INTERNSHIP REPORT

PROJECT TITLE: EARTHQUAKE PREDICTION

A Report Submitted to

Jawaharlal Nehru Technological University Kakinada,

Kakinadain partial fulfillment for the award of the

degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

(AI&ML)

Submitted by

B. VILOK SAI (20KN1A4206)
M. BHANU SRI (20KN1A4235)
T.SAI SUNDAR (20KN1AA4259)

Under the esteemed guidance of
Mr. Ankit

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NRI INSTITUTE OF TECHNOLOGY

Autonomous

(Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada) Accredited by NBA (CSE, ECE & EEE), Accredited by NAAC with 'A' GradeISO 9001: 2015 Certified Institution

Pothavarappadu (V), (Via) Nunna, Agiripalli (M), Krishna Dist., PIN: 521212, A.P, India.

2022-2023

NRI INSTITUTE OF TECHNOLOGY
(An Autonomous institution Approved by AICTE, Permanently Affiliated to JNTUK Kadinada)
Accredited By NBA(CSE, ECE & EEE), Accredited by NAAC with ‘A’ Grade
ISO 9001:2015 Certified Institution
Pothavarapu(V), (Via)Nunna, Agiripalli(M)Krishna Dist, PIN:521212 Ap,India.

CERTIFICATE

This is to certify that the “**Internship report”** submitted by **B.Vilok Sai (20KN1A4206), M. Bhanu Sri (20KN1A4235), T.Sai Sundar(20KN1A4259)** is work done by them and submitted during 2022-2023 academic year, in partial fulfillment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING,** at **BLACKBUCK ENGINEERS PVT LTD, Road No:36, Jubilee Hills, Hyderabad, Telangana**

Kathyayani

INTERNSHIP COORDIANTOR
(R KATHYAYANI)

Head of the Department
(Dr. D. SUNEETHA)

EXTERNAL EXAMINER

Internship Experience Letter

Certificate ID: BBNR0212 **Issued Date: 01st November 2022**

To Whom It May Concern:

This is to certify that **BODDAPATI VILOK SAI** has successfully completed internship at **Blackbuck Engineers Pvt Ltd**, and He worked with us from **13th June 2022** to **05th September 2022.**

He has worked on a project titled **Earthquake Prediction** by learning and incorporating Artificial Intelligence & Machine Learning concepts under the supervision of our project mentor.

We found that He is sincere, hardworking, technically sound and result oriented. He worked well as part of a team during the tenure.

We wish all the best for future endeavours.

Best regards,

Kathyayani

Kathyayani Rudravelli
Project Head
Blackbuck Engineers Pvt Ltd

Mounika Bezawada
HR Manager
Blackbuck Engineers Pvt Ltd

+91-9391002762 Jubilee Hills, Hyderabad, Telangana contact@blackbucks.me

BLACKBUCK ENGINEERS

Internship Experience Letter

Certificate ID: BBNR0261 **Issued Date: 01st November 2022**

To Whom It May Concern:

This is to certify that **THIRUMALA SAI SUNDAR** has successfully completed internship at **Blackbuck Engineers Pvt Ltd**, and He worked with us from **13th June 2022** to **05th September 2022.**

He has worked on a project titled **Earthquake Prediction** by learning and incorporating Artificial Intelligence & Machine Learning concepts under the supervision of our project mentor.

We found that He is sincere, hardworking, technically sound and result oriented. He worked well as part of a team during the tenure.

We wish all the best for future endeavours.

Best regards,

Kathyayani

Kathyayani Rudravelli
Project Head
Blackbuck Engineers Pvt Ltd

Mounika Bezawada
HR Manager
Blackbuck Engineers Pvt Ltd

Blackbuck Engineers Pvt Ltd • Hyderabad •

+91-9391002762 Jubilee Hills, Hyderabad, Telangana contact@blackbucks.me

BLACKBUCK ENGINEERS

Internship Experience Letter

Certificate ID: BBNR0237 **Issued Date: 01st November 2022**

To Whom It May Concern:

This is to certify that **M. BHANUSRI** has successfully completed internship at **Blackbuck Engineers Pvt Ltd**, and She worked with us from **13th June 2022** to **05th September 2022.**

She has worked on a project titled **Earthquake Prediction** by learning and incorporating Artificial Intelligence & Machine Learning concepts under the supervision of our project mentor.

We found that She is sincere, hardworking, technically sound and result oriented. She worked well as part of a team during the tenure.

We wish all the best for future endeavours.

Best regards,

Kathyayani

Kathyayani Rudravelli
Project Head
Blackbuck Engineers Pvt Ltd

Mounika Bezawada
HR Manager
Blackbuck Engineers Pvt Ltd

+91-9391002762 Jubilee Hills, Hyderabad, Telangana contact@blackbucks.me

ACKNOWLEDGEMENT

We take this opportunity to thank all who have rendered their full support to my work. The pleasure, the achievement, the glory, the satisfaction, the reward, the appreciation and theconstruction of my project cannot be expressed with a few words for their valuable suggestions.

We are expressing my heartfelt thanks to **Head of the Department, Dr. D. SUNEETHA**

garu for her continuous guidance for completion of my Project work.

We are extending our sincere thanks to **Dean of the Department, Dr. K. V. SAMBASIVA RAO** for his continuous guidance and support to complete my project successfully.

We are thankful to the **Principal, Dr. C. NAGA BHASKAR** garu for his encouragement tocomplete the Project work.

We are extending my sincere and honest thanks to the **Chairman, Dr. R. VENKATA RAOgaru & Secretary, Sri K. Sridhar** garu for their continuous support in completing the Project work.

Finally, We thank the Administrative Officer, Staff Members, Faculty of Department of CSE, NRI Institute of Technology and my friends, directly or indirectly helped us in the completion of this project.

NAME:
B. VILOK SAI (20KN1A4206)
M.BHANU SRI (20KN1A4235)
T. SAI SUNDAR (20KN1A4259)

ABSTRACT

Among the countless natural disasters, earthquakes are capable to inflict vast devastation to a large number of buildings and constructions at the blink of an eye. Lack of knowledge and awareness on earthquake as well as its comeback is conspicuous and results in disaster; leading to bitter memories. Therefore, earthquake forecast has been a polemical study theme that has defied even the most intelligent of minds.

In this session an attempt was made to predict the earthquake over an area. To predict earthquake over an area we created machine learning model. The approach used to create the machine learning model is tree-based ML model know as a random forest algorithm.

Earthquake magnitude prediction for a region has been carried out in this research using the temporal sequence of historic seismic activities in combination with the machine learning classifiers. Prediction has been made on the basis of mathematically calculated eight seismic indicators using the earthquake catalog of the region. These parameters are based on the well-known geophysical facts of Gutenberg–Richter's inverse law, distribution of characteristic earthquake magnitudes and seismic quiescence.

