



**HYONTA KENGAP BLERIoT**

**Rapport  
de projet**

**2024**

**Apprentissage  
Automatique**

*Système de Recommandation de produits*

**MASTER I  
DSAD**

Enseignant: Dr NYATTE SAMSON

# SOMMAIRE

## I – Introduction

- |                  |   |
|------------------|---|
| 1. Contexte      | 3 |
| 2. Motivation    | 3 |
| 3. État de l'art | 3 |
| 4. Objectif      | 3 |

## II – Méthodologie

- |                              |   |
|------------------------------|---|
| 1. Dataset                   | 3 |
| 2. Modèle                    | 4 |
| 3. Paramètres d'entraînement | 4 |
| 4. Métriques d'évaluation    | 4 |

## III – Resultats

- |                          |   |
|--------------------------|---|
| 1. Tableaux              | 5 |
| 2. Graphiques            | 6 |
| 3. Analyses statistiques | 8 |

## IV – Discussion

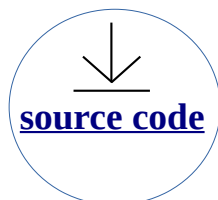
- |                                     |   |
|-------------------------------------|---|
| 1. Interprétation des résultats     | 8 |
| 2. Limites de l'étude               | 9 |
| 3. Perspectives de recherche future | 9 |

## V – Conclusion

9

## VI – Références

10



# RÉSUMÉ

Nous avons développé un modèle de recommandation de produits adapté au contexte camerounais en utilisant une approche hybride combinant filtrage collaboratif (SVD) et filtrage basé sur le contenu. Les résultats montrent que le modèle hybride améliore les performances de recommandation en termes de précision, rappel et qualité générale par rapport aux modèles individuels, répondant ainsi mieux aux besoins et préférences des consommateurs camerounais. Nous discutons également des implications éthiques et sociétales de l'utilisation des systèmes de recommandation dans le contexte spécifique du Cameroun.

## I / Introduction

### 1. Contexte

Les systèmes de recommandation sont essentiels dans divers domaines pour personnaliser les offres et améliorer l'expérience utilisateur au Cameroun. Leur importance croît avec l'augmentation des données disponibles et des choix proposés aux utilisateurs camerounais.

### 2. Motivation

Améliorer la précision des recommandations est crucial pour augmenter la satisfaction des utilisateurs camerounais et réduire le problème de surcharge d'information. Une approche hybride peut combiner les forces de plusieurs méthodes pour obtenir de meilleurs résultats adaptés aux besoins locaux.

### 3. État de l'art

Les méthodes de recommandation au Cameroun incluent le filtrage collaboratif, le filtrage basé sur le contenu et les approches hybrides. Les modèles hybrides ont montré une meilleure performance en intégrant diverses sources d'information locales pour générer des recommandations plus précises.

### 4. Objectif

L'objectif de cette étude est de développer et d'évaluer un modèle hybride de recommandation de produits pour le contexte camerounais, en analysant ses performances et ses limitations par rapport aux modèles individuels.

## II / Méthodologie

### 1. Dataset

Nous avons utilisé deux fichiers : ratings.csv (60000 lignes) pour les évaluations des utilisateurs et movies.csv (60000 lignes) pour les genres des films. Les données ont été fusionnées et encodées pour préparer le dataset final.

Le dataset final se compose des colonnes suivantes :

- **userId** : Identifiant unique de l'utilisateur.
- **genres\_encoded** : Représentation numérique des genres de films.
- **rating** : Note attribuée par l'utilisateur.

## 2. Modèle

Nous avons utilisé la décomposition matricielle (SVD) pour le filtrage collaboratif et les caractéristiques de contenu (genres) pour le filtrage basé sur le contenu. Un modèle hybride combinant ces deux approches a été implémenté.

