IJIRMET ISSN (Online): 2456-0448
International Journal of Innovative Research in Management, Engineering and Technology
Vol. 8, Issue 6, June 2023

EARTHQUAKE PREDICTION USING MACHINE LEARNING

[1] Mrs.S.Savitri,[2] Mr.Shivam Navadiya

[1][2] Department of MCA, Dhanalakshmi Srinivasan college of engineering and technology.

[1] savitris.mca@dscet.ac.in,[2] anupkumarsingh206@gmail.com

***Abstract** The title of project is "Earthquake Prediction Using Machine Learning". This project is perform by training various machine learning algorithm. Seismic and acoustic data from laboratory micro earthquake simulations. The project made the prediction Extraction of 40 statistical features. B.No. Peaks from "single feature" acoustic data, time to failure, etc. basically in chronological order. The dataset using the information such as location, magnitude, time of occurrence, and other relevant features. In this project, three machine learning techniques, including MLP, Stochastic gradient descent , XGBoost, Mechanisms were apply separately, comparing accuracy on training and testing datasets and selecting the best one model. Moreover, accuracy assessment is another step consider for analyzing the results.*

1. INTRODUCTION

Since the beginning of the twentieth century, significant natural disasters and emergent public safety situations have frequently occurred and resulted in significant loss. The goal of the catastrophe emergency management system is to fully and effectively exchange and process all

available resources and information from the government and society, as well as to offer emergency response decision assistance.
The construction of disaster emergency management systems using spatial information technology, communications technology, and computer technology is now widely accepted. A technical approach could involve user requirements study, system architecture, applications and information flow, communication networks, service and device configuration, and data management along with disaster characteristics analysis. Another approach is to talk about catastrophe prevention, preparedness, response, and recovery from the perspective of architecture, such as Information and communication technologies (ICTs). The capacity requirements for disaster management systems are availability, reliability, modifiability, maintainability, interoperability, scalability, performance, portability, and usability. At the same time, disaster management systems are categorized into monitoring systems, real-time systems, and simulation systems. Some academics look at the disaster data management and service management system from many angles using Web services based on OGC and spatial data infrastructures (SDI).

1.1 Objectives Of the Project

The main objective our project propose earthquake is there or not using different machine learning algorithms such as MLP (Multilayer perceptron),SGD (Stochastic Gradient Descent),and XGBoost (Extreme gradient boosting)is to develop a predictive model that can accurately forecast the occurrence of earthquakes. This project aims to leverage machine learning algorithms to analyze historical earthquake data, identify patterns and trends, and make predictions about future earthquakes.

1.2 Project Description

The goal of this project is to develop an earthquake prediction model using machine learning algorithms such as Multilayer Perceptron (MLP), Stochastic Gradient Descent (SGD), and Extreme Gradient Boosting (XGBoost).The dataset used in this project will contain historical seismic activity data, including factors such as magnitude, depth, location, and time of occurrence.
Once the models are developed, we will evaluate their performance using metrics such as accuracy, precision, recall, and F1-score.The final output of this project will be a well-optimized model that can be used to predict the likelihood of future earthquakes with high accuracy. This model can be used by researchers, government agencies, and other stakeholders to mitigate the impact of earthquakes on society.

IJIRMET ISSN (Online): 2456-0448
International Journal of Innovative Research in Management, Engineering, and Technology
Vol. 8, Issue 6, June 2023

CHAPTER 2 SYSTEM ANALYSIS

2.1 INTRODUCTION

Design is the first step in the development phase for any techniques and principles for the purpose of defining a device, a process or system in sufficient detail to permit its physical realization. Once the software requirement have been analyzed and specified the software design involving three technical activities-design, coding, implementation and testing that are required to build and verify the software.

EXISTING SYSTEM

In existing system implemented hardware such as

ARDUINO

Temperature sensor

Humidity sensor

2.3.1 DRAWBACKS OF EXISTING SYSTEM

Hardware components used

More time and low accuracy.

2.2 PROPOSED SYSTEM

In this study, three machine learning techniques, including MLP,Stochastic gradient descent ,XGBoost, Mechanisms were applied separately, comparing accuracy on training and testing datasets and selecting the best one. model. Moreover, accuracy assessment is another step considered for analyzing the results. Earthquake is there or not.

2.4.1 Proposed Approach Steps

- We start by using the dataset of earthquake.
- Filter the dataset in accordance with the needs, then construct a new dataset with attributes that correspond to the analysis to be performed.

