Analyse des Tendances Cybernétiques et d'Intelligence Artificielle

# Résumé du Projet

Ce projet vise à fournir à l'équipe CAF un outil d'analyse des tendances en matière de cybersécurité et d'Intelligence Artificielle. L'outil utilise Selenium pour extraire des posts LinkedIn en lien avec des mots-clés pertinents, stocke les données dans une base PostgreSQL sur Aiven, et présente les résultats sous forme de tableaux de bord interactifs avec Google Studio.

# Table des Matières

1. Introduction  
2. Prérequis  
3. Installation  
4. Utilisation  
5. Architecture du Projet  
6. Explications Techniques  
7. Data Modeling  
8. Dashboards et Visualisation  
9. Tests  
10. Conclusion  
11. Licence

# Introduction

Dans le cadre de la 4e édition du forum sur les "Risques cybernétiques et Intelligence Artificielle", l'équipe CAF souhaite comprendre les tendances récentes dans ces domaines. Ce projet consiste à construire un outil qui extrait automatiquement des posts LinkedIn relatifs à des mots-clés spécifiques, à traiter et stocker ces informations dans une base de données, et à générer des tableaux de bord pour aider à l'analyse des données.

# Prérequis

• Python 3.8 ou supérieur  
• Jupyter Notebook  
• Selenium et BeautifulSoup  
• Un compte Aiven avec une instance PostgreSQL  
• Un compte Google pour Google Studio

# Installation

1. Clonez le dépôt :  
```  
git clone https://github.com/Josuek1268/projet-CAF.git  
```  
2. Installez les dépendances Python nécessaires :  
```  
pip install -r requirements.txt  
```  
3. Connectez-vous à votre instance PostgreSQL sur Aiven :  
• Créez une instance Aiven PostgreSQL  
• Récupérez les identifiants de connexion  
• Ajoutez-les dans le fichier de configuration `.env`  
4. Configurez Selenium :  
• Téléchargez le driver Chrome adapté à votre version de navigateur.  
• Placez-le dans le répertoire de travail et mettez à jour son chemin dans le script.

# Utilisation

1. \*\*Exécution de l'extraction des données LinkedIn\*\* :  
• Ouvrez Jupyter Notebook et exécutez le fichier `data\_extraction.ipynb`.  
• Le script utilisera Selenium pour se connecter à LinkedIn et extraire les posts pertinents.  
2. \*\*Stockage des données dans la base PostgreSQL\*\* :  
• Les données seront formatées et envoyées vers la base de données PostgreSQL sur Aiven.  
3. \*\*Visualisation avec Google Studio\*\* :  
• Utilisez Google Studio pour vous connecter à votre base de données Aiven et créer des tableaux de bord.

# Architecture du Projet

/projet-CAF/  
│  
├── /notebooks/ # Dossier contenant les notebooks Jupyter  
│ ├── data\_extraction.ipynb # Script d'extraction des données LinkedIn  
├── /scripts/ # Dossier contenant les scripts Python  
│ ├── database.py # Script pour la connexion et le stockage dans PostgreSQL  
├── /dashboards/ # Dossier pour les fichiers de configuration des dashboards Google Studio  
├── requirements.txt # Liste des dépendances Python  
├── README.md # Documentation du projet

# Explications Techniques

1. \*\*Extraction des Données avec Selenium et BeautifulSoup\*\* :  
• Selenium a été utilisé pour automatiser la navigation sur LinkedIn et extraire des posts.  
• BeautifulSoup a analysé le HTML pour extraire les informations.  
2. \*\*Stockage des Données dans PostgreSQL\*\* :  
• Les données sont nettoyées et stockées dans une base de données PostgreSQL.  
3. \*\*Visualisation avec Google Studio\*\* :  
• Google Studio permet de connecter la base de données et de créer des visualisations interactives.

# Étude Comparative des Techniques de Data Modeling

1. \*\*Star Schema\*\* : Simplicité de mise en œuvre, performant pour des requêtes simples. Moins flexible pour des requêtes complexes.  
2. \*\*Snowflake Schema\*\* : Normalisation des données, stockage plus efficace, mais requêtes plus complexes.  
3. \*\*Data Vault 2.0\*\* : Flexibilité et évolutivité, mais complexité de mise en œuvre.

# Dashboards et Visualisation

Les tableaux de bord ont été créés avec Google Studio pour une présentation interactive des tendances sur LinkedIn en matière de cybersécurité et d'intelligence artificielle. Les graphiques incluent les tendances des posts par mot-clé et des analyses temporelles.

# Tests

• \*\*Tests de l'extraction\*\* : Vérification que le script récupère correctement les posts et que les données sont bien stockées.  
• \*\*Tests de performance\*\* : Évaluation du temps de traitement des données et des performances des tableaux de bord.

# Conclusion

Ce projet a permis de construire un outil d'extraction, de traitement et de visualisation des tendances en cybersécurité et IA. Les tableaux de bord créés avec Google Studio offrent une vue claire des tendances sur LinkedIn, aidant ainsi l'équipe CAF dans leur analyse.

# Licence

Ce projet est sous licence MIT.