IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304097 |

An Ensemble Earthquake Prediction System Using Ml

Ch. Siva Subrahmanyam (Guide) [1], N.V.Sai Pramodh [2], A. Harsha Vardhan [3], K. Raghu Ram [4], D. Milky [5]

Assistant Professor, Department of Information Technology, S.R.K.R. Engineering College (A), SRKR Marg, Bhimavaram, Andhra Pradesh, India [1]

U.G. Student, Computer Science and Business System, Department of Information Technology, S.R.K.R. Engineering College(A), SRKR Marg, Bhimavaram, Andhra Pradesh, India.[2345]

ABSTRACT: Due to the frequent incidence of natural disasters in everyday life, we are burdened with significant losses. This paper focuses on one such natural disaster: Earthquake. The lack of early earthquake detection leads to substantial loss of both human and animal life, as well as extensive damage to infrastructure, making recovery seemingly insurmountable at times. Therefore, the primary aim of this project is to devise an earthquake prediction system capable of forecasting earthquakes for the next seven days, providing information regarding the location, date, longitude, latitude, and probability of occurrence. Such a prediction system could serve as a crucial tool for early preparedness, potentially saving countless lives.

KEYWORDS: Forest Classifier and Decision Tree Classifier, XG Boost, Machine Learning, Earthquake Dataset.

I. INTRODUCTION

Predicting earthquakes is notoriously difficult due to the complex nature of geological processes underlying the earth's crust. Earth consists of three main layers: The crust, mantle, and core. The crust, covering the earth's surface, is made up of tectonic plates that slowly move atop the mantle. These movements can lead to cracks and faults, potentially triggering earthquakes in unstable regions.

The main obstacles to earthquake prediction include:
1. Complex geological processes: The interactions between tectonic plates and fault behaviour involve intricate physical and chemical processes that are challenging to observe or replicate directly.

2. Data limitations: Predicting earthquakes requires vast amounts of high-quality data from various sources. Analysing these data can be time-consuming and may not reveal subtle patterns or correlations necessary for prediction.

3. Temporal and spatial uncertainty: Predicting the timing and location of earthquakes is extremely difficult due to the unpredictable nature of seismic events.

4. Magnitude scaling: Accurately predicting the magnitude of earthquakes presents exceptional challenges.

5. Ethical and social implications: False alarms can cause unnecessary panic and economic costs, while missed predictions can result in loss of life and property.

This project aims to address these challenges using machine learning (ml) advancements. The primary objective is to forecast earthquake occurrences in the upcoming seven days based on dataset obtained from the usgs website, which is updated every 15 minutes and contains information on earthquakes over the past 30 days[1].

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304097 |

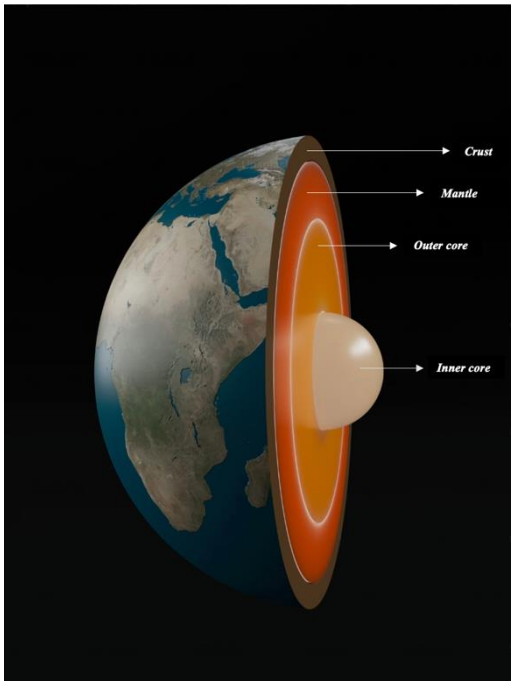

Fig1: Earth Layers

II. LITERATURE SURVEY

The research evaluates XGBoost's efficacy in predicting earthquakes within a 7-day period, emphasizing the importance of quality training data for model performance. Future directions involve exploring RNNs or LSTMs to address data quality limitations in earthquake forecasting[1].

The literature survey highlights a reliance on rule-based methods and shallow machine learning algorithms for earthquake prediction, pointing to the challenge of predicting high-magnitude earthquakes due to limited data. It suggests the need for advanced models to improve prediction capabilities[2].

