

## Clustering

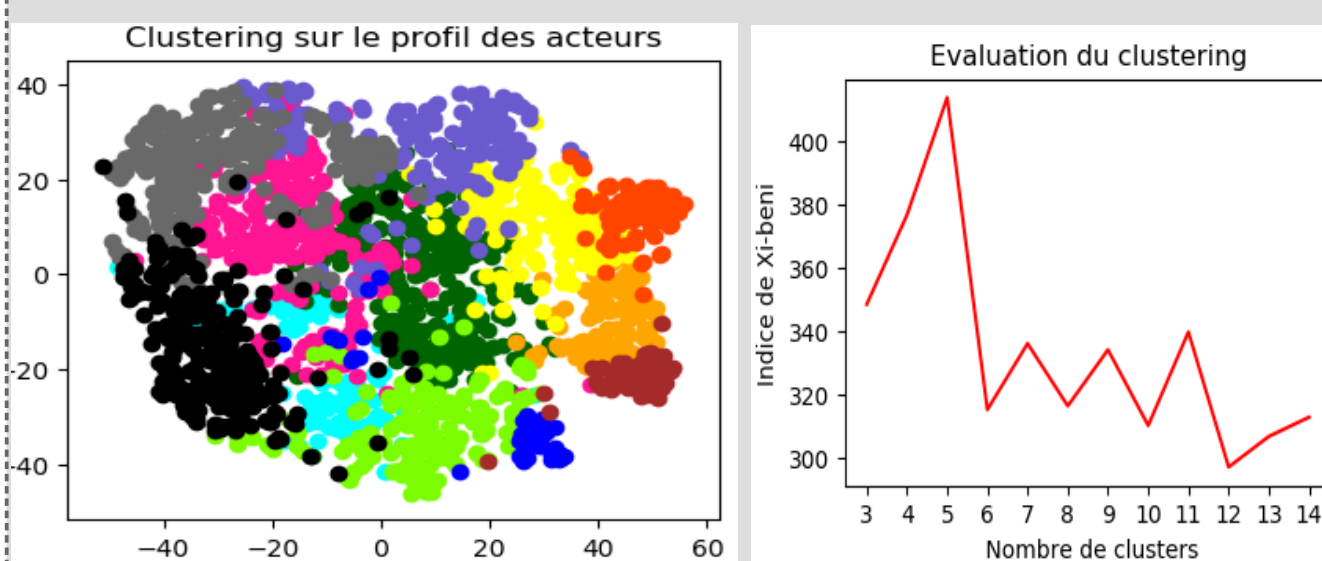
**Description :** Les acteurs de cinéma se retrouvent de plus en plus à jouer dans des films de genres différents. Ainsi, plusieurs profils d'acteurs semblent se détacher.

**Objectifs :**

- Identifier les différents profils d'acteurs.
- Déterminer le profil correspondant à un acteur en fonction des films dans lesquels il a joué.

