

Systèmes de recommandation : Applications aux livres, apports des transformers et impacts écologiques

Cécile Guillot, Ingénieur Machine Learning



"La radio marque les minutes de la vie ; le journal,
les heures ; le livre les jours."

Jacques de Lacretelle



Sommaire



01

Introduction

Définitions, contexte
et objectifs du projet

02

Méthodologie

Données, outils,
algorithmes

03

Résultats

Analyse des données
et des performances

04

Conclusion

Retour sur les objectifs,
perspectives et limites



Introduction

Définitions, contexte et
objectifs

01

Algorithmes de recommandation

- Omniprésent dans notre quotidien
- Netflix, YouTube, Spotify, Deezer, Amazon
- Nombreuses avancées technologiques



Algorithmes de recommandation



Collaborative Filtering

Modèles plus récents :
SVD, NMF, KNN,
OneSlope



Content-Based

Premiers algorithmes à
faire leur apparition



Hybride

Assez récent, nécessite
beaucoup de données
et font appel aux
méthodes de DL



Présentation de l'étude princeps

- Comparaison de 8 algorithmes de recommandation
- Prédiction de films
- Adaptation à nos données

Bhowmick et al. (2021)

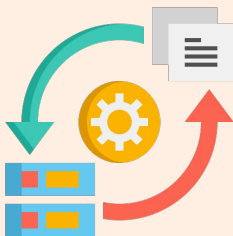


Les objectifs



Book recommender

Application
d'algorithmes de
recommandations
adaptés aux livres



Apport des transformers

Apport des
transformers dans le
content-based



Impact écologique

Mesurer les émissions
de CO₂ en lien avec des
outils du quotidien





02

Méthodologie

Données, outils,
algorithmes

Outils utilisés

arXiv.org

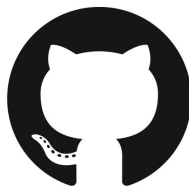
Towards
Data Science

Google Scholar

python

scikit
learn

seaborn



surprise

A Python scikit for
recommender systems.

CODE
CARBON

SBERT.net

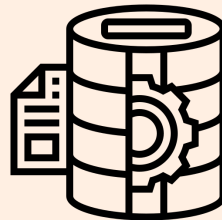
Les données



Book Crossing

Jusqu'à 2004

3 fichiers csv : books, ratings,
users



GoodReads API

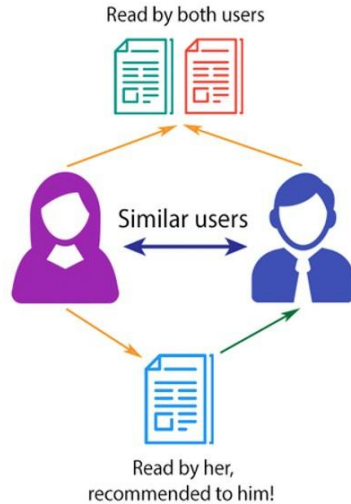
Jusqu'à 2019

20aine de fichiers csv : books et
ratings

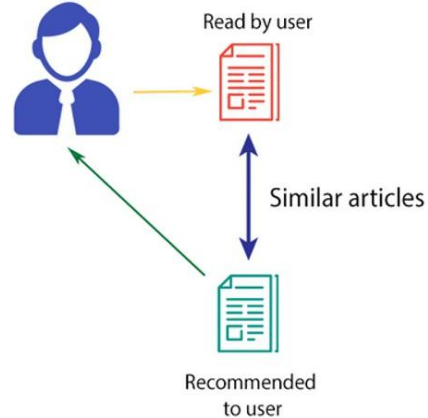
Les algorithmes utilisés



COLLABORATIVE FILTERING



CONTENT-BASED FILTERING



Ranking System

$$W = \frac{Rv + Cm}{v + m}$$

W= Weighted Rating

R = average rating for the movie

v = number of votes for the movie
(currently 3000)

C = the mean vote across the whole report (currently 6.9)

SVD, NMF, KNN, OneSlope,
CoClustering



RMSE

TFIDF vs Transformers
(BERT)

