

Heart disease prediction using the **K-Nearest Neighbors (KNN)** algorithm is a machine learning approach that helps assess the likelihood of heart disease based on medical data. Here's a breakdown of how it works:

1. Understanding KNN for Heart Disease Prediction

- KNN is a **lazy learning algorithm** that classifies a data point based on its nearest neighbors.
- It works by calculating the **distance** between a new data point and existing labeled data points, then assigns the most common label among the nearest neighbors.

2. Dataset Overview

- The dataset typically includes features like:
 - **Age, Sex, Chest Pain Type (cp), Resting Blood Pressure (trestbps), Cholesterol (chol), Fasting Blood Sugar (fbs), Maximum Heart Rate (thalach), Exercise-Induced Angina (exang), and more.**
- The target variable is **binary** (1 = presence of heart disease, 0 = absence).

3. Data Preprocessing

- **Handling Missing Values:** Replace missing entries with the mean value of each column.
- **Feature Scaling:** Standardization using **StandardScaler** to ensure all features contribute equally to distance computations.
- **Data Splitting:** Typically, **67% for training** and **33% for testing** to evaluate performance.

4. Model Training

- Set `n_neighbors=5` (meaning it checks 5 nearest neighbors for classification).
- Use **Euclidean distance** to measure similarity between data points.

5. Model Evaluation

- **Accuracy Score:** Measures correct predictions.
- **Classification Report:** Includes precision, recall, and F1-score.
- **Confusion Matrix:** Helps visualize true positives, false positives, and false negatives.

6. Optimization & Future Enhancements

- **Hyperparameter Tuning:** Use **GridSearchCV** to find the best `k` value.
- **ROC Curve & AUC Score:** Evaluate model performance.
- **Feature Importance Analysis:** Identify key predictors like **chest pain type, maximum heart rate, and exercise-induced angina.**