**AIM: Implementation of Bagging Algorithm: Decision Tree, Random Forest**

**THEORY:**

# Decision Tree:

A decision tree is a flowchart-like structure in which each internal node represents a test on a feature (e.g. whether a coin flip comes up heads or tails) , each leaf node represents a class label (decision taken after computing all features) and branches represent conjunctions of features that lead to those class labels. The paths from root to leaf represent classification rules. Below diagram illustrate the basic flow of decision tree for decision making with labels (Rain(Yes), No Rain(No)).

Decision tree is one of the predictive modelling approaches used in statistics, data mining and machine learning.

Decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a nonparametric supervised learning method used for both classification and regression tasks.

Tree models where the target variable can take a discrete set of values are called classification trees. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.

# Random Forest:

Random forest is an ensemble machine learning algorithm.

It is perhaps the most popular and widely used machine learning algorithm given its good or excellent performance across a wide range of classification and regression predictive modeling problems.

It is also easy to use given that it has few key hyperparameters and sensible heuristics for configuring these hyperparameters.

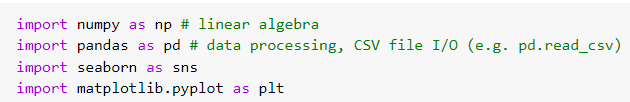
It is an extension of bootstrap aggregation (bagging) of decision trees and can be used for classification and regression problems.

In bagging, a number of decision trees are created where each tree is created from a different bootstrap sample of the training dataset. A bootstrap sample is a sample of the training dataset where a sample may appear more than once in the sample, referred to as sampling with replacement.

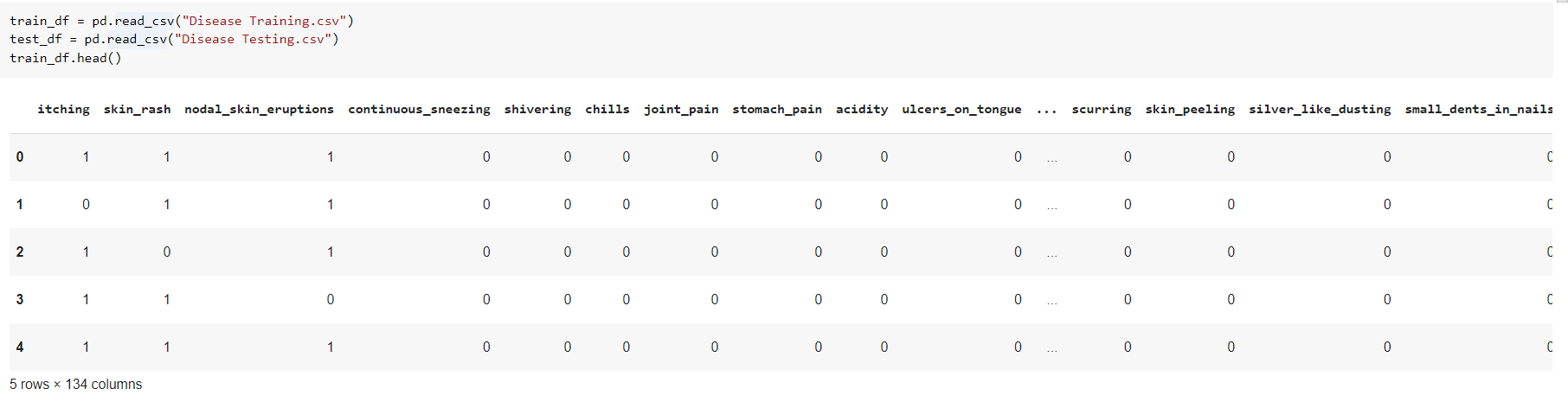
Bagging is an effective ensemble algorithm as each decision tree is fit on a slightly different training dataset, and in turn, has a slightly different performance. Unlike normal decision tree models, such as classification and regression trees (CART), trees used in the ensemble are unpruned, making them slightly overfit to the training dataset. This is desirable as it helps to make each tree more different and have less correlated predictions or prediction errors.

Predictions from the trees are averaged across all decision trees resulting in better performance than any single tree in the model.

