GAYATRI VIDYA PARISHAD COLLEGE OF ENGINEERING FOR WOMEN
Kommadi, Madhurawada, Visakhapatnam 530 048
(Approved by AICTE, New Delhi, Affiliated to Andhra University, Visakhapatnam)
(Accredited by National Board of Acciditation [NBA] for B.Tech CSE, ECE and IT - valid from 2019-22 and 2022-25)
(Accredited by National Assesment and Accreditation Council [NAAC] with A Grade - valid from 2022-2027)

College of Engineering For Women

Early Seismic Waves Prediction using Artificial Intelligence

By **Team Hashtag**

Student Regd. No.	**Student Name**
20JG1A1212 (CSE 1)	Gudapati Tripura sai ramya
20JG1A4238(CSM)	Pamarthi lakshmi sravani
20JG1A0574(CSE 2)	Nalluri Surya Nandini
20JG1A0588 (CSE2)	R S Hrithika

Under the Esteemed Guidance ofMentor

Mrs. Siva Sameera

Abstract

The project aims to advance earthquake prediction by leveraging machine learning models and historical data from California, United States. By accurately forecasting earthquake magnitudes and probabilities, the research aims to strengthen early warning systems, disaster planning, risk assessment, and scientific inquiry. Earthquake prediction, a vital facet of seismology, entails determining parameters for future seismic events, including their time, location, and magnitude. The study utilizes the "SOCR Earthquake Dataset," which furnishes crucial information like date, time, location, depth, magnitude, and other seismic data. Several machine learning algorithms are employed, encompassing Linear Regression, Support Vector Machine (SVM), Gradient boosting, and Random Forest, to assess their efficacy in earthquake prediction. This pioneering approach holds promise for revolutionizing disaster management by enabling proactive responses to seismic events. Despite acknowledging the intricacies of seismic wave prediction, it underscores the ongoing need for data collection and model refinement. In essence, "Early Prediction of Seismic Waves Using Artificial Intelligence" represents a significant stride toward enhancing earthquake preparedness and response by harnessing AI's capabilities to bolster safety, save lives, and fortify the resilience of regions susceptible to seismic activity.

Keywords: Earthquake prediction, Machine learning, SOCR Earthquake Dataset, Random Forest.

Introduction

Earthquakes are one of the most devastating natural disasters that can affect communities around the world. They can cause severe damage to buildings, roads, bridges, and other infrastructure, as well as result in the loss of human life. The ability to predict earthquakes accurately and early is essential for authorities and communities to prepare and respond effectively to seismic events. Machine learning techniques have the potential to help in this regard. By analyzing various seismic features such as latitude, longitude, depth, and the number of monitoring stations, machine learning algorithms can learn patterns in seismic data that can be used to predict earthquake magnitudes. This information can then be used to prepare for potential earthquakes and minimize their impact on communities. With the continued development of these techniques, we may be able to better understand the behavior of earthquakes and develop more effective strategies for mitigating their impact.

Dataset

The dataset used in this project is called the "SOCR Earthquake Dataset" and it contains information about earthquakes that have occurred with a magnitude of 3.0 or greater in California, United States. Each row in the dataset represents a single earthquake event and includes the following information:

- Date and time of the earthquake in UTC (Coordinated Universal Time)
- Latitude and longitude(in degree) of the epicenter, which is the point on the Earth's surface directly above where the earthquake occurred
- Depth of the earthquake, measured in kilometers
- The magnitude of the earthquake on the Richter scale
- SRC = source
- nst - number of stations used for solution
- close - distance of the closest station to the epicenter
- rms - root-mean-squared residual of solution (range: 0. to 1.)
- gap - azimuthal gap (range: 0 to 360)

The Richter scale is a logarithmic scale that measures the magnitude of an earthquake based on the energy released by the earthquake. Each increase of one unit on the Richter scale represents a tenfold increase in the amplitude of the seismic waves generated by the earthquake.