The review examines earthquake prediction and management using diverse data sources, including social media, but notes a gap in detailing data acquisition methods, especially regarding traditional sources beyond social media[3].

A qualitative analysis on earthquake disaster assessment, focusing on categorizing information rather than quantitative analysis. They found that deep learning is increasingly applied in earthquake disaster assessment, encompassing prediction, damage detection, and analysing various data sources such as remote sensing, seismic, and social media data. However, the paper lacks in-depth exploration of technical challenges, particularly in addressing limitations in training data and model interpretability. [4].

The study compared FPA-ELM and FPA-LS-SVM hybrid models for earthquake prediction in southern California, with FPA-LS-SVM showing superior accuracy. However, concerns arise regarding regional focus and generalizability, highlighting the need for diverse datasets and testing across seismic regions[5].

The study utilized data pre-processing techniques such as cleaning, normalization, and data splitting. Support Vector Machine (SVM) and Logistic Regression were among the machine learning algorithms employed for earthquake prediction. Results indicate the effectiveness of these models in analysing earthquake data. However, the study lacked comparison with existing methods and acknowledged limitations including the potential for enhancements with larger datasets and time-series analysis. [6].

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304097 |

In another research, a stacking model combining Random Forest and SVM algorithms achieved an 83% accuracy rate in earthquake prediction. While successful, future studies should explore scalability and real-time implementation across various seismic regions[7].

III. DATASET

The dataset has been acquired from the uses (u. s. geological survey), an organization engaged in focused research on the origins and impacts of earthquakes. Uses offers immediate notifications, streams, and online services regarding earthquakes shortly after their occurrence. We access the dataset from "https://earthquake.usgs.gov/earthquakes/feed/". The following outlines the features of the dataset, which consists of 22 attributes(features) and was trained on a sample size of 10561 instances.

- Time: Time stamp indicating when the event occurred(milliseconds)
- Latitude: Latitude at the time of event occurrence, where negative degree indicates location in southern hemisphere
- Longitude: Longitude at the time of event occurrence, where negative degree indicates location in western hemisphere
- Depth: The events depth is expressed in kilometres
- Mag:The events magnitude
- Magtype: The method or algorithm used to calculate the preferred magnitude.
- Nst: The number of seismic stations used to determine earthquake location.
- Gap: The angle between neighbouring seismic stations around the events location.
- Dmin:The horizontal distance, in degrees, from the epicentre to the nearest station.
- Rms: The root means square travel time residual(in seconds).
- Net: The identifier of the data source contributor for the occurred event.
- Id: A unique identifier assigned to the event
- Types: A comma separated list detailing the product types associated with the event
- Place: It is the place where the event has occurred
- Type: The type of event occurred
- Location source: The network that initially provides the location information for the event.
- Magsource: The network that initially provides the information relating to magnitude for the event.
- Horizontal error: The degree of uncertainty regarding the reported location of the event(in km)
- Depth error: The depth error is indicative of three principal errors along a vertical line.
- Magerror: The degree of uncertainty linked to the reported magnitude of the event.
- Magnst: The overall number o seismic stations employed in determining the earthquake magnitude.
- Status: A signal indicating whether the event has been subjected to human review.

IV. PROBLEM STATEMENT AND APPROACH

Predicting earthquakes is a critical challenge in earth science due to their profound and widespread consequences. The lack of awareness regarding date and location of impending earthquake contributes significantly to the substantial loss of life and property. This project aims to address the problem by leveraging machine learning techniques and utilize the selected dataset by developing the predictive models capable of forecasting probable locations and corresponding dates when earthquakes are likely to occur for the next upcoming seven days with improved accuracy. Considering it as a event series nature of problem, the project focuses on predicting earthquake events. The proposed approach for this project involves utilizing dichotomous classification (binary classification) to predict earthquake events. During the training period, a fixed rolling window is applied to calculate rolling averages of the past days data. The labels for training are generated by shifting a fixed window size forward in time. From this approach, we can estimate the probability of an earthquake occurring within a specified time frame as indicated. The model will undergo training using adaboost classifiers, specifically random forest classifier and decision tree classifier, and with svm(support vector machine), which then will be contrasted with xgboost. The evaluation will be based on auc -roc score and recall score, chosen due to binary classification nature of problem. The model exhibiting superior auc score and recall will be selected for further consideration.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304097 |

