Advancing Earthquake Risk Reduction through Machine Learning Enhanced Early Warning Systems

MSc Research Project

MSc Data Analytics

Pranav Ramakrishnan

Student ID: 23107979

School of Computing

National College of Ireland

Supervisor:Abdul Qayum

National College of Ireland

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ....Pranav Ramakrishnan..

Student ID: ....x23107979..

Programme: ....MSc Data Analytics........................ **Year** ......2024-25..........

Module: ......MSc Research Project..

Supervisor: ....Abdul Qayum...

Submission Due Date: ....12-12-2024..

Project Title: ...Advanced Earthquake Risk Reduction Through Machine Learning Enhanced Early Warning System...

Word Count: ....8348................................ **Page Count:**.....................20.................................

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ......

Date: ....12-12-2024...

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□

You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Advancing Earthquake Risk Reduction through Machine Learning Enhanced Early Warning Systems

Pranav Ramakrishnan

23107979

Abstract

Earthquakes are one of the most dangerous natural disasters.it is very important to manage the risk by making accurate predictions for an effective early warning system.This project uses the historical seismic data for the prediction of earthquake magnitude using different machine learning models., The aim is to determine the earthquake magnitude with the help of the machine learning algorithms integrated into the PyCaret Framework. Traditional approaches apply basic statistical patterns, which may fail to capture many-factor patterns inherent in earthquake data. In this research, we defined several features,like x-magnitude and then attempted to predict y-magnitude using our model and checked the accuracy of the model. The dependent variables were evaluated with Mean Absolute percentageError , Mean Squared Error , R-squared . The findings were that the Lasso regression performed best, the results are MSE 0.240, MAPE 7.01% and R-square -4.414200827151937e-05, which were more accurate than older ways of estimating. The present work points out that there is hope in developing the application of machine learning to make better forecasts of earthquakes.In future,we aim to add more features and find ways that can help predicting earthquakes for better results.

Keywords-Earthquake prediction, machine learning, early warning systems, PyCaret model,risk reduction, earthquake magnitude forecasting, predictive modeling, disaster mitigation, seismology,stremlit.

Report Organization

1.Introduction: The study establishes its operational objectives and reveals compelling reasons for accurate earthquake prediction with machine learning algorithms.
2.Related Work: The research explores former earthquake prediction approaches and analyzes the adoption of machine learning methods for this field.
3.Research Methodology: The research describes all operational steps starting from data collection through data cleaning until analysis completion and finalizing machine learning models with PyCaret.

- 3.1 Data Collection: The thesis explains the source of data along with its scope and its value in practice.
- 3.2 Data Preprocessing: A description outlines the data cleaning process which made the information ready for analytical purposes.
- 3.3 Exploratory Data Analysis: The analysis identifies data patterns alongside major dependencies holding between different components in the dataset.
- 3.4 Model Development: The article describes the selection process along with training and testing methods used for various machine learning models.

4.Design Specification: The document establishes details regarding the design features for both earthquake prediction algorithms and risk assessment frameworks.
5.Implementation: The section focuses on the devices and programming software that built models for development of the earthquake prediction web application.

5.1 Materials and Equipment Used: The research uses computers together with supporting libraries which appear in the list provided.

6.Results and Evaluation: This section details model testing outcomes with performance comparison followed by an explanation of the web application development for live earthquake predictions.

6.1 Discussion: Highlights the strengths and weaknesses of the results and compares them with other studies.

7.Conclusion and Future Work: The concluding section presents research outcomes and significance alongside proposed ways to advance and develop future applications of the study.

8.References: It presents all research sources which were employed within the investigation.

1 Introduction

The Main focus of the analysis is to utilize machine learning to predict the earthquake magnitude. Earthquakes are a natural calamity, which may be due to the variation in the magnitude, this can cause a lot of massive distraction and may put many people's lives at risk. Earthquake early warning (EEW) is the rapid detection and characterization of earthquakes and delivery of an alert so that protective actions can be taken(Richard M. Allen et al 2019).Despite the fact that machine learning can be used for earthquake magnitude prediction, there exist essential shortcomings in present practices. This brings the main challenge of feature engineering, which incorporates the determination of other useful patterns found in seismic data. This is challenging because there exists a huge correlation between the geophysical processes and the magnitudes of the earthquakes and a lot of these correlations are not linear but may involve a lot of 'noise'. Another problem is overfitting where the model performs well on data that has been trained but doesn't work well where new or distinct data is involved, especially from different geographical locations. Such difficulties make it complex for the development of models that can provide avenues for earthquakes in various locales and settings. These issues should be solved to enhance application of machine learning in developing more effective resolution systems. In this analysis many machine learning models have been used to improve the whole analysis. The main aim of this research is to predict earthquakes as fast and accurately as possible, which can help in preventing human casualties and building collisions(Hisahiko Kubo et al 2020) before the disaster.
Firstly, understand the earthquake patterns by examining the earthquake data that include the mapping of the region where there is a high possibility of an earthquake. This helped to identify the countries where there is high risk and helped me to improve my future prediction. Then the model has been cleaned and the data prepared for analysis by fixing the errors, filling in the missing values and organizing the data to be accurate and suitable for analysis. The data should be clean for getting accurate results in machine learning models. Then for doing overall analysis PyCaret framework has been used. (Ahmad Fadhil Naswir et al 2024)
This framework is used to speed up overall analysis and this framework will help in testing and comparing the machine learning models quickly. The main advantage of this model is that it can quickly identify which model has more potential in this analysis. This analysis has tested

19 different types of machine learning methods and deep learning models to find which model suits the best for earthquake prediction. The model used in this analysis are Extra Trees, Gradient Boosting, Random Forest, AdaBoost, KNN, Light Gradient Boosting Machine, passive Aggressive, Decision Tree, Extreme Gradient Boosting, Linear Regression, Lasso Regression, Elastic Net, Ridge regression, Least Angle Regression, Lasso least Angle Regression, Orthogonal Matching Pursuit, Bayesian Ridge, Dummy Regressor, Huber Regression. All these models are compared and took the best model for the earthquake prediction. These all models have their own advantages as well as some disadvantages, some models are good for small dataset and some others for large dataset and finding the best one from these is complex and choosing the right models for this evaluation is also a bit difficult.
Then comes the tuning that is done for the best performance of the model. The model which performs well in the comparison has been taken out and the analysis has been fine-tuned their settings and the hyperparameters to make sure they perform as well as possible. This is one of the important steps that is to be taken for the accuracy prediction. Then split the data into two which is training and testing 80% for the training of the model rest 20% is for testing the model. This is done for a model for predicting new or unseen data that will be helpful for real-world application. After this model has been evaluated several performance metrics. Those are Mean Square Error which shows the average squared difference between the actual and predicted values. Then Root Mean Square Error this will measure the predicted error, which is easy for to understand, and it shows the result in the same unit of the data. Mean Absolute Percentage Error from this it is easy to understand how accurate the overall models are and it shows the percentage error in the prediction as well. The last performance metrics that was evaluated was R square Error. This explains the performance of the variability in the data if the score is low which means the performance is low and if the score is high the performance is higher. Even created graphs for comparing the performance of each and every model. This visualization will also help in understanding which model is best for the earthquake prediction by understand their strength and weaknesses
Finally, a web app has been created with the best model from the evaluation with the help of streamlite app. Which can show the magnitude range and indicate whether the magnitude is high, low or normal.
By using machine learning techniques this model has proven that predicting earthquake magnitude is possible. My research objective is how machine learning techniques can be integrated to enhance the accuracy and the reliability of earthquake early warning systems. By implementing this more accurate and timely prediction has been made and Many lives can be saved, many can take precautions before the disaster. This study shows that the importance of modern technology in disaster management can help in predicting the earthquake magnitude.