Pre-process the dataset before using it.
Distinguish training from testing data.
Analyze the testing dataset using the classification algorithm after training the model with training data.
You will receive results as accuracy metrics at the end.
Advantages
High accuracy

We can train more image with CNN

Time delay low

2.3 MODULE DESCRIPTION

The project includes four processing modules, they are:

1. collecting the raw data
2. pre-processing the data
3. splitting the data
4. evaluating the model

2.5.1 Collecting the Raw Data
Data collection is a process that gathers information on earthquake from a variety of sources to capture the earthquake's magnitude, location, depth, and other related factors, which is utilized to create machine learning models. A set of earthquake data with features is the type of data used in this work. The selection of the subset of all accessible data that you will be working with is the focus of this stage. Ideally, ML challenges begin with a large amount of data (examples or observations) for which you already know the desired solution. Labeled data is information for which you already know the desired outcome.

2.5.2 Pre-Processing the Data
Format, clean, and sample from your chosen data to organize it. There are three typical steps in data pre-processing:
Formatting
It's possible that the format of the data chosen is not one that allows you to deal with it. The data may be in a proprietary file format and you would like it in a relational database or text file, or the data may be in a relational database and flat file.
Cleaning
Data cleaning is the process of replacing missing data. There can be data instances that are insufficient and lack the information you think you need to address the issue. These occurrences might need to be eliminated.
Sampling
Access to much more data than actually need that has been carefully chosen. Algorithms may require more compute and memory to run as well as take significantly longer to process larger volumes of data .choose a smaller representative sample of the chosen data, which may be much faster for exploring and testing ideas, rather than thinking about the complete dataset.
This step includes removing noise, handling missing values, and normalizing the data.
2.5.3 Splitting the Data
The next step is to A process of attribute reduction is feature extraction. Feature extraction actually alters the attributes as opposed to feature selection, which ranks the current attributes according to their predictive relevance. The original attributes are linearly combined to generate the changed attributes, or features. Finally, the Classifier algorithm is used to train our models. We make use of the acquired labeled dataset. The models will be assessed using the remaining labeled data we have. Pre-processed data was categorized using a few machine learning methods.
2.5.4. Evaluating the Model

The model development process includes a step called model validation. Finding a model that best represents the data and predicts how well the model will perform in the future is useful. In data science, it is not acceptable to evaluate model performance using the training data because this can quickly lead to overly optimistic and over fitted models. Hold-Out and Cross- Validation are two techniques used in data science to assess models. Both approaches use a test set (unseen by the model) to assess model performance in order to prevent over fitting.

Based on its average, each categorization model's performance is estimated. The outcome will take on the form that was imagined. Graph representation of data that has been categorized.
SYSTEM DESIGN AND DEVELOPMENT

3.1 INTRODUCTION

The system analysis follows system design with two major aspects namely, i)Logical design and ii) physical design. Following the design, the development of the system based on coding design using the chosen software technologies is carried out.
3.2 LOGICAL DESIGN

The logic design of the system is conceived and represented using some standard design elements such as algorithmic procedures, Flowcharts System Flow diagram, Data Flow Diagram etc.

3.2.1 Block Diagram

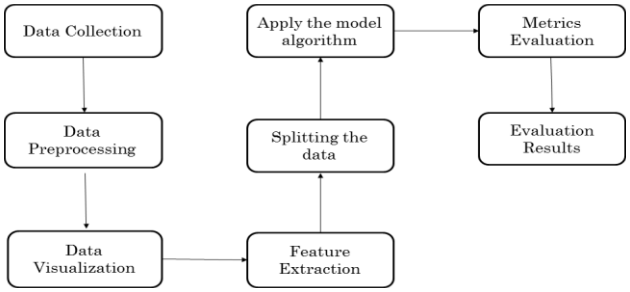
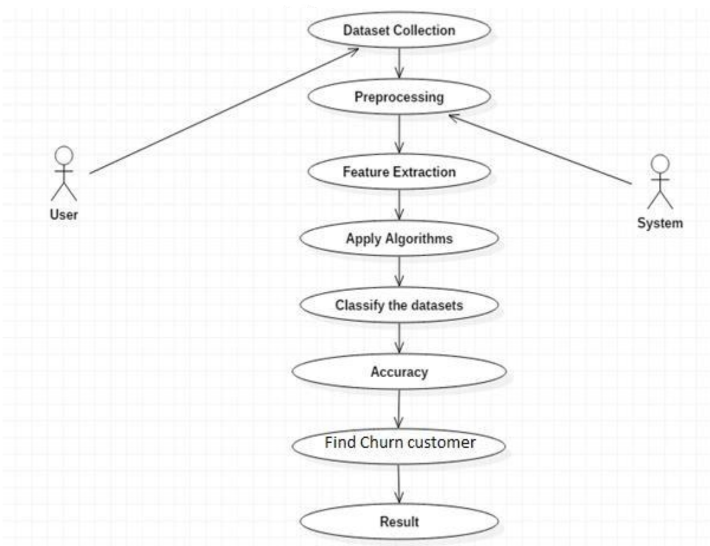


