Antoine Pizzetta



DATA SCIENTIST

105 B Rue de Genève, 74240 Gaillard, France · +33 7 50 87 60 39 · anta.pizzetta@gmail.com
linkedin.com/in/antoine-pizzetta · antoinepizzetta-portfolio.pages.dev · github.com/antaww
French national (EU) · Eligible for Swiss cross-border work permit (G permit) · Category B driving licence, own vehicle · Available immediately

PROFILE

Graduate engineer in data science, specialising in machine learning (Ynov Aix, RNCP 39586 qualification, level 7). Three years of uninterrupted work-study, including one year at a payment institution within the Crédit Mutuel Arkéa banking group and sixteen months as a data scientist at a hearing care clinic in the healthcare sector, where I took over the company's data to build forecasting models and management tools used every day. Organised, I run projects end to end both on my own and as part of a team, and I enjoy learning new tools. What drives me is seeing a model leave the notebook and genuinely help business teams. I am looking for a first data scientist role in Geneva.

TECHNICAL SKILLS

Machine learning: Python, pandas, NumPy, SciPy, scikit-learn, XGBoost, LightGBM, time series (Prophet), feature engineering, clustering, reinforcement learning (Stable-Baselines3)
Deep learning, NLP and LLMs: PyTorch, TensorFlow, YOLOv8, OpenCV, Hugging Face, LoRA, RAG (LlamaIndex, FAISS), spaCy
Data: SQL, PostgreSQL, Snowflake, MongoDB, Redis, Parquet, ETL, Airflow, web scraping, Google Apps Script, GeoPandas
BI and visualisation: Power BI, Streamlit, Plotly, Google Sheets dashboards
MLOps and cloud: FastAPI, Docker, CI/CD GitHub Actions, AWS, Azure, Terraform, MLflow, pytest, drift monitoring
Development: TypeScript, SvelteKit, React, Java, Spring, Git

PROFESSIONAL EXPERIENCE

Data Scientist (work-study) — Jul 2025 – Oct 2026

Correction Auditive Durance, hearing care clinic (healthcare), Vitrolles, France

- Apps Script tools and automations used daily by around 15 internal users: about 3,000 runs per week with a 0.2% error rate
- Centralised customer, sales and stock data in Google Sheets, collected automatically through Apps Script and the Google Sheets API and feeding the sales tracking dashboards
- Monthly sales forecasting model built on historical data (linear regression with seasonality), used for business planning
- Banking flows: generation and checking of SEPA XML credit transfer files, diagnosis and correction of bank rejections

Full-Stack Developer (work-study) — Oct 2024 – Jul 2025

Eliophot, communications agency, Aix-en-Provence, France

- RAG chatbot for a hotel: website content collected by scraping (Puppeteer) and indexed with LlamaIndex, FastAPI API, answer rules and a guardrail for out-of-scope questions
- Internal SvelteKit and TypeScript platform bringing together the agency's tools: web SSH terminal (xterm.js), WebDAV file management, Cloudflare API integration, real-time monitoring over WebSocket
- Docker containerisation and deployment of services on Cloudflare

Analyst Developer (work-study) — Sep 2023 – Oct 2024

Monext, payment institution (Crédit Mutuel Arkéa group), Aix-en-Provence, France

- Java and Spring (Maven) back-end development on Payline, Monext's online payment platform: CRUD, batch processing, SQL queries
- Automation of recurring tasks for the technical team and unit testing with Mockito, in a regulated payment environment

DATA PROJECTS

Real estate price prediction in France (DVF + OpenStreetMap) **Jan – Aug 2026**

Python, XGBoost, LightGBM, GeoPandas, OSMnx, FastAPI, Streamlit, Docker, GitHub Actions

Certification project, carried out alone. XGBoost model trained on 20 million French property sales records (DVF, five years, 97 French departments) enriched with OpenStreetMap indicators, reaching an R^2 of 0.70 on time-based validation. FastAPI inference API, Streamlit dashboard, GitHub Actions CI with a blocking quality gate, drift monitoring (PSI, Kolmogorov-Smirnov) and automated retraining.

Object detection with a mixture of experts (COCO 2017) **Jun 2026**

Python, PyTorch, YOLOv8 (Ultralytics), torchvision, pycocotools, Grad-CAM

Team project. A classification router sends each image to YOLOv8s detectors specialised by class family, fine-tuned on COCO 2017. Rare classes rebalanced with copy-paste augmentation, per-class AP evaluation, analysis of false positives and negatives, interpretability with EigenCAM.

Autonomous driving with reinforcement learning **Nov – Dec 2025**

Python, CARLA, Gymnasium, Stable-Baselines3, OpenCV, TensorBoard

School project. Training of a driving agent in the CARLA simulator through a custom Gymnasium environment, with a reward function that accounts for traffic rules and varied traffic situations. Training with Stable-Baselines3, learning curves tracked in TensorBoard.

Putting a Whisper pipeline into production (MLOps) **Feb – Mar 2026**

Python, FastAPI, Redis, Docker, Terraform, Azure, GitHub Actions, Locust, Hugging Face

Audio transcription with Whisper, summarisation and question answering on the recording, served through a FastAPI API with Redis caching. Azure virtual machine provisioned with Terraform, continuous deployment with GitHub Actions, load testing with Locust.

Uta, music popularity and recommendation **Nov 2024 – Feb 2025**

Python, scikit-learn, XGBoost, LightGBM, Flask, React

Team project on Spotify data from Kaggle. Comparison of regression models to predict track popularity (linear models, random forests, gradient boosting) and K-Means recommendation, served by a Flask API and a React interface.

EDUCATION

Engineering degree in data science, specialising in machine learning **2021 – 2026**

Ynov Campus, Aix-en-Provence, France. RNCP 39586 qualification, level 7, Master's equivalent. Work-study programme.

LANGUAGES

French (native) · English (professional working proficiency)