2 Related Work

The research deals with the application of machine learning algorithms in forecasting earthquake magnitude and depth, Key models include; Multiple Linear Regression (MLR) Support Vector Machines (SVMs), Long Short Term Memory (LSTM) neural networks, and the combined models.The dataset used for this analysis is from The US earthquake dataset has

been collected from the United States Geological Survey (USGS) online repository.The accuracy of the model was learned to perfection (R-square =1) during training, but when tested on new data it was moderately low because of over learning. Cross-validation re-validated this through high errors and low performance. The results of the SVM model were slightly higher in terms of accuracy compared to MLR, however, the overall percentage was low and the calculated R-square was far below 0.01 which suggests poor reliability of the current model.
According to the results of the prediction of discontinuity magnitude, the LSTM neural network had the highest accuracy. When training the model, the mathematical coefficient of determination R-square was 0.93; after cross-validation, it decreased to 0.92. MSE also was low and it justifies that on this basis LSTM model learned patterns in data and did not over-fit them.(Fardin Ahmed et al 2024)The results for depth prediction showed that if gradient boosting and neural networks are integrated, hybrid models outperforms the accuracy of individual models.
In general, the LSTM model was identified as one of the best for earthquake magnitude specification. However, one seems to limit the study by comparing only a few models and more comprehensive comparisons with other models could quite likely provide even better results. It was observed that using MLR and SVM in accomplishing this task was less appropriate.

This research study was completed in 2023 to understand and identify the applicability of deep learning models in relation to the identification of earthquake magnitude in early warning systems.The dataset used for this analysis is from National Research Institute for Earth Science and Disaster Resilience, 2019 From the said work, it emphasizes the weakness of the conventional approach in analyzing the seismic wave amplitude and frequency. Traditional methods are good for local application but lack versatility, they take time to provide predictions and are not as accurate at large numbers.This paper analyzes CNN, RNN, both CNN & RNN, and other tailor-made deep learning neural networks focused on making quicker predictions from limited seismic data.(Yanwei Wang et al 2023)These models proved to demonstrate a significant improvement in speed and accuracy over traditional empirical methods. The experiments with large scale earthquake data showed improved prediction accuracy and response time coupled with scalability irrespective of the geographical zone for earthquake activity. MAGNET, ranging from −0.02 to 0.43, EEWNet from 0.28 to 0.67, and r values, 0.83, reveals that EEWNet outperforms Pd in magnitude predictions and is highly effective in real-time warnings, where P-wave durations are often shorter than 0.5s, while Pd works well for 3.0s.
Nevertheless, several limitations of this study are identified as well: the necessity to employ large datasets for deep learning models, overfitting issues at limited data availability, and insufficient preprocessing although such issues are discussed in other studies with large data availability and various machine learning approaches..

This research is done in the year 2021 and it is related to the deep learning in short-term earthquake prediction with SVM, DT, SNN, DNN.The dataset used is January 1973 to July 2019, from USGS1 and IIEES It presents spatial parameters and fault density estimated using the kernel density estimation. The Decision Tree for the data achieved 82% accuracy and DNN achieved 79.6%.(Mohsen Yousefzadeh et al 2021) Considering high-magnitude, DNN was 94.7% accurate as compared to DT. However, SVM found difficulties in the real test as it achieved 66.6% only. These include low sensitivity of detecting small earthquakes due to noise, over fitting in the DNN model, and high computational overhead. Also, as it will be remembered the model was not implemented in real-time either. On the other hand, my study increased accuracy by training models selectively, attaining the overall best efficiency and utilizing streamlit to host the system as a web application.

The analysis is published in 2019 and emphasizes the aspect of machine learning within automatic seismic data processing.The dataset is seismogram data from Venezuelan stations (CUMV, CACV, Funvisis BAUV) for P and S wave identification.(Otilio Rojas et al 2019)As the problem of tools scarcity arises due to the growth in seismic data. Some of the popular techniques such as time and frequency analysis possess weaknesses, which makes it compulsory to use ANNs to acquire efficient results. Three models were used: The four types of deep architectures to be discussed include Feedforward Neural Network (FFNN), Recurrent Convolutional Neural Network (RcNN), Combined CNN and Recurrent Spiking Neural Network (RsNN). The models demonstrated satisfactory performance, the FFNN model had accuracy higher than 90% in detecting P and S waves, and RcNN is capable of detecting small earthquakes with low false alarm rates and CNN models including the ConvNetQuake has 94.8% precision. Some issues are related to a comparison that is not possible to make from a smaller dataset or for model-specific purposes. However, my study was based on a much larger matrix of data of over 300 thousand rows, but with higher precision across the broad range of various seismic zones..

The paper written in the year 2024 investigates the performance of Linear Regression (LR) and K-Nearest Neighbors (KNN) for the purpose of earthquake prediction. They point out that there is a need to have an early warning system towards minimizing the impact and effect on human lives(Sumanth Kalavakunta et al 2024). In the same scenario, the proposed LR model yielded a mean accuracy of 89.05 % with SD 4.062 as compared to KNN with an accuracy of 79.20% and SD 3.853. According to an independent sample t-test, the two samples differ significantly. However the accuracy range is not precise in both the models.As we can see the accuracy of LR model range from 85% to 95% while that of KNN is from 74% to 84%.

This present used dataset with only 10112 features can only make poor predictions and hence the results here are small. In contrast, my analysis is based on over 300,000 rows of a more contemporary data set, thus providing more timely and comprehensive rates in worldwide seismic zones..

The work was undertaken in 2021, which addresses the depth classification of micro-earthquake sources using machine learning algorithms.(De-He Yang et al 2021).The dataset used is from the ISC-EHB Bulletin It particularly targets weak motions of negative intensity to interpret subsurface characteristics, essential in the design of the liquid transport infrastructure. Feature extraction based on seismic templates and other conventional techniques such as template matching are weak in the face of noise and low signals. The results indicated that the SVM, Random Forest, KNN, and SCN are effective depth classification models in combination with PCA. CNN with all the features such as Continuous Wavelet Transform had better accuracies it stood 93.7%. Disadvantages contain, hyperparameters optimization and a non-real-time model. My model has more samples, real-time implementation, and advanced feature extraction, which makes it more accurate and more resistant than the models of this paper. This increases usability and practical applicability for earthquake study and evaluation.

The work has performed in 2022 to analyze the machine learning algorithm for predicting structural seismic impacts and early evaluation for risk diminution.The dataset used from European Strong-Motion Database and Pacific Earthquake Engineering Research Next Generation Attenuation.The analysis classified and evaluated the ground motion intensities, the study compares and finds out models that could predict the progression of damages after any earthquake.(Petros C. Lazaridis et al 2022)The research evaluated the 10 models separately. Etra Tree Regression revealed the highest R-square of 0.87 while GBM provided the highest accuracy of DIDC index cumulative damage prediction. Metrics of evaluation used were Mean Absolute Error, Root Mean Square Error and R-square values computed from 10-fold cross validated results. Unlike this, my analysis used PyCaret to compare 19 models at once which helped in efficient comparison and better scalability. This goes on to show that my proposed model helps in enhancing the prediction accuracy and reducing the computational cost.