Popularité





Analyses & Résultats

Analyse des données et
des performances



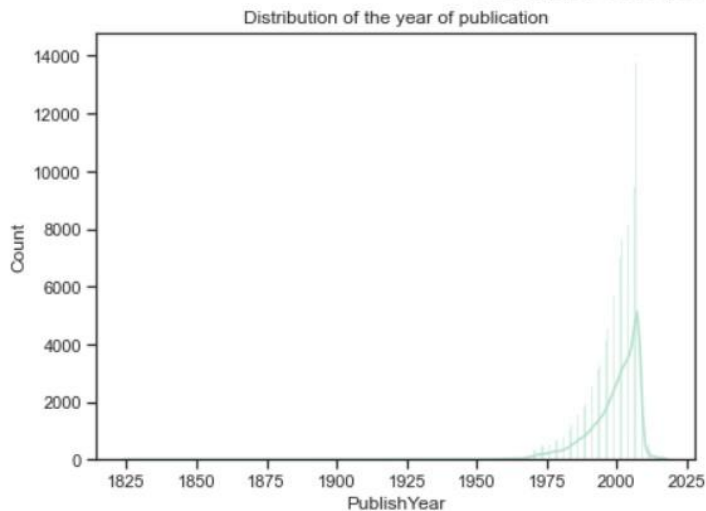
03



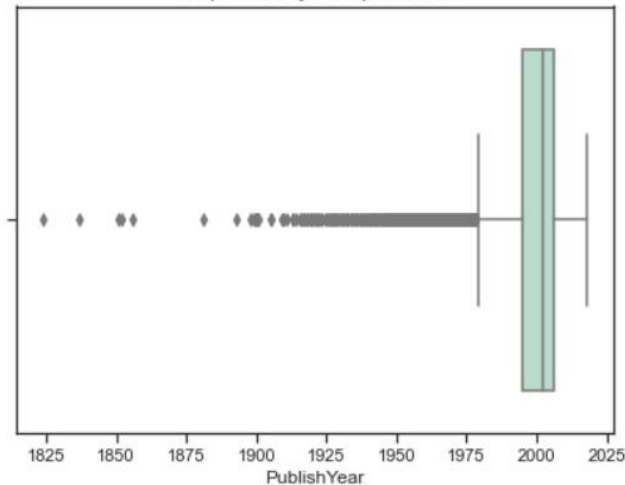
Années de publication:

- 1829 => 2019
- Moy. : 2000

Description of the year of publication of each book



Boxplot of the year of publication

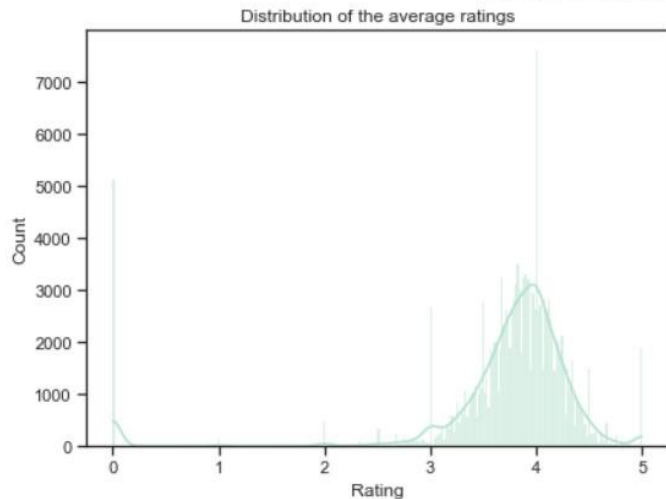




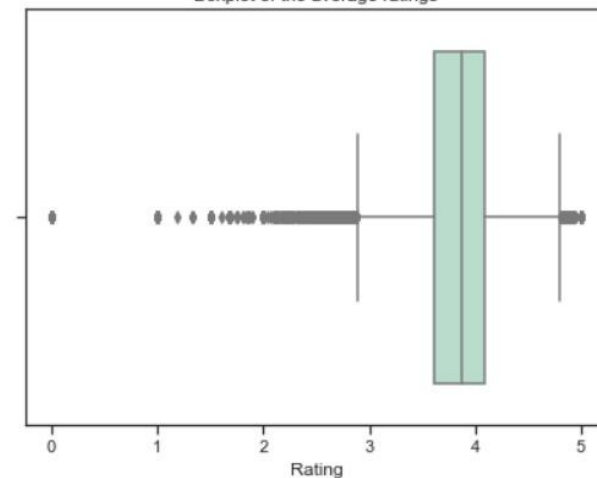
Evaluation :

