

# Ibai Irastorza Azcarate, Ph.D.

Berlin, Germany | ibai.irastorza@gmail.com | +49 16099858498 / +34 608498979

Portfolio: batxes.github.io | linkedin.com/in/ibai-irastorza | github.com/batxes

## Professional Summary

Senior Data Scientist with over 13 years of experience in software engineering, machine learning, bioinformatics and leadership. Proven track record in leading cross-functional teams, developing AI-driven solutions and securing research grants.

## Professional Experience

### Senior Data scientist | Bioinformatics

06/2018 - 07/2024

Max Delbrück Center · Berlin · Full time

- ❖ Applied machine learning and data visualization techniques to uncover insights in genome structure and gene regulation.
- ❖ Designed and implemented end-to-end ETL pipelines for large-scale, heterogeneous omics datasets. Automated data ingestion, cleaning, normalization, transformation, and feature engineering to ensure high-quality inputs for downstream data analysis and machine learning workflows. Optimized data storage and retrieval for efficient processing and scalability.
- ❖ Led end-to-end AI and data-driven research projects, applying advanced statistical and machine learning techniques, resulting in 8 publications in top-tier journals (200+ citations).
- ❖ Secured €100,000 in research funding and established collaborations with three research institutions, presenting findings at international conferences.

### Data Scientist and Software Engineer | Bioinformatics

10/2013 - 05/2018

CABD · Sevilla · Full time

- ❖ Applied machine learning, statistical and optimization algorithms in simulations to generate 3D models and predict protein and genome structures.
- ❖ Developed a novel open-source 3D modeling algorithm using AI techniques, adopted by 10+ international research projects, showcasing expertise in algorithm development and modeling.
- ❖ Led data analysis and software development in cross-functional teams, resulting in three first-author publications with a combined impact factor of 90+.
- ❖ Managed the High-Performance Computing (HPC) infrastructure at the university, providing technical support to 100+ users, optimizing computational workflows, and improving resource efficiency.

### Software and Web Developer

05/2011 - 08/2012

Accenture · Madrid · Full time

- ❖ Maintained and developed critical banking applications serving 100+ daily users with 99.9% uptime and compliance with IT security protocols.

## Projects

- ❖ **Satellite Picture Classification (Deep Learning):** Built a CNN model to classify satellite images. Deployed via Flask API, containerized with Docker, and used Python libraries like TensorFlow and Pandas.

- ❖ **Weather Prediction (Machine Learning):** Applied XGBoost model to predict daily rainfall. Deployed via Flask API, containerized with Docker, and hosted on the cloud with AWS Elastic Beanstalk.
- ❖ **Drought Prediction (MLOps):** Created an MLOps pipeline using XGBoost for drought prediction. Orchestrated with Prefect, Managed with MLflow, monitored via Prometheus/Grafana, containerized with docker and automated with Docker. Ensured high code quality through unit/integration tests, linting, formatting, pre-commit hooks and a CI/CD pipeline. Streamlined execution with a Makefile.
- ❖ **Chemical Production Forecasting (Deep Learning):** Developed an LSTM model to forecast chemical production for a company, leveraging time-series data and deep learning techniques with TensorFlow.
- ❖ **Steel Type Classification (LLM):** Developed a machine learning pipeline to classify steel products into categories fine tuning the pre-trained BERT model and evaluating its performance.
- ❖ **Weather End-to-end Pipeline (Data Engineering):** Developed a data pipeline that automates the collection, storage, transformation, and visualization of Berlin's hourly weather data. I used Apache Airflow for orchestration, stored raw data in Google Cloud Storage, processed it in BigQuery with dbt, and visualized insights in Metabase. Infrastructure is provisioned with Terraform and containerized using Docker.

## Education

**Ph.D. in Bioinformatics** (2018), **MSc in Bioinformatics** (2012), **MSc in Computer Engineering** (2011), **BSc in Computer Engineering** (2009)

**Technical skills:** Python (Pandas, Scikit-learn, Matplotlib), SQL, R, Git, Docker, PyTorch, TensorFlow, AWS, Google Cloud, BigQuery, Deep Learning, MLOps, Data Visualization, Pipeline Development, Big Data Processing, 3D Modeling, AI model deployment, Data Engineering, NLP, LLM.

**Soft Skills:** Cross-functional Team Leadership, AI Strategy Development and Implementation, Technical Teaching and Knowledge Transfer, Project Planning and Mentoring, Analytical Thinking and Problem Solving, Scientific Writing and International Presentation