Fake news detector

Groupe 13 - KUHN, TACHOIRES, HOURI, CHAUVARD, JORET DES CLOSIÈRES, JACQUIN

https://gitlab-cw6.centralesupelec.fr/augustin.tachoires/fakenewsdetector_group13

Contextualisation :

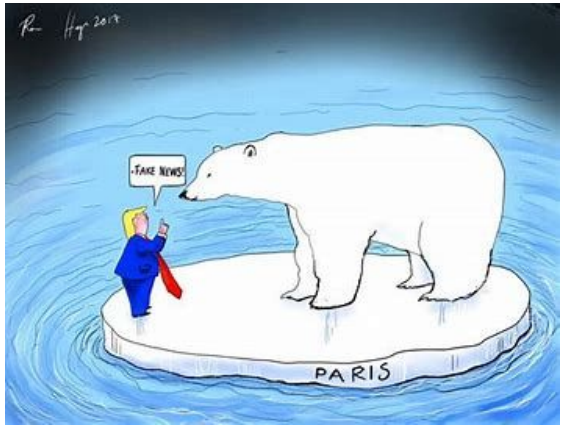
- Propagation de fake news sur les réseaux sociaux, en particulier sur twitter.

- Désinformation des populations sur des sujets à grands enjeux (climatiques, politiques voire scientifiques).

- Pas assez de modération sur les réseaux sociaux, et équipe en sous-effectifs notamment dans les pays en voie de développement.

Quelques exemples parlants…

Environnement :



Least trustable tweets

@Datti_Ahmed_PHD @APCPresCC2022 Yes ooooo. Maybe another fake CAN lager beer.
Posted by @Golibeem, got 0 retweets, credibility index of 0.8149333333333333.
RT @PeterDClack: 'Global warming' has been catastrophic for United Nations' credibility - because it didn't happen. It was all about money,…
Posted by @NakeDTrutH04, got 1928 retweets, credibility index of 0.8084932766149875.

Quelques exemples parlants…

Coronavirus :

Least trustable tweets

@andrevvsdavid @ZubyMusic India did we'll because of hydro chloroquine and Ivermectin, masks are useless.
Posted by @scharfmel, got 0 retweets, credibility index of 0.900533357758061.

Quelques exemples parlants...

Politique internationale :

Least trustable tweets

@UN @UNDPPA @DicarloRosemary For this USA and Europe should stop supporting ukrain. Ukraine is responsible for not stopping war now.
Posted by @drPawanduhan, got 0 retweets, credibility index of 0.8814409489418917.

Comment combattre la désinformation sur twitter ?

Le projet sur lequel nous avons travaillé cette semaine a pour ambition de **détecter les fake news par l'analyse générale de la crédibilité des tweets les plus populaires sur un sujet don**né.

Nous avons construit un tableau de bord capable de :

- afficher les 3 tweets qui ont la plus grande probabilité d'être fake.
- afficher en opposition les 3 tweets les plus crédibles.
- estimer la crédibilité d'un utilisateur.

En entrée : une liste de termes en rapport à un sujet donné.

Application concrète : évaluation de l'opinion publique sur twitter lors d'une campagne politique.

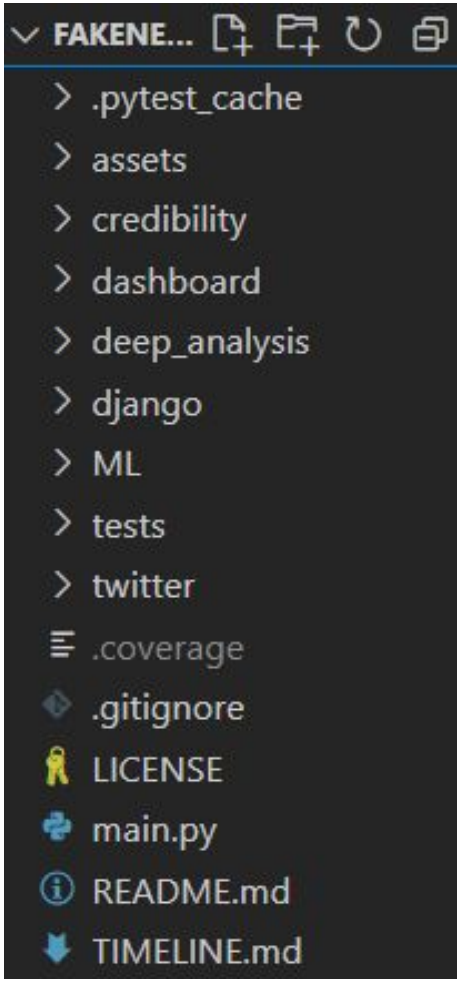
Application secondaire : estimation de la crédibilité d'un tweet donné.

