

Imane MOUSTAOU

Data Engineering

✉ imanemoustaoui11@gmail.com

☎ +212-669034758



Casablanca, Maroc



www.linkedin.com/in/imane-moustaoui



Expériences Professionnelles

Agenz (CASABLANCA) | Data Engineering / Data scientist :
Novembre 2021 – Présent | Télétravail

- Acquisition de données & Web Scraping : Création de pipelines robustes pour collecter et structurer des données dynamiques avec BeautifulSoup, Selenium, CrawlerAI et Requests.
- Data Processing & Transformation : Préparation et structuration de données pour analytics et modèles prédictifs, en utilisant Pandas, NumPy, SQL et techniques statistiques.
- Bases de données & Analytics : Gestion de SQL et NoSQL (MongoDB) avec pipelines d'agrégation, indexation optimisée et exports de données structurés.
- Modélisation prédictive & Machine Learning : Analyse de données, feature engineering et déploiement de modèles prédictifs (DBSCAN, XGBoost, Random Forest) en Python pour générer des insights exploitables et soutenir les décisions.
- Développement Backend & API : Conception et déploiement d'APIs REST scalables avec Flask et FastAPI, endpoints asynchrones, validation Pydantic et tests complets via Postman.
- Automation & Workflow : Conception et suivi de pipelines ETL via Apache Airflow pour un flux de données fiable et optimisé.
- Cloud & Conteneurisation : Conteneurisation d'applications Python avec Docker et déploiement sur environnements cloud AWS, avec bonnes pratiques DevOps pour assurer portabilité, scalabilité et automatisation des déploiements.
- Reporting & Visualisation : Développement de tableaux de bord interactifs et automatisés (Metabase, Power BI) pour supports décisionnels

Capgemini engineering (CASABLANCA) | Stage – Data Scientist :
Mars – Septembre 2021

Conception et implémentation d'un système ML pour détecter et corriger les anomalies dans des séries temporelles d'un système automobile.

- Analyse & Visualisation de Séries Temporelles : Traitement et visualisation de données temporelles avec NumPy, SciPy, Pandas, Matplotlib, Seaborn, incluant analyse de tendances et détection d'anomalies
- Détection d'anomalies non supervisée : Détection non supervisée : Création de modèles DBSCAN pour l'identification automatique d'anomalies et d'outliers.
- ML Supervisé pour correction d'anomalies : Implémentation de modèles XGBoost pour corriger les anomalies, incluant pipelines d'entraînement, réglage d'hyperparamètres et validation des performances pour des prédictions fiables.
- Web Applications & Python : Développement d'interfaces Flask avec intégration de modèles ML, visualisations et APIs.

Formation Académique

2019 - 2021 : Faculté des Sciences Ben M'sik – Casablanca
Master en Data Science et Big Data

2016 - 2019 : Faculté des Sciences Ben M'sik – Casablanca
Licence en Informatique – Bases de Données

2015 - 2016 : Lycée Al Khawarizmi – Casablanca
Baccalauréat – Mathématiques et Sciences B

Compétences

Bibliothèques Python

Pands, Numpy, Request, Sckitlearn, Scipy, Asynclo, Pydantic, Matplotlib, Seaborn, Geopandas

AI & LLM Tools

Github COPILOT, LangChain, LlamaIndex

Bases de Données

MONGODB Compass, MYSQL

WEB/APIs

FLASK, FAST API, DJANGO, RESTful APIs

Data Scrapping

Beautiful Soup, Selenium, Scrapy, CRAWLERAI

DEVOPS / Containerisation

Gitlab / GitHub, DOCKER, Kubernetes

Outils de Visualisation

METABASE, TABLEAU, Excel

Langages de Requête

SQL, NoSQL

Workflow & Orchestration

Apache Airflow

Data Science

Régression Linéaire, Random Forest, XGBoost, DBSCAN

Outils de Big Data

AWS

Langues

Arabe : Natale

Français : Courant

English : Professionnel