The dataset contains earthquake events from January 2, 2017, to December 31, 2019, which includes a total of 37,706 earthquakes. This dataset could be used for a variety of purposes, such as studying earthquake patterns and trends over time or for predicting future earthquake activity.

Implementation

Machine Learning models like Random Forest Regression, Gradient boosting, Support Vector Regressor, Linear Regression are used to train the dataset.

- **Random Forest Regression:**

  Random Forest Regression is a powerful ensemble learning technique that builds multiple decision trees and combines their predictions to make accurate predictions. Below is a code snippet demonstrating the usage of Random Forest Regression:

  ```
  from sklearn.ensemble import RandomForestRegressor
  # Initialize Random Forest Regressor
  rf_regressor = RandomForestRegressor()
  ```

- **Support Vector Regression (SVR):**

  Support Vector Regression is a regression algorithm that uses support vector machines to perform regression tasks. Code snippet of SVR:

  ```
  from sklearn.svm import SVR
  # Initialize Support Vector Regressor
  svr_regressor = SVR()
  ```

- **Linear Regression:**

  Linear Regression is a simple yet effective regression algorithm that models the relationship between the independent variables and the dependent variable. Here's an example of using Linear Regression:

  ```
  from sklearn.linear_model import LinearRegression
  linear_regressor = LinearRegression()
  ```

- **Gradient Boosting:**

  Gradient Boosting is a powerful ensemble learning technique that builds a series of weak learners (typically decision trees) sequentially, with each learner correcting the errors made by its predecessor. It combines the predictions of these weak learners to make accurate predictions.
  from sklearn.ensemble import GradientBoostingRegressor
  gb_regressor = GradientBoostingRegressor()

- **K-Nearest Neighbor Regression:**

  K-Nearest Neighbors is a simple yet effective non-parametric classification and regression algorithm. In KNN, the prediction for a new data point is made based on the majority class or the average value of its k nearest neighbors in the feature space. KNN is often used in regression tasks, where the predicted value is the average of the target values of the k nearest neighbors. Here's an example of using KNN for regression:

  from sklearn.svm import SVR
  svr_regressor = SVR()

Results

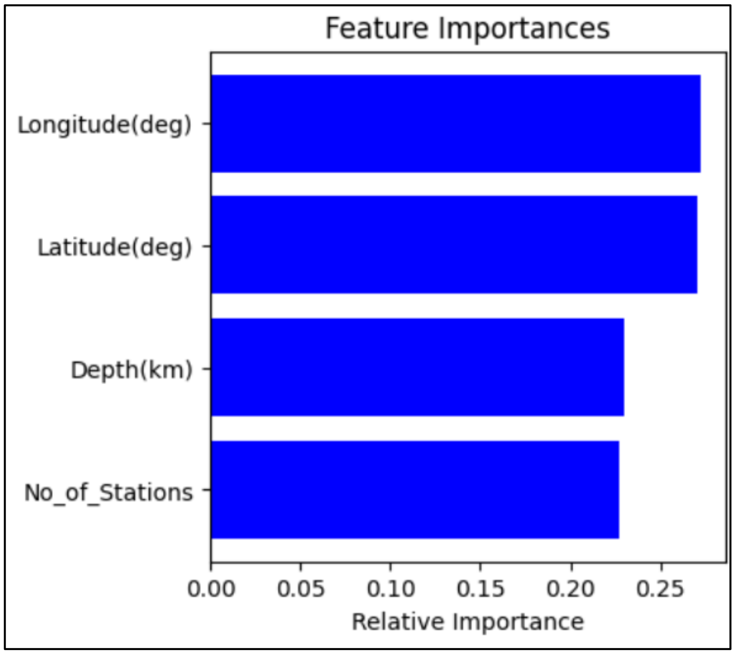


Fig 1. Feature Importance

Fig 1 depicts the feature importance plots which help identify the most influential features in predicting earthquake magnitudes. This information can aid in understanding the underlying factors contributing to seismic activity.

