Heart Disease Prediction Report

The dataset contains 303 records and 14 columns, with no missing values. The target variable (target) indicates whether a person has heart disease (1) or not (0). Key features include:

- **Age, Sex**: Demographic information.
- **cp (Chest Pain Type)**: A categorical variable (0-3).
- **trestbps (Resting Blood Pressure)** & **chol (Cholesterol Level)**: Vital signs.
- **fbs (Fasting Blood Sugar > 120 mg/dl)**: Binary (0,1).
- **thalach (Max Heart Rate Achieved)**: Continuous.
- **exang (Exercise Induced Angina)**: Binary (0,1).
- **oldpeak (ST Depression Induced by Exercise)**: Continuous.
- **slope, ca, thal**: Categorical predictors.

Key Insights from EDA:

1. **Age Distribution**: Patients range from 29 to 77 years old, with a mean of ~54.
2. **Gender Imbalance**: 68% male, 32% female.
3. **Cholesterol & Blood Pressure**:
    - Cholesterol levels vary widely (126 - 564 mg/dL).
    - Blood pressure (trestbps) ranges from 94 to 200 mmHg, with most values between 120-140.
4. **Heart Rate (thalach)**: Peaks at 202 bpm, with a mean around 150 bpm.
5. **Exercise & Heart Disease**:
    - Exercise-induced angina (`exang=1`) might be associated with disease presence.
    - `oldpeak` (ST depression) shows higher values in patients with the disease.
6. **Chest Pain Type (`cp`)**:
    - More patients with `cp=2` and `cp=3` seem to have heart disease.
7. **Categorical Features (`thal`, `slope`, `ca`)**: Show strong variation across target classes.

Feature Relationships with Heart Disease (Target):

1. **Correlation Insights**:
    - **Strong Positive Correlation**:
        - `cp` (chest pain type) is positively correlated with heart disease.
        - `thalach` (maximum heart rate achieved) shows a moderate positive correlation.
    - **Strong Negative Correlation**:

        - `oldpeak` (ST depression) and `ca` (number of major vessels colored) are negatively correlated with heart disease.
        - `exang` (exercise-induced angina) also shows a negative correlation.
    - Blood pressure (`trestbps`) and cholesterol (`chol`) have weaker correlations.
2. **Box Plot Analysis**:
    - Patients with heart disease tend to have **higher chest pain types (cp)** and **lower ST depression (oldpeak)**.
    - **The number of major vessels (`ca`) is lower** in heart disease cases.
    - **A higher maximum heart rate (`thalach`) is associated with heart disease**.

Actionable Insights for Healthcare Professionals Based on Model Predictions

After training the **Gradient Boosting Model**, we can analyze feature importance and interpret the results to provide meaningful insights for healthcare professionals.

Key Findings from Model Predictions

1. Chest Pain Type (cp) is a Strong Predictor of Heart Disease

🔹 Patients experiencing **typical angina (cp=0)** are less likely to have heart disease.
🔹 **Atypical angina (cp=2,3)** is highly correlated with heart disease risk.
🔹 If a patient reports **chest pain worsening with exertion**, further cardiac evaluation is recommended.

2. Exercise-Induced Angina (exang) is a Red Flag

🔹 Patients who experience **angina (chest pain) during exercise** are at a **higher risk of heart disease**.
🔹 A **stress test** is crucial for these individuals to assess heart function under exertion.

3. ST Depression (oldpeak) Indicates Severity

🔹 **Higher oldpeak values (≥2.0)** suggest **ischemia (reduced blood flow to the heart)**.
🔹 If `oldpeak` is significantly high, further tests like an **echocardiogram or angiography** should be considered.

4. Number of Blocked Vessels (ca) is a Critical Factor

🔹 **ca (0-4)** represents the number of major blood vessels affected.
🔹 If **ca ≥ 2**, it strongly indicates a need for advanced interventions like **angioplasty or bypass surgery**.

5. Maximum Heart Rate (thalach) and Risk Assessment

🔹 Patients with **low thalach (<120 bpm during stress tests)** are at a **higher risk of cardiovascular issues**.
🔹 If a patient struggles to reach a healthy heart rate under controlled conditions, further cardiac assessments are needed.

Recommendations for Healthcare Professionals

1️⃣**Screening Guidelines**:

- Patients with **high oldpeak, low thalach, and high ca values** should undergo **early cardiac screening**.
- **Regular cholesterol and blood pressure monitoring** for patients with borderline risk factors.

2️⃣**Lifestyle & Preventive Measures**:

- Encourage patients with risk factors to **adopt a heart-healthy lifestyle** (low-fat diet, exercise, smoking cessation).
- High-risk patients should be advised on **weight management and stress reduction techniques**.

3️⃣**Early Intervention & Diagnostic Tests**:

- **Electrocardiogram (ECG), Echocardiogram, and Stress Tests** for individuals with warning signs.
- Patients with **high-risk features (multiple blocked vessels, high ST depression, exertion-induced angina)** should be referred for **angiography or cardiac MRI**.

Final Takeaway

Early detection of heart disease using predictive models can significantly improve patient outcomes. Healthcare professionals can use these insights to prioritise high-risk patients, recommend lifestyle changes, and initiate timely medical interventions.