V. METRICS

Our project involves binary classification to predict earthquake occurrences (1) and non-occurrences (0) which is shown in fig 2. Using the Google Maps API, we aim to visualize the coordinates associated with these occurrences on a web application[1]. Our focus is on minimizing false negatives to ensure accurate earthquake identification. We evaluate model performance using the confusion matrix, AUC-ROC curve, and recall. A higher AUC-ROC score indicates better discrimination ability, while recall measures the proportion of true positives identified. We prioritize models with high AUC-ROC scores and recall values to achieve our goal of accurate earthquake detection

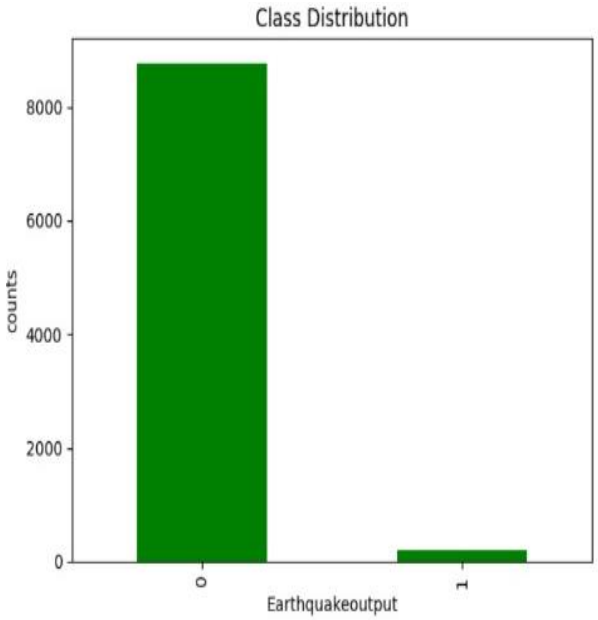

Fig 2

VI. PROPOSED SYSTEM:

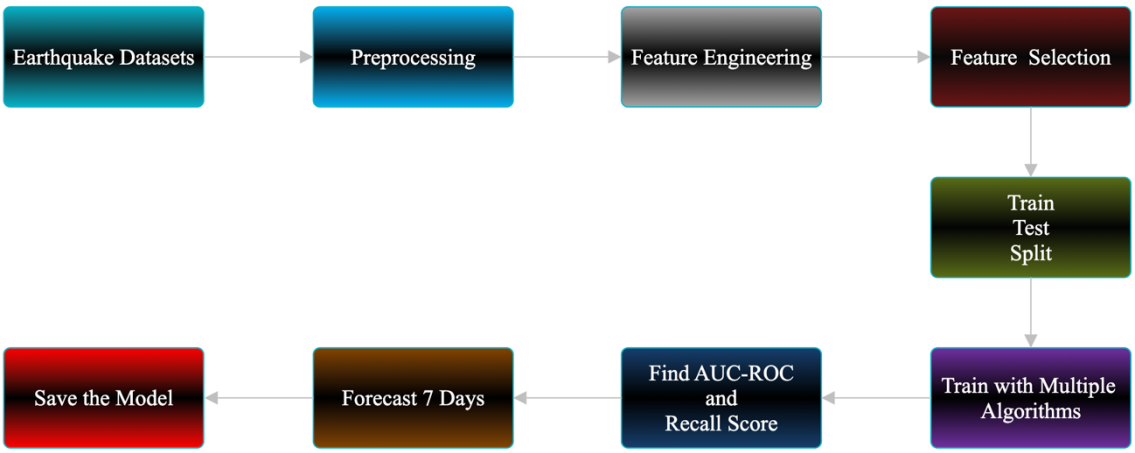

Fig 3

The above fig 3 showcases the proposed steps of our project. After analysing the dataset, we observed missing values in certain features, which required careful handling during data pre-processing to ensure accurate predictions. We

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304097 |

excluded features with more than 50% missing values from the model. The selected primary features included time, latitude, longitude, depth, magnitude, and place, as they contained reliable data without null values.

For improved visualization, we plotted bar graphs to analyse earthquake magnitudes, distinguishing between smaller (below 2) and larger (above 3) events across different countries.

In feature engineering, we implemented a fixed rolling window to compute average values for past days' data. Additionally, we introduced seven new features based on rolling window sizes for average depth and magnitude: 'dept22', 'dept15', 'dept7', 'magn22', 'magn15', 'magn7', and 'magnoutput'. These features allowed us to identify weekly, bi-weekly, and monthly trends, aiding in data understanding and forecasting.

The 'magnoutput' feature served as our target variable, representing forecasted values obtained by advancing the rolling window of past days forward in time, with a prediction span of 7 days.

