Ensemble BREAST CANCER CLASSIFICATION

ENSEMBLE MODEL

INTRODUCTION

When you want to purchase a new car, will you walk up to the first car shop and purchase one based on the advice of the dealer? It's highly unlikely.

WHAT WOULD YOU LIKE?

- Browse a few web portals
- Friends' reviews

In short, you will instead make a decision considering the opinions of other people as well.

ENSEMBLE MODELS:

- Similar idea
- Combine different models' decision to improve performance.

ENSEMBLE LEARNING EXAMPLE

Suppose you are a movie director and you have created a short movie on a very important and interesting topic. Now, you want to take preliminary feedback (ratings) on the movie before making it public. What are the possible ways by which you can do that?

A: *You may ask one of your friends to rate the movie for you.*

B: *Another way could be by asking 5 colleagues of yours to rate the movie.*

C: *How about asking 50 people to rate the movie?*

The responses, in the last case, would be more generalized and diversified since now you have people with different sets of skills. And as it turns out – this is a better approach to get honest ratings than the previous cases we saw.

Similar is true for a diverse set of models in comparison to single models. This diversification in Machine Learning is achieved by a technique called Ensemble Learning.

ENSEMBLE LEARNING

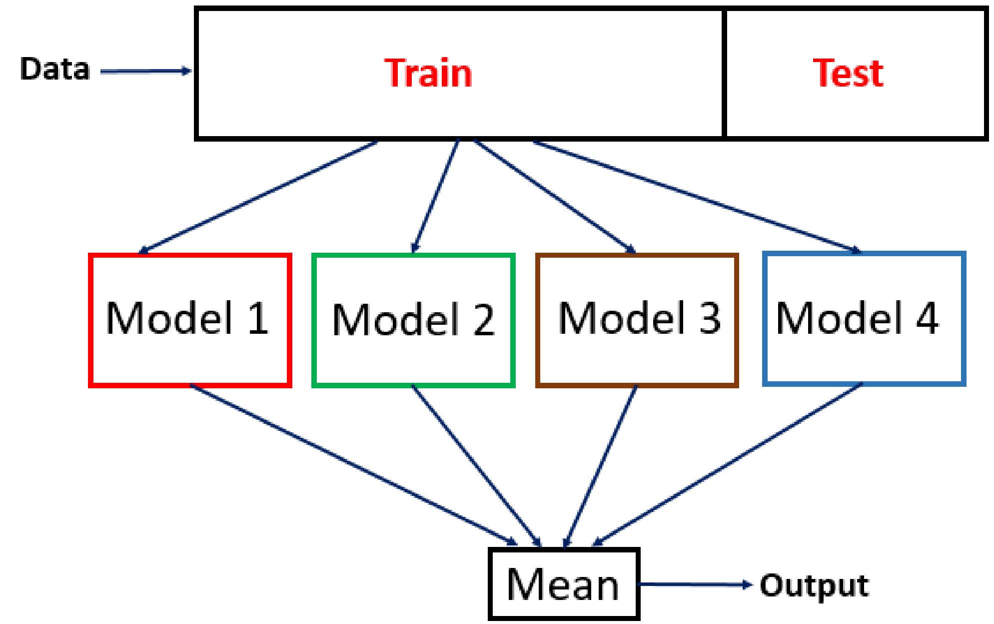
1. 1.Ensemble learning methods is applied to regression as well as classification.
2. 2. Ensemble learning for regression creates multiple repressors i.e. multiple regression models such as linear, polynomial, etc.
3. 3. Ensemble learning for classification creates multiple classifiers i.e. multiple classification models such as logistic, decision tress, KNN, SVM, etc.
4. 4. Ensemble Learning is a technique that create multiple models and then combine them to produce improved results.
5. 5. Ensemble learning usually produces more accurate solutions than a single model would.

Do It Skills
do it. enjoy it.

ENSEMBLE LEARNING

There are two steps in ensemble learning:

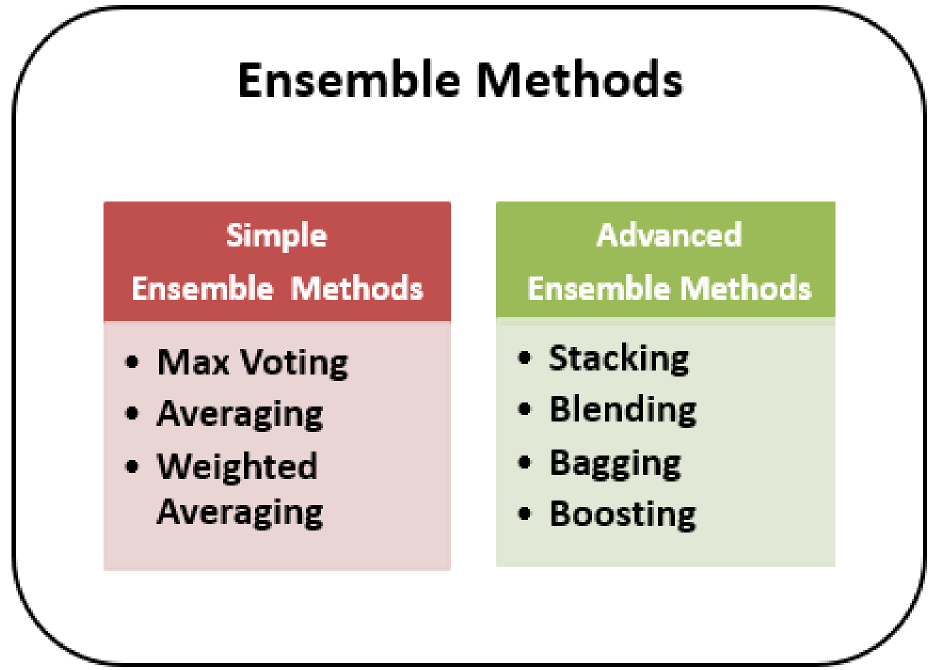
1. Multiples machine learning models were generated using same or different machine learning algorithm. These are called “base models”.
2. The prediction performed on the basis of base models.