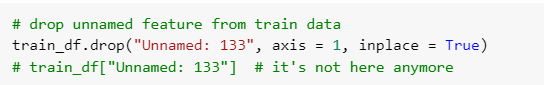
1. **IMPORTING LIBRARIES:**



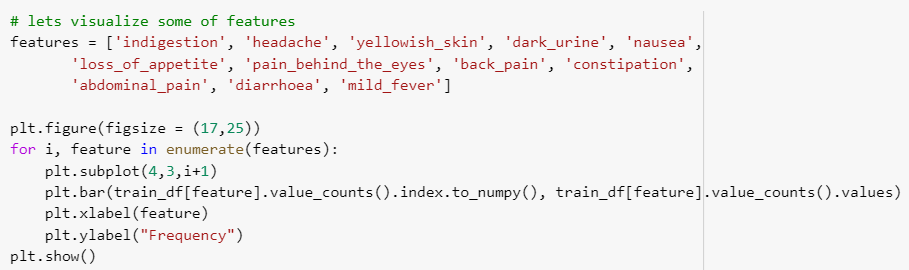
1. **READING DATASET [TRAINING & TESTING]:**

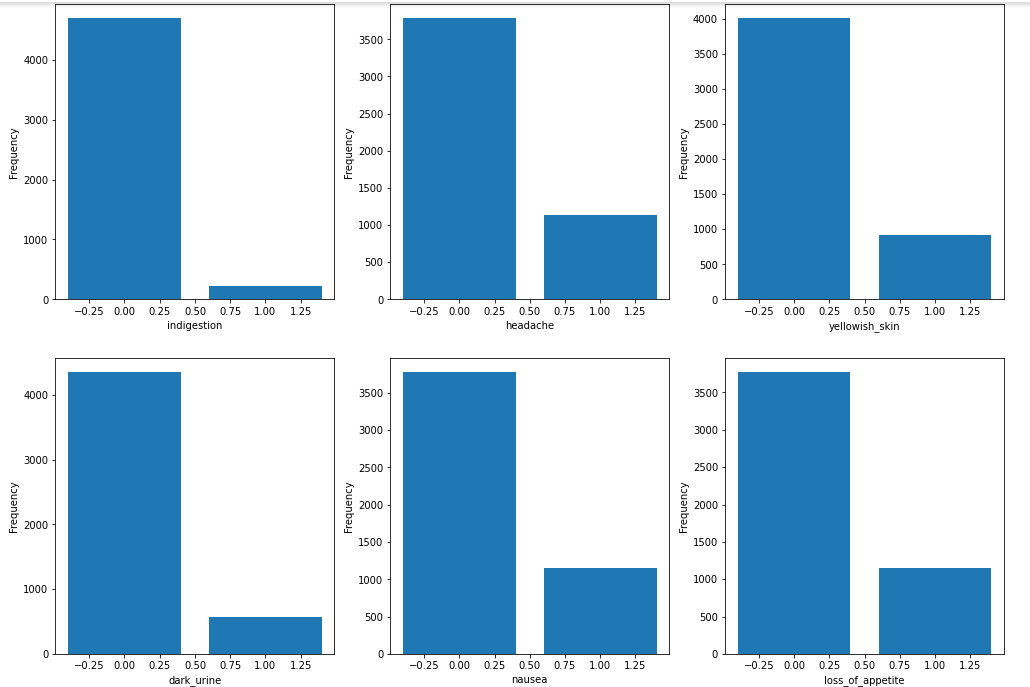


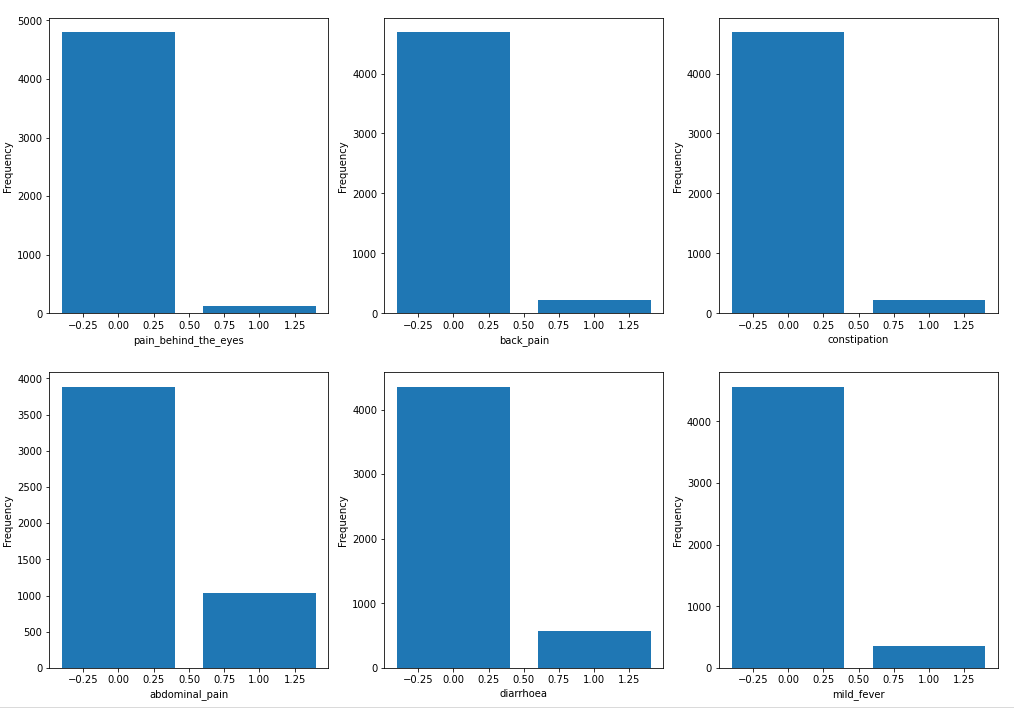
1. **DATA CLEANING:**

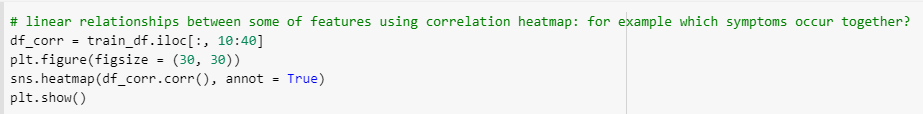


