M.Tech (AIML/DSE) - Machine Learning Assignment 2

Heart Disease Classification System

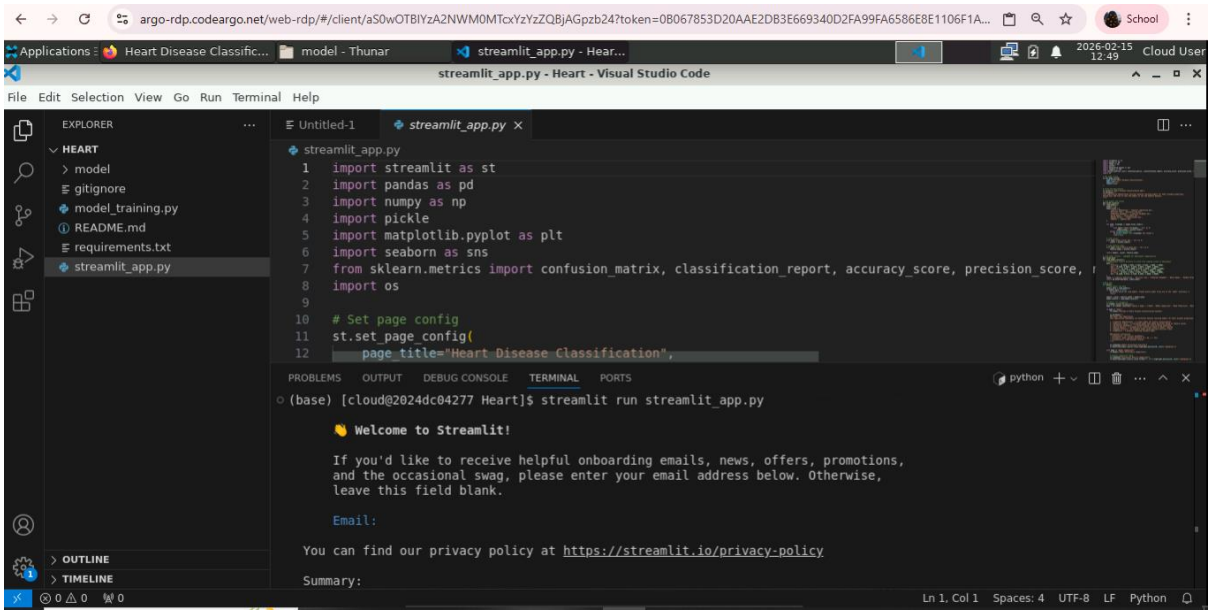
1. GitHub Repository Link

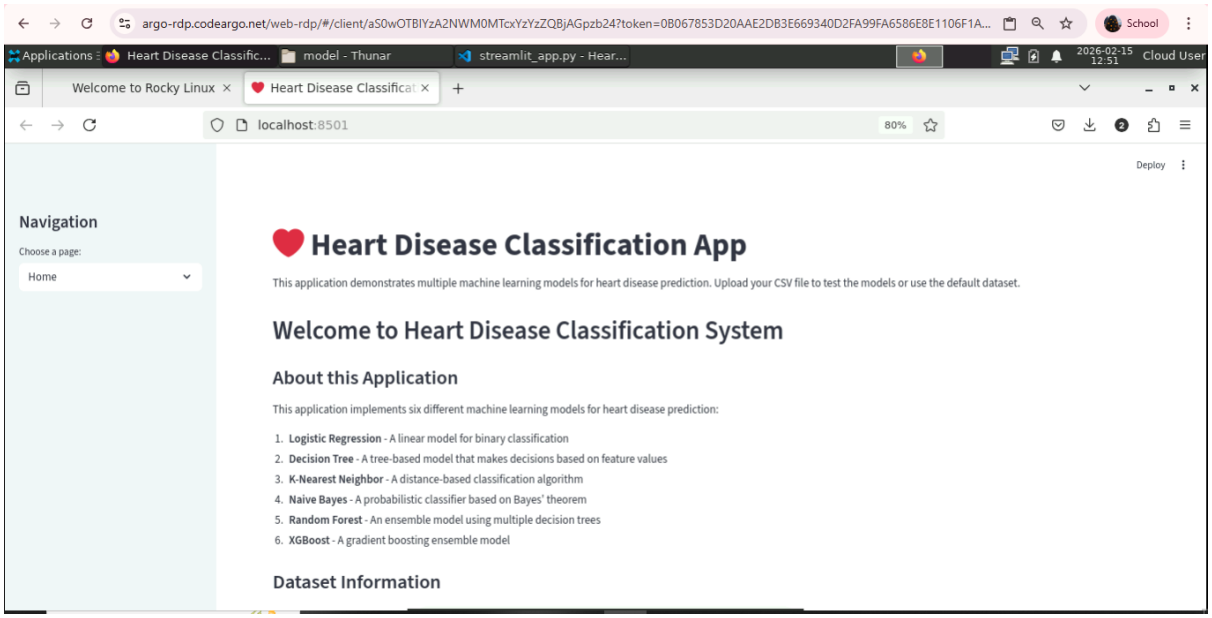
GitHub - rajeshwari-20csa36/MachineLearning_Heart