- De 1 à 5
- Moy. : 4/5

Description of the average ratings of each book

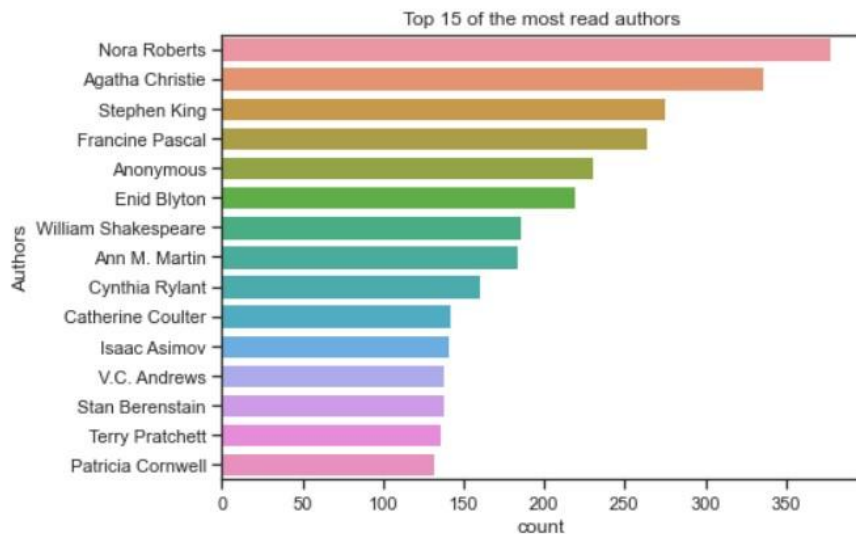
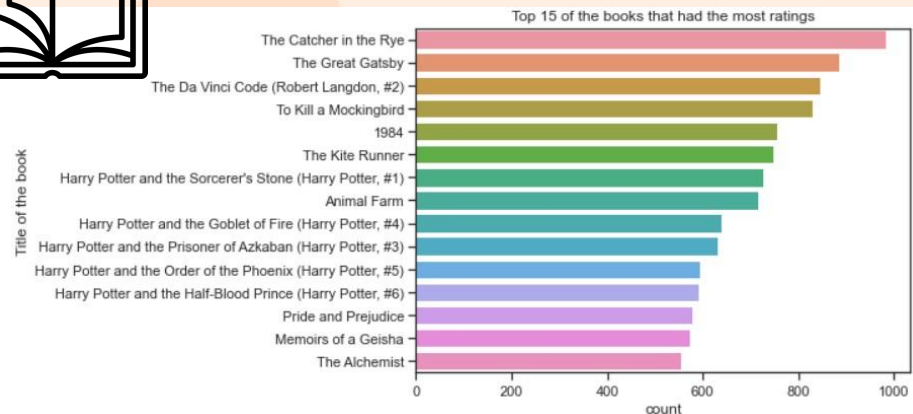


