IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125196

Earthquake Prediction using Machine Learning

Dr. Chethan H K[1], Mr. Harsha P T[2], Mr. Mahendra R[3], Mr. T Ram Gopal Reddy[4], Ms. Yashaswini K C[5]

Professor, Dept. of Computer Science and Engineering, Maharaja Institute of Technology, Thandavapura[1]

Students, Dept of Computer Science and Engineering, Maharaja Institute of Technology Thandavapura[2,3,4,5]

Abstract: In recent years, Earthquake prediction has grown more crucial in recent years, spurred by the pressing need to reduce the ruinous impacts of natural disasters on populations and infrastructure. This paper presents a machine learning-based method that uses past seismic records to predict seismic events as one of three events: earthquake warning, explosion, or no earthquake. The categorization is done on the basis of prominent features like magnitude, depth, root mean square (RMS), and depth error. Various models were trained and validated, i.e., Random Forest, Decision Tree, and XGBoost, with the highest prediction performance shown by XGBoost. For instances labeled as earthquake warnings, a K-Means clustering algorithm is also used to identify the severity level—Minimum, Moderate, or Severe. In order to achieve interpretability and reliability of the model, LIME (Local Interpretable Model-Agnostic Explanations) is implemented and makes each prediction comprehensible, and the intuitive user-friendly web application developed with Flask enables end users to provide seismic parameters to generate real-time and transparent results. The framework went through robust unit, integration, and acceptance testing, establishing confidence in reliability as well as usability. Generally, this solution provides a strong and interpretable means of early earthquake detection that can aid in improved preparedness and response approaches.

Key Words: Earthquake Prediction, XGBoost, Random Forest, K-means Clustering, LIME (Local Interpretable Model-Agnostic Explanations), Flask Application, Severity Classification.

1.INTRODUCTION

Earthquakes are unforeseeable natural disasters that tend to have devastating impacts on human lives, infrastructure, and economies. Early warning systems are essential to reduce these effects, but traditional seismic monitoring technologies often fail to deliver quick, precise, and interpretable data. With advances in artificial intelligence, especially machine learning, it is now possible to process vast amounts of seismic data from the past to determine patterns and make educated predictions. This project offers a machine learning-based solution to predicting earthquakes where seismic activity is classified into any of three categories: earthquake warning, explosion, or no earthquake. Through the analysis of parameters like magnitude, depth, root mean square (RMS), and depth error, the system is expected to provide timely predictions with increased reliability. In order to increase the practicability of this system further, a Flask-based web application has been created, allowing the input parameters to be entered and real-time results in an interpretable manner to be provided.

1.1 OBJECTIVE

The main goal of this project is to create a precise and interpretable machine learning model that can forecast seismic activity based on real-world data. It should not only label events as earthquakes but also separate them from explosions and meaningless severities. Furthermore, in situations where an earthquake warning is initiated, the system should determine its severity through clustering methods to classify it as minimum, moderate, or severe levels. The use of explainable AI (XAI) tools such as LIME guarantees that the user can comprehend why the model's predictions are made. The ultimate objective is to create an accessible and scalable interface supporting real-time decision making and disaster preparedness.

1.2 MOTIVATION TO TAKE UP THE PROJECT

The inspiration for the project stems from the critical international need to enhance the precision and functionality of earthquake detection systems. In most vulnerable areas, the conventional seismic monitoring infrastructure is non-existent or without predictive functionality. Through the application of machine learning, we can tap into patterns in past data that may not be apparent via manual analysis. Increasing strength of computational models provides an opportunity to enhance early warnings and response plans, which can save lives and minimize losses. Additionally, coupling such

DOI: 10.17148/IARJSET.2025.125196

models with interpretable systems and human-friendly interfaces gives even non-specialist users the ability to work with and trust the forecasts, leading to increased adoption and social influence.

1.3 CHALLENGES TO BE ADDRESSED

Throughout development, there were a number of technical and practical issues encountered. Preprocessing and feature selection were necessary for handling noisy or missing seismic data to ensure that only trustworthy inputs were utilized in training. A second significant issue was maintaining high prediction accuracy while not compromising interpretability; numerous high-performing models are black boxes, which was addressed by incorporating LIME for explainable outputs. Discerning between similar explosion and earthquake patterns also created classification difficulties, which were solved by employing strong ensemble algorithms such as Random Forest and XGBoost. Lastly, implementing the complete solution on a web-based application required cautious backend integration, session handling, and user interface to ensure responsiveness and usability. These challenges were systematically overcome to develop a stable, accurate, and interpretable earthquake prediction system.

