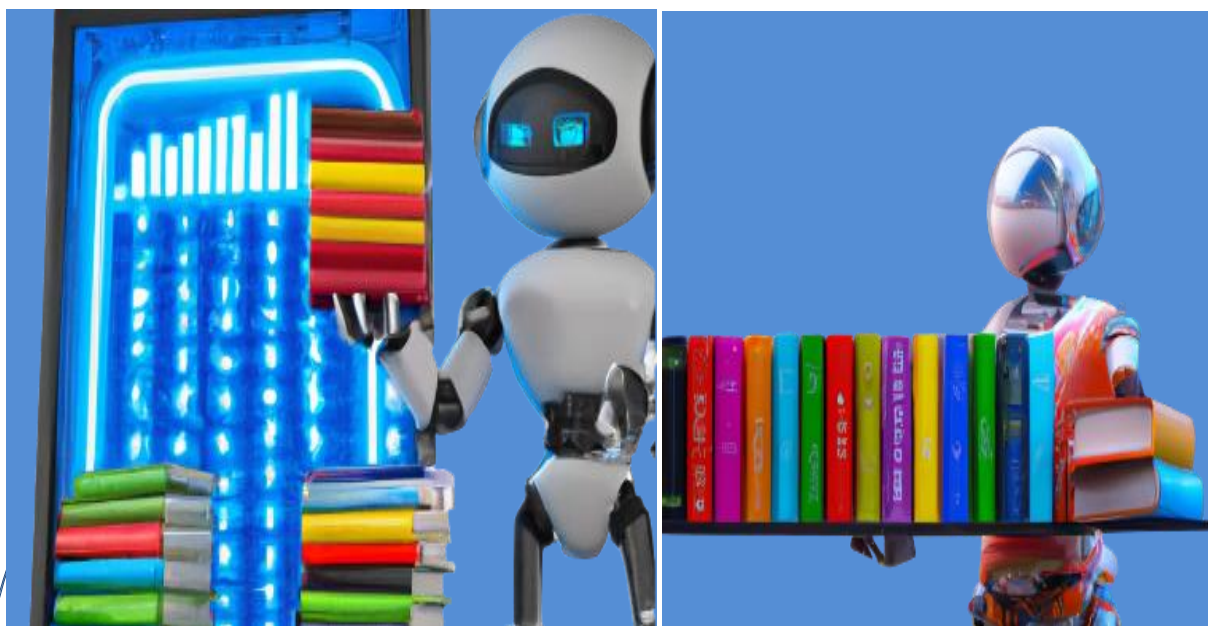


Rapport du Projet Libre:

<< *Système de Recommandation*

***De Livres* >>**



Réalisé par :

- **Abdellatif Satir**

A.U : 2022/2023

1- Introduction :

Un **système de recommandation** est un type de système de filtrage d'informations qui vise à prédire la "note" ou la "préférence" qu'un utilisateur donnerait à un élément.

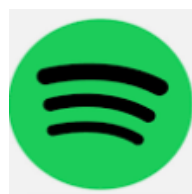
Le système fait des recommandations basées sur les comportements passés et les préférences de l'utilisateur, ainsi que sur les modèles et les similitudes entre les utilisateurs et les éléments. Les systèmes de recommandation sont largement utilisés dans diverses applications, notamment le commerce électronique, la vidéo en ligne, les services de streaming de musique et de livres, les articles de presse et les médias sociaux.

L'objectif d'un système de recommandation est d'améliorer l'expérience utilisateur en fournissant des recommandations personnalisées, conduisant à un engagement accru, à la satisfaction des clients et potentiellement à une augmentation des revenus.

Exemples :

Il existe de nombreux exemples de systèmes de recommandation dans notre vie quotidienne, dont certains incluent :

- Achats en ligne : les sites Web de commerce électronique tels qu'Amazon, Netflix et Spotify utilisent des systèmes de recommandation pour suggérer des produits ou des articles susceptibles d'intéresser un utilisateur.



- Médias sociaux : des plates-formes telles que Facebook, Instagram et Twitter utilisent des systèmes de recommandation pour suggérer de nouveaux amis, publications et pages aux utilisateurs.



- Actualités et médias : les sites Web d'actualités tels que CNN et la BBC utilisent des systèmes de recommandation pour suggérer des articles et des vidéos correspondant aux centres d'intérêt de l'utilisateur.



- Diffusion de musique : les services de diffusion de musique tels que Spotify et Pandora utilisent des systèmes de recommandation pour suggérer de nouvelles chansons et listes de lecture aux utilisateurs en fonction de leur historique d'écoute.

- Diffusion de films et de séries TV : des plates-formes telles que Netflix et Hulu utilisent des systèmes de recommandation pour suggérer de nouveaux films et émissions de télévision aux utilisateurs en fonction de leur historique de visionnage et de leurs préférences.

- Jeux : les plates-formes de jeux en ligne telles que Steam et Xbox Live utilisent des systèmes de recommandation pour suggérer de nouveaux jeux aux joueurs en fonction de leur historique de jeu et de leurs préférences.



- Voyages : les agences de voyages et les compagnies aériennes en ligne utilisent des systèmes de recommandation pour suggérer des hôtels, des vols

et des forfaits vacances en fonction de l'historique et des préférences de voyage d'un utilisateur.

Ce ne sont là que quelques exemples, mais les systèmes de recommandation peuvent également être appliqués à de nombreux autres domaines.