Usage : Préparation d'une campagne politique.

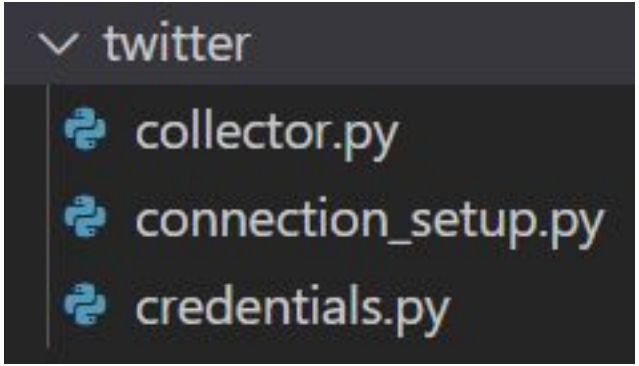
- Objectif: savoir ce qui occupe l'esprit des électeurs, sur un sujet donné.
- Identifie les rumeurs infondées pour les débunker.
- Identifie les nouvelles crédibles pour savoir sur quoi on peut être interrogé en interview.
- Identifie les notions associées au sujet.

Organisation du code

- Tests unitaires (TDD), couverture de 96%
- Documentation
  - README.md
  - docstrings

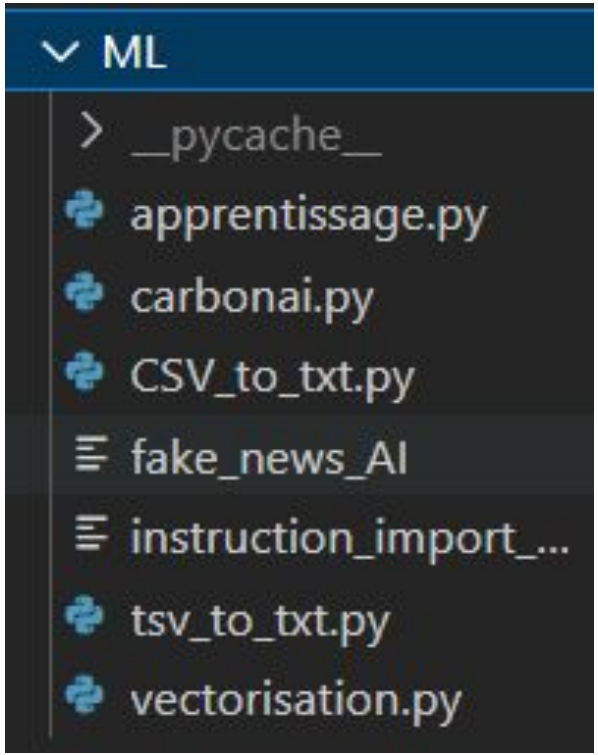


1. Interface avec twitter.



- **Connexion à l'API twitter**.
- **Extraction** des tweets.
- **Formatage des tweets** pour leur traitement ultérieur.

2. Modèle de Machine Learning.



Modèle de Machine Learning choisi

```
[[ 99 415]
 [136 617]]
              precision    recall  f1-score   support

       false       0.42      0.19      0.26       514
        true       0.60      0.82      0.69       753

    accuracy                           0.57      1267
   macro avg       0.51      0.51      0.48      1267
weighted avg       0.53      0.57      0.52      1267

0.5651144435674822
```

RandomForestClassifier, 1000 estimateurs

```
[[158 356]
 [226 527]]
              precision    recall  f1-score   support

       false       0.41      0.31      0.35       514
        true       0.60      0.70      0.64       753

    accuracy                           0.54      1267
   macro avg       0.50      0.50      0.50      1267
weighted avg       0.52      0.54      0.53      1267

0.5406471981057617
```

