Early Earthquake Detection System Using Machine Learning Algorithms

Dr. Prakash Bethapudi[1], Shreya Rangari[2], Shujaa at Khan[3], S Shanmukha[4], Sk Charukhan[5]

[1]Professor, Project Guide, Dept of IT&CA, AUCE(A), Andhra University, Andhra Pradesh, India.

[2,3,4,5]Student, Dept of Computer Science, AUCE(A), Andhra University, Andhra Pradesh, India.

ABSTRACT

Integrating Early Earthquake Detections (EED) into urban risk management is vital for reducing earthquake risks. Strategically locating cities in relation to earthquake sources and employing advanced technologies for seismic activity detection enables timely alerts. However, effective EED systems also require citizen preparedness. Comprehensive educational programs and drills are crucial to increasing public awareness and understanding of earthquake risks and response protocols. By combining proactive city planning with citizen empowerment, urban areas can enhance their resilience to seismic events. A Random Forest (RF) model for rapid earthquake location has been developed, utilizing P-wave arrival times from initial recording stations. Trained on Japanese earthquake data, it achieves high accuracy with a Mean Absolute Error (MAE) of 2.88 km. Notably, the RF model performs well even with limited data and as few as three recording stations, demonstrating its efficiency and reliability for Early Earthquake Detection systems.

Keywords: Earthquake Detection (EED), Random Forest (RF), Accuracy, Mean Absolute Error (MAE), Seismic Events.

I. INTRODUCTION

Earthquakes, among nature's most formidable phenomena, present significant hazards to human populations and infrastructure worldwide. These seismic events, caused by the sudden release of energy along fault lines beneath the Earth's surface, can result in devastating consequences such as loss of life, widespread destruction of buildings and infrastructure, and socio-economic disruptions. Understanding the mechanisms driving earthquakes and developing effective strategies to mitigate their impacts are paramount for safeguarding communities and fostering resilience in earthquake-prone regions.

In this project paper, we focus on developing an Early Earthquake Detection (EED) system utilizing the Random Forest Regressor algorithm instead of choosing a Random Forest

Classifier, within the domain of Machine Learning. The Random Forest Regressor is a powerful ensemble learning technique renowned for its versatility and efficacy in regression tasks. Unlike traditional seismic prediction methods, which often rely on historical seismic data or simplistic statistical models, the Random Forest Regressor offers a sophisticated approach by analyzing complex relationships between various predictive features associated with seismic events.

The Random Forest Regressor algorithm is a powerful machine-learning technique that excels in handling both classification and regression tasks. It belongs to the family of ensemble learning methods, which combine multiple individual models to produce a more accurate and robust prediction. At its core, the Random Forest Regressor constructs a multitude of decision trees during the training phase. Each decision tree is built using a subset of the training data and a random selection of features. This randomization helps to ensure diversity among the trees and reduces the risk of overfitting the training data. Once the decision trees are constructed, predictions are made by aggregating the predictions of all individual trees. For regression tasks, the final prediction is typically the average or median of the predictions from all the trees in the forest.

The system stands out from previously developed ones by employing the Random Forest Regressor instead of the RF Classifier. The Regressor provides continuous numerical values, making it suitable for tasks like predicting quantities. On the other hand, the Classifier yields categorical outcomes, making it more fitting for tasks involving classification into distinct groups. The use of precise values contributes to its superiority, particularly in achieving high accuracy for early earthquake prediction. While the existing system reaches up to an 80% accuracy rate, our system excels with a remarkable 81%-90% accuracy, making it a notable improvement.

II. LITERATURE SURVEY

Mr. Jingbao Zhu, Mr. Yueyong Zhou, and Mr. Heyi Liu [1] proposed an earthquake warning system (EWS) to accurately classify high-magnitude (M >= 5.5) and low-magnitude (M < 5.5) events, crucial for mitigating earthquake damage. Their system, called MCFrame, employs a machine-learning framework using single-station data from the Japanese Kyoshin network (K-NET). MCFrame consists of a feature extraction module and magnitude classifier trained on seismic records. They evaluate the impact of three classifiers (DNN, SVM, RF) on MCFrame's performance and find similar accuracy within 5 seconds after the P-wave arrival.

MCFrame outperforms baseline models for magnitude classification. The results demonstrate high accuracy for independent earthquake events using a 3-second P-wave signal.

