Ana Sofia Carmo

Data Scientist · Data-Driven Solutions · ML/AI · Digital Health

Bremen, Germany

anascacais@gmail.com | anascacais.github.io | in/anasofiacarmo | github.com/anascacais

Summary

Data Scientist with **5+ years** of experience in **Digital Health**, specializing in ML/AI, time series analysis, and translating patient data into actionable clinical insights. Experienced in **end-to-end workflows** across cross-functional projects, positioned at the **intersection of research and clinical translation**. Recognized for clear communication through 15+ invited talks and a 3MT ULisboa award.

Experience

Research Assistant & PhD Student at Telecommunications Institute (Lisbon, Portugal) *Feb 2021 - Present*

- Developed **probabilistic forecasting models**, namely logistic regression, LSTMs, and Bayesian phase modeling, for seizure risk estimation from **900+ hours** of physiological data, enabling accurate early warnings up to **24 hours before** events and informing algorithm adoption (codebase).
- Applied diverse **machine learning** techniques (such as KNN, SVMs, LR, and CNNs) across 3 other use cases, including seizure detection with **over 98%** sensitivity and precision (paper).
- Implemented **quantitative methods** (Wilcoxon signed-rank, Mann-Whitney U) to **justify the adoption** of a novel wearable device (codebase | paper). Performed mixed-effects meta-analysis to benchmark model performance across **40+ ML algorithms** (codebase | paper).
- **Led 5+ end-to-end cross-functional projects**, collaborating with clinical and technical stakeholders to align goals and share results; co-developed and **maintained 5 open-source**, community-driven Python projects.

Freelance Biomedical Data Scientist at Hospital de Santa Maria (Lisbon, Portugal) *Dec 2024 - Present*

- Engineered an automated pipeline for **extracting, cleaning, and transforming** time series recordings and patient metadata into structured datasets, enabling efficient analysis of 31 GB of data for downstream statistical modeling.
- Designed and conducted **hypothesis-driven analyses** using **mixed linear models** to evaluate HRV-based dysautonomia after surgical intervention, finding no significant improvement and informing clinical recommendations against the procedure.

Education

PhD in Biomedical Engineering at Instituto Superior Técnico (Lisbon, Portugal) *2021 - 2025*
Erasmus (Int'l Exchange) at Katholieke Universiteit Leuven (Leuven, Belgium) *2018 - 2019*
Integrated MSc in Biomedical Engineering at Instituto Superior Técnico (Lisbon, Portugal) *2015 - 2021*

Project Highlights

End-to-End Product Development

- Developed a **modular, automated, open-source pipeline** (GitHub + PyPI) for standardized development, evaluation, and **reproducibility of time-series ML models** in seizure forecasting, accelerating model deployment and ensuring methodological consistency (codebase).
- Designed a scalable **MySQL database** and FastAPI REST interface to provide secure, controlled access to physiological and clinical data (codebase). **Led migration of 2,500+ hours** of data from duplicated external drives.
- Engineered a **data collection system** featuring an autonomous Linux-based recording unit and a mobile app for real-time data visualization, enabling scaling of data collection (codebase | paper).

Communication and Leadership

- Co-authored 11 scientific publications (*IEEE T-BME*, *Frontiers*). Spoke at over 15 conferences, institutional briefings, and public events.
- Delivered over **80 hours of hands-on labs** in machine learning and biomedical instrumentation. **Mentored 100+** MSc and undergraduate research projects involving biosignal analysis and predictive modeling.
- Provided **cross-functional leadership** at ScientISST, guiding team operations, equipment use, and event planning while serving as the central point of contact for mission-aligned decision-making.

Skills

Programming Languages	Python, R, SQL, Bash
Data Science Libraries	ML/DL (Scikit-learn, PyTorch) | Manipulation (NumPy, Pandas) | Analysis (SciPy, Statsmodels, metafor) | Visualization (Matplotlib, Seaborn, Plotly)
Software Engineering	Versioning (Git) | Testing (unittest, pytest) | Packaging (PyPI) | CI/CD (GitHub Actions) | Containerization (Docker)
Backend Development	API Development (FastAPI, RESTful API) | Databases (MySQL) | Data Pipelines (ETL) | System Access (SMB)
Languages	Portuguese (native), English (C2), German (B1)

Note: Parts of this resume and cover letter were prepared with the assistance of generative artificial intelligence (AI) technologies, including OpenAI's ChatGPT, used to support language refinement and writing structure suggestions.