The study explores the ability to predict earthquake magnitude with Regression and Convolution Recurrent Neural Network (CRNN). done in 2021,(Asep Id Hadiana et al 2021)it shows that machine learning has a feature of identifying the seismic wave pattern. The study used three models. The ones we are most familiar with include Mag-net, Graph Neural Network (GNN), and CRNN. Comparing the algorithms CRNN achieved the RC-squared of 0.79772 and the lowest MSE 0.1909. Yet, GNN is highly complex to be a real-time model, while to implement the CRNN, requires a lot of computational power.
The dataset was STEAD and eliminating time domain information and averaging consecutive frames converted the raw data into 2D grayscale images. On the other hand, my model leverages segmentation and normalization methods which results in an RMSE of 0.4369. It outperforms their models in terms of prediction accuracy, computational cost and online feasibility.

This study deals with earthquake magnitude estimation and more particularly, the MagNet model. (S. Mostafa Mousavi et al 2019)This is estimated by directly using the raw seismic data to compute the magnitude and MagNet gave a sample mean of nearly 0 with standard deviation of 0.2. It uses convolutional layers only for feature extraction purposes and bidirectional LSTM for temporal analysis. However, there are some shortcomings when using this approach, such as limitations to single-station data, which results in delay when dealing with multis-station

data. This makes the model weak in reading lower frequency waveforms, and fewer training events affecting its use in warning systems.
However, my model is more accurate than MagNet because of the use of feature engineering and increased attention to X_magnitude. My approach also deals with the issue of regional sensitivity and allows for hybrid comparisons, making the approach more flexible and suitably scalable for real-time earthquake prediction.

This work was carried out in 2024 and aims at developing earthquake early warning systems based on machine learning approaches such as LSTM, MLP, XGBoost and LightGBM.(Li Pan et al 2024) The research employs entropy-based CEEMDAN algorithms in simplifying seismic signals and sorting natural and artificial earthquakes. The absolute magnitude was best predicted by the LSTM model with an accuracy of 6.2, yet the LSTM model underestimated other deep learning models . The understanding was done in a sequential manner which means the process could take time. Meanwhile, in my model, The PyCaret framework has been used to compare 19 models at once which are more efficient in their outcomes. Also, my model looks at factors such as magnitude and region, which the current study fails to consider due to its scope. My approach provides a better and more accurate estimate of earthquake magnitude than those used in previous studies.

Table 1 for summary of literature review

Authors & Years	Dataset	Technique/ Algorithm	Evaluation measures	Results	Limitations
Fardin Ahmed et al 2024	The US earthquake dataset has been collected from the United States Geological Survey (USGS) online repository	Multiple Linear Regression Support Vector Machines Long Short-Term Memory Hybrid Models	R-square,Mean square error	MLR-1 SVM-<0.01 LSTM-0.93	only comparing few models that limiting the performance of other better models
Yanwei Wang et al 2023	National Research Institute for Earth Science and Disaster Resilience, 2019	Convolutional Neural Networks Recurrent Neural Networks Combined version of CNN and RNN Custom Neural Networks	Speed and Accuracy, Error Margin Reduction ,Adaptability	EEWNet Mean error: −0.02 Standard deviation: 0.43 Correlation : r=0.83	It may suffer overfitting in smaller data because of deep learning
Mohsen Yousefzadeh et al 2021	January 1973 to July 2019, from USGS1 and IIEES	Feedforward Neural Network,Recurrent Convolutional Neural Network,Combination of Convolutional Neural Network and Recurrent Spiking Neural Network	Standard Deviation of Time Difference	FNN- Detect P and S waves with over 90% accuracy. CNN- ConvNetQuake achieved 94.8% precision	Model-Specific Purposes,Dataset Limitations,Need for Standardized Dataset
Otilio Rojas et al 2019	seismogram data from Venezuelan stations (CUMV, CACV, Funvisis BAUV) for P and S wave identification	Feedforward Neural Network Recurrent Convolutional Neural Network Combined Convolutional Neural Network and Recurrent Spiking Neural Network	Standard Deviation of Time Difference	FFNN:Detect P and S waves with over 90% accuracy CNN:ConvNetQuake achieved 94.8%	It is difficult to compare models and generalize results with smaller data.
Sumanth Kalavakunta et al	from kaggle with 10,112 entries.	Linear Regression Algorithm K-Nearest Neighbors (KNN) Algorithm	Mean accuracy, Standard deviation	LNR mean accuracy- 89.05% standard	It lacks consistent accuracy across the model and in prediction results in both of the models.

2024				deviation-4.062	
De-He Yang et al 2021	The dataset is derived from the ISC-EHB Bulletin	Support Vector Machine,Random Forest,K-Nearest Neighbors,Naive Bayes,Stochastic Configuration ,Logistic Regression, and some others	Accuracy	CWT: 93.70% STFT: 92.44% MFCC: 91.88%	Absence of real time implementation and hyperparameter tuning is not done for CNN
Petros C. Lazaridis et al 2022	European Strong-Motion Database,Pacific Earthquake Engineering Research Next Generation Attenuation	Extra Trees Regression Gradient Boosting Regressor	Mean Absolute Error,Root Mean Square Error,R-square values	R-square:0.87 MAE:~0.05 RMSE: ~0.1	Each of the models has been evaluated individually.
Asep Id Hadiana et al 2021	STanford EArthquake Dataset	Mag-net Graph Neural Network Convolutional Recurrent Neural Network	R-square values, Mean square Error	CRNN R-square:0.79772 MSE:0.1909	Uses GNN model which may be difficult in real time predictions
S. Mostafa Mousavi et al 2019	STanford EArthquake Dataset	MagNet (Convolutional Neural Network + Bidirectional LSTM layers)	Standard deviation, Mean error	SD:near 0 ME:0.2	It works on a single station and does not work on all the stations simultaneously.
Li Pan et al 2024	No dataset mentioned	LSTM MLP XGBoost LightGBM	Accuracy	LSTM:62%	It lacks real time comparisons of models in other deep learning approaches.

3 Research Methodology

3.1 Data Collection

The data is collected from kaggle in CSV format.The name of the dataset is World Earthquake data from 1906-2022.The dataset contains the features such as magnitude,depth, latitude longitude and other relevant columns .This dataset have the earthquake data of different regions and years of more than 300000 rows.

https://www.kaggle.com/datasets/garrickhague/world-earthquake-data-from-1906-2022

	time	latitude	longitude	depth	mag	magType	nst	gap	dmin	rms	...	place	type
0	2023-02-26 23:58:05.052000+00:00	41.8050	79.8675	10.000	5.0	mb	46.0	91.0	1.293	0.80	...	77 km NNW of Aksu, China	earthquake
1	2023-02-26 23:33:17.641000+00:00	18.7420	145.4868	200.365	4.8	mb	67.0	85.0	5.158	0.95	...	Pagan region, Northern Mariana Islands	earthquake
2	2023-02-26 21:42:14.541000+00:00	42.0857	79.9516	10.000	4.9	mb	45.0	77.0	1.223	0.82	...	NaN	earthquake
3	2023-02-26 21:35:01.303000+00:00	14.9364	-104.5563	10.000	4.6	mb	51.0	217.0	5.661	0.57	...	northern East Pacific Rise	earthquake
4	2023-02-26 18:58:54.828000+00:00	44.6730	146.5159	134.299	4.5	mb	108.0	62.0	2.866	0.82	...	84 km NE of Otrada, Russia	earthquake

5 rows × 23 columns

Fig 1: First 5 rows of the dataset

3.2 Data Preprocessing

The dataset contains the raw data.This dataset has gone through several steps to be good for the analysis(Tohari Ahmad et al 2019).The step includes Data cleaning this will remove the duplicate records and it will also handle the missing values.This Process will use mean or median for the numerical values,in this case all the values are numerical.This will also group the column which has lots of missing values. The next process is the feature scaling.This process will use standardscaler to normalise features like depth for improving the model performance.The final process that has been done is the feature selection, which will identify the weakly correlated features using correlation matrix and has remove the most non correlated columns. In this study the non correlated columns removed are nst,gap,magnet and some other columns.

