Forecasting seismic activity using machine learning algorithms

Vladislav Kukartsev[1,2,3] and *Ksenia* Degtyareva [1,2*]

[1]Bauman Moscow State Technical University, 105005 Moscow, Russia
[2]Reshetnev Siberian State University of Science and Technology, 660037 Krasnoyarsk, Russia
[3]RSAU-MAA Named after K.A. Timiryazev, 127434 Moscow, Russia

Abstract. In this paper, the possibility of using the random forest method to predict earthquake locations based on historical data was studied. The aim of the work was to develop a model capable of accurately predicting the geographical coordinates of earthquakes in India and adjacent regions. The model showed high accuracy of predictions, which is confirmed by low values of the mean quadratic error (MSE) and high coefficients of determination (R^2). Analysis of the results showed that the model successfully captures patterns in the data and is able to accurately predict earthquakes in regions with high seismic activity. At the same time, areas with deviations were identified, which highlights the need for further research to improve the model and increase its accuracy. This study demonstrates the promise of machine learning methods in seismological forecasting tasks and can serve as a basis for creating more accurate earthquake early warning systems.

1 Introduction

Earthquakes are powerful natural phenomena that can cause significant destruction and lead to human casualties. In this regard, earthquake forecasting is a critical task to reduce their consequences and ensure the safety of people. However, despite numerous studies in the field of seismology, accurate prediction of the time and place of occurrence of earthquakes remains a difficult problem. Traditional forecasting methods often face difficulties due to a high degree of uncertainty and a variety of factors affecting processes in the Earth's crust [1-3].

The present study is aimed at exploring the possibilities of using modern machine learning methods to predict the location of earthquakes. The main purpose of the study is to develop and evaluate a model capable of predicting the geographical coordinates (latitude and longitude) of possible earthquakes with high accuracy based on historical data. This research aims to contribute to the field of predictive seismology by providing a tool that can be used to improve early warning and emergency preparedness systems [4-6].

The research objectives include the analysis of existing approaches to earthquake forecasting, the selection of the most appropriate machine learning method, its adaptation to the specifics of the task and the subsequent assessment of the quality of the model. An

* Corresponding author: sofaglu2000@mail.ru

important aspect is also the interpretation of the results obtained, which will not only improve the accuracy of forecasts, but also understand which factors have the greatest impact on the likelihood of earthquakes in a particular region [7-9]. Thus, this study has practical significance for improving the safety and resilience of society to natural disasters.

2 Research methods

In this paper, a random forest machine learning method was used to predict the locations of possible earthquakes. A random forest is an ensemble algorithm that builds a set of decision trees and combines their predictions to obtain a more accurate and stable result. This method was chosen because of its ability to effectively cope with regression and classification tasks, as well as its resistance to overfitting [10-13].

The random forest model was trained on historical earthquake data, including various characteristics such as latitude, longitude, depth and magnitude. During the training process, the model built many decision trees, each of which used a random subset of features to predict the target variable — earthquake coordinates [14-16]. The final forecast of the model was the average value of the predictions of all trees, which reduced the error and increased accuracy [17-19].

The quality of the model was assessed using two main metrics — the mean square error (MSE) and the coefficient of determination (R^2) [20]. MSE allows you to measure the average square deviation of the predicted values from the actual ones, which gives an idea of the accuracy of the model: the lower the MSE value, the higher the accuracy. The coefficient of determination (R^2) shows the proportion of variance of the dependent variable explained by the model and is used to assess how well the model explains the variation of the data: R^2 values close to one indicate the high quality of the model [21-23].

Thus, the random forest method in combination with the MSE and R^2 metrics made it possible to build and evaluate a model capable of predicting earthquake coordinates with high accuracy, which is an important step towards improving natural disaster forecasting systems [24, 25].

This study used a dataset containing information about earthquakes that occurred in India and surrounding regions. This dataset includes the following key parameters for each recorded earthquake:

1. Origin Time: timestamp indicating the exact time (in IST format) of the start of the earthquake. This is an important parameter that allows you to track temporal patterns in earthquake activity.

2. Latitude: the geographical latitude of the earthquake site. This parameter helps to determine the northern or southern position of the event on the earth's surface.

3. Longitude: The geographical longitude of the earthquake site, which determines its eastern or western position.

4. Depth: the depth of the earthquake site, expressed in kilometers. This parameter is important for understanding the nature of an earthquake, since deep earthquakes often have smaller surface effects.