INDEX

WEEKLY OVERVIEW OF INTERNSHIP ACTIVITIES

1st WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	13.06.2022	Monday	Introduce the Topic & the Problem Statement
	14.06.2022	Tuesday	Introduce the Topic & the Problem Statement
	15.06.2022	Wednesday	Introduce the Topic & the Problem Statement
	16.06.2022	Thursday	Introduce the Topic & the Problem Statement
	17.06.2022	Friday	Introduce the Topic & the Problem Statement

2nd WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	20.06.2022	Monday	Abstract Building
	21.06.2022	Tuesday	Abstract Building
	22.06.2022	Wednesday	Abstract Building
	23.06.2022	Thursday	Abstract Building
	24.06.2022	Friday	Abstract Submission

3rd WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	27.06.2022	Monday	Abstract Submission
	28.06.2022	Tuesday	Abstract Submission
	29.06.2022	Wednesday	Explain your Approach to Solving Problem
	30.06.2022	Thursday	Explain your Approach to Solving Problem
	01.07.2022	Friday	Explain your Approach to Solving Problem

4th WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	04.07.2022	Monday	Explain your Approach to Solving Problem
	05.07.2022	Tuesday	Explain Structure of Project
	06.07.2022	Wednesday	Explain Structure of Project
	07.07.2022	Thursday	Explain Structure of Project
	08.07.2022	Friday	Explain Structure of Project

5th WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	11.07.2022	Monday	Data Preprocessing
	12.07.2022	Tuesday	Data Preprocessing
	13.07.2022	Wednesday	Data Preprocessing
	14.07.2022	Thursday	Data Preprocessing
	15.07.2022	Friday	Data Preprocessing

6th WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	18.07.2022	Monday	Perform Analysis
	19.07.2022	Tuesday	Perform Analysis
	20.07.2022	Wednesday	Perform Analysis
	21.07.2022	Thursday	Perform Analysis
	22.07.2022	Friday	Perform Analysis

7th WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	25.07.2022	Monday	PPT Preparation
	26.07.2022	Tuesday	PPT Preparation
	27.07.2022	Wednesday	PPT Preparation
	28.07.2022	Thursday	PPT Preparation
	29.07.2022	Friday	PPT Preparation

8th WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	01.08.2022	Monday	PPT Submission
	02.08.2022	Tuesday	PPT Submission
	03.08.2022	Wednesday	Mid Review
	04.08.2022	Thursday	Mid Review
	05.08.2022	Friday	Mid Review

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
9th WEEK	08.08.2022	Monday	Mid Review
	10.08.2022	Tuesday	Mid Review
	11.08.2022	Wednesday	Building & Applying Algorithm
	12.08.2022	Thursday	Building & Applying Algorithm

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
10th WEEK	16.08.2022	Tuesday	Building & Applying Algorithm
	17.08.2022	Wednesday	Building & Applying Algorithm
	19.08.2022	Friday	Building & Applying Algorithm

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
11th WEEK	22.08.2022	Monday	Concluding Project
	23.08.2022	Tuesday	Concluding Project
	24.08.2022	Wednesday	Concluding Project
	25.08.2022	Thursday	Concluding Project
	26.08.2022	Friday	Concluding Project

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
12th WEEK	29.08.2022	Monday	Final Review
	30.08.2022	Tuesday	Final Review
	01.09.2022	Wednesday	Final Review
	02.09.2022	Thursday	Final Review
	05.09.2022	Friday	Final Review

CHAPTER-1

INTRODUCTION

INTRODUCTION

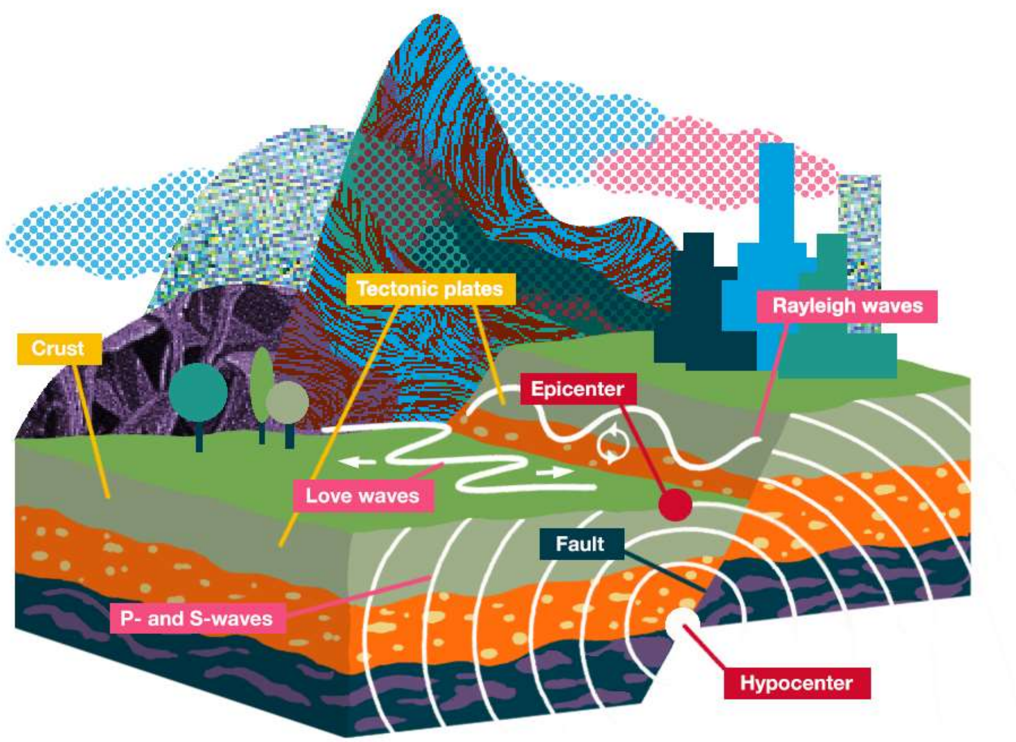
Introduction of the project:

An earthquake is a violent shaking of the ground produced by the sudden movement of rock materials below the earth's surface.

The earthquakes originate in "Tectonic plate boundary". The landscapes of our national parks, as well as geologic hazards such as earthquakes and volcanic eruptions, are due to the movement of the large plates of Earth's outer shell.

Tectonic plates, large slabs of rock that divide Earth's crust, move constantly to reshape the Earth's landscape. The system of ideas behind plate tectonics theory suggests that Earth's outer shell (lithosphere) is divided into several plates that glide over the Earth's rocky inner layer above the soft core (mantle). The plates act like a hard and rigid shell compared to Earth's mantle. The mantle sits between Earth's dense, very hot core and its thin outer layer, the crust.