ENSEMBLE LEARNING TECHNIQUES

Basic/Simple Ensemble Techniques

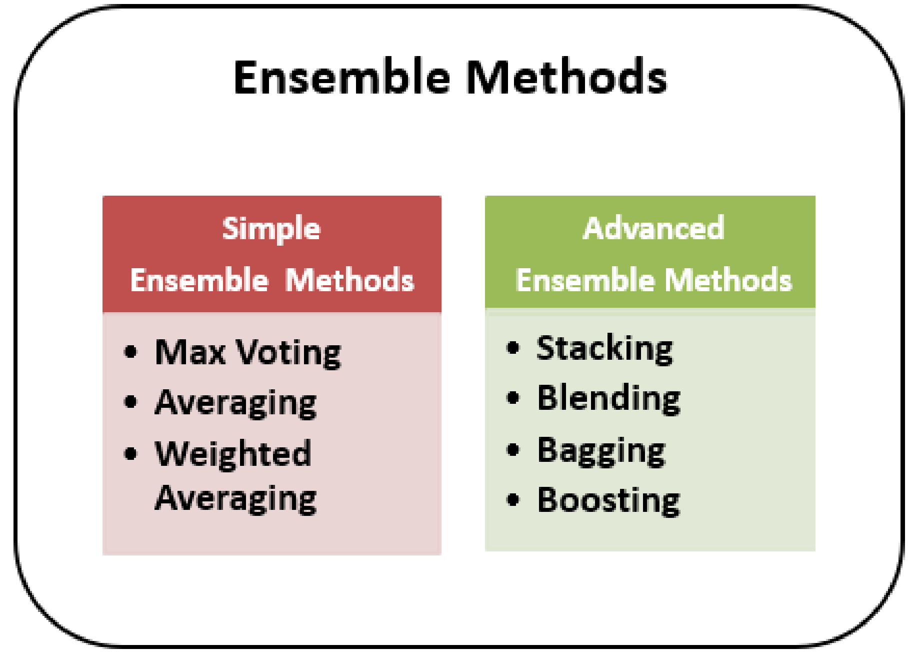
- Max Voting
- Averaging
- Weighted Average



ENSEMBLE LEARNING TECHNIQUES

Advanced Ensemble Techniques

- Stacking
- Blending
- Bagging
- Boosting

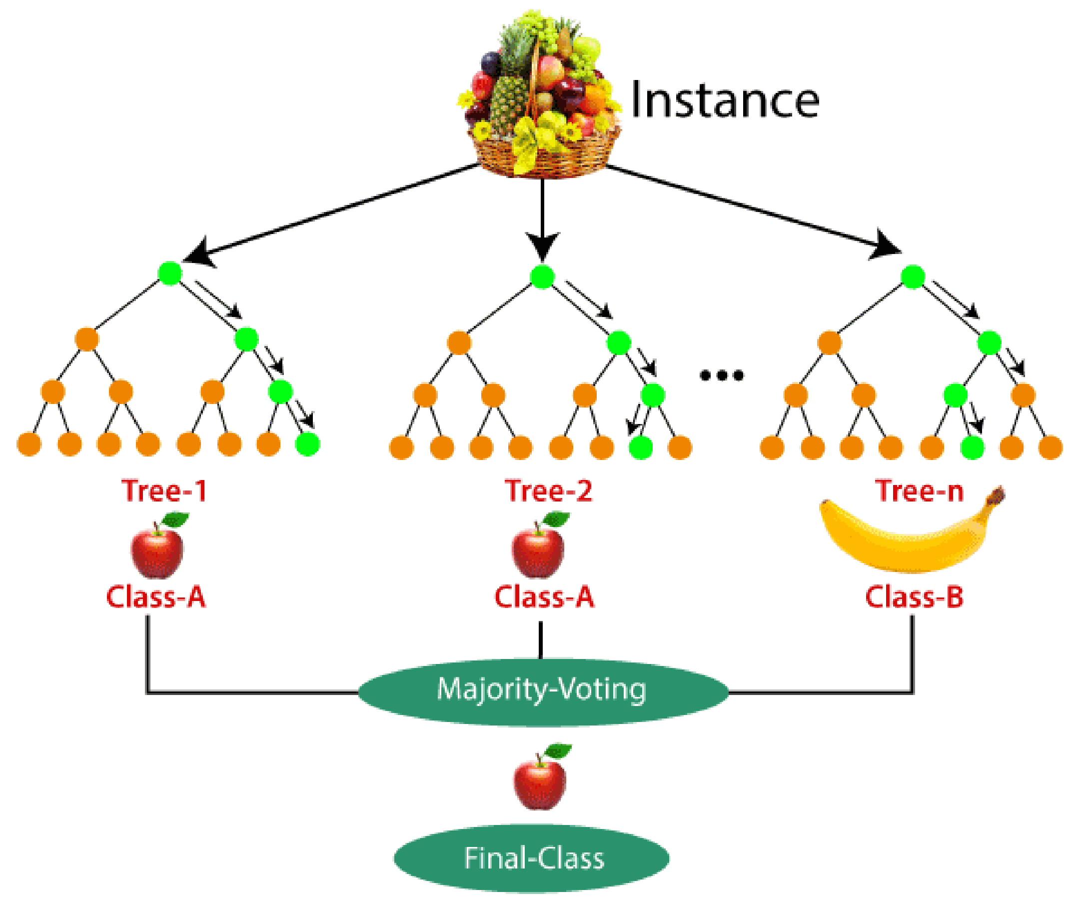


Do It Skills
do it. enjoy it.

ENSEMBLE LEARNING TECHNIQUES/METHODS

Max Voting

- Generally used for classification problems.
- Multiple models are used to make predictions for each data point.
- Considered as a ‘vote’.
- Majority predictions are used as the final prediction.



Do It Skills
do it. enjoy it.

ENSEMBLE LEARNING TECHNIQUES/METHODS

Max Voting

For example, when you asked 5 of your colleagues to rate your movie (out of 5); we'll assume three of them rated it as 4 while two of them gave it a 5. Since the majority gave a rating of 4, the final rating will be taken as 4. **You can consider this as taking the mode of all the predictions.**

The result of max voting would be something like this:

Colleague 1	Colleague 2	Colleague 3	Colleague 4	Colleague 5	Final rating
5	4	5	4	4	4

ENSEMBLE LEARNING TECHNIQUES/METHODS

Averaging

- Similar to the max voting technique
- Multiple predictions are made for each data point
- In this method, we take an average of predictions from all the models and use it to make the final prediction
- Used for making predictions in regression problems or while calculating probabilities for classification problems.

ENSEMBLE LEARNING TECHNIQUES/METHODS

Averaging

- For example, in the below case, the averaging method would take the average of all the values.

i.e. (5+4+5+4+4)/5 = 4.4

Colleague 1	Colleague 2	Colleague 3	Colleague 4	Colleague 5	Final rating
5	4	5	4	4	4.4

ENSEMBLE LEARNING TECHNIQUES/METHODS

Weighted Average

- An extension of the averaging method.
- All models are assigned different weights defining the importance of each model for prediction.
- For instance, if two of your colleagues are critics, while others have no prior experience in this field, then the answers by these two friends are given more importance as compared to the other people.

