Geoinformatics & Geostatistics: An Overview

Research Article A SCITECHNOL JOURNAL

Turkey's Earthquakes: Damage Prediction and Feature Significance Using a Multivariate Analysis

Shrey Shah[1*] and Alex Lin[2]

[1]*Nashua High School South, Nashua, New Hampshire, United States of America*

[2]*Walnut High School, California, United States of America*

***Corresponding Author:** Shrey Shah, Nashua High School South, New Hampshire, United States of America; E-mail: shreyshah1011@gmail.com*

Received date: 28 November, 2024 Manuscript No. GIGS-24-153503;

Editor assigned date: 02 December, 2024, PreQC No. GIGS-24-153503 (PQ);

Reviewed date: 16 December, 2024, QC No. GIGS-24-153503;

Revised date: 23 December, 2024, Manuscript No. GIGS-24-153503 (R);

Published date: 30 December, 2024, DOI: 10.4172/2327-4581.1000423.

Abstract

Accurate damage prediction is important for disaster preparedness and response strategies, particularly given the frequent earthquakes in Turkey. Utilizing datasets on earthquake data, infrastructural quality metrics contemporary socioeconomic factors, we tested various machine-learning architectures to forecast death tolls and fatalities per affected population. Our findings indicate that the random forest model provides the most reliable predictions. The model highlights earthquake magnitude and building stability as the primary determinants of damage. This research contributes to the reduction of fatalities in future seismic events in Turkey.

Keywords: Earthquakes; Natural disasters; Modified mercalli intensity; Peak ground acceleration

Introduction

Earthquakes serve as one of the most catastrophic natural disasters, claiming over 61,000 lives in 2023 alone [1,2]. Turkey, in particular, is struck by thousands of earthquakes annually because of the multiple major seismic fault lines running through its geography. Most notably, the recent 2023 Turkey-Syria earthquakes, of magnitudes 7.8 and 7.7, killed a confirmed 53,537 people [3]. Recent predictive models, such as the RECAST model, excel at forecasting future earthquakes, along with information such as power level, location time of formation [4]. However, in order to better prepare earthquake utilizing strategies, it is necessary to predict the damage outcome and feature significance of these earthquakes. Our study bridges the gap in understanding earthquake damage by predicting its severity based on major influencing factors such as building stability, earthquake depth population density per province. We also seek to determine the significance of these factors in contributing to the overall destruction caused by earthquakes. The primary objective of this research is to formulate a model that accurately predicts the death toll and fatalities per affected population in Turkish earthquakes by testing various machine learning architectures and using the most accurate one to assess the significance of each factor, thereby enhancing earthquake preparedness and resistance.

Materials and Methods

In our study, we analyzed four datasets of earthquakes occurring prior to 1950, each containing information on magnitude, earthquake depth (km), Modified Mercalli Intensity (MMI) or shaking intensity of earthquakes, death toll epicenter coordinates [5-8]. Additionally, we integrated a 2022 dataset on factors for each province in Turkey by matching epicenter coordinates with provinces [9,10].

Data sets of earthquakes

The factors used are population density, structural integrity and stability based on age, design, etc., of building and the susceptibility of buildings to earthquake damage [11]. We selected these variables since variations in them were hypothesized to influence death tolls. Our study requires two sets of machine learning models: One using death toll as the output variable and the other using death per capita of the affected population. Total deaths are useful in predicting the number of deaths from an upcoming earthquake in a specific region to help in immediate disaster response and resource allocation. Deaths per capita can be used for comparing the effectiveness of building codes and other infrastructural features more directly.

To calculate the death per capita of each earthquake, we first utilized Joyner and Boore's (1981) attenuation relations that predict Peak Ground Acceleration (PGA) using magnitude and the distance from the fault rupture, as shown in Formula 1 [12].

$$\text{Log (PGA(gravity))}=b_1+b_2M+b_3\log (R+b_4)\ldots\ldots (1)$$

The values of the coefficients b_1, b_2, b_3 and b_4 in the above formula were derived from a study by Kalkan et al., by performing regression analyses on ground motion data specific to Turkish seismic events [13]. In our study, we consider individuals in an area with a Modified Mercalli Intensity (MMI) of at least IV, where earthquakes are widely felt and fatalities are more likely, as affected by the event [9]. This threshold was chosen because fatalities are unlikely below an MMI of IV, which corresponds to a Peak Ground Acceleration (PGA) of 2.8 g [14]. By substituting the known values into the relations, the radius was found.

