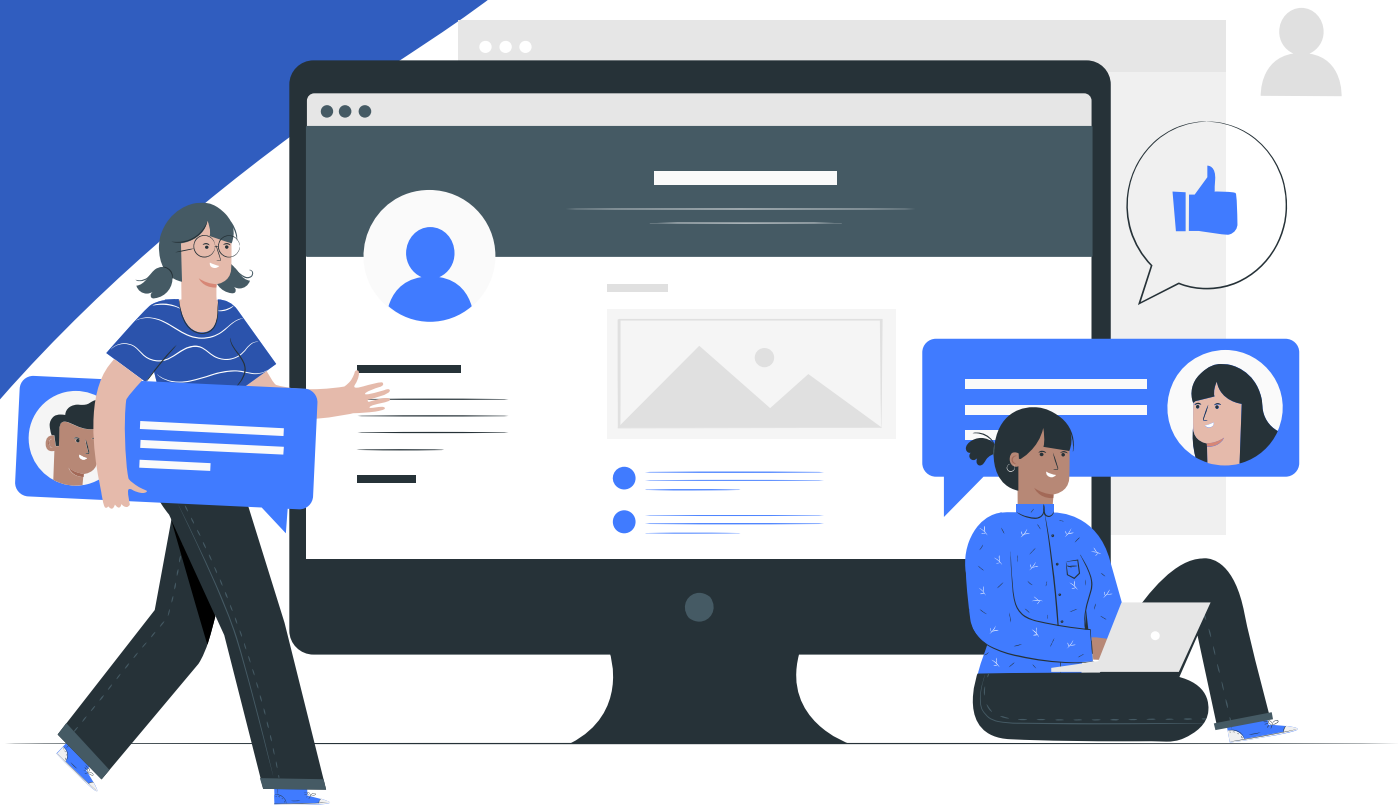
sentiments de Breaking Bad Movie Reviews

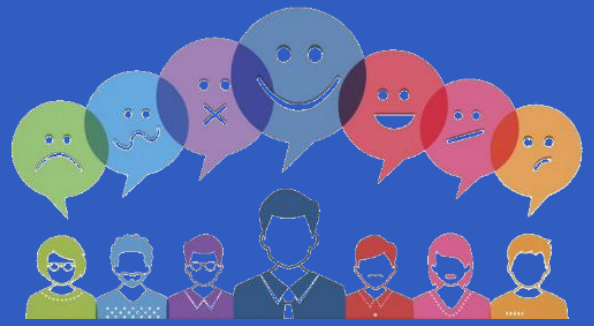
Fouille de données massive



Team Members

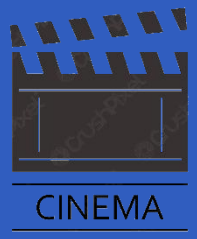
Khammeri Med Nour

Objectifs



Objectif

Scrapping de quantité massives de données en ligne de plusieurs Sources afin obtenir une grande Data Set contenant: review, rate ensuite faire les optimiser en faisant des prétraitements convenables selon la nature de donnes, Finalement utiliser ces donnés dans un modelé machine Learning pour faire analyser les reviews: positive ,négatif , neutre.

