

**Heart Disease Prediction Using Machine Learning**

Module: Special Topics in Software Engineering

# Team Members

Sarwar Mohammed (Team Leader)

Haryad Sabir

Bahez Mofaq

## Project Overview

This project is an end-to-end machine learning pipeline to predict the presence of heart disease in patients based on various health-related attributes. It includes data preprocessing, model training, and a user-friendly GUI application.

## Dataset

The dataset used contains 918 samples and 12 features related to cardiovascular health. Features include age, cholesterol levels and more. The target variable is Heart Disease, where 1 indicates presence and 0 indicates absence.

## Tools and Libraries

* Python
* Pandas, NumPy, Seaborn, Matplotlib
* Scikit-learn
* Tkinter for GUI
* Jupyter Notebook for EDA
* Pickle for model serialization

**Process Summary**

## 1. Data Preprocessing

Performed data cleaning and normalization:

* Removed missing values
* Scaled numerical features using StandardScaler
* Encoded categorical variables

## 2. Model Training

Trained a Random Forest Classifier:

* Train/test split: 80/20
* Achieved good accuracy (~85-90%) on test data
* Saved the model as model.pkl

## 3. GUI Application

Built with Tkinter:

* Loads the trained model
* Allows manual input of patient data
* Outputs a prediction (Heart Disease: Yes/No)

## Results

The Random Forest model provided high accuracy and robustness. The GUI enables non-technical users to interact with the model efficiently.

## Conclusion

This project demonstrates the full machine learning pipeline from raw data to deployment. It emphasizes clean code, user interaction, and practical deployment of ML models.