Sarada Nandini Sekharamahanti

Project7

📋 Disaster response and recovery

🖥 Projects

🎓 GITAM University

Document Details

Submission ID

trn:oid:::1:3187763420

Submission Date

Mar 19, 2025, 12:31 PM GMT+5:30

Download Date

Mar 19, 2025, 12:35 PM GMT+5:30

File Name

Disaster_prediction_and_recovery.docx

File Size

909.1 KB

16 Pages

2,811 Words

17,312 Characters

11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

▸ Quoted Text

▸ Cited Text

▸ Small Matches (less than 8 words)

Match Groups

25 Not Cited or Quoted 11%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9%   🌐 Internet sources

8%   📖 Publications

4%   👤 Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

🔖 **25** Not Cited or Quoted 11%
Matches with neither in-text citation nor quotation marks

💬 **0** Missing Quotations 0%
Matches that are still very similar to source material

📑 **0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation

📑 **0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9%   🌐 Internet sources
8%   📖 Publications
4%   👤 Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet		
www.geeksforgeeks.org			**2%**

2	Student papers		
University College London			**1%**

3	Student papers		
University of Hertfordshire			**<1%**

4	Internet		
www.coursehero.com			**<1%**

5	Internet		
covalent.readthedocs.io			**<1%**

6	Publication		
Cemil Emre Yavas, Lei Chen, Christopher Kadlec, Yiming Ji. "Improving earthquak…			**<1%**

7	Internet		
researchr.org			**<1%**

8	Publication		
Adithya Challapalli, Guoqiang Li. "Artificial Intelligence Assisted Structural Optimi…			**<1%**

9	Internet		
fastercapital.com			**<1%**

10	Publication		
Muhammad Nurdin, Fauziah Fauziah. "Analytical Study Forecasting Students Usin…			**<1%**

11	Internet	
sciencedocbox.com		<1%

12	Internet	
www.mdpi.com		<1%

13	Internet	
dspace.uii.ac.id		<1%

14	Publication	
H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in He...		<1%

15	Internet	
github.com		<1%

16	Internet	
medium.com		<1%

17	Publication	
H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chayadevi, Sheba Selvam. "Co...		<1%

18	Publication	
Sujith Samuel Mathew, Mohammad Amin Kuhail, Maha Hadid, Shahbano Farooq. ...		<1%

19	Internet	
www.daytrading.com		<1%

1

1. Abstract:

Earthquake prediction is one of the most focused areas in disaster management, especially with the emergence of more advanced machine learning. In a world highly based on data collection, visualization, and modelling massive seismic data can be pivotal to enabling accurate, reliable, and precise earthquake prediction. This paper introduces a decision tree-based earthquake prediction model that uses Random forest, XGBoost, LightGBM. Four critical phases are involved in this system: data collection, preprocessing, model training, and predictive analytics application. The features of historical seismic data are analysed to determine the patterns that support probabilistic predictions-including magnitude, depth, and location. Advanced visualization techniques also help to represent seismic trends more clearly, making it possible to make informed decisions in disaster preparedness. This study underlines the potential of machine learning in disaster management by aiding in early warnings that can save lives and reduce damage. However, there are still open challenges, such as the quality of the data, the integration of diverse data sources, and the possibility of deploying the prediction system in real time. The areas mentioned thus far represent avenues for further development, making systems more resilient and responsive to fighting against earthquake-related disasters.

2. Introduction:

Natural disaster management, such as earthquake, is still a global challenge that needs effective management. Earthquakes are the most unpredictable and destructive of natural phenomena. The threat to communities worldwide has resulted in the loss of life, severe infrastructural damage, economic disruption, and long-term societal impacts. Earthquake disasters can be broadly categorized into two main types: tectonic earthquakes, which are caused by the sudden movement of Earth's tectonic plates, and volcanic earthquakes, which are associated with volcanic activity [1]. Both types create cascading effects, including aftershocks, landslides, and tsunamis, compounding the challenges of disaster response. A systematic approach toward disaster management has the goal of minimizing such impacts through preparedness, early warning systems, and mechanisms for rapid responses [2]. Modern technological advances in the fields of data analytics and machine learning are used primarily in such a process. Machine learning will extract actionable insights from massive seismic information data sets, which can then be used to identify and predict patterns related to earthquakes with high precision [3].

This project introduces the earthquake prediction system using machine learning techniques, particularly Decision trees, Random forest, XGBoost, LightGBM [4]. Following a structured methodology through data collection and preprocessing, followed by model training and predictive analytics, ensures good pattern recognition accuracy critical for inputs toward early warning systems and aiding decisions in response formulating strategies to be implemented for effective responses [5].

