

**Nom :Souissi**

**Prénom : Salma**

**Adresse email : souissi.salma122000@gmail.com**

## **Rapport sur la Classification des Téléphones**

### **Introduction**

Dans ce projet, l'objectif était de classifier des téléphones en fonction de plusieurs caractéristiques (RAM, batterie, caméra, etc.). Deux algorithmes de classification non supervisés, **KMeans** et **GMM** (Gaussian Mixture Model), ont été utilisés pour regrouper les téléphones en différentes classes. La visualisation des résultats a été réalisée à l'aide de techniques de réduction dimensionnelle telles que **PCA** et **t-SNE**, et l'efficacité des clusters a été évaluée par l'**indice de silhouette**.

### **Pourquoi algorithmes non supervisés ?**

Je n'ai pas de labels pour les données et je souhaite regrouper les téléphones en catégories basées uniquement sur les caractéristiques fournies (RAM, batterie, taille de l'écran, etc.)

### **À propos de l'ensemble de données :**

#### **Contexte :**

Bob a lancé sa propre entreprise de téléphonie mobile. Il veut livrer un combat acharné aux grandes entreprises comme Apple, Samsung, etc.

Il ne sait pas estimer le prix des mobiles créés par son entreprise. Dans ce marché concurrentiel de la téléphonie mobile, vous ne pouvez pas simplement présumer des choses. Pour résoudre ce problème, il collecte des données sur les ventes de téléphones mobiles de diverses sociétés.

Bob souhaite découvrir une relation entre les fonctionnalités d'un téléphone mobile (par exemple : RAM, mémoire interne, etc.) et son prix de vente. Mais il n'est pas très doué en Machine Learning. Il a donc besoin d'aide pour résoudre ce problème.

Dans ce problème, on n'a pas besoin de prédire le prix réel mais une fourchette de prix indiquant le montant du prix.

**Source datasets :** <https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification/data>

### **Algorithmes Utilisés**

- **KMeans** : Partitionne les données en k groupes en minimisant la variance intra-cluster en utilisant la distance euclidienne.
- **GMM** : Modélise chaque cluster sous forme de gaussiennes et maximise la probabilité d'appartenance d'un point à un cluster.

### **Résultats des Algorithmes**

- **KMeans et GMM** ont donné des résultats similaires en termes de classification.
- **Indice de silhouette** :
  - **KMeans** : 0.06
  - **GMM** : 0.05

