

The Heart Disease Prediction Model

The Heart Disease Prediction Model project was a comprehensive exercise in predictive analytics, with the intent of diagnosing heart disease using various clinical parameters. The project utilized machine learning techniques to anticipate heart disease occurrence based on individual health metrics.

The dataset for this project was sourced from Kaggle, encompassing demographic information (age and sex), and clinical parameters, such as type of chest pain (cp), resting blood pressure (trtbps), serum cholesterol (chol), fasting blood sugar (fbs), and resting electrocardiographic results (restecg).

After data retrieval, an intensive data cleaning process ensued, with procedures like the removal of duplicates and null values, and data mapping. This essential phase ensured that the data quality was optimal for the subsequent analysis.

The project then progressed to the data preprocessing phase, which involved a robust exploratory data analysis. Univariate, bivariate, and multivariate analysis were performed to get an in-depth understanding of the data. Correlation analysis, data scaling using StandardScaler and MinMaxScaler, label encoding for categorical variables, and heatmaps for visualizing correlations were among the preprocessing tasks performed.

Once the data was adequately prepared, predictive analysis was performed using an array of sklearn machine learning models. These included Logistic Regression, Support Vector Classifier (SVC), Decision Trees, Random Forest, Gradient Boosting, k-Nearest Neighbors (KNN), Naive Bayes, and XGBoost. A model function was constructed and each model was looped through it to generate training and testing scores. The performance of each model was evaluated using confusion matrixes, providing detailed insight into the accuracy, precision, and recall of each model.

Finally, the performance of the models was compared with respect to accuracy and visually represented through plots. This allowed for a clear and understandable depiction of the best performing model for heart disease prediction.

The Heart Disease Prediction Model project successfully employed various data cleaning, preprocessing, and machine learning methods to predict heart disease occurrences, demonstrating the value of these tools in health-related predictive analytics.