Republic of the Philippines
Laguna State Polytechnic University
Main Campus, Sta. Cruz, Laguna

BASIC MACHINE LEARNING
Machine Problem 1

Name: Ramos, Jezreel R
Year/Section: BSCS3A **Instructor: Mark Bernardino**

FUNDAMENTALS OF MACHINE LEARNING

Topic: What is ML? Types of ML and Core Challenges

Lab Outline (3 hours)

Hour 1 – Setup & Dataset Exploration

- Install/verify Python, Jupyter/Colab, and Scikit-Learn.
- Load the Iris dataset (classification) or California Housing dataset (regression).
- Explore dataset (features, targets, summary statistics).

```
#HOUR 1: SETUP & DATASET EXPLORATION
from sklearn.datasets import load_iris
import pandas as pd
#load dataset
iris = load_iris(as_frame=True)
df = iris.frame
print(df.head())
#Explore
print(df.describe())
print("Target classes:",iris.target_names)
```

```
from google.colab import files
import pandas as pd

# Upload the CSV file
# uploaded = files.upload()

# Load the dataset from the uploaded CSV file
df = pd.read_csv('housing.csv')  # Adjust the filename if needed

# Show the first few rows
print(df.head())

# Describe the dataset
print(df.describe())
```

```python
# Display unique target classes (assuming 'species' column is present)
print("Target classes:", df['households'].unique())
```

```
   longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
0    -122.23     37.88                41.0        880.0           129.0
1    -122.22     37.86                21.0       7099.0          1106.0
2    -122.24     37.85                52.0       1467.0           190.0
3    -122.25     37.85                52.0       1274.0           235.0
4    -122.25     37.85                52.0       1627.0           280.0

   population  households  median_income  median_house_value ocean_proximity
0       322.0       126.0         8.3252            452600.0        NEAR BAY
1      2401.0      1138.0         8.3014            358500.0        NEAR BAY
2       496.0       177.0         7.2574            352100.0        NEAR BAY
3       558.0       219.0         5.6431            341300.0        NEAR BAY
4       565.0       259.0         3.8462            342200.0        NEAR BAY
          longitude      latitude  housing_median_age   total_rooms  \
count  20640.000000  20640.000000        20640.000000  20640.000000
mean    -119.569704     35.631861           28.639486   2635.763081
std        2.003532      2.135952           12.585558   2181.615252
min     -124.350000     32.540000            1.000000      2.000000
25%     -121.800000     33.930000           18.000000   1447.750000
50%     -118.490000     34.260000           29.000000   2127.000000
75%     -118.010000     37.710000           37.000000   3148.000000
max     -114.310000     41.950000           52.000000  39320.000000

       total_bedrooms    population    households  median_income  \
count    20433.000000  20640.000000  20640.000000   20640.000000
mean       537.870553   1425.476744    499.539680       3.870671
std        421.385070   1132.462122    382.329753       1.899822
min          1.000000      3.000000      1.000000       0.499900
25%        296.000000    787.000000    280.000000       2.563400
50%        435.000000   1166.000000    409.000000       3.534800
75%        647.000000   1725.000000    605.000000       4.743250
max       6445.000000  35682.000000   6082.000000      15.000100

       median_house_value
count        20640.000000
mean        206855.816909
std         115395.615874
min          14999.000000
25%         119600.000000
50%         179700.000000
75%         264725.000000
max         500001.000000
Target classes: [ 126. 1138.  177. ... 1767. 1832. 1818.]
```

```
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  \
0                5.1               3.5                1.4               0.2
1                4.9               3.0                1.4               0.2
2                4.7               3.2                1.3               0.2
3                4.6               3.1                1.5               0.2
4                5.0               3.6                1.4               0.2

  species
0  setosa
1  setosa
2  setosa
3  setosa
4  setosa
       sepal length (cm)  sepal width (cm)  petal length (cm)  \
count         150.000000        150.000000         150.000000
mean            5.843333          3.057333           3.758000
std             0.828066          0.435866           1.765298
min             4.300000          2.000000           1.000000
25%             5.100000          2.800000           1.600000
50%             5.800000          3.000000           4.350000
75%             6.400000          3.300000           5.100000
max             7.900000          4.400000           6.900000

       petal width (cm)
count        150.000000
mean           1.199333
std            0.762238
min            0.100000
25%            0.300000
50%            1.300000
```

Mini-task: Students answer:

1. What is the input (features)?
   - The flower measurements: sepal length, sepal width, petal length, petal width.
2. What is the output (label)?
   - The species of iris (Setosa, Versicolor, Virginica).
3. Is this supervised or unsupervised learning?
   - Supervised learning because we have labeled data: features, known target.

Hour 2 – Train-Test Split & Baseline Model

- Perform train-test split (80% train, 20% test).
- Train a simple baseline model: o Logistic Regression (for Iris) o Linear Regression (for Housing)
- Make predictions.

```
#Hour 2 - Train-Test Split & Baseline Model
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

X = df[iris.feature_names]
y = iris.target

X_train, X_test, y_train, y_test = train_test_split(X,y, test_size=0.2,
random_state=42)
model = LogisticRegression(max_iter=200)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
```

Mini-task: Students compute model accuracy.

Hour 3 – Evaluation & Reflection

- Evaluate model with different metrics: o Classification: Confusion matrix, precision, recall. o Regression: RMSE (Root Mean Squared Error).
- Discuss ML challenges: overfitting, underfitting, and bad data.
- Students reflect:

- "What would happen if the dataset had missing or wrong values?"
  - The model might fail to train properly or give poor predictions, since ML models rely on clean and consistent data.
- "How does this relate to real-world ML applications?"
  - In real-world problems (e.g., medical diagnosis, spam detection), data can often be messy. Handling missing values, noise, and biases is critical to building reliable ML systems.

```
#Hour 3 – Evaluation & Reflection

from sklearn.metrics import confusion_matrix, classification_report
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

```
Accuracy: 1.0
```

```
#Hour 3 – Evaluation & Reflection

from sklearn.metrics import confusion_matrix, classification_report
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

```
[[10  0  0]
 [ 0  9  0]
 [ 0  0 11]]
              precision    recall  f1-score   support

      setosa       1.00      1.00      1.00        10
  versicolor       1.00      1.00      1.00         9
   virginica       1.00      1.00      1.00        11

    accuracy                           1.00        30
   macro avg       1.00      1.00      1.00        30
weighted avg       1.00      1.00      1.00        30
```

Short Reflection (3–5 sentences)

In this lab, we used supervised learning, specifically a classification model (Logistic Regression) on the Iris dataset. A possible challenge that might affect the model is overfitting if the model memorizes training data instead of generalizing. Another issue is bad or missing data, which can reduce accuracy and reliability. This exercise connects to real-world ML because most datasets need cleaning and preprocessing before building trustworthy models.