



Système de recommandation de musiques

via les données de Twitter



Présentation du projet

Description : Prédire si une musique sera plus ou moins appréciée.

prédictions
popularité
dictionnaire de sentiments
régression Linéaire ?



Titre : Recommandations musicales par rapport aux similitudes entre morceaux.

recommandations

clustering

hashtag
vectorisation

Analyse des datasets

Jeux de données découpé en **3 datasets** :

<https://www.kaggle.com/datasets/chelseapower/nowplayingrs>

Context-Content-Features

Information sur les musiques,
les caractéristiques musicales,
le lieu d'écoute...

(22 variables, > 11 millions d'entrées, 2.19Go)

User-Track

Information sur le champ
lexical (hashtags) associé aux
morceaux par les utilisateurs...

(4 variables, > 17 millions d'entrées, 1.22Go)

Sentiments (dictionnaires)

Information reliant les hashtags
à un score de sentiment.

(21 variables, 5290 d'entrées, 373Ko)



Analyse des datasets

Un même morceau peut être renseigné par un même utilisateur :

- au minimum : 1 fois
- au maximum : 50 383 fois
- en moyenne : 5.8 fois

Plus de 5000 hashtags différents !

- certains peu informatifs
- répétitions d'un même utilisateur



Gros travail de nettoyage nécessaire.

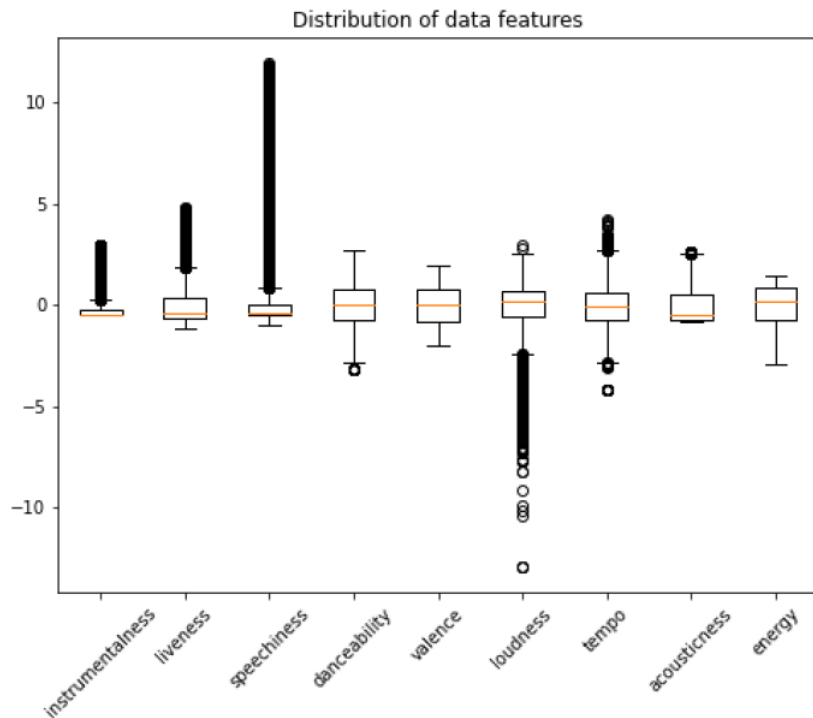


MODELES DE RECOMMANDATION



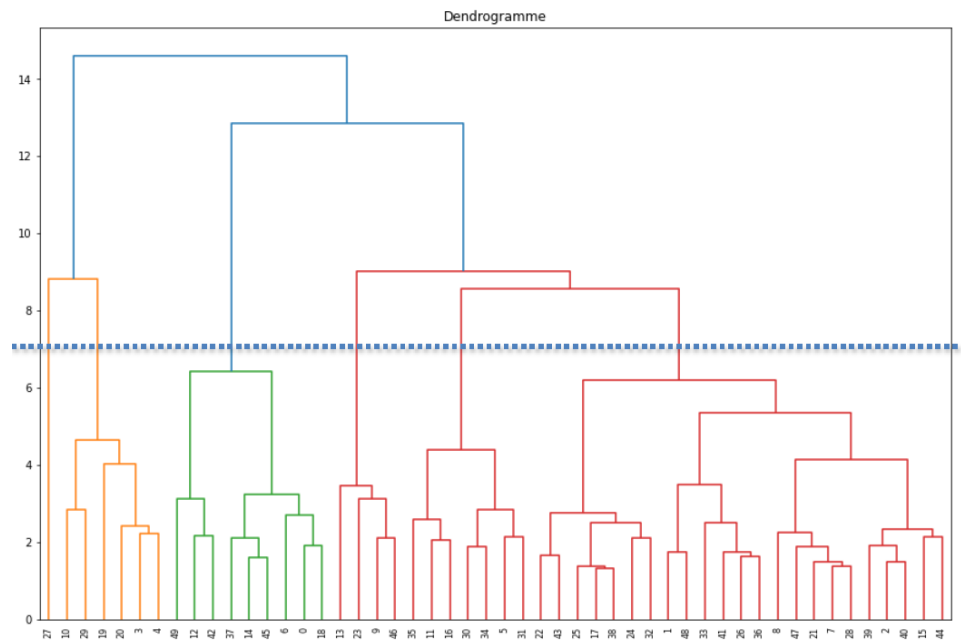
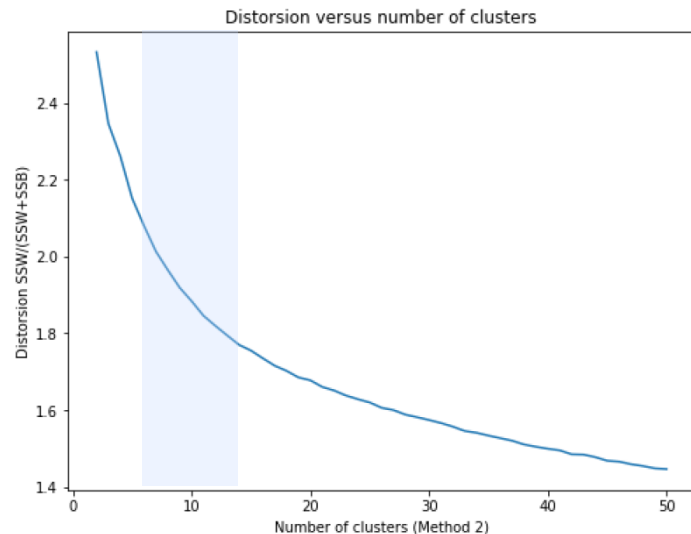
Modèle n°1

