Random Forest

I. Abstract

A random forest is a controlled machine learning method built from decision tree techniques. In machine learning, it may be utilised for both classification and regression issues. It is based on ensemble learning, which is a method of integrating several classifiers to solve a complicated issue and increase the model's performance. This algorithm is used to anticipate behaviour and results in a variety of sectors, including banking and e-commerce.

This document describes the algorithm and how it operates. The article will describe the algorithm's characteristics as well as how it is used in real-world situations. It also discusses the algorithm's benefits and drawbacks.

II. Introduction

A random forest is a machine learning approach for solving regression and classification issues. It makes use of ensemble learning, which is a technique that combines multiple classifiers to solve complicated problems. A random forest method is made up of a large number of decision trees. The random forest algorithm's 'forest' is trained via bagging or bootstrap aggregation. Bagging is a meta-algorithm ensemble that enhance the reliability of machine learning algorithms.Based on the predictions of the decision trees, the random forest algorithm determines the outcome. It forecasts by averaging or averaging the output from different trees. The precision of the result improves as the number of trees grows. The constraints of a decision tree algorithm are eliminated with a random forest. It reduces dataset overfitting and boosts precision. It provides forecasts without the need for several package settings [**1**].

III. Characteristics of a Random Forest Algorithm

Random Forest outperforms the decision tree algorithm in terms of accuracy. It gives an efficient method of dealing with missing data. Random forest is capable of producing a fair forecast without the need of hyper-parameter tweaking. It eliminates the problem of overfitting in decision trees. Random forest the node's splitting point in every random forest tree, a subset of characteristics is chosen at random [**2**].

IV. Working of algorithm

To have a better knowledge of random forest, we should first learn about decision trees. A random forest algorithm's building components are decision trees. A decision tree is a decision-making approach with a tree-like structure. A review of decision trees will assist us in comprehending how random forest algorithms function. A decision tree is made up of three parts: decision nodes, leaf nodes, and a root node. A decision tree method separates a training dataset into branches, which are then subdivided into subbranches. This procedure is repeated until a leaf node is reached. The leaf node cannot be further separated. The decision tree's nodes indicate qualities that are used to forecast the outcome. The decision nodes connect to the leaves, Figure 1 depicts the three types of decision tree nodes. [**3**].

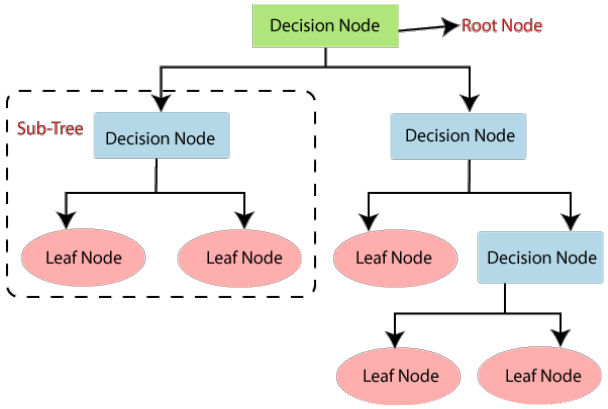

Fig. 1. General structure of a decision tree [**3**]

A supervised learning algorithm is random forest. The "forest" it creates is an ensemble of decision trees, which are often trained using the "bagging" approach. The bagging approach is based on the premise that combining learning models improves the final output. Simply said, random forest constructs many decision trees and blends them to get a more accurate and consistent forecast. Random forest has a significant benefit in that it can be utilised for both classification and regression tasks, which comprise the majority of contemporary machine learning systems. Let's look into random forest in classification, because classification is frequently said to be the foundation of machine learning. Random forest has roughly identical hyperparameters as decision trees and bagging classifiers. Fortunately, there is no need to combine a decision tree with a bagging classifier because the random forest classifier-class may be used instead. You may also use random forest to handle regression jobs by employing the algorithm's regressor. While growing the trees, Random Forest adds more randomization to the model. When splitting a node, it looks for the best feature from a random group of characteristics rather than the most essential feature. As a result, there is a wide range of variability, which leads to a better model in general [**4**].

V. Random forest classification

Random forest classification uses an ensemble technique to get the desired result. Various decision trees are trained using the training data. This dataset contains observations and characteristics that will be chosen at random when nodes are divided [**13**].

Various decision trees are used in a rain forest system. There are three types of nodes in a decision tree: decision nodes, leaf nodes, and the root node. Each tree's leaf node represents the ultimate result produced by that particular decision tree. The final product is chosen using a majority-voting procedure. In this situation, the ultimate output of the rain forest system is the output chosen by the majority of decision trees. A basic random forest classifier is shown in figure 2 below [**13**]. Because the random forest mixes numerous trees to forecast

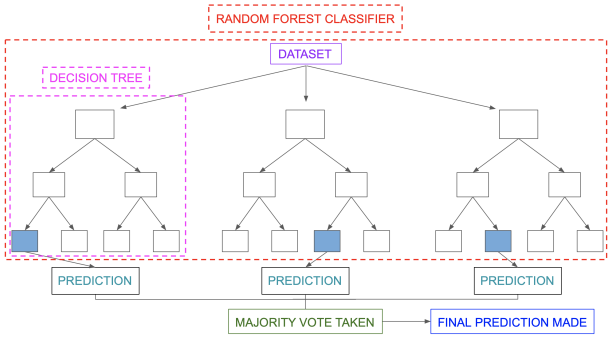

Fig. 2. basic random forest classifier [**13**]

the dataset's class, some decision trees may correctly predict the output while others may not. However, when all of the trees are combined, the proper result is predicted. As a result, two assumptions for a better Random forest classifier are as follows [**6**]:

- The dataset's feature variable should have some real values so that the classifier can predict correct outcomes rather than a guess.
- Each tree's predictions must have very low correlations.

VI. Example to Understand the Concept

VII. Random Forest Regression

VIII. Random forest applications

IX. Not ideal in the following circumstances

X. Random forest benefits and drawbacks

XI. Conclusion

XII. References

[1] Onesmus Mbaabu. (2020, December 11). Introduction to Random Forest in Machine Learning. Https://Www.Section.Io/. https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

[2] R, S. E. (2021, June 24). Random Forest — Introduction to Random Forest Algorithm. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

[3] Machine Learning Decision Tree Classification Algorithm - Javatpoint. (n.d.). Www.Javatpoint.Com. https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

[4] Donges, N. (2021, September 17). A Complete Guide to the Random Forest Algorithm. Built In. https://builtin.com/data-science/random-forest-algorithm

[5] Bakshi, C. (2021, December 14). Random Forest Regression - Level Up Coding. Medium. https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

[6] Meena, M. (2020, December 8). Applications of Random Forest. OpenGenus IQ: Computing Expertise and Legacy. https://iq.opengenus.org/applications-of-random-forest/

XIII. Declaration