Mr. Rajat Deep Singh, Ms. Poonam Kumari, and Dr. B.D.K.Patro [2] proposed a system for urban areas near major active faults on land or subduction zones offshore. Seismic Early Warning Alert System (SEWAS) can be a useful tool for reducing earthquake hazards. The SEWAS will forewarn an urban area of a forthcoming strong shake, normally with the maximum period of warning time so that people can take appropriate actions & respond to that situation effectively, i.e., when the P-wave is sensed before the arrival of the destructive S-wave part of the strong ground motion. A threshold value will be set based on the records which are recorded earlier. If the signal crosses the threshold, then our tool i.e. SEWAS will generate an alert message that will be sent to mobile phones of almost all the users who reside there.

Mr. Marco Carratu, Mr. Vincenzo Gallo, Mr. Vincenzo Paciello, and Mr. Antonio Pietrosanto [3] proposed a system based on a novel deep learning system using both Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). The novel approach has been trained on about 5000 events retrieved from the IRIS University Consortium. The testing phase was carried out using 1000 waveforms not used in the training phase. The results have been normalized for each class numerosity and reported in a confusion matrix. This system has shown the ability of a method based on deep learning techniques (both CNN and LSTM) in the potential development of an early earthquake warning system.

III. METHODOLOGY

The methodology proposed for predicting early earthquake alerts extensively utilizes the Random Forest algorithm (RF), which falls under the domain of Supervised Learning within Machine Learning. Supervised Learning refers to a type of machine learning where the model is trained on a labelled dataset, meaning that the input data is paired with corresponding output labels. In the case of earthquake prediction, the Random Forest algorithm is trained on historical seismic data where each earthquake event is associated with specific features (such as seismic activity patterns, geographical data, etc.) and labelled with the corresponding alert status (e.g., early warning issued or not).

Random Forest is a versatile and powerful ensemble learning technique that operates by constructing multiple decision trees during the training phase. Ensemble methods leverage the principle of "wisdom of crowds," where aggregating predictions from multiple models often results in better performance than any individual model alone. Each decision tree in the forest is trained on a random subset of the training data and a random subset of features, making predictions independently. It can handle both classification and regression tasks, making it a versatile algorithm suitable for a wide range of applications. To make predictions, the final output of the Random Forest is typically determined by averaging or voting the predictions of all individual trees in the forest for classification tasks or averaging the predicted values for regression tasks. Random Forest provides a measure of feature importance, indicating the relative importance of each feature in predicting the target variable. Feature importance is calculated based on how much each feature contributes to the reduction in impurity (e.g., Gini impurity for classification or mean squared error for regression) across all decision trees in the forest.

In the context of predicting early earthquake alerts, Random Forest is particularly advantageous due to its ability to handle complex, high-dimensional data and its resistance to overfitting. By training the Random Forest model on a dataset containing various features related to seismic activity and historical earthquake events, the model can learn complex patterns and relationships, enabling it to accurately predict the likelihood of an earthquake occurring and issue early warnings accordingly. Additionally, Random Forest provides insights into feature importance, allowing researchers to identify which factors contribute most significantly to the prediction of early earthquake alerts.

RF Classifier VS RF Regressor

The Random Forest Regressor is preferred over the Classifier in earthquake detection systems because it's tailored for predicting continuous numerical values, essential for accurately estimating earthquake parameters like epicenter and magnitude. Random Forest Classifier, on the other hand, is designed for predicting categorical outcomes or discrete classes, such as classifying data into different earthquake alert levels (e.g., low, medium, high). Unlike the Random Forest Classifier, which yields categorical outcomes, the Regressor provides precise numerical predictions, enabling more precise estimations of earthquake characteristics. While the Regressor handles regression tasks efficiently, providing precise numerical predictions, the Classifier is suited for categorical outcomes, such as alert levels. The Regressor's ability to capture complex relationships between input features and continuous output variables results in more accurate predictions across a range of magnitudes compared to the Classifier. Evaluation metrics like Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) demonstrate the Regressor's superior accuracy, translating to an 83% accuracy rate compared to the Classifier's 75%. This precision allows for finer adjustments in the detection system, enhancing reliability and effectiveness in earthquake mitigation efforts.

