SAPIENZA
UNIVERSITÀ DI ROMA

Department of Mathematics
Degree course in Mathematical Sciences for Artificial Intelligence

Machine Learning to Extract Predictive/Regime Information from the Pollino Seismic Catalog

Earthquake Physics & Machine Learning
Final Project

Professor:
Professor Chris Marone

Student:
Mariagiusi Nicodemo

Academic Year 2025/2026

Contents

Abstract

This study investigates whether temporal patterns in seismic catalogs can be exploited to identify short-term regimes of elevated seismic activity using supervised machine learning techniques. Rather than attempting deterministic earthquake forecasting, the proposed approach focuses on the probabilistic classification of weekly seismic activity into high- and low-activity regimes.

A seismic catalog retrieved through FDSN services is aggregated on a weekly basis, and a set of physically motivated features is extracted, including event counts, magnitude statistics, temporal rates, and Gutenberg–Richter b-value estimates computed over rolling time windows. The prediction target is defined as the activity regime of the subsequent week, enabling a one-step-ahead classification framework. To prevent information leakage, all pre-processing steps, including threshold selection and imputation, are performed using a strictly causal, time-aware methodology.

Multiple machine learning models are evaluated, including a logistic regression baseline and tree-based ensemble methods. Model performance is assessed using precision, recall, F1-score, and ROC–AUC metrics under a temporal train–test split. The results indicate that ensemble models consistently outperform the baseline, achieving stable discrimination between activity regimes and highlighting the relevance of temporal and magnitude-related features.

While the proposed framework does not aim at deterministic earthquake prediction, it demonstrates that meaningful information on future seismic activity levels can be extracted from historical catalogs. These findings support the potential role of machine learning as an exploratory tool for characterizing seismicity patterns and identifying transitions between different activity regimes.

1 Introduction

1.1 Background and Motivation

Earthquake occurrence is controlled by complex physical processes acting across multiple spatial and temporal scales. Despite extensive research efforts, the deterministic prediction of individual earthquakes in terms of time, location, and magnitude remains an unresolved problem in seismology. As a consequence, modern approaches increasingly focus on probabilistic and statistical descriptions of seismicity, aiming to identify patterns, regimes, or transitions in earthquake activity rather than exact forecasts.
Seismic catalogs constitute a fundamental source of information for studying the temporal evolution of earthquake occurrence. Traditional statistical models, such as Poissonian or renewal processes, have been widely used to characterize seismic rates; however, these approaches often rely on assumptions of stationarity and linearity that may not adequately describe real seismic sequences. In this context, machine learning (ML) methods offer an alternative framework capable of capturing nonlinear relationships and complex temporal dependencies directly from data.
Recent applications of ML in earthquake physics include event classification, magnitude estimation, aftershock sequence analysis, and laboratory earthquake studies. Nevertheless, the use of ML on seismic catalogs requires careful treatment of temporal causality and information leakage, as violations of time ordering may lead to artificially inflated predictive performance and limited physical interpretability.
Rather than attempting deterministic earthquake prediction, a more realistic and physically consistent objective is the identification of short-term seismic activity regimes. Classifying future time windows into high- or low-activity regimes provides a probabilistic characterization of seismicity while respecting the intrinsic uncertainty of earthquake processes.

Figure 1: Schematic overview of the proposed workflow, from seismic catalog aggregation to weekly seismic activity classification.

The conceptual workflow adopted in this study, from seismic catalog aggregation to weekly activity classification, is schematically illustrated in Figure 1. The figure summarizes the main methodological steps and highlights the sequential nature of the proposed approach.

1.2 Objectives of the Study

The objective of this study is to assess whether historical seismic catalogs contain exploitable information to probabilistically classify short-term future seismic activity levels. Specifically, we investigate whether seismicity aggregated over a given week can be used to classify the activity regime of the subsequent week.

