**Title:**

Analysis of Heart Disease Prediction Data: Insights and Implications

**Objective:**

The primary objective of this analysis is to utilize data science modeling techniques to predict heart disease. By analyzing a comprehensive dataset that combines two sources of heart-related health data, we aim to identify key factors that contribute to heart disease and develop a predictive model that can assist healthcare professionals in early detection and prevention strategies.

**Dataset Overview:**

The dataset used in this analysis comprises data from two different sources, merged into a single dataset with standardized column names and formats. It includes various features such as age, sex, types of chest pain, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate, exercise-induced angina, and more. The combined dataset offers a robust foundation for developing a predictive model for heart disease.

Methodology :

* Data Preprocessing: Handling missing values, normalizing data, and ensuring consistency across the dataset.
* Exploratory Data Analysis (EDA): Analyzing the dataset to understand distributions, correlations, and patterns.
* Feature Selection: Identifying the most relevant features that contribute to the prediction of heart disease.
* Model Development: Building and testing various machine learning models such as logistic regression, decision trees, random forests, and gradient boosting to find the best performer.
* Model Evaluation: Assessing the model's performance using metrics like accuracy, precision, recall, and the ROC-AUC curve.

Questions That can be answered using the dataset:

* Age and Heart Disease: How does age affect the likelihood of having heart disease? Investigate the distribution of heart disease across different age groups to determine if certain age groups are more prone to heart disease.
* Gender Differences: Are there significant differences in the prevalence of heart disease between males and females? Analyze the data to see if one gender is more likely to have heart disease than the other.
* Chest Pain Type and Heart Disease: What is the relationship between different types of chest pain (e.g., typical angina, atypical angina, non-anginal pain, asymptomatic) and the occurrence of heart disease? Determine if certain types of chest pain are more indicative of heart disease.
* Blood Pressure, Cholesterol, and Heart Disease: How do high blood pressure and high cholesterol levels correlate with the risk of heart disease? Examine the relationship between RestingBP, Cholesterol levels, and the incidence of heart disease.
* Impact of Exercise-Induced Angina: Does the presence of exercise-induced angina increase the risk of heart disease? Analyze the data to understand the impact of exercise angina on heart disease risk.
* Electrocardiographic Results and Heart Disease: What is the association between different resting electrocardiographic results and heart disease? Study the relationship between RestingECG results (e.g., normal, having ST-T wave abnormality, showing probable or definite left ventricular hypertrophy) and the prevalence of heart disease.
* Impact of Fasting Blood Sugar on Heart Disease: How does fasting blood sugar (FBS) influence the risk of heart disease? Analyze whether individuals with a fasting blood sugar level above 120 mg/dl have a higher incidence of heart disease compared to those with lower levels.
* Role of Maximum Heart Rate in Predicting Heart Disease: Is there a relationship between maximum heart rate achieved during exercise and the likelihood of having heart disease? Investigate how different ranges of maximum heart rate correlate with the presence or absence of heart disease.
* Influence of ST Depression Induced by Exercise on Heart Disease Risk: What is the significance of ST depression (measured in the Oldpeak variable) in predicting heart disease? Study the relationship between the degree of ST depression during exercise and the prevalence of heart disease.
* Analysis of Thallium Stress Test Results: How do different results from the thallium stress test (represented in the 'Thallium' column) relate to heart disease? This test assesses blood flow to the heart, and varying results may indicate different levels of risk for heart disease.

Illustrations:

Below presenting few illustrations on how different factors are important or play a role in detecting heart disease.

Age Distribution by Heart Disease Status

A graph of age distribution

Description automatically generated

Illustration 2: Gender vs Heart Disease

A graph of a person with different colored squares

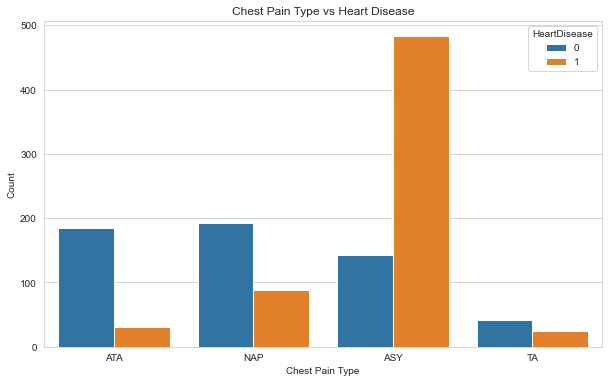
Description automatically generated with medium confidence

Illustration 3: Chest Pain Type vs Heart Disease

A graph of different colored squares

Description automatically generated

Illustration 4: Resting Blood Pressure Distribution by Heart Disease Status



Key Findings:

* Age, gender, types of chest pain, blood pressure, cholesterol levels, and exercise-induced angina are among the significant predictors of heart disease.
* Machine learning models demonstrate varying degrees of effectiveness, with some models showing high accuracy in predicting heart disease.
* The analysis reveals that certain combinations of factors significantly increase the likelihood of heart disease, providing valuable insights for targeted prevention strategies.
* Early Detection: The predictive model can assist in the early identification of individuals at high risk of developing heart disease, allowing for timely intervention.
* Personalized Care: Insights from the model can guide healthcare providers in tailoring preventive measures and treatments to individual patient profiles.
* Public Health Strategy: The analysis can inform public health initiatives focused on reducing the prevalence of heart disease risk factors in the population

Ethical Use of AI in Healthcare:

* Consent and Autonomy: Patients should be informed about how AI and predictive modeling may be used in their healthcare and given the choice to consent or opt-out.
* Avoiding Over-reliance: Caution should be exercised to avoid over-reliance on the model, recognizing that it is one of many tools available for healthcare decision-making.

References:

<https://www.kaggle.com/code/tanmay111999/heart-failure-prediction-cv-score-90-5-models/input?select=heart.csv>

<https://www.kaggle.com/code/caesarmario/listen-to-your-heart-a-disease-prediction>

<https://www.cdc.gov/heartdisease/risk_factors.htm>

<https://www.nhlbi.nih.gov/health/heart-healthy-living/risks>