ML Assignment 2 – Classification Models & Deployment

Live Streamlit App Link:
https://heartdiseasedataset.streamlit.app/#machine-learning-model-evaluation-classification-app

1. Problem Statement

The objective of this project is to implement, evaluate, and deploy multiple machine learning classification models on a real-world dataset. The project demonstrates the complete end-to-end ML workflow including data preprocessing, model training, evaluation using standard metrics, building an interactive web application using Streamlit, and deploying the application on Streamlit Community Cloud.

2. Dataset Description

Dataset Name: Heart Disease Dataset
Source: Public dataset (UCI / Kaggle)
Type: Binary Classification
Number of Instances: 900+
Number of Features: 13
Target Variable: 1 → Presence of heart disease, 0 → Absence of heart disease

3. Models Implemented

Logistic Regression, Decision Tree Classifier, K-Nearest Neighbors (KNN), Naive Bayes (Gaussian), Random Forest (Ensemble Model), XGBoost (Ensemble Model).

4. Evaluation Metrics

Accuracy, AUC Score, Precision, Recall, F1 Score, Matthews Correlation Coefficient (MCC).

Model Performance Comparison Table

Model	Accuracy	AUC	Precision	Recall	F1	MCC
Logistic Regression	0.85	0.88	0.84	0.83	0.83	0.59
Decision Tree	0.97	0.97	0.97	0.97	0.97	0.97
KNN	0.89	0.91	0.88	0.87	0.87	0.67
Naive Bayes	0.80	0.87	0.75	0.89	0.81	0.61
Random Forest	0.98	1.00	1.00	0.97	0.98	0.97
XGBoost	0.98	0.99	1.00	0.97	0.98	0.97