Name: Subodh Khanduri Roll No: 24IM61R02 Date: 20/01/25

1. Laboratory project details (Problem Statement):

1. Assignment2: Use linear regression model trained above, and apply both L1 and L2 regularization to eliminate overfitting in the model. Analyze the effectiveness of the above two regularization schemes.
2. Assignment3: Use the following dataset to develop a decision tree model to predict accidents occuring due to stormy weather in the current month. Apply L1 and L2 regularization to eliminate overfitting, and measure AICC, BIC metrics described in the second link given below.

2. Language Chosen: Python

3. Framework and libraries used to implement: Numpy, Pandas, Matplotlib, Scikit-learn

4. Dataset used (link): https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-red.csv

And California housing.

5. Codebase Developed (please share files using google drive or github):

6. Results obtained:   
a. Basic Model

Linear Regression Result

Training RMSE: 0.719676

Testing RMSE: 0.745581

Training R^2: 0.612551

Testing R^2: 0.575788

Ridge Regression Result

Training RMSE: 0.719676

Testing RMSE: 0.745579

Training R^2: 0.612551

Testing R^2: 0.575791

Lasso Regression Result

Training RMSE: 0.819658

Testing RMSE: 0.824396

Training R^2: 0.497419

Testing R^2: 0.481361

ElasticNet Regression Result

Training RMSE: 0.792002

Testing RMSE: 0.797406

Training R^2: 0.530762

Testing R^2: 0.514765

b. Logistic Regression Accuracy 0.72

Classification Report for Logistic Regression :

precision recall f1-score support

0 0.67 0.76 0.71 213

1 0.78 0.70 0.74 267

accuracy 0.72 480

macro avg 0.72 0.73 0.72 480

weighted avg 0.73 0.72 0.72 480

Decision Tree Accuracy 0.73

Classification Report for Decision Tree :

precision recall f1-score support

0 0.70 0.70 0.70 213

1 0.76 0.76 0.76 267

accuracy 0.73 480

macro avg 0.73 0.73 0.73 480

weighted avg 0.73 0.73 0.73 480

7. Conclusions and Lessons learnt:

a. ElasticNet is better than Ridge and Lasso.

b. Accuracy of Logistic Regression and Decision Tree is almost similar for this dataset.