

Optimisation de la stabilité des résultats de K-means en utilisant une approche de clustering hybride avec HAC

NGAM Amadou

SARR Serigne Abdou Lat



Plan

Introduction

Problématiques et Objectifs

Propositions

Expériences et résultats

Conclusion et perspectives

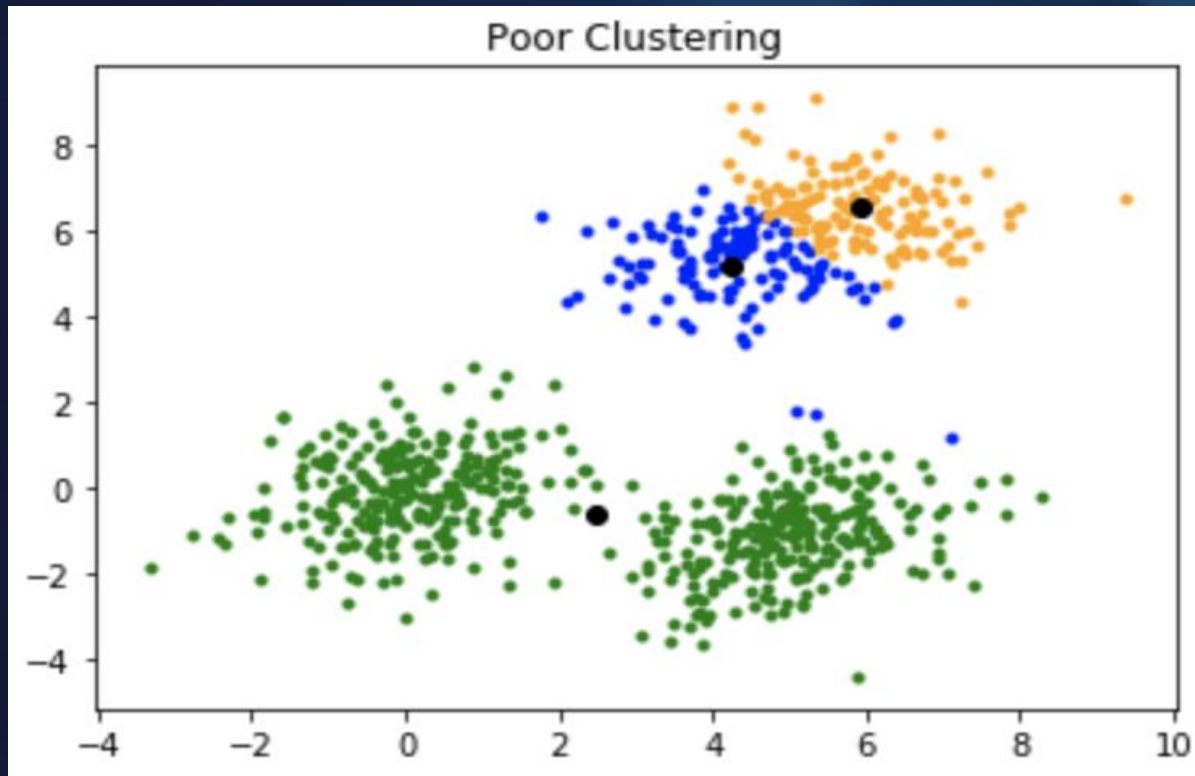




Introduction

- Définition de la notion de clustering
- Contexte
