

Project: Heart Disease Prediction using Logistic Regression

- **Objective:** Develop a machine learning model using Logistic Regression to predict the 10-year risk of Coronary Heart Disease (CHD) in patients based on health metrics.
 - **Dataset:** [Framingham Heart Disease Dataset](#)
-

Project Goals:

1. Importing Necessary Libraries and Dataset:

- **Load required Python libraries:**
 - Pandas for handling datasets.
 - NumPy for numerical operations.
 - Matplotlib/Seaborn for data visualization.
 - Sklearn for data preprocessing, model training, and evaluation.
- Load the dataset using Pandas and check its structure.
- Drop unnecessary columns (education).
- Rename columns for better readability.

2. Data Preprocessing:

- Handle missing values by removing rows with NaN values.
- Normalize numerical features using StandardScaler for uniform scaling.
- Convert categorical variables into numerical form if needed.
- Split dataset into training (70%) and testing (30%) sets.

3. Exploratory Data Analysis (EDA):

- Analyze class distribution of heart disease cases (CHD = 0 or 1).
- Visualize data distributions using histograms, count plots, and correlation heatmaps.
- Identify key risk factors such as age, cholesterol, smoking, blood pressure, and glucose levels.

4. Model Training using Logistic Regression:

- Define the Logistic Regression Model.

- **Train the model on the training dataset.**
- **Use Binary Cross-Entropy Loss as the loss function.**

5. Model Evaluation and Prediction:

- **Evaluate model performance using:**
 - **Accuracy Score**
 - **Precision, Recall, F1-Score**
 - **Confusion Matrix**
 - **ROC-AUC Curve**
 - **Predict the likelihood of heart disease in new patient data.**
-

Conclusion:

- **This model helps in early heart disease detection using key health indicators.**
- **Future improvements can include Ensemble Learning (Random Forest, XGBoost) for better accuracy.**