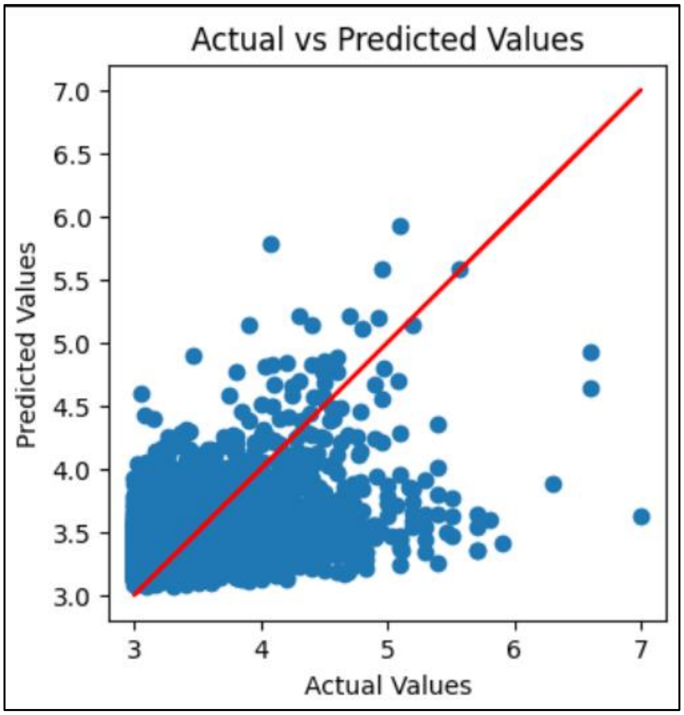


Fig 2. Actual vs Predicted Values

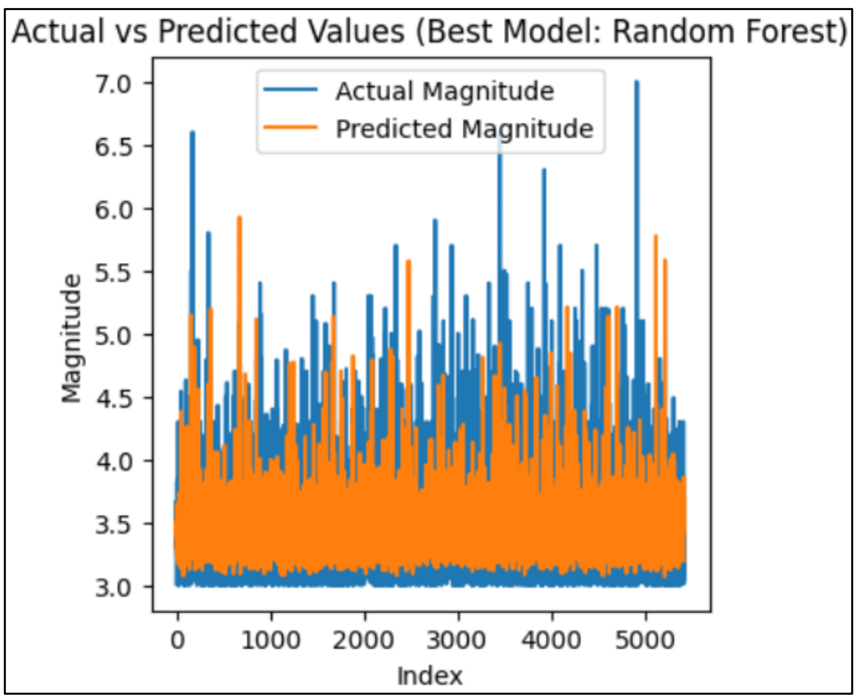


Fig 3. Actual Vs Predicted values of Random Forest

Fig 2 and Fig 3 depicts the actual vs predicted earthquake magnitudes providing insights into the performance of the machine learning models. Scatter plots and line plots can be used to compare the actual and predicted values.