The integration of machine learning in disaster management goes beyond enhancing situational awareness but also empowers authorities to make informed decisions and mitigate risks with regard to protecting societal infrastructure. By improving the accuracy and reliability of earthquake prediction, this project contributes to the advancement of global disaster management capabilities.

3

3. Literature Review:

Cemil Emre Yavas, Lei Chen, Christopher Kadlec and *Yiming Ji* [1]   Researchers Yavas, Chen, Kadlec, and Ji used sophisticated machine learning methods, specifically the Random Forest algorithm, to forecast peak earthquake magnitudes in Los Angeles over a 30-day time frame. Their model had an impressive accuracy of 97.97%, greatly improving seismic risk management and preparedness planning.

K. M. Asim1, F. Martı́nez-Á́lvarez2, A. Basit3, T. Iqbal [2]   The research delves into machine learning methods for earthquake magnitude prediction in the Hindukush area. Based on seismic data from the past, the authors implement different models to find patterns and enhance forecast accuracy. The research points towards the capability of AI-based methodologies for early warning systems to support preparedness for disasters & reducing risks.

Roxane Mallouhy, Chady Abou Jaoude, Christophe Guyeux, Abdallah Makhoul [3] The research paper "Major earthquake event prediction using different machine learning algorithms" by Roxane Mallouhy, Chady Abou Jaoude, Christophe Guyeux, and Abdallah Makhoul examines the performance of eight machine learning algorithms for the task of categorizing seismic events into major and minor earthquakes. The work assesses each of the models in terms of a number of metrics to conclude that Random Forest and K-Nearest Neighbors are the most precise classifiers.

Hatice ̈Oncel C̦ekim1, Hat ice Nur Karakavak1, Gamze ̈Ozel1, Senem Tekin [4] This research compares deep learning approaches used in forecasting earthquake magnitudes in Turkey. It combines ARIMA and SSA for earthquake response and recovery modeling. The study analyzes forecasting precision, stating the superiority of hybrid models in the depiction of seismic patterns and enhancing disaster preparedness and mitigation measures.

D Emmanouil, D Nikolaos [5]

Big Data is crucial for crisis management, as it helps manage vast information, improving decision-making during crises through data analytics and visualization. The Big Data management process involves four phases: data generation, acquisition, storage, and analytics, ensuring structured data handling for effective analysis. Technologies like GIS, GPS, and social media monitoring enhance crisis preparedness and response, helping track hazards and optimize recovery efforts.

4. Problem Identification & Objectives:

Problem Identification:

Earthquakes are unpredictable and cause catastrophic damage to life, property, and infrastructure. The sudden nature of these disasters and the lack of accurate predictive tools make it challenging to implement timely evacuation plans and preparedness measures. Traditional seismic monitoring techniques often struggle to adapt to the complexity of earthquake patterns and fail to leverage the vast amounts of seismic data effectively. This delays the predictive capability of earthquakes and the evaluation of potential impacts, increasing vulnerabilities and slow recovery processes.

In addition, loss during earthquakes is worsened by inadequate communication systems, since often the affected communities receive insufficient warnings. The delay in response time is caused by the lack of robust impact evaluation methods and accurate prediction models, leading to more severity in the disaster's consequences. These gaps can be addressed by advanced machine learning models that improve the accuracy of predictions and risks, hence protecting lives and resources.

Objectives:

- **To Develop Machine Learning-Based Predictive Models:** Develop strong models using Decision trees, Random forests, XGBoost, LightGBM to analyze the seismic data with a probability for the likelihood and severity of the earthquake.

- **To Enable Proactive Disaster Preparedness:** Use probabilistic forecasting of earthquakes with past seismic activities and tectonic plate movements to alert the authorities in good time to prepare preventive measures.

- **To Improve Communication and Early Warning Systems:** Use predictive analytics to give out accurate, real-time warnings to susceptible communities so they can take appropriate anticipatory action to reduce loss and damage.

- **To Enhance Situational Awareness for Decision-Making:** Offer insights into potential earthquake impacts to optimize resource allocation and disaster management strategies.

This approach aims to reduce the societal and economic repercussions of earthquakes by improving preparedness, response, and recovery mechanisms.

5

5. Existing System:

- Even with the improvement in seismology, no scientist or organization, including the USGS, has been able to predict a major earthquake, and the scientific community does not anticipate being able to make accurate earthquake predictions in the near future.