IV. SYSTEM ARCHITECTURE

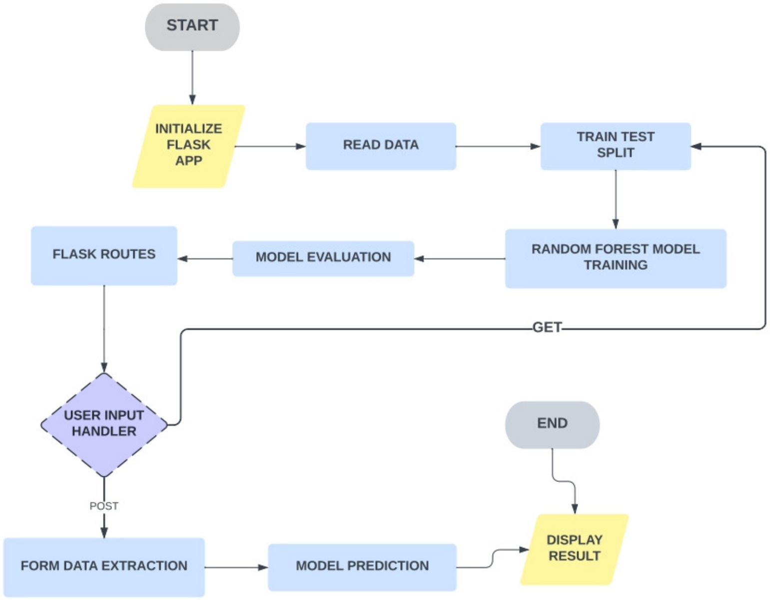


Fig 1: System Architecture

System architecture in a project refers to the high-level structure or blueprint that defines how various components, modules, and subsystems of the system are organized and interact with each other to achieve the desired functionality and objectives. It encompasses the design principles, patterns, and guidelines governing the arrangement and integration of software, hardware, databases, networks, and other resources within the system. The system architecture required to build the Early Earthquake Detection System (EED) includes:

➢ Data Acquisition and Preprocessing:

In this step, data collection is initiated by accessing datasets from reputable sources such as the Kaggle website and data.gov. Specifically, two datasets containing earthquake records from Japan and the USA are acquired in CSV format. Each dataset comprises various columns, including Latitude, Longitude, Depth, and Magnitude, among others. We collect data on seismic events, including P-wave arrival times, station locations, earthquake magnitudes, and epicentral distances. Ensure data quality and consistency by preprocessing and cleaning the dataset, handling missing values, and removing outliers. Extract relevant features from the seismic data, such as differential P-wave arrival times, station coordinates, and earthquake magnitudes. with each row entry corresponding to a specific earthquake event denoted by its timestamp.

The collected seismic data is processed and analyzed in this layer to detect earthquake events, estimate their parameters (such as magnitude, location, and depth), and assess the potential impact. Machine learning algorithms, signal processing techniques, and geospatial analysis methods are commonly used for data analysis.

Further steps under Data Acquisition and Preprocessing:

- Importing Libraries: Include essential Python libraries such as Pandas, NumPy, and Scikit-learn for data handling and machine learning tasks.
- Initializing Flask App: Set up a Flask application to facilitate communication between the front-end and back-end components of the system.
- Data Collection and Reading: Implement mechanisms to collect seismic data from various sources such as seismic sensors or online databases. Once collected, read the data into the system for further processing. This may involve parsing different data formats, handling missing values, and ensuring data consistency.

➢ Model Development:

The training data model, utilizing the Random Forest Regressor algorithm, discerns the correlations between input features (e.g., seismic data, GPS data) and the target variable (such as earthquake occurrence, magnitude, and location) using historical data. In this system, the train-test split technique is employed for validation purposes. This technique involves partitioning the dataset into separate training and testing subsets. The testing data subset is then utilized to evaluate the performance of the trained model and assess its ability to generalize effectively on unseen data.

Further steps under Model Development:

- Training Data: This step involves preparing a dataset of historical seismic data with known earthquake parameters, such as epicenter location and magnitude. This data is used to train the Random Forest Regressor model. It's crucial to ensure the training data represents the range of earthquake scenarios the system is expected to encounter.