We divided the data into training (70%) and testing (30%) segments and utilized 'gridsearchcv' for parameter tuning to enhance model accuracy through cross-validation. Evaluation metrics such as the confusion matrix, AUC score, and recall were used to assess the best model prediction, providing insights into true positives and false negatives

VII. ALGORITHMS

In our project, we utilized three ensemble methods for training our dataset: Decision Tree Classifier, Random Forest Classifier, and XGBoost Classifier, all imported from the `sklearn.ensemble` module of the scikit-learn Python library. Ensemble methods involve combining multiple machine learning models to enhance predictive performance.
Adaboost Algorithm: Adaboost aggregates numerous weak learners to perform classifications.
Decision Tree Classifier: This model resembles a tree structure, making decisions at each node (yes or no) to classify data.
Random Forest Classifier: This ensemble learning approach constructs multiple decision trees and aggregates their predictions to enhance accuracy.
XGBoost Classifier: Also, an ensemble learning technique, XGBoost is utilized for classification tasks, recognized for its superior speed and accuracy.

VIII. RESULTS

Results obtained for each classifier has shown in table 1:

CLASSIFER	AUC_ROC SCORE	RECALL
DECISION TREE	0.9165	0.833
RANDOM FOREST	0.9167	0.833
XGBOOST	0.9848	0.833

Table 1

Utilizing 5000 estimators and a learning rate of 0.003, xgboost showcases a notably elevated auc score of approximately 0.9848. It also presents merely 9 false negatives, a count comparable to that observed in random forest and adaboost. However, xgboost outperforms these alternatives regarding both true positives and false positives, yielding a recall score of 0.8335, akin to that of random forest and adaboost. Notably, xgboost demonstrates superior performance in distinguishing between positive and negative classes, as indicated by its remarkable auc score of 0.9847, surpassing other algorithms. This proficiency, coupled with its higher recall and auc score, renders xgboost a more resilient choice for handling class imbalance and mitigating false negative instances, solidifying its suitability for live data prediction and deployment within the application.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304097 |

Fig 4

IX. CONCLUSION

Xgboost algorithm gives us best auc_roc score and recall score.

- Currently, the model performance appears promising with XGBoost selected as the preferred algorithm for predictions in the web application, demonstrating superior AUC score and recall_score.
- Our primary objective is to forecast whether an earthquake will occur at a specific location on a given day. We prioritize minimizing False Negative predictions, as it's riskier to miss predicting an actual earthquake than incorrectly predicting one. While we can tolerate a higher rate of False Positives, prioritizing safety remains paramount.
- Furthermore, we utilized the Google API to predict earthquake occurrences for the next 7 days as shown in fig 5

