

جامعة محمد الأول بوجدة  
UNIVERSITÉ MOHAMMED PREMIER OUJDA  
+٣٨٠٦٤٢٩٨٣٦٦٦٦ | ٠٥٣٦٧١٨٣٦



المدرسة الوطنية للعلوم التطبيقية  
École Nationale des Sciences Appliquées d'Oujda  
+٣٨٠٦٤٢٩٨٣٦٦٦٦ | ٠٥٣٦٧١٨٣٦

ROYAUME DU MAROC  
UNIVERSITÉ MOHAMMED PREMIER  
**École Nationale des Sciences Appliquées**  
Oujda - Maroc

## Mémoire de Fin d'Année

Filière : Ingénierie du Data Science et Cloud Computing

## Système de recommandation amélioré

Réalisé par :

Mouad AIT HA  
Mohamed LAKBAKBI

Membres de jury :

M. Abdelmounim KERKRI  
M. NAJI

Encadré par :

Mme. chaimae CHOUIEKH

## **Dédicace**

À tous les acteurs ayant contribué à la réussite de ce projet,

C'est avec une profonde gratitude que je tiens à saluer votre dévouement, votre expertise et votre collaboration exceptionnelle tout au long de cette entreprise ambitieuse. Vos efforts soutenus, votre engagement sans faille et votre synergie ont été les moteurs indéniables de notre réussite collective.

Cette dédicace est un hommage sincère à votre apport inestimable et indispensable dans la concrétisation de ce projet d'envergure. Que cette réalisation soit le témoin éloquent de notre quête perpétuelle d'excellence, d'innovation et de dépassement de soi.

Recevez ici l'expression de ma reconnaissance la plus vive pour votre précieuse contribution et votre collaboration exemplaire. Puissent les fruits de nos efforts conjugués inspirer de nouvelles avancées et ouvrir la voie vers de nouveaux défis à relever.

## **Remerciement**

Avant toute chose, nous tenons à exprimer notre profonde gratitude envers le Très-Haut pour nous avoir accordé la santé, la force et la persévérance nécessaires à la concrétisation de ce projet d'envergure. Nous adressons nos sincères remerciements à Madame Chaymae Chouiekh,

encadrante de ce projet, pour son accompagnement empreint de rigueur, de bienveillance et de disponibilité. Ses conseils avisés et son expertise éclairée ont été un guide précieux tout au long de cette entreprise et ont grandement contribué à son succès. Nos remerciements les plus vifs

vont également à l'ensemble du corps enseignant pour leur dévouement, leur générosité et leur patience sans faille malgré leurs lourdes charges. Leurs enseignements de grande valeur et leur soutien indéfectible ont été les ferment de notre développement intellectuel et de notre progression académique. C'est grâce au concours inestimable et au soutien indéfectible de toutes

ces personnes dévouées que nous avons pu mener à bien ce projet. Nous leur exprimons notre gratitude la plus sincère et nous leur dédions ce travail avec une immense fierté.

# Table des matières

<b>Liste des figures</b>	<b>6</b>
<b>Liste des Abréviations</b>	<b>7</b>
<b>Liste des tableaux</b>	<b>8</b>
<b>Résumé</b>	<b>9</b>
<b>Abstract</b>	<b>10</b>
<b>Introduction générale</b>	<b>11</b>
<b>1 Etat d'art</b>	<b>13</b>
1.1 Introduction . . . . .	13
1.1.1 Domaines d'utilisation . . . . .	14
1.2 Algorithmes existants . . . . .	14
1.2.1 Filtrage collaboratif . . . . .	14
1.2.2 Filtrage basé sur le contenu . . . . .	15
1.2.3 Recommandation hybride . . . . .	16
1.2.4 Problèmes et limites des systèmes de recommandation . . . . .	16
1.2.5 Techniques Avancées et Nouvelles Approches . . . . .	17
1.2.6 Évaluation des Systèmes de Recommandation . . . . .	18
1.3 Conclusion . . . . .	18
<b>2 Étude préliminaire</b>	<b>19</b>
2.1 Introduction . . . . .	19
2.2 Algorithmes de Recommandation classiques . . . . .	19
2.2.1 Filtrage Collaboratif . . . . .	20
2.2.2 Filtrage Basé sur le Contenu . . . . .	20
2.3 Algorithmes de Recommandation Avancés . . . . .	20
2.3.1 K-Nearest Neighbors (k-NN) . . . . .	21
2.3.2 Décomposition en valeurs singulières (SVD) . . . . .	21
2.3.3 Factorisation matricielle non négative (NMF) . . . . .	21
2.4 Sources de données . . . . .	21
2.4.1 Dataset des livres : booksdf.csv . . . . .	21
2.4.2 Dataset des utilisateurs : usersdf.csv . . . . .	22
2.4.3 Dataset des interactions : ratingsdf.csv . . . . .	22
2.4.4 Prétraitement des données . . . . .	22

2.5	Outils utilisés . . . . .	22
2.5.1	Visualisation des données . . . . .	22
2.5.2	Python et Jupyter Notebook . . . . .	23
2.5.3	Bibliothèques de Machine Learning . . . . .	23
2.5.4	Environnement de Développement et d'Exécution . . . . .	23
2.6	Conclusion . . . . .	23
<b>3</b>	<b>Réalisation du projet</b>	<b>25</b>
3.1	Introduction . . . . .	25
3.2	Collecte et prétraitement des données . . . . .	25
3.2.1	Collecte des Données . . . . .	25
3.2.2	Nettoyage et Prétraitement des Données . . . . .	26
3.2.3	Structuration des Données . . . . .	26
3.3	Exploration et Analyse des Données . . . . .	26
3.3.1	Statistiques Descriptives . . . . .	26
3.3.2	Visualisation des Données . . . . .	27
3.4	Implémentation des Algorithmes de Recommandation . . . . .	27
3.4.1	Filtrage Collaboratif . . . . .	28
3.4.2	Filtrage Basé sur le Contenu . . . . .	28
3.4.3	Algorithmes de Voisinage (K-Nearest Neighbors) . . . . .	28
3.4.4	Factorisation Matricielle . . . . .	29
3.4.5	Réseaux de Neurones et Autoencodeurs . . . . .	29
3.5	Évaluation des Modèles . . . . .	29
3.5.1	Métriques d'évaluation . . . . .	30
3.5.2	Validation croisée . . . . .	30
3.5.3	Comparaison des Performances . . . . .	30
3.5.4	Évaluation qualitative . . . . .	30
3.6	Déploiement du Système de Recommandation . . . . .	31
3.7	Cas d'utilisation et Scénarios d'Application . . . . .	32
3.8	Défis et Limites . . . . .	32
3.9	Conclusion . . . . .	33
<b>Conclusion générale</b>		<b>34</b>
<b>Glossaire</b>		<b>34</b>
<b>Bibliographie</b>		<b>35</b>

# Liste des figures

1.1	Recommendation livres Amazon. . . . .	14
1.2	Recommendation hybride. . . . .	16
2.1	les différents types de Systèmes de Recommandation . . . . .	24
3.1	graphes - visualisation des données . . . . .	27
3.2	web application . . . . .	31
3.3	donner le resultat . . . . .	31
3.4	après resultat . . . . .	32

# Liste des Abréviations

**K-NN** K-nearest neighbors

**SVD** Singular values decomposition

**NMF** Non-negative matrix factorization

## **Liste des tableaux**

## Résumé

En premier lieu, nous souhaitons exprimer notre sincère reconnaissance envers Dieu, le Tout-Puissant, pour nous avoir donné la santé, la force et la persévérance indispensables pour réaliser ce projet.

Madame **Chaymae Chouiekh**, notre encadrante de projet, mérite notre sincère gratitude pour son précieux encadrement, sa patience, sa rigueur et sa disponibilité tout au long de ce projet. Son accompagnement éclairé et ses conseils avisés ont joué un rôle essentiel dans la réussite de cette tâche.

Nous souhaitons également exprimer notre gratitude envers tous les enseignants pour leur engagement, leur générosité et leur patience malgré leurs nombreuses responsabilités. Leurs enseignements précieux et leur soutien constant ont joué un rôle essentiel dans notre développement intellectuel et notre avancement tout au long de notre parcours scolaire.

C'est grâce à l'aide et au soutien de toutes ces personnes que nous avons pu réaliser ce projet. Nous leur sommes infiniment reconnaissants et nous leur dédions ce travail avec fierté.

## **Abstract**

First and foremost, we would like to express our sincere gratitude to God Almighty for granting us the health, strength, and perseverance necessary to accomplish this project.

Ms. **Chaymae Chouiekh**, our project supervisor, deserves our heartfelt thanks for her invaluable guidance, patience, rigor, and availability throughout this project. Her enlightened guidance and wise counsel played an essential role in the success of this task.

We would also like to express our gratitude to all the teachers for their dedication, generosity, and patience despite their many responsibilities. Their invaluable teachings and constant support played an essential role in our intellectual development and progress throughout our academic journey.

It is thanks to the help and support of all these individuals that we were able to accomplish this project. We are immensely grateful to them, and we dedicate this work to them with pride.