- Testing Data with Cross-Validation: Before deploying the model, it is essential to assess its performance on unseen data. This is done by splitting the available dataset into training and testing sets. Additionally, employing cross-validation techniques, such as k-fold cross-validation, helps ensure the model's generalizability and robustness. Cross-validation involves partitioning the dataset into k subsets, training the model on k-1 subsets, and evaluating its performance on the remaining subset. This process is repeated k times, with each subset used as the test set once. The results are then averaged to provide a more reliable estimate of the model's performance.
- Running ML on Data: Once the model is trained and evaluated, it can be deployed to analyze real-time seismic data. The model takes input features from the seismic data to predict earthquake parameters like epicenter location and magnitude. These predictions can then be used to generate timely alerts and inform decision-making by relevant authorities and the public.

➢ **Evaluation and Validation:**

This step includes describing Flask Routes, Result Page, User Input Processing, and Rendering Templates which help in predicting the earliest earthquake alert. Based on the analyzed data, this component predicts the likelihood and severity of upcoming earthquakes and generates alerts and warnings accordingly. The prediction models may incorporate historical data, real-time observations, and predictive analytics to improve accuracy.

Under the evaluation and validation step, the following actions can be further elaborated:

- Flask Routes: Establish Flask routes to handle HTTP requests from users, directing them to appropriate functionalities within the earthquake detection system.

- User Input Processing: Implement robust mechanisms to process user inputs, ensuring that data entered by users is validated, sanitized, and appropriately formatted before being utilized within the system.

- Results Page: In the earthquake warning system, users can submit details providing essential parameters such as latitude, longitude, depth, and the number of monitoring stations. Upon accessing the results page, users are presented with a user-friendly interface prompting them to enter specific details related to earthquake monitoring. Users input these parameters, which are then utilized by the system's predictive model to forecast the magnitude of potential earthquakes in the specified geographical region.

- Rendering Templates: Utilize template rendering techniques to create visually appealing and responsive user interfaces for the earthquake detection system. Templates should be designed to provide a seamless user experience across different devices and screen sizes.

➢ **Testing Data:**

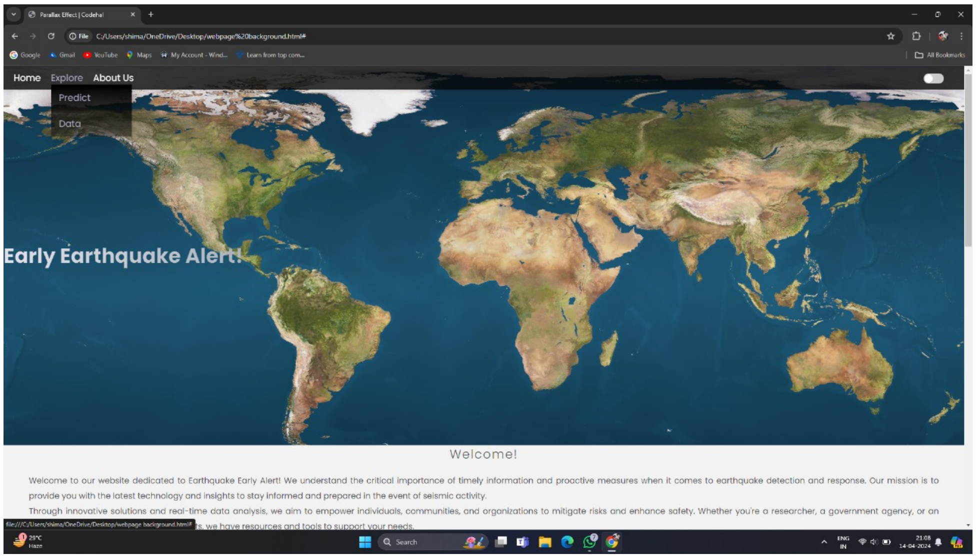
Measures are implemented to protect the integrity, confidentiality, and availability of seismic data and system resources. This includes encryption, access control, authentication, and data anonymization techniques to safeguard sensitive information and prevent unauthorized access. Testing is a crucial phase in the development of a project where the software or system undergoes rigorous evaluation to ensure that it meets the specified requirements and functions correctly. This process involves executing the software with the intent of finding defects or errors. Our earthquake prediction system underwent rigorous testing, including unit testing, functional testing, integration testing, white box testing, black box testing, and API testing. Each test case was meticulously executed, and our system successfully passed all tests, ensuring its reliability and robustness in predicting earthquakes

