

Iñaki Martín Sorribes

Data Science Student · Co-Founder @ Iceberg Datum Analytics · Valencia, Spain

imxrtins@gmail.com · imartinsorribes.github.io · linkedin.com/in/inaki-martin-sorribes · github.com/imartinsorribes

SUMMARY

Data Science student at Universitat de València and co-founder of Iceberg Datum Analytics, an AI startup delivering ML and NLP solutions for the legal and healthcare sectors. Experience deploying production systems on AWS. Seeking a September 2026 internship as ML/NLP engineer or data scientist, open to fintech, retail, industrial, healthcare or energy.

EXPERIENCE

Co-Founder — Iceberg Datum Analytics S.L.

Nov 2025 — Present

Valencia, Spain · Part-time

- Co-responsible for business development: client acquisition, commercial pipeline management, and contract negotiation under NDA.
- Co-lead technical development of Iceberg Health (healthcare data platform) and Iceberg Law (LLM-based document processing).
- Designed and deployed production infrastructure on AWS using Docker and FastAPI.
- Currently running pilots under NDA with potential clients in both sectors.

Research Intern — IDAL, Universitat de València

Summer 2026

Valencia, Spain · Upcoming

- Selected for a research stay at the Intelligent Data Analysis Laboratory, the Machine Learning research lab of Universitat de València.

PROJECTS

Fraud Detection System — Personal Project

Python · XGBoost · SHAP · FastAPI

- End-to-end fraud detection pipeline on the PaySim dataset (Kaggle, ~6M mobile money transactions), addressing extreme class imbalance.
- Compared sampling strategies and evaluated using precision-recall AUC.
- Model explainability with SHAP and deployment as REST API with FastAPI on Docker. 99.76% recall on test set.

Iceberg Health — Healthcare Data Platform

Python · FastAPI · Docker · AWS

- Data platform for nursing home management systems, deployed in production on AWS.
- Currently running pilots under NDA with clients.

Iceberg Law — Legal Document Processing

Python · LLMs · NLP

- Automated extraction and structuring of legal documents and invoices using Large Language Models.
- On-premise GDPR/LOPD anonymization engine for Spanish PDFs: checksum-based regex (DNI, NIE, IBAN), NER with BSC-RoBERTa, and local LLM (Qwen3 via Ollama).
- Reproducible pipeline with SHA-256 audit log and post-redaction OCR verification.

Iberia Geo-Alert — Personal Project

Python · FastAPI · Sentinel Hub · Leaflet

- Real-time dashboard of natural events aggregating four public APIs (USGS, NASA EONET, NASA FIRMS, Sentinel Hub).
- On-demand Normalised Burn Ratio (NBR) analysis on Sentinel-2 imagery for wildfires in the Iberian Peninsula.
- Modular FastAPI + Pydantic backend, Leaflet frontend with clustering.

EDUCATION

B.Sc. in Data Science — Universitat de València

2023 — 2027

GPA 7.58/10 · Top grades with honors in Machine Learning and Probability & Simulation.

Coursework: Machine Learning, Statistics, Databases, Time Series Analysis, Deep Learning.

Higher Vocational Degree in IT — IES Abastos

2020 — 2022

Network and Systems Administration. 4-month internship at Slimbook: hardware assembly and Bash scripting for automated hardware testing.

AWARDS

Honorable Mention — Valencia City Data Visualization Contest

2025

Team project on flood-prone areas in the Valencian Community, awarded by Valencia City Council and ETSE-UV in response to the DANA floods.

SKILLS

Languages & ML: Python, SQL, Machine Learning, NLP, Large Language Models.

Cloud & Tools: AWS, Docker, FastAPI, Bash, Git, Streamlit, Power BI.

Languages: Spanish (native), Valencian/Catalan (native), English (B2).