Republic of the Philippines
Laguna State Polytechnic University
Province of Laguna

Machine Problem No. 2			
Topic:	**Evaluating Machine Learning Model Performance**	**Week No.**	
Course Code:	**CSST102**	**Term:**	1st Semester
Course Title:	**Basic Machine Learning**	**Academic Year:**	2025-2026
Student Name		**Section**	
Due date		**Points**	

Evaluating Machine Learning Model Performance Using Logistic Regression

Learning Objectives

After completing this activity, students will be able to:

1. Apply data preprocessing, train-test split, and model training techniques.
2. Implement logistic regression for classification tasks.
3. Evaluate model performance using a confusion matrix and learning curves.
4. Apply cross-validation (5-fold) to validate model reliability.
5. Interpret and communicate model results accurately.

Task Description

You are tasked to **develop, evaluate, and interpret a classification model** using **Logistic Regression**.

This activity simulates a real-world data science workflow — from model training to performance evaluation.

Instructions

1. Dataset Selection

Choose any publicly available dataset appropriate for classification (e.g., Iris, Titanic, Breast Cancer, or your chosen dataset).

2. Data Preparation

- Load and explore the dataset.
- Handle missing values (if any).
- Encode categorical variables if necessary.
- Normalize or standardize the data if needed.

3. Train-Test Split

- Split your dataset into **training (80%)** and **testing (20%)** using train_test_split() from scikit-learn.

4. Model Building – Logistic Regression

- Train a **Logistic Regression** model on your training set.
- Evaluate the model's accuracy on both training and testing sets.

5. Cross-Validation (5-Fold)

- Apply **5-Fold Cross Validation** using cross_val_score() from scikit-learn.
- Compute and display the mean and standard deviation of the cross-validation scores.

6. Model Evaluation – Confusion Matrix

- Predict on the test data.
- Generate and visualize a **Confusion Matrix**.
- Compute **Accuracy**, **Precision**, **Recall**, and **F1-score**.

7. Learning Curve Visualization

- Plot a **Learning Curve** using learning_curve() from scikit-learn.
- Analyze whether the model is overfitting, underfitting, or well-fitted.

8. Interpretation & Discussion

Write a short report addressing the following:

- What do the results of the confusion matrix indicate?
- How consistent is the model's performance based on 5-Fold Cross Validation?
- What insights can be derived from the learning curve?
- How can the model be improved?

Rubrics (Total: 100 pts)

Criteria	Description	Points
Data Preparation & Splitting	Correctly applies preprocessing and train-test split	15 pts
Model Implementation	Properly implements logistic regression and training	20 pts
Cross Validation & Learning Curve	Correct computation and visualization	20 pts
Confusion Matrix Analysis	Correctly interprets results and metrics	20 pts
Report Clarity & Interpretation	Clear, logical, and well-written explanations	25 pts

Challenge (Optional +10 pts)
Compare Logistic Regression performance with another classifier (e.g., Decision Tree, KNN, or SVM) using the same dataset. Discuss which performs better and why.

Expected Outputs and Submission Instructions

1. **Create a folder named MP2** inside your **GitHub repository (CSST102)**.
   This folder will contain all your files related to **Machine Problem 2 (Logistic Regression and Model Evaluation)**.

2. Inside the MP2 folder, include the following required files:

Filename	Description
logistic_regression.ipynb	Jupyter Notebook containing your complete code, results, and explanatory comments
learning_curve.png	Visualization of the generated learning curve
confusion_matrix.png	Visualization or screenshot of the confusion matrix
cross_validation.txt	Summary of 5-Fold Cross Validation results (mean and standard deviation)
report.pdf / report.docx	Short reflection report (1–2 pages) summarizing your findings, interpretations, and model performance
README.md	Brief description of your experiment, dataset, and how to run your code

3. **Commit and push** all files to your GitHub repository.
   Ensure the folder structure is correct:

```
CSST102/
└── MP2/
  ├── logistic_regression.ipynb
  ── learning_curve.png
  ├── confusion_matrix.png
  ├── cross_validation.txt
  ├── report.pdf (or report.docx)
  └── README.md
```