Clustering sur les caractéristiques musicales



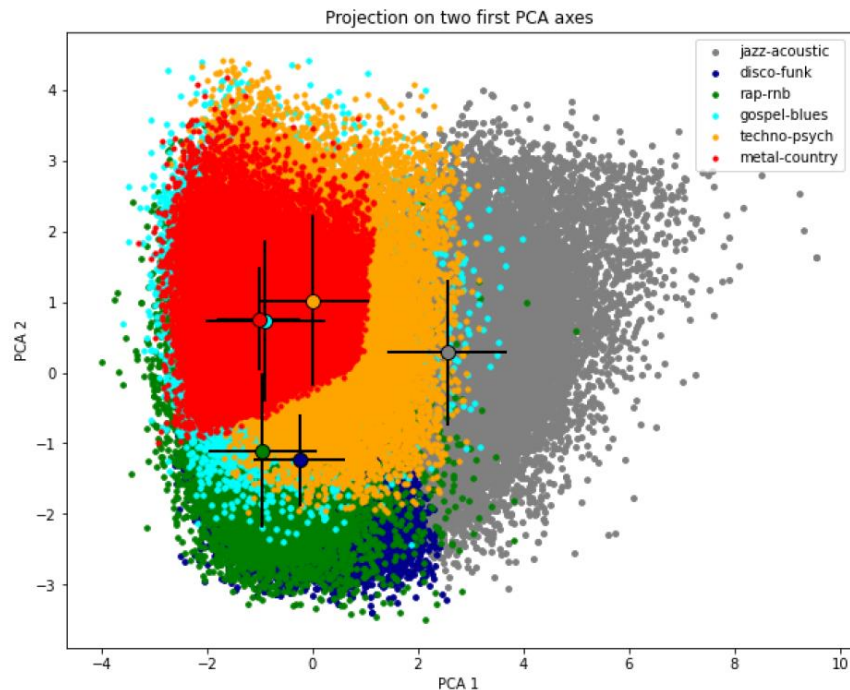
Modèle n°1

Clustering sur les caractéristiques musicales



Modèle n°1

Clustering sur les caractéristiques musicales



Group 0 : jazz-acoustic

acoustic
jazz lounge
indie gospel

Group 3 : gospel-blues

punk metal
gospel pop
latin

Group 1 : disco-funk

funk reggae
latin soul
disco

Group 4 : techno-psych

techno
psych triphop
house electronic

Group 2 : rap-rnb

rap classic
reggae hiphop rnb

Group 5 : metal-country

acoustic country
metal
rock punk

Modèle n°1 : Clustering sur les caractéristiques musicales

ERREURS DU MODELE

	track_id	hashtag	labels
120	67eeca3f067d0496881f4d283f0f626c	acoustic	5
121	0f6c3e42bc32d33710b9e066a14d6374	acoustic	1
123	54d0a6d0c80ffc3856c2729f63363a7b	acoustic	5
125	7d5f7749b21900870377d92d2c8ae0d9	acoustic	5
126	6a5aad35ef90a548ccf1fe7e6daf774a	acoustic	5

groupe 5 : metal-country

morceau de référence : d967db99cb358341ec06287b14c485c6
hashtag : ['metal']
label : [4 : techno-psych]

	track_id	hashtag	labels
49	3bb38331d88e6b6dfe627e15db190075	acoustic	4
57	d5fc0e7dd33205244c97bbf13a7fa822	acoustic	4
95	a5452c50af87568b010061c1f0e49d3a	acoustic	4
130	47ba1f3acab95e3cf8d30843abb5032c	alternative	4
138	e1ddf8e992f65333094683896072f95	alternative	4

morceau de référence : 844cd2570ddf0729bde295219cb25514
hashtag : ['pop']
label : [0 : jazz-acoustic]

	track_id	hashtag	labels
9	ee67b313db5a7fb1fdff43cc7f305e72	acoustic	0
10	bed12e9de9a2904aa7ee92088bed4096	acoustic	0
17	47919f438a3492115511f89ff25ac059	acoustic	0
19	47919f438a3492115511f89ff25ac059	indie	0
21	dd0657bc29856e0af8f4bbcfad8e0f7c	acoustic	0

