

# Rapport de Projet : Comparaison des Algorithmes d'Intelligence Artificielle

---

## Objectif du projet :

Ce projet vise à analyser et comparer les performances de trois algorithmes d'intelligence artificielle (**Arbre de décision**, **Clustering K-Means**, **Réseau de neurones**) sur une base de données réelle afin d'évaluer leurs forces, leurs faiblesses, et leur adéquation à différentes tâches.

---

## 1. Description des données

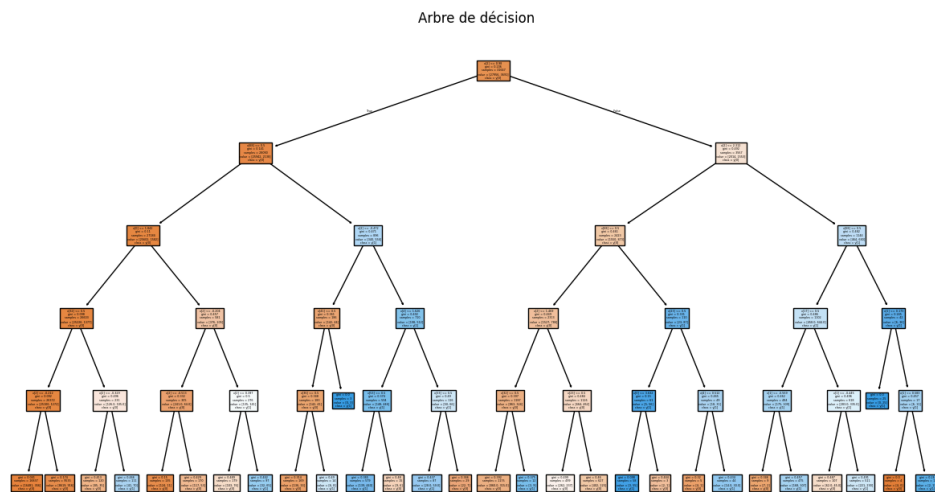
### Base de données utilisée : Bank Marketing Dataset

- **Source** : UCI Machine Learning Repository.
  - **Taille** :
    - Nombre d'échantillons : **45,211**.
    - Nombre de variables : **17** (16 features + 1 target).
  - **Objectif** :
    - Supervisé : Prédire si un client souscrit à un dépôt à terme bancaire ("yes"/"no").
    - Non supervisé : Découvrir des structures cachées dans les données via le clustering.
- 