Plate tectonics has become the unifying theory of geology. It explains the earth's surface movement, current and past, which has created the tallest mountain ranges and the deepest oceans.



There are three types of tectonic plate boundaries:

- Plates rip apart at a divergent plate boundary, causing volcanic activity and shallow earthquakes;
- At a convergent plate boundary, one plate dives ("subducts") beneath the other, resulting in a variety of earthquakes and a line of volcanoes on the overriding plate;
- Transform plate boundaries are where plates slide laterally past one another, producing shallow earthquakes but little or no volcanic activity.

Another large-scale feature is a hotspot, where a plate rides over a rising plume of hot mantle, creating a line of volcanoes on top of the plate. National Park Service lands contain not only active examples of all types of plate boundaries and hotspots, but also rock layers and landscapes that reveal plate-tectonic activity that occurred in the distant past.

The Earth's surface is active according to tectonic theory, moving as much as 1-2 inches a year. The many tectonic plates shift and interact all the time. This motion reshapes the Earth's outer layer. Earthquakes, volcanoes and mountains are the result of this process.

The focus is point inside the earth where the earthquake started, sometimes called the hypocenter, and the point on the surface of the earth directly above the focus is called the epicenter**.**
There are two ways by which we can measure the strength of an earthquake: magnitude and intensity. Magnitude is proportional to the energy released by an earthquake at the focus. It is calculated from earthquakes recorded by an instrument called **seismograph**. It is represented by Arabic Numbers (e.g. 4.8, 9.0). Intensity on the other hand, is the strength of an earthquake as perceived and felt by people in a certain locality. It is a numerical rating based on the relative effects to people, objects, environment, and structures in the surrounding. The intensity is generally higher near the epicenter. It is represented by Roman Numerals (e.g. II, IV, IX). In the Philippines, the intensity of an earthquake is determined using the PHIVOLCS Earthquake Intensity Scale (PEIS).

This research of earthquake magnitude prediction encompasses a set of input parameters extracted from temporal distribution of past earthquakes. Such temporal distributions illustrate the frequency of occurrence of seismic events as function of their magnitudes (Panakkat and Adeli 2007). These parameters show the underlying relations of geophysical facts of seismic quiescence (Hainzl et al. 2000), Gutenberg–Richter law (Christensen and Olami 1992) and frequency of foreshocks (McGuire et al. 2005). This relationship between seismic activity and geophysical facts needs to be modeled, irrespective of the degree of the nonlinearity that exists among them. Seismic quiescence is break in the normal seismic energy release from the fault region. This accumulation of energy in the faults may lead to the occurrence of an earthquake, and the amount of energy stored is related to the magnitude of upcoming earthquake (Wiemer and Wyss 1994). Similarly, foreshock frequency is considered to be a sign of a major earthquake (Boore 2001). Foreshocks are the series of earthquakes of magnitude slightly higher than the background seismic activity. The

Gutenberg–Richter inverse power law shows relation between the earthquake magnitudes and the cumulative frequency of events less than and equal to the corresponding magnitudes (Rundle 1989).

Machine Learning (ML) and Artificial Neural Networks (ANN) have been used in a variety of fields for prediction and classification purposes, like computer vision (Murtza et al. 2015), object recognition (Liang and Thorpe 2000), genetics (Zahur et al. 2014), bioinformatics (Larsen et al. 2012) and weather forecasting (Partal 2015). Researchers have considered using ANN for modeling of highly nonlinear and complex underlying relationship between geophysical facts and earthquakes (Adeli and Panakkat 2009; Morales-Esteban et al. 2013; Panakkat and Adeli 2007; Reyes et al. 2013) with quite meaningful results.

The core idea of this work is to predict earthquakes of magnitude 3.7 and above in a region on monthly basis using ML approaches in combination with eight seismicity indicators (Panakkat and Adeli 2007). The mathematically calculated seismicity indicators from the previously occurred seismic events show the seismic behavior of the region, which are used as input to the different ML approaches. These include Multi-Linear Regression,Decision Tree, random forest and n trees for earthquake prediction. The prediction results of all the mentioned techniques are discussed.

CHAPTER-2

ADVANTAGES& DISADVANTGES

ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

Random Forest is based on the bagging algorithm and uses Ensemble Learning technique. It creates as many trees on the subset of the data and combines the output of all the trees. In this way it reduces overfitting problem in decision trees and also reduces the variance and therefore improves the accuracy.

Random Forest can be used to solve both classification as well as regression problems. Random Forest works well with both categorical and continuous variables.

Random Forest can automatically handle missing values .No feature scaling required**:** No feature scaling (standardization and normalization) required in case of Random Forest as it uses rule based approach instead of distance calculation.

Handles non-linear parameters efficiently**:** Non linear parameters don't affect the performance of a Random Forest unlike curve based algorithms. So, if there is high non-linearity between the independent variables, Random Forest may outperform as compared to other curve based algorithms.

Random Forest can automatically handle missing values. Random Forest is usually robust to outliers and can handle them automatically.

Random Forest algorithm is very stable. Even if a new data point is introduced in the dataset, the overall algorithm is not affected much since the new data may impact one tree, but it is very hard for it to impact all the trees. Random Forest is comparatively less impacted by noise.

DISADVANTAGES:

Complexity: Random Forest creates a lot of trees (unlike only one tree in case of decision tree) and combines their outputs. By default, it creates 100 trees in Python sklearn library. To do so, this algorithm requires much more computational power and resources. On the other hand decision tree is simple and does not require so much computational resources.

Longer Training Period: Random Forest require much more time to train as compared to decision trees as it generates a lot of trees (instead of one tree in case of decision tree) and makes decision on the majority of votes.

Random forest may not get good results for small data or low-dimensional data (data with few features). Since the randomness becomes greatly reduced. Processing high-dimensional data and feature-missing data are the strengths of random forest.

Random forest may overfit for data with much noise. Decision trees tend to be overfitted in prediction, random forest reduces the degree of overfitting through voting, but its prediction is still overfitting compared to linear model, which is characterized by good matching of existing data. But very conservative with unknown data, and high probability of false negative error.

CHAPTER-3

SOFTWARE
REQUIREMENTS SPECIFICATIONS

SOFTWARE AND HARDWARE REQUIREMENTS

SOFTWARE REQUIREMENTS:

SCIKIT-LEARN:

Scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

SPYDER:

THE SCIENTIFIC PYTHON DEVELOPMENT ENVIRONMENT.

Spyder is an open-source cross-platform integrated development environment (IDE) for scientific programming in the Python language.

PYTHON:

Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems

MATPLOTLIB:

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

JUPYTER:

Jupyter notebook is an open-source IDE that is used to create Jupyter documents that can be created and shared with live codes. Also, it is a web-based interactive computational environment

PANDAS:

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license

NUMPY:

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

SEABORN:

Seaborn is a statistical plotting library that can read Pandas dataframes (as well as other data structures) and provides simple methods for adding regression lines to your scatter diagrams.