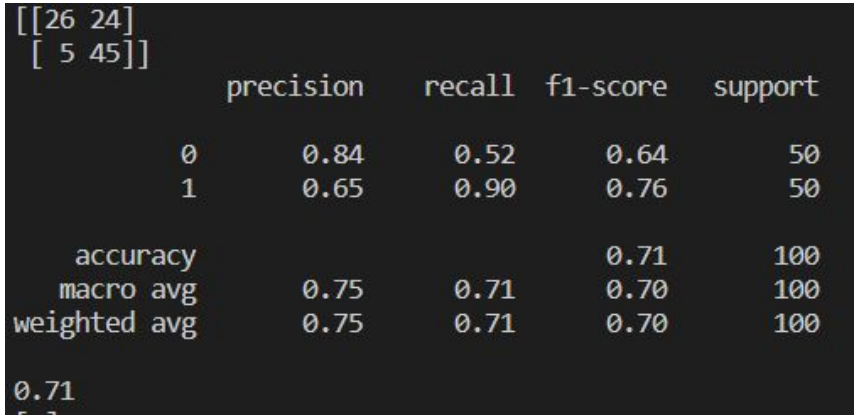
SVC, C=10, linéaire

Modèle de Machine Learning choisi

```
[[56 51 38 31 15 21]
 [41 74 44 33 14 43]
 [51 56 63 45 11 39]
 [43 36 51 51 18 42]
 [18 26 16  8 13 11]
 [26 45 47 41 14 35]]
              precision    recall  f1-score   support

 barely-true       0.24      0.26      0.25       212
       false       0.26      0.30      0.28       249
   half-true       0.24      0.24      0.24       265
 mostly-true       0.24      0.21      0.23       241
  pants-fire       0.15      0.14      0.15        92
        true       0.18      0.17      0.18       208

    accuracy                           0.23      1267
   macro avg       0.22      0.22      0.22      1267
weighted avg       0.23      0.23      0.23      1267

0.230465666929755 31
```

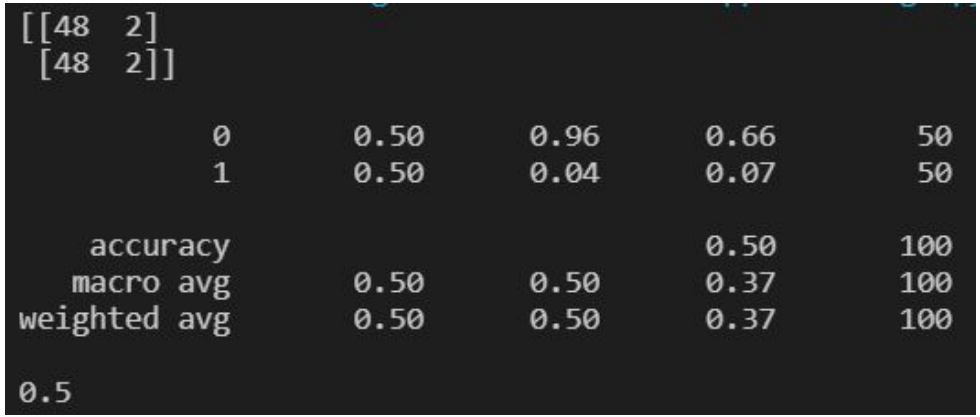
SVC, C = 100, linéaire, avec plus que 2 classes (faux, peu vrai, à moitié vrai, plutôt vrai, etc…)

Essai avec d’autres datasets (RandomForest)



```
[[26 24]
 [ 5 45]]
              precision    recall  f1-score   support

           0       0.84      0.52      0.64        50
           1       0.65      0.90      0.76        50

    accuracy                           0.71       100
   macro avg       0.75      0.71      0.70       100
weighted avg       0.75      0.71      0.70       100

0.71
```

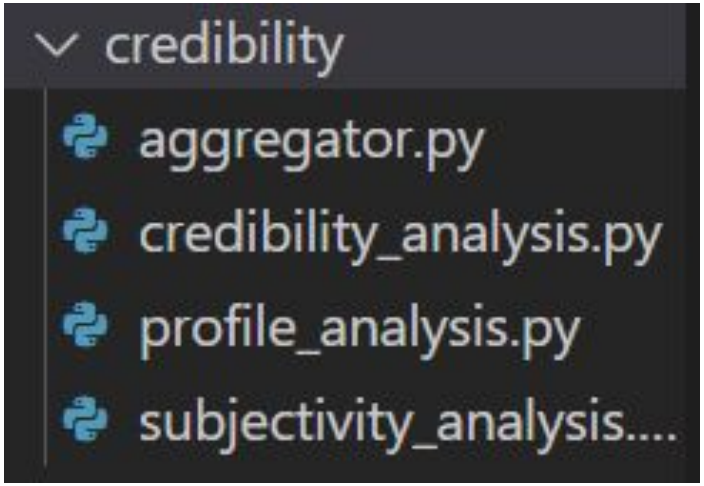
Dataset d’articles



```
[[48  2]
 [48  2]]
              precision    recall  f1-score   support

           0       0.50      0.96      0.66        50
           1       0.50      0.04      0.07        50

    accuracy                           0.50       100
   macro avg       0.50      0.50      0.37       100
weighted avg       0.50      0.50      0.37       100

0.5
```

Dataset de tweets

3. Analyse de la crédibilité.



- Analyse de la subjectivité et de la crédibilité d'un tweet.
- Attribution d'une note finale par **pondération des différents facteurs**.