## 2. Algorithmes implémentés

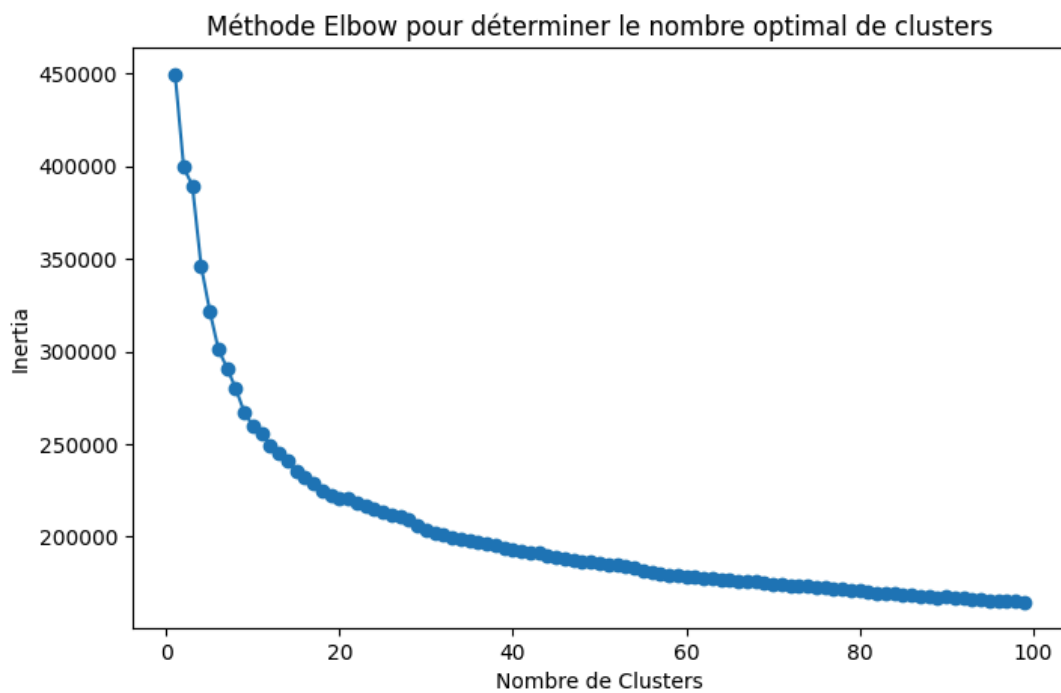
### 2.1. Arbre de décision

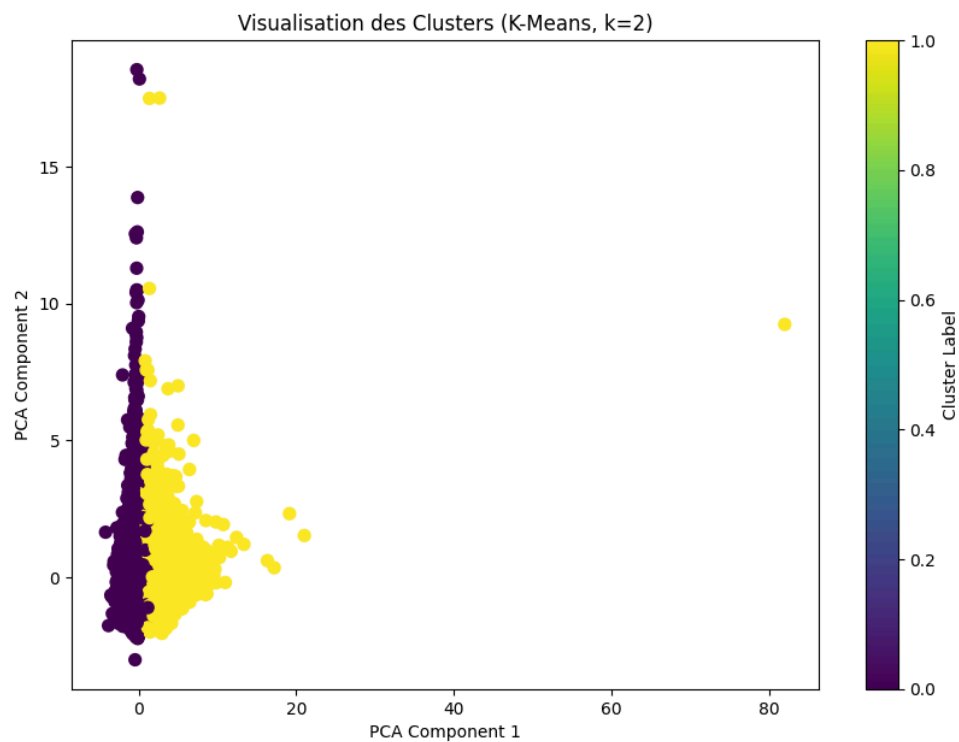
- Utilisé pour une tâche de classification supervisée.
- Critère : **Gini impurity**.
- Hyperparamètres optimisés : profondeur maximale.



## 2.2. Clustering K-Means

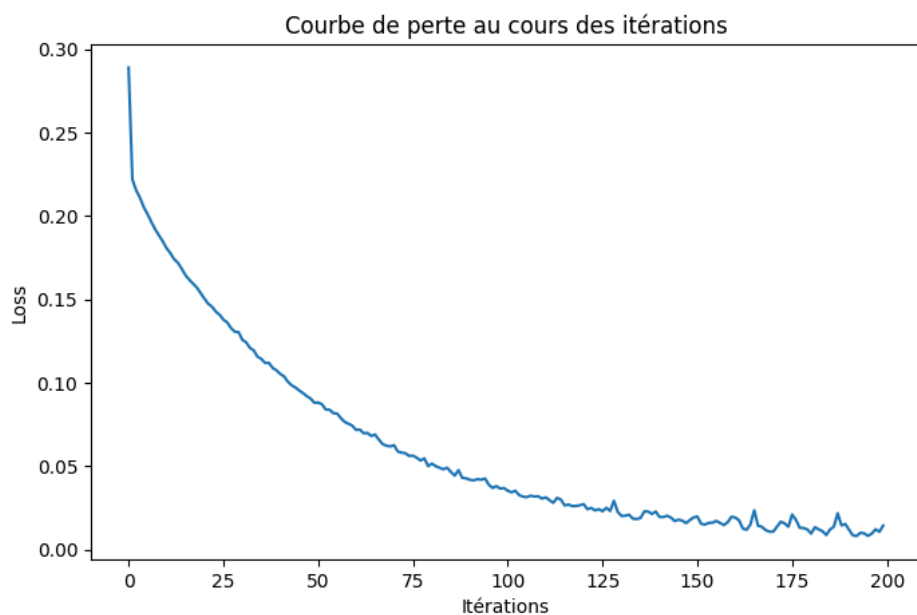
- Utilisé pour la segmentation des clients.
- Nombre de clusters : **2** (déterminé via la méthode Elbow).