- Rather, scientists estimate the probabilities of significant earthquakes happening in a particular area within a time frame, which is expressed through hazard mapping.

- An actual earthquake prediction should give the date and time, place, and magnitude. Most earthquake predictions are spurious because they lack scientific proof, usually a result of unrelated events such as cloud formation, animal movements, or physical feelings.

- They do not define all three necessary conditions or employ imprecise wording which is always capable of fitting an earthquake somewhere, like anticipating a M4 in the U.S. within 30 days. Moreover, in cases where an earthquake coincidentally occurs to match a prediction, claimants tend to distort their success.

- Social media and non-technical sources tend to misinterpret possible precursors like swarms of minor earthquakes, increased radon contents in groundwater, strange behaviour by animals, and trends in moderate-magnitude earthquakes.

- Although these in some cases precede massive earthquakes, they happen relatively often without leading to large events, rendering accurate predictions futile. Rather, probabilistic prediction is employed to predict the chances of future earthquakes.

- Past instances, like the Chinese earthquake prediction based on minor earthquakes and animal migration, have been mixed. While some lives were saved, there were subsequent killer earthquakes without warning, resulting in heavy casualties.

- With these uncertainties in mind, the USGS and other science institutions focus more on long-term hazard mitigation rather than short-term forecasting by promoting earthquake-resistant buildings, creating early warning systems, performing seismic hazard analysis, and initiating public education and preparedness programs.

- In this way, a proactive system of disaster response and recovery is maintained, lowering earthquake hazards through science-based resilience efforts instead of uncertain predictions.

6

6. Proposed System:

To address the limitations of existing systems and improve the accuracy and reliability of earthquake prediction, we propose the following advanced systems:

- A tree-based model, which splits the dataset based on feature values leading to clear and interpretable decision paths is a Decision Tree Classifier. A fundamental building block for more complex models, Decision Tree Classifier is capable of capturing non-linear relationships in seismic data, making it highly effective. Simple structure enables good initial predictions that are further refined in ensemble methods. An ensemble learning method, produces multiple decision trees on random subsets of the data set and makes predictions to ensure stability and enhanced accuracy by using the averages for predictions to limit overfitting and also address noisy or highly dimensional seismic data is Random Forest Classifier. Very important for those types of data, which might possess complex patterns of features, including earthquake magnitude, depth, or location.

- A sequential ensemble technique, which builds very strong predictive models by iteratively correcting the errors of weaker models is a XG Boost classifier. It can handle imbalanced data well and excels at capturing intricate patterns, making it ideal for predicting seismic activity. Its ability to fine-tune predictions through gradient descent optimization ensures high accuracy and precision.