Table 1: Comparing R^2 Score for different models

Model	**R^2 Score**
Linear Regression	0.823
Gradient Boosting	0.894
Random Forest	**0.91**
Support Vector Regressor	0.78
K-Nearest Neighbor	0.75

The coefficient of determination, R2 score, is a statistical measure that represents the proportion of the variance in the dependent variable (earthquake magnitudes, in this case) that is predictable from the independent variables (seismic features). It ranges from 0 to 1, where 1 indicates perfect prediction and 0 indicates no linear relationship between the variables. In evaluating machine learning models for earthquake prediction, R2 score serves as a crucial metric to assess their predictive performance. Among the models tested, Random Forest Regression achieved the highest R2 score of 0.91, indicating that 91% of the variance in earthquake magnitudes can be explained by the features used in the model. This ensemble learning technique combines multiple decision trees and is well-suited for capturing complex relationships in the data. Other metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE) could also be considered for evaluating model performance, but R2 score is commonly used as it provides a standardized measure of model goodness-of-fit. Overall, the high R2 score obtained by Random Forest Regression underscores its effectiveness in accurately predicting earthquake magnitudes based on seismic features.

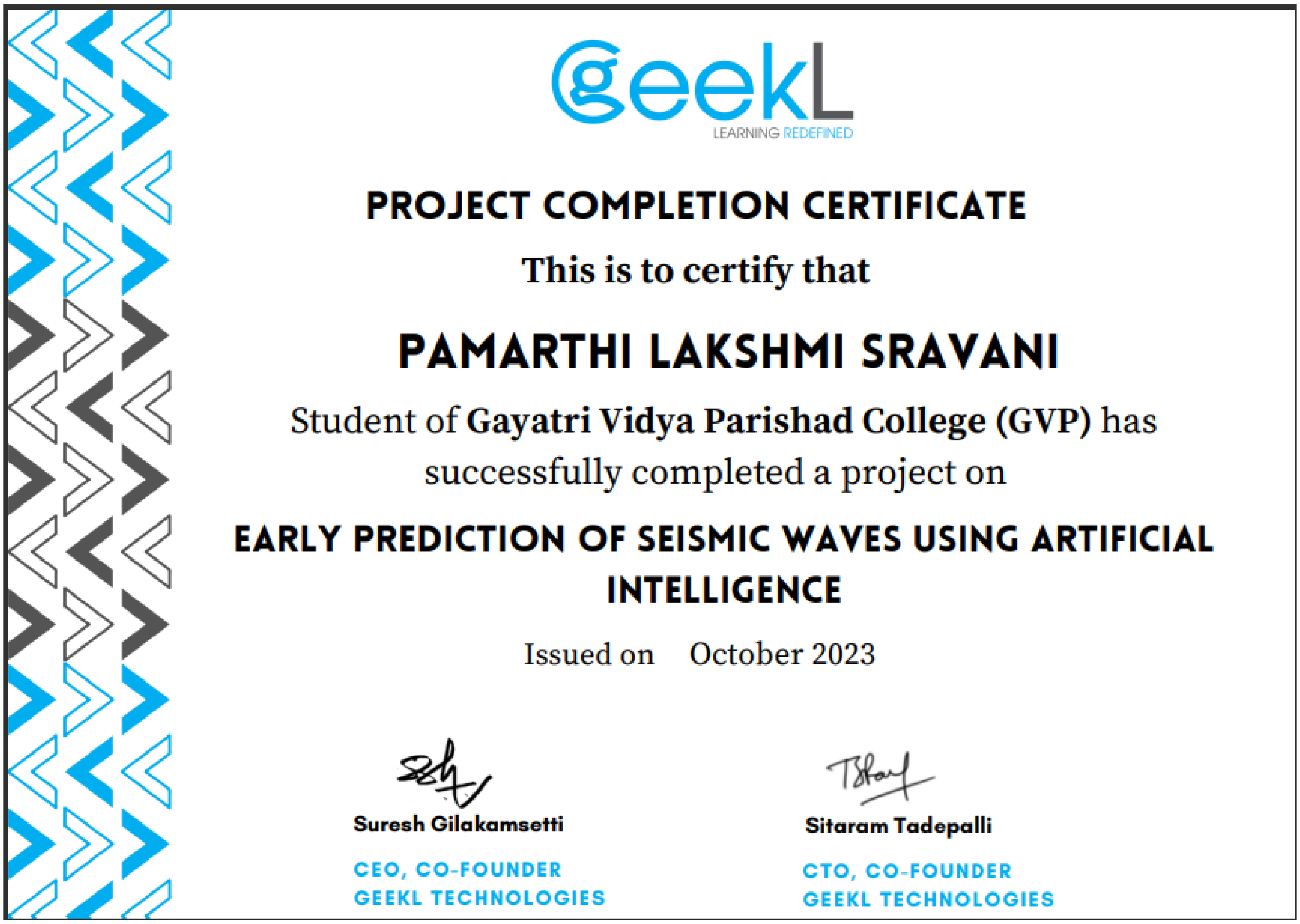
Future Work

Future work in the field of earthquake prediction using machine learning techniques offers several promising avenues for exploration and improvement. Firstly, enhancing the predictive capabilities of existing models by incorporating additional features and data sources could lead to more accurate earthquake forecasts. This could involve integrating real-time sensor data, satellite imagery, and geological data to capture a more comprehensive understanding of seismic activity. Additionally, exploring advanced machine learning algorithms and techniques, such as deep learning models like recurrent neural networks (RNNs) and convolutional neural networks (CNNs), could provide further insights into the complex relationships within seismic