



# HASSAN Youssef

Data Scientist / Analyst

En recherche d'un CDI

## Certificats

Décembre 2021- Microsoft Certified Data Analyst Associate - Power BI

Mai 2021- Data Analysis with Power BI, by Reach Academy and D CUBED (Canada).

## Expérience professionnelle

- Groupe BPCE (Bred Banque Populaire) Juillet 2023 - Septembre 2024  
Data Scientist / Analyst

Au sein de l'équipe Support & Prévention pour Fraudes Externes et sécurité du SI:

- Développement des modèles ML visant à optimiser la détection de la fraude.
- Responsable de l'exploration, de l'analyse et de la manipulation des données sur une base de données **SQL**.
- Analyse des **données massives** concernant l'atypisme des prélèvements SEPA en se basant sur des **séries temporelles** (via **PySpark** en utilisant **SQL**, et **ARIMA** pour les prédictions/analyse)
- Création de **Power BI** Dashboards pour l'analyse des performances.
- Environnement: **Dataiku, Python, Spark**.
- **Résultats:**
  - Développement d'une approche originale pour résoudre le grand problème de **déséquilibre des données**.
  - Mise en production de 2 modèles (**XGBoost** et **LightGBM**) en temps réelle sur des transactions journalières après un feature engineering extensive avec un anomaly score via **Isolation Forest** et une segmentation via **CAH**.

- NEUMANN AI Juillet - Décembre 2021  
Data Scientist - Stage

**Objectif:** Développer un outil d'Automatic Clustering pour la Segmentation Client:

- Extraire un échantillon aléatoire représentatif avec une approche automatique via python.
- Appliquer automatiquement le meilleur algorithme de clustering.
- Analyser les features les plus influentes et leurs poids pour chaque cluster.

- CMA CGM Mai - Juillet 2021  
Data Analyst - Stage

- Introduction de **Microsoft Power BI** dans l'équipe de finance.
- Responsable des requêtes **SQL** pour manipuler les données.

## Education

- Université Paris Cité 2022-2024  
Master - Machine Learning for Data Science

- Plusieurs projets de Machine Learning ( Python et R ) :
  - Classification supervisée : Regression - KNN - Random Forest - XGBoost
  - Classification non-supervisée : Clustering (Mixture Models dont K-Means - CAH - DBSCAN etc ...)
  - dimensionality reduction: PCA - UMAP- T-SNE - AFC- AFCM- AFDM ...
- Deep learning: CNN - RNN - LLM - LSTM - GRU - GAN, un Projet de self/semi-supervised learning avec **pytorch**.
- **NLP:** spaCy, NLTK, Embeddings ( word2vect, glove, jose)
- **Big Data Frameworks** (Hadoop mapreduce- Apache cassandra, Spark), **programmation distribuée** (infrastructure cloud **AWS/ GCP**, clusterisation de conteneurs Docker avec Kubernetes), **Linux/Git**.

- Université Saint-Joseph de Beyrouth 2019-2022  
Licence - Mathématiques appliquées à la science des données

Major de promotion avec une moyenne de 89.17/100

## Contact

### Numéro de téléphone

+33 6 56 69 77 06

### Email

youssefahassan1211@gmail.com

### Adresse

Paris, France

### Site web

in [youssefahassan12](#)

Q [YoussefHassan12](#)

## Compétences IT

- Python  
scikit-learn, Pandas, Numpy, pytorch, statsmodel, Matplotlib, Seaborn, Pyspark
- R
- C++
- Dataiku  
Importation des sources, Gestion de tables de fait, Nettoyage, Analyse et ML
- Microsoft Power BI
- SQL/PostgreSQL
- Linux
- Git
- Excel, Word, Power Point

## Langues

Anglais - TOEFL 101/120

Français - C1

## Divers

- Major de promotion en Licence
- Élu vice-président du conseil d'étudiants du campus Liban Nord de l'USJ (2020/2021).
- Président du Club de l'ONG Donner Sans Compter (2020/2021).
- 84 heures de volontariat au sein du DSC.
- Échecs – Guitariste – Footballeur