International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Performance Analysis of Earthquake Prediction Using Supervised Algorithm

[1]B.Sangeetha,[2]Y.Leela,[3]A.Siva Prasad,[4]K.Dilleswari

[1,2,3,4] student,Dept of Electronics And Communication Engineering(ECE)
AITAM(Aditya Institute Of Technology And Management),Tekkali.

ABSTRACT:

In order to lessen the catastrophic effects of earthquakes, accurate prediction techniques are required. Earthquakes are a crucial danger. This research uses a variety of machine learning models and historical seismic data to predict the likelihood, size, and frequency of earthquakes in a given area. With a specific focus, the study attempts to create predictive frameworks by utilizing relevant geological and geographical features in conjunction with historical seismic events. The results of this study project have important ramifications for the creation of sophisticated early warning systems, plans for disaster relief, risk assessment procedures, and seismic research projects. Through providing precise forecasts of earthquake intensity and probability, this research seeks to increase community adaptability and lessen the negative effects of seismic activity in the area of interest.

Introduction :

Developing statistical models and algorithms for computers to automatically learn from data and improve over time at a particular task without requiring explicit programming is known as machine learning in the field of artificial intelligence.

Several types of machine learning algorithms include reinforcement learning, supervised learning, and unsupervised learning. The computer system is trained on labeled data in supervised learning, where each data point has a target or output associated with it.

Unsupervised learning involves training a computer system on unlabeled data so that it may find patterns or groups without knowing what the output should be beforehand. earning, a computer system learns by making mistakes and getting rewarded or punished for its actions in a particular environment.

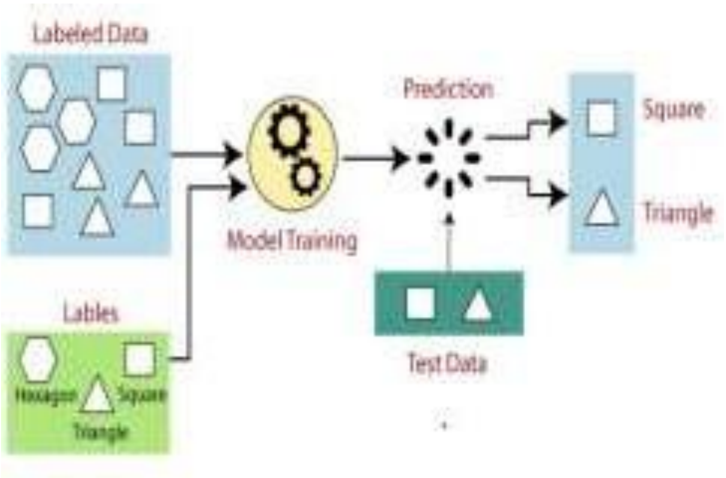
Among the many domains in which machine learning finds use are image identification, natural language processing, recommender systems, and fraud detection. Because computers can learn and become more efficient over time, machine learning has become a vital tool for companies and organizations trying to make decisions automatically and extract insights from their data.

Classification of Machine Learning:1. Supervised Learning

Machines are trained using properly "labelled" training data in supervised learning, and they make output predictions based on this data. Some input data has already been tagged with the appropriate output, as indicated by the labelled data. When machines are taught to predict an output properly, the supervisor in supervised learning is the training data that they are given. The same idea that a pupil learns under a teacher's guidance is applied here. Providing accurate input and output data to the machine learning model is known as supervised learning. The objective of an algorithm for supervised learning is to identify a mapping function that associates the input variable (x) with the output variable (y). Within the actual world.

How Supervised Learning Works:

In supervised learning, labelled datasets are used to train models, allowing them to learn about all kinds of data. After training is finished, the model is evaluated using testdata, which is a portion of the training set, to see how well it predicts the result.



- The shape is designated as a square if it has four sides and each side is equal.
- The shape will be designated as a triangle if it has three sides.
- If a form has six equal sides, it is referred to as a hexagon. We evaluate our model using the est set after training. It is the role of the model to identify the shape. Numerous forms have already been taught to the computer. When it comes across a new shape, it classifies it using multiple dimensions and predicts the outcome.

Methodologies:

1.Linear regression
2.Support Vector Machine(SVM)
3.NaiveBayes
4.Randomforest

1.Linear Regression

Here, latitude, longitude, depth, and the number of seismic stations that recorded the earthquake are examples of independent variables. The linear relationship between these variables and the dependent variable (here, earthquake magnitude) is modeled using a supervised machine learning algorithm called linear regression.
In this situation, we have used multiplelinear regression to model the relationship between earthquake magnitude andlatitude, longitude, depth, and the number of seismic stations that recorded the earthquake.

The multiple linear regression model postulates that the dependent variable and the independent variable have a linear relationship. (magnitude) and each of theindependent variables (latitude, longitude, depth, and number of seismic stations), andthat the relationship is additive (i.e., theeffect of each independent variable on the dependent variable is independent of the other independent variables).