2- Problématique :

Un système de recommandation de livres peut résoudre plusieurs problèmes, notamment :

Surcharge d'informations : avec la grande quantité de livres disponibles, il peut être difficile pour les utilisateurs de trouver les livres qui les intéressent. Un système de recommandation peut aider les utilisateurs à trouver rapidement et facilement des livres qui correspondent à leurs préférences et à leurs intérêts.

Difficulté à trouver des livres pertinents : les utilisateurs ont souvent du mal à trouver des livres qui correspondent à leurs préférences et à leurs intérêts, ce qui entraîne de la frustration et une mauvaise expérience utilisateur. Un système de recommandation peut aider les utilisateurs à découvrir de nouveaux livres correspondant à leurs centres d'intérêt.

Inefficacités du marketing : les éditeurs et les libraires ont souvent du mal à cibler efficacement leurs efforts de marketing, ce qui entraîne une utilisation inefficace des ressources et un impact limité. Un système de recommandation peut fournir des informations sur les préférences et les comportements des utilisateurs, ce qui permet de cibler plus efficacement les efforts de marketing.

Diminution des ventes : si les utilisateurs ne parviennent pas à trouver des livres correspondant à leurs centres d'intérêt et à leurs préférences, ils peuvent être moins enclins à effectuer un achat, ce qui entraîne une diminution des ventes. Un système de recommandation peut aider à augmenter les ventes en fournissant aux utilisateurs des recommandations personnalisées.

Dans l'ensemble, un système de recommandation de livres peut aider à résoudre ces problèmes en fournissant aux utilisateurs des recommandations

pertinentes et personnalisées, conduisant à un engagement accru, à la satisfaction des clients et potentiellement à une augmentation des ventes.

3- L'idée de projet :

L'idée de notre projet est de créer un site Web qui est un modèle d'apprentissage automatique pour prédire et recommander des livres similaires au livre entré par l'utilisateur que l'utilisateur pourrait aimer aussi



4- Types de system de recommandation:

Dans l'apprentissage automatique, un système de recommandation est un type d'algorithme qui utilise le comportement et les préférences passés de l'utilisateur pour prédire ce qu'un utilisateur pourrait aimer à l'avenir.

Il existe plusieurs types de systèmes de recommandation, notamment :

- **Filtrage collaboratif** : cette approche construit un modèle à partir du comportement passé d'un utilisateur (articles précédemment achetés ou sélectionnés et/ou notes numériques attribuées à ces articles) ainsi que des décisions similaires prises par d'autres utilisateurs. Le filtrage collaboratif peut être divisé en deux catégories : basé sur l'utilisateur et basé sur les éléments.

- **Filtrage basé sur le contenu** : cette approche utilise une série de caractéristiques discrètes d'un élément pour recommander des éléments supplémentaires avec des propriétés similaires. L'algorithme essaie de faire correspondre les éléments en fonction de leurs caractéristiques et attributs, tels que le genre, les acteurs, le réalisateur, etc.

- **Système de recommandation hybride** : Cette approche combine à la fois des méthodologies de filtrage collaboratives et basées sur le contenu afin de faire des recommandations plus précises.

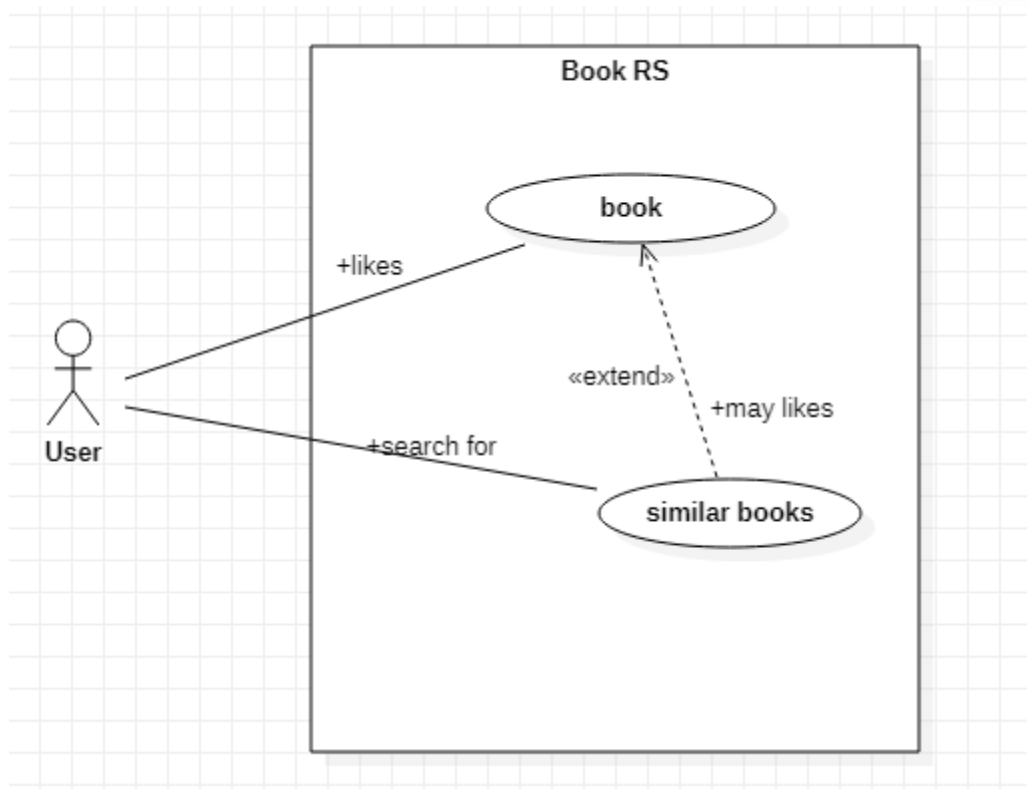
- **Systèmes de recommandation basés sur la popularité** : ces systèmes formulent des recommandations basées sur les éléments les plus populaires d'une population, quelles que soient les préférences individuelles des utilisateurs. Ces types de systèmes de recommandation peuvent être utiles dans certaines situations, par exemple lorsque les données des utilisateurs individuels ne sont pas disponibles ou lorsqu'il est plus pertinent de formuler des recommandations basées sur les grandes tendances de la population. Cependant, ils peuvent ne pas fournir des recommandations aussi personnalisées et précises que les systèmes de recommandation individualisés.

