Iliass Sijelmassi
Paris, France | *iliass.sijelmassi-idrissi@polytechnique.edu* | *github.com/iliassSjm*

Professional Summary

Joint MSc student at **École Polytechnique** and **HEC Paris** in Data Science for Business. Experience building end-to-end ML pipelines in **Python** (PyData stack, scikit-learn, PyTorch) and **SQL** on real-world datasets in finance and healthcare. Open to Data Scientist, ML Engineer, and Applied Scientist roles.

Education

École Polytechnique & HEC Paris — MSc Data Science for Business GPA: 3.79/4.00 *Sep 2024 – Jun 2026 (expected)*
- Courses: Probability, Statistics, Optimization, Machine Learning, Deep Learning, Time Series, Causal Inference, ML Opps.
- Tools: Python (NumPy, Pandas, scikit-learn, PyTorch), SQL, R, Jupyter, Git.

INP–ENSEEIHT, Toulouse — Dipl. in CS, Telecom & Applied Math *Sep 2020 – Sep 2023*
- Coursework: Algorithms, Data Structures, Statistics, Reinforcement Learning, Computer Architecture.

Experience

Data Scientist Intern | ***Crédit Agricole CIB (CACIB)**, Paris* *Mar 2025 – Sep 2025*
- Built predictive risk models (**XGBoost**, regularized logistic regression) to forecast overdue internal audit actions on **200k+** historical records, improving F1-score by **20%** vs. the existing heuristic baseline.
- Owned the full ML lifecycle: data cleaning, feature engineering, model selection, hyperparameter tuning, and deployment into a maintainable analytics codebase.
- Applied **time-aware validation** (PurgedGroupTimeSeriesSplit) and data-drift checks (**KS** / **PSI**) to ensure robustness under changing portfolios and reporting practices.
- Used **SHAP** to explain model behavior, identify key drivers, and help risk managers prioritize remediation efforts; presented findings to stakeholders with both technical and business backgrounds.
- Implemented reproducible pipelines and experiment tracking using **Python**, **pandas**, **scikit-learn**, **Dataiku**, and **MLflow**, with strict compliance and governance constraints.

Java Consultant Intern | ***Infosys – Renault**, Paris* *Mar 2023 – Oct 2023*
- Developed and integrated backend modules for supply-chain platforms (R3, EPO), improving data consistency and system reliability across logistics applications.
- Implemented business logic and database queries in **Java** (Maven), **PostgreSQL**, and **Oracle**, adhering to coding standards and review processes.
- Collaborated with product owners, QA, and other engineers to refine requirements, resolve bugs, and ship incremental improvements in a large codebase.

Software Developer Intern | ***Kuyper's Auto**, Lelystad, NL* *Jun 2022 – Aug 2022*
- Designed and built a modern web interface and online reservation system in JavaScript, reducing manual scheduling and phone calls.
- Worked directly with the business owner to gather requirements, iterate on UX, and deploy changes with version control.

Selected Machine Learning Projects

Leukemia Mortality Risk Prediction — ENS/QRT Challenge | *Python, RSF, CoxPH, SHAP* *2025*
- Predicted overall survival for leukemia patients using clinical and genomic features from **24 hospitals** (**4,500+** cases).
- Achieved **75.45% IPCW-C-index**, **+18%** vs. baseline, with Random Survival Forests and Cox models; used **SHAP** to interpret risk factors and support patient stratification.

Multi-Asset Allocation — Kaggle "Trust or Short?" | *Python, LightGBM, scikit-learn* *2025*
- Forecasted asset returns and allocation recommendations on panel time-series data using gradient boosting models.
- Designed **time-series-aware CV** (walk-forward / grouped splits) to avoid leakage; engineered lagged features and rolling stats; reached the **top quartile** of the public leaderboard.

Large Scale Log Anomaly Detection | *Python, PyTorch, Kafka, FastAPI* *2023*
- Built a synthetic log generator for microservice transaction traces with controllable failure modes (burst errors, incomplete flows), scalable to **tens of millions** of log lines.
- Trained an LSTM based sequence model to learn the normal "grammar" of logs and deployed it as a FastAPI service behind Kafka streaming, achieving **sub 50 ms** P99 latency for online anomaly scoring.

Skills and Interests

Programming: Python, C++, Java, SQL (PostgreSQL, MySQL), R, TypeScript, Bash
ML & Stats: Supervised & unsupervised learning, tree-based methods, gradient boosting, survival analysis, time-series modeling, reinforcement learning, hypothesis testing, model evaluation
Frameworks & Tools: scikit-learn, LightGBM, XGBoost, PyTorch, TensorFlow, Pandas, NumPy, Jupyter, Git, Docker, Dataiku, MLflow, Streamlit
Languages: French (native), English (TOEFL 105), Dutch (B2), Spanish, Portuguese, Arabic
Interests: Applied ML, finance, healthcare analytics, chess (1600 Elo), football, gym