HARDWARE REQIREMENTS:

Required ram: 4gb or more
Hard Disk: 30gb of free space
Processor: 64 bit, quad-core, 2.5 GHz minimum per core
Display: Dual XGA (1024 x 768) or higher resolution monitors

CHAPTER-4

MOTIVATION

MOTIVATION

The motivations behind the prediction of earthquakes are pretty clear: save human lives and minimize damage to the surrounding area. On paper, earthquakes seem like something we should at least be able to at least make reasonable guesses about when they might occur — earthquakes are a purely physical system, and thus, unlike economics or elections, are not affected by the "human" element. Moreover, half of the ten deadliest natural disasters in human history (excluding famines and epidemics) have been caused by earthquakes. Given the rapid population growth and urbanization in the last few decades, more than two billion people now live within 250 kilometers from a fault line. Consequently, predicting earthquakes has never been more important. Yet, despite that all the data we have, the physical nature of earthquakes, and the clear motives, we are still unable to predict when they will occur.

Any precise prediction of an earthquake needs to include three different pieces of information:

(1) the exact date and time
(2) the physical location
(3) the magnitude or severity of the earthquake.

To date, no major earthquake has been predicted by scientists that included all three of these pieces of information. The key difficulty, according to experts, is actually good news for most of us who are not seismologists: there have been only sixteen recorded earthquakes with a magnitude > 8.5 since 1901. Hence, a lot of the fundamental physics behind major earthquakes is still not fully understood due to the dearth of major earthquakes.

Models so far have not enjoyed a lot of success in predicting major earthquakes due to the enormous complexity of the physics involved. Moreover, GPS sensors today allow us to measure movements of less than 1/10th of a millimeter per year, which is much lower than the typical velocity of a tectonic plate. Consequently, scientists have recently been turning their attention to machine learning models to try to predict earthquakes, given the massive amount of sensor data available but the relative lack of sophistication in the physical models used for prediction.

CHAPTER-5

LITERATURE REVIEW

LITERATURE REVIEW

Each year the world faces thousands of earthquakes of magnitude 5.0 or greater, resulting in devastating property destruction and tragic loss of life. To help avert these catastrophes, scientists have long searched for ways to predict when and where earthquakes will happen. The earth science establishment in the US says that earthquake prediction still lies outside the realm of possibility. But recent scientific developments across the globe suggest that seismic forecasting is on the horizon.

An earthquake can strike without warning and wreak horrific destruction and death, whether it's the catastrophic 2010 quake that took a devastating toll on the island nation of Haiti or a future great earthquake on the San Andreas Fault in California, which scientists know is inevitable. Yet despite rapid advances in earthquake science, seismologists still can’t predict when the Big One will hit. Predicting the Unpredictable explains why, exploring the fact and fiction behind the science—and pseudoscience—of earthquake prediction.

Earthquake Prediction: Dawn of the New Seismology examines the latest scientific clues in hopes of discovering seismic precursors which may shed light on real earthquake prediction in the future. It is destined to be nothing less than an epoch-changing work, addressing this ancient enigma by joining the parts of a scientific detective story that ranges from the steppes of Russia to the coast of Chile, bringing to light astounding breakthroughs by researchers in Italy, India and elsewhere.

Governments in countries such as China and Japan provide support for seismic forecasting, and it is time for our country to do the same. *Earthquake Prediction* makes the case, with an important message for the tens of millions of Americans on the US West Coast, the Mississippi River Valley, and other seismically active zones.

A number of techniques have been presented in the literature that uses ANN in combination with some seismic precursor to predict earthquakes. Negarestani et al. (2002) used Back Propagation Neural Network (BPNN) to detect anomalous behavior in radon concentration induced by earthquakes. A concentration of radon gas is present in soil measured continuously which varies due to environmental changes. Seismic activity also causes the increase in radon concentration of soil, which is differentiated successfully from ordinary changes caused by environment, using neural network. Liu et al. (2004) predicted earthquakes in China by using ensemble of Radial Basis Function (RBF) neural networks. The past earthquake magnitude data are used as an input for the network. Ikram and Qamar. A number of techniques have been presented in the literature that uses ANN in combination with some seismic precursor to predict earthquakes. Negarestani et al. (2002) used Back Propagation Neural Network (BPNN) to detect anomalous behavior in radon concentration induced by earthquakes. A concentration of radon gas is present in soil measured continuously which varies due to environmental changes. Seismic activity also causes the increase in radon concentration of soil, which is differentiated successfully from ordinary changes caused by environment, using neural network. Liu et al. (2004) predicted earthquakes in China by using ensemble of Radial Basis Function (RBF) neural networks. The past earthquake magnitude data are used as an input for the network. Ikram and Qamar

CHAPTER-6

EXISTING SYSTEM
PROPOSED SYSTEM

EXISTING PROBLEM

- It is important to make the differentiation between earthquake forecasting and earthquake prediction. Forecasting utilizes existing data and trends to form estimates of the location, frequency and magnitude of Earthquakes.
- It becomes possible to focus on earthquake-prone areas and establish better and quicker acting safety measures, build earthquake resistant structures, and make the population more prepared and aware.
- To predict earthquake a machine learning algorithm model is developed using machine learning algorithms like Random forest , Linear regression , Light Gradient Boosting Mechanism , Support Vector Machine etc.
- Most of the scientists use polynomial regression to build machine learning model to predict Earthquake

PROPOSED SYSTEM

In Earthquake prediction Machine learning plays a major role. Machine Learning techniques are used to interpret and analyze previous Earthquake readings to find the major factor that causes earthquake.

Earthquake can strike any location at any time. But the history shows in same general pattern year after year. If we look at the pattern of where earthquakes occur around the world, it is clear that most of the activity is concentrated in a number of distinct earthquake belts; for instance the edge of the Pacific Ocean, or in the middle of the Atlantic Ocean. Machine learning has been used to predict the occurance of earthquake in India. Based upon Previous earthquake results in India.

We trained dataset using different regression techniques. We calculated the MAE, RMSE for each model. We chose the model which got least RMSE, MAE values. From our research the random forest model got the least values of MAE, RMSE. So, We chose Random forest regression model for predicting the earthquake.

CHAPTER-7

KEYWORDS AND DEFINITION

KEYWORDS AND DEFINITION

PANDAS:

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license

NUMPY:

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

SKLEARN:

Scikit-Learn, also known as sklearn is a python library to implement machine learning models and statistical modelling. Through scikit-learn, we can implement various machine learning models for regression, classification, clustering, and statistical tools for analyzing these models.

Mathplotlib:

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Create publication quality plots. Make interactive figures that can zoom, pan, update.

