



# Omar KHEMMI

## Data Scientist & AI Engineer

☎ +212 7 71 76 48 81 ✉ khemmiomar1@gmail.com

🌐 omarkhemmi 🌐 Omar KHEMMI 🌐 My Portfolio

Étudiant en dernière année de Master en Informatique et Data Science, passionné par la transformation des données en informations exploitables. Compétences en **apprentissage automatique**, **prétraitement des données** et **visualisation des données**. Expérimenté en **Python**, **SQL** et **Power BI**. Maîtrise de **Hadoop** et **MongoDB** pour l'analyse de Big Data. **Recherche un stage de fin d'études (PFE)** pour appliquer des solutions basées sur les données et approfondir ses compétences en **Intelligence Artificielle** et **Data Engineering**.

## FORMATION

<b>Mathématiques et Informatique pour la Data Science</b> , Master	2024 - 2026
<i>ENSA Khouribga, Université Sultan Moulay Slimane</i>	Khouribga
<b>Informatique Décisionnelle</b> , Licence Professionnelle	2023 - 2024
<i>École Supérieure de Technologie (EST), Université Mohammed Premier</i>	Oujda
<b>Génie Informatique</b> , Diplôme Universitaire de Technologie (DUT)	2021 - 2023
<i>École Supérieure de Technologie (EST), Université Cadi Ayyad</i>	Safi

## EXPÉRIENCE

<b>Ingénieur Logiciel - IA NLP</b>   Ensa – Khouribga (2 mois)	Juil 2025 – Aoû 2025
<ul style="list-style-type: none"><li>Création d'une <b>pipeline complète</b> pour l'entraînement de <b>Small Language Models (SLM)</b> sur des jeux de données spécifiques.</li><li>Application de <b>LoRA / PEFT</b> et <b>quantification</b> pour optimiser les performances et comparaison avec de grands LLM.</li><li>Livraison d'un <b>SLM prêt pour la production</b> avec scripts, documentation et rapports de performance.</li></ul>	
<b>Développeur Full Stack</b>   ADS – Rabat (3 mois)	Avr 2024 – Juin 2024
<ul style="list-style-type: none"><li>Développement d'une <b>plateforme digitale</b> pour l'intégration sociale des femmes avec <b>visualisations BI interactives</b> et <b>pipeline ETL</b>.</li><li>Développement front-end avec <b>Next.js</b> et back-end avec <b>Node.js</b>, assurant un flux de données fluide et interactif.</li><li>Intégration de <b>tableaux de bord Power BI</b> pour des insights en temps réel, fournissant une solution évolutive pour la prise de décision basée sur les données.</li></ul>	
<b>Développeur Web Full Stack</b>   Devosoft – Safi (2 mois)	Avr 2023 – Mai 2023
<ul style="list-style-type: none"><li>Conception et mise en œuvre d'une <b>plateforme d'annuaire web</b> utilisant <b>Laravel</b>, <b>HTML</b>, <b>CSS</b>, et <b>JavaScript</b>, améliorant l'accessibilité et l'organisation du contenu.</li></ul>	

## PROJETS

### WordCount Distribué sur Hadoop (Python Streaming)

- Implémentation d'un programme **WordCount MapReduce** avec **Python Streaming** sur **Hadoop**.
- Conception de scripts mapper et reducer pour traiter de grands ensembles de texte distribués.
- Expérimentations sur **Cloudera VM** pour analyser la scalabilité et explorer les concepts **Big Data** via des TPs sur **Pig**, **HDFS** et **MapReduce**.

### Système d'Extraction de Connaissances d'Articles Scientifiques

- Développement d'un **système de data science** pour collecter et analyser les publications scientifiques (**arXiv**, **PubMed**, **Semantic Scholar**).
- Utilisation de **NLP** pour extraire entités et relations entre auteurs, institutions et concepts.
- Création d'un **tableau de bord interactif (Streamlit, Neo4j)** pour visualiser graphes de connaissances et tendances.

### Plateforme de Prédiction de Matches de Football

- Création d'un **système de machine learning** pour prédire les résultats des matches de football à partir des données pré-match et de performance des équipes.
- Comparaison et optimisation des modèles **Régression Logistique**, **Random Forest** et **XGBoost** via validation croisée.
- Conception d'une **interface web interactive** pour la prédiction en temps réel et la visualisation des performances.

## TECHNOLOGIES

**Langages Core** : Python, Java, JavaScript, SQL, POO

**Data Science ML** : Pandas, NumPy, Scikit-learn, TensorFlow, Matplotlib, Seaborn, XGBoost

**Data Engineering Big Data** : Power BI, Hadoop, Spark, Airflow, Pipelines ETL, Entrepôt de données (SQL Server)

**Bases de données Stockage** : MySQL, PostgreSQL, MongoDB, Firebase, Neo4j