Dans cette projet nous allons créer deux types de recommandation system de livres :

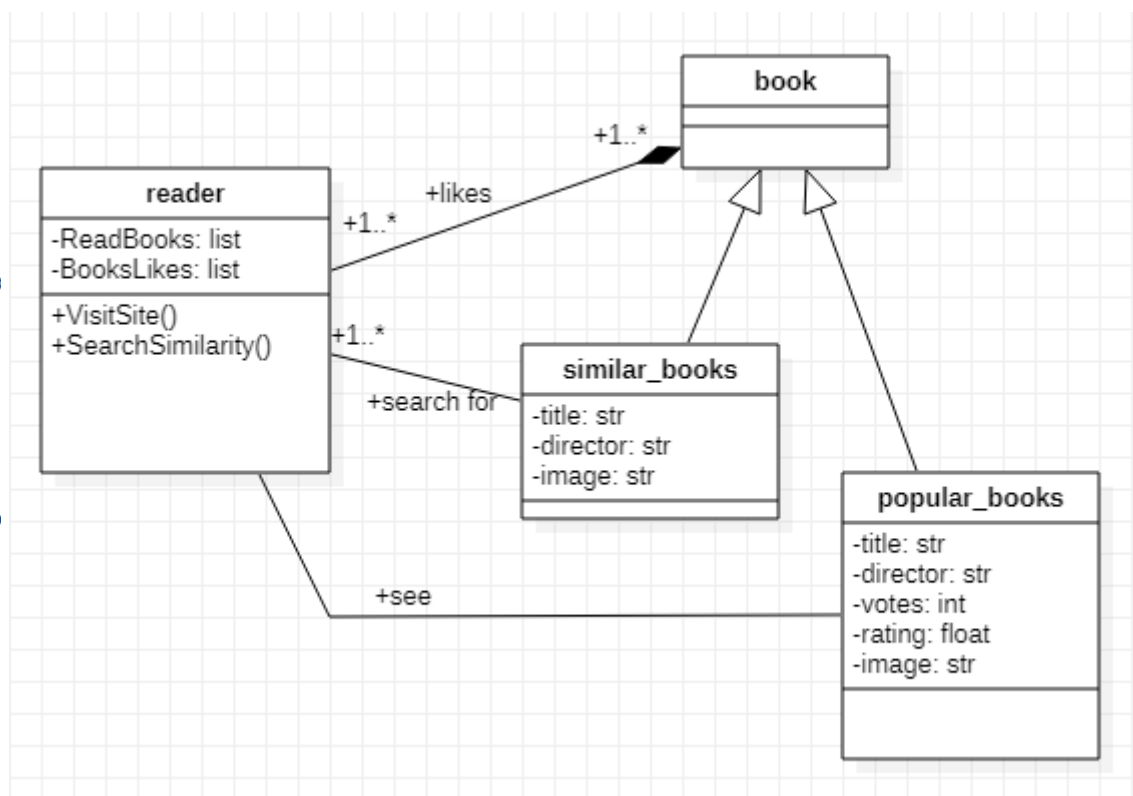
- Collaborative Filtering.
- Popularity based recommendation system.

5- Conception et technologies utilisé :

- **Use case diagram**



- Class diagram



- Technologie utilisé :



6- Implémentation :

- Data :

Dans ce projet, nous utiliserons l'ensemble de données de recommandations de livres de kaggle :

<https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset>

Data Explorer

Version 1 (106.94 MB)

Books.csv

Ratings.csv

Users.csv

Le jeu de données Book-Crossing comprend 3 fichiers.

- 1- Users : Contient les utilisateurs. Notez que les ID utilisateur (User-ID) ont été anonymisés et mappés sur des entiers. Les données démographiques sont fournies (localisation, âge) si elles sont disponibles. Sinon, ces champs contiennent des valeurs NULL.
- 2- Ratings : Contient les informations sur la notation du livre. Les notations (Book-Rating) sont soit explicites, exprimées sur une échelle de 1 à 10 (les valeurs les plus élevées dénotant une appréciation plus élevée), soit implicites, exprimées par 0. 1149780 rating

ratings

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6
...

- 3- Books : Les livres sont identifiés par leur ISBN respectif. Les ISBN non valides ont déjà été supprimés de l'ensemble de données. De plus, certaines informations basées sur le contenu sont fournies (titre du livre, auteur du livre, année de publication, éditeur), obtenues auprès d'Amazon Web Services. Notez qu'en cas de plusieurs auteurs, seul le premier est fourni. Les URL renvoyant aux images de couverture sont également données, apparaissant en trois versions différentes (Image-URL-S, Image-URL-M, Image-URL-L), c'est-à-dire petite, moyenne, grande. Ces URL pointent vers le site Web d'Amazon.

books

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...	http://images.amazon.com/images/P/0002005018.0...
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...	http://images.amazon.com/images/P/0060973129.0...
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...	http://images.amazon.com/images/P/0374157065.0...
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...	http://images.amazon.com/images/P/0393045218.0...
...