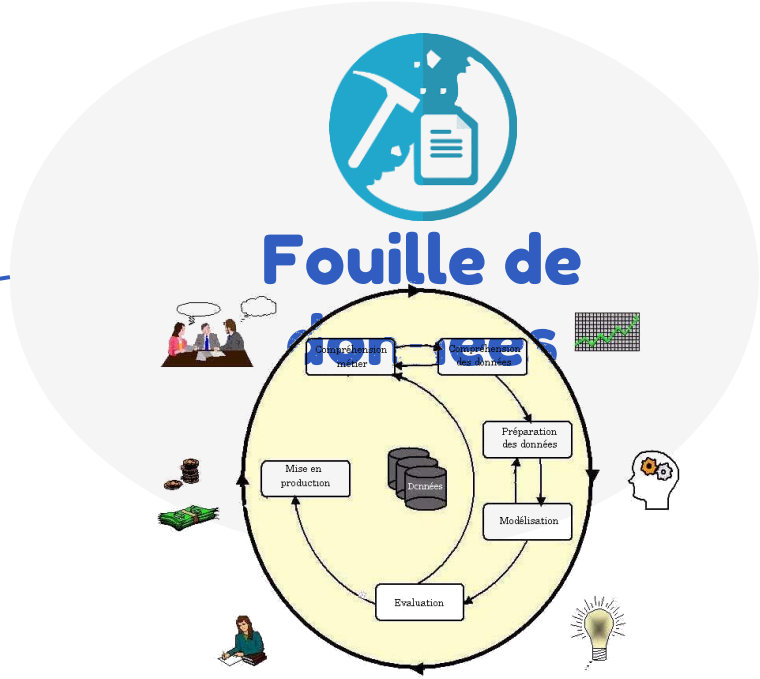


Introduction

Introduction

(1)

(2)



Scrapping Data

Premier Source de Scrapping: Amazon.com

Customer reviews

4.7 out of 5

4,929 global ratings

5 star		85%
4 star		7%
3 star		4%
2 star		1%
1 star		3%

How are ratings calculated?

Breaking Bad Season 1

by Vince Gilligan

Write a review

Top positive review

All positive reviews ›

Bare Bones

THE BARE BONES REVIEW

Reviewed in the United States on July 22, 2019

This item arrived quickly Via Amazon prime.

The picture quality of this Blu-ray is on par. I have personally read over the years that the picture quality for season one on Blu-ray was substandard. I have not found that to be the case. There is plenty of fine detail to be seen on the show in season one. It is a huge leap in terms of visual and audio quality over the DVD.

Top critical review

All critical reviews ›

Kindle Customer

Terrible Blu Ray commercials play all the time

Reviewed in the United States on December 24, 2014

While I am enjoying the show, Sony apparently feels that its paying customers must endure a 5 minute bluray commercial EVERY TIME you try and play the thing. And disable the menu so you can't get right in and watch. The loading time is about 3 or 4 minutes even before the commercial starts. And I am using a recent sony player. Absolutely horrible. Buy a DVD and suffer a little less quality for more speed!

7 people found this helpful

Code Python

```python
import scrapy
from bs4 import BeautifulSoup


class Test(scrapy.Spider):
    name = 'amazon'
    start_urls = ['https://www.amazon.com/product-reviews/B0012QRPU4/ref=atv_dp_cr_see_all?ie=UTF8&reviewerType=all_reviews']

    def parse(self, response):
        for products in response.css('div.a-section.a-spacing-none.review-views.celwidget div.a-section.review.aok-relative'):
            yield{
                'reviews': BeautifulSoup(products.css('span.a-size-base.review-text.review-text-content span').get(), 'html.parser').get_text(),
                'notes': products.css('span.a-icon-alt::text').get()[0],

            }

        next_page = response.css('ul.a-pagination li.a-last a').attrib['href']

        if next_page is not None:
            yield response.follow(next_page, callback=self.parse)
```

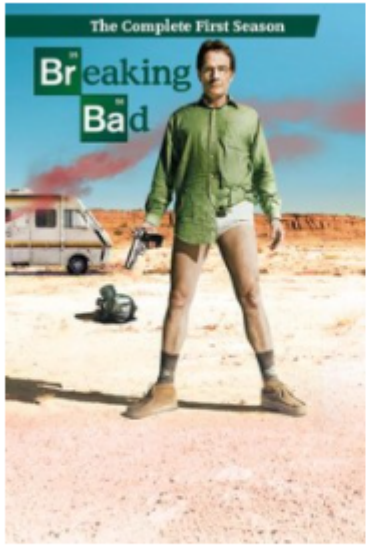
Directory Fichier source

Name	Date modified	Type	Size
breakingbad	11/30/2021 2:28 PM	File folder	
corpus.csv	11/30/2021 2:40 PM	Microsoft Excel C...	943 KB
scrapy.cfg	11/30/2021 2:25 PM	CFG File	1 KB

C:\Users\mednour\Desktop\Scrapp\breakingbad

A	B
reviews	notes
This item arrived quickly Via Amazon prime.The picture quality of this Blu-ray is on par. I have personally read over the years that the picture quality for season one on Blu-ray was substandard. I have not found that to be the case. There is plenty of fine detail to be seen on the show in season one. It is a huge leap in terms of visual and audio quality over the DVD.The plot of the show is rather simple. A middle-aged science teacher named Walter White is diagnosed with late stage lung cancer. With a second child on the way, and his family's debt climbing by the moment, he uses his knowledge of chemistry to dive into the lucrative world of drug manufacturing, to make sure that his family is financially secure after his death.There's a reason that Breaking Bad is considered one of the greatest shows in television history. The	

Deuxième Source de Scrapping: ottentomatoes.com



BREAKING BAD: SEASON 1

Genre:Crime, Drama
Network:AMC

BREAKING BAD: SEASON 1 REVIEWS

All Critics | Top Critics | **All Audience**

NEXT →

Lucas T ★★★★★ Jan 16, 2022

AMAZING.An amazing masterpiece.

