# Gabriele Paganelli

**Data Scientist & Statistician**

## CONTACTS


gabrielepaganelli13@gmail.com
https://GabriPaganelli.github.io
www.linkedin.com/in/gabriele-paganelli
https://github.com/GabriPaganelli


## PROFILE

Aspiring Data Scientist with a strong background in **statistical modeling** and **machine learning**, focused on turning complex and noisy data into insights that support **real decisions**. Interested in the structure behind data, I combine mathematical reasoning with domain knowledge to build interpretable and reliable models.

## SKILLS

**Statistics & Machine Learning**

- Bayesian and frequentist inference
- Regression and classification
- Time series, tree-based methods
- Deep learning, clustering, networks
- Model validation and interpretation

**Programming & Tools**

- R, Python, Stan
- LaTeX, Git

## LANGUAGES

Italian (native), English (C1–C2)
Spanish (basic), Latin (A1)

## ACTIVITIES

- Teaching Assistant, Dept. of Biology
- Stat Data Camp, University of Padua
- Vivi 3 Giorni da Manager (Accenture)
- AI interview on Canale Italia

## EDUCATION

### MSc in Statistics

University of Padua — 2024-present

- Thesis (ongoing): interpretable deep clustering with sparse VaDE
- Statistical modeling, machine learning and data-driven methods
- Exchange semester at **Boston University**

### BSc in Statistics for Technologies and Sciences

University of Padua — 2021-2024

- Final grade: 110/110 cum **laude**
- Thesis: “The Web of Fame: Network Analysis on Pantheon Data”

### Scientific High School Diploma

ISIS Einaudi, Dalmine — 2016-2021

- **Exchange year** in Canada (2019–2020) | 100/100 cum laude

## EXPERIENCES AND PROJECTS

### APPalti: PU Learning for Risk Detection — 2026

- **PU learning** for corruption risk on ~9.5M public contracts
- PU/PNU risk scoring models and Streamlit app

### Bayesian Soil Nutrient Profiles — 2026

- **Hierarchical model** of soil C, N and P profiles (40 fields)
- Multivariate random effects in Stan, projpred variable selection

### Italian Energy Market Analysis — 2026

- Energy, gas and renewable price time series analysis
- **SARIMA, VAR, TFM and cointegration** models

### IceCube Neutrino Clustering and Source Modeling — 2025

- **Bayesian mixture models** of neutrino sources (IceCube data)
- Clustering and inference on sparse, noisy data

### Mice LFP Data: GRU-Transformer — 2025

- Analysis of mouse local field potential (LFP) neural signals
- Hybrid **GRU–Transformer architecture** for temporal modeling

### Simpson vs Griffin: Text Mining — 2025

- Comparative **text analysis** of The Simpsons and Family Guy
- Sentiment analysis, LDA and STM topic models

## AWARDS

- **First Prize** – APPalti, Statistics Lab with Companies (2026)
- **Best Prediction** – Stats under the Stars, Italian Stat. Society (2026)
- **First Place** – Prima Math Fest (2026)