**SIR SYED UNIVERSITY OF ENGINEERING & TECHNOLOGY COMPUTER SCIENCE & INFORMATION TECHNOLOGY DEPARTMENT**

**Fall 2023**

**Machine Learning (CS-466) Assignment # 1**

Semester: 8th Batch: 2020S

Announced Date: 07/05/2024 Due Date:13/05/2024

Total Marks: 03 Instructor: Mr. Shardha Nand

**Group Member Names:**

**NAME: Muhib Mallick, ROLL: CS2020F – 153,**

**NAME: Ali Raza, ROLL: CS2020F – 180,**

**NAME: Saad, ROLL: CS2020F – 138,  
SEC: D,**

|  |  |  |  |
| --- | --- | --- | --- |
| **CLO #** | **Course Learning Outcomes (CLOs)** | **PLO Mapping** | **Bloom’s Taxonomy** |
| 1 | Describe the characteristics of different algorithms of Machine Learning and identify the functionalities of these in Classification and prediction | PLO\_2  (Knowledge for Computing Problem) | C2  (Understanding) |

**Question No.01 Marks.03**

1. **How the ensemble leaning helpful for avoiding overfitting. Discuss with suitable example. No marks**
2. Ensemble learning helps avoid overfitting by combining the predictions of multiple models to improve generalization over individual models. This is achieved by averaging the models' predictions to reduce variance without significantly increasing bias.

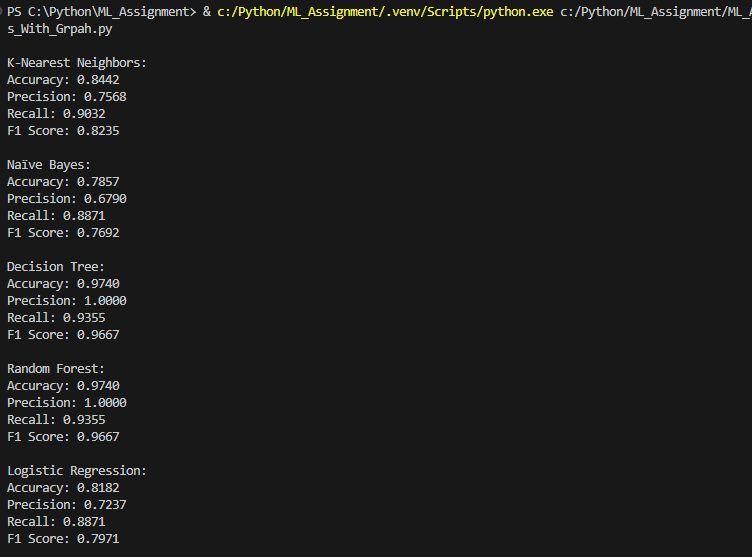
**Example: Random Forest**

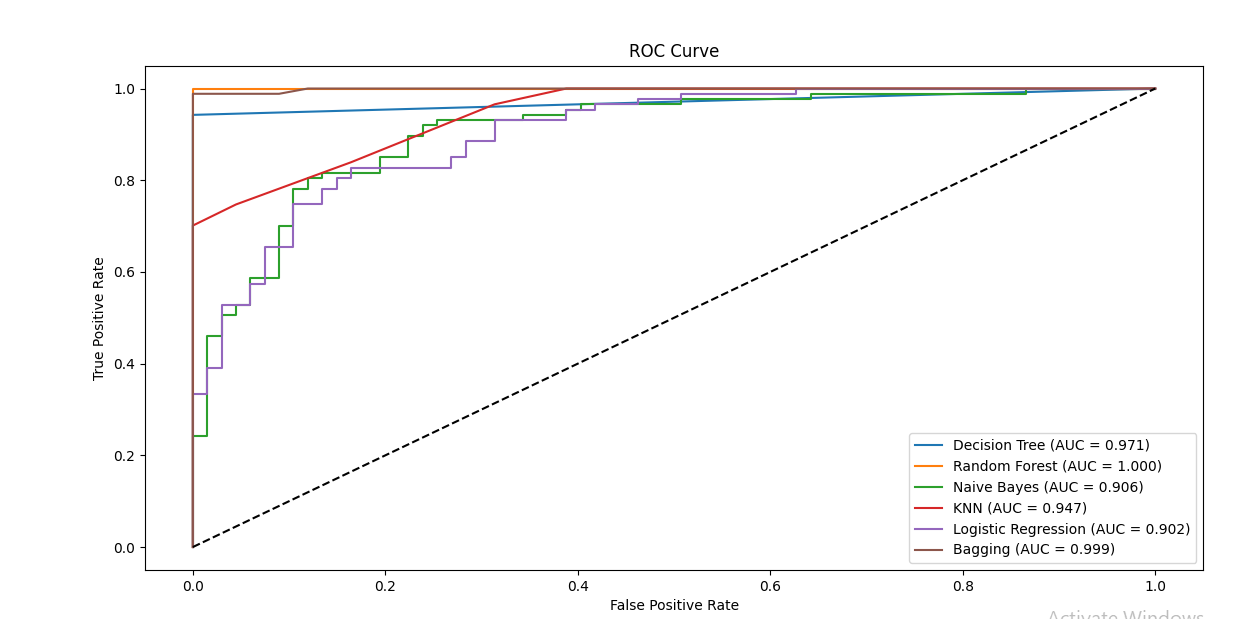
Random Forest uses bagging (bootstrap aggregating) to prevent overfitting. It involves creating multiple decision trees, each trained on a random subset of the training data. Each tree makes its own predictions, and the final output is determined by averaging these predictions (for regression) or by majority voting (for classification).