2.Support Vector Machine (SVM)

One kind of supervised machine learning method that may be applied to both regression and classification problems is Support Vector Machines (SVM). The fundamental notion of Support Vector Machines (SVM) is to determine the optimal boundary that either predicts a continuous output variable (here, earthquake magnitude) or divides the data into distinct classes.
A higher-dimensional space with easily discernible borders is created by the SVM mapping the data points.For each class, the ideal border is the one that maximizes the margin, or the distance between the boundary and the closest data points. We call this boundary the "hyperplane."

For regression tasks, SVM uses a similar approach but instead of a hyperplane, it finds a line (or curve in higher dimensions) that best fits the data while maximizing the margin. This line is the "support vector regression line.

3.NaiveBayes

The naive Bayes classifiers are a family of simple "probabilistic classifiers" used in statistics that make strong (naive) independence assumptions between the characteristics (see Bayes classifier).
These are some of the simplest models of Bayesian networks, but when combined with kernel density estimation, they can achieve very high accuracy.

Because naive Bayes classifiers require a linear number of parameters in relation to the number of variables (features/predictors) in a learning job, they are very scalable. An expensive iterative approximation, which is the strategy used for many other types of classifiers, can be replaced by evaluating a closed-form expression in linear time when training a maximum likelihood classifier.

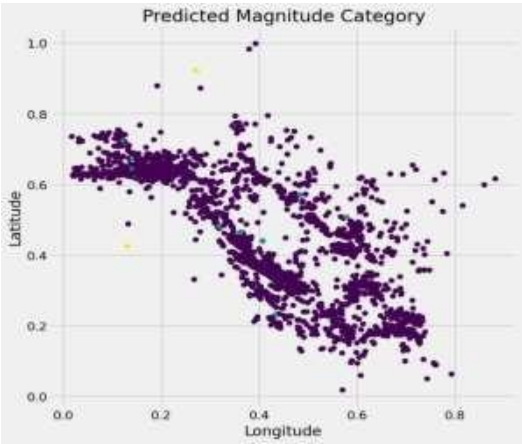
We employed the Naive Bayes classifier in the code to forecast earthquake magnitudes according to latitude, longitude, and the quantity of monitoring stations. The data was divided into training and testing sets. The Naive Bayes model's performance on the test data was evaluated using the accuracy score, confusion matrix, and classification report.

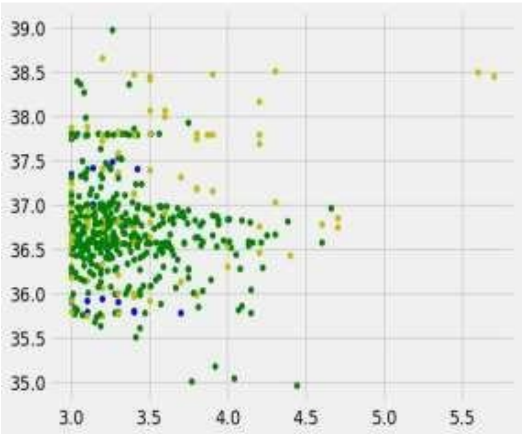
4.Random Forest

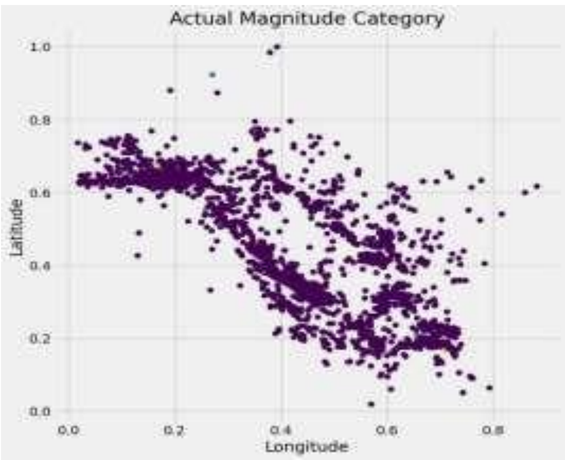
The fundamental notion behind random Random forest is based on the basic idea of building many decision trees, each trained on a random subset of attributes and data. Each tree generates its own forecast, and the total prediction is the mean (for regression) or the mode (for classification) of the individual projections from each tree.Random forests can lessen the effects of overfitting and increase the precision and stability of the model by growing a large number of trees and averaging them.

