# EARTHQUAKE FORECASTING AND RISK MODELING USING MACHINE LEARNING APPROACHES BASED ON LOCATION IN JAPAN

**Ruslan Umirbek ugli Berdiev**
Master's student at Urgench Branch of Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
bruslon2707@gmail.com

## ABSTRACT

This study analyzes earthquake risk in Japan based on the spatial and temporal distribution of earthquakes from 2001 to 2018. The earthquake dataset containing information on over 12,000 earthquakes in this 18-year time period is used to examine patterns in earthquake frequency, magnitude, and location. Spatial analysis is performed to identify areas of highest seismic activity and risk. The findings reveal specific regions of Japan that are particularly prone to earthquakes of varying magnitudes. The relationships between earthquake characteristics and the geophysical features of the Japanese archipelago are explored. This study provides an in-depth assessment of seismic hazard and risk at the national scale based on a comprehensive dataset of recent earthquakes in Japan.

**Keywords:** Earthquake risk, Japan, Earthquake dataset, frequency, magnitude, spatial analysis, activity, hazard, seismic risk, Kaggle.

## INTRODUCTION

Japan is one of the most seismically active countries in the world, with a long history of devastating earthquakes. The risk of earthquakes varies widely across the country, depending on factors such as geological conditions, proximity to active faults, and local building codes. Understanding the risk of earthquakes in different regions of Japan is essential for effective disaster preparedness and risk management. This thesis aims to investigate the relationship between earthquake risk and location in Japan, with a focus on identifying the factors that contribute to variation in earthquake risk across the country. The thesis begins with a review of the current state-of-the-art in earthquake risk assessment and an analysis of the historical earthquake data in Japan. It then presents a comprehensive evaluation of various factors that contribute to earthquake risk, including geological and geophysical conditions, building codes and construction practices, and population density and demographics. The thesis also proposes new methods for spatial analysis and modeling of earthquake risk, with the goal of improving the accuracy and precision of earthquake risk assessments in Japan. By providing a comprehensive analysis of earthquake risk based on location in Japan, this thesis aims to contribute to the development of more effective disaster preparedness and risk management strategies, and to promote the safety and well-being of the Japanese population.

This study examines the performance of machine learning models for predicting earthquakes in Japan. Section 2 reviews previous related work on earthquake forecasting and machine learning. The proposed machine learning models for earthquake prediction are detailed in Section 3, including training and evaluation procedures. Section 4 discusses the results, comparing the models' prediction accuracies. The implications of the findings and future directions for machine learning-based earthquake forecasting are also discussed. The work provides insights into the potential of machine learning methods to predict earthquake occurrence, location and magnitude.

## METHODS

This study analyzes the issues with information network systems during the 2011 Tohoku earthquake and discusses potential solutions[2]. Through efforts to reconstruct networks after the earthquake, network connectivity was found to be the most critical aspect of disaster information systems, rather than throughput or latency. In practice, satellite systems, wireless LANs, and cognitive wireless networks were effective at re-establishing network connectivity in the disaster areas. The analysis and discussion in this study provide insights into how to strengthen information network resilience and communications in the aftermath of a large-scale earthquake.

The scientific thesis analyzes the attenuation of ground motion parameters with distance from the fault for the Tokai earthquake in Japan. Strong motion data recorded during the earthquake is analyzed to examine how peak ground acceleration, peak ground velocity, Arias intensity, and response spectrum values vary with distance from the fault. The results show differences in attenuation rates for horizontal versus vertical ground motion components and for different response spectrum periods. These findings are compared with existing ground motion attenuation models for Japan earthquakes. The work provides further insights into ground motion patterns during large earthquakes that can inform seismic hazard analysis and disaster preparedness[1].

Overall, it develops an artificial neural network (ANN) model to predict earthquake magnitude in Japan. The optimal

ANN architecture is determined by evaluating different data splitting methods, network types, hidden layer sizes, learning algorithms, and performance metrics. The final ANN model uses the Levenberg-Marquardt backpropagation algorithm and a lower number of hidden neurons, achieving strong performance. The trained ANN is able to predict the magnitude of future earthquakes with minimal error. The results suggest that an ANN could serve as an effective tool for earthquake magnitude prediction, as ANNs can learn from data.[3] Overall, this study demonstrates that an ANN-based intelligent model can be used to predict earthquake magnitude.

