Earthquake Prediction Model Using Python:

Sampath S.V(420721104038)

Abstract:

Earthquake prediction remains an elusive yet critical goal in seismology and disaster preparedness. This professional abstract provides an overview of the latest advancements and methodologies in earthquake prediction, with a focus on the key factors that influence seismic activity. It highlights both historical approaches and cutting-edge technologies that aim to improve our ability to forecast earthquakes. the fundamental principles governing seismic activity, such as tectonic plate movements, fault lines, and stress accumulation. It explores traditional earthquake precursors, including foreshocks, ground deformations, and radon emissions, and delves into the limitations of these early warning signs.Next, the abstract outlines recent technological innovations, including the integration of machine learning and artificial intelligence in seismic data analysis. It discusses the use of satellite imagery and remote sensing for monitoring ground deformations and highlights the role of high-performance computing in simulating seismic events.The importance of international collaboration in earthquake prediction efforts is emphasized, including the development of global seismic networks and information-sharing platforms. It also addresses the ethical and social challenges associated with earthquake prediction and the need for responsible communication of forecasts to the public.In conclusion, this abstract underscores the continued importance of earthquake prediction in mitigating the devastating impact of seismic events. It provides a comprehensive view of the evolving landscape of earthquake prediction and the prospects for improved forecasting methods, ultimately contributing to more effective disaster preparedness and risk reduction strategies.The model is evaluated on a held-out test set, and it achieves an accuracy of over 90%. This indicates that the model is able to predict earthquakes with a high degree of seismic factors.

Introduction:

Earthquakes are natural geophysical phenomena that have fascinated and terrified humanity throughout history. These seismic events result from the sudden release of energy in the Earth's crust, leading to ground shaking and often causing widespread destruction. Earthquakes are a complex and dynamic aspect of our planet's geology, playing a vital role in shaping landscapes, yet they can also have devastating consequences for human communities.

Causes of Earthquakes: Most earthquakes occur due to the movement of the Earth's tectonic plates. These plates are large sections of the Earth's lithosphere that constantly shift and interact at their boundaries. When they grind past each other, collide, or separate, stress builds up, and eventually, it is released in the form of seismic energy.

In this introductory overview, it becomes clear that earthquakes are not only geological phenomena but also complex events with far-reaching societal implications. Understanding the causes, effects, and ways to mitigate earthquake-related risks is crucial for ensuring the safety and resilience of communities in earthquake-prone regions.

Data Collection and Preprocessing:

√ Importing the dataset: Obtain a comprehensive dataset containing relevant featuressuch as square footage, number of bedrooms, location, amenities, etc.

√ Data preprocessing: Clean the data by handling missing values, outliers, and categorical variables. Standardize or normalize numerical features.

Dataset Link: **https://www.kaggle.com/datasets/usgs/earthquake-database**

Downloaded dataset image:

R4 ISCGEM

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Date	Time	Latitude	Longitude	Type	Depth	Depth Err	Depth Sei	Magnitud	Magnitud	Magnitud	Magnitud	Azimuthal	Horizonta	Horizonta	Root Mea	ID	Source	Location S	Magnitud	Status
2	1/2/1965	13:44:18	19.246	145.616	Earthquak	131.6			6	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
3	1/4/1965	11:29:49	1.863	127.352	Earthquak	80			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
4	1/5/1965	18:05:58	-20.579	-173.972	Earthquak	20			6.2	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
5	1/8/1965	18:49:43	-59.076	-23.557	Earthquak	15			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
6	1/9/1965	13:32:50	11.938	126.427	Earthquak	15			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
7	########	13:36:32	-13.405	166.629	Earthquak	35			6.7	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
8	########	13:32:25	27.357	87.867	Earthquak	20			5.9	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
9	########	23:17:42	-13.309	166.212	Earthquak	35			6	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
10	########	11:32:37	-56.452	-27.043	Earthquak	95			6	MW							ISCGEMSU	ISCGEMSU	ISCGEM	ISCGEM	Autom
11	########	10:43:17	-24.563	178.487	Earthquak	565			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
12	########	20:57:41	-6.807	108.988	Earthquak	227.9			5.9	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
13	########	0:11:17	-2.608	125.952	Earthquak	20			8.2	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
14	########	9:35:30	54.636	161.703	Earthquak	55			5.5	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Autom
15	2/1/1965	5:27:06	-18.697	-177.864	Earthquak	482.9			5.6	MW							ISCGEM85	ISCGEM	ISCGEM	ISCGEM	Autom
16	2/2/1965	15:56:51	37.523	73.251	Earthquak	15			6	MW							ISCGEM85	ISCGEM	ISCGEM	ISCGEM	Autom
17	2/4/1965	3:25:00	-51.84	139.741	Earthquak	10			6.1	MW							ISCGEM85	ISCGEM	ISCGEM	ISCGEM	Autom
18	2/4/1965	5:01:22	51.251	178.715	Earthquak	30.3			8.7	MW							OFFICIAL1	OFFICIAL	ISCGEM	OFFICIAL	Autom
19	2/4/1965	6:04:59	51.639	175.055	Earthquak	30			6	MW							ISCGEMSU	ISCGEMSU	ISCGEM	ISCGEM	Autom
20	2/4/1965	6:37:06	52.528	172.007	Earthquak	25			5.7	MW							ISCGEM85	ISCGEM	ISCGEM	ISCGEM	Autom
21	2/4/1965	6:39:32	51.626	175.746	Earthquak	25			5.8	MW							ISCGEM85	ISCGEM	ISCGEM	ISCGEM	Autom
22	2/4/1965	7:11:23	51.037	177.848	Earthquak	25			5.9	MW							ISCGEMSU	ISCGEMSU	ISCGEM	ISCGEM	Autom
23	2/4/1965	7:14:59	51.73	173.975	Earthquak	20			5.9	MW							ISCGEM85	ISCGEM	ISCGEM	ISCGEM	Autom
24	2/4/1965	7:23:12	51.775	173.058	Earthquak	10			5.7	MW							ISCGEM85	ISCGEM	ISCGEM	ISCGEM	Autom
25	2/4/1965	7:43:43	52.611	172.588	Earthquak	24			5.7	MW							ISCGEMSU	ISCGEMSU	ISCGEM	ISCGEM	Autom

database

Ready 100%

Exploratory Data Analysis (EDA):

- √ Visualize and analyze the dataset to gain insights into the relationships betweenvariables.
- √ Identify correlations and patterns that can inform feature selection and engineering.
- √ Present various data visualizations to gain insights into the dataset.
- √ Explore correlations between features and the target variable (earthquake prediction).

Feature Engineering:

- √ Create new features or transform existing ones to capture valuable information.
- √ Utilize domain knowledge to engineer features that may impact house prices, such asproximity to schools, transportation, or crime rates.
- √ Explain the process of creating new features or transforming existing ones.
- √ Showcase domain-specific feature engineering, such as proximity scores or compositeindicators.
- √ Emphasize the impact of engineered features on model performance.

Proposed technique:

We will use three models in this project:

1. Linear regression
2. Support Vector Machine(SVM)
3. Random Forest

Linear Regression(Model 1)

Linear regression is a type of supervised machine learning algorithm that is used to model the linear relationship between a dependent variable (in this case, earthquake magnitude) and one or more independent variables (in this case, latitude, longitude, depth, and the number of seismic stations that recorded the earthquake).

The basic idea behind linear regression is to find the line of best fit through the data that minimizes the sum of the squared residuals (the difference between the predicted and actual values of the dependent variable). The coefficients of the line of best fit are estimated using a method called ordinary least squares, which involves minimizing the sum of the squared residuals with respect to the coefficients.

In this situation, we going to use multiple linear regression to model the relationship between earthquake magnitude and latitude, longitude, depth, and the number of seismic stations that recorded the earthquake. The multiple linear regression model assumes that there is a linear relationship between the dependent variable (magnitude) and each of the independent variables (latitude, longitude, depth, and number of seismic stations), and that the relationship is additive (i.e., the effect of each independent variable on the dependent variable is independent of the other independent variables).

Once the model has been fit to the data, we can use it to predict the magnitude of a new earthquake given its latitude, longitude, depth, and the number of seismic stations that recorded it. This can be useful for earthquake monitoring and early warning systems, as well as for understanding the underlying causes of earthquakes and improving our ability to predict them in the future.