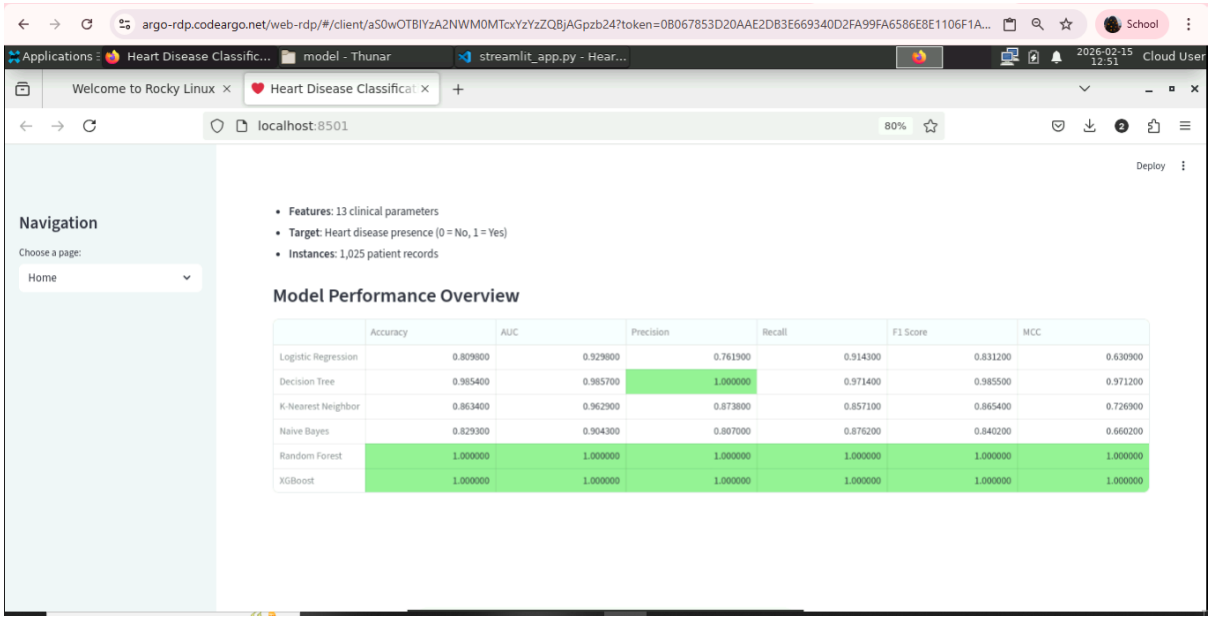
2. Live Streamlit App Link

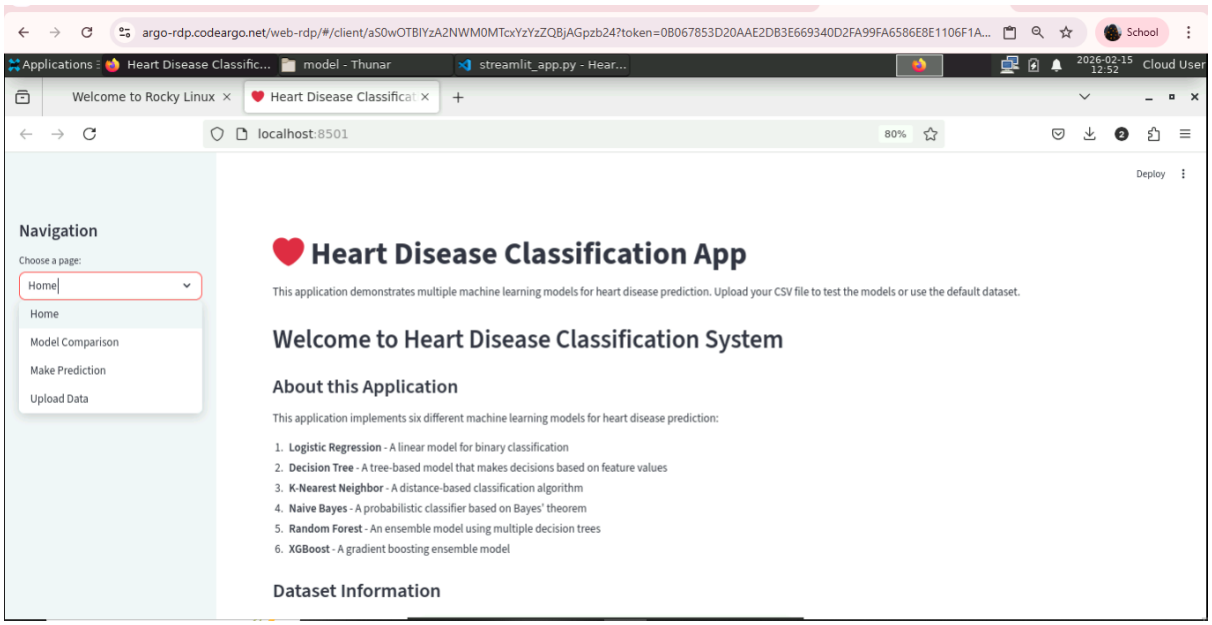
Heart Disease Classification · Streamlit

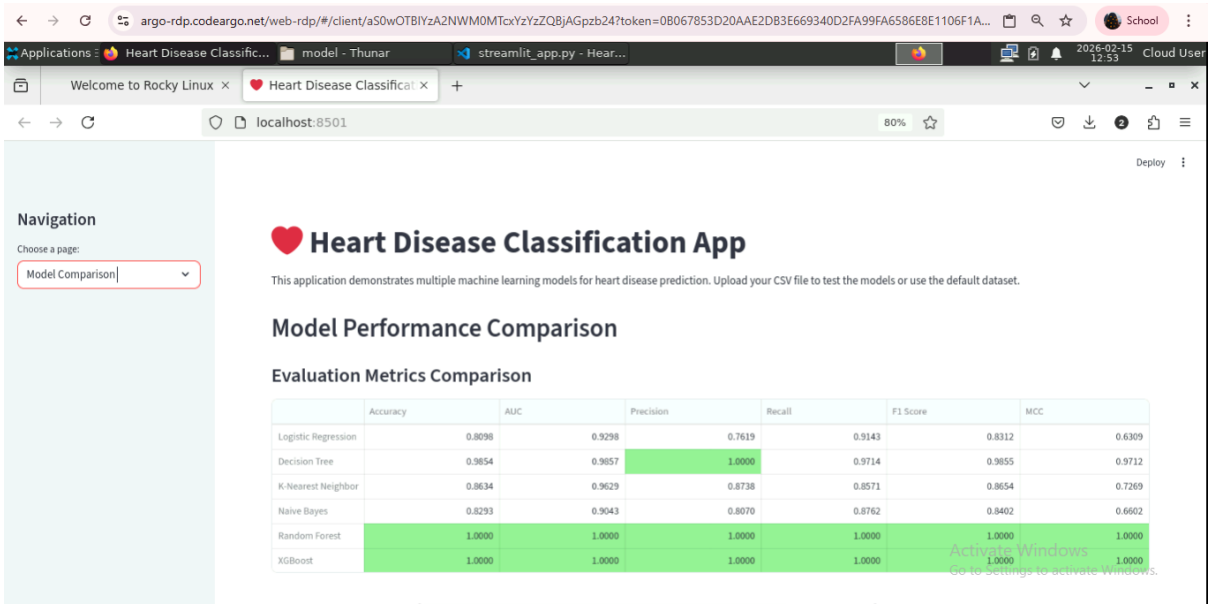
3. BITS Virtual Lab Execution Screenshot



