Assignment 4

In this assignment, you will work with a synthetic dataset to build, apply, and evaluate a simple linear regression model. The dataset consists of 15 observations with two variables: X (the predictor) and Y (the target). The goal is to understand how to build a linear regression model, make predictions, and evaluate the model's performance using various metrics.

(Round all your final answers to two decimal places)

Dataset (download as csv)

Index	X	Y
1	5	10.2
2	1	2.3
3	8	15.8
4	6	12.0
5	3	6.5
6	7	14.1
7	2	4.2
8	9	18.3
9	4	8.4
10	10	20.6
11	11	22.0
12	12	23.9
13	13	26.1
14	14	27.7
15	15	29.8

Training and Testing Sets

Training Set Label Indices: [9, 14, 8, 7, 15, 13, 6, 3, 10, 4]

Testing Set Label Indices: [5, 12, 1, 2, 11]

Note: Indices do not represent integer positions that begin at 0.

Task 1: (20 points)

1. Extract the training data using the given indices.
2. Build a simple linear regression model where **Y** represents the target variable and **X** denotes the predictor. Write your model in the answer section.

Answer: $Y=1.97 \cdot X+0.36$

Task 2: (40 points)

1. Apply the linear regression model to the testing data to predict Y values and calculate the error.
2. Present the results in the following table containing Index, X, Y, Predicted_Y, and Error.

Index	X	Y	Predicted_Y	Error
5	3	6.5	6.27	0.23
12	12	23.9	24.01	-0.11
1	5	10.2	10.21	-0.01
2	1	2.3	2.33	-0.03
11	11	22	22.03	-0.03

Task 3: (40 points)

1. Evaluate the model's performance on the test data using the following metrics and provide the evaluation score for each of them:
   - Mean Absolute Error (MAE) = 0.08
   - Mean Squared Error (MSE) = 0.01
   - Root Mean Squared Error (RMSE) = 0.12

Complete Tasks 1, 2, and 3, then upload a DOC or PDF file with the answers to Canvas, along with a Jupyter notebook containing the code you used to derive the answers.