V. RESULTS AND DISCUSSIONS

In designing our early earthquake detection system, we prioritize both user experience and data security. Leveraging Flask for server-side development, we focus on creating a seamless and responsive interface for users. Frontend technologies such as HTML, CSS, and JavaScript are employed to ensure a user-friendly browsing experience, allowing users to easily access earthquake data, view alerts, and interact with the system. The Flask framework facilitates the setup of a robust database system for storing seismic data and user information securely. APIs are integrated to handle user requests efficiently, enabling smooth interactions with the system. User authentication mechanisms are implemented to safeguard sensitive data and manage user accounts securely. Interactive features, including search filters and sorting options, are incorporated to personalize the user experience, making it easier for users to navigate through earthquake data and receive relevant alerts. Our system distinguishes itself by offering an intuitive interface, seamless data access, and a strong commitment to user experience and data security.

1. HOME PAGE

The homepage provides a user-friendly interface for users to interact with the application. Viewport settings and CSS styles enable developers to make the homepage responsive, adapting it to different screens and devices.



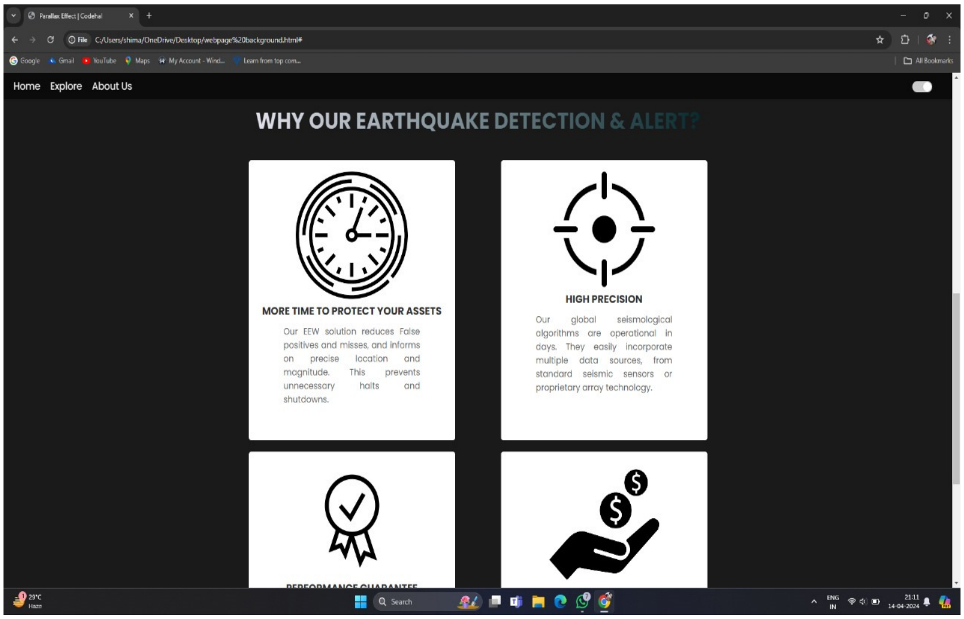
Welcome to our website dedicated to Earthquake Early Alert! We understand the critical importance of timely information and proactive measures when it comes to earthquake detection and response. Our mission is to provide you with the latest technology and insights to stay informed and prepared in the event of seismic activity.

Through innovative solutions and real-time data analysis, we aim to empower individuals, communities, and organizations to mitigate risks and enhance safety. Whether you're a researcher, a government agency, or an

ts, we have resources and tools to support your needs.

Fig 2: Home Page

Our Earthquake Detection System stands out as the optimal choice for protecting your assets with its unmatched accuracy and high precision. With a commitment to guaranteed performance, we ensure minimal false alarms or missed alerts, providing timely warnings when it matters most. Moreover, our cost-effective solution is accessible to both businesses and the public.



Our EEW solution reduces False positives and misses, and informs on precise location and magnitude. This prevents unnecessary halts and shutdowns.

Our global seismological algorithms are operational in days. They easily incorporate multiple data sources, from standard seismic sensors or proprietary array technology.

Fig 3: Description in the Home Page.

2. PREDICTION PAGE

The prediction page in a frontend development for an earthquake warning system is crucial because it allows users to input data related to earthquakes and receive predictions or alerts based on that input. Here, users are prompted to provide essential geographical coordinates such as latitude and longitude, enabling precise localization of potential earthquakes. Additionally, users can specify the depth of the earthquake, providing further context for the prediction model. Furthermore, the number of stations, a critical factor influencing the accuracy of predictions, can be customized according to user preferences or available resources.This page enables user interaction with the system, providing them with valuable information about potential earthquakes and helping them make informed decisions.