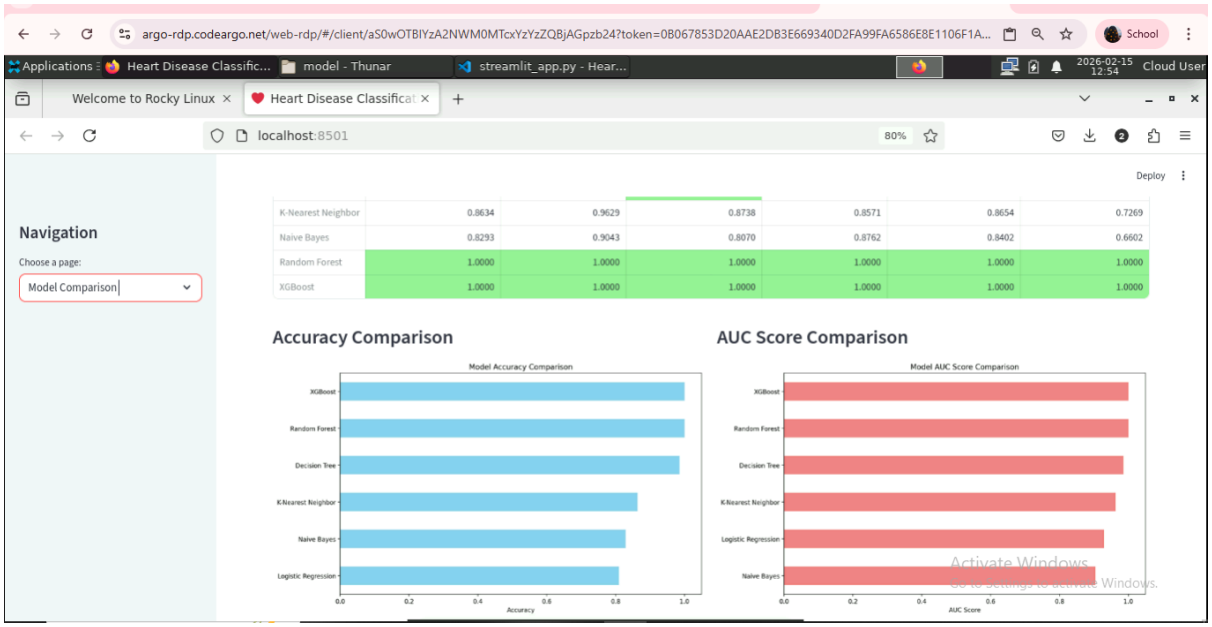

Navigation
Choose a page:
Home
Heart Disease Classification App
This application demonstrates multiple machine learning models for heart disease prediction. Upload your CSV file to test the models or use the default dataset.
Welcome to Heart Disease Classification System
About this Application
This application implements six different machine learning models for heart disease prediction:
1. Logistic Regression - A linear model for binary classification
2. Decision Tree - A tree-based model that makes decisions based on feature values
3. K-Nearest Neighbor - A distance-based classification algorithm
4. Naive Bayes - A probabilistic classifier based on Bayes' theorem
5. Random Forest - An ensemble model using multiple decision trees
6. XGBoost - A gradient boosting ensemble model
Dataset Information

