

Détection des Fake News avec l'IA

Projet d'intégration des modèles IA

Par: Hubert Chavasse, Hadrien LENNON, Gemima ONDELE, Mohamed GHARMAOUI, Niangoran BOKA





Objectif du Projet

- Développer une IA
Capable de détecter les fake news
- Intégrer la solution
Dans un outil pratique
- Tester et généraliser
Sur diverses sources d'actualité

Étapes du Projet

Prompter

Définir critères de classification

Modéliser

Entraîner un modèle NLP

Processer

Extraire et vectoriser les textes

Robotiser

Intégrer dans une application

Prompter : Critères de Classification

Sources douteuses

Absence de citations fiables

Tonalité émotionnelle

Langage sensationnaliste

Incohérences factuelles

Contradictions dans le texte



Modéliser : Entraînement du Modèle NLP



Processor : Extraction et Vectorisation

1

Nettoyage

Suppression HTML, stop-words

```
PS C:\Projet-Integration-des-modeles-IA> & C:/Projet-Integration-des-modeles-IA/.venv/Scripts/python.exe c:/Projet-Integration-des-modeles-IA/train_model.py
Nettoyage des textes du fichier fake.csv en cours...
Nettoyage des textes du fichier true.csv en cours...
Dataset combiné nettoyé sauvegardé dans 'data/all_news_cleaned.csv'.
Création et sauvegarde du TF-IDF vectorizer...
Vectorizer sauvegardé dans 'model/tfidf_vectorizer.pkl'.
PS C:\Projet-Integration-des-modeles-IA> & C:/Projet-Integration-des-modeles-IA/.venv/Scripts/python.exe c:/Projet-Integration-des-modeles-IA/Normalisation_data.py
```

2

Normalisation

Casse, correction typos

```
TERMINAL
0 Donald Trump Sends Out Embarrassing New Year'... ... December 31, 2017
1 Drunk Bragging Trump Staffer Started Russian ... ... December 31, 2017
2 Sheriff David Clarke Becomes An Internet Joke... ... December 30, 2017
3 Trump Is So Obsessed He Even Has Obama's Name... ... December 29, 2017
4 Pope Francis Just Called Out Donald Trump Dur... ... December 25, 2017

[5 rows x 4 columns]
PS C:\Projet-Integration-des-modeles-IA> & C:/Projet-Integration-des-modeles-IA/.venv/Scripts/python.exe c:/Projet-Integration-des-modeles-IA/script.py
Hello, Esther ! Python 3.11 fonctionne bien.
PS C:\Projet-Integration-des-modeles-IA> & C:/Projet-Integration-des-modeles-IA/.venv/Scripts/python.exe c:/Projet-Integration-des-modeles-IA/train_model.py
Nettoyage des textes du fichier fake.csv en cours...
```

3

Vectorisation

TF-IDF, word embeddings

```
PS C:\Projet-Integration-des-modeles-IA> & C:/Projet-Integration-des-modeles-IA/.venv/Scripts/python.exe c:/Projet-Integration-des-modeles-IA/Normalisation_data.py
Nettoyage des textes du fichier fake.csv en cours...
Nettoyage des textes du fichier true.csv en cours...
Dataset combiné nettoyé sauvegardé dans 'data/all_news_cleaned.csv'.
Création et sauvegarde du TF-IDF vectorizer...
Vectorizer sauvegardé dans 'model/tfidf_vectorizer.pkl'.
```