The proposes a deep learning model to predict the occurrence of earthquakes above a specified magnitude within 10 to 50 days at a given location. The model uses recurrent neural networks and convolutional neural networks to capture temporal and spatial dependencies in the data. A model combining these architectures achieves adequate performance. By using neural networks to automatically extract features, the proposed approach avoids the need for manual feature engineering for local mid-term earthquake prediction[4].

## DISCUSSION

The dataset contains information about earthquakes that occurred in Japan between 2001 and 2018. The data includes the date and time of the earthquake, the latitude and longitude of the epicenter, the magnitude of the earthquake, the depth of the earthquake, and the number of deaths and injuries caused by the earthquake.

### Data Collection

The Japan Meteorological Agency (JMA) is the government agency responsible for monitoring and recording earthquake activity in Japan. The agency operates a nationwide network of seismometers and other earthquake monitoring equipment, and is responsible for issuing earthquake alerts and warnings to the public.The JMA maintains a comprehensive earthquake database that includes information on all earthquakes that occur in Japan, including their location, magnitude, depth, and other characteristics. The agency also uses this data to generate seismic hazard maps and other earthquake risk assessment tools, which are used by local governments and other organizations to prepare for earthquakes and other natural disasters.

| | time | latitude | longitude | depth | mag | magType | nst | gap | dmin | rms | ... | updated |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2018-11-27T14:34:20.900Z | 48.3780 | 154.9620 | 35.00 | 4.9 | mb | NaN | 92.0 | 5.044 | 0.63 | ... | 2018-11-27T16:06:33.040Z |
| 1 | 2018-11-26T23:33:50.630Z | 36.0733 | 139.7830 | 48.82 | 4.8 | mww | NaN | 113.0 | 1.359 | 1.13 | ... | 2018-11-27T16:44:22.223Z |
| 2 | 2018-11-26T13:04:02.250Z | 38.8576 | 141.8384 | 50.56 | 4.5 | mb | NaN | 145.0 | 1.286 | 0.84 | ... | 2018-11-26T23:52:21.074Z |
| 3 | 2018-11-26T05:20:16.440Z | 50.0727 | 156.1420 | 66.34 | 4.6 | mb | NaN | 128.0 | 3.191 | 0.62 | ... | 2018-11-26T08:13:58.040Z |
| 4 | 2018-11-25T09:19:05.010Z | 33.9500 | 134.4942 | 38.19 | 4.6 | mb | NaN | 104.0 | 0.558 | 0.61 | ... | 2018-11-25T23:24:52.615Z |

**Fig. 1. First 5 rows in the training set.**

This Japan earthquakes 2001 - 2018.csv dataset contains 14092 training set categorized into 22 columns: time, latitude, longitude, depth, mag, magType, nst gap, dmin, rms, net and other columns. The distribution of the number and the tail of the dataset of each column can be seen in fig. 2.

| | time | latitude | longitude | depth | mag | magType | nst | gap | dmin | rms | ... | updated |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 14087 | 2001-01-04T04:18:21.430Z | 36.981 | 138.629 | 21.4 | 5.3 | mwc | 240.0 | NaN | NaN | 0.78 | ... | 2016-11-09T21:35:38.927Z |
| 14088 | 2001-01-03T14:47:49.540Z | 43.932 | 147.813 | 33.0 | 5.9 | mwc | 362.0 | NaN | NaN | 0.82 | ... | 2016-11-09T04:09:27.754Z |
| 14089 | 2001-01-03T09:32:54.710Z | 45.301 | 149.941 | 33.0 | 4.6 | mb | 40.0 | NaN | NaN | 0.82 | ... | 2014-11-07T01:11:51.270Z |
| 14090 | 2001-01-02T20:29:49.670Z | 32.239 | 141.508 | 33.0 | 4.9 | mb | 32.0 | NaN | NaN | 0.96 | ... | 2014-11-07T01:11:51.119Z |
| 14091 | 2001-01-02T10:53:23.280Z | 37.218 | 138.431 | 33.0 | 4.9 | mb | 43.0 | NaN | NaN | 0.93 | ... | 2014-11-07T01:11:50.950Z |

**Fig. 2. Last 5 rows in the training set.**

**Preprocessing dataset**

The whole dataset contains 14092 rows and 22 columns. This means that there are 14092 instances or records in the dataset without test dataset, with each record containing information about a single earthquake that occurred in Japan between 2001 and 2018.Each record contains 22 pieces of information, or columns, about the earthquake. These columns include the date and time of the earthquake, the latitude and longitude of the epicenter, the depth of the earthquake, the type and magnitude of the earthquake, the location of the earthquake and so on.