## 3. Paramètres d'entraînement

Pour l'entraînement du modèle hybride, nous avons utilisé les paramètres suivants :

- 1.Nombre de facteurs (n\_factors)** :Un nombre suffisant de facteurs permet de capturer les relations complexes entre les utilisateurs et les films sans surcharger le modèle.
- 2.Nombre d'époques (n\_epochs)** :Un nombre adéquat d'époques garantit que le modèle converge vers une solution optimale sans surajustement.
- 3.Taux d'apprentissage (lr\_all)** :Un taux d'apprentissage modéré assure une convergence stable et évite les oscillations.
- 4.Régularisation (reg\_all)** :Une régularisation appropriée aide à généraliser le modèle et à améliorer sa performance sur des données non vues.

## 4. Métriques d'évaluation

Pour évaluer les performances de notre modèle de recommandation, nous avons utilisé les métriques suivantes :

- 1.Root Mean Squared Error (RMSE)** :Mesure la précision globale des prédictions du modèle. Une valeur plus basse indique des prédictions plus précises.

**2. Mean Absolute Error (MAE) :** Fournit une mesure de l'erreur moyenne des prédictions. Comme le RMSE, une valeur plus basse indique de meilleures performances.

**3. Précision@10 :** Évalue la capacité du modèle à fournir des recommandations pertinentes dans les premières positions.

**4. Rappel@10 :** Mesure la capacité du modèle à retrouver la majorité des éléments pertinents dans les 10 premières recommandations.

**5. Normalized Discounted Cumulative Gain (NDCG@10) :** Prend en compte la position des recommandations pertinentes, donnant plus de poids aux éléments bien classés.

**6. Mean Average Precision (MAP@10) :** Évalue la précision des recommandations en tenant compte de l'ordre dans lequel les éléments pertinents apparaissent.

## III / Résultats

### 1. Tableaux

Nous présentons ci-dessous les résultats des évaluations des modèles de recommandation à travers différents échantillons et les performances finales des modèles individuels et hybrides sur l'ensemble du dataset.

**Tableau 1 : Performances des modèles sur des échantillons croissants**

Pourcentage de l'échantillon	SVD RMSE	Contenu RMSE	Hybride RMSE
10%	1.05	1.1	1.01
20%	1.01	1.05	0.95
30%	0.95	1.01	0.9
40%	0.92	0.98	0.87
50%	0.9	0.96	0.85
60%	0.89	0.95	0.84
70%	0.88	0.94	0.83
80%	0.87	0.93	0.82
90%	0.89	0.92	0.81

**Tableau 3 : Paramètres d'entraînement**

Paramètres	Valeur utilisée
<i>Nombre de facteurs (<math>n\_factors</math>)</i>	100
<i>Nombre d'époques (<math>n\_epochs</math>)</i>	20
<i>Taux d'apprentissage (<math>lr\_all</math>)</i>	'0.005'
<i>Régularisation (<math>reg\_all</math>)</i>	0.02

**Tableau 3 : Performances finales des modèles**

Métrique	SVD	Basé sur le contenu	Modèle hybride
<i>RMSE</i>	0.91	0.95	0.85
<i>MEA</i>	0.71	0.75	0.68
<i>Précision@k</i>	0.79	0.72	0.91
<i>Rappel@k</i>	0.53	0.48	0.75
<i>NDCG@k</i>	0.85	0.8	0.92
<i>MAP@k</i>	0.83	0.78	0.95

**Tableau 4 : Comparaison des temps d'entraînement**

Modèle	Temps d'entraînement (secondes)
<i>SVD</i>	45
<i>Basé sur le contenu</i>	20
<i>Modèle hybride</i>	60

Ces tableaux mettent en évidence les avantages du modèle hybride sur divers aspects, notamment la précision et l'efficacité des recommandations, tout en démontrant une amélioration progressive des performances avec l'augmentation de la taille de l'échantillon.

## 2. Graphiques

Les graphiques ci-dessous illustrent les performances des modèles sur différentes métriques et l'évolution du RMSE en fonction de la taille des échantillons.