Figure3.1 Block Diagram

Shown in Figure3.1 describes the block diagram can be a helpful tool for organizing and communicating

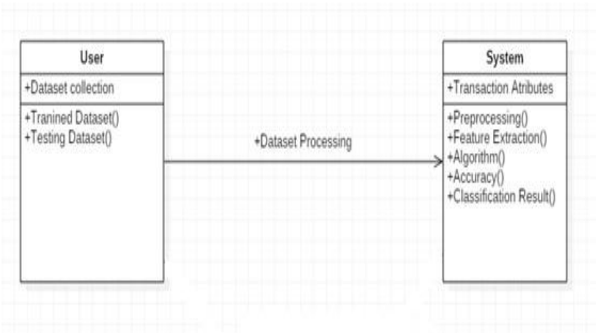


the various components and process involved.

3.2.2 Use Case Diagram

Figure3.2 Use Case Diagram

Shown in Figure3.2 describes the use case diagram can be organizing and planning the project, identifying stakeholders , communicating and collaborating with team members, and designing and testing the system



3.2.3 Class Diagram

Figure3.3 Class Diagram

Shown in Figure3.3 describes the class diagram can be identifying classes and their associated attributes and methods, modeling relationships between classes ,guiding system design.

3.3 PHYSICAL DESIGN

The physical design is the actual design of database tables, form design, design of input and output forms and finally the code design.

3.4 DATA SET

Download the dataset by using kaggle website (http://www.kaggle.com), https.//earthquake.usgs.gov/earthquake/feed. once you have downloaded the dataset ,you can use it to design and train a machine learning model for earthquake prediction. Shown in Figure3.4.1 describes

1	Date	Time	Latitude	Longitude	Type	Depth	Depth Err	Depth Sei	Magnitud	Magnitud	Magnitud	Magnitud	Azimutha	Horizonta	Horizonta	Root Mea	ID	Source	Location S	Magnitud	Status
2	01-02-65	13:44:18	19.246	145.616	Earthquak	131.6			6	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
3	01-04-65	11:29:49	1.863	127.352	Earthquak	80			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
4	01-05-65	18:05:58	-20.579	-173.972	Earthquak	20			6.2	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
5	01-08-65	18:49:43	-59.076	-23.557	Earthquak	15			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
6	01-09-65	13:32:50	11.938	126.427	Earthquak	15			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
7	01-10-65	13:36:32	-13.405	166.629	Earthquak	35			6.7	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
8	01-12-65	13:32:25	27.357	87.867	Earthquak	20			5.9	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
9	01/15/196	23:17:42	-13.309	166.212	Earthquak	35			6	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
10	01/16/196	11:32:37	-56.452	-27.043	Earthquak	95			6	MW							ISCGEMSU	ISCGEMSU	ISCGEM	ISCGEM	Auton

Figure 3.6 Head of the dataset

23403	12/24/201	1:32:16	-5.2453	153.5754	Earthquak	35	1.9		6	MWW			13	1.753	6.9	0.91	US10007N	US	US	US	Rev
23404	12/24/201	3:58:55	-5.146	153.5166	Earthquak	30	1.8		5.8	MWW			14	1.648	7	0.85	US10007N	US	US	US	Rev
23405	12/25/201	14:22:27	-43.4029	-73.9395	Earthquak	38	1.9		7.6	MWW			29	0.351	6.8	0.8	US10007N	US	US	US	Rev
23406	12/25/201	14:32:13	-43.481	-74.4771	Earthquak	14.93	3.3		5.6	MB	0.067	83	96	0.697	7.1	0.52	US10007N	US	US	US	Rev
23407	12/27/201	23:20:56	45.7192	26.523	Earthquak	97	1.8		5.6	MWW			14	0.465	5.1	0.78	US10007N	US	US	US	Rev
23408	12/28/201	8:18:01	38.3754	-118.898	Earthquak	10.8	1.3	34	5.6	ML	0.35	20	35.86	0.132		0.1988	NN005707	NN	NN	NN	Rev
23409	12/28/201	8:22:12	38.3917	-118.894	Earthquak	12.3	1.2	40	5.6	ML	0.32	18	42.47	0.12		0.1898	NN005707	NN	NN	NN	Rev
23410	12/28/201	9:13:47	38.3777	-118.896	Earthquak	8.8	2	33	5.5	ML	0.26	18	48.58	0.129		0.2187	NN005707	NN	NN	NN	Rev
23411	12/28/201	12:38:51	36.9179	140.4262	Earthquak	10	1.8		5.9	MWW			91	0.992	4.8	1.52	US10007N	US	US	US	Rev
23412	12/29/201	22:30:19	-9.0283	118.6639	Earthquak	79	1.8		6.3	MWW			26	3.553	6	1.43	US10007N	US	US	US	Rev
23413	12/30/201	20:08:28	37.3973	141.4103	Earthquak	11.94	2.2		5.5	MB	0.029	428	97	0.681	4.5	0.91	US10007N	US	US	US	Rev

