Lab-Assessment-3(16-4-2024)

Compare the performance of various clustering algorithms using a publicly available dataset. Your objective is to implement and evaluate the following clustering algorithms: K-Means, Hierarchical Clustering, and DBSCAN.

Tasks:

1. Load the dataset and perform any necessary preprocessing steps.

2. Implement K-Means, Hierarchical Clustering, and DBSCAN algorithms.

3. Evaluate each algorithm's performance using appropriate metrics such as silhouette score or Davies–Bouldin index.

4. Visualize the clustering results for each algorithm.

5. Compare the performance of the algorithms and discuss their strengths and weaknesses.

Instructions:

1. Load the dataset "customer_purchases.csv".

2. Preprocess the data by handling any missing values and scaling if necessary.

3. Implement the K-Means, Hierarchical Clustering, and DBSCAN algorithms using appropriate libraries (e.g., scikit-learn in Python).

4. Evaluate the performance of each algorithm using the silhouette score or Davies–Bouldin index.

5. Visualize the clustering results using scatter plots or any other suitable visualization techniques.

6. Provide a detailed comparison of the algorithms based on their performance metrics and visualization results.

7. Write a brief conclusion summarizing your findings.

Marking Scheme:

- Dataset loading and preprocessing: 15 marks

- Implementation of clustering algorithms: 30 marks (10 marks each)

- Evaluation of algorithm performance: 25 marks (10 marks for evaluation metric calculation, 15 marks for analysis)

- Visualization of clustering results: 20 marks (10 marks for clarity, 10 marks for insightfulness)

- Comparison and discussion: 10 marks

- Conclusion: 10 marks

Total: 100 marks