Marouane Ait El Faqir

Senior Data Scientist | Machine Learning & Computer Vision Specialist

Toulouse, France • +33 6 60 18 50 15 • m.aitelfaqir@gmail.com
Website: marouaneaef.github.io

Professional Summary

Senior Data Scientist with **10+ years** of experience delivering machine learning solutions for space, defense, and industrial applications. Led **10+ data science projects and R&D studies**, from feasibility to deployment-oriented prototypes, using satellite imagery and sensor data. Strong expertise in **deep learning**, **computer vision**, **signal processing**, and **statistical & probabilistic modeling**, with solid experience in **numerical optimization** and **end-to-end ML pipelines**. Currently strengthening **MLOps & cloud** skills (MLflow, Docker, CI/CD, AWS) through an **advanced Machine Learning Engineer (MLOps & Cloud) program** at DataScientest (2025–2026), combining continuing education and hands-on projects.

Work Experience

Data Scientist — Analytics Lead
CS Group, Toulouse, France — December 2020 – July 2025

- Led **5+ data science projects and technical proposals** for CNES, DGA, and ESA Copernicus on large-scale satellite imagery and multi-sensor time-series, delivering high-accuracy anomaly detection and pattern recognition models.
- Architected and implemented image processing pipelines for remote sensing imagery (preprocessing, normalization, denoising, feature extraction) using **Python**, **NumPy**, **Pandas**, and **Scikit-learn**.
- Delivered ML models (classification, regression, clustering) and integrated **CNNs** and **autoencoders** for satellite imagery and time-series sensor data, improving detection quality compared to baseline heuristic methods, with fewer false positives and missed anomalies on operational datasets.
- Optimized model performance through numerical optimization and rigorous evaluation (cross-validation, ROC-AUC, precision, recall, F1-score), ensuring robust integration in production processing chains.
- Established end-to-end data workflows (data ingestion, ETL, feature engineering, training scripts) and created reproducible, deployment-oriented pipelines reused across multiple projects, progressively applying MLOps practices (model versioning, experiment tracking, containerization).

Data Analyst — Statistical Modeling & Signal Processing Specialist
CHAUVIN-ARNOUX, Annecy, France — March 2018 – December 2020

- Developed statistical models for embedded electronics and industrial measurement devices using Python, C, and C++ to characterize device performance, reliability, and measurement uncertainty.
- Developed signal processing algorithms (digital filtering, Fourier transform, spectral analysis) for current, voltage, and thermal signals, improving the stability and reliability of key measurements by filtering out noise and correcting long-term drift.
- Built ML-based anomaly detection systems for thermal images and sensor data, combining image processing and machine learning techniques to enable proactive maintenance and quality control.
- Designed probabilistic and regression models for measurement uncertainty and calibration errors, contributing to an R&D project for a portable current measurement device (**patent filed**). This work combined signal processing, statistical inference, and numerical optimization.

Data Scientist — Analytics Consultant
ALTRAN, Toulouse, France — October 2015 – March 2018

- Developed ML and statistical modeling solutions for Airbus R&D projects, implementing models in **Python** and **Scikit-learn** on aeronautical time-series and sensor data.
- Built predictive maintenance models on multivariate time-series from aeronautical systems (e.g. temperature, pressure, vibration) to support early fault detection and maintenance planning.
- Designed classification and regression models with feature engineering and PCA-based dimensionality reduction, validated with statistical tests for safety-critical applications.

Research Data Scientist — Statistical Methods & Numerical Optimization
CNRS / INRA, Lyon–Paris, France October 2012 – October 2015

- Researched and developed stochastic numerical optimization algorithms for large-scale biological systems with nonlinear dynamics and noisy experimental data.
- Applied Bayesian methods and statistical inference to analyze complex biological datasets and experimental time-series.
- Published research on optimization and statistical methodologies (e.g. **IEEE CDC**), implementing algorithms (stochastic gradients, global optimization) in Python and MATLAB for model calibration and inverse problems.

Education

Machine Learning Engineer (MLOps & Cloud) — Continuing Education (in progress)
DataScientest (remote) 2025 – 2026
Advanced program focusing on **industrialization of ML models**, **MLOps practices** (MLflow, Docker, CI/CD), and deployment on **cloud platforms** (AWS).

PhD in Applied Mathematics & Systems Biology
École Centrale de Lyon & INRA Paris, France (EEA doctoral school) 2012 – 2015
Specialization in **Numerical Stochastic Optimization**, **Probability Theory**, **Statistical Modeling** and **Systems Biology**.

Skills

Core Expertise

- **Machine Learning**: Supervised and unsupervised learning, deep learning (CNNs, autoencoders), time-series forecasting (ARIMA/SARIMA), pattern recognition, anomaly detection, model evaluation and optimization.
- **Image & Signal Processing**: Satellite imagery, remote sensing, computer vision, digital filtering, spectral analysis, feature extraction for ML pipelines.
- **Statistical Modeling & Optimization**: Probabilistic modeling, Bayesian inference, uncertainty quantification, numerical optimization for model calibration and performance tuning.

MLOps & Production

- Model versioning and experiment tracking concepts with **MLflow**; reproducible training pipelines and environment management.
- Containerization and serving: packaging ML models into **Docker** images and exposing them through REST APIs (**FastAPI**, **Flask**) for batch inference.
- CI/CD and cloud foundations: Git-based workflows (GitHub/GitLab), introduction to CI/CD for testing and packaging; basic usage of **AWS** services (**S3**, **EC2**), monitoring and logging concepts. *Currently consolidating these skills through a dedicated Machine Learning Engineer (MLOps & Cloud) training and personal projects.*

Technical Stack

Python (NumPy, Pandas, SciPy, Scikit-learn, Statsmodels, TensorFlow, PyTorch), **OpenCV**, scikit-image; Matplotlib, Seaborn, Jupyter Notebooks; Git, GitHub; SQL, HDF5, Parquet; Docker, FastAPI, Flask.

Languages

French: Native/Bilingual • English: Advanced (technical and professional) • Arabic: Fluent