Project Summary: Heart Disease Prediction Using Machine Learning

Objective:
To develop a predictive model that can accurately determine whether a patient is likely to have heart disease based on clinical features.

Tools & Technologies:
Python, Pandas, Scikit-learn, Seaborn, Matplotlib, Logistic Regression, Random Forest, KNN, SVM

Dataset:
Heart disease dataset containing patient data such as age, sex, chest pain type, cholesterol, resting blood pressure, etc.

Approach:

- Conducted **Exploratory Data Analysis (EDA)** to identify trends and correlations.

- Preprocessed data (missing values, scaling, encoding) to prepare it for modeling.

- Trained multiple machine learning models including:

  - Logistic Regression (accuracy: 85.25%)

  - Random Forest

  - K-Nearest Neighbors (best accuracy: **91.80%**)

  - SVM and Naive Bayes

- Evaluated models using accuracy, precision, recall, F1-score, and confusion matrix.

- Visualized results and comparisons using heatmaps, bar plots, and pairplots.

Outcome:
K-Nearest Neighbors provided the highest accuracy, but Logistic Regression remained the most interpretable. The model provides a valuable **decision-support tool for doctors and healthcare providers** to identify potential heart disease cases early.

Future Enhancements:

- Deploy the model as a web or mobile application using Streamlit or Flask.

- Integrate real-time data from wearable devices.

- Use advanced ensemble models like XGBoost or LightGBM for improved performance.