3.3 Exploratory Data Analysis

The main purpose of the EDA is to understand the characteristics of the earthquake dataset, which the EDA will use to identify the trends and analyse the relation between feature and target variable which is magnitude,in this analysis(Mario Li Vigni et al 2013). The EDA will use the mean,median and standard deviation from their minimum to maximum values that are calculated for all the numerical features that includes the magnitude,which is to observe the tendency of the earthquake,depth is to find the distribution of the earthquake. Overall,The latitude and longitude will tell the precise place.

For visualising the earthquake more precisely an external file has been used which contains the spatial patterns of the earthquake.For this GeoPandas has been used for visualising the map. Two maps has been plotted, showing the intensity of the depth of the earthquake in the overall world.The magnitude map displays the location with the latitude and longitude(fig2).The colour represents the magnitude in the map and this map is based on continents.If the markers are bigger the earthquake appears bigger at that place.The colour indicates the intensity of the magnitude.The visualisation of the depth map is also similar to the magnitude but this map will highlights the spatial distribution of smaller vs bigger earthquakes.

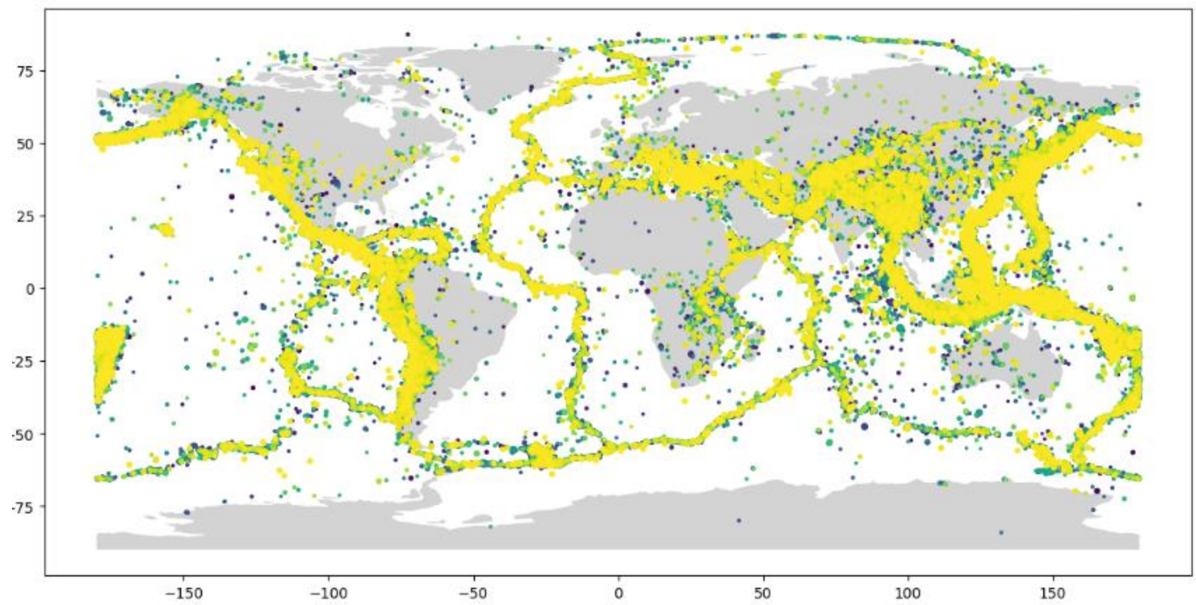
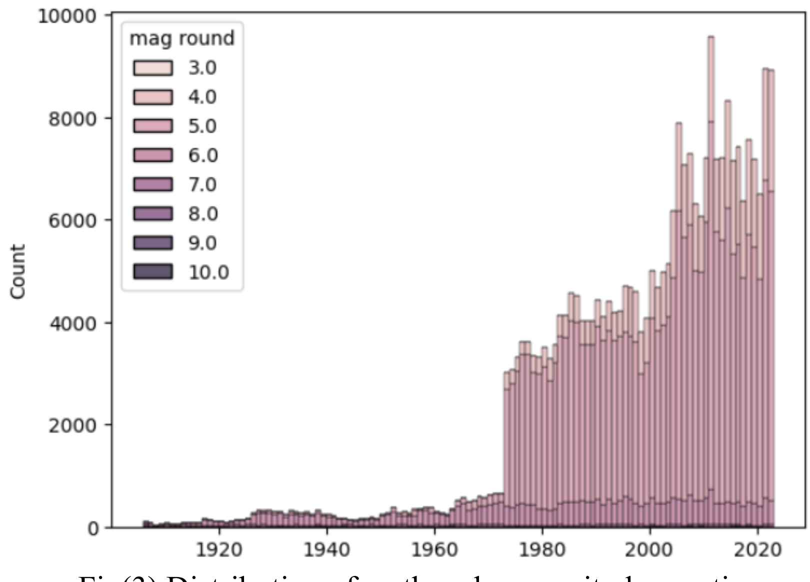


Fig 2 Geographic map of magnitude

The histogram has been created for visualise the earthquake happened the overtime on the basis of the magnitude.In this graph(fig3) it is very clear that, in the past 1920s the earthquake counts was less so doing an early warning system is not that necessary at that time but now the graphs represents that the earthquake count is more than 10000 in the year of 2020 is making an early warning system is necessary(Fanchun Meng et al 2023).The histogram highlights the maground but the intensity of the colour but the maground is almost in the same level from the past to now.



Fig(3) Distribution of earthquake magnitude overtime

3.4 Model Development

For the accurate prediction of magnitude. We have made a systematic approach for the development and evaluation of a variety of regression models(Yanwei Wang et al 2023). The method used is PyCaret framework, which is an automated machine learning framework.This framework is know for the low code interface.The main advantage of using this is modelling, this framework has build in different types of classification and regressions such as linear regression,logistic regression,decision tree,random forest,gradient boosting,KNN,naive bayes and so on.This will also compares each and every model in this framework according to our needs.Hyperparameter tuning will also be done by this model.