# Introduction Générale

Dans l'ère numérique actuelle, la quantité astronomique d'informations et de contenus disponibles en ligne représente à la fois une opportunité et un défi. Bien que cette abondance offre un vaste choix aux utilisateurs, elle peut également les submerger et rendre la découverte de contenus pertinents et adaptés à leurs préférences individuelles particulièrement ardue. C'est dans ce contexte que les systèmes de recommandation sont devenus des outils indispensables, guidant les utilisateurs vers du contenu susceptible de les intéresser et de répondre à leurs attentes.

Les systèmes de recommandation traditionnels analysent les données des utilisateurs, leurs interactions avec un système (site web, application, etc.) et leurs préférences pour générer des suggestions personnalisées. Cependant, ces systèmes présentent certaines limites, notamment en termes de précision, de diversité des recommandations et d'adaptation aux préférences dynamiques des utilisateurs. Face à ces défis, le concept de "système de recommandation amélioré" a émergé, promettant de révolutionner l'expérience utilisateur grâce à des recommandations plus précises, diversifiées et adaptées.

Les systèmes de recommandation améliorés exploitent des techniques d'apprentissage automatique avancées, telles que les réseaux de neurones profonds, les modèles de factorisation matricielle et l'apprentissage par renforcement, pour améliorer la pertinence des recommandations. Ils intègrent également des facteurs contextuels complexes, comme l'humeur, la situation géographique, les conditions météorologiques et les événements en temps réel, afin de fournir des suggestions adaptées au contexte spécifique de l'utilisateur.

L'un des principaux avantages de ces systèmes réside dans leur capacité à s'adapter aux préférences évolutives des utilisateurs. Grâce à des algorithmes d'apprentissage continu, ils peuvent détecter les changements de goûts et de comportements, ajustant ainsi les recommandations en conséquence pour éviter les suggestions obsolètes ou inadaptées. De plus, en intégrant des techniques de diversification, ces systèmes encouragent les utilisateurs à explorer de nouveaux contenus, évitant ainsi le problème de la "bulle de filtres" qui limite la découverte.

Au-delà de l'amélioration de l'expérience utilisateur, les systèmes de recommandation améliorés représentent un atout stratégique majeur pour les entreprises. En proposant des recommandations pertinentes et adaptées, elles peuvent fidéliser leur clientèle, encourager les achats répétés et augmenter leur part de marché dans un environnement hautement compétitif. Ces systèmes offrent un avantage concurrentiel indéniable.

Avec la montée en puissance des systèmes de recommandation améliorés dans divers secteurs de la société, il devient de plus en plus crucial de garantir qu'ils opèrent de manière éthique et transparente. Les questions de biais, de confidentialité des données et de manipulation des utilisateurs doivent être prises en compte. En mettant l'accent sur ces aspects, les développeurs et les entreprises peuvent non seulement renforcer la confiance des utilisateurs dans leurs systèmes, mais aussi favoriser un environnement numérique plus équitable et inclusif pour tous.

Ce rapport de projet représente une exploration approfondie et exhaustive des systèmes de recommandation améliorés, une technologie à la pointe de l'innovation qui redéfinit les standards de l'expérience utilisateur numérique. À travers une analyse minutieuse, il décortique les concepts clés, les principes fondamentaux, les avantages révolutionnaires et les défis complexes inhérents à ces systèmes d'avant-garde. S'appuyant sur une expertise de pointe, ce rapport offre un regard pénétrant sur les techniques d'apprentissage automatique de pointe exploitées, telles que les réseaux de neurones profonds et l'apprentissage par renforcement, ainsi que sur l'intégration stratégique de facteurs contextuels cruciaux pour garantir des recommandations véritablement personnalisées et pertinentes. De plus, il présente une dissection rigoureuse des méthodes d'évaluation les plus avancées, permettant de mesurer avec précision les performances et l'efficacité de ces systèmes novateurs. Allant au-delà d'une simple analyse théorique, ce rapport adopte une approche pragmatique en examinant les implications concrètes et les applications pratiques des systèmes de recommandation améliorés dans divers secteurs clés, tels que le commerce électronique, les médias numériques et les services de streaming. À travers des études de cas approfondies et des exemples concrets, il met en lumière les meilleures pratiques, les stratégies gagnantes et les tendances émergentes qui façonnent l'avenir de cette technologie révolutionnaire. En somme, ce rapport se positionne comme une ressource incontournable, offrant aux professionnels, aux chercheurs et aux entreprises avant-gardistes un aperçu exhaustif des perspectives transformatrices et du potentiel inexploité des systèmes de recommandation améliorés.

# Chapitre 1

## Etat d'art

L'explosion de la quantité de données au cours des dernières années a rendu le choix d'informations pertinentes une tâche complexe pour les internautes. Pour répondre à ces défis, de nouvelles techniques ont émergé, connues sous le nom de Systèmes de Recommandation.

### 1.1 Introduction

Les systèmes de recommandation sont devenus essentiels dans notre quotidien, facilitant ainsi nos achats, la découverte de nouveaux contenus et la recherche de produits adaptés à nos préférences. En effet, ces systèmes suggèrent à l'utilisateur des choix ou des produits qu'il est susceptible d'aimer en se basant sur les préférences d'autres utilisateurs, l'historique de l'utilisateur, et diverses autres informations permettant de corrélérer différents produits ou services.

Nés dans les années 90 avec l'essor du commerce électronique, les systèmes de recommandation ont évolué pour relever le défi de sélectionner des informations pertinentes au sein d'une quantité croissante de données. Essentiellement, le problème consiste à estimer la note qu'un utilisateur donnerait à un produit (livre, film, musique, restaurant, etc.) qu'il ne connaît pas, en se basant sur ses propres notes antérieures ainsi que celles d'autres utilisateurs pour des produits similaires. Par conséquent, le système recommande le produit qui a le plus de chances de plaire à l'utilisateur ou qui maximise une certaine fonction de profit pour le vendeur.

Selon les notations de Adomavicius et Tuzhilin,  $C$  désigne l'ensemble des utilisateurs et  $S$  l'ensemble des items pouvant être recommandés. En outre, la fonction  $u$  de  $C \times S$  dans  $R$  mesure l'utilité de l'item  $s$  pour l'utilisateur  $c$ , cette utilité étant souvent représentée par une note, comme un nombre d'étoiles de 1 à 5 pour un livre.

L'émergence du Web 2.0 a marqué un tournant dans l'histoire des systèmes de recommandation. Grâce à l'abondance de données générées par les utilisateurs et à l'essor des plateformes sociales, il est désormais possible de développer des approches plus sophistiquées basées sur le filtrage collaboratif.

De plus, l'avènement du machine learning a révolutionné le domaine des systèmes de recommandation. En effet, des algorithmes puissants comme le K-Nearest Neighbors (KNN), la factorisation matricielle et les réseaux de neurones profonds ont permis de traiter des volumes massifs de données et d'identifier des patterns complexes avec une précision inégalée.

Ainsi, ces avancées ont donné naissance à des systèmes de recommandation plus performants et personnalisables, capables de prendre en compte des facteurs contextuels tels que la localisation, l'heure de la journée ou l'humeur de l'utilisateur pour proposer des recommandations en temps réel et ultra-pertinentes.

Aujourd’hui, les chercheurs s’efforcent de développer des systèmes de recommandation plus explicables, transparents et éthiques. En effet, il s’agit de comprendre les raisons pour lesquelles un système propose une recommandation particulière et d’éviter les biais discriminatoires provenant des données d’apprentissage.

Par ailleurs, l’exploration de systèmes hybrides combinant différentes approches, comme le filtrage collaboratif et le machine learning, est également un domaine de recherche actif. L’objectif est de tirer parti des avantages de chaque méthode afin de créer des systèmes de recommandation encore plus performants et adaptables.

### 1.1.1 Domaines d'utilisation

Les systèmes de recommandation sont utilisés dans une variété de domaines pour améliorer l’expérience utilisateur et augmenter l’engagement. Dans le commerce électronique, ils suggèrent des produits pertinents, augmentant ainsi les ventes et la satisfaction client.

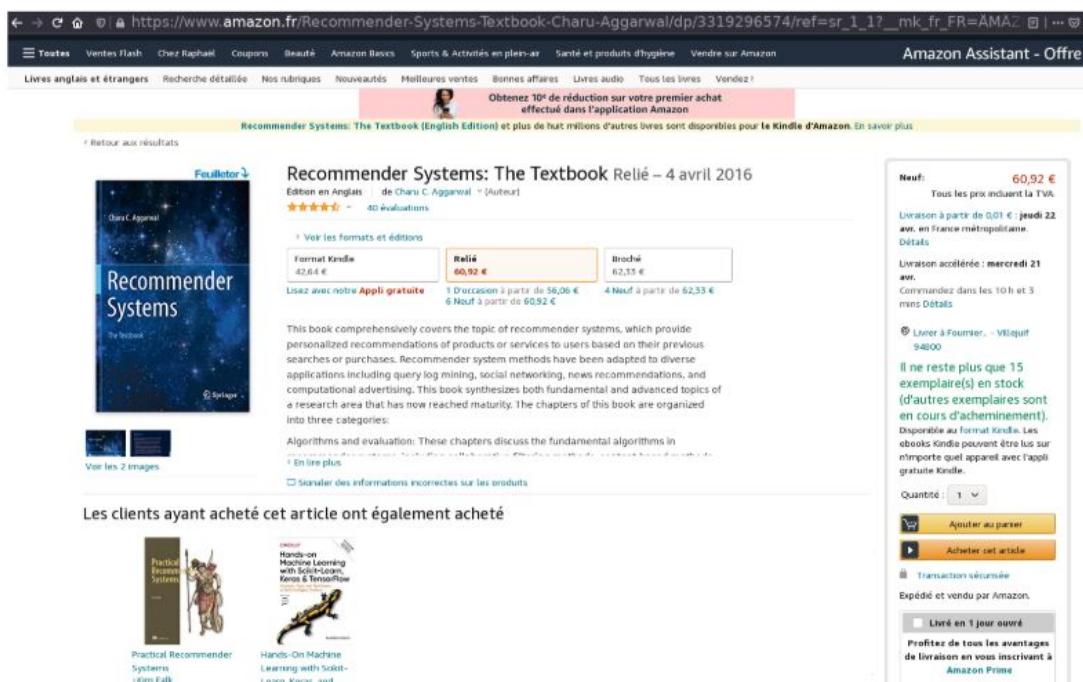


FIGURE 1.1 – Recommendation livres Amazon.

