Healthcare

Course-end Project 2

Problem Statement Scenario:
Cardiovascular diseases are the leading cause of death globally. It is therefore necessary to identify the causes and develop a system to predict heart attacks in an effective manner. The data below has the information about the factors that might have an impact on cardiovascular health.

Summary:

This write-up presents a machine learning project focused on predicting the heart attacks based on the Cardiovascular dataset and develop a predictive model to identify potential cases of heart disease.

The Cardiovascular dataset consists of 303 samples, each represented by 14 attributes, including demographic, clinical, and laboratory features. These attributes include age, sex, cholesterol levels, resting blood pressure, maximum heart rate achieved, and presence of risk factors such as smoking, hypertension, and diabetes. The dataset also provides information about the presence or absence of heart disease, classified into two categories: 0 (no presence) and 1 (presence).

Methodology:

Data Preprocessing:

Handle missing values: Identify and handle any missing values in the dataset using suitable techniques such as imputation or removal.

Feature scaling: Normalize or standardize the numeric features to ensure all variables have a similar scale.

Feature Selection:

Perform exploratory data analysis to understand the relationship between variables and their importance in predicting heart disease.

Utilize techniques such as correlation analysis, feature importance, or domain knowledge to select relevant features for training the model.

Model Development:

The dataset was split into training and testing sets (e.g., 80:20 ratio).

Appropriate machine learning algorithm for heart disease prediction, such as logistic regression, random forests were chosen and performed.

The model was trained using the training set and evaluate its performance using suitable metrics (e.g., accuracy, precision, recall, and F1 score).

The model was fine tuned by adjusting hyperparameters, employing techniques like cross-validation, or considering ensemble methods to enhance performance.

Model Evaluation:

The model's performance was evaluated using the testing set and compare it against benchmark results or clinical standards.

The confusion matrix was analyzed to assess the model's ability to predict true positives, true negatives, false positives, and false negatives.

Conclusion:

Predicting heart disease using machine learning techniques can assist in early detection, risk stratification, and timely intervention. By analyzing the cardiovascular dataset, we demonstrated the development of a predictive model to identify individuals at risk of heart disease. Through proper data preprocessing, feature selection, and model development, we achieved promising results in predicting heart disease presence. However, it is crucial to continuously validate and improve such models using diverse datasets and collaborate with medical experts for effective integration into clinical practice.