EarthGPT: Earthquake Prediction with GPT2

Deja Scott[1], **Kolipaka Pranavi**[1]
[1]University of South Carolina
ds17@email.sc.edu, kolipakv@email.sc.edu

Abstract

The design of earthquake-resistant structures is a multidisciplinary field where advancements in Machine Learning (ML) and Artificial Intelligence (AI) are driving significant innovations. Peak Ground Acceleration (PGA), commonly referred to as earthquake magnitude, is a critical parameter for designing structures capable of withstanding seismic forces. This study introduces a regressive GPT-2 model specifically designed to predict earthquake magnitudes in California, United States, based on historical seismic data. The model leverages the natural language processing capabilities of GPT-2, repurposed for numerical regression tasks. Our results demonstrate the efficacy of the regressive GPT-2 model, achieving a Mean Absolute Error (MAE) of 0.0228, Mean Squared Error (MSE) of 0.0016, Root Mean Squared Error (RMSE) of 0.0399, Mean Absolute Percentage Error (MAPE) of 0.0061, and a coefficient of determination (R^2) of 0.9917. These metrics highlight the model's exceptional predictive accuracy and potential as a robust tool for seismic risk assessment. By providing precise magnitude predictions, this work contributes to the development of safer, earthquake-resilient infrastructure and highlights the potential of AI-driven approaches in geophysics.

Introduction

An earthquake is a natural disaster characterized by the sudden shaking of the Earth's surface, caused by the abrupt release of energy in the Earth's lithosphere. This energy release typically results from stress accumulation along geological faults or volcanic activity. Earthquakes can vary significantly in intensity, duration, and the extent of the affected area, posing a substantial threat to human life, infrastructure, and economies worldwide.

How Earthquakes Are Traditionally Measured? The measurement of earthquakes has evolved significantly over time. Traditionally, scientists and researchers utilized seismometers (or seismographs) to detect and record seismic waves, facilitating the study of these natural phenomena. These instruments capture ground motion in real-time, providing critical data for understanding seismic activity. The Richter Scale, developed in 1935 by Charles F. Richter, was one of the earliest tools for quantifying earthquakes. Using a logarithmic scale, it measures the amplitude of seismic waves, with each whole-number increase corresponding to a tenfold rise in measured amplitude and approximately 31.6 times more energy released. Modern seismology, however, relies more on the Moment Magnitude Scale (Mw), which provides a more accurate representation of the total energy released during an earthquake, especially for large or distant events. In addition, the Modified Mercalli Intensity Scale is employed to assess the effects of an earthquake based on observed damage and human experiences. Unlike the Richter and Mw scales, this method is inherently subjective, focusing on qualitative observations rather than precise numerical measurements.

Why Magnitude Prediction? Magnitude prediction is a cornerstone of earthquake preparedness and disaster mitigation. For populations living near fault zones, it is essential to implement measures for minimizing loss of life and property. Accurate predictions enable the development of evacuation plans, the shutdown of critical infrastructure such as nuclear reactors and gas pipelines, and the timely mobilization of emergency response teams. From an engineering perspective, understanding the potential magnitude of an earthquake helps design buildings and infrastructure that can endure anticipated seismic forces, particularly in high-risk regions. Economically, earthquakes can cause widespread disruption by damaging property, halting business operations, and impacting industries like tourism. Effective magnitude prediction reduces these economic shocks by informing preventive measures and emergency planning. Even a few seconds of early warning can be life-saving, providing crucial time for individuals to take protective actions or move to safer locations. Moreover, large-scale earthquakes often trigger secondary disasters such as tsunamis or landslides. Magnitude prediction plays a vital role in initiating global alerts for such catastrophic events, enhancing overall disaster response strategies.

Despite advancements in seismology, accurately predicting the timing and magnitude of earthquakes remains one of the greatest challenges in geophysics. This difficulty arises from the complex and dynamic nature of Earth's tectonic systems, which involve a multitude of interacting variables.

How can AI help? Advancements in computing have brought Artificial Intelligence (AI) and Machine Learning

(ML) to the forefront of earthquake prediction and measurement. These technologies offer robust tools for designing efficient early warning systems, with the potential to save millions of lives. The integration of AI and ML provides several advantages. These systems can analyze vast amounts of historical data, uncovering patterns and anomalies that traditional methods might overlook. AI-based models can also be customized to meet the specific needs of a region, offering flexibility for future enhancements and scalability for large-scale applications. Many regions prone to seismic activity lack adequate instruments to measure and monitor earthquakes comprehensively. AI presents an ideal solution, as it can compensate for these gaps by using data-driven approaches tailored to local conditions. Additionally, as AI models continuously improve with new data, their predictive accuracy and reliability are expected to increase over time. In light of these benefits, we propose a regressive GPT-2 (Radford et al. 2019) model designed to predict earthquake magnitudes. This model leverages prior event features such as depth, date, time, and location, providing a scalable and adaptive solution to one of geophysics' most persistent challenges.