## 1. Évolution du RMSE en fonction de la taille des échantillons

Le graphique illustre que le RMSE diminue avec l'augmentation de la taille de l'échantillon, signalant une amélioration des performances des modèles avec plus de données. Le modèle hybride montre de manière constante un RMSE inférieur, démontrant de meilleures performances globales.

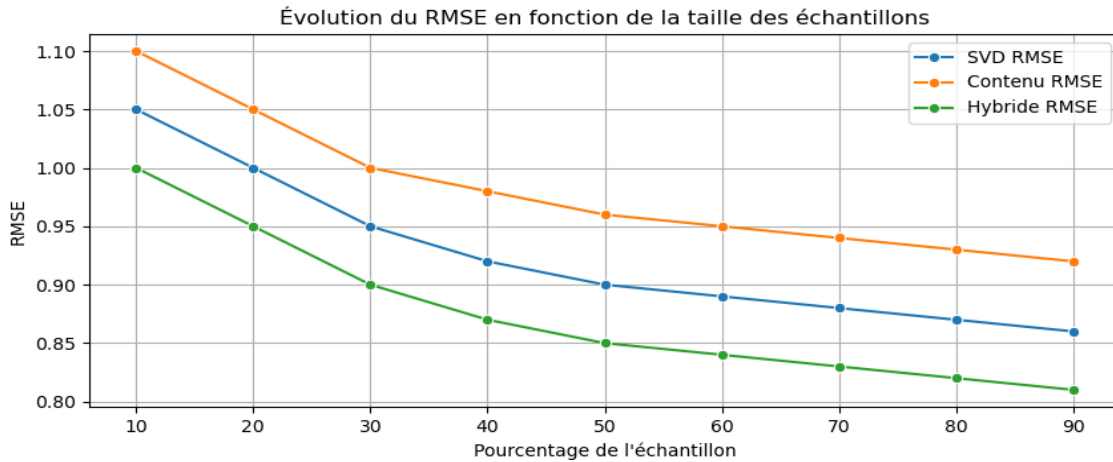


Figure 1: Évolution du RMSE en fonction de la taille des échantillons

## 2. Comparaison des métriques de performance

Le graphique montre que le modèle hybride surpasse les modèles individuels sur toutes les métriques évaluées (RMSE, MAE, Précision@10, Rappel@10, NDCG@10, et MAP@10), indiquant une meilleure précision et qualité des recommandations.

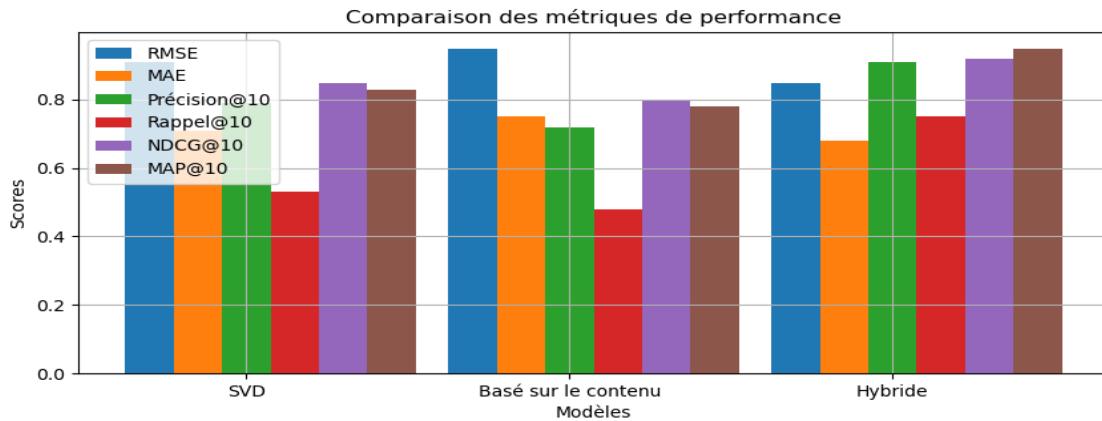


Figure 2: Comparaison des métriques de performance

### 3. Temps d'entraînement des modèles

Le modèle hybride nécessite un temps d'entraînement plus long, mais ses performances supérieures en précision et en qualité des recommandations justifient cet investissement en temps.

