**Heart Disease Prediction**

**Introduction:**

Cardiovascular diseases represent a global health challenge, emphasizing the critical need for early detection and intervention. This project, "Heart Disease Detection using Logistic Regression," endeavors to contribute to proactive healthcare by employing logistic regression, a powerful tool for binary classification tasks. The primary objective is to predict the probability of heart disease in individuals based on a comprehensive set of health parameters, ultimately providing a valuable resource for healthcare professionals and individuals alike.

**Data Collection**:

Data Source The dataset, sourced from kaggle, will encompass a broad spectrum of health information, including age, sex, chest pain type, blood pressure, cholesterol levels, and other relevant features. The target variable will denote the presence or absence of heart disease. Data Features Key features include:

• Age

• Sex

• Chest pain type

• Resting blood pressure

• Cholesterol levels •

Fasting blood sugar

•Resting electrocardiographic results

• Maximum heart rate achieved

• Exercise-induced angina

• ST depression induced by exercise

• Number of major vessels colored by fluoroscopy

• Thall

**Data Collection and Processing:**

- Importing Libraries:

- NumPy and Pandas are imported for data manipulation and analysis.

- scikit-learn's `train\_test\_split` is utilized for splitting the dataset.

- Logistic Regression model, accuracy\_score for evaluation, and pickle for model serialization are imported.

- The dataset is loaded using Pandas from a CSV file named 'heart\_disease\_data.csv'.

- Features (X) and the target variable (Y) are separated.

- A stratified split is applied to ensure an even distribution of target classes in both the training and testing sets.

**Model Training:**

- Logistic Regression Model:

- Logistic regression is chosen as the predictive model due to its suitability for binary classification tasks.

- The logistic regression model is instantiated with parameters such as `max\_iter` and `random\_state` for control and reproducibility.

- The model is trained on the training data (`X\_train` and `Y\_train`).

- Model Evaluation:

- The script evaluates the logistic regression model's performance using accuracy scores.

- On the training set, predictions are made using `model.predict(X\_train)`.

- On the testing set, predictions are made using `model.predict(X\_test)`.

- Accuracy scores are calculated using `accuracy\_score` to quantify the model's effectiveness.

**Building a Predictive System:**

- Sample Input Data:

- A sample input data point is manually defined, simulating the data provided by a user.

- The logistic regression model is used to predict heart disease likelihood based on this input.

- Model Serialization:

- The trained logistic regression model is serialized using `pickle`.

- Serialization involves converting the model to a byte stream, which can be saved to a file.

- The serialized model is stored in 'Model.sav' for future use, allowing it to be easily loaded and reused without retraining.

---

Streamlit Web Application:

- Creating the Streamlit App:

- Streamlit is employed to build an interactive web application.

- Text input fields are provided for users to input various health parameters, mimicking a real-world scenario.

- Loading Trained Model:

- The Streamlit application loads the pre-trained logistic regression model using `pickle`.

- The model is loaded at the beginning of the application, ensuring quick and efficient predictions without the need for retraining.

- Predictions and Display:

- Upon clicking the 'Heart Disease Test Result' button, the application processes user input.

- The logistic regression model predicts the likelihood of heart disease based on the entered parameters.

- The application then displays the prediction, providing valuable information to the user.

This detailed breakdown offers a comprehensive understanding of the steps involved in data processing, model training, and the creation of an interactive web application using Streamlit. Feel free to ask for further clarification or additional details on specific aspects.

**Summary:**

In summary, this project aims to leverage logistic regression to create a robust model for predicting the likelihood of heart disease. By combining advanced machine learning techniques with a user-centric approach, the project seeks to provide accurate, interpretable, and accessible insights into heart disease risk. Through comprehensive data analysis, model development, and interface design, the project aspires to contribute to early detection and empower individuals to make informed decisions about their cardiovascular health.

**Made by- Harsh Pandey**

**Sec-M**

**Btech cse -5 th sem**

**Roll no.20**

**Uni roll no.2118601**

**Submitted to- Mr. Animesh Shrivastav**