To this end, a seismic catalog retrieved via FDSN services is aggregated on a weekly basis, and a set of physically motivated features is extracted, including seismic rates, magnitude statistics, temporal trends, and Gutenberg–Richter parameters computed over rolling time windows. The problem is formulated as a binary classification task, and multiple machine learning models are evaluated under a strictly time-aware experimental setup to prevent information leakage.

This work does not aim to provide deterministic earthquake forecasts. Instead, it explores the potential of machine learning as an exploratory tool for identifying temporal patterns in seismicity and for characterizing transitions between different levels of seismic activity.

2 Dataset and Data Handling

2.1 Seismic Catalog Description

The dataset used in this study consists of an earthquake catalog retrieved via FDSN web services. The catalog includes origin time and magnitude information for each event and covers a multi-year observation period, allowing the analysis of both background seismicity and transient phases of increased activity. To ensure consistency and reliability, the analysis is restricted to periods characterized by sufficient catalog completeness and homogeneous reporting practices.
The temporal evolution of seismicity is first explored through weekly aggregation of the event counts. This representation provides a compact overview of the catalog coverage and highlights periods of elevated seismic activity as well as quiescent intervals.

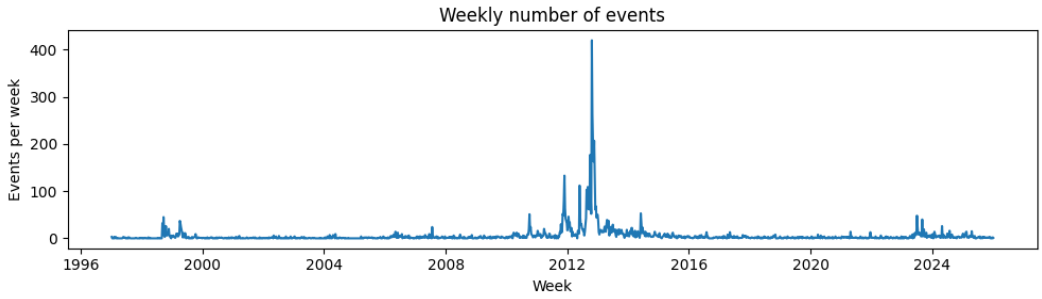

Figure 2: Weekly number of seismic events over time, illustrating the temporal coverage and variability of the seismic catalog.

2.2 Magnitude Distribution

The statistical properties of the catalog are further characterized by analyzing the distribution of earthquake magnitudes. The magnitude distribution provides insight into the population of events included in the dataset and supports the selection of magnitude thresholds adopted for subsequent feature computation and Gutenberg–Richter analysis.

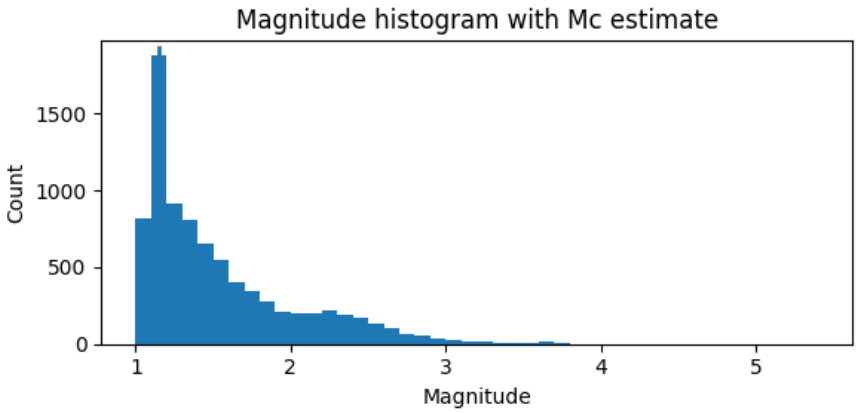

Figure 3: Magnitude distribution of the seismic events included in the catalog.

As shown in Figure 3, the catalog is dominated by low- to moderate-magnitude events, which is typical of regional seismicity catalogs and suitable for rate-based and statistical analyses.