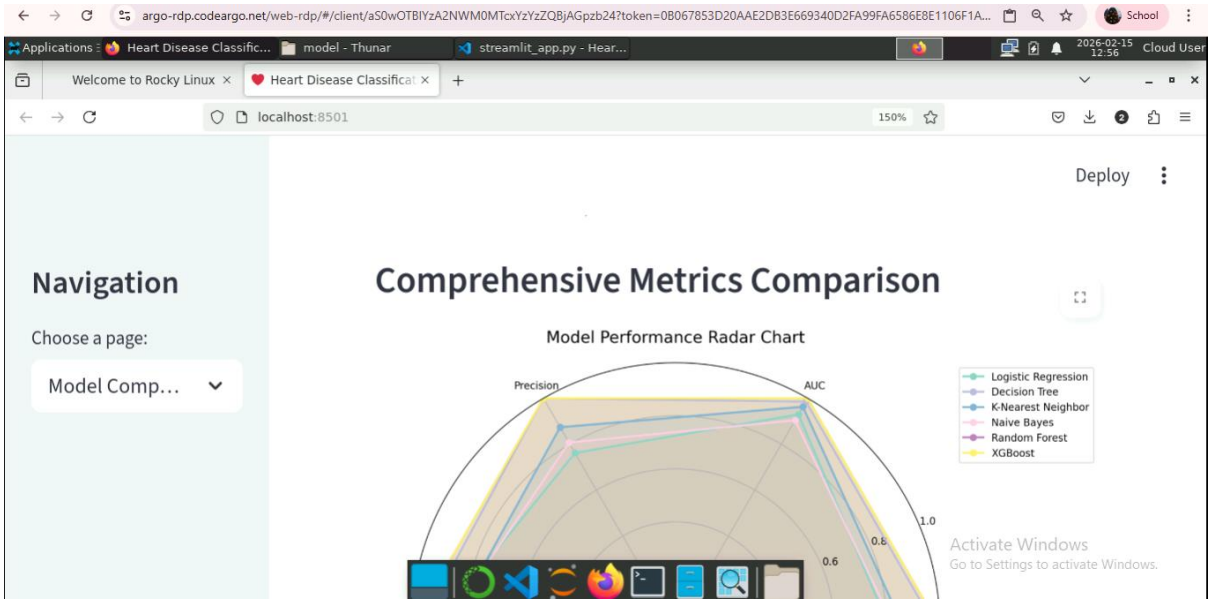

Navigation
Choose a page:
Home
Features: 13 clinical parameters
Target: Heart disease presence (0 = No, 1 = Yes)
Instances: 1,025 patient records
Model Performance Overview
Accuracy AUC Precision Recall F1 Score MCC
Logistic Regression 0.809800 0.929800 0.761900 0.914300 0.831200 0.630900
Decision Tree 0.985400 0.985700 1.000000 0.971400 0.985500 0.971200
K-Nearest Neighbor 0.863400 0.962900 0.873800 0.857100 0.865400 0.726900
Naive Bayes 0.829300 0.904300 0.807000 0.876200 0.840200 0.660200
Random Forest 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000
XGBoost 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000