1.4 Random Forest and XGBoost

Random Forest and XGBoost are two popular machine learning algorithms that are renowned for their strength, precision, and flexibility in dealing with structured data. Random Forest is an ensemble learning method that constructs many decision trees while training and predicts the class that is the mode of predictions of all the individual trees. It minimizes overfitting by aggregating the output of many weak learners (decision trees), thus enhancing generalization and performance. As it can process high-dimensional data sets and noisy data effectively, Random Forest is especially suited for real-world data classification problems such as seismic activity.

XGBoost (Extreme Gradient Boosting) is a powerful boosting algorithm that constructs decision trees sequentially, wherein each subsequent tree tries to improve on the mistakes of the earlier ones. It is an optimized version of gradient descent and has regularization parameters to avoid overfitting and is thus both accurate and computationally efficient. XGBoost allows for parallelism and internal handling of missing values, which contributes significantly towards its efficiency on large datasets. Random Forest as well as XGBoost were used and compared in the present study for classifying seismic events. Whereas Random Forest had produced consistent baseline outputs, XGBoost performed better in predictive precision and speed to emerge as the model of choice for final rollout in the earthquake prediction system.

1.5 LITERATURE SURVEY

Over the past few years, machine learning methods have gained popularity for earthquake forecasting, providing a data-based alternative to classical geophysical models. [1]Shrote et al. (2024) introduced a model using Random Forest, LightGBM, and XGBoost classifiers to predict seismicity. Their model attained an accuracy of 0.69, precision of 0.63, and an F1-score of 0.65. Even though the ensemble approaches were found to have reasonable performance, the authors did recognize one significant limitation—the model could not predict the real impact or effect of the earthquake, which is essential for disaster preparedness and mitigation.

Along the same lines, [2]Babu et al. (2024) created an earthquake forecast model based on Random Forest and Gradient Boosting techniques. The model produced a score of 0.6967 for Random Forest Regressor and 0.7037 for Gradient Boosting Regressor, reflecting slightly better performance. Nevertheless, the research also pointed to the increasing problem of managing large-scale and intricate seismic datasets, underlining the necessity for scalable machine learning applications.

[5]Kumar et al. (2023) performed a comprehensive analysis on Random Forest and Support Vector Machine (SVM) classifiers for predicting seismic events. Their proposed model attained an impressively high F1-score of 0.965, indicating excellent prediction ability and classification accuracy. However, the work was confined to event classification alone and did not investigate severity estimation or model interpretability—two factors critical for real-time deployment in practice.

These existing studies demonstrate the superiority of kernel-based and ensemble-based models over other classification strategies for seismic classifications. But in the process, they expose gaping loopholes when it comes to interpretability, real-time assessment of severity, and deploy ability that users have access to. This work redresses this disparity by unifying classification, cluster-based severity forecasting, and transparent AI within an elastic web-supported system.

DOI: 10.17148/IARJSET.2025.125196

2. EXISTING SYSTEM

Current earthquake prediction systems rely primarily on conventional seismological methods, where ground motion is measured with seismographs and waveforms are examined to detect seismic activity. They are typically event detection-based reactive systems, focusing on event detection once the major waves have been detected and not on delivering predictive values. While useful in providing real-time alerts, such systems are unable to forecast events prior to their occurrence with high precision. Conventional models are physically and geologically interpretation-dependent and hence inflexible in dealing with noisy or complicated data. With the advent of machine learning in this context, various models have been proposed that utilize historical seismic data for event classification. Most machine learning-based methods currently, however, are limited to binary classification, predicting the occurrence or non-occurrence of an earthquake, without assessing the magnitude of the event or distinguishing it from other seismic anomalies such as explosions. Moreover, these models are black boxes with poor interpretability and are not components of real-time, user-accessible systems. This limits their use in real-time applications where instant decision support and explainability are necessary.

3. PROPOSED SYSTEM

The proposed system provides a machine learning-based, intelligent earthquake event prediction system that is more advanced than binary classification by including event detection and severity analysis. Unlike traditional systems, the proposed model uses past seismic data and applies advanced classification algorithms such as Random Forest and XGBoost to classify seismic events into earthquake warning, explosion, and no earthquake. These classifiers analyze a number of features such as magnitude, depth, root mean square (RMS), and depth error to detect patterns reflecting the probability and nature of a seismic event. If the model identifies an "earthquake warning," the system also evaluates its potential impact through a K-Means clustering algorithm to determine severity levels, labeled as Minimum, Moderate, or Severe.

