**National University**

**of Computer & Emerging Sciences-Islamabad Chiniot-Faisalabad Campus**

**Assignment Submission Guidelines:**

1. Submit your assignment in **soft form (Code + Report)** within the due date and time. Soft form does not mean submitting photos of the hardcopy. Late submissions will result in a deduction of marks.
2. The **report** must include a discussion, comments, and a conclusion about your solution. Submitting without a report will result in a loss of full marks.
3. Name the zip or other folder/file that you submit using the following format:

**AML\_A1\_RollNo\_FirstName**.

1. Ensure that you solve each task of the assignment on your own.
2. There is no restriction on the programming language used for the tasks.
3. Any question with Error is not acceptable and marked as zero.

**Question #01: (K-Mean And DBSCAN Clustering)**

1.**Load and Preprocess the Dataset**

* Write a Python code to load the dataset and handle missing values appropriately.
* Normalize the numerical features to ensure proper clustering.

2. **Determine the Optimal Number of Clusters (K)**

* Implement the Elbow Method to find the optimal value of K.
* Use the Silhouette Score to validate the number of clusters.

3.**Apply K-Means Clustering**

* Perform K-Means clustering on the dataset after preprocessing.
* Visualize the clustering results using a scatter plot (for relevant features).

4.**Cluster Interpretation**

* Assign cluster labels to the dataset and analyze the characteristics of each cluster.
* Find the average values of each feature for different clusters and interpret the results.

5.**Handling Outliers**

* Detect and remove outliers from the dataset before clustering.
* Compare the clustering results before and after outlier removal.

6.**Comparison with Other Clustering Algorithms**

* Implement DBSCAN Clustering and compare the results with K-Means.
* Discuss when K-Means performs better and when DBSCAN clustering methods might be preferable.

**Question #02: (Linear Regression)**

**1. Data Preprocessing and Exploration**

* Load the dataset and handle missing values appropriately.
* Perform exploratory data analysis (EDA) by visualizing feature distributions and relationships with the target variable (**House\_Price**).
* Identify and handle outliers in the dataset.

**2. Feature Engineering and Selection**

* Normalize or scale numerical features appropriately for regression.
* Check for feature correlation and identify which features contribute the most to **House\_Price** prediction.
* Create polynomial features to capture non-linearity in the dataset.

**3. Train a Linear Regression Model**

* Split the dataset into training and testing sets (e.g., 80-20 split).
* Train a **Linear Regression** model to predict **House\_Price**.
* Evaluate the model using **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, and **R² score**.

**4. Implement Linear Regression using Gradient Descent**

* Implement Linear Regression from scratch using **Gradient Descent** instead of using LinearRegression from sklearn.
* Compare the performance of the Gradient Descent implementation with Scikit-learn's Linear Regression model.

**5. Predict House Prices for New Data**

* Create a function that takes new house data as input and predicts the house price using the trained model.
* Test your model on unseen data points and analyze its predictions