

Aicha Ndiaye
DIC2 IABD

Le code:

```
# Importation des bibliothèques nécessaires

import pandas as pd
import numpy as np

from sklearn.cluster import KMeans, DBSCAN
from sklearn.mixture import GaussianMixture
from sklearn.metrics import silhouette_score
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA


# Charger les données depuis un fichier CSV
# Remplacez 'dataset.csv' par le nom de votre fichier
data_path = "uci_malware_detection nonSup.csv" # Indiquez le chemin vers votre
fichier
data = pd.read_csv(data_path, header=None)


# Définition des modèles
kmeans = KMeans(n_clusters=2, random_state=4) # 2 clusters
dbscan = DBSCAN(eps=0.5, min_samples=2)
gmm = GaussianMixture(n_components=7, random_state=4) # Exemple avec 2
clusters


# Application des modèles et calcul des Silhouette Scores
# KMeans
```

Aicha Ndiaye
DIC2 IABD

```
kmeans_labels = kmeans.fit_predict(data)
kmeans_silhouette = silhouette_score(data, kmeans_labels)

# DBSCAN
dbscan_labels = dbscan.fit_predict(data)
# Vérifier que tous les points ne sont pas bruités (éviter silhouette_score si un seul
cluster)
if len(set(dbscan_labels)) > 1:
    dbscan_silhouette = silhouette_score(data, dbscan_labels)
else:
    dbscan_silhouette = -1 # Cas où DBSCAN ne forme pas de clusters valides

# GMM
gmm_labels = gmm.fit_predict(data)
gmm_silhouette = silhouette_score(data, gmm_labels)

# Affichage des résultats
print("Silhouette Scores:")
print(f"KMeans: {kmeans_silhouette}")
print(f"DBSCAN: {dbscan_silhouette} (peut être -1 si pas de clusters valides)")
print(f"Gaussian Mixture Model: {gmm_silhouette}")

# Comparaison des modèles
best_model = max([(kmeans_silhouette, 'KMeans'),
                  (dbscan_silhouette, 'DBSCAN'),
                  (gmm_silhouette, 'GMM')], key=lambda x: x[0])
```

Aicha Ndiaye
DIC2 IABD

```
print(f"Le meilleur modèle basé sur le Silhouette Score est : {best_model[1]} avec un  
score de {best_model[0]}")
```

Le résultat:

```
Silhouette Scores:  
KMeans: 0.5459617768735687  
DBSCAN: 0.26068637130449684 (peut être -1 si pas de clusters valides)  
Gaussian Mixture Model: 0.07849190832268581  
Le meilleur modèle basé sur le Silhouette Score est : KMeans avec un score de 0.545  
9617768735687
```

ps: Nous voulions afficher les cluster formé mais les méthode utilisé étaient peu efficace et le téléchargement de librairies nécessaire nous à pris beaucoup de temps et de connection.