Les services de streaming, tels que Netflix et Spotify, les utilisent pour recommander des films, des séries et de la musique en fonction des préférences des utilisateurs. Sur les réseaux sociaux, ces systèmes aident à connecter les utilisateurs avec des amis potentiels et à proposer du contenu pertinent. De plus, dans les moteurs de recherche, où ils aident à suggérer des articles, des livres et des cours adaptés aux intérêts et aux besoins des utilisateurs.

## 1.2 Algorithmes existants

### 1.2.1 Filtrage collaboratif

Le filtrage collaboratif est l’une des méthodes les plus efficaces pour recommander des contenus à des utilisateurs déjà connus. Il repose sur le principe suivant : si deux personnes ont

aimé des contenus similaires dans le passé, elles ont de fortes chances d'apprécier les mêmes contenus à l'avenir.

Les recommandations personnalisées issues du filtrage collaboratif peuvent être générées de différentes manières : en se basant sur le profil des utilisateurs (**User-based**), en utilisant les profils des contenus (**Item-based**). Examinons comment ces deux algorithmes de filtrage collaboratif fonctionnent.

**USER-BASED :** Le principe sur lequel repose le filtrage collaboratif est le suivant : si des lecteurs ont eux des comportements et goûts de lecture similaires par le passé, alors, ils devraient également en avoir dans le futur. L'avantage du filtrage collaboratif est qu'il est agnostique au contenu. Il se base uniquement sur le profil de navigation des lecteurs pour faire des recommandations personnalisées.

**ITEM-BASED :** Le filtrage collaboratif basé sur les articles est une méthode de recommandation qui recherche des articles similaires en fonction des articles que les utilisateurs ont déjà aimés ou avec lesquels ils ont eu des interactions positives. Développée par Amazon en 1998, cette méthode a joué un rôle majeur dans le succès d'Amazon..

#### Item-Based vs User-Based, quelles différences ?

Les deux algorithmes ont un fonctionnement globalement similaire. Leurs différences résident dans le fait que l'Item-based et le User-based ont un point de départ inversé pour le calcul des recommandations personnalisées. L'un commence par regarder les profils utilisateurs (User-based), l'autre débute le calcul en utilisant les contenus (Item-based).

### 1.2.2 Filtrage basé sur le contenu

La recommandation basée sur le contenu est une méthode utilisée pour suggérer des articles, des films, des livres, ou tout autre type de contenu à un utilisateur en se basant sur les caractéristiques des contenus que cet utilisateur a déjà appréciés. Contrairement au filtrage collaboratif qui se base sur les comportements des autres utilisateurs, la recommandation basée sur le contenu se concentre sur les attributs et les descriptions des éléments eux-mêmes.

#### Comment ça fonctionne ?

**1- Profil Utilisateur :** Tout d'abord, un profil utilisateur est créé. Ce profil est une représentation des préférences de l'utilisateur, souvent basé sur les articles qu'il a consultés, aimés, ou notés positivement dans le passé. Par exemple, si un utilisateur aime des articles de science-fiction, son profil reflétera cette préférence.

**2- Caractéristiques des Contenus :** Chaque article ou élément de contenu est décrit par un ensemble de caractéristiques ou de "tags". Par exemple, un film pourrait être décrit par son genre (science-fiction, drame), ses acteurs, son réalisateur, son année de sortie, etc.

**3- Correspondance Profil-Contenu :** L'algorithme de recommandation compare les caractéristiques des nouveaux contenus disponibles avec le profil de l'utilisateur. Il calcule la similarité entre les descriptions des nouveaux articles et les préférences de l'utilisateur. Les articles qui ont des caractéristiques similaires à celles que l'utilisateur a appréciées dans le passé sont alors recommandés.

### 1.2.3 Recommandation hybride

Les deux modèles de filtrage présentent des limitations car le modèle de filtrage basé sur le contenu repose sur les métadonnées concernant les articles de l'utilisateur, tandis que le filtrage collaboratif s'appuie sur les données de notation des articles par l'utilisateur. Un modèle de recommandation hybride a été proposé pour résoudre les limitations des deux modèles de filtrage de recommandation et améliorer les performances de recommandation.

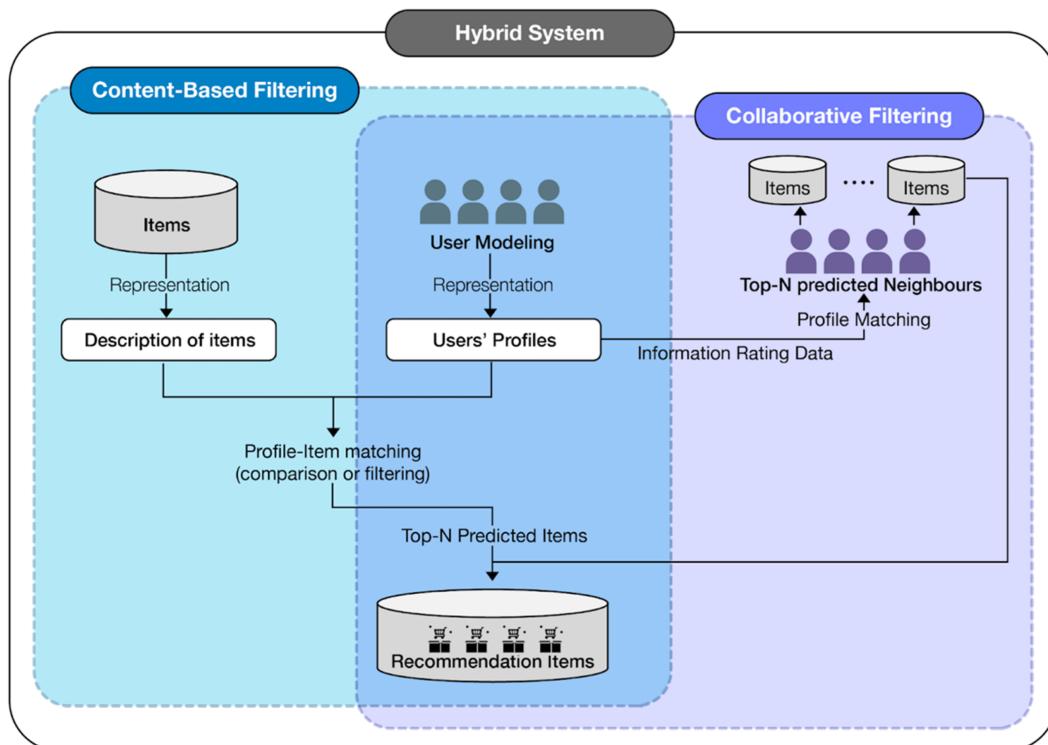


FIGURE 1.2 – Recommandation hybride.

### 1.2.4 Problèmes et limites des systèmes de recommandation

Les systèmes de recommandation se basant sur les techniques précédemment expliquées ont certaines limites. Plusieurs approches ont été proposées dans la littérature pour pallier ces limites. Ci-dessous les problèmes les plus importants sont décrits.

**Démarrage à froid :** Les systèmes de filtrage collaboratif dépendent des évaluations des items par les utilisateurs. Ainsi, un nouvel item ne peut pas être recommandé tant qu'aucun utilisateur ne l'a évalué. Dans les systèmes de recommandation basés sur le filtrage collaboratif et les systèmes basés sur le contenu, il est impossible de prédire les préférences des utilisateurs sans connaître leurs historiques d'évaluations d'items. Ainsi, les nouveaux utilisateurs ne recevront pas de recommandations précises avant d'avoir évalué un certain nombre d'items.

**Sparsity :** Un système de recommandation souffre de la sparsity quand le nombre d'items évalués par les utilisateurs est très faible par rapport au nombre d'items total présent dans le système. Ce fait conduit à avoir une très faible densité dans la matrice d'évaluation utilisée.

teurs/items. Cela a des conséquences sur la capacité du système de recommandation à recommander toutes les items disponibles et sur l'exactitude des recommandations générées.

