



Université Mohammed V  
Faculté des Sciences  
Rabat

Master IDDL

Réalisé par :

HOUDA KHABBALI &

OUSAMA ZARROK

## Rapport mini-projet NLP

### Introduction:

Dans ce mini-projet, nous allons utiliser l'analyse des sentiments sur les titres d'articles publiés dans différents domaines (politique, économie, ...) sur la plate-forme HESPRESS.



Si au lieu d'avoir à parcourir chaque titre et de comprendre qu'il s'agit d'une nouvelle positive ou négative nous pouvons utiliser Python pour analyser « fr.hespress.com » et rassembler tous ces titres d'articles, puis utiliser l'analyse des sentiments pour comprendre si la moyenne des nouvelles de chaque domaine est positive ou négative.

### 1) Obtenir et manipuler les données de HESPRESS

- HESPRESS est un journal électronique marocain et les titres d'articles sont aussi dans le code source HTML et on a utilisé « BeautifulSoup » pour analyser ce code HTML et d'obtenir les titres pertinents afin d'analyser et d'appliquer sentiment analysis sur le texte de ces titres.

- On a importé `urlopen` et `request` de « `urllib.request` » pour ouvrir URL et demander les données HTML et `BeautifulSoup` de `bs4` pour intégrer et analyser le contenu HTML.
- On a importé `pandas` pour nous permet de manipuler les données dans un tableau et d'appliquer très rapidement sentiment analysis sur notre ensemble de données.

```
from urllib.request import urlopen, Request
from bs4 import BeautifulSoup
import pandas as pd

hespress_url = 'https://fr.hespress.com/'
news = ['politique', 'economie', 'sport', 'societe', 'monde', 'culture', 'opinions']

news_tables = {}
for new in news:
    url = hespress_url + new

    req = Request(url=url, headers={'user-agent': 'my-app'})
    response = urlopen(req)

    html = BeautifulSoup(response, features='html.parser')
    news_table = html.find(id='listing')
    news_tables[new] = news_table

parsed_data = []

for new, news_table in news_tables.items():
    for row in news_table.findAll('h3'):
        title = row.text

        parsed_data.append([new, title])

df = pd.DataFrame(parsed_data, columns=['new', 'title'])
```

## 2) Application l'analyse des sentiments

- Avant d'appliquer sentiment analysis on a commencé par installer la bibliothèque « `transformers` » qui nous permet d'importer, de télécharger et d'installer facilement le modèle nlp qu'on a utilisé c'est le modèle BERT multilingual « **bert-base-multilingual-uncased-sentiment** » qui nous permet de

donner un score de sentiment entre un et cinq et d'effectuer sentiment analysis qu'il fonctionne en six langues : anglais, français, néerlandais, allemand, espagnol et italien et cela nous convient car les titres qu'on veut analyser est en français.

```
[ ] !pip install transformers
```

```
Collecting transformers
  Downloading transformers-4.9.2-py3-none-any.whl (2.6 MB)
    |████████████████████████████████████████| 2.6 MB 4.1 MB/s
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from transformers) (21.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (from transformers) (3.0.12)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.7/dist-packages (from transformers) (2019.12.20)
Collecting huggingface-hub==0.0.12
  Downloading huggingface_hub-0.0.12-py3-none-any.whl (37 kB)
Collecting tokenizers<0.11,>=0.10.1
  Downloading tokenizers-0.10.3-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux2_12_x86_64.manylinux2010_x86_64.whl (3.3 MB)
    |████████████████████████████████████████| 3.3 MB 39.8 MB/s
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.7/dist-packages (from transformers) (4.62.0)
Collecting sacremoses
  Downloading sacremoses-0.0.45-py3-none-any.whl (895 kB)
    |████████████████████████████████████████| 895 kB 45.6 MB/s
Collecting pyyaml>=5.1
  Downloading PyYAML-5.4.1-cp37-cp37m-manylinux1_x86_64.whl (636 kB)
    |████████████████████████████████████████| 636 kB 54.6 MB/s
Installing collected packages: tokenizers, sacremoses, pyyaml, huggingface-hub, transformers
  Attempting uninstall: pyyaml
    Found existing installation: PyYAML 3.13
    Uninstalling PyYAML-3.13:
      Successfully uninstalled PyYAML-3.13
Successfully installed huggingface-hub-0.0.12 pyyaml-5.4.1 sacremoses-0.0.45 tokenizers-0.10.3 transformers-4.9.2
```

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification
import torch

tokenizer = AutoTokenizer.from_pretrained('nlpTown/bert-base-multilingual-uncased-sentiment')

model = AutoModelForSequenceClassification.from_pretrained('nlpTown/bert-base-multilingual-uncased-sentiment')
```

- Après qu'on a instancié et configuré notre modèle On a créé une fonction qui nous permet de prendre les titres et le transmettre à tokenizer et ensuite au modèle qu'on a choisi et obtenir le score de sentiment.