271360 books

• Population Filtering

Un système de recommandation basé sur la population fait référence à un système de recommandation qui prend en compte les préférences et les habitudes d'un grand groupe de personnes, plutôt que des utilisateurs individuels. L'objectif de ce type de système de recommandation est de faire des recommandations basées sur les préférences et les comportements collectifs de la population, plutôt que sur les utilisateurs individuels.

Pour ça nous avons combiné les deux datasets books et ratings, et compté nombre de Books-Ratings et puis on nous calcule moyen pour chaque livre, et nous avons filtré le résultats pour afficher juste les 100 premiers.

popular_df

	Book-Title	Num_ratings	Avg_ratings
80434	Harry Potter and the Prisoner of Azkaban (Book 3)	428	5.852804
80422	Harry Potter and the Goblet of Fire (Book 4)	387	5.824289
80441	Harry Potter and the Sorcerer's Stone (Book 1)	278	5.737410
80426	Harry Potter and the Order of the Phoenix (Boo...	347	5.501441
80414	Harry Potter and the Chamber of Secrets (Book 2)	556	5.183453
...
176827	The Beach House	435	3.347126
191926	The Hours: A Novel	294	3.346939
89780	Icy Sparks	309	3.346278
14351	Angela's Ashes (MMP) : A Memoir	326	3.337423
86511	How to Be Good	304	3.305921

100 rows × 3 columns

Top 100 Book

• Collaborative Filtering :

Définition :

Le filtrage collaboratif est un type de système de recommandation qui fait des recommandations basées sur le comportement passé et les préférences des utilisateurs. Il fonctionne en utilisant les données générées par les utilisateurs pour identifier les modèles et les relations entre les éléments, puis en utilisant ces modèles pour faire des recommandations.

Le filtrage collaboratif est un système de recommandation populaire et efficace, car il est basé sur les préférences et les comportements réels des utilisateurs, plutôt que sur des hypothèses ou des généralisations sur la population.

Il existe deux principaux types de filtrage collaboratif : basé sur l'utilisateur et basé sur les éléments.

- **Filtrage collaboratif basé sur l'utilisateur :** ce type de filtrage collaboratif émet des recommandations basées sur les similitudes entre les utilisateurs.

Cela fonctionne en trouvant d'autres utilisateurs qui ont des préférences similaires à l'utilisateur actuel, puis en recommandant des éléments que ces utilisateurs similaires ont aimés dans le passé.

- **Filtrage collaboratif basé sur les éléments** : ce type de filtrage collaboratif émet des recommandations basées sur les similitudes entre les éléments. Cela fonctionne en trouvant des éléments similaires aux éléments qu'un utilisateur a aimés dans le passé, puis en recommandant ces éléments similaires à l'utilisateur.

Préparation :

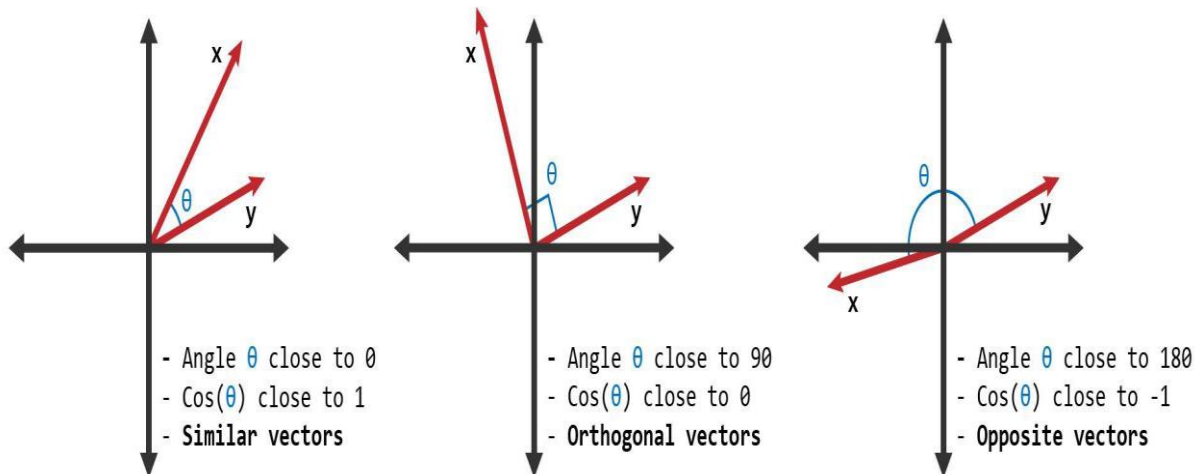
Pour ça nous avons combiné les deux datasets books et ratings, et sélectionner uniquement les utilisateurs évalués à plus de 200 livres , et après sélectionner uniquement les livres qui ont été évalués par plus de 50 utilisateurs 'famous_book' . Puis la création d'un tableau croisé dynamique (pivot table) contient chaque livre et le score d'évaluation des utilisateurs :

User-ID	254	2276	2766	2977	3363	4017	4385	6251	6323	6543	...	271705
Book-Title												
1984	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	10.0
1st to Die: A Novel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	...	0.0
2nd Chance	0.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
4 Blondes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
A Bend in the Road	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
...
Year of Wonders	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
You Belong To Me	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
Zen and the Art of Motorcycle Maintenance: An Inquiry into Values	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
Zoya	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0

706 rows × 810 columns

Cosine Similarity :

La similarité cosinus est une mesure de similarité entre deux vecteurs qui peut être utilisée dans les systèmes de recommandation pour déterminer la similarité entre les utilisateurs ou les éléments, et est particulièrement utile dans les scénarios de données clairsemées.