The earthquake dataset was split into training and testing subsets. The training subset contained 80% of the data (14092 batches) and was used to train the machine learning model. The testing subset contained 20% of the data (2818 batches) and was used to evaluate the performance of the final trained model. This split into training and testing sets allowed the assessment of the model's ability to generalize to new, unseen data and provided a more realistic estimate of the model's performance on real-world earthquake data.

## RESULTS

### Support vector machines model.

A support vector machine with a radial basis function kernel was trained on the training data. The trained model achieved 83.76% accuracy in predicting earthquake locations on the training data. The model was then used to generate predictions on the test data, demonstrating its ability to generalize to unseen data.

### Linear SVC

A linear support vector machine (SVM) was trained on the available earthquake data. The trained linear model achieved 83.76% accuracy in predicting earthquake locations using the training data. The ability of the model to generalize to new, unseen data was demonstrated by generating predictions on held-out test data. The success of the linear SVM highlights the potential of machine learning methods for learning predictive patterns in earthquake data. The linear nature of the model also allows for easier interpretation of the patterns learned relative to more complex nonlinear models. These results suggest linear SVMs could serve as a viable tool for earthquake forecasting and analysis.

### RandomForest Model

A random forest model was trained on the available earthquake data. The trained model achieved 91.12% accuracy in predicting earthquake locations using the training data. The ability of the model to generalize to new, unseen data was demonstrated by generating predictions on held-out test data. Random forest models combine multiple decision trees to produce more accurate predictions than a single tree alone. They have the additional advantage of providing estimates of variable importance, which can be used to determine the most predictive input features. The success of the random forest approach suggests that ensemble methods could play an important role in earthquake forecasting and analysis. The variable importance scores may also yield insights into the factors that influence earthquake occurrence, magnitude, and location.

**KNN classification**

A k-nearest neighbors (KNN) classifier was trained on the available earthquake data. The KNN model identified the k closest training examples to a given test example and made a prediction based on the labels of the nearest neighbors. A KNN model with k = 5 neighbors achieved 82.73% accuracy in predicting earthquake locations on the test data. The KNN approach is a simple non-parametric method that can provide accurate predictions without assumptions about the distributions of features. The results suggest KNN could be an effective baseline approach for earthquake forecasting, with potential to incorporate more powerful distance metrics and alternative values of k.

**LGBM Classifier**

A LightGBM classifier was trained on the available earthquake data. LightGBM is a gradient boosting framework that uses tree based learning algorithms and can handle large-scale data efficiently. The trained LightGBM model achieved 84.22% accuracy in predicting earthquake locations on the training data. The ability of the model to generalize to new, unseen data was demonstrated by generating predictions on held-out test data. LightGBM has been shown to produce more accurate predictions than traditional tree boosting methods while training faster. The strong performance of LightGBM in this study suggests it could be a powerful approach for earthquake forecasting, especially given its ability to handle large datasets.

## CONCLUSION

The results show that the random forest, LightGBM, and support vector machine models achieved similar accuracy in predicting earthquake locations, with the random forest model achieving the highest accuracy. Given its strong performance and ability to provide variable importance estimates, the random forest model is selected as the top approach among the models compared in this study. The perfect prediction of a magnitude category by the SVM is likely due to an imbalance in the training data for that category, indicating the need for more balanced data. The confusion matrix for the random forest model shows it produces the fewest errors relative to the other models. Overall, these results demonstrate the feasibility of machine learning methods for earthquake forecasting and suggest that ensemble methods like random forests could provide particularly powerful tools for this task.

**REFERENCES**

1. Dong Yinfeng, Peng Bo, Li Xupeng. Preliminary analysis of Japan yokai earthquake characteristics. Yichang, China. 2011.
2. Yoshitaka Shibata; Noriki Uchida; Yuji Ohashi: Problem Analysis and Solutions of Information Network Systems on East Japan Great Earthquake. Fukuoka, Japan. 2012.
3. R.S. Kamath; R. K. Kamat: Magnitude Prediction Model for Japan Seismic Tremors Using Artificial Neural Network. Pune, India. 2022.
4. Roman Kail; Evgeny Burnaev; Alexey Zaytsev: Recurrent Convolutional Neural Networks Help to Predict Location of Earthquakes. Ministry of Science and Higher Education (**Grant Number**: 075-10-2021-068). 2022.