In our study, we assume a circular pattern for earthquake damage prediction, which simplifies the complex nature of seismic wave propagation and distribution. Seismic waves radiate outward in all directions from the earthquake's epicenter [15]. While this circular approximation might not capture all the details of fault geometries, the impact on our model's accuracy is minimal due to the general nature of seismic wave dispersion. Using the radius derived from the attenuation relations, we applied the formula for the area of a circle to estimate the surface area of the affected region, which was then multiplied by the population density of the province where the earthquake occurred to find the total affected population. We then divided the total number of fatalities by the total affected population to calculate the death per capita for the earthquake.

Evaluation metrics

In training our models, we deviate from the standard Mean Squared Error (MSE) loss function due to its interpretative limitations with our

All articles published in Geoinformatics & Geostatistics: An Overview are the property of SciTechnol and is protected by copyright laws. Copyright © 2024, SciTechnol, All Rights Reserved.

data. Instead, we utilize Mean Absolute Percent Error (MAPE), which presents loss as a percentage error rather than a numerical discrepancy, making it more suitable for our context. Additionally, we employ Mean Absolute Error (MAE) as a complementary metric. MAE is beneficial as it uniformly weights all values, irrespective of direction, providing a robust assessment of model performance.

Model selections

In this study, our various machine learning models use the factors described as x-values and either the number of fatalities per population affected or the death toll as y-values.

The linear regression model is this study's baseline model. It implements Ordinary Least Squares (OLS) regression to fit the dataset and generate a model summary with the metrics described.

The neural network's architecture features an input layer, multiple hidden layers an output layer. The model is compiled with the Adam optimizer and MAPE loss function. The data undergoes an 80/20 train-test split the model is trained for 4500 epochs.

For the decision tree, random forest, ridge Least Absolute Shrinkage and Selection Operator (LASSO) machine learning models, the data is first split in an 80/20 train test split. Hyper-parameter optimization is conducted on the training data using grid search with 5-fold cross-validation, employing negative MAPE as the scoring metric. The model is then trained and evaluated on the test data.

In the decision tree regressor, the hyperparameters are the maximum depth, the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node the number of features to consider when looking for the best split. In the random forest regressor, the hyperparameters are the number of trees in the forest, the maximum depth of each tree, the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node whether bootstrap samples are used when building trees. Both Ridge and LASSO Regression models use one hyperparameter for alpha.

Results

The linear regression models exhibited a suboptimal performance, with a MAPE of 23.25 and a MAE of 0.0030 for per capita predictions and a MAPE of 88.92 and an MAE of 2737.01 for the total death toll. Despite the seemingly favorable per capita predictions, fitting the model directly to Ordinary Least Squares (OLS) without splitting the data into training and testing sets suggests potential overfitting. The testing values for differing machine learning models grouped together due to their shared evaluation methods, as detailed (Tables 1 and 2).

Based on the performance metrics, the decision tree model and random forest model provide the best results due to having low MAPE and MAE values compared to the other data. The Random forest model would be a better predictive model, not only due to the better MAPE scores, is an ensemble model combining multiple decision trees. Since the forest model averages many tree models, this leads to a lower risk of overfitting and better feature variability. Therefore, the Random forest model is the best damage-predictive model for this study (Figures 1 and 2).

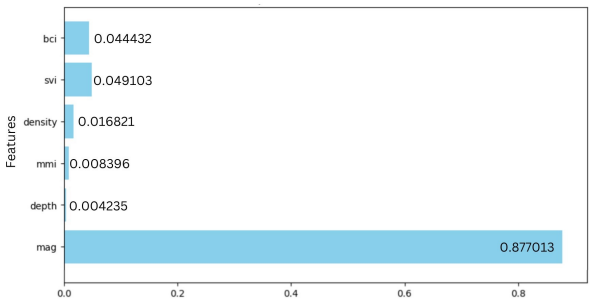

Figure 1: Factor significance in influencing earthquake deaths.

