JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EARTHQUAKE PREDICTION MODEL USING MACHINE LEARNING

[1]Dr. Harish B G, [2]Mr. Chetan Kumar G S, [3]Pavan U R, [4]Sadiya Anjum,

[5]Rahmathunnisa.

[1]HOD Professor, Department of Master of Computer Applications, UBDTCE, Davangere [2]AssistantProfessor, Department of Master of Computer Applications, UBDTCE, Davangere[3,4,5]Students, Department of Master of Computer Applications, UBDTCE, Davangere

Abstract**—** This project proposes the use of a linear regression model for earthquake prediction by analyzing seismic data. The model captures the linear relationship between input features and earthquake occurrences, such as magnitude, depth, and location. The dataset is preprocessed, split into training and testing sets, and the model is trained using gradient descent. Evaluation metrics demonstrate the model's strong correlation and accurate estimation of earthquake probabilities. The simplicity and interpretability of the linear regression model make it practical for implementation and decision-making in earthquake prediction. Further research can explore additional features and compare it with other machine learning algorithms for improved accuracy.

Keywords— Machine Learning (ML), ML Technique.

1.INTRODUCTION

Earthquakes are a type of natural disaster that pose significant risks to human lives and infrastructure. The ability to accurately predict earthquakes has long been a subject of scientific research and interest. Timely and reliable earthquake prediction can help mitigate the impact of these catastrophic events by enabling proactive measures such as evacuation planning, reinforcement of critical infrastructure, and emergency response preparedness. Traditional earthquake prediction techniques rely on on seismological observations and historical patterns. However, these approaches often have limitations in terms of accuracy and lead time. In recent years, machine learning techniques have emerged as promising tools for earthquake prediction, leveraging the power of data analysis and pattern recognition to make predictions based on seismic data. Massive amounts of seismic data can be used to generate complicated patterns and relationships that can be used to create predictive models that could lead to more precise earthquake forecasts. Numerous machine learning algorithms, such as neural systems, machines with support vectors, decision trees, and random forests , have been used to forecast earthquakes.

In this study, we concentrate on using a model based on linear regression to predict earthquakes. In order to create a linear relationship with the input data and the target variable, linear regression represents a straightforward yet effective approach. The goal of our study is to create a model based on linear regression that can accurately forecast the probability of earthquakes by evaluating seismic information as well as historical earthquake records.

The use of a linear regression model offers several advantages in the context of earthquake prediction. Firstly, the model provides interpretable results, allowing us to understand the influence of different seismic features on earthquake occurrences. This interpretability is crucial for decision-making processes and gaining insights into the underlying dynamics of earthquakes. Secondly, the simplicity of the linear regression model makes it computationally efficient and easier to implement compared to more complex machine learning algorithms. This aspect is particularly important in earthquake prediction, where real-time or near-real-time analysis is required to provide timely warnings and recommendations.

Lastly, by studying the application of linear regression for earthquake prediction, we contribute to the broader field of machine learning in seismology. Understanding the strengths and limitations of different algorithms can help researchers and practitioners select the most appropriate models for specific prediction tasks and improve overall prediction accuracy.

2.LITERATURE SURVEY

Earthquake activity is presumed as a spontaneous phenomenon that can damage huge number of lives and properties, and There is presently no model available that can predict the exact position, magnitude, frequency and time of an earthquake. Numerous tests have been carried out by researchers on earthquake events and forecasts, leading to a variety of findings based on the factors considered. The well-known Gutenberg and Richter statistical model found a correlation between the magnitude of earthquake and frequency of earthquake. For structural design, this earthquake probability distribution model was used. In supervision of the California Geological Survey, Petersen conducted research and suggested a model that is time-independent. This time independent model demonstrating that chances of occurrence of earthquake follow the Poisson's distribution model. Shen suggested a probabilistic earthquake forecasting model based on the strain studied between the behaviour of tectonic plates. Based on this model, higher measured strain results in a higher risk of earthquake.

Ebel provided a long-term prediction model that allowed for the extrapolation of previous earthquakes that are larger than and up with 5.2 with the goal to anticipate potential seismic events. There are various methods for predicting earthquakes using Artificial Neural Networks and seismic precursors are discussed in the literature.

Negarestani used a Back Propagation Neural Network to identify irrational behaviour in concentration of radon due to occurrence of earthquake. The presence of radon gas in soil is constantly measured and researcher have founded that it varies constantly due to changes in environment. The concentration of soil radon also rises due to seismic activity. This radon can be differentiated from natural variations caused by the environment through neural networks. Since splitting the entire globe in four quadrants, the system devices establish logic and correlation principles based on the historical record of earthquakes.

