Republic of the Philippines
Laguna State Polytechnic University
Province of Laguna

Activity 5			
Topic:	**Topic 5: Logistic Regression**	**Week No.**	6
Course Code:	**CSEL302**	**Term:**	2nd Semester
Course Title:	**Introduction to Intelligent Systems**	**Academic Year:**	2023-2024
Student Name		**Section**	
Due date		**Points**	

Assessment Task 5: Bank Customer

Objective:
Predict whether a customer will subscribe to a term deposit based on their demographic and account information using logistic regression.

Dataset Overview
Imagine we have a dataset named bank_customers.csv with the following columns:
Dataset: https://www.kaggle.com/datasets/kidoen/bank-customers-data

- **age:** The age of the customer.
- **job:** Type of job (e.g., admin, technician, services, management).
- **marital:** Marital status (e.g., married, single, divorced).
- **education:** Level of education (e.g., secondary, tertiary, primary, unknown).
- **default:** Has credit in default? (yes, no).
- **balance:** Average yearly balance, in euros.
- **housing:** Has housing loan? (yes, no).
- **loan:** Has personal loan? (yes, no).
- **contact:** Communication type (e.g., unknown, cellular, telephone).
- **day:** Last contact day of the month.
- **month:** Last contact month of the year (e.g., jan, feb, mar, ...).
- **duration:** Last contact duration, in seconds.
- **campaign:** Number of contacts performed during this campaign for this client.
- **pdays:** Number of days that passed by after the client was last contacted from a previous campaign (999 means client was not previously contacted).
- **previous:** Number of contacts performed before this campaign and for this client.
- **poutcome:** Outcome of the previous marketing campaign (e.g., unknown, other, failure, success).
- **subscribed:** Has the client subscribed a term deposit? (yes, no) - Target Variable.

Steps:

1. Data Preprocessing:

- Load the dataset into a Pandas DataFrame.
- Convert categorical variables into dummy variables.
- Handle missing values if any.
- Convert the target variable subscribed into a binary format (1 for yes, 0 for no).

2. Feature Selection:

- Decide which features to include in the model. You might exclude highly correlated features to avoid multicollinearity.

3. Data Splitting:

- Split the dataset into training and testing sets (typically a 70-30 or 80-20 split).

4. Model Training:

- Train a logistic regression model on the training set.

5. Model Evaluation:

- Evaluate the model's performance on the testing set using metrics such as accuracy, precision, recall, F1-score, and the confusion matrix.

6. Conclusion:

- Summarize the model's performance and discuss any insights or implications for the bank's marketing strategies.

This activity not only provides hands-on experience with linear regression but also allows students to explore the factors influencing student performance, making it more engaging and relevant to their academic context.

Submission Instruction:

- Share the Google Collab Activity to markbernardino@lspu.edu.ph
- Filename Format: **2A-BERNARDINO-EXER5**

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.