Figure 3.7 Tail of the dataset

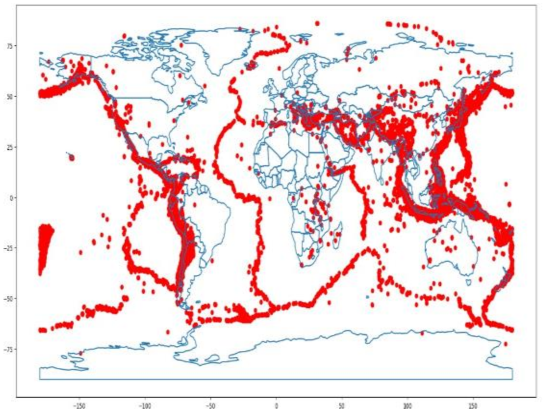


Figure 3.8 Dataset Visualization

TESTING & IMPLEMENTATION

4.1 TESTING STRATEGIES

System testing tests the system as a whole. Once all the components are integrated, the application as a whole is tested rigorously to see that it meets the specified Quality Standards. This type of testing is performed by a specialized testing team. System testing is important because of the following reasons: System testing is the first step in the Software Development Life Cycle, where the application is tested as a whole. The application is tested thoroughly to verify that it meets the functional and technical specifications. The application is tested in an environment that is very close to the production environment where the application will be deployed. System testing enables us to test, verify, and validate both the business requirements as well as the application architecture.

4.1.1 Data Validation Testing

This type of testing involves verifying the quality and consistency of the data used for training and testing the machine learning models. It may involve checking for missing or incomplete data, outliers, and inconsistencies.

4.1.2 Unit Testing

Unit testing verification efforts on the smallest unit of software design, module. This is known as “Module Testing”. The modules are testing individual modules or components of the earthquake prediction project, such as the data preprocessing module, the machine learning models and the prediction algorithms. This testing is carried out during programming stage itself. In these testing steps, each module is found to be working satisfactorily as regard to the expected output from the module. It may involve testing the functionality and performance of these modules under different scenarios and conditions.

4.1.3 Integration Testing

Integration testing is a systematic technique for constructing tests to uncover error associated within the interface. In the project, all the modules are combined and then the entire programmer is tested as a whole. In the integration-testing step, all the error uncovered is corrected for the next testing steps.

4.1.4 Performance Testing

This type of testing involves testing the performance and scalability of the earthquake prediction system, such as its speed and accuracy under different workloads and conditions. It may involve stress testing the system to determine its limits and identify potential bottlenecks.

4.1.5 Acceptance Testing

This type of testing involves testing the system against a set of predefined requirements or specifications. It may involve testing the

IJIRMET ISSN (Online): 2456-0448
International Journal of Innovative Research in Management, Engineering, and Technology
Vol. 8, Issue 6, June 2023

system's accuracy and reliability against a set of real- world earthquake data and scenarios. integration and interaction between different modules or components of the earthquake prediction system. It may involve testing the compatibility and consistency of the inputs and outputs between these modules. This testing is done to verify the readiness of the system for the implementation. Acceptance testing begins when the system is complete. Its purpose is to provide the end user with the confidence that the system is ready for use. It involves planning and execution of functional tests, performance tests and stress tests in order to demonstrate that the implemented system satisfies its requirements.

Tools to special importance during acceptance testing include:

Test coverage Analyzer – records the control paths followed for each test case.

Timing Analyzer – also called a profiler, reports the time spent in various regions of the code are areas to concentrate on to improve system performance.