Related Works

Scientists and researchers previously used traditional ML methods to predict earthquake magnitudes like Linear Regression, Support Vector Regression (SVR) and Random Forest Regression models. Linear Regression seeks to establish a linear relationship between the input features and the target variable. However, while this model is the simplest it underperforms, achieving an R^2 score of 0.000067, which indicates that it explains a negligible amount of variance in the dataset. Support Vector Regression (SVR), attempts to identify a hyperplane that best fits the data in a high-dimensional space. However, it yields the poorest results among the three models, with a negative R^2 score of -38.618141, suggesting that its predictive capability is even less effective than that of a horizontal line. Conversely, Random Forest Regression, employs an ensemble of decision trees to generate predictions. This model exhibits the best performance of the three regression models, attaining an R^2 score of 0.212287, explaining approximately 21.23% of the variance in the target variable. Furthermore, each model is trained on the earthquake dataset and evaluated using Mean Squared Error (MSE) and R-squared (R^2) metrics. The findings reveal that the Random Forest Regressor outperforms the other two models, achieving the lowest Mean Squared Error (MSE) of 0.14336 and the highest R^2 score of 0.212287. While these results are superior, the Random Forest Regressor requires extensive feature engineering and domain expertise to identify the relevant patterns within the data.

Recently, researchers have begun incorporating deep learning methods, such as Deep Neural Networks (DNN) (Jipan et al. 2018) and the Earthquake Early Warning Network (EEWNet) (Wang et al. 2023), to enhance earthquake magnitude prediction. While these models demonstrate strong performance and improve on traditional methods, they often require significant feature engineering and careful preprocessing of seismic data. Additionally, their fixed architectures can limit flexibility, making it challenging to adapt to diverse datasets or evolving seismic patterns without substantial retraining and reconfiguration. These limitations highlight the need for more adaptable and scalable approaches in this domain. (Laurenti et al. 2022) employed models tailored for time-series data, such as Long Short-Term Memory (LSTM), Temporal Convolution Networks (TCN), and Transformer models, to forecast stress levels at fault zones. While these models are effective for capturing temporal patterns and trends, they often require substantial computational resources due to their complex architectures. Additionally, they may struggle with fully grasping the global context of the data and long-term dependencies, limiting their ability to identify broader trends or nuanced patterns within seismic datasets.

To address these limitations, we propose a regressive GPT-2 model fine-tuned on earthquake data for predicting earthquake magnitudes. Unlike Random Forest, GPT-2 is capable of handling complex data structures and uncovering intricate patterns and trends that may be hidden in the data, offering a more efficient approach to prediction.

Methodology

Dataset To train and evaluate our approach, we employed the SOCR Earthquake Dataset (SOCR), which contains data on earthquakes with a magnitude of 3.0 or greater that occurred in California, United States, between July 1st 1966 and December 28th 2007. This dataset consists of 18,030 earthquake events, with each entry representing an individual seismic occurrence. It serves multiple analytical purposes, such as investigating patterns and trends in seismic activity over time and forecasting future earthquakes. The dataset is organized into 14 variables, each providing critical information about the events. These variables include the date and time of the earthquake in Coordinated Universal Time (UTC), the geographic coordinates (latitude and longitude) of the epicenter—the point on the Earth's surface directly above the earthquake's origin—and the depth of the earthquake in kilometers. Additional features include the earthquake's magnitude on the Richter scale, the source (SRC), the number of stations used for the solution (NST), the distance from the nearest station to the epicenter (CLOSE), the root mean square (RMS) residual of the solution, and the azimuthal gap (GAP), which ranges from 0° to 360°. This comprehensive dataset provides a valuable foundation for analyzing seismic events.