Code Python

```python
import scrapy
from scrapy.contrib.spiders import Rule
from scrapy.linkextractors import LinkExtractor
from lxml import html


class sidatascraper(scrapy.Spider) :
    name='Reviews'
    start_urls=['https://www.rottentomatoes.com/tv/breaking_bad/s01/reviews?type=user']
    def parse(self, response):
        for review in response.css('ul.audience-reviews li.audience-reviews__item'):
            print("oui")
            yield{
                    'Name': review.css('div.audience-reviews__name-wrap a::text').get().replace("\n", "").strip(),
                    'Comment': review.css('p.audience-reviews__review.js-review-text.clamp.clamp-8.js-clamp::text').get(),
                    'Date': review.css('span.audience-reviews__duration::text').get().strip(),

                    'Rate': len(review.css('span.star-display span.star-display__filled').getall()),



                }
        Rules = (Rule(LinkExtractor(allow=(), restrict_xpaths=('//a[@class="button next"]',)), callback="parse", follow= True),)
            # follow next page links
        next_page = response.xpath('.//a[@class="js-prev-next-paging-next.btn.prev-next-paging__button.prev-next-paging__button-right"]/@href').e
        if next_page:
            next_href = next_page[0]
            next_page_url = 'https://www.rottentomatoes.com/tv/breaking_bad/s01/reviews?type=user' + next_href
            request = scrapy.Request(url=next_page_url)
            yield request
```

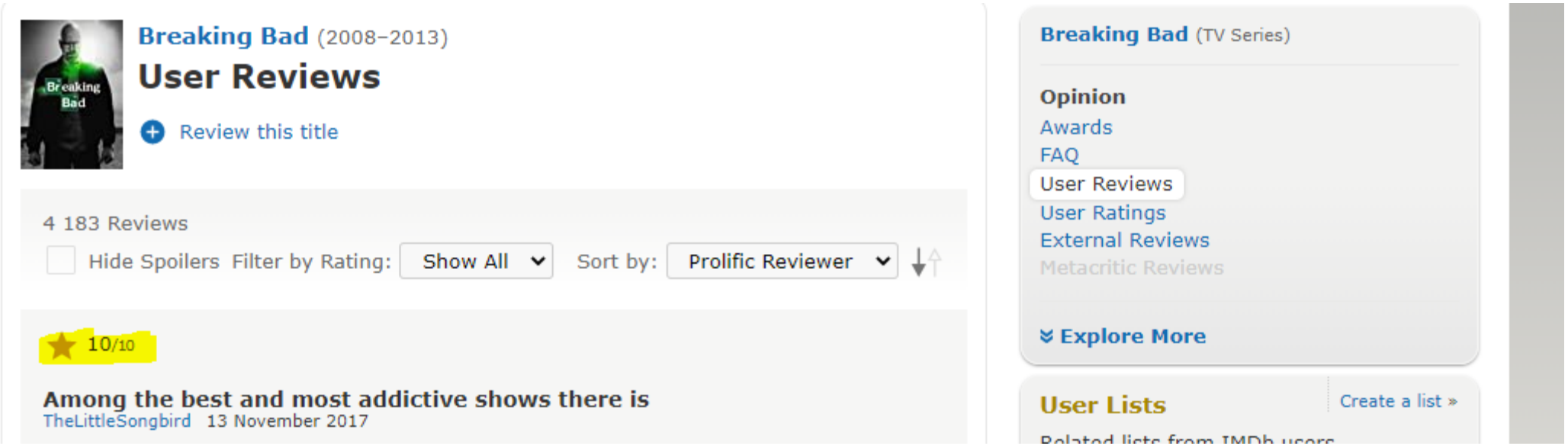
Directory Fichier source

Name	Date modified
BreakingBadReviews	11/26/2021 8:43 PM
corpus24.csv	11/27/2021 5:46 PM

C:\Users\mednour\Desktop\Scrapp\BreakingBadReviews

	A	B	C	D	E	F	G	H	I	J
1	Name	Comment	Date	Rate						
2	Roofy D	Season 1 of Breaking Bad often feels like a slow set-up, but it's a such an interesting story and Bryan Cranston's gripping work in the lead ro								
3	richard m	My name is Walter Hartwell White. I live at 308 Negra Ar	########	4						
4	Daniel B	Sin duda lo que hace excepcional esta primera temporac	25-Oct-21	4						
5		Ø¬Ù…Ø§Ù„ Ù‡Ø°Ø§ Ø§Ù„Ù…Ø³Ù„Ø³Ù„ ÙŠØ¬Ø¹Ù„Ù†Ù†Ù Ø	14-Oct-21	5						
6	Yang D	Best show in history of mankind	2-Sep-21	5						
7	Camille C	Do you remember your chemistry teacher? If you do	1-Sep-21	5						

Troisième Source de Scrapping: imdb.com



Code Python

```python
#%%
import requests
from bs4 import BeautifulSoup
import pandas as pd
Run Cell | Run Above | Debug Cell
#%%

requete = requests.get("https://www.imdb.com/title/tt0903747/reviews?ref_=tt_urv")
page = requete.content
page = BeautifulSoup(page)


comments = page.find_all('div', attrs={'class':'review-container'})
reviews=[]
notes=[]
for comment in comments :
    if len(comment.find('div', attrs={'class':'lister-item-content'}))==9:
        reviews.append(comment.find('div', attrs={'class':'text show-more__control'}).text)
        notes.append(comment.find('span', attrs={'class':'rating-other-user-rating'}).span.text)


if len(page.find('div', attrs={'class':'lister'}))==9:
    next= page.find('div', attrs={'class':'load-more-data'})['data-key']
    url=f'https://www.imdb.com/title/tt0903747/reviews/_ajax?ref_=undefined&paginationKey={next}'
else:
    next=False


while next:
```

