Projet CAF - Forum sur les Risques Cybernétiques et l'Intelligence Artificielle

# Description

Ce projet est un projet fictif qui a pour objectif de fournir aux équipes CAF un outil permettant d'identifier les tendances sur le thème 'Risques cybernétiques et Intelligence Artificielle : quelles stratégies de défense face aux nouvelles menaces numériques ?', à l'occasion de la 4e édition du forum.

L'outil extrait automatiquement des posts LinkedIn liés à des mots-clés spécifiques, les traite et les stocke dans une base de données PostgreSQL. Un tableau de bord a été créé dans Google Data Studio pour visualiser les données.

# Fonctionnalités

1. **Extraction de données LinkedIn** :

- Recherche automatique de posts liés à des mots-clés pertinents (ex. 'cybersécurité + Afrique') via Selenium et BeautifulSoup.  
- Extraction d'informations des posts comme le texte, la date, l'auteur, etc.

2. **Stockage des données** :

- Stockage des posts extraits dans une base de données PostgreSQL pour un usage futur.

3. **Visualisation des données** :

- Création de tableaux de bord interactifs dans Google Data Studio pour l'équipe CAF.

# Prérequis

Avant de commencer, assurez-vous d'avoir les éléments suivants :

- Un compte Linkedin  
- Python 3.x  
- Jupyter Notebook ou tout autre éditeur de code  
- Google Data Studio (pour visualiser les données)

# Installation

## 1. Cloner le dépôt

**bash**  
git clone https://github.com/votre-utilisateur/projet-caf.git

## 2. Créer un environnement virtuel et l'activer

**bash**python -m venv venv  
source venv/bin/activate # Sur macOS/Linux  
venv\Scripts\activate # Sur Windows

## 3. Installer les dépendances

**bash**  
pip install -r requirements.txt

## 4. Configurer les variables d'environnement

Créez un fichier `.env` à la racine du projet et y ajouter les informations de configuration, comme l'exemple suivant :  
**bash**  
# Fichier .env  
DB\_HOST=localhost  
DB\_NAME=your\_db\_name  
DB\_USER=your\_username  
DB\_PASSWORD=your\_password  
IN\_USERNAME=your\_linkedin\_usernameoremail  
PASSWORD=your\_linkedin\_password  
MOTS\_CLES=exemple (cybersécurité,IA,Afrique)

## 5. Configuration de la base de données

Créez une base de données PostgreSQL locale ou utilisez un service gratuit comme <https://aiven.io/> pour stocker les données. Vous pouvez adapter les informations de connexion dans le fichier .env.

## 6. Lancer le script d'extraction

Vous pouvez exécuter le notebook dans Jupyter Notebook ou le script Python classique pour commencer l'extraction des posts LinkedIn en fonction des mots-clés.  
  
Via **Jupyter notebook**

Ou directement via le fichier python :  
**bash**  
python getdata.py

Lorsque vous exécutez le script, vous devrez valider la connexion à votre compte LinkedIn. Vous avez deux options :

1. **Validation par l'application :** Si vous choisissez cette option, valider d’abord sur votre téléphone et une fois cela fait il vous suffira d'appuyer sur la touche **Entrée** dans le champ du script.
2. **Validation par SMS :** Si vous optez pour cette méthode, entrez sms dans le premier champ , puis dans le deuxième champ suivant saisissez le code à six chiffres que vous avez reçu par SMS puis valider et c’est tout le reste se fera de manière automatique.

# Utilisation

- Extraction des données : Le script utilise Selenium et BeautifulSoup pour récupérer des posts LinkedIn basés sur des mots-clés. Les données sont ensuite stockées dans PostgreSQL.  
- Accès aux données : Les données peuvent être visualisées via le tableau de bord Google Data Studio connecté à la base de données.

# Dépendances

Voici les principales dépendances utilisées dans ce projet :  
  
- selenium  
- webdriver-manager  
- beautifulsoup4  
- psycopg2-binary  
- python-dateutil  
Voir le fichier requirements.txt pour une liste complète.

# Tableau de bord

Les données extraites peuvent être visualisées sur un tableau de bord interactif via Google Data Studio. Voir ce lien : <https://lookerstudio.google.com/reporting/c51a22c4-d73d-45ac-87af-8d38fe0b7e84>

# Auteur

Josué Kouassi  
email : [josue.kouassi@epitech.eu](mailto:josue.kouassi@epitech.eu)

Contact : +225 0703743457