



UNIVERSITÉ PARIS 8 - VINCENNES À SAINT-DENIS

Licence informatique & vidéoludisme

**Rapport - Ingénierie des langues
Le scribe de la mort**

Bastien Capiaux

Date de rendu : le 16/04/2024

Contents

0.1	Résumé	2
1	Scraper	3
1.1	scraper_url	3
1.2	scraper_avis	4
2	IA	5
2.1	prediction.py	5

0.1 Résumé

Ce projet d'ingénierie des langues, développé dans le cadre de la Licence Informatique et Vidéoludisme à l'Université Paris 8 - Vincennes à Saint-Denis, vise à extraire des données à partir d'un site en particulier que j'ai choisi : Jeuxvideo.com (JVC).

Ces données incluent notamment les titres des articles, les avis, les auteurs et les notes. L'objectif principal est de recueillir des avis sur les jeux populaires de jeuxvideo.com afin de constituer une base de données.

Cette base de données sera ensuite utilisée afin d'évaluer d'autres avis et déterminer s'ils sont positifs ou négatifs.

Chapter 1

Scraper

Dans notre projet, nous avons deux scrapers différents, tous deux présents dans le dossier `src`.

1.1 `scraper_url`

Le premier, `scraper_url.py`, nous sert à récupérer les liens des pages d'avis sur la page populaire de JVC. Pour utiliser le programme et stocker les informations, vous devez exécuter la commande suivante :

```
python3 src/scraper_url.py > database/url.txt
```

Cela créera un fichier contenant les URL des derniers jeux populaires de JVC.

```
database > url.txt
1 https://www.jeuxvideo.com/jeux/mac/jeu-1299145/avis/note/
2 https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/
3 https://www.jeuxvideo.com/jeux/ps5/jeu-1716397/avis/note/
4 https://www.jeuxvideo.com/jeux/xbox-series/jeu-1791483/avis/note/
5 https://www.jeuxvideo.com/jeux/stadia/jeu-1164242/avis/note/
6 https://www.jeuxvideo.com/jeux/xbox-series/jeu-1590768/avis/note/
7 https://www.jeuxvideo.com/jeux/pc/jeu-1590761/avis/note/
8 https://www.jeuxvideo.com/jeux/ps5/jeu-1590767/avis/note/
9 https://www.jeuxvideo.com/jeux/pc/jeu-1417171/avis/note/
10 https://www.jeuxvideo.com/jeux/xbox-series/jeu-1843411/avis/note/
11 https://www.jeuxvideo.com/jeux/one/jeu-1843412/avis/note/
12 https://www.jeuxvideo.com/jeux/ps5/jeu-1752701/avis/note/
13 https://www.jeuxvideo.com/jeux/pc/jeu-1752699/avis/note/
14 https://www.jeuxvideo.com/jeux/pc/jeu-1567915/avis/note/
15 https://www.jeuxvideo.com/jeux/ps5/jeu-1567916/avis/note/
16 https://www.jeuxvideo.com/jeux/xbox-series/jeu-1567919/avis/note/
17 https://www.jeuxvideo.com/jeux/switch/jeu-1567918/avis/note/
18 https://www.jeuxvideo.com/jeux/ps4/jeu-1567917/avis/note/
19 https://www.jeuxvideo.com/jeux/one/jeu-1567920/avis/note/
20 https://www.jeuxvideo.com/jeux/mac/jeu-1711003/avis/note/
21 https://www.jeuxvideo.com/jeux/ps5/jeu-1591523/avis/note/
22 https://www.jeuxvideo.com/jeux/pc/jeu-670687/avis/note/
23 https://www.jeuxvideo.com/jeux/xbox-series/jeu-1591524/avis/note/
```

1.2 scraper_avis

Le second, `scraper_avis.py`, nous sert à récupérer les informations contenues sur les pages d'avis contenues dans `database/url.txt` sur les pages populaires de JVC. Pour utiliser le programme et stocker les informations dont nous aurons besoin par la suite.

Vous devrez avoir le fichier :

```
"database/url.txt"
```

Vous pouvez exécuter la commande suivante :

```
python3 src/scraper_avis.py
```

Cela créera un fichier du nom de "jeux_avis_populaire.csv" contenant le Nom, l'URL, la Note, le pseudo et l'avis dans un CSV. Afin de ne pas récupérer une page en double, une vérification par pseudo est réalisée. Cela signifie que si la page précédente contient les mêmes pseudos que la page actuelle, alors cela signifie que nous avons récupéré tous les avis du jeu et choisi une autre URL si disponible.

Attention, cette opération est coûteuse en temps !!!