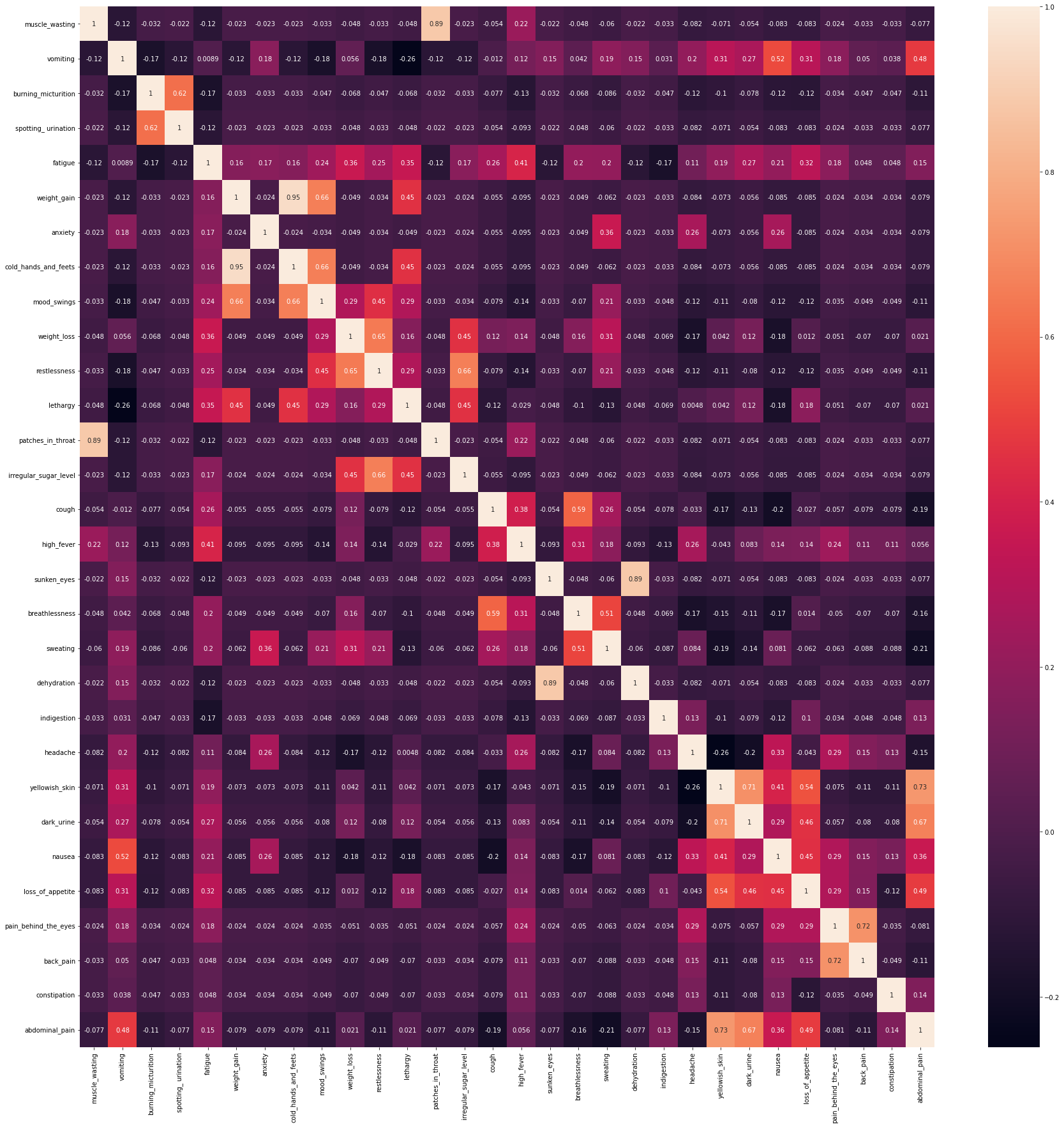
1. **FEATURES VISULALIZATION:**



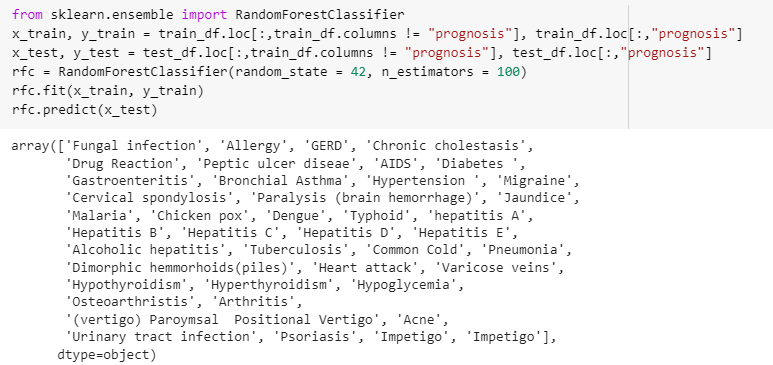




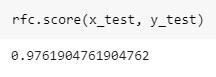




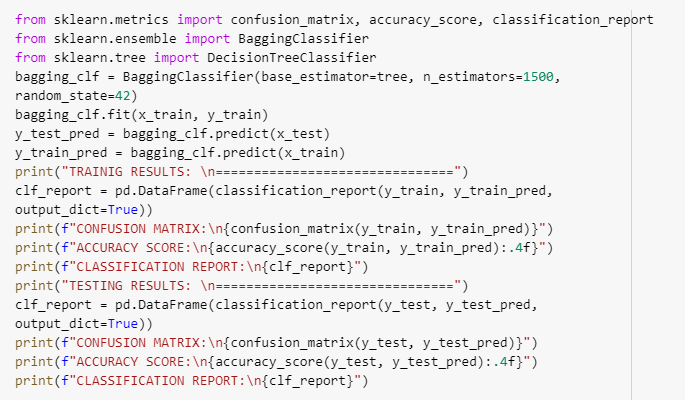
1. **BUILDING RANDOM FOREST MODEL:**

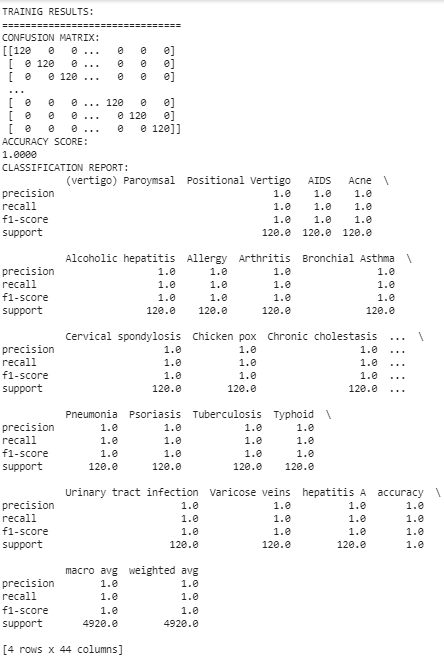