Directory Fichier source

BreakingBadReviews	11/26/2021 8:43 PM	File folder	
corpus24.csv	11/27/2021 5:46 PM	Microsoft Excel C...	4 KB
corpus111.csv	11/28/2021 3:56 PM	Microsoft Excel C...	1,655 KB

C:\Users\mednour\Desktop\Scrapp\BreakingBadReviews

	A	B	C
1		reviews	notes
2	0	'Breaking Bad' is one of the most p	10
3	1	Bryan Cranston shows his acting s	9
4	2	The outline is clear. Most people	10
5	3	Some years after writing for "The	10
6	4	To provide for his family (pregnar	8
7	5	It is a wonder, and a great mistake	9
8	6	It's hard for me to be super object	10
9	7	I have never seen a show that I lo	10
10	8	I like the bit where they make cry	10
11	9	Brilliant - one of the greatest drar	10
12	10	This is one of those top of the lea	10
13	11	Having just watched the finale of	10
14	12	PERFECTSeason 1: 8	10

Quatrième Source de Scrapping:Kaggle

Search

Sign In

Arpit Verma

Topic Author

Breaking Bad TV show: All seasons, episodes dataset

Posted in General 7 months ago

Beginner | Intermediate | Advanced | Exploratory Data Analysis | Movies and TV Shows

Below is the link for Breaking Bad TV show dataset:

https://www.kaggle.com/varpit94/breaking-bad-tv-show-all-seasons-episodes-data

Quote | Follow | Bookmark | Report | 3 Upvoters

Directory Fichier source

corpus24.csv	11/27/2021 5:46 PM	Microsoft Excel C...	4 KB
corpus111.csv	11/28/2021 3:56 PM	Microsoft Excel C...	1,655 KB
got.csv	11/28/2021 3:30 PM	Microsoft Excel C...	2,015 KB
scrapy.cfg	11/26/2021 8:31 PM	CFG File	1 KB

C:\Users\mednour\Desktop\Scrapp\BreakingBadReviews

	A	B	C
1		reviews	notes
2	0	reviewers. Plus with such a great cast of talent and a brilliant book series, how could it possibly go wrong? Th	
3	1	television show that does its original source material justice and treats it with respect but it is on its own me	
4	2	television show so brilliant that one has to actually check that it was made for television when everything is	
5	3	strongest examples of an acclaimed show that deserves every ounce of the praise it's garnered.Visually, 'Gar	
6	4	atmospheric and beautiful on the eyes with a real meticulous eye for detail and the costumes suit the charac	
7	5	programme and are not overused or abused, the scale, the detail and how they actually have character and s	
8	6	and editing, which are cinematic quality as well.One cannot talk about 'Game of Thrones' without mentionin	
9	7	unforgettable main theme. Again, worthy of a high-budget fantasy/action/drama film.It is hard not to be bov	
10	8	how good the writing is. It always has a natural flow, is layered and thought-provoking and demonstrates a w	
11	9	story-lines are paced so beautifully, structured with such nuance and attention to coherence, a high emotion	
12	10	there's a set-piece or more action-oriented scene there's always a reason, never there for the sake of it. Not	
13	11	tension but underneath all the scale and flashy attention to detail there is a lot of heart and a multi-layered	
14	12	the appeal too. 'Game of Thrones' has characters that are so well developed and as close to real life as one ca	
15	13	(Joffrey is the only one close to that, the difference though is that he is an extremely interesting one with a l	
16	14	they have much more to them and have strengths and flaws. Decisions are logical and one doesn't like any ch	
17	15	learnt from.'Game of Thrones' cast is full of talented names and, thanks to so well rounded characters and su	
18	16	favourites of mine. Big acting standouts are Peter Dinklage, Sean Bean, Lena Headey and Jack Gleeson (Joffre	
19	17	conclusion, absolutely outstanding and a rare television show worthy of being a cinematic modern classic. Th	
20	18	Bethany Cox	

Fusion Data Set

Source1

```python
path = r'C:\Users\mednour\Desktop\Scrapp\breakingbad' # use your path
all_files = glob.glob(path + "/*.csv")
all_files
```

✓ 0.1s

```
['C:\\Users\\mednour\\Desktop\\Scrapp\\breakingbad\\corpus.csv']
```

```python
li = []

for filename in all_files:
    df = pd.read_csv(filename,index_col=None, header=0)
    li.append(df)
li
```

✓ 1.4s

```
[                                                 reviews  notes
0     \n  This item arrived quickly Via Amazon prime...      4
1     \n  What can be said about this spectacular sa...      5
2     \n  If you're just starting this series then t...      5
3     \n  My husband and I are clearly behind the 8 ...      5
```

Source2

```python
path2 = r'C:\Users\mednour\Desktop\Scrapp\BreakingBadReviews' # use your path
all_files2= glob.glob(path2 + "/*.csv")

li2 = []
```

✓ 0.1s

```python
for filename in all_files2:
    df = pd.read_csv(filename)
    li2.append(df)
li2
```

```
Output exceeds the size limit. Open the full output data in a text editor
[      Unnamed: 0                                          reviews  notes
 0              0  'Breaking Bad' is one of the most popular rate...     10
 1              1  Bryan Cranston shows his acting skills portray...      9
 2              2  The outline is clear. Most people will know wh...     10
 3              3  Some years after writing for "The X Files", Vi...     10
 4              4  To provide for his family (pregnant wife Anna ...      8
```

Concaténer (1) et (2)

```python
df2= pd.concat(li2, axis=0, ignore_index=True)
```