- Motivations

Problématique et Objectifs



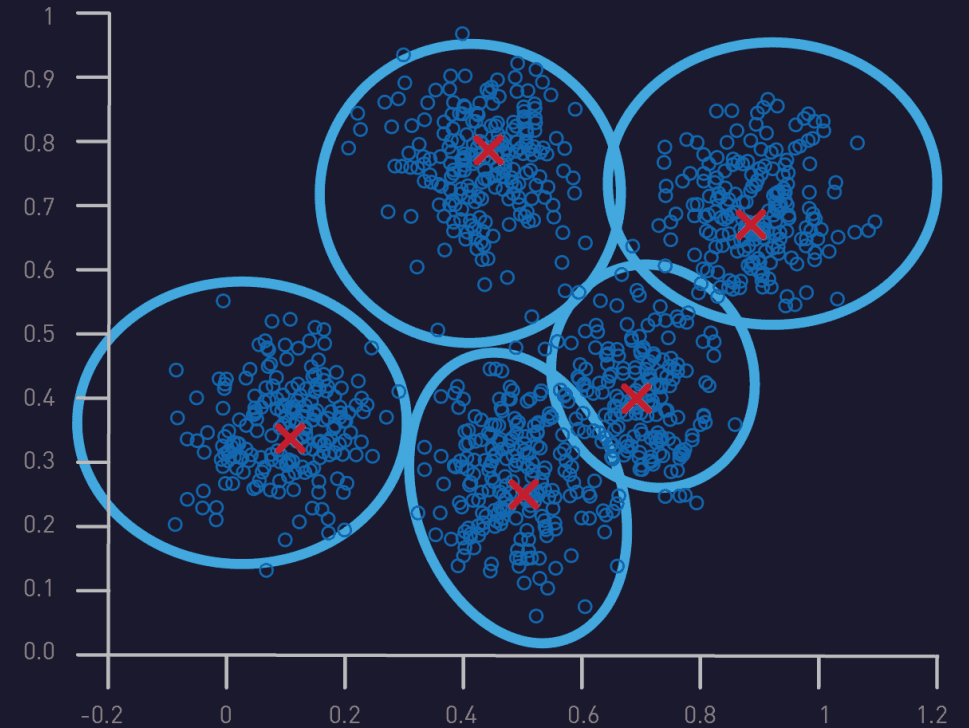
Proposition

- Fonctionnement de HCA et de K-means
- Combinaison des deux algorithmes



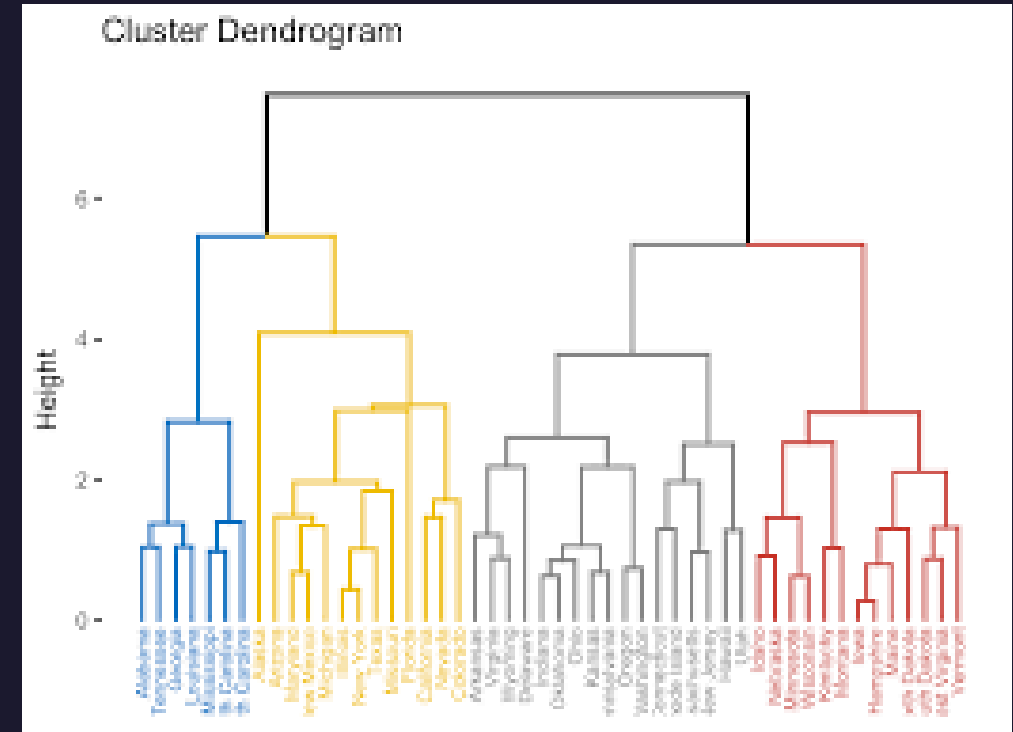
Proposition : Fonctionnement de K-means

- Initialisation des centroïdes
- Affectation des données aux centroïdes les plus proche
- Mise à jour des centroïdes