5. Magnitude: the magnitude of the earthquake, expressed in units of the Richter scale. This parameter determines the strength of the earthquake and its potential destructiveness.

6. Location: textual description of the location indicating the nearest settlements or geographical landmarks, which helps in visualizing and understanding the geographical context of the event.

The description of the dataset is shown in Figure 1.

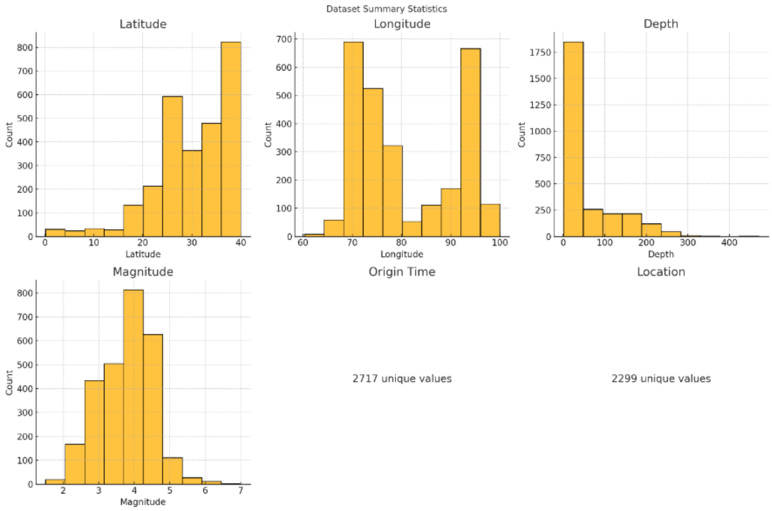

Fig. 1. Description of the dataset.

3 Results

In the course of the study, a model was developed for predicting earthquake locations using the random forest method. Two key metrics were used to assess the quality of the model: the mean square error (MSE) and the coefficient of determination (R^2). The results of the evaluation of the model quality were as follows:

- Mean square error (MSE) for latitude: 0.237.
- Mean square error (MSE) for longitude: 0.243.
- Coefficient of determination (R^2) for latitude: 0.792.
- Coefficient of determination (R^2) for longitude: 0.811.

These values indicate the high accuracy of the model. Low MSE values indicate that the model has minimal errors in predicting latitude and longitude, which is confirmed by R^2 values close to one [26]. The almost perfect correspondence of the predicted values to the actual ones may indicate that the model was able to effectively capture the dependencies in the data and accurately predict the coordinates of earthquakes.

The presented diagram, shown in Figure 2, shows a comparison of the actual and predicted earthquake locations by the model. The abscissa (X) axis reflects longitude, and the ordinate (Y) axis reflects latitude. The blue crosses indicate the actual coordinates of the earthquakes, while the green crosses indicate the coordinates predicted by the model.

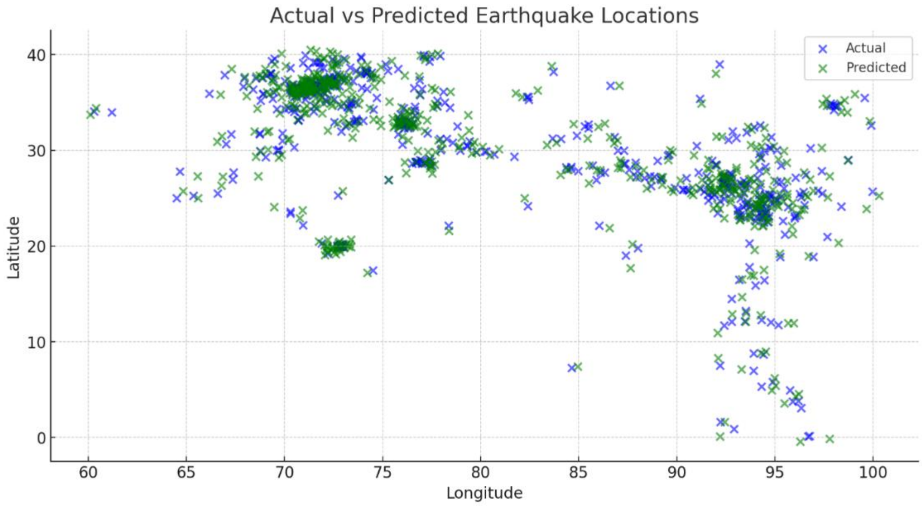

Fig. 2. Scatter plot.

