Hafsa Ouajdi

Madrid, Spain

+336 12 78 48 71 hafsa.ouajdi.fr@gmail.com Linkdin Github Kaggle Gitlab Portfolio

Education

Ecole Centrale de Nantes **Sep. 2021 – Oct. 2024**

Master of Engineering (GPA: 3.94/4) *Nantes, France*

- Applied Mathematics, Data science, machine Learning, Signal & Image Processing
- International Business Development (Finance, Marketing, Sales).

Lycée Mohammed VI d'Excellence **Sep. 2019 – Jul. 2021**

Preparatory classes (CPGE) *Bengurir, Morocco*

- Intensive theoretical mathematics and physics courses to prepare for French engineering college competitions.

Professional Experience

Data Scientist & Mobility Consultant **Dec. 2024 – present**

Nommon Solutions and Technologies. *Madrid, Spain*

- Lead and implement projects to provide mobility indicators and insights using big data analytics and geolocation data.
- Develop and deploy machine learning models for mobility analytics, optimizing travel demand forecasting and transport planning.
- Work with transport authorities, operators, and consultants across the world.
- Process and analyze large-scale mobility data (e.g., anonymized mobile network data, GPS data) to extract insights using statistical modeling and AI techniques.

Machine Learning Engineer Intern — 6-Months **Apr. 2024 – Sep. 2024**

Paul Wurth (SMS Group) *Luxembourg, Luxembourg*

- Researched and developed advanced AutoML models leveraging Autoencoders and Transformers for anomaly detection and time series forecasting, contributing to novel machine learning methodologies.
- Automated the end-to-end machine learning pipeline, including data preprocessing, model training, and prediction deployment, through a web-based platform using PyTorch, Next.js, Django, and Docker.
- Engineered a data exploration chatbot integrating the OpenAI API and LangChain, facilitating intuitive data analysis for non-technical stakeholders.
- Designed and implemented interactive dashboards for results visualization, enabling effective communication of findings and fostering decision-making grounded in data-driven insights.

R&D Intern: Deep Learning & Virtual Reality — 5-Months **Apr. 2023 – Aug. 2023**

SEGULA Technologies *Nantes, France*

- Developed a collaborative VR simulation for ship design based on prior research.
- Used deep learning for 3D object position prediction to improve simulation accuracy.
- Explored innovative VR solutions for the industrial sector using Unity, C#, and Git.

Academic Projects & Summits

Mena Machine Learning Winter School **Feb. 2025**

Middle East and North Africa Machine Learning School *Doha, Qatar*

- Selected for the prestigious Mena Machine Learning Winter School in Qatar.
- Attended lectures and hands-on workshops on cutting-edge ML techniques, deep learning, and AI applications.
- Collaborated on a project tackling real-world AI challenges.

Deep Learning Project: Environmental Audio Deep Fake Detection. **Nov. 2023 – Mar. 2024**

Collaborative Project: Centrale Nantes and Carnegie Mellon University. *Nantes, France*

- Developed an advanced deep learning classification system achieving 98% accuracy using embeddings and PyTorch
- Co-authored a conference paper for the European Conference on Signal Processing (Eusipco 2024).

Business Development & AI Project **Nov. 2023 – Mar. 2024**

Mantu *Nantes, France*

- Evaluated AI solutions for optimizing sales processes, including recommendation, scoring, and matching systems.
- Conducted in-depth ROI analysis to guide strategic decision-making.

Publication

Detection of Deepfake Environmental Audio **2024**

European conference on signal processing 2024 *Lyon, France*

Technical Skills

Programming Languages: Python, C++, C#, JavaScript
Machine Learning & Deep Learning: Supervised & Unsupervised Models (Random Forest, SVM, Regression), Dimensionality Reduction (PCA), Clustering (K-Means), Neural Networks (CNN, RNN, LSTM), NLP (BERT, GPT), Transformers, GANs, GNN, Optimization (Gradient Descent), Cross-Validation, Statistical Tests (t-test, ANOVA)
Libraries/Frameworks: PyTorch, Scikit-Learn, TensorFlow, Keras, Pandas, OpenCV
Web development: HTML/CSS, JavaScript, NodeJS, ReactJS, NextJS, TypeScript
Tools: Poetry, Git/GitHub/GitLab, Azure Cloud, Streamlit, Databricks, Apache Spark, Docker, Kubernetes, Linux, SQL, Unity3D, LaTeX

Languages

English (Advanced), French (Fluent), Arabic (Native), Spanish (Beginner)

Personal Projects

Azure-Based Data Engineering Project

Tools: Azure Cloud, Databricks, PowerBI, Data Lake, SQL, Pyspark

- Designed and implemented a serverless ETL pipeline on Azure for AdventureWorks datase.
- Ingested raw data from GitHub, transformed using PySpark on Databricks.
- Stored processed data in Azure Data Lake Gen2 using Parquet format.
- Queried and analyzed data using Synapse Serverless SQL.
- Delivered interactive visualizations via Power BI dashboards.

AI-Powered SOC Agent with Kafka and Streamlit Dashboard

Tools: Kafka, Python, Streamlit, Docker, Pytorch, Scikit-Learn, Pandas, JSON, Pickle

- Developed an end-to-end AI Security Operations Center (SOC) system for real-time cyber threat detection and alerts.
- Trained a machine learning model (TabTransformer) using the UNSW-NB15 dataset to classify normal vs. attack traffic across multiple categories.
- Simulated synthetic network logs by sampling and encoding real attack data with label encoders, enabling live testing and streaming.
- Implemented Kafka-based stream to generate, process, and classify log data in a continuous flow.
- Built an interactive Streamlit dashboard to monitor security alerts, visualize attack categories, and inspect network behaviors.
- Designed a Dockerized microservice architecture using Docker Compose for Kafka, Zookeeper, the AI agent, and the dashboard.

Cancer Prediction using Machine Learning

Tools: Streamlit, Python, Scikit-Learn, Random Forest, Logistic Regression, Pandas

- Developed a machine learning model to predict cancer outcomes using patient data.
- Implemented multiple classification models, including Random Forest and Logistic Regression, to predict cancer stages.
- Used Pandas for data preprocessing and feature engineering.
- Deployed the model as a web application using Streamlit for real-time predictions.
- Visualized model results and performance metrics through interactive plots in the web app.

AI Assistant for Data Scientists

Tools: Langchain, OpenAI API, Streamlit, Python

- Created an AI assistant tool for data scientists to streamline data exploration and manipulation.
- Integrated Langchain and OpenAI API for natural language processing to assist with dataset queries and visualizations.
- Developed a user-friendly web interface with Streamlit for interacting with the assistant.
- Enabled model to respond to user queries about datasets, preprocessing, and data analysis.

Image Captioning using Transformers

Tools: Python, PyTorch, Transformer, CNN, COCO Dataset

- Developed an image captioning model using Transformer architecture and CNNs for feature extraction.
- Used the COCO dataset for training and fine-tuned a pre-trained Transformer model for caption generation.
- Achieved significant performance improvement with fine-tuning strategies and data augmentation.
- Implemented the model in Python with PyTorch for deep learning training and inference.

Financial Transaction Analysis Dashboard

Tools: Power BI, SQL, Pandas, Python, Data Visualization

- Created an interactive Power BI dashboard for analyzing financial transaction data.
- Incorporated SQL queries for data extraction and Pandas for data manipulation and cleaning.
- Used advanced data visualization techniques in Power BI to create insightful and actionable visualizations.
- Enabled drill-down and filter options for users to explore transaction data in-depth.