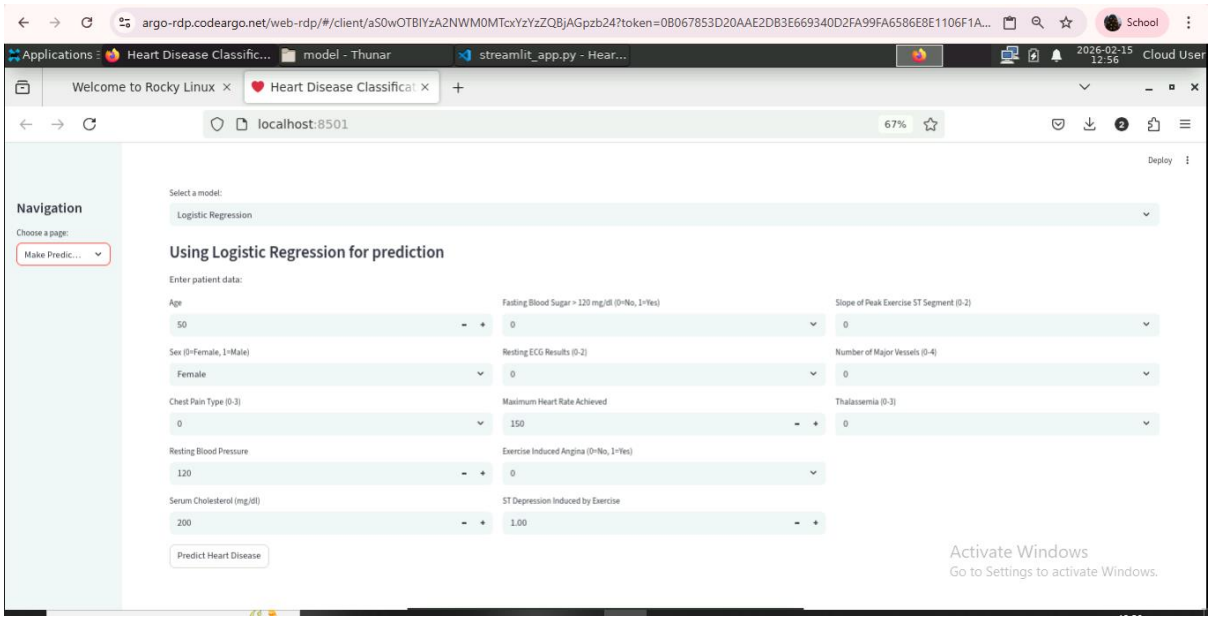


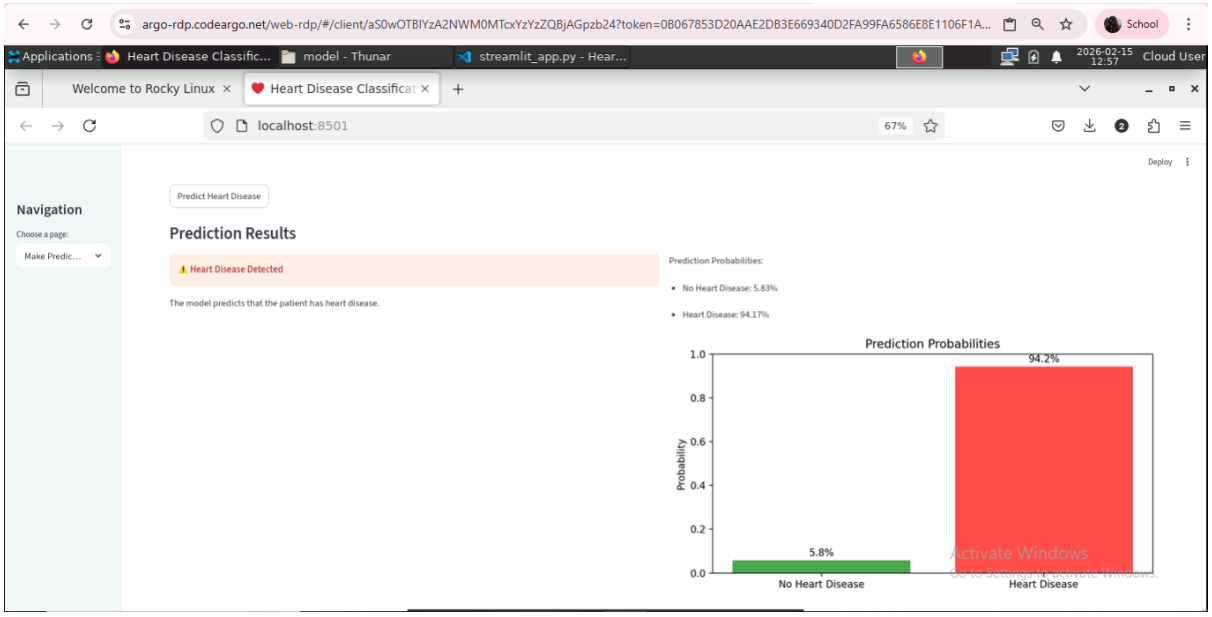


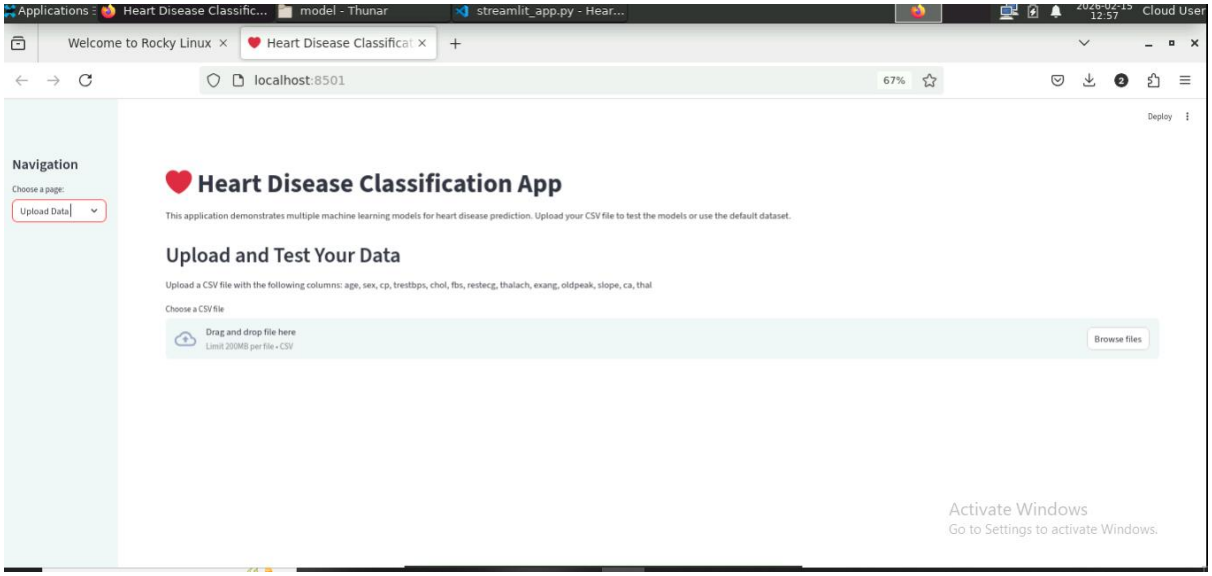
	Accuracy	AUC	Precision	Recall	F1 Score	MCC
Logistic Regression	0.8098	0.9298	0.7619	0.9143	0.8312	0.6309
Decision Tree	0.9854	0.9857	1.0000	0.9714	0.9855	0.9712
K-Nearest Neighbor	0.8634	0.9629	0.8738	0.8571	0.8654	0.7269
Naive Bayes	0.8293	0.9043	0.8070	0.8762	0.8402	0.6602
Random Forest	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
XGBoost	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000