ENSEMBLE LEARNING TECHNIQUES/METHODS

Weighted Average

- The result is calculated as [(5*0.23) + (4*0.23) + (5*0.18) + (4*0.18) + (4*0.18)] = 4.41.

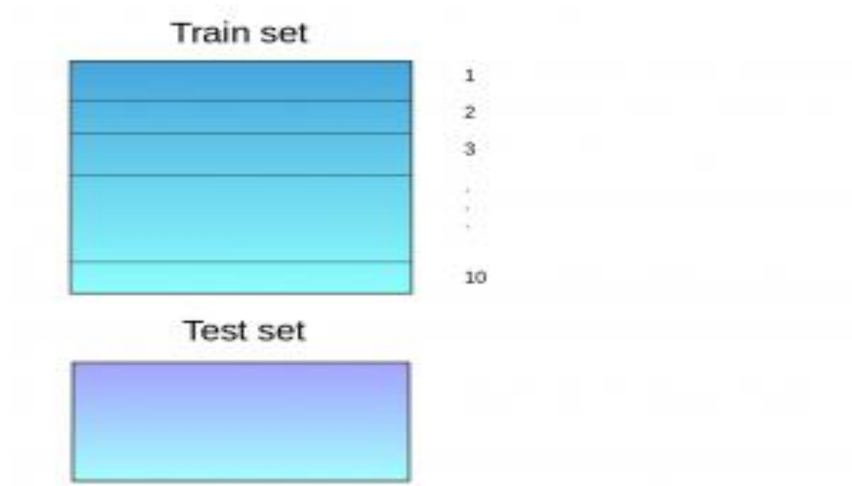
	Colleague 1	Colleague 2	Colleague 3	Colleague 4	Colleague 5	Final rating
weight	0.23	0.23	0.18	0.18	0.18	
rating	5	4	5	4	4	4.41

ENSEMBLE LEARNING TECHNIQUES/METHODS

Stacking

Stacking is an ensemble learning technique that uses predictions from multiple models (for example decision tree, knn or svm) to build a new model. This model is used for making predictions on the test set. Below is a step-wise explanation for a simple stacked ensemble:

1. The train set is split into 10 parts.

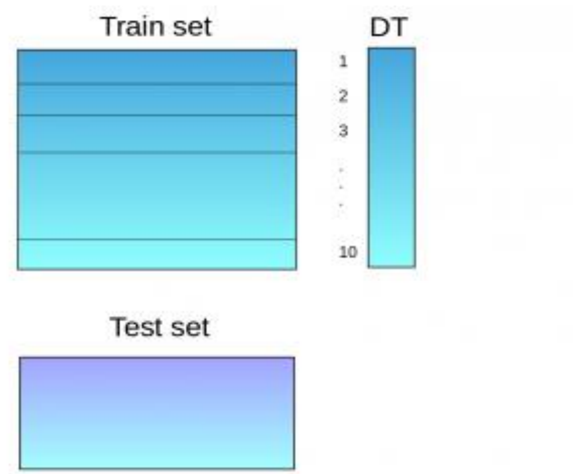


Do It Skills
do it. enjoy it.

ENSEMBLE LEARNING TECHNIQUES/METHODS

Stacking

2. A base model (suppose a decision tree) is fitted on 9 parts and predictions are made for the 10th part. This is done for each part of the train set.

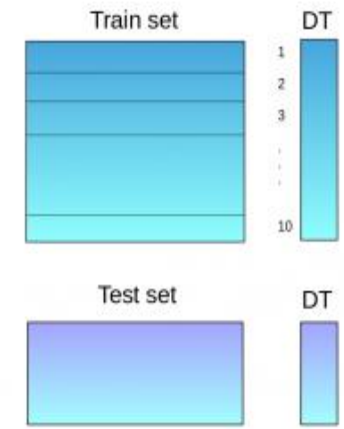


3.The Base model (in this case, decision tree) is then fitted on the whole train dataset.

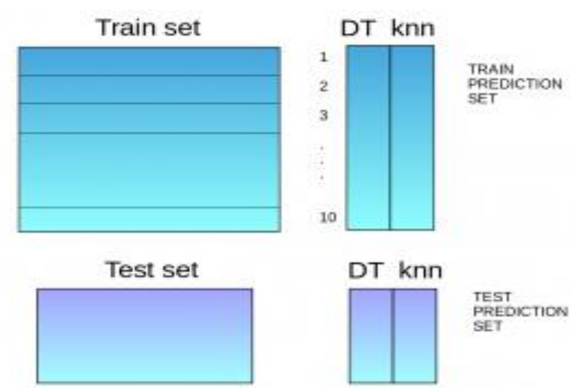
ENSEMBLE LEARNING TECHNIQUES/METHODS

Stacking

4. Using this model, predictions are made on the test set.



5.steps 2 to 4 are repeated for another base model (say knn) resulting in another set of predictions for the train set and test set.

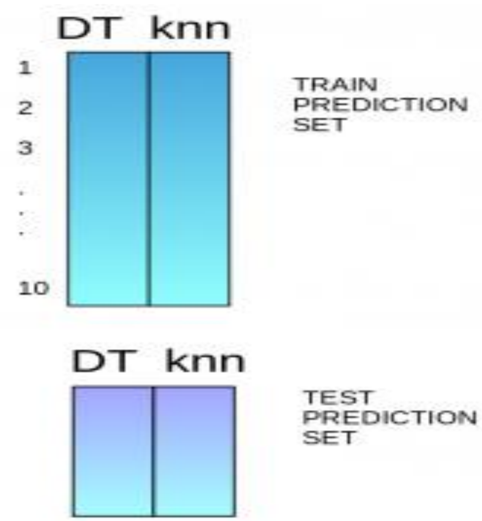


Do It Skills
do it. enjoy it.

ENSEMBLE LEARNING TECHNIQUES/METHODS

Stacking

6. The predictions from the train set are used as features to build a new model



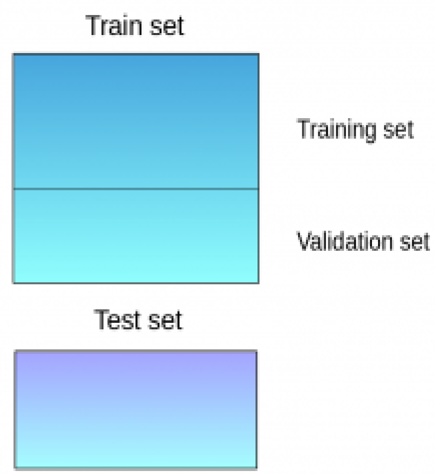
7. This model is used to make final predictions on the test prediction set.