The PyCaret regression module has been initialised with the processed earthquake dataset.The magnitude is taken as the target variable for the analysis.The process like normalising numerical features is also done at the beginning. Then PyCaret framework automatically evaluated all the Regression models such as Extra Trees,Gradient Boosting,Random Forest,AdaBoost,KNN,Light Gradient Boosting Machine,passive Aggressive,Decision Tree,Extreme Gradient Boosting,Linear Regression,Lasso Regression,Elastic Net,Ridge regression,Least Angle Regression,Lasso least Angle Regression,Orthogonal Matching Pursuit,Bayesian Ridge,Dummy Regressor, Huber Regression.Then these models has been compared with each other.The linear Regression is baseline model which is good for its performance.The decision tree will capture the linear relationship by splitting to small.The random forest will improve in the accuracy by averaging their prediction.KNN is know for the predict the magnitude based on the patterns of the similar events.The Elastic Net Regression which combines the models like lasso and ridge, good in balancing the features selection and reduces the overfitting by using the large coefficients.Once this is compared this will give an output as a form of table(fig 6) with all the values.

Hyperparameter tuning has been done with the top performing model that has been selected form the comparison of the models and this tuning will be conducted by the PyCaret framework by a simple code (tune_model()).This method contains randomised search to improve the parameters those are learning rate,tree depth and number of estimators.Then the data has been spited for training and testing, 80% for training(fig4) and the rest 20% is for the testing.For model development and tuning training data has been used and for model performance and unseen data test data has been used.

	latitude	longitude	depth	magType	net	id	updated	horizontalError	depthError
97234	-35.3870	-73.3850	29.9	20	18	usp000ha3d	2022-08-09T04:10:23.550Z	NaN	NaN
126154	29.1470	139.3440	415.5	8	18	usp000e7fk	2014-11-07T01:28:02.867Z	NaN	NaN
62329	-20.5511	-173.8695	10.0	8	18	usb000sh45	2014-12-19T00:05:21.040Z	5.9	1.9
20596	-6.5322	153.4455	10.0	8	18	us6000b97i	2020-10-10T17:11:13.040Z	10.8	1.9
280551	-20.2730	-71.2270	15.0	19	5	iscgem908756	2022-04-25T22:54:03.963Z	NaN	25.0
...	...	...	...	...	...	...	...	...	...
119879	46.3580	154.7320	10.0	8	18	usp000eyg2	2014-11-07T01:30:55.412Z	NaN	NaN
259178	14.1290	56.3530	33.0	8	18	usp0000apz	2014-11-06T23:21:29.858Z	NaN	NaN
131932	10.4100	-62.3900	8.5	8	18	usp000djtu	2014-11-07T01:25:17.810Z	NaN	NaN
146867	-21.4280	169.4890	33.0	8	18	usp000bdrg	2014-11-07T01:16:41.498Z	NaN	NaN
121958	-18.9460	-177.8660	628.5	8	18	usp000equb	2014-11-07T01:29:57.429Z	NaN	24.6

226505 rows × 19 columns

Fig 4 Dataset for training

The Performance metrics has been evaluated by the Root Mean Square Error,Mean Absolute Error, R-square value.The RMSE will calculate the average prediction error of the magnitude and it is sensitive to outliers.The MAE will get the absolute error by providing an interpretable measure of the average error. The Final R-square will indicate the proportion of the variance which is explained by the model.The higher value of the R-square will indicate the better performance and the lower the value that shows the performance is lesser.The PyCaret model will compare all their performance and give the output as the best performance model.

Using the best performing model the web app has been created with the help of streamlit(fig9).Once we enter the latitude,longitude,depth,magtype,month and day the app will show the magnitude level and also indicate whether this magnitude is dangerous or not by colours.This has been done by creating repository in github and connecting the github account to the same streamlit account and by deploying it.

This analysis confirms that the models are performed well by achieving its performance metrics.

4 Design Specification

The analysis is used to forecast earthquake magnitudes using most of the earthquake characteristics for instance latitude, longitude, depth, year, month among others and the amount of energy released by an earthquake. The data applied for this study is obtained from highly credible seismic databases. First the data is pre-processed, then unwanted features are removed and more important features for analysis are created. The parameters of interest, the latitude, and longitude, depth, and temporal components, are extracted from the data set for investigation.

In order to explain the earthquake magnitude dependence on geographic location, we use Geopandas for spatial distributions and histograms for temporal distributions computed with Seaborn. It also assists in establishing a relationship between past data and current and future data over a period and space.

Subsequently, to discover the possible correlates of the ‘magnitude’, the machine learning regression models are checked to approximate the target variable as comprised of latitude, longitude, depth, year, month, energy. The accuracy and measure of the built models is done using common regression metrics such as; Mean Absolute Error (MAE), Mean Squared Error (MSE), and Coefficient of determination (R-squared) (R^2). This validation creates many different testing sets to help determine if the several models will accurately predict unseen data. Considering features, feature importance is also calculated and this displays the relevant features which impact most in the prediction.

Once the model with the best set of evaluation metrics is determined, it is then expanded and integrated into a live web application with Streamlit. This web app enables users to enter new seismic data including latitude, longitude, depth and and get an instant earthquake magnitude. Also, application of spatial and temporal visualizations is also included, thus giving the users some of the ways through which earthquake trends can be understood. The realistic graphical user interface style will allow the use of the app for live prediction where it will be valuable for earthquake activity tracking.

5 Implementation

The output produced is that the preprocessed data of the earthquake will be ready to use for the analysis then the machine learning model will be trained and tested by dividing it by 80:20 ratio.This models will be saved using pickles that are used for web app development. We have also done the visualisation of the feature impotence by plotting with the residual analysis.The tools and language used mainly is jupyter notebook, this has been used for all the analysis in the means of development and testing the libraries used for this analysis is Pandas,Scikit-learn,Matplotlib,Seaborn,XGBoost.These are the backbone of the implementation.
The Linear Regression is the main model that helped in providing the baseline performance metrics.The model such as Random forest,Gradient boosting,XGBoost has done the robust prediction with improved accuracy.The neural networks is also been used for exploring the non linear dependencies that will achieve the good results.The baseline of this all models is the PyCaret model,all the model comparison has been done through this model(M Apriani et al 2021), even the hyperparameter tuning has been done with the help of PyCaretl.In final the best model that is good in predicting the magnitude has taken out and that model has been implemented through streamlit pipeline for an efficient earthquake magnitude web app.

5.1 Materials and Equipment used

The Jupyter notebook has been used for iterative development and analysis.The python libraries used for this analysis are Pandas and Numpy,this is used mainly for data manipulation and in the part of preprocessing.the Scikit-learn has been used for model development and evaluation.Matplotlib and Seaborn are used for the visualisation of the data as well as exploratory data analysis of this study.For performance evaluation data scaling like standardscaler have been used.This whole analysis is done using a personal computer,which has eight gigabytes of ram has been used.

