Task 19

Machine Learning

Comparative Analysis of Clustering Techniques Across Datasets

Task Description:

1. Dataset Selection and Initial Analysis:
Step 1: Choose two datasets from the list provided (Iris, Mall Customers, Wholesale Customers).
Step 2: Conduct an initial exploratory data analysis (EDA) for each dataset to understand its characteristics, including data distribution, feature correlations, and potential outliers.

2. Implementing Clustering Algorithms:
Step 3: Apply K-Means clustering to both datasets. Determine the optimal number of clusters using methods such as the Elbow Method and Silhouette Score.
Step 4: Apply Hierarchical Clustering to both datasets, choosing an appropriate linkage criterion (e.g., single, complete, average) and visualizing the dendrogram to determine the number of clusters.

3. Cluster Visualization and Interpretation:
Step 5: Visualize the clusters obtained from both K-Means and Hierarchical Clustering. Use dimensionality reduction techniques like PCA or t-SNE to help in visualizing the clusters, if necessary.
Step 6: Compare the clustering results qualitatively (e.g., cluster compactness, separation) and quantitatively (e.g., Silhouette Score, Davies-Bouldin Index).

4. Exploratory Analysis and Insights:
Step 7: Analyze the clusters in the context of the original features. For each dataset, interpret the clusters to identify any patterns or insights (e.g., customer segments, species differentiation).

Step 8: Explore the impact of different clustering parameters (e.g., number of clusters in K-Means, linkage criteria in Hierarchical Clustering) on the results.

5. Comparison and Reporting:
Step 9: Compare the effectiveness of K-Means and Hierarchical Clustering across the two datasets. Discuss which algorithm performed better for each dataset and why, considering factors such as data distribution and feature space.

Step 10: Prepare a comprehensive Article summarizing the findings, including visualizations, cluster interpretations, and a comparative analysis of the clustering techniques used.