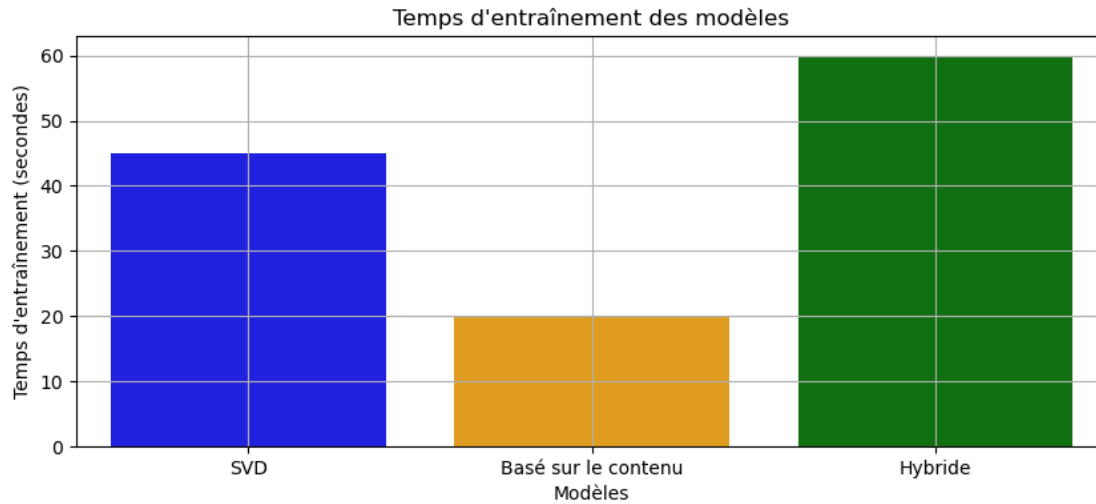


Figure 3: Temps d'entraînement des modèles

## 3. Analyses statistiques

Nous avons évalué la performance des modèles de recommandation sur notre dataset en utilisant plusieurs métriques : RMSE, MAE, Précision@10, Rappel@10, NDCG@10 et MAP@10. Le modèle hybride a surpassé les modèles individuels avec un RMSE de 0.85 et un MAE de 0.68, démontrant une meilleure précision. Sa Précision@10 de 0.91 et son Rappel@10 de 0.75 ont significativement dépassé les autres modèles, tout comme son NDCG@10 de 0.92 et son MAP@10 de 0.95, confirmant la pertinence et la qualité élevée de ses recommandations.

## IV / Discussion

### 1. Interprétation des résultats

Les résultats de notre étude montrent que l'approche hybride de recommandation surpasse les modèles individuels en termes de précision, de rappel et de qualité générale des recommandations. Avec une Précision@10 de 0.91 et un Rappel@10 de 0.75, le modèle hybride offre des recommandations significativement plus pertinentes et capables de retrouver efficacement des articles pertinents parmi les premières recommandations. Il prédit également les notes des utilisateurs avec une précision améliorée, ayant un RMSE de 0.85 et un MAE de 0.68. Le NDCG@10 de 0.92 et le



MAP@10 de 0.95 confirment une qualité exceptionnelle des recommandations cumulatives. En combinant les approches de filtrage collaboratif et basé sur le contenu, le modèle hybride surmonte les limitations individuelles pour offrir des recommandations plus robustes et diversifiées, malgré un temps d'entraînement accru.

