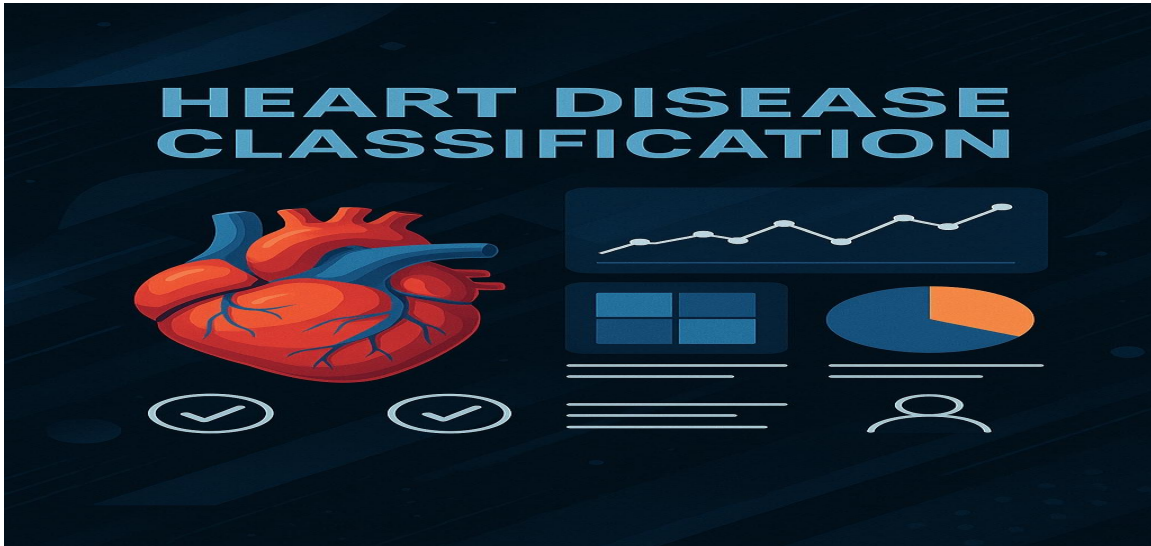


Task Statement

Project Title:

CardioNet: Intelligent System for Heart Disease Prediction



Objective:

You are required to refine and enhance the existing classification project by applying advanced feature selection, feature engineering, and evaluation techniques. This assignment will help you explore how different techniques impact model performance and build a deeper understanding of how to justify model evaluation in real-world health prediction scenarios.

Part 1: Feature Selection Comparison

1. Apply at least five feature selection techniques in addition to SelectKBest. You may choose from:

- Recursive Feature Elimination (RFE)
- Mutual Information
- Tree-based feature importance (e.g., from Random Forest or XGBoost)
- L1 Regularization (Lasso)
- Variance Threshold
- Boruta (optional)

2. Compare selected features from each method:

- Identify and list selected columns from all five techniques
- Determine the intersection of features selected by all techniques
- Train a model using only those common selected features

3. Evaluate model performance using Accuracy, F1 Score, and ROC-AUC
4. Plot the ROC Curve of the best model on this refined feature set

Part 2: Feature Engineering

1. Apply at least two additional feature engineering techniques, such as:
 - Polynomial or interaction terms
 - Binning continuous variables
 - Creating domain-based ratios or transformations
 - Clustering-based group labels (e.g., KMeans bins)
2. Retrain the model using engineered features (with or without previously selected ones)
3. Compare the model's performance before and after feature engineering
4. Provide a brief explanation of what helped and why

Part 3: Evaluation Metric Justification

1. Based on your experiments and the nature of the healthcare use case, identify:
 - The most appropriate evaluation metric (Accuracy, Precision, Recall, F1 Score, ROC-AUC)
 - Provide a justification with examples (e.g., cost of false negatives in heart disease detection)

Deliverables

You must submit the following:

- Jupyter Notebook or Python script with well-commented code
- Comparison table of feature selection results
- Before & after model evaluation results
- ROC curve plot(s)
- Final reflection on feature engineering
- Written justification for the chosen evaluation metric

Grading Criteria (Total: 25 Marks)

Component	Marks
Feature Selection Comparison (5 methods)	5
Common feature intersection + ROC Curve	5
Feature Engineering Implementation	5
Evaluation Metric Justification	5
Code Quality + Explanation & Structure	5