

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique

ECOLE SUPÉRIEURE EN INFORMATIQUE
8 Mai 1945 - Sidi-Bel-Abbès



الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

المدرسة العليا للإعلام الآلي
8 ماي 1945 - سيدي بلعباس

ECOLE SUPÉRIEURE EN INFORMATIQUE 08-MAI-1945 SIDI BEL ABBES

RAPPORT DE MINI PROJET

Système de Recommandation de Livres

Présenté par :

- Ghandouz Amina
- Benghenima Hafsa
- Benahmed Firdaws

Encadrant :

Dr. Nassima DIF

Date de publication : Décembre 2024

Table des matières

1	Introduction	2
2	Type de Système de Recommandation	3
3	Sources de Données Utilisées	3
4	Prétraitement et Nettoyage des Données	4
5	Techniques et Algorithmes Utilisés	6
5.1	Apriori	6
5.2	Eclat	6
5.3	FP-Growth	6
6	Étude Comparative et Sélection du Meilleur Modèle	7
6.1	Résultats Comparatifs	7
6.2	Meilleur Modèle Sélectionné	7

1 Introduction

Les systèmes de recommandation font partie intégrante de nos vies, ils nous aident à mieux acheter, trouver du nouveau contenu, etc. Pour les entreprises, ces systèmes sont un outil essentiel pour augmenter leurs ventes et la satisfaction des clients.

La majorité des bibliothèques n'ont pas de systèmes de recommandation de livres, ce qui les empêche de faire découvrir aux utilisateurs des livres qui pourraient les intéresser. Par conséquent, si les bibliothèques pouvaient recommander automatiquement des livres, elles pourraient mieux promouvoir la lecture.

Le but de ce projet est de créer un système de recommandation de livres basé sur les livres que les utilisateurs lisent et qui ont des intérêts similaires.

Les retombées du projet seraient d'avoir un système fonctionnel pouvant être intégré aux bibliothèques, permettant aux utilisateurs de découvrir des livres intéressants qu'ils ne connaissaient pas.

Dans ce projet, nous explorons les techniques d'apprentissage automatique non supervisé, en particulier les algorithmes d'association, pour développer un système de recommandation de livres. Nous avons expérimenté trois algorithmes : **Apriori**, **ECLAT**, et **FP-Growth**, afin de découvrir les associations les plus pertinentes entre les livres. Après une analyse comparative, nous avons sélectionné le meilleur modèle et l'avons déployé pour fournir des recommandations efficaces.

Le code complet du projet est disponible sur GitHub via le lien :

<https://github.com/VotreLien/nom-du-projet>

2 Type de Système de Recommandation

Le système de recommandation que nous avons développé utilise des algorithmes basés sur les règles d'association, en particulier Apriori, Eclat, et FP-Growth, qui sont des techniques de découverte d'ensembles fréquents.

Dans ce type de système, nous identifions des ensembles de livres fréquemment lus ensemble par plusieurs utilisateurs. Lorsque qu'un utilisateur a lu un livre qui fait partie de ces ensembles fréquents, nous lui recommandons l'autre livre du même ensemble. Par exemple, si un ensemble fréquent est formé des livres A et B (lisent ensemble souvent par les utilisateurs), et qu'un utilisateur a déjà lu le livre A, le système recommande automatiquement le livre B.

3 Sources de Données Utilisées

Nous avons utilisé deux datasets principaux pour ce projet : `books_df` et `ratings_df`.

- Le dataset `book_df` contient 271 360 lignes et 8 colonnes, représentant les informations relatives aux livres, telles que le titre, l'auteur et le genre.
- Le dataset `ratings_df` contient 1 149 780 lignes et 3 colonnes, représentant les évaluations des utilisateurs. Chaque ligne associe un identifiant d'utilisateur User-ID, un identifiant de livre ISBN, et une évaluation (rating).

Nous avons filtré le dataset `ratings_df` pour ne conserver que les évaluations dont la note est supérieure à 5. Cette étape permet de se concentrer uniquement sur les livres ayant reçu des évaluations relativement positives, afin d'éviter d'inclure des évaluations faibles qui pourraient biaiser les recommandations. Ensuite, nous avons fusionné ce sous-ensemble avec le dataset `books_df` en utilisant l'ISBN comme clé de jointure. Cela nous a permis d'obtenir un dataset final contenant à la fois les titres des livres et leurs évaluations positives, essentiel pour générer des recommandations précises.

	User-ID	ISBN	Book-Rating	Book-Title
0	276729	0521795028	6	The Amsterdam Connection : Level 4 (Cambridge ...
1	276744	038550120X	7	A Painted House
2	276747	0060517794	9	Little Altars Everywhere
3	276747	0671537458	9	Waiting to Exhale
4	276747	0679776818	8	Birdsong: A Novel of Love and War

FIGURE 1 – Extrait du dataset final fusionné.

4 Prétraitement et Nettoyage des Données

Après la fusion des datasets, nous avons effectué un nettoyage pour préparer les données avant de les utiliser pour les recommandations. Plusieurs étapes ont été réalisées pour garantir la qualité du dataset fusionné :

- **Doublons** : Des doublons ont été identifiés, notamment des évaluations répétées pour un même utilisateur et un même livre. Ces doublons ont été supprimés afin de garantir l'intégrité des données.
- **Valeurs manquantes** : Aucune valeur manquante n'a été trouvée dans les colonnes essentielles du dataset. Par conséquent, ce problème n'a pas nécessité de traitement particulier.
- **Filtrage des évaluations** : Nous avons sélectionné uniquement les évaluations supérieures à 5, afin de nous concentrer sur les livres ayant reçu des évaluations positives.