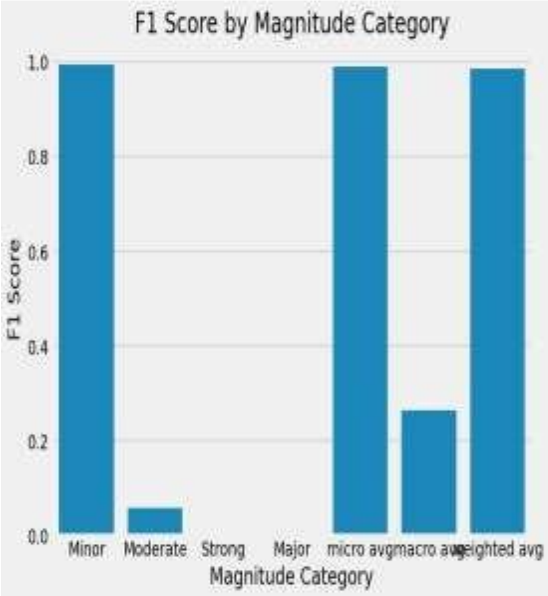
Based on an earthquake's location, longitude, depth, and number of monitoring stations, we were able to predict its magnitude using the random forest approach in the code we previously released. The random forest model was trained on the training data after the data was divided into training and testing sets. The mean squared error (MSE) and R- squared (R2) score were used to assess the models performance.

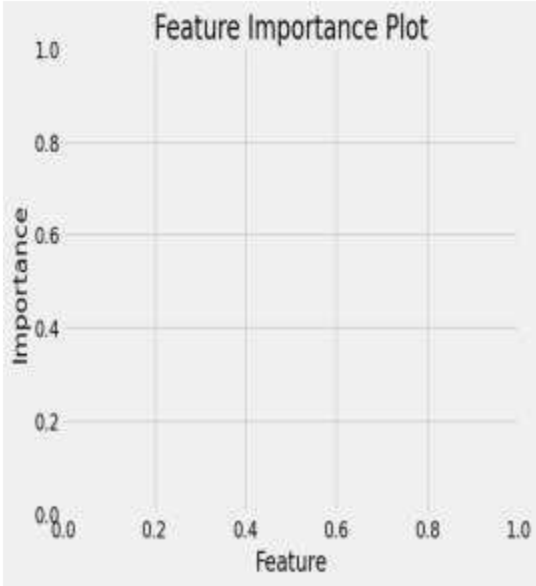
Results:

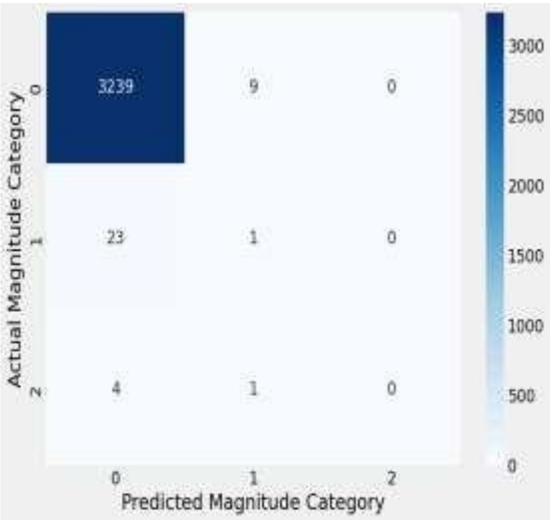


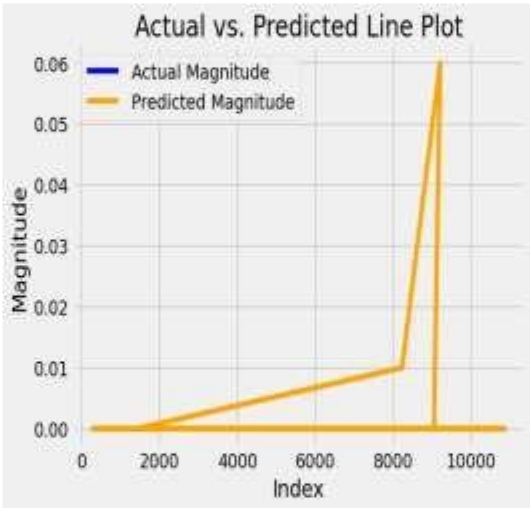


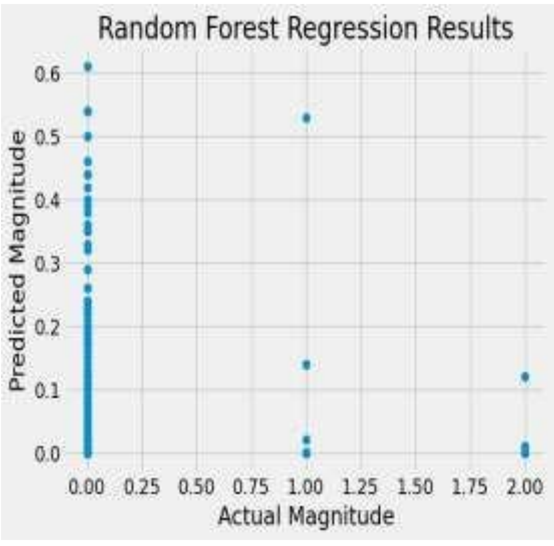


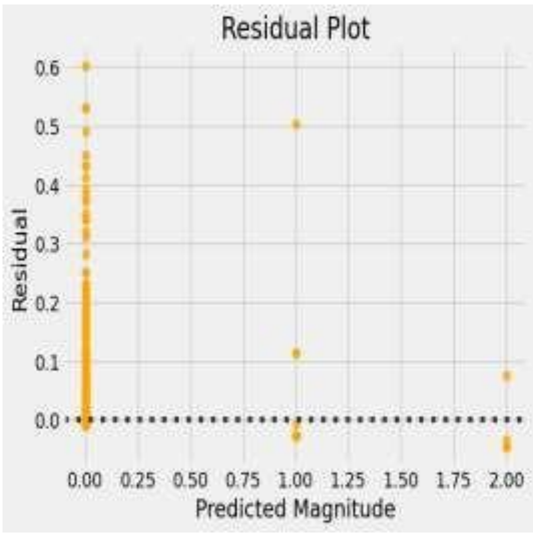

F1 Score by Magnitude Category
F1 Score
Minor
Moderate
Strong
Major
micro avg
macro avg
weighted avg
Magnitude Category

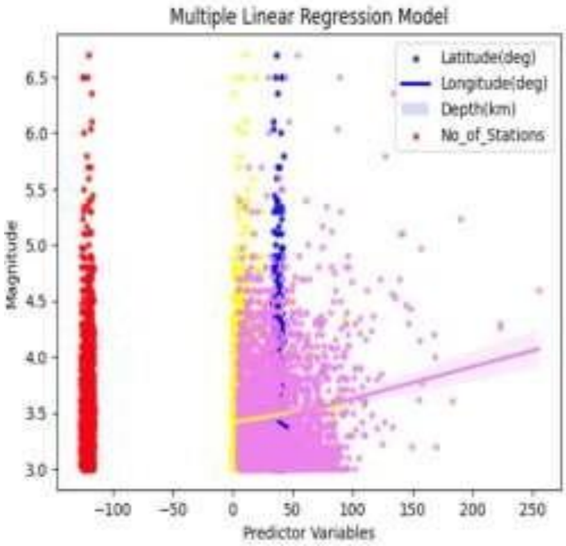

Feature Importance Plot
Importance
Feature


3239
9
0
23
1
0
4
1
0
Actual Magnitude Category
Predicted Magnitude Category


Actual vs. Predicted Line Plot
Actual Magnitude
Predicted Magnitude
Magnitude
Index


Random Forest Regression Results
Predicted Magnitude
Actual Magnitude


Residual Plot
Residual
Predicted Magnitude



Conclusion:

Both the R-squared (R2) score and the mean squared error (MSE) can be used to assess how well two models perform when compared.
A model is generally seen as being better if it has a higher R2 score and a lower MSE. This is so because a lower mean square error (MSE) denotes a more accurate prediction made by the model. The MSE calculates the average difference between the projected and actual values. A higher R2 score suggests that the model is able to explain more of the variability in the target variable. The R2 score represents the proportion of the variance in the target variable that is explained by the model. We can infer from this project's outcomes that random forest is the most.

REFERENCES:

1. JYH-WOEI LIN , CHUN-TANG CHAO, AND JUING-SHIAN CHIOU Department of Electrical Engineering, Southern Taiwan University of Science and Technology, Tainan 71005, Taiwan, under Grant MOST 106-2221-E- 218-001-MY2-2018
2. LEI ZHANG 1, LANGCHUN SI2, HAIPENG YANG2, YUANZHI HU2, AND JIANFENG QIU11 Key Laboratory of Intelligent Computing and Signal Processing, Ministry of Education,
3. Anhui University, Anhui 230039, China 2School of Computer Science and Technology, Anhui University, Anhui 230039, China-2019
4. MD. HASAN AL BANNA1, (Associate Member, IEEE), KAZI ABU TAHERI ,M. SHAMIM KAISER 2, (Senior Member, IEEE), MUFTI MAHMUD 3,
5. (Senior Member, IEEE), MD. SAZZADUR RAHMAN2, (Member, IEEE), A. S. M. SANWAR HOSEN 4, (Member, IEEE), AND GI HWAN CHO 4, (Member, IEEE) -2020 KAISER 2, (Senior Member, IEEE), MUFTI MAHMUD 3,
6. (Senior Member, IEEE), MD. SAZZADUR RAHMAN2, (Member, IEEE), A. S. M. SANWAR HOSEN 4, (Member, IEEE), AND GI HWAN CHO 4, (Member, IEEE) -2020