

# Analyse et Prédiction des Offres d'Emploi Frauduleuses

Projet Data Sciences

ESPRIT 2024

# Projet

Développer une solution complète pour analyser et prédire les offres d'emploi frauduleuses.



Développer une solution complète pour analyser et prédire les offres d'emploi frauduleuses.



## Offres d'emploi frauduleuses

sont des annonces d'emploi trompeuses créées dans le but de duper les chercheurs d'emploi. Elles peuvent avoir diverses formes, telles que des escroqueries financières, des offres non existantes, ou des pratiques illégales.

# Statistiques sur les Offres d'Emploi au Monde

## Données Globales (2022)

- ▶ 1,9 milliard de personnes en emploi dans le monde.
- ▶ Taux de chômage mondial : 6,3% (baisse de 0,5% par rapport à 2021).

## Pays Développés vs. Pays en Développement

- ▶ Croissance forte dans les pays développés.
- ▶ Taux de chômage aux États-Unis : 3,9% (novembre 2023).
- ▶ Taux de chômage en Chine : 3,6% (octobre 2023).

# Sources de Recrutement

- Sites d'emploi en ligne.
- Agences de recrutement.
- Réseaux sociaux.

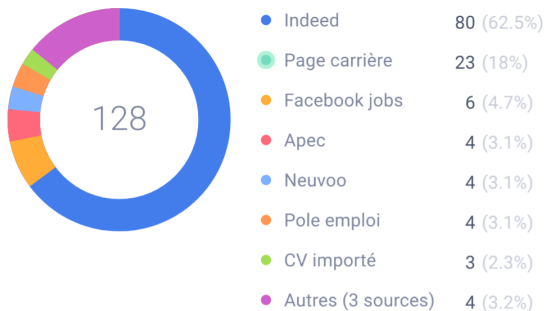


Figure – source : wererecruit.com

# Offres d'Emploi Frauduleuses

Selon une étude récente, environ 10% des offres d'emploi en ligne peuvent être considérées comme frauduleuses.



## Les secteurs les plus touchés incluent les opportunités de travail

- ▶ à domicile,
- ▶ les emplois hautement rémunérés
- ▶ les offres nécessitant un investissement initial.
- ▶ ...

La cybercriminalité liée aux faux emplois a augmenté de plus de 25% au cours des deux dernières années.

# Statistiques sur les Offres d'Emploi Frauduleuses

Selon la Federal Trade Commission (FTC) des États-Unis (2022)

- ▶ Victimes : Perte moyenne de 1 200 dollars.
- ▶ Troisième type d'escroquerie en ligne le plus courant.
- ▶ Plus de 35 000 plaintes d'escroqueries à l'emploi.
- ▶ Toutes catégories d'âges et milieux socio-économiques.
- ▶ Évidentes : Salaires ou avantages exorbitants.
- ▶ Subtiles : Frais de candidature, demandes d'informations sensibles.



Il est important de savoir identifier/analyser et prédire les offres d'emploi frauduleuses.

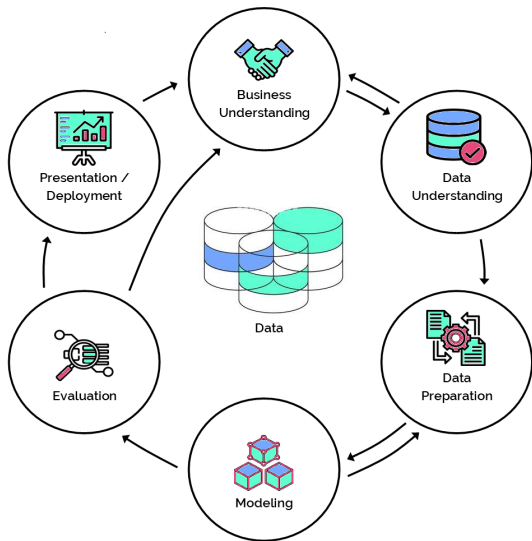
## Travail demandé

- **Objectif** : Développer une solution complète pour analyser et prédire les offres d'emploi frauduleuses.
- **Approches** : Data science, machine learning, analyse de sentiment.
- **Implique** : Développement d'une interface utilisateur, déploiement du modèle prédictif, validation continue.





# Méthode de travail : Cross-Industry Standard Process for Data Mining (CRISP-DM)



La méthode CRISP-DM est itérative, ce qui signifie que les phases peuvent être revisitées en fonction des résultats obtenus

# planning

## Sprint 1 : Data Scraping

**Durée :** Semaines 1-2

**Etape 1**

**Technologies :**

BeautifulSoup



# planning

## Sprint 1 : Data Scraping

**Durée :** Semaines 1-2

**Etape 1**

**Technologies :**

BeautifulSoup



Scrapy  
Python Library



SELENIUM  
with  
Python

**Etape 2**

- ▶ Utiliser SQL server integration service (SSIS) pour automatiser le processus de collecte de données.
- ▶ Mise en place de flux de données pour extraire, transformer et charger (ETL) les informations dans la base de données.



Microsoft®  
**SQL Server®**  
Integration Services

**Durée :** Semaines 3-4

Il y a plusieurs techniques statistiques et visuelles peuvent être utilisées pour comprendre la distribution des données et identifier des caractéristiques significatives

**Technologies :**



# Planning

## Sprint 3 : Modélisation

**Durée :** Semaines 5-6

création, entraînement et évaluation de modèles prédictifs basés sur les données disponibles...

**Technologies :**



# Planning

Sprint 4 : Analyse de Sentiment

**Durée :** Semaine 7

Compréhension approfondie des nuances émotionnelles présentes dans les descriptions d'offres d'emploi.

Utilisation de techniques de traitement du langage naturel ...



**POSITIVE**



**NEUTRAL**



**NEGATIVE**

# Planning

Sprint 5 : Développement de l'Interface Utilisateur

**Durée :** Semaines 8-9

Conception et implémentation...

Intégration de fonctionnalités interactives...

**Technologies :**

The Django logo, featuring the word "django" in a bold, dark green, lowercase sans-serif font.The Flask logo, which includes a black silhouette of a flask on the left and the word "Flask" in a large, black, serif font on the right. Below "Flask" is the tagline "web development, one drop at a time" in a smaller, black, sans-serif font.The Plotly Dash logo, consisting of a blue square icon with white vertical bars and dots on the left, followed by the text "plotly | Dash" in a blue, sans-serif font.The Streamlit logo, featuring a red, stylized crown icon on the left and the word "Streamlit" in a large, dark blue, sans-serif font on the right.

# Planning

Sprint 6 : Déploiement sur une Plateforme

**Durée :** Semaine 10





# Livrables

- ▶ Rapport de data scraping avec les données collectées.
- ▶ Rapport d'analyse exploratoire des données.
- ▶ Modèle de prédiction des fake jobs avec un rapport d'évaluation.
- ▶ Intégration de l'analyse de sentiment dans le modèle.
- ▶ Interface utilisateur fonctionnelle.
- ▶ Modèle déployé sur une plateforme avec documentation.