Omar KHEMMI

Data Science Student

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

 omarkhemmi   Omar KHEMMI

Final-year Master's student in Computer Science and Data Science, passionate about transforming data into valuable insights. Skilled in **machine learning**, **data preprocessing**, and **data visualization**. Experienced with **Python**, **SQL**, and **Power BI**. Proficient in **Hadoop** and **MongoDB** for big data analytics. **Seeking a final-year internship (PFE)** to apply data-driven solutions and advance expertise in **Artificial Intelligence** and **Data Engineering**.

EDUCATION

Mathematics and Computer Science for Data Science, Master's Degree	2024 - 2026
ENSA Khouribga, Sultan Moulay Slimane University	Khouribga
Decision Support Systems (Business Intelligence), Professional Bachelor's Degree	2023 - 2024
Higher School of Technology (EST), Mohammed First University	Oujda
Computer Engineering, University Technology Diploma (DUT)	2021 - 2023
Higher School of Technology (EST), Cadi Ayyad University	Safi

EXPERIENCE

Software Engineer - AI NLP | Ensa – Khouribga (2 months)   Jul 2025 – Aug 2025
- Built a **full pipeline for Small Language Model (SLM) fine-tuning** on domain-specific datasets.
- Applied **LoRA / PEFT and quantization** to optimize performance and benchmarked against larger LLMs.
- Delivered a **production-ready SLM** with scripts, documentation, and performance reports.

Full Stack Developer | ADS – Rabat (3 months)   Apr 2024 – Jun 2024
- Developed a **digital platform for women's social integration** with **interactive BI visualizations** and an **ETL data pipeline**.
- Built front-end using **Next.js** and back-end with **Node.js**, ensuring seamless data flow and interactivity.
- Integrated **Power BI dashboards** for real-time insights, delivering a scalable solution to enhance data-driven decision-making.

Full-Stack Web Developer | Devosoft – Safi (2 months)   Apr 2023 – May 2023
- Designed and implemented a **web directory platform** using **Laravel, HTML, CSS,** and **JavaScript** improving user accessibility and content organization.

PROJECTS

Distributed WordCount on Hadoop (Python Streaming)
- Implemented a **MapReduce-based WordCount** program using **Python Streaming on Hadoop**.
- Designed and executed mapper and reducer scripts to process large text datasets in a distributed environment.
- Performed experiments on **Cloudera VM** to analyze the effect of multiple reducers and verify scalability.
- Explored **Big Data concepts** through practical TPs on **Pig, HDFS, and MapReduce** as part of the course .

Scientific Articles Knowledge Extraction System
- Developed a **data science system** to collect and analyze scientific publications from **arXiv**, **PubMed**, and **Semantic Scholar**.
- Applied **NLP techniques** for named entity extraction and relationship detection between authors, institutions, and research concepts.
- Built an **interactive dashboard** using **Streamlit** and **Neo4j** to visualize knowledge graphs and emerging research trends.

Football Match Prediction Platform
- Built a **machine learning system** to predict football match outcomes using pre-match and team performance data.
- Compared and optimized **Logistic Regression**, **Random Forest**, and **XGBoost** models through cross-validation.
- Designed an **interactive web interface** for real-time prediction and performance visualization.

TECHNOLOGIES

Languages & Core: Python, Java, SQL, JavaScript, OOP, Git

Data Science & ML: Pandas, NumPy, Scikit-learn, TensorFlow, XGBoost, Matplotlib, Seaborn, NLP, Hugging Face

Data Engineering & Big Data: Power BI, Hadoop, Spark, Airflow, ETL Pipelines, Data Warehouse (SQL Server)

Databases & Storage: MySQL, PostgreSQL, MongoDB, Neo4j