Critères pris en compte

- Véracité du tweet : Machine Learning

Critères pris en compte

- Véracité du tweet : Machine Learning
- Analyse de l’auteur :

  - activity_credibility

$$\log\left(\frac{\text{nombre de tweets et retweets}}{\text{age du compte}}\right)$$

Critères pris en compte

- Véracité du tweet : Machine Learning
- Analyse de l’auteur :

  - activity_credibility

  - follow_credibility

$$\log\left(\frac{\text{nombre de followings}}{\text{nombre de followers}}\right)$$

Critères pris en compte

- Véracité du tweet : Machine Learning
- Analyse de l’auteur :

  - activity_credibility

  - follow_credibility

  - age_credibility

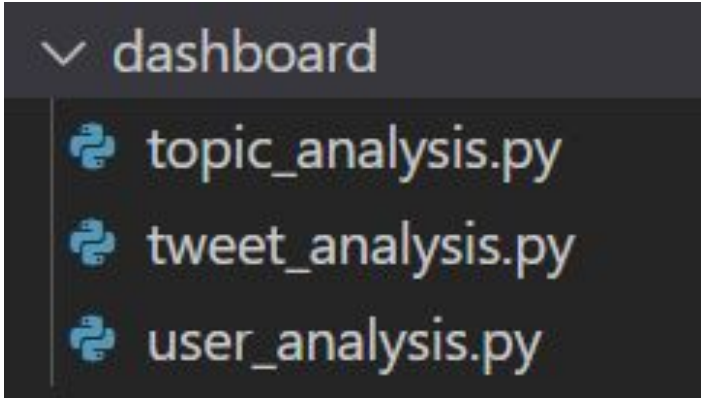
    1 - log(âge du compte)

Prise en compte de ces critères

Critère	Coefficient
Machine Learning	6
activity_credibility	1
follow_credibility	1
age_credibility	1

4. Mise en place du site par Dash.

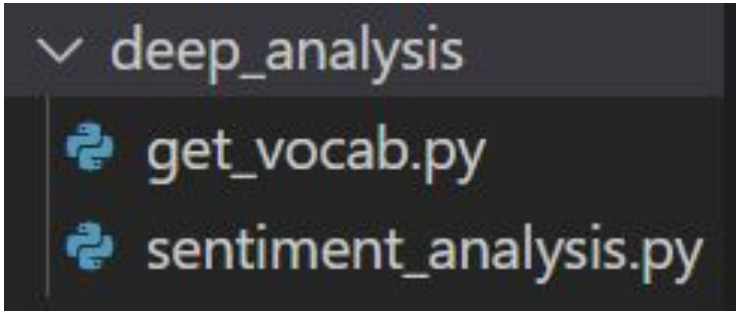
plotly | Dash



- **Création de l'application** avec champs d'entrée.
- Display de l'**histogramme de répartition de crédibilité** des fake news.
- Display des 3 tweets les plus fake et des 3 tweets les plus crédibles.

interface mise à jour depuis…

5. Analyse complémentaire.



- Récupération du vocabulaire associé à un tweet.
- **Sentiment analysis.**
- Récupération des retweeteurs.

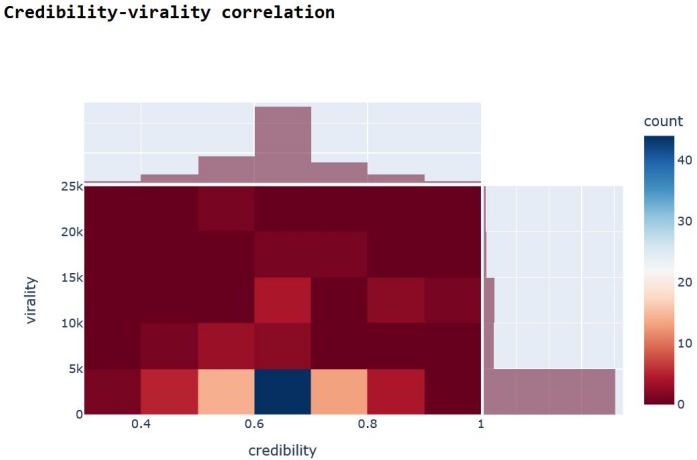
6. Enrichissement du dashboard.

plotly | Dash

- Affichage d'un **nuage de mots** associés aux 30% des tweets les moins crédibles

- Affichage de **l'analyse de sentiments**.