### 2.3. Réseau de neurones (Perceptron multicouche)

- Utilisé pour une tâche de classification supervisée.
- Architecture : 2 couches cachées (100 neurones, 50 neurones), activation **ReLU**, optimiseur **Adam**.



---

### 3. Résultats

#### 3.1. Résultats globaux

Critères	Arbre de décision	K-Means	Réseau de neurones
Tâche	Classification supervisée	Clustering non supervisé	Classification supervisée
Performance globale	<b>90.0% accuracy (test)</b>	Score de silhouette : 0.23	<b>88.1% accuracy (test)</b>
Interprétabilité	Très bonne	Moyenne	Faible
Complexité	Faible	Faible	Élevée

#### 3.2. Comparaison des performances supervisées (Classification)

Critères	Arbre de décision	Réseau de neurones
Accuracy (test)	<b>90.0%</b>	<b>88.1%</b>
Recall (classe "yes")	<b>35%</b>	<b>46%</b>
F1-Score (classe "yes")	<b>46%</b>	<b>48%</b>
Complexité computationnelle	Faible	Élevée

#### 3.3. Résultats non supervisés (K-Means)

Critères	Valeur
Score de silhouette	<b>0.23</b> (faible)
Score de Davies-Bouldin	<b>1.96</b> (moyen)
Clusters détectés	<b>2 clusters</b>
Cluster 0	Clients ayant peu d'activité passée.
Cluster 1	Clients plus actifs, avec plusieurs contacts.

---

## 4. Analyse des résultats

### 4.1. Arbre de décision

- **Forces :**
  - Interprétabilité : La structure de l'arbre permet d'identifier les règles de décision.
  - Bonne précision globale (90% sur le test).
- **Limites :**
  - Mauvaise détection de la classe minoritaire "yes" (recall de 35%).
  - Biais vers la classe majoritaire.

### 4.2. Clustering K-Means

- **Forces :**
  - Capacité à segmenter les clients sans supervision.
  - Identification de 2 segments : Cluster 0 (clients peu actifs), Cluster 1 (clients plus actifs).
- **Limites :**
  - Faible qualité des clusters (score de silhouette : 0.23).
  - Non adapté pour des données complexes.

### 4.3. Réseau de neurones

- **Forces :**
  - Haute capacité d'apprentissage, notamment sur l'ensemble d'entraînement (99.5% accuracy).
  - Meilleure performance pour la classe "yes" que l'arbre de décision.
- **Limites :**
  - Surapprentissage évident (écart de performance entre entraînement et test).
  - Complexité computationnelle élevée.

---

## 5. Recommandations

### 5.1. Pour des tâches supervisées (classification) :

- L'**arbre de décision** est recommandé si l'interprétabilité est importante et que les performances légèrement inférieures sont acceptables.
- Le **réseau de neurones** est préférable si des performances maximales sont recherchées et que les ressources computationnelles ne sont pas un problème. Cependant, il nécessite une optimisation pour réduire le surapprentissage.

### 5.2. Pour des tâches non supervisées (segmentation) :

- Les résultats de **K-Means** indiquent une faible qualité des clusters. Il est recommandé de :
  - Tester d'autres algorithmes comme DBSCAN ou Gaussian Mixture Models.
  - Effectuer une analyse approfondie des variables pour améliorer les clusters.

### 5.3. Améliorations possibles :

- **Arbre de décision** :
  - Optimiser les hyperparamètres (profondeur, critère).
- **Réseau de neurones** :
  - Augmenter le nombre d'itérations pour améliorer la convergence.
  - Ajuster le taux d'apprentissage.
- **K-Means** :
  - Augmenter le nombre de clusters pour explorer plus de segments.
  - Réduire les variables bruitées ou redondantes pour améliorer la qualité des clusters.

---

## 6. Conclusion

Chaque algorithme a des forces spécifiques :

- **Arbre de décision** : Bonne précision et interprétabilité pour les tâches supervisées.
- **Réseau de neurones** : Haute performance brute, mais risque de surapprentissage.
- **K-Means** : Utile pour des tâches exploratoires, mais améliorable pour des résultats exploitables.

Ce projet montre que le choix de l'algorithme dépend du contexte et des besoins spécifiques (interprétabilité, performance brute, complexité).