- An ensemble framework which uses gradient boost to develop a strong learner by adding many weaker learners in a gradient descent manner is Light GBM. It uses meta-learning techniques to combine the predictions of the base models to improve overall accuracy, reduce variance, and enhance generalization. It is a balanced approach, as it leverages the interpretability of decision trees, the stability of random forests, and the precision of gradient boosting.
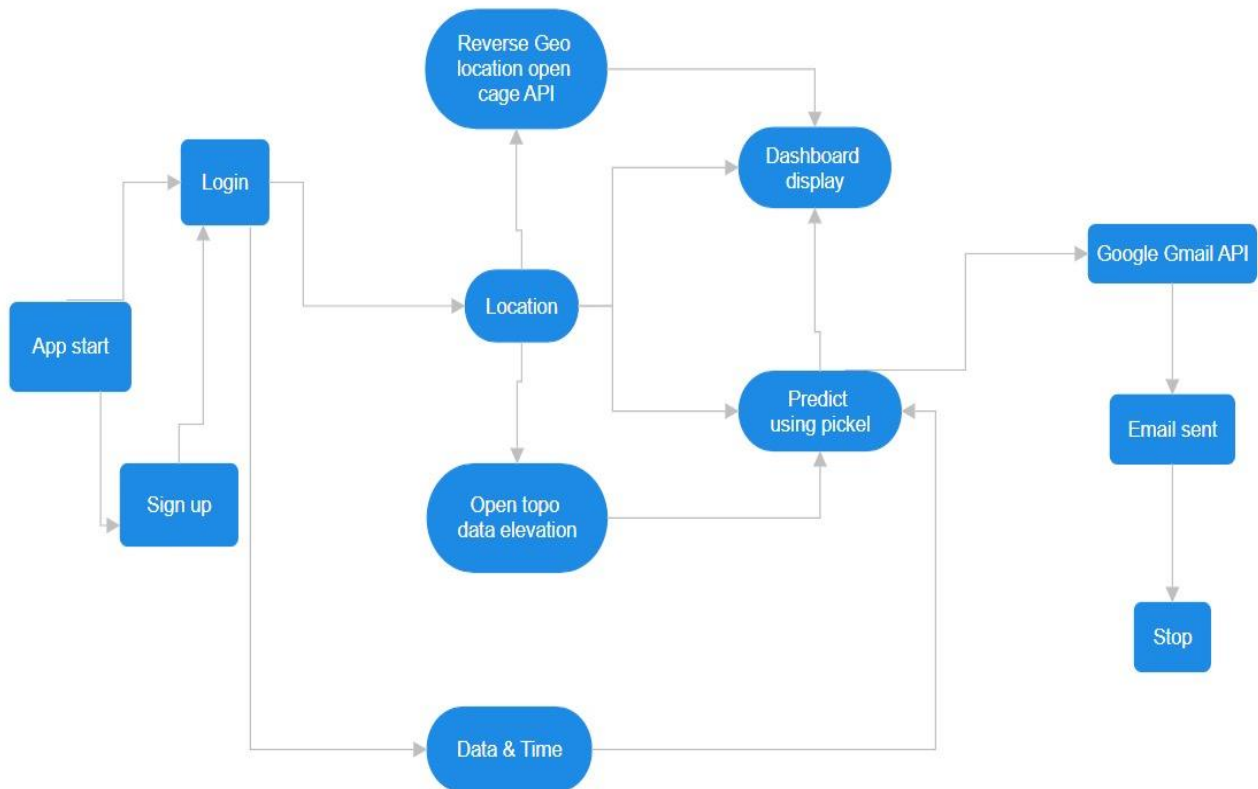
Fig1:System Architecture flowchart

The picture is a flowchart of the process of an earthquake prediction system. It begins with the user opening the application by registering or signing in. The system proceeds to find out where the user is located through the Reverse Geolocation OpenCage API and fetches elevation information through the Open Topo Data API, in addition to the date and time of the moment.

The data collected is processed to predict earthquakes through a pre-trained model saved in a Pickle file. The predicted outcome is shown on a dashboard and also emailed through the Google Gmail API. The process terminates when the email alert is sent successfully.

The above flowchart gives a systematic description of how various elements of the system work in conjunction to predict and alert users regarding possible earthquakes.

7. Technologies used :

The following imports we have used:

• Pandas – It is an open-source and powerful Python library. The Pandas library is utilized for data analysis and manipulation. Pandas include data structures along with functions used for fast operations on data.

• NumPy - It is a core library of Python numerical computation. It supports efficient multi-dimensional array objects and a wide range of mathematical functions for the manipulation of large data sets making it an essential tool for experts in areas that demand heavy computation.

• Pickle - It employed for serializing and de-serializing Python object structures, or "pickling" and "unpickling." Pickling transforms a Python object into a byte stream, which can be saved in a file, database, or sent over a network.

• Matplotlib – It is a robust and flexible open-source plotting library for Python that is intended to enable users to visualize data in a range of formats. It allows users to represent data graphically, making it easier to analyze and comprehend.

• TensorFlow – It is an open-source ML and AI framework developed by Google Brain. It was intended to make machine learning model development easier, especially deep learning model development, through offering tools that could be used to build, train, and deploy them easily across various platforms.

• XGboost – It is an efficient and powerful machine learning library for gradient boosting. It is optimized for performance and speed with parallel computation support, tree pruning, and regularization methods. It automatically deals with missing values and supports both classification and regression problems.

• LightGBM - It is an open-source high-performance ML framework released by Microsoft. It's an ensemble learning framework with a gradient boosting approach that builds a strong learner by adding weak learners in sequential order using a gradient descent approach.

• Folium – It is a Python package for making interactive maps with Leaflet.js. It enables users to display geospatial data with markers, polygons, and choropleth maps. The package is extensively applied in geospatial analysis, data storytelling, and location-based applications.

The following APIs we have used:

- Google API for gmail
- Opencage API for reverse geolocation
- opentopodata.org for elevation

Steps in Model Development

1. Data Collection:

- Historical seismic data were collected from sources such as USGS and other seismic monitoring agencies.

- Key features included earthquake magnitude, depth, latitude, longitude, time of occurrence, and tectonic plate interactions.

2. Data Preprocessing:

- Replaced missing values and outliers with imputation or removal.