Ensemble:

An ensemble is a machine learning model that combines the predictions from two or more models. The models that contribute to the ensemble, referred to as ensemble members, may be the same type or different types and may or may not be trained on the same training data.

SEABORN:

Seaborn is a statistical plotting library that can read Pandas dataframes (as well as other data structures) and provides simple methods for adding regression lines to your scatter diagrams.

REGRSSION:

Regression is a technique for investigating relationship between dependent variable or outcome and independent features or variables.

DATASET:

A collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer.

MACHINE LEARNING:

Regression is a technique for investigating relationship between dependent variable or outcome and independent features or variables.

CHAPTER-8
IMPLEMENTION AND ARCHITECTURE

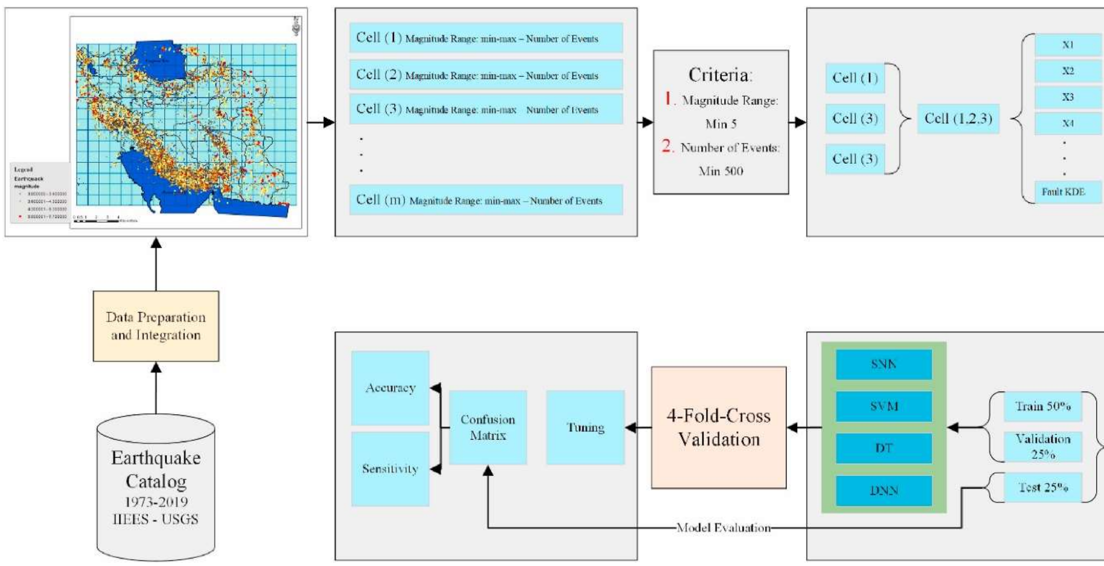
IMPLEMENTATION AND ARCHITETURE

Implementation

Testing earthquake prediction methods requires statistical techniques that compare observed success to random chance. One technique is to produce simulated earthquake catalogs and measure the relative success of predicting real and simulated earthquakes. The accuracy of these tests depends on the validity of the statistical model used to simulate the earthquakes. This study tests the effect of clustering in the statistical earthquake model on the results. Three simulation models were used to produce significance levels for a VLF earthquake prediction method. As the degree of simulated clustering increases, the statistical significance drops. Hence, the use of a seismicity model with insufficient clustering can lead to overly optimistic results. A successful method must pass the statistical tests with a model that fully replicates the observed clustering. However, a method can be rejected based on tests with a model that contains insufficient clustering. U.S. copyright. Published in 1997 by the American Geophysical Union

Earthquake prediction research must meet certain standards before it can be suitably evaluated for potential application in decision making. For methods that result in a binary (on or off) alarm condition, requirements include (1) a quantitative description of observables that trigger an alarm, (2) a quantitative description, including ranges of time, location, and magnitude, of the predicted earthquakes, (3) documented evidence of all previous alarms, (4) a complete list of predicted earthquakes, (5) a complete list of unpredicted earthquakes. The VAN technique [Varotsos and Lazaridou, 1991; Varotsos et al., 1996] has not yet been stated as a testable hypothesis. It fails criteria (1) and (2) so it is not ready to be evaluated properly. Although telegrams were transmitted in advance of claimed successes, these telegrams did not fully specify the predicted events, and all of the published statistical evaluations involve many subjective ex post facto decisions. Lacking a statistically demonstrated relationship to earthquakes, a candidate prediction technique should satisfy several plausibility criteria, including: (1) a reasonable relationship between the location of the candidate precursor and that of the predicted earthquake, (2) some demonstration that the candidate precursory observations are related to stress, strain, or other quantities related to earthquakes, and (3) the existence of co-seismic as well as pre-seismic variations of the candidate precursor. The VAN technique meets none of these.

Architecture:



CHAPTER-9

CHALLENGES FACED

CHALLENGES FACED

- Earthquakes happen at depths that are 10-15km from the earth's surface, or deeper. These are not observable.
- We collected data from observation stations when earthquakes take place and then use inverse modelling to produce results.
- we are hoping that we can find some "precursor" that can be used to predict earthquakes. A precursor is a characteristic pattern of seismic activity or some other physical, chemical or biological change, which would be used as a marker for high probability of an earthquake taking place within days or even months.
- When the stored energy beneath the crust is suddenly released as an earthquake, the crust's response to the changing stress beneath it is not directly proportional. This makes it hard to predict the strength of the earthquake and the behaviour of the crust. This behaviour is dependent on the crust's complex and highly variable and poorly understood geological properties.
- Sometimes the accuracy after training a model is very low.
- After entering the latitude and longitude the model didn't predict the magnitude and will the earthquake will occur or not.
- The model didn't recognize the tectonic areas sometimes. The model showed the same magnitude for some places.
- If an earthquake happens to occur that remotely fits their prediction, they claim success even though one or more of their predicted elements is wildly different from what actually occurred, so it is therefore a failed prediction.
- Unfortunately, most such precursors frequently occur without being followed by an earthquake, so a real prediction is not possible. Instead, if there is a scientific basis, a forecast might be made in probabilistic terms

CHAPTER-10

TESTING

TESTING

Libraries used:

- Pandas
- Sklearn
- Metrics
- Numpy

SOURCE CODE:

```
import pandas as pd
import numpy as np

data=pd.read_csv("C:\\Users\\Vilok Sai Boddapati\\Downloads\\dataset.csv")
data

data.info()

data.isnull()

data.isnull().sum()

X=data[['Latitude','Longitude','Depth']]
Y=data['Magnitude']
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size = 0.2, random_state = 100)

#multi linear regression
from sklearn.linear_model import LinearRegression
mlr=LinearRegression()
mlr.fit(x_train,y_train)
print("Intercept:",mlr.intercept_)
print("coefficients:")
list(zip(X,mlr.coef_))

y_pred_mlr=mlr.predict(x_test)
print(y_pred_mlr)
#y_pred_mlr=mlr.predict([[33,89,45]])
#print(y_pred_mlr)


from sklearn.metrics import mean_squared_error
np.sqrt(mean_squared_error(y_test,y_pred_mlr))

from sklearn.metrics import r2_score
r2_score(y_test,y_pred_mlr)
```

```
from sklearn.metrics import mean_absolute_error as mae
mae(y_test,y_pred_mlr)

#support vector
from sklearn import svm
sv=svm.SVR()
sv.fit(x_train,y_train)
y_pred_svr=sv.predict(x_test)

from sklearn.metrics import mean_squared_error
np.sqrt(mean_squared_error(y_test,y_pred_svr))

from sklearn.metrics import r2_score
r2_score(y_test,y_pred_svr)

from sklearn.metrics import mean_absolute_error as mae
mae(y_test,y_pred_svr)

#random forest regression
from sklearn.ensemble import RandomForestRegressor
rfr=RandomForestRegressor(n_estimators = 100, random_state = 0)
rfr.fit(x_train,y_train)
y_pred_rfr=rfr.predict(x_test)


from sklearn.metrics import mean_squared_error
np.sqrt(mean_squared_error(y_test,y_pred_rfr))

from sklearn.metrics import r2_score
r2_score(y_test,y_pred_rfr)

from sklearn.metrics import mean_absolute_error as mae
mae(y_test,y_pred_rfr)

#decisiontreeregression
from sklearn.tree import DecisionTreeRegressor
dtr=DecisionTreeRegressor(random_state = 0)
dtr.fit(x_train,y_train)
y_pred_dtr=dtr.predict(x_test)

from sklearn.metrics import mean_squared_error
np.sqrt(mean_squared_error(y_test,y_pred_dtr))

from sklearn.metrics import r2_score
r2_score(y_test,y_pred_dtr)

from sklearn.metrics import mean_absolute_error as mae
mae(y_test,y_pred_dtr)


#lasso regression
from sklearn.linear_model import Lasso
las = Lasso(alpha=1)
las.fit(x_train, y_train)
y_pred_las=las.predict(x_test)
```

```
from sklearn.metrics import mean_squared_error
np.sqrt(mean_squared_error(y_test,y_pred_las))

from sklearn.metrics import r2_score
r2_score(y_test,y_pred_las)

from sklearn.metrics import mean_absolute_error as mae
mae(y_test,y_pred_las)
```

```
import pandas as pd
import numpy as np
data=pd.read_csv("C:\\Users\\Vilok Sai Boddapati\\Downloads\\dataset.csv")
data.isnull().sum()
X=data.iloc[:,:-1]
Y=data.iloc[:,-1]
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test =train_test_split(X,Y,random_state=100)
from sklearn.ensemble import RandomForestRegressor
rf=RandomForestRegressor()
rf.fit(x_train,y_train)
y_pred=rf.predict([[18,94,5]])
if(y_pred<4):
   print("no chance for earthquake")
else:
   print("earthquake can occur")
```

```
import matplotlib.pyplot as plt
scores = [mse_mlr,mse_svm,mse_rfr,mse_dtr,mse_las]
algorithms = ["Multi-Linear","Support Vector","Random Forest","Decision Tree","Laso "]
sns.set(rc={'figure.figsize':(8,5)})
plt.xlabel("Algorithms")
plt.ylabel("Accuracy score")
sns.barplot(algorithms,scores)
```

CHAPTER-11

RESULTS

RESULTS

	Latitude	Longitude	Depth	Magnitude
0	29.06	77.42	5.0	2.5
1	19.93	72.92	5.0	2.4
2	31.50	74.37	33.0	3.4
3	28.34	76.23	5.0	3.1
4	27.09	89.97	10.0	2.1
...	...	...	...	...
2714	12.30	94.80	10.0	4.8
2715	24.70	94.30	40.0	4.1
2716	22.50	88.10	10.0	3.6
2717	24.60	94.20	54.0	3.5
2718	14.50	92.90	10.0	4.6

2719 rows × 4 columns

```
count    2719.000000
mean        3.772196
std         0.768076
min         1.500000
25%         3.200000
50%         3.900000
75%         4.300000
max         7.000000
Name: Magnitude, dtype: float64
```

```
array([2.5, 2.4, 3.4, 3.1, 2.1, 5.2, 3. , 5.5, 4.3, 2.6, 4.4, 4.2, 4.9,
       4. , 4.5, 3.2, 3.5, 2.9, 4.8, 2.8, 3.6, 5.3, 4.1, 2.3, 3.8, 3.9,
       3.7, 2.2, 4.6, 4.7, 5.9, 5.1, 3.3, 2.7, 5.4, 5. , 7. , 5.8, 1.9,
       6.6, 6.4, 6. , 5.6, 6.3, 1.8, 1.5, 2. , 6.2, 5.7, 6.1])
```

```
4.3    174
4.2    171
4.0    150
3.6    141
4.1    135
3.9    133
3.8    126
3.5    123
4.5    123
4.4    110
3.7     99
2.8     95
3.4     94
4.6     87
3.1     80
3.2     80
2.9     76
4.7     70
3.3     67
2.6     63
3.0     63
4.8     62
2.7     57
2.5     52
2.4     50
...
1.5      1
6.2      1
5.7      1
Name: Magnitude, dtype: int64
```

```
Latitude     0
Longitude    0
Depth        0
Magnitude    0
dtype: int64
```

