Target Definition

Document ID: 2.8

Project: Earthquake EDA & Prediction

Date: 2025-12-27

Document Control

Document ID	2.8
Title	Target Definition
Version	[v1.0]
Status	Draft
Owner	Donghun Kim / Head of Development and Research
Last Updated	2025-12-27

Revision History

Version	Date	Author	Change Summary
v1.0	2025-12-27	Donghun Kim	Initial draft.

Approvals

Name	Role	Date	Signature
Donghun Kim	Head of Development and Research	2025-12-27	Donghun Kim

Table of Contents

Purpose

This documentation precisely defines the project's prediction task (target) and serves as a 'Single Source Of Truth(SSOT)' for labeling, dataset configuration, and dashboard output.

Prediction Outputs

The definite output that this service and dashboard must provide includes following 3 things:

1) Location (Area): hazard map(cell, local risk score or probability) + Top-K candidates.
2) Probability (occurrence): P(event | cell, horizon)
3) Magnitude: conditional magnitude prediction (E[M] or interval probability when occurring)

Task Formulation

Task A - Occurrence (Classification)

- Label: $y\_occ(c, t) \in \{0, 1\}$

- Definition: 1 if at least one event of magnitude ≥ M_th occurs in cell c during the next T, relative to time t, otherwise 0.

- Output: $p\_occ(c, t) \in [0, 1]$ (probability)

Task B - Magnitude (Conditional)

Definition of conditional: Predict magnitude only for cases where the event occurred(positive).

- Option B1 (Regression): $y\_mag(c, t)$ = max magnitude within horizon (or next event magnitude) [Need to choose a definition]

- Option B2 (Interval Classification): bins = {[5.0, 6.0), [6.0,7.0), [7.0, +)}, etc.

- Output: $p\_occ(c, t) \in [0, 1]$ (probability)

Location Handling

- In this project, location prediction is defined not as continuous one-point regression, but rather as cell-level probability map and top-K cells presented as location candidates.

- Reason: Evauable in academic and practical settings, directly connected to dashboard, and advantageous for expressing uncertainty.

Spatial/Temporal Resolution (Location and time unit)

1. Spatial Unit (Grid and local unit)
- Unit Type: Grid / Admin region / Radius-based

- Grid Resolution: 0.5° x 0.5° (Draft, candidate: 1.0° x 1.0°)

- Coverage Boundary: global (Draft, candidates: specific country or region)

- Coordinate Reference: WGS84 (lat, lon), if needed, use projected coordinates.

2. Temporal Unit (Time unit)
- Base time step: 1 day (Draft)

- Horizon T: 7 days (Draft)

- Timezone: unify UTC standard (if raw data is UTC)

Event Definition (Definition of event and earthquake)
- Magnitude threshold M_th: M ≥ 5.0 (Draft, candidates: 4.0, 4.5)

- Depth constraints: [TODO] (not exist or limited)

- Duplicate handling: [TODO] (Rules for handling duplicate events and events occurring at the same time)

- Aftershock and declustering policy: separating for research experiments (use/not use, separable for research experiments)

Label Generation

1. Occurrence Label y_occ(c, t)

- Input: earthquake catalog events with (time, lat, long, magnitude, depth, etc.)

- Procedure:

  1) Iterate over time step t
  2) Search for events belonging to cell c within horizon (t, t + T]
  3) t_occ=1 if the max magnitude ≥ M_th else 0

2. Magnitude Label y_mag(c, t)

- Only for samples where y_occ=1

- Select definition

  (a) Max magnitude within horizon
  (b) First event magnitude within horizon
  (c) Expected magnitude (statistics)
- Choose: expected magnitude (Draft)

- Decision gate: If it is possible to realize the practical implementation, the aim of this project can be adjusted to predict the whole flow of the earthquake which includes not only maximum magnitude, but also the plenty of aftershocks.

Feature Timing Rules (Leakage)

- Features must be computed using information available at or before time t.

- Any source updated after event occurrence must be aligned to 'as-of t' snapshot.

Output Schema

1. Map Layer

- Cell_id, cell_polygon/centroid

- p_occ(c, t)

- risk_level(c, t) (color step: defined in the policy document)

2. Top-K candidates

- Ranked list of (cell_id, p_occ, optional p_mag_bin / E[M|occ], confidence/calibration info)

Success Criteria

- Research: AUPRC + [TODO]% improvement over baseline in the walk-forward test, calibration improvement

- Product: Dashboard update cycle 60s, response time [TODO], notification overload limit policy [TODO]

Open Decisions

- Grid resolution: 0.5° x 0.5°

- Horizon T: 7 days

- M_th: 5.0 or higher

- y_mag definition: expected magnitude

- Declustering usage: separating for research experiments

- Region coverage: global

Change Log

- V1.0: initial draft