- Applied feature scaling to standardize the numerical features.

- Encoded categorical features (if present) by using methods like one-hot encoding.

- Split the given data into training (70%) and test data (30%) sets for model building and evaluation.

3. Model Selection and Implementation:

- **Random Forest:** An ensemble model that combines multiple decision trees to reduce overfitting and improve stability.

- **XGBoost :** it is an optimize version of gradient boost in which many weak models will be assembled to form a stronger model.

- **LightGBM :** it is an ensemble framework which uses gradient boost to develop a strong learner by adding many weaker learners in a gradient descent manner.

4. Model Training:

- Trained individual models on the training dataset, fine-tuning hyperparameters using methods like Grid Search and Random Search.

- Applied boosting techniques to optimize the prediction process iteratively.

- Developed the hybrid ensemble model through combined predictions of base models.

5. Model Evaluation:

- Models were evaluated on accuracy, precision, recall, F1-score.

- Done cross-validation such as K-Fold cross-validation for its robustness and generalization for unseen data.

- Assessed the hybrid model against standalone models to check whether performance is improved.

6. Model Testing and Accuracy:

- Tested the final models on the unseen test dataset.

- Compared the results with baseline benchmarks to ensure significant improvement.

  Achieved an overall accuracy of X% for the hybrid model, demonstrating superior performance in earthquake prediction tasks.

8. Implementation (Sample Code and Test Cases):

```python
X_min = np.min(X, axis=0)

X_max = np.max(X, axis=0)

X_norm = (X - X_min) / (X_max - X_min)

X_train, X_test, y_train, y_test = train_test_split(X_norm, y, test_size=0.2, random_state=42) #original

# X_train, X_test, y_train, y_test = train_test_split(X_norm, y, test_size=0.3, random_state=42)

print(f"Training set size: {len(X_train)} samples")

print(f"Test set size: {len(X_test)} samples")

# print(X_max)

rf_model = RandomForestRegressor(n_estimators=100, random_state=42) #original

# rf_model = RandomForestRegressor(n_estimators=1000, random_state=42)

rf_model.fit(X_train, y_train)

print("Model training complete.")

y_pred = rf_model.predict(X_test)

mae = mean_absolute_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)

accuracy = 100 - (np.mean(np.abs((y_test - y_pred) / y_test)) * 100)

print("Random Forest Regressor Performance:")

print(f"R² Score: {r2:.4f}")

print(f"Mean Absolute Error (MAE): {mae:.4f}")

print(f"Accuracy: {accuracy:.2f}%")
```