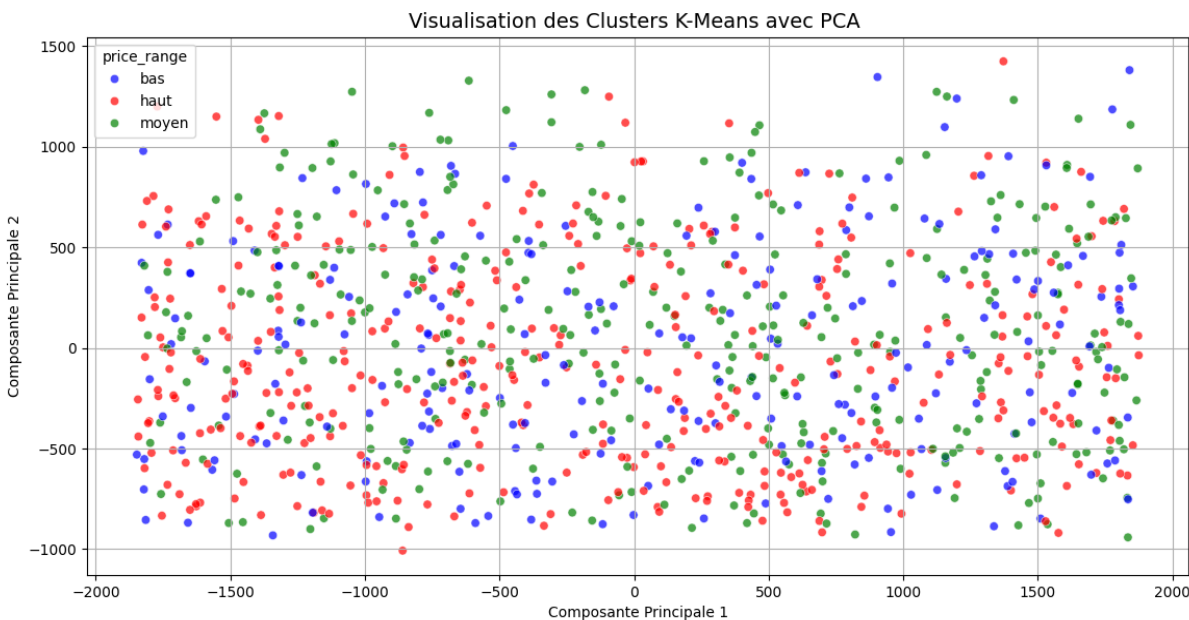
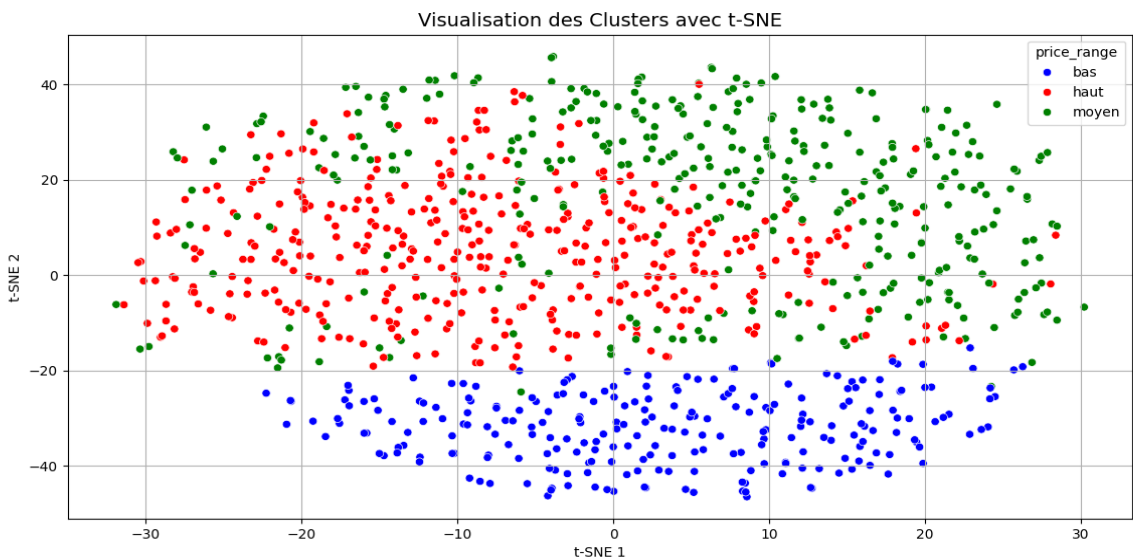
→ *Cet indice faible indique que la séparation entre les clusters est faible, ce qui suggère que les frontières entre les classes ne sont pas très nettes.*

### **Visualisation des Clusters**

- **PCA** : Une méthode linéaire pour réduire les dimensions des données. Ici, les classes obtenues n'étaient pas bien séparées avec PCA, ce qui peut indiquer une certaine complexité des données.
  - **T-SNE** : Une méthode non linéaire plus puissante pour visualiser des clusters. Avec t-SNE, les classes étaient bien séparées, ce qui montre que, même si les algorithmes
- *Peinent à séparer linéairement les données, les structures sous-jacentes sont visibles dans un espace réduit non linéaire.*

## Interprétation des Résultats et Visualisation

- **Indice de silhouette** : Le faible score de silhouette signifie que, bien que des clusters aient été formés, les distances intra-cluster et inter-cluster ne sont pas suffisamment distinctes. Les points sont donc mal répartis par rapport à leurs clusters respectifs.
- **Visualisation avec le nuage de points (scatter plot)** :
  - **PCA** n'a pas montré une bonne séparation, probablement en raison de la nature non linéaire des données.
  - **T-SNE** a révélé des clusters plus clairement séparés, suggérant que les données peuvent être mieux interprétées avec des méthodes non linéaires.



## Conclusion

Même si les résultats avec KMeans et GMM sont similaires et que les indices de silhouette sont faibles, la visualisation avec t-SNE montre qu'il existe des structures intéressantes dans les données. Cela peut indiquer qu'une analyse plus approfondie ou des algorithmes plus adaptés à des données complexes sont nécessaires pour améliorer la classification.