Ces étapes ont permis d'obtenir un dataset propre et prêt à être utilisé pour les recommandations.

Afin de visualiser les effets du nettoyage, nous avons comparé la distribution des évaluations avant et après le prétraitement :

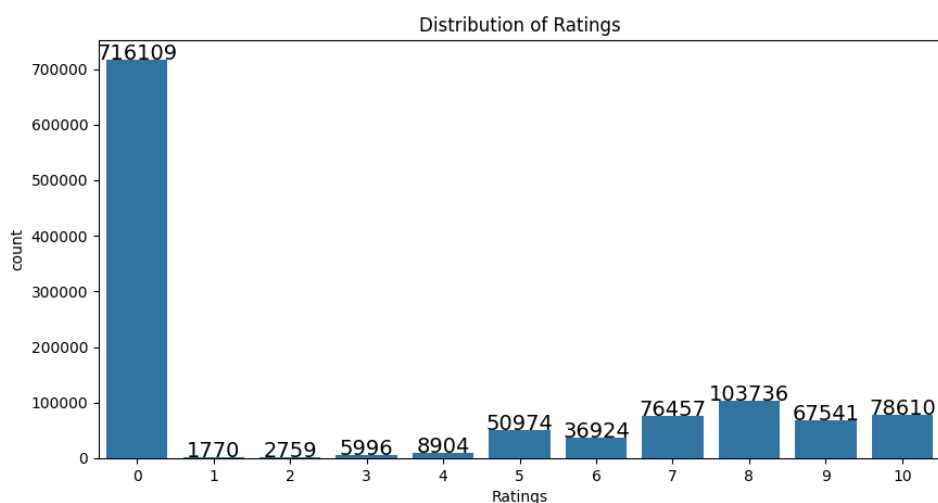


FIGURE 2 – Distribution des évaluations avant le prétraitement

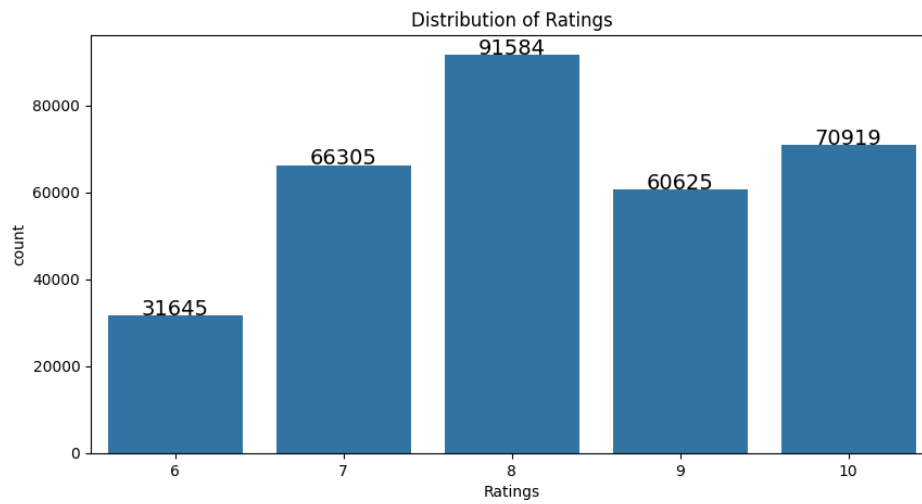


FIGURE 3 – Distribution des évaluations après le prétraitement

5 Techniques et Algorithmes Utilisés

Pour les recommandations, nous avons utilisé trois techniques populaires basées sur les règles d'association : **Apriori**, **Eclat**, et **FP-Growth**.

5.1 Apriori

5.2 Eclat

5.3 FP-Growth

6 Étude Comparative et Sélection du Meilleur Modèle

Après avoir appliqué les algorithmes Apriori, Eclat, et FP-Growth, nous avons comparé leurs performances afin de sélectionner le modèle le plus adapté pour notre système de recommandation. Cette comparaison s'est basée sur plusieurs critères essentiels :

Temps d'exécution : Chaque algorithme a été évalué en termes de temps nécessaire pour traiter le dataset final et générer les règles d'association.

Nombre de règles extraites : Nous avons analysé la quantité de règles générées par chaque algorithme, en prenant en compte leur pertinence et leur diversité.

Simplicité des règles : Un autre critère a été la lisibilité et la simplicité des règles d'association pour leur interprétation par les utilisateurs.

Support et confiance minimaux : Les valeurs de support et de confiance minimaux influencent directement la qualité des recommandations. Ces paramètres ont été ajustés pour chaque algorithme afin de garantir des résultats optimaux.

6.1 Résultats Comparatifs

Apriori : Cet algorithme a généré des règles d'association pertinentes, mais il était le plus lent en termes de temps d'exécution, surtout avec des datasets de grande taille.

Eclat : Bien que plus rapide qu'Apriori, Eclat a parfois produit des règles légèrement redondantes, nécessitant un filtrage supplémentaire.

FP-Growth : Cet algorithme s'est distingué par son efficacité, générant rapidement des règles pertinentes avec une utilisation optimale des ressources.

6.2 Meilleur Modèle Sélectionné

Sur la base des résultats obtenus, FP-Growth a été identifié comme le meilleur modèle. Les critères principaux pour ce choix ont été :

- Son temps d'exécution rapide sur des datasets volumineux.
- La pertinence des règles générées.
- Sa capacité à s'adapter aux variations des paramètres comme le support et la confiance minimaux.