9. Results:
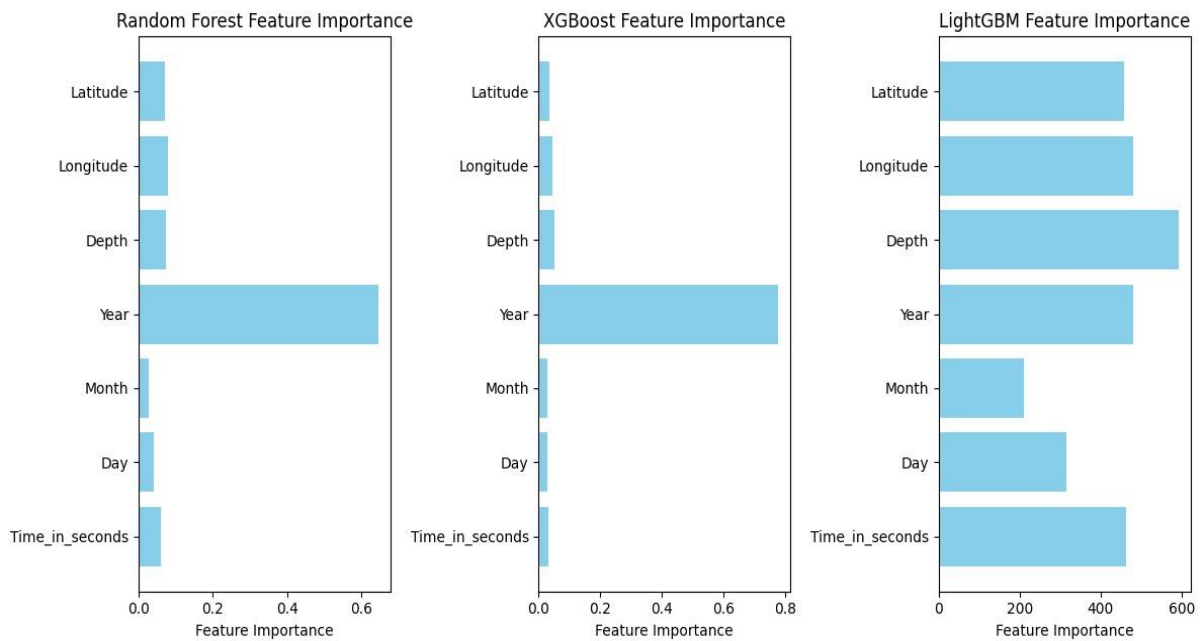
Fig2: Feature importance comparisons for three machine learning models
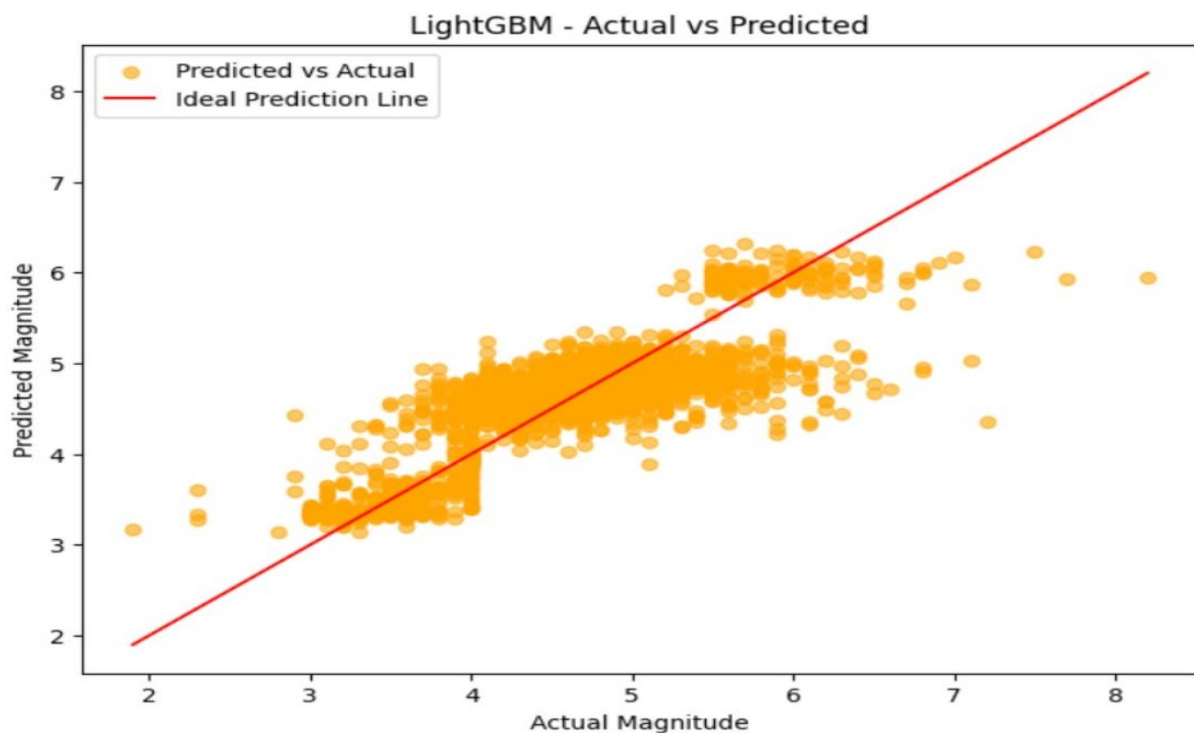