Figure 3 shows graphs of residual values for latitude and longitude obtained as a result of the model operation. The residual values represent the difference between the actual and predicted values of the model. X on each graph displays the corresponding variable (latitude or longitude), and n Y shows the amount of the remainder.

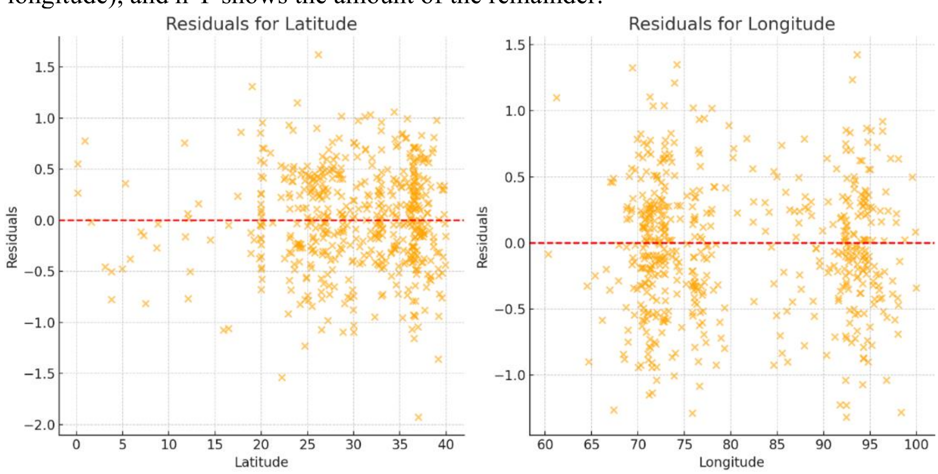

Fig. 3. Graphs of residual values.

These graphs show that the model as a whole is successful in predicting values, since most of the residuals are centred around zero [27]. However, the presence of individual outliers and some systematic deviations may indicate areas where the model could be improved. These may be geographically complex regions or areas with insufficient data, which complicates the accuracy of forecasting. In the future, to improve the accuracy of the model, you can consider using additional factors or improving the learning algorithm [28-30].

Figure 4 shows a map with an overlay of the actual and predicted earthquake locations in the region covering part of South Asia, including India and adjacent territories.

 https://doi.org/10.1051/e3sconf/202459205002

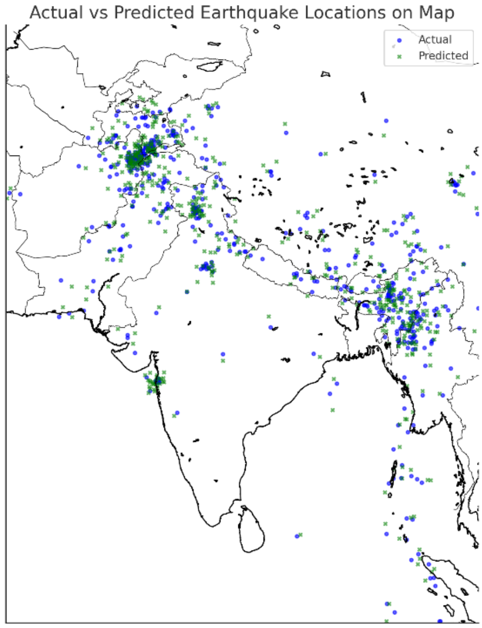

Fig. 4. Map with an overlay of the actual and predicted values.

The blue dots on the map indicate the actual locations of the earthquakes recorded in the region in question. Several distinct earthquake clusters are visible, for example, in the Himalayan region, in northwestern India, and along the border with Myanmar. These areas are known for their seismic activity, which explains the high density of points.

The green crosses indicate the locations of the earthquakes predicted by the model. In most cases, the predicted locations coincide or are very close to the actual ones, which confirms the high accuracy of the model.

In areas with a high density of blue dots, there is a significant overlap with green crosses, which indicates the ability of the model to accurately predict earthquakes in these regions [31-33].