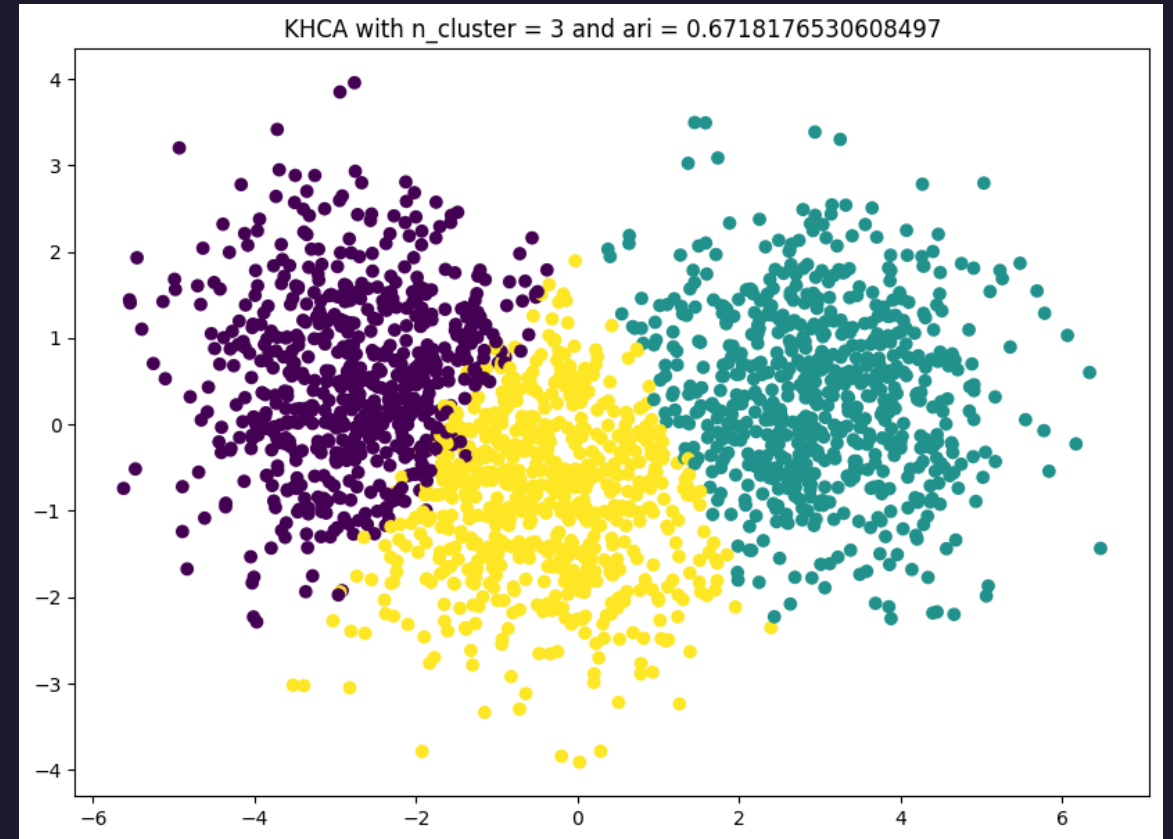
Proposition : Fonctionnement de HCA

- Calcul de distance entre chaque paire d'observation
- Création de dendrogramme
- Sélection du nombre de clusters



Proposition : Combinaison des deux algorithmes

- Récupération des centroïdes du HCA
- Initialisation du k-means avec ces centroïdes