✓ 0.1s

Résultat

```python
df2
```

✓ 0.1s

	Unnamed: 0	reviews	notes	Name	Comment	Date	Rate
0	0.0	'Breaking Bad' is one of the most popular rate...	10.0	NaN	NaN	NaN	NaN
1	1.0	Bryan Cranston shows his acting skills portray...	9.0	NaN	NaN	NaN	NaN
2	2.0	The outline is clear. Most people will know wh...	10.0	NaN	NaN	NaN	NaN
3	3.0	Some years after writing for "The X Files", Vi...	10.0	NaN	NaN	NaN	NaN
4	4.0	To provide for his family (pregnant wife Anna ...	8.0	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...
7559	4084.0	What a final? Wonderful tv show, character and...	8.0	NaN	NaN	NaN	NaN
7560	4085.0	Many of the reviews below reflect exactly how ...	5.0	NaN	NaN	NaN	NaN
7561	4086.0	.....	10.0	NaN	NaN	NaN	NaN
7562	4087.0	Drop down season 8\nIt's very disappointed sea...	1.0	NaN	NaN	NaN	NaN
7563	4088.0	Best tv serie of the times, i just simply love...	10.0	NaN	NaN	NaN	NaN

```
1  result = pd.concat([df2, df1], ignore_index=True)
✓ 0.7s
```

Concaténer (1) (2) et (3)

Résultat Final

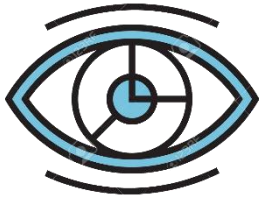
```
1  result = pd.concat([df2, df1], ignore_index=True)
✓ 0.7s
```

```
1  result
✓ 0.2s
```

	reviews	notes
0	'Breaking Bad' is one of the most popular rate...	10.0
1	Bryan Cranston shows his acting skills portray...	9.0
2	The outline is clear. Most people will know wh...	10.0
3	Some years after writing for "The X Files", Vi...	10.0
4	To provide for his family (pregnant wife Anna ...	8.0
...	...	...
11082	\n I watched the first two episodes of this. ...	1.0
11083	\n This isn't a review on the item as i know ...	1.0
11084	\n Great, as described and arrived ahead of s...	5.0
11085	\n Worked great- no issues! It came in pristi...	5.0
11086	\n Breaking Bad ist auch meiner Meinung nach ...	5.0

11087 rows × 2 columns

Visualisation



Données initiales data Set

	reviews	notes
0	'Breaking Bad' is one of the most popular rate...	10.0
1	Bryan Cranston shows his acting skills portray...	9.0
2	The outline is clear. Most people will know wh...	10.0
3	Some years after writing for "The X Files", Vi...	10.0
4	To provide for his family (pregnant wife Anna ...	8.0
...	...	...
11082	\n I watched the first two episodes of this. ...	1.0
11083	\n This isn't a review on the item as i know ...	1.0
11084	\n Great, as described and arrived ahead of s...	5.0
11085	\n Worked great- no issues! It came in pristi...	5.0
11086	\n Breaking Bad ist auch meiner Meinung nach ...	5.0

11087 rows × 2 columns

Défaillance et problèmes

```python
F.isnull().sum()
```

✓ 0.1s

```
reviews    11
notes      10
dtype: int64
```

```python
# total missing values
total_missing = F.isna().sum().sum()
print(f'Total missing values: {total_missing}.')
```

✓ 0.9s

```
Total missing values: 21.
```

10595	\n Totalmente en español, y si te gusta la se...	5.0
10876	\n This was a good buy.It was a great season ...	5.0
10883	\n Wer mit "breaking bad" anfängt, wird NICHT...	5.0
11043	\n habe mir die staffeln von breaking bad bes...	3.0
11083	\n This isn't a review on the item as i know ...	1.0

Beaucoup Des valeurs Nulles

Des colonnes inutiles


Donnés contenant: /n des ponctuations
des balises HTML **de bruits**…etc.

Nettoyage et pré-traitement des données

Eliminer les colonnes inutiles

```python
# Delete Null Values
File.dropna( inplace = True)
File.isnull().sum()
```

✓ 0.1s

```
reviews    0
notes      0
dtype: int64
```

Eliminer les commentaires dupliqués…

```python
File.drop_duplicates(subset=['reviews'])
```

✓ 0.1s

	reviews	notes
0	'Breaking Bad' is one of the most popular rate...	10.0
1	Bryan Cranston shows his acting skills portray...	9.0
2	The outline is clear. Most people will know wh...	10.0
3	Some years after writing for "The X Files", Vi...	10.0
4	To provide for his family (pregnant wife Anna ...	8.0

Eliminer les Balises HTML ,/n ,chiffres…etc

```python
File = File.replace('<[^<]+?>', '', regex = True)
File = File.replace('\n', '', regex = True)
File = File.replace('\t', '', regex = True)
#\d remove numbers
File = File.replace('\d', '', regex = True)
#File = File.replace('\s', '', regex = True)
```

Nettoyage et pré-traitement des données

Transformer les reviews en minuscules…

```python
File['reviews'] = File['reviews'].str.lower()


```

✓ 0.2s

```
APL=A Programming Language
OMGAGA=Oh My God
OMG=Oh My God
AFAIK=As Far As I Know
AFK=Away From Keyboard
ASAP=As Soon As Possible
ATK=At The Keyboard
ATM=At The Moment
A3=Anytime, Anywhere, Anyplace
BAK=Back At Keyboard
BBL=Be Back Later
BBS=Be Back Soon
BFN=Bye For Now
B4N=Bye For Now
BRB=Be Right Back
BRT=Be Right There
BTW=By The Way
B4=Before
B4N=Bye For Now
CU=See You
CUL8R=See You Later
CYA=See You
FAQ=Frequently Asked Questions
FC=Fingers Crossed
FWIW=For What It's Worth
FYI=For Your Information
GAL=Get A Life
```

