Mini Project Report
on

Heart Disease Prediction Model

Submitted By
Pradhyumn Vijay Jadhao

Course Name: Machine Learning

Department of Computer Science and Engineering

(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)

Shri Ramdeobaba College of Engineering & Management, Nagpur 440013

(An Autonomous Institute affiliated to Rashtrasant Tukdoji Maharaj Nagpur University Nagpur)

October 2025

1. Problem Statement

Heart disease is a leading cause of mortality worldwide. Many of its risk factors are complex and interconnected, making it difficult for clinicians to accurately assess a patient's risk based on manual inspection of data alone. The lack of accessible and reliable predictive tools can lead to delayed diagnosis and treatment.

This project aims to solve this problem by developing a robust machine learning model that can accurately predict the presence of heart disease in a patient based on a set of standard medical attributes. The final model will serve as a valuable decision-support tool, helping to identify high-risk patients earlier.

2. Project Objectives

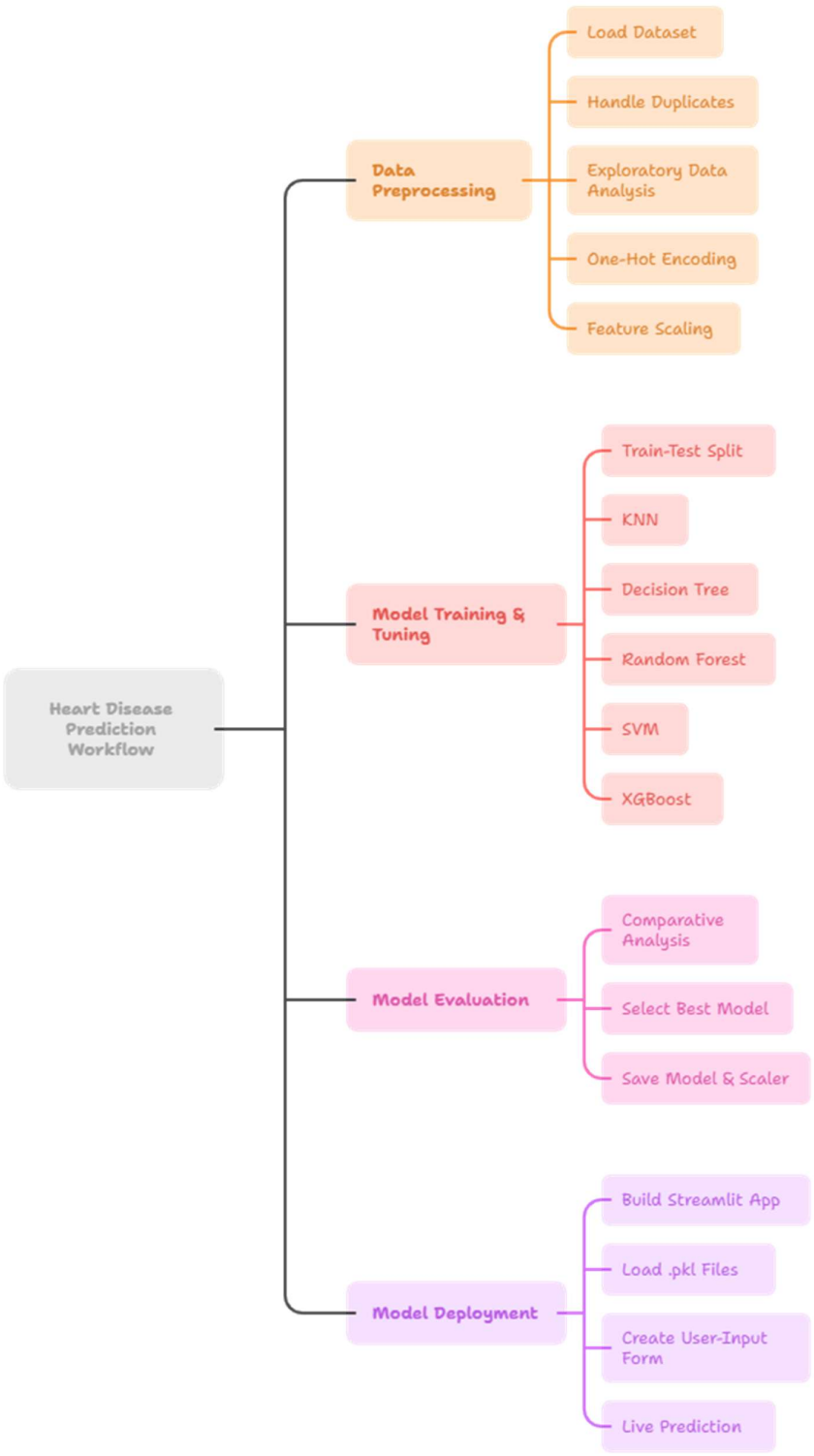
This project is guided by the following key objectives:

1. To perform comprehensive Exploratory Data Analysis (EDA) on the UCI Heart Disease dataset to identify key features and relationships.
2. To apply robust data preprocessing techniques, including one-hot encoding for categorical data and RobustScaler to handle clinical outliers, preparing the data for modeling.
3. To implement, train, and fine-tune multiple classification algorithms: K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost.
4. To conduct a comparative analysis of all trained models using metrics like Accuracy, Precision, Recall, and F1-Score to identify the single best-performing model.
5. To develop and deploy a functional, user-friendly web application using Streamlit that integrates the best-trained model, allowing users to input patient data and receive a real-time risk prediction.