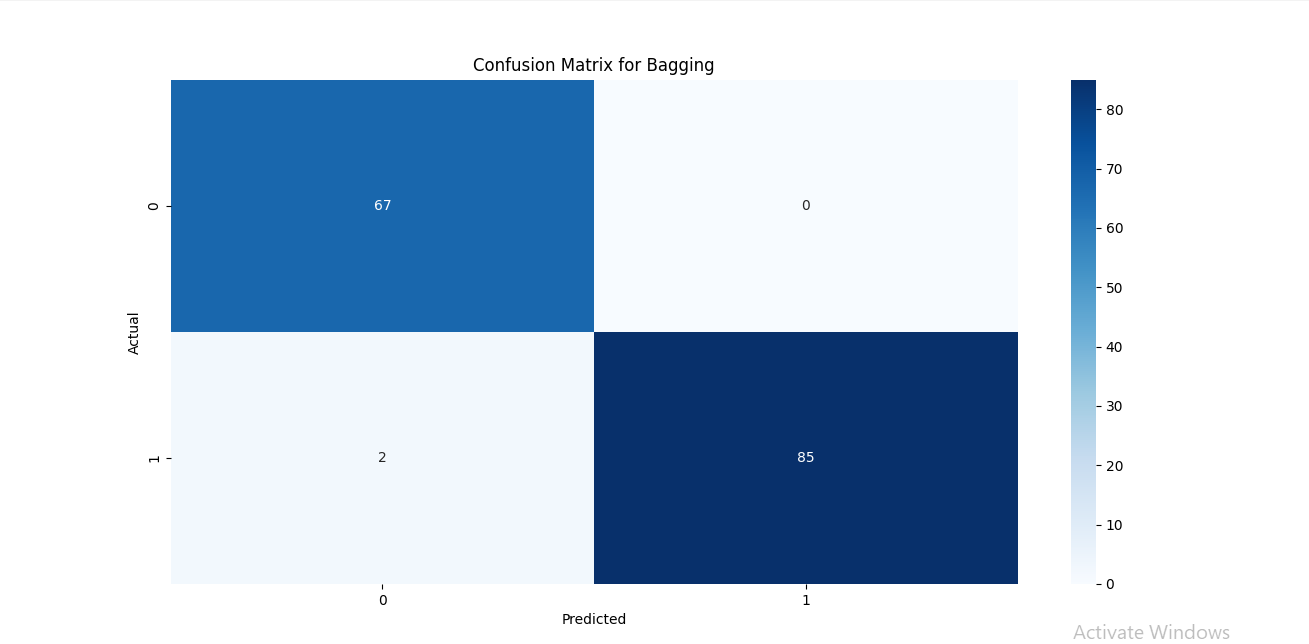
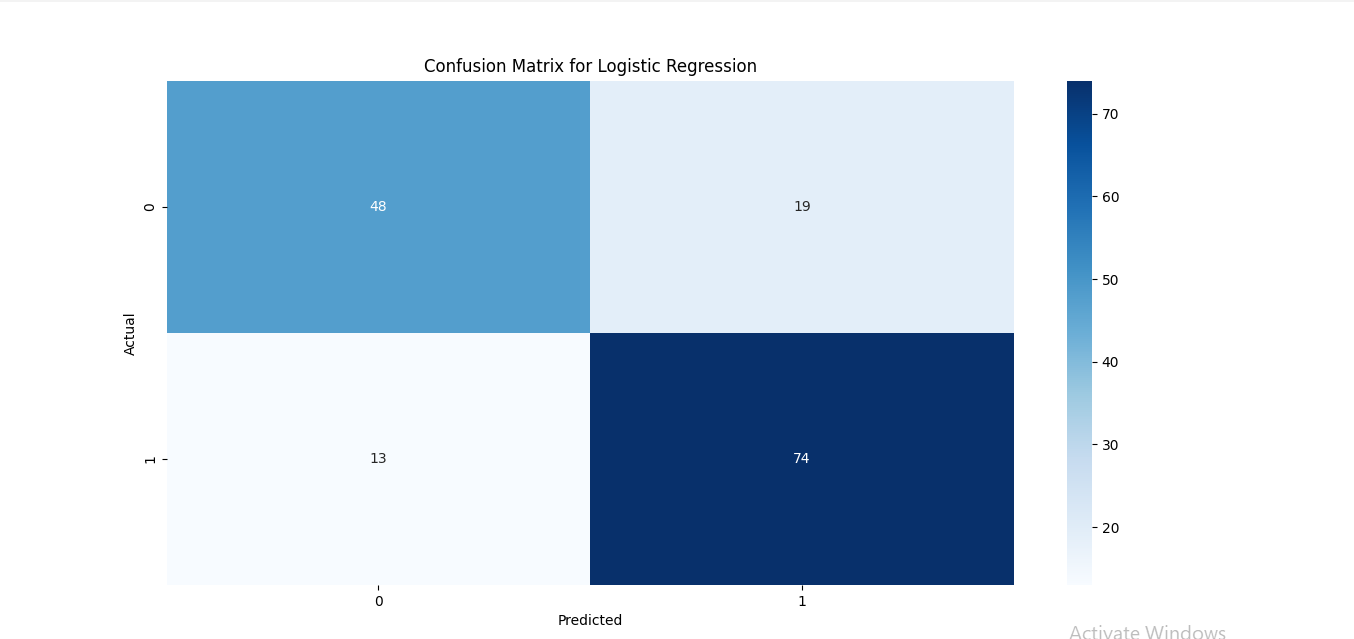
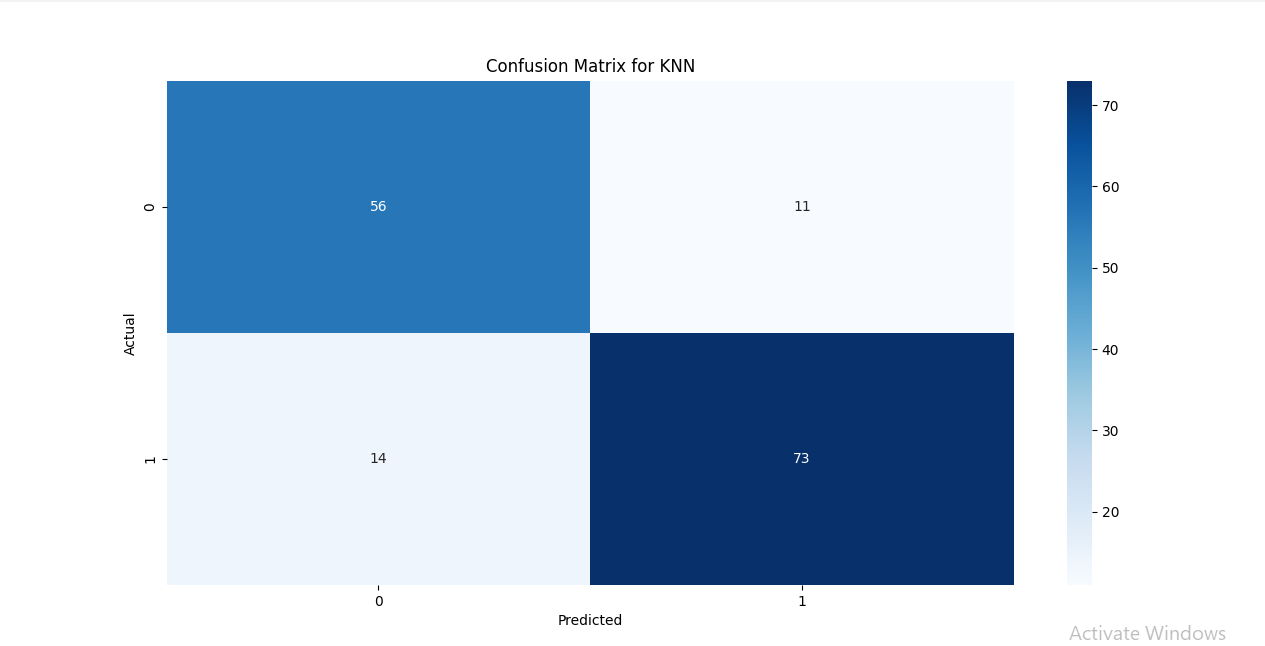
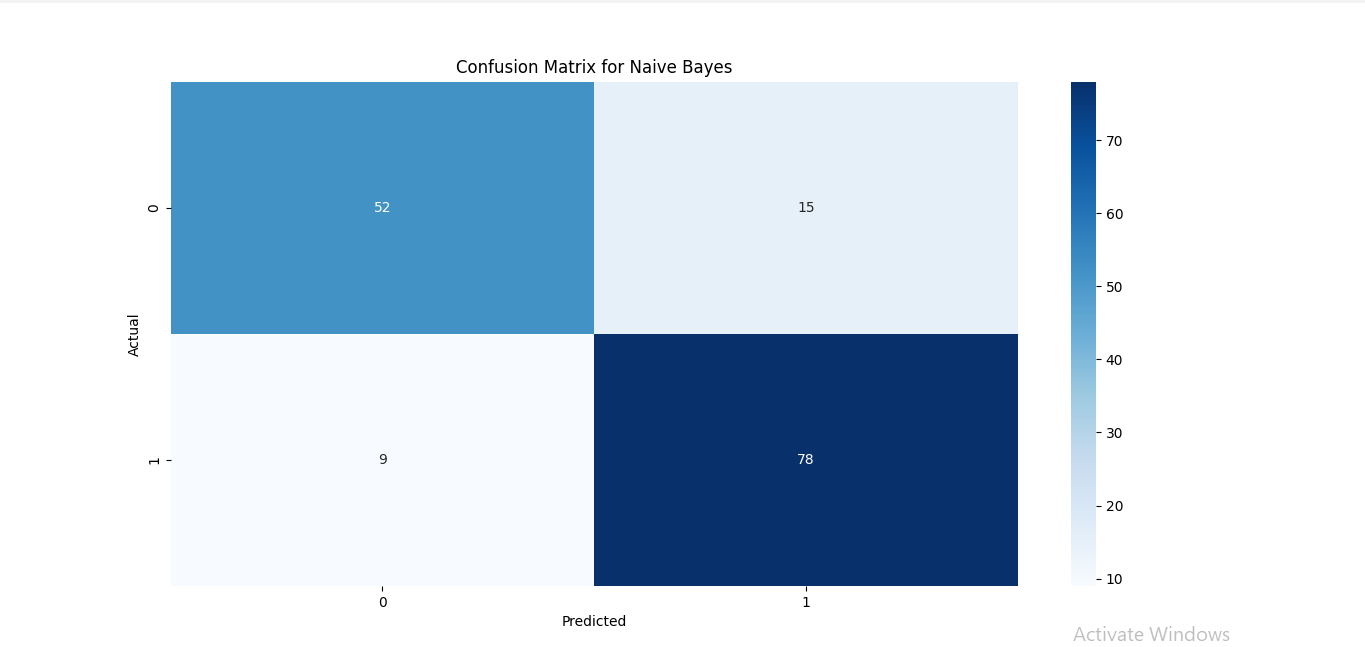
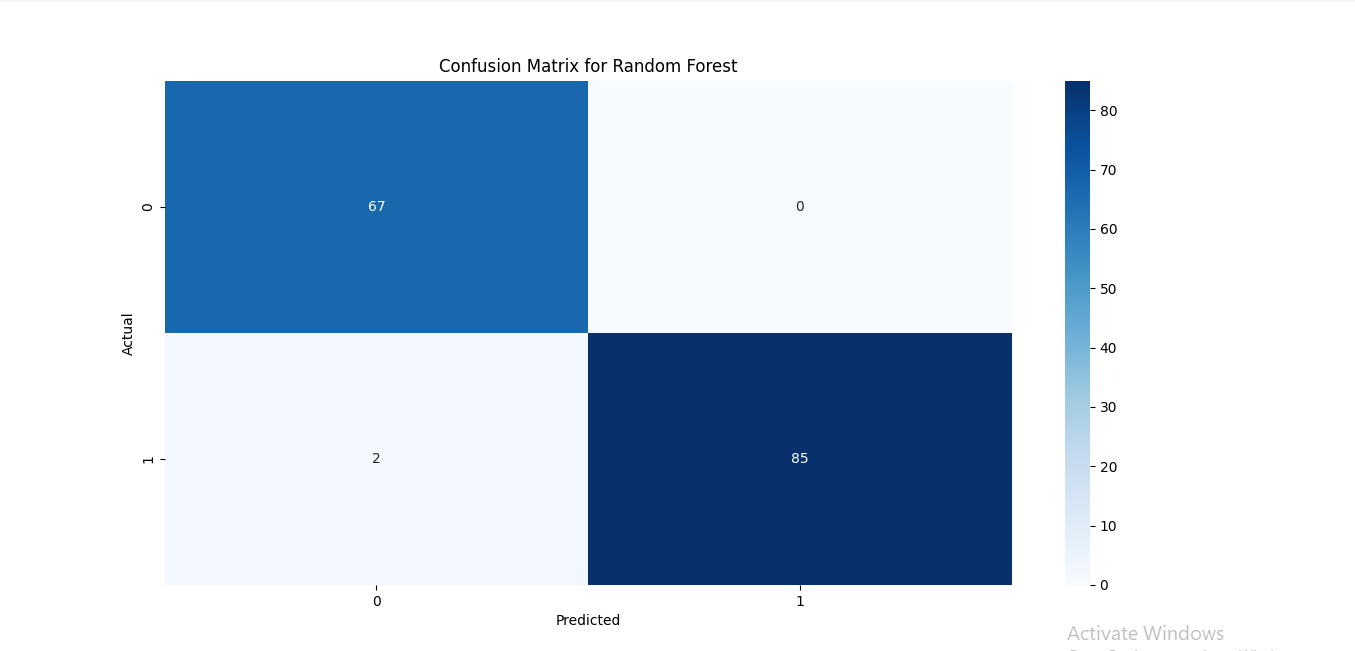
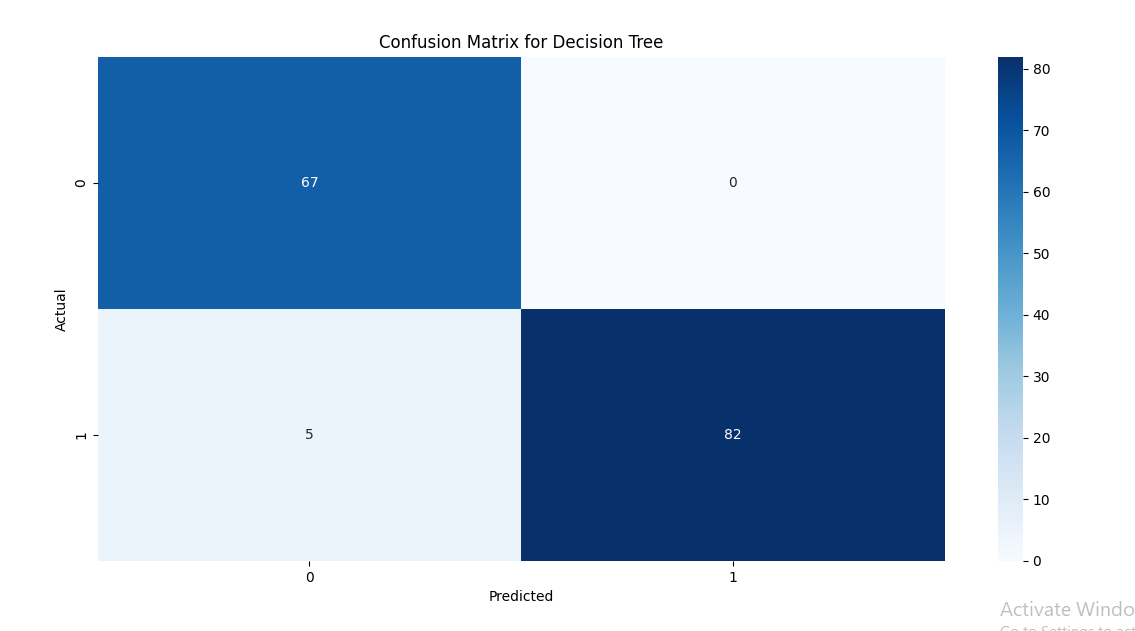