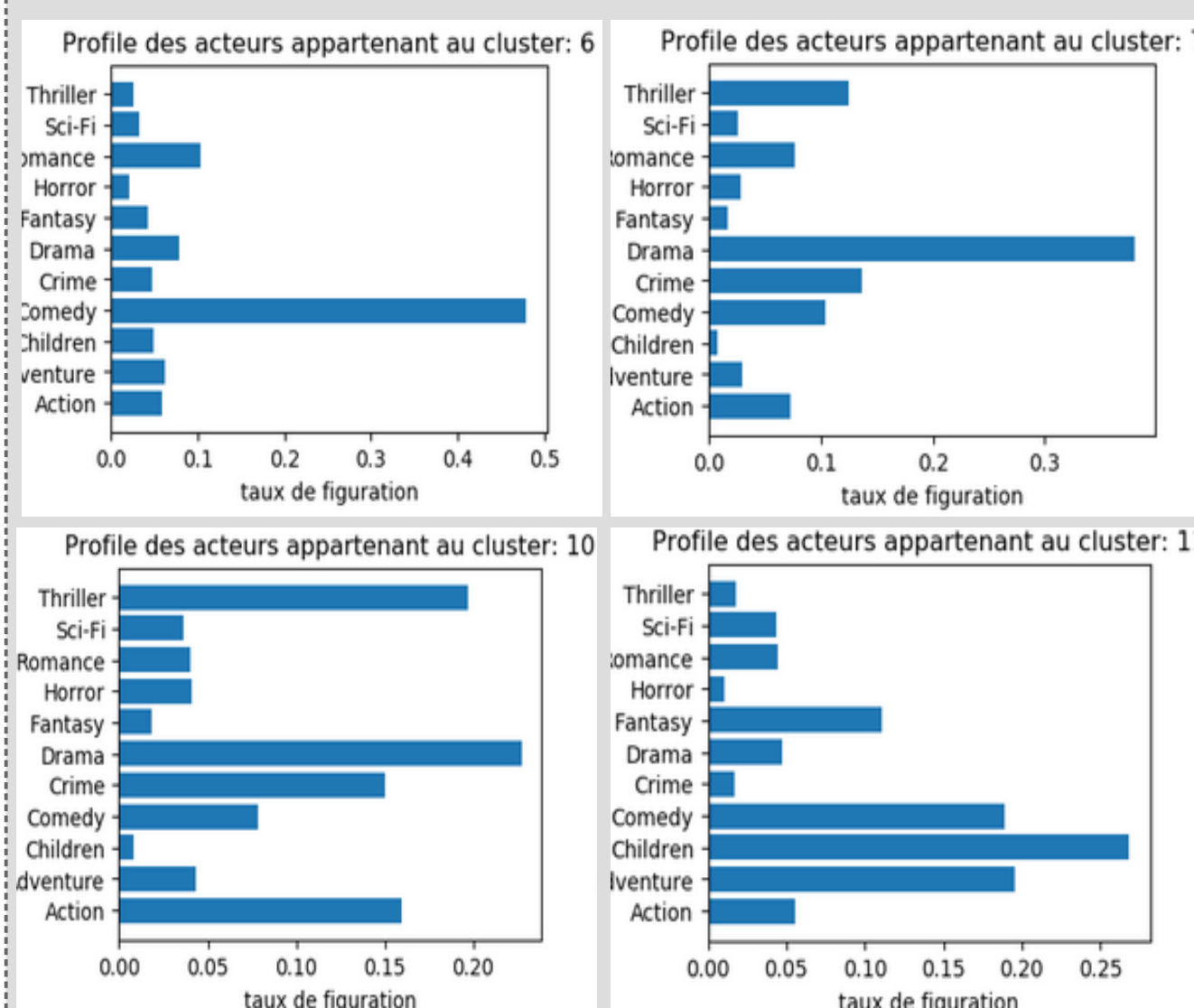
**Méthode :** Clustering sur les profils d'acteurs en utilisant l'algorithme des K-moyennes.

**Résultats :**



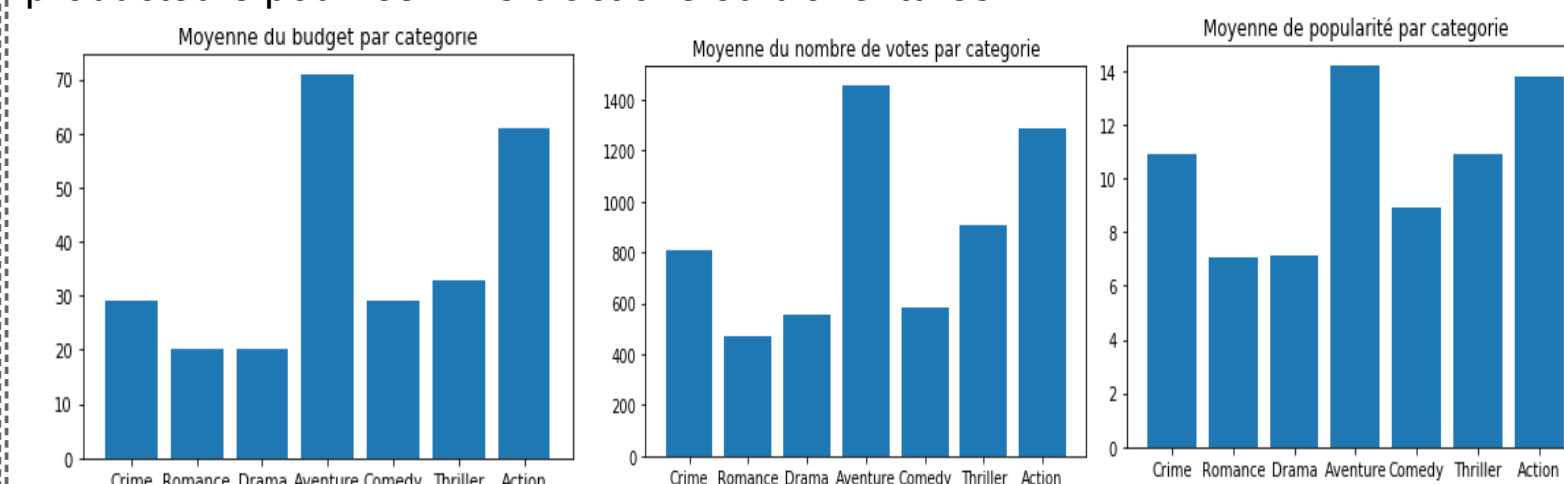
Ce clustering fait apparaître 12 groupes correspondant chacun à un profil d'acteur.