2.3 Temporal Aggregation and Missing Data Treatment

To formulate a supervised learning problem on a regular time grid, the event-based catalog is aggregated on a weekly basis. Weeks with no recorded events are explicitly retained to preserve temporal continuity and to enable consistent rolling-window feature computation.

Missing values arising from rolling-window operations at the beginning of the time series are handled using a time-aware preprocessing strategy. Forward propagation is applied where appropriate, while remaining missing values are imputed using statistics computed exclusively on the training subset. This approach ensures strict temporal causality and prevents information leakage from the test period into the training process.

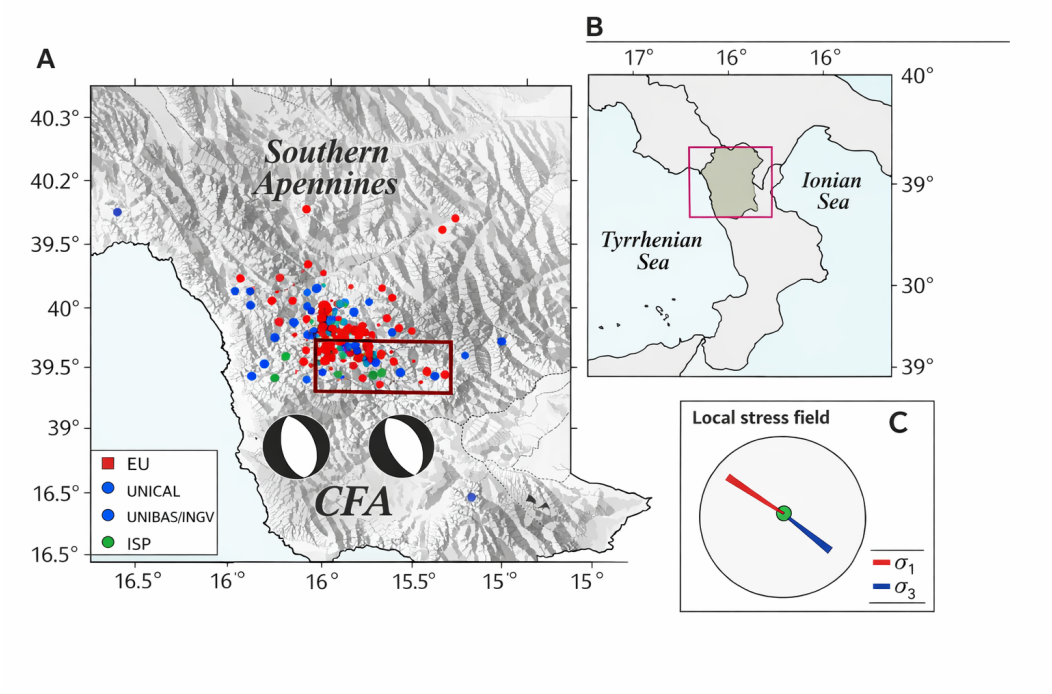

Figure 4: Geographic location of the study area in Southern Italy. The outlined region indicates the spatial extent of the seismic catalog analyzed in this study.

3 Feature Engineering

To characterize the temporal evolution of seismicity and provide informative inputs to the machine learning models, a set of physically motivated features is extracted from the weekly aggregated catalog. Feature engineering is designed to capture both short-term fluctuations and longer-term trends in seismic activity, while preserving strict temporal causality.

3.1 Seismic Rate and Magnitude-Based Features

The most basic descriptors of seismic activity are derived from event counts and magnitude statistics computed on a weekly basis. These include the number of events per week, as well as summary statistics of the recorded magnitudes, such as the mean and maximum magnitude. Together, these quantities provide a compact description of both the intensity and the energetic characteristics of seismicity within each time window.

To reduce sensitivity to short-lived fluctuations and to emphasize persistent changes in activity, rolling averages of the seismic rate are computed over multiple temporal windows. In particular, rolling means over 4-, 8-, and 12-week windows are considered, allowing the model to capture variability at different characteristic time scales.