```
def sentiment_score(title):
    tokens = tokenizer.encode(title, return_tensors='pt')
    result = model(tokens)
    return int(torch.argmax(result.logits))+1

df['sentiment'] = df['title'].apply(lambda x: sentiment_score(x[:512]))
```

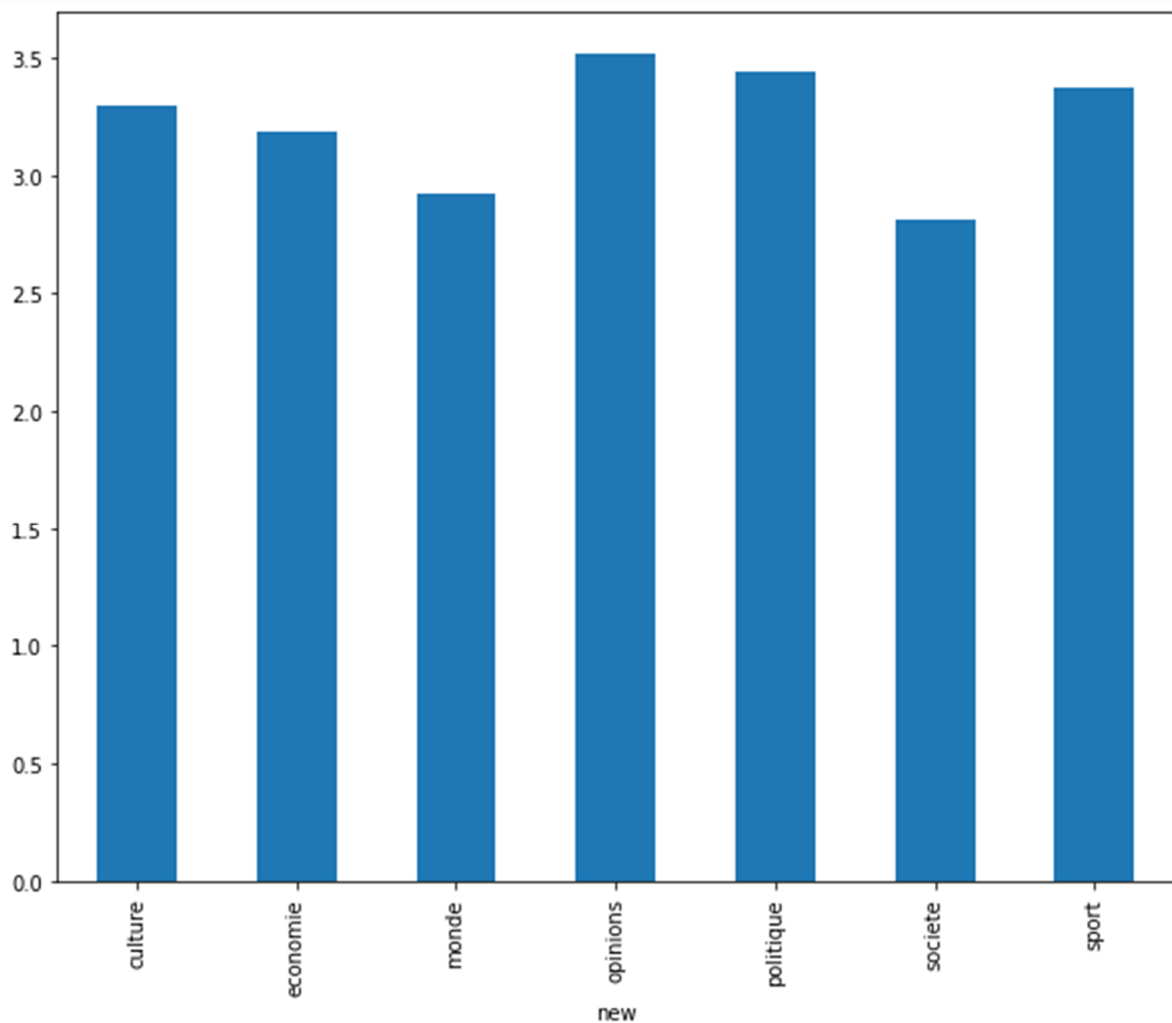
### 3) Visualisation de l'analyse des sentiments

- Après qu'on a appliqué sentiment analysis et on a vu à quel point les titres sont positives ou négatives, on a pris tous les

scores de sentiment de tous les titres des articles de chaque domaine et on a calculé la moyenne de cela pour visualiser ces données dans un diagramme en barres en utilisant « matplotlib »

```
plt.figure(figsize=(10,8))
mean_df = df.groupby(['new']).mean()
mean_df=mean_df.unstack()
mean_df = mean_df.xs('sentiment').transpose()
mean_df.plot(kind='bar')
plt.show()
```

- Résultat :



#### 4) Conclusion

Dans ce projet nous avons étudié comment faire une analyse des sentiments à partir des sources textuelles dématérialisées sur de grandes quantités de données (big data) comme "hespress", en

utilisant les « transformers » avec le modèle "BERT" et les visualisés en graphes.

Et on peut aussi utiliser analyse des sentiments pour mieux comprendre les avis et l'envie des clients par rapport à un produit quelconque, et donner de nouvelles perspectives aux équipes afin d'obtenir un travail de meilleure qualité et plus productif qui va aider à créer un bon plan de marketing.