Préparer un dictionnaire de donnés pour remplacer des abréviations des mots en mo complètes afin de faciliter l'apprentissage d sentiment de reviews.

```python
a_dictionary = {}
a_file = open("slang.txt")
for line in a_file:

    print(line)
    p=line.split("=")
    print(p)
    a_dictionary[p[0]]=p[1]
```

✓ 0.5s

Nettoyage et pré-traitement des données

Breaking Bad

```
U2=You Too
U4E=Yours For Ever
WB=Welcome Back
WTF=What The F...
WTG=Way To Go!
WUF=Where Are You From?
W8=Wait...
7K=Sick:-D Laugher
breaking bad= this movie
```

On remarque que le **mot Breaking Bad**(Nom de série) **se répète** plusieurs fois dans les reviews……

Le risque que le mot **Bad** a une apparition négatif dans la fonction **polarity** alors elle a un effet **pas complètement correcte** dans l'apprentissage

La solution qu'on remplace le mot Breaking Bad par **This movie**.

Nettoyage et pré-traitement des données

Supprimer les ponctuations et les mots bruits…

STOP

```
File['reviews']=File['reviews'].astype(str)
File['reviews']=File['reviews'].apply (lambda x : remove_punctuation (x))
File['reviews']=File['reviews'].apply (remove_stop)
```

Supprimer les liens | https://

```
1  File['reviews']=File['reviews'].apply (lambda x : remove_http (x))
```

Stemming

adjustable → adjust
formality → formaliti
formaliti → formal
airliner → airlin ⚠

Lemmatization

was → (to) be
better → good
meeting → meeting

Lemmatiser les commentaires ..(sous forme canonique)…

```
1  File['reviews']=File['reviews'].astype(str)
2  File["lemmreviews"]=File["reviews"].apply(lambda row: " ".join([w.lemma_ for w in nlp(row)]))
3
```

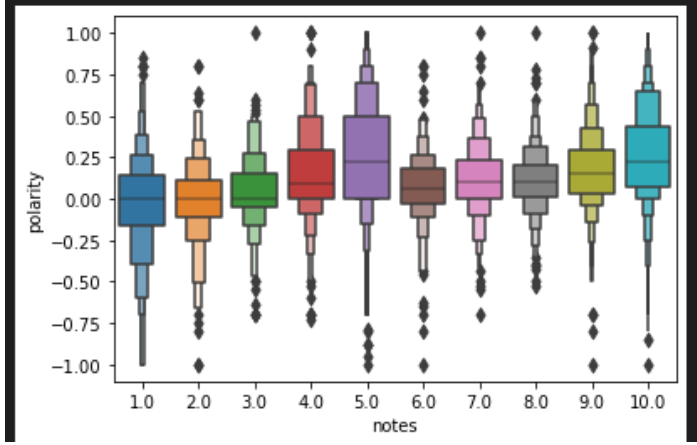
Nettoyage et pré-traitement des données

Appliquer la fonction **polarity** qui va détecter **la subjectivité** de commentaire et le retourner de valeurs entre **[-1..1]**. En appliquant quelques **Visualisations graphiques**

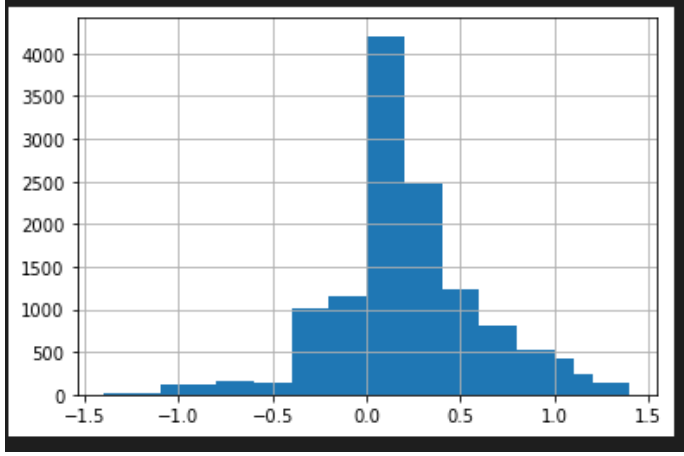
```
File['polarity']= File.lemmreviews.apply(detect_polarity)
```

```
File['polarity']= File.lemmreviews.apply(detect_polarity)

###########visualisation


import seaborn as sns
#plt.figure(figsize=(10,6))
sns.boxenplot(x='notes', y='polarity', data=File)
```

```
<AxesSubplot:xlabel='notes', ylabel='polarity'>
```




```
File['polarity'].hist()
plt.bar(File.polarity.value_counts().index, File.polarity.value_counts())
ddd=File.copy()
```



Nettoyage et pré-traitement des données

La fonction **polarity** retourne des valeurs comme -0.2, 0.5, 1, 0.7 mais lorsque on est dans un problème de **classification** alors on a besoin de **deux classes** par exemples: 0: pour les reviews negatives et 1: pour les reviews positives.
Alors on va régler colonne polarity du data set comme suit:

```
File.loc[ (File.polarity<0), 'polarity'] = -1
File.loc[ (File.polarity>=0), 'polarity'] = 1
```

```
File.polarity
```