3. Methodology

The project followed a structured machine learning pipeline, as illustrated in the flowchart below. The core of the methodology involved five distinct phases: Data Acquisition, Data Preprocessing, Model Training & Tuning, Model Evaluation, and Deployment.

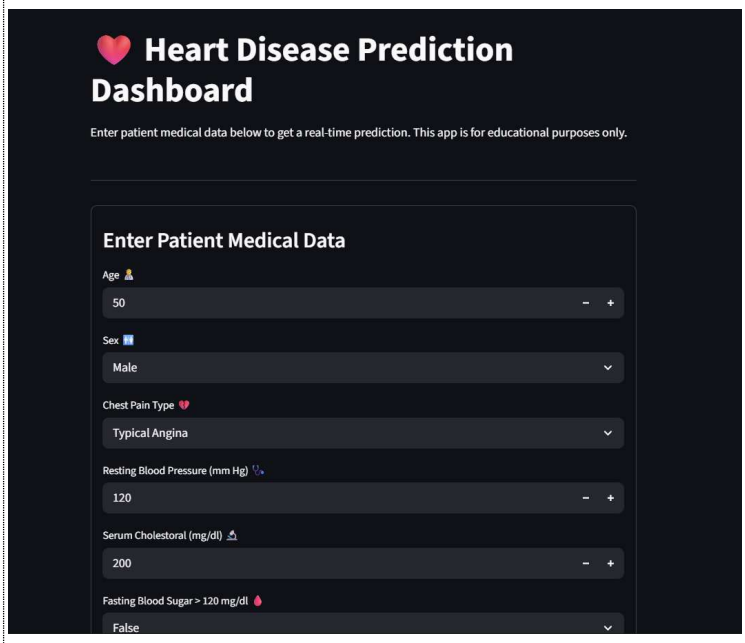
Heart Disease Prediction Workflow

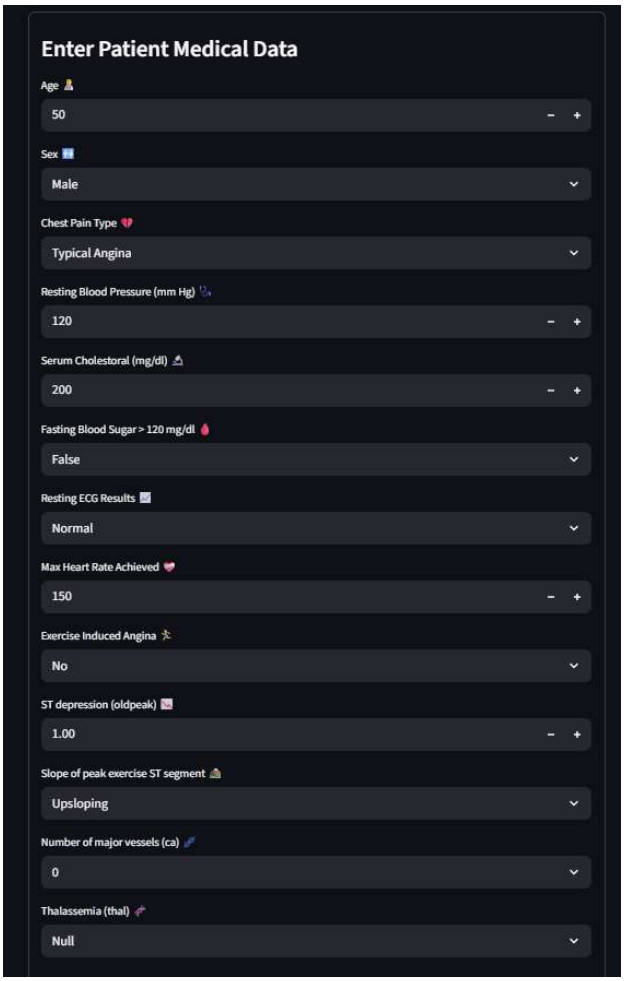


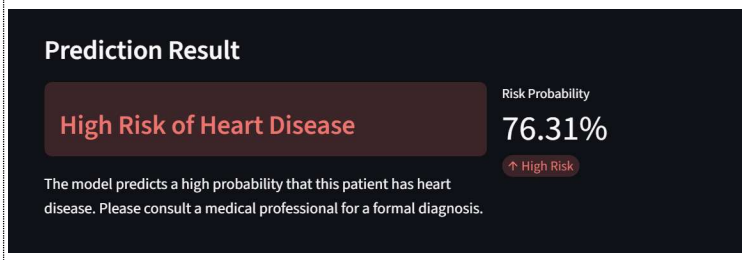
4. Technology Stack

- Programming Language: Python
- Data Analysis & Manipulation: Pandas, NumPy
- Machine Learning: Scikit-learn (for KNN, Decision Tree, Random Forest, SVM, Preprocessing, and Metrics), XGBoost
- Data Visualization: Matplotlib, Seaborn
- Development Environment: Google Colab
- Web App & Deployment: Streamlit
- Model Saving: Pickle

5. Result







6. Conclusion

This project successfully demonstrated the complete end-to-end development of a machine learning model for heart disease prediction. We performed thorough data preprocessing, compared multiple classification algorithms, and identified Random Forest as the most robust model for this dataset.

The project culminated in a functional and user-friendly Streamlit web application that can serve as a valuable tool for educational or illustrative purposes, demonstrating how ML can be applied to aid in medical decision-making. It is important to note that this tool is not a substitute for professional medical advice. Future work could involve training the model on a larger, more diverse dataset and exploring more complex deep learning models to further improve predictive accuracy.