

## Mohamed KEITA

Data Scientist | ML Engineer | Data Analyst | Data Engineer | Python Developer  
0033638141242 | contact@mohamedkeita.pro | mohamedkeita.pro

### COMPÉTENCES

<b>Machine learning</b>	Tensorflow, Pytorch, Scikit-Learn
<b>Big Data</b>	Hadoop, Spark, Kafka, Databricks
<b>Développement</b>	Python, JavaScript
<b>ETL / BDD</b>	Talend, MySQL, PostgreSQL, MongoDB
<b>Cloud</b>	GCP, Google BigQuery, AWS (S3, Sagemaker)
<b>Génie Logiciel</b>	Conteneurisation (Docker), CI/CD, APIs, Airflow
<b>Web</b>	HTML, CSS, Django, Flask, FastApi
<b>Autres</b>	Git, PowerBI, Tableau, Microsoft Office
<b>Langues</b>	Français, Anglais, Espagnol

### EXPERIENCE PROFESSIONNELLE

**Data Analyst (CDI)** Depuis Septembre 2024  
SOS Oxygène Nice, France

- Automatisation de toutes les tâches administratives répétitives
- Création de requêtes SQL (PostgreSQL, SQL Server) pour optimiser certains processus
- Création de script de traitement de données en PowerBI DAX
- Création de tableaux de bord pour les agences métier

**Environnement technique:** Python, Windows, PostgreSQL, SQL Server, PowerBI

**Senior Data Engineer (CDI)** Mars 2024 - Août 2024  
SpikeeLabs Nice, France

- Participation à la création du back-end de l'intranet en Django
- Mettre en place un ETL pour récupérer les données RH et les charger dans une base PostgreSQL, source de l'intranet
- Mise en place d'une automatisation de l'ETL
- Mise en place des tableaux de bords avec PowerBI pour afficher les KPI dans l'intranet
- Configurer les projets et services Google Cloud (Cloud SQL/BigQuery), en veillant à mettre en place des mesures de sécurité robustes.
- Adapter et transférer le schéma de base de données et les données à l'aide d'outils ETL.
- Effectuer une validation complète et des tests de performance pour garantir l'intégrité des données et des performances optimales.

**Environnement technique:** Python, Django, SQL, Git, Windows, GCP, PostgreSQL, Scrum, PowerBI

**Data Science Teacher (Freelance)** Déc 2023 - Déc 2023  
Le Wagon Nice Nice, France

- Dispense de cours sur le workflow MLOps
- Encadrer et accompagner les étudiants du BootCamp Data Science dans le cadre de leur projet
- être un référent technique. Assister les étudiants dans le développement des modèles de machine learning, leur mise en production et leur déploiement.

**Encadrant Projet de fin d'études (Vacataire)** Oct 2023 - Fév 2024  
Université Côte d'Azur Nice, France

- Encadrer et accompagner un groupe de trois étudiants en deuxième année de Master Ingénierie Mathématique dans le cadre de leur projet de fin d'études.
- Guider les étudiants à travers le processus et les étapes nécessaires pour mener une recherche scientifique. Les aider à définir un objectif, une problématique et à formuler des hypothèses.
- Organiser des réunions régulières en fonction de l'avancement du projet et des besoins des étudiants.

- être un référent technique et théorique. Superviser les étudiants dans le développement des modèles mathématiques nécessaires à la résolution des problèmes et les soutenir dans la rédaction d'un rapport scientifique.

### Consultant Data Engineer

Travel Planet

Oct 2023 – Dec 2023

Valbonne, France

- Identifier parmi les factures fournisseurs et factures ventes, les différentes incohérences
  - Création d'une pipeline qui permet de récupérer les factures depuis la boîte mail, de gérer les erreurs de facturation puis de stocker les lignes corrigées dans une base de données MySQL
  - Automatisation de cette pipeline pour permettre son exécution dès la réception de nouvelles factures dans la boîte mail
- Création d'une interface graphique en Python Flask pour permettre à la comptabilité de gérer manuellement les erreurs nécessitant une intervention humaine.

**Environnement technique:** Python, Flask, SQL, Git, MacOS, AWS (S3), MySQL, Cassandra, Scrum

### Consultant Data Engineer / Data Scientist

MyHotelMatch

Avril 2023 – Oct 2023

Biot, France

Participation au développement d'une application de matching entre voyageurs et hôtels

- Participation au développement d'une application de matching entre voyageurs et hôtels
- Développement de l'algorithme de matching, recommander une liste d'hôtels à l'utilisateur en fonction de son profil voyageur
- Développement d'un système de recommandation avec pondération
- Mise en production des algorithmes et leur déploiement
  - Création d'une API (Routes et Endpoints)
  - Tests unitaires, puis tests d'intégration en continu
  - Déploiement en continu
  - Evaluation du modèle en continu
- Maintenance de la base de données MongoDB
  - Création d'une pipeline qui permet récupérer des données d'hôtels d'une part depuis le backoffice (base de données PostgreSQL) et d'autre part les disponibilités des chambres depuis l'API d'Amadeus.
  - Stocker toutes ces données dans la base de données MongoDB, base de données source pour l'entraînement des différents modèles
  - Veiller à mettre à jour la base de données à chaque changement de disponibilité des chambres
- Création de tableaux de bords (Dashboards) avec Tableau
- Construction de nouveaux indicateurs de performance (KPI)
- Indexation des données dans une base de données PostgreSQL.

**Environnement technique:** Python, Scikit-learn, Tensorflow, SQL, NoSQL, FastAPI, Git, Linux, GCP, Kubernetes, PostgreSQL, MongoDB, Agile, Tableau

### Data Scientist, CDI

Skema Business School

Jan 2020 – Mars 2023

Biot, France

- Exploitation en architecture big data de la base de données mondiale des brevets (PATSTAT) regroupant plus de 500 millions d'observations
- Identification des brevets en relation avec l'intelligence artificielle en analysant les titres et résumés des brevets grâce à des méthodes de text mining et de traitement de langage naturel
  - Création d'une pipeline en PySpark pour récupérer les données depuis le disque dur, d'effectuer toutes les transformations nécessaires à l'identification des brevets en relation avec l'IA
  - Utilisation de la librairie MLib pour appliquer des modèles sur les données dans Spark
  - Stocker les brevets identifiés IA dans une base de données SQLite

- Classification des brevets dans les différentes catégories d'intelligence artificielle
- Analyse exploratoire pour déterminer les technologies et acteurs clés du domaine de l'intelligence artificielle
- Publication d'un rapport scientifique et d'un policy brief sur l'intelligence artificielle.

**Environnement technique:** Python, SQL, Spark, Hadoop, Git, Linux, Windows, SQLite, Scrum, PowerBI

## **FORMATION**

---

### **Master Mathématiques Appliquées et Statistiques**

Aix Marseille Université

Sep 2017 – Oct 2019

Marseille, France

### **Licence Mathématiques générales (options Informatique)**

Aix Marseille Université

Sep 2012 – Jan 2016

Marseille, France