Digital Empowerment Network

Artificial Intelligence

Week 02

□ **Clustering and Dimensionality Reduction in Machine Learning**

Mentor: Hussain Shoaib

Task Title:

Clustering and Dimensionality Reduction in Machine Learning

Objective:

To explore and implement unsupervised learning techniques including clustering and dimensionality reduction, using a real-world dataset.

Step 1: Dataset Selection

Choose any of the following datasets :

- Mall Customer Segmentation Dataset
- Wholesale Customer Data
- Fashion MNIST (for dimensionality reduction)
- Any dataset suitable for clustering tasks

Step 2: Task Instructions

- **A. Data Preparation & Preprocessing:**
- Load and explore the dataset

- Handle missing or duplicate values

- Encode categorical features (if any)

- Scale or normalize the data

- **B. Dimensionality Reduction (PCA or t-SNE)**:
- Apply PCA or t-SNE to reduce dimensionality

- Visualize the reduced dataset using scatter plots

- Interpret variance explained by components (in PCA)

- **C. Clustering Implementation:**
- Apply at least two clustering algorithms (e.g., K-Means, DBSCAN, Hierarchical Clustering)

- Use Elbow Method or Silhouette Score to find optimal cluster count

- Visualize clusters with reduced features (2D plot)

- **D. Evaluation:**
- Analyze the quality of clusters (Silhouette Score, Davies-Bouldin Index)

- Interpret and explain the clusters formed

- Compare clustering results between different algorithms

Deliverables:

• Well-documented Jupyter Notebook or Python script
• A short report explaining:
 - Preprocessing steps
 - Dimensionality reduction insight
 - Clustering approach and evaluation
• (Optional) A short video explaining your clustering project

Deadline:

Submit your task within 7 days from assignment.
Last Date: 31-July-2025