# Report – Week 01: Heart Disease Prediction

## 1. Introduction

Heart disease is one of the leading causes of death worldwide. Early prediction using machine learning can help doctors make informed decisions. This study uses the UCI Heart Disease dataset to build models that classify patients as having or not having heart disease.

## 2. Dataset

- Source: UCI Machine Learning Repository (Dataset ID: 45)  
- Attributes:  
 • Numerical: age, cholesterol, resting blood pressure, maximum heart rate, ST depression, etc.  
 • Categorical: sex, chest pain type, fasting blood sugar, exercise-induced angina, etc.  
- Target Variable: Presence (1) or absence (0) of heart disease.

## 3. Methodology

### 3.1 Data Preprocessing

- Handled missing values in 'ca' and 'thal' using median imputation.  
- Normalized numerical features using Min-Max scaling.  
- Encoded categorical variables with Label Encoding (binary) and One-Hot Encoding (multi-category).

### 3.2 Exploratory Data Analysis (EDA)

- Checked dataset statistics.  
- Correlation heatmap revealed strong relationships between certain features and heart disease.

### 3.3 Model Training

- Train-test split: 80/20.  
- Models used:  
 • Logistic Regression (baseline).  
 • Random Forest Classifier (ensemble).

### 3.4 Model Evaluation

Metrics considered: accuracy, precision, recall, F1-score, confusion matrix.

## 4. Results

- Logistic Regression: Accuracy ~86.9%, Recall ~84.4%, F1 ~0.871.  
- Random Forest: Accuracy ~88.5%, Recall ~87.5%, F1 ~0.889.  
- Random Forest misclassified fewer patients with heart disease compared to Logistic Regression.

## 5. Conclusion

- Both models performed well, but Random Forest achieved better accuracy, recall, and F1-score.  
- Random Forest is selected as the final model for this dataset.  
- Future work: Explore hyperparameter tuning and additional algorithms such as Gradient Boosting and SVM for further improvement.

## 6. References

- UCI Machine Learning Repository: Heart Disease Dataset  
- Scikit-learn Documentation: https://scikit-learn.org