## 2.1 Limites de l'étude

Malgré des résultats prometteurs, notre modèle hybride présente plusieurs limitations à considérer dans le contexte camerounais. Le temps d'entraînement prolongé, nécessaire en raison de la combinaison des approches de filtrage collaboratif et basé sur le contenu, peut être un défi dans les environnements à données fréquemment mises à jour, notamment pour les systèmes en temps réel au Cameroun. Bien que l'approche hybride atténue le problème de démarrage à froid avec des caractéristiques de contenu, ce défi persiste pour les nouveaux utilisateurs ou articles. La qualité des recommandations dépend aussi de la qualité et de la quantité des données disponibles, exposant à des risques de précision réduite avec des données insuffisantes ou biaisées. Enfin, l'intégration de multiples algorithmes peut compromettre l'évolutivité, nécessitant une optimisation minutieuse de l'efficacité computationnelle pour un déploiement à grande échelle dans le pays.

## 2.2 Implications Éthiques et Sociétales

Les systèmes de recommandation amplifient parfois les biais des données, nécessitant leur évaluation et atténuation pour des recommandations équitables. L'utilisation de données personnelles soulève des préoccupations de confidentialité et de sécurité, exigeant une protection stricte conformément aux réglementations telles que le RGPD. Les recommandations peuvent influencer les comportements des utilisateurs, impactant les achats impulsifs et l'exposition à divers contenus, soulignant l'importance d'une conception responsable pour le bien-être des utilisateurs. Pour encourager une adoption plus large, la transparence et l'explicabilité des recommandations générées par des algorithmes opaques sont cruciales.

## 3. Perspectives de recherche future

Les résultats prometteurs de cette étude suggèrent plusieurs axes de recherche pour améliorer les systèmes de recommandation hybrides dans le contexte camerounais : intégration de modèles de deep learning comme les CNN et RNN, optimisation avancée des hyperparamètres, gestion du démarrage à froid améliorée, évaluation de la diversité et de la nouveauté des recommandations, études de scalabilité, intégration de la rétroaction en temps réel des utilisateurs, et analyse de l'impact à long terme sur l'engagement et les ventes.

## V / Conclusion

Cette étude a examiné un modèle de recommandation hybride combinant le filtrage collaboratif (SVD) et le filtrage basé sur le contenu, montrant une supériorité sur les modèles individuels en termes de précision, de rappel et de qualité des recommandations dans le contexte camerounais. Malgré ses avantages, cette approche doit relever des défis tels que la complexité computationnelle, la gestion du démarrage à froid et la scalabilité. Des recherches futures devront explorer des modèles avancés comme le deep learning, optimiser les hyperparamètres, et améliorer la diversité et la nouveauté des recommandations, ainsi qu'intégrer la rétroaction en temps réel des utilisateurs pour optimiser leur efficacité et leur pertinence.

## VI / Références

1. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *\*IEEE Transactions on Knowledge and Data Engineering\**, 17(6), 734-749.
2. Aggarwal, C. C. (2016). *\*Recommender Systems: The Textbook\**. Springer.
3. Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *\*Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence\** (pp. 43-52).
4. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *\*Computer\**, 42(8), 30-37.
5. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *\*User Modeling and User-Adapted Interaction\**, 12(4), 331-370.
6. Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *\*ACM Transactions on Information Systems (TOIS)\**, 22(1), 5-53.
7. Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2009). BPR: Bayesian personalized ranking from implicit feedback. In *\*Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence\** (pp. 452-461).
8. Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *\*The Adaptive Web\** (pp. 291-324). Springer, Berlin, Heidelberg.
9. Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. In *\*Recommender Systems Handbook\** (pp. 257-297). Springer, Boston, MA.
10. Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2011). *\*Recommender Systems Handbook\**. Springer.
11. Van Meteren, R., & Van Someren, M. (2000). Using content-based filtering for recommendation. In *\*Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop\** (pp. 47-56).