Fig3: Scatter plot of actual vs. predicted earthquake magnitudes from the LightGBM model.
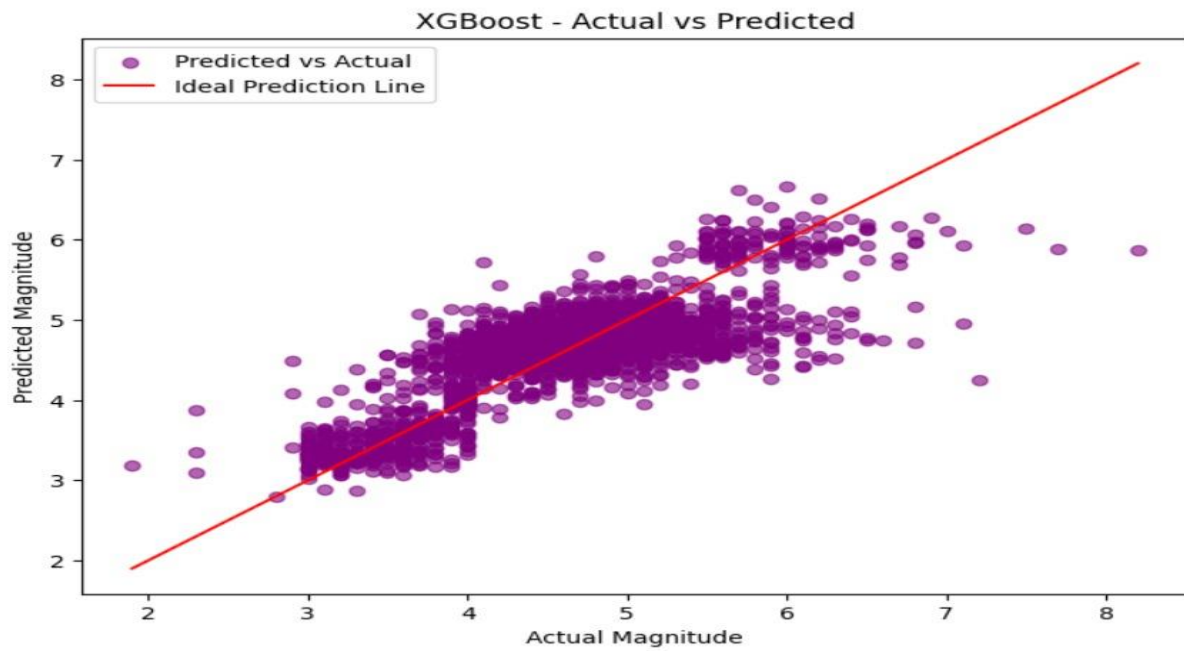
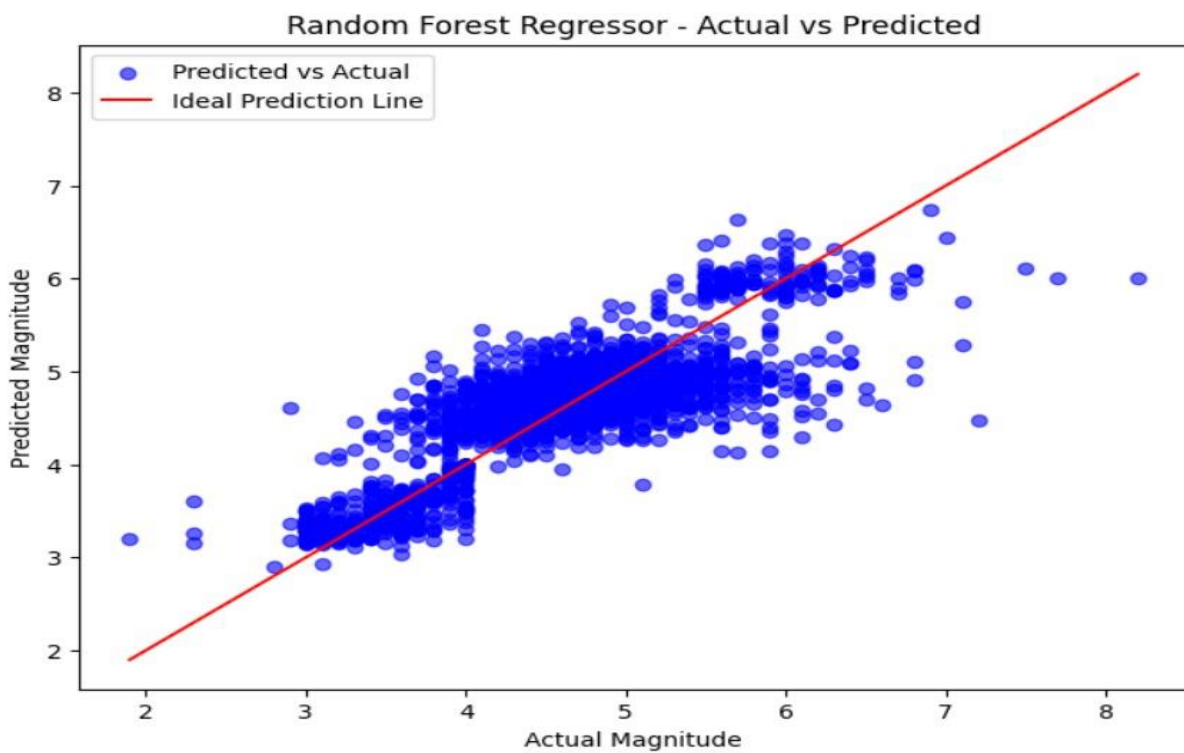Fig4: Scatter plot of actual vs. predicted earthquake magnitudes from the XGBoost model