1	Titre	URL	Note	Pseudo	Avis
2	Baldur's Gate III	https://www.jeuxvideo.com/jeux/mac/jeu-1299145/avis/note/?p=1	20/20		"EA qui servira à faire un jeux nickel pou
3	Baldur's Gate III	https://www.jeuxvideo.com/jeux/mac/jeu-1299145/avis/note/?p=1	20/20		"Vivement que la version finale soit fin p
4	Baldur's Gate III	https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Deux heures après le démarrage de l'EA d
5		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Une note relativement élevée afin de, je
6		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Mon avis s'adresse au hater, le fais pari
7		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Un 20 pour compensé les note injustifiée
8		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Un très bon jeu. Les graphismes sont bien
9		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"J'ai décidé de profiter de l'early access
10		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Deux soles, quand on aime le genre, les
11		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Je sais c'est nul de mettre 20/20, mais f
12		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Magnifique, juste frustrée que ce ne soit
13		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Quelle idée absurde de noter un Early Acc
14		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Le Masterclass des rpg depuis sa sortie e
15		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"De part son gameplay viscéral qui complète
16		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Jeu tout simplement exceptionnel... Je ne
17		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Ecriture stupéfiement riche, perso stupide
18	Baldur's Gate III	https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=1	20/20		"Tout simplement le meilleur jeu auquel j'
19		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Au ou du nombre d'heures de jeu que j'y p
20		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Si ce jeu n'est pas le GOTY ce sera unqu
21		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"On peut difficilement faire mieux que ce r
22		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Un très bon jeu. Les graphismes sont bien
23		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Ecriture stupéfiement riche, perso stupide
24		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Une valeur sûre du RPG, riche, généreux,
25		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Le meilleur jeu de tous les temps, de mon
26		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Vous voulez du vrai Zelda mature? en voil
27		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Jamais un RPG n'aura fait autant parler
28		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Après 20ans d'attente on voilà de retour
29		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Un RPG incroyable qui établit une norme d
30		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Baldur's Gate 3 est un jeu de rôle épique
31		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Baldur's Gate 3 est un chef-d'œuvre du RP
32		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Si je peux être d'accord avec la plupart
33		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=2	20/20		"Expérience incroyable à consommer sans mo
34	Baldur's Gate III	https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Le dieu du RPG IIIIES COMBATS, une nouvel
35		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Le jeu est tout bonnement incroyable, pro
36		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"C'est clairement le GOTY pour moi.Seuils d
37		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Baldur's Gate 3 s'établit solidement par
38		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"C'est dommage d'avoir laissé les notes d
39		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Jeu absolument incroyable si on aime les R
40		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"C'est un parfait sans fautes pour Larian
41		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Une écriture parfaite. Une quête principal
42		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Univers d'une richesse, d'une narration,
43		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"On s'en fiche des notes, j'habite pas un
44		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Ce jeu est pour moi une œuvre réalisée sa
45		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Immense, intense, et incroyablement bien
46		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Comment on peut mettre 0 ou 1, sérieux...
47		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Jeu de l'année, jeu de l'année, jeu de l'ann
48		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=3	20/20		"Jeu exceptionnel, à faire avec un ami ou
49	Baldur's Gate III	https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=4	20/20		"J'ai été tout de suite immergé par la mis
50		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=4	20/20		"Aucun souci notoire pour ma part. Riche,
51		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=4	20/20		"LE jeu de l'année, beau, accessible même
52		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=4	20/20		"L'histoire principale se déploie avec une
53		https://www.jeuxvideo.com/jeux/pc/jeu-5571/avis/note/?p=4	20/20		"Meilleur RPG de la décennie ! Un très bon

Chapter 2

IA

Notre IA est basée sur la vectorisation Tf-idf de nos avis avec comme modèle une régression logistique car nous avons 2 valeurs à prévoir : positif ou négatif.

2.1 prediction.py

Notre programme va récupérer les notes et donner un label "+" si la note est supérieure à 15 et si la note est en dessous de 5, alors le label sera "-". Tout cela sera mis dans "database/label_avis.csv".

Vous devrez avoir le fichier :

```
"database/jeux_avis_populaire.csv"
```

Vous pouvez exécuter la commande suivante afin de lancer le programme :

```
python3 src/prediction.py
```

