Review - 1

Earthquake Magnitude Prediction from Precursor Signals

August 11, 2025

Course:

Foundations of Data Science (BCSE206L)

Guided by:

Prof. Nalini N

Submitted By:

Ashwin Anand (22BCE2721)

Yagnit Mahajan (22BCE2588)

INDEX

PROBLEM STATEMENT:-

Earthquakes cause significant loss of life and property. Accurate early detection of earthquake magnitude based on precursor signals can improve disaster preparedness. This project uses real-world earthquake datasets to analyze precursor–magnitude relationships and build detection models.

ABSTRACT:-

This project focuses on early detection of earthquake magnitudes by analyzing precursor signals using a real-world dataset of over 9,000 earthquake events. The approach involves extensive data cleaning, exploratory data analysis, and predictive modeling to understand the relationship between precursor signals and earthquake magnitude. Python is used for data processing and analysis, SQL helps in data aggregation and querying, and Tableau is employed for visualizing spatial patterns and trends. This multidisciplinary workflow aims to enhance disaster preparedness by providing actionable insights from seismic data through effective data analysis and machine learning techniques. This project aligns well with SDG 3 – Good Health and Well Being.

OBJECTIVES:-

• Clean and preprocess a dataset of 9000+ records to ensure data quality and consistency.

• Perform exploratory data analysis to uncover key trends and patterns.

• Use SQL for data aggregation and manipulation to support analysis.

• Create interactive visualizations in Tableau to communicate insights effectively.

• Predict approximate earthquake magnitude based on precursor signals.

LITERATURE SURVEY:-

This section summarizes the key findings and limitations of 10 research papers focused on earthquake magnitude detection using machine learning approaches. Based on this review, the critical research gaps are identified to justify the need for the current study.

RESEARCH PAPERS:-

S. No.	Paper	Author	Journals and Publication and Year	Methodology	Key Findings	Limitations
1.	Improving earthquake prediction accuracy in Los Angeles with machine learning	Cemil Emre Yavas, Lei Chen, Christopher Kadlec, Yiming Ji	Scientific Reports, 2024	Comprehensive feature matrix synthesis using sixteen algorithms to optimize seismic pattern classification.	Random Forest achieved nearly 98% accuracy predicting maximum earthquake magnitude within 30 days.	Model trained on Los Angeles data; performance on other seismic regions remains unverified.

2.	Earthquake Prediction using Hybrid Machine Learning Techniques	Mustafa Abdul Salam, Lobna Ibrahim, Diaa Salama Abdelminaam	International Journal of Advanced Computer Science and Applications (IJACSA), 2021	FPA optimization integrates Extreme Learning Machine with Support Vector Machine parameters.	Hybrid FPA-LS-SVM outperformed other models, improving 15-day magnitude prediction accuracy significantly.	Study limited to Southern California datasets; feature selection optimization needed for better accuracy.
3.	Long-Term Forecasting of Strong Earthquakes in North America, South America, Japan, Southern China and Northern India with Machine Learning	Victor Manuel Velasco Herrera, Eduardo Antonio Rossello, Maria Julia Orgeira, Lucas Arioni, Willie Soon, Graciela Velasco, Laura Rosique-de la Cruz, Emmanuel Zúñiga, Carlos Vera	Frontiers in Earth Science, 2022	Wavelet transforms decompose seismic signals for Bayesian probabilistic temporal pattern recognition.	Bayesian ML identified cycles of seismicity, enabling probabilistic rather than precise earthquake forecasting.	Does not support accurate short-term forecasts; physical–statistical model integration remains unfinished.
4.	Universal Neural Networks		Nature Comm	Generalized earthquake data trains	Neural networks predicted earthquake location and	Requires real-world deployment;

	for Real-Time Earthquake Early Warning	Xiong Zhang, Miao Zhang	unication, 2024	universal networks for cross-regional P-wave magnitude estimation.	magnitude within 4 seconds of P-wave detection.	latency reduction essential for high-density urban safety applications.
5.	Local Magnitude Estimation via an Attention-Based Machine Learning Model	Zhang, Aitaro Kato, Huiyu Zhu, Wei Wang	Seismological Research Letters, 2024	Convolutional layers with attention mechanisms process STEAD waveform features automatically.	Attention-based model rapidly estimated local magnitudes from real-time seismic waveforms for early warning.	Validation limited to certain regions; integration with national-scale seismic networks still pending.
6.	Magnitude Estimation System based on Convolutional Neural Networks (MESCNN)	Ji'an Liao, Siran Yang, Yanwei Wang	Results in Engineering, 2025	Multi-layer CNN architecture extracts hierarchical features from raw seismic time-series data.	CNN-based system surpassed traditional algorithms for real-time earthquake magnitude estimation in U.S. networks.	Detection accuracy drops for small events.
7.	Analysis of earthquake detection using deep learning	Sebastián Gamboa-Chacón, Esteban Meneses, Esteban J. Chaves	Applications in Advanced Mathematical Concepts and Disco	EQTransformer model processes three-component waveforms for automated P and S detection.	Deep learning (EQTransformer) improved detection rates and reliability compared to classical detection methods.	Slow in rejecting false positives; needs transfer learning for diverse geographical regions.

