Heart Disease Prediction Report

Main Objective of the Analysis

The main objective of this analysis is prediction-building a model that can accurately predict whether an individual is at risk of heart disease based on several clinical and lifestyle attributes.

The predictive model benefits stakeholders by enabling risk stratification in patient populations, supporting diagnostic decision-making processes, and informing public health strategies.

Dataset Description

The dataset contains 303 observations with 14 attributes, including:

- Age, Sex, Chest Pain Type, Resting BP, Cholesterol, Fasting Blood Sugar, Resting ECG, Max Heart Rate, Exercise-Induced Angina, ST Depression, Slope, and the Target variable.

The goal is to predict the 'Target' variable (1 = heart disease, 0 = no disease).

Data Exploration and Preprocessing

No missing values were found in the dataset. The target variable is moderately imbalanced. Categorical variables were encoded, and numerical features scaled. Feature engineering involved one-hot encoding and standardization to prepare the data for training.

Classifier Models

Three classifiers were trained with a 70-30 train-test split:

1. Logistic Regression: Accuracy ~84%, easy to interpret.
2. Random Forest: Accuracy ~90%, excellent feature importance.
3. Support Vector Machine: Accuracy ~88%, strong precision but less interpretable.

Recommended Model

The Random Forest Classifier is recommended as the final model due to its high accuracy (~90%) and

interpretability. It provides a strong balance between performance and insights.

Key Findings and Insights

- Chest pain type and exercise-induced angina are top predictors.
- Patients with lower max heart rate and higher oldpeak have greater risk.
- Sex and age show moderate impact.
- Cholesterol is less predictive after accounting for other variables.

Suggested Next Steps

- Add features such as BMI, smoking status, and family history.
- Expand the dataset for better generalization.
- Experiment with advanced models (e.g., XGBoost).
- Use SHAP values for clinical interpretability.