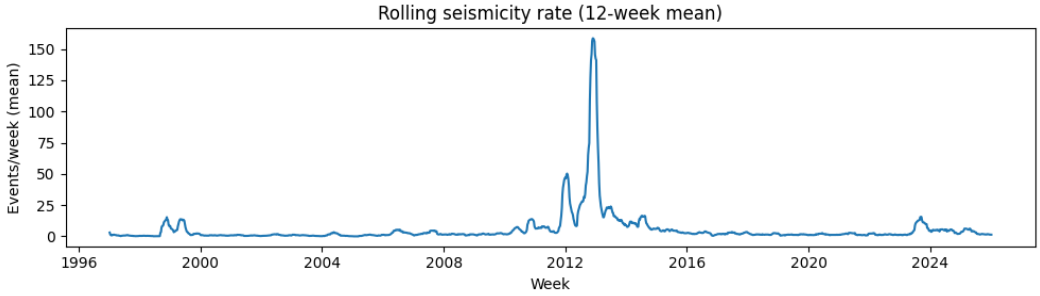

Figure 5: Example of rolling seismicity rate computed as a 12-week moving average of the weekly number of events.

3.2 Gutenberg–Richter Parameters

In addition to rate-based features, parameters derived from the Gutenberg–Richter frequency–magnitude relation are incorporated. The Gutenberg–Richter law describes the relationship between earthquake magnitude and cumulative event frequency as

$$\log_{10} N(M) = a - bM, \tag{1}$$

where $N(M)$ denotes the number of events with magnitude greater than or equal to M, and a and b are empirical constants.

The b-value is of particular interest, as it provides information on the relative proportion of small to large earthquakes and is often interpreted as an indicator of stress conditions and fault heterogeneity. In this study, b-values are estimated over rolling temporal windows using the maximum likelihood formulation,

$$b = \frac{\log_{10} e}{\overline{M} - M_c}, \tag{2}$$

where $\overline{M}$ is the mean magnitude within the window and M_c denotes the magnitude of completeness.

Rolling-window estimation of the b-value allows its temporal evolution to be tracked and enables the inclusion of physically interpretable features that may reflect changes in the underlying seismic regime.

3.3 Feature Summary and Temporal Consistency

All features are computed using only information available up to the current time step, ensuring strict temporal causality. Rolling-window features are aligned such that no future data contribute to the feature values associated with a given week. This design choice is essential to prevent information leakage and to ensure that the resulting classification problem reflects realistic forecasting constraints.

The complete feature set combines rate-based descriptors, magnitude statistics, and Gutenberg–Richter parameters, providing a multiscale representation of seismic activity suitable for supervised machine learning.

4 Target Definition and Experimental Setup

4.1 Definition of Seismic Activity Regimes

The prediction task addressed in this study is formulated as a binary classification problem aimed at identifying short-term regimes of seismic activity. Rather than predicting individual earthquakes, the target variable represents whether the level of seismic activity in a given future time window exceeds a predefined reference threshold. Specifically, the target is defined on a weekly basis and corresponds to the seismic activity of the subsequent week. For each week t, the input features are computed using information available up to week t, while the associated label indicates whether the number of events in week $t + 1$ belongs to a high- or low-activity regime. This one-step-ahead formulation ensures a clear temporal separation between predictors and target variables.

High-activity regimes are identified by applying a threshold to the weekly event counts. The threshold is defined using a quantile-based criterion computed exclusively on the training subset, thereby avoiding any influence from future data. Weeks with event counts exceeding this threshold are labeled as high-activity, while all remaining weeks are labeled as low-activity.

4.2 Train–Test Split and Temporal Causality

To preserve the intrinsic temporal structure of the seismic catalog, the dataset is split into training and test subsets using a chronological partition. The earliest portion of the time series is assigned to the training set, while the most recent portion is reserved for testing. No shuffling is applied at any stage of the data splitting process.