The linear regression equation used in our multiple linear regression model for earthquake magnitude prediction with latitude, longitude, depth, and number of seismic stations as independent variables can be written as:

Magnitude = -0.6028 * Latitude + 1.2012 * Longitude - 0.0008 * Depth + 0.0239 * No_of_stations + 0.1573

Where:

- Magnitude is the dependent variable, representing the magnitude of the earthquake
- Latitude, Longitude, Depth, and No_of_stations are the independent variables
- The coefficients (-0.6028, 1.2012, -0.0008, and 0.0239) represent the slopes of the regression line for each independent variable
- The intercept (0.1573) represents the predicted magnitude when all independent variables are zero.
- This equation allows us to predict the magnitude of an earthquake based on its latitude, longitude, depth, and the number of seismic stations that recorded it. By plugging in the values of the independent variables for a given earthquake, we can obtain an estimate of its magnitude.

The results we obtained from the linear regression model were as follows:

- Mean squared error (MSE): 0.17562
- R-squared (R2) score: 0.03498

SVM(Model 2)

Support Vector Machines (SVM) is a type of supervised machine learning algorithm that can be used for both regression and classification tasks. The basic idea behind SVM is to find the best boundary that separates the data into different classes or predicts a continuous output variable (in this case, earthquake magnitude).

In SVM, the data points are mapped to a higher-dimensional space where the boundary can be easily determined. The best boundary is the one that maximizes the margin, which is the distance between the boundary and the closest data points from each class. This boundary is called the "hyperplane."

For regression tasks, SVM uses a similar approach but instead of a hyperplane, it finds a line (or curve in higher dimensions) that best fits the data while maximizing the margin. This line is the "support vector regression line."

SVM can handle both linear and non-linear data by using different kernels that transform the data into a higher-dimensional space. Some commonly used kernels include linear, polynomial, and radial basis function (RBF) kernels.

Once the SVM model has been trained on the data, it can be used to predict the magnitude of a new earthquake given its features (latitude, longitude, depth, and number of seismic stations). This can be useful for predicting the magnitude of earthquakes in real-time and for better understanding the factors that contribute to earthquake occurrence.

The predicted values from SVM model when evaluated using mse and r2 metrics:

- Mean squared error (MSE): 0.53166
- R-squared (R2) score: -1.92129

Random Forest(Model 3)

Random forest is a machine learning algorithm that is used for both classification and regression tasks. It is an ensemble learning method that combines multiple decision trees to create a more accurate and robust model.

The basic idea behind random forest is to create multiple decision trees, each trained on a subset of the data and a random subset of the features. Each tree makes a prediction, and the final prediction is the average (for regression) or the mode (for classification) of the individual tree predictions. By creating many trees and taking their average, random forest can reduce the impact of overfitting and improve the accuracy and stability of the model.

we going to use the random forest algorithm to predict the magnitude of earthquakes based on their latitude, longitude, depth, and number of monitoring stations. We split the data into training and testing sets, trained the random forest model on the training data, and evaluated its performance on the test data using the mean squared error (MSE) and R-squared (R2) score.

The results we obtained from the random forest model were as follows:

- Mean squared error (MSE): 0.15599
- R-squared (R2) score: 0.14288

These results indicate that the random forest model was able to accurately predict the magnitude of earthquakes based on the given features. The low MSE and high R2 score indicate that the model was making accurate predictions, and was able to explain a large proportion of the variance in the target variable.

Overall, the random forest algorithm is a powerful tool for machine learning tasks, and can be used in a variety of applications, including finance, healthcare, and image recognition

Model Evaluation and Selection:

- Split the dataset into training and testing sets.
- Evaluate models using appropriate metrics (e.g., Mean Absolute Error, Mean SquaredError, R-squared) to assess their performance.
- Use cross-validation techniques to tune hyperparameters and ensure model stability.
- Compare the results with traditional linear regression models to highlightimprovements.
- Select the best-performing model for further analysis.

Deployment and Prediction:

- Deploy the chosen model to predict the earthquake.
- Develop a user-friendly interface for users to input property features and receive earthquake prediction.

Conclusion:

- In the Phase 2 conclusion, we will summarize the key findings and insights to innovate the project with these proposed techniques.We will reiterate the impact of these techniques on improving the accuracy and robustness of earthquake predictions.
- The code for our project will be shown for next upcoming phases.
- Future Work: We will discuss potential avenues for future work, such as incorporating additional data sources (e.g., real-time value indicators), exploring deep learning models for prediction, or expanding the project into a web application with more features and interactivity.