

# Système de Recommandation

Apprentissage chez Junglebike

KOMLAN JEAN-MARIE DANTODJI

Université Paris 8, LIASD

Encadrante : Mme Rakia JAZIRI

Tutrice : Mme Alice Battarel

# Plan

2/24

- 1 Introduction
- 2 Contexte
- 3 Problématique
- 4 État de l'art
- 5 Conclusion

# JungleBike

3/24

- Start Up de B2C.
- Spécialisée dans le secteur du Vélo.
- Mise en relation des clients avec les réparateurs.

# Contexte RH

4/24

- Equipe Data
- Formé de deux data scientists
- Intégration de donnée
- Construction des algorithmes de catégorisation et d'extraction de données.
- Construction des modèles de recommandation

# Contexte technique

5/24

- Outils : Dataliku, DBeaver, .
- Langages et librairies : Python, Scikit Learn, Keras

# Problème

6/24

- Recommandation des produits basée sur les avis des clients.

# Problématique des données

7/24

- L'information sur l'avis client.
- Biais de popularité : non diversité des produits recommandés
- Données issues du scrapping des sites des fournisseurs.
- Manque d'avis clients en comparaison au nombre d'article à recommander

## Données

8/24

	product_id	product_name	brand	user	city	age	activity	level	vote	avis	description
0	0	Pneu Route Continental GP 5000 700 mm Tubetype...	CONTINENTAL	Br74	Anancy	45-54	VTT - XC	Eclairé	3	Très déçu par le poids réel	Connaissant très bien ce pneu car utilisé en 7...
1	0	Pneu Route Continental GP 5000 700 mm Tubetype...	CONTINENTAL	StM21	Dijon	45-54	Route - Cyclosporive	Eclairé	5	Très satisfait	Après plus de 15000 km parcourus avec ces pneu...
2	0	Pneu Route Continental GP 5000 700 mm Tubetype...	CONTINENTAL	boddishiva	barcares	45-54	Route - Cyclosporive	Amateur	2	deçu peut etre un default	j ai acheté ses pneu l an dernier j ai pas par...
3	0	Pneu Route Continental GP 5000 700 mm Tubetype...	CONTINENTAL	Cont2021	None	35-44	Route - Cyclosporive	Eclairé	4	Une fissure après 1200 km	J'avais fait bcp de bornes avec le GP5000 en 2...
4	0	Pneu Route Continental GP 5000 700 mm Tubetype...	CONTINENTAL	Thibj	Strasbourg, France	25-34	Route - Cyclosporive	Eclairé	1	Mauvaise usure	À peine une dizaine de sortie (courte) et l'on...
...	...	...	...	...	...	...	...	...	...	...	...
61819	9221	Chaussures de Trail Femme Asics Gel Sonoma 6 G...	ASICS	None	None	None	None	None	5	Au top !	[Cet avis a été recueilli en réponse à une off...
61820	9221	Chaussures de Trail Femme Asics Gel Sonoma 6 G...	ASICS	None	None	None	None	None	5	Souple et confortable	[Cet avis a été recueilli en réponse à une off...
61821	9221	Chaussures de Trail Femme Asics Gel Sonoma 6 G...	ASICS	None	None	None	None	None	4	Confort	[Cet avis a été recueilli en réponse à une off...
61822	9221	Chaussures de Trail Femme Asics Gel Sonoma 6 G...	ASICS	None	None	None	None	None	4	chaussure confortable	[Cet avis a été recueilli en réponse à une off...
61823	9221	Chaussures de Trail Femme Asics Gel Sonoma 6 G...	ASICS	None	None	None	None	None	5	Asics gel sonoma trade 6 G TX	[Cet avis a été recueilli en réponse à une off...

