Machine Learning to Extract Predictive Signals from the Pollino Seismic Catalog

Mariagiusi Nicodemo (2114171)
Rome, December 18, 2025

SAPIENZA
UNIVERSITÀ DI ROMA

Key topics and main questions

01
Data availability
What seismic catalog data were available for the Pollino area?

02
Data handling
How was the seismic catalog selected and cleaned?

03
Data pre-processing
How was the dataset prepared for exploratory and predictive analysis?

04
Features engineering
What catalog-based features were used for learning?

05
Model and Training
Which machine learning models were tested and how were they trained?

06
Performance and conclusions
What does the model performance tell us about the predictive information in the catalog?

Project overview

Why?

Traditional statistical methods struggle to provide reliable short-term predictions of earthquakes when applied to **real seismic catalogs**. Machine Learning offers an alternative, exploratory approach to investigate **whether real seismic data contain predictive information**, even if limited.

My motivating idea

Rather than aiming at deterministic earthquake prediction, Machine Learning can be used as a **diagnostic tool** to test the information content of seismic catalogs.
The key question is whether catalog-based variables encode **statistical patterns** that can be partially captured by data-driven models.

The goal of this project is to apply Machine Learning techniques to the **Pollino seismic catalog** to assess the presence of predictive information in real seismic data and to discuss the intrinsic limitations related to data sparsity, noise, and system complexity.

Machine Learning Models

To explore the predictive information contained in the Pollino seismic catalog, I tested simple and robust machine learning models.

•**Random Forest**
Used as a baseline model to capture nonlinear relationships without strong assumptions.

•**Gradient Boosting**
Applied to evaluate whether sequential ensemble methods improve performance.

•**Extra Trees**
Used to test model stability and sensitivity to randomness.
Ensemble methods are well suited for exploratory analysis of complex and noisy geophysical data.