Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

ANALYSIS AND PREDICTION OF EARTHQUAKE USING MACHINE LEARNING ALGORITHMS

Mrs.M.Padmaja Assistant Professor, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India. padmaja.malarowthu@raghuenggcollege.in
T. Lakshmi Lahari, V. Shiva Jyothi, Y. Daya Sagar Students, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India.
19981a05g0@raghuenggcollege.in, 19981a05h6@raghuenggcollege.in, 19981a05i3@raghuenggcollege.in

Abstract-*Two fundamental types of earthquake forecasts exist: forecast predictions and short-term predictions. In contrast to long-term forecasts, which are made months to years in advance, short-term predictions are made just hours or days in advance. The main objective of this research is to identify whether a significant classification of an earthquake as a bad or good occurrence will be done using different machine learning algorithms. On a real seismic dataset, several algorithms have been applied, including Random Forest Classifier, Naive Bayes, Logistic Regression, Multilayer Perceptron, Adaptive Boosting, K-Nearest Neighbours, and Support Vector Machine. A variety of hyperparameters were selected for each model, and the resulting prediction results were fairly compared using a variety of metrics. As a consequence, for prediction, the algorithm with the highest accuracy is used.*

Keywords: ***Random Forest Classifier, Naive Bayes, Logistic Regression, Adaptive Boosting, KNN, Support Vector Machine, and Multi-Layer Perceptron Classifier.***

I.INTRODUCTION

One of the most devastating natural disasters is an earthquake, which results in a large number of fatalities and property losses [12]. The ability to predict earthquakes can help mitigate their impact on society by providing early warning and allowing for the implementation of appropriate safety measures. This research uses a variety of machine learning algorithms to predict earthquake events based on various factors such as geological features, seismic activity patterns, and geographical location. The study also explores the use of feature selection techniques and model evaluation methods to develop an accurate earthquake prediction model [13]. The research assesses the effectiveness of each algorithm and identifies key characteristics that are crucial for earthquake prediction.

The remainder of the essay is organized as follows: Using machine learning algorithms, Section 2 reviews relevant research on earthquake prediction. In Section 3, the study's approach and algorithms are described. Section 4 of the research discusses the mathematical concepts used in this work, how different machine learning algorithms performed, and the key characteristics that were discovered for earthquake prediction. Finally, Section 5 discussed the results.

II.RELATED WORK

Because earthquakes have a random element, it may never be feasible to pinpoint their precise timing, despite the best efforts of researchers. In this study, earthquake magnitude prediction for the Hindukush area was done by combining machine learning classifiers with the temporal sequence of historical seismic activity. Every machine learning classifier is also subjected to a study of earthquake prediction findings in predictive values. Another success indicator taken into account when analyzing the outcomes is accuracy. [1]

Another study is based on self-organization in disaster response. With a focus on emergency medical services, this research examines the process of self-organization in the larger disaster response system for this earthquake. Effective disaster response depends on developing a community's capacity for quick transition in the case of a disaster. There is evidence of spontaneous reallocation of resources and reorganization of action to satisfy

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

urgent human requirements in every disaster environment. [2]

Another investigation done through a thorough examination of building damage detection due to earthquakes using remote sensing methods is highlighted by Laigen Dong and Jie Shan. In this paper, we address this issue in two novel steps, utilizing a fresh two-step process. [3].

A technique done by Guillaume Royer, Christophe Guyeux, and Jean-François Couchot This research is about data on locations that were initially anonymized using a local differential privacy method. The reconstruction is a collection of artificial data that is decoupled from the users and then done using statistical estimators. Finally, the forecasts were made using a supervised learning strategy that employed extreme gradient boosting. [4].

III. METHODOLOGY

Several machine learning algorithms were put forth to determine whether an earthquake was likely to occur or not. No study has properly addressed the issue of misdiagnosis, whether from an earthquake or not. Therefore, we propose a Random Forest Classifier and compare it with algorithms like support vector machines, Naive Bayes, logistic regression, adaptive boosting, and multi-layer perceptron to predict. The datasets are taken from Kaggle in the format of CSV files.

In this context, we design a web page that contains the Home, About, user registration, and login details. After logging in successfully, we have to upload datasets that were downloaded from Kaggle. Next, we have to pre-process the data to get accurate results. The data is split into a training set and a testing set. Datasets are divided into 70% for training and 30% for testing. By testing, we determine the accuracy of all algorithms. Based on the accuracy, the one with the highest accuracy is used for prediction. Different parameters are considered for the prediction to be accurate.

Different features that are considered for this work and are available on datasets are latitude, longitude, azimuthal gap, depth, depth error, etc. These values are taken from datasets, and submitting them gives the result of whether there is a chance of an earthquake occurring or not. We have to create a web page using HTML.