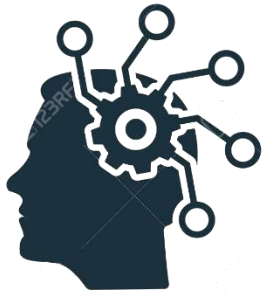
```
0        1.0
1        1.0
2        1.0
3       -1.0
4        1.0
         ...
11082   -1.0
11083    1.0
11084    1.0
11085    1.0
11086   -1.0
Name: polarity, Length: 11076, dtype: float64
```

DataSet Après Traitement

```
File.head()
```

	reviews	notes	lemmreviews	polarity
0	breaking bad one popular rated shows imdb one ...	10.0	break bad one popular rate show imdb one rarit...	1.0
1	bryan cranston shows acting skills portraying ...	9.0	bryan cranston show acting skill portray compl...	1.0
2	outline clear people know even havent seen sho...	10.0	outline clear people know even have not see sh...	1.0
3	years writing x files vince gilligan created s...	10.0	year write x file vince gilligan create show s...	-1.0
4	provide family pregnant wife anna gunn teenage...	8.0	provide family pregnant wife anna gunn teenage...	1.0

Modèle Machine Learning



Elaborer un modèle de machine learning

Build Machine Learning Model

```python
import pickle
from sklearn.svm import LinearSVC
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
vect = TfidfVectorizer(ngram_range=(1,1),use_idf=True,stop_words=stopwords.words('english'))

X = vect.fit_transform(File['lemmreviews'])
Y=File['polarity']
X_train, X_test, y_train, y_test = train_test_split(X,Y, random_state = 0,test_size=0.1)
model = LinearSVC().fit(X_train, y_train)
model.score(X_test,y_test)
pickle.dump(model, open('model1.pkl','wb'))

model1 = pickle.load(open('model1.pkl','rb'))

text=["breaking bad is boring and violence "]

vectorize=vect.transform(text)

print(model1.predict(vectorize))

```

```
[-1.]
```

Matrice de confusion, Accuracy Score et Report

```python
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.metrics import accuracy_score
y_pred=model1.predict(X_test)
results = confusion_matrix(y_test, y_pred)
print ('Confusion Matrix :')
print(results)
print ('Accuracy Score :',accuracy_score(y_test, y_pred) )
print ('Report : ')
print (classification_report(y_test, y_pred))
```

```
Confusion Matrix :
[[114  71]
 [ 27 896]]
Accuracy Score : 0.9115523465703971
Report :
              precision    recall  f1-score   support

        -1.0       0.81      0.62      0.70       185
         1.0       0.93      0.97      0.95       923

    accuracy                           0.91      1108
   macro avg       0.87      0.79      0.82      1108
weighted avg       0.91      0.91      0.91      1108
```

Enregistrer le modèle .PKL

Enregistrer le modèle .pkl avec la bibliothèque joblib afin de l'exploiter dans une application web avec plus de performance de sécurité de notre modèle…

Save Model

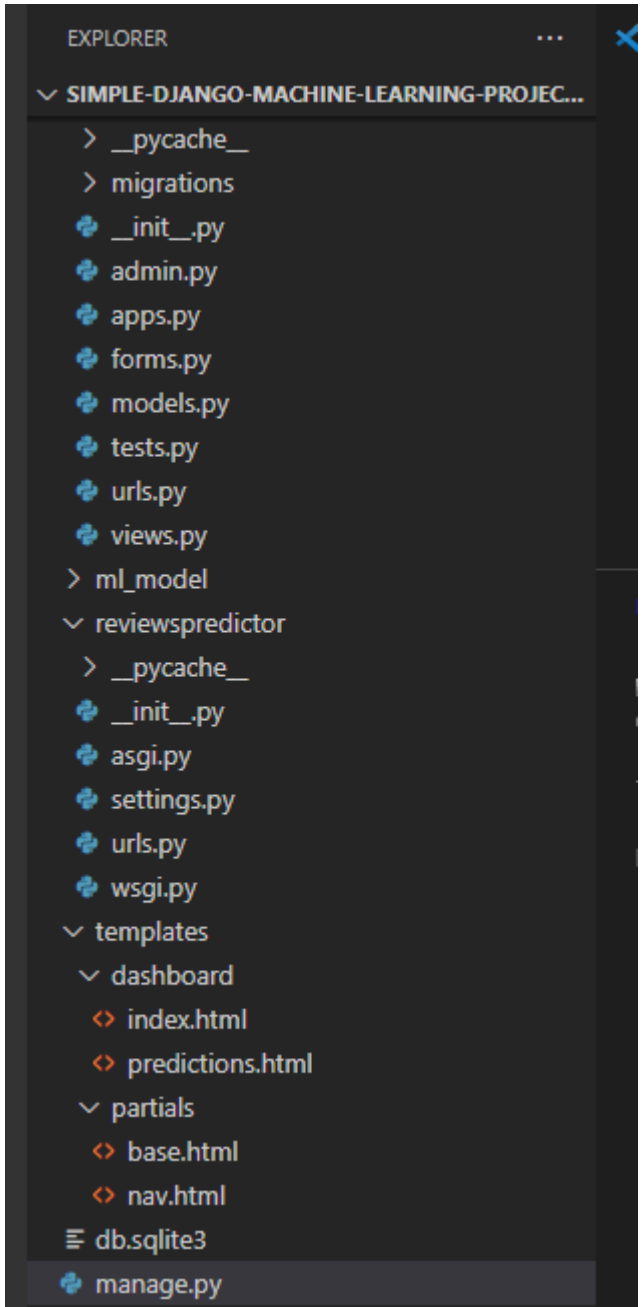
```
[ ]
1  import joblib
```

```
[ ]
1  joblib.dump(model,'ml_reviews_model.joblib')
2  joblib.dump(vect, 'vectorizer.pkl')
```

```
...  ['vectorizer.pkl']
```