To deal with the uninterpretability of traditional machine learning systems, the proposed framework uses LIME (Local Interpretable Model-Agnostic Explanations), giving users explicit explanations of how each prediction was made. The overall system is deployed through a Flask-based web application, enabling real-time user interaction where users input seismic parameters and receive predictions, severity evaluations, and cautionary recommendations in real time. Through the integration of classification, clustering, explainable AI, and web integration, the proposed system is a scalable, interpretable, and highly actionable tool for earthquake monitoring and early warning, aimed at enhancing disaster preparedness.

4. SYSTEM ARCHITECTURE

The proposed earthquake prediction system follows layered and modular design principles for maintainability, scalability, and user-friendliness. The system architecture is separated into five main components: the User Interface, Data Preprocessing Layer, Machine Learning Prediction Engine, Severity Classification Module, and the Output and Notification Layer. These layers are integrated into a pipeline and work together to process user input, analyze seismic parameters, and generate real-time, interpretable earthquake-related event predictions.

On top of the stack is the User Interface, which is managed through Flask and HTML technologies. The interface provides users with the option to input seismic attributes such as magnitude, depth, RMS value, and geographic coordinates. These are passed to the backend server as HTTP requests, where they are converted in the Data Preprocessing Layer. The layer normalizes and encodes raw input with pre-trained utilities such as scaler.pkl and label_encoder.pkl to prepare the data to be in the format required by the machine learning models that have been trained.

The Machine Learning Prediction Engine is the core analytical component of the system. Here, classification models such as Random Forest and XGBoost operate on the features and predict the input to be one of three classes: earthquake warning, explosion, or no threat. If the classification result is an earthquake warning, the system once more sends the input to the Severity Classification Module. This module employs a K-Means clustering model to classify the detected earthquake event to one of three severity classes: Minimum, Moderate, or Severe. These classifications are statistical patterns learned from the historical data. Finally, the Output and Notification Layer presents the prediction result to the user via the web interface. It includes the classified event type, severity level (if specified), and contextual safety advice. The layer also addresses real-time response formatting and is designed to be able to accommodate notification mechanisms for integration with early-warning or public alert systems.

The system as a whole has a defined data flow. Historical seismic datasets like magnitude, depth, RMS value, and depth error are initially collected in order to train the machine learning models. Data is preprocessed using techniques such as normalization and cleaning to preserve uniformity. Feature extraction is utilized to collect the most important parameters for prediction and learning. These features are fed into machine learning models, which are trained and tuned

DOI: 10.17148/IARJSET.2025.125196

on metrics such as accuracy, precision, recall, and F1-score. For deployment, real-time input is fed into the same preprocessing and prediction pipeline, with meaningful output used for well-informed decision-making.

This architecture enables a pipeline for full-cycle analysis, from raw seismic data ingestion through to real-time, actionable alerts. It not only facilitates systematic and efficient forecasting but also explainability, and the system is thus suitable for public consumption as well as expert-level disaster response applications.

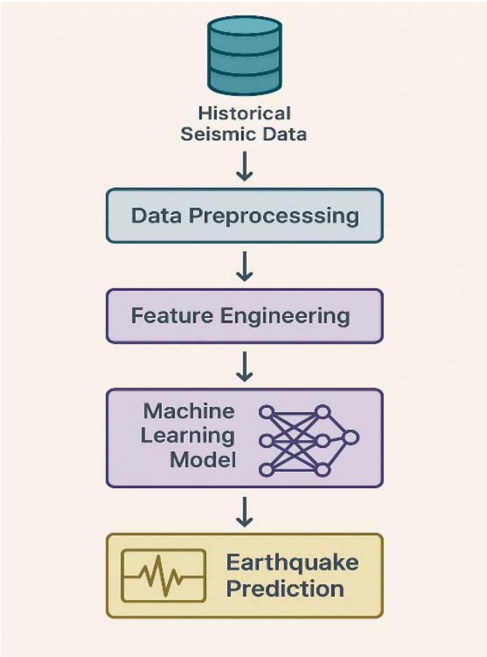

Fig-1: System Architecture

5. DATASETS

The information for this study come from publicly available and very reliable seismic data archives, mostly the United States Geological Survey (USGS). USGS offers complete, real-time histories of worldwide seismic activity and is the most authoritative and dependable source of information on earthquakes. The information consists of precise parameters of earthquake events, such as magnitude, depth, latitude, longitude, root mean square (RMS) value, depth error, and other measurement-based parameters. The parameters are based on a worldwide network of seismic monitoring stations and sensors, producing high fidelity and consistency to the data.