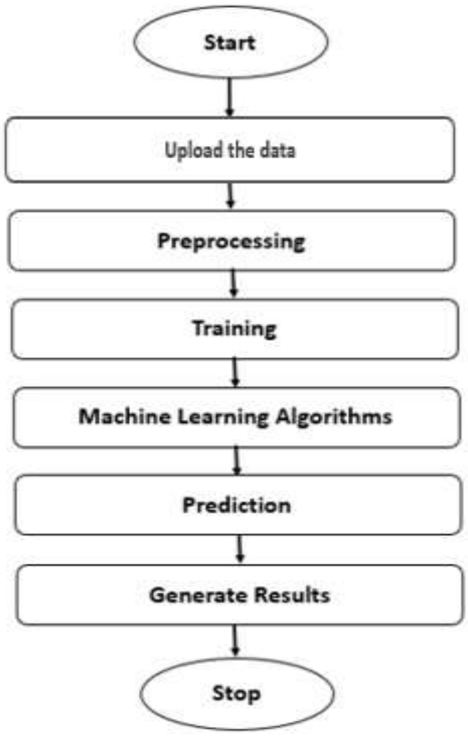

Fig1: Block Diagram

The datasets that are taken from Kaggle are pre-processed to get accurate values. Empty spaces or columns are filled with values using the fill function and by calculating the root mean square.

Fig2:Datasets

ALGORITHMS USED

1. Random Forest Classifier:

A random forest classifier is a type of supervised learning algorithm used in machine learning. It works by creating multiple decision trees. Each decision tree in the forest is constructed using a random subset of the training data and a random subset of the features. During prediction, the Random Forest Classifier outputs the class predicted by the majority of the trees in the forest. [5]

2. Logistic Regression

It is a popular algorithm used in machine learning for binary classification problems. The algorithm aims to predict a binary outcome (0 or 1) based on a set of independent variables or features.

The algorithm works by first computing a weighted sum of the features using a set of coefficients, or weights. After calculating the weighted sum, it is fed into a sigmoid function that assigns a value between 0 and 1 to the output. This resulting value denotes the anticipated probability of the positive class. [6]

3. Naive Bayes classifier

Naive A well-liked method in machine learning for classification issues is called Bayes. The algorithm computes the product of the prior probabilities and the conditional probabilities for each characteristic, given the new data point, in order to categorize the new data point. It selects the highest accuracy class as the predicted class. The Bayes theorem has the following formula:

P(Y|X) = P(X|Y) * P(Y) / P(X), where:
P(Y|X) is the probability of class Y given input X.
P(X|Y) is the probability of input X given class Y.
P(Y) is the prior probability of class Y.
P(X) is the probability of input X. [7]

4.AdaBoost Classifier**:**

The Adaboost (short for adaptive boosting) algorithm works by iteratively training a set of weak classifiers on a training dataset and combining them to form a strong classifier. Errors that are not rectified by the first step are forwarded to the next phase to be cleared. [8]

5. KNN Classifier

It is a supervised learning technique. The KNN method can be applied to both classification and regression. The Euclidean distance between the two points will then be determined [9]. It can be calculated as:

The Euclidean distance between two points is

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

6. Support Vector Machine

SVM is a supervised machine learning algorithm used for both classification and regression. SVM works by finding a hyperplane in a high-dimensional space that maximally separates different classes of data points. [10]

7. Multi-Layer Perceptron Classifier

A type of feedforward artificial neural network is called a multilayer perceptron. A minimum of three levels of nodes make up an MLP: the input layer, the hidden layer, and the output layer. Backpropagation is a supervised learning method that the multilayer perceptron uses in its training process. [11]

IV.MATHEMATICAL CONCEPTS

By testing the performance of ML algorithms in earthquake prediction, the dataset is typically split into training and testing sets. The algorithm is trained on the training set and evaluated on the testing set using various evaluation metrics such as accuracy and F1-score. Root mean squared error can also be calculated to measure the algorithm's predictive performance. Accuracy is the proportion of correctly classified samples, and the formula is defined as the number of correct predictions divided by the total number of predictions.
Root Mean Square Error is,
RMSE = sqrt((1/n) * (yi - i)2)
It's important to note that these evaluation metrics can be affected by factors such as class imbalance, dataset size, and the choice of hyperparameters.

V.RESULTS AND DISCUSSIONS

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

1. Naive Bayes: When the Naive Bayes algorithm was used, 98.24% of instances were properly classified, but 1.76% were not.

2. Logistic Regression: When the Logistic Regression algorithm was used, 99.91% of instances were properly classified, but 0.09% were incorrectly classified.

3. KNN Classifier: By using the K-nearest neighbor algorithm, 99.88 percent of instances were properly classified, but 0.12 percent were not.