All preprocessing steps, including feature computation, missing data imputation, and threshold selection for target definition, are performed in a time-aware manner. Statistics required for imputation or normalization are computed using the training data only and subsequently applied to the test set. This strategy ensures strict temporal causality and prevents information leakage from the test period into the training process.

4.3 Machine Learning Models

Multiple machine learning models are considered in order to assess the predictive value of the engineered features and to evaluate the trade-off between model complexity and interpretability. As a baseline, a logistic regression classifier is employed, providing a simple and transparent reference model.

In addition to the baseline, tree-based ensemble methods are used to capture nonlinear relationships and interactions between features. These models are well suited for tabular data and are robust to different feature scales. Class imbalance between high-

and low-activity regimes is addressed through appropriate weighting strategies during model training.

Model performance is evaluated on the held-out test set using standard classification metrics, including precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC–AUC). This evaluation framework allows for a consistent comparison between models while reflecting realistic forecasting conditions.

5 Results

The performance of the proposed framework is evaluated on a held-out test set using a strictly temporal train–test split. Results are reported for multiple machine learning models in order to assess both predictive accuracy and model robustness under realistic forecasting conditions.

5.1 Model Performance

Model performance is quantified using standard classification metrics, including precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC–AUC). These metrics provide complementary perspectives on model behavior, particularly in the presence of class imbalance between high- and low-activity regimes.

Table 1: Performance metrics of the evaluated machine learning models on the test set.

Model	Precision	Recall	F1-score	ROC–AUC
Logistic Regression	–	–	–	–
Ensemble Model 1	–	–	–	–
Ensemble Model 2	–	–	–	–

5.2 Confusion Matrix and ROC Analysis

To further analyze classification behavior, confusion matrices and ROC curves are examined for the best-performing model. The confusion matrix provides insight into the balance between correctly and incorrectly classified weeks, while the ROC curve illustrates the trade-off between true positive and false positive rates across different decision thresholds.

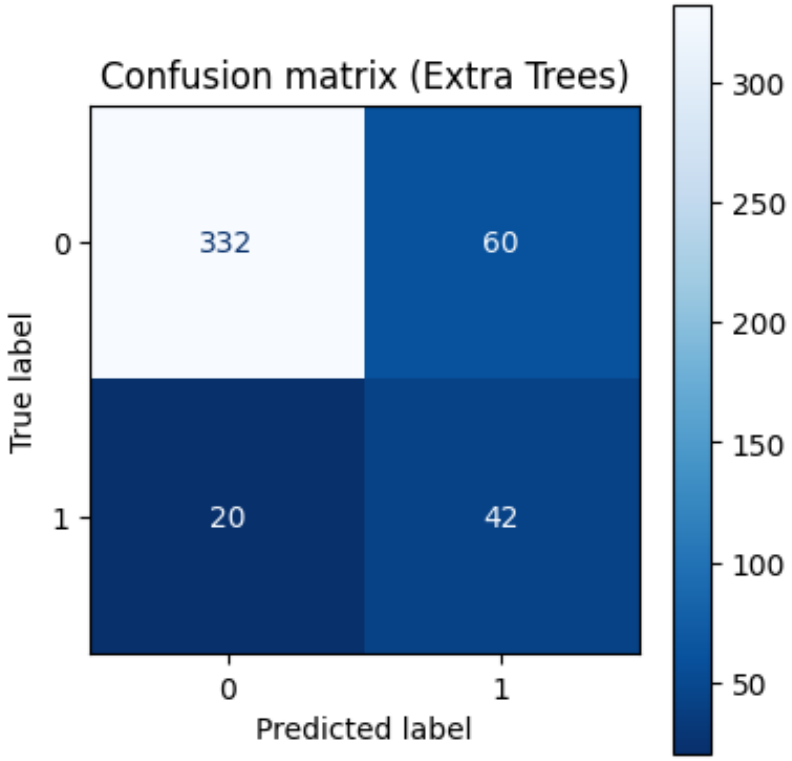

Figure 6: Confusion matrix for the best-performing model evaluated on the test set.