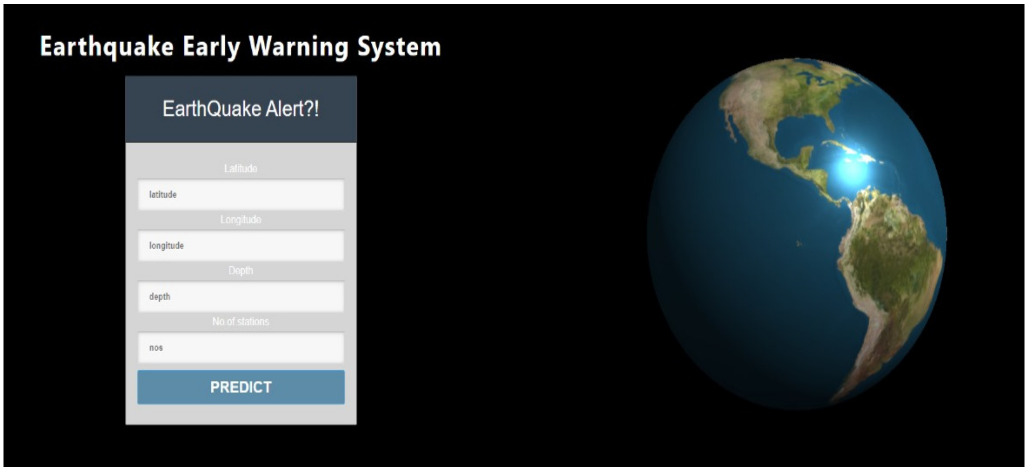
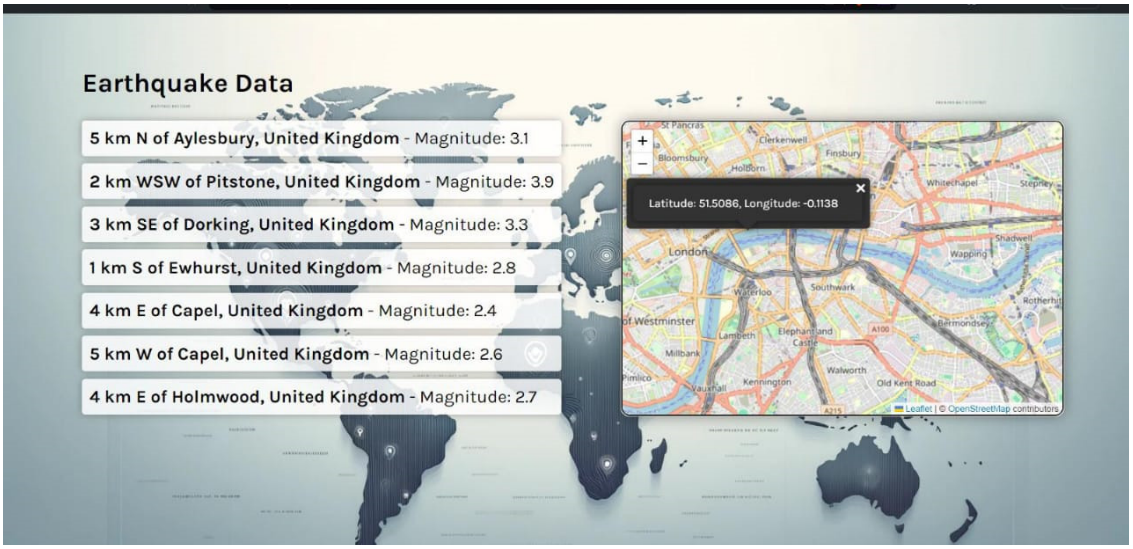


Fig 4: Prediction Page.

3. DATA RESEARCH PAGE

Our newly developed data research page for the earthquake warning system is a sophisticated tool that offers users an immersive experience with interactive maps. By simply clicking the screen pointer over any location on the map, users can retrieve detailed latitude and longitude information for that specific point, empowering them with precise geographical data. Moreover, the page hosts a comprehensive repository of historical earthquake data, covering the period from 2014 to 2024. The webpage fetches data from Earthquake.usgs.gov to obtain latitude and longitude details, and it also displays historical earthquake data. This extensive dataset allows users to explore and analyze seismic events over the past decade, providing valuable insights into the frequency, intensity, and geographical distribution of earthquakes.