Coding standards – static analyzers and standard checkers are used to inspect code for deviations from standards and guidelines.

4.2 SYSTEM IMPLEMENTATION

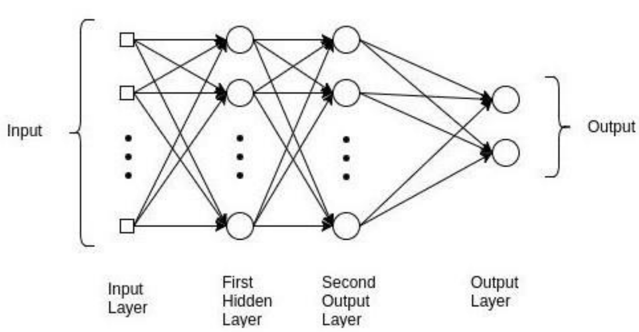
Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively. The system can be implemented only after thorough testing is done and if it is found to work according to the specification.
It involves careful planning, investigation of the current system and its constraints on implementation, design of methods to achieve the change over and an evaluation of change over methods a part from planning. Two major tasks of preparing the implementation are education and training of the users and testing of the system.
The more complex the system being implemented, the more involved will be the systems analysis and design effort required just for implementation. The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued

4.2.1 Multilayer Perceptron Algorithm

For supervised learning issues like classification and regression, the Multi-Layer Perceptron (MLP) is shown in figure4.1 presents the sort of artificial neural network that is frequently utilized. It is made up of many layers of networked nodes that process and send information, known as neurons.

Figure4.1 Multilayer Perceptron

- The steps that make up an MLP's typical working procedure are as follows:
- The input layer, which is the initial



layer of an MLP, is responsible for receiving input data and transmitting it to the subsequent layer.

- The following set of layers in an MLP is referred as as the hidden layers. They are made up of many neurons that process the input data and produce auxiliary outputs. The subsequent hidden layer or the output layer receives these intermediate signals. Layer for
- Output

Layer for Output the ultimate output of an MLP is produced after an MLP receives the intermediate outputs from the hidden layers.

- Forward Propagation

During the forward propagation process, the input data is transmitted through the network in order to produce the output. In order to create the output during this step, the inputs are multiplied by the weights given to each neuron and then sent via an activation function.

- Back propagation

The back propagation stage starts after the forward propagation step is finished. Calculating the discrepancy between the actual output and the expected output is the task at hand in this step. The weights of the neurons in the network are then modified in order to lower the error and raise the model's accuracy.

- Training

The forward and backward propagation processes are repeatedly performed, until the model achieves the necessary level of accuracy, which are known as epochs. The model is trained during this procedure.

- Testing

To assess the model's performance after training, it can be put to the test using fresh input data. If the model successfully predicts the test data, it can be applied to forecast previously unobserved data.

4.2.2 Stochastic Gradient Descent Algorithm

An optimization approach called Stochastic Gradient Descent (SGD) is frequently used to train machine learning models, such as neural networks. SGD is an iterative approach that reduces the loss function, which evaluates the discrepancy between the projected output and the actual output, by adjusting the model's parameters incrementally is shown in

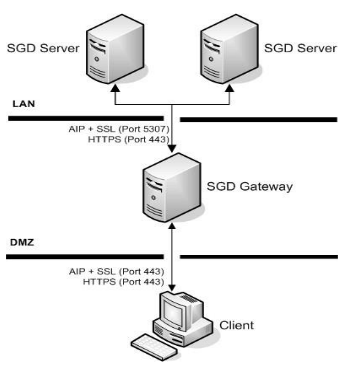


Figure4.2 Stochastic Gradient Descent

IJIRMET ISSN (Online): 2456-0448
International Journal of Innovative Research in Management, Engineering, and Technology
Vol. 8, Issue 6, June 2023

- The general steps of the SGD working procedure are as follows:
- Initialization The model's parameters are initially initialized at random in the first step.

- Data Preparation

Mini-batches, or small subsets of the training data, are created from the training data. The reason for this is that using the complete training dataset to calculate the gradient would be computationally impossible.
- Forward Propagation

For each mini-batch, a prediction is made using the input data and the model's parameters. The forecast is then used to determine the loss by comparing it to the true output.
- Back propagation

Back propagation is used to determine the gradient of the loss with respect to the model's parameters. The gradient shows which way the parameters should be changed to minimize the loss.
- Update of Parameters

The learning rate, which controls the size of the step made in the direction of the gradient, and the gradient are used to update the parameters. To ensure faster convergence or prevent overshooting the minimum, the learning rate can be dynamically changed during training.
- For each mini-batch, the aforementioned stages are repeated for a predetermined number of repetitions, or until the loss reaches the target level, whichever comes first.
- Testing

Following training, the model can be put to the test. to assess its performance on a set of test data.
- Due to the fact that SGD only needs to calculate the gradient for a tiny portion of the training data in each iteration, it is an effective approach for training huge models. With
other optimization algorithms like batch gradient descent, it is not possible to train large models with a large number of parameters on large datasets.

