



# TAFAHI Soufiane

Data Scientist

✉ +212 655995512 | 📩 itssoufianetafahi@gmail.com | LinkedIn | GitHub | Morocco

## PROFIL

---

Étudiant en **Master de Mathématiques et Informatique**, spécialisé en **Science des Données**. Passionné par le **machine learning** et l'**analyse prédictive**, avec une expérience dans la conception de **modèles statistiques** et la gestion de **données massives**. Compétent dans les environnements **Cloud** et les architectures **Big Data** pour le traitement de données à grande échelle. Je recherche un **stage de fin d'études (PFE)** en **Data Science** pour contribuer à des projets innovants et apporter des solutions basées sur l'analyse de données.

## COMPÉTENCES

---

**Langages & Outils :** Python, SQL, Git

**Data Science & Machine Learning :** Pandas, NumPy, Scikit-learn, TensorFlow, XGBoost, LSTM, RNN

**Bases de données & Stockage :** MySQL, PostgreSQL, MongoDB, Neo4j

**Visualisation & Reporting :** Power BI, Tableau, Matplotlib, Seaborn, Plotly

**Cloud Computing & DevOps :** Azure, Google Cloud, Docker

## EXPÉRIENCE

---

### Stage de Fin d'Année, Ensa Khouribga ([GitHub](#))

Juil 2025 – Sep 2025

*Développement d'un système de génération de logs synthétiques et détection d'anomalies*

Génération de logs synthétiques par **LLM (DistilGPT-2)** pour enrichir les jeux de données de **40%**

Modèle **LSTM** de détection d'anomalies atteignant **99% de précision**

Fine-tuning du modèle pour injection contrôlée d'anomalies sémantiques

**Technologies :** DistilGPT-2, LSTM, Python, TensorFlow, PyTorch, NLP

### Stage de fin d'études, EST Oujda

Avr 2024 – Jul 2024

*Solution Business Intelligence pour la production agricole mondiale*

Plateforme **BI complète** sur Azure analysant **+500k enregistrements**

Modèles prédictifs avec **90% de précision** pour les rendements agricoles

**10+ dashboards** Power BI pour l'analyse décisionnelle temps réel

**Technologies :** Python, Azure (Data Factory, Databricks, Synapse), Power BI, Scikit-learn

## PROJETS

---

### Pipeline d'Analyse Temps Réel - Données Financières

*Architecture Big Data pour l'analyse de marchés financiers et cryptomonnaies*

Ingestion de données temps réel via **Apache Kafka** depuis +2 APIs financières (Binance, Alpha Vantage)

Traitements de flux avec **Spark Streaming** et calcul d'indicateurs techniques (RSI, moyennes mobiles)

Détection d'anomalies sur +10k transactions/jour avec stockage **PostgreSQL**

Orchestration des pipelines avec **Apache Airflow** et déploiement conteneurisé **Docker**

Visualisation temps réel dans **Power BI** pour analyse décisionnelle immédiate

**Technologies :** Kafka, Spark, Airflow, Docker, PostgreSQL, Power BI, APIs financières

## **Plateforme d'Exploration de Publications Scientifiques ([GitHub](#))**

*Système d'analyse NLP et graphes de connaissances pour recherche académique*

Extraction automatique depuis **API arXiv** de +20k publications scientifiques

Construction d'un **graphe de connaissances Neo4j** avec +80k relations entre entités

Clustering thématique avec **KMeans/TF-IDF** identifiant +10 domaines de recherche

Tableau de bord **Streamlit** permettant l'exploration intuitive des connexions scientifiques

**Technologies** : Python, Neo4j, NLP, spaCy, KMeans, Streamlit, arXiv API

## **Platform d'Analyse de Ventes Internationales - Google Cloud**

*Solutions Cloud pour l'analyse de données de ventes multi-sources*

Automatisation de l'ingestion de +500k lignes/jour vers **Google Cloud Storage**

Pipeline ETL optimisé vers **BigQuery** réduisant le temps de traitement de 60%

Analyse des performances identifiant les top 10 produits et tendances mensuelles

Tableau de bord **Looker Studio** avec +5 KPI pour suivi décisionnel temps réel

**Technologies** : GCP, BigQuery, SQL, Python, Looker Studio, Data Studio

## FORMATION

---

### **Master en Mathématiques et Informatique pour la Science des Données**

*École Nationale des Sciences Appliquées de Khouribga*

Khouribga, Maroc

*Sep 2024 - En cours*

### **Licence Professionnelle Informatique Décisionnelle**

*École Supérieure de Technologie Oujda*

Oujda, Maroc

*Sep 2023- Juin 2024*

### **DUT Génie Informatique**

*École Supérieure de Technologie Mèknes*

Mèknes, Maroc

*Sep 2020- Juin 2022*

### **Baccalauréat en Sciences Mathématiques B**

*Lycée Technique Taza*

Taza, Maroc

*Sep 2017- Juil 2020*

## CERTIFICATIONS

---

**Machine Learning avec K-Nearest Neighbors:** [365 Data Science](#)

**Git et GitHub:** [365 Data Science](#)

**Nettoyage et Prétraitement des Données avec pandas:** [365 Data Science](#)

**Programmation Python**

**Intermédiaire:** [365 Data Science](#)

## LANGUES

---

**Arabe** : Langue maternelle | **Français** : Niveau intermédiaire avancé | **Anglais** : Niveau intermédiaire

## DISTINCTIONS

---

**4<sup>ème</sup> place** à Hackathon – Open Source Days 13.0 DATAVERSE

**3<sup>ème</sup> place** à Problem Solving – Open Source Days 13.0 DATAVERSE