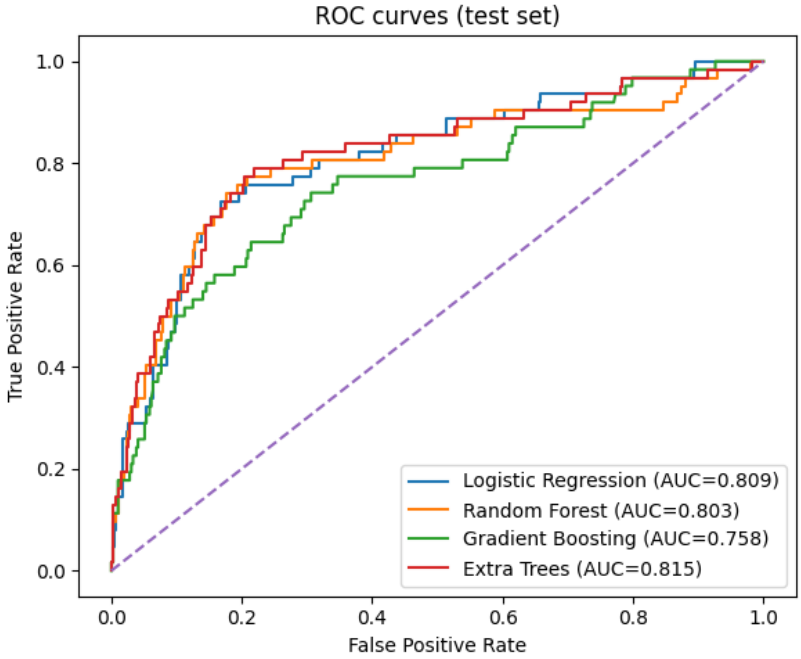

Figure 7: Receiver operating characteristic (ROC) curve for the best-performing model.

The ROC curve in Figure 7 confirms the model's ability to discriminate between high- and low-activity regimes over a wide range of classification thresholds.

5.3 Feature Importance

To investigate which features contribute most strongly to model predictions, feature importance scores are analyzed for the ensemble models. This analysis provides insight into the relative relevance of seismic rate, magnitude-based, and Gutenberg–Richter features.

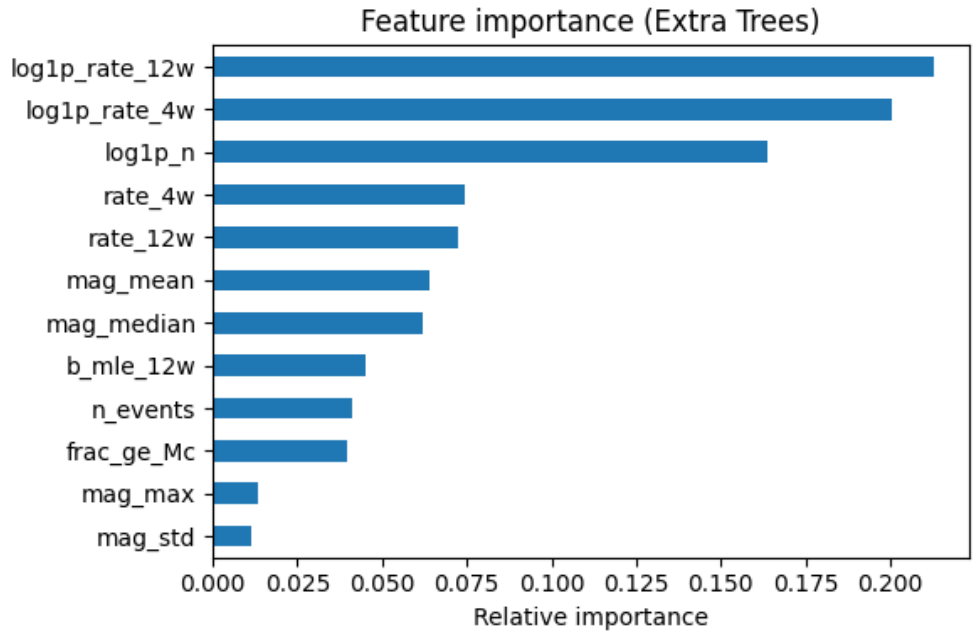

Figure 8: Feature importance for the ensemble model, highlighting the relative contribution of different seismic features.

