

PROJECT: 3

Heart Disease Prediction

Name: Mani P

Email: maniprabu991@gmail.com

Program: Unified Mentor ML Internship Program

1. Problem Statement

Heart disease is one of the leading causes of death worldwide. Early detection is critical for effective treatment and prevention. However, diagnosing heart disease accurately requires analyzing multiple clinical parameters, which can be complex and time-consuming. This project aims to build a machine learning system that assists healthcare professionals by predicting the presence of heart disease based on patient clinical data.

2. Objective

The objective of this project is to develop a reliable machine learning model that can predict whether a patient has heart disease using structured patient health data such as vitals, lab test results, and ECG readings.

3. Tech Stack Used

- **Programming Language:** Python
- **Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
- **Machine Learning Algorithm:** Random Forest Classifier
- **Model Optimization:** GridSearchCV
- **Domain:** Healthcare / Medical Diagnosis

4. Project Architecture / Workflow

1. Dataset collection
2. Exploratory data analysis (EDA)

3. Data preprocessing and outlier handling
4. Feature selection and preparation
5. Model training using Random Forest
6. Hyperparameter tuning
7. Model evaluation
8. Heart disease prediction

5. Implementation Details

- Dataset contains patient clinical parameters such as age, cholesterol, blood pressure, ECG results, and exercise-induced angina
- Outliers were handled using the **IQR method**
- Feature scaling was not required as Random Forest is not sensitive to scale
- Dimensionality reduction techniques (PCA, LDA, RFE) were avoided to maintain interpretability
- Model trained using **Random Forest Classifier**
- Hyperparameters tuned using **GridSearchCV with 5-fold cross-validation**
- Model optimized for **Recall**, which is critical in medical diagnosis to reduce false negatives

Best Hyperparameters:

- criterion: entropy
- max_depth: None
- max_features: sqrt
- min_samples_split: 5
- min_samples_leaf: 1
- n_estimators: 100

6. Output / Results

Model Performance:

- **Accuracy:** 92.02%

Classification Report:

- Class 0 (No Heart Disease):
 - Precision: 0.91
 - Recall: 0.92
 - F1-Score: 0.92
- Class 1 (Heart Disease):
 - Precision: 0.93
 - Recall: 0.92
 - F1-Score: 0.92

Confusion Matrix:

[[103, 9],

[10, 116]]

Sample Prediction:

Input Patient Data:

- Age: 55
- Sex: Male
- Chest Pain Type: Atypical angina
- Blood Pressure: 140
- Cholesterol: 250
- ECG: ST-T abnormality

Prediction: No Heart Disease

Probability: 77.89%

7. Challenges Faced

- Handling outliers in sensitive medical data
- Selecting performance metrics suitable for healthcare use cases
- Preventing false negatives while maintaining good accuracy

8. Future Enhancements

- Integration with hospital management systems
- Use of advanced models such as XGBoost or LightGBM
- Adding SHAP for model explainability
- Deploying the system as a web application using Streamlit or Flask

9. Conclusion

This project successfully demonstrates how machine learning can assist in early detection of heart disease. By using a Random Forest model optimized for recall, the system provides accurate and reliable predictions, making it a valuable decision-support tool for healthcare professionals.