ANTONIO GONZÁLEZ SUÁREZ

antoniogsav20@gmail.com • antoniogonzalez.dev
linkedin.com/in/antonio-gonzalez-suarez-93055218a • github.com/AntonioGonzalezSuarez

ABOUT ME

Senior Data Scientist and AI project lead with 5 years taking machine learning and generative AI from the first client meeting to production. Spent two and a half years forward-deployed at a leading multinational brewer as the lead in-house data scientist, delivering a churn model with 2–3× the precision of the vendor model it replaced, and alerting products the sales force uses every day. Builds full-stack LLM applications and runs the DevOps behind them (CI/CD, Cloudflare, Hetzner with Coolify). Lead author of a chapter of the *Springer Handbook of Data Engineering*.

WORK EXPERIENCE

baobab AI (formerly baobab soluciones) Madrid, Spain

Senior Data Scientist & AI Project Lead 2024 – Present

Forward-deployed at a leading multinational brewer (2024 – 2026): lead in-house data scientist and project manager, owning each project from business definition to production and hypercare.

- **Distributor next-best-action alerts:** owned end to end the engagement's highest-value product, daily alerts for independent wholesalers' sales reps built on sell-out data. Agreed the API contract with the offshore integration team and each distributor, and fed the reps' answers back from their ERPs to silence or re-raise alerts.
- **Churn prediction:** replaced an external vendor's model with an in-house one scoring tens of thousands of customers monthly (LightGBM, calibrated, SHAP reasons): 2–3× the alert precision, and flagged customers churn 7–10× more often than average.
- **Sales alerting platform:** led the platform that turns sales, agreement, equipment, churn and P&L data into thousands of monthly alerts pushed to Salesforce.

Other client projects

- **Airport assistance demand:** forecast reduced-mobility passenger assistance at airports (LightGBM, MSTL, Prophet, web-scraped covariates), deployed with Docker.
- **Solar data quality (project manager):** rule-based anomaly detection plus autoencoder reconstruction of solar measurement data, shipped as v1.0 with CI and SonarQube.
- **Social services demand (project manager):** forecast demand for a national disability-support organisation's services (financial aid, private tutoring, guide dogs) to improve how resources are allocated.

Platform, generative AI and pre-sales

- Rolled out **MLflow** company-wide as an internal service with Google SSO and group access control, plus a starter project template.
- Built baobab's **time series framework**: backtesting, leakage-checked features and per-series model routing (Kedro, darts, polars).
- Shipped **generative AI marketing tools** end to end, including a production **lead-generation assistant** for an energy campaign that turns a conversation into a tailored solution summary, with lead analytics for marketing.
- **Pre-sales:** first client meetings, scoping, proposals, mockups, presentations and demos that help close deals.
- Agentic development with Claude Code and a custom multi-agent workflow, to iterate with the business much faster.

Data Scientist 2022 – 2024

- **Credit risk:** default-probability models for financial guarantees, and loan default forecasting so a lender could start collection and legal processes early.
- **Computer vision:** classified electrical grid assets from aerial imagery (CNNs, YOLO, GroundingDINO).
- **Quality analytics:** explained defects in electrical transformer manufacturing with CatBoost and SHAP.
- **Forecasting and causal inference:** demand for new retail products and accessory stock; causal impact of campaigns on beverage sales.

It&People, Gijón, Spain: Data Analyst 2021 – 2022

Data analysis of workplace surveys.

PUBLICATIONS

Predictive Analytics Models and Machine Learning 2026

A. González Suárez, G. González-Santander de la Cruz, Á. García-Sánchez. In F. Xhafa (ed.), *Springer Handbook of Data Engineering*, Springer Handbooks, pp. 159–170. doi:10.1007/978-3-031-92372-2_10

Survey of classification, regression and time series forecasting, with a retail case study benchmarking classical (AutoARIMA, ETS, MSTL) against deep learning (N-BEATS, N-HiTS) models on the M5 Walmart dataset.

TECHNICAL SKILLS

Programming Languages: Python, TypeScript, Go, SQL, Bash, MATLAB

ML & Data: NumPy, pandas, polars, PySpark, scikit-learn, LightGBM, CatBoost, Optuna, darts, Prophet, TensorFlow, SHAP, statsforecast, neuralforecast, causalimpact, Ultralytics, GroundingDINO

Generative AI: LangChain, PydanticAI, Claude API, Gemini (Google Gen AI SDK), Claude Code

MLOps & Orchestration: MLflow, Kedro, Apache Airflow, Celery, Redis, SonarQube

DevOps & Infrastructure: Docker, CI/CD (GitHub Actions), Hetzner VPS, Coolify, Cloudflare (DNS, Tunnel, Access), domain and Google Workspace administration

Web & APIs: FastAPI, Flask, Go, Vue.js, React Native (Expo), Hono

Data Platforms & Cloud: Databricks, Azure (Data Factory, DevOps), GCP, AWS, PostgreSQL, MongoDB

Others: Git, Jira, LaTeX

PROJECTS

cocinita.app

Meal-planning and recipe app for iOS, Android and web, built solo in TypeScript (Expo, Hono, PostgreSQL). An LLM pipeline turns recipe blogs and social-media captions into structured recipes. Self-hosted on Hetzner with Coolify, Cloudflare and CI/CD.

Open source

Contributor to baobab's mango library and Google's tfp-causalimpact; author of check-formatting-python, a GitHub Action that runs Black on pull requests.

AutoML (Master's thesis)

Web application for creating and automating machine learning models in the cloud.

EDUCATION

Master's in Big Data: Technology and Advanced Analytics 2023

Universidad Pontificia Comillas, Madrid, Spain

ICAI School of Engineering

Double Degree in Mathematics and Physics 2016 – 2022

University of Oviedo, Oviedo, Spain

Erasmus exchange at Sapienza Università di Roma, Italy (2019 – 2020)

ACADEMIC AWARDS

Quantum Computing Competition 2021

First place in the quantum computing competition organized by ArcelorMittal.

LANGUAGES

Spanish: Native • **English:** B2 (Cambridge) • **Italian:** B1