

# MOHAMED JARBOUI

Étudiant en IA et Data Science

Sousse, Tunisie | mohamed.jarboui@polytechnicien.tn | +216 99 746 810

## PROFILE

Élève-ingénieur en dernière année, spécialisé en Data Science. Expérience pratique dans le développement de prototypes de A à Z, le nettoyage de données, les processus ETL et la modélisation prédictive. Passionné par l'application du Machine Learning à des problématiques concrètes pour une prise de décision guidée par la donnée. Reconnu pour mon esprit d'analyse, ma capacité à résoudre des problèmes complexes et à livrer des solutions performantes.

## COMPÉTENCES TECHNIQUES

- Langages & libs** : Python, SQL, Java, Pandas, NumPy, Scikit-learn, TensorFlow/Keras
- Data Engineering** : ETL, ingestion API, web scraping (Selenium, undetected\_chromedriver), Snowflake (intégration)
- Machine Learning / Deep Learning** : CNN, SVM, classification/régression, NLP (prétraitement), feature engineering, Computer Vision, LLMs, RAG
- MLOps / Infra** : MLflow, Docker, Kafka, gRPC, CI/CD (notions DevOps)
- Développement Web** : MERN (MongoDB, Express, React, Node.js), MEAN (MongoDB, Express, Angular, Node.js)
- Outils & Visualisation** : Power BI, Matplotlib, Git
- Autres** : REST APIs, automatisation, gestion de workflows

## EXPÉRIENCE PROFESSIONNELLE

### Stagiaire Data Science / Data Engineering, DNEXT Intelligence SA

Juin- Août 2025

- Identification et évaluation des sources de données** : Recherche et analyse de sources de données pertinentes (sites web, APIs) pour répondre aux objectifs métier du projet, en évaluant leur qualité, leur fiabilité et leur valeur potentielle.
- Collecte et ingestion de données** : Développement et mise en œuvre de pipelines robustes pour scraper et collecter des données fiscales et des historiques d'échanges commerciaux (codes HS) à partir des sources identifiées.
- Traitement et modélisation des données** : Nettoyage, transformation et enrichissement des données via des processus ETL pour sélectionner, structurer et valoriser les informations essentielles.
- Gestion de l'infrastructure data** : Intégration des jeux de données traités dans la plateforme interne et gestion des stockages sur cloud (Snowflake, etc.) pour assurer leur disponibilité, leur sécurité et leur intégrité.
- Automatisation et résolution de problèmes** : Implémentation d'une solution IA pour contourner les systèmes CAPTCHA, automatisant ainsi la collecte et garantissant son efficacité.
- Visualisation et restitution** : Création de tableaux de bord interactifs avec Power BI pour offrir une vision claire et actionable des données aux équipes métier, facilitant la prise de décision.

### FORMATION PROFESSIONNELLE

Juin- juillet 2024

Participation à un programme intensif de formation axé sur les fondamentaux et les pratiques avancées de la data. Cette formation m'a permis de développer une base solide en :

- Manipulation, nettoyage et préparation de données (Python, Pandas, NumPy)
- Analyse exploratoire et visualisation (Matplotlib, Power BI)
- Traitement de données issues de multiples sources (CSV, APIs, web scraping)
- Initiation au Machine Learning (modélisation, validation, métriques)
- Notions de Data Engineering : pipelines, ETL, automatisation, bonnes pratiques
- Mise en place de projets pratiques simulant des cas réels d'entreprise

## FORMATION

Septembre 2023 – Juin 2026 (prévu)

École Polytechnique de Sousse — Ingénierie Informatique (Data Science & AI)

- Cours principaux : Entrepôt de données, Big Data, Base de Données Avancée, JAVA Avancé, Python Avancé, MLOPS, DevOps, Deep Learning, Computer Vision (CNN), NLP

Septembre 2021 – Juin 2023

École Polytechnique de Sousse — Cycle préparatoire intégré (MTIC)

- Cours : Analyse, Algèbre, Physique, Intelligence Artificielle, Algorithmique, Développement Web, Programmation Orientée Objets

## PROJETS PERSONNELS

### PFA - Dermatological Age Prediction (CNN) — Face-based Age Regression

Built a prototype that estimates a person's dermatological age from face photos.

- Data & prep: Combined UTKFace with an additional dataset; resized images, converted to grayscale, and applied  $\pm 40^\circ$  rotations to improve generalization and keep compute light.
- Modeling: Tried several TensorFlow/Keras CNNs and selected the best using early stopping; trained on GPU with Adam.
- Outcome: ~3-year RMSE on a held-out set.
- Scope & limits: Noted fairness limits (skewed ages, noisy images, limited diversity). Technologies: Python, Tensorflow, Keras, Git, Streamlit

### Intelligent Content Moderation

- Built a real-time content moderation system with microservices communicating via gRPC and Kafka.
- Integrated a Hugging Face transformer model to score toxicity and classify content as toxic or acceptable.
- Added a policy service to enforce business rules (flag, block, send to human review).
- Implemented logging and audit trails so decisions can be inspected later.
- Outcome: Automatically filters harmful content before it reaches users or moderators. Technologies: Kafka, gRPC, Docker, Hugging Face Model Api, Nodejs, ExpressJs

### EU Trade Tariff Data Scraper & Updater (Selenium + APIs)

Developed an automated Python tool that keeps international trade tariff data continuously updated by extracting information from both web sources (EU Access2Markets portal) and external APIs.

Built a Selenium bot (undetected Chrome) that:

- Navigates to tariff result pages for multiple HS codes, origins, and destinations
- Waits for dynamic JS content, scrolls to load full results, and parses tariff formulas & descriptions
- Saves all retrieved data into a clean CSV dataset

Added an update layer that pulls supplementary information from APIs and site endpoints, ensuring the dataset stays fresh, complete, and synchronized over time.

Designed with a clear, modular workflow:

configure → scrape website → fetch API data → merge → clean → export

Key Outcomes:

- Fully automated daily/weekly data refresh
- No manual collection needed
- Reliable structured dataset for analysis or integration into other systems

Technologies:

Python, Selenium (undetected\_chromedriver), Pandas, WebDriverWait/EC, REST APIs, CSV automation

### Développement Web Full-Stack – MERN & MEAN (Travaux Académiques)

- **Réalisation de plusieurs projets académiques en MERN et MEAN :**  
CRUD, APIs REST, gestion d'authentification, manipulation de MongoDB, interfaces dynamiques avec React ou Angular
- **Maîtrise des concepts fondamentaux :** routing, state management, services, modularité.

### Mini-Projets Machine Learning & IA Générative

- Mise en œuvre de classifieurs SVM pour différents jeux de données.
- Expérimentations en Computer Vision : prétraitement d'images, extraction de caractéristiques.
- Construction d'un pipeline RAG combinant embeddings + base vectorielle pour améliorer les réponses d'un LLM.

- 
- **Languages:** Arabe (Langue maternelle) — Français (Intermédiaire) — Anglais (Compétent)

- **Certifications:** Python Certiport , Intro to machine learning (Kaggle) , Feature Engineering (Kaggle), ReactJs Certification from Udemy , Google Data Analytics Professional Certificate (Coursera) AI For Everyone – Andrew Ng (Coursera)