

# Benchmark des Algorithmes de Recommandation et Prédiction de Popularité

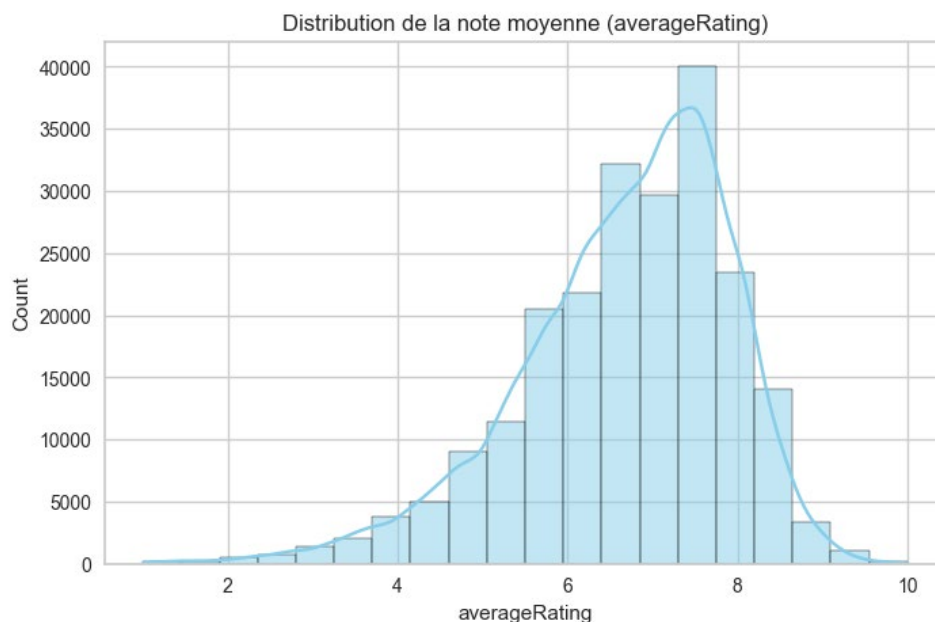
Analyse comparative des modèles de prédiction de la popularité des films en utilisant PyCaret.

## 1. Analyse Exploratoire

### 1.1 Distribution de la Note Moyenne

L'histogramme ci-dessous montre la distribution de la variable cible 'averageRating'.

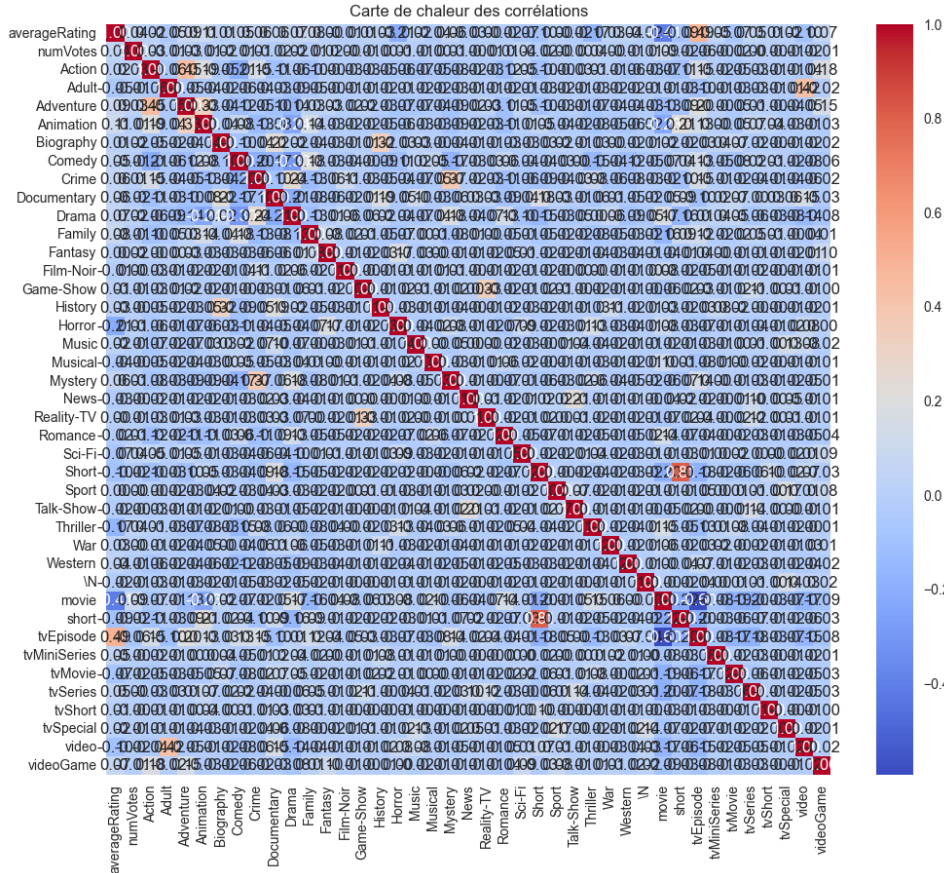
Analyse : Ce graphique permet d'observer la répartition des notes attribuées aux films. Une distribution avec une concentration autour d'une valeur moyenne indique une homogénéité dans les évaluations. Toute dispersion ou asymétrie pourrait suggérer des biais ou des groupes distincts dans les préférences du public.