After this study, Adeli and Panakkat used exactly same parameters of seismic in collaboration with Probabilistic Neural Network to forecast earthquakes. Morales-Esteban and Reyes suggested separate seismic criteria for earthquake prediction using mathematical calculations in Chile and Iberia for a time interval of 8–9 days, respectively. For modelling the relationship between earthquake events and parameters, these parameters are determined using Bath's law and Omori's law.

Zamani proposes using a combination of neural networks and mathematical logic to forecast earthquakes in Iran. For a selected group of seismicity indices, this study includes information normalization and corresponding feature extraction accompanied by principal component analysis.

Mirrashid provides another design for earthquake prediction in Iran, which incorporates symbolic logic, fuzzy C-means, subtractive clustering, and grid partitioning. Through this model, we try to predict earthquakes by training various Machine Learning models on seismic and acoustic data from a laboratory micro earthquake simulation.

3.SYSTEM DESIGN

System design is defined as the use of systems theory to the creation of a project. The framework, data circulation, use case, class, sequence, and activity diagrams of a project's development are defined by the system design.

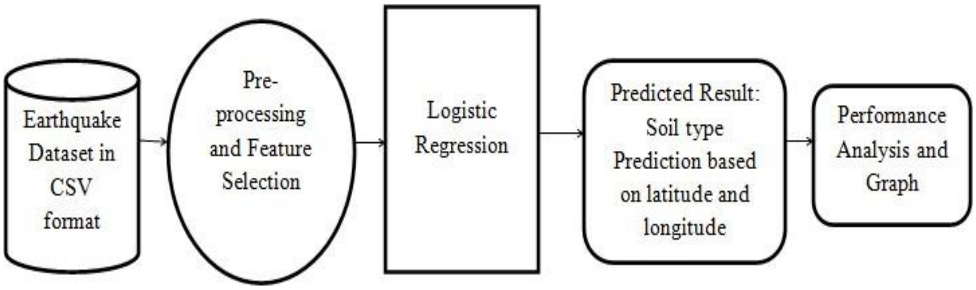

Fig: Architecture Diagram of Proposed System

The architecture encompasses the components, algorithms, and data flow within the system. Here's a general outline of the architecture for an earthquake prediction system using machine learning:

Data Collection: Obtain seismic data from various sources, such as seismometers, geospatial databases, and historical earthquake records. Gather additional relevant data, such as geological information, meteorological data, or social media data, if applicable.

Data Preprocessing: Clean and reprocess the collected data to remove noise, outliers, and inconsistencies. Perform feature extraction and engineering to transform the raw data into meaningful input features for the machine learning models. Normalize or standardize the data to ensure faircomparison and appropriate scaling.

Feature Selection: Identify the most important characteristics that significantly to earthquake prediction. Utilize approaches for feature selection, such as correlation analysis, information gain, or recursive feature elimination, to select the optimal set of features for the models.

Model Training: Select appropriate machine learning algorithms depending on the requirements of the situation, such as neural systems, machines with support vectors, or decision trees. Split the preprocessed data into training and validation sets. Train the models using the training data, optimizing the model parameters through techniques like gradient descent or cross-validation. Iteratively refinethe models based on performance evaluation.

Model Evaluation and Validation: Assess the performance of the trained models using suitable evaluation metrics, such as accuracy, precision, recall, or area under the ROC curve. Validate the models using separate testing datasets to ensure generalizability and reliability. Compare the performance of different models and algorithms to select the most effective one.