**Visualisation des caractéristiques des profils : (4 / 12)**



## Classification

**Introduction :** La moyenne par catégories du budget, du nombre de votes ainsi que de la popularité semblent indiquer un plus grand intérêt des spectateurs et des producteurs pour les films d'actions ou d'aventures.



**Objectifs :** Notre approche consiste donc à consolider ou à banaliser ces observations basées sur la moyenne au travers d'une classification binaire supervisée. Une première classe sera constituée des films des genres qui semblent les plus attractifs (Action & Aventure) et l'autre classe sera constituée des films des genres les moins attractifs (Drame & Romance).

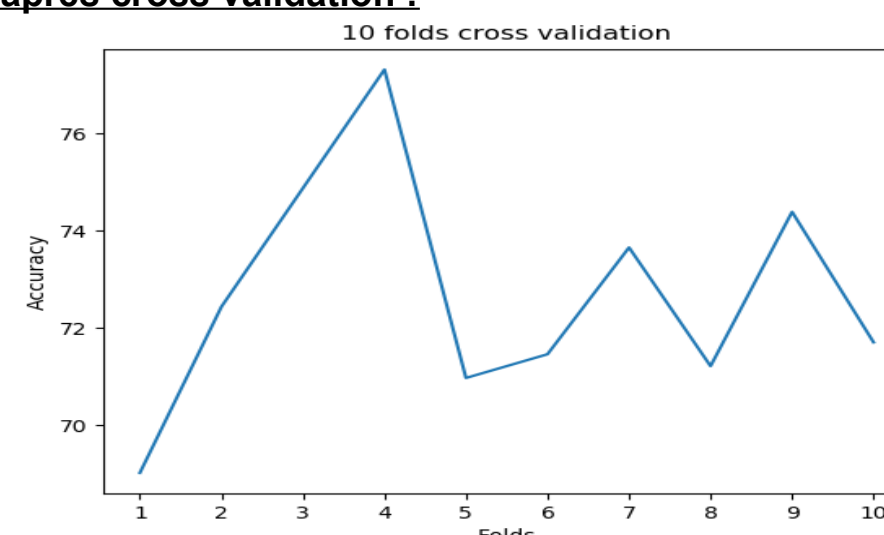
**Méthodes :** Algorithme des K plus proches voisins & arbres de décision.

**Résultats Obtenus pour les KPPV après cross validation :**

*Accuracy moyen : 72,7 %*

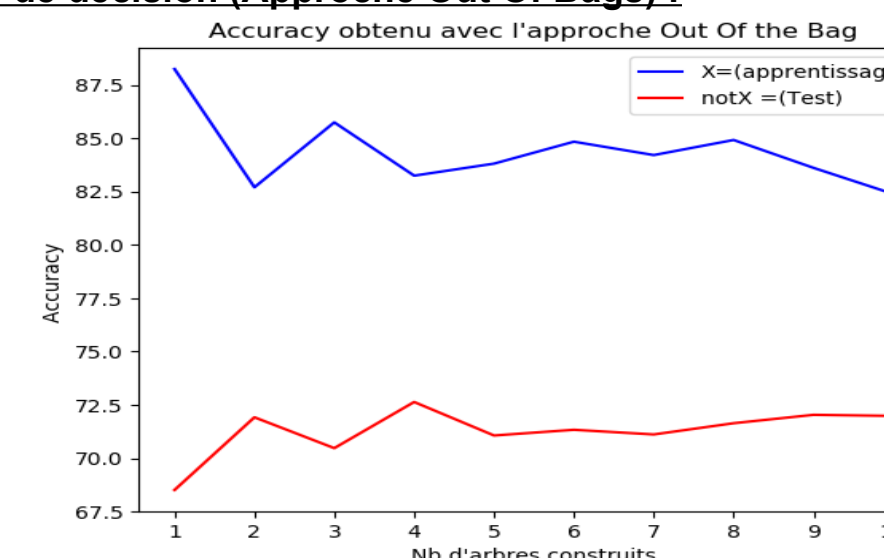
*Ecart type : 2,26*

*Best value of K : 5*



**Résultats Obtenus avec les arbres de décision (Approche Out Of Bags) :**

*Accuracy with 10 trees: 72,5 %*



**Conclusion :**

L'évaluation de nos algorithmes nous donne globalement un taux de bonne classification autour de 72 %. Ainsi, dans 3/4 des cas, les films au gros budgets et qui jouissent d'une grande popularité seront des films d'action & d'aventure. Une explication à ce phénomène consisterait à dire que les films d'actions et d'aventure touchent autant le public masculin que le public féminin tandis que les films Drame et Romance n'ont généralement d'écho qu'après des femmes.