### 1.2 Carte de Chaleur des Corrélations

La carte de chaleur suivante illustre les corrélations entre les variables numériques du dataset.

Analyse : Cette visualisation permet d'identifier les relations linéaires entre les variables. Les zones aux coefficients élevés indiquent une forte corrélation, ce qui peut être utile pour repérer des redondances ou déterminer l'influence de certaines variables sur la cible.



## 2. Modélisation et Comparaison des Modèles

Le modèle sélectionné est **CatBoostRegressor**. Ce modèle a été choisi sur la base de ses performances, notamment en termes de  $R^2$  et de RMSE, après une comparaison parmi plusieurs modèles.

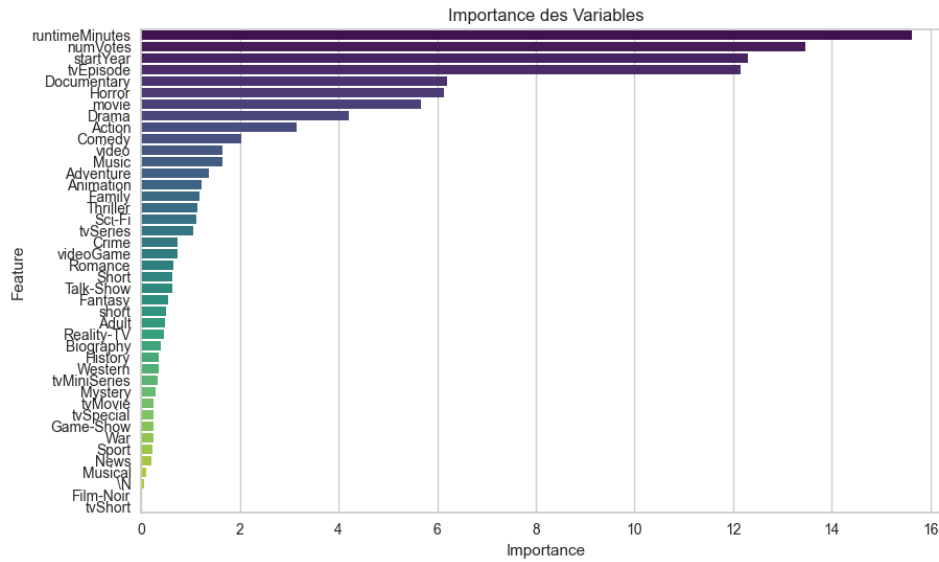
### 2.1 Classement des Modèles

Le tableau ci-dessous présente le classement des modèles selon leurs performances :

Rang	Modèle	MAE	MSE	RMSE	R2	RMSLE	MAPE
1	3	0.6174	0.7234	0.8505	0.5406	0.1311	0.1106
2	Mean	0.6161	0.7231	0.8503	0.5344	0.1308	0.1102
3	2	0.6155	0.7188	0.8478	0.5343	0.1299	0.1095
4	4	0.6165	0.7211	0.8492	0.5334	0.1296	0.1093
5	1	0.6133	0.7184	0.8476	0.5321	0.1305	0.1097
6	0	0.618	0.7336	0.8565	0.5314	0.1329	0.1117
7	Std	0.0017	0.0056	0.0033	0.0033	0.0012	0.0009

## 3. Importance des Variables

Le graphique ci-dessous présente l'importance des variables tel qu'extraites du modèle optimisé. Les variables classées en tête indiquent leur impact majeur sur la prédiction de la note moyenne.



#### 4. Justification du Modèle Sélectionné

Le modèle sélectionné est **CatBoostRegressor**. Ce choix est justifié par sa capacité à maximiser le coefficient de détermination ( $R^2$ ) tout en minimisant l'erreur quadratique moyenne (RMSE). Les résultats obtenus après tuning indiquent que ce modèle offre un bon compromis entre performance et généralisation.