La similitude cosinus entre deux vecteurs (X et Y) est calculée comme le cosinus de l'angle entre les deux vecteurs, représenté comme le produit scalaire des deux vecteurs divisé par le produit des grandeurs des deux vecteurs. La similarité cosinus va de -1 à 1, où une similarité cosinus de 1 indique que les deux vecteurs sont identiques, et une similarité cosinus de -1 indique que les deux vecteurs sont complètement différents. Une similitude cosinus de 0 indique que les vecteurs sont orthogonaux et n'ont aucune similitude.

```
cos_sim = cosine_similarity(pivot)
print(cos_sim)

(706, 706)
[[1.          0.10255025 0.01220856 ... 0.12110367 0.07347567 0.0431604
 6]
 [0.10255025 1.          0.2364573 ... 0.07446129 0.16773875 0.1426339
 7]
 [0.01220856 0.2364573 1.          ... 0.04558758 0.04938579 0.1079611
 9]
 ...
 [0.12110367 0.07446129 0.04558758 ... 1.          0.07085128 0.0196177
 1]
 [0.07347567 0.16773875 0.04938579 ... 0.07085128 1.          0.1060296
 2]
 [0.04316046 0.14263397 0.10796119 ... 0.0196177 0.10602962 1.
 2]]
```

706 x 706

String Matching :

Est une technique utilisée pour rechercher des modèles dans des chaînes et peut être utilisée dans des systèmes de recommandation pour faire correspondre des éléments en fonction d'attributs textuels, tels que des noms de produits ou des titres de livres. Par exemple, les algorithmes de correspondance exacte des chaînes (string matching) : *FuzzyWuzzy*

FuzzyWuzzy est une bibliothèque Python qui utilise Fuzzy string matching algorithms

Dans le contexte des systèmes de recommandation, *FuzzyWuzzy* peut être utilisé pour effectuer une correspondance de chaîne floue entre les éléments afin d'identifier et de recommander des éléments similaires aux utilisateurs.

Par exemple, si un utilisateur recherche un livre avec un titre similaire à celui qu'il a déjà lu, *FuzzyWuzzy* peut être utilisé pour effectuer une correspondance de chaîne floue entre le titre recherché et les titres des livres dans la base de données du système de recommandation pour identifier et recommander des livres similaires.

FuzzyWuzzy est particulièrement utile dans les systèmes de recommandation où les données peuvent contenir des erreurs, des fautes de frappe ou des variations d'orthographe ou de formulation, car il peut toujours identifier les correspondances entre les chaînes même lorsqu'elles ne sont pas exactes.

Exemple :

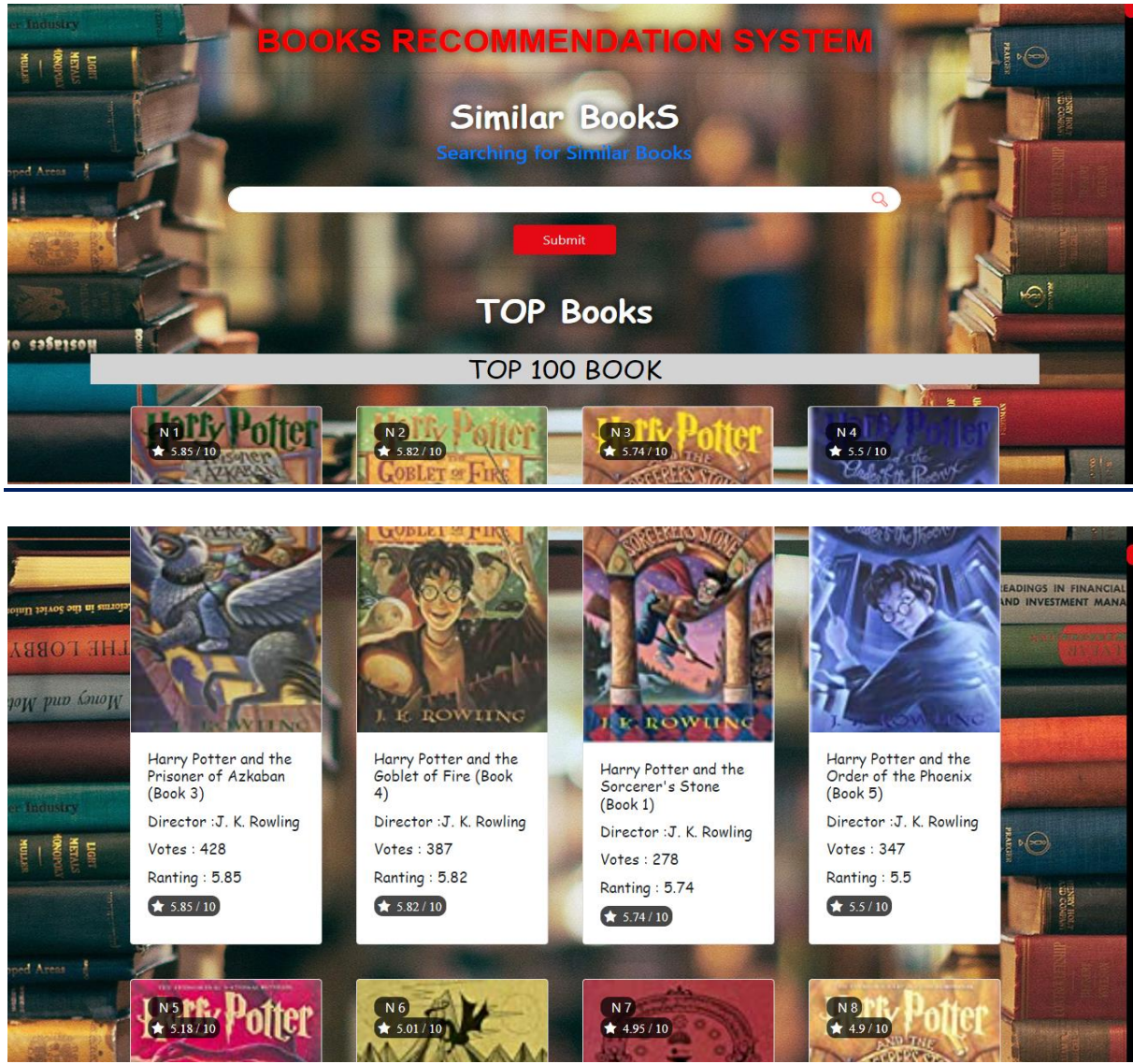
```
In [104]: from fuzzywuzzy import process # match the input text to the data
def test_matching(name):
    match_name = process.extractOne(name , pivot.index)[0]
    return match_name

test_matching("Year of ")

Out[104]: 'Year of Wonders'
```