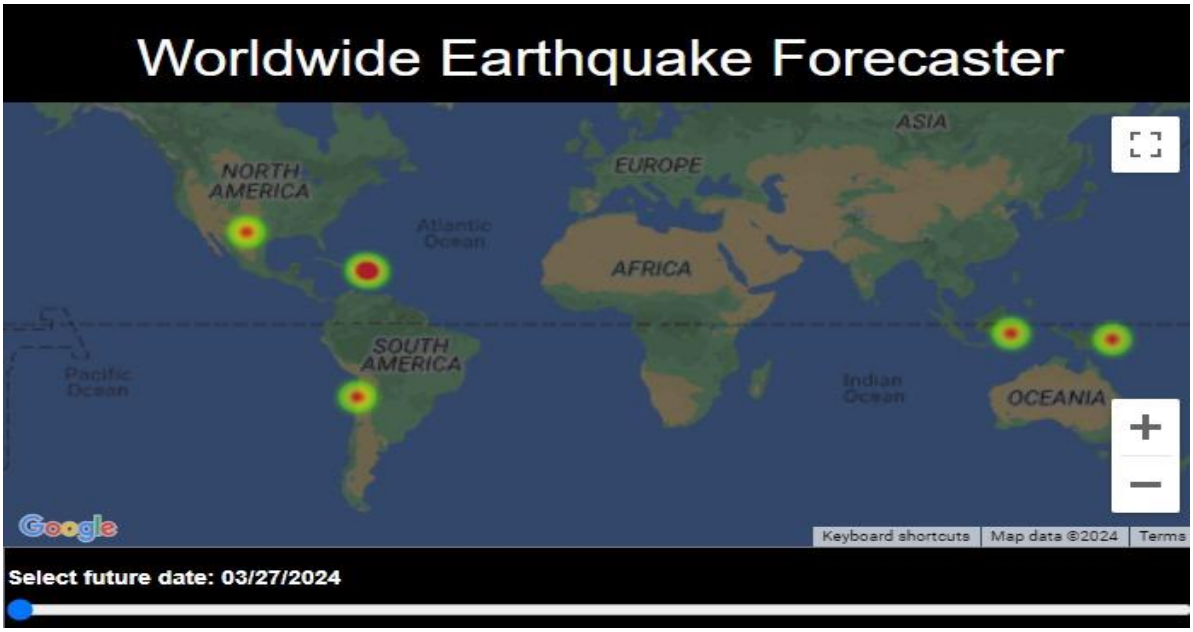

Fig 5

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304097 |

X. FUTURE WORKS

To extend our project for predicting earthquakes up to 15 days in the future, we must address several challenges that currently limit our prediction window to only 7 days:
1.Limited Availability of Reliable Data
2. Complexity of Models
3.Need for Sophisticated Algorithms
To tackle these challenges, we can leverage various algorithms, such as:
- Neural Network Architectures like CNNs and RNNs, which effectively capture complex patterns in seismic data.
- Graph Neural Networks (GNNs), suitable for capturing spatial relationships in seismic data, such as fault lines and seismic activity clusters.
- Hidden Markov Models (HMMs) for analysing seismic time series data and detecting earthquake precursors.
This extension can be achieved through:
- Advanced mathematical techniques to improve prediction accuracy.
- Gathering more seismic data from diverse sources.
- Enhancing computing power to handle the increased computational demands of extended prediction horizons.

REFERENCES

[1] Kulkarni, M., Mulay, C., Marathe, S., & Itankar, P. (2021). Earthquake Prediction using Machine Learning. International Journal of Advance Research, Ideas and Innovations in Technology (IJARIIT).
[2] Al Banna, M. H., Taher, K. A., Kaiser, M. S., Mahmud, M., Rahman, M. S., Hosen, A. S. M. S., & Cho, G. H. (2022). Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges.
[3] Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A., & Chakraborty, B. (2018). A review on application of data mining techniques to combat natural disasters. Ain Shams Engineering Journal.
[4] Jia, J., & Ye, W. (2023). Deep Learning for Earthquake Disaster Assessment: Objects, Data, Models, Stages, Challenges, and Opportunities. Remote Sensing, 15(16).
[5] Abdul Salam, M., Ibrahim, L., & Abdelminaam, D. S. (2021). Earthquake Prediction using Hybrid Machine Learning Techniques. International Journal of Advanced Computer Science and Applications.
[6] Faqueeh, S. U. R., Uddin, M. S., Halais, M. B. A., & Farooq, S. O. (2023). Earthquake Prediction Using Machine Learning. Mathematical Statistician and Engineering Applications.
[7] Bangar, P., Gupta, D., Gaikwad, S., Marekar, B., & Patil, J. (2020). Earthquake Prediction using Machine Learning Algorithm. International Journal of Recent Technology and Engineering (IJRTE), March 2020.
[8] P.Jiao and A.H.Alavi,"Artificial intelligence in seismology: Advent,performance and future trends," Geosci. Frontiers, vol. 11, no. 3,pp. 739–744, May 2020.
[9] Q. Wang, Y. Guo, L.Yu, and P. Li, "Earthquake prediction based on spatio-temporal data mining:An LSTM network approach,"IEEE Trans.Emerg.Topics Comput.,vol.8,no.1,pp.148–158, Jan. 2020.
[10] S. K. Majhi, S. S. Hossain, and T. Padhi,"MFOFLANN:Moth flame optimized functional link artificial neural network for prediction of earthquake magnitude," Evolving Syst., vol. 11, no. 1, pp. 45–63, Mar. 2020.
[11] K. M. Asim, S. S. R. Moustafa, I. A. Niaz, E.A.Elawadi,T.Iqbal,and F. Martínez-Álvarez, "Seismicity analysis and machine learning models for short-term low magnitude seismic activity predictions in Cyprus," Soil Dyn. Earthq. Eng., vol. 130, Mar. 2020, Art. no. 105932.
[12] Jena, R.; Pradhan, B.; Al-Amri, A.; Lee, C.W.; Park, H.-J. Earthquake Probability Assessment for the Indian Subcontinent Using Deep Learning. Sensors 2020,20, 4369.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details