**Problème du mouton gris :** Les utilisateurs d'un système de recommandation peuvent avoir des goûts particuliers et des préférences très inhabituelles par rapport aux autres. Ces utilisateurs sont à la frontière entre deux ou plusieurs clusters d'utilisateurs. Il leur est donc difficile de trouver des utilisateurs similaires et des recommandations pertinents.

**Sérendipité :** Vu que les systèmes de recommandation basés sur le contenu ne recommandent que les items correspondants au profil de l'utilisateur, ce dernier ne recevra que des recommandations similaires à celles qu'il a déjà rencontrées. Il n'aura aucune chance de recevoir des recommandations inattendues. Cela peut amener l'utilisateur à se lasser des recommandations..

## 1.2.5 Techniques Avancées et Nouvelles Approches

### 1.2.5.1 Deep Learning

Les systèmes de recommandation évoluent constamment, et de nouvelles techniques et approches sont développées pour améliorer leur précision, leur scalabilité et leur capacité à s'adapter aux besoins des utilisateurs. Parmi ces techniques, on trouve :

- **Auto-encodeurs** : Utilisés pour réduire la dimensionnalité des données et extraire des caractéristiques essentielles, les auto-encodeurs peuvent aider à identifier des patterns complexes dans les préférences des utilisateurs.
- **Réseaux de neurones convolutifs (CNN)** : Bien que généralement utilisés pour le traitement des images, les CNN peuvent aussi être appliqués aux systèmes de recommandation pour capter des caractéristiques locales dans les données séquentielles ou temporelles.
- **Réseaux de neurones récurrents (RNN)** : Particulièrement efficaces pour traiter des données séquentielles, les RNN (y compris les variantes comme LSTM et GRU) permettent de modéliser les séquences d'interactions des utilisateurs pour des recommandations plus précises.
- **Réseaux de neurones de graphes (GNN)** : Ces réseaux exploitent la structure des graphes pour capturer des relations complexes entre les utilisateurs et les items, ce qui permet de fournir des recommandations plus fines et contextualisées.

### 1.2.5.2 Approches contextuelles

Les recommandations contextuelles prennent en compte des informations supplémentaires sur le contexte de l'utilisateur, comme l'heure, le lieu, l'appareil utilisé, ou l'état émotionnel. Par exemple :

- **L'heure** : Proposer des restaurants pour le déjeuner à midi ou des films relaxants en soirée.
- **Le lieu** : Recommander des événements ou des lieux à proximité de l'utilisateur.
- **L'appareil utilisé** : Suggérer des articles courts sur un smartphone et des contenus plus longs sur une tablette ou un ordinateur.
- **L'état émotionnel** : Adapter les recommandations de musique ou de films en fonction de l'humeur de l'utilisateur (détectée via des capteurs ou des interactions précédentes).

Ces informations contextuelles permettent de fournir des recommandations plus pertinentes et personnalisées, augmentant ainsi la satisfaction des utilisateurs.

### 1.2.5.3 Apprentissage par renforcement

L'apprentissage par renforcement (Reinforcement Learning) permet aux systèmes de recommandation de s'adapter dynamiquement aux interactions en temps réel des utilisateurs. Voici comment cela fonctionne :

- **Agent et Environnement** : Le système de recommandation agit comme un agent qui prend des actions (fournir des recommandations) dans un environnement (les utilisateurs et leurs interactions).
- **Récompenses** : L'agent reçoit des récompenses en fonction de la qualité des recommandations (par exemple, clics, achats, temps passé sur une page).
- **Optimisation** : Au fil du temps, l'agent apprend à optimiser ses recommandations pour maximiser les récompenses cumulées, s'adaptant ainsi aux préférences changeantes des utilisateurs.

Cette approche permet de créer des systèmes de recommandation plus réactifs et adaptatifs, capables de s'améliorer continuellement à mesure qu'ils interagissent avec les utilisateurs.

### 1.2.6 Évaluation des Systèmes de Recommandation

L'évaluation des systèmes de recommandation est essentielle pour mesurer leur efficacité et leur pertinence. Elle repose sur diverses métriques, telles que la précision, le rappel, le F1-Score, et des métriques de classement comme NDCG, MAP et MRR, qui évaluent la qualité et le positionnement des recommandations. Les métriques d'erreur comme RMSE et MAE sont utilisées pour évaluer la précision des prédictions de notes. Les évaluations peuvent être menées hors ligne, via des techniques comme la validation croisée, ou en ligne, avec des tests A/B pour mesurer l'impact en temps réel sur les utilisateurs. Une évaluation rigoureuse permet d'améliorer continuellement les algorithmes de recommandation, garantissant ainsi une meilleure expérience utilisateur.

## 1.3 Conclusion

Les systèmes de recommandation sont essentiels dans notre quotidien numérique, facilitant la découverte de produits et de contenus. Depuis leur émergence dans les années 90, ils ont évolué grâce à des techniques comme le filtrage collaboratif, le filtrage basé sur le contenu et les approches hybrides. L'intégration du deep learning et des approches contextuelles a permis de développer des modèles plus précis et personnalisés.

Malgré des défis tels que le démarrage à froid et la sparsity, des solutions innovantes comme l'apprentissage par renforcement ont amélioré l'adaptabilité et l'efficacité des recommandations. Une évaluation rigoureuse à l'aide de diverses métriques est cruciale pour mesurer et améliorer ces systèmes.

À l'avenir, les systèmes de recommandation continueront d'évoluer avec de nouvelles technologies et approches, se concentrant sur la transparence, l'explicabilité et l'éthique, pour offrir une expérience utilisateur enrichie et satisfaisante.

# **Chapitre 2**

## **Étude préliminaire**

### **2.1 Introduction**

Dans le paysage contemporain où l'abondance de livres disponibles peut s'avérer à la fois une richesse et un défi, notre projet vise à optimiser les systèmes de recommandation sur la plateforme Amazon. Ce chapitre explore les fondements théoriques et les méthodologies employées dans cette quête d'amélioration, essentielle pour guider les lecteurs vers les ouvrages les plus en phase avec leurs préférences et leurs centres d'intérêt.

Nous commencerons par une présentation des techniques de recommandation traditionnelles, à savoir le filtrage collaboratif et le filtrage basé sur le contenu, qui constituent les pierres angulaires de ces technologies. Ces approches classiques seront examinées en profondeur, mettant en lumière leur fonctionnement, leurs avantages et leurs limites inhérentes.

Ensuite, nous porterons notre attention sur les algorithmes avancés intégrés dans notre projet novateur. Parmi ces techniques d'avant-garde figurent les k-Nearest Neighbors (k-NN), la décomposition en valeurs singulières (SVD), la factorisation matricielle non négative (NMF) et les autoencodeurs. Chacune de ces méthodes sera disséquée minutieusement, révélant ses mécanismes intrinsèques, ses forces et ses faiblesses, ainsi que son implémentation spécifique dans le contexte de la recommandation de livres.

L'objectif primordial de ce chapitre est de fournir une compréhension approfondie des diverses approches et technologies de recommandation, jetant ainsi des bases solides pour les développements pratiques et les expérimentations que nous présenterons ultérieurement. En explorant ces études préliminaires, nous mettons en lumière les innovations qui permettent de proposer des recommandations de livres toujours plus précises et personnalisées, enrichissant ainsi l'expérience utilisateur sur Amazon de manière significative.

### **2.2 Algorithmes de Recommandation classiques**

Les systèmes de recommandation modernes reposent sur des algorithmes sophistiqués, mais les fondations ont été bâties sur des techniques plus classiques. Ces méthodes de base constituent encore aujourd'hui des piliers importants dans de nombreux systèmes de recommandation, souvent combinées à des approches plus avancées.

## 2.2.1 Filtrage Collaboratif

Le filtrage collaboratif est une méthode de recommandation populaire qui exploite les interactions passées des utilisateurs avec les éléments pour générer de nouvelles recommandations. Il existe deux variantes principales de cette approche :

### 2.2.1.1 Filtrage collaboratif utilisateur

Cette technique recommande des éléments à un utilisateur cible en identifiant d'autres utilisateurs ayant des préférences similaires, sur la base de leurs interactions passées avec les éléments. Les éléments appréciés par ces utilisateurs "voisins" sont ensuite recommandés à l'utilisateur cible.

### 2.2.1.2 Filtrage Collaboratif Élément

À l'inverse, cette méthode recommande à l'utilisateur des éléments similaires à ceux qu'il a appréciés dans le passé. La similarité entre les éléments est calculée en fonction des préférences des utilisateurs qui les ont déjà consultés ou notés.

Le filtrage collaboratif offre l'avantage de ne pas nécessiter d'informations détaillées sur les caractéristiques des éléments, mais repose uniquement sur les interactions utilisateur-élément. Cependant, cette approche peut souffrir du problème du "démarrage à froid" lorsque peu d'interactions sont disponibles pour un nouvel utilisateur ou un nouvel élément.

## 2.2.2 Filtrage Basé sur le Contenu

Le filtrage basé sur le contenu est une autre technique fondamentale de recommandation. Elle recommande des éléments en se basant sur les caractéristiques des éléments et sur les préférences passées de l'utilisateur.