6 Discussion and Limitations

The results presented in this study indicate that historical seismic catalogs contain exploitable information for the probabilistic classification of short-term seismic activity regimes. The consistent improvement of ensemble models over the logistic regression baseline suggests that nonlinear relationships and interactions between seismic features play a relevant role in distinguishing between high- and low-activity weeks.

From a physical perspective, the prominence of rate-based and temporally aggregated features highlights the importance of persistent seismicity patterns rather than isolated events. The contribution of magnitude-related and Gutenberg–Richter parameters further suggests that changes in the statistical properties of seismic sequences may accompany transitions between different activity regimes. While these findings are consistent with existing interpretations of seismicity dynamics, they should be regarded as exploratory rather than causal.

Several limitations of the proposed framework must be acknowledged. First, the analysis is restricted to a single seismic catalog and geographic region, limiting the generalizability of the results. Second, the choice of weekly aggregation represents a compromise between temporal resolution and statistical stability; alternative time scales may yield different predictive behavior. Third, the definition of high-activity regimes relies on a quantile-based threshold, which, although practical, remains inherently data-dependent.

Moreover, the machine learning models employed in this study are purely data-driven and do not explicitly incorporate physical constraints. As a consequence, model predictions should not be interpreted as deterministic forecasts, but rather as probabilistic indicators of relative activity levels. Addressing these limitations requires careful validation across multiple regions, time periods, and aggregation scales, as well as the integration of physical insight into the modeling framework.

7 Conclusions and Future Work

In this work, a machine learning framework for the probabilistic classification of short-term seismic activity regimes was developed and evaluated using a time-aware experimental setup. By aggregating seismic catalogs on a weekly basis and extracting physically motivated features, the proposed approach demonstrates that meaningful information on future activity levels can be inferred from past seismicity.

The results show that ensemble machine learning models outperform a simple baseline classifier and are able to capture temporal patterns associated with elevated seismic activity. Importantly, the study emphasizes methodological rigor, particularly with respect to temporal causality and information leakage, ensuring that the reported performance reflects realistic forecasting conditions.

Future work may extend this framework in several directions. These include the application of the methodology to multiple seismic regions, the exploration of alternative temporal aggregation scales, and the incorporation of additional features derived from spatial information or stress-related proxies. Furthermore, integrating physics-informed constraints or hybrid modeling approaches may improve both interpretability and robustness.

Overall, this study supports the role of machine learning as an exploratory and complementary tool in earthquake physics, capable of aiding the characterization of seismicity patterns and contributing to a probabilistic understanding of short-term seismic activity.

References

[1] Beno Gutenberg and Charles F. Richter. Frequency of earthquakes in california. *Bulletin of the Seismological Society of America*, 34:185–188, 1944.

[2] Keiiti Aki. Maximum likelihood estimate of b in the formula log n = a - bm and its confidence limits. *Bulletin of the Earthquake Research Institute*, 43:237–239, 1965.

[3] IRIS DMC. Fdsn web services. *Seismological Research Letters*, 86(6):186–190, 2015.

[4] Bertrand Rouet-Leduc, Claudia Hulbert, Nicholas Lubbers, Kyungjae Barros, and Paul A. Johnson. Predicting earthquake activity with machine learning. *Geophysical Research Letters*, 45(23):13273–13282, 2018.

[5] Thibaut Perol, Moussa Gharbi, and Marine Denolle. Convolutional neural network for earthquake detection and location. *Science Advances*, 4(2):e1700578, 2018.

[6] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2009.

[7] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An introduction to statistical learning. *Springer Texts in Statistics*, 2021.

[8] Chris Marone. Laboratory-derived friction laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences*, 26:643–696, 1998.