**MCSE 666: Assignment 10**

**Roll: 079 05536**  **Name: Shyed Shahriar Housaini**

**What is Random Forest Algorithm? Why the name random forest given to the Random forest Algorithm?**

**Random Forest Algorithm** is an ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees at training time. The output of the random forest is the class selected by most trees for classification tasks, while for regression tasks, the mean or average prediction of the individual trees is returned [1](https://en.wikipedia.org/wiki/Random_forest). Random forests correct for decision trees’ habit of overfitting to their training set [1](https://en.wikipedia.org/wiki/Random_forest). They generally outperform decision trees, but their accuracy is lower than gradient boosted trees [1](https://en.wikipedia.org/wiki/Random_forest).

Random forest is a supervised learning algorithm that can be used for both classification and regression tasks. It works by constructing a multitude of decision trees at training time and using them to predict the output for new data points. The final prediction is made by averaging the predictions of all the decision trees.

**The name "Random Forest" is derived from two key aspects of the algorithm Data Randomness and Feature Randomness:**

1. **Random Subsampling (Bootstrap Aggregating):** The "Forest" part of the name comes from the fact that the algorithm creates a collection of decision trees, typically a large number of them. Each tree is trained on a random subset of the training data with replacement. This process is known as bootstrap aggregating or bagging. These subsets are referred to as "bootstrap samples."
2. **Random Feature Selection:** The "Random" part of the name comes from the fact that when each tree is being constructed, it doesn't consider all features for splitting nodes. Instead, at each node, a random subset of features is considered for finding the best split. This introduces an element of randomness into the tree construction process.

The combination of these two sources of randomness, both in terms of data sampling and feature selection, helps make Random Forest more robust and less prone to overfitting than individual decision trees. It also allows the algorithm to capture complex relationships in the data by aggregating the results from multiple trees.

The randomness in the selection of features and data subsets helps to reduce overfitting and improve generalization performance [1](https://en.wikipedia.org/wiki/Random_forest).

**Advantages of Random Forest Algorithm**

Random forest has a number of advantages over other machine learning algorithms, including:

* It is very accurate and robust to overfitting.
* It can be used for both classification and regression tasks.
* It is easy to interpret and understand.
* It can handle large datasets with many features.

**Disadvantages of Random Forest Algorithm**

Random forest also has some disadvantages, including:

* It can be computationally expensive to train.
* It can be difficult to tune the hyper-parameters.
* It is not as interpretable as some other machine learning algorithms, such as linear regression.

Overall, random forest is a powerful and versatile machine learning algorithm that can be used to solve a wide variety of problems.

Resource:

(1) Random forest - Wikipedia. https://en.wikipedia.org/wiki/Random\_forest.

(2) Random Forests Definition | DeepAI. https://deepai.org/machine-learning-glossary-and-terms/random-forest.

(3) Random Forest Algorithm - How It Works and Why It Is So Effective - Turing. https://www.turing.com/kb/random-forest-algorithm.

(4) Random forest - Simple English Wikipedia, the free encyclopedia. https://simple.wikipedia.org/wiki/Random\_forest.

(5) What Is Random Forest? A Complete Guide | Built In. https://builtin.com/data-science/random-forest-algorithm.