Par exemple, pour recommander des livres, on peut utiliser des caractéristiques telles que le genre, l'auteur, les mots-clés du résumé, ou d'autres métadonnées pertinentes. Un profil de préférences de l'utilisateur est construit à partir des éléments qu'il a appréciés dans le passé, et de nouveaux éléments similaires à ce profil lui sont recommandés.

Le filtrage basé sur le contenu présente l'avantage de pouvoir recommander des éléments nouveaux ou impopulaires, tant qu'ils correspondent aux préférences de l'utilisateur. Cependant, il nécessite une représentation riche des caractéristiques des éléments, ce qui peut être difficile à obtenir pour certains types de données.

Ces approches classiques ont jeté les bases des systèmes de recommandation modernes, mais présentent des limitations inhérentes. C'est pourquoi nous avons intégré des algorithmes plus avancés, capables de capturer des relations complexes et de fournir des recommandations plus précises et personnalisées.

## 2.3 Algorithmes de Recommandation Avancés

Dans cette section, nous explorons en profondeur les algorithmes d'avant-garde intégrés à notre système de recommandation novateur pour la plateforme Amazon. Ces techniques sophistiquées permettent de capturer les relations complexes entre les utilisateurs et les éléments, offrant ainsi des recommandations de livres plus précises et personnalisées.

### 2.3.1 K-Nearest Neighbors (k-NN)

L'algorithme des k plus proches voisins (**k-NN!**) est une technique non paramétrique puissante pour le filtrage collaboratif. Il calcule la similarité entre les utilisateurs ou les éléments en utilisant des mesures de distance appropriées, telles que la distance euclidienne ou la distance cosinus. Les recommandations sont ensuite générées en se basant sur les préférences des k voisins les plus proches de l'utilisateur cible ou de l'élément cible.

Nous avons optimisé l'implémentation de k-NN en explorant différentes techniques de pondération, de normalisation et de sélection des voisins. Cette approche garantit une meilleure précision et une plus grande robustesse face aux données manquantes ou bruitées.

### 2.3.2 Décomposition en valeurs singulières (SVD)

La décomposition en valeurs singulières (SVD) est une méthode de factorisation matricielle puissante qui décompose la matrice utilisateur-élément en trois matrices distinctes. Cette décomposition permet de capturer les relations latentes entre utilisateurs et éléments, améliorant considérablement la précision des recommandations.

Nous avons exploré diverses variantes de SVD, notamment la SVD tronquée et la SVD régularisée, afin d'optimiser les performances et de réduire le sur-apprentissage.

### 2.3.3 Factorisation matricielle non négative (NMF)

La factorisation matricielle non négative (NMF) est une extension de SVD avec la contrainte supplémentaire que les valeurs des matrices résultantes doivent être non négatives. Cette méthode est particulièrement adaptée aux données de recommandation, où les relations entre utilisateurs et éléments sont intrinsèquement positives.

Nous avons implémenté différentes variantes de NMF, notamment la NMF alternée et la NMF régularisée, afin d'améliorer la robustesse et la stabilité de l'algorithme face aux données bruitées ou manquantes.

## 2.4 Sources de données

La qualité et la pertinence des données sont cruciales pour le succès d'un système de recommandation. Dans le cadre de notre projet visant à améliorer les recommandations de livres sur Amazon, nous avons exploité trois ensembles de données riches et complémentaires, obtenus à partir de la plateforme Kaggle. Cette section détaille la nature de ces données et les étapes de prétraitement nécessaires pour les rendre exploitables par nos algorithmes.

### 2.4.1 Dataset des livres : booksdf.csv

Le premier ensemble de données rassemble des informations détaillées sur les livres disponibles sur Amazon. Chaque entrée comprend des métadonnées telles que le titre, l'auteur, l'éditeur, la date de publication. Ces caractéristiques permettent de décrire chaque livre de manière précise et de construire des profils de contenu riches pour les recommandations basées sur le contenu.

### 2.4.2 Dataset des utilisateurs : users sdf.csv

Le deuxième ensemble de données contient des informations sur les utilisateurs de la plateforme. Les données incluent des identifiants uniques d'utilisateurs, leur localisation géographique et d'autres informations démographiques pertinentes. Bien que ces données soient limitées en comparaison avec d'autres sources potentielles, elles fournissent une base pour le filtrage collaboratif basé sur les utilisateurs.

### 2.4.3 Dataset des interactions : ratings sdf.csv

Le troisième ensemble de données est essentiel pour le filtrage collaboratif, car il répertorie les interactions entre les utilisateurs et les livres sous forme de notes attribuées. Chaque entrée contient l'identifiant de l'utilisateur, l'identifiant du livre, et la note attribuée par l'utilisateur et autre . Ces interactions servent de fondement pour entraîner et évaluer nos modèles de recommandation.

### 2.4.4 Prétraitement des données

Avant de pouvoir utiliser ces données pour entraîner nos algorithmes de recommandation, nous avons effectué un prétraitement minutieux pour assurer leur qualité et leur cohérence.

Les étapes de prétraitement incluent :

- **Nettoyage des données** : Suppression des entrées dupliquées, traitement des valeurs manquantes, et gestion des valeurs aberrantes.
- **Normalisation** : Mise à l'échelle des caractéristiques des livres et des notes pour garantir une cohérence dans les calculs.
- **Encodage des caractéristiques** : Conversion des caractéristiques catégorielles en représentations numériques pour une utilisation dans les algorithmes.
- **Division des données** : Séparation des ensembles de données en ensembles d'entraînement et de test pour une évaluation rigoureuse des modèles.

Ces étapes de prétraitement sont cruciales pour garantir que les données soient prêtes pour l'entraînement des modèles de recommandation, maximisant ainsi leur efficacité et leur précision.

## 2.5 Outils utilisés

Pour mener à bien notre projet, nous avons utilisé un ensemble d'outils et de technologies modernes qui facilitent le traitement des données, la construction des modèles et l'évaluation des performances. Cette section décrit les principaux outils utilisés et leur rôle dans le projet.

### 2.5.1 Visualisation des données

La visualisation des données est une étape cruciale dans le développement de systèmes de recommandation, car elle permet de mieux comprendre les données et les résultats des modèles. Nous avons utilisé plusieurs bibliothèques de visualisation pour explorer et présenter nos données :

- **Matplotlib** : Pour créer des graphiques statiques, simples et personnalisables.

- **Seaborn** : Pour des visualisations statistiques attractives et informatives, facilitant la compréhension des distributions et des relations entre les variables.
- **Plotly** : Pour des graphiques interactifs qui permettent une exploration plus dynamique des données.

## 2.5.2 Python et Jupyter Notebook

Python est le langage de programmation principal utilisé pour ce projet en raison de sa flexibilité et de sa richesse en bibliothèques spécialisées pour l'apprentissage automatique et le traitement des données. Jupyter Notebook est utilisé comme environnement de développement interactif, permettant de documenter, tester et visualiser le code de manière efficace.

## 2.5.3 Bibliothèques de Machine Learning

Nous avons exploité plusieurs bibliothèques Python dédiées à l'apprentissage automatique et à l'analyse de données :

- **Pandas, Numpy, Scipy** : Pour la manipulation et l'analyse des données tabulaires.
- **Scikit-learn** : Pour la mise en œuvre d'algorithmes de recommandation classiques et avancés.
- **Surprise** : Une bibliothèque spécialisée pour les systèmes de recommandation, utilisée pour les algorithmes de filtrage collaboratif et de factorisation matricielle.

## 2.5.4 Environnement de Développement et d'Exécution

Le projet a été développé et exécuté dans un environnement local ainsi que sur des plateformes de cloud computing pour bénéficier de ressources de calcul supplémentaires lorsque nécessaire.

## 2.6 Conclusion

En synthèse, ce chapitre a fourni une vue d'ensemble complète des techniques de recommandation traditionnelles et avancées utilisées, des sources de données exploitées et des outils utilisés pour le développement de notre système de recommandation de livres sur Amazon. Cette étude préliminaire a posé les bases théoriques et pratiques nécessaires pour aborder les phases de réalisation et d'évaluation que nous explorerons dans les chapitres suivants. En adoptant une approche méthodique et rigoureuse, nous visons à améliorer la précision et la personnalisation des recommandations, offrant ainsi une expérience utilisateur enrichie et sur mesure.