Robotiser : Intégration Application

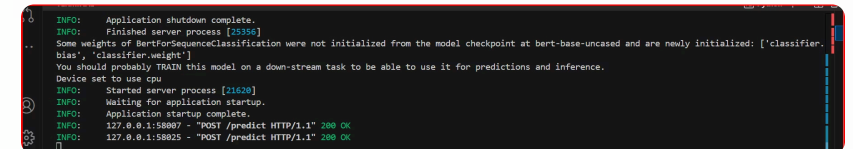
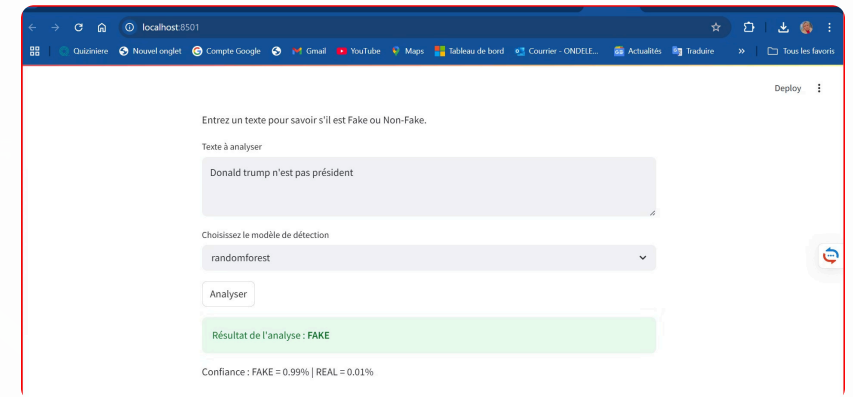
Plugin navigateur (web driver, selenium)

Extension StreamLit

Analyse articles en temps réel

Fast API

Analyse des textes provenant de données scrappées et des deux DataSets
(fake et true)



Généraliser : Tests et Validation

```
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:  
The secret `HF_TOKEN` does not exist in your Colab secrets.  
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings).  
You will be able to reuse this secret in all of your notebooks.  
Please note that authentication is recommended but still optional to access public models or datasets.  
warnings.warn(  
tokenizer_config.json: 100% ██████████ 48.0/48.0 [00:00<00:00, 3.40kB/s]  
vocab.txt: 100% ██████████ 232k/232k [00:00<00:00, 4.42MB/s]  
tokenizer.json: 100% ██████████ 466k/466k [00:00<00:00, 7.55MB/s]  
config.json: 100% ██████████ 570/570 [00:00<00:00, 37.0kB/s]
```

← → ↻ 🏠 ⓘ 127.0.0.1:8000 ☆

🗖 | 🌐 Quiziniere 🔄 Nouvel onglet 🌐 Compte Google 📧 Gmail 📺 YouTube 🗺 Maps 📊 Tableau de bord 📧 Courrier - ONDELE... 📰 Actualités 🗣 Traduire

Impression élégante ☐

```
{"message": "Fake News Detection API is running 🚀"}
```

```
✓ TERMINAL Python + Python ...  
🔗 5 liens extraits du flux : https://feeds.bbc1.co.uk/news/rss.xml  
✓ Article extrait : Trump tariffs: US orders from Chinese small businesses on hold  
✓ Article extrait : Toby Carvery owner admits cutting down ancient Enfield oak tree  
Created TensorFlow Lite XNNPACK delegate for CPU.  
✓ Article extrait : Birmingham City Council 'on track' to clear waste backlog  
✓ Article extrait : Rare footage from Sudanese city under siege  
Attempting to use a delegate that only supports static-sized tensors with a graph that has dynamic-sized tensors (tensor#58 is a dynamic-sized tensor).  
✓ Article extrait : Gene Hackman: Rodent nests found near house where Betsy Arakawa died  
🔗 0 liens extraits du flux : https://www.reutersagency.com/feed/?best-topics=politics  
📄 Résultats enregistrés dans 'results/rss_selenium_articles.csv'  
✓ Scrapping terminée !  
C:\Projet-Integration-des-modeles-IA>
```

1 Diversité des sources

Sites d'info, blogs, réseaux sociaux

2 Évaluation performance

Précision sur différents types de contenu

3 Ajustements

Affinage critères et modèle

Data-driven : Amélioration Continue

1

Feedback utilisateurs

Signalement fausses
classifications

2

Mise à jour dataset

Intégration nouvelles données

3

Réentraînement

Ajustement périodique du modèle

4

Analyse performance

Suivi indicateurs en temps réel

Statistiques des feedbacks





Conclusion et Perspectives

Solution intégrée

De la modélisation au
déploiement

Impact sociétal

Lutte contre la désinformation

Évolutions futures

IA explicable, multilingue