ENSEMBLE LEARNING TECHNIQUES/METHODS

Blending

Blending follows the same approach as stacking but uses only a holdout (validation) set from the train set to make predictions. In other words, unlike stacking, the predictions are made on the holdout set only. The holdout set and the predictions are used to build a model which is run on the test set. Here is a detailed explanation of the blending process:

1. The train set is split into training and validation sets.

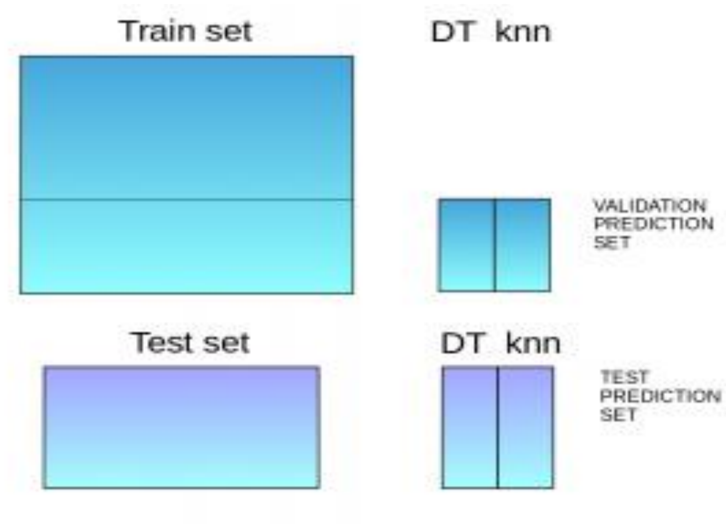


ENSEMBLE LEARNING TECHNIQUES/METHODS

Blending

2. Model(s) are fitted on the training set.

3. The predictions are made on the validation set and the test set



ENSEMBLE LEARNING TECHNIQUES/METHODS

Blending

4. The validation set and its predictions are used as features to build a new model.

5. This model is used to make final predictions on the test and meta-features.

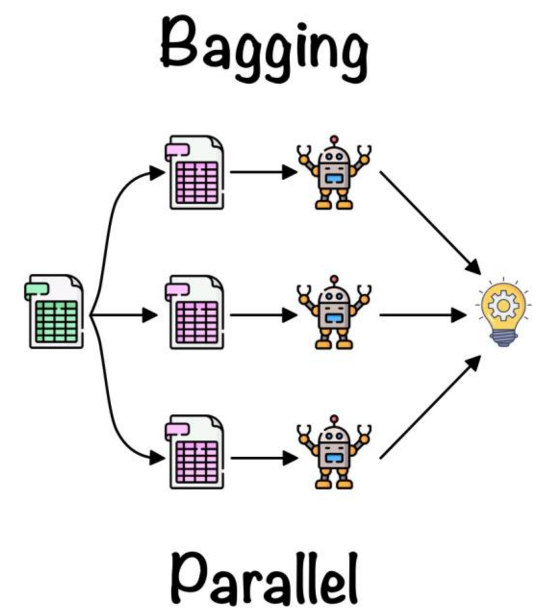
ENSEMBLE LEARNING TECHNIQUES/METHODS

Bagging

- Combining the results of multiple models (for instance, all decision trees) to get a generalized result.
- Here's a question: If you create all the models on the same set of data and combine it, will it be useful? There is a high chance that these models will give the same result since they are getting the same input. So how can we solve this problem? One of the techniques is bootstrapping.

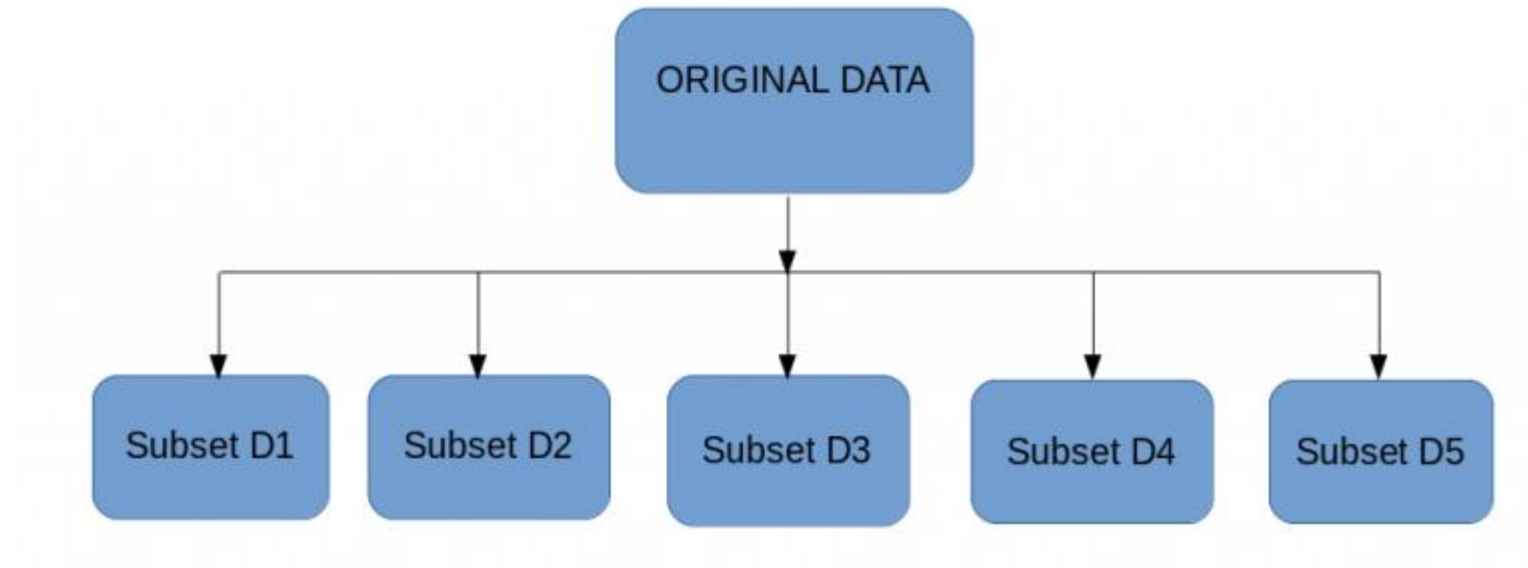
Bootstrapping : sampling technique in which we create subsets of observations from the original dataset, **with replacement**. The size of the subsets is the same as the size of the original set.