- Affichage de la **viralité** en fonction de la crédibilité



7. Analyse poussée des tweets extrêmes.

Décomposition de l'analyse, avec affichage du score de chaque critère.

Input : global warming.

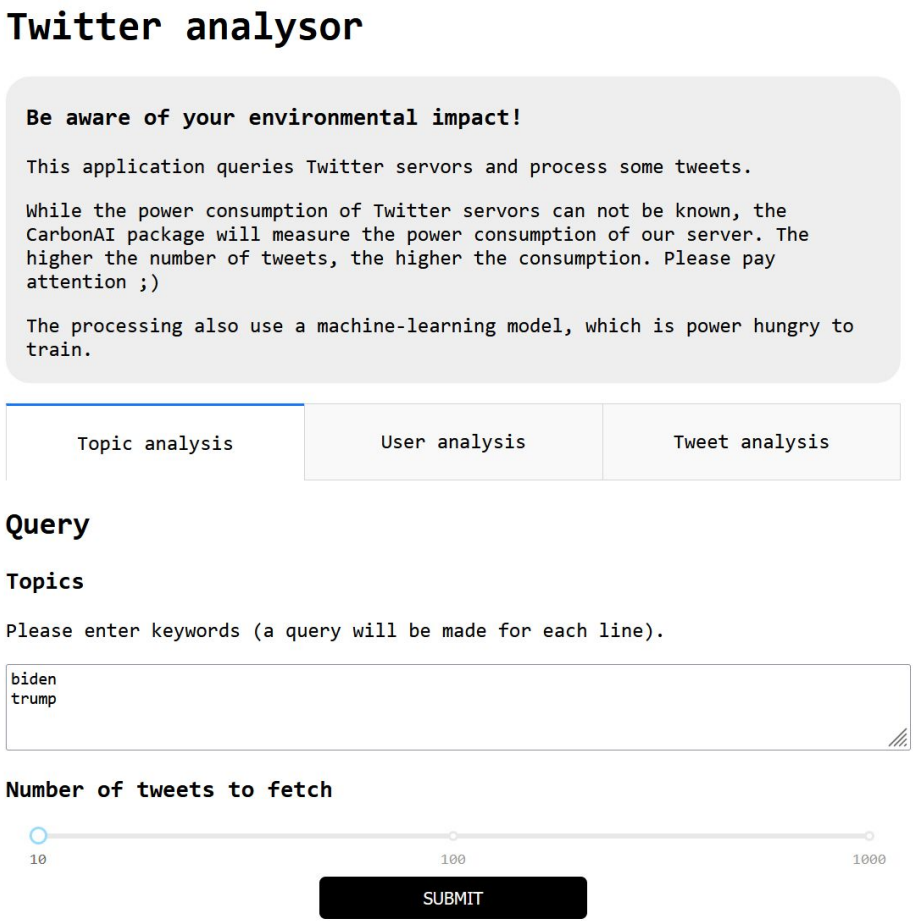
RT @PeterDClack: Global warming has been faked, based on climate studies for 10K yrs. Massive glacier fields across Europe & America are fa…

▼ Tweeted by @hallermunt.

- Credibility 0.86
  - Machine learning-based estimation 0.84
  - Author credibility 0.89
    - Account age **21 day(s)** 1
    - Following/followers ratio **0** 1
    - Activity **8 status(es) per day** 0.68
- Virality **1855 retweets**
- Polarity **0.0** (-1: negative, 1: positive)
- Subjectivity **0.5** (0: objective, 1: subjective)
- Tweet ID **1593514263406170115** (use it in the "Tweet analysis" tab)

8. Requête spécifiques.

- Refonte du site en onglets.
- Onglet crédibilité d'un tweet.
- Onglet crédibilité d'un utilisateur.



Critiques et pistes d'amélioration

- **Risque de laisser passer certains types de Fake News**.
- Risque de filtrer certains **comptes qui dénoncent des fake news (over-fitting)**.
- Précision du modèle de machine learning.
- Optimisation des coefficients dans le calcul de la crédibilité nécessaire.
- Affichage de l'**impact écologique (carbonAI)**.

Annexe - Répartition du travail

Travail en équipe sur chaque sprint.

Yann -----> Machine learning.
David -----> Django et extraction données ML
Eliott -----> Analyse de la crédibilité, choix des critères.
Théo -----> CarbonAI
Arthur -----> Réalisation de l'interface.
Augustin ------> Analyse de la crédibilité, choix des critères.