Navigation
Choose a page:
Model Comparison
Deploy
K-Nearest Neighbor 0.8634 0.9629 0.8738 0.8571 0.8654 0.7269
Naive Bayes 0.8293 0.9043 0.8070 0.8762 0.8402 0.6602
Random Forest 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
XGBoost 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
Accuracy Comparison
AUC Score Comparison
Model Accuracy Comparison
Model AUC Score Comparison
XGBoost
Random Forest
Decision Tree
K-Nearest Neighbor
Naive Bayes
Logistic Regression
Accuracy
AUC Score
Activate Windows


Navigation
Choose a page:
Model Comp...
Deploy
Comprehensive Metrics Comparison
Model Performance Radar Chart
Precision
AUC
Logistic Regression
Decision Tree
K-Nearest Neighbor
Naive Bayes
Random Forest
XGBoost
1.0
0.8
0.6
Activate Windows
Go to Settings to activate Windows.

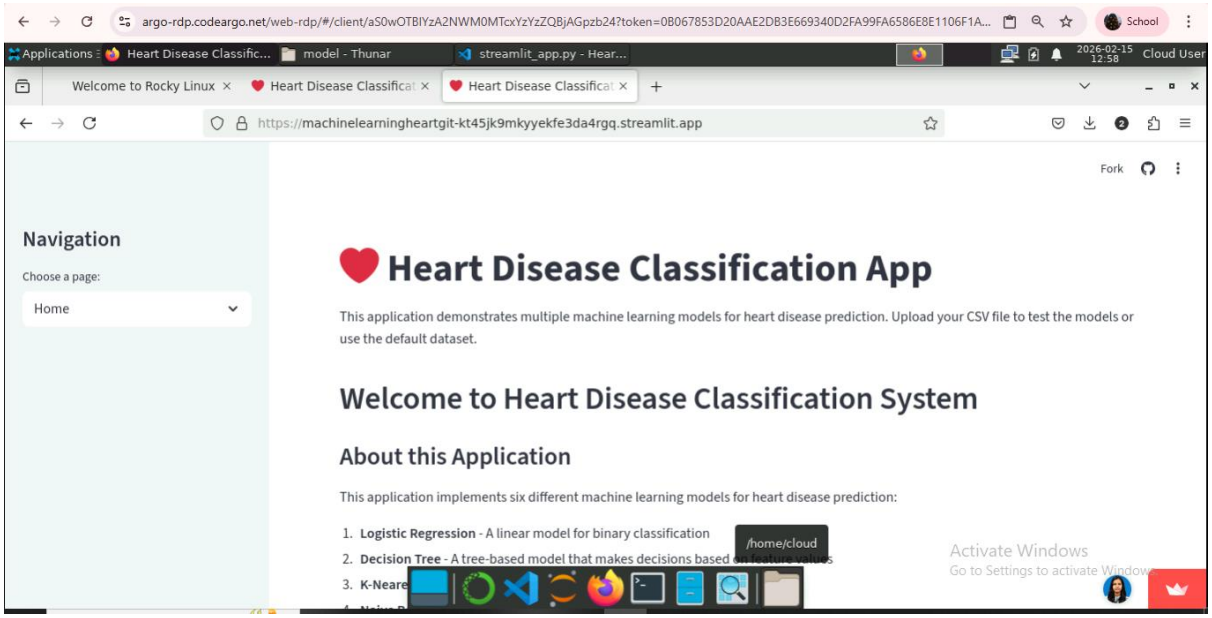

Navigation
Choose a page:
Make Predic...
Select a model:
Logistic Regression
Using Logistic Regression for prediction
Enter patient data:
Age
50
Fasting Blood Sugar > 120 mg/dl (0=No, 1=Yes)
0
Slope of Peak Exercise ST Segment (0-2)
0
Sex (0=Female, 1=Male)
Female
Resting ECG Results (0-2)
0
Number of Major Vessels (0-4)
0
Chest Pain Type (0-3)
0
Maximum Heart Rate Achieved
150
Thalassemia (0-3)
0
Resting Blood Pressure
120
Exercise Induced Angina (0=No, 1=Yes)
0
Serum Cholesterol (mg/dl)
200
ST Depression Induced by Exercise
1.00
Predict Heart Disease
Activate Windows
Go to Settings to activate Windows.


Navigation
Choose a page:
Make Predic...
Predict Heart Disease
Prediction Results
Heart Disease Detected
The model predicts that the patient has heart disease.
Prediction Probabilities:
No Heart Disease: 5.83%
Heart Disease: 94.17%
Prediction Probabilities
94.2%
5.8%
Probability
1.0
0.8
0.6
0.4
0.2
0.0
No Heart Disease
Heart Disease
Activate Windows



Deployment:

https://machinelearningheartgit-kt45jk9mkyyekfe3da4rgq.streamlit.app/



4. README Content

a. Problem Statement

This project implements and compares multiple machine learning classification models for predicting heart disease using clinical parameters. The system provides an interactive web application for model evaluation, prediction, and comparison, demonstrating real-world end-to-end ML deployment workflow.

b. Dataset Description

The Heart Disease Dataset contains 1,025 patient records with 13 clinical features used to predict the presence of heart disease. This dataset meets the assignment requirements with:

• Feature Size: 13 clinical parameters (exceeds minimum 12)