Model	Neural network	Decision tree	Random forest	Ridge	LASSO
Training MAPE	85.7	5.79	7.79	19.39	35.01
Training MAE	0.0024	0.0026	0.0026	0.003	0.0031
Testing MAPE	83.97	9.92	10.61	32.09	31.64
Testing MAE	0.0013	0.0026	0.0026	0.0022	0.0018

Table 1: Deaths per population predictor evaluation metrics.

Model	Neural network	Decision tree	Random forest	Ridge	LASSO
Training MAPE	84.37	3.99	1.76	35.01	86.29
Training MAE	2379.86	464.98	187.83	2193.49	2453.03
Testing MAPE	83.78	12.16	6.2	84.91	63.93
Testing MAE	2091.89	802.43	1175.13	3041.39	2858.44

Table 2: Death toll predictor evaluation metrics.

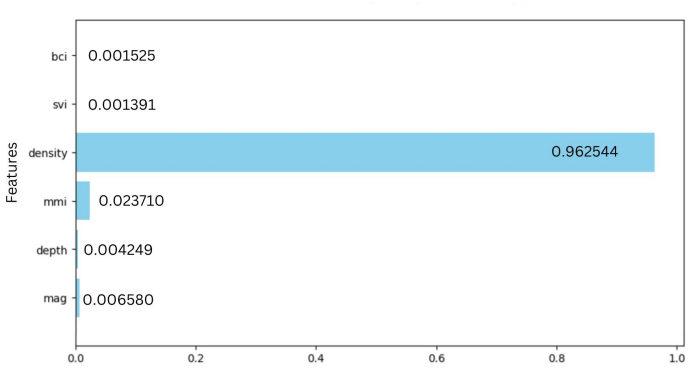

Figure 2: Factor significance in influencing earthquake deaths per capita.

In Random forest models, factor importance applies a score to each x-coordinate, reflecting how relatively significant each factor is in predicting the damage. This is calculated by adding the reduction in error that each feature contributes at each split across all nodes and trees in the forest and then averaging these scores. The importance values are normalized to sum to one, providing a clear comparison of how much each feature contributes to the model. The feature significance of both predicting deaths and predicting deaths per population affected is shown in the graphs above.

For the model predicting deaths per population affected, the feature importance values indicate that population density is majorly the most significant factor. This is because it is directly involved in calculating the target variable, as it is multiplied by surface area affected to get the number of people affected.

In the death toll prediction model, magnitude is the most significant factor in earthquake damage. Notably, SVI and BCI have similar importance scores, suggesting that earthquake vulnerability and building stability have comparable influence on predicting the death toll. Understanding these importance scores helps in identifying which factors are most critical to focus on when developing models and strategies for earthquake impact mitigation and response planning.

Discussion

In this article, the models were designed in the field of earthquake damage prediction and determined the significance of socioeconomic factors to earthquake damage. In the future, combining the results from earthquake prediction models such as the RECAST and ETAS model could make our model extremely useful to Turkish disaster preparation agencies, as we would be predicting vital information on earthquakes to come, leading to better reducing strategies [4,16].

Nonetheless, there are important limitations we hope to address in future iterations of this model. First, conceptually, there are many reasons earthquake damage prediction is difficult. Often, the aftermath of the earthquake can have a more devastating effect than the seismic event itself. Unpredictable variables such as possible fires or explosions can create a difficult environment for damage prediction as it is impossible to know what other catastrophes a simple earthquake could lead to. Our model provides a useful conceptual foothold to build upon. Theoretically, as technology advances, more sophisticated methods will enable us to conquer this issue. Additionally, since we only used data preceding 1950, there were certain limitations, as most data collected on earthquakes before that point would be inaccurate or incomplete. This led to a limited training dataset of only 99 values, but all with complete data in the most important variables.

