

RAFAEL CAETITÉ

📍 Minas Gerais, Brazil 🇧🇷 Brazilian

☎ +55 28 99993 5364 ✉ rafaelmcaetite@gmail.com
🔗 github.com/rafaelcaetite 🔗 linkedin.com/in/rafaelcaetite
🌐 rafaelcaetite.dev

Summary

Computer Science undergraduate student at the Federal University of Viçosa (expected 01/2028) with solid experience in **Data Engineering**, **Machine Learning**, and **Artificial Intelligence**. Hands-on experience building end-to-end data pipelines, including automated ingestion via APIs (Copernicus CDS, Google Earth Engine), geospatial and tabular data processing (NetCDF, PySpark), and predictive modeling (XGBoost, LightGBM, Random Forest, Neural Networks) with Bayesian Optimization and Explainable AI (**SHAP**). Proven ability to develop REST APIs with **FastAPI**, concurrent architectures, and dynamic front-ends in **React**. Seeking an internship to deliver scalable impact in data engineering and applied AI solutions.

Experience

Federal University of Viçosa (UFV)

📅 09/2025 – Present

Undergraduate Research Fellow (PIBIC/CNPq) – Machine Learning & Data Engineering

📍 Viçosa, MG, Brazil

- Engineered and automated data ingestion and processing pipelines (Python) to extract large-scale spatial climate series (NetCDF) via Copernicus CDS and Google Earth Engine APIs, structuring over 40 years of historical data.
- Trained Machine Learning models (Random Forest, XGBoost, MLP) with dimensionality reduction (PLS) and Bayesian Optimization (*skopt*), validated via 6-year *Out-of-Time* testing and statistical hypothesis tests (ANOVA, Tukey HSD with 30 seeds).
- Developed an interactive full-stack agroclimatic dashboard (React, Vite, FastAPI) to deliver real-time metrics and forecasts to decision-makers, ensuring end-to-end product delivery.

Talon Tech

📅 2023 – Present

Freelance Software Developer (Full-Stack & Backend)

📍 Minas Gerais, Brazil

- Designed and implemented scalable RESTful APIs in Python and FastAPI with 15+ robust endpoints supporting business rules, authentication, and client data workflows.
- Modeled and optimized relational databases (PostgreSQL, MySQL), applying indexing and normalization strategies to accelerate operational query response times.
- Built responsive web interfaces using React, TypeScript, and Tailwind CSS, delivering administrative panels and scheduling systems with high reliability.

Featured Projects

Breast Cancer Mortality Prediction (METABRIC)

Python, LightGBM, Scikit-Learn, SHAP, Pandas

- Developed a predictive survival pipeline over 1,900+ patients and 600+ genomic/clinical features, applying early partitioning and causal *de-biasing* with LightGBM (AUC-ROC 0.68) for baseline-only early prognosis.
- Implemented Explainable AI with **SHAP** values and feature importance, identifying deterministic biomarkers (*ugt2b17*, *stat5a*, *mmp1*) to assist clinical interpretation.

RefManager Pro – Concurrent Data & RAG Engine

React, TypeScript, Web Workers, Gestalt LCS, Reger

- Engineered a high-performance client-side data engine for parsing, standardizing, and deduplicating massive bibliographic datasets (Scopus, WoS, IEEE) using *Web Workers* for multithreaded processing with 0 UI blocking.
- Implemented text similarity algorithms (Gestalt LCS) and intelligent batching logic for data ingestion into RAG pipelines and LLM agents.

Scientific Production & Leadership

- Organizing Director and Proceedings Editor of the 1st Tech & Science Exhibition (UFV):** Led the organizing committee, peer-review process, and official proceedings editorial for research in AI, Software Engineering, and Algorithms.
- Systematic Literature Reviews (UFV):** Author of 2 scientific literature reviews in: (1) Retrieval-Augmented Generation (RAG) architectures and (2) Machine Learning for climate time-series forecasting.

Education

Federal University of Viçosa (UFV)

📅 05/2022 – 01/2028

Bachelor's Degree in Computer Science

📍 Viçosa, Minas Gerais, Brazil

Technical Skills

- Programming Languages:** Python, SQL, JavaScript, TypeScript, Java, C++, C#
- Machine Learning & Data:** Scikit-Learn, LightGBM, XGBoost, Random Forest, Neural Networks (MLP, LSTM), SHAP (XAI), Bayesian Optimization (*skopt*), Pandas, NumPy, Statistics (ANOVA, Tukey HSD)
- Data Engineering & AI:** PySpark, Databricks, Delta Lake, ETL/ELT, APIs (Copernicus CDS, Google Earth Engine), NetCDF, RAG, Web Workers
- Backend, Cloud & Tools:** FastAPI, REST APIs, PostgreSQL, MySQL, Docker, Git, Linux, React, Vite, Tailwind CSS, LaTeX

Languages

- Portuguese:** Native | **English:** Advanced / Fluent | **Spanish:** Advanced