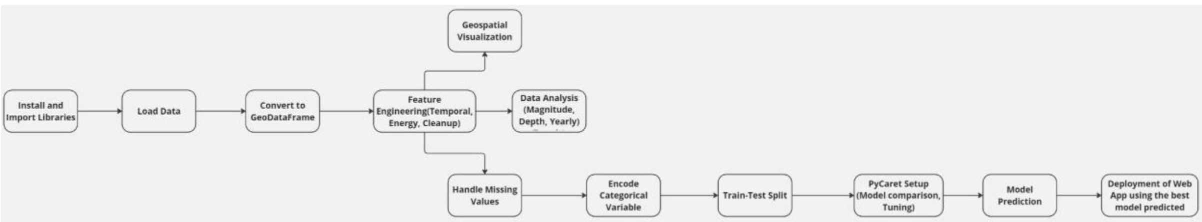


Fig 5 Workflow diagram

The research steps involved in earthquake response are well-defined and structured in order to have accurate research(fig 5). First, basic libraries like Pandas, Geopandas, and PyCaret among others are called and imported as data manipulation, geospatial analysis tools and machine learning tools respectively. The earthquake data set is then read into the program and converted to GeoDataFrame format in order to geographically analyze and visualize them using geographic coordinates. Feature engineering is performed to improve the dataset by converting

temporal attributes into features (year, month) generating new characteristics of the earthquake energy, or removing inconsistent and meaningless variables.
Next, statistical analysis is conducted to employ trends of features such as magnitude, depth and year for earthquake data and geospatial representations to analyze spatial patterns of earthquakes. Data quality issues including missing data are also dealt with to ensure that the resultant data set is complete.As for the categorical variables, an attempt is made to convert them to numerical form for ready association with the machine learning algorithms. The dataset is then split into the training and testing areas in order to make accurate model assessment on new data.
At one place, PyCaret helps in the modeling process and training of models along with the process of comparing models and tuning hyperparameters to attain the best model.Then selected model is applied to make the predicted value in the test dataset and the factors used to assess the set depend on MSE, RMSE, & R^2. Lastly, using the best model with assistance of streamlite the web app has been developed for early warning systemThis workflow offers a detailed approach to perform earthquake prediction study following a step-by-step sequence, highlights on pre-processing data, spatial analysis and model calibration to enhance the robustness and accuracy of earthquake prediction.

6 Results and Evaluation

The main aim of this study is to predict the earthquake magnitude using a machine learning model that is connected to seismic feature data and their target.In the time of analysis ,it has been found that the important features in predicting earthquake magnitude are depth,latitude, and longitude.These features provides very important information about the earthquake occurrence, how strong the earthquake is. which will be the direct things connected to the strength of the earthquake. With the help of the PyCaret farework data processing,feature engineering and model training for optimised prediction performance has been done. The earthquake magnitude prediction analysis has developed and tested for the maximum accurate prediction as well as the partial usability.In the time of data preprocessing many of the features contains lots of the missing value and had lots of similar values.Those were found and removed for avoiding confusion in the model.By doing this the model accuracy has lassie been improved and also overfitting was reduced.

The analysis carried out a comparison of all nineteen regression models namely, Extra Trees, Gradient Boosting, Random Forest, AdaBoost, KNN, Light Gradient Boosting Machine, Passive Aggressive, Decision Tree, Extreme Gradient Boosting, Linear Regression, Lasso Regression, Elastic Net, Ridge Regression, Least Angle Regression, Lasso Least Angle Regression, Orthogonal Matching Pursuit, Bayesian Ridge, Dummy Regressor and Huber Regression. All these regression models were assessed using PyCaret which is an automated framework which will compare all the models at the same time on its own.It also helps in finding the best model by feature selection and comparison The findings were articulated in different evaluation criteria such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), etc.(Fig 6).

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
gbr	Gradient Boosting Regressor	0.1951	0.0819	0.2862	0.6514	0.0436	0.0375	45.7700
et	Extra Trees Regressor	0.2066	0.0834	0.2888	0.6446	0.0443	0.0401	4.3700
rf	Random Forest Regressor	0.1860	0.0844	0.2905	0.6406	0.0436	0.0353	11.2500
ada	AdaBoost Regressor	0.1731	0.0907	0.3008	0.6136	0.0454	0.0321	16.2567
knn	K Neighbors Regressor	0.2005	0.0990	0.3146	0.5786	0.0482	0.0379	10.0367
lightgbm	Light Gradient Boosting Machine	0.2931	0.1612	0.4015	0.3134	0.0633	0.0579	2.7167
dt	Decision Tree Regressor	0.2752	0.1832	0.4250	0.2209	0.0660	0.0525	3.2100
par	Passive Aggressive Regressor	0.3256	0.1890	0.4348	0.1952	0.0683	0.0633	2.1200
xgboost	Extreme Gradient Boosting	0.3431	0.2299	0.4795	0.0212	0.0755	0.0661	2.2033
ridge	Ridge Regression	0.3583	0.2349	0.4846	0.0001	0.0766	0.0696	1.9133
lr	Linear Regression	0.3583	0.2349	0.4847	-0.0000	0.0766	0.0696	2.0967
en	Elastic Net	0.3583	0.2349	0.4847	-0.0000	0.0766	0.0696	2.0033
lasso	Lasso Regression	0.3583	0.2349	0.4847	-0.0000	0.0766	0.0696	1.9367
lar	Least Angle Regression	0.3583	0.2349	0.4847	-0.0000	0.0766	0.0696	1.9700
llar	Lasso Least Angle Regression	0.3583	0.2349	0.4847	-0.0000	0.0766	0.0696	2.0033
omp	Orthogonal Matching Pursuit	0.3583	0.2349	0.4847	-0.0000	0.0766	0.0696	1.9533
br	Bayesian Ridge	0.3583	0.2349	0.4847	-0.0000	0.0766	0.0696	2.1733
dummy	Dummy Regressor	0.3583	0.2349	0.4847	-0.0000	0.0766	0.0696	1.7533
huber	Huber Regressor	0.3607	0.2538	0.5035	-0.0809	0.0774	0.0701	5.1533

```
Processing:   0%|          | 0/81 [00:00<?, ?it/s]
```

Fig 6 Evaluation result of each model

The models which are performing better have been hyperparameter tuned one by one with the help of PyCaret framework in a single line code.The main purpose of hyperparameter tuning is to find the best combination of hyperparameters of the model.By doing over 10 cross validation this model has ensured the selection of hyperparameter that has performed consistently across the dataset.The things helped in the performance of the model but it did not made any significant overfitting issue in some model.The output metrics include MAE,MSE,RMSE,R-square, RMSLE and MAPE.These mesearses has given the detailed feedback on each model performance on each candidate that allows to find the best performing hyperparameter combinationThese models has also been gone through 10-fold cross validation where the dataset has been splitted into 10 parts.The 9 folds has been used for the training purpose and the rest 1 fold is for the validation purpose.The same process will be repeated 10 times with each fold has used once for validation.This process has done on all the better performing models,the output of one of the model which is tuned lasso is in figure 7.

