

Heart Disease Prediction Project Report

Abstract

This project predicts the likelihood of heart disease using patient health indicators. The model was trained on a heart disease dataset and deployed using Streamlit for interactive predictions.

Introduction

Heart disease is one of the leading causes of death worldwide. Predicting heart disease early can save lives and help in preventive measures. This project aims to build a machine learning model to predict heart disease based on patient data.

Dataset Description

- **Source:** Kaggle
 - **Number of rows and columns:** (918 , 12)
 - **Features:**
 - Age
 - Sex
 - Chest Pain Type
 - Resting Blood Pressure
 - Cholesterol
 - Fasting Blood Sugar
 - Resting ECG
 - Maximum Heart Rate
 - Exercise-Induced Angina
 - ST Depression (Oldpeak)
 - ST Segment Slope
-

Data Preprocessing & EDA

- Data cleaning and handling missing values
 - Feature encoding (categorical → numerical)
 - Exploratory Data Analysis: Histograms, Correlation plots, Visualizations
-

Model Training

Multiple machine learning models were applied, including Logistic Regression, KNN, Decision Tree, Gaussian Naive Bayes, SVM, Random Forest, AdaBoost and XGBoost.

The best model was **XGBoost Classifier** with an accuracy of **85%**, which was further fine-tuned to reach **89%** accuracy.

Model Evaluation

- Accuracy before tuning: 85%
 - Accuracy after tuning: 89%
 - The model predicts:
 - 0 → No Heart Disease
 - 1 → Heart Disease
-

Streamlit Web App

The model was deployed using **Streamlit**, creating an interactive web app with three tabs:

1. Single Prediction

Heart Disease Predictor

Predict Bulk Predict Model Information

Age (years)

45 - +

Sex

Male ▼

Chest Pain Type

Non-Anginal Pain ▼

Resting Blood Pressure (mm Hg)

120 - +

Serum Cholesterol (mm/dl)

170 - +

Fasting Blood Sugar

<=120 mg/dl ▼

Resting ECG Results

ST-T Wave Abnormality ▼

Maximum Heart Rate Achieved

60 - +

Exercise-Included Angina

Yes ▼

Oldpeak (ST Depression)

0.00 - +

Slope of Peak Exercise ST Segment

Up Sloping ▼

Submit

Result...

No Heart Disease Detected

2. Bulk Prediction via CSV

Heart Disease Predictor

Predict Bulk Predict Model Information

Upload CSV File

Upload a CSV File

Drag and drop file here
Limit 200MB per file • CSV

heart_sample_data.csv 413.0B

Predictions:

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAng
0	45	0	3	130	230	0	0	150	
1	54	1	2	140	260	1	1	135	
2	60	0	2	150	300	1	2	120	
3	39	1	0	120	210	0	0	170	
4	48	0	1	138	240	0	1	160	
5	65	1	2	160	320	1	2	110	
6	52	0	3	135	225	0	0	155	
7	58	1	1	145	280	1	1	140	
8	42	0	0	118	190	0	0	175	
9	70	1	2	170	350	1	2	105	

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	Prediction
0	45	0	3	130	230	0	0	150	0	1	0	0
1	54	1	2	140	260	1	1	135	1	2.5	1	1
2	60	0	2	150	300	1	2	120	1	3	2	1
3	39	1	0	120	210	0	0	170	0	0	0	0
4	48	0	1	138	240	0	1	160	0	1.2	1	0
5	65	1	2	160	320	1	2	110	1	4	2	1
6	52	0	3	135	225	0	0	155	0	0.8	0	0
7	58	1	1	145	280	1	1	140	1	2	1	1
8	42	0	0	118	190	0	0	175	0	0	0	0
9	70	1	2	170	350	1	2	105	1	5	2	1

3. Model Information

Heart Disease Predictor

[Predict](#) [Bulk Predict](#) [Model Information](#)

Model Information

Model Used: XGBoost Classifier

Overview:
This application uses an XGBoost Classifier to predict the likelihood of heart disease based on patient health indicators. XGBoost is an efficient and powerful gradient boosting algorithm known for its high performance on structured data.

Dataset:
The model was trained on a heart disease dataset containing medical attributes such as age, blood pressure, cholesterol, ECG results, and exercise-related features.

Input Features:

- Age
- Sex
- Chest Pain Type
- Resting Blood Pressure
- Cholesterol
- Fasting Blood Sugar
- Resting ECG
- Maximum Heart Rate
- Exercise-Induced Angina
- ST Depression (Oldpeak)
- ST Segment Slope

Prediction Output:

- 0 → No Heart Disease Detected
- 1 → Heart Disease Detected

Disclaimer:
This application is intended for educational purposes only and should not be used as a substitute for professional medical advice.

Conclusion

This project demonstrates a machine learning approach to predict heart disease. The XGBoost model provided high accuracy, and the Streamlit app enables interactive predictions. Future improvements could include:

- Using larger datasets
 - Adding more medical features
 - Calculating additional evaluation metrics
-

References

- Kaggle dataset: (insert link)
- Python Libraries: pandas, numpy, scikit-learn, xgboost, streamlit
- Streamlit documentation: <https://docs.streamlit.io>