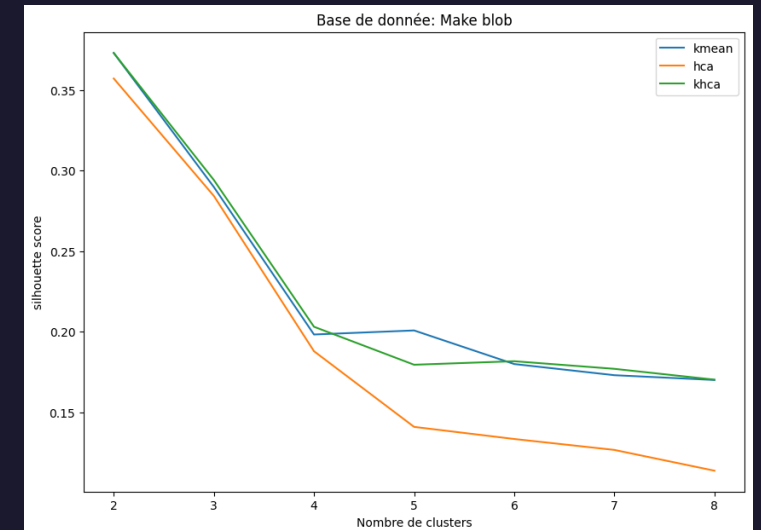
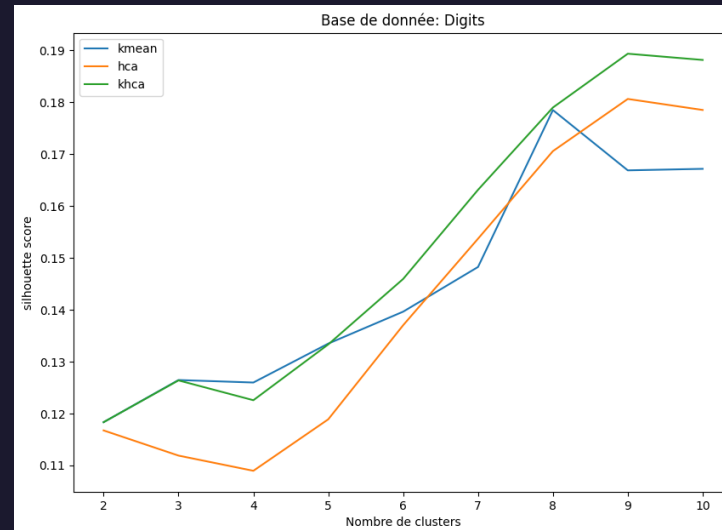
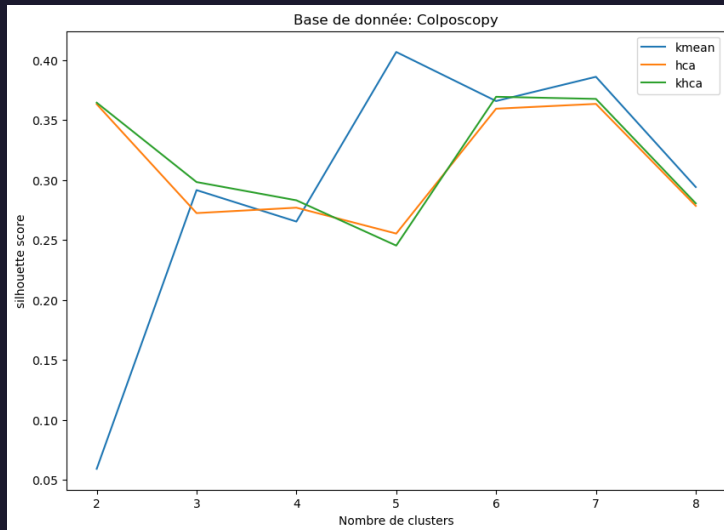
Expériences

- Tests sur quelques bases de données
- Calcul des scores pour chaque algorithme
 - ❑ Silhouette score
 - ❑ Adjusted Rand Index (Ari)



