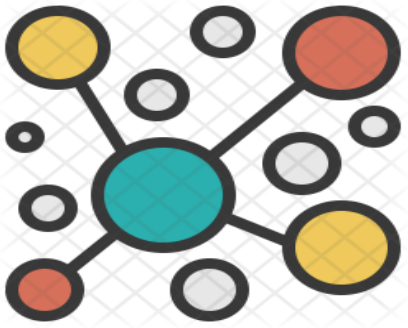
Data Mining

Lab sheet N°7: Clustering

ENSIA 2023-2024



Resources

Clustering

- **Clustering algorithms:** API Reference — scikit-learn 1.3.2 documentation
- **Distance metrics:** scipy.spatial.distance.cdist

Hierarchical clustering

- **Scipy:** Hierarchical clustering (scipy.cluster.hierarchy) — SciPy v1.11.4 Manual
- **Dendogram:** scipy.cluster.hierarchy.dendrogram
- **Agglomerative clustering:** sklearn.cluster.AgglomerativeClustering — scikit-learn 1.3.2 documentation

DBSCAN clustering

- **ScikitLearn:** sklearn.cluster.DBSCAN — scikit-learn 1.3.2 documentation

Clustering - Evaluation metrics

- **Clustering metrics:** 2.3. Clustering — scikit-learn 1.3.2 documentation
- **Silhouette Score:**sklearn.metrics.silhouette_score — scikit-learn 1.3.2 documentation
- **Homogeneity score:** sklearn.metrics.homogeneity_score