

RÉSUMÉ DU PROFIL

Ingénieur d’État en Data Engineering et en Master 2 de Data Science, je recherche un stage de fin d’études pour travailler sur des pipelines de données et des architectures Big Data.

ÉDUCATION

Université Claude Bernard Lyon 1	2025 - 2026
Master en Data Science   Cours pertinents: Deep Learning, Big Data, Machine Learning, Cloud Computing.	
École nationale des sciences appliquées d'Al Hoceima, Université Abdelmalek Essaadi	2020 - 2025
Diplôme d’ingénieur d’état en Data Engineering   Cours pertinents: Databases, Big Data, Machine Learning, Cloud.	

EXPÉRIENCE PROFESSIONNELLE

ENTREPRISE SHIFTBRICKS	Al Hoceima, Maroc
Stagiaire en data engineering	Juil. 2024 – Sept. 2024

- Développé une architecture Medallion Big data pour l’ingestion et la traçabilité des documents juridiques arabes.
  - Transformé les données et créé des vecteurs d’embeddings par l’OCR (Tesseract, EasyOCR) et LLM (Cohere).
  - Automatisé le flux de données avec Airflow et Conteneurisé la solution à l'aide de Docker.
  - Créé une interface utilisateur avec le stack FARM (FastAPI, React) pour la validation humaine.
- Tech stack: Python, MongoDB, Airflow, FastAPI, React, Docker, Git, GitLab, OCR, GenAI.

ENTREPRISE DIGITAL PLACE	Tanger, Maroc
Stagiaire en data science	Fév. 2025 – Juil. 2025

- Conçu un pipeline vision pour l’automatisation, la détection, la segmentation et le suivi des semi-remques.
  - Entraîné YOLOv11 pour détecter les plaques, atteignant 97,8 % de précision.
  - Entraîné EfficientNet-B2 pour la classification des plaques marocaines (arabes) et européennes (latines).
  - Mis en œuvre PaddleOCR pour la lecture des plaques avec une précision de 87,5 % en arabe et 93,75 % en latin.
  - Appliqué DINOv2 pour générer des embeddings des composants des camions et enrichir leur ADN visuel.
  - Développé et déployé l’application FastAPI + React, conteneurisée (Docker) et livrée via Bitbucket CI/CD sur VPS (DevOps).
- Tech stack: YOLOv11, PaddleOCR, Grounded SAM2, DINOv2, PyTorch, OpenCV, FastAPI, React.js, Docker, Bitbucket, Hostinger VPS.

COMPÉTENCES TECHNIQUES

- Languages: Python, Java, SQL, PL/SQL, Spring Boot.
- Big Data & BI Tools: Spark, Hadoop, Kafka, Airflow, Power BI.
- Bases de données: MySQL, Oracle, PostgreSQL, MongoDB, Cassandra, ClickHouse.
- DevOps: Git, GitLab, GitHub, Docker, Kubernetes.
- Machine Learning & Deep Learning: Classification, Clustering, Regression, scikit-learn, LLM, pyTorch.
- Services Cloud: S3, Aws EC2, Databricks,Cosmos DB, Synapse Analytics, Azure Data Lake, Azure Data Factory, Azure Functions.

DERNIERS PROJETS

Conception d’une architecture Big Data pour un Chatbot des tendances IA basé sur RAG – Azure
<ul style="list-style-type: none"><li>- Conçu une architecture Medallion Big Data sur Azure pour automatiser le pipeline des données (bronze, argent, gold) pour un chatbot .</li><li>- Transformé les données dans Databricks avec création d’embeddings, stockés dans Cosmos DB et AI Search pour la recherche avancée.</li><li>- Automatisé les workflows avec Data Factory et déployé la solution complète sur une infrastructure Azure via CI/CD (DevOps).</li></ul> <p>Tech Stack: Azure Functions,Cosmos DB, Databricks, AI Search, Data Factory, Container Apps, FastAPI, Cohere API, Next.js, BeautifulSoup.</p>
Pipeline de données en architecture Big Data Kappa – traitement temps réel des données d’appartements (Streaming)
<ul style="list-style-type: none"><li>- Ingesté et diffusé des données avec Kafka, stocké les textes bruts dans HDFS et transformé en temps réel avec PySpark Streaming.</li><li>- Stocké les données traitées dans ClickHouse pour des requêtes rapides et visualisations temps réel via React.js et FastAPI (Docker).</li></ul> <p>Tech Stack: Hadoop HDFS, Apache Kafka, PySpark Streaming, Cohere API, ClickHouse, Docker, React.js, FastAPI.</p>
Prédiction des prix et recommandation d’appartements (2 <sup>e</sup> place – Compétition ML, Oujda, Maroc)
<ul style="list-style-type: none"><li>- Collecté des données immobilières avec BeautifulSoup et Selenium, et construit un modèle de régression hybride et recommandation.</li><li>- Développé une application web avec Flask et JavaScript pour afficher les prédictions et recommandation</li></ul> <p>Tech Stack: BeautifulSoup, Selenium, Python, Scikit-learn, pandas, Flask, HTML/CSS, JavaScript</p>
Pipeline de données d'analyse des brevets agricoles
<ul style="list-style-type: none"><li>- Collecté et préparé des données de brevets depuis Google Patents, stockées dans MongoDB, transformées avec PySpark, puis chargées dans PostgreSQL (Data Warehouse).</li><li>- Analysé les données avec Power BI et développé une interface utilisateur avec Flask pour faciliter l’interaction..</li></ul> <p>Tech Stack: MongoDB, PySpark, PostgreSQL, SQLAlchemy, Power BI, Flask.</p>

CERTIFICATIONS

- Big Data with Spark and Hadoop Essentials (IBM).
- ETL and Data Pipelines with Shell, Airflow, and Kafka (IBM)
- Machine Learning with Apache Spark (IBM).
- Data Warehousing and BI Analytics Essentials (IBM).
- NoSQL Databases Essentials (IBM).

SOFT SKILLS

- Travail en équipe éprouvé sur des projets collaboratifs, adaptabilité, résolution de problèmes, communication.

LANGUES

- Français (Avancé), Anglais (Avancé), Arabe & Tamazight (Langue maternelle)