**Predicting Heart Diseases using KNN**

**Contents**

1. Introduction
2. Required Libraries
3. Exploratory Data Analysis
   * Loading the Dataset
   * Dataset Overview
   * Correlation Matrix
   * Visualizing Correlations
   * Target Variable Distribution
4. Data Preprocessing
   * Preparing the Dataset
5. Model Development and Training
   * K-Nearest Neighbors (KNN)
   * Random Forest Classifier
6. Conclusion

**Introduction**

Heart diseases include a variety of conditions, such as:

1. Blood vessel disease
2. Heart rhythm problems
3. Congenital heart defects (present at birth)

In this project, we leverage Machine Learning techniques to analyze a Heart Disease dataset and classify whether a person is suffering from heart disease.

**Note**

Sections in **purple text** throughout this project present carefully constructed reports. These are derived directly from statistical analyses (e.g., t-tests, chi-square tests) and graphical visualizations.

**Required Libraries**

We begin by importing essential libraries for data manipulation, visualization, and model building.

**Exploratory Data Analysis**

**Loading the Dataset**

The dataset is loaded into a Pandas DataFrame to inspect its structure and contents.

**Dataset Overview**

General information and descriptive statistics of the dataset are displayed to understand the data's shape, types, and summary metrics.

**Correlation Matrix**

A correlation matrix is generated to examine relationships between features.

**Visualizing Correlations**

A heatmap is plotted to visualize the strength and direction of correlations among features. Scatterplots and pairwise comparisons further illustrate these relationships.

**Target Variable Distribution**

The distribution of the target variable is analyzed to check for balance between classes.

**Data Preprocessing**

**Preparing the Dataset**

Categorical variables are converted into dummy variables, and numerical features are scaled to standardize the data. The dataset is then split into fea