Date: 07/03/2025

Submission Date: 14/03/2025

Assignment 2:

Course Code: 202046702

Course Title: Artificial Intelligence and Machine Learning

Predict Housing Prices with Machine Learning

Dataset:

Use the **"California Housing Prices" dataset**, available on Kaggle **Dataset link:** California Housing Data. This dataset contains information about housing features and prices, making it a good regression problem.

Assignment Steps:

Part 1: Data Loading and Exploration

1. Load the California Housing Prices dataset from Kaggle.
2. Convert the dataset into a Pandas DataFrame for easier manipulation.
3. Display the first five rows and get basic statistics like mean, median, and standard deviation for each feature.
4. Explore the relationships between features using:
    - Histograms for individual features.
    - A heatmap to show correlations between features.

Part 2: Preprocessing

1. Check for missing values and handle them (e.g., fill, drop, or impute).
2. Normalize/standardize numerical features using techniques like Min-Max scaling or Standard Scaling.
3. Split the dataset into training (80%) and testing (20%) sets.

Part 3: Model Training

1. Train a Linear Regression model on the dataset.
2. Use the training data to fit the model.

Part 4: Model Evaluation

1. Predict the target variable (housing prices) for the test data.
2. Evaluate the model using metrics like:
    - Mean Absolute Error (MAE)
    - Mean Squared Error (MSE)
    - R-squared Score
3. Plot the predicted vs. actual housing prices using a scatter plot.

Part 5: Experimentation

1. Train at least one additional regression model (e.g., Decision Tree Regressor or Random Forest Regressor).
2. Compare the performance of both models using the evaluation metrics.

Submission Requirements:

- A Jupyter Notebook (.ipynb file) with:
    - Clear and organized code, including comments.
    - Outputs for each step, including visualizations.
    - A short conclusion summarizing the findings and comparing the models.