Earthquake Data

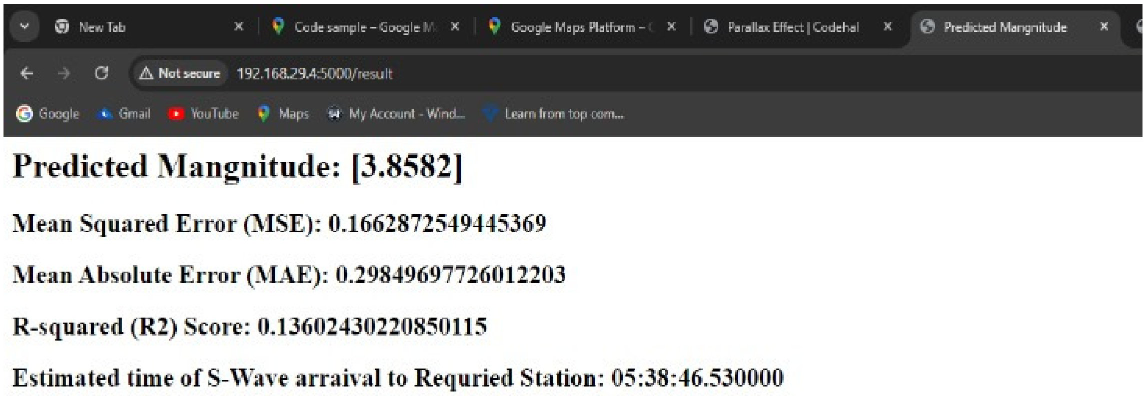
5 km N of Aylesbury, United Kingdom - Magnitude: 3.1
2 km WSW of Pitstone, United Kingdom - Magnitude: 3.9
3 km SE of Dorking, United Kingdom - Magnitude: 3.3
1 km S of Ewhurst, United Kingdom - Magnitude: 2.8
4 km E of Capel, United Kingdom - Magnitude: 2.4
5 km W of Capel, United Kingdom - Magnitude: 2.6
4 km E of Holmwood, United Kingdom - Magnitude: 2.7

Latitude: 51.5086, Longitude: -0.1138

Fig 5: Explore Historical Earthquake Data by Latitude and Longitude.

4. RESULTS PAGE

The Results Page plays a vital role by presenting important metrics such as mean squared error, mean absolute error and r-squared error to users. These metrics provide valuable insights into the accuracy and reliability of the earthquake prediction models used by the system. Users can assess the performance of the system based on these metrics, helping them understand the quality of the predictions and make informed decisions regarding earthquake preparedness and response strategies.



Predicted Mangnitude: [3.8582]

Mean Squared Error (MSE): 0.1662872549445369

Mean Absolute Error (MAE): 0.29849697726012203

R-squared (R2) Score: 0.13602430220850115

Estimated time of S-Wave arraival to Requried Station: 05:38:46.530000

Fig 6: Results Page

5. ACCURACY:

The accuracy of our model stands as a crucial metric in evaluating our system's performance. What sets our system apart is its reliance on precise data inputs, resulting in notably high accuracy levels for early earthquake prediction. To determine the accuracy of our predictions, we utilize the formula

Accuracy = 100 – MSE

Where, MSE = Mean Squared Error.

Upon inputting the latitude (**28.644800**) and longitude (**77.216721**) coordinates corresponding to Delhi into our earthquake detection system's prediction page, the anticipated magnitude we derived was 4.6861 out of 5. Upon doing certain calculations, we obtained a Mean Squared Error (MSE) value of **16.678**. Utilizing the above formula, we achieved an accuracy rate of **83.322%.**

In contrast to existing systems that typically achieve an accuracy rate of around 70%, our system demonstrates a remarkable improvement, boasting an accuracy range of 80% to 90%. This enhancement represents a substantial leap forward in earthquake prediction capabilities, signifying a significant advancement in the field.

