

Antoine Pizzetta

DATA SCIENTIST

Marignane (13), France · +33 7 50 87 60 39 · anta.pizzetta@gmail.com

linkedin.com/in/antoine-pizzetta · antoinepizzetta-portfolio.pages.dev · github.com/antaww

Nationalité française, éligible au permis G (frontalier) · Permis de conduire B, véhiculé · Disponible en novembre 2026

PROFIL

Diplômé Ingénieur en science des données spécialisé en apprentissage automatique (Ynov Aix, titre RNCP 39586, niveau 7). Trois ans d'alternance sans interruption, dont seize mois sur un poste de data scientist où j'ai repris les données de l'entreprise pour en tirer des modèles de prédiction et des outils de pilotage utilisés au quotidien. Je cherche un premier poste de data scientist à Genève, disponible en novembre 2026 et prêt à m'installer dans la zone frontalière française.

COMPÉTENCES TECHNIQUES

Machine learning : Python, pandas, NumPy, SciPy, scikit-learn, XGBoost, LightGBM, feature engineering, clustering, apprentissage par renforcement (Stable-Baselines3)

Deep learning, NLP et LLM : PyTorch, TensorFlow, YOLOv8, OpenCV, Hugging Face, LoRA, RAG (LlamaIndex, FAISS), spaCy

Données : SQL, PostgreSQL, MongoDB, Redis, Parquet, ETL, web scraping, Google Apps Script, GeoPandas, Streamlit, Plotly

MLOps et cloud : FastAPI, Docker, CI/CD GitHub Actions, Azure, Terraform, MLflow, pytest, suivi de dérive

Développement : TypeScript, SvelteKit, React, Java, Spring, Git

EXPÉRIENCES PROFESSIONNELLES

Data Scientist (alternance)

juil. 2025 – oct. 2026

Correction Auditive Durance, Vitrolles, France

- Structuration et centralisation des données clients, ventes et stock dans Google Sheets, avec collecte automatisée par Google Apps Script et l'API Google Sheets
- Modèle de prévision des ventes mensuelles à partir de l'historique (régression linéaire avec saisonnalité), intégré au pilotage
- Tableaux de bord de suivi commercial alimentés automatiquement et automatisation des tâches récurrentes en Apps Script
- Génération et contrôle des fichiers de virement SEPA XML, diagnostic et correction des rejets bancaires

Développeur Full-Stack (alternance)

oct. 2024 – juil. 2025

Eliophot, agence de communication, Aix-en-Provence, France

- Plateforme interne SvelteKit et TypeScript centralisant les outils de l'agence : terminal web SSH (xterm.js), gestion de fichiers WebDAV, intégration de l'API Cloudflare, supervision en temps réel par WebSocket
- Chatbot RAG pour un hôtel : contenus du site collectés par scraping (Puppeteer) et indexés avec LlamaIndex, API FastAPI, règles de réponse et garde-fou sur les questions hors périmètre
- Conteneurisation Docker et déploiement des services sur Cloudflare

Analyste développeur (alternance)

sept. 2023 – oct. 2024

Monext, établissement de paiement, Aix-en-Provence, France

- Développement back-end Java et Spring (Maven) sur la solution de paiement Payline : CRUD, traitements batch, requêtes SQL
- Automatisation de tâches récurrentes pour l'équipe technique et tests unitaires avec Mockito

PROJETS DATA

Prédiction des prix de l'immobilier en France (DVF + OpenStreetMap)

janv. – août 2026

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

Projet de certification mené seul. Modèle XGBoost entraîné sur cinq ans de transactions DVF (97 départements) enrichies d'indicateurs OpenStreetMap, avec un R^2 de 0,70 en validation temporelle. API d'inférence FastAPI, tableau de bord Streamlit, CI GitHub Actions avec seuil de qualité bloquant, suivi de dérive (PSI, Kolmogorov-Smirnov) et réentraînement automatique.

Détection d'objets par mélange d'experts (COCO 2017)

juin 2026

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

Projet d'équipe. Un routeur de classification oriente chaque image vers des détecteurs YOLOv8s spécialisés par famille de classes, affinés sur COCO 2017. Rééquilibrage des classes rares par augmentation copy-paste, évaluation de l'AP par classe, analyse des faux positifs et négatifs, interprétabilité par EigenCAM.

Conduite autonome par apprentissage par renforcement

nov. – déc. 2025

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

Projet d'école. Entraînement d'un agent de conduite dans le simulateur CARLA à travers un environnement Gymnasium personnalisé, avec une fonction de récompense qui intègre le respect du code de la route et des situations de trafic variées. Entraînement avec Stable-Baselines3, suivi des courbes d'apprentissage sous TensorBoard.

Mise en production d'un pipeline Whisper (MLOps)

févr. – mars 2026

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

Transcription audio avec Whisper, résumé et réponse à des questions sur l'enregistrement, exposés par une API FastAPI avec cache Redis. Machine virtuelle Azure provisionnée avec Terraform, déploiement continu par GitHub Actions, tests de charge avec Locust.

Uta, popularité et recommandation musicale

nov. 2024 – févr. 2025

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

Projet d'équipe sur des données Spotify issues de Kaggle. Comparaison de modèles de régression pour prédire la popularité des titres (linéaires, forêts aléatoires, gradient boosting) et recommandation par K-Means, servies par une API Flask et une interface React.

FORMATION

Ingénieur en science des données spécialisé en apprentissage automatique

2021 – 2026

Ynov Campus, Aix-en-Provence, France. Titre RNCP 39586 de niveau 7, équivalent master. Cours suivi en alternance.

LANGUES

Français (langue maternelle) · Anglais courant (B2)