

1. TASK 2

2. Features:

- age, sex, cp (chest pain type), trestbps (resting blood pressure), chol (cholesterol), fbs (fasting blood sugar), restecg (resting ECG results), thalach (maximum heart rate achieved), exang (exercise-induced angina), oldpeak (ST depression), slope (slope of peak exercise ST segment), ca (number of major vessels), thal (thalassemia).

3. Target:

- target: Indicates the presence of heart disease (1) or absence (0).

Next Steps:

1. Perform **Exploratory Data Analysis (EDA)** to understand feature distributions and correlations.
2. Preprocess the data (e.g., scaling, encoding if necessary).
3. Train models: Logistic Regression, Random Forest, XGBoost.
4. Evaluate using accuracy, precision, recall, and F1-score.

Key Insights:

1. Performance Comparison:

- XGBoost and Random Forest models outperform Logistic Regression in predicting readmissions.
- XGBoost achieved the highest **ROC-AUC score**.

2. Feature Importance:

- XGBoost and Random Forest provide insights into feature importance. Key features like age, cholesterol levels, and other patient vitals strongly influence predictions.

3. Recommendations:

- Deploy the XGBoost model for identifying high-risk patients.
- Focus on optimizing critical patient features to reduce readmissions.
- Future improvements: Include more features like treatment duration, comorbidities, and demographics for better predictions.

Potential Use Cases:

- Hospitals can use this model to allocate resources efficiently for high-risk patients.
- Improved patient follow-ups based on model predictions to reduce readmission rates.