Bagging (or Bootstrap Aggregating) technique uses these subsets (bags) to get a fair idea of the distribution (complete set). The size of subsets created for bagging may be less than the original set.



ENSEMBLE LEARNING TECHNIQUES/METHODS

Bagging

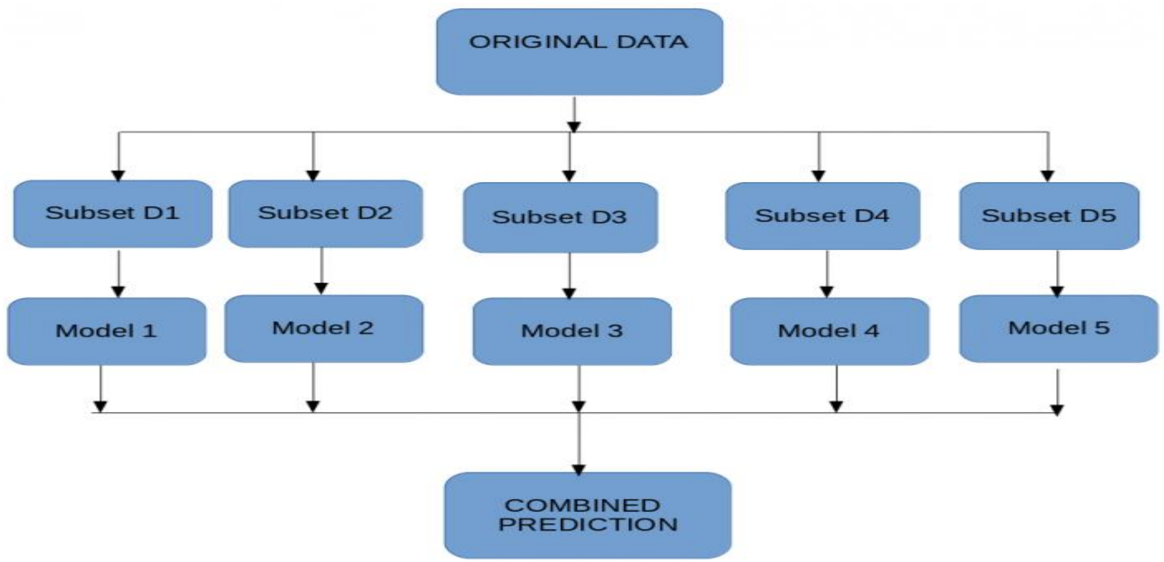


Do It Skills
do it. enjoy it.

ENSEMBLE LEARNING TECHNIQUES/METHODS

Bagging

1. Multiple subsets are created from the original dataset, selecting observations with replacement.
2. A base model (weak model) is created on each of these subsets.
3. The models run in parallel and are independent of each other.
4. The final predictions are determined by combining the predictions from all the models.



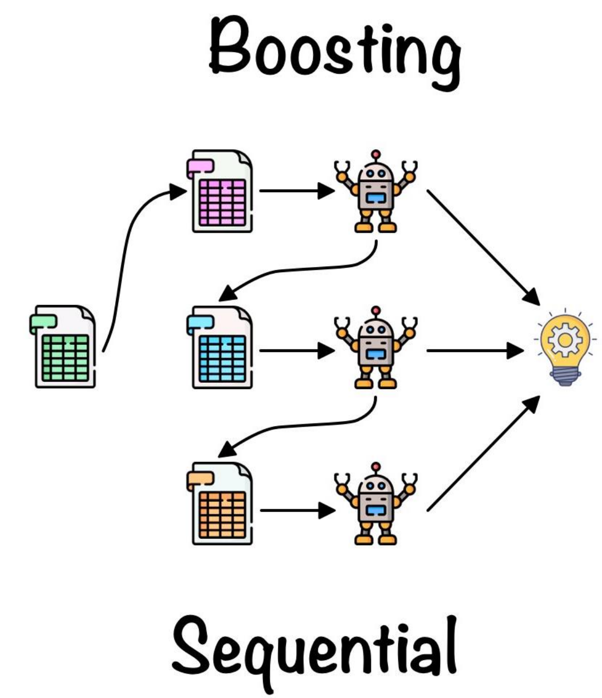
Do It Skills
do it. enjoy it.

ENSEMBLE LEARNING TECHNIQUES/METHODS

Boosting

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.

1. A subset is created from the original dataset.
2. Initially, all data points are given equal weights.
3. A base model is created on this subset.
4. This model is used to make predictions on the whole dataset



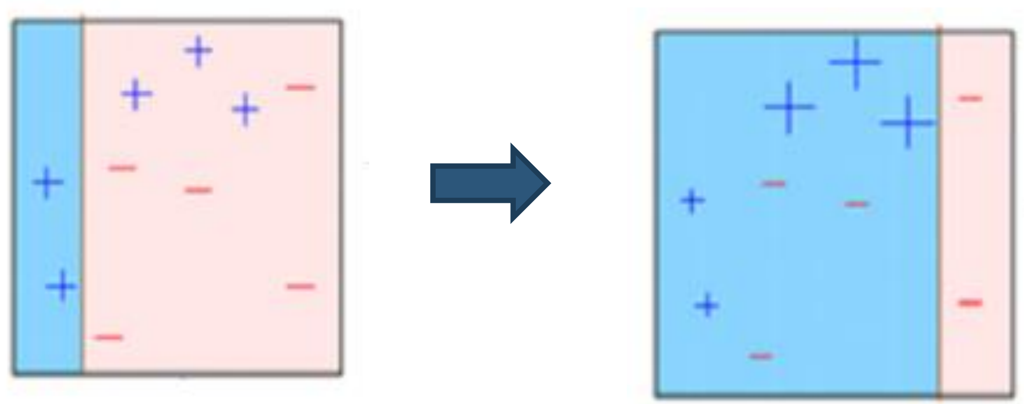
ENSEMBLE LEARNING TECHNIQUES/METHODS

Boosting

5. Errors are calculated using the actual values and predicted values.

6. The observations which are incorrectly predicted, are given higher weights.
(Here, the three misclassified blue-plus points will be given higher weights)

7. Another model is created and predictions are made on the dataset.
(This model tries to correct the errors from the previous model)