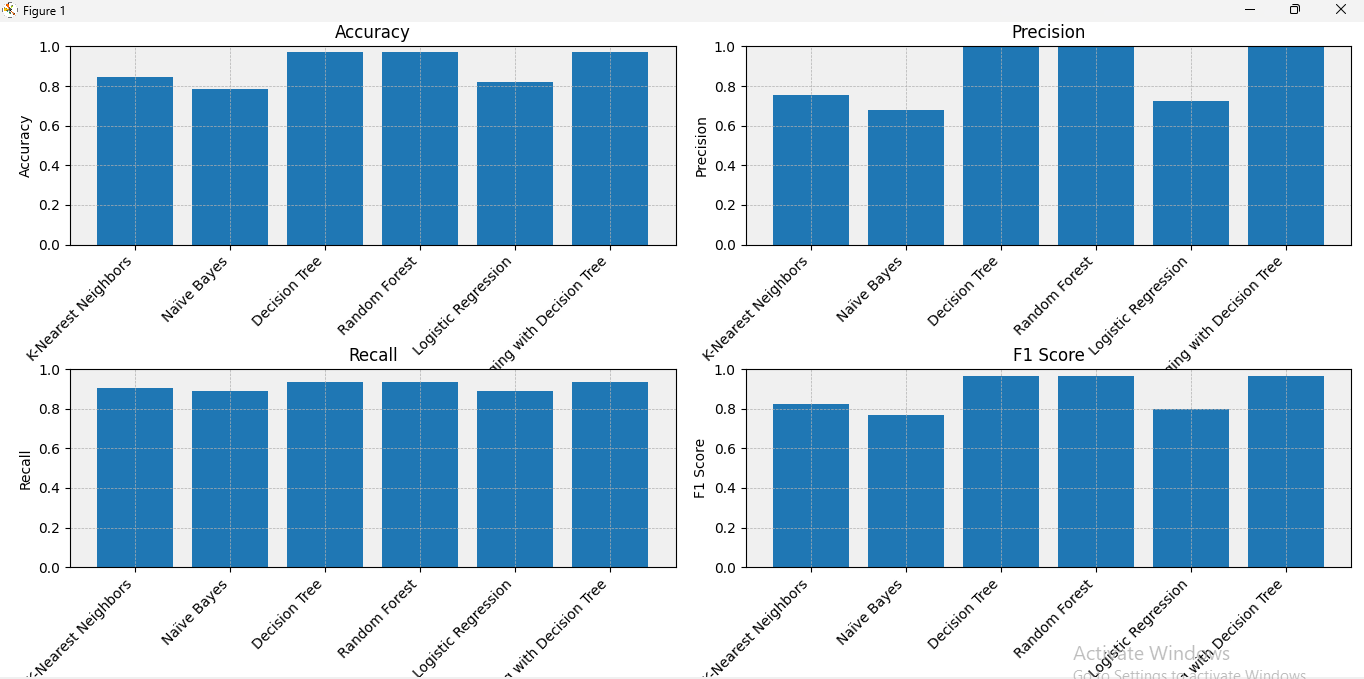
Since each tree sees only a part of the data, the risk of learning noise from the entire dataset (as a single tree might do) is reduced. This results in a model that is robust and less prone to overfitting, yet maintains a low bias by leveraging the power of multiple learners.

1. **Implement and compare the performance of following algorithms for classification: 03 Marks**
   1. **KNN**
   2. **Naïve Bayes**
   3. **Decision tree**
   4. **Random forest**
   5. **Logistic regression**

**OUTPUTS:**

****

****

****

Download public dataset of heart disease from Kaggle, GitHub or UCI Machine Learning repository. Implement and evaluate performance of each algorithm using accuracy, precision, recall, F1- score and AUC curve metrics.

**Bonus:** Implement a simple ensemble method (e.g. bagging or boosting) using one of the above algorithms and evaluate its performance.

**Note:** You can use Python with scikit-learn and TensorFlow libraries to implement the algorithms and evaluate their performance,