Data Analysis This dataset comprises 18,030 earthquake records with 14 variables, offering a detailed overview of seismic activity in the region from 1st July 1966 to 28th December 2007. Geographically, it spans latitudes 31.73°N to 45.56°N and longitudes 127.51°W to 112.11°W. Key earthquake characteristics include depths ranging from 0 to 121.31 km (mean: 8.88 km, median: 7.07 km) and magnitudes from 3.0 to 7.39 (mean: 3.43, median: 3.30), with magnitude types primarily comprising Md, alongside others such as ML, Mw, and Mx. Data quality indicators include the number of stations per event (4 to 327, mean: 31.94, median:

Model	MAE ↓	MSE ↓	RMSE ↓	MAPE ↓	R2 ↑
Linear Regression	—	0.182	—	—	0.000067
SVM	—	7.210	—	—	-38.618
Random Forest	—	0.143	—	—	0.212
KAN	0.303	0.189	0.434	0.083	0.083
GRU	0.305	0.175	0.419	0.085	0.007
LSTM	0.326	0.203	0.449	0.091	0.005
Transformer	0.299	0.218	0.467	0.079	0.250
TimeMixer (SOTA)	0.301	0.212	0.461	0.081	0.219
PacthTST	0.298	0.202	0.449	0.080	0.159
EQGPT (Ours)	0.0228	0.0016	0.0399	0.0061	0.9917

Figure 1: Comparative Analysis - Our approach was benchmarked against various methods, including traditional and state-of-the-art (SOTA) models. EQGPT consistently outperformed all others across all evaluation metrics.

26), gap values ranging from 12° to 354° (mean: 147.56°, median: 127°), and RMS values from 0 to 45.4 (mean: 0.14, median: 0.08). The dataset is sourced exclusively from the Northern California Seismic Network (NCSN), with each event uniquely identified by an EventID. This comprehensive dataset serves as a robust resource for analyzing seismic activity patterns and characteristics across a 41-year period.

Data Cleaning The dataset was divided into training and testing subsets, with 80% of the dataset allocated for training and 20% reserved for testing, ensuring the temporal sequence of the data remained intact. The resulting subsets, labeled as "train_earthquake_data" and "test_earthquake_data," were saved as CSV files while excluding row indices. To prepare the data for the proposed methodology, six essential features—Date (YYYY/MM/DD), Time (UTC), Latitude (°), Longitude (°), Depth (km), and Magnitude (ergs)—were combined to generate structured input prompts for both training and evaluation.

EarthquakeGPT We introduce EarthquakeGPT (EQGPT), a regression model based on the GPT-2 architecture, tailored for earthquake magnitude prediction. The model extends GPT-2 with a custom regression head composed of dropout layers for enhanced generalisation and a multi-layer perceptron (MLP) to transform GPT-2's final hidden states into numerical predictions. A Mean Squared Error (MSE) loss function is employed to optimise magnitude predictions. The model is trained and evaluated using the SOCR Earthquake dataset, split into temporally consistent training and testing subsets. To ensure robust evaluation, a 5-fold cross-validation approach is implemented. Each fold is independently trained and validated, calculating metrics such as Mean Absolute Error (MAE), MSE, Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R2 score on the validation data. The fold yielding the lowest RMSE is selected as the best-performing model. The training pipeline leverages the Hugging Face Trainer API, allowing for hyperparameter tuning (e.g., learning rate, batch size, and epochs) via an external YAML configuration. Tokenisation and data formatting are managed using a custom dataset class. The model is finetuned during training, with early stopping applied based on validation loss to prevent overfitting. For the final evaluation, the best-performing model is tested on a hold-out dataset. Predicted earthquake magnitudes are compared to actual values, and performance metrics are recalculated. EarthquakeGPT effectively combines natural language processing capabilities with regression-based outputs, capturing complex seismic patterns and addressing the limitations of conventional machine learning methods. The model demonstrates adaptability, scalability, and an enhanced ability to learn from diverse temporal and spatial data representations.

Evaluation

Experimental Environment We benchmarked our model against both traditional and state-of-the-art models, including Linear Regression, Transformer (Zhou et al. 2021), PatchTST (Nie et al. 2023), Kolmogorov-Arnold Network (KAN) (Liu et al. 2024), TimeMixer (Wang et al. 2024), Gated Recurrent Unit (GRU) (Cho et al. 2014), Support Vector Machines (SVM), Random Forest and Long Short-Term Memory (LSTM) networks, to provide a thorough evaluation of its performance. To assess the effectiveness of our approach and the comparative methods, we employed several metrics including: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2). MAE and RMSE offer insights into the average accuracy and the magnitude of prediction deviations, respectively. MSE serves to penalise errors in proportion to their size, assigning greater penalties to larger errors. MAPE quantifies the percentage deviation in our model's predictions, while R-squared evaluates the extent to which the model accounts for the variability in the data. Our analysis was conducted using the testing portion of the SOCR dataset, with all comparative methods considering the same six features utilised in training EQGPT, and combining the Date (YYYY/MM/DD) and Time (UTC) columns. All models were trained using five-fold cross-validation on an MSI computer equipped with 49 GB of RAM, a 16-GB AMD Radeon RX 6700 XT GPU, and a Ryzen 5 CPU.