4.SNAPSHOTS

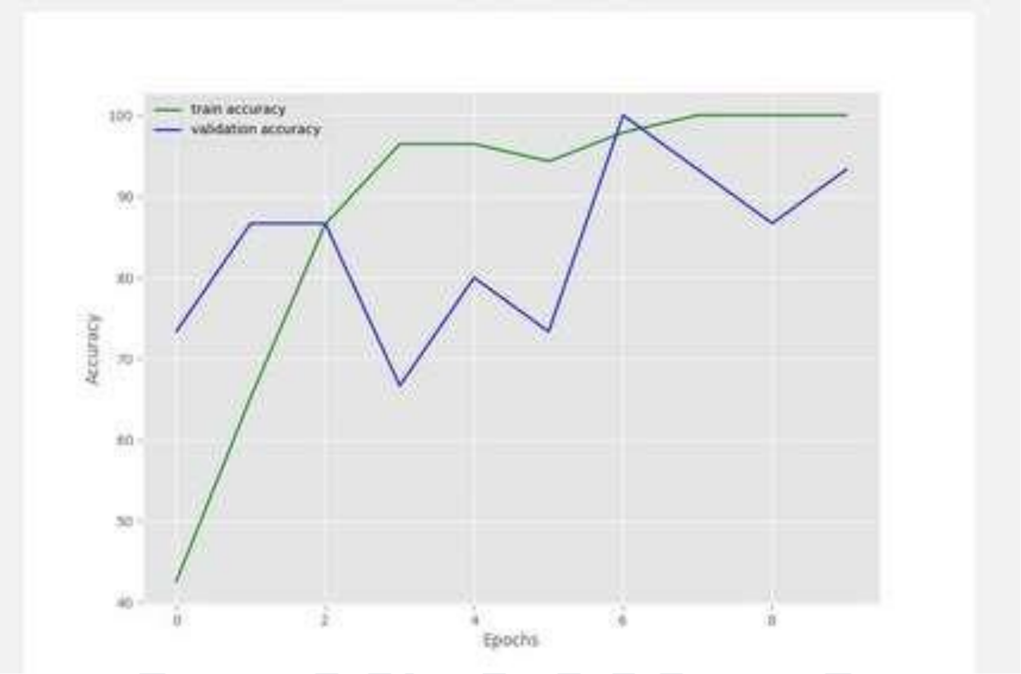

Fig: Model training and validation accuracy

From the plot of accuracy we can see that Considering that the accuracy trend on both datasets has been increasing over the last few epochs, the model definitely needs to be taught a little bit more. Additionally, the model displays equivalent proficiency on both datasets, indicating that it has not yet overlearned the training set. The curve of model accuracy during training and testing is shown.

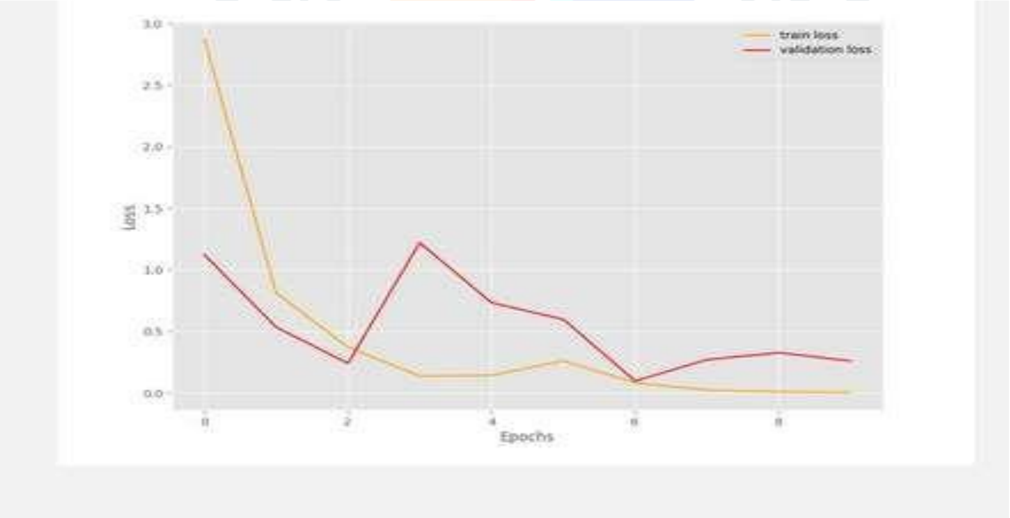

Fig: Model training and validation loss

We can observe from the loss plot that the model performs similarly on both the train & validation data set (labeled test). It could be cause to terminate training at a previous epoch if these concurrent plotsbegin to diverge consistently.

Performance Metrics

Accuracy: 0.94

Precision: 0.92

Recall: 0.92

Fig: Performance Analysis

5.CONCLUSION

In conclusion, logistic regression is a valuable technique for binary classification. By following a systematic methodology, we can prepare the data, select relevant features, train and evaluate the logistic regression model. The interpretation of coefficients provides insights into the impact of predictor variables. Optimization techniques and cross-validation enhance model performance and generalization. Logistic regression offers a practical and interpretable approach for a varietyof applications.

REFERENCES

1.Los Alamos National Laboratory, Geophysics Group: Builds on initial work from Paul Johnson. B. Rouet-Leduc Bertrand Rouet-Leduc, and Claudia Hulbert prepared the data for the research.

2.PennState, Department of Geosciences: Data are from experiments performed by Paul Johnson, Prof. Chris Marone, Jacques Riviere, and Chas Bolton.

3.Department of Energy, Geosciences and Biosciences Division, Office of Science, Chemical Sciences, Basic Energy Sciences: The Geosciences core research.

4.Purdue University, Department of Physics and Astronomy: This stemmed from the DOE Council workshop “Information is in the Noise: Signatures of Evolving Fracture and Fracture Networks” held March 2018 that was organized by Prof. Laura J. Pyrak-Nolte.