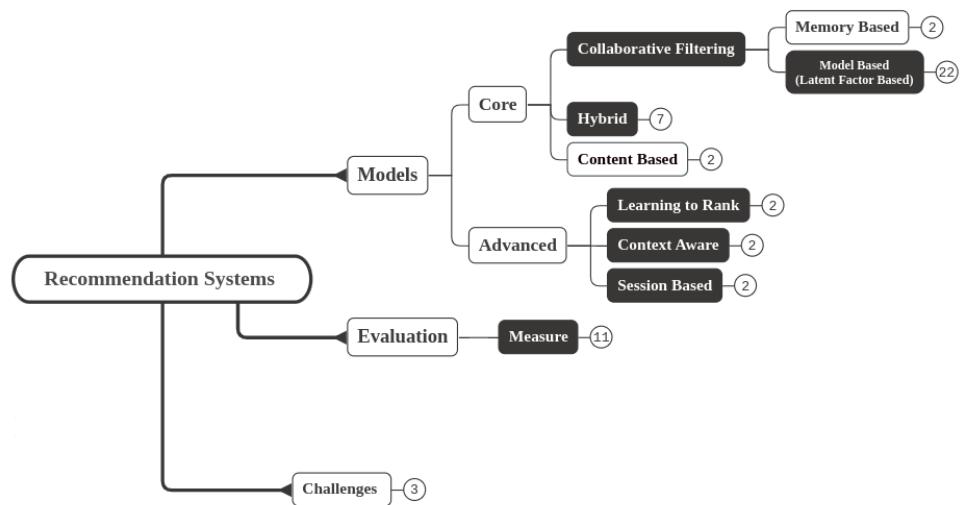


FIGURE 2.1 – les différents types de Systèmes de Recommandation

# Chapitre 3

## Réalisation du projet

### 3.1 Introduction

Dans ce dernier chapitre, nous abordons en détail la mise en oeuvre concrète de notre système de recommandation de livres, en mettant l'accent sur les différentes techniques et algorithmes employés pour offrir des suggestions personnalisées aux utilisateurs. Notre projet s'appuie sur un vaste ensemble de données de notation de livres provenant de la plateforme Amazon, que nous avons minutieusement nettoyé et préparé en vue de l'analyse.

Tout au long de ce chapitre, nous explorons un éventail d'approches de recommandation, allant des méthodes classiques telles que le filtrage collaboratif et le filtrage basé sur le contenu, jusqu'aux techniques plus avancées comme les algorithmes des(K-NN), la décomposition en valeurs singulières (SVD) et la factorisation non négative de matrices (NMF).

Pour chacune de ces approches, nous fournissons une description approfondie de l'algorithme, accompagnée d'une présentation détaillée de son implémentation pratique, illustrée par des extraits de code pertinents. Nous analyserons ensuite les résultats obtenus pour chaque méthode, en mettant en évidence leurs forces et leurs limites dans le contexte de notre projet.

L'objectif principal de ce chapitre est de démontrer l'efficacité et la pertinence des différentes techniques de recommandation mises en oeuvre, tout en offrant une évaluation comparative rigoureuse de leurs performances respectives. Cette analyse approfondie nous permettre de tirer des conclusions éclairées quant à l'approche la plus adaptée pour notre système de recommandation de livres, en tenant compte des spécificités de notre jeu de données et des attentes de nos utilisateurs.

En somme, ce chapitre final constitue une synthèse des aspects théoriques et pratiques de notre projet, mettant en lumière les défis rencontrés ,les solutions apportées et les résultats obtenus. Il témoigne de notre démarche rigoureuse et de notre engagement à fournir un système de recommandation performant et innovant, capable de satisfaire les besoins de lecteurs en quête de nouvelles découvertes littéraires.

### 3.2 Collecte et prétraitement des données

#### 3.2.1 Collecte des Données

Nous avons obtenu notre jeu de données principal à partir de la plateforme Kaggle, qui propose trois ensembles de données distincts : *usersdf.csv*, *booksdf.csv* et *ratingsdf.csv*. Ces

fichiers contiennent respectivement des informations sur les utilisateurs, les livres et les notes attribuées par les utilisateurs aux livres.

### 3.2.2 Nettoyage et Prétraitement des Données

Après avoir collecté les données, nous avons procédé à une série d'opérations de nettoyage et de prétraitement pour garantir leur qualité et leur exploitabilité. Ces étapes comprenaient notamment :

- La suppression des doublons et la gestion des valeurs manquantes.
- La normalisation des notes pour assurer une comparabilité entre les utilisateurs.
- Le filtrage des utilisateurs et des livres ayant un nombre insuffisant d'interactions.
- L'identification et le traitement des valeurs aberrantes.
- L'identification et la suppression des colonnes ou lignes redondantes.
- L'encodage des variables catégorielles.
- La réduction de la dimensionnalité.
- La transformation logarithmique et la normalisation.

### 3.2.3 Structuration des Données

Enfin, nous avons organisé les données sous forme de matrices pour faciliter leur utilisation par les différents algorithmes de recommandation. Nous avons construit une matrice de notation, où chaque ligne représente un utilisateur et chaque colonne un livre, avec les entrées représentant les notes attribuées.

Cette structuration des données nous a permis de mettre en œuvre efficacement différentes techniques de recommandation, telles que le filtrage collaboratif, le filtrage basé sur le contenu et les approches basées sur la factorisation matricielle.

Avec ce jeu de données de qualité, nous avons pu procéder à la mise en œuvre de notre système de recommandation de livres, dont nous détaillons les différentes approches dans les sections suivantes.

## 3.3 Exploration et Analyse des Données

Avant de plonger dans la mise en œuvre des algorithmes de recommandation, il est crucial de procéder à une exploration et une analyse approfondies des données. Cette étape nous permet de mieux comprendre la structure et les caractéristiques de notre jeu de données, d'identifier d'éventuels biais ou anomalies, et de dégager des insights précieux pour orienter notre stratégie de recommandation.

### 3.3.1 Statistiques Descriptives

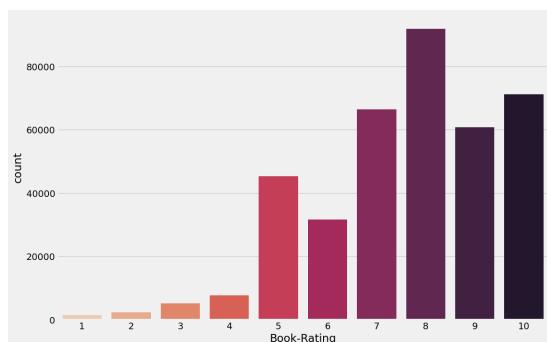
Nous avons commencé par examiner les statistiques descriptives de nos 3 ensembles de données : usersdf.csv, booksdf.csv et ratingsdf.csv. Cette analyse nous a permis de mieux apprendre la distribution des données, la présence de valeurs aberrantes et les éventuelles disparités entre les différentes catégories d'utilisateurs ou de livres.

Pour les utilisateurs, nous avons étudié des indicateurs tels que l'âge, le sexe, la localisation géographique et le nombre de notes attribuées. Concernant les livres, nous avons analysé la répartition par genre, la distribution des années de publication, ainsi que la moyenne et la

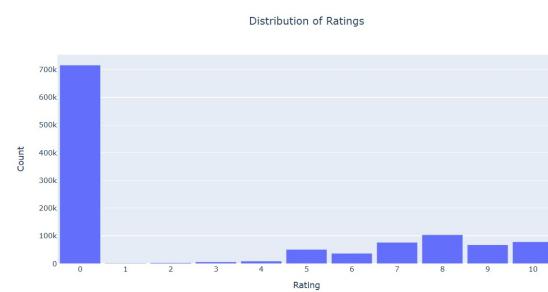
dispersion des notes reçues. Enfin, pour les notes, nous avons examiné leur distribution globale, leur évolution dans le temps et les éventuelles corrélations avec les caractéristiques des utilisateurs et des livres.

### 3.3.2 Visualisation des Données

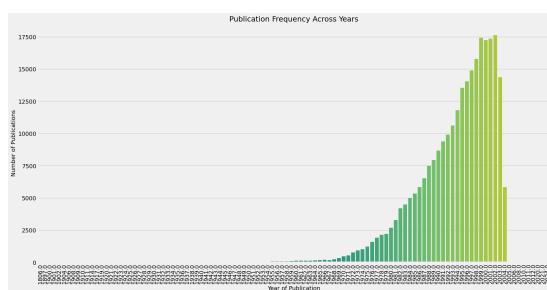
Afin de faciliter l'interprétation des résultats de notre analyse exploratoire, nous avons eu recours à diverses techniques de visualisation des données. Des graphiques tels que des histogrammes, des diagrammes en boîte, des nuages de points et des matrices de corrélation nous ont permis de représenter de manière claire et concise les principales caractéristiques de notre jeu de données.



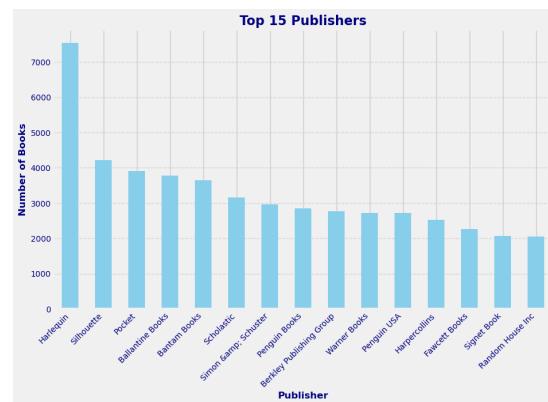
(a) Distribution of Ratings except '0'



(b) Distribution of Ratings



(c) Publication Frequency Across Years



(d) top 15 publishers

FIGURE 3.1 – graphes - visualisation des données

## 3.4 Implémentation des Algorithmes de Recommandation

Dans cette section, nous présentons en détail les différents algorithmes de recommandation que nous avons mis en œuvre dans le cadre de notre projet. Nous abordons à la fois les approches classiques et les méthodes plus avancées, en soulignant les défis rencontrés et les stratégies adoptées pour optimiser les performances de chaque algorithme.

