

Soufian Boukir

soufianboukir.contact@email.com | linkedin.com/in/soufianboukir | soufian-boukir.vercel.app

SUMMARY

Junior Data Scientist building end-to-end ML pipelines, NLP systems, and analytics dashboards from raw data to deployed products. Strong in Python, statistics, model evaluation, and cloud deployment.

EDUCATION

Higher School of Technology

2025 – 2026

Bachelor of Science in Data Science & AI Engineering | GPA: 16/20

Beni Mellal, Morocco

- **Relevant Coursework:** Machine Learning, Statistics, Deep Learning, Data Mining & BI, Big Data Analytics

Specialized Institute of Applied Technology

2023 – 2025

Technical Diploma in Digital Development | GPA: 18/20

Tiznit, Morocco

- **Relevant Coursework:** Web Development, Databases (MySQL, MongoDB), Application Deployment

EXPERIENCE

Data Science Intern

Apr. 2026 – June 2026

YOOL Education

Agadir, Morocco

- Built a **full NLP pipeline** (TF-IDF, multilingual BERT, risk scoring) to analyze 2,339 multi-source learner feedbacks.
- Developed an **interactive Streamlit dashboard** surfacing KPIs, trends, weak signals, and LLM recommendations for pedagogical teams.
- Flagged **8 at-risk courses out of 93** using a weighted composite scoring model combining ratings, sentiment, trend degradation, and complaint frequency.

PROJECTS

Food Delivery Intelligence System ([Link](#))

- Built an **end-to-end ML pipeline** on 15K delivery records with preprocessing, feature engineering, training, and evaluation.
- Developed a delivery time regressor (R^2 : 0.998) and cancellation classifier with 55.7% recall; deployed via **AWS EC2** with **Flask, Docker and CI/CD** integration.

E-commerce Analytics & Revenue Forecasting ([Link](#))

- Built an **end-to-end analytics platform** on 100K transactions with cleaning, feature engineering, EDA, and KPI tracking.
- Benchmarked **4 forecasting models**; selected **XGBoost** with RMSE 61,570 and MAPE 3.71%; delivered Streamlit forecasting dashboard.

Robust Vision Benchmark ([Link](#))

- Evaluated **5 deep learning architectures** including MLP-3, MLP-5, LeNet-5, AlexNet, and ResNet-18 on CIFAR-10 under image corruptions.
- Tested **6 corruption types** across 5 severity levels; found ResNet-18 had 85.9% clean accuracy but 50.9% average corruption drop.

SKILLS

Programming : Python, TypeScript, SQL

Development & Databases : Next.js, Flask, MySQL, MongoDB

ML & Data Science : Pandas, NumPy, scikit-learn, PyTorch, TensorFlow, Streamlit, Apache Spark

Data Visualization : Power BI, Excel, Plotly, matplotlib, seaborn

Cloud & Tools : Git, GitHub, Docker, Linux, CI/CD, AWS EC2, GitHub Actions, LaTeX

Soft Skills & Languages : Team Collaboration, Problem-Solving, Communication

Languages : Arabic(Native), English(Advanced), French(Proficient)