**Heart Disease Prediction Project**

**1. Aim**

To build a predictive system using machine learning to classify individuals as having heart disease or not, based on their health data.

**2. Objectives**

* Understand the dataset and its features.
* Preprocess the data by handling missing values, scaling features, and splitting the data.
* Train a classification model using **Logistic Regression**.
* Evaluate the model's performance on both training and testing datasets.
* Develop a system to predict heart disease for new input data.

**3. Theory**

Heart disease prediction involves classifying individuals into one of two categories:

* **1**: Defective heart (heart disease present).
* **0**: Healthy heart (no heart disease).  
  Machine learning models like Logistic Regression are well-suited for binary classification tasks. The model predicts the probability of an individual belonging to one of the two classes based on input features like age, cholesterol levels, blood pressure, and more.

**4. Data Collection and Processing**

* **Dataset**: The dataset was loaded from a CSV file (heart\_disease\_data.csv) and consisted of multiple health-related features and a target variable (target).
* **Exploratory Data Analysis**:
  + Inspected the first and last five rows of the dataset to understand its structure.
  + Checked for missing values: No missing values were found.
  + Statistical analysis revealed the ranges and distributions of the features.
  + Distribution of the target variable:
    - 1: Defective Heart
    - 0: Healthy Heart
* **Feature Selection**: The target column was separated as the dependent variable, and the remaining columns were used as independent variables.
* **Train-Test Split**: The data was split into training and testing sets (80%-20%), with stratification ensuring balanced class distributions.

**5. Model Training**

* **Algorithm Used**: Logistic Regression.
* **Training**: The model was trained using the training dataset (x\_train, y\_train).

**6. Model Evaluation**

* **Training Accuracy**: The accuracy of the model on the training dataset was computed.
* **Testing Accuracy**: The accuracy of the model on the testing dataset was computed to assess its generalization performance.
* **Results**:
  + Training Accuracy: **(insert computed accuracy here)**.
  + Testing Accuracy: **(insert computed accuracy here)**.

**7. Building a Predictive System**

* The predictive system was built to classify individuals based on input data.
* **Example Predictions**:
  + For input data (41, 0, 1, 130, 204, 0, 0, 172, 0, 1.4, 2, 0, 2):
    - Output: **The person has a heart disease**.
  + For input data (58, 0, 0, 170, 225, 1, 0, 146, 1, 2.8, 1, 2, 1):
    - Output: **The person does not have a heart disease**.

**8. Observations**

* Logistic Regression was effective in classifying individuals with high accuracy.
* The model demonstrated better accuracy on the training data, indicating it learned the patterns in the dataset well.
* The testing accuracy was slightly lower, but the model showed good generalization ability.

**9. Conclusion**

The project successfully built a heart disease prediction system with satisfactory accuracy. The system can be used to predict the likelihood of heart disease in individuals based on their health parameters. Future improvements could include experimenting with other machine learning algorithms, hyperparameter tuning, and feature engineering.