ANASS MAJJI

anassmajji34@gmail.com | amajji (Anass Majji) · GitHub | + 33 6 23 45 85 37

Skills

Programming: Python | R | C / C++ | JavaScript | SQL | Spark | Matlab | VHDL
AI Librairies and Tools: PyTorch | TensorFlow | ONNX | DeepSpeed | Scikit-learn | Scrapy / Selenium | React | HTML5 | CSS3 | Flask / FastAPI | Streamlit | Git | Docker | Jenkins | Gitlab CI/CD | MLflow | DVC | OpenShift | Kubernetes | Prometheus | Grafana | AWS / DataiKu / Domino.
Languages: English – Fluent | French, Arabic – Native | German – Beginner.

Professional Experience

Data Scientist | Full-time job

BNP PARIBAS — *Paris - France* — 10/2022 – Current | **2 years and 5 months**

Within the **Data Analytics & Research Team**, I contributed to various key projects:

- **Question Answering LLM Model for Annual Report Analysis**
  - Led the end-to-end development and deployment of a **Question Answering LLM model** using **Mistral (7B parameters)** to automate data extraction from company annual reports, reducing team workload by 30%.
  - Utilized **GROBID API** for text extraction from PDFs and created a **vector database for RAG** with **Langchain**.
  - Optimized model inference performance with techniques such as **quantization**, **KV caching**, and **Flash attention**.
  - Deployed the model on a **Streamlit web application** for seamless user interaction.
- **Energy Performance Prediction Model for Real Estate**
  - Developed and deployed a **Machine Learning model** to predict building energy performance labels, achieving **92% F1-score** and saving **€250k for BNP Paribas**.
  - Conducted **unit tests**, monitored model performance, and adhered to best software development practices during deployment.
  - Deployed the model on a **FastAPI web interface** for end-users.
- **Teaching Assistant (CentraleSupelec)**
  - Collaborated with research students at **CentraleSupelec** to develop an advanced time-series **forecasting** model to predict CO2 emission intensities for commercial real estate assets within the BNP Paribas portfolio.
- **Conducted technical interviews for candidates to grow C2A team and mentored several new hires.**

Data Scientist | Internship + Full-time job

LCL Bank — *Paris - France* — 06/2020 – 10/2022 | **2 years and 4 months**

Within the **Data Factory's AI team**, I developed and deployed **4 Deep & Machine Learning models** in production, following best practices in unit testing, model monitoring, and software development.

- **Document Classification and Information Extraction**
  - Developed and deployed a **CamemBERT deep learning model** for automatic classification of client documents (e.g., ID cards, passports, Kbis) and extraction of key informations (e.g., surname, first name, date of birth, address) using **Pytesseract**.
  - Processed over **20 million documents** for classification and information extraction.
- **Fraud Detection Model**
  - Built and deployed a **fraud detection model** using **XGBoost** and a **FastAPI web interface**, achieving an **F1-score of 89%** and saving **€400k** for **LCL Bank**.
  - Extracted data using **SAS** and **SQL Teradata**, pre-processed it, and deployed the model on the **internal cloud platform** for real-time fraud detection.
- **Client Review Classification & Sentiment Analysis (Awarded Best Innovative Project)**
  - Scraped **client reviews from Google** using **Scrapy** and **Selenium**, then classified them into satisfaction topics using an **NLP model (CamemBERT)**.
  - Developed and deployed an interactive **Streamlit dashboard** for visualizing client feedbacks.
- **Credit Scoring Model**
  - Developed a **CatBoost-based credit scoring model** to predict clients' probability of default.
  - Performed **data preprocessing** and deployed the model on the **Dataiku platform** for accurate credit risk assessment.

Research trainee | Internship

IRIT Labs — *Toulouse - France* — 06/2019 – 09/2019 | **3 months**

Developed and deployed a **Deep Learning model** on **AWS** to predict **gas density** from thermal images and estimate its pressure.

Education

Master of Science in Artificial Intelligence — ENSEEIHT & Paul Sabatier University — *Toulouse, France* — 2019 – 2020

Master of Engineering – MEng — ENSEEIHT — *Toulouse, France* — 2017 – 2020

- Major in Computer Science and Applied Mathematics

CPGE MPSI/MP — Lycée Med Reda Slaoui — *Agadir, Morocco* — 2015 – 2017

- Preparatory classes for engineering schools, Major in Maths and Physics

Projects

- **Project 1: Smart Human Tracking Camera**
  - Developed a **real-time human tracking system** using **Arduino UNO**, **Servomotors**, and the **YOLOv5 model** for human detection and tracking (Github)
- **Project 2: Aircraft Failure Prediction (A320)**
  - Trained **Autoencoder** and **LSTM models** to classify sensor outputs and predict failures on **A320 aircrafts**, enabling **anomaly detection** for proactive maintenance

Certifications

- Computer Architecture (by **Princeton University**)
- Big Data & AI Certificate from **ISAE-SUPAERO** (*Toulouse, France*)
- Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization **On Coursera**