OEL REPORT

GROUP MEMBERS:

NAME	ROLL NO
1. MUHAMMAD ALI	CS-21104
2. SAMEED FAIZ	CS-21127

Overview

In this project, we performed an in-depth analysis of the Heart Attack dataset, including Exploratory Data Analysis (EDA), data preprocessing, and model implementation using logistic regression and neural networks. This work was done using Streamlit for an interactive and user-friendly presentation, and the final project was hosted for easy access and demonstration.

Dataset Description

The Heart Attack dataset includes the following features related to patient heart health indicators:

- Age: Age of the patient
- Sex: Gender of the patient (1 = male, 0 = female)
- cp: Chest pain type (4 values)
- trestbps: Resting blood pressure (in mm Hg)
- chol: Serum cholesterol in mg/dl
- fbs: Fasting blood sugar > 120 mg/dl (1 = true, 0 = false)
- restecg: Resting electrocardiographic results (0, 1, 2)
- thalach: Maximum heart rate achieved
- exang: Exercise-induced angina (1 = yes, 0 = no)
- oldpeak: ST depression induced by exercise relative to rest
- slope: The slope of the peak exercise ST segment
- ca: Number of major vessels (0-3) colored by fluoroscopy
- thal: Thalassemia (3 = normal; 6 = fixed defect; 7 = reversible defect)
- target: Presence of heart disease (1 = yes, 0 = no)

Exploratory Data Analysis (EDA)

1. Data Summary

We began by loading the dataset and summarizing its structure:

- Shape: The dataset consists of 303 rows and 14 columns.
- Missing Values: There were no missing values in the dataset.
- Descriptive Statistics: We computed the mean, standard deviation, and other descriptive statistics for each feature.

2. Data Visualization

- Histograms: We plotted histograms for each feature to understand their distributions.
- Correlation Heatmap: We created a heatmap to visualize correlations between features and the target variable.
- Pairplot: A pairplot was used to explore relationships between features and the target variable.

3. Feature Relationships

- Age and Heart Attack: We observed that the likelihood of a heart attack increases with age.
- Sex and Heart Attack: Males had a higher prevalence of heart attacks compared to females.
- Chest Pain Type: Certain types of chest pain were more strongly associated with heart attacks.

Data Preprocessing

1. Encoding Categorical Variables

We encoded categorical variables such as `cp`, `restecg`, `slope`, `ca`, and `thal` using one-hot encoding.

2. Feature Scaling

We applied standardization to the dataset to ensure that all features contribute equally to the model's performance.

3. Splitting the Data

The data was split into training and testing sets using an 80-20 split.

Model Implementation

1. Logistic Regression

We implemented a logistic regression model to predict the presence of heart disease.

- Training: The model was trained using the training dataset.

- Evaluation: We evaluated the model using accuracy, precision, recall, and F1-score.

2. Neural Network

We designed and trained a neural network for the prediction task.

-Architecture: The neural network consisted of an input layer, two hidden layers with ReLU activation, and an output layer with sigmoid activation.

- Training: The model was trained using the training dataset with backpropagation and gradient descent optimization.

- Evaluation: The model was evaluated using the same metrics as the logistic regression model.

Results

Model Performance

-Logistic Regression:

- Accuracy: 84.2%
- Precision: 84.2
- Recall: 84.2
- F1-score: 84.2

- **Neural Network**:

- Accuracy: 86.8%
- Precision: 87
- Recall: 86.84
- F1-score: 86.84

The neural network model outperformed the logistic regression model in terms of accuracy and F1-score.

Streamlit Application

To make the project interactive and user-friendly, we developed a Streamlit application that allows users to explore the dataset, visualize relationships, and interact with the models. The application was hosted and made accessible for demonstration purposes.

Features of the Streamlit Application

- **Data Exploration**: Users can explore the dataset and view descriptive statistics.

- **Visualization**: Interactive plots and charts to visualize data distributions and relationships.

- **Model Predictions**: Users can input new data and get predictions from the logistic regression and neural network models.

Conclusion

This project successfully demonstrated the process of EDA, data preprocessing, and model implementation using logistic regression and neural networks on the Heart Attack dataset. The Streamlit application provided an interactive platform for users to engage with the analysis and predictions. Future work could involve optimizing the neural network architecture and exploring other advanced machine learning techniques to further improve prediction accuracy.