Modèle n°2 basé sur les hashtags renseignés par les utilisateurs

track_id	hashtag
0000e47c1207e2c637a44753a713456f	rock
0000e47c1207e2c637a44753a713456f	pops
0000e47c1207e2c637a44753a713456f	70s
0000e47c1207e2c637a44753a713456f	music
0000e47c1207e2c637a44753a713456f	music

track_id	hashtag
0000e47c1207e2c637a44753a713456f	rock pops 70s music

Hashtags sans intérêt :

'listenlive', 'class95', 'kiss92', 'music', '6music', '234radio', 'jammin105', 'radio2', 'audio5cafe', 'radio', 'bbc6music', 'tweetlink', 'myplaylist', 'tbfmonline', 'vkscrobbler', 'wzbt', 'nowlistening', 'spotify', 'onlineradio', 'tunein', ...

track_id	hashtag	hashtag_cleaned
0000e47c1207e2c637a44753a713456f	rock pops 70s music	rock pops 70s

Modèle n°2

basé sur les hashtags renseignés par les utilisateurs

TF-IDF

TF-IDF is a measure of originality of a word by comparing the number of times a word appears in a doc with the number of docs the word appears in.

$$\text{TF-IDF} = \text{TF}(t, d) \times \text{IDF}(t)$$

Term frequency

Number of times term t appears in a doc, d

Inverse document frequency

$\log \frac{1 + n}{1 + \text{df}(d, t)}$

n ← # of documents

Document frequency of the term t


TF-IDF + linear_kernel

pour identifier les similarités entre les champs de hashtags

Modèle n°2 basé sur les hashtags renseignés par les utilisateurs

```
morceau : [ 83847 ] : track_id = c0a3db38bd9f8f8c44ed0c2d0e1b60af  
hashtags: rock pops 70s
```

résultats:



	idx	track_id	hashtags
0	1	0000e47c1207e2c637a44753a713456f	rock pops 70s
1	83847	c0a3db38bd9f8f8c44ed0c2d0e1b60af	rock pops 70s
2	30054	45138db5d3e84a423b12650a0f65d4a6	rock 70s pops
3	44206	65929a3541868bf88b85bdc16e86e1d0	rock pops 60s
4	24288	37bc24b5c16cc1f91a1049d6ebed410b	pops 00s
5	71372	a4196978d57dd77ce85628bc86cb27d1	rock 70s
6	60588	8b4146285e3ca27693c3b1cf81851e1c	rock hardrock 70s
7	43678	646a1e0d1b8ceb54ce87f017d4357071	rock hardrock 70s
8	1786	03e508ca8989f648f305283100388b6c	rock hardrock 70s
9	102600	ec3ac9eb8e3df884700b76c70c252b5f	rock hardrock 70s

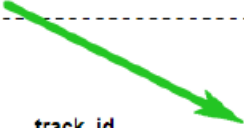
Modèle n°2 basé sur les hashtags renseignés par les utilisateurs

ERREURS DU MODELE

morceau : [104841] : track_id = f16e9029afe5e148d2dffcf88a55603b

hashtags: hiltonhead savannah progressiveradio toulouse sanjose phoenix word beatbc ontheair winnipeg belfast brunswick freshly squeezed yourmusic aracityradio rock itunes metal bringinitback manchester rheims orleans kamloops pittsburgh anchorage takeover rtext 2002年の音楽 radiocidadeoficial tocandonacidade fuerzamartes cancióndeviernes cancióneldía musicislife nanaimo liverpool köln lille tacoma fragradio gaming foofighters turnitup turnup rockon tocando radio98rock brazil atacalaradio allmylife tsgonai r rcbs

résultats:



	idx	track_id	hashtags
0	104841	f16e9029afe5e148d2dffcf88a55603b	hiltonhead savannah progressiveradio toulouse ...
1	35674	51f31c2eb4aa9fbcb9f6ddc6f716416	savannah hiltonhead ontheair belfast brunswick...
2	65829	9774a311af1b0e64e092e50f4a35ec2a	worldclassmusic savannah hiltonhead auckland b...
3	108216	f92726d6d470a7f911257132863bdf5c	progressiveradio savannah hiltonhead orleans l...
4	83168	bf1468972cc9e5360fb681d9c6bcda5	progressiveradio savannah hiltonhead buffalo o...



QUESTIONS / REPOSSES





MERCI