# Heart Disease Prediction using Machine Learning

## Internship Submission Report

**Internship Organization:** Digital Empowerment Network  
**Mentor:** Ali Mohiuddin Khan  
**Intern:** Shahzaib Shafique  
**Project Week:** Week 01  
**Date:** July 19, 2025

## 1. Project Title

**Heart Disease Prediction using Machine Learning Models**

## 2. Objective of the Project

The primary goal of this project was to build a machine learning model that can predict whether a person is likely to develop heart disease based on several clinical and demographic features. The problem addressed is a binary classification problem using a supervised learning approach.

## 3. Dataset Information

* **Source:** Open dataset
* **Total Records:** 1025
* **Features:** 'age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach', 'exang', 'oldpeak', 'slope' , 'ca', 'thal', 'target'
* **Target Variable:** Heart Disease (0 = No, 1 = Yes)

## 4. Technologies Used

* Python
* Pandas, NumPy
* Scikit-learn
* Imbalanced-learn (SMOTE)
* Joblib
* Matplotlib, Seaborn (for visualization)

## 5. Workflow Summary

### Step 1: Data Preprocessing

* Removed missing or inconsistent data
* Converted categorical variables into numeric using encoding techniques
* Normalized data using StandardScaler

### Step 2: Handling Imbalance

* The dataset was imbalanced with fewer positive heart disease cases.
* Used **SMOTE (Synthetic Minority Over-sampling Technique)** to balance the training data.

### Step 3: Train-Test Split

* 80% Training, 20% Testing using train\_test\_split with stratify=y for balanced class proportions.

### Step 4: Model Building

Three machine learning models were trained and evaluated: - **Logistic Regression** - **Decision Tree Classifier** - **Random Forest Classifier**

### Step 5: Hyperparameter Tuning

* Performed **GridSearchCV** on the Random Forest model to find the best parameters.

## 6. Evaluation Metrics

Used the following metrics for evaluation: - Accuracy - Precision - Recall - F1 Score - Confusion Matrix - Classification Report

## 7. Model Results

| Model | Accuracy | Precision | Recall | F1 Score |
| --- | --- | --- | --- | --- |
| Logistic Regression | 0.83 | 0.84 | 0.78 | 0.81 |
| Decision Tree | 0.67 | 0.7 | 0.5 | 0.58 |
| Random Forest | 0.83 | 0.87 | 0.75 | 0.80 |

## 8. Exported Files

The final trained models and preprocessing tools were exported using joblib: - logistic\_regression.joblib - decision\_tree.joblib - random\_forest.joblib - scaler.joblib

These can be reloaded and used directly for predictions.

## 9. Conclusion

This project successfully implemented a machine learning-based solution to predict heart disease. Among the three models tested, both Decision Tree and Random Forest showed excellent results. Random Forest was further optimized using GridSearchCV. The project demonstrated key ML workflow concepts including preprocessing, handling imbalance, model evaluation, and deployment preparation.

## 10. Implementation Context

The heart disease prediction model developed in this project can be effectively integrated into real-world healthcare applications to assist in early diagnosis and timely intervention. By incorporating this model into clinical decision support systems or digital health platforms, healthcare providers can input patient data—such as age, cholesterol levels, blood pressure, and other vital indicators—and receive a quick, data-driven prediction about the likelihood of heart disease. This can help doctors prioritize high-risk patients for further testing or specialist referrals, improving patient outcomes through proactive care. Moreover, the model can be embedded in remote health monitoring systems, enabling early warnings for at-risk individuals even outside clinical settings, particularly in under-resourced or rural areas. Overall, it has the potential to enhance diagnostic accuracy, reduce healthcare costs, and save lives through early detection.

## 11. GitHub Repository

All code, models, and project documentation are available on GitHub:  
[Shahzaib30/Heart-Disease-Prediction-using-Machine-Learning](https://github.com/Shahzaib30/Heart-Disease-Prediction-using-Machine-Learning)