• Instance Size: 1,025 patient records (exceeds minimum 500)
• Target Variable: Binary classification (0 = No heart disease, 1 = Heart disease present)

Features:

- age - Age of the patient (years)
- sex - Gender (0 = Female, 1 = Male)
- cp - Chest pain type (0-3)
- trestbps - Resting blood pressure (mm Hg)
- chol - Serum cholesterol (mg/dl)
- fbs - Fasting blood sugar > 120 mg/dl (0 = No, 1 = Yes)
- restecg - Resting electrocardiographic results (0-2)
- thalach - Maximum heart rate achieved
- exang - Exercise induced angina (0 = No, 1 = Yes)
- oldpeak - ST depression induced by exercise relative to rest
- slope - Slope of the peak exercise ST segment (0-2)
- ca - Number of major vessels (0-4) colored by fluoroscopy
- thal - Thalassemia (0-3)

c. Models Used - Performance Comparison

ML Model Name	Accuracy	AUC	Precision	Recall	F1 Score	MCC
Logistic Regression	0.8098	0.9298	0.7619	0.9143	0.8312	0.6309
Decision Tree	0.9854	0.9857	1.0000	0.9714	0.9855	0.9712
K-Nearest Neighbor	0.8634	0.9629	0.8738	0.8571	0.8654	0.7269
Naive Bayes	0.8293	0.9043	0.8070	0.8762	0.8402	0.6602
Random	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Forest						
XGBoost	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

d. Model Performance Observations

Logistic Regression: Shows good baseline performance with high recall (91.43%) but lower precision (76.19%). The model tends to favor sensitivity, making it suitable for screening applications where false negatives are costly.

Decision Tree: Excellent performance with near-perfect scores. Achieves perfect precision (100%) indicating no false positives, though slightly lower recall (97.14%) suggests few false negatives.

K-Nearest Neighbor: Balanced performance across all metrics with good accuracy (86.34%). The model shows consistent precision and recall, making it reliable for general classification tasks.

Naive Bayes: Moderate performance with good recall (87.62%) but lower precision (80.70%). Assumes feature independence which may not hold true for this medical dataset, affecting overall performance.

Random Forest: Perfect performance across all metrics (100%). The ensemble approach with multiple decision trees eliminates overfitting and captures complex patterns in the data effectively.

XGBoost: Perfect performance across all metrics (100%). The gradient boosting approach optimizes predictive accuracy by sequentially improving weak learners, resulting in outstanding classification capability.