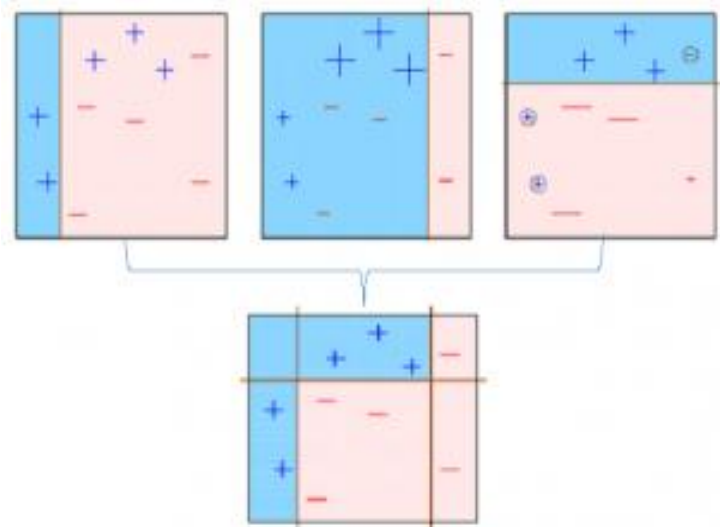


ENSEMBLE LEARNING TECHNIQUES/METHODS

Boosting

8. Similarly, multiple models are created, each correcting the errors of the previous model.

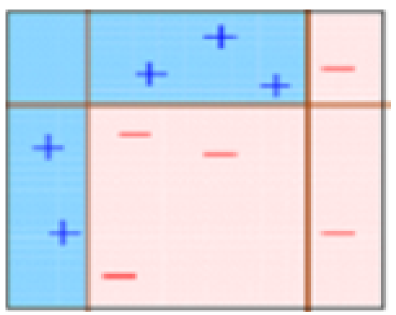
9. The final model (strong learner) is the weighted mean of all the models (weak learners)



ENSEMBLE LEARNING TECHNIQUES/METHODS

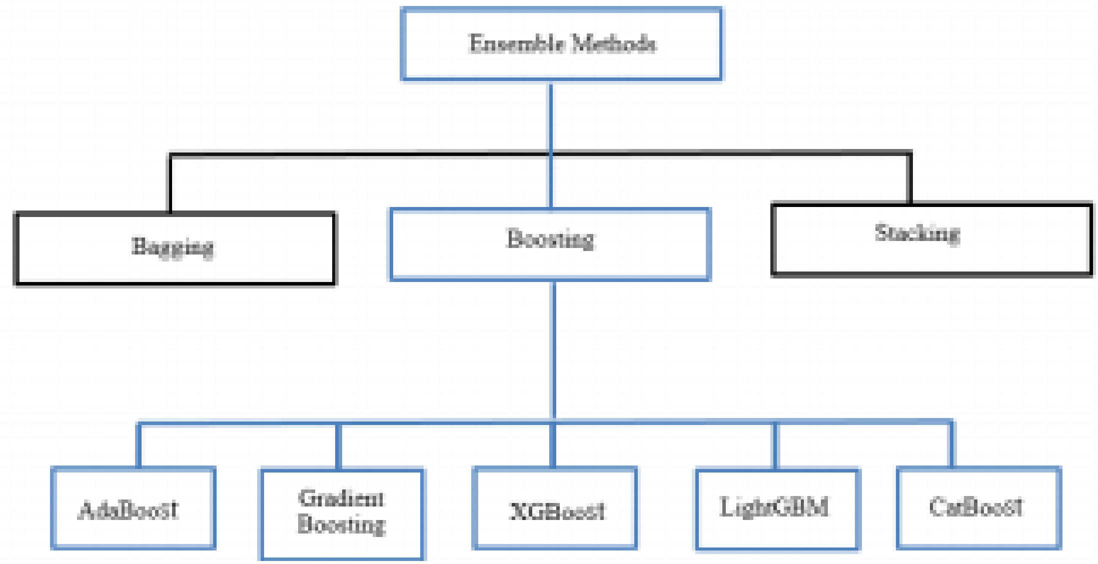
Boosting

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.



ALGORITHMS BASED ON BAGGING AND BOOSTING

- Bagging meta-estimator
- Random Forest
- AdaBoost
- GBM
- XGB
- Light GBM
- CatBoost



Do It Skills
do it. enjoy it.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Bagging and Boosting are two of the most commonly used techniques in machine learning. In this section, we will look at them in detail. Following are the algorithms we will be focusing on:

Bagging algorithms:

- Bagging meta-estimator
- Random forest

Boosting algorithms:

- AdaBoost
- GBM
- XGBM
- Light GBM
- CatBoost

ALGORITHMS BASED ON BAGGING AND BOOSTING

Bagging meta-estimator

- An ensembling algorithm
- Can be used for both classification (BaggingClassifier) and regression (BaggingRegressor) problems.
- Follows the typical bagging technique to make predictions.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Bagging meta-estimator

- Following are the steps for the bagging meta-estimator algorithm:

1. Random subsets are created from the original dataset (Bootstrapping).
2. The subset of the dataset includes all features.
3. A user-specified base estimator is fitted on each of these smaller sets.
4. Predictions from each model are combined to get the final result.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Bagging meta-estimator

Parameters used in the algorithms:

- **base_estimator**:
- **n_estimators**:
- **max_samples**:
- **max_features**:
- **n_jobs**:
- **random_state**:

Do It Skills
do it. enjoy it.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Random Forest

- Random Forest is another ensemble machine learning algorithm that follows the bagging technique.
- It is an extension of the bagging estimator algorithm.
- The base estimators : decision trees.
- It randomly selects a set of features which are used to decide the best split at each node of the decision tree.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Random Forest

Looking at it step-by-step, this is what a random forest model does:

1. Random subsets are created from the original dataset (bootstrapping).
2. At each node in the decision tree, only a random set of features are considered to decide the best split.
3. A decision tree model is fitted on each of the subsets.

   The final prediction is calculated by averaging the predictions from all decision trees.