Comparative Analysis We present a comparative analysis of the performance of various models, including both traditional and state-of-the-art approaches, as summarised in Figure 1. The models were evaluated across several key metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2).

Among the traditional models, Linear Regression demonstrated the lowest R^2 value of 0.000067, suggesting minimal effectiveness in capturing the variability of the data. Support Vector Machine (SVM) performed poorly, with an extremely negative R^2 value of -38.618, indicating a poor fit to the data. The Random Forest model showed a modest improvement, with an R^2 value of 0.212, indicating some ability to model the data but still lagging behind more advanced models.

In the realm of deep learning models, KAN achieved an MAE of 0.303, MSE of 0.189, and an R^2 of 0.083, reflecting a relatively average performance. GRU and LSTM both showed similar results, with GRU slightly outperforming LSTM in terms of MAE (0.305 vs. 0.326) and MSE (0.175 vs. 0.203), but both models still exhibited lower performance in comparison to more advanced models in terms of R^2, which were 0.007 and 0.005, respectively. The Transformer model performed similarly to the GRU and LSTM, with an MAE of 0.299 and R^2 of 0.250, indicating a somewhat better fit than the other deep learning models, but still far from optimal.

Among the state-of-the-art models, TimeMixer and PatchTST demonstrated competitive results. TimeMixer, with an MAE of 0.301, MSE of 0.212, and R^2 of 0.219, showed a slightly better performance than PatchTST, which had an MAE of 0.298, MSE of 0.202, and R^2 of 0.159. Despite this, these models still fell short of the performance exhibited by our proposed model, EQGPT.

The EQGPT model outperformed all other methods across all metrics. With an exceptionally low MAE of 0.0228, MSE of 0.0016, and RMSE of 0.0399, EQGPT demonstrated superior prediction accuracy. Furthermore, it achieved an outstanding MAPE of 0.0061 and a significantly high R^2 value of 0.9917, indicating that it captures almost all the variability in the data. This highlights the effectiveness of EQGPT in comparison to both traditional machine learning models and other state-of-the-art models.

Conclusion and Future Works

In conclusion, this study effectively demonstrates the application of a regressive GPT-2 model, EarthquakeGPT, for predicting earthquake magnitudes in California, utilizing historical seismic data. The model achieved notable predictive performance, with an R^2 value of 0.9917 and a Mean Absolute Error (MAE) of 0.0228, emphasizing its potential as a reliable tool for seismic risk assessment. This research contributes to the advancement of earthquake-resistant infrastructure and highlights the increasing relevance of AI-driven methodologies in the field of geophysics.

Future research is essential for further refining and expanding the predictive capabilities of the EarthquakeGPT model. A key area for exploration is the evaluation of the model against additional, previously unseen datasets to evaluate its ability to generalize across diverse seismic conditions and maintain high accuracy despite variations in input variables. Furthermore, the integration of multimodal data, such as fault line characteristics and other geological features, could enhance the model's predictive power by enabling it to synthesize a broader array of data sources, thereby improving the comprehensiveness and reliability of its forecasts. Lastly, conducting real-time prediction assessments is imperative to evaluate the model's operational feasibility, particularly concerning latency and the speed at which it generates accurate results. Such experiments are crucial for ensuring the model's practical applicability in seismic risk management, ultimately advancing its potential utility in real-world geophysical applications.

References

Cho, K.; van Merrienboer, B.; Bahdanau, D.; and Bengio, Y. 2014. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. arXiv:1409.1259.

Jipan, H.; Wang, X.; Zhao, Y.; Chen, X.; and Han, X. 2018. LARGE EARTHQUAKE MAGNITUDE PREDICTION IN TAIWAN BASED ON DEEP LEARNING NEURAL NETWORK.

Laurenti, L.; Tinti, E.; Galasso, F.; Franco, L.; and Marone, C. 2022. Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress. *Earth and Planetary Science Letters*, 598: 117825.

Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T. Y.; and Tegmark, M. 2024. KAN: Kolmogorov-Arnold Networks. arXiv:2404.19756.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. arXiv:2211.14730.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language Models are Unsupervised Multitask Learners.

(SOCR), U. S. O. C. R. 2016. Earthquakes Over 3.0 Dataset. Accessed: 2024-12-09.

Wang, S.; Wu, H.; Shi, X.; Hu, T.; Luo, H.; Ma, L.; Zhang, J. Y.; and Zhou, J. 2024. TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting. arXiv:2405.14616.

Wang, Y.; Li, X.; Wang, Z.; and Liu, J. 2023. Deep learning for magnitude prediction in earthquake early warning. *Gondwana Research*, 123: 164–173. Data driven models.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. arXiv:2012.07436.