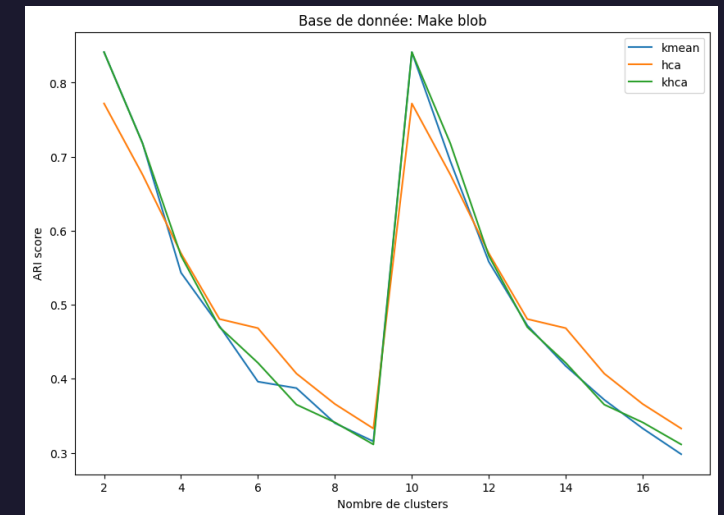
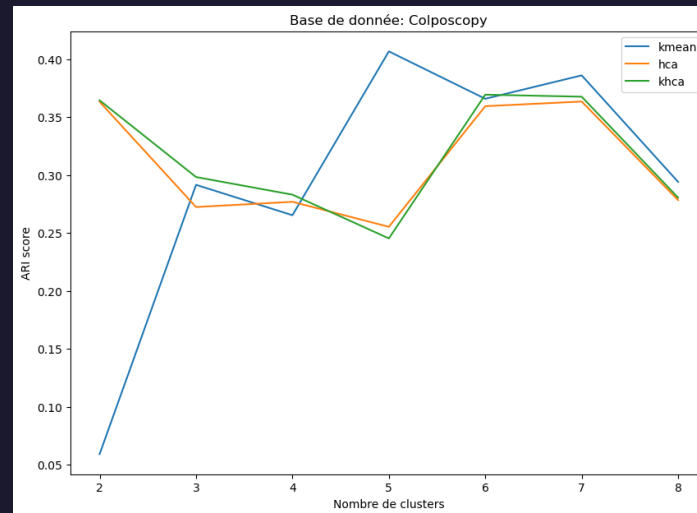
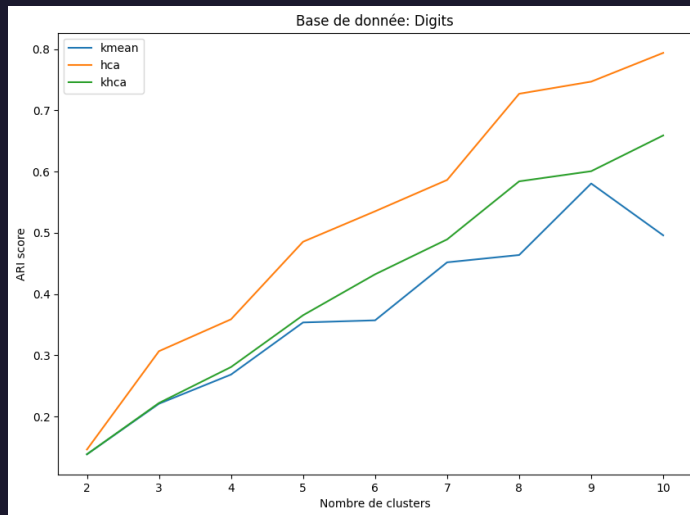
Résultats

Silhouette score



Résultats

Adjusted Rand Index (Ari)



Résultats

Base de donnée /Score ARI	Kmeans	HCA	KHCA
Digits	0,39	0,52	0,42
Make blob	0,501	0,509	0,504
Colposcopy	0,439	0,309	0,315



Résultats

Base de donnée /Score silhouette	Kmeans	HCA	KHCA
Digits	0,136	0,141	0,151
Make blob	0,226	0,191	0,225
Colposcopy	0,149	0,161	0,173





Conclusion et Perspectives