Dataset Overview

The dataset titled **"Datasetapproved.csv.xlsx"** contains medical records of **1,025 patients**, each described by **14 features** such as age, sex, chest pain type, resting blood pressure, cholesterol level, maximum heart rate, and other cardiovascular indicators. The final column, **`target`**, indicates whether a patient has **heart disease (1)** or **not (0)**. There are no missing values, and the dataset includes both numerical and categorical data types.

Purpose of the Project

The goal of this project is to **analyze and predict the likelihood of heart disease** based on patient health attributes. Using this dataset, we plan to explore the relationship between various medical features and the presence of heart disease, identifying which factors contribute most to cardiovascular risk.

Objectives

1. **Data Exploration and Visualization:**

    - Perform statistical analysis and visualize distributions of key variables (e.g., cholesterol, age, blood pressure).
    - Identify patterns and correlations between health indicators and heart disease outcomes.

2. **Data Preprocessing:**
    - Handle categorical encoding, normalization, and feature selection.
    - Prepare the dataset for machine learning modeling.

3. **Model Development:**
    - Build and compare classification models such as **Logistic Regression**, **Decision Tree**, and **Random Forest** to predict heart disease.
    - Evaluate model performance using accuracy, precision, recall, and F1-score metrics.

4. **Outcome:**
    - Develop a predictive system that can classify new patient data as "at risk" or "not at risk" for heart disease.
    - Summarize insights that could help in early disease detection and preventive healthcare decisions.