For the project, a judiciously selected dataset of 10,100 records and 8 key attributes was utilized. The selected features possess both the spatial and physical characteristics of seismic activity and are hence best suited for supervised machine learning applications such as classification and severity analysis. The records are structured such that a single row represents a single seismic event, and it possesses its corresponding attributes along with a labeled outcome, "earthquake warning," "explosion," or "no earthquake." This label is used as the target variable to train and test the predictive models. Before inputting the dataset into the machine learning pipeline, a series of preprocessing steps were carried out to enhance its quality and compatibility with the modeling framework. These included missing value management through imputation or deletion, outlier removal through statistical techniques such as z-score and boxplot analysis, and feature selection through correlation analysis to incorporate only the most important features. For numerical feature preparation for optimal model performance, standardization was carried out using StandardScaler, which offers uniformity between features of different scales. The categorical target labels were encoded using LabelEncoder for model interpretability. The dataset was split into training and testing subsets in an 80:20 ratio to facilitate efficient learning and generalization.

Overall, the data set provides a high-diversity, informative basis for predicting earthquakes using machine learning. By having a multivariate structure made up of varying geographic, physical, and measure-derived features combined together, the models can learn to identify robust patterns in seismic activity and present clear, accurate classifications.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125196

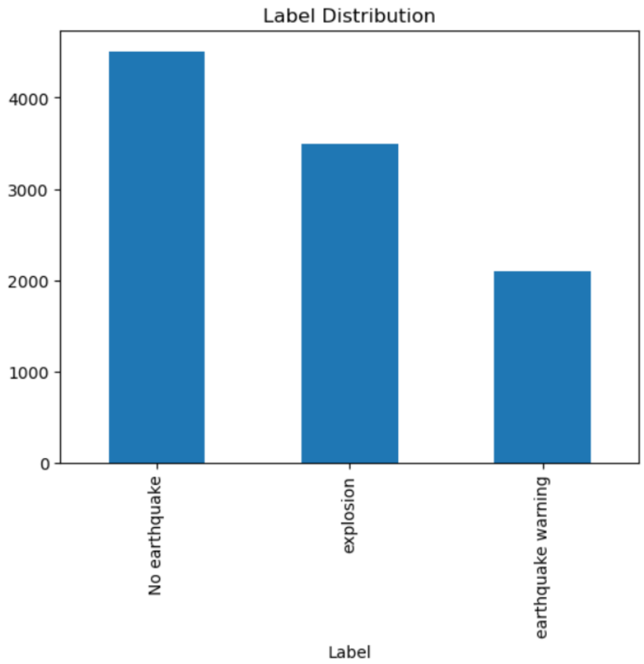

Fig-2: Dataset Label Distribution

6. IMPLEMENTATION

The earthquake forecasting model was created following a mix of supervised and unsupervised machine learning approaches that were used for varying purposes at various stages of the project. The approach is modular with a clear function for preprocessing data, classification, estimation of severity, and explainability.

The primary classification issue was addressed using a set of supervised learning methods including Random Forest, XGBoost, Support Vector Machine (SVM), Decision Tree, Naive Bayes, and K-Nearest Neighbors (KNN). The algorithms were trained to classify seismic events as earthquake warning, explosion, or no earthquake. XGBoost was selected as the final classifier after comparing the algorithms on a number of performance metrics since it was the most accurate and can handle noisy and class-imbalanced data. Random Forest was also doing well and was a good baseline since it is stable and can give feature importance.

For earthquake warning severity evaluation, the project utilized K-Means Clustering, an unsupervised method. If the classifier returns "earthquake warning," the input is passed to the K-Means model, which assigns the event to one of three classes of severity: Minimum, Moderate, or Severe. It uses features such as magnitude, depth, RMS, and depth error, and helps to give actionable insight into the intensity of the event that was detected.

To increase the explainability of the model, LIME (Local Interpretable Model-Agnostic Explanations) was added. LIME generates a human-interpretable explanation of each prediction by locally approximating the model with an interpretable model, thereby offering end users increased trust and transparency. All the models were created and trained with Python libraries like scikit-learn and XGBoost and were serialized using joblib for deployment. The backend as well as the user interface was written in Flask to enable live input of data and real-time output immediately using an internet-based system.

