Khadija Kriaa

Data & Decisional Systems Engineering Student – 6-Month End-of-Studies Internship from February 2027
+216 28 831 913 | kriaakhadija784@gmail.com | portfolio
linkedin | github

Profile

Final-year Data & Decisional Systems Engineering (IDSD) student at ENET'Com Sfax, seeking a 6-month end-of-studies internship (PFE) starting February 2027. Three Data and AI internships, two of them in banking: an end-to-end evaluated generative AI assistant (RAG, LLM, LangGraph), predictive modelling on web-scraped data (scikit-learn, XGBoost) and automated document processing (OCR, LLM).

Professional Experience

Union Internationale de Banques (UIB) **August 2026**
AI / Data Intern – Intelligent Assistant for Banking Circulars *Tunisia – 1 month*

- Developed an internal question-answering assistant over banking circulars (RAG architecture, Python, LLM) to give teams easier access to regulations.
- Built the ingestion pipeline: PDF extraction, article-aware chunking and embeddings (Sentence Transformers, BGE-M3) indexed in Qdrant.
- Implemented hybrid search (dense + BM25, Reciprocal Rank Fusion).
- Orchestrated the generation pipeline with LangGraph and built the user interface in Streamlit.
- Evaluated the system on 282 questions (MRR, Recall@K): 13 chunking strategies compared, 0.87 citation precision, 15/15 correct refusals on out-of-corpus questions.

Société Tunisienne de Banque (STB) **June – July 2026**
Data Engineering / Data Science Intern – Real Estate Price Revaluation *Tunisia – 2 months*

- Scraped over 16,000 real estate listings (Tayara.tn, NewKey), cleaned down to 4,413 sale listings (Pandas); exploratory data analysis and feature engineering (TF-IDF on descriptions).
- Built reproducible ML pipelines (scikit-learn Pipeline, ColumnTransformer, RobustScaler).
- Compared and tuned 7 regression models (Ridge, Lasso, ElasticNet, Random Forest, Gradient Boosting, XGBoost, LightGBM) with RandomizedSearchCV; best model XGBoost ($R^2 = 0.72$ on the test set).

Coconsult Nearshore **July 2025**
Software Development / AI Intern – Intelligent Payslip Processing *Tunisia – 1 month*

- Developed a payslip extraction platform: Tesseract OCR and field structuring via the Cohere API.
- Designed the PostgreSQL data model and the Spring Boot (Java) back-end API, secured with Spring Security and JWT.
- Contributed to the Angular (TypeScript) interface for reviewing and validating the extracted data.

Academic Projects

Tunisian Dialect Classification (team project) – *MARBERT, Transformers, Python* *Mar. – Apr. 2026*

- Fine-tuned MARBERT on about 48,000 tweets (Arabic and Arabizi) for binary sentiment classification: 92% accuracy, 0.90 macro F1 on the test set.

Hybrid Driver-Assistance Agent (team project) – *Python, Machine Learning, VEINS* *Jan. – Apr. 2026*

- Agent combining a reactive layer with a BDI model: driving-situation detection (ML) and contextual alerts.

Education

National School of Electronics and Telecommunications of Sfax (ENET'Com) **Sept. 2024 – 2027**
Engineering Degree – Data & Decisional Systems Engineering (IDSD) *Sfax, Tunisia*
Faculty of Sciences of Sfax **Sept. 2022 – June 2024**
Preparatory Classes for Engineering Schools – Mathematics-Physics (MP) *Sfax, Tunisia*

Technical Skills

- **Languages:** Python, SQL, Java, TypeScript
- **Data Science & Machine Learning:** Pandas, NumPy, scikit-learn, XGBoost, LightGBM, exploratory data analysis, feature engineering, regression
- **Generative AI & NLP:** RAG, LLM, LangGraph, embeddings, Sentence Transformers, BGE-M3, BM25, Hugging Face Transformers, Cohere API, Tesseract OCR
- **Databases:** PostgreSQL, Qdrant (vector database)
- **Development & Tools:** Streamlit, Spring Boot, Angular, Docker, web scraping (BeautifulSoup), Git, GitHub, Jupyter
- **Spoken Languages:** French (B2), English (B2), Arabic (native)

Volunteering

- **AIESEC Sfax** (2024 – 2025) – Information Management Lead: tracking internal data and building dashboards (Google Sheets).