   To sum up, Random forest **randomly** selects data points and features, and builds **multiple trees (Forest)**

ALGORITHMS BASED ON BAGGING AND BOOSTING

Random Forest

Parameters

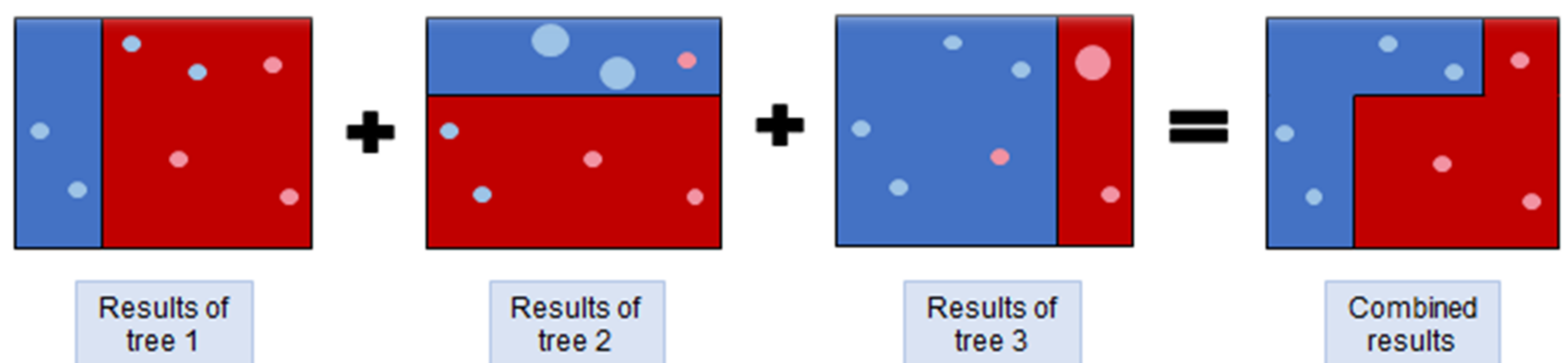
- **n_estimators:**
- **criterion**:
- **max_features** :
- **max_depth**:
- **min_samples_split:**
- **min_samples_leaf:**
- **max_leaf_nodes:**
- **n_jobs**:
- **random_state**:

Do It Skills
do it. enjoy it.

ALGORITHMS BASED ON BAGGING AND BOOSTING

AdaBoost

- Adaptive boosting or AdaBoost is one of the simplest boosting algorithms.
- Usually, decision trees(weak learners) are used for modelling.
- Multiple sequential models are created, each correcting the errors from the last model.
- AdaBoost assigns weights to the observations which are incorrectly predicted and the subsequent model works to predict these values correctly.



Do It Skills
do it. enjoy it.

ALGORITHMS BASED ON BAGGING AND BOOSTING

- **AdaBoost**

Below are the steps for performing the AdaBoost algorithm:

1. Initially, all observations in the dataset are given equal weights.
2. A model is built on a subset of data.
3. Using this model, predictions are made on the whole dataset.
4. Errors are calculated by comparing the predictions and actual values.
5. While creating the next model, higher weights are given to the data points which were predicted incorrectly.
6. Weights can be determined using the error value. For instance, higher the error more is the weight assigned to the observation.
7. This process is repeated until the error function does not change, or the maximum limit of the number of estimators is reached.

ALGORITHMS BASED ON BAGGING AND BOOSTING

- **AdaBoost**

Parameters

- **base_estimators**:
- **n_estimators:**
- **learning_rate:**
- **max_depth**:
- **n_jobs**
- **random_state** :

Do It Skills
do it. enjoy it.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Gradient Boosting (GBM)

Gradient Boosting or GBM is another ensemble machine learning algorithm that works for both regression and classification problems. GBM uses the boosting technique, combining a number of weak learners to form a strong learner. Regression trees used as a base learner, each subsequent tree in series is built on the errors calculated by the previous tree.

We will use a simple example to understand the GBM algorithm. We have to predict the age of a group of people using the below data:

ID	Married	Gender	Current City	Monthly Income	Age (target)
1	Y	M	A	51,000	35
2	N	F	B	25,000	24
3	Y	M	A	74,000	38
4	N	F	A	29,000	30
5	N	F	B	37,000	33

Do It Skills
do it. enjoy it.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Gradient Boosting (GBM)

1. The mean age is assumed to be the predicted value for all observations in the dataset.
2. The errors are calculated using this mean prediction and actual values of age

ID	Married	Gender	Current City	Monthly Income	Age (target)	Mean Age (prediction 1)	Residual 1
1	Y	M	A	51,000	35	32	3
2	N	F	B	25,000	24	32	-8
3	Y	M	A	74,000	38	32	6
4	N	F	A	29,000	30	32	-2
5	N	F	B	37,000	33	32	1

Do It Skills
do it. enjoy it.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Gradient Boosting (GBM)

3. A tree model is created using the errors calculated above as target variable. Our objective is to find the best split to minimize the error.

4. The predictions by this model are combined

with the predictions 1

ID	Age (target)	Mean Age (prediction 1)	Residual 1 (new target)	Prediction 2	Combine (mean+pred2)
1	35	32	3	3	35
2	24	32	-8	-5	27
3	38	32	6	3	35
4	30	32	-2	-5	27
5	33	32	1	3	35

5. This value calculated above is the new prediction.

Do It Skills
do it. enjoy it.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Gradient Boosting (GBM)

6. New errors are calculated using this predicted value and actual value

ID	Age (target)	Mean Age (prediction 1)	Residual 1 (new target)	Prediction 2	Combine (mean+pred2)	Residual 2 (latest target)
1	35	32	3	3	35	0
2	24	32	-8	-5	27	-3
3	38	32	6	3	35	-3
4	30	32	-2	-5	27	3
5	33	32	1	3	35	-2

7. Steps 2 to 6 are repeated till the maximum number of iterations is reached (or error function does not change).

ALGORITHMS BASED ON BAGGING AND BOOSTING

Gradient Boosting (GBM)

