# Chammami Yahya

## Data Analyst - ML/AI Engineer

+216 46291123 | Chammemiyahia@gmail.com | Linkedin | GitHub

## SUMMARY

Dynamic and solution-oriented Data Scientist with a robust background in statistics and machine learning. Experienced in building predictive models, orchestrating RAG flow systems, and developing LLM-based solutions. Skilled in Python, TensorFlow, and Power BI, with a passion for AI-driven insights and automation.

## EDUCATION

**Engineering Degree in Statistics and Information Analysis.**  Ariana, Tunisia
*Higher School of Statistics and Information Analysis Tunis,* **ESSAI**  *Aug. 2020 – May. 2025*

**Pre-Engineering Cycle, Mathematics and Physics.**  Bizerte, Tunisia
*Preparatory Institute for Engineering Studies Bizerte,* **IPEIB**  *Sep. 2020 – Jun. 2022*

## EXPERIENCE

**AI Engineer, Intern**  Jun. 2024 – Sep. 2024
*Oratio*  *Tunis, Tunisia*
- Developed a **RAG** Inference pipeline using Ollama and FastAPI to generate response from Documents.
- Participated in code architecture and production deployment with Git, Docker and Azure AI Services.
- Used Sentence-Transformers embeddings and Qdrant vector database for document indexing and storage.
- Evaluated the performance of the Rag system and **Llama2** model with Ragas framework (Context relevancy, Context recall, Faithfulness, Answer relevancy).

**Data Scientist, Intern**  Jul. 2023 – Aug. 2023
*Bridges S.A*  *Tunis, Tunisia*
- Developed a **Web application** for scoring and fraud detection using Python and Machine Learning algorithms.
- Automated the analysis of medical insurance data for 4,000+ doctors and 2,200+ pharmacists.
- Designed a user-friendly graphic interface with **Dash** and **Plotly** for real-time insurance recovery tracking.

## PROJECTS

**LLM-Based Chatbot** | *Python, Flan-T5, LoRA, SFTTrainer, LangChain*
- Preprocessed and cleaned over 3.000 rows of data using Pandas and Regex to ensure accurate data input.
- Fine tuned a **Flan-T5** model using Low-Rank-Adaptation (**LoRA**) to enhance performance of Q&A system.
- Evaluated the model to mesure the relevance of generated responses of the LLM using Blue score metrics.

**Time Series Forecasting - Energy Consumption** | *SARIMA, SARIMAX, Prophet, RNN, CNN*
- Cleaned and preprocessed data and visualize Energy Consumption patterns over time.
- Engineered a benchmarking framework to evaluate statistical models against deep learning models over time.
- Created a feature importance analysis to understand which variables have the most impact on forecasting enhancing model interpretability.

**Sales Insights Analytics Dashboard** | *Python, HaggingFace, Streamlit*
- Designed a PowerBI Dashboard to understand company hardware goods sales trend.
- Maintained SQL queries to analyze user behavior and engagement metrics, enhancing user satisfaction.
- Created DAX functions to create visualizations and KPIs that highlighted key business trends and patterns.

## TECHNICAL SKILLS

**Language**: Python, SQL, R.
**Developer Tools**: Git, Docker, Flask, FastAPI.
**Concepts**: Machine Learning, Generative AI, RAG, Time Series, Applied Mathematics.
**Soft Skills**: Problem-Solving, Writing, Public Speaking, Leadership, Agile methodology.

## CERTIFICATIONS

**DataCamp Professional Data Scientist - DataCamp**
**IBM Data Science Professional Certificate - DataCamp**
**Deep Learning Specialization - Machine Learning Specialization - Coursera**
**Building RAG Agents with LLMs - NVIDIA**