## Régression

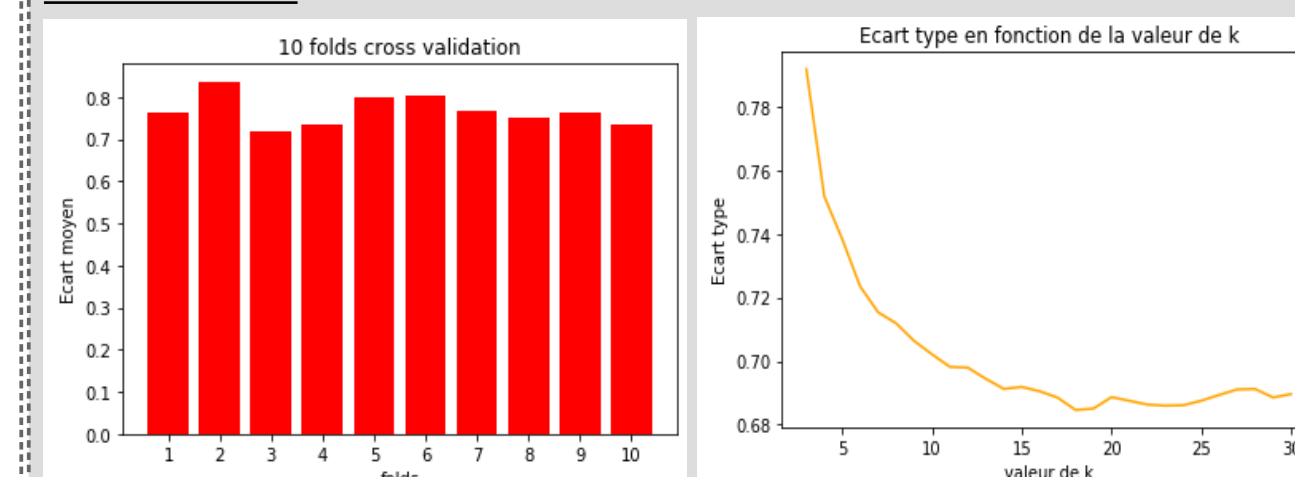
**Introduction :** La note moyenne donnée à un film sur internet peut être influencée par plusieurs facteurs. Ainsi, le budget, l'année de sortie, la popularité ou encore la durée d'un film peuvent servir d'indicateurs pour la prédiction de la note moyenne que recevra un film sur les plateformes de streaming.

**Objectifs :** Notre approche consiste donc, au travers de la résolution d'un problème de régression, à :

- Identifier quelles caractéristiques ont le plus d'influence sur la note moyenne.
- Approcher au mieux la note associée à un film en se basant sur ces caractéristiques.

**Méthodes:** Algorithme des K plus proches voisins & algorithme du Perceptron.

**Résultats obtenus avec les KPPV après cross validation :**



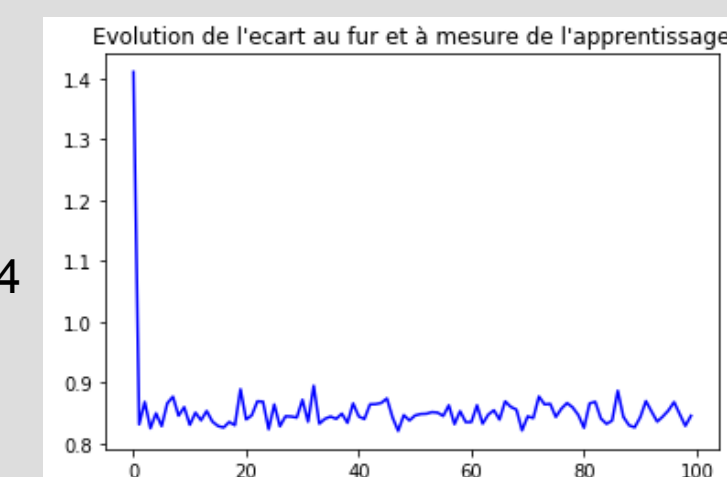
*Meilleure valeur de K : k>20*

*Écart moyen : 0,76*

**Résultats obtenus avec l'algorithme du perceptron :**

**Kernel utilisé:  $(x_1, \dots, x_d) \rightarrow (1, x_1, \dots, x_d)$ .**

Écart moyen obtenu : 0,84



**Conclusion :**

L'application de nos algorithmes nous donnent des résultats peu satisfaisants. Ainsi, nous considérons qu'un écart moyen autour de 0,8 est très approximatif. Cependant il se pourrait qu'on ait de meilleurs résultats en proposant une autre méthode de noyau dans le cas du perceptron ou une autre mesure de « distance » pour les k plus proches voisins.