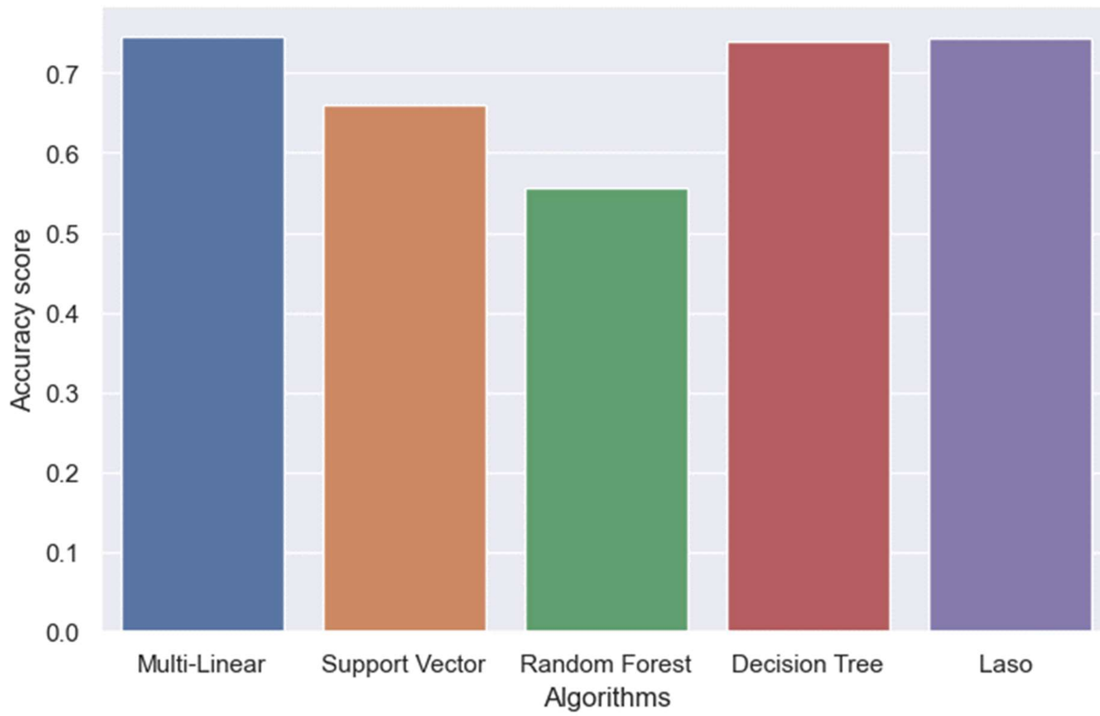
```
Intercept: 3.032452254124305
coefficients:

[('Latitude', 0.009256706024398686),
 ('Longitude', 0.0033837970689275776),
 ('Depth', 0.0035547131236077553)]
```

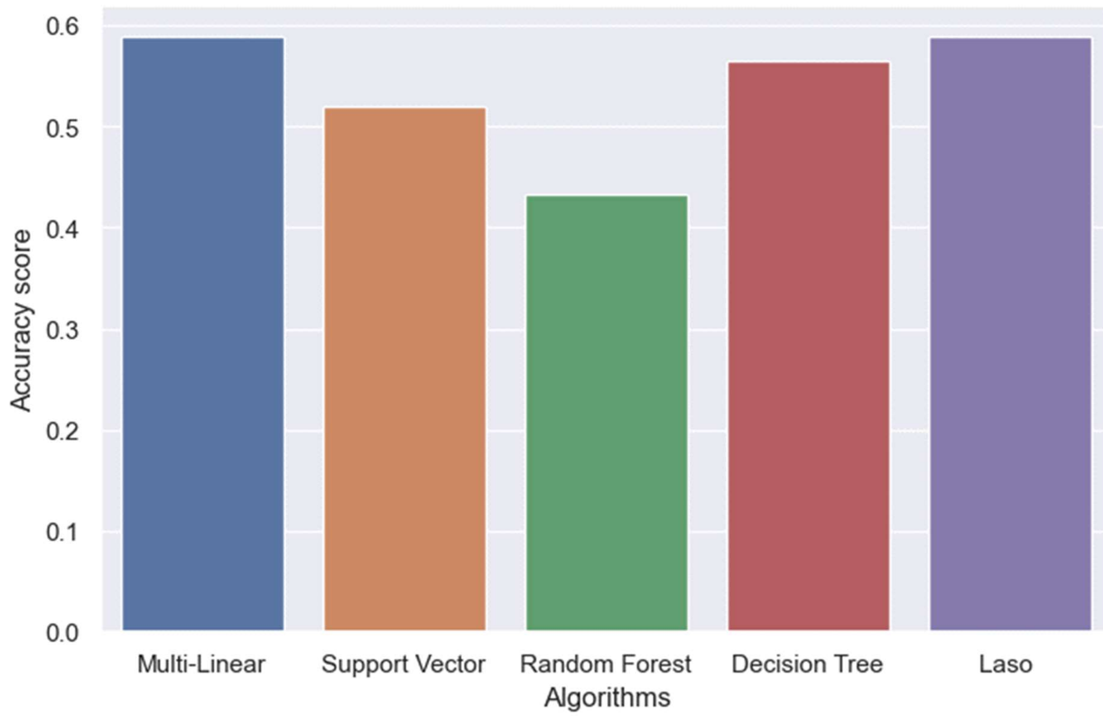
RandomForest regression Predicted values :

```
Output exceeds the size limit. Open the full output data in a text editor
[4.22166667 3.322      4.148      3.777      3.281      4.18533333
 4.243      3.22307143 3.56       4.728      3.929      4.204
 4.192      3.147      2.93793333 2.613      4.01       3.789
 2.852      4.429      3.78061667 4.47       3.639      3.068
 3.211      2.982      3.95       4.198      4.352      2.811
 4.225      3.137      3.326      3.874      4.319      4.34
 3.788      4.043      4.386      3.523      2.932      4.282
 4.191      3.054      3.781      4.143      3.467      4.021
 4.044      2.65186667 4.246      3.745      3.32       4.411
 4.077      2.879      4.216      3.08393333 4.125      4.196
 3.157      3.30857857 2.85       4.31       4.541      4.253
 3.826      3.024      3.985      4.005      2.93       3.737
 4.388      4.408      2.64081429 4.27913333 3.677      4.422
 4.165      3.787      3.371      4.794      3.652      3.15
 4.247      3.653      4.131      2.75       3.868      2.697
 3.062      3.794      2.77776667 3.3        4.421      4.078
 4.534      3.979      3.618      4.014      2.73141667 4.496
 3.962      4.525      3.04486905 4.233      2.714      3.943
 2.743      4.142      3.473      3.558      3.308      2.981
 3.851      3.941      4.458      4.117      3.7655     3.667
 3.55       5.013      4.155      4.237      3.853      3.979
 3.003      3.254      4.392      4.461      2.7905     3.922
 4.374      4.23633333 2.9215     3.737      2.932      4.446
 4.438      3.032      4.289      3.259      4.309      3.351
 3.22307143 3.973      2.969      2.982      2.82       4.117
...
 4.146      2.928      4.575      3.012      3.688      4.421
 3.322      3.933      2.68091667 4.148      4.11266667 2.775
 3.902      3.068      4.286      3.397      4.413      4.3405
 2.816      4.466     ]
```