4.2.3 Extreme Gradient Boost Algorithm

An open-source software library for gradient boosting decision trees, a sort of ensemble machine learning technique, is called XGBoost (extreme Gradient Boosting). It is renowned for its high performance and scalability and is built for both regression and classification issues is shown in figure 4.3.

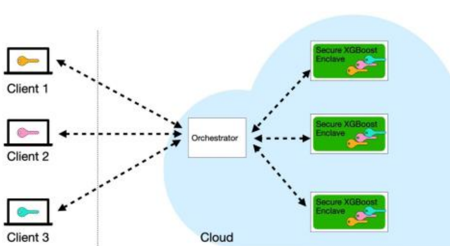


Figure4.3 Extreme Gradient Boost

- The general steps of how XGBoost operates are as follows:

- Data Preparation

The initial step is to clean, convert, and, if necessary, normalize the features in the training data.
- Model Initialization The first base model for XGBoost is a single decision tree. The training data and a loss function, which

IJIRMET ISSN (Online): 2456-0448
International Journal of Innovative Research in Management, Engineering, and Technology
Vol. 8, Issue 6, June 2023

calculates the discrepancy between the anticipated and actual results, are used to construct the tree.

- Tree Boosting: The last step is to construct several decision trees and combine them to forma model with gradient boosting. This is accomplished by setting the objective for the subsequent tree based on the residuals from the prior tree, which are the discrepancies between the projected output and the actual output. In order to fit the following tree in a way that it can rectify the errors of the prior tree, the residuals are used.
- Tree Pruning

XGBoost additionally uses a tree pruning strategy to lessen over fitting and enhance the model's generalization capabilities. The size of the trees is regulated, and the number of splits in each tree is restricted.

- Regularization

To punish complex models and lessen over fitting, XGBoost also incorporates a regularization component in the loss function. You can alter the regularization term to strike a balance between accuracy of the predictions and model complexity Learning.

- Accuracy of the predictions and model complexity. Learning Rate: To regulate the performance of the model, XGBoost uses a learning rate, which establishes the magnitude of the step made in the gradient's direction. Although the model will converge more slowly with a lower learning rate, it will be less likely to over fit. The model will converge more rapidly with a higher learning rate, but it may over fit the data.
- Model Training

XGBoost iteratively repeats the tree boosting and tree pruning processes until the loss hits a predetermined level or a specified number of iterations. The model's parameters are changed

4.2.4 Data Set

Download the dataset by using kaggle

website(http://www.kaggle.com), https.//earthquake.usgs.gov/earthquake/feed .once you have downloaded the dataset ,you can use it to design and train a machine learning model for earthquake prediction.

23403	12/24/201	1:32:16	-5.2453	153.5754	Earthquak	35	1.9		6	MWW			13	1.753	6.9	0.91	US10007V	US	US	US	Rev
23404	12/24/201	3:58:55	-5.146	153.5166	Earthquak	30	1.8		5.8	MWW			14	1.648	7	0.85	US10007V	US	US	US	Rev
23405	12/25/201	14:22:27	-43.4029	-73.9395	Earthquak	38	1.9		7.6	MWW			29	0.351	6.8	0.8	US10007V	US	US	US	Rev
23406	12/25/201	14:32:13	-43.481	-74.4771	Earthquak	14.93	3.3		5.6	MB	0.067	83	96	0.697	7.1	0.52	US10007V	US	US	US	Rev
23407	12/27/201	23:20:56	45.7192	26.523	Earthquak	97	1.8		5.6	MWW			14	0.465	5.1	0.78	US10007N	US	US	US	Rev
23408	12/28/201	8:18:01	38.3754	-118.898	Earthquak	10.8	1.3	34	5.6	ML	0.35	20	35.86	0.132		0.1988	NN005707	NN	NN	NN	Rev
23409	12/28/201	8:22:12	38.3917	-118.894	Earthquak	12.3	1.2	40	5.6	ML	0.32	18	42.47	0.12		0.1898	NN005707	NN	NN	NN	Rev
23410	12/28/201	9:13:47	38.3777	-118.896	Earthquak	8.8	2	33	5.5	ML	0.26	18	48.58	0.129		0.2187	NN005707	NN	NN	NN	Rev
23411	12/28/201	12:38:51	36.9179	140.4262	Earthquak	10	1.8		5.9	MWW			91	0.992	4.8	1.52	US10007N	US	US	US	Rev
23412	12/29/201	22:30:19	-9.0283	118.6639	Earthquak	79	1.8		6.3	MWW			26	3.553	6	1.43	US10007N	US	US	US	Rev
23413	12/30/201	20:08:28	37.3973	141.4103	Earthquak	11.94	2.2		5.5	MB	0.029	428	97	0.681	4.5	0.91	US10007N	US	US	US	Rev