### 3.4.1 Filtrage Collaboratif

Nous avons commencé par implémenter les algorithmes de filtrage collaboratif, qui reposent sur l'idée de recommander des éléments similaires à ceux qui ont été appréciés par des utilisateurs ayant des préférences similaires. Nous avons exploré deux variantes principales :

#### 3.4.1.1 Filtrage Collaboratif Basé sur l'Utilisateur

Cette approche calcule les similarités entre les utilisateurs en se basant sur leurs profils de notation, puis recommande à un utilisateur cible les éléments préférés par ses voisins les plus proches. Nous avons utilisé des mesures de similarité telles que la corrélation de Pearson et la similarité cosinus pour identifier ces voisins.

#### 3.4.1.2 Filtrage Collaboratif Basé sur l'Élément

Cette méthode se concentre sur les similitudes entre les éléments (livres dans notre cas), en calculant des similarités basées sur les notes attribuées par les utilisateurs. Les recommandations pour un utilisateur cible sont alors générées à partir des éléments similaires à ceux qu'il a déjà appréciés.

Ces approches de filtrage collaboratif ont donné des résultats prometteurs, mais ont également soulevé des défis tels que le démarrage à froid (cold start) pour les nouveaux utilisateurs ou éléments, ainsi que le problème de la dispersion des données (sparsity).

### 3.4.2 Filtrage Basé sur le Contenu

Nous avons ensuite implémenté des algorithmes de filtrage basé sur le contenu, qui exploitent les caractéristiques intrinsèques des éléments (livres) pour générer des recommandations. Cette approche consiste à construire des profils de préférences pour chaque utilisateur en analysant les caractéristiques des éléments qu'il a appréciés, puis à recommander des éléments similaires.

Pour mettre en œuvre cette méthode, nous avons extrait diverses caractéristiques des livres, telles que le genre littéraire, les mots-clés de la description, les auteurs et les thèmes abordés. Nous avons ensuite utilisé des techniques de traitement du langage naturel (NLP) et de pondération TF-IDF pour construire des vecteurs de caractéristiques représentant chaque livre.

Les recommandations pour un utilisateur cible ont été générées en calculant la similarité entre son profil de préférences et les vecteurs de caractéristiques des livres, en utilisant des mesures telles que la similarité cosinus ou la distance euclidienne.

### 3.4.3 Algorithmes de Voisinage (K-Nearest Neighbors)

Nous avons également exploré les algorithmes de type K-plus proches voisins (K-NN), qui combinent à la fois les aspects collaboratifs et les caractéristiques intrinsèques des éléments. Cette approche consiste à identifier les voisins les plus proches d'un utilisateur cible en tenant compte à la fois de ses préférences passées et des similitudes entre les caractéristiques des éléments.

Nous avons implémenté différentes variantes de l'algorithme K-NN, en utilisant des mesures de distance adaptées à notre jeu de données, telles que la distance de Minkowski ou la

distance de Jaccard. Nous avons également expérimenté différentes stratégies de pondération pour combiner les aspects collaboratifs et basés sur le contenu dans le calcul des similarités.

### 3.4.4 Factorisation Matricielle

Les algorithmes de factorisation matricielle, tels que la décomposition en valeurs singulières (SVD) et la factorisation non-négative de matrices (NMF), ont constitué une partie importante de notre travail. Ces techniques visent à décomposer la matrice d'interaction utilisateur-élément en produit de deux matrices de rang inférieur, capturant ainsi les principales dimensions latentes sous-jacentes aux préférences des utilisateurs.

Pour la SVD, nous avons utilisé des implémentations optimisées basées sur des décompositions en valeurs singulières tronquées, ce qui nous a permis de réduire la dimensionnalité du problème et d'améliorer les performances de calcul. Concernant la NMF, nous avons exploré différents algorithmes d'optimisation, tels que les descentes de gradient et les multiplicateurs de Lagrange alternés, pour résoudre efficacement le problème de factorisation non-négative.

Ces approches de factorisation matricielle ont démontré leur capacité à capturer des structures complexes dans les données et à générer des recommandations pertinentes, même pour les utilisateurs et les éléments ayant peu d'interactions observées.

### 3.4.5 Réseaux de Neurones et Autoencodeurs

Enfin, nous avons tenté d'exploiter les capacités des réseaux de neurones, en particulier des autoencodeurs, pour effectuer des recommandations. L'idée principale était d'entraîner un autoencodeur à reconstruire la matrice d'interaction utilisateur-élément, en capturant les motifs latents dans une représentation compacte.

Cependant, la mise en œuvre de cette approche s'est avérée complexe en raison de la nature creuse de notre matrice d'interaction et des défis liés à l'entraînement de réseaux de neurones sur des données très volumineuses. Nous avons dû nous attaquer à des problèmes tels que le sur-apprentissage, la convergence lente et l'instabilité des gradients.

Bien que nous ayons pu obtenir des résultats encourageants sur des sous-ensembles de données, l'application des autoencodeurs à l'ensemble de notre jeu de données s'est avérée difficile et n'a pas permis d'atteindre des performances compétitives par rapport aux autres approches.

Dans les sections suivantes, nous présentons une analyse détaillée des résultats obtenus avec chacun de ces algorithmes, en mettant en évidence leurs forces et leurs limites respectives dans le contexte de notre projet de recommandation de livres.

## 3.5 Évaluation des Modèles

Une étape cruciale dans le développement de notre système de recommandation a été l'évaluation rigoureuse des performances de chacun des modèles implementés. Cette évaluation nous a permis de comparer objectivement les différentes approches, d'identifier leurs forces et leurs faiblesses respectives, et de sélectionner les modèles les plus prometteurs pour une intégration dans notre système final.

### 3.5.1 Métriques d'évaluation

Nous avons utilisé métriques standard pour évaluer la qualité de recommandations générées par nos modèles. Ces métriques incluent :

- **Précision** : Cette métrique mesure la proposition de recommandations pertinentes parmi toutes les recommandations effectuées.
- **Rappel** : Le rappel évalue proposition d'éléments pertinents qui ont été effectivement recommandés parmi tous les éléments pertinents.
- **Score F1** : Le score F1 combine la précision et le rappel en une seule métrique harmonique, offrant un équilibre entre ces 2 mesures.
- **Erreur quadratique Moyenne (RMSE)** : La RMSE quantifie l'écart entre les notes prédites par le modèle et les notes réelles attribuées par les utilisateurs.

### 3.5.2 Validation croisée

Pour renforcer la robustesse de notre évaluation, nous avons mis en oeuvre une stratégie de validation croisée (cross-validation). Cette technique consiste à diviser aléatoirement le jeu de données en plusieurs sous-ensembles, puis à entraîner et à évaluer le modèle de manière itérative sur ces sous-ensembles, puis à entraîner et à évaluer le modèle de manière itérative sur ces sous-ensembles. Les résultats finaux sont obtenus en moyennant les scores des différents itérations.

La validation croisée nous a permis de réduire le biais potentiel lié à une division spécifique des données, et mieux estimer les performances généralisées de nos modèles.

### 3.5.3 Comparaison des Performances

Après avoir évalué chaque modèle individuellement, nous avons procédé à une comparaison approfondie de leurs performances respectives. Cette analyse comparative nous a permis d'identifier les forces et les faiblesses de chaque approche en fonction des différentes métriques utilisées.

...

### 3.5.4 Évaluation qualitative

Parallèlement à l'évaluation quantitative basée sur les métriques standard, nous avons également mené une évaluation qualitative en impliquant directement des utilisateurs réels. Cette approche nous a permis de recueillir des retours précieux sur l'expérience utilisateur globale de notre système de recommandation.

Nous avons organisé des sessions de test utilisateur, au cours desquelles les participants ont interagi avec notre système et ont évalué la pertinence des recommandations proposées.

Ces retours qualitatifs ont été précieux pour affiner notre compréhension des attentes et des besoins réels des utilisateurs, et pour identifier des pistes d'optimisation au-delà des performances purement quantitatives.

...

En combinant les résultats de l'évaluation quantitative et qualitative, nous avons pu sélectionner les modèles les plus prometteurs et les intégrer dans notre système de recommandation final. Les sections suivantes détaillent cette intégration, ainsi que les performances et les caractéristiques du système complet.

## 3.6 Déploiement du Système de Recommandation

Après avoir évalué et sélectionné les modèles les plus performants, nous avons procédé au déploiement de notre système de recommandation. Cette étape cruciale a impliqué l'intégration des modèles sélectionnés dans une interface utilisateur conviviale et évolutive. Nous avons opté pour une architecture client-serveur, avec une application web en front-end et une infrastructure back-end robuste pour gérer les calculs intensifs des modèles de recommandation. Nous avons mis en place des mécanismes de mise en cache et de parallélisation pour garantir des temps de réponse rapides, même avec une charge d'utilisateurs importante. Pour faciliter la maintenance et les mises à jour futures, nous avons adopté une approche modulaire, permettant de remplacer ou d'ajouter des modèles de recommandation sans perturber le reste du système. Nous avons également intégré des outils de suivi et de surveillance pour détecter les éventuels problèmes et optimiser les performances en conséquence.

