Geothermal Energy
Final Sprint Review - 28 November, 2025

Thierry Fotabong, Muhammad Ammad, Laiba Tahir, Tanjim Hossain, Berhe Kiflom, Alain Patrick

Table of contents

- Research questions
- Last sprint backlog
- New insights and findings/results
- (Demo of dashboard and show status github site)
- High Level planning: on track or not?
- Risks and challenges
- Team retrospective
- Questions and answers (Q&A)

Research questions

For previous sprint,

- To develop a predictive traffic light system that forecasts seismic risk hours to days ahead, enabling proactive operational adjustments to prevent damaging earthquakes while maximizing geothermal energy production efficiency.

For this sprint,

To maintain the same predictive traffic light system for seismic risk forecasting, focusing primarily on:

- Validating model outcomes
- Refining prediction accuracy
- Documenting results and methodology

Last Sprint Backlog

Backlog item	Status	Comments
Explore data sets	Done	Successfully analyzed seismic events and operational metrics datasets
Clean the data	Done	Handled missing values, null values, and timestamp inconsistencies
Model Selection	In Progress	

Preprocessing Operational Data

Dataset: 695,275 rows × 25 columns

Missing Values: Empty cells in operational data filled using linear interpolation

Merging & Seismic Data Interpolation

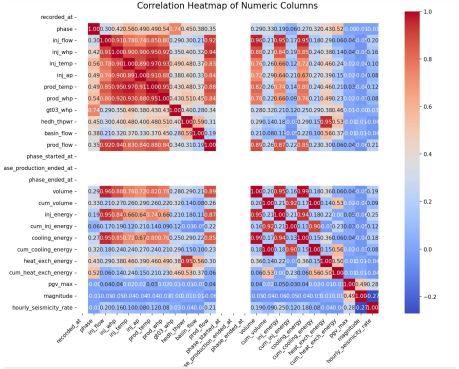
Merge: Combine cleaned operational data with seismic columns

(magnitude, pgv_max, hourly_seismicity_rate)

Feature Selection & Modeling Prep

Feature Selection: Only columns relevant to prediction (operational + seismic)

- inj_flow
- inj_whp
- inj_temp
- inj_ap
- prod_temp
- prod_whp
- gt03_whp
- basin_flow
- pgv_max
- hourly_seismicity_rate
- volume



Model Training

Model Architecture:

- LSTM (Sequential)
- Input Features: 11 operational variables
- Output: Seismic magnitude (continuous regression)
- Sequence Length: 12 time steps (~1 hour lookback)
- Prediction Horizon: 1 week ahead

Handling Data Imbalance:

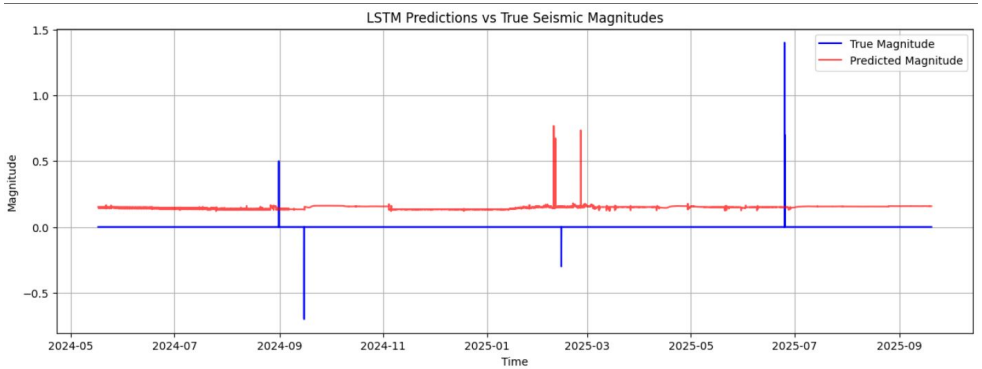
- Rare seismic events vs abundant non-events
- Used `sample_weight` during training to give **higher weight to seismic events**.

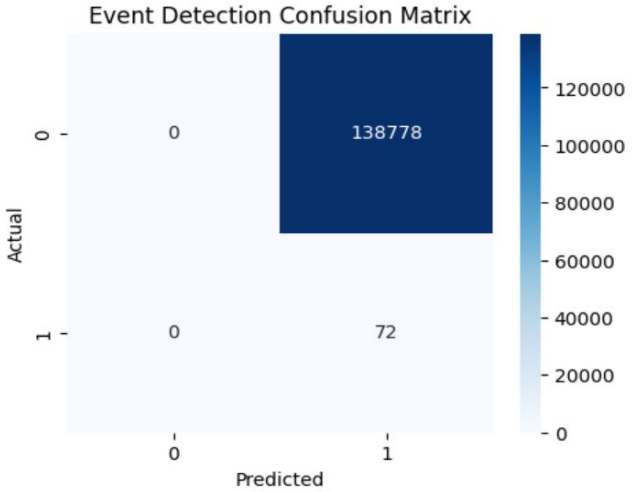
Model Results

Model Results

True Positives (TP = 72): Only 72 actual seismic events were correctly identified.

False Positives (FP = 138,778): Over 138K non-events were falsely flagged as events (massive false alarm rate).





Demo (of dashboard and status github site) - shared separately

High-level planning: on track or not?

Week	Task	Description	On track or not
1	Preprocessing	Handle missing values & timestamps; merge datasets; create lagged & aggregated features	Completed
2	Exploratory Analysis	Descriptive statistics; visualize trends; check collinearity & correlations	Completed
3-4	Baseline Modeling	Temporal train-test split; train Logistic Regression & Random Forest; map outputs to traffic light system	Completed
5-6	Advanced Modeling	Train LSTM/GRU; feature importance & sensitivity analysis; evaluate early-warning capability	Completed
7	Reporting & Visualization	Dashboards (forecasts, confusion matrices, feature importance); final report & recommendations	In progress

Risks and challenges

Risk	Impact	Severity	Owner	Status
Data Cleaning and Merging	*Potential delays in analysis if inconsistencies or missing values remain in the dataset.*	Low	Team	Resolved
Data Balancing	*Imbalanced classes can lead to biased model predictions and reduced accuracy for rare events.*	High	Team	
Model Selection	Incorrect or suboptimal model choice could limit predictive performance and interpretability.	Medium	Team	
Frontend	Delays or integration issues in displaying results and visualizations through the user interface.	Medium	Team	
Model Deployment	Technical challenges in hosting or integrating the predictive model with the operational platform.	Medium	Team	

Team Retrospective

What went well

- Handled missing values, null values, and timestamp alignment
- Visualized distribution and time series trends
- Analyzed relationships between operational and seismic features
- Team worked effectively on assigned components

Challenges faced

- Better balance of workload across weeks required
- LSTM and other Machine Learning architectures requires more investigation time than initially planned

Thank you!