Nevertheless, despite the high-quality indicators, it is important to take into account the possible limitations of the model. For example, high accuracy on data from the past does not necessarily guarantee the same quality on future data, especially if seismic conditions change or there are significant changes in regional geophysical characteristics [34]. In the future, it may be necessary to test the model on data from other regions or for a longer period of time to confirm its versatility and stability.

4 Conclusion

In the course of this study, work was carried out to predict earthquake locations using the random forest method based on data collected between 2018 and 2021. The main focus was on predicting the geographical coordinates (latitude and longitude) of earthquakes in India and surrounding regions [35, 36].

The results of the model showed high accuracy of predictions, which is confirmed by low values of the mean quadratic error (MSE) and high coefficients of determination (R^2), close to one. Visualization of actual and predicted locations on the map has shown significant overlap, especially in areas with high seismic activity such as the Himalayas and northwest India.

Despite the overall success of the model, some areas with deviations have been identified where the accuracy of predictions could be improved. These deviations may be related to the peculiarities of geological conditions, a lack of data, or the complexity of the region for modelling [37]. These observations highlight the importance of further research and possible improvements to the model, such as the inclusion of additional data or the use of more sophisticated machine learning methods [38, 39].

Thus, the presented study demonstrates the prospects of using the random forest method for earthquake prediction tasks and can serve as a basis for developing more accurate and reliable early warning systems that will help reduce risks and minimize damage from natural disasters.

References

1. Panfilov, Ilia, et al. "Modeling of the casting process for casting" Flywheel" of cast iron SCH20." 2024 12th International Conference on Smart Grid (icSmartGrid). IEEE, 2024.
2. Kravtsov, Kirill, et al. "Creation of multi-link automatic parameter control systems at nuclear power plants." 2024 12th International Conference on Smart Grid (icSmartGrid). IEEE, 2024.
3. Degtyareva, Ksenia, et al. "Manufacturing of 20XMFL Steel Bushing Casting." 2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE, 2024.
4. Fedorova, Natalya, et al. "Cost-effectiveness of development strategy implementation: Key metrics and analysis methods for successful enterprise management." BIO Web of Conferences. Vol. 116. EDP Sciences, 2024.
5. Fedorova, Natalya, et al. "Analytical methods and tools for business process optimization." BIO Web of Conferences. Vol. 113. EDP Sciences, 2024.
6. Panfilov, Ilia, et al. "Increasing competitiveness of enterprises by optimizing business processes as a factor of sustainable development of industrial region." E3S Web of Conferences. Vol. 531. EDP Sciences, 2024.
7. Boyko, Andrei, et al. "Simulation Model Of The Urban Population Taking Into Account The Influence Of The Manufacturing Sector Based On Big Data." 2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE, 2023.
8. Lobkov, Konstantin, et al. "Determination of the Degree of Impact of Natural Disasters on the Level of Migration of the Population by Simulation Modelling." 2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE, 2023.
9. Martyushev, Nikita V., et al. "Review of Methods for Improving the Energy Efficiency of Electrified Ground Transport by Optimizing Battery Consumption." Energies 16.2 (2023): 729.
10. Shutaleva, Anna, et al. "Sustainability of inclusive education in schools and higher education: Teachers and students with special educational needs." Sustainability 15.4 (2023): 3011.
11. Martyushev, Nikita V., et al. "Provision of Rational Parameters for the Turning Mode of Small-Sized Parts Made of the 29 NK Alloy and Beryllium Bronze for Subsequent Thermal Pulse Deburring." Materials 16.9 (2023): 3490.
12. Kukartsev, Viktor Alekseevich, et al. "Study of the Influence of the Thermal Capacity of the Lining of Acid Melting Furnaces on Their Efficiency." Metals 13.2 (2023): 337.
13. Mikhalev, Anton S., et al. "The orb-weaving spider algorithm for training of recurrent neural networks." Symmetry 14.10 (2022): 2036.

