



DHIA HMILA

Data Scientist Senior chez AXA France

@ dhiahmila@gmail.com
hmladhia.github.io
dhia-hmila

+33610201606
hmladhia

92700, Colombes, France
dhia-hmila

Dhia Hmila



EXPÉRIENCE PROFESSIONNELLE

Data Scientist Senior

Équipe Data Science & IA, AXA France

📅 Juin 2021 - Aujourd'hui

📍 Nanterre, France

- Développement de systèmes d'IA d'extraction d'information basé pour lire des documents (permis de conduire, carte d'identité nationale, etc.). Les modèles construits ont permis de traiter plus de 6 millions de documents.
- Création d'une gamme de packages Python réutilisables pour le traitement de documents (OCR, détection d'objets, etc.). Les packages sont conçus pour expérimenter rapidement avec différents modèles (Startegy Pattern).
- Automatisation de l'indexation de 70% des documents entrants d'AXA France (flux sinistre) en développant un système d'IA de classification de documents (exposé en tant qu'API).
- Élaboration de pipelines Azure ML réutilisables (CT - Continuous Training) pour réentraîner automatiquement les systèmes d'IA et les publier dans le Registre des modèles.
- Conception de modèles de pipelines CI/CD Azure DevOps à utiliser dans les projets de l'équipe.
- Supervision du processus de recrutement, intégration des nouveaux membres de l'équipe et facilitation du transfert efficace des connaissances.
- Coordination du processus d'annotation avec les annotateurs et les experts métiers pour garantir des datasets cohérents et de qualité.
- Direction de l'initiative "Python COP" d'AXA, organisant une série d'événements mensuels sur la programmation portant sur des sujets tels que les tests, le web scraping, le packaging, la qualité du code, etc.
- Animation d'une série de formations (sur Python, la manipulation des données avec pandas & pyspark, les bonnes pratiques de code) pour plus de 30 collaborateurs d'AXA.

Graduate Program - Data Scientist

Équipe Data, Fraud, Waste & Abuse, AXA France

📅 Mai 2020 - Juin 2021

📍 Nanterre, France

- Conception et développement d'un outil de détection avancée des fraudes, gaspillage et abus piloté par l'IA, agissant à différentes échelles (sinistres, prestataires de santé, bénéficiaires). Cet outil novateur a joué un rôle clé dans la réalisation d'économies dépassant les 150 000 euros en 2022.
- Conception et entraînement de divers modèles de détection de fraudes pour repérer les patterns frauduleux, dans le cadre de l'outil de détection des FWA(Fraud, Waste & Abuse).

Data Scientist

Shift Technology

📅 Avril 2019 - Mars 2020

📍 Paris, France

- Mis en œuvre des techniques de Traitement du Langage Naturel (NLP) pour regrouper les notes des gestionnaires de sinistres, en utilisant des méthodes de visualisation avancées afin de déceler et d'identifier les patterns de fraude émergents.

A PROPOS DE MOI

Je suis un Data Scientist Senior spécialisé dans la détection de fraudes/anomalies et la classification de documents. J'ai travaillé sur une panoplie de sujets ML dont des sujets Computer Vision ou encore NLP/NLU. Je possède également une solide expérience autour des pratiques de MLOps, pour garantir une intégration fluide des modèles dans les environnements de production.

CURSUS

Double diplôme d'ingénieur

🎓 ENSTA ParisTech

🎓 ENIT

📅 2016 - 2019

📍 France - Tunisia

Spécialisé en Intelligence Artificielle

Cycle préparatoire

🎓 IPEIT

🎓 Esprit Prépa

📅 2014 - 2015

📍 France - Tunisia

Spécialisé en Math & Physique Rang 10 au concours d'entrée aux écoles d'ingénieurs en Tunisie parmi plus de 3000 candidats.

COMPÉTENCES

Intelligence Artificielle

Anomaly Detection

Tensorflow

TFX

scikit-learn

NLP

nlTK

spaCy

Gensim

pyod

Data Wrangling

pandas

pyspark

SQL

ETL

Langages de programmation

Python

Rust

JavaScript

C/C++

C#

MLOps

- Élaboration de nouveaux modèles d'IA pour la solution de détection de fraude de Shift et déploiement de la solution pour un nouveau client.

Stage de recherche

U2IS - ENSTA Paris

📅 Mai 2018 - Août 2018

📍 Palaiseau, France

- Simulation d'un agent dans un environnement artificiel et création d'un dataset d'images où on peut percevoir l'effet de l'oubli catastrophique en utilisant le Raytracing pour rendre des images réalistes..
- Évaluation de la méthode "Elastic Parameters Consolidation" proposée par un article de recherche pour remédier à l'oubli catastrophique qui surgit lors de l'apprentissage incrémental.

PROJETS

📦 mlflow: Outil d'orchestration pour les projets MLflow.

[mlflow](#) [orchestration](#) [experiment-tracking](#)

📅 Novembre 2023 - Aujourd'hui

Outil d'orchestration de jobs d'entraîneement basé sur MLflow Projects.

📦 pysira: publication de CVs dans différents formats.

[Github Actions](#) [Jinja2 Templates](#) [jsonschema](#)

📅 Février 2023 - Aujourd'hui

Outil en ligne de commande permettant d'exporter des CVs au format 'jsonresume' vers différents formats (HTML, TeX, PDF, etc.) et langues (Anglais, Français, etc.).

- Création d'un pipeline CI/CD (Actions Github) pour publier des CV dans différents thèmes, formats et langues.
- Ce CV a été créé en utilisant pysira : hmladhia.github.io, hmladhia.github.io/cv.pdf, hmladhia.github.io/fr/kendall ou hmladhia.github.io/resume.pdf

📦 nbmanips: Découper, fusionner et convertir vos Notebooks

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

📅 Mai 2021 - Aujourd'hui

Un package open-source d'outils de manipulation de Notebooks via des scripts python ou une CLI.

📄 Spot-Language: Prédiction du langage de programmation

[Classification](#) [Flask](#) [NLP](#) [Compression de Modèles](#)

📅 Janvier 2020

Spot-Language est un modèle de classification des langages utilisés dans des sources de code.

- Construction de dataset d'entraînement provenant de repositories github publiques.
- Déploiement d'une application Web pour servir de démo au modèle IA.

[Github Actions](#)

[Azure Pipelines](#)

[Git](#)

[Linux](#)

[Docker](#)

[Azure ML](#)

[Databricks](#)

Web Scraping

[selenium](#)

[requests](#)

[BeautifulSoup4](#)

LANGUES

Arabe
Anglais
Français



CERTIFICATIONS



Machine Learning Engineering for Production (MLOps) Specialization

DeepLearning.AI

📅 2022 2022



Natural Language Processing Specialization

DeepLearning.AI

📅 2024 2024



Cloud-Native Development with OpenShift and Kubernetes Specialization

Red Hat

📅 2023 2023



Rust Fundamentals

Duke University

📅 2024 2024



Rust for DevOps

Duke University

📅 2024 2024



Rust for LLMops (Large Language Model Operations)

Duke University

📅 2024 2024



DevOps, DataOps, MLOps

Duke University

📅 2024 2024



MLOps Tools: MLflow and Hugging Face

Duke University

📅 2023 2023



Introduction to Large Language Models

Google Cloud

📅 2023 2023



TOEIC Listening and Reading Test (Score 985/990)

ETS Global

📅 2018 2018