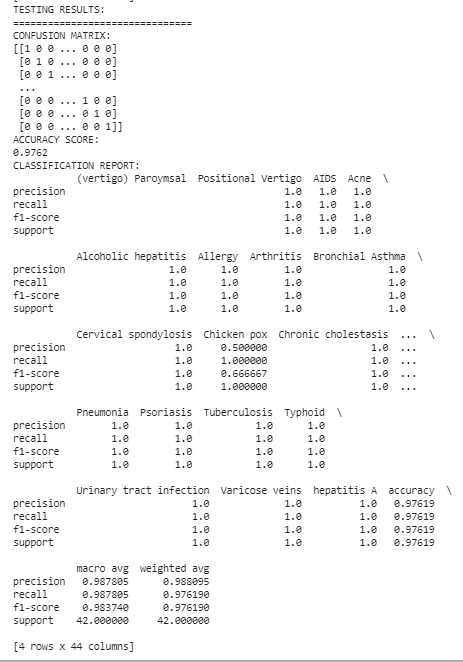
1. **CHECKING RANDOM FOREST SCORE:**



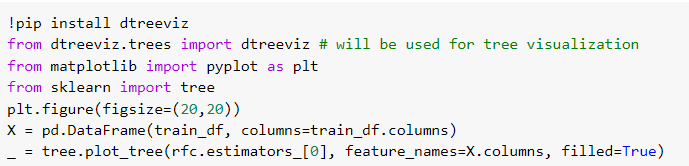
1. **BUILDING BAGGING CLASSIFIER:**







1. **PLOTTING TREE:**





**CONCLUSION:**

From this practical, I have learned and implemented the random forest algorithm in python.