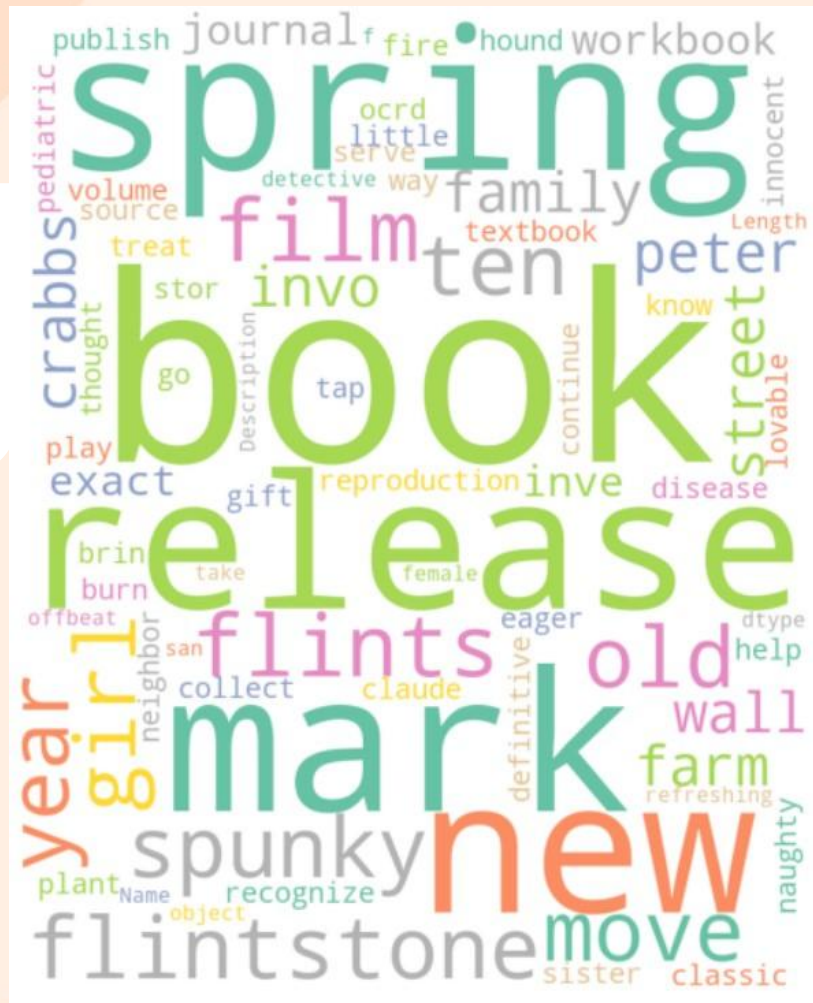
Boxplot of the average ratings







Top 15 des livres et auteurs





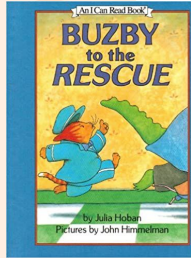
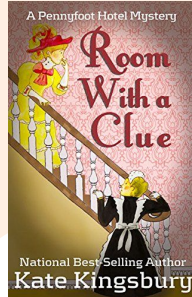
Performances



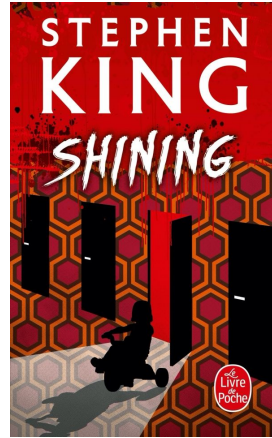
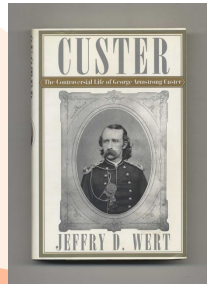
	Popularité	SVD	NMF	KNN	OneSlope	CoClust	TFIDF	BERT
RMSE	0.108	0.765	0.875	0.770	0.891	0.891		
GHG*		0.027	0.029	0.015	0.034	0.015	0.1060	1.454

*g eq. CO2

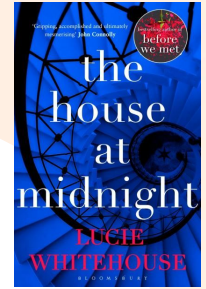
Exemples de prédictions content-based



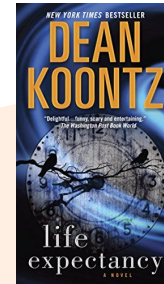
TF-IDF



Livre lu



Transformers







04

Conclusion

Retour sur les objectifs,
perspectives et limites

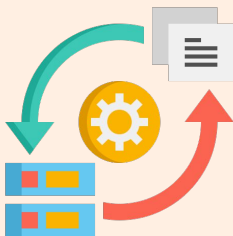


Retour sur les objectifs



Book recommender

Application
d'algorithmes de
recommandations
adaptés aux livres



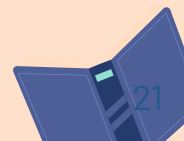
Apport des transformers

Apport des
transformers dans le
content-based

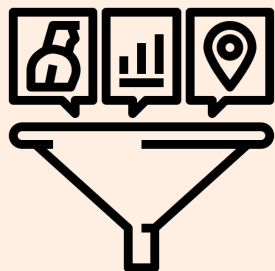


Impact écologique

Mesurer les émissions
de CO2 en lien avec des
outils du quotidien



Perspectives et limites



La collecte des données

- Age, pays, genre
- Gamification ?
- Respect RGPD & CNIL



Les modèles hybrides

- Utilisation Words Embedding
- Trouver les poids à attribuer
- Développer un outil performant et écologique



Merci!

Des questions ?

cecile.gltslmcs@protonmail.com || cecileguillot.com

CREDITS: This presentation template was created by **Slidesgo**, including icons from **Flaticon**, infographics & images by **Freepik**.

Please keep this slide for attribution.