From that some of the model does not performs well in the tuning, which shows in the output that the model where better without the tuning so the model will less performance have removed and selected some of the other models which has done better in tuning those are Randomforest,LightGBM,LinearRegression,RidgeRegression,LassoRegression,XGBoost,AdaBoost and Decision tree,these all models are tuned.There is no model which perform good without tuning in this analysis.While performing the models a warning message in the LightGBM has shown up, which was not a error.The warning indicates potential conflicts in the hyperparameter tuning,which has been removed at the same time.This warning was because of the overlapping of this model,after resolving the warning by narrowing the rage the model became normal.Once after the tuning the testing has been done. On the basis of that these models have been selected.

```
lasso = create_model('lasso')
```

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.3567	0.2343	0.4841	-0.0001	0.0765	0.0692
1	0.3578	0.2347	0.4844	-0.0000	0.0766	0.0695
2	0.3590	0.2370	0.4868	-0.0000	0.0768	0.0697
3	0.3569	0.2313	0.4810	-0.0000	0.0762	0.0694
4	0.3600	0.2387	0.4885	-0.0000	0.0771	0.0698
5	0.3584	0.2314	0.4811	-0.0001	0.0762	0.0698
6	0.3572	0.2324	0.4820	-0.0000	0.0763	0.0695
7	0.3565	0.2316	0.4812	-0.0001	0.0762	0.0694
8	0.3581	0.2334	0.4831	-0.0000	0.0764	0.0696
9	0.3621	0.2443	0.4943	-0.0001	0.0778	0.0701
Mean	0.3583	0.2349	0.4846	-0.0000	0.0766	0.0696
Std	0.0017	0.0039	0.0040	0.0000	0.0005	0.0003

```
Processing:   0%|          | 0/4 [00:00<?, ?it/s]
```

Fig 7 Tuned lasso output

From those selected models, we have again evaluated for a better clarity of the models. By the evaluation we got to know that the advanced models like Random forest,LightGBM and XGBoost perform worse than normal disputes their complexity.That shows that these model have overfitting, poor parameter tuning or because of lack of features to support these models.Then the Linear Regression and Ridge Regression has been performing well in terms of MSE,RMSE and MAPE but not in the R-square value,this value is slightly lower than the best performing model

The Lasso Regression model was the best performing model among all these models which has very less MSE of 0.240 and the RMSE of 0.490 that make this model best in terms of error minimizing.Lasso Regression also achieve the lowest MAPE that is the accuracy of the prediction which is 7.01% error on average that means the prediction is closest to the actual values in relative term.All of the models R-square values where negative that indicated the models where performing worse than predicting the mean.Even though Lasso regression has the highest least negative R-Squared value that is -4.414200827151937e-05 that shows that this model performs closest to useful model that when compared to other models.Overall Lasso Regression has the lowest MSE and RMSE and the highest accuracy(MAPE) even the R-squared value is negative this model outperforms the other models which shows that this model is string compared to other model in predicting the earthquake magnitude.

Additionally,There are some reason that Lasso model outperformed well compared to other models, which is this model includes the feature selection by shrinking the irrelevant coefficient to zero which help in making it zero of irrelevant or redundant features in the dataset.Even this model avoids the overfitting which make it robust even with limited data.

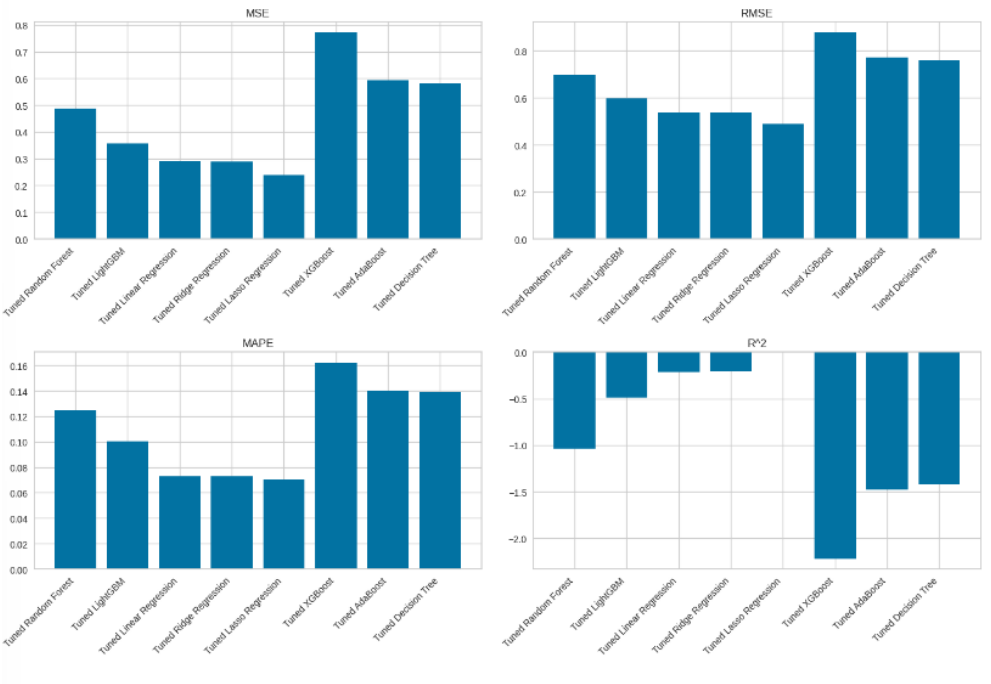


Fig 8 Performance of each model

The metrics use for finding the performance of each mode are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and R-Square these highlights the significance and effectiveness of each model(fig8).The model lasso regression has demonstrated the best performance by achieving the lowest error values and highest R-Square value.this proves that the lasso model is suitable for reliable and more accurate prediction and this model suits the best for the dataset.The main reason is the ability of the model to capture the underlying relationship in the data.

In addition,tree based models such as tunes XGBoost,AdaBoost and Decision Tree has shown significantly low performance compared to other models.These models has the negative R-Squared values and high error values.These model fails to explain the variability in the data and has proven that these models are less reliable for the prediction with this data.Among these the model tuned XGBoost shows the highest MSE and RMSE .That indicates the larger deviation between the predicted and actual values.

The best model that is Lasso Regression has been saved by a pickled file.Using this model.A web app has been created for the real time prediction of earthquake magnitude.This has been done with the help of streamlite.The best model has been picked and done the coding for the web app which then copied to github, made a repository that connected to deploy the code in streamlit

The input parameters of the web app such as latitude,longitude,depth,magnitude type which is a dropdown column,month and day.The risk categories has been divided in to 3 that is low risk when the magnitude is below 4.0,the moderate risk when the magnitude is between 4.0 and 6.0.The last one is high risk above the magnitude of 6.0 and this will indicted in colour too.This application has been tested in edge case to ensure the robust and handling of input variables.The extreme input of 700 km of depth or unusual latitude and longitude values.The application predicted without any errors that shows the flexibility and haindling diverse data of the application.Even with the missing inputs the prediction has been made by this application but the prediction where less accurate.

This application performed consistently across different scenarios.The Streamlit interface(fig 10)helps the users to put the input easily and view the result in the real time.The response time of this application is actually less than one second which is suitable for real time testing. An example prediction in this web app for ensuring the smooth and fast run has been done(fig9).For the input parameter,the latitude as 34.05,longitude as -118.25,depth of 10 km,magnitude as ML then the month is 5 and the day is 15 The model predicted the magnitude as 4.94 and this values was categorised as moderate risk represented in blue colour.This prediction was aligned well with expected threshold.The link for the web app is given below.