Parameters

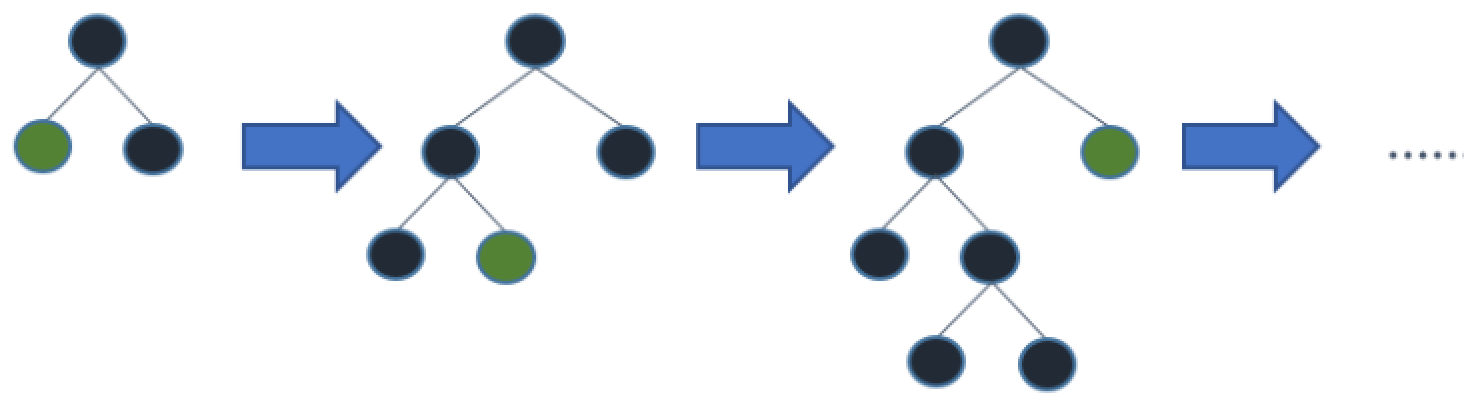
- **min_samples_split**
- **min_samples_leaf**

min_weight_fraction_leaf
max_depth
max_leaf_nodes
max_features

ALGORITHMS BASED ON BAGGING AND BOOSTING

XGBoost

- XGBoost (extreme Gradient Boosting) is an advanced implementation of the gradient boosting algorithm.
- Proved to be a highly effective ML algorithm, extensively used in machine learning competitions and hackathons.
- Has high predictive power
- Almost 10 times faster than the other gradient boosting techniques.
- Includes a variety of regularization which reduces overfitting and improves overall performance.
- Also known as ‘**regularized boosting**‘ technique.



Leaf-wise tree growth

ALGORITHMS BASED ON BAGGING AND BOOSTING

XGBoost

How XGBoost is comparatively better than other techniques:

1. **Regularization:**
2. **Parallel Processing:**.
3. **High Flexibility:**
4. **Handling Missing Values:**
5. **Tree Pruning:**
6. **Built-in Cross-Validation:**

ALGORITHMS BASED ON BAGGING AND BOOSTING

XGBoost

Parameters

- nthread
- eta
- min_child_weight
- max_depth
- max_leaf_nodes
- gamma
- subsample
- colsample_bytree

ALGORITHMS BASED ON BAGGING AND BOOSTING

Light GBM

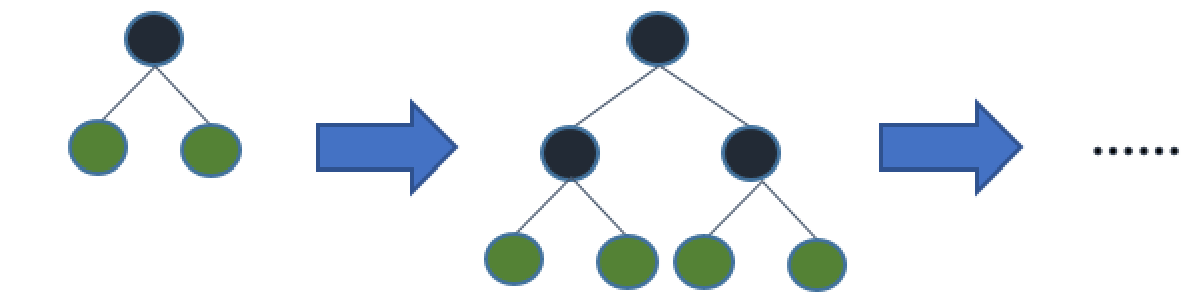
LightGBM is a gradient boosting framework that uses tree-based algorithms and follows leaf-wise approach while other algorithms work in a level-wise approach pattern.

WHY LIGHT GBM???

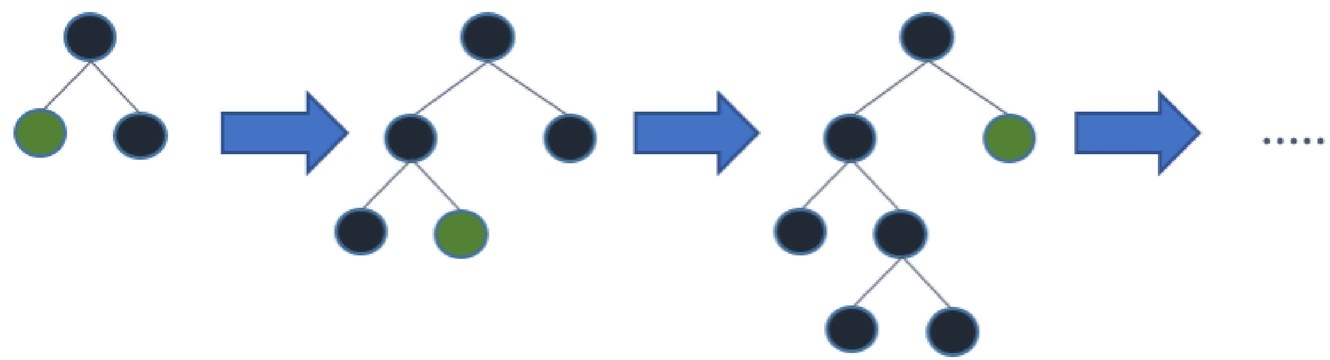
- **Light GBM beats all the other algorithms when the dataset is extremely large**.
- Takes lesser time to run on a huge dataset.

ALGORITHMS BASED ON BAGGING AND BOOSTING

Light GBM



Level-wise tree growth



Leaf-wise tree growth

ALGORITHMS BASED ON BAGGING AND BOOSTING

Light GBM

Leaf-wise growth may cause overfitting on smaller datasets but that can be avoided by using the ‘max_depth’ parameter for learning.

Parameters

- num_iterations:
- num_leaves :
- min_data_in_leaf :
- max_depth:
- bagging_fraction:
- max_bin :

ALGORITHMS BASED ON BAGGING AND BOOSTING

CatBoost

- Handling categorical variables is a tedious process
- Performing one-hot-encoding on them exponentially increases the dimensionality
- It becomes really difficult to work with the dataset.

CatBoost can automatically deal with categorical variables and does not require extensive data preprocessing like other machine learning algorithms.

ALGORITHMS BASED ON BAGGING AND BOOSTING

CatBoost

Parameters

- loss_function:
- iterations:
- learning_rate:
- border_count:
- depth:
- random_seed:

This brings us to the end of the ensemble algorithms

CONCLUSION

Ensemble modeling can exponentially boost the performance of your model and can sometimes be the deciding factor between first place and second!

THANK YOU