

6) Random Forest (C/R) .

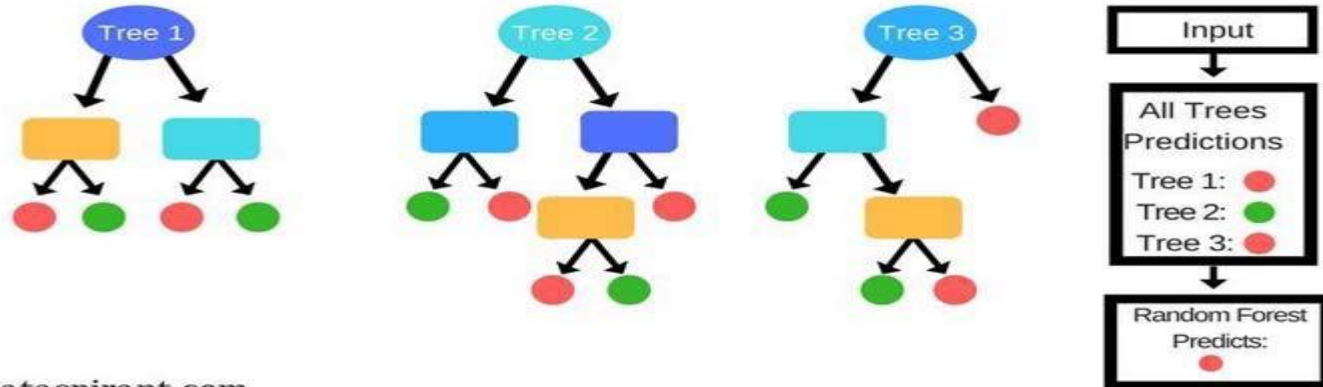
--> The Random forest or Random Decision Forest is a supervised Machine learning algorithm used for classification, regression, and other tasks using decision trees.

--> In the case of a classification problem, the final output is taken by using the majority voting classifier.

--> In the case of a regression problem, the final output is the mean of all the outputs.

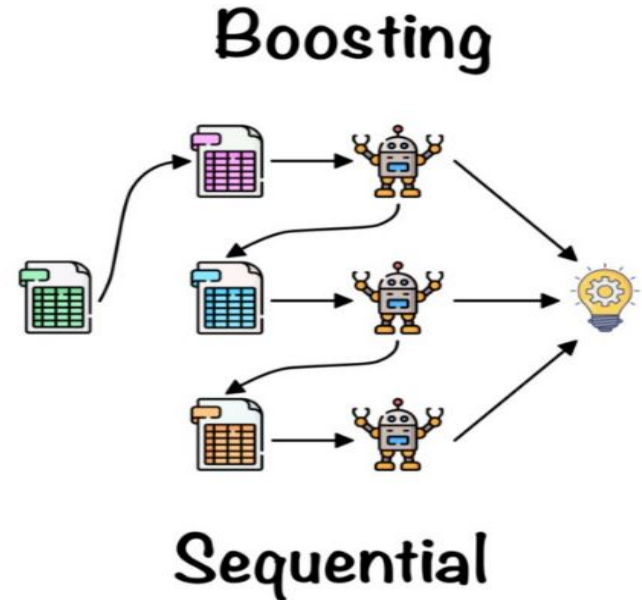
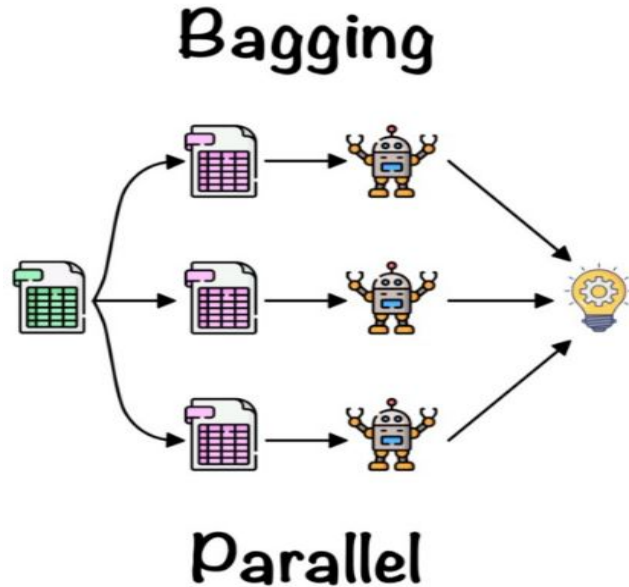
--> A Random Forest is an ensemble technique (Uses bagging method) capable of performing both regression and classification tasks with the use of multiple decision trees. Random Forest has multiple decision trees as base learning models.

Introduction To Random Forest Algorithm



--> The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

--> Every decision tree has high variance, but when we combine all of them together in parallel then the resultant variance is low as each decision tree gets perfectly trained on that particular sample data and hence the output doesn't depend on one decision tree but multiple decision trees.



--> A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging.

--> Bagging decreases variance, not bias, and solves over-fitting issues in a model.

--> In Bagging, training data subsets are drawn randomly with a replacement for the training dataset.

--> A boosting is an approach to increase the complexity of models that suffer from high bias, that is, models that underfit the training data.

--> Boosting decreases bias, not variance.

--> In Boosting, every new subset comprises the elements that were misclassified by previous models.

Advantages:

- > It is used to solve both classification as well as regression problems.
- > It works well with both categorical and continuous variables.
- > It can automatically handle missing values.
- > It is comparatively less impacted by noise.

Disadvantage:

- > Complexity.
- > Longer Training Period.

Applications:

--> Credit Card Fraud Detection.

--> Diabetes Prediction.

--> Stock Market Sentiment Analysis.

--> Product Recommendation.