Rasoul **Samei**

Data Scientist · Clinical & Health Data Science

Bergamo, Italy | samei@live.com | linkedin.com/in/rsamei | rasoulsamei.com | ORCID

Data scientist in clinical research. At the Mario Negri Institute I apply statistics and machine learning to intensive care data, and contributed to the development and validation of the SOFA-2 organ failure score (*JAMA*, 2025). Most of my work is making fragmented clinical databases usable for analysis. I am looking for a doctoral position or a research role in health data, and I am open to industry roles across Europe.

EDUCATION

MA in Economics and Data Analysis

2021 – 2024

Università degli Studi di Bergamo, Italy · taught in English

Thesis: *Anomaly Detection in Financial Advisory Services: A Machine Learning Approach for Mortgage Conduct of Business Advisers*, supervised by Prof. Vittorio Moriggia ABSTRACT, ICSEM 2025

Selected coursework: Coding and Machine Learning (28/30), AI and Machine Learning for Finance (29/30), Empirical Methods in Impact Evaluation, Advanced Econometrics

BA in Business Management

2015

Islamic Azad University of Neyshabur, Iran

PUBLICATIONS

1. Ranzani OT, *et al.* (incl. **Samei R**) (2025). Development and Validation of the Sequential Organ Failure Assessment (SOFA)-2 Score. *JAMA* 334(23):2090–2103. doi:10.1001/jama.2025.20516
2. Shalchilar M, **Samei R**, *et al.* (2026). Assessing the Efficacy of a Sequential General Variational Mode Decomposition-Based Combination Model for US Wind Power Forecasting. *Modeling Earth Systems and Environment* 12(2):74. doi:10.1007/s40808-025-02693-5
3. **Samei R** (2025). Anomaly Detection in Financial Advisory Services: A Machine Learning Approach for Mortgage Conduct of Business Advisers. Conference abstract, *ICSEM 2025*, Ottawa. SOLE AUTHOR
4. Avigdor R, Giugni A, **Samei R** (2026). Refeeding syndrome in long-stay ICU patients: nearly universal, and harder to measure than we think. Poster, *GiViTI meeting 2026*, Pesaro, Italy, 4 November 2026.

EXPERIENCE

Data Scientist

Nov 2024 – Present

Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Ranica, Italy

- Contributed to the development and validation of the SOFA-2 score (*JAMA*, 2025): extracted and processed data from the MargheritaTre (GiViTI) ICU database and helped with statistical analysis and manuscript review
- Built a 900+ line SQL pipeline on Google BigQuery that turns a fragmented multi-table clinical database into a single validated analytical dataset, now reused by other teams and external researchers
- Added an LLM assistant to the study dashboard so researchers can filter data in plain language (in beta, being tested by some clinicians)
- Automated recurring processing and reporting workflows across multiple studies
- Data engineer on an ICU nutritional-therapy residency thesis (University of Bologna): built a no-code extraction workbook so the clinician could run the analysis without a data-management background

Data Scientist, MA Traineeship

Dec 2023 – Jun 2024

The Openwork Partnership, Swindon, United Kingdom

- Designed an unsupervised anomaly detection system (Azure Machine Learning) for regulated mortgage advisory data; the methodology became my MA thesis and a conference abstract (ICSEM 2025)
- Built data pipelines (Azure Python SDK v2, Spark SQL) and validated 28+ KPIs for senior supervisors

Machine Learning Teaching Assistant

Mar – Jun 2023

University of Bergamo, Italy

Led group tutoring on model building, evaluation and data analysis.

METHODS & TOOLS

Statistics: regression modelling, hypothesis testing, time series forecasting, multivariate analysis (PCA, MCA, FAMD), model calibration and validation

Machine learning: supervised and unsupervised learning, anomaly detection, clustering, NLP (spaCy), LLMs and RAG (LangChain, pgvector)

Health data: GiViTI and MIMIC ICU databases · HL7 FHIR R4 · OMOP Common Data Model

Programming: Python (pandas, NumPy, scikit-learn, XGBoost), R, SQL, SAS · BigQuery, PostgreSQL, Databricks · Git, MLflow, Docker

AWARDS & LANGUAGES

Award: First Place, Stat-Hackathon Competition, University of Bergamo (2022)

Languages: Persian (native) · English (C1) · Italian (A2) · German (basic)

REFERENCES

Available on request.