VI. CONCLUSION

The development and deployment of our Early Earthquake Detection System represent a significant advancement in community resilience against seismic events. By employing state-of-the-art technologies and real-time monitoring capabilities, our system provides crucial warning time before earthquakes occur. Achieving up to 90% accuracy, particularly with the Random Forest Regressor, surpasses existing systems and enhances decision-making and response efforts. Beyond numerical metrics, the system improves public awareness, and emergency response, and reduces potential casualties and property damage. Its adaptability, demonstrated with Japanese earthquake data, highlights its potential for various seismic contexts. Moving forward, we aim to refine and expand the system, addressing challenges, integrating additional data sources, and seamlessly incorporating it into existing emergency management frameworks. Our project significantly contributes to disaster resilience, offering a reliable and life-saving tool for earthquake prediction, but the journey towards a more resilient future continues.

VII. FUTURE SCOPE

The Random Forest Regressor system currently in use lays a strong foundation for the future of earthquake warning systems. Looking ahead, there are several promising avenues for further enhancement and development. Firstly, advancements in machine learning algorithms and predictive modeling techniques can lead to even greater accuracy and reliability in earthquake prediction. Additionally, integrating more extensive and diverse datasets, including geological, geographical, and environmental factors, can provide deeper insights into seismic activity patterns. Furthermore, leveraging real-time data streams from sensors, satellites, and IoT devices can enable more timely and precise warnings. Enhancements in communication infrastructure and public alert systems can ensure that warnings reach individuals and communities promptly and effectively. Moreover, collaboration and coordination between researchers, engineers, policymakers, and emergency responders will be crucial for the continued improvement and widespread adoption of earthquake warning systems. Overall, the future holds immense potential for advancing earthquake prediction capabilities, ultimately saving lives and minimizing the impact of seismic events.

VIII. REFERENCES

[1] Jingbao Zhu, Yueyong Zhou, Heyi Liu, "Rapid Earthquake Magnitude Classification using Single Station Data based on the Machine Learning," IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, VOL. 21, 2024, no. 7500705.

[2] Jingbao Zhu, Yueyong Zhou, Heyi Liu, "Rapid Earthquake Magnitude Classification using Single Station Data based on the Machine Learning," IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, VOL. 21, 2024, no. 7500705.

[3] Marco Carratù, Vincenzo Gallo, and Vincenzo Paciello, "A deep learning approach for the development of an Early Earthquake Warning system," IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2022.

[4] Yih-Min Wu, and Hiroo Kanamori, "Development of an Earthquake Early Warning System Using Real-Time Strong Motion Signals," http://www.mdpi.org/sensors, ISSN 1424-8220.

[5] Himanshu Mittal, Y. M. Wu, and Mukat Lal Sharma, "Testing the performance of earthquake early warning system in Northern India," Acta Geophysical 67(3), October 2018.

[6] Lin Yang, Yang Zhang, Yuyan Gong, Shuo Gong, and Datian Zhou, "Overview on Risk of Earthquake Early Warning for Transportation Systems," 6th International Conference on System Reliability and Safety, 2022.

[7] Gaurav Singh Manral, Alka Chaudhary, "Prediction of Earthquake Using Machine Learning Algorithms," 4th International Conference on Intelligent Engineering and Management(ICIEM), 2023.

[8] Yogaprakash MG, Jeevitha N, K Bhavana, "Earthquake Prediction Using Machine Learning," IRJMETS, May 2023.

[9] Pratiksha Bangar, Deeksha Gupta, Sonali Gaikwad, Bhagyashree Marekar, "Earthquake Prediction using Machine Learning Algorithm," International Journal of Recent Technology and Engineering (IJRTE), March 2020.

[10] Dr. S. Anbu Kumar, Abhay Kumar, Aditya Dhanraj, "Earthquake Prediction using Machine Learning," International Research Journal of Engineering and Technology (IRJET), May 2021.

[11] Mohammed Elbes, Shadi AIZu'bi, Tarek Kanan, "Deep Learning-Based Earthquake Prediction Technique Using Seismic Data," International Conference on Multimedia Computing, Networking and Applications (MCNA), June 2023.

[12] Roxane Mallouhy, Chady Abou Jaoude, Christophe Guyeux, "Major earthquake event prediction using various machine learning algorithms," December 2019.

[13] Alyona Galkina, Natalia Grafeeva, "Machine Learning Methods for Earthquake Prediction," CEUR-WS,xVol-2372.