61824 rows x 11 columns

FIG. 1 : Jeu de donnée



# Détail des colonnes

9/24

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 61824 entries, 0 to 61823
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   product_id            61824 non-null  int64
1   product_name          61818 non-null  object
2   brand                 58570 non-null  object
3   user                  52904 non-null  object
4   city                  47642 non-null  object
5   age                   52554 non-null  object
6   activity              52554 non-null  object
7   level                 52554 non-null  object
8   vote                  54624 non-null  object
9   avis                  54624 non-null  object
10  description            54624 non-null  object
dtypes: int64(1), object(10)
memory usage: 5.2+ MB
```

FIG. 2 : Détail des colonnes

# Méthodes élémentaires

10/24

- Recommandation aléatoire

# Recommandation Objet

11/24

- Caractéristiques des produits
- Projection des produits dans un repere

# Methode du Cosin Similarity

12/24

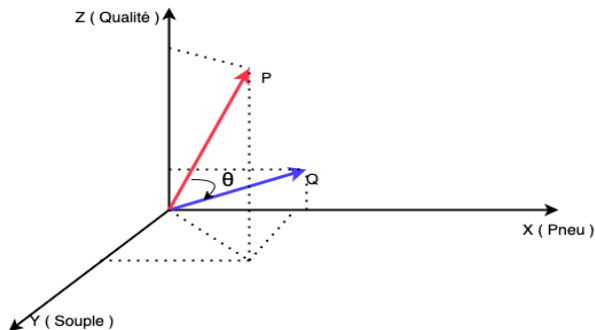


FIG. 3 : Projection des produits

# La Matrice de Factorisation

13/24

- Décomposer la matrice de votes en deux

$$M = U \times I$$

# Décomposition

14/24

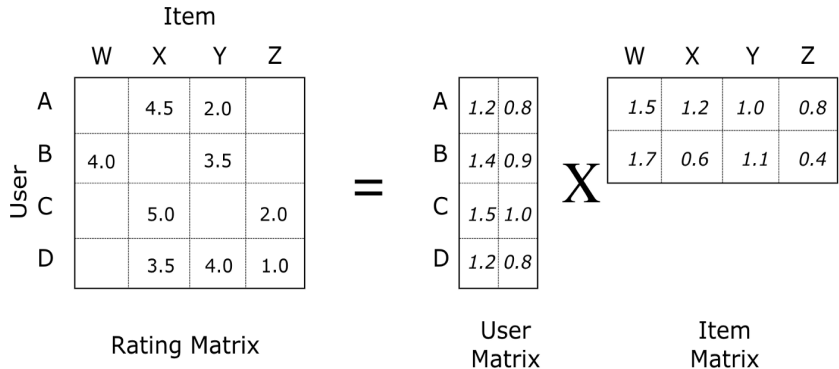


FIG. 4 : Décomposition de la matrice

# Matrice similaire

15/24

3.16	1.92	2.08	1.28
3.63	2.22	2.39	1.48
3.95	2.4	2.6	1.6
3.16	1.92	2.08	1.28

FIG. 5 : Décomposition de la matrice

# Neural Collaborative Filtering (NCF)

16/24

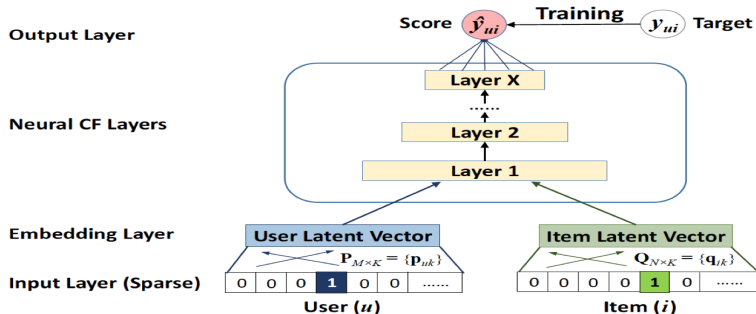


FIG. 6 : Neural Colaborative Filtering,

<https://towardsdatascience.com/neural-collaborative-filtering-96cef1009401>



# Généralisation du NFC

17/24

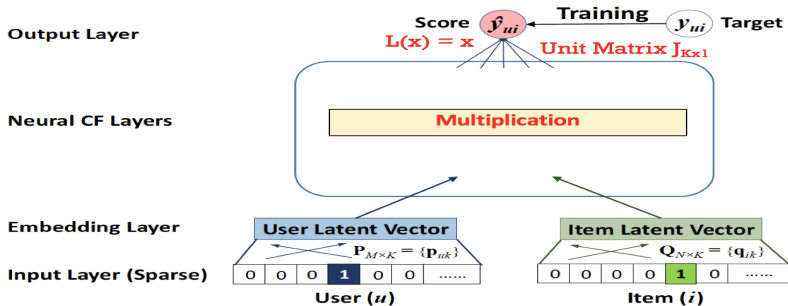


FIG. 7 : Généralisation du NFC,

<https://towardsdatascience.com/neural-collaborative-filtering-96cef1009401>

# Conditions de généralisation

18/24

- initialiser le poids de la couche de sortie à une matrice  $J$  dont toutes les valeurs sont égales à 1.
- Considérer une fonction d'activation  $L$  linéaire :

$$L(x) = x$$

# Généralisation du NFC

19/24

$$\hat{y}_{ui} = L(p_u \odot q_i \times J_{K \times 1})$$

$$\hat{y}_{ui} = L(p_u^T \cdot q_i)$$

$$\hat{y}_{ui} = p_u^T \cdot q_i$$

FIG. 8 : Généralisation du NFC

# Combinaison des méthodes

20/24

- Combinaison du Collaborative Filtering au Content Based.

# Conclusion

21/24

- La méthode de Deep Learning généralise au mieux les autres.

# Références

22/24

- ▶ Sumit Sidana. Recommendation systems for online advertising. Computers and Society [cs.CY]. Université Grenoble Alpes, 2018. English. ffNNT : 2018GREAM061ff. fftel-02060436ff
- ▶ D Gunawan et al. The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents. 2018 J. Phys. : Conf. Ser. 978 012120

# Références

23/24

- ▶ Chakrabarti S, van den Berg M, Dom B 1999 Focused crawling : a new approach to topic-specific Web resource discovery Comput. Networks 31 11–16 pp 1623–1640

Merci pour votre attention