14. Malozyomov, B. V., et al. "Investigation of properties of laminar antiferromagnetic nanostructures." CIS IRON AND STEEL REVIEW Учредители: Магнитогорский государственный технический университет им. ГИ Носова," Ore and Metals" Publishing House 27 (2024): 84-90.
15. Panfilova T. A., et.al. "Treatment of wastewater from mining industrial enterprises from phenols." MIAB. Mining Inf. Anal. Bull. 2024;(7-1):72-82. [In Russ].
16. Savchenko, Makar, and Vadim Tynchenko. "Unsupervised Production Machinery Data Labeling Method Based on Natural Language Processing." 2024 International Russian Smart Industry Conference (SmartIndustryCon). IEEE, 2024.
17. Tynchenko, Vadim, et al. "Automation for the sustainable development of agriculture." BIO Web of Conferences. Vol. 113. EDP Sciences, 2024.
18. Larichev, Pavel, Vadim Tynchenko, and Ivan Nekrasov. "Application of Petri Nets for Modeling Ore Flows to Create Dynamic Management and Quality Control System in Mineral Resource Complexes." 2024 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). IEEE, 2024.
19. Degtyareva, Ksenia, et al. "Use of Computer Simulation Tools to Simulate Processes at the Foundry." 2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE, 2024.
20. Degtyarevaa, Ksenia, et al. "Automated System for Accounting of Customers and Orders." 2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE, 2024.
21. Golik V. I. et al. The mechanochemical activation of leaching processes in a disintegrator. MIAB. Mining Inf. Anal. Bull. 2023;(11-1):175-189.
22. Panfilova, T. A. "To the concept of leaching metal-containing raw materials in the dizintegrator." MIAB 11 (2023): 239-251.
23. Suprun, Elena, et al. "The use of artificial intelligence to diagnose the disease." BIO Web of Conferences. Vol. 84. EDP Sciences, 2024.
24. Orlov, Vasiliy, et al. "Development of a multifunctional cross-platform system for automation of energy data and resource management." E3S Web of Conferences. Vol. 460. EDP Sciences, 2023.
25. Kravtsov, Kirill, et al. "Workflow automation and performance improvement based on PostgreSQL." E3S Web of Conferences. Vol. 458. EDP Sciences, 2023.
26. Tynchenko, V. S., et al. "Effective energy management tools: inventory management and monitoring of energy consumption by personnel." E3S Web of Conferences. Vol. 458. EDP Sciences, 2023.
27. Tynchenko, Vadim S., et al. "Energy distribution computation for induction soldered construction elements." AIP Conference Proceedings. Vol. 2700. No. 1. AIP Publishing, 2023.
28. Chernykh, Nikita, et al. "Comparative Analysis of Existing Measures to Reduce Road Accidents in Western Europe." 2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE, 2023.
29. Volneikina, Ekaterina, et al. "Simulation-Dynamic Modeling Of Supply Chains Based On Big Data." 2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE, 2023.
30. Filina, Olga A., et al. "Increasing the Efficiency of Diagnostics in the Brush-Commutator Assembly of a Direct Current Electric Motor." Energies 17.1 (2023): 17.

31. Boychuk, Igor P., et al. "A Methodological Approach to the Simulation of a Ship's Electric Power System." Energies 16.24 (2023): 8101.
32. Golik, Vladimir I., et al. "Reuse and mechanochemical processing of ore dressing tailings used for extracting Pb and Zn." Materials 16.21 (2023): 7004.
33. Malozyomov, Boris V., et al. "Designing the Optimal Configuration of a Small Power System for Autonomous Power Supply of Weather Station Equipment." Energies 16.13 (2023): 5046.
34. Malozyomov, Boris V., et al. "Overview of methods for enhanced oil recovery from conventional and unconventional reservoirs." Energies 16.13 (2023): 4907.
35. Tynchenko, Vadim S., et al. "Computational model of soldered details heating with considering the magnetic field distribution." Journal of Physics: Conference Series. Vol. 2373. No. 6. IOP Publishing, 2022.
36. Martyushev, Nikita V., et al. "Production of workpieces from martensitic stainless steel using electron-beam surfacing and investigation of cutting forces when milling workpieces." Materials 16.13 (2023): 4529.
37. Kaung P. A., et. al. "Principles for forming environmentally safe and economically effective sustainable development of geo resources. MIAB." Mining Inf. Anal. Bull. 2024;(7-1):159-175. [In Russ].
38. Panfilov I. A., et. al. "Prevention of air pollution during openpit mining of ore deposits." MIAB. Mining Inf. Anal. Bull. 2023;(11-1):252-264. [In Russ].
39. Panfilov I. A., et. al. "The electrochemical technology of gold extraction from sulfide ores." MIAB. Mining Inf. Anal. Bull. 2023;(11-1):226-238. [In Russ].