

# BERKANI Yacine



## À LA RECHERCHE D'UN STAGE EN DATA

### CONTACT



+33 0757 77 59 08



yacineberkani32@gmail.com



[Linkedin](#)



Région Île-de-France

### COMPÉTENCES

#### Gestion de projet agile

- Jira, Git

#### Langages de programmation

- Python, R, SQL, Scala, Java, C

#### Visualisation de données et BI

- PowerBI, Tableau

#### Outil ETL

- Talend Open Studio, Logstash

#### Machine Learning/AI

- Apprentissage supervisé/ Non supervisé/ Par renforcement
- IA générative/ NLP/ Prompt Engineering

#### BigData

- Hadoop, HDFS, MapReduce, YARN, Hive, Spark.

#### Plateforme cloud

- Amazon web service (AWS)
- Google Cloud Platform (GCP)

#### Bases de données

- SQL (Mysql, PostgreSQL, MariaDB)
- NoSQL( MongoDB, Cassandra)

#### Langues

- Français
- Anglais
- Kabyle

### PROFIL

- **Diplôme préparé:** Master 2 en Big Data à l'Université Paris8. de février.
- **Rythme :** Une semaine à l'escale et une semaine en entreprise, puis à prtir de février, temps plein en entreprise.

### EXPÉRIENCE PROFESSIONNELLE

#### Stage Data Scientist/Gen AI , Laboratoire Paragraphe Paris 8

03/2024 – 08/2024 | Saint-Denis

- Scraping (beautifulsoup, request, NER, SpaCy)
- Génération de texte (RAG, Fine Tuning, LLM, Hugging Face)

#### Stage Data Analyst, Traitement/Visualisation, IRIS

04/2023 – 07/2023 | Paris

- **Extraire et nettoyer les données** à l'aide d'**Excel**.
- Transformer les données.
- Créer des tableaux de bord interactifs avec Power BI.

#### Stage Data Engineering, Architecture Big Data , E N I E M

04/2022 – 09/2022 | Paris

- Architecture distribuée cluster Hadoop sur **GCP, Dataproc**
- Architecture pseudo distribuée cluster Hadoop sur **Docker**

### FORMATIONS

#### Master Big Data, Université Paris 8

09/2023 – 09/2025 | Saint-Denis

- Machine Learning, réseaux de neurones, Big Data.
- Traitement de données massives avec Hadoop et Spark

#### Licence en Mathématiques Informatique, UMMTO

09/2020–09/2023|Tizi Ouzou

- Maîtrise des algorithmes et des structures de données.
- Création, gestion et optimisation de bases de données.
- Analyse de données
- Modélisation statistique appliquées à la science des données.

### PROJET

#### Détection de Tumeurs Cérébrales à partir d'IRM

- **Modèles utilisés :** CNN, VGG16, Random Forest.
- **Outils :** Python, scikit-learn, TensorFlow, Keras, PyTorch

#### Analyse de Sentiment sur les Tweets

- **Techniques :** NLP, Machine Learning, Deep Learning
- **Modèles utilisés :** RNN, LSTM, BLSTM

#### Prédire le prix de l'immobilier

- **Techniques :** Régression, Machine Learning
- **Modèles utilisés :** Régression Linéaire, Régression Non Linéaire, Random Forest Regressor