Application Demonstration

Structure application Django



Structure application Django

models.py

ashboard > models.py > ...

```python
class Data(models.Model):
    reviews = models.CharField(max_length=100, null=True)
    #Rate = models.PositiveIntegerField(validators=[MinValueValidator(0), MaxValueValid
    #height = models.PositiveIntegerField(null=True)
    #sex = models.PositiveIntegerField(choices=GENDER, null=True)
    predictions = models.CharField(max_length=100, blank=True)
    date = models.DateTimeField(auto_now_add=True)

    def save(self, *args, **kwargs):
        ml_model = joblib.load('ml_model/ml_reviews_model.joblib')
        v = joblib.load('ml_model/vectorizer.pkl')
        news_reviews2=""
        v1=v.transform([self.reviews])
        new_reviews=ml_model.predict(v1)
        if (new_reviews[0]==-1.0) :
            news_reviews2=news_reviews2+"Negative"
        else:
            news_reviews2=news_reviews2+"Positive"
        self.predictions = news_reviews2
        return super().save(*args, *kwargs)

    class Meta:
        ordering = ['-date']

    def __str__(self):
        return self.name
```

Structure application Django

python

django

predictions.html ×

mplates > dashboard > predictions.html > div.container > div.row.mt-4 > div.col-md-8.off

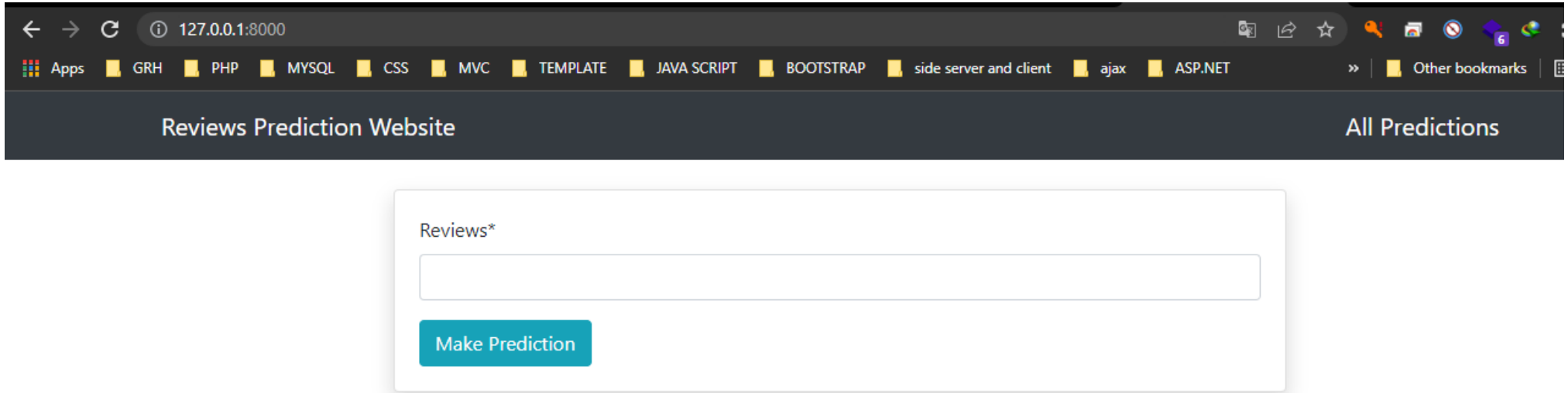
```html
{% extends 'partials/base.html' %}
{% block title %}All Prediction{% endblock %}

{% block content  %}
<div class="container">
    <div class="row mt-4">
        <div class="col-md-8 offset-md-2">
            <table class="table">
                <thead>
                  <tr>
                    <th scope="col">reviews</th>

                    <th scope="col">Prediction</th>
                  </tr>
                </thead>
                <tbody>
                  {% for data in predicted_reviews %}
                  <tr>
                    <th scope="row">{{ data.reviews }}</th>

                    <td>{{ data.predictions }}</td>
                  </tr>
                  {% endfor %}
                </tbody>
              </table>
        </div>
    </div>
</div>
{% endblock %}
```

DEMO

python

django

BOOTSTRAP



DEMO

python

django

BOOTSTRAP

Reviews*

breaking bad is the best serie i know ever and ever in my life!! really i enjoy all actor...

Make Prediction

reviews	Prediction
breaking bad is the best serie i know ever and ever in my life!! really i enjoy all actor...	Positive

Reviews*

really breaking bad is boring..more scene are violents...

Make Prediction

really breaking bad is boring..more scene are violents...

DEMO

python

django

BOOTSTRAP

Reviews Prediction Website

All Predictions

reviews	Prediction
breaking bad is the best serie i know ever and ever in my life!! really i enjoy all actor...	Positive
best series in ever and ever	Positive
best series in ever and ever	Positive
breaking bad is boring noisy	Negative
best series ever	[1.]
bad boring series	[-1.]
bad borings series ever	[-1.]
bad series ever	[-1.]
best show	[1.]

Conclusion

À l'avenir, l'exploration de données inclura des types de données plus complexes. De plus, pour tout modèle qui a été conçu, un raffinement supplémentaire est possible en examinant d'autres variables et leurs relations. La recherche en data mining aboutira à de nouvelles méthodes pour déterminer les caractéristiques les plus intéressantes des données.

Merci pour votre attention