			veries (AAMCD), 2025			
8.	Feature-Based Magnitude Estimates for Small Earthquakes in Yellowstone, USA	Alysha D. Armstrong, Ben Baker, Keith D. Koper	Bulletin of the Seismological Society of America, 2024	Automated pipeline extracts waveform characteristics for rapid small-earthquake magnitude estimation systems.	Feature-based ML pipeline rapidly estimated magnitudes for thousands of small earthquakes in Yellowstone.	Model tuned for Yellowstone; adaptation for tectonically diverse areas remains challenging.
9.	A New Algorithm for Earthquake Prediction Using Machine Learning	Nada Badr Jarah, Abbas Hanon Hassin Alasadi, Kadhim Mahdi Hashim	Journal of Computer Science, 2024	Ensemble boosting combines multiple base learners with weighted voting for prediction.	Ensemble machine learning significantly improved prediction accuracy using boosting over traditional classifiers.	Real-time deployment and integration with USGS hazard tools are still undeveloped.
10.	Machine learning–powered earthquake early warning system	Vijaya Saraswathi R	Journal of Engineering Research and Reports, 2024	Multi-algorithm framework integrates real-time data streams for automated seismic event classification.	Random Forest-based early warning achieved 96% accuracy, reducing false positives in simulations.	Lacks validation in nationwide operational scenarios and during multiple simultaneous seismic events.

RESEARCH GAPS:-

Key research gaps emerging from this literature review are:

1. **Regional Generalizability:** Many models are developed using data from specific geographic areas, limiting their applicability across diverse seismic regions.

2. **Physical model integration:** ML rarely combined with physical/geophysical models for better forecasts.

3. **Feature selection & data diversity:** Limited use of diverse, optimized features reduces model robustness.

4. **Operational deployment:** Few models tested for real-time or multi-event scenarios in warning systems.

5. **Model Transferability:** The adaptation and transfer of models across regions with differing tectonic characteristics and data sparsity remain underexplored.

6. **Benchmarking Against Established Methods:** Several studies lack rigorous comparison with traditional statistical seismology models, limiting assessment of practical utility.

DATASET:-

The dataset consists of over 9,000 recorded earthquake events from across various states in the USA, spanning from December 27, 2023, to January 26, 2024. It includes detailed information such as the time and location (latitude and longitude), depth, magnitude, and other relevant seismic parameters, enabling comprehensive analysis of earthquake characteristics and patterns.

In this dataset, precursor signals refer to measurable seismic parameters that may indicate the likelihood or intensity of an impending earthquake. These include features like the number of reporting stations (nst), measurement gaps (gap), minimum distances to stations (dmin), and error margins in location and magnitude estimations. Monitoring changes or patterns in these signals helps in detecting early signs of seismic activity before a major event occurs.

- **Raw Dataset**

time	latitude	longitude	depth	mag	magType	nst	gap	dmin	rms	net	id	updated	place	type	horizontalError	depthError	magError	magNst	status	locationSource	magSource
2024-01-26T04:52:42.967Z	31.604	-104.213	4.4198	1.7	ml	18	69	0.1	0.5	tx	tx2024bubg	2024-01-26T05:08:27.774Z	51 km NW of Toyah, Texas	earthquake	0	1.292058988	0.1	13	automatic	tx	tx
2024-01-26T04:42:50.711Z	64.501	-146.9058	4.2	1.4	ml				0.75	ak	ak024172lv8v	2024-01-26T05:01:12.516Z	2 km S of Salcha, Alaska	earthquake		0.2			automatic	ak	ak
2024-01-26T04:32:51.471Z	63.529	-147.5543	13.1	1.6	ml				0.62	ak	ak024172jq57	2024-01-26T04:34:54.160Z	71 km ESE of Denali Park, Alaska	earthquake		0.3			automatic	ak	ak
2024-01-26T04:29:01.180Z	38.833168	-122.7971649	1.73	0.4	md	9	65	0.007468	0.02	nc	nc73994061	2024-01-26T04:46:12.828Z	6 km W of Cobb, CA	earthquake	0.34	0.97	0.31	10	automatic	nc	nc
2024-01-26T04:23:14.444Z	63.5462	-150.9719	0	1.2	ml				0.8	ak	ak024172hpuc	2024-01-26T04:25:43.831Z	37 km E of Denali National Park, Alaska	earthquake		0.4			automatic	ak	ak

- **Cleaned Dataset**

latitude	longitude	depth	mag	magType	gap	rms	net	id	updated	place	type	horizontalError	depthError	magError	status	locationSource	magSource	date	time_only
31.604	-104.213	4.4198	1.7	ml	69	0.5	tx	tx2024bubg	2024-01-26T05:08:27.774Z	51 km NW of Toyah, Texas	earthquake	0	1.292058988	0.1	automatic	tx	tx	1/26/2024	52:43.0
64.501	-146.9058	4.2	1.4	ml	89.91	0.75	ak	ak024172lv8v	2024-01-26T05:01:12.516Z	2 km S of Salcha, Alaska	earthquake	0.46	0.2	0.173	automatic	ak	ak	1/26/2024	42:50.7
63.529	-147.5543	13.1	1.6	ml	89.91	0.62	ak	ak024172jq57	2024-01-26T04:34:54.160Z	71 km ESE of Denali Park, Alaska	earthquake	0.46	0.3	0.173	automatic	ak	ak	1/26/2024	32:51.5
38.833168	-122.7971649	1.73	0.4	md	65	0.02	nc	nc73994061	2024-01-26T04:46:12.828Z	6 km W of Cobb, CA	earthquake	0.34	0.97	0.31	automatic	nc	nc	1/26/2024	29:01.2
63.5462	-150.9719	0	1.2	ml	89.91	0.8	ak	ak024172hpuc	2024-01-26T04:25:43.831Z	37 km E of Denali National Park, Alaska	earthquake	0.46	0.4	0.173	automatic	ak	ak	1/26/2024	23:14.4