenge to many predictive models.

Bagging is classified into two types, i.e., bootstrapping and aggregation. **Bootstrapping** is a sampling technique where samples are derived from the whole population (set) using the replacement procedure. The sampling with replacement method helps make the selection procedure randomized. The base learning algorithm is run on the samples to complete the procedure.

**Aggregation** in bagging is done to incorporate all possible outcomes of the prediction and randomize the outcome. Without aggregation, predictions will not be accurate because all outcomes are not put into consideration. Therefore, the aggregation is based on the probability bootstrapping procedures or on the basis of all outcomes of the predictive models.

Bagging is advantageous since weak base learners are combined to form a single strong learner that is more stable than single learners. It also eliminates any variance, thereby reducing the overfitting of models. One limitation of bagging is that it is computationally expensive. Thus, it can lead to more bias in models when the proper procedure of bagging is ignored.

**2. Boosting**

Boosting is an ensemble technique that learns from previous predictor mistakes to make better predictions in the future. The technique combines several weak base learners to form one strong learner, thus significantly improving the predictability of models. Boosting works by arranging weak learners in a sequence, such that weak learners learn from the next learner in the sequence to create better predictive models.

Boosting takes many forms, including gradient boosting, Adaptive Boosting (AdaBoost), and XGBoost (Extreme Gradient Boosting). [AdaBoost](https://medium.com/@desardaakash/understanding-adaboost-2f94f22d5bfe) uses weak learners in the form of decision trees, which mostly include one split that is popularly known as decision stumps. AdaBoost’s main decision stump comprises observations carrying similar weights.

[Gradient boosting](https://corporatefinanceinstitute.com/resources/data-science/gradient-boosting/) adds predictors sequentially to the ensemble, where preceding predictors correct their successors, thereby increasing the model’s accuracy. New predictors are fit to counter the effects of errors in the previous predictors. The gradient of descent helps the gradient booster identify problems in learners’ predictions and counter them accordingly.

XGBoost makes use of decision trees with boosted gradient, providing improved speed and performance. It relies heavily on the computational speed and the performance of the target model. Model training should follow a sequence, thus making the implementation of gradient boosted machines slow.

**3. Stacking**

Stacking, another ensemble method, is often referred to as stacked generalization. This technique works by allowing a training algorithm to ensemble several other similar learning algorithm predictions. Stacking has been successfully implemented in regression, density estimations, distance learning, and classifications. It can also be used to measure the error rate involved during bagging.

**Variance Reduction**

Ensemble methods are ideal for reducing the variance in models, thereby increasing the accuracy of predictions. The variance is eliminated when multiple models are combined to form a single prediction that is chosen from all other possible predictions from the combined models. An ensemble of models combines various models to ensure that the resulting prediction is the best possible, based on the consideration of all predictions.