

DHIA HMILA

Senior Data Scientist at AXA France

@ dhiahmila@gmail.com +33610201606 92700, Colombes, France
hmladhia dhia-hmila Dhia Hmila dhia-hmila

hmladhia.github.io



PROFESSIONAL EXPERIENCE

Senior Data Scientist

Data Science Team, AXA France

June 2021 - Present Nanterre, France

- Development of document classification models using Computer Vision (Vision Transformer, VGG16, openCV, ...) & OCR (tesseract, EasyOCR, ...) to automatically index AXA's Document management system (GED).
- Development of text extraction AI Systems using Computer Vision (Yolov5, tesseract, VGG16, Vision Transformer, ...) to read documents (Driver's license, National ID card, Passeports, "Relevé d'information" ...)
- Development of reusable assets (packages, models, micro-services, ...) to use across the Team's projects.
- Building reusable Azure ML Pipelines (Continuous Training) to automatically retrain AI Systems & publish to Model Registry.
- Defining CI (Continuous Integration), CD (Continuous Deployment) & CT (Continuous Training) Templates for the Team's projects.
- Leading AXA's "Python COP" and organising a series of monthly programming talks/events (e.i. testing, Web Scraping, packaging, code quality, ...)
- Leading a series of training sessions/courses on Python, Data Manipulation (pandas & pyspark), Software Engineering best practices for AXA collaborators

Graduate Data Scientist

Data, Fraud, Waste & Abuse Team, AXA France

May 2020 - June 2021 Nanterre, France

- Development of an AI-based Fraud, Waste, and Abuse Detection Tool that acts on various scopes (Claims, Health Providers, Beneficiaries, ...) & Deployment on Databricks
- Creation of various fraud detection models for fraudulent patterns as part of the FWA detection tool.

Data Scientist

Shift Technology

April 2019 - March 2020 Paris, France

- Use of NLP technologies for clustering Claim managers' notes and visualizing results for detecting new patterns of fraud
- Build AI Models for Shift's fraud detection solution & Deployment for a new customer

PROJECTS

pysira: publish resumes in different formats

Github Actions Jinja2 Templates jsonschema

February 2023 - Present

CLI tool to export "jsonresume" files to different formats (html, tex, pdf, ...) and different languages.

ABOUT

I'm a passionate Data Scientist and a programming enthusiast. I've worked on fraud/anomaly detection and document classification problems

EDUCATION

Engineering double degree

ENSTA ParisTech

ENIT

2016 - 2019

France - Tunisia

Specialized in Artificial Intelligence

Preparatory School

IPEIT

Esprit Prépa

2014 - 2015

France - Tunisia

Specialized in Mathematics & Physics Ranked 10th in the entrance exam for engineering schools among 3,000 candidates.

SKILLS

Machine Learning

scikit-learn Tensorflow/keras trax NLP
 nltk spaCy Gensim pyod

Data Wrangling

pandas pyspark SQL ETL

Programming Languages

Python C/C++ C# JavaScript

MLOps

Github Workflows Azure Pipelines
 Git Linux Docker Azure ML
 Databricks huggingface

Web Scraping

- Creation of a CI/CD pipeline (Github Actions) to publish resume in different themes, formats, languages
- This resume was created using pysira: hmladhia.github.io/en or hmladhia.github.io/resume.pdf

nbmanips: Split, merge and convert IPython Notebooks

[Jupyter](#) [Databricks](#) [Zeppelin](#)

 May 2021 - Present

An open-source package containing a collection of tools to manipulate Notebooks via a python or CLI.

- Creation of a python package to manipulate (Split, Merge, ...) Notebooks
- Use of CI/CD pipelines (Github Actions) to Lint, Test and publish the package to Pypi

Spot-Language: Programming Language Detection

[Classification](#) [Flask](#) [NLP](#) [Lime](#)

 January 2020

Spot-Language is a classification model to detect the language used in code snippets.

- Construction of a Training dataset from github public repositories.
- Building an ML Experiment to train an NLP model (Random Forest, Neural Networks, ...) for language classification.
- Deployment of Web Application to serve as demo of the ML Model.

[selenium](#)

[requests](#)

[BeautifulSoup4](#)

Web

[HTML](#)

[CSS](#)

[JavaScript](#)

[fastapi](#)

[flask](#)

[Dash](#)

LANGUAGES

Arabic
English
French



CERTIFICATES



Cloud-Native Development with Open-Shift and Kubernetes Specialization

Coursera

 2023 2023



Machine Learning Engineering for Production (MLOps)


Coursera

 2022 2022



Machine Learning & General Coding Logic

Triplebyte

 2023 2023