Heart Disease Dataset

Abstract:

Heart disease remains a leading cause of mortality worldwide, necessitating a comprehensive understanding of its risk factors for effective prevention and management. This project aims to conduct an exploratory analysis of a heart disease dataset using various data science techniques. The dataset comprises demographic information, clinical attributes, and diagnostic test results of patients, collected from diverse sources.

The project begins with data preprocessing to handle missing values, outliers, and data imbalances. Exploratory data analysis (EDA) techniques such as data visualization and statistical summaries are employed to uncover patterns, correlations, and potential insights into the relationships between various risk factors and the occurrence of heart disease.

Machine learning models, including logistic regression, decision trees, random forests, and neural networks, are implemented to predict the likelihood of heart disease based on the available features. Model performance is evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score.

Furthermore, feature importance analysis is conducted to identify the most significant predictors of heart disease, aiding in risk stratification and personalized intervention strategies. The project concludes with recommendations for healthcare practitioners and policymakers based on the findings, emphasizing the importance of early detection and targeted interventions in mitigating the burden of heart disease.

Top of Form