

## Experiment 7

### Objective: Feature Selection and Engineering on the Heart Disease Dataset in R

#### Dataset:

##### Cleveland Heart Disease Dataset

- Source: [UCI Machine Learning Repository](#) or directly available via mlbench or caret packages in R.
  - Contains medical records with features like age, sex, chest pain type, cholesterol level, fasting blood sugar, etc.
  - Target variable: presence of heart disease.
- 



#### Experiment Goals:

1. Load and explore the dataset.
    1. `data("PimaIndiansDiabetes2")` # Alternatively use Heart dataset from UCI
  2. Perform feature engineering (categorical encoding, scaling, transformations).
    1. Normalize numerical features
    2. Create a new feature: BMI-to-Age Ratio
    3. Correlation matrix
  3. Use correlation, recursive feature elimination (RFE), and LASSO for feature selection.
    1. Feature Selection - Recursive Feature Elimination
  4. Train and compare model performance (with and without engineered/selected features).
    1. Train model with selected features
    2. "glucose", "mass", "bmi\_age\_ratio"
-

**Expected Outcomes:**

- Visualization of correlation between features.
- Identification of top contributing features using RFE and LASSO.
- Impact of engineered features (like bmi\_age\_ratio) on model performance.
- Practice on using caret, glmnet, and correlation visualization.