ANAS MEFTAH

Tunis, Tunisia | +216 97 846 830 | anas.mefteh@etudiant-fst.utm.tn
linkedin.com/in/anas-meftah | github.com/anas-meftah

PROFILE

Data Science Engineering student (expected 2028) with hands-on experience building ML pipelines, drift detection systems, and RAG-based LLM applications. Proficient in Python, Scikit-learn, PyTorch, and MLOps tooling.

EDUCATION

National Engineering Diploma — Data Science — Faculty of Science of Tunis, Univ. El Manar | *2025 – 2028 (expected)*
Coursework: Algorithms, Machine Learning, Probability & Statistics, Distributed Systems, Graph Theory, AI

Integrated Preparatory Cycle (Engineering) — Faculty of Science of Tunis | *2023 – 2025*

Baccalaureate in Mathematics — Highest Honors — Sakiet Ezzit Pioneer High School, Sfax | *2023*

EXPERIENCE

R&D Member — Cybersecurity — Securinets FST Student Club, Faculty of Science of Tunis | *2024 – Present*

- Processed and cleaned the NSL-KDD intrusion detection dataset (~125,000 records, 42 features), performing EDA and correlation analysis to guide feature selection.
- Trained a Random Forest classifier with Bayesian hyperparameter tuning (skopt), achieving competitive F1 scores across attack categories; evaluated via confusion matrix, Precision and Recall per class.
- Delivered a 3-page interactive Streamlit dashboard (Overview · Feature Importance · Model Comparison) enabling non-technical stakeholders to explore model results.

PROJECTS

Feature Drift Detection System — Personal Project | *2024*

- Engineered an end-to-end pipeline comparing reference vs. production data streams using Kolmogorov–Smirnov tests with configurable alert thresholds; optimized for large datasets via vectorized NumPy/Pandas operations.
- Built modular preprocessing (imputation, scaling, encoding) and a visualization dashboard (histograms, density plots, drift scores) for feature-wise drift interpretation and automated retraining triggers.

MCP Server for PDF Q&A (RAG Pipeline) — Personal Project | *2025*

- Built a Model Context Protocol server integrating PDFs with LLMs via an automated ETL pipeline (extraction → chunking → embedding); configured a Vector Database for similarity-based retrieval using LangChain.
- Designed a conversational interface allowing multi-turn Q&A over large document corpora with context-aware responses.

Vision Transformer (ViT) Reproduction — Personal Project | *2024*

- Reproduced "An Image is Worth 16×16 Words" from scratch in PyTorch using modular OOP design; implemented custom DataLoader pipelines and optimized training with Adam and learning rate scheduling.

TECHNICAL SKILLS

Languages: Python (Advanced), SQL, Bash/Shell, Java (Basic)
ML & AI: Scikit-learn, PyTorch, LangChain, RAG, LLMs, Random Forest, Drift Detection, Anomaly Detection
MLOps & Stats: SciPy (KS test), skopt (Bayesian tuning), MLflow, Streamlit, Confusion Matrix / F1 analysis
DevOps & Tools: Git/GitLab, Docker, Linux CLI, Vector Databases, Pandas, NumPy, Matplotlib/Seaborn

LANGUAGES

Arabic (Native) **English** (Professional) **French** (Advanced)