- Interface web :

Home page TOP 100 BOOKS :



BOOKS RECOMMENDATION SYSTEM

Similar Books
Searching for Similar Books

Submit

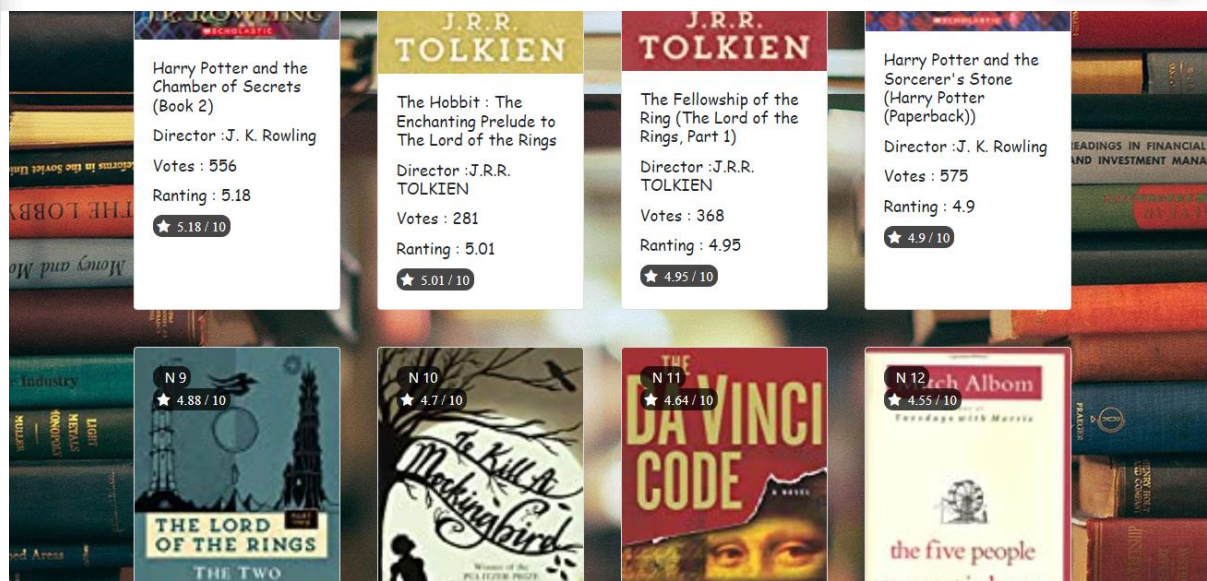
TOP Books

TOP 100 BOOK

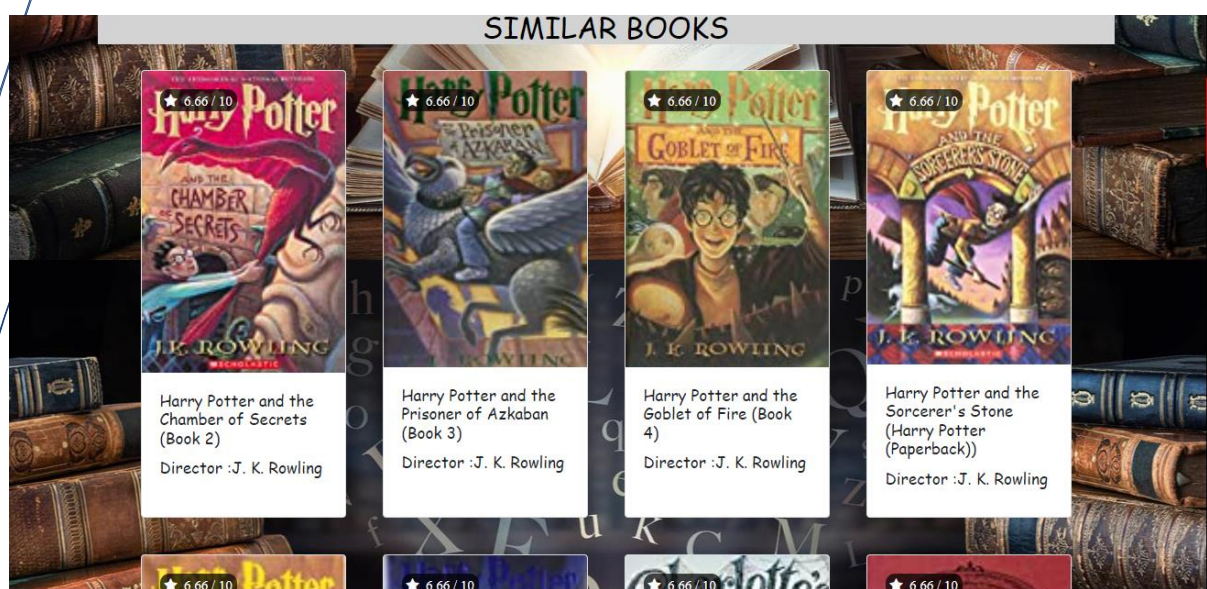
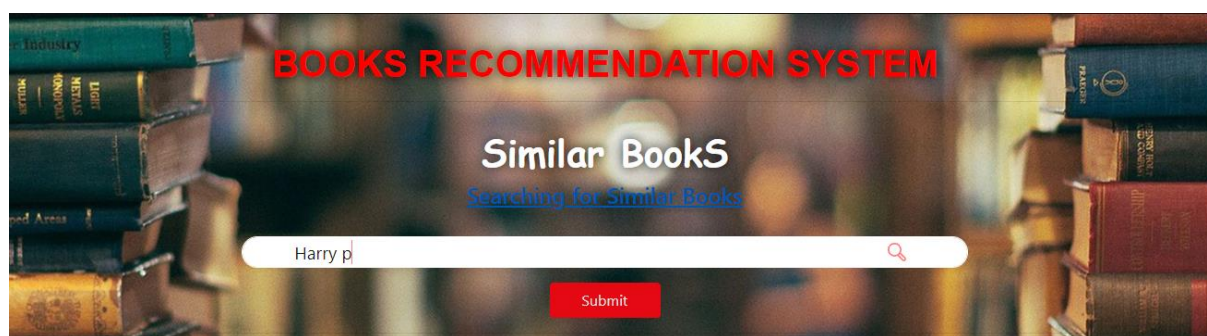
Rank	Book Title	Rating
N1	Harry Potter and the Prisoner of Azkaban (Book 3)	5.85 / 10
N2	Harry Potter and the Goblet of Fire (Book 4)	5.82 / 10
N3	Harry Potter and the Sorcerer's Stone (Book 1)	5.74 / 10
N4	Harry Potter and the Order of the Phoenix (Book 5)	5.5 / 10
N5	Harry Potter and the Chamber of Secrets (Book 2)	5.18 / 10
N6	Harry Potter and the Half-Blood Prince (Book 6)	5.01 / 10
N7	Harry Potter and the Deathly Hallows (Book 7)	4.95 / 10
N8	Harry Potter and the Philosopher's Stone (Book 1)	4.9 / 10

Detailed view of book cards:

