

ASSIGNMENT #4

1. Apply the below clustering algorithms using Python:

- a. **Partition based:** K-means, K-medoids/PAM
- b. **Hierarchical:** Dendrogram, AGNES, BIRCH
- c. **Density based:** DBSCAN, OPTICS

on the following UCI datasets (can be loaded from the package itself):

- a. **Iris plants dataset:** <https://archive.ics.uci.edu/ml/datasets/Iris/>
- b. **Wine Dataset:** <https://archive.ics.uci.edu/ml/datasets/wine>

Additionally, implement **K-means++** and **Bisecting K-means**.

2. Evaluate and compare the performances of the algorithms for each type of clustering, based on the following metrics:

- a. **Rand index:** rand score, adjusted rand score
- b. **Mutual Information based scores:** mutual info, adjusted mutual info, normalized mutual info
- c. **Silhouette Coefficient, Calinski-Harabasz Index and Davies-Bouldin Index**

During evaluation, replace the class/category names with numerical values starting from 0, 1, 2,...

3. Also determine the **Cohesion** and **Separation** performance scores using **Sum of Squared Error (SSE)** and **Sum of Squares Between groups (SSB)**.

Try to achieve accuracy $\geq 80\%$.

Show the performance comparison for each category of clustering algorithms in a tabular form.

Save the assignment in a single pdf file with the naming convention “Full Class Roll No_Full Name.pdf” and upload the report by using the Google form link:

<https://forms.gle/rJ59q1T8kg9tJmpA9>

Submission Deadline: 27th September 2022 Tuesday (11:59 pm) EOD