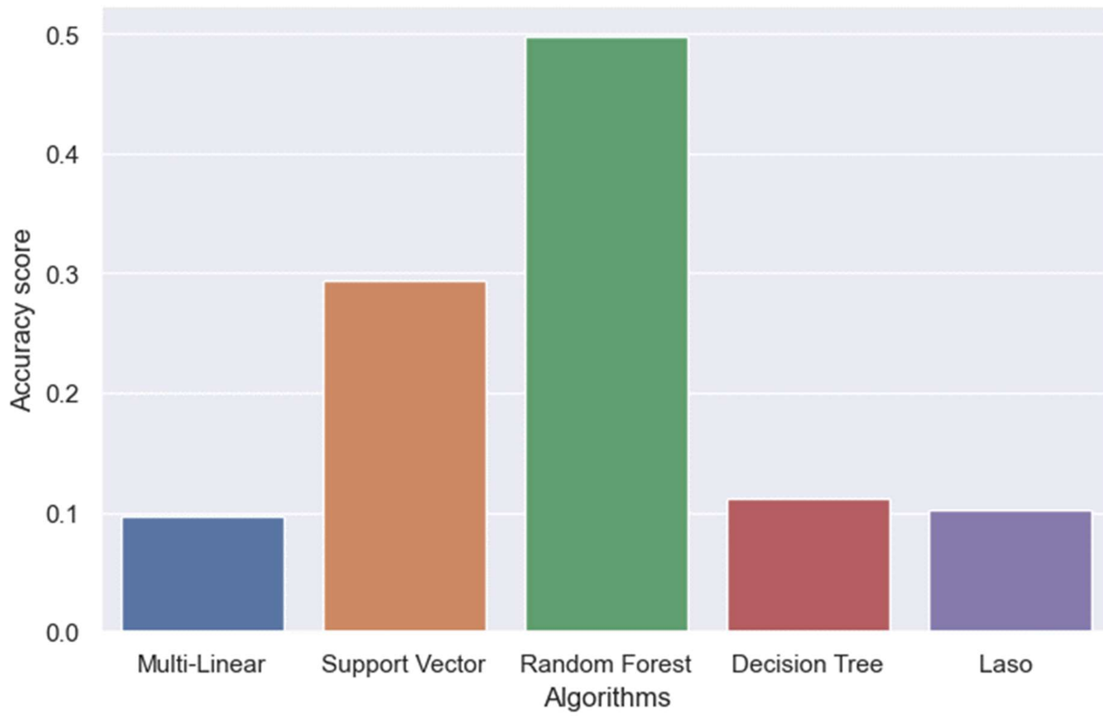
Root mean square error:



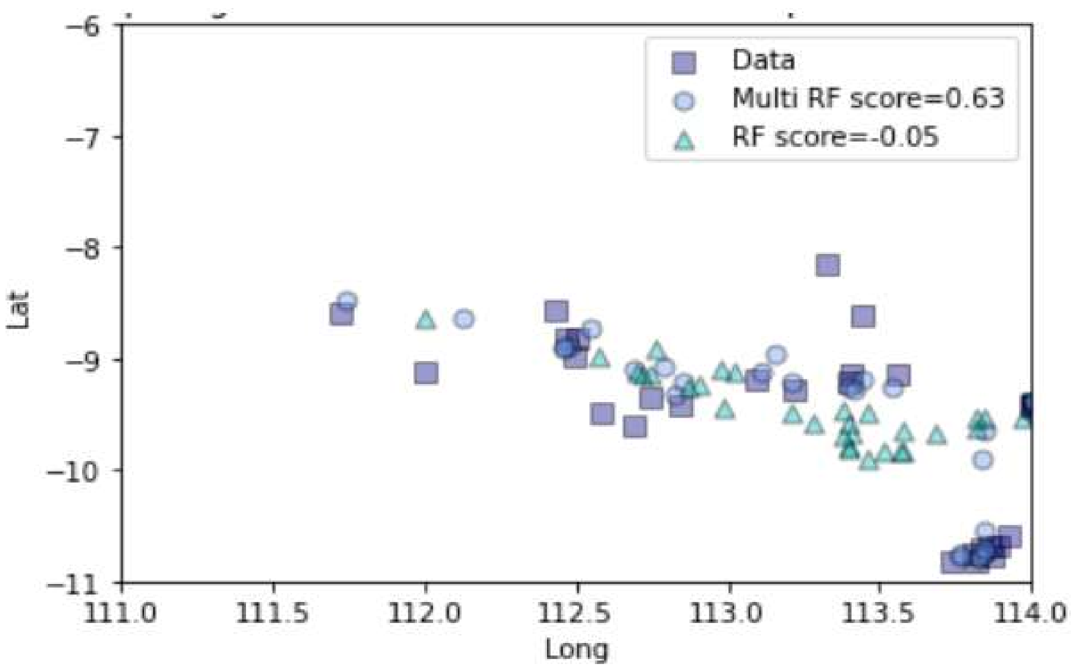
Mean Absolute Error:



R2 score :



Accuracy score
0.0
0.1
0.2
0.3
0.4
0.5
Multi-Linear
Support Vector
Random Forest
Decision Tree
Laso
Algorithms



It can be seen more clearly that the prediction results are close to the data, so it can be said that the prediction results are comparable to accurate. It's just that some predictions are wrong, as in the magnitude at point 11; according to the data, the result is 4.7 SR while the forecast shows 4.03. To overcome this problem, further iterations are needed to obtain the most fantastic accuracy of results.

CHAPTER-12

CONCLUSION

CONCLUSION

As we can see from the above results we have the least error rate and highest R2 score for Random Forest Model. So we will use Random Forest for predicting Earthquake which has highest accuracy rate.

We have proposed a model to predict the earthquake at a location. It is believed that the performance of the model can be improved or the model can give more accurate data if more datasets are available. The model give the results on the basis of data given to it. Thus, forecasting may not be 100% accurate, but it can surely be used as a corrective measure

From the two methods, it can be seen that the accuracy of the two calculations is relatively similar, which is close to 55%. However, the accuracy of manual calculations can be improved by finding the minimum child information gain value from other features such as deep, latitude, and longitude. As for the measures using Google Collab, it can be iterated again to get the highest accuracy

CHAPTER-13

REFERNCES

References

- Geller et al. 1997, p. 1616, following Allen 1976, p. 2070, who in turn followed Wood & Gutenberg 1935.

- Kagan 1997b, p.

- ICEF 2011, p.

- Aceves, Richard L.; Park, Stephen K.; Strauss, David J. (27 May 1996), "Statistical evaluation of the VAN Method using the historic earthquake catalog in Greece", Geophysical Research Letters, **23** (11): 1425–1428, Bibcode:1996GeoRL..23.1425A, doi:10.1029/96GL01478, ISSN 1944-8007.

- Varotsos, P.; Eftaxias, K.; Lazaridou, M. (27 May 1996), "Reply I to "VAN: Candidacy and validation with the latest laws of the game" and "Precursor candidacy and validation: The VAN Case so far"", Geophysical Research Letters, **23** (11): 1331–1334, Bibcode:1996GeoRL..23.1331V, doi:10.1029/96GL01436, ISSN 1944-8007

- https://www.irjet.net/archives/V8/i5/IRJET-V8I5574.pdf

- https://www.liveai.eu/earthquake-prediction-using-deep-learning/

- https://medium.com/marionete/predicting-earthquakes-using-machine-learning-21689435dc52