Notre programme vas succesivement créer une base de données nettoyer de tout bruit parasite, vectoriser a l'aide de Tfidf notre jeu de donnée et l'entrainer sur notre Regression logistique.

```

database > label_avis.csv > data
1 Text ,Avis
2 + , "ça qui servira à faire un jeux nickel pour sa sortie lon est de suite dans l'ambiance d&d,pour le moment, c'est une bonne surprise... vivement la version gold !"
3 + , "vivement que la version finale soit fin prête ! ce jeu déchire, tout est de qualité, et j'adore les cinématique, qui me font penser à dragon age origin. vivement
4 + , "deux heures après le démarrage de l'ea du jeu, certains crient déjà au scandale... "oui je veux avoir 6 persos dans ma team, je peux pas donc ça vaut 0/20...""
5 + , "une note relativement élevée afin de , je l'avoue, contre balancer des notes négativement abusées. ok les puristes bg n'y trouvent peut-être pas leur compte, mais
6 + , "mon avis s'adresse au haters.je fais partie de ces joueurs de la première heure de bg. les deux bg ont été une révélation pour moi et dans mon panthéon, ils trône
7 + , "un 20 pour compenser les note injustifiées des méchants.le jeu est un early access .travail impeccable, toutes les promesses y sont. il est évident qu'il y a un l
8 + , "un très bon jeu. les graphismes sont bien travaillés et j'adore les cinématiques qui nous propulsent dans l'histoire. le choix des dialogues aussi n'est pas s'e
9 + , "j'ai décidé de profiter de l'early access pour me créer un personnage pas trop gentil W. hier je suis tombé sur un couple de monstre qui était en plein déb.
10 + , "jeux sublime, quand on aime le genre, les musiques sont folle, le jeu est beau et bien fournis, il y as de quoi faire, dépasse la durée de vie de bon nombres de j
11 + , "je sais c'est nul de mettre 20/20, mais faut pas déconner là... il y a trop de notes sanctions totalement injustifiées.en l'état, de ce que je vois, pour moi c'es
12 + , "magnifique, juste frustrée que ce ne soit que l'early access, hâte d'être à la sortie officielle avec tout le contenu et les succès ! une vraie réussite !"
13 + , "quelle idée absurde de noter un early access... en tenant compte de cet aspect fondamental, je donne un 20/20.en effet hormis quelques soucis de stabilité, quelq
14 + , "la masterclass des rpg depuis sa sortie on en a, tous les joueurs de jdr devraient apprécier cette qualité assez rare de nos jours, des centaines d'heures à remodel
15 + , "de part son gameplay viscéral qui complète à merveille son level design organique, les petits gars de chez larian studio nous prouvent encore une fois qu'ils en on
16 + , "jeu tout simplement exceptionnel... je ne suis pas déçu d'avoir attendu aussi longtemps ! une véritable pépite ! merci larian studio, j'en attendais pas moins de
17 + , "écriture stupéfiement riche, perso stupéfiement charismatiques, système un poil simplifié par rapport au jeu sur table mais seulement sur les parties inutilement co
18 + , "tout simplement le meilleur jeu auquel j'ai pu jouer. et quelle bouffée d'air frais de pas avoir de battle pass, de dic, d'abonnement mensuel, de boutique bidule
19 + , "au vu du nombre d'heures de jeu que j'y passe comment mettre moins tout est dans ce jeu, je ne comprends pas du tout la note moyenne si faible...le jeu est beau,
20 + , "si ce jeu n'est pas le gty ce sera uniquement à cause de sa relative difficulté d'accès et le public peut-être relativement restreint auquel il s'adresse.pour l
21 + , "on peut difficilement faire mieux que ce nouvel opus d'une densité assez incroyable. dommage que les éternels chômeurs du site spamment encore de faux avis pour fa

```

Aperçu de cette nouvelle base de données nettoyée.

	precision	recall	f1-score	support
+	0.87	0.96	0.91	231
-	0.93	0.77	0.84	145
accuracy			0.89	376
macro avg	0.90	0.86	0.87	376
weighted avg	0.89	0.89	0.88	376

	Avis	Sentiment
0	J'ai adoré ce jeu, les graphismes sont superbes.	+
1	Je recommande ce jeu à tous mes amis, c'est un...	+
2	Les développeurs ont vraiment fait du bon trav...	+
3	Graphismes décevants, le jeu est daté.	-
4	Évitez ce jeu à tout prix, c'est une véritable...	-

Pourcentage de réussite : 100.0%

Aperçu de la precision de nos réponse.

Puis nous amuser à le tester par nous-même avec une petite interface utilisateur dans le terminal où nous pourrons écrire nos propres avis.

```
Entrez votre avis sur un jeu vidéo (ou tapez 'q' ou 'quit' pour quitter) :  
Ce jeu me paraît plus que convenable pour le prix.  
Entrez le sentiment associé à votre avis (positif ou négatif) :  
positif  
Le sentiment prédit pour votre avis est : positif  
La prédiction a réussi !
```

Sentiment utilisateur positif. (vraie)

```
Entrez votre avis sur un jeu vidéo (ou tapez 'q' ou 'quit' pour quitter) :  
Ce jeu me paraît plus que convenable pour le prix.  
Entrez le sentiment associé à votre avis (positif ou négatif) :  
négatif  
Le sentiment prédit pour votre avis est : positif  
La prédiction a échoué...
```

Sentiment utilisateur négatif. (faux)

Comme nous pouvons le voir "Ce jeu me paraît plus que convenable pour le prix." a une connotation positive et notre modèle est capable de trouver le bon label.