Recent advancements in machine learning have improved the accuracy of many earthquake prediction models. These models, such as the recent RECAST models, rely on dense seismic networks and automated data processing techniques [4]. The RECAST Model utilizes deep learning neural networks to expand upon the ETAS model's temporal point processing model. By adding features such as predicted location as well as other geophysical data, the Rapid Earthquake Analysis and Simulation Tool (RECAST) model improves the accuracy and quantity of data in earthquake prediction. The development in machine learning and earthquake prediction technology has been paralleled by a rise in earthquake damage prediction. For example, Driven Data began hosting a damage prediction competition, where participants predict the level of damage caused to buildings from Nepal's Gorkha earthquake, based on building data and socioeconomic statistics [17]. However, the competition overlooks dynamic variables having the largest effects on the loss of human life. Our team realized that with different data, we could use a similar approach to predict the damage of earthquakes that will happen in the future, thus connecting earthquake prediction to damage prediction.

Conclusion

This study has concentrated on predicting earthquake damage severity in Turkey by analyzing factors such as building stability, earthquake depth population density. Through the evaluation of various machine learning architectures, our objective has been to accurately forecast death tolls and fatalities per affected population. Additionally, performing feature importance on the final model improves earthquake preparedness strategies. We hope that this research will prompt increased efforts toward reducing fatalities caused by earthquakes.

Acknowledgement

All source code and the text of this paper were authored by Shrey Shah, Alex Lin, Scott Lin Josh Patel, who designed the project following an extensive literature review. Scott Lin and Josh Patel are high school students at Foothill High School and The Peddie School, respectively we extend our gratitude to Mike Lam and Kevin Zhu for their contributions through lectures on machine learning and research skills, suggested readings, high-level guidance constructive comments on the manuscript.

References

1. NCEI/WDS Global Significant Earthquake Database. Noaa National Centers for Environmental Information. 2024.
2. Herece E (1990) The fault trace of 1953 Yenice-Gönen earthquake and the westernmost known extension of the NAF System in the Biga Peninsula. Bulleti Min Res Expl 111(111): 31-42.
3. Presse AF. Nearly 60,000 Killed in 2023 Turkey, Syria Quake: New toll.
4. Dascher CK, Shchur O, Brodsky EE, Günnemann S (2023) Using deep learning for flexible and scalable earthquake forecasting. Geophy Res Lett 50(17): e2023GL103909.
5. Kandilli Observatory and Earthquake Research Institute: Large Turkish Earthquakes, Instanbul, Turkey. 2002

6. Ergin K (1967) A catalogue of earthquakes for Turkey and surrounding area (11 AD to 1964 AD). Tech Univ Mining Eng Fac Publ 24:189.
7. Department of The Interior U.S. Geological Survey New Information Resources of the U.S. Geological Survey Library System. 1986.
8. Sezen H, Altunisik M, Emin A, Caglar N, Demir A, et al. (2023) StEER 2022 Mw 6.1 Duzce, Turkey, Mw 6.1 Earthquake. DesignSafe CI.
9. Modified Mercalli Intensity Scale. USGS General Interest Publication.
10. Türkiiye Country Data, Links and Map by Administrative Structure. Republic of türkiiye.
11. Turkish Statistical Institute.
12. Joyner WB, Boore DM (1981) Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. Bul Seismol Soc Am 71(6):2011-38.
13. Kalkan E, Gülkan P (2004) Site-dependent spectra derived from ground motion records in Turkey. Earthquake Spectra 20(4): 1111-38.
14. Wald DJ, Quitoriano V, Heaton TH, Kanamori H (1999) Relationships between peak ground acceleration, peak ground velocity modified Mercalli intensity in California. Earthquake spectra 15(3):557-64.
15. The Science of Earthquakes. U.S. Geological Survey.
16. Lombardi AM (2015) Estimation of the parameters of ETAS models by Simulated Annealing. Scienti Rep 5(1):8417.
17. Open Machine Learning Competition for Earthquake Damage Prediction (2019). Driven data.