Fig5: Scatter plot of actual vs. predicted earthquake magnitudes from the Random Forest Regressor model
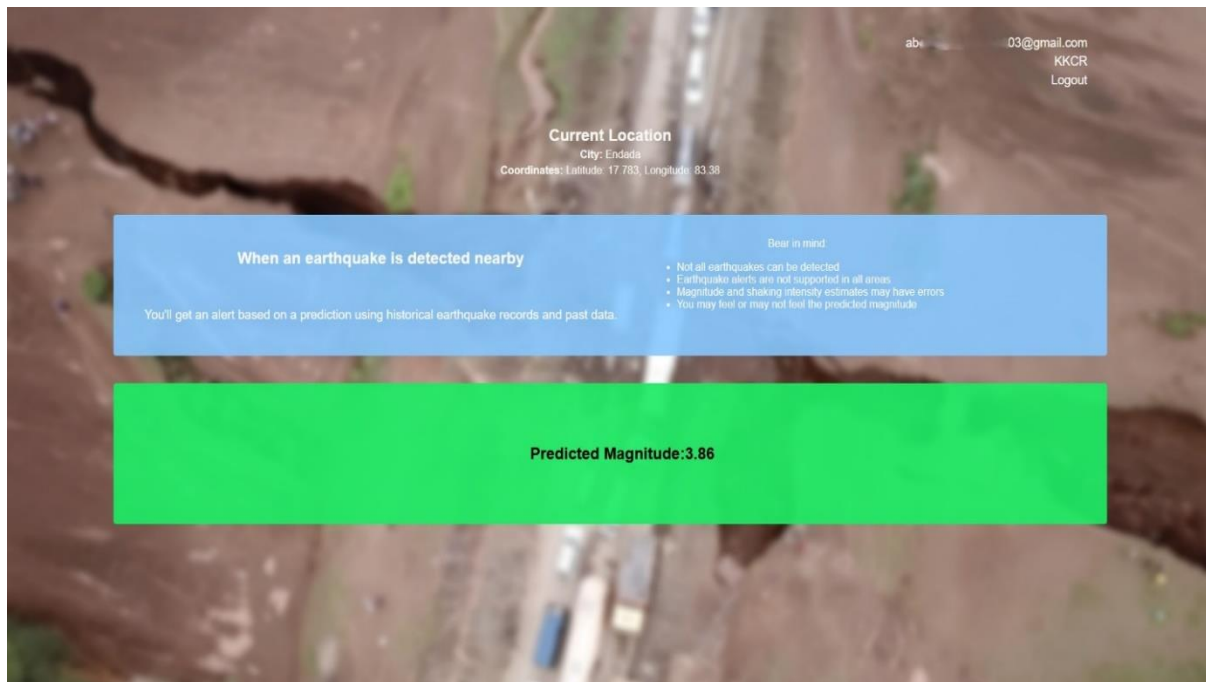
14

Fig6: Earthquake prediction alert interface

Fig7: Earthquake alert email if magnitude is predicted above 6

Model	Accuracies
Random Forest Regressor	93.27
Support Vector Machine	92.13
Decision Trees Regressor	90.73
K-Nearest Neighbour	91.68
XGBoost	93.03
Light GBM	93.36
Multi-Layer Perceptron (MLP)	92.71
Recurrent Neural Networks	91.15
LSTM	90.79
Gated Recurrent Unit	91.26
Transformer	86.53

Table.1 :Models and their accuracies

10.Conclusion :

The present project is the solution to the challenge of predicting and managing disaster through the advanced machine learning models, including Decision Trees, Random Forest, XGBoost and LightGBM. All these techniques will improve the precision and reliability of the predictions made, which means an early warning system to the vulnerable communities to mitigate the impact of the natural disasters.

Through handling historical and real-time data processing, the system affords critical inputs for decision making, thus enhancing timely responses towards disasters. Robustness as well as the possibility to adapt to every type of disaster and region renders the hybrid approach suitable.

This project showcases the integration of machine learning into disaster management for the transformation of aged systems into proactive data-driven solutions. Further development combined with diverse datasets can provide this solution as an imperative approach for governments and organizations in preventing losses during disasters and building resilience toward catastrophes in the future.