https://earthquake-magnitude-prediction-axs7w3lt3hwvryxzwv6w3b.streamlit.app/

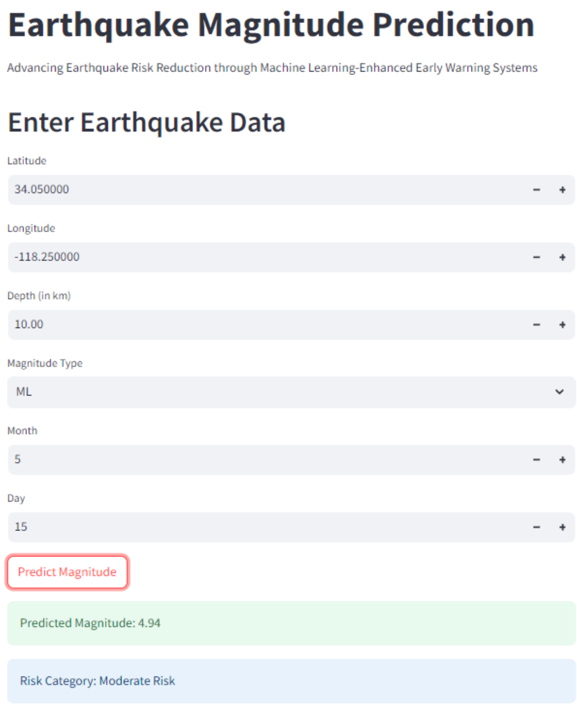


Fig 9 Interface of web

This application is efficient and accurate for earthquake magnitude prediction and risk assessment.For the future enhancement this can be built as an visualising prediction on a map.This can also integrate additional features for the prediction and also can improve the utility of this application.This application can also be connected to real time seismic waves monitoring websites with some slight changes in the analysis that results this to give alarm when there is an earthquake in any specific region.

6.1 Discussion

In this study, we have aimed to predict the earthquake using a simple PyCaret farmwork with the help of features and target in this case the feature is x_magnitude and target is y_magnitude.This study shows the both strength and the weakness of the approach.
The key strength of this application shows that the machine learning is capable of predicting the earthquake in an efficient manner.The use of PyCaret framework shows the robustness of the analysis and also analyzes the relationship between the key features and target column.This approach has also developed a early warning system in the form of web app and can predict the magnitude in lease than one second which will potentially improve the disaster management and risk reduction.

However there are some limitation to this study.The dataset used for this analysis does not have sufficient large or diverse data because of the limited hardware specification.That leads to the limitation in the generalization of my findings.In addition the performance of the model has been influenced by the chosen model and the steps used for preprocessing the data.In future studies it is possible integrate more advanced datasets and can do more advanced feature engineering for more accurate prediction.
In comparison with the other studies,my findings were better with the prior research that highlights the potential in the data driven approaches in earthquake prediction.There are even differences in the experimental setups and dataset used compared to other studies.This proves the need for a standardized benchmark in the field.Apart from the limitations, this approach has more more different knowledge in machine learning application.This also have the option in the future to enhance the exploration and refinement in predictive models.

7 Conclusion and Future Work

This research has shown the usage of machine learning in specific the PyCraret framework to predict the earthquake magnitude.The main goal is to see, how can machine learning techniques be integrated to enhance the accuracy and the reliability of earthquake early warning systems.Which are very important for reducing the earthquake risk and saving lives.
The Results showed that machine learning is capable of finding the earthquake magnitude in real time. Magnitudes are the main reason for the earthquake using the magnitude the earthquake severity can be calculated.By using the best model to build the real time web app this model has able to find the magnitude of the earthquake within less than a second.This is the tool that the people can use for finding the earthquake in real time.
On the other hand,The study has some limitations that can be improved; those are the dataset does not contain the large diverse data for predicting the full range earthquake patterns and also this analysis accuracy can be improved by better feature engineering and by using advanced techniques.The web app that has been developed will only give the result when we manually put the inputs but this can be connected with the real time seismic monitoring stations for predicting the magnitude automatically and give the alarm of the earthquake when the magnitude rises.By addressing these challenge in future this model can find the best and accurate earthquake prediction.
In conclusion,Even this study does not provide a complete solution for the earthquake prediction but this study provides a foundation and valuable insights that can be improved in the future researchers.By enhancing this finding the researchers can build better performing models that will help to protect the communities from the danger of earthquakes.This work has proved that the importance of the technology in disaster risk reduction.

References

Ahmed, F., Akter, S., Rahman, S.M., Harez, J.B., Mubasira, A. and Khan, R., 2024, March. Earthquake Magnitude Prediction Using Machine Learning Techniques. In *2024 IEEE*

Ahmad, T. and Aziz, M.N., 2019. Data preprocessing and feature selection for machine learning intrusion detection systems. ICIC Express Lett, 13(2), pp.93-101.

International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) (Vol. 2, pp. 1-5). IEEE.

Allen, R.M. and Melgar, D., 2019. Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Review of Earth and Planetary Sciences*, *47*(1), pp.361-388.

Apriani, M. and Wijaya, S.K., 2021, June. Earthquake magnitude estimation based on machine learning: application to earthquake early warning system. In *Journal of Physics: Conference Series* (Vol. 1951, No. 1, p. 012057). IOP Publishing.

Hadiana, A.I., Sukma, R.M. and Putra, E.K., 2024. Advanced Earthquake Magnitude Prediction Using Regression and Convolutional Recurrent Neural Networks. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *8*(4), pp.571-578.

Kalavakunta, S. and Parthipan, V., 2024, May. Natural Disaster Earthquake Prediction using Linear Regression Algorithm Comparing with K-Nearest Neighbors Algorithm. In *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)* (pp. 51-54). IEEE.

Kubo, H., Kunugi, T., Suzuki, W., Suzuki, S. and Aoi, S., 2020. Hybrid predictor for ground-motion intensity with machine learning and conventional ground motion prediction equation. *Scientific reports*, *10*(1), p.11871.

Lazaridis, P.C., Kavvadias, I.E., Demertzis, K., Iliadis, L. and Vasiliadis, L.K., 2022. Structural damage prediction of a reinforced concrete frame under single and multiple seismic events using machine learning algorithms. *Applied Sciences*, *12*(8), p.3845.

Meng, F., Ren, T., Liu, Z. and Zhong, Z., 2023. Toward earthquake early warning: A convolutional neural network for rapid earthquake magnitude estimation. *Artificial Intelligence in Geosciences*, *4*, pp.39-46.

Mousavi, S.M. and Beroza, G.C., 2020. A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), p.e2019GL085976.

Naswir, A.F. and Fahmi, H., 2024. Feature Importance and Binary Classification using PyCaret. *Journal of Data Science, Technology, and Artificial Intelligence*, *1*(1), pp.29-32.

Pan, L., Liu, M., Chen, R. and Ma, S., 2024, August. Research on Seismic Signal Identification and Magnitude Prediction Model Based on Sample Entropy and Machine Learning. In *2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE)* (pp. 1586-1592). IEEE.

Rojas, O., Otero, B., Alvarado, L., Mus, S. and Tous, R., 2019. Artificial neural networks as emerging tools for earthquake detection. *Computación y Sistemas*, *23*(2), pp.335-350.

Vigni, M.L., Durante, C. and Cocchi, M., 2013. Exploratory data analysis. In *Data handling in science and technology* (Vol. 28, pp. 55-126). Elsevier.

Wang, Y., Li, X., Wang, Z. and Liu, J., 2023. Deep learning for magnitude prediction in earthquake early warning. *Gondwana Research*, *123*, pp.164-173.

Yousefzadeh, M., Hosseini, S.A. and Farnaghi, M., 2021. Spatiotemporally explicit earthquake prediction using deep neural network. *Soil Dynamics and Earthquake Engineering*, *144*, p.106663.

Yang, D.H., Zhou, X., Wang, X.Y. and Huang, J.P., 2021. Mirco-earthquake source depth detection using machine learning techniques. *Information Sciences*, *544*, pp.325-342.