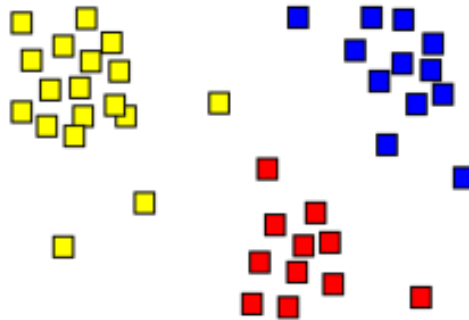


Objective : l'objectif principal de cet atelier est de pratiquer les concepts du clustering, en traitant les données d'une Data Sets.



Outils : Python , Pandas, Sklearn, matplotlib.

Data Sets :

Credit Card Data Set : <https://www.kaggle.com/arjunbhasin2013/ccdata>

Partie 1 (Data Visualisation):

1. En utilisant pandas essayer d'explorer les données du Data set.
2. Afficher le résumer statistique du Data Sets avec une interprétation des résultats obtenues.
3. Afficher les nuages des points du data set selon les propriétés « Features » en utilisant matplotlib et pandas « scatter_matrix ».
4. Appliquer les deux techniques PCA et Tsne sur les features du Dataset, que ce que vous constatez.

Partie 2 (Clustering):

1. Essayer de construire les modèles de clustering en utilisant Kmeans (avec les nouvelles features (Un modèle basé sur PCA et l'autre sur Tsne) Question 4 de la partie 1.



Université Abdelmalek Essaadi
Faculté des Sciences et techniques de Tanger
Département Génie Informatique
Cycle Ingénieur: LSI s4
Machine Learning
Pr . EL AACHAK LOTFI
2023/2024



2. Définir le K nécessaire pour les deux modèles en utilisant la méthode d'Elbow.
3. Présenter les clusters obtenues dans un graphe en utilisant matplotlib.
4. Interpréter les résultats obtenus des deux modèles .
5. Refaire la même chose en utilisant l'algorithme fuzzy cmeans « il faut utiliser la bibliothèque skfuzzy » DBSCAN , EM , SOM et Hierarchical clustering .
6. Comparer les cinq algorithmes et faire une conclusion globale sur ces méthodes de clustering.

Note : un rapport bien détailler doit être rédigé et imprimé.

Références :

<https://towardsdatascience.com/k-means-clustering-with-scikit-learn-6b47a369a83c> http://eneskemalergin.github.io/blog//blog/Fuzzy_Clustering/
<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>
<https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>