

# ANASS MAJJI



[anassmajji34@gmail.com](mailto:anassmajji34@gmail.com)



[amajji \(Anass Majji\) · GitHub](#)



[+ 33 6 23 45 85 37](tel:+33623458537)

## Compétences

**Programming:** Python | R | C / C++ | JavaScript | SQL | Spark | Matlab | VHDL

**AI Libraries and Tools:** PyTorch | TensorFlow | ONNX | DeepSpeed | Scikit-learn | Scrapy / Selenium | 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.

## Expérience professionnelle

### Data Scientist | CDI

BNP PARIBAS

Paris - France

10/2022 – Présent | **2 ans et 5 mois**

Au sein de l'équipe **Climate Analytics & Alignment**, j'ai développé et déployé **2 projets** en production:

- **Modèle de Question Answering LLM** : Développement d'un modèle Mistral 7B pour extraire automatiquement des informations des rapports annuels des entreprises, réduisant la charge de travail de l'équipe de 30%.
  - Extraction du texte à partir des PDF avec **GRODBID API** et création d'un vector database pour le **RAG** avec **Langchain**.
  - Optimisation des performances d'inférence du **LLM** (quantisation, KV caching, Flash attention ... )
  - Déploiement du LLM sur une streamlit webapp.
- **Prédiction de la performance énergétique des bâtiments** : Développement et déploiement d'un modèle de **Machine Learning** pour la prédiction les étiquettes énergétiques des bâtiments commerciaux afin d'évaluer le bilan carbone du portefeuille. Le modèle a atteint une accuracy de **92%** économisant **250 k€**.
  - Conduite des tests unitaires et model monitoring ainsi que les bonnes pratiques de Software development.
  - Automatisation du process de geocoding des adresses des biens immobiliers à l'aide d'une Fastapi web application puis imputing des données manquantes avec du web scraping.
  - Développement et déploiement via une Fastapi web interface, un modèle XGBOOST pour la prédiction des étiquettes énergétiques des logements.

### Data Scientist | Stage + CDI

Crédit Lyonnais (LCL)

Paris - France

06/2020 – 10/2022 | **2 ans et 4 mois**

Au sein de l'équipe **IA de la Data Factory**, j'ai développé et déployé en production **4 modèles DL et ML** :

- Conduite des tests unitaires et model monitoring ainsi que les bonnes pratiques de Software development.
- **Classification des documents et extraction des informations** : Développement et déploiement d'un modèle de deep Learning de type **CamemBERT** pour la classification automatique des documents clients (CNI, passeport, kbis...) et extraction des champs textes (nom, prénom, date de naissance, adresse ..) avec un pyesseract. Le modèle a été utilisé pour classifier et extraire les informations de **+20M documents**.
- **Détection de fraude**: Développement d'un modèle XGBOOST pour la détection des fraudes et d'une application web: extraction des données avec du SAS/ SQL Teradata, pré-processing et déploiement du modèle. L'implémentation de cette solution a permis, au LCL, d'éviter des fraudes d'une valeur totale de **+400 k€**.
- **Modèle de scoring pour l'octroi de crédit** : Développement d'un modèle Catboost pour le scoring des crédits afin de prédire la probabilité de défaut des clients: pré-processing des données, développement et déploiement sur la plateforme Dataiku.
- **Avis clients et sentiment analysis : (Best Innovative Project Award)** Récupération automatique des avis des clients à partir de **Google** à l'aide du web scraping (Scrapy & Selenium) puis classification en 5 grandes thématiques à l'aide d'un modèle de Deep Learning de type CamemBERT.
  - Développement et déploiement d'un dashboard Streamlit.

### Data Scientist | Stage R&D

IRIT Labs

Toulouse - France

06/2019 – 09/2019 | **3 mois**

Développement et Déploiement sur AWS d'un modèle de Deep Learning afin de prédire la densité et la pression d'un gaz à partir des images thermiques.

## Éducation

### Master de recherche en Artificial Intelligence

ENSEEIH T & Paul Sabatier University

Toulouse, France

2019–2020

### Diplôme d'ingénieur

ENSEEIH T

Toulouse, France

2017–2020

- Filière informatique et mathématiques appliquées

### CPGE MPSI/MP

Lycée Med Reda Slaoui

Agadir, Maroc

2015–2017

- Classes préparatoires aux grandes écoles d'ingénieurs. Option mathématiques et physique.

## Projets

- **Caméra intelligente pour le suivi des individus**
  - Développement d'une caméra intelligente pour le suivi des individus en temps réel, utilisant Arduino UNO et le modèle YOLOv5 ([Github](#))
- **Prédiction des pannes sur les avions A320**
  - Entraînement d'un Auto-encodeur et d'un LSTM pour la classification des séries temporelles des capteurs et la détection des anomalies sur les avions A320.

## Certifications

- Computer Architecture (de **Princeton University**)
- Big Data & AI Certificate de **ISAE-SUPAERO** (Toulouse, France)
- Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization **sur Coursera**
- Docker for Data Scientists **sur LinkedIn**