

## **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. **Homogeneity, Completeness and V-measure**

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 "Class Roll No\_Full Name.pdf" and mail the report to us through email by 8<sup>th</sup> November, 2021 EOD to:

- [pksingh.it@jadavpuruniversity.in](mailto:pksingh.it@jadavpuruniversity.in)
- [neelotpalcse.rs@jadavpuruniversity.in](mailto:neelotpalcse.rs@jadavpuruniversity.in)