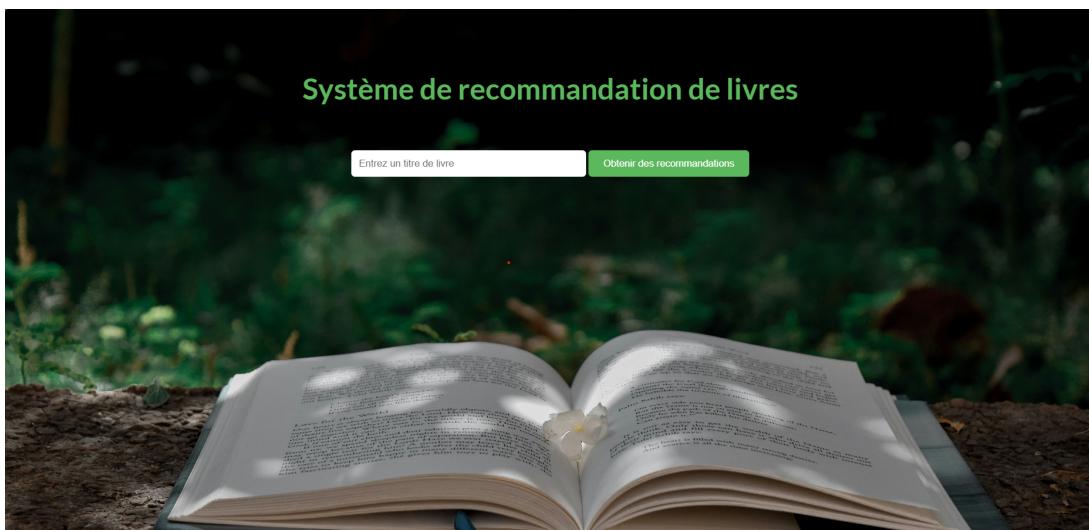


FIGURE 3.2 – web application



FIGURE 3.3 – donner le resultat

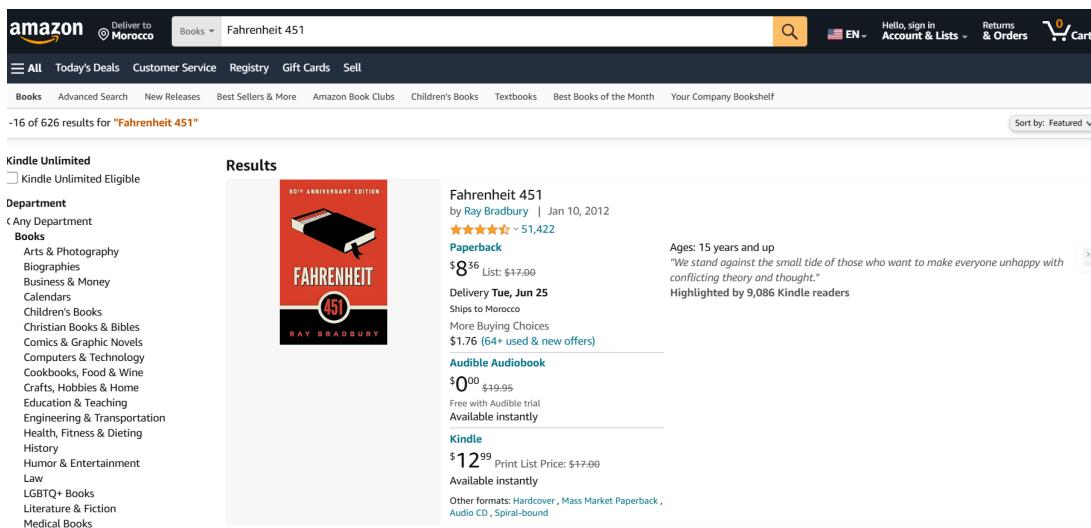


FIGURE 3.4 – après résultat

## 3.7 Cas d'utilisation et Scénarios d'Application

Notre système de recommandation de livres trouve des applications dans divers contextes, offrant une expérience personnalisée et enrichissante aux utilisateurs. Parmi les cas d'utilisation les plus pertinents, on peut citer :

Découverte de nouveaux livres : Le système permet aux utilisateurs de découvrir de nouveaux titres en phase avec leurs préférences et leurs lectures passées, élargissant ainsi leurs horizons littéraires. Suggestions de lecture pour les clubs de lecture : Les recommandations peuvent être utilisées pour suggérer des livres à discuter lors des réunions de clubs de lecture, en tenant compte des goûts diversifiés des membres. Recommandations personnalisées pour les libraires : Les libraires peuvent utiliser notre système pour offrir des recommandations sur mesure à leurs clients, améliorant ainsi l'expérience d'achat et la satisfaction des clients. Analyse des tendances de lecture : Les données collectées par notre système peuvent être exploitées pour analyser les tendances de lecture, identifier les genres populaires et aider les éditeurs à prendre des décisions éclairées sur les futurs lancements de livres.

## 3.8 Défis et Limites

Malgré les performances prometteuses de notre système de recommandation, nous avons identifié certains défis et limites qu'il convient de prendre en compte pour des améliorations futures :

- Biais dans les données** : Comme la plupart des systèmes de recommandation, notre approche peut être influencée par les biais présents dans les données d'entraînement, tels que la surreprésentation de certains groupes démographiques ou de certains genres littéraires.
- Évolutivité à long terme** : Bien que notre architecture soit conçue pour être évolutive, l'augmentation constante du volume de données et du nombre d'utilisateurs pourrait nécessiter des ajustements supplémentaires pour maintenir des performances optimales.
- Explicabilité des recommandations** : Certains modèles, comme les réseaux de neurones, peuvent être considérés comme des "boîtes noires", rendant difficile l'explication

des recommandations générées. L'amélioration de l'explicabilité pourrait renforcer la confiance des utilisateurs dans le système.

4. **Prise en compte du contexte** : Notre système actuel ne tient pas compte des facteurs contextuels, tels que l'humeur de l'utilisateur, l'environnement de lecture ou les événements récents, qui peuvent influencer les préférences de lecture.

## 3.9 Conclusion

Ce chapitre a détaillé le développement et la mise en œuvre de notre système de recommandation de livres, depuis la collecte et le prétraitement des données jusqu'à l'évaluation des modèles et leur déploiement. Nous avons exploré diverses techniques, allant des méthodes classiques aux approches avancées, pour offrir des suggestions personnalisées et pertinentes aux utilisateurs. Malgré certains défis et limites identifiés, notre démarche rigoureuse et nos évaluations quantitatives et qualitatives ont permis de créer un système performant et innovant, prêt à enrichir l'expérience de lecture de nombreux utilisateurs.

# Conclusion générale

Ce rapport a mené une exploration approfondie du domaine prometteur des systèmes de recommandation améliorés, soulignant leur potentiel à transformer de manière significative l'expérience utilisateur numérique. En exploitant des techniques d'apprentissage automatique de pointe telles que la factorisation matricielle et les algorithmes de voisinage, conjuguées à une intégration judicieuse de facteurs contextuels cruciaux, ces systèmes offrent des recommandations d'une précision remarquable et d'un niveau de personnalisation inégalé.

Notre projet, dédié à l'optimisation des recommandations de livres sur la plateforme de commerce électronique Amazon, a démontré l'efficacité de ces approches avant-gardistes. En combinant intelligemment des algorithmes éprouvés comme le filtrage collaboratif et le filtrage basé sur le contenu avec des méthodes avancées telles que les k-plus proches voisins, la décomposition en valeurs singulières et la factorisation non négative de matrices, nous avons réussi à développer un système de recommandation performant, capable de capturer avec une grande finesse les préférences complexes des utilisateurs.

Cependant, nos travaux ont également mis en exergue les défis intrinsèques à ce domaine, notamment la nécessité de gérer efficacement les problématiques liées au démarrage à froid, à la rareté des données et aux biais potentiels. Nous avons souligné l'importance primordiale d'une évaluation rigoureuse des systèmes de recommandation, reposant sur des métriques standard reconnues et des tests utilisateurs qualitatifs approfondis, afin de garantir une expérience optimale aux utilisateurs finaux.

En déployant notre solution au sein d'une architecture évolutive et modulaire, nous avons démontré la faisabilité concrète d'intégrer ces techniques novatrices au cœur d'applications réelles, ouvrant ainsi de multiples perspectives d'application prometteuses, allant de la découverte de nouveaux contenus littéraires à l'analyse fine des tendances de lecture.

Bien que notre projet ait atteint des résultats remarquables, nous reconnaissions qu'il subsiste des défis à relever, notamment en termes de gestion des biais inhérents aux données, d'explicabilité des recommandations générées et de prise en compte des facteurs contextuels propres à chaque utilisateur. Ces aspects représentent des pistes de recherche stimulantes pour l'avenir, offrant l'opportunité de repousser encore les limites de ces systèmes révolutionnaires.

En définitive, ce projet a confirmé le potentiel transformateur des systèmes de recommandation améliorés, capables de guider avec discernement les utilisateurs à travers l'océan d'informations disponibles en ligne, pour leur offrir une expérience sur mesure, enrichissante et époustouflante. Au fur et à mesure que ces technologies continueront d'évoluer, elles promettent de redéfinir en profondeur notre rapport au contenu numérique, en harmonisant de manière transparente nos préférences individuelles avec un vaste éventail de possibilités passionnantes, au service d'une expérience utilisateur toujours plus immersive et personnalisée. . . .

# **Bibliographie**