Republic of the Philippines
Laguna State Polytechnic University
Province of Laguna

Laboratory Exercise No. 1			
Topic:	**Topic 2: Supervised Learning Fundamentals**	**Week No.**	3
Course Code:	**CSST102**	**Term:**	1st Semester
Course Title:	**Basic Machine Learning**	**Academic Year:**	2024-2025
Student Name		**Section**	
Due date		**Points**	

Laboratory Exercise 1: Linear Regression Implementation

Objective:
To apply the concepts of linear regression learned in lectures by implementing a simple linear regression model from scratch in Python.

Task Overview:
You are provided with a dataset containing information about house prices in a particular city. The dataset includes features such as the size of the house (in square feet), the number of bedrooms, and the age of the house. Your task is to predict the price of a house based on these features using a linear regression model.

Steps:

1. Data Preprocessing:
- Load the dataset into a Pandas DataFrame.
- Check for any missing values and handle them appropriately.
- Normalize the features to ensure they are on a similar scale.

2. Model Implementation:
- Implement the linear regression model using Python (do not use any pre-built libraries like Scikit-learn for this part).
- Derive the model parameters (slope and intercept) using the least squares method.
- Write a function that predicts the house price based on input features.
-

3. Model Training:
- Split the dataset into training and testing sets.
- Train your linear regression model on the training set.
- Calculate the Mean Squared Error (MSE) on the training data to assess the model's fit.

4. Model Evaluation:

- Test your model on the testing set and compute the MSE for the test data.
- Plot the regression line along with the test data points to visualize the model's performance.

5. Report:

- Write a short report (2-3 pages) summarizing your findings.
- Include sections on data preprocessing, model implementation, training, evaluation, and conclusions.
- Discuss any challenges you encountered and how you addressed them.

Submission Instruction:

- After completing the laboratory exercise in Google Colab, download your notebook as a .ipynb file. Upload the. ipynb file to the GitHub repository designated for our subject.
- Filename Format: **2A-BERNARDINO-EXER1**

Inability to follow this instruction will be deducted 5 points each for filename format and late submission per day. Also, cheating and plagiarism will be penalized.

Rubric for Laboratory Exercise 1: Linear Regression Implementation

Criteria	Excellent (90-100%)	Good (75-89%)	Satisfactory (60-74%)	Needs Improvement (0-59%)
Data Preprocessing	Complete and thorough data cleaning, normalization; no errors.	Good data cleaning; minor issues in handling or normalization.	Basic cleaning; some errors or inconsistencies in processing.	Little to no preprocessing; significant errors.
Model Implementation	Model implemented from scratch; code is clear, efficient, and accurate.	Model implemented with minor issues; code is mostly clear and functional.	Basic model implemented; some errors or inefficiencies in the code.	Major errors in implementation; unclear or non-functional code.
Model Training	Model is correctly trained; MSE is appropriately calculated and interpreted.	Model is trained with minor errors; MSE is calculated but with some misinterpretations.	Model is trained with some errors; MSE is calculated but not well interpreted.	Model training is incorrect; MSE calculation is incorrect or missing.
Model Evaluation	Model performance is thoroughly evaluated; plots are clear and informative.	Evaluation is mostly correct; plots are provided but may lack clarity.	Basic evaluation provided; plots are unclear or not fully accurate.	Little to no evaluation; no or very unclear plots.
Report Quality	Report is well-organized, clear, and complete; professional presentation.	Report is organized and mostly clear; minor issues in completeness or presentation.	Report is somewhat clear but lacks organization or completeness.	Report is unclear, disorganized, and lacks critical information.