# Predicting Heart Disease Using Machine Learning

## Problem Statement

Heart disease is one of the leading causes of death worldwide. Predicting heart disease at an early stage can help in taking timely preventive measures. This project aims to predict the presence of heart disease in patients using a machine learning model. The dataset used is the Heart Disease dataset from the UCI Machine Learning Repository. The objective is to classify patients as having heart disease or not, based on various medical features.

## Our Approach

We use the Random Forest classification model to predict heart disease. The process involves loading the dataset, performing exploratory data analysis (EDA), preprocessing the data, training the model, and evaluating its performance. Additionally, we optimize the model using GridSearchCV to identify the best hyperparameters.

## Steps in the Project

1. \*\*Load Dataset\*\*: The Heart Disease dataset is loaded, and missing values are handled.

2. \*\*Exploratory Data Analysis (EDA)\*\*: We analyze the dataset to understand class distribution, correlation between features, and other patterns.

3. \*\*Data Splitting\*\*: The dataset is split into training and testing sets to evaluate the model on unseen data.

4. \*\*Model Training\*\*: A Random Forest Classifier is trained on the training data.

5. \*\*Evaluation\*\*: The model is evaluated using metrics like accuracy, recall, precision, F1 score, and AUC. A confusion matrix is plotted for further insights.

6. \*\*Hyperparameter Tuning\*\*: We use GridSearchCV to optimize the model's hyperparameters for better performance.

7. \*\*Final Evaluation\*\*: The best-tuned model is evaluated, and its performance is compared with the baseline.

## Conclusion

This project demonstrates the effectiveness of Random Forest in predicting heart disease. The tuned model achieved high accuracy, precision, recall, and F1 score, making it suitable for assisting medical professionals in diagnosis.