Heart Disease Prediction Using Logistic Regression

This project focuses on predicting heart disease using logistic regression, analyzing the **Heart Disease UCI dataset** with **303 entries** and **14 features**. The model achieved an **accuracy of 85.25%**, demonstrating strong predictive capability.

Key Insights & Analysis

1. **Dataset Overview**
    - **Class Distribution**:
        - **54.5% (165 patients)** had heart disease (target = 1).
        - **45.5% (138 patients)** were healthy (target = 0).
    - **Feature Correlation**:
        - Chest pain type (cp) and maximum heart rate (thalach) showed strong correlation with heart disease.
        - Age and cholesterol (chol) had moderate influence.
2. **Exploratory Data Analysis (EDA)**
    - **Visualizations**:
        - **Countplot**: Confirmed balanced distribution of heart disease cases.
        - **Boxplot**: Revealed that younger patients (median age ~50) were more prone to heart disease than older ones (median age ~55).
        - **Heatmap**: Highlighted key correlations—e.g., negative correlation between thalach (max heart rate) and disease presence.
3. **Model Performance**
    - **Logistic Regression** was trained on **80% of data (242 samples)** and tested on **20% (61 samples)**.
    - Despite a **convergence warning**, the model performed well with **85.25% accuracy**.
4. **Predictive System**
    - A sample input ([62, 0, 0, 140, 268, 0, 0, 160, 0, 3.6, 0, 2, 2]) was classified as **healthy (0)** with the model outputting: *"You are healthy."*

Conclusion & Recommendations

- The model effectively predicts heart disease, with **chest pain type (**cp**)** and **heart rate (**thalach**)** being the most influential features.
- **Improvement Suggestions**:
    - Address convergence issues by scaling features or increasing max_iter.
    - Experiment with other models (e.g., Random Forest, SVM) for comparison.
    - Deploy as a **web/mobile app** for real-time health assessments.

The visualizations (countplot, boxplot, heatmap) provided clear insights into feature importance and disease distribution, reinforcing the model's reliability.

Next Steps:

- Expand dataset for better generalization.
- Optimize hyperparameters for higher accuracy.
- Develop a user-friendly interface for medical diagnostics.