4. Support Vector Classifier: According to the Support Vector Classifier algorithm, 99.90% of instances were properly classified, but 0.1% weren't.

5. Multilayer Perceptron Classifier: When the Multilayer Perceptron Classifier algorithm was used, 99.23% of instances were properly classified, but 0.77% of them were not.

6.AdaBoost Classifier: By using an algorithm, 83.71% of instances were properly classified, but 16.29% weren't.

7. Random Forest: 99.98% of instances were properly classified using the Random Forest Classifier algorithm; the error rate was 0.02%. Among all algorithms, this one provides the greatest accuracy.

Fig:3 accuracy of random forest Classifier

By entering different values the output changes. Here we have three cases they are whether there is a chance of an Earthquake occurring or not and the third case is severe if happens nuclear explosion.

Fig:4 prediction of an earthquake occurring

Fig:5 Prediction of an earthquake not happening

By choosing algorithms we can see the accuracy of each algorithm that is shown in fig:3 next we have to predict the earthquake by giving the necessary features. The result is shown in fig:4 above. In fig:4 it is shown that the predicted result is there is a chance of happening Nuclear Explosion and fig:5 shows the output that there is no chance of happening Earthquake.

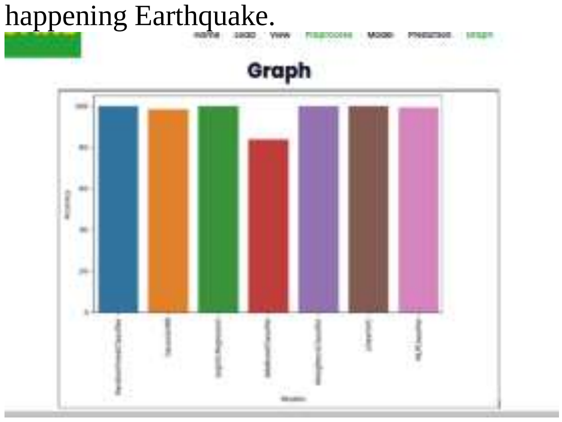

Fig:6 Comparison graphs between algorithms

In this research, we use seven algorithms, and their accuracy comparison is shown in the graph above figure 6.

 DOI: 10.36893.IEJ.2023.V52I04.234-238

VI.CONCLUSION

To categorize the main earthquake events as negative or positive, seven machine learning algorithms have been tested. The random forest classifier outperformed the other methods in this study in terms of accuracy. Therefore, the Random Forest classifier is best suited for earthquake forecasts. This algorithm is mainly suitable for high complexity. Future work will entail developing case studies based on actual intervention data and taking feature selection techniques into account.

VII.FUTURE SCOPE

There are a lot of items that could be improved or added in the upcoming work. In this undertaking, the ID3 and Naive Bayes classifiers are the two data mining classifiers we have chosen to use. There are additional classifiers, including the Bayesian network, neural network, and C4.5 classifiers. Such classifiers could be tallied in the future to provide more data to be compared with, but they were not included in this study.

VIII.REFERENCES

[1] KM Asim, F. Martnez-Alvarez, A. Basit, and T. Iqbal Using machine learning techniques, the Hindukush region can forecast the magnitude of earthquakes. Natural Hazards, 85(1):471–48, 2017.
[2] Louise K. Comfort, Disaster reaction by self-organization: the great Hanshin, Japan earthquake of January 17, 1995.
[3] Jean-François Couchot, Christophe Guyeux, and Guillaume Royer predicting anonymously the quantity and type of firefighting activities 23rd International Database Applications Conference Proceedings & Engineering Symposium, pages 30:1–30:8 ACM, June 2019.
[4] Laigen Dong and Jie Shan provide a detailed analysis of the use of remote sensing to identify earthquake-related building damage. 2013; 84:85–99, ISPRS Journal of Photogrammetry and Remote Sensing Ratios: A Decision Tree Approach Expert Systems with Applications, 40(10):3970–3983, 2013.
[5] machine-learning-random-forest-algorithm of March 28, 2023.
[6] logistic regression in machine learning as of March 28, 2023.
[7] logistic regression in machine learning as of March 28, 2023.
[8] adaboost-algorithm/March 28, 2023.
[9] k-nearest-neighbor-algorithm-for-machine-learning of March 28, 2023.
[10] support-vector-machine-algorithm as of March 28, 2023.
[11] topics (engineering or perception) of March 28, 2023.
[12] The University of Management and Technology's Department of Computer Science is run by Rabia Tehseen, Muhammad Shoaib Farooq, and Adnan Abid. Lahore, Pakistan (54770).
[13] Alyona Galkina and Saint Petersburg State University, Russia, Machine Learning Methods for Earthquake Prediction: A Survey, April 2019.