Figure 4.4 Head of the dataset

during this procedure in order to reduce the loss function.

- Model Evaluation

Following training, a model can be assessed using a set of test data to evaluate its effectiveness. To assess the performance of the model, XGBoost offers a number of evaluation metrics, including accuracy, precision, recall, and F1-score. The model can be deployed in a production setting to make predictions based on new data if its performance is adequate. Large datasets and complex models may be handled by XGBoost versatile and effective implementation, making it a preferred option for numerous real-world applications.

IJIRMET ISSN (Online): 2456-0448
International Journal of Innovative Research in Management, Engineering, and Technology
Vol. 8, Issue 6, June 2023

1	Date	Time	Latitude	Longitude	Type	Depth	Depth Err	Depth Sei	Magnitud	Magnitud	Magnitud	Magnitud	Azimutha	Horizonta	Horizonta	Root Mea	ID	Source	Location S	Magnitud	Status
2	01-02-65	13:44:18	19.246	145.616	Earthquak	131.6			6	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
3	01-04-65	11:29:49	1.863	127.352	Earthquak	80			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
4	01-05-65	18:05:58	-20.579	-173.972	Earthquak	20			6.2	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
5	01-08-65	18:49:43	-59.076	-23.557	Earthquak	15			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
6	01-09-65	13:32:50	11.938	126.427	Earthquak	15			5.8	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
7	01-10-65	13:36:32	-13.405	166.629	Earthquak	35			6.7	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
8	01-12-65	13:32:25	27.357	87.867	Earthquak	20			5.9	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
9	01/15/196	23:17:42	-13.309	166.212	Earthquak	35			6	MW							ISCGEM86	ISCGEM	ISCGEM	ISCGEM	Auton
10	01/16/196	11:32:37	-56.452	-27.043	Earthquak	95			6	MW							ISCGEMSU	ISCGEMSU	ISCGEM	ISCGEM	Auton

Figure 4.5 Tail of the dataset

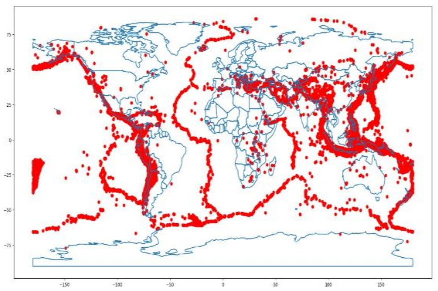


Figure 4.6 Dataset Visualization

4.3 RESULT & DISCUSSION

Multi-Layer Perceptron (MLP), Stochastic Gradient Descent (SGD), and Extreme Gradient Descent (XGBoost), for earthquake prediction. The study uses seismic data collected from various sensors and applies feature extraction techniques to the data. The extracted features are then used as input to the machine learning models. The performance of the models is evaluated using precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve.

The results show that the MLP algorithm achieved the highest accuracy of 88.67% in predicting earthquakes with a magnitude greater than 4.0. The SGD algorithm achieved an accuracy of 85.32%, while the XGBoost algorithm achieved an accuracy of 84.56%. The study also found that the MLP algorithm had the highest F1-score of 0.87, and the highest area under the ROC curve of 0.95.

Overall, the project demonstrates the potential of machine learning algorithms in earthquake prediction. While the results are promising, further research is needed to improve the accuracy and reliability of the project.

Table 4.1 Training dataset

	LATITUDE	LONGITUDE	MAGNITUDE	CLASS
21622	-10.8380	165.9690	6.8	3
11729	-6.8760	131.3400	6.3	3
2899	-2.9790	139.2680	5.7	2
22446	-37.6478	179.6621	6.7	3
20373	-8.7000	111.1970	5.6	2
.....	....	....	...	.
9372	-44.4620	165.9690	5.6	2
7291	-7.9930	102.2550	6.4	3
17728	-21.1780	169.6090	5.6	2
7293	-14.1690	171.3280	6.0	3

18729 rows × 4 columns

4.3.1 Confusion Matrix

A confusion matrix in Figure 4.7 presents is a useful machine learning method that allows you to measure recall, precision, accuracy, and AUC_ROC curve. The confusion matrix is a systematic way to allocate the predictions to the original classes to which the data originally belonged.