With this multi-algorithmic solution, the system not only identifies seismic events with accuracy but also offers severity understanding and interpretability and can be applied to actual disaster warning systems.

7. RESULT

The earthquake prediction model was developed using a labeled dataset comprising essential seismic parameters, including magnitude, depth, RMS, and depth error. To identify the most accurate model for predicting the occurrence of earthquakes, multiple machine learning algorithms were trained and evaluated on this dataset. The dataset underwent thorough preprocessing, which involved cleaning and normalizing the data before it was divided into training and testing subsets to ensure robust model development and evaluation.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125196

Among the algorithms tested, the Random Forest Classifier achieved an accuracy of 0.98, whereas the XGBoost Classifier demonstrated superior performance with an accuracy of 0.99. These results highlight the effectiveness of the XGBoost model in learning the complex relationships within the seismic data and reliably predicting whether an earthquake is likely to occur based on specific input features.

Fig-3: Home Page

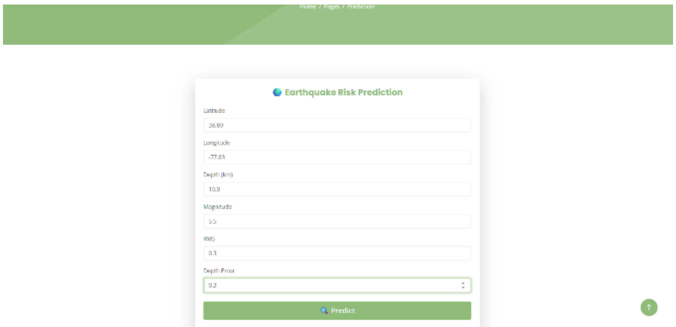

Fig-4: Earthquake Prediction User Interface

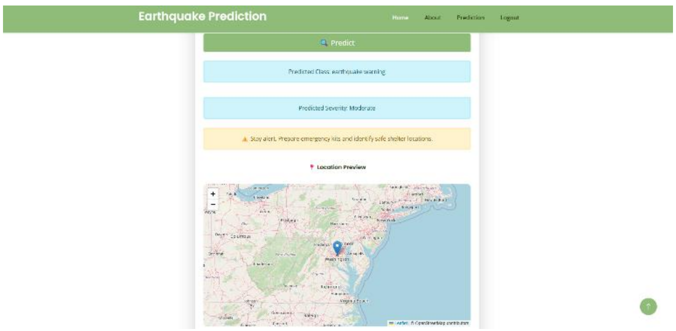

Fig-5: Earthquake Prediction result as ‘Moderate’

8. CONCLUSIONS

The Earthquake Prediction System project demonstrates the effective implementation of machine learning methods for forecasting seismicity, of extremely high precision in classification. With Random Forest and XGBoost algorithms, the system finds earthquakes and explosions by identifying patterns in seismic data. The project highlights how machine learning participates in emergency planning, offering dynamic predictive tools and decision-making tools. It also speaks of the model explainability and performance metrics in delivering reliable results. Future improvements may be directed towards improving prediction accuracy, incorporating additional data sources, and maximizing real-time capabilities for more extensive application in risk management.

DOI: 10.17148/IARJSET.2025.125196

REFERENCES

[1]. Shrote, P., Dasarwar, P., & Dongre, S. (2024). Earthquake Prediction through Machine Learning Approach. International Journal of Advanced Technology & Research.

[2]. Babu, D. B., Revathi, M. L. N., & Senthil, M. (2024). Earthquake Prediction Model using Random Forest & Gradient Boosting Algorithms. Journal of Engineering Science & Technology.

[3]. Giridhar, U. S., Prajapati, N., & Sonkusa, R. (2023). Analysis & Determination of magnitude of earthquake using STA/LTA algorithm. Journal of Earthquake Science & Technology.

[4]. Patil, R., Devkar, A., Patil, S., Raut, R., & Todakari, N. (2023). Earthquake Depth & Magnitude Prediction Model Using Artificial Neural Network. Journal of Earthquake Prediction Research.

[5]. Kumar, R., Gupta, M., Bedi, K. S., & Upadhyay, A. (2023). Earthquake Prediction using Machine Learning. International Journal of Data Science.

[6]. Xinyu Liu, Jia Chen, Linxio Xu., (2023). Study on Longwave Radiation Anomalies in the Dual Earthquakes in Turkey Based on LSTM.

[7]. Sultan Lok, Murat karabatak. (2021). Earthquake Prediction by using time series and analysis.