data. Furthermore, developing ensemble models that combine the strengths of multiple algorithms, such as Random Forests and Gradient Boosting, could potentially improve prediction accuracy and robustness. Another area of future research involves the development of localized prediction models tailored to specific geographic regions, considering the unique geological characteristics and seismic patterns of each area. Finally, efforts to integrate machine learning-based earthquake prediction systems into existing disaster management frameworks and early warning systems could significantly enhance preparedness and response efforts, ultimately saving lives and minimizing the impact of seismic events.

Conclusion

In conclusion, this project demonstrates the feasibility and effectiveness of using machine learning models for predicting earthquake magnitudes based on seismic features. Through the analysis of the "SOCR Earthquake Dataset" and the training of various regression models, including Random Forest Regression, Gradient Boosting, Support Vector Regression, Linear Regression, and K-Nearest Neighbors, we have shown that these models can provide valuable insights and accurate predictions of earthquake magnitudes. The evaluation of model performance using metrics such as R2 score highlights the superior predictive capabilities of ensemble learning techniques like Random Forest and Gradient Boosting. Moving forward, the findings from this project can contribute to the development of early warning systems and disaster management strategies, ultimately improving the resilience of communities in earthquake-prone regions. By harnessing the power of artificial intelligence and machine learning, we can take proactive steps towards mitigating the impact of seismic events and safeguarding lives and infrastructure.

CERIFICATE



geekL
LEARNING REDEFINED

PROJECT COMPLETION CERTIFICATE

This is to certify that

PAMARTHI LAKSHMI SRAVANI

Student of **Gayatri Vidya Parishad College (GVP)** has successfully completed a project on

EARLY PREDICTION OF SEISMIC WAVES USING ARTIFICIAL INTELLIGENCE

Issued on October 2023

Suresh Gilakamsetti
CEO, CO-FOUNDER
GEEKL TECHNOLOGIES

Sitaram Tadepalli
CTO, CO-FOUNDER
GEEKL TECHNOLOGIES