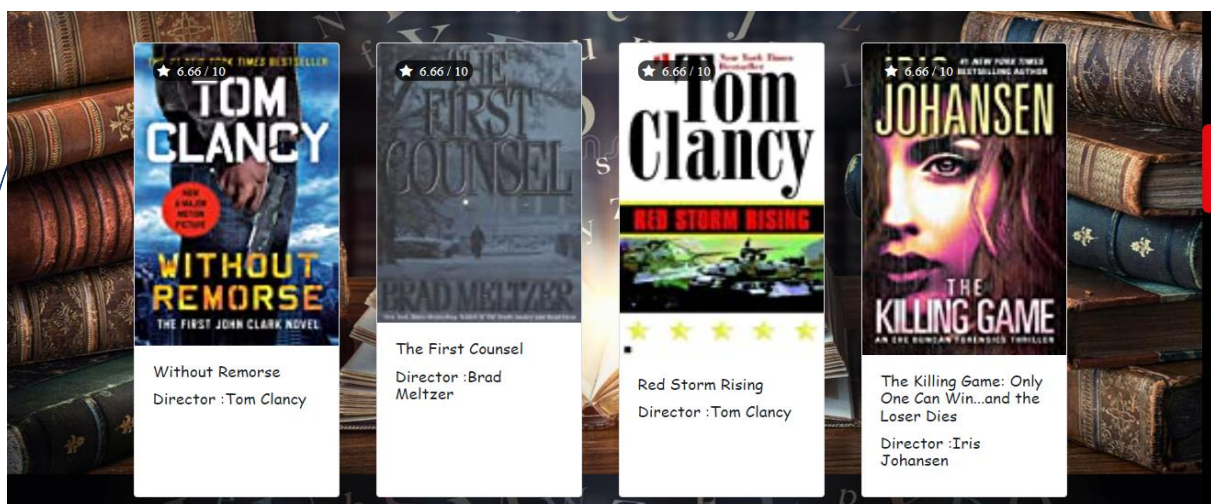
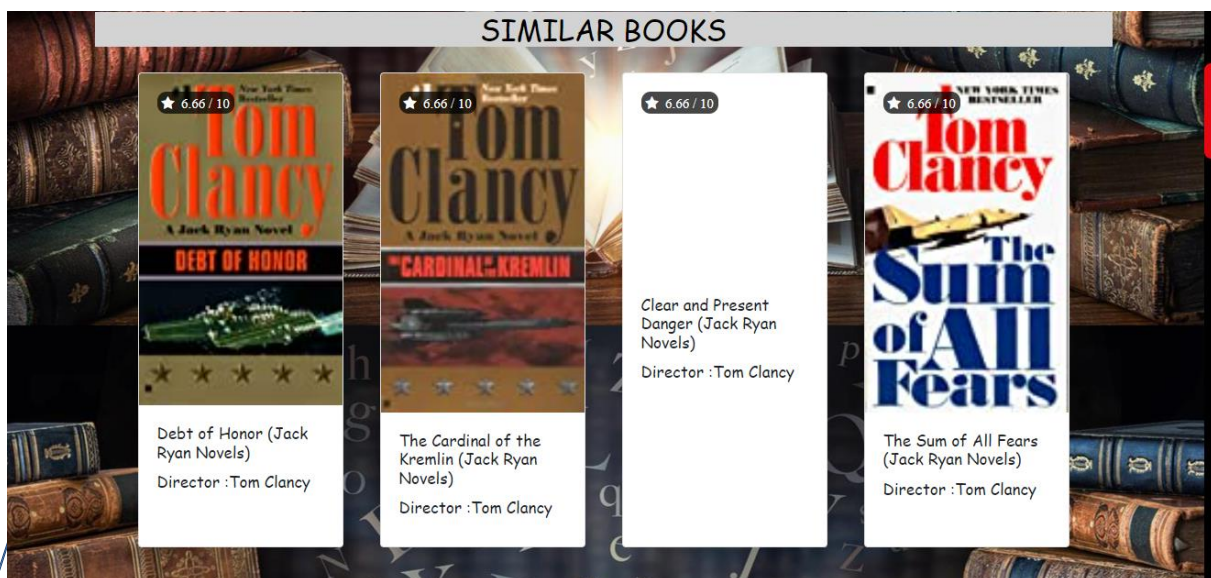
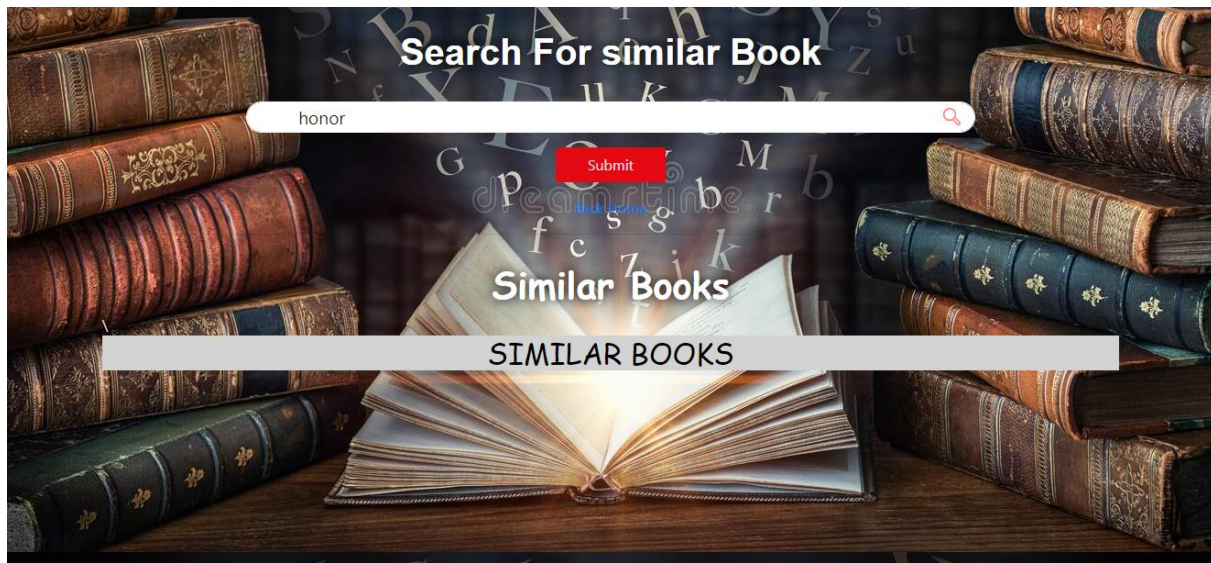
- Harry Potter and the Prisoner of Azkaban (Book 3)**
Director: J. K. Rowling
Votes: 428
Rating: 5.85
- Harry Potter and the Goblet of Fire (Book 4)**
Director: J. K. Rowling
Votes: 387
Rating: 5.82
- Harry Potter and the Sorcerer's Stone (Book 1)**
Director: J. K. Rowling
Votes: 278
Rating: 5.74
- Harry Potter and the Order of the Phoenix (Book 5)**
Director: J. K. Rowling
Votes: 347
Rating: 5.5



Searching for similar books :



Testing example :



7- Conclusion :

En conclusion, les systèmes de recommandation sont des outils importants pour personnaliser l'expérience utilisateur et aider les utilisateurs à trouver des articles, tels que des livres, des produits ou des films, qu'ils pourraient aimer. Il existe plusieurs types de systèmes de recommandation, y compris les systèmes de recommandation basés sur le contenu, de filtrage collaboratif et hybrides, et différents algorithmes peuvent être utilisés pour mettre en œuvre ces systèmes, tels que la similarité cosinus et la factorisation matricielle.

L'apprentissage automatique joue un rôle crucial dans le développement des systèmes de recommandation, car il permet au système d'apprendre du comportement et des préférences des utilisateurs pour fournir des recommandations plus précises et personnalisées. De plus, des technologies telles que Flask peuvent être utilisées pour créer une interface Web pour le système de recommandation, le rendant accessible et facile à utiliser pour les utilisateurs.

L'importance des systèmes de recommandation réside dans leur capacité à aider les utilisateurs à trouver les éléments qui les intéressent, tout en réduisant le temps et les efforts qu'ils doivent consacrer à la recherche de ces éléments. Cela peut entraîner une augmentation de l'engagement et de la satisfaction des utilisateurs, ainsi qu'une augmentation des ventes pour les entreprises qui utilisent des systèmes de recommandation pour promouvoir leurs produits.