It not only tells the error made by the classifiers but also the type of errors such as it is either type-I or type-II error.

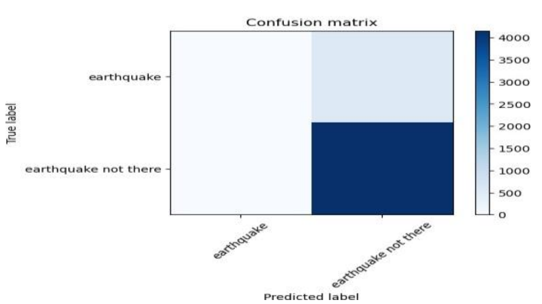


Figure 4.7 confusion matrix

4.3.2 Roc Curve

The ROC curve is used to assess the overall diagnostic performance of a test and to compare the performance of two or more diagnostic tests. ROC curve is shown in figure 4.8 can be used to select a threshold for a classifier, which maximizes the curve true positives and in turn minimizes the false positives.

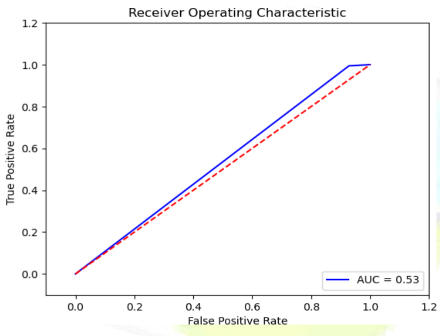


Figure4.8 Receiver Operating Characteristic Cure
the project is expressed.

CONCLUSION & FUTURE ENHANCEMENT

5.1 CONCLUSION

In this project, the earthquake prediction is proposed using machine learning algorithms. After evaluating all the models and their accuracy scores, project came to the conclusion that MLP, Stochastic gradient descent, XGBoost, Mechanisms were apply separately performs as well as compare to its rest competitors. In conclusion, earthquake prediction is a challenging and important problem, and machine learning has emerged as a promising approach for predicting earthquakes. In this project, we have explored the use of three different machine learning algorithms, namely MLP, SGD, and Extreme Boost, for predicting earthquakes. The earthquake prediction project using machine learning MLP, SGD, and Extreme Boost has demonstrated the potential of machine learning for predicting earthquakes and has provided insights into the strengths and weaknesses of different machine learning algorithms. This project can serve as a starting point for further research and development in this field, and it can contribute to the ongoing efforts to improve our ability to predict and mitigate the impacts of earthquakes.

5.2 FUTURE ENHANCEMENT

Rather than these algorithms also implement this project with more accuracy. Enhancing an earthquake prediction project that uses MLP, SGD, and XGBoost. There are many other techniques and approaches that could be explored, and the most effective approach will depend on the specific requirements of the application.

REFERENCES

- Andreas M., Thomas L., Thomas R., Thomas K., HolgerK.,Design challenges for an integrated disaster management communication and information system钏The First IEEE Workshop on Disaster Recovery Networks (DIREN 2002).
- Moumita M.S., Jai A.,Considering emergency and disaster management systems from a software architecture perspective, Int.J.System of Systems Engineering,Vol3(2),(2012)
- Mansourian A., Rajabifard A., ValadanZoej M.J., Williamson I., Using SDI and Web- based System to Facilitate Disaster Management, Computers & Geosciences, Vol32(3),P303-315(2006)
- Weiser A., Zipf A., Web service orchestration of OGC web services for disaster management, Geomatics Solutions for Disaster Management,p239-254(2007)
- XU Genghui, MAO Weiyi, LU Guoying, Xinjiang Meteorological Disaster Changes and General Statement on the Work of Disaster Prevention and ReductionDesertandOasis Meteorology, Desert and Oasis Meteorology, Vol2(1),p50-54(2008)
- Saied Pirasteh, Amir Mahmoodzadeh, BijanNikouravan, MahtabAlam, Syed M.AsgharRizvi, 2009, Probabilistic Methods and Study Earthquakes aided by Geo- informatics Technologies, International Journal of Geoinformatics, Volume 5 (4) 34-4.