Homework – Week 5: Clustering Techniques

Course: Introduction to Machine Learning

Objective

Use unsupervised learning techniques to analyze a dataset by applying clustering algorithms. You will explore patterns in the data without using any labels.

Instructions

1. Visit **https://www.kaggle.com/datasets** and select a dataset that interests you.
   - It must contain at least two numerical variables that can be used for clustering.
   - You may use public datasets such as wine quality, mall customers, social network ads, etc.

2. Load the dataset into a Jupyter notebook and apply basic preprocessing:
   - Drop missing or irrelevant columns
   - Normalize or scale features as needed

3. Apply at least **two clustering algorithms** from the following list:
   - K-Means
   - Hierarchical Clustering (Agglomerative)
   - DBSCAN

4. For each algorithm, include:
   - A scatter plot showing the clusters
   - A short explanation of your results (e.g., number of clusters, shape, noise)
   - Optional: Use silhouette score or another metric to evaluate the quality of the clustering

5. Submit your notebook or export it as a PDF. Your submission should include:
   - The name and link of the dataset

- Your code and visualizations

- Your observations and interpretation

Tips

- Use 'StandardScaler' from 'sklearn.preprocessing' to standardize your features.

- For hierarchical clustering, try using a dendrogram to choose the number of clusters.

- For DBSCAN, experiment with different 'eps' and 'min samples' values.